



Tomorrowland Data Maestro: unlocking festival magic with data-driven & ML insights

Realisation

Bachelor Applied Computer Science

Wouter Selis

Academic year 2023-2024

Campus Geel, Kleinhoefstraat 4, BE-2440 Geel

Table of contents

1	Abstract	3
2	Acknowledgements	4
3	Introduction	5
4	Presenting EpicData	6
5	Project scope	7
5.1	Description	7
6	Contact persons	8
7	Data analysis	9
7.1	Data Analysis Process	9
7.2	Conclusion	10
8	Objectives for visualisations	11
8.1	Merchandising	11
8.2	Financial	11
8.3	Consumption	11
8.4	Individual visitors.....	11
8.5	Social media activity.....	12
8.6	LogisticsCamping	12
8.7	PodiaSetup	12
9	Power BI	13
9.1	Usage	14
9.2	Data model	15
9.3	Dashboard	17
9.3.1	Home.....	17
9.3.2	Consumption.....	18
9.3.3	Merch – popular items.....	19
9.3.4	Merch – Financial.....	20
9.3.5	Logistics Camping.....	21
9.3.6	Social Media	22
9.3.7	Individual visitors	23
9.3.8	Podia	24
9.3.9	Podia – Detail.....	25
9.3.10	Safety Security.....	26
10	Data Engineering	27
11	Hackaton	30

1 Abstract

This work is written based on the internship of Wouter Selis and was made to graduate as a bachelor Applied Computer Science at Thomas More Hogeschool Geel, Belgium. The subject for this internship is to provide actionable insights into various aspects of the Tomorrowland festival, aimed at enhancing operational efficiency, maximizing revenue, and improving attendee experience utilizing Power BI.

The project involves processing and analysing data sourced from various synthetic data sources associated with the Tomorrowland festival. This includes ingesting, cleaning, and preparing the data for in-depth analysis. Additionally, natural language processing techniques were implemented to develop valuable data and to gain deeper insights. Utilizing the data engineering platform Azure Databricks, the data was efficiently combined and stored to facilitate integration and management, enabling in-depth analysis.

Once the data analysis & processing was done, comprehensive visualizations were created to highlight key insights and trends, thereby supporting data-driven decision-making processes, with a leveraging tool like Power BI.

An essential part of the internship is learning to effectively present data findings and recommendations. At the conclusion of the internship, these insights and actionable strategies are presented to the EpicData management team.

2 Acknowledgements

I am deeply grateful for the incredible opportunity to work at EpicData as an intern. This experience has been invaluable in shaping my professional skills and knowledge. I would like to extend my heartfelt thanks to Rutger Mols and Frédérick Vissers for their mentorship throughout my internship. Their guidance, support, and insights have been instrumental in my growth and learning.

I also want to express my sincere gratitude to Kathleen Renders for being my internship supervisor. Her encouragement and constructive feedback have greatly enhanced my internship experience.

Sincere appreciation is directed towards the teachers at Thomas More Geel. Over the past years, their dedication and expertise have provided me with a solid foundation and have significantly contributed to my academic and professional development.

Finally, my deepest gratitude goes to my parents for their unwavering support and encouragement throughout my academic journey. Their belief in me has been a constant source of motivation and strength. From the late-night study sessions to the moments of doubt, they have always been there to guide and uplift me. This achievement would not have been possible without their love, sacrifices, and dedication.

3 Introduction

Tomorrowland, valued as one of the globe's largest and most prestigious music festivals, annually attracts hundreds of thousands of attendees. As the festival's complexity continues to expand, data-driven decision-making has emerged as an essential element for optimizing its operations, revenue streams, and attendee satisfaction. This internship at EpicData is dedicated to employing data analytics to elevate various facets of Tomorrowland.

The project involves processing and analysing data sourced from various synthetic data sources associated with the Tomorrowland festival. This includes ingesting, cleaning, and preparing the data for in-depth analysis. Additionally, artificial intelligence and machine learning techniques are implemented to develop valuable data, such as predictive models and natural language processing applications. Utilizing platforms such as Azure Synapse, Databricks, and Snowflake, the data is efficiently combined and stored to facilitate integration and management, enabling in-depth analysis.

With a leveraging tool like Power BI, comprehensive visualizations are created to highlight key insights and trends, thereby supporting data-driven decision-making processes.

An essential part of the internship is learning to effectively present data findings and recommendations. At the conclusion of the internship, these insights and actionable strategies are presented to the EpicData management team.

By focusing on these objectives, the internship project aims to contribute significantly to Tomorrowland's optimization efforts, leveraging data analytics to enhance its operations, revenues, and overall attendee satisfaction.

4 Presenting EpicData

EpicData is a dynamic data consultancy firm, emerged in 2022 from the merger of two esteemed companies, AtOnce and DataMotive, both established in 2004. Based in Kontich, Belgium, EpicData operates under the umbrella of the Cronos Group, a prominent IT corporation.

EpicData positions itself as the “go-to partner to turn your organisation into a data-driven powerhouse.” They invite their clients to “embrace the potential of data and achieve unparalleled insights for sustainable growth.”(epicdata.be)

With a mission focussed on unlocking the full potential of data for companies and their employees, EpicData collaborates closely with clients to develop tailored solutions addressing specific data needs. Their goal is to lay a robust foundation for data-driven success, fostering a culture where data plays a pivotal role.

The core expertise of EpicData lies in various aspects of data management and analytics. From crafting comprehensive data strategies to building robust data platforms and engineering solutions, their team excels in guiding organizations towards effective data utilization. Additionally, they specialize in creating visually compelling data representations, implementing BI tools, and conducting insightful business analytics.

To deliver their services, EpicData harnesses a diverse array of advanced technologies and tools, including Astrato, Microsoft Azure Synapse Analytics, Power BI, SAP Analytics Cloud, SAP Business Objects, Snowflake, Tangent Works, Qlik Sense, and QlikView.

EpicData's commitment to tailoring data solutions and their extensive proficiency across divers data technologies position them as a valuable partner for any organization seeking to leverage their data effectively. Through collaboration and customized strategies, EpicData ensures that their clients can achieve sustainable data-driven success.

5 Project scope

5.1 Description

The project will start at the 26th of February and end on the 24st of May. This time frame will be divided in three major parts.

During my internship at EpicData, I had the opportunity to work on a comprehensive project that involved several key stages, each contributing to the successful analysis and visualization of data.

The first phase of my internship will focused on analysing the data. This initial step is crucial for understanding the data set and generating preliminary ideas for potential insights. During this phase, I will conduct an in-depth exploration of the data, identifying patterns, trends, and anomalies. This foundational work provides the necessary context and direction for the subsequent stages of the project.

The second phase will involve cleaning and transforming the data. This will be a detailed process where I address any inconsistencies, missing values, and errors within the data set. By the end of this phase, I will ensure that the data is clean, reliable, and ready for analysis. Additionally, I will create fact and dimension tables, which organize the data into a structured format, facilitating efficient querying and analysis in the next phase.

The final phase of my internship will be dedicated to creating insights and visualizations using Power BI. I will develop a comprehensive dashboard that provided a detailed report covering various key aspects such as financial metrics, consumer behaviour, operational logistics, social media data, attendance, and safety/security. The financial metrics analysis will include revenue, expenses, profit margins, and other financial indicators to assess the company's financial health. Understanding customer preferences, purchase patterns, and engagement levels are essential components of the consumer behaviour analysis. Evaluating supply chain efficiency, inventory management, and other logistical operations will be covered under operational logistics. The social media data section will examine social media trends, customer feedback, and sentiment analysis. Attendance metrics will track participation levels and related data, while safety and security will monitor incidents and breaches, implementing preventive measures.

The last two weeks of this project will be spent working on the end report. This report will be the documentation of my internship explaining everything I did. Additionally I will be working on a presentation to present to the EpicData team.

6 Contact persons

During my internship at EpicData, I had the privilege of working with several assigned contact persons and fictive clients, each responsible for a different dataset. These individuals played a crucial role in guiding me through my tasks and providing valuable insights.

