DATA 602 Project Presentation by James Ding

This is a presentation about Flight Distance & Airtime in 2019

Introduction

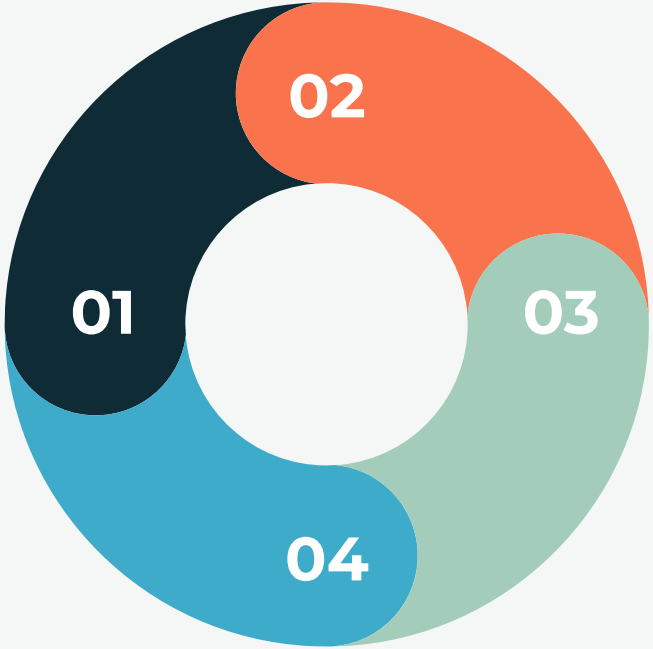
Objective:

Analyze & Visualize & Predict Flight Distance & Airtime

Motivation:

Save $ for airlines (Fuel Consumption)

Entertainment Preparation



Data Set:

Flight Delays & Cancellation Dataset (2019-2023)

How?

Descriptive Statistics

Inferential Statistics

Domain Introduction

Analyzing and visualizing Flight Distance & Air Time in the Aviation Industry: Why is it important?

- Customer Satisfaction & Loyalty
- Operational Efficiency
- Cost Management
- Safety and Compliance
- Strategic Planning and Scheduling
- Risk Management
- Competitive Advantage
- Visual Communication

Dataset

Flight Delay and Cancellation Dataset (2019-2023)

Source: Kaggle (origin: data was collected by the US Department of Transportation, Bureau of Transportation Statistics)

