IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Manjusha NG
17-March-2023

Outline

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

Executive Summary

- The following methodologies were used to analyze data
  - Data collection using web scrapping and SpaceX API
  - Exploratory data analysis (EDA) , including data wrangling ,data visualization, interactive visual analytics.
  - Machine Learning Prediction
- Summary of all results
  - Collected Valuable data from various sources
  - Effective EDA to identify the feature which is best to predict the success of launchings
  - Machine learning prediction showed the best model to predict

Introduction

- Objective : To evaluate the Viability of the new company Space Y to compete with Space X

- Desirable Answers :
  - Estimate the total cost for launches, by predicting successful landings of the first stage of rockets.
  - Predict the best place to make launches

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data from Space X was obtained from two sources
    - SpaceX API (https://api.spacexdata.com/v4/rockets/)
    - Webscapping (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 the features.

Methodology

Executive Summary

- 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 that was collected until this step were normalized, divided to training and test dataset and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters

Data Collection

Data from Space X was obtained from two sources

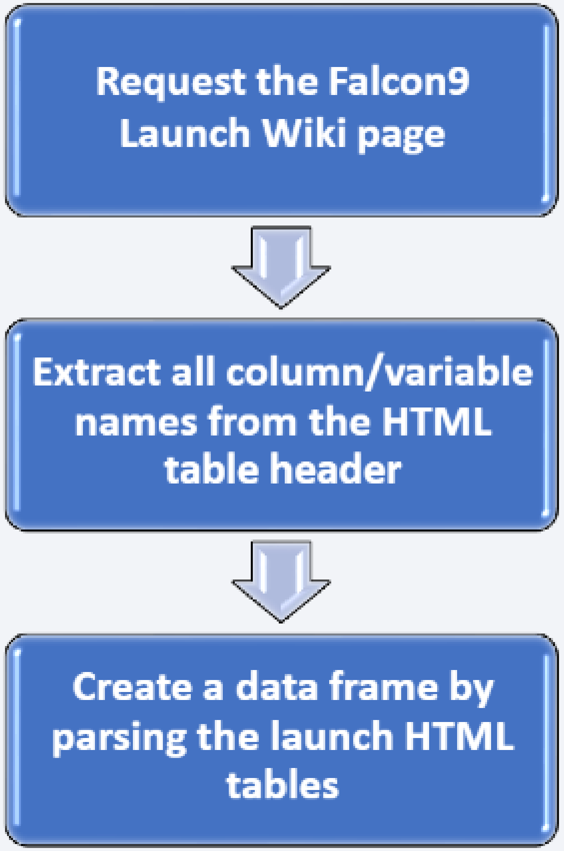
- SpaceX API (https://api.spacexdata.com/v4/rockets/)
- Webscapping (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)

Data Collection – SpaceX API

- Space X offers a public API from where data can be obtained and then used.
- This API was used as shown in the flowchart and data was persisted.
- Source Code: [SPACE-Y/Data Collection API.ipynb at master · manjushang-git/SPACE-Y (github.com)]

Data Collection - Scraping

- Data from space X launches can be obtained from Wikipedia.
- Data is obtained using web scrapping as shown in the flowchart.
- Source Code :[SPACE-Y/Data Collection with Web Scraping.ipynb at master · manjushang-git/SPACE-Y (github.com)]



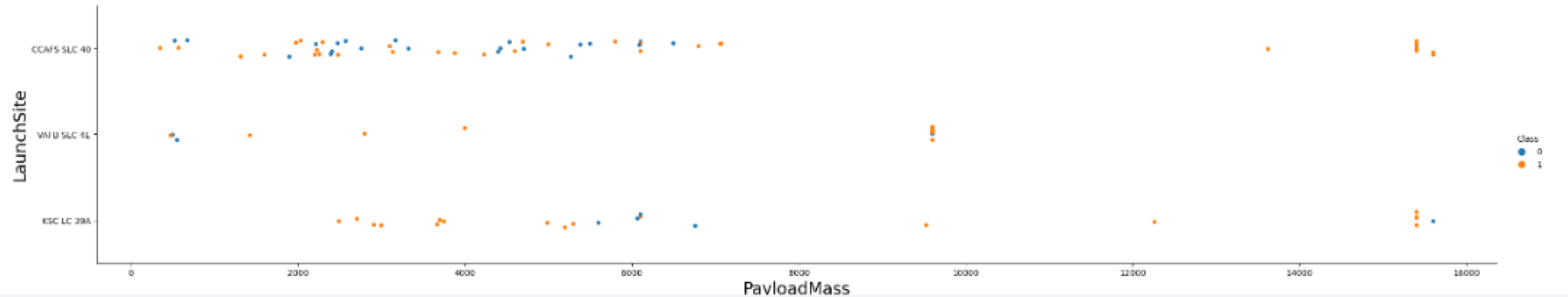
Data Wrangling

- Initially some Exploratory data analysis was performed on the dataset .
- The summary analysis such as Launches per site, Occurrences of each orbit and occurrences of mission outcome per orbit , were performed
- Finally, the landing outcome label was created from Outcome column.

- Source Code :[SPACE-Y/EDA.ipynb at master · manjushang-git/SPACE-Y (github.com)]

EDA with Data Visualization

- To explore data ,scatterplots and bar plots were used to visualize the relationship between pair of features.
  - **Visualization of the relationship between Payload and Launch Site**



- Source Code :[SPACE-Y/EDA with Data Visualization.ipynb at master · manjushang-git/SPACE-Y (github.com)]()

EDA with SQL

- Performed the following SQL queries.
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first successful landing outcome in ground pad was acheived.
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

EDA with SQL

- Performed the following SQL queries.
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
  - List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
  - Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order
- Source code :SPACE-Y/SQL.ipynb at master · manjushang-git/SPACE-Y (github.com)

Build an Interactive Map with Folium

Markers, circles, lines and marker clusters were used with folium maps.

