João Vitor Vargas Soares

Marketing Technology: a study on compliance, acquisition, and usage of first-party data in digital marketing

São Paulo, SP

2023

João Vitor Vargas Soares

Marketing Technology: a study on compliance, acquisition, and usage of first-party data in digital marketing

End of studies paper presented to the Computer Engineering and Digital Systems Departament at the Escola Politécnica of the Universidade de São Paulo to obtain the Title of Engineer.

Universidade de São Paulo – USP

Escola Politécnica

Departamento de Engenharia de Computação e Sistemas Digitais (PCS)

Supervisor: PhD Jorge Rady de Almeida Junior

São Paulo, SP

2023

Catalogação-na-publicação

Vargas Soares, João Vitor
Marketing Technology: a study on compliance, acquisition, and usage of first-party data in digital marketing / J. V. Vargas Soares -- São Paulo, 2023.
71 p.

Trabalho de Formatura - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1.Análise de Dados 2.Marketing Digital 3.GDPR 4.First-party Data 5.LGPD I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II.t.

This paper is dedicated to all my friends,
teachers and family members that were the foundation
of my life and my experiences acquired during my time
at the Universidade de São Paulo.

Acknowledgements

I would like to express my sincere thanks, first and foremost, to my parents, Neide and João, and my sister, Carolina, for shaping my character and providing unwavering support, enabling me to achieve the title of Computer Engineer. Next, I extend my gratitude to my friends, who were with me throughout my college years, serving as my second family away from home. I also thank my supervisor, Professor Jorge Rady, for guiding me during the execution of this project, along with all the other professors I had the privilege of learning from, both at USP and Grenoble INP. Without them, I would not have been able to receive an excellent higher education. I express my appreciation to Procter & Gamble for providing me with the opportunity to work at one of the world's largest consumer goods companies. Finally, I thank all others who were present during my education, from university staff to internship mentors who guided me in my professional life.

Abstract

This paper explores the important role of first-party data in digital marketing, examining its usage, the impact of Europe's GDPR, relating it with Brazil's LGPD. The study includes real-world cases of GDPR-compliant user data processing in a major consumer goods company. It is based on the student's internship at Procter & Gamble in France, covering the period from May 2022 to November 2022. The initial phase, supervised by P&G's Marketing Technology manager Vera Glukhenkaya and Grenoble INP professor Antoine Frenoy, focused on the implementation aspect. The second phase, undertaken at the Universidade de São Paulo for the graduation project, transitioned to an academic focus. The paper is divided into two projects that represent real-world use cases of the usage of first-party data in digital marketing. The first one validated the practicality of companies procuring first-party data from internal datasets, underlining the escalating emphasis on regulatory adherence. Analysis results showed the presence of potential consumers' data that could come from internal disorganized data from the company's databases. The subsequent project showcased the ability to extract meaningful insights through in-depth analysis of first-party data, employing Amazon Marketing Cloud to answer pivotal business questions. This approach demonstrated the potential for unveiling new insights from existing, untreated, and disorganized data, all aligned with the company's interests. Through comprehensive analyses and real-world cases, this paper underscores the challenges and transformative potential of strategically utilizing first-party data. The shift away from traditional reliance on second and third-party data is evident, showcasing businesses' adaptability to regulatory frameworks. The study also explores global variations in data protection regulations, emphasizing the need for a nuanced approach. Significantly, GDPR is portrayed not just as a regulatory framework but as a powerful tool empowering users to regain control over their personal information. Ultimately, the research envisions a future where the judicious use of first-party data, aligned with evolving regulations, plays a central role in achieving business goals and building consumer trust in the dynamic digital marketing landscape.

Keywords: First-party Data. GDPR. Digital Marketing. LGPD.

Resumo

Este artigo explora o papel importante de first-party data no marketing digital, examinando sua utilização, o impacto da *General Data Protection Regulation* (GDPR) da Europa e a relacionando com a Lei Geral de Proteção de Dados (LGPD) do Brasil. O estudo inclui casos do mundo real de processamento de dados de usuários em conformidade com a GDPR em uma grande empresa de bens de consumo. Foi baseado no estágio do aluno na Procter & Gamble na França, de maio de 2022 a novembro de 2022. A primeira fase foi supervisionada por Vera Glukhenkaya, gerente de Tecnologia de Marketing da P&G, e o professor Antoine Frenoy da Grenoble INP, e focou na implementação. A segunda fase, realizada na Universidade de São Paulo para o projeto de graduação, teve uma abordagem acadêmica. O artigo é dividido em dois projetos que representam casos de uso do mundo real do uso de first-party data no marketing digital. O primeiro validou a praticidade de empresas adquirirem first-party data de bases de dados internas, destacando a crescente ênfase na conformidade regulatória. Os resultados da análise mostraram a presença de first-party data de potenciais consumidores que poderiam vir de dados internos desorganizados nos bancos de dados da empresa. O projeto subsequente demonstrou a capacidade de extrair insights significativos por meio da análise aprofundada de first-party data, utilizando a Amazon Marketing Cloud para responder a perguntas de negócios. Essa abordagem demonstrou o potencial para descobrir novos insights a partir de dados existentes, não tratados e desorganizados, alinhados aos interesses da empresa. Por meio de análises abrangentes e casos do mundo real, este artigo destaca os desafios e o potencial transformador da utilização estratégica de first-party data. A mudança do tradicional uso de second e third-party data é evidente, mostrando a adaptabilidade das empresas às regulamentações. O estudo também explora as variações globais nas regulamentações de proteção de dados, enfatizando a necessidade de uma abordagem nuanceada. De forma significativa, a GDPR é retratada não apenas como um framework regulatório, mas como uma ferramenta poderosa que capacita os usuários a recuperar o controle sobre suas informações pessoais. Em última análise, a pesquisa mostra um futuro onde o uso judicioso de first-party data, alinhado às regulamentações em evolução, desempenha um papel central na conquista de metas empresariais e na construção da confiança do consumidor no dinâmico cenário do marketing digital.

Palavras-chave: First-party Data. GDPR. Marketing Digital. LGPD.

List of Figures

List of Tables

List of abbreviations and acronyms

GDPR	General Data Protection Regulation
EU	European Union
1PD	First-Party Data
2PD	Second-Party Data
3PD	Third-Party Data
P&G	Procter & Gamble
AMC	Amazon Marketing Cloud
C360	Consumer 360
USP	Universidade de São Paulo
MPN	Marketing Program Number
DBMS	Data Base Management System
KPI	Key Performance Indicator
PR	Purchase Rate
ROAS	Return On Advertising Spend
CSV	Comma-Separated Values
NTB	New To Brand

Contents

1 Introduction

1.1 Motivation

In the landscape of digital marketing, the usage of first-party data has become a cornerstone for personalized and targeted campaigns. This paper explores the intricate intersection of first-party data usage, the influence of the General Data Protection Regulation (GDPR) on digital marketing practices, and an examination of Brazil's Lei Geral de Proteção de Dados (LGPD) and Europe's GDPR. Furthermore, the study provides real-world use cases of processing user data in strict adherence to GDPR within the European context in a big consumer goods company in the area of Marketing Technology.

This paper offers a comprehensive exploration of the dynamic intersection between Marketing Technology, Data Analytics, and Consumer Behavior through the lens of an international internship undertaken during a double degree program at the University of Grenoble INP. The primary focus is on the internship experience at Procter & Gamble in Paris, France, where the student assumed the role of a Data Analyst for grooming brands. Drawing upon this rich experience, the paper navigates the intricacies of data-driven marketing strategies and their implications for the grooming industry. Furthermore, it discusses the adaptation of these insights in shaping a graduation project at the Universidade de São Paulo.

1.2 Objectives

This paper is based on the student's internship during their double degree exchange program in France at the University of Grenoble INP. This is the adaptation of the End of Studies project done by the student at the company Procter & Gamble, focusing on the research of the relationship between data analysis and digital marketing in the modern world where new digital marketing compliance rules, such as GPDR, are now in effect.

The internship was divided into two main projects: consumer first-party data acquisition and media activation; and retail data and collaboration using Amazon Marketing Cloud.

1.2.1 Consumer first-party data acquisition and media activation

This project intended to collect and activate consumer first-party data for P&G's grooming brands based on the company's fiscal year targets. Data activation, according to (ORACLE, n.d.), is unlocking value in data through the development of insights and turning those insights into action. The discussion of the change in data collection mechanisms and types in digital marketing is discussed here too, once new data regulations are in effect now in Europe.

1.2.2 Retail data and collaboration using Amazon Marketing Cloud

The second project had the goal of using the digital marketing platform Amazon Marketing Cloud to collect and analyze consumer data to answer Procter & Gamble's predefined business questions, using Python to analyze the data. AMC is a cloud-based clean room that contains information about sales and consumer activity on Amazon's website and can be accessed using SQL queries.

1.3 Justification

Understanding the dynamics of first, second, and third-party data in digital marketing is crucial as it forms the foundation of effective strategies and personalized user experiences. The influence of data protection regulations like GDPR and LGPD significantly impacts how organizations handle user data, making it imperative to navigate legal frameworks for compliance. Real-world use cases provide practical insights into successful implementations, challenges faced, and ethical considerations, offering valuable lessons for professionals and researchers. This knowledge is vital for informing strategic decision-making, ensuring responsible data processing practices, and maintaining trust in an ever-evolving digital landscape.

1.4 Organization of the Paper

The paper is organized into six chapters that will be described concisely now. The first chapter is Chapter 1, which introduces the research motives and objectives. In Chapter 2, the conceptual aspects used in the paper are explained, going from the definition of each type of data in digital marketing, to the data regulations that exist in Europe (GDPR) and in Brazil (LGPD), to the link between first-party data and cookies. Chapter 3 defines how the work was organized during its development, divided into two parts: the work done during the internship and the adaptation of it for the student's graduation project at the Universidade de São Paulo. Chapter 4 defines the system requirements for both of the projects that were developed during the internship. In Chapter 5, the development of the project is explained, going from the used technologies to the implementation of the project and its results. The final Chapter 6, gives an overview of the results, the conclusions reached with the study, and the continuation prospectives of the work done.

2 Conceptual Aspects

2.1 Types of data used in digital marketing

There are three main types of data used in digital marketing: First-Party Data, Second-Party Data, and Third-Party Data. These are terms used to describe different types of data that organizations can benefit from in their marketing and business activities. The definitions of each type of data, according to (ONAUDIENCE, 2021; SIGNAL, 2021), are discussed as follows.

- First-Party Data (1PD): consumer data that is collected and owned by the company. The collection of the data is done by software and systems created and owned by the respective company. This is the best kind of data a company can have about its consumers since it is more accurate and precise in terms of data quality, and it reduces the need for other kinds of consumer data. With this kind of data, the company can create user-personalized experiences and display them relevant advertisements. Some examples of First-Party Data collection are web and mobile app behavior, purchase history, in-store or call center interaction, surveys created by the company that are filled directly by the consumer, etc.

- Second-Party Data (2PD): it is First Party Data collected from other companies, normally business partners, e.g., consumer data that a supermarket collects from their consumers that a consumer goods company buys.

- Third-Party Data (3PD): Different from the two other types of data explained, Third-Party Data does not come from the direct relationship between consumer and company, but from an outside source that collected the data. This kind of data is the least precise and accurate since it does not involve directly the company's interests, but a more general perspective about the consumer. Some examples of 3PD are: what websites the consumer visits on the web, which products the consumer sees, what brands are interesting for this consumer, etc.

2.2 General Data Protection Regulation in Europe

The General Data Protection Regulation, referred also as GPDR, is a comprehensive data privacy law that came into effect in May 2018. It has the purpose of protecting users' personal data by establishing strict rules in terms of data collection, processing, and storage, imposing fines for non-compliance, and granting individuals more control over their data. These are the legal terms defined by (WOLFORD, n.d.):

- Personal Data: Personal data is any kind of information related to an individual that can identify them directly or indirectly. Some examples are location, ethnicity, gender, web cookies, etc.
- Data Processing: any action performed on data, automated or not. Examples: collecting, organizing, recording, etc.;
- Data Subject: the person whose data is processed;
- Data Controller: the person who decides why and how the data will be processed. In the case of an organization, it refers to the employees;
- Data Processor: a third party that processes personal data on behalf of a data controller.

2.3 First-Party data and cookies

A definition of HTTP cookies is given by Peters e Sikorski (1997): "Cookies are small data structures sent from a Web server to your browser and saved on your hard drive in a text file. They are nothing more than a string of characters (letters and numbers) that store certain pieces of information about you." This is a definition from 1997, so today this could be interpreted as small text files that websites place on a user's device, normally in the web browser, to track and store information about the user and their online behavior. This data is usually used for different purposes like user authentication, personalization, and data analytics (CISNA; PENG, 2000).

With all the definitions stated by the GDPR, third and second-party data collection becomes obsolete. Web browsers, such as Firefox, Google Chrome, and Safari, have started to phase out third-party cookies, which are trackers placed on a website by someone other than the owner, to be compliant with the new regulations. The collection of First-Party Data is the way for companies to continue collecting consumer data, which is a win for the consumer in terms of privacy and data protection, once the data is collected now needs to be directly asked to the consumer, with the "Allow cookies" pop-up for example. If the companies do not follow these new regulations, very high fines could be applied, which max out at €20 million or 4% of global revenue, and data subjects have the right to seek compensation for damages. Each country in the European Union can have its own specific rules for data collection and processing, but all of them have the GDPR as a base (HERMAN, 2022).

2.4 *Lei Geral de Proteção de Dados* in Brazil

The *Lei Geral de Proteção de Dados*, or General Data Protection Law in English, according to MPF (n.d.), is a Brazilian regulation created to safeguard the fundamental rights of privacy, personal data protection, and the individual's right to personal development. It was constructed based on Europe's GDPR and it has the objective of creating a legal framework to ensure data security and to standardize practices to protect the personal data of individuals in Brazil, aligning with international norms. These norms apply not only to organizations in Brazil but also to those abroad that process any data of individuals within national borders.

The *Autoridade Nacional de Proteção de Dados*, or Nation Data Protection Authority, is responsible for the implementation and regulation of LGPD in Brazil. The law also defines how risk management and prompt response to security breaches must be implemented to be compliant with the regulation. Security breaches may lead to penalties of a maximum of 2% of the organization's yearly revenue in the country, capped at R$50 million for violation.

