Winning Space Race with Data Science

Gerson Hauwanga
12.09.2023

Outline

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

Executive Summary

Using data science mechanisms to predicting the success rate of SpaceX Falcon 9 rockets

Summary of methodologies

- Data collection from different data sources
- Data processing and cleaning
- Exploring Data Analysis with data visualization
- Exploratory Data Analysis with SQL
- Using Folium to build interactive map
- Creating Dashboard using Plotly Dash
- Predictive analysis using Classification algorithm

Summary of all results

- Analysis of results
- Presentation of interactive data analysis results
- Predictive analysis result data insights

Introduction

Project background and context

The mail goal for this project is to employ data science mechanisms and methodologies such as data gathering, processing, and virtualization with the help of different machine learning algorithms and apply it to a really world project SpaceX Falcon 9 in order to get insights from data to be able to predict its first stage launchers success rate and thus reusal since SpaceX reuse their lauchers. The results of this analysis would help in determining cost for individual launches.

Problems you want to find answers

Determine whether first stage launchers of Falcon 9 will be successful.
Determine the how different variables impact the success rate.
Determine if success rate increase with time.
Determine whether launch sites and success rates.

Section 1

Methodology

Methodology

Data collection methodology:

Using SpaceX API and Web Scraping SpaceX Falcon 9 launch records data from Wikipedia pages

Perform data wrangling

Cleaning and structuring data in order to deal with missing data

Filter data according and eliminate all unnecessary data to suit the project needs

Perform exploratory data analysis (EDA) using visualization and SQL

Perform interactive visual analytics using Folium and Plotly Dash

Perform predictive analysis using classification models

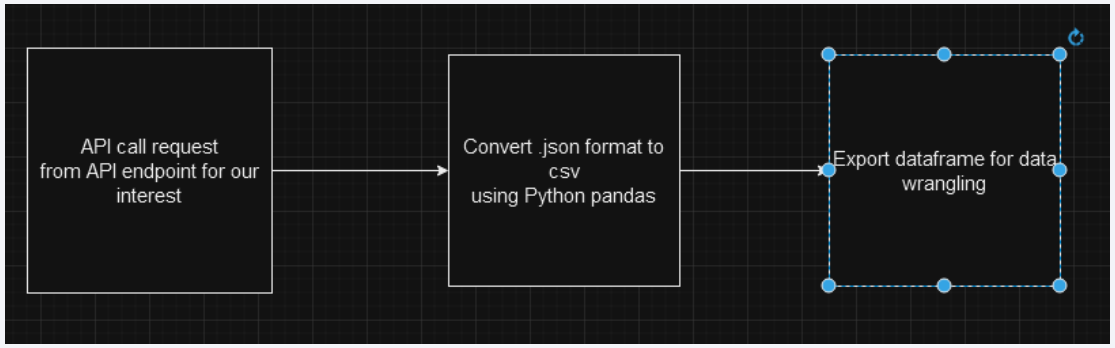
Build, tune, evaluate classification models for better results

Data Collection

Data was collected from SpaceX launch records API end points and Web scraping falcon 9 records.

The data results are in json format and need to be converted into CSV before using them in the project.

The process of SpaceX data collection

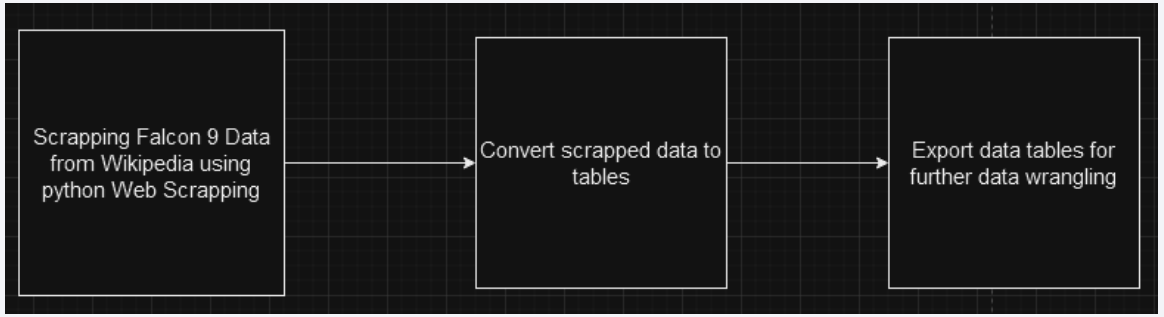


API calls on SpaceX falcon 9 data end points

NoteBook

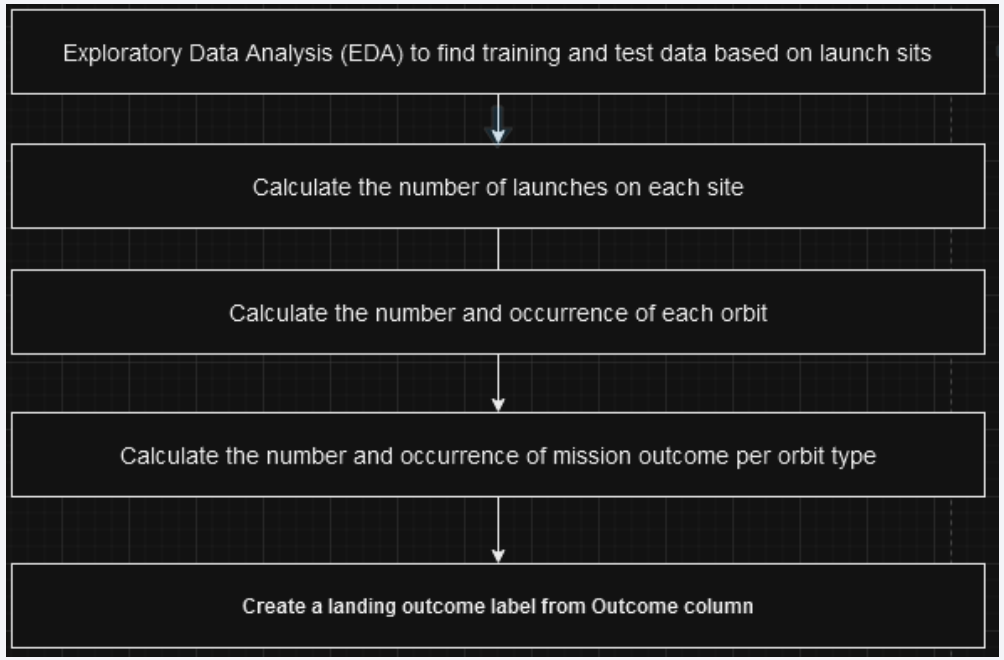
Wikipedia web Scraping

NoteBook



Data Wrangling

At first the data from the API and web scraping are always in raw form. Contain too much noise such missing columns, duplicates and wrongly labeled column data types therefore, the data need to be data need to be cleaned and structured and replace missing data. In that process, the tuned data would then be divided into 2 categories as training and test data.



API calls on SpaceX
falcon 9 data end points

NoteBook

EDA with Data Visualization

Using pandas and Matplotlib, I was able to plot chart and get insight on the data and find success.

