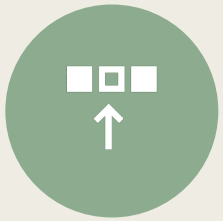
ANALYSIS OF REUSABLE ROCKETS

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?

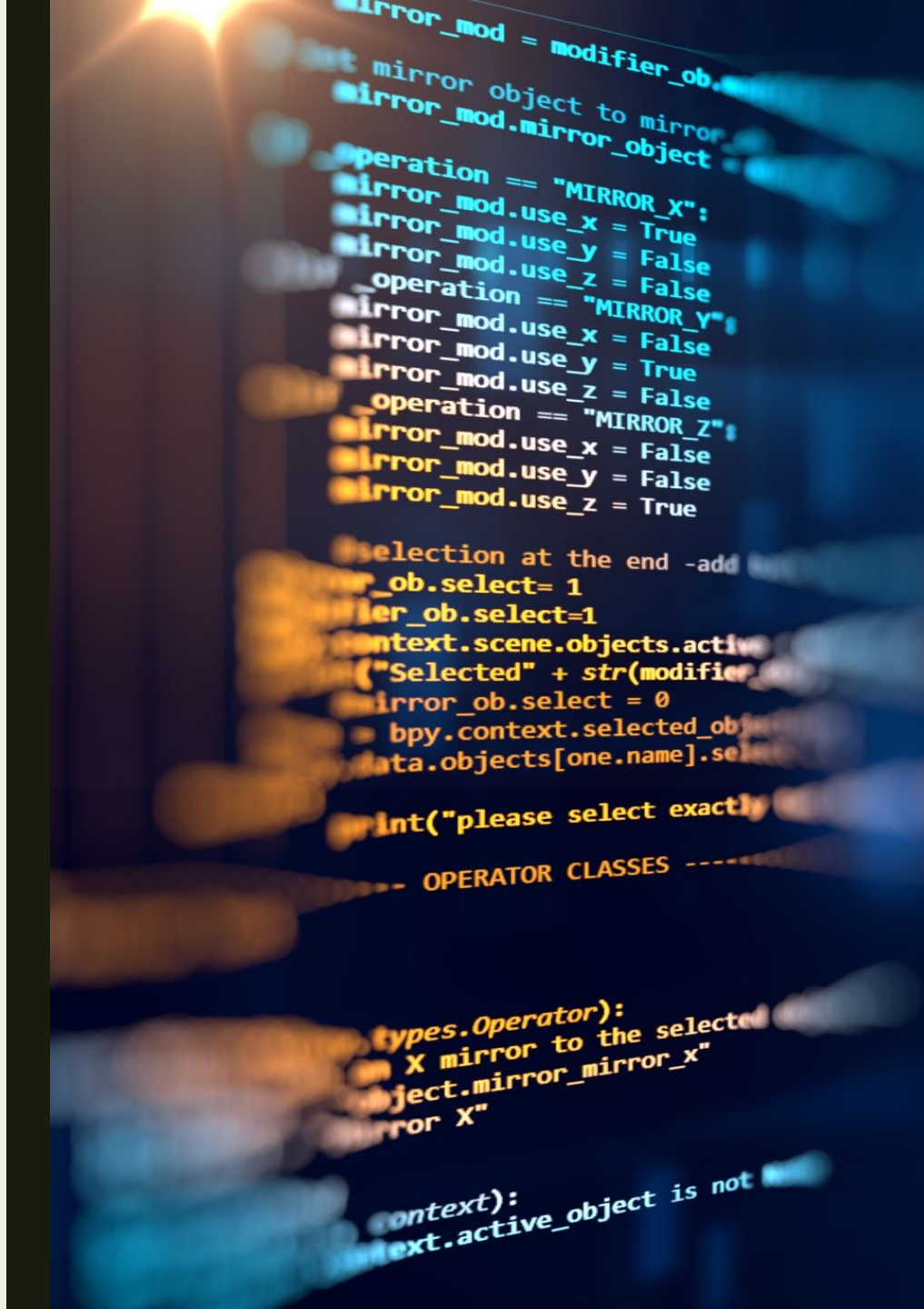
METHODOLOGY

Data Collection SpaceX API

- SpaceX API - "https://api.spacexdata.com/v4/launches/past"
- SpaceX API is a public API where data can be collected and used in analysis
- Process of using the API data is listed below

Request	Filter	Handle
Request the API and parse the data	Filter the data to only include Falcon 9 launches	Handle missing values

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/Data%20Collection%20API.ipynb

