Winning Space Race with Data Science

Anthony Doyle
10th July 2024

Outline

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

Executive Summary

Summary of methodologies

- **Data Collection** – gathering historical data on the Space X Falcon 9 launches with a view to understanding the success or failure of first stage landings.
- **Feature Selection** – Extract relevant data that impact on successful landings such as payload details, orbit types, and location of launch.
- **Model Selection** – Use supervised learning algorithms to classify the data (logistic regression, support vector machines, decision tree, and K nearest neighbor) to train predictive models.

Summary of all results

- The decision tree algorithm scored the highest in terms of percentage of correct predictions on the test data.

Introduction

Project background and context

Space Y would like to compete with Space X.

SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each.

The purpose of this project is to gather information about Space X from past launch data. Space X are able to significantly reduce the cost of missions compared to their competitors, due to the ability to reuse the first stage of the launch.

The project hopes to define a predictive model that can determine the cost of a launch based on specific parameters.

Problems you want to find answers

Prediction of successful first landing

Cost estimation for future launches

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

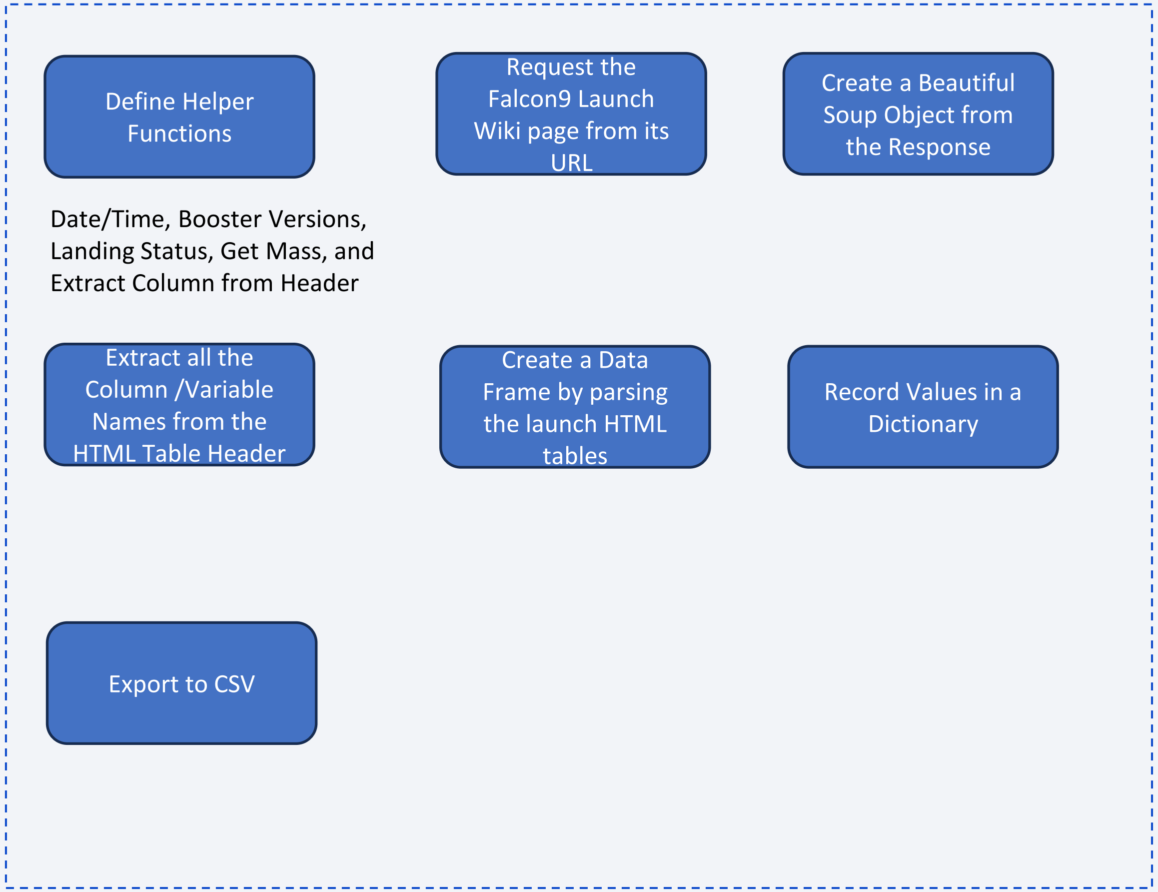
- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts
- The GitHub URL of the completed SpaceX API calls notebook.
- The Final CSV Output

Data Collection - Scraping

- Web scraping process
- The GitHub URL of the completed web scraping notebook
- The final CSV Output



Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts
- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

EDA with Data Visualization

- Summary of key charts
  - Scatterplot showing the relationship between Payload mass and flight numbers, overlaid with the outcome of each launch
  - Scatterplot showing the relationship between the launch site and the flight number, overlaid with the outcome of each launch.
  - Scatterplot showing the relationship between the launch site and the payload mass, overlaid with the outcome of each launch.
  - Bar chart plotting the success rate of each Launch by Orbit Type
  - Scatterplot showing the relationship between the payload mass and the orbit type, overlaid with the outcome of each launch.
  - A line plot showing the average success rate trend
- The GitHub URL of completed EDA with data visualization notebook.
- The final CSV output