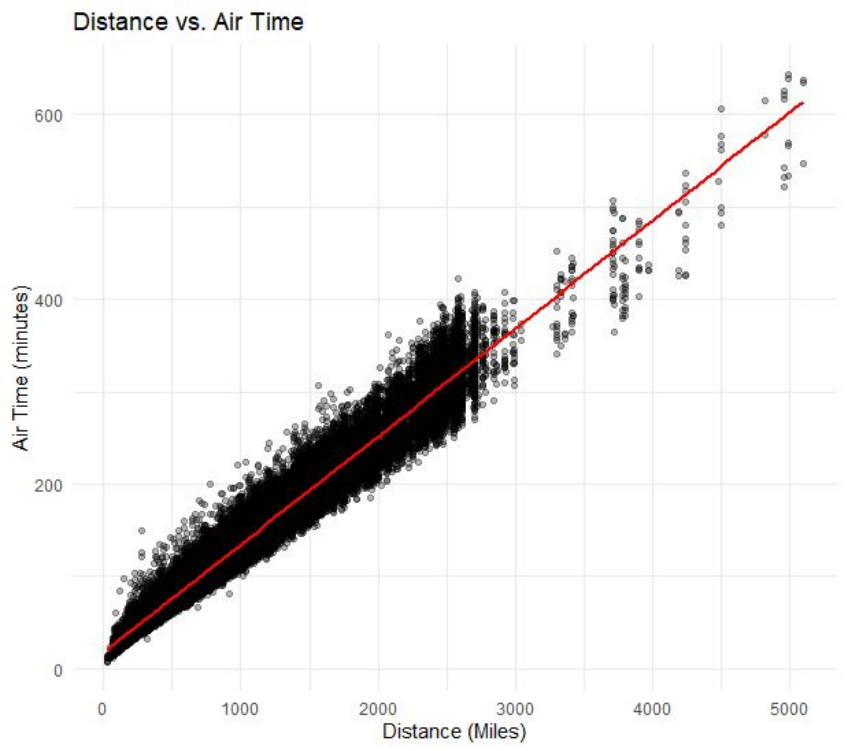
Format: csv

Original sample size = 1,048,576 flights

Guiding Questions

- Distance average? By month?
- Air time average? By month?
- Proportion average? (Velocity)
- Linear relationship between distance & airtime

