Winning Space Race with Data Science

Magno Melendez
5/29/2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data wrangling
  - Exploratory Data analysis using SQL
  - Interactive visual analytics Exploratory Data Analysis
  - Predictive analysis
- Summary of all results
  - Analysis of data
  - Prediction of outcomes
  - Interactive tools

Introduction

- The objective of this study and analysis is to determine if using the SpaceX Falcon 9 rockets will be a cost saving. When compared to other rocket manufacturers.
- The analysis needs to determine The Falcon 9 success rate in association with payload mass

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using the following methods:
    - SpaceX Api
    - Web Scraping
- Perform data wrangling
  - Data was normalized. Adding indicators to help determine success rates.
- Perform exploratory data analysis (EDA) using visualization and SQL

Methodology

Executive Summary

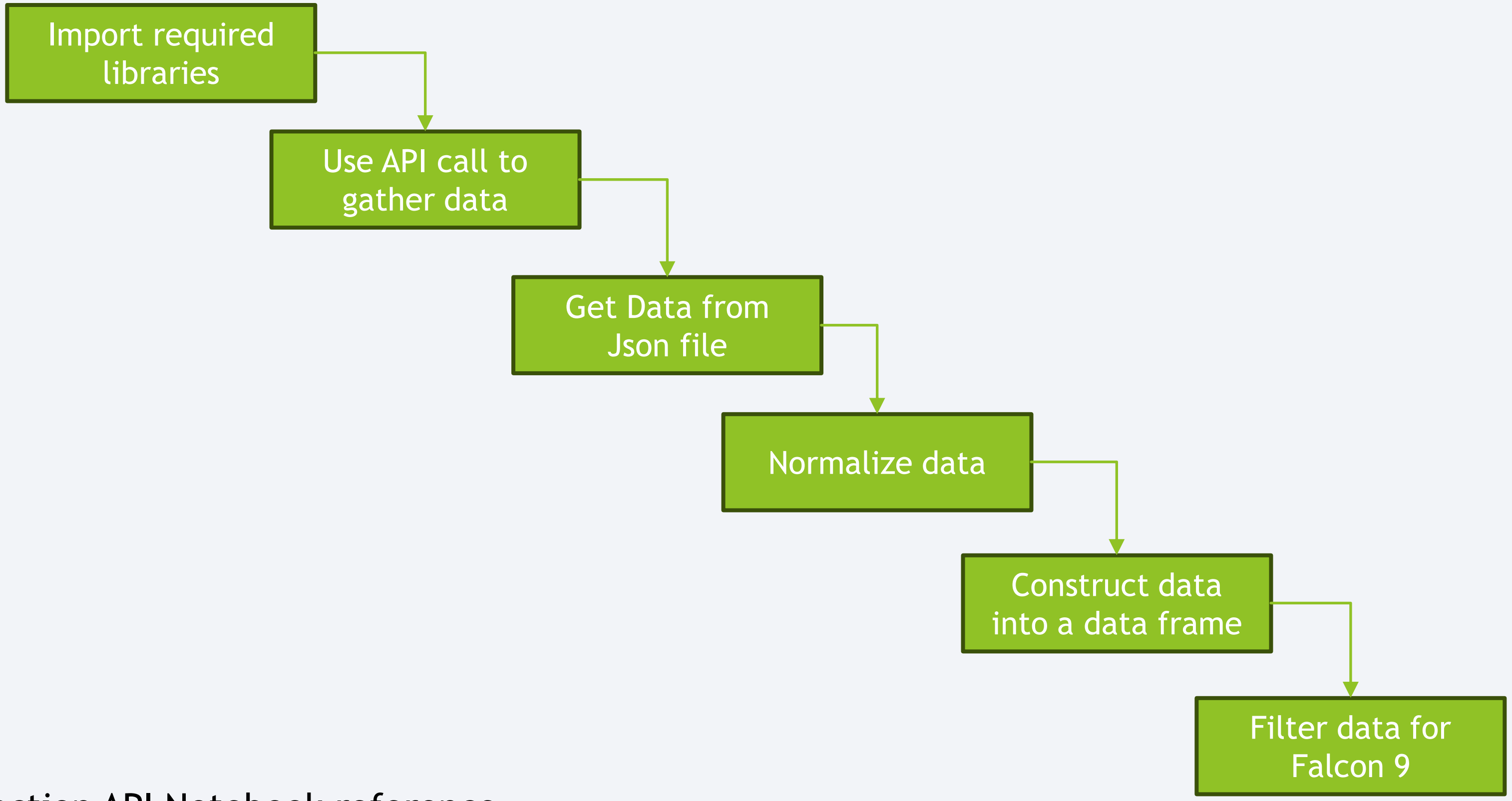
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data was normalized
  - Train and test sample created
  - Evaluation of data was made different class models

Data Collection

- Data was collected using the SpaceX REST API. Data stored in Json file.
- Web scraping was also used for additional Falcon 9 data.

Data Collection – SpaceX API

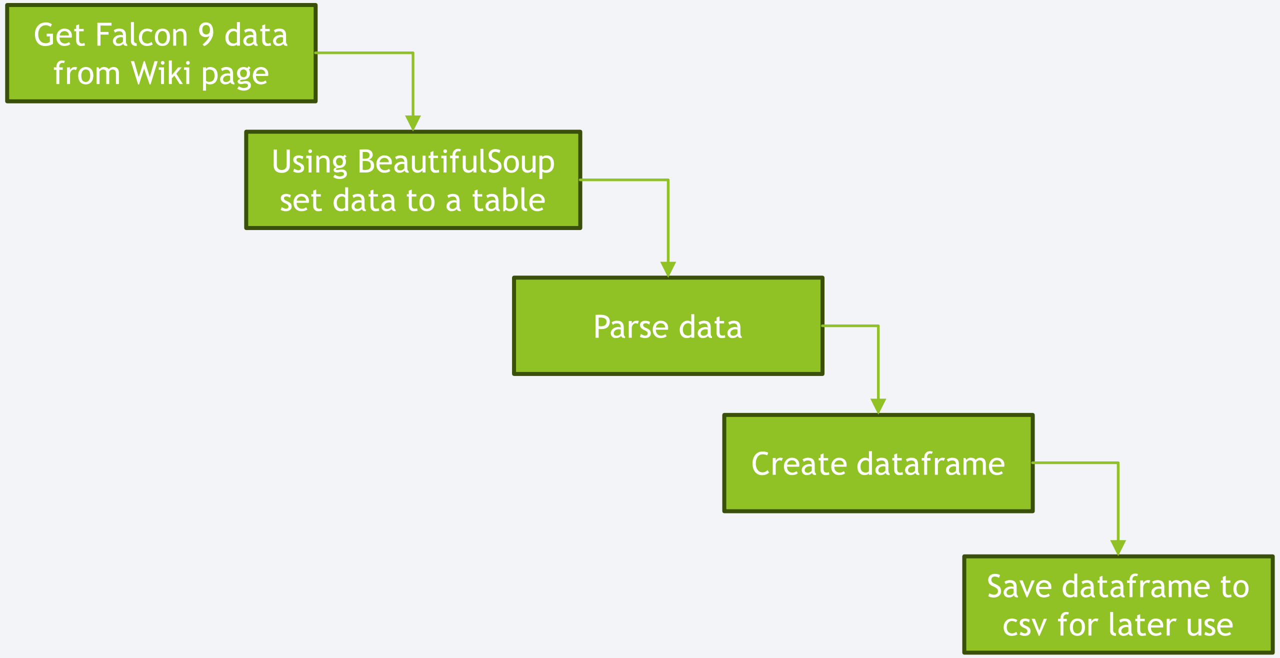
Process flow



Link to collection API Notebook reference
https://github.com/MagMel8860/Mags_Rep_Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

