IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mauricio Antonio Del Palacio Ortega

10/09/2022

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- This Data Science Project and methodologies were based on a SpaceX API

Data were collected using web scraping on a SpaceX API

Exploratory Data Analysis was used for: data wrangling, data visualization and data analytics

Machine Learning techniques were used to this data

- Summary of all results

Data collection was possible thanks to open-source extraction

Exploratory Data Analysis let us identify features for successful launchings and landings

Machine Learning was a useful tool for predicting outcomes of all collected data

Introduction

- The scope of this study is to evaluate the feasability of the introduction to the market of Space Y,which is an aerospacial company designed to compete with Space X
- We would like to know the best way to estimate the cost of launches and predict the possibility for successful landings of the first stage of rockets, because they are the most costly parts of the full assemble, also we would like to know were to make launches.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

Data was obtained from SpaceX API (https://api.spacexdata.com/v4/rockets/)
Data was also obtained through Web Scraping on Wikipedia
(https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches/)

- Perform data wrangling

Data was summarized and analyzed through python to get reliable outcome data

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data normalized and train/test splitted for ML models for varying parameters

Data Collection

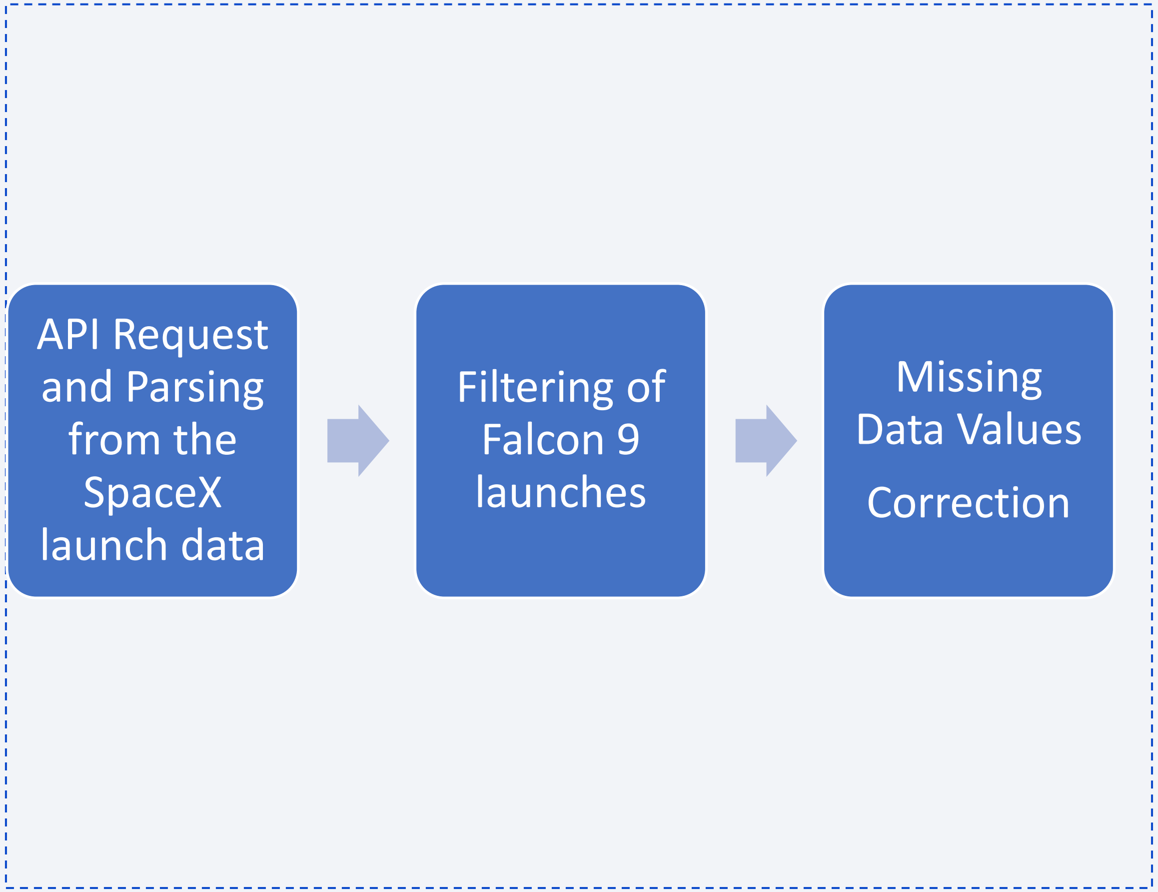
Data was obtained from SpaceX API (https://api.spacexdata.com/v4/rockets/)

Data was also obtained through Web Scraping on Wikipedia

(https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches/)

