A Work Project presented as part of the requirements for the Award of a Master's degree in Business Analytics from the Nova School of Business and Economics

Scraping the Web for evidence of Price Steering

Maria Garcia da Fonseca Veiga

Work project carried out under the supervision of:

Qiwei Han

Fabrizio Esposito

13-12-2022

Acknowledgements: To professor Qiwei Han for all the continuous support and availability throughout the whole thesis and to professor Fabrizio Esposito for the valuable input.

Abstract: This paper exhibits a study about the practice of price personalisation on e-commerce websites. With this goal, eleven websites were analysed from three different industries: Fashion retail, General Retail and Travel. Moreover, to find evidence of this practice, the differentiated features tested were the device, operating system (OS), browser, and geolocation of the visiting user. Furthermore, two dimensions of price personalisation were conducted in-depth: price discrimination and price steering. Finally, summarised research about the existing laws and their implications is also presented in this report.

Keywords: Price Personalisation, Price Discrimination, Price Steering, Data Science, Web Scraping, Data analysis, GDPR

Table of Contents

1. Structure of the paper

With the goal of searching for evidence of price personalisation on e-commerce websites, the analysis and, subsequently, this paper is presented in the following structure:

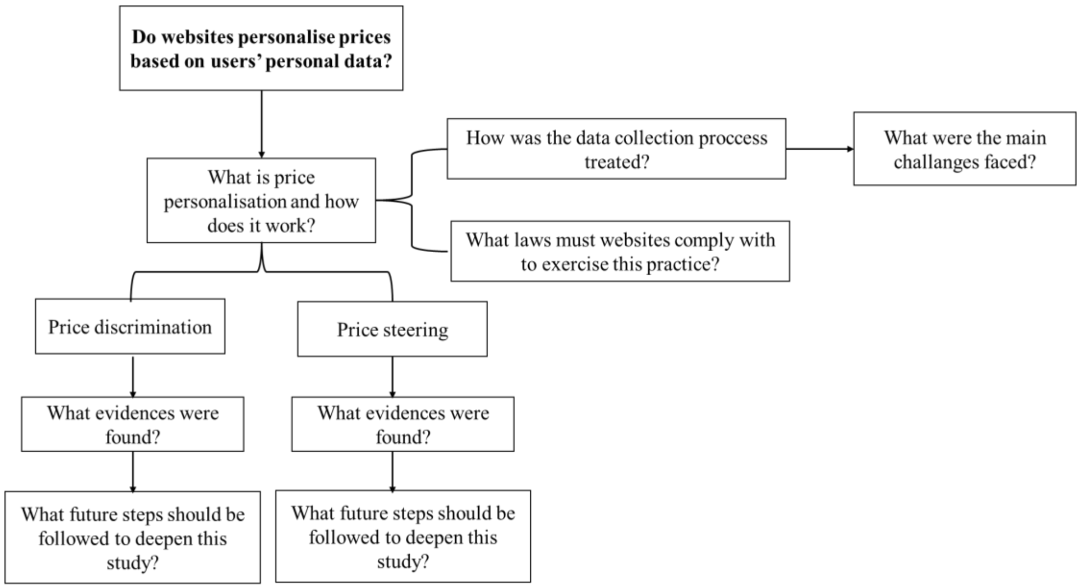

FIGURE 1 - PAPER STRUCTURE

2. Introduction

A good salesperson will say that understanding the customer is one of the most critical steps in a sales strategy. Who they are, what they like, and what drives them. As data has become an increasingly abundant resource, businesses have been developing ways to know their online customers. Tailoring their services to each customer, today's internet is personalised (OECD 2021, 7-9). Personalisation can take many different forms. Spotify, for instance, creates custom weekly playlists with music that matches users' listening habits. Netflix changes the cover art of the entertainment to drive a specific user to a click. The use cases are almost endless.

This paper focuses on a specific type of personalisation, price personalisation, on e-commerce websites. A custom pricing algorithm can be implemented using customer data (e.g. personal information, search history and browser from which they access a retailer's website) together with big data and artificial intelligence tools. The great push for online personalisation indicates a potential for more extensive adoption of these types of algorithms (OECD 2021, 7-9). Currently, the General Data Protection Regulation (GDPR) clearly states when it is legal to process personal data and the necessity for consent from a data subject (Wolford n.d.). As such, the effectiveness of these rules is also a relevant topic for this study.

This research aims to understand how broad of an issue price personalisation is in the current e-commerce European market. Because websites don't publicly disclose their code, custom scripts that scrape the pricing information for each specific website were set up to look for evidence of personalised pricing. In each website, these scripts mimic different user settings (devices or locations) and store which products were returned and the displayed price. Results obtained for different user profiles are then compared.

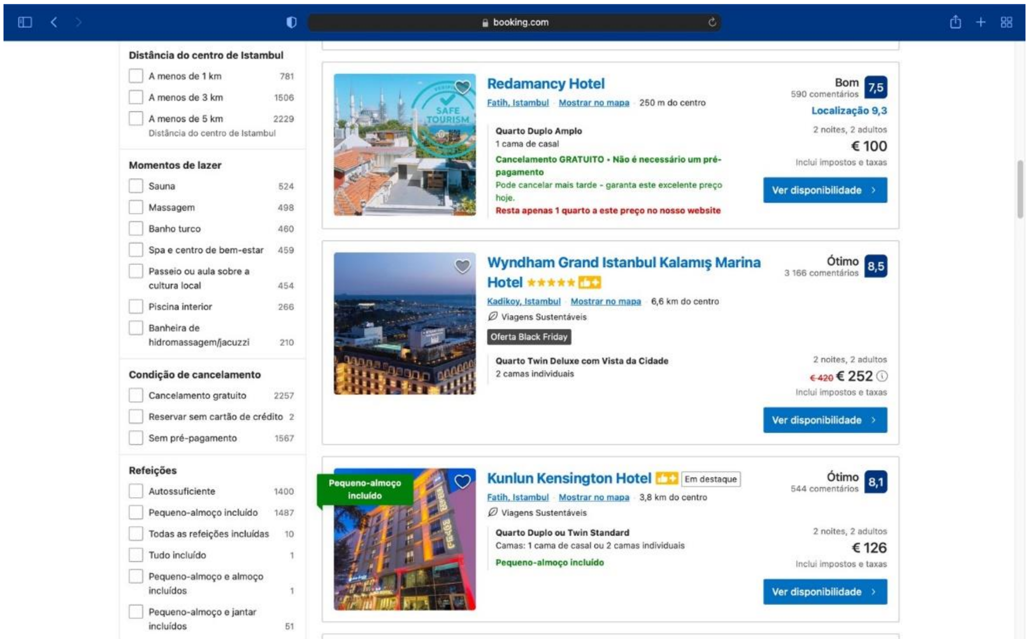

FIGURE 2 - EXAMPLE OF PAGE SCRAPED

This study considers the pricing data for eleven major e-commerce websites from three industries (travelling, apparel, and general retail). Moreover, two different perspectives were investigated – price discrimination and price steering (ranking of offers). Evidence of both was positively identified in some of the explored businesses.

3. Price personalisation

Online personalisation can be defined as the practice of dynamic web experiences that are achieved using data collected from an individual's online activity (European Comission 2018, 30-32). Under the umbrella of which is personalised pricing. A practice where computer algorithms automatically price goods depending on one or more features (UK government 2018) that they were programmed to take as input.

3.1 Pricing algorithms

Dynamic pricing strategies are not a novelty. In the 1990s, Coca-Cola experimented with temperature-sensitive vending machines that would increase products' prices depending on how hot the day was. Online custom pricing can be achieved by implementing pricing algorithms - algorithms that use price as an input (together with other features) and uses a computational procedure to determine the price as an output. These algorithms have been in use for some time with varying complexity (e.g. algorithms that estimate demand or "Match Low Price"). However, there has been a significant increase in data collected by businesses. From 2016 to 2018, over 90% of all data available until that point was created in those two years (UK government 2018). This fact makes the implementation of even more complex pricing algorithms a possible reality. "It is not the algorithmic practice that has significantly changed, although new and more efficient algorithms are invented all the time. It is the data that made the major impact here" (Gal 2017, 3).

As pricing algorithms become more prevalent in the market it is expected that, potentialized by the growing amounts of data, more and more cases of price personalisation will emerge (OECD 2021, 7-9). This paper explores two distinct angles: price discrimination and price steering.

3.2 Price discrimination

Price discrimination can be defined as a customized pricing strategy that charges customers different prices for the same product or service based on what the seller believes the customer will accept (Twin 2022). Online price discrimination is possible since e-commerce websites can leverage the available large size of consumers' data to estimate their willingness to pay (OECD 2021, 7-9).

Discriminatory pricing has been implemented across many industries and can take different forms. Some cinemas offer senior discounts for movie tickets. Spotify has a student plan that cuts the cost of a monthly subscription in half. In these cases, the customer decides freely and out of their conscience to knowingly exchange a limited part of their personal information for different, better pricing. The seniors share their ages with the cinema, and the students share their occupations (and academic information).

3.3 Price steering

Price steering, or personalised ranking of offers, refers to changing the order of search results for the same product based on processing data collected from a customer's online activity.

It is a subject to which online consumers are no strangers. When googling "restaurants", we receive, at the top results, options that are near us. We acknowledge that Google's algorithm uses our personal information (location) to return a list of results that will increase our likelihood of clicking. In many cases, personalised offers are not so "customer friendly". In 2011, The Wall Street Journal found that Orbitz (the American travelling aggregator) steered customers who use apple products towards different and costlier options (Mattioli 2012).

4. Legality behind price personalisation

4.1 Brief notes about the history of price personalisation

The personalisation of prices by sellers has always been present in the economy. The goal of a seller is to extract the maximum willingness to pay from each customer. Therefore, it is in their interest that the price of a good may differ from customer to customer.

A prominent example of price personalisation is in fairs, where different people may buy the same product for a different price. This means that the seller tries to evaluate and takes advantage of the maximum price a buyer is willing to pay. For this, aspects like the image of a person and the way they talk may have an impact.

Finally, as internet use has been increasing at lightning speed, with close to 5 billion users (Pasquali 2022), this concept of price personalisation has also been adopted there. This represents an advantage for sellers with the automated collection of users' data but also presents a challenge regarding its protection and ethical use.

4.2 Welfare economic effects

In general terms, price personalisation benefits the economy as a whole. On the one hand, sellers can extract the whole consumer surplus by charging each customer or group of customers their maximum willingness to pay. On the other hand, it can also benefit customers since a person willing to pay lower than the uniform price may buy the good at a personalised price, lower than the uniform one and equal to their willingness to pay.

Therefore, price personalisation reduces deadweight losses that exist when all customers are charged the same price. However, it cannot be neglected that some customers, the ones willing to pay higher than the uniform price, turn out to pay a higher price, becoming individually harmed (Zuiderveen Borgesius, Poort 2017).

