



Optimizing-Airline-Expansion- A-Data-Driven-Approach

Table of contents

1. Business Overview
2. Tools Used
3. Data Overview
4. Data Cleaning and Preparation
5. Data Munging
6. Data Processing
7. Data Quality Issues
8. Analysis
9. KPI's
10. Recommendations

01

Business Overview





About the Company

The airline company is planning to enter the United States domestic market. The company has decided to initiate operations with 5 round trip routes between medium and large US airports. The company's motto is "On time, for you," emphasizing the importance of punctuality in its brand image.

Objective

To determine the 5 round trip routes for the company entering the U.S. domestic market, focusing on busiest routes, maximizing profitability and achieving a quick breakeven. The goal is to analyze 2019 quarter 1 data to make data driven decisions in route selection and future performance monitoring.



Key Tasks



Identify the 10 busiest round trip routes.



Calculate breakeven points for recommended routes.



Identify the 10 most profitable round trip routes.



Suggest KPIs for future tracking.



Recommend 5 routes for investment.



02

Tools Used

Tools used for analysis:

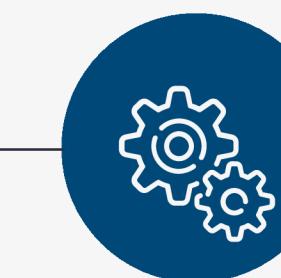
Software: Python-Google Collab

Libraries and
Packages Used
are:



Pandas

For data manipulation and analysis.



Numpy

For numerical operations.



Seaborn

For statistical data visualization.



matplotlib.pyplot

For interactive visualizations.



plotly.express

For interactive visualizations.



03

Data Overview



Data Sets Used:



Flights Dataset

Contains data about flights and available routes from origin to destination, including occupancy rates and flight details.



Tickets Dataset

Contains data about ticket prices and related information for round trip flights.



Airports Dataset

Contains data about whether an airport is considered medium or large sized.