- Markers indicate points like launch sites.
- Circles indicate highlighted areas around specific coordinates , like NASA Johnson space Center.
- Marker clusters indicate groups of events in each coordinate ,like launches ina launch site
- Lines are used to indicate the distances between two coordinates

- Source Code : [SPACE-Y/Interactive Visual Analytics with Folium lab.ipynb at master · manjushang-git/SPACE-Y (github.com)]

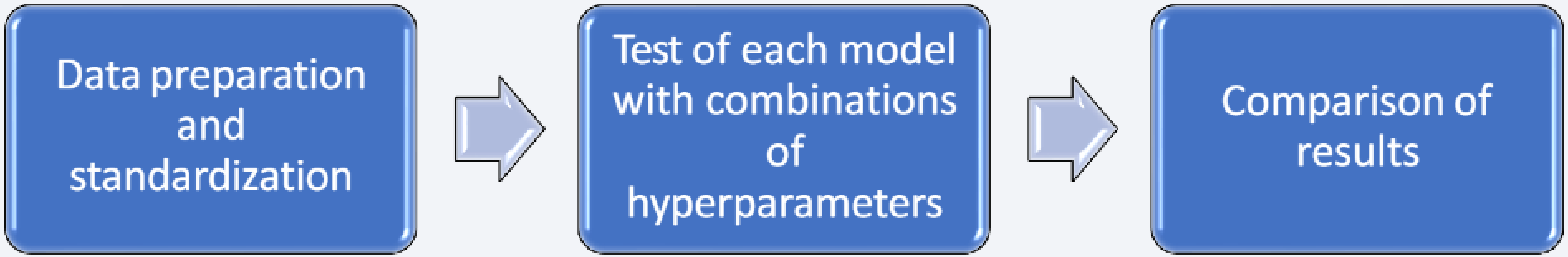
Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize data
  - Percentage of Launches by site
  - Payload range
- This combination allowed to quickly analyze the relation between payloads and launch sites ,helping to identify where is best place to launch according to payloads.

- Source Code : SPACE-Y/spacex_dash_app.py at master · manjushang-git/SPACE-Y (github.com)

Predictive Analysis (Classification)

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



- Source Code :[SPACE-Y/SpaceX_Machine Learning Prediction_Part_5.ipynb at master · manjushang-git/SPACE-Y (github.com)]

Results

- Exploratory data analysis results
  - Space X used four different launch sites
  - The first launches were done to NASA and Space X itself
  - The average payload of F9 v1.1 Booster is 2928 kg.
  - The first success landing outcome happened in 2015 , five year after the first launch
  - Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average
  - Almost 100% of mission outcomes were successful
  - Two booster versions were failed at landing in drone ships in 2015,F9 v1.1 B1012 and F9 v1.1 B1015
  - The number of landing outcomes become better as years passed

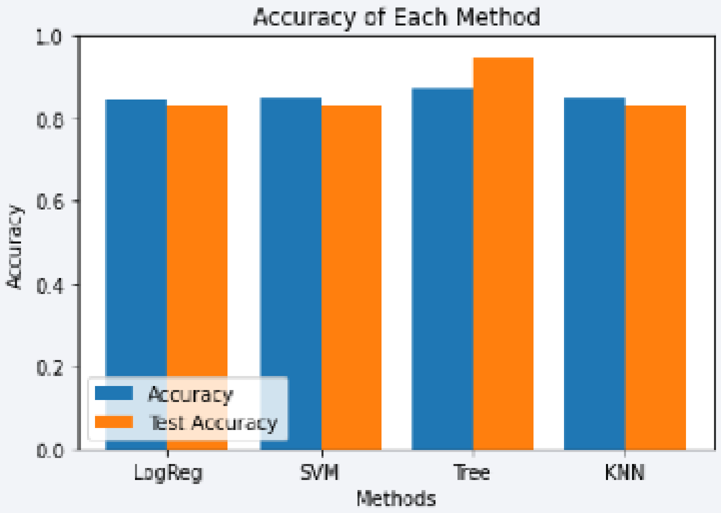
Results

Interactive analytics demo in screenshots

Results

- Predictive analysis results

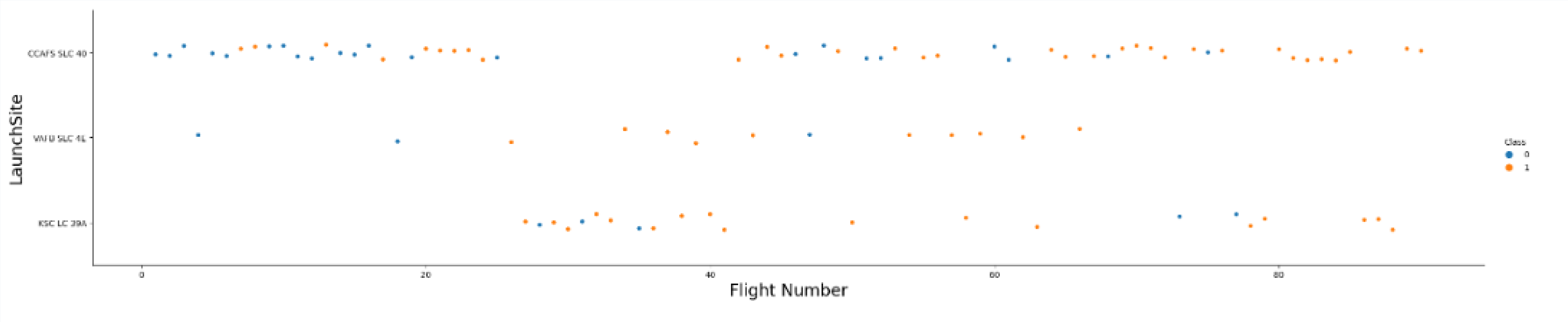
Predictive analysis showed that Decision Tree Classifier is the best model to predict successful landings.



