SpaceX Falcon 9

First Stage Landing Prediction

Farah Dahir

Outline

1. Executive Summary
2. Introduction
3. Methodology
4. Results
5. Conclusion

Executive Summary

Summary of Methodology

- Data Acquisition
- Exploratory Data Analysis
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive analysis (Classification)

Summary of Results

- Exploratory Data Analysis Results
- Geospatial Maps
- Interactive Dashboard
- Predictive analysis Models

Introduction

Project Background

The cost of a Falcon 9 launch, at $62 million, is notably lower than competitors' offerings, which can exceed $165 million per launch. SpaceX's competitive advantage lies in the reusability of the first stage, resulting in significant cost savings compared to other providers in the space launch industry. This cost information can be invaluable for potential competitors looking to bid against SpaceX for rocket launch contracts.

Introduction

Key Questions to Address

- The best way to estimate the total cost for launches, by predicting successful landings of the first stage of rockets.

- How do variables such as payload mass, launch site, number of flights, and orbits affect the success of the first stage landing.

- Where is the best place to make rockets launches.

Methodology

- Data Acquisition
- Exploratory Data Analysis
- Geospatial Analysis
- Interactive Dashboard
- Predictive Analysis

Data Acquisition

From SpaceX API:

- Data was obtained from the SpaceX API. Filtered to obtain useful columns, filled missing values and Store the cleaned data in a CSV file.

1. GET request made to the API and response converted to a Pandas DataFrame.

2. Filtering the response and retaining only the Falcon 9 rocket launches and important columns.

3. Filling missing values in the ‘PayloadMass’ column using the mean value.

1

```
spacex_url="https://api.spacexdata.com/v4/launches/past"

response = requests.get(spacex_url)

# Use json_normalize method to convert the json result into a dataframe
data = pd.json_normalize(response.json())
```

2

```
# Call getBoosterVersion
getBoosterVersion(data)

# Call getLaunchSite
getLaunchSite(data)

# Call getPayloadData
getPayloadData(data)

# Call getCoreData
getCoreData(data)
```

```
launch_dict = {'FlightNumber': list(data['flight_number']),
'Date': list(data['date']),
'BoosterVersion':BoosterVersion,
'PayloadMass':PayloadMass,
'Orbit':Orbit,
'LaunchSite':LaunchSite,
'Outcome':Outcome,
'Flights':Flights,
'GridFins':GridFins,
'Reused':Reused,
'Legs':Legs,
'LandingPad':LandingPad,
'Block':Block,
'ReusedCount':ReusedCount,
'Serial':Serial,
'Longitude': Longitude,
'Latitude': Latitude}
```

3

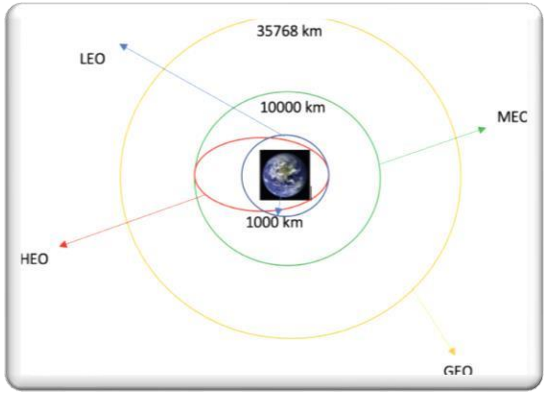
```
# Calculate the mean value of PayloadMass column and Replace the np.nan values with its mean value
data_falcon9 = data_falcon9.fillna(value={'PayloadMass': data_falcon9['PayloadMass'].mean()})
```

GitHub Link

Exploratory Data Analysis

Using PANDAS:

- The SpaceX dataset contains several Space X launch facilities, and each location is in the LaunchSitecolumn.
- Each launch aims to a dedicated orbit, and some of the common orbit types are shown in the figure below. The orbit type is in the Orbitcolumn.



Initial Data Exploration:

- Using the .value_counts()method to determine the following:

1. Number of launches on each site
2. Number of and occurrence of each orbit
3. Number and occurrence of landing outcome per orbit type

1

```
# Apply value_counts() on column LaunchSite
df['LaunchSite'].value_counts()

CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
Name: LaunchSite, dtype: int64
```

2

```
# Apply value_counts on Orbit column
df['Orbit'].value_counts()

GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
ES-L1     1
GEO       1
SO        1
HEO       1
Name: Orbit, dtype: int64
```

3

```
# landing_outcomes = values on Outcome column
landing_outcomes = df['Outcome'].value_counts()
landing_outcomes

True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
None ASDS       2
False Ocean     2
False RTLS      1
Name: Outcome, dtype: int64
```

GitHub Link

Exploratory Data Analysis

Labeling the Outcome:

- To ascertain the likelihood of a successful booster landing, it is advisable to employ a binary column, denoted by values 1 or 0, to signify the outcome of the landing. This process involves:

1. Establishing a collection of unsuccessful outcomes termed as "bad_outcome.“
2. Generating a list, named "landing_class," where each element assumes the value 0 if the corresponding entry in the "Outcome" falls within the "bad_outcome" set; conversely, it takes on the value of 1 if it does not.
3. Constructing a "Class" column that incorporates the values derived from the "landing_class" list, effectively categorizing the outcomes as either successful (1) or unsuccessful (0) landings.

1

```
bad_outcomes=set(landing_outcomes.keys()[[1,3,5,6,7]])
bad_outcomes
```

```
{'False ASDS', 'False Ocean', 'False RTLS', 'None ASDS', 'None None'}
```

2

```
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise

landing_class = []

for outcome in df['Outcome']:
    if outcome in bad_outcomes:
        landing_class.append(0)
    else:
        landing_class.append(1)
```

3

```
df['Class']=landing_class
```

Exploratory Data Analysis

SCATTER CHARTS

Scatter charts were used to visualize the relationships between:

- Flight Number and Launch Site
- Payload and Launch Site
- Orbit Type and Flight Number
- Payload and Orbit Type

BAR CHART

A bar chart was used to visualize the relationship between:

- Success Rate and Orbit Type

LINE CHARTS

Line chart was produced to visualize the relationships between:

- Success Rate and Year (i.e. the launch success yearly trend)

[GitHub Link]

Geospatial Analysis

Using Folium:

- Circles, Markers and lines were drawn on Folium Maps:

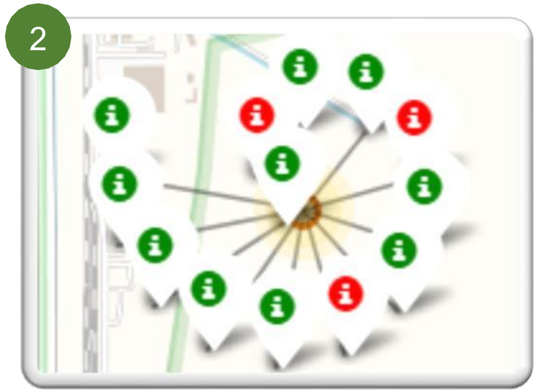
1. Used circles to indicate highlighted areas around specific coordinates, like NASA Johnson Space Center;

2. Added colored markers of successful (green) and unsuccessful (red) launches at each launch site to show which launch sites have high success rates

