NOVA IMS Information Management School

MDSAA

Master Degree Program in

Data Science and Advanced Analytics

Protecting legacy consumer behavior and traffic analytics data with a Universal Analytics transition

A Business Intelligence Solution for INCANTO

Anastasiia Tagiltseva

Project Work

presented as partial requirement for obtaining the Master Degree Program in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

A BUSINESS INTELLIGENCE SOLUTION FOR INCANTO: PROTECTING LEGACY CONSUMER BEHAVIOR AND TRAFFIC ANALYTICS DATA WITH A UNIVERSAL ANALYTICS TRANSITION

by

Anastasiia Tagiltseva

Project Work presented as partial requirement for obtaining the Master's degree in Advanced Analytics, with a Specialization in Business Analytics

Supervisor: Miguel de Castro Ferreira Simões Neto

Co-Supervisor: João Bruno Morais de Sousa Jardim

February 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledge the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Anastasiia Tagiltseva
Lisbon, February 2024

ABSTRACT

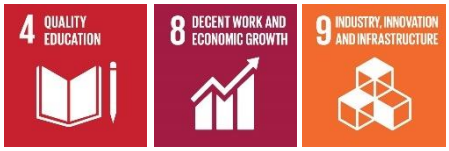
Given the abandonment of Universal Analytics and subsequent deletion of historical data from mid-2024, there is an urgent need to migrate and preserve historical data for future use in the context of utilizing detailed time series and digital marketing data that is crucial to measure the effectiveness of digital content and optimize multichannel marketing campaigns. In digital marketing, removing historical data presents a significant challenge to retaining legacy consumer behavior and traffic data. Recognizing a gap in knowledge about effective lossless data migration, this study uses a scientific design research methodology to develop a systematic approach by exploring a business intelligence solution for INCANTO, an online lingerie retailer. In order to analyze and efficiently move from Universal Analytics to a more sustainable analytics system without losing valuable historical data, the marketing communications of INCANTO and conducting advertising campaigns as efficiently as possible were developed and implemented Business Intelligence solution for the INCANTO online store.

The project identifies vital consumer trends, behaviors, and traffic patterns, helping key users make decisions to optimize INCANTO's marketing strategy. The implementation demonstrates Power BI's ability to work with UA data, ensuring the continuity of data-driven decision-making processes beyond the UA era. The developed model provides valuable guidance for companies going through similar transitions, emphasizing the importance of preserving legacy data to understand and analyze future data with a historical foundation.

KEYWORDS

Web Analytics; Data visualization; Business Intelligence; Digital Marketing; Google Analytics; Universal Analytics; Data-Driven Marketing

SUSTAINABLE DEVELOPMENT GOALS (SDG)



INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

BI	Business Intelligence
CRM	Customer Relationship Management
CTR	Click-through rate
CPC	Cost per click
DAX	Data Analysis Expressions
DSRM	Design Research Methodology
GA	Google Analytics
GATC	Google Analytics Tracking Code
GIS	Geographic Information System
ROI	Return on investment
SEO	Search Engine Optimization
UA	Universal Analytics
URL	Uniform Resource Locator
UTM	Urchin Tracking Module
UX/UI	User Experience/User Interface

1. INTRODUCTION

1.1. CONTEXT

Due to the growth in internet traffic and the continuous advancement of digital technology, digital marketing has emerged as the predominant marketing strategy. 31st edition of The CMO survey shows that “Digital marketing transformations show progress post-pandemic as companies shift from the nascent phase (from 27.1% to 8.8%) and into the emerging (from 52% to 54.2%), integrated (from 13.9% to 24.2%) and institutionalized (7% to 12.8%) phases of the transformation journey.” (The CMO Survey, 2023).

W3Techs, a web technology survey firm, reports that Google Analytics is currently the most widely used web analytics platform on the internet (W3Techs, 2024). This data indicates that Google Analytics is the preferred choice for tracking website traffic and analyzing user behavior by website owners and marketers alike. Introduced in the fall of 2012, Universal Analytics – a version of Google Analytics - offered new tracking codes for websites and features that could more accurately measure user behavior by that time (Google, 2012).

In March 2022, Google announced that its flagship Universal Analytics (UA) platform would stop processing new website sessions on July 1, 2023 (Ketchum, 2022). Even though Google is offering users of UA the option to transition to their newer Google Analytics 4 (GA4) platform, officially launched in October 2020, the sunsetting of the UA platform is causing great consternation in the web analytics and marketing analytics communities. Millions of businesses rely on UA to perform detailed analyses of their websites, providing insights into their website visitors' makeup and engagement patterns and the performance of their digital marketing campaigns. (McGuirk, 2023)

1.2. MOTIVATION

The motivation for this study is related to the removal of Universal Analytics historical data permanently according to Google's plans. INCANTO, an online lingerie shop, faced the challenge of migrating to GA4 without losing this valuable historical data that is essential for trend analysis, understanding customer behavior and strategic planning in decision making. Moving to GA4, with its emphasis on user privacy and event-driven cross-platform tracking, presents an opportunity to expand INCANTO's analytical capabilities. However, it requires a strategy to adapt to these changes while maintaining data integrity. The purpose of this research is to develop a practical and effective methodology for such a transition so that INCANTO can take advantage of the enhanced marketing analytics capabilities provided by GA4 without compromising the foundation of historical data.

1.3. CONTRIBUTIONS

This study aims to develop a systematic approach for preserving historical data using Power BI for INCANTO. INCANTO is a company specializing in the fashion segment – a very competitive market with many players. INCANTO products are focused primarily on underwear, home wear, and swimwear, which entails seasonality and a predominantly female target audience aged 20-45. There are both online and offline points of sale. Starting in 2019, the company launched an online store, which showed significant growth, supported by the impact of increased online activity during the pandemic. To attract

and retain traffic, the company actively uses Internet promotion tools. The study aims to develop a systematic approach that utilizes a Business Intelligence (BI) solution. Power BI[1] was chosen for its robust data integration, analysis, and visualization capabilities. It allows for the efficient aggregation of data from various sources and enables the creation of comprehensive reports and dashboards. By using Power BI, INCANTO can not only maintain continuity in accessing historical data but also enhance their ability to derive actionable insights through advanced data analytics, thus effectively bridging the gap created by the transition to GA4.

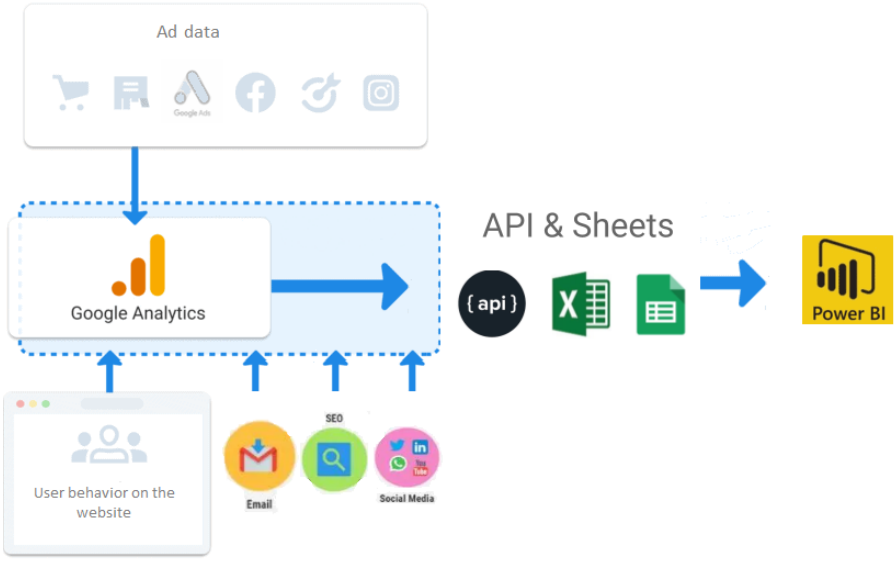


Figure 1.1 – Overview of INCANTO marketing data pipeline for website traffic data.

This project contributes to the field of digital analytics by providing a methodology for preserving historical UA data, adapted to the needs of online shops such as INCANTO. It includes a step-by-step guide on migrating data using Power BI to preserve important insight and historical patterns and trends of website performance, thus ensuring continuity of analytics and decision-making processes. It proposes the collection of data from advertising services, user behavior on the website, email, SEO, and social media data, and then transferring that data to Power BI to build the reports and dynamic indicators necessary for INCANTO's management based on the data obtained. Figure 1.1 provides an overview of this pipeline. This study will not only assist INCANTO in its analytics transition but will also serve as a valuable resource for other companies facing this tipping point, facilitating a seamless transition to a Google's new analytics platform, Google Analytics 4.

1.4. Thesis Structure

The rest of the project is organized as follows: **Literature Review** reviews existing research on digital analytics, the evolution and importance of BI in enhancing business decision-making, particularly in the digital marketing and e-commerce sectors. **Methodology** defines the Design Science Research Methodology approach adopted for the needs of INCANTO. This section also presents the implemented dimensional model for INCANTO website traffic data. **Results and Discussion** presents the results of the BI solution, including interpretation, discussion of the practical application of the solution for INCANTO and similar businesses, and comparison of the results with existing literature.