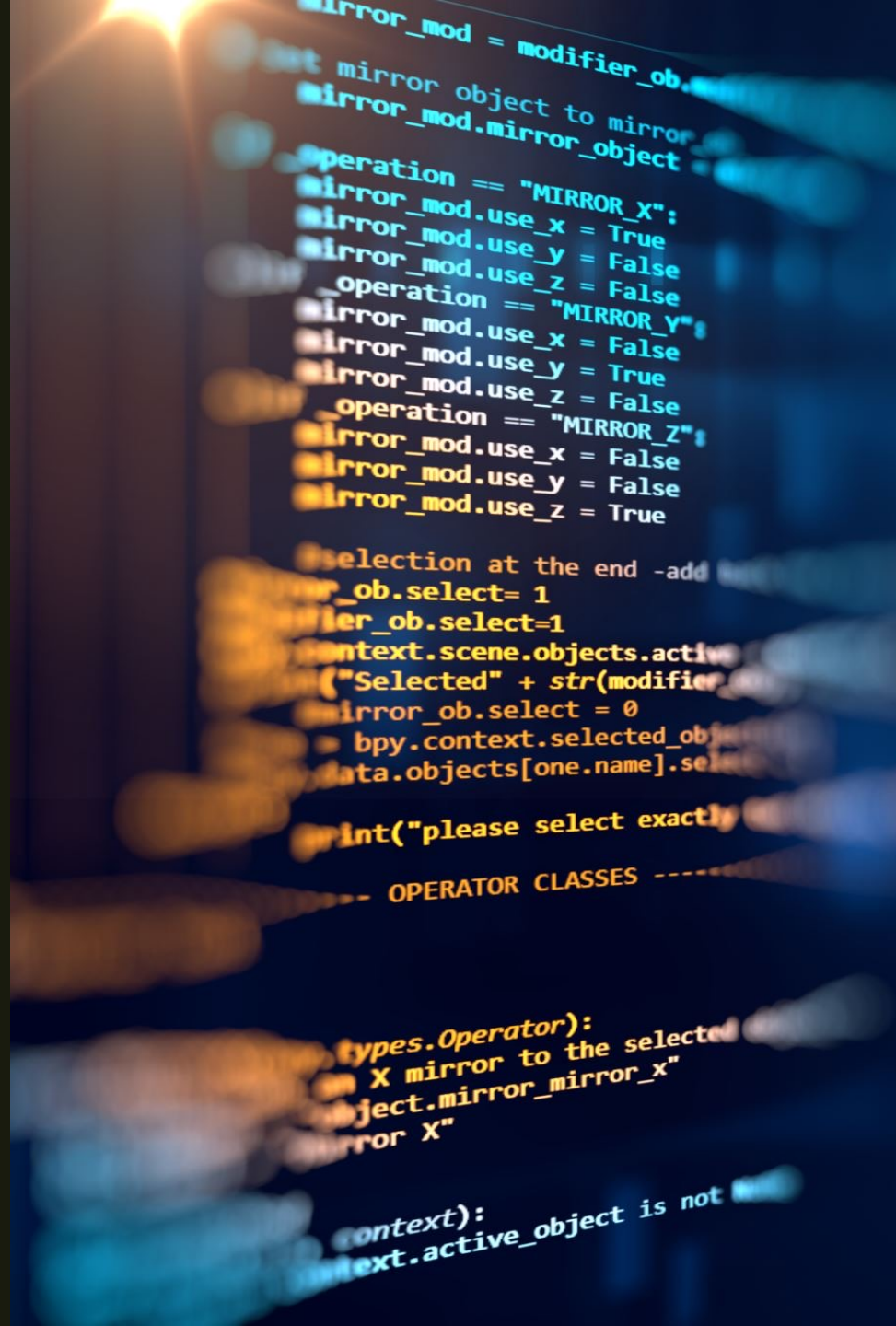


Data Collection Webscraping

- Data is obtained from Wikipedia
  - *https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922*”
- Data is collected according to graphic below

Request	Extract	Create
Request the Falcon9 Launch Wiki page	Extract column and variable names from HTML header	Parse the launch HTML tables to turn them into a dataframe

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/Data%20Collection%20with%20WebScraping.ipynb



Data Wrangling

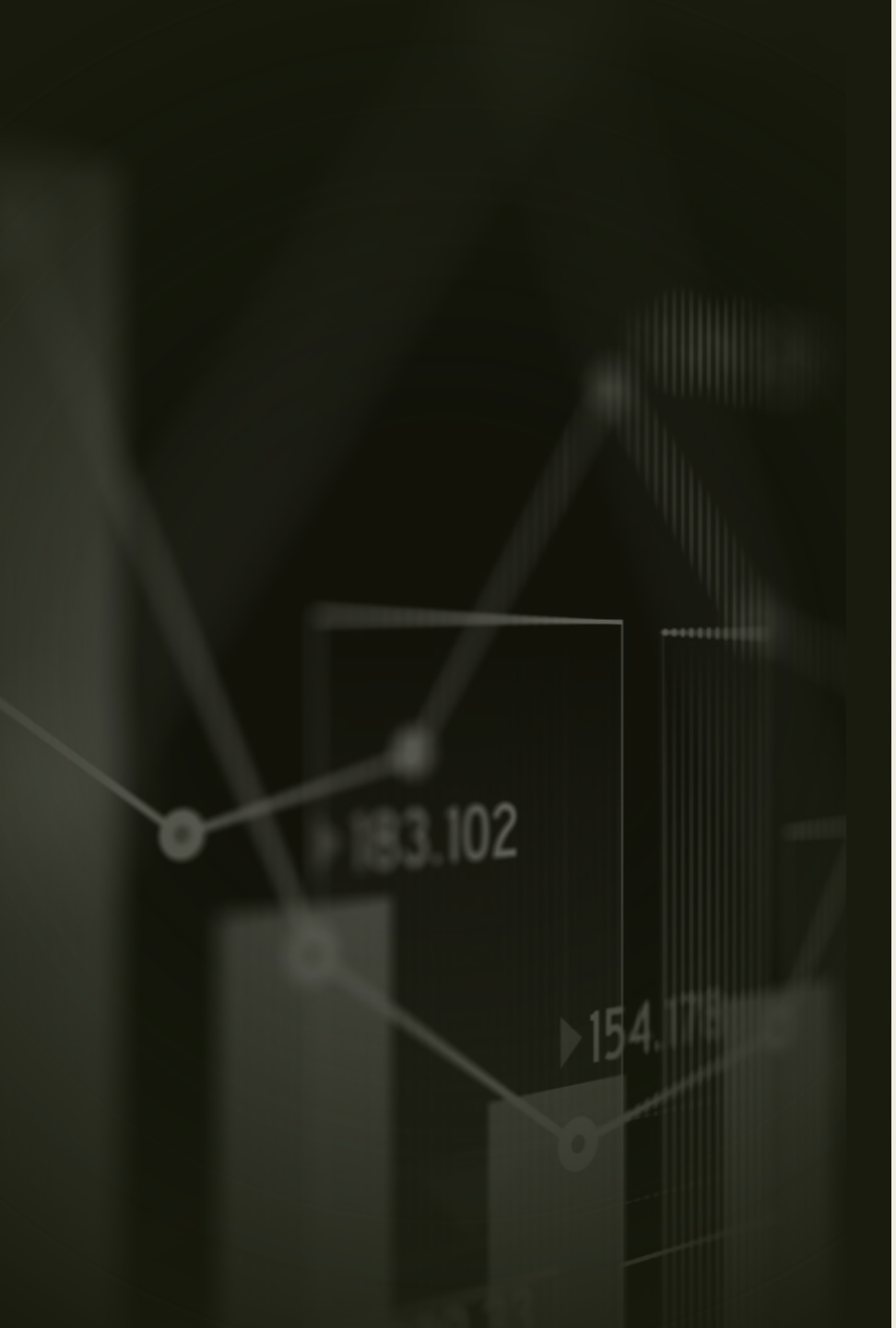
Initial EDA was performed on the collected data

