IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jesper Thomsen
16.08.2022

Outline

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

Executive Summary

- **This project follows these steps:**
- Data Collection
- Data Wrangling
- Exploratory Data Analysis
- Interactive Visual Analytics
- Predictive Analysis (Classification)
- Summary of Results:
- This project produced the following outputs and visualizations:
- Exploratory Data Analysis (EDA) results
- Geospatial analytics
- Interactive dashboard
- Predictive analysis of classification models

Introduction

- SpaceX launches Falcon 9 rockets at a cost of around $62m. This is considerably cheaper than other providers (which usually cost upwards of $165m), and much of the savings are because SpaceX can land, and then re-use the first stage of the rocket.
- If we can make predictions on whether the first stage will land, we can determine the cost of a launch, and use this information to assess whether or not an alternate company should bid and SpaceX for a rocket launch.
- This project will ultimately predict if the Space X Falcon 9 first stage will land successfully.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected via spaceX api and via web scraping
- Perform data wrangling
  - Data was wrangled using fromjson, fiillNA(), onehotencoding.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Standardize the data and split into training and test set
  - Models were built using scikitlearn and several models were tried. Gridsearch was used to determine the best hyperparameters

Data Collection

1

- Make a GET request to the SpaceX REST API
- Convert the response to a .json file then to a Pandas DataFrame

2

- Use custom logic to clean the data
- Define lists for data to be stored in
- Call custom functions to retrieve data and fill the lists
- Use these lists as values in a dictionary and construct the dataset

3

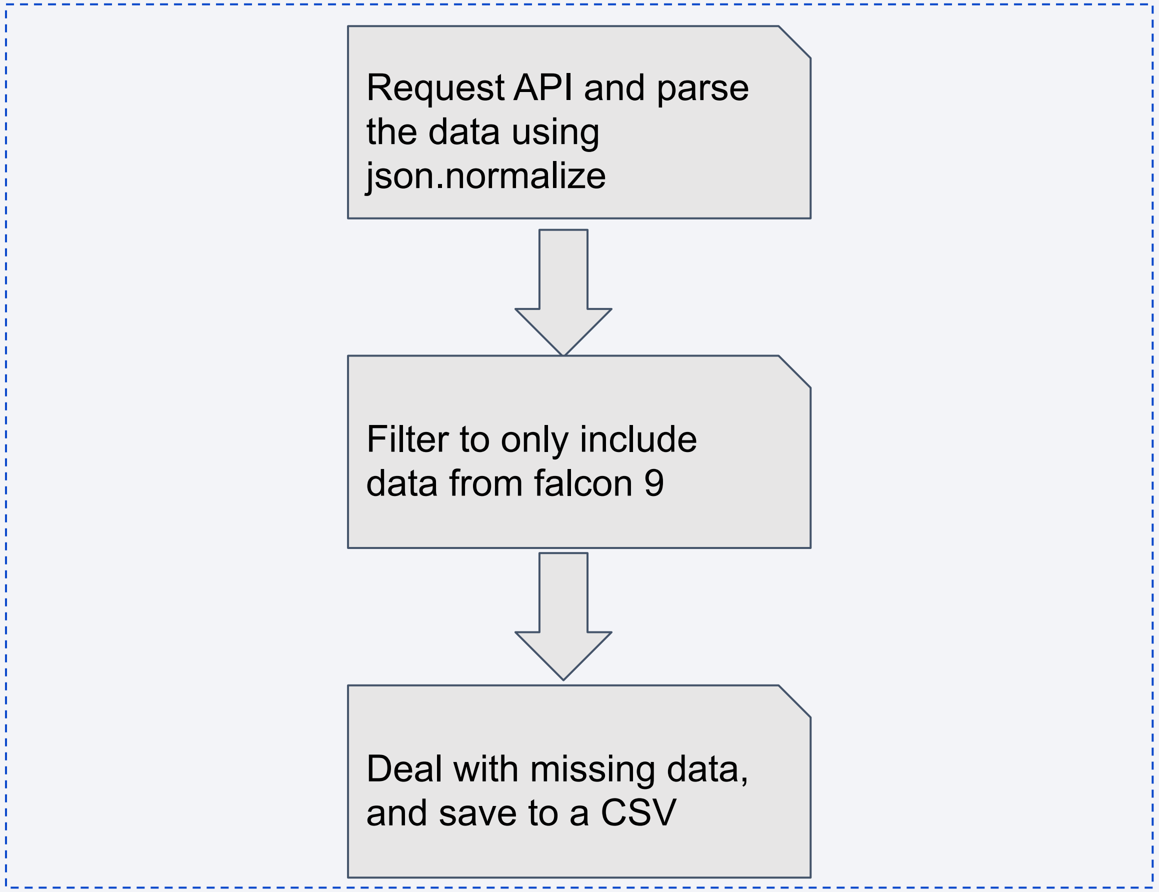
- Create a Pandas DataFrame from the constructed dictionary dataset

4

- Filter the DataFrame to only include Falcon 9 launches
- Reset the FlightNumber column
- Replace missing values of PayloadMass with the mean PayloadMass value

