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

Project plan

Bachelor Applied Computer Science

Wouter Selis

Academic year 2023-2024

Campus Geel, Kleinhoefstraat 4, BE-2440 Geel

LID VAN ASSOCIATIE KU LEUVEN

THOMAS MORE

Project plan

Content

1 Company introduction

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.

2 Introduction and background of the internship project

The datasets are AI generated, they are not real datasets from Tomorrowland.

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.

The project is performed for a fictitious client. Several EpicData employees have been assigned as responsible for a particular department (EpicEvents, EpicSocials, EpicDrinks, EpicLogistics, EpicMerch and EpicPodia). Listed below are all my contacts for this project with their positions.

Henny.speelman@epicdata.be
Event manager of 'EpicEvents', facilitating all data from Tomorrowland.

Timmy.diricx@epicdata.be
Chief Logistics officer of 'EpicLogistics', making sure people of Tomorrow can sleep and rave before the stages.

Neil.Vets@epicdata.be
Marketing guru of EpicSocials', making sure the web analytics are up to date.

Bart.vanmulders@epicdata.be
Chief Brand Ambassador of 'EpicMerch', making sure people buy Tomorrowland merchandise.

Anthony.coppens@epicdata.be
Operations officer of 'EpicDrinks', making sure people of Tomorrow can eat and drink on the festival.

Niels.vandingenen@epicdata.be
General manager of 'EpicPodia', delivering the right podium for your festival.

3 Objectives

The objective of my internship is to provide actionable insights into various aspects of the Tomorrowland festival, utilizing Power BI. By leveraging the data visualization and analysis capabilities of Power BI, I aim to enhance operational efficiency, maximize revenue, and improve the overall attendee experience. This involves not only creating comprehensive visualizations that highlight key trends and insights but also developing data-driven strategies that can be implemented to address specific challenges faced by the festival. Through detailed data analysis and interpretation, my goal is to support Tomorrowland in making informed decisions that lead to a more efficient and enjoyable festival for all involved.

4 Business Case

The business case aims to provide actionable insights to enhance operational efficiency, maximize revenue, and improve the overall attendee experience. The scope of the project includes creating a comprehensive Power BI dashboard that covers a wide range of metrics and data points.

The report will focus on financial metrics to track revenue streams and expenditures, providing a clear picture of the festival's economic performance. Consumer behaviour analysis will offer insights into attendee preferences and spending patterns, helping to tailor offerings and marketing strategies. Operational logistics will be examined to identify areas for efficiency improvements, ensuring smooth and effective festival operations. Social data will be analysed to measure public sentiment and engagement, offering a thorough understanding of the festival's impact on its audience.

Additionally, attendance data will be analysed to monitor crowd dynamics and optimize space utilization, while safety and security metrics will be included to ensure the well-being of all attendees. By integrating these diverse data sets into a cohesive and interactive dashboard, the project aims to support Tomorrowland in making informed decisions that drive success.

5 Information and reporting

To ensure a transparent communication structure, I provide weekly updates to both my internship supervisor and my mentor. These updates provide a detailed overview of my daily work, keeping them constantly informed of my progress and the specific tasks I am working on.

Furthermore, weekly meeting with my internship mentor will be held so that he too is aware of the progress of my work.

In addition, I will implemented a Jira board to manage the project according to the agile methodology. This board will act as a visual aid showing the progress of tasks, assigned responsibilities and any obstacles. The use of Jira will not only promote streamlined collaboration, but it will also allow my mentor and supervisor to access the current status of the project at any time.

During the third week of the internship, the internship supervisor will visit EpicData to discuss the progress and future steps of the internship with me and my mentor. Additionally, there are two online group meetings scheduled with the internship supervisor. In the first meeting, a brief presentation will be given about the internship company and the project to be carried out.

In the second meeting, which will take place around the seventh week, the Plan of Action will be reviewed. Additionally, a brief update will be provided on the completed tasks and the planned activities.

Through this combined approach of regular updates and the use of Jira, I strive for effective communication that provides a clear understanding of project progress, potential challenges and milestones achieved. This strategy will increase the involvement of all stakeholders and ensure a smooth and efficient execution of the BI project.

6 Elaboration & Techniques

6.1 Data engineering

In this document, I describe the data engineering processes performed in Databricks. The goal is to create a scalable and efficient data infrastructure that ensures accurate and up-to-date data representation in the dashboard.

The first step in this process is cleaning and preparing the data, ensuring the data is usable and error-free through data cleaning and transformation processes in Databricks. Next, we apply Natural Language Processing (NLP) techniques to the data to gain deeper insights. We then develop fact and dimension tables that serve as fundamental components of a star schema, which helps make analytical queries more efficient.

These data engineering processes will result in centralized storage that we can connect with the Power BI dashboard. By centralizing the cleaned and processed data, we create a unified source of truth that facilitates seamless integration with Power BI. This integration enables the visualization and analysis of data through interactive dashboards, providing comprehensive insights and supporting data-driven decision-making.

In the first phase, we import raw data into Databricks, perform data cleaning by removing duplicates, handling missing values, and standardizing data types. We use PySpark and SQL to transform and prepare the data for further modelling. The result is clean and structured data in Databricks.

In the second phase, we analyse text data using Python and NLP libraries to gain insights. We perform sentiment analysis, topic modelling, and other relevant NLP tasks, integrating the obtained insights into the dataset, resulting in an enriched dataset with NLP insights.

In the third phase, we identify the necessary fact and dimension tables, design and create these tables in Databricks, and ensure the integrity and relationships between the tables. This results in fact and dimension tables ready for use in the data warehouse.

6.2 Visualisations

Here, I describe how I will create a report of various visualizations to help the client gain insights into their data. I have chosen to develop my visualizations with Power BI because it is a widely used tool and I want to take this opportunity to improve my skills in it. Since I have not worked with this tool before, I see this as an additional challenge that I am eager to take on to expand my knowledge.

The main goal of the report is to improve decision-making by providing insights into merchandise sales, consumption patterns, and visitor behaviour. This includes identifying poorly performing consumption and merchandise items, determining the most popular items, and analysing the differences in consumer behaviour between the two weekends.

Additionally, I aim to increase revenue through optimized pricing and strategies for food, drinks, and merchandise. The report will also provide insights into camping logistics, such as the average arrival time of festivalgoers, the most chosen transportation option, and the most popular camping option, with the goal of attracting more people to the premium camping area.

Another key aspect is social media analysis, where I will identify the platforms with the most posts, which can be useful for future advertising campaigns. I will also evaluate the visitors' sentiment about certain artists and identify which countries post the most about the festival.

The analysis of festivalgoers will provide a better understanding of the target audience, such as young adults, adults, and seniors, as well as different genders. The report will also examine the distribution of ticket sales and the satisfaction scores of the festivalgoers.

Furthermore, I will analyse the stages, investigating the average setup time and the factors influencing it. This will help cater to the favourite music genres of the festivalgoers.

Finally, the report will focus on improved safety by monitoring security and medical incidents, security breaches, and criminal activity, such as drugs, weapons, and arrests.