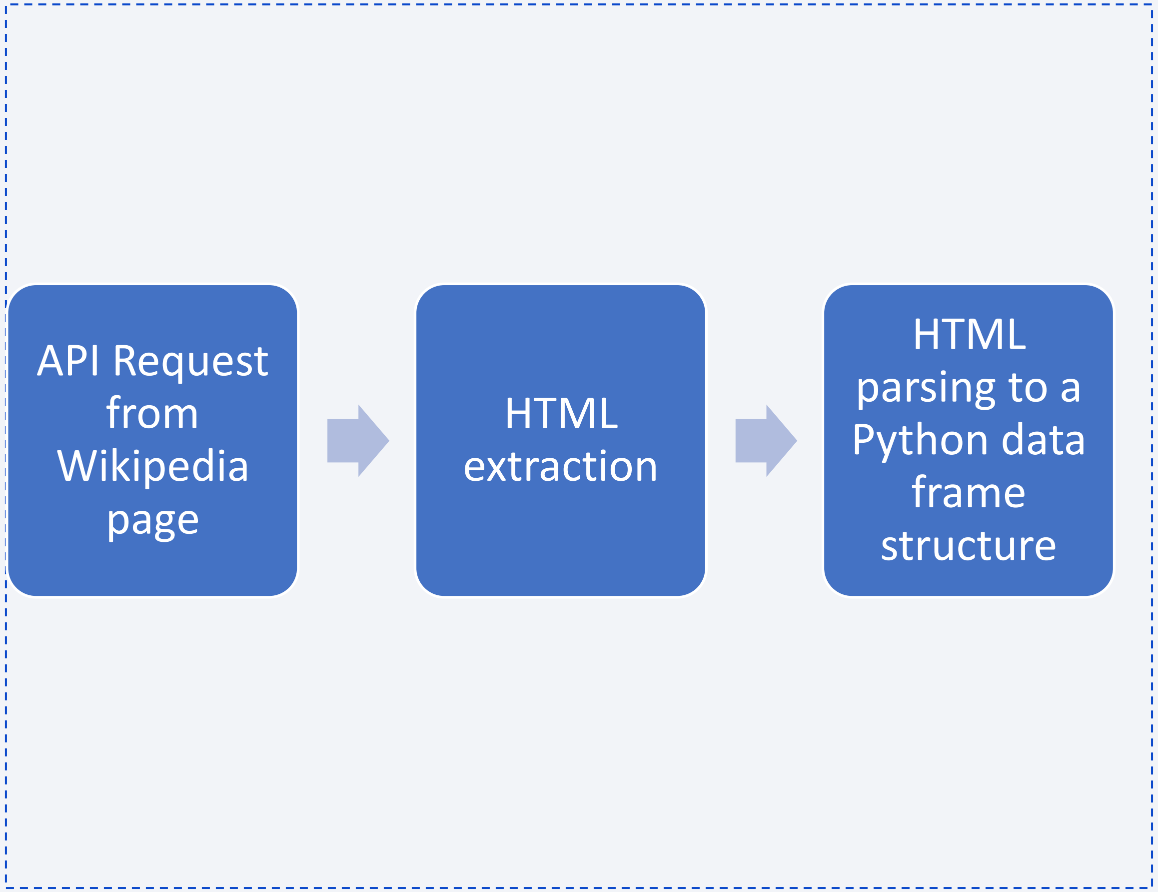
Data Collection – SpaceX API

- SpaceX data can be obtained by a public API
- This data were collected through json and python libraries like pandas
- Source code: https://github.com/mauriciodelpalacio/Applied_Data_science_Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Data obtained through Wikipedia and analyzed through html extraction and parsing into a dataframe
- Source code: https://github.com/mauriciodelpalacio/Applied_Data_science_Capstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Exploratory Data Analysis was performed on the datasets,analyzing launches per site, orbit ocurrences and outcome of all missions, given this a landing outcome was created for labeling
- Source code: https://github.com/mauriciodelpalacio/Applied_Data_science_Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Scatterplots, data and barplots were used to explore data to analyze paired variables
- Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, Orbit and Flight Number, Payload and Orbit
- This can be viewed through github, for further analysis
- Source code:

https://github.com/mauriciodelpalacio/Applied_Data_science_Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Loaded data set into IBM DB2 Database.
- Queried using SQL Python integration.
- Queries were made to get a better understanding of the dataset.
- Queried information about launch site names, mission outcomes, various pay load sizes of customers and booster versions, and landing outcomes
- Source code: https://github.com/mauriciodelpalacio/Applied_Data_science_Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Folium maps helps us mark launch sites, successful and unsuccessful landings, and a proximity example to key locations: Railway, Highway, Coast, and City. This allows us to understand why launch sites may be located where they are. Also visualizes successful landings relative to location. Markers indicate launch sites, with circles that highlight areas, and clustering is also applied to group of events in each coordinates, also we can measure distance.