3 Work Method

This paper is based on the projects undertaken by the student during their internship at Procter & Gamble in France, which spanned from May 2022 to November 2022. The first part of the paper was done based on a project development plan created by the student and their internship mentor, which was P&G's Marketing Technology manager Vera Glukhenkaya. She was responsible for the marketing campaigns done in Europe by the company, focusing on the data analysis and science aspect of it. The area in which the student worked was Marketing Technology, which is the area of marketing that involves designing and operating technology solutions in the service of marketing. It leverages technology to plan, execute, and measure campaigns and other marketing tactics by using software applications and solutions (BRINKER; HELLER, 2021; STEWART, 2012). The project was also supervised and approved by the Grenoble INP professor Antoine Frenoy, and what was done in it represents the implementation part of this paper. The schedule followed is shown in Table 1 below.

Table 1 – Work schedule - Internship

Order	**Date**	**Phase**
1	Mid May 2022	Internship start and onboarding on the company
2	Start of June 2022	Understand all P&G's technical tools and models (C360, AMC, etc.)
3	Mid June 2022	Start building AMC's queries and analytics models
4	Mid July 2022	Start testing AMC's models created and analyze data
5	Mid August 2022	Gather data collected and build presentation for business partners
6	End of August 2022	Start researching about Marketing Data Types for the FPD project
7	Start of September 2022	Analyze databases provided for the FPD project and start building queries
8	End of September 2022	Analyze data collected to get insights
9	End of October 2022	Finalize both projects and build final internship presentation
10	Mid November 2022	Final internship presentation and end of internship
11	End of November 2022	End writing the first part of the project

The adaptation of the end of studies project was done at the Universidade de São Paulo as the graduation project of the student. In this second part, the original paper was modified and the focus was on researching the usage of data in digital marketing, the new data regulations both in Europe and in Brazil, being more academic-driven, rather than business-driven. The phases of this work were defined by the student and their supervisor at USP and are shown in Table 2. It started with defining which conceptual aspects based on the implementation done were going to be discussed. The student then proceeded with the research to complement the work that was done. The conclusions were also a part of this in the paper's development.

Table 2 – Work schedule - Paper development

Order	**Date**	**Phase**
1	Start of August 2023	Defining themes related to the project made to complement the paper
2	End of August 2023	Research about themes and project requirements specification
3	Start of September 2023	Start paper adaptation to USPś format (ABNT)
4	End of November 2023	End of paper writing and adaptation
5	Mid of December 2023	Final project presentation

4 Requirements specification

4.1 Consumer first-party data acquisition and media activation project

Before defining the functional requirements of the system, some terms used to describe the system functionalities must be defined.

- Marketing Program: Marketing plan made by the company concerning all activities taken by the company to increase sales related to a brand or a collection of brands. It is divided by country;
- Marketing Program Number: A number related to a marketing program used to identify it when querying on P&G's databases. E.g. Gillette is part of the Grooming Brands, and in France, it has a unique Marketing Program Number (or MPN) that is 357;
- Corporate database: A database inside GCP that gathers all P&G's brands that do not have their marketing program number. It gathers information about its consumers. Each country has its corporate database;
- Grooming database: A database that includes all brands that are related to grooming, including Gillette, Gillette King C, and Braun. It gathers information about its consumers. Each country has its grooming database;
- Grooming golden trait: A Grooming Golden Trait is a consumer attribute that is directly related to grooming, for example, the type of shaver that the user owns, which brand of shaver
- Active consumer: a definition made by P&G of a consumer that has been active on events defined on the database. This will be further explained later.
- Potential grooming consumer: definition made by P&G of a consumer inside a database that is a potential buyer of grooming products. The consumer has to comply with the following parameters:
  - The consumer must have age and gender information;
  - The consumer needs to have an ID and be unique, meaning not collecting user IDs that are repeated;

  - The consumer must not be present on any other grooming-related databases since these consumers are already grooming products consumers;
  - The consumer must have at least one grooming golden trait;
  - The consumer needs to be active for the past 12 months;
  - The consumer must have agreed to receive email marketing communications, which means that they must have agreed to be contacted by email according to GDPR's regulations.

4.1.1 Functional Requirements

- The system has to authenticate with GCP's BigQuery to access all databases
- The system has to query all the consumers inside the corporate database who are potential grooming consumers for three different brands: Gillette, Gillette King C, and Braun
- The system has to query all the traits inside the corporate database that could be related to grooming products
- The system has to query all the events inside the corporate database that could be related to grooming products
- The system has to query all the data sources inside the corporate database that could be related to grooming consumers
- The system has to analyze all data collected related to traits, consumer events, and data sources to get insights based on the definition of a potential grooming consumer
- The system has to generate charts related to each analysis result
- The system has to do all the querying and analysis activity on five different corporate databases, divided by country (Italy, France, Germany, Spain, and the United Kingdom) and marketing program number
- The system must only query consumers with age and gender information
- The system must not query duplicated consumers, traits, sources, or events
- The system must not query consumers that are present on the grooming database
- The system must not query consumers that are not active in the past 12 months
- The system must only query consumers that have grooming golden traits

- The system must only query consumers that agree with the email communication regulation defined by GDPR
- The user has to be able to input the country selected for analysis and get as output the data related to that country analyzed, in chart and table formats

4.1.2 Non-Functional Requirements

- The system has to use Python programming language to treat the collected data;
- The system has to use SQL to recover data from the relational databases provided;
- The system has to use Google Cloud Platform's cloud services, such as databases and cloud computing environments;
- The system has to execute the analysis using Jupyter Notebooks and has to generate the reports directly from them;
- The connection with GCP's databases has to be done internally, not being able to connect to external networks;
- All data management has to comply with GDPR's regulations;
- The system has to be able to perform analysis for five different countries: Italy, France, Germany, Spain, and the United Kingdom;
- The system has to be able to be reused for future analysis of different countries.

4.2 Retail data and collaboration using Amazon Marketing Cloud project

On this project, the student was provided with a set of business questions by their supervisor, and consequently, all the analytical work conducted was predicated upon addressing them. The business questions were the following.

- What are the optimal advertisement campaign settings for frequency caps and day-parting for Amazon Display Ads?
- How do the different media vehicles, Display Ads, Search Ads, and SoV placements, work together?
- What are the different campaign paths in order of occurrence considering display and search advertisements?

- What is the time passed between the first and last impression/click made by a consumer on advertisements until they make a purchase?
- What consumer audiences are driving the most sales? Are there any age-based differences in terms of what works best for conversions (purchases)?

Within the context of this academic paper, it is imperative to provide definitions for certain terms pertinent to the business questions and subsequent implementation discussed later in the document. Further elucidation on terms associated with media vehicles will be presented in Chapter 5 under the heading "Used Technologies".

- Frequency caps: The optimal number of times a user is exposed to an advertisement in a defined period, e.g., in 15 days, 30 days, etc
- Day-parting: The division of the advertisement's exposure during a day, for example, how many times an advertisement is shown to consumers during the night;
- Campaign paths: The sequence of events or campaigns a consumer gets exposed to from the first advertisement exposure until the last exposure before purchase;
- Purchase rate: It is a KPI that represents the number of purchases related to a marketing campaign set up divided by the number of impressions or views the same advertisement has received;
- Return On Advertising Spend (ROAS): It is a KPI that represents the total value earned by sales related to a specific marketing campaign divided by the total value spent in advertisements related to the same campaign.
- Consumer audiences/segments: They represent the specific groups of individuals that have similar purchasing patterns and that fit into a common category. Marketers aim to reach and engage with them through online channels, to promote products, services, or messages in a more targeted and effective manner.
- Advertisement impressions: The number of views an advertisement has received in a period;
- Unique reach: Number of unique consumers that have seen an advertisement for that specific brand, time frame, and exposure group.

4.2.1 Functional Requirements

- The system must query the Display Advertisements information to generate reports related to the day parting of them.

- The system has to define what is the optimal frequency cap for Display Advertisements in a month;
- The system has to define what is the time of the day in which most of the sales related to a marketing campaign;
- The system must use the purchase rate, ROAS, number of impressions, and number of users as KPIs for evaluating what is the best frequency cap for Display Advertisements;
- The system must divide the result for frequency cap and day parting by the selected brands in analysis, Braun and Gillette;
- The system must generate charts related to day parting, dividing the results by hour of the day;
- The system must generate charts for frequency caps comparing the number of users that have seen an advertisement, the number of times the advertisement has been seen by the same users, with the purchase and ROAS rate related to them;
- The system must analyze how the different types of media vehicles work together and separately
- The system must compare the media vehicles analysis results using the following parameters: the number of impressions, the number of users that have purchased a product, the purchase amount for the campaign in Euros, the total cost of the advertisements, and the unique reach;
- The system must use ROAS and purchase rate as KPIs for the media vehicles analysis;
- The system must generate charts and tables with the results for the media vehicles analysis divided by the period of operation of the advertisement and by the brand in analysis;
- The system must compare the performance of the different media vehicles in peak periods and non-peak periods;
- The system must compare the different campaign paths that occur on Amazon's website and generate tables with all the results;
- The system must analyze how the consumers interact with advertisements after their first impression until their last impression on a predefined period;

- The system must generate charts for the user impressions analysis using as parameters the time passed after the impression, the purchases made up until that moment, and the actions made by the consumer related to that advertisement. All this must be done for each brand in analysis;
- The system must analyze what are the best-performing consumer audiences on Amazon for each brand in analysis and generate tables with this comparison on a predefined period;

4.2.2 Non-Functional Requirements

- The system has to use Python programming language to treat the collected data;
- The system has to use SQL to recover data from the relational databases provided on the Amazon Marketing Cloud website, generating Excel files with the results;
- The system has to use Jupyter Notebooks to treat the data and generate insights related to the business questions;
- The system has to use Google Cloud Platform's cloud services for treating the data, specifically AI Vertex;
- The connection with GCP's databases has to be done internally, not being able to connect to external networks;
- All data management has to comply with GDPR's regulations;
- The analysis has to be adaptable to generate insights for other countries than the one in the analysis, which is France.

5 Research Development

5.1 Used Technologies

5.1.1 Python libraries

5.1.1.1 *pydata_google_auth*

This library provides helpers for authenticating to Google APIs. It was used to get the data from the databases containing user information that are stored on GCP.

5.1.1.2 *pandas*

Open source data analysis manipulation tool built on top of Python programming language. It was used to create the tables that were extracted from the databases and to create the final analysis tables with the findings after the analysis.

5.1.1.3 *numpy*

A Python library used for array manipulation, linear algebra, matrices manipulation, and many other uses.

5.1.1.4 *datetime*

A module that supplies classes for manipulating date and time in Python.

5.1.1.5 *sys*

This is a module that gives access to variables used or maintained by the interpreter and to functions that relate to it.

5.1.1.6 *google.cloud*

A library used to access Google Cloud APIs programmatically via Python. It was used to access BigQuery, GCP's data querying tool.

5.1.1.7 *matplotlib*

Data visualization library that allows the creation of charts based on data that is given as input to the commands available. It was used to generate all charts with the final or partial result of the analysis done.

5.1.1.8 *openpyxl*

A library used to generate Excel files in the following formats: xlsx/xlsm/xltx/xltm.

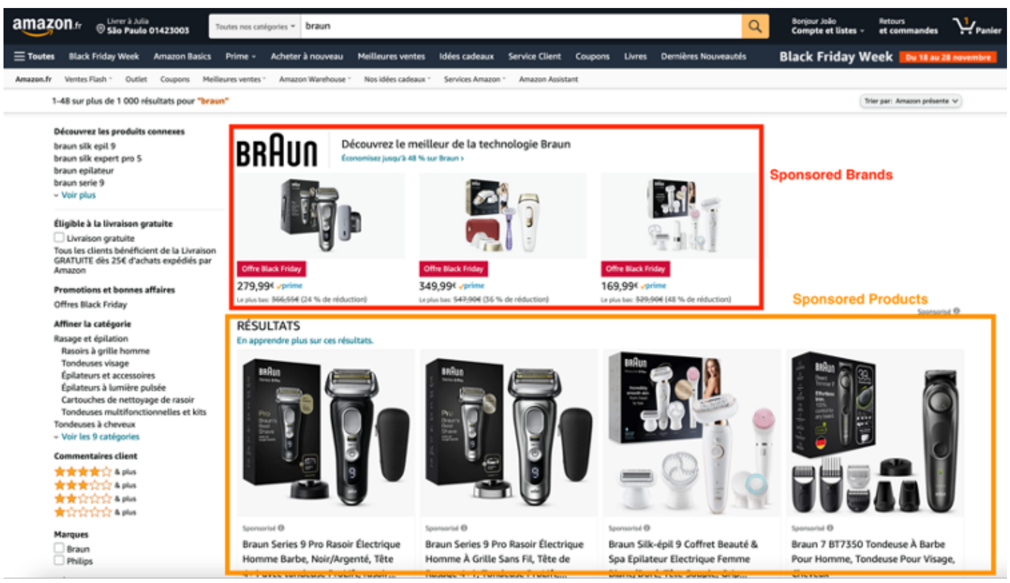
5.1.2 Amazon Marketing Cloud

Amazon is an important online consumer sales platform in Europe. In 2019, the company created a platform called Amazon Marketing Cloud that contains a database with granular data about user behavior on their website. This data can be accessed with SQL queries run inside their platform, generating output files in CSV format. The data can then be treated and used to generate conclusions and direct companies to their business goals (AMC, n.d.).

Inside Amazon, companies can define their marketing campaigns in two different formats, which will be defined as follows.

- *Sponsored Ads*: this type of marketing campaign includes the advertisements shown to a consumer when they search for products on Amazon. When a user searches for a product on Amazon, on the search results different types of advertisements appear. The ones related to *sponsored ads* are placed on the website in a high-visibility spot, which is good for user visibility. There are two types of *sponsored ads* on the platform: sponsored products, the type of advertisement that promotes a predefined set of products chosen by a company; and sponsored brands, which are shown on the search page to promote a brand as a whole, also showing brand-related products.

