POLITECNICO DI TORINO

Master's Degree in Data science and engineering

Master's Degree Thesis

Auditing online software for bias: A controlled experiment in the domain of car insurance

Supervisors

Prof. ANTONIO VETRO'

Prof. RICCARDO COPPOLA

Candidate

OUMAIMA REGRAGUI

APRIL 2024

Summary

As automation increasingly becomes a part of society, processes that reverse engineer and uncover key aspects of algorithms and automated decision systems become essential. Algorithmic auditing plays a crucial role in ensuring fairness, accountability, and transparency. It allows communities to monitor technology and decision-making systems, ensuring they align with specific values and requirements.

The Italian car insurance industry plays a critical role in the country's economic, social, and political landscape. With Italy ranking second only to Luxembourg in car ownership among EU states, auto insurance is vital in making vehicle ownership and usage less financially hazardous for individuals. The industry is primarily governed by the Motor Vehicle Liability system (Responsabilità Civile Autoveicoli - RCA), which mandates the purchase of RCA coverage for all motor vehicles used or kept on public roads. Comparison websites, also known as aggregators, have gained prominence as primary channels for RCA subscriptions, accounting for a significant portion of the total gross written premiums in the Italian vehicle insurance market. These websites have consequently faced increased scrutiny and regulation, especially concerning how they present insurance options and pricing to consumers.

In this thesis work, we carried out an audit of an Italian car insurance comparison website by collecting quotes. We show that some protected attributes such as birthplace have a considerable impact on the prices received by each profile. This feature is used to the disadvantage of non-Italian profiles but also to Italians born in some specific cities. We also showed that riskier profiles (for instance profiles born in a foreign country) were given fewer quotes in the result page of the comparison website. This work confirms the presence of strategic behavior by insurers, affecting some users and putting at advantage other types of profiles.

Acknowledgements

To my dearest mom and brother, my two strongest allies endlessly showering love and believing in me, I say, thank you. You were the ones who made the bumps along the way easier, bringing light to this challenging journey.

To my dearest friends, you are the architects of the joy that has graced this journey. The laughter we shared, the challenges we faced together, and the memories we created are memorable. You have made the corridors of learning resonate with warmth transforming them into spaces of cherished experiences.

And to my professor who laced this opportunity of doing this project, I am profoundly grateful. This project has not just been a task; it has been a journey that I have thoroughly enjoyed, it's a path of growth and discovery.

Each of you has an imprint on my story which I'll never erase, for which I remain eternally thankful.

Oumaima Regragui

Table of Contents

List of Tables

List of Figures

Acronyms

AI
: artificial intelligence

ML
: Machine learning

RCA
: Responsabilità Civile Auto

IVASS
: Istituto per la Vigilanza sulle Assicurazioni

IAIS
: International Association of Insurance Supervisors

EIOPA
: European Insurance and Occupational Pensions Authority

FTU
: Fairness Through Uniformity

Chapter 1

Introduction

1.1 Motivation

The application of Machine Learning (ML) technologies across software products has put the question of fairness at the center of technological discussions. In the era of fast-developing digital technologies, the threat potential of those technologies which are characterized by a significant impact on the life and society of people is enormous. Uncovering and dealing with unfairness issues that may be present across different application domains has become a pressing matter. Several research have indicated examples of errors and undeserving behaviors in industries that have a direct influence on individual lives (For example: loan approvals, licensing, criminal sentencing, employment hiring, and so forth).

This study reveals a crisis in the car insurance industry in Italy, which has experienced many revolutions in recent times. A major part of it can be attributed to the groundbreaking breakthrough in the field of digital technologies and the legislative system that keeps developing. They symbolize a departure from the old ways towards the modern technology-driven approach. Such a move is restructuring the intermediation, pricing, and process of insurance products that affect all consumers.

The principal motivation of the research is to determine the far-reaching magnitude of such paradigm changes in the Italian car insurance sector and measure the impact and value of professional critical audit skills assessment in this context. Algorithmic audits have now gotten more attention as they serve as a crucial method for revealing biases and maintaining the fair nature of the algorithms that drive digital car insurance tools. The concern here is complex, considering it has both a technical and an ethical dimension. On a technical level, this requires us to understand how these algorithms sort through and analyze a lot of data to make choices. On the ethical side, we should make sure these decisions are balanced, without discrimination, and of course, if everyone is treated equally as a stakeholder.

This research is aimed at uncovering and analyzing the triad of technological development, regulatory changes, and market dynamics through a detailed exploration of each risk factor to provide a complete and balanced view of the situation of car insurance in the current Italian market. The aim is not only the actual performance evaluation but also the forecasting of a multitude of future scenarios with their corresponding consequences. This research aims to provide useful results that can lead to the development of practical solutions or approaches for the stakeholders in handling the rapidly changing business arena. It attempts thus to make sure that technological developments are consistent with ethical standards and legally mandated requirements, thus creating an equitable competitive environment designed to benefit all market participants.

1.2 Context of Use

The Italian market of car insurance has been profoundly changed by the emergence of digital comparison tools and algorithmic decision-making. These technologies have not only affected consumers' tastes but also changed the course of markets. The Italian regulatory framework lies at the core of this study because it provides an important background from where the findings can be examined. The Italian insurance market has witnessed a paradigm shift in recent years, mainly due to the more extensive use of digital technologies and algorithms in the processes of pricing and risk assessment. This shift has brought up issues of fairness and transparency especially because the algorithms usually function as "black boxes" meaning users are not fully aware or able to understand how decisions are made. Especially, the incorporation of features like gender, and nationality to set insurance premiums has been a sensitive matter. These sensitive characteristics, such as gender or age, distinguish privileged and unprivileged groups and are typically not appropriate for use in decision-making systems. European Union legislation forbids the direct use of gender as a factor for rate setting. Yet, although some regulations were introduced, there were many indications of discrimination in the audit of pricing algorithms in the Italian car insurance market.

In this study, the place of birth for example of drivers greatly affected the prices quoted which in turn worked against foreign-born drivers and those born in some Italian cities.

Besides that, comparison websites currently (aggregators) serve as a default gateway for subscribing to RCA (Responsabilità Civile Auto - mandatory liability insurance for motor vehicles in Italy). These channels though accounted for about 50% of the overall gross written premium in the Italian vehicle insurance market in 2017. Their increasing significance has led to them being under more scrutiny and regulation. Previous investigations carried out by Italian insurance regulators have revealed anecdotal evidence of output variability associated with the risk profile of these comparison websites. They found out that websites displayed fewer quotes to high-risk driver profiles which could have resulted in differential treatment and unequal opportunities for different driver segments. This context plays a crucial role in understanding how auditing of algorithms in the Italian car insurance industry is undertaken. It shows the importance of these audits in a dynamic digital terrain where regulatory frameworks are struggling to cope with the technology growth. The research will focus on this context to give a general appreciation of the industry status and the effects of algorithmic decision-making on market dynamics and consumer choices.

1.3 Limitations

The limitations of this study are plural and are important factors to be accounted for. Most importantly, in the research, we are limited by the range and depth of the data available, yet we cannot capture the entire Italian insurance industry completely with it. Not collecting full data may lead to such mistakes as missing out some aspects of market movement. Not only does this employ the flexibility of technology, but also the rapid change in the digital realm and algorithmic decision-making makes it a dilemma. The findings of the study, as of the time this is written, could rapidly become outdated because more new technologies and regulations are being introduced. They are subjected to such rapid change that they could seriously jeopardize the applicability of the study's results over time.

The other grand hurdle is in the complexity of the algorithmic decision analysis and interpretation. Algorithms that are complex and opaque such as the ones that are driven by machine learning and artificial intelligence are the most popular ones. Figuring out their decision-making process, looking for an internal bias, and measuring the fairness of their decisions is a hard task. This multi-faceted aspect, however, is accentuated by the enclosed algorithms that are used in the insurance industry, even though it limits outside perception as well as complete comprehension.

Another significant factor is the ever-changing national regulatory framework, particularly in Italy, where new laws and rules are being proposed almost daily. In this dynamic regulatory environment, the findings of a study are influenced by the way that they stay relevant and valid over time given the continued potential for changes that may render some conclusions less relevant or accurate.

Moreover, the study's particularity on the Italian market though it provides an in-depth of local insights, also limits the generalizability to other regions or markets. The car insurance markets in different countries tend to have very different traits mainly because of the differences in culture, economic situation, and rule-making. Thus, the implications of the Italian instance may not be identical to or immediately adaptable to other settings.

This study offers detailed coverage of an aggregator website of car insurance but still doesn't offer detailed coverage of the Italian insurance market, due to the database size. A study involving different aggregator websites and an inspection of all features of each website will have a precise characterization of the pricing algorithm within the Italian market. Also, the field requires continuous research and updating with the swiftly changing sector of technology regulation and market dynamics of car insurance.

1.4 Thesis structure

The thesis is organized as follows:

1. The first chapter introduces the motivation, the context of use, and the limitations and this work.

2. The second chapter reviews the Italian car insurance market. It introduces the concept of algorithmic audit in the context of bias and fairness while presenting state of art studies in both contexts.

3. The third chapter introduces the research method adopted by presenting the data collection method and the analysis adopted to highlight the features involved in bias.

4. The fourth chapter presents the findings and offers a detailed comparison study with the initial study of the car insurance industry.

5. The fifth chapter concludes the thesis with a summary of the research study done and future explorations.

Chapter 2

State of Art

2.1 Background

2.1.1 Overview of the Italian Car Insurance

The way the car insurance system is structured in Italy is featured with several original features. First, the legislation makes it obligatory for all vehicles in Italy to have compulsory civil liability insurance (RCA: Responsabilità Civile Auto, or Mandatory Third-Party Liability Insurance), which ensures the compensation of damages to other cars and the person in case of accidents. This global requirement provides a common shield of protection for all road users and facilities. Another benefit is that, unlike other insurances that tie the people responsible to the insurance holder, Italian insurance is tied to the vehicle. Yet, this factor can also raise complications in the evaluation of risk and calculation of premiums.

Italian car insurances use a system where people are put into different groups, numbered from 1 to 18 (with class 1 being the best and class 18 being the worst) depending on how often they have had to make insurance claims. In the case of a new driver, or a foreign profile requesting car insurance for the first time in Italy, the class attributed is 14. This feature is named insurance classes, and it plays a significant role in determining insurance premiums since it aims to fairly assess the risk associated with insuring different profiles.

Lastly, it is worth noticing that vehicles from EU and some other countries, such as Switzerland or Norway, are saved from the need to sign new policies for Italy as they are protected by their current policies. This demonstrates that there is a certain integration and standardization in the European insurance framework. While this offers EU drivers some privileges over non-EU drivers, it creates differences too.

The Italian car insurance market is not limited to RCA, but offers comprehensive and personal accident policies as well, giving the consumers broader options, but possibly creating a divide in protection and price as well. The regulatory authority, formed by the combination of national and EU laws, ensures the standardization of practices and norms of the entire market.

2.1.2 Digital comparison tools and their impact

The emergence of digital comparison tools has largely reshaped the Italian car insurance market and it has brought more justice and consumers' power. Online comparison sites are incredibly important links that connect consumers to insurers. The platforms offer individualized access to analyze different policies in their coverage, price, and available extras. The provision of this level of transparency offers a positive development towards fairness as it provides consumers with the necessary information to make the right decisions. It is in this informed manner that consumers will be able to acutely see the market rates and all the specifications of the coverage they are purchasing.

The influence of these on the insurance companies is no less than an expectancy. The higher publicity and lower chances of providers to escape scrutiny by customers drive insurers to be more competitive both in pricing and the services they offer. This greater competition can lead to more favorable conditions for consumers, which are essential for creating a fair competitive environment.

Nevertheless, the degree of how fair the digital tools are in presenting the information is contingent on their capability of making the information unbiased and complete. If these platforms collaborate with car insurance companies and in the meantime omit some key insurers from the comparisons, the consumer may be misled by the resulting distorted view. Therefore, the integrity and inclusiveness of these comparison platforms are critical. They must ensure that all significant insurance providers are represented and that their rankings or recommendations are based solely on objective criteria relevant to consumer needs.

Moreover, the ease of use and high degree of availability of these tools are also very important elements of their effectiveness. They should be built in such a manner that the user experience should be good for a wide group of users. Inclusivity means all the population groups benefit from the policies which only strengthens the principle of equality in the insurance industry.

2.1.3 Regulatory framework in Italy

The Italian car insurance regulation is comprehensive and multidimensional, creating adequate protection for both the insurers and the insured and making all the parties subjected to the control of the regulatory bodies. What forms the basis of this framework is the Italian Insurance Supervisory Authority (Istituto per la Vigilanza sulle Assicurazioni – IVASS), which is responsible for regulating it. This body operates together with international organizations such as IAIS (International Association of Insurance Supervisors) and EIOPA (European Insurance and Occupational Pensions Authority) to provide stability for the market and to adhere to international standards.

This framework functions under a tighter regulatory framework that safeguards consumers and fair practice. One of the key features of this framework is the prohibition of the utilization of gender and place of birth for discrimination purposes in designing insurance premiums directly. Even though the insurers could collect this type of data, the EU legislation and Italian regulations limit their direct use in risk models. As an illustration, women and men have not been allowed to be considered by insurers in different premium rates since a judgment by the European Court of Justice in 2011. Likewise, rules that used either nationality or birthplace as risk factors were criticized and subject to regulatory action, and the abolition of premium variation by the country of the driver's birth was proposed. Such measures are developed to avoid discrimination and guarantee fair treatment within the insurance sector. This may be associated with some flaws in terms of transparency and the possibility of quote ratings disparity and therefore constant surveillance is required to correspond with regulatory requirements.

2.2 Algorithmic audit

2.2.1 Understanding Algorithmic audits

Algorithmic audits are a necessary approach for evaluating and assessing automated systems, especially at a time when algorithms dominate decision-making. These audits aim to ensure the appliance of ethical standards, fairness, and legality with decision-making algorithms. The process entails a thorough audit of how inputs (including the user data) are transformed into outputs (like insurance premiums and credit scores) and whether there are any biases or unfair practices involved in this transformation process. Algorithm audits, which involve dissecting and understanding the decision-making logic behind algorithms, help in identifying biases, thereby preventing the spread of discrimination and inequity. However, this is more important in sectors such as finance, insurance, and healthcare where algorithmic decisions have profound individual implications. Through algorithmic audits, companies can remain transparent and answerable to their automated processes hence fostering trust from users and stakeholders.

