ANALYSIS OF REUSABLE ROCKETS

Noah McMahon

11-21-2023

Outline

EXECUTIVE SUMMARY

INTRODUCTION

METHODOLOGY

RESULTS

CONCLUSION

APPENDIX

Executive Summary

Methodologies Used in this Analysis

Data Collection by web scraping and usage of Spacex API

Exploratory Data Analysis(EDA)

- Includes data wrangling, visualization, and interactive visual analytics

Machine Learning Prediction

Summary of all results

Collecting public data was possible

EDA methods found which features are best for predicting landing success

Machine Learning provided the best model for predicting impactful characteristics on outcome

Introduction

Objective of Analysis

To evaluate the possibility for new company SpaceY to compete with SpaceX

Questions to Solve

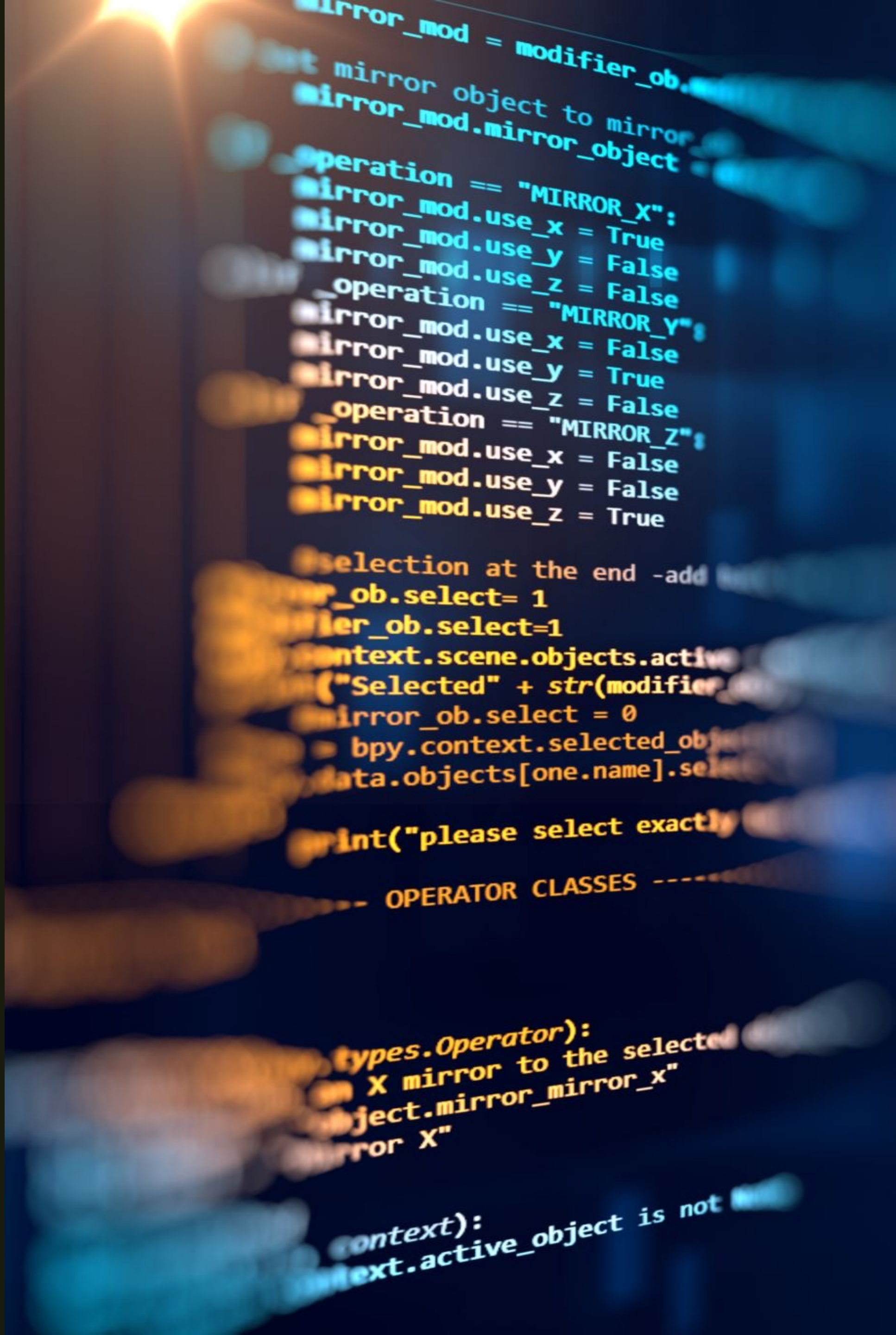
Where is the best place to have launches?