Section 2

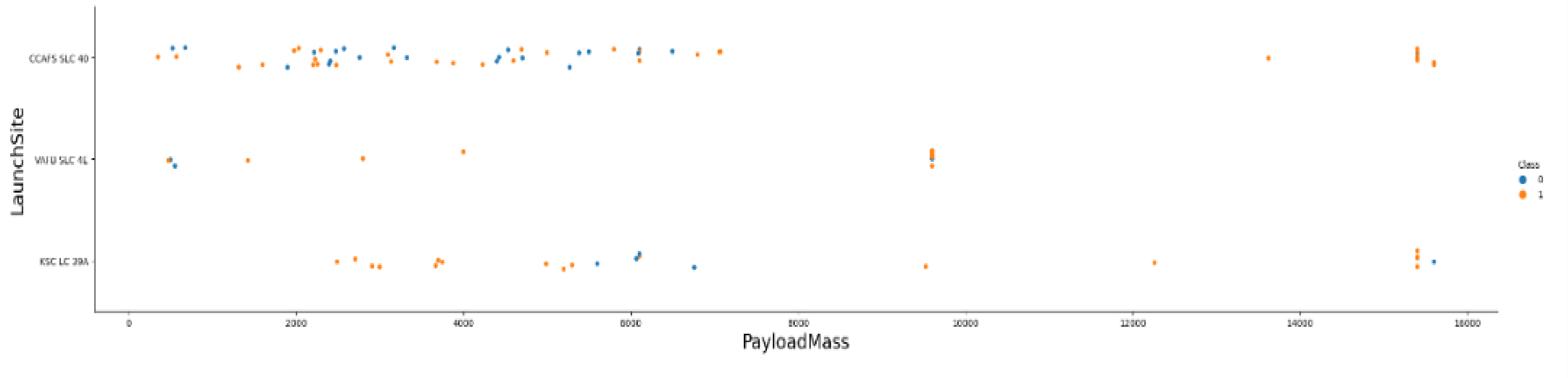
Insights drawn from EDA

Flight Number vs. Launch Site



- According to the plot above ,its possible to verify that the best launch site nowadays is CCAF5 SLC40, where most of the recent launches were successful
- In second place,VAFB SLC 4E and third place KSCLC 39A
- Its also possible to see that the general success rate improved over time.

