Sky Route

Business STRATEGY

Byung Jae Bae

Table of CONTENTS

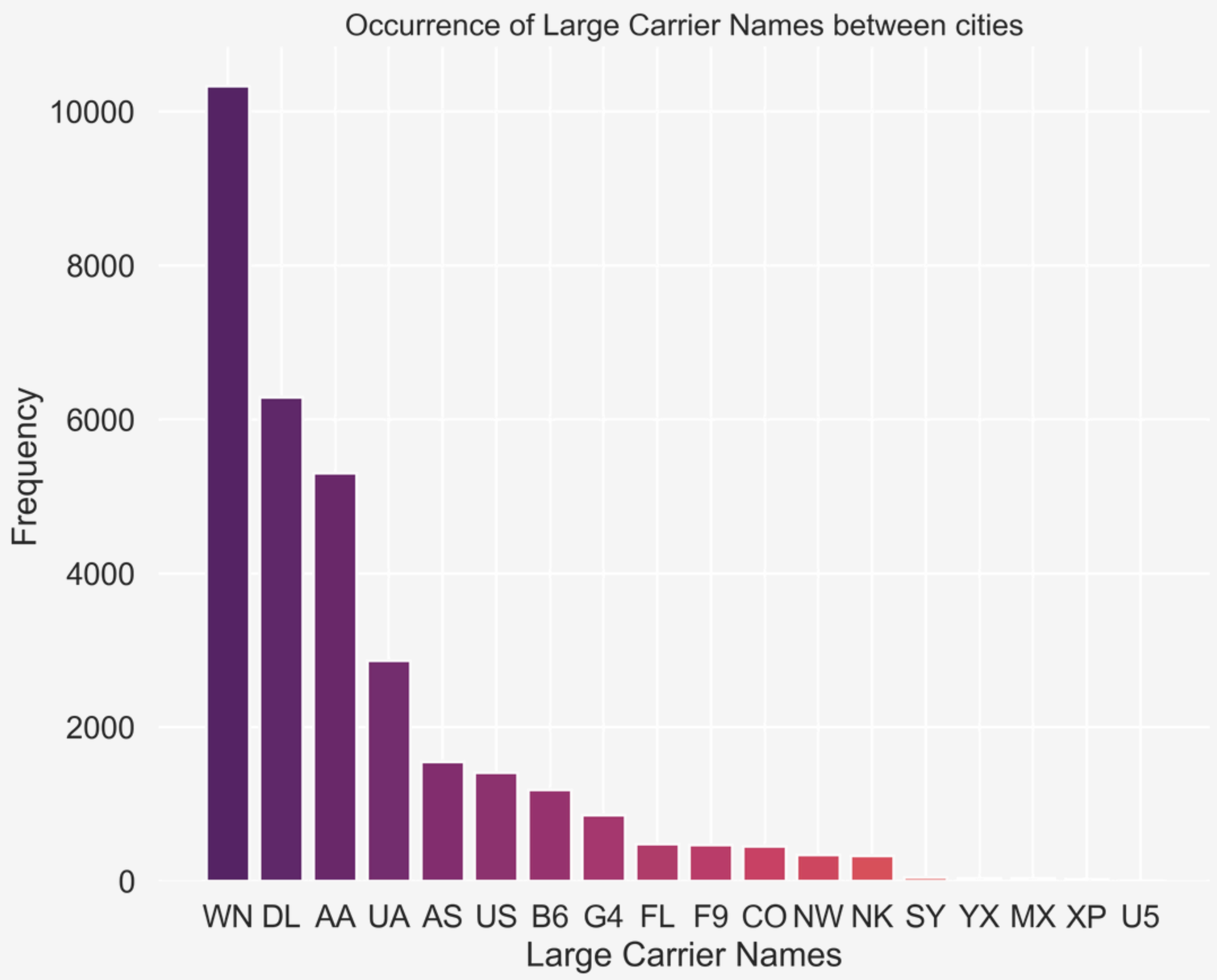
OBJECTIVE

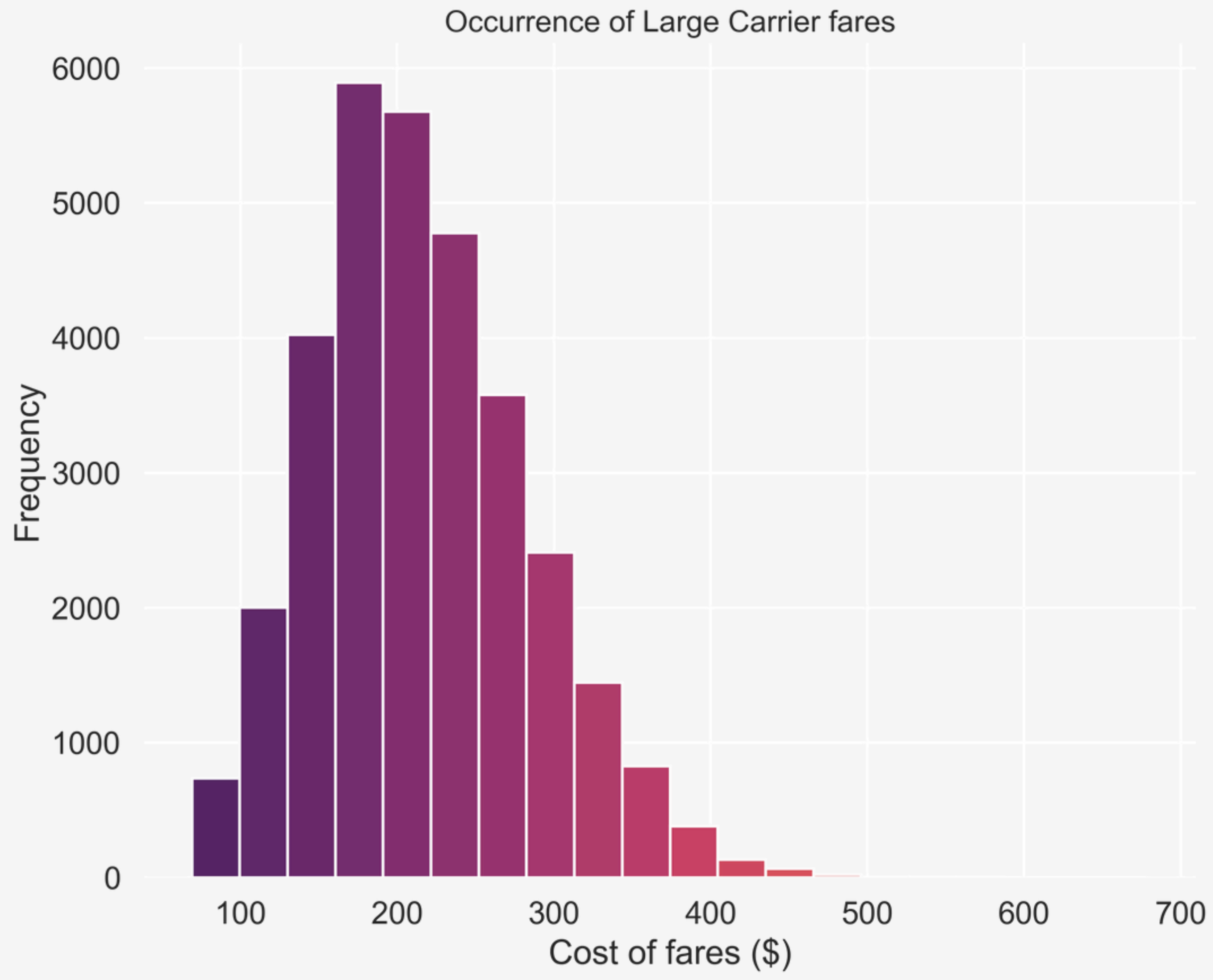
Identify trends in fare prices using route data to determine what factors contribute to higher or lower fares and create a market strategy.

INTRODUCTION

The airline industry is servicing 1510 unique routes to 163 cities. There are 21 total carriers competing in this space and 18 carriers compete in both the large and low cost carrier market. Out of the 21 total carriers, only 3 carriers uniquely compete only in the low cost market.

In this highly competitive market, is there still room for disruption?



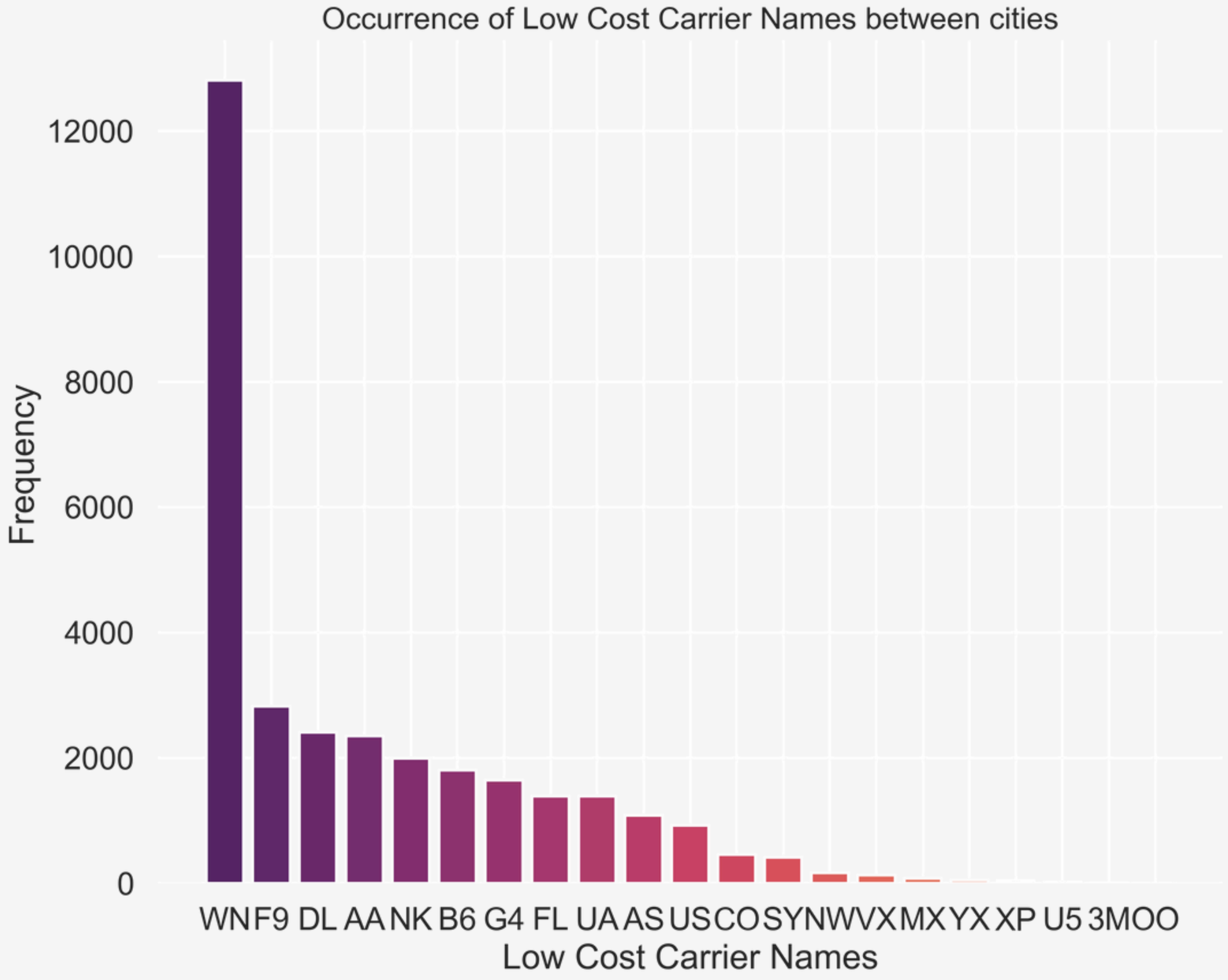


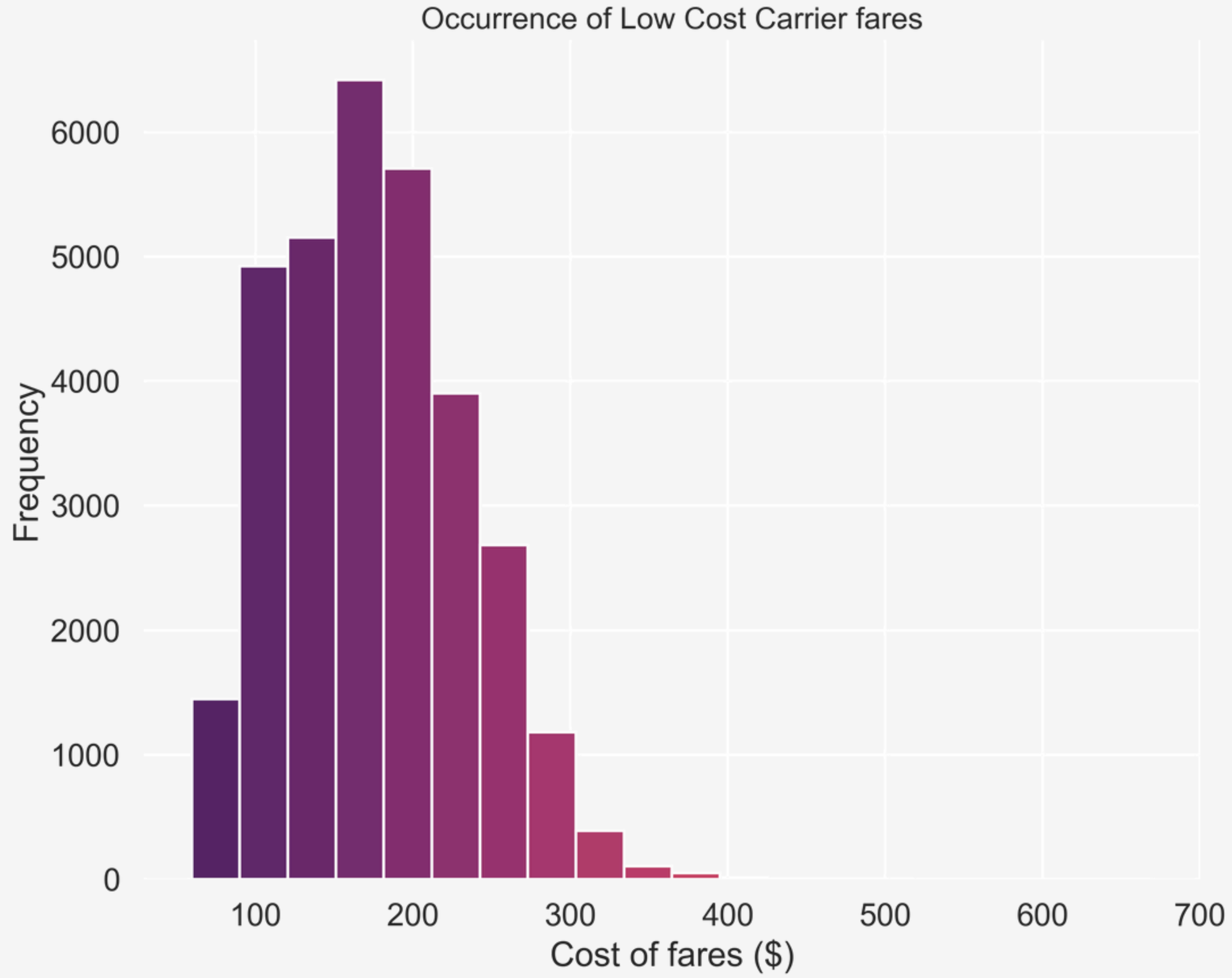
LARGE AIRLINES

Among the large airlines, there is relatively healthy competition between the top 3 major carriers. WN Airline still holds a massive lead of around 200% from the 2nd largest airline, DL.

