IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Boluwatife Awoyemi
October 9, 2024

Outline

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

Executive Summary

- This study looks into predicting the first stage successful landing of SpaceX Falcon 9 spacecraft after launch. Unlike other spacecraft manufacturers, SpaceX boats of a much less costly rocket launch, and this is largely attributed to their ability to reuse the first stage launch crafts. This emphasizes the importance of SpaceX knowing the possibilities of their rockets landing successfully at the first-stage. This can help SpaceX invest more in the attributes that fetch more success that failures while also giving them an edge over other spacecraft manufacturers.
- We employ data science methodologies to:
    - extract existing data about SpaceX Falcon 9 spacecrafts,
    - query the data to understand what we are working with,
    - analyze and visualize the data to extract predictions for successful landings.

Introduction

Rockets are usually launched into space for many important reasons which include transporting payload and or humans into space to explore the planetary systems, deliver satellites, help conduct research and many other uses. However, manufacturing rockets, seeing rockets to a state of safety, successful launching and landing requires a lot of money.

Space X designed Falcon 9 rockets which serve as a two-stage rocket that boasts re-use provided the rocket landed successfully after its first launch stage. These rockets have been launched many times and SpaceX recorded the data for these events.

We want to study this existing data of first-stage Falcon 9 launches and to draw insights on the attributes that lead to successful and unsuccessful landings of the rockets at the first stage. These insights and predictions that help SpaceX understand the correlation and causal relationships between different attributes that constitute for successful landings and the company can focus on improving their rockets landing successes and reputation with this information.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - We use the Python BeautifulSoup library to scrap the Falcon 9 launch records HTML tables from Wikipedia, we parse the data and convert it to an accessible data frame using the Pandas library. We also use the requests library to make HTTP requests to the SpaceX API to get data..
- Perform data wrangling
  - After organizing all the collected data into a data frame, we used Pandas and Numpy libraries to deal with missing values, identify patterns in the data and determine the response variable label for training unsupervised models.

Methodology Cont.

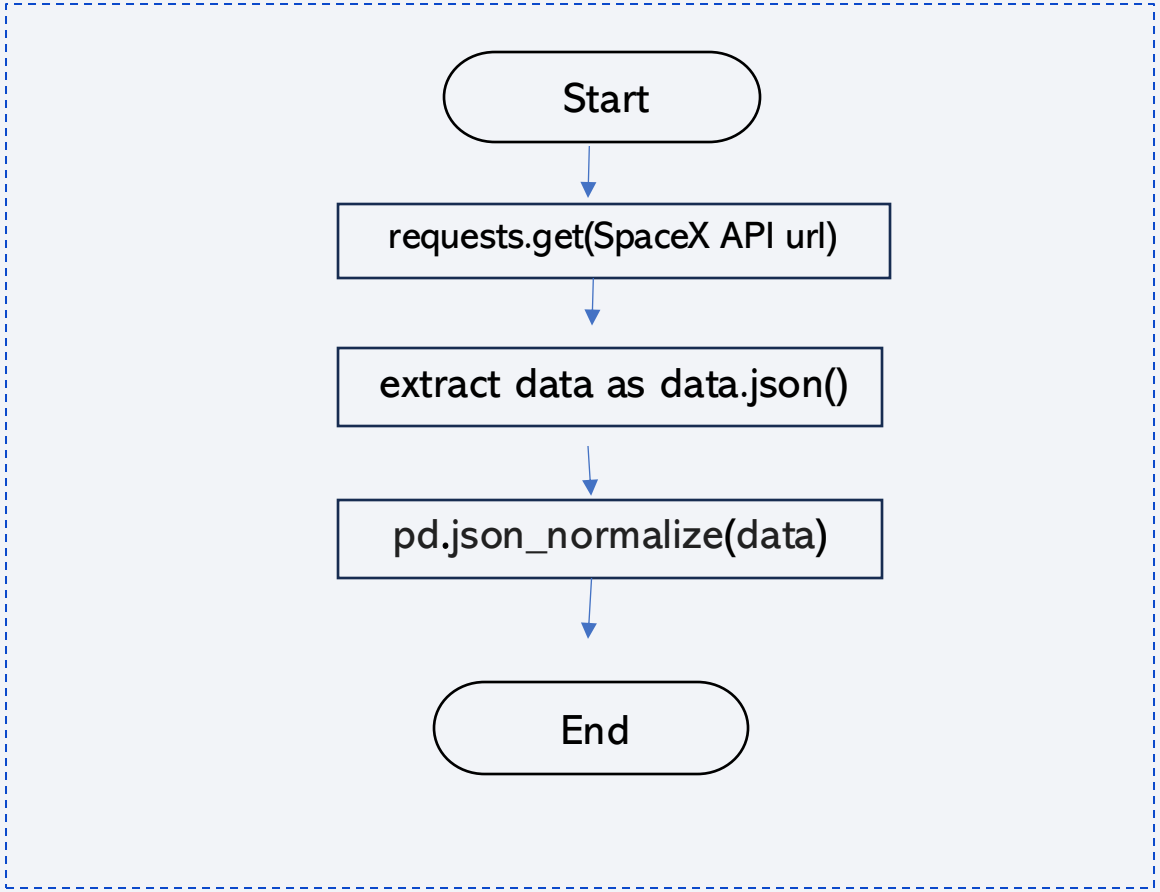
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - For the data frame with engineered features, we separate the attributes and response variables into different arrays, we normalize the feature array, split the arrays into training and test data, and we apply Machine Learning classification models to the data while incorporating search for best parameters.

Data Collection

- We collect data from Wikipedia by scraping HTML content using BeautifulSoup library in Python source for data We also supplemented this data by using get requests to source for data with API calls on the SpaceX API.
- We combine the data from these sources to one data frame before proceeding to wrangling the data.