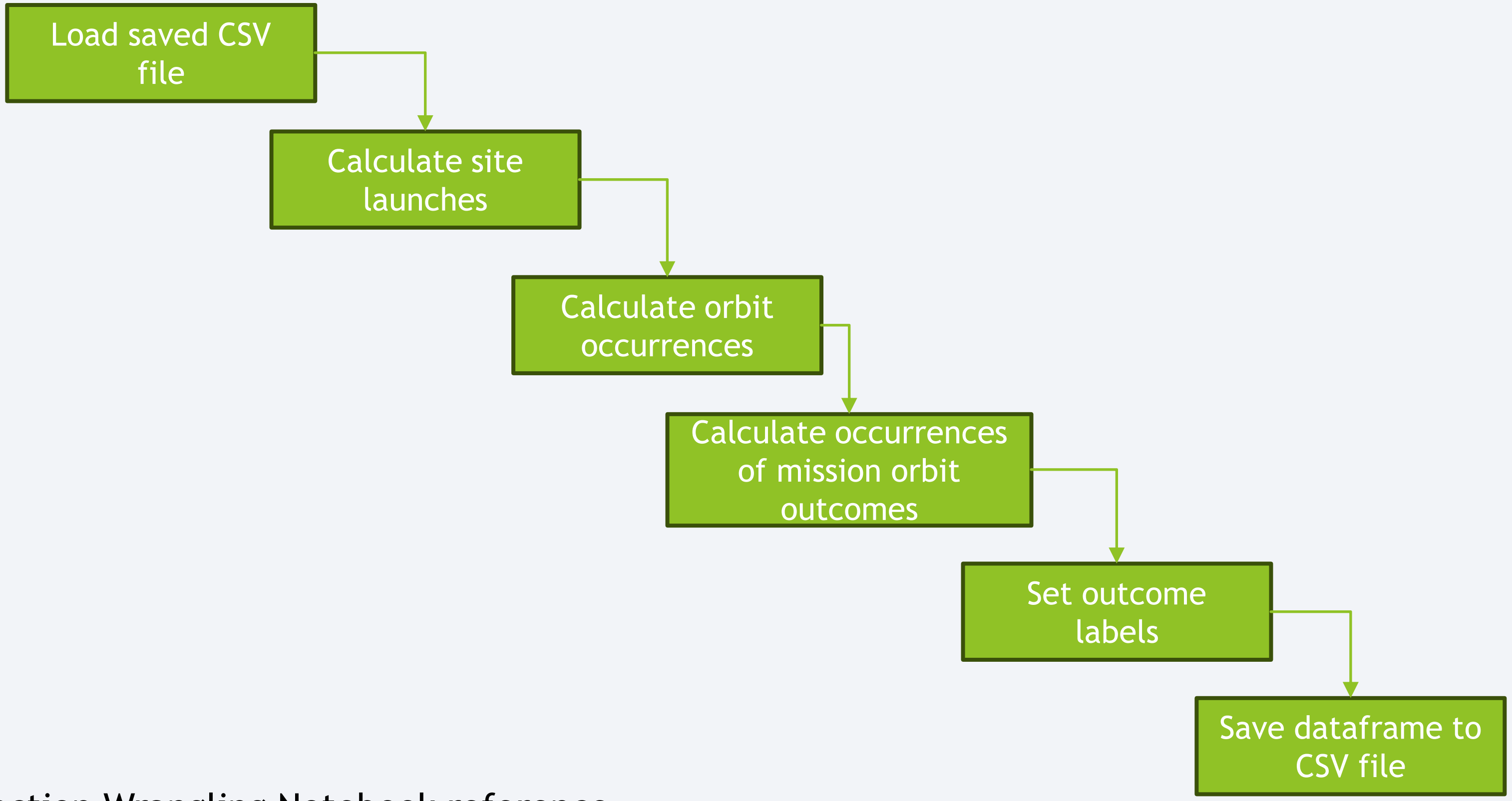
Process flow



Link to collection Scraping Notebook reference
https://github.com/MagMel8860/Mags_Rep_Capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

Process flow



Link to collection Wrangling Notebook reference
https://github.com/MagMel8860/Mags_Rep_Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Charts types include
  - Bar Chart
  - Scatter Plot
  - Line Plot

- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

Link to Data Visualization reference
https://github.com/MagMel8860/Mags_Rep_Capstone/blob/main/edadataviz.ipynb

EDA with SQL

- Queries performed
  - Create Table
  - Select distinct launch sites
  - 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 failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015
  - Rank the count of landing outcomes

Link to EDA with SQL reference
https://github.com/MagMel8860/Mags_Rep_Capstone/blob/main/jupyter-labs-eda-sql.ipynb

Build an Interactive Map with Folium

- Added to Folium map several Markers.
  - Circles
  - Distance
  - Lines
  - Launch site locations
- Added these markers to identify distance to from launch site to other key locations on the map.

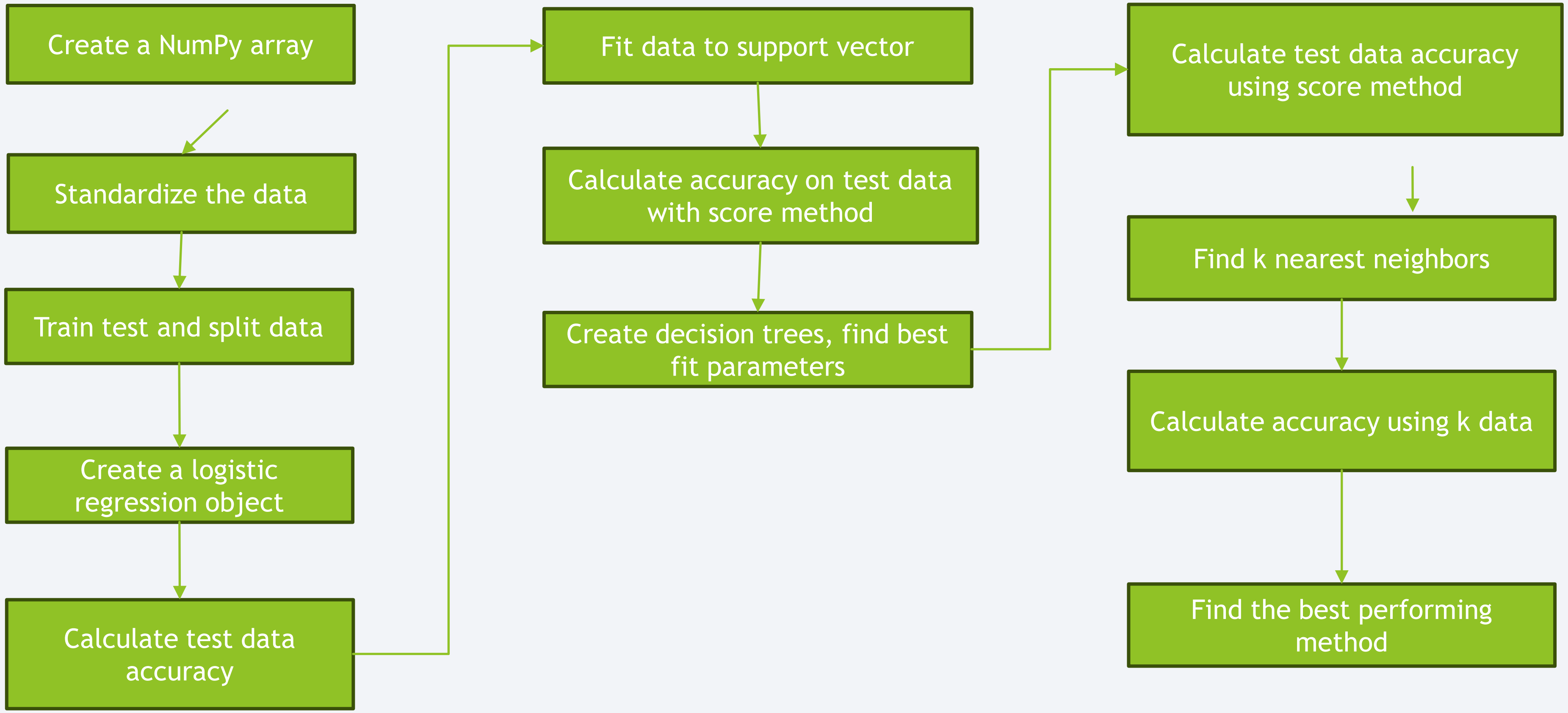
Link to Interactive Map Folium reference
https://github.com/MagMel8860/Mags_Rep_Capstone/blob/main/Interactive%20Folium.ipynb

