IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mohammad Al Balki
2/9/2022

Outline

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

Executive Summary

- By using python data science libraries and web scraping libraries we've collected the datasets needed and then we performed data wrangling using pandas, some EDA and visualizations using Seaborn library, and then we made a map and dashboard using Folium and Plotly, finally we've built some classification models to predict if the landing will success or not.

- Many results were interesting, the datasets were clean but not big enough to determine a certain results, by using Grid search, every model performed the same.

Introduction

- We're trying to study SpaceX Falcon 9 landing success rates, using some information about many past tries through 2010-2014
- We need to answer these questions :
    1. Is there any relationship between the landing outcome of Falcon9 attempts and any other factors?
    2. Can we use pre-calculated and defined information to predict the outcome of any spaceX launch?
    3. What's the best machine learning model to predict the landing outcome In this case?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using SpaceX API and Web Scraping via Python Beautiful Soup module.
- Perform data wrangling
  - After exploring the datasets, there were some missing values on many columns, so we fixed that, and then we added some important new columns such as Class.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - We've built many classification models, and then used Grid Search CV to find the best hyperparameters and to achieve the best possible accuracy.

Data Collection

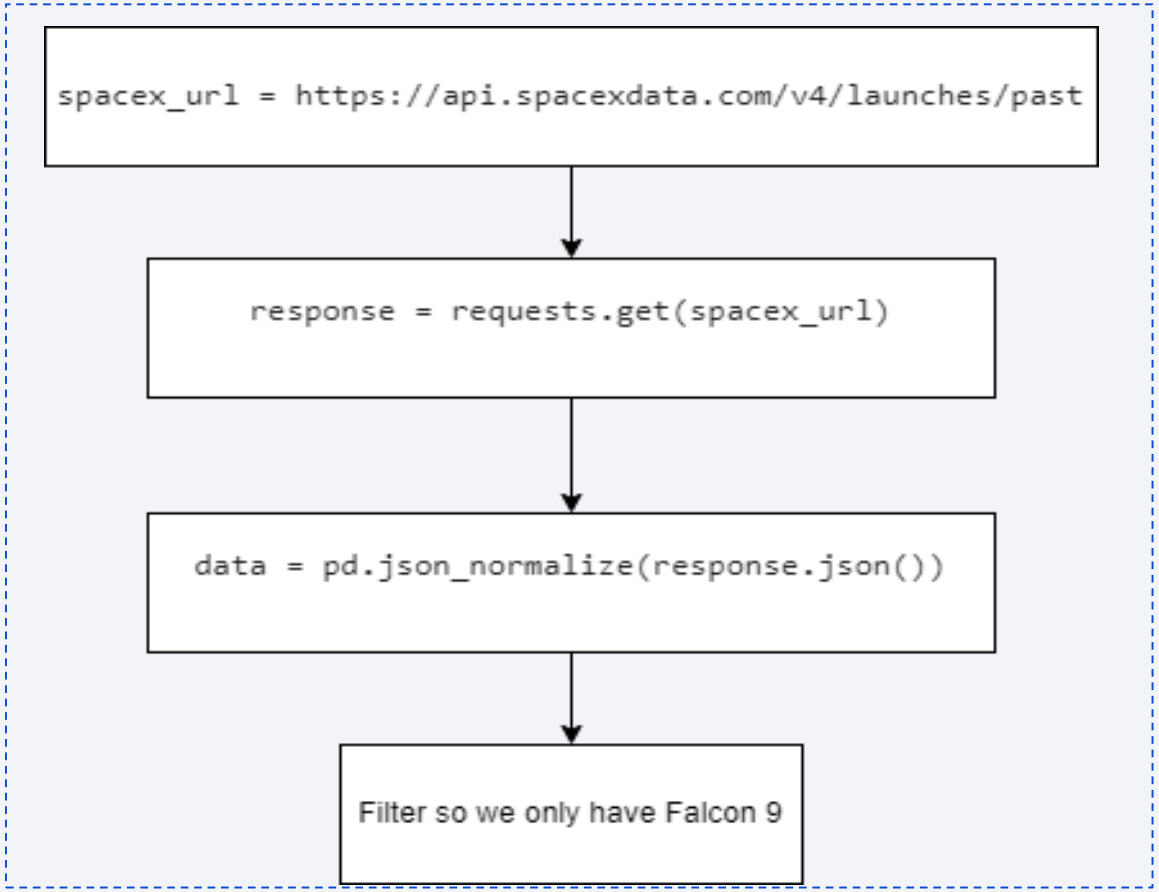
- Datasets were collected from SpaceX API.
- Also some of the data were collected using webscraping 'Beautiful Soup'.
- Using requests module of python, we preformed a Get on SpaceX API, and got our response which was converted to json later on and stored as Pandas Dataframe.
- Using Beautiful Soup module of python, we've taken many various information about many launches from Wikipedia.

Data Collection – SpaceX API

- Using requests.get we've got the response from spaceX API, and then stored it into a Dataframe.

