



AIRFLY INSIGHTS

DATA VISUALIZATION AND ANALYSIS OF AIRLINE & AIRPORT OPERATIONS

PREPARED BY: YUVADHARANIE M

TABLE OF CONTENTS

- 1. PROJECT OVERVIEW**
- 2. PROBLEM STATEMENT**
- 3. DATASET OVERVIEW**
- 4. METHODOLOGY & WORKFLOW**
- 5. EXPLORATORY DATA ANALYSIS (EDA)**
- 6. DATA ACQUISITION & UNDERSTANDING**
- 7. DATA CLEANING & FEATURE ENGINEERING**
- 8. UNIVARIATE & BIVARIATE ANALYSIS**
- 9. DELAY CAUSE ANALYSIS**
- 10. ROUTE & AIRPORT-LEVEL EXPLORATION**
- 11. CANCELLATION & SEASONAL TRENDS**
- 12. FINAL DASHBOARD**
- 13. CHALLENGES ENCOUNTERED**
- 14. CONCLUSION**

PROBLEM STATEMENT

THE PROJECT FOCUSES ON ANALYZING A LARGE AND DETAILED FLIGHT DATASET TO DISCOVER TRENDS IN DELAYS, DISRUPTIONS, AND AIRLINE PERFORMANCE.

THROUGH DATA VISUALIZATION AND STATISTICAL EXPLORATION, IT SEEKS TO PROVIDE MEANINGFUL INSIGHTS THAT SUPPORT BETTER DECISION-MAKING FOR AIRLINES AND AIRPORTS.

PROJECT OVERVIEW

AIRFLY INSIGHTS AIMS TO STUDY LARGE-SCALE U.S. AIRLINE DATA TO UNCOVER PATTERNS RELATED TO FLIGHT DELAYS, CANCELLATIONS, AND ROUTE PERFORMANCE.

USING ANALYTIC METHODS AND INTERACTIVE VISUAL DASHBOARDS, THE PROJECT HIGHLIGHTS OPERATIONAL INEFFICIENCIES AND HELPS IMPROVE OVERALL FLIGHT RELIABILITY AND SCHEDULING.

DATASET OVERVIEW

DATASET SIZE

- **ROWS: 484,549**
- **COLUMNS: 33 (TIMINGS, DELAYS, AIRLINES, AIRPORTS, STATUS)**
- **TYPE: MULTI-YEAR U.S. DOMESTIC FLIGHT DATASET**

WHY THIS DATASET?

- **REAL, LARGE, AND DIVERSE DATASET SUITABLE FOR DEEP ANALYSIS**
- **RICH DETAILS ABOUT DELAY CAUSES AND CANCELLATIONS**
- **COVERS ROUTES, TIMINGS, AIRLINES, SEASONS, AND AIRPORTS**
- **SUPPORTS MULTI-DIMENSIONAL INSIGHTS AND PERFORMANCE COMPARISON**

METHODOLOGY & WORKFLOW

THE WORKFLOW FOLLOWED A SYSTEMATIC ANALYTICAL PROCESS:

1. DATA LOADING & EXPLORATION – STUDIED STRUCTURE, DATA TYPES, AND MISSING VALUES.
2. CLEANING & PREPROCESSING – HANDLED NULLS, STANDARDIZED DATE/TIME FIELDS, REMOVED DUPLICATES.
3. FEATURE ENGINEERING – CREATED MONTH, DAYNUMBER, HOUR, ROUTE, AND ONTIME COLUMNS.
4. EXPLORATORY DATA ANALYSIS – USED CHARTS AND GRAPHS TO DETECT PATTERNS AND TRENDS.
5. DASHBOARD DEVELOPMENT – BUILT INSIGHTS INTO POWER BI FOR INTERACTIVE VISUALIZATION.
6. FINAL DOCUMENTATION – COMPILED FINDINGS INTO REPORTS AND PRESENTATIONS.

EXPLORATORY DATA ANALYSIS

EDA HELPED UNDERSTAND THE DATASET'S STRUCTURE AND KEY CHARACTERISTICS.

IMPORTANT OBSERVATIONS:

- DATA SPANS SEVERAL YEARS OF DOMESTIC U.S. FLIGHTS.
- DETAILED DELAY CATEGORIES ENABLE CAUSE-BASED ANALYSIS.
- TIME-RELATED FEATURES HELP IDENTIFY HOURLY, MONTHLY, AND SEASONAL TRENDS.
- ROUTE AND AIRPORT DATA REVEAL GEOGRAPHICAL PERFORMANCE.
- CANCELLATION AND DIVERSION FIELDS HIGHLIGHT DISRUPTION PATTERNS.

📌 CLEANED Dataset Shape: (484551, 33)
📌 FINAL Dataset Shape: (484549, 33)

📌 Null values in FINAL dataset:
0 total nulls (should be 0)

📌 Columns in FINAL dataset: 33
Extra derived columns: set()

📌 Sample 3 rows from FINAL dataset:

| | DayOfWeek | Date | DepTime | ArrTime | ... | Month | DayNumber | Hour | Route |
|---|-----------|------------|---------|---------|-----|-------|-----------|------|---------|
| 0 | 4 | 2019-03-01 | 1829 | 1959 | ... | 3 | | 18 | IND-BWI |
| 1 | 4 | 2019-03-01 | 1937 | 2037 | ... | 3 | | 19 | IND-LAS |
| 2 | 4 | 2019-03-01 | 1644 | 1845 | ... | 3 | | 16 | IND-MCO |

[3 rows x 33 columns]

DATA ACQUISITION & UNDERSTANDING

SOURCE: KAGGLE (HIGH-QUALITY U.S. FLIGHT RECORDS)