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

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/Data%20Wrangling.ipynb



EDA With SQL

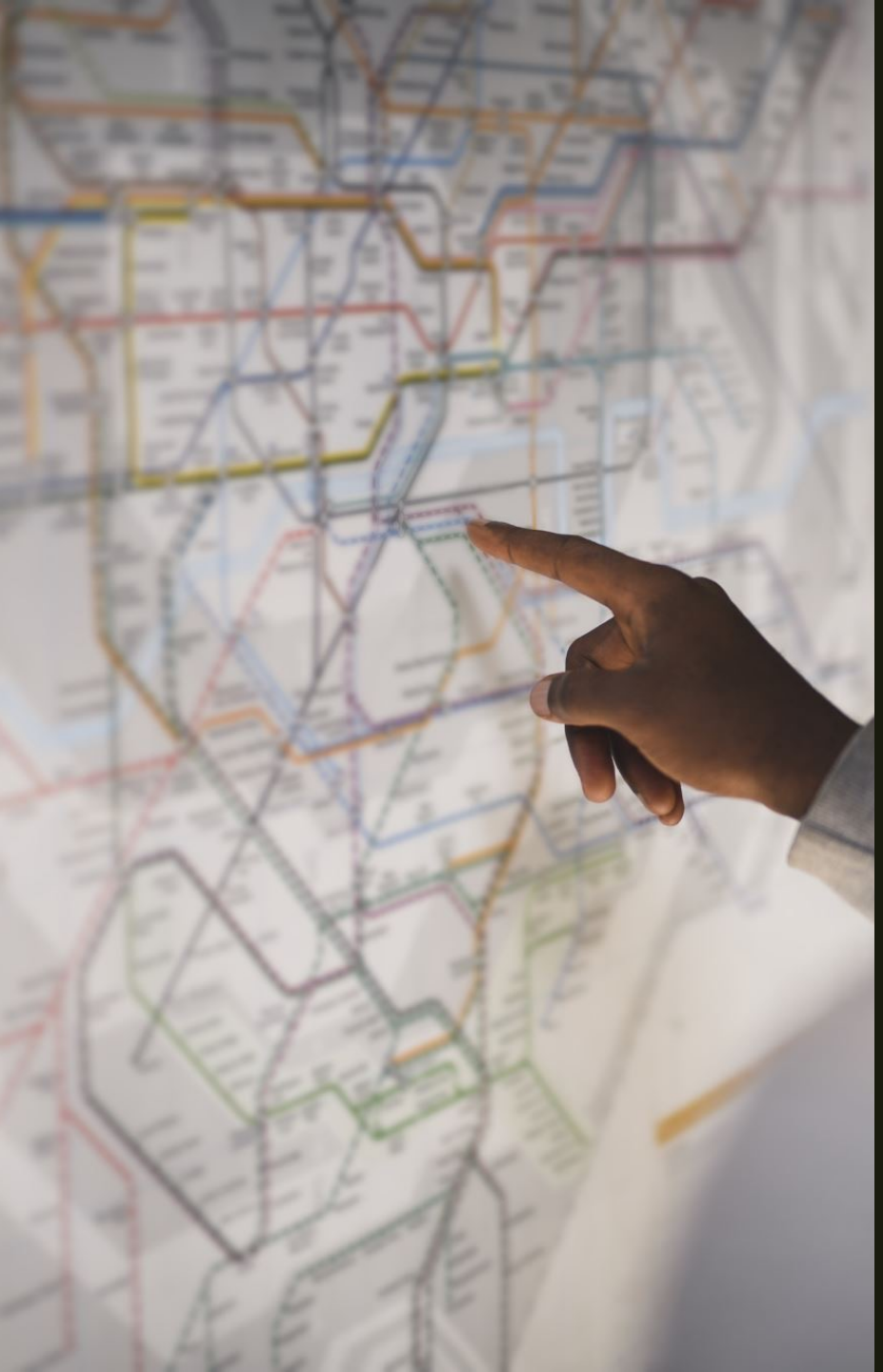
- Performed the following queries:
  - *Names of unique launch sites*
  - *Top 5 launch sites starting with 'CCA'*
  - *Total Payload Mass for NASA (CRS)*
  - *Average payload mass of booster F9 v1.1*
  - *Date of first successful landing outcome*
  - *Boosters with success in drone ship and have payload mass between 4000 and 6000kg*
  - *Total number of successful and failed outcomes*
  - *Boosters that have carried the max payload mass*
  - *Failed landing outcomes in drone ship with the booster versions and launch site name in 2015*
  - *Rank of the count of landing outcomes between 06-04-2010 and 03-20-2017*

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/EDAWithSQL.ipynb

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

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/EDA%20With%20Visualization.ipynb



Visualization with Folium

- Created a map with multiple features including:
  - *Used markers to signify specific points on the map*
    - Launch sites
  - *Circles represent specific highlighted areas around specific coordinates – ex) NASA Johnson Space Center*
  - *Marker clusters for groupings of events in each coordinate*
    - Used for launches at launch site
  - *Lines used for distances between proximities*
    - Roadways
    - Railways
    - Cities

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

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 to determine the best site to launch based on payload mass

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/Interactive%20Dashboard%20with%20Plotly%20Dash.py

Machine Learning

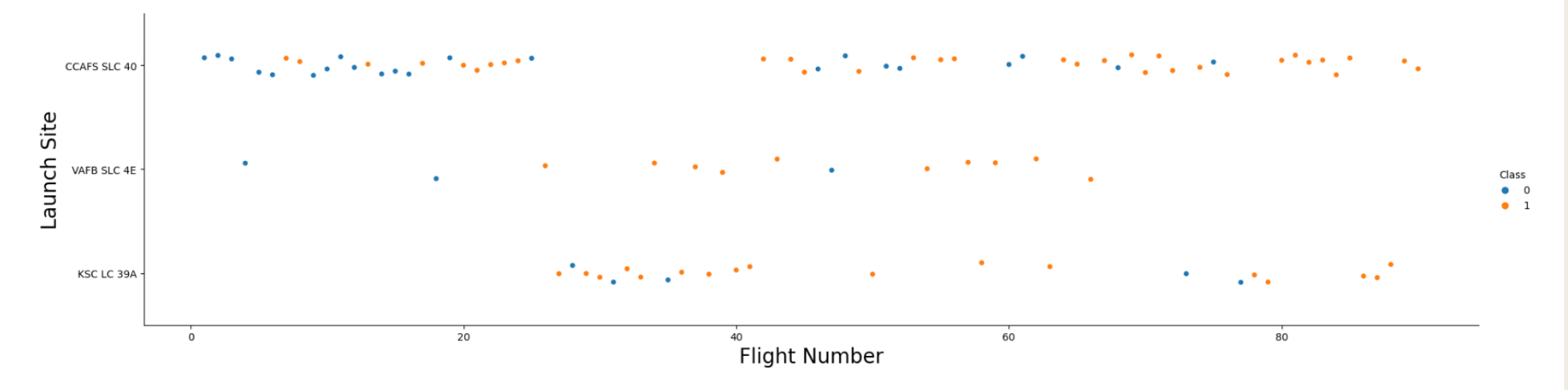
Used multiple Classification techniques including:

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

EDA WITH VISUALIZATION

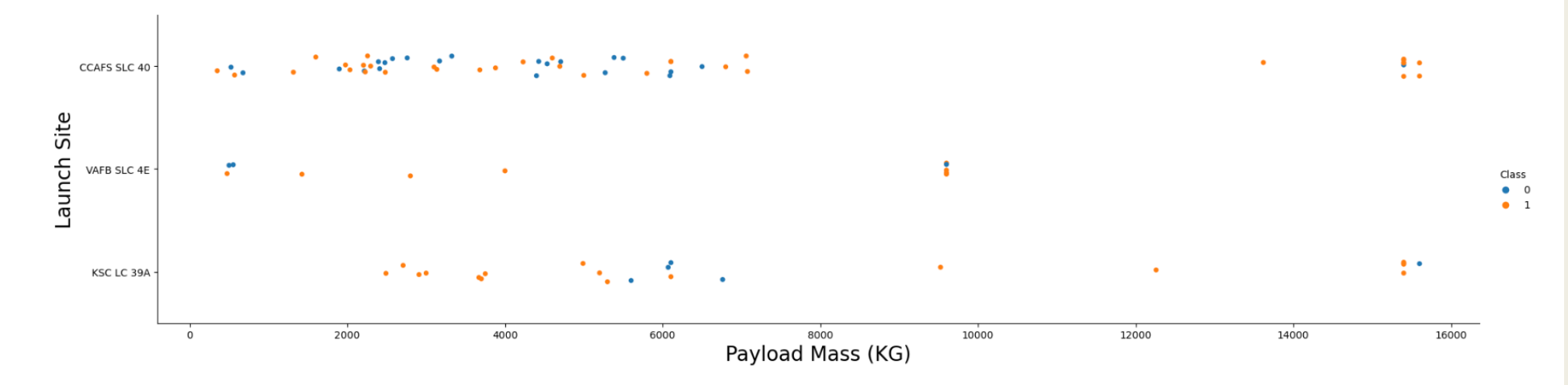
Flight Number vs. Launch Site



- The best launch site as of most recently is CCAFS SLC 40
  - *This site has the most successful launches of the most recently launched missions*
- The general success rate has improved over time as well

