IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ehab Elsayed
2022-12-22

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection: using SpaceX API and web scraping.
  - Exploratory Data Analysis (EDA): including data wrangling, data visualization and interactive visual analytics.
  - Machine Learning Prediction.
- Summary of all results
  - EDA allowed to identify which features are the best to predict success of launchings
  - Machine Learning Prediction showed the best model to predict which characteristics are
  - important to drive this opportunity by the best way, using all collected data.

Introduction

The main object is to evaluate the ability of new company Space Y to compete with Space X company.

To do so we need to answer the following questions:

- What is the best way to estimate the total cost for launches?
- Where is the best place to make launches?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

  Data about Space X was obtained from two sources:

  - Space X API: https://api.spacexdata.com/v4/rockets/
  - Web Scraping: https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches

- Perform data wrangling
  - creating a landing outcome label based on outcome data after summarizing and analyzing features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

Methodology

Executive Summary

- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

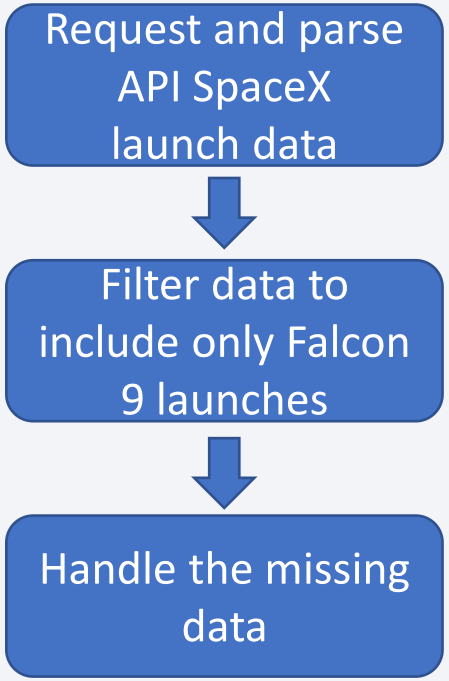
Normalize collected data, divided in training and test data sets and evaluated by four different classification models, analyze the accuracy of each model evaluated using different combinations of parameters.

Data Collection

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

Data Collection – SpaceX API

- SpaceX offers a public API where data can be obtained and used
- API was used according to the flowchart

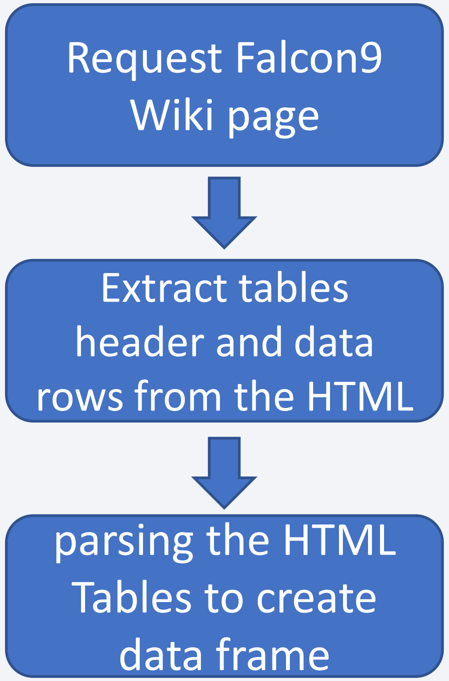


Source code: https://github.com/devalees/Applied-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- SpaceX launches data can also be obtained from Wikipedia
- Data are downloaded from Wikipedia according to the flowchart

Source code: https://github.com/devalees/Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb



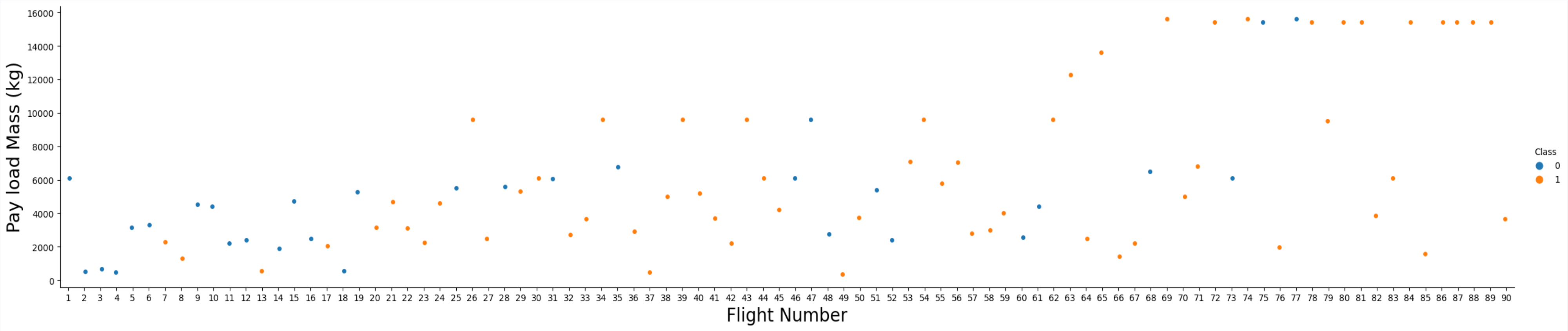
Data Wrangling

- Initiated Exploratory Data Analysis (EDA) was performed on the dataset.
- Then summaries launches per site, occurrences of each orbit and calculate occurrences of mission outcome per orbit type
- Finally, create the landing outcome label from Outcome column.

Source code: https://github.com/devalees/Applied-Data-Science-Capstone/blob/main/Data-wrangling.ipynb

EDA with Data Visualization

- Using scatterplots and barplots to visualize the relationship between pair of features :
- 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/devalees/Applied-Data-Science-Capstone/blob/main/Complete-the-EDA-with-Visualization-lab.ipynb

EDA with SQL

Performing some SQL queries as the following:

- 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 achieved.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- 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: https://github.com/devalees/Applied-Data-Science-Capstone/blob/main/Complete-the-EDA-with-SQL-lab.ipynb

Build an Interactive Map with Folium

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- For each launch site, add a Circle object based on its coordinate (Lat, Long) values. In addition, add Launch site name as a popup label
- Draw a line between a launch site to the selected coastline point
- Display the distance between coastline point and launch site
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Source code: https://github.com/devalees/Applied-Data-Science-Capstone/blob/main/Interactive-Visual-Analytics-with-Folium-lab.ipynb

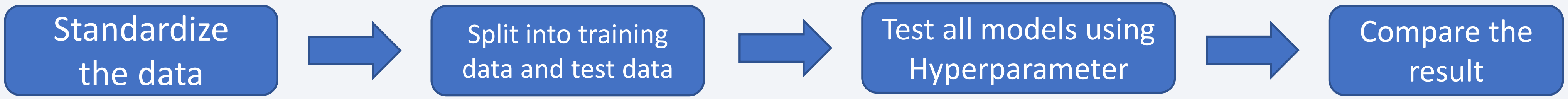
Build a Dashboard with Plotly Dash

- Use pie chart to show the total successful launches count for all sitesExplain why you added those plots and interactions.
- Use scatter chart with slider to show the correlation between payload and launch success, so we can identify the best place to launch based on payloads.
- This dashboard allows to analyze the relationships between payloads and launch sites

