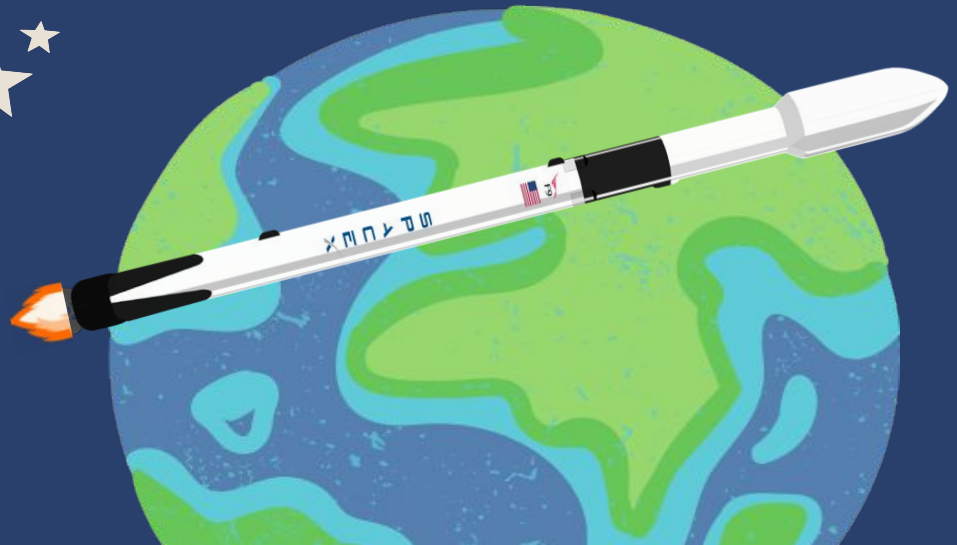
Back for More

SpaceX Falcon 9 first stage Landing Prediction



Ian Batangan

09/17/2024

Slidedeck Outline

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

Executive Summary

Summary of Methodologies

- **Data Collection** using web scraping and SpaceX API and **Data Wrangling** using Pandas and NumPy
- **Exploratory Data Analysis (EDA) with Data Visualization** using SQL , Pandas, Seaborn, and Matplotlib
- **Location Analysis** using Folium
- **Data Dashboard** using Plotly Dash
- **Predictive Analysis** using Machine Learning

Summary of All Result

Using the data collected from SpaceX, EDA yielded a clean database to analyze. Derived from this is the Interactive Map created with Folium to aid in the Location Analysis of the launch sites and the features needed for the Machine Learning in order to predict if the SpaceX Falcon 9 rocket can land back safely.

Introduction

In this project, we will predict if the Falcon 9 first stage will land successfully. 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, much of the savings is because SpaceX can reuse the first stage. With the available public information from SpaceX Rocket Launches we will assess the viability of a new company dubbed “Space Y” by evaluating the launch locations and determining if the first stage will land, therefore we can determine the cost of a launch.

Methodology

Methodology Executive Summary

- **Data collection methodology:**
  - Data was taken from SpaceX API (https://api.spacexdata.com/v4/rockets/) and web scraped from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)
- **Perform data wrangling**
  - Managing the missing values in the data and determining training labels
- **Perform exploratory data analysis (EDA) using visualization and SQL**
- **Perform interactive visual analytics using Folium and Plotly Dash**
- **Perform predictive analysis using classification models**
  - Collected data is standardized and classified , splitting the data into training set and test set, and finding best Hyperparameter for SVM, Classification Trees and Logistic Regression.

Data Collection

Data collection process is a combination of API requests from SpaceX REST API and Web Scraping the dataset from SpaceX's Wikipedia site. Cleaning the data with Pandas to remove extraneous data. The combination of the 2 datasets gives the basis for the data analysis.

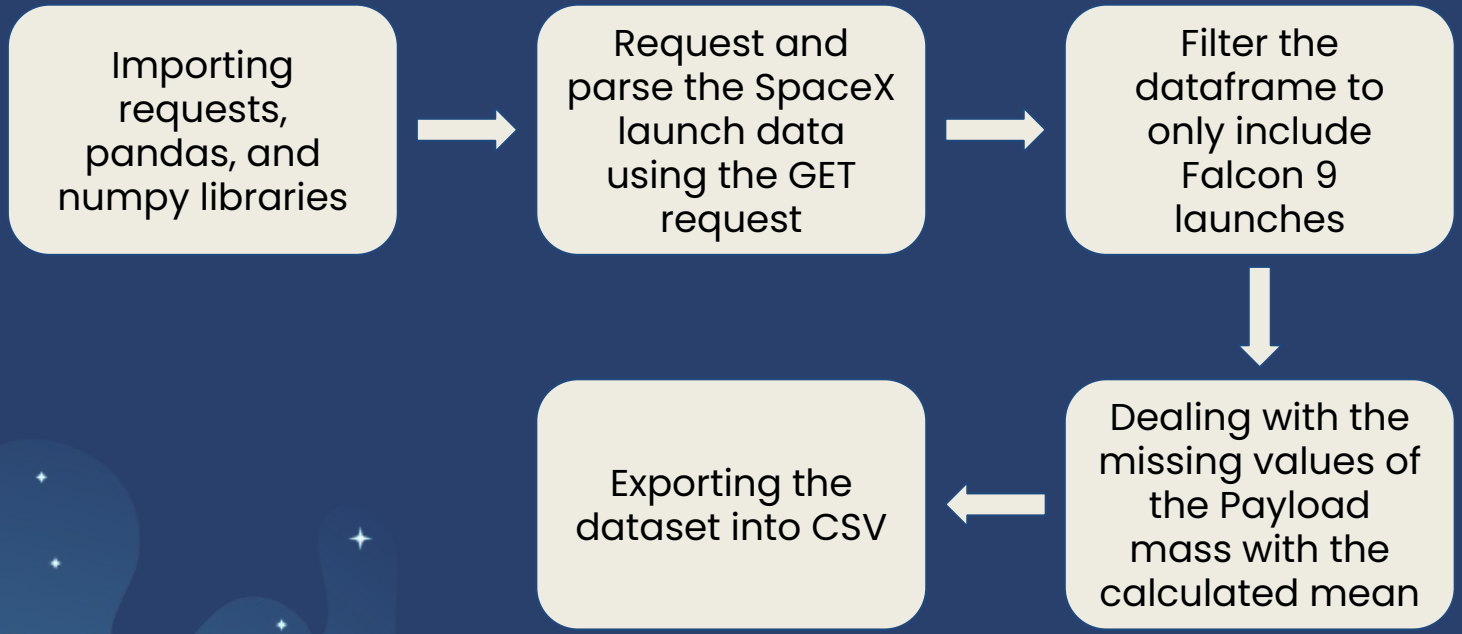
In where the Data Columns that are obtained by using SpaceX REST API are

FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude

and the Data Columns that are obtained by using Wikipedia Web Scraping are

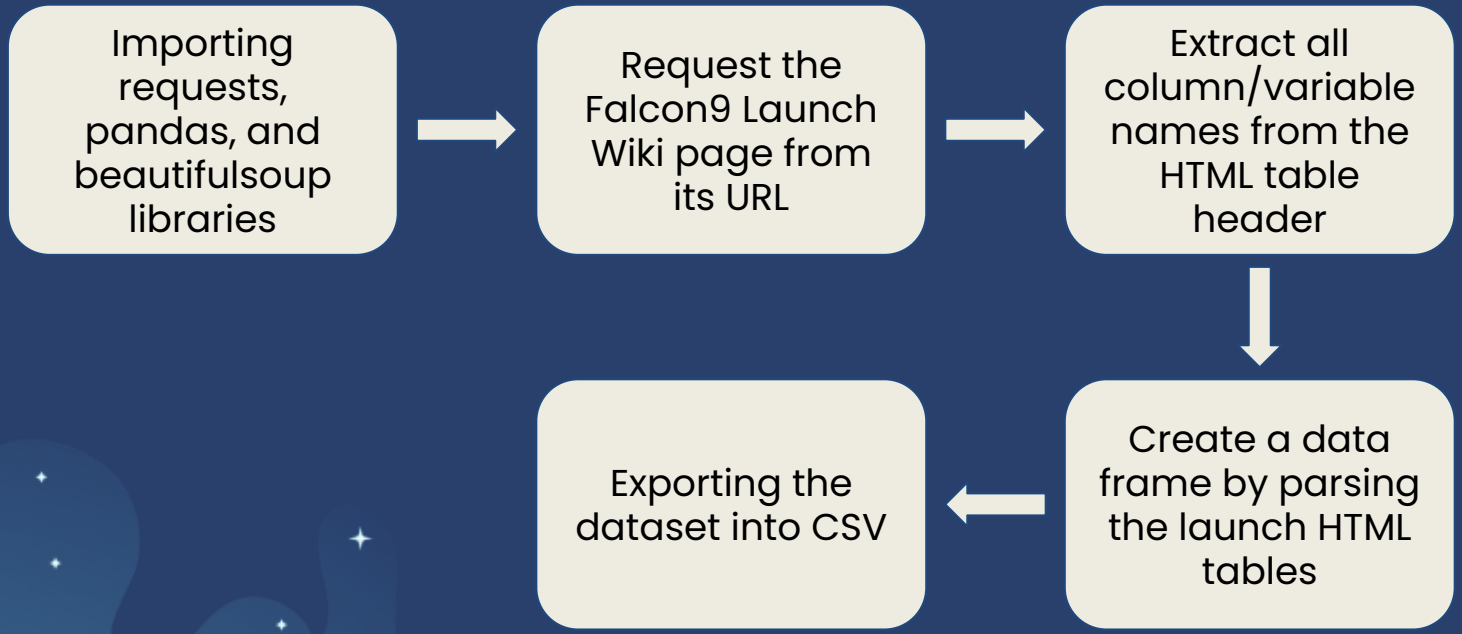
Flight No., Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time

Data Collection - SpaceX API



Github Code: 1 API Data Collection

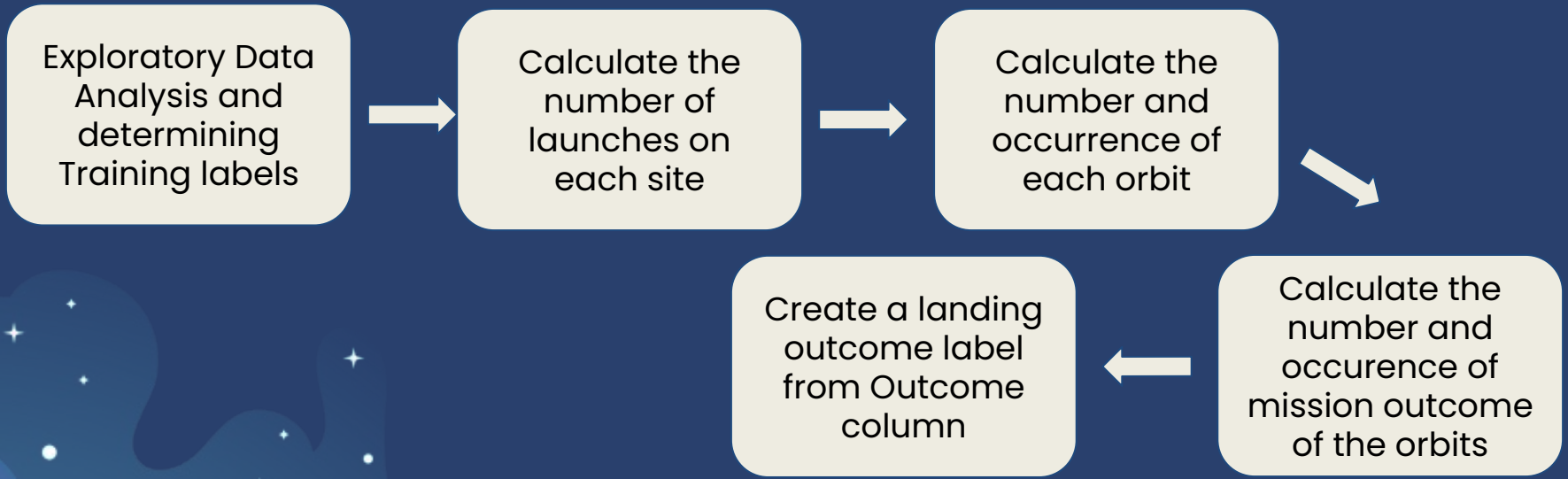
Data Collection - Scraping



Github Code: [2 Web Scraping Data Collection]

Data Wrangling

This is the start of the Exploratory Data Analysis, initially the data is grouped is grouped by launch site, occurrence of each orbit, and occurence of mission outcome of the orbits. Finally the creation of landing outcome labels.