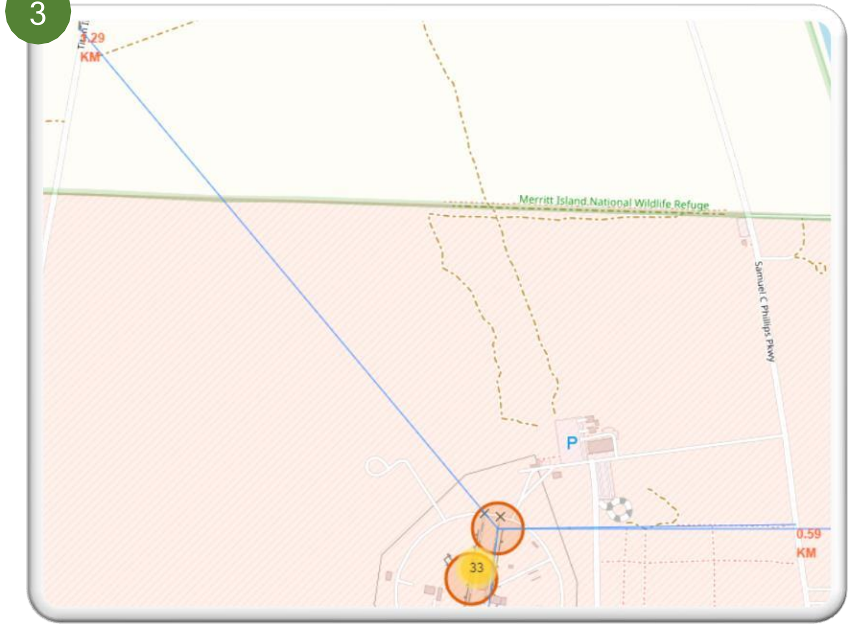
3. Added coloured Lines to show distances between Launch Site and its proximities like Railway, Highway, Coastline and Closest City.

1

2



3



[GitHub Link](#)

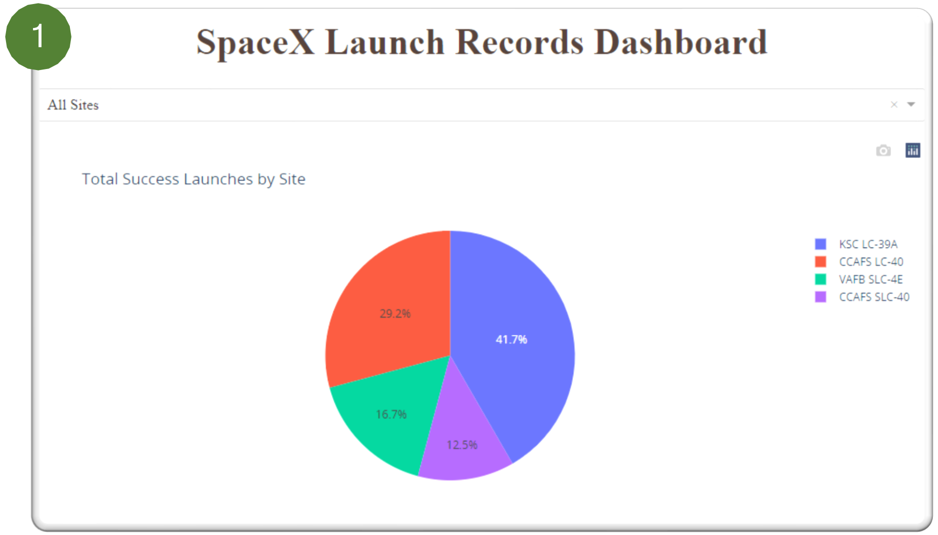
Interactive Dashboard

Using Plotly Dash:

- A dropdown list was used to enable Launch Site selection, pie chart for Success Launches of all sites and a Slider of Payload Mass Range:

1. Added a pie chart to show the total successful launches count for all sites and the Success vs. Failed counts for the site, if a specific Launch Site was selected.

2. Used Scatter Chart to show correlation between Payload Mass and Success Rate for different Booster Versions.



GitHub Link

Predictive Analysis

ML Classification Models:

- The following steps were taking to develop, evaluate, and find the best performing classification model:

1. Standardize the data with StandardScaler. Fit and transform the data.

2. Split the data using train_test_split

3. Create a GridSearchCV object with cv=10 for parameter optimization

4. Apply GridSearchCV on different algorithms: logistic regression, support vector machine, decision tree and KNN

5. Calculates accuracy scores and compared to Identified the best model for the project.

	Algorithms	Accuracy Score	Best Score
0	Logistic Regression	0.833333	0.821429
1	Support Vector Machine	0.833333	0.848214
2	Decision Tree	0.944444	0.887500
3	K Nearest Neighbors	0.777778	0.833929

Results

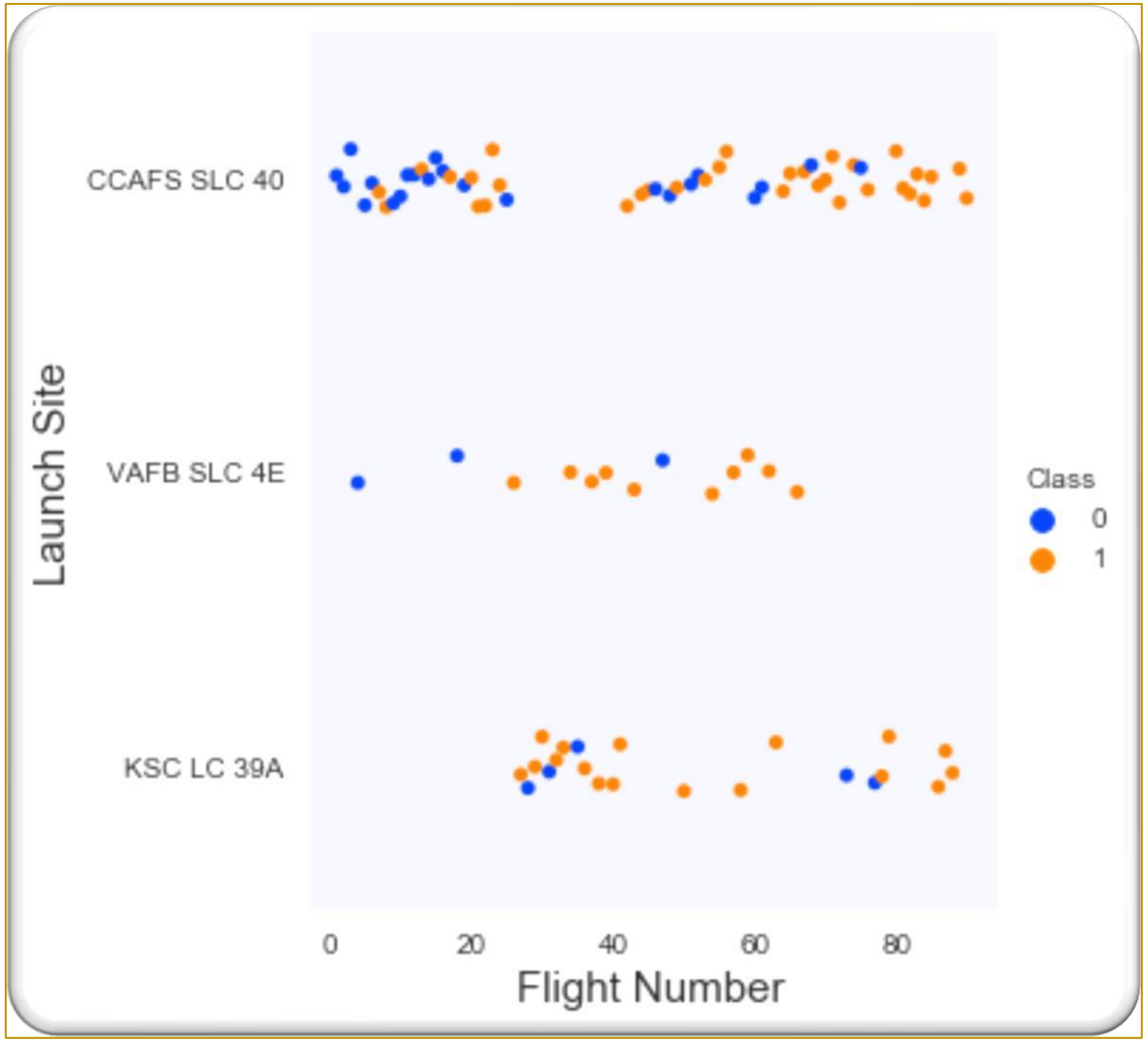
- EDA Results
- Interactive Analytics
- Predictive Analysis

