1. **INTRODUCTION**

**System Overview**

In our modern era data is very important, A corporation depends on data to make accurate decisions and procure a competitive edge. The Airline industry, with its large amounts of data generated from its customer's flight bookings Thanks to the ability of Data Warehousing and Business Intelligence this data can be put to good use. The aim of this project is to **Design and Implement a Data Warehouse/Data Mart (proof of concept) and develop BI reports based on the data of a flight booking company**, gathering data from a source system.

**1.2 REASON FOR SELECTING THE SUBJECT AREA AND DATA**

The flight booking industry is an ever-changing industry with large-scale data on flight bookings, flight cancellations, customer preferences, and revenue. This subject area offers promising opportunities for gathering business insights.

**Source System**

The source system that I chose is a flight booking database. This database records all data that is related to flights booked, this includes information on the passengers, sales channels, trip types, purchase lead times, length of stay, flight schedules, routes, and customer preferences such as baggage requirements, seat preferences, and in-flight meal requests. This dataset mimics real-world flight booking systems realistically, the columns or key attributes of the data are as such i listed them the way they are on the dataset:

**num\_passengers**: This is the number of passengers that are booked in a single transaction.

**sales\_channel**: The medium through which the booking service was done, it could be via the internet

**trip\_type**: The type of trip it could be one-way, round-trip etc.

**purchase\_lead**: The period between booking and flight departure in days.

**length\_of\_stay**: How long the passenger stays at the destination.

**flight\_hour**: The hour of the flight departure.

**flight\_day**: The particular day of the week which the flight is scheduled.

**route**: The route of the flight.

**booking\_origin**: The location where the booking was made.

**wants\_extra\_baggage**: states if the passenger requested extra baggage.

**wants\_preferred\_seat**: Indicates if the passenger requested a preferred seat.

**wants\_in\_flight\_meals**: Shows if the passenger requested meals on flight.

**flight\_duration**: The duration of the flight.

**booking\_complete**: Indicates if the booking process was completed successfully.

**Reference for Source System**

The source system i made use of is inspired by the operation of real-world flight booking systems and datasets available for research purposes.

Source: British Airways

**1.3 VISION AND GOALS**

The vision for building this data warehouse is to integrate flight booking data from my chosen source system into a single data mart. With this, I'll be able to formulate BI reports and dashboards that offer necessary information concerning booking trends, the behavior of customers, efficiency, and the pattern of revenue. By utilizing this information, stakeholders can make decisions driven by data to make operations more productive, boost the satisfaction of customers, and improve business growth.

**1.4 KEY STAKEHOLDERS**

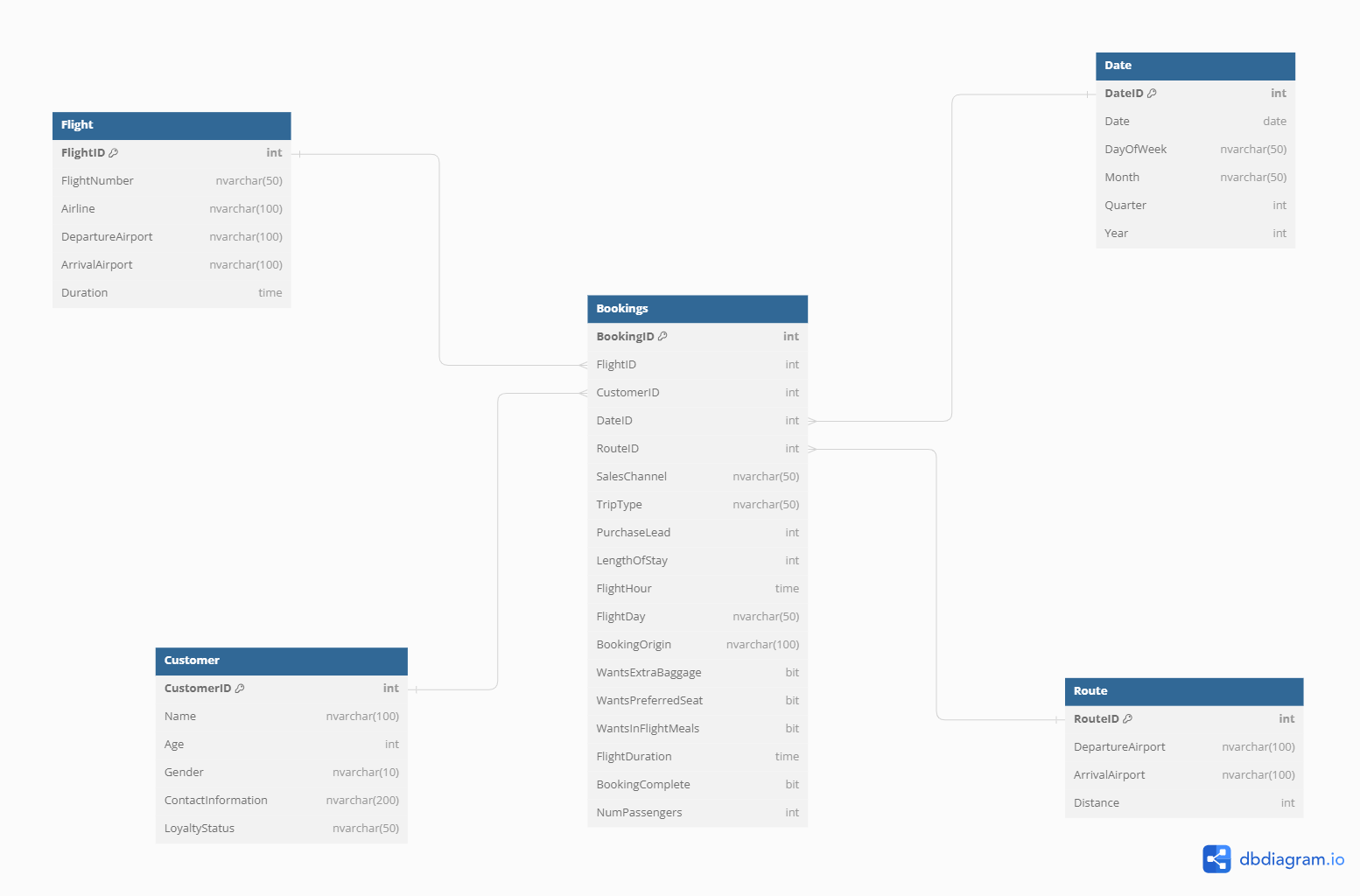
The key stakeholders that I identified for this project are:

* **The Finance Team**: They are concerned with revenue, costs, and profitability.
* **The Executives**: Interested in the overall flow of the business, revenue, and business strategies.
* **Marketing Team**: They deal with customer demographics and trends.
* **Operations Team**: They require data such as flight occupancy, cancellations, and punctuality.

**1.5 INSIGHTS FROM DATA**

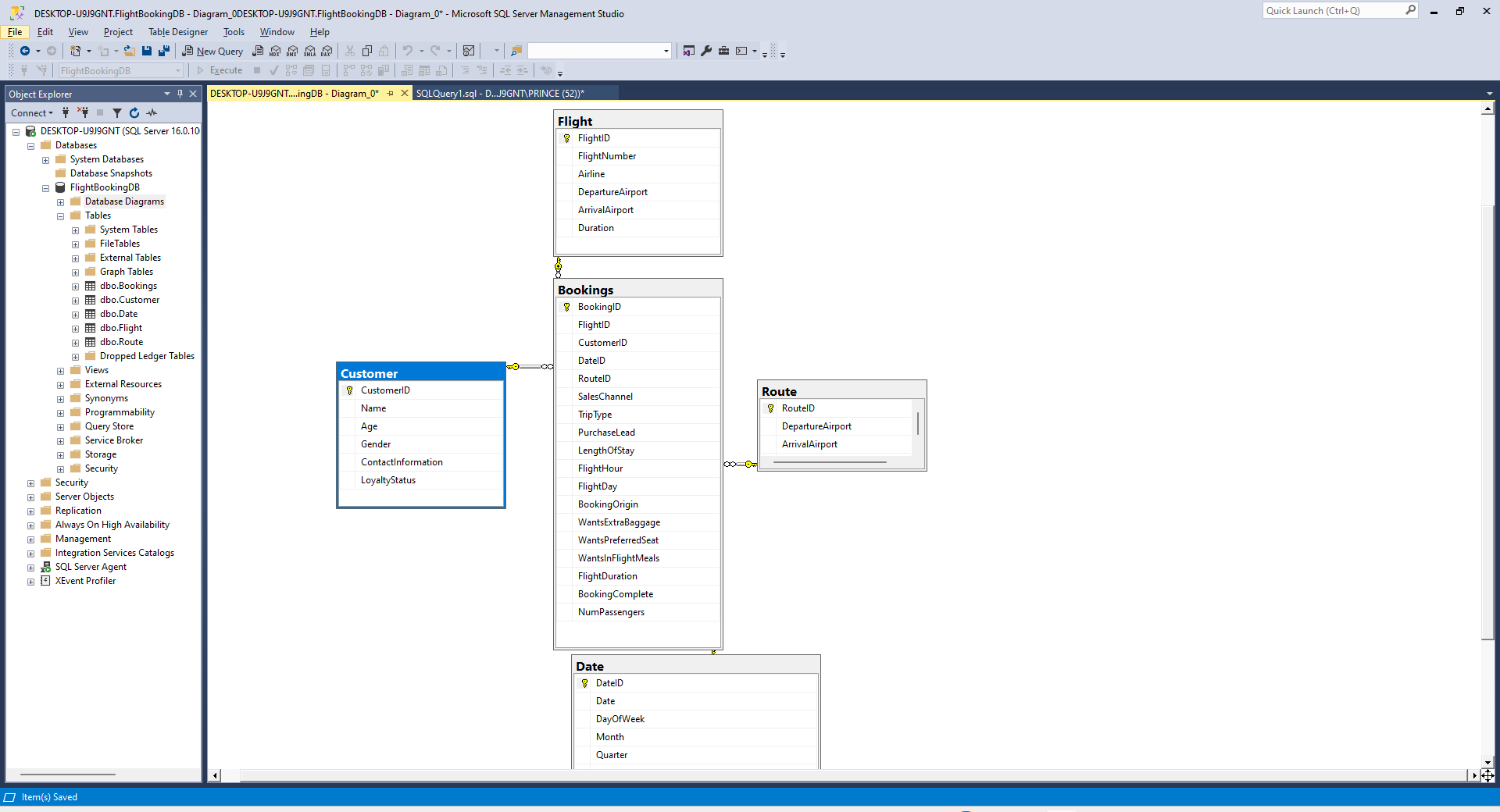
* **Analysis of sales and revenue**: Grasping the knowledge of booking trends, peak periods, and generation of revenue.
* **The behavioral patterns of customers**: Studying customer demographics, selections, and feedback.
* **Operational Efficiency**: Tracking the rate of flight occupancy, cancellation, and on-time performance.
* **Market Trends**: Identifying popular routes

1. **SCHEMA / DIMENSIONAL DATA**



My reason for using this design is the fact that **Querying data is simple and straightforward.** The star schema is designed to enhance the simplicity of querying data by organizing data into fact and dimension tables. This structure allows for easier analytical queries by users without dealing with complex joins, as the star schema involves fewer tables and more accurate relationships.

A good example in the flight booking business is a query to find the total number of passengers and revenue per route and month, this would be easy due to the straightforward relationships between the fact table [Bookings] and the dimensions table [Route, Date]

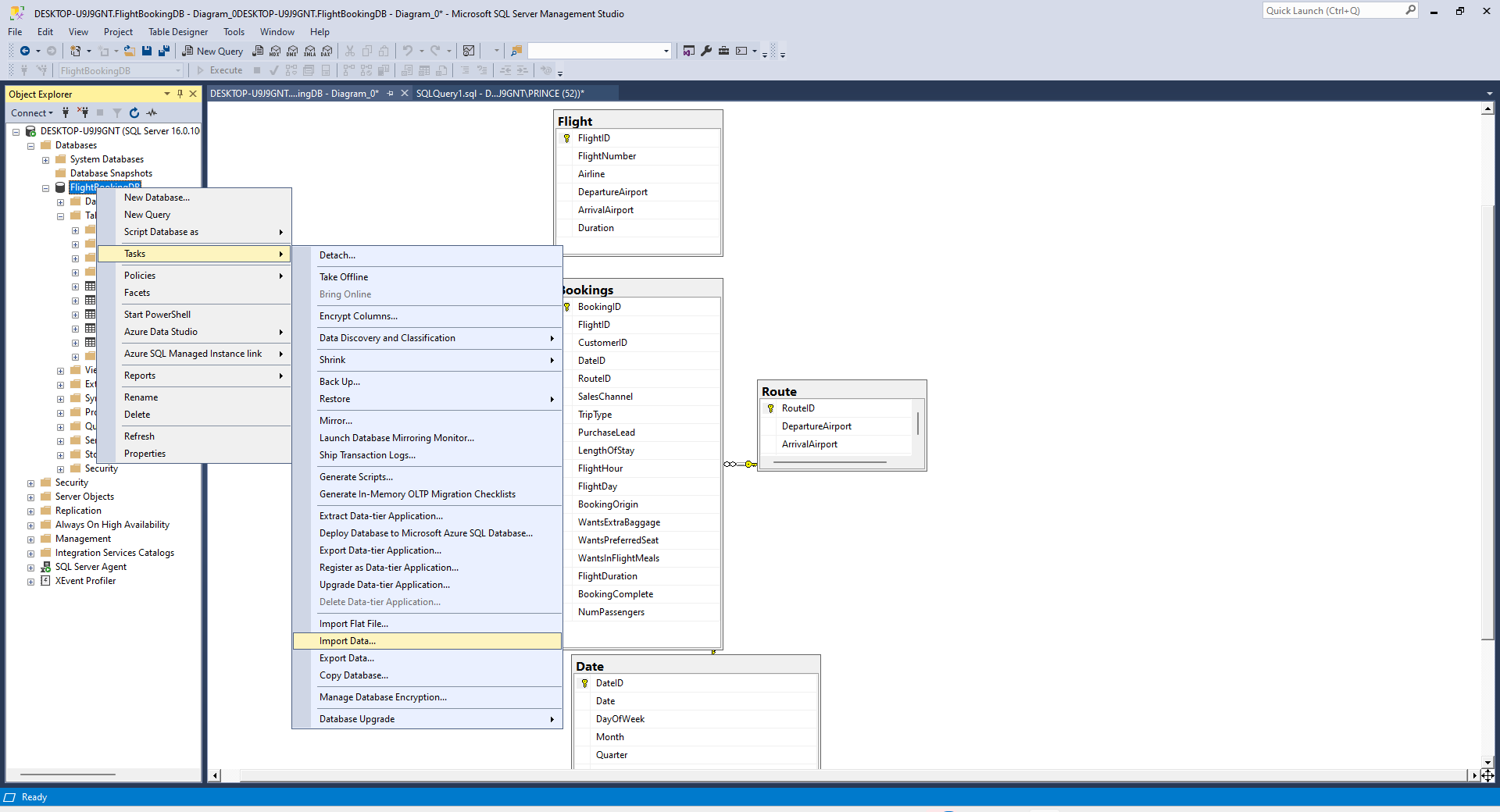
1. **IMPLEMENTATION OF DATA WAREHOUSHE/DATA MART**
2. 
3. **ETL TO POPULATE THE DATA WAREHOUSE / DATA MART**
4. **Source Data**

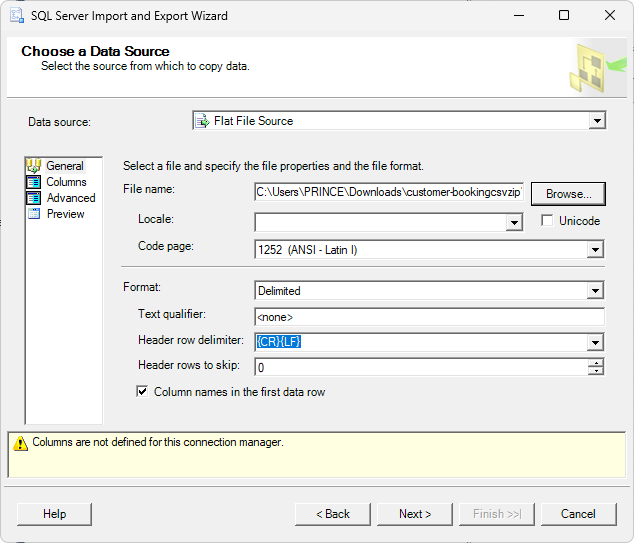
In my case the source data was in a csv file, the stage involves me pulling my data

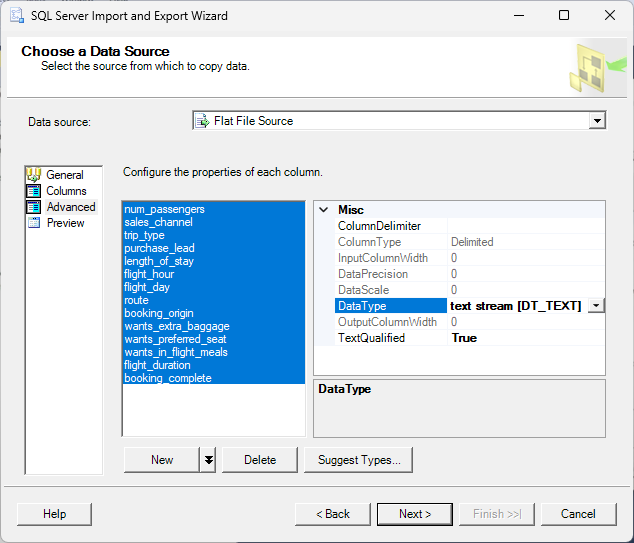
1. **Transforming the data**

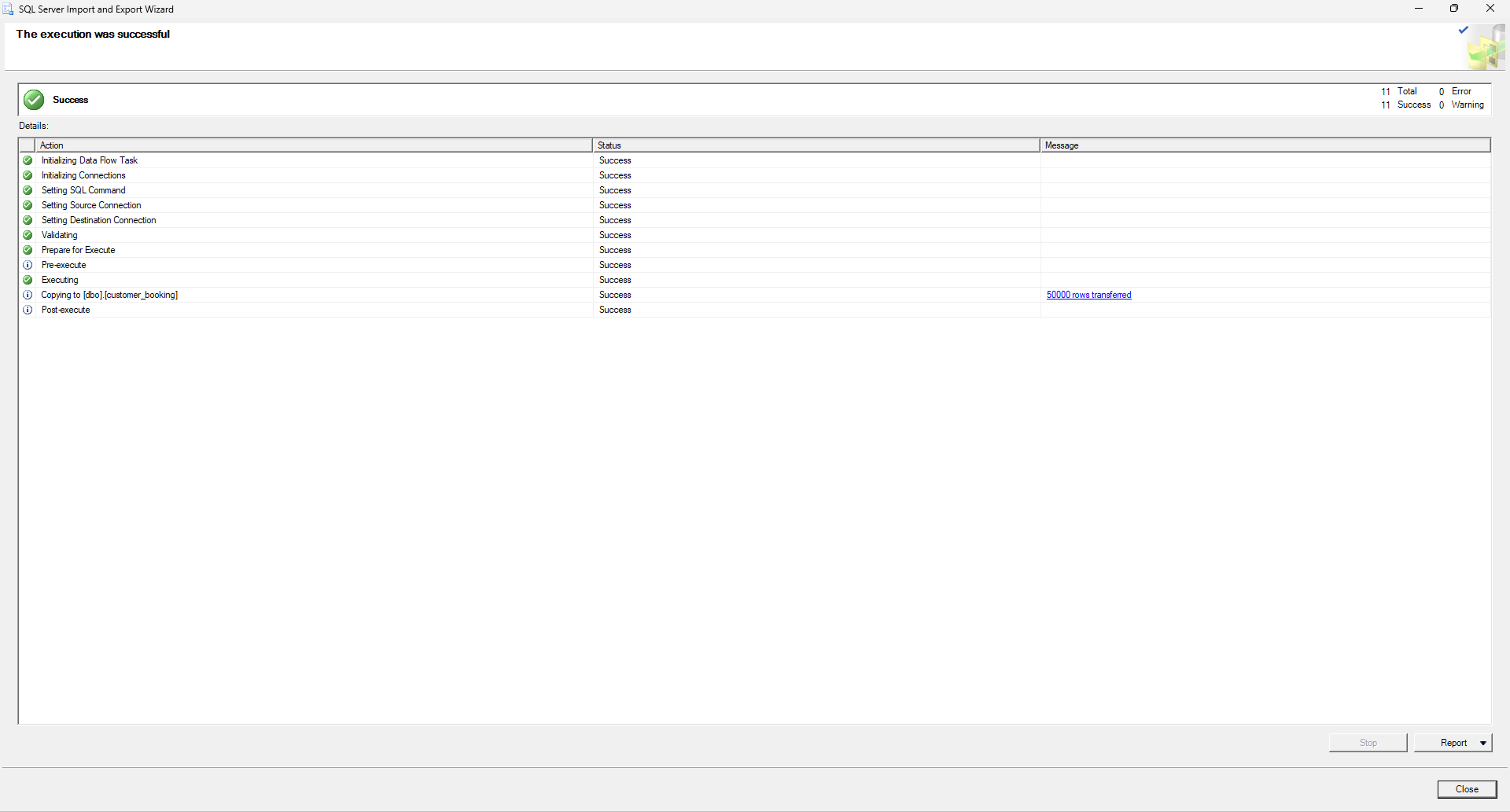
This stage involves cleaning up the data and formatting it

**3. Loading the data**

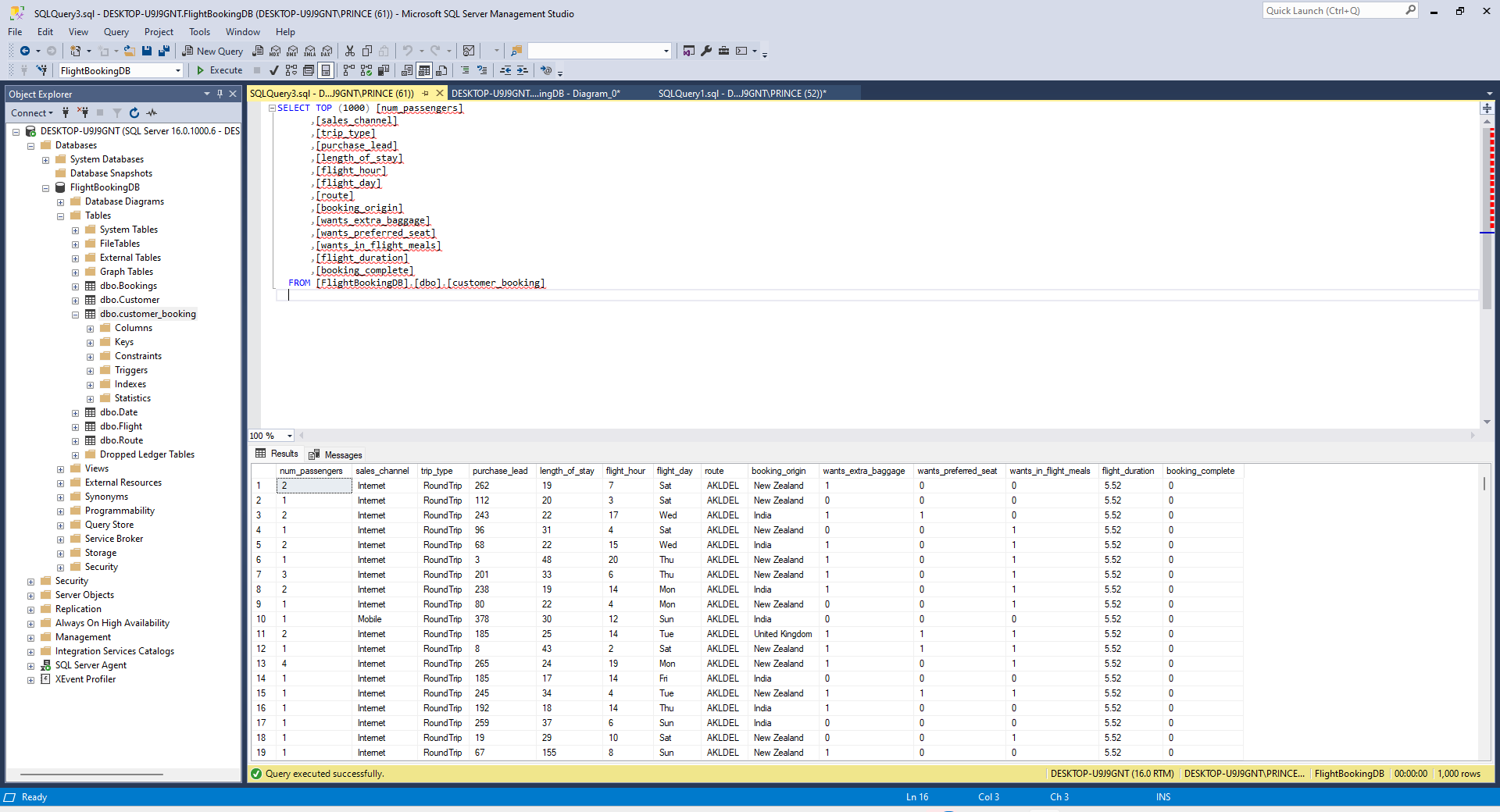
Here I’m loading the data into my SSMS database.

Im using the select flat file source option cause my data file is a csv file

I changed the data type from string to text stream



I execution was successful with 50,000 rows transferred

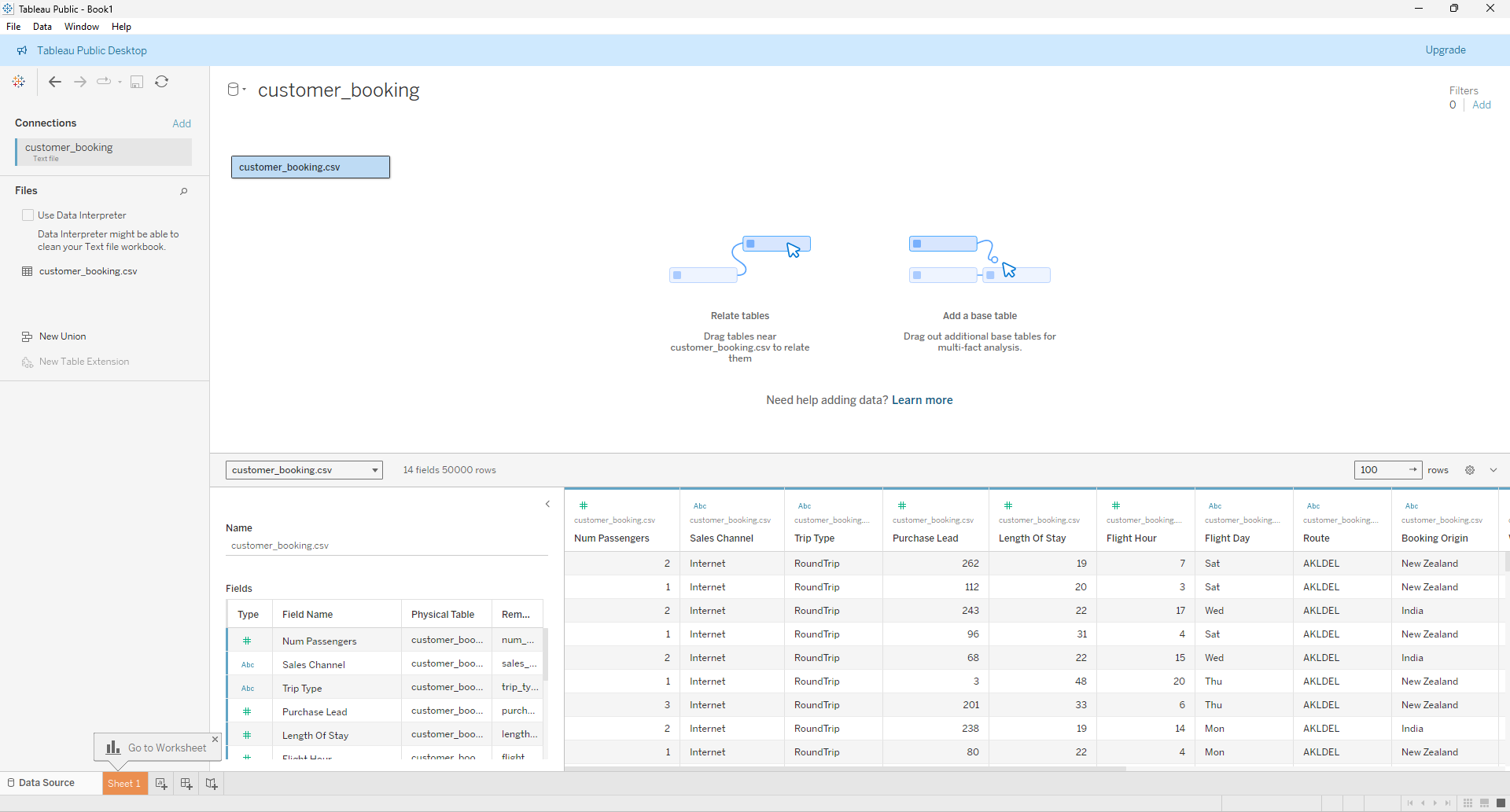


**4. VISUALIZATION AND REPORTS**

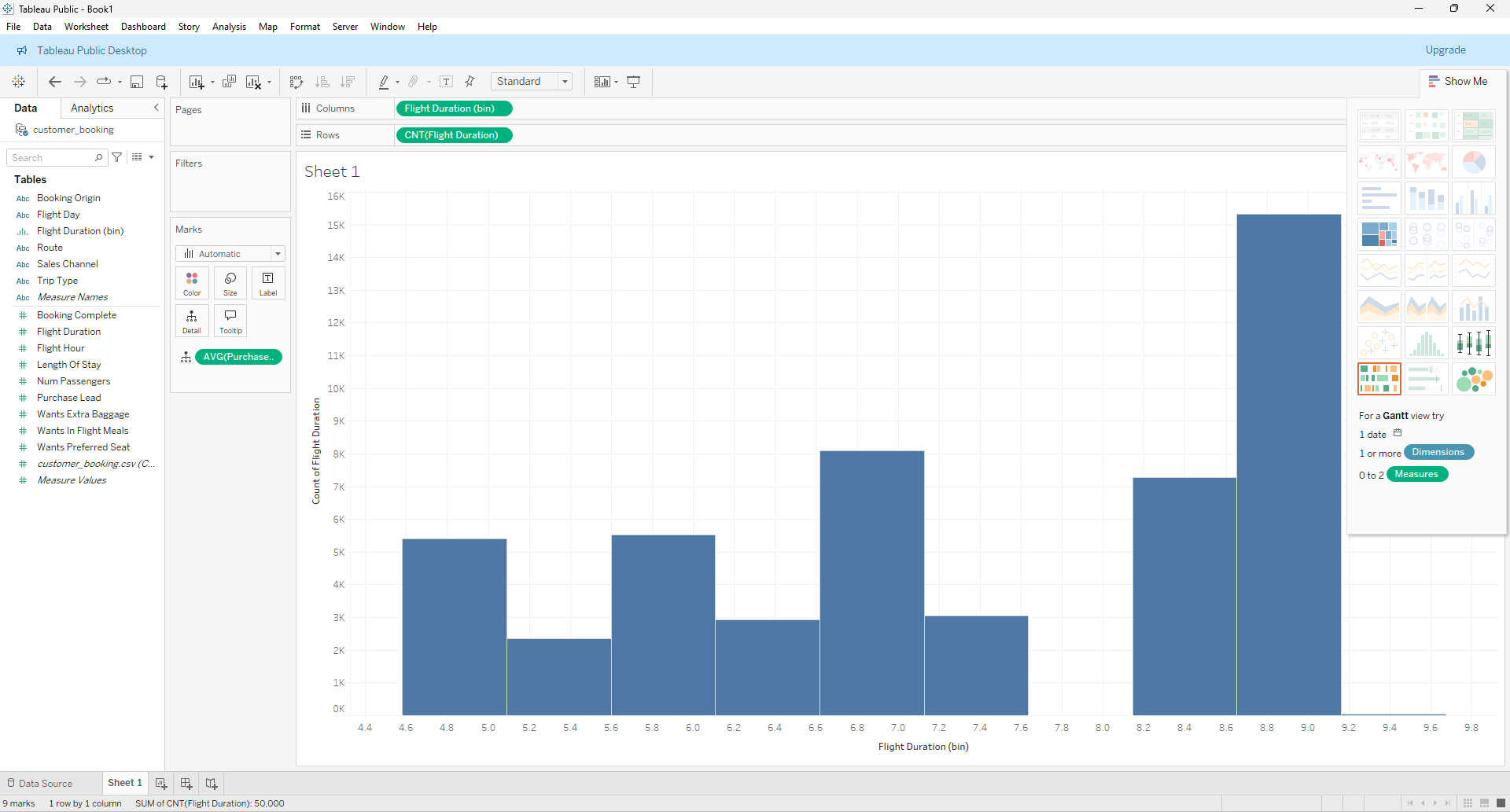
**4.1 TABLEAU VISUALIZATION**

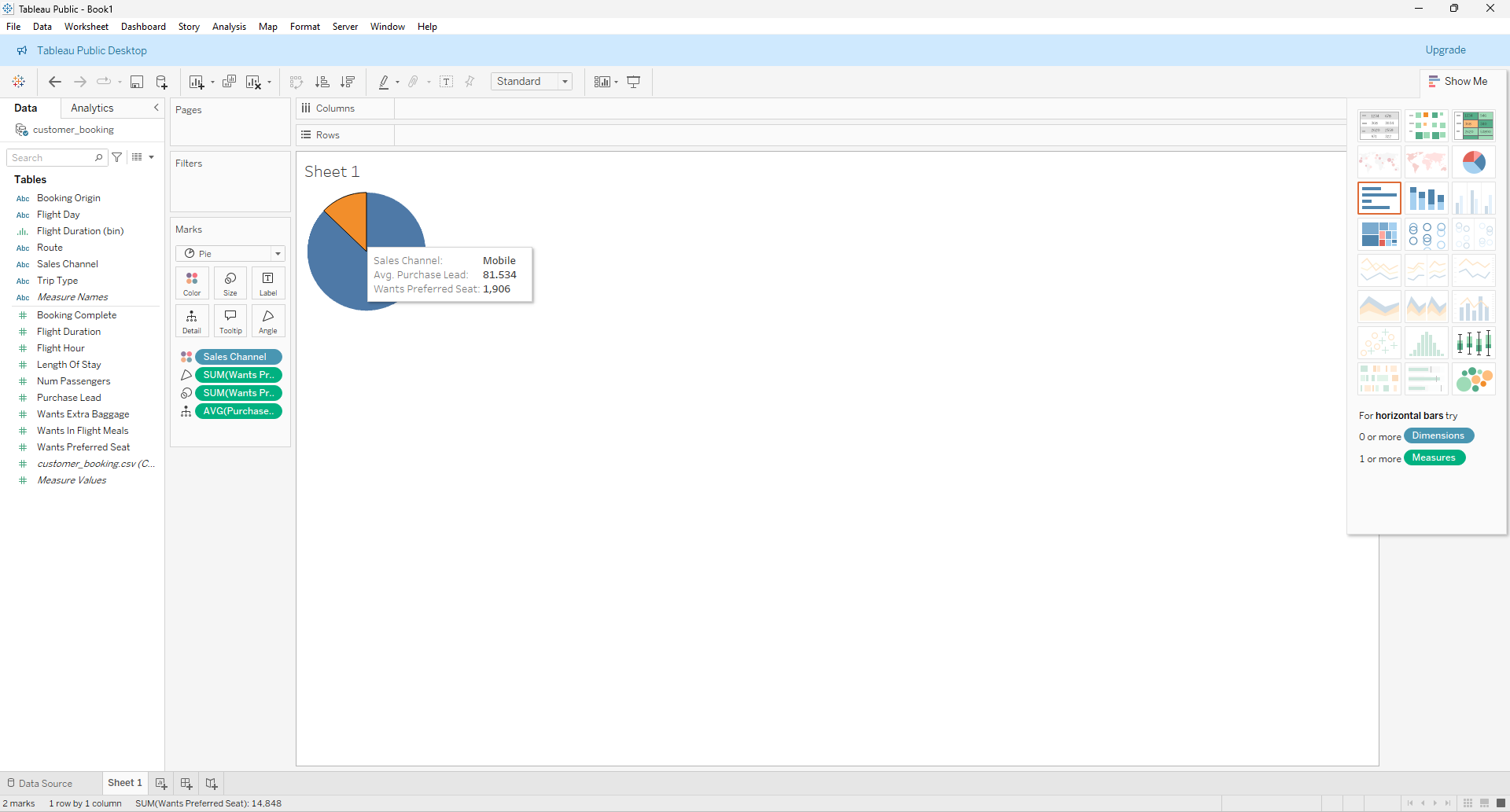
The business requirements for the data mart are

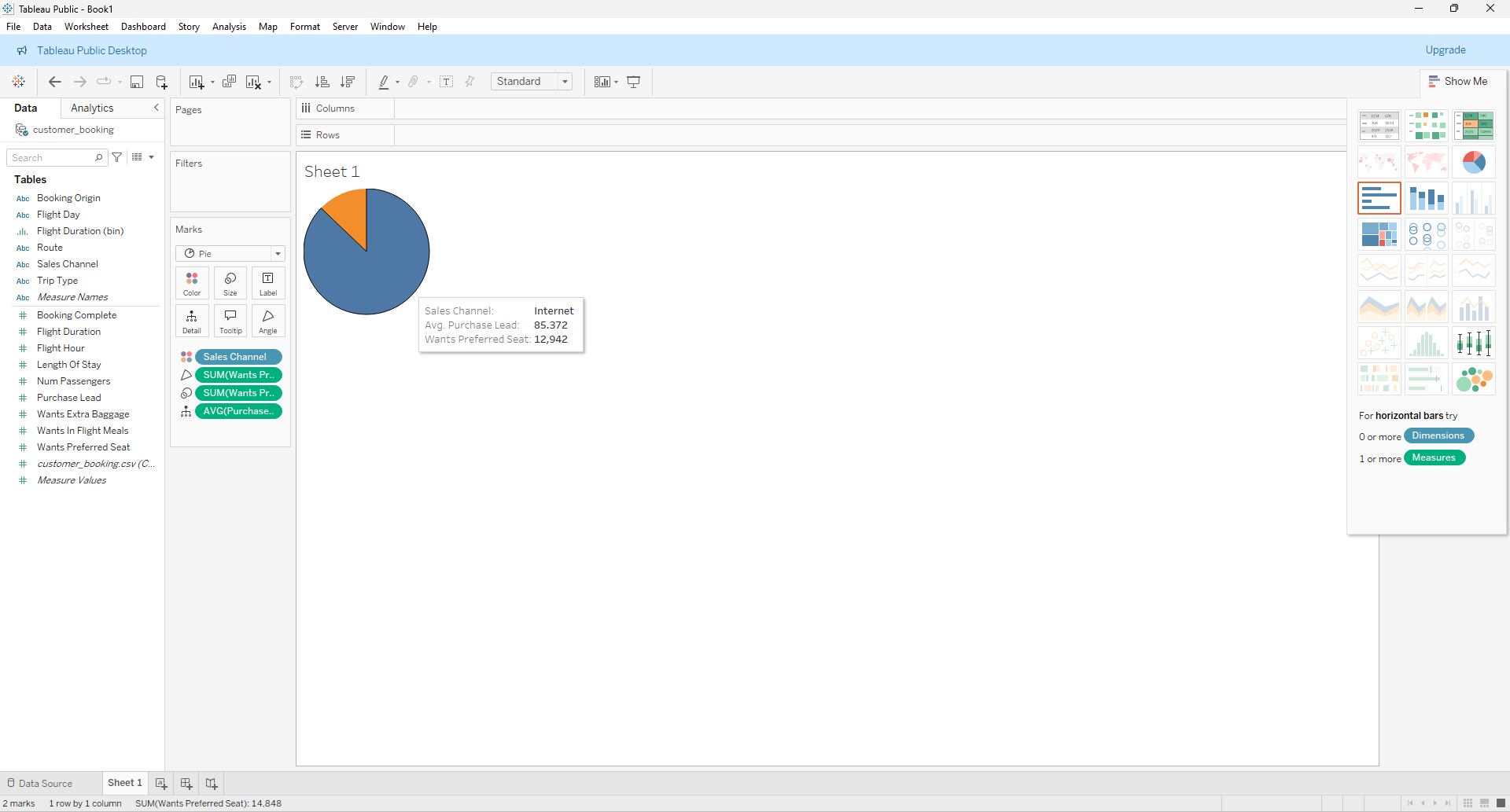
* **Revenue Analysis**: A bar chart to show revenue by flight duration, with calculated fields for average revenue.
* **Sales Trend**: A line chart to display booking trends accompanied by filters for sales channel and trip type.
* **Customer Insights**: A pie chart for analyzing customer preferences.
* **Flight Occupancy**: A heat map to show occupancy rates by route and time.
* **Revenue Analysis**: A bar chart to depict revenue by route and calculated fields for average revenue.

