Applied Data Science
Southeast Airlines NLP

Statistical Analysis using R

Funmi Esuruoso

Introduction

- Tasked with transforming, analyzing, and creating visualizations for the Southeast Airlines dataset containing 194,833 observations, my goal was to answer the “big question” of what factors significantly affect NPS score and what can be done by the airline to increase this score?
- Some tools and techniques used for this project are as follows:
  - Geomaps
  - Barcharts
  - Histogram
  - Confusion Matrix
  - Two-Way Frequency Matrix
  - Regression Analysis

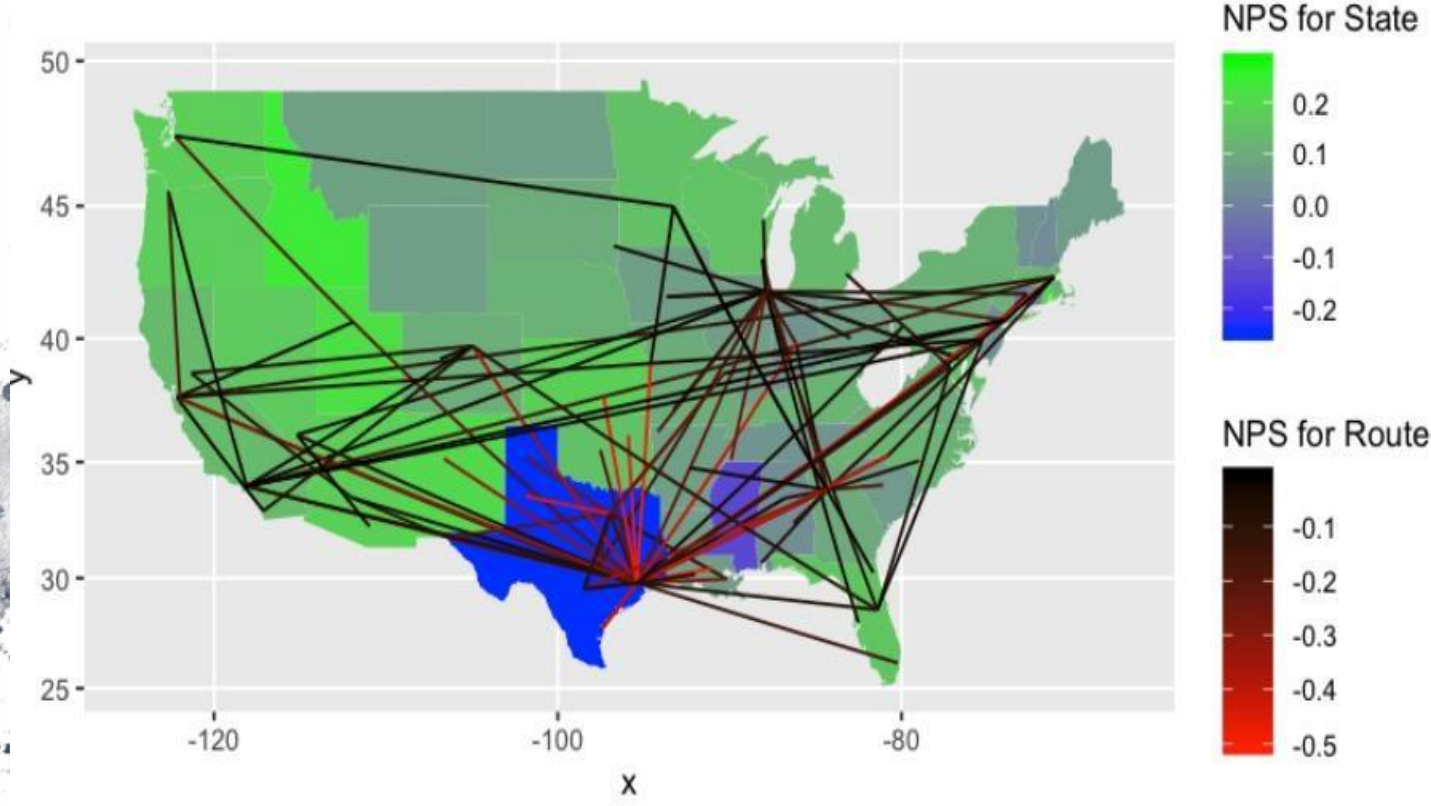
Exploratory

Map View:

- Origin.State
- Routes between any two (2) cities

Key Takeaways:

- Texas State has the worst average NPS score
- The Houston Airport received a significant amount of customer complaints



Modeling

Two-way Frequency Table:

- Examine the relationship between Gender & Detractor.
- Examine the relationship between Airline.Status & Detractor

Key Takeaways:

- Females have a higher probability of being detractors than males
- Flyers in the “Blue” Airline Status have a higher probability of being detractors than any other flyer

Gender	Female	Male	Sum
Detractor			
FALSE	0.38	0.32	0.70
TRUE	0.18	0.11	0.30
Sum	0.56	0.44	1.00

