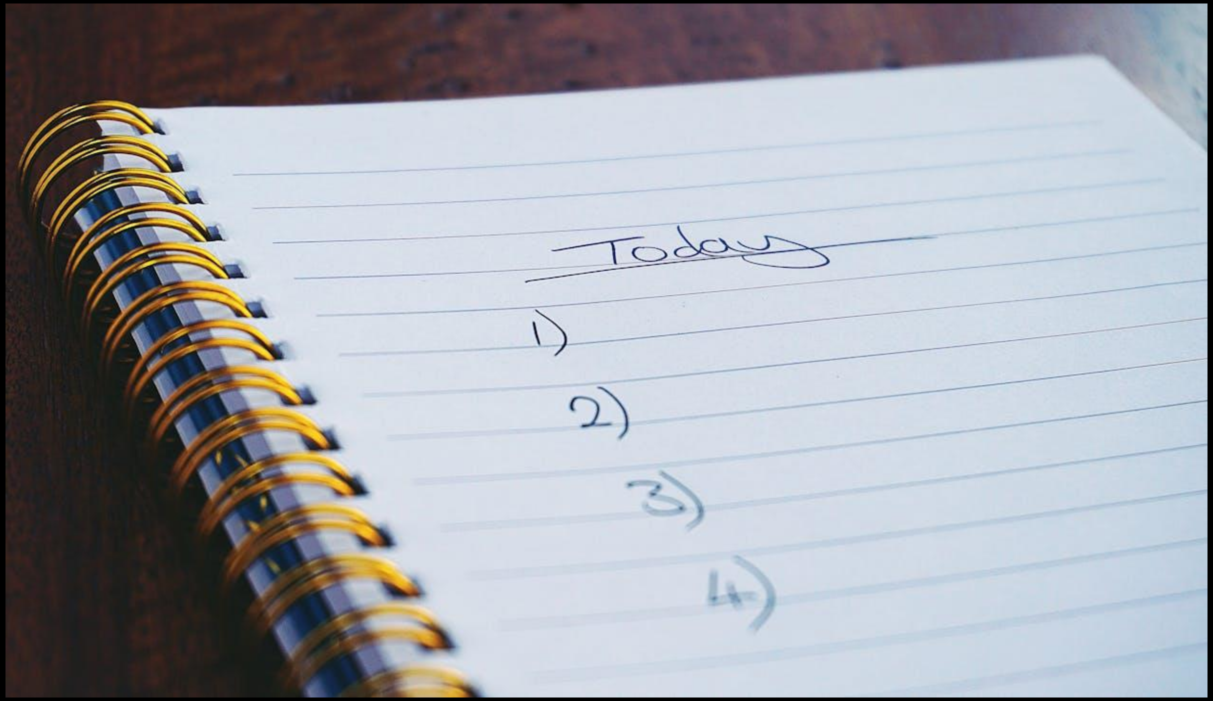
IBM Developer
SKILLS NETWORK

SPACEX FALCON 9 FIRST STAGE ANALYSIS

Anirudh Chillarige

09/03/24



Outline

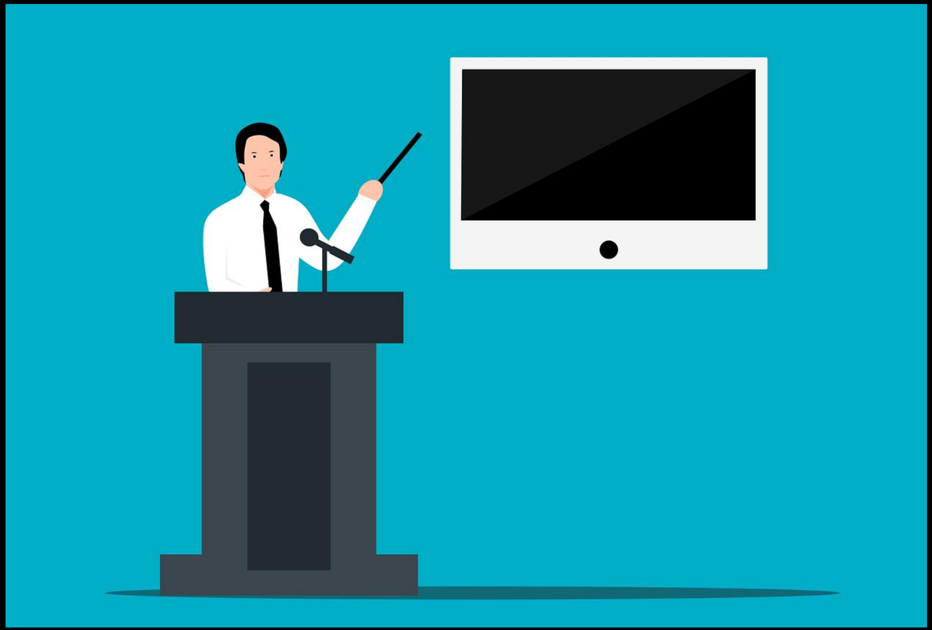
- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- This project utilizes SpaceX's Data API and Falcon 9's Wikipedia page to collect data on successful and failed first stage launches and the variables that played a part in that outcome. Using tools such as SQL, Data Visualization and Machine Learning the following was learned
- The variable with the highest impact on successful landing was Payload Mass, with most successful landing happening with a payload mass between 2000 to 6000 kg. The most successful landing site was KSC LC-39A with a 76.9% success rate. Lastly our machine learning algorithms outputted an R2 score of 0.83 and F1 score of about 0.89 displaying their successful ability to predict future landing outcomes.

Introduction

- SpaceX has a strong grip on space exploration and travel due to its sharp decrease in costs compared to its competitors. This decrease is a result of their ability to reuse the First Stage.
- This project assumes a new client wishes to replicate SpaceX's success. In order to do so it predicts the success rate of the first stage landing in an attempt to help determine the cost of a launch.



SECTION 1: METHODOLOGY

Methodology Overview

Data collection methodology:

- Data was collected through Falcon 9's Wikipedia page and SpaceX's data API

Performed data wrangling

- Data was processed to create a new variable based on landing outcomes that depicted a 1 in the case the landing outcome was successful and 0 when it failed

Performed exploratory data analysis (EDA) using visualization and SQL

Performed interactive visual analytics using Folium and Plotly Dash

Performed predictive analysis using classification models

- The following models were used: Logistic Regression, SVM, Decision Tree and KNN
- Each Model had gone through a hyperparameter optimization process to select its most efficient parameters

Data Collection - SpaceX API

Import libraries and use get request to connect and obtain raw data from SpaceX API

Decode response content using .json()

load into pandas dataframe using .json_normalize()

Remove unnecessary columns and use data ID's and premade functions to request launch data from API

Load launch data into a dictionary and use that to create a dataframe

Restrict data in the dataframe to only hold information about Falcon 9

Replace null values in Payload Mass with mean

Import dataframe into csv file to use for later analysis

[GitHub Link: Data Collection API]

Data Collection – Web Scraping

1. Import libraries and use get request to connect and obtain html response from Falcon 9 Wikipedia Page
2. Create a BeautifulSoup object from the html response and load all the data tables from the wiki page into a list
3. Select the third table from the list and extract column names
4. Create an empty dictionary with the column names as the keys
5. Iterate through the table rows and load the data into the dictionary using the premade functions
6. Use the data filled dictionary to create a pandas dataframe
7. Import dataframe into csv file to use for later analysis

GitHub Link: [Data Collection Web Scraping]()

Data Wrangling

Data Formatting

Look for null values, and incorrect data types within each column of the data set

Using .value_counts() calculate the number and occurrence of each Launch Site, Orbit, and Landing Outcome

Create a new column called landing class within the data set that outputs a 1 if the landing outcome was successful and a 0 if it failed

Calculate the mean of this new column to find the success rate of all the missions

After ensuring the data is ready further analysis save it so a csv file

Data Wrangling is an essential part of any data science project

In order for further analysis to result in accurate insights, data must be formatted in a way that does not skew or misinterpret information.

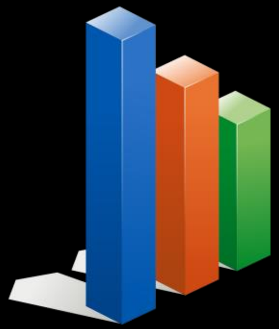
In this scenario the column "landing class" was formatted into a binary class that would later allow Machine Learning Algorithms to understand it.