What is the best way to predict successful launches and thus estimate the total cost of the launch?

METHODOLOGY

Data Collection

- Sources used to obtain SpaceX data
  - *SpaceX API - "https://api.spacexdata.com/v4/launches/past"*
  - *Webscraping – https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922*
- SpaceX API
  - *Used data wrangling and formatting methods to create usable data*
- Webscraping
  - *Took data from online website (Wikipedia) and created a data frame to be used for analysis*



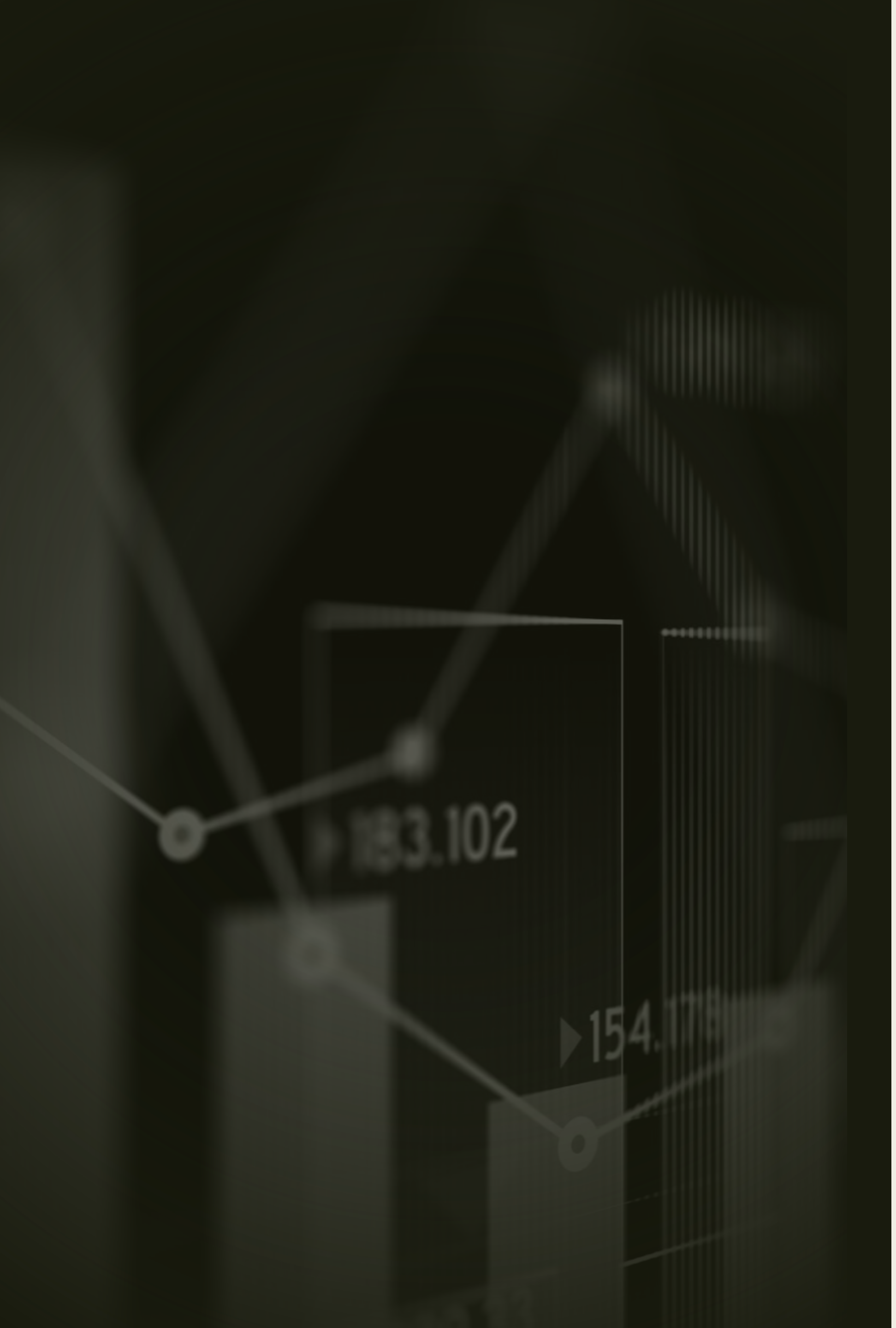
Data Wrangling

Initial EDA was performed on the collected data

Several data points were found including:

- *The number of launches at each site*
- *The number and occurrence of each orbit*
- *The number and occurrence of mission outcomes of orbits*

A landing outcome label was then created to organize the mission outcomes



EDA With SQL

- SQL allowed for further analysis to find:
  - *Commonalities between features in the data frame*
  - *Calculations such as Total Payload Mass for certain customers and booster versions*
  - *Which boosters had the most success with certain features*