- Source code:

https://github.com/mauriciodelpalacio/Applied_Data_science_Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Graphs and plots based on number of launches per site and payload range were obtained to analyze and identify the relation of payloads and launch sites, hinting us the best place to launch rockets depending on the payload size.
- Source code:
- https://github.com/mauriciodelpalacio/Applied_Data_science_Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

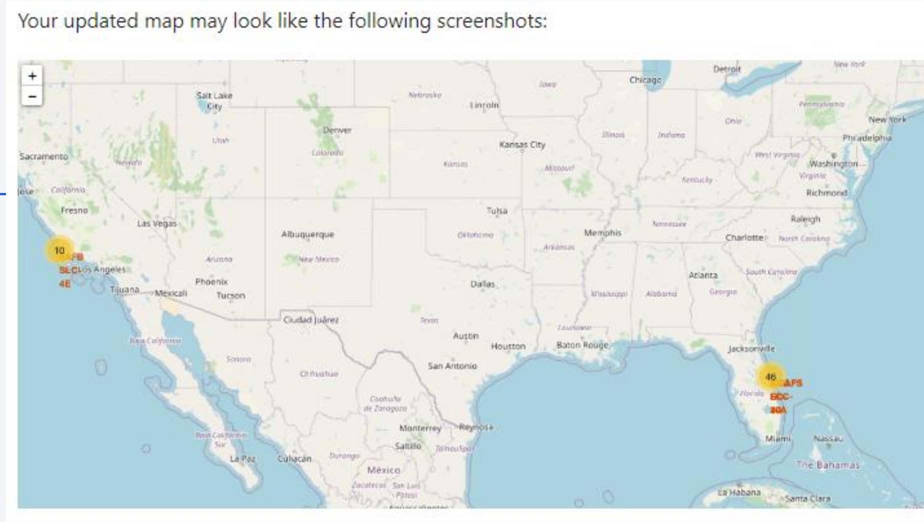
- Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbors, using the next strategy:
- Source code:
- https://github.com/mauriciodelpalacio/Applied_Data_science_Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

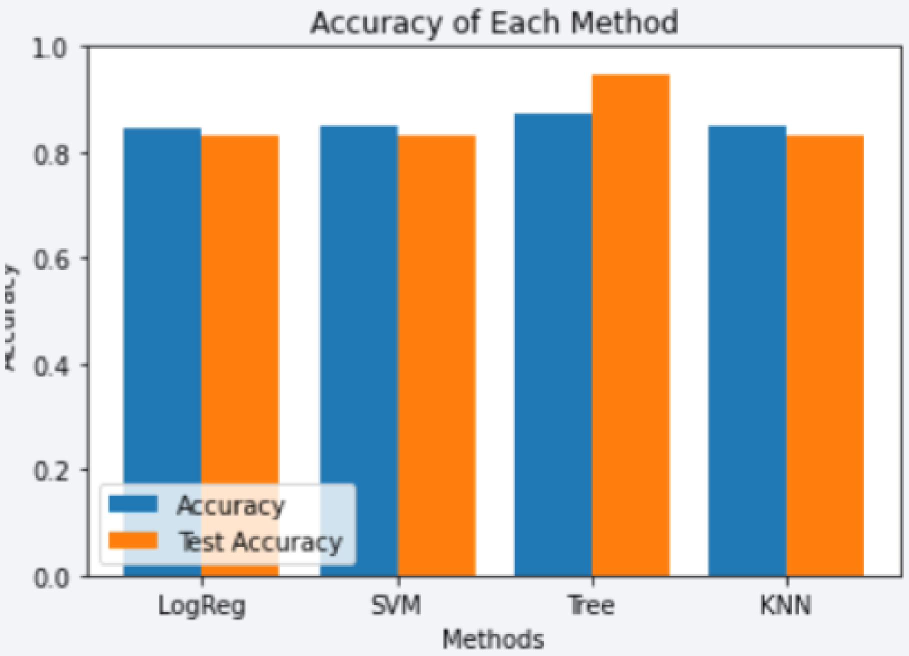
- Exploratory data analysis results:

Space X has 4 launch sites used by NASA and Space X, and the first successful landing occurred in 2015, with the F9 v1.1 booster average payload being 2928 kgs. Also the landing outcomes improved as years passed by, even having payloads above average, with almost the entire mission outcomes being successful, implying a great improvement over few years

- Interactive analytics
- The best launch sites were places near sea with the east coast being the preferred location.



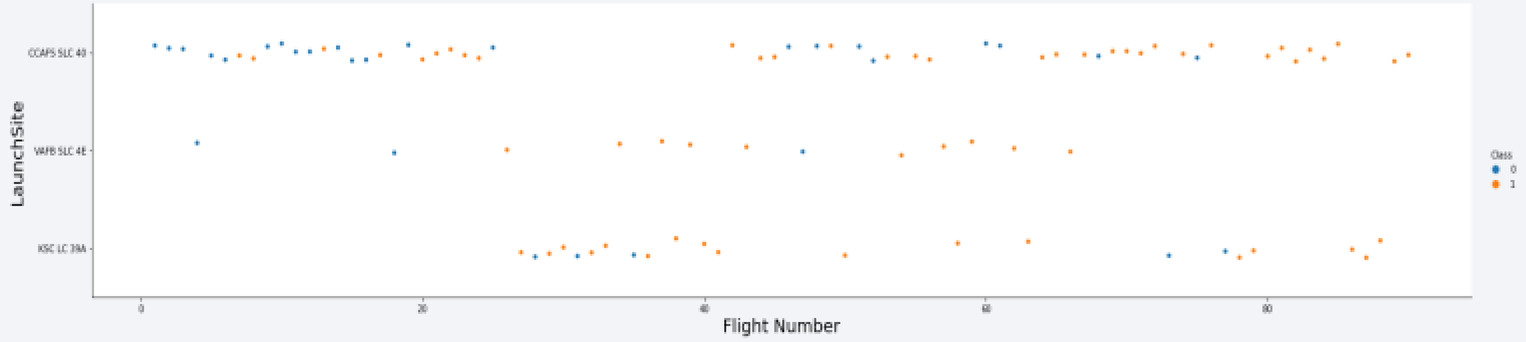
- Predictive analysis showed us that Tree classification were the best model available with 87% accuracy for training and 94% for test data



