**PROJECT REPORT**

**AeroUS**

**Airline Management System**

**Professor: Sreeram Chavali**

**Submitted By:**

**Aishwarya Chaubal**

**Bhupinder Jagwani**

**Nivedita Thapa**

**Shruti Garg**



Course: Business Intelligence and Data Warehousing

**TABLE OF CONTENTS**

| **Phase 1: Business Analysis** | 3 |
| --- | --- |
| 1.1 Business Scenario | 3 |
| 1.2 Key Information | 3 |
| 1.3 Tasks and Goals | 4 |
| 1.4 SwimLane Diagram | 4 |
| 1.5 Our Approach | 5 |
| 1.6 Technology Overview | 6 |
| 1.7 Business Analytics Questions | 7 |
| **Phase 2: Data Modeling** | 8 |
| 2.1 OLTP Schema Design | 8 |
| 2.2 Dimensional Data Model | 9 |
| **Phase 3: ETL Implementation** | 10 |
| 3.1 Extract | 10 |
| 3.2 Transform | 11 |
| 3.3 Load | 13 |
| **Phase 4: Tableau Implementation** | 14 |
| 4.1 Descriptive Analysis | 16 |
| 4.2 Prescriptive Analysis | 17 |
| 4.3 Key Dashboard Concepts | 17 |
| **Conclusion** | 18 |
| Key Learnings | 18 |
| Future Scope | 18 |

**Phase 1: Business Analysis**

**1.1 Business Scenario:**

John wanted to book a flight from San Francisco to New York to visit his parents for Christmas. He knew that AeroUS Airlines was having a sale for Christmas and decided to book the flight 4 months in advance. He logged on to the site, signed in with his email ID and password, and started searching for the flights. He entered the source and destination, the required travel date, the number of passengers, and the class he wanted to travel in. He found a few flights departing at night(which he preferred), was within his budget and seemed like a good option for him. However, he had a few doubts regarding baggage allowance and decided to use the AeroUS chatbot to resolve his doubts. He asked the agent about the check-in baggage allowance and the charges for an extra bag. After getting all his queries answered from the chatbot, John selected the preferred flight. He navigated to check-out, entered his personal information, billing and mailing address, credit card details, and finally purchased the ticket. His ticket details were emailed to him. Satisfied with the services provided and the ease of the entire process, he logged off.

On the day of his flight departure, he received an email update from AeroUS stating that the flight would be delayed by 3 hours due to bad weather. John went on AeroUS’s website to confirm the delay, and decided to leave for the airport at a later time than previously planned.

**1.2 KEY INFORMATION:**

**Stakeholders**: Customer, Chatbot/Customer Rep

**Department:** Advertising, Sales, customer service, IT

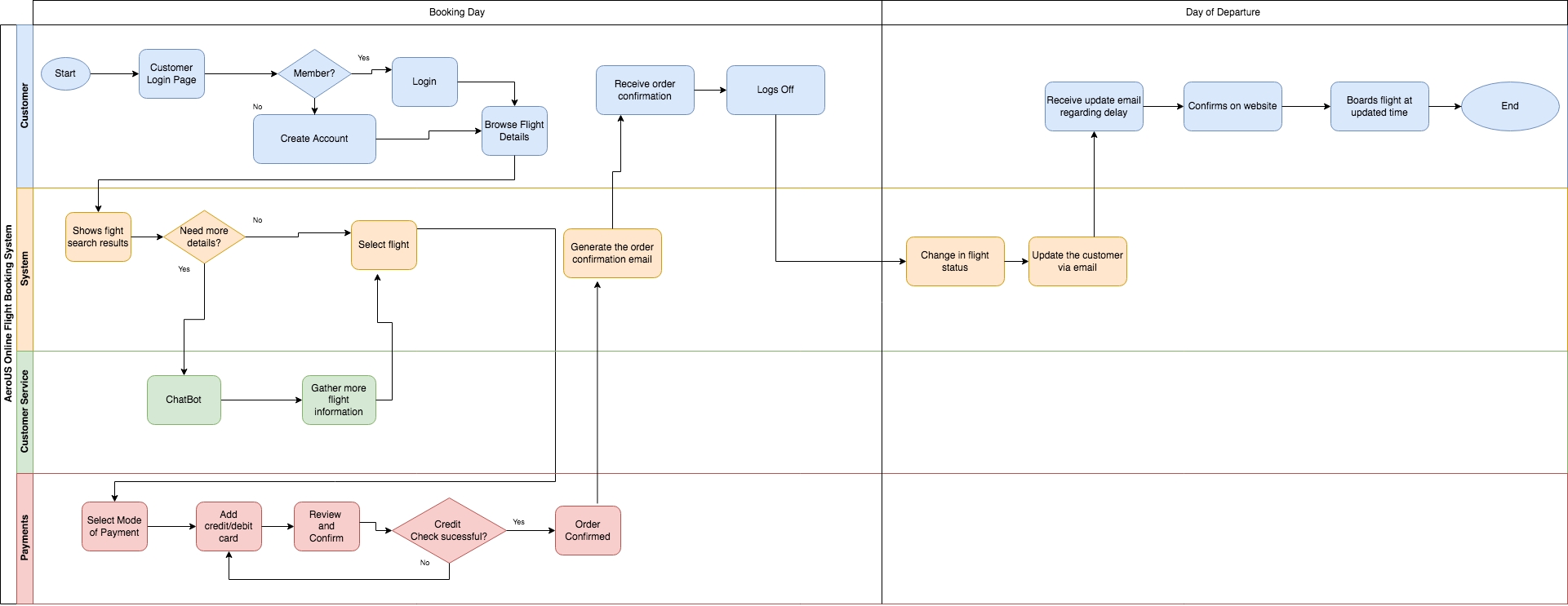
**Products**: Airline Tickets, Service

**Systems involved:** Google, Website, Email

**1.3 Tasks & Goals:**

|  | **Customer** | **System** |
| --- | --- | --- |
| **Task** | * Launch AeroUS website * Log In/Sign-up * Browse over flight details * Selects preferred flight * Need more details? * Chat with AeroUS chatbot * Go to checkout process * Select mode of payment * Add credit/debit card * Review/ confirm * Get confirmation email * Receives delay update email * Confirms on AeroUS website * Boards flight at updated time | * Help with login/sign-up process * Provide the flight and their details * Adds selected flight to cart * Redirect to chatbot in case of more details * Prompts options for modes of payment * Prompts review of the flight selected * Credit check * Provides flight confirmation * Sends email regarding flight details * Checks for flight delay * Sends email regarding flight delay |
| **Goal** | Purchase correct flight   * Get the right flight information * Pay right price | Enable smooth experience through the process to avoid abandonment |