04

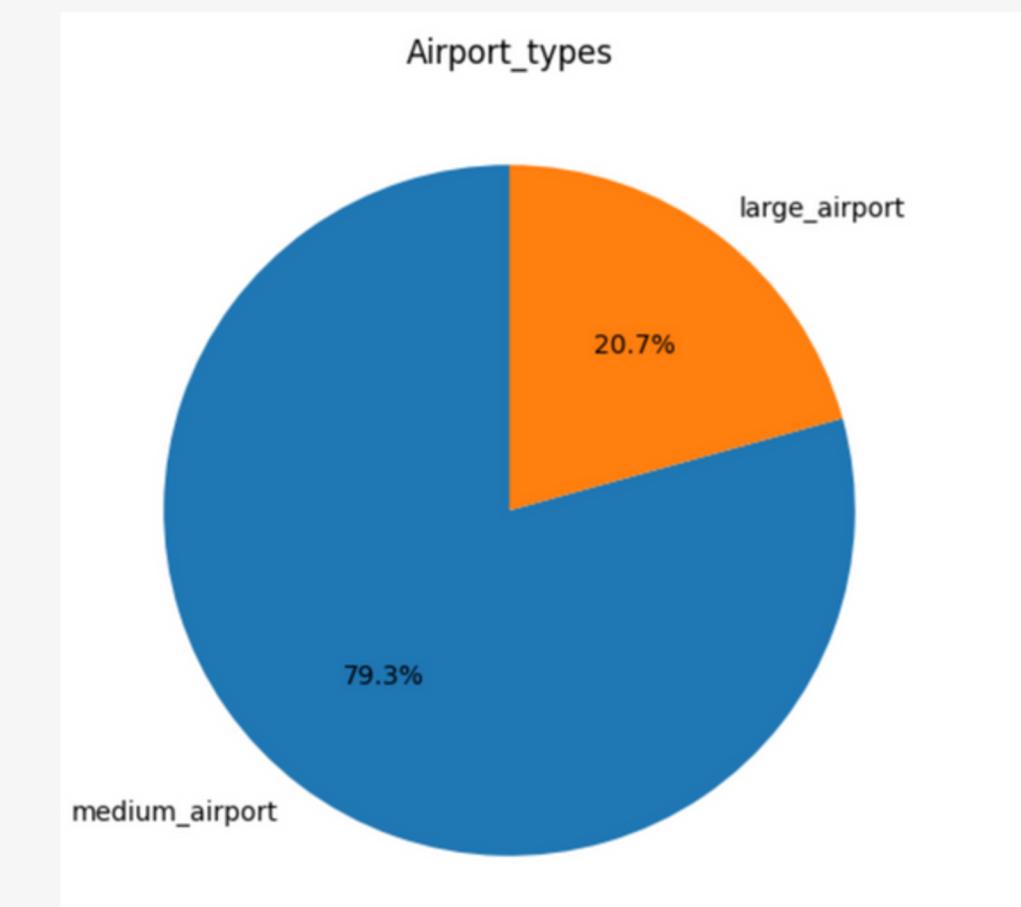
Data Cleaning And Preparation

Tickets Data set

- Filtered only round trip flights (ROUNDTRIP == 1.0).
- Dropped the rows with null values in ITIN_FARE and PASSENGERS columns.
- Ensured ITIN_FARE column is in numeric format for calculations.
- Created ITIN_FARE_PER_PASSENGER as the fare divided by the number of passengers.
- Aggregated data by ORIGIN and DESTINATION to compute the average fare per passenger.
- Dropped rows with null values in the ITIN_FARE_PER_PASSENGER column.
- Created Normalized_ROUTE to standardize route representation (e.g., "JFK-ORD").
- Applied outlier removal to clean the dataset further.

Airports Data set

- Filtered data to only airports in the United States (ISO_COUNTRY == "US").
- Filtered to medium and large airports (TYPE in ['medium_airport', 'large_airport']).
- Removed the ISO_COUNTRY column as it contains only "US".
- Dropped rows where IATA_CODE is null.
- Create a pie chart to visualize the distribution of medium vs. large airports.



Flights Data set

- Filtering only flights that were not cancelled (`CANCELLED == 0`).
- Creating `Normalized_ROUTE` to standardize route representation (e.g., "JFK-ORD").
- Removing columns that are not needed for analysis (`IATA_CODE_origin`, `IATA_CODE_destination`, `CANCELLED`, `ORIGIN_f`, `DESTINATION_f`).
- Dropping duplicate rows to ensure data uniqueness.
- Ensuring `DEP_DELAY` and `ARR_DELAY` are in absolute terms.
- Using existing routes data to fill missing values in the `DISTANCE` column.
- Converting `AIR_TIME` to numeric and fill missing values with the mean air time for each normalized route.
- Replacing missing values in `ARR_DELAY` with 0.
- Apply outlier function to remove the outliers.



05

Data Munging

Step1: Merging Flights with Airports:

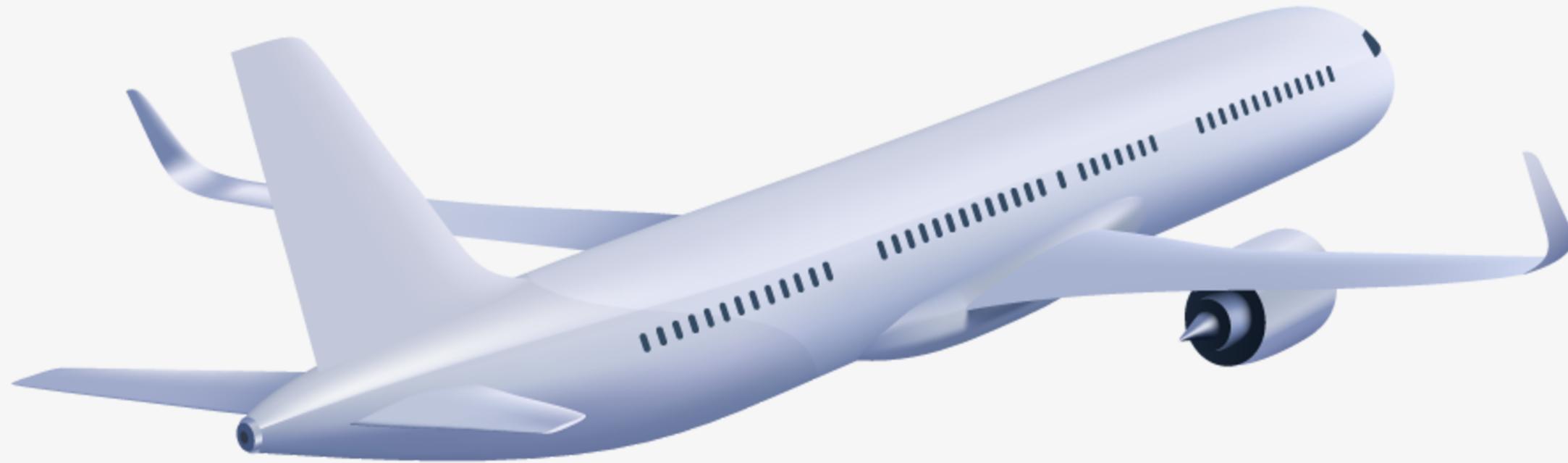
Merged the flights dataset with the airports dataset to include only medium and large origin airports, matching on 'ORIGIN' and 'IATA_CODE' using Inner join'.

Step2:Merging Flights with Destination Airports:

Merged the result from step 1 with the airports dataset again to include only medium and large destination airports, matching on 'DESTINATION' and 'IATA_CODE' using Inner join

Step3: Merging with Tickets Data:

Merged the resulting dataset with the tickets dataset to display ITIN FARE for each round trip, matching on 'Normalized_ROUTE' using Inner join



06

Data Processing

Meta Data

- ITIN_FARE_PER_PASSENGER: The itinerary fare per passenger, calculated by dividing the total itinerary fare by the number of passengers.
- Normalized_ROUTE: Combines the origin and destination of the trip, representing a round trip route.
- TOTAL_REVENUE: Total revenue from ticket fare and baggage fare, calculated using occupancy rate, total occupancy, and itinerary fare per passenger.
- TOTAL_COST: Total cost for one round trip, including fuel, oil, maintenance, crew, depreciation, insurance, other costs, airport operational costs, and delay charges.
- AIRPORT_CHARGES_origin: Airport charges at the origin, based on the type of airport.
- AIRPORT_CHARGES_destination: Airport charges at the destination, based on the type of airport.
- DELAY_CHARGE_origin: Delay charges for departure delays.
- DELAY_CHARGE_destination: Delay charges for arrival delays.
- TOTAL_PROFIT: Profit generated per round trip, calculated by subtracting total cost from total revenue.
- BREAK_EVEN: Number of flights required on a particular route to break even.

07

Data Quality Issues





Missing IATA Codes for some medium and large airports, leading to their removal. Potential changes in analysis outcomes and risk of drawing incorrect conclusions.



Columns with inconsistent data types (e.g., ITIN FARE with strings and numbers). Requires data type conversions, causing disruptions in the analysis process.



Negative values in arrival and departure delay columns. Can confuse analysis and potentially skew results, needing clarification or correction.