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/EDA%20With%20Visualization.ipynb

Payload Mass vs. Launch Site

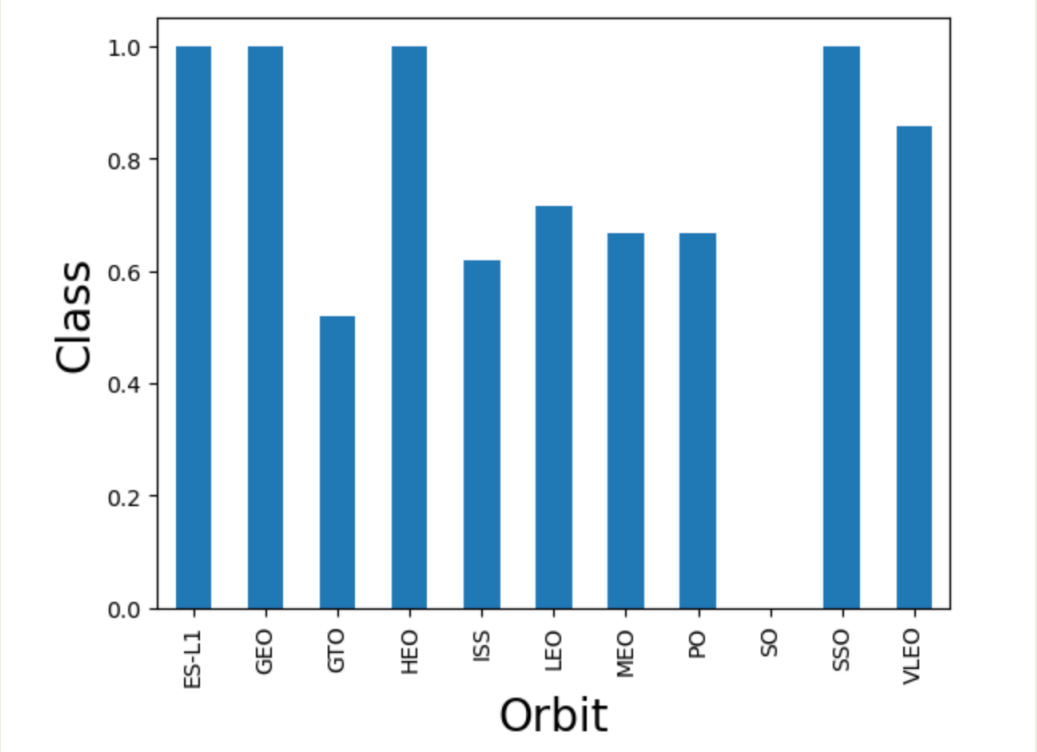


- The only launch sites capable of 12,000 kg + payloads are CCAFS SLC 40 and KSC LC 39A
- Launches at VAFB SLC 4E are relatively smaller masses for the most part

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/EDA%20With%20Visualization.ipynb

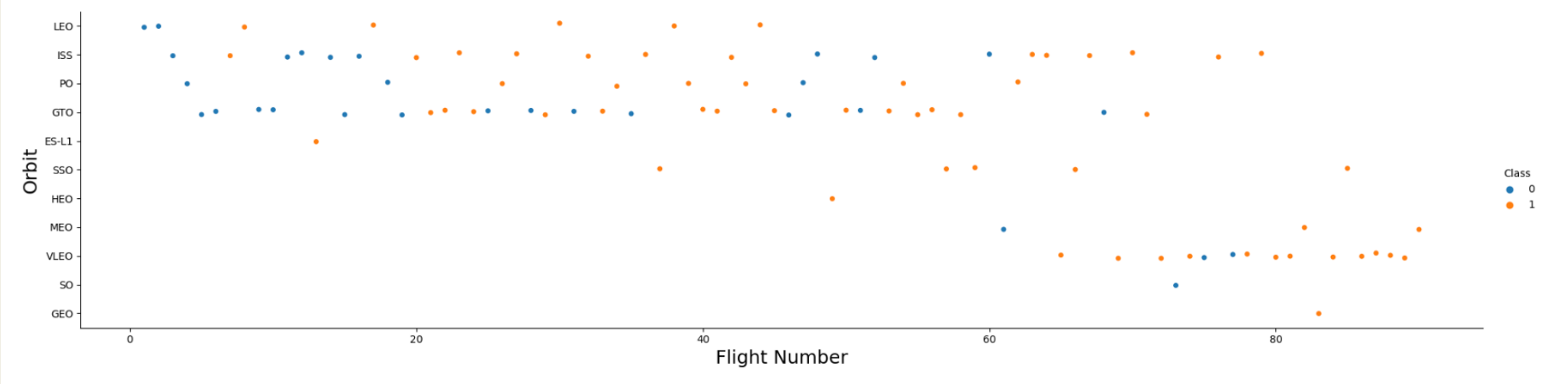
Success Rate vs. Orbit Type

- Some orbits do not have failures
  - *ES-L1, GEO, HEO, and SSO*
- The riskiest orbit appears to be GTO



https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/EDA%20With%20Visualization.ipynb

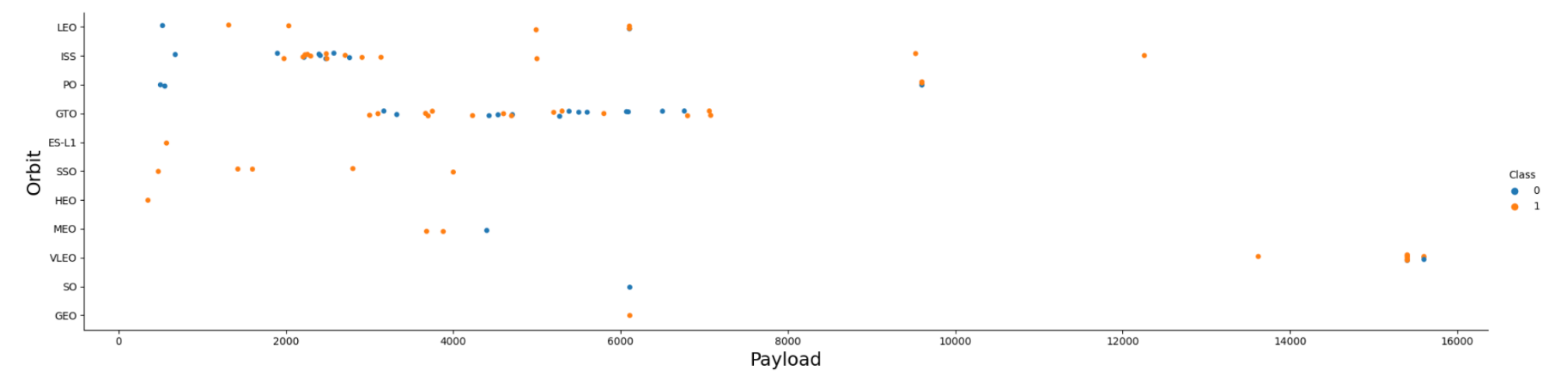
Flight Number vs. Orbit



- The SO orbit type has not been tested much and has only failed
- Success rate improves over time across all orbits
- VLEO is the new most popular orbit type