- GitHub:

https://github.com/mohamedBalkhi/AI_Capstone/blob/main/DataCollectionAPI.ipynb

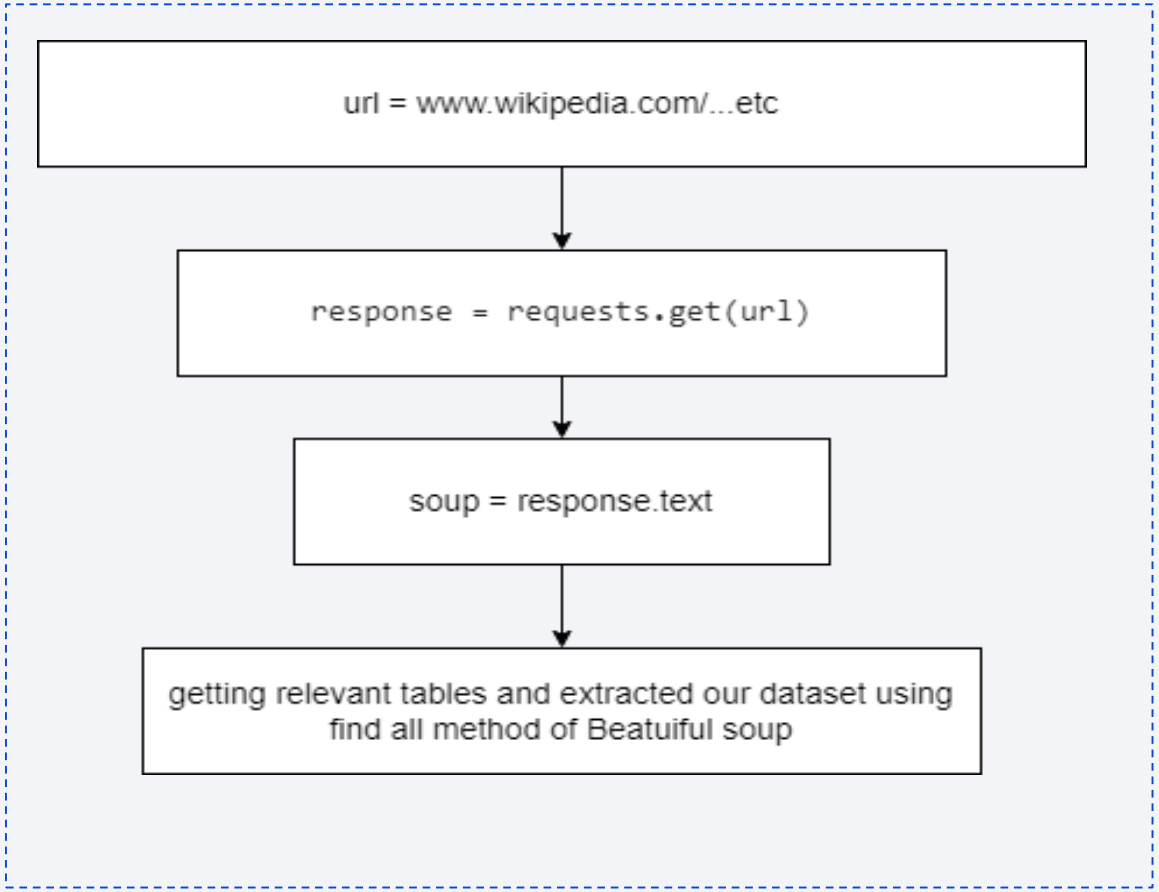


Data Collection - Scraping

- After getting the response we've used BeautifulSoup to scrape tables and information from HTML content.

- GitHub:

https://github.com/mohamedBalkhi/AI_Capstone/blob/main/Data_Collection_with_Web_Scraping.ipynb



Data Wrangling

- Missing values were dropped or replaced with mean, some columns were added to help in classification later on.
- Using Pandas methods we checked for null values and dealt with them
- We added Class column which determines the success of a Launch.

- GitHub:

https://github.com/mohamedBalkhi/AI_Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- We've made a scatter plot between many columns such as :
    1. Flight Number Vs Payload Mass
    2. Payload Mass Vs Launch Site
    3. Flight Number Vs Launch Site
    4. And many others.
- Also we've made a Bar plot to show the success rates of each Orbit.
- Line plot to show the launch success rates trend yearly.
- GitHub: https://github.com/mohamedBalkhi/AI_Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

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

- GitHub : https://github.com/mohamedBalkhi/AI_Capstone/blob/main/jupyter-labs-eda-sql-coursera.ipynb

Build an Interactive Map with Folium

- We've marked all the Launch Sites, and made a Circle around them, also added some markers for each Launch try 'red' for failure and 'green' for success tries.
- It gives a really good perspective how were the launches distributed around those launch sites and whether the location has any effect on the results.
- GitHub: https://github.com/mohamedBalkhi/AI_Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- We've added a pie chart to show success rates of different launch Sites, with a dropdown menu to control which site to show.
- That was done to give an outline and interactive plots and Dashboard.

- GitHub:

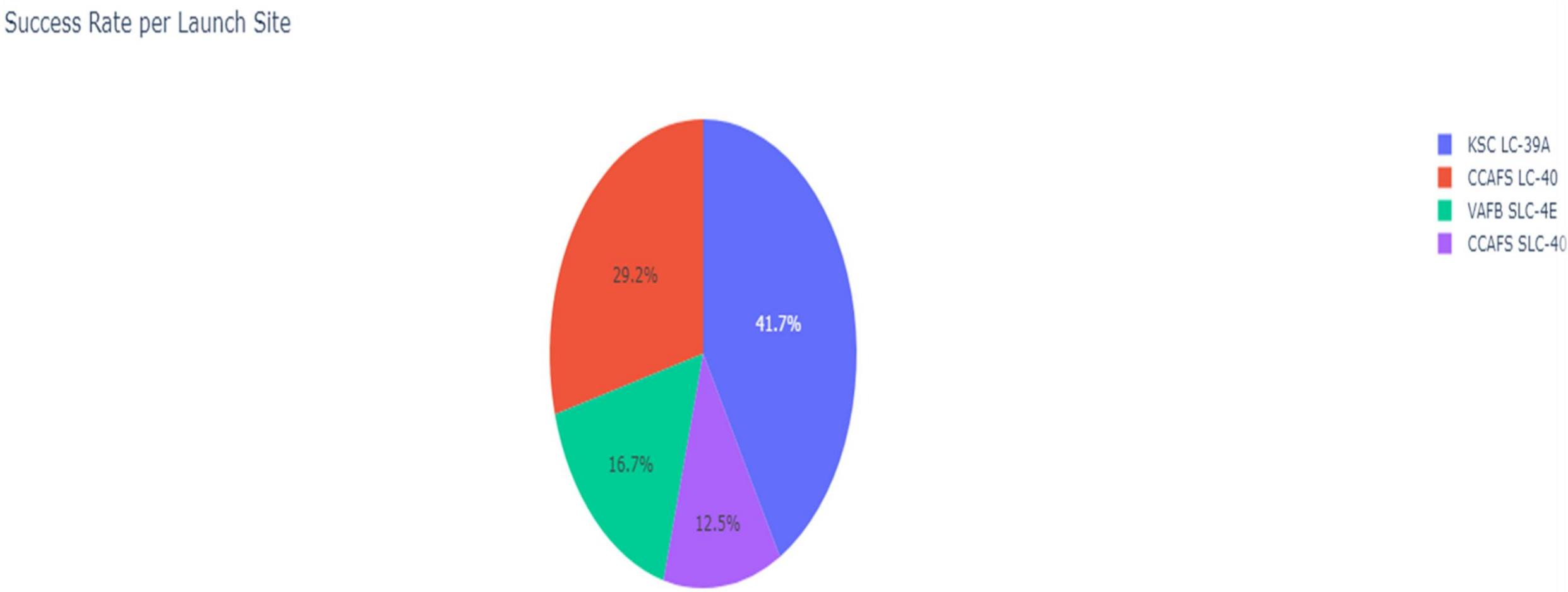
https://github.com/mohamedBalkhi/AI_Capstone/blob/main/dash.py

Predictive Analysis (Classification)

- I've used Grid Search CV to find the best parameters for each Model of the following : (Logistic Regression,Decision Tree,SVM,KNN) and then calculate the accuracy for each model.
- Model object creation => Grid Search CV with cv = 10, fit the model with X_train and Y_train, calculate the accuracy using score method on X_Test.
- GitHub: https://github.com/mohamedBalkhi/AI_Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results (EDA)

- We see that as the flight number increases, the first stage is more likely to land successfully.
- For VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).
- The sucess rate since 2013 kept increasing till 2020



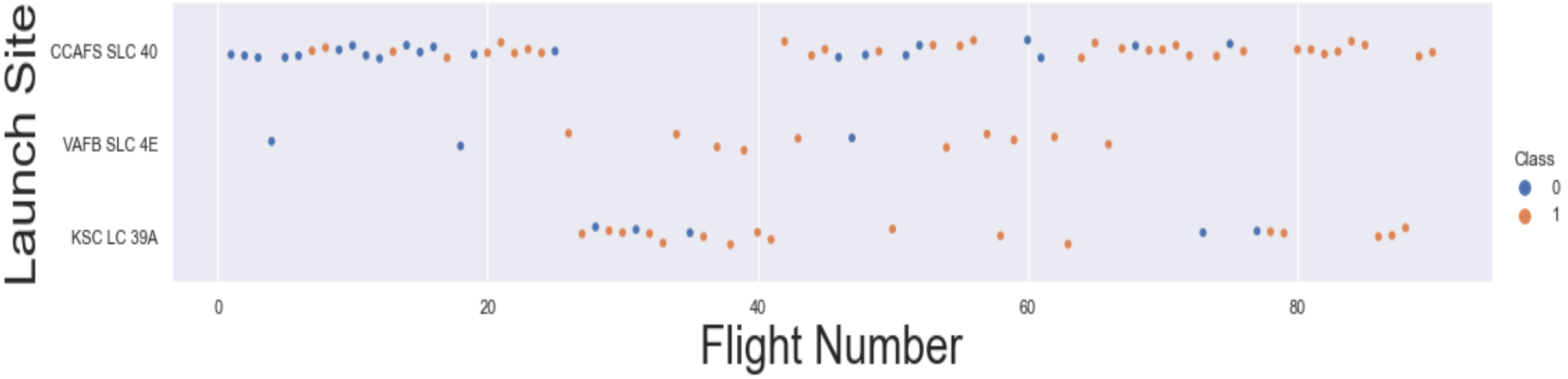
Results (Interactive analytics)

Results (Predictive analysis)

- We've found out that, we can predict the outcome of a launch depending on many attributes, so after using many models to perform that.
- All models got a 83.3% accuracy, and successfully predicted every landed launch but failed 3 class 0 (failure) launches.