The average cost of most fares of large carriers hover around $200.

The large airlines also consistently have the higher market share over their low cost carrier counterparts in all 1297 competitive routes showing that most people prefer to fly with larger airlines that with smaller ones in all routes.





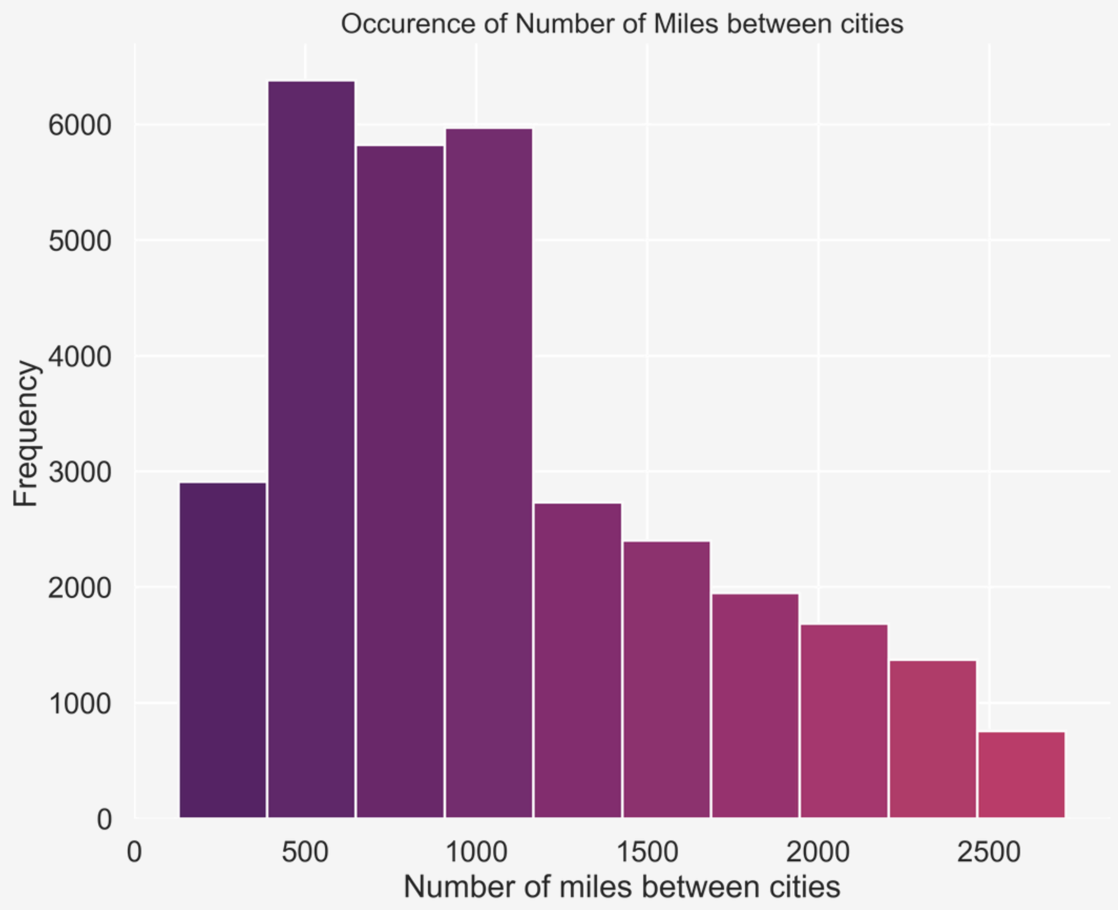
LOW COST AIRLINES

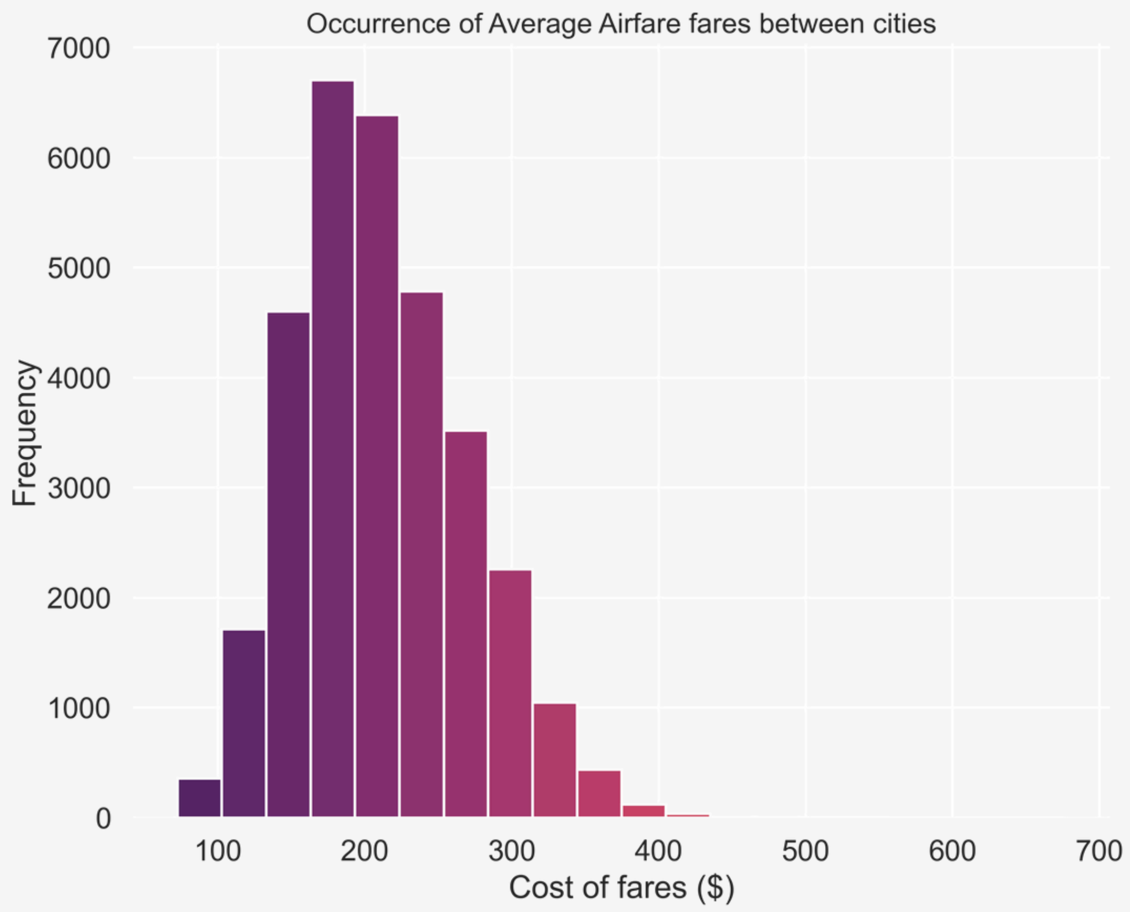
WN Airline dominates the low cost airline industry with almost a 6 times lead from the next largest competitor, F9 airline.

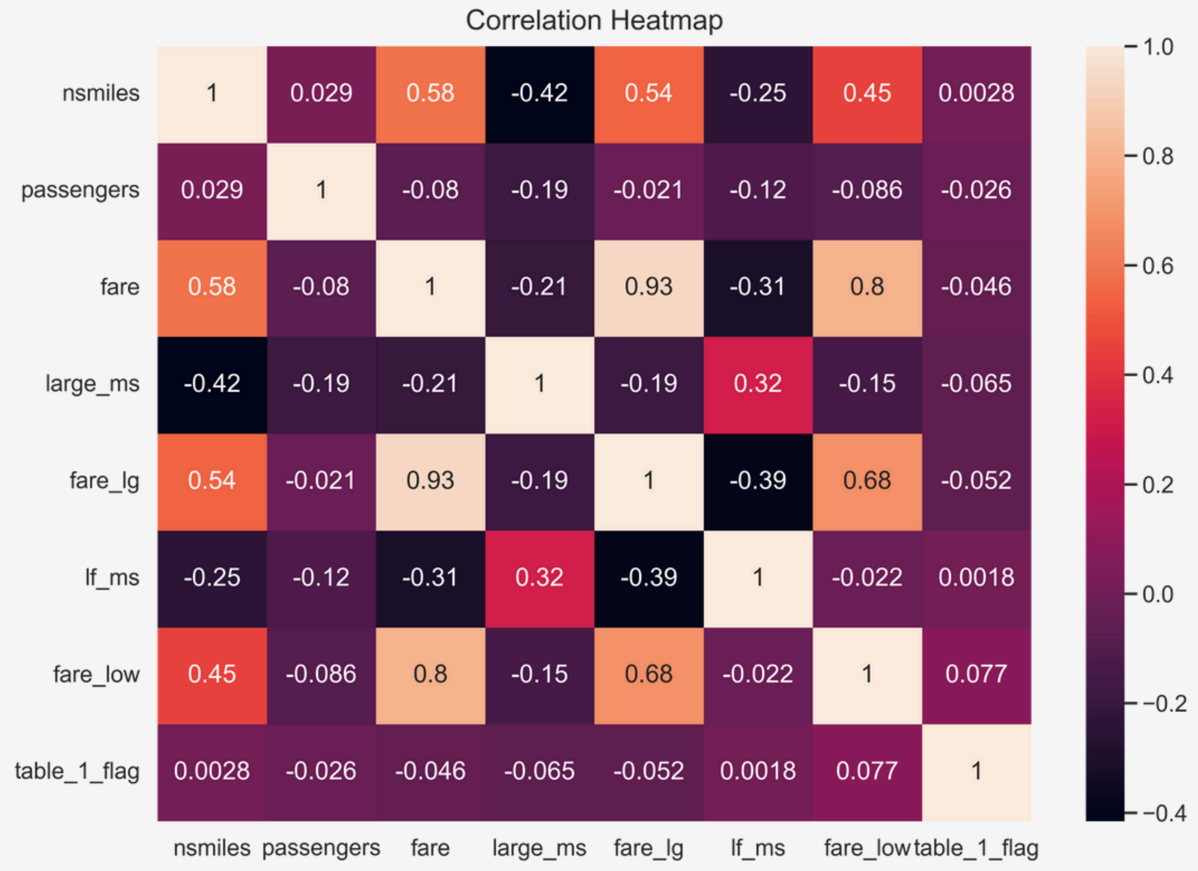
Only 3 low cost airlines: VX, OO, 3M do not compete in the large airline sector.

The majority of low cost airline fares range between $150- $200, undercutting the large airlines in fares, but also often overlapping in the higher price tier fares.

