PART 1

Data Analytics and Visualization

Implementing several techniques for Data Analytics and

Visualization

San Francisco Airport Passenger Data

Student Name: Jeevanandh Vijayan

Student No: 20033071

Table of Contents

1. Introduction

The aim of this project is to deliver a comprehensive insight to passenger trends of San Francisco airport by leveraging a structured data warehouse and data visualization tools. This data warehouse solution is centered around the official dataset of airline passenger data, encompassing various aspects such as airline details, passenger counts, and geographical data.(*Air Traffic Passenger Statistics | DataSF | City and County of San Francisco,* 2024) By transforming this raw data into a structured, easily queryable format the is to aim to unlock insights that can drive strategic decision-making and operational improvements in airport operations.

1.2. Reasons for selecting the subject area and data

The subject area was selected due to several reasons:

- **Personal Preference**: I am interested in airport operations and learning about airline operations.
- **Industry Relevance**: Understanding passenger trends is important for airlines to optimize operations, manage resources, and improve customer satisfaction.
- **Data Availability**: While researching for the topics the data was found readily available with good quality.
- **Challenging and Rewarding**: The complexity of airline operations offers a challenging and rewarding area for data warehousing and business intelligence solutions.

1.3. Vision and Goals

The vision of this project is to create a comprehensive data warehouse that transforms raw airport passenger data into actionable insights through interactive visualizations and reports.

Goals:

- Develop a robust dimensional model for efficient data analysis
- Implement an ETL process for data integration using SSIS
- Create insightful reports and visualizations using SSRS and Tableau
- Enable data-driven decision making for airport management and airlines

1.4. Key Stakeholders

- Airport Management
- Airline Companies
- Marketing Teams
- Customer Service Department

- IT Team

1.5. Business requirements

- Track passenger traffic trends over time
- Analyze airline performance and market share
- Monitor terminal and boarding area utilization
- Compare domestic and international travel patterns
- Identify peak travel periods

2. Schema / Dimensional Model

Schema was developed using Lucid chart and it follows star schema model which includes fact and dimesnion tables for capturing data from the dataset.
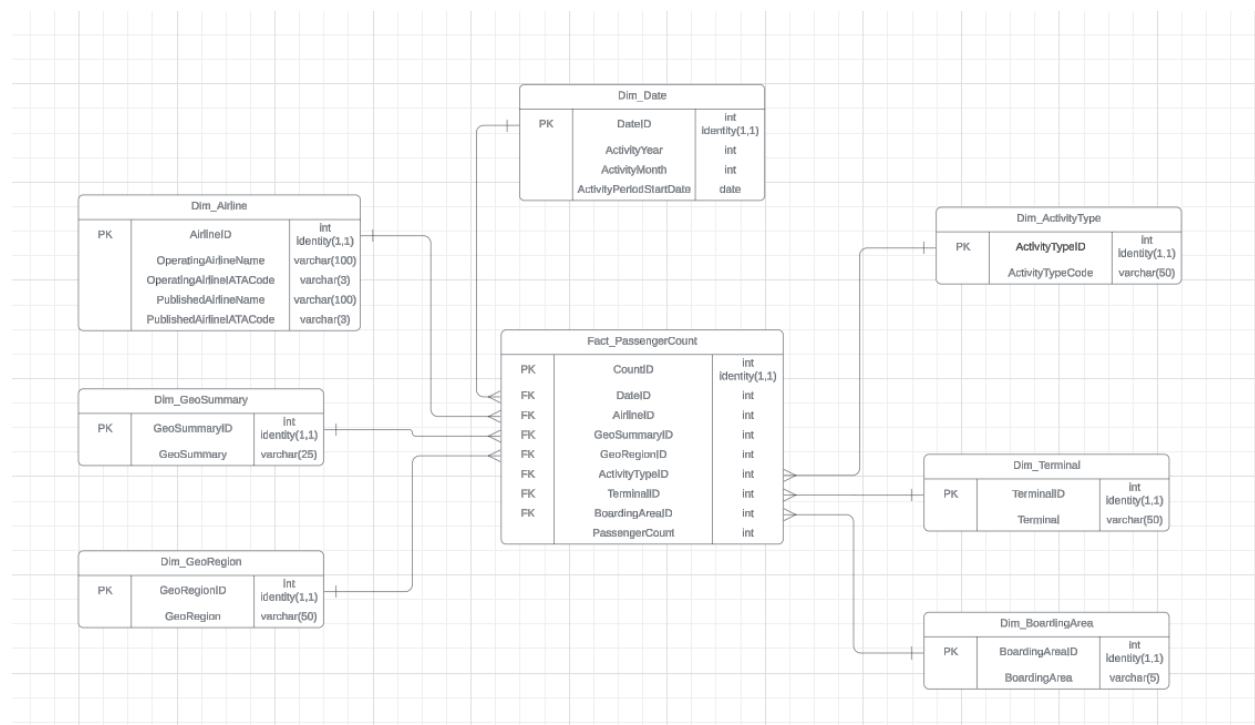
Fact Table:
Fact_PassengerCount
Stores passenger counts and references to dimension tables.

Dimension Tables:

Dim_Date: Contains date-related information like month and year.

Dim_GeoRegion: Stores geographic regions.
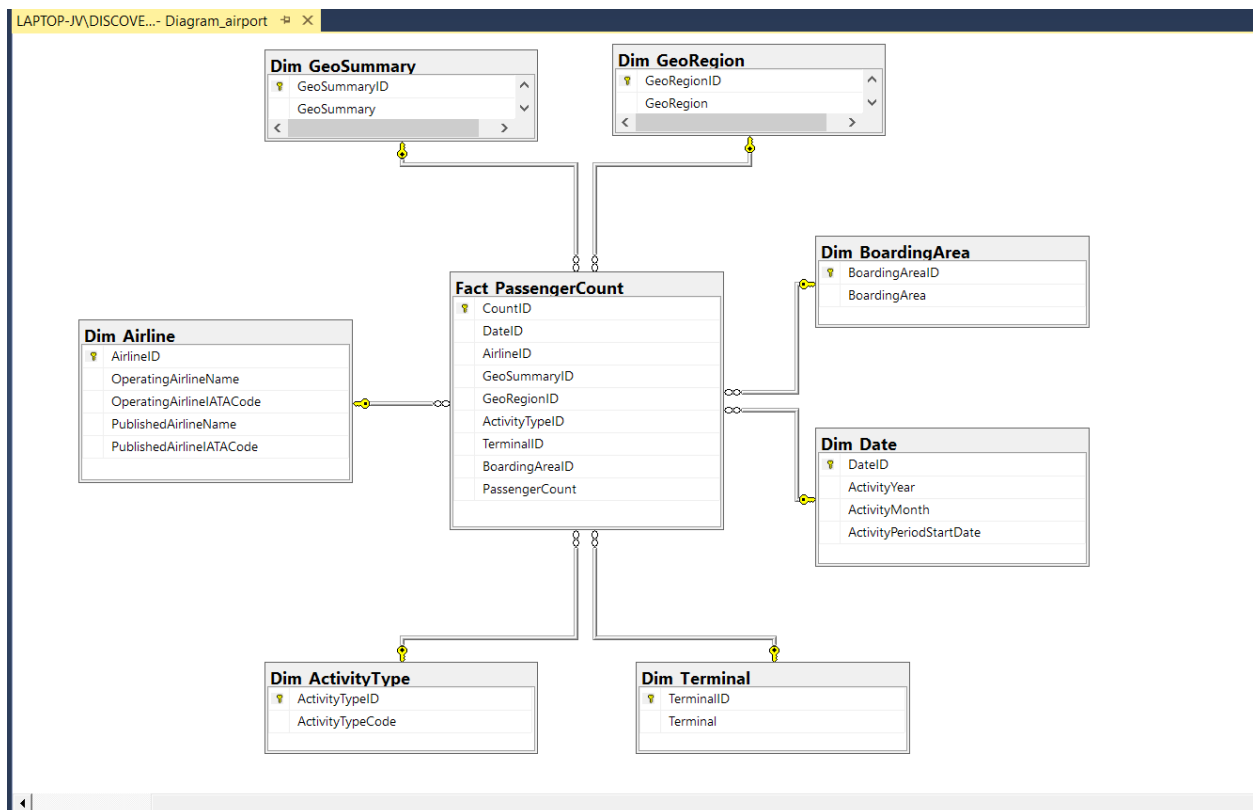
Dim_Airline: Captures airline names and codes.
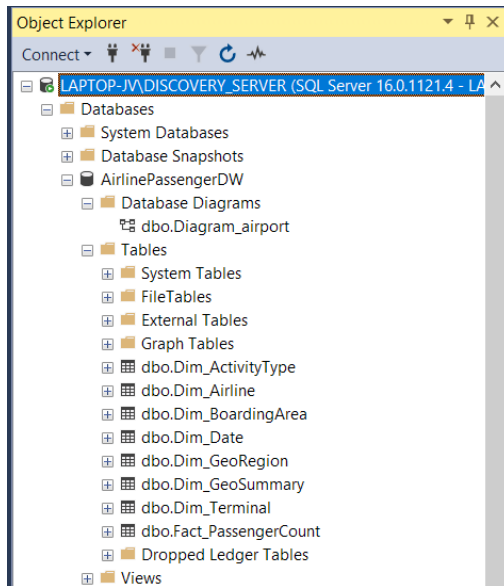
Dim_BoardingArea: Contains boarding areas.

Dim_ActivityType: Stores activity types.

Dim_GeoSummary: Stores geographic regions.

Dim_Terminal: Terminals of the airport.

3. Implementation of Data Warehouse

4. ETL to Populate the Data Warehouse / Data Mart

The ETL (Extract, Transform, Load) process is important for ensuring that data is accurately and efficiently loaded into the data warehouse.(chugugrace, 2023) The process of loading data into the data ware house database was done using Visual Studio 2022.(*Visual Studio 2022 IDE - Programming Tool for Software Developers*, *2024*)
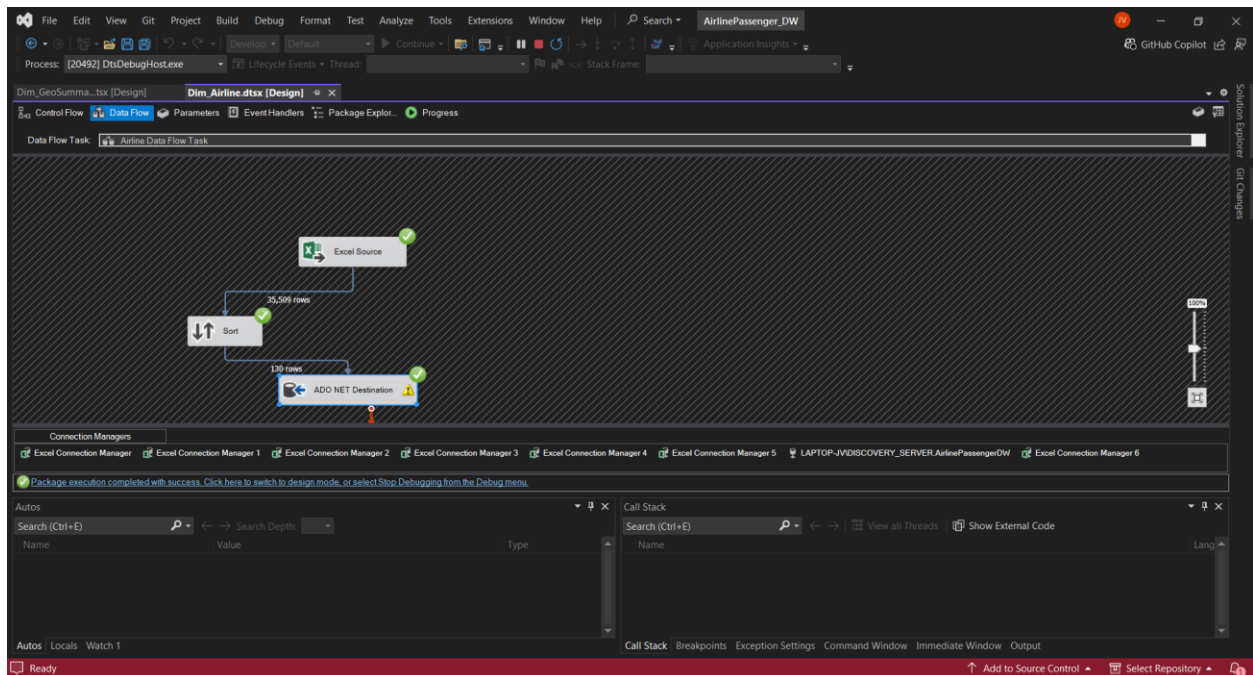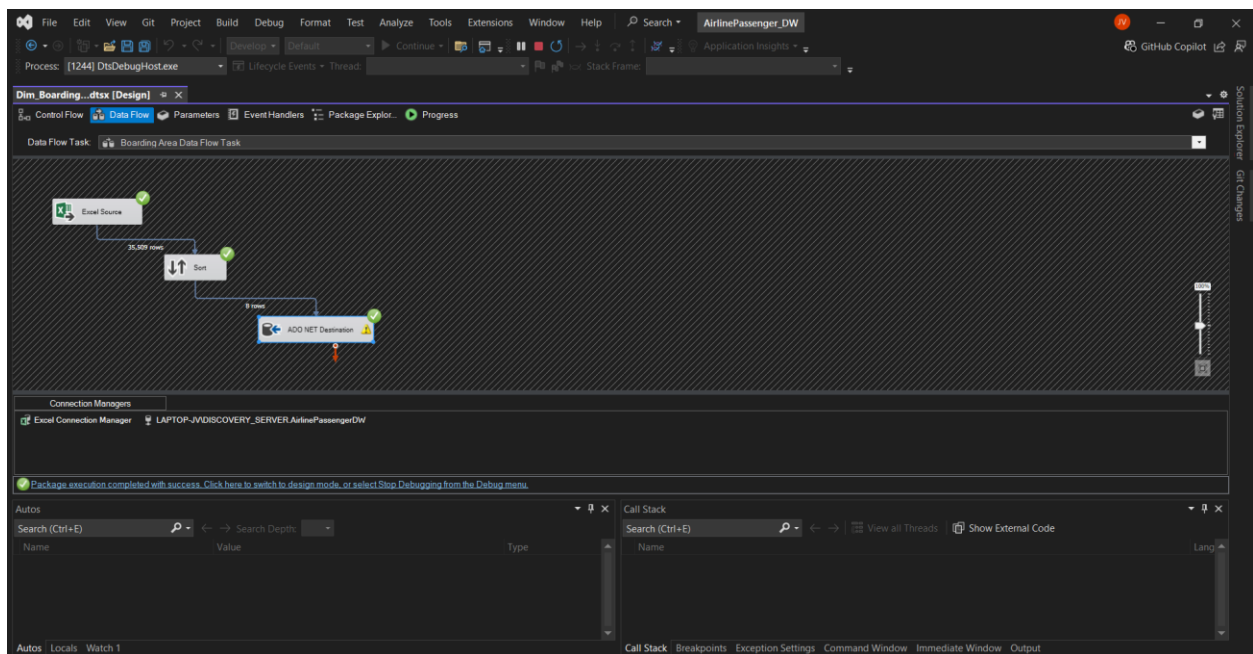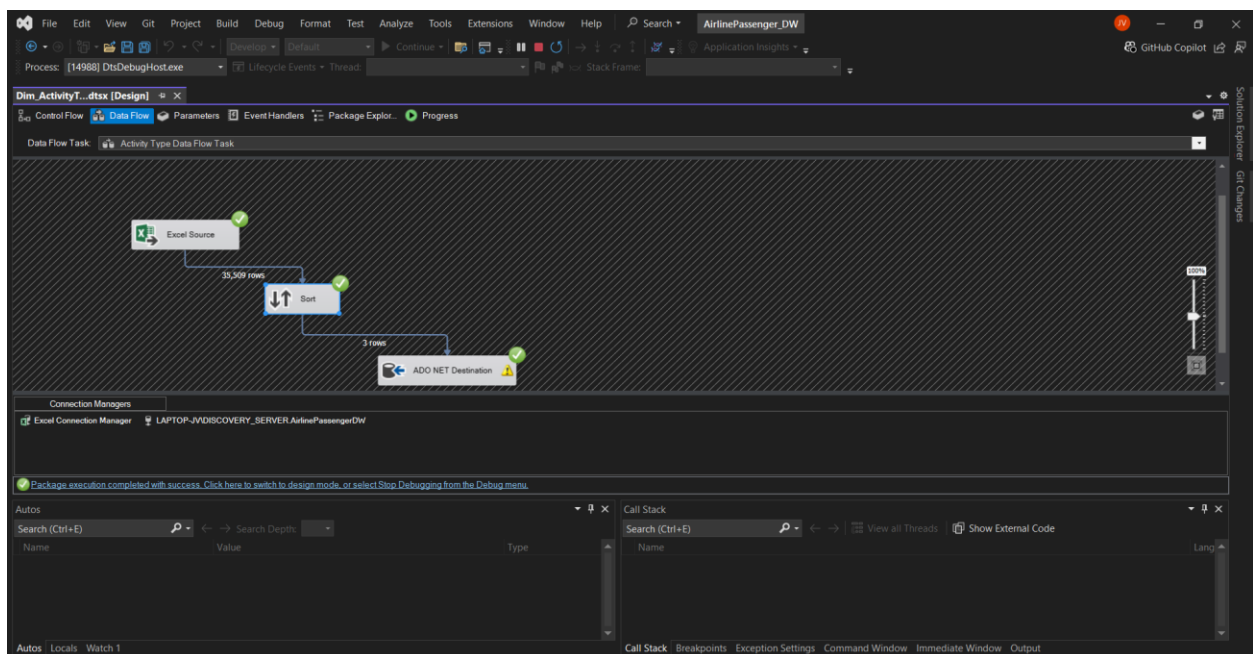
Dim_Date
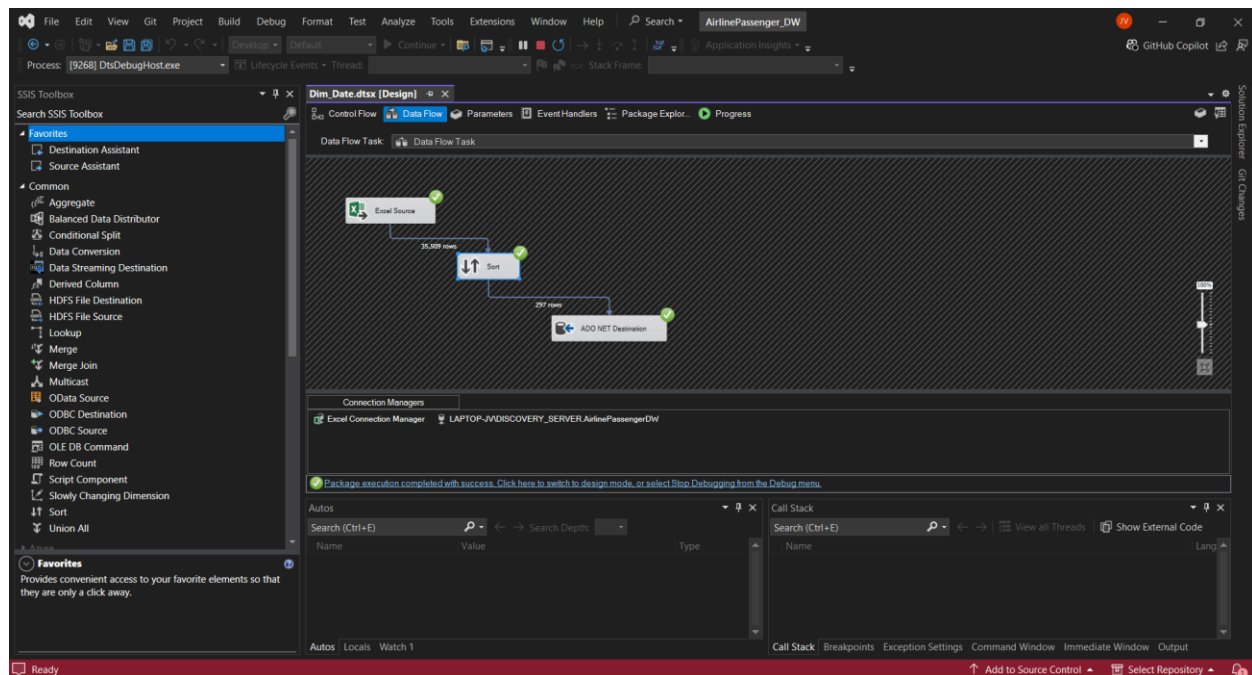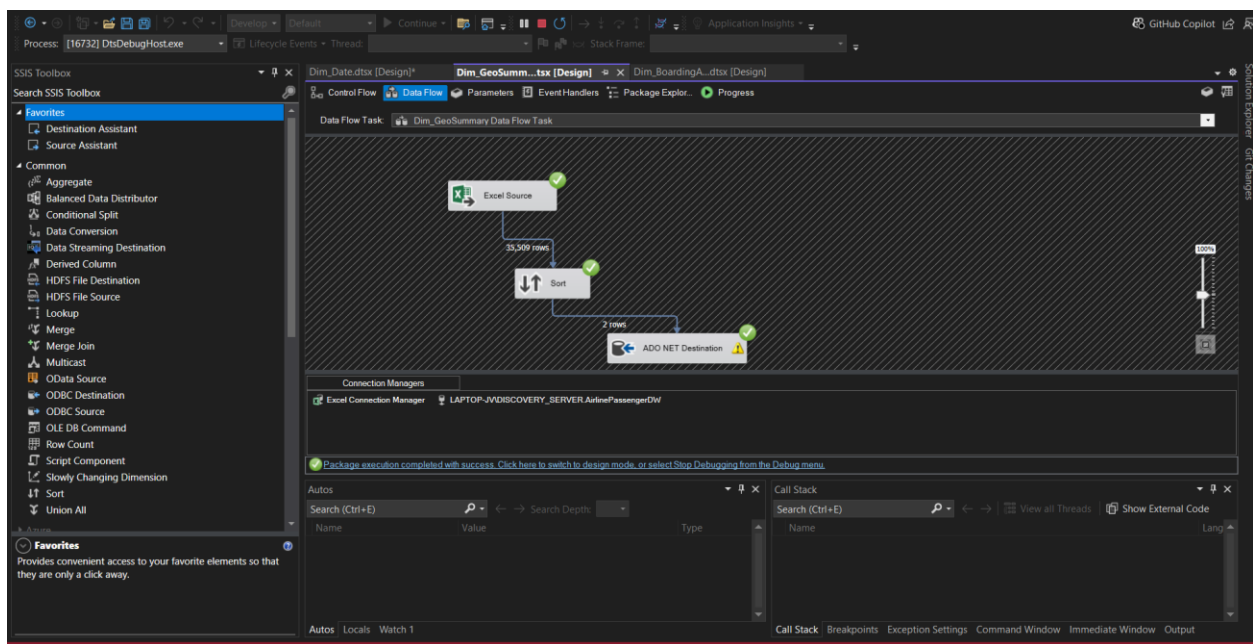
Dim_GeoRegion
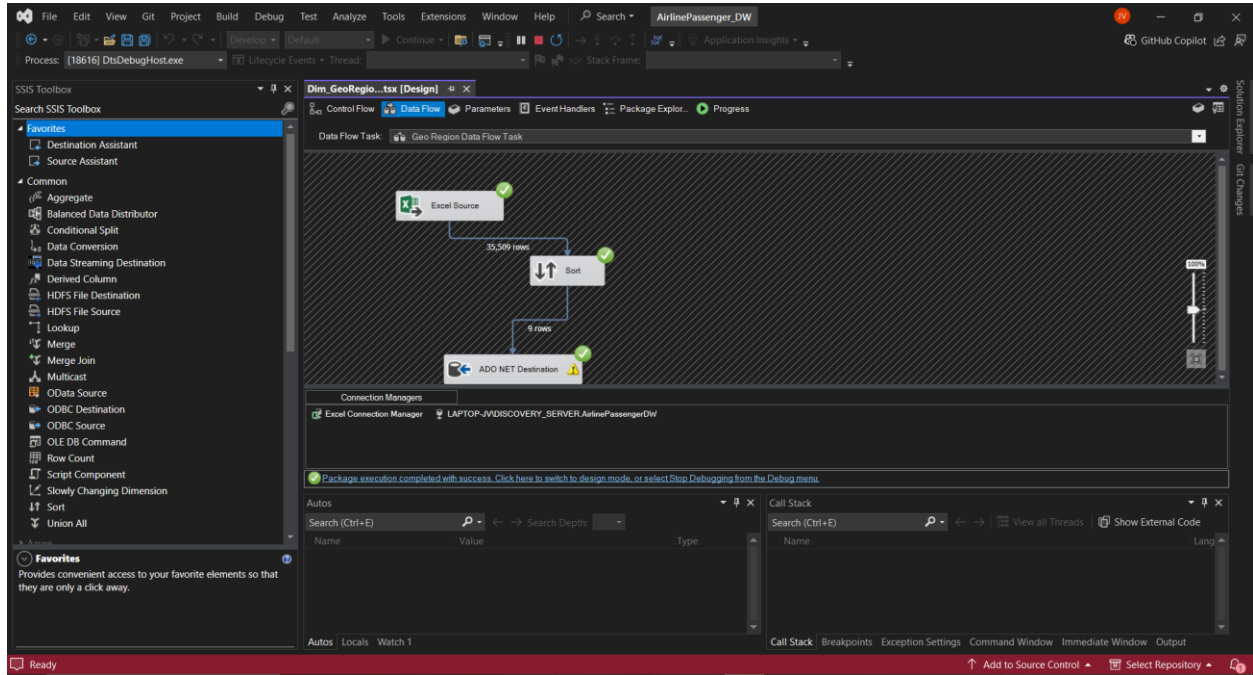
Dim_Airline

Dim_BoardingArea

Dim_ActivityType
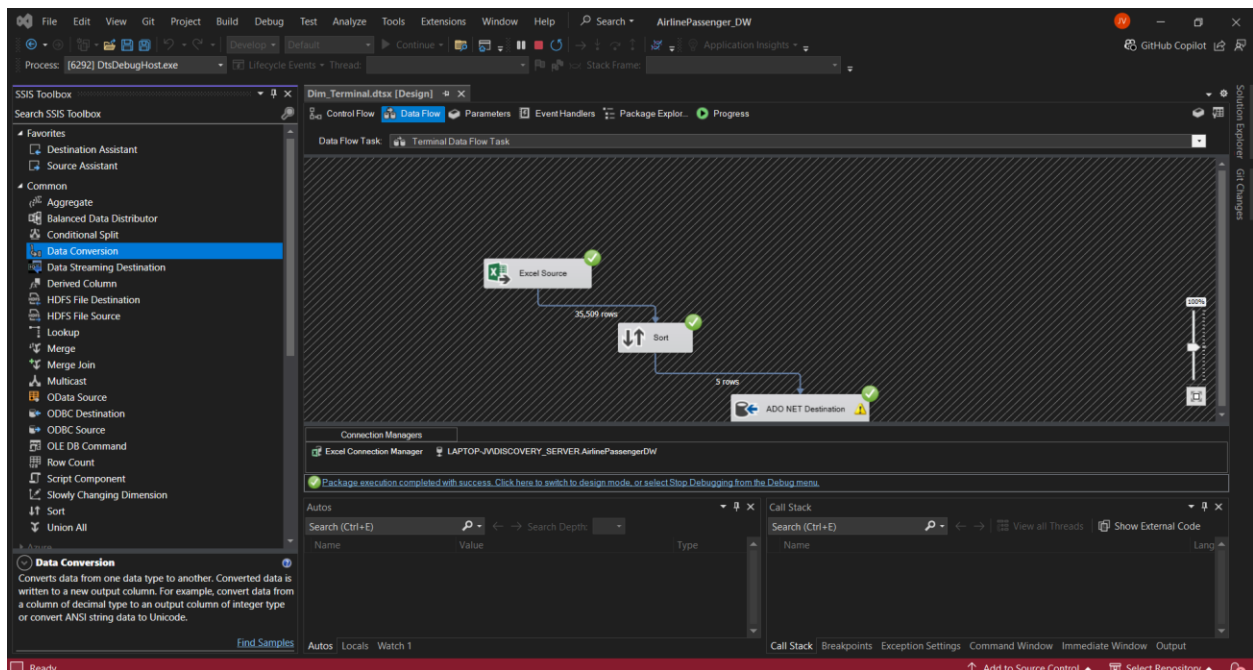
Dim_GeoSummary

Dim_Terminal

Fact_PassengerCount

Used excel datsource with lookup function in ssis for each dimension for adding passenger count in fact table. Data conversion used to data mismatch error.
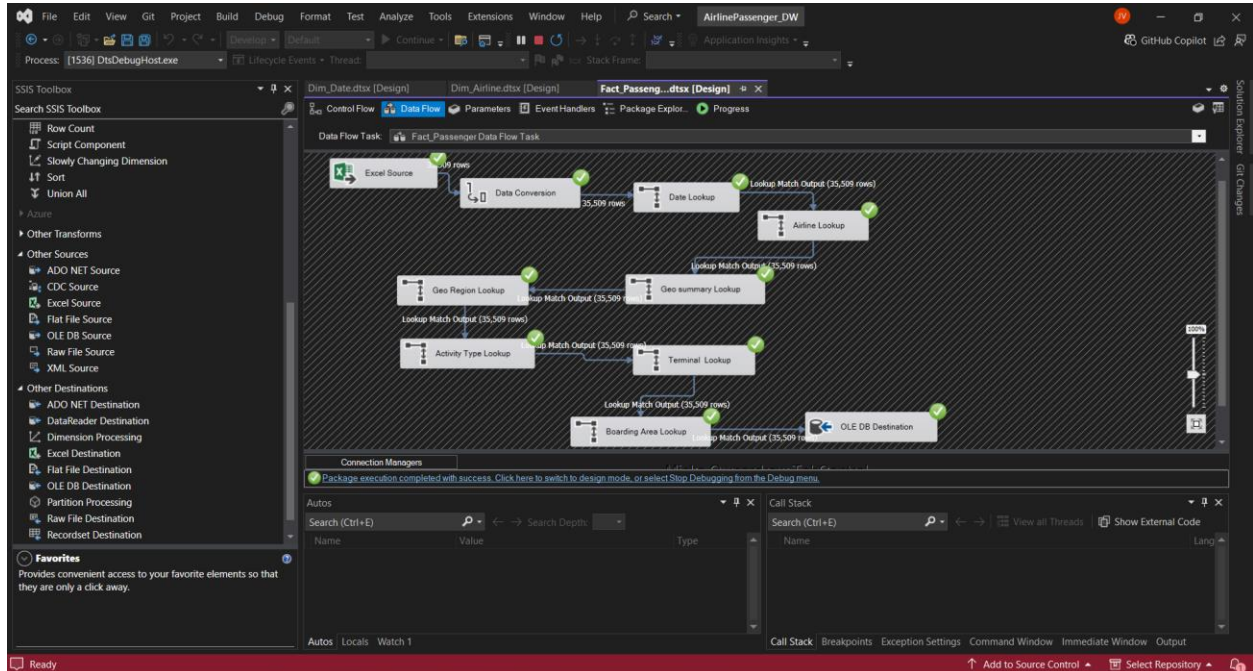
4. Visualizations and Reports

Four SSRS reports were created for passenger count analytics.

1.Airline Monthly Passenger Count

This report shows the monthly passenger count of each airline at the San Franciso airport with a column of year for refernce.

2.Geographic Regional Passenger Count

This report shows total passenger count according to geographical region with activity whether entry, exit or transfer.

3.Terminal and Boarding Area Passenger Count

Total passenger count terminal wise with boarding area.

Terminal and Boarding Area Passenger Count

Terminal	Boarding Area	Total Passenger Count
International	A	119229637
International	G	130848681
Other	Other	182
Terminal 1	A	8503529
Terminal 1	B	128311526
Terminal 1	C	94915967
Terminal 2	D	97357905
Terminal 3	E	106477320
Terminal 3	F	314156310

4. Yearly Passenger Count of Domestic and International

This ssrs report shows early passenger count domestic and international in San Francisco airport.

Yearly Passenger Count

Activity Year	Geo Summary	Total Passenger Count
1999	Domestic	17180397
1999	International	3770059
2000	Domestic	32469008
2000	International	8155234
2001	Domestic	26665077
2001	International	7535557
2002	Domestic	23699764
2002	International	7364310
2003	Domestic	22259958
2003	International	6763119
2004	Domestic	24778094
2004	International	7668040
2005	Domestic	24906743
2005	International	8192449
2006	Domestic	24745933

4.1. TABLEAU Visualizations

An interactive dashboard was created with four tableau sheets to visualize the data in warehouse showcasing passenger count of airline, geo region, terminal, and yearly period wise at San Francisco airport. On clicking on certain visualizations data will be filtered out to show data depending on that filter.

Microsoft SQL server with warehouse database was connected to tableau for configuring the dashboard.
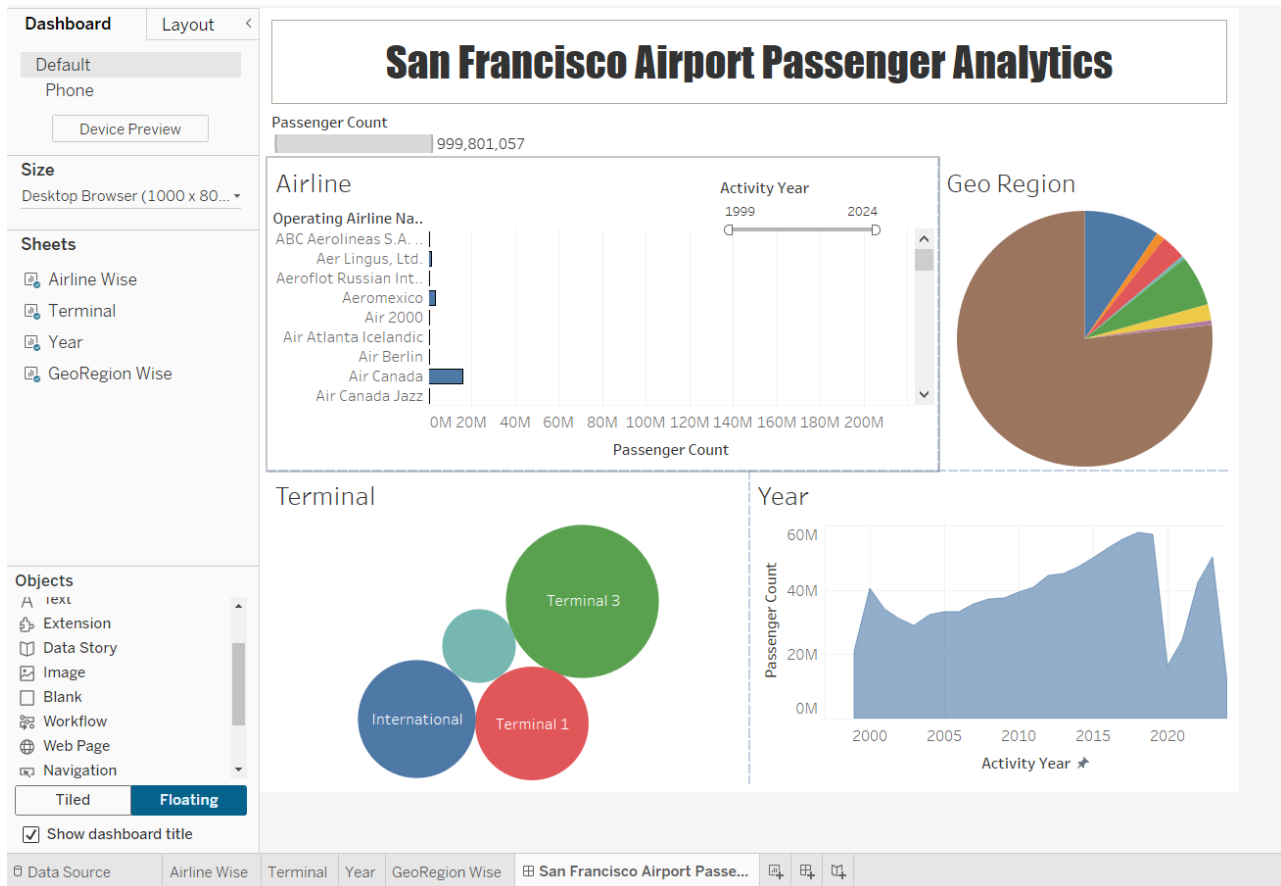
5. Conclusions

The "Airport Passenger Data Analysis" project demonstrates the importance of data warehousing and visualization in the airline industry. By structuring complex data into a robust warehouse and creating insightful visualizations and reports, stakeholders can make informed decisions that drive operational efficiency and strategic growth. rephrase more human tone. This project enabled to learn SSIS and SSRS in depth.

Bibliography

Air Traffic Passenger Statistics | DataSF | City and County of San Francisco (no date). Available at: https://data.sfgov.org/Transportation/Air-Traffic-Passenger-Statistics/rkru-6vcg/data_preview (Accessed: 12 July 2024).

chugugrace (2023) *SQL Server Integration Services - SQL Server Integration Services (SSIS)*. Available at: https://learn.microsoft.com/en-us/sql/integration-services/sql-server-integration-services?view=sql-server-ver16 (Accessed: 12 July 2024).

Visual Studio 2022 IDE - Programming Tool for Software Developers (no date). Available at: https://visualstudio.microsoft.com/vs/ (Accessed: 12 July 2024).