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/EDA%20With%20Visualization.ipynb

Payload vs. Orbit

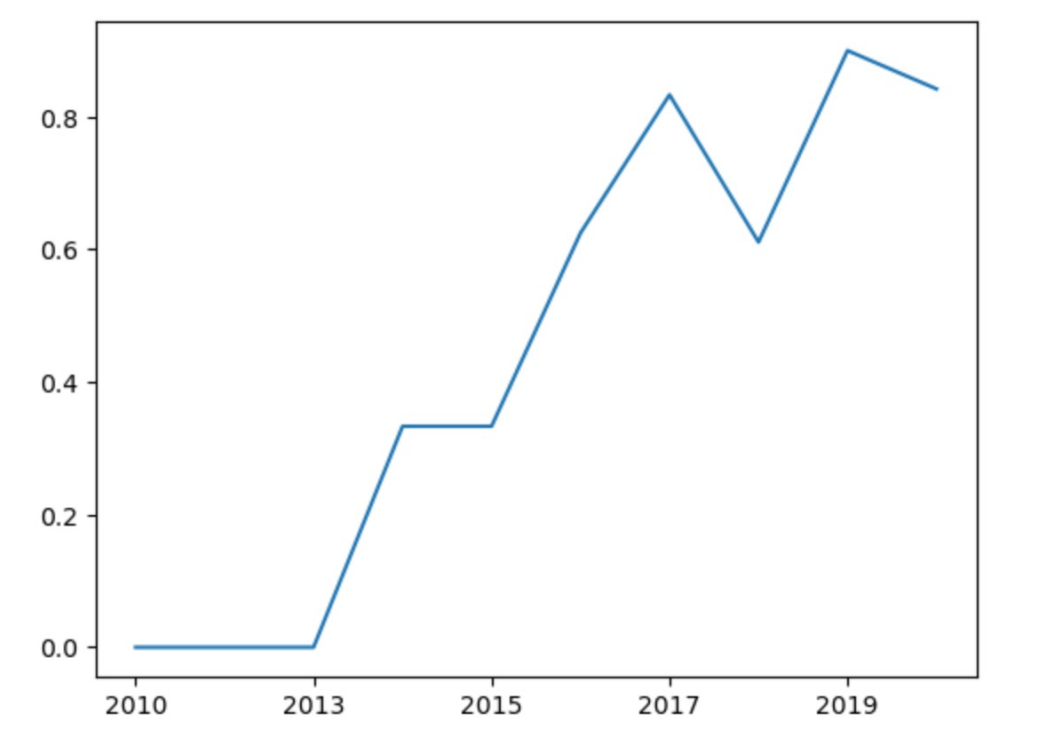


- Larger payloads over 13000 kg are only used for VLEO orbits
- ISS orbits have the widest range of payload capabilities

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/EDA%20With%20Visualization.ipynb

Launch Success Yearly Trend Line

- This graph shows the launch success average by year from 2013 to 2020
- Rate increases starting in 2013
- Plateau in 2014
- Slight dip in success rate in 2018
- Overall trending upwards



https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/EDA%20With%20Visualization.ipynb

EDA WITH SQL

All Launch Site Names

- There are 4 distinct Launch Sites in the table

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/EDAWithSQL.ipynb

Launch Site Names that start with ‘CCA’

- 5 records where the launch Site starts with ‘CCA
- These are the Cape Canaveral launches

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (p
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (p
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	N
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	N
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	N

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/EDAWithSQL.ipynb

Total Payload Mass carried by NASA (CRS)

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/EDAWithSQL.ipynb

TOTAL_PAYLOAD
45596

Average Payload Mass Carried by F9 v1.1

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/EDAWithSQL.ipynb

AVG_MASS
2928.4

Boosters with Successful Drone Ship Landing with Payload between 4000 and 6000 kg

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/EDAWithSQL.ipynb

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failed Mission Outcomes

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/EDAWithSQL.ipynb

Mission_Outcome	count(*)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Which Have Carried the Max Payloads

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/EDAWithSQL.ipynb

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- Failed Drone Ship landing outcomes
  - Included is booster version and launch site

month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/EDAWithSQL.ipynb

Ranking Landing Outcomes between 06-04-2010 and 03-20-2017

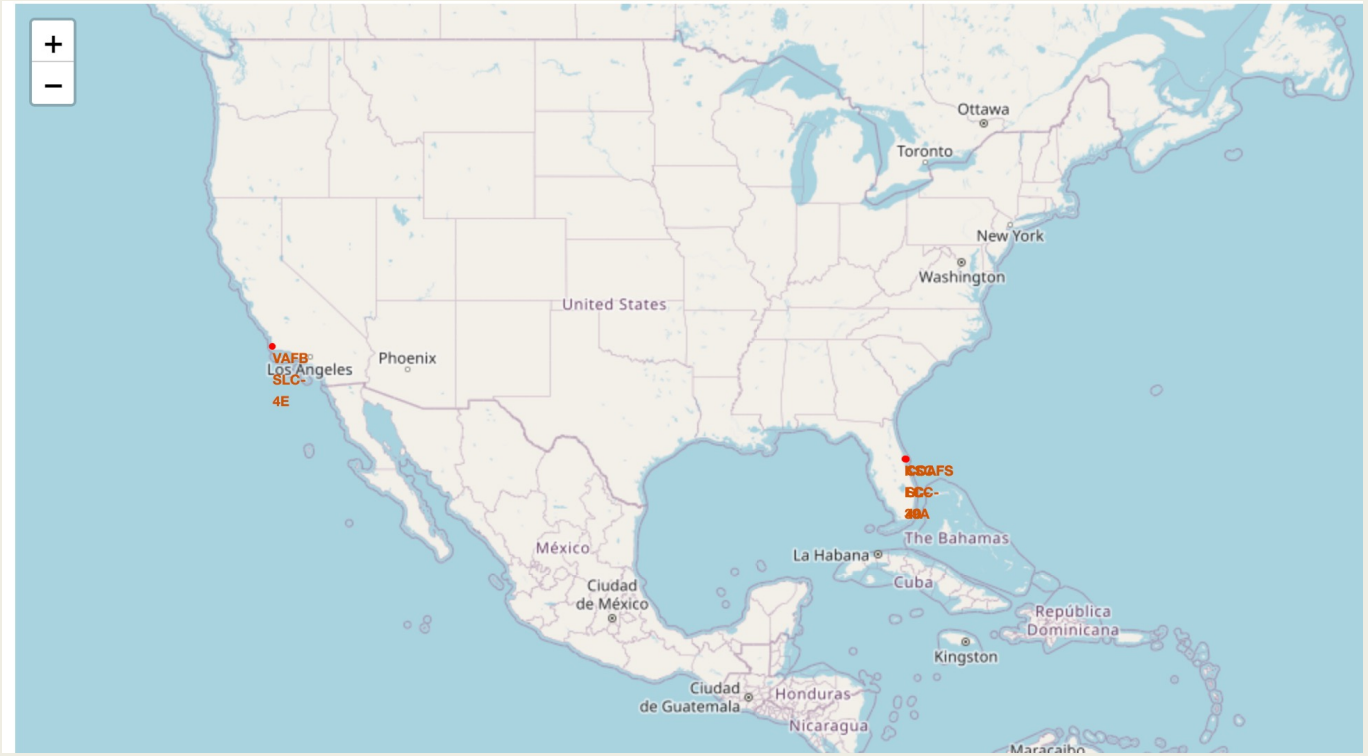
https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/EDAWithSQL.ipynb