Finally, price personalisation encourages market power. This happens because a monopolist may capture the whole market by being able to charge different prices. This, consequently, makes the entrance unattractive for new competitors.

4.3 Personal data

For websites to be able to personalise prices for a specific user, it has to have access to information about the same, i.e., personal data.

According to the GDPR, Chapter 1, article 4, personal data shall mean “any information relating to an identified or identifiable natural person”, and this data shall include information such as “name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person” (European Union 2016, L119, p. 35).

Firstly, the data must be collected by the website in question or by a third party, then it is processed, and finally, it is ready to be used for price personalisation. With this being said, it can be concluded that online users are constantly being monitored and investigated and, consequently, treated differently (Christl 2017).

4.3.1 Sources of data:

Different sources can collect personal data, such as voluntarily and involuntarily, knowingly and unknowingly, and, finally, by a third party (Zuiderveen Borgesius, Poort 2017).

When a customer creates an account on a website, they know that whenever they log on, the website has access to their past purchases and the personal data they provided when creating the account. In this case, the customer voluntarily provides and knowingly their data.

Alternatively, a customer can provide information involuntarily and unknowingly. This happens, for instance, when a website takes advantage of the IP address, browser, or cookie with a unique identifier of a specific user.

At last, personal data can be obtained by a third party. By the third party, it is meant another website or entity that is not the one in the matter. This can be an advertising network, for example. Using cookies, these entities can create a profile for a specific user, permitting a website to personalise prices for them. Fundamentally, cookies are files which record information such as personal settings of users and interactions with a specific website. Then it can be used by a website to identify a user and provide a customized experience.

4.3.2 Processing personal data

Data protection law ensures that the processing of personal data happens transparently, i.e. with customers' consent. According to the GDPR, personal data must be "processed lawfully, fairly and in a transparent manner about the data subject" (European Union 2016, L119, p. 37). Additionally, in Chapter 2, Article 7, it is reinforced that this process "is based on consent" and "the controller shall be able to demonstrate that the data subject has consented to the processing of his or her personal data" (European Union 2016, L119, p. 39).

Therefore, data protection law grants both rights to users, those from whom the data is collected, and obligations to the websites (controller), those that collect the data.

4.4 People's behaviour facing price personalisation

According to a survey conducted in the US in 2016, 72% of the participants considered that practices such as price discrimination should be prohibited (Poort 2019). Additionally, more than 80% think that it is unfair and unacceptable.

Finally, about 80% of the inquired agree that they must be informed if they are on a website that personalises prices (Poort 2019).

Furthermore, in 2000, Amazon experimented with price discrimination. It turned out that different loyal users realised that once they removed any signs of being regular Amazon customers from their computers, they were given better deals. Consequently, the practice stopped as soon as the complaints began (Streitfeld 2000).

It can be concluded that people generally do not feel comfortable or enthusiastic about buying on a website that states the practice of price personalisation.

4.5 Data protection and transparency

Article 13 of the GDPR presents a list of the information the website must provide. This includes, for instance, "the purposes of the processing for which the personal data are intended and the period for which the personal data will be stored" (European Union 2016, L119, p. 43). When a website practices price personalisation, it must inform each customer explicitly.

Moreover, to comply with Data Protection law, it is not enough for websites to have statements like "we use personal data to offer our customers better-personalized services". This would benefit the website since the user would not have complete knowledge and may end up staying on the website besides the practice of price personalisation.

Overall, transparency could mitigate the information asymmetry that exists. Customers could choose a website that does not personalise prices or, if it benefits them, could delete cookies (Zuiderveen Borgesius, Poort 2017).

4.6 Unfair Contract Terms Directive

The Unfair Contract Terms Directive "protects consumers against unfair terms in all types of business-to-consumer contracts" (European Union 2019, C232, p.7). This shall be applied to all contracts for purchasing goods or services, including financial services and e-commerce or offline commerce. In addition, it states the conditions that make an agreement not fair.

Regarding this topic, Article 4(2) of the Directive mentioned above must be considered. This Article states, in simple words, that if a contract operates with transparency, it complies with the requirements to not be regarded as unfair. In other words, for a contract to not fall into unfair conditions, it must present the price practices in a "plain intelligible language" (European Union 2019, C232, p. 21).

4.7 Portuguese Law

The framework in which this topic falls in Portuguese law is known as *Regime das Cláusulas Contratuais Gerais*. This, combined with the *Lei da Defesa do Consumidor*, define and governs the fairness of contracts and the transparency in the seller-buyer relationship (PGDL n.d.).

4.8 Overall notes

To be by the existing law, websites must inform their customers if they collect and use their personal data. Moreover, they must explicitly notify the purpose of it. This means that a website that personalises prices must unequivocally inform the buyer that it is presenting prices that were explicitly personalised for that user based on their data.

All these legal requirements may not be attractive for sellers since customers are uncomfortable and do not find this practice acceptable. They even may end up purchasing on another website because they do not think this practice is fair and feel disadvantaged.

5. Methodology for data collection

As mentioned before, this paper aims to understand how standard pricing personalisation practices in the European market are. To do so, it is necessary to analyse the results that users with different settings receive, on an e-commerce website, when searching for the same product. Several web scraping techniques were relied upon to get to this quantitative data. Web scraping is automatically extracting data available on the internet (European Commission 2018). Most commonly, web scrapers download the code that composes the visited page which in turn, contains the embedded data. Many services rely on web scrapers. Sporting results pages, stock evaluation services, and price comparisons between competitors are examples, and web scraping can be used in several more use cases.

In this paper, for each product search in each website scraped, different iterations of the same query were run, changing the settings of the user who made the request. Settings that mimic other user profiles that were previously set up. Before getting into the approaches taken and how these different settings were implemented, exploring the design choices taken in this research is crucial.

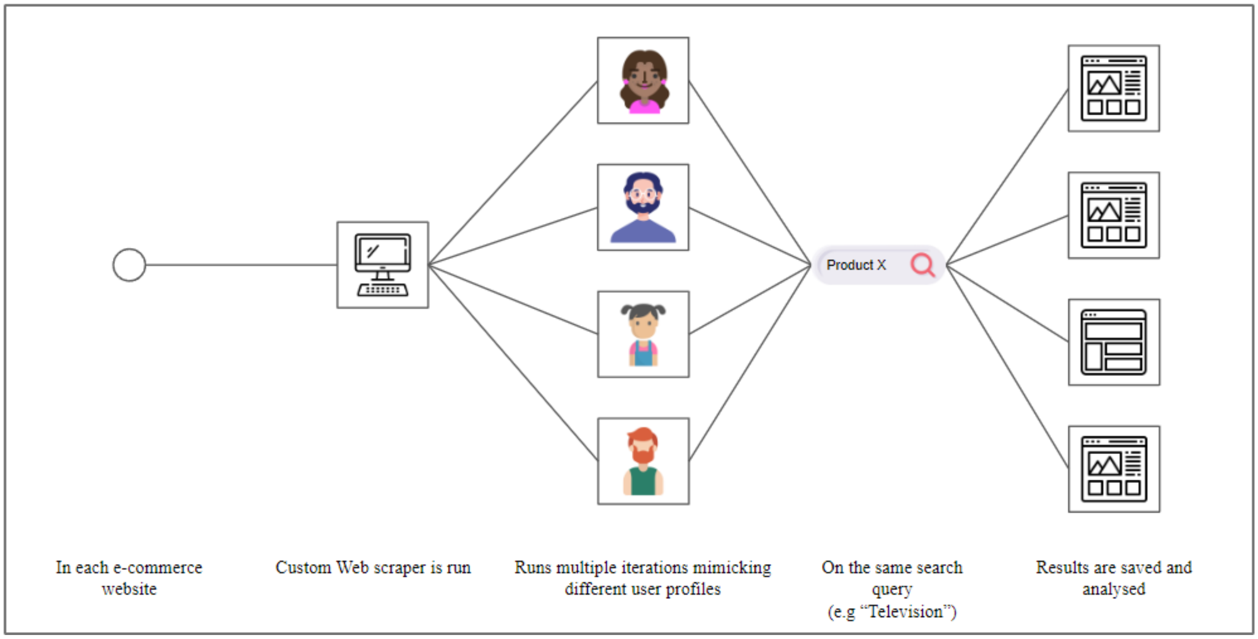

FIGURE 3 - OVERVIEW OF THE DATA COLLECTION PROCESS

5.1 E-commerce Websites

As objects for this study, eleven e-commerce websites were chosen. The criteria for the selection of these websites differ. Firstly, general retailers and fashion retailers specifically. These two industries were identified as highly detrimental to the customer (European Commission 2017). Also, the travelling industry has been found to have performed price personalisation in other related studies (Hannak, Mislove, Wilson, Lazer, and Soeller 2014). Then, the businesses selected must be of considerable size in the European/Portuguese market. Larger companies have access to more data and the infrastructure to implement complex algorithms. These are the e-commerce websites that are going to be investigated when looking for price personalisation.

			Tested characteristics			
Industry	**Website name**	**Website URL**	**Device**	**OS**	**Browser**	**Location**
Fashion Retail	Nike	https://www.nike.com/pt/	X	X	X	X
	Adidas	https://www.adidas.pt/	X	X	X	X
	New Balance	https://www.newbalance.pt/	X	X	X	X
	Farfetch	https://www.farfetch.com/pt	X	X	X	X
General Retail	Worten	https://www.worten.pt/	X	X	X	X
	Decathlon	https://www.decathlon.pt/	X	X	X	X
	Pingo Doce	https://mercadao.pt/store/pingo-doce	X	X	X	
Travel	Booking	https://www.booking.com/	X	X	X	
	Expedia	https://www.expedia.com	X	X	X	X
	Tripadvisor	https://www.tripadvisor.com/	X	X	X	
	Airbnb	https://www.airbnb.com	X	X	X	

TABLE 1 - WEBSITES INVESTIGATED AND USER CHARACTERISTICS TESTED

5.2 User profiles

The selection of the user settings that this paper aims to test is a crucial step in this process. These features will be what the existing (or not) personalisation algorithm receives as input. If no evidence for personalisation were to be found, it only means that no pricing algorithm is based on the selected features. Personalisation could still exist but be based on characteristics not tested in this paper.

Similar research on the effects of online price personalisation has taken place, and although the “evidence of the prevalence of online personalised pricing remains sparse” (OECD 2021, 7) when pricing variations exist, they are mostly found when changing location or browser (European Comission 2018, 40-46). Based on the results from those previous studies, the selected user settings tested in this paper are device, operating system (OS), browser, and geolocation.