**1.4 Swimlane Diagram:**



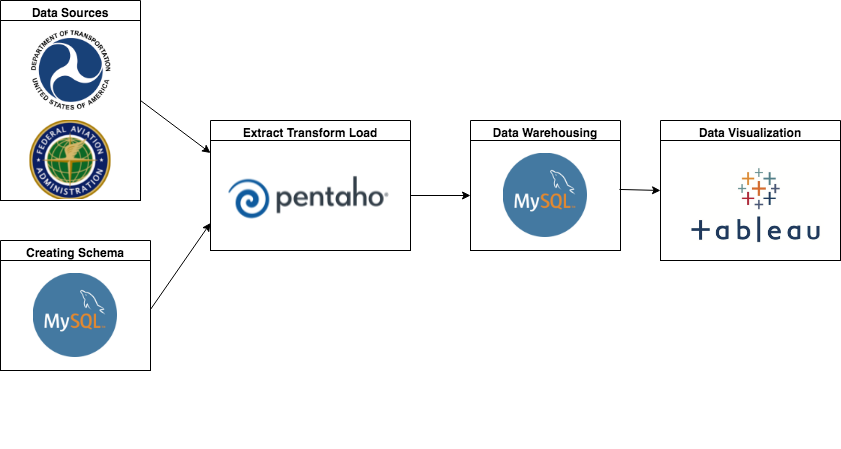
**1.5 Our Approach:**



For gathering key business insights for our airline and its performance, we followed the below steps:

1. **Data Collection & Profiling**: Before we collected any data, we focused on the business scenario and the different users of such data. The types of methodologies we used for data collection were going through different observations, documentations and records generated so far as well as creating questionnaires. Once all the information gathering and measuring on targeted variables was completed, we performed data profiling by reviewing source data, understanding the structure, content and interrelationships, and identifying potential for data projects. Data profiling which is a crucial part of Data warehouse and business intelligence (DW/BI) projects in the sense that data profiling can uncover data quality issues in data sources, and what needs to be corrected in ETL.
2. **Data Cleansing:** The next step was to perform data cleansing. We were using data from multiple sources. While combining multiple data sources, we observed that some data was duplicated or mislabeled. So we had to fix the data and remove the incorrect, corrupt, incorrectly formatted, duplicate, or incomplete data within a dataset.
3. **Data Modeling:** Once the data was cleansed, we performed data modeling where we identified entities, their properties, the relationship, mapped the attributes, decided upon the degree of normalization and finally finalized upon the data models. We created an OLTP and a Dimensional Model for our airline system.
4. **Data Integration:** The next step required us to perform data integration. We had data from multiple sources, and of various types like JSON, EXCEL, CSV and MySQL Table. Hence, there was a requirement to combine this data from multiple separate business systems into a single unified view, called data warehouse. We used Pentaho Data Integration tool for it
5. **Data Analytics:** While performing data analytics, we analyzed the data from the warehouse to find trends and answer key business questions. We focused on the lead, lag measures as well as descriptive and predictive analytics questions, which would provide key business insights to the different teams within our organization.
6. **Data Insights:** The final step was to visually represent the insights gained so far and we used Tableau Desktop for the data visualization. We graphically represented the information and data depending on the data analytics we performed in the earlier steps. We implemented various charts such as bar chart, stacked bar chart, trend lines, maps, etc. and also performed predictive analytics using forecasting mechanisms. This provided an accessible way to view and understand the trends, outliers, and patterns in data for our users, enabling them to take the key business decisions, benefitting the company.

**1.6 Technology Overview:**



**Data Sources:** The data was collected from multiple sites such as the Department of Transportation, USA and Federal Aviation Administration. Also, data sets for a few tables were manually created.

**Schema**: The entire database schema, OLTP data model and the scripts for AeroUS were created using MySQL Workbench.

**Extract Transform and Load:** We had data from multiple sources, and of various types like JSON, EXCEL, CSV and MySQL Table. We performed the entire data extraction process, extracting data from the source files and loaded it into the ODS layer. This was followed by performing various transformations such as joins, lookup, filtering, etc and finally loading this data into the data warehouse. All these steps were performed using the Pentaho Data Integration tool.

**Data Warehousing:** The data was combined from multiple sources and the ETL process was performed. The final data which will be used to draw meaningful business insights was loaded into the warehouse schema, creating a single unified view.