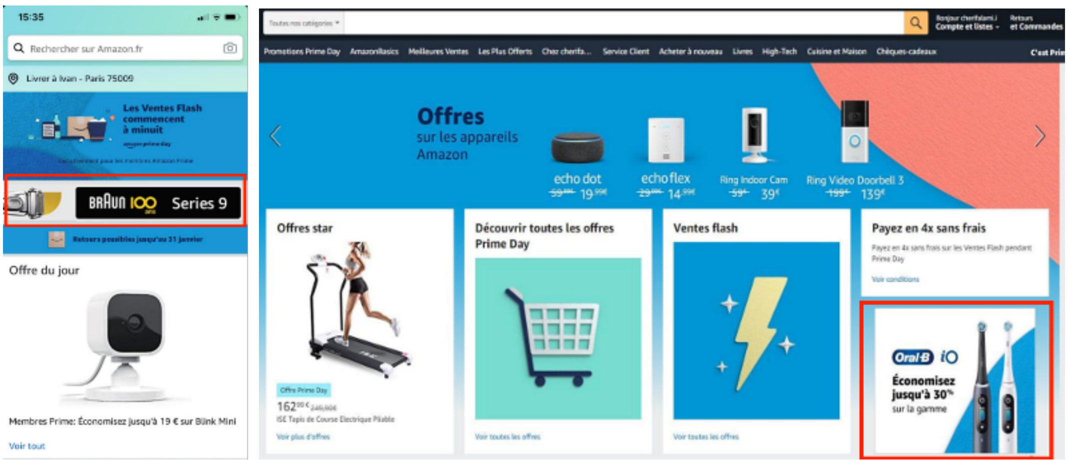
Figure 1 – Sponsored Brands and Products on Amazon



Source: Amazon's website

- *Display Ads*: this type of marketing campaign is related to the advertisements that appear on the top or sides across Amazon's website. It is usually a message with a call-to-action type of button and it has high visibility across the website. There are two types of *display ads* on the platform: Always-On, which are advertisements that are always running on the website; and the Share-of-Voice package, marketing packages for visibility that can be purchased by companies to boost marketing performance during seasonal periods, in which sales as typically higher, for example Black Friday or Amazon's Prime Day.

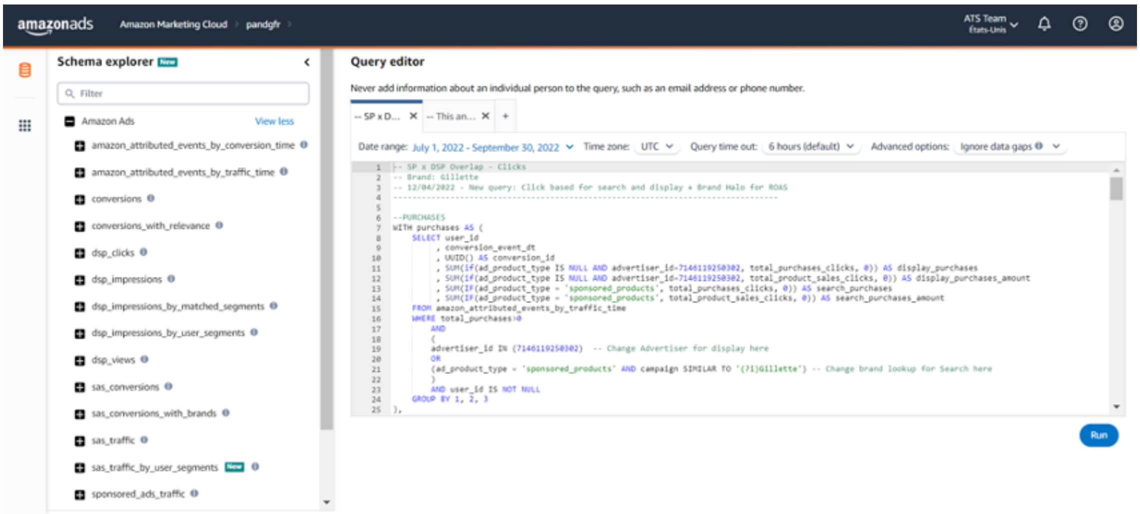
Figure 2 – Always-on (left figure) and Share of Voice (right figure) Display Ads



Source: Amazon's website

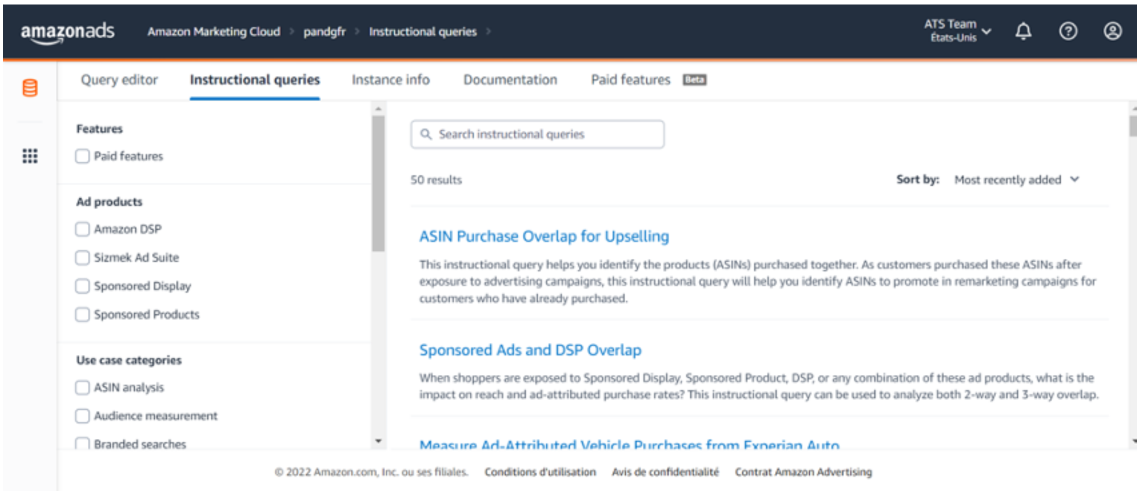
On the Amazon Marketing Cloud interface, it is possible to see all the tables related to the consumer data available on the platform. To consult those tables, SQL queries are used and executed using the interface provided. On AMC's website, there is a part (Instructional Queries) in which it is possible to find pre-made queries by Amazon that could relate to the questions that are going to be answered. Most of the queries used on all analyses were based on these pre-made queries, with minor modifications, and on queries created by other Procter & Gamble IT teams that have done similar analyses in the past.

Figure 3 – AMC's interface for query editing and execution



Source: Amazon Marketing Cloud website

Figure 4 – Instructional queries available on AMC



Source: Amazon Marketing Cloud website

5.1.3 Google Cloud Platform

Google Cloud Platform is a public cloud vendor such as Amazon Web Services or Microsoft Azure. It offers a suite of computing services that can be used to deal with all kinds of data. This is a pay-per-use type of platform and it is highly used by companies today. The tools provided by GCP that were used in this paper are: BigQuery, a fully managed enterprise data warehouse that uses SQL for data querying; and Vertex AI

Workbench, a Jupyter notebook-based development environment used for data science purposes.

5.1.4 Procter & Gamble's custom tools

Data is stored at P&G using custom tools developed by the company. The methodology used to create this system is called Consumer 360 in which all consumer data is stored by creating a master record of them, which has all data related to that consumer across the organization.

All of the personally identifiable data is treated differently, once it has to comply with GPDR's rules. To do that, the company stores this data in a platform called Segment. Segment is a consumer data platform used to store and activate consumer data, and it has the capability of anonymizing consumer data so that it can be used in marketing without concerns. This data can be consulted using Google Cloud Platform's tool BigQuery using SQL queries.

The non-personally identifiable data is stored by P&G on GCP's data lake tool. A data lake is used to store structured and non-structured data.

The C360 methodology used by P&G has some key concepts that need to be defined, and they will be referred to later in this paper.

- Touchpoint: any point of interaction with a customer or potential customer at any stage of the customer journey, some examples are websites, emails, and mobile apps;

- Events: user activity (clicks, views, orders) on a touchpoint. Example: for an email sent by P&G, an event related to that email is one that tracks if the email was opened;

- Sources: the source from where the information available about the consumer came from;

- Traits: user attributes such as name, gender, and city, but also computed traits, like the answers of a survey;

- Persona: a unified consumer profile with all the information about a consumer.

5.2 Project and Implementation

5.2.1 Consumer first-party data acquisition and media activation project

5.2.1.1 Files created for the analysis

During the project implementation, four Jupyter notebooks were created for this analysis, listed in Table 3.

Table 3 – Files created for the analysis and MPN selection

File name	Function
Traits analysis.ipynb	Analyze the user traits related to Grooming
Events analysis.ipynb	Analyze the events related to Grooming
Sources analysis.ipynb	Analyze the sources related to Grooming
All users analysis.ipynb	A complete analysis, considering all three categories listed above

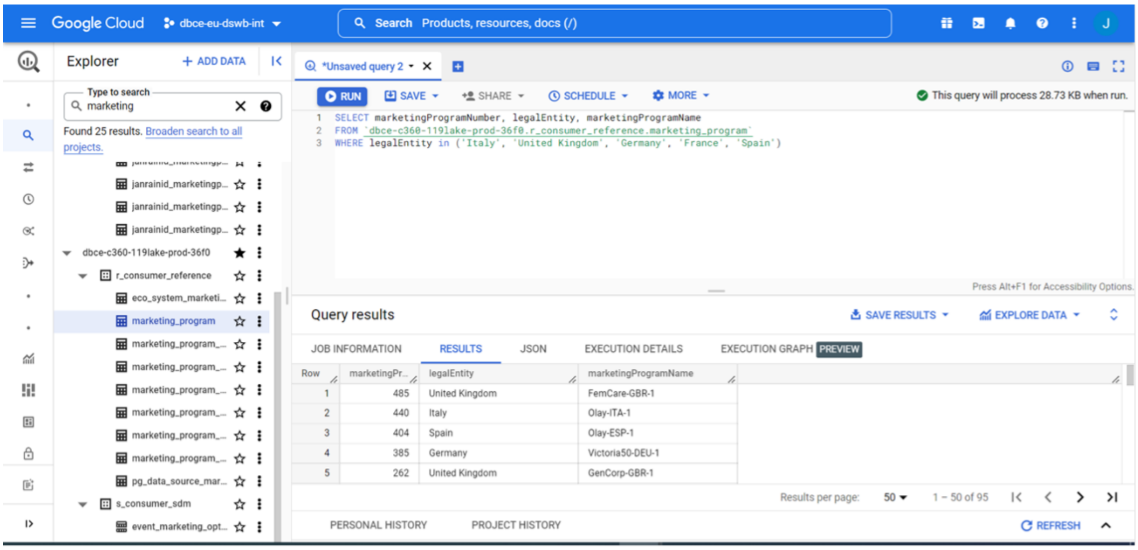
Inside all four files, the first blocks of code were used to import all Python libraries that were used on the execution, and the connection to the BigQuery client tool. This can be seen in the code below.

```python
# Google Cloud Tools
import pydata_google_auth
from pydata_google_auth.cache import CredentialsCache,
    ReadWriteCredentialsCache
from google.cloud import bigquery as bq

# Data Manipulation Libraries
import pandas as pd
import numpy as np
from datetime import timedelta,datetime,date
import sys

# Notebook tools
from IPython.display import HTML

# Create credentials
credentials_cache = ReadWriteCredentialsCache('/home/jupyter/.config/
    pydata/', 'pydata_google_credentials.json')
credentials = pydata_google_auth.get_user_credentials(
['https://www.googleapis.com/auth/cloud-platform'],
credentials_cache=credentials_cache
)
# Construct a BigQuery client object.
client = bq.Client(project='dbce-eu-dswb-int-cda2', credentials=
    credentials)
```

5.2.1.2 Recovering Marketing Program Numbers

This analysis was done in five different countries: Italy, France, Germany, Spain and the United Kingdom. Each country has its own Marketing Program Number (MPN) for the corporate database, also called the Growing Families database, and for the Grooming brands. The Grooming brands used in the analysis were Gillette and Braun, each with its own MPN respective to the country in the analysis. To recover the Marketing Program Numbers required, a query was executed directly from BigQuery's interface, shown in Figure 5 below. The columns *marketingProgramNumber*, *legalEntity* and *marketingProgramName* refer to the Marketing Program Number, the country, and the Brand's MP name, respectively.

Figure 5 – BigQuery's interface with MPN query execution and results



Source: Google Cloud Platform website

The first country analyzed was France, and then the analysis was expanded to the others mentioned. To select the correct MPN related to the brand and country found on the query present in Figure 5, the block of code shown below was also added to all of the analysis files.

```
if(sys.argv[1] in ['spain', 'germany', 'united kingdom', 'italy', '
    france']):
    country = sys.argv[1].lower()
else:
    country = str(input('Type the country you wish to analyze. Options:
    Spain, Germany, United Kingdom, France or Italy\n')).lower()
    while(country not in ['spain', 'germany', 'united kingdom', 'italy',
     'france']):
```

```
        country = str(input('Invalid country!\nType the country you wish to analyze. Options: Spain, Germany, United Kingdom, France or Italy\n')).lower()
# MPN set for the select country
if(country == 'spain'):
    mpn_gf = [290]
    mpn_grooming = [518, 308]
elif(country == 'germany'):
    mpn_gf = [293]
    mpn_grooming = [235, 331]
elif(country == 'united kingdom'):
    mpn_gf = [288]
    mpn_grooming = [322, 466]
elif(country == 'italy'):
    mpn_gf = [289]
    mpn_grooming = [350, 353]
elif(country == 'france'):
    mpn_gf = [291]
    mpn_grooming = [357, 507]
```

5.2.1.3 Executing queries on BigQuery via Jupyter Notebook

To execute a query on BigQuery inside a Jupyter Notebook and define the query parameters, the following lines of code were used as base code. To use it, it is only needed to define the query parameters and modify the *sql* variable by filling it up with the SQL code that is going to be executed. The query parameters mentioned before are variables created on the Jupyter Notebook that are passed as parameters to the SQL execution code written inside the *sql* variable.

```
# Base code for executing BigQuery queries from Jupyter Notebook
query_results = pd.DataFrame()
sql = """
INSERT HERE YOUR QUERY
"""
job_config2 = bq.QueryJobConfig(
    query_parameters=[
    # Insert here your query parameters, for example:
    bq.ArrayQueryParameter("mpn_gf", 'INT64', mpn_gf),
    bq.ArrayQueryParameter("mpn_grooming", 'INT64', mpn_grooming)
    ]
)
query_job = client.query(sql, job_config=job_config2) # Execution of the query with the parameters set above
query_results = query_job.result().to_dataframe() # Dataframe with the query results
```

5.2.1.4 Tables used in the analysis

The GCP's DBMS tables used in the analysis are defined in Table 4 below.

