IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mehmet Kizilaslan
03/02/2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection with Web Scraping and Data Wrangling for the SpaceX dataset
  - Exploratory Data analysis with SQL Data Visualization
  - Data Visualization and Interactive Visual Analytics
  - Machine Learning Prediction
- Summary of all results
  - Exploratory Data Analytics results
  - Interactive Analytics results
  - Machine Learning Prediction results

Introduction

- Project background and context

  Space X can reuse the first stage of their Falcon 9 rocket launches and make it cost 165 million dollars less than other providers. In this project is our objective is to analyze the Space X data to determine the cost of each launch and the probability of reuse of the first stage of the rockets. Finally, using all the gathered information for our competing company Space Y.

- Problems you want to find answers
  - The best machine learning method to estimate total cost for the launches by determining the successful landings of the first stage of the rockets
  - Best location is to be picked for maximizing the successful landings

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Get request used SpaceX API
  - BeautifulSoup used for Web Scrapping on Wikipedia url
- Perform data wrangling
  - Collected data group, manipulated. Success rate calculated.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Dataset normalized, divided into train and test data, 4 ML models tested for accuracies.

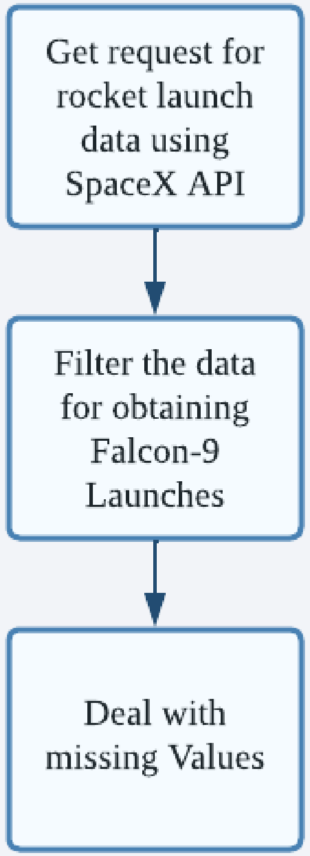
Data Collection

- Get request used SpaceX API: https://api.spacexdata.com/v4/launches/past
  - The response from request has decoded as json file and transformed into a pandas dataframe.

- BeautifulSoup used for Web Scrapping on Wikipedia_url : https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
  - Launch records are extracted from the html sourced data. A brand-new dictionary created and transformed into a new dataframe.

Data Collection – SpaceX API

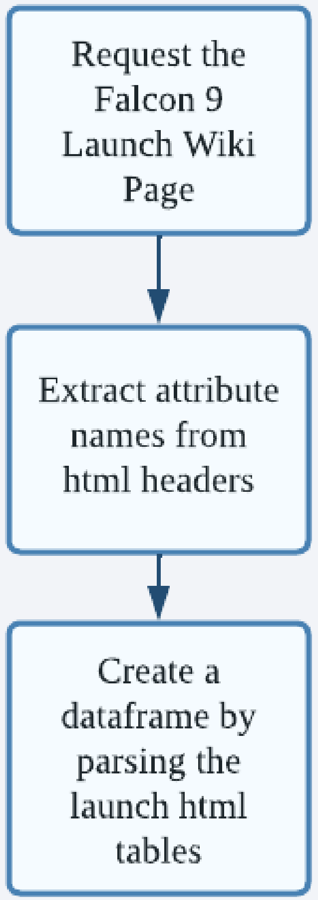
- SpaceX API's data is gathered with the get request. Data has cleaned, data wrangling and formatting applied.



Source Code: https://github.com/mkizi1/SPACE-Y/blob/master/Data_Collection_API.ipynb

Data Collection - Scraping

- Headers from html response of Wikipedia url web scrapped and used for creating a pandas dataframe

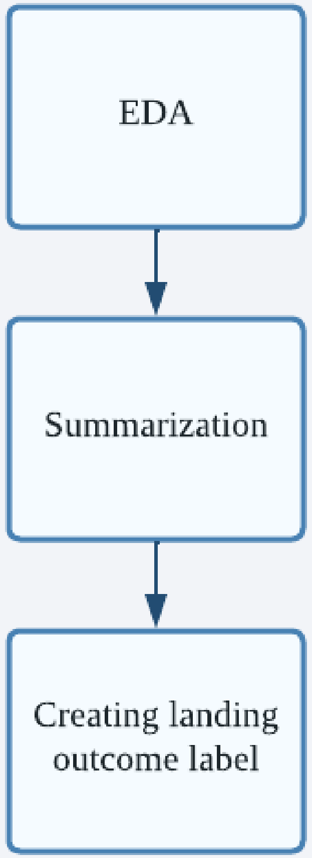


Source Code: https://github.com/mkizi1/SPACE-Y/blob/master/Data_Collection_with_Web%20Scraping.ipynb

Data Wrangling

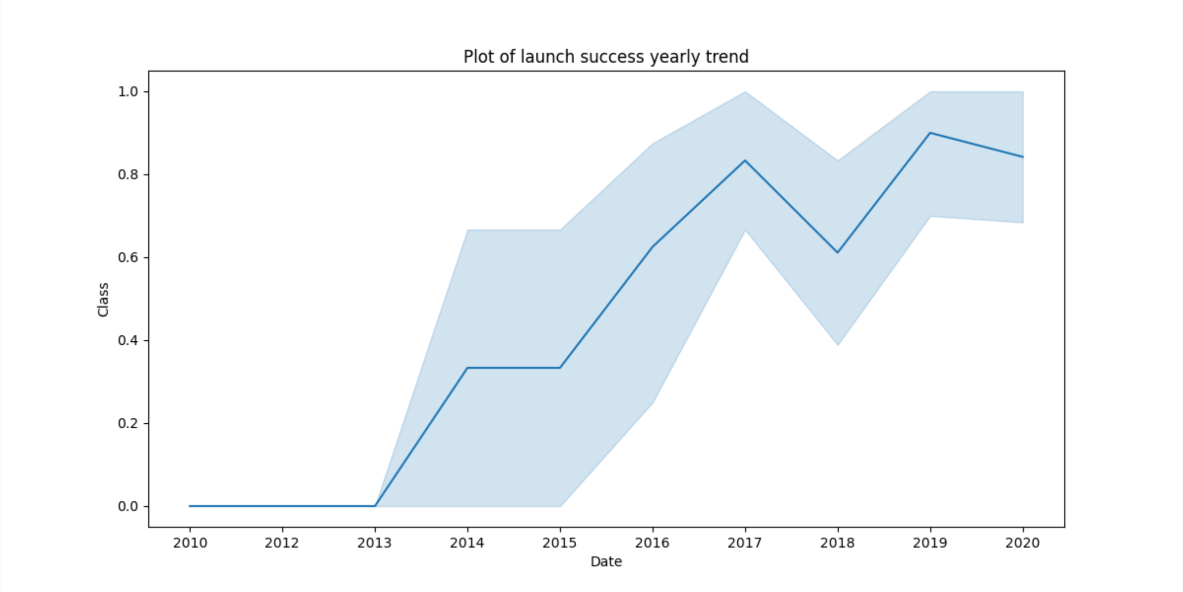
- Exploratory data analysis performed
- Quantity of launches on each site and occurrences of each orbit are calculated
- Landing outcome attribute added to the output dataframe, and results exported as a csv file.