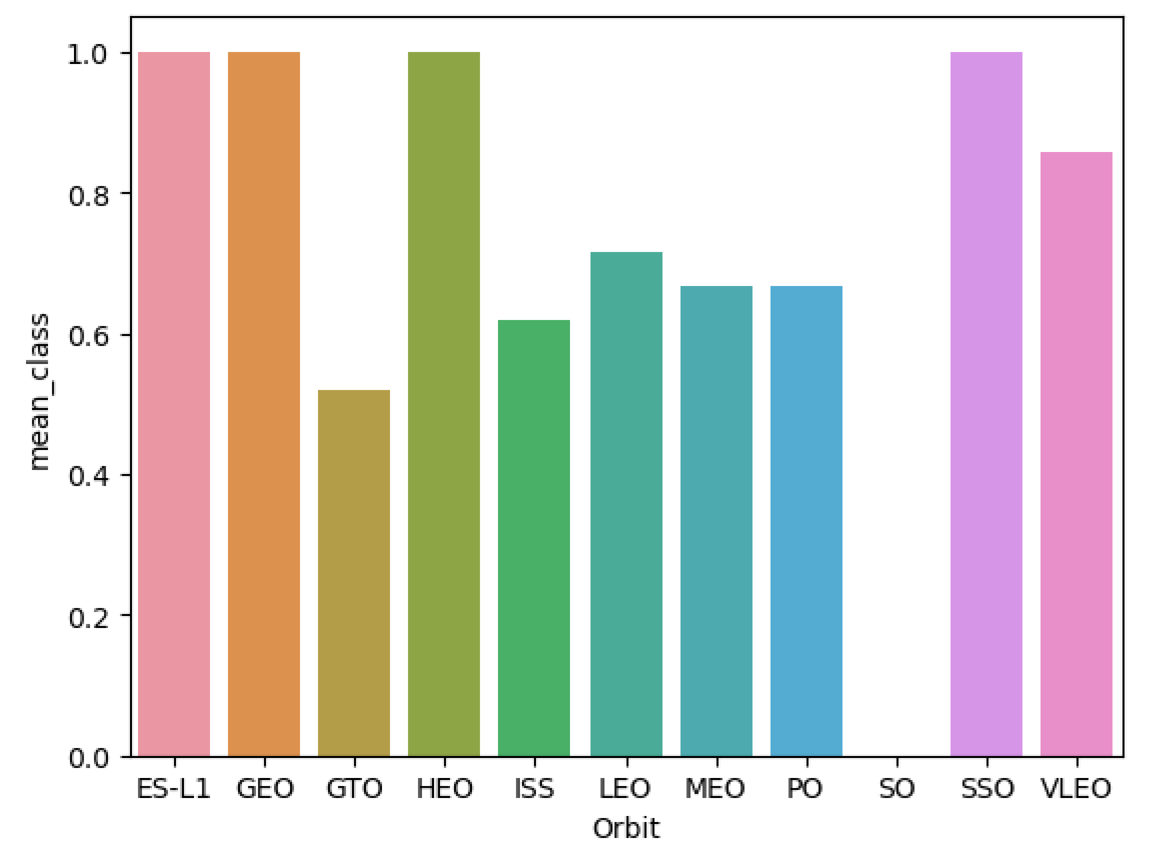
Payload vs. Launch Site



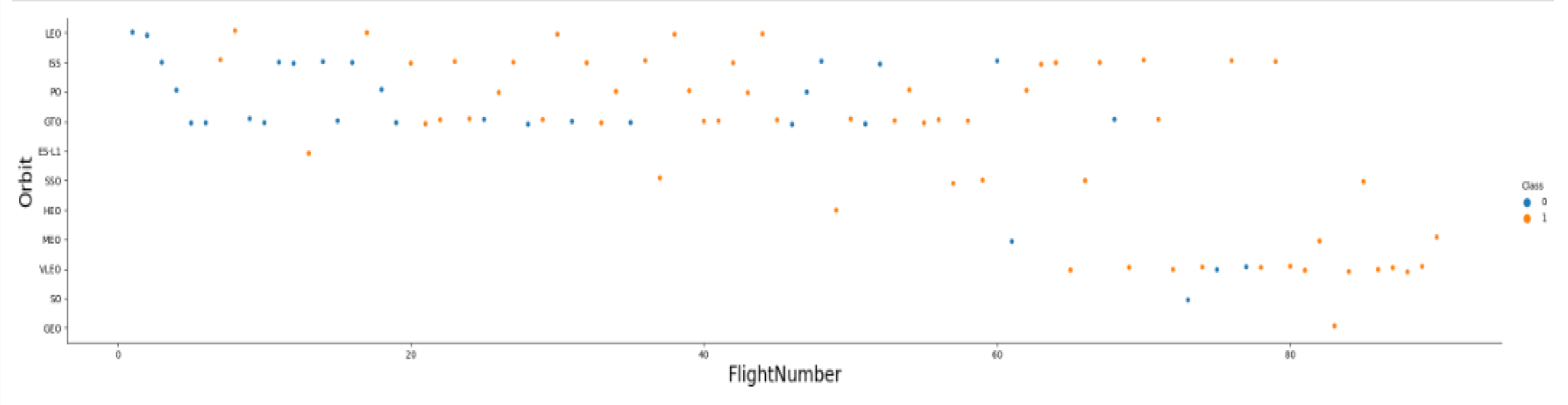
- Payloads over 9000kg has excellent success rate
- For the VAFB-SLC launch site there are no rockets launched for heavy pay load mass(greater than 10000).

