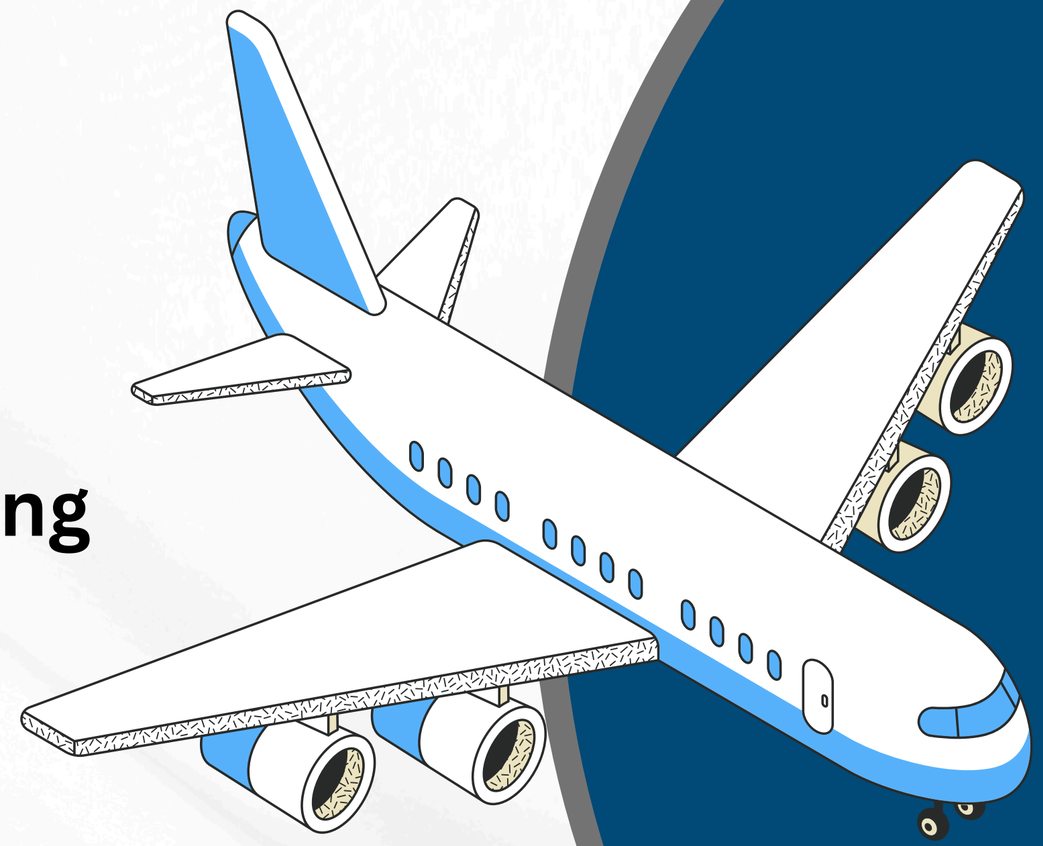
AIRLINE DATA ANALYSIS

Strategic Entry into the US Domestic Airline Market

Presented By: Patrick Ishimwe

AGENDA

1. Business Understanding
2. Data Understanding
3. Data Quality Check & Munging
4. Analysis & Insights
5. KPIs & Recommendations



BUSINESS UNDERSTANDING

The airline company intends to enter the US domestic airline market starting with five trip routes connecting medium and large airports. With the motto **"On time, for you"** the company focuses on punctuality, excellent service; and profitability.

Objectives

1

Identify busiest and most profitable round-trip routes in US domestic Airline market.

2

Recommend 5 round-trip routes to invest in based on various factors and calculate the number of flights required to break even.

3

Identifying Key Performance Indicators (KPIs) and provide recommendations

DATA UNDERSTANDING

1 Flights Dataset 

Each entry represents a distinct flight from the first quarter of 2019 and provides detailed information.

2 Tickets Dataset 

Each entry represents a distinct travel itinerary and provides detailed information, including fare and whether the ticket is for a one-way or round-trip flight.

3 Airport Codes 

Each entry represents a distinct airport, detailing attributes such as its type, name, location, and identification codes

Datasets	Observations	Columns
Flights	1, 915, 886	16
Tickets	1, 167, 285	12
Airport Code	55, 369	8

DATA QUALITY CHECK

Data Quality Issues	Resolutions
• Inconsistent Date Formats	Uniform date format conversion
• Incorrect Values	Mode imputation for incorrect values
• Missing Values	Mean/mode imputation for missing values
• Duplicate Values	Dropped duplicate values
• Incorrect Numeric Entries	Use Regular expressions for numeric values
• Outliers	Removed outliers using Interquartile Range (IQR) Method.

