IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

MOLLETI CHIRU VISHAL
26-2-2002

Outline

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

Executive Summary

- Data was analysed using the following methodologies:

1)Machine Learning Prediction.

2)Exploratory Data Analysis (EDA), including data wrangling, data visualization and interactive visual analytics;

3)Data Collection using web scraping and SpaceX API;

- Summary of all results

1)Using the collected data, Machine Learning Prediction demonstrated the best model for predicting which attributes are necessary to drive this opportunity.

2)EDA enabled the identification of which features best predict the success of launches.

3) It was possible to collected valuable data from public sources;

Introduction

- The goal is to assess the viability of the new firm Space Y in competing with Space X.
- Desirable answers:

1)The best technique to estimate total launch costs is to predict successful first-stage rocket landings.

2)Where is the best place to make launches.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
    - Data from Space X was obtained from 2 sources:
    - Space X API (https://api.spacexdata.com/v4/rockets/) • WebScraping
    - (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform data wrangling

• Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features

Methodology

- 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 acquired up to this point were normalised, divided into training and test data sets, and evaluated by four different classification models, with the accuracy of each model evaluated using different parameter combinations.

Data Collection

- Data sets were collected from Space X API (https://api.spacexdata.com/v4/rockets/) and from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using web scraping technics.

– SpaceX API

SpaceX provides a public API through which data may be downloaded and used.

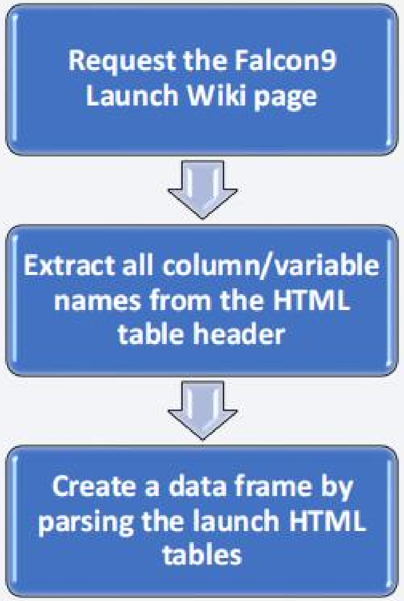
This API was used according to the flowchart beside and then data is persisted.

Source Code : https://github.com/chiruvishal/Datascience/blob/main/Spacex.ipynb

Data Collection – Scraping

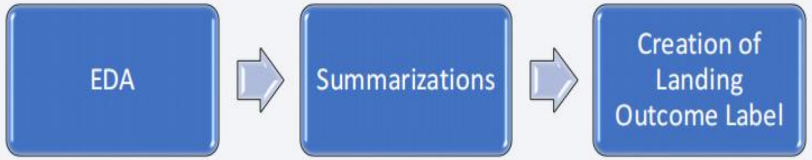
- *Data from SpaceX launches can also be obtained from Wikipedia;*
- *According to the flowchart, data is downloaded from Wikipedia* and then persisted.

Source Code- https://github.com/chiruvishal/Datascience/blob/main/Web%20Scraping.ipynb



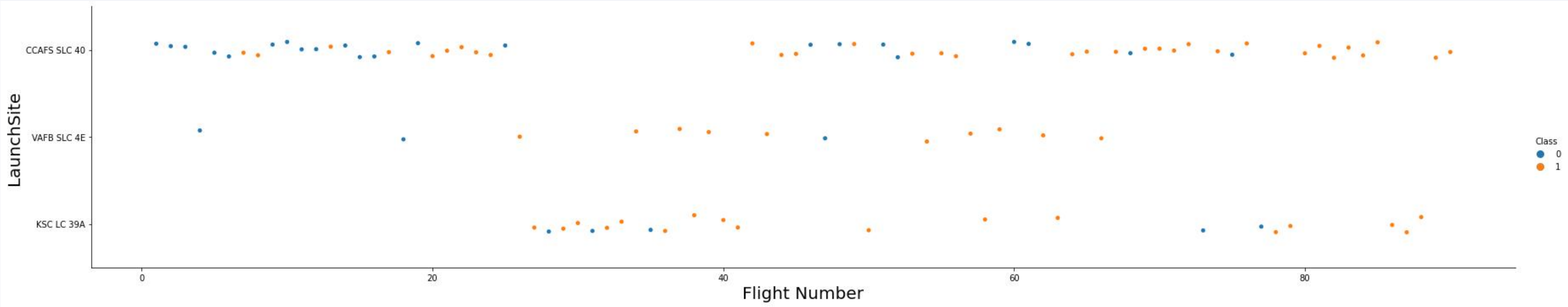
Data Wrangling

- The dataset was first subjected to some exploratory data analysis (EDA). The sums of launches per site, occurrences of each orbit, and occurrences of mission outcome per orbit type were then computed. Finally, the Outcome column was used to generate the landing outcome label.
- Source Code- https://github.com/chiruvishal/Datascience/blob/main/Data%20Wrangling.ipynb



EDA with Data Visualization

- Scatterplots and barplots were used to visualise the relationship between two features when exploring data.Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, Orbit and Flight Number, Payload and Orbit.