5.2.1 Device, OS, and browser profiles

In the context of this paper, it is essential to differentiate mobile from non-mobile devices. Mobile devices are portable computing devices (i.e. smartphones, tablets) and non-mobile devices (or desktop devices) are customer-owned personal computers (i.e. laptops, desktop computers). Moreover, an Operating system is a software that acts as an intermediator between the user and the device hardware, for example, Windows (University of Wollongong n.d). Finally, the browser is a piece of software that "retrieves information from other parts of the web and displays it on your desktop or mobile device" (Mozilla 2022). It connects the user to the internet, for example, Firefox.

The choice for the device-browser combinations is based on what are the most used in today's market. Using standard devices, operating systems and browser profiles increases the chances that an e-commerce website has tailored a pricing algorithm based on those (or part of those) specific characteristics. Nowadays, online consumers use mobile devices as much as desktop machines when visiting online e-commerce websites (Smith 2022). However, the conversion rate for mobile visitors is almost half the rate of desktop users (Jehanne 2022). As such, online businesses could see custom pricing models targeted at mobile customers as an opportunity to increase their likelihood of buying.

Seven different device-OS-browser combinations were selected, which cover both mobile and non-mobile devices. For desktop devices, 90% of the non-mobile devices used today are running either Windows or macOS. Android and IOS cover more than 98% of the operating systems market for mobile devices. The browsers chosen represent over 95% of users (gs.statcounter 2022). Besides the individual market share, the combinations' commonness was considered. For example, Firefox on mobile devices has a market share of less than 1% in 2022. As such, it is not tested in this research.

	Operating System	**Browser**
Mobile devices	Android	Google Chrome
	IOS	Safari
		Google Chrome
Desktop devices	macOS	Safari
	Windows	Google Chrome
		Microsoft Edge
		Firefox

TABLE 2 - DEVICE, OS AND BROWSER USER PROFILES TESTED

The selected profiles allow for the exploration of possible personalisation focused on specific characteristics like the browser used or on a more general approach, for example, comparing mobile to non-mobile devices.

5.2.2 Geolocation

The user's geolocation while navigating e-commerce sites has been found to be a factor that some businesses use for price personalisation (OECD 2021, 7-9). As such, it is also a user setting tested on this paper. Since this research focuses on the European market, four geolocations and European countries have been selected. These are **Portugal, Spain, the Netherlands, and the United Kingdom** region. The selected regions cover a sparse area of Europe and allow for the exploration of price personalisation practices in regions with varying purchasing power.

5.3 Web scraping

Understanding how web scraping works is essential to acknowledge the approaches taken to answer the proposed challenge. Web scrapers are versatile and can take many forms, but their ultimate function is clear: To extract content from internet pages. Before comprehending web scrapers, another step back must be taken. Because before pulling the content from a web page, one must understand how that content is retrieved in the first place.

The web has immense amounts of websites and information. Websites are hosted on servers containing the data and information seen when visiting an internet location. When accessing a website, on the back end, the client (in this case, the scraper/ browser) is requesting that the host (website server) returns, as a response, the resources needed to build the content up. The request-response dynamic occurs under the Transmission Control Protocol (TCP), which permits the exchange of resources between the two. HTTP – Hypertext Transfer Protocol – is how the server and the client communicate with each other when these transactions of information occur (sematext n.d). The requests are known as HTTP requests. After an HTTP request, the server will return an HTTP response. If the request is successfully validated, the response will return, as a result, the resources requested by the client (IBM 2021). The resources are then translated by the browser to a visual form and presented to the user as we are familiarised. When web scraping, these resources aren’t usually transformed into visual assets. The focus is mainly on the data embedded into the page’s response.

5.3.1 HTTP requests

HTTP requests play an essential role in this paper. They allow for defining different user characteristics for the user profiles described previously. As such, it is crucial to understand how that can be achieved and, consequently, what composes an HTTP Request. An HTTP request comprises three elements: A request line, HTTP headers, and a message body (if needed).

The request line is composed of three other elements – an HTTP method, URL, and the HTTP version number being used. The HTTP method indicates to the server the action to be taken. For instance, the GET method is used to retrieve data, and DELETE is used to delete a particular resource on the server. The URL serves as a path to the resources the client wishes to access.

The header provides the server with information about the message and the client itself. It permits adding information about the client such as cookies, and user agent string. The user agent string contains details about the client, allowing the server to identify **the device, operating system, and browser when making a request**. Lastly, the server uses the message body to return the response back to the client. In cases where the client wants to use methods that add to the server's resources, also described in the message body.

This paper uses the header to mimic different user profiles when scraping e-commerce websites. The illusion that a user with a specific, pre-chosen characteristic is visiting the website is enabled by the customisation of the header before making the request. This works for both different geolocations and different device-OS-browser combinations.

Different geolocations are accomplished using IP addresses from other countries. IP address stands for Internet Protocol Address. It is associated with a specific computer network that allows for transferring information by connecting the client to the server when connected to the internet. When making a request, the server receives the client's IP address so both devices can communicate. The server can then receive the request and send back the appropriate response. Once the IP is shared, the e-commerce website can trace the user's location in real-time.

The understanding of whether price personalisation exists will be made based on whether the response from the website is different for requests that only differ in the user-agent string or IP address passed.

5.3.2 Creating the user profiles

To mimic the different device-OS-browser combinations, different header settings are used. For each combination selected, a profile was created. In each, three user agent elements are changed.

- Sec-ch-ua-mobile: A flag which is either “1” if the device is mobile or “0” if non-mobile.
- Sec-ch-ua-platform: Represents the OS. It can be, for example, “Windows” or “macOS”.
- User-agent: Represents the Device/Os and browser used.

To simulate different geolocations, a service called smartproxy was used. Smartproxy provides IP addresses from other locations.

5.4 Web scraping for price personalisation

It is necessary to extract the pricing data for different products to understand if personalised pricing practices are in motion on each website. The data necessary is a product identifier and product price for different user profiles on the same search query.

There are many options for web scrapers. Pre-built scrapers or browser extensions are available online that help the average user extract online content. For this paper, specialised tools were needed. The websites where personalised pricing practices were investigated belonged to different industries and were built using diverse techniques. Also, there was a need for scrapers that allow for the passing of custom HTTP request headers. With these requisites in mind, custom web scrapers were created for each website.

For this search, the scrapers were developed in Python. Python is a high-level programming language widely used for data science, machine learning and data analysis. It has grown in popularity over the last few years and is now the second-most-used programming language in the world. Many Python libraries have been made available by the community. These libraries, collections of existing code, can be leveraged when implementing new web scraping solutions.

Because the chosen websites are built in different ways, no solution fits all. Therefore, three different scraping techniques were used.

5.4.1 Direct request methods

Using the *requests* Python library, it is possible to directly request the website for the resources available on a page using a specific URL. Because the URLs often have information on the query made by the user (for example, “*https://www.worten.pt/search?query=iphone*”), this approach allows for the retrieval of information using different URLs for different products. Such a method works for websites where the data is embedded directly into the page’s HTML code – static websites. HTML – Hypertext Markup Language – is a markup language that tells the browser how to logically translate the HTTP responses' content into being displayed in a web browser.

The output of scraping a website using this method is the HTML code for the page with the specified URL.

5.4.2 Direct request to the API

Not all websites, especially in 2022, have the data directly embedded into the HTML code. Most complex websites have functions that perform different tasks on the website itself. These dynamic websites have, commonly, JavaScript code. JavaScript is a programming language used by most websites to handle a page’s behaviour on the client’s side (for example, on what clicking a button does). Dynamic websites cannot be scraped using direct requests. The returned HTML code does not contain data but the JavaScript code on the page. This happens because the page hasn’t had time to load its content.

On dynamic websites, the approach followed was to take advantage of the page’s API use. APIs - Application Programming Interface - allow two software components to communicate with each other using a specific set of settings. Moreover, they perform tasks between the client and the server. The APIs leveraged were the ones that are responsible for the fetching of data. A dynamic e-commerce website where a user searches for “iPhone” will call for the API responsible for retrieving the information on “iPhone”, and the API shall return the data that is then displayed.

The network activity (where the Requests activity is logged in a browser) was analysed to find the API responsible for returning the data. Once the appropriate API was found, using software called *Insomnia* allowed for a better understanding of the API and how to manipulate it. For example, to return data equivalent to two pages of results, not just one. *Insomnia* also facilitates the translation of the API request into Python code.

Using this translation, for most dynamic websites, the approach was to make requests directly to the API. Dodging the need for loading the page and going directly to the software responsible for fetching the data. The output of this approach was the HTML code for some pages with the data embedded in them and, for others, the data in an already structured format ready for analysis.

5.4.3 Puppeteering a browser

On some dynamic websites making the request directly to the API does not work. Either it can’t be found, or extra layers of security exist, such as not allowing similar requests to be made from the same client side. Or, in the travelling industry, it is common for websites to redirect the user to another page with the requested information.

In these cases, the *Playwright* library was used. *Playwright*, initially designed for testing feature testing, allows for the launching of browser instances that will follow specific instructions on what steps to take next using HTML selectors. In a website built using HTML, each element will have a unique identifier - element selectors. Examples of instructions given are clicking a specific button, scrolling a particular number of pixels, or filling an input box.

To retrieve pricing data, the browsers were instructed to search for a specific product and then extract the HTML content present in the newly loaded page, which is the final output for this approach.

The service used for achieving different geolocations was not compatible with *Playwright*. As such, Booking.com, Adidas, TripAdvisor and Airbnb were not tested on whether user location impacts price personalisation.

5.4.4 Extracting the content from the HTML code

After scraping the HTML code, its content is extracted using the *beautifulsoup4* library. It facilitates the navigation of HTML code and easier retrieval of the desired data. For this paper, the product identifier and product name were required. This data was extracted using *beautifulsoup4* and the HTML selector for where this data is embedded on the pages HTML code.

5.5 General methodology

For all websites, the general methodology is the same. A page is to be scrapped on a website with no purchase history, no user logged in and complete acceptance of cookies permission requests. All iterations for the same website and product were done in the same conditions, apart from the variable that is to be tested.

A minimum of four product queries searched was established per website. The pricing data extracted is restricted to the first and second page of the results. When scraping a website, both variables are used at different times. The website is first scraped using the different device-OS-browser profiles and then the different geolocations, keeping the other variable fixed throughout the process. When the content is scraped, in case it is HTML code, it is parsed to extract a product identifier and the lowest shown product price.

For each website and product searched, iterations of the same request are made with the different profiles. The result of each iteration is then saved on a CSV file for future analysis.

6. Price steering

Price steering is the manipulation of the search results page on a website, after a query made by a user. This means that two people, on the same website searching for the same product, may end up on pages with different products or with the same products in a different order.

Essentially, when a user searches for a product on a website, the search result may differ based on factors such as the device and browser used or the location. Moreover, this concept of price steering is about the products that appear on a page and their order and not differences in prices.

This analysis was conducted to understand what kind of differences exist when searching for a specific product on a specific website with different devices, operating system, browsers, and geolocations.