Build a Dashboard with Plotly Dash

- Contains following chart
  - Pie chart, showing the successful launches. Easy to determine percentages visually
  - Scatter chart, showing relation between launch success rate and payload mass. Easy to compare between booster versions

Link Dashboard with Plotly reference
https://github.com/MagMel8860/Mags_Rep_Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)



Link Predictive Analysis reference
https://github.com/MagMel8860/Mags_Rep_Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

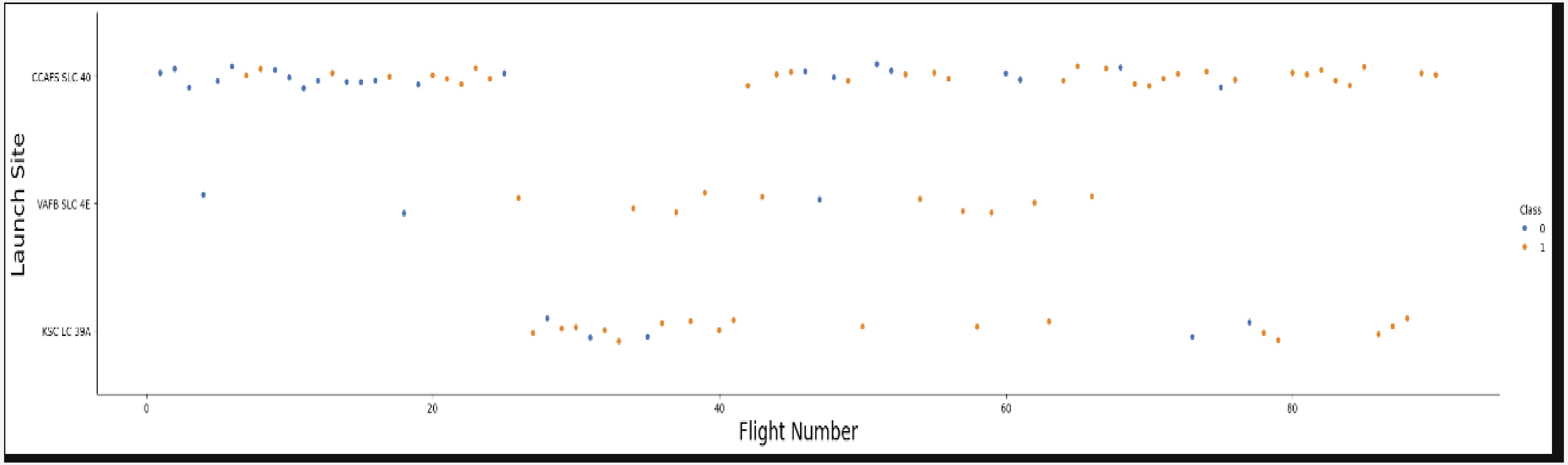
Results

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

Section 2

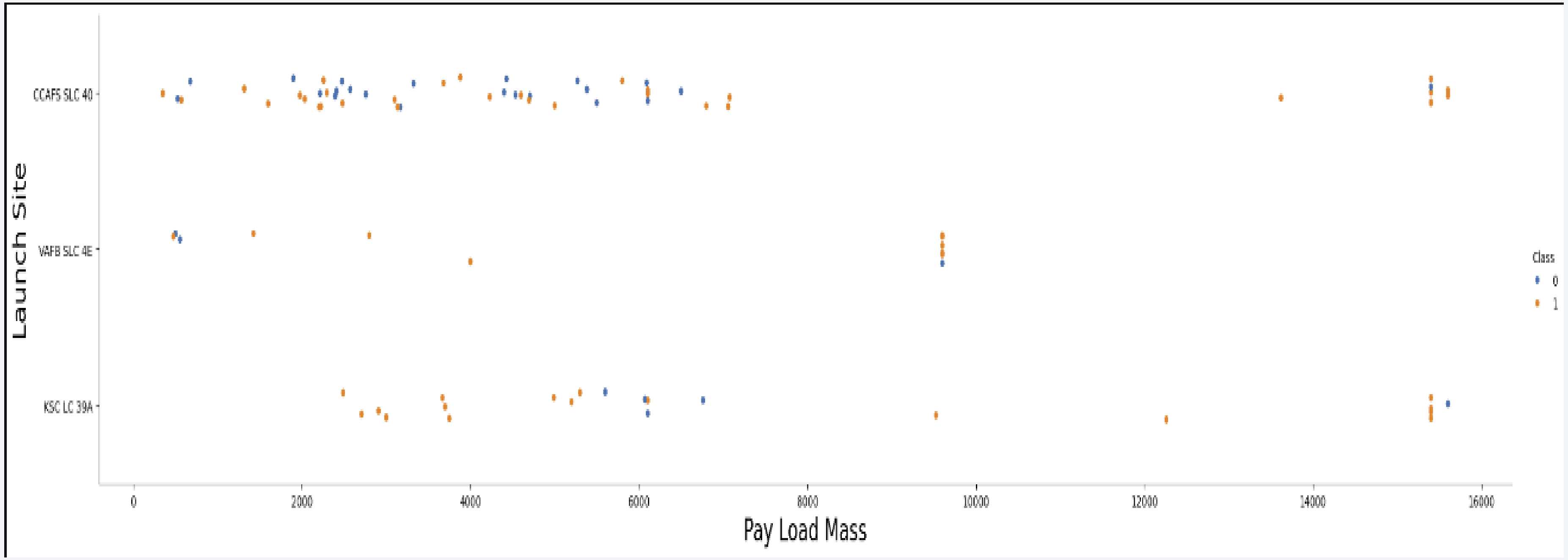
Insights drawn from EDA

Flight Number vs. Launch Site



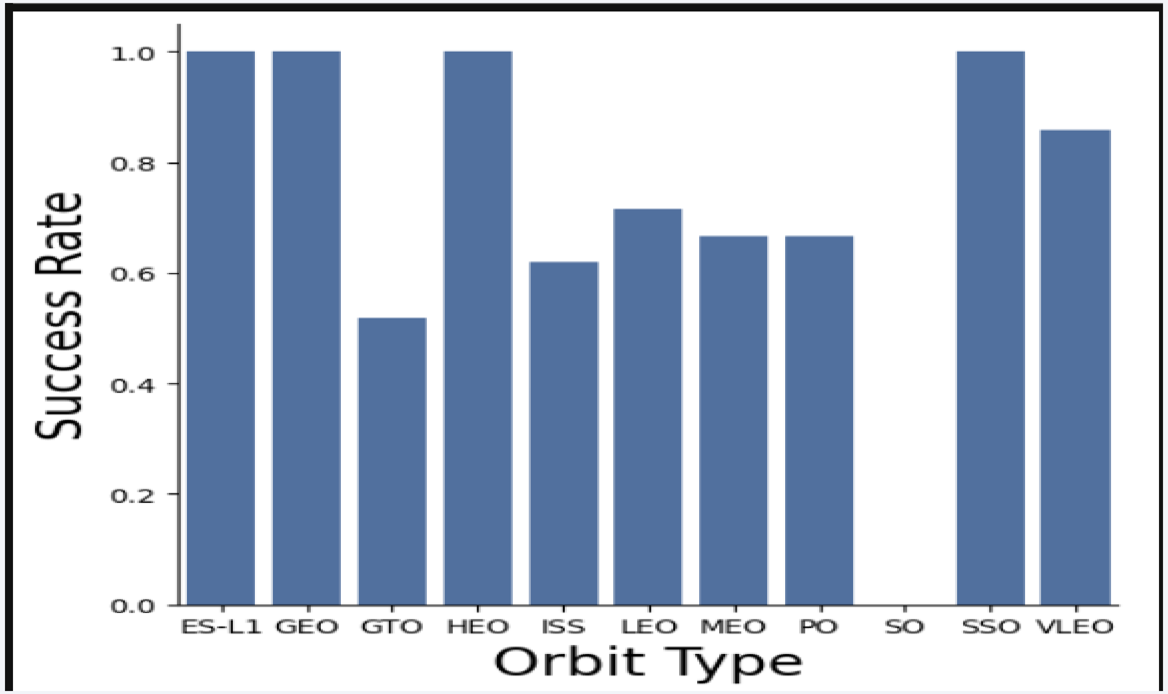
- The chart above can tell us the following
  - The more flights the greater the success rate
  - Site CCAFS SLC-40 has had overall much more launches than other sites.
  - Site VAFB SLC-4E has had more successes but overall, less flights