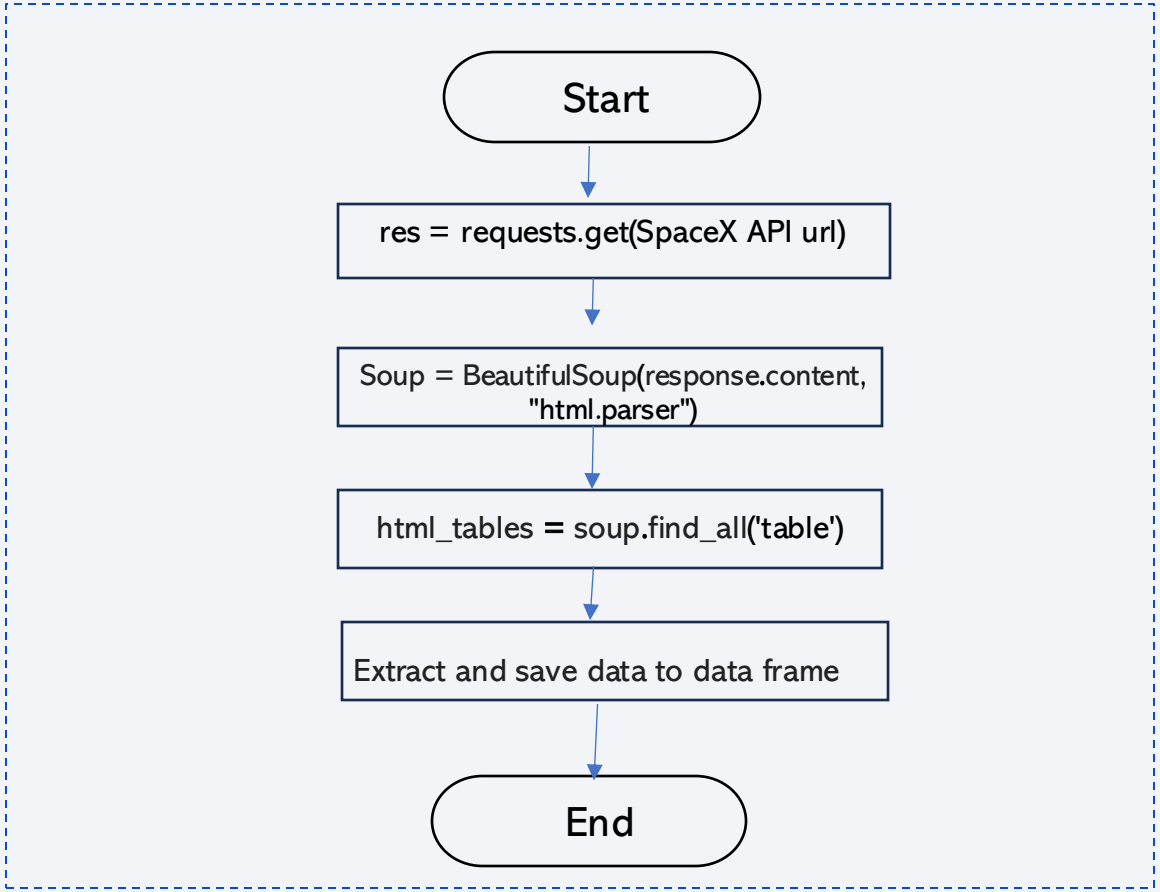
Data Collection – SpaceX API

- https://github.com/TifeAwoyemi/IBM-Data-Science/blob/main/1.%20SpaceX%20Webscraping.ipynb



Data Collection - Scraping

- https://github.com/TifeAwoyemi/IBM-Data-Science/blob/main/1.%20SpaceX%20Webscraping.ipynb



Data Wrangling

- We start by identifying the total number of a. launches per launch site, b. orbits per orbit and c. landing outcomes per landing outcome type. Next, we identified and stored the set of outcomes where landing was unsuccessful at the second stage. We then use this saved data to properly identify create a class column that has 0 where land was unsuccessful at the second stage and 1 where landing was successful.
- This helps us to determine the success rate of launches for all sites and per site. By applying .mean() to the class column of the data frame filtered by launch site where applicable.
- https://github.com/TifeAwoyemi/IBM-Data-Science/blob/main/3.%20SpaceX%20Data%20Wrangling.ipynb

EDA with Data Visualization

- We use the seaborn library to visualize the data with appropriate charts.
- We use scatter plots to visualize the relationship between two attributes and their landings.
- We use a bar chart to show the relationship between attributes like different orbits and their landing success rate.
- We use a line chart to show the trend of successful landings over the years.
- https://github.com/TifeAwoyemi/IBM-Data-Science/blob/main/5.%20SpaceX%20EDA%20Visualizations%20and%20Feature%20Engineering.ipynb

EDA with Feature Engineering

- We apply the pandas.get_dummies() function to categorical features that we use in success prediction. This function encodes the categorical variables for easy application to classification.

EDA with SQL

Some of the SQL queries we run include:

- We show the unique Launch Sites in the data with %sql select distinct Launch_Site from SPACEXTBL
- We display 5 records where launch sites begin with the string 'CCA' with %sql select * from SPACEXTBL where Launch_Site like "CCA%" limit 5
- We display the total payload mass carried by boosters launched by NASA (CRS) with %sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer = "NASA (CRS)"
- We display average payload mass carried by booster version F9 v1.1 with %sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version = "F9 v1.1"

EDA with SQL

- We select the date when the first succesful landing outcome in ground pad was achieved with %sql select min(Date) from SPACEXTBL where Landing_Outcome like "%ground%"
- We list the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000 with %sql select Booster_Version from SPACEXTBL where Landing_Outcome = "Success (drone ship)" and PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000
- https://github.com/TifeAwoyemi/IBM-Data-Science/blob/main/4.%20SpaceX%20EDA%20SQL.ipynb

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- In this section, we continue the exploratory data analysis by looking at interactive maps using the Folium library in Python
- We added markers to the map to identity the different launch sites for easy access and zooming.
- We added circles to the map to highlight each launch site with a text label based on its coordinates
- We use lines to connect a launch site to a selected coastline point to see the proximaity of some landmarks to the launch sites.
- https://github.com/TifeAwoyemi/IBM-Data-Science/blob/main/6.%20SpaceX%20Analysis%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

- We add pie charts to the graph that show the percentage of success landing to failed landing when a launch site is selected. The pie chart adjusts to show the percentage of successful landings for all launch sites if all sites are chose from the dropdown menu.
- We also add a scatter plot that reacts to the dropdown selection of launch sites and a Payload mass range slider. This scatter plot shows the relationship between payload mass and landing outcomes by booster version category. This helps us visualize performance of different booster versions when considering the rocket's payload mass.
- https://github.com/TifeAwoyemi/IBM-Data-Science/blob/main/7_SpaceX_Dash_App.py

Predictive Analysis (Classification)

- We employ four different classification models to predict successful landings. We use the cross-validation technique "GridSearchCV" to select the optimal parameters for each model from a given set of parameters. We set a 10-fold cross validation (that is, cv = 10) with the GridSearchCV() for all models for fairness in comparison.
- https://github.com/TifeAwoyemi/IBM-Data-Science/blob/main/8.%20SpaceX%20Machine%20Learning%20Prediction.ipynb

Results

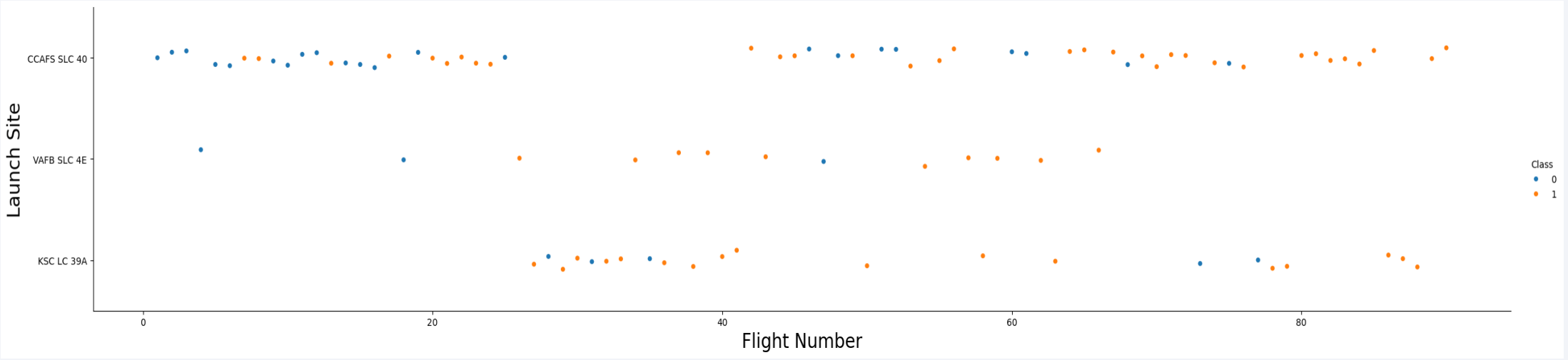
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