EDA with SQL

- Summary of the SQL queries
  - Select Distinct to provide the unique list of launch sites in the space mission
  - Select Statement Limited to the first 5 launch sites beginning with 'CCA'
  - Select Statement to display the total payload mass carried by boosters launched by NASA (CRS)
  - Select Statement to display the average payload mass carried by booster version F0 v1.1
  - Select Min Statement to list the date when the first successful landing outcome in ground pad was achieved
  - Select Statement listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - Select Statement listing the total number of successful and failure mission outcomes
  - Select Statement listing the names of the booster versions which have carried the maximum payload mass.
  - Select Statement listing the records which will display the month names, failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015.
  - Select Statement 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 GitHub URL of your completed EDA with SQL notebook.

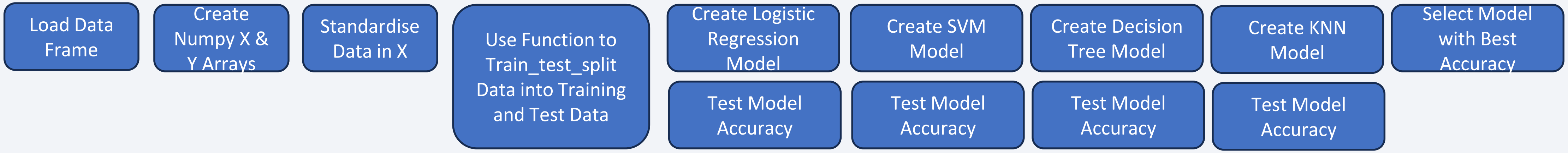
Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
  - Marked the 4 launch sites on the folium map, using coordinates, map object, folium circles, and folium markers.
  - Marked the success and failure launches for each site on the map, using the class column indicator, and red & green folium markers to signify success or failure
  - Calculated the distance between launch sites and proximities, using the coordinates, the distance function, and polylines to illustrate the distance.
- The [GitHub URL](#) of your completed interactive map with Folium map.

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)



The GitHub URL of your completed predictive analysis lab.

Results

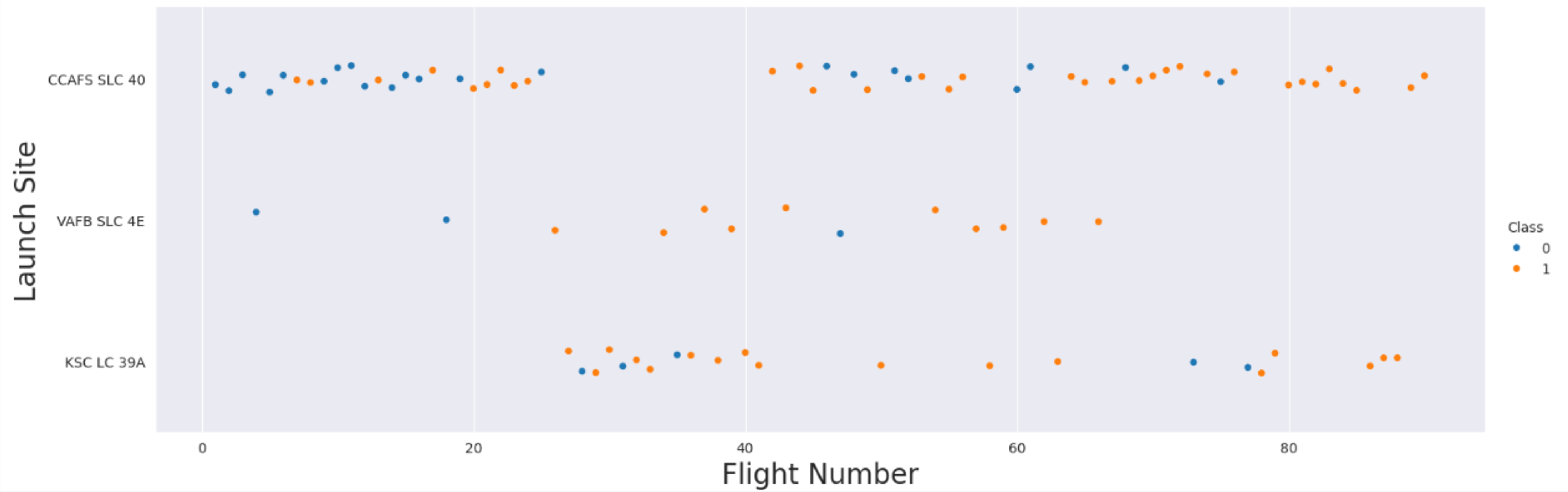
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