Henny Speelman acted as the Event Manager. He was an expert on the data, and I could always reach out to him with any questions. His extensive knowledge of data visualization also made him an invaluable resource, as he taught me a lot about best practices in this field.

Neil Vets was my second contact person, responsible for the social media data. He was the go-to person for any questions related to social media, helping me understand and navigate the intricacies of this dataset.

Anthony Coppens handled the consumption dataset. We had detailed discussions about the insights he wanted to see regarding food and beverage data, ensuring that my analyses aligned with his expectations and needs.

Timmy Diricx was responsible for logistics, specifically ensuring that attendees of Tomorrowland could sleep and rave before the stages. His role was critical in understanding the logistical aspects and their impact on the event's overall success.

Bart Van Mulders managed the merchandise data. His responsibility was to ensure that Tomorrowland merchandise sales were optimized. His insights helped me focus on the key metrics and trends that mattered most for merchandise sales.

Lastly, Niels van Dingenen was in charge of the podia data. His role was essential in overseeing the stages, and he provided the necessary information to ensure that my analyses were relevant and accurate.

Working with these individuals not only enriched my internship experience but also provided me with a comprehensive understanding of the various aspects of data management and visualization at EpicData.

7 Data analysis

During the first few weeks I focused primarily on the data analysis aspect. This step is crucial to get started and to get to know the data you are working with. The analysis was an ongoing, iterative process, crucial for uncovering deeper insights and enhancing my understanding of the festival's dynamics.

The primary objective of my data analysis was to understand/explore the data. This involved handling missing values, correcting inconsistencies, and eliminating duplicates, as well as converting data into suitable formats such as parsing dates, encoding categorical variables and recalculating values like revenue and cost.

Python notebooks were the primary tool used for data analysis, due to their interactive nature and integration with powerful data analysis libraries such as Pandas for data manipulation and NumPy for numerical computations.

7.1 Data Analysis Process

The data analysis process began with a comprehensive review of all datasets obtained from the CSV files provided by EpicData. These CSV files contain information about financial metrics, consumption, visitors, logistics of camping, merchandise, podia, safety & security, social media activity and pearls tokens. This initial step involved loading the datasets into a suitable data analysis environment, such as Python to facilitate easy manipulation and exploration.

When the data is loaded into my notebook environment I could start explore the data. it became evident that there were numerous aspects that were not immediately clear. The datasets contained ambiguous entries, missing values, wrong labelled rows, and inconsistencies that required further clarification to ensure accurate analysis. By organizing meetings with the client, I gained a deeper understanding of the data and start cleaning and transforming the data. This involved handling missing values, correcting inconsistencies, and eliminating duplicates, as well as converting data into suitable formats such as parsing dates and encoding categorical variables.

Data analysis was a continuous process due to the evolving nature of insights. As new patterns emerged, further analysis was required to validate and explore these findings. This iterative approach ensured that each round of analysis led to new questions and hypotheses, prompting additional data exploration and validation. The insights gained from one phase of analysis were used to refine subsequent analyses, deepening my understanding of the data.

7.2 Conclusion

The data analysis process provided a comprehensive understanding of the datasets and their underlying patterns. Through meticulous exploration, cleaning, and transformation, I was able to address ambiguities and ensure data integrity. Collaborating with the client through interactive Python notebooks and organized meetings further clarified the data's context and significance.

As a result, I was able to collect valuable insights and key trends from the data. These insights have been documented and will serve as the foundation for the creation of an informative and actionable dashboard. This dashboard will effectively communicate the findings, support data-driven decision-making, and provide a clear visualization of the critical metrics and trends identified during the analysis.

The collaborative and systematic approach to data analysis not only enhanced my understanding of the data but also ensured that the final deliverables align closely with the client's objectives and expectations.

8 Objectives for visualisations

After completing the data analysis, I outlined several objectives and ideas to guide what could be visualized. These objectives and ideas were then discussed with the client to ensure they aligned with their expectations and needs.

8.1 Merchandising

- Which merchandise items are sold the most and generate the most revenue?
- Does the design type influence the sales volume (limited edition/artist collab sold more)?
- Is there a difference in sales volume between the two weekends?
- Which items are sold the least/What percentage of the total stock is sold? Revenue per product segment.

8.2 Financial

- What are the costs per component per year? Do we observe increases/decreases?

8.3 Consumption

- What are the best-selling food items and drinks? Over the years, popular items.
- Visualize sales performance per year to determine if there is growth or decline in the consumption of items.
- At which food items and drinks is the most profit generated?
- Visualize the distribution of waste per item and identify items with a high waste percentage.

8.4 Individual visitors

- From which countries do the festival-goers come?
- What is the average age of festival-goers? Split into age categories.
- What mode of transportation do the festival-goers use? Do we see a change over the years?
- What is the most used payment method?
- What was the favorite stage?
- Distribution of festival tickets (VIP, Regular, Day pass)
- Where do most people stay overnight?
- Key Performance Indicator (KPI) for visitors compared to the previous year in percentage.

8.5 Social media activity

- On which platforms are the messages viewed the most? Interesting for advertising posts.
- Is there a link between artists and the positivity/likes of a post?
- Does the platform, tagged artists, and hashtags influence the views/likes of a post?
- Where are most of the posts posted from?
- Do likes influence the visual content or length of a post?

8.6 LogisticsCamping

- Look into suggestions regarding transportation to see if improvements can be made?
- Transportation cost only with metric transportation.
- Provide alternative suggestions when cost is high.
- Cost and pearls spent.

8.7 PodiaSetup

- Factors that influence setup time such as crew size, amount of equipment.
- Evolution of setup times, examining crew size cost per stage.
- Influence of weather conditions on setup time.
- Food traffic: number of people who can comfortably go to the beverage stand from the stage within X number of minutes in 1 minute.
- What are the most popular stages?
- Most popular music type?

9 Power BI

Power BI is a powerful business intelligence tool developed by Microsoft that allows users to visualize and analyse data from various sources. It provides a suite of features for data preparation, modelling, visualization, and collaboration, making it a comprehensive solution for both business users and data professionals.

One of Power BI's key strengths is its ability to connect to a wide range of data sources, including databases, spreadsheets, cloud services, and online services. Users can easily import data into Power BI and transform it using intuitive tools for cleaning, shaping, and modelling.



Once the data is prepared, Power BI offers a variety of visualization options to help users create insightful reports and dashboards. These visualizations include bar charts, line graphs, pie charts, maps, and more, which can be customized to suit specific needs. Users can also create interactive elements such as slicers, filters, and drill-downs to explore data in depth.

Power BI's capabilities extend beyond individual reports to enable users to create comprehensive dashboards that consolidate multiple visualizations and insights into a single view. These dashboards can be shared with others within the organization or embedded into other applications for broader accessibility.

Furthermore, Power BI incorporates advanced analytics features such as predictive modelling, natural language queries, and machine learning integration, allowing users to uncover deeper insights and make data-driven predictions.

9.1 Usage

Using Power BI, I crafted a comprehensive dashboard featuring a diverse array of graphs and visualizations. This dashboard encompasses a wide range of topics, ensuring that all pertinent aspects are covered to generate valuable insights.

The dashboard comprises various types of graphs, including bar charts, line graphs, pie charts, and more, each meticulously selected to effectively represent different data sets and metrics. These visualizations offer a holistic view of the underlying data, facilitating a deeper understanding of trends, patterns, and correlations.

With interactive features such as filters, slicers, and drill-down capabilities, users can explore the data in detail, uncovering hidden insights and identifying key drivers of performance and success. The dashboard is designed to be user-friendly and intuitive, allowing stakeholders to effortlessly navigate through the information and extract meaningful insights.

Whether it's optimizing marketing strategies, improving operational efficiency, or enhancing customer satisfaction, the insights derived from this Power BI dashboard empower decision-makers to make informed choices and drive business growth.

Through the self-paced training course provided by Microsoft, I gained my first-hand experience with Power BI. Prior to this, I had no prior exposure to the tool. However, by completing the Power BI Data Analyst Associate course, I acquired a solid foundation in Power BI's functionalities and capabilities.

