SPACEY

MUSOKE JOSEPH JOHN MUGAYA

Wednesday 13 Nov 2024

Outline

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

Executive Summary

This report aims to predict the success of Falcon 9 first stage landings. SpaceX has reduced launch costs to $62 million by reusing the first stage of its rockets, compared to $165 million from other providers.

Predicting landing outcomes helps estimate launch costs and supports competitive bids against SpaceX. Using machine learning, this analysis models the probability of successful landings based on historical data and key features like launch conditions, payload mass, and mission type.

Results show that these factors significantly impact landing success, with the best model achieving high accuracy and predictive power. This aids stakeholders in forecasting outcomes and making informed decisions, enhancing market intelligence and assessing potential savings.

Introduction

Space Exploration Technologies Corp., commonly referred to as SpaceX, is an American space technology company. Since its founding in 2002 the company has made numerous advancements in rocket propulsion, reusable launch vehicle, human spaceflight and satellite constellation technology. It's the most successful in its industry.

One reason SpaceX can do rocket launches relatively inexpensive is because it can reusable launch vehicles, one of them being the Flacon 9. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. Spaces X's Falcon 9 launch like regular rockets.

We are going to also determine if SpaceX will reuse the first stage. Instead of using rocket science to determine if the first stage will land successfully, you will train a machine learning model and use public information to predict if SpaceX will reuse the first stage.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was using SpaceX REST API and web scraping techniques
- Perform data wrangling
  - Filtering the data, One-hot encoding, handling missing values and convert landing outcomes into Training Labels with 1 means the booster successfully landed 0 means it was unsuccessful.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Create a machine learning pipeline to predict if the first stage will land

Data Collection

Data collection was done using get request to the SpaceX API.

Next, we decoded the response content as a Json using .json() function call and turn it into a pandas dataframe using .json_normalize().

We then cleaned the data, checked for missing values and fill in missing values where necessary.

In addition, we performed web scraping from Wikipedia for Falcon 9 launch records with BeautifulSoup.

The objective was to extract the launch records as HTML table, parse the table and convert it to a pandas dataframe for future analysis.

Data Collection – SpaceX API

Requested rocket launch data from SpaceX API
url=”`https://api.spacexdata.com/v4/launches/past`”
Requested and parsed SpaceX data with GET request

Decoded response as a json file
Turned it into a Pandas data frame

Used the API with functions to get info from IDs:

- BoosterVersion
- Payloads
- LaunchSite
- Cores

Stored them in a List

Assigned list to dictionary
Converted to Dataframe

Filtered data and remained with Falcon 9 data only, Replaced missing payload mass value with mean

Saved Dataframe into a csv file

https://github.com/Musoke-Joseph/IBM-SpaceX-Capstone/blob/0756954952f9297cfa2a3bbdb11c2c8961166d51/Data%20collection%20api.ipynb

Data Collection - Scraping

Request the Falcon9 Launch Wiki page from its URL
HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response

Extract all column/variable names from the HTML table header
collect all relevant column names from the HTML table header

Create a data frame by parsing the launch HTML tables
create an empty dictionary with keys from the extracted column names. Later, this dictionary will be converted into a Pandas dataframe

fill up the dictionary with launch records extracted from table rows

Saved Dataframe into a csv file

https://github.com/Musoke-Joseph/IBM-SpaceX-Capstone/blob/0756954952f9297cfa2a3bbdb11c2c8961166d51/Webscraping.ipynb

Data Wrangling

Calculate the number of launches on each site. Using the method value_counts()

Calculate the number and occurrence of each orbit
Use the method value_counts()

Calculate the number and occurence of mission outcome of the orbits
Use the method .value_counts()