[1] https://powerbi.com/

This chapter also evaluates the benefits and limitations of using the solution in the context of digital analytics. **Conclusion** summarizes key finding in this project, **and Limitations and Recommendations for Future Work** discusses the limitations of the work and possible research directions that can be pursued to apply the resulting solution for in-depth digital marketing analytics.

2. LITERATURE REVIEW

in 1989, Howard Dresner promoted and extended the term Business Intelligence (BI) as a “set of concepts and methods to improve business decision making by using fact-based support systems.” (Power, 2007).

Dresner's terminology continues to hold relevance and remain prevalent within academic discourse, despite variations in its interpretation that may arise as a consequence of differing authorial perspectives. But two main things remain the same: 1. Business intelligence is defined firstly as a decision-making process. 2. Business intelligence is supported by the integration and analysis of an organization’s data resources (Tavera Romero, 2021).

Data from digital marketers translate further to various levels of the company and BI makes it more comprehensive by breaking the matter into charts, graphs which can be understood by both technical teams and nontechnical managerial strata in the workforce. Fundamentally all challenges faced in the digital marketing environment can be resolved using BI tools as it possesses the capability to be in tune with the dynamic changes experienced by digital media and can provide real-time solutions to reach a final, more promising output. (Bhosale, 2020)

According to Maggy (2020), BI it is widely applied in e-commerce because traditional web analytics techniques usually implemented in businesses aiming to digitalize their working methods are becoming outdated and ineffective. These activities are typically performed manually and are highly time-consuming. Google Analytics is an online tracking tool that is implemented alongside the website by adding a “page tag.” It is a tracking code of Google (GATC, Google Analytics Tracking Code), a block of JavaScript code that a user who wants to track its website activity implements in all the pages of its web domain. This code interacts directly with the Google server, allowing the collection of the browsing data of users, which is elaborated, developed, and shown within the platform afterward. The GACT, other than transferring the data to Google’s server, sets the first-party cookie in every visitor’s computer to memorize anonymous information such as the type of visitor (if it’s new or returning), the duration of the visit, and the source type. In this way, the tool can collect information regarding the user behavior on the website to provide the website owner with some insights regarding its performance. Keeping this in mind, Google Analytics uses sampling and aggregating data techniques to make reports faster and more efficient for the final user, sometimes by compromising the integrity of the information. The result of this approach is that Google Analytics allows the exportation of data from the user interface with the lowest granularity level of information: data can be extracted per group of users based on age, geographical information, source of the user visits between other dimensions.

A study conducted by Azizi (2023) explores the critical aspects of technology adoption in the e-commerce industry, focusing on user characteristics, customization options, financial constraints, software complexity, and data quality. Through semi-structured interviews with IranServer managers and marketing specialists, the study aims to provide information applicable to Iran and the broader e-commerce sector. The findings contribute to a better understanding of marketing dashboard adoption, offering valuable insights for organizations seeking to implement such tools effectively.

For instance, to demonstrate how data can be presented in a clear way and speed up decision-makers' ability to make decisions Kalboussi, et al. (2023) proposed approach based on building Power BI dashboard. It acquired data from multiple traffic sources including Facebook, Instagram, e-mails, and Twitter to provide general track information. Power BI dashboard presented an overview of the most important metrics and employed a clustering algorithm to get better ideas about data and its structure.

Prikrylova (2016) explored the possible uses of BI in digital marketing to automate decision-making processes and minimize the risk of judgmental errors. It was said that when utilized by skilled professionals, BI tools can lead to superior digital marketing strategies. One of the primary functions of marketing analytics involves the analysis of customer sales, necessitating a comprehensive approach that encompasses forecasting, segmentation, and the collaborative analysis of Customer Relationship Management (CRM) data and diverse data sources such as the Internet, mobile platforms, and Geographic Information Systems (GIS). This comprehensive approach to analytics is crucial for understanding the effectiveness of marketing campaigns (Talaoui, 2020).

Moreover, BI extends the perspective of digital marketing analysts beyond just the product in focus (Kursan, 2010). It ensures that the solutions devised are seamlessly integrated with the broader objectives of the entire organization. The insights garnered by digital marketers are disseminated across various organizational levels, with BI tools enhancing this information through clear and accessible visualizations like tables, charts, and graphs (Pucka, 2020). Such visual tools make the data comprehensible to technical teams and non-technical management, facilitating a universal understanding across the company (Ivanov, 2022).

Study made by Ye (2023) shows that designing and implementing a solid BI framework is essential to overcoming the struggle of e-commerce companies to use the enormous amounts of data collected from diverse sources in the big data era. At the same time BI tools offer the flexibility to analyze various metrics such as the duration spent by potential customers in a marketing environment, their responsiveness to specific ads, click-through rates (CTR), and the cost per click (CPC). This level of detailed analysis is instrumental in shaping effective marketing strategies (Bhosale, 2020).

Table 2.1 summarizes related works on BI-solutions in e-commerce and digital marketing with their corresponding to the result of the study.

Table 2.1 – Summary of Related Works on BI-solutions in e-commerce and digital marketing

Author(s)	Year	Title	Result
Azizi	2023	IranServer´s Digital Marketing Transformation: A Study of the Adoption of Marketing Dashboards	The study concluded that implementing marketing dashboards in IranServer significantly improved decision-making processes by integrating data-driven information, thereby increasing overall marketing effectiveness.
Maggi	2020	Enhanced Web Analytics for Health Insurance	Solution for enhancing the web analytics tasks and supporting the improve of the online user digital experience

Kalboussi et al.	2023	Machine Learning for Decision Making in the Improvement of Digital Press Services	Dashboard that presents data in a clear way and speed up decision-makers' ability to make decisions
Prikrylova	2016	Business intelligence models for capturing and analysis of enterprise marketing data	Framework supporting effective marketing decisions utilizing the advantages of several BI methods and integrating data from different sources of semi-structured marketing data with a time-related component
Talaoui	2020	Of BI research: a tale of two communities	The theoretical underpinnings of the BI literature and shedding light upon the evolution, the contributions, and the research gaps for each of the six clusters composing the BI body of knowledge
Kursan	2010	Business Intelligence: The Role of the Internet in Marketing Research and Business Decision-Making	The paper examines the role of the Internet in marketing research and its implications on the business decision-making processes
Ivanov	2022	The Impact of Business Intelligence Implementation on Digital Marketing Analysts' Effectiveness	The author's research focuses on the relationship between digital marketing and aspects of business intelligence (BI), which allows for more efficient processing of data
Bhosale	2020	Role of business intelligence in digital marketing	The paper replicates upon Digital marketing, the requirement for using Business Intelligence in Digital Marketing, the framework for Business Intelligence performs and Business Intelligence from Web Analytics.
Ye	2023	Unleashing the Power of Big Data: Designing a Robust Business Intelligence Framework for E-commerce Data Analytics	Comprehensive BI platform designed especially for e-commerce research to unleash the power of big data

After reviewing the literature, it has been found that Business Intelligence (BI) implementation has been widely used in e-commerce, digital marketing, and web analytics. However, there needs to be more practical research on using BI solutions for storing and utilizing data from Universal Analytics. Therefore, this paper aims to demonstrate a BI solution focusing on data warehousing and data visualization for INCANTO.

3. METHODOLOGY

Design Science Research Methodology (DSRM) is an approach utilized primarily in information systems and computer science research to address complex problems through the creation and evaluation of innovative artifacts. DSRM focuses on the development of novel solutions or artifacts to address practical problems encountered in various domains.

This methodology emphasizes the creation of new knowledge in the form of designed artifacts, such as software systems, models, processes, methodologies, or frameworks. These artifacts are not just academic theories but tangible creations that are rigorously developed and evaluated to solve real-world problems.

DSRM typically involves several iterative phases (Peffers et al., 2007):

1. Problem Identification and motivation: Clearly defining and understanding the problem or opportunity that requires a solution.
2. Objectives Definition: Setting the goals and objectives for the design process.
3. Design and Development: Creating the artifact or solution based on existing theories, frameworks, or methodologies.
4. Demonstration: Presenting the solution and demonstrating its potential
5. Evaluation: Assessing the effectiveness and functionality of the designed artifact against predefined objectives.
6. Communication: Reflecting on the outcomes, refining the artifact, and identifying areas for improvement.

3.1. PROBLEM IDENTIFICATION

Every time a user interacts with the INCANTO website, a vast amount of data is collected: the name of the page with which the user interacted, the relevant section of the page, the URL, the browser, and device he is using, the country of his location and the time he spent on the page and the site as a whole and much more. All data is collected and sent to Google Analytics every time a user interacts with a tracked website element. At the same time, Google Analytics uses data sampling and aggregation techniques to make reports faster and more effective for the end user, sometimes at the expense of compromising the integrity of the information. The result of this approach is that Google Analytics allows data to be exported from the user interface at the lowest level of information granularity: data can be extracted for each user group based on age, geographic information, source of user visits, and other parameters. There may need to be more, as website traffic reports provide invaluable information about various aspects of a website's performance, user behavior, and customer interactions. By analyzing these reports, INCANTO can understand it audience, optimize site through this understanding, and achieve better results, whether tracking conversion rates, measuring SEO performance, or monitoring user demographics.