With this newfound knowledge, I was able to create a dynamic and informative dashboard using Power BI. This dashboard has empowered me to make better data-driven decisions by visualizing insights in a clear and actionable manner.

9.2 Data model

Creating a data model is a crucial step before starting any report in Power BI. A well-structured data model serves as the foundation for your analysis, ensuring that your data is organized, accurate, and efficient to work with. This involves defining relationships between different data tables, typically categorized into fact tables and dimension tables. Fact tables store quantitative data and metrics, such as sales amounts or transaction counts, while dimension tables contain descriptive attributes related to the facts, like customer information, product details, or time periods.

Establishing these relationships and organizing your data into fact and dimension tables allows for seamless data integration and enhances performance. It simplifies the creation of insightful visualizations by enabling easy filtering, grouping, and drilling down into data. A robust data model ensures that your reports are reliable, scalable, and easy to maintain, ultimately leading to more effective data-driven decision-making.

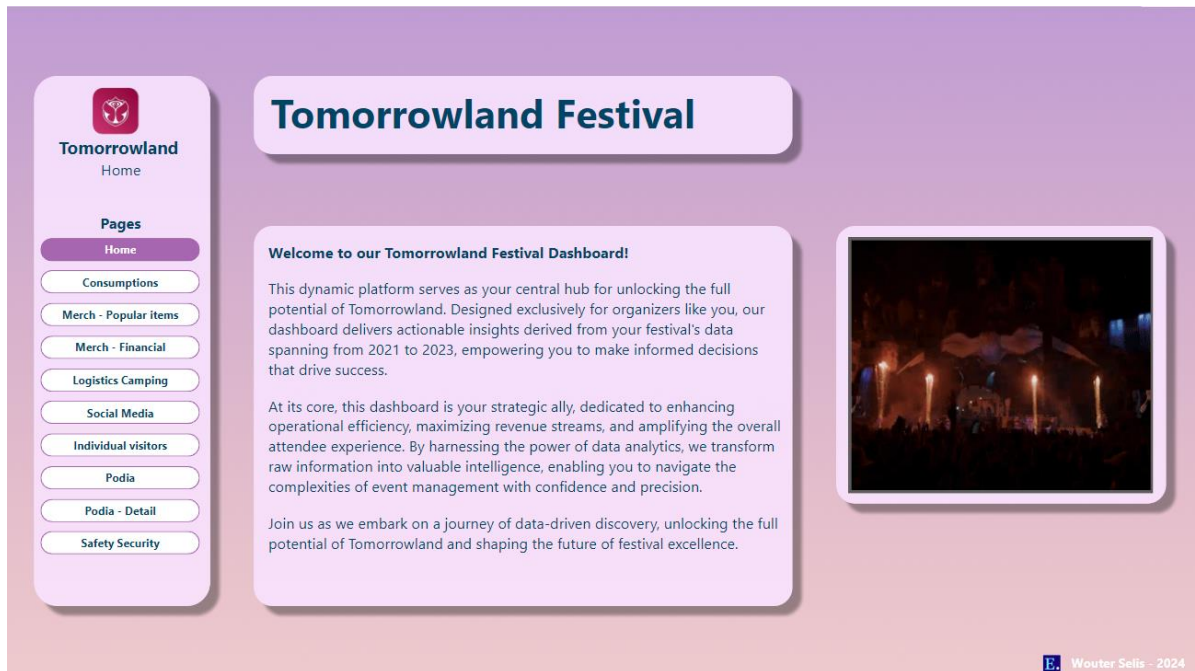
Without a solid data model, reports may suffer from inconsistencies, inaccuracies, and inefficiencies, undermining the value of your analysis. Therefore, investing time in building a strong data model with clear distinctions between fact and dimension tables is essential for maximizing the potential of Power BI reports.

(See data model on next page)

9.3 Dashboard

9.3.1 Home

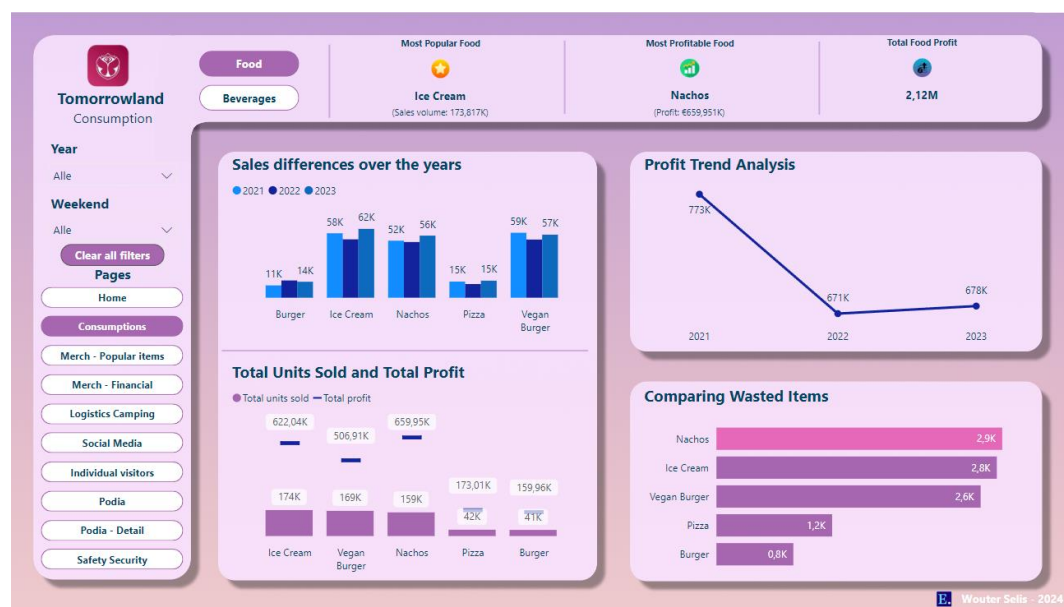
The user first lands on the Home page of my dashboard. On this page there is an introduction and description of my dashboard. On the left side you can navigate to all the other pages.



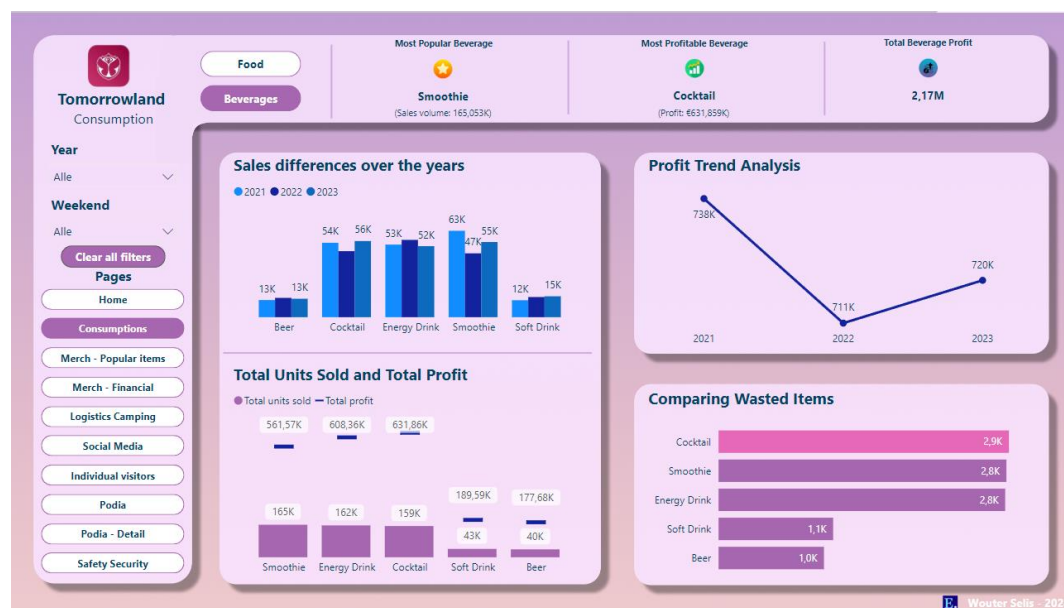
9.3.2 Consumption

By understanding consumption trends across multiple years, Tomorrowland can identify shifting preferences, popular food and beverage items, and areas for improvement in its offerings. This analysis enables Tomorrowland to adjust its consumption item selections, pricing strategies, and inventory management to meet the evolving tastes and demands of festival attendees, ultimately enhancing the overall food and beverage experience and driving increased sales revenue. By identifying trends in wasted inventory, such as unsold food items or expired beverages, Tomorrowland can implement measures to reduce waste and optimize inventory management processes.