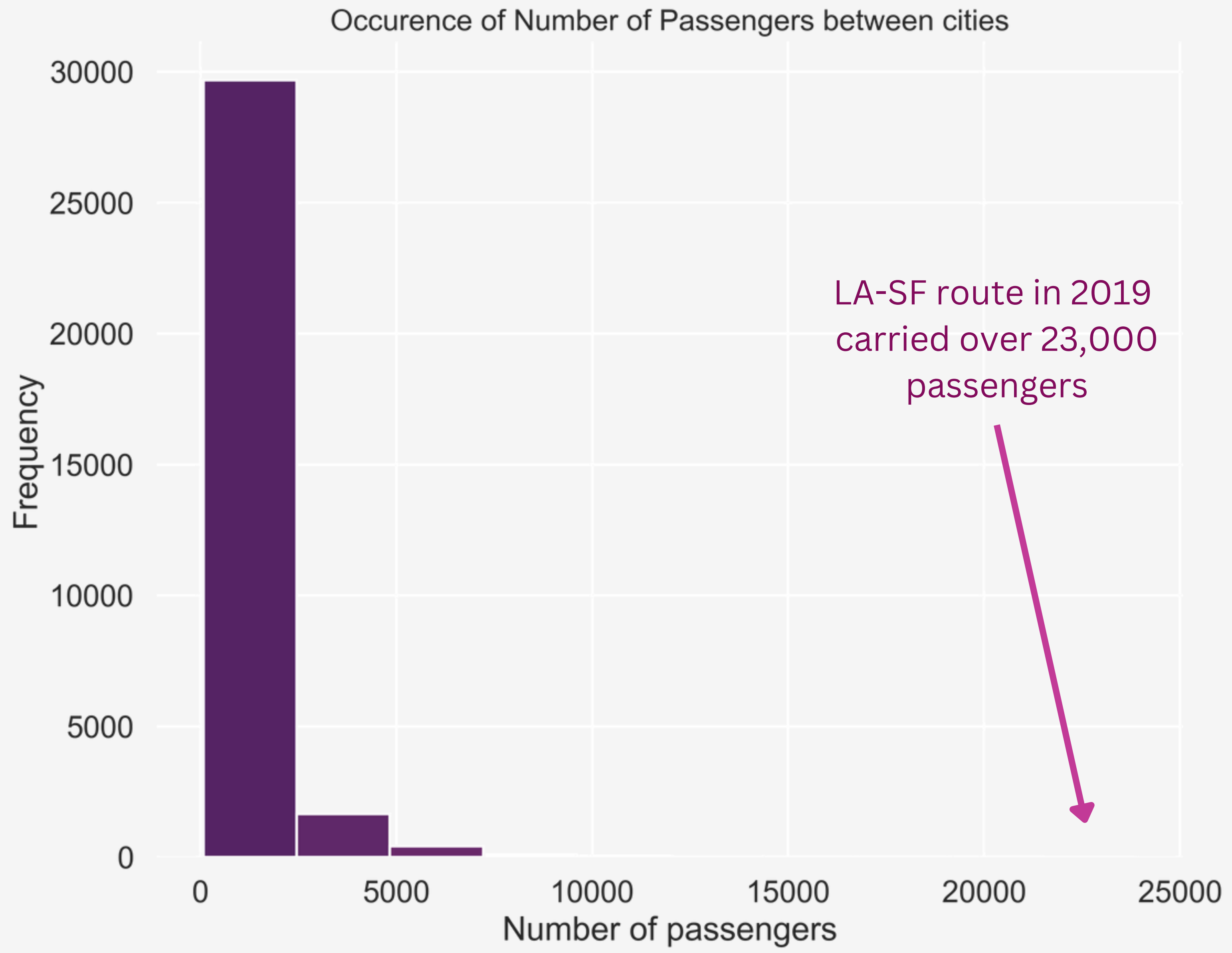
DATA OVERVIEW







DATA OVERVIEW

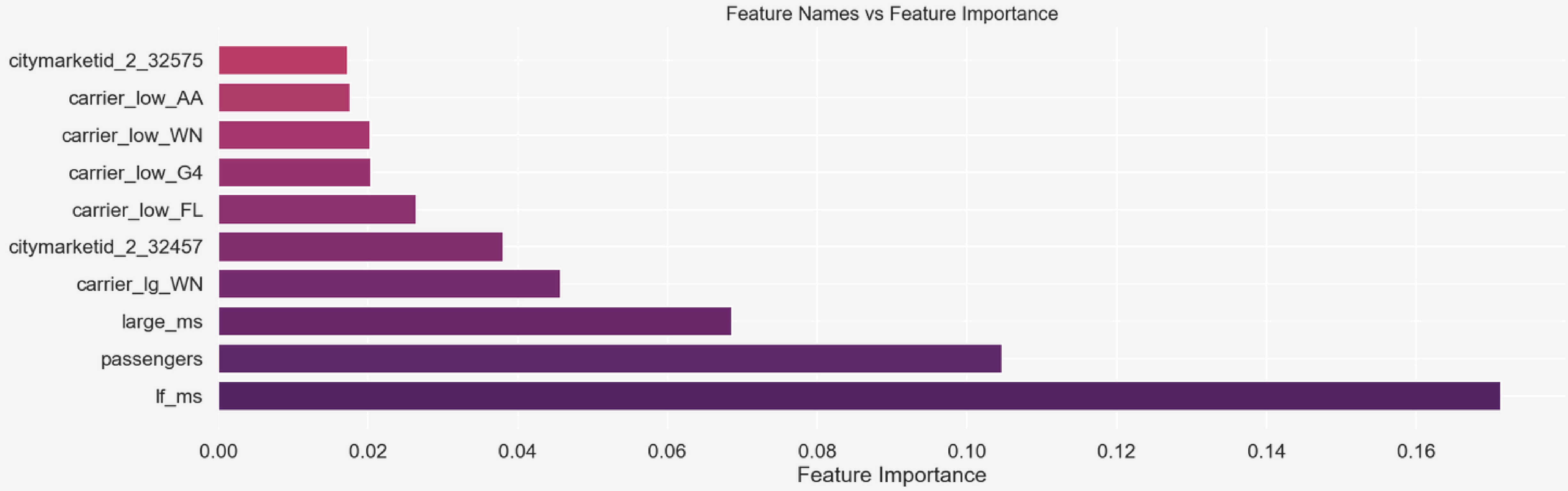


Market OPPORTUNITIES

There seem to be some market inefficiencies with budget airlines.

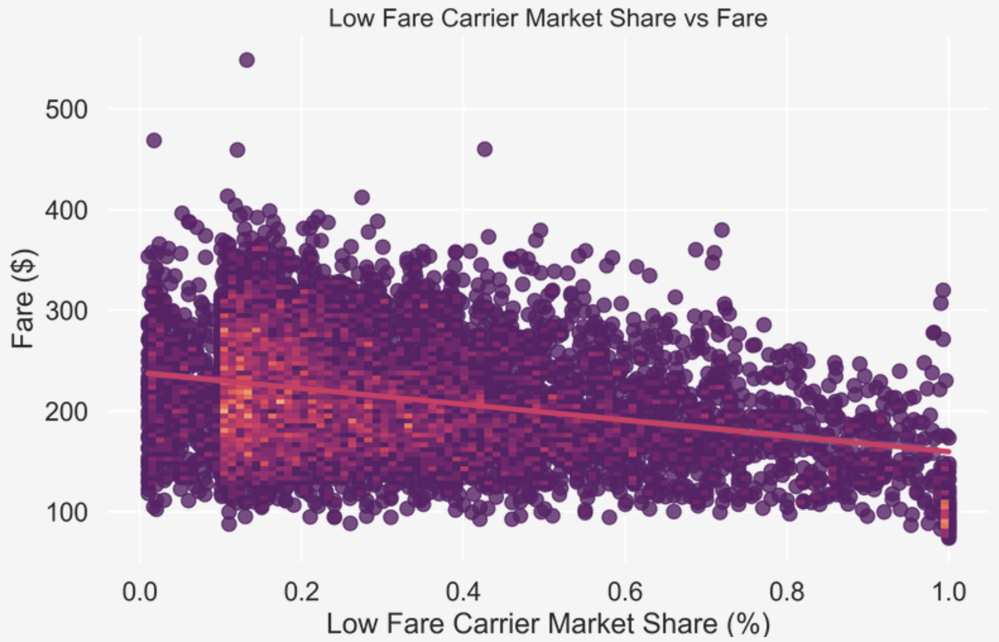
- For routes with many passengers and short flights, why is everyone opting to pay more for the larger carriers than budget ones?
- For example, looking at the market share of low cost carriers for the 2019 route from LA-SF that transported over 23,000 passengers, is only 14% compared to the large carrier's 59%.

Budget Airlines are supposed to leverage large number of passengers to offer the flights for much cheaper than larger airlines, but none of the low cost carriers have a market share greater than traditional carriers in any of the competitive routes (where large carrier and low cost carriers are different companies).



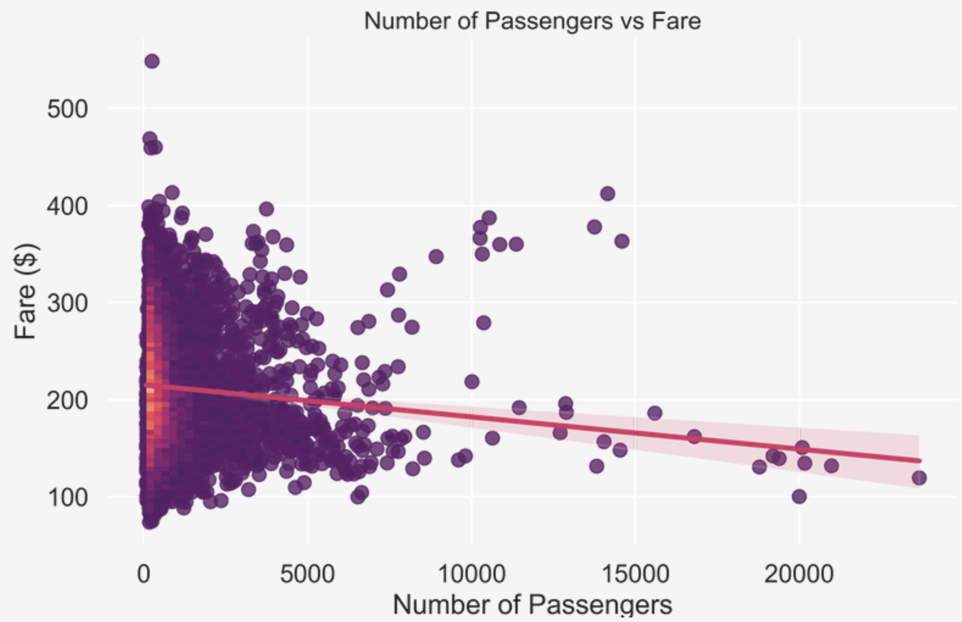
LEVERAGING SCALE

01



More Budget Carrier Market Share leads to cheaper fares

02



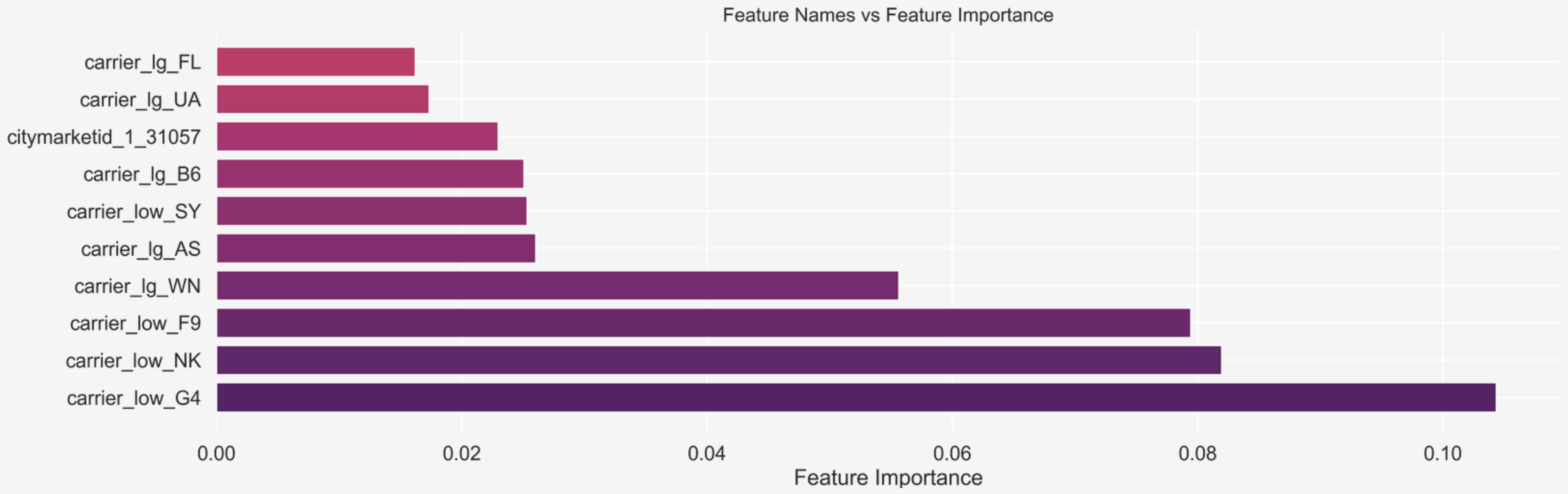
More Passengers lead to cheaper fares

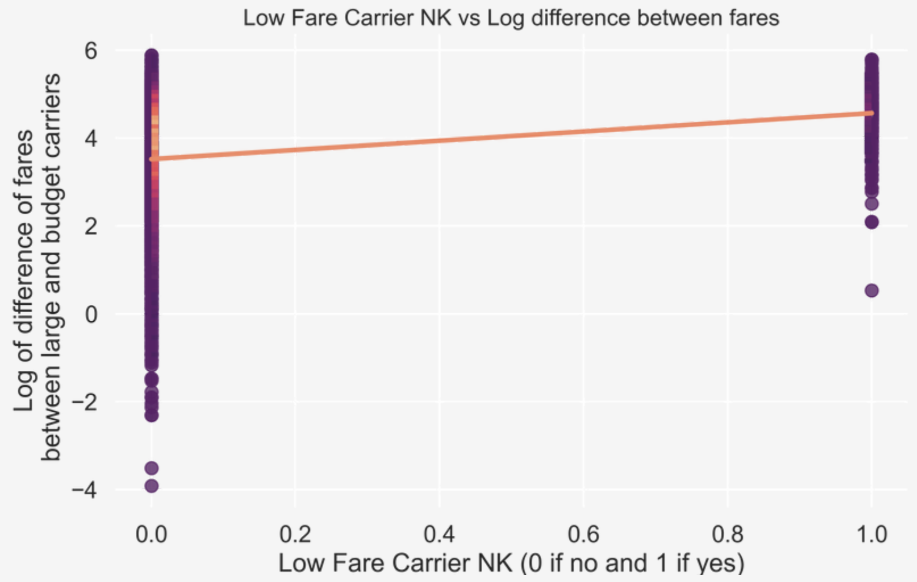
Analysis of PRICING DIFFERENCES

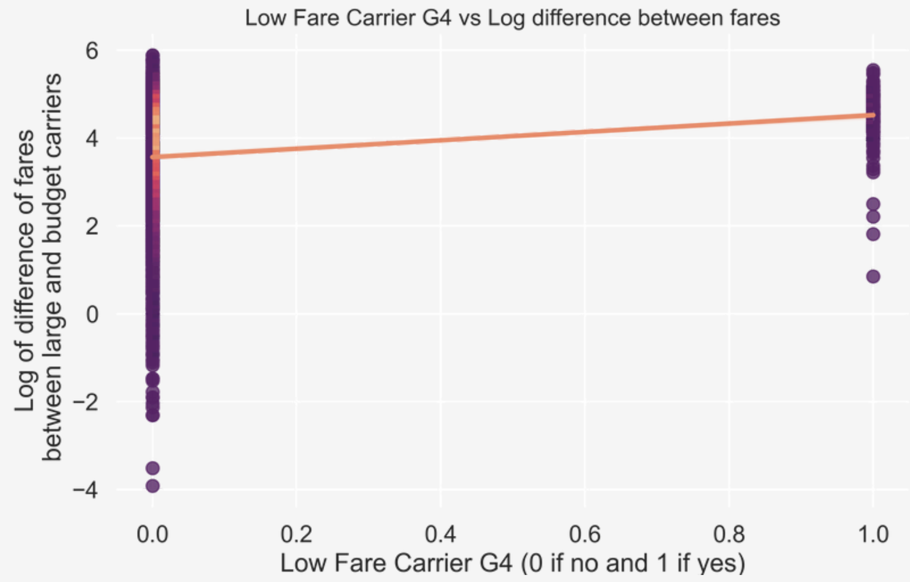
I analyzed what could be the driving factors for the difference between the large carrier fare and the low cost fares and to see if it was larger airlines dictating the price differences or low cost ones.