Success Rate vs. Orbit Type

- The biggest Success Rates happens to orbits:
  - ES-L1
  - GEO
  - HEO
  - SSO

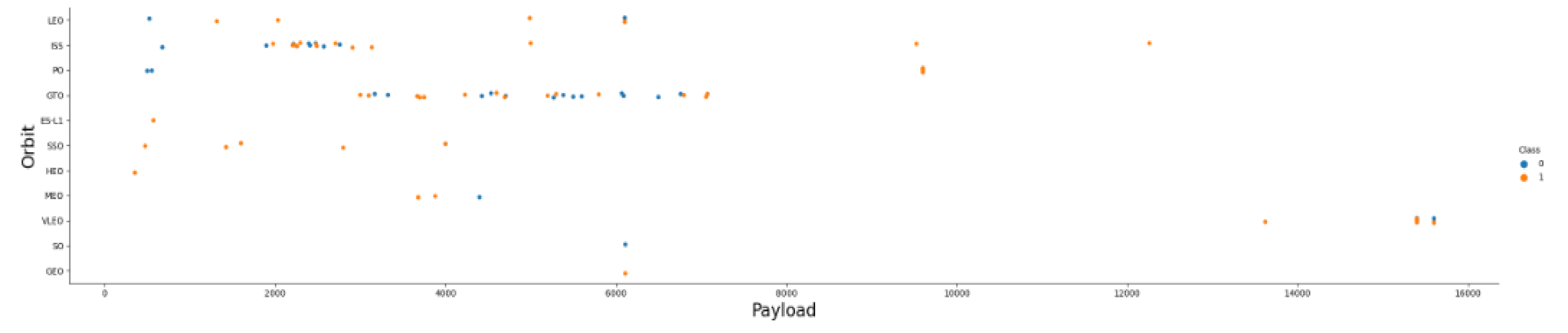


Flight Number vs. Orbit Type



- The LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

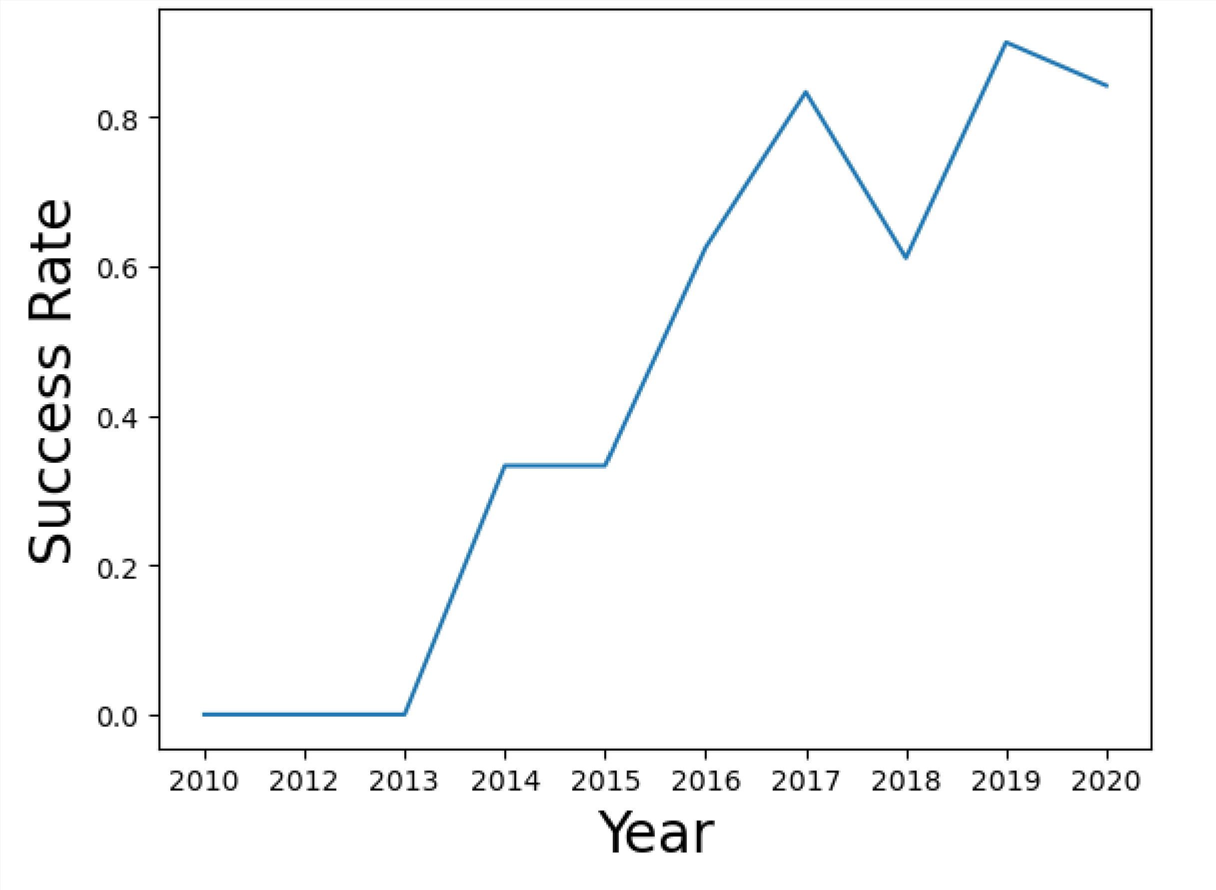
Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend

- The success rate since 2013 kept increasing till 2020



All Launch Site Names

- According to data, there are four launch sites

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

- They are obtained by selecting unique occurrences of “launch_site” values from the dataset.

Launch Site Names Begin with 'CCA'

- Find 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	**CCA**FS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	**CCA**FS 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	**CCA**FS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	**CCA**FS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	**CCA**FS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attemp

- Here we can see five samples of Cape Canaveral launches.

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:

Avg Payload (kg)
2.928

- Filtering data by the booster version above and calculating the average payload mass we obtained the value of 2,928 kg.

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad

Min Date
2015-12-22

- By filtering data by successful landing outcome on ground pad and getting the minimum value for date it's possible to identify the first occurrence, that happened on 12/22/2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

- Selecting distinct booster versions according to the filters above, these 4 are the result

Total Number of Successful and Failure Mission Outcomes

- The total number successful and failure mission outcomes

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

- Grouping mission outcomes and counting records for each group led us to the summary above.

Boosters Carried Maximum Payload

- The names of the booster 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 which have carried the maximum payload mass registered in the dataset.

2015 Launch Records

- List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

- Both the Failed landing happened on the same Launch site.

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

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

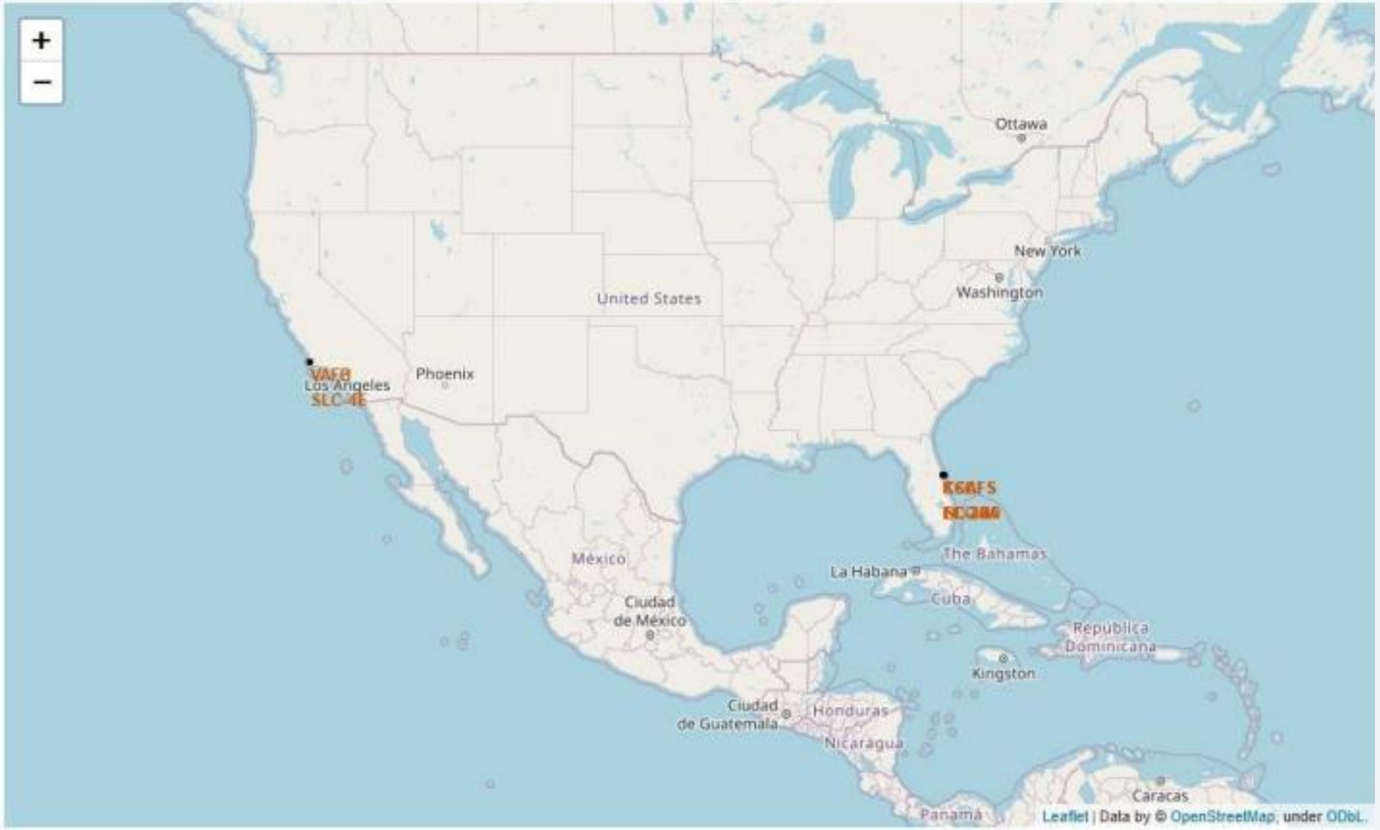
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- This view of data alerts us that “No attempt” must be taken in account.