Given that in some cases, the mobile user agents, during the web scraping process, collected a lower sample, it was necessary to reduce the others' sample sizes. Therefore, by matching the number of observations collected by each user agent, the results present a greater accuracy. For this reason, Adidas was excluded from this analysis, since there were user agents that only collected two products.

Finally, with the goal of measuring price steering, three metrics were used: Jaccard Index, Kendall's tau and Normalized Discounted Cumulative Gain (nDCG).

7.1 Jaccard Index

7.1.1 Clarification on the metric.

The Jaccard Index measures the size of the intersection over the size of the union, between two sets. In simple words, it is the ratio between the common elements between two sets over the sum of the elements of both (Hannak, Mislove, Wilson, Lazer, and Soeller. 2014).

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A|+|B| - |A \cap B|}$$

FIGURE 4 - JACCARD INDEX FORMULA

In this case, the Jaccard Index was used to compare and find differences between pairs of two search results pages, with different user agents or proxy locations, regardless of the products' order. In uncomplicated words, this index answers the following question: Are the products that appear on a search results page the same, regardless of their order, across different user agents or proxy locations?

For instance, to exemplify, consider that a person is using their windows computer and goes to Google Chrome and searches Booking.com. On this website, searches for a house in Barcelona for the next weekend and receive a page with all the offers. The point is: Would this person receive the same houses (same products) if the search had been done on a Mac using Safari? Or, if this person was in Barcelona, instead of being in Portugal?

7.1.2 Methodology followed.

Starting from the CSV files with the products' names and prices, and after transforming these files into dataframes and dropping the price column, all dataframes concerning the same website, were merged by its indexes. This means having a dataframe with products' names of all products that appear on the results' page for each user agent / proxy.

Then, these columns with the products' names were transformed into lists, to make it possible to do comparisons between sets. Moreover, since the goal was to compute the Jaccard Index for each pair of two user agents / proxy, it was created all possible combinations between them.

Finally, the Jaccard Index was computed for each combination of two sets. It was also computed the mean Jaccard Index score for each website.

7.1.3. Results and conclusions.

On one hand, if the Jaccard index is equal to 1, it means that the 2 sets have the same products. However, nothing about the order of the products on the search results page can be concluded. The products can be in the exact same order or in a completely different order. On the other hand, if the index is equal to 0, it implies that the results do not have any product in common.

a) About user agents:

According to the Jaccard Index, there were found differences on the websites of Farfetch, Expedia, TripAdvisor, Airbnb and Booking, since their Jaccard Index is, on average, lower than one.

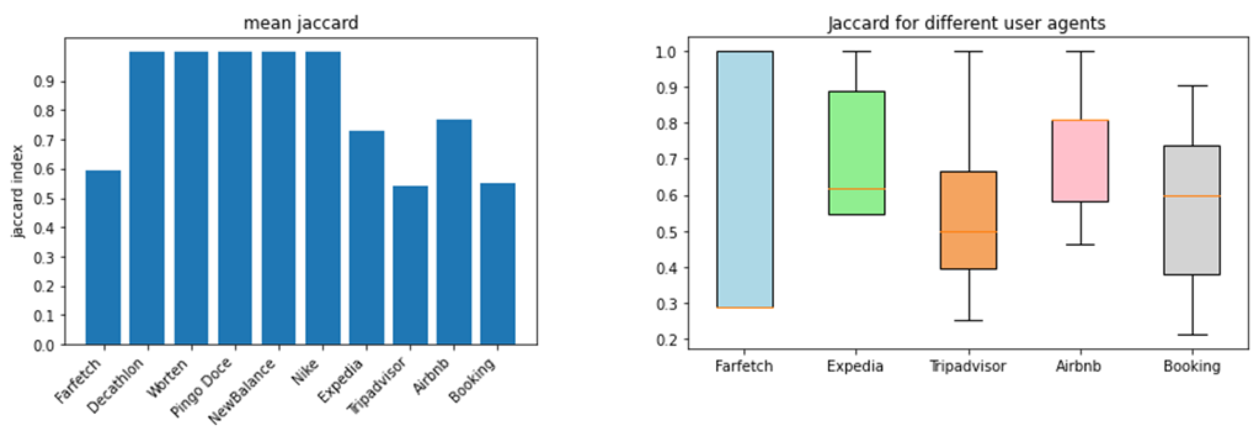

FIGURE 5 - MEAN JACCARD SCORE OF DIFFERENT USER AGENTS FOR EACH WEBSITE AND VARIANCES OF THE JACCARD INDEX FOR EACH WEBSITE THAT WAS FOUND DIFFERENCES WHEN ITERATING THE USER AGENT.

As it can be observed on the figures, the first conclusion that can be taken is that all websites concerning the travel industry (Expedia, TripAdvisor, Airbnb and Booking) show differences, according to this Index. Moreover, Farfetch is the only website from the fashion retail industry that presents variations, and it is also the only one from luxury ecommerce.

Also, TripAdvisor is the website that shows a higher variance between scores. Its values are between 0.3 (in fact, lower than 0.3) and 1.

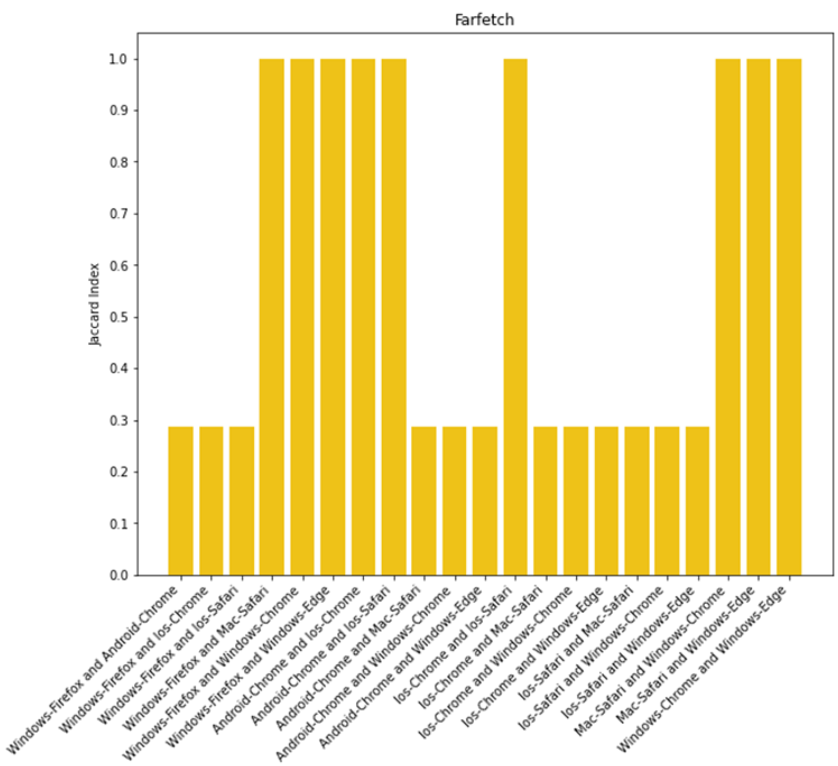

FIGURE 6 - FARFETCH JACCARD INDEX FOR EACH COMBINATION OF USER AGENTS.

About Farfetch, it can be concluded that the differences are between mobile and non-mobile versions. Meaning that all desktop user agents present the same products and all mobile user agents present the same products. Finally, between these mobile and desktop the Jaccard Index is approximately equal to 0.3.

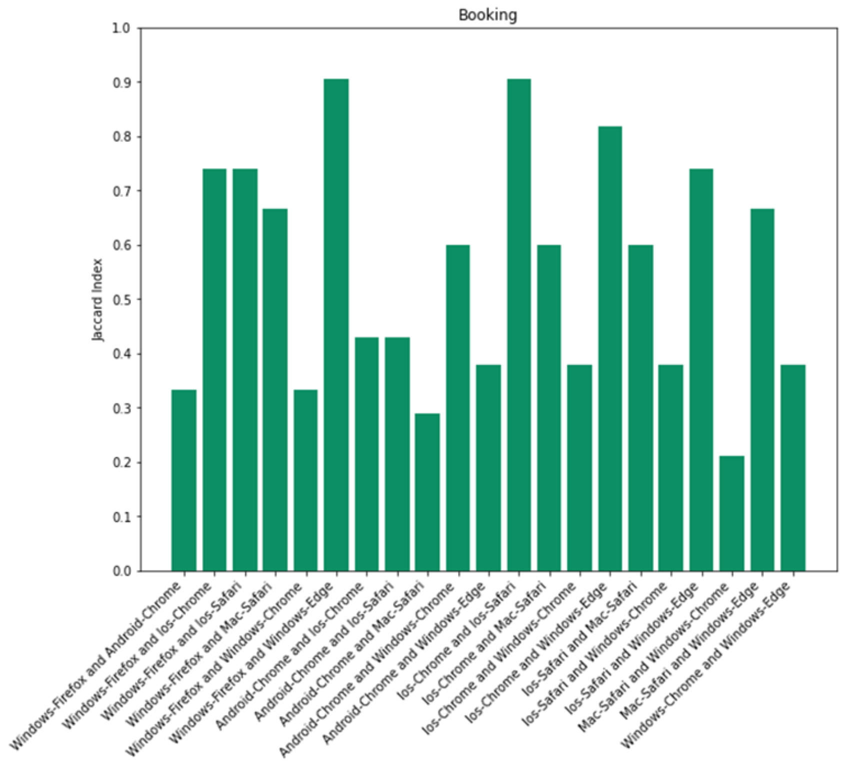

FIGURE 7- BOOKING JACCARD INDEX FOR EACH COMBINATION OF USER AGENTS.

Regarding Booking, there is no user agent with the exact same products as another. Moreover, the ones that share more similarities (i.e., have more products in common) are the iOS ones and Windows Firefox and Edge. On the other hand, the ones that share less similarities are Mac Safari and Windows Chrome, with a Jaccard Index of approximately 0.2.

b) About proxy:

Regarding the analysis using the different proxy locations, it was found evidence on price steering only on Nike and Expedia.