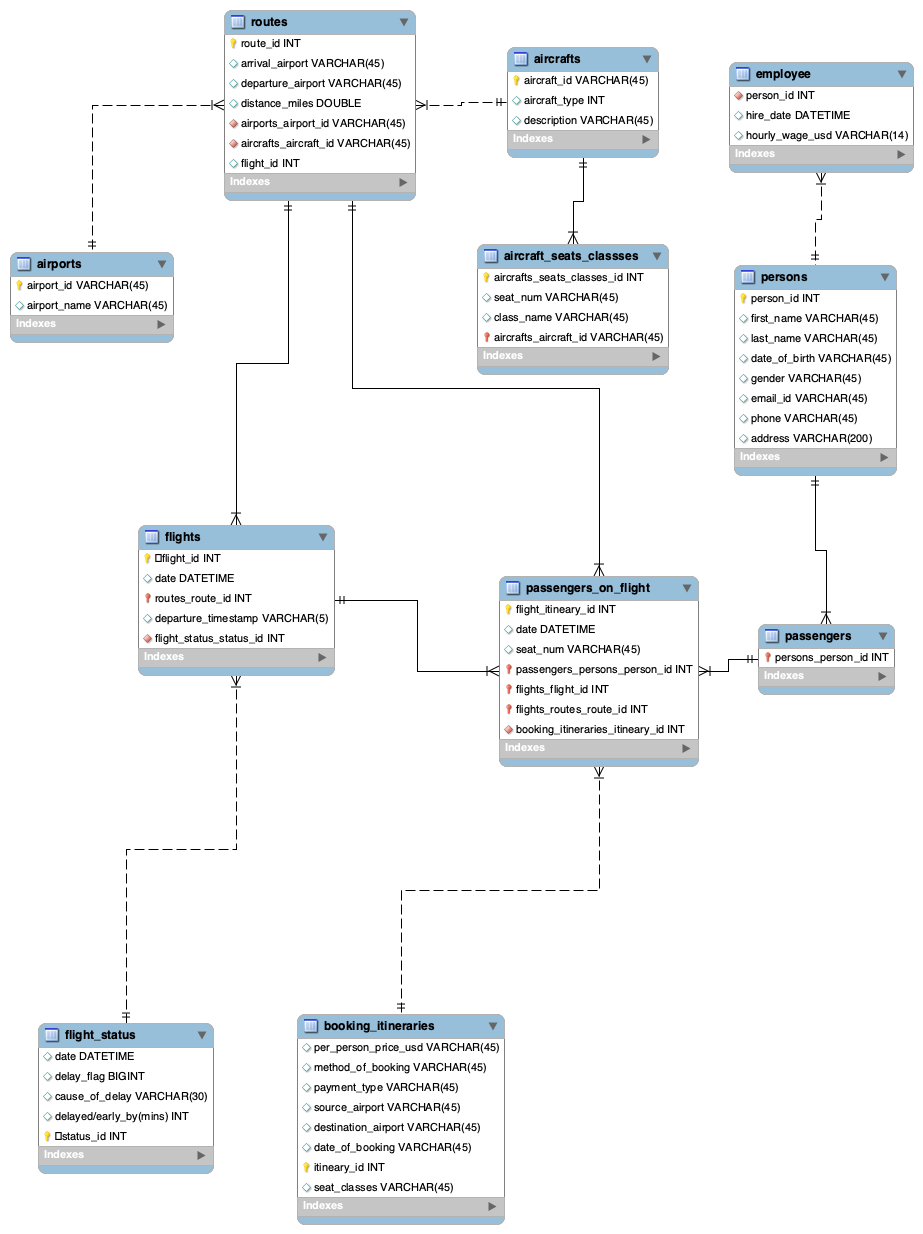
**Data Visualization:** The key business insights were visually represented using Tableau Desktop. This provided an accessible way to view and understand the trends, outliers, and patterns in data for our users, enabling them to take the key business decisions, benefitting the company.

**1.7 Business Analytics Questions:**

1. **Descriptive analytics:**
2. Revenue increase/decrease over the period of 1 year
3. Increase in the number of flights in the last 1 year
4. Increase in the number of new routes
5. Top 5 fully-booked flight routes in the last year
6. Revenue generated by each airline service category (cargo/passenger) in the past 1 year.
7. **Predictive analytics:**
8. Cost prediction based on average aircraft delay projections
9. Offers or upgrades based on customer history
10. Increase in the flight frequency based on analysing overbooked flights for certain routes.
11. Forecast sales based on past promotions (Eg: sales during Christmas,Thanksgiving promotion)
12. Reduction in expenses by analysing past marketing costs ( Eg: targeted marketing based on demographics/ economic status)
13. Analyze flight data based on aircraft type & weight, weather and the route, etc. to predict the optimal amount of fuel needed for the flight
14. **Lead Measures:**
15. How many left on time in a day?
16. What are the causes of delay per day?
17. What are the airports where delays were caused?
18. Maximum and Minimum Total Sales per Week
19. **Lag Measures:**
20. Increase in the number of flights in the last 1 year?
21. How many left on time in a month?
22. Cause of delay per year?
23. Top 10 Airport names where delays were in a month?
24. Number of bookings via third-party website in the past month
25. Total yearly sales
26. Total sales per quarter