Landing_Outcome	count(Landing_Outcome)
Success (drone ship)	5
Success (ground pad)	3
Precluded (drone ship)	1
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
No attempt	10
Failure (parachute)	1

FOLIUM

All Launch Site Markers

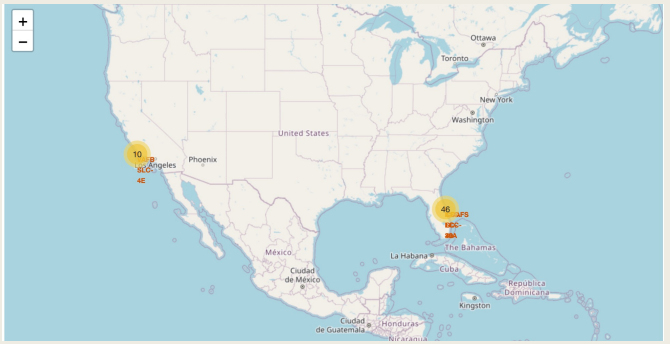
- The markers on this map are all the 4 launch sites found in the data
  - On the west coast in California, we have VAFB SLC-4E
  - On the east coast in Florida, we have
    - CCAFS LC-40
    - CCAFS SLC-40
    - KSC LC-39A

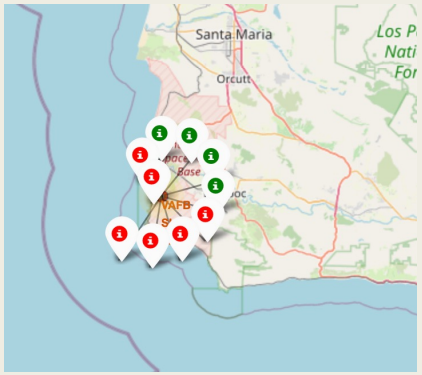


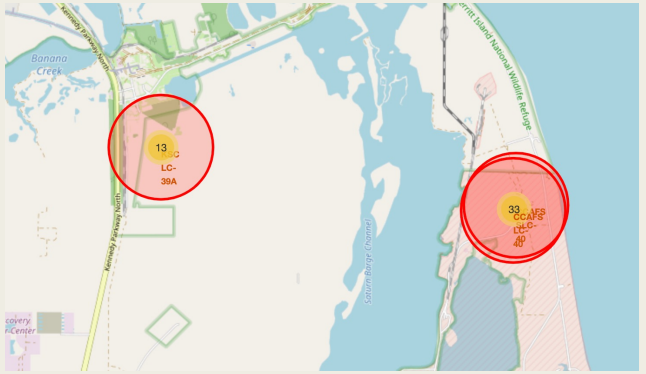
https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

All Launch Records per Site

- There are 10 launches on the west coast site and 46 east coast launches
  - The east coast launches can be broken down into two clusters
- The marker clusters were made using the MarkerCluster method



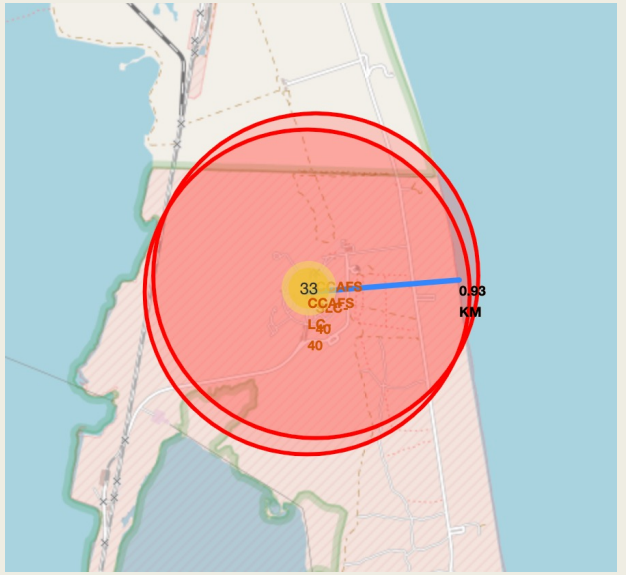


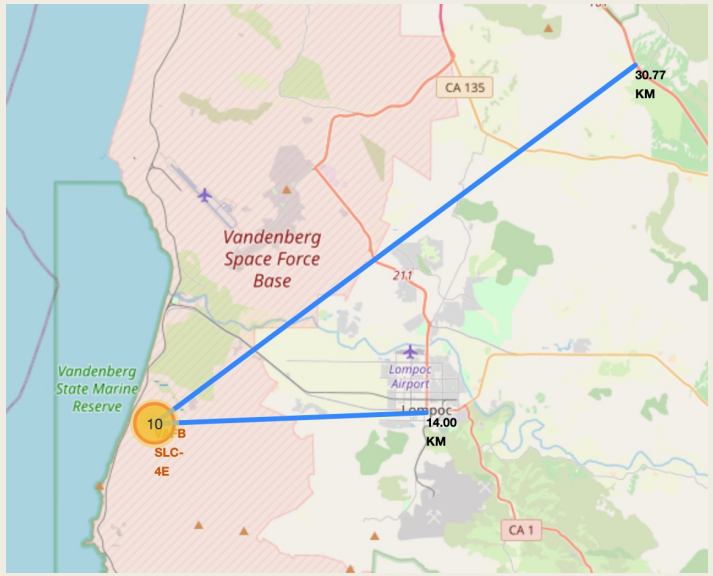


https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Launch Site Proximities

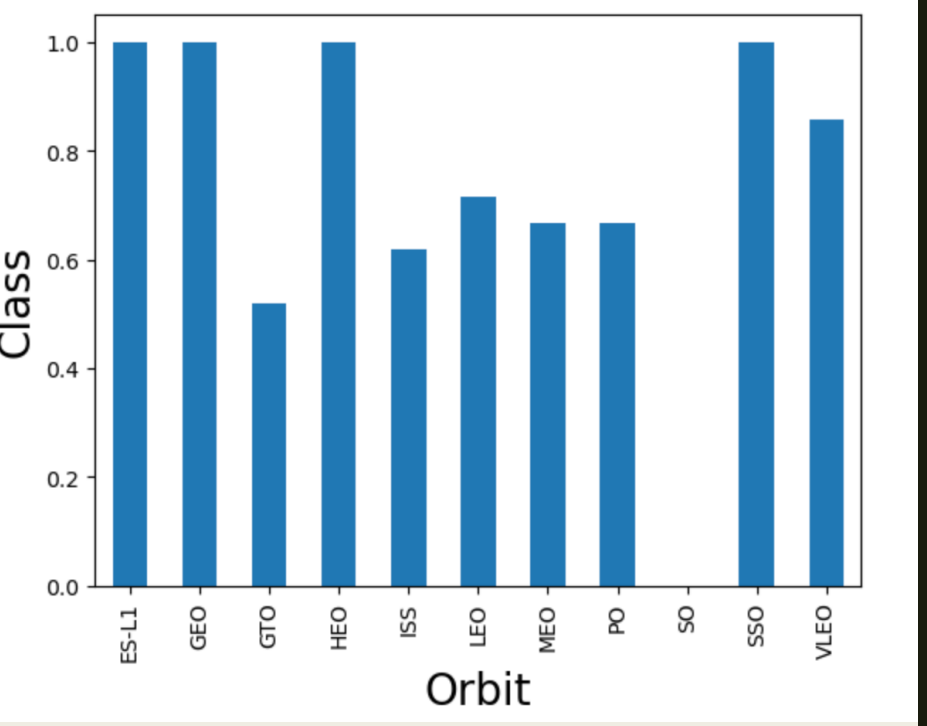
- The lines on the map are proximities to some notable landmarks
- In the top map, the distance from launch site to coastline is shown
- In the bottom map, the top line is showing the distance to a nearby highway while the bottom line is showing the distance to the nearest city

