POLITECNICO DI TORINO

Corso di Laurea in Ingegneria Informatica

Tesi di Laurea Magistrale

Analisi di Dati Sanitari

Analisi delle spese mediche dei contratti pubblici piemontesi

Relatore
prof. Maurizio Morisio

Laureando
Edoardo Emmolo
matricola: 236239

Anno accademico 2018 – 2019

Ringraziamenti

Eccomi arrivato al momento che per tanto tempo mi è sembrato così lontano e inarrivabile e per cui non mi sono mai sentito nemmeno lontanamente preparato. Ho sempre parlato della fine di questo percorso utilizzando sempre il futuro, come se fosse un giorno ancora molto distante prima del quale sarebbero cambiate così tante cose che non avrei potuto nemmeno immaginare la mia situazione arrivati a quel giorno. E così è stato. E voglio ringraziare tutte le persone che mi hanno accompagnato alla fine di questo percorso. Sono persone che direttamente e indirettamente mi hanno messo a disposizione la loro esperienza, il loro sostegno, il loro aiuto aiuto o anche solo uno sguardo o una parola involontaria che per me invece ha significato tantissimo. Tante persone che menzionerò hanno intrapreso una strada diversa dalla mia per scelte personali o meno che posso condividere oppure no, o semplicemente perché è così che doveva andare. Ma il fatto che anche a loro dedicherò un pensiero, vuol dire che hanno contribuito a loro modo a farmi diventare la persona che sono e che ha finito questo progetto di tesi, e per questo vi dico un semplice *grazie* che per me vale *oro*.

Grazie *mamma*, perché sei il pilastro che tiene insieme tutto quello che faccio. Osservi da vicino e da lontano tutto quello che faccio, sei silenziosa quando non c'è niente da aggiungere o quando non è il momento, mentre sei la voce più forte di tutte quando sai che c'è bisogno di te e di quello che hai da dire. Non potrò mai ringraziarti abbastanza, e non lo faccio quanto meriteresti, ma spero nei pochi momenti in cui lo faccio con i fatti o con le parole parole di trasmetterti tutto il bene che ti voglio.

Grazie *papi*, perché da dietro osservi tutto silenzioso, sapendo tutto quello che c'è da sapere, e parlando solo quando hai qualcosa di utile o pratico da dire che fa la differenza nelle situazioni. Non ti ho mai detto grazie per tutto quello che hai fatti, fai e continuerai a fare per noi, perché sono un timidone e non potrei cominciare a parlare senza cominciare a piangere come un bambino. Sai trovare una soluzione a tutto e non hai paura di niente, e spero un giorno di diventare la metà di quello che sei diventato tu. Grazie.

Grazie *Ugi*, il mio fratellino minore che in realtà è il maggiore dei due. Io ti ho insegnato a camminare mentre tu mi hai insegnato a studiare, cosa che prima di andare a vivere insieme a Torino non sapevo fare e infatti avevo dei voti schifosi.

Ogni giorno sei la perfezione e la cosa più giusta in quello che fai, quando c'è da studiare studi, quando c'è da sbrigare sbraghi, sei la cosa migliore che poteva uscire da mamma e papà. Sei un sostegno in ogni situazione, sai dire la tua per ogni cosa e se non conosci di cosa si sta parlando ti ci addentri e mi restituisci il punto di vista che mi mancava. Grazie anche a te, sei fondamentale.
Grazie *Alice*, perché sei la mia metà in ogni cosa che faccio e per ogni cosa che provo. Tesi (che non è stata di coppia ma è come se lo fosse stata), scuola, lavoro, serate, aperitivi, mare, montagna, festa, paranoie, ansia, felicità, rabbia, tristezza. Sai capirmi in ogni situazione, sei sempre pronta a spalleggiarmi, mi difenderesti davanti alla peggior merda mai fatta. Non ti approfitti delle mie debolezze, le comprendi e ti piacciono tutte, e non hai paura di mostrarti per come sei, e io so di poter fare lo stesso. Abbiamo finora condiviso così tante cose di così tanti tipi diversi che non saprei da dove cominciare, neanche tutti gli #ABvideo dell'#ABloop potrebbero descrivere in minima parte quello che siamo. Sai veramente trasformare e farmi vedere tutti i miei lati più deboli nei miei punti di forza. Senza di te la fine del mio periodo universitario non sarebbe stato così il top.
Grazie *Cene Ignoranti*, perché siete dei cazzoni (alcuni dei cazzoni studiati, altri cazzoni che sanno comunque fare i soldi e alcuni cazzoni e basta che però hanno anche dei difetti) che mi fate sognare e mi avete sempre dato la giusta carica per affrontare le settimane di studio impegnative. Cene ignoranti nei locali di spicco (Tapas, Fa Fumme, Grazia RIP, La Fermata etc.), serate in loop, pomeriggi a giocare a calcio tennis al Tapas, partite di serie A domenicali, aperitivi da Mimmo il pomeriggio, e potrei andare avanti per secoli. Per secoli spero che possa durare l'amicizia e l'ignoranza che ci lega.
Grazie *Zanga*, che mi avete fatto imparare a osservare ogni situazione da dietro, che mi avete insegnato che Timur sta dietro ad Amur ribaltando le osservazioni superficiali in cui ricadono i pecoroni a primo impatto. Grazie delle 3-4 pizze a settimana, dei consigli studiati su esami e università, per le risate in palestra, in vacanza. Avete reso Torino un posto meraviglioso. FIEROLOCCHIO.
Grazie Rodolfo, che mi hai sempre mostrato un punto di vista nettamente diverso dal mio e che ha smorzato il mio essere troppo rigido in certe situazioni.
Grazie alla *Dojo Miura*, alla *Red Steel* e all'*Accademia della Boxe* che hanno sempre accompagnato la mia grande passione per gli sport da combattimento durante tutto il mio periodo universitario, facendomi conoscere delle persone che mi hanno fatto ridere, hanno condiviso il ring con me, e poi hanno anche festeggiato vittorie e leccato le ferite insieme a me nelle sconfitte. Un ringraziamento particolare a *Mauri* e *Umbi*, grazie anche a tutti i compagni di allenamento.
Grazie *Giorgio*, l'unico professore diventato anche amico personale, che anche se non ci vediamo tanto so che è sempre a vegliare su di me e qualsiasi cosa accada e per qualsiasi cosa ci sarà sempre. E io per te. Grazie per *tutte le persone* che sono state importanti per me e che ora hanno preso una strada diversa qualsiasi sia la ragione, avete contribuito a farmi diventare la persona che sono oggi facendomi

ridere, piangere, buttandomi giù il morale e tirandomi su nei momenti di debolezza e mettendo a disposizione tutta la vostra esperienza.
Grazie ai gestionali della triennale (*Citta, Simo Albe, Albi Disto, Impe*) che anche se hanno preso un percorso diverso dal mio dimostrano come anche le persone conosciute da quasi "adulti" siano persone in grado di rimanere nel tempo e dei grandi amici, ignoranti, professionali con cui smazzare i risultati e traguardi scolastici e non, e condividere feste in montagna e al mare.
Grazie a *Simo Albe* per l'infinta esperienza su ogni sport o interesse che mi è mai venuto in mente, su cui mi ha spalleggiato e istruito dal principio.
Grazie a *Rox*, una delle poche ragazze che posso considerare mie amiche che anche se non ci sentiamo tanto ultimamente (sorry :D) rimane sempre una persona su cui potrò fare sempre affidamento e viceversa.
Grazie al *TeamF* che mi accompagna dal primo giorno di università con serate a qualsiasi tema, sessioni di studio, festoni di laurea e chi più ne ha più ne metta. Anche voi mi dimostrate come sia bello cambiare città e trovare dei punti fissi su cui poter contare sempre.
Grazie a *Stefano*, che con i suoi consigli mi ha permesso di conoscermi meglio come mai prima d'ora e trasformare le mie debolezze nei maggiori punti di forma.
Grazie a tutti *quelli che non sto menzionando* in questi ringraziamenti, scusate ma scrivo tutto di getto come mio solito,
Un ultimo ringraziamento speciale *a chi mi ha sempre voluto male, a chi mi ha messo i bastoni fra le ruote, a chi ha cercato di complicarmi la vita a scuola e fuori da essa* perché è proprio grazie a voi che ho superato le sfide più difficili e complicate che mi hanno fatto superare i miei limiti e fare il salto di qualità, il traguardo di cui vado più fiero da quando ho iniziato i miei studi a Torino.

Indice

Elenco delle tabelle

Elenco delle figure

Capitolo 1

Introduzione generale

Questo progetto di tesi è stato svolto presso Synapta s.r.l., un'azienda nata a Torino nel 2016 che si è occupata fino ad oggi di analisi di dati sopratutto in ambito di Linked Data usando tecnologie open source e innovative nel settore. Uno dei maggiori lavori dell'azienda è ContrattiPubblici.org [1] : un database che raccoglie tutti i dati dei contratti pubblici resi disponibili online dalle pubbliche amministrazioni italiane a partire dal 2013.

Il mio ruolo in questo panorama è stato quello di analizzare un sottoinsieme di tutti i contratti pubblici, in particolare, si concentra sull'analisi delle spese pubbliche di un sottoinsieme di questi documenti, più precisamente quelli che coinvolgono tutte le strutture appaltanti che hanno sede nella regione Piemonte; l'obiettivo principale è quello di evidenziare in quali categorie verte la maggior parte della spesa e quali strutture sanitarie spendono di più.

Legge 6 novembre 2012, n. 190 Per inquadrare l'origine dei dati usati in questo progetto, si rimanda alla legge n.190 del 6 novembre 2012 che ha per la prima volta imposto alle Pubbliche Amministrazioni di pubblicare queste informazioni. In questo paragrafo ci concentreremo solo sugli aspetti della legge di maggior interesse ai fini di capire il contesto delle analisi svolte.

L'articolo 1 comma 32 della legge 190 / 2012 [2] prevede per le Pubbliche Amministrazioni, in seguito indicate in questo testo come testo PA, che queste «*sono in ogni caso tenute a pubblicare nei propri siti web istituzionali: la struttura proponente; l'oggetto del bando; l'elenco degli operatori invitati a presentare offerte; l'aggiudicatario; l'importo di aggiudicazione; i tempi di completamento dell'opera, servizio o fornitura; l'importo delle somme liquidate.*», e che poi queste informazioni devono essere «*pubblicate in tabelle riassuntive rese liberamente scaricabili in un formato digitale standard che consenta di analizzare e rielaborare, anche a fini statistici, i dati informatici.*» Questi dati sono pubblicati come file XML che seguono le specifiche dello schema descrittivo XSD approvato dalla legge redatta

dall'Autorità Nazionale Anticorruzione (ANAC) [3]. I dati sono pubblicati in seguito nelle sezioni "Amministrazione Trasparente" nel sito ufficiale di ogni pubblica amministrazione. Grazie alla licenza CC BY 3.0 [4] applicata a questi dati, essi sono scaricabili e consultabili a scopo personale e di analisi.

contrattipubblici.org La realizzazione di questa base dati, che include tutti i contratti delle Pubbliche Amministrazioni italiane, è stata portata a termine in due passi successivi. Innanzitutto sono stati elaborati dei programmi specifici che visitassero e scaricassero in modo automatico da tutte le sezioni "Amministrazione Trasparente" di tutte le PA fino ad ora incluse nella ricerca, i file XML delle gare d'appalto. Poi, dopo averli raccolti, sono stati strutturati, indicizzati, puliti e connessi tra di loro e infine resi consultabili sul portale web. Per integrare informazioni non presenti sono state ricercate anche altre fonti dati tra cui la Gazzetta Ufficiale[**?**] per recuperare anche gli allegati ai bandi che riguardano: il dirigente responsabile del procedimento che indica lo stato in cui si trova la gara, lo stato di aggiudicazione della gara con l'ente aggiudicatario, il bando di gara, l'elenco di tutti i lotti, il verbale delle assemblee dalle quali risultano le domande di partecipazione al bando di gara, chi è ammesso e chi invece non è ammesso perché i requisiti non sono rispettati.

Classificazione delle spese sanitarie. Per operare un'analisi critica sulla distribuzione delle spese sanitarie della regione Piemonte, serve classificare tutti gli appalti in un insieme di categorie merceologiche, e quindi trovare un vocabolario che le contenga tutte adatto all'analisi da svolgere. In ambito ospedaliero si sostengono generi di spese molto eterogenei tra di loro, quali i lavori di edilizia, spese per medicinali e apparecchiature mediche, arrivando fino alle spese alimentari e a quelle per la pulizia.

Organizzazione della sanità piemontese. L'Azienda Sanitaria Locale (A.S.L.) è un ente dotato di autonomia imprenditoriale e personalità giuridica pubblica che opera nel servizio sanitario nazionale (S.S.N.) (art. 3, comma 1 bis, del Decreto Legislativo 30 dicembre 1992, n. 502, come modificato dal Decreto Legislativo 19 giugno 1999, n. 229 [6]). Il servizio sanitario nazionale affianca alle strutture locali ASL anche altre strutture chiamate Aziende Ospedaliere. Un' Azienda Ospedaliera, in Italia, è una struttura di ricovero e assistenza pubblica che fa anch'essa parte del servizio sanitario nazionale. Essa svolge la funzione di ospedale ed è adibita anche a prestazioni specialistiche. In particolare assicurano attività sanitaria di specializzazione con dotazioni di tecnologie diagnostico-terapeutiche avanzate ed innovative e svolgono i compiti specificamente attribuiti dagli atti della programmazione regionale.
Ogni regione italiana è suddivisa in zone servite da Aziende Sanitarie Locali (A.S.L.), che fanno parte del Servizio Sanitario Nazionale (S.S.N.). Il Piemonte in particolare

è suddiviso in dodici ASL: l'ASL di Torino 1, Torino 2, Torino 3, Torino 4 e Torino 5 (abbreviate rispettivamente ASL TO1, TO2, TO3, TO4, TO5), l'ASL di Vercelli (ASL VC), quella di Biella (ASL BI), quella di Novara (ASL NO), quella del Verbano Cusio Ossola (ASL VCO), quella di Cuneo 1 (ASL CN1), quella di Cuneo 2 (ASL CN2 di Alba-Bra), quella di Asti (ASL AT) e infine quella di Alessandria (ASL AL) [7].

Nei capitoli successivi, il testo presenta lo studio sulle spese sanitarie piemontesi nella sua interezza; a partire dallo studio del dataset e della successiva classificazione degli appalti in categorie merceologiche, si prosegue quindi con un'attenta analisi dei dati per poi illustrare le conclusioni ottenute.

Capitolo 2

Tecnologie adottate

In questo capitolo verranno introdotte e spiegate tutte le applicazioni, i framework, i formati e i linguaggi di interrogazione utilizzati per rappresentare, manipolare i dati e infine portare a compimento le analisi.

2.1 Formati di rappresentazione dei dati

2.1.1 XML

XML (acronimo di EXtensible Markup Language) è un linguaggio che consente la scrittura e rappresentazione di documenti e di dati strutturati sui supporti digitali. Questo formato ha una sintassi insieme rigorosa e flessibile che lo rende perfetto per rappresentare dati di grande complessità. Lo standard XML è stato sviluppato dal W3C e la sua prima pubblicazione è stata nel 1998. I documenti scritti in formato XML hanno una struttura gerarchica composta da elementi e attributi.Tutti gli elementi sono organizzati un una struttura ad albero con una radice singola. Un elemento rappresenta un componente logico del documento, mentre un attributo è un'informazione descrittiva che può essere associato ad un elemento. Riportiamo un esempio:

```
<utenti>
        <utente anni="20">
                <nome> Emanuele </nome>
                <cognome> Principe </cognome>
                <indirizzo> Torino </indirizzo>
        </utente>
        <utente anni="24">
                <nome> Giorgio </nome>
                <cognome> Scala </cognome>
                <indirizzo> Milano </indirizzo>
        </utente>
</utenti>
```

utenti è l'elemento radice, detto anche *root*, il quale contiene un array di elementi *utente*; ogni *utente* è formato dagli elementi *nome*, *cognome* e *indirizzo* i quali sono ad un livello gerarchicamente inferiore a *utente*, e hanno rispettivamente i valori *Emanuele*, *Principe* e *Torino* il primo utente, *Giorgio*, *Scala* e *Milano* il secondo utente . Gli attributi nell'esempio sono *anni*, che hanno come valori il numero *20* e il numero *24*.

2.1.2 JSON

JSON (acronimo di Javascript-Oriented Notation) è un formato simile a XML che permette di descrivere dati strutturati e scambiare dati in rete, ma più leggero. I dati sono rappresentati come coppie nome/valore. Rappresentiamo lo stesso esempio XML in formato JSON:

```
{
"utenti": [
   {
   "utente": {
        "anni": 20 {
             "nome": "Emanuele",
             "cognome": "Principe",
             "indirizzo": "Torino"
              }
        }
     },
     "utente": {
        "anni": 24 {
             "nome": "Giorgio",
             "cognome": "Scala",
             "indirizzo": "Milano"
              }
        }
}
```

La sintassi cambia rispetto a quella del formato xml: non abbiamo tag di apertura e di chiusura, la coppia elemento - valore è separata dal carattere *due punti* (':'), e il nome degli elementi e dei valori è racchiuso tra doppi apici. Cambia anche la rappresentazione degli array; un elemento che contiene un array di altri elementi è strutturato come nell'esempio: dopo il nome dell'elemento viene aperta una parentesi quadra e poi una graffa, e all'interno di quest'ultima vengono scritti gli elementi che costituiscono il vettore uno dopo l'altro, separati da una virgola.

2.1.3 RDF

RDF è una tecnologia proposta dal W3C per codificare, scambiare e riutilizzare metadati. Consente inoltre interoperabilità tra le applicazioni che condividono

informazioni sul Web. Questo standard è costituito da due diversi componenti:

1. l'*RDF model and Syntax* che espone la struttura del modello pdf e ne descrive una possibile sintassi
2. l'*RDF Schema* invece espone la sintassi per definire schermi e vocabolari per i metadati

RDF Data Model L'RDF Data Model si basa su tre principi chiave:

1. qualunque cosa/entità può essere identificata da una URI
2. the *least-power*, ovvero utilizzare il linguaggio meno espressivo per definire qualunque cosa
3. qualunque cosa può dire qualunque cosa su qualunque cosa

Un modello RDF è rappresentabile da un grafo orientato i cui nodi rappresentano risorse o tipi primitivi, mentre gli archi rappresentano proprietà che esistono fra diverse risorse (nodi). Un grafo RDF è serializzabile in diversi modi:

- RDF/XML
- N-Triples: il grafo viene serializzato come un insieme di triple soggetto - predicato - oggetto
- Notation3: descrizione una per volta ogni risorsa insieme tutte le sue proprietà

RDF Schema Uno schema RDF (RDFS) è un insieme di classi e proprietà rdf per creare un vocabolario che estende il vocabolario di base RDF e che descrive in genere concetti più complessi a partire da semplici concetti di base

2.2 Linguaggi di interrogazione

2.2.1 SPARQL

SPARQL è l'acronimo ricorsivo di *SPARQL Protocol and RDF Query Language*. Si tratta di un linguaggio di interrogazione per dati rappresentati secondo lo standard RDF. Sono definiti due tipi di interrogazioni possibili:

1. le interrogazioni di tipo *select*: esse permettono di selezionare dei dati da un database RDF scegliendo delle condizioni e manipolando i contenuti dei nodi. Questo tipo di interrogazioni restituiscono i dati selezionati in formato tabulare;

2. Le interrogazioni di tipo *construct*: a differenza delle precedenti, queste interrogazioni restituiscono un insieme di dati in formato RDF. Permettono anch'esse di applicare delle condizioni per filtrare il set di nodi e archi di interesse e manipolare i dati racchiusi in essi, cambiando il loro formato o il loro valore.

L' interrogazione di tipo *construct* è quella usata insieme al tool JARQL per trasformare un file JSON in un file RDF, mentre quella di tipo *select* è quella che viene usata invece per interrogare il database a grafo su cui verranno caricati tutti i file XML trasformati in file RDF.

Riportiamo di seguito una query di tipo SELECT che commenteremo in modo tale da avere un'interrogazione spiegata a cui rifarsi per capire le query che verranno trattate nelle sezioni successive:

```
1 PREFIX : <http://example.com/>
2 SELECT ?title
3 WHERE {
4         <http://example.org/book/book1> :title ?title
5         }
```

Nella riga 1 troviamo la parola chiave *PREFIX* che permette di abbreviare la URI a cui fa riferimento in modo da non dover scrivere il path completo nelle parti seguenti della query. Nella seconda riga troviamo la keyword *SELECT* a cui seguono le variabili restituite in formato tabulare dall'interrogazione. Nella riga 3 troviamo invece la parola chiave *WHERE* che è seguita nelle righe 4 e 5 dal pattern dei nodi e degli archi del grafo a cui siamo interessati per far ritornare il risultato desiderato. Questo pattern lo scriviamo come *lista di triple* dato che ci stiamo riferendo ad una struttura dati a grafo.

2.2.2 Ontologia

Questo termine letteralmente significa lo studio dell'essere in quanto tale, e delle sue categorie fondamentali, ma adattato ai contratti pubblici prende il significato di modo di descrivere in formato RDF ogni file relativo a un contratto pubblico riguardante la pubblicazione della legge 190/2012. La query di tipo *construct* che sarà mostrata nelle successive sezioni è la nostra ontologia per trasformare in formato RDF i contratti pubblici.

2.3 Applicazioni utilizzate

In questa sezione verranno approfondite le applicazioni utilizzate per compiere le analisi e le elaborazioni sui dati.

2.3.1 JARQL

JARQL è un progetto open source che permette di eseguire le query SPARQL di tipo *construct* su dei file in formato JSON. Il progetto JARQL si è ispirato da TARQL che ha una funzione simile, ma come oggetto della trasformazione ha un file in formato CSV. Con JARQL quindi è possibile trasformare semplicemente dei file scritti in formato JSON in file con formato RDF. Perciò, prima si è reso necessario tradurre i dato dal formato xml al formato JSON, dopodiché tramite l'utilizzo di JARQL e di un'apposita query scritta in linguaggio SPARQL si è arrivati ad avere lo stesso documento iniziale scritto come un insieme di triple in formato RDF. Vediamo un esempio della trasformazione di un file di esempio scritto in formato JSON (*topolino.json*) mediante una query SPARQL (*topolino_query.sparql*) in un file in formato RDF (*topolino.ttl*).

topolino.json questo è il file di partenza in formato JSON:

```
{
   "parent": {
             "name": "Topolina",
             "children": ["Topolino1", "Topolino2", "Topolino3"],
             "fiance": {"name": "Topolona"}
         },{
             "name": "Topolino",
              "children": ["Tip", "Tap", "Top"],
              "fiance": {"name": "Minnie"}
        } ]
}
```

topolino_query.sparql invece questa la query di tipo *construct*, o *ontologia*, per trasformare *topolino.json* in *topolino.ttl*:

```
PREFIX : <http://example.com/>
PREFIX jarql: <http://jarql.com/>

CONSTRUCT {
     ?p :name ?n;
        :child [:name ?cn].
}
WHERE {
     jarql:root jarql:parent ?p.
     ?p jarql:name ?n.
     ?p jarql:children ?cn.
}
```

topolino.ttl eseguendo il comando : *java -jar jarql-1.0-pre1.jar paperino.json paperino.query > paperino.ttl* otteniamo il nostro file finale in turtle:

```
[ <http://example.com/child>  [ <http://example.com/name>
                    "Topolino1"^^<http://www.w3.org/2001/XMLSchema#
                        string> ] ;
  <http://example.com/child>  [ <http://example.com/name>
                    "Topolino2"^^<http://www.w3.org/2001/XMLSchema#
                        string> ] ;
  <http://example.com/child>  [ <http://example.com/name>
                    "Topolino3"^^<http://www.w3.org/2001/XMLSchema#
                        string> ] ;
  <http://example.com/name>   "Topolina"^^<http://www.w3.org/2001/
      XMLSchema#string>
] .

[ <http://example.com/child>  [ <http://example.com/name>
                    "Top"^^<http://www.w3.org/2001/XMLSchema#string> ]
                        ;
  <http://example.com/child>  [ <http://example.com/name>
                    "Tap"^^<http://www.w3.org/2001/XMLSchema#string> ]
                        ;
  <http://example.com/child>  [ <http://example.com/name>
                    "Tip"^^<http://www.w3.org/2001/XMLSchema#string> ]
                        ;
  <http://example.com/name>   "Topolino"^^<http://www.w3.org/2001/
      XMLSchema#string>
] .
```

JARQL presenta qualche limitazione come riportato in seguito, che però non è stata rilevante per l'uso che ne è stato fatto:

1. non supporta la traduzione di array annidati nei file di input JSON
2. l'ordine degli array del file JSON è irrilevante

2.3.2 Blazegraph Database

I database sono insiemi strutturati di dati persistenti, di cui esistono una grande varietà di tipologie: relazionale, orientati ai documenti, strutturati, non relazionali e a grafo. In questo progetto di tesi i dati trattati sono stati trasformati in formato RDF, perciò per memorizzarli si è adottato un database a grafo. I database a grafo utilizza nodi e archi per rappresentare ed archiviare l'informazione. Rispetto alle altre tipologie, i database a grafo sono spesso più veloci nell'associazione di insiemi di dati e mappano in maniera più diretta le strutture di applicazioni che fanno uso di oggetti. In aggiunta, questi database sono più scalabili e non richiedono costose operazioni di join come invece accade per i database di tipo relazionale. In questo progetto di tesi è stato scelto come database a grafo Blazegraph, che supporta un

RDF/SPARQL APIs. E' capace di gestire fino a 50 miliardi di archi su un singolo calcolatore, prestazioni più che sufficienti per la mole di dati che si è dovuto gestire. Inoltre supporta un API REST, che si è resa molto utile per interrogare il database da terminale e poter scaricare i risultati delle interrogazioni in locale per poter essere manipolati.

2.3.3 JAXB Java Framework

Per analizzare tutti i lotti del dataset e per in seguito trasformarli in altri formati, si è scelto di costruire applicazioni in tecnologia Java sfruttando tutte le potenzialità del framework JAXB (acronimo di Java Architecture for XML Binding).
L'Extensible Markup Language (XML) e la tecnologia Java sono tecnologie spesso usate insieme nel supporto allo scambio di dati in Internet, per realizzare web services e relative applicazioni d'accesso.
JAXB mette a disposizione un insieme di API per semplificare la lettura e la scrittura di documenti XML attraverso applicazioni scritte direttamente in Java. Dalla versione 1.6, JAXB è già incluso in Java SE. Per accedere ai documenti XML, questo framework dispone il binding del corrispondente XML Schema. Questo significa generare un insieme di classi Java che rappresentano questo schema, che mette in relazione gerarchica tutti gli elementi che costituiscono il file XML che si vuole parsificare. Una volta redatto, si procede alla generazione automatica delle classi JAXB, che sono classi Java Beans dotate di metodi getters e setters. Tramite la classe ObjectFactory i Java Beans per generare degli oggetti Java in cui memorizzare e poi manipolare tutto il contenuto di ogni documento XML. Il JAXBContext è l'oggetto che fornisce l'entry point per le API JAXB. Occorre specificare il contesto, che rappresenta una lista di nomi dei package che contengono le interfacce generate da un binding compiler. Se gli schemi utilizzati sono ben strutturati, è sufficiente passare il nome del package contenente la classe contenitore del documento come vedremo successivamente. Il framework JAXB permette di fare due principali operazioni: l'unmarshalling e il marshalling di un documento.

Unmarshalling Effettuare l'unmarshalling di un documento significa creare un albero di oggetti Java (istanze delle classi JAXB precedentemente create dallo schema) che rappresentano il contenuto e l'organizzazione del documento XML. Una volta completato, si ha a disposizione l'elemento root del documento target dell'operazione di unmarshalling, dal quale è possibile accedere a tutti gli altri sottoelementi e ai relativi attributi.

Marshalling L'operazione inversa alla precedente prende il nome di Marshalling, grazie alla quale è possibile generare un file XML a partire da un oggetto root delle classi JAXB opportunamente riempito di informazioni.

2.3.4 Classificatore CPV

I programmatori di Synapta avevano sviluppato un'applicazione web chiamata *classificatore CPV* che sfrutta un albero decisionale per categorizzare un lotto a partire da 3 dati di input: la stringa dell'elemento "oggetto" del documento xml relativo all'appalto in questione, il codice fiscale della stazione appaltante e l'importo concordato per la realizzazione del servizio e/o opera. Questa operazione di categorizzazione produce come risultato una categoria merceologica scelta a partire da un vocabolario. Il vocabolario a cui il classificatore CPV Synapta si chiama appunto CPV, acronimo di *vocabolario comune per gli appalti pubblici*.

2.4 Linked Data

I linked data sono una modalità di pubblicazione di dati strutturati che facilita a essere collegati tra di loro, e quindi ad essere utilizzabili attraverso interrogazioni semantiche. I linked data si basano su tecnologie e standard web open come HTTP e URI, e ne estendono l'applicazione per fornire ulteriori informazioni machine-readable. Uno dei grossi vantaggi dei linked data è il fatto che diventa possibile collegare e utilizzare dati provenienti da diverse sorgenti. I linked data si basano su 4 criteri fondamentali:

1. Ogni oggetto è identificato da una URI
2. Questi oggetti possono essere referenziati e cercati da persone e user-agent usando HTTP URI
3. Viene adottato lo standard RDF come formato di rappresentazione dei dati
4. Includere link ad altre URI relative ai dati esposti per migliorare la ricerca di altre informazioni relative nel Web

2.5 Distribuzione cumulativa

La distribuzione cumulata permette di osservare la velocità di crescita di una data variabile in funzione di un'altra rispetto alla quale se ne stabilisce un ordinamento. Ogni punto del grafico rappresenta la frequenza assoluta di quello specifico elemento sommata alla frequenza cumulata di tutti gli elementi che lo precedono.

2.6 Struttura dei documenti pubblicati dalle pubbliche amministrazioni

Ogni documento XML è pubblicato da uno specifico ente (ad esempio ASL di Torino 3) in fase di adempimento della legge 190/2012 e contiene un array di lotti che si

riferiscono tutti all'ente stesso. Ogni *lotto* identifica l'aggiudicazione di una gara d'appalto. Ogni lotto è identificato da un codice univoco chiamato CIG (acronimo di *codice identificativo gara*), ha una *struttura proponente* (azienda appaltante), un array di *partecipanti* (tutte le aziende che partecipano alla gara per l'acquisizione dell'appalto), un altro array di *aggiudicatari* (azienda o aziende appaltatrici). Infine ogni lotto contiene un elemento che tiene traccia della data di inizio e la data di fine dell'opera o servizi come definito dall'appalto, e un altro elemento formato da due importi: il primo definisce il totale concordato per il compimento dell'opera, chiamato *importo aggiudicazione*, e il secondo che indica la somma cumulativa di quanto è già stato pagato rispetto all'importo nominale, chiamato *importo somme liquidate*.
L'elemento *legge190:pubblicazione* è l'elemento root di ogni documento xml. Questo elemento si articola in due sotto elementi, *metadata* e *data*. Il primo sotto elemento contiene tutte le informazioni comuni a tutti i lotti, mentre il secondo elemento contiene l'array di tutti i lotti.
L'elemento *metadata* si articola nei seguenti sottoelementi:

- *titolo* : stringa che descrive il titolo della pubblicazione
- *abstract* : contiene una stringa che rappresenta una breve descrizione della pubblicazione
- *dataPubblicazioneDataset* : stringa contenente la data in cui questo documento è stato pubblicato
- *entePubblicatore* : ente che ha pubblicato questo documento xml
- *dataUltimoAggiornamentoDataset* : data dell'ultima modifica della pubblicazione del documento (spesso coincide con *dataPubblicazioneDataset*)
- *annoRiferimento* : anno in cui si riferisce il documento e i suoi lotti pubblicati
- *urlFile* : indirizzo internet in cui è pubblicato il dataset
- *licenza* : tipo di licenza che è applicata al documento

L'elemento *data* contiene un numero variabile di elementi *lotto*. Ogni elemento *lotto* è formato da altri sottoelementi:

- *cig*, stringa alfanumerica di 10 cifre che rappresenta il Codice Identificativo Gara del lotto in questione;
- *strutturaProponente*, elemento che contiene due elementi al suo interno, il primo sotto-elemento *codiceFiscaleProp* in cui è scritto il codice fiscale dell'azienda appaltante e il secondo *denominazione* dove è presente la ragione sociale dell'azienda appaltante;

- *oggetto*, elemento che contiene la descrizione della prestazione richiesta dall'appalto, è questa la stringa che verrà in seguito classificata in una categoria;
- *sceltaContraente*, stringa che descrive il tipo di contratto adottato (un glossario di tutte le scelte contraente è riportato su www.contrattipubblici.org);
- *partecipanti*, elemento che contiene un array formato da un numero variabili di elementi *partecipante*, ogni elemento *partecipante* indica un'azienda che ha preso parte alla gara di questo specifico appalto (lotto);
- *aggiudicatari*, elemento che contiene un array formato da un numero variabile di elementi *aggiudicatario*;
- *tempiCompletamento*, elemento che contiene due sottoelementi relativi alla data di inizio e la data di fine dell'appalto (*dataInizio* e *dataUltimazione*, entrambe sono due stringhe);
- *importoAggiudicazione*, stringa che indica l'importo nominale della prestazione definita dall'appalto;
- *importoSommeLiquidate*, stringa che definisce un importo minore o uguale all'importo aggiudicazione, in cui è indicato la somma di denaro pagata rispetto all'importo nominale.