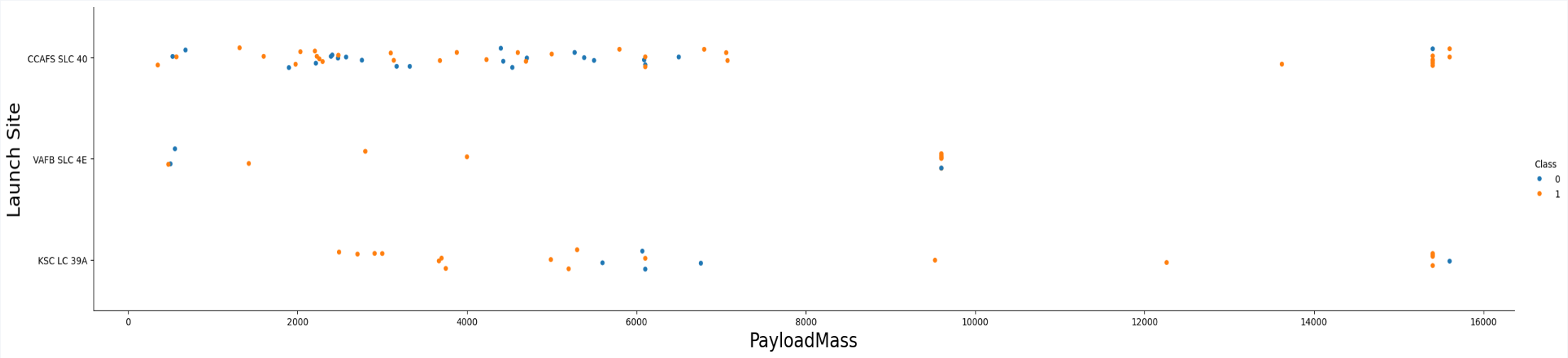
Boluwatife
Awoyemi's Project

Flight Number vs. Launch Site



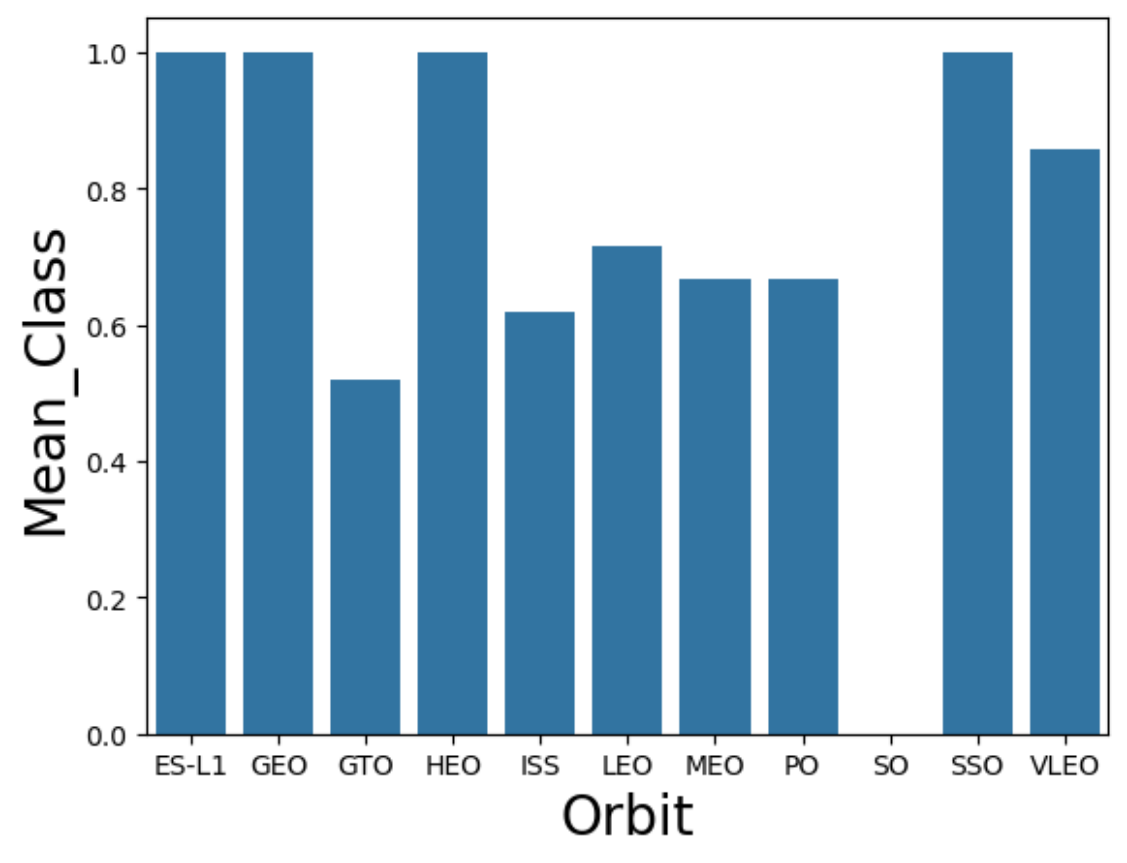
- Each launch site sees a good amount of successful launch.
- If we consider the launches per site, VAFB SLC 4E has the highest percentage of successful with respect to the total launches they had.

Payload vs. Launch Site



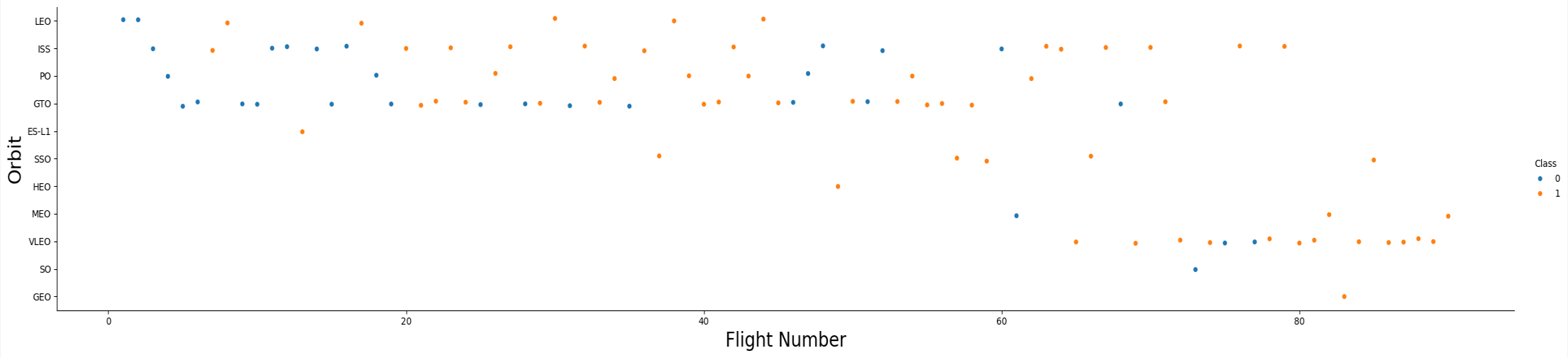
- The total number of launches for all sites reduced drastically as payload mass exceeded 7000kg.
- VAFB SLC 4E gets affected the most because they do not launch any Falcon 9 rocket with payload mass higher than 10000kg.

