



Decision Tree Classifier for Market Share Prediction



Spirit Airline's Financials

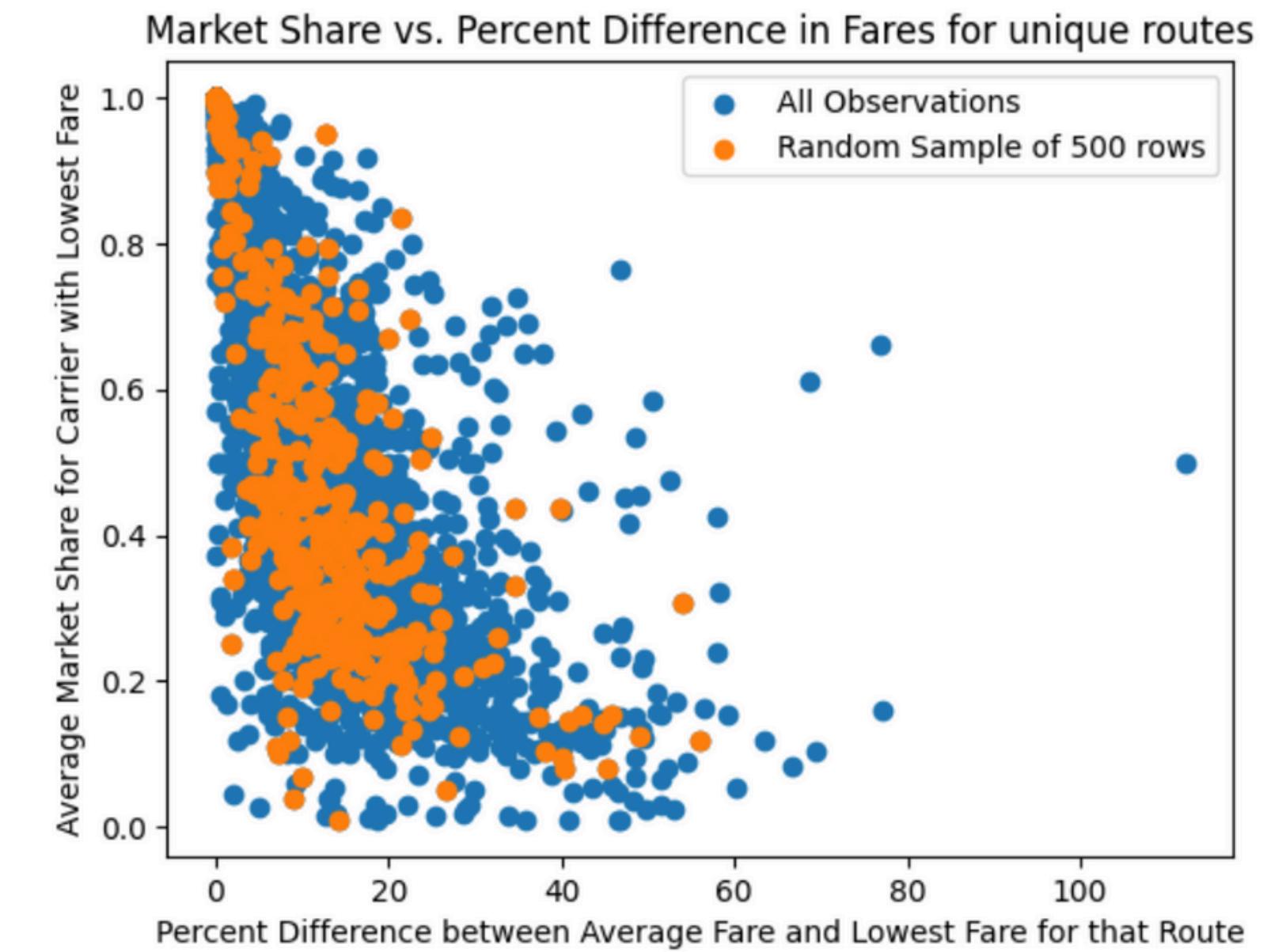
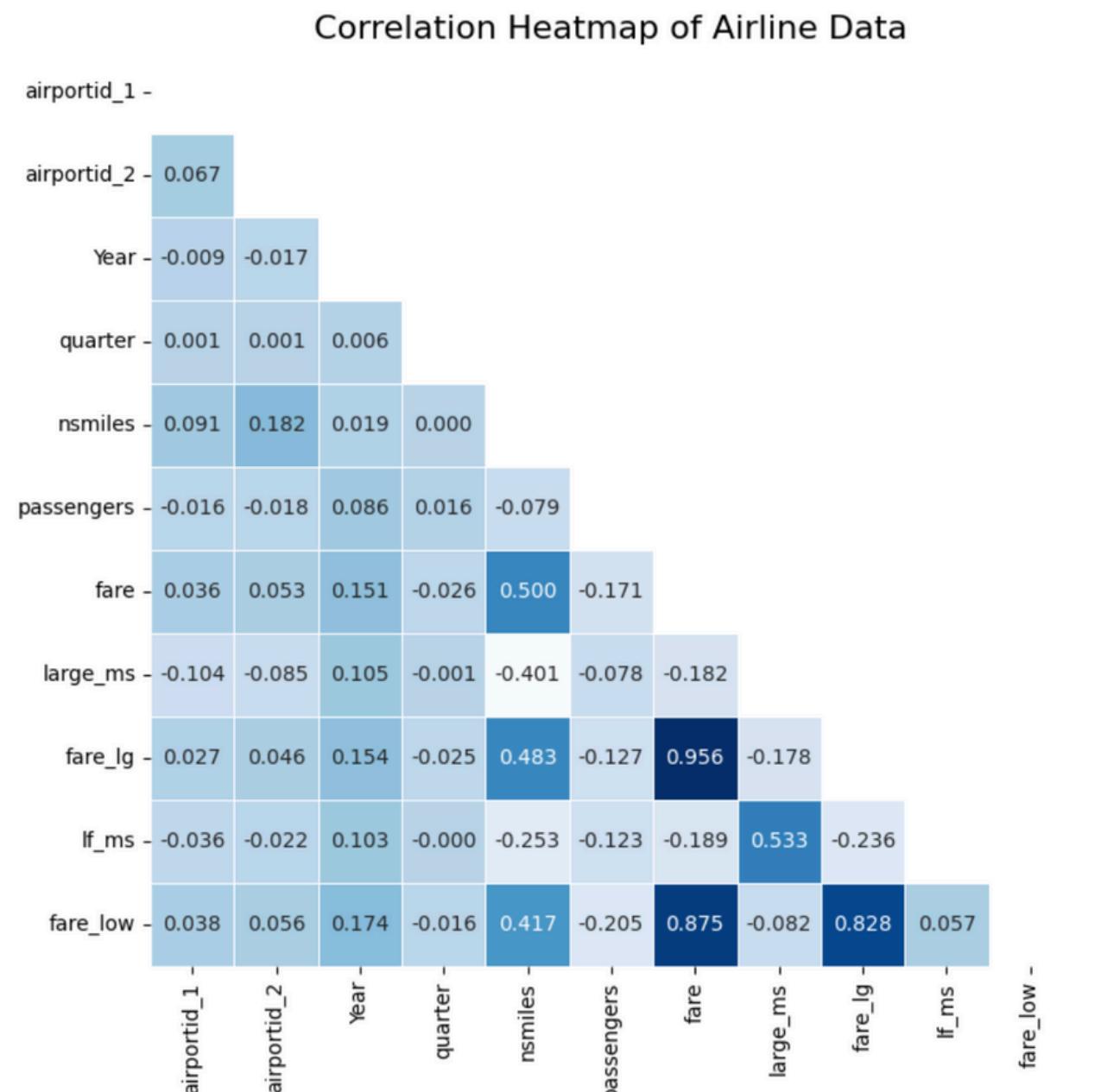
- Low profits
 - 5/6 last quarters were reported as losses
- Stock prices dropped by 83%
- Failed merger with Jet Blue
- 2024 Q2:
 - Net Income (losses): (192, 927)
 - Total Operating expenses: 1,433,401
 - Earnings per share (loss) : \$(1.76)