  - *Spread of outcome type across time spans*

EDA With Visualization

- Used multiple types of charts and plots to see patterns in the data
  - *Scatter plots included:*
    - Flight Number vs. Payload Mass
    - Flight Number vs. Launch Site
    - Payload Mass vs. Launch Site
  - *Bar Graph included:*
    - Success rate by orbit type
  - *Line Graph included:*
    - Launch Success Rate as a yearly trend line

Visualization with Folium

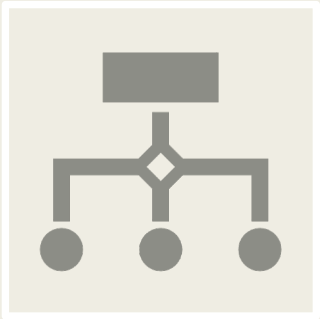
- Created a map with multiple features including:
  - *All launch site locations*
  - *Successful launch outcomes and failures per launch site*
    - Created clusters of these marks on the map
  - *Calculated distances between launch sites and their proximities*
    - Proximities included:
      - *Highway*
      - *Coastline*
      - *Railway*

Dashboard by Plotly Dash

Created two charts on an interactive dashboard

Pie chart displaying Percentage of launches by site

Scatter plot displaying the launches by payload in a specified range



The interactive dashboard allows for quick adjustments to be made to the data and filtering to be applied on the spot