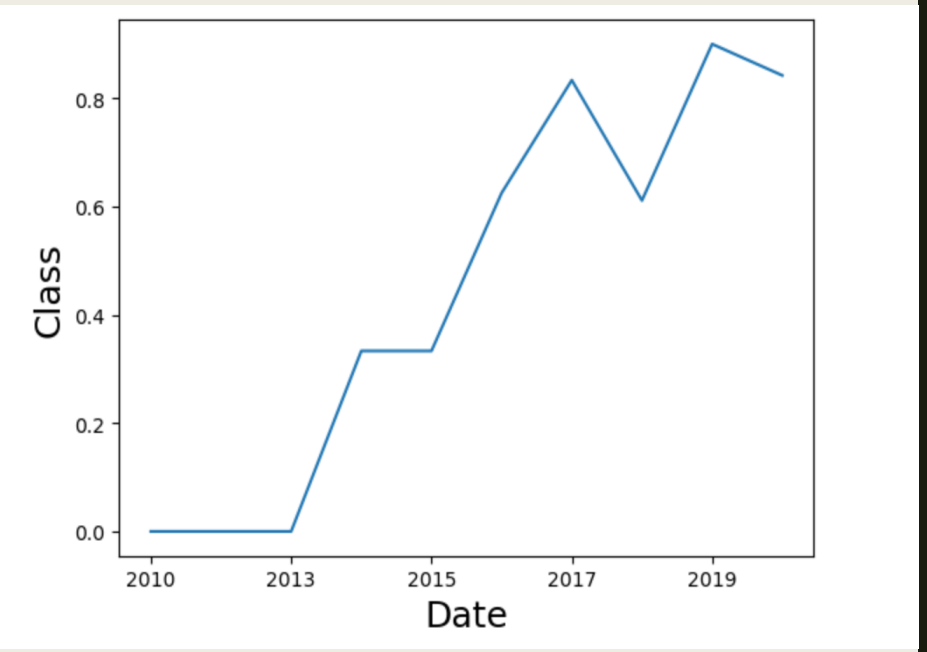




https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/7c374dda2be97ee846f60c875c5833b6ab3752ea/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