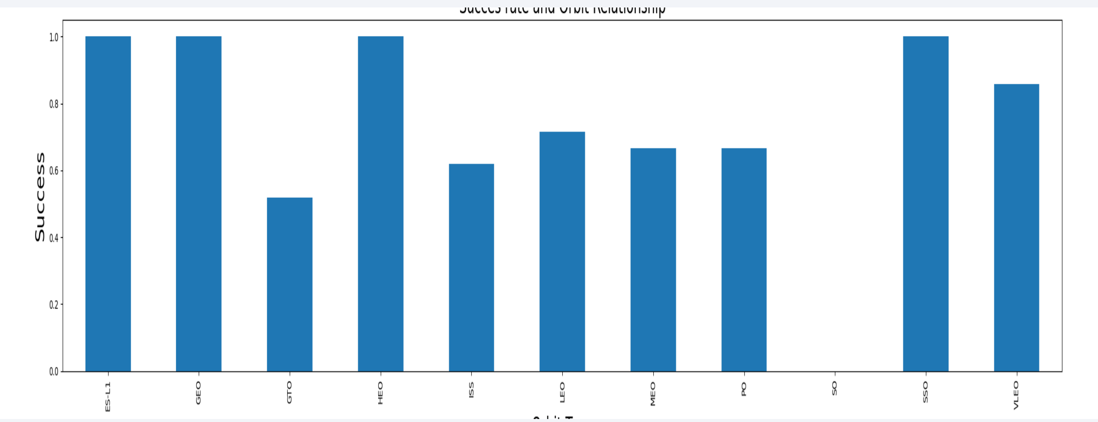
Source Code: https://github.com/mkizi1/SPACE-Y/blob/master/Data_Wrangling.ipynb



EDA with Data Visualization

- Data explored by visualizing with bar charts and scatterplots between pair of features:
  - Payload Mass-Flight Number, Launch Site-Flight Number, Launch Site-Payload Mass, Orbit-Flight Number, Payload-Orbit





Source Code: https://github.com/mkizi1/SPACE-Y/blob/master/EDA_with_Data_Visualition.ipynb

EDA with SQL

- The SQL queries you performed:
  - The names of the unique launch sites in the space mission.
  - Top 5 launch sites whose name begin with string ”CCA”
  - 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, and their booster versions, and launch site names in 2015
  - The ranking of landing outcomes due to their success rates between specific dates

Source Code: https://github.com/mkizi1/SPACE-Y/blob/master/EDA_with_SQL.ipynb

Build an Interactive Map with Folium

- Using the folium map, we identified every launch site and added items to indicate the success or failure of launches for each site, such as markers, circles, and lines.
- We categorize feature launch outcomes into classes 0 and 1 depending on whether they succeed or fail.
- We determined which launch sites had a comparatively high success rate using the clusters of colored marker markers.
- Distances between two coordinates are shown by lines.

Source Code: https://github.com/mkizi1/SPACE-Y/blob/master/Interactive_Visual_Analytics_with%20Folium.ipynb

Build a Dashboard with Plotly Dash

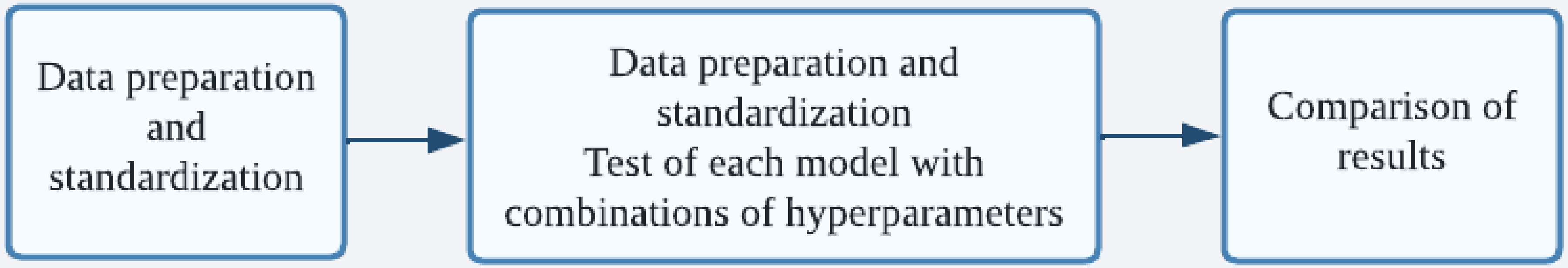
- Using Plotly dash, we created an interactive dashboard.
  - We created pie graphs that display the total number of launches by particular sites.
  - For each booster version, we created a scatter graph to show the relationship between the outcome and the payload mass (Kg).

- In order to determine the optimal launch site for a given payload, this combination made it possible to swiftly analyze the relationship between payloads and launch locations.

Source Code: https://github.com/mkizi1/SPACE-Y/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

- Using Numpy and Pandas, we imported the data, transformed it, and divided it into training and testing sets.
- Using GridSearchCV, we constructed various machine learning models and adjusted various hyperparameters.
- We measured the performance of our model using accuracy and then enhanced it using feature engineering and algorithm tuning.
- We discovered the classification model that performed the best.