These audits can be divided into several types, e.g., functional audits which check how good an algorithm is in performing its intended task, and process audits which are targeted at the factors influencing the development of the algorithm that includes the data selection and model training. Evaluation of the consequences of algorithmic decisions becomes particularly relevant in the context of bias and fairness; they identify any bias or disparity in outcomes between different groups. This means evaluating whether some groups such as those based on gender, race, and age tend to be discriminated against by the algorithms. Through identifying and addressing these biases algorithmic audits contribute to the operation of keeping the automated systems fair and just, consequently gaining trust and being in line with ethical and legal standards.

2.2.2 Process of Conducting an Algorithmic audit

Running an algorithmic audit is a complex and multi-step procedure and a key success factor in practically assessing fairness and bias in automated systems. The audit starts with a clarification of objectives, and identification of the specific portions of the operation to be audited like ‘bias, transparency, and adherence to ethical principles. The next stage in the process is data gathering during which the auditors analyze the data inputs and outputs of the algorithm, including the training set it was taught upon and the decisions it makes.

During the analysis stage, auditors take the algorithm’s decision-making mechanism apart and scrutinize its logic, rules, and criteria of decision-making. This includes understanding how the algorithm was constructed — which attributes are considered, and the weightage given to each of them. This stage is rather crucial because, at this point, features involved in bias can be detected and prevented.

Impact assessment is the next stage of the audit that focuses on the major influences the algorithm has on a wide range of real-world issues. It is also a statistical analysis that is used to detect patterns of bias, such as when specific groups experience more of the impacts of the issue. The auditors utilize equity metrics as well as benchmarks to assess the outcomes in the context of established standards of fairness.

For the final step, the audit ends up in a conclusive report that explains the results and the suggestions. This report gives a comprehensive evaluation of the bias issues, areas of doubt, and recommendations on the areas that need to be addressed. One of the main functions of the algorithm is to provide stakeholders with an insight into the implementation of the algorithm and advise on modifying or enhancing the algorithm to ensure compliance with ethics, law, and social norms. This circular approach keeps the systems in control and not subject to bias.

2.2.3 Challenges in Algorithmic auditing

Conducting algorithmic audits is a multifaceted task that brings a lot of problems. Along with this comes the problem of restricted access to corporate algorithms and data sets as companies consider them as their secret intellectual property, which minimizes their accountability to the public. Together with the intricate nature of machine learning models and their underlying math, comes another level of complexity that calls for domain expertise for the correct interpretation and analysis. The dynamics of algorithms by their nature alter according to the incoming data and the interactions. As a result, it is challenging to audit the behavior of the system at any given point in time.

Moreover, fairness in algorithmic decision–making is typically dependent on perspective besides being culturally sensitive, socially sensitive, and legalistic. This variation is an impediment to the implementation of justice in an international setting. Loyalty accuracy and discrimination constitute a different challenge, especially in applications where decision accuracy is of paramount importance while bias may give rise to irrevocable repercussions.

Furthermore, biases in training data laid inadvertently can cause bias in the systems. It also complicates auditing. The most technical solutions are not enough. To solve the problem, it is important to know not only how the data was collected but also the socio-cultural context. Besides, AI and machine learning technologies are evolving at an extremely fast speed which implies the auditors need to stay informed about the latest updates in terms of methodologies and skills. Such issues shed light on the importance of persistent continuous work in research, policy designing, and training to carry out efficient algorithmic audits.

2.2.4 Case Studies in Algorithmic auditing

Case studies in algorithmic auditing have been instrumental in highlighting the importance and challenges of this process. Different studies were conducted in different domains:

1. In the field of social media, an audit [1] was conducted to evaluate Twitter's algorithmic curation system. The research team formed fake accounts (audit bots) to observe the impact of the timeline algorithms on the visibility and distribution of tweets. They identified biases in how the curated timeline presents information, particularly in terms of popularity, recency, and exposure. The main finding of the study was that the Twitter algorithm is likely to show more popular content which may be a reason to give a distorted impression of users' friends' activities and the whole information environment. The methodology

and results showcase several meaningful implications of algorithms in curation for social media users.

2. In the field of social justice, [2] showed an example that illustrates how artificial intelligence in algorithms can cause unfairness with Amazon's AI recruitment tool as an example. This tool was found to disadvantage candidates whose resumes included phrases such as "chess club captain women's ", which revealed the gender of the candidate even if the tool didn't explicitly use this feature in its algorithm definition. This example pinpoints the implicit bias, which is encoded into algorithms, however, there is no intentionality in it to discriminate against women. Another example in the field of social justice is illustrated by [3], the paper focuses on the gender bias in recommendation algorithms for jobs on Chinese job boards. The behavioral experiment constructed male and female workers' identical profiles on these platforms and then examined the job recommendations they were given. It puts into evidence some significant gender-related patterns in the recommendations. For instance, females received jobs that paid the lowest wages and required fewer years of experience, as opposed to males. This clearly shows a gender imbalance. This study highlights the need for algorithmic audits to bias identification and appropriate suggestions, therefore making the systems more transparent and fair.

3. In the field of healthcare, [4] examined in terms of racial bias in healthcare. This research discovered that the algorithm with the highest probability of flagging patients for additional healthcare assistance also exhibited substantial racial bias. It demonstrated that black patients were severely sicker than white patients. Thus, this was made due to the algorithm estimating healthcare costs rather than illness, combined with unequal access to care leading to lower spending on black patients. This study emphasizes the fact that algorithms in healthcare have a racial bias, which indicates the need for a conceptual framework and auditing of the algorithms so that they can serve for proper and equitable healthcare management.

2.2.5 Algorithmic audits of car insurance

The study "Algorithmic Audit of Italian Car Insurance: Evidence of Unfairness in Access and Pricing" [5] which is the base study of this work, focuses on the Italian car insurance industry. It reveals significant findings about the influence of birthplace and gender on car insurance pricing. The audit shows that despite regulations against such practices, birthplace, and gender impact quoted prices, with foreign-born drivers and those from certain Italian cities facing financial disadvantages. This study emphasizes the need for algorithmic audits in the car insurance industry to ensure fairness and adherence to regulatory standards. The study has been done by collecting data on a comparison website, which shows that even though there are national and international regulations against their use, nationality, and gender influence insurance quotes for drivers. Notably, foreign-born drivers and those born in certain Italian cities face disadvantages. For example, in the case of Laos drivers, they were charged up to 1000€ more than those in Milan who had a similar profile. Furthermore, as the research showed, the users who were regarded as risky received fewer quotes on the aggregator search results pages which are aligned with previous complaints about unfair actions reported by insurance regulators from Italy.

Other studies in this domain but in different markets have also highlighted similar issues, Harrington, and Niehaus utilized a dataset from Missouri to verify whether insurance profits are greater for profiles that live in minority-dominated ZIP codes. Another work conducted by Ong and Stoll revealed different findings. They gathered 836 quotes changing only the ZIP code and the other conditions were kept constant. The researchers concluded that some risk factors, and socioeconomic factors in a neighborhood, such as the percentage of poor residents and black residents, correlated with higher premiums. Also, ProPublica analyzed car insurance payouts and premiums in California, Illinois, Texas, and Missouri, coming to similar conclusions that affect minority neighborhoods unfavorably.

2.2.6 Impact of Algorithmic audits on regulations

The effect of auditing algorithms on the regulatory field is profound and extensive. This, in turn, highlights the importance and role played by audits in influencing and shaping the formulation of regulations on algorithms and automated decision systems. Biases and unfair practices can be uncovered that can also provide concrete evidence that can lead to stricter implementation and enforcement of current anti-discrimination laws. For example, algorithmic audits in the areas of hiring, lending, or insurance can show bias against certain demographic groups, and not only identify the areas for immediate improvement but also give proof to regulatory entities about more strict guidelines.

Additionally, these audits must be utilized to create new regulations. With technology development, comes new ways of introspection which open the agenda to greater forms of bias and improper behavior. Algorithmic audits will identify these problems in the early stages allowing regulators to reformulate adequately the policies. This is of critical importance in cases where algorithmic decision-making can make profound differences in people's lives, as it is in healthcare, criminal justice, and financial services.

These audits are key factors in establishing transparency and accountability in the system as well. They decompose algorithms' decision procedures, which may cast light on the alignment with ethical and legal standards. This can result in regulations that ask companies to reveal their algorithm's functioning mechanism and make sure that the algorithm is auditable and interpretable. It is only by transparency that fairness is attained in addition to the trust of automated systems by the public.

On the other hand, algorithmic audits, to some extent, provide standards for managing and creating qualitative data with the assurance that the data is representative and without biases. The latter has an immediate effect on the data collection, storage, and usage rules.

The flexibility of technology requires regulations to be adapted and renewed regularly. Policy adjustments that technology regulators need for the reforms are provided by algorithmic auditing, which enables them to give timely feedback.

Overall, the increase of using algorithmic audits is leading to more informed regulators' strategy, where decisions become data-driven and tasked with achieving fairness and accountability in the digital age. This way is very critical to have the values of people and society close to the uses of automation and AI.

2.3 What is bias?

2.3.1 Definition of bias

Bias includes a range of unfairness and systemic problems in automated decision systems. It depicts a deviation from objective, fair, and equal treatment being replaced with a dispensation of preferential outcomes due to features like race, gender, age, or socio-economic status. Several sources of bias can develop during the algorithmic life cycle as the data used to train such systems can contribute to this. For example, if training data is not a good representation of the general population or includes historical biases, the algorithm will more likely perpetuate or amplify such biases in its predictions or decisions.

The repercussions of algorithmic bias are extreme and can intensify societal inequalities. Biased algorithms in aspects like criminal justice, healthcare, hiring, and lending can result in prejudiced outcomes, like unjustified sentencing, ineffective medical care, unfair hiring, and unfair credit denial. Beyond the personal, this enhances systemic inequalities at a larger scale as well.

Discrimination and its elimination represent a complex interaction of technical and ethical factors. Bias if carefully looked at may originate from the subjective nature of the data, the assumptions and the values of the designers integrated in the design, and the context in which the data was collected, and the algorithm is implemented. Addressing bias also means engaging with the multifaceted nature of fairness and the realization that different definitions of fairness might be appropriate in different contexts.

Additionally, the dynamic nature of algorithms, which learn and evolve, makes the detection and mitigation of bias a continual challenge. When algorithms adapt to new data, there is a possibility that such adaptations can lead to the creation of new biases or the aggravation of existing ones if not properly monitored and adjusted.

2.3.2 Different types of bias

Concerning decision-making and algorithms, there is more than just one type of bias, each with its own wholly distinctive characteristics and possible outcomes.

1. Measurement Bias: This happens when the tools or method is employed to get faulty data. Such as, if facial recognition processes the data with some inaccuracy while recognizing the ethnic groups, the final data will reflect these errors. Bias of such kind may be illustrated by the example of the recidivism risk prediction tool COMPAS. This tool uses previous arrests and arrests of friends and relatives as proxies to define riskier profiles. These proxies may be perceived as crude indicators and they do not provide the proper meaning on their own. This is partially because minority communities endure an overburden of law enforcement and surveillance. If people from minority groups are usually the ones who are arrested, this does not necessarily imply they are committing more crimes [10].

2. Omitted variable Bias: The omission of one or more essential factors from the model causes omitted variable bias. This kind of bias poses a significant risk in data-driven decision-making, for instance, predictive modeling or machine learning algorithms. For example, in the field of financial modeling, if important economic indicators are left out, the forecasts would considerably miss the target. Therefore, in predictive analytics in healthcare, neglecting the relevant patient's history and/or demographic information can cause misunderstandings regarding the effectiveness of treatment or patient risk profiles. Omitted variable bias is the issue here because it can shift the outcomes in an insidious way that might not be obvious to the observer until it is too late, hence a model may seem correct on the surface but might lead to the incorrect decision. Therefore, the complete data analysis process and development of models is an important requirement to look at all the relevant variables to make it reliable and accurate [11] [12] [13].

3. Representation bias: This happens when the data set used to train an algorithm doesn't fully represent the target population. In car insurance, if the training data is heavily skewed towards urban drivers, rural drivers' risk might not be accurately assessed [10].

4. Aggregation bias happens when the conclusions drawn about an individual member in each group are made from aggregate data obtained about the group. This bias can lead to a situation in which the group, not the group's members is incorrectly assumed, and wrong decisions are taken. For instance, in the public health domain, many assumptions are made about individuals' health practices based on the aggregate health data of the community. These assumptions may be misleading. Should the community have low disease levels, that doesn't mean that each person in that community is at low risk of this disease. Schools have the same problem when test scores are averaged. It could be that the aggregate test scores of a school might not tell whether the students can learn or what their learning needs are. This type of bias underscores the need to investigate the individual data and refrain from overseeing the general statistics of the group, especially in those areas in which those personal differences are considerable [14].

5. Sample Bias: This happens when the data applied to train a model is not general enough. For example, for credit scoring, if an AI model is trained in large data mainly from middle-aged people, it fails to assess credit worthiness of younger or older people accurately [14].

6. Historical Bias: Algorithm or data-driven system bias stems from system inequalities, prejudices, or practices that are inadvertently included in the data used [10]. Such prejudice may carry severe consequences, as it fuels the discrimination and injustice handed down from the past to the present and even future choices made by such systems. As in the case of hiring algorithms, if the older enduring data indicates a predilection for hiring applicants usually from a specific gender or ethnicity, then the algorithm will in another way continue to lean towards hiring the applicants from that group, and therefore, historical discrimination will still be practiced. In addition, likewise in predictive policing, if previous crime data shows historically biased policing against some communities, then this data could be used to train algorithms to still over-police these communities. Historical bias is the reason for critical examination and possible corrections in case historical data is used for training algorithms and such algorithms don't reinforce outdated and unjust practices.

7. Population Bias: If a sample used to train an algorithm cannot represent the target group accurately, then the algorithm will give biased results. This form

of prejudice may greatly influence the result of a prediction model in terms of accuracy and fairness. For example, in the health sector, taking the case of drug development, a drug design based on data primarily from one ethnic group, might fail to work for other ethnicities also. In the same manner, if the training dataset for facial recognition technology is mainly populated by pictures of people from certain racial groups, the system may not be capable enough to correctly identify people belonging to the underrepresented groups. Social media platforms may also show disproportionately age bias; for example, if most of the data is gathered from a particular group of people then the output of content may not capture the interests of users out of that border. Structuring bias out of the population is essential that AI systems are equitable and exact in the entire groups of users intended.