Data Collection – SpaceX API

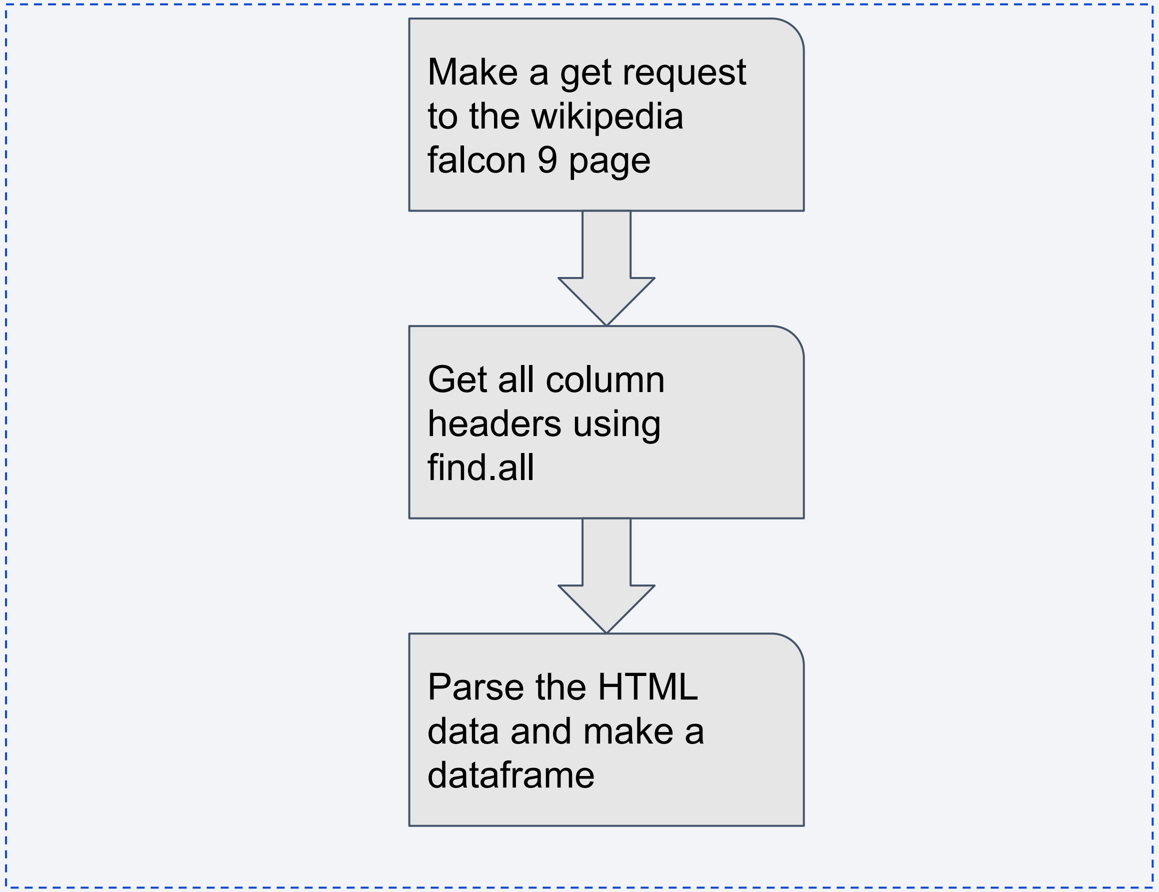
- We used SpaceX's online API to gather data
- github URL:



Data Collection - Scraping

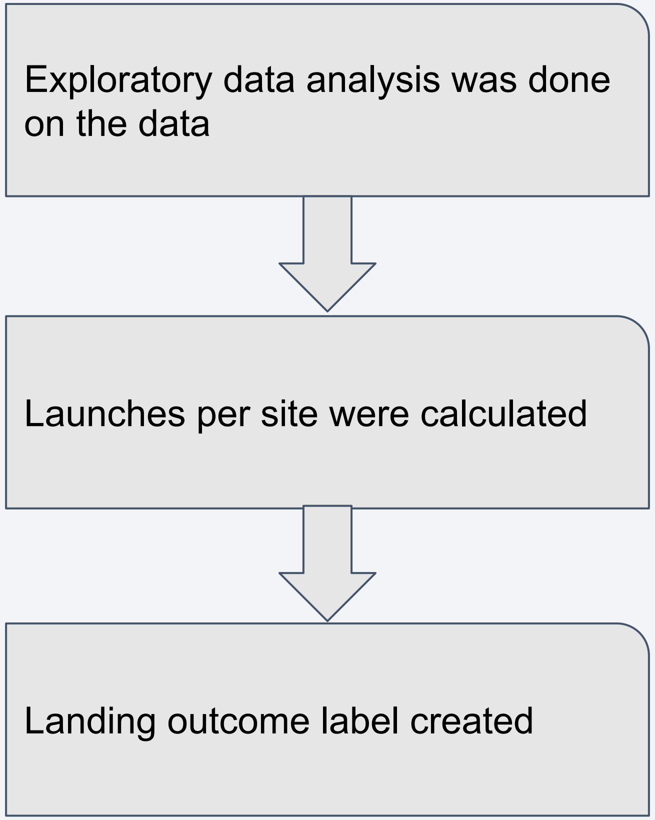
- We also obtained data from Wikipedia using a get request

- Github URL:



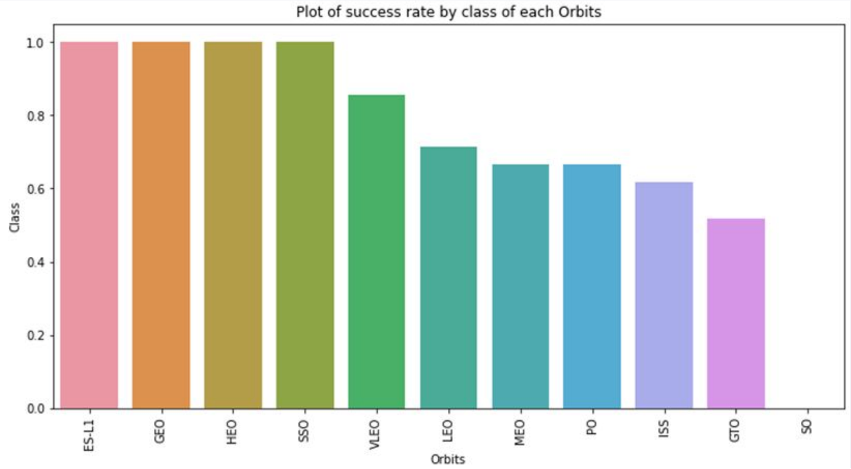
Data Wrangling

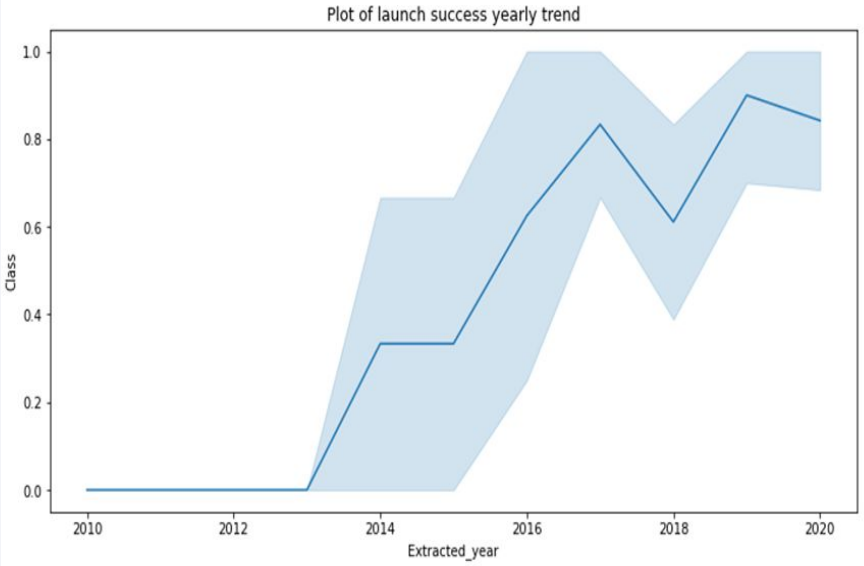
- •We performed exploratory data analysis and determined the training labels.
- •We calculated the number of launches at each site, and the number and occurrence of each orbits
- •We created landing outcome label from outcome column and exported the results to csv.

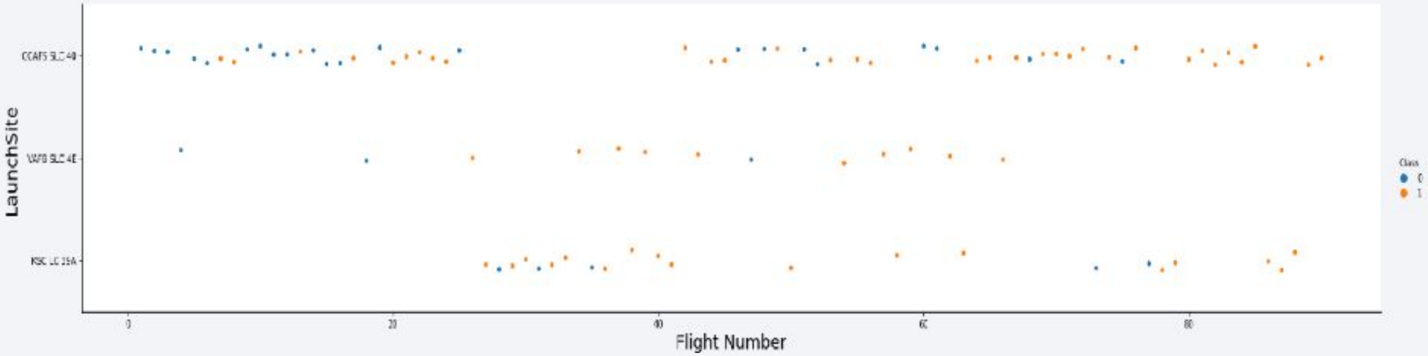


EDA with Data Visualization

- Success rate of orbits
- Yearly trend
- Different scatter plots showing relationships between factors such as launchsite and flight number, mass,payload, etc
- Github URL:







EDA with SQL

- Exploratory data analysis with SQL. We wrote queries to find:
    - The names of unique launch sites in the space mission.
    - The total payload mass carried by boosters launched by NASA (CRS)
    - The average payload mass carried by booster version F9 v1.1
    - The total number of successful and failure mission outcomes
    - The failed landing outcomes in drone ship, their booster version and launch site names
    - Date of first successful launch
    - names of successful launches between 4000 and 6000 mass
    - Names of boosters carrying the highest mass
- Github URL:

Build an Interactive Map with Folium

- Folum map indicators
  - Launch sites
  - Successful vs unsuccessful launches
  - Proximity to key objects such as oceans, railways and highways
- Explain why you added those objects
  - To gain a visual representation of factors that could be important for launch success
- GitHub URL:

Build a Dashboard with Plotly Dash

- We built an interactive dashboard with Plotly dash
- We plotted pie charts showing the total launches by a certain sites
- We plotted scatter graph showing the relationship with outcome an payload mass for the different booster version.
- GitHub URL:

Predictive Analysis (Classification)

•We loaded the data using numpy and pandas, transformed the data, split our data into training and testing.

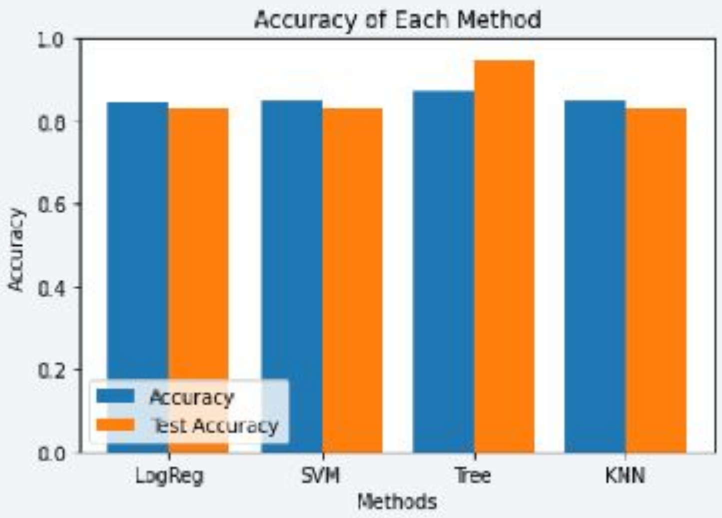
•We built different machine learning models and tune different hyperparameters using GridSearchCV.

•We used accuracy as the metric for our model, improved the model using feature engineering and algorithm tuning.

•We found the best performing classification model.