Source Code: https://github.com/mkizi1/SPACE-Y/blob/master/Machine_Learning.jupyterlite.ipynb

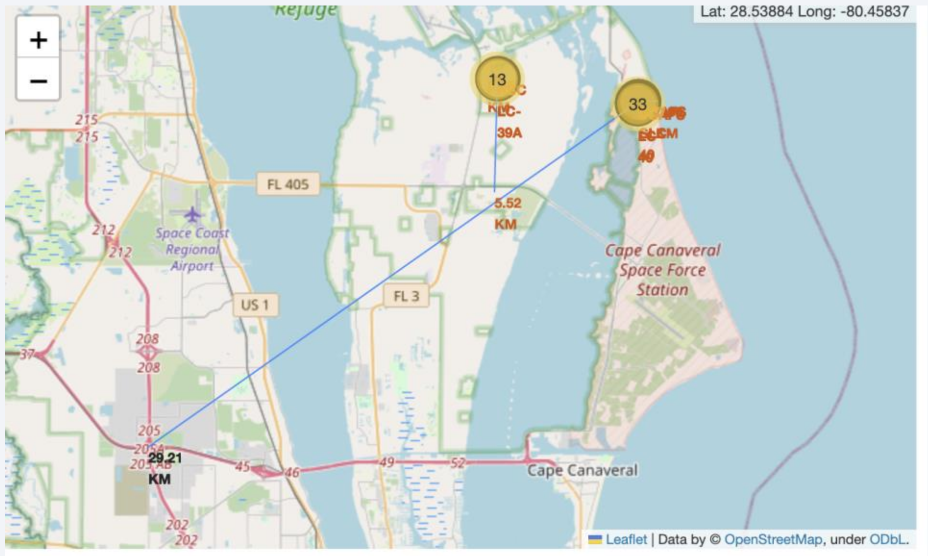
Results

- Exploratory data analysis results:
  - Space X uses 4 different launch sites
  - The first launches were done to Space X itself and NASA
  - 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
  - The number of landing outcomes became as better as years passed.

Results

- It was feasible to determine through interactive analytics that launch sites previously were in secure locations, such as close to the sea and with a good logistic infrastructure nearby.

- East Coast launch locations host the majority of launches.

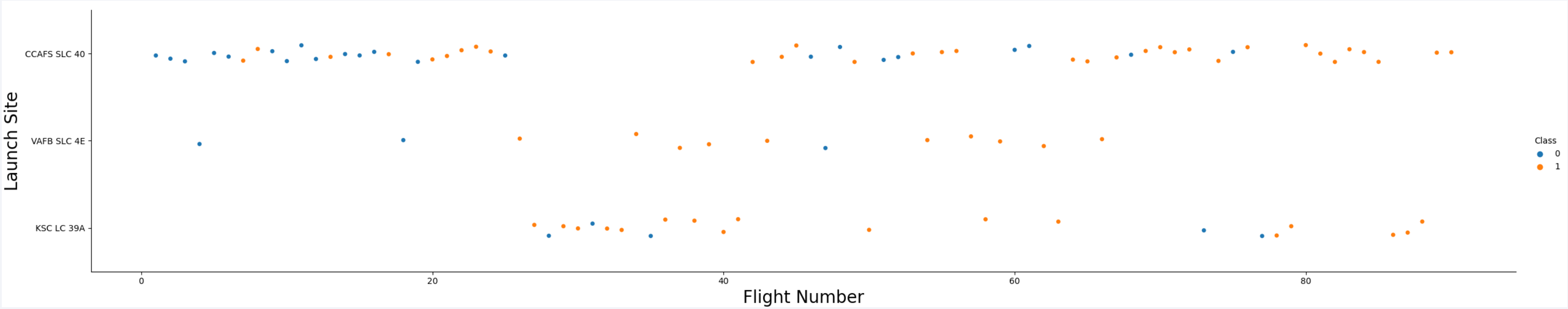


Section 2

Insights drawn from EDA

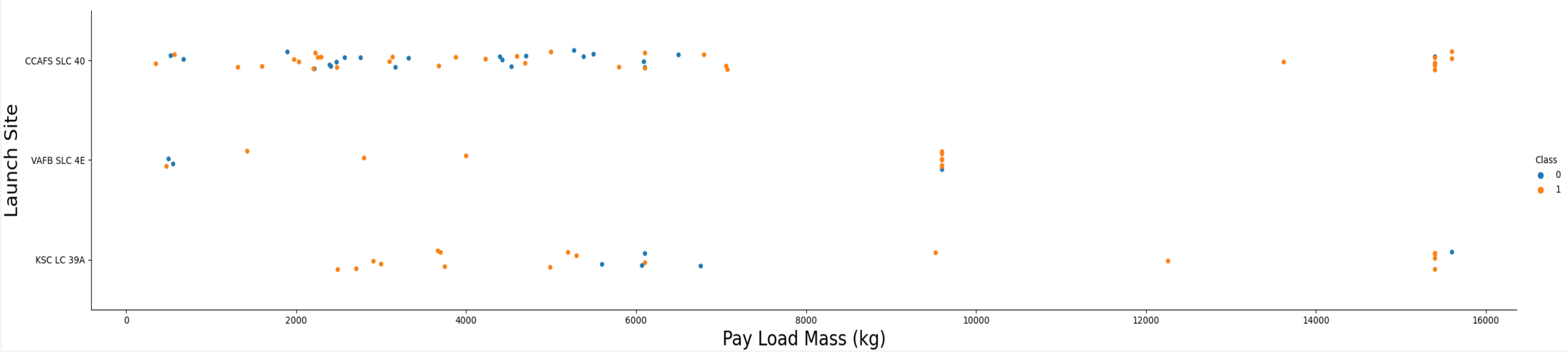
Flight Number vs. Launch Site

- The general success rate has increased over time, which is clearly apparent.
- The plot above makes it possible to confirm that CCAF5 SLC 40, where the majority of recent launches were successful, is the best launch site right now;
- KSC LC 39A came in third, followed by VAFB SLC 4E in second;