2.3.3 Use cases of bias

In the arena of data science and algorithmic decision-making, biases can show up in different use cases; as a result, such biases can have significant repercussions across different sectors. Here, we explore some prominent examples where biases have been observed and their impact on real-world scenarios:

1. Healthcare: The historical bias in healthcare data may be cause for the bias in AI models. One instance is that if clinical trial data have been largely with one gender or race, predictive models that are developed will be less accurate for the other groups. Such a development can lead to overdiagnosis or inappropriate treatment for the underrepresented populations [18].

2. Criminal Justice: Predictive policing and risk assessment tools can mix aggregation bias and historical bias, which can reinforce existing societal issues. Such as in case the arrest data reflecting the bias of police will be used to predict future criminal activity, one may end up with advanced surveillance of some communities [17].

3. Hiring and Employment: Biased decisions on job potential in AI–run recruiting systems may reinforce unequal hiring practices. If variables that fall into the non-technical skill or the unconventional education background basket are ignored, the candidates who do not fit the traditional profile but may be skilled for the job might be passed [16].

4. Credit and Lending: Population biases are present in financial services that bias algorithms in credit scoring. If the training data only contains data from a certain social group, the algorithm might unjustly favor or punish only those social groups leading to unfair lending decisions [20].

5. Education: Educational tools also are prone to aggregation bias that has adverse implications for better teaching practices. If a learning platform analyzes general data to provide a customized curriculum that does not consider the different requirements of every student, it will not help the students to effectively deal with their learning problems [21].

6. Marketing and Advertising: An advertising algorithm's measurement bias can lead to stereotype formation and bias reproduction as part of societal stereotypes. This can be demonstrated by citing an example like the case where an ad targeting algorithm merely tallies the level of user interaction without considering the context and presenting the same stereotypical content to a specific demographic [22].

7. Social Media: What an individual sees as a part of the content is affected by the representation bias on social media. If training the algorithm is done on a data set predominantly from the other user groups, the algorithm will not be able to cater appropriately to the varied interests of the entire user base [19].

8. Facial Recognition: Sampling bias in facial recognition technology results in uncertainties, particularly in situations where there is poor representation. If the training data does not relate to the different identifiable categories such as gender, age, or ethnicity, the system can fail to correctly identify the individuals from those groups [23].

2.3.4 Consequences of bias

The consequences of bias in the algorithms and data-driven systems reach further than just direct impacts, creating chronological societal and ethical difficulties. During recruitment, biases result in employment inequality, leaving income differentials and workforce diversity less diverse. In the criminal justice system, biased algorithms create racial inequalities even worse, they impact bail, sentence, and parole decisions, so that systemic injustices are continued. Biases inherent in diagnostic tools and treatment algorithms in healthcare can lead to misdiagnoses and unjust treatment allocations, inequitably affecting marginalized groups and aggravating health disparities.

The role biases have is not just limited to these domains. It is in the financial services sector that biased credit scoring algorithms can be an explanation for unequal access to loans and financial products that are affecting wealth accumulation and economic mobility. The bias in student assessment tools can also limit educational opportunities and outcomes, which can have adverse effects on their academic career.

Another factor besides biases in algorithms is the loss of trust in technology and the entities that implement these systems whether they are government entities, corporate entities, etc. Such erosion of trust, no matter how small it is, can have big consequences since this may lead to suspicion of technological developments among the public.

The indirect damages resulting from algorithmic bias should raise alarm as well. They sustain and normalize the existing prejudices and stereotypes, which, ultimately, turn into an inequality and discrimination cycle. Overcoming these biases is not only a technical problem but also a moral issue that calls for a multi directional approach that gets the technologists, policymakers, and social scientists to work together to guarantee fair, equitable, and socially just outcomes in this AI and data era [24][25][26].

2.4 What is fairness?

2.4.1 Definition of fairness

Ethics of AI and algorithm audits is one of the most important issues in this context, which covers vast areas of ethical, technical, and social aspects. In AI systems, fairness means avoiding biases and eliminating the perpetuation of biases that already exist in society. Therefore, AI algorithms must function without manifesting preference or discrimination against specific people groups based on race, gender, age, and social value. Holding this level of fairness requires a profound knowledge of both the technical elements of AI like the data sorting mechanics, model training, and output interpretation, and the social and ethical consequences of such technologies [27].

Algorithmic auditing is invaluable in providing insights into the fairness of an AI system. These audits go for the watertight analysis of each phase of the AI development process, starting with the initial data collection and preparation to the final decision-making outputs. Auditors investigate if the dataset fed into the algorithms to be trained encompasses both bias-free and representative data, and whether the algorithms are fairly and unbiasedly processing this data. They measure the transparency and accountability of the AI systems, which makes the decisions understandable and vacuous.

The aspect of fairness in AI is not only technical accuracy but also civilization principles, i.e. to make AI technologies trustworthy. For instance, biased artificial intelligence algorithms can have serious real-life consequences, for instance, supporting prejudices, worsening social injustices, and decreasing the faith of the people in vital institutions. As an example, while unfair AI used in criminal justice could result in biased sentencing; same in the field of healthcare it may cause unequal treatment of patients and in finance, it could cause unfair loan or insurance premiums assessments.

Fairness in AI is a dynamic process and must be monitored regularly, especially with the proliferation of machine learning systems. It necessitates a multidisciplinary perspective with the involvement of specialists from the fields of technology, ethics, sociology, and the law. Along with AI becoming more embedded in the different sectors of society, fairness in these systems does not only constitute a technical challenge but a moral imperative, providing the basis for the equal distribution of benefits from AI and ensuring that these technologies serve the good of the society [28][29].

2.4.2 Previous related work

Fairness in algorithmic audits has a long and opulent history of previous studies and research that have founded the current view of the field and its current practices. Before these were associated with bias detection and mitigation models, the early works in this discipline were primarily based on statistical models. In this regard, research on "Big Data's Disparate Impact" by Barocas and Selbst [30], which was foundational, demonstrated that discrimination could be reinforced as a result of data mining as well as algorithmic decision-making.

Afterwards, additional studies were carried out to examine many possible implications of fairness in AI, including the works of Dwork et al., on "Fairness Through Awareness," [31] which introduced a new framework for considering individual fairness in algorithms. This strategy underlined the need for reasonable equality in the treatment of individuals, rather than flat distribution among the demographic groups.

In the practical world of implementation, Buolamwini and Gebru's foundational "Gender Shades" [32] work illustrates the role of discriminatory algorithms in commercial facial recognition systems, which underscores the relevance of fair algorithms. This study serves as a milestone in shaping people's views about the drawbacks of the current AI technology and that of requiring suitable and widely representational training data.

The multidisciplinary dimension is also involved in the research of this area. Eubanks (2018) in "Automating Inequality" and O'Neill (2019) in "Weapons of Math Destruction" [33] highlighted the sociological and ethical dimensions of algorithmic decision-making; in specific, how these systems can impact minority groups of society.

However, what has recently become a focal point is the development of frameworks for algorithmic audits and algorithm tools that can be operationalized to implement validity. Studies, such as 'A Framework for Understanding Unintended Consequences of Machine Learning' [34], were fundamental in the realization of the fairness principles following the AI systems for both practitioners and policymakers.

No single work has ended the discussion on fairness in AI, it has rather opened it up to further complexity and the need for researchers from multiple fields to work together to ensure AI systems are technically sound as well as ethically responsible and socially equitable.

2.4.3 Importance of fair dataset

Fairness in AI works can be approached ethically, technically, and in practice. Representative data sets are those which are balanced and include varying elements that make up the target population, to make sure that AI models are trained on a wide spread of input data. This diversity is needed for the making of AI systems that are impartial on any basis and equitable in their usage. For instance, in the use of facial recognition technology, diversification that includes ethnicities, ages, and genders is crucial for the system not to misidentify people or fail to recognize certain groups. While, in the case of AI in healthcare, datasets that contain differing patient medical backgrounds, genes, and lifestyle variations will be critical for developing accurate and universally applicable diagnostic tools.

The component of fair data is so important in avoiding algorithmic bias - a significant concern where AI systems inadvertently perpetuate societal biases. Discrimination towards certain groups can arise from biases in datasets. This can happen in the context of employment, credit approval, or criminal justice. To this, AI in organizations raises questions both of ethics and risks for the company in juridical and reputation terms.

Moreover, the quality of the datasets dictates the credibility and applicability of AI systems. Such models trained with detailed databases are expected to operate with consistency in various scenes and groups, thus increasing their reliability and efficiency. Particularly in the case of global applications, this is important because such systems need to operate in a manner that considers the different cultural and demographic environments.

Furthermore, keeping biases out of data presents societal norms and future technological development which are oriented towards equity motive. It also reveals the obligation on the part of technologists and multinational corporations that the technologies designed by them should not widen social divides. Fair datasets are not merely technical prerequisites but a touchstone of ethical and responsible AI development that creates trust in the AI systems to harness their potential for societal progress [35][36].

adjustbox

Chapter 3

Methodology

3.1 Research design

3.1.1 Wohlin's Framework

Wohlin's Framework is a purpose-directed methodology for carrying out empirical investigations, treating them carefully and precisely through the design of studies based on evidence as well as the selection process of variables. Key components of this framework include:

1. Defining Research Questions: Especially defining the purpose and goals of the research clearly, particularly which aspects would be explored and commonly occurring facts to be learned.

2. Variable Selection and Operationalization: Judiciously choosing the correct variables, both for intervention and for outcome in a manner that is measurable. Also, the decision must stick to the theme of the research.

3. Designing the Experiment: First, decide on the experiment type (experimental, quasi-experimental, observational) and then, racialize the population, and sample size and outline the data collection and analysis techniques.

4. Addressing Validity: Designing an experiment should allow for the study's validity through the evaluation of cause-and-effect relationships inherent to the study, generalizability (the applicability of the study's results to populations beyond the sample group), accurate concept interpretation, and reliable measurements (using the same measurement tools in the same way throughout the study).

5. Data Collection and Analysis: Systematically gathering and thoroughly subjecting to analysis the data that is produced within the specified analytical

framework.

6. Interpreting Results: The research results are discussed systematically and the connection to the research questions and the current literature is clarified. The scope of the study is considered as the consequences of the results might be outlined.

7. Reporting and Documentation: Full transparency documents all processes namely designing, executing, results, and interpretations as a means of replicating and scrutinizing the methods.

Wohlin's Framework stands out for ensuring that empirical study is as methodologically strong as possible but is also meaningful and relevant to the field of study. The latter does prompt researchers to critically engage with their research questions and try designing studies that will form a spread of knowledge for their domain of specialization. The structure is especially significant because it gives researchers a way to navigate through experimental design issues in technology-related disciplines and guarantees that the statistical studies they undertake are resilient, replicable, and ethically sound.

In our study, we implemented this framework through a full factorial experiment, carefully selecting variables that are both protected (such as gender and birthplace) and recognized as significant for pricing in car insurance (including driver age, municipality of residence, car yearly mileage, and claim history). This method allowed for a rigorous examination of each variable's impact.

Wohlin's Framework provides a comprehensive structure for designing and interpreting research studies, particularly in fields where multiple variables can influence outcomes, such as in the case of the car insurance pricing study. The different classes of variables in this framework are defined as follows:

1. Independent Variables: In Wohlin's Framework the dependent variables are the things the researcher actively alters or inters to create an effect. In the situation when certain experimental research talking about car insurance pricing is carried out, such factors as the age of a driver, gender, and birthplace, for instance, can be considered. Amongst the studies, these factors are shifted to observe their impacts on other important variables, for instance, the insurance pricing rates in the studies.

2. Dependent Variables: These are precisely the bottlenecks the study is set to identify and specify. It is the outcomes it intends to clarify. Writer Wohlin's Framework involves that dependent variables are conditioned by the alterations of independent variables. A car insurance study could have the dependent

variable include the insurance quote given to different drivers. This is an effect measurement that can be quantitatively reflected by the changing of the independent variables.

3. Context Variables: Wohlin's Framework represents context variables that do not directly influence the results in an experimental study but can play a more significant role in its outcomes. The factors may subsist with large economic environments, wider legal tapestry related to car insurance or the more particular conditions of the car insurance industry. These variables cannot have any effect on the study being conducted by the researcher, but their impact is acknowledged while analyzing the results.

4. Control Variables: According to Wohlin's Framework, control variables are constant factors that are used to maintain the study with the change of independent variables, and therefore the results come from the manipulation of the independent variable not from any other extraneous influences. For example, we can structure our case according to the number of competencies of the driver while keeping the vehicle type a control variable. This guarantees the tangibility of the impact of our research as any impact on the pricing is relevant to the age of a driver and not the car insured.

3.1.2 Design of experiment

The experimental plan is of a fully factorial design of the experiment involving different defined features, including both protected features and those that the businesses commonly rely upon in insurance pricing, including the age of driver, municipality of residence, car type, yearly mileage, and overall claim history. This format will make it possible to achieve a ground-level in understanding of the impact of various independent and combined covariant effects on pricing. Based on Wohlin's Framework in the context of the study of car insurance study, variables are classified differently relying on how they influence pricing.

1. Independent Variables:

   (a) Birthplace: Such differences may affect insurance premium rates on the grounds of the risk perception that some blame certain regions. The driving cultural difference for a single country can be significant. Road conditions and accident rates can also vary from one country to another. Such situations might contribute to the risk analysis assessments by insurance companies.

   (b) Gender: In the past, car insurance rates have often been set based on the gender of the driver. This is because gender was commonly used to differentiate between men and women as well as the accident statistics associated with each of them.

(c) Marital Status: Marital status may suggest that a person is stable and diligent and that would seem to be helpful to the insurance companies in discerning what kind of risk they represent as drivers.

(d) Educational Qualification: Higher educational degrees might explain associated driving behaviors with generally more responsible driving habits as well as lower risks for filing claims, which regularly could reflect insurance premiums.

(e) Profession: Several jobs may be deemed as either more or less of a security risk due to external influences such as levels of stress, time spent on the road, and working hours.

2. Context Variables:

   (a) Age: Age is the principal context factor, given that the youth might be less affected while the older population would have higher risks of dying. Young drivers are likely to come across steeper insurance bills because of their novice presence on the roads and higher risk of getting into an accident.In fact, the variable of age is the main determinant of the amount to be paid as a rate for premiums. The data from the last RCA (Responsabilità civile autoveicoli) contracts reveals the sensor in premiums depending on the age group. Particularly, the age groups 18-24, especially, face the highest premiums upon their approval, which are two times more than the national average. This is an example of a common belief of the insurance industry which is based on a general view that those who are younger are considered as representing a high risk due to their less experience and not uncommon risky driving behaviors. The age of drivers is inversely proportional to their premiums, as the latter gets lower. Continuing to decrease until the age category of 35-44 the premiums begin to fit the national average. The fact that older drivers are in demand for a share offers the idea that middle-aged drivers are considered less risky as they possess a lot of driving experience and probably highly disciplined lifestyles. As a result, this concentrates the insights on the relatively young segments of the population, for example, 18, 25, and 32 to figure out how age affects the insurance premiums. Such age groups have been a goal since they reflect different degrees of personal driving experience and/or life stages which replicate the Italian population. Consequently, we find out that by 18, they behave as new drivers, by 25, their driving experience is about 7 years and by 32, they are used to driving approximately 14.