Scatter plots based to show how Flight Number vs. Payload Mass,Flight Number vs. Launch Site, Orbit vs. Payload Mass, and Payload Mass vs. Lauch Site variables determine the success rate of the successful landing.

Bar Chart for determining the success rate of landing by comparing Class and Orbit.

Line Chart to indicate a trend of success landings in a year.

NoteBook

EDA with SQL

SQL data manipulation was carried out to

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 sub query
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.

NoteBook

Build an Interactive Map with Folium

In this task, the goal was to mark launch sites on the the interactive map and indicate the success/failed launches using folium python package markers, circles and lines.

The main tasks where to:

Mark all launch sites on a map

Mark the success/failed launches for each site on the map

Calculate the distances between a launch site to its proximity

NoteBook

Build a Dashboard with Plotly Dash

Using Plotly Dashboards and dash pi-charts

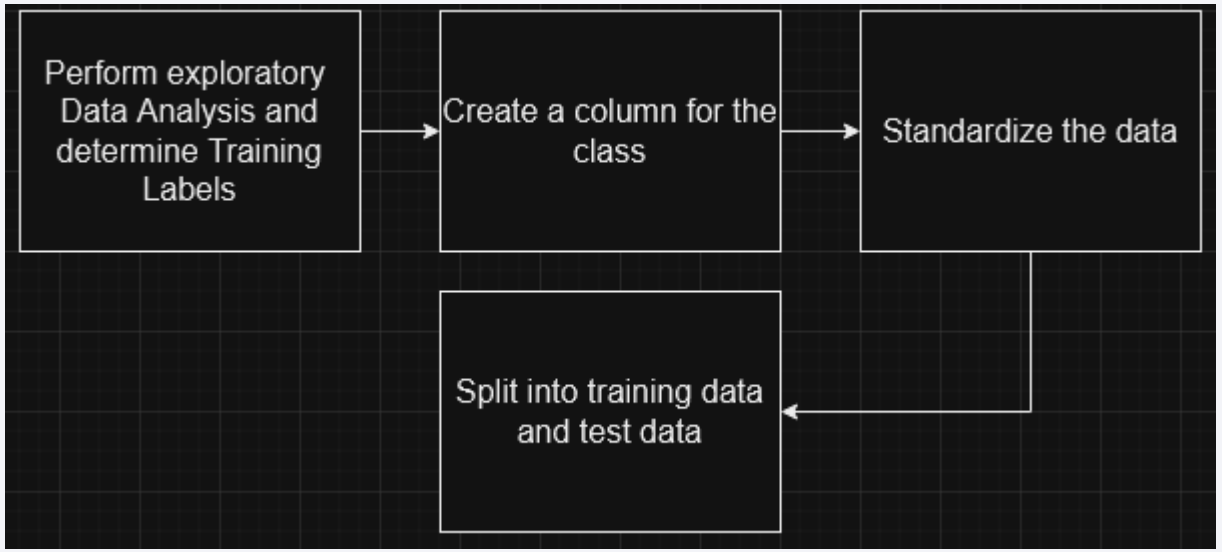
to indicate the success rate for all sites and comparing that to the failed rates.

This was done to show which site has the most success launches.

NoteBook

Predictive Analysis (Classification)

Using data analysis classification model to be able to predict future launch success rates.