Table 4 – Description of the most important tables used in the analysis

Table name	**Definition**
dbce-c360-119lake-prod-36f0. s_consumer_sdm.**persona_trait**	Contains all the information on traits related to a persona (consumer profile)
dbce-c360-119lake-prod-36f0.r_consumer_reference. **marketing_program**	Contains all marketing programs information
dbce-c360-119lake-prod-36f0. s_consumer_sdm.**persona**	Contains a unified view of all consumer profiles
dbce-c360-119lake-prod-36f0.r_consumer_reference. **pg_data_source**	Contains information about the source from where a persona's data has come from
dbce-c360-119lake-prod-36f0.s_consumer_sdm. **eventg_event_email_clt**	Contains information about email events (clicks, opens, etc.)
dbce-c360-119lake-prod-36f0.s_consumer_sdm. **eventg_event_consumer_action**	Contains information on events related to consumers' actions (coupon events, cash-back events, etc.)
dbce-c360-119lake-prod-36f0.s_consumer_sdm. **eventg_event_loyalty**	Contains information about events that are related to loyalty programs
dbce-c360-119lake-prod-36f0.s_consumer_sdm. **eventg_event_website_event**	Contains information about events that happened on P&G's websites

5.2.1.5 Defining queries' constraints

Since all consumers need to have gender and age information as defined in the system requirements, this verification needs to be added to the queries. This can be done by selecting from the table *persona* the users that have that information, which is stored in the columns *p_key* and *p_value* present in the table mentioned. The column *p_key* contains the name of the information wanted, for example, "age" or "gender", and *p_value* contains the value related to that information, e.g., if *p_key* is "gender", one possible value of *p_value* is "Male". Within the database, there are various potential gender values. Consequently, before choosing consumers based on this data, it is essential to identify and combine all the available gender values in the database. As a final result, the only possible values for gender are *'m'* or *'f'*. The *userId*'s of the consumers need to then be selected after applying these constraints. The gender and age constraints SQL code are shown below.

```
persona_age as (
    select s.p_value as age,
           p.userId
    FROM `dbce-c360-119lake-prod-36f0.s_consumer_sdm.persona` as p,
   unnest(s_value_ots) as s
    where p.marketingProgramNumber in unnest(@mpn_gf)
    and (p_key = 'age')
),
persona_gender as (
```

```
    select case
                 WHEN REGEXP_CONTAINS(p_value, '(?i)Homme|H|M|Hombre|G|
    Garcon|Male|Manl|Man|Mannlich|Monnlich') THEN 'H'
                 WHEN REGEXP_CONTAINS(p_value, '(?i)Femme|F|Female|Mujer|
    G|Ms|Femle|Fenale|Fem|Femalw|femminile|donna|femmina|Weiblich|Mrs|
    Frau') THEN 'F'
                 ELSE 'N/A or Unknown'
             end as gender,
            p.userId
    FROM `dbce-c360-119lake-prod-36f0.s_consumer_sdm.persona` as p,
    unnest(s_value_ots) as s
    where p.marketingProgramNumber in unnest(@mpn_gf)
    and (p_key = 'gender')
),
persona_infos as (
    select a.age,
            g.gender,
            g.userID
    from persona_gender as g, persona_age as a
    where g.userID = a.userID
),
users_with_age_gender as (
    select pi.userId, trait, gender
    from persona_infos as pi, unduplicated as u
    where (gender = 'F' or gender = 'H')
    and age IS NOT NULL
    and u.userId = pi.userId
),
```

Another constraint defined by the system requirements is that a consumer must be active for the past 12 months. To select the active consumers, the events that the consumer participates in that are *not* considered as active need to be defined. For that, a variable was created on the Jupyter Notebook with a list containing those events, named *event_blacklist*, and passed as a parameter to the query. The variable created is present in the code below.

```
event_blacklist=[ "Campaign List", "Changed Opt Status", "Content Card
    Sent", "Email Hard Bounced", "Email Miscellaneous", "Email
    Miscellaneous", "Email Sent", "Email Soft Bounced", "Email Soft
    Bounce", "Invite Sent", "Postal Sent", "Push Notification Sent", "
    Push Notification Bounced", "Referral Sent Bonus", "SMS Sent", "
    Webhook Sent", "Invite Sent"]
```

Then, the last time the consumer has been active needs to be verified on all tables related to events. This is done by verifying if the column *originalTimestamp* contains a date that is inside the interval of the past 12 months. After that, the blacklisted events were excluded from this verification, resulting in a list of *userId*'s that are active in the past 12 months. The SQL code can be seen as follows.

```
all_events as (
select userId, marketingProgramNumber, event, originalTimestamp
from `dbce-c360-119lake-prod-36f0.s_consumer_sdm.eventg_event_app_event`
union all
select userId, marketingProgramNumber, event, originalTimestamp
from `dbce-c360-119lake-prod-36f0.s_consumer_sdm.
    eventg_event_consumer_action`
union all
select userId, marketingProgramNumber, event, originalTimestamp
from `dbce-c360-119lake-prod-36f0.s_consumer_sdm.eventg_event_email_clt`
union all
select userId, marketingProgramNumber, event, originalTimestamp
from `dbce-c360-119lake-prod-36f0.s_consumer_sdm.eventg_event_iot`
union all
select userId, marketingProgramNumber, event, originalTimestamp
from `dbce-c360-119lake-prod-36f0.s_consumer_sdm.eventg_event_loyalty`
union all
select userId, marketingProgramNumber, event, originalTimestamp
from `dbce-c360-119lake-prod-36f0.s_consumer_sdm.
    eventg_event_replenishment`
union all
select userId, marketingProgramNumber, event, originalTimestamp
from `dbce-c360-119lake-prod-36f0.s_consumer_sdm.
    eventg_event_sales_transaction`
union all
select userId, marketingProgramNumber, event, originalTimestamp
from `dbce-c360-119lake-prod-36f0.s_consumer_sdm.
    eventg_event_website_event`
union all
select userId, marketingProgramNumber, event, originalTimestamp
from `dbce-c360-119lake-prod-36f0.s_consumer_sdm.eventg_event_sms_clt`),
active_consumers as (
select
userId
from all_events
where date(originalTimestamp) > DATE_ADD(CURRENT_DATE(), INTERVAL -365
    DAY)
and marketingProgramNumber in unnest(@mpn_gf)
and event not in unnest(@event_blacklist)),
```

The next constraint is regarding email marketing communications. The consumers that have agreed to receive email marketing communications need to be selected, and this is stored in the *persona* table as well. This is done by selecting the *p_key* column that contains the email opt-in information and verifying if the *p_value* is set to *True*. It is also necessary to verify the date regarding that information to see if it is updated, once the consumer can decide to opt out of the email communications. This verification is done by adding the following lines of SQL code to the query.

```
opt_in as (
SELECT userId
FROM `dbce-c360-119lake-prod-36f0.s_consumer_sdm.persona` as b , UNNEST(
    s_value_ots) s
WHERE marketingProgramNumber IN unnest(@mpn_gf)
and p_key like "%email%Opt%" and (p_value = "true" or p_value = "1")
and p_originalTimestamp < @today),
```

One final constraint needs to be added to the analysis, which is related to the analysis done in Italy. In Italy, the databases have been through a cleaning regarding GDPR's regulations, which is related to the consent given by the user to allow the usage of their data. This is stored on the *persona_trait* table, where the *p_key* is equal to *"profiling_consent"* and the *p_value* is equal to *True*. This final constraint SQL code is available below.

```
profiling_consent as (
select distinct userId
FROM
`dbce-c360-119lake-prod-36f0.s_consumer_sdm.persona_trait`, UNNEST(
    s_value_ots) AS s
WHERE marketingProgramNumber in unnest(@mpn_gf)
AND (p_key = 'profilingConsent' OR p_key="profiling_consent")
and s.p_value='true'),
```

5.2.1.6 Consumer traits query

The first query created had the intent of looking for consumer traits that have grooming-related terms on its name. The names of the traits and their descriptions are present in the table *persona_trait* inside the variable *p_key*. Regular expression functions were used to search for grooming-related traits. Some terms were excluded from the search as it is shown on line 7 of the query shown below. The Grooming traits and all the unique *userId*'s related to them that are present on the Growing Families database were selected. After selecting them, there was an exclusion of: the *userId*'s that were present on any of the grooming databases; and the traits that have less than 10000 userId's associated with them. The final result is the number of unique consumers related to each consumer trait.

```
1 SELECT p_key, COUNT(DISTINCT userId) AS cnt
2 FROM (
3 SELECT DISTINCT p_key, userId
4 FROM `dbce-c360-119lake-prod-36f0.s_consumer_sdm.persona_trait`, UNNEST(
    s_value_ots) AS v
5 WHERE marketingProgramNumber IN UNNEST(@mpn_gf)
```

```
AND REGEXP_CONTAINS(p_key, '(?i)shav|razor|beard|gillette|braun|epilator
    |IPL|grooming|blade|facialhair|venus')
AND NOT REGEXP_CONTAINS(p_key, '(?i)score|lastupdate|enrich|bisnode|
    propensity|traitvalue|traitdate|ita_braun_presence|traitNumber|
    traitSequenceNumber|Number$')
EXCEPT DISTINCT
SELECT p_key, userId
FROM `dbce-c360-119lake-prod-36f0.s_consumer_sdm.persona_trait`
WHERE marketingProgramNumber IN UNNEST(@mpn_grooming))
GROUP BY p_key
HAVING COUNT(DISTINCT userId) > 10000
ORDER BY cnt DESC
```

After selecting the traits, all the constraints mentioned before need to be applied. The complete query is present in the *Traits analysis.ipynb* file was done by joining all of the constraints SQL codes. The query's output was stored in a Pandas Dataframe and reorganized to get the final results containing the number of unique *userId*'s. For each country analyzed, one Excel file was created to store the outputs of the analysis using the Python library *openpyxl*.

5.2.1.7 Consumer sources query

The next query created was regarding the sources from where the consumer data has come from. The first thing done was to get a list of all of the sources of information available, which are present in the *pg_data_source* table. With the sources selected, the next step is to select the ones that are related to grooming and get all of the consumers that have them as a source. To verify if the source is grooming-related, the columns used were *sourceName*, which contains the name of the source, and *description*, which contains its description. The following SQL code shows how this selection was made.

```
with src as (
SELECT sourceId, sourceName, description
FROM `dbce-c360-119lake-prod-36f0.r_consumer_reference.pg_data_source`
),
unduplicated as (
SELECT distinct userId, sourceName, description
FROM `dbce-c360-119lake-prod-36f0.s_consumer_sdm.persona` as a, UNNEST(
    s_value_ots) s
INNER JOIN src as b on safe_cast(s.p_value AS INT64)= b.sourceId
WHERE marketingProgramNumber in unnest(@mpn_gf) and p_key = 'sourceId'
AND (
REGEXP_CONTAINS(sourceName, '(?i)shav|razor|beard|gillette|braun|
    epilator|IPL|grooming|blade|facialhair|venus')
OR
REGEXP_CONTAINS(description, '(?i)shav|razor|beard|gillette|braun|
    epilator|IPL|grooming|blade|facialhair|venus'))
```

```
EXCEPT DISTINCT
select distinct userId, sourceName, description
FROM ‘dbce-c360-119lake-prod-36f0.s_consumer_sdm.persona‘,UNNEST(
    s_value_ots) s
INNER JOIN src as b on safe_cast(s.p_value AS INT64)= b.sourceId
WHERE marketingProgramNumber IN unnest(@mpn_grooming)),
```

As it was done for the traits query, the constraints defined before were also applied to get our final result on the *Sources analysis.ipynb* file, generating as output Excel files for each of the countries in the analysis.

5.2.1.8 Consumer events query

The events query was done based on the events selected in partnership with each country's e-commerce team. All of the selected consumers are unique, meaning that their *userId* appears only once in the results of the analysis. The selected events were the ones that were interesting commercially to each country and were related to grooming brands. The queries for each of the events selected and their definition are shown below.

- Email events: select all unique consumers that have events that track actions made by them related to email marketing sent to them (clicked, opened, etc);

```
unduplicated_email as (
SELECT distinct userId, event
FROM ‘dbce-c360-119lake-prod-36f0.s_consumer_sdm.eventg_event_email_clt
    ‘, unnest(products_struct) as ps
WHERE marketingProgramNumber IN unnest(@mpn_gf)
AND date(originalTimestamp) > DATE_ADD(@today, INTERVAL -1 YEAR)
and date(originalTimestamp) < @today
and event in ("Email Opened", "Email Clicked")
AND REGEXP_CONTAINS(payload.campaignname, '(?i)shav|razor|beard|gillette
    |braun|epilator|IPL|grooming|blade|facialhair|venus')
EXCEPT DISTINCT
select distinct userId, event
FROM ‘dbce-c360-119lake-prod-36f0.s_consumer_sdm.eventg_event_email_clt
    ‘, unnest(products_struct) as ps
WHERE marketingProgramNumber IN unnest(@mpn_grooming)),
```

- Coupon events: select all unique consumers that have events that track their usage of P&G's purchase coupons;

```
unduplicated_coupons as (
SELECT distinct userId, event
FROM ‘dbce-c360-119lake-prod-36f0.s_consumer_sdm.
    eventg_event_consumer_action‘
```

```sql
where marketingProgramNumber in unnest(@mpn_gf)
and (((
event = "Request Coupon Print" or
event = "Cashback Request Submitted" or
event = "Cashback Redeem Intent" or
event = "Cashback Request Processed" or
event = "Request Coupon Receipt Scan" or
event = "Coupon Redemption Receipt Scan" or
(event = "Coupon Redemption" and payload.sourceid in ("11644", "12076")))
and payload.brandname IN ('Braun', 'Gillette', 'King C Gillette', 'Venus'))
or
(
event = "Product Review Approved"
and REGEXP_CONTAINS(payload.endingweburl, '(?i)shav|razor|beard|gillette|braun|epilator|IPL|grooming|blade|facialhair|venus')))
EXCEPT DISTINCT
select distinct userId, event
FROM `dbce-c360-119lake-prod-36f0.s_consumer_sdm.eventg_event_consumer_action`, unnest(products_struct) as ps
WHERE marketingProgramNumber IN unnest(@mpn_grooming)),
```

- Receipt events: select all unique consumers that have events that track receipts from purchases made by them;

```sql
unduplicated_receipt as (
select distinct userId, event
from `dbce-c360-119lake-prod-36f0.s_consumer_sdm.eventg_event_loyalty`, unnest(products_struct) as ps
where event = "Receipt Verified"
AND marketingProgramNumber in unnest(@mpn_gf)
AND REGEXP_CONTAINS(payload.brandname, '(?i)shav|razor|beard|gillette|braun|epilator|IPL|grooming|blade|facialhair|venus')
EXCEPT DISTINCT
select distinct userId, event
from `dbce-c360-119lake-prod-36f0.s_consumer_sdm.eventg_event_loyalty`, unnest(products_struct) as ps
WHERE marketingProgramNumber IN unnest(@mpn_grooming)),
```

- Cashback events: select all unique consumers that have events that track cashbacks requested from purchases made by them;

```sql
unduplicated_cashback as (
select distinct userId, event
```

```sql
from `dbce-c360-119lake-prod-36f0.s_consumer_sdm.eventg_event_website_event`
where marketingProgramNumber in unnest(@mpn_gf) and
REGEXP_CONTAINS(payload.url, '(?i)shav|razor|beard|gillette|braun|epilator|IPL|grooming|blade|facialhair|venus')
and event='Coupon Redemption Cashback'
EXCEPT DISTINCT
select distinct userId, event
from `dbce-c360-119lake-prod-36f0.s_consumer_sdm.eventg_event_website_event`
 WHERE marketingProgramNumber IN unnest(@mpn_grooming)),
```

After selecting all types of events, the result of each query was combined to make sure each of the *userId*s collected were unique, as shown on the SQL code above.

```sql
unduplicated as (
SELECT *
FROM unduplicated_email
UNION DISTINCT
SELECT *
FROM unduplicated_coupons
UNION DISTINCT
SELECT *
FROM unduplicated_cashback
UNION DISTINCT
SELECT *
FROM unduplicated_receipt),
```

The constraints defined before were also applied, resulting in a list of events and the number of unique consumers that are related to those events. The analysis full analysis is on the file *Events analysis.ipynb*, and the results of it were saved on the Excel files created for each country.