Section 3

Launch Sites Proximities Analysis

All Launch Sites



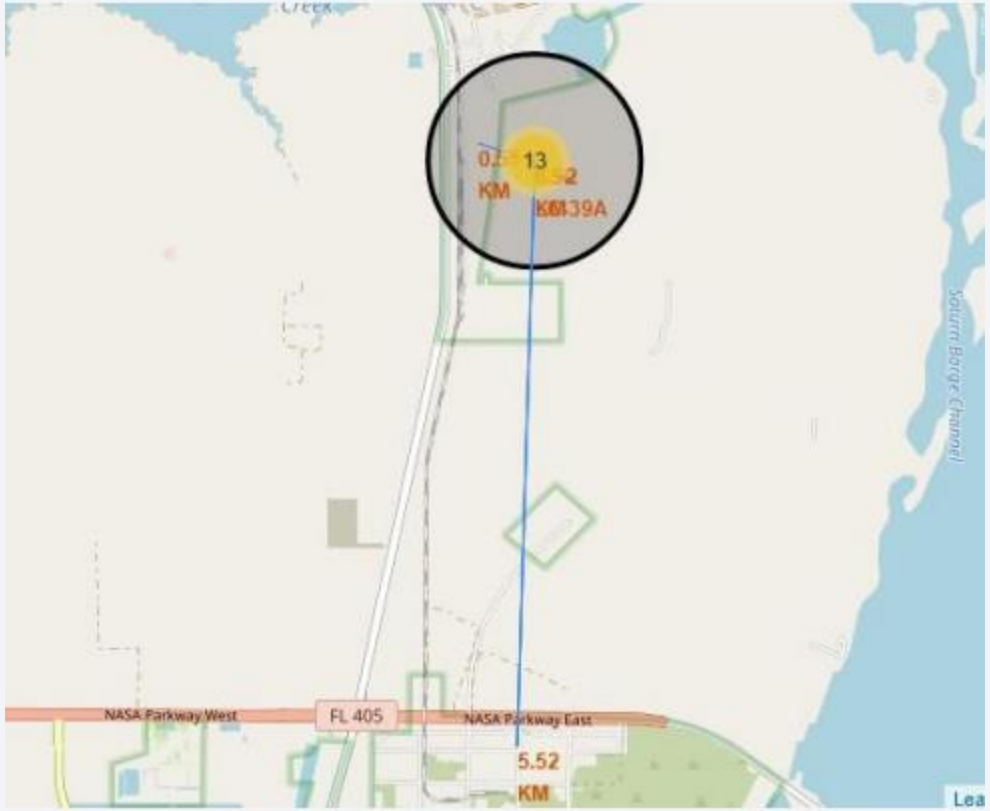
- Launch sites are near sea, probably by safety, but not too far from roads and railroads.