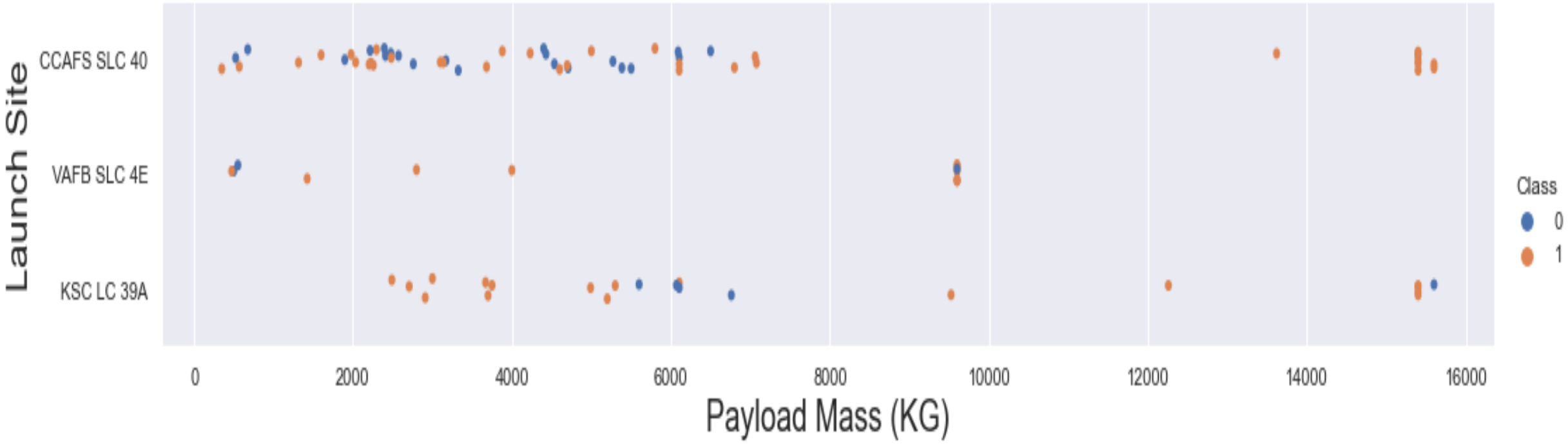
Section 2

Insights drawn from EDA



Flight Number vs. Launch Site

- This is a scatter plot between Flight Number and Launch Sites.
- As we can see in the late flight numbers, the success rate increased, and most of the flights were class 1(succussed).

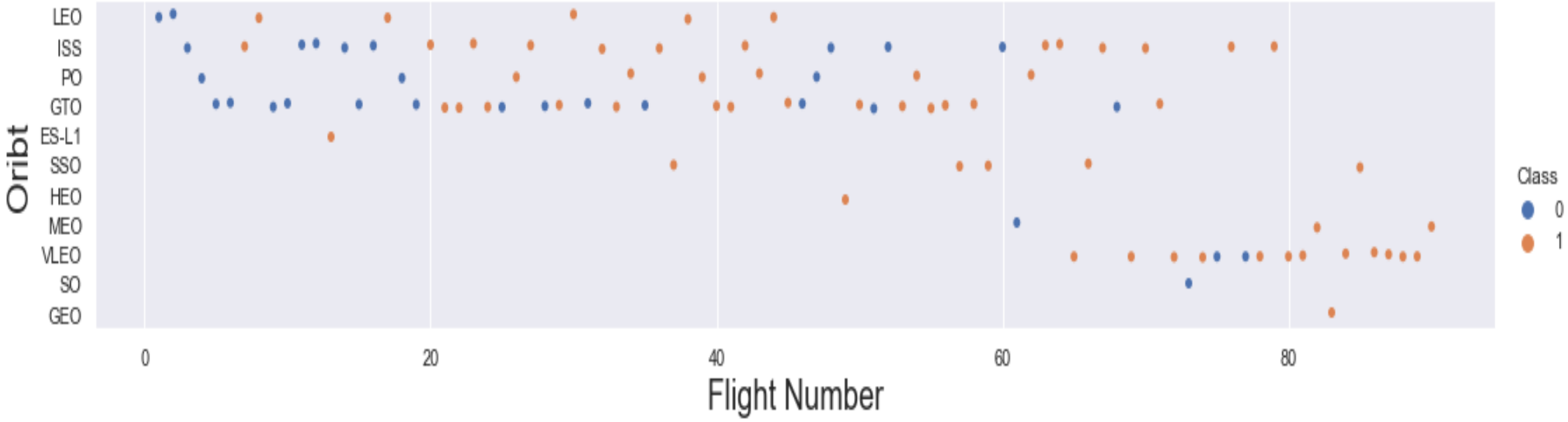


Payload vs. Launch Site

- This is a scatter plot between Payload Mass and Launch Sites.
- As we can see KSC LC 39A has a high success rate for light weights.
- Other sites have a better success rate for heavy weights.

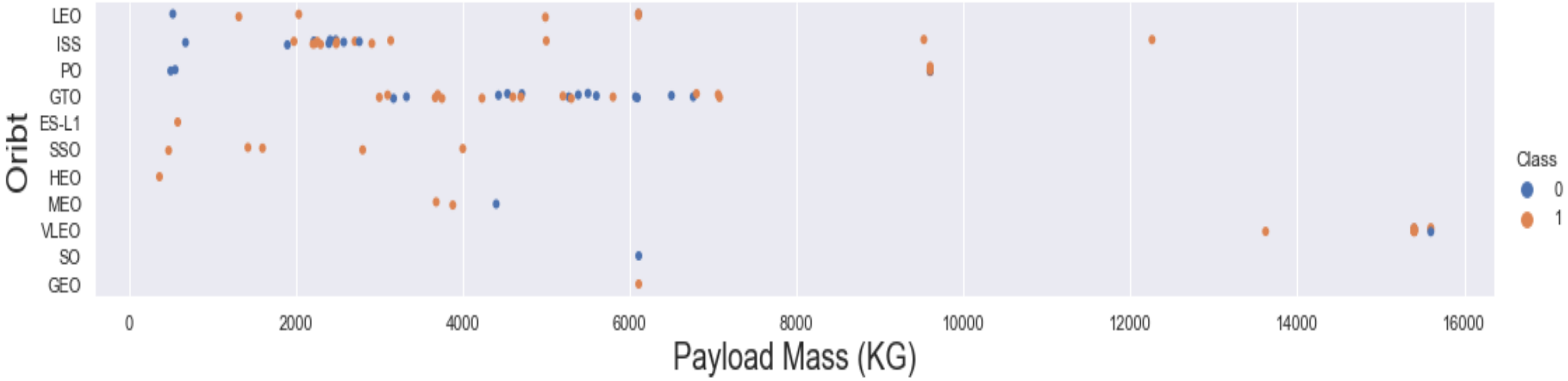
Success Rate per Orbit

- This is a bar plot to show the success rate per Orbit.
- As we can see LEO,ISS,GTO,SO, have a perfect success rate.
- On the other hand, PO have a 50% success rate, while VLEO has never got a successful Launch.