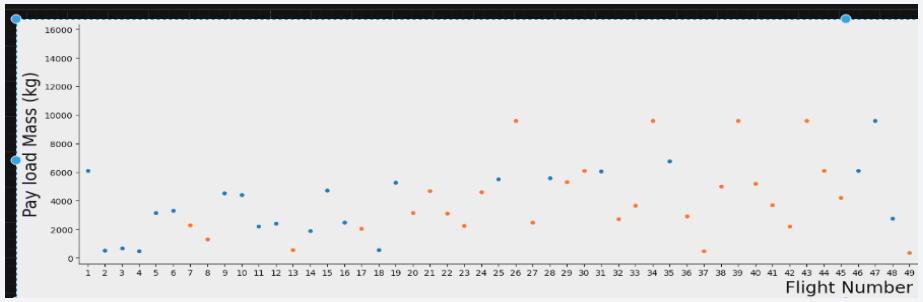
NoteBook

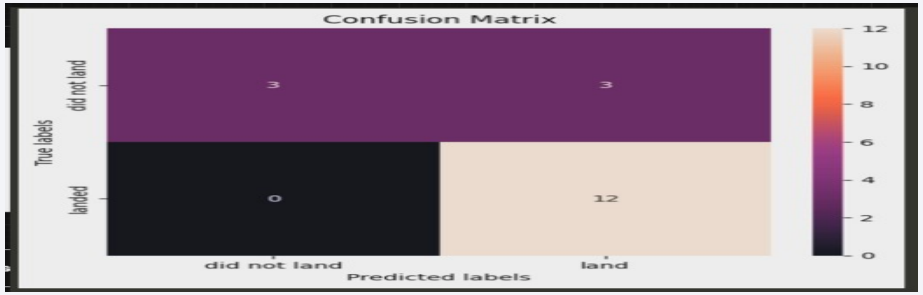
Results

Exploratory data analysis results

Interactive analytics demo in screenshots

Predictive analysis results

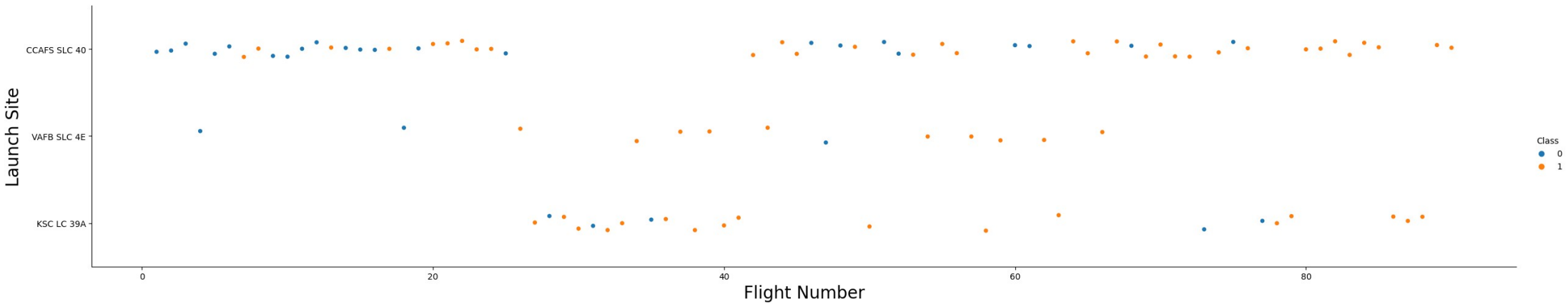




Section 2

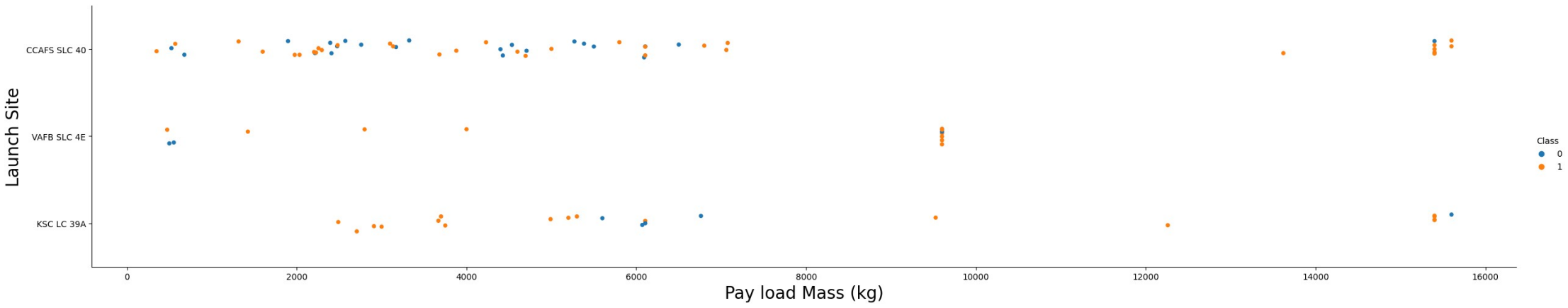
Insights drawn from EDA

Flight Number vs. Launch Site



We see that different flight numbers have different success rates per each launch sites. CCAFS LC-40, has better success flight, while KSC-LC-39A and VAFB-SLC-4E are slightly lower

Payload vs. Launch Site

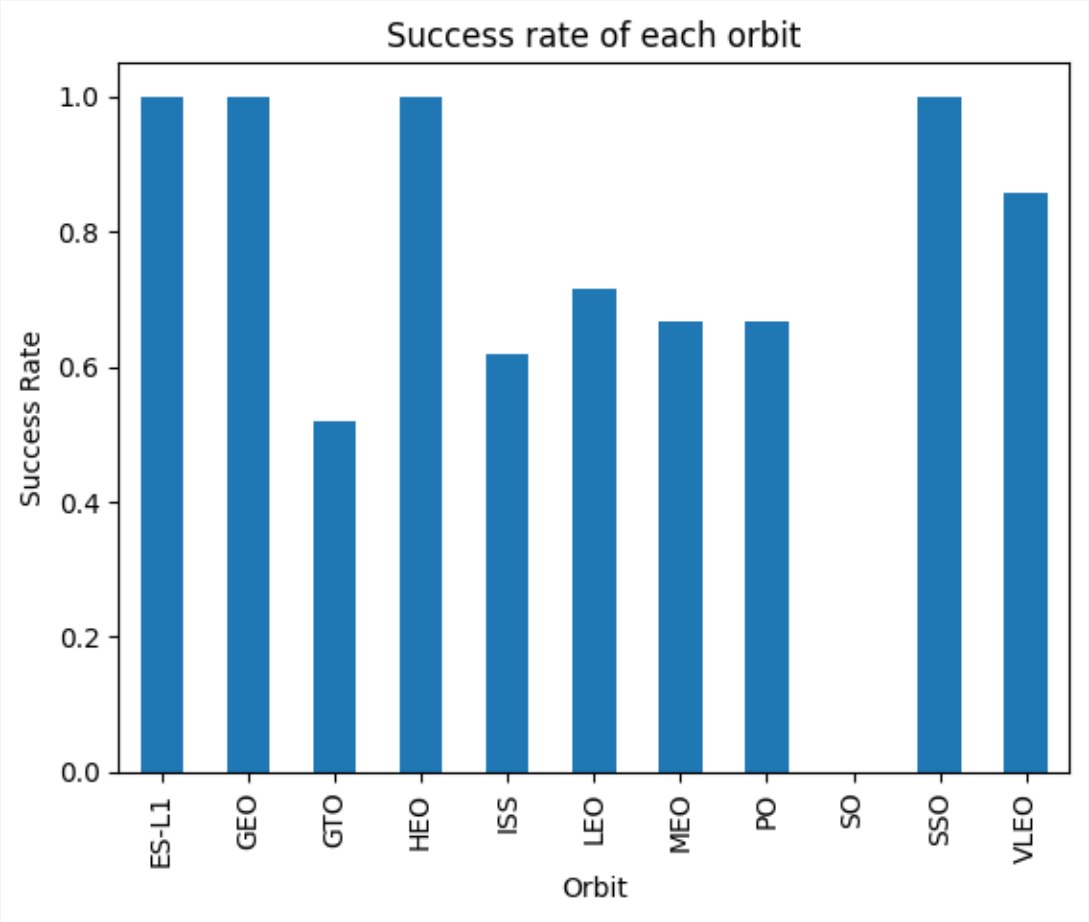


For launch site VAFB-SLC there were no rockets launched for heavy payload mass(greater than 10000)

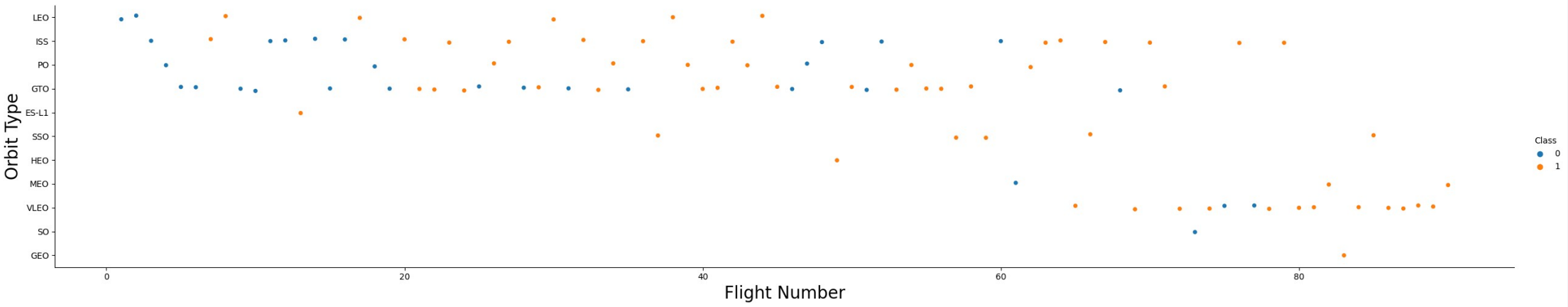
For launch site CCAFS SLC 40, there were a higher number of successful flights of lower payload mass less than ~7000.

Success Rate vs. Orbit Type

Orbits ES-LI, GEOm HEO, and SSO have highest the success flight rate while GTO is lowest.



Flight Number vs. Orbit Type



There was only one success landing for GEO at flights after 80, GTO, ISS, LEO PO had great but random successful landing.