Purpose: Ensuring data quality and improved analysis accuracy, which significantly contribute to providing better-recommended routes for investment

DATA MUNGING & PREPARATION

Assumptions Made

- Distances values were recorded correctly and consistently for each route in most of the case; mode was used to fill missing and negative values.
- Negative values for delays indicate punctuality and were retained.
- Bidirectional routes (eg., ATL-LAX and LAX-ATL) are considered round trips.

DATA MUNGING & PREPARATION

Merging Datasets

To prepare dataset for analysis:

1. The Airport Codes dataset merged with the Flights dataset on ORIGIN and DESTINATION airport codes (IATA Code) using an INNER JOIN, filtering to include only flights between medium and large airports.

2. Then, the Tickets dataset was merged with the combined Flights dataset on a newly created TRIP_ROUTE (Origin with Destination) column, also using an INNER JOIN.

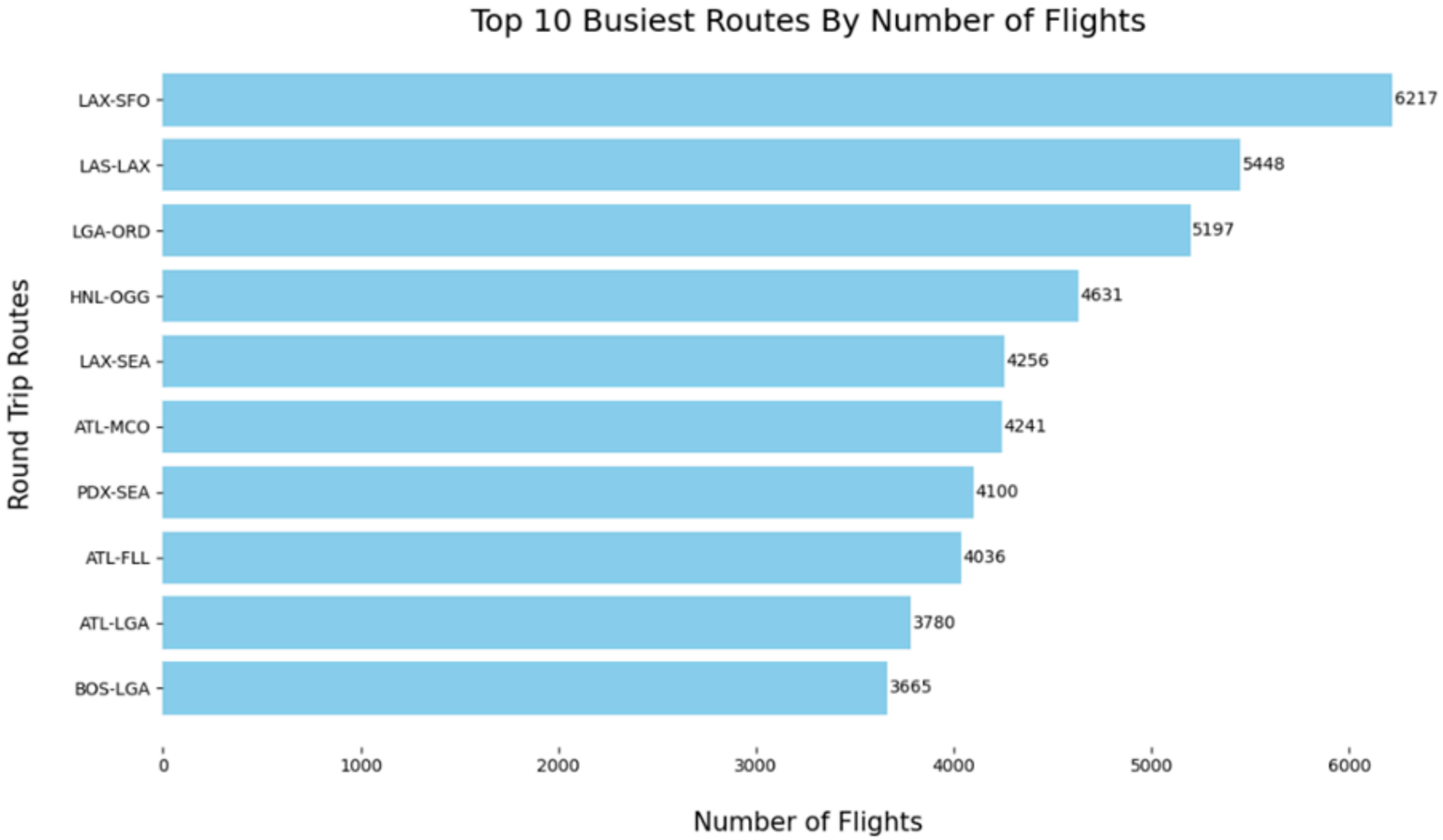
Final Dataset

3. Created new columns based on given columns and calculations. These created with purpose of responding to business questions.