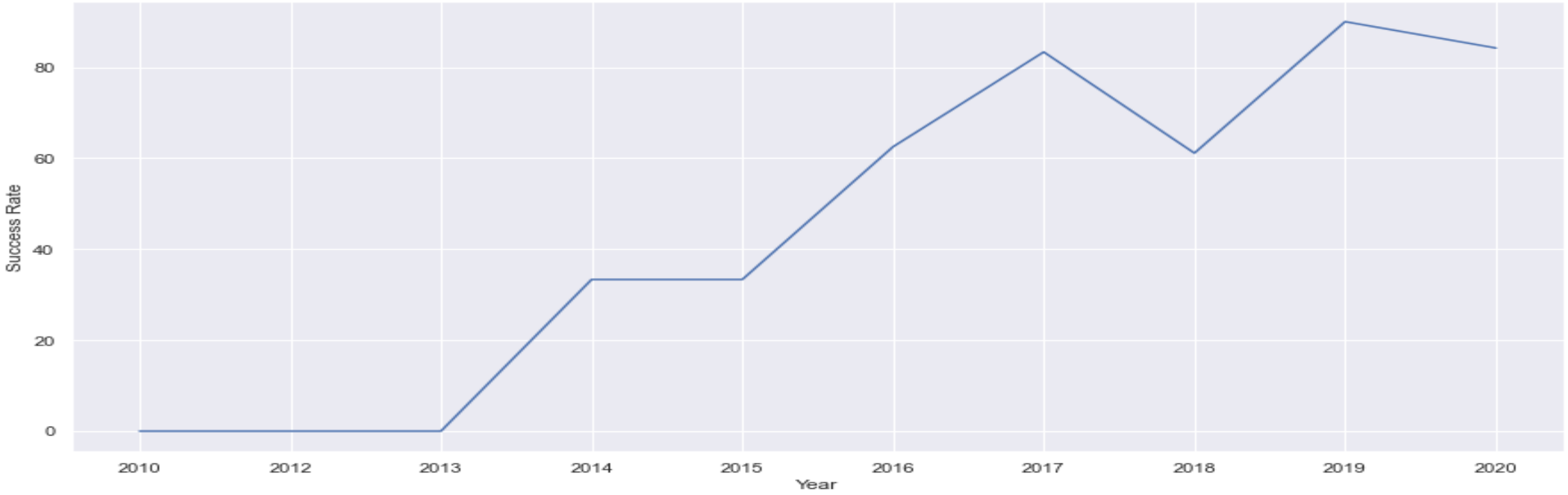
Flight Number vs. Orbit Type

- This is a scatter plot between Flight Number and Orbit.
- As we can see on 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

- This is a scatter plot between Payload Mass and Orbit.
- As we can see 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

- This is a line plot to show the Yearly trend of Launch Success Rate.
- We can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

- Find the names of the unique launch sites
- Using SQL, we've queried for the distinct Launch Sites.

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

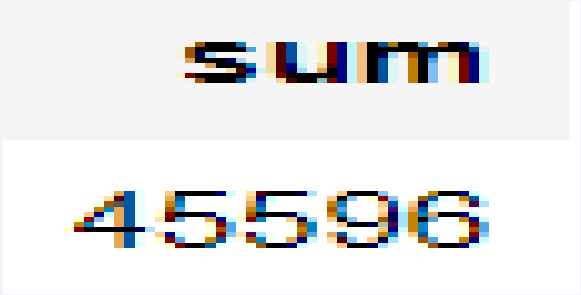
Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Using SQL LIKE and LIMIT We've found 5 the records.

index	date	time(utc)	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing_outcome
0	04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	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)
2	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
3	08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	01-03-2013	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
- Using sum Function of SQL and WHERE clause. we've found that the total payload carried by boosters from NASA is 45596 (KG).



Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Using Avg function of SQL and Where Clause with LIKE.
- The Average Payload Mass by F9 v1.1 is : 2534.67 (KG)

avg_payload
2534.6666666666666667

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Using min function of SQL, We've found out that The first successful Ground Ladning was in 1/5/2017.

first_landing
01-05-2017

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
- Using WHERE Clause and BETWEEN from SQL.

customer
SES
SES EchoStar
SKY Perfect JSAT Group

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Using count function of SQL and Group BY.

mission_outcome	count
Success	1
Success (payload status unclear)	1
Success	98
Failure (in flight)	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Using max function from SQL in a sub query.

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Using WHERE Clause from SQL.

booster_version	launch_site	landing_outcome
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)
F9 v1.1 B1017	VAFB SLC-4E	Failure (drone ship)
F9 FT B1024	CCAFS LC-40	Failure (drone ship)

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
- Using count function and GROUP BY and ORDER BY.