Spirit Airline's 2024 Q2 Financial summary

Second Quarter 2024 Financial Results

	(unaudited)	
	<u>As Reported</u>	<u>Adjusted¹</u>
Total operating revenues	\$1,280.9 million	\$1,280.9 million
Operating income (loss)	\$(152.5) million	\$(166.9) million
Operating margin	(11.9)%	(13.0)%
Net income (loss)	\$(192.9) million	\$(157.9) million
Diluted earnings (loss) per share	\$(1.76)	\$(1.44)

Exploratory Data Analysis



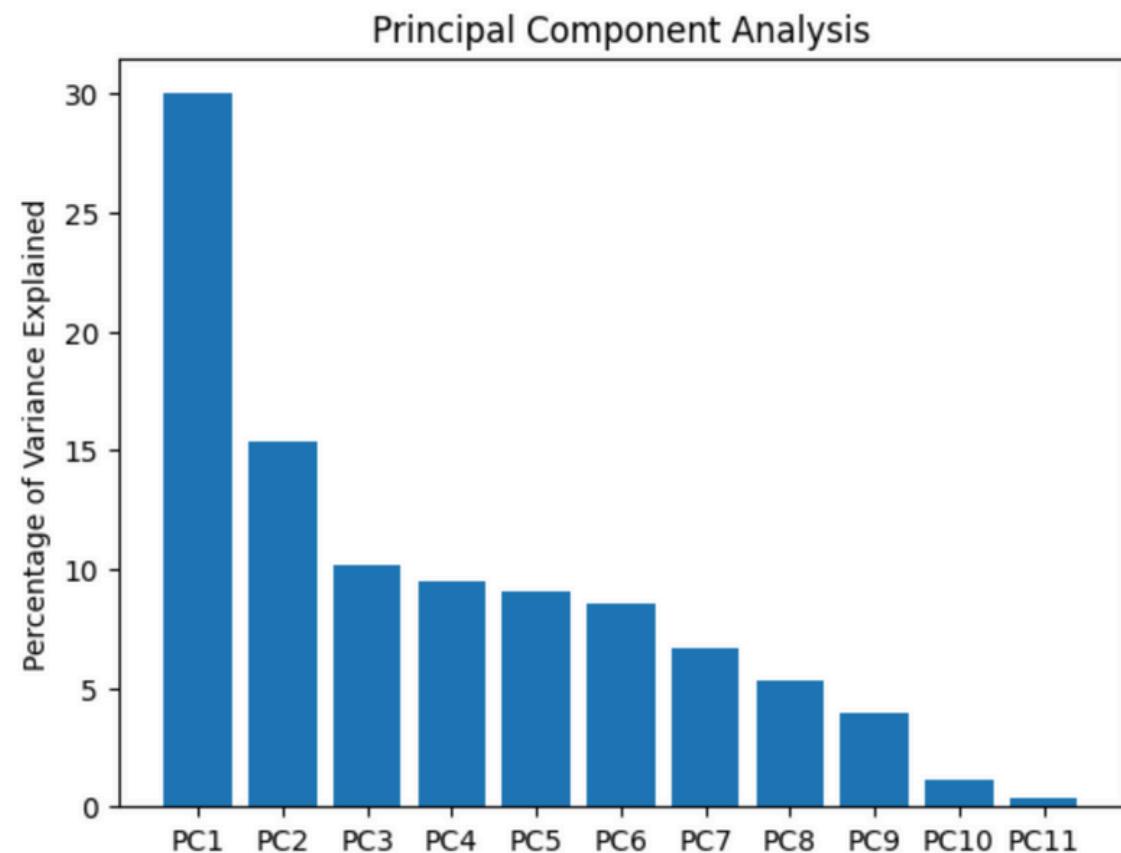


Goal & Process

Is there a way to determine which routes have the most **opportunity** for Spirit Airlines to hold a sufficient market share? Can we optimize which routes to fly to optimize revenues?

