

Project Title:

Exploring Insights from
Synthetic Airline Data
Analysis with Qlik

Submitted By:

ASHNOOR MAAN

SRM Institute of Science and
Technology

Kattankulathur

Contents

1. Problem statement
2. Loading the dataset
3. Data preprocessing
4. Data visualization
5. Dashboard
6. Story
7. Project documentation and demonstration
8. Conclusion

Chapter 1 Problem statement

The project "Exploring Insights from Synthetic Airline Data Analysis with Qlik" involves utilizing synthetic airline data to derive valuable insights using Qlik, a business intelligence and data visualization tool. In this project, the synthetic airline data simulates various aspects of airline operations, including flight schedules, passenger demographics, ticket sales, and performance metrics. The objective is to leverage Qlik's analytical capabilities to uncover patterns, trends, and correlations within this data, aiding in decision-making processes for airlines, airports, and related stakeholders.

Airline data holds immense importance as it offers insights into the functioning and efficiency of the aviation industry. It provides valuable information about flight routes, schedules, passenger demographics, and preferences, which airlines can leverage to optimize their operations and enhance customer experiences. By analyzing data on delays, cancellations, and on-time performance, airlines can identify trends and implement strategies to improve punctuality and mitigate disruptions. Moreover, regulatory bodies and policymakers rely on this data to ensure safety standards, enforce regulations, and make informed decisions regarding aviation policies. Researchers and analysts use airline data to study market trends, assess environmental impacts, and develop strategies for sustainable growth within the industry. In essence, airline data serves as a foundation for informed decision-making, operational efficiency, and the overall advancement of the aviation sector.

Chapter 2 Loading the dataset

1. The project flow provided the link to the dataset.
2. The link redirected to a kaggle page about the airline dataset.
3. The airline dataset contains information about:
 - Passenger ID
 - First Name
 - Last Name
 - Gender
 - Age
 - Nationality
 - Airport Name
 - Airport Country Code
 - Country Name
 - Airport Continent
 - Continents
 - Departure Date
 - Arrival Airport
 - Pilot Name
 - Flight Status

Chapter 3 Data Preprocessing

Data Preprocessing can be defined as a process of converting raw data into a format that is understandable and usable for further analysis. It is an important step in the Data Preparation stage. It ensures that the outcome of the analysis is **accurate**, **complete**, and **consistent**.

Data preprocessing is an important step in the data mining process that involves cleaning and transforming raw data to make it suitable for analysis. Some common steps in data preprocessing include:

Data Cleaning: This involves identifying and correcting errors or inconsistencies in the data, such as missing values, outliers, and duplicates. Various techniques can be used for data cleaning, such as imputation, removal, and transformation.

Data Integration: This involves combining data from multiple sources to create a unified dataset. Data integration can be challenging as it requires handling data with different formats, structures, and semantics. Techniques such as record linkage and data fusion can be used for data integration.

Data Transformation: This involves converting the data into a suitable format for analysis. Common techniques used in data transformation include normalization, standardization, and discretization. Normalization is used to scale the data to a common range, while standardization is used to transform the data to have zero mean and unit variance. Discretization is used to convert continuous data into discrete categories.

Data Reduction: This involves reducing the size of the dataset while preserving the important information. Data reduction can be achieved through techniques such as feature selection and feature extraction. Feature selection involves selecting a subset of relevant features from the dataset, while feature extraction involves transforming the data into a

lower-dimensional space while preserving the important information.

In the given dataset, all the columns needed to be renamed. the values needed to be sorted and some were missing a value wich needed to either be filled or the entire row needed to be removed, or you can just leave it as is, if its not an important value.

Chapter 4 Data visualization

Raw data is not insightful. Data needs to be converted into information for it to be useful. one of the easiest ways of turning raw data into information is data visualization.