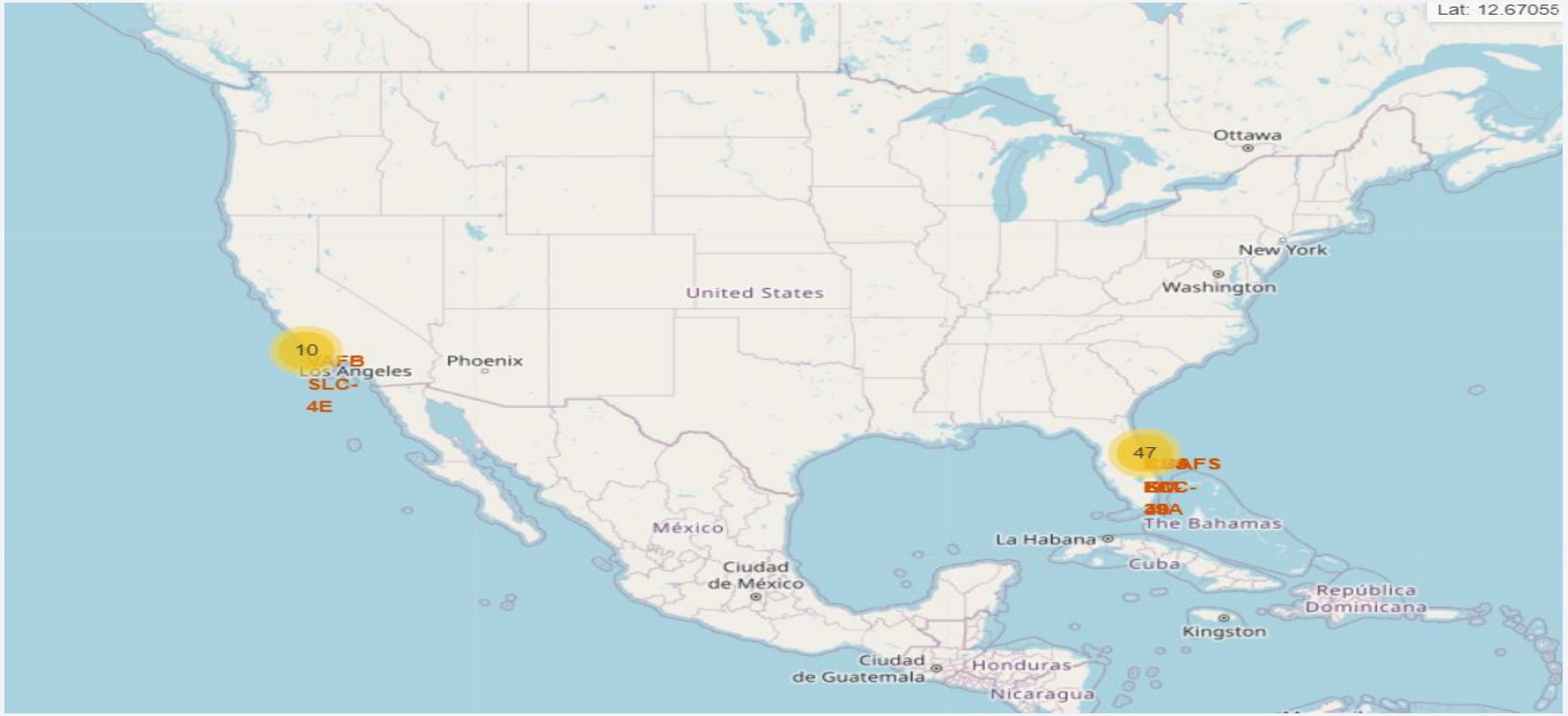
landing_outcome	count
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

Section 3

Launch Sites Proximities Analysis

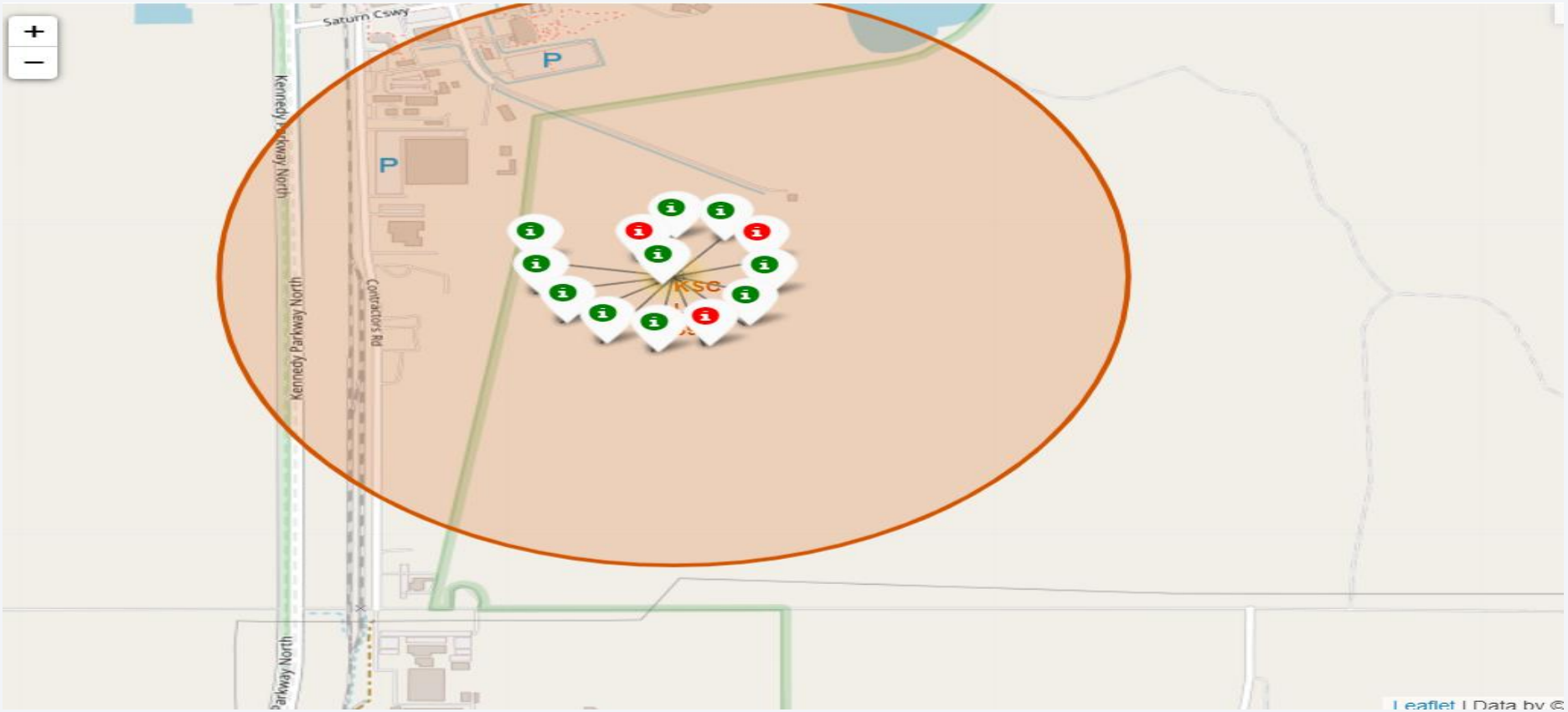
Launch Sites On Map

Those red marks are our Launch Sites, as we can see All of the sites are near the coasts.