Payload vs. Launch Site



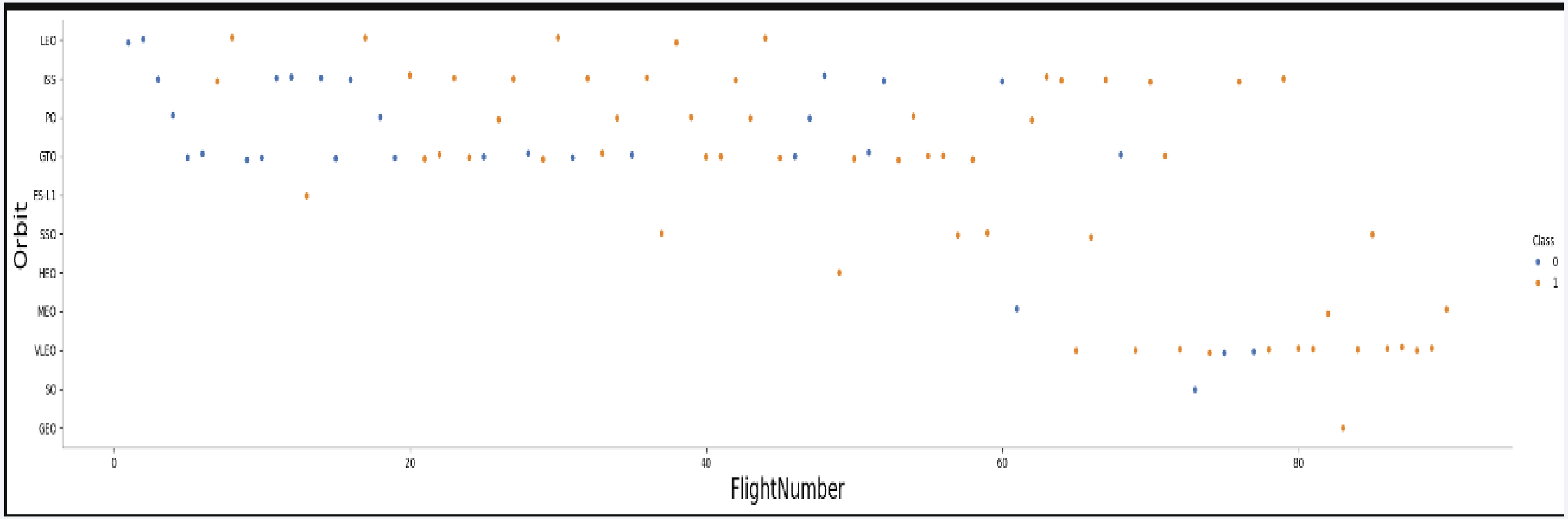
- The chart above can tell us the following
  - Very few launches with payloads between 8000 and 1400
  - Site KSC LC-39A is best with payload below 6000
  - Fir midrange payload of around 3000, site VAFB SLC-4E would be a good choice

Success Rate vs. Orbit Type



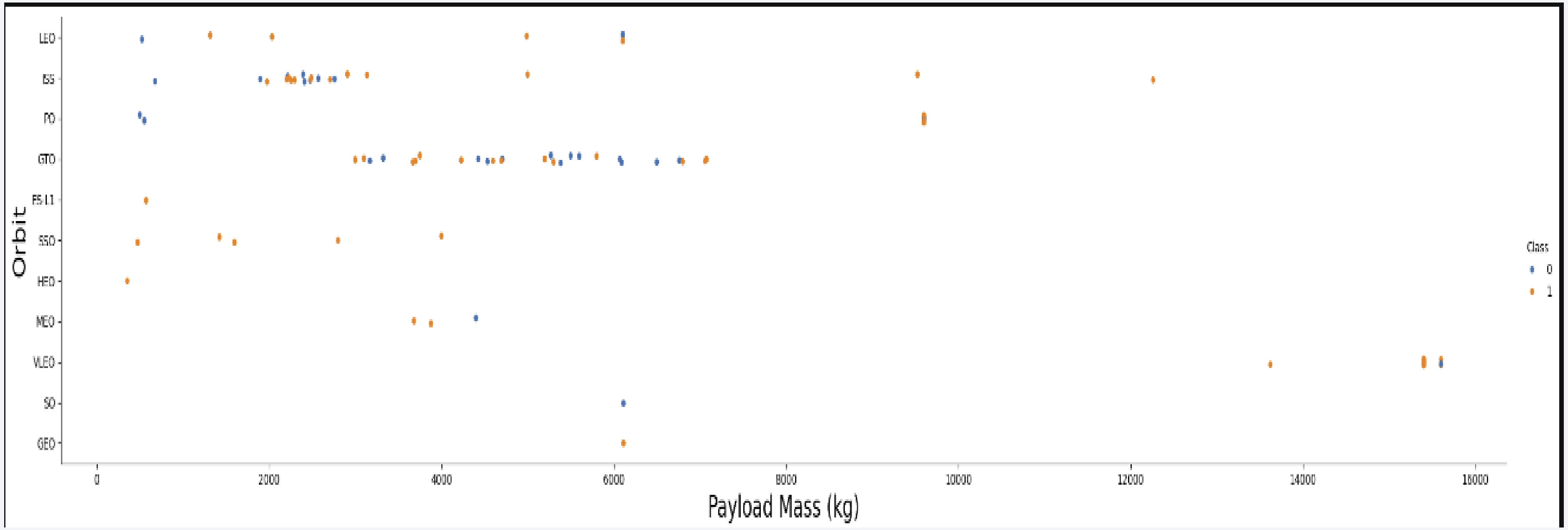
- The chart above can tell us the following
  - Best Orbit are ES-L1, GEO, HEO, SSO
  - Next best orbit type is VLEO
  - Orbit type SO has no successes