Airline.Status	Blue	Gold	Platinum	Silver	Sum
Detractor					
FALSE	0.43	0.07	0.02	0.19	0.70
TRUE	0.25	0.02	0.01	0.01	0.30
Sum	0.68	0.08	0.03	0.20	1.00

SVM Classifier

Objectives:

- Understand how the following variables affect who is a detractor
  - Age
  - Airline.Status
  - Class
  - Eating.and.Drinking.at.Airport

Key Takeaways:

- Airline.Status & Age take priority
- SVM doesn't work well for categorical data

Confusion Matrix:

Accuracy	0.7521
95% CI	(0.7491, 0.755)
No Information Rate :	0.7061
P-Value [Acc > NIR] : <	2.2e-16

Feature Importance:

1. Airline.Status: 100.00
2. Age: 56.122
3. Eating.and.Drinking.at.Airport: 5.826

Random Forest Classifier

Objectives:

- Understand how the following variables affect who is a detractor
  - Age.bin, Airline.Status, Class
  - Origin.State, Origin.City
  - Eating.and.Drinking.at.Airport.bin
  - Partner.Name

Key Takeaways:

- Works well for both categorical and numeric data

Confusion Matrix:

Accuracy	0.7851
95% CI	(0.7791, 0.7821)
No Information Rate :	0.7061
P-Value [Acc > NIR] : <	2.2e-16

Feature Importances: (descended order)

1. Airline.Status
2. Eating.and.Drinking.at.Airport.bin 50~100
3. Origin.State Texas
4. Origin.City Houston, TX
5. Age.bin 30~45
6. Partner.Name FlyFast Airways Inc.

Xgboost Classifier

Objectives:

- Understand how the following variables concern to whom became a detractor
  - Age.bin, Airline.Status, Class
  - Origin.State, Origin.City
  - Eating.and.Drinking.at.Airport.bin
  - Partner.Name

Key Takeaways:

- Need to cut the numeric value to bins, then one-hot encoding the matrices
- Better Accuracy

Confusion Matrix:

Accuracy	0.8221
95% CI	(0.7991, 0.815)
No Information Rate :	0.7061
P-Value [Acc > NIR] : <	2.2e-16

Feature Importances: (descended order)

1. Airline.Status.Blue
2. Airline.Status.Silver
3. Age.bin.0~18
4. Eating.and.Drinking.at.Airport.bin
5. Partner.Name.FlyFast Airways Inc
6. Origin.State.Texas

Recommendation No. 1

Food and drink prices matter to flyer's trip satisfaction

- Manage the price to ensure everyone's consumption of food and drink falls into the range of $25 to $50 at airport:
  - Provide more choices for quick or instant food & drink
  - Reduce high end & expensive restaurants
  - Price regulations & regular checks at the airport

Evidence Support:

- 95% of the passengers buy food and drink at airport
- $25 ~ $50 proves to have the best value for the price, which helps lift the overall trip satisfaction
- Passengers who spent over $150 for food & drinks are not satisfied with their trips
- Top three (3) most expensive airports for their food and drink prices produce the worst customer satisfaction.
  - Killeen, TX, Houston, TX, San Antonio, TX

Recommendation No. 2

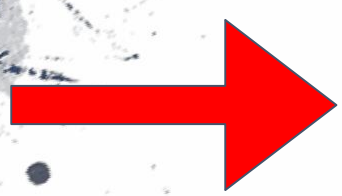
Assess the packages for each tier, with considerable revisions made to the Blue Airline Status package

- Southeast has implemented a three (3) tier Airline Status system for their customers which are platinum, gold, silver, and blue. The Blue Airline status has a 25% detractor value. Southeast Airline should assess the tier's package and revise the Blue Airline Status's package.
- Should Offer more points, complimentary same-day flight changes or complimentary upgrades to the "Blue" flyers.

Recommendation No. 3

Reevaluate FlyFast Airlines Partnership and other “Low Performers”

- **FlyFast Airlines**: Worst customer satisfaction among our partners — 40% of all flyers are "Detractors"
- **Options for Immediate action:**
  - Renegotiate contract: performance-based terms/NPS as a KPI
  - Form intercompany team to develop performance improvement plan
  - Evaluate whether other partners flying out of FlyFast hub can pick-up additional flights if we terminate FlyFast contract
  - Contract "probation"
  - Termination of contract



Only 25% of FlyFast flyers rate as “promoters” — FlyFast produces the highest level of “Detractors” and the fewest “Promoters”

Recommendation No. 4

Segment flyers based on “Age” to boost airport experience

- **Age**
  - The average Likelihood to recommend score (under 7) are for the ages 15-20 and 60 and older.
- **Ages: 15-20**
  - Create an arcade in the waiting area to improve the LTR score. This allows the customers to be occupied during long wait or layover times. Also, if a flight is delayed the airlines could give out vouchers to the passengers as they wait.
- **Ages: 60+**
  - Create an option that provides an assistance to passengers over 60 to navigate the airport from arrival to departure. The assistant will assist the passenger with luggage, transport to the appropriate terminal and answer any questions order concerns. The assistant will be a personal travel guide while at the airport.