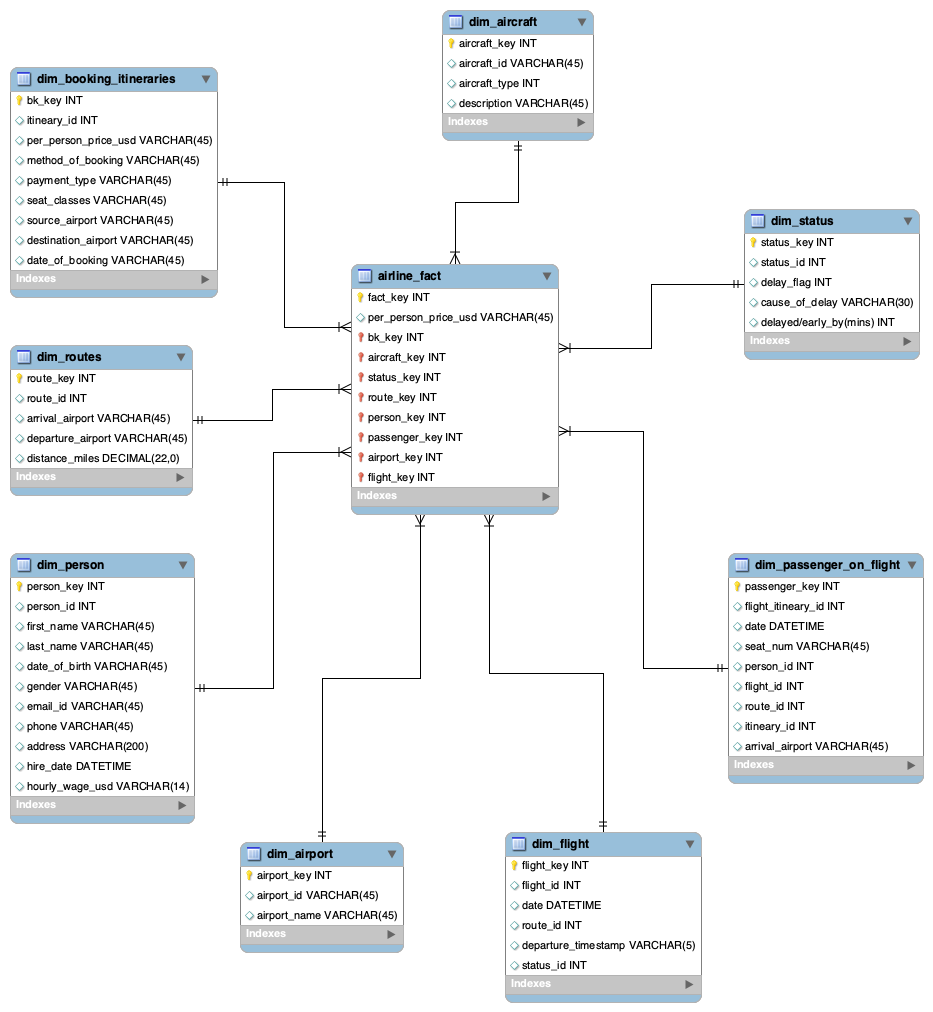
**Phase 2: Data Modeling**

**2.1 OLTP Schema Design:**

****

The OLTP data model was created after gathering all our data from various sources.The model shows that passenger\_on\_flight is one of the tables that is central to most of the tables (contains foreign keys) such as routes, flights and booking\_itineraries.

**2.2 Dimensional Data Model:**

****

The Dimensional Data model on the other hand adopts a star schema and so there is one fact table that has the transactional value (per\_person\_price\_usd) and all other information is stored in the dimension tables. This model was created to streamline the data and make it easy to read and perform analysis using BI tools such as Tableau.

**Phase 3 : ETL Implementation**

**Data Sources used for Implementation**

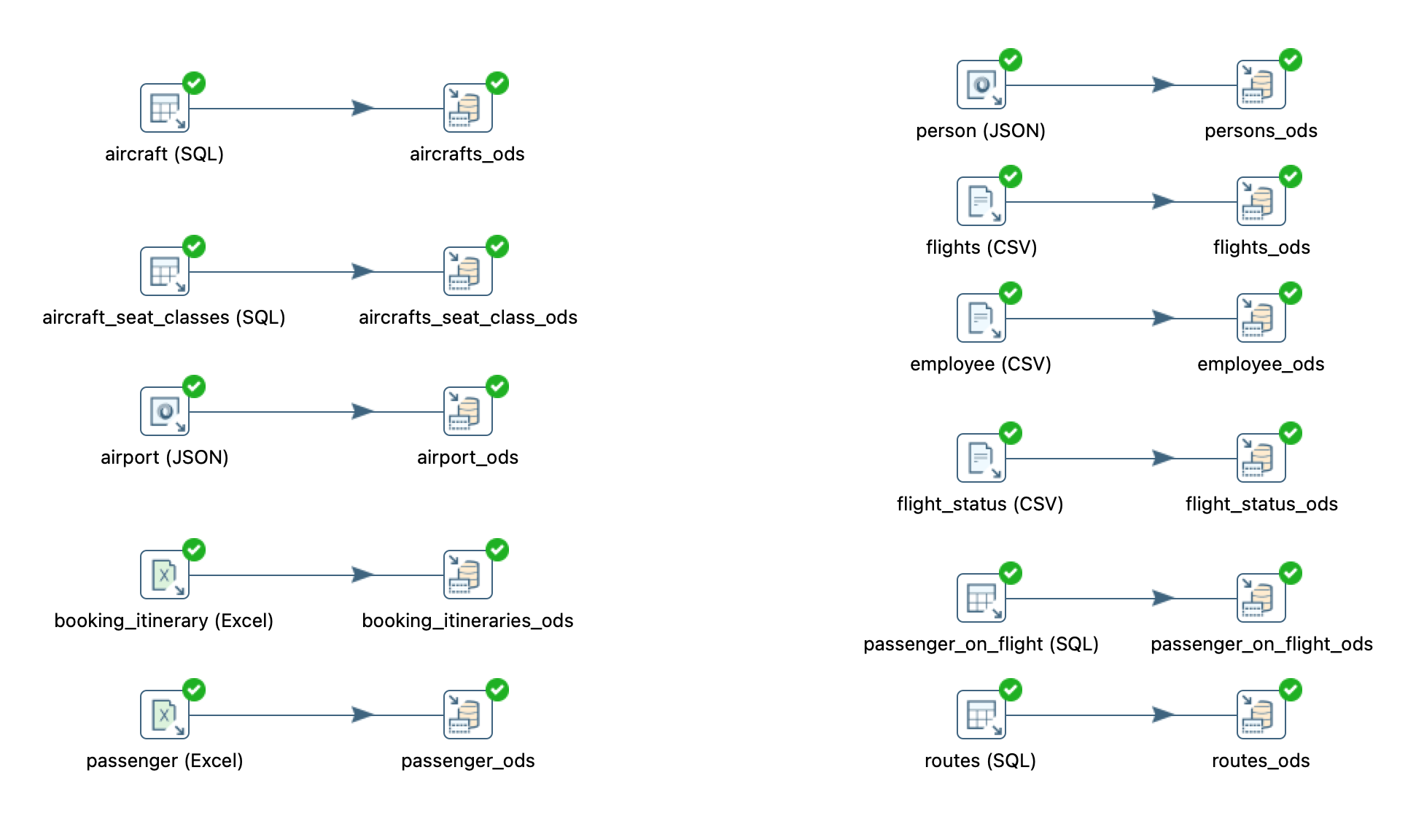
1. MySQL Database Tables
2. JSON files
3. CSV files
4. Excel files