Source code: https://github.com/devalees/Applied-Data-Science-Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Find the method performs best using Hyperparameter for SVM, Classification Trees and Logistic Regression



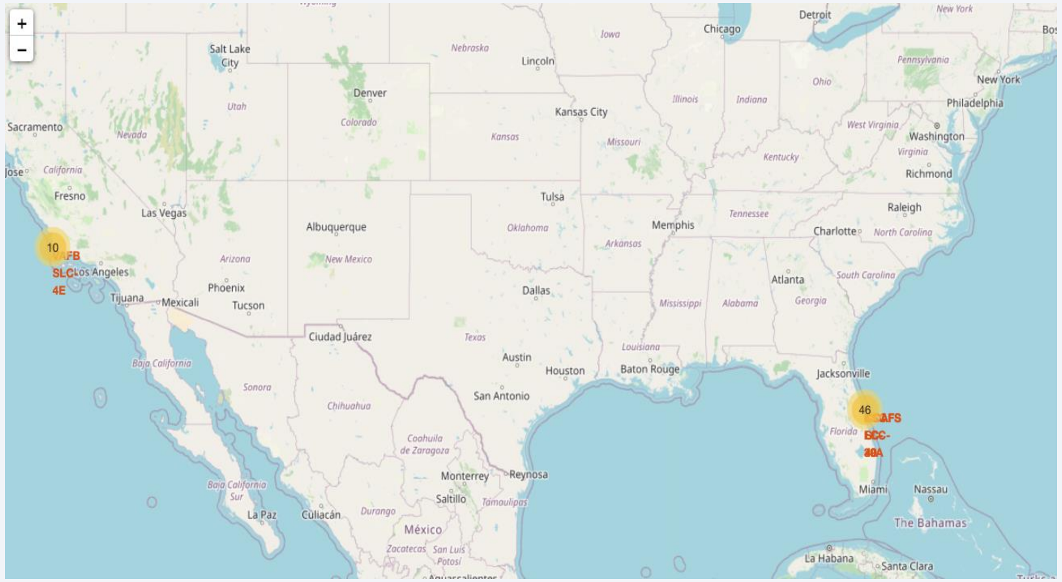
Source code: https://github.com/devalees/Applied-Data-Science-Capstone/blob/main/Machine-Learning-Prediction-lab.ipynb

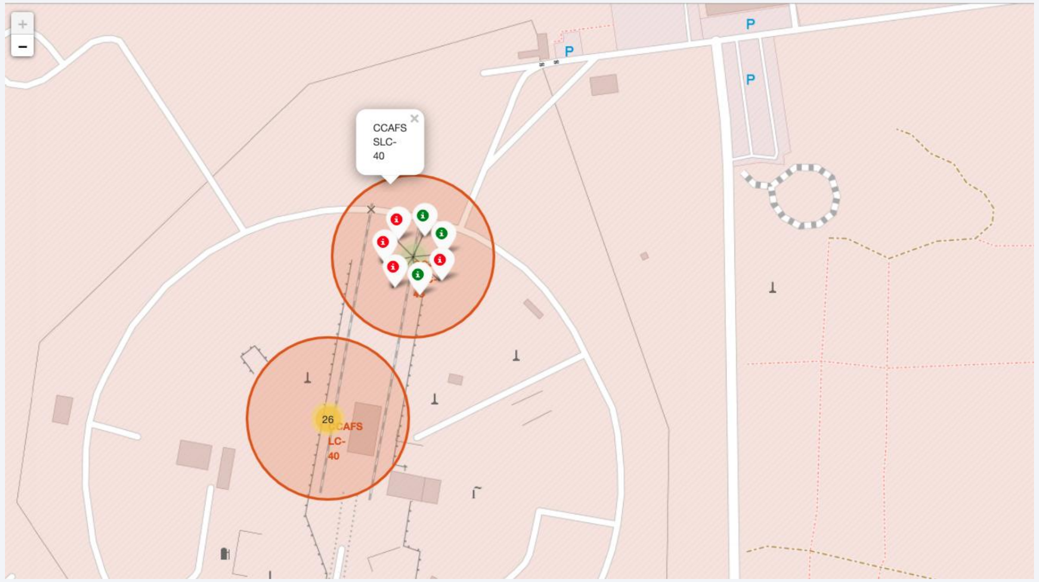
Results

- Exploratory data analysis results:
  - Space X uses 4 different launch sites
  - The average payload of F9 v1.1 booster is 2,928 kg
  - The first success landing outcome happened in 2015 fiver 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 failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015

Results

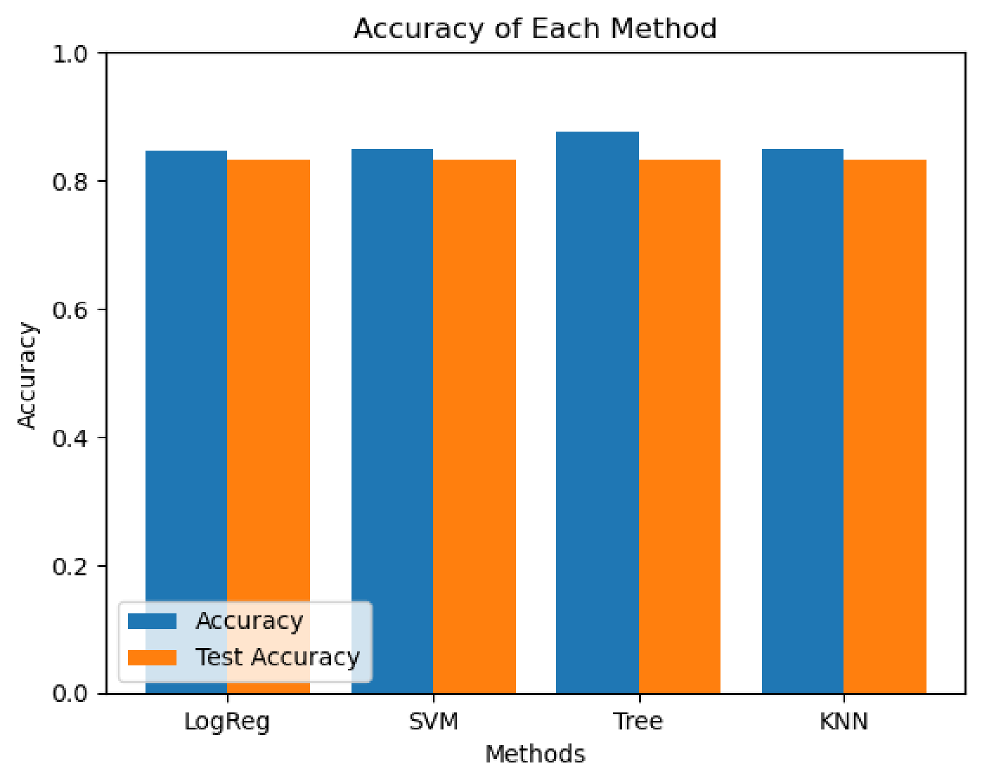
Using interactive analytics to identify that launch sites use to be in safety places, near sea.