Github Code: [3 Data Wrangling]

EDA with Data Visualization

- Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch Site, Orbit Type vs. Success Rate, Flight Number vs. Orbit Type were charted as scatter plots to show the relationship between variables and to be used for the machine learning model.
- Payload Mass vs Orbit Type is charted as a bar chart to show comparisons among discrete categories.
- Success Rate Yearly Trend is charted as a line chart to show the trend of the data over time

Github Code: 4 EDA Data Visualization

EDA with SQL

SQL queries:

- 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 succesful landing outcome in ground pad was acheived.
- 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 subquery
- 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
- 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

Github Code: 5 EDA using SQL

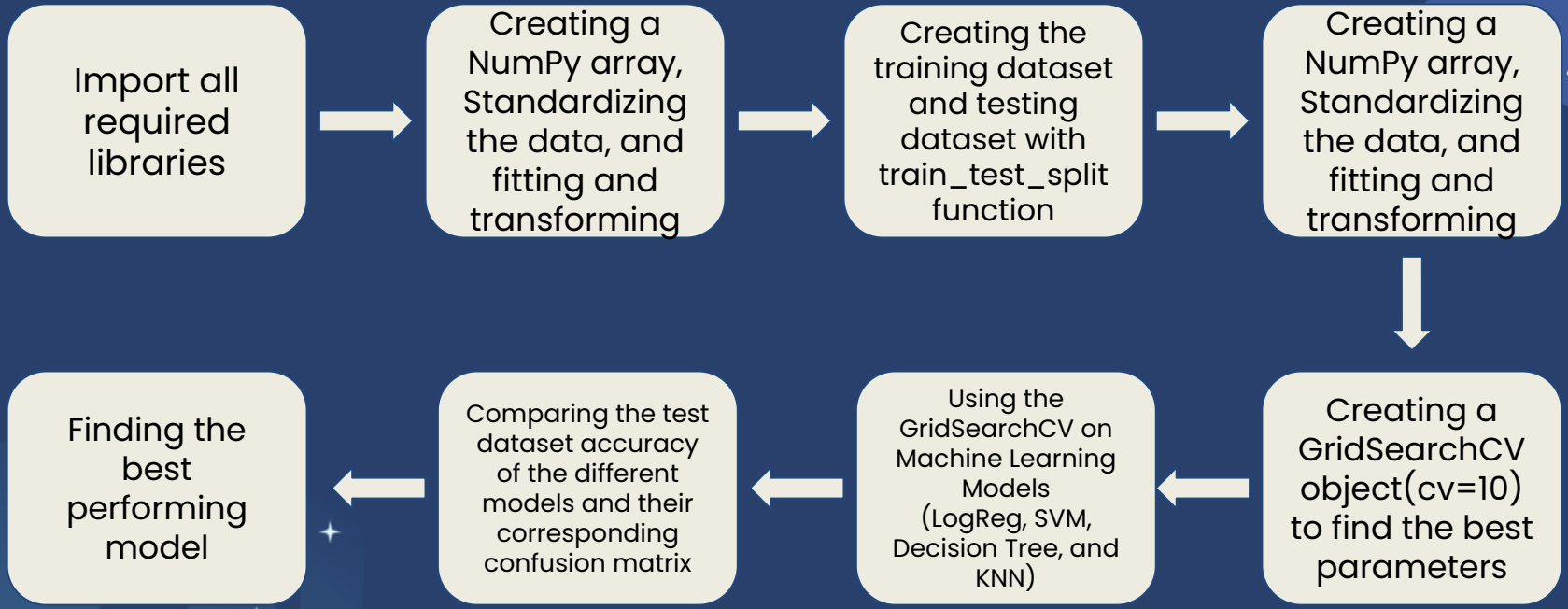
Build a Dashboard with Plotly Dash

4 Main Dashboard Elements

- Launch Sites Dropdown List:
  - Gives the option to choose which site data to present
- Pie Chart showing Success Launches (All Sites/Certain Site):
  - Shows the difference of the successful launch count and changes depending on the selected site.
- Slider of Payload Mass Range
  - Give the option to finetune the payload mass so sort the data shown in the scatter chart
- Scatter Chart of Payload Mass vs. Success Rate for the different Booster Versions
  - Shows the correlation between Payload and Launch Success

Github Code: 7 DASH_app.py

Predictive Analysis (Classification)



Github Code: [8 SpaceX_ Machine Learning Prediction]

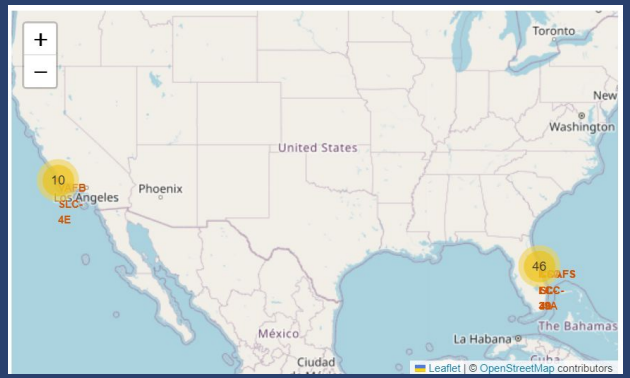
Results

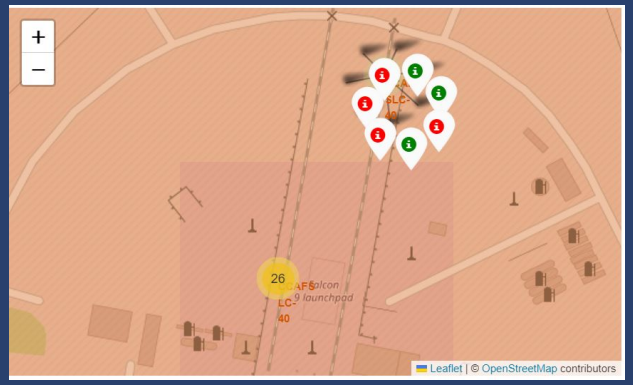
Exploratory data analysis results

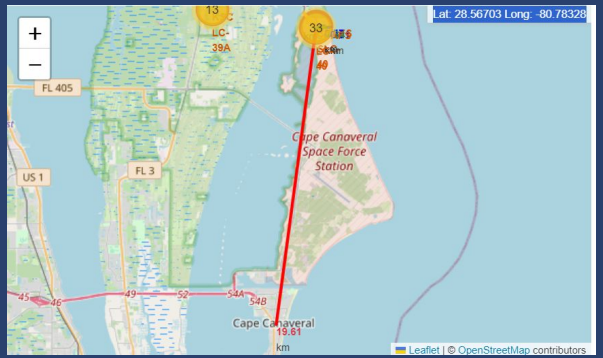
- There are 4 unique launch sites used by SpaceX Falcon 9 Rocket
- There was a collaboration between SpaceX and NASA on some of the launches
- Booster version f9 v1.1 has an average payload of 2534.66 kg
- The first successful landing was in 2015
- Falcon 9 has a high percentage approximately 98% of successfully landing
- Number of landings increased over time