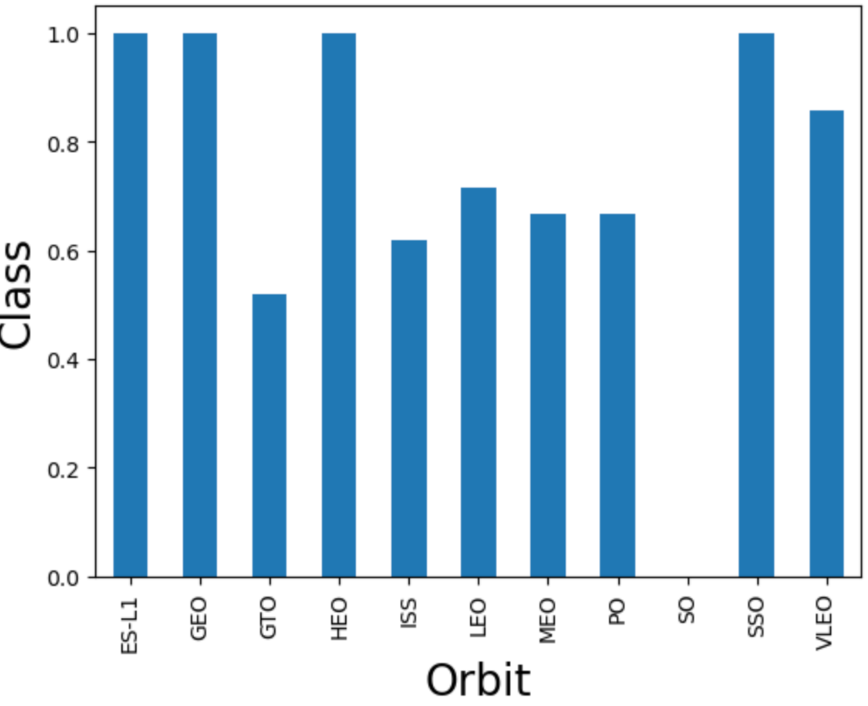
Machine Learning

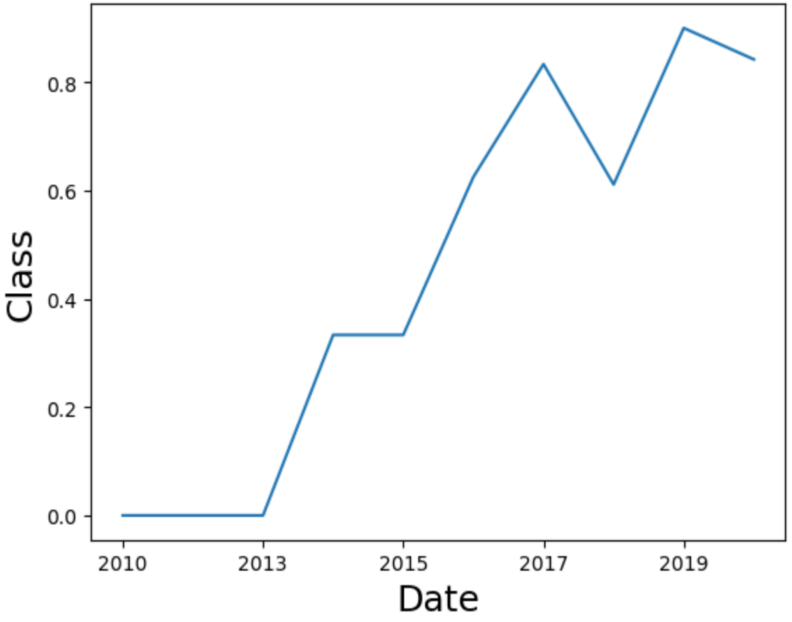
Used multiple Classification techniques including:

- *Logistic Regression*
- *Support Vector Machine*
- *Decision Tree Classifier*
- *K Nearest Neighbors*

Each model type was trained and tested with prepared and standardized data. The results were then compared