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%

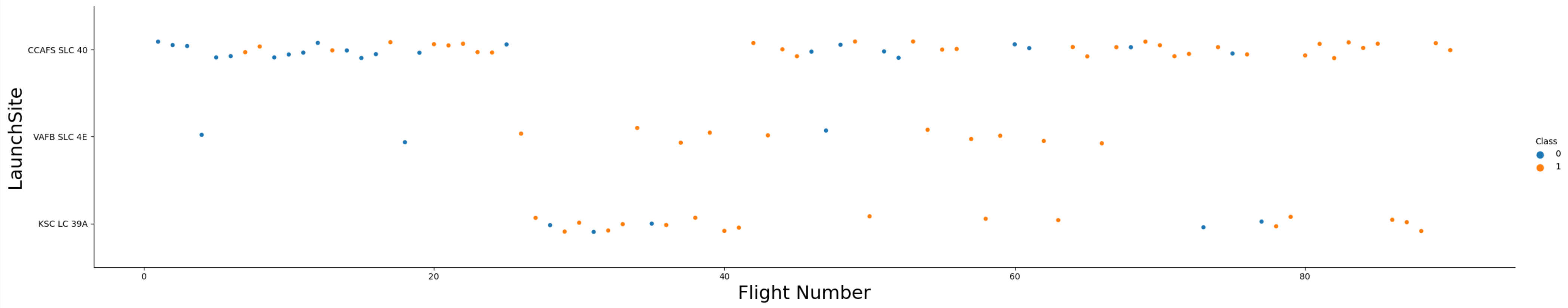


Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

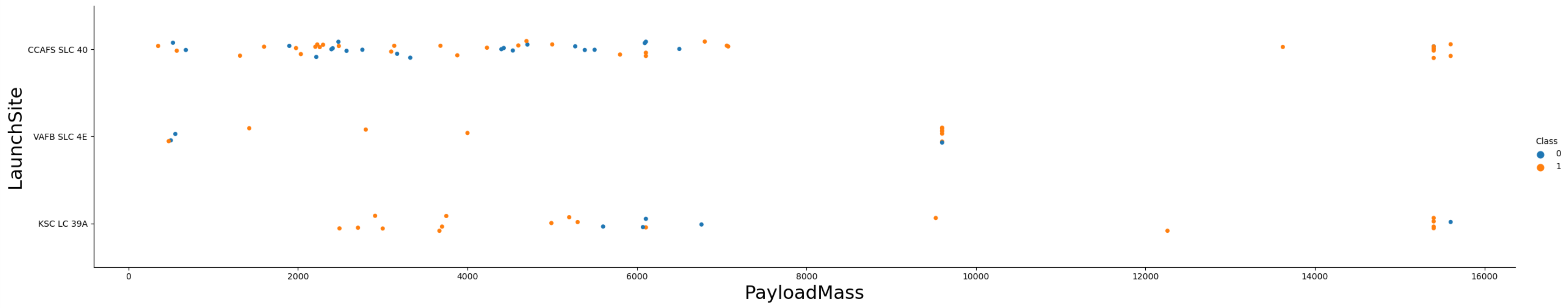
Number vs. Launch Site



- According to the plot above, the most of recent launches were successful site is CCAF5 SLC 40
- In second place VAFB SLC 4E and third place KSC LC 39A

Payload vs. Launch Site

Payload vs. Launch Site

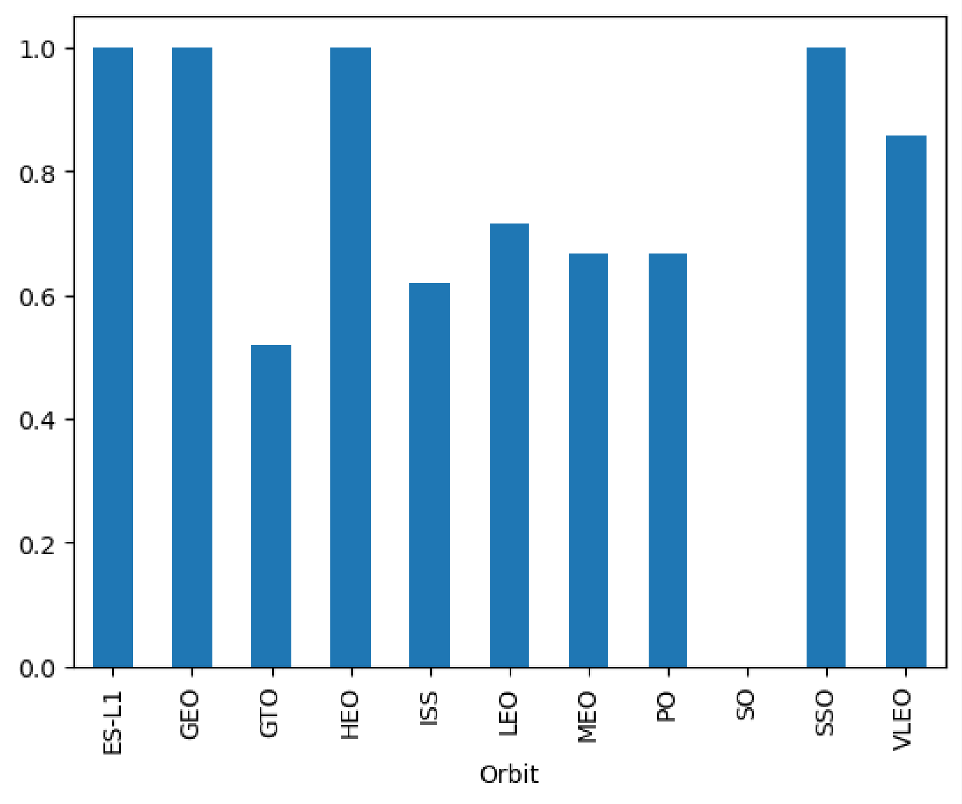


- Payloads over 9,000kg (about the weight of a school bus) have excellent success rate
- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