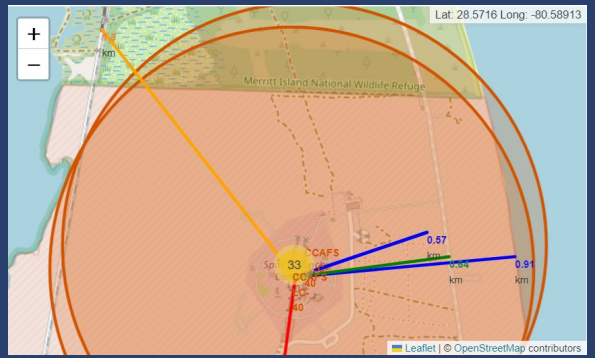
Results

Interactive analytics demo in screenshots









Results

Predictive analysis results

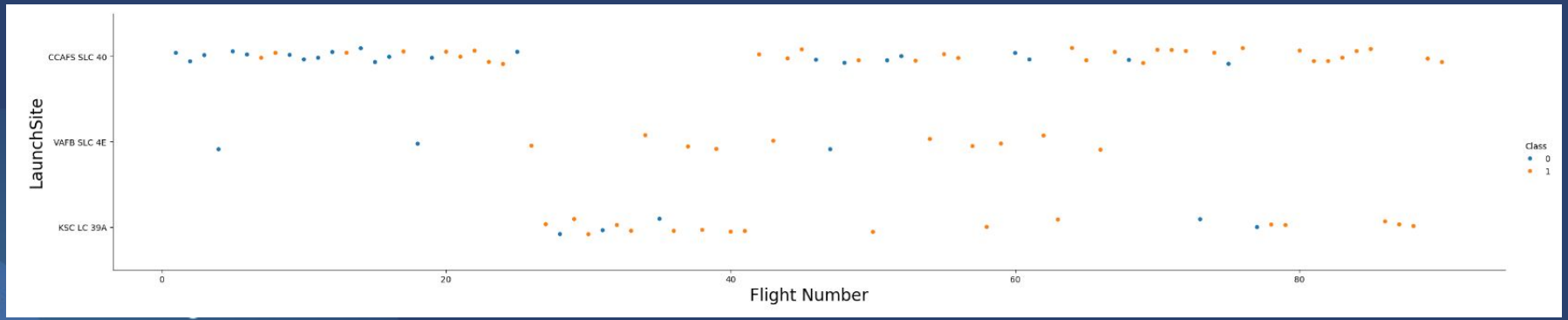
- Overall all Machine Learning Models are more or less equal in Test accuracy, Jaccard Score, and F1 Score but the Decision Tree Clasifier is the best available model because of its Best Score Accuracy

	LogReg	SVM	Tree	KNN
Test Accuracy	0.833333	0.833333	0.833333	0.833333
Best Score Accuracy	0.846429	0.848214	0.887500	0.848214
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889

Insights drawn from EDA

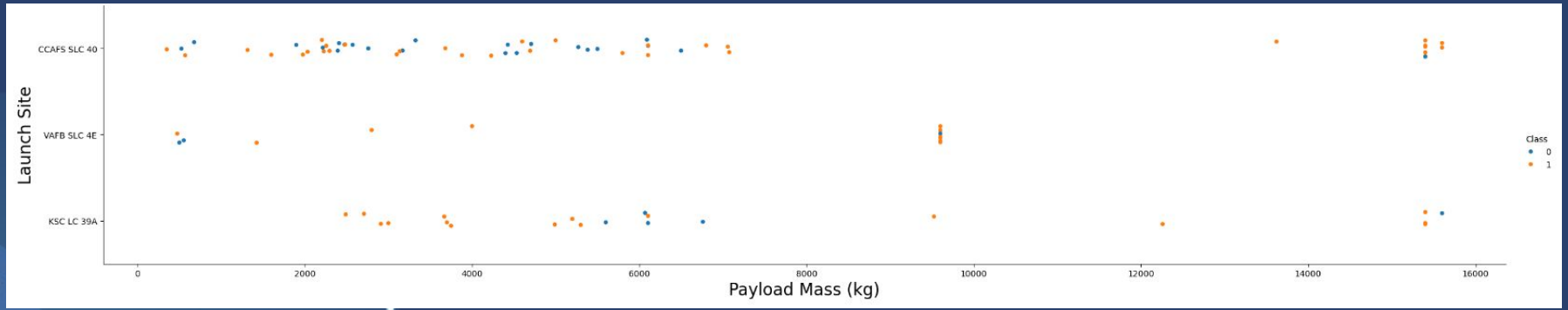
Flight Number vs. Launch Site

- Earliest Flights flights all failed but over time the number of successful flights steadily increased
- Most of the launches happened at CCAFS SLC 40
- While VAFB SLC 4E and KSC LC 39A had fewer launches, individually they have a higher success rate vs CCAFS SLC 40