Payload vs. Orbit Type

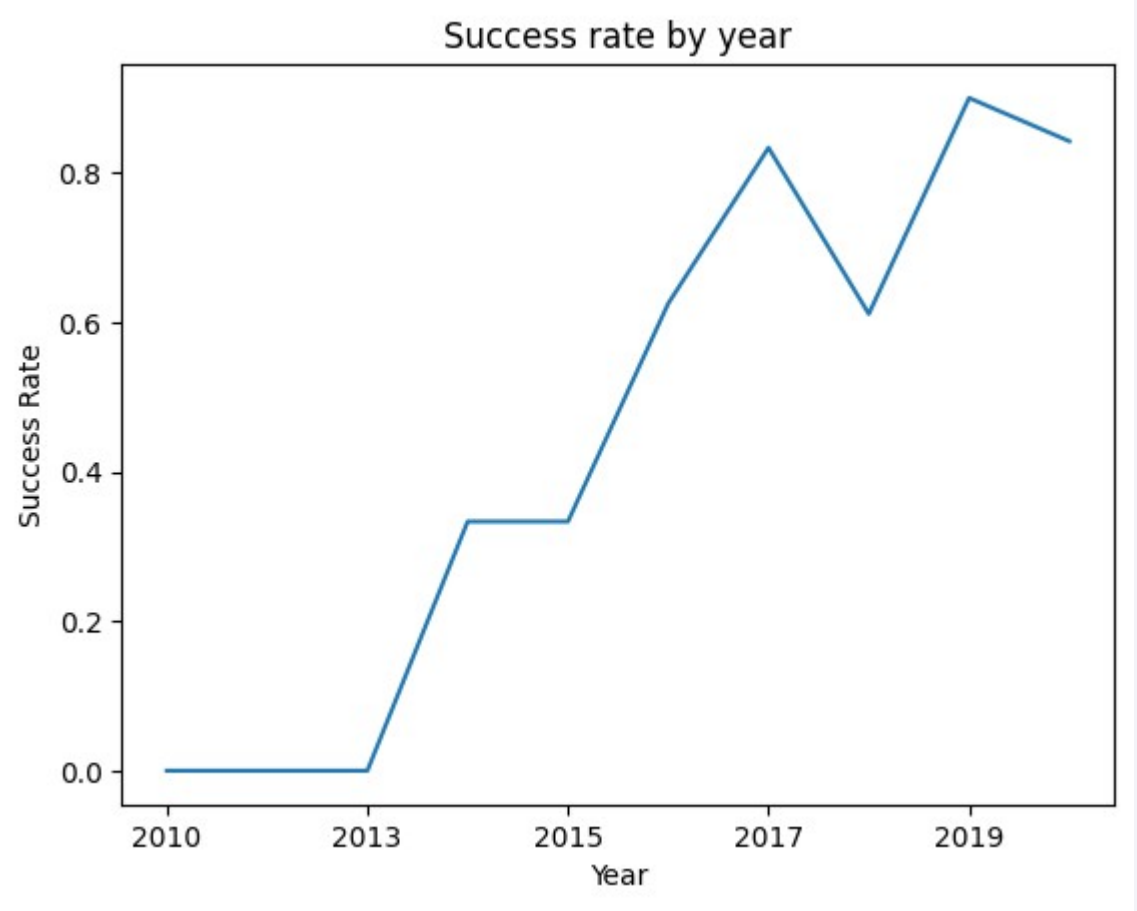
Heavy payload mass had lower success rate for all orbits.

There was no successful landing for weight over 16000.

With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

Launch Success Yearly Trend

The success rate since 2013 kept increasing till 2020. That means success was a result of improvements and repetition



All Launch Site Names

```
%sql select distinct launch_site from spacextable order by 1
```

```
 * sqlite:///my_data1.db
Done.
```

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

Launch Site Names Begin with 'CCA'

```
%sql select * from SPACEXTABLE WHERE spacextable.launch_site like 'CCA%' limit 5
```

Python

```
 * sqlite:///my_data1.db
Done.
```

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)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	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

Total Payload Mass

```
%sql ALTER TABLE spacextable RENAME PAYLOAD_MASS__KG to PAYLOAD_MASS_KG
```

```
%sql select sum (PAYLOAD_MASS__KG) from spacextable where customer = 'NASA (CRS)'
```

```
 * sqlite:///my_data1.db
Done.
```

sum (PAYLOAD_MASS__KG)
45596

Average Payload Mass by F9 v1.1

```
%sql select AVG (PAYLOAD_MASS_KG) from spacextable where Booster_Version = 'F9 v1.1'
```

```
 * sqlite:///my_data1.db
Done.
```

AVG (PAYLOAD_MASS_KG)
2928.4

First Successful Ground Landing Date

```
%sql select MIN (Date) from SPACEXTABLE where Mission_outcome LIKE 'Success%'
```

```
 * sqlite:///my_data1.db
Done.
```

MIN(Date)
2010-04-06

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT Booster_Version from spacextable where Landing_Outcome like 'Success (drone ship)' AND PAYLOAD_MASS_KG BETWEEN 4000 AND 6000
```

Python

```
 * sqlite:///my_data1.db
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

```
%sql select mission_outcome, count(mission_outcome) from spacextable where mission_outcome like 'Success%'
```

```
%sql select mission_outcome, count(mission_outcome) from spacextable where mission_outcome like 'Failure%'
```

```
 * sqlite:///my_data1.db
Done.
```

Mission_Outcome	count(mission_outcome)
Failure (in flight)	1

Boosters Carried Maximum Payload

```
%sql select date, Booster_Version, payload_mass_kg,landing_outcome
from spacextable where payload_mass_kg =
(select max(payload_mass_kg) from spacextable)
```

```
 * sqlite:///my_data1.db
Done.
```

Date	Booster_Version	PAYLOAD_MASS_KG	Landing_Outcome
2019-11-11	F9 B5 B1048.4	15600	Success
2020-07-01	F9 B5 B1049.4	15600	Success
2020-01-29	F9 B5 B1051.3	15600	Success
2020-02-17	F9 B5 B1056.4	15600	Failure
2020-03-18	F9 B5 B1048.5	15600	Failure
2020-04-22	F9 B5 B1051.4	15600	Success
2020-04-06	F9 B5 B1049.5	15600	Success
2020-03-09	F9 B5 B1060.2	15600	Success
2020-06-10	F9 B5 B1058.3	15600	Success
2020-10-18	F9 B5 B1051.6	15600	Success
2020-10-24	F9 B5 B1060.3	15600	Success
2020-11-25	F9 B5 B1049.7	15600	Success

2015 Launch Records

```
%sql SELECT substr(Date, 6, 2) AS Month, substr(Date, 0, 5) AS Year,
BOOSTER_VERSION, Landing_Outcome,launch_site FROM SPACEXTABLE
where substr(Date, 0, 5) = '2015' AND Landing_Outcome = 'Failure (drone ship)'
```

```
* sqlite:///my_data1.db
Done.
```

Month	Year	Booster_Version	Landing_Outcome	Launch_Site
10	2015	F9 v1.1 B1012	Failure (drone ship)	CCAFS LC-40
04	2015	F9 v1.1 B1015	Failure (drone ship)	CCAFS LC-40

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

```
%sql select MIN(payload_mass_kg) from spacextable
```

```
 * sqlite:///my_data1.db
Done.
```

MIN(payload_mass_kg)
0

```
%sql select Landing_Outcome, COUNT(*) as qty
from spacextable where substr(Date, 0, 5) || substr(Date, 6, 2) || substr(Date, 8, 2)
between '20100604' and '20170320' AND Landing_Outcome LIKE'%Success%'
GROUP BY Landing_Outcome ORDER BY qty DESC
```