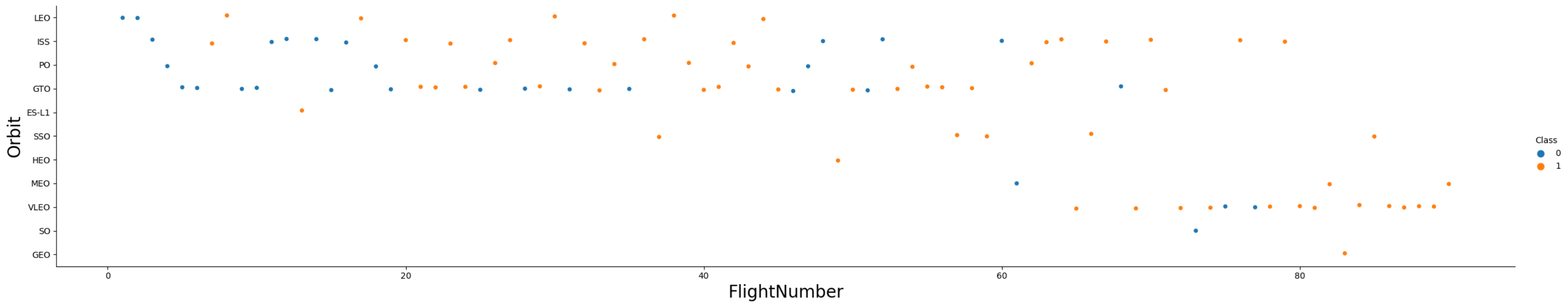
Success Rate vs. Orbit Type

The biggest success rates happens to orbits:

- ES-L1
- GEO
- HEO
- SSO

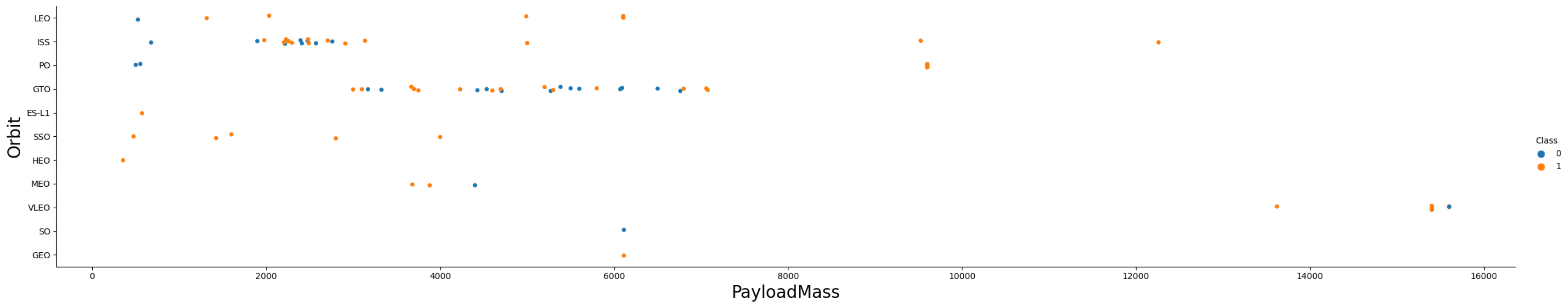


Flight Number vs. Orbit Type



- Success rate improved over time to all orbits

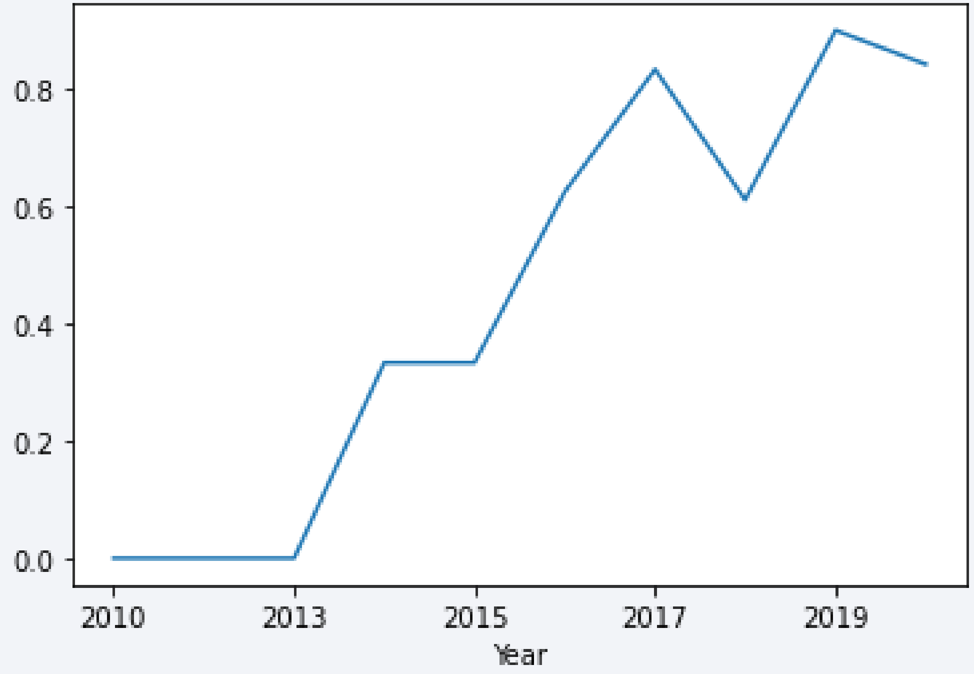
Payload vs. Orbit Type



- There seems to be no relationship between payload and success rate when in GTO orbit
- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS

Launch Success Yearly Trend

- Success rate started increasing in 2013 and kept until 2020
- It seems that the first three years were a period of adjusts and improvement of technology



All Launch Site Names

- There are four lunch sites according to the data:

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

The result above obtained by selecting unique values from “launch_site” 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
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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)
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-12	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170	GTO	SES	Success	No attempt

Total Payload Mass

- Total payload carried by boosters from NASA

total_payload
56479

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
3676

Filtering the data by F9, then average the payload mass

First Successful Ground Landing Date

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

first_success_gp
2017-01-05

Sorting data set by date after filtering the successful landing outcome

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

booster_version
F9 FT B1031.2
F9 FT B1022

Distinct booster versions, after filtering payload mass that is greater than 4000 but less than 6000 and drone ship is succussed

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

mission_outcome	qty
Success	44
Success (payload status unclear)	1

Counting the mission outcomes, then group them.