Flight Number vs. Orbit Type



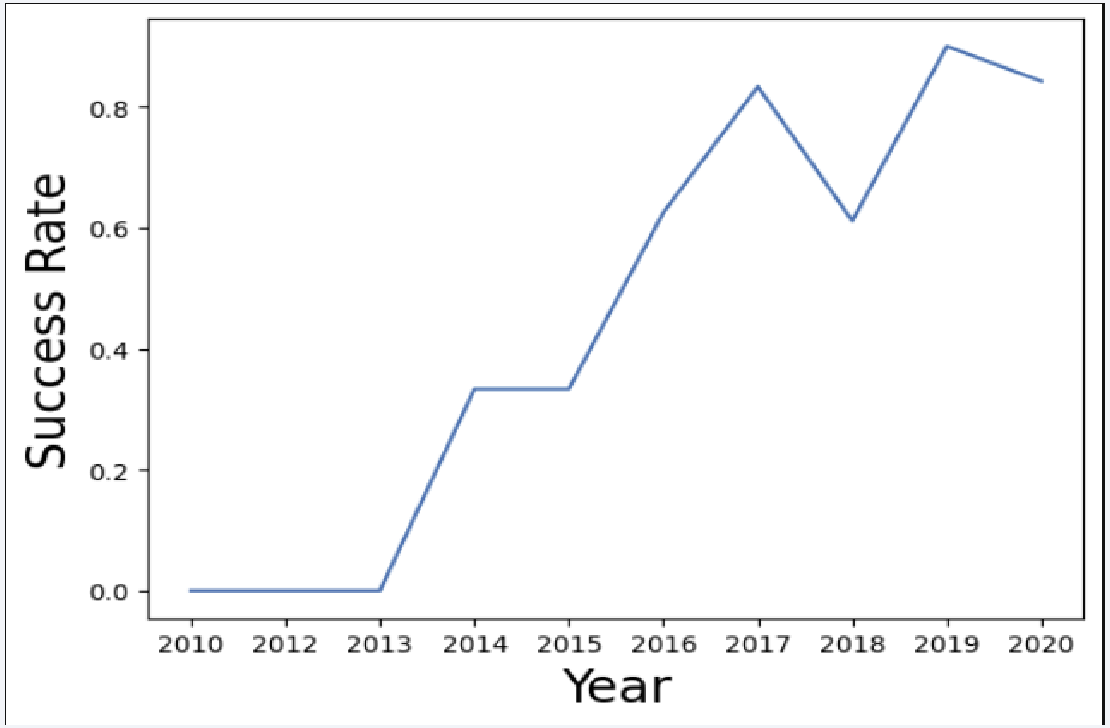
- The chart above can tell us the following
  - Orbit ISS increased success with increased flights
  - Orbit SO lacks success data for first 60 flights
  - Most orbits used are PO and GTO

Payload vs. Orbit Type



- The chart above can tell us the following
  - Orbit SO used for payloads over 13000 kg
  - Payload between 2000 and 800 have a mixed success rate
  - For payload below 4000 site SSO has more successes

Launch Success Yearly Trend



- The chart above can tell us the following
  - Success rate usually increase with each year.
  - Year 2018 had a large dip in success rate

All Launch Site Names

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

Query used

```
%sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL ORDER BY 1;
```

Launch Site Names 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	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

- The chart above can tell us the following
  - Payload range from 0 – 677 KG
  - Orbit if LEO and (ISS)

Total Payload Mass

TOTAL_PAYLOAD for payloads like CRS

111268

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD FROM SPACEXTBL
WHERE PAYLOAD  LIKE '%CRS%';
```

Filter payload with % % as to capture any records where CRS is in the payload text

Average Payload Mass by F9 v1.1

AVG_PAYLOAD
928

%sql SELECT AVG(PAYLOAD_MASS__KG_) AS AVG_PAYLOAD FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1';

Used AVG function to calculate the average

First Successful Ground Landing Date

FIRST_SUCCESS_GP
2015-12-22

```
%sql SELECT MIN(DATE) AS FIRST_SUCCESS_GP FROM SPACEXTBL WHERE
LANDING_OUTCOME = 'Success (ground pad)';
```

Used the MIN function to get record with oldest (first) date in dataset

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000 AND LANDING_OUTCOME = 'Success (drone ship)';
```

Filtered data based on payload size and success

Total Number of Successful and Failure Mission Outcomes

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

%sql SELECT MISSION_OUTCOME, COUNT(*) AS QTY FROM SPACEXTBL GROUP BY MISSION_OUTCOME ORDER BY MISSION_OUTCOME;

Data Grouped by success outcomes

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

```
%sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE
PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
ORDER BY BOOSTER_VERSION;
```

Used a sub query to first find the mass payload. Then selected records matching that criteria

2015 Launch Records

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

```
%sql SELECT substr(Date, 6,2) as MONTH, LANDING_OUTCOME ,BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL WHERE LANDING_OUTCOME = 'Failure (drone ship)' AND substr(Date,0,5)='2015';
```

Selected failed launch for the year 2015

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

Landing_Outcome	QTY
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

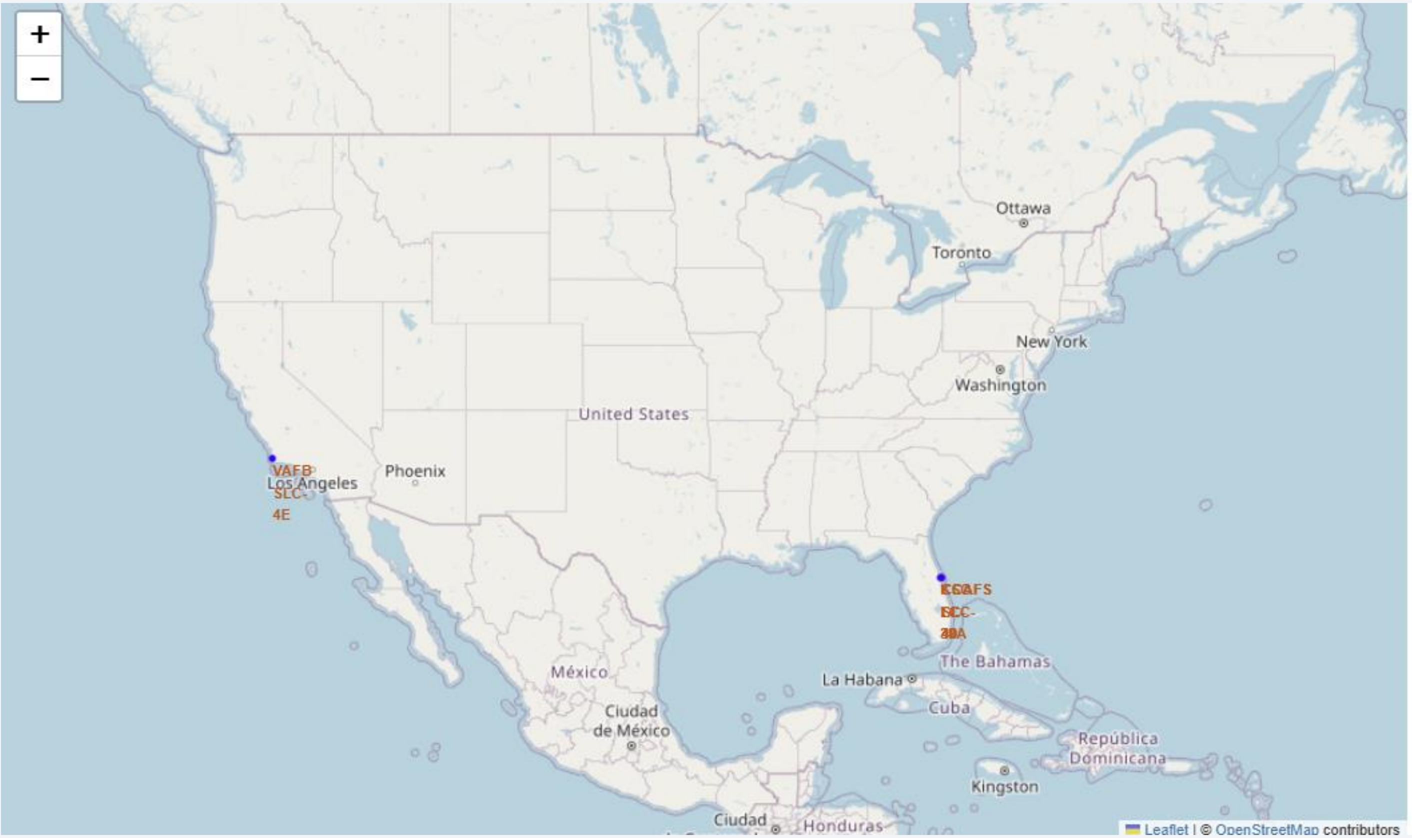
```
%sql SELECT LANDING_OUTCOME, COUNT(*) AS QTY FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING_OUTCOME ORDER BY QTY DESC;
```