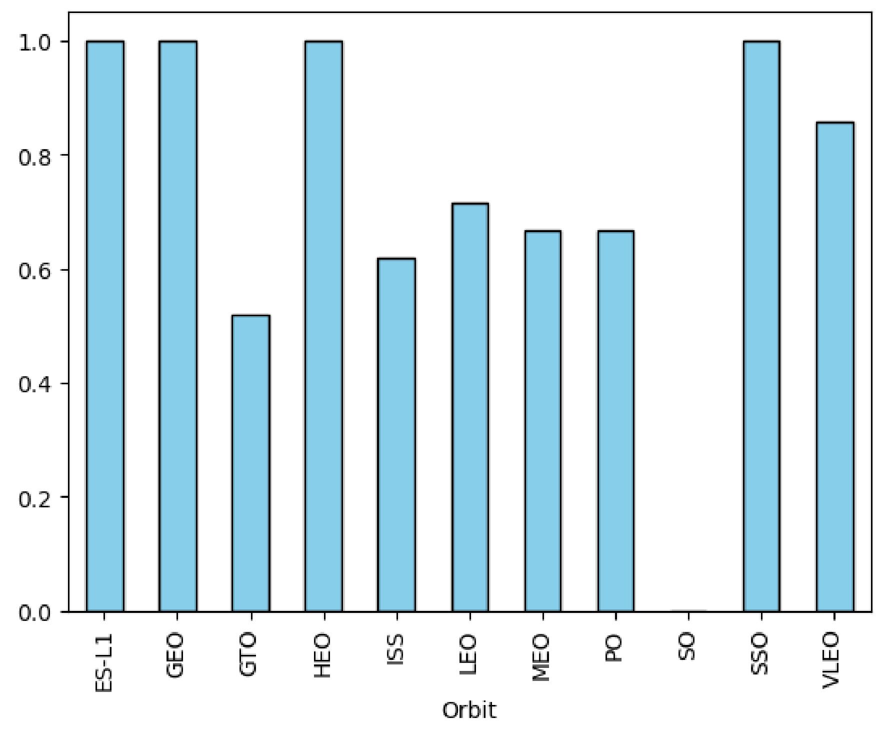
Create a set of outcomes where the second stage did not land successfully

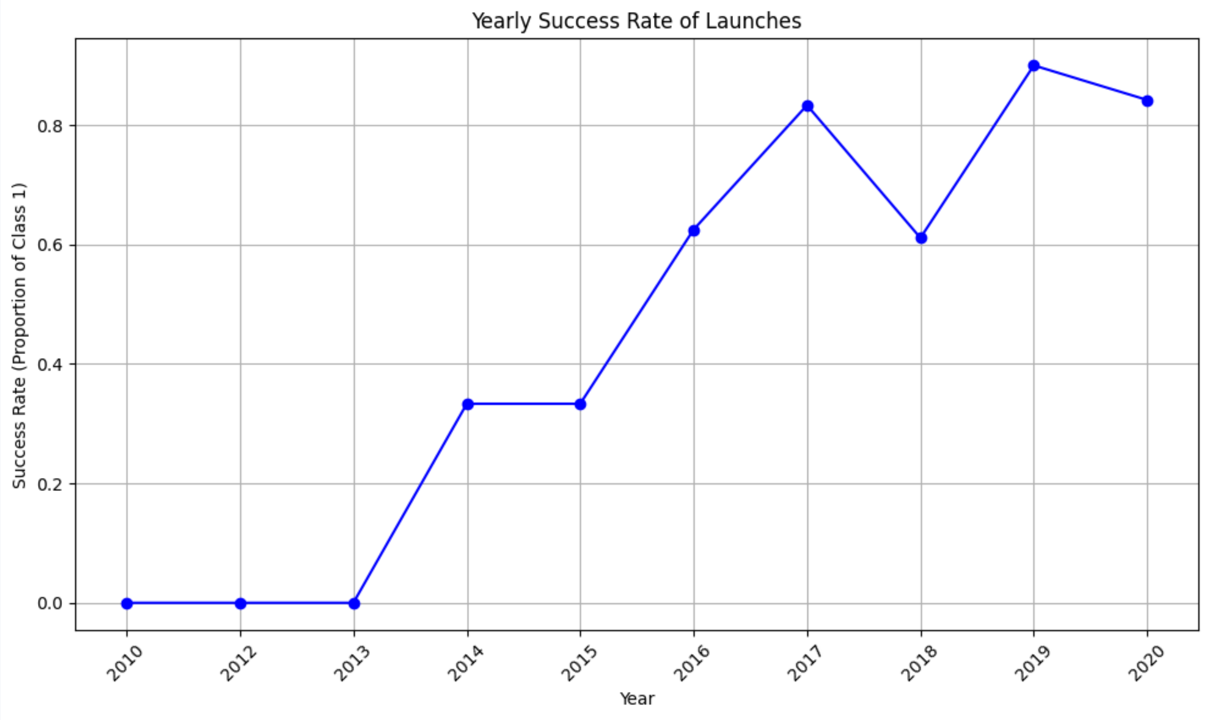
Create a landing outcome label from Outcome column
Using the Outcome, create a list where the element is zero if the corresponding row in Outcome is in the set bad_outcome; otherwise, it's one. Then assign it to the variable landing_class