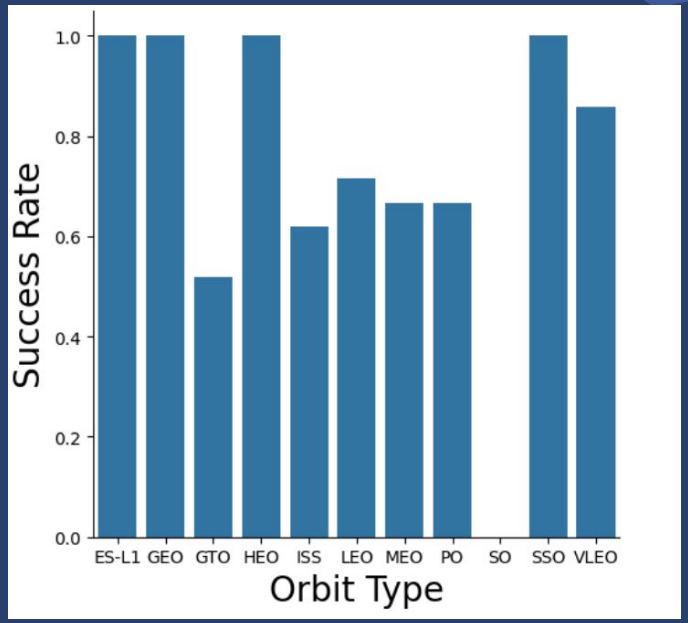
Payload vs. Launch Site

- Payloads above 7000 kg have a higher success rate compared to Payloads below 7000 kg
- KSC LC 39A retains a 100% success rate for Payloads under 5500 kg
- CCAFS SLC 40 and KSC LC 39A launch sites are the launch site used for Payloads above 10000 kg



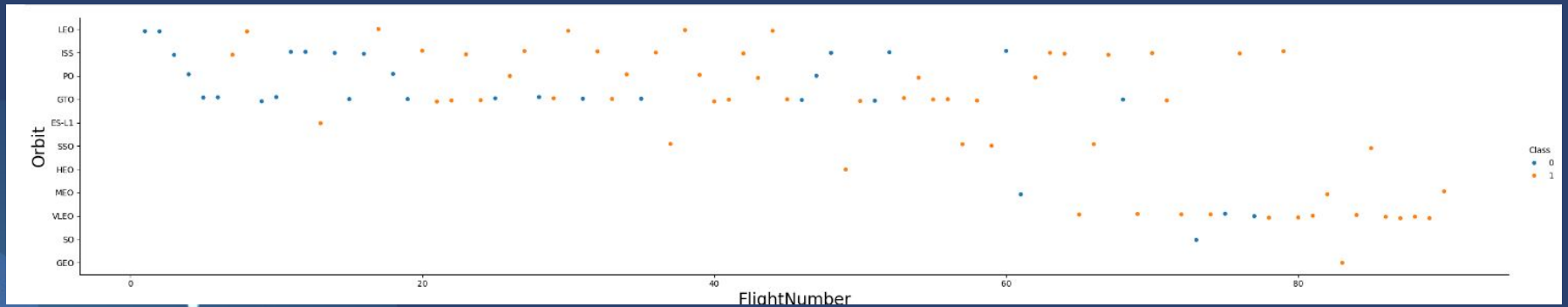
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, and SSO have a 100% success rate
- VLEO has a 85% success rate
- LEO has a 70% success rate
- MEO and PO have a 65% success rate
- ISS have a 60% success rate
- GTO has a 50% success rate
- SO has a 0% success rate



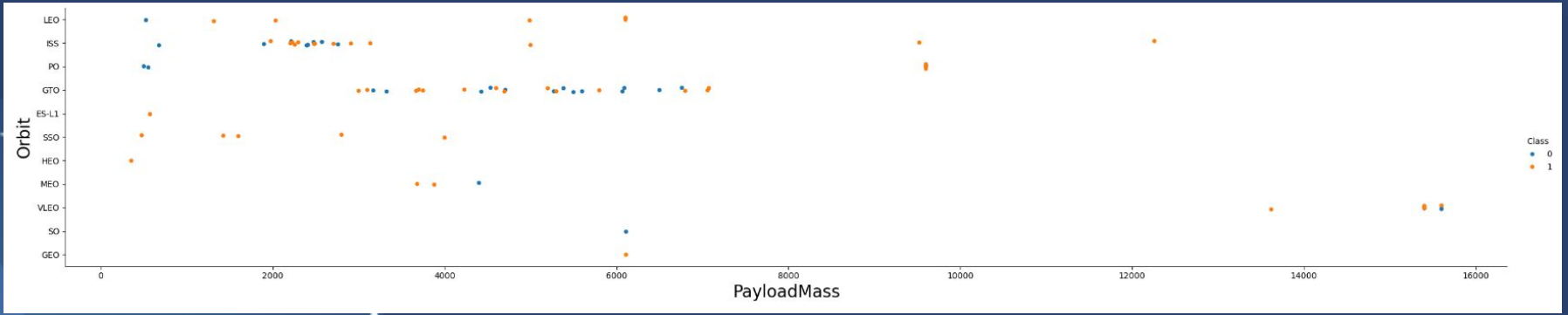
Flight Number vs. Orbit Type

- VLEO orbit seems to be most of the flights in the later stages
- Success seems to be related to the number of flights in LEO orbit
- In the GTO orbit, there appears to be no relationship between flight number and success



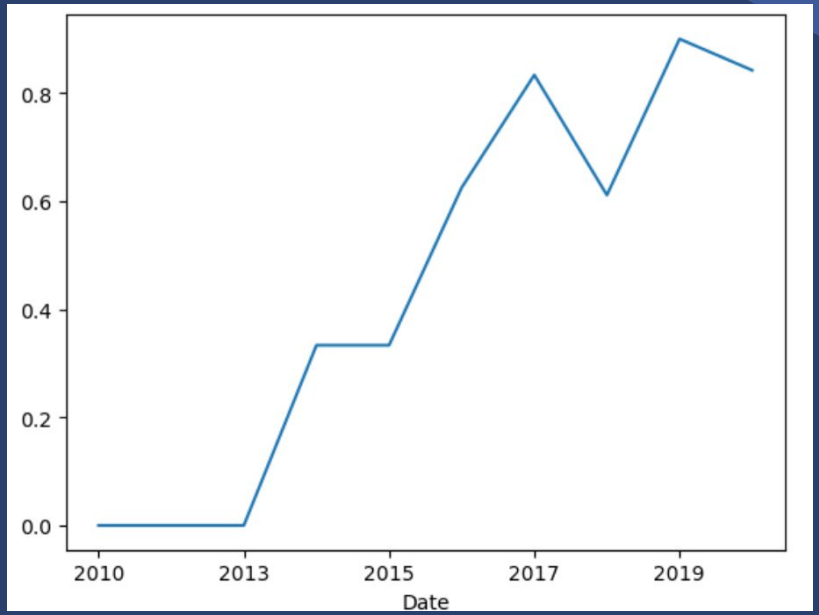
Payload vs. Orbit Type

- Heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- For GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present



Launch Success Yearly Trend

- Success rate started to increase in 2013 until 2017
- SpaceX was able to make a recovery after 2018 and continued to increase their success rate