Total landing outcomes between 2010-06-04 and 2017-03-20 listed in descending order

Section 3

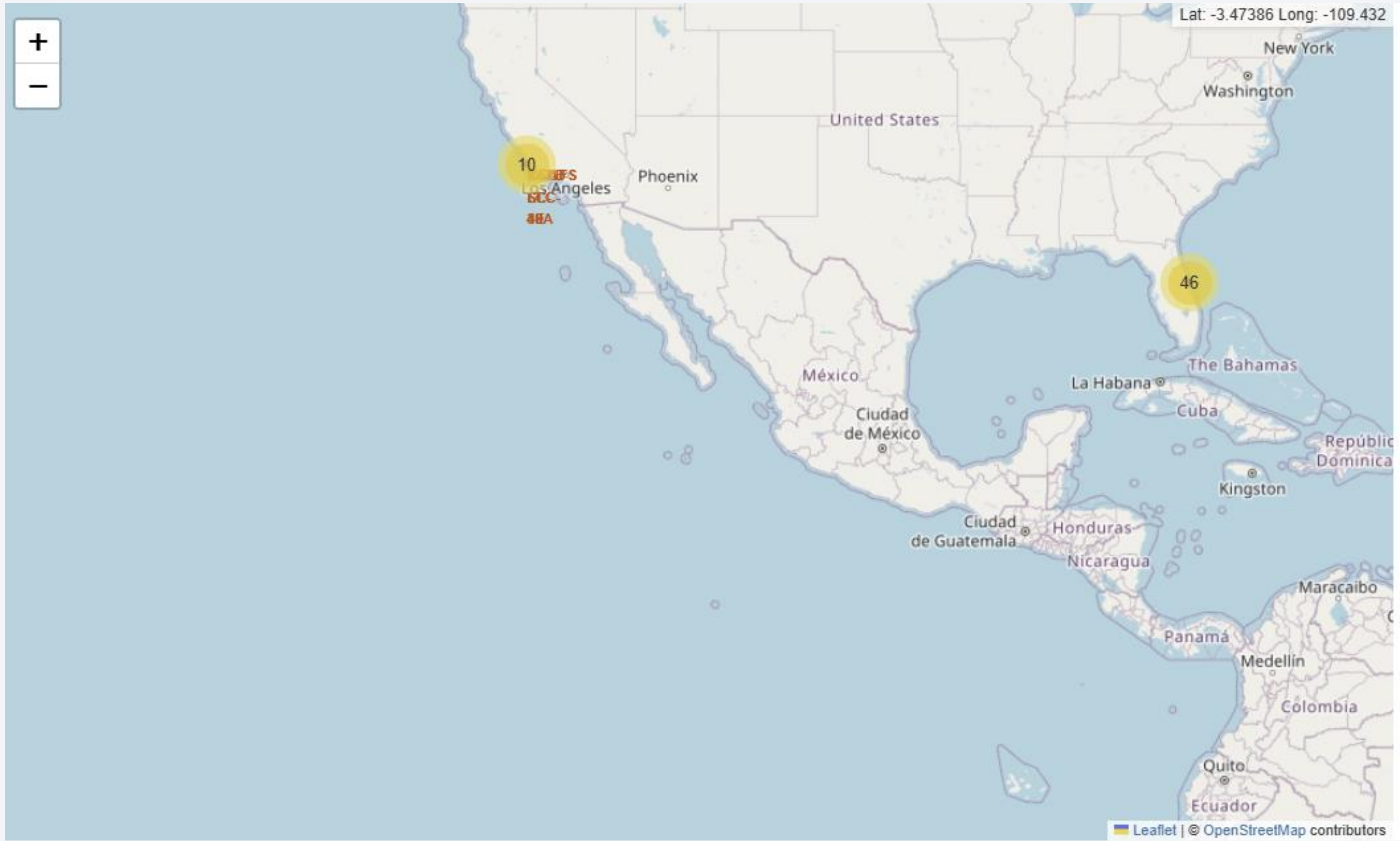
Launch Sites Proximities Analysis

Map of launch site locations



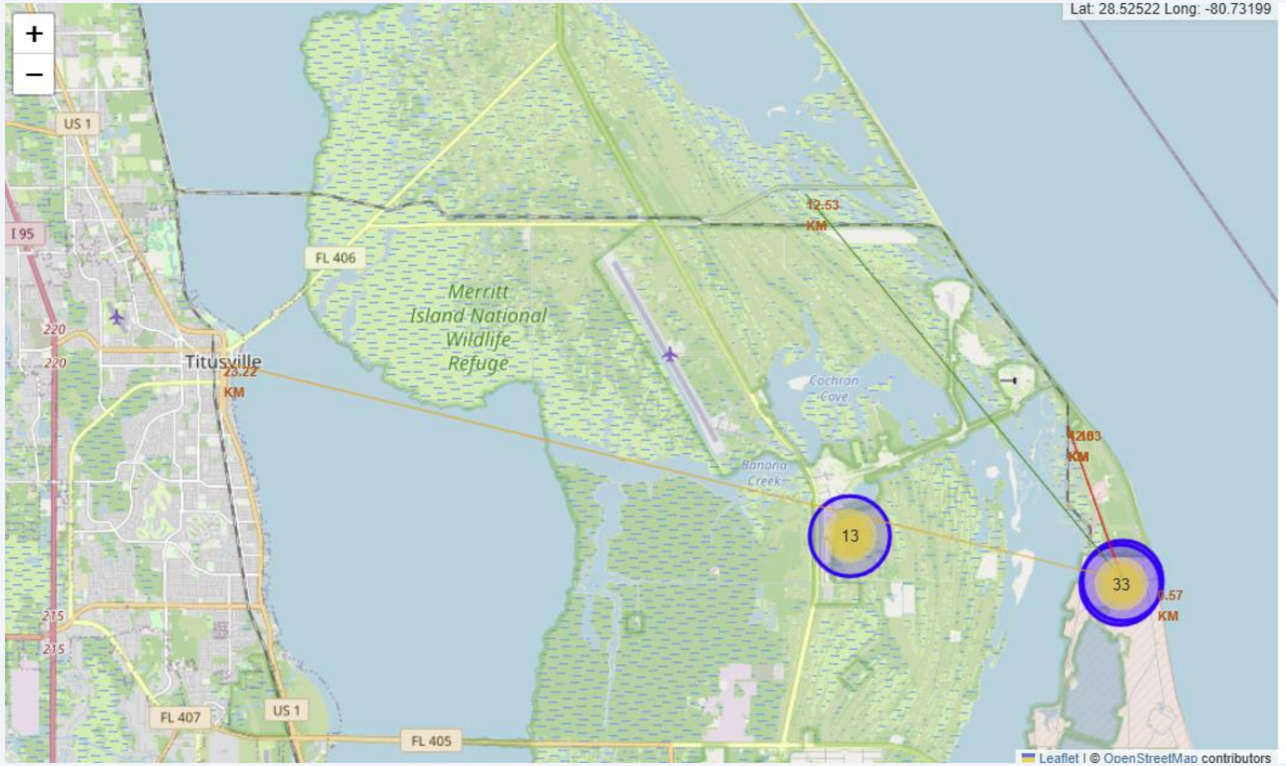
The map marks the locations of the launch site.
From that marks we can see sites are located near large bodies of water.
Sites are on the southern part of the US

Map of location site with latitude and longitude indicator



On this map we can see that latitude and longitude has been added to the upper right. Changes as mouse pointer moves on map