Un esempio completo di un documento XML riportato in seguito:

```
<?xml version="1.0" encoding="utf-8"?>
<legge190:pubblicazione>
  <metadata>
    <titolo>GENERATORE DI CALORE</titolo>
    <abstract>GENERATORE DI CALORE</abstract>
    <dataPubbicazioneDataset>2018-04-03</dataPubbicazioneDataset>
    <entePubblicatore>ISTITUTO GAUDENZIO DE PAGAVE</entePubblicatore>
    <dataUltimoAggiornamentoDataset>2018-04-03</
        dataUltimoAggiornamentoDataset>
    <annoRiferimento>2013</annoRiferimento>
    <urlFile>http://anac.robyone.net/xml/Dataset.aspx?cid=177&amp;year
        =2014&amp;id=19412</urlFile>
    <licenza xsi:type="xsd:string">IODL</licenza>
  </metadata>
  <data>
    <lotto>
      <cig>Z0E0D24B6E</cig>
      <strutturaProponente>
        <codiceFiscaleProp>00429870033</codiceFiscaleProp>
        <denominazione>ISTITUTO GAUDENZIO DE PAGAVE</denominazione>
      </strutturaProponente>
      <oggetto>GENERATORE DI CALORE</oggetto>
      <sceltaContraente>08-AFFIDAMENTO IN ECONOMIA</sceltaContraente>
```

```
      <partecipanti>
        <partecipante>
          <codiceFiscale>00499970036</codiceFiscale>
          <ragioneSociale>DUOTERMICA SRL</ragioneSociale>
        </partecipante>
        <partecipante>
          <codiceFiscale>03264450168</codiceFiscale>
          <ragioneSociale>HOVAL SRL</ragioneSociale>
        </partecipante>
      </partecipanti>
      <aggiudicatari>
        <aggiudicatario>
          <codiceFiscale>00499970036</codiceFiscale>
          <ragioneSociale>DUOTERMICA SRL</ragioneSociale>
        </aggiudicatario>
      </aggiudicatari>
      <importoAggiudicazione>20002.00</importoAggiudicazione>
      <tempiCompletamento>
        <dataInizio>2013-12-16</dataInizio>
        <dataUltimazione>2014-01-31</dataUltimazione>
      </tempiCompletamento>
      <importoSommeLiquidate>18124.00</importoSommeLiquidate>
    </lotto>
  </data>
</legge190:pubblicazione>
```

Capitolo 3

Stato dell'arte

Come già anticipato nel capitolo introduttivo, per analizzare le spese è necessario inserire ogni lotto in una categoria merceologica, in modo tale da poter dividere la spesa totale per categorie oltre che per struttura appaltante o appaltatrice.
Questo passaggio fa emergere due necessità: la prima è quella di trovare un modo per classificare ogni lotto analizzato, mentre la seconda è quella di avere un insieme di categorie in cui inserire ogni lotto del dataset.
Nella ricerca della soluzione più adatta si è cercato in letteratura l'esistenza di opere in cui si sono trovate soluzioni a problemi simili, o se in azienda si fosse sviluppata una soluzione a questa problematica, in quanto i contratti pubblici sono una parte importante di tutti i progetti attualmente lavorano.

3.1 Vocabolari merceologici esistenti

In questa sezione analizziamo i vocabolari valutati per cercare un insieme di categorie da usare per classificare gli appalti del dataset.

3.1.1 Vocabolario CPV

Il CPV è un sistema di classificazione unico per gli appalti pubblici con lo scopo di unificare i riferimenti utilizzati dalle amministrazioni e dagli enti appaltanti per la descrizione dell'oggetto degli appalti.
Il vocabolario comune per gli appalti pubblici (il cui acronimo è CPV), è in vigore dal 17.09.2008 ed è stato adottato dal regolamento (CE)n. 213/2008. Il CPV comprende un vocabolario principale per la descrizione dell'oggetto degli appalti pubblici e un vocabolario supplementare per aggiungere informazioni a livello qualitativo all'oggetto. Il vocabolario principale è composto da una struttura ad albero di codici composti da 9 cifre (un codice di 8 cifre più una di controllo), ai quali corrisponde una descrizione delle forniture, i lavori o servizi, oggetto del contratto.

Di seguito riportato un esempio di ramificazione verticale di un albero di codici reale. Maggiore è il numero di cifre maggiore è il livello di dettaglio della categoria merceologica:

- 30000000-9 : Macchine per ufficio ed elaboratori elettronici, attrezzature e forniture, esclusi i mobili e i pacchetti software
- 33000000-0 : Apparecchiature mediche, prodotti farmaceutici e per la cura personale
- 33100000-1 : Apparecchiature mediche
- 33130000-0 : Strumenti e dispositivi odontoiatrici e di sottospecialità
- 33131000-7 : Strumenti portatili odontoiatrici
- 33131100-8 : Strumenti chirurgici odontoiatrici
- 33131110-1 : Tenaglie, spazzole, divaricatori e brunitori odontoiatrici
- 33131112-5 : Spazzole odontoiatriche operative

Ad ogni livello è possibile trovare anche ramificazioni di tipo orizzontale, che esplorano tutte le diverse categorie merceologiche che si trovano ad uno stesso livello con la stessa radice di codice:

- 33131111-8 : Tenaglie odontoiatriche
- 33131112-5 : Spazzole odontoiatriche operative
- 33131113-2 : Divaricatori odontoiatrici
- 33131114-9 : Brunitori odontoiatrici

3.1.2 Vocabolario CND

CND, acronimo di *Classificazione Nazionale dei Dispositivi medici*, è un vocabolario che opera una classificazione molto dettagliata delle apparecchiature mediche, come descrive l'acronimo. Esso divide i dispositivi in un albero in categorie, di cui il primo livello, formato da ventidue categorie, prendono il nome dalle lettere dell'alfabeto italiano dalla A alla Z (X esclusa).
Ogni categoria del primo livello si ramifica fino ad una profondità di quattro cifre in aggiunta alla lettera del primo livello, per un totale di 6985 diverse categorie.
L'unica limitazione di questo vocabolario è che si riferisce soltanto ai dispositivi medici, che sono solamente una parte delle spese ospedaliere.

3.2 Dataset dell'Autorità Nazionale AntiCorruzione

L'ANAC [1] è l'ente predisposto a raccogliere in dei documenti tutti i contratti pubblici dagli anni 2015 in poi [2]che hanno un importo nominale uguale o superiore a 40.000 euro [3]. Questi appalti per legge devono essere classificati in una categoria del vocabolario CPV. Di conseguenza questo risulta un dataset di tutti i contratti pubblici italiani già classificato da un ente riconosciuto a livello nazionale.

[1](acronimo di *Autorità Nazionale Anticorruzione*)

[2]Questo progetto è stato iniziato a Ottobre 2018 e concluso ad Aprile 2019, e visto che i contratti pubblici sono pubblicati sempre un anno dopo rispetto a quello a cui si riferiscono, durante lo svolgimento questa tesi si sono avuti a disposizione i contratti pubblici classificati dall'ANAC appartenenti agli anni 2015, 2016 e 2017.

[3]Questa è la regola nominale. Come vedremo in seguito, le analisi condotte hanno svelato che sono presenti anche una buona parte dei lotti con importo inferiore ai 40.000 €.

Capitolo 4

Metodi

In questo capitolo analizzeremo in ordine come è stato scelto il dataset su cui in seguito sono state condotte le analisi e la classificazione dei lotti del dataset. Nel mentre si metteranno a fuoco tutti i problemi di Qualità dei Dati che sono stati incontrati, con le relative soluzioni e accorgimenti adottati.

4.1 Introduzione ai file della legge 6 Novembre 2012, n.190

Nella fase iniziale del progetto di tesi gli esperti di Synapta mi hanno dato accesso ai contratti pubblici raccolti nel database aziendale che si riferiscono agli appalti pubblici degli anni che partono dal 2013 fino ad arrivare al 2017 [1]. La normativa 190/2012 impone infatti che i dati dell'anno X debbano essere pubblicati entro la fine del mese di Febbraio dell'anno X+1, ma è stato rilevato che negli anni X+2, X+3 è solito per le Pubbliche Amministrazioni pubblicare aggiornamenti di appalti (lotti) che sono stati ultimati (nel caso di opere la cui realizzazione necessita di tempi sull'ordine di svariati mesi o addirittura anni) o il cui importo pattuito è stato finito di pagare, oppure lotti che semplicemente non erano stati pubblicati nell'anno di scadenza per problemi tecnici o di altro genere. Questi risultati sono stati confermati dai test e dalle analisi condotte come vedremo in seguito.

[1] Questo progetto è stato iniziato a Ottobre 2018 e concluso a fine Marzo 2019. I dati del 2018 sono stati pubblicati dalle PA a Febbraio 2019 e quindi conseguentemente inseriti solo a partire da quella data nel dataset analizzato in questa tesi.

4.2 Sviluppo dell'ontologia per i contratti pubblici

Come attività parallela allo studio dei contratti pubblici si è cominciato a prendere dimestichezza la tecnologia SPARQL (vedi sezione 2.2.1) e le sue interrogazioni per due diverse ragioni. La prima è che le analisi sarebbero state condotte su un database a grafo interrogabile solo utilizzando query SPARQL, la seconda ragione è che in azienda era in corso un progetto volto a sviluppare un'ontologia [2] che permettesse di rappresentare ogni contratto pubblico in un formato compatibile con il database a grafo. Questa ontologia era ancora incompleta, perciò si è terminata la sua stesura e la si è testata per capire se tutte le informazioni dei contratti pubblici fossero correttamente tradotte nel nuovo formato.
In primo luogo si sono studiati tutti i namespaces [3] costruiti per esprimere concetti più complessi come quello di organizzazione, azienda, codice fiscale, partita IVA, denominazione sociale, che sono tutti concetti che ricorrono nei file XML dei contratti pubblici. I namespaces usati nell'ontologia sono:

- *<http://www.w3.org/2001/XMLSchema#* , namespace atto a definire principalmente le diverse tipologie di dato (e.g. *xsd:int* si riferisce al tipo di dato intero, *xsd:date* a una data, *xsd:float* a un numero decimale di tipo float)
- *https://w3id.org/italia/onto/COV/*, namespace che definisce il concetto di *Organization*, Company, e anche di attributi come il codice fiscale aziendale e la ragione sociale
- *https://w3id.org/italia/onto/l0/*, un namespace che è stato usato per aggiungere delle descrizioni a delle entità di tipo Classe
- *http://jarql.com/*, utilizzato per definire delle relazioni 'temporanee' per permettere di costruire e manipolare gli elementi dei contratti pubblici originari;
- *https://w3id.org/italia/onto/PublicContract*, che è un namespace che definisce un insieme di relazioni e classi utilizzati per un'ampia moltitudine di contratti spendibili anche per i contratti analizzati in questo progetto;
- *<http://test.yo/>*, un altro namespace di "appoggio" utilizzato per costruire delle URI identificative per tutti gli elementi che vengono creati dall'ontologia.

Analizziamo ora l'ontologia sviluppata. Questa verrà spiegata dividendo la query CONSTRUCT in tre parti:

[2] più dettagli nella sezione 2.2.2

[3] Un *namespace* è un riferimento ad uno spazio di nomi che indica delle risorse di tipo ontologico. Il nome completo è composto dalla coppia namespace+nome locale.

Prima parte. Namespaces Nella prima parte vengono inseriti i prefissi utilizzati nelle altre parti per riferirsi ai concetti racchiusi nei nampespaces e adattarli alle nostre esigenze:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dcterm: <http://purl.org/dc/terms/>
PREFIX : <https://w3id.org/italia/onto/PublicContract/>
PREFIX jarql: <http://jarql.com/>
PREFIX l0: <https://w3id.org/italia/onto/l0/>
PREFIX covapit: <https://w3id.org/italia/onto/COV/>
PREFIX test: <http://test.yo/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
```

Seconda parte. Output dell'ontologia La seconda parte entra nel vivo della query. Nella sezione *construct* viene costruito lo scheletro del file RDF che verrà restituito come risultato della query.

```
construct {
    ?URI_lot a :Lot;
        :actualStartDate      ?dataInizio;
        :actualEndDate        ?dataUltimazione;
        :totalAmountPaid  ?importoSommeLiquidate;
        :CIG                        ?cig;
        l0:description         ?oggetto;
        :hasAwardNotice  ?URI_award_notice;
        :isIncludedInProcedure ?URI_procedure.
    ?URI_award_notice a :AwardNotice;
         :selectedTender ?URI_tender;
         :agreedAmount ?importoAggiudicazione;
         :winner ?URI_winner_company.
    ?URI_tender a :Tender;
         :appliesToLot ?URI_lot;
         :tenderer ?URI_partecipante.
    ?URI_procedure a :Procedure;
         :includesLot ?URI_lot;
         :hasProcedureType ?URI_procedure_type;
         :hasProcuringEntity ?URI_proponent_struct.
    ?URI_procedure_type a :ProcedureType;
          rdfs:label ?sceltaContraente.
    ?URI_partecipante a covapit:Organization;
          covapit:legalName ?ragSocPartecipante;
          covapit:VATnumber ?codFiscPartecipante.
    ?URI_proponent_struct a covapit:Organization;
          covapit:legalName ?denominazioneStruttura;
          covapit:VATnumber ?codFiscStruttura.
```

```
    ?URI_winner_company a covapit:Organization;
          covapit:legalName ?ragSocAggiudicatario
          covapit:VATnumber ?codFiscAggiudicatario.
}
```

Terza parte. Elaborazione e manipolazione delle informazioni di input In questa parte invece viene fatto il binding tra le variabili che sono state usate nella seconda parte e i dati del file di input (nel nostro caso il file JSON). In aggiunta al binding in questa parte vengono anche manipolate queste variabili del file di input:

```
where {
   ?lotti jarql:lotto ?lotto.
   ?lotto jarql:cig ?cig.
   ?lotto jarql:strutturaProponente
                                ?strutturaProponente.
   ?lotto jarql:oggetto ?oggetto.
   ?lotto jarql:sceltaContraente ?sceltaContraente.
   ?lotto jarql:partecipanti ?partecipanti.
   ?lotto jarql:aggiudicatari ?aggiudicatari.
   ?lotto jarql:importoAggiudicazione ?importoAggiudicazioneStr.
   ?lotto jarql:tempiCompletamento ?tempiCompletamento.
   ?lotto jarql:importoSommeLiquidate ?importoSommeLiquidateStr.

   ?strutturaProponente jarql:codiceFiscaleProp ?codFiscStruttura.
   ?strutturaProponente jarql:denominazione ?denominazioneStruttura.
   ?partecipanti jarql:partecipante ?partecipante.
   ?partecipante jarql:codiceFiscale ?codFiscPartecipante.
   ?partecipante jarql:ragioneSociale ?ragSocPartecipante.
   ?aggiudicatari jarql:aggiudicatario ?aggiudicatario.
   ?aggiudicatario jarql:codiceFiscale ?codFiscAggiudicatario;
       jarql:ragioneSociale ?ragSocAggiudicatario.
   ?tempiCompletamento jarql:dataInizio ?dataInizioStr;
       jarql:dataUltimazione ?dataUltimazioneStr.

   bind(uri( concat( concat("http://test.yo/lotto/" , ?cig),
          ?codFiscStruttura)) as ?URI_lot ).
   bind(uri( concat( concat("http://test.yo/award_notice/",
           ?cig) , ?codFiscStruttura)) as ?URI_award_notice).
   bind(uri( concat( concat("http://test.yo/tender/" , ?cig),
           ?codiceFiscalePart)) as ?URI_tender ).
   bind(uri( concat( concat("http://test.yo/procedure/" , ?cig),
           ?codFiscStruttura)) as ?URI_procedure ).
   bind(replace(?sceltaContraente, '[s/]', '_') as ?
      sceltaContraenteUnder).
   bind(uri(concat("http://test.yo/",concat("procedure_type/",
          ?sceltaContraenteUnder))) as ?URI_procedure_type)
   bind(uri(concat("http://test.yo/partecipant/",
```