08

Analysis



Top 10 Busiest Round Trip

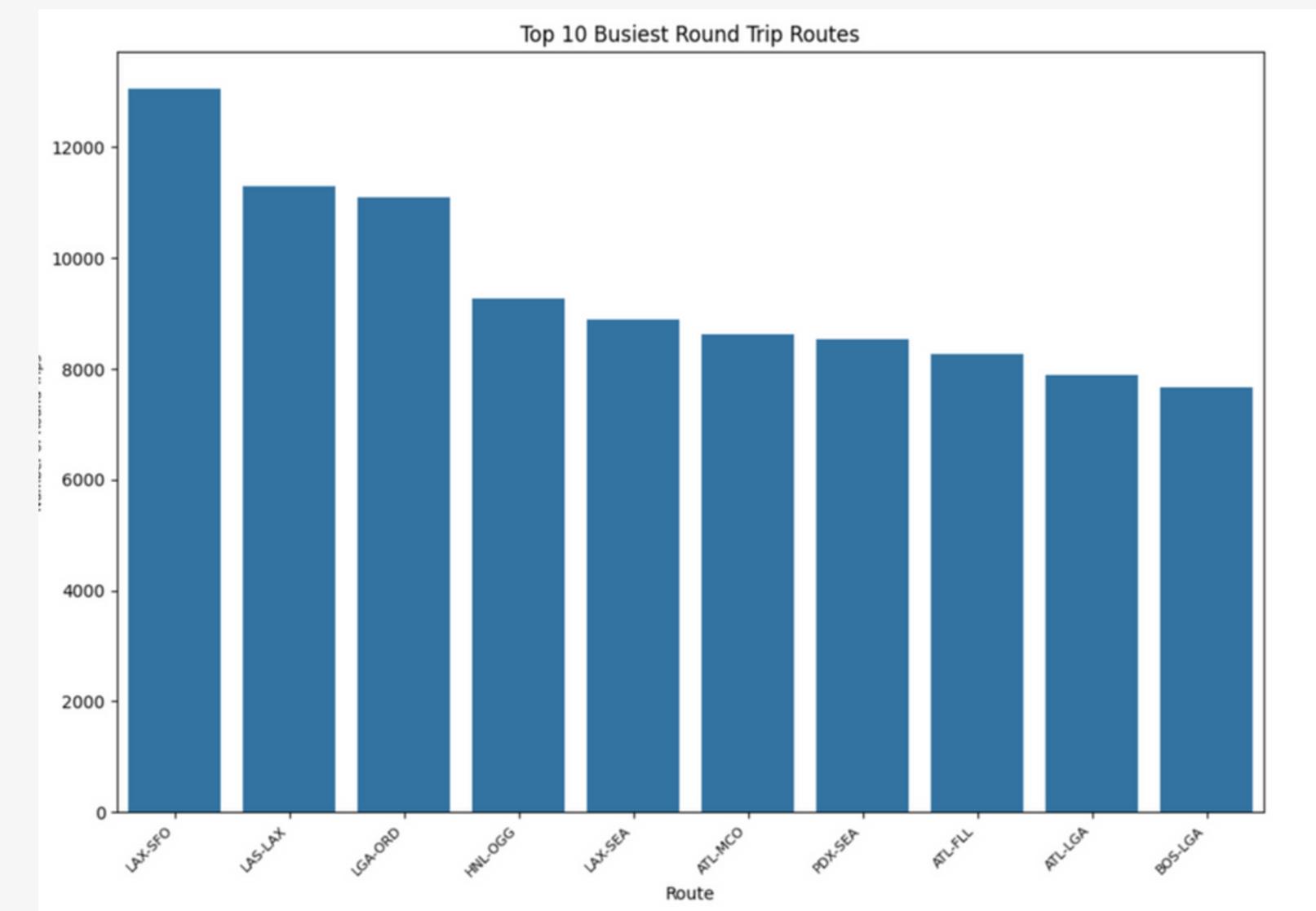


Methodology:

Count the number of flights per normalized route, sort by the number of flights, and identify the top 10 busiest round trip routes.

Interpretation:

The busiest route is between Los Angeles (LAX) and San Francisco (SFO), followed by Las Vegas (LAS) to Los Angeles (LAX). Other heavily traveled routes include New York (LGA) to Chicago (ORD) and Honolulu (HNL) to Los Angeles (LAX).