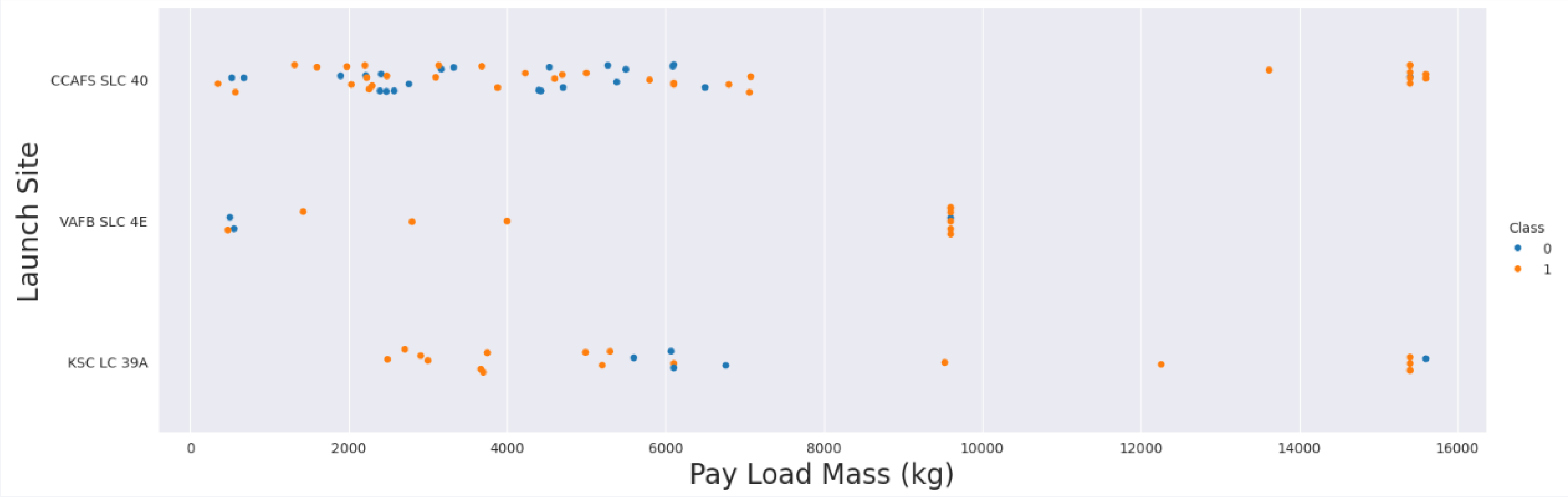
The scatter plot of Flight Number vs. Launch Site indicates CFAS SLC 40 has a higher success rate after 60 continuous flights



Payload vs. Launch Site

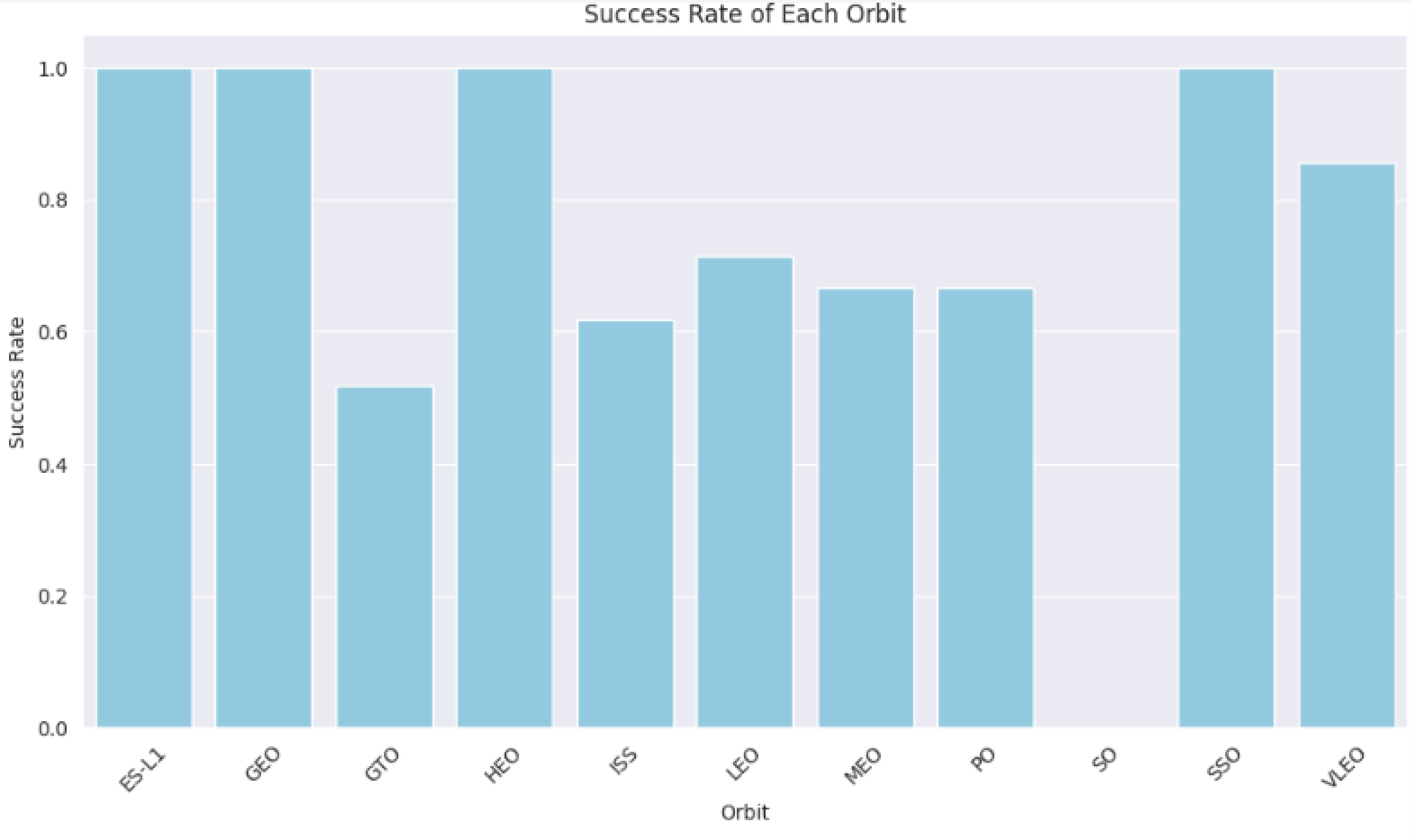
The Payload Vs. Launch Site scatter point chart indicates for the VAFB-SLC