Food:



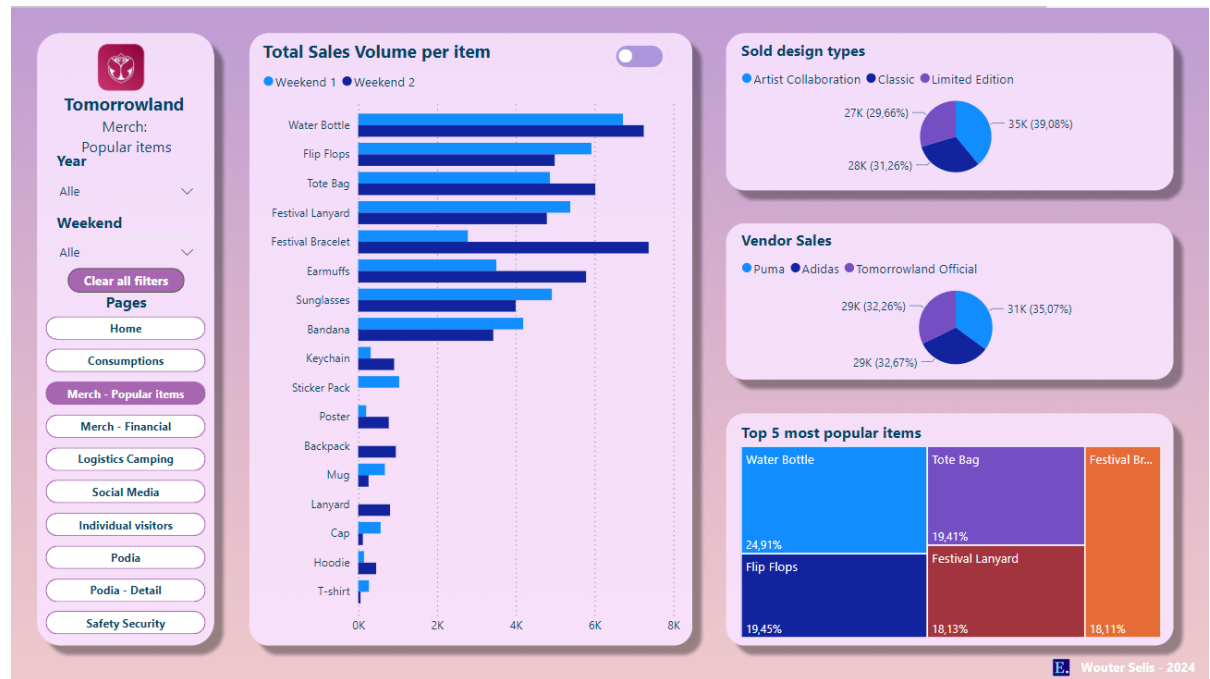
Beverages:



9.3.3 Merch – popular items

Analysing merchandise sales data at Tomorrowland provides invaluable insights into consumer preferences, inventory management, and vendor performance. By understanding their data Tomorrowland can optimize its merchandising strategy, enhance the festival experience for attendees, and drive increased revenue through targeted inventory selection, strategic partnerships, and personalized marketing efforts.

Toggle off:



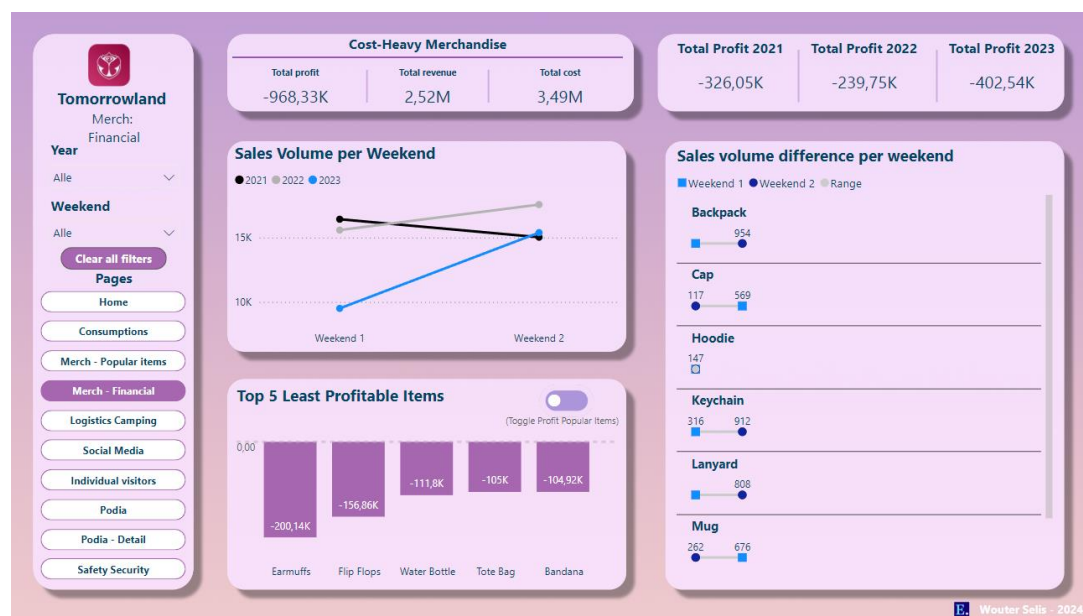
Toggle on:



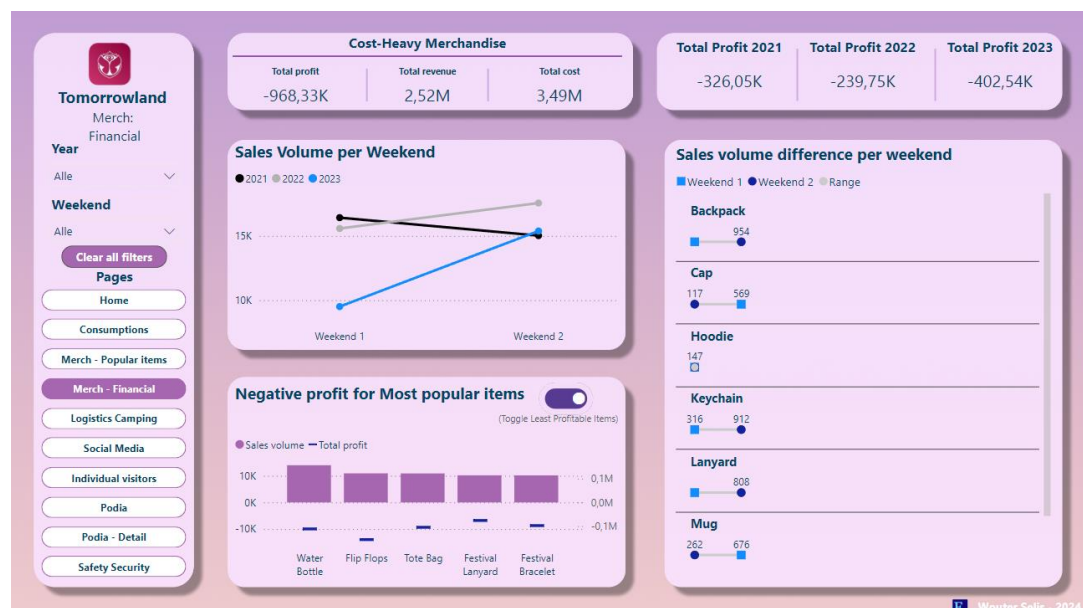
9.3.4 Merch – Financial

Understanding the financial performance of merchandise sales at Tomorrowland is supreme for strategic decision-making. Analysing total profit, revenue, and cost, alongside year-over-year comparisons, provides insights into profitability trends and operational efficiency. Sales volumes per weekend and differences in sales volume offer granular insights into consumer behaviour and demand fluctuations, informing inventory management and marketing strategies. Identifying the top 5 least profitable items allows Tomorrowland to decrease losses, optimize product offerings, and enhance overall profitability, ensuring sustainable growth and financial success in the merchandise sector.