The next major issue is that on July 1, 2023, Universal Analytics stopped collecting and processing data. Moreover, it has already been announced that as of July 1, 2024, Google will no longer be able to access Google Analytics 3. Viewing reports in the interface or accessing data through the API will no longer be possible. At the same time, transferring data from the Universal Analytics interface to the

new site counter Google Analytics 4 is impossible, primarily due to the different data collection models in the two versions. In this case, the problem is protecting legacy consumer behavior and traffic analytics data during the transition from Universal Analytics to GA4. The motivation behind this problem is to ensure that INCANTO maintains access to valuable historical data while leveraging the benefits of the latest analytics solution.

3.2. Objectives Definition

The study aims to efficiently transition INCANTO from UA to GA4 without losing valuable historical data, ensuring the continuity of data-driven decision-making processes. By extracting data from UA, transforming it, and visualizing it in Power BI, the research will demonstrate the effectiveness of the DSRM approach in addressing the challenge of preserving legacy data. The project will identify key consumer trends, behaviors, and traffic patterns typical for the previous period. This research will contribute to understanding effective lossless data migration strategies, emphasizing the importance of preserving legacy data for future analysis with a historical foundation. The developed model will provide valuable guidance for companies facing similar transitions, highlighting the need for a systematic approach to data preservation and the benefits of leveraging a BI solution for a seamless transition from UA to GA4.

3.3. Design and Development

The design was crafted using the Microsoft Power BI platform. The solution involved a combination of exporting data from UA, storing it in a suitable format, importing it into Power BI, and analyzing the data. The entire layout, encompassing graphics and other design elements, was completed within this platform. The development process involved mapping out a structured development flow, which will be presented in the demonstration. This demonstration will showcase the transformations applied and the flow of data, illustrating the process from data acquisition to its utilization in the final solution.

3.4. Demonstration

The demonstration of the challenge of protecting legacy consumer behavior and traffic analytics data with a Universal Analytics transition in a Business Intelligence solution for INCANTO is accompanied by a working project that shows the whole development process from the beginning and will be described in part 4 of this thesis.

3.5. Evaluation

The evaluation will involve gathering feedback and conducting assessments from INCANTO regarding the presented solution. It will revolve around determining their satisfaction levels with the current solution. This involves assessing the efficiency of the data migration process, the quality of the data imported into Power BI, and the usefulness of the insights gained from the analysis.

3.6. Communication

Communication will be done through this work project where all the information about the solution will be described. This step involves communicating the results of the research to relevant stakeholders, such as INCANTO's management, marketing team, and data analysts. The communication involves presenting the findings, discussing the implications, and recommending any

further actions. This approach ensures a comprehensive and well-structured research process that ultimately leads to a solution that meets INCANTO's needs.

3.7. DATA MODEL

Power BI stores data in separate tables that can be related to each other. The organized structure of these tables and their relationships is called a data model. Currently, the most recommended Power BI model is the Star schema (Mihiranga, 2022). This schema organizes data into two types: measurement tables and fact tables. Measurement tables contain descriptive attributes that characterize an object, while fact tables contain information about events in which particular objects are involved. This method of organizing data based on logical divisions is considered optimal for Power BI.

Data for the analysis of visitor behavior on the website was used for the subdomain it.incanto.eu. In order to get data with minimal granularity, was used the markup of site visitors and their sessions using custom parameters where unique identifiers was stored. This is since Google Analytics does not return the identifiers we need through the API. The data was downloaded from when the company started using user parameters as markup of site visitors and their sessions, namely from 22 February 2022 until UA's end of data collection. In the case of the subdomain in question, it was 16 July 2023.

When connecting UA to Power BI via the built-in connector, there are other issues that could occur as well, such as: automatic UA data sampling[2] for high-traffic web pages often result in incorrectly displayed values. Automatic aggregation often creates average values, various dimensionality constraints, and metrics once connected. For the problem solved in INCANTO, it is critical to eliminate erroneous decisions built on too small a sample of data because once sampled, the patterns that emerge are representative of only some of the entire data set. Therefore, it was decided to use a different way of loading the data. There are multiple ways to save UA counter data and avoid sampling problem: manually by exporting the data from the interface, using a special add-on in Google Tables, using the UA Query Explorer tool, via Google Analytics API, exporting data from Google Analytics to Big Query. In this work querying Google Analytics API with the help of the client library and Python programming language was used. The data thus obtained was saved in Excel files, transformed, and ready for creating a data model. As a result, we have a model built on a classical star schema with multiple fact tables. Table 3.1 shows description of each table.

Table 3.1 – List of dimension tables

Table Name	Description
Fact_Sessions	Fact table logs each session, representing a group of interactions that take place on INCANTO website signifying a single visit
Fact_Pageviews	Fact table that captures data on each page view, tracking every time a user views a page on INCANTO website, indicative of website content engagement
Dim_Devices	Measurement table which stores details the device characteristics used by visitors, such as type (mobile, tablet, desktop) and information about

[2] https://support.google.com/analytics/answer/2637192?hl=en#zippy=%2Cin-this-article

	operation system and browser used by device, to understand device-based user experiences
Dim_Geo	Measurement table contains geographic information about visitors, including country, city, and region, to analyze location-based user behavior and preferences
Dim_User	Measurement table stores user-specific attributes like user ID and user category (new vs. returning), facilitating personalized user analysis and segmentation
Dim_Traffic_sources	Measurement table identifies where INCANTO traffic comes from, categorizing sources into channels such as organic search, direct, social, or referral, and other channel grouping, campaigns, and keywords to gauge the effectiveness of different acquisition strategies
Dim_Date_parameters	Measurement table organizes date-related attributes, providing a structured way to analyze data by time periods such as day, week, month, quarter, or year
Dim_Time	Measurement table organizes time-related attributes, providing a structured way to analyze data by time periods such as hour or tome of the day (night, morning, day, or evening), crucial for understanding user activity patterns throughout the day
scaleDatePeriod	Date options (scale) measurement table offers options for scaling date ranges in the analysis, with specific granularities such as year, month, week, and day

We have defined a hierarchical structure underlying the relationships between entities that are assumed to be one-to-many relationships in the database structure. Hierarchies presented in the appendix have been defined in the model to enable more detailed analysis and make it easier to manage.

We also utilized Data Analysis Expressions (DAX) to create more efficient measures, which helped us build a comprehensive Power BI dashboard. DAX is a valuable feature of Power BI that provides a library of operators and functions to enable users to create new measures in Power BI (Sousa et al., 2021).

Figure 3.1 shows the relational diagram of the INCANTO website traffic data for this project.