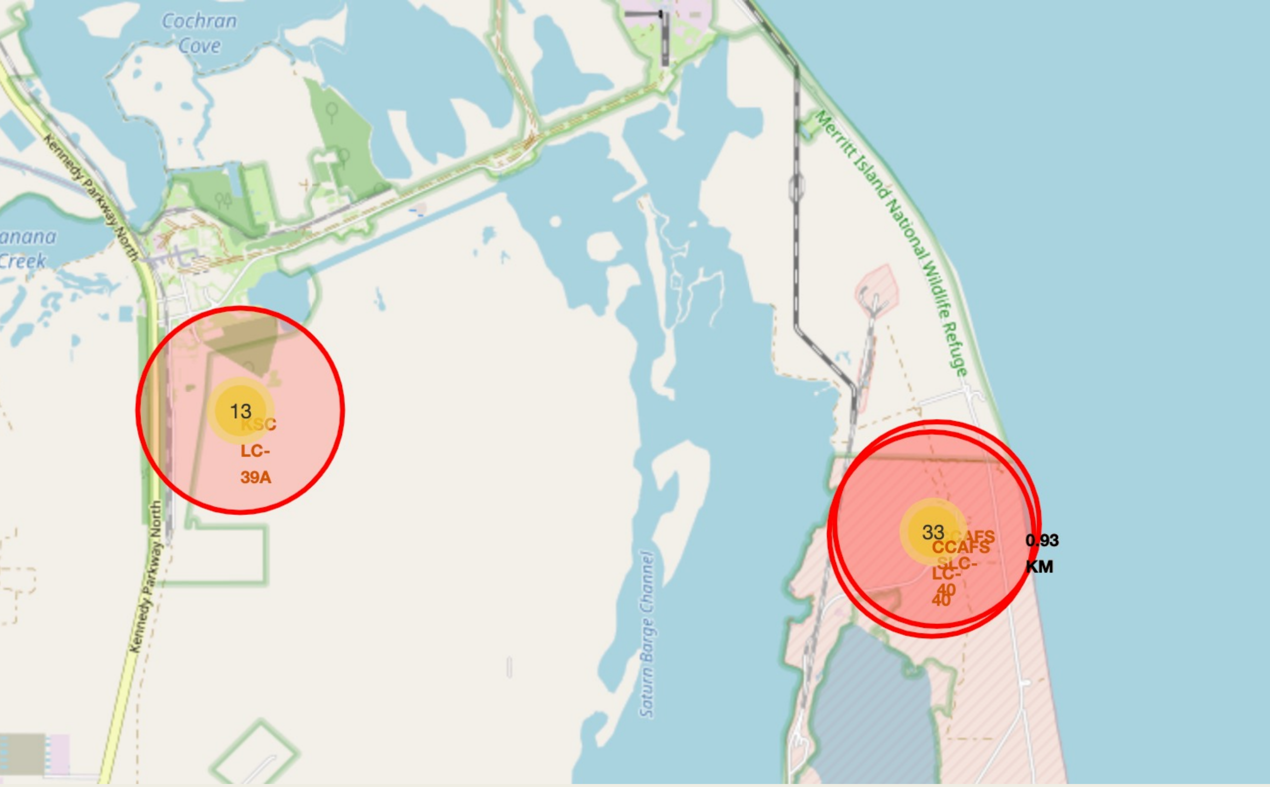
RESULTS





EDA Results

- Initial Findings
  - *4 launch sites*
  - *Average payload mass for F9 v1.1 booster is 2,928.4 kg*
  - *The first successful landing was in 2015*
  - *There has been 99 successful mission outcomes, 1 failure in flight and 1*
- Further Analysis found
  - *Some orbits succeed far more than others*
    - GTO orbits only succeed around half the time
    - Meanwhile ES-L1 and GEO orbits have never failed
  - *The average success rate has increased as years have passed between 2013 and 2020*