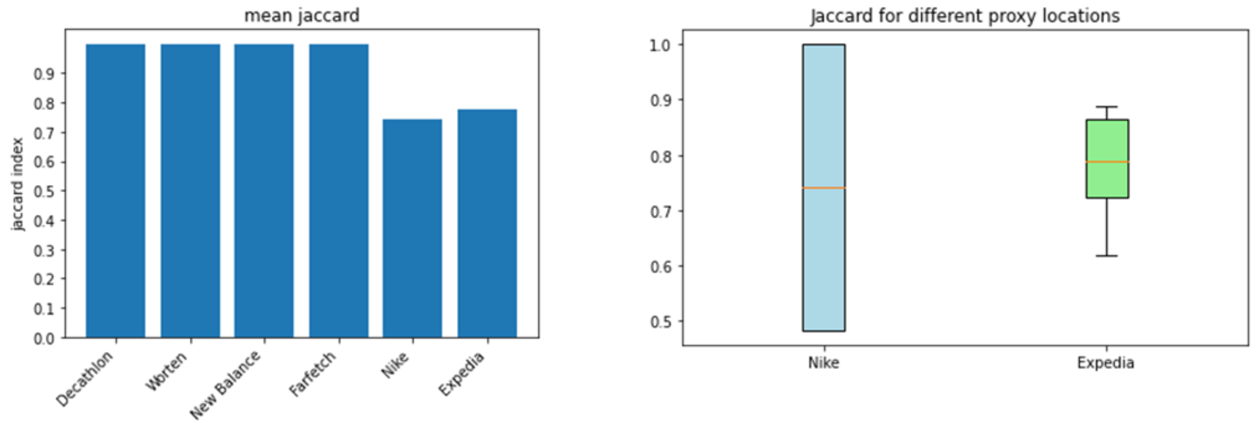

FIGURE 8- MEAN JACCARD SCORE OF DIFFERENT PROXY LOCATIONS FOR EACH WEBSITE AND VARIANCES OF THE JACCARD INDEX FOR EACH WEBSITE THAT WAS FOUND DIFFERENCES WHEN ITERATING THE PROXY.

With the right plot, it can be said that Nike only presents two different Jaccard scores and Expedia shows more variation, with values between, approximately, 0.6 and 0.9.

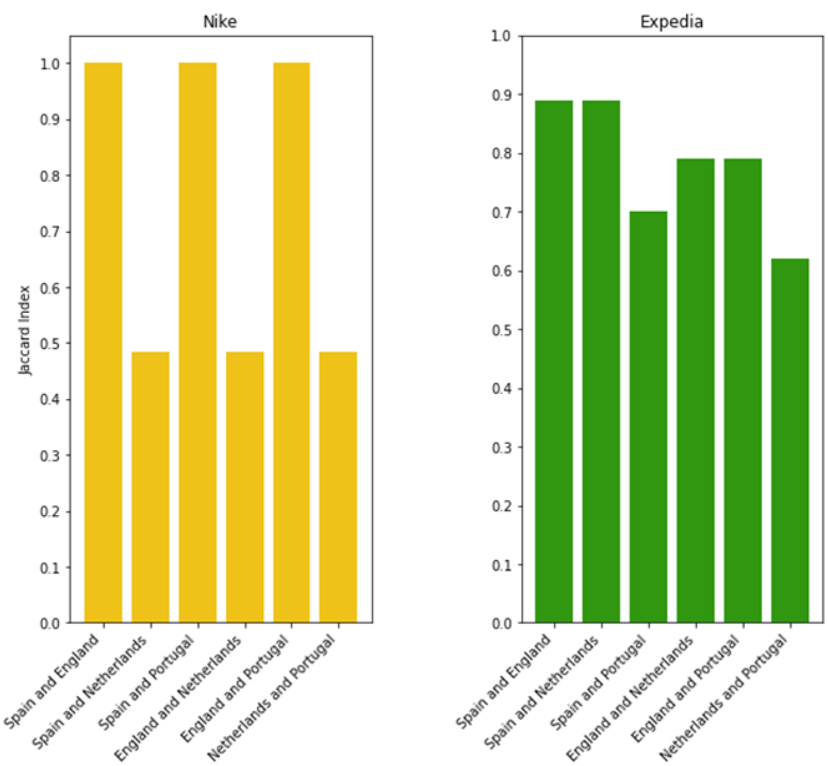

FIGURE 9- NIKE AND EXPEDIA JACCARD INDEX FOR EACH COMBINATION OF PROXY LOCATIONS.

Regarding Nike, all countries present the same products except Netherlands. The Jaccard index of this country with another one in the study is 0.5, meaning they only share half of the products.

About Expedia, there is no country with the exact same products as another. The ones that have more products in common are Spain and England and Spain and Netherlands. On the other side, the ones that show a lower index are Portugal and Netherlands.

7.2 Kendall's tau

7.2.1 Clarification on the metric.

Kendall's tau is a correlation coefficient that measures reordering between two lists. In practice, this metric compares the rank of each product in each search result.

The Jaccard index only measured if the products of two different queries were the same. The Kendall's tau goes even further and checks if the products are in the same order.

7.2.2 Methodology followed.

After transforming the CSV files into dataframes and dropping the price column, a column "rank" was created that is equivalent to the index. Afterwards, it was merged all dataframes concerning the same website, on the product name column, ending up with a dataframe with the rank of all products that appear on the results page for each user agent or proxy location.

After that, it was calculated the standard deviation of the ranks of each product. If the standard deviation equals to zero, it means that the product appears always in the same place, i.e., at the same position on the page. In addition, if the standard deviation equals Nan, it means that the product does not even appear when using some user agent or proxy.

Finally, as close as the standard deviation is to zero, the less fluctuations the product rank has. Therefore, it was assumed that there might be some evidence of price steering both if the Kendall's tau is equal to Nan or different than 0.

7.2.3. Results and Conclusions.

a) About user agents:

According to the following plot, all websites from the travel industry show evidence of price steering, according to the Kendall's tau metric. Moreover, Farfetch again and Worten for the first time, also present some differences in the products' rank, writing down the presence of price steering.

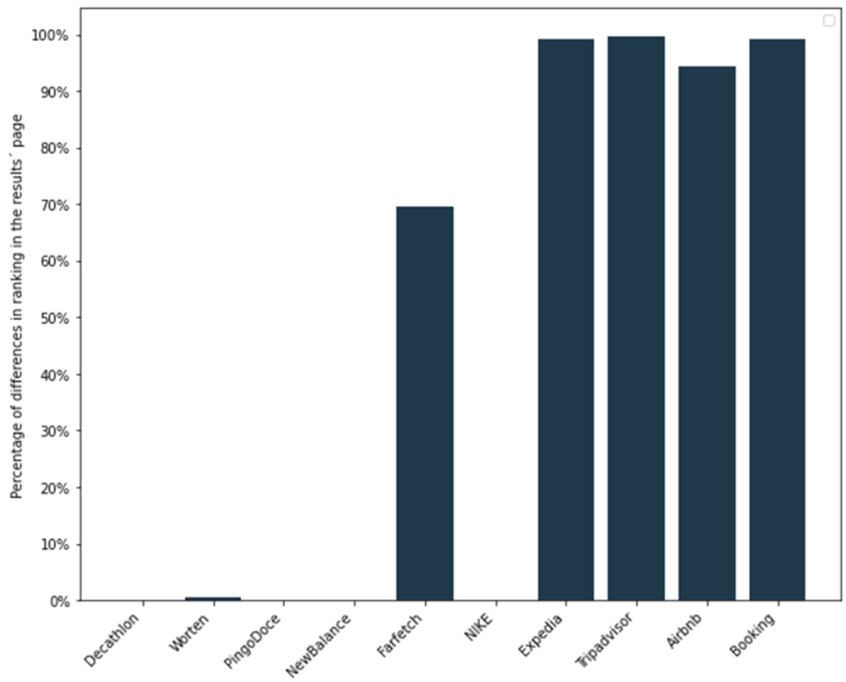

FIGURE 10- PERCENTAGE OF DIFFERENCES IN THE RANK OF A PRODUCT, DEPENDING ON THE USER AGENT.

First, it is straightforward to understand that all websites with a Jaccard Index lower than one, i.e., all websites that showed different products, must also show price steering according to this metric. This is because these websites are going to have products with a rank equal to Nan and, consequently, the standard deviation of the same will also be equal to Nan.

About the Worten case, since the Jaccard Index for this website was always equal to one, it shows that the existing differences in rank do not come from differences in the products, i.e., Nan values, but from real differences in the order of some products. However, the relative number of differences is so small (only 0.54% of products present differences in the rank), that it would be inaccurate to affirm that Worten shows evidence of price steering, based on this metric.

b) About proxy:

Regarding the differences found in the rank of products related to the different proxy locations, the websites with evidence were Worten, Farfetch and Expedia.

The reasoning behind it is the same as before. About Worten not much can be concluded as very few differences were found. Farfetch and Expedia, as proved before, had a Jaccard Index lower than one, and, as a result, makes sense to also show evidence of price steering with Kendall's tau metric.

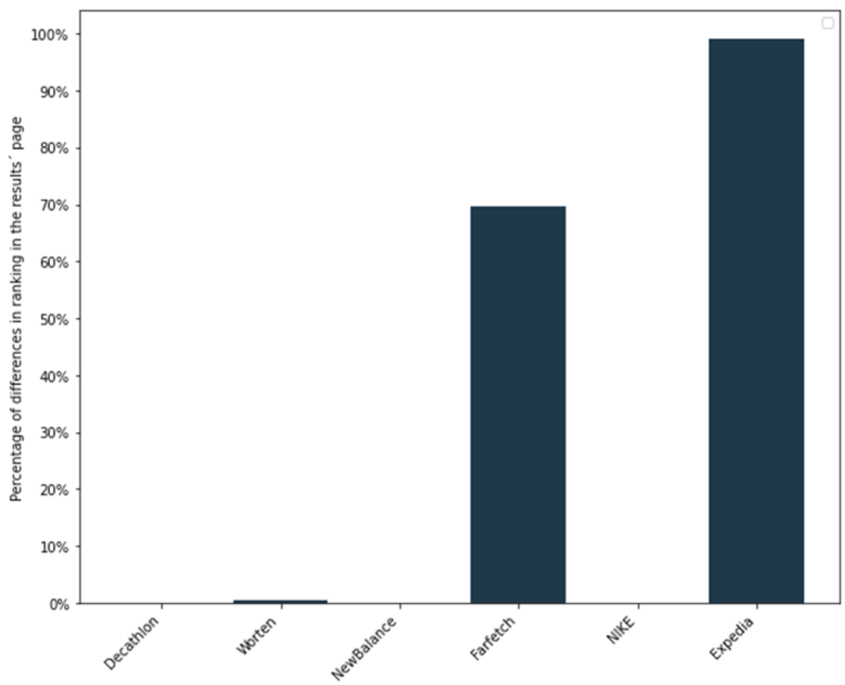

FIGURE 11 - PERCENTAGE OF DIFFERENCES IN THE RANK OF A PRODUCT, DEPENDING ON THE PROXY LOCATION.

7.3 Normalized Discounted Cumulative Gain (nDCG)

7.3.1. Clarification on the metric.

The Normalized Discounted Cumulative Gain is a metric that measures how strongly two variables are correlated. In this case, it was used to understand the correlation between the price of a product and its rank on the search results page.

7.3.2 Methodology followed.

After transforming the CSV files into dataframes, the product name column was dropped, leaving the dataframe with only the needed columns: rank and price. Later, the dataframes for each firm were merged by the index. Also, the dataframes were separated by each search term, to be able to consider only the relation rank – price for a specific search and not multiple searches at once.