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

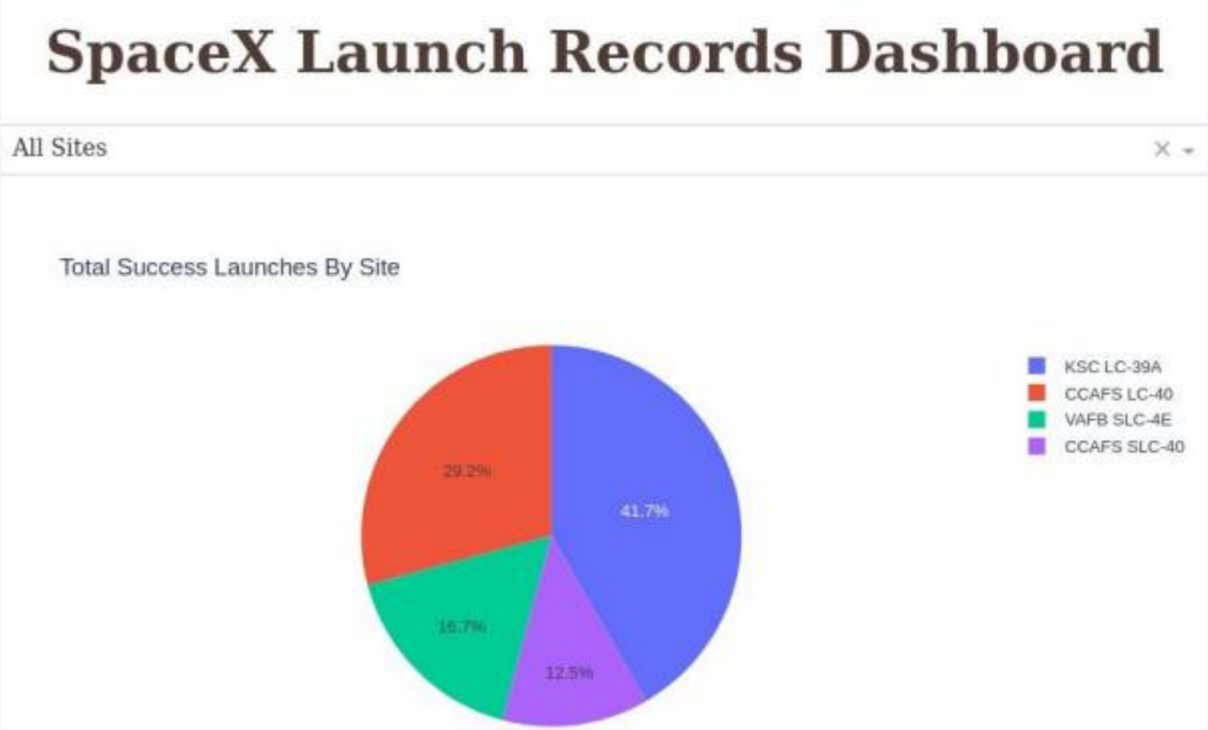


- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas.