This gradual analysis provides the ability to break down profile plans, which change with the driver's experience.

(b) Class: The class system, primarily the Bonus-Malus System (BMS), plays the main role in the completion of insurance ratings according to the driver's claim history. Under this system, a driver's class is determined by their driving history every year. Those with fewer tickets or no tickets get fewer penalties while those who get ticketed repeatedly have higher charges added to their class.

   i. Range of Classes: BMS rating scales runs from 1 to 18 where 1 is the best score and 18 the lowest. These classes are made according to the prior history of my driver which is the number of claims occurred previously.

   ii. Starting Point for New Drivers: Class 14 generally is a place where we start to learn. The provision stipulates that a newbie driver can acquire their BMS license at the expense of their relative who is living in the same household as them when they purchase their first car read more This option might well be a good starting point for new drivers, and the learners may have the same good record as their relative.

   iii. Yearly Class Adjustment: The driver's record of the given year is 1 more if the driver has no accident record and 2 more if the car accident is caused by the driver. This system is an excellent motive for safe driving as due to class changing to a more favorable one, the premiums get lower over time without any claims of responsibility.

   iv. Special Classes: The system recognizes a distinction between class 1 and "class 1 for one year or more"; the latter is labeled as "class 0" for clarity in some contexts. Additionally, profiles with no driving record are tested as class None, which is considered equivalent to class 14.

   v. Impact on Insurance Premiums: The BMS class is a significant factor in determining insurance premiums. A lower class (closer to 1) typically results in lower insurance premiums due to the perceived lower risk, while a higher class (closer to 18) indicates a higher risk and therefore higher premiums. This class system in Italian car insurance underlines the importance of maintaining a good driving record. It directly affects the cost of insurance premiums, making it a crucial variable for both insurers in assessing risk and for drivers in managing their insurance costs.

(c) Km Traveled in One Year: The higher this mileage, the greater amount of road time will be spent, rendering the accidents more plausible, and therefore raising the insurance costs.

   (d) License Plate: Plate number can tangibly point out the licensing area, which may in turn affect the insurance premiums due to the disparity in traffic congestion, occurrence rate, and vehicle thefts in different regions.
   (e) City: In other words, choosing between plots of land in the city or country will bring huge differences in the price stability of the insurance rates, it is the traffic level, road conditions, accident rate, and so on that matter.

3. Dependent Variables:
   (a) Insurance Premium Rates: The study focuses on the dependent variable, caused by the interaction of the independent and context factors.

4. Control Variables:
   (a) Where the Car is Kept, Vehicle Ownership, and Regular Driver Status: These factors are controlled to eliminate external influences on the premium rates, such as the risk of theft or accidents.
   (b) Number of Cars in the Household and Presence of Other Drivers: More cars or drivers in a household could imply higher risk, impacting the premiums.
   (c) Youngest License Holder's Age in the Household and Cohabiting Children: These factors are controlled as they could significantly influence the risk assessment due to the presence of younger, potentially less experienced drivers.
   (d) Driver's License Status: Demerit points, suspensions, and the date of the license matter a lot to the insurance company while evaluating the individual, they have to bear the costs/risk of the person.
   (e) Claim History: A claims history of the last 6 years is a direct reflection of risk and a top item in the underwriters' decision on the insurance premiums.
   (f) Cohabiting Children: The occurrence of kids, mainly driving-age teens, in a particular household, acts as a warning sign in a risk assessment. Consequently, this results in an increase in insurance rates.

Each of these variables, in the context of Wohlin's Framework, plays a specific role in understanding and predicting how various factors impact car insurance pricing. By manipulating independent variables and observing their effects on the dependent variables (insurance premium rates), while controlling for other influencing factors, the study can provide valuable insights into the determinants of car insurance pricing and flagging where the bias is occurring.

Independent variables	Context variables	Dependent variables	Control variables
Birthplace	Age	Insurance Premium Rates	Where is the car kept? Whether the contractor is also the vehicle owner? Whether the contractor is the regular driver? Number of cars in the household? Presence of other drivers? Youngest license holder's age in the household, Is the car owned by you?
Gender	Class		Driver's license status Year of obtaining license Claim history (past 6 years) Usage of merit class of another vehicle
Marital Status	Km traveled in one year		Year of purchase; Car setup features (e.g., Burglar alarm); Usual use of the vehicle
Education qualification	Licence plate		Cohabiting children
Profession	City		Usual use of the vehicle

Table 3.1: Classification of variables by Wohlin's Framework

The values of the variables used in this analysis are summarized in the table below:

Variables	New study
Birthplace	Milan, Naples, Rome, Morocco, China
Gender	Male, Female
Marital Status	Married, Single, Widow
Educational qualification	Master, Without a qualification
Profession	Employee, looking for a job
Age	25,32
Class	1,4,9,18
Miles traveled in one year	10000, 30000
License plate	OLED, NSEP
City	Milan, Naples
Total Number of queries	7680

Table 3.2: Values of all different variables

3.2 Data collection

Data was collected from a comparison website, the same comparison website as the original study: facile.it. The dataset was collected by combining different methods:

1. Web scrapping: This process is mainly carried out by using scripts that are

made in Python language to extract large quantities of data from different web pages. The scripts of web scraping programs are structured so they mimic human browsing action, move from one web page to another, and as a result collect the precise data parts. That is why this approach can be useful in instances where the data is not easily accessible through an API or a database. Scientists can capitalize on web scraping to tap into a variety of online resource streams that offer a larger pool of data and consequently enable them to conduct research that is more extensive and robust. Also, attention is imperative to the legal and ethical consequences of web scraping as you must be in line with copyright laws and website terms of use. The incorporation of web scraping in data collection not only enriches the dataset but also significantly expands the analytical capabilities, paving the way for more comprehensive and insightful data-driven studies.

2. Manual Data collection: A different technique that is opposed to the web scraping methods is called manual data collection. This traditional method encompasses direct observation, surveys, interviews, or manually extracting information from sources where automated techniques are not feasible or ethical. Manual collection allows for nuanced understanding, particularly in qualitative research where the richness of data cannot be captured through automated means. It also plays a critical role in verifying the accuracy and validity of data obtained through other methods, acting as a crucial step in maintaining the integrity of the research. However, this method is typically more time-consuming and labor-intensive, which can limit the scale and speed of data gathering. Despite these challenges, manual data collection remains indispensable in various scenarios, providing depth and context that automated methods might overlook.

In the data collection adopted for this study, we combined both methods to ensure minimal disruption to the comparison website's service. Three primary potential sources of variability in the pricing data were identified both in the original study and the new replicated study:

1. Evolution of Actuarial Models and Pricing Schemes: Over time, insurance companies have improved their actuarial models and costing methods, which will surely impact the quotes they offer.

2. Session Duration Effects: The duration of time spent on the comparison website could be one of the parameters that can be utilized and embedded in the pricing model, hence it will influence the final rates provided.

3. A/B Testing by Insurance Companies and the Comparison Website: To determine the most effective pricing strategies, insurance companies or the

comparison website itself might employ A/B testing, leading to variations in the pricing presented to different users.

To eliminate the sources of variation and to ensure the robustness of the collected dataset, we carried out a doubly nested randomization approach with a control group. Developed to limit information overlaps, this method was designed to minimize the effects of these disturbances, enabling a more accurate assessment of the pricing strategies employed by the insurance groups. The doubly nested randomization approach, which is implemented in the original study, is a complex and efficient design intended to respond to possible output interferences in the price information received from a car insurance rate comparison website. Thus, a mechanism was put in place, so data collected is properly reflected as the true pricing patterns and not deliberately or indirectly altered by external factors. Here's a detailed explanation of each component of the doubly nested randomization:

1. Identification of Disturbance Sources: Before initiating data collection, the original study identified the three primary sources of potential disturbance in the insurance pricing signal.

2. First Level of Randomization (Inner Loop): This involved randomizing the order of profiles queried on the website. For example, if you were comparing insurance quotes for different driver profiles (like different ages, genders, etc.), the order in which these profiles were queried would be randomized. This step aims to ensure that each profile combination has an equal chance of being queried at any given time, thereby reducing biases that could arise from fixed-order querying.

3. Second Level of Randomization (Outer Loop): This step added another layer of randomization to the process. It involved randomizing the order of blocks of queries. Each block consisted of a set of profiles that were to be queried together. By randomizing the order of these blocks, the study aimed to minimize the impact of any time-dependent factors (like changes in pricing models or A/B testing strategies that might be implemented during the data collection period).

4. Inclusion of a Control Group: A control group was integrated into the study design. This group likely involved collecting data that would not be affected by the variables being tested (like querying for insurance quotes using a standard, unchanging profile). The purpose of this control group was to have a baseline against which to compare the results from the randomized profiles.

To ensure a reliable data collection of the dataset on which the pricing differentials were due to the available options and not by any uncontrollable factors, we followed the same doubly nested randomization approach defined in the original study by creating a Python script that will create the database that will be adopted while collecting all the necessary information from the website. This database respects the principle of randomization and all possible combinations of all variables.

3.3 Data Sources and Acquisition

For this study, we collected our dataset through a popular Italian website (facile.it) that helps people compare different car insurance prices. The website is a very popular comparison Italian website to obtain different car insurance prices, obtained from the first two results by searching “comparatore RCA” on different search engines especially (google). Choosing this website to conduct this analysis is a pertinent choice since the results offered in it regroup more than 60% of the RCA market. According to the findings of the original study on setting prices for car insurance in Italy, which we used as a base reference, we took our research in the same way. We deliberately took a reasonable number of insurance policy quotes every day - less than 700 - in order not to waste the server resources, for two weeks. The data we collected is very high-quality and gives us a good vision of the various factors that influence car insurance premiums in Italy. This includes multiple prices, which we can use to observe the methods companies apply when calculating the car insurance premium. This way of data gathering guarantees compliance with the reality of car insurance in Italy.

Comparative websites have become the best choice for data collection due to two main aspects. At first, we were interested to see how the gender and place of birth were affecting insurance prices, with the driver as the center. Utilization of the comparison website as one of the popular ways for people to get information about the market, was an effective tool. Then, we attempted to check if preferential treatment exists among different clients as indicated by IVASS, an Italian insurance regulator. Back in 2014, the IVASS investigation revealed that elderly driver profiles are underrepresented on comparison websites and that it is an alarming situation taking into consideration the issue of disproportional treatment.

The role of comparison websites is to coordinate the relations between the customers and the insurance providers, generally, the insurance companies cover the charges while the customers use the services for no fee. In the European insurance market, their influence, especially in Italy, increased greatly. As of 2017, they had become the major insurance providers in the country, wherein they were responsible for 48% of the total motor gross written premiums in Italy. The sizeable market share, which makes them irreplaceable providers of that information, demonstrates their significance in this regard. Moreover, these websites are a valuable tool for drivers to select the right car insurance from a cost and quality perspective while minimizing the confusion over the vast number of online options, thus fostering fair competition between car insurance providers.

3.4 Algorithmic Fairness measures

For the sake of this analysis, we decided to utilize "Fairness through Unawareness" (FTU) as the key principle, similarly to the original study. The concept of fairness in the algorithms known as the FTU may be defined as developing algorithms in such a way that they do not explicitly use or consider to some extent the characteristics, such as race, gender, or age in their decision-making process. The core of this idea is that if these features are not among the algorithm factors, the general tendency would be to treat all parties in a fair and non-discriminatory way.

In the formal definition of FTU, the structure of functions or algorithms is taken, and S represents a sensitive feature, X is the other covariates, and Y represents an output space. The algorithm approves FTU as far as feature S relates to its input if and only if the output of the algorithm is preserved without having either S on or off the input. More precisely, it signifies that the sensitive characteristic is left at no intersection with an algorithm at all. If a sensitive feature is absent from the factors mathematically modeled in the algorithm in the first place or program, then naturally the algorithm will not incorporate any such, therefore meeting the FTU requirements for that feature.

The selection of the FTU principle as the basis for our fairness standard does so because it complies with the regulating standard in most jurisdictions that usually do not allow or forbid the use of sensitive factors in decision-making. Adopting FTU verifies our study's ethical commitment and makes it to be within the legal request. Besides, it must be mentioned that although the idea of FTU (Fairness Through Uniformity) is considered as a chosen type of fairness for this analysis, there might be other notions of fairness, that could be more appropriate in other contexts. The usage of FTU in our study reveals that we are aiming for an impartial and fair representation, which takes into consideration the interests of all and as well produces outcomes that are not only logical but also ethically and legally sound.

3.5 Statistical Methods

A comprehensive approach has been used in analyzing the impact of a protected attribute, which includes birthplace, gender, marital status, educational qualification, and profession. The analysis was segmented into two distinct methods: looking at average prices as a function of features' level as well as price changes for profiles that differ only in one protected characteristic like gender or city/region of birth (Milan-China). These different profiles are termed 'protected partners'.

FTU or Fairness Through Unawareness was the principle used as a guide in the methodology. The results exhibited in profiles contained within the protected pairs should be uniform to ensure the successful maintenance of FTU. We made a special arrangement to set aside the effect of external factors like A/B testing by considering the minor differences as similar ones seen when two queries are compared both of which are the same thus known as 'control pairs'.

Data analysis was conducted using two primary statistical methods: we focused on 'top1' and 'top5' analyses. The 'top1' analysis emphasized the cheapest quote for each profile which is mainly the first choice for those who are focused on obtaining the bottom-line price. However, the 'top3' technique averaged the three lowest offers for every profile, similarly to the drivers searching for what is best in different policy options. Each set of query results was collapsed into a single price for these analyses.

For both methods, the statistical approach involved computing the median value of the price differences between all protected pairs with two given factor levels (e.g., female and male for gender). This median value has, therefore, subsequently been tested with a sign test, whose null hypothesis states that the median changes for each pair profile are equal to zero. This means that, as a logical procession, one would expect as many cases of women getting cost benefits as men would. The acceptance of this alternate hypothesis implies the rejection of the null hypothesis, which means FTU is not held; however, the failure of the rejection does not suggest that FTU is upheld. One condition in which the gap between average prices equals 0 should be assumed, given that it guarantees no case of the overall deprivation of a protected group, but does not ensure 100% of equity.