Finally, after calculating the nDCG score for each search term, it was computed the mean of all of them, to only have one score per website.

7.3.3. Results and Conclusions

An nDCG of 1 means that the search results page is equal to the ideal one, i.e., products are shown in ascending order. On the other side, an nDCG equal to 0 means that no conclusions can be taken on the relationship of these two variables.

a) About user agents:

First, it can be concluded that all firms in the study have shown to have a strong price – rank relationship. This means that all websites present the products by prices in a somewhat ascending order.

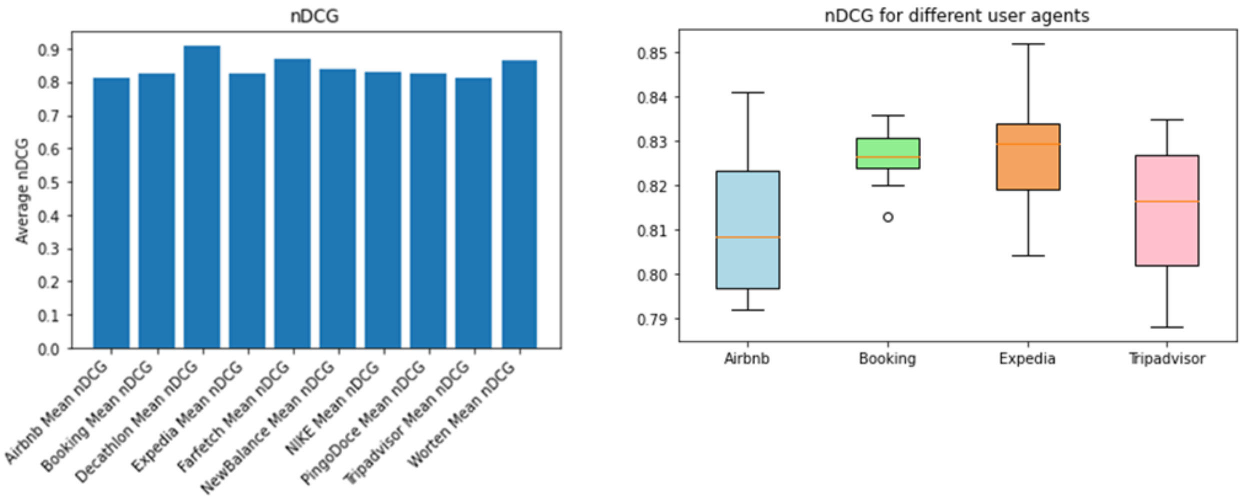

FIGURE 12- MEAN NDCG SCORE OF DIFFERENT USER AGENTS FOR EACH WEBSITE AND VARIANCES OF THE NDCG FOR EACH WEBSITE THAT WAS FOUND DIFFERENCES WHEN ITERATING THE USER AGENT.

Furthermore, concerning the different user agents, only the websites from the travel industry revealed to have different scores among the different users. However, the differences found were so small (i.e., lower than 5%) that it would be incorrect to assume that, according to this metric, evidence of price steering has been found.

b) About proxy:

Among the five websites that were used as case studies for the different proxy locations, Expedia, Farfetch and Nike were the ones that exhibited fluctuations in the nDCG scores across different countries.

Compared to the fluctuations seen when iterating the user agent, these were even less accentuated and, so, no evidence of price steering can be taken from this.

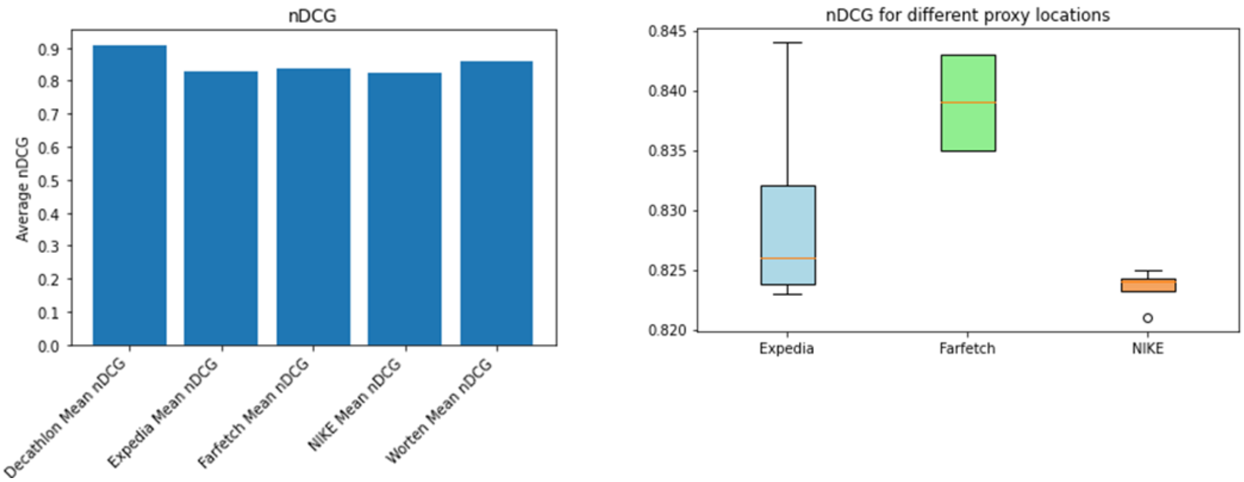

FIGURE 13- MEAN NDCG SCORE OF DIFFERENT PROXY LOCATIONS FOR EACH WEBSITE AND VARIANCES OF THE NDCG FOR EACH WEBSITE THAT WAS FOUND DIFFERENCES WHEN ITERATING THE PROXY.

7.4 Future steps and Recommendations

First of all, since the only website from the fashion retail industry that demonstrated to practice price steering was Farfetch, which is also the only from luxury industry as well, it would be interesting to analyze some other websites that sell luxury goods.

Also, it would be curious to expand the proxy locations study to more regions inside and outside Europe. Moreover, since most of the travelling websites were not tested with the proxy, it would also be noteworthy to discover a way to do it.

Finally, and the most obvious recommendation is to test other features, such as the browser history and the user being logged in or not, with the goal of finding evidence of price personalization on websites.

7.5 Conclusion

First, it can be concluded that there was found proof of price personalization on some of the analyzed websites, i.e., evidence that these websites use personal data concerning a specific user to customize the search results page for them, after they make a query. Therefore, the website tries not only to extract the maximum willingness to pay of each customer but also may end up capturing the entire market by personalizing the offer to each specific user.

Secondly, it is important to acknowledge that about the websites that it was not found evidence of this practice, does not mean that they do not do it. It simply indicates that these websites do not personalize prices based on the settings that were tested on this research: device, operating system (OS), browser, and geolocation. They may customize their offers, considering other criteria not tested on this paper, such as the users' browser history and the user being logged in or not.

Regarding price steering, the first conclusion that can be taken is that all websites from the travel industry showed evidence of this practice, considering two metrics: Jaccard Index and Kendall's tau. Booking is the most significative example since the results page for a product search is different for each user agent.

Also, it is noticeably clear that about the proxy location, it was not found as much evidence of price steering as regarding user agents. Although it cannot be neglected that fewer websites were tested for different proxy locations, including the travelling ones. Expedia was the only website that showed proof of price steering depending on the proxy location, according to the same two metrics.

Moreover, Farfetch, which is the only website from the luxury fashion retail industry that was tested, also showed evidence of price steering according to the same two of the metrics when iterating the user agents. The most relevant conclusion about this website is that it has two different results page after a search query: one for mobile devices and other for desktop ones.

It was also visible that, according to the Normalized Discounted Cumulative Gain metric, no clear evidence of price steering was found in any website. The differences in the nDCG scores that were found were too tiny to be considered.

Finally, the websites that were found to engage in pricing steering were all found to display a banner warning consumers that their personal information was handled and retained by the website.

Booking, Expedia, Farfetch, Airbnb, and TripAdvisor all display an alert informing users that the use of personal information is done so that the online experience can be improved and customized. TripAdvisor also emphasizes the collecting of geolocation information. It is also critical to understand that a website's content warning may vary depending on the browser and device used.

Therefore, it can be concluded that although all websites warn users about the handling and use of their personal data, it is not entirely and explicitly clear that the prices and ordering of products may be affected.

7. Challenges and limitations

The main challenge in developing this thesis was the data collection step. More websites were investigated, but some constraints were found. The scraping was sometimes very limited. Websites like FNAC would ban the exact search from being done multiple times in a limited period. Other e-commerce platforms, Trivago, for example, have in their terms and conditions that web scrapping is not allowed on their websites. Also, when the scrapping was possible, extracting the data embedded into the page's code was only sometimes possible. In some websites, it did not follow a standard pattern which made the collection of data impossible.

Another limitation mentioned before is that the scrapping accomplished using *Playwright* could not consider the geolocation variable. Since it was mainly on the websites of the travelling industry where price discrimination was found, important insights might have been lost

8. References

OECD. 2021. "The effects of online disclosure about personalised pricing on consumers: Results from a lab experiment in Ireland and Chile.". Paris.
https://doi.org/10.1787/1ce1de63-en.

European Commission. 2018. “Consumer market study on online market segmentation through personalised pricing/offers in the European Union.” London.
https://ec.europa.eu/info/sites/default/files/aid_development_cooperation_fundamental_rights/aid_and_development_by_topic/documents/synthesis_report_online_personalisation_study_final_0.pdf

UK government. 2018.“Pricing algorithms Economic working paper on the use of algorithms to facilitate collusion and personalised pricing”. London.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/746353/Algorithms_econ_report.pdf

Wolford, Ben. n.d. “What is GDPR, the EU’s new data protection law?”. GDPR.EU
https://gdpr.eu/what-is-gdpr/

Twin, Alexandra. 2022.“What Is Price Discrimination, and How Does It Work?” Investopedia.
https://www.investopedia.com/terms/p/price_discrimination.asp

Mattioli, Dana. 2012. “On Orbitz, Mac Users Steered to Pricier Hotels.” Wall Street Journal.
https://www.wsj.com/articles/SB10001424052702304458604577488822667325882

Bertini, Marco, Koenigsberg, Oded. 2021. “The Pitfalls of Pricing Algorithms.” Harvard Business Review.
https://hbr.org/2021/09/the-pitfalls-of-pricing-algorithms

Gal, Avigdor. 2017.“It's a Feature, not a Bug: On Learning Algorithms and what they teach us - Note by Avigdor Gal”
https://one.oecd.org/document/DAF/COMP/WD(2017)50/en/pdf

Ho, Shuk Ying. 2010. “The Attraction of Internet Personalization to Web Users.”
http://www.electronicmarkets.org/fileadmin/user_upload/doc/Issues/Volume_16/Issue_01/V16I1_The_Attraction_of_Internet_Personalization_to_Web_Users.pdf

Pasquali, Marina. 2022. “E-commerce worldwide – statistics & facts”.
https://www.statista.com/topics/871/online-shopping/#topicHeader__wrapper

Zuiderveen Borgesius, Frederik, & Poort, Joost. 2017. “Online Price Discrimination and EU Data Privacy Law”. Journal of Consumer Policy, 40(3), 347-366.