Section 2

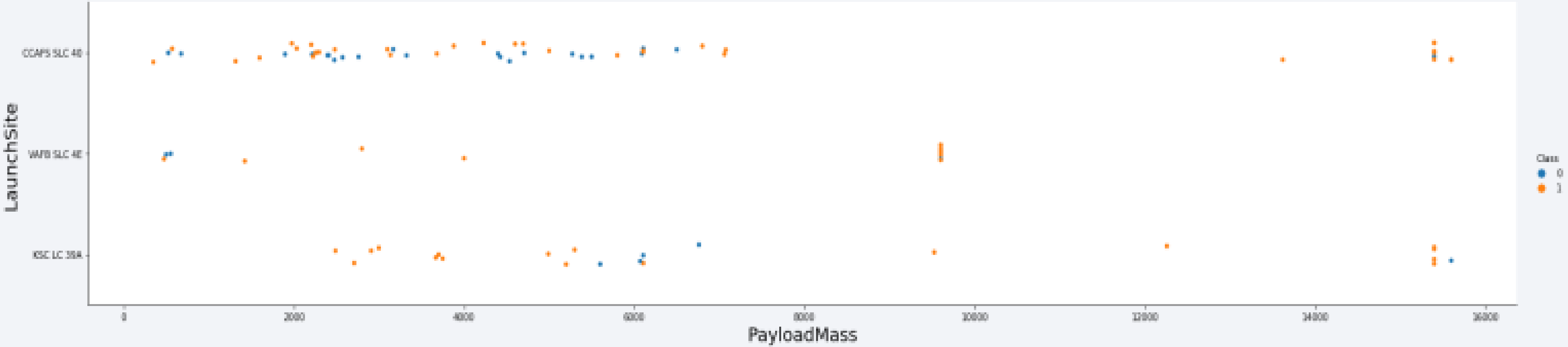
Insights drawn from EDA

Flight Number vs. Launch Site



- The best launch site nowadays is CCAF5 SLC 40, where most of recent launches were successful; In second place VAFB SLC 4E and third place KSC LC 39A;
- It's also possible to see that the general success rate improved over time.

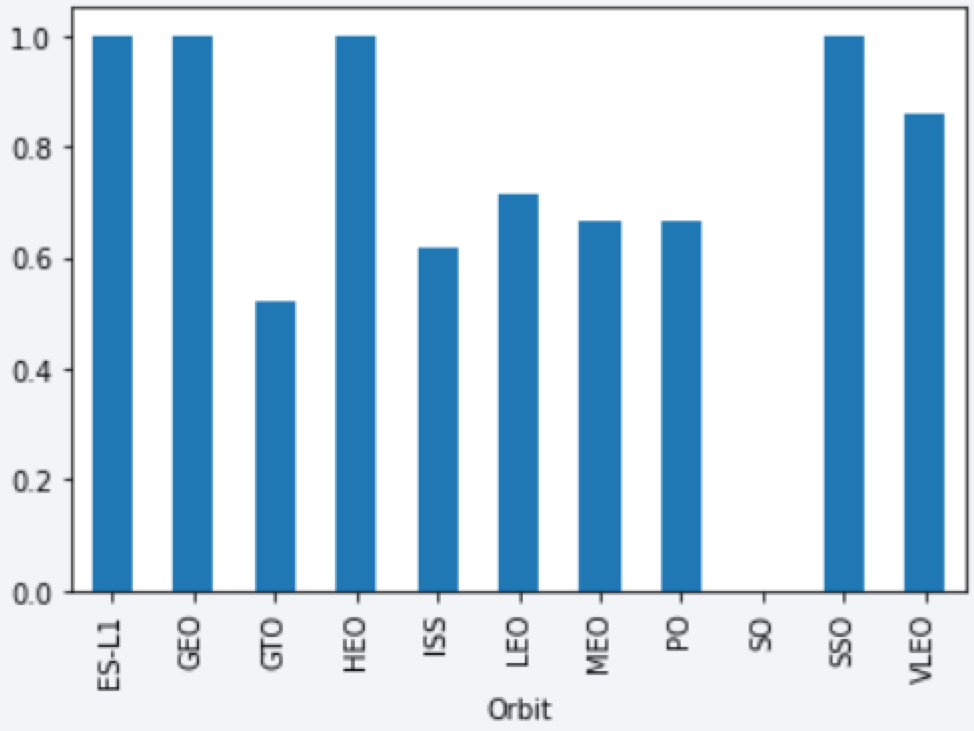
Payload vs. Launch Site



- Payloads over 9,000kg have excellent success rate, and over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites. Hinting us that these locations could be exclusively used for heavy weight launches