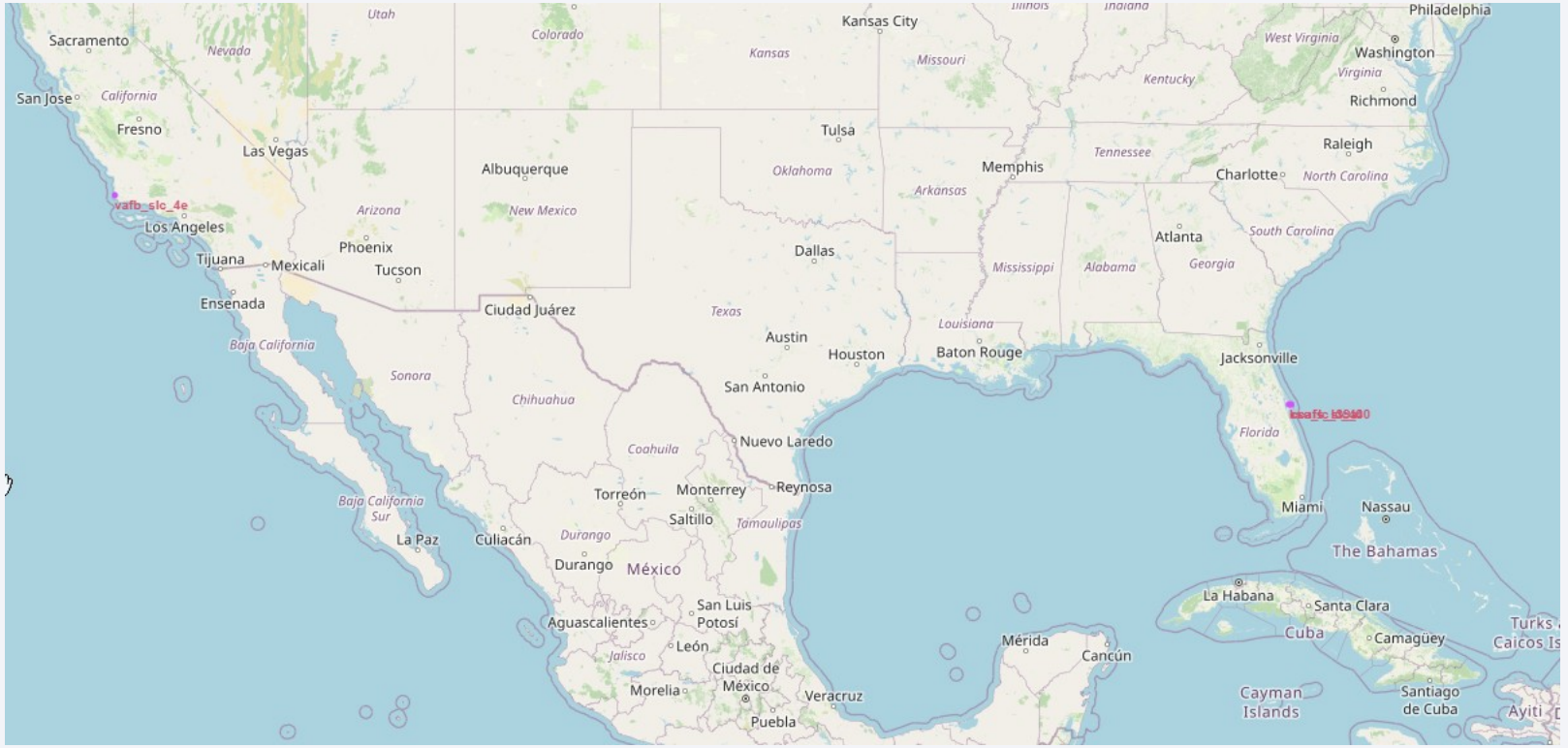
```
 * sqlite:///my_data1.db
Done.
```

Landing_Outcome	qty
Success (drone ship)	6
Success (ground pad)	5

Section 3

Launch Sites Proximities Analysis

Mark all launch sites on a map

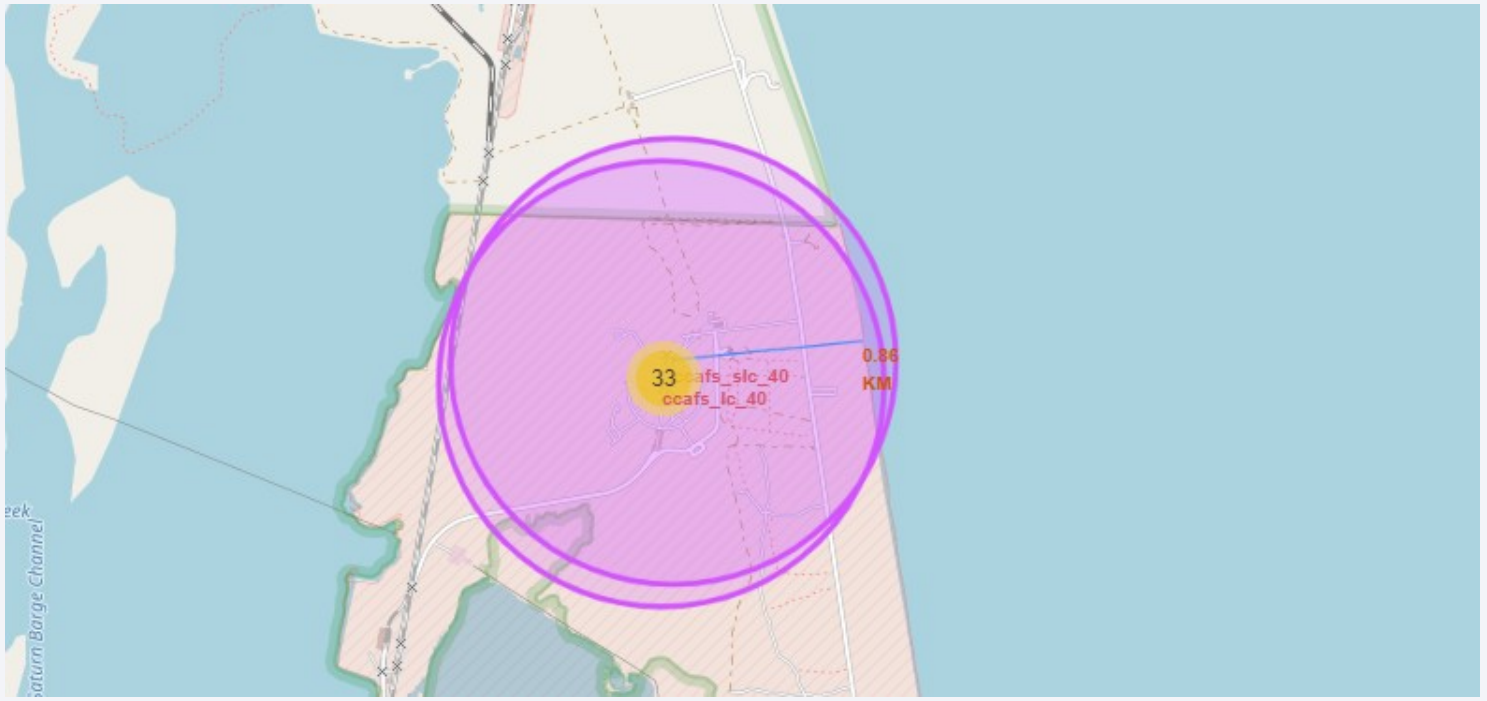


All launch sites were nearby the coastal line

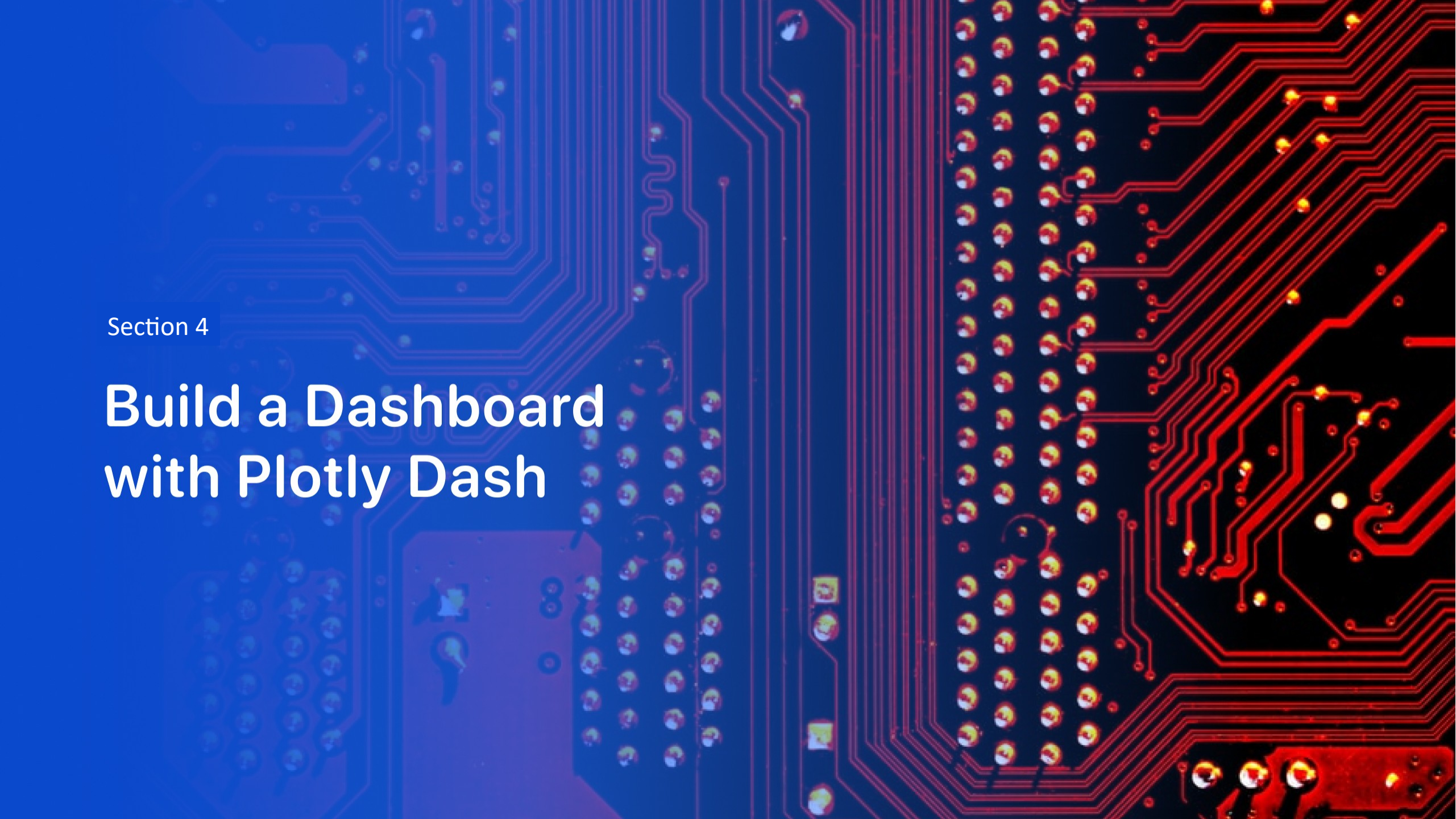