Payload vs. Launch Site

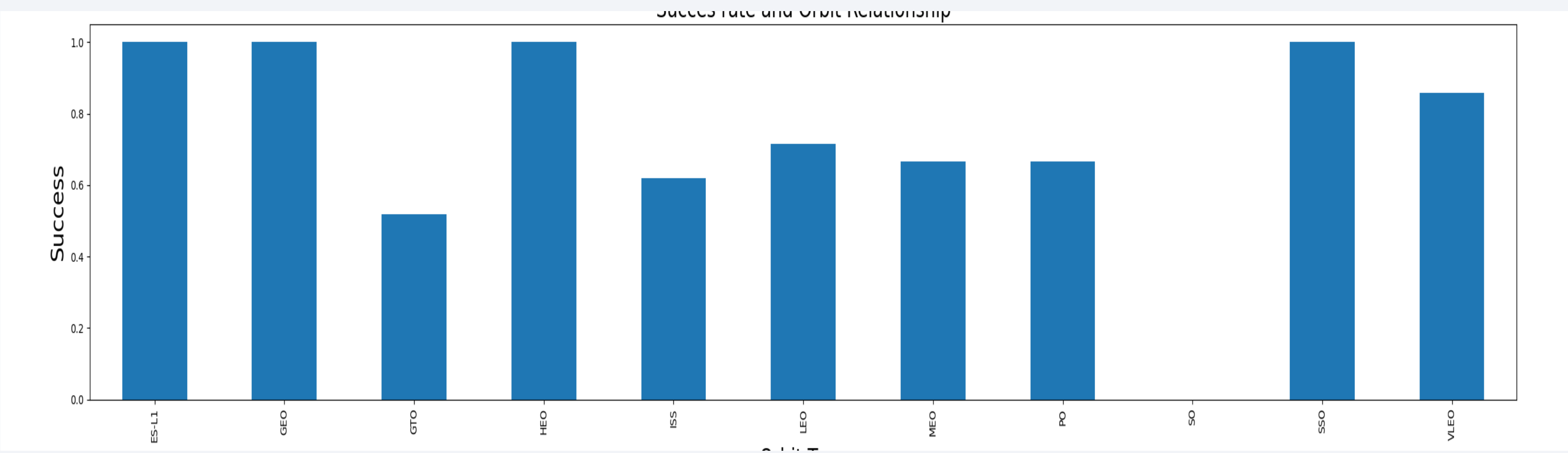
- Only at the CCAFS SLC 40 and KSC LC 39A launch sites do payloads over 12,000 kg appear to be feasible.
- At payloads exceeding 9,000 kg, the success rate is very high



Success Rate vs. Orbit Type

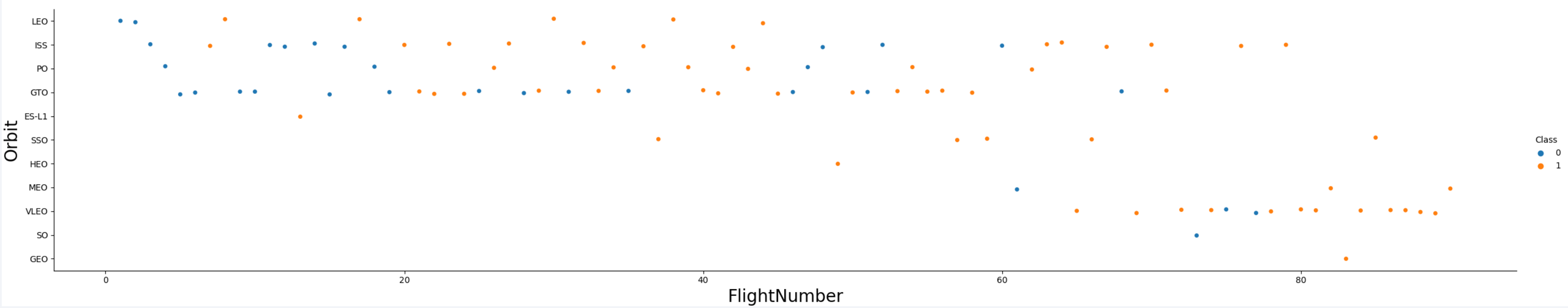
- The biggest success rates happens to orbits:
- ES-L1
- GEO
- HEO
- SSO

Followed by:
- VLEO (above 80%)
- LEO(above65%)



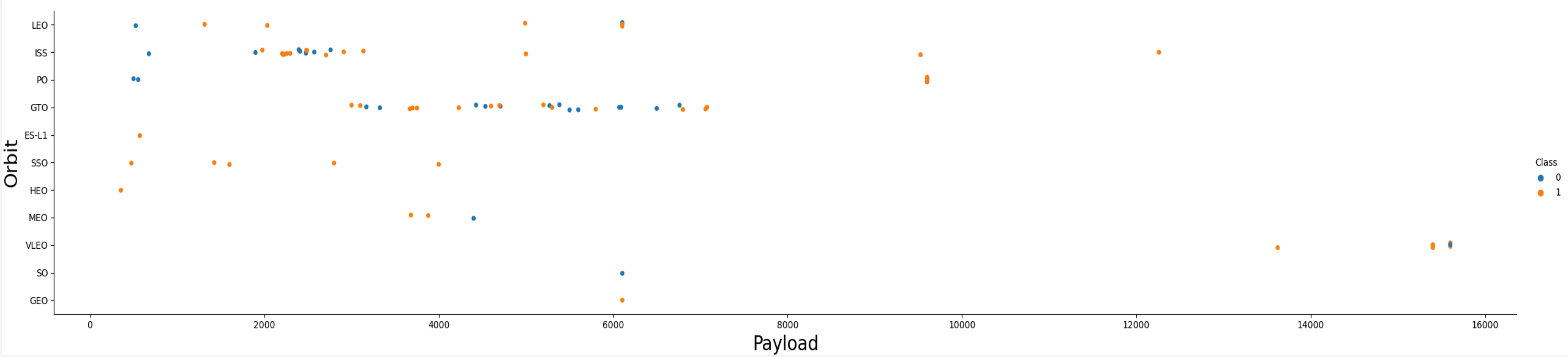
Flight Number vs. Orbit Type

- Evidently, all orbits' success rates increased with time;
- Because to the recent rise in frequency, VLEO orbit appears to be a new economic prospect.