EDA Results

Launch Site Vs. Flight Number

Scatter plot of Launch Site vs. Flight Number

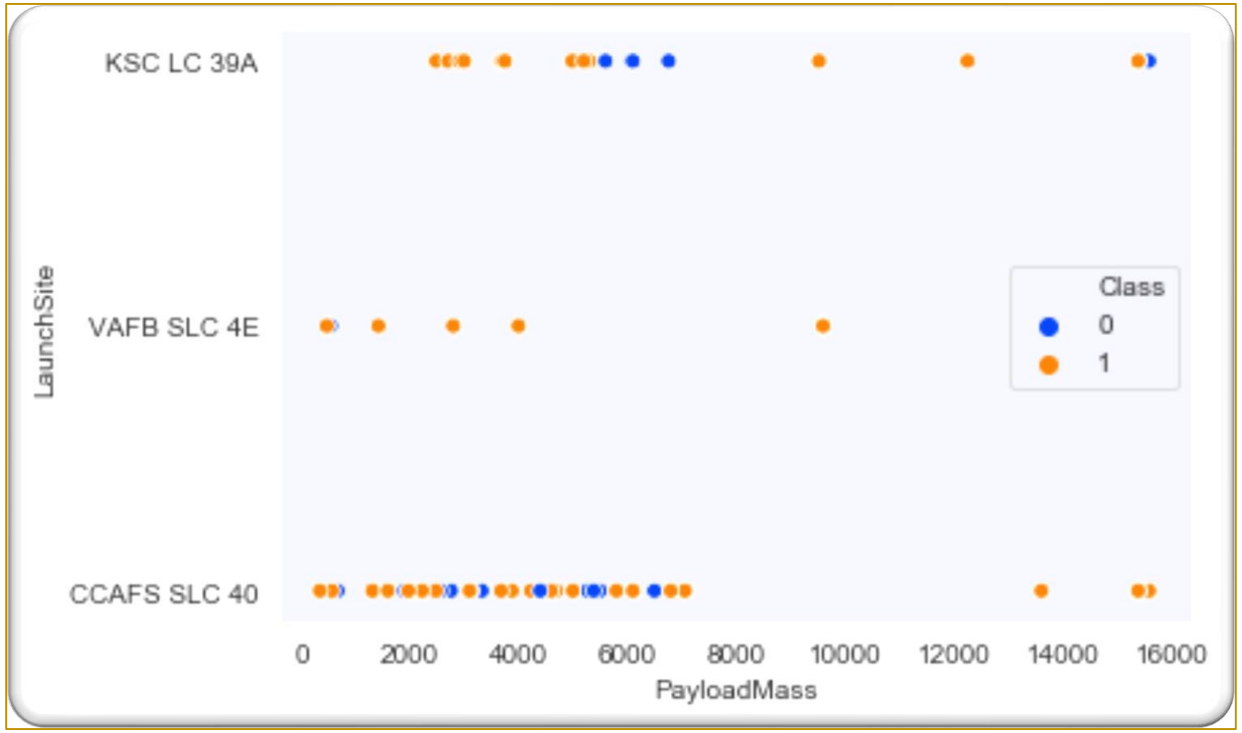
- Flight numbers less than 30 were generally unsuccessful. Can be seen in sites CCAFS SLC 40 and VAFB SLC 4E.
- The CCAFS SLC 40 launch site has about half of all launches.
- VAFB SLC 4E and KSC LC 39A are the sites that have higher success rates.
- Above flight number of 80, there is a 100% success rate for sites CCAFS SLC 40 and KSC LC 39A.



Launch Site Vs. Payload Mass

Scatter plot of Launch Site vs. Payload Mass

- KSC LC 39A has a 100% success rate for launches less than 5,500 kg.
- Most of the launches from CCAFS SLC40 are of lighter payloads (with some outliers).
- Most launces with a payload greater than 7,000 kg were successful.
- For VAFB-SLC there are no rockets launched for heavy payload mass(greater than 10000).

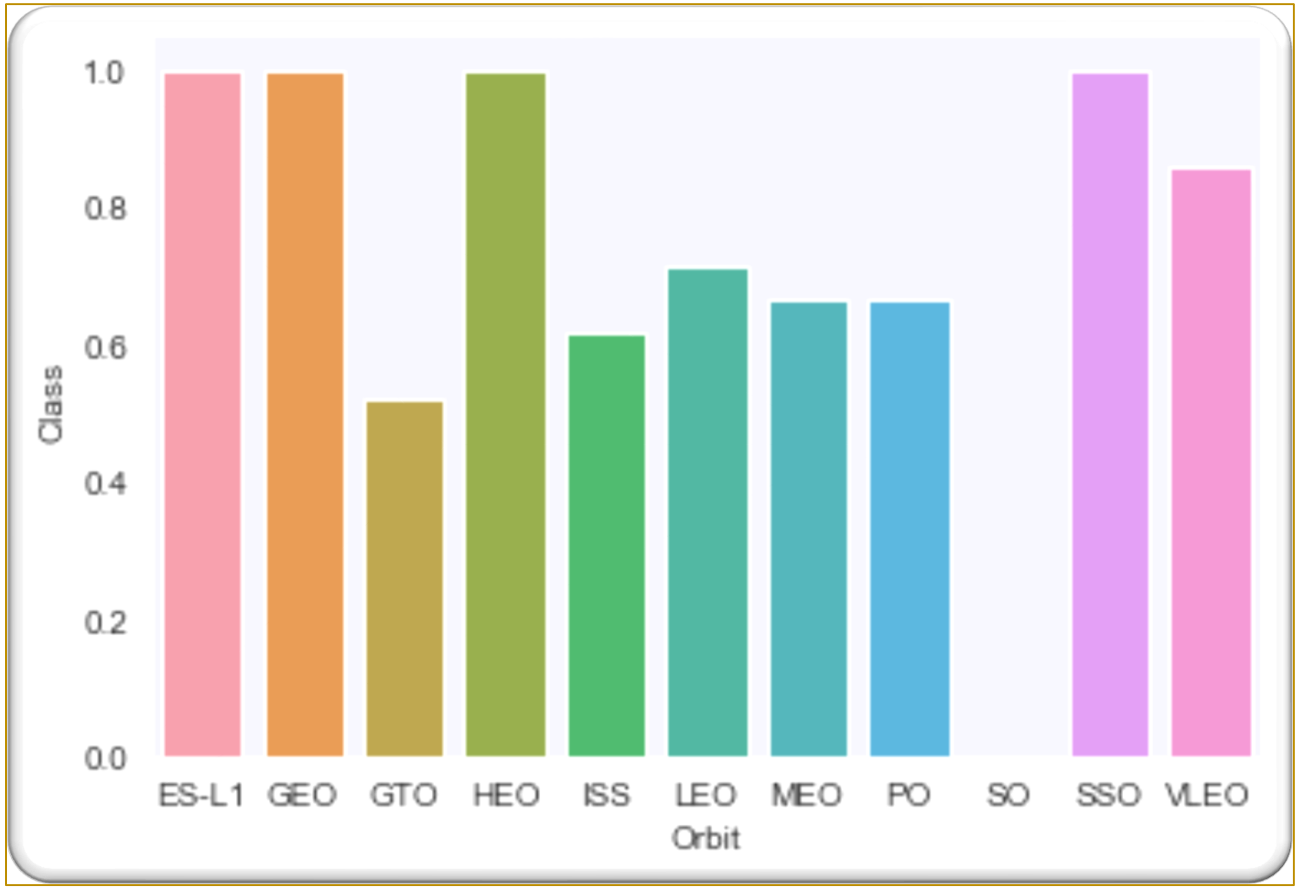


Success Rate Vs. Orbit Type

Bar chat of Success Rate vs. Orbit Type

- Orbits with 100% success rate:

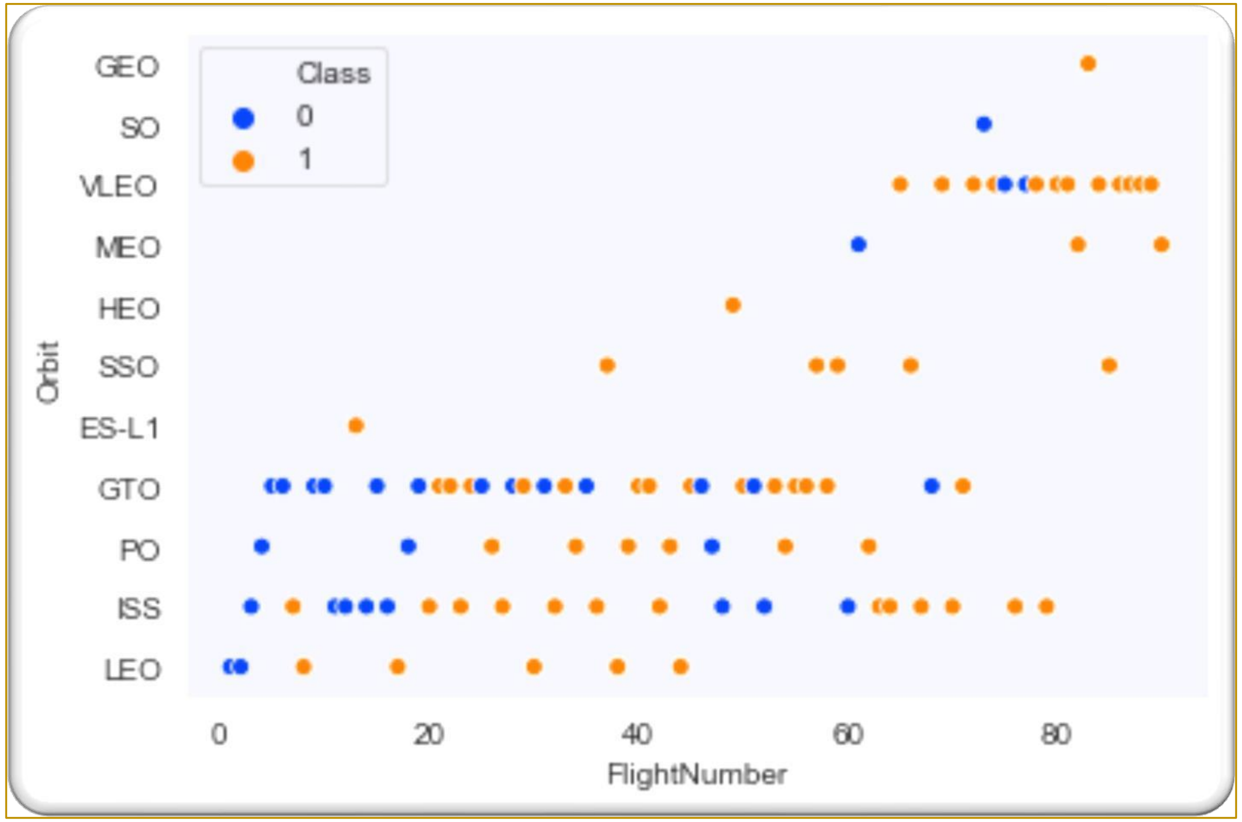
  ES-L1, GEO, HEO and SSO
- Followed by VLEO (above 80%);
- The orbit with the lowest (0%) success rate is SO



Orbit Type Vs. Flight Number

Scatter plot of Orbit Type vs. Flight Number

- 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.
- GEO, HEO, and ES -L1 only had 1 flight each which were all successful
- 100% success rate in SSO

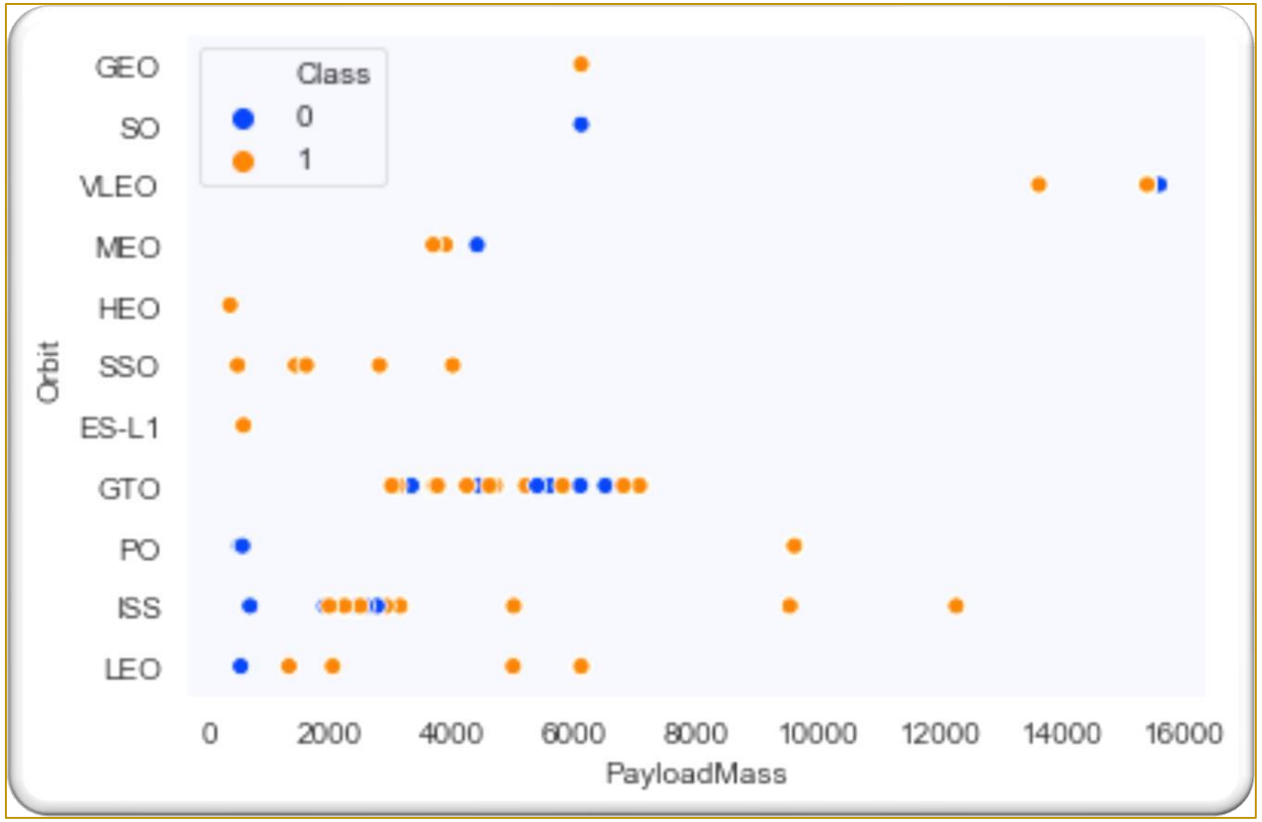


EDA Results

Orbit Type Vs. Payload Mass

Scatter plot of Orbit Type vs. Payload Mass

- With heavy payloads the successful landing rate are more for PO, LEO and ISS.
- However for GTO we cannot distinguish this well as both successful and unsuccessful landings are mixed up with an increase in payload Mass.
- ISS has the widest range of payload and about 67% success rate.