Toggle off:



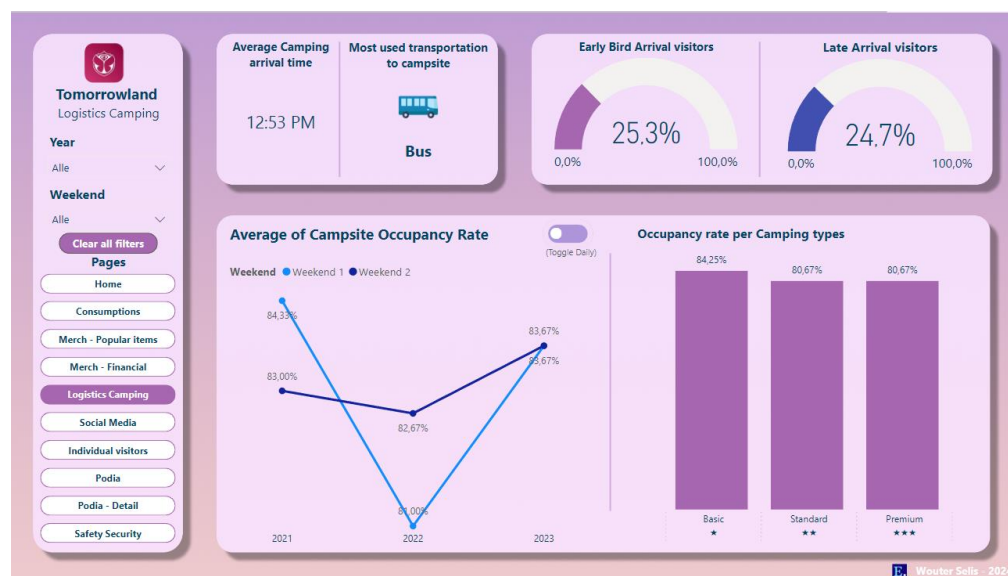
Toggle on:



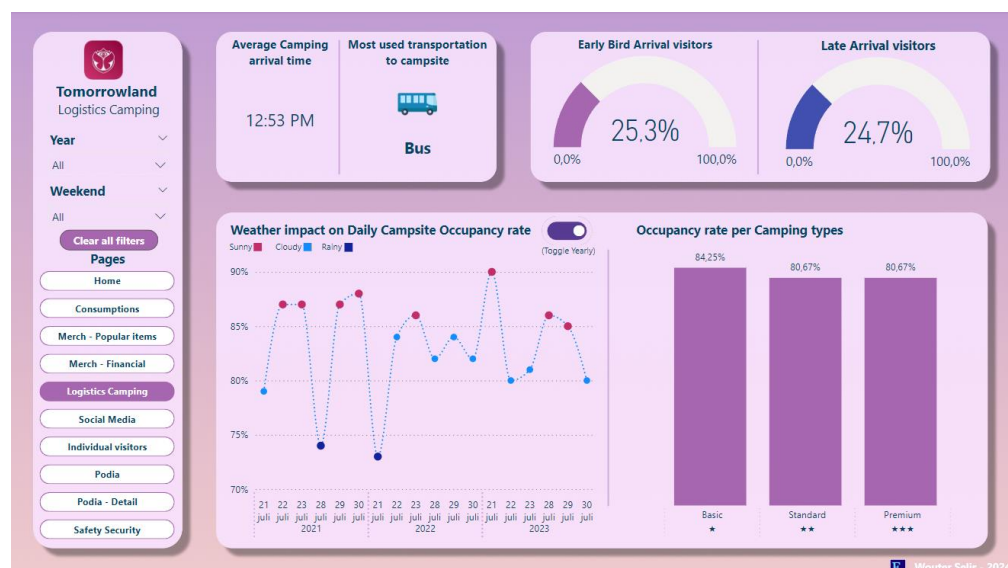
9.3.5 Logistics Camping

Understanding logistics and camping trends at Tomorrowland is crucial for optimizing the attendee experience and operational efficiency. Analysing average camping arrival time, transportation preferences, early bird and late arrival visitor percentages, and campsite occupancy rates provides insights into attendee behaviour and demand patterns. Assessing the impact of weather on daily campsite occupancy rates allows for proactive planning and resource allocation, ensuring smooth operations despite external factors. Furthermore, analysing occupancy rates for the different camping types enables Tomorrowland to tailor amenities and services to different camper preferences, ultimately enhancing satisfaction and ensuring a seamless camping experience for all attendees.

Toggle off:



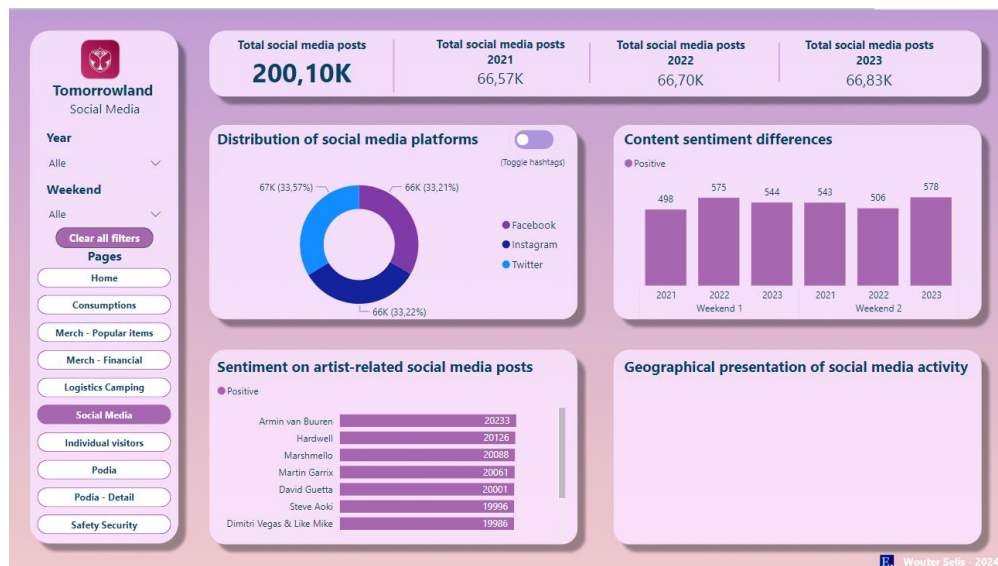
Toggle on:



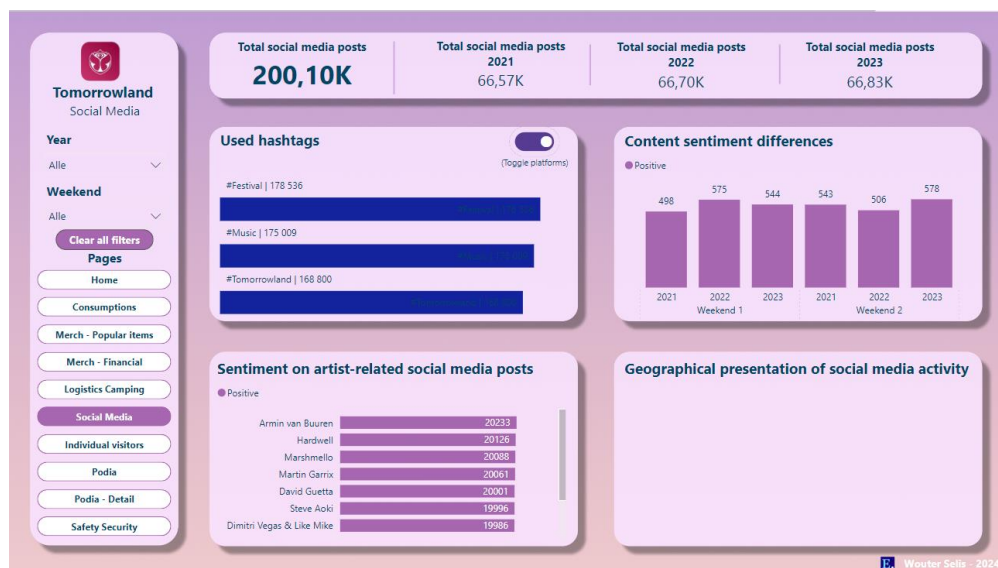
9.3.6 Social Media

Analysing social media activity data provides Tomorrowland with valuable insights into its digital presence and audience engagement. Understanding the total social media posts per year, platform distribution, and content sentiment differences helps gauge overall performance and identify trends. Sentiment analysis on artist-related posts offers deeper insights into fan engagement and preferences. Geographical analysis of social media activity reveals regional interest and potential markets, while hashtag usage patterns highlight popular themes and topics. These insights enable Tomorrowland to refine its social media strategy, target advertising effectively by knowing which platforms to focus on, enhance audience interaction, and drive targeted marketing efforts, ultimately boosting brand visibility and fan loyalty.

Toggle off:

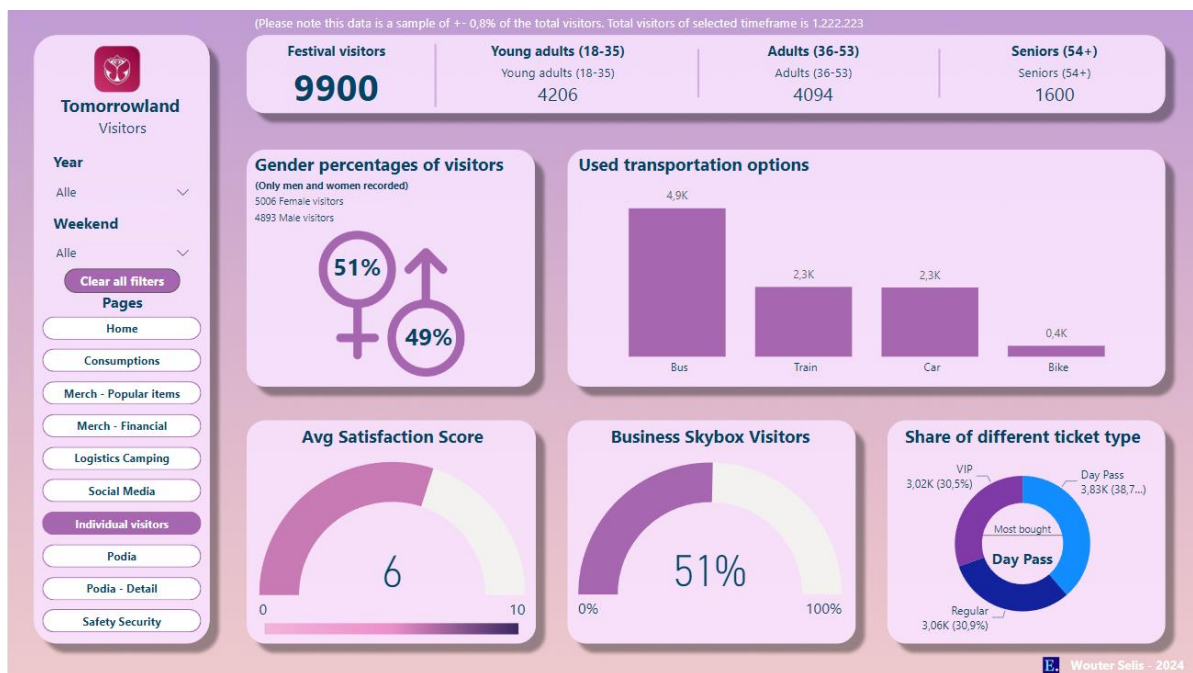


Toggle on:



9.3.7 Individual visitors

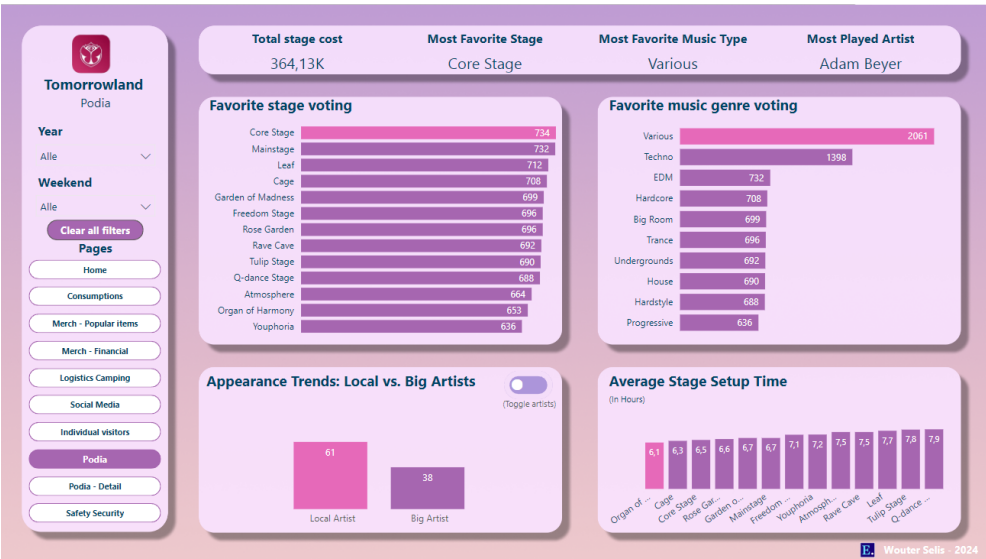
Analysing visitor data provides Tomorrowland with crucial insights into attendee demographics and behaviour. Understanding total festival visitors, age group distribution, and gender distribution helps tailor marketing and event experiences to diverse audience segments. Examining transportation options used by attendees and average satisfaction scores informs logistical planning and overall festival improvements. Data on business skybox visitors, different ticket type sales, and the most bought ticket type enables Tomorrowland to optimize ticketing strategies and enhance revenue. These insights collectively support targeted marketing, efficient operations, and improved attendee satisfaction, driving the festival's success and growth.



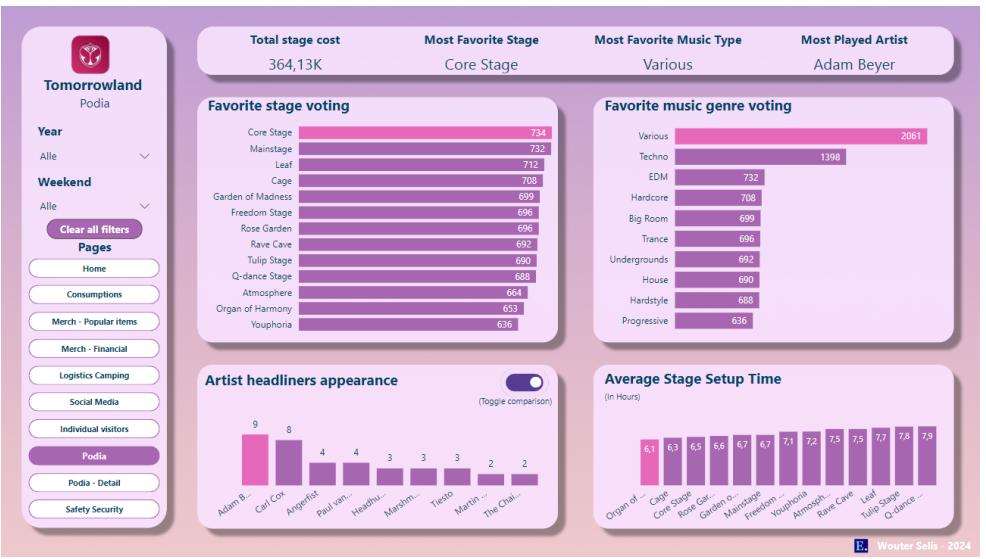
9.3.8 Podia

Analysing stage data at Tomorrowland provides essential insights into the festival's production and audience preferences. Understanding the total stage costs helps optimize budgeting and resource allocation. Identifying the most favourite stages and music types, as well as the most played artists and voting results for favourite stages and genres, reveals attendee preferences and enhances programming decisions. Tracking artist appearances and stage setup times informs logistical planning and operational efficiency. These insights enable Tomorrowland to enhance the attendee experience, streamline operations, and allocate resources effectively, ensuring a memorable and well-organized festival.