- Source Code- https://github.com/chiruvishal/Datascience/blob/main/Exploratory%20Analysis%20using%20matplotlib.ipynb

EDA with SQL

- The following SQL queries were performed:
- Names of the unique launch sites in the space mission;
- Top 5 launch sites whose name begin with the string 'CCA’;
- Total payload mass carried by boosters launched by NASA (CRS);
- Average payload mass carried by booster version F9 v1.1;
- Date when the first successful landing outcome in ground pad was achieved;
- Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg;
- Total number of successful and failure mission outcomes;
- Names of the booster versions which have carried the maximum payload mass;
- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015; and
- Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20.

source code-
https://github.com/chiruvishal/Datascience/blob/main/Exploratory%20Analysis%20(1).ipynb

Build an Interactive Map with Folium

- Folium Maps were utilised with markers, circles, lines, and marker clusters. Markers represent points, such as launch sites; Circles represent highlighted areas around specific coordinates, such as NASA Johnson Space Center; Marker clusters represent groups of events in each coordinate, such as launches at a launch site; and Lines represent distances between two coordinates.
- Source Code- https://github.com/chiruvishal/Datascience/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

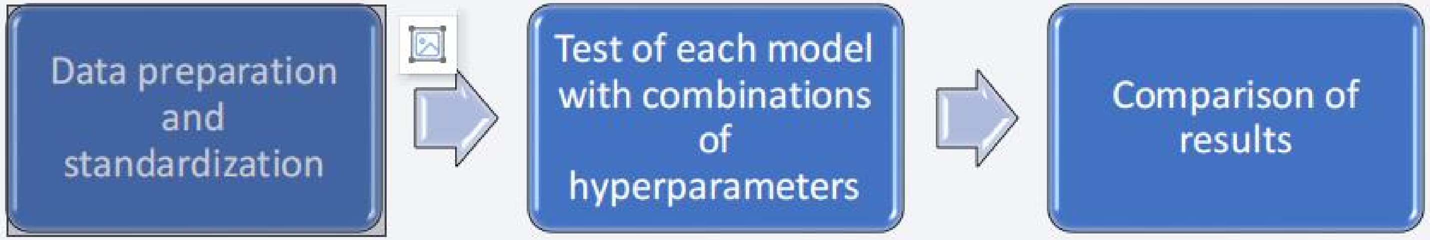
- The graphs and plots shown below were used to visualise data:
  - Payload range • Percentage of launches by site .

This combination allows for a fast analysis of the relationship between payloads and launch locations, assisting in determining the optimum spot to launch based on payloads.

Source Code-
https://github.com/chiruvishal/Datascience/blob/main/DashboardwithPlotly.py

Predictive Analysis (Classification)

- Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbors.



Source Code-
https://github.com/chiruvishal/Datascience/blob/main/Machine%20Learning%20Prediction.ipynb

Results

- Exploratory data analysis results:

1) Space X uses 4 different launch sites;

2) The first launches were done to Space X itself and NASA;

3)The average payload of F9 v1.1 booster is 2,928 kg;

4)The first success landing outcome happened in 2015 fiver year after the first launch;

5)Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;

6)Almost 100% of mission outcomes were successful;

7)Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;

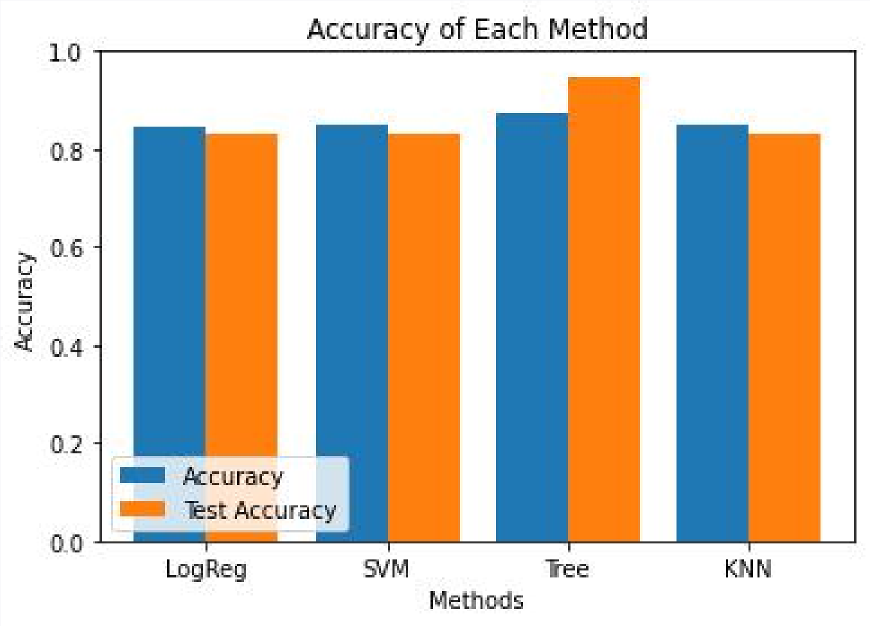
8) The number of landing outcomes became as better as years passed.

Results

- Using interactive analytics, it was possible to determine that launch sites used to be in safe locations, such as near the sea, and had a good logistic infrastructure surrounding them.

Results

- Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%.