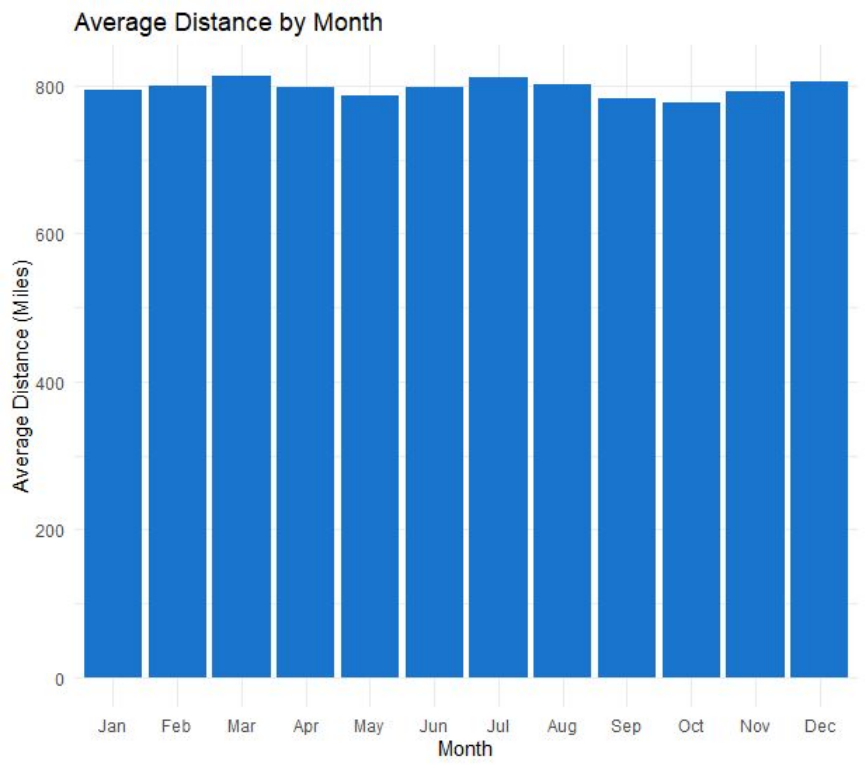
Visualization

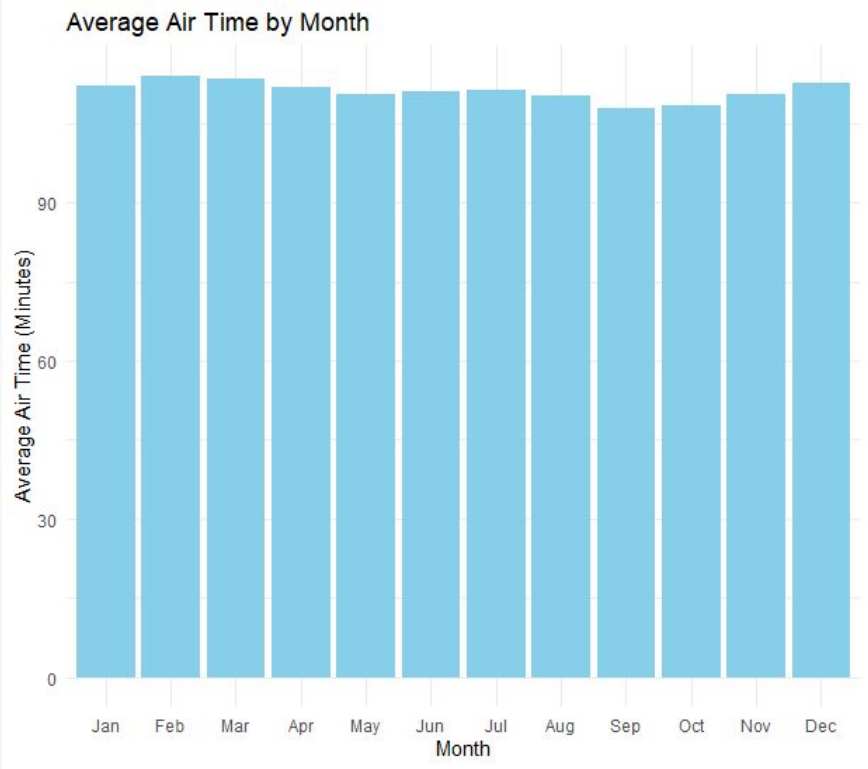


Descriptive Analysis (Average Distance/Air Time)

	Distance	Air time	velocity
n	100,000	97945	97945
mean	799.59	111.43	6.74
sd	590.79	70.35	1.2
median	640	93	6.85
max	5095	663	10.55
min	31	10	1.42
range	5064	653	9.14

Visualization P.2
(Avg Distance/Air Time by Month)