Christl, Wolfie. 2017. “Corporate surveillance in everyday life: How companies collect, combine, analyse, trade, and use personal data on billions”.
http://crackedlabs.org/dl/CrackedLabs_Christl_CorporateSurveillance.pdf

Poort, Joost. 2019. “Does everyone have a price? Understanding people’s attitude towards online and offline price discrimination.” https://policyreview.info/articles/analysis/does-everyone-have-price-understanding-peoples-attitude-towards-online-and-offline

Streitfeld, David. 2000. “On the Web, Price Tags Blur”. The Washing Post. https://www.washingtonpost.com/archive/politics/2000/09/27/on-the-web-price-tags-blur/14daea51-3a64-488f-8e6b-c1a3654773da/

European Union. 2016. L119. Official Journal of the European Union. 59. 35-43.

European Union. 2019. C232. Official Journal of the European Union. 62. 7-21.

European Commission. 2020. “Consumer market study on online market segmentation through personalised pricing/offers in the European Union.” https://ec.europa.eu/eurostat/cros/system/files/04_-_web_scraping_policy.doc

European Commission. 2017. “STUDY ON MEASURING CONSUMER DETRIMENT IN THEEUROPEAN UNION.” https://ec.europa.eu/info/sites/default/files/consumer-detriment-study-executive-summary_en.pdf

Hannak, Mislove, Wilson, Lazer, and Soeller. 2014. “Measuring Price Discrimination and Steering on E-commerce Web Sites.” https://mislove.org/publications/Ecommerce-IMC.pdf

n.d. “Understanding Operating Systems – University of Wollongong.” UOW. https://www.uow.edu.au/student/learning-co-op/technology-and-software/operating-systems/

“What is a web browser?”, Mozilla, n.d. https://www.mozilla.org/en-US/firefox/browsers/what-is-a-browser/

Smith, Justin. 2022.“Mobile eCommerce Stats in 2022 and the Future Online Shopping Trends of mCommerce.” outerboxdesign. https://www.outerboxdesign.com/web-design-articles/mobile-ecommerce-statistics

Jehanne, Marie. 2022. “Mobile vs Desktop Usage: What Are the 2022 Consumer Trends?” contentsquare. https://contentsquare.com/blog/mobile-vs-desktop/

gs.statcounter. 2022. “Desktop Operating System Market Share Worldwide.” https://gs.statcounter.com/os-market-share/desktop/worldwide

gs.statcounter. 2022. “Mobile Operating System Market Share Worldwide.” https://gs.statcounter.com/os-market-share/mobile/worldwide

gs.statcounter. 2022. “Browser Market Share Worldwide.” https://gs.statcounter.com/browser-market-share

sematext. n.d. “HTTP Requests.”
https://sematext.com/glossary/http-requests/#what-is-an-http-request

IBM. 2021. “HTTP responses.”
https://www.ibm.com/docs/en/cics-ts/5.2?topic=protocol-http-responses

Hicks, Kristen. 2020. “How IP Addresses Are Tracked.” hostgator.
https://www.hostgator.com/blog/how-ip-addresses-are-tracked/

n.d. 2022. “Cookie: Definition”. Google Ads Help.
https://support.google.com/google-ads/answer/2407785?hl=en

10. Annexes

10.1 Price Steering

	Farfetch	Expedia	Tripadvisor	Airbnb	Booking	User-agents
0	0.286	0.545	0.500	0.520	0.333	Windows-Firefox and Android-Chrome
1	0.286	0.619	0.364	1.000	0.739	Windows-Firefox and Ios-Chrome
2	0.286	0.545	0.429	1.000	0.739	Windows-Firefox and Ios-Safari
3	1.000	1.000	0.667	0.810	0.667	Windows-Firefox and Mac-Safari
4	1.000	1.000	1.000	0.810	0.333	Windows-Firefox and Windows-Chrome
5	1.000	0.889	1.000	0.810	0.905	Windows-Firefox and Windows-Edge
6	1.000	0.889	0.333	0.520	0.429	Android-Chrome and Ios-Chrome
7	1.000	1.000	0.500	0.520	0.429	Android-Chrome and Ios-Safari
8	0.286	0.545	0.538	0.462	0.290	Android-Chrome and Mac-Safari
9	0.286	0.545	0.500	0.583	0.600	Android-Chrome and Windows-Chrome
10	0.286	0.545	0.500	0.583	0.379	Android-Chrome and Windows-Edge
11	1.000	0.889	0.250	1.000	0.905	Ios-Chrome and Ios-Safari
12	0.286	0.619	0.395	0.810	0.600	Ios-Chrome and Mac-Safari
13	0.286	0.619	0.364	0.810	0.379	Ios-Chrome and Windows-Chrome
14	0.286	0.619	0.364	0.810	0.818	Ios-Chrome and Windows-Edge
15	0.286	0.545	0.463	0.810	0.600	Ios-Safari and Mac-Safari
16	0.286	0.545	0.429	0.810	0.379	Ios-Safari and Windows-Chrome
17	0.286	0.545	0.429	0.810	0.739	Ios-Safari and Windows-Edge
18	1.000	1.000	0.667	0.810	0.212	Mac-Safari and Windows-Chrome
19	1.000	0.889	0.667	0.810	0.667	Mac-Safari and Windows-Edge
20	1.000	0.889	1.000	1.000	0.379	Windows-Chrome and Windows-Edge

TABLE 3 – JACCARD SCORES BETWEEN COMBINATIONS OF DIFFERENT USER AGENTS.

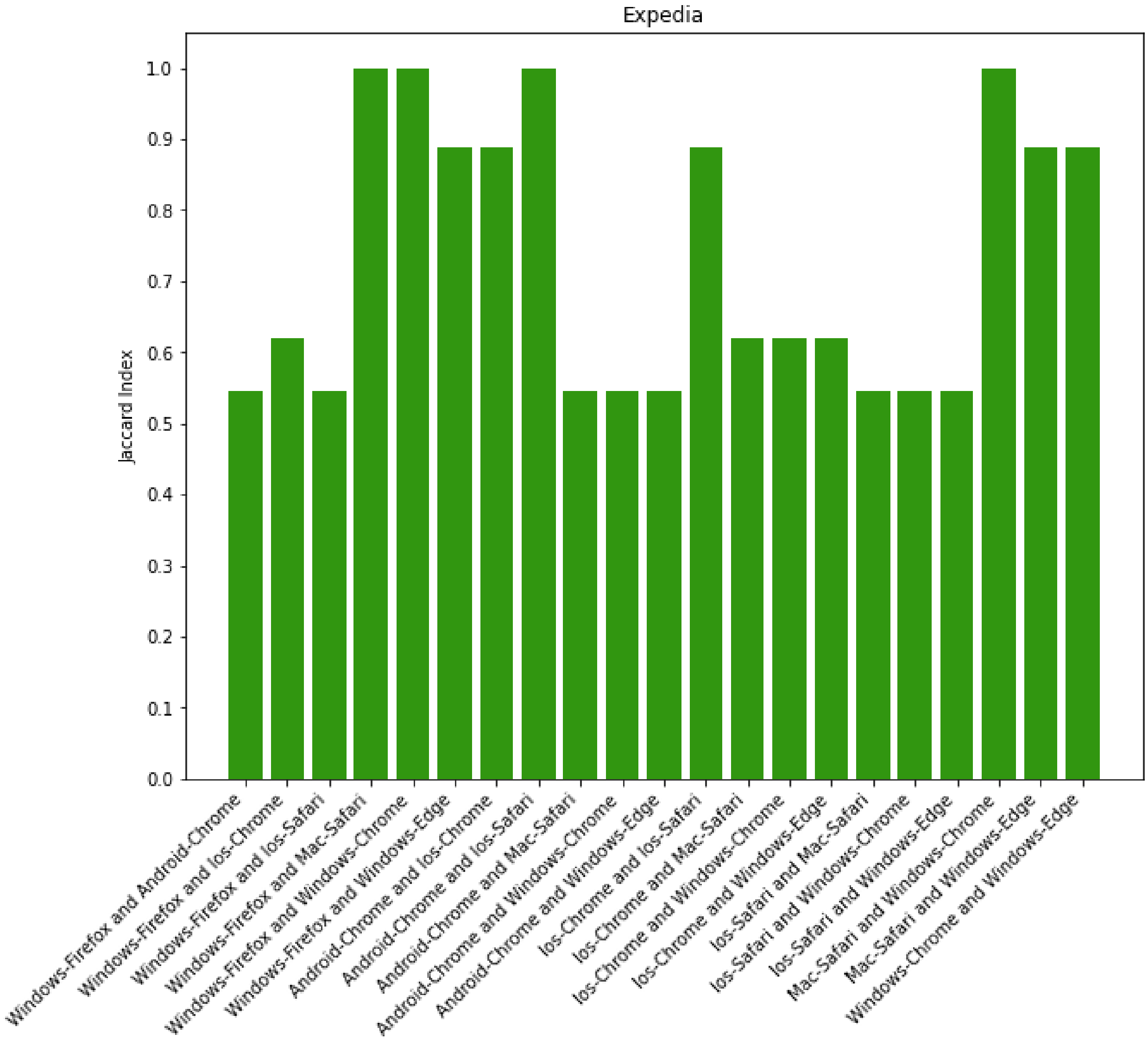

FIGURE 14 – EXPEDIA JACCARD INDEX PER COMBINATION OF USER AGENTS.