**Data Sets:** Few of the table's data are manually generated.

**Sources we will be using for this project:**

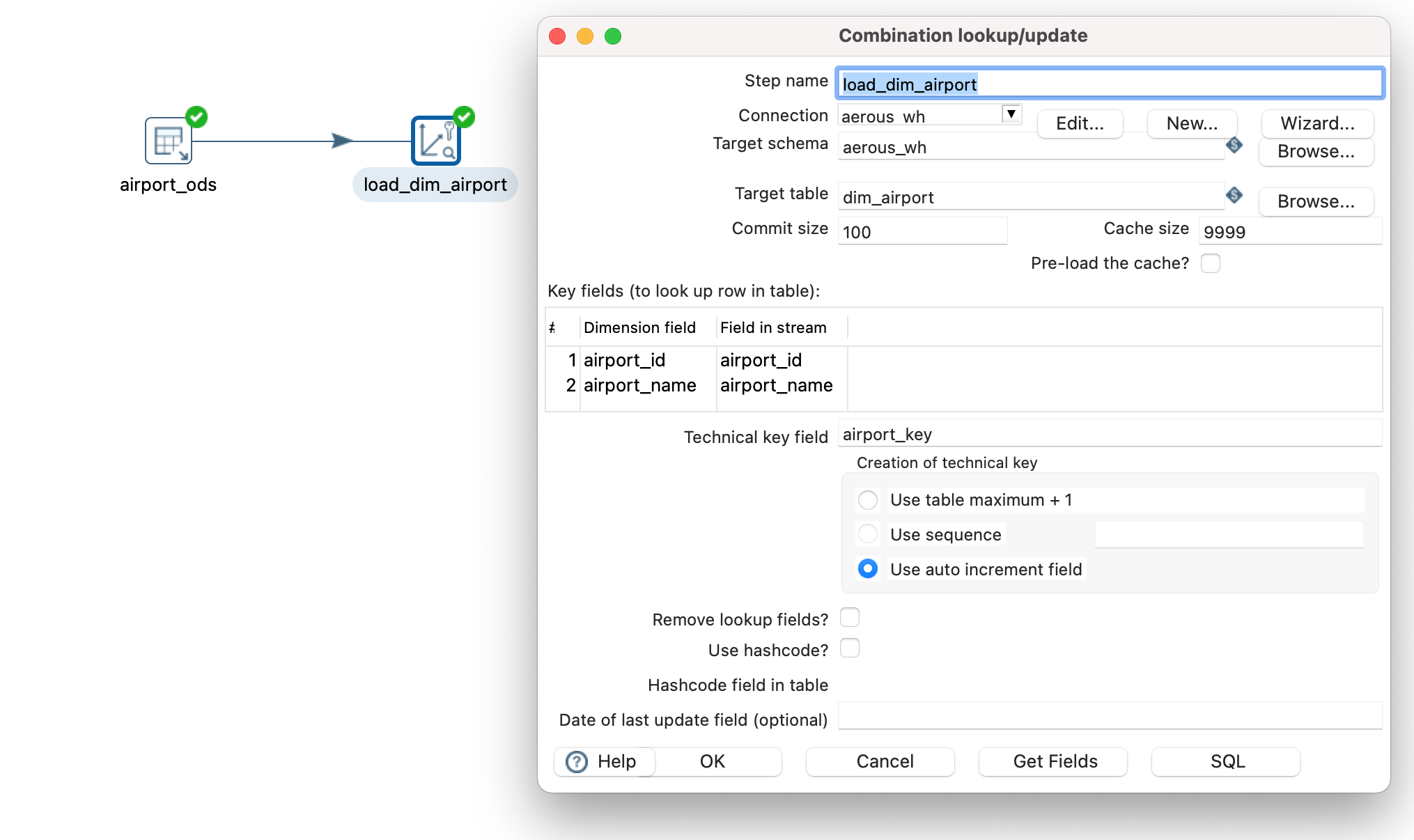
1. Airline schedule: [Airline Schedule](https://transtats.bts.gov/ONTIME/Departures.aspx)
2. Delay causes: [Delay Causes](https://www.bts.dot.gov/explore-topics-and-geography/topics/airline-time-performance-and-causes-flight-delays)