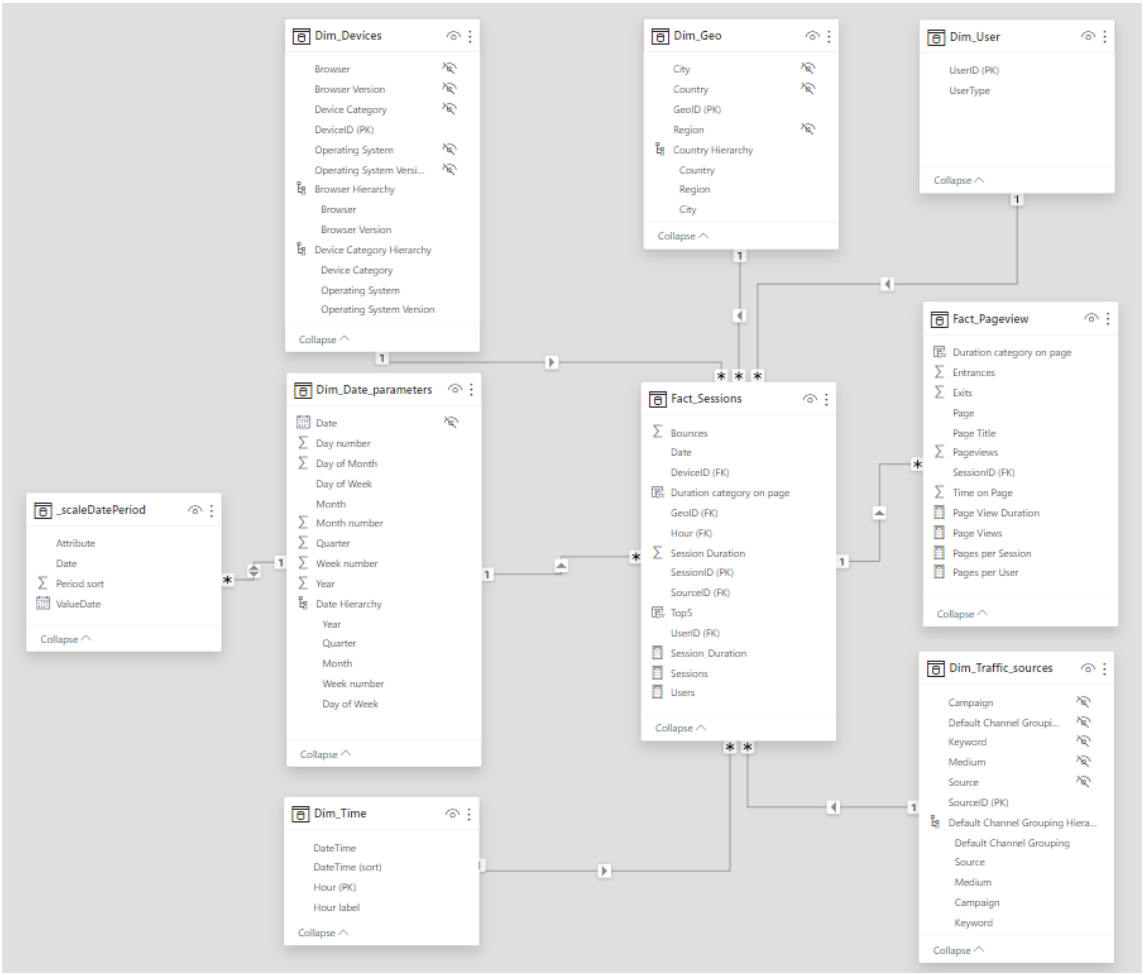


Figure 3.1 – The implemented dimensional model for INCANTO website traffic data.

4. RESULTS AND DISCUSSION

During the final stage of the Business Intelligence solution, the data was visualized, and the results were obtained. To effectively analyze and improve our website, we must understand how our visitors interact with it. It's important to keep track of these dynamics as search engines evolve, and new marketing tools become available. To preserve historical data, we created a Power BI dashboard that measures the site's performance in terms of audience activity. The dashboard is divided into six semantic sections, each of which provides insights into different aspects of our website's performance, such as traffic, time and number of visits, transitions, and geography. To make the materials easier to present to our company's management, we used the color scheme specified in INCANTO's brand book and to visually separate the sections from each other, their shades are different. Since the number of colors in Power BI themes is limited to eight, and we needed a minimum of nine colors, we used a custom theme created by loading a JSON file.

The following indicators are available across all pages of the report:

- Sessions: a record of a single visitor browsing a website during a given time period. This can include multiple screens or pageviews, events, or ecommerce transactions. Sessions end at midnight on the day a session was initiated or after 30 minutes of inactivity.
- Page Views: the total number of pages viewed during each session. For example, if a user opens five pages on the site during a single session, this counts as five page views.
- Users: the total number of unique users who have accessed the site.
- Session/Page Duration: the length of time a session or page view lasted. For example, if a user viewed one page for 5 seconds and another for 10 seconds, the session duration would be 15 seconds.
- Pages per User/Session: the ratio of pages to users or sessions. For example, if a user opens five pages during one session, this represents a 5/1 ratio.

On the first four pages of the report, switching chart axes using sliders (day, week, month, year) is used. This is a convenient alternative to the built-in drill-down function for switching between hierarchy levels on charts.

4.1. OVERVIEW REPORT

The Main Page (illustrated in figure 4.1) displays the leading indicators, which can be viewed in more detail on other pages. The panel with filters is always visible to avoid the need for multiple clicks when using slicers and to provide a clear view of the active filters. Slicers are presented in tile format, which helps optimize the report's performance by not creating queries when updating visualizations.

Based on the information in the Universal Analytics Overview Report for INCANTO, here are some meaningful insights:

- User Engagement: With an average of 3.38 pages per session and 4.50 pages per user, users are relatively engaged with the content, indicating that the website likely offers a good user experience that encourages browsing.

- Channel Traffic: Paid Search leads to the highest number of sessions, 39 000, indicating that the advertising budget drives most of the traffic. Organic Search follows closely with 36 000 sessions, which suggests strong SEO performance. Referral and Email channels are significantly lower, suggesting these areas may need more attention or improved strategies.
- Session Duration: The refusal category has the most significant proportion of sessions at 65.7K, 50.47% of all sessions. This is a concerning metric as it may imply that many users leave the site before it has fully loaded or interact with it minimally. Sessions lasting > 5 minutes are significantly fewer, at 13.1K (10.03% of sessions), which might suggest that fewer users are deeply engaged with the website.
- User Type: New Visitor sessions are at 91.5K, accounting for 70.35% of all sessions, indicating the site continually attracts new users. Returning Visitor sessions are at 38.6K, making up 29.65% of all sessions, which is a healthy number but still presents an opportunity to improve user retention and loyalty.
- Device Type: Most sessions occur on mobile devices, accounting for 113.6K sessions, or 87.35% of the total. This high percentage indicates that the website is predominantly accessed through mobile devices, consistent with global internet usage trends favoring mobile.
- Browser Usage: Safari (in-app) has the highest number of sessions among browsers, with 36,629. This suggests that many users are accessing the site through iOS apps.
- Strategic Insights: The high mobile usage emphasizes the importance of having a mobile-optimized website. The site should be responsive, with fast loading times and easy navigation on mobile devices. Given the high rate of mobile sessions, ensuring that all features, content, and marketing strategies are tailored for a mobile-first experience would also be necessary. The user experience on mobile should be a primary focus, including page load speed, mobile design elements, and seamless navigation to keep the high volume of mobile users engaged.

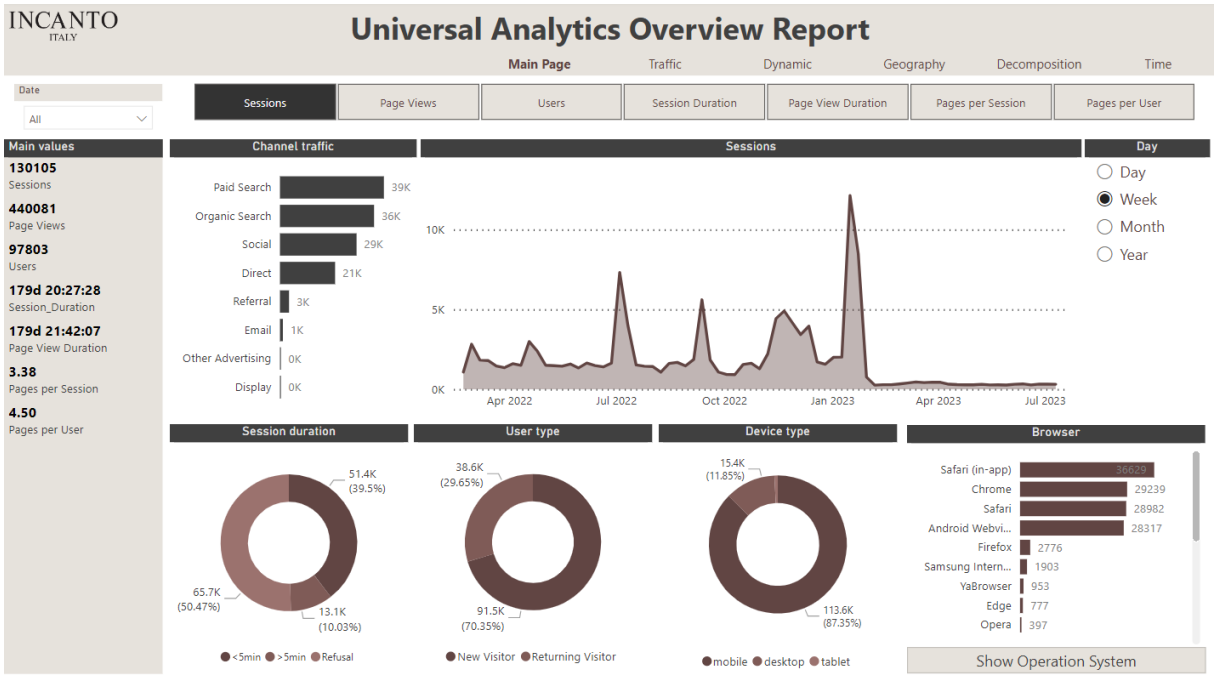


Figure 4.1 – Universal Analytics Overview Report – Main Page

Again, the high percentage of sessions classified as a refusal (50.5%) is concerning and should be addressed, as it could be indicative of technical issues or poor user experience, which could be even more pronounced on mobile devices due to factors like slower internet speeds or smaller screen sizes. Immediately after the extensive January 2023 peak, the period shows a sharp decline in sessions. Further technical investigation into this problem showed that UA had started not tracking some session IDs by that time.

4.2. TRAFFIC

The Traffic Page shows the leading indicators in more detail by grouping channels, sources, and campaigns. Channel traffic is a channel that shows where users come from. For example, most of them came from paid searches through paid links. Device type - accordingly, it is a phone, a computer, or a tablet. The ratio by device type is important as well for ux/ui designers, technician specialist and marketing specialists. Figure 4.2 displays an example of the dashboard for the Sessions metric and Month granularity.