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass

booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1058.3
F9 B5 B1060.2

Distinct the booster versions that have maximum payload mass

2015 Launch Records

- 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

Selecting booster versions and launch site where the drone ship was failed in 2015

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

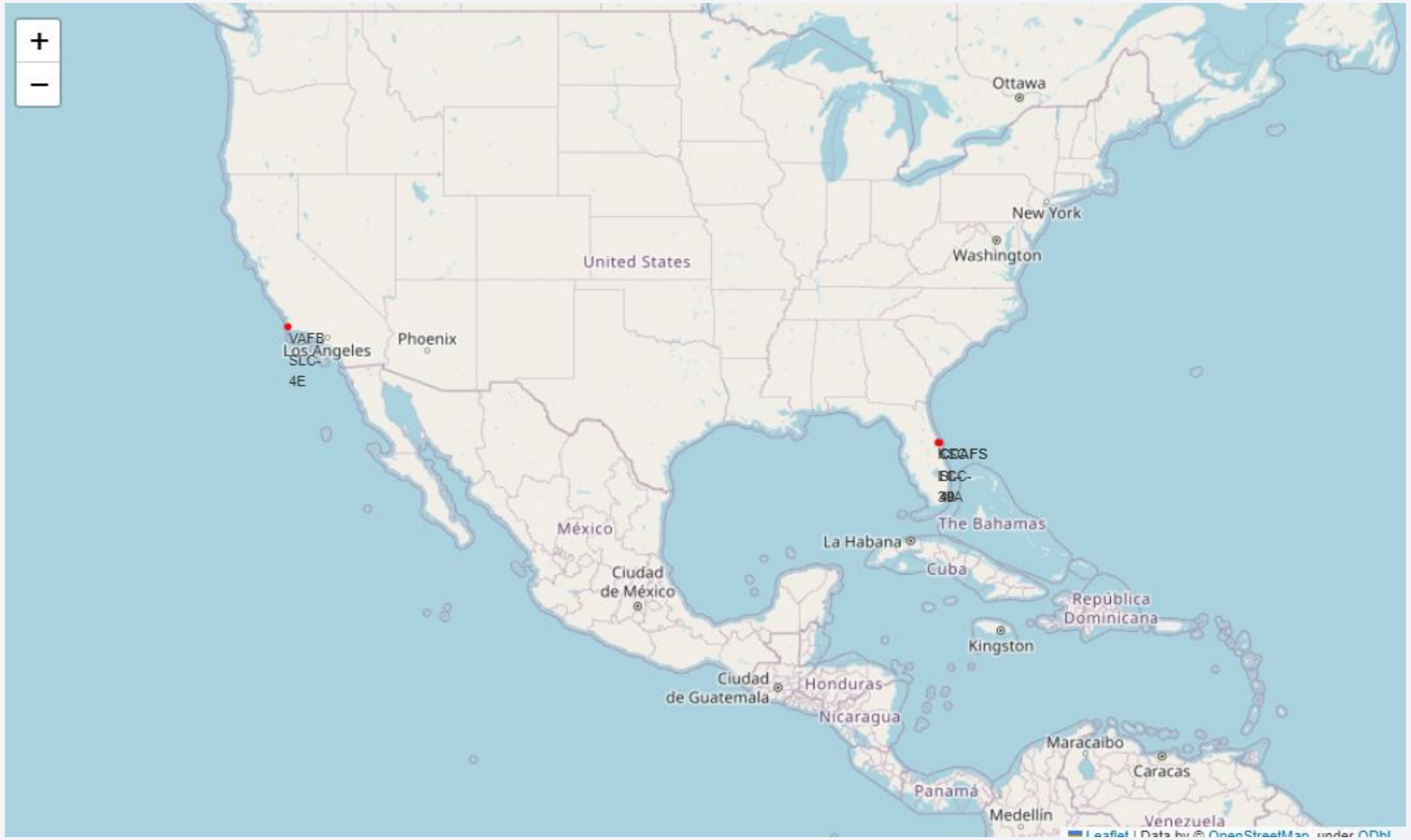
landing__outcome	qty
No attempt	7
Failure (drone ship)	2
Success (drone ship)	2
Success (ground pad)	2
Controlled (ocean)	1
Failure (parachute)	1