Results

- Exploratory data analysis results
  - We found out that launch sites should be in good sites, close to the sea and close to good infrastructure
- Predictive analysis results
  - Testing different classification models a decision tree model performed best

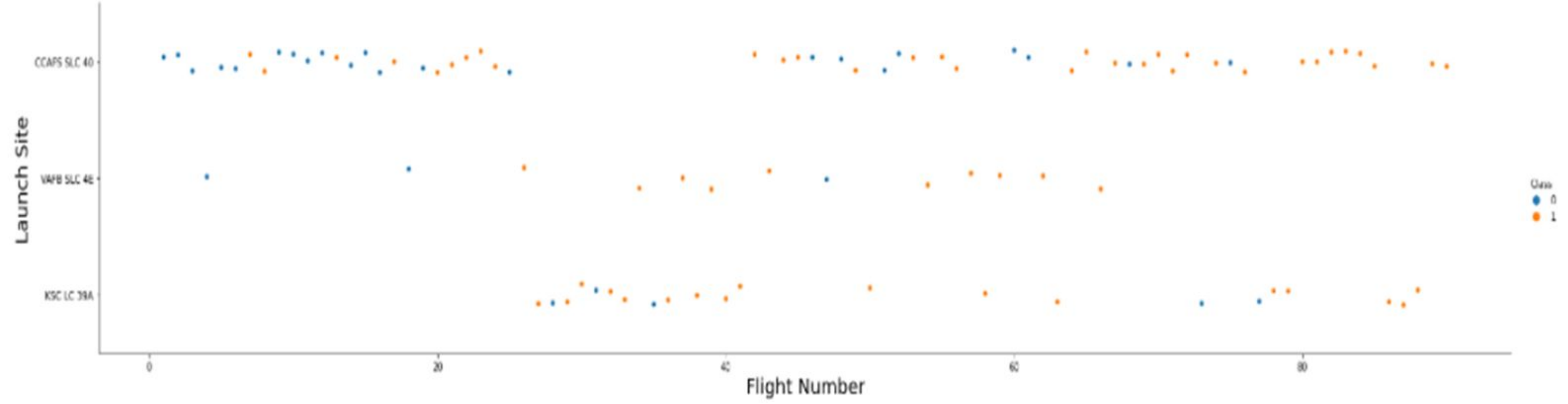


Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- The plot shows that the larger the flight amount at a test site, the higher the success rate

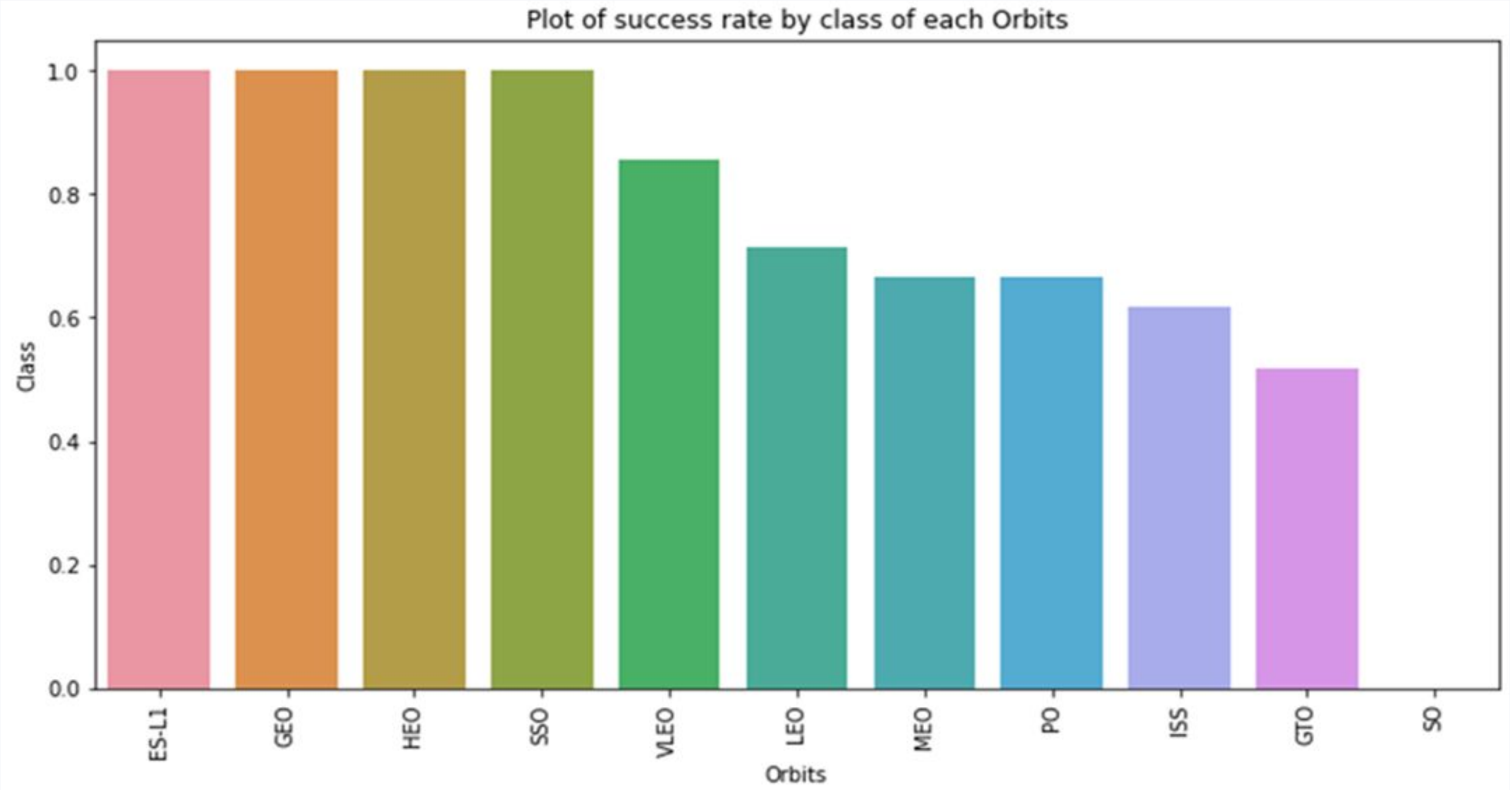


Payload vs. Launch Site

- The scatter plot shows that higher payload mass has a higher success rate
- It also indicates that only two launch sites have the capacity to launch rockets with mass over 10000 kg