5.2.1.9 All consumers query

The final file *All users analysis.ipynb* includes the merge of all of the sources, traits, and events SQL queries, as well as the constraints that were defined, resulting in a final number of unique users that could be potential grooming consumers. For the final number of users, the only events that were not considered were the ones related to email events, which was a decision that was made in discussion with the e-commerce team.

5.2.1.10 Executing the analysis

To execute all the analysis automatically for each country defined, a Makefile was created in the same directory as the analysis files. This Makefile file converts all Jupyter notebooks into Python files, executes all of them for each country, and then deletes the Python files. The code for the Makefile is shown below.

```
.PHONY: all convert execute remove
all: convert execute remove
convert:
	jupyter nbconvert --to=python "All users analysis.ipynb"
	jupyter nbconvert --to=python "Events analysis.ipynb"
	jupyter nbconvert --to=python "Source analysis.ipynb"
	jupyter nbconvert --to=python "Traits analysis.ipynb"
execute:
	python "All users analysis.py" italy
	python "Events analysis.py" germany
	python "Source analysis.py" "united kingdom"
	python "Traits analysis.py" spain
	python "All users analysis.py" spain
	python "Events analysis.py" italy
	python "Source analysis.py" germany
	python "Traits analysis.py" "united kingdom"
	python "All users analysis.py" "united kingdom"
	python "Events analysis.py" spain
	python "Source analysis.py" italy
	python "Traits analysis.py" germany
	python "All users analysis.py" germany
	python "Events analysis.py" "united kingdom"
	python "Source analysis.py" spain
	python "Traits analysis.py" italy
	python "All users analysis.py" france
	python "Events analysis.py" france
	python "Source analysis.py" france
	python "Traits analysis.py" france
remove:
	rm *.py
```

5.2.1.11 Main results

As mentioned before, the results were divided by country and saved on Excel files. In this section, only the results containing the total number of potential grooming consumers by country will be discussed, which were the outputs related to the file *All users analysis.ipynb*. The results related to the other files are available in the section *Other results*, in which the results are divided by traits, events, and sources analyzed separately.

The results for each country are shown in Table 5 below. *Total Users* is the number of users that have gender and age information and that are related to at least one grooming golden trait, event, or source. *Opt-in users* are the *Total Users* with the restriction of having agreed to receive email marketing communications. *Active users* are the *Total Users* with the restriction of being active in the past 12 months. *Active and Opt-in Users* is the number of users with both restrictions. The percentage of Male and Female consumers with both restrictions is also quantified in Table 5.

Table 5 – Potential grooming consumers analysis results

Country	**Total Users**	**Opt-in Users**	**Active Users**	**Active and Opt-in Users**	**Active and Opt-in Female %**	**Active and Opt-in Male %**
France	119186	88316	80977	66979	94%	6%
Italy	454006	407995	242937	239469	69%	31%
Spain	456780	305224	250391	202782	91%	9%
Germany	206550	82771	77083	58705	86%	14%
UK	145965	89591	65589	54906	89%	11%

By examining the results of the analysis, it is possible to see that for all countries the majority of consumers are female and that for most of them, most of the consumers are active and accepted to be contacted by email marketing. This means that most of the consumers are considered potential grooming consumers, which is the main point of this analysis. Germany and Italy are the only countries that have less than half the total consumers as potential grooming consumers, meaning that their consumers are not as active as in the other countries and that most of them do not agree to be contacted via email marketing.

Each of the countries analyzed had a target defined for first-party data consumer acquisition for the Fiscal Year of 2022-2023, which was defined by each country's e-commerce team. These numbers can be found in Table 6 below and they will be the parameter of success of the analysis. As it is shown, with this analysis, the targets related to FPD acquisition are partially fulfilled by the users found, which means that this analysis was a success. It is also important to keep in mind that this analysis will continue to be made at Procter & Gamble periodically, and it will allow the company to track the evolution of the prospective grooming consumers that are present on the growing families databases.

Table 6 – Fiscal year FPD target and percentage of target acquired per country analyzed

Country	**Target**	**Target acquired %**
France	270k	24.8%
Italy	240k	99.8%
Spain	200k	101.4%
Germany	520k	11.3%
UK	390k	14.1%

5.2.1.12 Other results

Each of the countries analyzed had its own set of consumer traits and events, as well as the sources of data. Since the analysis was made per country, the results will follow the same pattern. The first country analyzed was France and the results are shown in Tables 7, 8, and 9. There were only two grooming golden traits found in this analysis, but most of the consumers found come from these traits. The sources have the least number of consumers found of the three analyses done on the country, followed by the analysis of the events. In all countries analyzed, the final numbers of consumers found did not consider the email events, as they are not considered a strong factor in the definition of a potential grooming consumer. Knowing that the event that had the most potential grooming consumers was the one related to requesting purchase coupons.

Table 7 – Traits found in the French market analysis

Trait	**Total Users**	**Opt-in Users**	**Active Users**
traitHowShaveLegsIdValue	92333	68741	68164
traitHowDoYouShaveLegsValue	46668	31486	25076

Table 8 – Sources found in the French market analysis

Source	**Total Users**	**Opt-in Users**	**Active Users**
Belgium Gillette website	24	16	20
FRA Gillette DTC Clickstream V2 (10978)	192	0	1
France Braun website	22	14	16
Gillette Website	383	18	25

Table 9 – Events found in the French market analysis

Source	**Total Users**	**Opt-in Users**	**Active Users**
Coupon Redemption	3159	2463	3046
Coupon Redemption Receipt Scan	1430	1210	1430
Email Clicked	23891	18381	23891
Email Opened	457508	402646	457508
Product Review Approved	2077	1521	1725
Request Coupon Print	18398	14748	17329
Request Coupon Receipt Scan	6141	5027	6116

The analysis results for the Italian market are shown in Tables 10, 11, and 12 below. It was the country that had the most traits found, but most of them had the same consumers overall. The traits also represent most of the potential consumers found. The sources had a low impact on the analysis, with few potential consumers found. The coupon events were the most significant in the events analysis.

Table 10 – Traits found in the Italian market analysis

Trait	Total Users	Opt-in Users	Active Users
traitUseElectricShaverMaleIdValue	454815	408054	243550
traitWhatShaveOrBeardIdValue	454937	408176	243656
traitPurchaseFemaleRazorP12MIdValue	454816	408055	243551
traitShavingFrequencyIdValue	454823	408061	243551
traitNbrBladesUsedIdValue	454815	408054	243550
traitUseRazorWetShavingIdValue	454815	408054	243550
traitBrandElectricShaverIdValue	454815	408054	243550
traitIPLbrandUsedIdValue	454805	408045	243550
traitTypeOfRazorWetShavingIdValue	454816	408055	243551
traitEpilatorBrandUsedIdValue	454806	408046	243551
traitBrandPurchasedFemaleRazorP12MIdValue	20662	20570	20118
traitHowShaveLegsIdValue	44248	43990	41515
traitCampaignsIdValue	23970	23552	22747

Table 11 – Sources found in the Italian market analysis

Source	Total Users	Opt-in Users	Active Users
Braun Website Registration	7	6	4
Braun website	224	199	139
Gillette Website Registration	14	13	10
Legacy from old DB	4	4	3

Table 12 – Events found in the Italian market analysis

Source	Total Users	Opt-in Users	Active Users
Coupon Redemption Cashback	5802	5670	5802
Coupon Redemption Receipt Scan	215	214	215
Email Clicked	6369	6056	6369
Email Opened	64598	63011	64598
Product Review Approved	116	115	114
Request Coupon Receipt Scan	3283	3252	3281

The next country analyzed was Spain and the results are shown in Tables 13, 14, and 15 below. The traits also represent most of the consumers found, followed by the events, which had only one significant event considered on the final number, and finally, the sources, with few consumers found.

Table 13 – Traits found in the Spanish market analysis

Trait	Total Users	Opt-in Users	Active Users
traitHowShaveLegsIdValue	404395	255394	214156
traitShaveCarePrepUserHabitIdValue	17371	16098	16026
traitHowOftenFacialHairStyleChangeIdValue	27479	24972	23820
traitWhereDoYouBuyFemaleRazorsIdValue	12756	11864	12429
traitBeardCareProductsIdValue	16831	15499	15332
traitFacialHairStyleIdValue	67636	58406	52825
traitWhereUsuallyBuyRazorsIdValue	18448	17496	18176
traitRazorYouUseIdValue	35978	32422	32621
traitUseNonElectricRazorBrandIdValue	45076	39428	37015

Table 14 – Sources found in the Spanish market analysis

Source	Total Users	Opt-in Users	Active Users
Braun website	10	9	3
Lead generation web	4716	2212	1439

Table 15 – Events found in the Spanish market analysis

Source	Total Users	Opt-in Users	Active Users
Email Clicked	52594	44016	52594
Email Opened	629972	570086	629972
Product Review Approved	2626	2386	2360

The analysis was then made for Germany, with its results displayed in Tables 16, 17, and 18 below. It had a larger quantity of consumers found in the sources analysis compared with the three countries mentioned before. The traits represent the most consumers, and the events the least.

Table 16 – Traits found in the German market analysis

Trait	Total Users	Opt-in Users	Active Users
traitBraunProductIdValue	21468	857	845
traitCampaignsIdValue	16668	12485	15593
traitHowShaveLegsIdValue	150158	75695	74325

Table 17 – Sources found in the German market analysis

Source	Total Users	Opt-in Users	Active Users
Braun Cashback	4	1	1
DEU Growing Families Clickstream	40421	4053	2231
HP Atos Gillette	4046	560	457
I2	319	61	129
P&G Gillette website	173	48	33
i2/GRS/25/DEU/EU/Venus/MndTr	2	0	0

Table 18 – Events found in the German market analysis

Source	Total Users	Opt-in Users	Active Users
Email Clicked	40621	34015	40621
Email Opened	290481	261968	290481
Product Review Approved	36	23	22

The final region analyzed was the United Kingdom, resulting in the values present in Tables 19, 20, and 21 that follow. The sources analysis on this country had the most consumers found on all analyses. The traits, as in the other analysis, had the most consumers. The events had a very low impact on the final numbers.

Table 19 – Traits found on the UK market analysis

Trait	Total Users	Opt-in Users	Active Users
traitBeardGrowerInTheHouseholdIdValue	54858	46187	37583
traitHowShaveLegsIdValue	74418	43650	36424
traitShavingConcernsIdValue	29099	22190	20616
traitFrequencyShaveIdValue	7423	5845	7203
traitGroomingSurveyForIdValue	7616	5982	7377

Table 20 – Sources found in the UK market analysis

Source	Total Users	Opt-in Users	Active Users
Braun Website	4	2	1
Gillette Website Registrations	780	394	195
Gillette website	37280	12995	8481
Venus website Registrations	4	4	1

Table 21 – Events found in the UK market analysis

Source	Total Users	Opt-in Users	Active Users
Email Clicked	53825	32411	53825
Email Opened	957699	835397	957699
Product Review Approved	35	29	29

5.2.2 Retail data and collaboration using Amazon Marketing Cloud project

The implementation of this project was divided into five topics that will be discussed here, each of them answering the business questions defined in section 4.2.

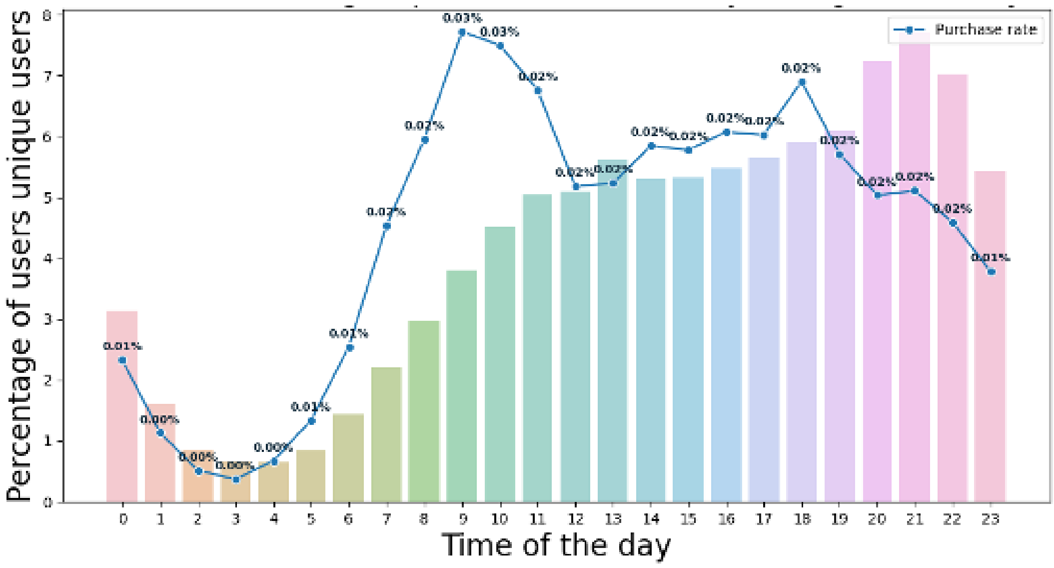
5.2.3 Frequency cap and dayparting

The objective of this part of the analysis was to understand, for Display Advertisements, how is the division of advertisements displayed during the day (dayparting) and what is the optimal frequency cap to be set for displaying them, in other words, how many advertisements they should be displayed to a consumer for each marketing campaign set up on a period of 30 days to get the best results in terms of purchases. This is important to know because Amazon charges their advertisements per impression and clicks (how many times the advertisement was clicked), meaning that a maximum value of times an advertisement is displayed needs to be set up so that the ratio between the money spent on a campaign and the return in sales is the best.