Count landing outcomes results from 2010-06-04 to 2017-03-20

Section 3

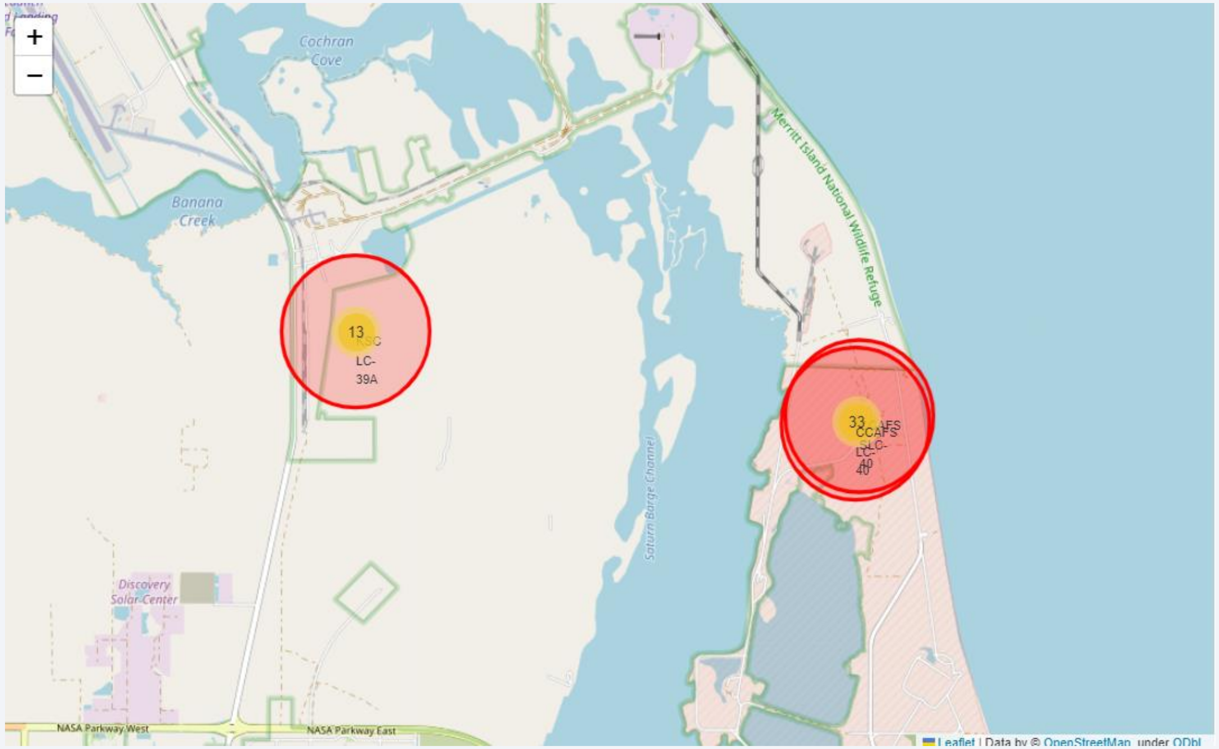
Launch Sites Proximities Analysis

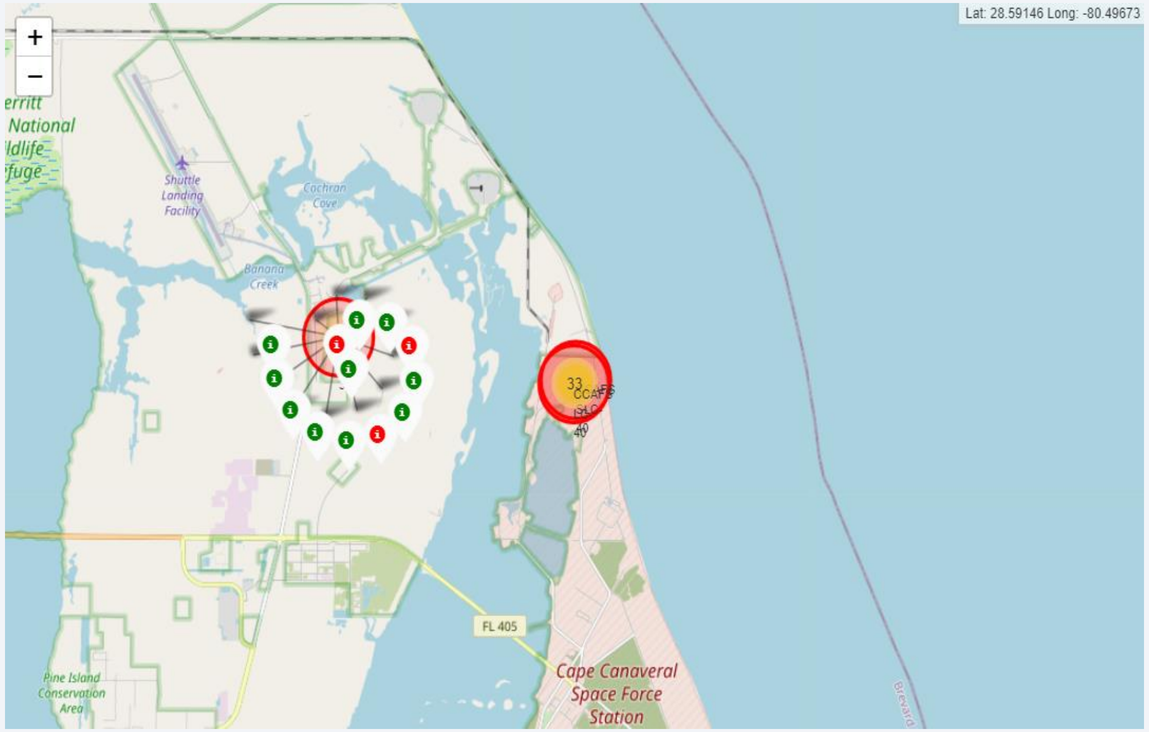
All Lunch Sites



As show all lunching sites near the sea

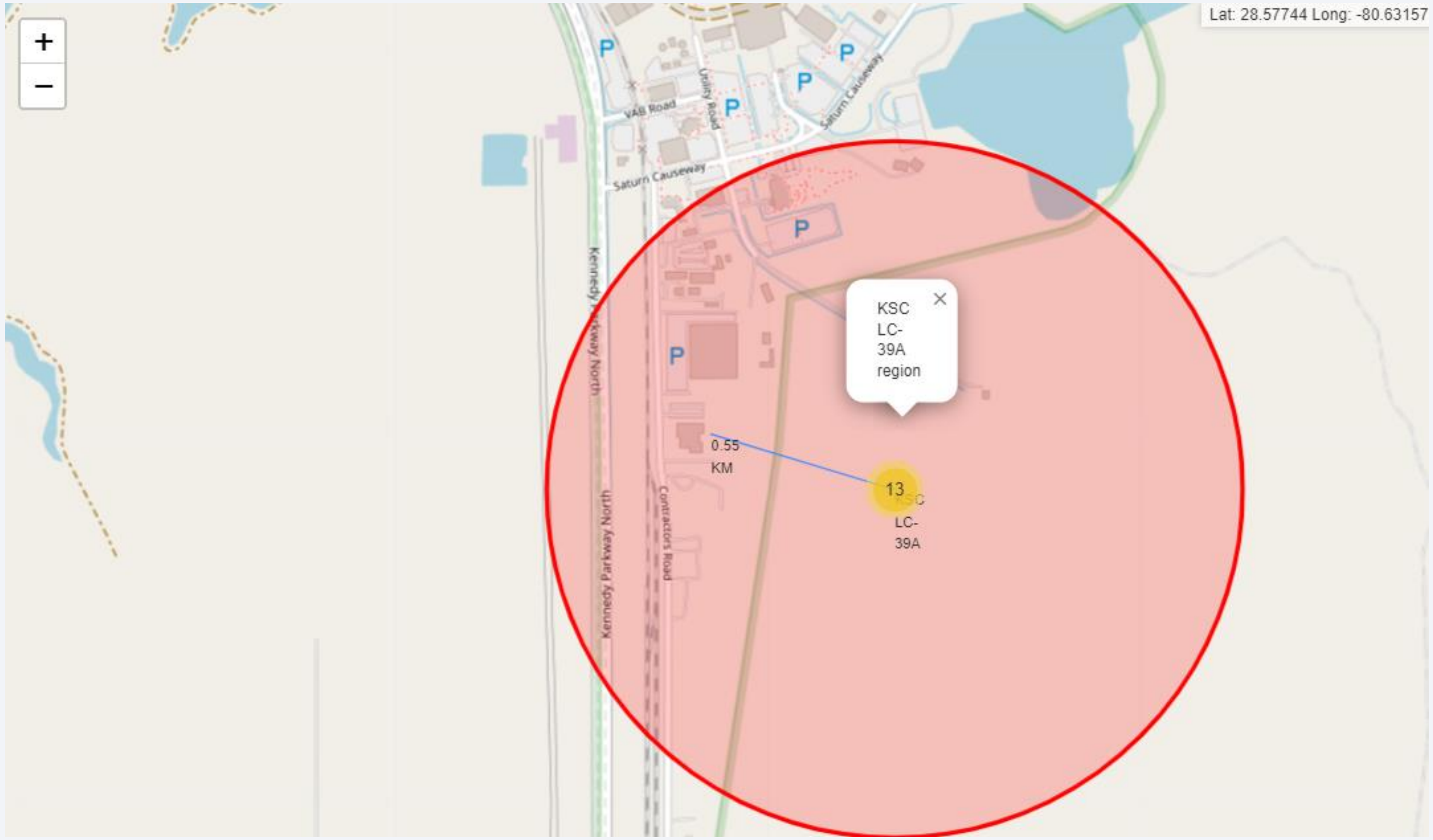
Launch Outcomes by Site





Illustrate lunch outcomes per site