Mark the success/failed launches for each site on the map

East coast launch sites, KSC-LC-39A, CCAFS-LC-40 and CCAFS-SLC-40 had a many failed/success launches compared to the western coast.

Calculate the distances between a launch site to its proximities

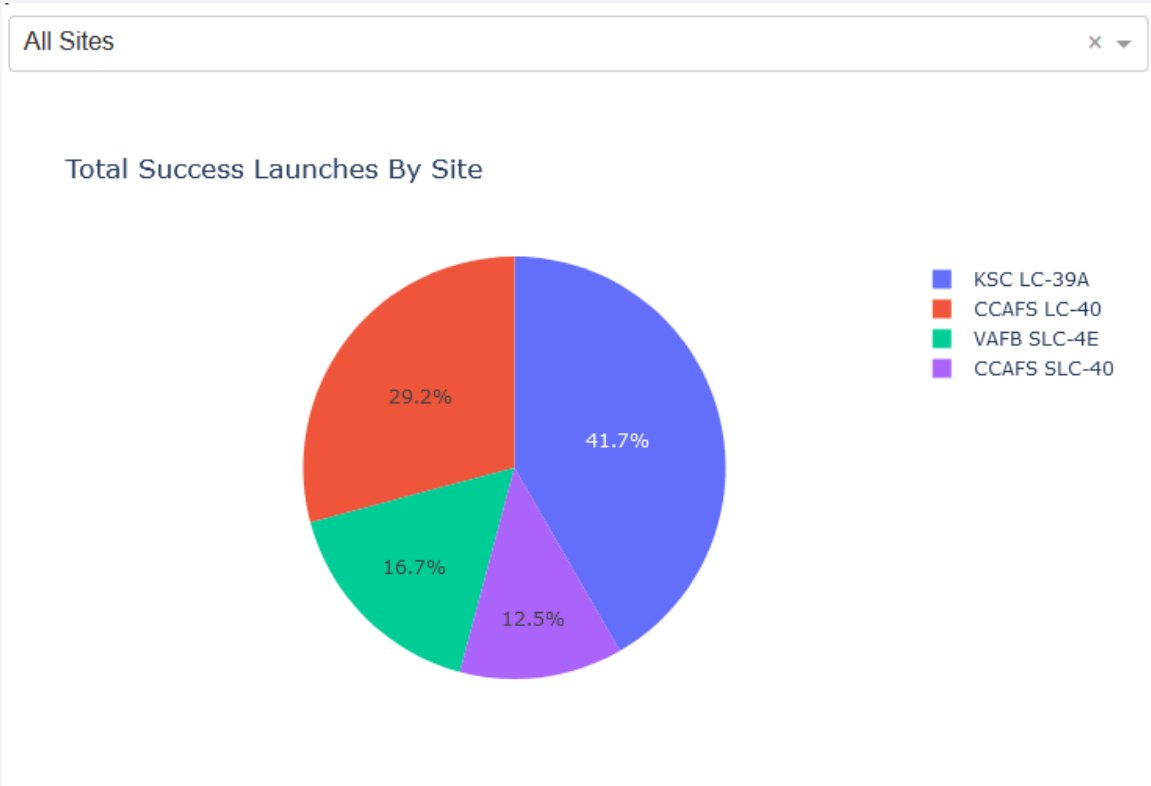


2x East coastal launch sites are in proximity with about 33 point of interest and also nearby the ocean.