Saved Dataframe into a csv file

https://github.com/Musoke-Joseph/IBM-SpaceX-Capstone/blob/0756954952f9297cfa2a3bbdb11c2c8961166d51/Data%20wrangling.ipynb

EDA with Data Visualization





Bar Chart. To visually check if there are any relationship between success rate and orbit type

Line Chart. To visually check to get the average launch success trend. We can observe that the success rate since 2013 kept increasing till 2020

https://github.com/Musoke-Joseph/IBM-SpaceX-Capstone/blob/0756954952f9297cfa2a3bbdb11c2c8961166d51/Exploring%20and%20Preparing%20Data%20Viz.ipynb

EDA with SQL

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

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

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

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

- List the date when the first successful landing outcome in ground pad was achieved

```
%%sql
SELECT MIN(DATE) FROM SPACEXTBL
    WHERE Landing_Outcome == 'Success (ground pad)'
    AND Mission_Outcome == 'Success'
```

- List the total number of successful and failure mission outcomes

```
%%sql
SELECT Mission_Outcome, COUNT(Mission_Outcome) FROM SPACEXTBL
    GROUP BY Mission_Outcome
```

- 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)
```

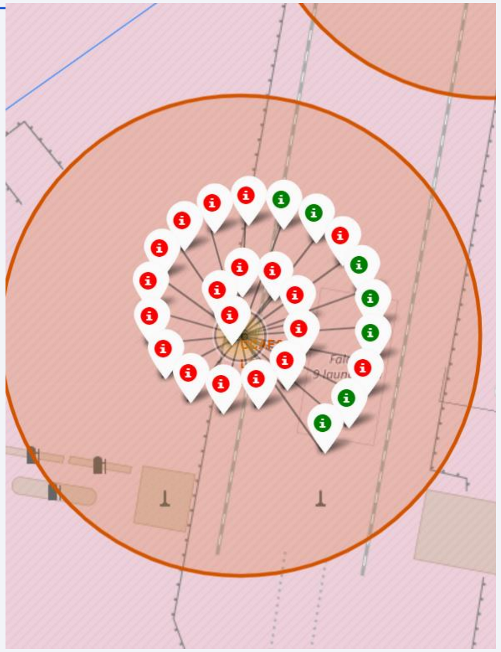
https://github.com/Musoke-Joseph/IBM-SpaceX-Capstone/blob/0756954952f9297cfa2a3bbdb11c2c8961166d51/Exploring%20and%20Preparing%20Data%20SQL.ipynb

Build an Interactive Map with Folium

- Launch success rate may depend on the location and proximity of a launch site. Folium Interactive Map was used for visualizing and analyzing SpaceX Launch Sites.
- Used Interactive mapping library called Folium
- Identified all SpaceX launch sites on a map: Florida, California
- Included longitude and latitude info.
- Identified successful/failed launches for each site on map

Calculated the distance between a launch site (CCAFS_SLC40 in Cape