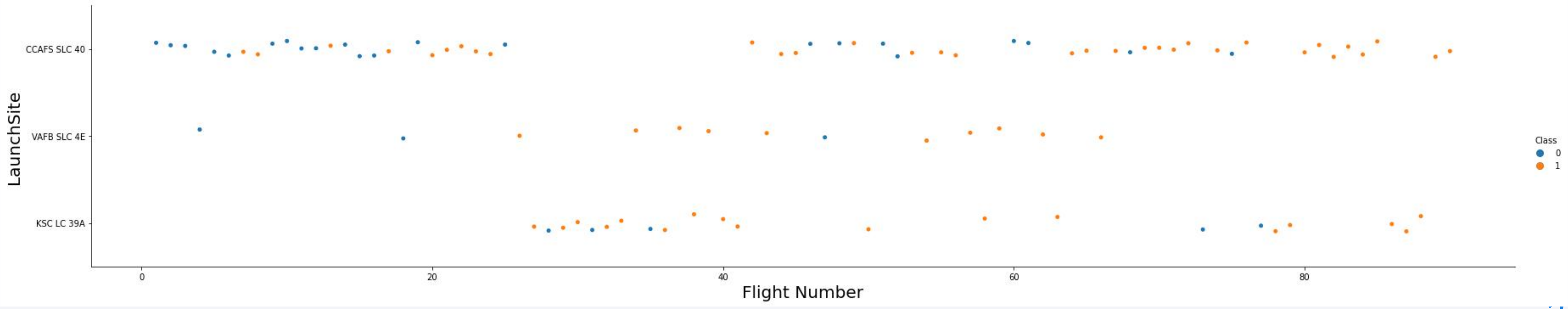


Section 2

Insights drawn from EDA

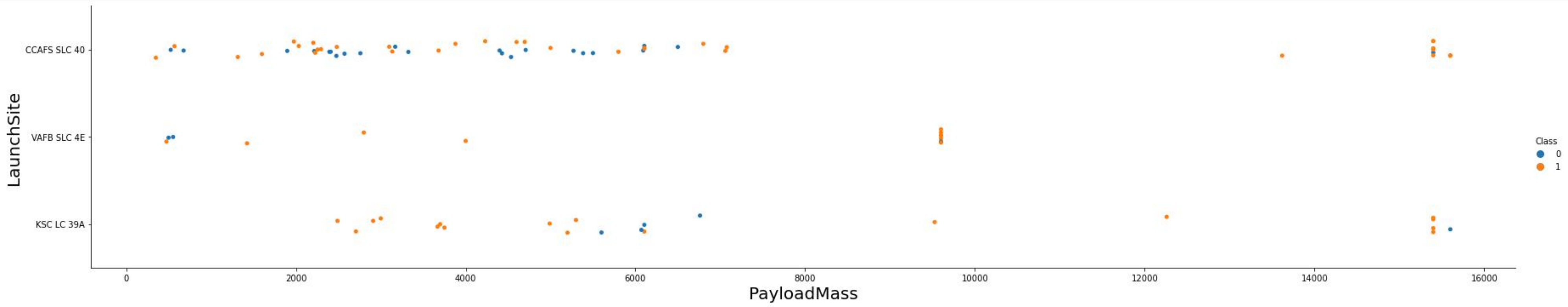
Flight Number vs. Launch Site

- According to the plot above, the best launch site nowadays is CCAF5 SLC 40, where the majority 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 weighing more over 9,000kg (about the weight of a school bus) have a high success rate.
- Payloads weighing more than 12,000kg appear to be limited to the CCAFS SLC 40 and KSC LC 39A launch sites.



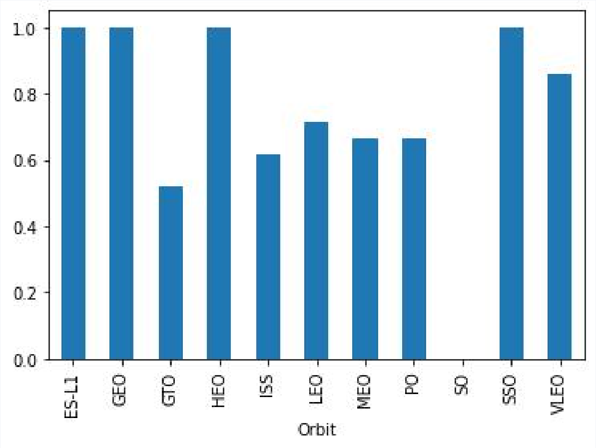
Success Rate vs. Orbit Type

The biggest success rates happens to orbits:

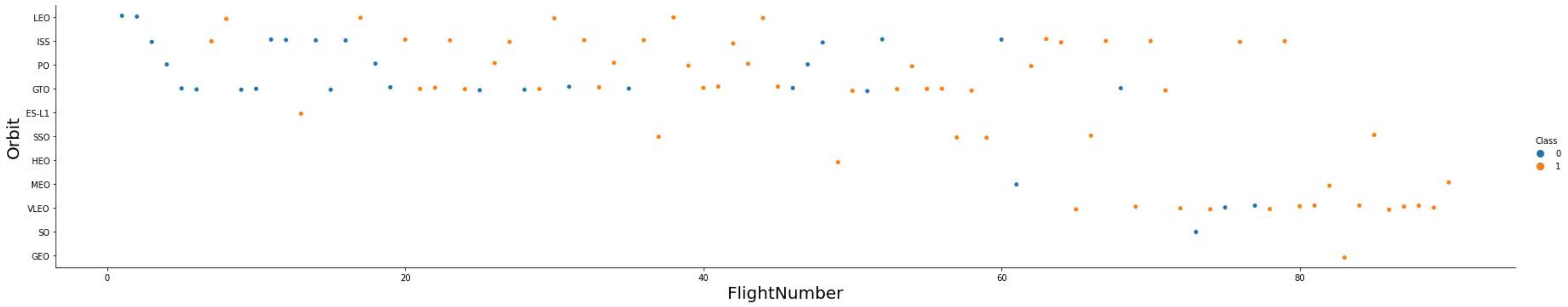
- ES-L1;
- GEO;
- HEO; and
- SSO.