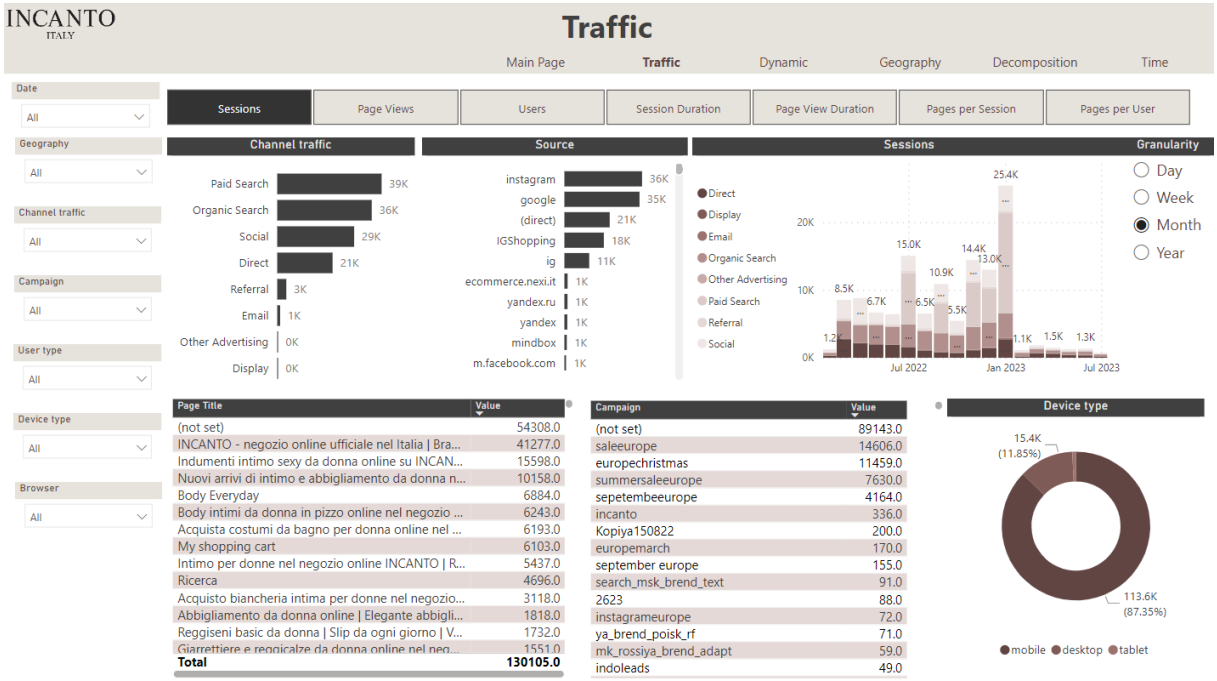


Figure 4.2 – Universal Analytics Overview Report – Traffic Page

Based on the information in the Traffic page, here are some meaningful insights:

- Channel Traffic Distribution: The primary traffic source is Paid Search, then Organic Search. As a rule, paid traffic is more expensive and should not be in the lead. If we analyze only 2022 separately, it had the proper ratio. We recommend going deeper into analyzing the changes in marketing strategies in 2023 and the ROI of advertising investments in this period.
- Source of Traffic: Instagram and Google are the top traffic sources, driving 36K and 35K sessions, respectively, indicating strong social media presence and search visibility. Google is the second largest source of visits with 35 thousand, which coincides with many sessions from organic search, indicating effective SEO strategies. At the same time, we see some disunity of urchin tracking module (UTM) tags, e.g., ig, IGShopping, and Instagram resources should have

been combined if their separation was not intentional by marketing professionals. It is recommended to pay attention to this to have a consistent approach to tagging.

- Page Performance: The page with the title "(not set)" has the highest value, indicating tracking issues where some sessions or visits are not attributed to a specific page. This should be clarified to ensure accurate data collection. The official online shop page ("INCANTO - negozio online ufficiale nel Italia | Brand Intimo Italiano") has the second-highest value, indicating that it will likely be the main landing page and a compelling entry point for e-commerce. By a significant margin, the capsule collection page "Indumenti intimo sexy da donna online su INCANTO" is ranked third in visits. This indicates effective SEO or marketing for this product category relative to other categories.
- Trends over time: The "Sessions" trend graph with granularity applied on a monthly basis shows a sharp increase in the number of sessions in January 2023, followed by a sharp decline, as previously noted, indicative of the technical aspects of tracking individual sessions from February 2023 onwards. A further decision on whether the data can be used from February 2023 onwards will need to be confirmed by marketing experts. This issue could not be tracked on standard UA charts, which proves the value of the analysis performed.
- User preferences by device: Most users access the site via mobile devices (87.35%), as shown on the previous page. The need to repeat this graph is due to the high value of information on device statistics by the campaign for further image optimization by designers according to the available data. For example, when using filters, it can be found that the share of users opening email newsletters from the desktop is 16%. At the same time, it is only 1.3% on social networks.
- Campaigns: There are several campaigns with relevant values in the list, but the campaign labeled "(not set)" has the highest value of 89 143, indicating that some campaigns may be incorrectly marked or tracked. The campaigns "saleeurope" and "europechristmas" seem to be the most prominent among the correctly tracked campaigns, indicating effective campaign strategies.

4.3. DYNAMIC

The Dynamic page shows leading metrics, but dynamically, meaning everything is displayed on a date graph with a selected degree of granularity. This dashboard analyses the site's performance over time, focusing on user engagement and traffic metrics. Unlike other pages, there are no metrics to select at the top here, as the goal is to compare comparable metrics over time. To do this, the user is presented with filtering options. On the right side of the panel, several sliders implement the option to turn the display of graphs on or off by metrics to "unload" the visualizations and make them easier to understand. Figure 4.3 displays the dashboard with the Week type of granularity without any filter implemented.

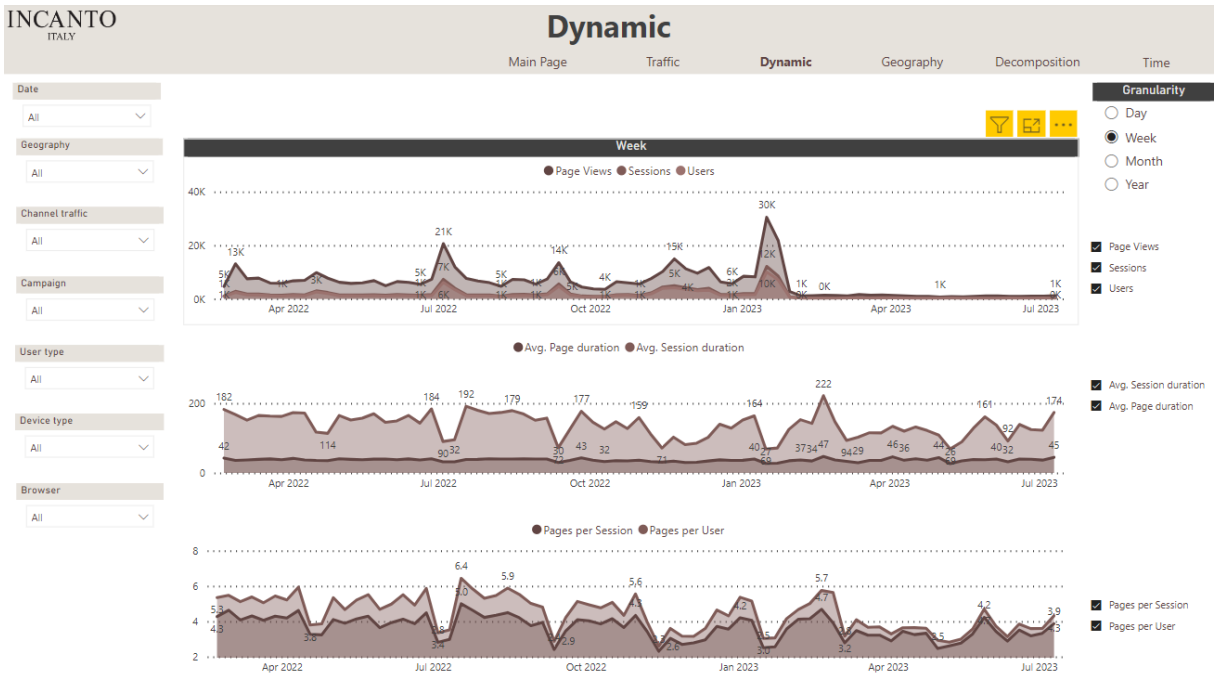


Figure 4.3 – Universal Analytics Overview Report – Dynamic Page

Here are some conclusions for INCANTO based on the analysis of this page:

- Seasonal trends: The data is seasonal. For example, a significant peak in all indicators (page views, sessions, and users) occurs in January 2023, as well as seasonal peaks characteristic of fashion retail, including swimsuit sales - an increase in activity in July and September, as well as during sales periods - on Black Friday.
- User engagement: Average page duration and average session duration were highest in February 2023. This suggests that users were more interested in content during this period. However, given the previously reported drop in sessions due to a lack of Google tracking, caution should be exercised when interpreting all metrics after January 2023.
- Consistency in engagement: Average session duration shows more variability over time than average page view duration. This may indicate that although users spend the same amount of time on a page, the total time they spend per session is more variable.
- Consistency in interaction: The number of pages per session and pages per user remains relatively constant over time, suggesting that although the number of users and sessions varies, user behavior in terms of the number of pages they view remains stable.
- Relation between users and sessions: The peaks in the number of users and sessions coincide closely, suggesting that the website does not have many repeat visits in the same period. This needs to be carefully considered to encourage users to come back more often.
- Relationship between sessions and interaction: we can see a decrease in view duration and average page count during peak periods. This may be due to the fact that during active advertising campaigns the number of random new and unfamiliar users with the brand increases, which affects the quality of traffic.