Top 10 Profitable Round Trip

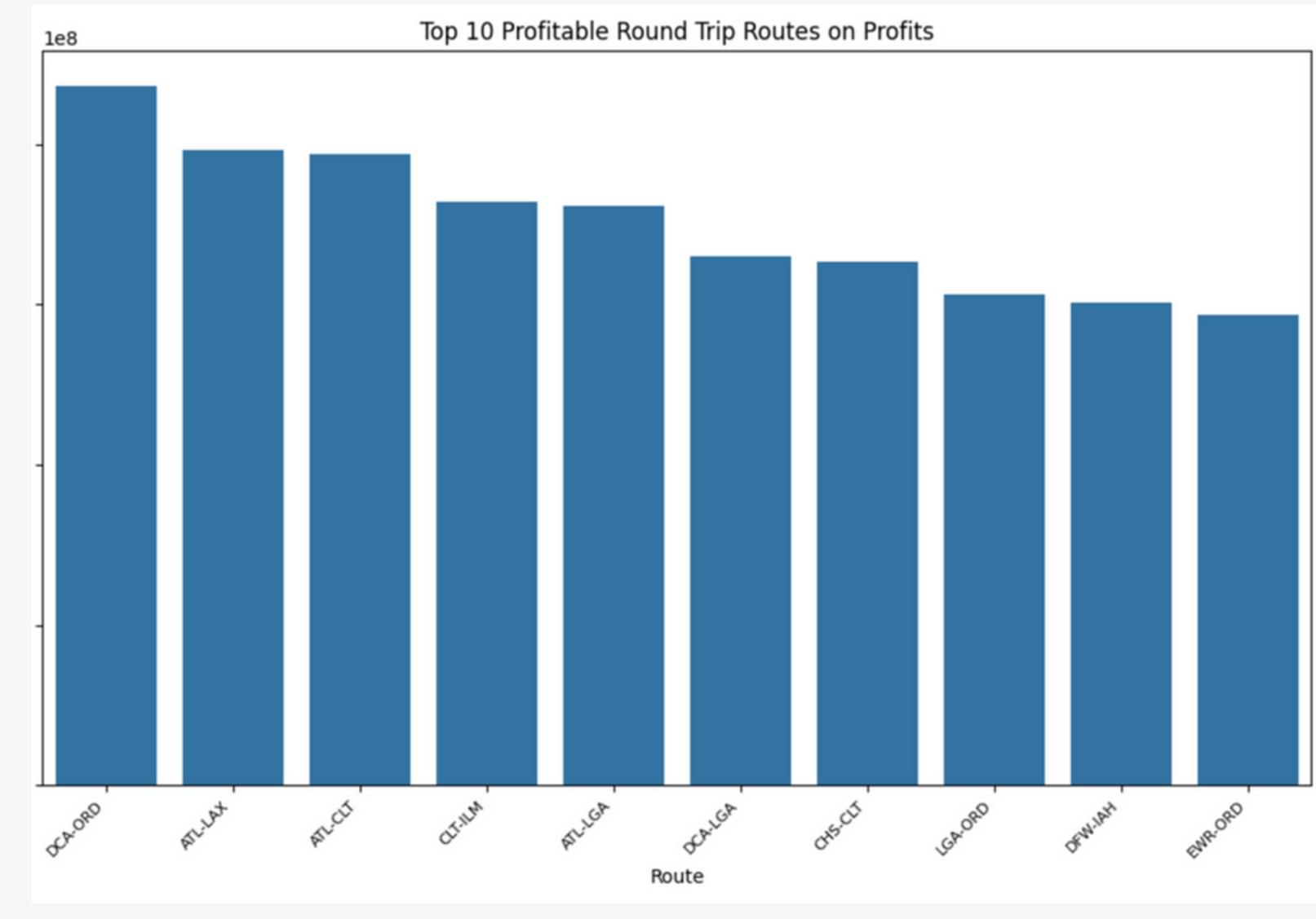


Methodology:

The data was aggregated by normalized route, calculating total revenue, total cost, total profit, occupancy rate, departure delay, and arrival delay, total profit(revenue-cost). The top 10 most profitable routes were identified by sorting the total profit in descending order, and a bar plot was created to visualize these routes.