KEY STEPS IN UNDERSTANDING THE DATA:

- DEFINED CLEAR PROJECT KPI'S AND OBJECTIVES**
- LOADED DATA USING PANDAS**
- VERIFIED SCHEMA, MISSINGS, AND VARIABLE TYPES**
- SAMPLED AND OPTIMIZED FOR MEMORY DUE TO LARGE FILE SIZE**
- CHOSE IMPORTANT COLUMNS SUITABLE FOR ANALYSIS AND FEATURE CREATION**

DATA CLEANING & FEATURE ENGINEERING

CLEANING ACTIVITIES:

- FILLED MISSING VALUES IN AIRPORT, DELAY, AND CANCELLATION FIELDS
- REMOVED DUPLICATES AND CORRECTED INCONSISTENT FORMATS
- STANDARDIZED TIME FIELDS TO USABLE DATETIME FORMATS

FEATURE ENGINEERING:

- ADDED MONTH, DAYNUMBER, DEPARTUREHOUR, ROUTE, AND ONTIME INDICATORS
- CREATED MEANINGFUL COMBINATIONS FOR DEEPER ANALYSIS
- STORED CLEANED DATASET TO SPEED UP FUTURE PROCESSING

```
# Load final dataset
df_final = pd.read_csv(final_path)

print("\n📌 CLEANED Dataset Shape:", df_cleaned.shape)
print("📌 FINAL Dataset Shape:", df_final.shape)

# Check nulls
print("\n📌 Null values in FINAL dataset:")
print(df_final.isnull().sum().sum(), "total nulls (should be 0)")

# Compare extra columns
print("\n📌 Columns in FINAL dataset:", len(df_final.columns))
print("Extra derived columns:", set(df_final.columns) - set(df_cleaned.columns))

# Quick sample
print("\n📌 Sample 3 rows from FINAL dataset:")
print(df_final.head(3))
```

```
# Month, Day of Week, Day Number
df["Month"] = df["Date"].dt.month
df["DayOfWeek"] = df["Date"].dt.dayofweek # Monday=0, Sunday=6
df["DayNumber"] = df["Date"].dt.day

# Convert DepTime (HHMM format) into Hour (HH)
df["Hour"] = (df["DepTime"] // 100).astype(int)

# Create Route as Origin-Dest
df["Route"] = df["Origin"] + "-" + df["Dest"]

# ✅ Double check
print("📌 Sample with new features:")
print(df[["Date", "Month", "DayOfWeek", "DayNumber", "Hour", "Origin", "Dest", "Route"]].head())
```

UNIVARIATE & BIVARIATE ANALYSIS

UNIVARIATE ANALYSIS:

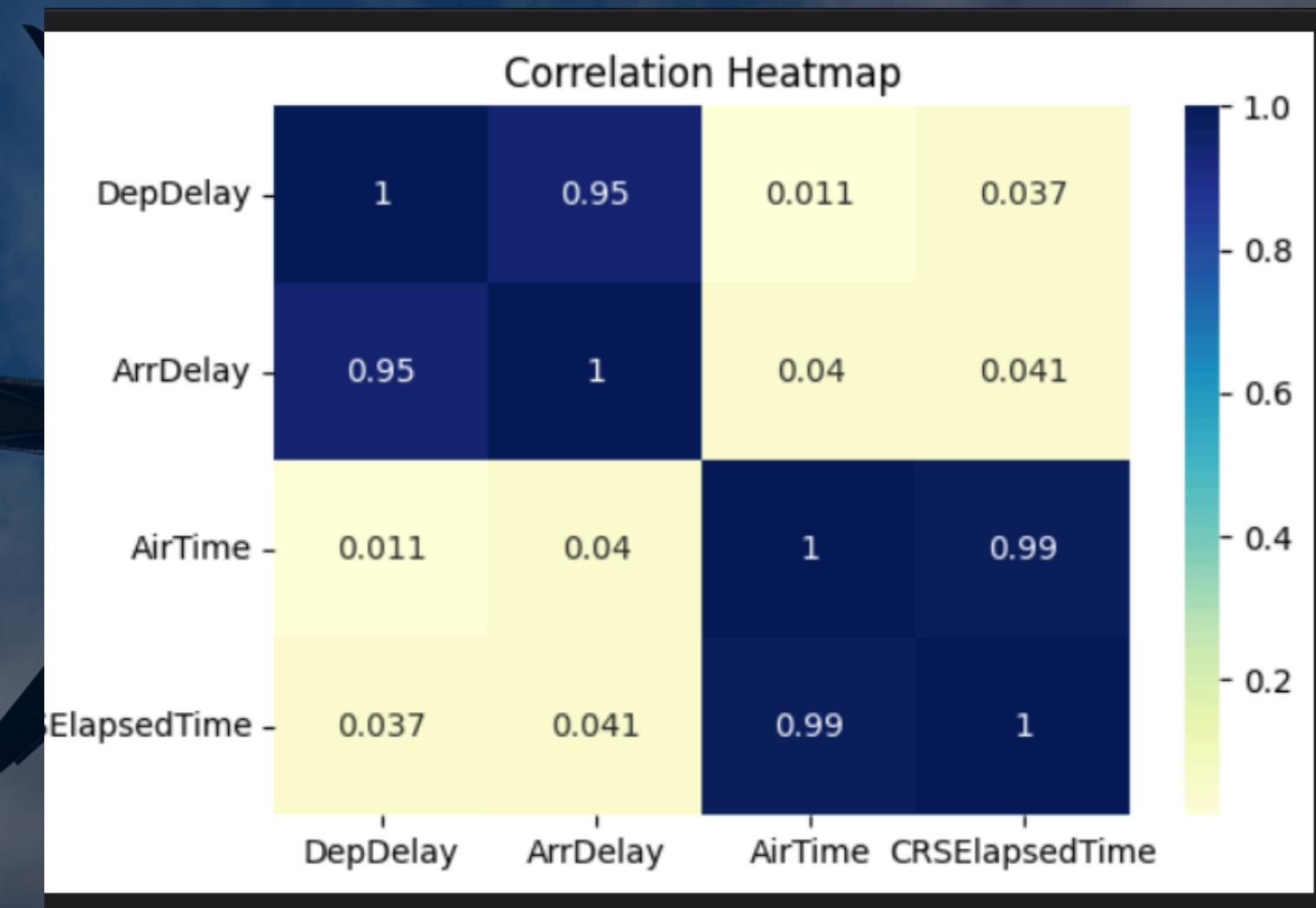
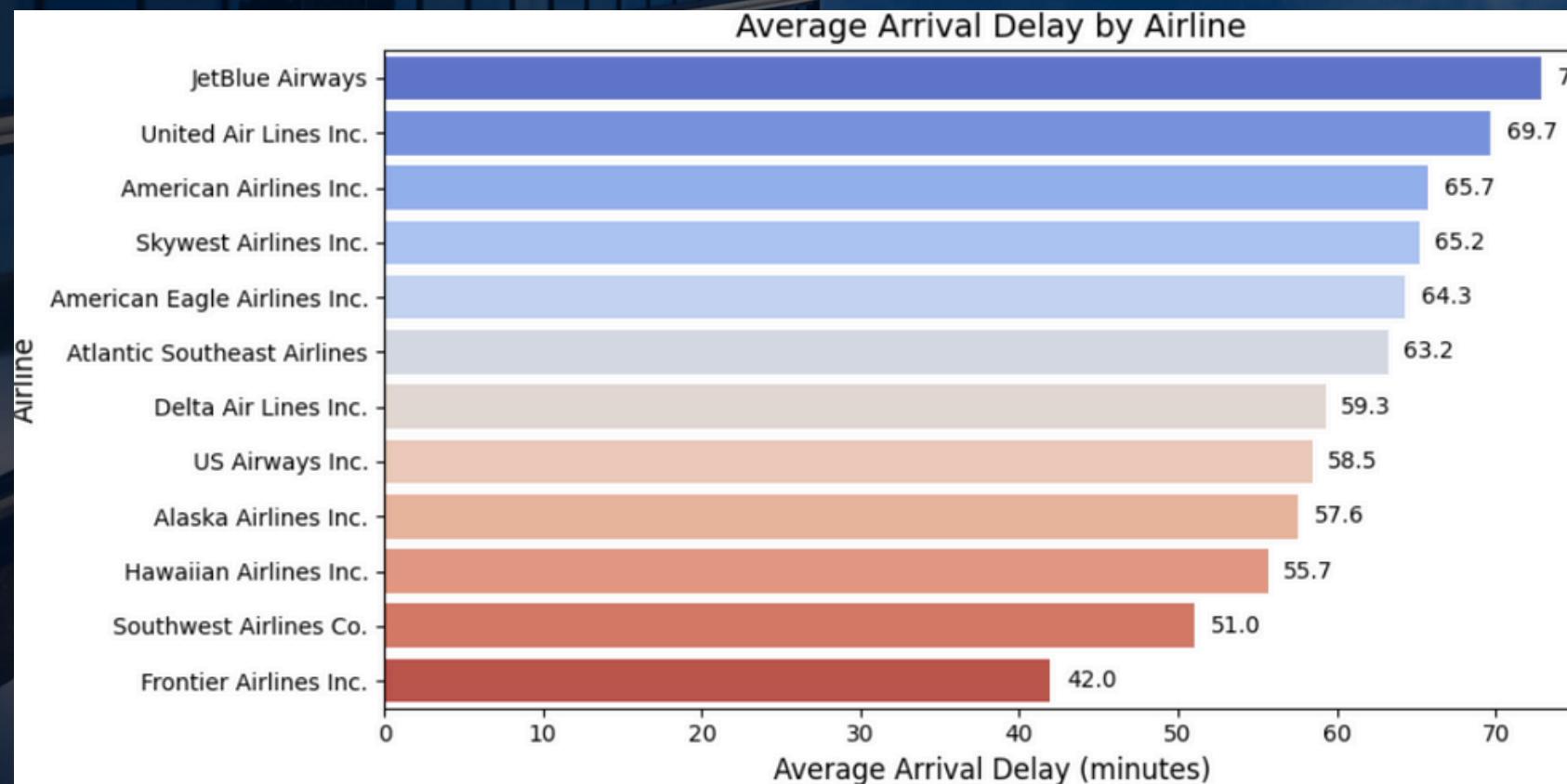
- STUDIED SINGLE COLUMNS SUCH AS AIRLINES, MONTHS, DELAY TYPES, HOURS
- UNDERSTOOD DATA DISTRIBUTION, PEAK TIMINGS, AND COMMON CATEGORIES

BIVARIATE ANALYSIS:

- EXAMINED RELATIONSHIPS LIKE AIRLINE VS DELAY, MONTH VS CANCELLATIONS, AIRPORT VS DELAY
- USED BAR CHARTS, SCATTER PLOTS, LINE GRAPHS, HISTOGRAMS, AND BOXPLOTS

OUTCOME:

- FOUND PEAK DELAY HOURS AND BUSY AIRLINES
- IDENTIFIED SEASONAL PATTERNS AND DISRUPTION SPIKES

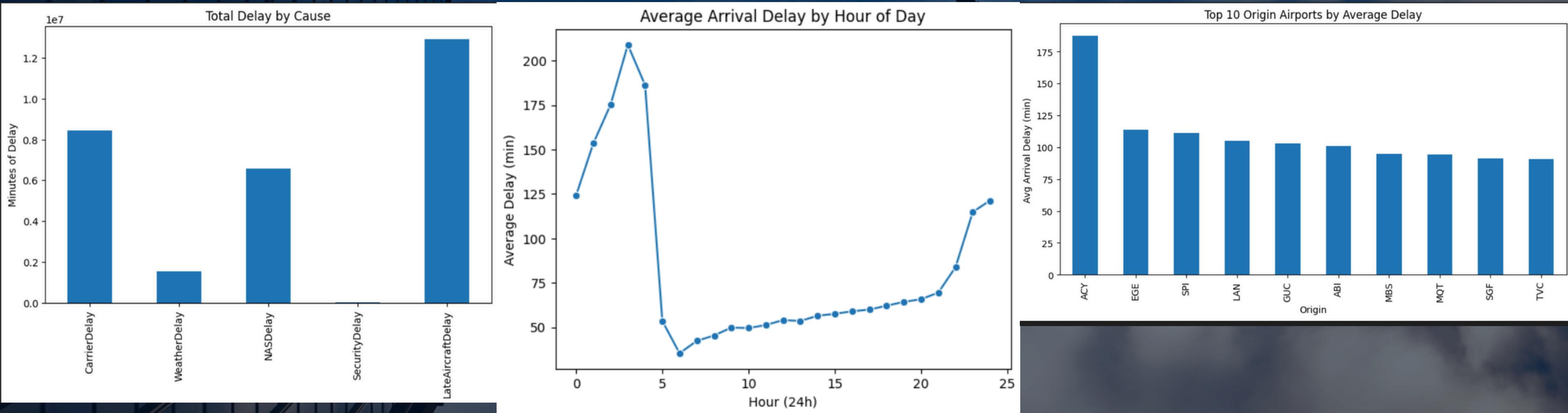


DELAY CAUSE ANALYSIS

ANALYSIS WAS PERFORMED ACROSS THE FIVE MAJOR DELAY CATEGORIES:
CARRIER, NAS, WEATHER, SECURITY, AND LATE AIRCRAFT.

KEY INSIGHTS:

- COMPARED DELAY SEVERITY ACROSS DIFFERENT AIRLINES
- MAPPED DELAY TRENDS BY AIRPORT AND DEPARTURE HOUR
- IDENTIFIED PROBLEMATIC ROUTES WITH CONSISTENTLY HIGH DELAYS
- HIGHLIGHTED TOP TIME PERIODS MOST PRONE TO DISRUPTIONS