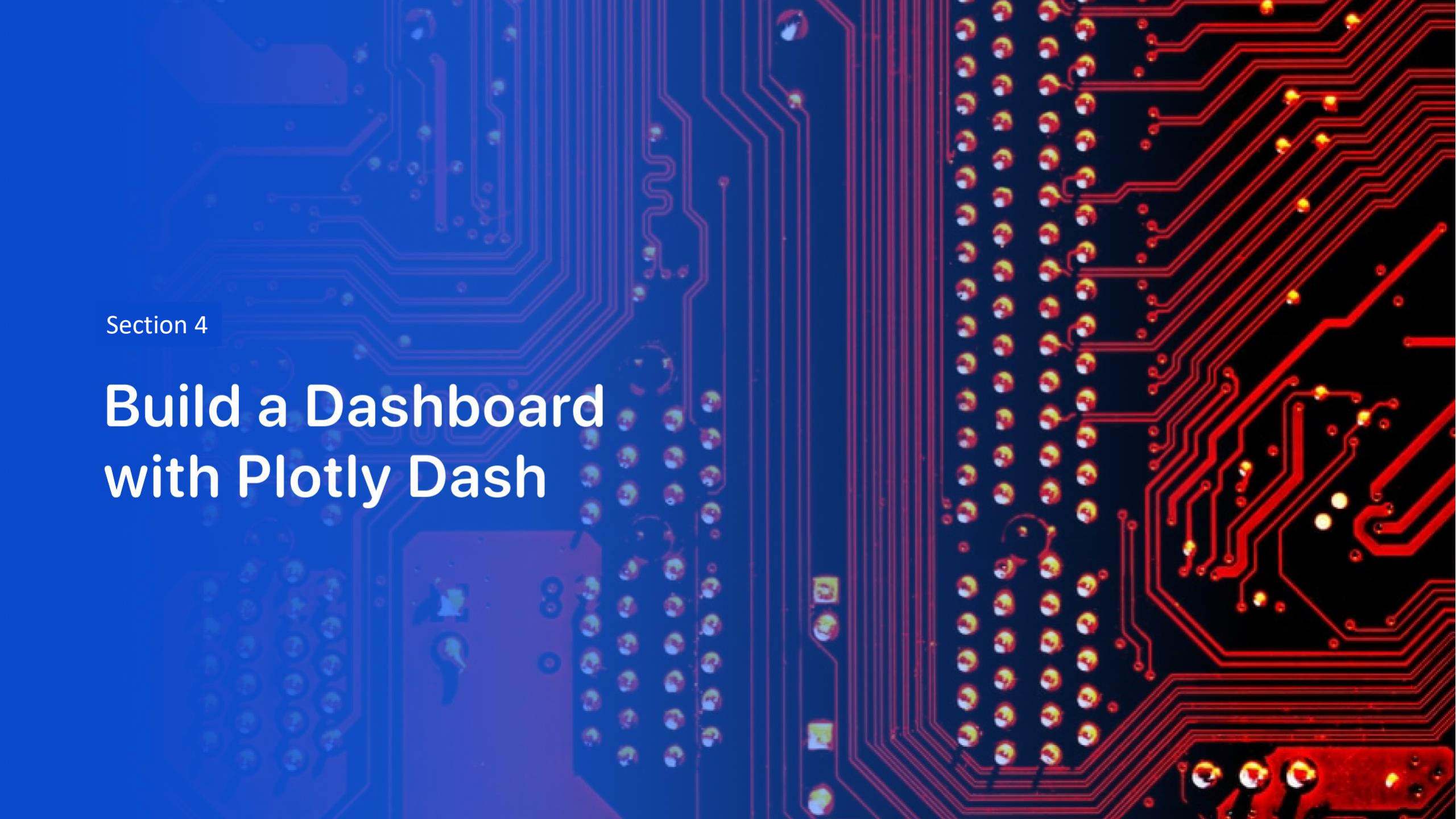
Launch Outcomes on Map

Red marks means failure launches, and green means Successful ones.