These statistical methods were employed to answer three primary research questions (RQs) about car insurance pricing and access in the Italian market:

RQ1: What are the factors that play a major role in setting RCA premiums? The purpose of answering this question is to evaluate the influence of different factors affecting driving on car insurance rates. The goal is to give a complete picture of how each criterion affects the final pricing decision. Another portion of the study is about the role of these sensitive features like birthplace and gender in deciding upon pricing as well as regulation against using them directly on pricing models.

RQ2: Do gender and birthplace directly influence quoted premiums? This question would perceive whether sensitive attributes are directly associated with the insurance quotes provided to the users. It included implementing a price comparison scheme between driver profiles with equally sensitive attributes. This approach is constituted by computing the price discrepancies for the "protected pairs" passing through some statistic and using a sign test to make sure there is any arithmetic precision variation from zero. This choice is made to see if there is a possible systemic advantage of disfavor for any given group of the sensitive attributes that they harbor.

RQ3: Do riskier driver profiles see fewer quotes on comparison websites? The third research question is to understand the presentation of riskier driver profiles on the comparison websites. Here, this study's interpretation is focused on how often companies for car insurance are seen in the search results for different driver profiles. The rationale is to find out whether higher-risk profiles being deemed as such by the insurers, imply that these people are likely to be exposed to fewer bids, suggesting discrimination when it comes to the availability of quotes among insurance options for them.

3.6 Ethical considerations

The ethical dynamics underlying the algorithmic audit of the Italian car insurance study are complex and mainly concentrated on the classified data's element, primarily the data associated with gender and place of birth. The parameters were picked because they are permitted for insurance pricing by current regulations, but the RCA portals require inputting these data at the moment to get a quote. However, the study is mindful of the broader ethical landscape, especially concerning European and Italian regulations concerning discrimination.

European legislation, especially the Charter of Fundamental Rights of the European Union, contains the principle of gender equality, in addition to several necessary legal procedures. This principle had been channeled into the insurance business as the Council Directive 2004/113/EC which banned unequal premiums due to gender consideration. So, though it is not allowed to discriminate against a man or woman directly, indirect differentiation on gender lines may still occur justifiably, for example, the divergence of the car plate prices which slightly correlates with gender.

This controversy also relates to the application of nationality-based variables, such as the date of birth, in actuarial models. This occurred in Italian regions until the surge of legal challenges and subsequent investigations woke up the authorities to rethink the tactics. Along with the guidelines of the Italian agencies, the National Anti-Racial Discrimination Office, and the Institute for the Supervision of Insurance (UNAR and IVASS), insurance companies were recommended to remove the birthplace from their pricing models. Such supervisory guidance by IVASS stakes out the nationality as a sensitive factor and excludes it from directly being used as a basis of insurance payment.

Among the sensitive attributes that were added to the study were professional status, marriage, and the level of educational qualification. So additional consideration of the ethical implication of such attributes was considered. This precautionary principle encapsulates the issue of adopting these factors in the calculation of insurance premiums, which might result in the unfairness of bias and discrimination as an outcome. Thus, the study is navigating not an easy ethical path and trying so balance the necessity of researching possible unfair playing practices and the described laws that regulate the use of important attributes, considering underwriting. The study is based upon the principle of legal and ethical provisions by focusing on how European and Italian laws regulate the pricing of insurance according to protected attributes which means adhering to the principles of fairness and non-discrimination.

adjustbox pdflscape booktabs siunitx

Chapter 4

Analysis

4.1 Descriptive Statistics of the dataset

The dataset gathered in the procedure of this study offers a comprehensive descriptive statistical overview of the premiums that various companies offer. The data included six quotes that were from companies that were offering up to four insurance products. The companies that provide the insurance are labeled by c1, c2, c3, c4, c5, and c6, with the corresponding services being differentiated as /a, /b, /c, and /d. These companies include Admiral Europe Compania de seguros S.A. (labeled c1), Compania Assicuratrice Linear S.P.A (c2), Genertel S.P.A (c3), Iptiq EMEA P&C S.A (c4), Quixa Assicurazioni S.P.A (c5), and Allianz Direct S.P.A (c6).

Examination of the compiled insurance quotes demonstrates the divergence of frequencies and number of quotes offered by each insurance company and product set. For example, one of the companies (corporation C9) offered 7680 quotes; all of which have been recorded by the dataset. In contrast, the latter combinations such as C1/a, and C1/b produced just 1831 (24%) and 7660 (99%) quotes (Please refer to table 4.1 for more details). The diversity in the number of quotes and the frequency in which they are mentioned across different companies and product labels suggest a wide variance in the representation of these companies in the selection of those quotes. Such diversity is indispensable in the process of understanding the depth of market behavior and the challenging competitive conditions in the Italian market for car insurance.

The dataset, with a total of 15 price points, further allows performing a comparative study of pricing strategies employed by the said car insurance companies in the Italian market. The outcome of a similar undertaking allows us to point out the trends, patterns, biases, as well as possible disparities in the insurance business seen as different prices that different insurance companies and their products charge.

Company	Num Quotes	Frequency
C9	7680	100%
C1/a	1831	24%
C1/b	7660	99%
C1/c	2792	35%
C2/a	2818	36%
C2/b	1960	24%
C2/c	177	3%
C3/a	1894	25%
C3/b	482	6%
C3/c	1844	24%
C3/d	1419	19%
C4/a	3912	64%
C5/a	5853	76%
C5/b	1825	24%
C6/a	499	6%

Table 4.1: Frequency of quotes of each Company

4.2 Average Impact Analysis

"Average Impact Analysis" is an essential element of understanding the factors responsible for the definition of insurance rates. The analysis is done by determining the average rate of variables that influence the cost of car insurance. We intend to evaluate the contribution of the major factors and define those that are most significant regarding car insurance premiums.

The Average Impact Analysis approach conducts a detailed analysis of each variable encompassing factors such as age, the type of vehicle, location, and other elements of demography that could be sensitive like gender, occupation, marital status, education, and birthplace. The report first estimates changes in insurance premiums due to fluctuations in these factors. As an example, it may assess the spread between the average premium for drivers of different ages or in different kinds of cars.

Through this strategy, the whole valuation scheme in Italian car insurance can be better observed. It brings out the hidden patterns and the fundamental base of what the insurers take into consideration while setting prices for their insurance products. This analysis, however, can also be crucial in identifying potential biases or discriminatory practices. The average impact evaluation, for instance, could point to situations where anomalies in premiums arise as a result of attributes such as gender or birthplace, which could suggest the existence of unfair pricing practices.

This analysis is supplemented by two specific analyses: TOP1 and TOP5. These analyses were crucial for understanding the nuances of insurance pricing from both a cost-focused and a choice-focused consumer perspective:

1. TOP1 Analysis: This analysis concentrates exclusively on the most affordable quote for each particular profile, as it often represents the viewpoint of a person who is primarily concerned with the lowest insurance cost. It is attempting to determine which of the influencing factors specifically led to the establishment of the lowest affordable premium quoted. The "cheapest quote option" in TOP1 Analysis will enable the customers to understand the minimum pricing level among insurance companies, and factors that can affect this baseline.

2. TOP3 Analysis: In contrast, TOP3 enables a comparison of the averages of the three cheapest quotes for every profile. It is like a situation where customers explore multiple policy options before deciding the way forward. Such a method, at the same time, is more comprehensive providing an overview of the level of the market competitiveness and the diversity of pricing strategies applied in the top three options. This observation helps outline for the customers what is the whole range of goods available on the market and what factors influence their choices.

4.2.1 By Age

Focusing on the variable of age, specifically comparing profiles at two different ages: 25 years, and 32 years. To assess the impact of age on insurance premiums, we employed two distinct approaches: An average analysis of the IMPACT for the TOP1 and TOP3 prices. In the case of TOP1, this meant zooming in on the lowest price to compare all the quotes for each age category separately. The price seasonality of the TOP3 analysis has been performed using the average price of three of the cheapest quotes for each profile. According to the graphs, we can see the price variance in the Top1 and Top3 analyses. In the Top 1, the price difference between 32-year profiles and 25-year profiles shows a 30 euros difference which is not a higher amount. In the Top 3, the price difference between the two profiles is the same.

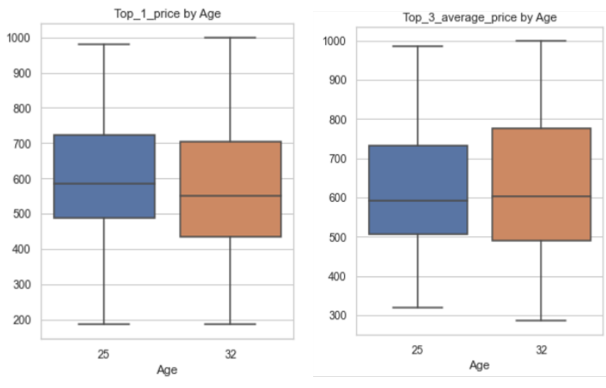

Figure 4.1: Top 1 and TOP 3 average analysis for age

4.2.2 By Birthplace

In this section, we focus on the variable of birthplace, specifically comparing profiles at the different values of this variable: Morocco, China, Milan, Naples, and Rome. In the two graphs, we can see the price variation in both the TOP 1 and TOP 3 analyses. In both cases, we can see that profiles born in Milan are at an advantage compared to other profiles, while profiles born in Morocco are the ones at a disadvantage.

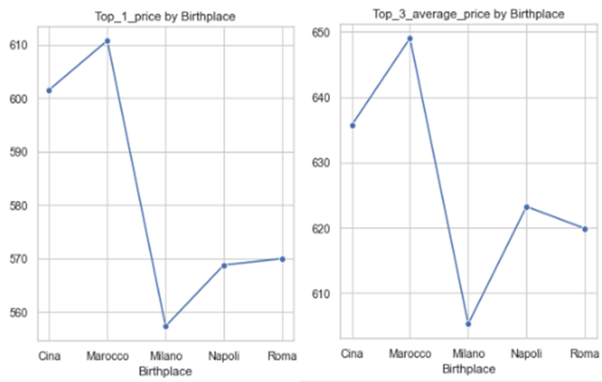

Figure 4.2: Top 1 and TOP 3 average analysis for birthplace

4.2.3 By Gender

In this section, we focus on the variable of gender, specifically comparing profiles at the different values of this variable: Female and Male. In the two graphs, we can see the price variation in both the TOP 1 and TOP 3 analyses. We can see that profiles in the TOP 3 analysis are equal, while in the TOP 1 analysis, the difference is 30 euros.

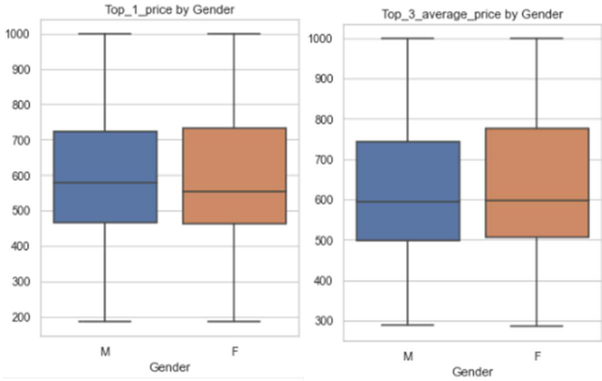

Figure 4.3: Top 1 and TOP 3 average analysis for gender

4.2.4 By City

In this section, we focus on the variable of the city, specifically comparing profiles at the different values of this variable: Milan and Naples. In the two graphs, we can see the price variation in both the TOP 1 and TOP 3 analyses. We can see that profiles in the TOP 3 analysis are equal, while in the TOP 1 analysis, the difference is less than 10 euros.

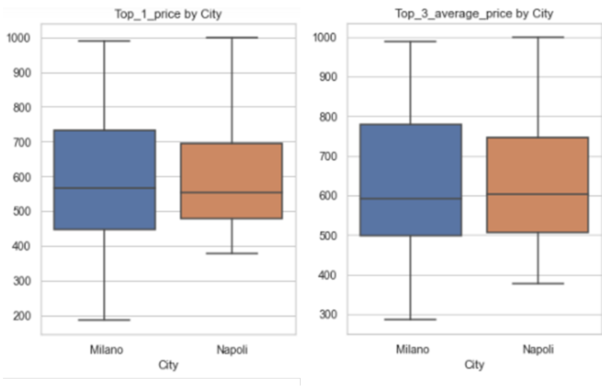

Figure 4.4: Top 1 and TOP 3 average analysis for city

4.2.5 By Car

In this section, we focus on the variable of the city, specifically comparing profiles at the different values of this variable: NSEP and OLED. In the two graphs, we can see the price variation in both the TOP 1 and TOP 3 analyses. We can see in both analyses that profiles with an OLED type of car have higher insurance prices than profiles with NSEP profiles.

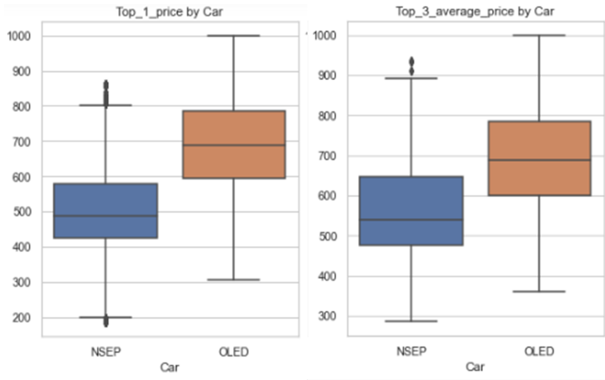

Figure 4.5: Top 1 and TOP 3 average analysis for car

4.2.6 By Km driven

In this section, we focus on the variable of the km-driven, specifically comparing profiles at the different values of this variable: 10000 and 30000. In the two graphs, we can see the price variation in both the TOP 1 and TOP 3 analyses. We can see in both analyses that profiles with 30000 km car have lower insurance prices.