4.4. GEOGRAPHY

The Geography page shows essential metrics associated with locations and regions. This is especially important when planning and analyzing local promotional events and considering the need to enter separate localized language and currency versions. On this page, the metric selection slicers, which were removed from the Dynamic page, were returned to make tracking regional features in various slices convenient. In addition to visualizations with the top 5 cities, a table by country, and a bubble map, it was decided to add a histogram on a timescale to understand how stable the relationship between brand and site visitors is in different regions. Taking into account the previously identified problems with taking into account sessions on this page, we will consider the data taking into account the filter for 2022 as it can be seen in figure 4.4.

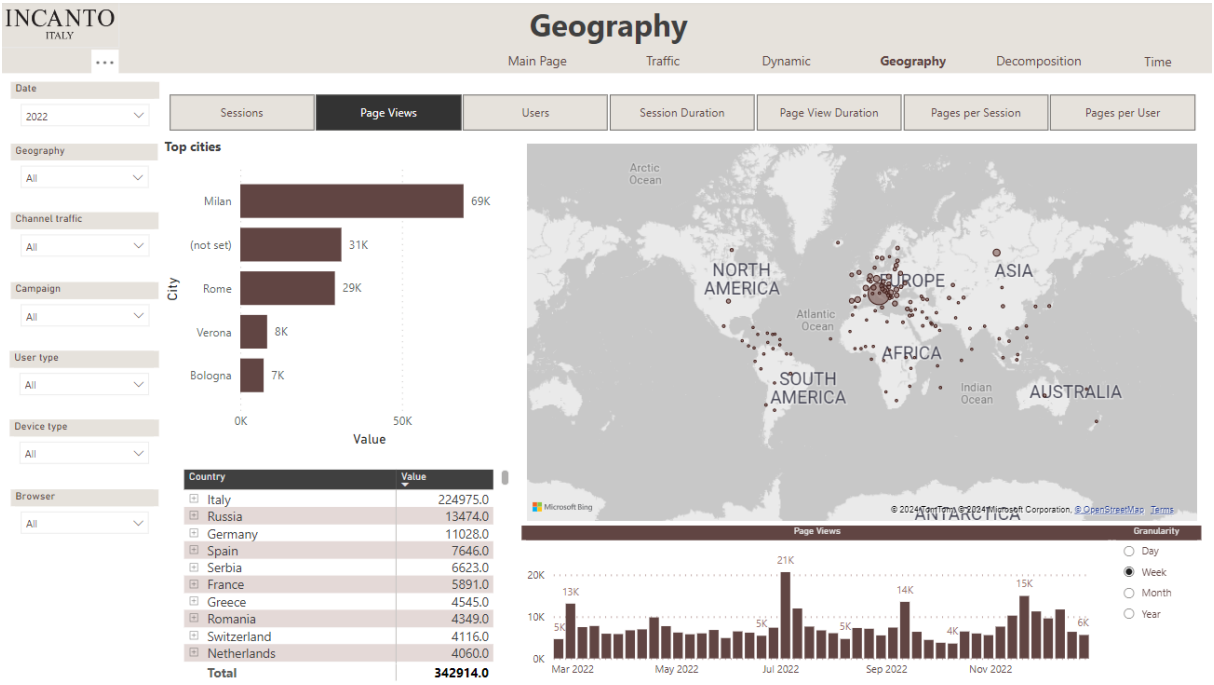


Figure 4.4 – Universal Analytics Overview Report – Geography Page

Let's analyze the geographic data, taking into account the filter applied by date:

- Top cities by number of page views: Milan has the highest page views, indicating a solid user base in the town and high brand awareness, which is also due to the presence of offline points of sale. It stands out significantly with 69K page views, more than double the number of page views in the next city.
- Undefined Segment: A significant number of page views (31K) is associated with "(unidentified)," suggesting there were sessions in which city data was not collected or was not available. This could indicate problems with tracking and possible anonymization of data to protect users.
- Distribution by country: Italy, as expected, has the highest page views, with more than 65% of views coming from the country. It is followed by a significant lag by Russia and Germany, with 4 and 3% shares, respectively. Many sessions from Italy correspond to the company name and subdomain localization.

- Participation in European markets. The concentration of bubbles over Europe on the world map suggests greater involvement in European countries. However, it should be noted that web traffic is observed on all inhabited continents.
- Distribution by Session: The histogram below shows the distribution of page views over time, with notable peaks in July and November 2022. This is consistent with previous data on promotional activity during these periods and seasonal interest in INCANTO's offerings.

4.5. DECOMPOSITION

The Decomposition tab (shown in figure 4.5) offers a detailed breakdown of web sessions, providing information on user types, geographic distribution, city-level interactions, traffic sources, and associated campaigns.

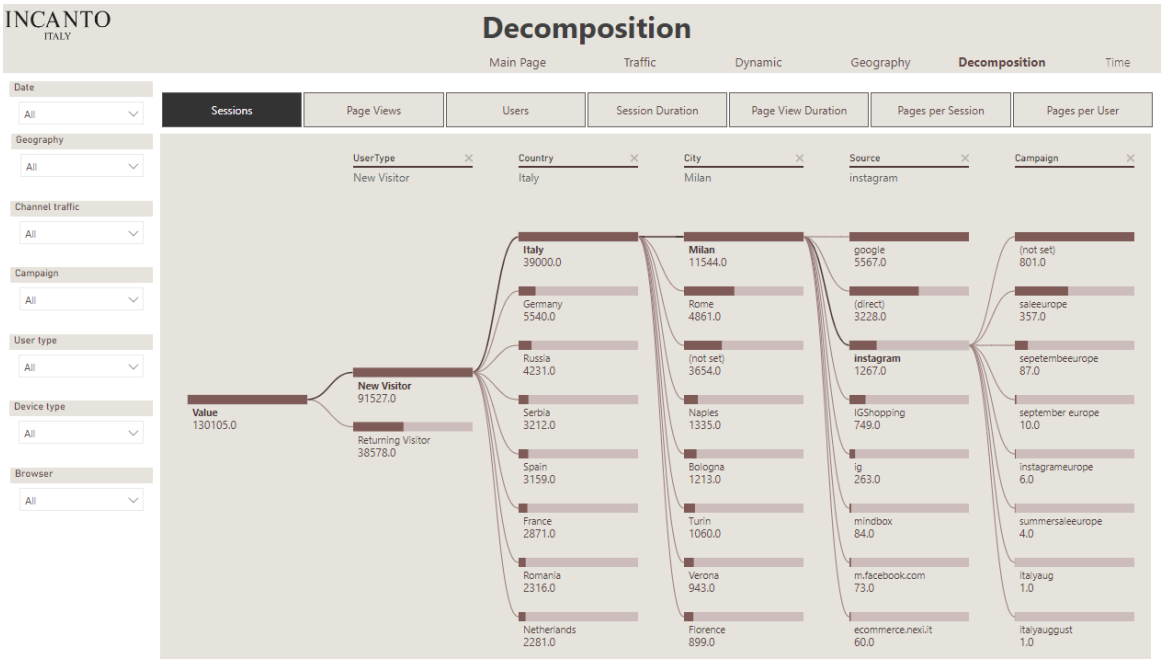


Figure 4.5 – Universal Analytics Overview Report – Decomposition Page

The decomposition tree displays the main metrics but is segmented based on the selected parameters. It illustrates how the total can be broken down into other categories. For example, in the one hundred thirty thousand sessions, the decomposition tree breaks down how this number is distributed among user types, countries, cities, sources, and campaigns. Here again, various slices and filters can be used, but let's consider the analysis of the session indicator without using a filter by date: it is very clearly demonstrated from the top step of decomposition how the maximum values are first divided at the user type level by majority into new users. Then again, we go to the top of the list - the majority of new users are from Italy; in turn, the maximum of them are from Milan. The maximum number of sessions for the study period for new users from Milan was attracted through Instagram. Still, the label for this campaign was not created - the campaign value is reflected "(not set)." It is necessary to bring to the attention of responsible employees the importance of designating and monitoring utm-tags to track the effectiveness of ongoing campaigns.

4.6. TIME

We have created a page that displays the distribution of site visits based on the day of the week and time of day. By utilizing separate calendars and time-of-day directories, we were able to create a revealing visualization. However, to achieve this graphic using only default visualizations, we had to resort to the layout of separate elements. This information can be useful for technicians in analyzing peak site loads, as well as for marketers and analysts who want to identify how metrics are distributed throughout the day or week, based on factors such as dates, locations, campaigns, and other filters.