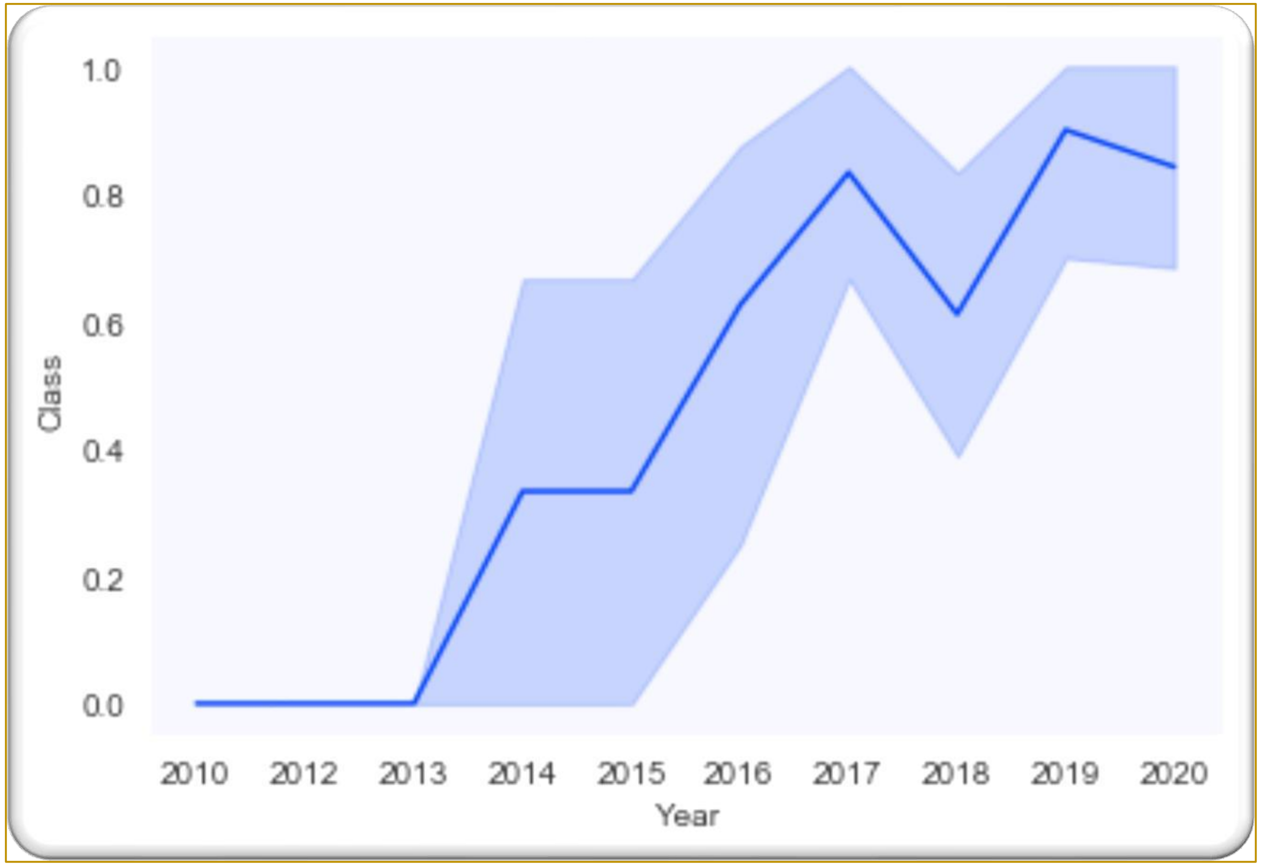


EDA Results

Launch Success Yearly Trend

Line chart of launch success for 2010-2020

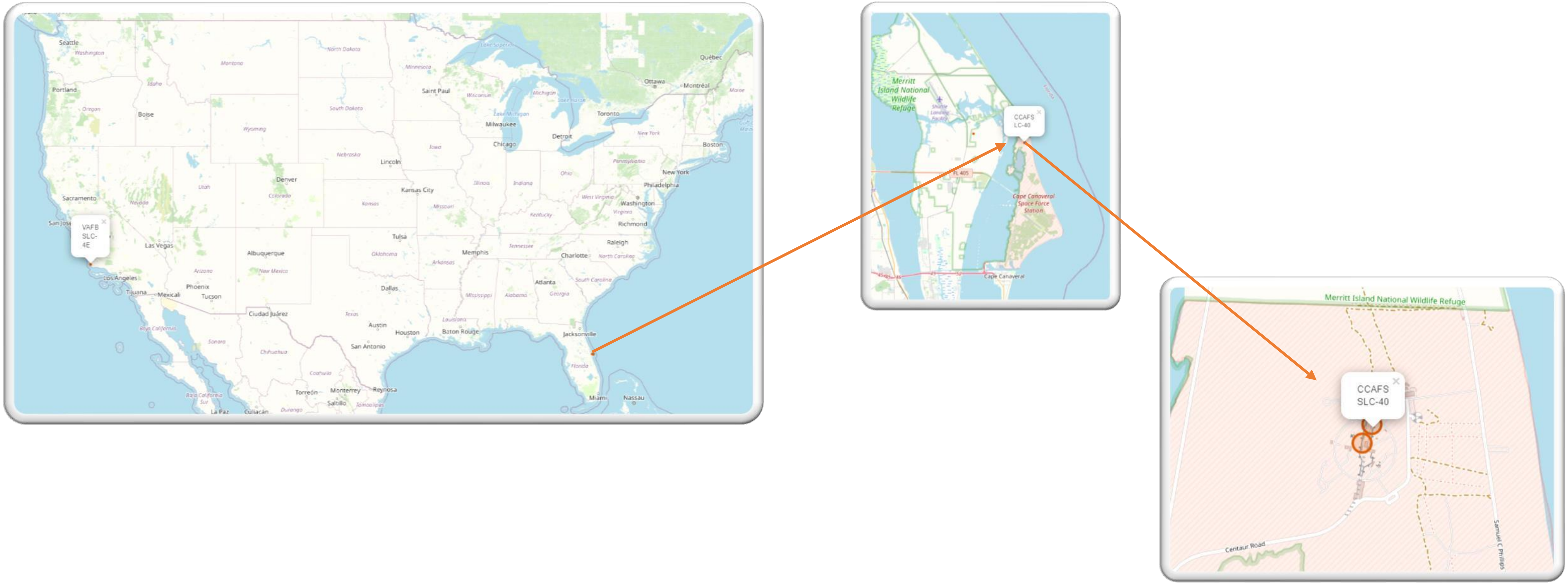
- First 3 years (2010-2013), all landings were unsuccessful.
- From 2013 success rate kept increasing till 2020.
- From 2016 onwards, there was always more than a 50% chance of success.



Interactive Analysis (Folium)

All Launch Sites

All sites are close proximity to the coast. Most of Launch sites are in proximity to the Equator line and rockets launched from these sites get an additional natural boost due to the rotational speed of earth.



Interactive Analysis (Folium)

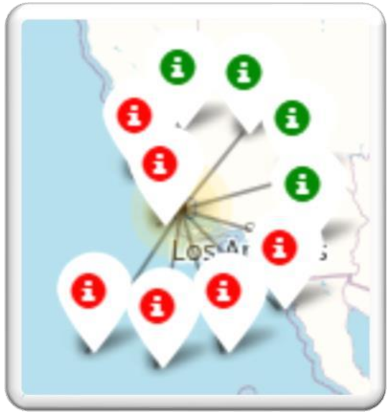
Launch Outcomes by Site

Green markers indicate successful and red ones indicate unsuccessful launches.