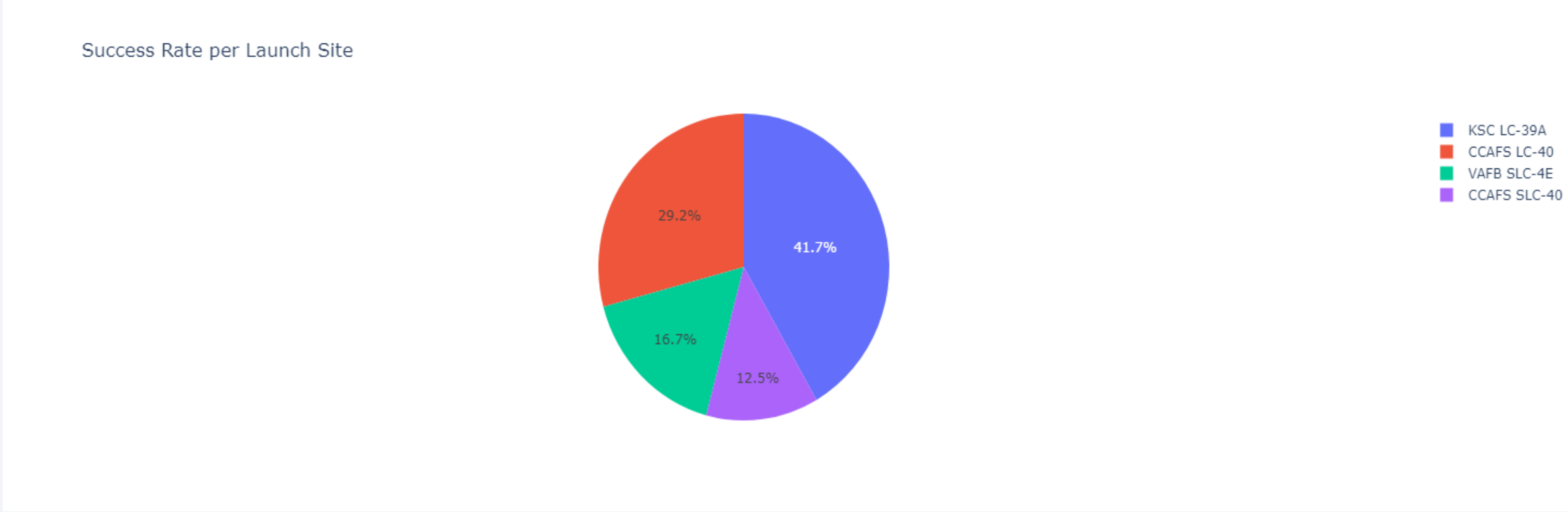
Launch Site proximity

The distance between CCAFS SLC-40 launch site and nearest coastal line is only 0.85 KM


Section 4
Build a Dashboard with Plotly Dash

Pie Chart for All Sites

KSC LC has the highest successful launches



Section 5

Predictive Analysis (Classification)

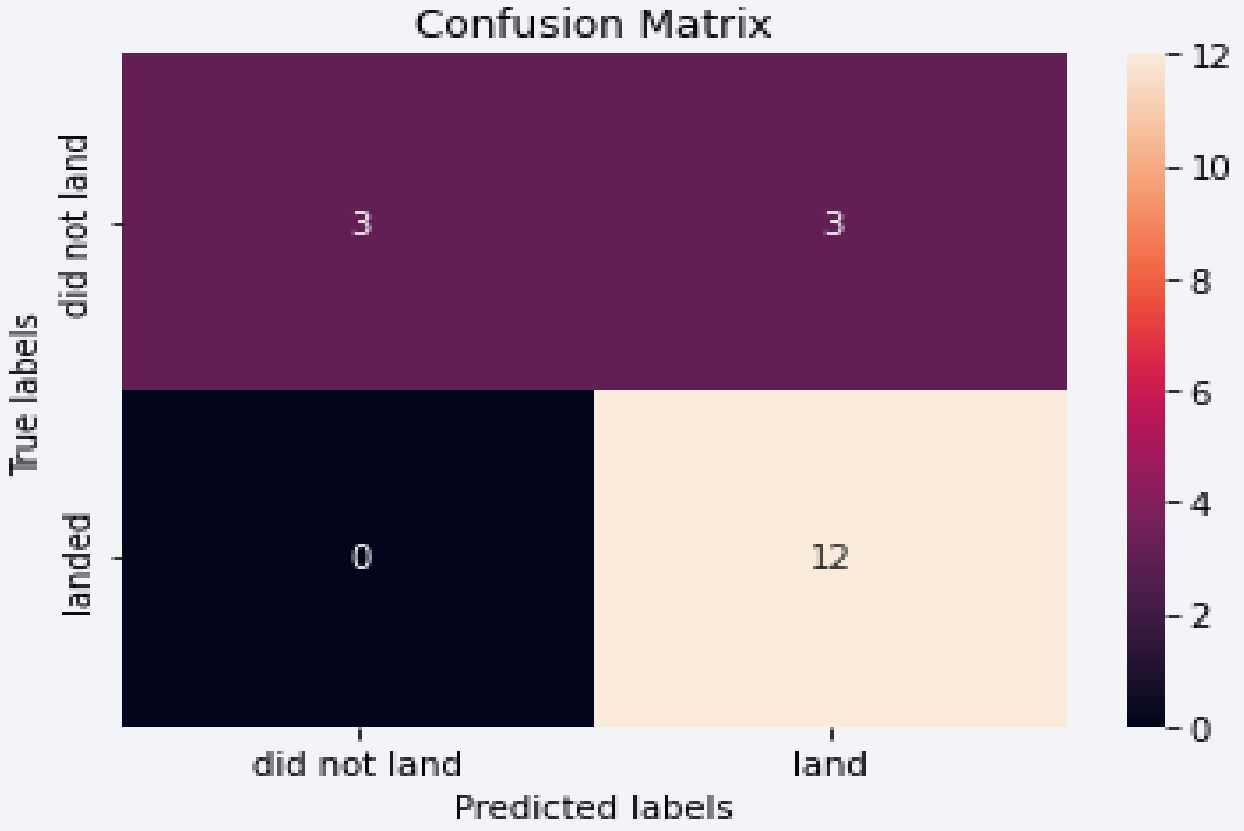
Classification Accuracy

All Models performed the same!

Confusion Matrix

All Successfully landed Launches were predicted successfully.

Only 50% Accuracy on the Failure landings.



Conclusions

- It was obvious that through the time and years, the success rates have increased rapidly.
- The test set was kind of small that’s why we got only 50% Accuracy on Failure landings.
- The next launch of Falcon 9 have a really high chance to success!
- We can use Any of the provided models to predict the result of any future launch.

Appendix

To check all of the files and notebooks

Refer to this repo : https://github.com/mohamedBalkhi/AI_Capstone

Thank you!