Followed by:

- VLEO (above 80%); and
- LFO (above 70%).

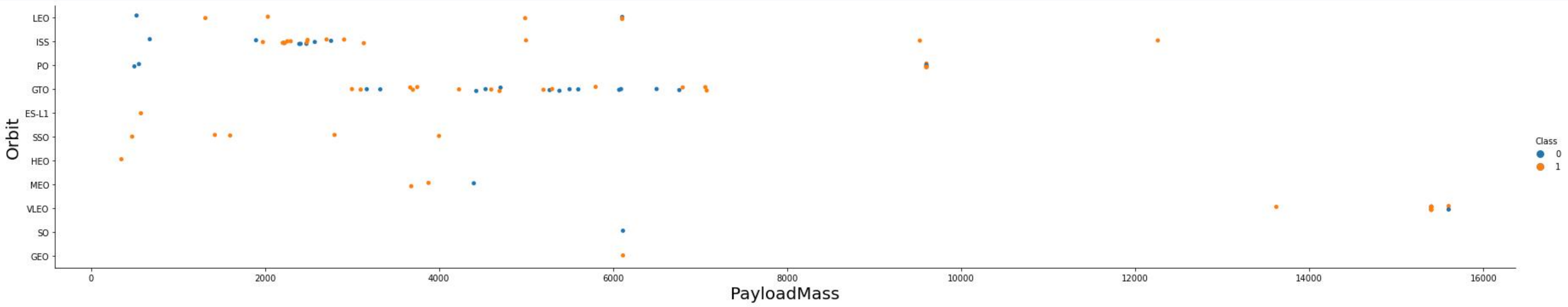


Flight Number vs. Orbit Type



- Although the success rate for all orbits appears to have improved over time, the VLEO orbit appears to be a new business opportunity due to the recent increase in its frequency.

Payload vs. Orbit Type

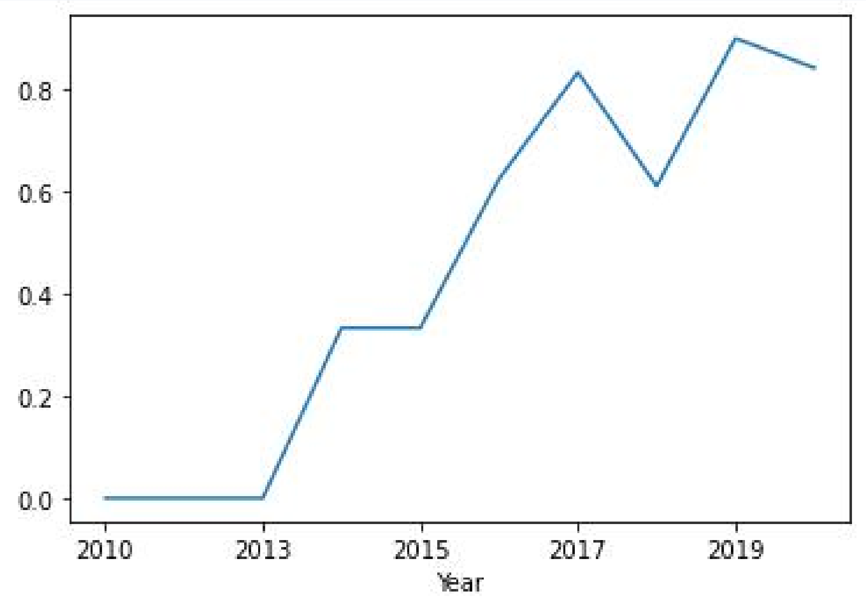


Apparently, there is no relation between payload and success rate to orbit GTO;

• ISS orbit has the widest range of payload and a good rate of success;

• There are few launches to the orbits SO and GEO.

Launch Success Yearly Trend



- The success rate began to rise in 2013 and will continue to rise until 2020. It appears that the first three years were a period of technological adjustment and advancement.

All Launch Site Names

- According to data, there are four launch sites:
- They are obtained by selecting unique occurrences of “launch_site” values from the dataset.

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`:

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

- Present your query result with a short explanation here

Total Payload Mass

- Total payload carried by boosters from NASA:

Total Payload (kg)
111.268

- Total payload calculated above, by summing all payloads whose codes contain ‘CRS’, which corresponds to NASA.

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1:
- Filtering data by the booster version above and calculating the average payload mass we obtained the value of 2,928 kg.

Avg Payload (kg)
2.928

First Successful Ground Landing Date

- First successful landing outcome on ground pad:

Min Date
2015-12-22

- By filtering data by successful landing outcome on ground pad and obtaining the minimum value for date, the first occurrence, which occurred on 12/22/2015, can be identified.

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters that successfully landed on a drone ship with a payload mass greater than 4000 but less than 6000.

- These four booster versions were selected using the filters listed above.

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

- We arrived at the summary above by grouping mission outcomes and counting records for each group.

Boosters Carried Maximum Payload

- Boosters which have carried the maximum payload mass

Booster Version (...)
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3

Booster Version
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

- These are the boosters that have transported the most payload mass in the dataset.