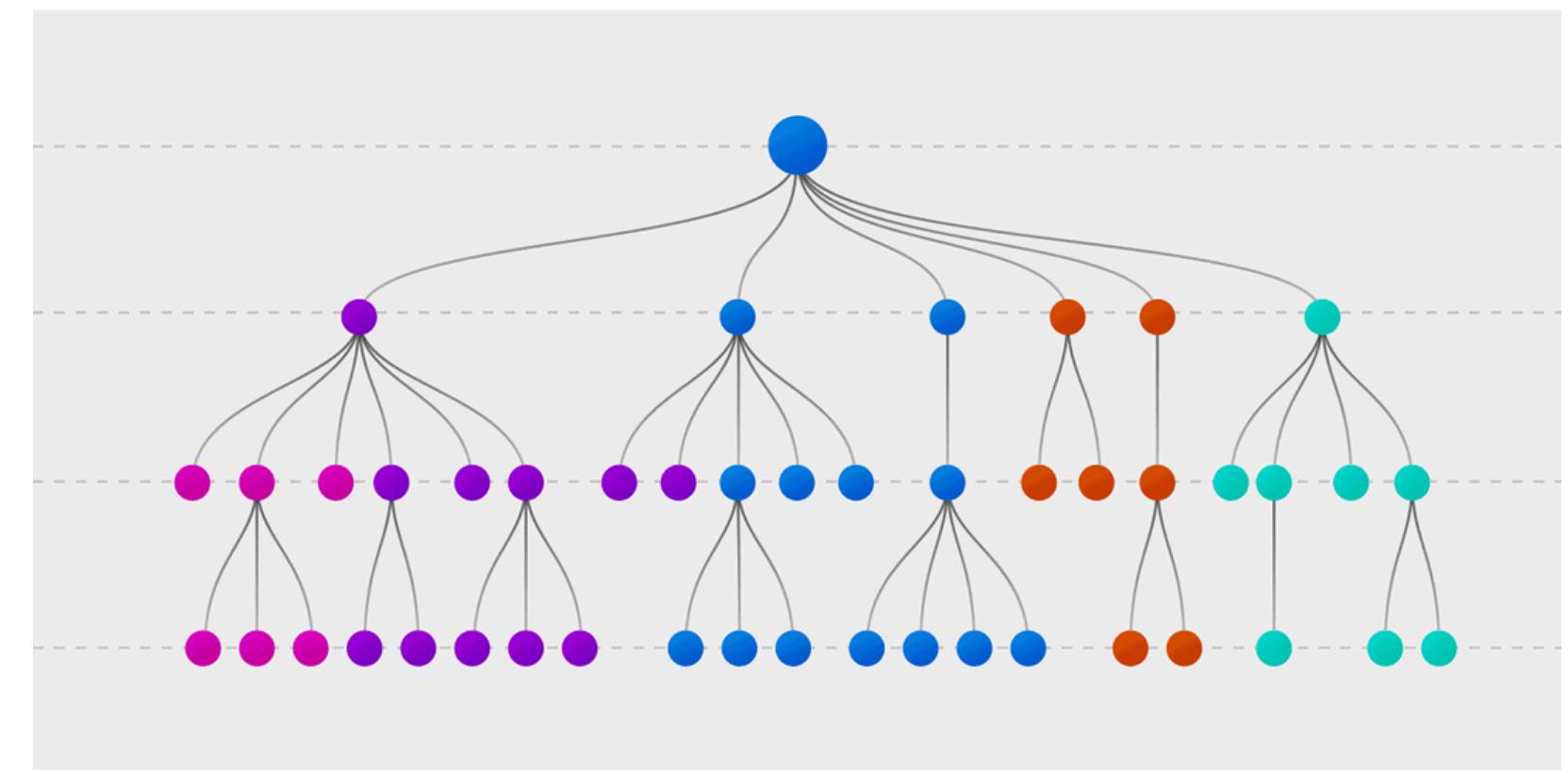
Allows smaller airline companies like Spirit to compete against larger companies in certain routes.

Is our data linear or non-linear in structure?



A linear technique used traditionally for dimensionality reduction, but whose performance can tell you if your data is linear or non-linear in structure.

Could we predict the market share of the largest carrier?



A Decision Tree Classifier: a supervised machine learning model that uses a tree-like structure to make decisions based on input features, splitting data into subsets to classify or predict outcomes.

Model and Results

```
#Grouping 'large_ms' into 'Low', 'Medium', 'High' classifications.  
bins = [0, 0.3, 0.6, 1]  
groups = ['Low', 'Medium', 'High']  
#  
correlated_variables['large_ms_groupings'] = pd.cut(correlated_variables['large_ms'], bins, labels = groups)
```

```
from sklearn.model_selection import KFold, train_test_split, cross_val_score  
from sklearn.tree import DecisionTreeClassifier  
from sklearn import tree  
#  
features = correlated_variables.drop(['large_ms', 'large_ms_groupings'], axis=1)  
target = correlated_variables['large_ms_groupings']  
#
```

```
X_train, X_test, y_train, y_test = train_test_split(features, target, test_size=0.3, random_state=30)
```

```
decision_tree = DecisionTreeClassifier()  
decision_tree = decision_tree.fit(X_train, y_train)  
#  
predictions = decision_tree.predict(X_test)
```

```
accuracy = sum(predictions == y_test)/(len(y_test))  
accuracy
```

```
0.8820512120922746
```

Model Cross-Validation

```
kfold = KFold(n_splits=5, shuffle=True, random_state=30)  
scores = cross_val_score(decision_tree, features, target, scoring='accuracy', cv =kfold)  
scores
```

```
array([0.88501913, 0.88399599, 0.88168368, 0.88571253, 0.88687894])
```



Impact

The model predicts to a high degree of accuracy whether the market share of the largest airline carrier will be high, medium, or low for each route.

Notably, when the market share is identified as low (or even medium), it presents a strategic opportunity for Spirit to enter the market.