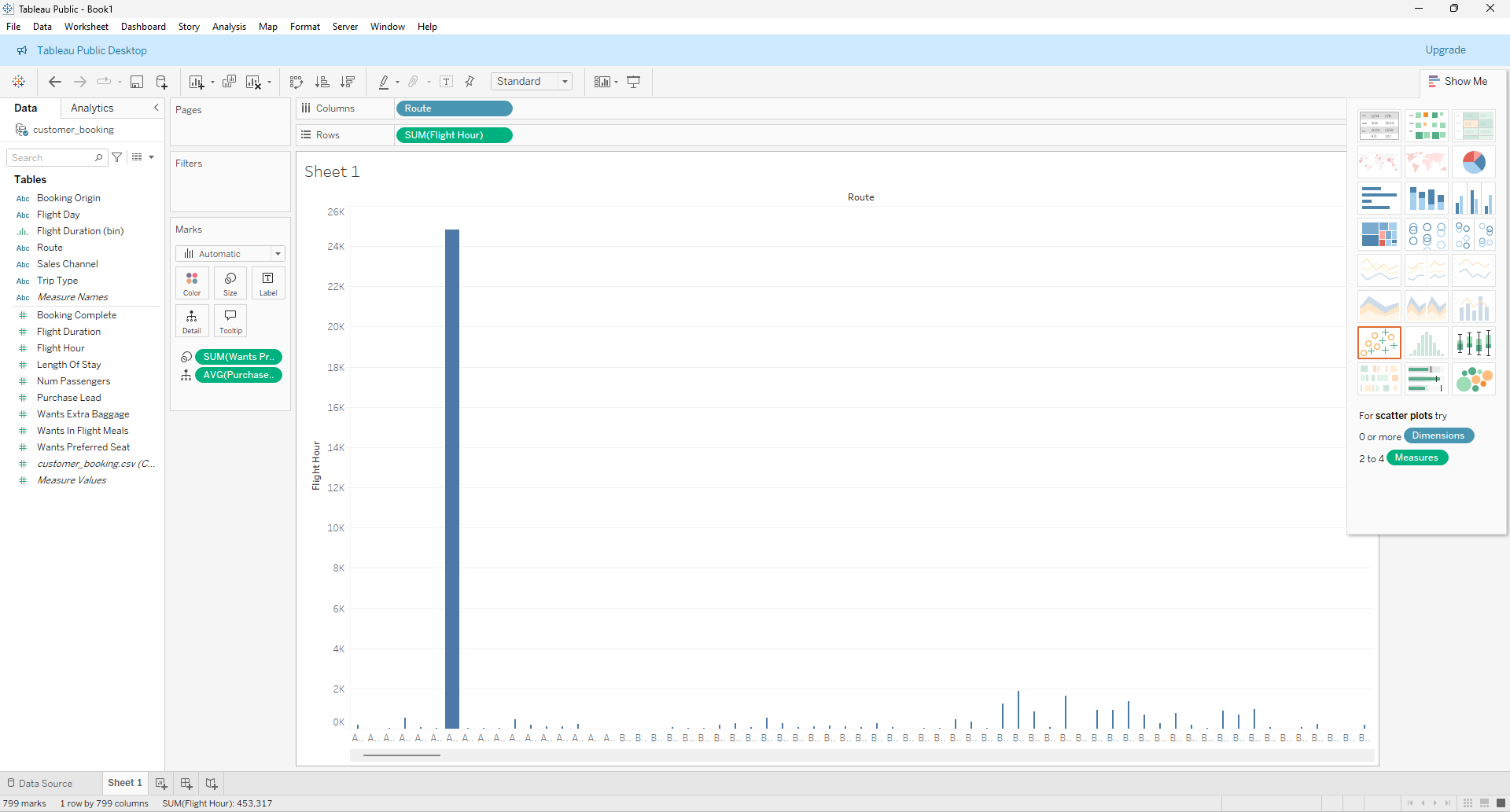
Loading the csv file data in tableau

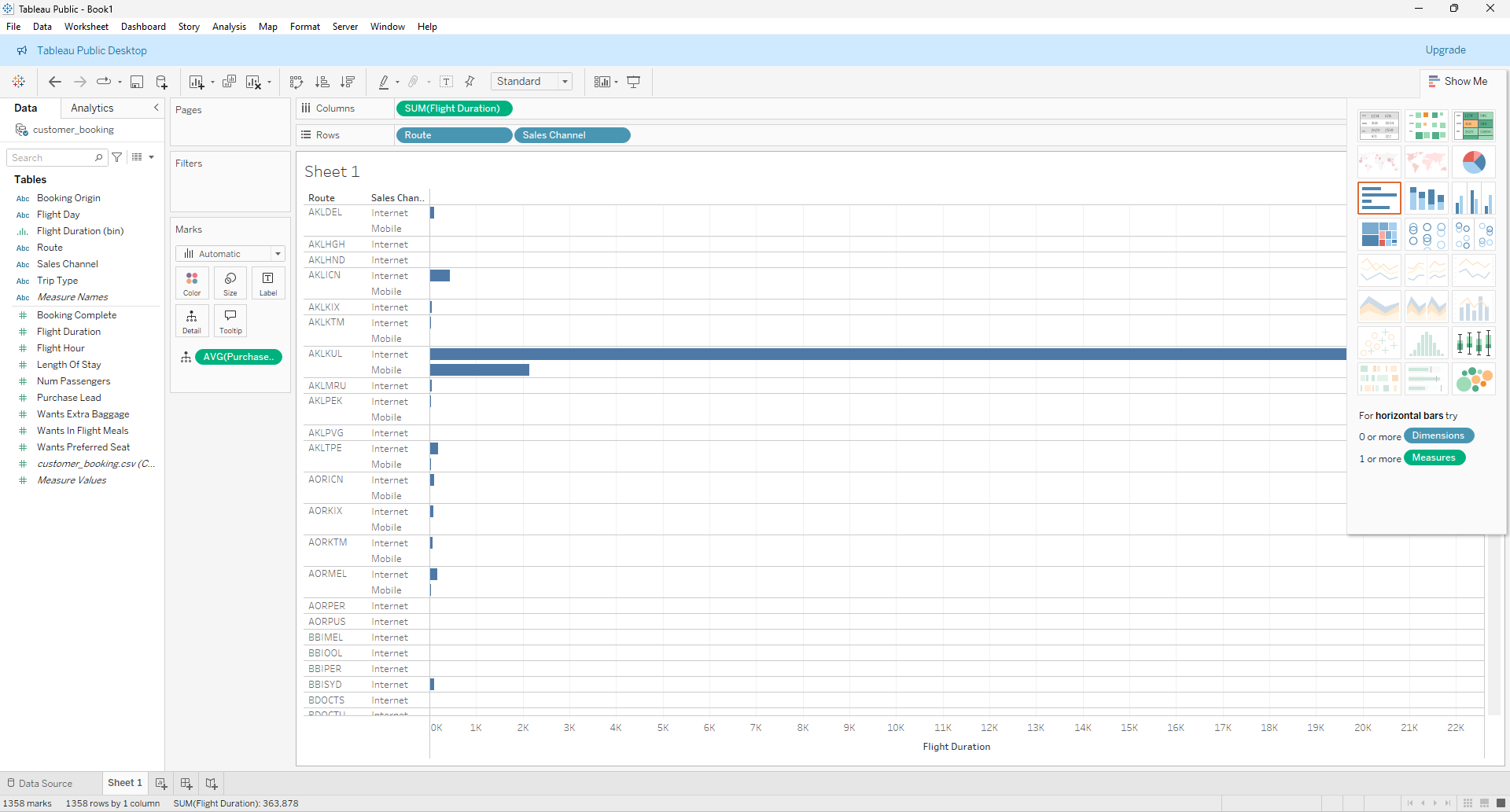
**4.2 REPORTS - SSRS**











These are the derived reports

**Sales Report**: Total number of bookings, revenue, booking cancellations, organized by sales channel and trip type.

**Customer Preferences Report**: Tendency for customer to request extra baggage, or for a customer to request preferred seats, also in-flight meals.

**Operational Efficiency Report**: The occupancy rate of flights, punctuality, reasons why customers chose to book or cancel a flight.

**Route Popularity Report**: The route travelled on the most, length of stay.

**5. CONCLUSIONS**

The data mart has given very useful insights into the flight booking business, this will allow stakeholders to make calculated decisions. The SSRS reports and Tableau dashboard points out key metrics such as operational efficiency.

**6. BIBLIOGRAGHPY**

1. **Kimball, R., & Ross, M. (2013)**. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. John Wiley & Sons.

2. **Inmon, W. H. (2005)**. Building the Data Warehouse. John Wiley & Sons.

3. **Microsoft SQL Server Documentation. (n.d.). Retrieved from [Microsoft Docs] -**https://docs.microsoft.com/en-us/sql/sql-server/.

**7. APPENDICES A CREATE Table queries for Data Warehouse / Data Mart**

These are the SQL Scripts used for constructing the sql server report

--Scripts for dimension tables

CREATE TABLE Customer (

CustomerID INT PRIMARY KEY,

Name NVARCHAR(100),

Age INT,

Gender NVARCHAR(10),

ContactInformation NVARCHAR(200),

LoyaltyStatus NVARCHAR(50)

);

CREATE TABLE Flight (

FlightID INT PRIMARY KEY,

FlightNumber NVARCHAR(50),

Airline NVARCHAR(100),

DepartureAirport NVARCHAR(100),

ArrivalAirport NVARCHAR(100),

Duration TIME

);

CREATE TABLE Date (

DateID INT PRIMARY KEY,

Date DATE,

DayOfWeek NVARCHAR(50),

Month NVARCHAR(50),

Quarter INT,

Year INT

);

CREATE TABLE Route (

RouteID INT PRIMARY KEY,

DepartureAirport NVARCHAR(100),

ArrivalAirport NVARCHAR(100),

Distance INT

);

-- script for fact table

CREATE TABLE Bookings (

BookingID INT PRIMARY KEY,

FlightID INT,

CustomerID INT,

DateID INT,

RouteID INT,

SalesChannel NVARCHAR(50),

TripType NVARCHAR(50),

PurchaseLead INT,

LengthOfStay INT,

FlightHour TIME,

FlightDay NVARCHAR(50),

BookingOrigin NVARCHAR(100),

WantsExtraBaggage BIT,

WantsPreferredSeat BIT,

WantsInFlightMeals BIT,

FlightDuration TIME,

BookingComplete BIT,

NumPassengers INT,

FOREIGN KEY (FlightID) REFERENCES Flight(FlightID),

FOREIGN KEY (CustomerID) REFERENCES Customer(CustomerID),

FOREIGN KEY (DateID) REFERENCES Date(DateID),

FOREIGN KEY (RouteID) REFERENCES Route(RouteID)

);