Success Rate vs. Orbit Type



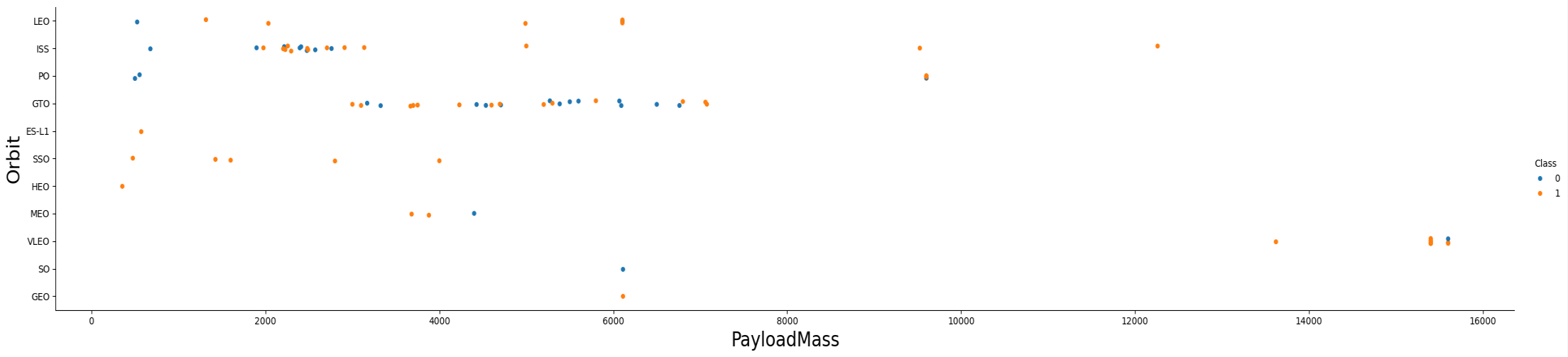
- The Orbits ES-L1, GEO, HEO and SSO seem to have the highest success rates of landing.
- On the other hand, the Orbit SO seems to not have had any success in Falcon 9 landing.

Flight Number vs. Orbit Type



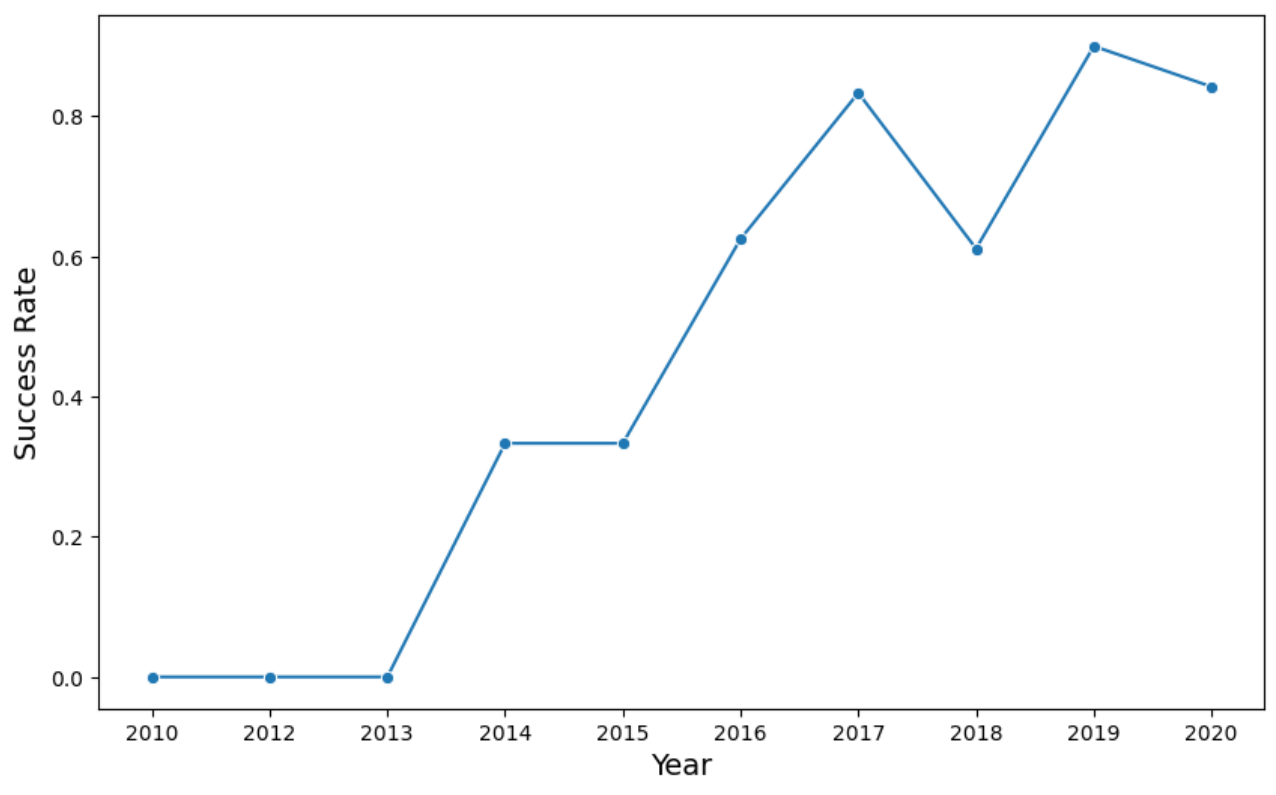
- The scatter plot helps us understand the bar chart better. From the bar chart, of all the orbits with about 100% success rate, only Orbit SS) has more than one launch. The other three orbits only seem to see one rocket launch which happen to land successfully.
- The SO Orbit that saw no successful launch as seen on the bar chart is confirmed with the scatter plot. This Orbit only saw one flight that did not land successfully.
- The Orbits LEO, ISS, PO, GTO and VLEO launched the most flights and SSO is the only orbit with more than one launch that saw only successful landings.

Payload vs. Orbit Type



- Many orbits do not launch flights for high payload mass.
- ES-L1 and HEO only saw rockets launch with a payload mass less than 1000kg.
- VLEO is the only orbit that does not launch any flight for any payload mass up to about 13000kg but saw flight launches with successful landing for payload mass higher than 13000kg.

Launch Success Yearly Trend



- We see that no successful launch was recorded between 2010 and 2013, but the launch success gained an upward trend after 2013.
- The year 2019 saw the highest launch success rate.

All Launch Site Names

Find the names of the unique launch sites

- As seen in the picture, there are four different launch sites in the data and two of these launch sites are at Cape Canaveral Space Force Station in Florida.

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

Launch Site Names Begin with 'CCA'

Find 5 records where launch sites begin with `CCA`

- The image shows 5 records with launch sites located at the Cape Canaveral Space Force Station in Florida.

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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 attempt

Total Payload Mass

Calculate the total payload carried by boosters from NASA

- The result for the total payload carried by NASA boosters was 45596kg. This seems to be a lot of payload.

Average Payload Mass by F9 v1.1

Calculate the average payload mass carried by booster version F9 v1.1

- The result for the average payload carried by booster version F9 v1.1 was 2928.4kg. This seems to be on the low end of average payload.

First Successful Ground Landing Date

Find the dates of the first successful landing outcome on ground pad

- The first day a ground landing was successful (based on the data) was the December 22, 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 B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

- There are four total boosters which saw successful landing on drone ships for payload mass between 4000kg and 6000kg.