All Launch Site Names

There are 4 distinct launch sites

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

[28]:
```
%%sql
    SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL ORDER BY 1;
```

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

[28]:

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

Launch Site Names Begin with '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

Samples of Cape Canaveral Launches

Total Payload Mass

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

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

TOTAL_PAYLOAD
111268

Total payload mass carried by boosters launched by NASA (CRS) is 111268 kg

Average Payload Mass by F9 v1.1

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

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

AVG_PAYLOAD
2534.6666666666665

Average payload mass carried

by booster version F9 v1.1

Is 2534.66 kg

First Successful Ground Landing Date

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

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

FIRST_SUCCESS_GP
2015-12-22

First successful landing outcome in ground pad was achieved in December 22, 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

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

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

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

The boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000 are 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(*) AS QTY FROM SPACEXTBL
    GROUP BY MISSION_OUTCOME ORDER BY MISSION_OUTCOME;
```

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

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

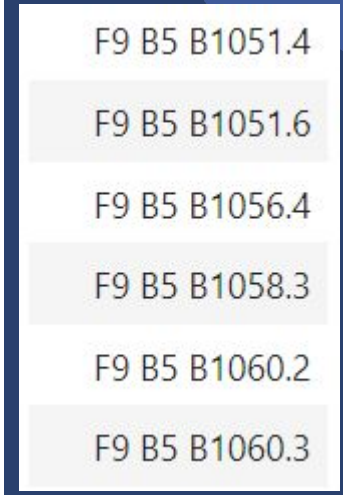
Total number of successful and failure mission outcomes

Boosters Carried Maximum Payload

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

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

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



The names of the booster versions which have carried the maximum payload mass

2015 Launch Records

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, DATE, 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	Date	Landing_Outcome	Booster_Version	Launch_Site
01	2015-01-10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	2015-04-14	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

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

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

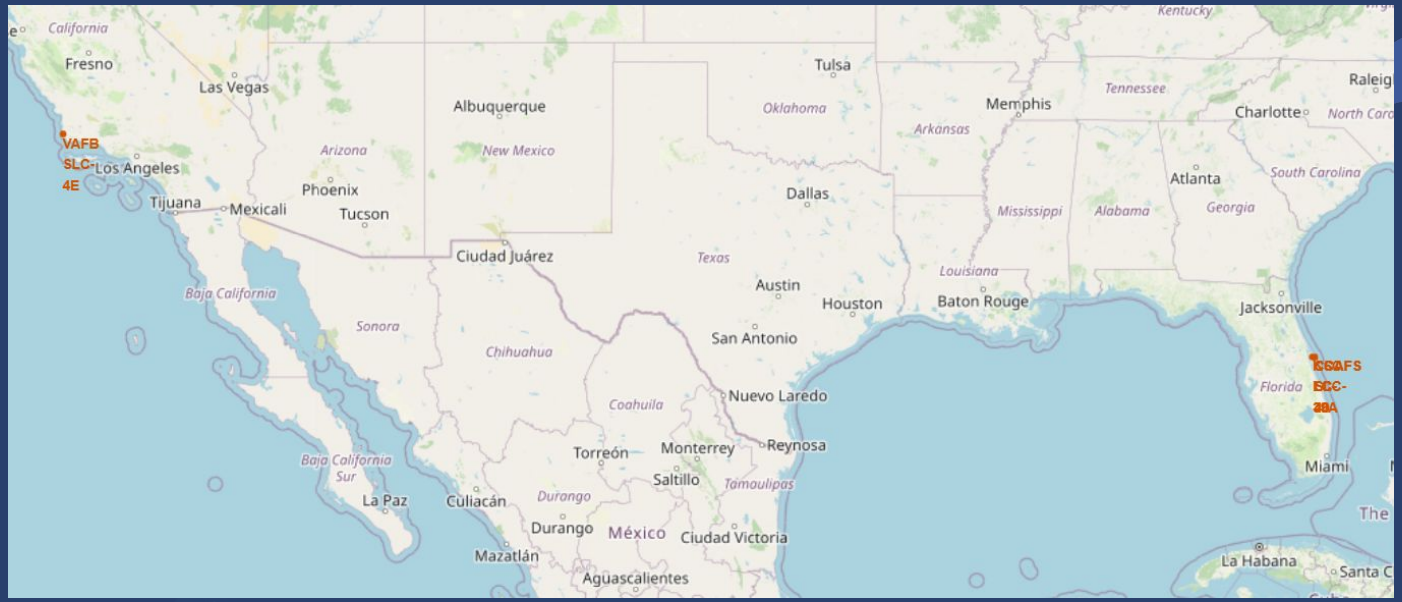
```
 * sqlite:///my_data1.db
Done.
```

Landing_Outcome	COUNT_OUTCOME
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

Ranking of 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.

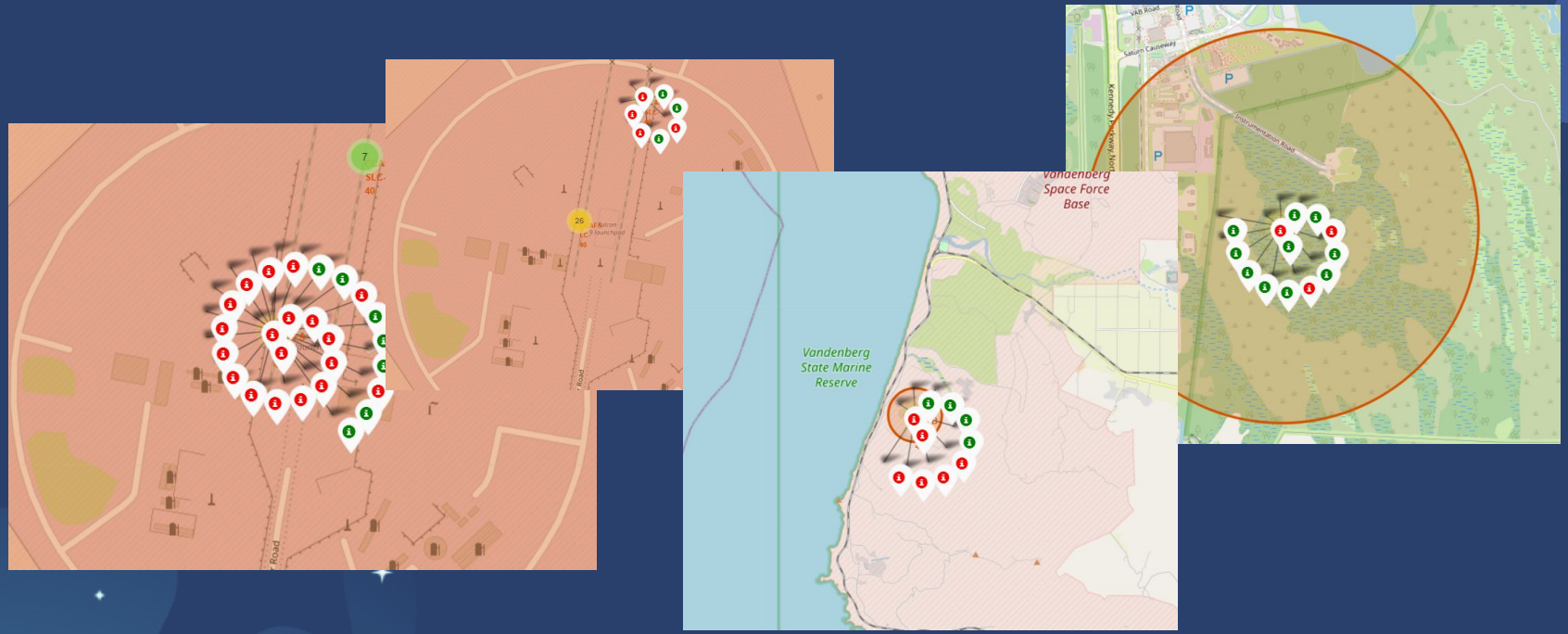
Launch Sites Proximities Analysis

Launch Sites



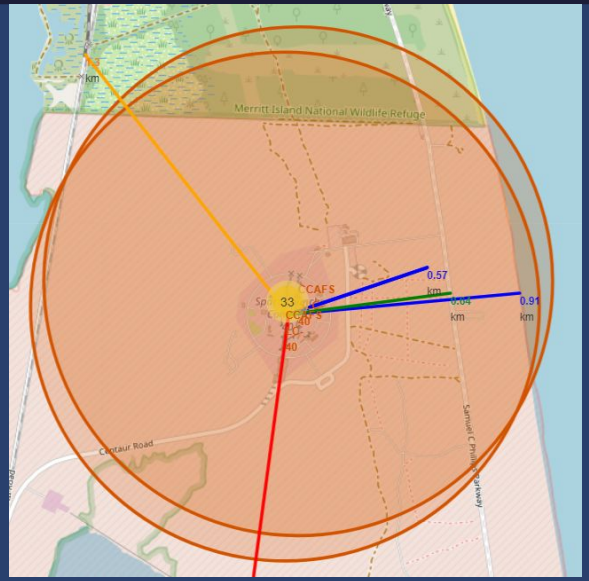
It is quite often that launch sites are close to the ocean/sea in order to minimize the damage in case of failure.