```
        ?codFiscPartecipante)) as ?URI_partecipant).
  bind(uri(concat("http://test.yo/struct/", ?codFiscStruttura))
                                    as ?URI_proponent_struct).
  bind(uri(concat("http://test.yo/agg/", ?codFiscAggiudicatario))
                                     as ?URI_winner_company).
  bind(strdt(?importoAggiudicazioneStr, xsd:float) as
                                     ?importoAggiudicazione).
  bind(strdt(?importoSommeLiquidateStr, xsd:float) as
                                     ?importoSommeLiquidate).
  bind(strdt(?annoRiferimentoStr, xsd:date) as ?annoRiferimento).
  bind(strdt(?dataInizioStr, xsd:date) as ?dataInizio).
  bind(strdt(?dataUltimazioneStr, xsd:date) as ?dataUltimazione).
}
```

4.3 Analisi del dataset dei contratti pubblici

Una volta completata l'ontologia l'attenzione si è completamente rivolta allo studio del dataset dei contratti pubblici, per capire la distribuzione dei lotti secondo determinati parametri e per evidenziare eventuali problemi o anomalie che non sono emerse a prima vista.
In questa fase sono cominciati ad emergere i problemi di qualità dei dati che caratterizzano questi file. Purtroppo non è stata prevista una revisione e/o manutenzione dei dati pubblicati dalle Pubbliche Amministrazioni e ricevuti dall'ANAC, in quanto la funzione di questi documenti è di rendere trasparenti le spese finanziate con fondi statali, mentre sono utilizzati a scopo di analisi in questo progetto di tesi. La conseguenza è che questi dati siano di qualità mediocre, e senza un attento studio e revisione i risultati potrebbero mostrarsi poco precisi e fedeli rispetto alla realtà.
Una volta cominciato il processo di unmarshalling [4] dei file XML per tradurre questi dati in oggetti Java, su 442.524 lotti analizzati sono stati individuati 20 file XML con numero di lotti variabile in cui la procedura di unmarshalling è fallita a causa di errori di sintassi dei file (ad esempio i tag XML non erano aperti e poi chiusi correttamente, oppure mancavano entrambi i campi dell'importo o soltanto uno dei due, stessa cosa per le date di inizio e di completamento dei lotti). Ancora ci sono casi in cui l'elemento dei partecipanti è vuoto, e in queste situazioni se il campo <aggiudicatari> ha almeno un elemento <aggiudicatario>, si assume che la gara d'appalto abbia come partecipanti le aziende (o l'azienda) vincitrice dell'appalto.
Per quanto riguarda invece i dati mancanti degli importi, in seguito si è trovata una soluzione per cercare di pulire dati dei lotti, ma ne discuteremo nelle successive sezioni.

[4] si faccia riferimento alla sezione 2.3.3

Una volta convertiti i dati in oggetti Java, si è potuto procedere con l'analisi vera e propria del dataset, composto soltanto da dati esenti da errori sintattici [5].

4.3.1 Analisi della distribuzione dei dati per anno e per prezzo

Le analisi fatte sui lotti hanno come scopo quello di calcolarne la loro distribuzione per anno e per prezzo.

Numero di lotti per ogni anno. La prima analisi fatta è stata quella per evidenziare il numero di lotti presenti in ogni anno. Di seguito è riportato l'istogramma 4.1 che mostra i risultati ottenuti:

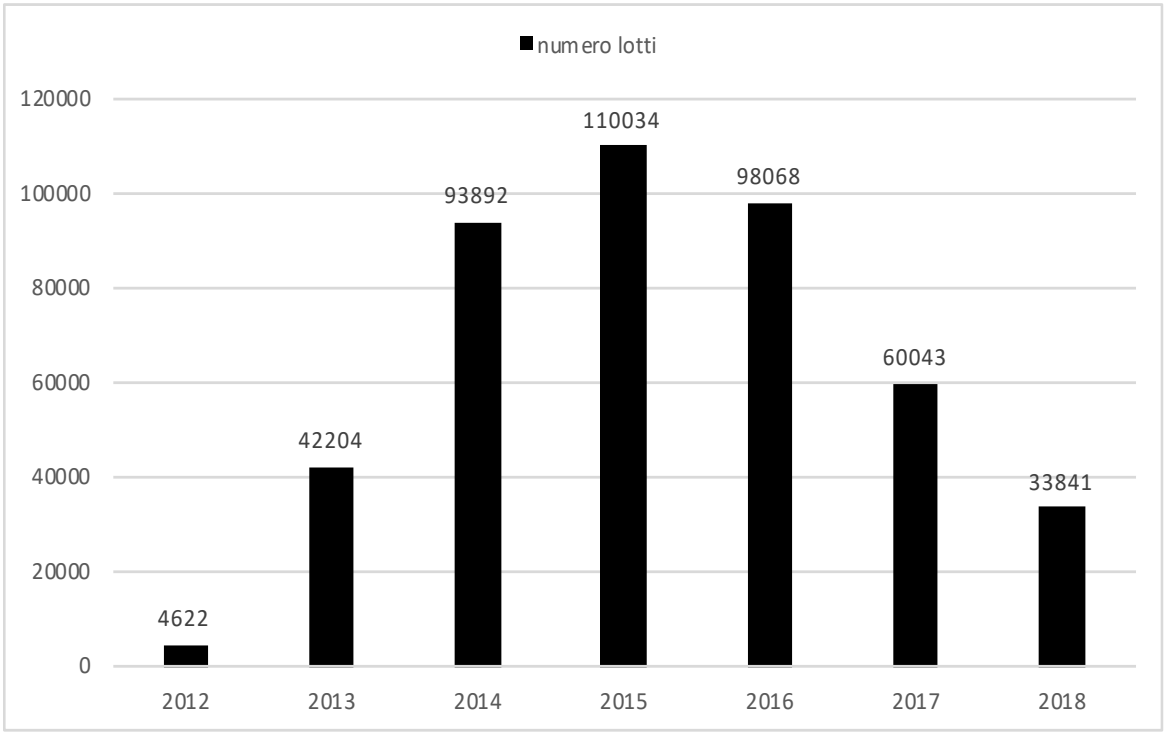

Figura 4.1. Ogni barra indica il numero totale di lotti per il corrispondente anno sull'asse delle ascisse.

[5] Gli errori di tipo semantico sono stati opportunamente trattati in seguito

Negli anni 2014, 2015 e 2016 siano presenti il maggior numero di lotti, mentre nel 2013, 2017 e 2018 ce ne sono circa la metà rispetto ai primi 3. Nel 2012 troviamo un numero molto minore rispetto agli altri anni.
L'anno 2015 è quello in cui sono presenti più lotti, circa 110.000, a seguire il 2016 con circa 98.000 lotti e poi il 2014 con 93.000 lotti. Seguono il 2017, 2013 e 2018 con rispettivamente 60.000, 42.000 e 33.000 lotti, e infine abbiamo il 2012 in cui troviamo solo 4622 lotti.
La prima considerazione da fare su questo questo risultato è che sembrerebbe che gli anni che contengono più informazioni siano il 2014, 2015 e 2016. Per dare credito a questa considerazione, procediamo con le successive analisi.

Spesa totale per anno. Raggruppiamo ora la spesa totale per anno, e verifichiamo in quali anni sono contenute più informazioni in termini di denaro speso. Ci si potrebbe aspettare una possibile correlazione tra numero di lotti presenti in un anno e spesa sostenuta. In figura 4.2 riportiamo i risultati trovati:

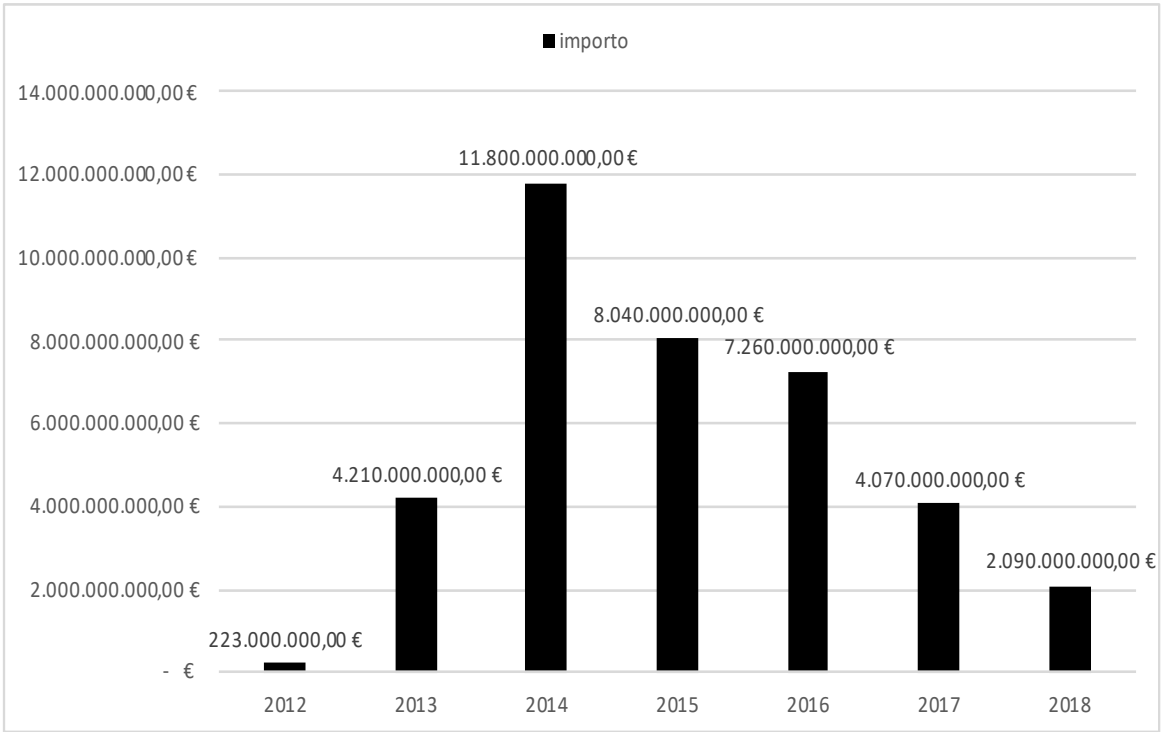

Figura 4.2. Ogni barra indica l'importo totale di tutti i lotti del corrispondente anno rappresentante sull'asse delle ascisse.

Osservando il grafico si può verificare l'ipotesi fatta precedentemente: il numero di lotti corrisponde alla spesa maggiore per tutti gli anni considerati tranne per il 2014. In questo anno è concentrata la spesa maggiore rispetto agli altri anche se il numero di lotti è minore rispetto al 2015; per quanto riguarda gli anni rimanenti, abbiamo la spesa distribuita in maniera proporzionale al numero di lotti. In particolare nel 2014 abbiamo poco più di 11 miliardi di euro di spesa, nel 2015 circa 8 miliardi di €, nel 2016 7,2 miliardi, mentre nel 2013 registriamo 4,2 milioni di € di spesa, nel 2017 poco più di 4 miliardi di € di spesa, scendiamo a 2 miliardi nel 2018 e 2 milioni di € nel 2012.
Dopo aver valutato i risultati di questa seconda analisi possiamo dar credito alla considerazione fatta precedentemente: sembrerebbe che gli anni che contengono maggiori informazioni siano il 2014, 2015 e 2016, mentre il 2013, 2017 e 2018 contengono circa la metà della spesa, e il 2012 si conferma l'anno più povero di informazioni anche in termini di denaro speso.

Distribuzione dei lotti per importo Come ultima analisi sul dataset si è calcolata la distribuzione dei lotti per importo. Gli importi dei lotti spaziavano da 0 a quasi 1 milione di euro. In base a questi due dati si sono scelte 7 classi di importo in cui distribuire i lotti: la prima va da 0 a 10 €, la seconda da 11 a 100 €, la terza da 101 a 1000 €, la questa da 1001 a 10000 €, la quinta da 10001 a 100000 €, la sesta da 100001 a 500000 € e l'ultima da 500001 a 1000000 €.
Commentiamo i risultati dell'analisi in figura 4.3.

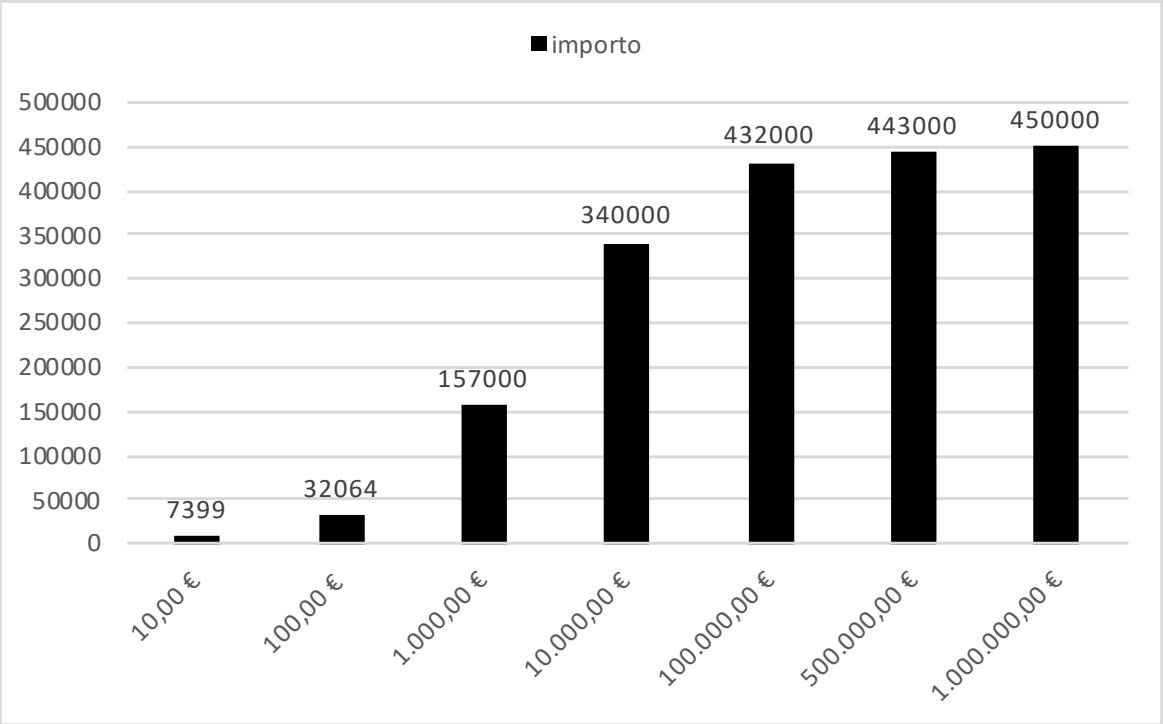

Figura 4.3. Ogni barra indica il numero di lotti con importo minore o uguale all'importo sull'asse x.

In questo grafico salta all'occhio come ci siano pochissimi lotti rispetto al totale con valore sotto i 10 e sotto i 100 €, mentre a partire dai 100 € fino ad arrivare ai 100000 € abbiamo raggruppato la maggior parte dei lotti. Dai 100.000 ai 10.000.000 € rimangono poche migliaia di lotti.
L'unica considerazione su questo risultato è il dubbio set di lotti caratterizzati da un importo inferiore ai 10 €, in quanto sembra strano che si indica una gara d'appalto per un servizio o la realizzazione di un'opera con importo pattuito di una decina di euro o meno.

4.3.2 Anomalia del dataset e soluzione proposta

Proseguendo nell'analisi del dataset è stata rilevata un'anomalia: contando il numero totale di cig e di lotti, che data la definizione di cig [6] sono due numeri che

[6]CIG, acronimo di Codice Identificativo Gara, è un numero univoco composto da 10 caratteri che ha la funzione di identificare in modo assoluto un appalto per permettere quindi la tracciabilità

dovrebbero essere coincidenti, i risultati rivelano l'opposto. Infatti sono risultati 454.332 lotti e 153.493 cig diversi, in altre parole per ogni cig sono presenti in media 3 lotti.
Questo problema era giá stato riscontrato dai programmatori e dagli analisti in azienda, e quando è stata presentata questa anomalia si è deciso di considerare per ogni codice identificativo [7] il lotto con importo di aggiudicazione maggiore, in quanto questo problema è causato dal fatto che le strutture sanitarie in molti casi riutilizzano lo stesso cig per lotti che vengono acquistati periodicamente e ripetutamente nel tempo, oppure altre volte è soltanto dovuto a errori di disattenzione da parte degli operatori responsabili della scrittura dei documenti dei lotti.
A causa di questo problema di qualità dei dati e della soluzione adottata, si è dovuto ridurre il numero dei lotti nel dataset.
Nelle successive sezioni verrà spiegato come mai si è voluto approfondire questa problematica relativa al cig dei lotti e invece di accettare di lavorare su dei lotti con codice identificativo non univoco si è preferito ridurre notevolmente la dimensione del dataset.
Ripropongo in seguito le distribuzioni ottenute sul nuovo set di dati messe a confronto con quelle fatte precedentemente.
Vediamo la distribuzione del numero di lotti per anno (figura 4.4).

dei pagamenti effettuati dalla pubblica amministrazione italiana. Dunque, per ogni lotto dovrebbe corrispondere un cig diverso

[7] si fa riferimento al codice identificativo gara

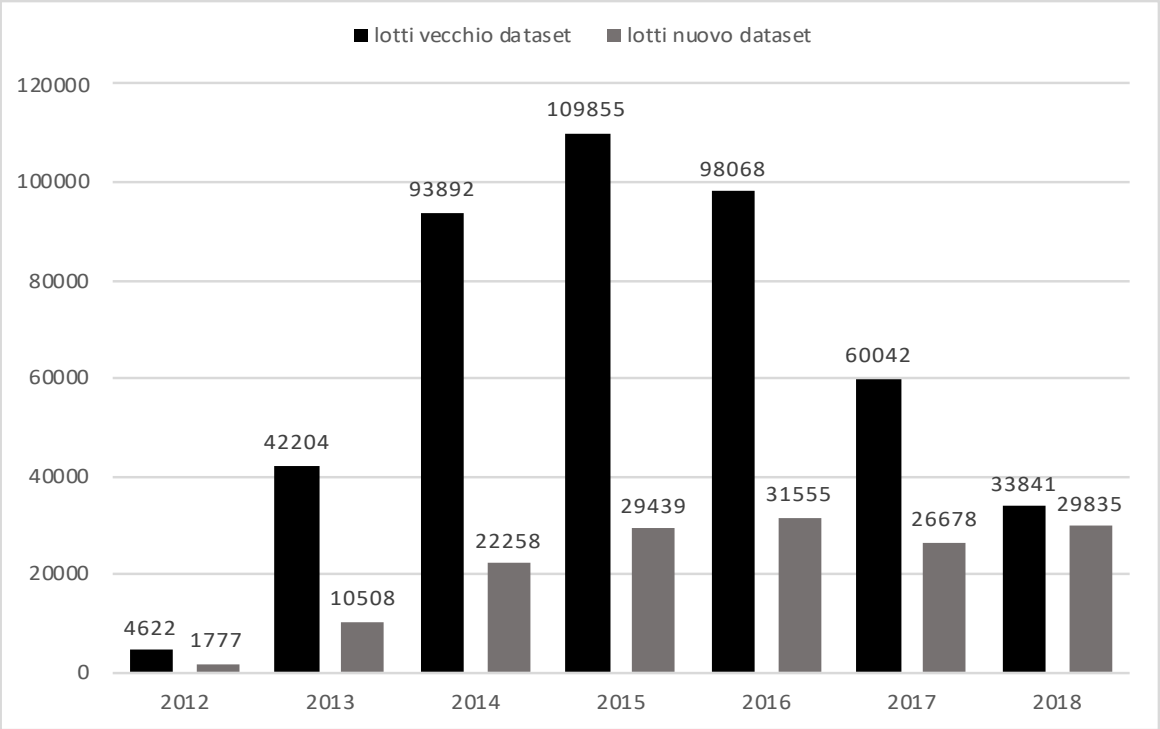

Figura 4.4. Le barre grigie indicano il numero totale di lotti per il corrispondente anno sull'asse delle ascisse del nuovo dataset, mentre quelle nere indicano il numero di lotti per il corrispondente anno del vecchio dataset

Il numero totale di lotti del nuovo dataset è di 152.050 rispetto ai 442.524 del vecchio dataset, circa ⅔ in meno.
Questa diminuzione è presente abbastanza uniformemente in tutti gli anni tranne il 2017 e il 2018, rispettivamente la metà e il 90% rispetto ai lotti del vecchio set.
Una seconda osservazione è che dal 2015 al 2018 abbiamo un numero abbastanza omogeneo di lotti nel nuovo dataset, molto diverso rispetto alla distribuzione di quello vecchio. In più il 2016 e il 2018 sono i due anni con il numero di lotti maggiore, mentre il 2012 si riconferma quello con meno informazioni di tutti. Un'ultima riflessione è fatta sull'anno 2014, che è quello che ha perso più lotti di tutti e che quindi conteneva più lotti duplicati (e quindi dati di qualità peggiore rispetto agli altri anni).
Esaminiamo ora la nuova spesa dei lotti del nuovo dataset raggruppata per anno nella figura 4.5 per vedere se i risultati rispecchiano quelli della figura 4.4 :

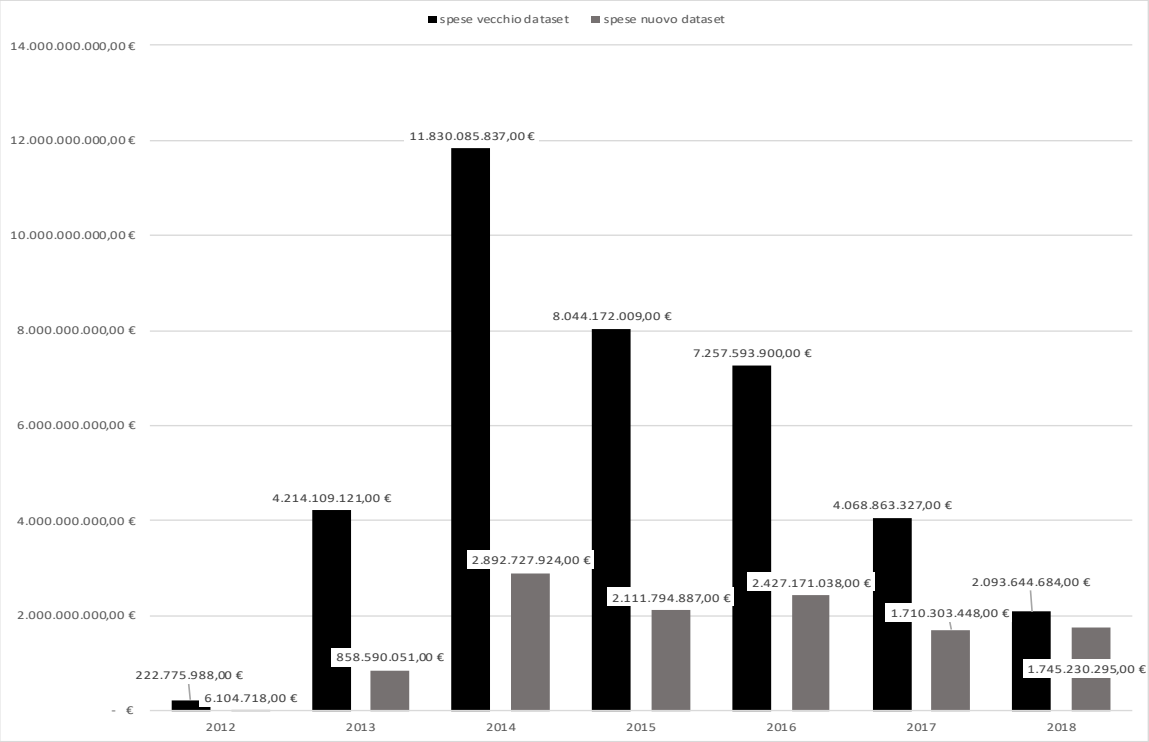

Figura 4.5. Le barre grigie indicano l'importo totale dei lotti per il corrispondente anno del nuovo dataset, mentre quelle nere indicano la spesa totale dei lotti per il corrispondente anno del vecchio dataset

La spesa totale del nuovo dataset è di 11.806.864.829 €, rispetto alla precedente di 37.731.244.866 €, una diminuzione di circa ⅔ come quella della precedente analisi.
Questa diminuzione è abbastanza omogenea su tutti gli anni, tranne per gli anni 2017 che la ha diminuita di circa metà e 2018 che ha perso circa il 10%. Questa stessa considerazione è stata fatta per il grafico precedente sul numero di lotti per anno.
Consideriamo la distribuzione generale del trend delle spese per tutti gli anni rispetto al vecchio set di dati: troviamo un'omogeneità nei valori degli importi dal 2014 al 2018, mentre gli anni 2013 e 2012 in particolari sono quelli con minori informazioni. Il trend rispetto ai risultati ottenuti in figura 4.4 è simile tranne per l'anno 2014: infatti è quello che nel nuovo dataset (come in quello vecchio) registra la spesa più alta di tutti.

Ripetiamo l'ultima distribuzione sull'importo di ogni lotto del nuovo dataset in confronto a quello vecchio, figura 4.6 :

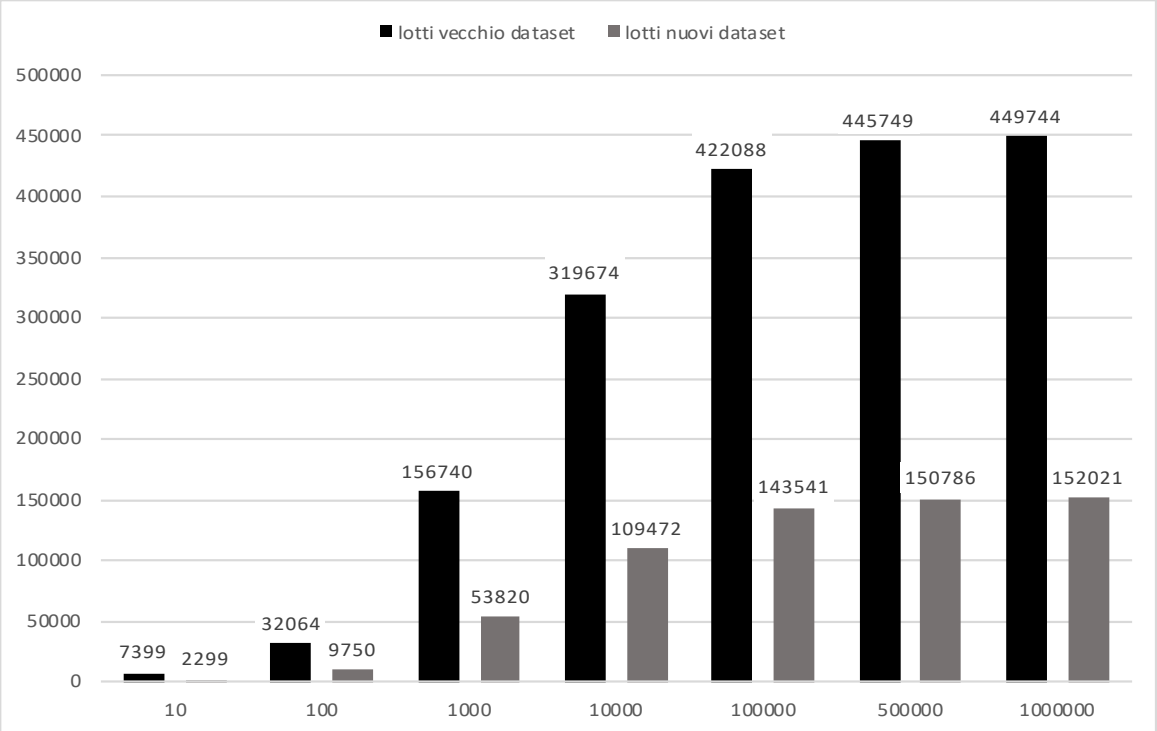

Figura 4.6. Ogni barra grigia indica il numero di lotti del nuovo dataset con importo minore o uguale al quello sull'asse delle ascisse, mentre ogni barra nera si riferisce al numero di lotti con importo minore o uguale all'importo sull'asse x del vecchio dataset

Analizzando quest'ultimo grafico notiamo che il numero di lotti in ogni fascia di prezzo è diminuito in media di circa due volte, con qualche perdita maggiore nei lotti con importo maggiore di 100.000 €. Il trend della distribuzione è rimasto pressoché inalterato: la maggior parte dei lotti rimane concentrato nel range di importo che va dai 100 ai 100.000 €. Si sono dunque perse informazioni in maniera abbastanza uniforme, con qualche picco nei lotti con importo maggiore a 100.000 €.

Per riassumere, abbiamo circa perso il ⅔ delle informazioni a causa della soluzione che si è scelti di adottare per risolvere il problema di data quality circa la relazione tra lotto e cig.

4.4 Classificazione del campo oggetto dei lotti

Dopo aver approfondito la composizione del mio dataset, l'attenzione si è rivolta a una delle maggiori sfide incontrate nello svolgimento di questo progetto: la classificazione del campo oggetto dei lotti. Per decidere quale alternativa fosse meglio adottare si è valutata l'accuratezza del classificatore CPV (sezione 2.3.4) sviluppato da Synapta e quella della classificazione dei dati ANAC (sezione 3.2).

4.4.1 Dataset ANAC

Esaminiamo la composizione del dataset ANAC: questo set di dati si riferisce soltanto agli anni che vanno dal 2015 al 2017; questi dati sono dei file pubblicati in formato CSV, e sono formati da 48 colonne.
Questi dati presentano molti problemi di qualità, il primo dei quali risiede nel fatto che ogni colonna è sprovvista di un header di descrizione. Si è dovuto quindi dedurre il contenuto delle colonne per potere utilizzare il dataset: è stato individuato il campo che contiene il codice CPV in quanto questo è formato da esattamente otto caratteri alfanumerici. Confrontando l'unico campo formato da otto caratteri con i codici CPV del dataset dei contratti pubblici si è trovata una percentuale molto alta di corrispondenza, che ha confermato il contenuto della colonna. Si è inoltre dedotto il campo che contiene la data di inizio e di ultimazione del lotto ANAC, in quanto il formato della data si differenzia da quello di tutte le altre colonne. In seguito sono stati riconosciuti due campi che contengono gli importi del lotto, che sembravano corrispondere agli importi di aggiudicazione e quelli liquidati che caratterizzano ogni lotto dei contratti pubblici. Purtroppo non è stato possibile capire a quale importo le colonne del dataset si riferiscono: queste due colonne di importi sono uguali nella stragrande maggioranza dei casi, e nei casi in cui sono diversi differiscono di esattamente tre ordini di grandezza, fatto che risultava sospetto oltre che inutile per distinguerli.
L'ultimo campo che si è riuscito a distinguere è stato quello in cui veniva riportato il codice CPV del lotto, in quanto seguito direttamente dalla corrispondente stringa di descrizione della categoria. In più un rapido confronto tra i presunti codici del dataset ANAC e il vocabolario reperibile in rete non ha lasciato alcun dubbio.
Infine un ultimo problema riscontrato durante l'analisi di questo file è la presenza di campi che non erano indentati nella propria colonna e la presenza di valori nulli in tante colonne.
La classificazione ANAC è molto precisa, infatti la precisione minima con cui ogni lotto è classificata è del quarto livello, mentre la massima precisione a cui si arriva a classificare una parte dei lotti è l'ottavo livello.

4.4.2 Scelta della classificazione

Come anticipato in precedenza, uno dei criteri per scegliere tra i due classificatori è stato quello di valutare la bontà della loro classificazione, confrontando il campo oggetto da classificare con il risultato della classificazione ottenuta con entrambi i metodi.
Chi scrive ha valutato la loro precisone personalmente, limitandosi a considerare come corretta una classificazione molto chiara e dettagliata con presenza di parole chiave o sinonimi delle parole chiave ripetute sia nel campo oggetto che nella classificazione risultante; ha anche considerato sbagliata la classificazione troppo generica. Ad esempio, l'oggetto descritto dalla stringa *"sostituzione coledocoscopio digitale"* a cui è stata attribuita la categoria CPV *"331412 - cateteri"* è considerata sbagliata, in quanto non sono presenti ripetizioni di parole chiave o sinonimi nella stringa dell'oggetto e nella descrizione della categoria. Rientra nelle classificazioni sbagliate anche la categoria CPV *"6311 - servizi di movimentazione e magazzinaggio"* per l'oggetto *"Servizio di gestione delle attività di supporto alla logistica saniaria ed economale"*, in quanto non c'è nessuna ripetizione o riferimento a parole o sinonimi che tolgano ogni dubbio circa la classificazione. Un esempio di classificazione corretta per l'oggetto *"fornitura di sonde e cateteri - cateteri autolubrificanti in PVC di lunghezza cm 40"* è *"331412 - cateteri"*, in quanto è presente la ripetizione della parola chiave *"cateteri"*.
Questo approccio è stato obbligato dal fatto che chi scrive non è un esperto di contratti pubblici, e il suo giudizio riguardo la classificazione di un lotto è anch'esso altamente discutibile e opinabile. Adottando questo metro di valutazione oltre a ridurre la possibilità di errore nel giudizio positivo di una classificazione che invece è errato, si cerca di fissare un limite minimo di garanzia di accuratezza che paragonato al giudizio di una persona esperta e qualificata nel settore può soltanto migliorare.
Per valutare la precisione della classificazione si sono classificati 500 lotti presi casualmente fra il dataset con il classificatore CPV di Synapta, e altri 500 lotti casuali fra quelli presenti incrociando il dataset dei contratti pubblici con quello ANAC, in modo da poter confrontare il campo oggetto con la classificazione ANAC. I risultati della classificazione dei 500 campioni casuali è stata la seguente:

Classificazione	**Totali**	**Giuste**	**Sbagliate**	**Accuratezza**
Classificatore CPV	500	276	224	55,2 %
Classificazione ANAC	500	388	112	77,6 %

Tabella 4.1. Confronto tra le accuratezze delle classificazioni ANAC e del Classificatore CPV.

La valutazione del campionamento è risultata favorevole per la classificazione ANAC, in quanto l'accuratezza è superiore rispetto a quella del Classificatore CPV di più del 20%, oltre al fatto che la precisione minima della classificazione ANAC è al quarto livello CPV, mentre quella del classificatore CPV è soltanto al secondo livello.
Anche se i numeri sembrano indicare ANAC come la classificazione migliore tra le due opzioni, sono state fatte altre considerazioni: in primo luogo i dati ANAC sono legati ai lotti dei file XML della legge 190/2012 dal codice cig del lotto, e gli esperimenti sui dati hanno rivelato che nel dataset ANAC sono presenti soltanto una parte dei lotti presenti nel dataset dei contratti pubblici; bisogna quindi valutare se il sottoinsieme di lotti presente in ANAC e nel dataset della normativa 190/2012 è rappresentativo o meno rispetto al totale. In secondo luogo, usare la classificazione ANAC vuol dire ridurre l'analisi ai soli anni 2015, 2016, 2017, non considerando gli anni 2012, 2013, 2014 e 2018. Perciò per scegliere la classificazione ANAC si è proceduto a verificare che la spesa coperta dal sottoinsieme comune ai due dataset è rappresentativa.
Osserviamo dunque i numeri: l'importo totale di riferimento per valutare la fetta di spesa del sottoinsieme ANAC è la somma degli importi aggiudicazione di tutti i lotti appartenenti agli anni 2015, 2016 e 2017 del dataset dei contratti pubblici, che ammonta a 6.249.269.373 €. La somma degli importi dei lotti presenti nel dataset ANAC il cui cig è presente nel dataset dei contratti pubblici nel triennio 2015-2017 risulta essere di 5.192.738.352 €, ovvero l'83% del totale. Si può quindi affermare che la somma degli importi presente nel dataset ANAC ricopre più di gran parte del totale degli importi del dataset della normativa 190/2012.
Per quanto riguarda l'analisi di soli 3 anni rispetto ai 6 a disposizione nel dataset, i quali sarebbero tutti classificabili utilizzando il classificatore CPV, si p scelto comunque di utilizzare la classificazione ANAC per il risultato della classificazione del campione di 500 elementi, in quanto il classificatore CPV in media sbaglia 1 classificazione su 2, che è un risultato che renderebbe i risultati delle analisi poco affidabili.
Facciamo riferimento a quanto promesso nella sezione 4.3.2 dove avevamo anticipato il motivo per cui si è controllato che il numero di lotti totali e il numero di cig diversi coincidessero. Per classificare i lotti dei contratti pubblici utilizzando la categoria CPV dei lotti nel dataset ANAC si sono incrociati questi due insiemi di dati utilizzando il cig come attributo comune. Se non si fosse provveduto ad avere un solo cig per ogni lotto nei contratti pubblici, un lotto ANAC corrisponderebbe in media a tre lotti dei contratti pubblici, e non si avrebbe una corrispondenza univoca tra i due dataset[8].

[8]Questo si tradurrebbe in un errore concettuale: se si cercasse il corrispondente lotto ANAC nel dataset dei contratti pubblici, il lotto trovato potrebbe differire in ricerche differenti.

4.5 Operazione di integrazione dati e classificazione con dataset ANAC

Il dataset ANAC oltre a essere stato scelto per classificare il campo oggetto dei lotti del dataset della normativa 190/2012, è stato anche utilizzato come dataset da cui attingere il campo importo quando il campo importo aggiudicazione dei lotti è nullo.
Come si è visto nella figura 4.3 ,durante l'analisi della distribuzione degli importi dei lotti è emerso che circa 2500 di questi hanno come valore dell'importo aggiudicazione 0. Questo fatto risulta chiaramente un errore in quanto è impossibile che un appalto venga aggiudicato a una società a gratis o per una decina di euro.
Nel dataset ANAC sono riportati due importi, i quali sono stati usati per cercare di recuperare il lotto con importo nullo e renderlo quindi utilizzabile per le successive analisi. In seguito viene spiegato il procedimento che è stato adottato per cercare di recuperare i dati reali degli importi nel dataset dei contratti pubblici.
In primo luogo quando si trova un importo nullo, si guarda il corrispondente importo somme liquidate. Questo nel 90% dei casi risulta nullo, e quindi non può essere utilizzato come sostitutivo.
In secondo luogo in presenza di importo pari a zero si guarda l' importo presenti nel dataset ANAC[9]. Questo importo si è dimostrato reale e ottimo come sostituto in quanto sono state condotte alcune analisi per verificarne la bontà. Considerando tutti i lotti con qualsiasi importo, nel 90% dei confronti con l'importo aggiudicazione dei file della legge 190/2012, l'importo ANAC è dello stesso ordine di grandezza, e nell'81% dei casi non lo supera in eccesso o in difetto neanche del doppio. Perciò si è deciso di approssimare tutti gli importi che nei file XML erano a 0 con l'importo corrispondente ANAC.
Dopo aver risolto il problema degli importi nulli, si è proceduti a integrare il campo oggetto dei file XML con la classificazione presente nei file ANAC.
Come già detto precedentemente, il dataset ANAC classifica ogni lotto con precisione minima garantita del livello 4 CPV. Ciò significa che se si volesse ottenere una classificazione più accurata, ad esempio scendendo al livello 5 del codice, sorgerebbe la necessità di classificare a mano i lotti che non arrivano a questo livello di dettaglio, cercando di trovare la categoria CPV a livello 5 che più si avvicina al campo oggetto del lotto. Le analisi hanno mostrato che circa 9.000 lotti non raggiungono la precisione di classificazione al livello 5, dunque si è deciso di accettare come precisione massima di classificazione quella a livello 4 del CPV offerta dal dataset ANAC per due motivi: il primo è che io avrei dovuto classificare tutti questi lotti senza il parere di un esperto di contratti pubblici, di conseguenza la

[9]Tra i due importi presenti nel dataset ANAC si sta considerando quello minore, per le motivazioni spiegate in seguito.

classificazione non sarebbe stata giustificata da nessuna autorità. Il secondo motivo è che il numero di lotti da classificare è troppo alto, e si sarebbe impiegato troppo tempo per completare questo passaggio.
Nelle analisi in questo lavoro di tesi, si sono analizzate le spese prima utilizzando le prime 3 cifre del codice CPV, per poi valutare la fattibilità di un'analisi più dettagliata utilizzando anche la quarta cifra del codice.
Per scrupolo il campo oggetto dei file XML dei contratti pubblici non è stato sostituito con il codice e la descrizione della categoria CPV, ma è stato integrato con essi. Più precisamente la categoria CPV è stata concatenata (separato dal campo oggetto con il carattere "_") in modo tale da poter verificare eventuali anomalie.
Questa operazione di integrazione è stata effettuata con un algoritmo ricorsivo [10] scritto in linguaggio Java; dopo aver caricato in memoria tutti i dati del dataset ANAC e tutti i codici CPV con le corrispondenti descrizioni, l'algoritmo compie uno scan della directory in cui sono contenuti tutti i file del dataset dei contratti pubblici, sostituendo i valori nulli dell'importo aggiudicazione e integrando il campo oggetto con la classificazione ANAC. Durante l'esecuzione del programma vengono anche calcolate le spese per anno e il relativo numero di lotti, in modo da confrontare questi risultati con quelli ottenuti una volta caricati i dati sul database.
Terminata l'integrazione dei dati originari con le nuove informazioni ANAC, i nuovi file XML sono stati tradotti in formato JSON utilizzando un programma in tecnologia Python, che utilizza la libreria XMLTODICT [11].
Questi file JSON non sono però direttamente traducibili con JARQL in quanto XMLTODICT in caso di elementi XML vuoti genera elementi JSON con valore nullo che non sono supportati da JARQL. Viene perciò effettuato un passaggio intermedio in cui gli elementi nulli sono sostituiti da elementi aventi valore vuoto (*nome_elemento : ""*).
Questi nuovi file JSON possono essere direttamente trasformati in formato RDF che ha come estensione *ttl*. Questa conversione viene effettuata utilizzando l'applicativo di nome JARQL (sezione 2.3.1). JARQL è un programma capace di trasformare un file JSON alla volta in formato RDF, quindi per automatizzare la conversione si è ricorsi ad un programma che lancia un nuovo processo ogni volta che viene individuato un file avente estensione *.json*.
Terminata la trasformazione da JSON a RDF, i dati di input sono processati da un nuovo programma che lancia un processo ogni volta che un file con estensione *.rdf* viene trovato. Ogni processo esegue un comando CURL che prende come parametri l'indirizzo IP del database Blazegraph a cui mandare la richiesta HTTP POST

[10] Un algoritmo ricorsivo è un algoritmo che richiama se stesso generando un numero di chiamate non determinato a priori, dipendente dalle variabili di input e dal genere del problema in questione, che termina quando si verifica una condizione particolare chiamata *condizione di terminazione*.

[11] Questa libreria di Python mette a disposizione delle funzioni per tradurre il formato da XML a JSON e per gestire eccezioni dovute a errori di sintassi del file da tradurre.

contenente il file RDF da caricare sul database.
Terminate tutte le operazioni descritte il nostro database a grafo è pronto ad essere interrogato.

4.6 Analisi del risultato delle query sul database

Blazegraph è un database che si appoggia al browser durante l'esecuzione. Una volta caricati i dati, è possibile eseguire delle query sul dataset per aggregare e visualizzare i dati contenuti. Il risultato di queste interrogazioni viene rappresentato utilizzando un formato tabulare visualizzato nel browser, ma non comodamente esportabile su altri applicativi come Excel o simili che permettono di dare una rappresentazione più efficace ai risultati.
Sfruttando la REST API di Blazegraph è possibile eseguire una query SPARQL dal Terminale desktop tramite il comando CURL, e il risultato viene fornito non più in formato tabulare ma bensì in formato JSON. Manipolando il i dati risultanti è possibile convertirli in altri formati in base alle proprie esigenze.
Per ottenere questo risultato, che in questo progetto risulta estremamente comodo e utile in quanto i risultati delle query devono essere raggruppati in tabelle più complesse per costruire grafici comparativi, si è sviluppata un'applicazione Java che riceve come parametri la query in formato SPARQL da eseguire sul database e che crea come risultato una cartella al cui interno si viene creato un file JSON contenente il risultato in formato JSON della query eseguita, un file che contiene la risposta opportunamente trasformata dal formato JSON al formato CSV e infine un file di testo per tenere memoria della query eseguita.
Grazie a questa applicazione è stato possibile rappresentare più efficacemente i risultati, non solo in formato numerico ma anche in formato grafico.

Capitolo 5

Risultati

Come già accennato precedentemente, le spese sono state categorizzate e analizzate prima a livello 3 e dopo a livello 4 del codice CPV. Esse si riferiscono agli anni 2015, 2016 e 2017 (si veda sezione 3.2). Purtroppo come vedremo nelle prossime sezioni, i dati di certe categorie già a livello 3 si sono dimostrati piuttosto sparsi, e dunque l'analisi a livello 4 del codice CPV non è stata effettuata in quanto si sarebbero trovati dati ancora più sparsi, senza poter effettuare adeguate e interessanti considerazioni sui risultati.
Il tema centrale di queste analisi è quello di mettere in evidenza quali sono le categorie merceologiche a livello 3 in cui si spende di più, e quali sono le ASL e le aziende ospedaliere a incidere di più in questa spesa. Nelle prossime sezioni verranno ricapitolati i risultati ottenuti, che verranno forniti come risposte a domande.
Prima di concentrarci su dove siano stati spesi i soldi, confrontiamo i nostri dati, in particolare le spese del biennio 2015-2016 divise per ASL e Azienda Ospedaliera con i rendiconti finanziari ufficiali dei due anni [8], in modo da paragonare i numeri del dataset utilizzato con i dati ufficiali del Ministero della Salute. Purtroppo non è stato pubblicato ancora il rendiconto dell'anno 2017, anno che è stato escluso da questa analisi di confronto.

5.1 Confronto delle spese del dataset con Rendiconti Finanziari

I risultati di queste analisi sono stati raggiunti eseguendo una query sul database per raggruppare tutta la spesa del 2015 e poi del 2016 per ospedale. La query è la seguente:

```
PREFIX : <https://w3id.org/italia/onto/PublicContract/>
PREFIX jarql: <http://jarql.com/>
PREFIX l0: <https://w3id.org/italia/onto/l0/>
```

```
PREFIX covapit: <https://w3id.org/italia/onto/COV/>
PREFIX test: <http://test.yo/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

select ?denominazioneStruttura (sum(?importoAggiudicazione) as ?
    sommaSpesa)

where {

       ?lotto     a                                        :Lot;
                  l0:description                ?oggetto;
                  :actualStartDate              ?dataInizio;
                  :totalAmountPaid              ?importoSommeLiquidateFloat
                      ;
                  :isIncludedInProcedure ?URI_procedure;
                  :CIG                                  ?cig;
                  :hasAwardNotice             ?URI_award_notice.

       ?URI_award_notice :agreedAmount ?importoAggiudicazione.