launch site there are no rockets launched for heavy payload mass(greater than 10000)



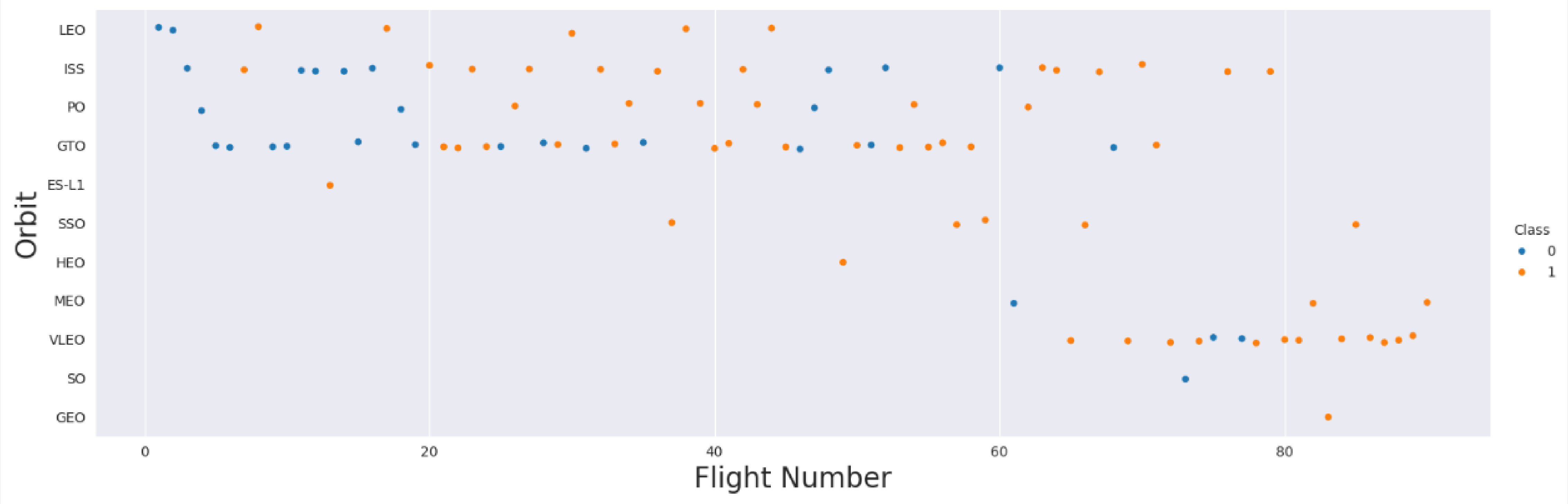
Success Rate vs. Orbit Type

The ES L1, GEO, HEO, and SSO have the highest chance of success



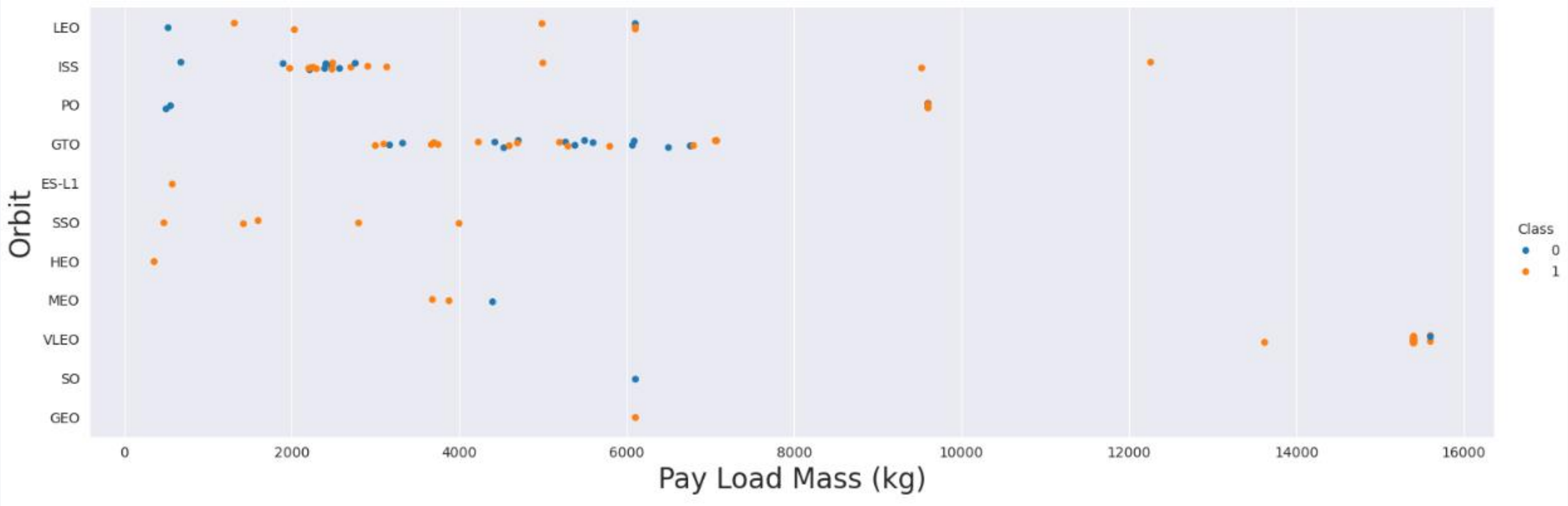
Flight Number vs. Orbit Type

You should see that in 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.



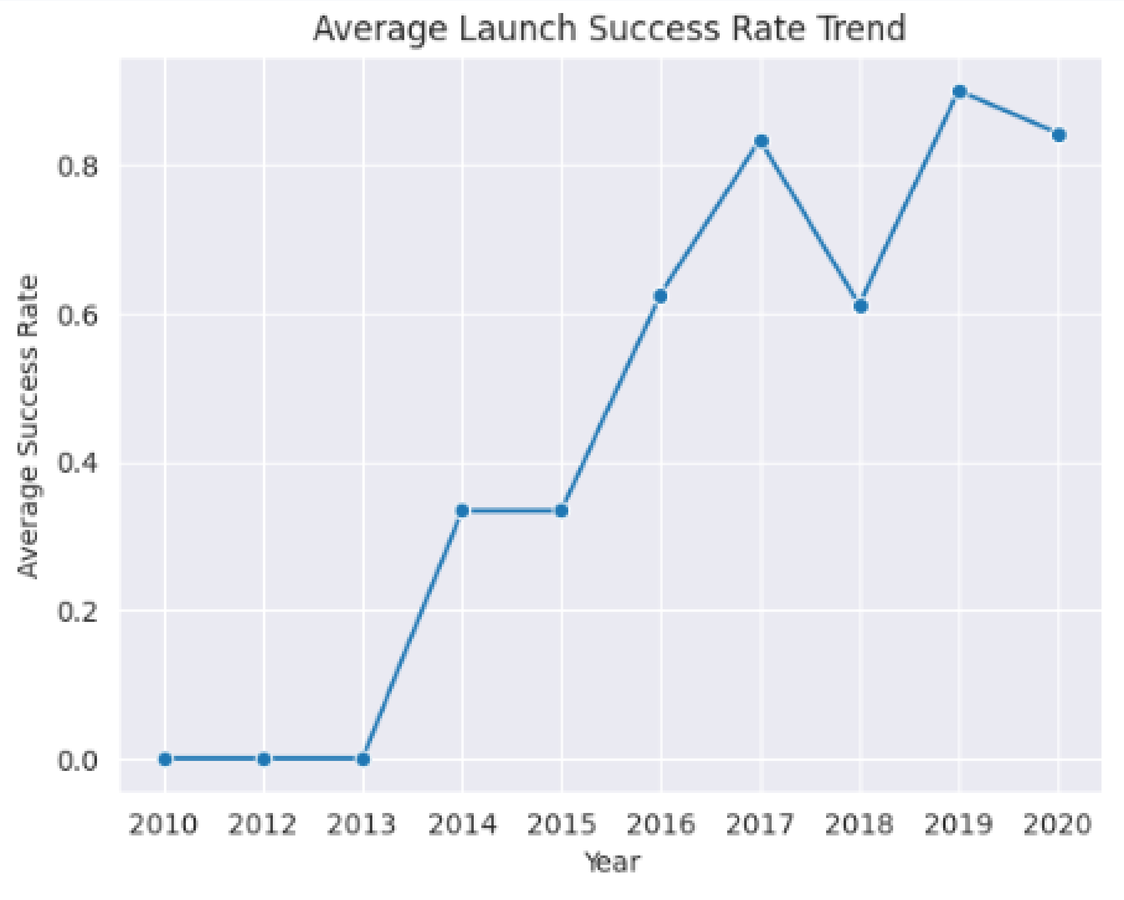
Payload vs. Orbit Type

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

You can observe that the success rate since 2013 has continued to improve



All Launch Site Names

The names of the unique launch sites

Display the names of the unique launch sites in the space mission

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTBL;
```

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

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

Launch Site Names Begin with 'CCA'

5 records where launch sites begin with `CCA`

Display 5 records where launch sites begin with the string 'CCA'