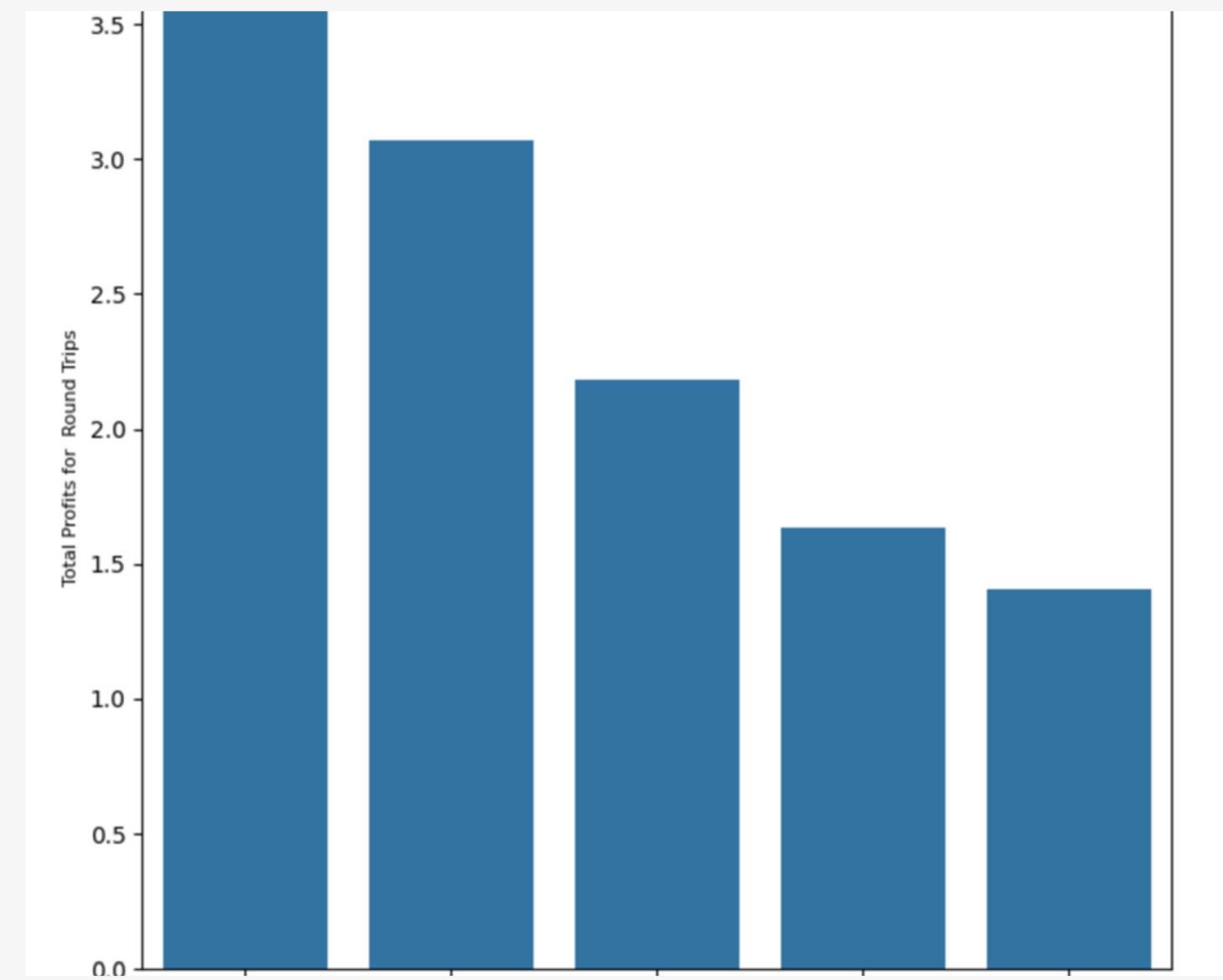
Interpretation:

The bar chart reveals that the route from Washington, D.C. (DCA) to Chicago (ORD) is the most profitable, followed by Atlanta (ATL) to Los Angeles (LAX). Other highly profitable routes include Atlanta (ATL) to Charlotte (CLT) and Cincinnati (CIT) to Wilmington (ILM)



5 Recommended Routes

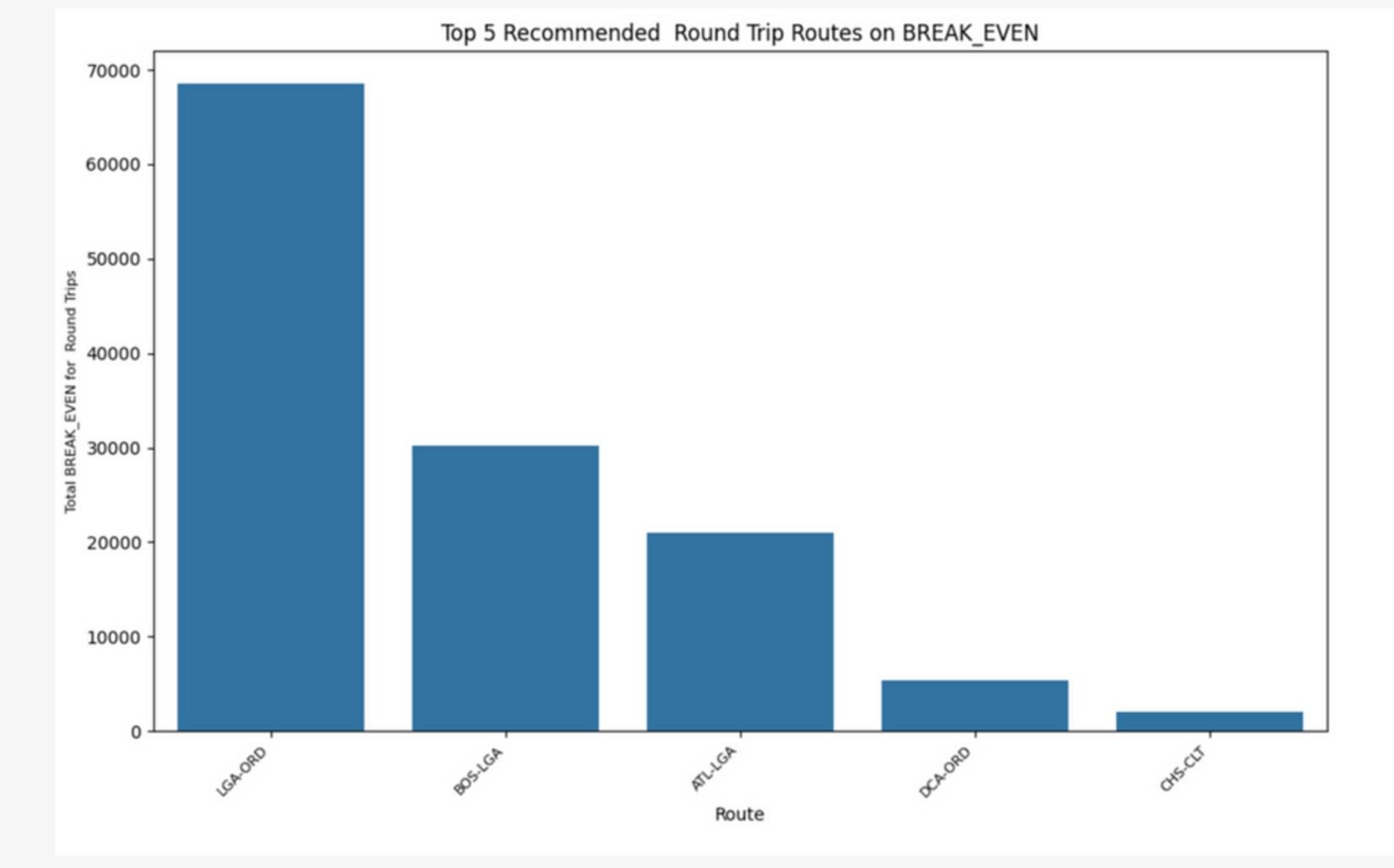
By considering factors like total profit, break-even points, and delays, the five recommended routes are Atlanta to New York LaGuardia (ATL-LGA), Washington D.C. to Chicago O'Hare (DCA-ORD), New York LaGuardia to Chicago O'Hare (LGA-ORD), Minneapolis to Chicago O'Hare (MSP-ORD), and Boston to New York LaGuardia (BOS-LGA). These routes not only take into account the financial perspective but also consider brand image by addressing delays and operational efficiency by considering break-even points.