The Time page can be handy for content management and marketing teams to determine the optimal time to publish new content and schedule social media posts, email campaigns, or online advertising to increase user engagement and conversion rates. Figure 4.6 shows the Time page for the Page Views metric.

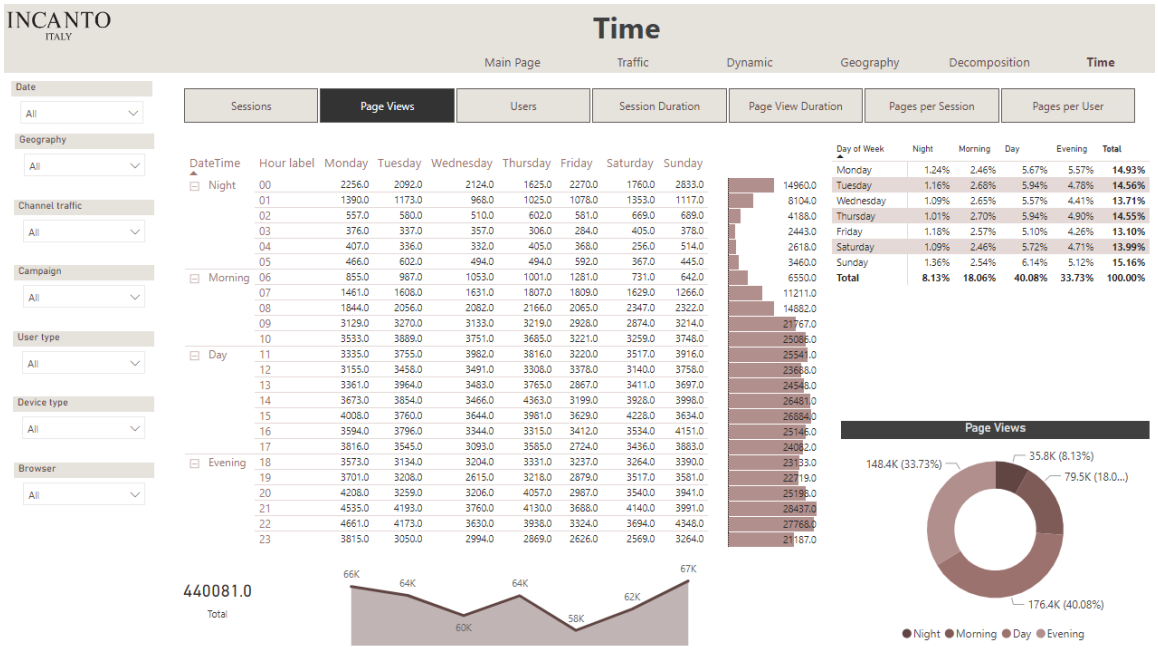


Figure 4.6 – Universal Analytics Overview Report – Time Page

Based on the page views metric, we see the following observations:

- Characteristics of the time of day: User activity varies significantly throughout the day, typically peaking in the late evening (9:00–11:00 p.m.). That said, evening hours tend to show slightly higher page views, especially around 9:00 pm, which may indicate that users are more likely to browse the website in their free time after work or daily routines. At the same time during the daytime, there is consistent activity with less fluctuation than in the morning and evening, suggesting sustained interest and a potentially good time to attract users.
- Daytime Dominance: The donut chart indicates that most page views occur during the day (40.08%), followed by views in the evening (33.73%). When deciding when to release new content or launch live campaigns for maximum reach, this information should be considered.

- Activity by day of the week: The table shows that Monday, Tuesday, and Thursday have relatively similar dynamics, but there is a slight decline on Wednesday, Friday, and Saturday, followed by a slight increase on Sunday. Sunday and Monday have higher page views than other days of the week, which may indicate that users have more free time to browse the website at the beginning of the week or after the weekend. This suggests that any urgent content updates or marketing efforts may be most effective if they are timed around this time.

4.7. DISCUSSION

Returning to this work's motivation and problem identification, two main tasks can be identified - preserving critical historical data and creating a convenient and easy-to-use tool for marketers, technical specialists, analysts, and management. The human brain is incapable of processing the increasingly significant volumes of complex data that user activity on the site generates. Manual results are too simple and dependent on the analyst, and predictions can be crudely made (Maggie, 2020). The proposed solution is a BI-tool for providing analytical support for data-driven decision-making for stakeholders. This work is essential for the company, whose online strategy now consists of a more detailed study of web traffic and the effectiveness of existing campaigns to clarify positioning and strengthening in the market.

Today, a company can get a complete picture of a website's performance in the simplest way. The analytics for INCANTO Universal Analytic data reveal a website with solid traffic from paid and organic search, high mobile usage, and notable international reach, peaking during specific campaigns and evening hours. To maximize effectiveness, strategies should focus on enhancing mobile experience, leveraging high-traffic periods, and potentially expanding international marketing efforts. Incorporating filters for various dimensions such as geography, device type, and time can enable a deeper analysis, revealing more nuanced user behavior patterns and potentially uncovering opportunities for further optimization and targeted marketing strategies.

It should be noted that it was possible not only to complete the tasks set in the work but also to create a template that can be used both by INCANTO with other domains (en, us) and by anyone interested in quickly solving a problem, the solution period of which is limited to just a few months. Assessing the role of this solution, one cannot fail to note the labour savings on the part of marketers, as well as the difficult-to-assess importance of the detected technical problems in the data due to the transparency of the data and the lack of aggregation and sampling that are characteristic of the Universal Analytics tool. Moreover, the skills acquired in solving this problem will help build similar and updated dashboards based on GA4. And plans in the future include the task of combining them into a single marketing analysis tool.

We evaluated the project using the Balanced Scorecard methodology (Kaplan & Norton, 1992). The essence of this method is that both financial and non-financial components are checked by identifying four separate areas that need to be analyzed:

1. Business processes - transferring historical data from Universal Analytics to Power BI was completed within the planned time frame (3 months) without data loss, verified by random sampling.

2. Finance - this assessment is currently being processed. However, according to the team, analyzing the effectiveness of historical campaigns can reduce marketing costs by at least 10-15%. In the future, tests will allow us to evaluate income growth due to the optimization of strategies and calculate the dynamics of changes in the return on investment indicator.
3. Users - using a proposed scale of zero (not at all useful) to ten (extremely useful) to quantify the use of the dashboard, the marketing team and management rated it an average of eight.
4. Training and Growth - INCANTO trained the marketing team and analysts after implementing the dashboard, focusing on data analytics capabilities and Power BI features, fostering a culture of continuous learning and innovation in data analytics.

In this way, we can recognize the usefulness of the work that went into creating this dashboard and ensure that it meets the needs of the company, which was the project's primary goal.

At the same tome the research not only offers a practical solution to the acute problem of online stores in the context of digital marketing but also integrates the theoretical foundations of using Design Science Research methodology into this process. The project bridges the gap between digital marketing and data analytics by demonstrating how BI solutions can optimize marketing strategies through data-driven decision-making. This approach encourages further research at the intersection of marketing, data analytics, and e-commerce, fostering innovation in an academic field that prepares researchers to better understand the capabilities of Power BI in tandem with the ever-evolving Google Analytics solutions.

5. CONCLUSION

This work presents the development of a Business Intelligence solution for INCANTO. The solution aims to store and analyze valuable historical data related to consumer behavior on the website. The data was previously stored in Universal Analytics, which had stopped collecting and processing data on 1 July 2023. In 2024, Google will no longer be able to access Google Analytics 3, and it will not be possible to view reports in the interface or access data via API.

The six dashboards created for INCANTO within the Business Intelligence solution offer insights into various aspects of the website's performance, leveraging data from Universal Analytics. Thanks to the created solution, INCANTO can now store and analyze data on the number of users, sessions, and geolocation of visitors. Additionally, it can identify the most frequently visited pages, traffic sources that generate the maximum number of sessions, and the dynamics of users during the week and the day. This information allows for the identification of peak periods when the site has the most active sessions and users, as well as the most viewed pages that generate the most traffic. The continuity of data provides valuable information for improving website content and design.

Furthermore, our study showcases the potential of Business Intelligence solutions in web analytics and presents a universal methodology that can be adopted by other companies grappling with the issue of preserving historical data. This is a critical and pressing concern, as Google Analytics, which holds a market share of 85.4% (Gartner, 2022), is the most widely used tool for monitoring website activity.

As per McGuirk's research in 2023, numerous organizations had utilized UA to analyze their websites comprehensively. This helped them gain insights into the composition of their visitors, their engagement patterns, and the effectiveness of their digital marketing campaigns. However, the announcement of sunsetting the UA platform has caused a lot of distress among the web analytics and marketing communities. This work is critical in providing a roadmap to move to an analytics platform while maintaining the strategic value of historical data, ultimately improving business intelligence, and maintaining a competitive advantage in digital marketing. The proposed solution offers data preservation and advanced reporting capabilities, enabling the extraction of valuable insights from the data. It is suitable for companies of any size and has a user-friendly interface that simplifies reporting for end users.

6. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORKS

Important to note some limitations to the analysis performed in this work. Firstly, the analysis is only useful for a limited time since all preservation work for historical data from the Universal Analytics counter must be completed by July 1, 2024. However, if the data can be saved using the methods outlined in this work as a database, the analysis can continue without any restrictions.

It is also worth mentioning that currently, it is not possible to connect Universal Analytics data to GA4 due to the difference in data collection models: common metrics like users and sessions are measured differently in UA and GA4 (DiTomaso, 2023). However, it's possible that the analysis can be considered continuous in certain sections. Additionally, with some modifications to the data model, the report can easily be adapted to the new version of Google Analytics, thanks to the model's universality.

Although the solution provides a strong foundation for INCANTO, its applicability, and effectiveness in other contexts remains to be thoroughly studied. It includes assessing the solution's flexibility to use different data structures and business models, critical to its wider adoption. In addition, the discussion may address potential issues during the implementation phase, such as user training and system integration issues.

It was out of the scope of this analysis, but for future work, we recommend using Transactions and Cost tables to calculate the return on investment (ROI) of all the web marketing actions a business puts in place to determine its digital competition in the market. This will allow analyzing transactional and ROI advertising across channels, which remains highly important.

Additionally, in future work, several directions could significantly improve the understanding and applicability of the Business Intelligence (BI) solution developed for INCANTO. There are some recommendations:

- Integration with real-time analytics. Future research could explore integrating a BI solution with real-time analytics platforms. This will allow INCANTO to monitor and respond to user behavior and website performance in real time, offering opportunities for immediate optimization.
- Advanced predictive analytics. Implementing predictive analytics models within BI can provide INCANTO with forward-looking information to predict user behavior, sales trends, and website traffic patterns. This will help in strategic planning and proactive decision-making.
- Machine learning for personalization. Future research could explore the application of machine learning algorithms to personalize user experience on a website. A BI solution can facilitate dynamic content customization by analyzing user behavior and preferences, increasing user engagement and satisfaction.

7. REFERENCES

Azizi, E. (2023). IranServer's Digital Marketing Transformation: A Study of the Adoption of Marketing Dashboards. NOVA Information Management School, Universidade Nova de Lisboa. Retrieved from http://hdl.handle.net/10362/159699

Bhosale, S. S., Sharma, Y. K., & Kurupkar, F. (2020, March). Role of business intelligence in digital marketing. In International Journal of Advance and Innovative Research, 7(1), Part-1. Mumbai. Retrieved from https://www.researchgate.net/publication/339676330_ROLE_OF_BUSINESS_INTELLIGENCE_IN_DIGITAL_MARKETING

DiTomaso, D. (2023). How to Blend Universal Analytics (UA) and Google Analytics 4 (GA4) Data in Looker Studio. Kick Point Playbook. Retrieved from https://kpplaybook.com/resources/how-to-blend-ua-and-ga4-data-with-looker-studio/

Gartner. (2022). Web Product and Digital Experience Analytics Market Review. Gartner. Retrieved from https://www.gartner.com/reviews/market/web-product-and-digital-experience-analytics

Google. (2012). Universal Analytics. Retrieved from https:// Universal Analytics analytics.googleblog.com/2012/10/universal-analytics.html

Google. (2023). Google Analytics 4 has replaced Universal Analytics Retrieved from https://support.google.com/analytics/answer/11583528?hl

Ivanov, E. (2022). The Impact of Business Intelligence Implementation on Digital Marketing Analysts' Effectiveness. Bachelor's thesis, Tallinn University of Technology, School of Business and Governance, Department of Business Administration. Retrieved from https://digikogu.taltech.ee/en/Download/04b3fcf3-7660-4fed-99bb-ccd910331e2b

Kalboussi, E., Ndhaief, N., & Rezg, N. (2023). Machine Learning for Decision Making in the Improvement of Digital Press Services. In CIGI QUALITA MOSIM 2023. Laboratoire de Génie Informatique de Production et de Maintenance (LGIPM), Université de Lorraine, France. Retrieved from https://oraprdnt.uqtr.uquebec.ca/pls/public/docs/GSC6979/O0005294075_CIGI_QUALITA_MOSIM_2023_paper_4616.pdf

Kaplan, R. S., & Norton, D. P. (1992). The Balanced Scorecard - Measures That Drive Performance. *Harvard Business Review*, 70(1), 71-79. Retrieved from https://hbr.org/1992/01/the-balanced-scorecard-measures-that-drive-performance-2

Kursan, I., & Mihic, M. (2010). Business Intelligence: The Role of the Internet in Marketing Research and Business Decision-Making. Management, 15(1), 69-86. Retrieved from https://hrcak.srce.hr/53607

McGuirk, M. (2023). Performing web analytics with Google Analytics 4: a platform review. Journal of Marketing Analytics, 11(4), 1-15. https://doi.org/10.1057/s41270-023-00244-4

Mihiranga, N. (2022). Power BI Data Modeling: Build Interactive Visualizations, Learn DAX, Power Query, and Develop BI Models (English Edition). BPB Publications. Retrieved from https://books.google.pt/books?id=JpxlEAAAQBAJ

Maggi, P. (2020). Enhanced web analytics for health insurance. NOVA Information Management School (NIMS). http://hdl.handle.net/10362/101010

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Prikrylova, D. (2016). Business intelligence models for capturing and analysis of enterprise marketing data [Master's thesis, Masaryk University Faculty of Informatics]. Theses.cz. Retrieved from https://theses.cz/id/c65nr9/

Power, D. J. (2007, March 10). A brief history of decision support systems. Retrieved from http://DSSResources.COM/history/dsshistory.html

Pucka, V. (2020). Marketing Data Visualization - A Complete Guide with Examples and Tool Recommendations. Whatagraph. Retrieved from https://whatagraph.com/blog/articles/what-is-data-visualization

Sousa, R., Miranda, R., Moreira, A., Alves, C., Lori, N., & Machado, J. (2021). Software tools for conducting real-time information processing and visualization in industry: An upto-date review. Applied Sciences, 11(11), 4800

Talaoui, Y., & Kohtamaki, M. (2020). Of BI research: a tale of two communities. Management Research Review, Retrieved from https://doi.org/10.1108/MRR-10-2019-0452

Tavera Romero, C. A., Ortiz, J. H., Khalaf, O. I., & Rios Prado, A. (2021). Business Intelligence: Business Evolution after Industry 4.0. *Sustainability*, 13(18), 10026. Retrieved from https://doi.org/10.3390/su131810026

The CMO Survey. (2023). The CMO Survey: Highlights and Insights Report - Fall 2023. Duke University's Fuqua School of Business. Retrieved from https://faculty.fuqua.duke.edu/cmosurveyresults/The_CMO_Survey-Highlights_and_Insights_Report-Fall_2023.pdf).

W3Techs. (2024). Usage statistics of traffic analysis tools for websites. Retrieved from https://w3techs.com/technologies/overview/traffic_analysis

Ye, X., & Jonilo, M. (2023). Unleashing the Power of Big Data: Designing a Robust Business Intelligence Framework for E-commerce Data Analytics. Journal of Information Systems Engineering and Management, 8(3), 22638. http://dx.doi.org/10.55267/iadt.07.13946

APPENDIX

Annex 1 – Query example to the Google Analytics API using a client library and the Python programming language in Google Colab


```
{
  "name": "Custom Theme",
  "dataColors": [
    "#614543",
    "#7F5B57",
    "#9B726E",
    "#B08F8C",
    "#C5ACAA",
    "#DACBC8",
    "#E4D9D7",
    "#EEE7E6",
    "#F8F6F5"
  ],
  "good": "#1AAB40",
  "neutral": "#C5ACAA",
  "bad": "#D64554",
  "maximum": "#614543",
  "center": "#C5ACAA",
  "minimum": "#F8F6F5",
  "null": "#F8F6F5"
}
```


Annex 2 – JSON theme

Measure	DAX Formula
Page View Duration	DIVIDE(sum('Fact_Pageview'[Time on Page]),24*60*60))&"d"&" "&format(DIVIDE(sum('Fact_Pageview'[Time on Page]),24*60*60),"hh:mm:ss")
Page Views	sum('Fact_Pageview'[Pageviews])
Pages per Session	[Page Views]/[Sessions]
Pages per User	[Page Views]/[Users]
Session Duration	DIVIDE(sum('Fact_Sessions'[Session Duration]),24*60*60))&"d"&" "&format(DIVIDE(sum('Fact_Sessions'[Session Duration]),24*60*60),"hh:mm:ss"
Sessions	COUNT('Fact_Sessions'[SessionID (PK)])
Users	DISTINCTCOUNTNOBLANK(Fact_Sessions[UserID (FK)])
Avg. Page view duration	sum('Fact_Pageview'[Time on Page])/sum('Fact_Pageview'[Pageviews]))
Avg. Session view duration	sum('Fact_Sessions'[Session Duration])/[Sessions]

Annex 3 – DAX expressions

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa