

SAMSUNG

TOGETHER FOR TOMORROW!
Enabling People

AirSense

SIC 7

BD Course
G9

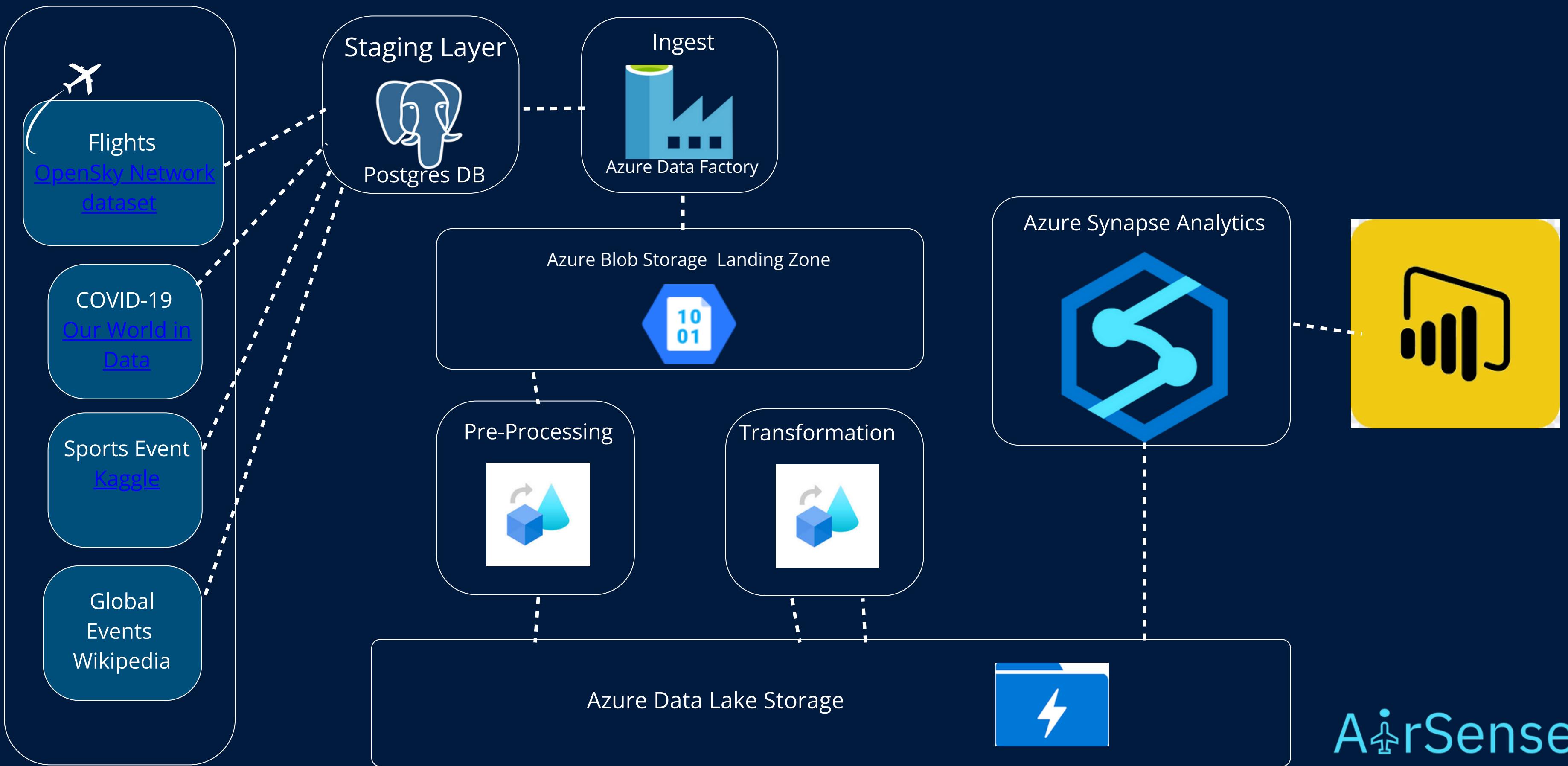


AirSense

Agenda

- | Team
- | **Introduction**
- | Problem Statement
- | Methodology
- | Key Tasks
- | Results and Insights
- | Conclusion
- | Q&A

Workflow



Introduction

This project analyzes how global events shape international air travel by building an end-to-end data pipeline using Azure. Flight data, pandemic indicators, and major events were collected, cleaned, and transformed through PostgreSQL and Azure Data Factory, then stored in Azure Synapse for advanced analysis. Insights were visualized in Power BI to show how flight activity responds to global disruptions and major happenings. The result is a system that connects worldwide trends with travel behavior and provides a foundation for informed decision making.



Agenda

- | Team
- | Introduction
- | **Problem Statement**
- | Key Tasks
- | Methodology
- | Results and Insights
- | Conclusion
- | Q&A

Problem Statement

In today's world, global events—whether health crises, sporting events, political developments, or cultural phenomena—have a direct impact on human mobility and air travel. However, decision-makers often lack tools to quickly and accurately understand these effects, particularly in international air travel, one of the fastest-responding sectors to global events.



Agenda

- | Team
- | Introduction
- | Problem Statement
- | **Methodology**
- | Key Tasks
- | Results and Insights
- | Conclusion
- | Q&A

Methodology

- AirSense follows a structured workflow to convert raw data into meaningful insights:
- - Data Collection
 - Pre-processing
 - Modeling
 - Transformation
 - Pipelines
 - Visualization

Agenda

- | Team
- | Introduction
- | Problem Statement
- | Methodology
- | **Key Tasks**
- | Results and Insights
- | Conclusion
- | Q&A

Key Tasks

- **Data Acquisition:** Collected global flight records, COVID-19 statistics, and major event data for 2021–2022.
- **Data Cleaning & Preprocessing:** Standardized formats, removed duplicates, handled missing values, and validated data quality.
- **Dimensional Modeling:** Built a Galaxy Schema with conformed dimensions and fact tables to integrate heterogeneous datasets.
- **Transformation & ELT:** Applied aggregations, derived metrics (flight volumes, estimated passenger capacity, event impact), and linked facts to dimensions using Azure Data Factory.
- **Pipeline Orchestration:** Automated ETL workflow to maintain referential integrity and optimize performance.
- **Analysis & Visualization:** Produced dashboards and insights showing the impact of global events on air travel and tourism trends.

Key Tasks - Data Acquisition

Description:

- Collected global flight records, COVID-19 statistics, and major event data for 2021–2022.

Outcome:

- Flight data contains detailed flight movements, aircraft info, origin/destination, timestamps, and positions.
- COVID-19 data contains cumulative excess deaths and confirmed cases by country and date.
- Events and sports data captured major global and sports events with relevant attributes.
- Next Step: Move data to PostgreSQL for structured storage.

Flights
[OpenSky Network dataset](#)

COVID-19
[Our World in Data](#)

Sports Event
[Kaggle](#)

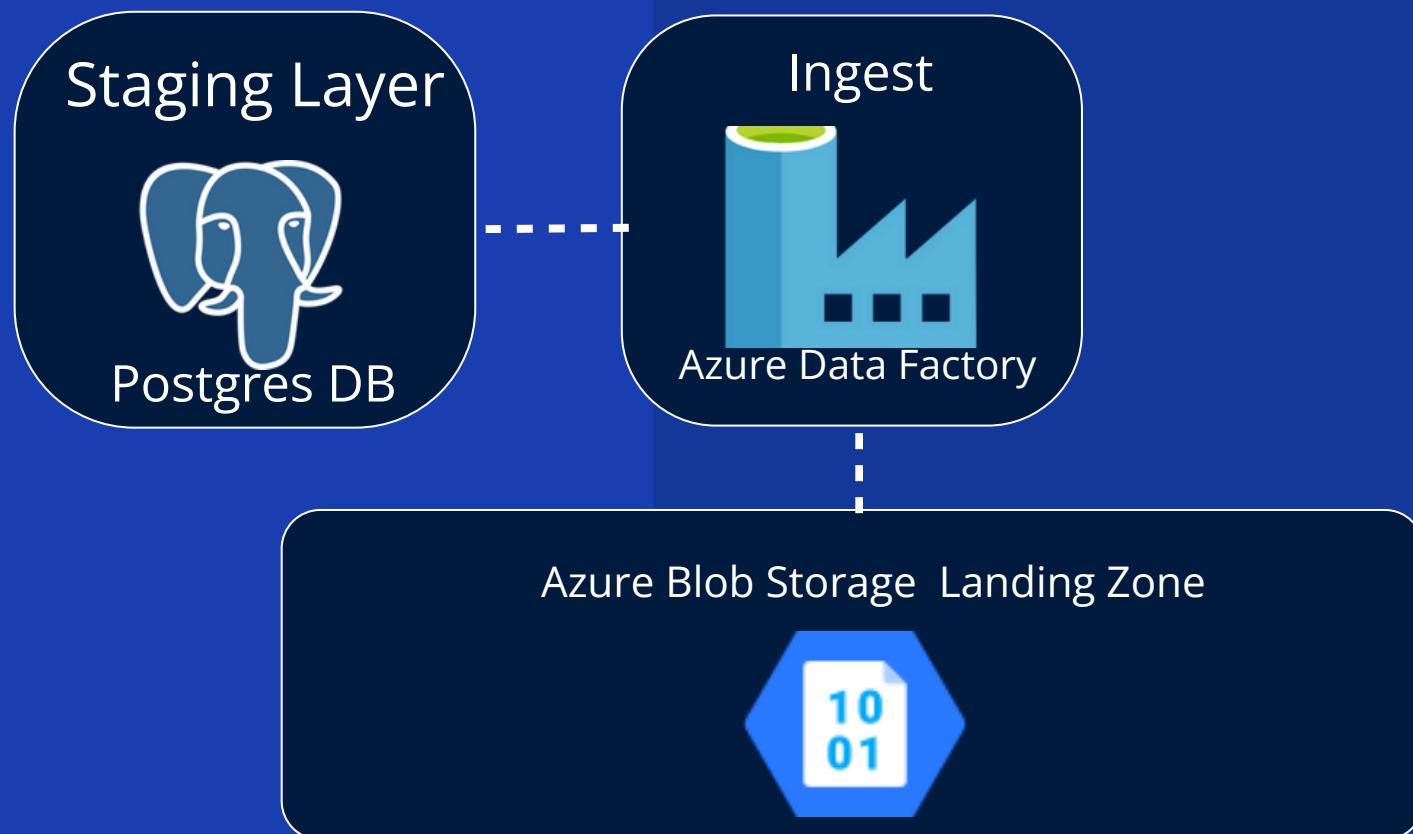
Global Events
[Wikipedia](#)

Key Tasks - Data Ingestion

Goal: Move raw flight & event data from local sources to Azure for processing.

Steps:

1. Create Azure Blob Storage
 - Container raw-data for all raw datasets.
2. Connect Sources via ADF
 - PostgreSQL (local) → Self-Hosted IR
 - Blob Storage (cloud) → Sink
3. Define Datasets
 - Source: PostgreSQL tables
 - Sink: Blob Storage, format Parquet
4. Build Pipeline & Copy Data
 - ADF Copy Activity moves data from local → cloud
5. Run Pipeline
 - Raw data ingested in cloud, compressed & ready for transformations



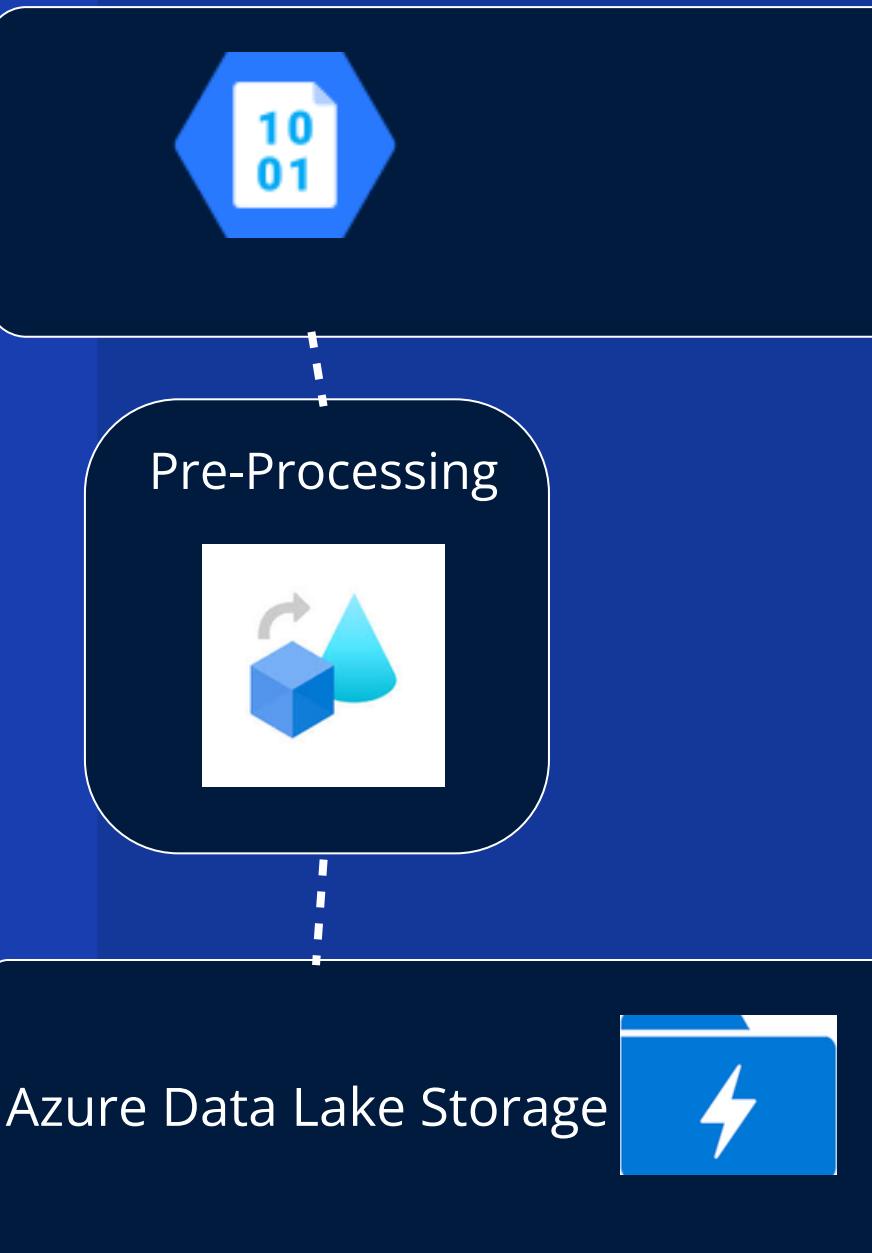
Key Tasks - Data Cleaning & Preprocessing

Description:

- Standardized formats, handled missing values, validated data quality.

Outcome:

- Flights: kept essential fields (flight_id, icao24, typecode, origin, destination, date), removed nulls.
- COVID: standardized date, removed records with missing total_confirmed, kept key columns.
- Sports: removed invalid entries, normalized fields, extracted match_id, date, attendance, stadium info.
- Global Events: standardized date, normalized categories, removed duplicates.



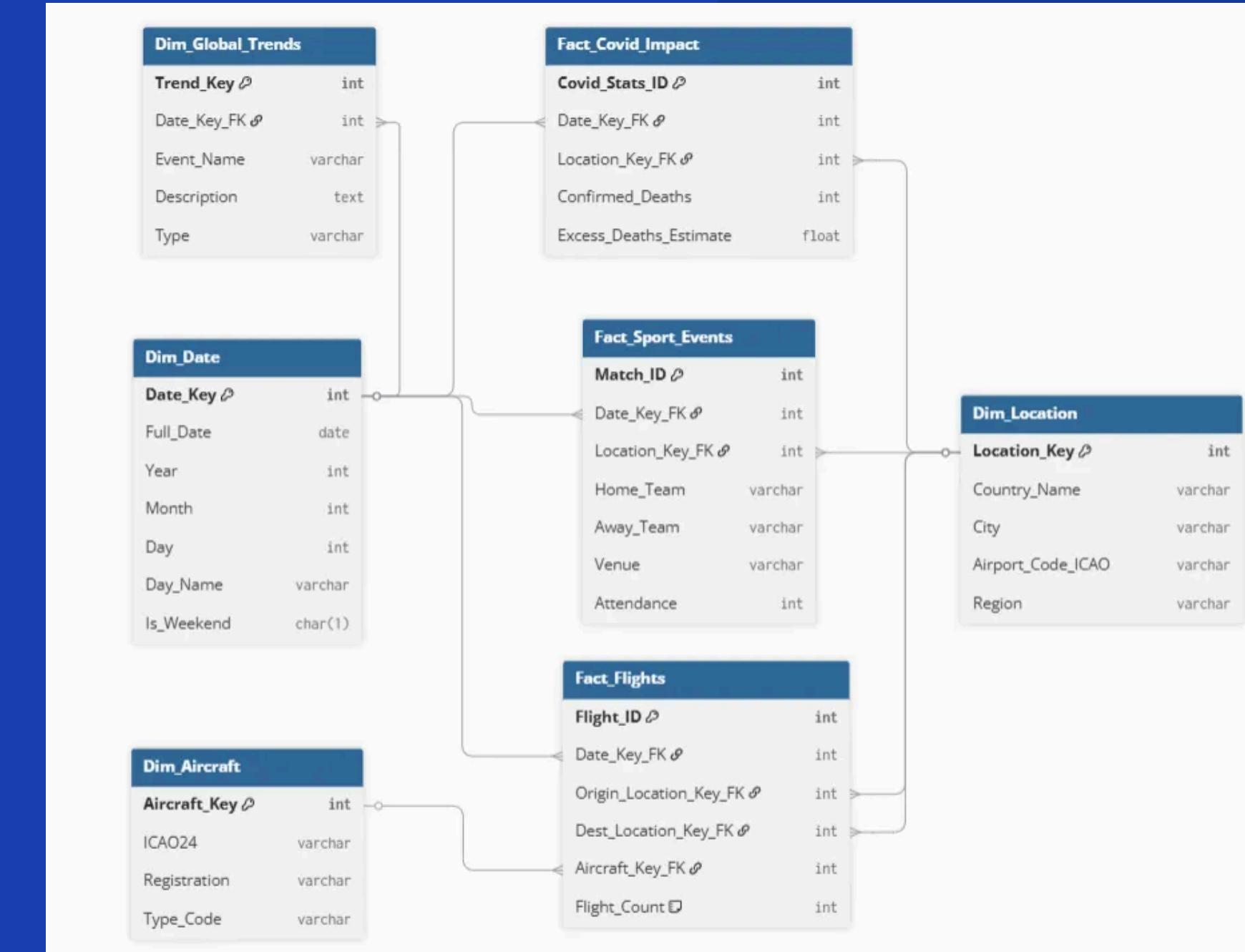
Next Step: Load cleaned data into Azure Data Lake for downstream processing.

Key Tasks - Dimensional Modeling

Goal: Organize heterogeneous data for easy analysis and linking.

Steps:

1. Identify Fact & Dimension Tables
 - Fact Table: Flights (all flight movements)
 - Dimensions: Date, Location, Events, Sports, COVID
2. Design Galaxy Schema
 - Conformed dimensions allow linking across datasets
 - Supports analysis across multiple perspectives (time, location, event type)



Key Tasks - Transformation

Goal: Convert raw data into meaningful, analysis-ready datasets.

Steps:

1. Aggregations & Metrics

- Compute flight volumes per day/month/country
- Derive metrics like estimated passenger capacity
- Calculate event impact on flight movements

2. Link Facts to Dimensions

- Join fact table (Flights) with dimensions (Date, Location, Events, Sports, COVID)

3. Store Processed Data

- Save in curated zone in Data Lake
- Format: Parquet (optimized for size & performance)

Transformation



Azure Data Lake Storage



Key Tasks - Orchestration

Goal: Automate ELT workflow for consistent, reliable data movement

Steps:

1. Create Pipelines in Azure Data Factory (ADF)
 - Separate pipelines for Flights, Events, Sports, COVID
2. Self-Hosted Integration Runtime (SHIR)
 - Connect on-premises PostgreSQL securely to ADF
3. Orchestration
 - Validate → Clean → Transform → Load automatically
 - Ensure referential integrity and performance



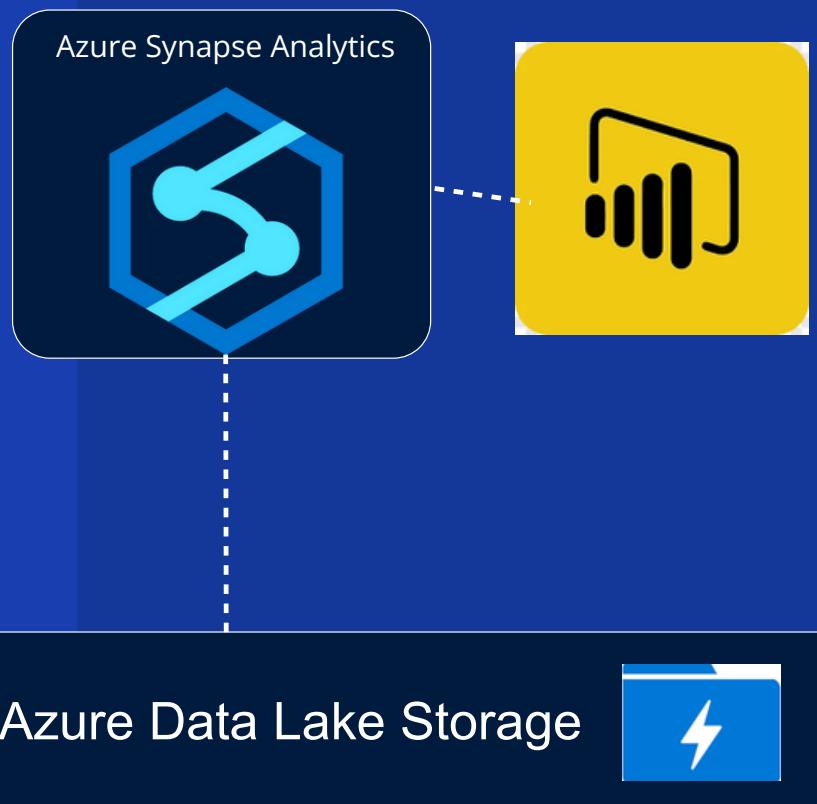
Key Tasks - Analysis & Visualization

Step 1 — Data Analysis in Synapse

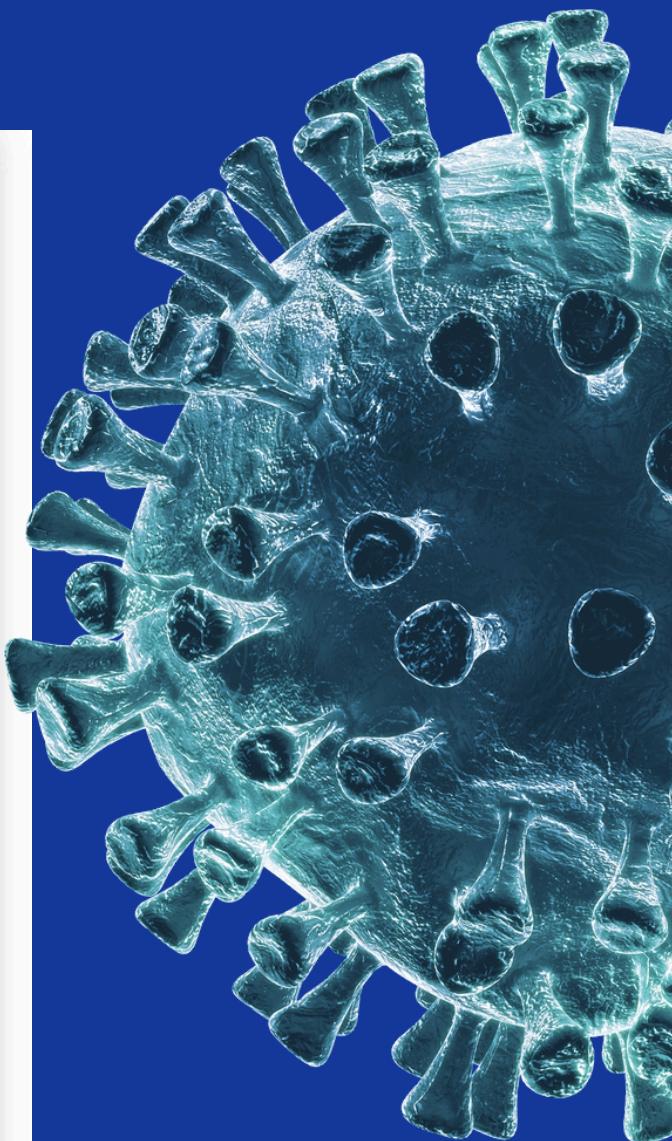
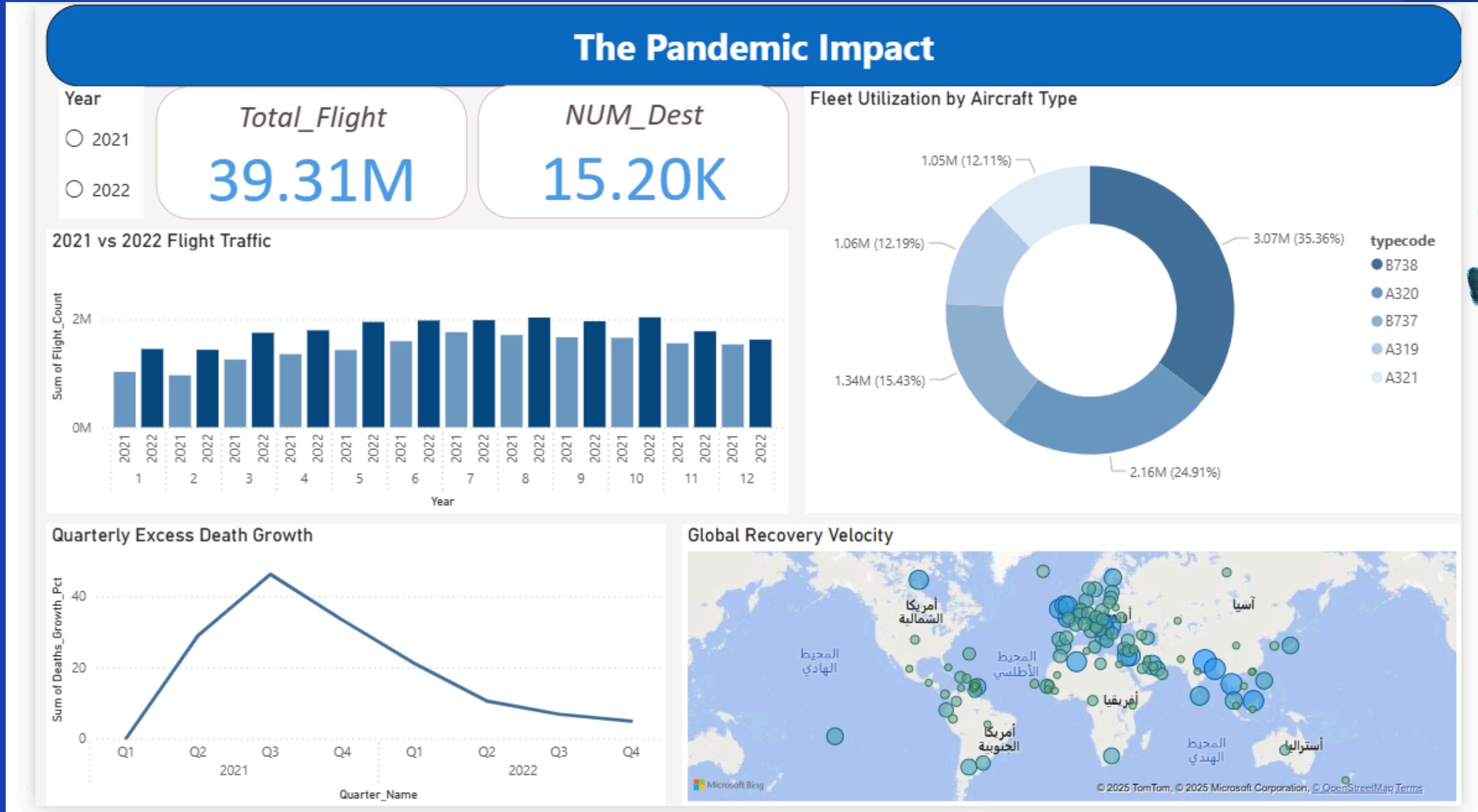
- Used Azure Synapse Analytics to query and analyze our processed datasets.
- Created External Tables on top of the curated Parquet files in Data Lake.
- Applied SQL queries to calculate metrics, aggregate data, and explore trends.