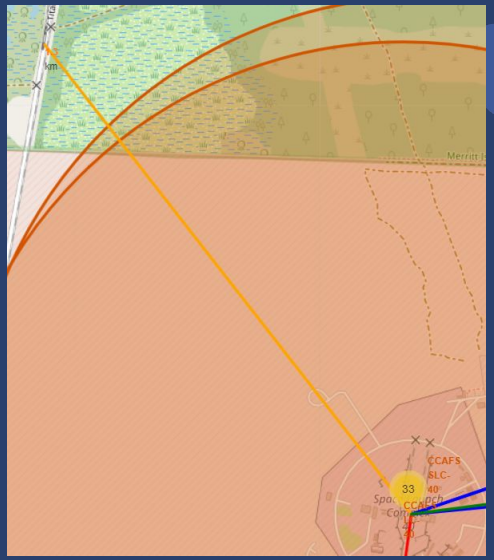
Launch Outcomes



Green Markers indicate a successful launch while red signifies failure

Examining Falcon 9 Launchpad

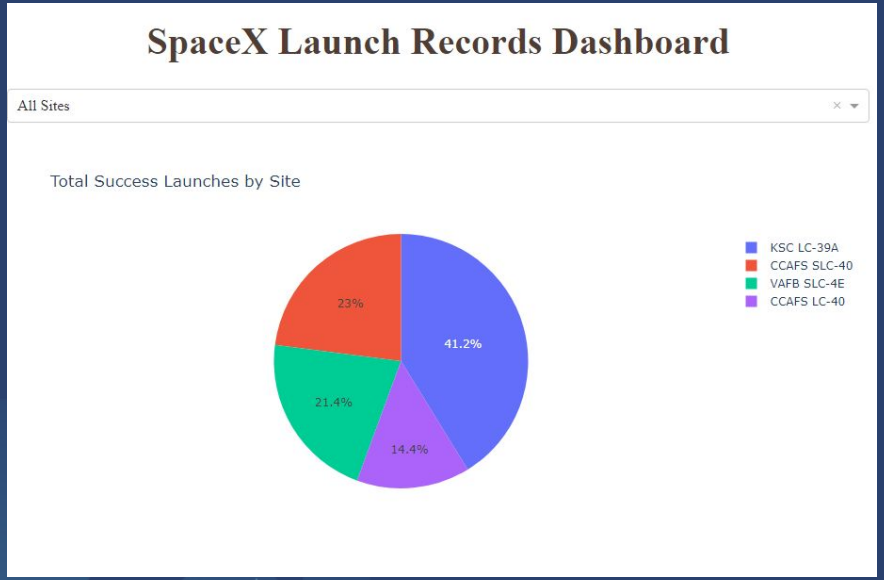




The area is very near the coast and is very accessible by on site railways and highways. It is also relatively far from inhabited areas.

Build a Dashboard with Plotly Dash

Total Successful Launches

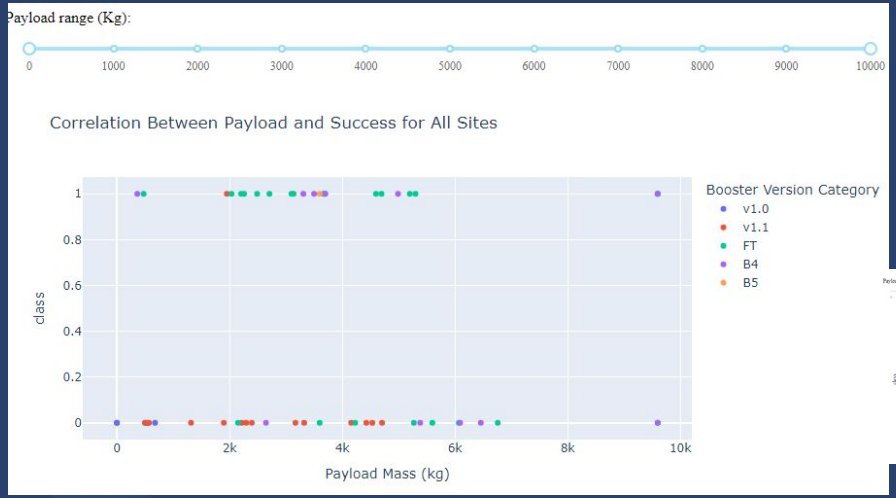


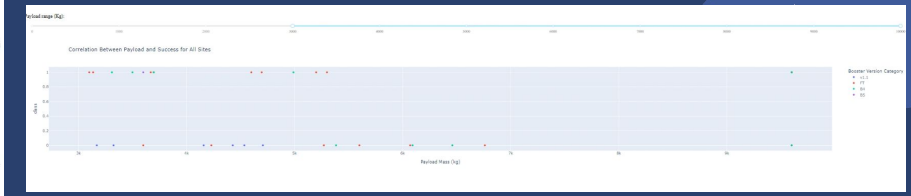
KSC LC-39A has the most successful launches.

Total Successful Launches

KSC LC-39A has the highest success rate with a rate of 76.9%

Payload vs. Launch Outcome



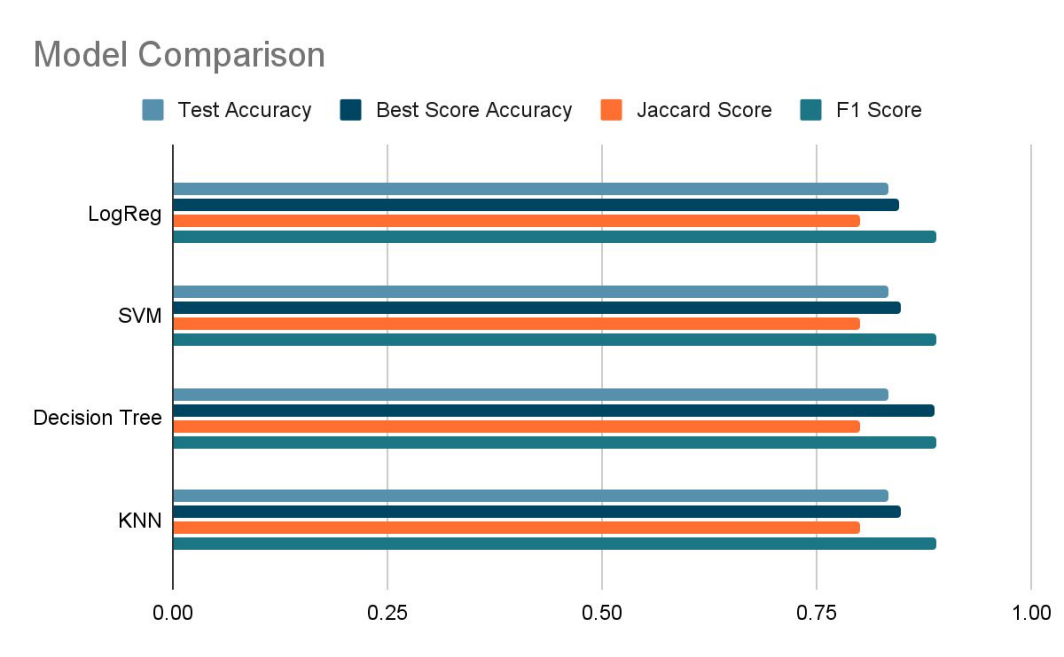


FT Boosters with a sub 6000 kg Payload have the highest success rate out of the other booster categories

Predictive Analysis (Classification)

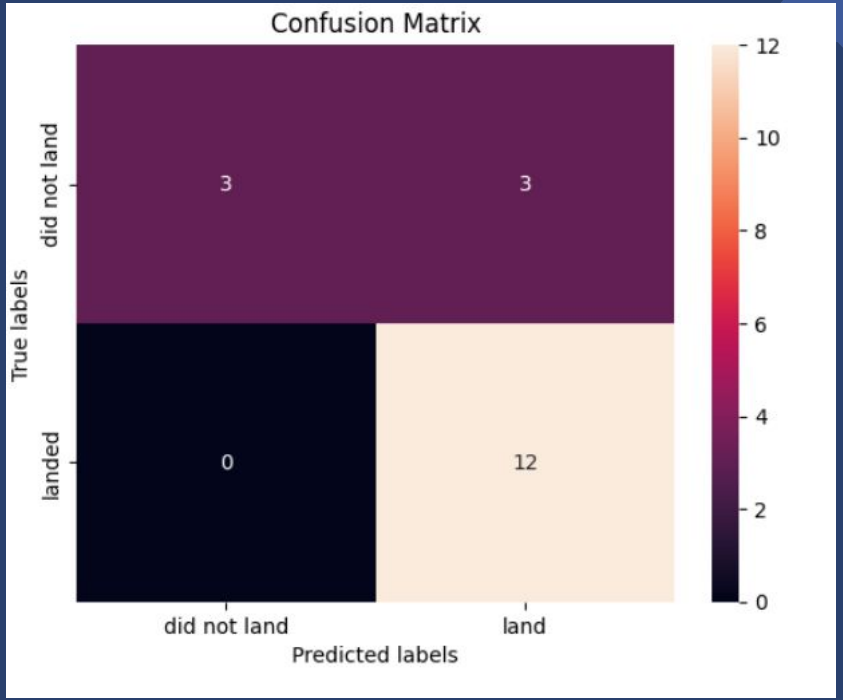
Classification Accuracy

Out of the four models Decision Tree Classifier has the highest accuracy with a score of 89%



Confusion Matrix

Decision Tree Classifier yielded 3 false positives but no false negatives



Conclusion

- Decision Tree Model is the best algorithm for the Predictive Analysis
- The high percentage of success of the Falcon 9 booster makes it a viable venture
- With an access to launchsite KSC LC-39A and using an FT booster with a sub 6000 kg Payload, should have the highest percentage of success
- Although there are failures, it seems that success rates improve over time

Appendix

Thank You