- Average price of ticket
- Number of passengers
- Cost of operation
- Cost of delays
- Baggage fee
- Cost of maintenance and depreciation
- Trip route (Origin to Destination)

ANALYSIS & INSIGHTS

Top 10 Busiest Routes



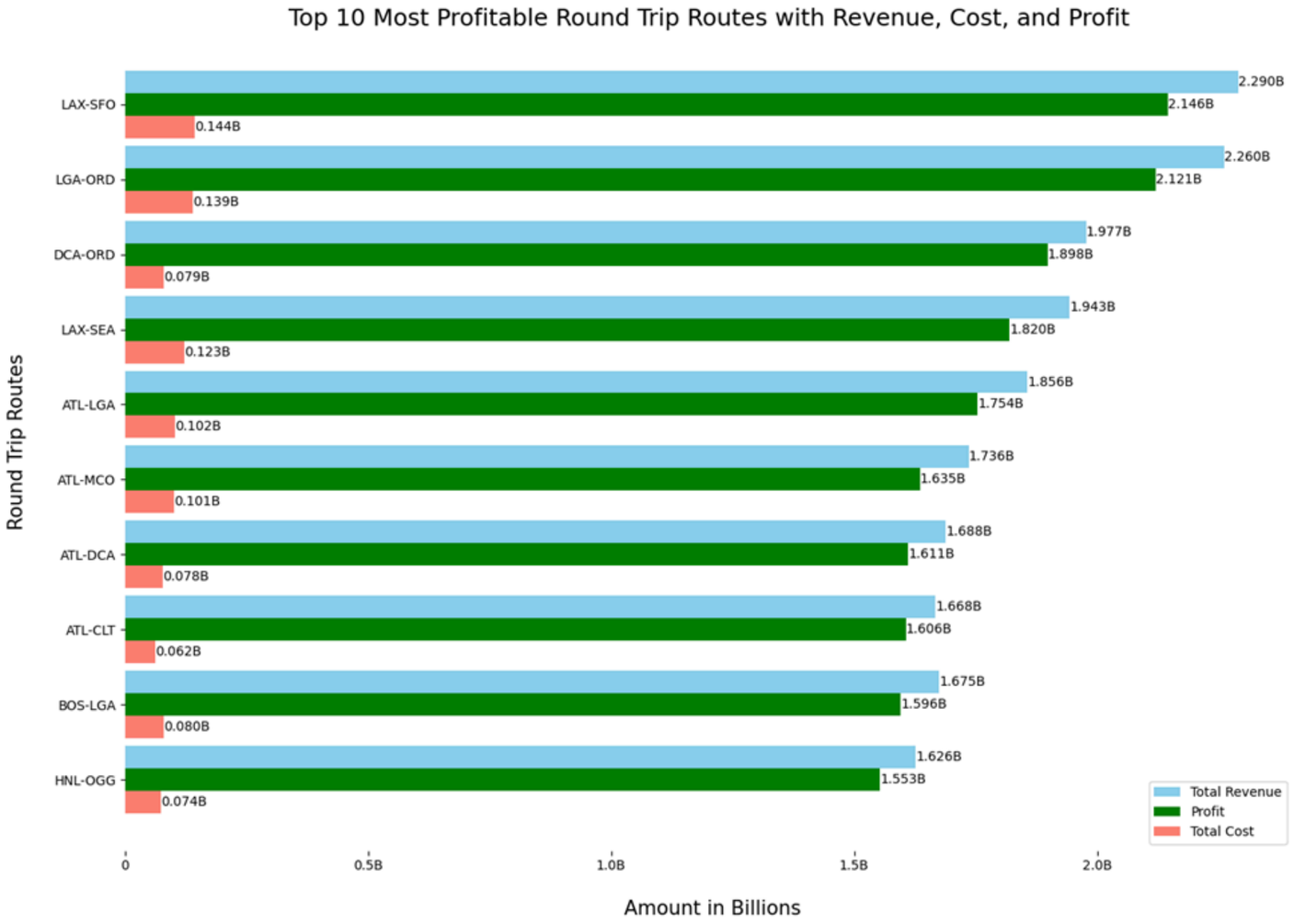
LAX-SFO: (Los Angeles to San Francisco) ranks first in demand, with a total of 6,217 round-trip flights.