Two queries were used on AMC to get the information needed for the analysis, one for day parting and one for frequency cap. The queries were executed per month, meaning the date range selected on AMC was a month, and also per brand. To get the results per brand, the variable called *advertiser_id* was used on the *WHERE* clause inside the SQL query, which contains the id related to a brand. The two brands in the analysis were Braun and Gillette, so two *advertiser_id* values were used, one for each respective brand.

For day parting, the main KPI used to understand how the campaign has performed is the purchase rate. It was calculated for each hour of the day on the Jupyter Notebook created for this analysis. After calculating the purchase rate per hour, a chart was plotted for each marketing campaign in analysis. On this chart, besides the purchase rate per hour, the percentage of the total daily impressions a campaign has received is displayed too, also divided by the hour of the day. One of the results of this analysis is shown in Figure 6 below. It is possible to see that the time of the day that this specific campaign receives the most impressions is during the night, but the best purchase rate happens during the morning period.

Figure 6 – Day parting chart for a Braun marketing campaign



For the frequency cap, the need was to understand what is the optimal number of advertisements that need to be displayed to a consumer to get the best results. This was also done by dividing by marketing campaign and by brand, using the variables *campaign_id* and *advertiser_id* on the queries. After getting the queries' results, the KPIs purchase rate and ROAS were used to understand what are the optimal values.

This analysis was done in two parts: In the first part, the number of impressions was aggregated in groups of 5, e.g., if a user has seen an ad one time in the period of 30 days, the user is going to be part of the *1 to 5 ads* group; on the second part, the analysis was expanded by modifying this query to get more granular values, and instead of using frequency groups, the frequencies were considered as individual groups, so for the same example given before, the user would be fit into the *1 ad* group. For each of the groups, the percentage of users that fit into them was calculated. Two different charts were plotted for each campaign in analysis, one comparing the purchase rate for each group and the number of users in the group, and the other comparing the purchase rate for each group with the ROAS related to that group.

Figures 7 and 8 show the results of the first part of the analysis for Braun shavers campaigns and Figures 9 and 10 show the second part of the analysis done for the same campaigns. It is possible to acknowledge that in the first part of the analysis, the results are not granular, meaning that it is not possible to find exact values of frequencies that are optimal. For the first part of the analysis, the campaign chosen has its optimal frequency cap value between 6 to 10 advertisements per month, since it is where the ROAS and purchase rate are the highest. In the second part, more granular values of frequencies are defined, and for the specific campaign displayed on the charts, the optimal value would be 10 advertisements per month, as it is when the ROAS and purchase rate are the highest.

Figure 7 – Frequency cap for Braun shavers in November 2021 – Purchase rate vs Percentage of users in frequency bucket (groups of 5)

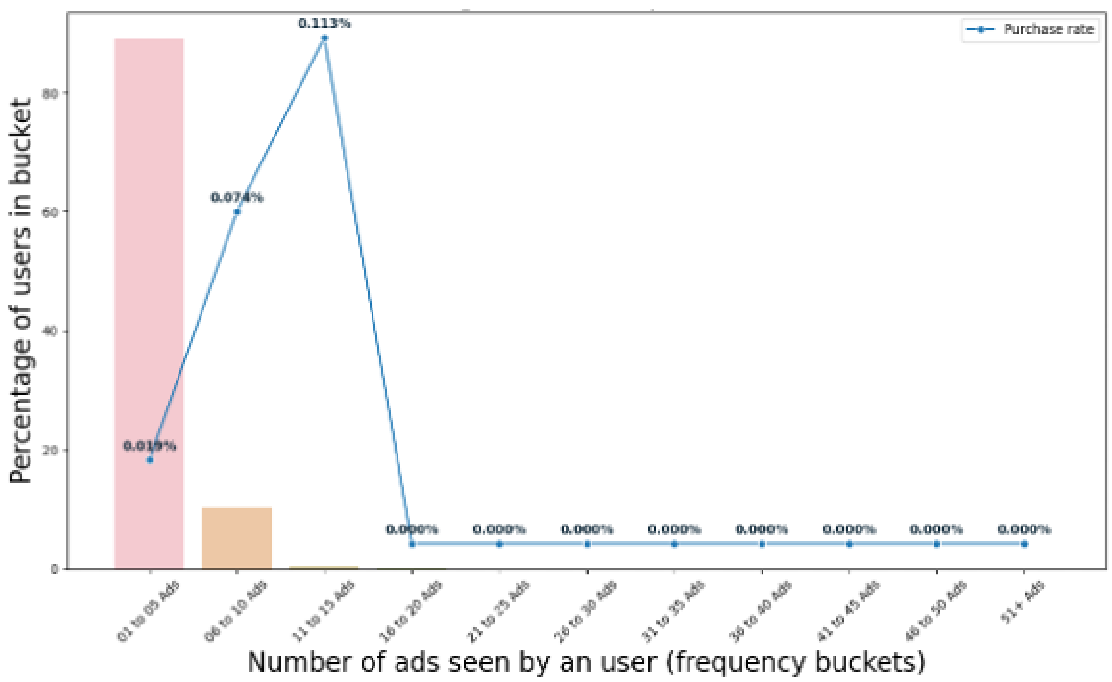


Figure 8 – Frequency cap for Braun shavers in November 2021 – Purchase rate vs ROAS per frequency bucket (groups of 5)

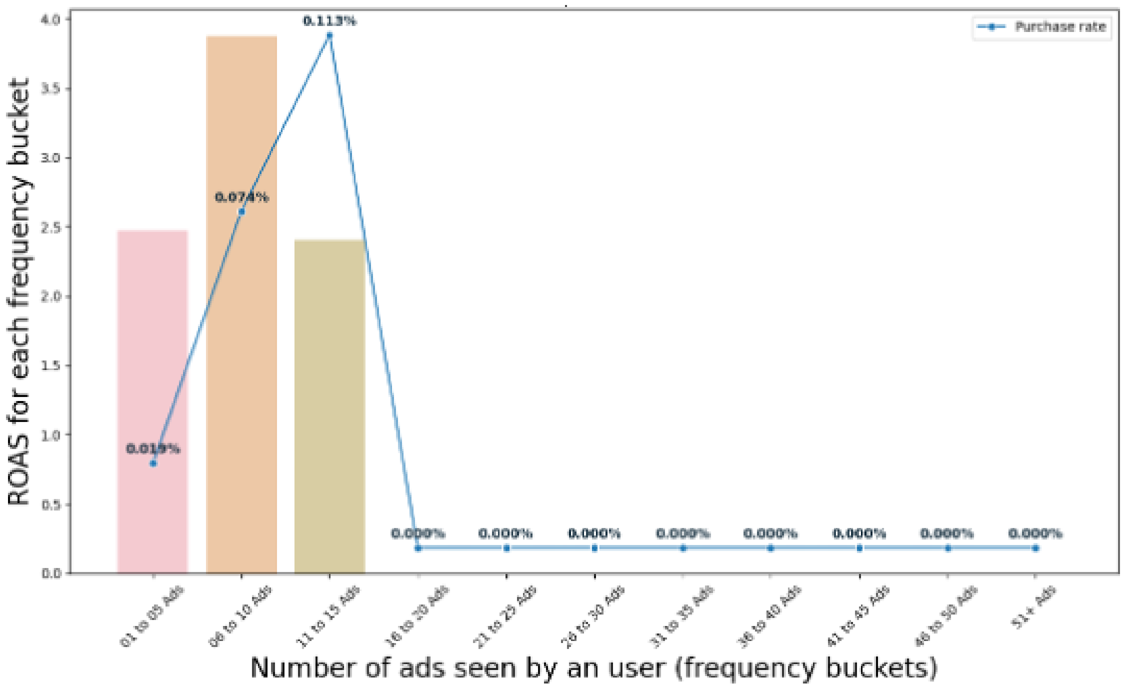


Figure 9 – Frequency cap for Braun shavers in November 2021 – Purchase rate vs ROAS per frequency bucket (groups of 5)

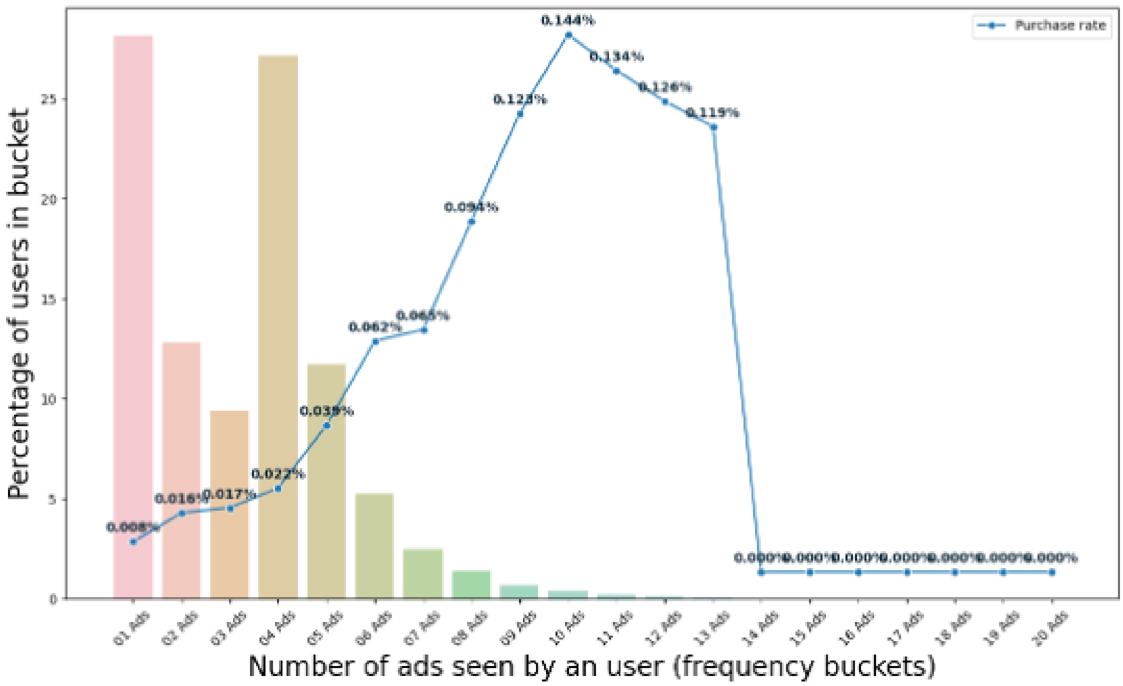
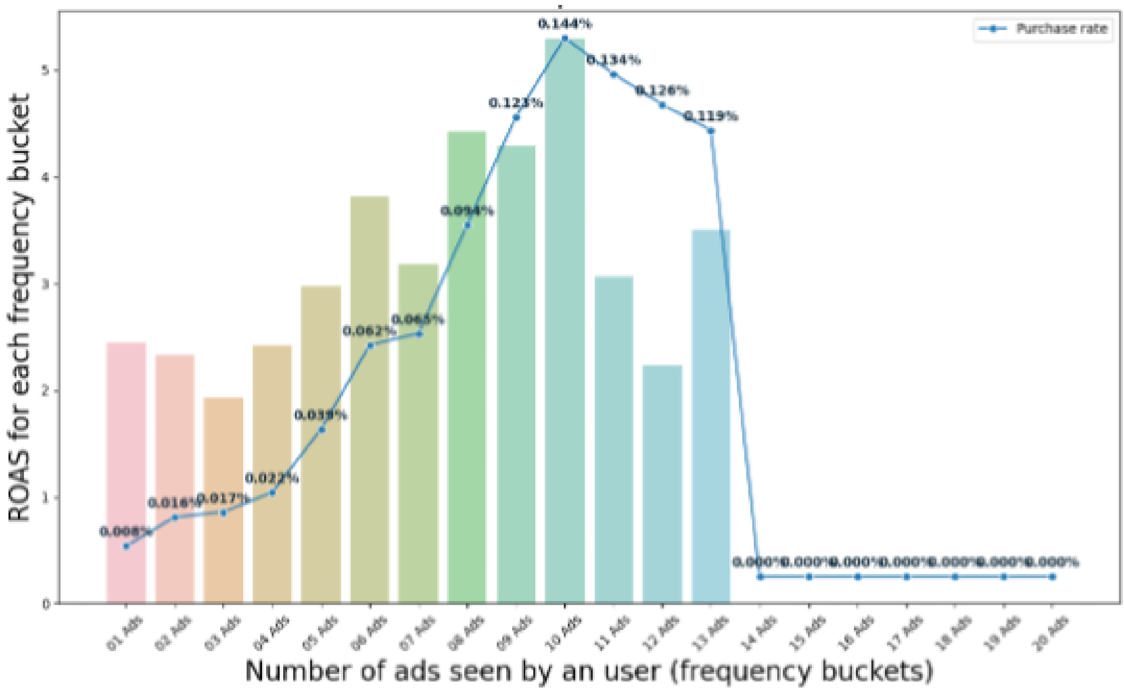


Figure 10 – Frequency cap for Braun shavers in November 2021 – Purchase rate vs ROAS per frequency bucket (groups of 5)



5.2.3.1 Media vehicles

In this part of the analysis, the goal was to find out how the different types of advertisements work together and separately. The exposure groups, or media types, used for comparison on this were the Sponsored Advertisements (Search Ads), Display Advertisements, and Share of Voice packages. The objective was to know how these three types of advertisement perform during peak periods (Black Friday and Prime Day), and on off-peak periods, to find out if their use is worth the spent. The analysis was done by querying on AMC for all of the selected months wanted on the analysis. As a first analysis, only Sponsored Advertisements and Display Advertisements were compared, leaving Share of Voice packages for future analysis. Each comparison was done by the brand, and the selected brands were Gillette and Braun. The variables selected in the queries were:

- Exposure group: Display, Search, SoV, Display vs Purchase, Display vs SoV, SoV vs Seach, and Search vs Display vs SoV. This means that if a user fits into the exposure group Search vs Display vs SoV, they were exposed to all three types of advertisements before converting.