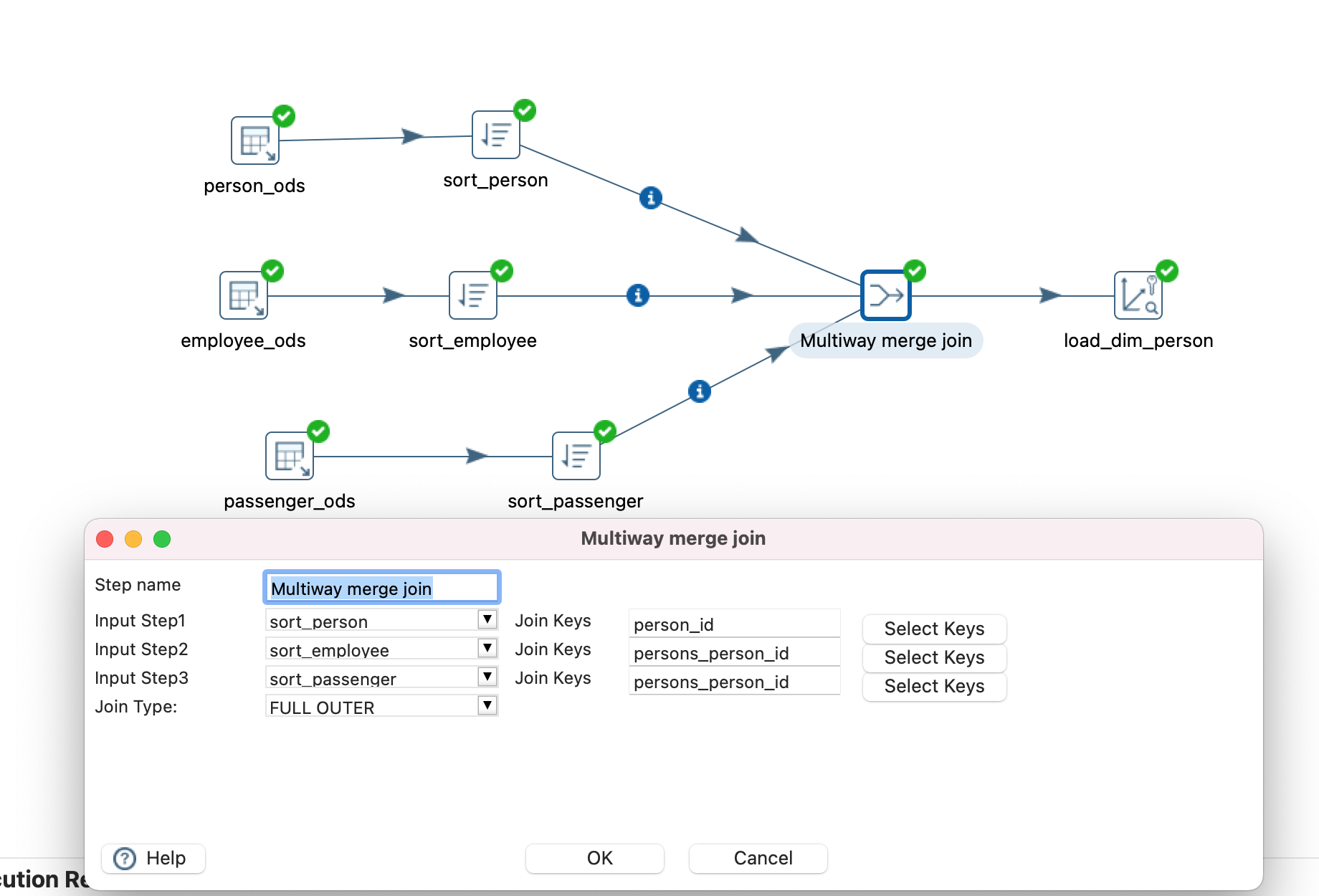
**3.1 Extract**

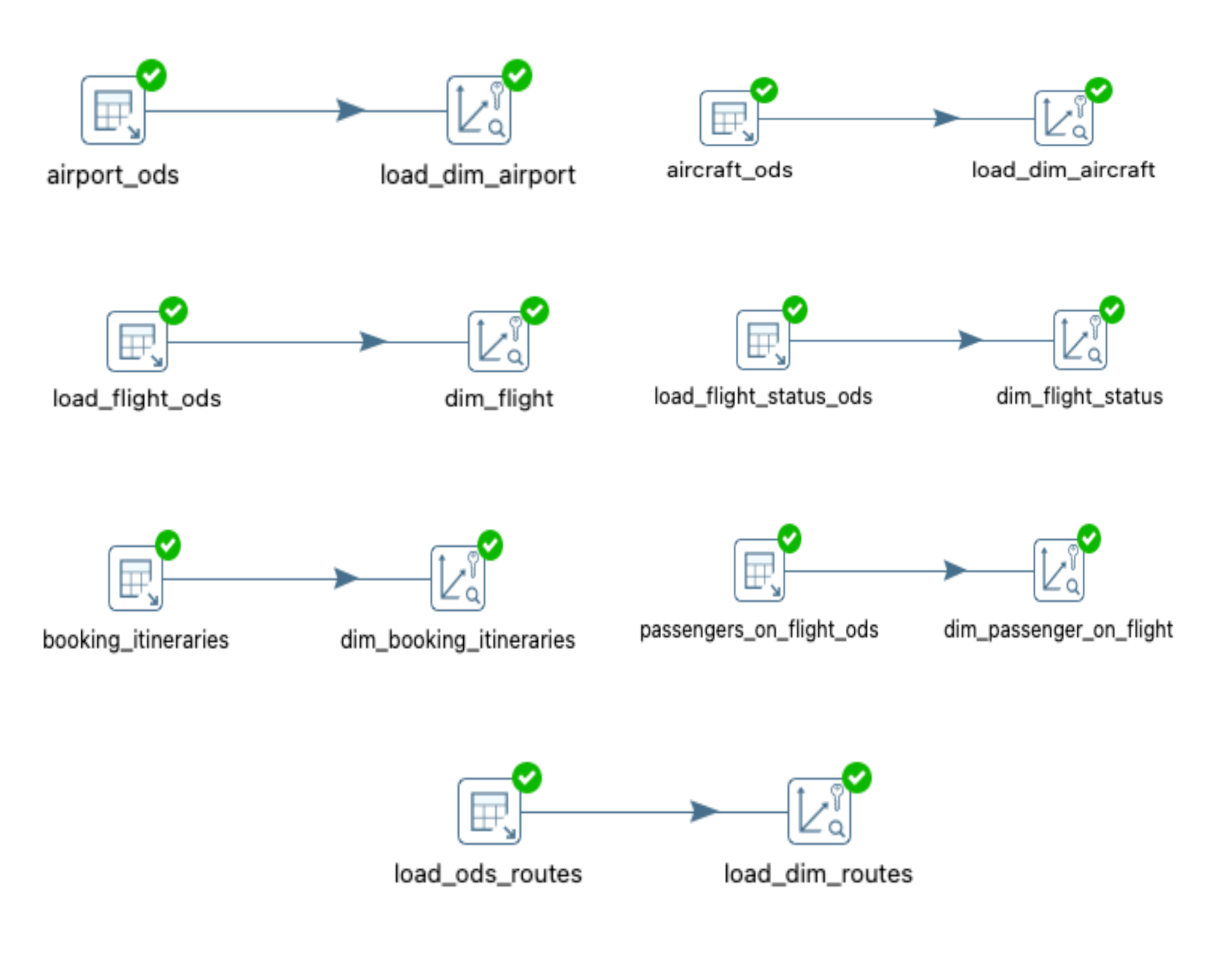
****

The Extract phase in the ETL Implementation involved extracting data from different sources (SQL database tables, Excel files, JSON, CSV,etc.) and storing it in a staging area called ods ( aerous\_ods schema in mySQL). In this stage we basically did an **insert/update** transformation using Pentaho.