The high demand is driven by factors such as :

- Economic Ties
- Tourism Attraction
- Major Hubs

ANALYSIS & INSIGHTS

Top 10 Most Profitable Routes



With $2.146 billion, Los Angeles (LAX) and San Francisco (SFO) stand out as the most profitable routes in the US domestic airline market.

It's likely driven by its high demand with over 6000 flights.

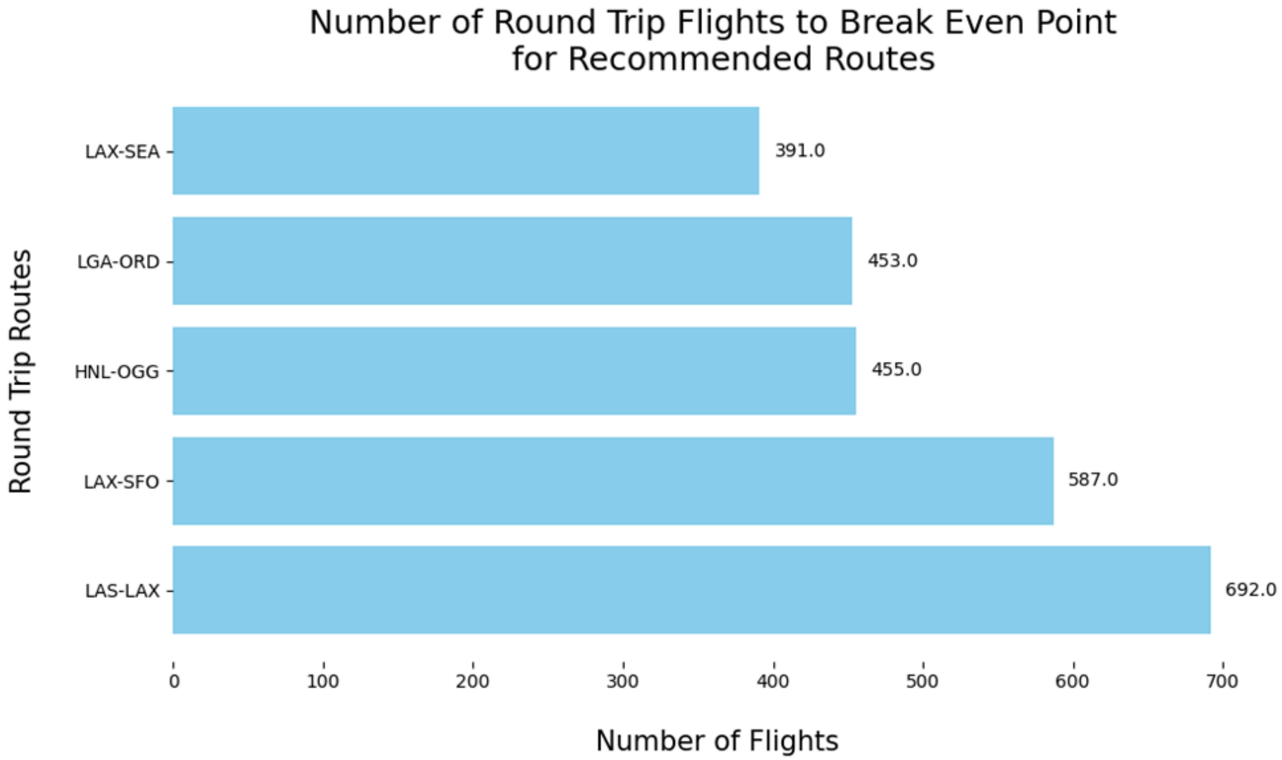
LAX-SFO: Los Angeles to San Francisco
LGA-ORD: New York LaGuardia to Chicago O'Hare
DCA-ORD: Washington National to Chicago O'Hare
LAX-SEA: Los Angeles to Seattle
ATL-LGA: Atlanta to New York LaGuardia
ATL-MCO: Atlanta to Orlando
ATL-DCA: Atlanta to Washington National
ATL-CLT: Atlanta to Charlotte
BOS-LGA: Boston to New York LaGuardia
HNL-OGG: Honolulu to Maui (Kahului)