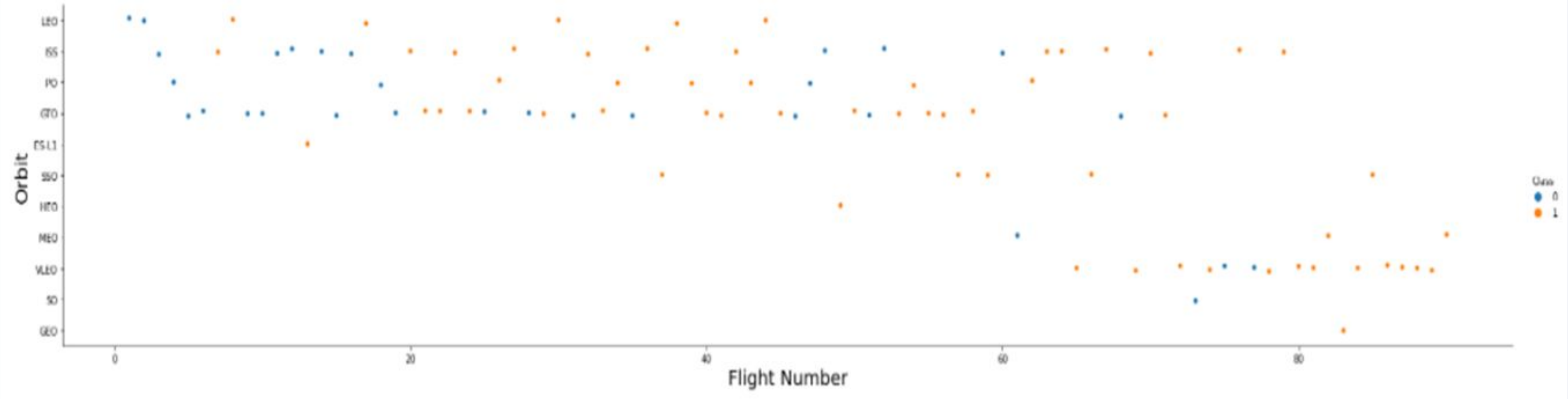
Success Rate vs. Orbit Type

- From the plot, we can see that ES-L1, GEO, HEO, SSO, VLEO had the most success rate



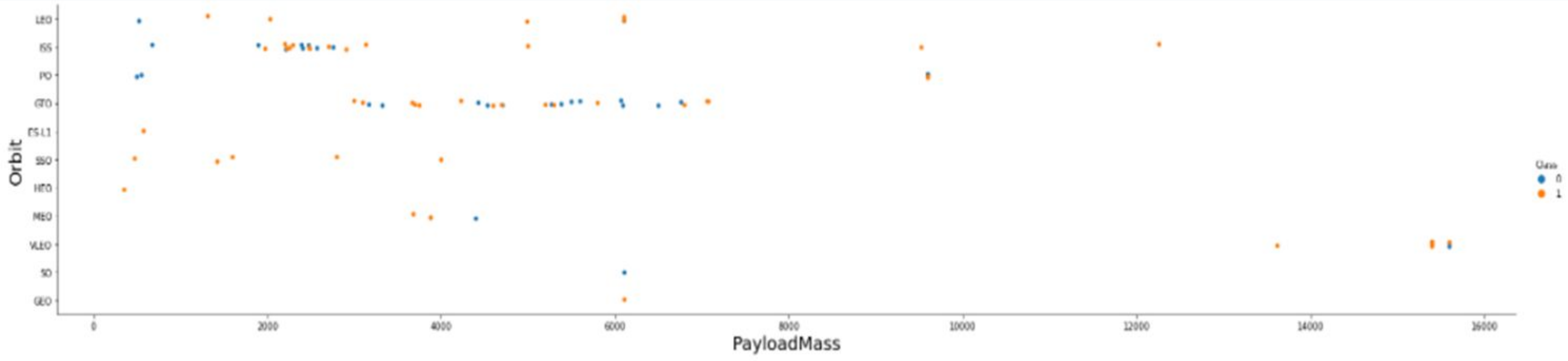
Flight Number vs. Orbit Type

- Success seems to increase with more flights for most orbits
- Some orbits seem more successful even from the first launch, such as the VLEO orbit