2015 Launch Records

- In 2015, failed landing results in drone ships, their booster variants, and launch site names

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

- The list above has the only two occurrences

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

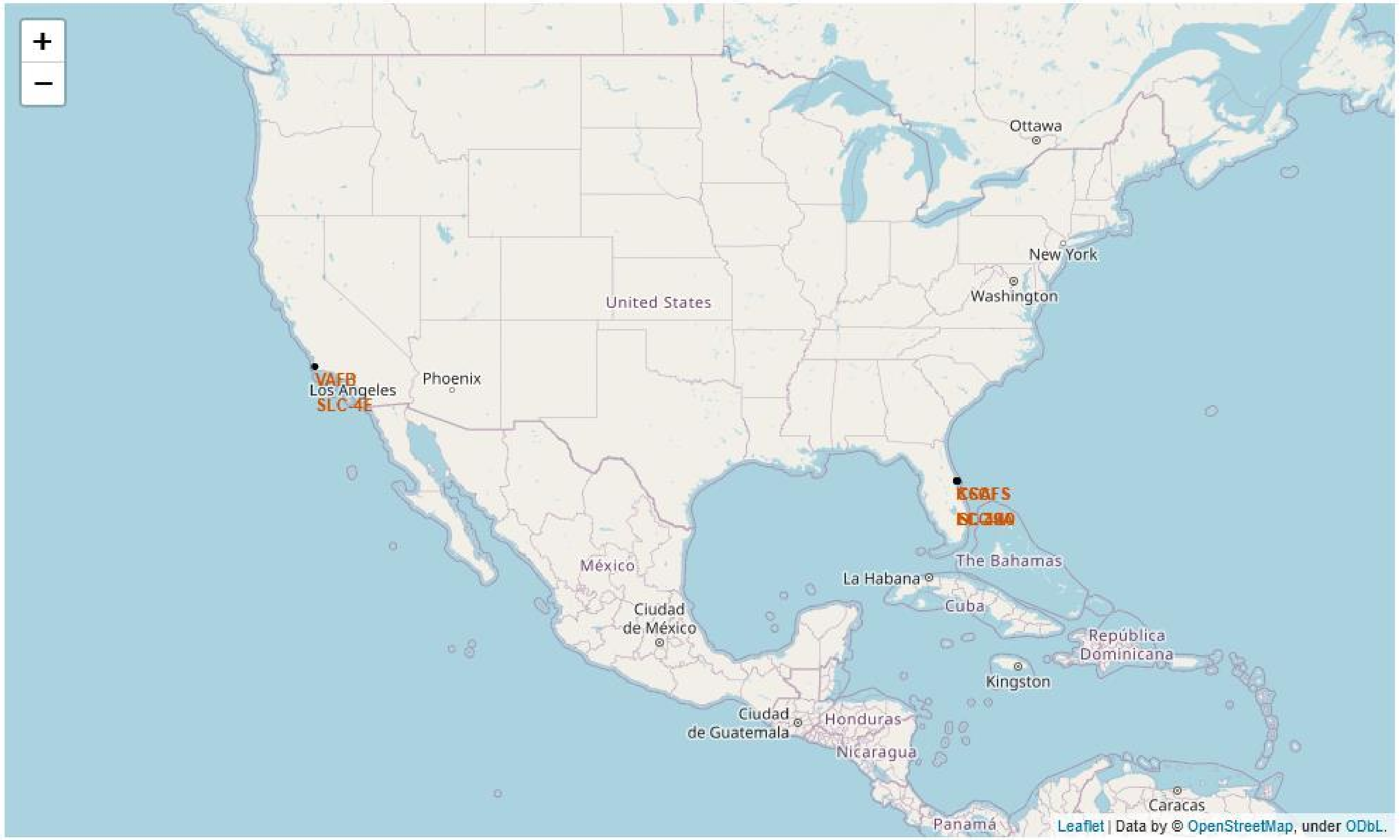
- Ranking of all landing outcomes between the date 2010-06-04 and 2017-03-20:
- This view of data alerts us that “No attempt” must be taken in account