ANALYSIS & INSIGHTS

1 5 Routes Recommended to Invest

- LAX-SEA (Los Angeles to Seattle)
- LGA-ORD (New York LaGuardia to Chicago O'Hare)
- HNL-OGG (Honolulu to Kahului)
- LAX-SFO (Los Angeles to San Francisco)
- LAS-LAX (Las Vegas to Los Angeles)

3 Flights to Break Even



2 Selection Metrics

The top 5 investment routes were selected using a weighted recommendation score based on ROI (60%), delay cost (20%), flights (10%), and passengers (10%). These metrics align with our focus on profitability, demand, utilization and punctuality.

Recommendation Score = ROI * 0.6 - Delay Cost * 0.2 + Passengers * 0.1 + flights * 0.1

RECOMMENDATIONS

Key Performance Indicators (KPIs)

- Profitability

  Assess profit after costs and average income per flight to understand earning potential

- Operational Efficiency

  Track flight delays and costs to ensure on-time performance and find ways to save money.

- Customer Satisfaction

  Monitor punctuality of flights and gather passenger feedback to improve satisfaction.

- Occupancy and Utilization

  Assess flight capacity to understand demand and optimize the use of available seats

Recommendations

- Perform Cost-Benefit Evaluation

  Evaluate each route's costs and benefits, considering factors like seasonal demand and pricing for better decisions.

- Integrate Predictive Analytics

  Use advanced analytics to predict future profitability and optimize flight schedules.

- Boost Customers Satisfaction

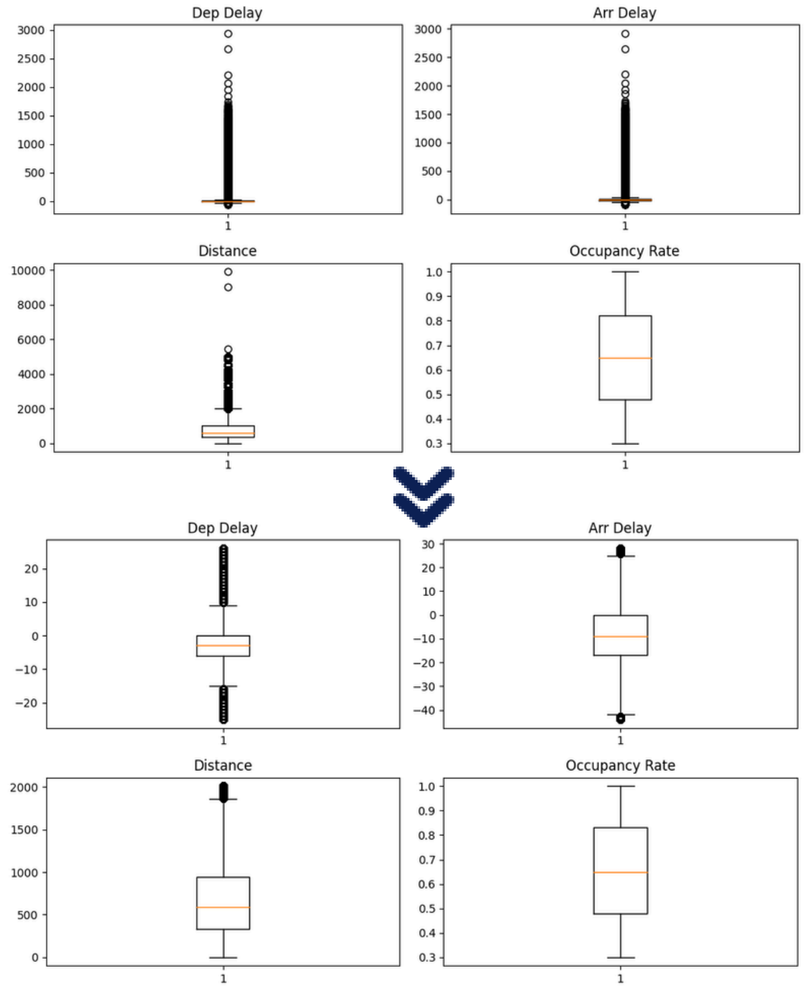
  Enhance the overall passenger experience by gathering feedback and improving service quality based on customer insights.

Thanks for Your Attention!

Happy to Take Questions & Feedback!

APPENDIX

Plot showing outlier with columns in Flight dataset



Plot showing outlier with columns in Ticket dataset