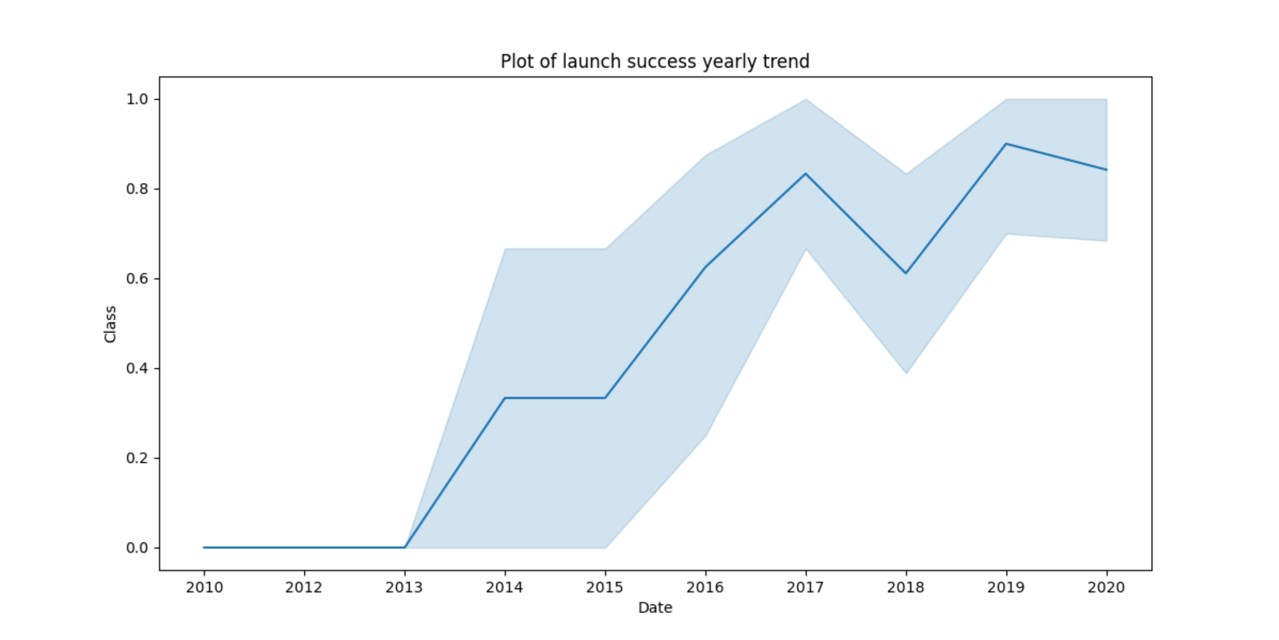
Payload vs. Orbit Type

- It appears that there is no connection between payload and GTO success rate.
- The greatest payload range and success rate are in ISS orbit.
- The SO and GEO orbits see infrequent launches.



Launch Success Yearly Trend

- Success rate began to rise in 2013 and continued until 2020.
- It appears that the first three years were a time of adjustments and technological advancement.



All Launch Site Names

- According to data, there are four launch sites:

- To display only distinct launch sites from the SpaceX data, we applied the keyword DISTINCT.

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

Launch Site Names Begin with 'CCA'

- There are 5 records where launch sites begin with `CCA` :

[22]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- From the table, it appears that no successful landing attempt has been made.

Total Payload Mass

- The total payload carried by boosters from NASA

TOTAL_PAYLOAD
111268

- The total payload was determined above by adding up all of NASA's payloads.

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1

AVG_PAYLOAD
2928.4

- We determined the average payload mass to be 2,928 kg by filtering the data according to the booster version mentioned above.

First Successful Ground Landing Date

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

First_Successfull_Landing_Date
01-05-2017

- We noted that the first successful landing took place on the ground pad on May 1, 2017.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

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

- These 4 boosters from among various booster versions using the specified filters.

Total Number of Successful and Failure Mission Outcomes

- The number of successful and failure mission outcomes:

Mission_Outcome	Total
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

- The summary above was generated by grouping mission results and calculating the records for each category.

Boosters Carried Maximum Payload

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

- The List of the booster which have carried the maximum payload mass:
- Using a subquery in the WHERE clause and the MAX() method, we were able to identify the booster that had carried the most payload.

2015 Launch Records

- 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

- The 2 failed landing outcomes can be observed in the table above

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:

- The landing results were categorized using the GROUP BY clause, and they were then put in decreasing order using the ORDER BY clause.