- Total impressions: Total impressions a brand has received on the selected period and grouped by exposure group;

- Users that have purchased: the number of unique users that have purchased any products from the selected brand and related to an exposure group;

- Purchase amount: Sum value of all purchases done for the selected brand, time frame, and exposure group, in Euros;

- Total cost: How much was spent in advertising for that specific time frame, brand, and exposure group;

- Unique reach: Number of unique consumers that have seen an advertisement for that specific brand, time frame, and exposure group.

After recovering all these values for the selected months, the KPIs used for success measurement could be calculated, which were ROAS and purchase rate, and some other values that could be interesting for comparison, such as the percentage of total impressions and percentage of total sales an exposure group has received for that period. The KPIs were calculated per brand and exposure group. With all the values calculated, a chart was plotted to show for each brand the corresponding ROAS for the selected month and exposure group. The results are shown in Figures 11 and 12 from August 2021 until June 2022.

For Braun, it is possible to see that for most months, the use of both Display and Search advertisements is beneficial in terms of ROAS. Display advertisements always have lower values of ROAS when used alone, since the number of impressions is always high because it is used mostly for reach. For Gillette, the use of both Search and Display advertisements is beneficial only in specific months. There are some months that Display advertisements are not even used, which are the months where ROAS is zero on the chart for Display.

Figure 11 – ROAS for each exposure group by month – Braun

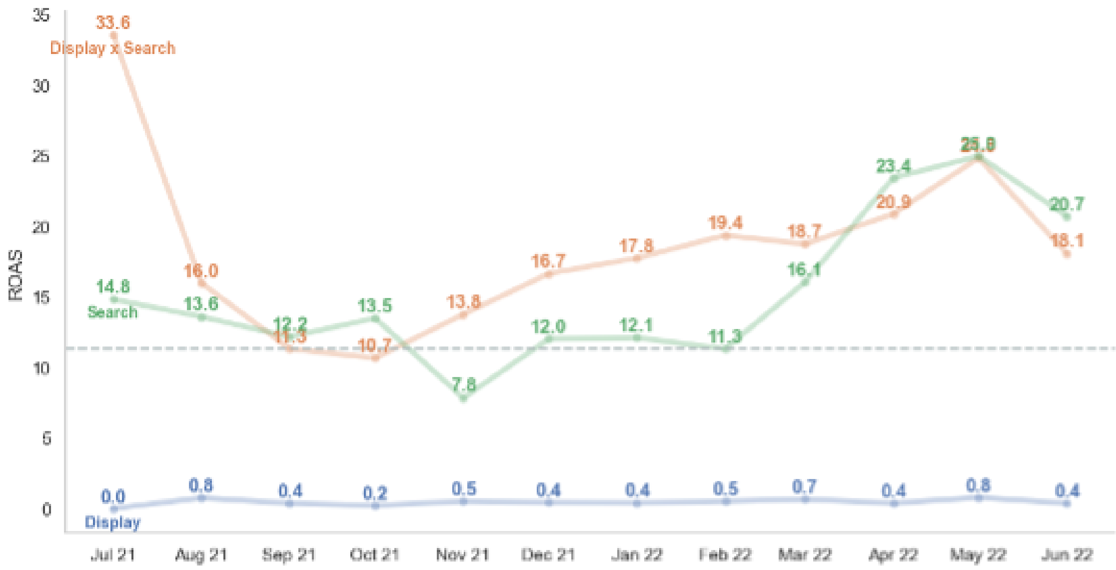
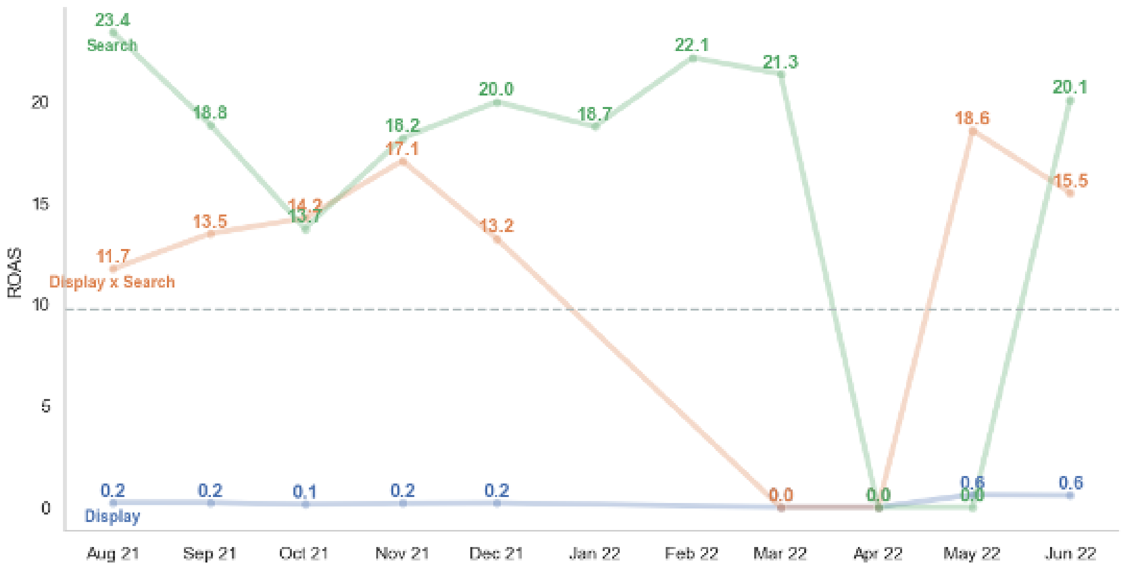


Figure 12 – ROAS for each exposure group by month – Gillette



To get a general picture of the use of both types of advertisements, the comparison values were summed for all the months in the analysis, and the results were put in a table for each brand. For both Braun and Gillette, most of the impressions come from Display advertisements, but the ROAS and purchase rate for this exposure group is very low. For Braun, as shown in Table 22, using the combination of both types of advertisements gives the best results. For Gillette, observing the results in Table 23, the purchase rate is better when overlapping the advertisements, but the ROAS is better when using only Search ads.

As a conclusion, it is noticeable that for Braun, the use of both advertisements is important. For Gillette, the recommendation given to the agencies that control Amazon advertisements was that the overlap of media vehicles should be done, but not in all months. They should focus on using both of them during peak periods. For all tables, M stands for thousand and MM for million.

Table 22 – KPIs for Braun split by exposure groups from August 2021 until June 2022

Exposure Group	Display	Display x Search	Search
Total Impressions	107.3MM	70.5MM	96.9MM
Users that Purchased	2.3M	38.3M	51.4M
Purchase Amount	149.5M	6.0MM	5.5MM
Total Cost	325.1M	364.6M	415.5M
Unique Reach	33.6MM	2.3MM	5.9MM
Total Impressions %	39.04%	25.67%	35.28%
Purchase Amount %	1.28%	51.73%	46.98%
ROAS	0.46	16.51	13.16
Purchase Rate	0.007%	1.679%	0.875%

Table 23 – KPIs for Gillette split by exposure groups from August 2021 until June 2022

Exposure Group	Display	Display x Search	Search
Total Impressions	34.0MM	13.8MM	35.8MM
Users that Purchased	1.5M	13.5M	41.9M
Purchase Amount	26.7M	488.7M	957.5M
Total Cost	106.7M	34.4M	56.3M
Unique Reach	14.6MM	586.7M	3.6MM
Total Impressions %	40.68%	16.49%	42.84%
Purchase Amount %	1.81%	33.18%	65.01%
ROAS	0.25	14.22	17
Purchase Rate	0.010%	2.297%	1.173%

As a second part of this analysis, there is a need to understand how adding SoV as a new exposure group would interfere with sales performance. For that, knowing which were the periods when SoV packages were bought by Procter & Gamble for the selected brands is needed since this type of advertisement is seasonal and is bought normally in peak periods. The only brand that was analyzed that used this type of advertisement is Braun, so two peak months were selected to add this new variable, July 2022 (Prime Day) and October 2022 (second Prime Day). New queries were executed on AMC adding the new exposure group, one for each of the months. The results are shown in Tables 24 and 25, and it is possible to conclude that the use of SoV packages is beneficial for peak periods, having higher ROAS and Purchase Rates when combined with Search ads. It also shows good results when there is a triple overlap between the three exposure groups.

Table 24 – KPIs for Braun split by exposure groups in July 2022

Exposure Group	Total Impressions	Purchase Amount	Total Cost	ROAS	Purchase Rate
Display	12.5MM	20.7M	38.9M	0.53	0.005%
Display x Search	189.8M	548.1M	39.7M	13.80	16.178%
Display x Search x SoV	132.2M	262.4M	27.9M	9.40	9.900%
Display x SoV	3.6MM	16.8M	31.5M	0.53	0.011%
Search	43.8M	522.9M	41.2M	12.69	18.591%
Search x SoV	3.7M	84.5M	5.3M	15.92	22.721%
SoV	213	0.0	25.0M	0.00	0.003%

Table 25 – KPIs for Braun split by exposure groups in October 2022

Exposure Group	Total Impressions	Purchase Amount	Total Cost	ROAS	Purchase Rate
Display	12.8MM	0.0	41.3M	0.00	0.002%
Display x Search	80.5M	195.4M	20.2M	9.67	12.730%
Display x Search x SoV	62.6M	122.8M	14.5M	8.50	8.251%
Display x SoV	4.3MM	0.0	41.1M	0.00	0.006%
Search	20.6M	199.5M	22.0M	9.07	15.621%
Search x SoV	1.3M	19.3M	2.0M	9.50	15.977%
SoV	50	3.8M	20.2M	0.19	0.001%

5.2.3.2 Campaign paths

The analysis done on this part concerned the different campaign paths a user can follow until a purchase. This means that the goal is to understand what types of advertisement a user has seen (Search or Display), in what order, and how many times until they purchase a product, all related to a specific marketing campaign. The values used for comparison in this analysis were:

- Starts with: The path followed by the consumer until the purchase starts with this type of advertisement. The options are Display or Search ads;
- Ends with: The path followed by the consumer until the purchase ends with this type of advertisement. The options are Display or Search ads;
- Path occurrences: How many times the path to purchase started and ended with the selected start and end advertisements;
- Purchase rate: Purchase rate related to that type of path;
- ROAS: ROAS related to that type of path;
- Average path length: the average number of advertisements seen by the consumer in that type of path;
- Path occurrence %: Percentage of total occurrences for that type of path;

As shown in Tables 26 and 27, the paths that have the most occurrences are the ones that start and end with display advertisements, but these are the ones that have the worst purchase rate. The path in which the user has the best results is the one that starts with Search and ends with Display advertisements, in terms of ROAS and purchase rate.

Table 26 – Path to purchase for Braun campaigns

Starts With	Display	Display	Search	Search
Ends With	Display	Search	Display	Search
Path Occurrence	38MM	48M	134M	245M
ROAS	7.31	8.24	10.69	4.95
Purchase Rate	0.016%	1.114%	1.808%	4.749%
Average Path Length	1.4	3.1	3.1	1.3
Path Occurrence %	98.89%	0.13%	0.35%	0.64%

Table 27 – Path to purchase for Gillette campaigns

Starts with	Display	Display	Search	Search
Ends with	Display	Search	Display	Search
Path occurrence	12MM	5.8M	19M	85M
ROAS	2.76	15.36	19.31	16.44
Purchase rate	0.021%	4.079%	5.522%	17.5%
Average path length	1	2.5	2.4	1.4
Path occurrence %	99.11%	0.05%	0.15%	0.69%

5.2.3.3 Time lag for purchase

In this analysis, the objective was to find out how the consumer interacts with the advertisements over the next few days after seeing them. Periods of 15 days were analyzed, meaning that after a user saw an advertisement, their actions on the Amazon website during those 15 days were examined. The three main consumer actions were:

- Purchase: The user has purchased a product after seeing an advertisement related to it;
- Add to shopping cart: The user has added a product to Amazon's shopping cart after seeing an advertisement related to it;
- See detail page: The user has seen the product's detail page on Amazon after seeing an advertisement related to it.

This analysis was done per brand, and the results of 12 months of analysis were queried, and divided into periods of 15 days. For the results, the chart plotted shows: the period passed after the user has seen an advertisement, starting from 10 minutes until 15 days (360 hours); the percentage of users that have taken an action after the respective time; the actions taken by the user on that time; the percentage of total purchases done up until the moment in 15 days. The brands analyzed were Gillette and Braun, from June 2021 until April 2022. As shown in Figures 13 and 14, both Gillette and Braun have similar behaviors, with most consumers converting up to two hours after seeing an advertisement related to the respective brand. It is also possible to notice that two days after seeing an advertisement, a peak of conversions happens, and more than 50% of the sales are done up until this day.