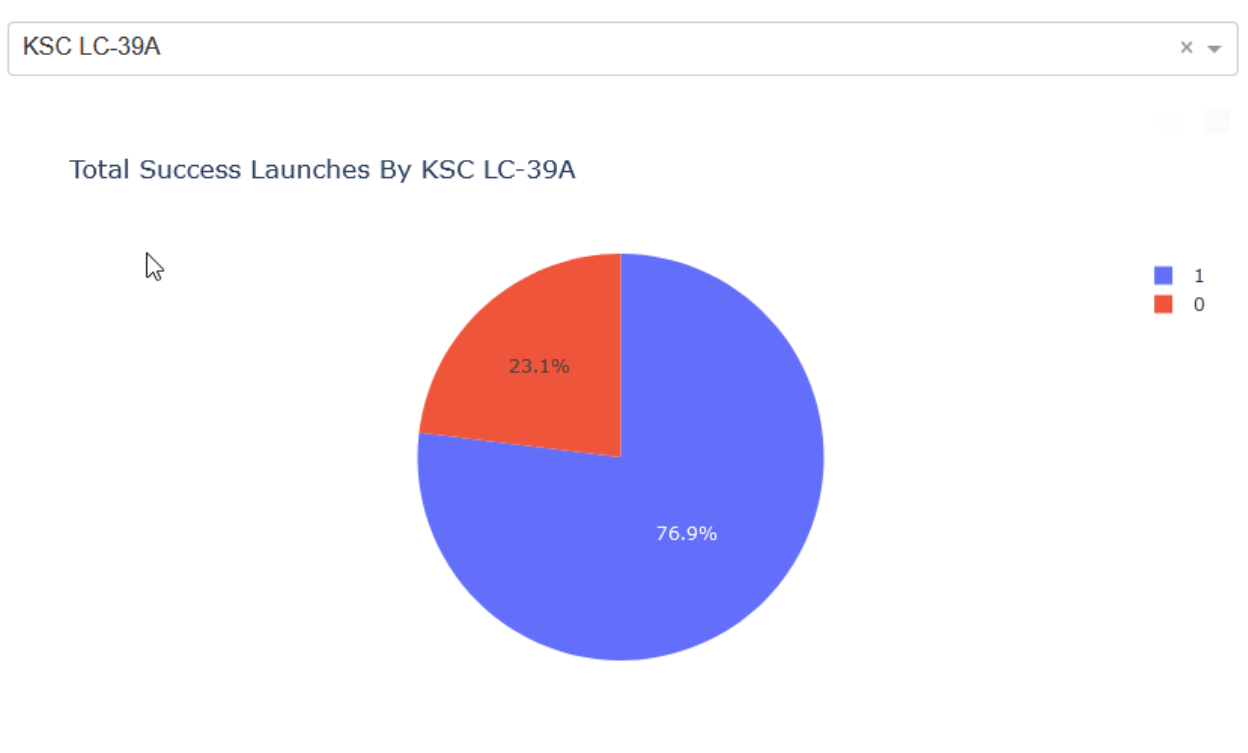

Section 4
Build a Dashboard with Plotly Dash

SpaceX Launch Records Dashboard

KSC LC – 39A had the highest successful launches and CCAFS SLC- 40 had the lowest

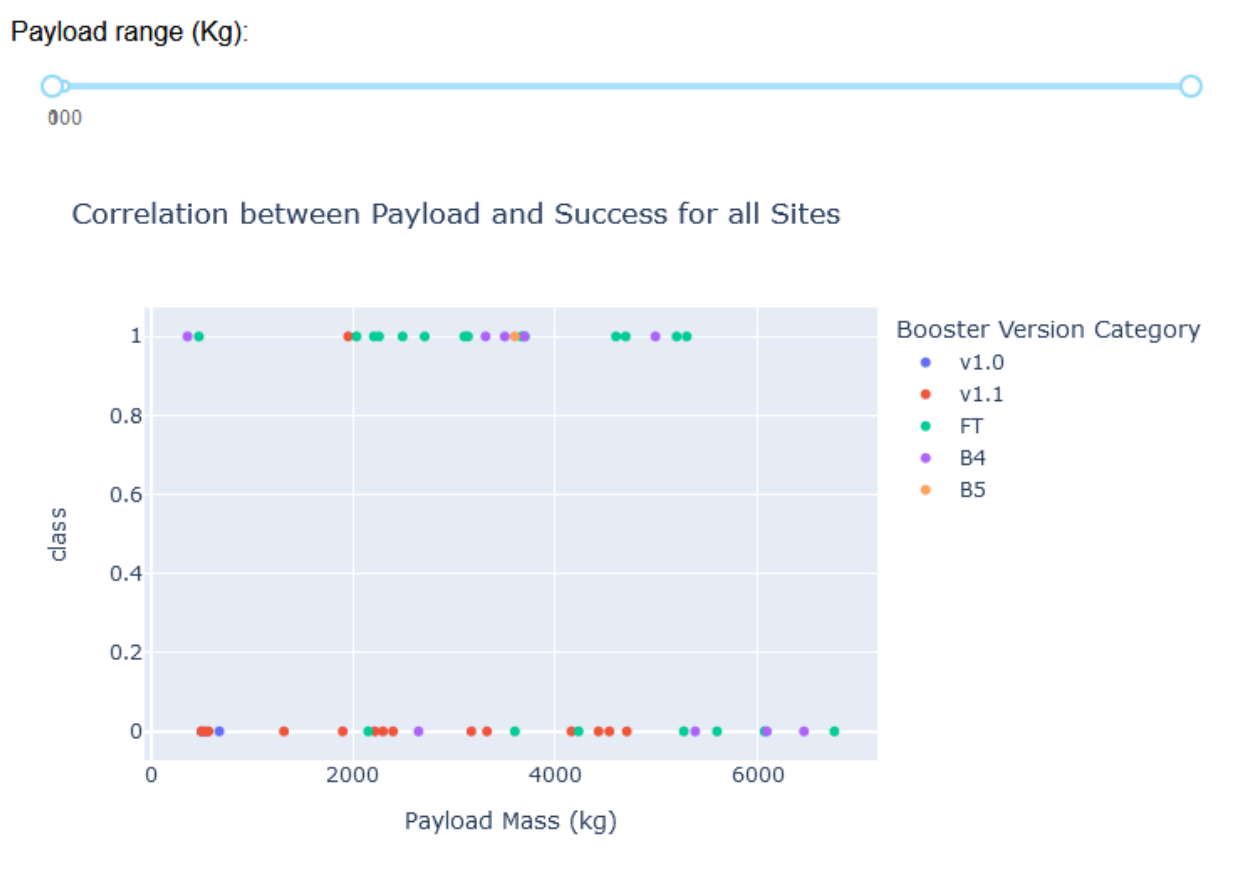


Total success Launches by KSC LC-39A



High success was of class 1 compered to class 0 with the lowest success rate.