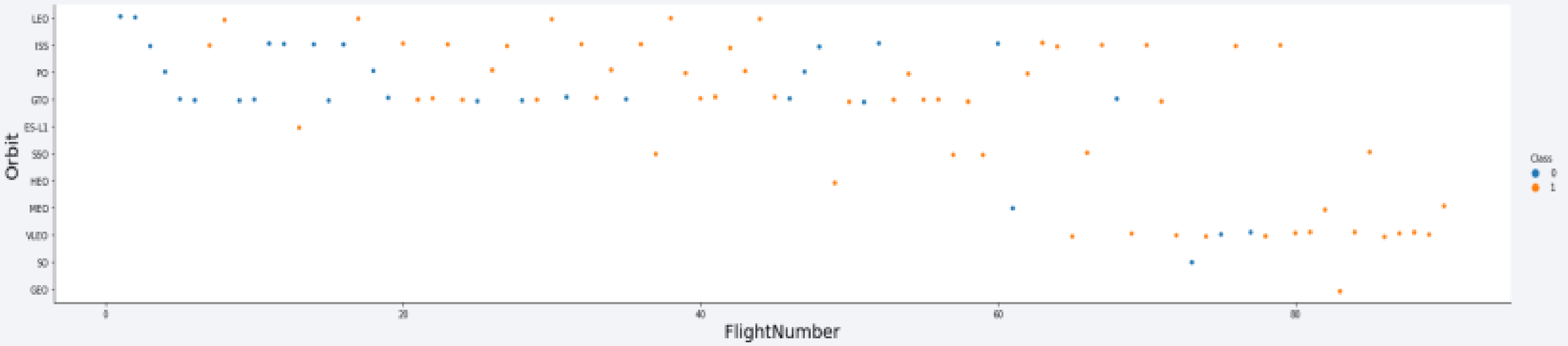
Success Rate vs. Orbit Type

- • The biggest success rates by orbits were: ES-L1,GEO, HEO,SSO.
- The worst success rates were:
- SO,ISS,GTO.

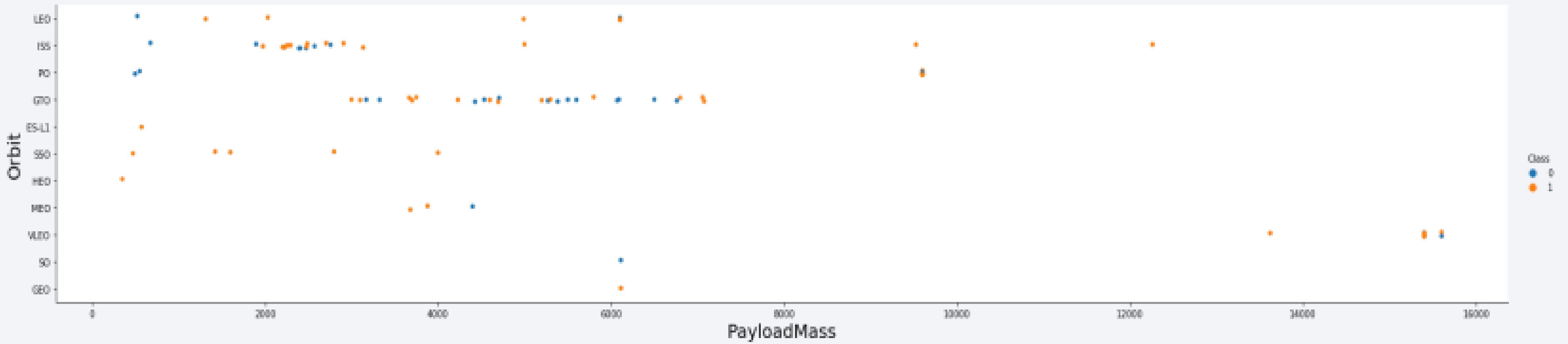


Flight Number vs. Orbit Type

Success rate improved over time to all orbits, use of VLEO increased over time showing that preferences changed by financial reasons.