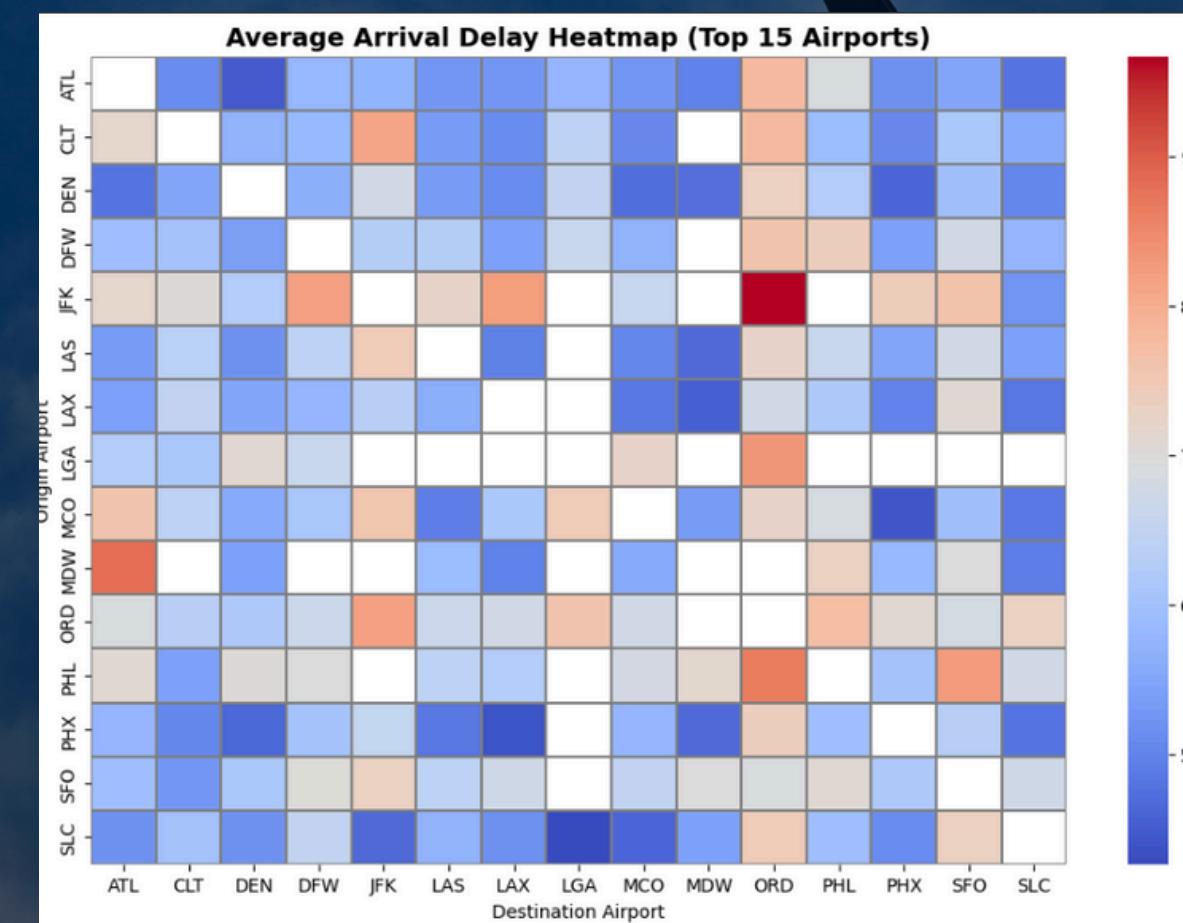
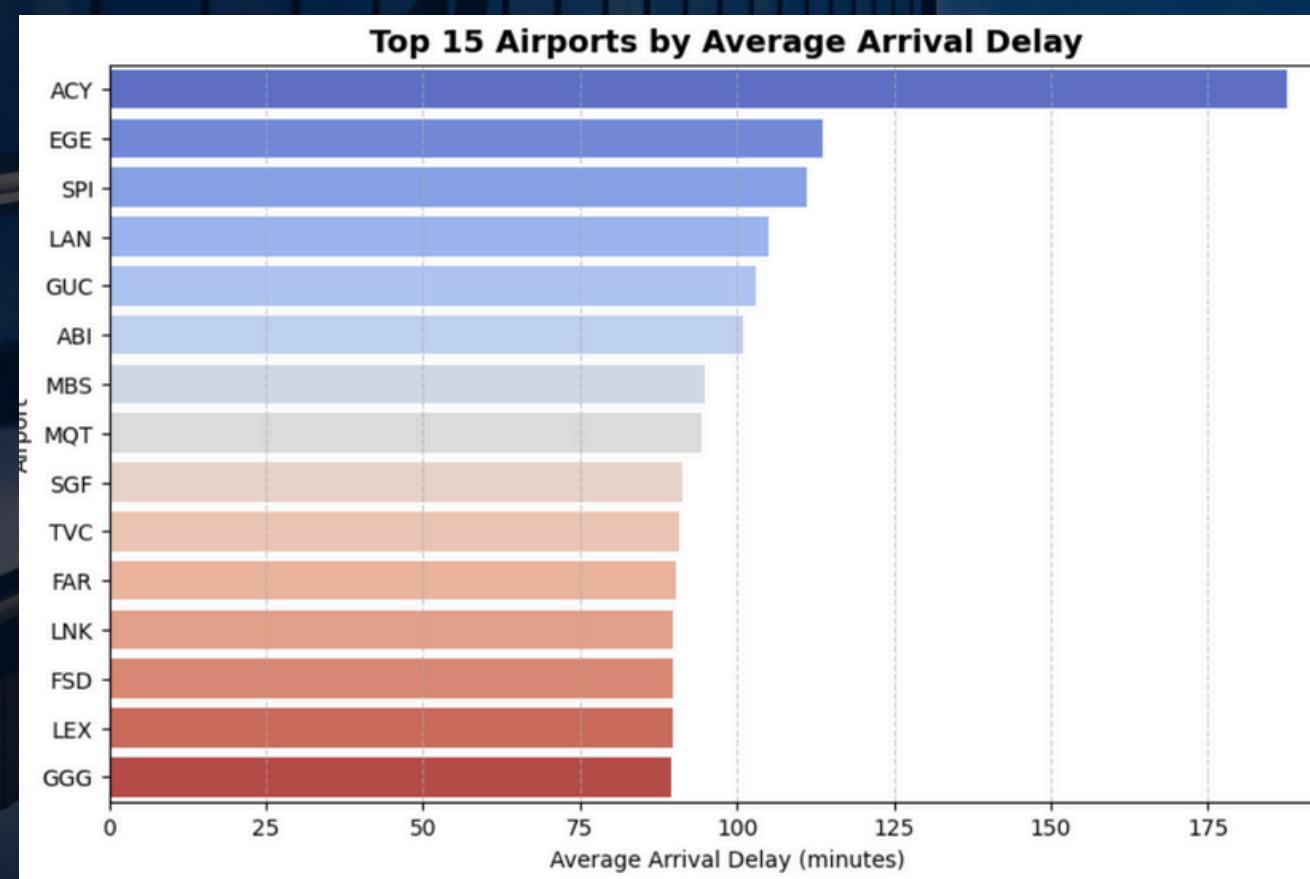
ROUTE & AIRPORT-LEVEL EXPLORATION

ROUTE ANALYSIS:

- RANKED THE BUSIEST FLIGHT ROUTES USING OD PAIRS
- IDENTIFIED HIGH PERFORMING AND LOW PERFORMING ROUTES
- EVALUATED HOW CONGESTION AND DISTANCE AFFECT DELAYS

AIRPORT-LEVEL STUDY:

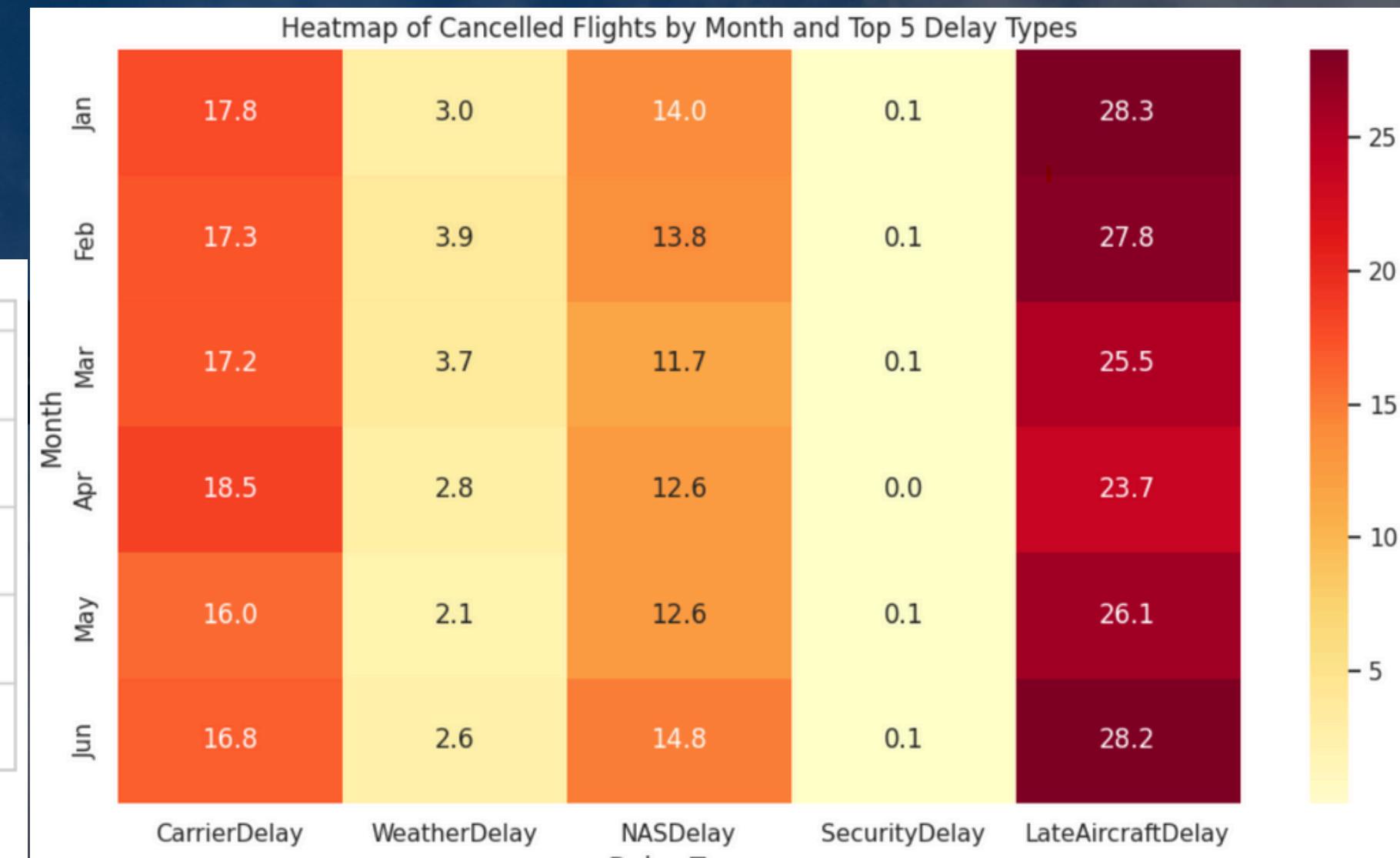
- CHECKED DELAY PATTERNS ACROSS MAJOR AIRPORTS
- IDENTIFIED AIRPORTS WITH FREQUENT CONGESTION OR WEATHER ISSUES
- COMPARED ORIGIN-DESTINATION PERFORMANCE FOR OPERATIONAL CLARITY



CANCELLATION & SEASONAL TRENDS

CANCELLATION STUDY:

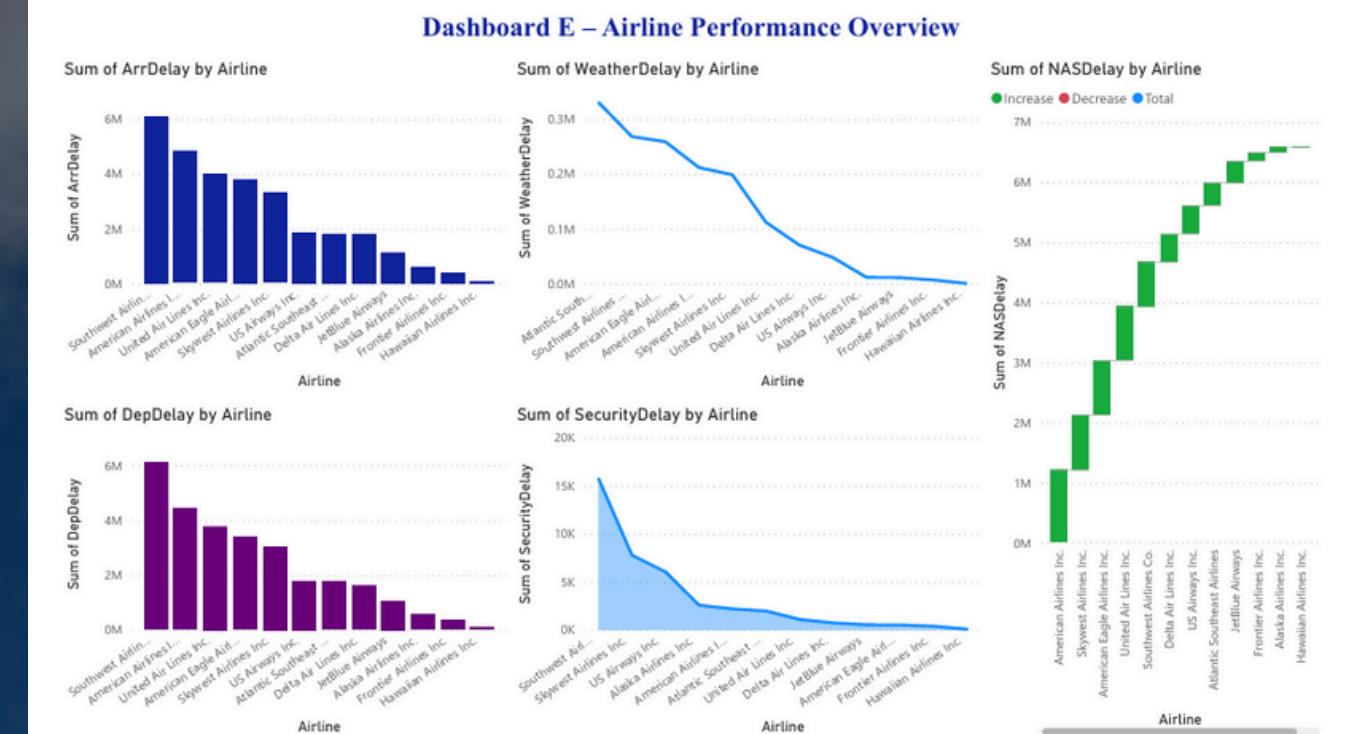
- ANALYZED MONTHLY AND SEASONAL CANCELLATION SPIKES
- CATEGORIZED CANCELLATIONS BY OFFICIAL REASON CODES
- COMPARED CANCELLATION RATES ACROSS AIRLINES AND ROUTES
- IDENTIFIED MONTHS WITH HIGHEST DISRUPTION DUE TO WEATHER OR NAS ISSUES