[GitHub Link: Data Wrangling]

EDA with Data Visualization

Types Of Graphs:

Bar plots: Great at comparing categorical values with a numerical value.



Bar Plots:
- Orbit Type vs Success Rate

Scatter plots: Excel at comparing two independent variables with each other and can show possible correlation

Scatter Plots:
- Flight Number vs Payload Mass
- Flight Number vs Launch Site
- Payload Mass vs Launch Site
- Flight Number vs Orbit Type
- Payload Mass vs Orbit Type

Line plots: Help show trends in data a period of time

Line Plots:
- Date vs Success Rate

[GitHub Link: Data Visualization]

EDA with SQL

SQL Queries:

- Load data set into pandas data frame then used dataframe.to_sql() to create SQL Table
- Displayed Distinct Launch Sites
- Displayed 5 Launch Sites whose name begin with CCA
- Displayed Total Payload Mass carried by boosters launched by NASA(CRS)
- Displayed Average Payload Mass carried by booster version F9 v1.1
- Displayed the date in which the first successful ground pad landing occurred
- Displayed the names of the boosters which successful land in drone ships and have a payload mass greater than 4000 kg but less than 6000
- Displayed the total number of Successful and Failure mission Outcomes
- Using a Subquery displayed the names of all booster versions which have carried the maximum payload mass
- Displayed the month of failed landing outcomes in drone ships, booster versions, and launch site in the year 2015
- Displayed the different types of landing outcomes between 2010-06-04 and 2017-03-20 and the number of times they occurred in descending order

[GitHub Link: EDA with SQL]()

Interactive Map with Folium

Main Map

- Centered at Nasa Johnson Space Center (JSC)

Markers

- Marker positioned on NASA headquarters that displays NASA JSC in red
- 4 Markers positioned on each rocket launch site that display their name in black

Circles

- Red Circle positioned on NASA headquarters that shows displays a popup with its name when clicked
- 4 different colored circles positioned on each rocket launch site that display their names as a popup when clicked

Clusters

- Marker clusters positioned on each rocked launch site that display the total number of landings
- When zoomed in and clicked on a spiral of clusters is displayed with a green marker for each successful landing and a red marker for each failed landing

Distance Lines

- A line that connects launch site VAFB SLC-4E to the coastline and shows its distance in km
    - When hovered over a popup is shown that displays the lines purpose
- A line that connected launch site KSC LC-39A with the city of Orlando, Florida and shows its distance in km
    - When hovered over a popup is shown that displays what the line is connecting

Mouse Position

- An add-on that displays the current latitude and longitude of your cursor as it hovers over the map

[GitHub Link: SpaceX Folium Map]

Build a Dashboard with Plotly Dash

Dashboard Outline

Title: SpaceX Launch Records Dashboard

- Title is bolded and centered

Equipped with a Dropdown menu

- Allows you to choose All Launch Site statistics or an individual Launch Site's statistics

Dashboard Visualizations

Pie Charts

- The All Sites dropdown shows a pie chart that compares the number of successful landings between all the launch sites
- Each individual site dropdown shows a pie chart that compares is landing success rate

Scatter Plot and Range Slider

- A scatter plot is shown that compares the Payload Mass (x-axis) and whether or not the launch successfully landed (y-axis)
    - The points are colored based on Booster Version Category
- The slider allows you to control the Payload Mass range

[GitHub Link: Interactive Dashboard Analysis]

Predictive Analysis (Classification)

Create a NumPy array that holds the target "Class" column's data and a dataframe X that holds the independent attributes data

Standardize X data using .StandardScalar() then proceed to split data into 80% train and 20% test using train_test_split()

Create Logistic Regression Model

Create Support Vector Machine Model

Create Decision Tree Classifier Model

Create K-Nearest-Neighbors Model

Create a dictionary for each model with multiple hyperparameters then create a GridSearchCV object to find the best possible parameters

Find R^2 Score for each Machine Learning model and create a Confusion Matrix
Compare each models score and select the highest performing algorithm

[GitHub Link: Machine Learning Prediction Algorithms]

Results

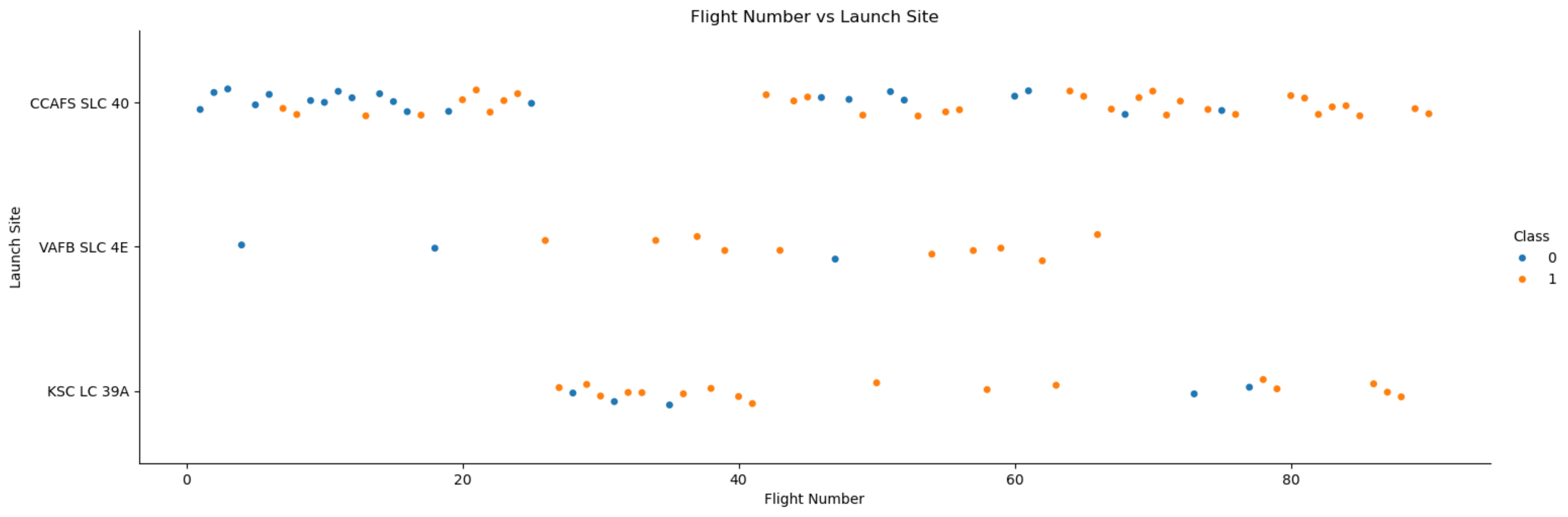
- Exploratory data analysis results
- Interactive analytics demo
- Predictive analysis results

SECTION 2:
EXPLORATORY DATA ANALYSIS RESULTS

Exploratory data analysis results:
Data Visualization

Flight Number vs. Launch Site

The class is used to represent if the landing was successful or not.
1 (Orange) for Success and 0 (Blue) for Failure

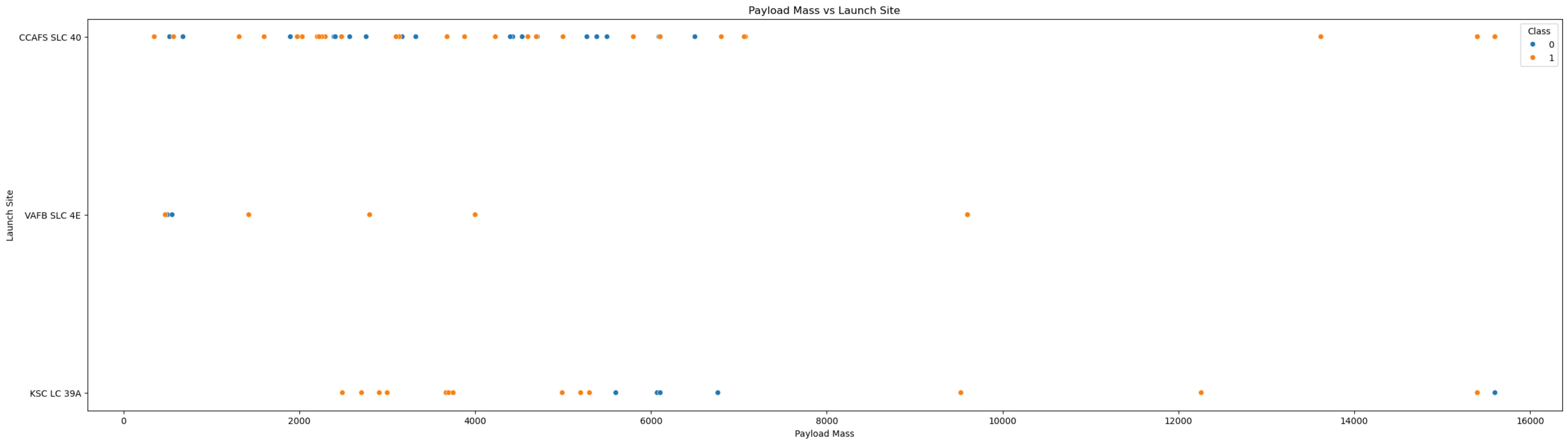


Insight:

- CCAFS SLC 40 has the highest amount of launches
- Success rate increases as flight number increases

Payload Mass vs. Launch Site

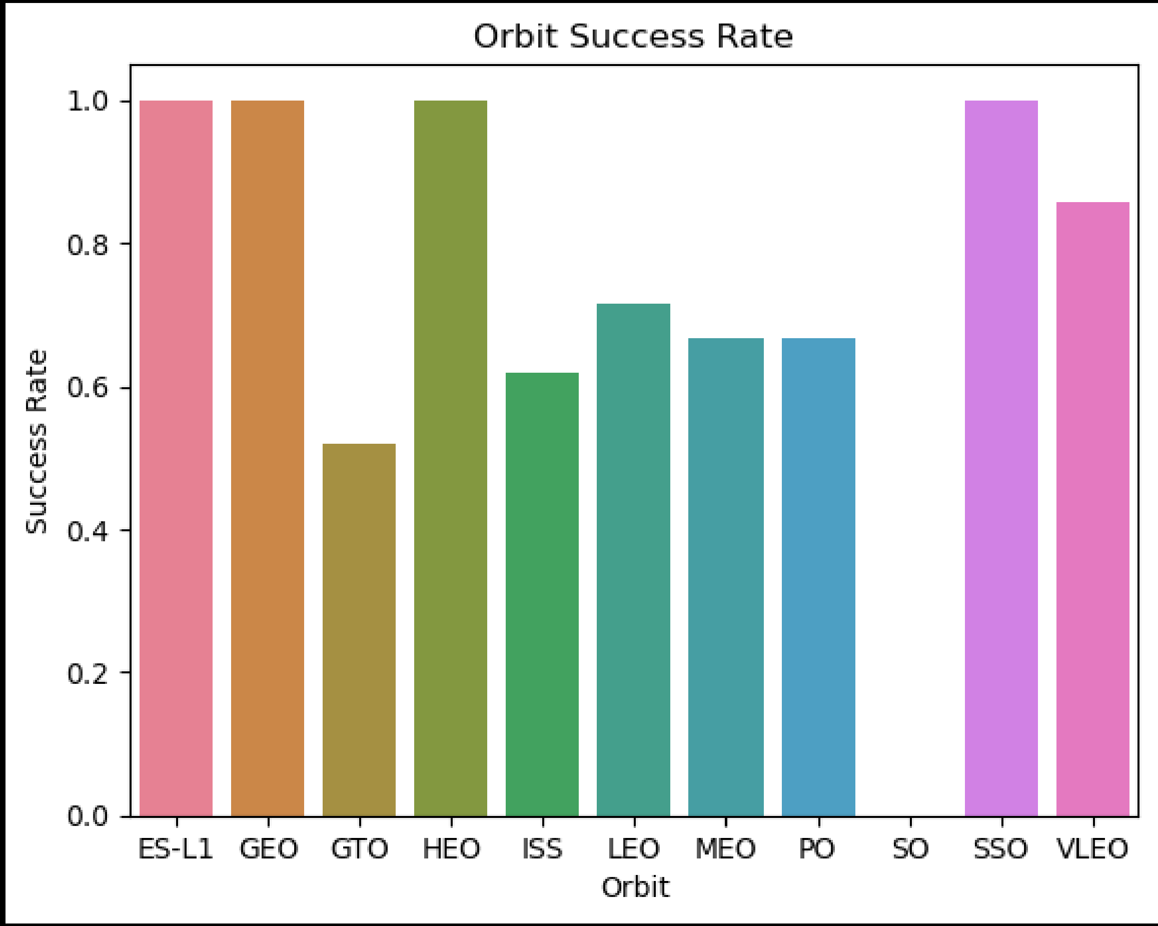
The class is used to represent if the landing was successful or not.
1 (Orange) for Success and 0 (Blue) for Failure



Insight:

- VAFB SLC 4E has no launches for rockets with a payload mass greater than 10000
- Majority of launches with a high payload mass had a higher success rate than those with low payload mass

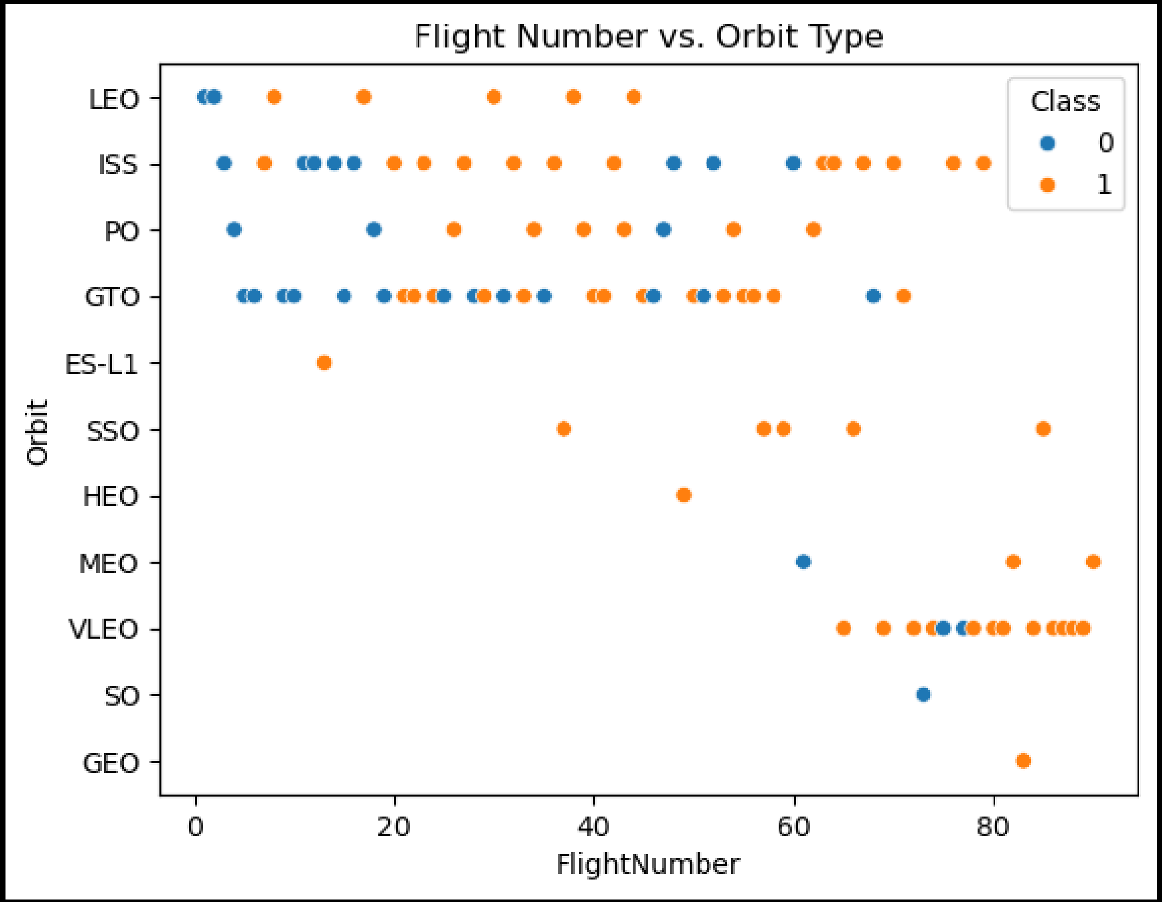
Success Rate vs. Orbit Type



Insights:

- Orbits ES-L1, GEO, HEO, and SSO have a 100% success rate
- Orbit VLEO has the second highest success rate
- Orbit SO has a 0% success rate
- Orbit GTO has the second lowest success rate

Flight Number vs. Orbit Type



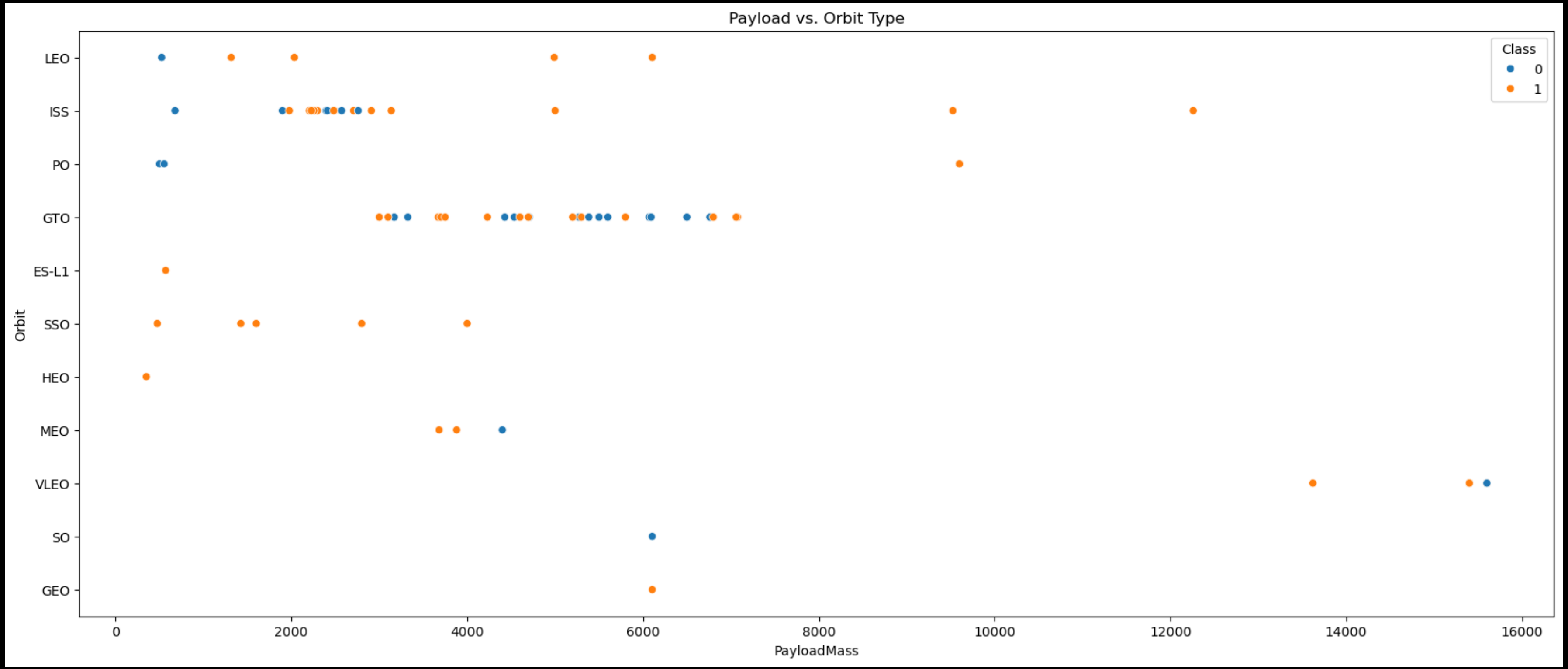
The class is used to represent if the landing was successful or not.
1 (Orange) for Success and 0 (Blue) for Failure

Insights:

- Orbit VLEO only has launches with a flight number above 60
- Orbits ES-L1, GEO, and HEO only have a single launch
    - All of which are successful
- Orbit LEO has a 100% success rate after its first two failures
- Orbits ISS, GTO, and VLEO have the most launches
- Orbit GTO's success rate has no relationship with flight number

Payload vs. Orbit Type

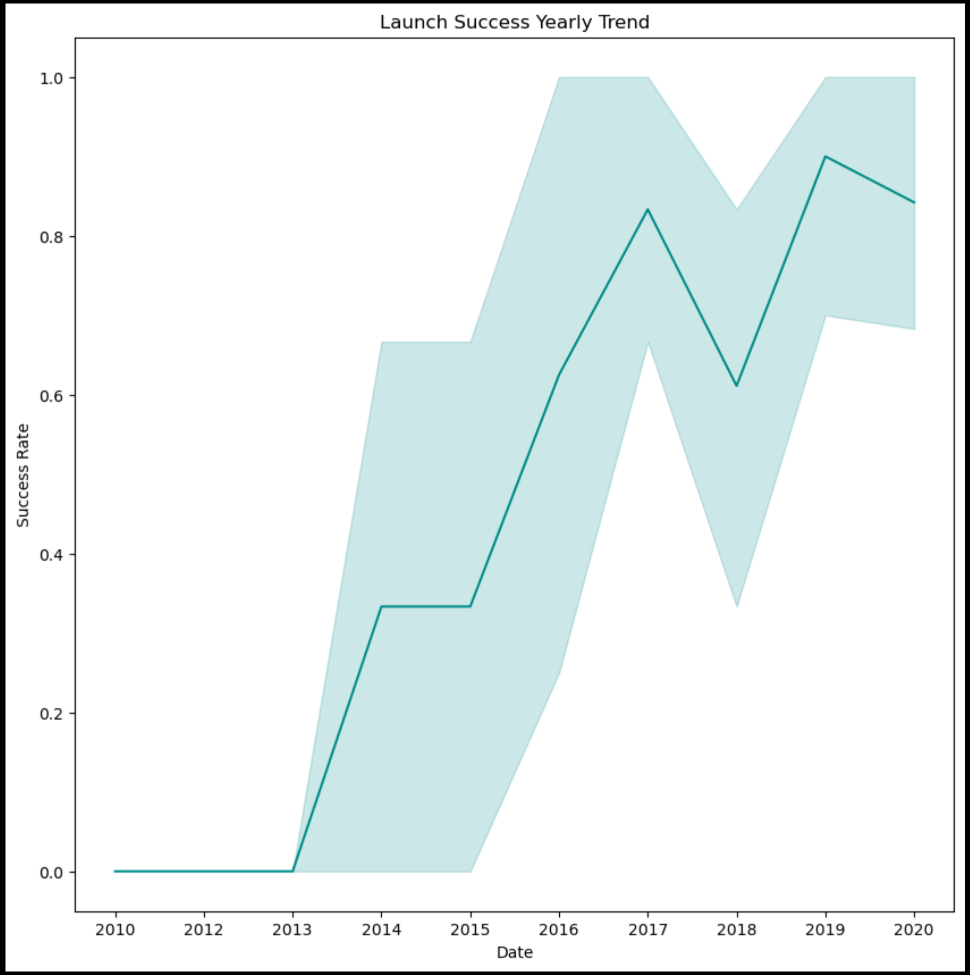
The class is used to represent if the landing was successful or not.
1 (Orange) for Success and 0 (Blue) for Failure



Insight:

- Orbit GTO only has launches with a payload mass between 2000 kg and 8000 kg
- Orbit BLEO only has launches with payload mass above 13000 kg
- Orbits PO, LEO, and ISS have a strong positive relationship with payload mass

Launch Success Yearly Trend



Insights:

- Success rate had periods of sharp increases between 2013-2014, 2015–2017, and 2018-2019
- Success rate had periods of stagnation between 2010–2013 and 2014-2015
- Success rate had periods of decline between 2017–2018, and 2019–2020
- Overall success rate has increased drastically since 2013

Exploratory data analysis results:
SQL

All Launch Site Names

```
%sql select distinct(Launch_Site) from SPACEXTBL ;
```

Running query in 'sqlite:///my_data1.db'

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

Search query that displays each unique launch site name

Launch Site Names Begin with 'CCA'

```
%sql select * from SPACEXTBL where Launch_Site like ("%CCA%") limit 5;
```

Running query in 'sqlite:///my_data1.db'

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

Search query that displays 5 records where Launch Site includes "CCA"

Total Payload Mass

```
%sql select Customer, sum(PAYLOAD_MASS__KG_) as Total_Payload_Mass from SPACEXTBL where Customer like '%NASA (CRS)%';
```