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

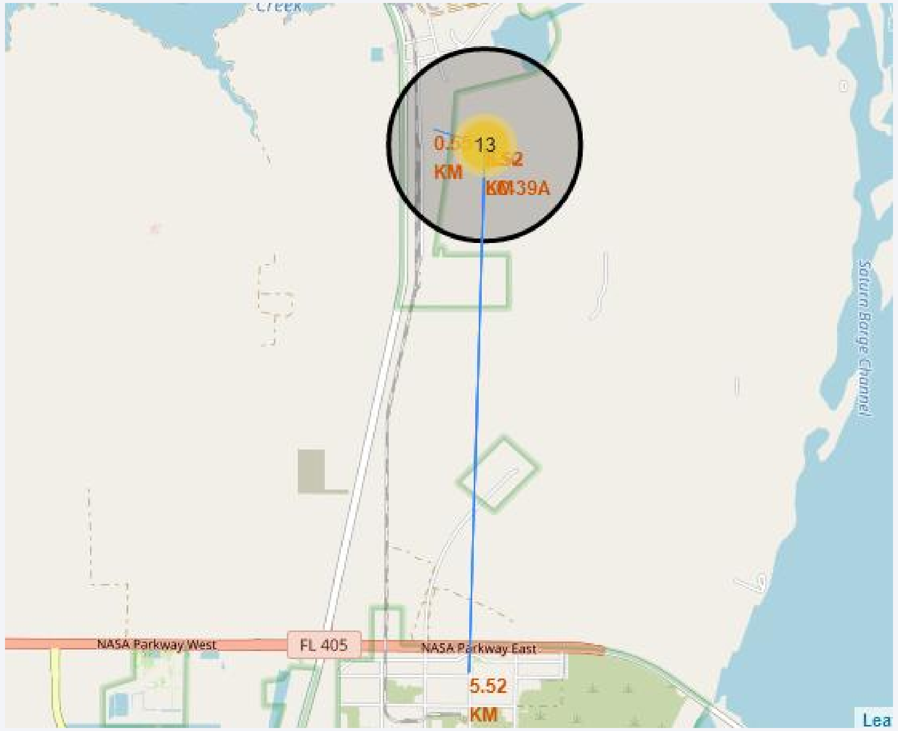


Launch sites are near sea, probably by safety, but not too far from roads and railroads.

Launch Outcomes by Site

- Example of KSC LC-39A launch site launch outcomes, Green markers indicate successful and red ones indicate failure.

Logistics and Safety

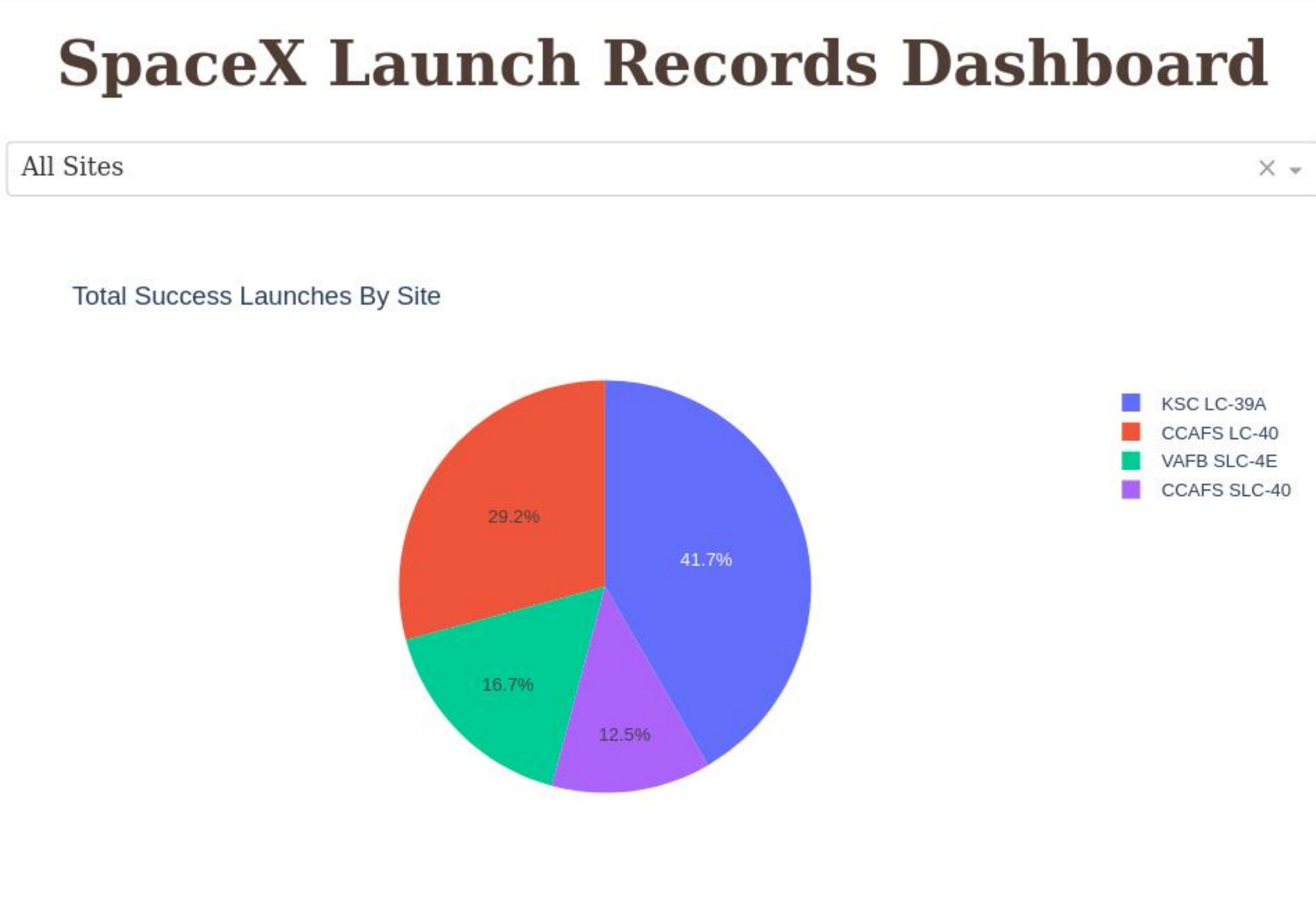


Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas.