Data visualization is the graphical representation of information and data in a pictorial or graphical format (Visualization of Data could be: charts, graphs, and maps). Data visualization tools provide an accessible way to see and understand trends, patterns in data, and outliers. Data visualization tools and technologies are essential to analyzing massive amounts of information and making data-driven decisions. The concept of using pictures is to understand data that has been used for centuries. General types of data visualization are Charts, Tables, Graphs, Maps, and Dashboards.

The process of putting data into easily readablegraphics such as graphs, plots, animations and values (such as min, max, average) is data visulaization. Data visualization dashboards can impart a lot of information in v ery little time and space.

While creating data visualizations we must be careful to use colours which are interesting to look at but also do not strain the eyes.

Why is Data Visualization Important?

Data Visualization Discovers the Trends in Data

Data Visualization Provides a Perspective on the Data

Data Visualization Puts the Data into the Correct Context

Data Visualization Saves Time

Chapter 5 Dashboard

Creating the dashboard with all the data visualizations is arguably the most important part of the whole project.

Dashboards contain all different kinds of data visualizations. Dashboards can impart a lot of knowledge in a matter of seconds.

A dashboard is a way of displaying various types of visual data in one place. Usually, a dashboard is intended to convey different, but related information in an easy-to-digest form. And oftentimes, this includes things like key performance indicators (KPI)s or other important business metrics that stakeholders need to see and understand at a glance.

Both dashboards and reports are commonly utilized to collect and analyze data. So what makes them different?

Broadly speaking, **reports** usually have a more narrow focus. They serve the purpose of providing a deep-dive view into a data set and tend to concentrate on a single item or event.

On the other hand, **dashboards** tend to have a high-level view of broad amounts of data and are created to answer a single question. That question can be broad, such as, “how was our site performance last month?” Or more specific, such as, “how many units did we sell?” Or perhaps something that’s a little harder to track without specialized expertise, such as, “is our overall efficiency improving?”

Two common terms when it comes to analytics and reporting are “data dashboard” and “data visualization.” What’s the difference?

Data visualization is a way of presenting data in a visual form to make it easier to understand and analyze.

Data dashboards are a summary of different, but related data sets, presented in a way that makes the related information easier to understand. Dashboards are a type of data visualization, and often use common visualization tools such as graphs, charts, and

tables.

Chapter 6 Story

Data storytelling is very similar to human storytelling but provides the added benefits of deeper insights and supporting evidence through graphs and charts. Through data storytelling, complicated information is simplified so that your audience can engage with your content and make critical decisions quicker and more confidently.

Constructing a data story that moves a person to take action can be a very powerful tool. Effective data storytelling can have a positive impact on people and your organization. Some benefits of successful data storytelling include:

- Adding value to your data and insights.
- Interpreting complex information and highlighting essential key points for the audience.
- Providing a human touch to your data.
- Offering value to your audience and industry.
- Building credibility as an industry and topic thought leader.

Data visualization plays a significant role in determining how receptive your audience is to receiving complex information. Data visualization helps transform boundless amounts of data into something simpler and digestible. Here, you can supply the visuals needed to support your story. Effective data visualizations can help:

- Reveal patterns, trends, and findings from an unbiased viewpoint.
- Provide context, interpret results, and articulate insights.
- Streamline data so your audience can process information.
- Improve audience engagement.

Chapter 7 Project documentation and demonstration

The last part of the project is documentation and demonstration.

The documentation part is this document which takes you through the project flow step by step explaining the importance of every step and explaining how each step is executed.

The demonstration part of the project is the video. It takes you through the project step by step with audio description. Starting from the raw dataset, to the processed data, data visualization, dashboard and story telling, everything is included in the project demonstration video.

Chapter 8 Conclusion

In the end, we can conclude that this project provided a great insight into how business analytics works in the real world. Going through with this project, every step gave informative insights into every step of the process.

In this project, we aimed to analyze airline data through comprehensive data collection, rigorous analysis, and the application of advanced business analytics techniques. Our analysis revealed several key insights.

This project underscores the critical role of business analytics in driving informed decision-making and fostering a data-driven culture within our organization. By continuously integrating analytics into our strategic framework, we can maintain a competitive edge and achieve sustainable growth.