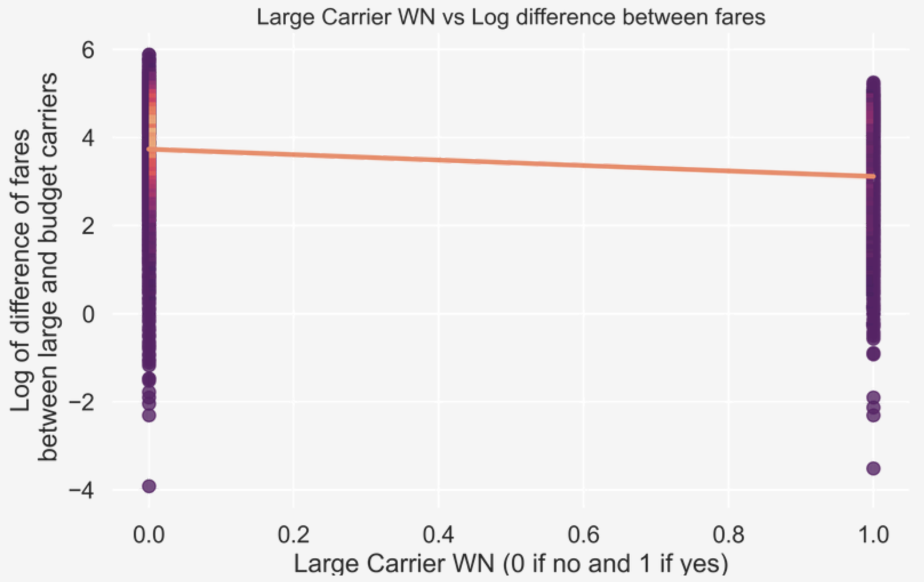
I created a metric, log(large carrier fare - low cost fare), where log is used for technical reasons. This metric can be interpreted as a larger difference in fares as number goes up and smaller differences for numbers going down.

Low cost carriers create a bigger margin between the low cost and large carriers but still do not capture as much market share.









Proposed STRATEGIES

Compete with Budget Carriers

There are less people willing to fly with budget carriers over larger carriers, even with the lower prices, reducing the market share and lowering the effectiveness of scale to provide lower prices. Without scale, the budget airlines are very limited in their ability to lower prices further. If Sky Route is able to scale to transport more passengers, there is disruption that can be made.

Look further into WN’s success

WN has reached an inflection point where they are able to operate their business in a way that caters to both large and low cost airlines without having to lower their cost as much as the other budget airlines do, like F9 or G4. This might be a good future topic to look into

Focus on High Volume Routes

To meet scaling demands, the best way is to focus on serving a few high volume routes. These routes are highly competitive but the market share data shows us that there is practically no difference between low cost carriers’ market share in highly populated routes vs lowly populated routes. The budget carrier’s expertise should be focused on serving these routes well to maximize passenger capacity

Sky Route

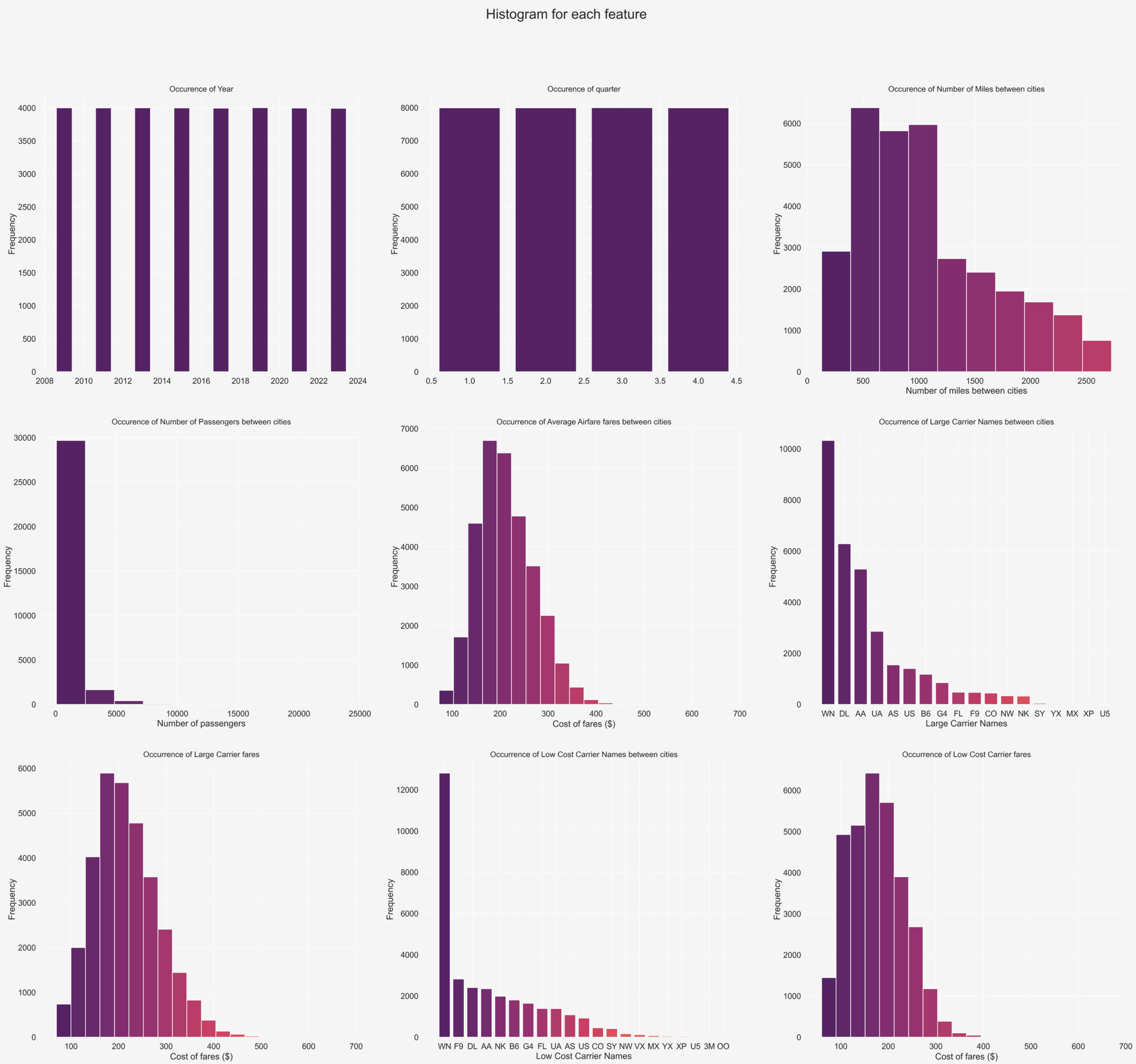
CONCLUSION

1. Compete with budget Carriers
2. Look in to WN’s success
3. Focus on High Volume Routes

And most importantly leverage scale!

APPENDIX

Histograms for each feature



APPENDIX

Data Cleaning Process

1. Checked if all city names pair with unique city id.
2. Checked if sum of market share percentage between low cost and large carriers were less than or equal to 100% (if they were different airlines)
3. Check for NaN values (excluding Geocoded city)
4. Check if there are no floating point quarters or years
5. Delete all invalid rows

Left with 31995 rows for experiment 1 (Predicting Fare)

APPENDIX

Data Cleaning Process (Cont.)

For Experiment 2 (Predicting delta of fares):

1. Repeat all data cleaning steps from experiement 1.
2. Get rows that are competitive routes (i.e. more than 5000 passengers)
3. Create new data column named “delta” for log of difference of fares
    a. subtract “fare_lg” from “fare_low”
    b. if result is 0, add 0.001 (so that there is no zero values input into log)
    c. apply log the the results

Left with 21410 rows for experiment 2 (Predicting delta of fares)

APPENDIX

Train Test Split

Applied Train Test split of (80/20) using the “year” feature to stratify the two datasets.

This makes sure there are equal number of data points according to the year in each dataset. This helps reduce fluctuations in yearly differences and only focuses more on seasonality.

APPENDIX

Model 1.x was used to predict average fares given following features:

Categorical features were One Hot Encoded

quarter: category
citymarketid_1: category
citymarketid_2: category
passengers: int64
fare: float64
carrier_lg: category
large_ms: float64
carrier_low: category
lf_ms: float64
table_1_flag: category

APPENDIX

Model 2.x was used to predict Log Difference of Average Large carrier Fare and Averge Low Cost Carrier fare for a given route using the following features:

quarter: category
citymarketid_1: category
citymarketid_2: category
nsmiles: int64
passengers: int64
carrier_lg: category
large_ms: float64
carrier_low: category
lf_ms: float64
table_1_flag: category
delta: float64

Categorical features were One Hot Encoded

Log Difference of Average Large carrier Fare and Averge Low Cost Carrier fare = Log(Average Large carrier Fare-Averge Low Cost Carrier fare)

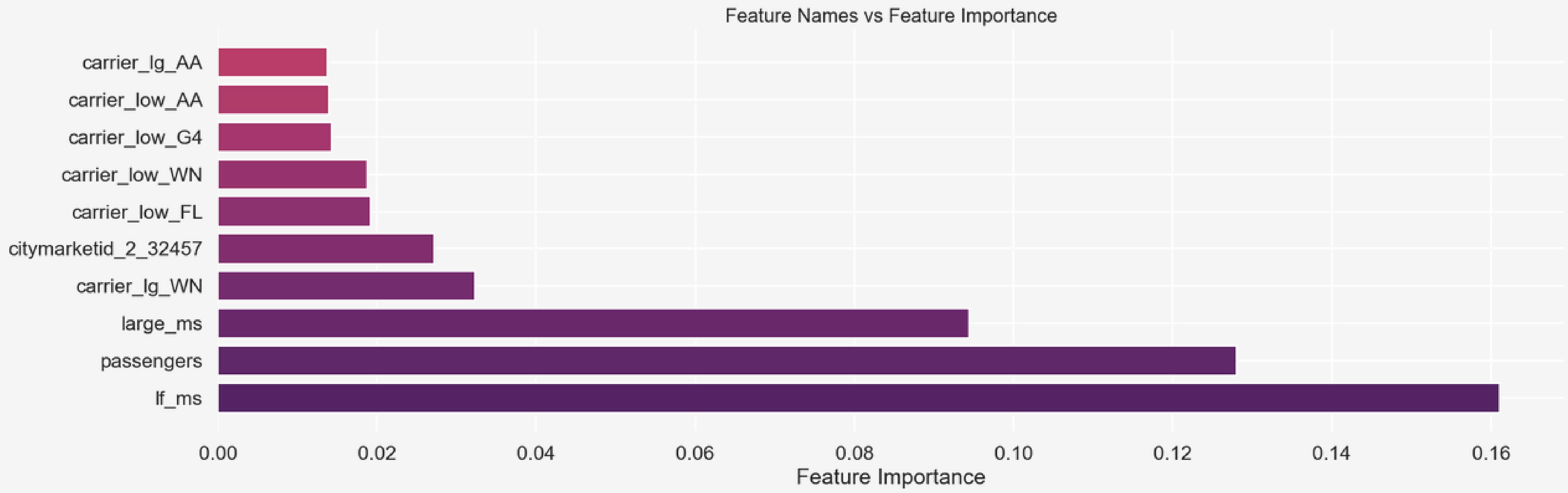
APPENDIX

Feature Descriptions

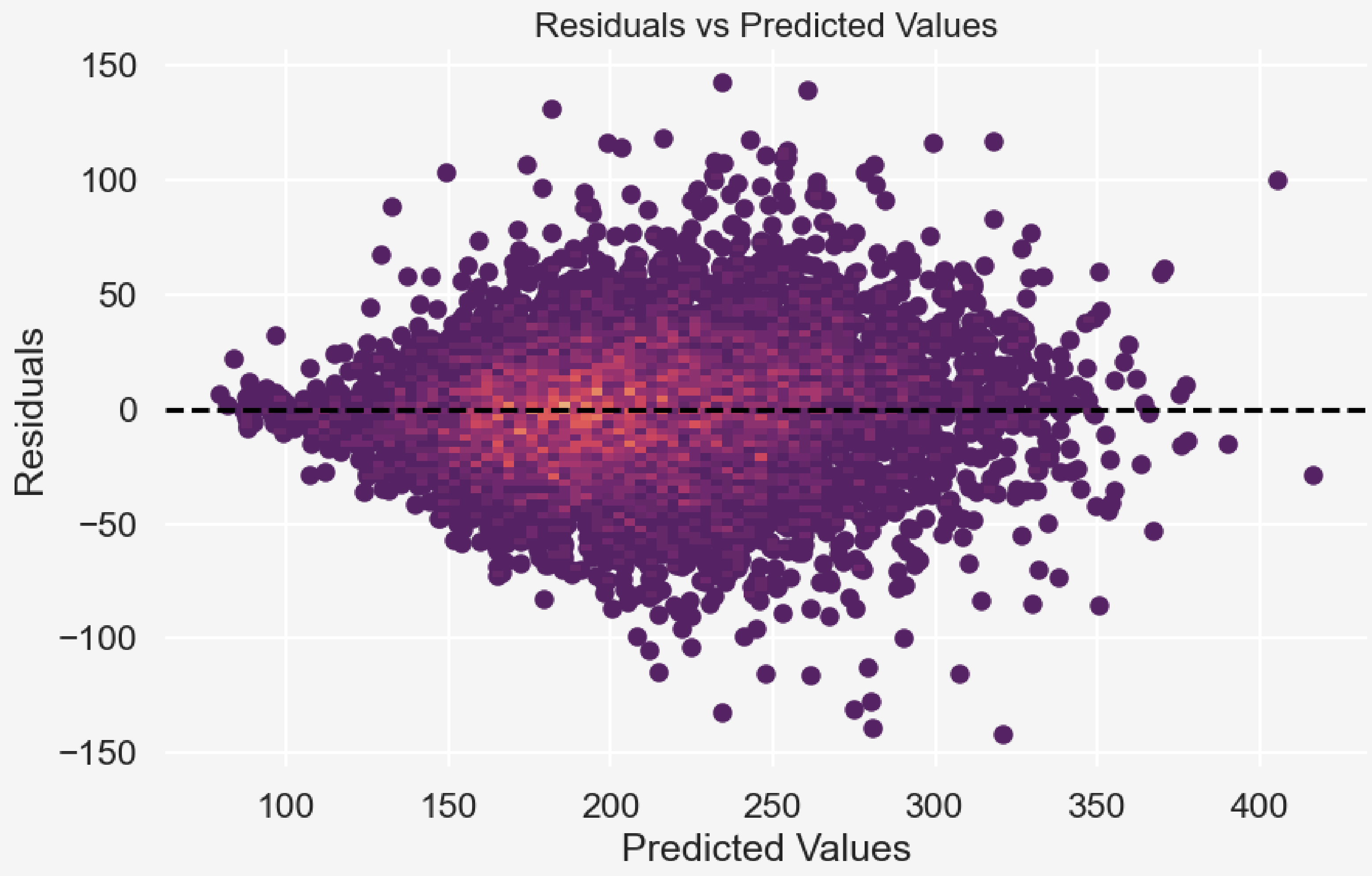
fare
The average airfare for the city pair.
passengers
The number of passengers traveling between the two cities.
nsmiles
The non-stop mileage between the cities.
large_ms
The market share of large carriers on the route.
fare_lg
The average fare charged by large carriers.
lf_ms
The market share of low-fare carriers on the route.
fare_low
The average fare charged by low-fare carriers.
Geocoded City1, Geocoded City2
Geocoded location data for the two cities.
carrier_lg
The identifier for the large carrier on the route.
carrier_low
The identifier for the low-fare carrier on the route.
table_1_flag
A binary indicator showing whether the route is part of the top 1,000 contiguous state city-pair markets.
delta
Log Difference of Average Large carrier Fare and Averge Low Cost Carrier fare