Section 4

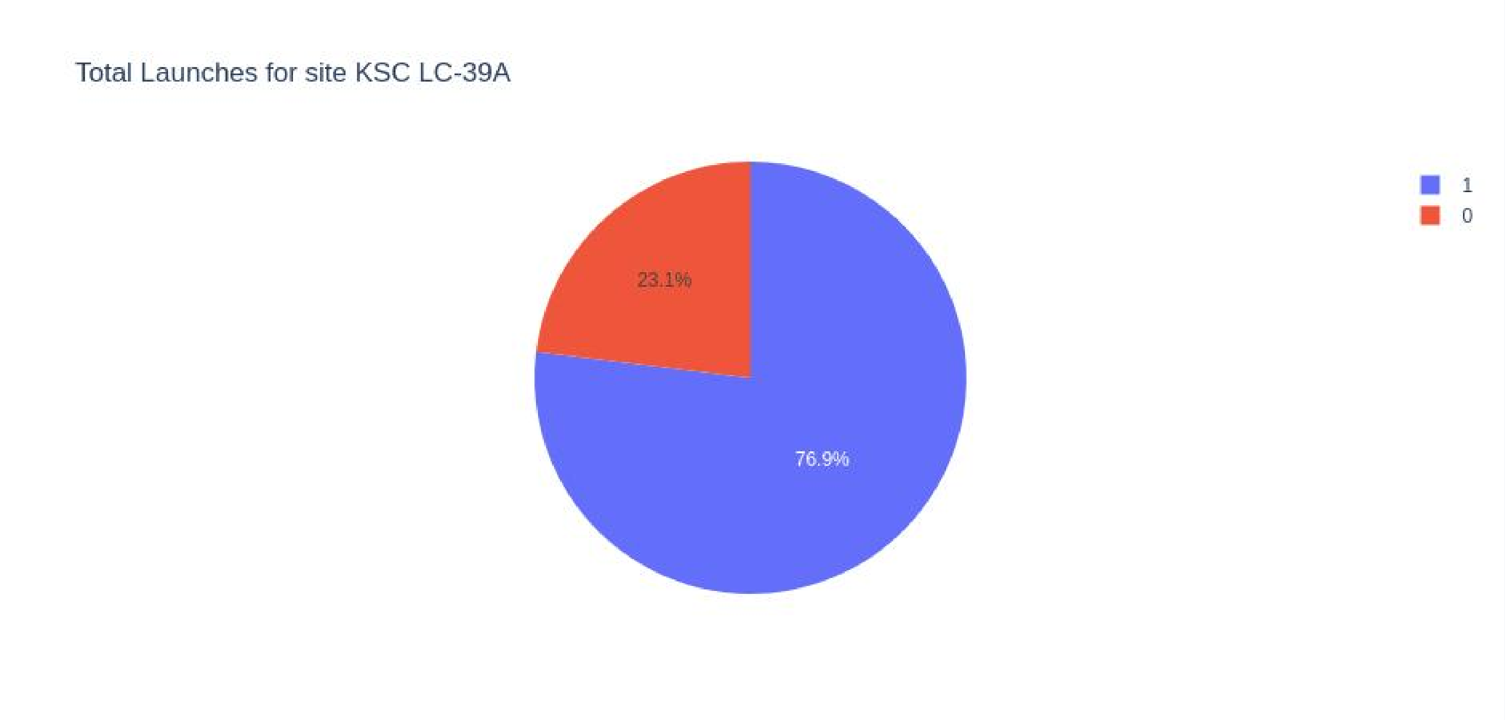
Build a Dashboard with Plotly Dash

Successful Launches by Site



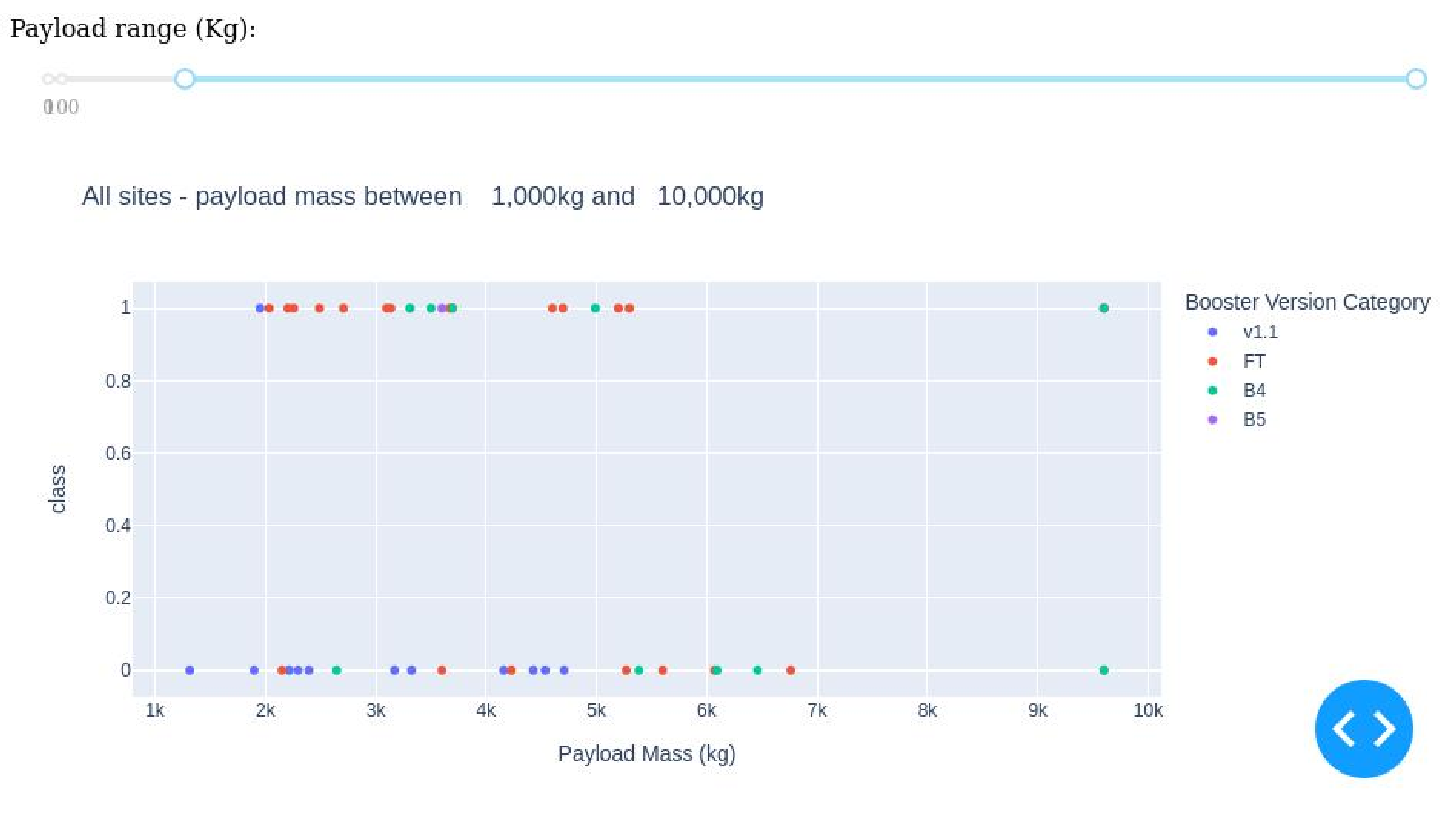
The location of mission launches appears to be a critical determinant in mission success.