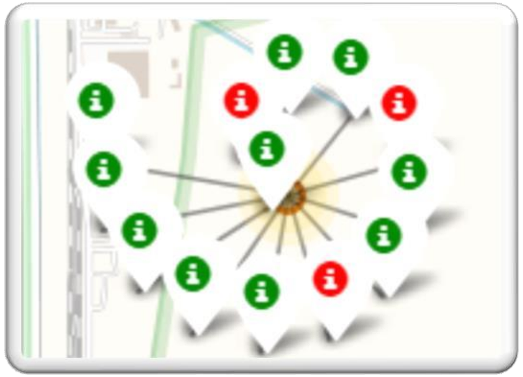
Launch Site KSC LC-39A has a very high Success Rate.

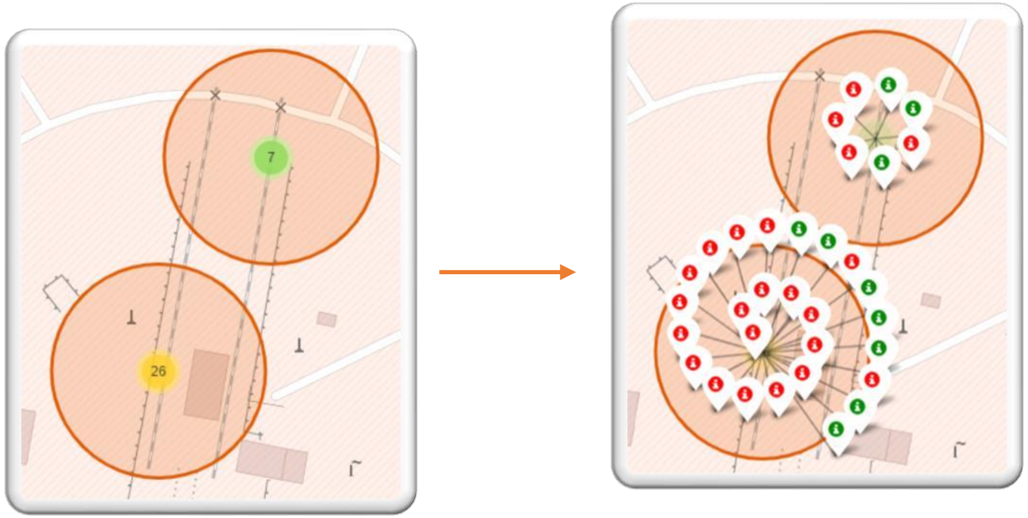
CCAFS SLC-40 and CCAFS LC-40

VAFB SLC-4E



KSC LC-39A



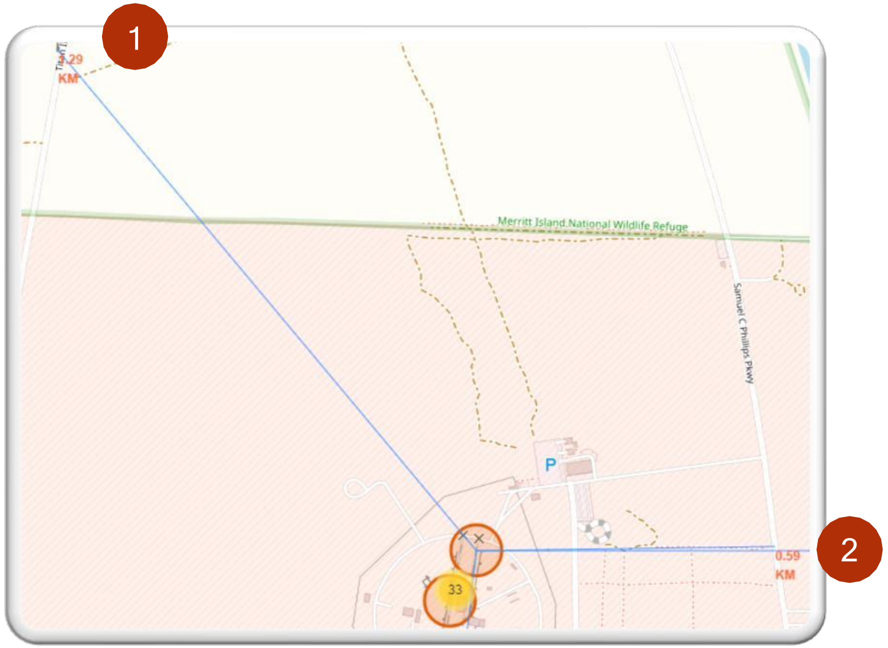


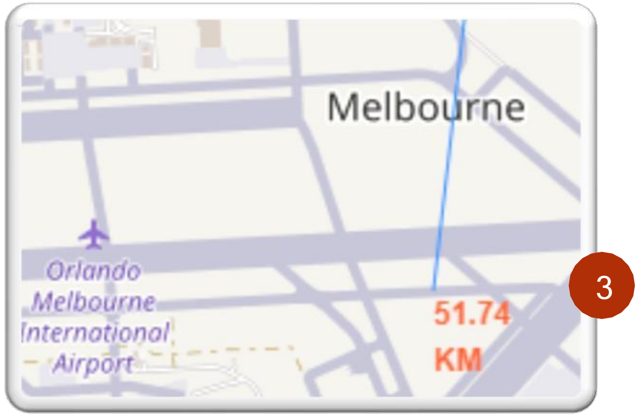
Interactive Analysis (Folium)

Distance to Proximities

This shows distance proximities of the CCAFSSLC-40 launch site as an example and it can be see that it is:

1. Close to railway (1.29 km)
2. Close highway (0.29 km)
3. Very far from the nearest city (51.74 km)

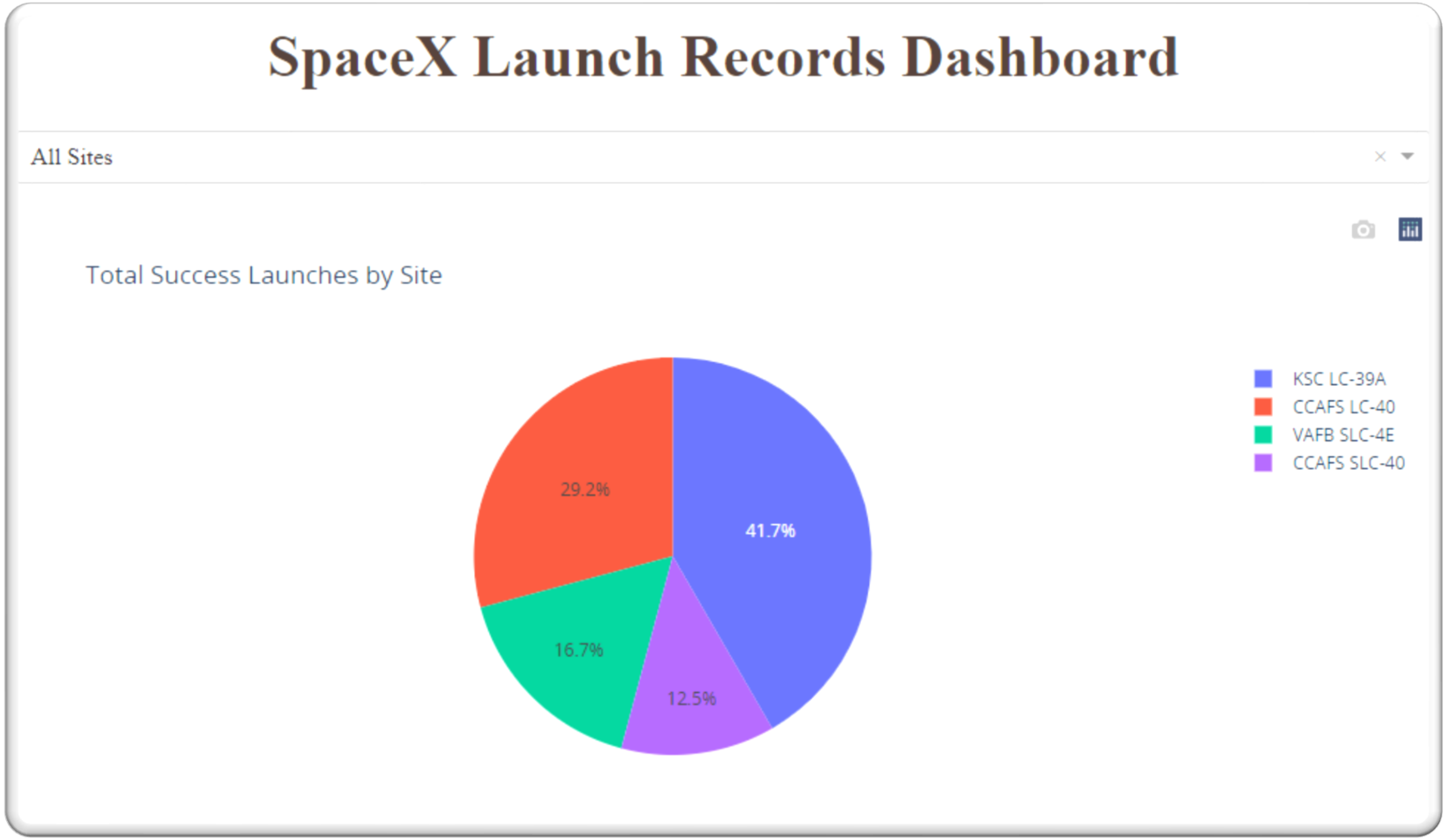




Launch Success by Site

Total Success Launch by Site

- KSC LC-39A has the most successful launches amongst launch sites (41.2%)

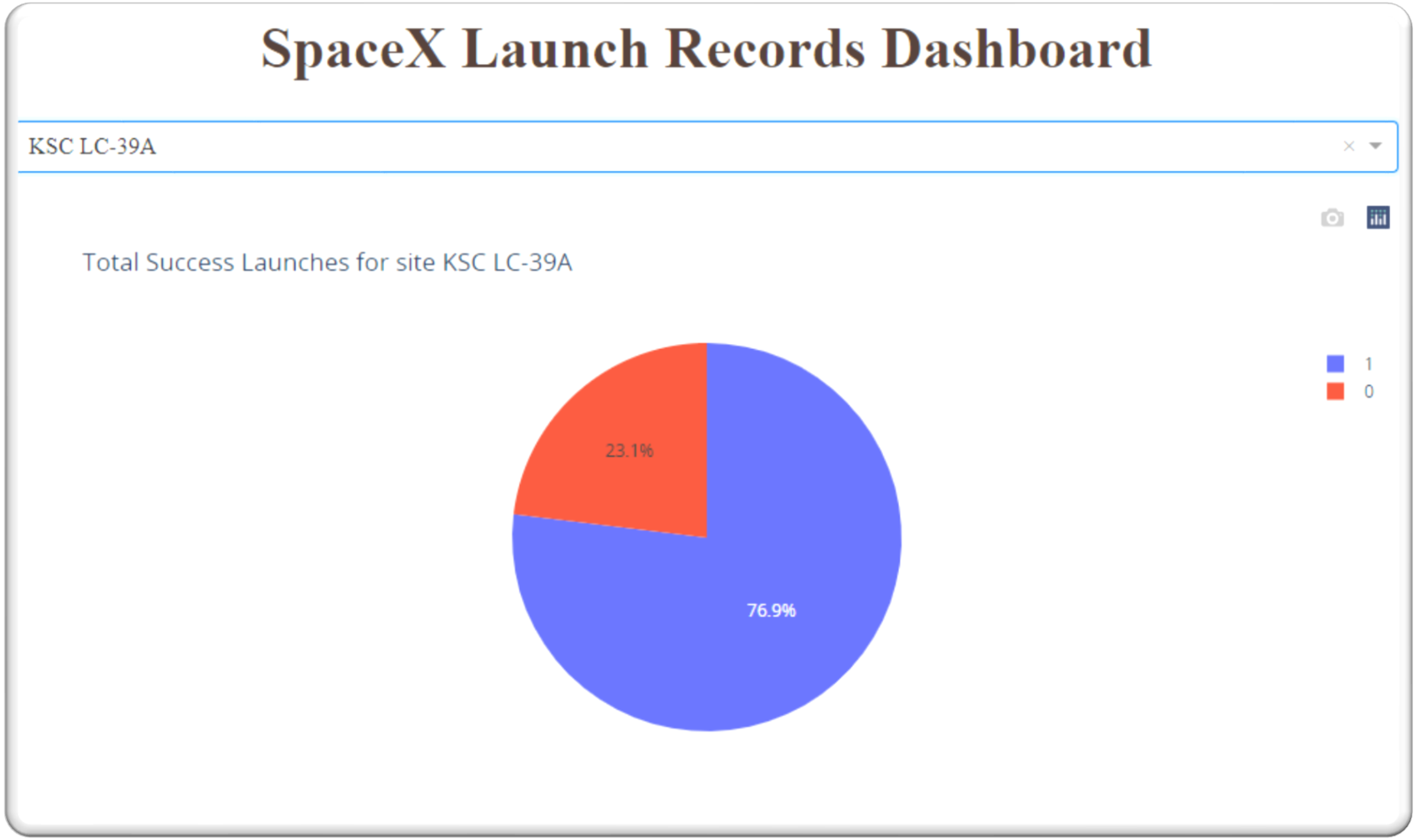


Interactive Analysis (Plotly Dash)

Highest launch success ratio

Total Success Launch for KSC LC-39A

- KSC LC-39A has the highest launch success rate (76.9%) with 10 successful and only 3 failed landings.



Interactive Analysis (Plotly Dash)

Payload vs. Launch Outcome

Correlation between Payload and Landing

- Payloads between 2,000 kg and 4,000 kg (low payload) have the highest success rate.
- Payloads of 4,000-10,000 (massive payload), have less success rate.