Running query in 'sqlite:///my_data1.db'

Customer	Total_Payload_Mass
NASA (CRS)	48213

Search query that displays the sum of NASA (CRS)'s payload mass over all its launches

Average Payload Mass by F9 v1.1

```
%sql select Booster_Version, avg(PAYLOAD_MASS__KG_) as Average_Payload_Mass from SPACEXTBL where Booster_Version like ("%F9 v1.1%");
```

Running query in 'sqlite:///my_data1.db'

Booster_Version	Average_Payload_Mass
F9 v1.1 B1003	2534.6666666666665

Search query that displays Booster Version F9 v1.1's average payload mass

First Successful Ground Landing Date

```
%%sql
select min(DATE(Date)) as First_Succesful_Landing_Date from SPACEXTBL
where Landing_Outcome like ("%Success%") and Landing_Outcome like ("%ground pad%");
```

Running query in 'sqlite:///my_data1.db'

First_Succesful_Landing_Date
2015-12-22

Search query that displays the date in which the first successful ground pad landing occurred

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql
select Distinct(Booster_Version), PAYLOAD_MASS__KG_, Landing_Outcome from SPACEXTBL
where Landing_Outcome like ("%Success (drone ship)%") and PAYLOAD_MASS__KG_ between 4000 and 6000;
```

Running query in 'sqlite:///my_data1.db'

Booster_Version	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Search query that displays all successful drone ship landing with a payload mass between 4000 and 6000 kg

Total Number of Successful and Failure Mission Outcomes

```
%sql select Mission_Outcome, count(Mission_Outcome) from SPACEXTBL group by Mission_Outcome;
```

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

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

Search query that displays all possible Mission Outcomes and the number of times they occurred

Boosters Carried Maximum Payload

```
%sql select Booster_Version, PAYLOAD_MASS__KG_ from SPACEXTBL where PAYLOAD_MASS__KG_ == (select max(PAYLOAD_MASS__KG_) from SPACEXTBL) ;
```

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

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

Search query that displays all Booster Versions that carried the maximum payload

2015 Launch Records

```
%%sql

select substr(Date,6,2) as Month_in_2015, Landing_Outcome, Booster_Version, Launch_Site from SPACEXTBL
where Landing_Outcome like ("%Failure%") and Landing_Outcome like ("%drone ship%") and Date like ('%2015%') ;
```

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

Month_in_2015	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

Search query that displays all failed drone ship landings in 2015

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