Launch Success Ratio for KSC LC-39A



- 76.9% of launches are successful in this site.

Payload vs. Launch Outcome



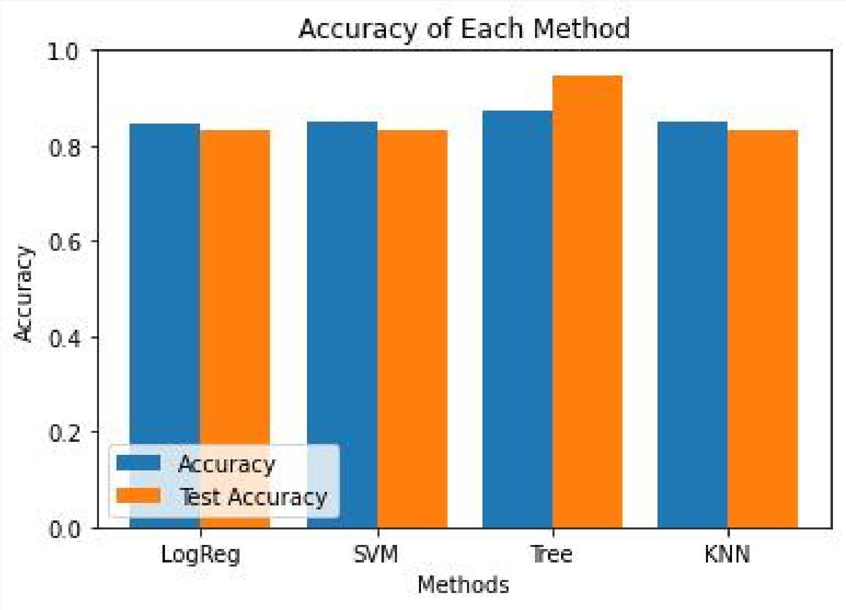
- The most successful combination is payloads under 6,000 kg and FT boosters.

Section 5

Predictive Analysis (Classification)

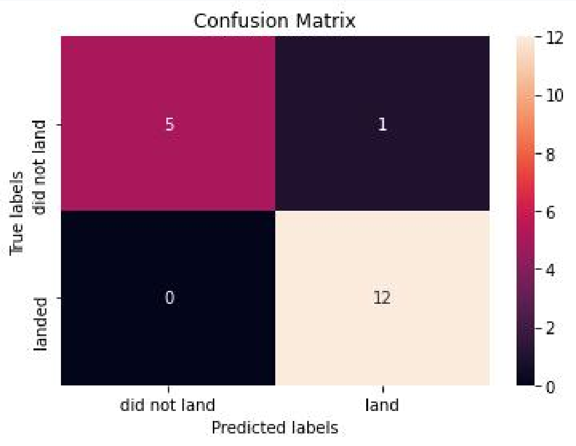
Classification Accuracy

- Four classification models were tested, and their accuracies are plotted beside;
- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87%



Confusion Matrix

- The Decision Tree Classifier's confusion matrix demonstrates its accuracy by displaying a large number of true positive and true negative results compared to false results.



Conclusions

- Profits can be increased by using Decision Tree Classifier to predict successful landings.
- Although the majority of mission outcomes are successful, successful landing outcomes appear to improve over time as processes and rockets evolve.
- Different data sources were analyzed, refining conclusions along the process;
- The best launch site is KSC LC-39A . Launches above 7,000kg are less risky;

Appendix

- As an improvement for model tests, it’s important to set a value to

  np.random.seed variable;

Thank you!