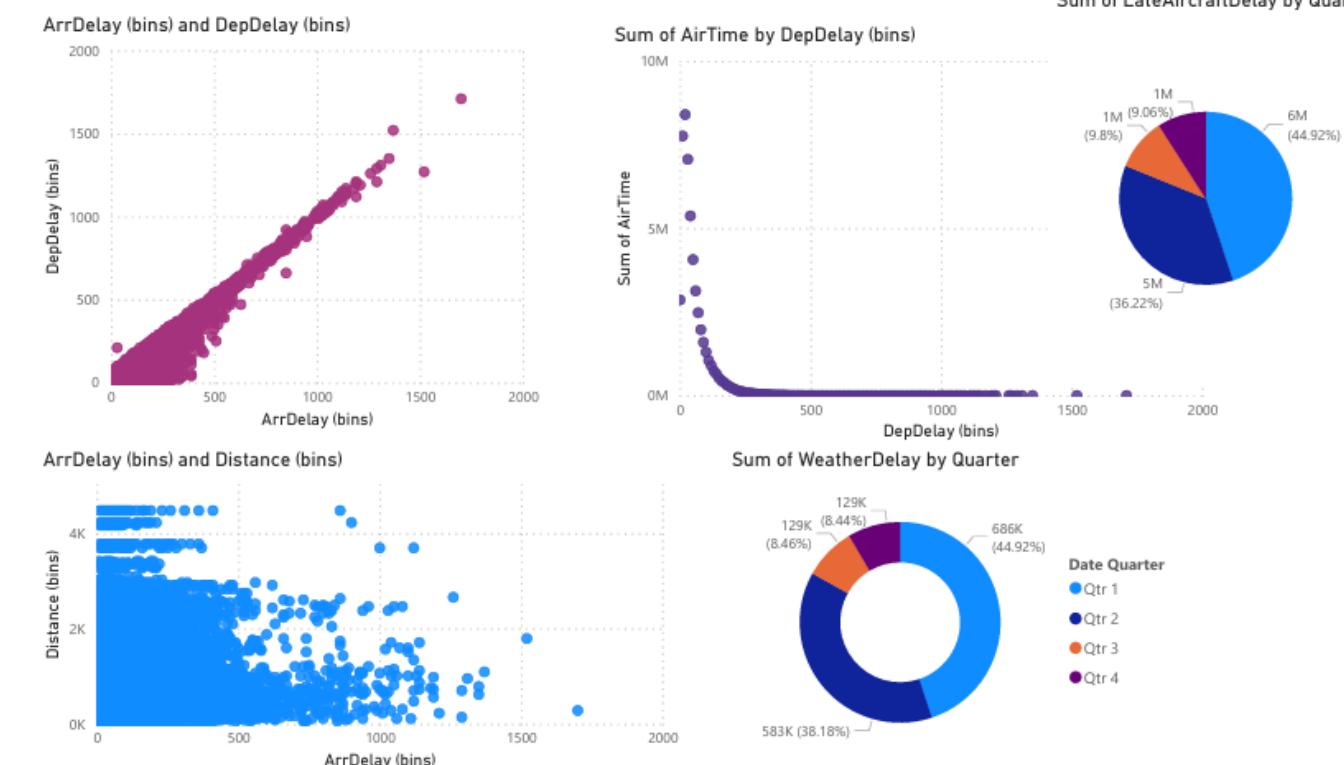
FINAL DASHBOARD

DASHBOARD FEATURES (POWER BI):