Logestics

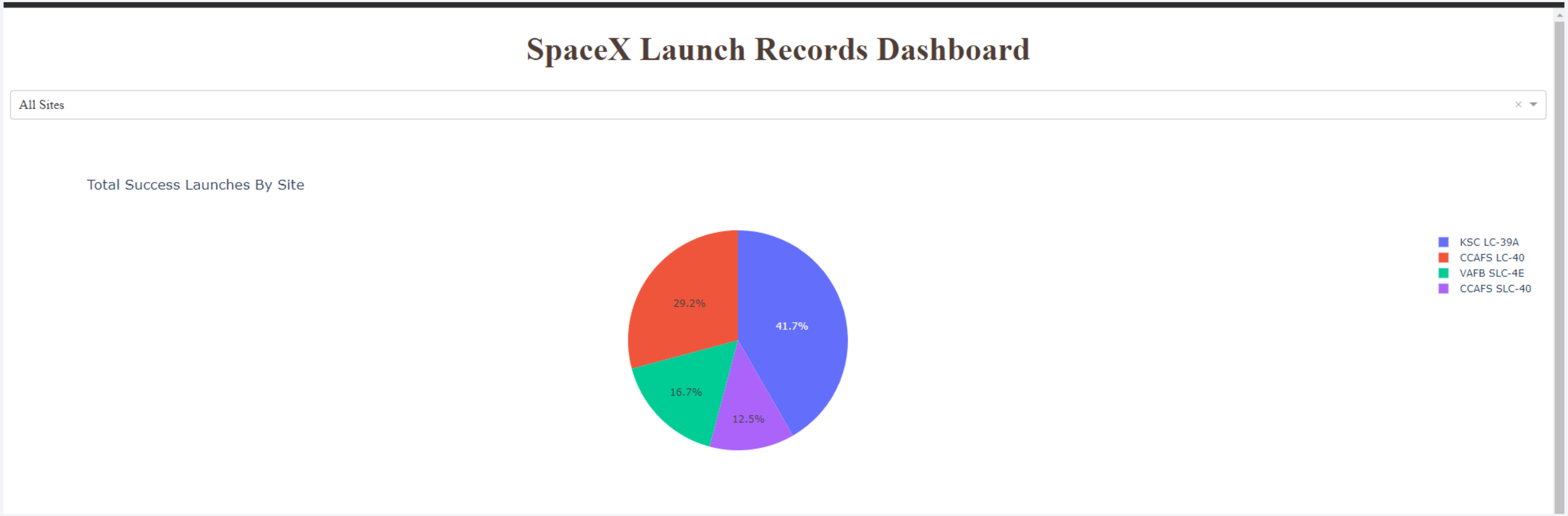


KSC LC-39A is near to Contractors road.

Section 4

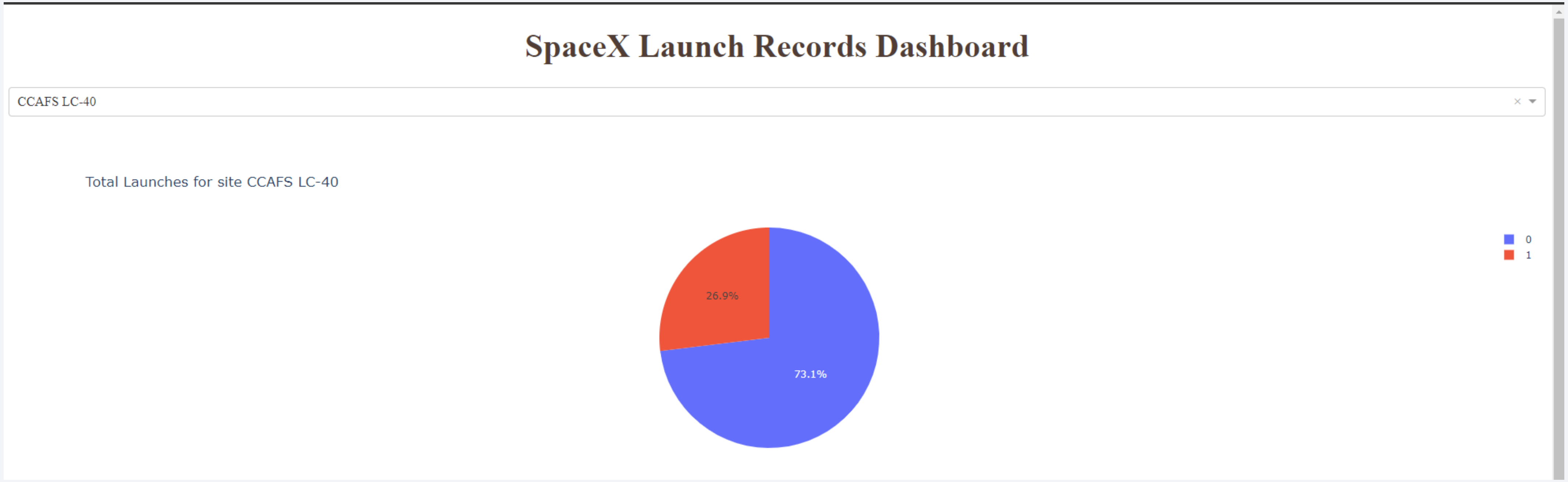
Build a Dashboard with Plotly Dash

Launch success for all sites



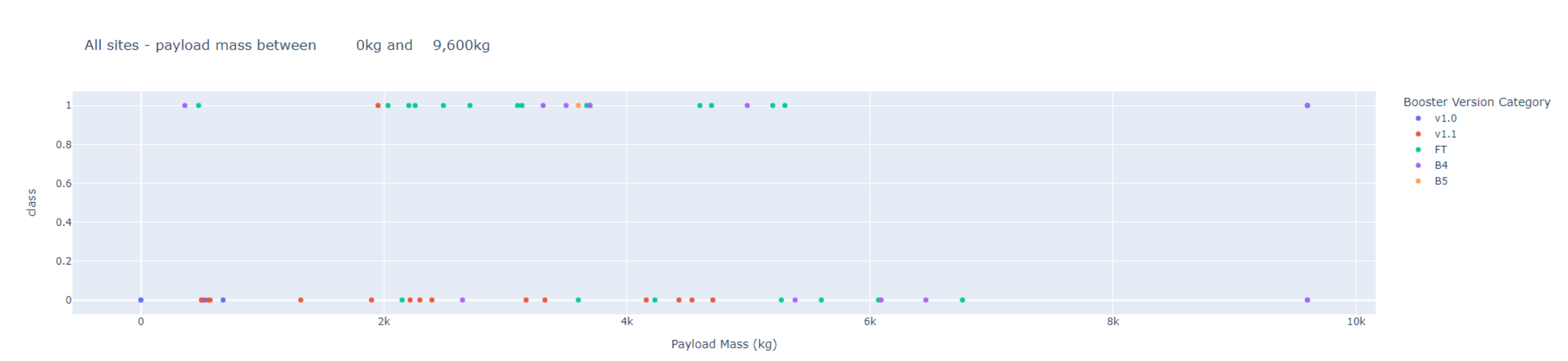
Picking the place for launch is very important for success of missions