```
%sql SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
```

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

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

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") AS Total_Payload_Mass FROM SPACEXTBL WHERE Customer LIKE '%CRS%';
```

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

Total_Payload_Mass
48213

Average Payload Mass by F9 v1.1

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

Display average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") AS Average_Payload_Mass FROM SPACEXTBL WHERE Booster_Version LIKE '%F9 v1.1%';
```

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

Average_Payload_Mass
2534.6666666666665

First Successful Ground Landing Date

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

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
%sql SELECT MIN(Date) AS First_Successful_Landing_Date, Landing_Outcome FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground pad)';
```

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

First_Successful_Landing_Date	Landing_Outcome
2015-12-22	Success (ground pad)

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

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql SELECT Booster_Version FROM SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)'AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

```
 * 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

Calculate the total number of successful and failure mission outcomes

List the total number of successful and failure mission outcomes

```
%sql SELECT Mission_Outcome, COUNT(*) AS Outcome_Count FROM SPACEXTBL WHERE Mission_Outcome IN ('Success', 'Failure (in flight)') GROUP BY Mission_Outcome;
```

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

Mission_Outcome	Outcome_Count
Failure (in flight)	1
Success	98

Boosters Carried Maximum Payload

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

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%sql SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

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

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 in year 2015

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
%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';
```

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

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

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

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.

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

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

Landing_Outcome	Outcome_Count
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

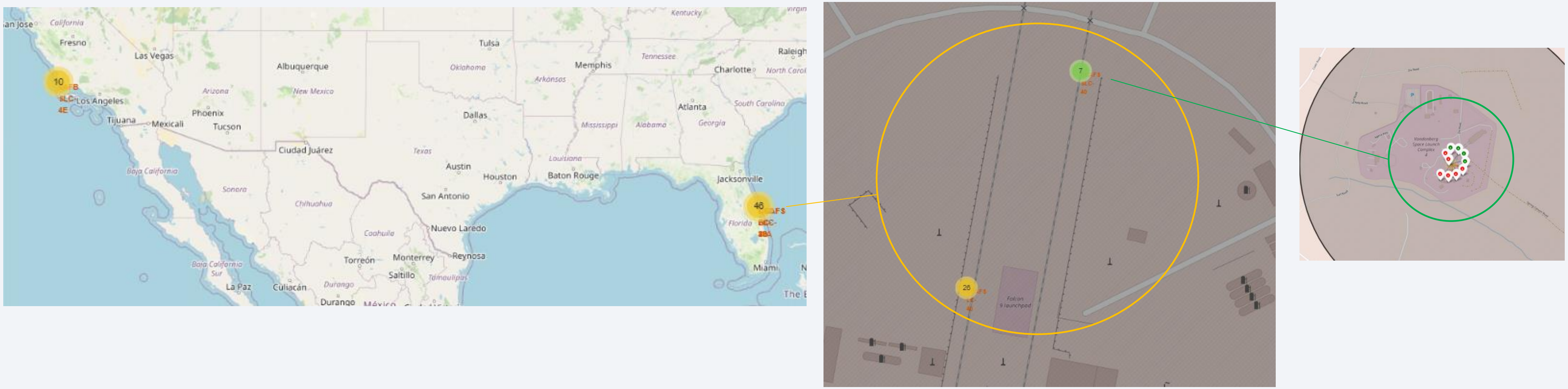
Launch Sites Proximities Analysis

Launch site Locations

- All the launch sites are situated approx. 30 degrees above the Equator line.
- All the launch sites are within close proximity to the coast. There are 4 launch sites, 1 situated on the West Coast, 3 situated on the East Coast