       ?URI_procedure                        :hasProcuringEntity ?
           URI_proponent_struct.
       ?URI_proponent_struct      covapit:legalName    ?
           denominazioneStruttura.

       bind(strbefore(?oggetto, "__") as ?cpv)
       bind(strafter(str(?dataInizio), "_") as ?data)
       bind(strbefore(?data, "-") as ?anno)


       FILTER(
         ?anno = "2015" #?anno = "2016"
         &&
         ?importoAggiudicazione >= 0
         &&
         ( contains(?denominazioneStruttura, "ASL")
             ||
           contains(?denominazioneStruttura, "Azienda") )
       )

   }

group by ?denominazioneStruttura
order  by ?denominazioneStruttura
```

Nelle due query l'unico parametro che cambia è quello relativo all'anno a cui ci si riferisce, che nel primo caso è "2015" mentre nel secondo caso è "2016", come commentato nella query riportata.

5.1.1 Spese a rendiconto considerate per il confronto

I rendiconti finanziari utilizzati per il confronto sono pubblicati dal ministero della salute sul proprio sito. Questi rendiconti considerano molte spese, di cui solo un sottoinsieme è stato scelto per fare il confronto, in quanto i lotti di contratti pubblici si riferiscono a degli appalti e certi tipi di spese non rientrano in prestazioni lavorative che necessitano di un appalto per poter essere assegnate. Le spese che sono state considerate in ogni rendiconto sono le seguenti:

- Acquisti di beni
- Acquisti di servizi
- Manutenzione e Riparazione

Nelle successive sezioni sono mostrati i risultati dei confronti tra le spese sottoforma di istogramma comparativo del valore assoluto delle spese del dataset e del rendiconto, e sottoforma di istogramma relativo alla percentuale di spesa coperta dal dataset rispetto al rendiconto finanziario.

5.1.2 Confronto anno per anno

Anno 2015

Dopo aver interrogato il database formulando una query che restituisse la spesa totale dei lotti che sono state fatte nel 2015 raggruppata per struttura sanitaria, si sono confrontati i dati del dataset utilizzato con i dati ufficiali. In seguito è riportato il grafico 5.1 che mette a confronto le due tipologie di spese prima per le ASL e poi per le Aziende Ospedaliere, e in seguito il grafico 5.1 che mette in evidenza la percentuale di spesa coperta dagli ospedali del dataset rispetto alla spesa a rendiconto.

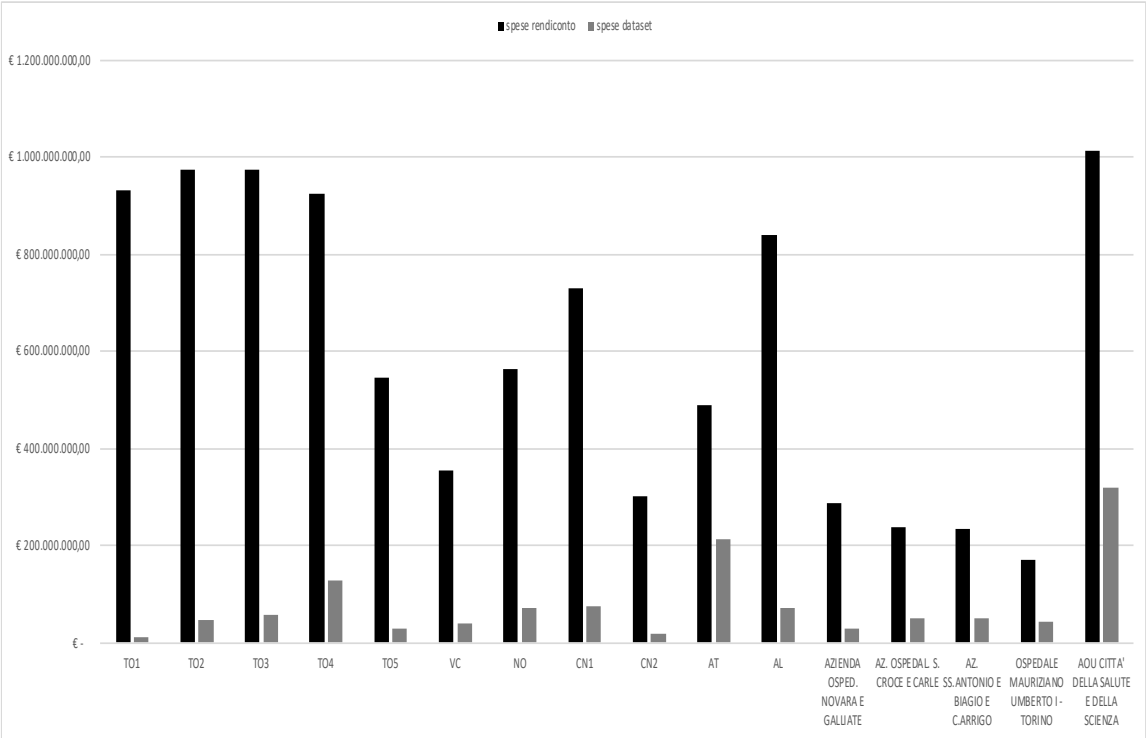

Figura 5.1. Le barre nere si riferiscono alle spese del rendiconto finanziario del 2015, mentre le barre grigie si riferiscono alle spese del dataset dello stesso anno

Nella figura 5.1 si nota come le spese rendicontate delle ASL sono in media molto maggiori di quelle delle Aziende Ospedaliere (tranne per quanto riguarda L'Azienda Ospedaliero Universitaria Città della salute e della Scienza di Torino). In percentuale, la spesa del dataset copre solo circa il 13% di quella a rendiconto.
Il trend delle spese del rendiconto rispecchia il trend delle spese del dataset tranne per degli outlier: l'ASL di Torino 4 (TO4), l'ASL di Alessandria (AL) e l'Azienda Sanitaria Universitaria Città della Salute e della Scienza di Torino (Salute e Scienza).
Di seguito è riportato il grafico che mostra la percentuale di spesa coperta dal dataset rispetto a quella a rendiconto, raggruppata per ASL e Azienda Ospedaliera.

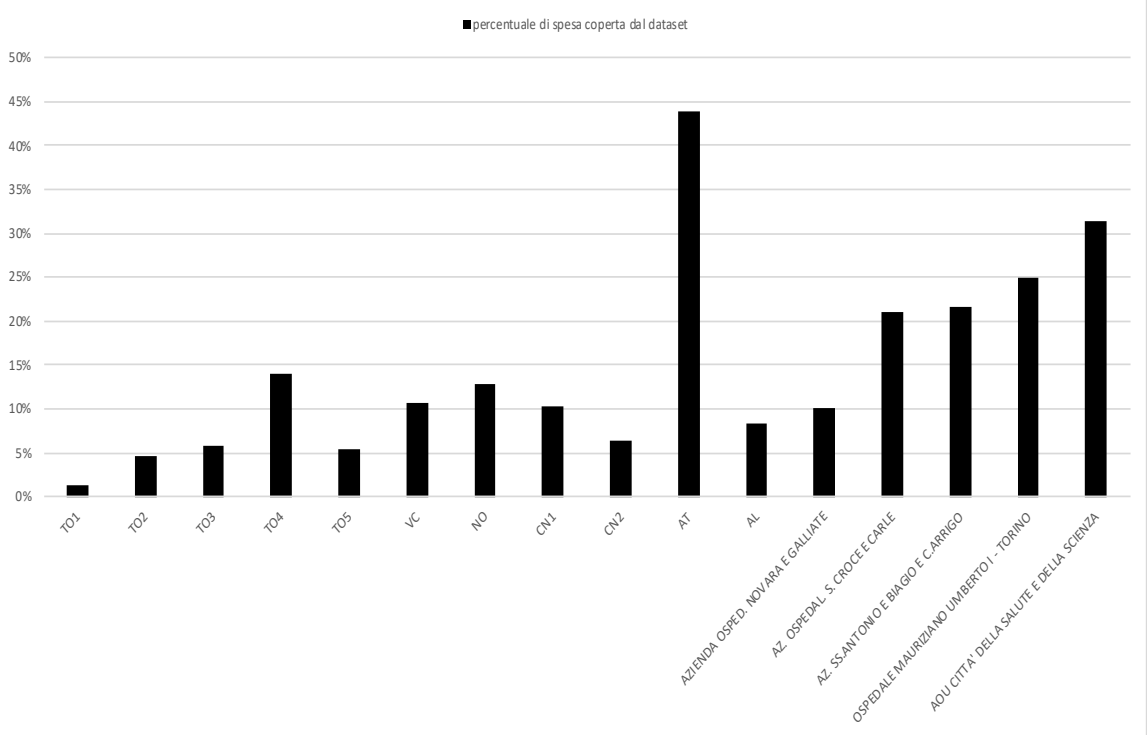

Figura 5.2. Ogni barra indica la percentuale di spesa coperta da ogni ospedale nel dataset rispetto alla spesa a rendiconto nell'anno 2015

Nel grafico 5.2 si nota che la percentuale di spesa coperta dalle ASL è molto minore di quella coperta dalle Aziende Ospedaliere, a eccezione dell'ASL di Asti e dell'Azienda Ospedaliera Novara e Galliate. In particolare, la spesa del dataset delle ASL non supera il 15% rispetto alla spesa a rendiconto, mentre quella delle Azienda Ospedaliera parte dal 20% fino ad arrivare al 30%. I due outlier sono la ASL di Asti copre il 43% della spesa mentre l'Azienda Ospedaliera di Novara e Galliate copre il 10%.

Anno 2016

Spostiamo ora la nostra attenzione sull'anno 2016, per il quale procederemo nello stesso modo con cui abbiamo fatto per l'anno 2015. Nel grafico 5.3 è rappresentato il confronto fra le spese del dataset e quelle a rendiconto.

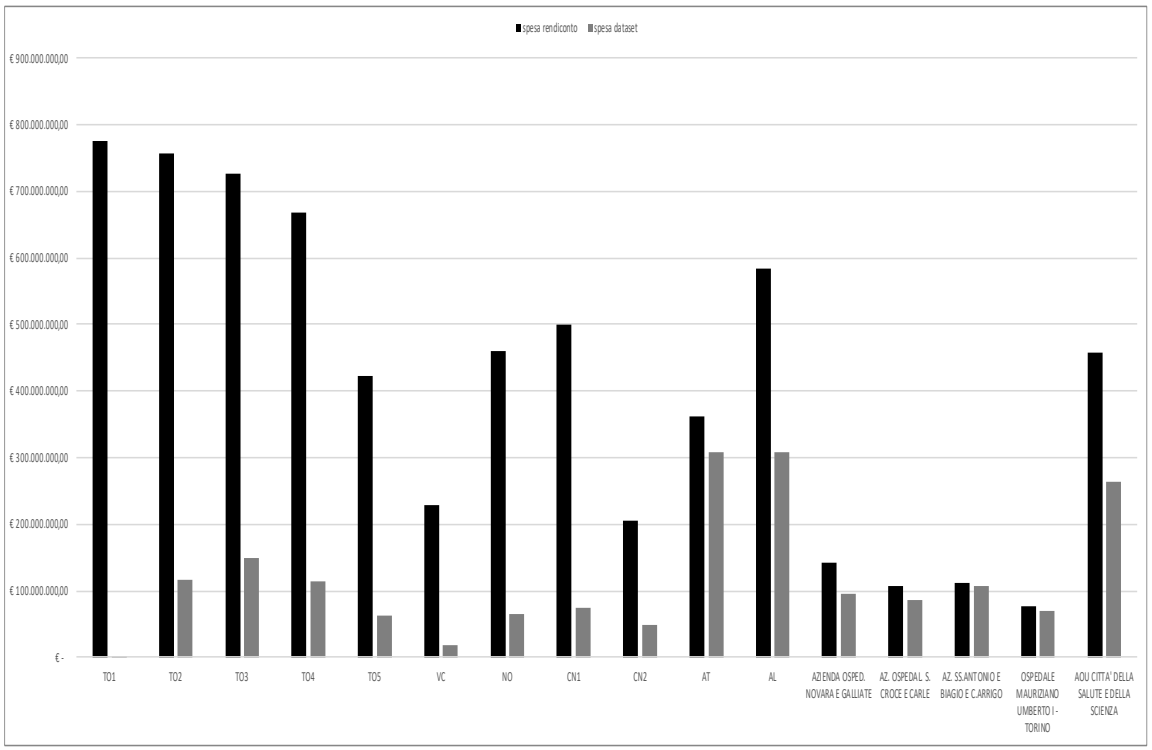

Figura 5.3. Le barre nere si riferiscono alle spese del rendiconto finanziario del 2016, mentre le barre grigie si riferiscono alle spese del dataset dello stesso anno

La spesa totale a rendiconto nel 2016 ammonta a 6.006.205.000 €, mentre quella del dataset si attesta a 1.577.414.744 €. Se nell'anno 2015 il dataset copriva il 13% della spesa che troviamo a rendiconto, nel 2016 la copertura arriva a coprire circa il 26% (circa il doppio).
Per quanto riguarda il trend, le spese degli ospedali del dataset seguono abbastanza bene l'andamento delle spese a rendiconto fatta eccezione per le ASL di Torino. Concentriamoci ora sulla copertura delle spese del dataset rispetto a quella a rendiconto (figura 5.4).

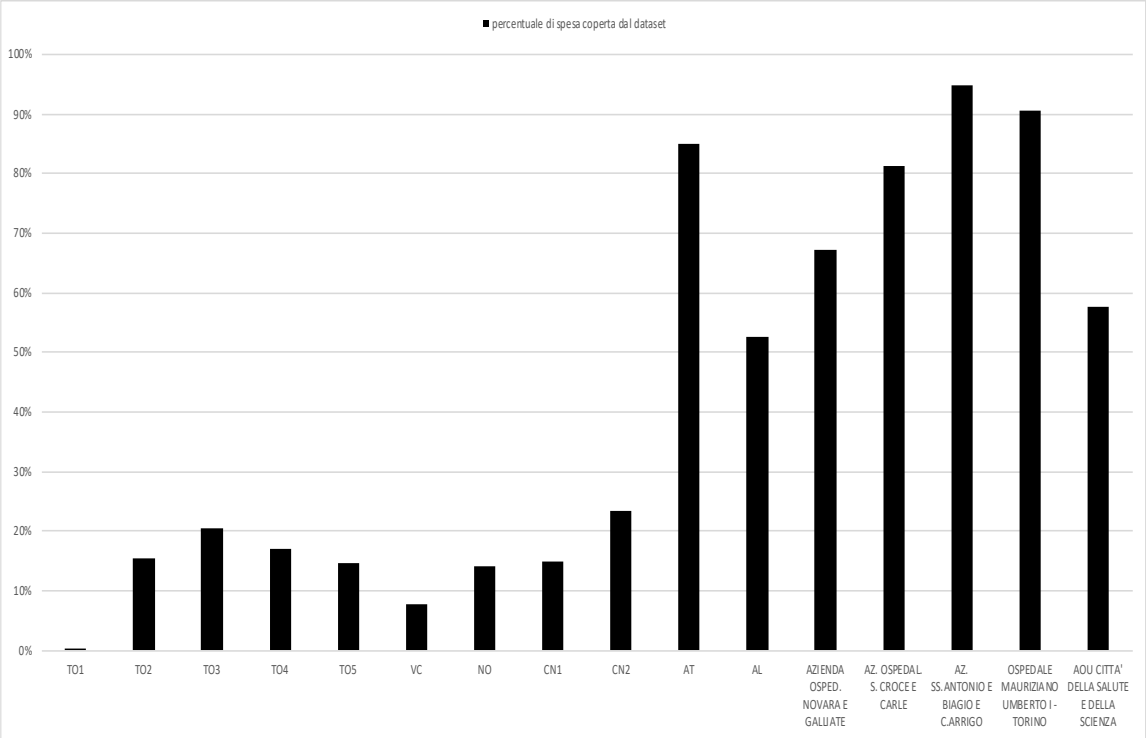

Figura 5.4. Ogni barra indica la percentuale di spesa coperta da ogni ospedale nel dataset rispetto alla spesa a rendiconto nell'anno 2016

In generale vale la stessa considerazione fatta per l'anno 2015, ovvero che la percentuale di spesa coperta dalle ASL è minore rispetto a quella coperta dalle Aziende Ospedaliere, fatta eccezione per l'ASL di Asti.
Confrontiamo la percentuale di spesa coperta nel 2015 con quella coperta nel 2016 per fare ulteriori considerazioni (figura 5.5) .

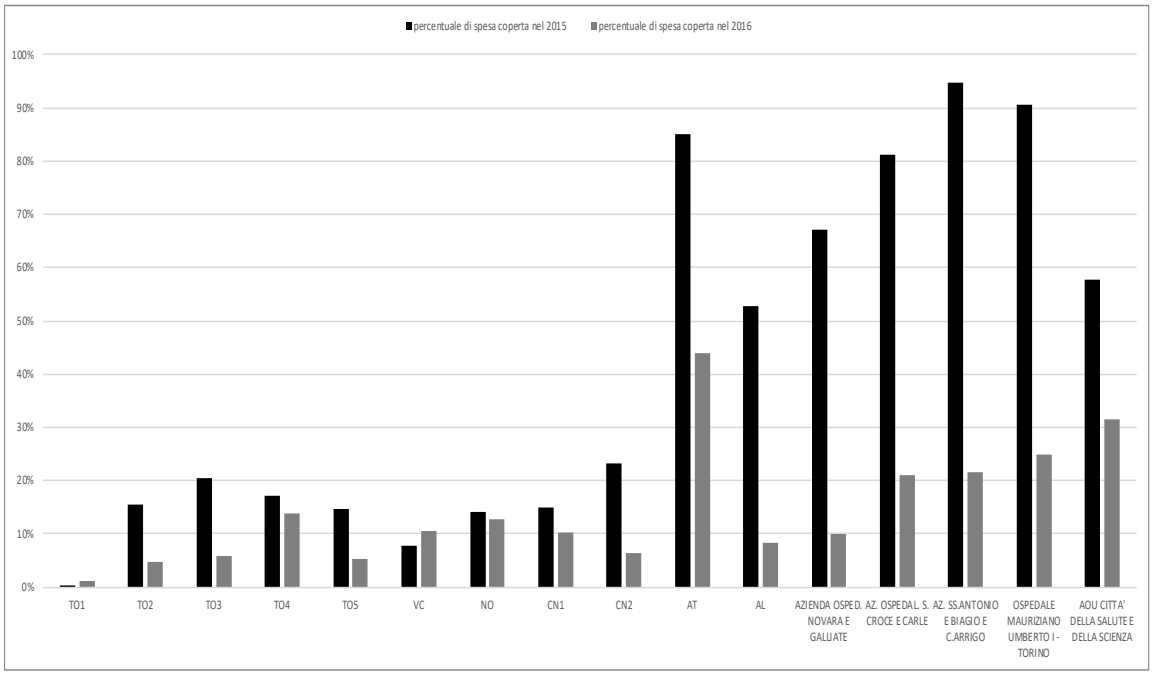

Figura 5.5. Le barre nere indicano la percentuale di spesa coperta dagli ospedali del dataset rispetto a quella del rendiconto nel 2015, mentre le barre grigie si riferiscono all'anno 2016

Si nota immediatamente che la spesa coperta nel 2015 è minore rispetto a quella coperta nel 2016, fatta eccezione per l'ASL di Vercelli (VC), Novara (NO) e di Torino 1 (TO1). Se la spesa coperta dalle ASL è cresciuta di circa 3-4 punti percentuali (fatta eccezione per gli outliers), lo stesso non si può dire per le Aziende Ospedaliere: la crescita è stata di circa il 30-40%. Nelle successive sezioni in cui verranno analizzati in maniera più approfondita i singoli anni si cercherà di dare risposta a questi risultati.
Per quanto riguarda il trend, non sembra esserci relazione fra i due anni.
In generale, i risultati sembrano affermare che il dataset su cui si stanno effettuando queste analisi sia un campione rappresentativo di circa il 20-25% delle spese in media, ma a livello quantitativo le spese non sono proporzionali nei due dataset, tranne per 2-3 ospedali nel solo anno 2016. Per quanto riguarda l'anno 2017 non è possibile fare alcun confronto con le spese a rendiconto. Proseguiamo la nostra analisi concentrandoci sul raggruppamento delle spese per anno e categoria merceologica.

5.2 Analisi delle spese per categoria CPV, triennio 2015-2017

In questa sezione proseguiamo le analisi cercando di dare risposta a certe domande, che verranno poste a inizio di ogni paragrafo. Utilizziamo come precisione la terza cifra della categoria cpv, come già anticipato in precedenza.

5.2.1 Come sono distribuite le spese maggiori negli anni 2015-2017?

La query eseguita per ottenere i risultati in figura 5.6 è la seguente:

```
PREFIX : <https://w3id.org/italia/onto/PublicContract/>
PREFIX jarql: <http://jarql.com/>
PREFIX l0: <https://w3id.org/italia/onto/l0/>
PREFIX covapit: <https://w3id.org/italia/onto/COV/>
PREFIX test: <http://test.yo/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

select ?categoria_cpv (sum(?importoAggiudicazione) as ?somma_importo)
where {

      ?lotto     a                                          :Lot;
                 l0:description              ?oggetto;
                 :actualStartDate            ?dataInizio;
                 :totalAmountPaid            ?importoSommeLiquidateFloat
                     ;
                 :isIncludedInProcedure ?URI_procedure;
                 :CIG ?cig;
               :hasAwardNotice   ?URI_award_notice.
      ?URI_award_notice :agreedAmount ?importoAggiudicazione.
      ?URI_procedure     :hasProcuringEntity ?URI_proponent_struct.
      ?URI_proponent_struct      covapit:legalName ?
         denominazioneStruttura.

      bind(strbefore(?oggetto, "__") as ?categoria_cpv)
      bind(strafter(str(?dataInizio), "_") as ?data)
      bind(strbefore(?data, "-") as ?anno)

      FILTER(
        (?annoANAC = "2015"
           ||
         ?annoANAC = "2016"
           || ?annoANAC = "2017")
        &&
        ?importoAggiudicazione >= 0
        &&
```

```
        (contains(?denominazioneStruttura, "ASL") || contains(?
            denominazioneStruttura, "Azienda"))
    )
}

group by ?categoria_cpv
order by desc (?tot)
```

I grafici 5.7 e 5.8 sono uno zoom del grafico precedente per evidenziare più in dettaglio il comportamento delle curve cumulate per un sottinsieme di categorie.

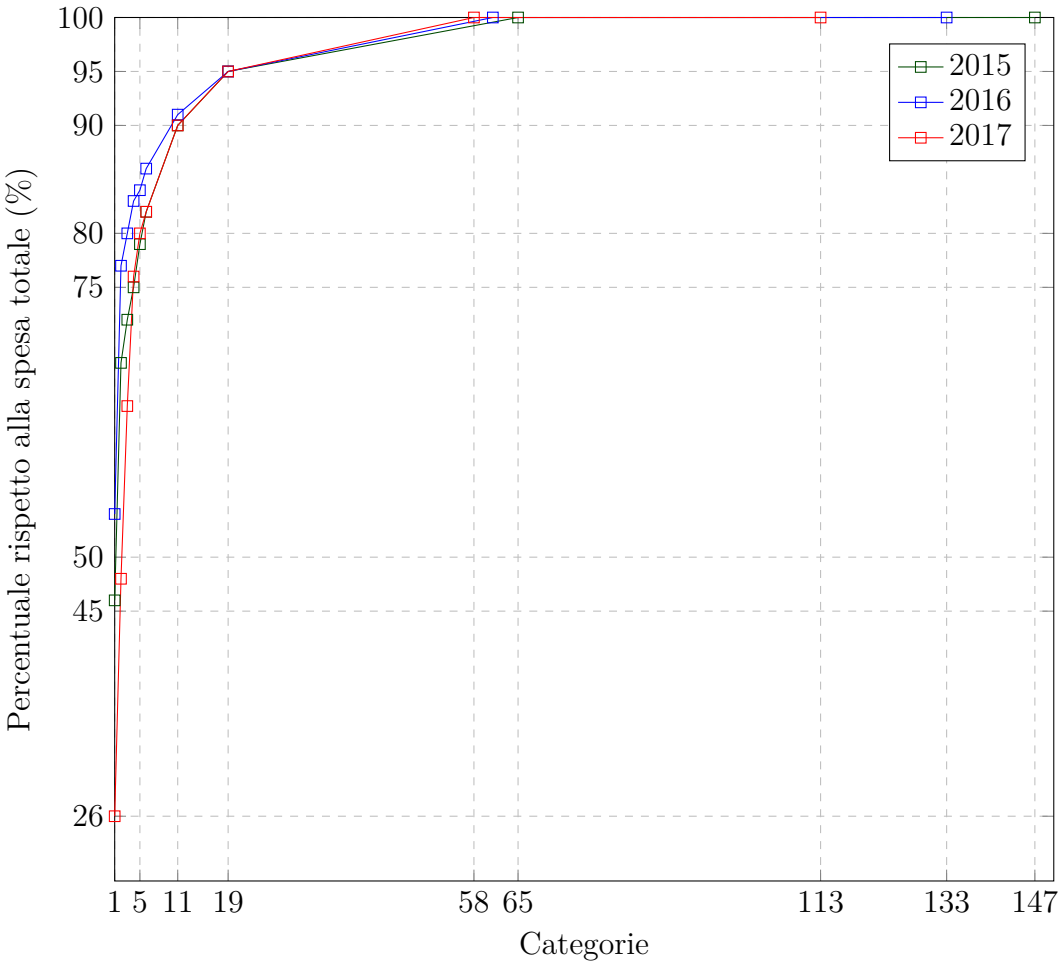

Figura 5.6. In verde la curva cumulata delle prime 147 categorie più costose e della loro spesa percentuale del 2015,in blu quella rispetto al 2016 e in rosso quella riferita all'anno 2017

Avendo scelto di usare la classificazione ANAC, abbiamo ristretto gli anni di

analisi al triennio 2015-2017, perciò senza applicare un filtro alle interrogazioni fatte sul database, i risultati si riferiranno a questi tre anni.
Raggruppando la spesa per anno e categoria merceologica CPV, e ordinando i risultati in ordine decrescente di prezzo, selezioniamo le categorie merceologiche e la loro l'importo totale speso in ognuna di esse. Questi risultati sono stati utilizzati per costruire un grafico avente sull'asse delle ascisse le categorie merceologiche ordinate in ordine decrescente, e sull'asse delle ordinate la percentuale di spesa coperta rispetto al totale della spesa del triennio. Questo tipo di grafico prende il nome di curva cumulata (sezione 2.5).Nel caso in esame, la funzione di distribuzione cumulata è usata per avere chiarezza sulla distribuzione della spesa. L'asse delle ascisse si riferisce alle categorie merceologiche ordinate dalla più costosa alla meno costosa (senza per ora precisare di quale categoria stiamo parlando) mentre sull'asse delle ordinate è rappresentata la spesa, in valore assoluto o come percentuale rispetto alla spesa totale. La pendenza della curva è una misura della velocità di crescita della spesa: più i costi sono assorbiti da una specifica categoria, più l'andamento sarà in ripida salita. Spostandosi a destra sull'asse delle ascìsse, ovvero passando alle categorie con costi inferiori, si può notare una crescita della spesa totale più lenta che corrisponde a una pendenza inferiore della curva. Nel caso in cui la spesa sia espressa in percentuale, la distribuzione cumulata cresce fino a tendere al 100%, il valore massimo è raggiunto in corrispondenza dell'ultima categoria sull'asse x. Il fine è quello di capire qual è la fetta di spesa coperta dalle categorie più costose la distribuzione delle spese maggiori delle categorie merceologiche. Nei grafici seguenti 5.7 5.8 riassumiamo i risultati.

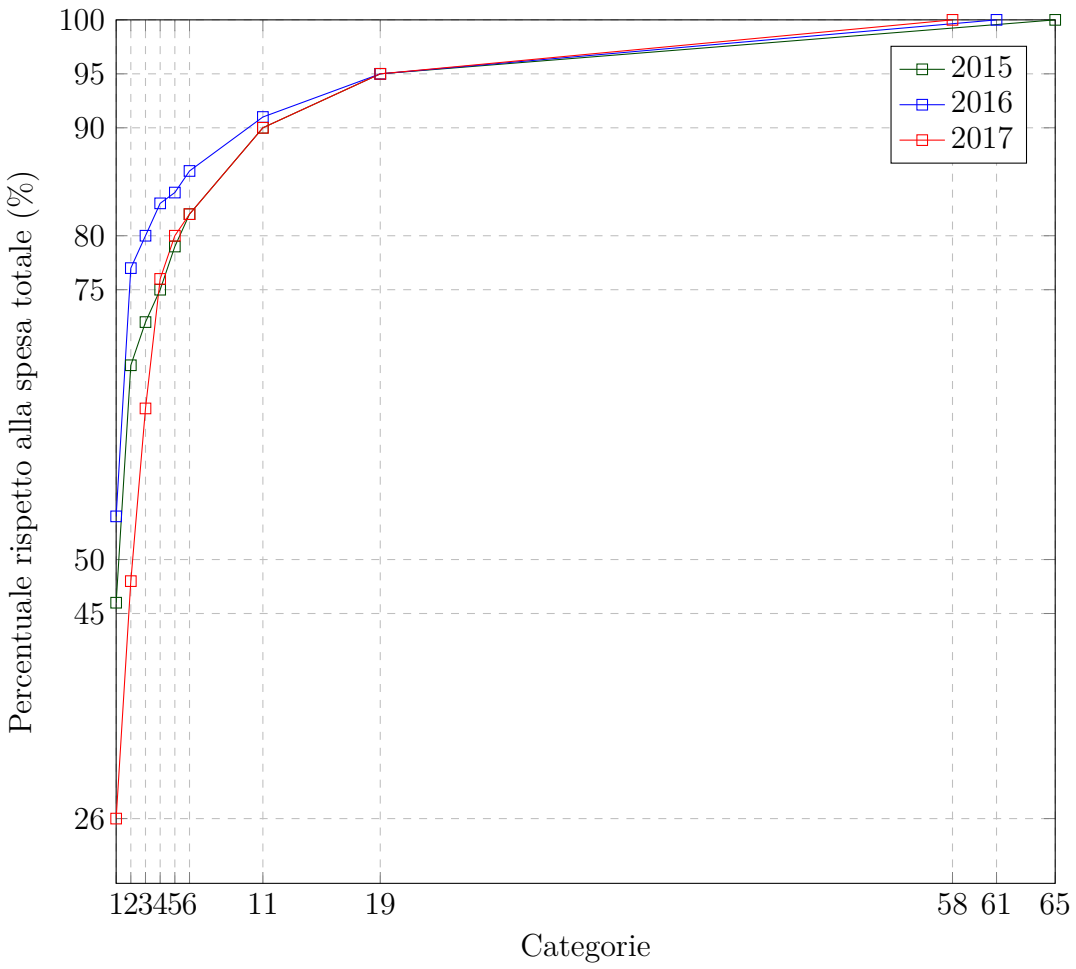

Figura 5.7. In verde la curva cumulata delle prime 65 categorie più costose e della loro spesa percentuale del 2015,in blu quella rispetto al 2016 e in rosso quella riferita all'anno 2017

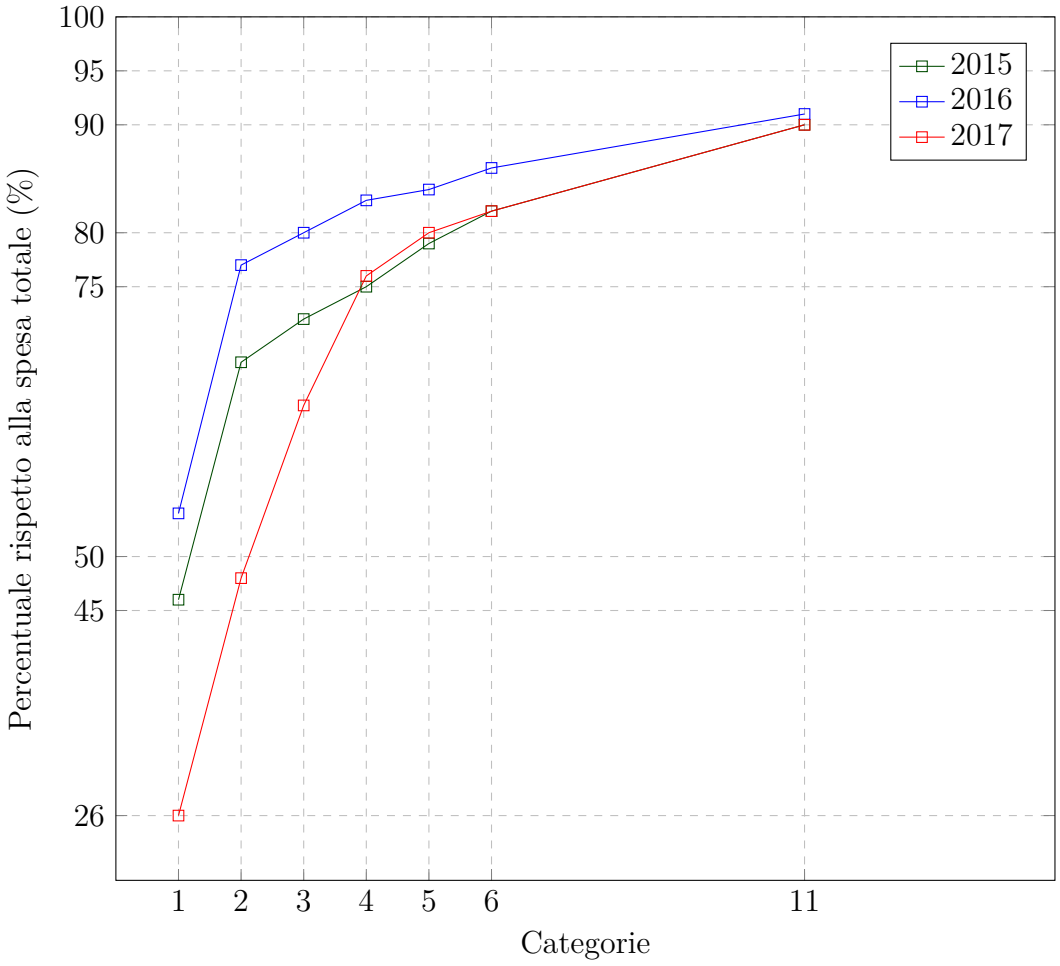

Figura 5.8. In verde la curva cumulata delle prime 11 categorie più costose e della loro spesa percentuale del 2015,in blu quella rispetto al 2016 e in rosso quella riferita all'anno 2017

Il trend generale delle tre curve dei tre anni è molto simile come si può vedere nelle figure 5.6 e 5.7. In particolare sono molto simili le cumulate degli anni 2015 e 2016, nelle quali le prime 3 categorie racchiudono più spesa rispetto a quella del 2017 (come si vede meglio in figura 5.8). A partire dalla 4a categoria le curve del 2015 e del 2017 arrivano quasi a sovrapporsi, restando però al di sotto di quella del 2016 che fino all'11a categoria racchiude in proporzione una maggior spesa rispetto alla totale annuale. Riassumiamo con qualche macronumero i risultati racchiusi nei 3 grafici appena analizzati:

L'informazione più significativa della tabella che si vuole evidenziare è che il 90% della spesa totale di ogni anno analizzato è racchiuso nelle prime 10 categorie cpv.

percentuale di spesa coperta	categorie 2015	categorie 2016	categorie 2017
50%	2	1	3
75%	4	2	4
90%	11	10	11
95%	19	18	18
100%	65	61	58

Tabella 5.1. In tabella sono riportate il numero di categorie CPV a livello 3 che coprono il 50, 75, 90, 95 e 100% della spesa

Nelle successive sezioni, quando andremo ad analizzare quali sono gli attori responsabili delle spese maggiori, ci concentreremo sulle prime 10 categorie merceologiche (sempre a livello 3) di ogni anno, in modo da non analizzare un numero troppo ampio e quindi dispersivo di categorie che non sarebbe riassuntivo e conciso.

5.2.2 Quali sono le 10 categorie del triennio in cui è racchiuso il 90% della spesa?

Le prime 10 categorie in ordine decrescente di importo degli anni 2015-2017 sono mostrate in questi due grafici, quello in figura 5.9 che mostra la loro distribuzione rispetto alla spesa totale del triennio, e 5.10 che mostra il nome della categoria e le corrispondente spesa.

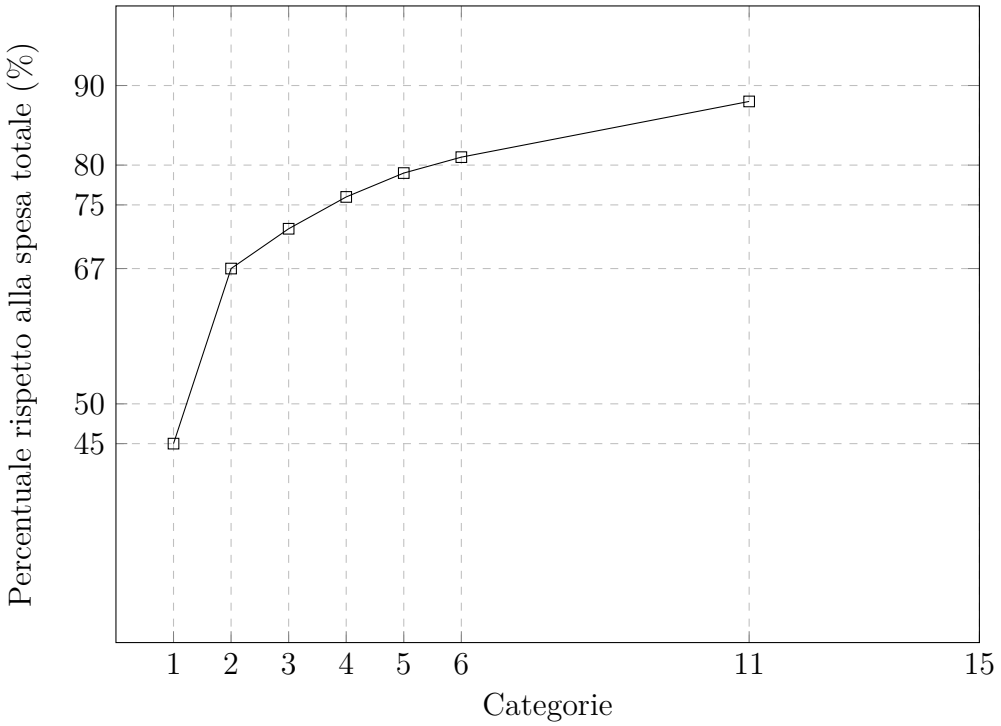

Figura 5.9. Curva cumulata delle prime 11 categorie più costose e della loro spesa percentuale rispetto al totale del triennio 2015-2017

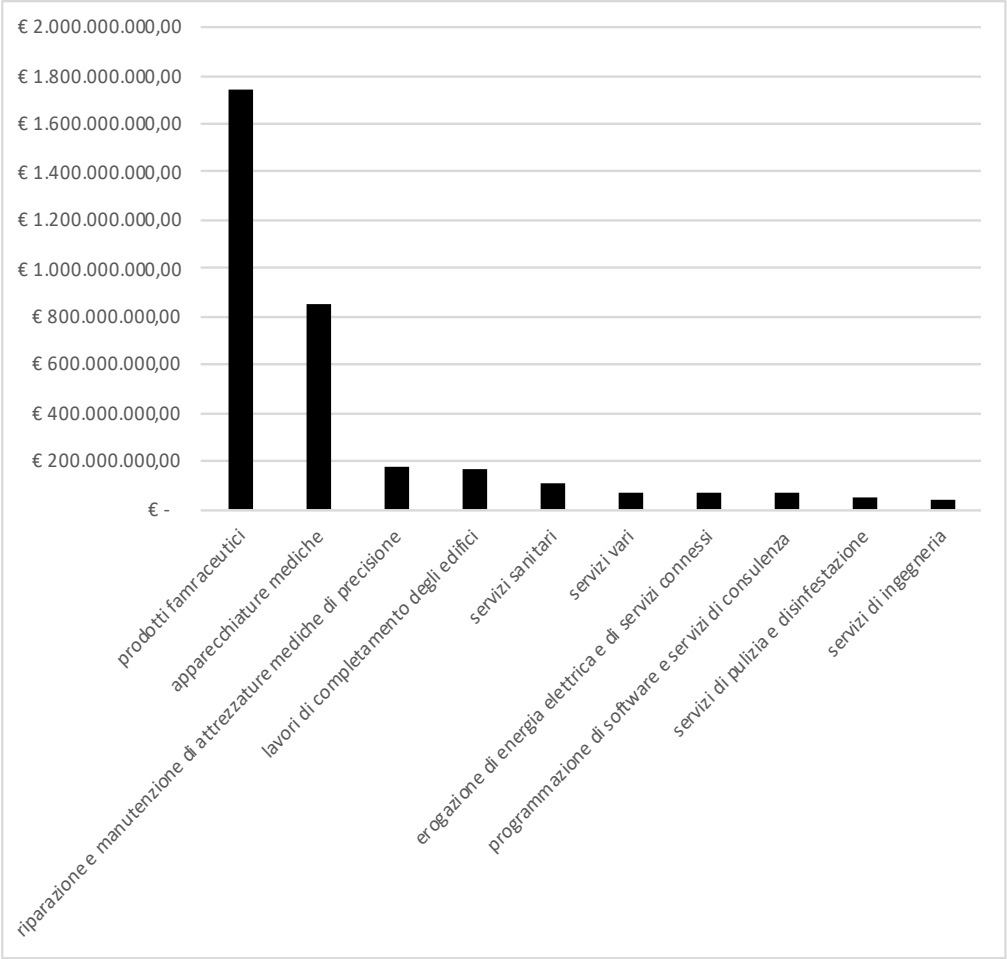

Figura 5.10. Sull'asse delle ascisse abbiamo le 10 categorie merceologiche ordinate in ordine decrescente di prezzo, sull'asse delle ordinate la corrispondente spesa nel triennio 2015-2017

La categoria con spesa di maggior spicco è quella per *prodotti farmaceutici* che racchiude il 45% della spesa totale del triennio, a seguire la categoria delle *apparecchiature mediche* in cui troviamo il 22% della spesa totale. Al terzo posto troviamo la categoria *servizi di riparazione e manutenzione di attrezzature mediche e di precisione* con il 6% dell'importo totale speso nel triennio; seguono le categorie di *lavori di completamento degli edifici (4%), servizi sanitari (3%), servizi vari (2%), erogazione di energia elettrica e servizi connessi (2%), programmazione di software e servizi di consulenza (2%), servizi di pulizia e disinfestazione (1%) e infine servizi di ingegneria (1%).* Notiamo che l'ultima categoria delle prime 10 contribuisce solo all'1% della spesa totale, come tutte quelle a seguire.

5.2.3 Quali sono gli ospedali che spendono di più nelle categorie che racchiudono il 90% della spesa?

In figura 5.11 troviamo la spesa totale degli ospedali per le categorie considerate. Gli ospedali sono ordinati da sinistra a destra in ordine decrescente di prezzo.
La query utilizzata per ottenere questi risultati è la seguente:

```
PREFIX : <https://w3id.org/italia/onto/PublicContract/>
PREFIX jarql: <http://jarql.com/>
PREFIX l0: <https://w3id.org/italia/onto/l0/>
PREFIX covapit: <https://w3id.org/italia/onto/COV/>
PREFIX test: <http://test.yo/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>


select ?denominazione_ospedale (sum(?importoAggiudicazione) as ?
    somma_importo)

where {
      ?lotto     a                                          :Lot;
                 l0:description              ?oggetto;
                 :actualStartDate            ?dataInizio;
                 :totalAmountPaid            ?importoSommeLiquidate;
                 :isIncludedInProcedure ?URI_procedure;
                 :CIG ?cig;
               :hasAwardNotice   ?URI_award_notice.

      ?URI_award_notice :agreedAmount ?importoAggiudicazione.
      ?URI_procedure     :hasProcuringEntity ?URI_proponent_struct.
      ?URI_proponent_struct      covapit:legalName ?
          denominazione_ospedale.

      bind(strbefore(?oggetto, "__") as ?categoria_cpv)
      bind(strafter(str(?dataInizio), "_") as ?data)
      bind(strbefore(?data, "-") as ?anno)

      FILTER(
        ( ?annoANAC = "2015" || ?annoANAC = "2016" || ?annoANAC =
            "2017"  )
        &&
        ?importoAggiudicazione >= 0
        &&
        ( contains(?denominazione_ospedale, "ASL") || contains(?
            denominazione_ospedale, "Azienda")  )
      )
   }
group by ?denominazione_ospedale
order by desc (?tot)
```

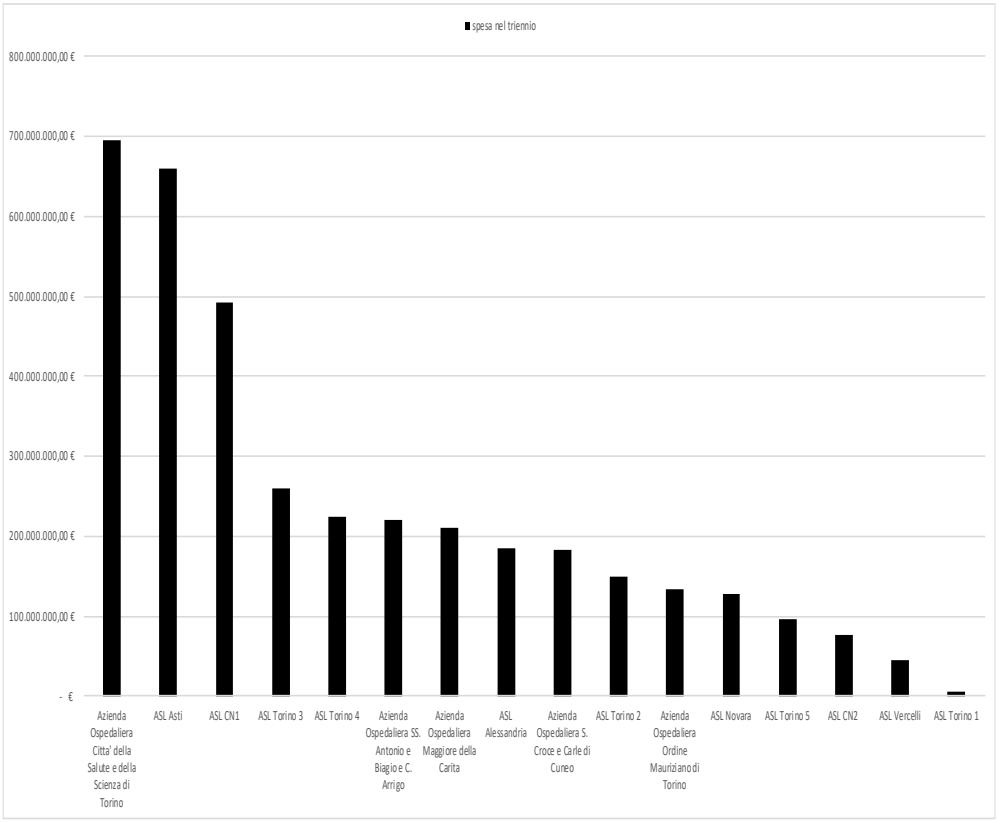

Figura 5.11. Sull'asse delle ascisse troviamo gli ospedali presenti nel dataset ordinati in ordine decrescente di spesa per le categorie che racchioudono il 90% della spesa del triennio 2015-2017

Spiccano la ASL di Asti e Azienda Ospedaliero Universitaria città di Torino e ASL di CN1, mentre quelle che spendono di meno sono l'ASL di Vercelli e di Torino 1. Si può notare come ci sia una netta differenza di prezzo fra le prime 3 e il resto degli ospedali, e come l'ASL di Torino 1 sembra spendere pochissimo in confronto alla penultima per spesa.

5.2.4 Il trend di spesa nel triennio è in crescita?

Raggruppando la totalità delle spese per anno si evidenzia il trend della spesa totale per ogni anno (grafico 5.12). La query eseguita raggruppa tutte le spese soltanto per anno filtrando a turno sull'anno 2015, 2016 e 2017. I dati sono stati poi messi a confronto per costruire l'istogramma seguente:

```
PREFIX : <https://w3id.org/italia/onto/PublicContract/>
PREFIX jarql: <http://jarql.com/>
PREFIX l0: <https://w3id.org/italia/onto/l0/>
```

```
PREFIX covapit: <https://w3id.org/italia/onto/COV/>
PREFIX test: <http://test.yo/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

select ?anno (sum(?importoAggiudicazione) as ?somma_importo)

where {
      ?lotto     a                                          :Lot;
                 l0:description                ?oggetto;
                 :actualStartDate              ?dataInizio;
                 :totalAmountPaid              ?importoSommeLiquidate;
                 :isIncludedInProcedure ?URI_procedure;
                 :CIG ?cig;
                :hasAwardNotice   ?URI_award_notice.

      ?URI_award_notice :agreedAmount ?importoAggiudicazione.
      ?URI_procedure     :hasProcuringEntity ?URI_proponent_struct.
      ?URI_proponent_struct      covapit:legalName ?
          denominazione_ospedale.

      bind(strbefore(?oggetto, "__") as ?categoria_cpv)
      bind(strafter(str(?dataInizio), "_") as ?data)
      bind(strbefore(?data, "-") as ?anno)

      FILTER(
        ( ?annoANAC = "2015" ) # 2016 e 2017 nelle query successive
        &&
        ?importoAggiudicazione >= 0
        &&
        ( contains(?denominazione_ospedale, "ASL") || contains(?
            denominazione_ospedale, "Azienda")  )
      )
   }
group by ?anno
```

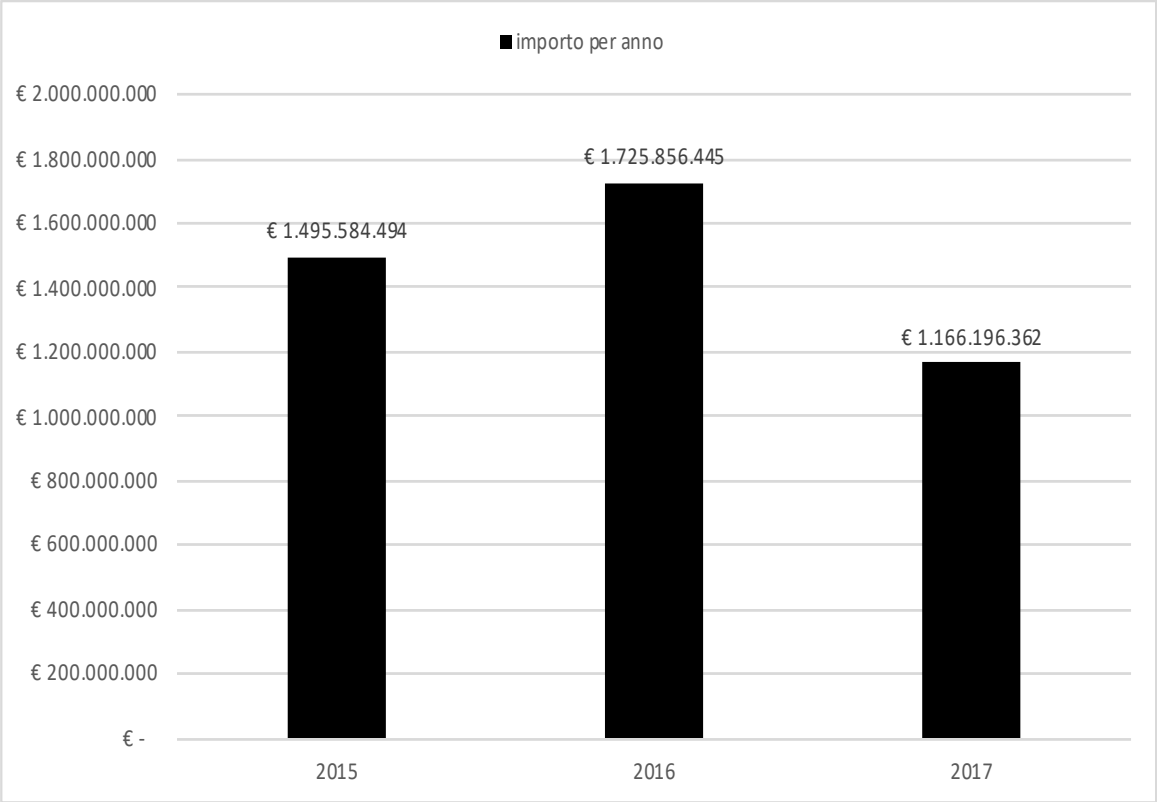

Figura 5.12. Le barre indicano ognuna la spesa totale di ognuno dei tre anni analizzati.

Nel 2015 si è speso circa 1 miliardo e mezzo, nel 2016 la spesa è aumentata ma nel 2017 è diminuita di poco più di mezzo miliardo di € rispetto al 2016 e di 300 milioni di € rispetto al 2015. Per esprimerci in numeri percentuali, nel 2016 si è registrato un incremento della spesa tra il 13 e il 14%, mentre nel 2017 una diminuzione del 34% rispetto alla spesa del 2016 e del 23%.

5.2.5 Quali sono le spese in crescita nel 2016? E quali quelle in diminuzione nel 2017?

Spese in crescita nel 2016

Le spese in crescita nel 2016 responsabili dell'aumento della spesa rispetto all'anno precedente sono riportate nel grafico 5.13. I dati per ottenere questi risultati sono stati ottenuti grazie alla seguente query:

```
PREFIX : <https://w3id.org/italia/onto/PublicContract/>
```

```
PREFIX jarql: <http://jarql.com/>
PREFIX l0: <https://w3id.org/italia/onto/l0/>
PREFIX covapit: <https://w3id.org/italia/onto/COV/>
PREFIX test: <http://test.yo/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>


select ?categoria_cpv (sum(?importoAggiudicazione) as ?somma_importo)

where {
      ?lotto     a                                          :Lot;
                 l0:description              ?oggetto;
                 :actualStartDate            ?dataInizio;
                 :totalAmountPaid            ?importoSommeLiquidate;
                 :isIncludedInProcedure ?URI_procedure;
                 :CIG ?cig;
                :hasAwardNotice   ?URI_award_notice.

      ?URI_award_notice :agreedAmount ?importoAggiudicazione.
      ?URI_procedure     :hasProcuringEntity ?URI_proponent_struct.
      ?URI_proponent_struct      covapit:legalName ?
          denominazione_ospedale.

      bind(strbefore(?oggetto, "__") as ?categoria_cpv)
      bind(strafter(str(?dataInizio), "_") as ?data)
      bind(strbefore(?data, "-") as ?anno)

      FILTER(
        ( ?annoANAC = "2015" ) # "2016"
        &&
        ?importoAggiudicazione >= 0
        &&
        ( contains(?denominazione_ospedale, "ASL") || contains(?
            denominazione_ospedale, "Azienda")  )
        &&
        ( contains(?cpv, "336") || contains(?cpv, "331") || contains(?
            cpv, "504") || contains(?cpv, "454") || contains(?cpv,
            "851") || contains(?cpv, "983") || contains(?cpv, "653") ||
             contains(?cpv, "722") || contains(?cpv, "909") || contains
            (?cpv, "713") )
      )
   }
group by ?categoria_cpv
order by desc (?somma_importo)
```

Essa restituisce tutte le spese delle categorie CPV filtrate riferite all'anno filtrato. Eseguendo due volte questa interrogazione filtrando prime per l'anno 2015 e poi per l'anno 2016, e selezionando solo le spese che aumentano nel 2016, otteniamo i risultati rappresentati nel prossimo istogramma:

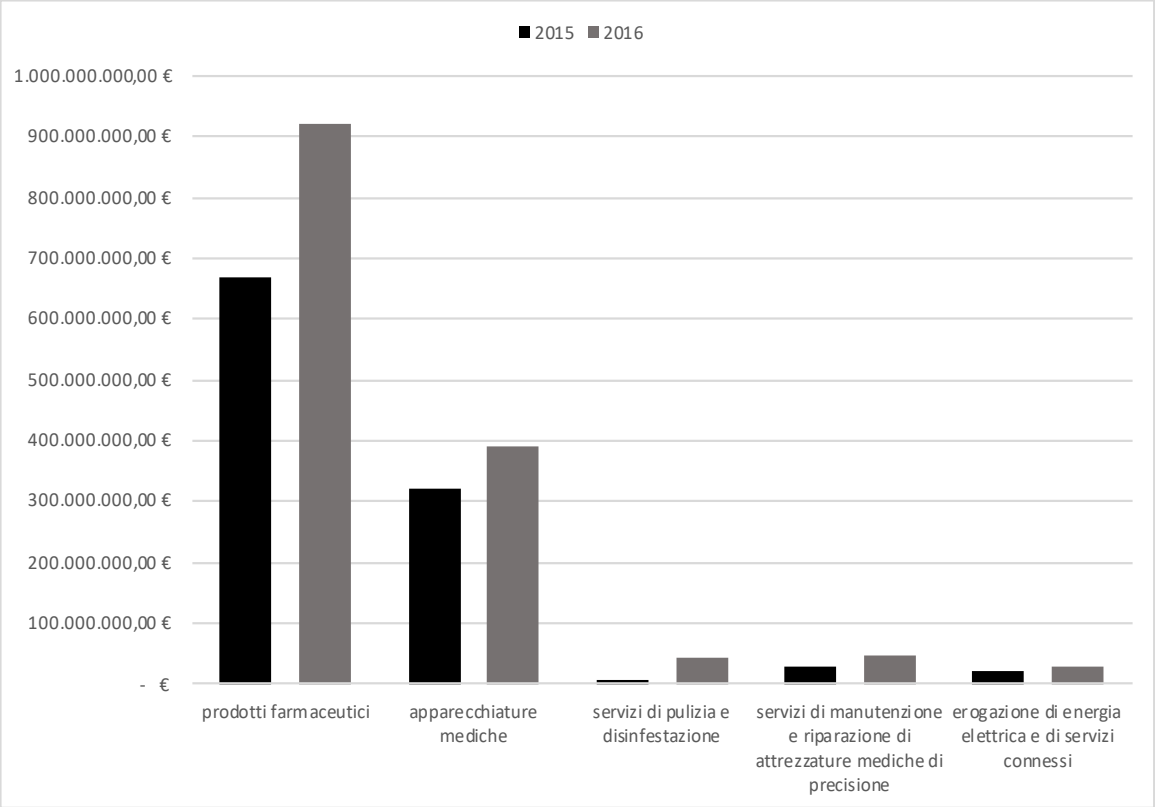

Figura 5.13. L' istogramma rappresenta il confronto delle spese del 2015 e del 2016 nelle categorie merceologiche che hanno aumentato la loro spesa.

Mettiamo in evidenza l'aumento della spesa delle categoria di figura 5.12 nella tabella 5.2.

Categoria	Delta spesa 2015-2016	Aumento %
prodotti farmaceutici	255.158.720 €	+38%
apparecchiature mediche	70.721.472 €	+22%
servizi di pulizia e disinfestazione	35.401.683 €	+500%
manutenzione di attrezzature mediche	19.727.174 €	+69%
erogazione di energia elettrica e servizi connessi	5.621.874 €	+24%

Tabella 5.2. Spese aumentate nel 2016 rispetto al 2015, focus sul delta e sulla percentuale di aumento.

Per riassumere, l'aumento di spese nel 2016 è dovuto sopratutto ai 255 milioni

di € della categoria dei prodotti farmaceutici e ai 70 milioni di € delle apparecchiature mediche, con un significativo +38% e +22% di spesa che pesa molto di più rispetto all'aumento delle altre categorie, dato che il valore della spesa dei prodotti farmaceutici e delle apparecchiature mediche è maggiore di 1 ordine di grandezza rispetto alle altre.

Spese in diminuzione nel 2017

Il procedimento per ottenere i dati del prossimo grafico è lo stesso usato per ottenere i dati del grafico 5.13 con la differenza che gli anni considerati sono il 2016 e il 2017 e che sono selezionate solo le spese che rispetto al 2016 subiscono una diminuzione. Nel grafico5.14 osserviamo le categorie per cui si è speso di meno nel 2017:

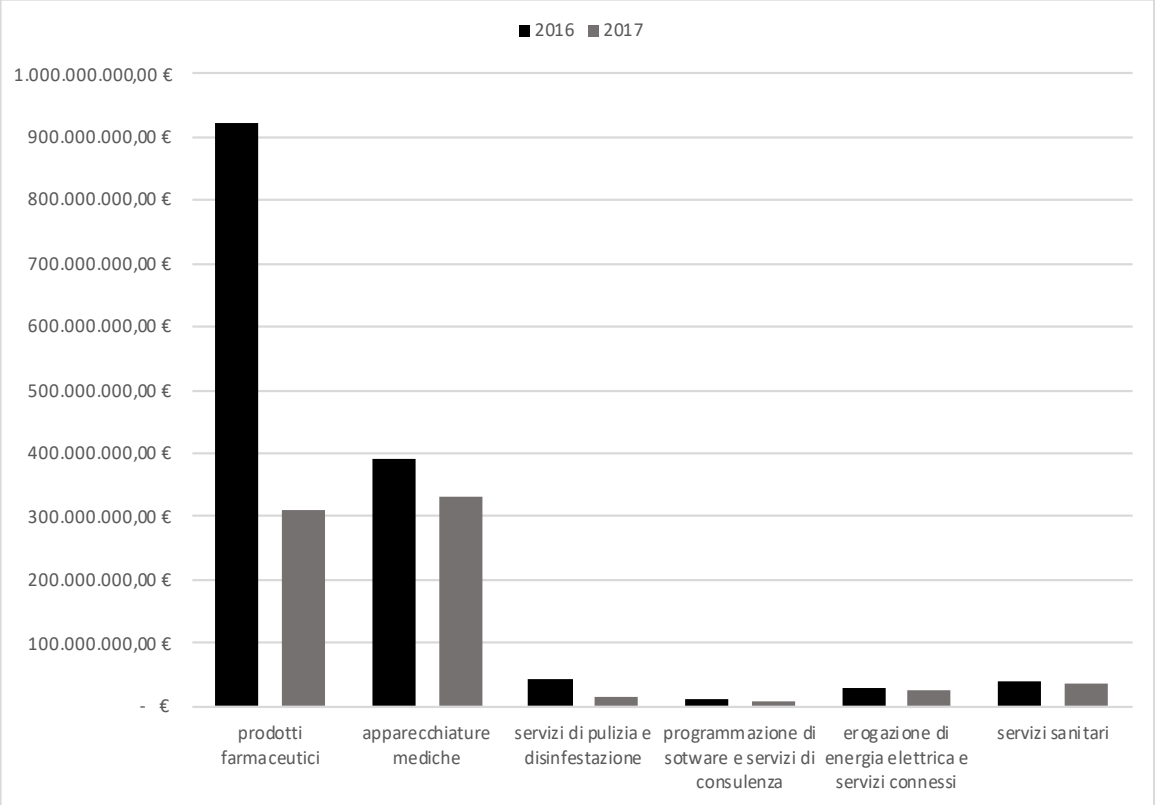

Figura 5.14. L' istogramma rappresenta il confronto delle spese del 2016 e del 2017 nelle categorie merceologiche che hanno diminuito la loro spesa.

Mettiamo in risalto il delta di spesa e la diminuzione percentuale nella tabella 5.3:

Categoria	Delta spesa 2016-2017	Diminuzione %
prodotti farmaceutici	611.646.780 €	-66%
apparecchiature mediche	61.238.400 €	-16%
servizi di pulizia e disinfestazione	28.584.308 €	-66%
software e servizi di consulenza	3.208.723 €	-32%
erogazione di energia elettrica e servizi connessi	3.153.280 €	-11%
servizi sanitari	2.920.856 €	-8%

Tabella 5.3. Spese diminuite nel 2017 rispetto al 2016, focus sul delta e sulla percentuale di diminuzione.

Le categorie che hanno registrato una diminuzione significativa della spesa sono: i prodotti farmaceutici, le apparecchiature mediche ed i servizi di disinfestazione, i cui delta di spesa sono nell'ordine delle decine e centinaia di milioni di euro, rispettivamente 611.646.780 €, 61.238.400 € e 20.584.308 €. Le altre 3 categorie in ribasso (la programmazione di software e servizi di consulenza, l'erogazione di energia elettrica e i servizi connessi e infine tutto ciò che riguarda i servizi sanitari) sono nell'ordine dei milioni di euro, e quindi hanno un peso molto minore nella diminuzione della spesa.

Notiamo come sia per quanto riguarda l'aumento della spesa nel 2016 e la diminuzione nel 2017 i prodotti farmaceutici e le apparecchiature mediche hanno un peso molto maggiore rispetto a tutte le altre categorie cpv.

5.2.6 Quali ospedali hanno speso di più nel 2016? E quali hanno speso meno nel 2017?

Nel prossimo grafico analizziamo le spese divise per anno e per ospedale, in modo da capire quali ospedali hanno speso di più nel 2016 e quali di meno nel 2017. La seguente query è stata utilizzata per ottenere i dati necessari:

```
PREFIX : <https://w3id.org/italia/onto/PublicContract/>
PREFIX jarql: <http://jarql.com/>
PREFIX l0: <https://w3id.org/italia/onto/l0/>
PREFIX covapit: <https://w3id.org/italia/onto/COV/>
PREFIX test: <http://test.yo/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

select ?denominazione_ospedale (sum(?importoAggiudicazione) as ?
    somma_importo)

where {
      ?lotto     a                                                :Lot;
```

```
                l0:description                ?oggetto;
                :actualStartDate              ?dataInizio;
                :totalAmountPaid              ?importoSommeLiquidate;
                :isIncludedInProcedure ?URI_procedure;
                :CIG ?cig;
              :hasAwardNotice  ?URI_award_notice.

    ?URI_award_notice :agreedAmount ?importoAggiudicazione.
    ?URI_procedure    :hasProcuringEntity ?URI_proponent_struct.
    ?URI_proponent_struct     covapit:legalName ?
        denominazione_ospedale.

    bind(strbefore(?oggetto, "__") as ?categoria_cpv)
    bind(strafter(str(?dataInizio), "_") as ?data)
    bind(strbefore(?data, "-") as ?anno)

    FILTER(
      ( ?annoANAC = "2015" ) # "2016" , "2017"
      &&
      ?importoAggiudicazione >= 0
      &&
      ( contains(?denominazione_ospedale, "ASL") || contains(?
          denominazione_ospedale, "Azienda")  )
     )
  }
group by ?denominazione_ospedale
order by desc (?denominazione_ospedale)
```

La query è stata eseguita tre volte ognuna filtrando con un anno diverso del triennio, in modo da poter poi confrontare i dati ottenuti e metterli nel prossimo istogramma 5.15:

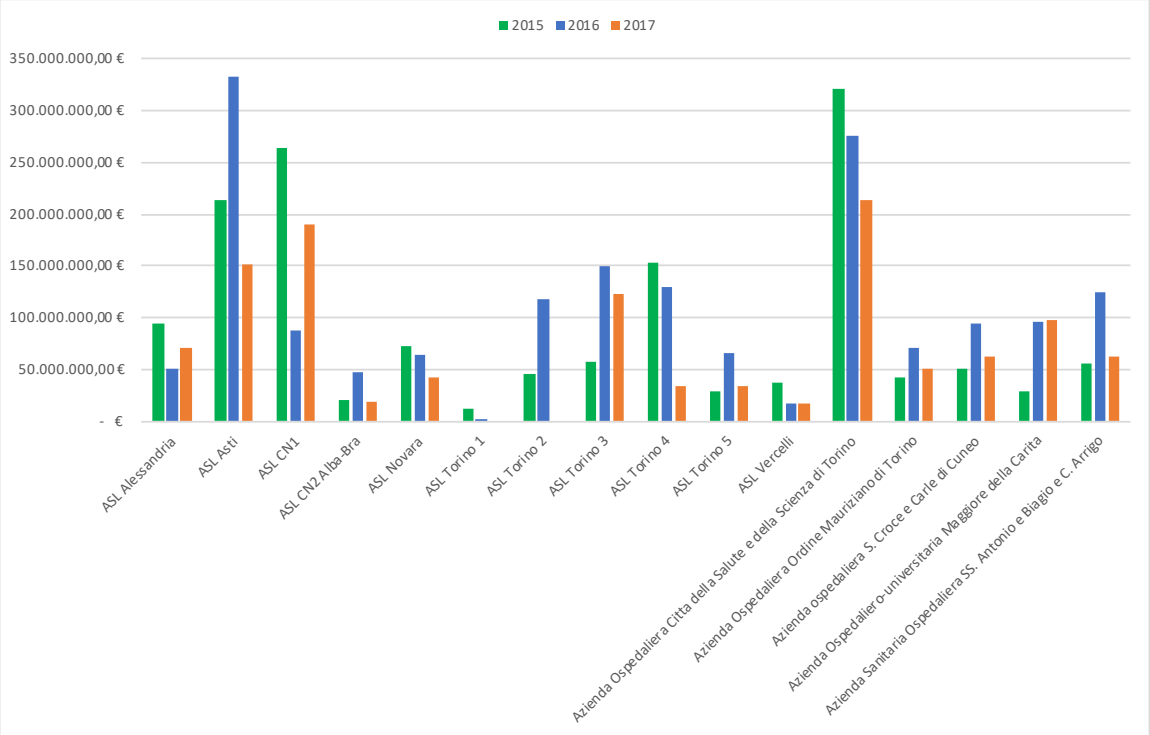

Figura 5.15. Confronto nel triennio 2015-2017 delle spese sostenute da ogni ospedale per ogni anno.

Nel grafico osserviamo che gli ospedali che hanno speso di più nel 2016 rispetto al 2015 sono l'ASL di Asti, Torino 2, Torino 3, l'Azienda Ospedaliero-Universitaria Maggiore della Carità e SS. Antonio e Biagio e Arrigo. Invece quelle che hanno diminuito di più la loro spesa nel 2017 rispetto all'anno precedente sono l'ASL di Asti, Torino 4, Torino 5 e l'Azienda Ospedaliera S.Croce e Carle di Cuneo, Città della Salute e della Scienza di Torino e SS. Antonio e Biagio e Arrigo.
Per quanto riguarda la diminuzione di spesa nel 2017 è evidentissimo come non siano registrate spese per le ASL di Torino 1 e 2, e sopratutto la spesa mancante di Torino 2 sembra incidere di circa 50-100 milioni di euro nel trend decrescente. Però il delta di spesa tra il 2016 e il 2017 è di circa 560 milioni di euro, quindi a meno che la spesa di Torino 1 e 2 complessivamente si aggiri intorno a quella cifra nel 2017, non basta a giustificare la diminuzione di spesa registrato.
Inoltre interessante notare come l'ASL di Asti sia l'ago della bilancia che pilota il trend della spesa nel triennio: nel 2016 ha aumentato la propria spesa di quasi 150 milioni di euro, mentre nel 2017 la ha diminuita quasi 200 milioni, il delta maggiore in assoluto registrato nelle analisi svolte. Gli altri ospedali che seguono il trend totale complessivo ma con un delta non così importante come quello dell'ASL di

Asti sono Torino 3, Torino 5, l'Azienda Ospedaliera Mauriziano di Torino, Croce e Carle di Cuneo e SS. Antonio e Biagio e C. Arrigo. Controcorrente invece le ASL di Cuneo 1 e Alessandria.
Un'altra considerazione che possiamo fare guardando questo grafico è che gli ospedali che spendono di più sono l'ASL di Asti, di CN1 e l'Azienda Ospedaliero-Universitaria Città della Salute e della Scienza di Torino. Quelli ad aver speso meno invece sono l'ASL di Torino 1 e di Torino 2, che sono anche le uniche due ad avere registrato spese proporzionali a quelle degli altri ospedali in soli due dei tre anni considerati.
Dopo aver valutato la spesa complessiva nel triennio e divisa per anno e per ospedale, raggruppiamo la spesa per categoria e continuiamo a condurre le analisi.

5.3 Analisi delle spese per categoria e ospedale

Considerando sempre gli anni 2015, 2016 e 2017, adesso analizziamo le spese sostenute da ogni ospedale per ogni anno. Per quanto riguarda quest'ultimi, viene posta attenzione a verificare se gli ospedali che hanno speso di più e di meno nel triennio (figura 5.15) sono le protagoniste in termini di spesa maggiore o minore anche nelle singole categorie che analizziamo. In particolare, gli ospedali ad aver speso di più sono l'ASL di Asti, quella di CN1, l'Azienda Sanitaria Città della Salute e della Scienza di Torino, mentre quelli ad aver speso di meno sono l'ASL di Torino 1 e 2 (che non registrano spese significative nel 2017) e e di Vercelli.
Analizziamo le categorie che rientrano nelle prime 10 a livello 3 del codice cpv in ordine di spesa sostenuta nel triennio e che sono presenti in tutti e 3 gli anni, in modo da poter osservare il loro trend. Inoltre le prime 10 sono rappresentative della spesa totale, infatti come già osservato nelle precedenti sezioni il 90% della spesa è racchiuso appunto nelle prime 10 categorie più costose a livello 3 cpv.

Le 10 categorie che consideriamo sono:

1. i **prodotti farmaceutici**, categoria che comprende tutti i tipi di medicinali, articoli di farmacia, vaccini, reattivi chimici;

2. le **apparecchiature mediche**, come i dispositivi per per terapie mediche, materiali medici (bisturi, guanti, etc.), apparecchiature da sala operatoria, strumenti da anestesia, strumenti per il sostegno funzionale come il macchinario per dialisi e defibrillatori, e infine i prodotti ad uso ospedaliero come letti medici e sedie a rotelle;

3. la **riparazione e manutenzione di attrezzature mediche di precisione**, che comprende la riparazione di tutte le attrezzature e i prodotti della categoria trattata al punto precedente;

4. i **lavori di completamento di edifici**, che include comprende lavori di intonacatura, falegnameria, installazione di porte e finestre, lavori di carpenteria, piastrellamento, pavimentazione, tinteggiatura, facciata, riparazione edilizia e ristrutturazione;

5. i **servizi sanitari** ovvero i servizi di riabilitazione offerti dalle strutture ospedaliere, i servizi specialistici medici ospedalieri come quelli polmonari e cardiaci, i servizi chirurgici specialistici, quelli di fisioterapia, quelli di ambulanza e di analisi mediche;

6. i **servizi vari**, cioè tutti i servizi di lavanderia, tintoria, portineria, servizi funerari, di barbiere e parrucchiere;

7. l' **erogazione di energia elettrica e servizi connessi**, ovvero le spese di elettricità;

8. la **programmazione di software e servizi di consulenza**, categoria che racchiude la programmazione di ogni genere di applicazione, servizi di archiviazione di dati e servizi per la manutenzione e la riparazione di applicativi;

9. i **servizi di pulizia e disinfestazione** di qualsiasi genere;

10. i **servizi di ingegneria**, categoria che comprende tutti i servizi di consulenza di ingegneria civile, i servizi per la gestione energetica degli impianti, i servizi di consulenza in ambito delle telecomunicazioni, di consulenza sanitaria e di sicurezza, servizi di progettazione di impianti di riscaldamento, di ventilazione e di impianti idraulici.

Le analisi sono svolte osservando dei grafici che si riferiscono ognuno ad una categoria merceologica diversa, con gli ospedali analizzati nel database sull'asse delle ascisse, e il valore della spesa sostenuta da ogni ospedale sull'asse delle ordinate. Ogni ospedale sull'asse x è caratterizzato da tre barre, ognuna delle quali si riferisce alla spesa per ognuno dei tre anni.
Infine, prima di passare ai numeri, forniamo lo scheletro delle query sparql che sono state utilizzate per interrogare il database:

```
select ?denominazione_ospedale (sum(?importoAggiudicazione) as ?
    somma_importo)

where {

      ?lotto     a                                          :Lot;
                 l0:description                ?oggetto;
                 :actualStartDate              ?dataInizio;
                 :totalAmountPaid              ?importoSommeLiquidateFloat
                     ;
```

```
                :isIncludedInProcedure ?URI_procedure;
                :CIG ?cig;

            :hasAwardNotice  ?URI_award_notice.

    ?URI_award_notice :agreedAmount ?importoAggiudicazione.

    ?URI_procedure     :hasProcuringEntity ?URI_proponent_struct.
    ?URI_proponent_struct      covapit:legalName ?
        denominazione_ospedale.

    bind(strbefore(?oggetto, "__") as ?cpv)
    bind(strafter(str(?dataInizio), "_") as ?dataANAC)
    bind(strbefore(?dataANAC, "-") as ?annoANAC)
    bind(strbefore(str(?dataInizio), "-") as ?anno190)
    bind(strdt(?annoANAC, xsd:integer) as ?annoANACint)
    bind(strdt(?anno190, xsd:integer) as ?anno190int)
    bind(strafter(?cpv, "_") as ?cpvL3)

    FILTER(
      (?annoANAC = "2015" || ?annoANAC = "2016" || ?annoANAC =
          "2017" )
      &&
      ?importoAggiudicazione >= 0
      &&
      ( contains(?denominazione_ospedale, "ASL") || contains(?
          denominazione_ospedale, "Azienda") )
      &&
      ( contains(?cpv, "336") )
    )
  }

group by ?denominazione_ospedale
order by desc (?denominazione_ospedale)
```

Questa query è stata ripetuta filtrando una per una ogni categoria merceologica di interesse.

5.3.1 Spesa per la categoria dei prodotti farmaceutici

La prima categoria in ordine di spesa è quella dei prodotti farmaceutici. Di seguito è riportato il grafico 5.3.1 per confrontare le spese sostenute da ogni ospedale per ogni anno:

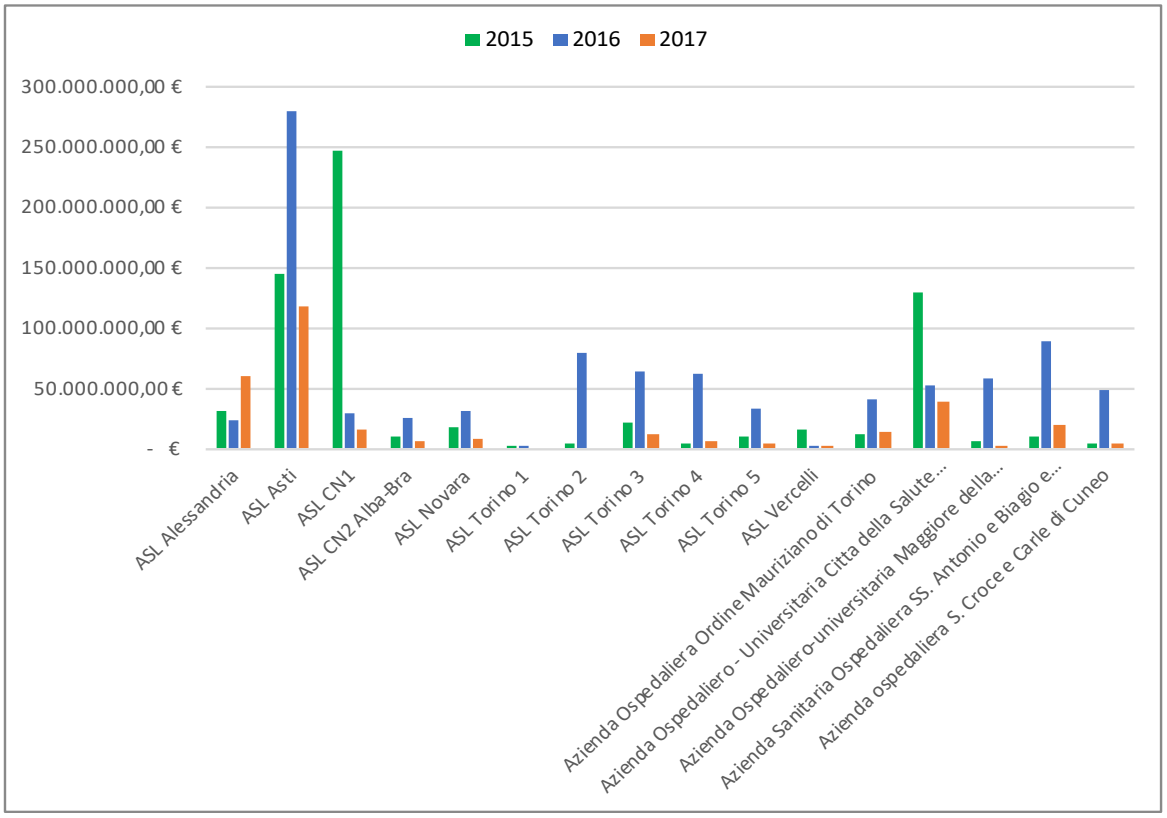

Figura 5.16. Categoria dei prodotti farmaceutici: confronto delle spese sostenute da ogni ospedale per ogni anno del triennio 2015-2017.

Gli ospedali che spendono di più per i prodotti farmaceutici sono l'ASL di Asti e CN1 e l'Azienda Ospedaliero-Universitaria Città della Salute e della Scienza di Torino: Asti spende quasi 150 milioni di euro nel 2015, circa 290 milioni nel 2016 per poi ridurre la spesa a 120 milioni nel 2017; Città della Salute di Torino invece registra spese a decrescere: 146 milioni di euro nel 2015, 50 milioni nel 2016 e 45 milioni nel 2017.
Questo risultato non ci sorprende: i prodotti farmaceutici in tutti e 3 gli anni comprendono circa il 50 % della spesa, tranne per il 2017 che raggruppa solo il 27%, e dato che Asti e Città della Salute e della Scienza nel triennio sono quelle che spendono di più, è logico anche che spendano di più nella categoria più costosa del triennio. Lo stesso discorso è possibile farlo per l'ASL di Torino 1 e di Vercelli, che sono quelle che in questo grafico hanno la spesa è più bassa di tutti. Torino 2 spende pochissimo nel 2015 mentre nel 2016 sostiene una spesa sopra la media, intorno agli 85 milioni di euro. L'ASL di Asti insieme alla maggior parte degli altri ospedali (esclusa Città della Salute e della Scienza di Torino) sostengono

delle spese che seguono il trend seguito dalla spesa totale annuale: nel 2016 si ha un forte aumento della spesa registrata, mentre nel 2017 una forte diminuzione rispetto ai due anni precedenti.

5.3.2 Spesa per la categoria di apparecchiature mediche

La seconda categoria che analizziamo è quella delle apparecchiature mediche. Nel prossimo grafico 5.3.2 riportiamo l'andamento delle spese:

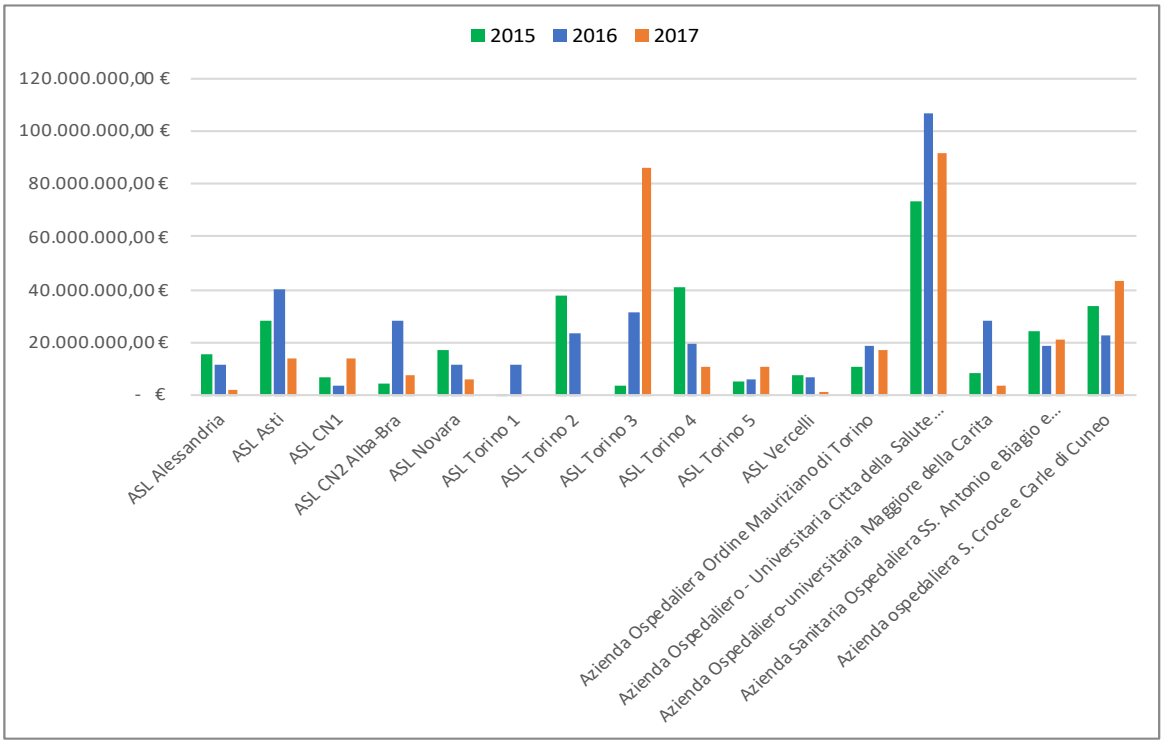

Figura 5.17. Categoria delle apparecchiature mediche: confronto delle spese sostenute da ogni ospedale per ogni anno del triennio 2015-2017.

In questa categoria merceologica l'ASL di Torino 3 e l'Azienda Ospedaliero-Universitaria Città della Salute e della Scienza di Torino sono quelle che spendono di più. Contrariamente a quanto ci si poteva aspettare l'ASL di Asti non rientra negli ospedali con maggiore spesa sostenuta, anche se la categoria delle apparecchiature mediche in media racchiude il un quarto della spesa totale annuale. Gli ospedali a spendere meno sono Torino 1 e Vercelli come ci si aspettava, accompagnati anche da Torino 5 che nel triennio registra spese totali appena sotto la media. L'Azienda Ospedaliero-Universitaria Città della Salute e della Scienza di Torino è

l'unica insieme all'ASL di Asti a seguire il trend seguito anche dalla spesa complessiva annuale. Gli altri ospedali pareggiano o diminuiscono la spesa per apparecchiature mediche nel 2016. Nel 2017 invece circa la metà la aumenta e la metà la diminuisce.

5.3.3 Spesa per la categoria di riparazione e manutenzione di attrezzature mediche di precisione

Di seguito riportato l'istogramma 5.18 per la terza categoria in ordine di spesa:

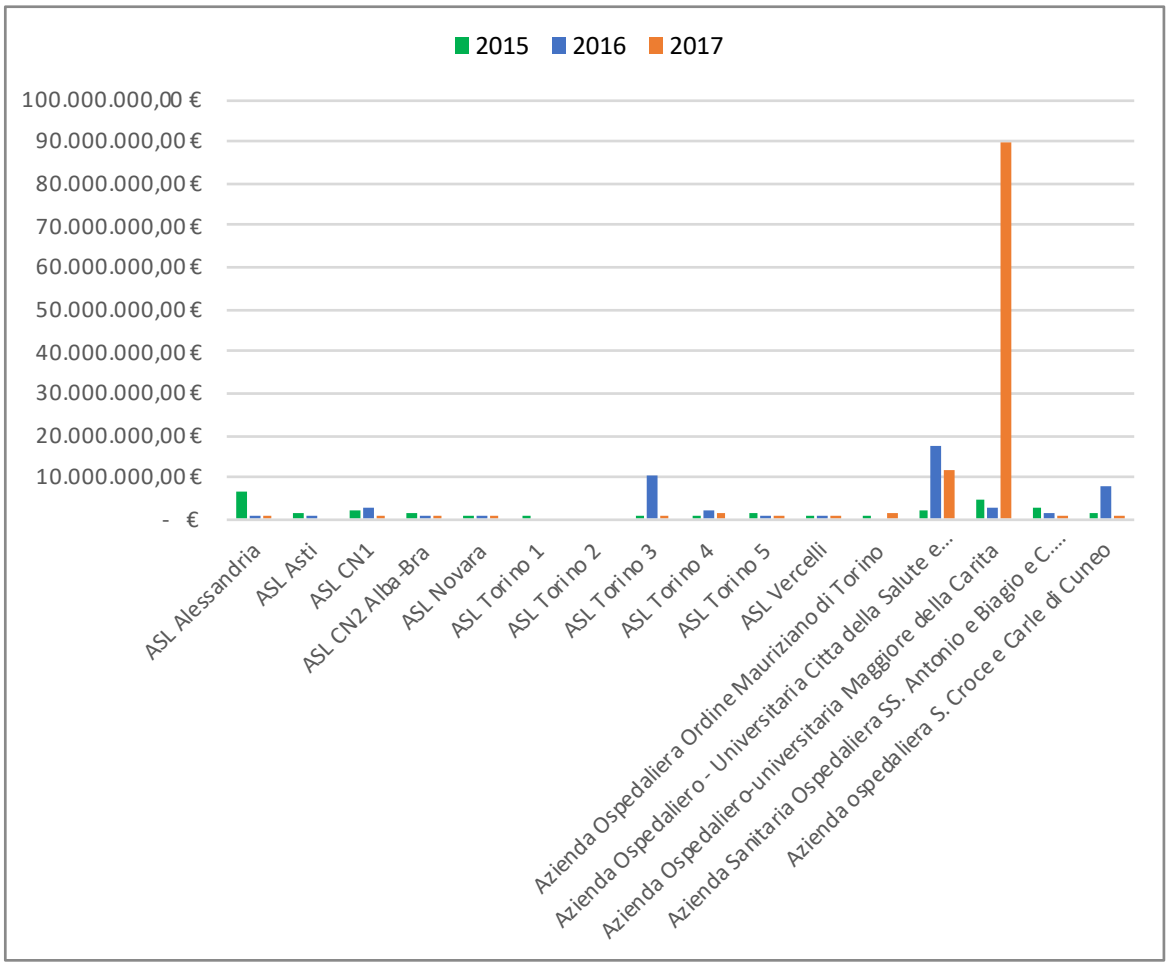

Figura 5.18. Categoria della manutenzione e riparazione delle apparecchiature mediche di precisione: confronto delle spese sostenute da ogni ospedale per ogni anno del triennio 2015-2017.

La variabilità della spesa sostenuta è alta, quindi facciamo un'analisi ad alto livello in questa prima parte e poi in un successivo grafico 5.19 risaltiamo le spese minori.
Questa grande variabilità è causata dalla spesa sostenuta dell'Azienda Ospedaliero-Universitaria Maggiore della Carità che nel 2017 spende 89 milioni di euro. E' un dato strano e anomalo per due motivi: il primo è che questo ospedale sui 16 analizzati è il settimo in ordine di spesa, che è 3 - 4 volte più bassa rispetto alle prime 4; il secondo motivo è che negli anni 2015-2016 le spese di questo ospedale per questa categoria si aggirano intorno ai 5 milioni di euro, cioè più di 1 ordine di differenza rispetto a quella del 2017.
In questa figura sono apprezzabili anche le spese di Alessandria nell'anno 2015 che non raggiungono i 10 milioni di €, quelle di Torino 3 nel 2016 di 10 milioni di euro, ancora quelle di Città della Salute e della Scienza di Torino nel 2016-2017 con una spesa di rispettivamente di 17 e 11 milioni di euro, e infine l'Azienda Ospedaliera S.Croce e Carle di Cuneo con una spesa di quasi 8 milioni di € nel 2016.
Ora facciamo uno zoom sulle spese minori per evidenziarne trend e differenze nella figura 5.19:

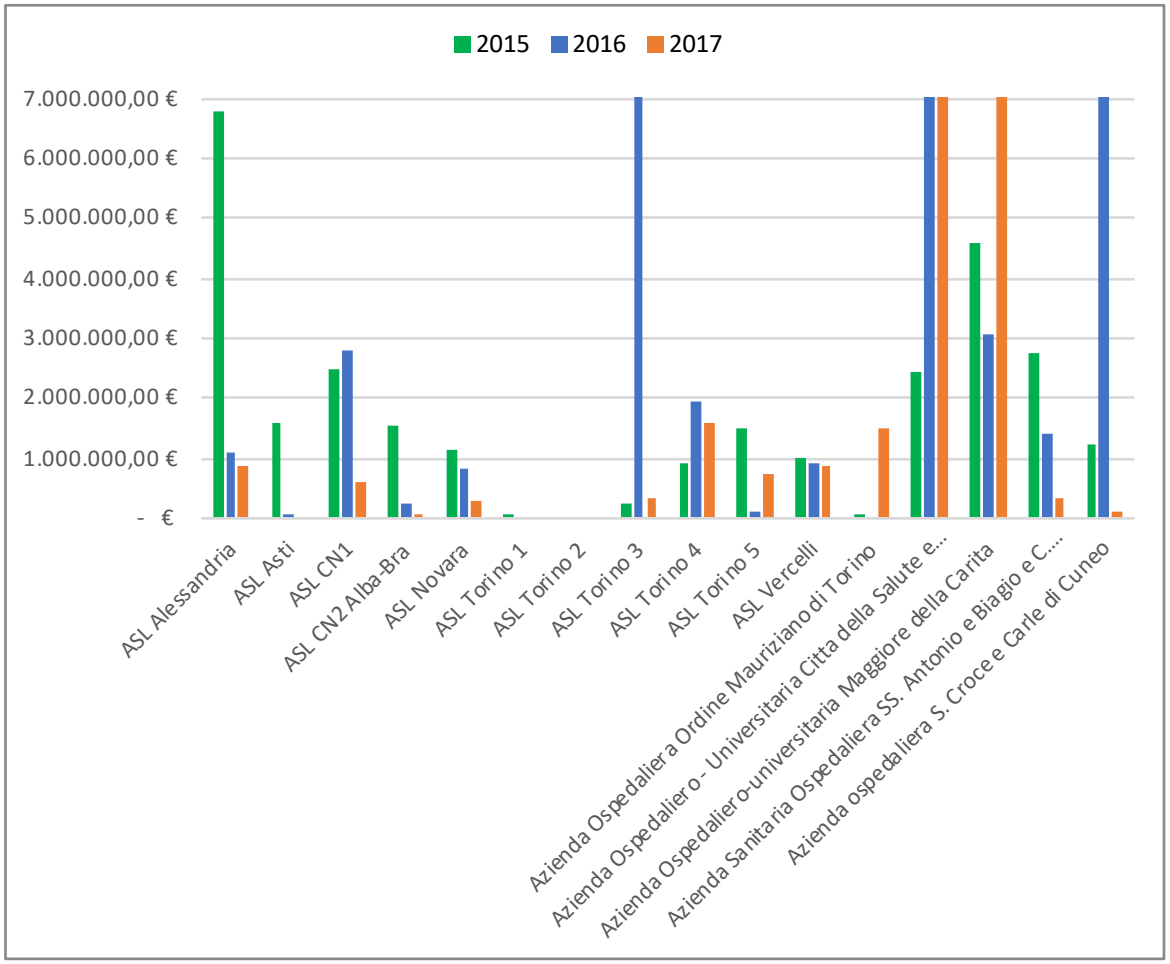

Figura 5.19. Focus sulle spese minori della categoria riparazione e manutenzione di attrezzature mediche di precisione.

Rispetto ai numeri del grafico precedente, questi sono poco significativi. Le uniche informazioni rilevanti sono che comunque sono registrate le spese per quasi tutti gli ospedali tranne che per Torino 2, e che Torino 1 e Torino 2 sono come da ipotesi le due ASL che spendono di meno. L'ASL di Vercelli invece registra spese appena sotto la media degli ospedali nel grafico zoommato.
Non si evidenzia un trend ricorrente nelle spese durante il triennio.

5.3.4 Spesa per i lavori di completamento degli edifici

La prossima categoria che analizziamo è quella che racchiude in se tutti i lavori di completamento degli edifici. Come per la precedente categoria della riparazione e manutenzione di attrezzature mediche di precisione, conduciamo l'analisi in due grafici, uno ad alto livello e una più a basso livello, per valorizzare i due set di valori di spesa che non sono proporzionali fra di loro. Cominciamo con la prima analisi ad alto livello con il grafico in figura 5.20:

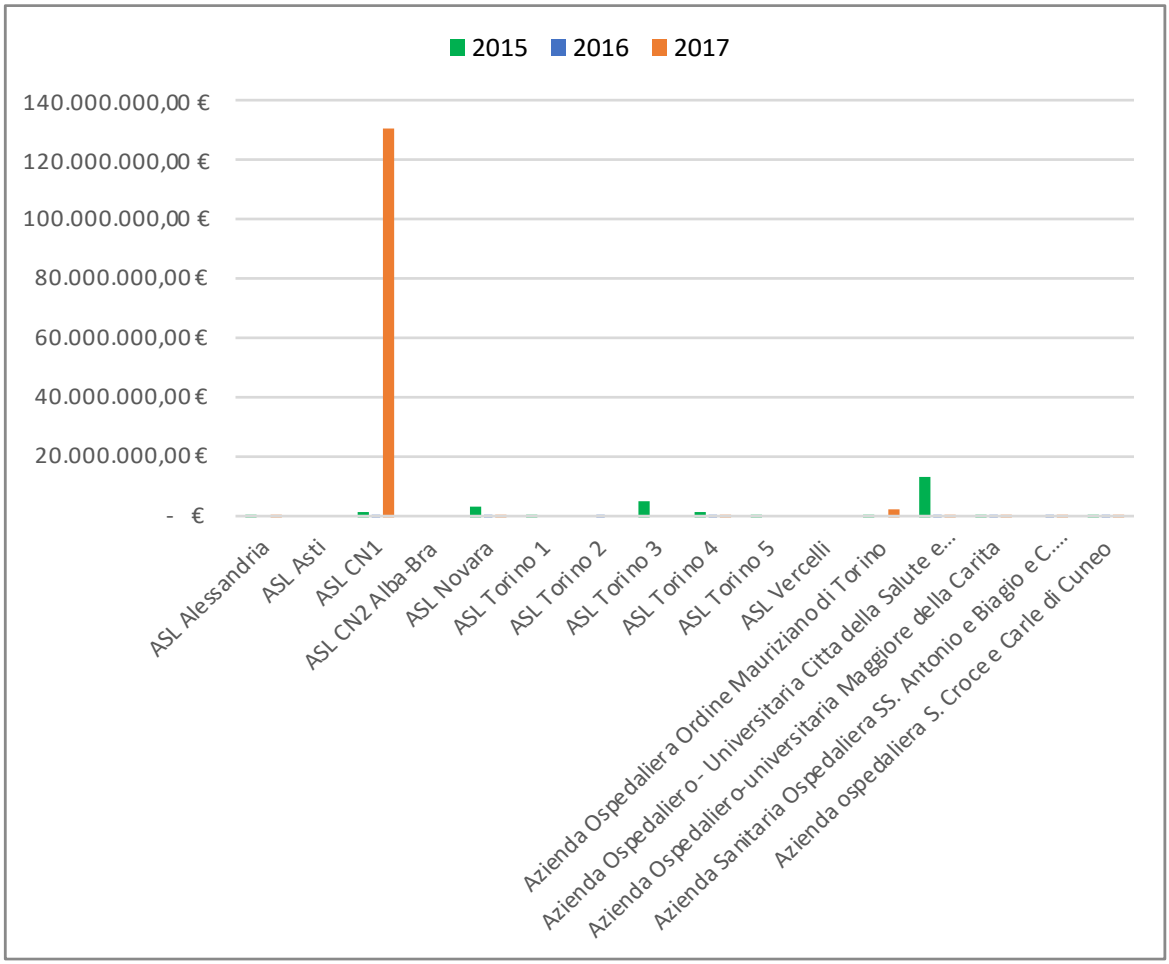

Figura 5.20. Categoria dei lavori di completamento degli edifici: confronto delle spese sostenute da ogni ospedale per ogni anno del triennio 2015-2017.

Una prima osservazione sembrerebbe far presupporre che i dati raccolti sulle spese per questa categoria siano molto sparsi. L'unica spesa valorizzata in questo primo grafico è quella dell'ASL CN1 che nel 2017 ha speso circa 130 milioni di euro,

che è molto sospetta: l'ASL in questione registra infatti nel 2015 e nel 2016 spese per meno di 1 ordine di grandezza di valore. La motivazione di questo outlier potrebbe risiedere nel fatto che nel 2017 ha dovuto investire questa somma di denaro a prima vista anomala per rimodernare la struttura.
Evidenziamo ora le spese minori nel grafico 5.21:

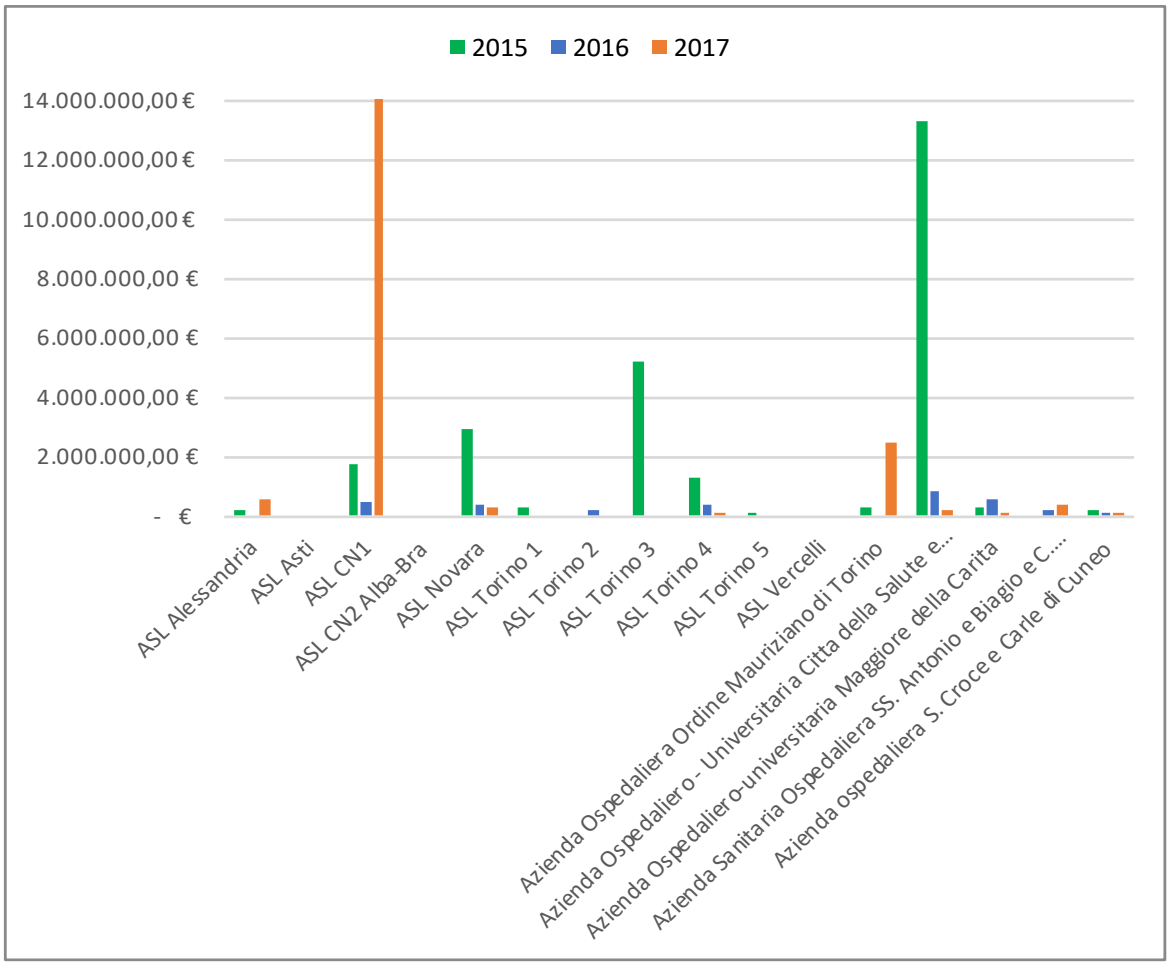

Figura 5.21. Focus sulle spese minori della categoria dei lavori di completamento degli edifici.

I dati non sono così sparsi come sembrava nel precedente istogramma, anche se rimane molto maggiore rispetto alle precedenti categorie. La motivazione risiede nel fatto che o le pubbliche amministrazioni hanno pubblicato solo una parte dei dati, o queste spese non erano presenti nel dataset ANAC usato per classificarle o ancora non sono state sostenute spese per questa categoria merceologica. Un esempio

significativo sono l'ASL di Asti, di Vercelli e CN2, per le quali non sembrano essere presenti spese per lavori di completamento degli edifici in nessuno dei tre anni. l'ASL di Torino 2 invece registra solo delle spese per il 2016, mentre quella di Alessandria registra spese nel 2015 e nel 2017, ma non per il 2016.
Rimaniamo stupiti dal fatto che Asti, l'ASL che nel triennio si classifica prima per spesa sostenuta, addirittura non abbia nessuna spesa registrata in questa categoria. Città della Salute invece almeno nel 2015 si distingue per spesa dalla maggioranza, Cuneo 1 a parte.
Dopo questa prima considerazione, ora ci concentriamo sui numeri. l'ASL di CN1 e l'Azienda Ospedaliero-Universitaria Città della salute e della scienza di Torino si classificano rispettivamente al primo e secondo posto in ordine di spesa media sostenuta nel triennio. Per CN1 abbiamo già discusso della spesa anomala nel 2017, e con questo focus notiamo che nel 2015-2016 non arriva a spendere neanche 2 milioni l'anno. Per quanto riguarda Città della Salute e della Scienza di Torino, registriamo nel 2015 quasi 14 milioni di spesa, per poi scendere a circa 1 milione nel 2016 e 500 mila euro nel 2017. Un andamento la cui spiegazione potrebbe essere simile a quella per Cuneo 1.
Per quanto riguarda le altre ASL, Novara nel 2015 registra circa 3 milioni di euro di spesa per poi scendere a circa 1 milione negli anni successivi; Torino 3 registra solo poco più di 5 milioni di spesa nel 2015, mentre per gli altri anni mancano i dati. L'Azienda Ospedaliera del Mauriziano nel 2017 spende poco più di 2 milioni di euro nel 2017, mentre mancano dati del 2016; nel 2015 la spesa sostenuta non arriva nemmeno a 1 milione di euro.
Gli altri ospedali che non sono stati menzionati, oltre ad avere dati sparsi come già evidenziato, hanno spese con valori poco significativi rispetto a quelli descritti nel precedente paragrafo, e non sono accumunati da un andamento degno di nota, anche a causa della sparsità dei dati.

5.3.5 Spesa per i servizi sanitari

Veniamo ora ai servizi sanitari, la quinta categoria che osserviamo con la lente d'ingrandimento. Di seguito riportato il grafico delle spese per anno e ospedale (figura 5.22) :

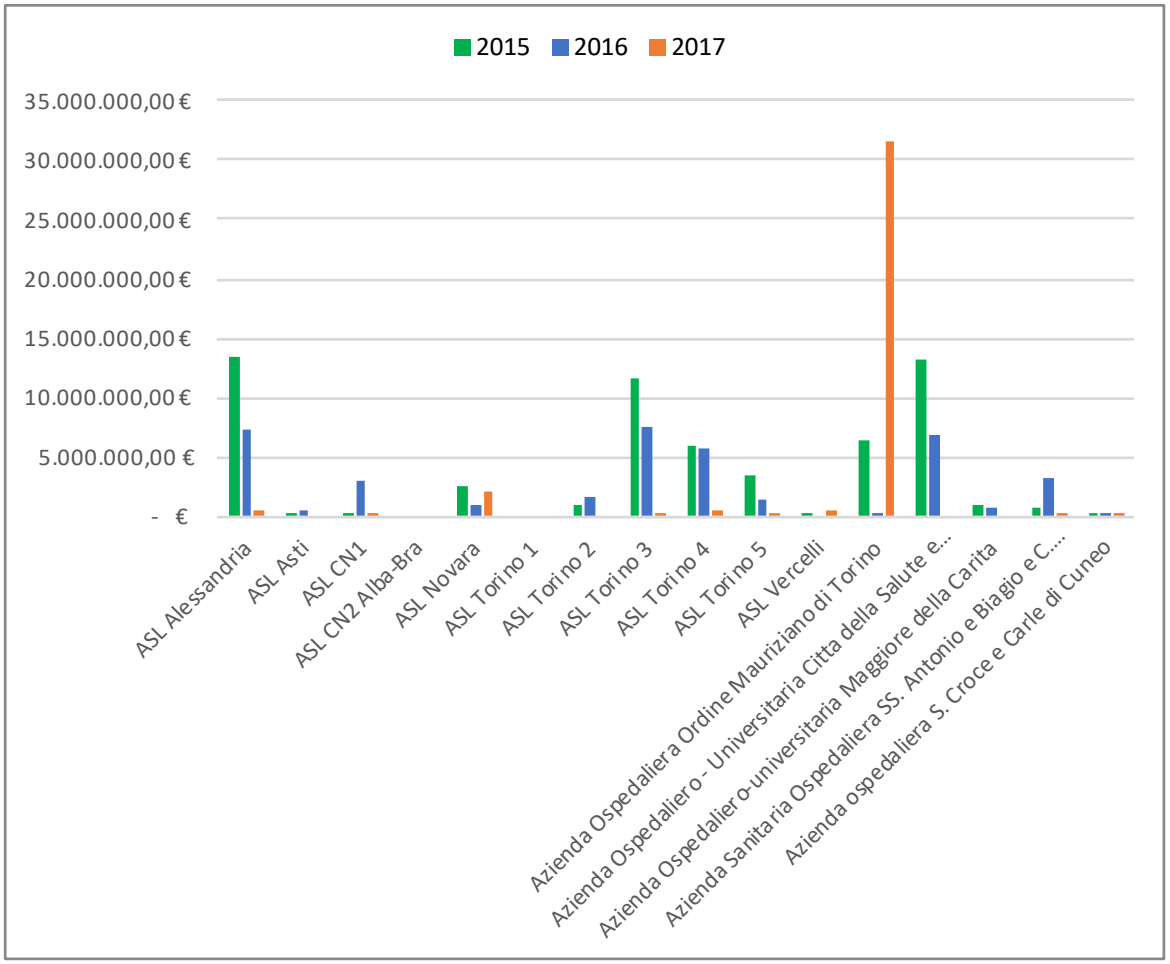

Figura 5.22. Categoria dei servizi sanitari: confronto delle spese sostenute da ogni ospedale per ogni anno del triennio 2015-2017.

Cominciamo osservando il livello di sparsità dei dati. CN2 e Torino 1 non registrano spese per servizi sanitari in tutti e 3 gli anni, fatto che risulta sospetto in quanto questa categoria racchiude tutte le spese che riguardano i servizi specialistici offerti come anche quelli di fisioterapia (come spiegato precedentemente).
L'ASL di Asti, Torino 2, Vercelli e l'Azienda Ospedaliera Maggiore della Carità invece sono accomunate dal fatto che non registrano spese per servizi sanitari nel solo anno 2017. La motivazione potrebbe risiedere le fatto che stiamo parlando di appalti che saranno rendicontati nelle prossime pubblicazioni dei dati della normativa 190, o dal fatto che non siano stati pubblicati dalle ASL oppure che non sono ancora stati inseriti nei dati anac.
Passiamo ad analizzare gli ospedali che hanno speso di più nella categoria in esame.

L'azienda ospedaliera Mauriziano di Torino nel 2017 registra la spesa più alta di tutti e tre gli anni (32 milioni di euro), ma nel 2016 è una di quelle che spende meno in assoluto (poche centianaia di migliaia di euro, 3-4 ordini di grandezza in meno rispetto al 2016), mentre nel 2015 rispecchia la media degli ospedali. Dopodichè troviamo Alessandria, Torino 3, Torino 4 e Città della Salute e della Scienza di Torino che nel 2015 e nel 2016 sono quelle che spendono di più (nel 2015 spendono intorno ai 13-14 milioni di euro tranne Torino 4 che ne spende circa 6, mentre nel 2016 spendono tutte intorno ai 6 milioni di euro).
Tutti gli altri ospedali registrano spese molto inferiori ai 5 milioni di euro e quindi non particolarmente significative.
L'ASL di Asti ci sorprende di nuovo in quanto è quella che insieme all'Azienda Ospedaliera Croce e Carle di Cuneo registrano la minore spesa nella categoria in esame. Città della Salute e della Scienza si mantiene in linea con la media degli altri ospedali nel 2015 e 2016, nel 2017 non abbiamo dati come già detto in precedenza. Torino 1 non registra spese, Torino 2 spende molto più rispetto ad atri ospedali mentre Vercelli è coerente: infatti è uno degli ospedali a spendere di meno nella categoria.

5.3.6 Spesa per servizi vari

Veniamo ora alle spese per servizi vari. Osserviamo il grafico di seguito 5.23 :

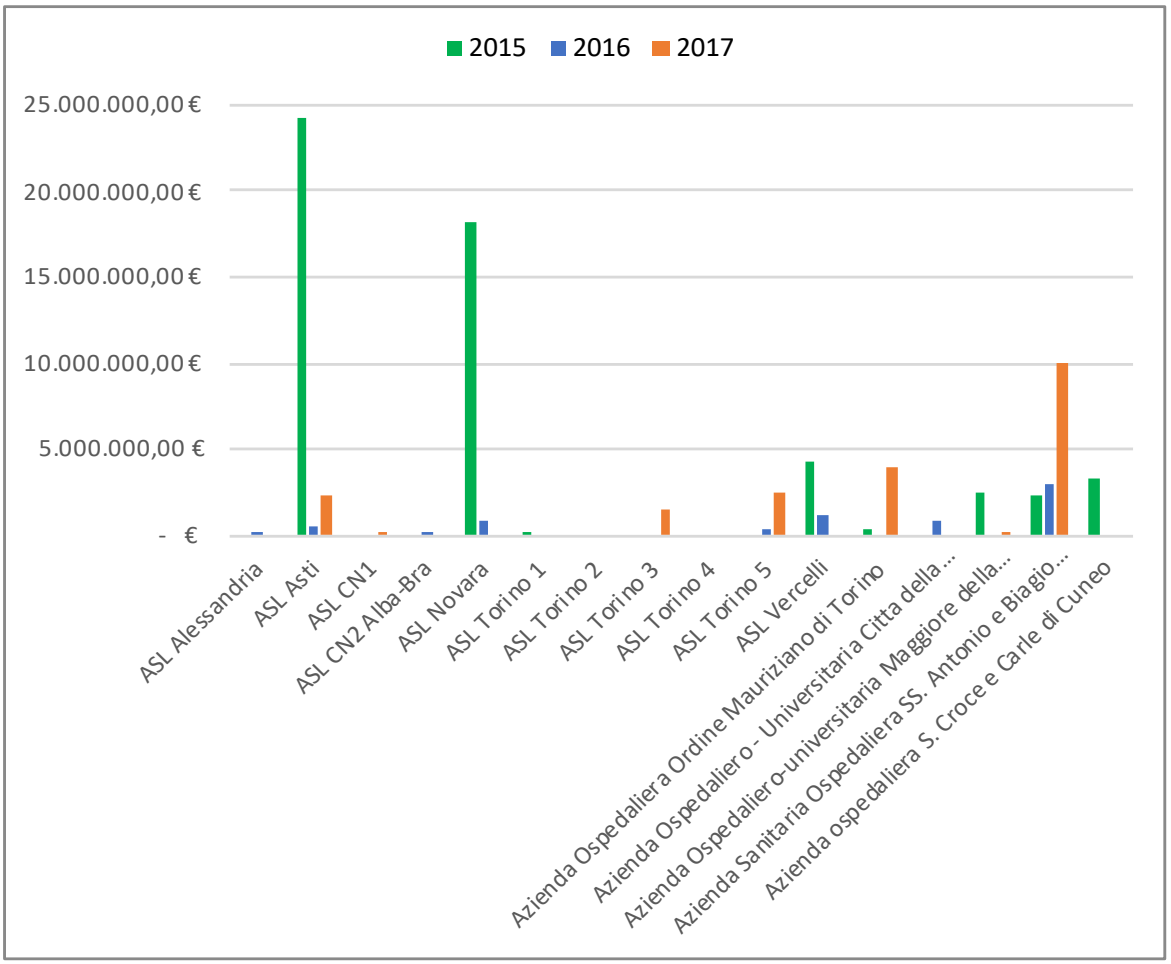

Figura 5.23. Categoria dei servizi vari: confronto delle spese sostenute da ogni ospedale per ogni anno del triennio 2015-2017.

In questa categoria troviamo dati con un livello di sparsità medio-alto. Mancano i dati del triennio per l'ASL di Torino 2 e Torino 4, mentre Alessandria, CN1, CN2, Torino 1, Torino 3, Città Della Salute e della Scienza di Torino e l'Azienda Sanitaria S. Croce e Carle di Cuneo hanno dati per un solo anno. Le uniche complete di dati per tutti e tre gli anni sono Asti e l'Azienda Ospedaliera SS.Antonio e Biagio e Arrigo, mentre le rimanenti hanno dati per due dei tre anni.
Asti e Novara sono le due ASL che registrano la spesa più alta di tutte nel triennio, e queste due spese si riferiscono al 2015. Sono di 24 milioni abbondanti per Asti e di 18 milioni per Novara. Negli anni 2016-2017 invece sostengono spese al di sotto della media e per niente proporzionale a quella del 2015 (tutte sotto i 2 milioni di euro, quindi ben un ordine di grandezza in meno) . La terza spesa più alta è

registrata dall'Azienda Ospedaliera SS.Antonio e Biagio e Arrigo nel 2017 che vale 10 milioni di euro, mentre la sua spesa nel 2015-2016 è intorno alla media (circa 3 milioni di euro per anno).
Asti in questa categoria riconferma il suo primo posto per spesa sostenuta, mentre sorprende ancora Città della Salute e della Scienza di Torino per la quale si hanno solo i dati del 2016, che sono comunque molto bassi (quasi 1 milioni di euro).
L'ASL di Vercelli contrariamente alle aspettative spende di più rispetto ad altri ospedali, mentre Torino 1 e 2 oltre a registrare spese sempre bassissime, sono i due ospedali per i quali i dati sono quasi sempre mancanti.
Non è evidenziabile nessun trend degno di nota.

5.3.7 Spesa per l' erogazione di energia elettrica e per i servizi connessi

Passiamo ora ad analizzare un'importante categoria in termini di qualità dei dati: la categoria per le spese di energia elettrica. Infatti il consumo di energia elettrica dovrebbe sempre essere costante nel tempo a meno di ampliamenti o di altri lavori connessi alla costruzione / smantellamento di edifici. Riportiamo il grafico dei costi 5.24:

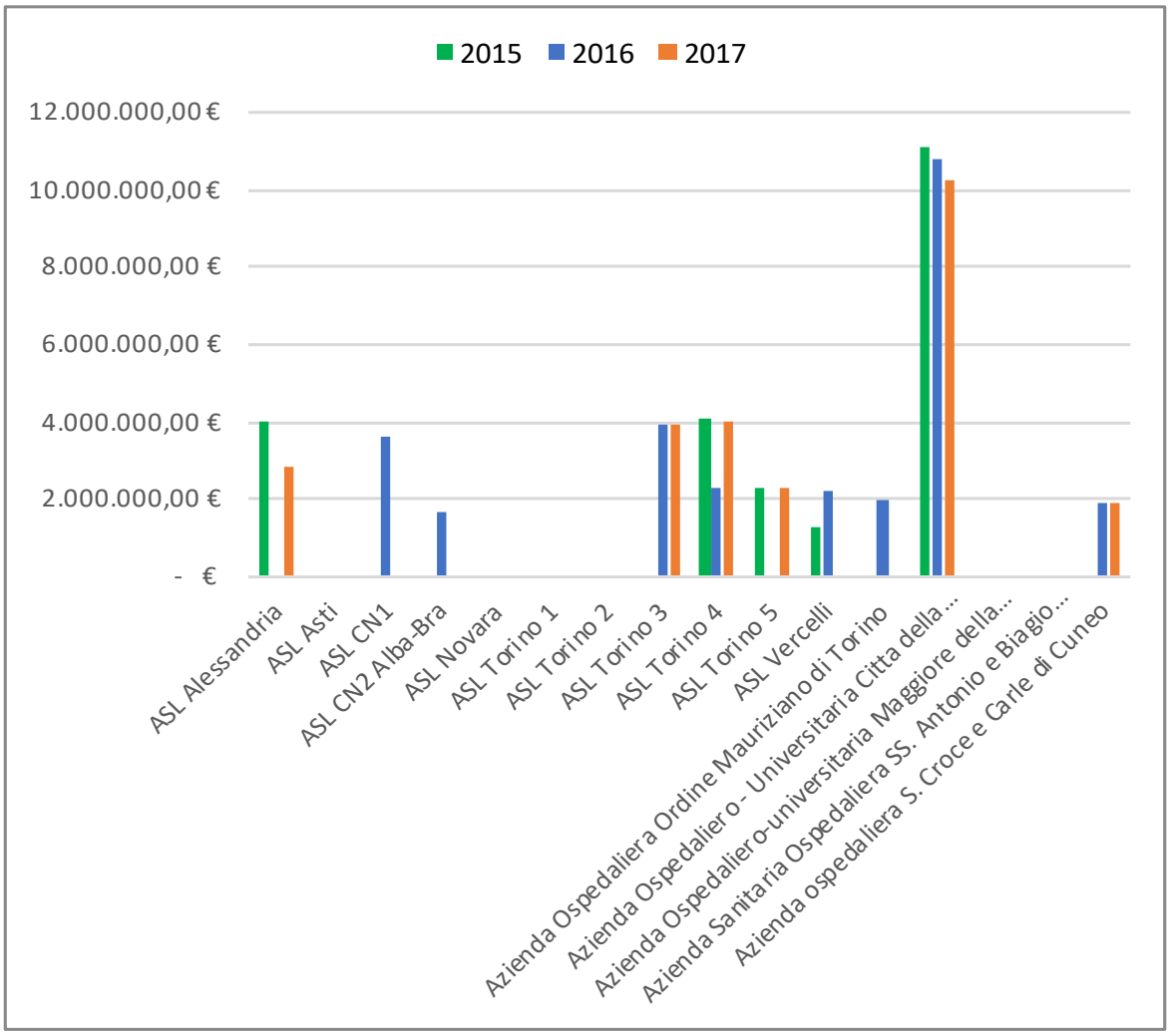

Figura 5.24. Categoria dell'erogazione di energia elettrica e dei servizi connessi: confronto delle spese sostenute da ogni ospedale per ogni anno del triennio 2015-2017.

Molto sospetta è l'assenza di spese legate all'elettricità in tanti ospedali, che indica che effettivamente ci sono casi in cui i dati non sono correttamente pubblicati dalle Pubbliche Amministrazioni oppure che non sono presenti nel dataset anac. Gli ospedali che non registrano spese per l'elettricità consumata sono ASL di Asti, che è l'azienda sanitaria locale che spende più di ogni altro ospedale nel triennio, l'ASL di Novara, Torino 1 e 2, che sono quelle che spendono di meno, l'Azienda Ospedaliero Maggiore della Carità e quella SS: Antonio e Biagio e Arrigo.
Passiamo ad analizzare gli ospedali che invece hanno registrato queste spese: solo per l'ASL di Torino 4 e Città della Salute e della Scienza di Torino abbiamo i tre

anni di spesa, l'ASL di CN1, CN2 e l'Azienda Ospedaliera Mauriziano registrano spese soltanto per uno dei tre anni, il 2016. Le rimanenti hanno registrato spese per 2 anni su 3, senza maggiore frequenza di un anno rispetto ad un altro.
Le spese di ogni ospedale sono tutte molto simili da un anno al successivo. L'ASL di Alessandria che nel 2017 ha diminuito la spesa di circa 1 milione rispetto al 2015, da 4 a 3 milioni circa. Non abbiamo i suoi dati del 2016, quindi non possiamo osservare se il trend della spesa è decrescente oppure se il 2017 è l'anno in cui è stato registrato un calo netto nella spesa di elettricità. Torino 4 invece nel 2015 e nel 2017 spende 4 milioni all'anno per spese legate al consumo di elettricità, mentre nel 2016 la vede quasi dimezzata. Molto probabilmente questo calo è un segnale della mancanza di dati. L'ASL di Vercelli invece spende 1 milione abbondante nel 2015 e 2 milioni abbondanti nel 2016, vedendo la spesa quindi raddoppiata.
Torino 3, Torino 5, Città della Salute e della scienza di Torino e S.Croce e Carle di Cuneo spendono quasi esattamente la stessa cifra per tutti gli anni registrati, un ottimo segnale che indica che 5 dei 7 ospedali con più di una spesa registrata hanno numeri molto simili in anni successivi, come ci si aspetterebbe.

5.3.8 Spesa per la programmazione di software e per servizi di consulenza

Osserviamo ora una categoria che di discosta un po' da quelle trattate finora: la categoria che ingloba tutte le spese per la programmazione di software e i servizi di consulenza connessi. Vediamo come sono distribuite le spese negli ospedali in figura 5.25:

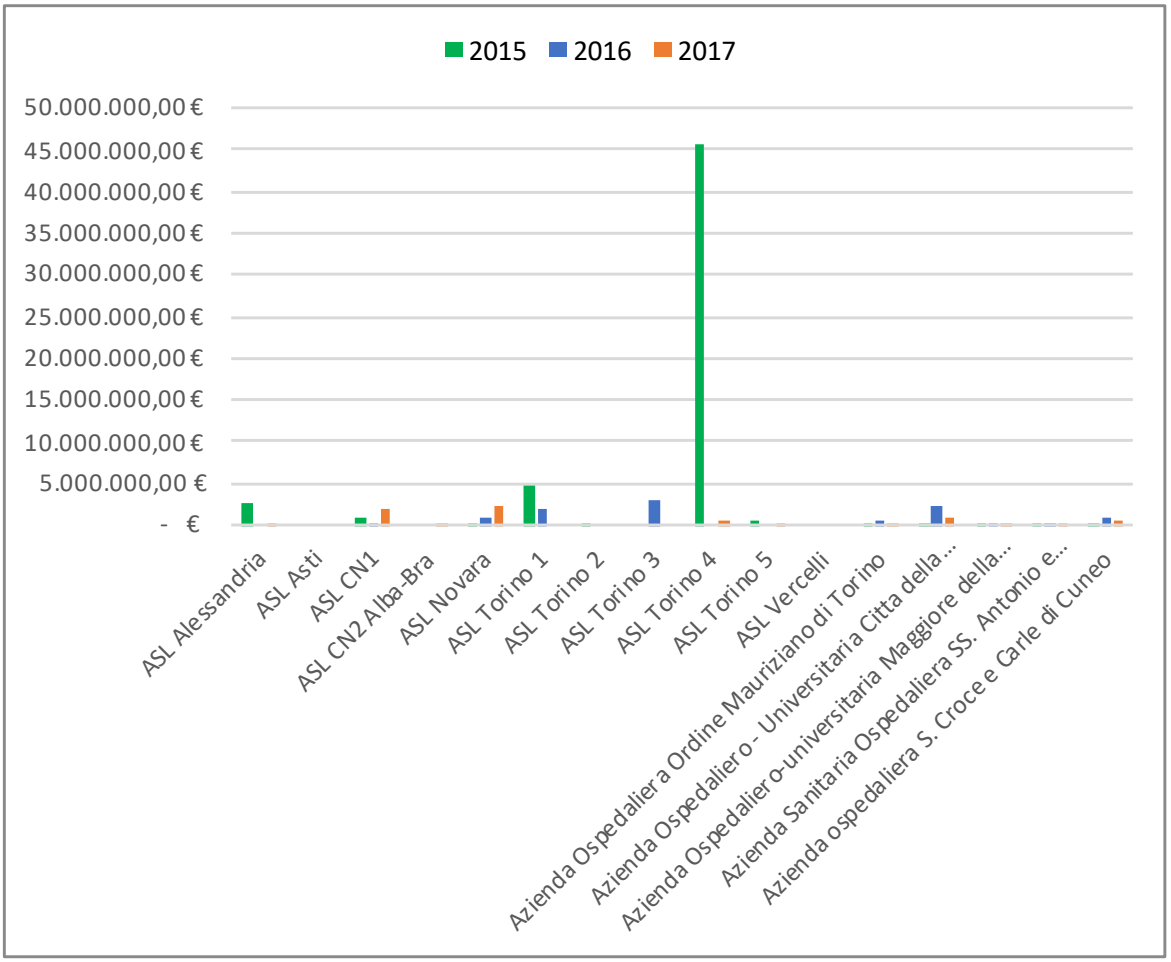

Figura 5.25. Categoria della programmazione di software e dei servizi di consulenza: confronto delle spese sostenute da ogni ospedale per ogni anno del triennio 2015-2017.

Questa sembrerebbe a prima osservazione la categoria con più dati sparsi finora analizzati: ben 6 ospedali su 16 sembrano non registrare spese per nessuno dei 3 anni! E non sembra esserci neanche una struttura in cui sono presenti dati per tutti i tre anni. Nel prossimo grafico facciamo un focus sulle spese minori in modo da chiarire se i dati a disposizione sono veramente così pochi e tanto sparsi. Considerando solo questa figura, il 2015 è l'unico anno che registra due spese degne di nota: Torino 4 spende la cifra di 45 milioni di euro mentre Torino 1 circa 5 milioni. Sorprende trovare Torino 1 tra le struttura che spendono maggiormente in quanto nelle categorie precedenti non aveva spesso dati da analizzare, come sorprende il fatto che l'ASL di Asti e l'Azienda Ospedaliera Città della Salute e della Scienza di Torino non risaltino in questo istogramma, anzi quella di Asti non registra alcuna

spesa, mentre Città della Salute spende circa 100 mila euro nel 2015, 2,2 milioni nel 2016 e quasi 1 milioni nel 2017, che sono numeri nella media delle spese di questo grafico. Per contro stiamo comunque parlando di una categoria merceologica molto differente dalle solite, perciò potrebbe anche essere questa la giustificazione di spese così variabili e sparse.
Evidenziamo le spese minori che non sono proporzionali numericamente parlando con quelle appena trattate, si veda la figura 5.26 :

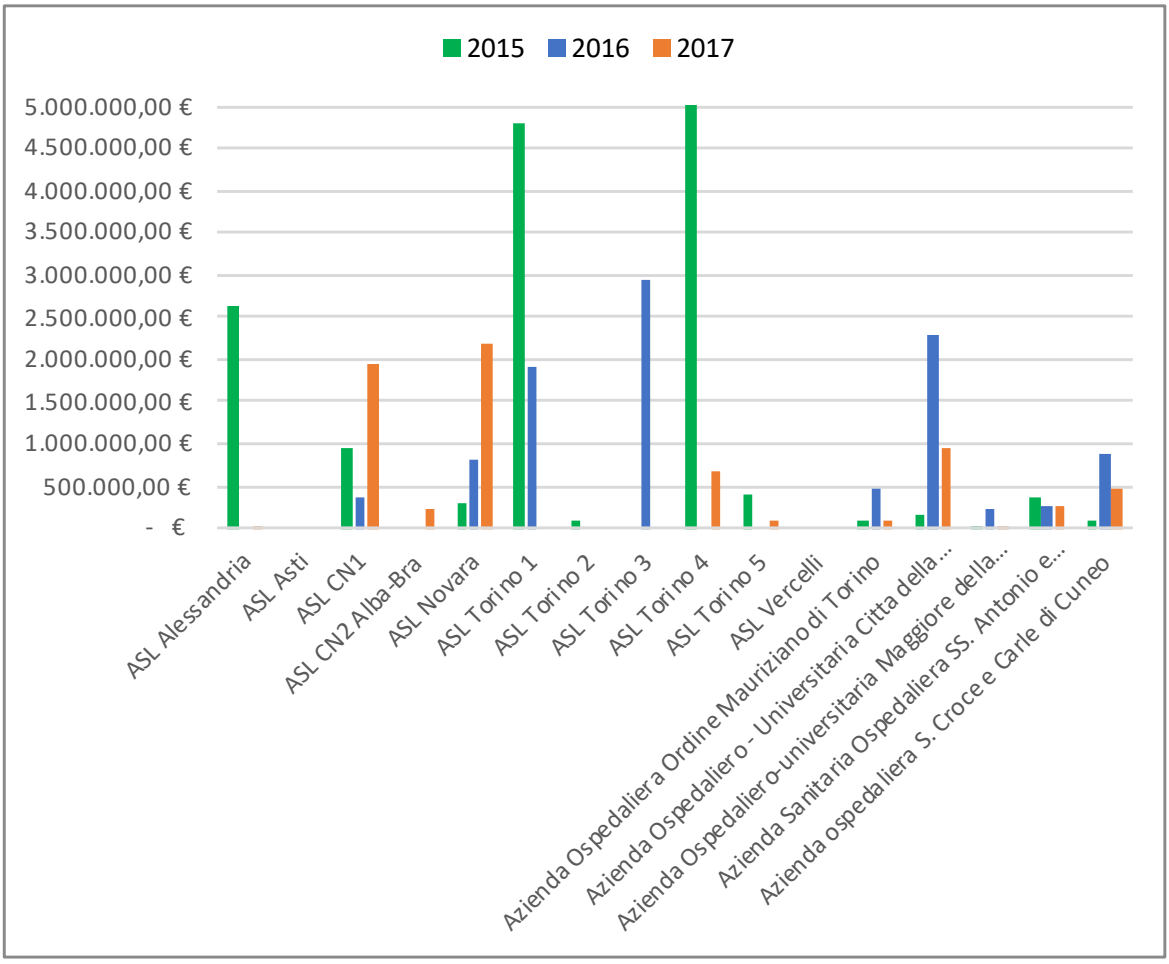

Figura 5.26. Focus sulle spese minori della categoria della programmazione di software e dei servizi di consulenza.

Questo zoom ribalta parte delle osservazioni ad alto livello fatte precedentemente sul primo grafico: i dati non sono così sparsi come si era prima detto, infatti solamente le ASL di Asti e Vercelli non riportano alcun lotto per questa categoria, mentre le altre strutture presentano almeno una spesa registrata sulle tre.

Oltre agli ospedali citati nell'analisi ad alto livello, troviamo le spese di Alessandria, che è la terza maggiore se guardiamo solo le spese del grafico nel 2015 e la quarta se consideriamo le spese di tutti e tre gli anni. Non abbiamo altri dati per questa ASL, quindi non possiamo valutare trend nel tempo. Salta all'occhio la spesa di Torino 3, la più alta del 2016, ma anch'essa non ha termini di paragone per anni differenti. Un'altra spesa che risalta è quella di Novara, che nel 2017 ha speso circa 2,2 milioni di euro e si classifica come la prima in ordine di spesa tra quelle del 2017. L'ASL di Novara registra anche le spese del 2015 e del 2016: il trend è in forte crescita, forse sinonimo di un processo di digitalizzazione dei sistemi informatici. Le sue spese nel 2015-2016 sono proporzionali a quelle delle altre strutture, tranne Alessandria, Torino 1 e 4 di cui abbiamo già parlato prima.
Per completare l'analisi degli ospedali iniziato nelle osservazioni ad alto livello, aggiungiamo che Vercelli e Torino 2 registrano in totale 1 sola spesa su 6. Se ipotizziamo che le spese ci siano ma manchino nel nostro dataset, non sarebbero comunque significative in quanto per ipotesi dovrebbero essere le minori di tutte, e quindi nell'ordine delle decine di migliaia di euro.

5.3.9 Spesa per i servizi di pulizia e disinfestazione

La prossima categoria che analizziamo è quella in cui troviamo le spese per la pulizia e la disinfestazione. Conduciamo le analisi osservando il grafico 5.27:

I dati risultano molto sparsi: l'ASL di Alessandria non sembra sostenere spese in nessuno dei tre anni, come anche quella di Torino 2, di Torino 3, l'Azienda Ospedaliera Mauriziano e Maggiore della Carità.
La spesa maggiore è sostenuta da l'Azienda Ospedaliera Universita della Salute e della Scienza di Torino, di importo poco superiore ai 25 milioni di euro, mentre stupisce l'ASL di Asti che registra spese apparentemente nella media. In questo grafico risaltano anche la spesa nel 2016 dell'ASL di CN2 e quella dell'Azienda Ospedaliera Croce e Carle di Cuneo, che valgono entrambe 6,5 milioni di euro.
Facciamo uno zoom per evidenziare le spese minori, figura (5.28):

In quest'altro grafico riusciamo a valorizzare le spese minori dell'ASL di Asti, CN1 e dell'Azienda Ospedaliera SS.Antonio e Biagio e Arrigo, di cui prima non si poteva apprezzare l'andamento nel triennio. Inoltre osserviamo più da vicino le spese dell'ASL di Novara, Torino 1, Torino 4 e Vercelli, che sono molto minori rispetto alle altre analizzate (siamo nell'ordine delle migliaia di euro), quasi trascurabili.
Si osserva inoltre che Asti è l'unica insieme all'Azienda Ospedaliera SS. Antonio e Biagio e Arrigo ad avere i dati completi dei tre anni; seguono lo stesso trend le spese di entrambi gli ospedali. Torino 1 registra una spesa (esigua ma la registra) mentre Torino 2 si conferma una delle ASL i cui dati sono i più sparsi in assoluto. L'ASL di Vercelli rientra nelle ultime tre in ordine di spesa anche per questa

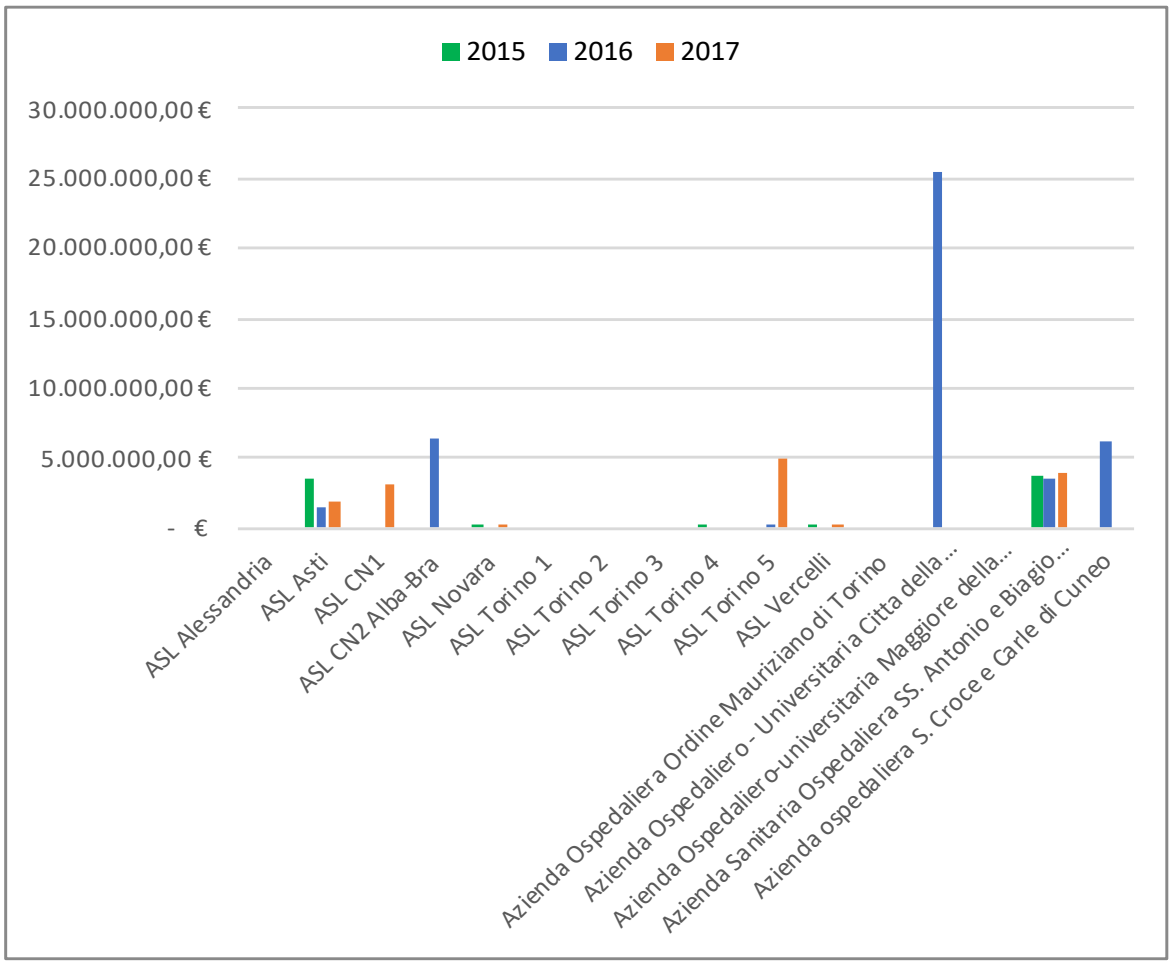

Figura 5.27. Categoria dei servizi di pulizia e disinfestazione: confronto delle spese sostenute da ogni ospedale per ogni anno del triennio 2015-2017.

categoria, perfettamente in linea con il risultato della maggior parte delle categorie analizzate finora.

5.3.10 Servizi ingegneria

L'ultima categoria che analizziamo, che vale circa l'1% della spesa per ogni anno, è quella che comprende tutti i servizi di ingegneria elencati precedentemente. Vediamo come si distribuiscono le spese nei vari ospedali in figura 5.29 :

Si possono notare dati molto sparsi e con un'escursione altissima, la maggiore registrata nelle categorie finora analizzate. L'unica spesa apprezzabile in questo grafico è quella dell'Azienda Sanitaria Città della Salute e della Scienza di Torino, di

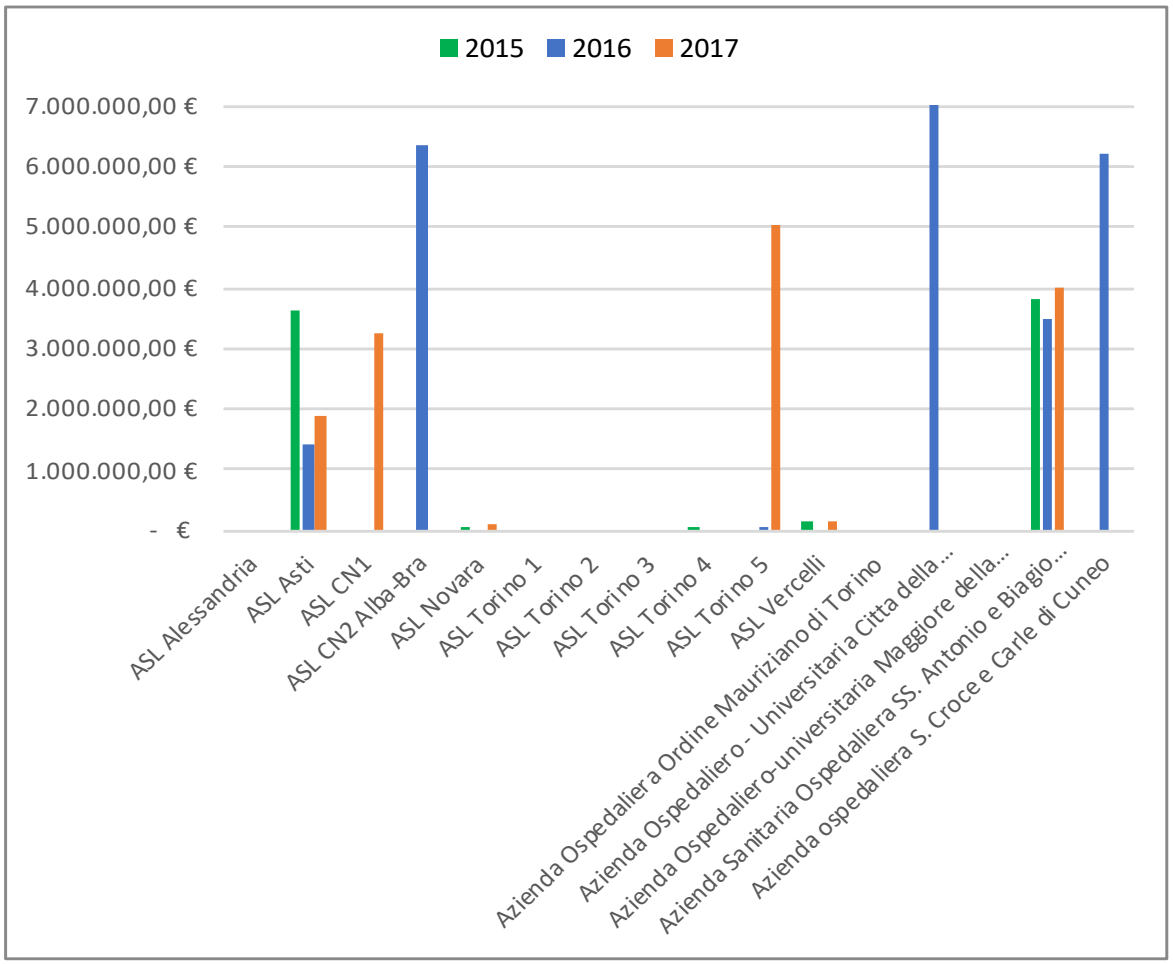

Figura 5.28. Focus sulle spese minori della categoria dei servizi di pulizia e di disinfestazione.

42,5 milioni di euro. L'ASL di Asti, a differenza dell'Azienda Sanitaria precedente, non ha nemmeno una spesa registrata, contrariamente alle aspettative.
Facciamo uno zoom per evidenziare le spese minori non osservabili in questo primo grafico, si veda figura 5.30:

Ora è possibile guardare più da vicino le altre spese ,seppur molto esigue e quasi trascurabili. Tutte quante a eccezione di Torino 4 nel 2016 registrano una spesa inferiore ai 200 mila euro. Impossibile valorizzare un trend in quanto i dati sono eccessivamente sparsi e tranne che per Torino 4 e l'Azienda Sanitaria Maggiore della Carità, le altre non hanno dati per tutti e 3 gli anni. L'ASL di Torino 1 e 2 non hanno dati registrati, in conformità con la sparsità dei loro dati che li ha caratterizzati nell'analisi finora condotta.

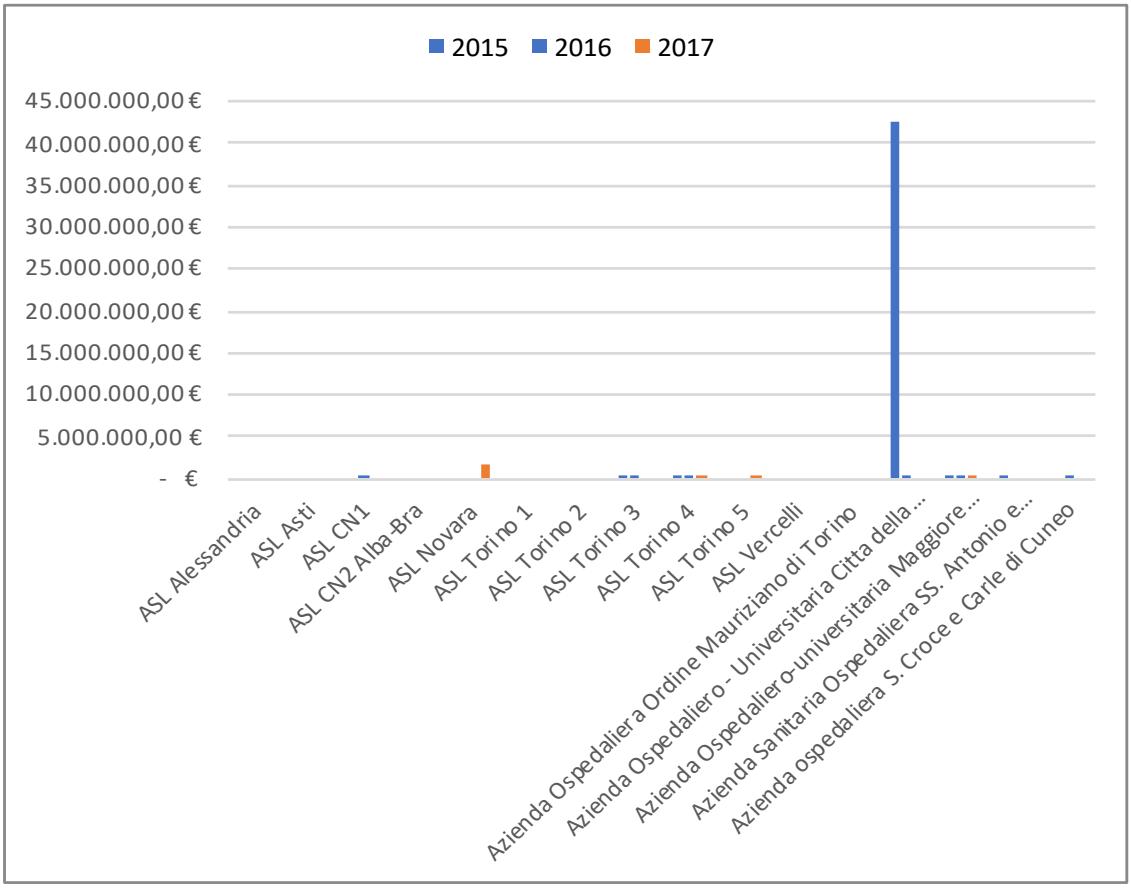

Figura 5.29. Categoria dei servizi di ingegneria: confronto delle spese sostenute da ogni ospedale per ogni anno del triennio 2015-2017.

5.4 Comparazione della spesa pro-capite nel triennio

In questa sezione consideriamo soltanto la spesa sostenuta dalle ASL presenti nel dataset, in quanto possiamo rapportare la somma spesa in relazione alla popolazione servita da ogni Azienda Sanitaria Locale e quindi valutare quale ASL spende di più per curare un singolo cittadino. Inoltre le 5 ASL di Torino sono considerate tutte insieme in un'unica ASL che chiamiamo ASL Torino, in modo da avere dati meno sparsi (le analisi per categoria e ospedale della sezione precedente hanno mostrato quanto possono essere sparsi i dati, basti pensare che nel 2017 non abbiamo dati relativi alle ASL di Torino 1 e 2) e anche per il fatto che rispetto alle altre ASL che sono geograficamente distanti, quelle di Torino non lo sono affatto e un abitante

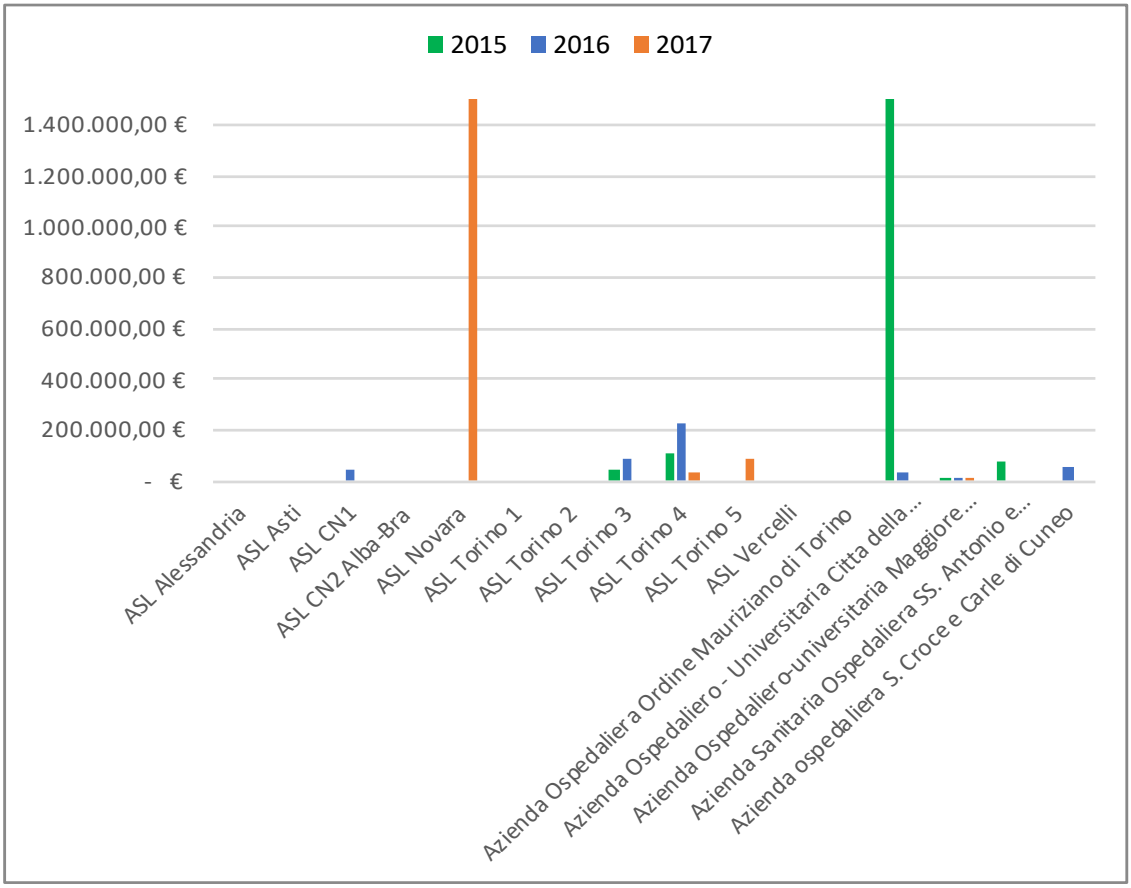

Figura 5.30. Focus sulle spese minori della categoria dei servizi di ingegneria.

residente sotto una determinata Azienda Sanitaria può facilmente accadere che in base a dove si trovi si vada a far curare in una struttura diversa dalla sua ASL di riferimento.
Per ottenere questi risultati sono stati utilizzati tutti i dati ottenuti con le query precedenti: si sono utilizzate le spese raggruppate per le sole ASL (e non comprendendo più ogni Azienda Ospedaliera) e per anno, che sono state divise per la popolazione servita da ogni ASL.
Riportiamo nella figura 5.31 che mostra il numero di persone servite da ognuna di esse e la spesa sostenuta divisa per ogni anno.

Nell' istogramma 5.32 mettiamo a confronto invece l'andamento della spesa pro capite per ogni ASL durante il triennio, e ne valutiamo l'andamento rispetto al trend osservato precedentemente nel triennio. Considerazioni sui risultati: Asti e CN1 sono quelle che in media spendono di più per curare ogni cittadino (più di

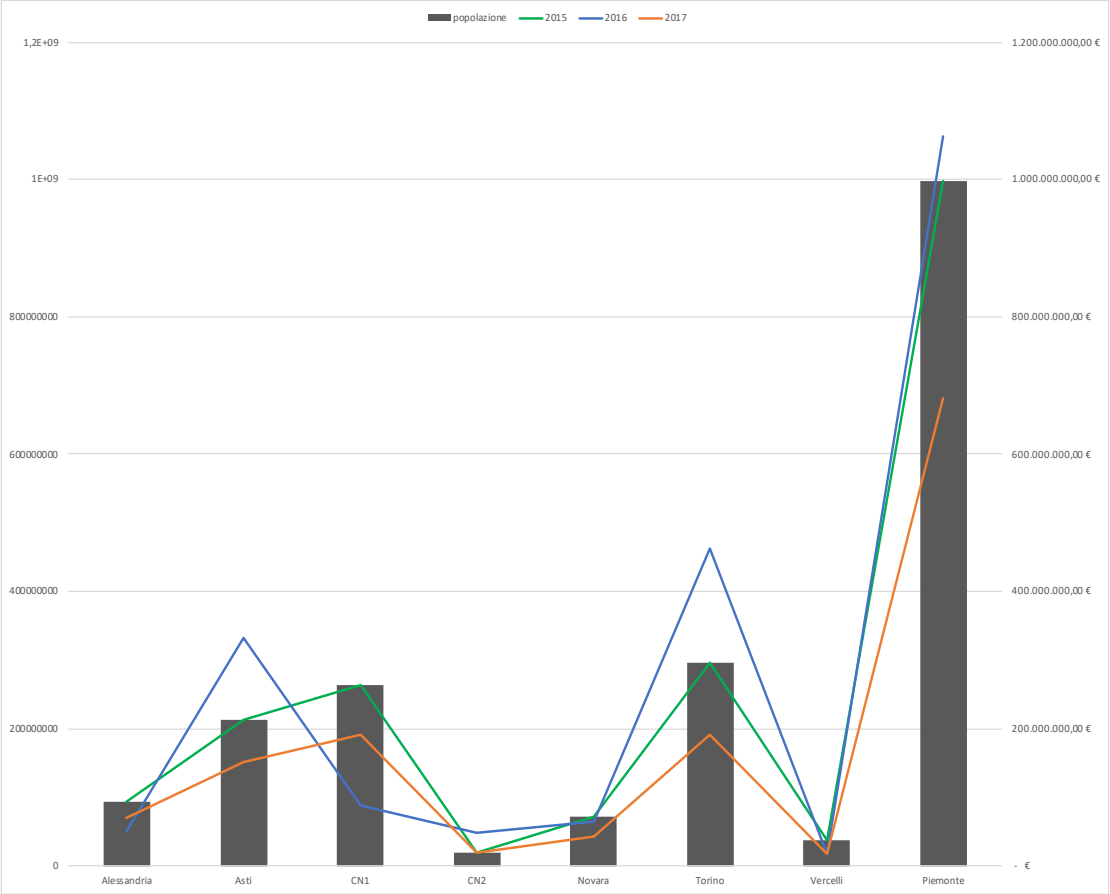

Figura 5.31. Le barre grigie rappresentano la popolazione servita da ogni ASL, mentre le tre linee di colore verde, blu e rossa rappresentano la spesa totale per ogni ASL rispettivamente per gli anni 2015, 2016 e 2017.

1000 euro nel 2015, 1650 € nel 2016 e 750€ nel 2017 per Asti; 600 € nel 2015, 200 € nel 2016 e 450 € nel 2017 per Cuneo 1), mentre Alessandria è quella che spende meno (spesa pro capite annuale sempre minore o uguale a 200 €). Gli abitanti serviti dalle ASL di Asti, CN2 e Torino hanno subito un forte incremento di spesa dal 2015 al 2016 mentre sono anche quelle che hanno subito una forte decrescita di spesa nel 2017.
Prendiamo in esame la spesa pro capite media di tutte le ASL analizzate (l'ultimo set di colonne con valore 'Piemonte' sull'asse delle ascisse), che registra una spesa pro capite di 251 € nel 2015, 267 € nel 2016 e 171 € nel 2017. Per quanto riguarda il 2015, solo l'ASL di Torino e CN2 spendono di meno rispetto alla media, mentre Asti e CN1 spendono dalle 2 alle 4 volte in più. Simile invece la spesa per Alessandria, Novara e Vercelli. Nel 2016 solo Asti spende molto di più rispetto alla media

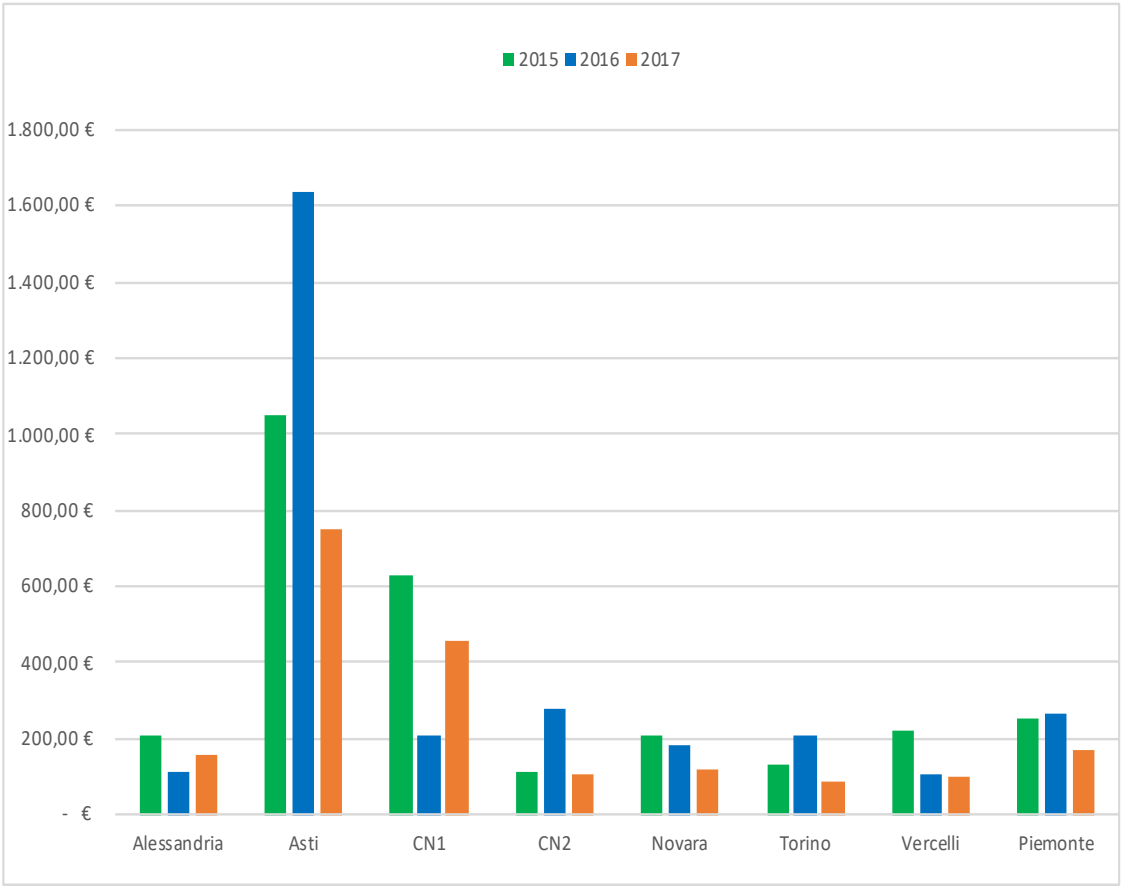

Figura 5.32. Spesa pro capite annuale per abitante per ogni ASL, le barre verdi si riferiscono alla spesa nel 2015, le barre blu si riferiscono alla spesa nel 2016 mentre quelle rosse alle spese del 2017.

piemontese, mentre Alessandria e Vercelli che prima avevano valori intorno alla media hanno visto la loro spesa diminuire di quasi la metà. Infine nel 2017 Asti e CN1 come nel 2015 hanno il primato della spesa mentre tutte le restanti ASL spendono tutte poco meno della media piemontese.

Capitolo 6

Conclusioni e lavori futuri

Dopo aver condotto tutte le analisi, possiamo concludere che le categorie merceologiche in cui si spende di più sono quelle dei prodotti farmaceutici e delle apparecchiature mediche. Le ASL di Asti e di Cuneo 1 insieme all'Azienda Ospedaliero-Universitaria Città della Salute e della Scienza di Torino sono le strutture ospedaliere che spendono di più nei tre anni analizzati, mentre l'ASL di Torino 1, Torino 2 e quella di Vercelli sono quelle che spendono di meno in tutte le categorie merceologiche. L'ASL di Asti inoltre è quella che spende di più per curare ogni abitante tra tutte le ASL, per contro quella che spende di meno è l'ASL di Vercelli. Interessante notare come Asti e Vercelli siano al primo e all'ultimo posti di spesa anche rapportando il totale alla popolazione servita.
Durante questo lavoro di tesi sono sopraggiunti molti problemi come la qualità dei dati analizzati, a partire da quelli ANAC sprovvisti di headers e con un sacco di campi con valore nullo o non significativo, come nel caso degli importi di appalti aggiudicati uguali a zero. I dati della normativa 190/2012 hanno un sacco di dati mancanti, nulli, imprecisi o ancora senza significato, come l'assenza di partecipanti alla gara d'appalto e invece la presenza di società vincitrici e la presenza di codici identificativi gara duplicati per più lotti. Un altro problema ancora è quello che riguarda la mancata o la parziale pubblicazione dei dati della normativa da parte di certe ASL, come l'ASL di Biella e l'ASL VCO, che rendono l'analisi delle spese della sanità piemontese incompleta.
Il risultato di questi problemi di qualità dei dati è che si analizza un numero ridotto di lotti rispetto alla mole di informazioni a disposizione. Di conseguenza i risultati ottenuti sono più interessanti e fedeli alla realtà a livello qualitativo che quantitativo, come testimonia anche la percentuale di spesa coperta dal dataset rispetto al rendiconto negli anni 2015 e 2016. Si preferisce esprimere gli esiti delle analisi in termini di percentuale di spesa coperta da una categoria merceologica, di confronto dei costi sostenuti da un ospedale rispetto alla media piemontese o rispetto agli altri ospedali rispetto a focalizzarsi su quanto è stato speso.

Tutti questi problemi di qualità dei dati che sono stati presentati, potrebbero essere presenti in minor volume se soltanto fosse prevista una sorta di controllo e/o validazione dei dati pubblicata e una manutenzione da parte di chi raccoglie questa mole di dati, come ad esempio l'ANAC.
L'ulteriore problema è quello che riguarda la precisione della classificazione effettuata. Il massimo livello di dettaglio raggiunto è stato quello del livello 4 del codice CPV delle categorie merceologiche. Il livello 4 è adeguato ad un'analisi ad alto livello come quella effettuata, ma se si volesse conoscere l'apparecchiatura medica più acquistata dalle ASL piemontesi o il prodotto farmaceutico più costoso, questo livello di dettaglio non basta.
Questa maggiore precisione sarebbe ottenibile costruendo da zero un classificatore degli oggetti dei contratti pubblici oppure migliorando la precisione del classificatore CPV di Synapta. Un altro vantaggio sarebbe quello di poter classificare molti più lotti rispetto a quelli classificati con questa analisi, includendo ad esempio l'anno 2014 che comprende un numero di lotti e una spesa paragonabile a quella degli anni analizzati in questo progetto.
Un altro lavoro molto utile e interessante una volta riusciti a classificare più anni e dati di quanto fatto in questa tesi sarebbe quello di monitorare anno per anno le categorie merceologiche in cui effettivamente si registra un aumento delle spese, fornire una sorta di storico non solo a livello della singola ASL ma a livello regionale (espandendo la ricerca non solo al Piemonte) e in più monitorare il trend e stimare le spese per gli anni a venire.
Un altro lavoro che sarebbe utile e interessante svolgere su questi dati potrebbe essere quello di formare un campione rappresentativo della spesa sanitaria Piemontese di un determinato intervallo temporale, in modo da ridurre il numero di dati su cui lavorare per effettuare un altro tipo di analisi e stime.

Bibliografia

[1] ContrattiPubblici.org, Contratti Pubblici. Contratti Pubblici, [online]. https://contrattipubblici.org

[2] Legge, 6 novembre 2012, n. 190, Gazzetta ufficiale. Gazzetta Ufficiale, [online]. http://www.gazzettaufficiale.it/eli/id/2012/11/13/012G0213/sg

[3] Associazione Nazionale AntiCorruzione. ANAC, [online]. https://www.anticorruzione.it/

[4] Attribuzione 3.0 Italia, Creative Commons. Creative Commons, [online]. https://creativecommons.org/licenses/by/3.0/it/legalcode

[5] Gazzetta Ufficiale, [online]. https://www.gazzettaufficiale.it

[6] Decreto Legislativo, 19 giugno 1999, n.229 Gazzetta ufficiale. Gazzetta Ufficiale, [online]. https://www.gazzettaufficiale.it/eli/id/1999/07/16/099G0301/sg

[7] Elenco ASL Piemonte, Ministero della Salute. Ministero della Salute, [online]. https://bit.ly/2uB0kLb

[8] Rendiconti finanziari del Ministero Della Salute, anni 2015-2016, [online] http://www.salute.gov.it/portale/temi/p2_6.jsp?id=1314&area=programmazioneSanitariaLea&menu=vuoto