Appendix A – CREATE Table queries for Data Warehouse / Data Mart

```sql
-- Create the database
CREATE DATABASE AirlinePassengerDW;
GO
-- Use the new database
USE AirlinePassengerDW;
GO
-- Create the dimension tables
CREATE TABLE [Dim_Date] (
  [DateID] int IDENTITY(1,1),
  [ActivityYear] int,
  [ActivityMonth] int,
  [ActivityPeriodStartDate] date,
  PRIMARY KEY ([DateID])
);
CREATE TABLE [Dim_GeoRegion] (
  [GeoRegionID] int IDENTITY(1,1),
  [GeoRegion] varchar(50),
  PRIMARY KEY ([GeoRegionID])
);
CREATE TABLE [Dim_Airline] (
  [AirlineID] int IDENTITY(1,1),
  [OperatingAirlineName] varchar(100),
  [OperatingAirlineIATACode] varchar(3),
  [PublishedAirlineName] varchar(100),
  [PublishedAirlineIATACode] varchar(3),
  PRIMARY KEY ([AirlineID])
);
CREATE TABLE [Dim_BoardingArea] (
  [BoardingAreaID] int IDENTITY(1,1),
  [BoardingArea] varchar(5),
  PRIMARY KEY ([BoardingAreaID])
);
CREATE TABLE [Dim_ActivityType] (
  [ActivityTypeID] int IDENTITY(1,1),
  [ActivityTypeCode] varchar(50),
  PRIMARY KEY ([ActivityTypeID])
);
CREATE TABLE [Dim_GeoSummary] (
  [GeoSummaryID] int IDENTITY(1,1),
  [GeoSummary] varchar(25),
  PRIMARY KEY ([GeoSummaryID])
);
CREATE TABLE [Dim_Terminal] (
  [TerminalID] int IDENTITY(1,1),
  [Terminal] varchar(50),
  PRIMARY KEY ([TerminalID])
);
-- Create the fact table
CREATE TABLE [Fact_PassengerCount] (
  [CountID] int IDENTITY(1,1),
  [DateID] int,
  [AirlineID] int,
  [GeoSummaryID] int,
  [GeoRegionID] int,
  [ActivityTypeID] int,
```

```sql
    [TerminalID] int,
    [BoardingAreaID] int,
    [PassengerCount] int,
    PRIMARY KEY ([CountID]),
    CONSTRAINT [FK_Fact_PassengerCount.GeoSummaryID]
      FOREIGN KEY ([GeoSummaryID])
        REFERENCES [Dim_GeoSummary]([GeoSummaryID]),
    CONSTRAINT [FK_Fact_PassengerCount.DateID]
      FOREIGN KEY ([DateID])
        REFERENCES [Dim_Date]([DateID]),
    CONSTRAINT [FK_Fact_PassengerCount.ActivityTypeID]
      FOREIGN KEY ([ActivityTypeID])
        REFERENCES [Dim_ActivityType]([ActivityTypeID]),
    CONSTRAINT [FK_Fact_PassengerCount.BoardingAreaID]
      FOREIGN KEY ([BoardingAreaID])
        REFERENCES [Dim_BoardingArea]([BoardingAreaID]),
    CONSTRAINT [FK_Fact_PassengerCount.TerminalID]
      FOREIGN KEY ([TerminalID])
        REFERENCES [Dim_Terminal]([TerminalID]),
    CONSTRAINT [FK_Fact_PassengerCount.GeoRegionID]
      FOREIGN KEY ([GeoRegionID])
        REFERENCES [Dim_GeoRegion]([GeoRegionID]),
    CONSTRAINT [FK_Fact_PassengerCount.AirlineID]
      FOREIGN KEY ([AirlineID])
        REFERENCES [Dim_Airline]([AirlineID])
);
```

Appendix B – SSRS Report Queries

1.Airline Monthly Passenger Count

```sql
SELECT d.ActivityYear, d.ActivityMonth, a.OperatingAirlineName, SUM(f.PassengerCount) AS
TotalPassengerCount
FROM     Fact_PassengerCount AS f INNER JOIN
                Dim_Date AS d ON f.DateID = d.DateID INNER JOIN
                Dim_Airline AS a ON f.AirlineID = a.AirlineID
GROUP BY d.ActivityYear, d.ActivityMonth, a.OperatingAirlineName
ORDER BY d.ActivityYear, d.ActivityMonth, a.OperatingAirlineName
```

2.Geographic Regional Passenger Count

```sql
SELECT   d.ActivityYear,   g.GeoRegion,   at.ActivityTypeCode,   SUM(f.PassengerCount)   AS
TotalPassengerCount
FROM     Fact_PassengerCount AS f INNER JOIN
                Dim_Date AS d ON f.DateID = d.DateID INNER JOIN
                Dim_GeoRegion AS g ON f.GeoRegionID = g.GeoRegionID INNER JOIN
                Dim_ActivityType AS at ON f.ActivityTypeID = at.ActivityTypeID
GROUP BY d.ActivityYear, g.GeoRegion, at.ActivityTypeCode
ORDER BY d.ActivityYear, g.GeoRegion, at.ActivityTypeCode
```

3.Terminal and Boarding Area Passenger Count

```sql
SELECT
    t.Terminal,
    b.BoardingArea,
```

```sql
    SUM(f.PassengerCount) AS TotalPassengerCount
FROM
    Fact_PassengerCount f
JOIN
    Dim_Terminal t ON f.TerminalID = t.TerminalID
JOIN
    Dim_BoardingArea b ON f.BoardingAreaID = b.BoardingAreaID
GROUP BY
    t.Terminal, b.BoardingArea
ORDER BY
    t.Terminal, b.BoardingArea;
```

4. Yearly Passenger Count of Domestic and International

```sql
SELECT
    d.ActivityYear,
    gs.GeoSummary,
    SUM(f.PassengerCount) AS TotalPassengerCount
FROM
    Fact_PassengerCount f
JOIN
    Dim_Date d ON f.DateID = d.DateID
JOIN
    Dim_GeoSummary gs ON f.GeoSummaryID = gs.GeoSummaryID
GROUP BY
    d.ActivityYear, gs.GeoSummary
ORDER BY
    d.ActivityYear, gs.GeoSummary;
```