```
%%sql
select Landing_Outcome, count(Landing_Outcome) as Count from SPACEXTBL
where Date between '2010-06-04' and '2017-03-20' group by Landing_Outcome order by Count DESC;
```

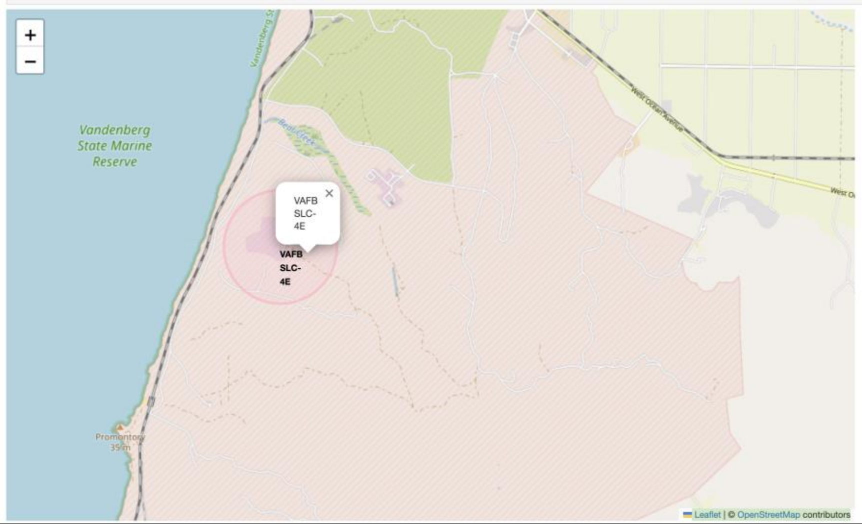
```
 * sqlite:///my_data1.db
Done.
```

Landing_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

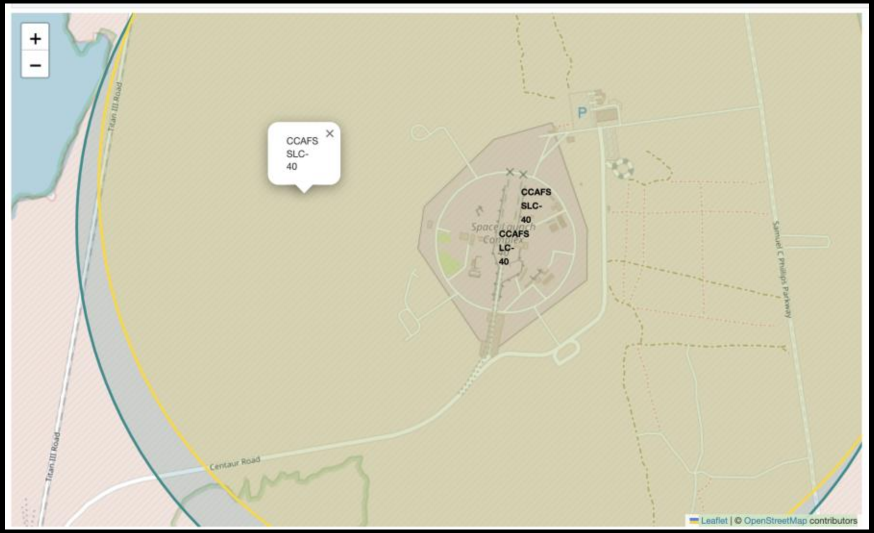
Search query that displays each possible landing outcome and the number of times they occurred

SECTION 3:
INTERACTIVE ANALYTICS DEMO

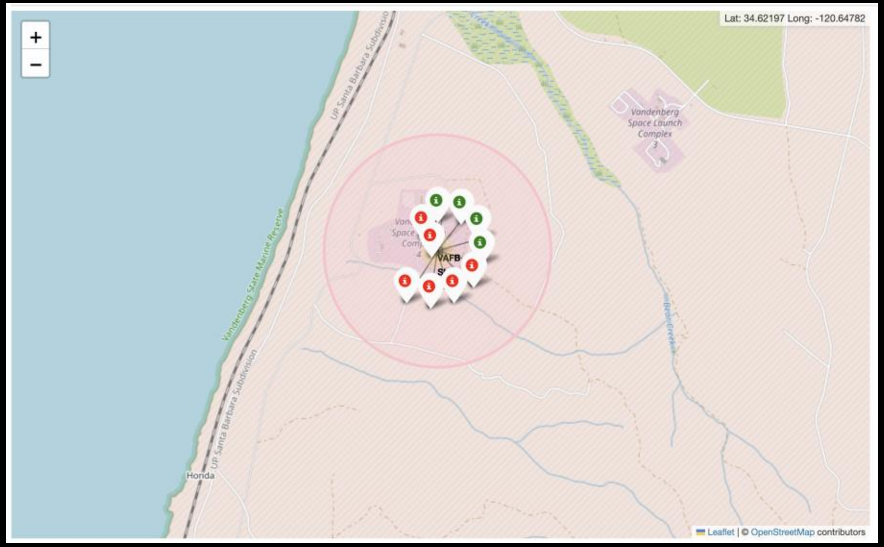
Folium Map - Launch Sites

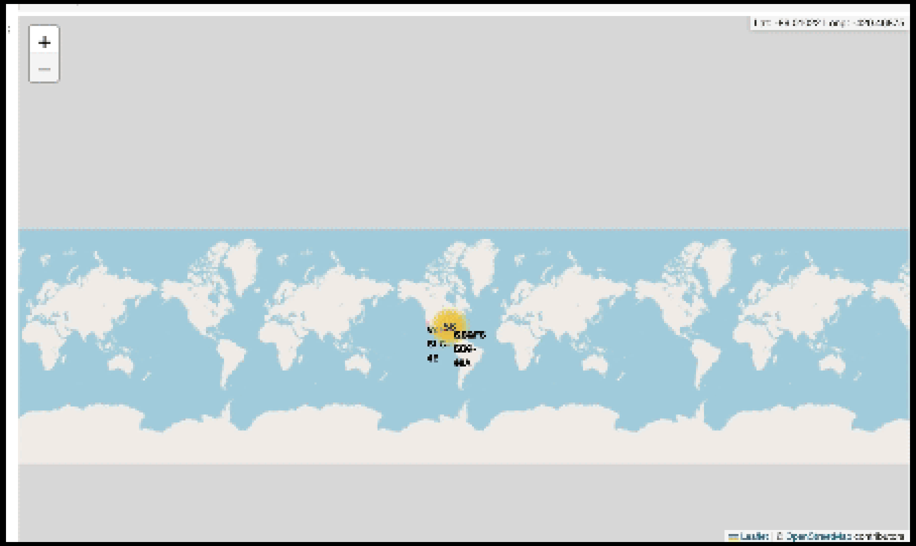


An interactive map displaying the 4 rocket launch sites



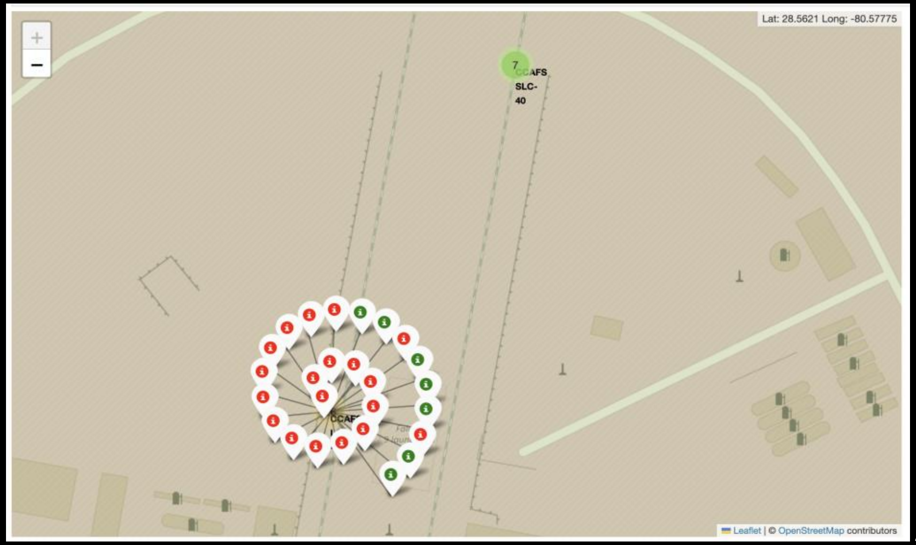
Folium Map – Launch Outcome



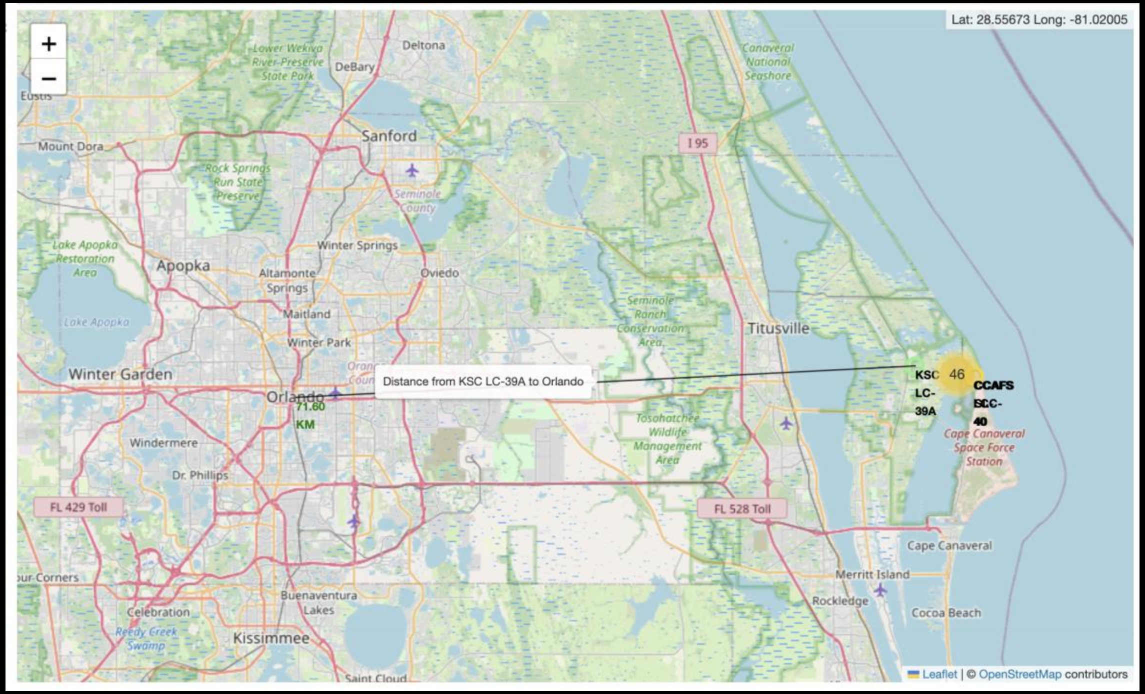


An interactive map displaying the total number of successful and failed launch outcomes.

Successful launches are green while failed launches are red



Folium Map – Distance Lines



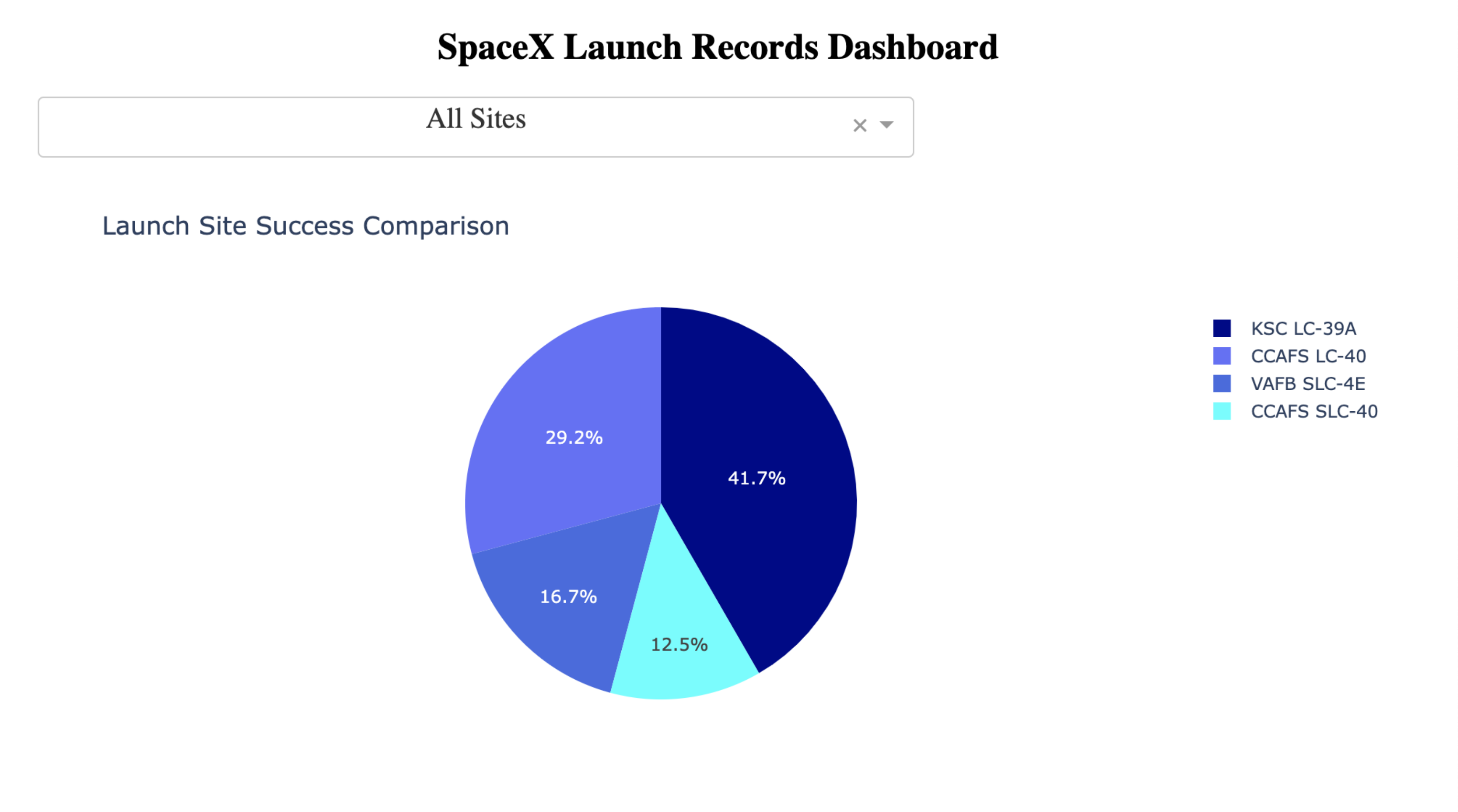
An interactive map displaying the distance between:

- Launch Site VAFB SLC-4E and its nearest Coastline
- Launch Site KSC LC-39A and the city of Orlando, Florida

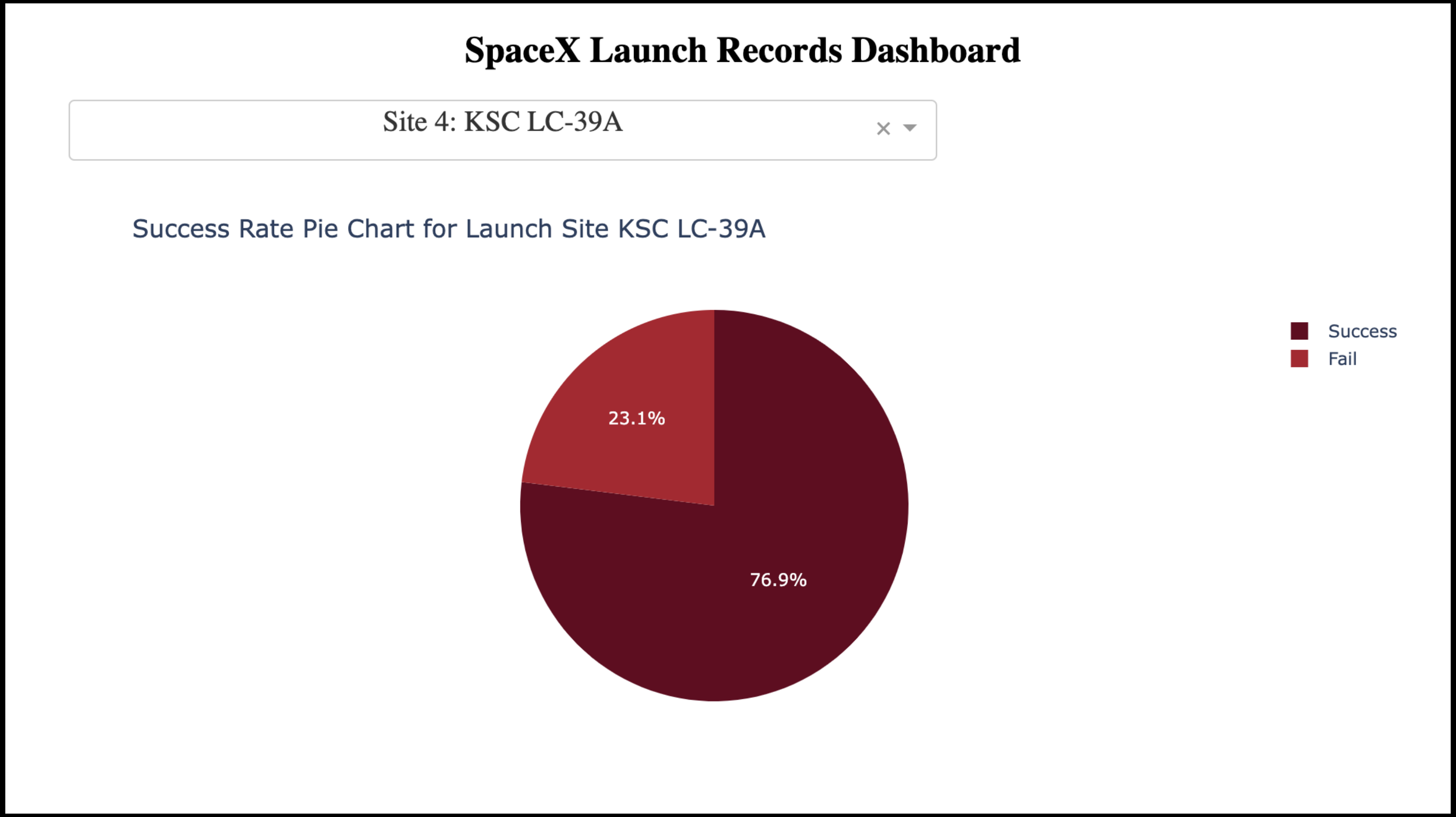
Dash – Launch Site Success Rates