Figure 13 – Chart showing time lag until conversion event after seeing an advertisement for Braun

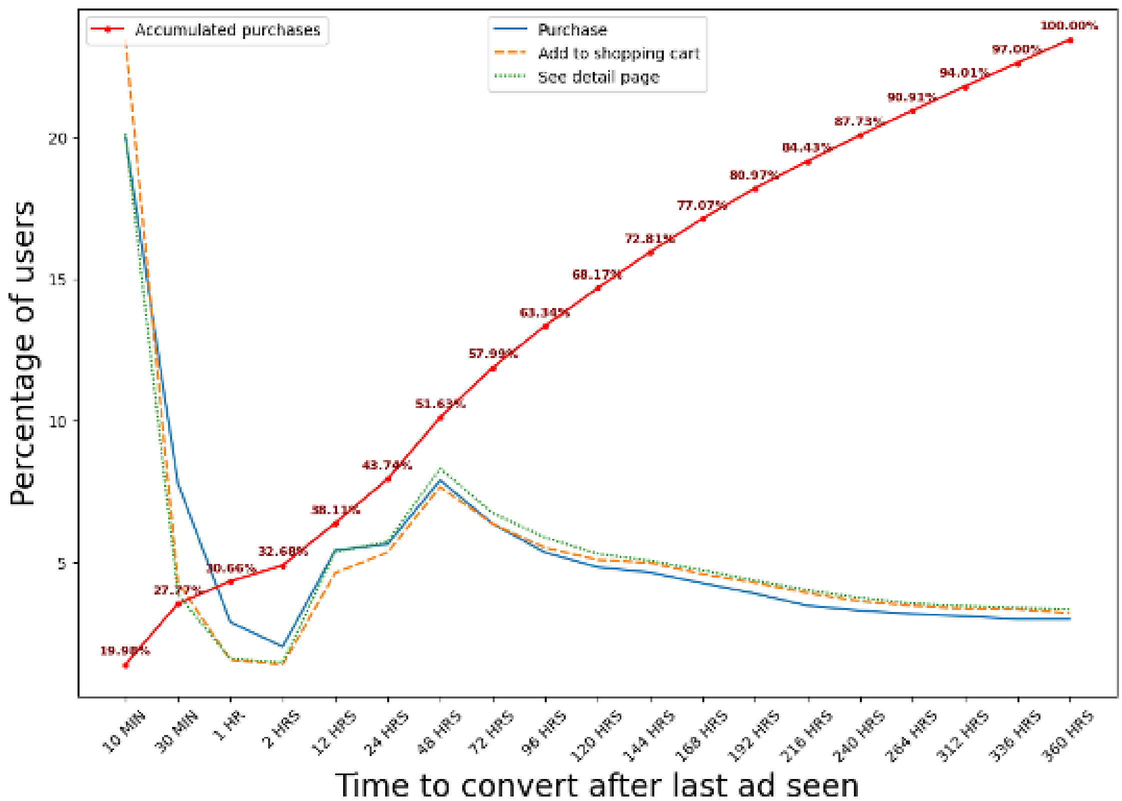
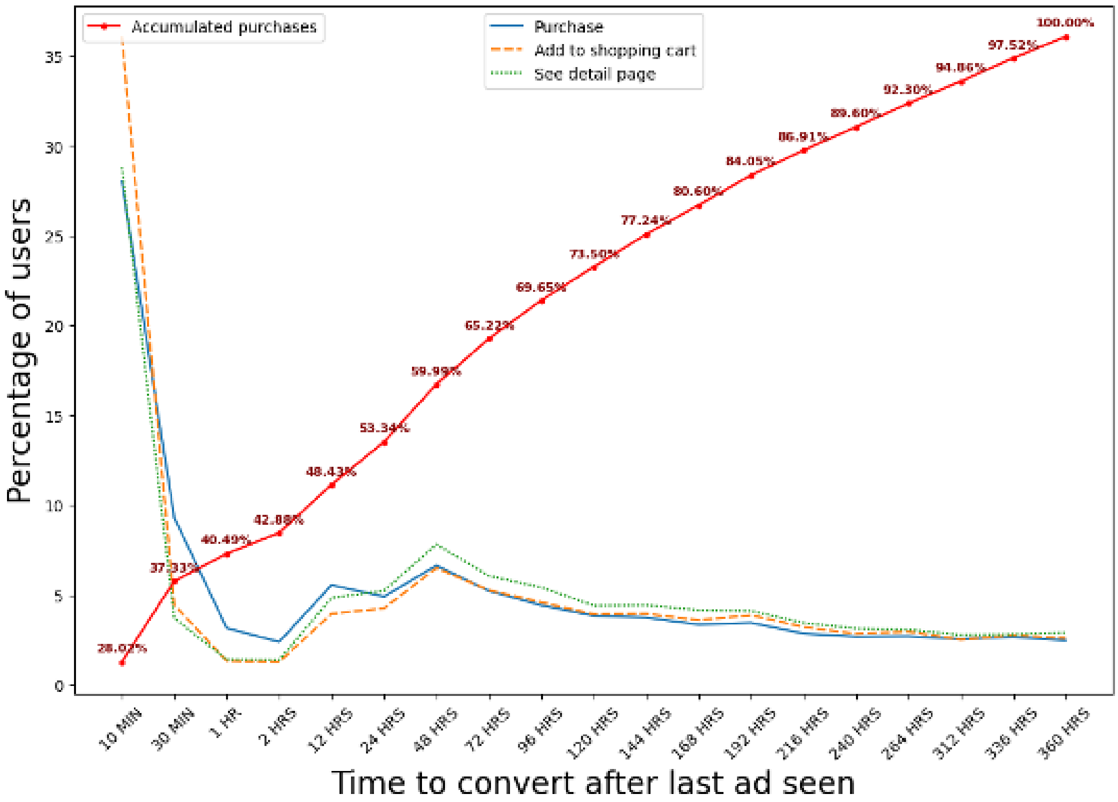


Figure 14 – Chart showing time lag until conversion event after seeing an advertisement for Gillette



5.2.3.4 Consumer audiences

This last analysis was regarding the consumer audiences, also called segments, available on Amazon Marketing Cloud. A consumer audience is a group of consumers that have similar purchasing patterns and that fit into a common category. This is used for directing the correct types of advertisements to the correct consumers, depending on the consumer's interests, and it shows to have better performance than displaying the same advertisements to all consumers without discretion. With this analysis, the best-performing audiences available on Amazon for each brand, Gillette and Braun, needed to be found. For each brand, the total value of product sales, in euros, was displayed and then the ROAS, the purchase rate, and the percentage of sales that were done by new-to-brand (NTB) users were calculated. The definition of a new-to-brand user in Amazon is a user that has not purchased any products from a respective brand in the last 12 months. The period of the analysis was from August 2021 until June 2022.

Braun's best-performing segments are shown in Table 28, and Gillette's in Table 29. As is shown, the two first segments are very broad and general, and the ones that are important for the brands are the others since they have more specific descriptions. It is also noticeable that Braun has more consumers who are new to the brand, in percentage, than

Gillette, which means that Gillette's consumers are more prone to repurchasing products.

Table 28 – Best-performing segments for Braun from August 2021 until June 2022

Segment Name	**Total Product Sales**	**Purchase Rate**	**Percentage of NTB sales**	**ROAS**
Demo - clients francais	2.98MM	0.06%	76%	3.6
Demo - Clients	2.74MM	0.06%	76%	3.7
LS - Acheteurs d'articles peu couteux	2.61MM	0.06%	76%	3.6
LS - Produits d'Entertainment	2.56MM	0.06%	76%	3.5
LS - Beauté (Santé et Soins du corps)	2.58MM	0.07%	75%	3.5

Table 29 – Best-performing segments for Gillette from August 2021 until June 2022

Segment Name	**Total Product Sales**	**Purchase Rate**	**Percentage of NTB sales**	**ROAS**
Demo - clients francais	302M	0.11%	55%	2.3
Demo - Clients	271M	0.12%	54%	2.3
LS - Acheteurs d'articles peu couteux	265M	0.13%	53%	2.3
LS - Produits d'Entertainment	263M	0.13%	53%	2.2
LS - Beauté (Santé et Soins du corps)	263M	0.16%	52%	2.2

The next step in this analysis was to find out which age groups perform the best for each brand in terms of sales value. This is done by selecting the segments that exist on AMC that split the consumer into age groups. The types of devices (Personal Computer, Phone, Tablet) that perform best for each age group were also analyzed. The results are shown in Table 30, for Braun, and Table 31, for Gillette. As it is shown, the age groups that consume the most for both brands are the ages from 35 to 44 and 45 to 54 years old. The device used the most for sales is a Phone for Braun and a Personal Computer (PC) for Gillette. As mentioned before, Gillette has a bigger partition of consumers who tend to repurchase their products. A final finding is that, even though Tablets have the least sales attributed to them, they are the ones that have the highest ROAS in most of the groups.

Table 30 – Best-performing age groups, divided by device, for Braun from August 2021 until June 2022

Segment Name	**Device Type**	**Total Product Sales**	**Purchase Rate**	**Percentage of NTB sales**	**ROAS**
Demo - Age 18-24	PC	65M	0.07%	77%	3.6
	Phone	90M	0.05%	81%	3.6
	Tablet	1.7M	0.08%	67%	5.2
Demo - Age 25-34	PC	142M	0.06%	78%	3.5
	Phone	250M	0.05%	81%	3.5
	Tablet	1.2M	0.06%	77%	3.4
Demo - Age 35-44	PC	212M	0.06%	76%	3.5
	Phone	300M	0.05%	81%	3.6
	Tablet	3.3M	0.04%	82%	4.4
Demo - Age 45-54	PC	221M	0.07%	74%	3.9
	Phone	185M	0.06%	76%	3.8
	Tablet	2.9M	0.05%	86%	4
Demo - Age 55-64	PC	110M	0.07%	69%	3.9
	Phone	70M	0.06%	75%	3.5
	Tablet	2.5M	0.06%	84%	5.8

Table 31 – Best-performing age groups, divided by device, for Gillette from August 2021 until June 2022

Segment Name	**Device Type**	**Total Product Sales**	**Purchase Rate**	**Percentage of NTB sales**	**ROAS**
Demo - Age 18-24	PC	12M	0.15%	53%	1.9
	Phone	3.7M	0.11%	64%	1.8
	Tablet	71	0.17%	22%	1.4
Demo - Age 25-34	PC	23M	0.16%	56%	2.1
	Phone	11M	0.12%	63%	2.1
	Tablet	28	0.11%	100%	0.7
Demo - Age 35-44	PC	37M	0.15%	55%	2.2
	Phone	15M	0.11%	58%	2.1
	Tablet	136	0.11%	13%	2.8
Demo - Age 45-54	PC	43M	0.16%	51%	2.4
	Phone	10M	0.13%	59%	2
	Tablet	141	0.12%	63%	4.2
Demo - Age 55-64	PC	23M	0.17%	47%	2.5
	Phone	4.7M	0.15%	57%	2
	Tablet	134	0.07%	67%	2.8

6 Final considerations

6.1 Conclusions

In conclusion, this paper underscores the significance of first-party data within the dynamic landscape of digital marketing, particularly against the backdrop of the stringent General Data Protection Regulation (GDPR) in Europe. The multiple analysis conducted in this paper reveals a nuanced understanding of the role, challenges, and transformative potential associated with the utilization of first-party data.

The real-world case studies presented provide tangible evidence of how businesses strategically leverage first-party data to propel their digital marketing objectives. These cases show the adaptability and resilience required in response to the regulatory frameworks, especially GDPR, shaping the contemporary data-driven marketing paradigm.

Furthermore, the research explores the intricacies of retrieving and analyzing first-party data from untapped internal sources within companies. This approach not only unlocks so far unexplored insights but also establishes the foundation for a more robust and personalized marketing strategy. The emphasis on using first-party data sourced internally reflects a departure from traditional reliance on second and third-party data, presenting a paradigm shift in the industry.

GDPR's implications are evident in its role as a trigger for reevaluating data collection and usage practices in the digital marketing sphere. The paper articulates the consequential downfall of second and third-party data, as businesses change their strategies to align with GDPR compliance. This shift shows the important role of regulations in reshaping the digital marketing landscape and fostering a more ethical and privacy-centric approach.

Moreover, this research brings attention to the global diversity of data protection regulations, exemplified by the contrast between GDPR in Europe and LGPD in Brazil. The variations in these regulations underscore the need for a nuanced and adaptable approach in the global digital ecosystem.

An important dimension explored in this paper is the significance of GDPR not just as a regulatory framework but as a potent tool giving users the power to reclaim control over their personal information. By limiting the overuse of personal data by major corporations in digital marketing, GDPR contributes to a more equitable and user-centric online experience.

In essence, the summary of insights from this research points toward a future where the judicious use of first-party data, aligned with evolving regulations, is integral to achieving business goals, gathering consumer trust, and navigating the complexities of the contemporary digital marketing landscape.

6.2 Contributions

This paper was based on the End of Studies paper done by the student at the University of Grenoble INP during their double degree exchange program in France. It was modified and completed to serve as the End of Studies project at the Universidade de São Paulo. All of the work is authored by the student, their supervisor, and the company Procter & Gamble.

6.3 Continuity prospectives

As continuity prospects, future papers based on this research could involve a deeper exploration of evolving data protection regulations worldwide in the landscape of digital marketing. Further investigations could go into the specific adaptations and strategies employed by companies to navigate the stringent GDPR and similar data protection laws globally. Examining emerging technologies and methodologies for secure and ethical data collection, especially in the context of first-party data, would contribute to the ongoing discourse. Additionally, the paper could be extended to explore the long-term implications of the decline of second and third-party data in digital marketing. It could include an analysis that investigates how businesses are innovating and adjusting their marketing strategies to remain effective within the constraints set by data protection laws.

Bibliography

AMC. Amazon marketing cloud - discover advertising insights. *Amazon Marketing Cloud*, n.d. Accessed on: Sep. 20, 2023. Cited on page 34.

BRINKER, S.; HELLER, J. Marketing technology: What it is and how it should work. *McKinsey & Company*, 2021. Accessed on: Sep. 3, 2022. Cited on page 25.

CISNA, J.; PENG, W. Http cookies – a promising technology. *Online Information Review*, v. 24, n. 2, p. 150–153, 2000. Cited on page 22.

HERMAN, P. The cookiepocalypse: Why first-party data is going to matter to your newsroom. *Reuters Institute for the Study of Journalism*, v. 1, n. 1, p. 2–8, 2022. Cited on page 22.

MPF. O que é a lgpd? — lei geral de proteção de dados. *Ministério Público Federal*, n.d. Accessed on: Sep. 20, 2023. Cited on page 23.

ONAUDIENCE. 1st party data vs 2nd party data vs 3rd party data. *OnAudience*, 2021. Accessed on: Oct. 21, 2022. Cited on page 21.

ORACLE. What is data activation? *Oracle*, n.d. Accessed on: Nov. 10, 2023. Cited on page 19.

PETERS, R.; SIKORSKI, R. Cookie monster? *Science*, v. 278, p. 1486-7, 1997. Cited on page 22.

SIGNAL. What is first-party data? *Signal*, 2021. Accessed on: Oct. 21, 2022. Cited on page 21.

STEWART, D. S. S. What is martech? a guide to marketing technology. *Amazon*, 2012. Accessed on: Sep. 12, 2022. Cited on page 25.

WOLFORD, B. What is gdpr, the eu's new data protection law? *GDPR.eu*, n.d. Accessed on: Nov. 10, 2023. Cited on page 22.