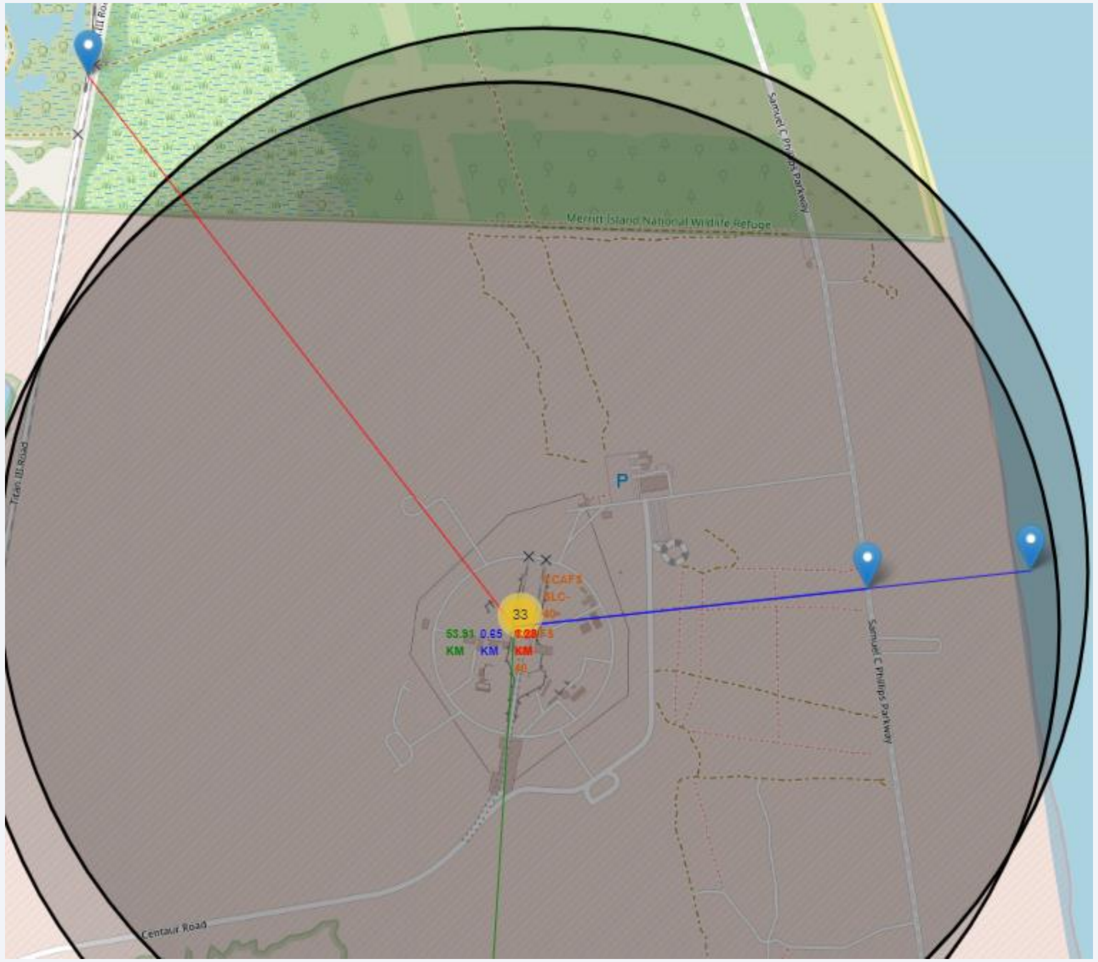
Launch Success Rates

- CCAFS SLC-40 has a higher rate of success indicated by the green marker cluster on the map.



Launch Site Proximity to Prominent Features

The distance from the launch site to the closest coastline is approximately 0.95 kilometers.
The distance from the launch site to the closest railway is approximately 1.28 kilometers.
The distance from the launch site to the closest highway is approximately 0.65 kilometers.
The distance from the launch site to the closest city is approximately 53.91 kilometers.



Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title

- Show the screenshot of launch success count for all sites, in a piechart

- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title

- Show the screenshot of the piechart for the launch site with highest launch success ratio

- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title

- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider

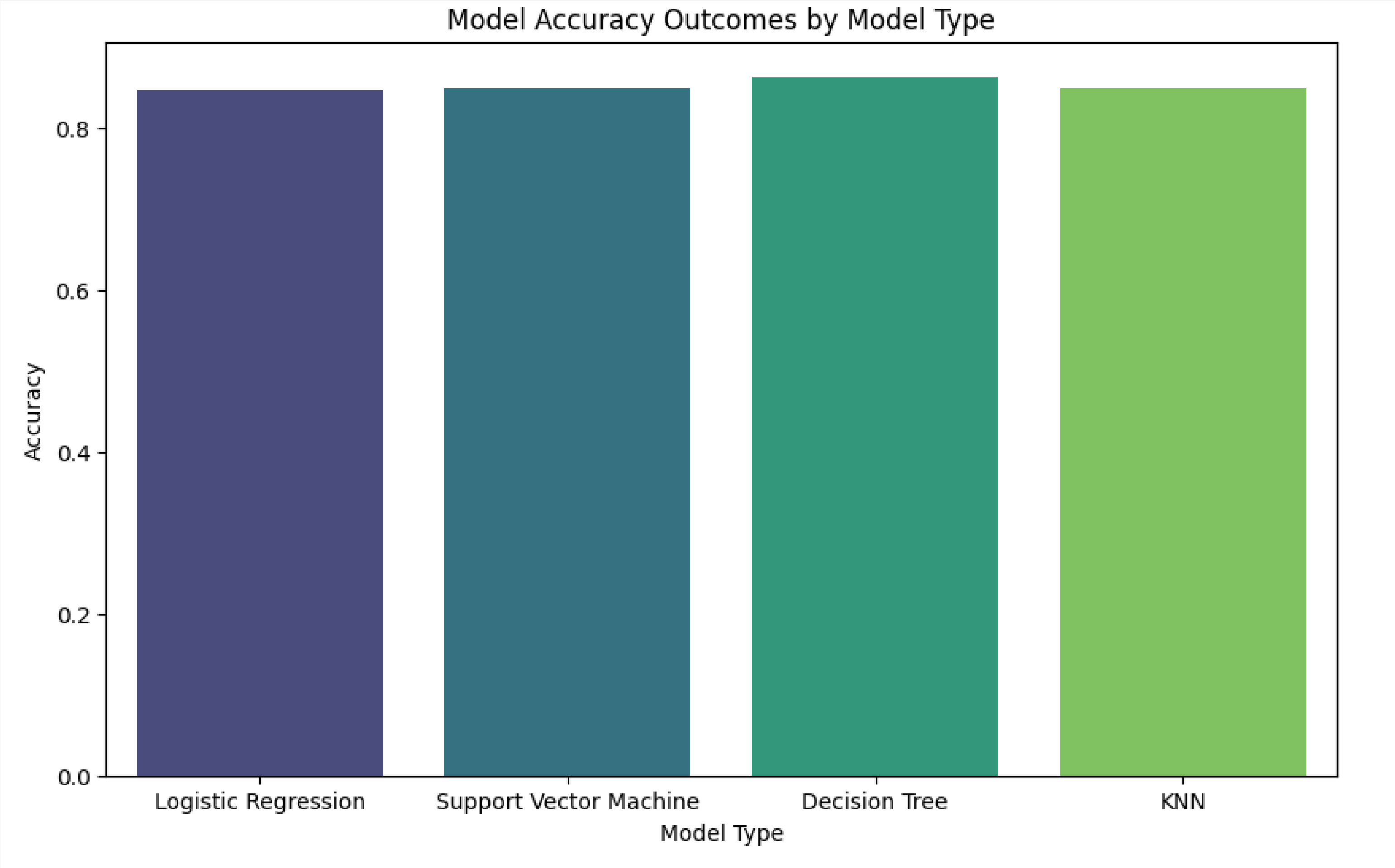
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