Step 2 — Data Visualization

- Visualized key insights: flight trends, event impact, and mobility patterns.
- Dashboards allow temporal and geographical analysis for strategic decision-making.



Key Tasks - Analysis & Visualization



Key Tasks - Analysis & Visualization

The World Cup Effect

Country_Name: Qatar

Attendance: 3M

AVG_Attention: 53.19K

MAX_Attention: 89K

Total_match: 64

Flights Volume (Q4 2021 vs 2022)

Detailed description: A bar chart comparing monthly flight volumes between Q4 2021 (dark blue) and Q4 2022 (light blue). The Y-axis represents the sum of total flights, ranging from 0K to 5K. The X-axis shows months 10, 11, and 12. In month 10, 2022 has slightly more flights than 2021. In month 11, 2022 has significantly more flights. In month 12, 2022 has more flights than 2021.

Home_Team	Away_Team	Day	Venue
Canada	Morocco	1	Al Thumama Stadium
Costa Rica	Germany	1	Al Bayt Stadium
Croatia	Belgium	1	Ahmed bin Ali Stadium
Japan	Spain	1	Khalifa International Stadium
Cameroon	Brazil	2	Lusail Iconic Stadium
Ghana	Uruguay	2	Al Janoub Stadium
Korea Republic	Portugal	2	Education City Stadium
Serbia	Switzerland	2	Stadium 974
Argentina	Australia	3	Ahmed bin Ali Stadium
Netherlands	USA	3	Khalifa International Stadium
England	Senegal	4	Al Bayt Stadium
France	Poland	4	Al Thumama Stadium (Neutral Site)
Brazil	Korea Republic	5	Stadium 974
Japan	Croatia	5	Al Janoub Stadium
Morocco	Spain	6	Education City Stadium
Portugal	Switzerland	6	Lusail Iconic Stadium
Croatia	Brazil	9	Education City Stadium
Netherlands	Argentina	9	Lusail Iconic Stadium
England	France	10	Al Bayt Stadium
Morocco	Portugal	10	Al Thumama Stadium
Argentina	Croatia	13	Lusail Iconic Stadium
France	Morocco	14	Al Bayt Stadium

Daily Flight Trends During Tournament

Detailed description: A line chart showing the sum of total flights per day from Day 0 to 30. The Y-axis ranges from 200 to 500. The line shows a general upward trend with significant fluctuations, particularly around Day 15 and Day 25, where flight volumes drop sharply before recovering.



Key Tasks - Analysis & Visualization



Conclusion

AirSense shows that behind every flight and every movement in the sky, there's a story about the world: how it moves, reacts, and adapts to major events.

Data is not just numbers it's a mirror of the world, giving deeper insights and awareness of global changes.

The project highlights the power of understanding data as a source of vision, not just information, and opens the door to analyzing any phenomenon that impacts human mobility and society.





TOGETHER FOR TOMORROW!
Enabling People

©2020 SAMSUNG. All rights reserved.

Samsung Electronics Corporate Citizenship Office holds the copyright of book.

This book is a literary property protected by copyright law so reprint and reproduction without permission are prohibited.

To use this book other than the curriculum of Samsung innovation Campus or to use the entire or part of this book, you must receive written consent from copyright holder.