Total Number of Successful and Failure Mission Outcomes

Calculate the total number of successful and failure mission outcomes

- There were a total of 100 successful mission outcomes and 1 failure.

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

Boosters Carried Maximum Payload

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

- A total of 12 different boosters listed in the picture carried maximum payload during a launch.

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

2015 Launch Records

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

- In January and April 2015, one failure to land on drone ships was recorded for each month for the CCAFS LC-40 launch site but the rockets had different booster versions.

Month	Booster_Version	Landing_Outcome	Launch_Site
01	F9 v1.1 B1012	Failure (drone ship)	CCAFS LC-40
04	F9 v1.1 B1015	Failure (drone ship)	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

- The picture shows a list of total landing outcomes for all 8 outcomes types in descending order between June 4, 2010 and March 20, 2017.

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

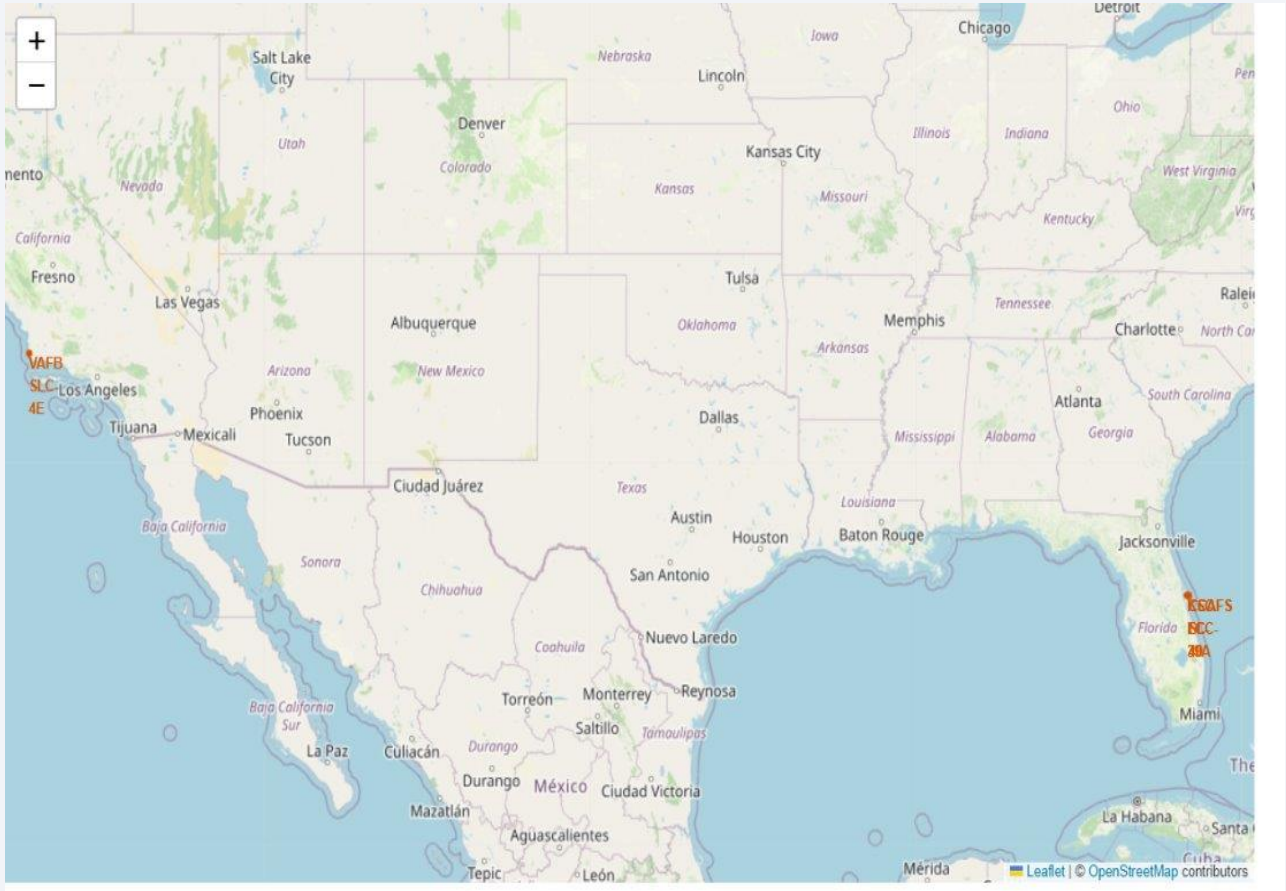
Section 3

Launch Sites Proximities Analysis

Boluwatife Awoyemi's Project

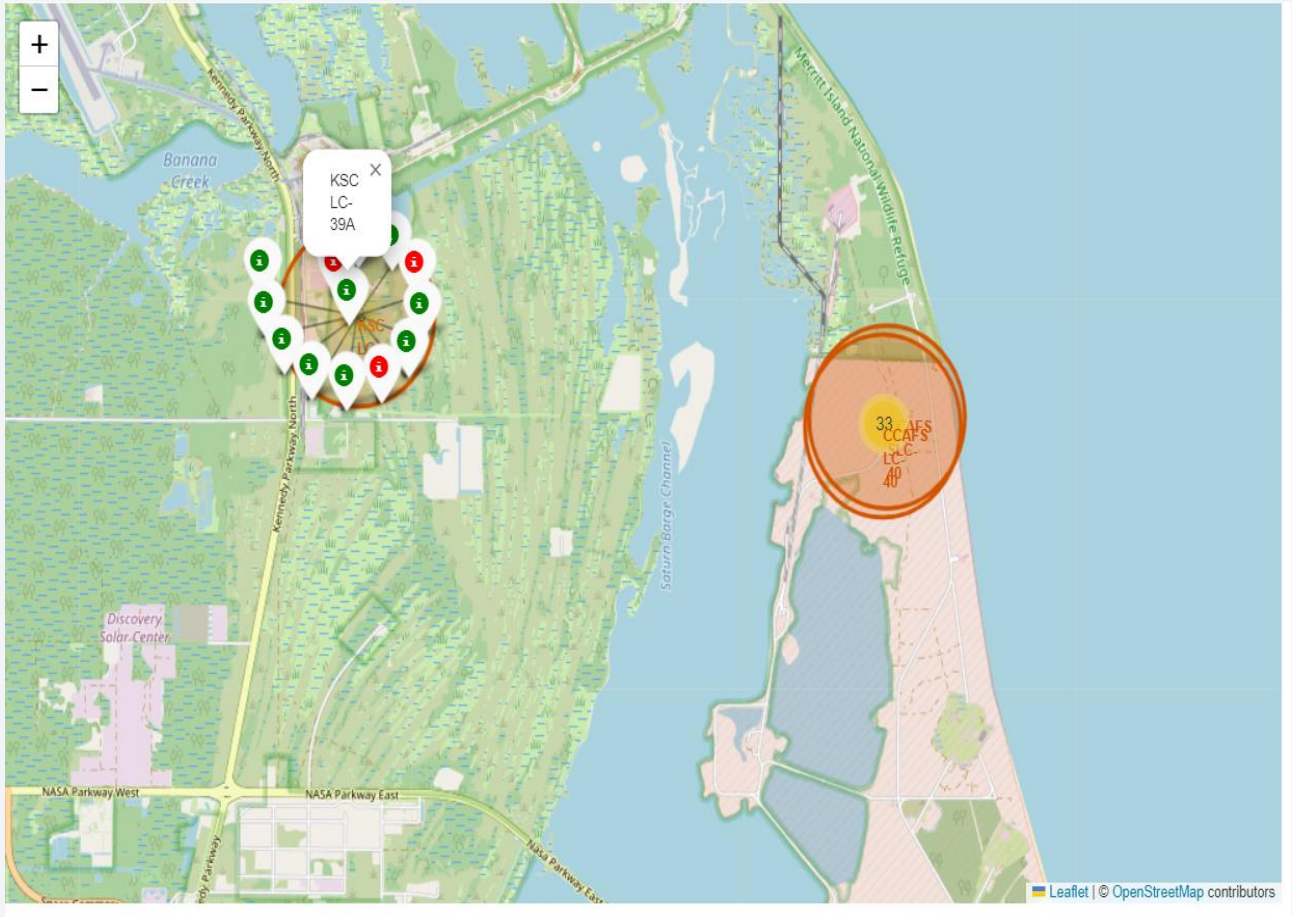
Launch Site Location Markers on World Map