Landing _Outcome	Quantity
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

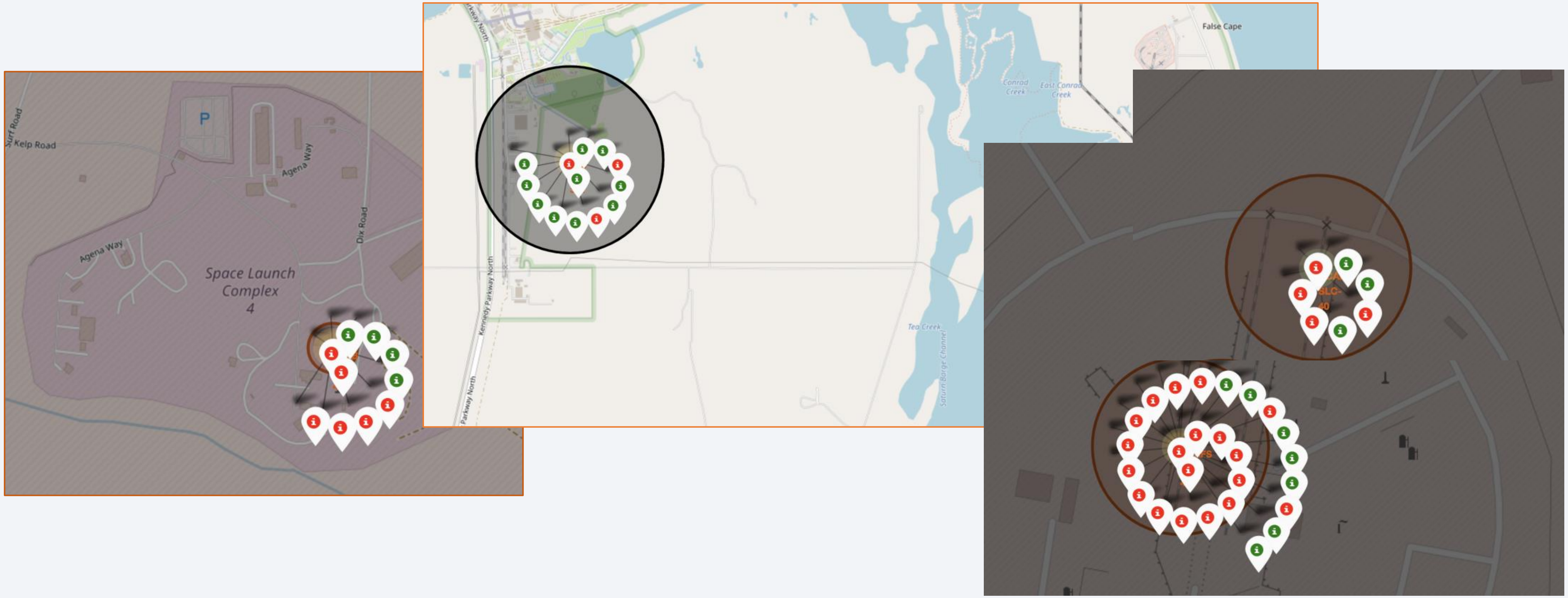
Section 3

Launch Sites Proximities Analysis

All SpaceX Launch Sites

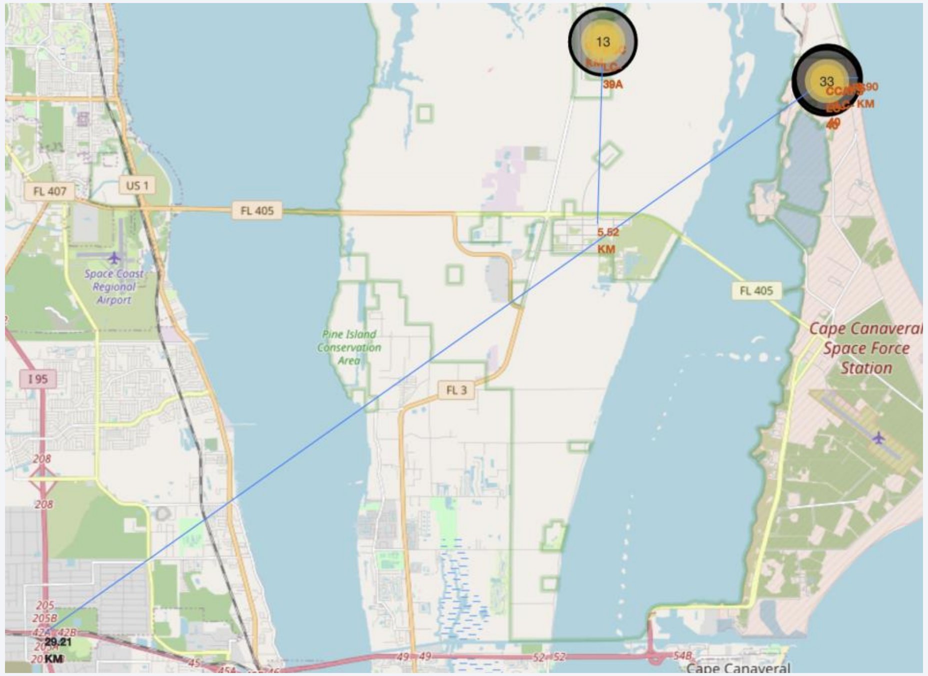
- In order to be safe, launch sites are close to the ocean yet still relatively close to highways and trains.

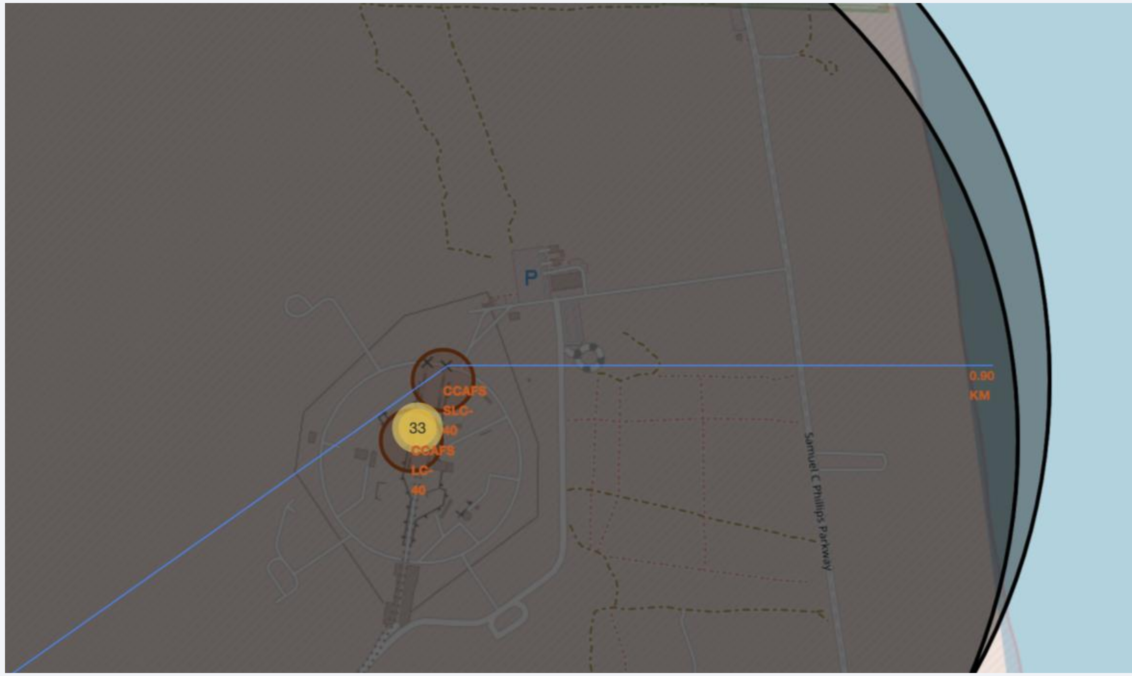
Launch Sites with the color labels



- Red and green marks indicate failure and success, respectively.