Map showing distance between site location and points of interest

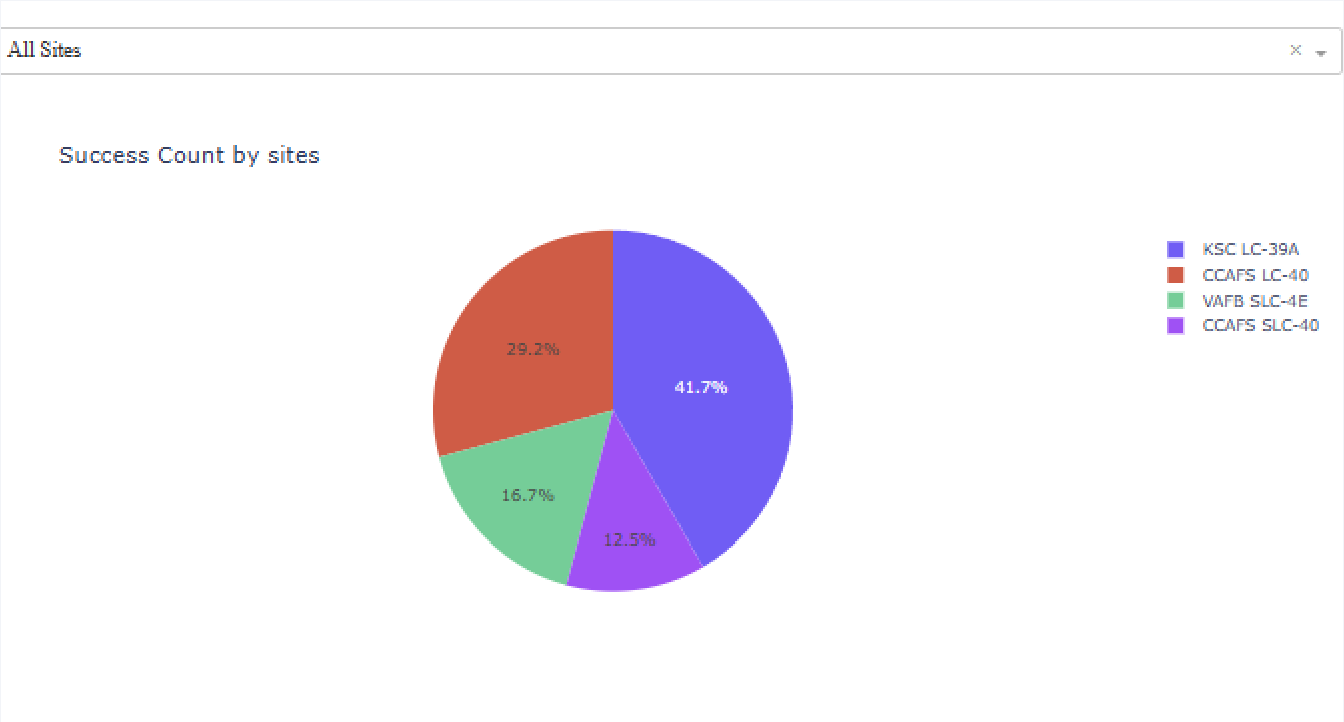


Map list the distance between a point of interest and launch site. The two points distance is calculated for are shown by a line market

Section 4

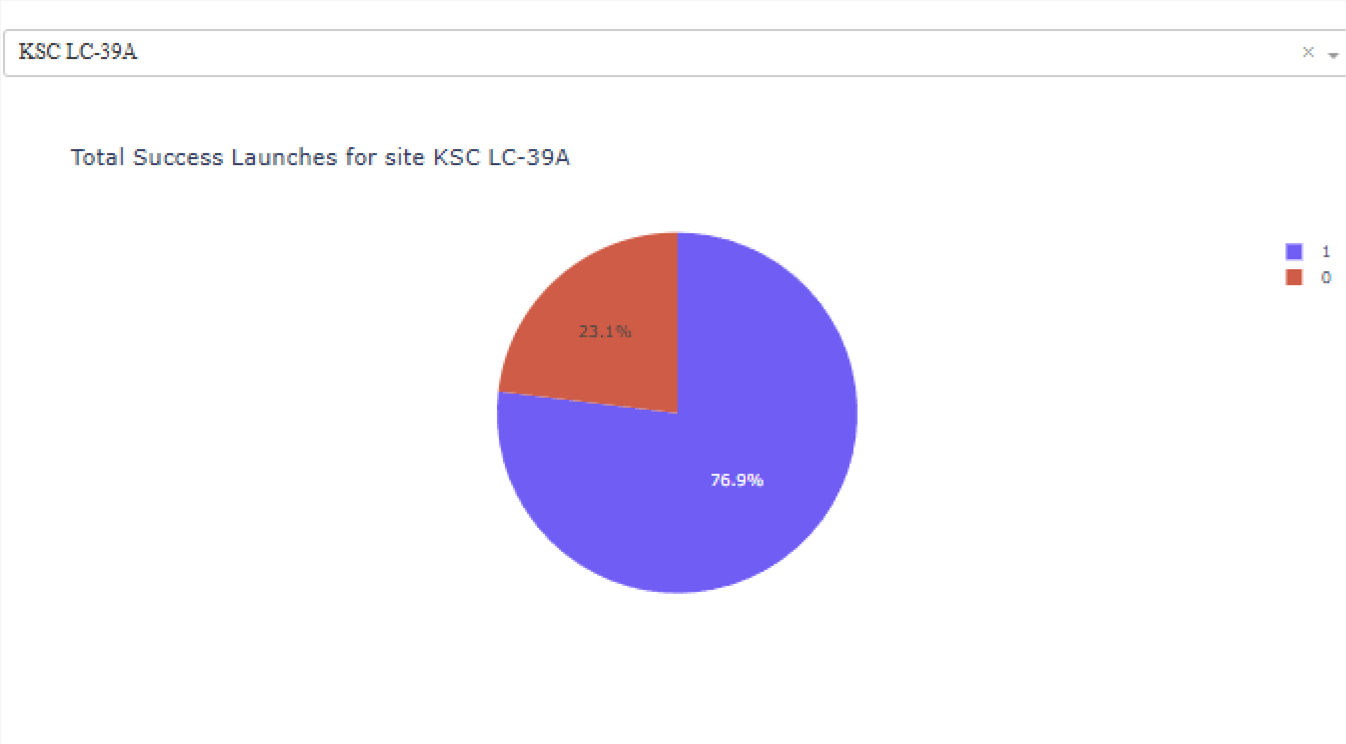
Build a Dashboard with Plotly Dash

Pie chart of success rates



We can see from the chart site KSC LC-39A has the best success rate.
Can also see how it compares to the other sites

Pie Chart of Best Success Rate



We can see the rate of the of the site with the best rate.