- The launch site are clustered in two regions which are northwest and southeast of each other.
- The launch sites are also very close to the coastlines which is reasonable due to launches with drone ships.



Color-labeled Launch Sites Outcomes by Launch Sites

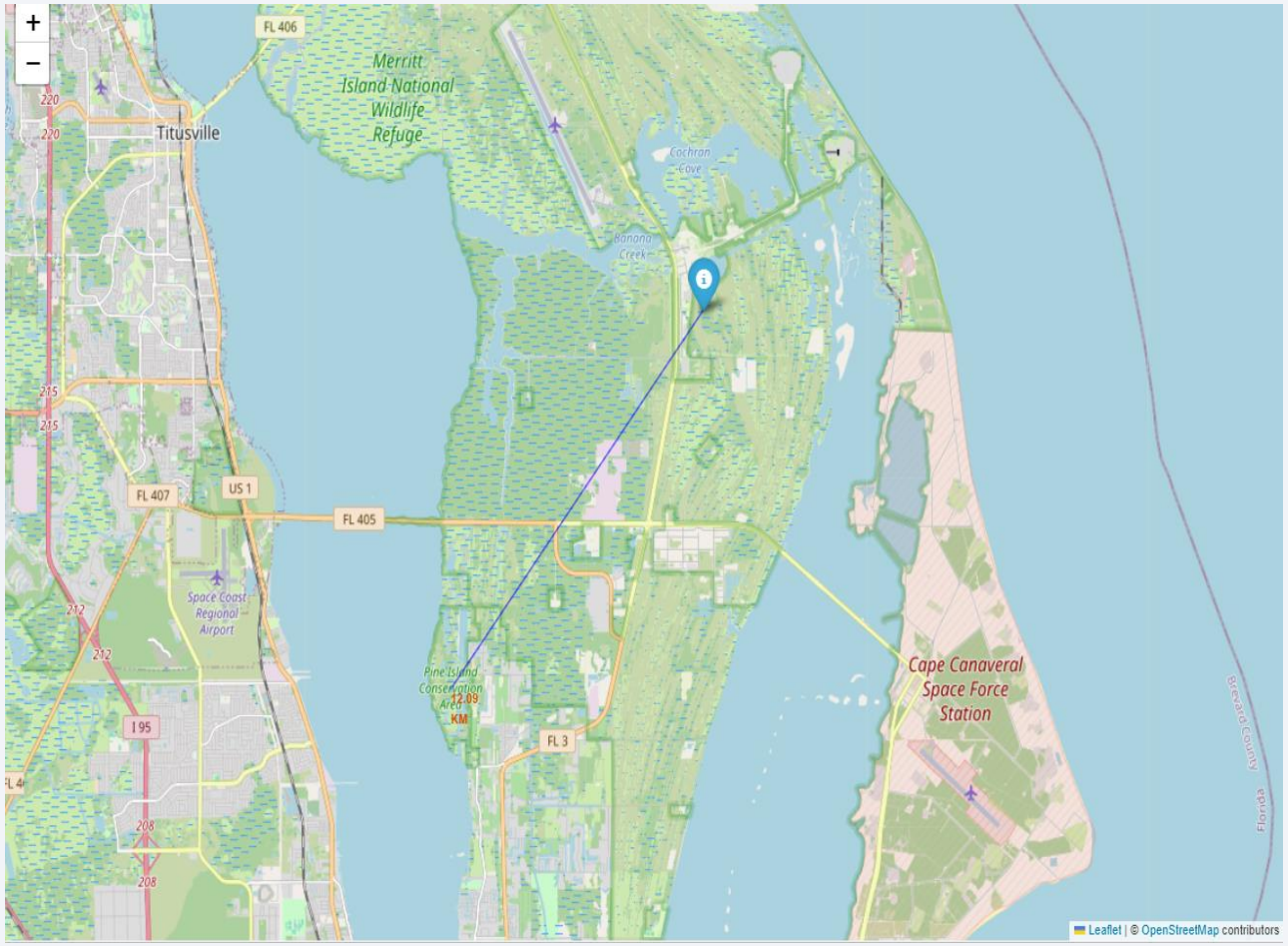
- In the cluster selected, many successful landing outcomes (with green markers) are recorded.
- One of these launch outcomes come from the KSC LC-39A launch site.



Proximity of CCAFS SLC-40 to the Pine Island Conservation Area

- The CCAFS SLC-40 launch site is about 12.01 KM away from the Pine Island Conservation area.

- This highlights how launch sites are significantly far away from key places like conservation areas, and generally cosmopolitan areas.