Predictive Analysis (Classification)

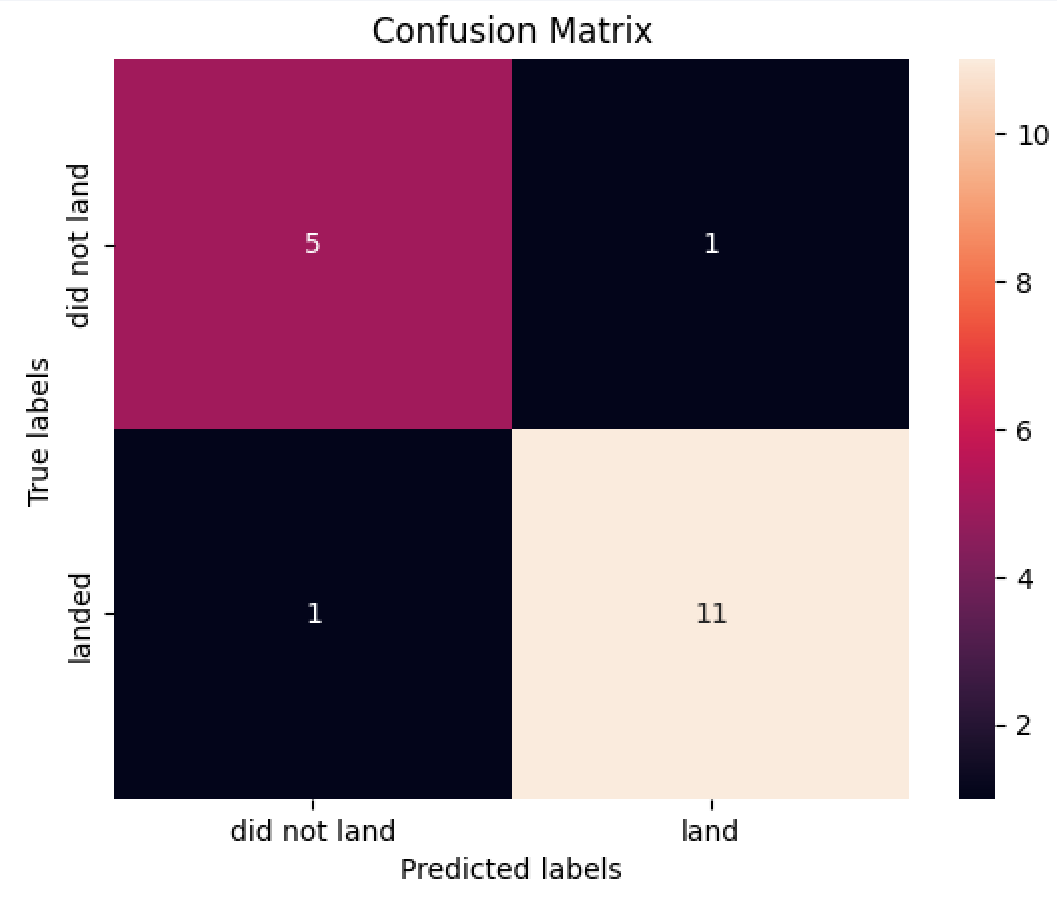
Classification Accuracy

The Decision Tree Model is the most accurate with a score of 0.863



Confusion Matrix

The confusion matrix of the best performing model. The model has a high number of true positives, and accurately predicted true negatives in most cases.



Conclusions

- The CCAFS SLC 40 site has the highest success rate for recovery of first stage launches
- The success rate of the CCAFS SLC 40 site improves with each flight

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!