Chart Showing Correlation Between Payload and Class

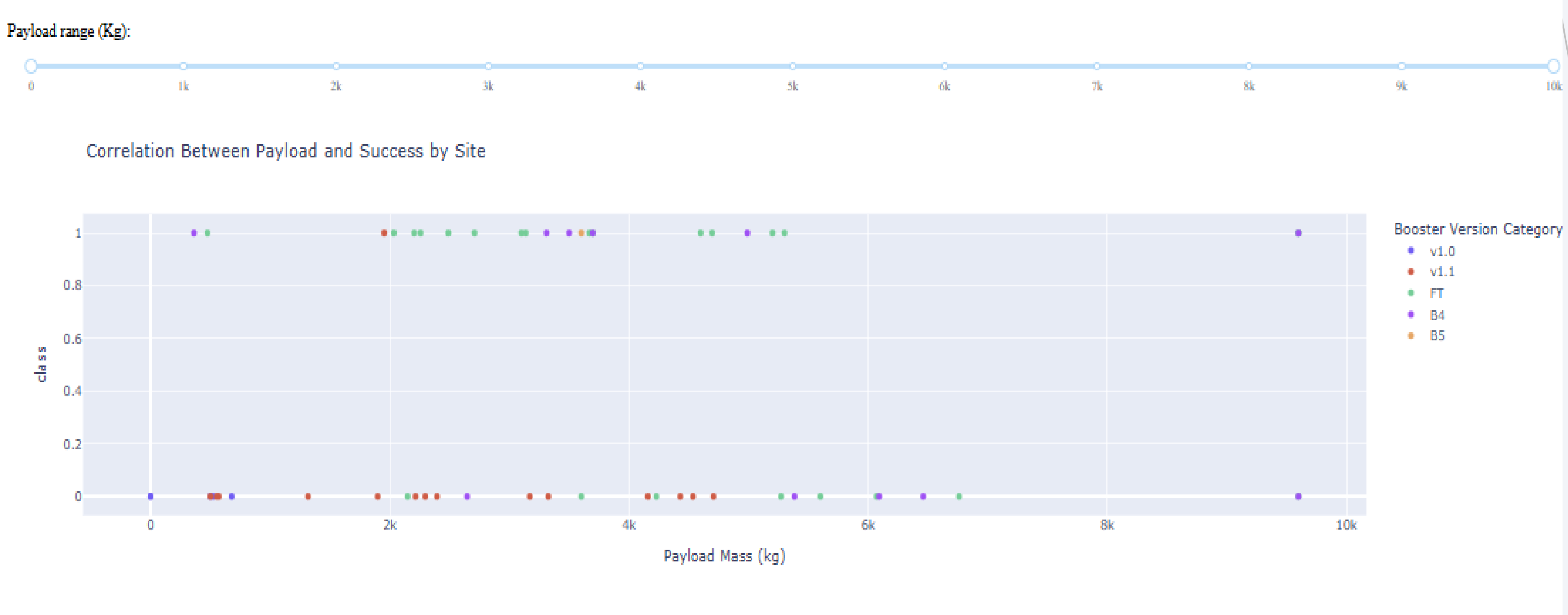
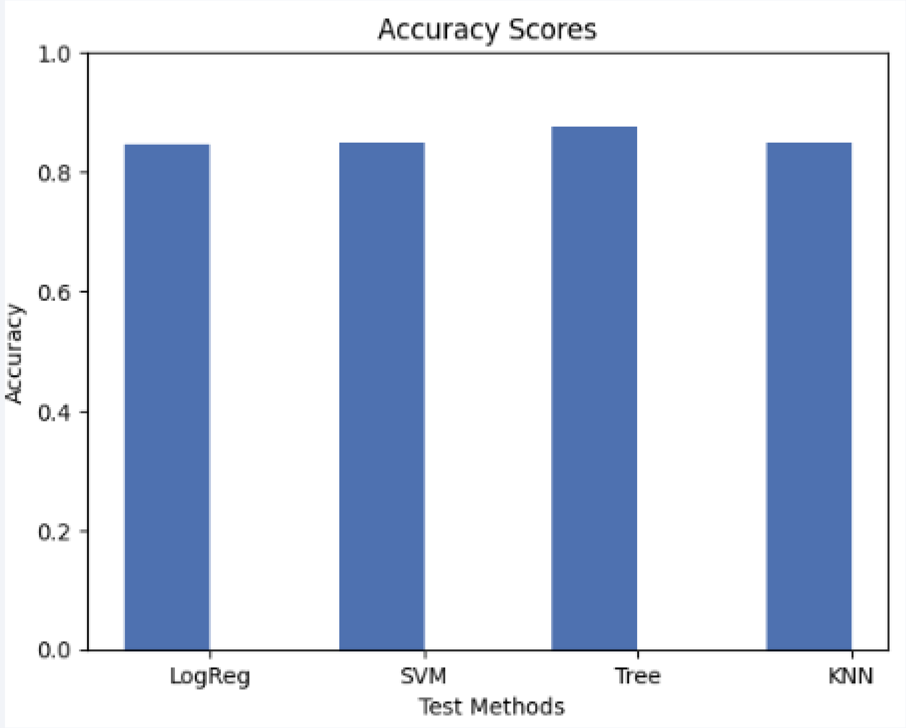


Chart shows booster FT has best success rate

Section 5

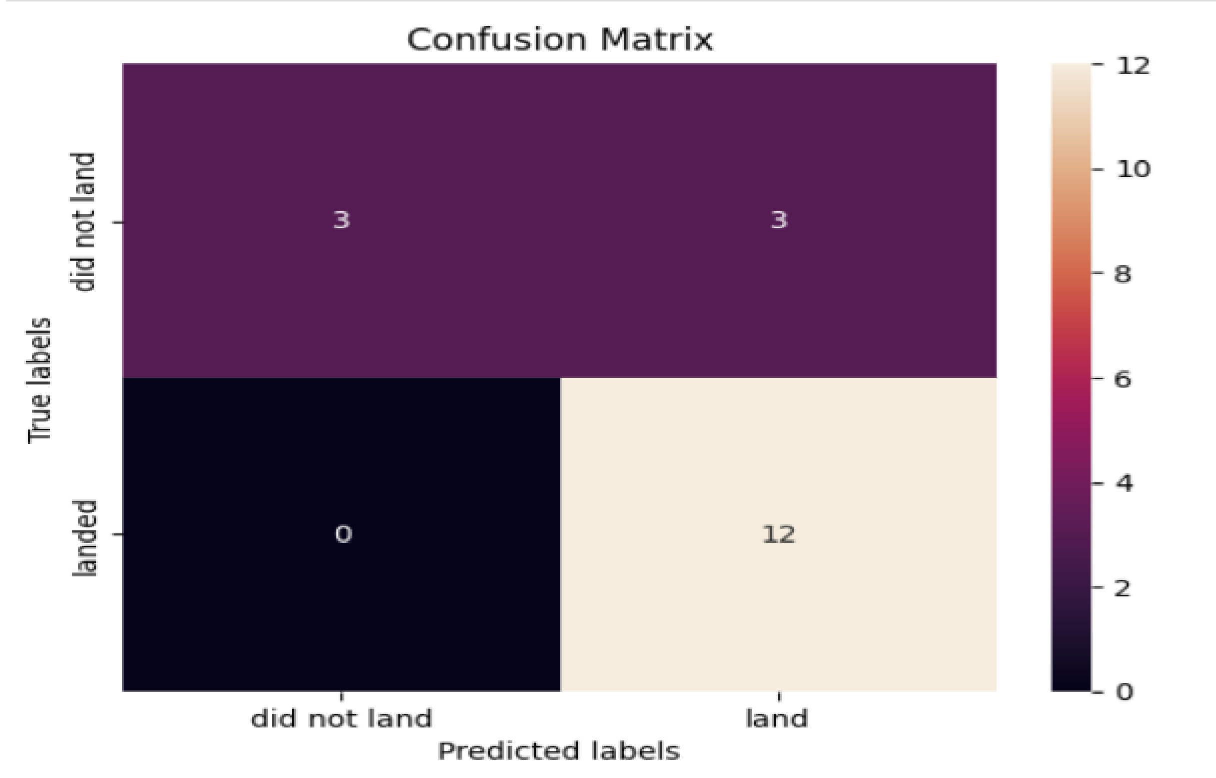
Predictive Analysis (Classification)

Classification Accuracy



The chart shows the Tree classification method has the best accuracy

Confusion Matrix



From the confusion matrix we can see most landing fall with the true positive section
Equal amounts between true negative and false positive

Conclusions

- Success rates keep increasing with time. Some years may have small dips but increase rate bones back up
- Best change for a successful lunch is with site CCAFS LC-40 or CCAFS SLC-40.
- Increase chance of successful launches increases with payload below 6000 Kg
- Combining above mentioned launch site with orbits ES-L1, GEO, HEO, SSO produce the best chance of a successful launch

Appendix

Resources used for preparation

www.coursera.org

Jupyter Notebooks in Skills Network Labs

www.wikipedia.org for launch data

Microsoft Power point

Thank you!