An interactive Dashboard with a dropdown menu that displays different Launch site statistics

This section compares the success rate of each launch site in a Pie Chart

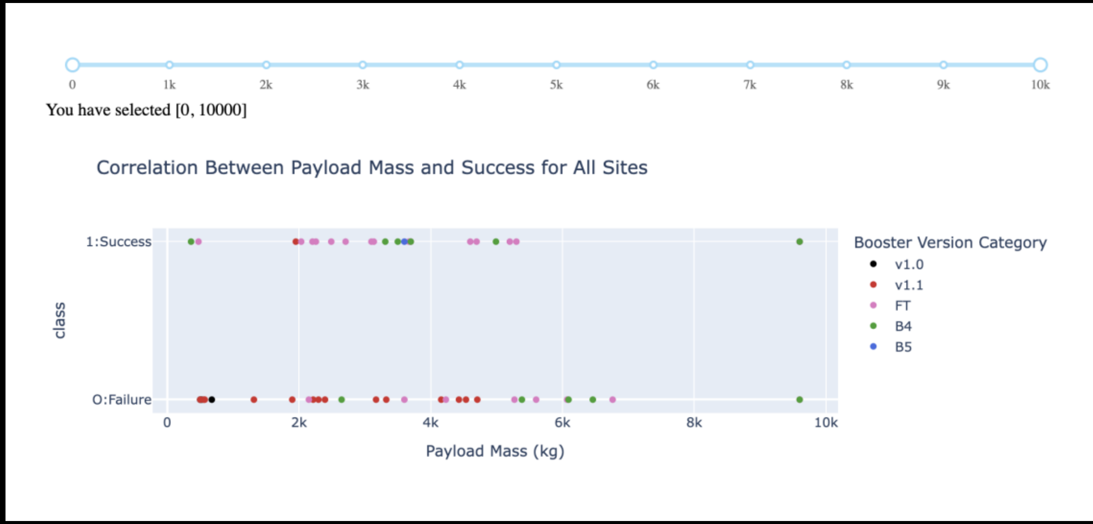


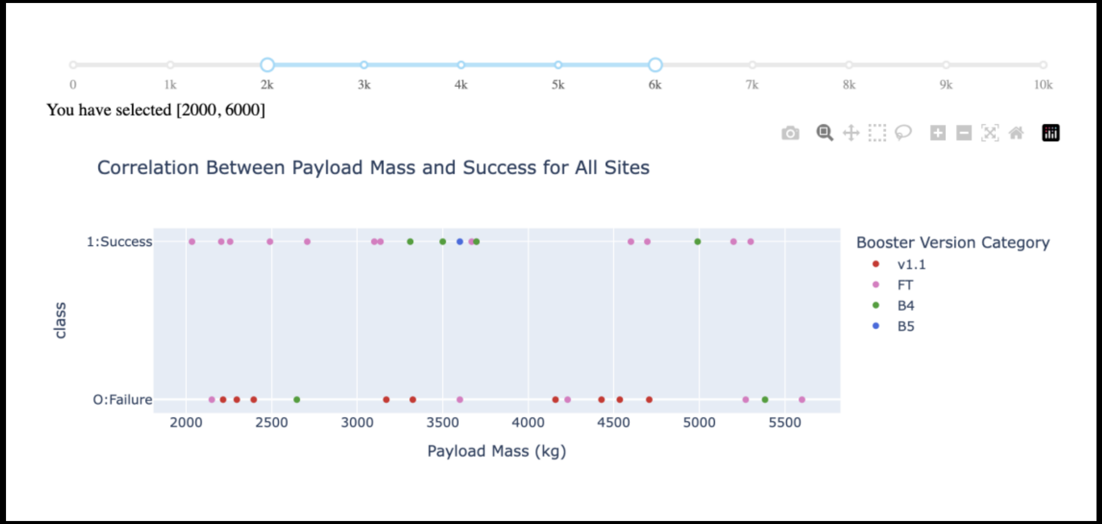
Dash – Highest Launch Site Success Rate



Compared to the other 3 launch sites, Launch Site KSC LC 39-A boasts the highest launch site success rate at 76.9%

Dash – Payload Mass vs Success Rate Scatter Plot





A Scatter Plot that helps show the correlation between payload mass and success rate

- Booster Version Category types are separated by color
- The range slider above the graph allows you to customize the range of the x-axis

Most successful launches seem to be with a payload mass between 2k to 6k kg

SECTION 4:
PREDICTIVE ANALYSIS RESULTS

Classification Accuracy

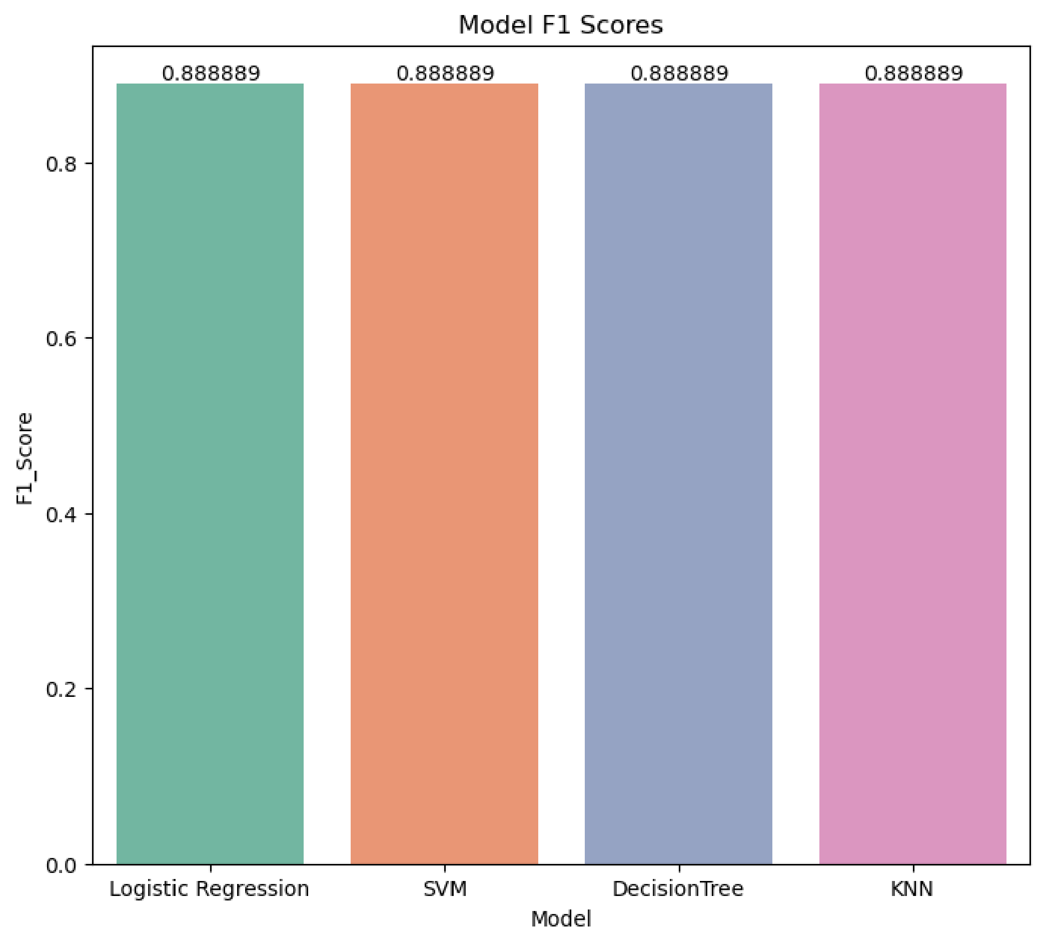
The following scoring methods were used to identify each models accuracy :

- Jaccard Index
- F1 Score
- R^2 Score

When tested for accuracy using the test set data, all scores were the same

- A possible reason for this result is the sample size
- Test sample size is 20% of the full data set (only 18 samples) making it hard to differentiate accuracy between the different models