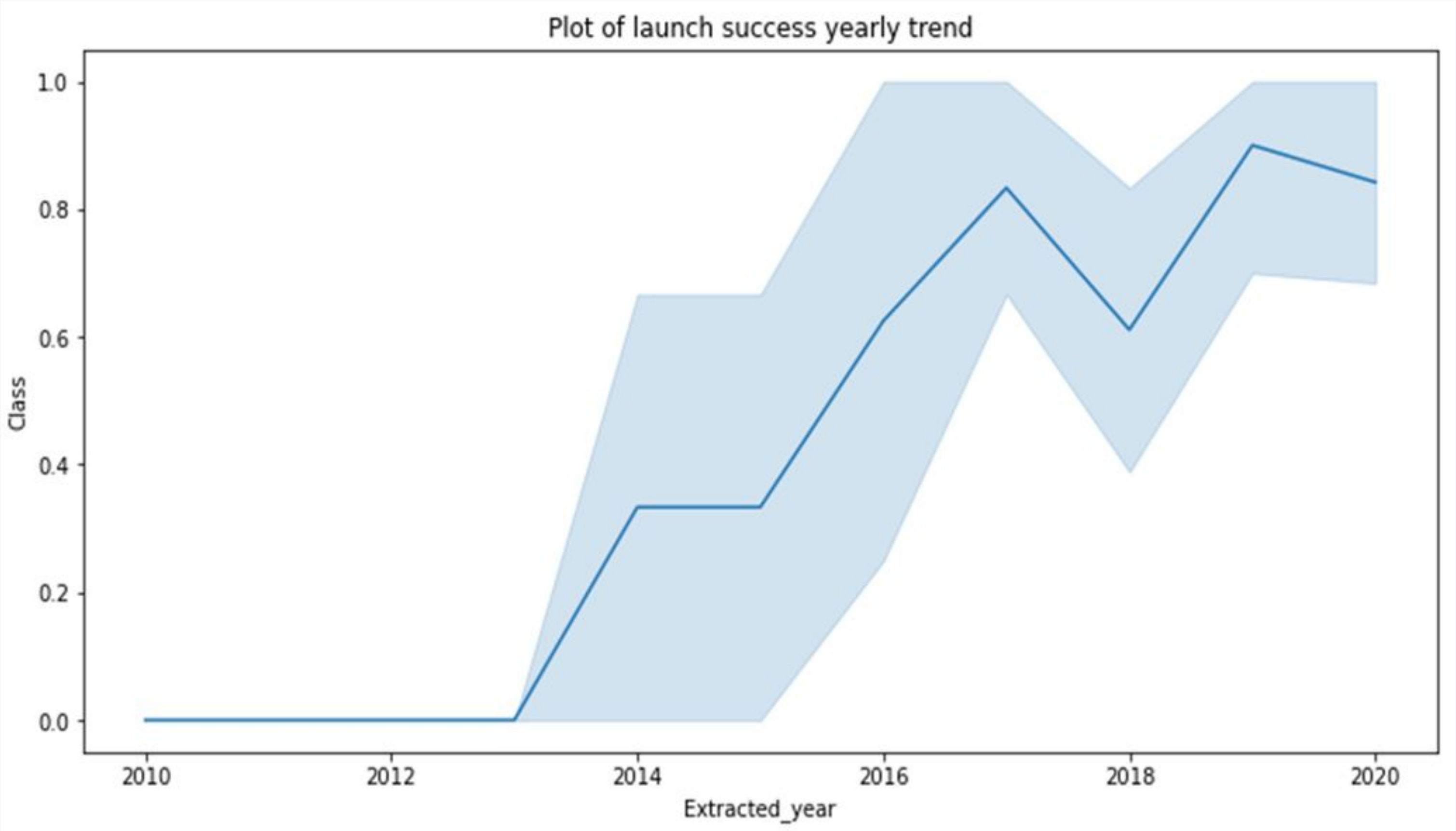
Payload vs. Orbit Type

- Higher payloads seem to have higher success rates, but also only some orbits are used for these payloads
- Some orbits only have successes, and are only used for low payloads, making them seem excellent for that purpose



Launch Success Yearly Trend

- Success rates have gradually increased since 2013, despite having a dip in 2018



All Launch Site Names

- There are 4 unique launch sites

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

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	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 mass is here

Total Payload (kg)
111.268

Average Payload Mass by F9 v1.1

- The total payload carried by F9 v1.1 from NASA is 2928.4

First Successful Ground Landing Date

- Date of the first successful ground landing was 22.12.2015

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

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
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3

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

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

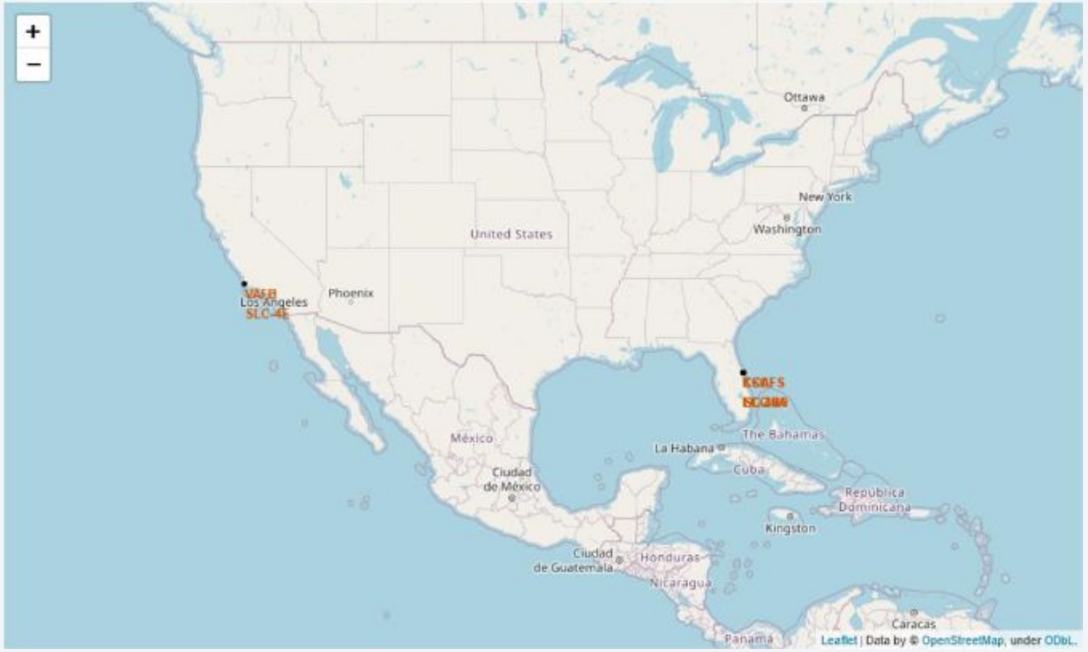
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