Section 4

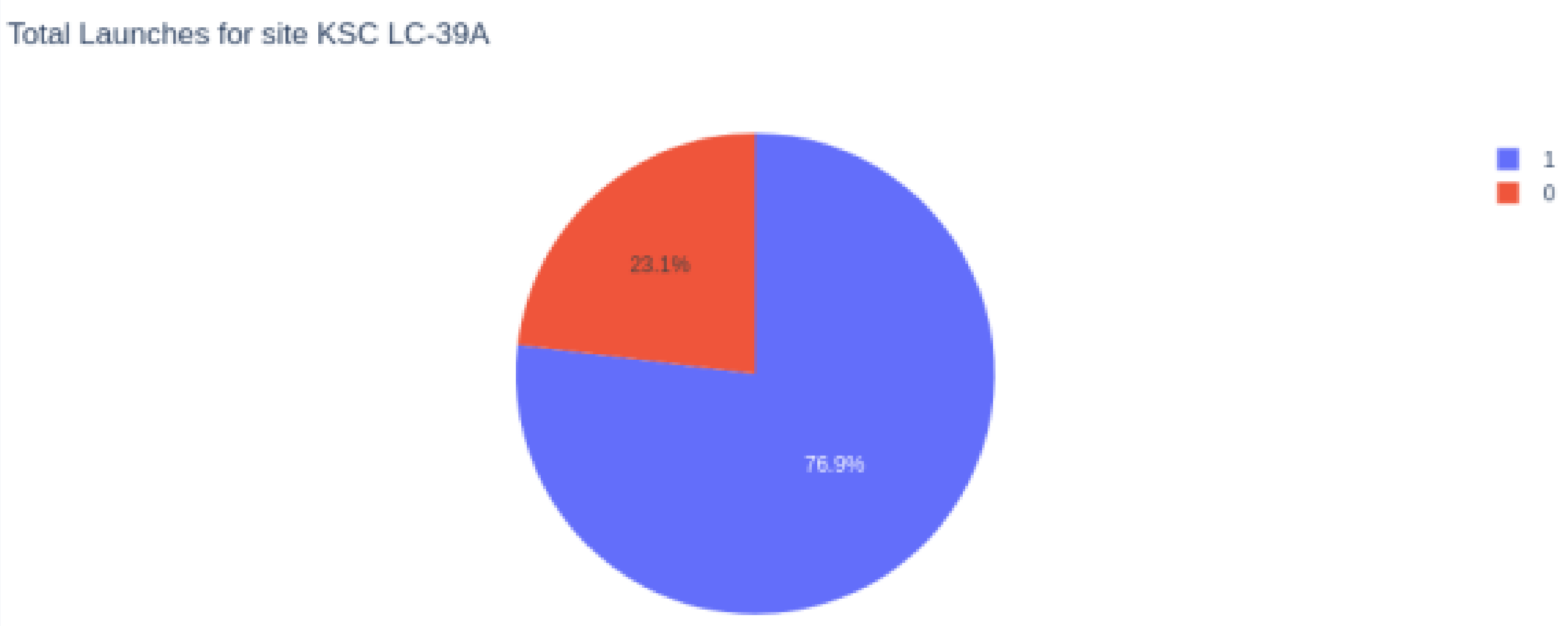
Build a Dashboard with Plotly Dash

Successful Launches by Site



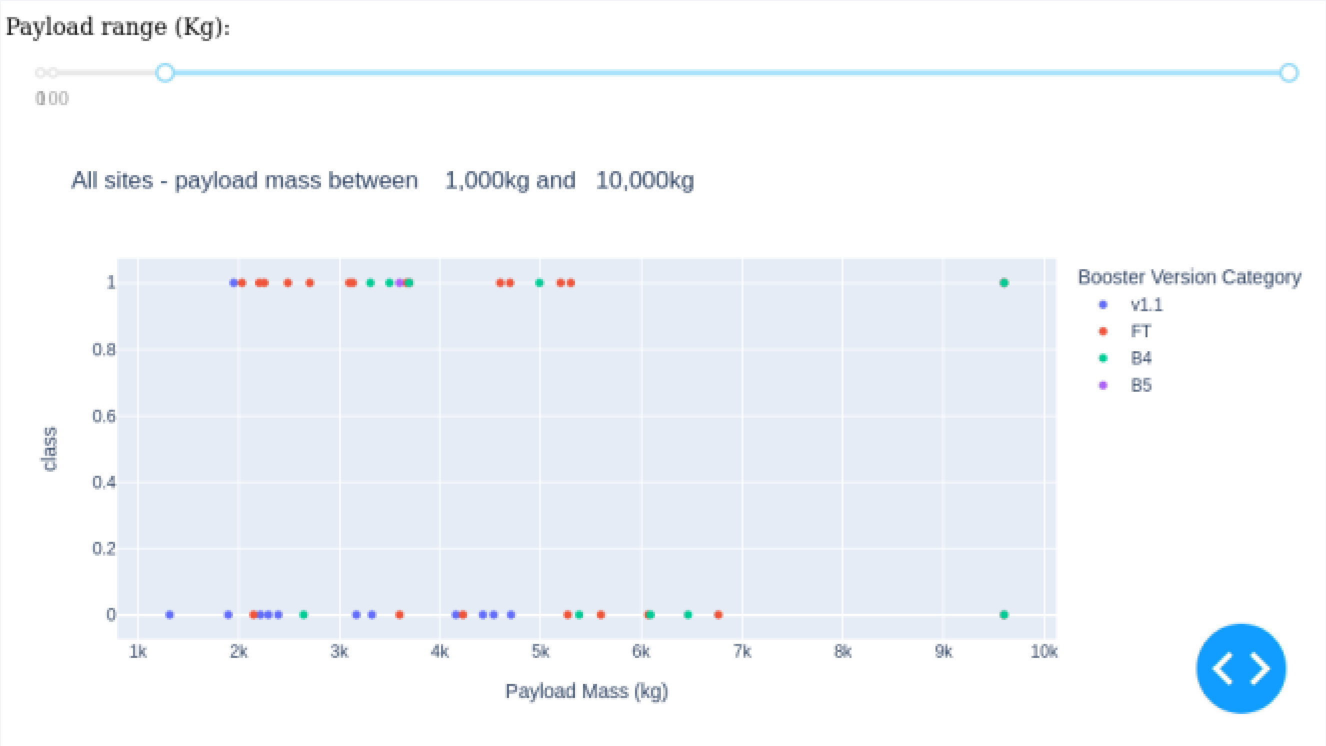
- The place from where launches are done seems to be a very important factor of success of missions.

Launch Success Ratio for KSCLC-39A



- 76.9% of launches are successful in this site.

Payload vs. Launch Outcome



- Payloads under 6,000kg and FT boosters are the most successful combination.

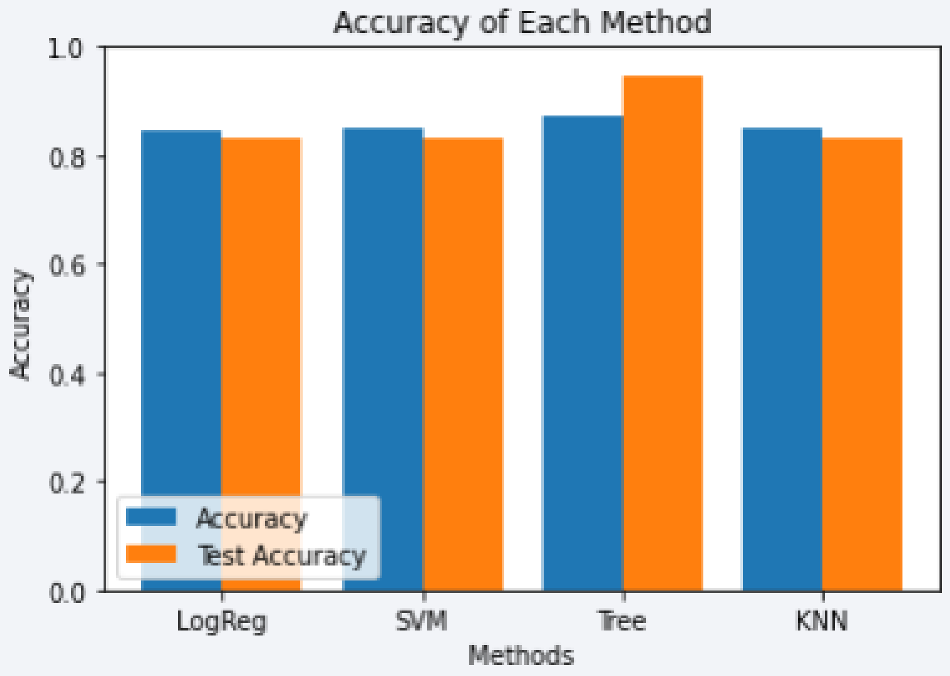
Section 5

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 of Decision Tree Classifier

- Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.

Conclusions

- Different data sources were analyzed, refining conclusions along the process;
- The best launch site is KSCLC-39A;
- Launches above 7,000kg are less risky;
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets;
- Decision Tree Classifier can be used to predict successful landings and increase profits.

Appendix

- As an improvement for model tests, it’s important to set a value to np.random.seed variable;
- Folium didn’t show maps on Github, so I took screenshots.

Thank you!