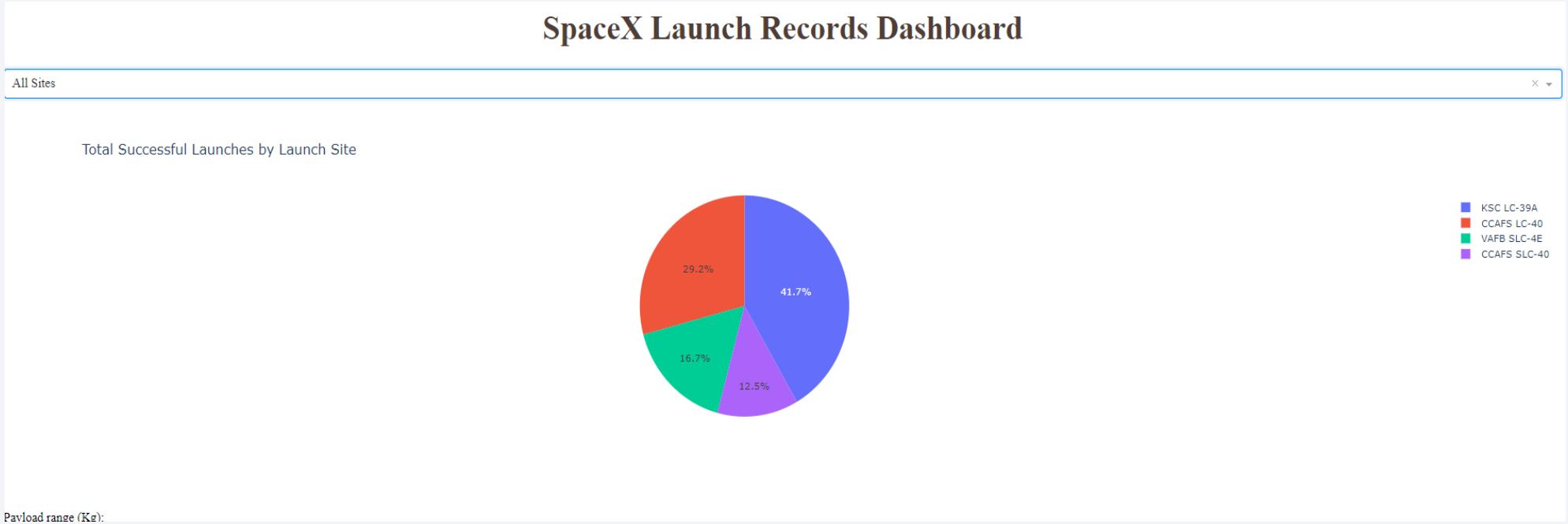


Section 4

Build a Dashboard with Plotly Dash

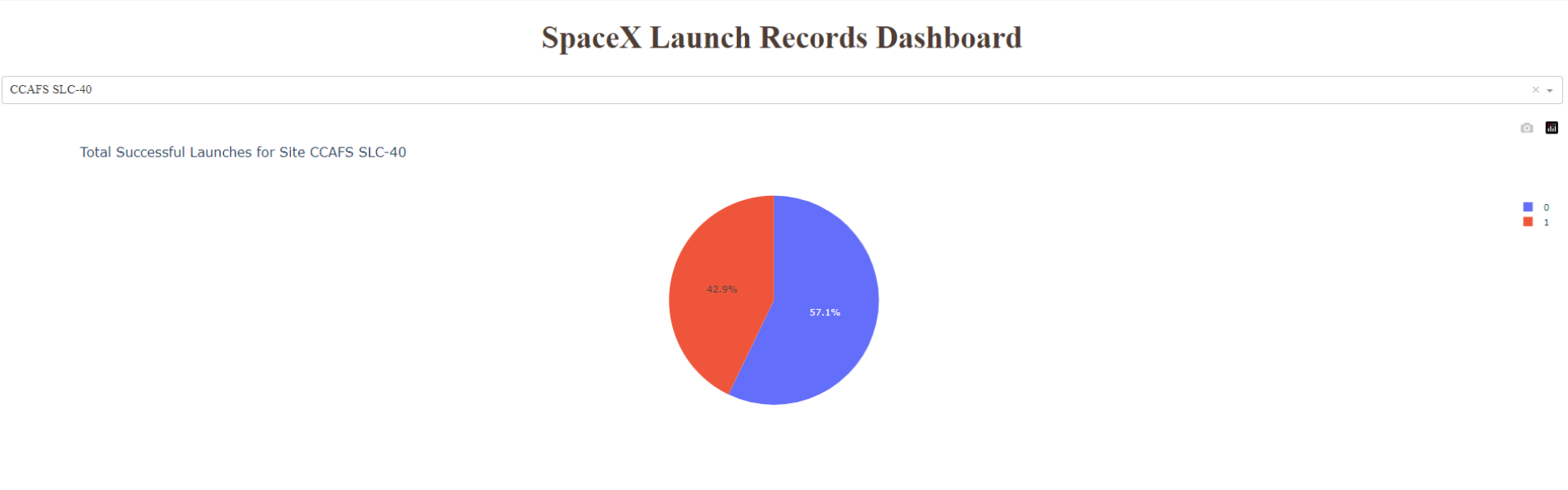
Boluwatife Awoyemi's Project

Percentage of Success Rates for Launch Sites



- The Pie chart shows the percentage of successful launches for each site when "All sites" menu is selected.
- KSC LC-39A had the highest landing success rate while CCAFS SLC-40 has the least.

Launch Site with the Highest Launch Success Ratio

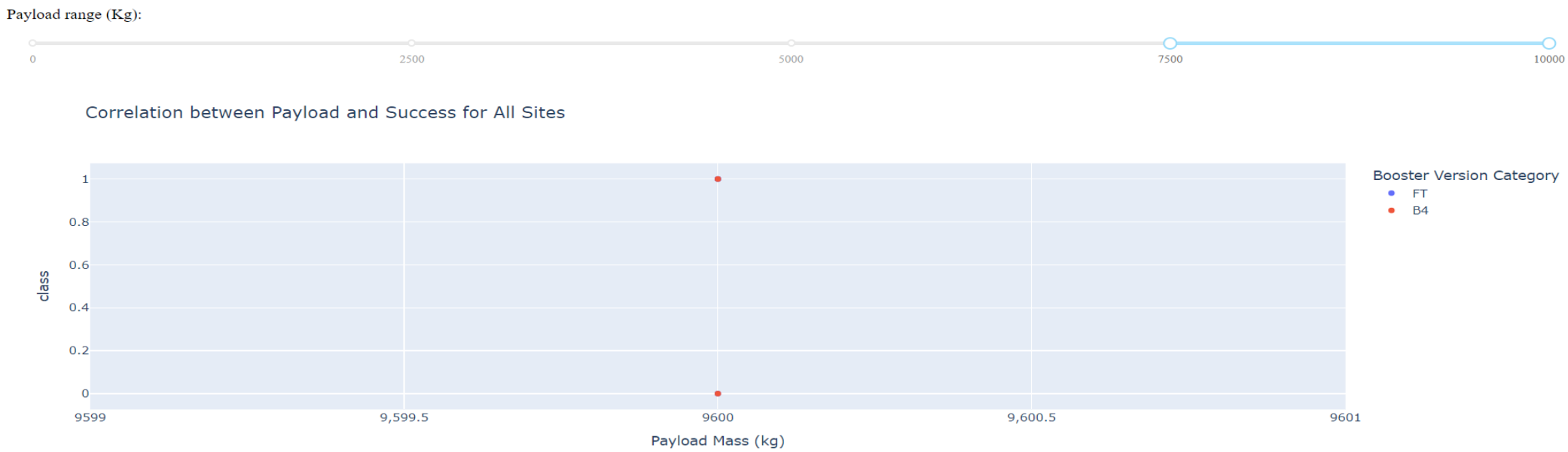


- Though, none of the sites has a higher success rate than failure rate, when compared to the other launch sites, CCAFS SLC-40 has the higher launch success ratio of 0.42.

Payload vs. Launch Outcome for All sites

- Payload range between 2000kg and 4000kg boast of slightly higher success rate that the other payload ranges. However, the ratio of success to failure in this range seems close too.
- The FT booster version seems to enjoy the highest rate of successful outcomes.

Payload vs. Launch Outcome for All sites



- At very high payload mass, only the FT and B4 booster versions see any rocket launch and these launches record both successes and failures to land at exactly 9600kg payload mass.