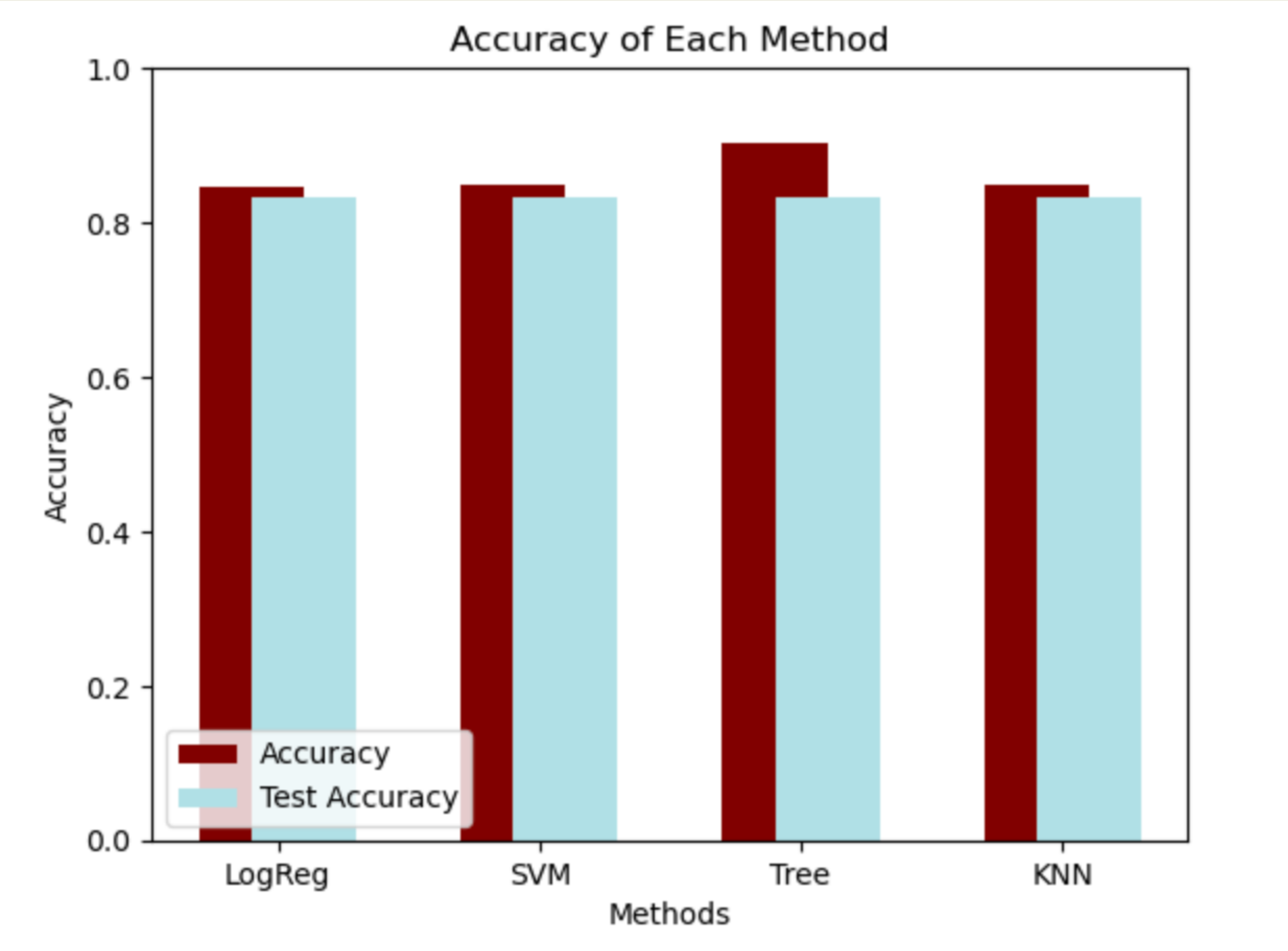


Folium Results

- The generated maps provided numerous key insights
  - *Most launches occurred on the east coast*
  - *Launch Sites are close to the coastline and are a considerable distance from nearby cities, i.e. safer locations*
  - *All sites have similar latitudes i.e. distances from equator*

Plotly Dash Results

- Most of the Payload Masses were in the 2,000kg to 6,000kg range
- The v1.1 failed the most often
- KSC LC-39A in Florida had the largest percentage of launches
- Florida sites held the most launches at 83.3%



Machine Learning Results

- The Decision Tree model was most accurate
  - *Accuracy of 90.4% and Test Accuracy of 83.3%*
- The other models performed very similarly
  - *The Logistic Regression model was slightly worse*

Conclusion

- Through this analysis, a few important factors have been understood for a successful mission
  - *Having a mission with an orbit of ES-L1, GEO, HEO, or SSO have the greatest probability of success*
  - *Choosing a launch site on the east coast in a safe area*
    - Example: In Florida, near the coastline, away from cities, area with access to industrial transport like railways or highways
    - Best site is KSC LC-39A
  - *Keeping payload in the range of 2,000 kg to 6,000kg is safest*
  - *A Decision Tree Classifier is the best way to predict successful landings and improve company profits*


THANK YOU!