**3.2 Transform**

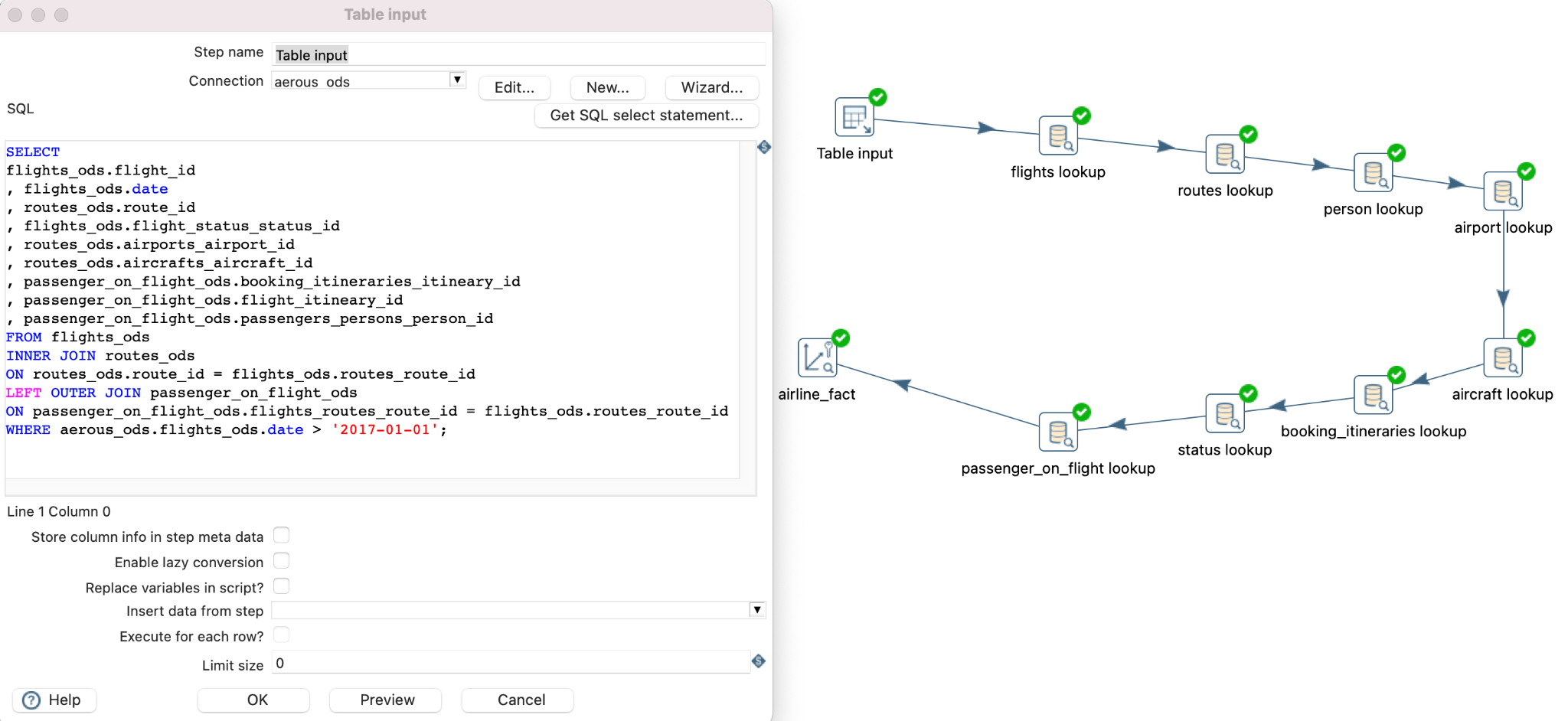
****

****



The Transform stage of the ETL Implementation involved loading all dimensions to the data warehouse. This involved using **combination lookup/update** transformation in Pentaho. We also did a multiway merge for getting all the information from person, employee and passenger tables to one table.

**3.3 Load**

****

The last step in the ETL process is Loading. This is basically combining all the dimensions that we loaded and looking those up (getting the key for each dimension) and finally loading the fact table (that connects all dimension tables). In this step, we wrote a simple query to have our table input based on which we performed database lookup for each dimension. Finally, the last step was combination lookup/update where in we loaded all the dimension table keys as well as our transaction data (per\_person\_price\_usd)

**Phase 4 : Tableau Implementation**

Based on the Business Scenario, we generated a series of questions for lead, lag, descriptive & predictive analytics. Further, we identified our key user personas who require the dashboards and the different sheets they would need. Hence, below are the 4 main user personas we identified who will need the 2 dashboards (descriptive & predictive) which we have created:

User Persona:

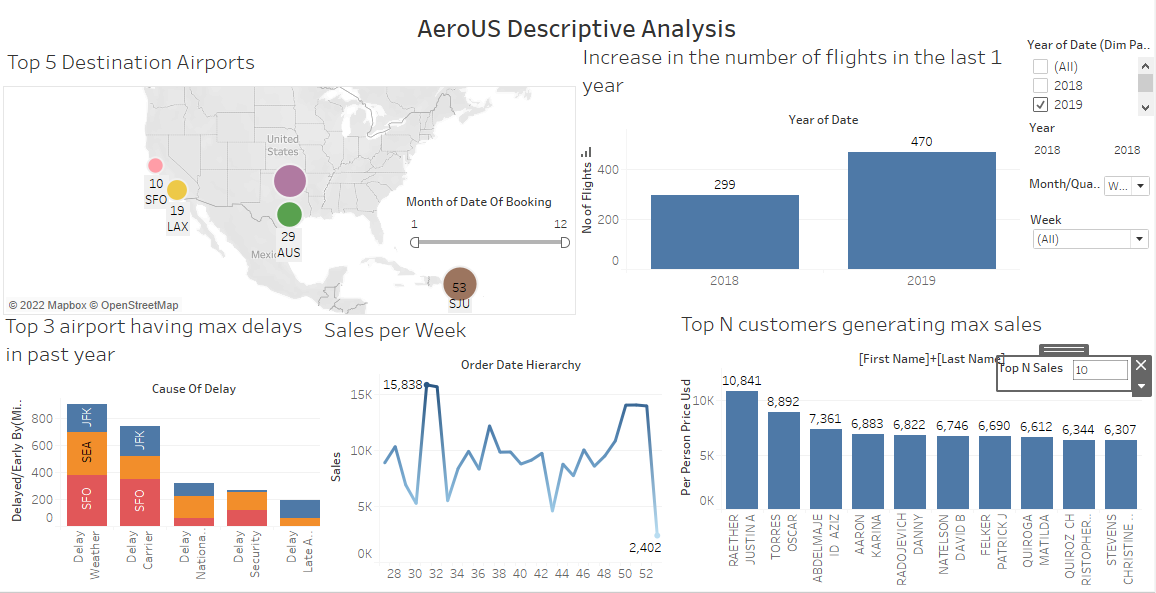
1. AeroUS CFO
2. Marketers
3. Revenue Managers
4. Regional Managers