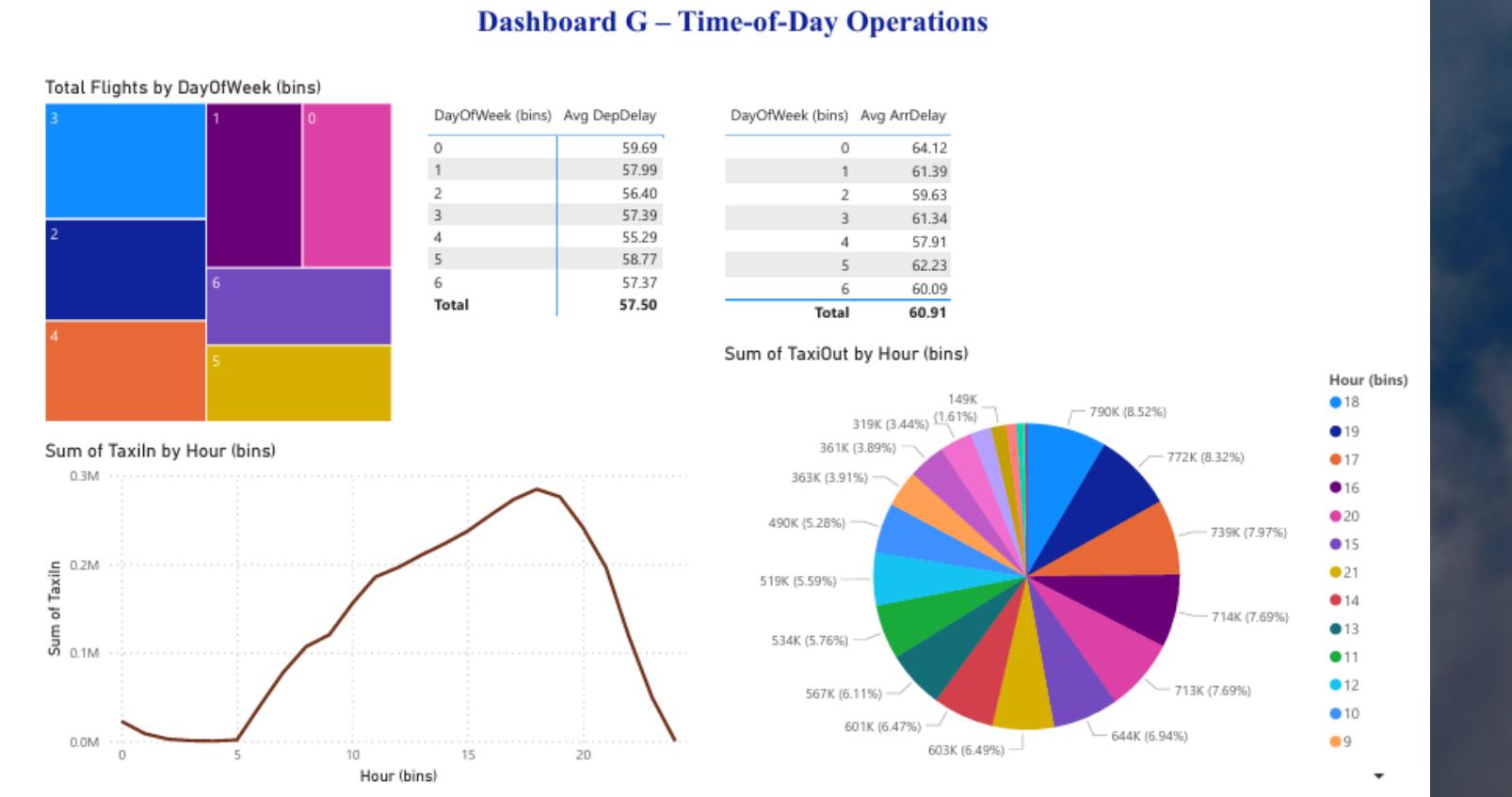
- IMPORTED CLEANED DATASET
- CREATED CALCULATED FIELDS SUCH AS ROUTE, DELAY TYPE, MONTH
- BUILT KPI INDICATORS: TOTAL FLIGHTS, DELAY RATE, CANCELLATION RATE
- VISUALS INCLUDE BAR CHARTS, LINE CHARTS, MAPS, DONUT CHARTS
- SLICERS FOR AIRLINE, AIRPORT, DELAY TYPE, AND MONTH
- FINAL DASHBOARD PUBLISHED FOR INTERACTIVE USER EXPLORATION



Dashboard J – Delay Correlations



Dashboard G – Time-of-Day Operations



CHALLENGES FACED

- HANDLING 480K+ ROWS REQUIRED MEMORY OPTIMIZATION
- MULTIPLE COLUMNS CONTAINED MISSING OR INCONSISTENT VALUES
- TIME FIELDS IN HHMM FORMAT WERE DIFFICULT TO CONVERT
- SIGNIFICANT DUPLICATES AND NOISY ENTRIES
- DELAY CATEGORIES REQUIRED CAREFUL MAPPING
- COMPLEX TRANSFORMATIONS FOR FEATURE ENGINEERING
- HEAVY VISUALIZATION LOAD SLOWED PERFORMANCE
- EXTRACTING MEANINGFUL INSIGHTS FROM NOISY TRENDS REQUIRED DEEPER ANALYSIS
- POWER BI DASHBOARDS NEEDED TUNING FOR SMOOTH PERFORMANCE

CONCLUSION

THIS PROJECT SUCCESSFULLY TRANSFORMED A LARGE RAW DATASET INTO ACTIONABLE INSIGHTS ABOUT DELAYS, CANCELLATIONS, AND AIRLINE PERFORMANCE. THROUGH THOROUGH CLEANING, VISUALIZATION, AND DASHBOARDING, IT REVEALED KEY OPERATIONAL PATTERNS THAT CAN HELP IMPROVE SCHEDULING EFFICIENCY AND REDUCE DISRUPTIONS ACROSS ROUTES AND AIRPORTS.