Toggle off:



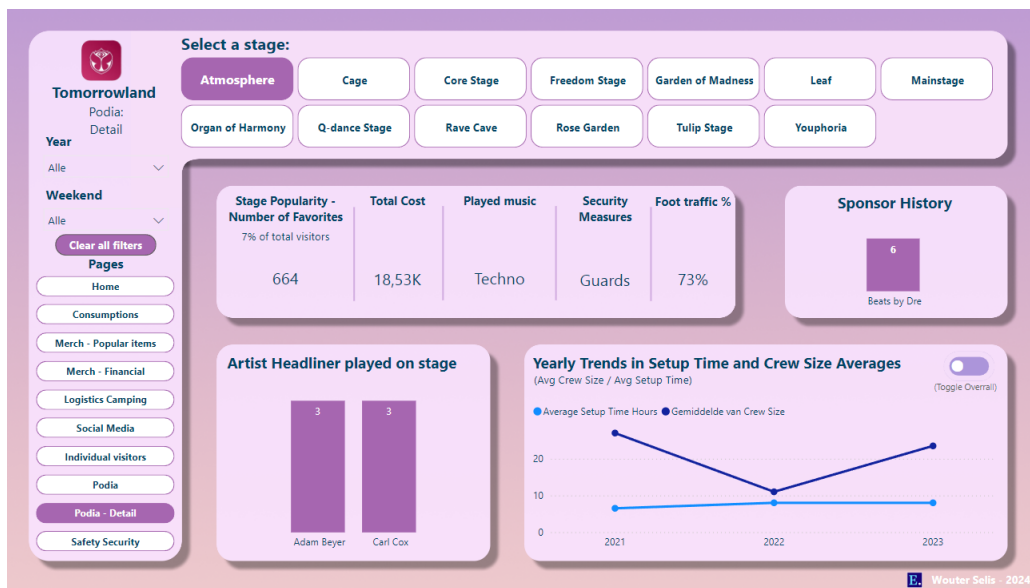
Toggle on:



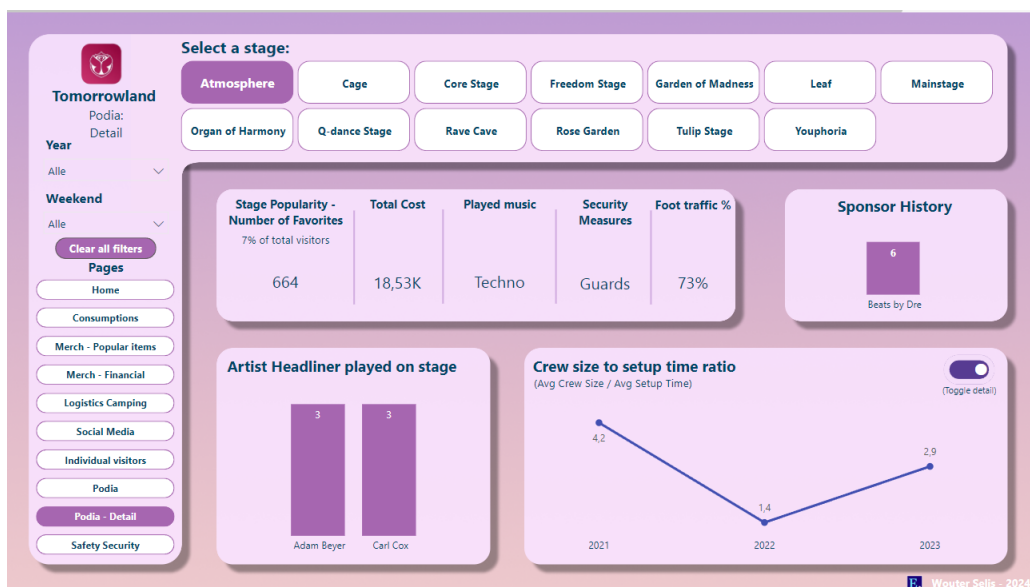
9.3.9 Podia – Detail

The stage selection page provides Tomorrowland with detailed insights into each stage's performance and operational details. By allowing users to view information such as stage popularity, total cost, music genres played, security measures, foot traffic, sponsor history, headlining artists, and yearly trends in setup time and crew averages, this page offers a comprehensive view of each stage's impact and requirements. These insights help Tomorrowland optimize stage management, improve audience satisfaction, enhance security protocols, attract sponsors, and ensure efficient stage setups, ultimately contributing to a smoother and more successful festival experience.

Toggle off:



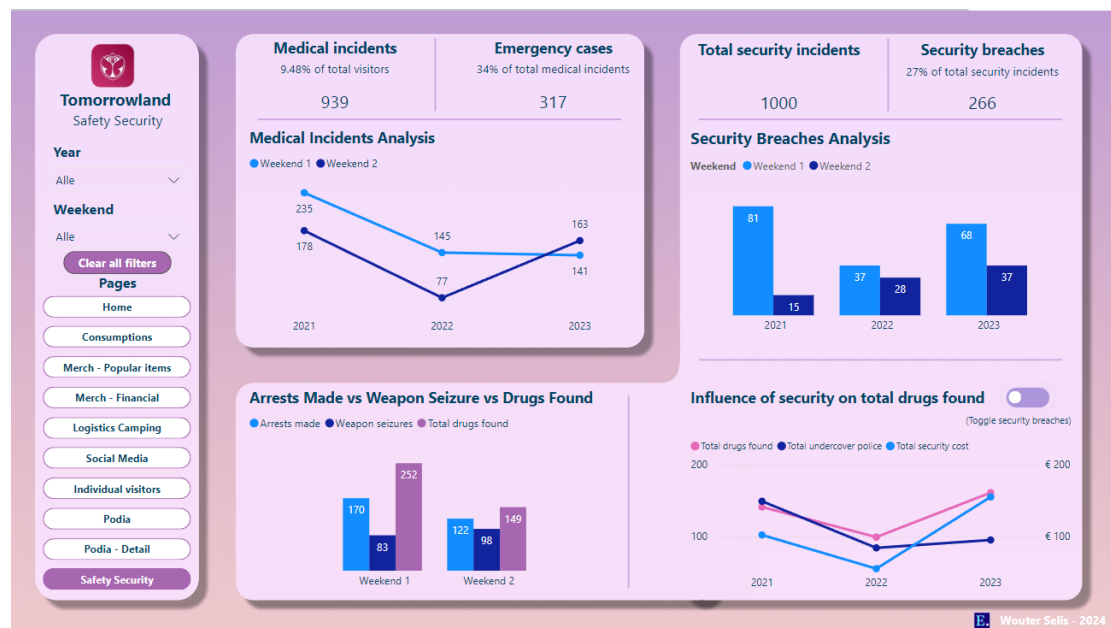
Toggle on:



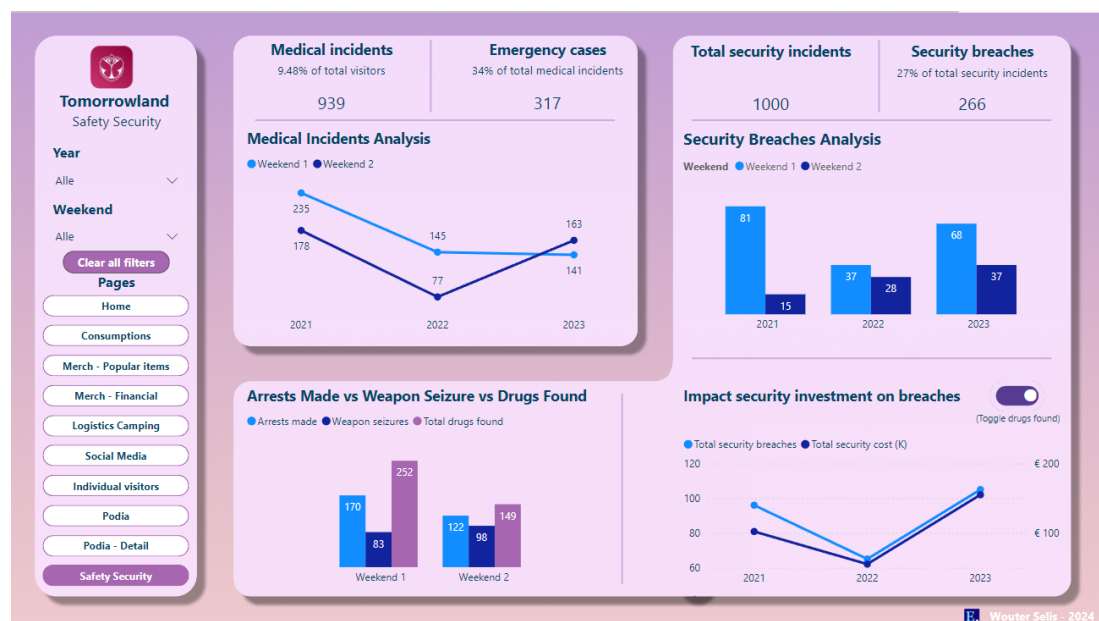
9.3.10 Safety Security

The safety and security insights page provides Tomorrowland with comprehensive data on festival-wide safety and security performance. By allowing users to view information such as medical insights and security insights, this page offers a holistic understanding of the festival's safety landscape. Additionally, analysing the impact of security investments on breaches helps Tomorrowland optimize safety protocols, enhance emergency response, allocate resources effectively, and ensure a secure environment for all attendees, ultimately improving overall festival safety and satisfaction.

Toggle off:



Toggle on:



10 Data Engineering

Data engineering is essential for modern data analysis and business intelligence, involving the design, construction, and maintenance of infrastructure and systems for collecting, storing, processing, and analysing data. Its main goals include ensuring efficiency, scalability, data quality, and integration across various sources.



databricks

Databricks, a powerful platform built on Apache Spark, facilitates these processes by providing a collaborative environment for data engineers and data scientists to work together seamlessly. With features like scalability, collaboration, integrated data processing, security, and machine learning support, Databricks streamlines the entire data lifecycle from extraction to analysis. Leveraging data engineering with tools like Databricks will be essential in preparing and processing data for analysis and reporting, enhancing efficiency, collaboration, and innovation in your data engineering workflows.

During my internship, I utilized the powerful data engineering tool Databricks to streamline my data processing workflows. One of the initial steps involved loading my local CSV files directly into the tool, storing them under the default folder in the `hive_metastore`. This `hive_metastore` acts as a central metadata repository in the Apache Hive data warehouse system, managing schema information and facilitating data querying and processing. Once the data was imported, I proceeded to create a separate notebook for each dataset within Databricks. These notebooks served as collaborative workspaces where I could clean and transform the data using a combination of PySpark and SQL. Leveraging PySpark's powerful data manipulation capabilities and SQL's intuitive querying syntax, I structured the data into fact and dimension tables, ensuring its readiness for subsequent analysis and visualization tasks. Through this approach, I was able to efficiently manage and prepare the data for further exploration and insights generation.

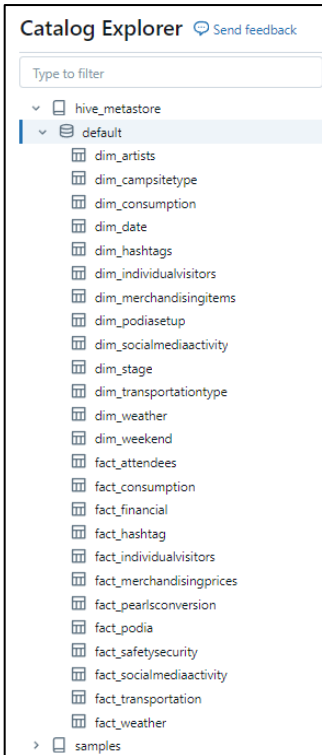
Cluster:

The image below shows the "Compute" section of my Databricks workspace, highlighting an all-purpose compute cluster named "Wouter Selis's Cluster." This cluster is configured with runtime version 13.3 and is currently active with a DBU usage of 4. All-purpose clusters in Databricks are used for interactive data exploration, running ad-hoc queries, and executing notebooks, providing the necessary computational resources for a variety of data processing tasks. This interface allows users to create, monitor, and manage their clusters efficiently.
































Catalog:

The image below shows the "Catalog Explorer" within the Databricks workspace, highlighting the tables I created under the hive_metastore in the default schema. These tables include both dimension (dim) and fact (fact) tables. I organized these tables to support various aspects of my data analysis project, from tracking individual visitors and their activities to analyzing financial and weather data. The Catalog Explorer provides an intuitive interface to navigate and manage these datasets, making it easier for me to perform efficient data exploration and querying within the Databricks environment.



Notebooks:

The image below displays a collection of notebooks I created in the Databricks workspace. Each notebook is dedicated to a specific aspect of my data analysis project, such as analyzing artist data, attendee information, camping details, and weather conditions. These notebooks serve as interactive documents where I can combine code, visualizations, and narrative text to explore and analyze my datasets comprehensively. They allow me to execute data processing tasks and perform complex queries. By organizing my work into these notebooks, I ensure a structured and efficient workflow, facilitating collaboration and reproducibility within the Databricks environment.

Name 	Type	Owner	Created at	
 TML-Artists	Notebook	Wouter Selis	2024-05-07 20:27:14	
 TML-Attendees	Notebook	Wouter Selis	2024-05-14 15:29:49	
 TML-Camping	Notebook	Wouter Selis	2024-05-15 11:35:41	
 TML-Consumption	Notebook	Wouter Selis	2024-05-10 10:43:30	
 TML-Date	Notebook	Wouter Selis	2024-05-08 14:08:28	
 TML-Financial	Notebook	Wouter Selis	2024-05-14 15:41:35	
 TML-Individual_Visitors	Notebook	Wouter Selis	2024-05-08 14:35:05	
 TML-Merchandise	Notebook	Wouter Selis	2024-05-15 10:49:10	
 TML-Pearls	Notebook	Wouter Selis	2024-05-14 16:09:26	
 TML-Podia	Notebook	Wouter Selis	2024-05-10 15:17:05	
 TML-SafetySecurity	Notebook	Wouter Selis	2024-05-15 16:15:05	
 TML-SocialMedia	Notebook	Wouter Selis	2024-05-14 10:57:33	
 TML-Weather	Notebook	Wouter Selis	2024-05-08 11:18:40	
 TML-Weekend	Notebook	Wouter Selis	2024-05-08 14:28:36	

11 Hackaton

As part of the SMEP! - Shared Mobility Equity Principles project, SMEP! challenged students from Belgian universities and colleges to formulate policy proposals that could make the use of public transportation and shared mobility more inclusive and equitable for women. As part of this project, Mpact organized an online hackathon for students on February 29, 2024, to help narrow the gender gap in public transportation and shared mobility. Following the event, the visions were presented to key national and international experts, and the top 3 groups received awards.

With three other interns at EpicData we formed a group to participate to this Hackaton. We created an action plan to enhance gender-inclusive access to shared mobility solutions and public transport.

To address these challenges and promote gender-inclusive access to transportation, our team proposed MOBUDDY, an innovative AI-powered application system that integrates policy reform and technological innovation to ensure safety, accessibility, and affordability in transportation. This holistic approach aligns with Sustainable Development Goal 5 (SDG5), aiming to empower women and girls.