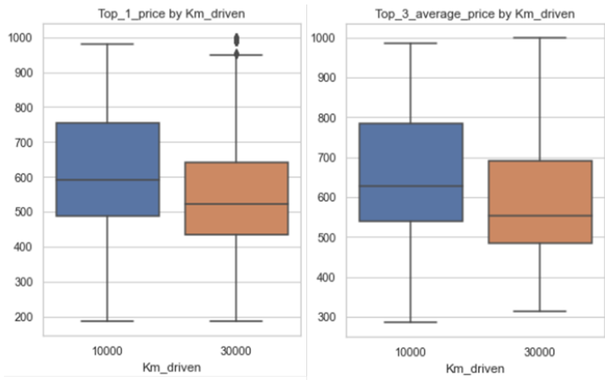

Figure 4.6: Top 1 and TOP 3 average analysis for km driven

4.2.7 By Class

In this section, we focus on the variable class, specifically comparing profiles at the different values of this variable: 1,4,9,18. In the two graphs, we can see the price variation in both the TOP 1 and TOP 3 analyses. We can see in both analyses that class 18 profiles have the highest price and class 1 profiles have the lowest price, which is under the class law definition in the insurance sector.

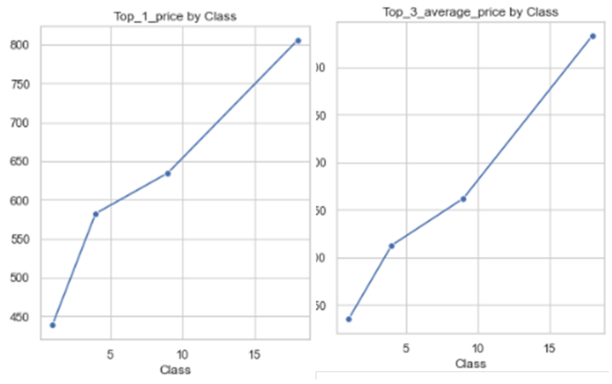

Figure 4.7: Top 1 and TOP 3 average analysis for class

4.2.8 By Profession

In this section, we focus on the variable of the profession, specifically comparing profiles at the different values of this variable: Looking for a Job and Employee. In the two graphs, we can see the price variation in both the TOP 1 and TOP 3 analyses. We can see in both analyses that the price difference is less than 10 euros.

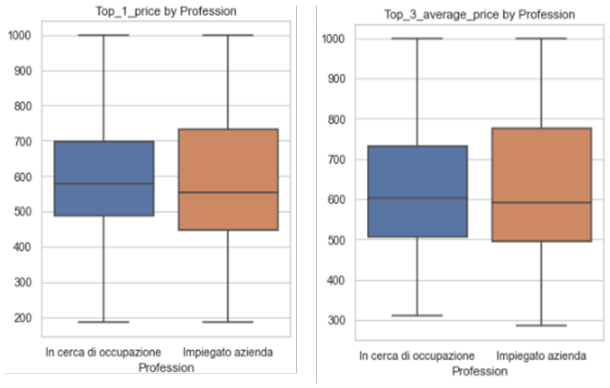

Figure 4.8: Top 1 and TOP 3 average analysis for profession

4.2.9 By Marital status

In this section, we focus on the variable of marital status, specifically comparing profiles at the different values of this variable: married, single, and widow. In the two graphs, we can see the price variation in both the TOP 1 and TOP 3 analyses. We can see in both analyses that married profiles have less than 10 euros higher insurance rates compared to single and widow profiles.

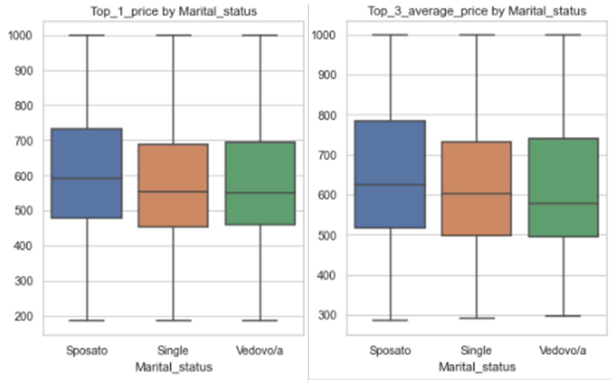

Figure 4.9: Top 1 and TOP 3 average analysis for marital status

4.3 Distribution Analysis of Price Differences

The second analysis focuses on answering the second question: Do sensitive attributes directly influence quoted premiums? This study is very important for revealing the diversity of insurance prices as well as their distribution from one profile to another. Instead of just focusing on average impacts, this analysis delves deeper into how prices vary for similar profiles that differ only in one or two specific attributes.

The examination implies the examination of 'protected pairs,' among which all the profiles are similar except for one feature that is under protection. Such comparison makes it possible to know which feature has the most powerful impact on the pricing.

However, 'control pairs' must immediately be highlighted just like the other parameters if there is a probability of determining average risk and severity. All the pairs that are formed on this setup made of such profiles are similar to the absolute background for the evaluation of the degree of deviation from the prices. Comparing a protected group with a control group is a strategy that can help remove confounding factors and shift the spotlight only on the effect of protected attributes on pricing. This study is centered on several statistical indicators, such as the median and the 90th percentile of this price gap. The median reflects the average influence of protected attributes, while the 95th percentile shows the extent of extreme impacts. The adoption of such strategies can firmly pinpoint and then remove any signs of discrimination and bias in pricing strategies.

By evaluating these measures, the analysis seeks to detect systemic biases against certain groups based on protected attributes. If the price differences for protected pairs significantly diverge from those of control pairs, especially at the 95th percentile, it might indicate embedded bias in the insurance pricing mechanisms. This method provides a comprehensive understanding of fairness and discrimination in car insurance pricing.

According to the results obtained in this analysis, we can conclude that birthplace is used only to the advantage of drivers born in Milan. The p-values associated with this attribute allow us to reject the null hypothesis since they are significant. Analyzing the other sensitive attributes, we can notice that the median is centered on zero. Focusing on median differences, we find no systematic gender bias, marital status bias, profession bias, or education bias. Therefore, we need to analyze the other statistical values obtained. To satisfy FTU, we would expect protected pairs and control pairs to exhibit values with comparable frequency summarized by ties5 and 5th and 95th percentile. Considering the 5th percentile and 95th percentile of these features:

1. Looking at the 5th percentile, values are all negative, showing that the baseline of each feature is rarely at a disadvantage.
2. Looking at the 95th percentile, we can see that values are always larger than 100 euros.

Although the bias expressed in these features is not systematic, there is a direct influence on prices obtained. We can then conclude that all these sensitive features are evidence of the direct influence of prices.

Attribute	Ties5 (%)	5th percentile	95th percentile	Median	p-value
Birthplace Rome vs Milan	1	-123	139	8	6.21×10^{-36}
Birthplace Naples vs Milan	4	-106	162	28	6.74×10^{-140}
Birthplace Morocco vs Milan	4	-50	229	102	3.2×10^{-16}
Birthplace China vs Milan	6	-78	182	99	3.2×10^{-16}
Gender F vs M	24	-151	153	0	6.88×10^{-4}
Profession Employee vs Looking	23	-188	145	0	2.5×10^{-2}
Education Master vs No degree	23	-173	145	0	4.83×10^{-1}
Marital Status Single vs Widow	24	-165	147	0	7.3×10^{-2}
Control Pairs – Noise control	92	-9	6	0	6×10^{-1}

Table 4.2: Top1 distribution analysis of prices differences

Attribute	Ties5 (%)	5th percentile	95th percentile	Median	p-value
Birthplace Rome vs Milan	2	−238	255	8	8.7×10^{-20}
Birthplace Naples vs Milan	3	−252	240	29	3.36×10^{-82}
Birthplace Morocco vs Milan	4	−187	244	102	5.6×10^{-32}
Birthplace China vs Milan	6	−185	189	92	5.8×10^{-29}
Gender F vs M	18	−217	214	0	1.14×10^{-1}
Profession Employee vs Looking	16	−221	242	0	6.3×10^{-1}
Education Master vs No degree	16	−184	269	0	2.0×10^{-2}
Marital Status Single vs Widow	18	−199	212	0	2.5×10^{-2}
Control Pairs – Noise control	92	−7	14	0	7×10^{-1}

Table 4.3: Top3 distribution analysis of prices differences

4.4 Frequency Analysis of Insurer Quotes

Regarding the third analysis, which targets the frequency of quotes placed by different insurers in the dataset as well as the distribution of these quotes, the objective is to determine how often the appearance of different insurers in the dataset takes place. By doing so, this plays the role of establishing or identifying their territorial position and place in the market. The relative occurrence of each insurer's quotes is tallied from amongst the total dataset and this proportion is analyzed to unravel the rank of each insurer and how important they are in the market.

It is performed by the series of additions during which every insurer's quotes are counted separately in the dataset, the task that is involved in counting the number of times when each insurance entity appears with their quotes across different profiles. After that data is brought together to show how many quotes are available for the insurance companies.

The method of tallying and modeling the market share of every insurer helps to establish a precise evaluation of the position of each operator. The comparison of the frequencies would carry out the identification of dominant insurers who are most active in the provision of quotes and those who are less active. Besides looking at numbers and patterns; it delves into the market dynamics of these frequencies as well. For illustration, a fairly high or rapidly increasing number of quotes can be a sign of an influential or belligerent company strategy, while a lower frequency of quotes can testify to a more accurate and selective approach.

We focus this analysis on four companies: C1/a, C2/a, C4/a, and C6/a. These four companies give us a good illustration of how frequency companies vary along the results pages of all profiles selected, by selecting the companies with highest frequency and lowest frequency. Through the graphs presented above, we can note some relevant observations:

1. Company C6 is never present in result pages for profiles living in Napoli, aged 25, with class 1, 4, and 9, and with OLED cars.

2. Company C4 is less frequent for profiles with OLED cars and class 18.

3. Company C2 is never present in the result pages for profiles living in Napoli, born in Morocco, with class 18, and with an NSEP car.

4. Company C1 is never present in result pages for profiles living in Napoli and with class 9.

We can conclude that class, type of car, birthplace, and city are deterministic in defining if a company appears on the result page of the aggregator website.

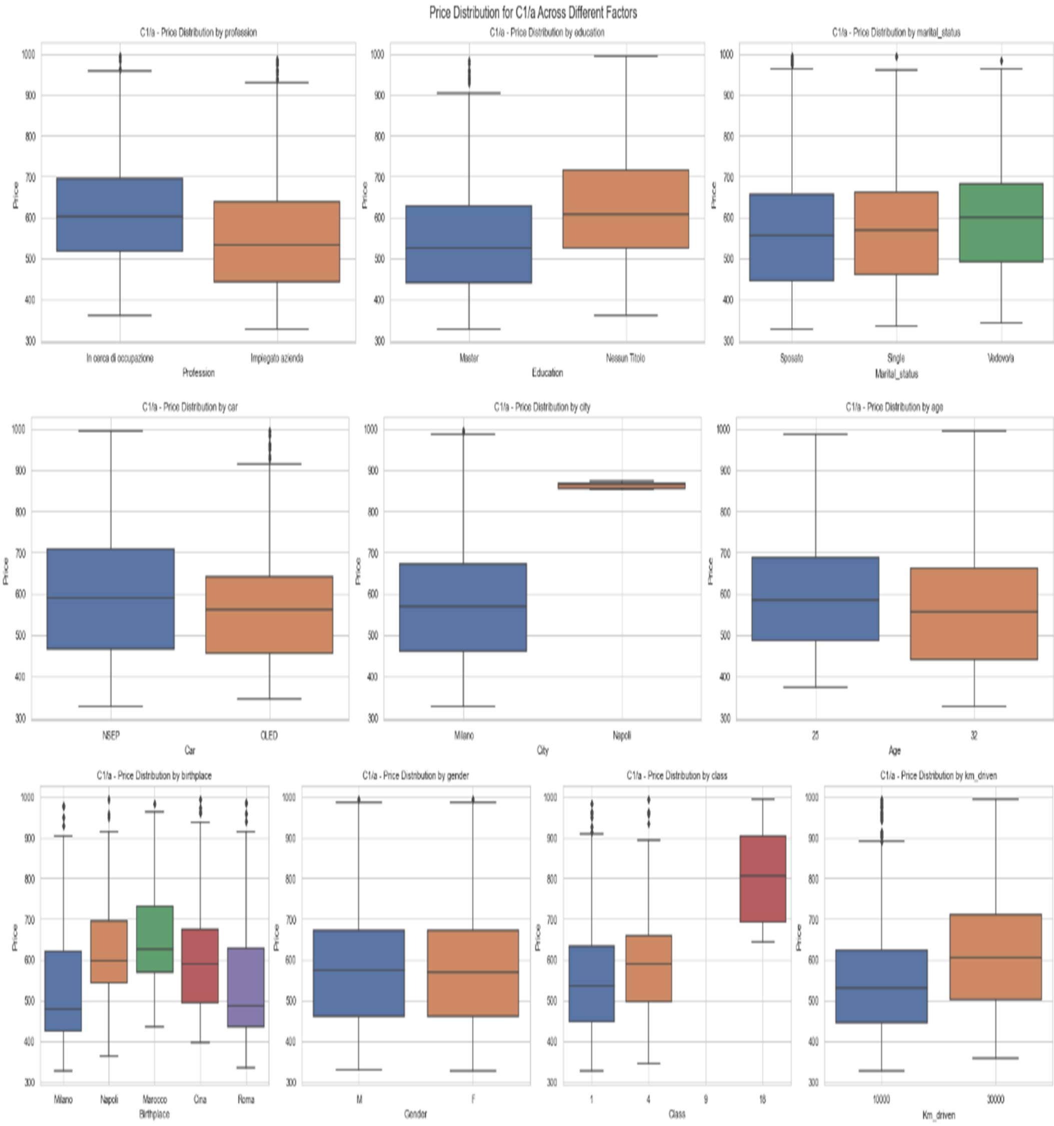

Figure 4.10: Frequency analysis of company C1/a

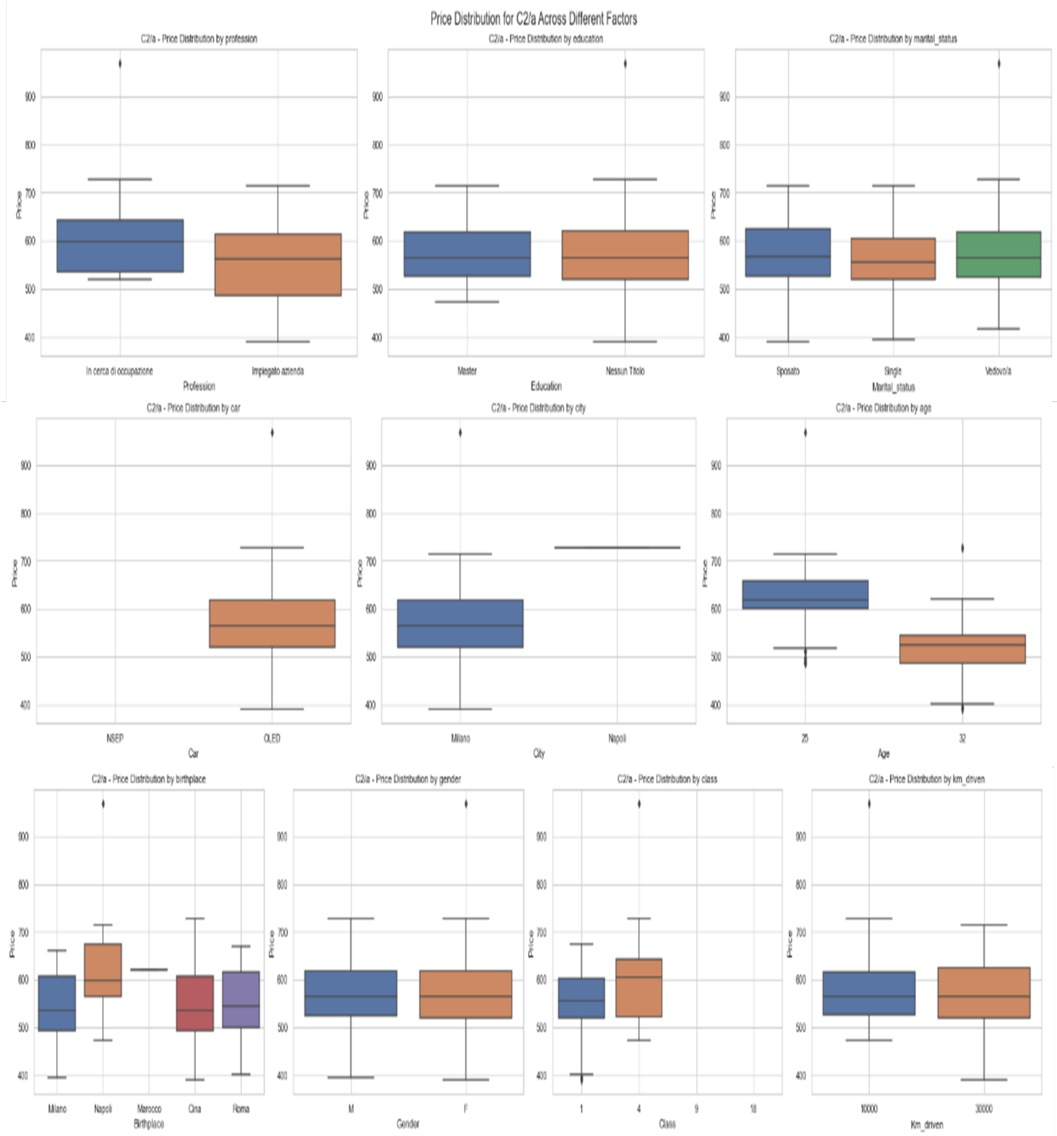

Figure 4.11: Frequency analysis of company C2/a

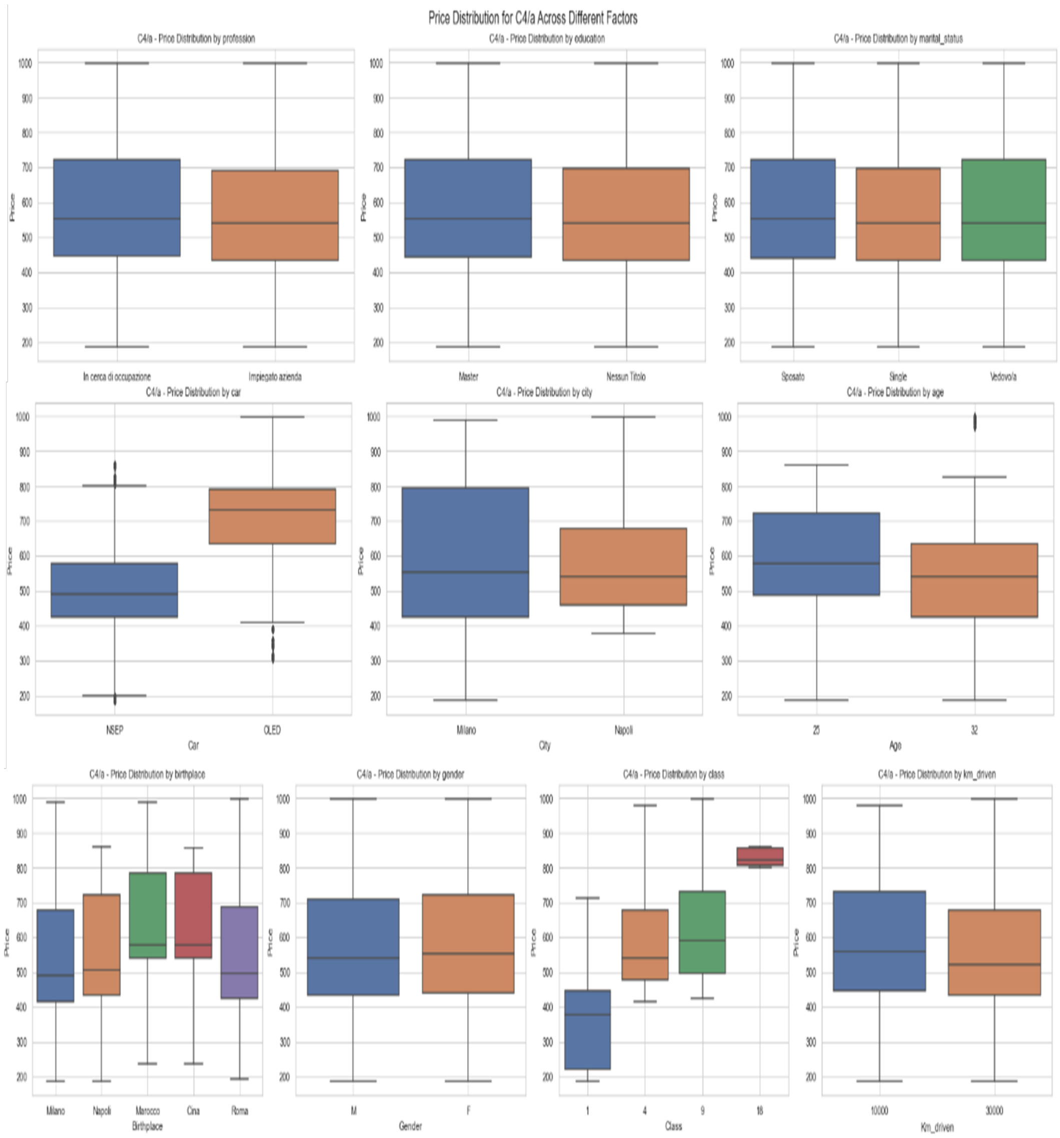

Figure 4.12: Frequency analysis of company C4/a

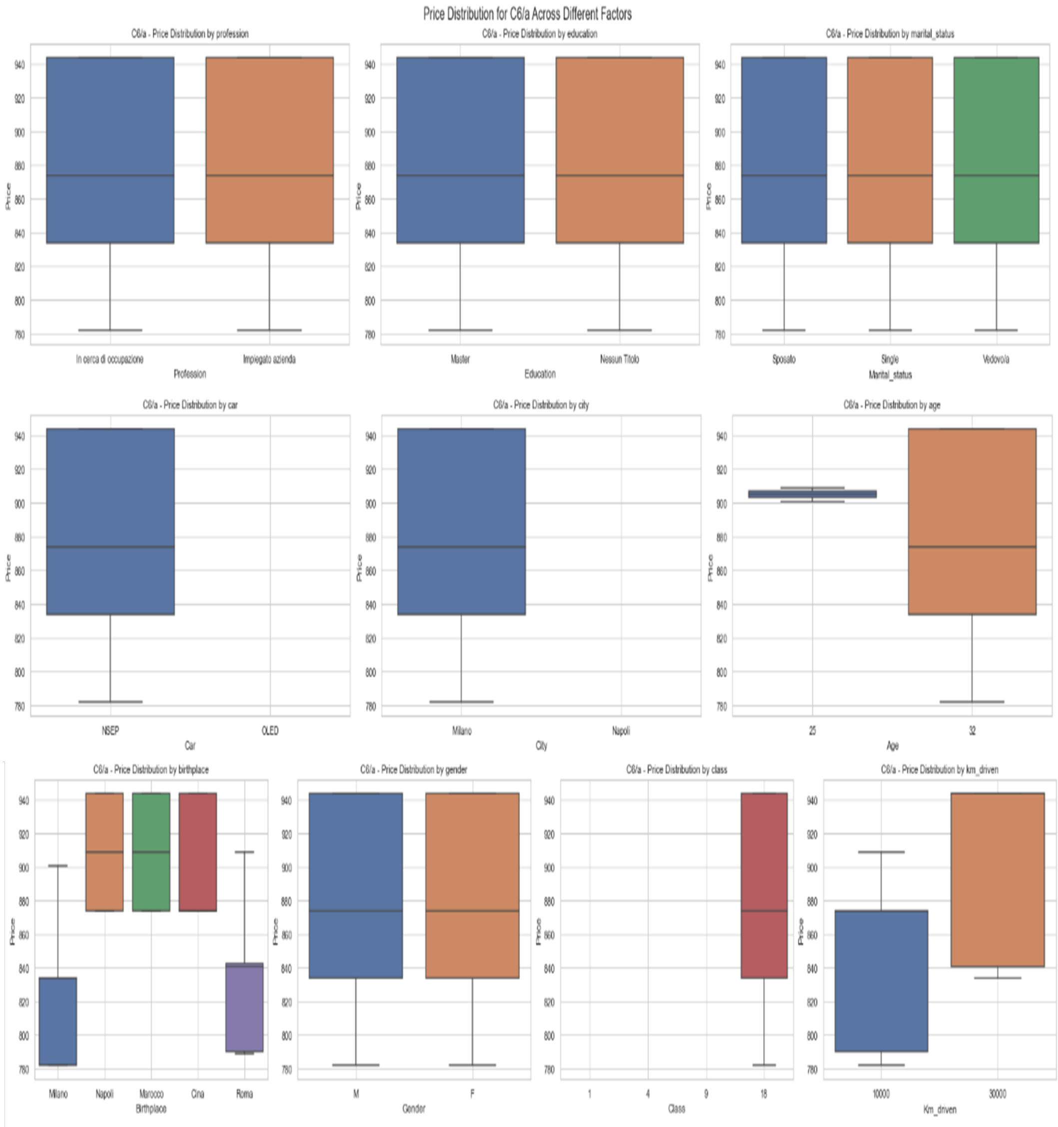

Figure 4.13: Frequency analysis of company C6/a

4.5 Comparative Analysis with Original Study

The current thesis expands on the original study by incorporating a broader range of sensitive attributes and providing a more comprehensive examination of insurer quote frequencies. While both studies employ similar methodologies, the current thesis extends the scope and depth of the analysis, offering a more nuanced understanding of fairness and bias in the Italian car insurance market.

Comparison of Results in the Three Analyses:

1. Average Impact Analysis: This analysis of various factors like gender, city, car type, and km-driven, was conducted in both analyses to explore their impact on insurance prices.

   (a) Original Study: Identified age, city, car type, and Bonus-Malus System (BMS) class as significant factors influencing insurance premiums. Mileage and protected factors (like gender and birthplace) had a lesser impact on insurance premiums.

   (b) Current Study: Our analysis is all about how different variables affect the average car insurance cost. The data revealed that the geographical location, type of vehicle, and history of claims have a significant impact on the rates of obligatory vehicle insurance as tested in our example dataset. We did not discover a significant correlation between the mileage and the pricing probably because the publishers of the sales grid had problems checking this parameter. And in addition, place of birth, and gender were less likely to significantly influence price variation. The added attributes in this new analysis: Some other variables, such as Marital Status, Profession, and Education also seem to have a negligible effect on the price quotes. Due to the sensitivity of those issues and the unlawful character of their usage in this area, the main section was complemented by a more in-depth and critical discussion. In summary, the new dataset got almost the same results as the first analysis except for the age factor that disclosed that those users who are 18 years old are the ones with a high risk in which the bias has been previously confirmed in the first analysis.

2. Distribution Analysis of Price Differences:

   (a) Original Study: This study examines how sensitive attributes (birthplace and gender) affect the price. It has been found that gender does not yield to any disadvantage for every group when comparing median values; however, researched other statistics to surmise that it has a direct impact on cost. Another significant price gap indicated by the result of the study was connected with drivers coming from different areas. In particular, clients from Milan were often quoted lower prices by the insurance company compared with other drivers.

   (b) Current Study: We considered the distribution of the price gap for the protected categories and found that gender, place of birth, marital status, profession, and education directly reflected the variations in quotes. Gender, marital status, profession, and education-related disparities are at zero confirming the findings of the first analysis that these features are not at a disadvantage by default. Interestingly, we discovered significant disparities in the course of this research indicating that the features mentioned above can directly affect the costs. Differences among birthplaces reflect the developed inequality, particularly among the natives of Naples, and drivers born abroad.

3. Frequency Analysis of Insurer Quotes:

    (a) Original Study: Analyzed frequency breakdown of how frequently the different insurance companies appeared in quotes for specific profiles. Noticed the complete absence of the companies' involvement in specific groups of drivers; the reason may be the higher risk connected with the particular driver peers.

    (b) Current Study: We analyzed the numbers of how often are insurance companies featured on the result page of the comparison website based on a variety of customer profiles. Our findings show that there is an existing pattern whereby certain insurers do not appear for specific categories of drivers. First of all, residents of Napoli don't have many invasive quotes, which is also the case for profiles whose birthplace is Morocco, or profiles with a higher class. Consequently, the data from the initial analysis and this dataset tend to support the idea that insurance companies are making biased strategic decisions from the start.

The table below summarizes the different values of the variables used in both analyses:

Variables	**New study**	Original study
Birthplace	Milan, Naples, Rome, Marocco, China	Milan, Rome, Naples, Romania, Ghana, Laos
Gender	Male, Female	Male, Female
Marital Status	Married, Single, Widow	None
Educational qualification	Master, Without a qualification	None
Profession	Employee, looking for a job	None
Age	25,32	18, 25, 32
Class	1,4,9,18	1,4,9,14,18,None
Miles traveled in one year	10000, 30000	10000, 30000
License plate	OLED, NSEP	OLED, NSEP
City	Milan, Naples	Milan, Rome, Naples
Total Number of queries	7680	2160

Table 4.4: Values of variables in both analyses

Chapter 5

Conclusion and Future Work

5.1 Summary of Findings

The comprehensive audit of the Italian car insurance comparison website conducted in this study has traced several essential factors that have a diverse impact on driving the insurance premium level. Most significantly, it demonstrates the huge impact of locational aspects, especially birthplace, on quote prices. This geographic bias manifests in two distinct ways: in the first place, during the insurance quote, foreign-born individuals are charged more compared to the Italians coming originally from the home country, and in the second place, even among natives, the city of birth plays a pivotal part in pricing premiums. For instance, citing the case of people of particular towns finding themselves in a more complex position than those of other areas shows a bit intricate but comprehensible regional discrimination.

In the process of deeper research that has been conducted in this study, another important aspect also surfaces, strategic selection in the offering of insurance quotes based on the perceived risk. Consumers who are deemed to be high-risk customers, because they are non-nationals or have a complicated background history, will have a limited number of insurance options to pick from on the insurance website comparison. The use of a quote that mentions this selective tactic indicates that under the surface we can see some kind of evaluation algorithm. It plays a role in the analysis that seems heavily dependent on the geographical origin of the profile. This way some people can enjoy insurance while others are denied insurance based on this factor. Apart from that, the study indicates that having several quotes offered leads to a higher risk being attributed to this individual. Insurers allocate a wider range of quotes to a profiled entity as seen as less risky, while an insurer can provide fewer options for a perceived high-risk profile. This difference in access to quotes may cause a shortage of competitive pricing that could further lead to more expensive insurance.

Furthermore, the result of the study is evidence for the complexity of the insurance market where among other things the pricing is determined not only by the insured individual risk but also by the interdependence of insurance companies in making decisions, possible existence with regulations, and others. The differential treatment due to geographical factors produces ethical issues of equal treatment and fairness in the insurance industry. This is the discrimination that can be linked to deeper issues at the systemic level, whose nature is well beyond risk assessment, and also touches upon the philosophical principles of insurance as fair treatment and non-discrimination.

Overall, this research provides profound knowledge about the biases in premium pricing in the Italian auto insurance market. This fact that birthplace plays a role in premium rates exposes a discriminatory practice which raises questions like how fair and equal the insurance is. The strategic behavior of insurers in the provision of quote options noted for anticipated risk factors becomes a pointer for reevaluation of current pricing strategies that match ethical requirements of fairness and non-discrimination.

5.2 Interpretation of Results

The findings from the in-depth review of the Italian car insurance comparison website in this study therefore are the basics for understanding the fundamental principles of insurance pricing and the different factors that impact these mechanisms. The research unveils that spatial factors such as place of birth influence insurance premiums and therefore a complex interpretation plotting the reasons and follow-up effects of this is necessary.

Geographical Bias and its Implications: The influence of geographical dynamics, in particular, birthplace, on insurance mark-ups is a pertinent outcome. Such ethnocentric tendency, leading to Italians from certain cities and foreign-born people being charged higher premiums, demonstrates that risk estimation methods in the insurance business suffer from an underlying problem. It sparks the question of a fair environment in the insurance market and evokes doubts about the lawful risk criteria used. The differentiation of how that treatment is done based on location points to the institutional bias which goes beyond physical attributes, but also covers the society's social perceptions and stereotypes of certain regions and groups of people.

Strategic Quote Provision and Market Dynamics: Insurers generally take a selective approach in terms of the quoted coverage, with acceptances being fewer for higher risks. We suppose that insurers most likely analyze confidential risk score data similarly to sophisticated algorithms that classify profiles based upon many criteria such as geographical origin. Cheap insurance companies are not just offering low premiums to certain customer segments but also attempting to create risk profile segmentation there. Insurance product selection is likely to be segmented out. This may cause disruptions in insurance access and differences in consumer profiles which can lead to fewer options and even higher costs in some cases.

Risk Perception and Insurance Practices: The influence, in the number and diversity of quotes resulting from the perceived risk, shows the insurance industry's practices in risk management. One might argue that insurance companies may be inclined to present more variants to those who are safe by nature, as it has a fast increase in demand. In comparison, higher-risk profiles receive limited choices of plans, and this clearly shows caution on the part of insurers who don't want to extend their exposure to people perceived as high-risk. The reason for that is risk aversion which is very clear from the risk management perspective standpoint, however, the indignation about fairness and discrimination is associated mainly with the assessment of risk, for example, if the relative risk is calculated individually different from the place of birth.

Ethical and Equity Considerations: This study's findings provide a basis for ascertaining the ethical and equity issues in the insurance industry as well. The insurance premiums being seemingly geographically biased result in an unfair and unequal distribution of resources beyond the issue of disparity in the insurance industry. This too emphasizes the possibility of unintentional discrimination in algorithmically driven decision-making processes, where certain groups get discriminated based on factors that supposedly don't matter to such algorithms.

Overall, the conclusion of this research shows that there is a very complex insurance market in Italy, whereby the details of a person's geographical origin have a sizable impact on the pricing of an insurance contract. This gives rise to critical issues about the fairness and ethics of current insurance practice and shows that the time is right for re-examining how risks are evaluated and priced. The findings of the study give rise to conversations about justice, transparency, and responsibility in the insurance field, and make people factor in the proportion of risk management to equitable treatment.

5.3 Implications for Policy and Practice

The findings from the audit of the Italian car insurance comparison website carry substantial implications for both policy and industry practice. The issue of geographical biases with the impact on the existence of one particular location being a significant factor when it comes to the price of the insurer's quotes becomes a critical question both for the regulators and the ethical code of conduct.

1. Policy Implications and Regulatory Response:
   (a) Need for Stricter Regulatory Oversight: The study's empirical data on gender-based discrimination call for more rigorous regulatory action. Policymakers have to refer to regenerated policies in which discriminatory (based on geographical factors) practices are implemented and outlawed.
   (b) Revisiting Insurance Pricing Models: The effect of born place on the premiums increases the need for policy reassessment. The establishment of rates of impartial discrimination might be required to maintain a fair situation that does not favor disparity.
   (c) Enhanced Transparency Requirements: As one of the demands of policymakers they can think about introducing regulations in the field of the insurance industry that will oblige them to transparency in calculating the premium amounts. Such an action might comprise identifying obligatory ones for insurers to reveal the dynamics of their pricing.
   (d) Consumer Protection Laws: The data reinforce the point that it is not enough to just put in place consumer protection rules only. This is because some consumers may be unfairly priced based on this criterion for example age.
2. Implications for Insurance Industry Practice:
   (a) Ethical Reassessment of Risk Assessment Models: Insurance companies are those who may reconsider compliance issues and improper risk assessment, especially geographic-based ones.
   (b) Developing Fairer Pricing Strategies: Insurers should design and develop competitive pricing instituted on actual risk rather than perceived risk and determined based on people's features they can moderately influence. It might require the implementation of a more detailed risk assessment method that would not only rely heavily on a broader behavioral spectrum including individual habits and circumstances but also other aspects.
   (c) Emphasis on Corporate Responsibility: The discovery of this is a wake-up call for the insurance business to put more emphasis on the social side of business. Particularly, fairness and non-discrimination should be a priority in all of their services to the people.
   (d) Training and Awareness: This involves capacity building and informing the insurance pros about the entailed situation of discriminating geographical areas and the essence of giving every customer a fair chance.

3. Broader Implications:
   (a) Consumer Trust and Market Dynamics: Such results, due, may affect consumers' belief in the insurance sphere. Insurers who adopt comprehensive measures on this point can be likely to achieve superiority over their competitors and satisfy those consumers who appreciate fair and honest companies.
   (b) Innovation in Risk Assessment: Insurers could be likely to see innovations as the industry moves forward in terms of risk assessment technologies and methods as the industry evolves and seeks ways to assess risks more justly and accurately.
   (c) Global Relevance: The result of the study only concentrates on the Italian market, in which lessons drawn may be extrapolated and used globally to influence the insurance ways and policies of other countries.

5.4 Limitations of the Study

On the one hand, though this study delivers useful information it is essential to note its limitation. The range of the research is within the limits of the Italian car insurance website, the truth is that it cannot reflect all of the Italian insurance market. These results, in turn, may have limited applicability beyond the anonymous website context. The rather uncertain nature of the insurance sector together with changing, complex, and probably short-term reasons for the findings suggests that these might be limited in the timeframe studied. Studies with large scales including all latter and former policies will be needed, and this is just the start of understanding this field in depth.

5.5 Limitations of the Study

Developing future studies to include different carriers and price comparison websites in the scope is necessary to get a complete picture of the industry. Tracking the process over time is enabled by the longitudinal studies and this also reveals areas that need to be regulated significantly and areas with market dynamics that are changing all the time. Furthermore, a comparative study could provide a world view on insurance practices and prejudices involvement all over the world based on the insurance market in different countries. Similarly, investigations into the technicalities of this issue, such as the function of AI and machine learning in the calculation of policy premiums, may lead to other critical insights into current insurance practices that could even be applied to improve technology and ethics.

Bibliography

[1] Bartley, N., Abeliuk, A., Ferrara, E., Lerman, K. (2021, June). Auditing algorithmic bias on twitter. In Proceedings of the 13th ACM Web Science Conference 2021 (pp. 65-73).

[2] Vecchione, B., Levy, K., Barocas, S. (2021, October). Algorithmic auditing and social justice: Lessons from the history of audit studies. In Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (pp. 1-9).

[3] Zhang, S. (2021). Measuring algorithmic bias in job recommender systems: An audit study approach. Technical report. 10.1257/rct. 6101-1.1 [CrossRef][Google Scholar].

[4] Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453.

[5] Fabris, A., Mishler, A., Gottardi, S., Carletti, M., Daicampi, M., Susto, G. A., Silvello, G. (2021, July). Algorithmic audit of italian car insurance: Evidence of unfairness in access and pricing. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (pp. 458-468).

[6] Mustard, D. B. (2003). Reexamining criminal behavior: the importance of omitted variable bias. Review of Economics and Statistics, 85(1), 205-211.

[7] Dobbie, W., Liberman, A., Paravisini, D., Pathania, V. (2021). Measuring bias in consumer lending. The Review of Economic Studies, 88(6), 2799-2832.

[8] Wan, M., Ni, J., Misra, R., McAuley, J. (2020, January). Addressing marketing bias in product recommendations. In Proceedings of the 13th international conference on web search and data mining (pp. 618-626).

[9] Leslie, D. (2020). Understanding bias in facial recognition technologies. arXiv preprint arXiv:2010.07023.

[10] Duan, H. (2019, January). Causes, Consequences and Solutions of Gender Bias. In 2018 International Workshop on Education Reform and Social Sciences (ERSS 2018) (pp. 163-167). Atlantis Press.

[11] Suresh, H., Guttag, J. (2021, October). A framework for understanding sources of harm throughout the machine learning life cycle. In Proceedings of

the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (pp. 1-9).

[12] Kevin A Clarke. «The phantom menace: Omitted variable bias in econometric research». In: Conflict management and peace science 22.4 (2005), pp. 341–352 (cit. on p. 13).

[13] Stephanie K Riegg. «Causal inference and omitted variable bias in financial aid research: Assessing solutions». In: The Review of Higher Education 31.3 (2008), pp. 329–354 (cit. on p. 13).

[14] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. «A survey on bias and fairness in machine learning». In: ACM Computing Surveys (CSUR) 54.6 (2021), pp. 1–35 (cit. on pp. 14–16, 20).

[15] Eszter Hargittai. «Whose space? Differences among users and non-users of social network sites». In: Journal of computer-mediated communication 13.1 (2007), pp. 276–297 (cit. on p. 15).

[16] Jeffrey Dastin. «Amazon scraps secret AI recruiting tool that showed bias against women». In: Ethics of Data and Analytics. Auerbach Publications, 2018, pp. 296–299 (cit. on p. 16).

[17] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. «Machine bias». In: Ethics of Data and Analytics. Auerbach Publications, 2016, pp. 254–264 (cit. on p. 17).

[18] Ziad Obermeyer and Sendhil Mullainathan. «Dissecting racial bias in an algorithm that guides health decisions for 70 million people». In: Proceedings of the conference on fairness, accountability, and transparency. 2019, pp. 89–89 (cit. on p. 17).

[19] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. «Discrimination through optimization: How Facebook's Ad delivery can lead to biased outcomes». In: Proceedings of the ACM on human-computer interaction 3.CSCW (2019), pp. 1–30 (cit. on p. 16).

[20] Warikoo, N., Sinclair, S., Fei, J., Jacoby-Senghor, D. (2016). Examining Racial Bias in Education: A New Approach. Educational Researcher, 45(9), 508-514. https://doi.org/10.3102/0013189X16683408

[21] Thornton A, Lee P. Publication bias in meta-analysis: its causes and consequences. J Clin Epidemiol. 2000 Feb;53(2):207-16. doi: 10.1016/s0895-4356(99)00161-4. PMID: 10729693.

[22] Heijungs, Reinout, et al. "Bias in normalization: causes, consequences, detection and remedies." The International Journal of Life Cycle Assessment 12 (2007): 211-216.

[23] Madaio, Michael, et al. "Assessing the Fairness of AI Systems: AI Practitioners' Processes, Challenges, and Needs for Support." Proceedings of the ACM on Human-Computer Interaction 6.CSCW1 (2022): 1-26.

[24] Zhou, Jianlong, Fang Chen, and Andreas Holzinger. "Towards explainability

for AI fairness." International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers. Cham: Springer International Publishing, 2020.

[25] Mehrabi, Ninareh, et al. "A survey on bias and fairness in machine learning." ACM computing surveys (CSUR) 54.6 (2021): 1-35.

[26] Barocas, Solon, and Andrew D. Selbst. "Big data's disparate impact." California law review (2016): 671-732.

[27] Dwork, Cynthia, and Christina Ilvento. "Fairness under composition." arXiv preprint arXiv:1806.06122 (2018).

[28] Buolamwini, Joy Adowaa. Gender shades: intersectional phenotypic and demographic evaluation of face datasets and gender classifiers. Diss. Massachusetts Institute of Technology, 2017.

[29] Herzog, Lisa. "Algorithmic bias and access to opportunities." The Oxford Handbook of Digital Ethics. Oxford: Oxford Academic, 2021.

[30] Suresh, Harini, and John V. Guttag. "A framework for understanding unintended consequences of machine learning." arXiv preprint arXiv:1901.10002 2.8 (2019).

[31] Ding, Frances, et al. "Retiring adult: New datasets for fair machine learning." Advances in neural information processing systems 34 (2021): 6478-6490.

[32] Zemel, Rich, et al. "Learning fair representations." International conference on machine learning. PMLR, 2013.