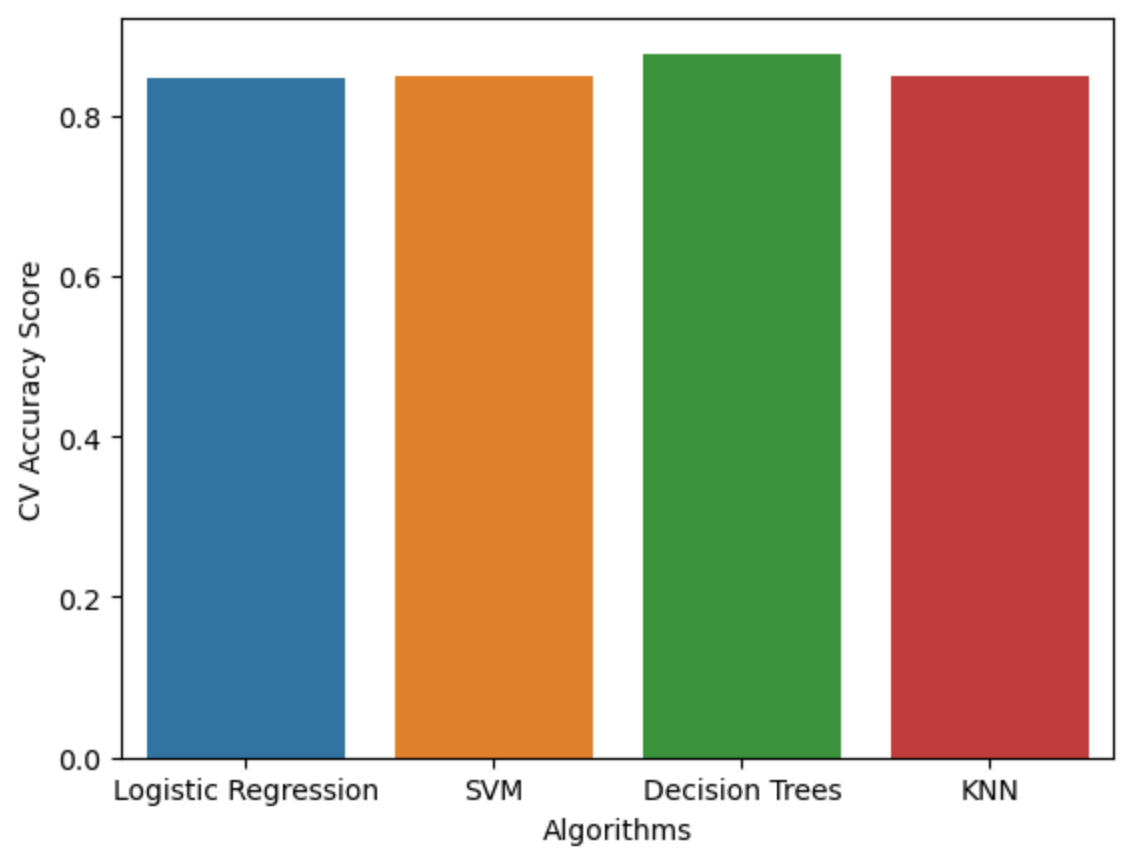
Section 5

Predictive Analysis (Classification)

Boluwatife
Awoyemi's Project

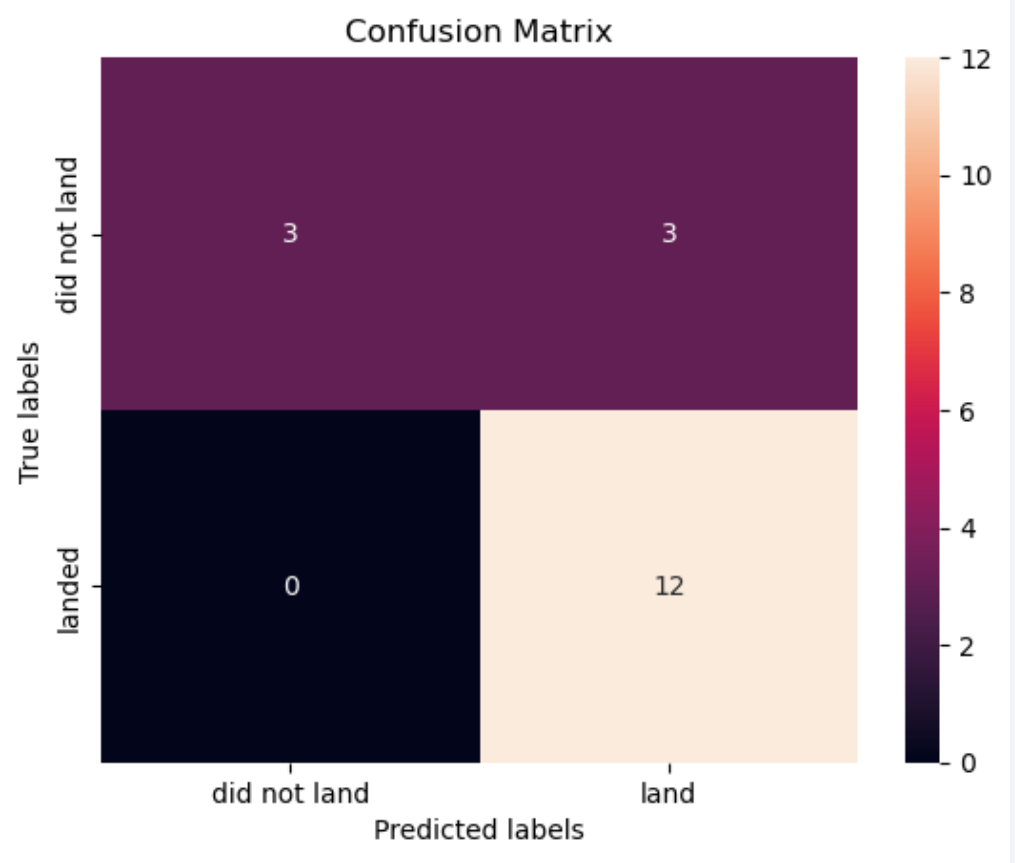
Classification Accuracy

- From the bar chart, the Decision Trees model has the highest classification accuracy (0.88).
- This means that the Decision Trees model avoids overfitting and generalizes to unseen data a little better than the other models.



Confusion Matrix

- This is the confusion matrix for the test data using the Decision Trees model.
- The size of the test data is 18 meaning the sum of all the data on the matric is 18.
- In this test data, there are 0 false negatives, 12 true positives, 3 true negatives and **3 false positives**.
- The problem lies with the 3 false positives where the model predicted successful landings where there were actually failed landings. This means 3 models incorrectly predicts 3 successful outcomes.



Conclusions

- We studied the SpaceX Falcon 9 rocket launch past data to predict successful landings of rockets at their first-stage.
- In terms of payload and orbits, the VLEO orbits was seen to be the only orbit that saw rocket launches for very high payload.
- We noted that there has been an upward trend of successful launch rates since 2013, and with the advancement in studies and technologies, this is likely to continue.
- We noted that launch sites and payload mass have a relationship such that launch sites like VAFB SLC-4E don't venture into launching rockets with high payload mass while others do.
- Lastly, the Decision Trees model seems to work best in predicting successful landing outcomes for the SpaceX Falcon 9 because it boasts a higher cross validation accuracy score than other models.

Appendix

- https://github.com/TifeAwoyemi/IBM-Data-Science/blob/main/dataset_part_1.csv
- https://github.com/TifeAwoyemi/IBM-Data-Science/blob/main/dataset_part_2.csv
- https://github.com/TifeAwoyemi/IBM-Data-Science/blob/main/dataset_part_3.csv
- https://github.com/TifeAwoyemi/IBM-Data-Science/blob/main/spacex_web_scraped.csv

Thank you!

Boluwatife
Awoyemi's Project