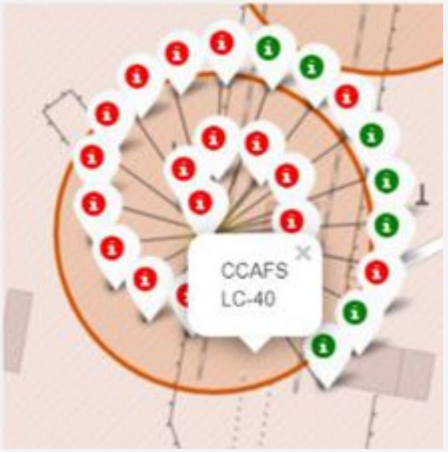
Launch sites

- Launch sites are close to water, and not too far from important infrastructure



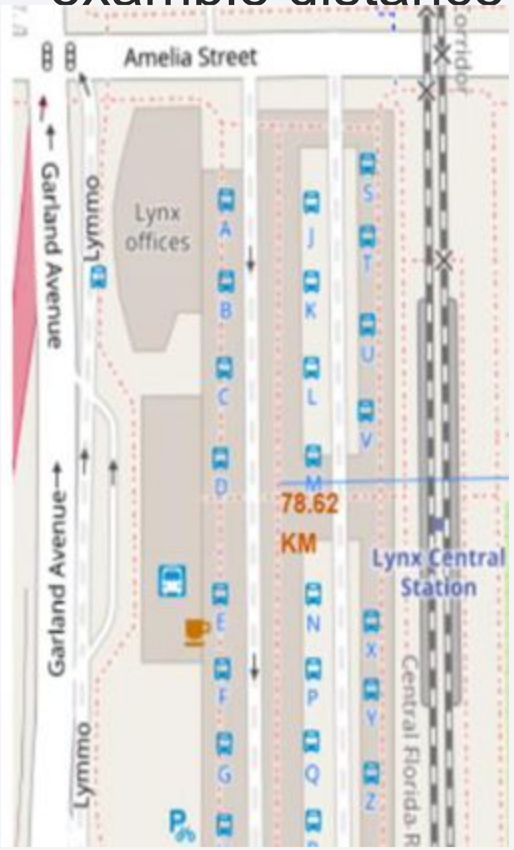
Launch success

- Green shows successful launches and red unsuccessful



Distances

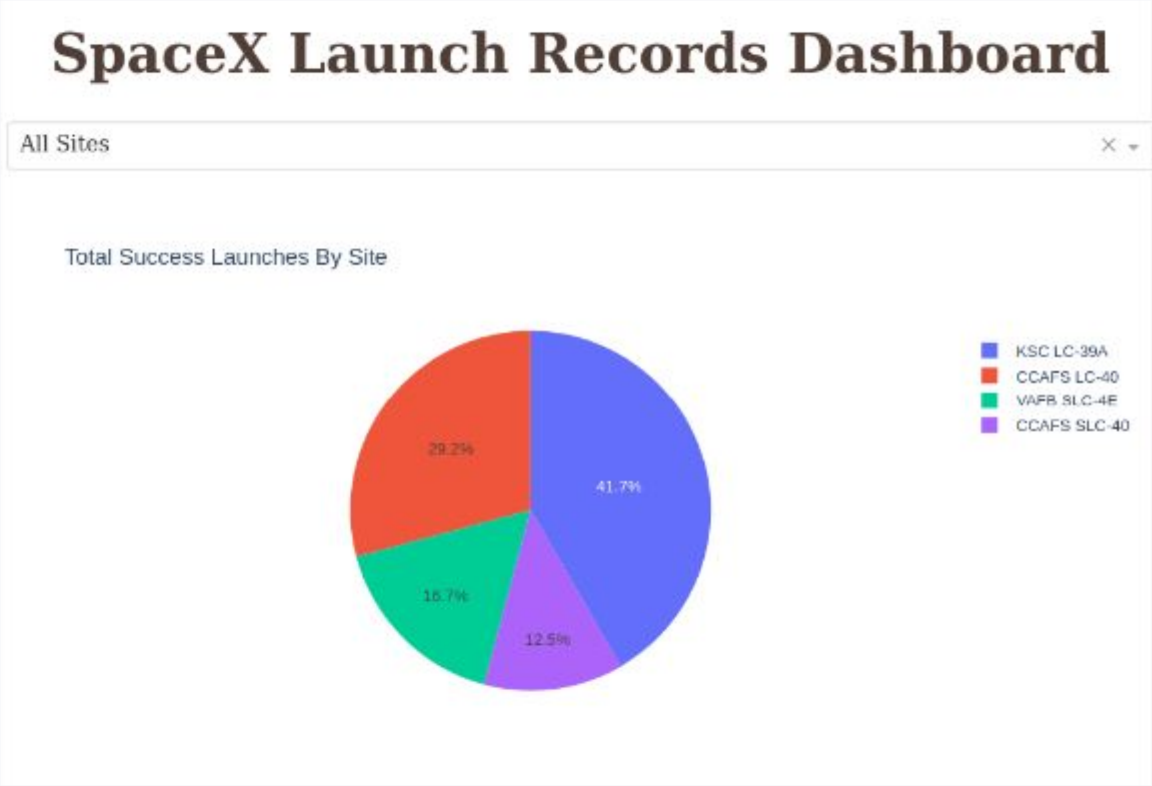
- example distance to station



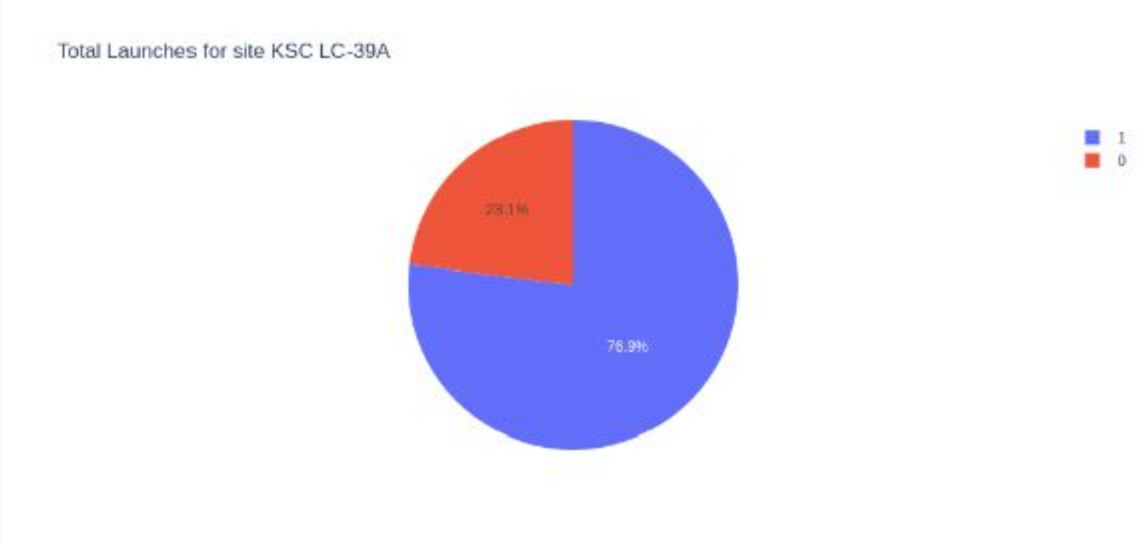
Section 4

Build a Dashboard with Plotly Dash

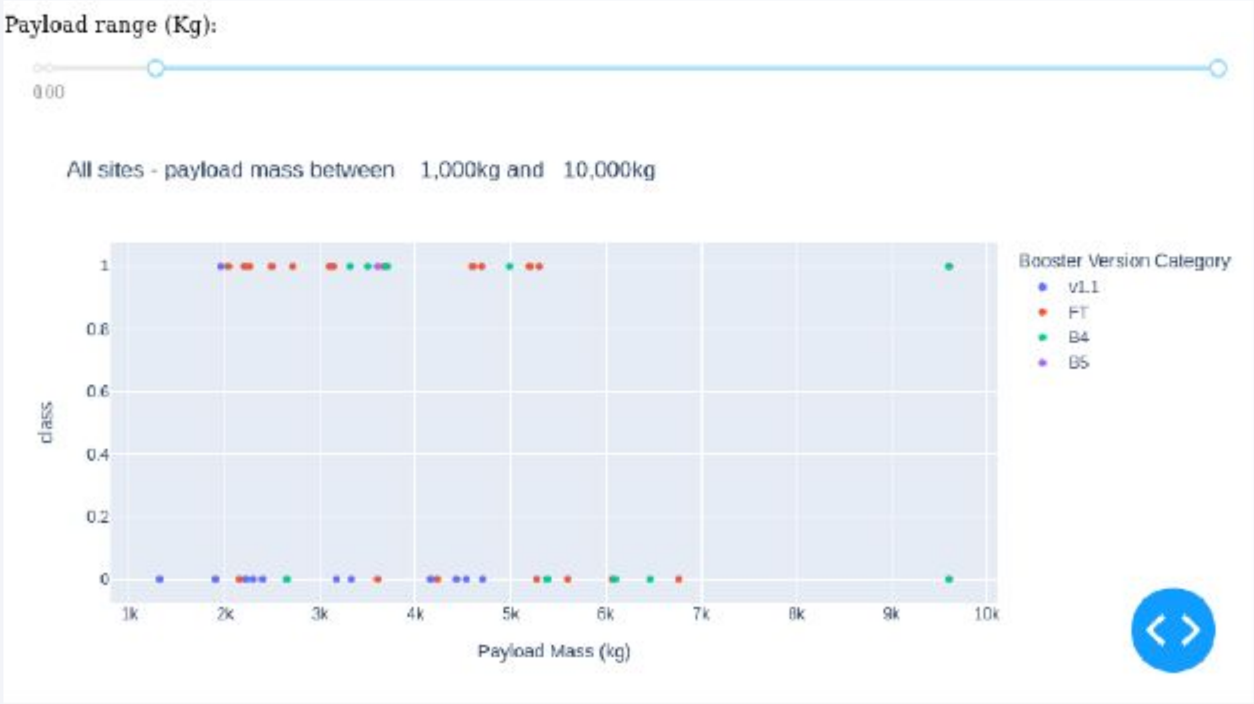
SpaceX launch records Dashboard



Launch success rate



Payload launch outcome



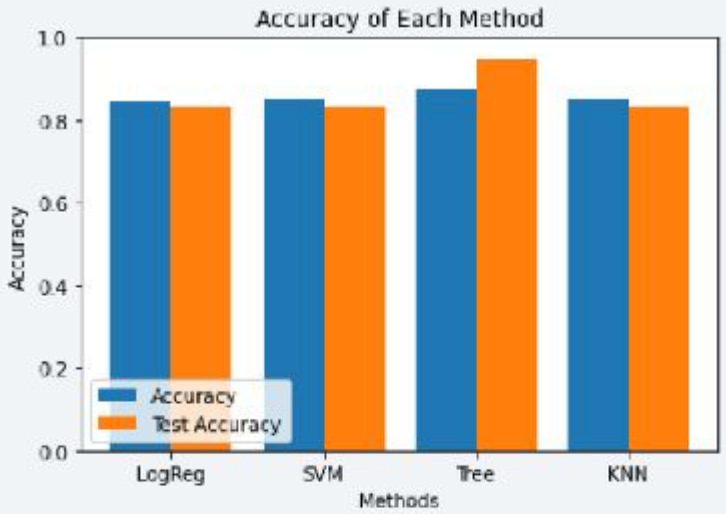
Section 5

Predictive Analysis (Classification)

Classification Accuracy

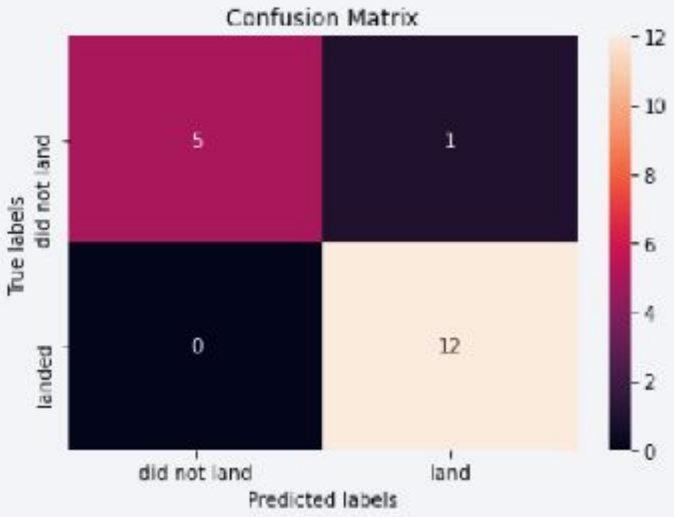
- Visualize the built model accuracy for all built classification models, in a bar chart

- Decision tree model has the highest accuracy



Confusion Matrix

- This shows the number of predictions and the models prediction in 12 cases in predicted land when it did in fact land and in 1 in predicted a land when it did not it has a tendency to underpredict how many did not land



Conclusions

- Several factors go into building a successful launch site
- Experience seems to help, as success has increased with time
- It is possible to model with reasonable accuracy. Best accuracy we got was with a decision tree model
- Some sites are more successful than others, showing room for improvement

Thank you!