APPENDIX

Model 1.0: Random Forest



- 100 trees
- standard scikit-learn library



Metrics

- Train R^2: 0.9619763557500659
- Test R^2: 0.7296550118701761
- Train MSE: 129.26674688773625
- Test MSE: 891.4454422406401
- Train RMSE: 11.369553504326204
- Test RMSE: 29.857083619145392

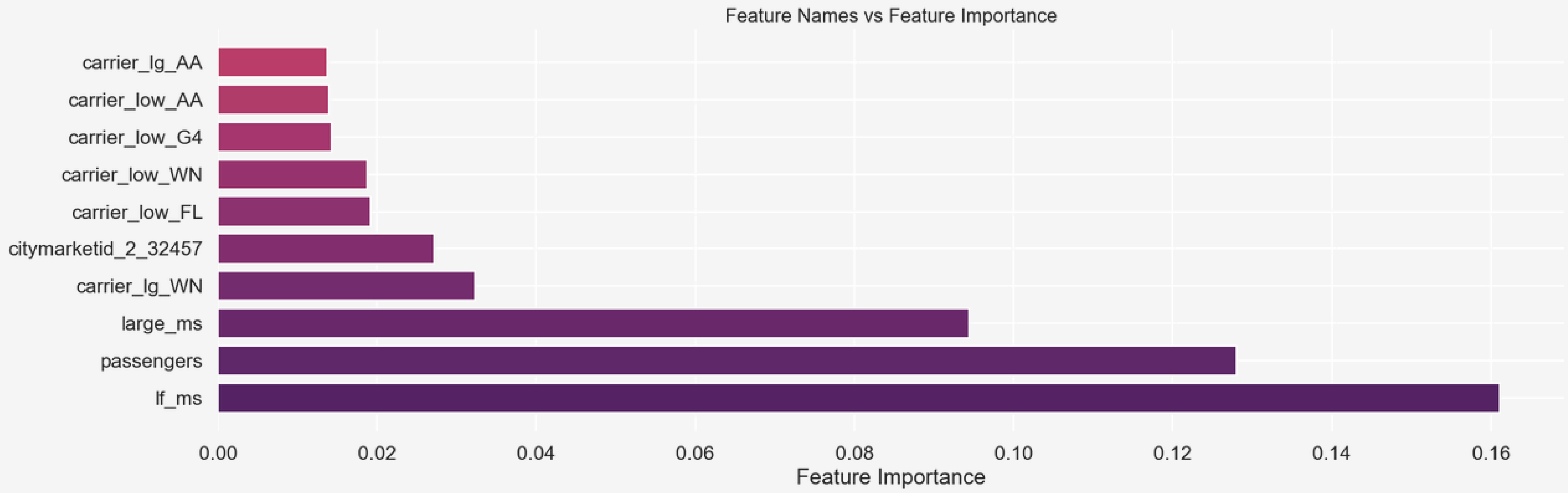
APPENDIX

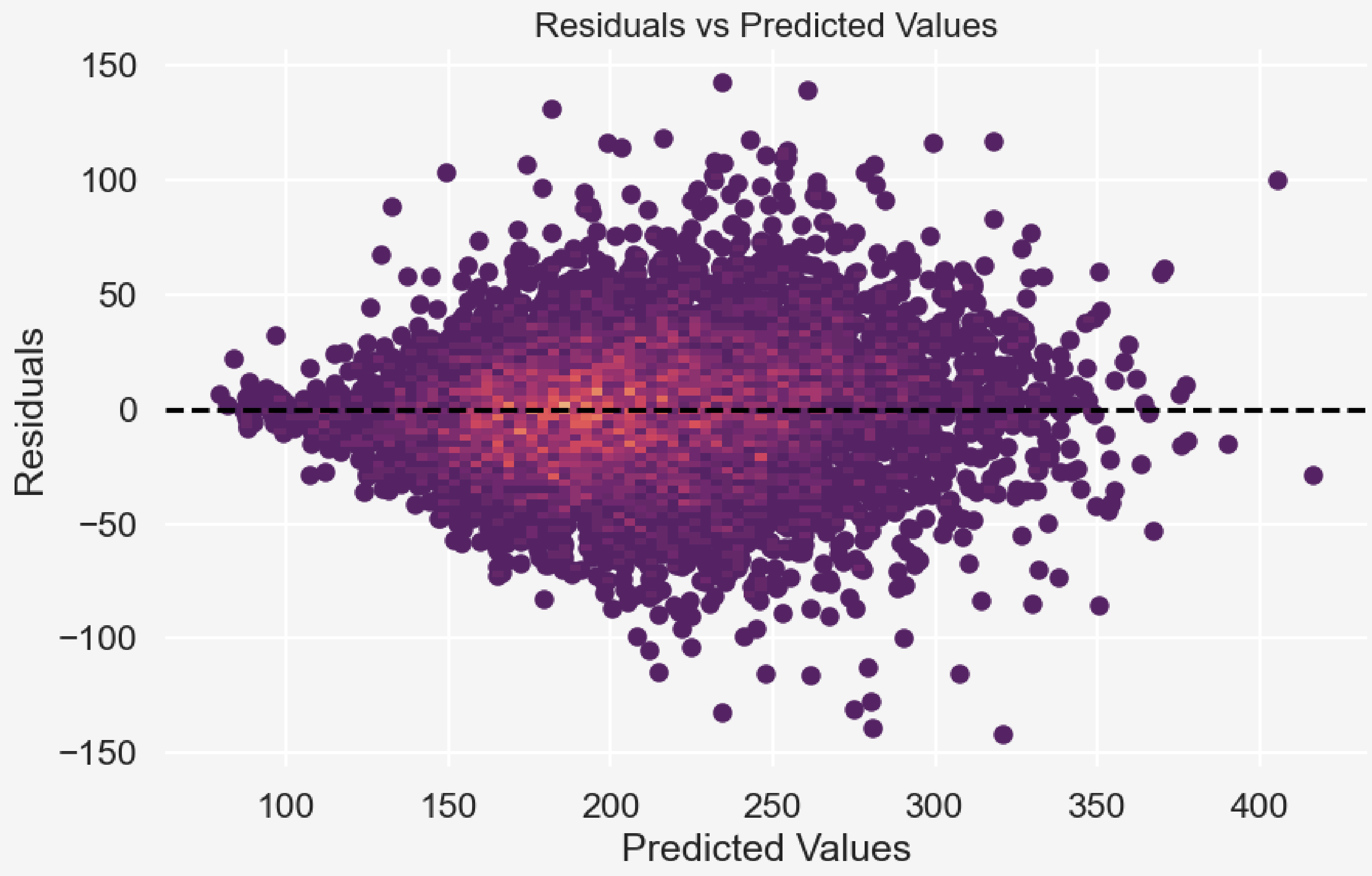
Model 1.1: Random Forest

- 100 trees
- minimum sample split = 65

Metrics

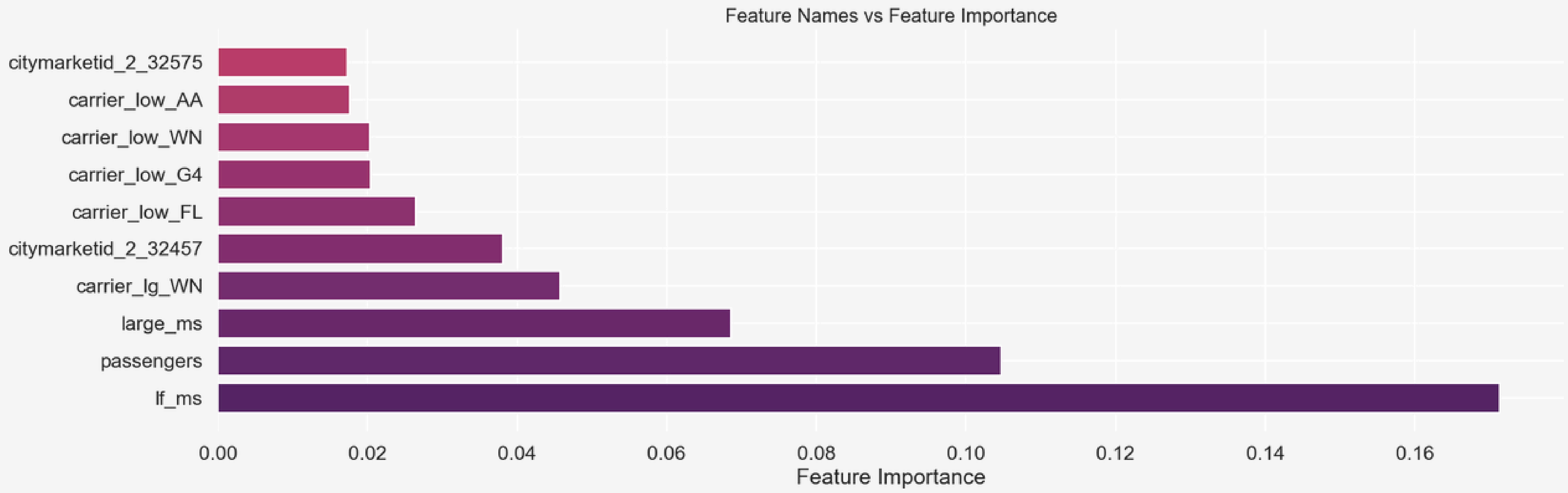
- Train R^2: 0.7419420356532337
- Test R^2: 0.647576771632218
- Train MSE: 877.3044829766858
- Test MSE: 1162.0932307327287
- Train RMSE: 29.619326173576024
- Test RMSE: 34.08948856605525