Launch Success Ratio of KSC LC-39A



73.1% of launches are successful in this site

Payload vs. Launch Outcome



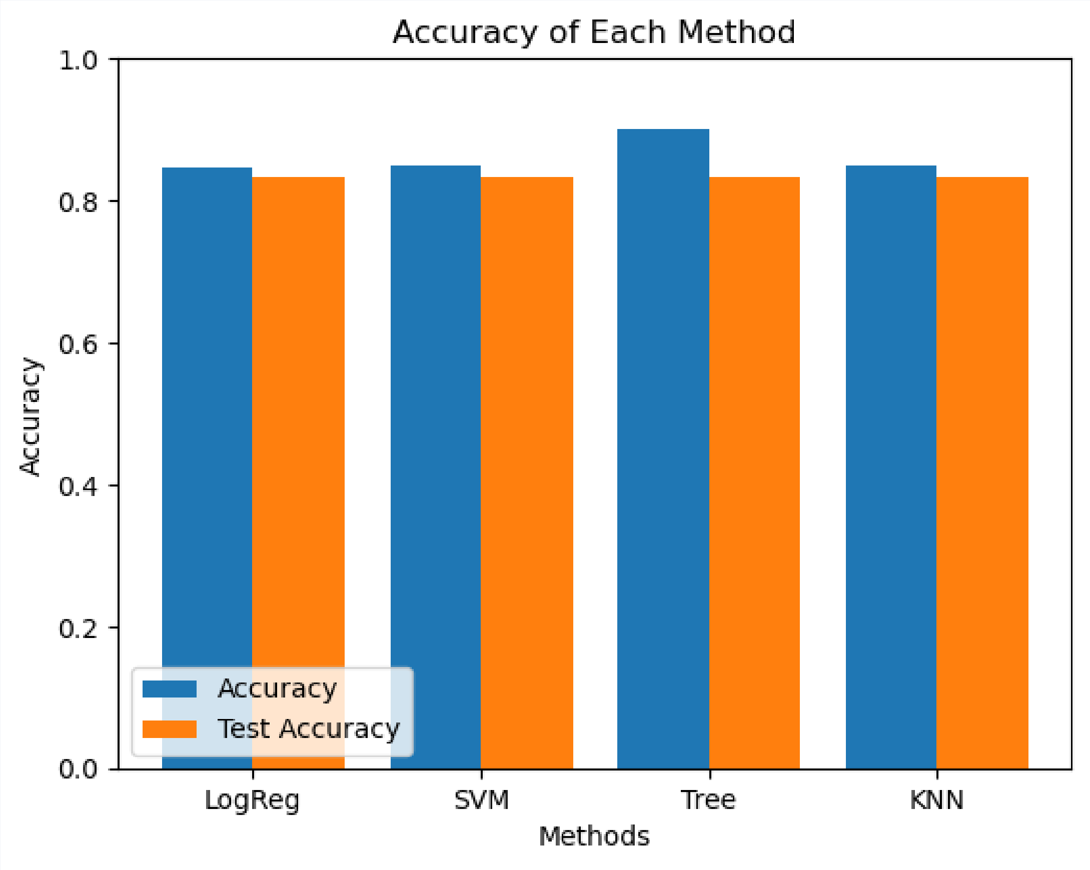
Payloads under 6,000kg and FT boosters are the most successful combination

Section 5

Predictive Analysis (Classification)

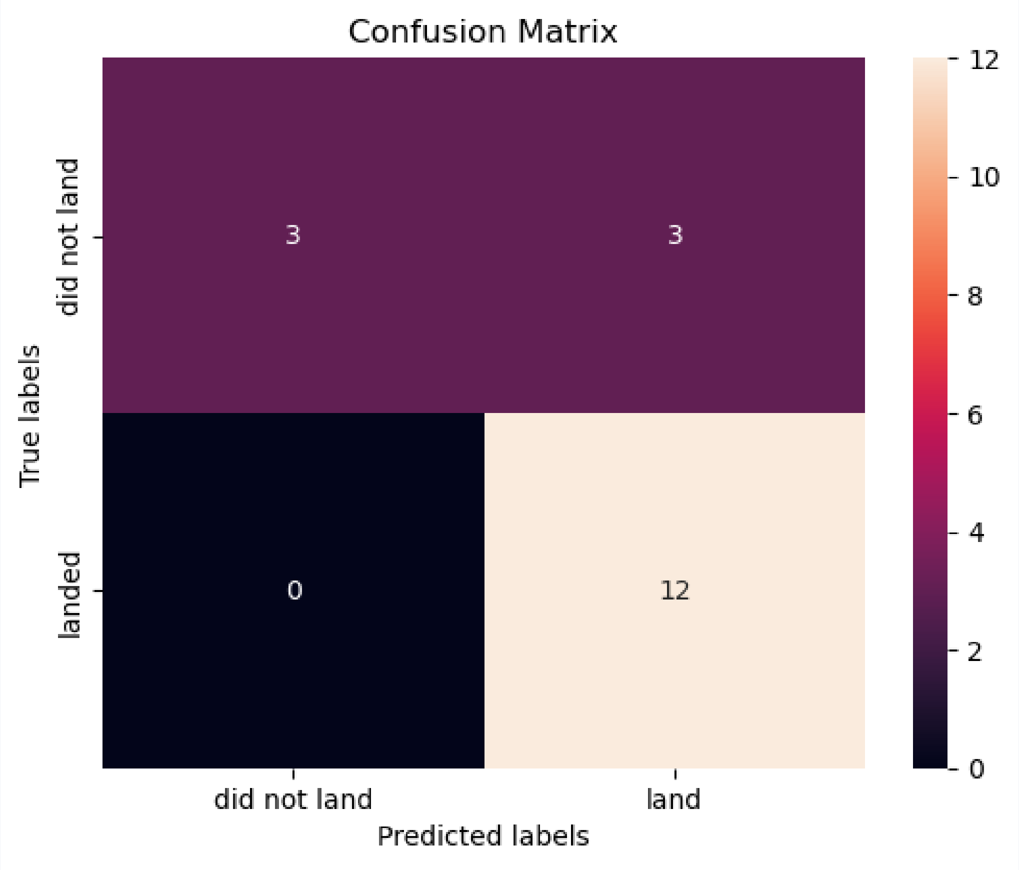
Classification Accuracy

- Four classification models were applied, as shown plotted beside
- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87%



Confusion Matrix

- Confusion matrix of Decision Tree Classifier proves its accuracy by comparing true positive and true negative.



Conclusions

- The most successful and recommended lunch site is KSC LC-39A.
- Lunches above 7,000 kg are less risky
- Decision Tree Classifier might be used to predict successful landings and

Appendix

Note: Folium did not show maps on Github, so I took screenshots.

Thank you!