| **As a** | **I want to** | **So that** |
| --- | --- | --- |
| Chief Financial Officer | Have a dashboard which includes what-if analysis for monthly sales | I can estimate the budget forecast for the organization |
| Marketer | 1. Have a dashboard which includes Model Percentile for Top 5 Destination Airports and Method of Booking 2. Have a dashboard which includes Increase in the number of flights in the last 1 year 3. Have a dashboard which includes Top N customers generating max sales | I can identify and cater for the customers booking preferences according to their geographical location and also improve marketing strategies  I can identify and cater for those customers and provide them with reward points, flight-concessions, etc |
| Financial Analyst | 1. Have a dashboard which includes Model Quantile for Sum of Sales 2. Have a dashboard which includes Sales per week/quarter/year | 1. I can generate a confidence interval and analyze the posterior predictive quantile, or the expected value at a specified quantile for Sales 2. I can analyze the data and suggest insights & mitigation for the next week/quarter/year |
| Regional Manager | 1. Have a dashboard which includes Delay Forecast per Cause of Delay 2. Have a dashboard which includes Top 3 airport having max delays in past year 3. Have a dashboard which includes Top 5 customers location | 1. I can analyze and take measures to mitigate future delays at my regional airport 2. I can analyze the busy airports to increase/decrease the flight frequencies in my region |

The Tableau file (.twbx) has descriptive and predictive analysis.

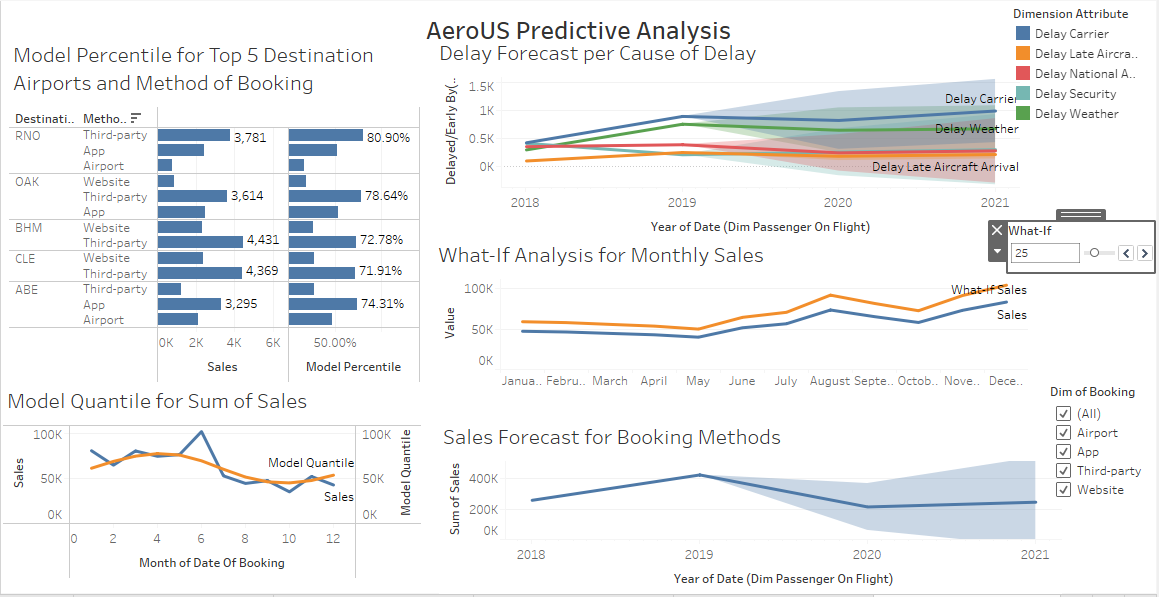
**4.1 Dashboard 1: AeroUS Descriptive Analysis:**

* Top 5 Destination Airports
* Increase in the number of flights in the last 1 year
* Top 3 airport having max delays in past year
* Sales per Quarter/Month/Week
* Top N customers generating maximum sales



**4.2 Dashboard 2: AeroUS Predictive Analysis:**

* Model Percentile for Top 5 Destination airports and Method of Booking
* Delay forecast per Cause of delay
* What-If-Analysis for monthly sales
* Model Quantile for sum of sales
* Sales forecast for booking Methods

****

**4.3 Key Dashboard Concepts:**

The key dashboard concepts that we implemented were as follows:

1. Created and Implemented various parameters such as what-if analysis, Top N Sales, Top N Customers,etc. This enabled us to create an interactive and user friendly dashboard where the user has the ability to get maximum insights by selecting the dimensions of their preference within the same dashboard. It gives the user the ability to dynamically modify values rather than manually setting the number. Adding this feature into our dashboard made it dynamic & highly insightful.
2. For data visualization, we have used multiple charts such as bar graphs, line charts, maps, etc. Each of these charts have different interpretations and annotations. A bar graph helps us to compare data across multiple categories. We used line charts in both descriptive and predictive analysis. This chart helps connect individual data points in a single view. They provide a simple way to visualize a sequence of values and are useful when we want to see trends over time, or to forecast future values. Map charts helped us view data with respect to the different geographic locations

**Conclusion**

**Key Learnings:**

* Importance of good (clean, consistent) data: This does not just mean clean and consistent but it also means to collect data that helps answer business questions better. This can be applied to the real world also, that organizations should not just collect all data, but data that is useful.
* Ease that Data Transformation tools like Pentaho and Tableau have brought: These tools have brought a lot of ease to data transformation activities, saving time of data engineers and the likes.
* Data Visualization helps in comprehending data at a glance and also brings focus to the quality of data

**Future Scope:**

More data would help us answer the following business questions:

* Number of booking cancellations 24 hrs before departure (Lead Measure)
* Customer/Employee Satisfaction (Lag Measure)
* Offers or upgrades based on customer history (Predictive)
* Top 5 fully-booked flight routes in the last year (Descriptive)