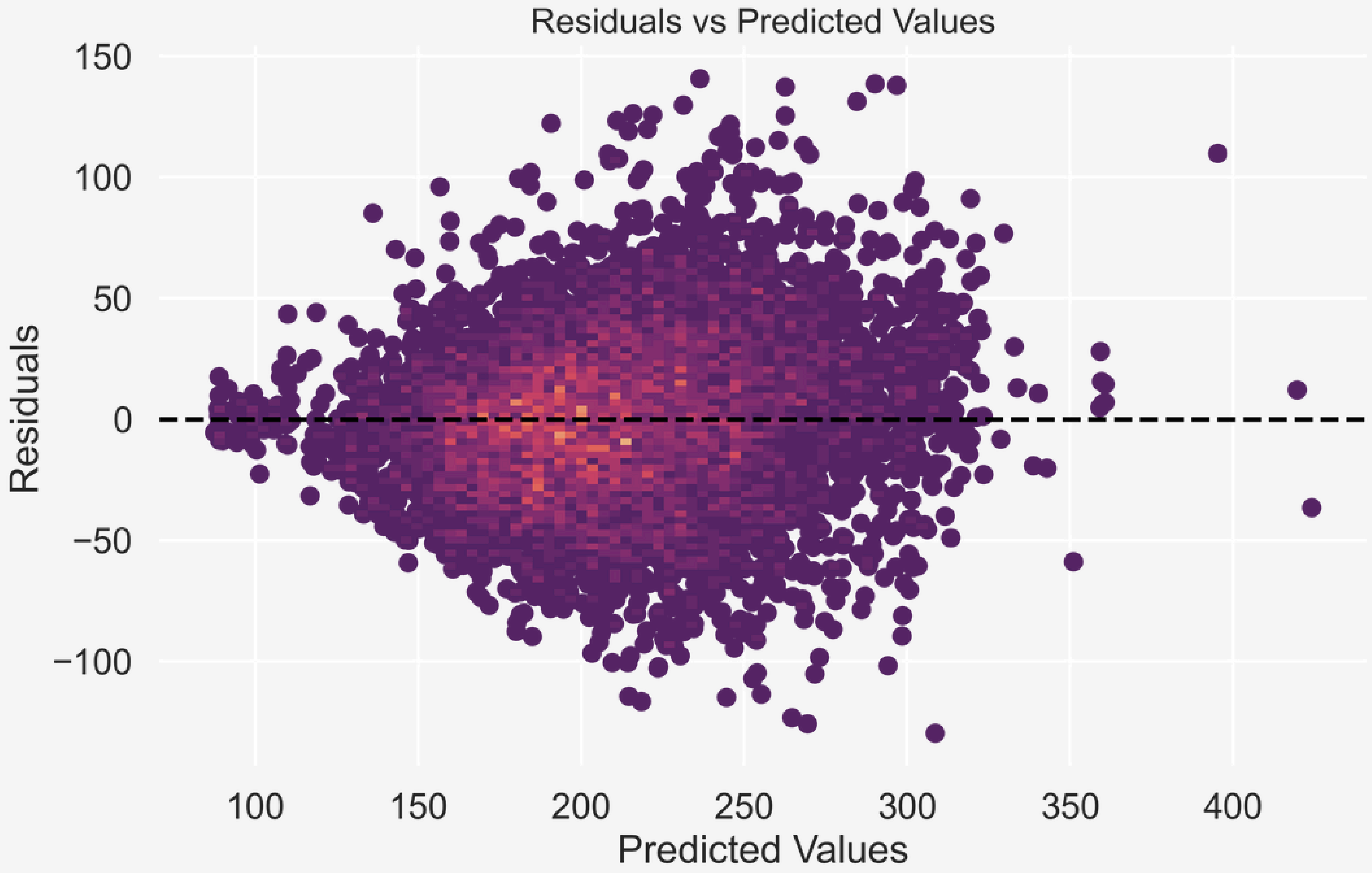


APPENDIX

Model 1.1: Random Forest



- 100 trees
- minimum sample split = 65



Metrics

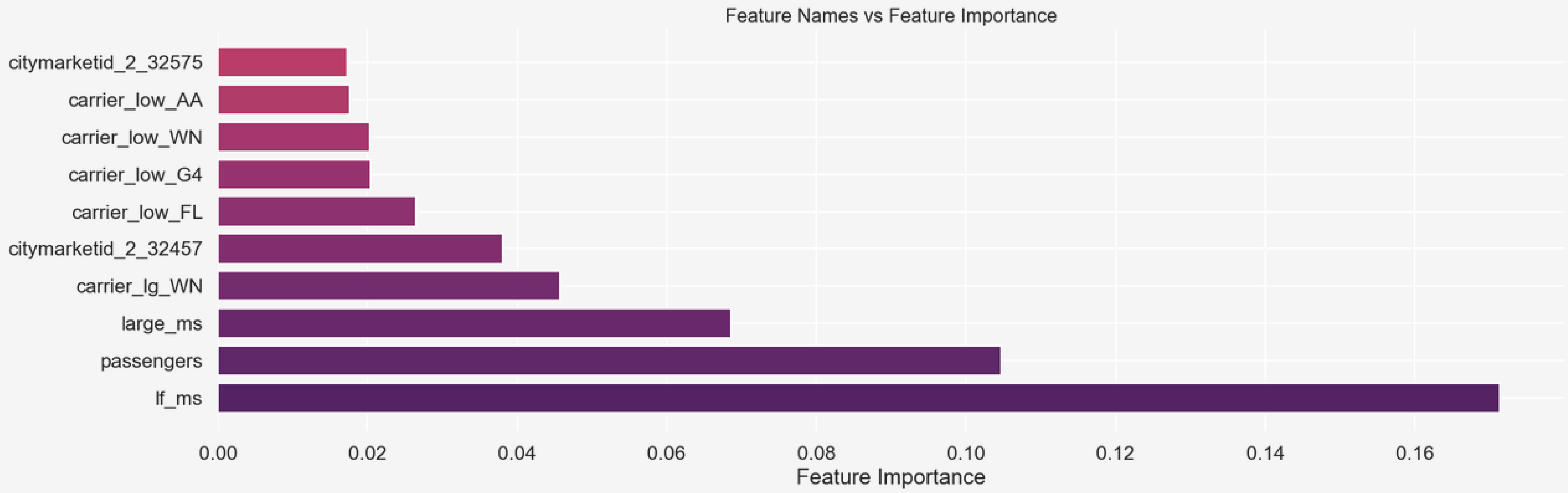
- Train R^2: 0.7419420356532337
- Test R^2: 0.647576771632218
- Train MSE: 877.3044829766858
- Test MSE: 1162.0932307327287
- Train RMSE: 29.619326173576024
- Test RMSE: 34.08948856605525

Used this in presentation due to lower overfitting (slides 9 and 10)

APPENDIX

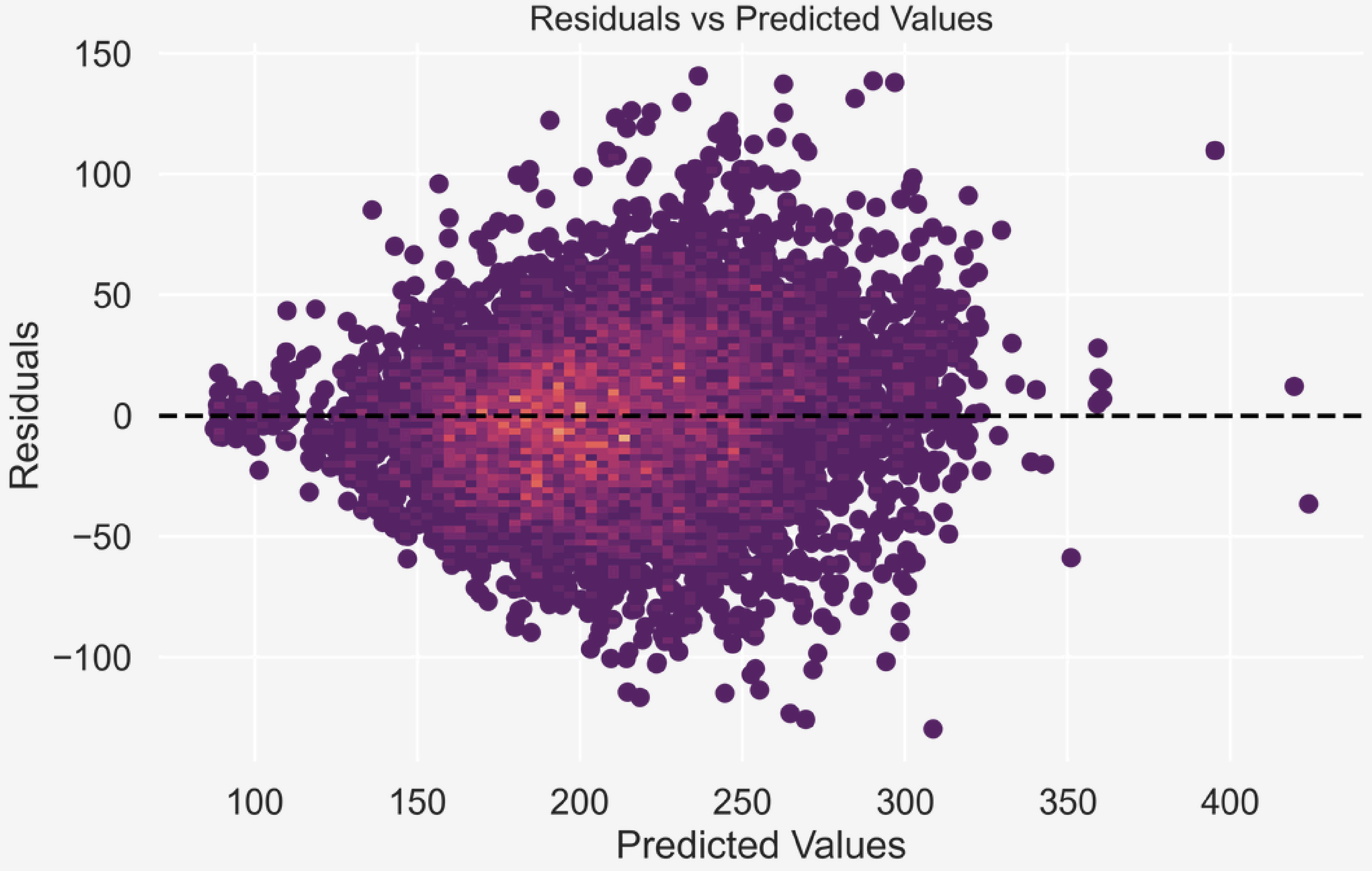
Model 1.1: Random Forest

- 100 trees
- minimum sample split = 65



Metrics

- Train R^2: 0.7419420356532337
- Test R^2: 0.647576771632218
- Train MSE: 877.3044829766858
- Test MSE: 1162.0932307327287
- Train RMSE: 29.619326173576024
- Test RMSE: 34.08948856605525

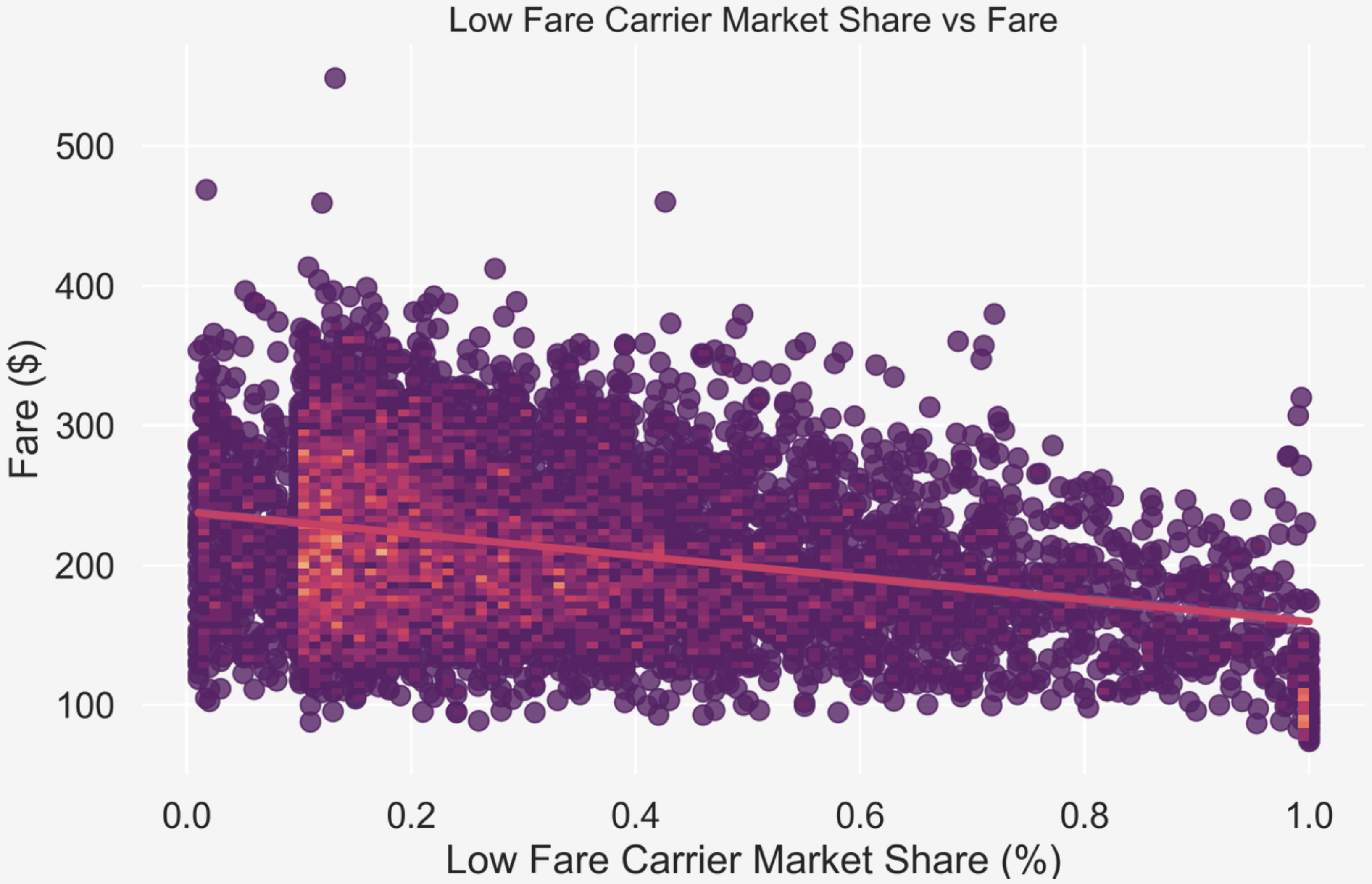


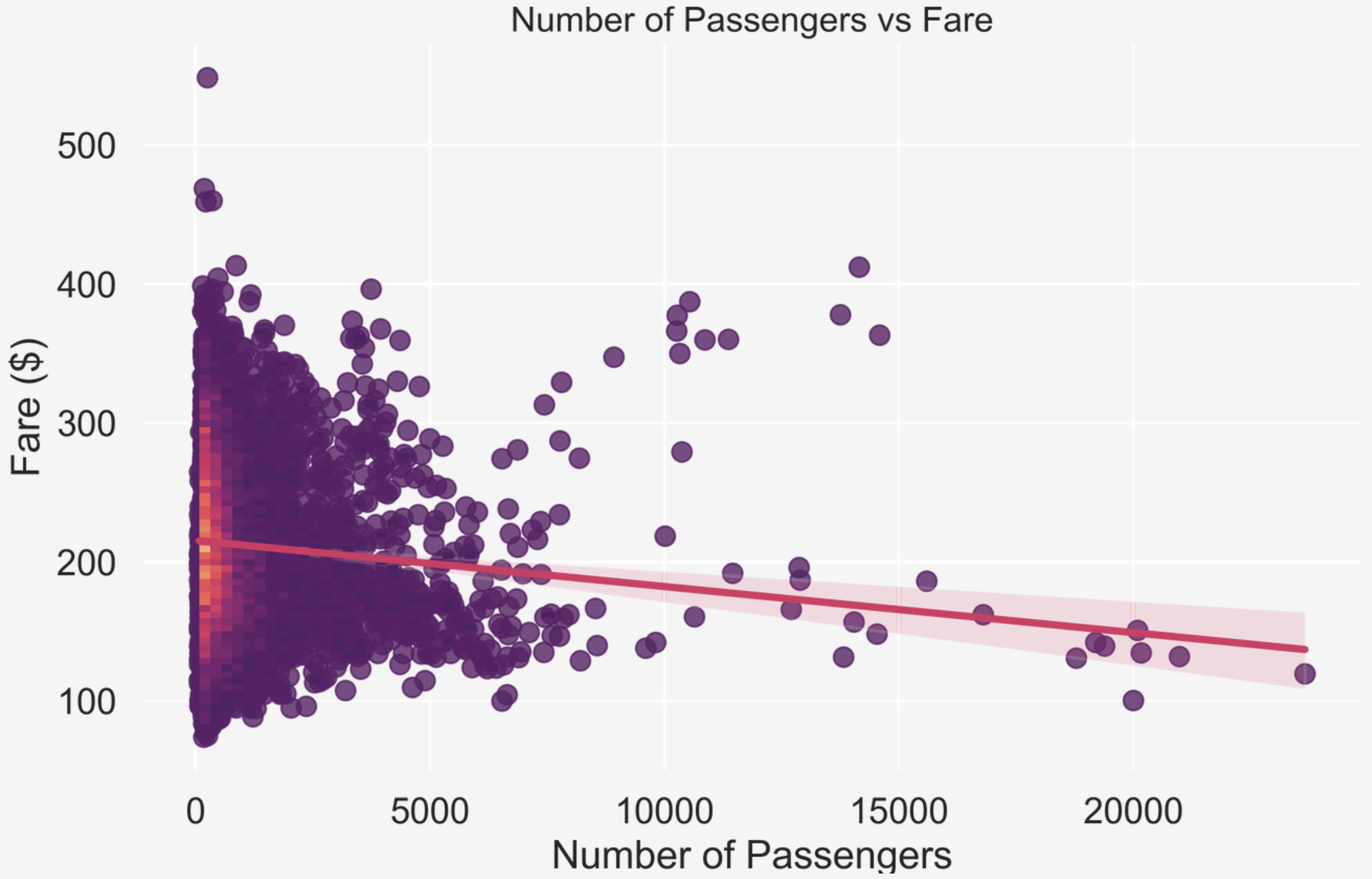
Used this in presentation due to lower overfitting (slides 9 and 10)

APPENDIX

Model 1.1: Random Forest (Cont.)