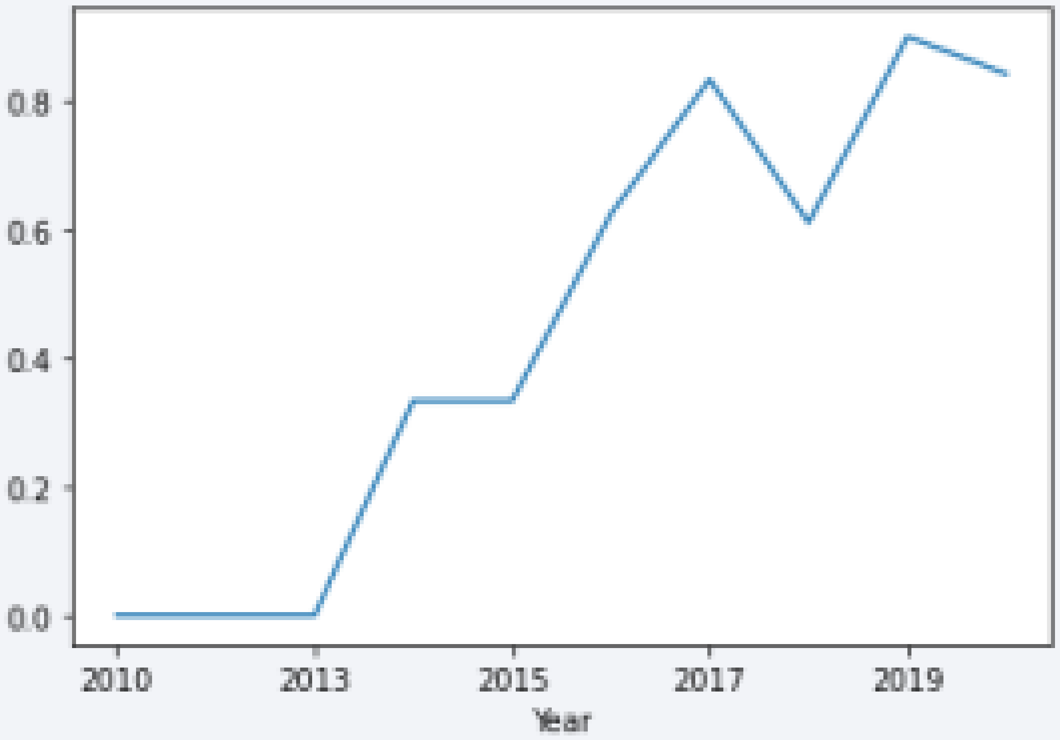
Payload vs. Orbit Type



- ISS orbit has the widest range of payload and a good rate of success, also there is no relation between payload and success rate to orbit GTO being a really pragmatic orbit, finally SO and GEO orbits don't get a lot of launches.

Launch Success Yearly Trend

Between 2010 and 2013 were no important success rate increasing from 2013 until 2020, in a fast pace. Showing us that from 2010 to 2013 a training timelapse were used to improve on further years



All Launch Site Names

- There are 4 launch sites.

Launch Site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`:
- Cape Canaveral launches are shown.

Date	Time UTC	Booster Version	Launch Site	Payload	Payload Mass kg	Orbit	Customer	Mission Outcome	Landing Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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 (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attemp

Total Payload Mass

- Total payload carried by boosters from NASA, by accumulative analysis:

Total Payload (kg)
111.268

Average Payload Mass by F9 v1.1

- • Average payload mass carried by booster version F9 v1.1:

Avg Payload (kg)
2.928

First Successful Ground Landing Date

- • First successful landing outcome on ground pad:
- This were found by looking for successful landing outcomes and getting the earliest date.

Min Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Getting values between 4000 and 6000 kgs and looking for succesful landings gave us this result

Booster Version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

Total Number of Successful and Failure Mission Outcomes

Number of successful and failure mission outcomes:

Mission Outcome	Occurrences
Success	99
Success (payload status unclear)	1
Failure (in flight)	1

A Summarization by mission outcomes gave us this table, showing only 2 failures and 1 by flight problems.

2015 Launch Records

- The failures occurred only at CCAFS LC-40

Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Ranking of all landing outcomes between the date 2010-06-04 and 2017-03-20: showing us that no attempt were the primarily outcome

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

Section 3

Launch Sites Proximities Analysis

All Launch sites

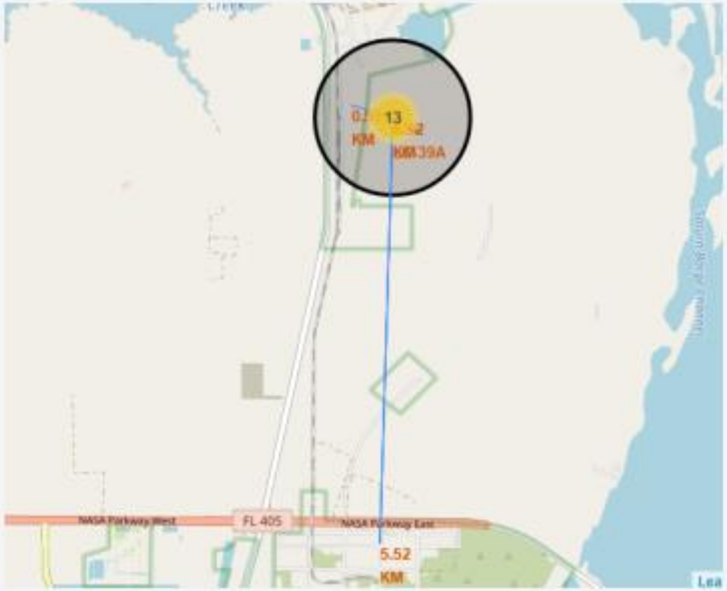
- Launch sites are based near coast, probably by atmosphere and combustion requirements.

Launch Outcomes by Site

- • Example of KSC LC-39A launch site launch outcomes. Note that green and red dots show failure and success spots

Logistics and Safety

- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas.