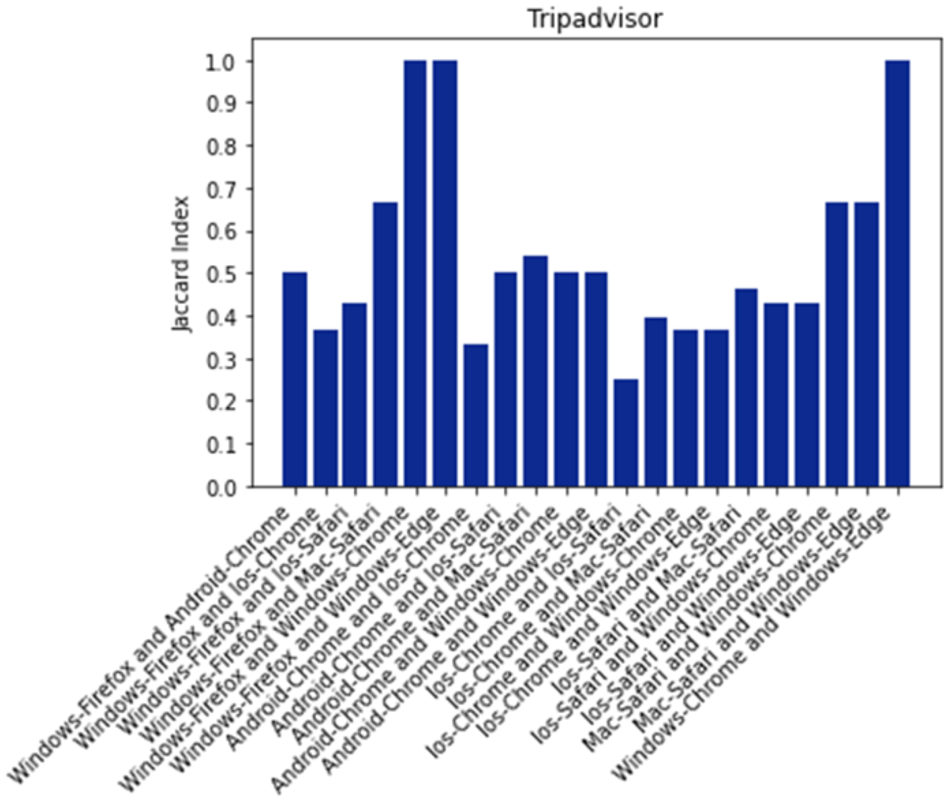

FIGURE 15 – TRIPADVISOR JACCARD INDEX PER COMBINATION OF USER AGENTS.

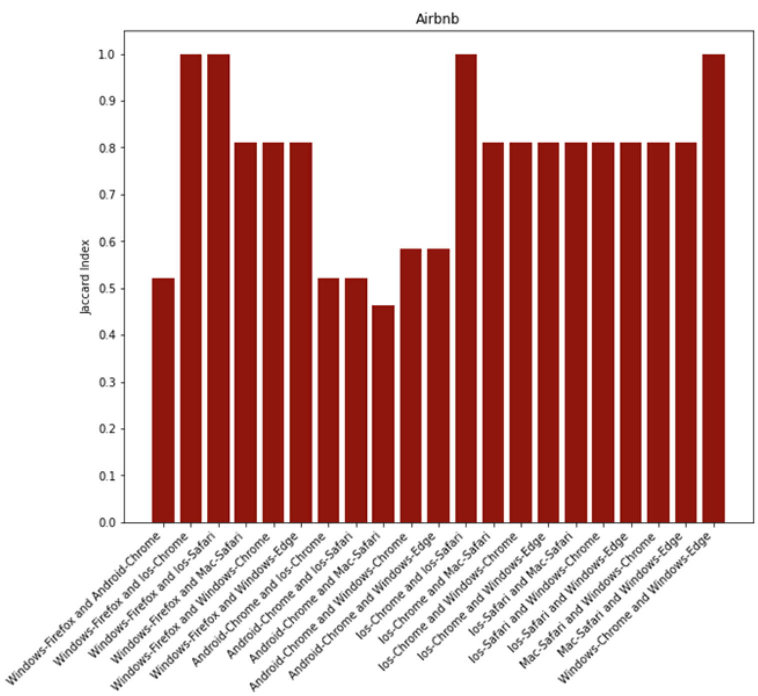

FIGURE 16 – AIRBNB JACCARD INDEX PER COMBINATION OF USER AGENTS.

	Decathlon	Worten	New Balance	Farfetch	Nike	Expedia	Proxy
0	1.0	1.0	1.0	1.0	1.000	0.889	Spain and England
1	1.0	1.0	1.0	1.0	0.483	0.889	Spain and Netherlands
2	1.0	1.0	1.0	1.0	1.000	0.700	Spain and Portugal
3	1.0	1.0	1.0	1.0	0.483	0.789	England and Netherlands
4	1.0	1.0	1.0	1.0	1.000	0.789	England and Portugal
5	1.0	1.0	1.0	1.0	0.483	0.619	Netherlands and Portugal

TABLE 4 – JACCARD SCORES BETWEEN COMBINATIONS OF DIFFERENT PROXY LOCATIONS.

	Airbnb Mean nDCG	Booking Mean nDCG	Decathlon Mean nDCG	Expedia Mean nDCG	Farfetch Mean nDCG	NewBalance Mean nDCG	NIKE Mean nDCG	PingoDoce Mean nDCG	Tripadvisor Mean nDCG	Worten Mean nDCG
Android-Chrome	0.792	0.830	0.909	0.804	0.867	0.837	0.832	0.826	0.829	0.865
Ios-Chrome	0.805	0.833	0.909	0.821	0.867	0.837	0.832	0.826	0.789	0.865
Ios-Safari	0.797	0.836	0.909	0.813	0.867	0.837	0.832	0.826	0.788	0.865
Mac-Safari	0.833	0.825	0.909	0.833	0.871	0.837	0.832	0.826	0.806	0.865
Windows-Chrome	0.841	0.813	0.909	0.837	0.871	0.837	0.832	0.826	0.826	0.865
Windows-Edge	0.796	0.827	0.909	0.832	0.871	0.837	0.832	0.826	0.835	0.865
Windows-Firefox	0.820	0.820	0.909	0.852	0.871	0.837	0.832	0.826	0.820	0.865

TABLE 5 – NDCG SCORES FOR EACH USER AGENT.

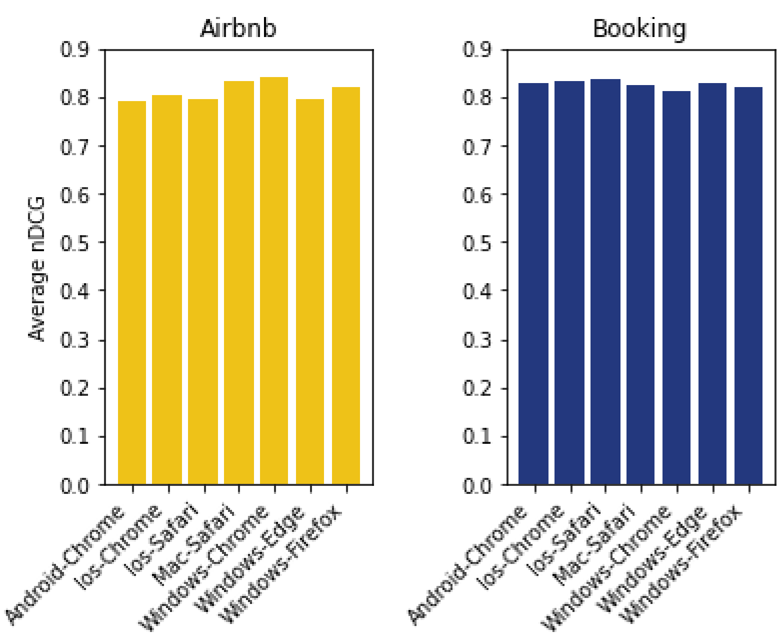

FIGURE 17 – NDCG SCORES OF AIRBNB AND BOOKING FOR DIFFERENT USER AGENTS.

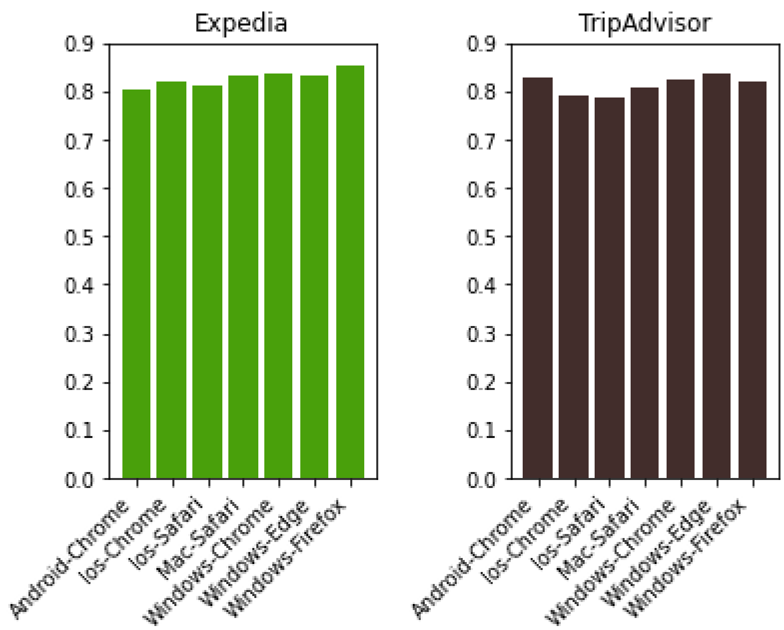

FIGURE 18 – NDCG SCORES OF EXPEDIA AND TRIPADVISOR FOR DIFFERENT USER AGENTS.

	Decathlon Mean nDCG	Expedia Mean nDCG	Farfetch Mean nDCG	NIKE Mean nDCG	Worten Mean nDCG
England	0.909	0.824	0.843	0.825	0.859
Netherlands	0.909	0.844	0.835	0.821	0.859
Portugal	0.909	0.823	0.843	0.824	0.859
Spain	0.909	0.828	0.835	0.824	0.859

TABLE 6 – NDCG SCORES FOR EACH PROXY LOCATION.

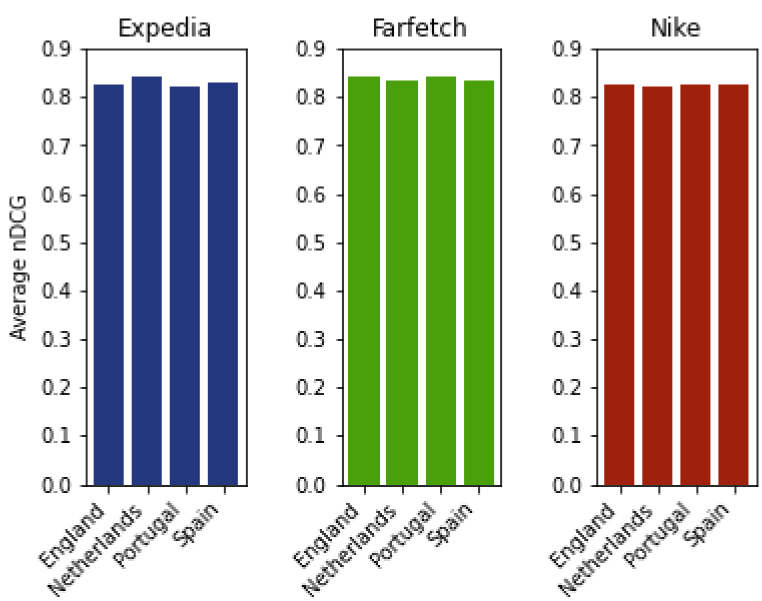

FIGURE 19 – NDCG SCORES OF EXPEDIA, FARFETCH AND NIKE FOR DIFFERENT PROXY LOCATIONS.