Correlation between Payload and Success for all Sites

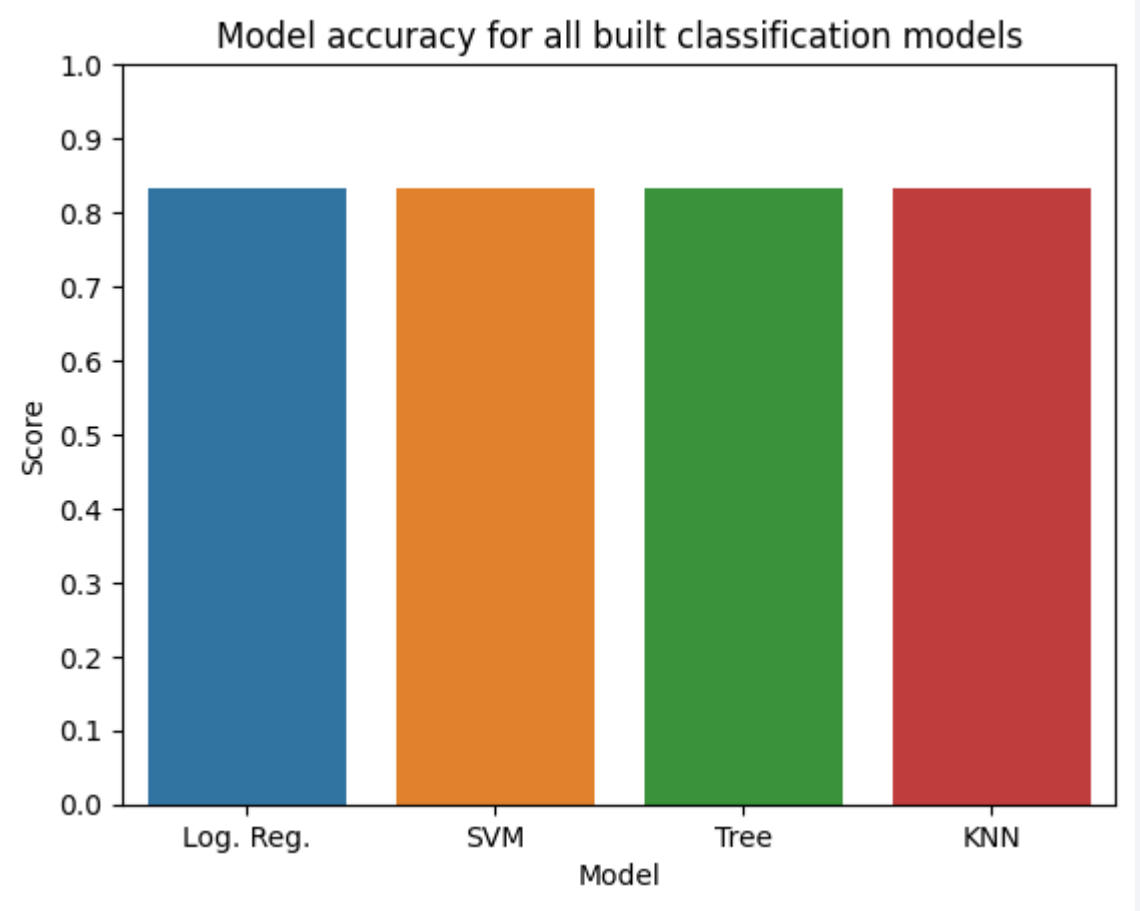


Section 5

Predictive Analysis (Classification)

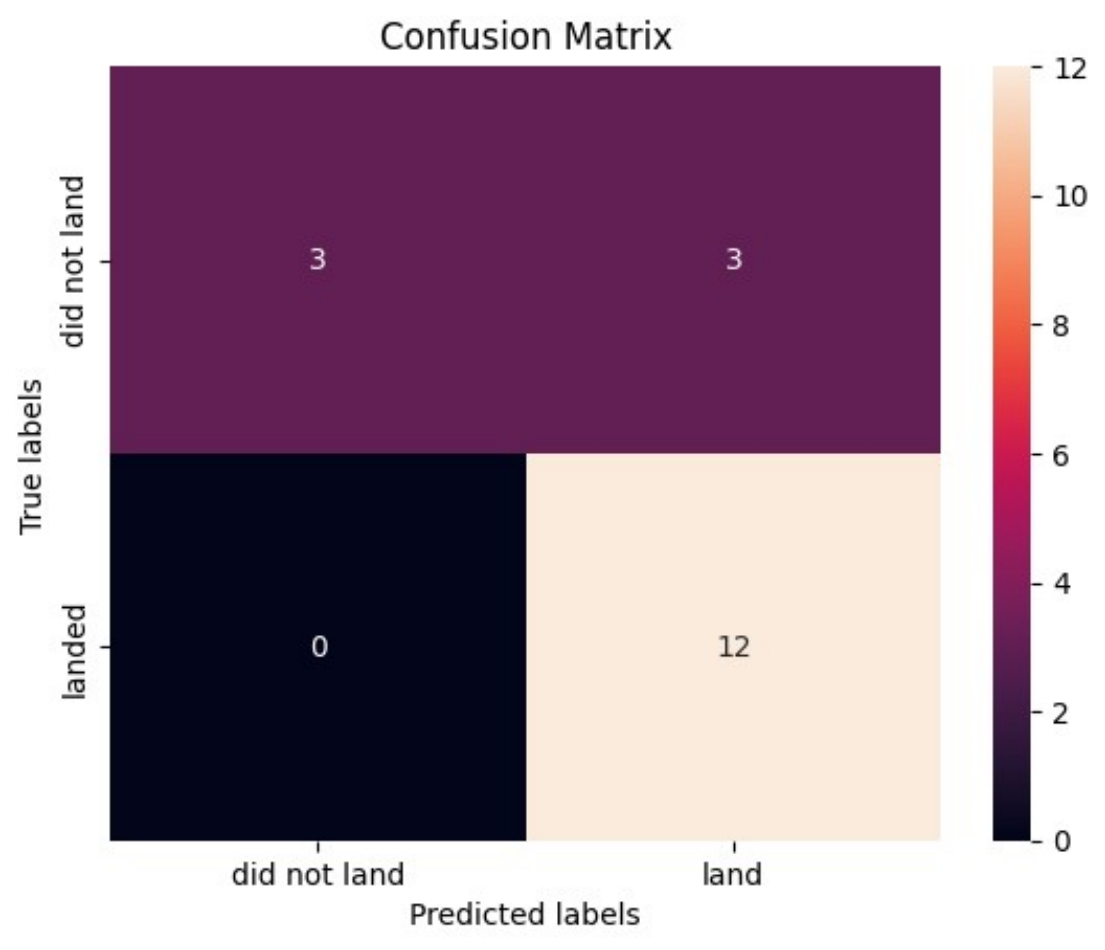
Classification Accuracy

All models have similar accuracy



Confusion Matrix

Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is false positives.



Conclusions

On classification algorthims, the decision tree is a suitable choice for this project.

Based on the results, the lower the payload mass the higher the success rate. We can rule that the lighter the better.

Another point is that, the rate of success seem to increase with time and repetition. As for the launch sites, KSC LC-39A has many success launch/landing compared to others.

Appendix

Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!