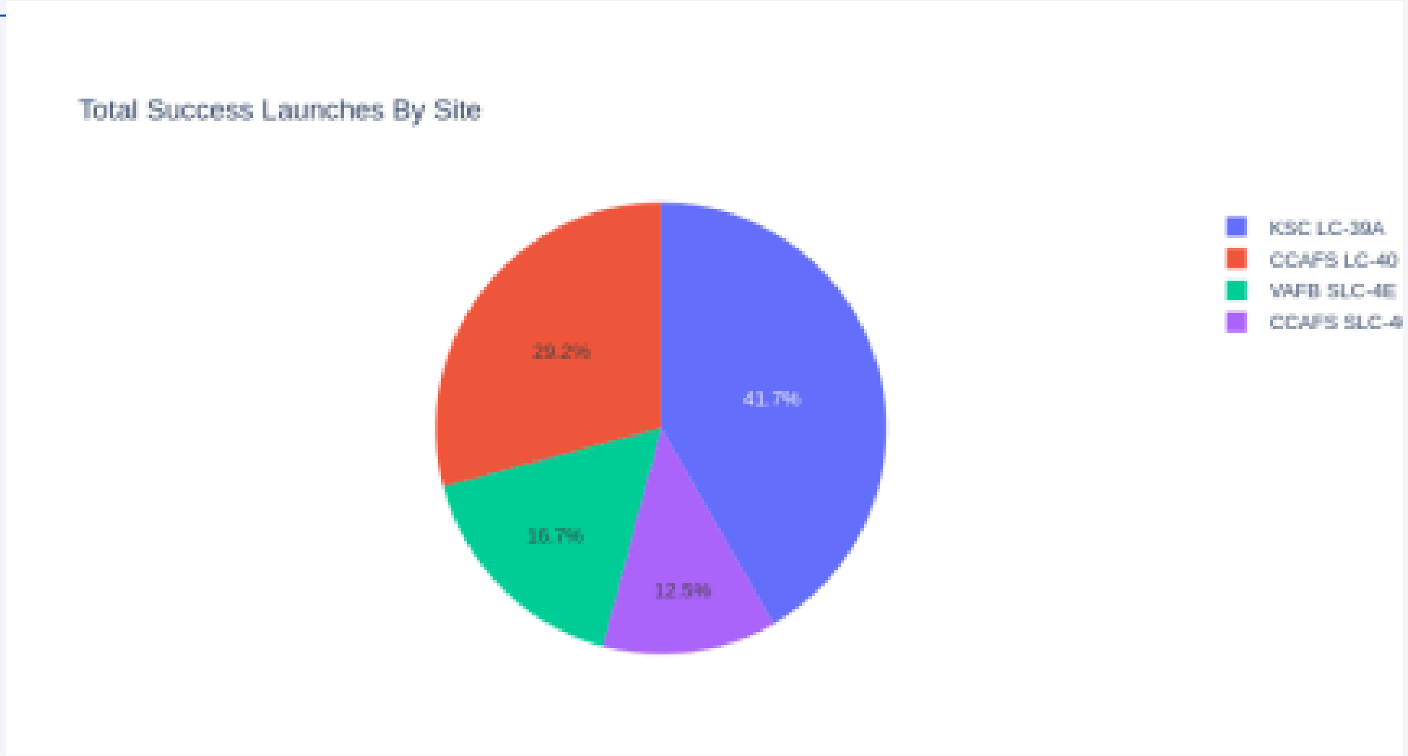


Section 4

Build a Dashboard with Plotly Dash

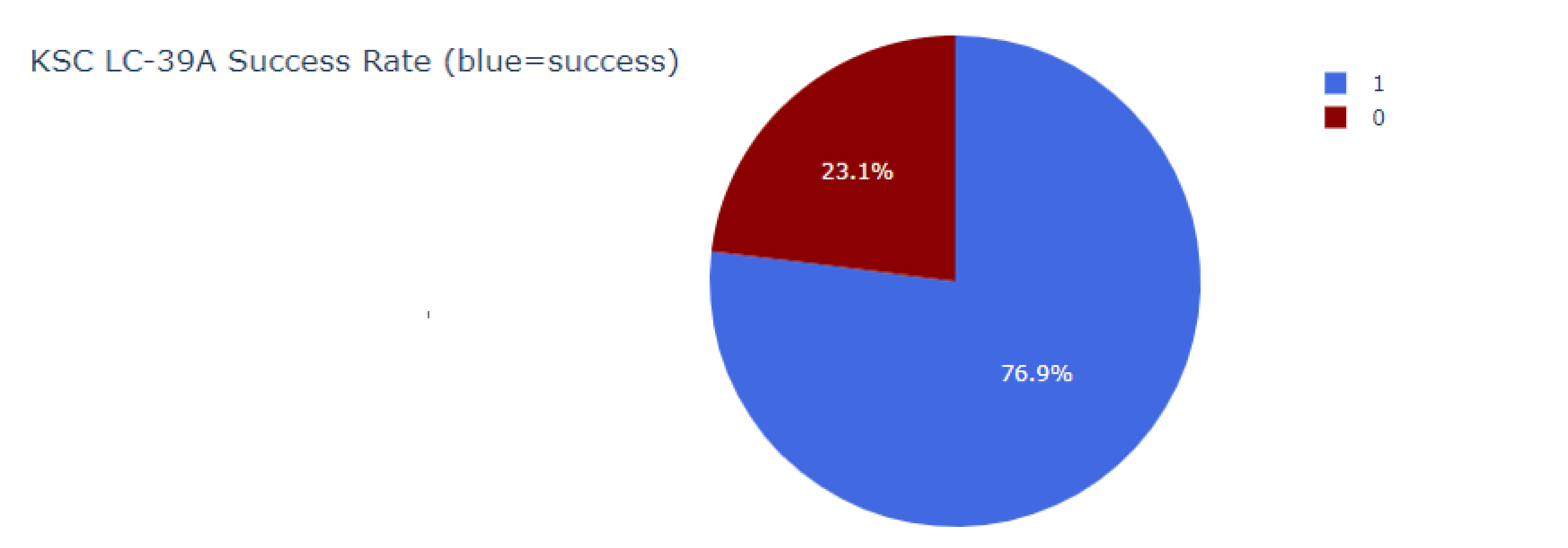
Successful Launches by Site

- This is the distribution of successful landings across all launch sites. CCAFS LC-40 is the old name of CCAFS SLC-40 so CCAFS and KSC have the same amount of successful landings, but a majority of the successful landings where performed before the name change. VAFB has the smallest share of successful landings. This may be due to smaller sample and increase in difficulty of launching in the west coast.



Highest Success Rate Launch Site and Payload vs Lauch Outcome

- KSC LC-39A has the highest success rate with 10 successful landings and 3 failed landings.

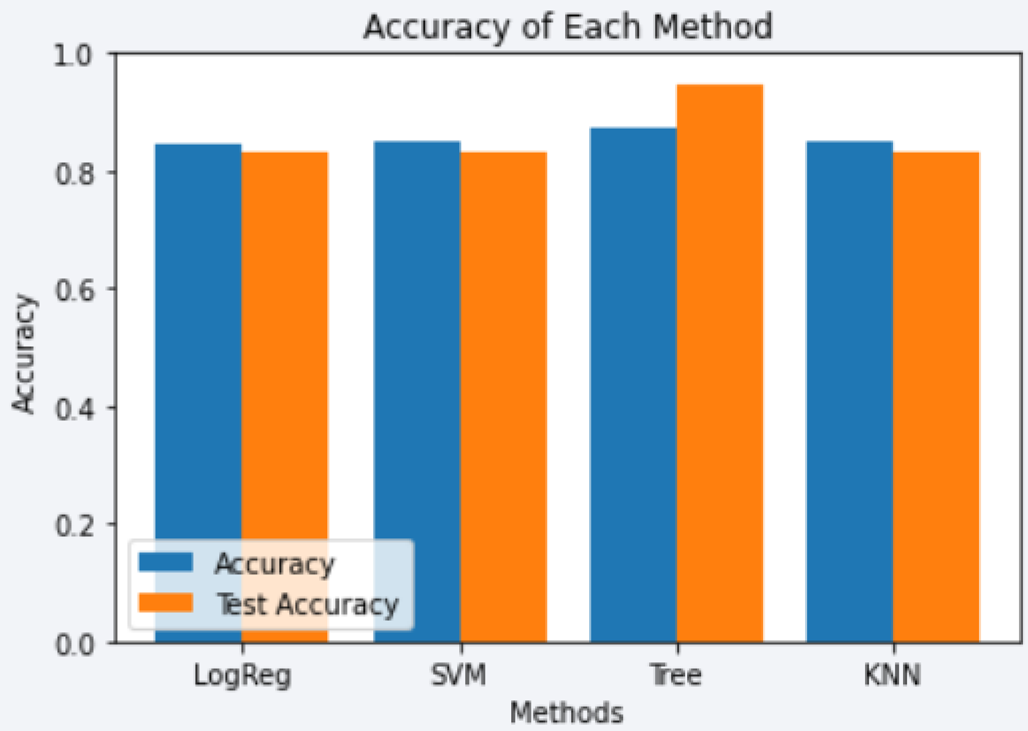


Section 5

Predictive Analysis (Classification)

Classification Accuracy

Four classification models were tested. The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87% and 94% for the test accuracy.



Confusion Matrix

- Confusion matrix of Decision Tree Classifier show us that the diagonal of the classifier(upper left and lower right) for true classification was achieved, and misclassification occurred only in 1 event (predicting a landing were there were non existing)

Conclusions

- The use of multiple data sources gave strength and refinement of the ML processing.
- The top launch site found was KSC LC-39A
- Launches over 7 tons present less risk.
- Successful landing depends on the repetition and evolution of rockets and strategies
- Decision tree classifiers could be used to predict successful landings and lessen error prone launches

Appendix

- Machine Learning notebook had trouble showing on github given the vast deprecation of some methods, its available for download and display on an alternative notebook API like Jupyter or Google Colab

Thank you!