1

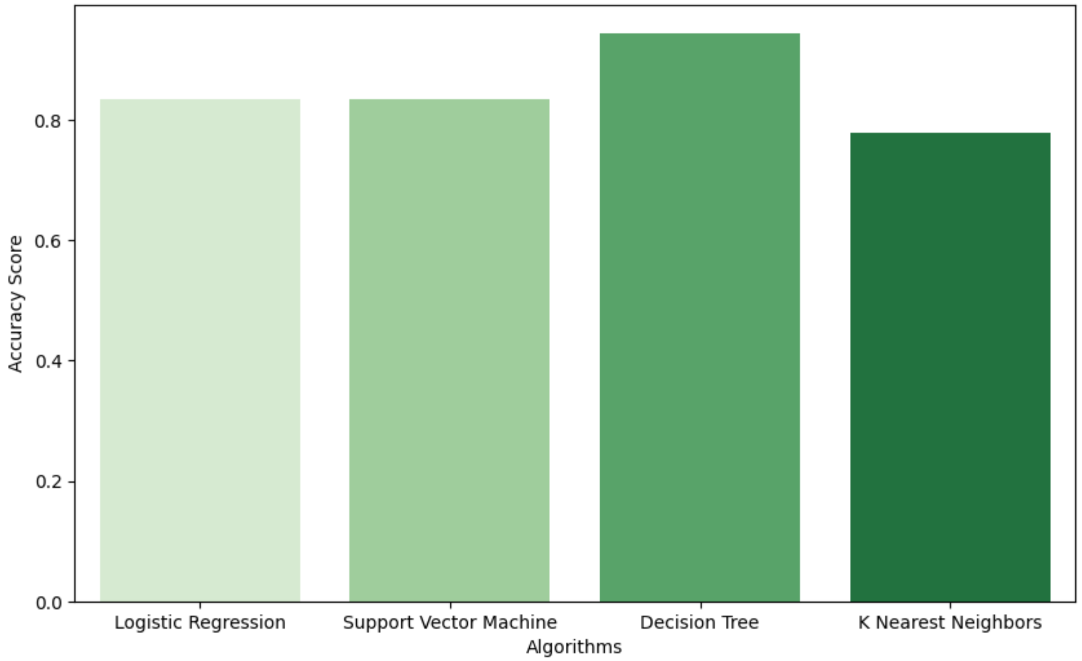
2

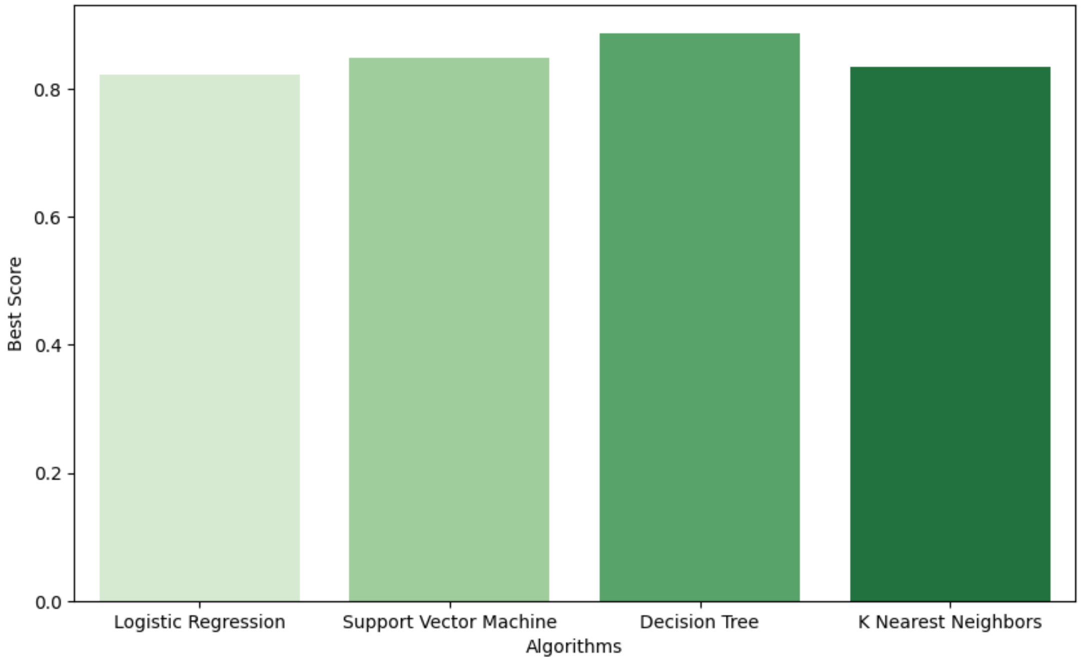
Predictive Analysis

Classification Accuracy

Accuracy Score

- The model with the highest accuracy was Decision Tree Classifier, which had an accuracy_score of 0.94 and a best_score of 0.89 after hyper-parameter tuning.



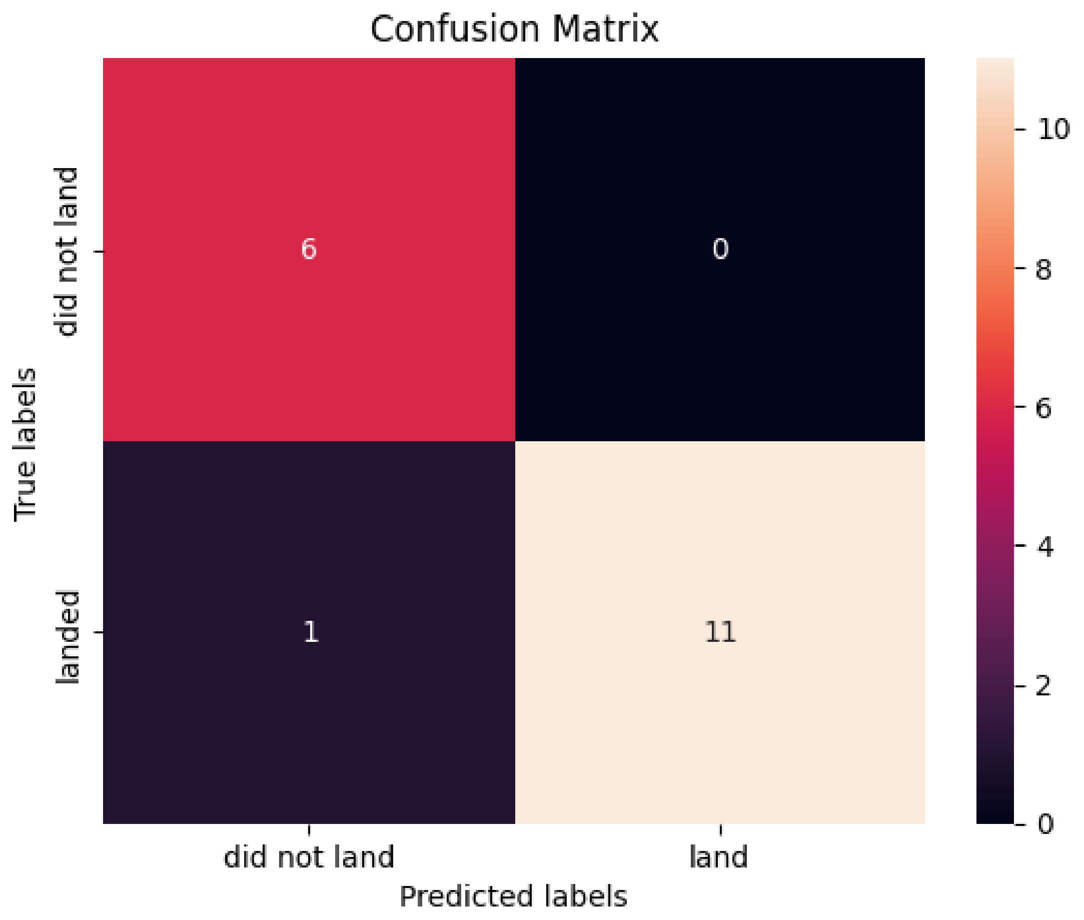


Predictive Analysis

Confusion Matrix

True +ve/–ve VS False +ve/-ve

- The confusion matrix of the Decision Tree Classifier shows only 1 out of 18 total results classified incorrectly (a false negative, shown in the bottom-left corner). The other 17 are correctly classified (6 did not land, 11 did land).



Conclusion

Conclusion

- Most of the launch sites are near the equator for an additional natural boost - due to the rotational speed of earth, which helps save the cost of putting in extra fuel and boosters.
- The success rate of launches increases over the years.
- The launch site KSCLC-39 A had the most successful launches, with 41.7% of the total successful launches, and also the highest rate of successful launches, with a 76.9% success rate.
- Launches with a low payload mass (less than 4,000)show better results than launches with a larger payload mass.
- Orbits ES-L1, GEO, HEO, and SSO have a 100% success rate
- Decision Tree Classifier can be used to predict successful landings hence estimate the total cost for a launch.

THANK YOU!