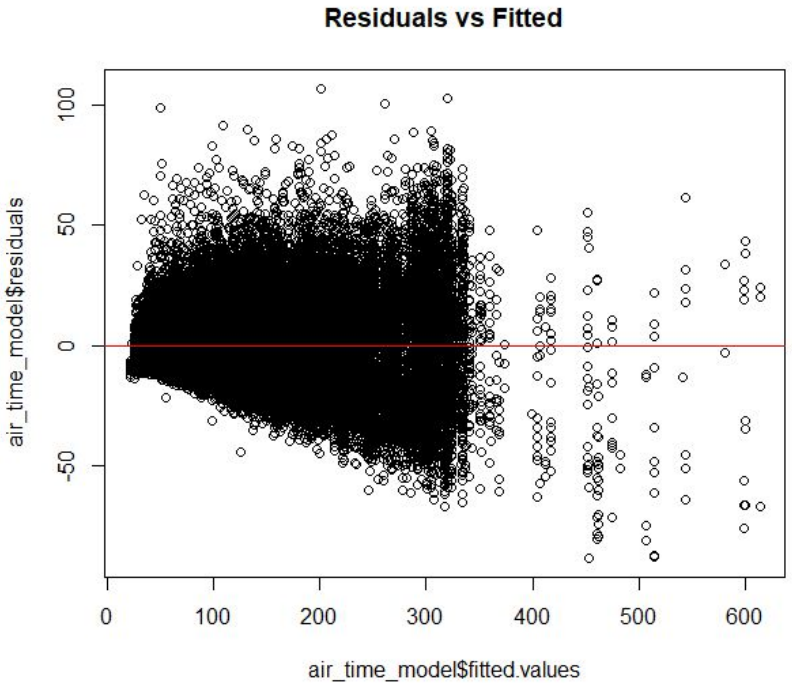
Inferential Analysis

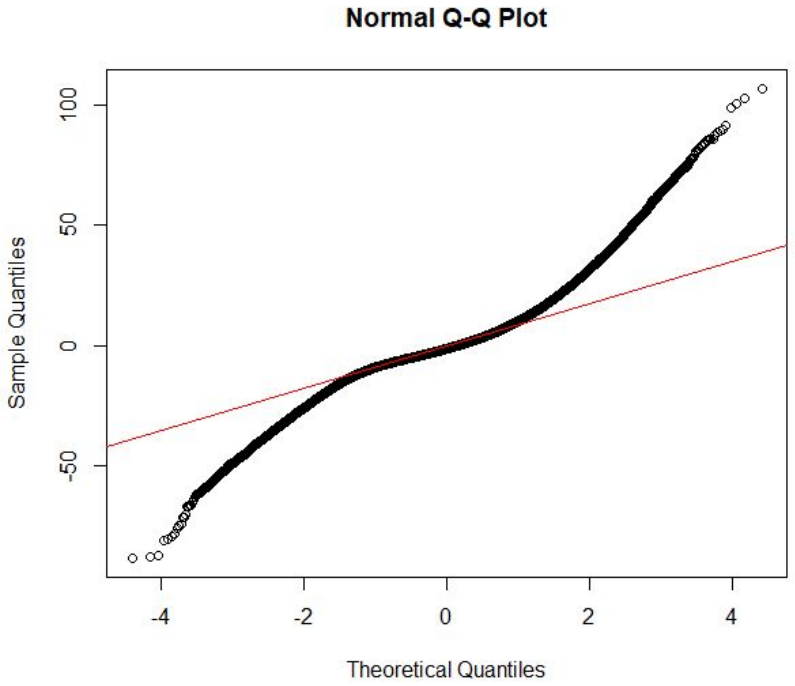
- Null Hypothesis (Ho)
  - There is no statistically significant relationship between Distance & Air Time
    - Ho: B=0
- Alternative Hypothesis (Ha)
  - As the flight distance increases, the air time increases
    - Ha: B>0

Inferential Analysis

```
Call:
lm(formula = AIR_TIME ~ DISTANCE, data = df_2019_sorted)

Residuals:
    Min      1Q  Median      3Q     Max
-88.362  -6.441  -1.394   5.398 106.571

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.769e+01  6.929e-02   255.3   <2e-16 ***
DISTANCE    1.170e-01  6.973e-05  1677.7   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.88 on 97950 degrees of freedom
  (2048 observations deleted due to missingness)
Multiple R-squared:  0.9664,  Adjusted R-squared:  0.9664
F-statistic: 2.815e+06 on 1 and 97950 DF,  p-value: < 2.2e-16
```

Inferential Analysis





Inferential Analysis

- Intercept: 17.69 minutes
  - When Distance = 0
- Every Distance (1 mile): 0.117 ~7 seconds
  - Change in minutes
- P_value : 2e-16 < 0.05
  - We reject the Null Hypothesis test
- RSE: 12.88
  - Observed Vs prediction
- R Square:96.64%
  - A very strong fit of the model
  - 96.64% of variability in air time can be explained by distance
- F-statistic: 2.815e+06
  - Extremely high, strong evidence against null hypothesis
- T-distribution: (Confint())
  - Intercept: (17.553360,17.8249668)
  - Distance: (0.116854,0.1171273)

Prediction

predicted_air_time ← predict(air_time_model, newdata = new_data)

Distance = 5000
Predicted Airtime: 602.6423
Predicted Interval: (577.3874,627.8971)

rows_with_long_distance ← subset(df_2019_sorted, DISTANCE > 5000, AIR_TIME)

3 flights with 5095 distances
Airtimes are (638,634,547)

638 - 0.117 * 95 = 626.885
634 - 0.117 * 95 = 622.885
547 - 0.117 * 95 = 535.885
Avg: **595.2183** (Predicted **602.6423**)

Conclusion

- There is a strong linear relationship between distance & air time
- Around 800 miles are travelled per flight
- Average air time is 112 minutes
- Based on our inferential statistics
  - We reject the null hypothesis
- There is a significant statistical relationship between distance & air time that as the flight distance increase by a mile, air time increase by 0.117 minutes (7 seconds)

What's next?

- Comfortably packing enough entertainment for my next flight
- Important statistics to calculate fuel required for aircraft
- Save money for airlines

References:

- Patrick Zel, (2023). Flight Delay and Cancellation Dataset 2019-2023 [Data set]. Kaggle. Retrieved from https://www.kaggle.com/datasets/patrickzel/flight-delay-and-cancellation-dataset-2019-2023
- In the preparation of this work, I received assistance from OpenAI's ChatGPT, which provided guidance on syntax for the R programming language and advice on citation formats.

Thank you. Please feel free to ask any questions. 😄