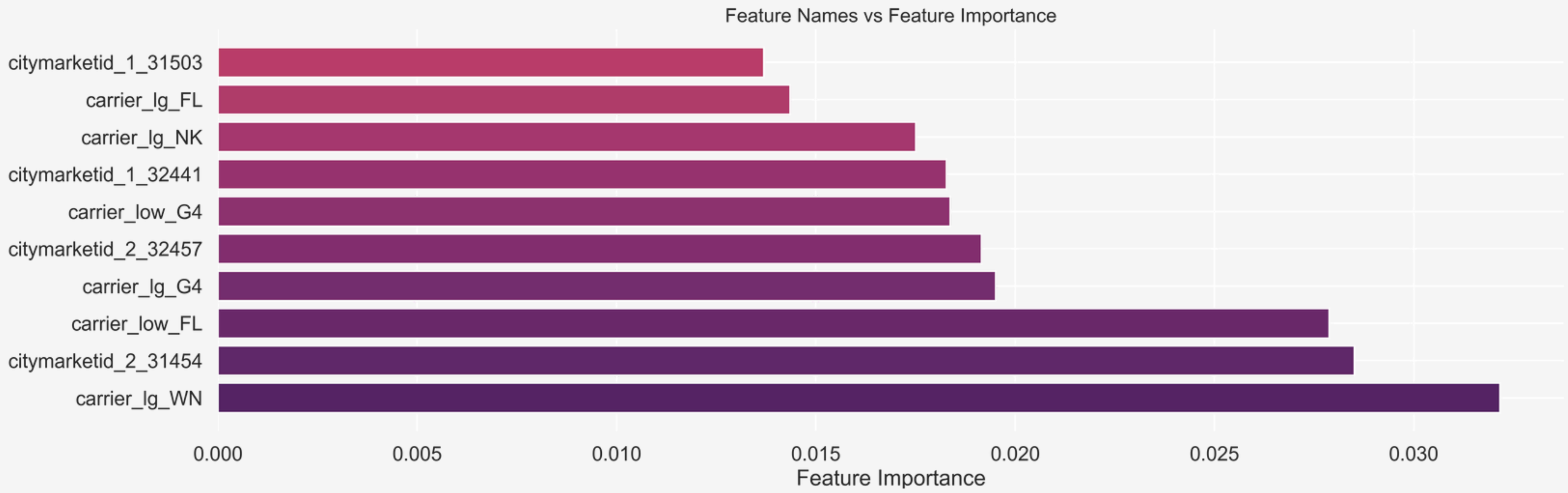
Plotted 2 Variables against dependant variable



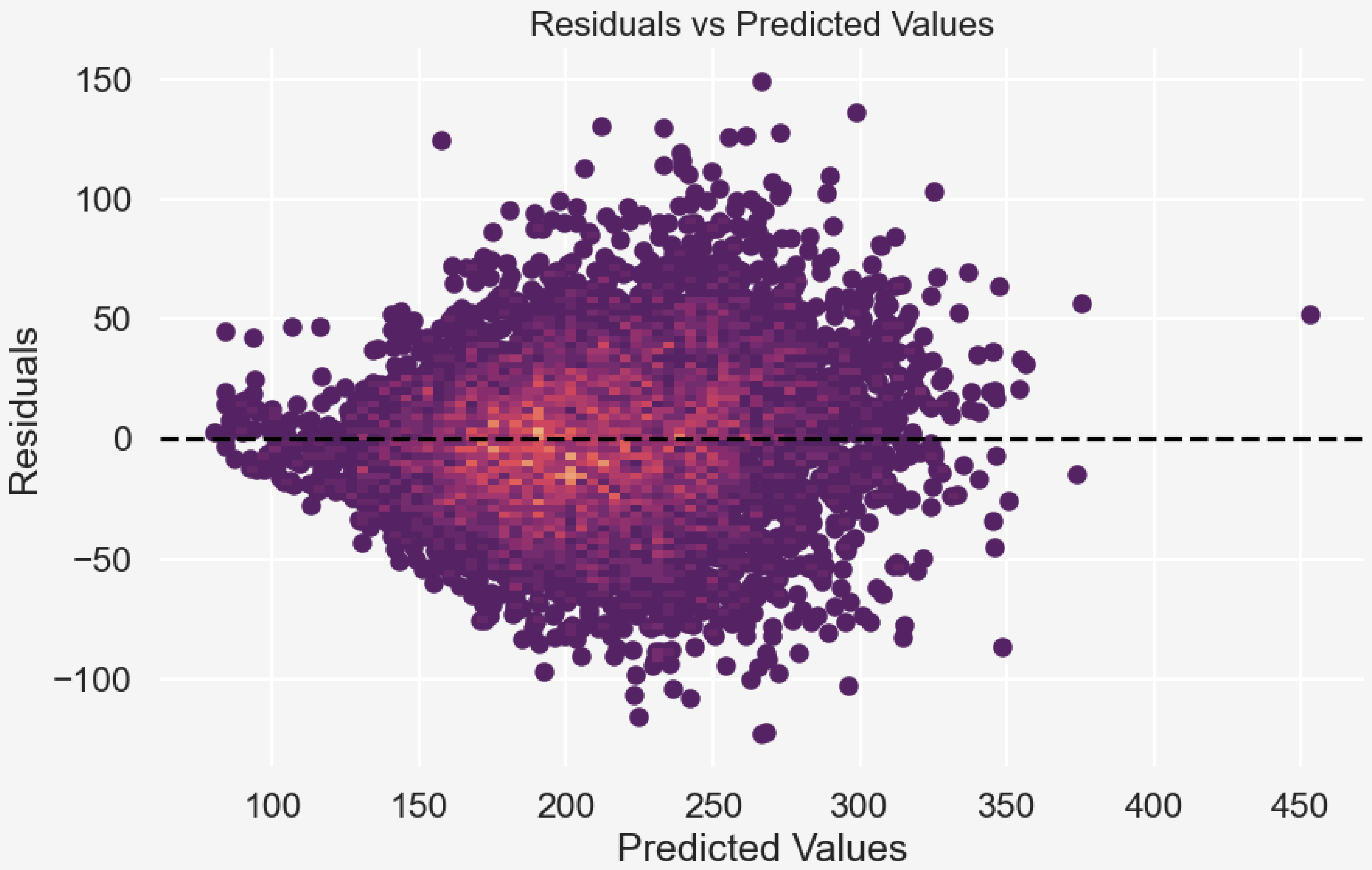


APPENDIX

Model 1.3: XGBoost



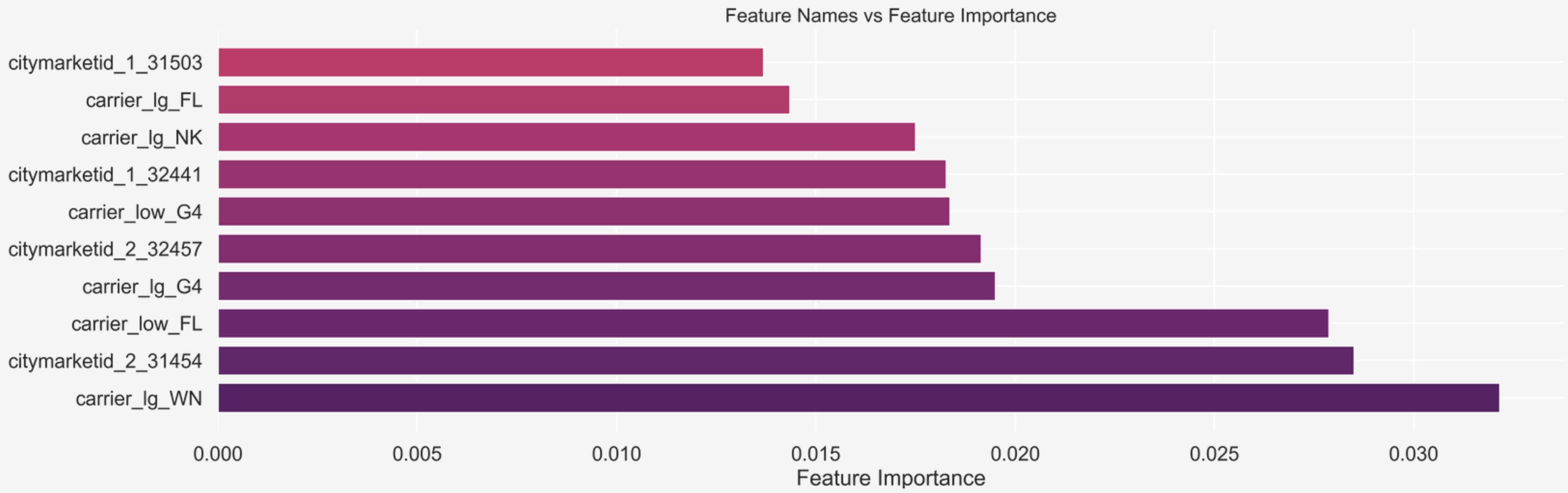
- all xgboost library defaults



Metrics

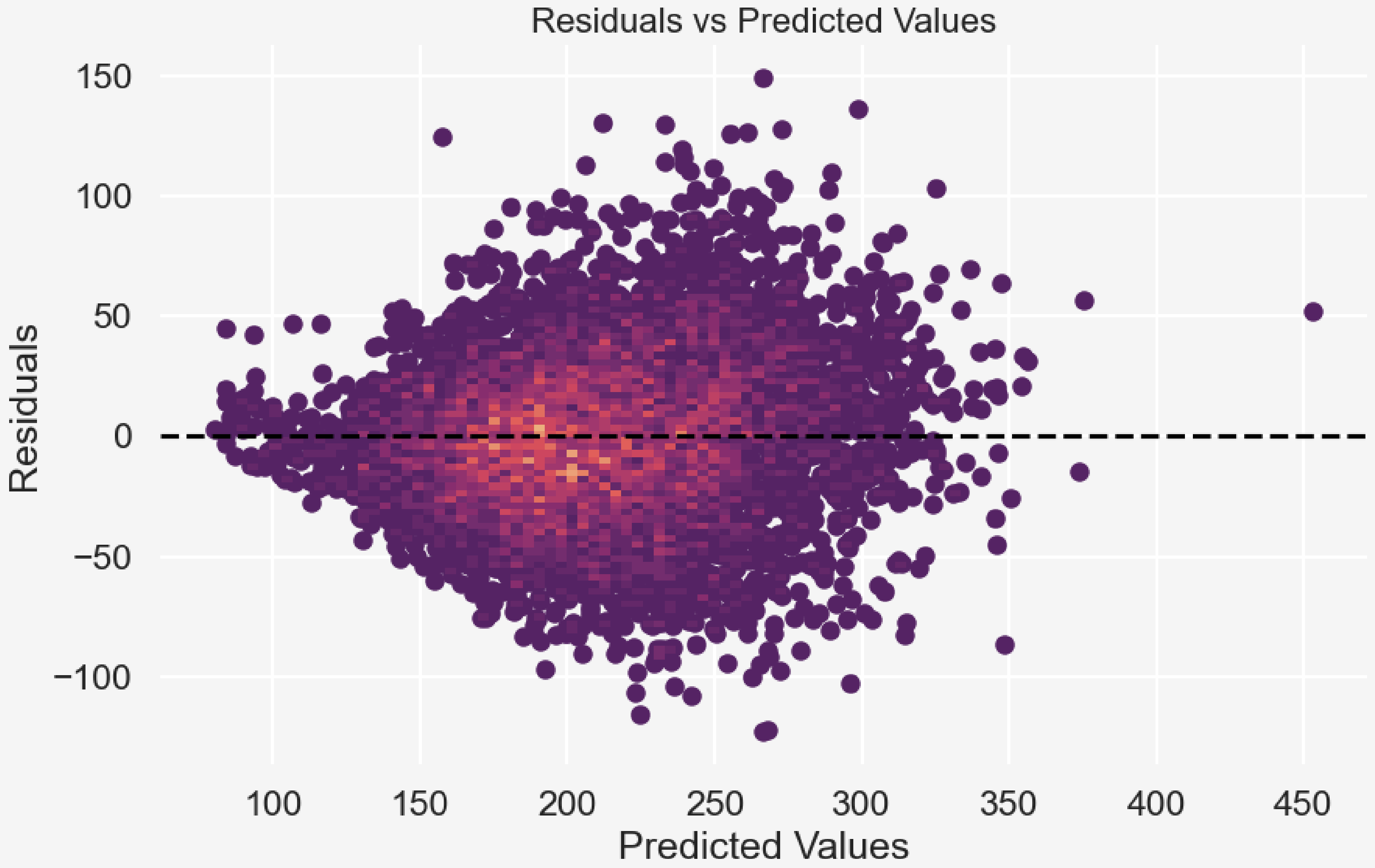
- Train R^2: 0.7401233714006803
- Test R^2: 0.6770486357933511
- Train MSE: 883.4872888661819
- Test MSE: 1064.9116289485762
- Train RMSE: 29.723514073308724
- Test RMSE: 32.632983757979844