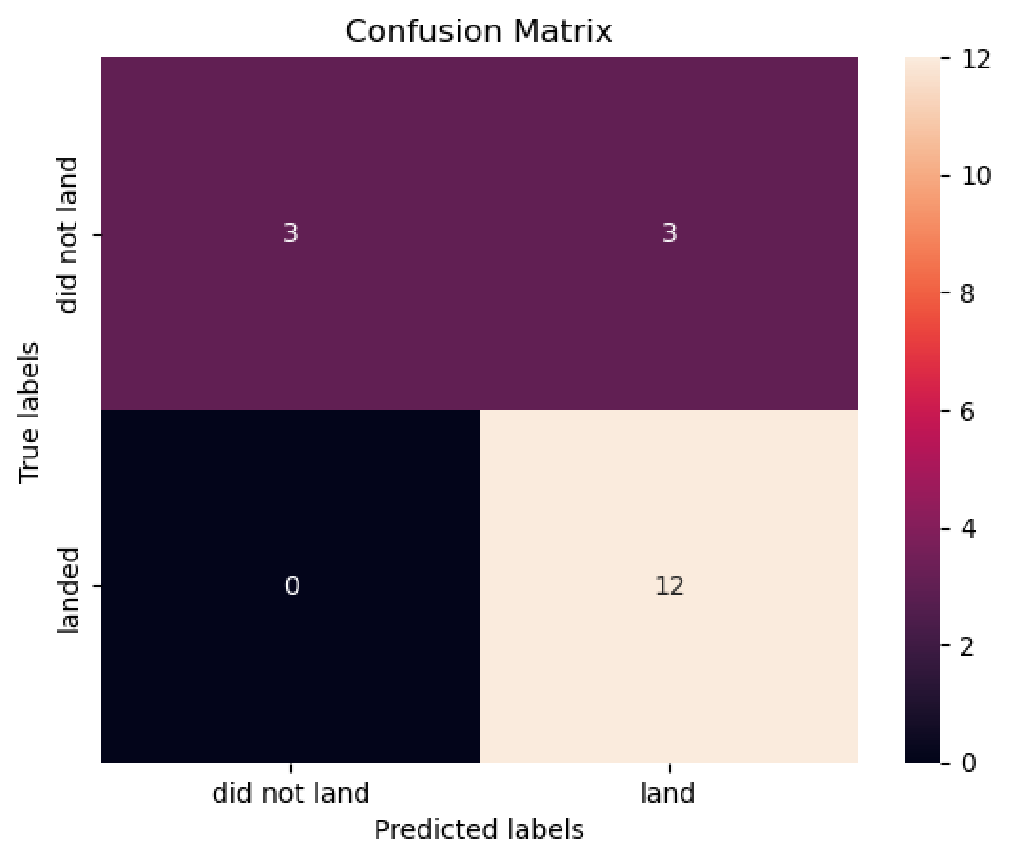
Model	Jaccard_Score	F1_Score	R2_Score
Logistic Regression	0.8	0.888889	0.833333
SVM	0.8	0.888889	0.833333
DecisionTree	0.8	0.888889	0.833333
KNN	0.8	0.888889	0.833333



Confusion Matrix

A confusion matrix displays 4 different quadrants of information

- Quadrant 1: False Positive
  - Incorrectly predicts rocket successfully landed
- Quadrant 2: True Negative
  - Correctly predicts rocket failed to land
- Quadrant 3: False Negative
  - Incorrectly predicts rocket failed to land
- Quadrant 4: True Positive
  - Correctly predicts rocket successfully landed



Due to the small sample size, all machine learning algorithms had the same level of accuracy and outputted the same values for their Confusion Matrix

Conclusions

- The machine learning algorithms have an F1 score of 0.89 which shows it's skill in accurately predicting both successful and failed launches
- Launches with a payload mass between 2000 and 6000 kg show the highest rate of success
- Launch Site KSC LC 39-A has the highest success rate among the other launch sites
- Orbits ES-L1, GEO, HEO, and SSO have a 100% success rate
    - Orbits ISS, GTO, and VLEO have the most launches
- Launch Sites are located in close proximity to coastlines