Launchsite and Proximities



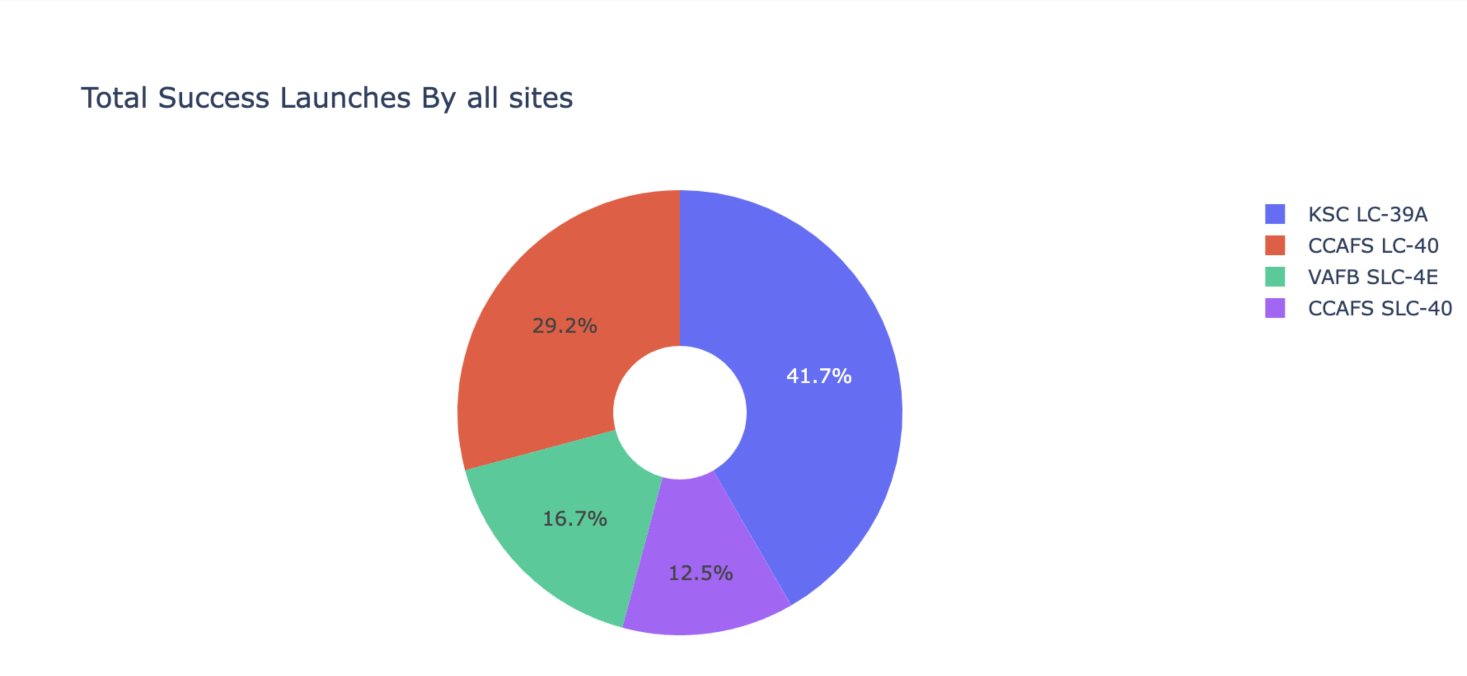


- CCAFS LC-40
- Highway distance: 29.21km, Coastline distance: 0.9km

Section 4

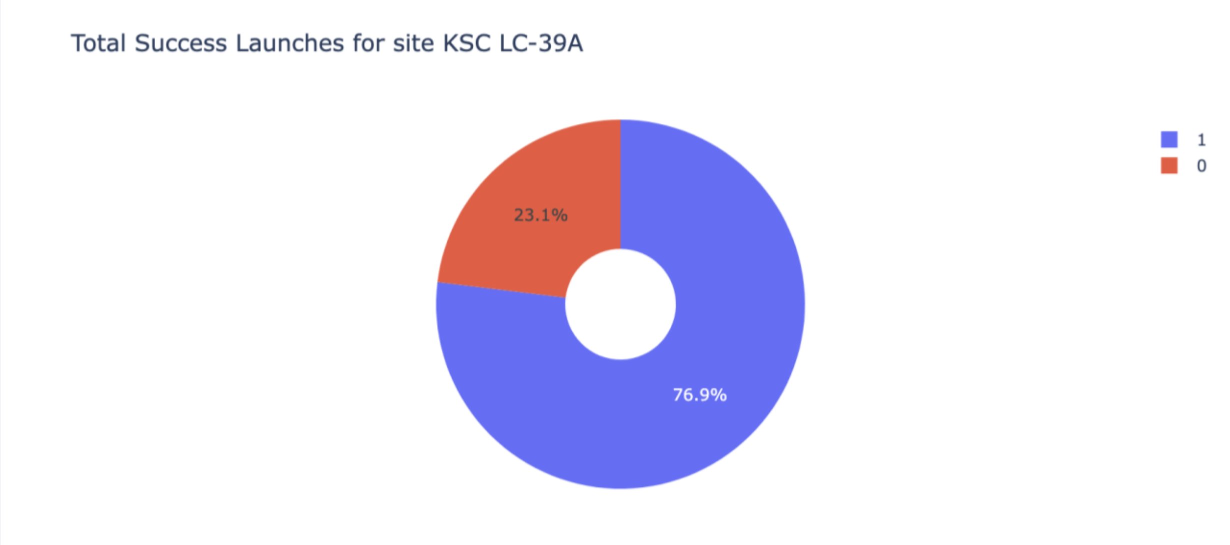
Build a Dashboard with Plotly Dash

Total Successful Launches by Site



- KSC LC-39A is the most successful site with %41.7 among all successful outcomes

<Dashboard Screenshot 2>



- KSC LC-39A achieved %76.9 success rate and %23.1 failure rate for all the rocket launches done from this site.