APPENDIX



Model 1.4: ElasticNet (L1 & L2)

- all scikit learn defaults



Metrics

- Train R^2: 0.08331707309349023
- Test R^2: 0.08405199362566884
- Train MSE: 3116.3930292909326
- Test MSE: 3020.2804248758403
- Train RMSE: 55.82466327073485
- Test RMSE: 54.95707802345245

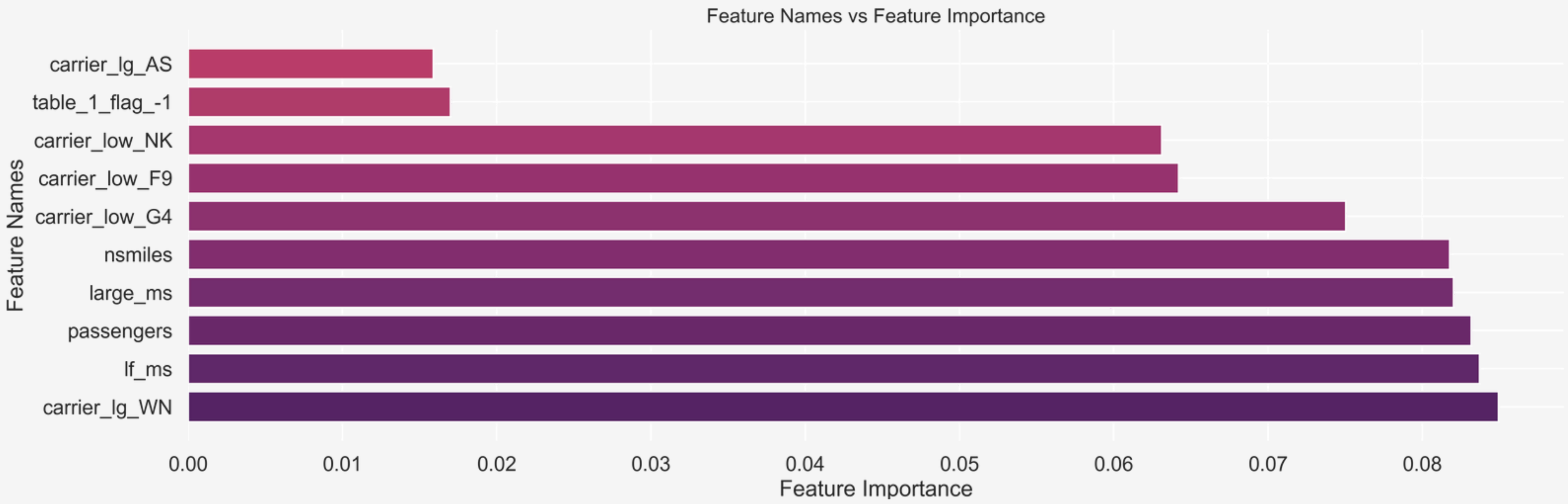
APPENDIX

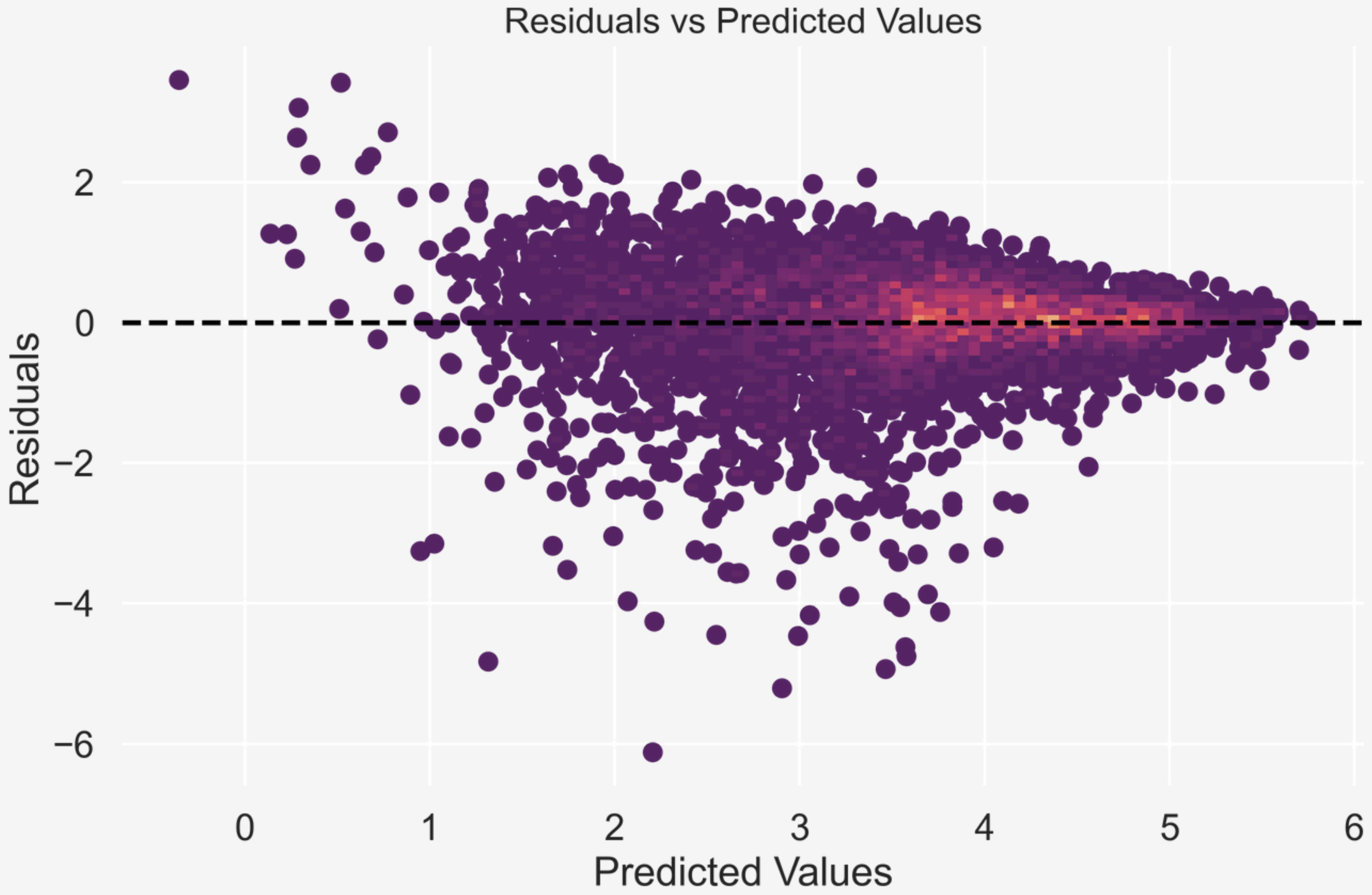
Model 2.1: Random Forest

- 100 trees
- all scikit learn defaults

Metrics

- Train R^2: 0.9424527168892749
- Test R^2: 0.5900086504291013
- Train MSE: 0.08009582983127352
- Test MSE: 0.5775695826990117
- Train RMSE: 0.28301206658245776
- Test RMSE: 0.7599799883543064

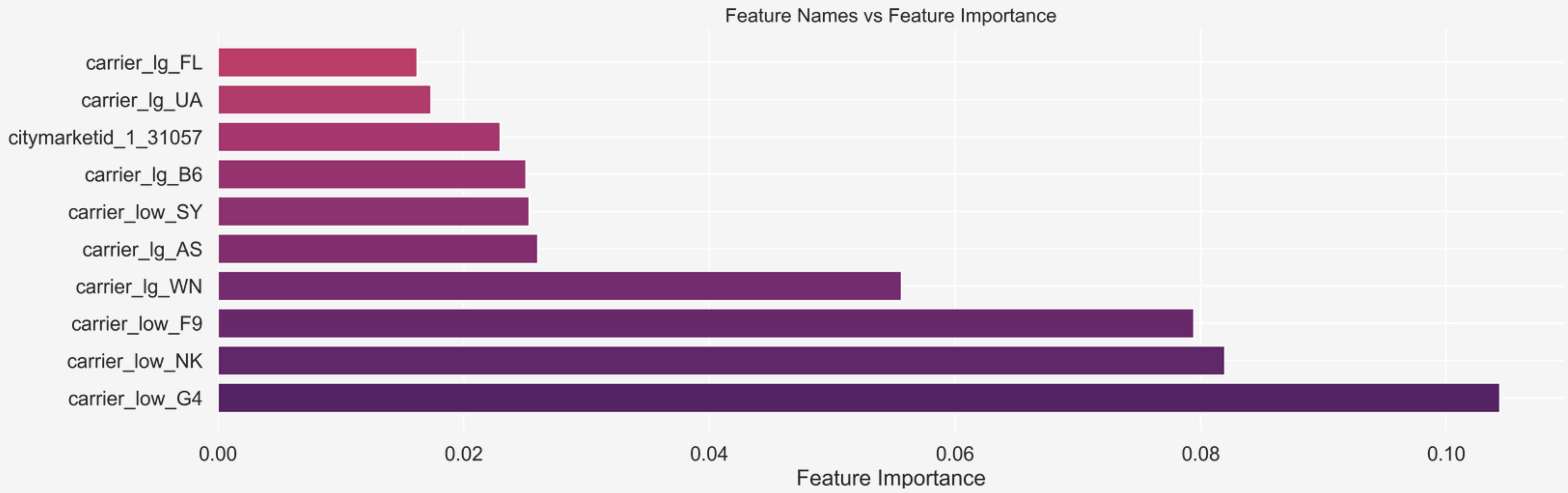




APPENDIX

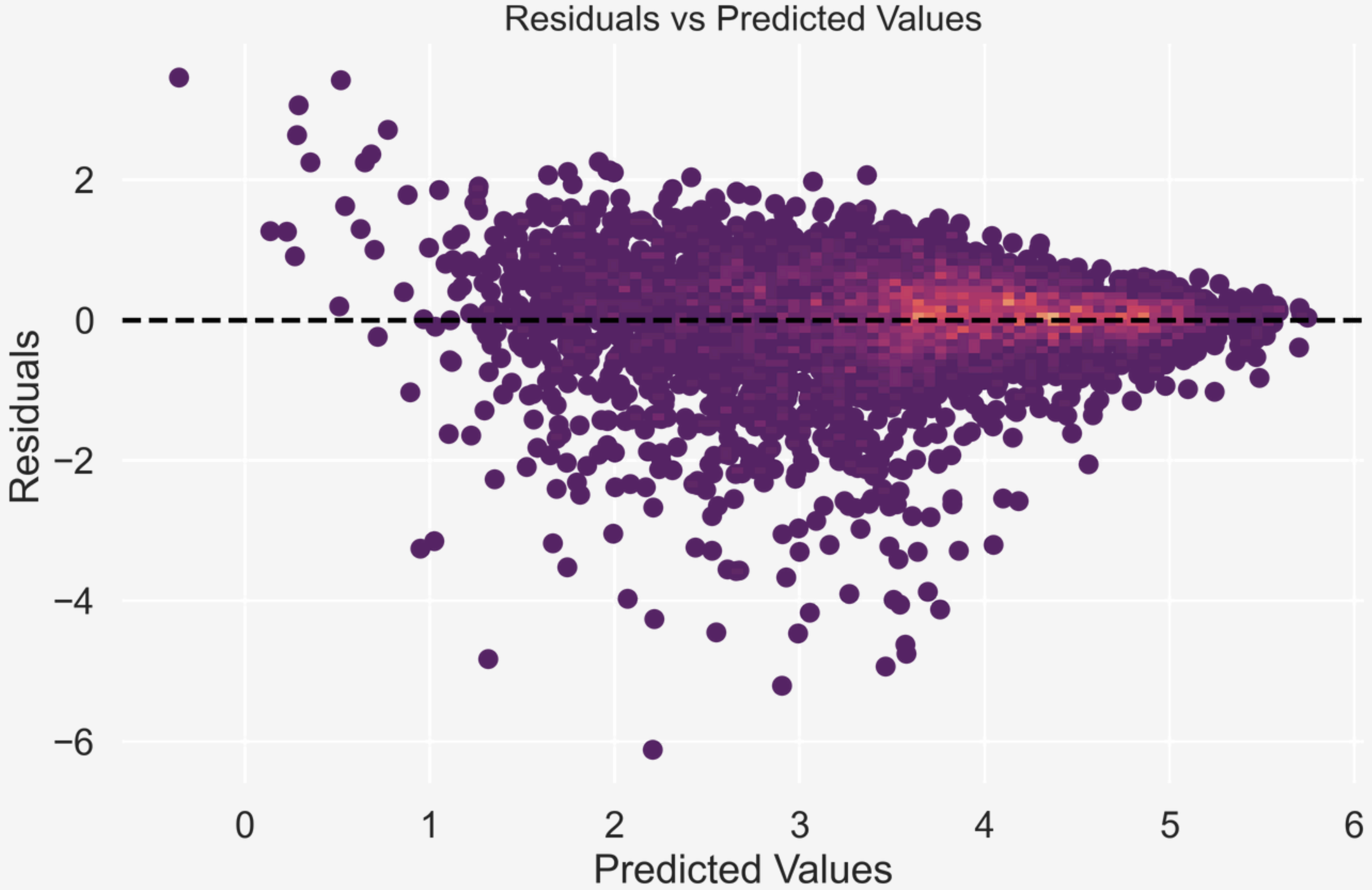
Model 2.2: XGBoost

- all xgboost library defaults



Metrics

- Train R^2: 0.9426144065314304
- Test R^2: 0.5933991979067752
- Train MSE: 0.07987078591323754
- Test MSE: 0.5727931963341509
- Train RMSE: 0.2826141997283087
- Test RMSE: 0.756831022312214



APPENDIX

Model 2.2: XGBoost (Cont.)