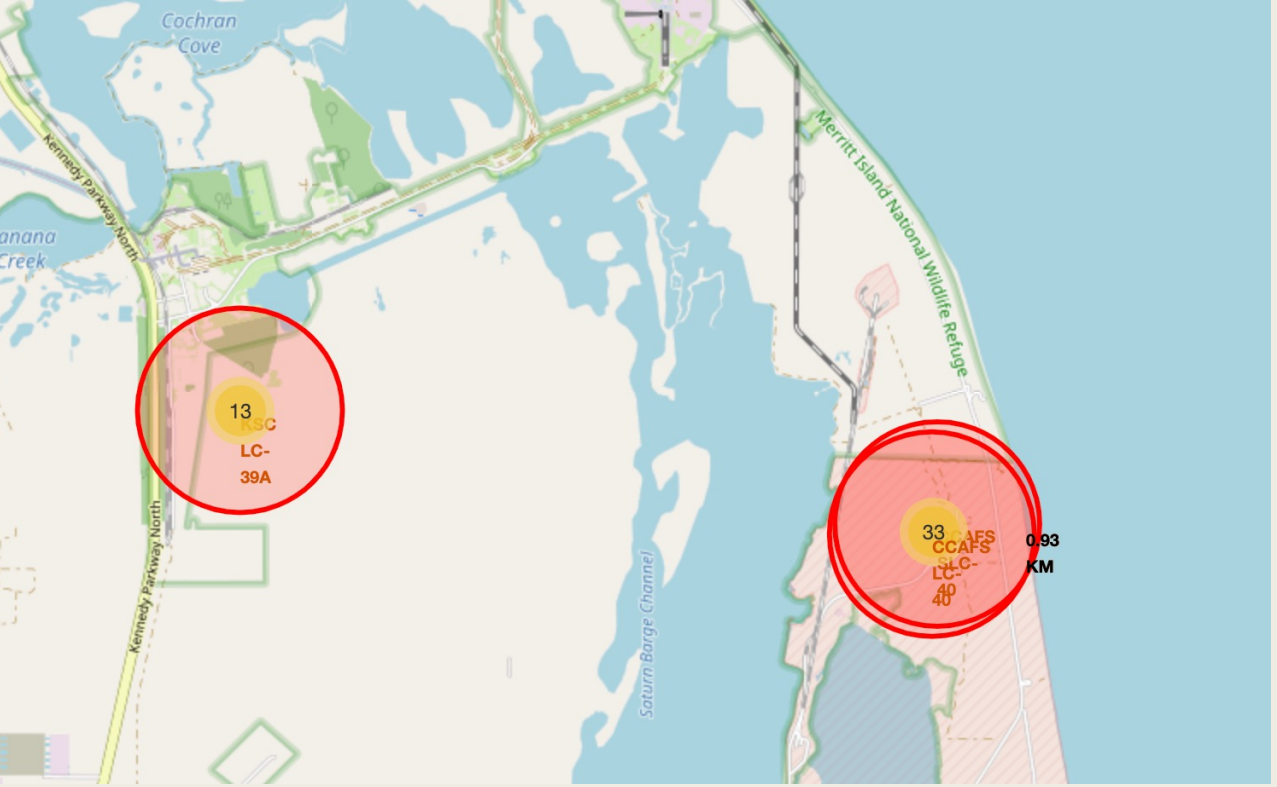

RESULTS





EDA with Visualization 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*

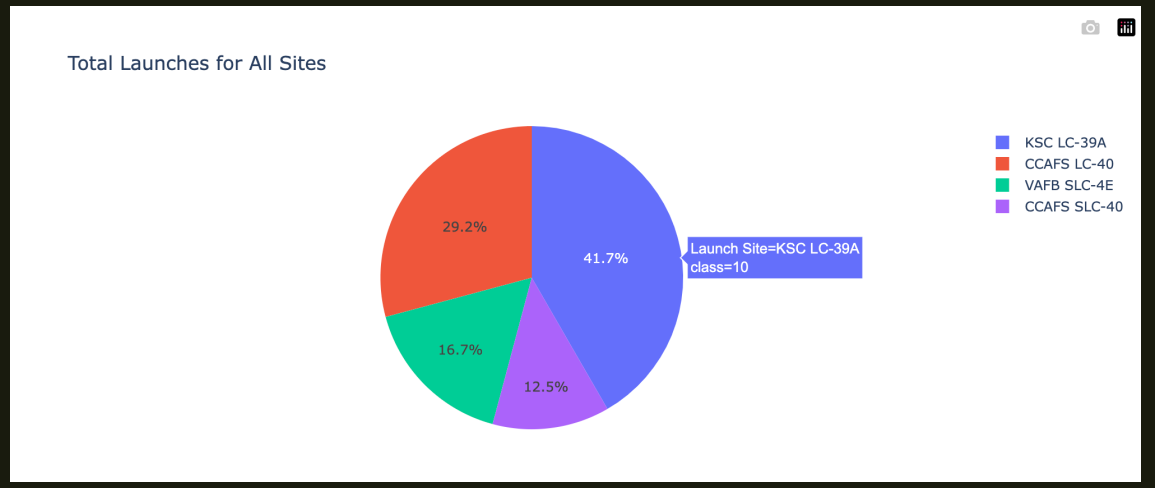


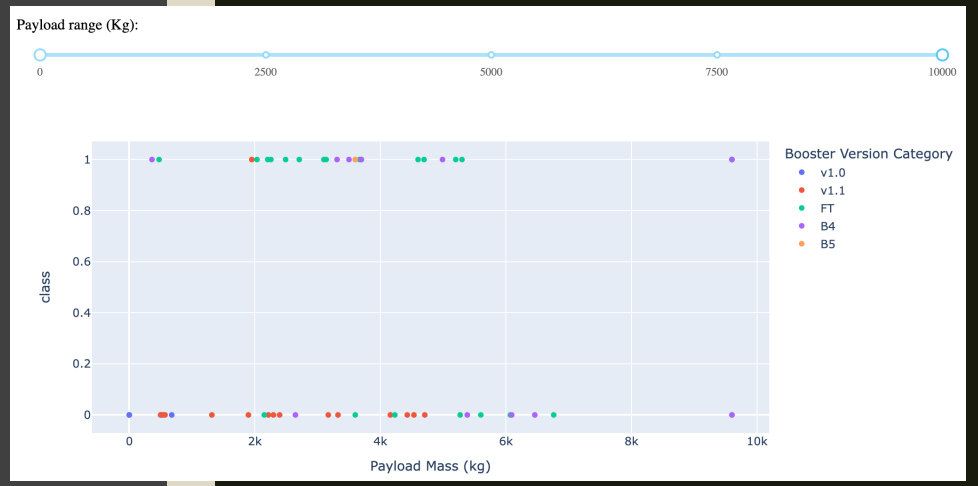
Folium Results

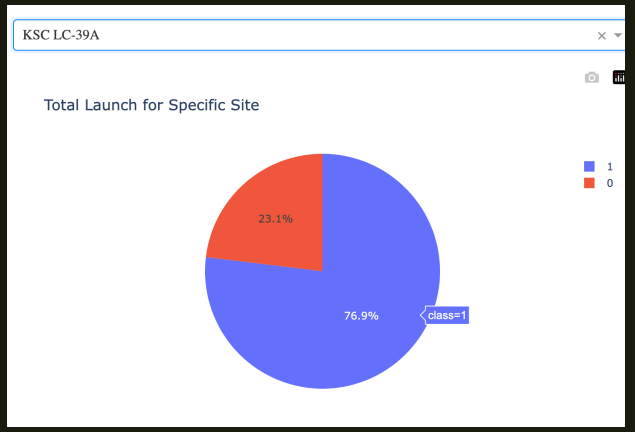
- The generated maps provided numerous key insights
  - *Most launches occurred on the east coast*
  - *Launch Sites are close to the coastline and are a considerable distance from nearby cities, i.e. safer locations*
  - *All sites have similar latitudes i.e. distances from equator*

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%
- KSC LC-39A has the highest rate of success out of all launch sites





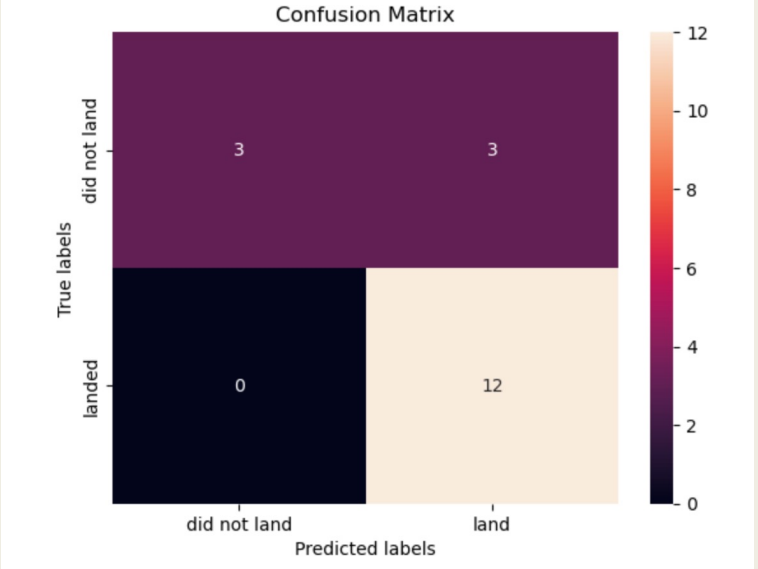


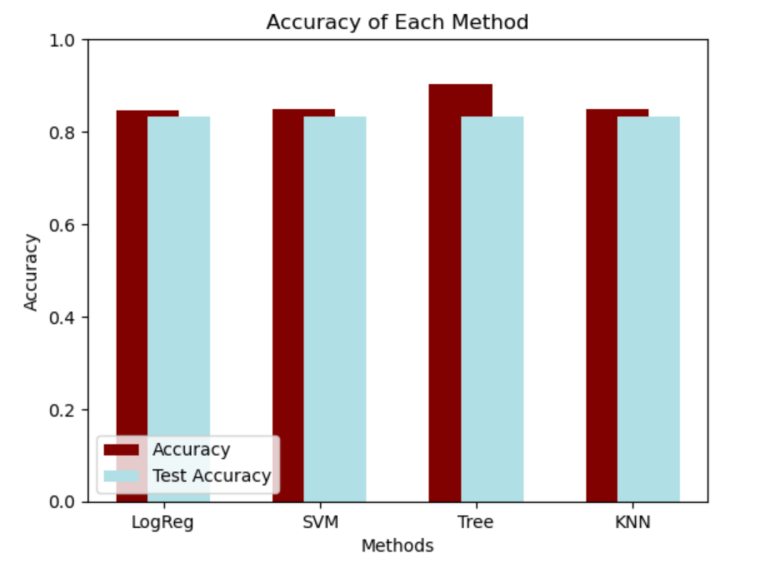
https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/22de8f7d384470ab4d44139dc3b72af06fb70536/Interactive%20Dashboard%20with%20Plotly%20Dash.py

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*
- The confusion matrix is for the Decision Tree model
  - *There were 3 false positives when running the model with the test data*
  - *The rest of the 15 test data points were predicted correctly*

https://github.com/noahtmcmahon/SpaceX-Capstone-Project/blob/22de8f7d384470ab4d44139dc3b72af06fb70536/Machine%20Learning%20Prediction.ipynb





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!