<Dashboard Screenshot 3>

- Better success rates can be observed for payloads less than 6000kg
- There are not many launches can be observed for payloads more than 6000kg

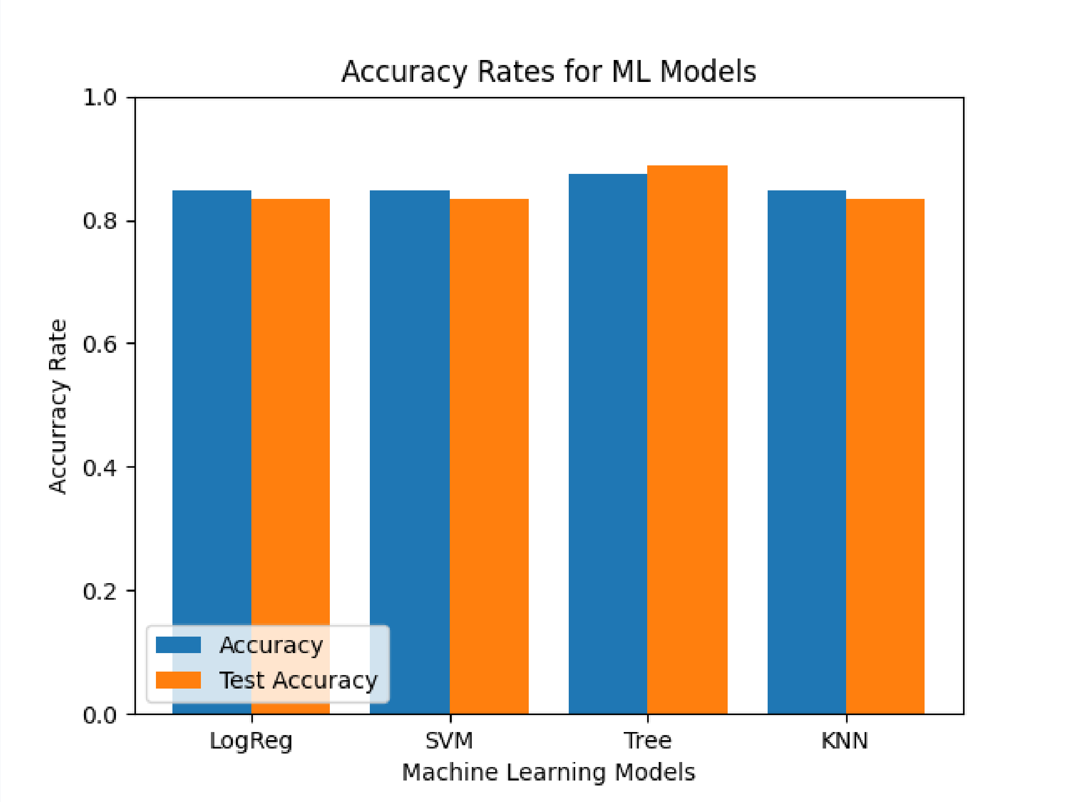
Section 5

Predictive Analysis (Classification)

Classification Accuracy

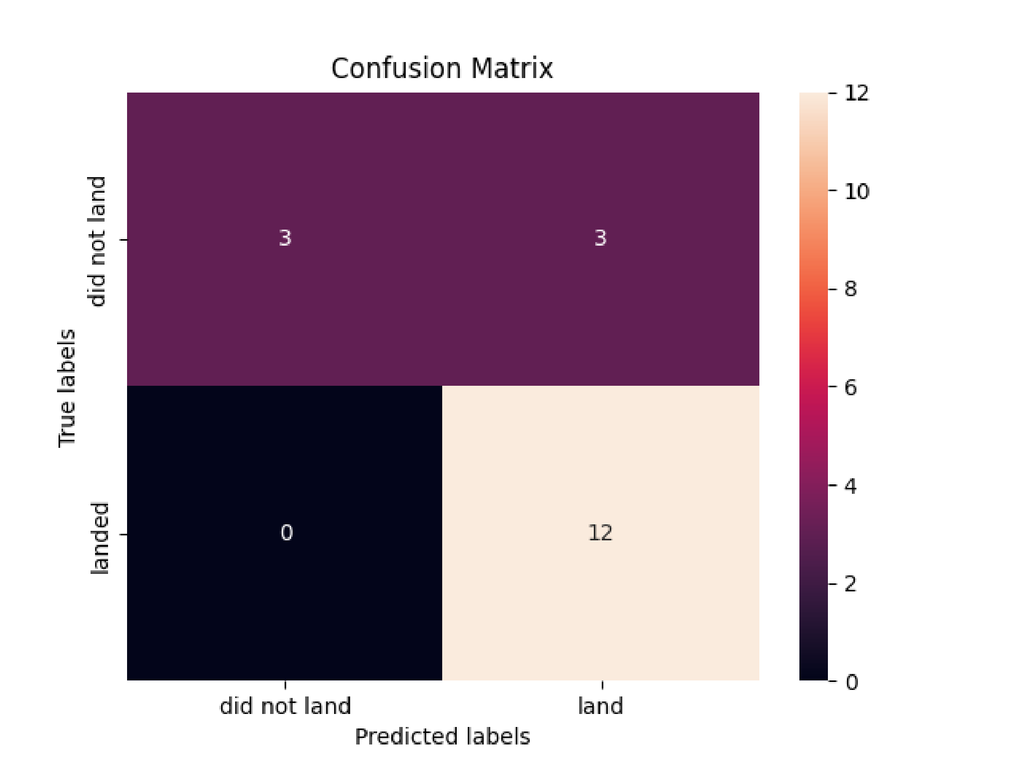
- Four categorization models were evaluated, and the accuracies of each are shown next to one another;

- Decision Tree Classifier, which has accuracy levels higher than 87%, is the model with the highest classification accuracy.



Confusion Matrix

- The decision tree classifier can discriminate between the several classes, as shown by the confusion matrix for the classifier. The main issue is false positives. For instance, unsuccessful landing marked as successful landing by the classifier.



Conclusions

- Launches with less than 6,000 kg payload weight have better success rate
- The launch success rate increased from 2013 to 2020.
- The highest success rate was in the ES-L1, GEO, HEO, SSO, and VLEO orbits.
- Among all the locations, KSC LC-39A had the most successful launches.
- The most effective machine learning method for this problem is the decision tree classifier.

Thank you!