Break-Even analysis



The graph shows the top 5 recommended round trip routes based on break-even points. The New York LaGuardia to Chicago O'Hare route has the highest break-even point, indicating it requires the most revenue to cover costs. Boston to New York LaGuardia and Atlanta to New York LaGuardia also have high break-even points, reflecting significant operational expenses. Washington D.C. to Chicago O'Hare and Charleston to Charlotte have lower break-even points, suggesting lower costs. This analysis highlights the financial thresholds needed for profitability on these routes.



Recommended Routes



01

Atlanta to New York
LaGuardia

02

Washington D.C. to
Chicago O'Hare

03

New York LaGuardia
to Chicago O'Hare
(LGA-ORD),

04

Minneapolis to
Chicago O'Ha

05

Boston to New York
LaGuardia

09

Key Performance Indicators





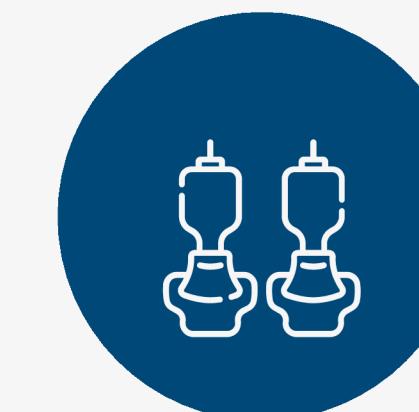
On-Time Departure and Arrival Rates

Focuses on punctuality to improve brand loyalty and customer satisfaction.



Profits

Measures financial performance and efficiency by tracking revenues and associated costs.



Break-Even Factor

Determines the minimum load factor needed for a no-loss, no-profit situation to ensure operational efficiency.



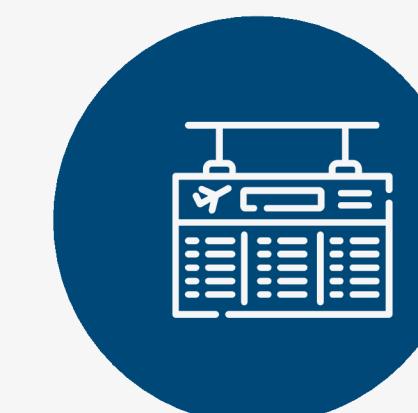
Occupancy Rate

Indicates operational efficiency and capacity maximization by measuring seat fill rates.



Customer Satisfaction Index

Assesses customer satisfaction to enhance repeat business and positive word of mouth.



10

Recommendations



- **Invest in Recommended Routes:** Profitable and efficient routes identified for optimal capacity utilization and operational efficiency.
- **Ensure Punctuality:** Focus on on-time departures and arrivals to enhance brand positioning and attract business travelers.
- **Promote Sustainable Travel:** Implement eco-friendly initiatives to appeal to environmentally conscious customers.
- **Gather Customer Feedback:** Use surveys and reviews to identify areas for improvement and gauge customer satisfaction.
- **Conduct Market Research:** Understand the US airline market, identify strengths and weaknesses, and strategically position the airline.



Questions?