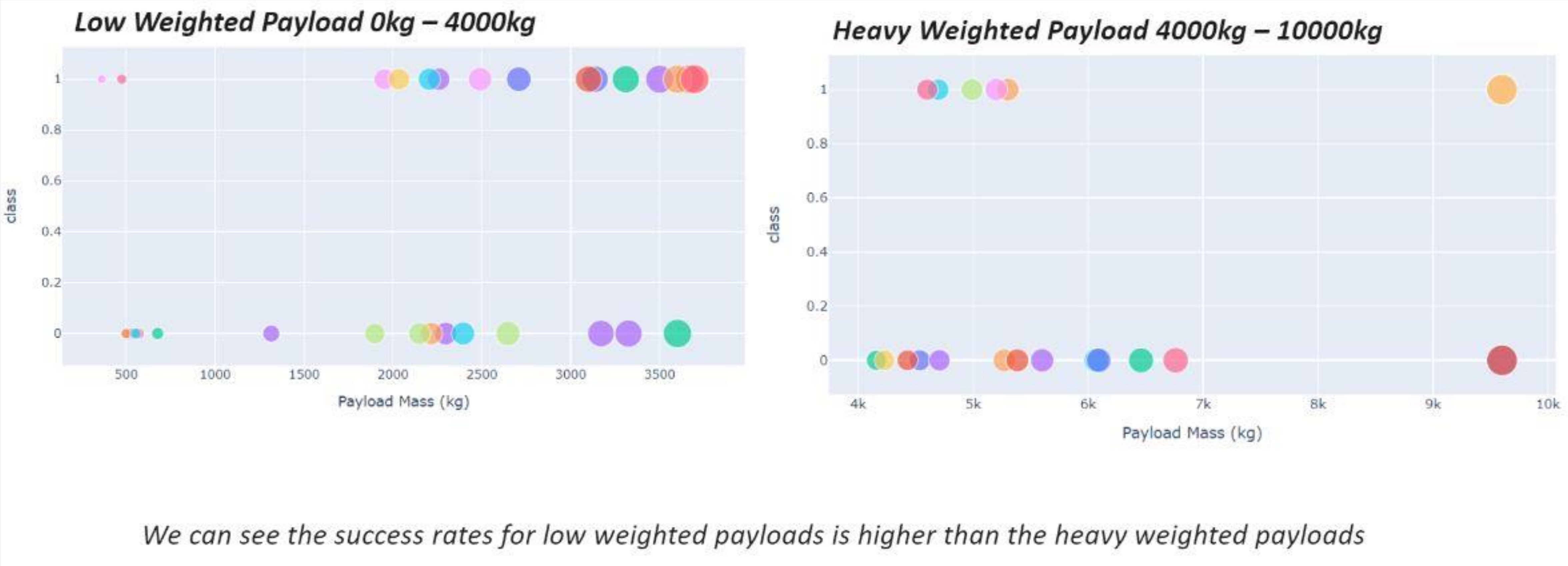
- Canaveral, FL) and:
- Closest coastline
- Closest high traffic density railway: Florida East Coast Railway
- Closest high traffic density highway: Interstate I95
- Closest high density urban area: Orlando (FL)



https://github.com/Musoke-Joseph/IBM-SpaceX-Capstone/blob/0756954952f9297cfa2a3bbdb11c2c8961166d51/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

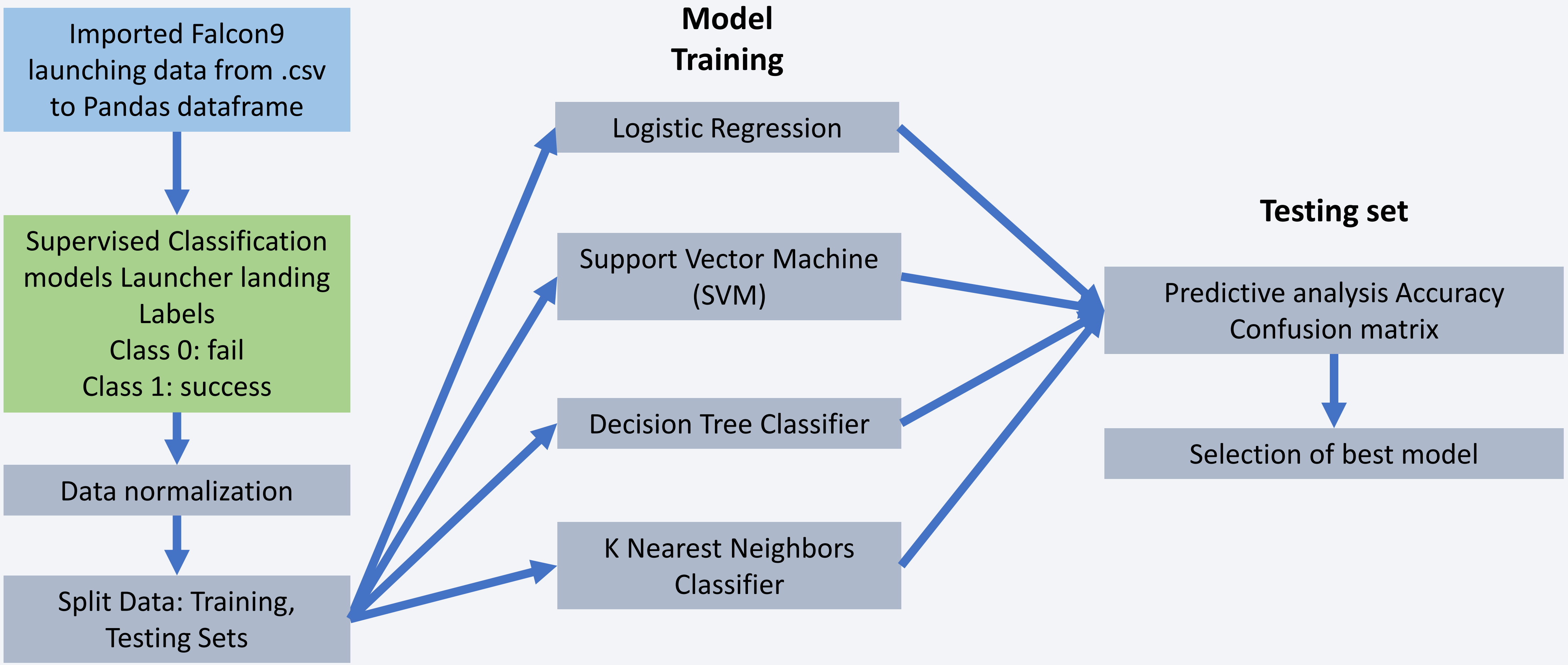
Build a Dashboard with Plotly Dash

- We built an interactive dashboard with Plotly dash



https://github.com/Musoke-Joseph/IBM-SpaceX-Capstone/blob/0756954952f9297cfa2a3bbdb11c2c8961166d51/dashboard.py

Predictive Analysis (Classification)



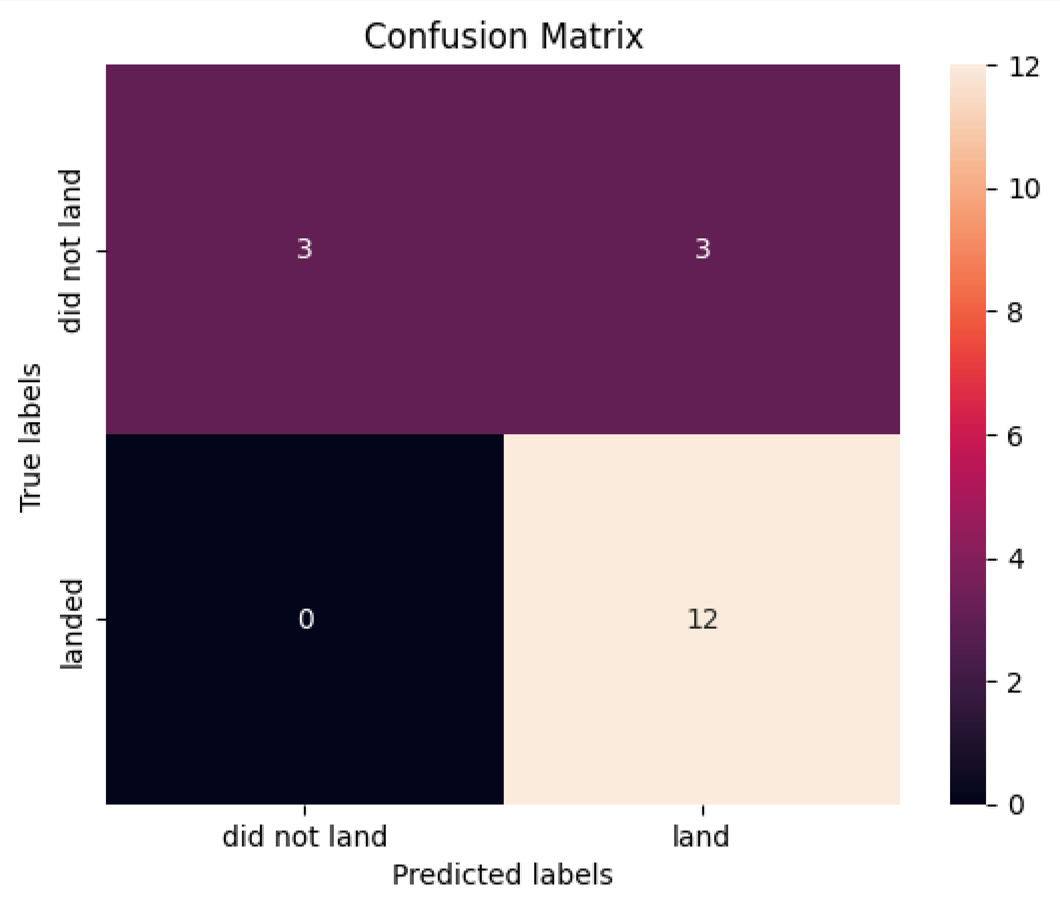
https://github.com/Musoke-Joseph/IBM-SpaceX-Capstone/blob/10221212a17cf779cddc98ba901ad2728d3c033d/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

The best model was the Decision Tree model with the highest accuracy score of 0.8333

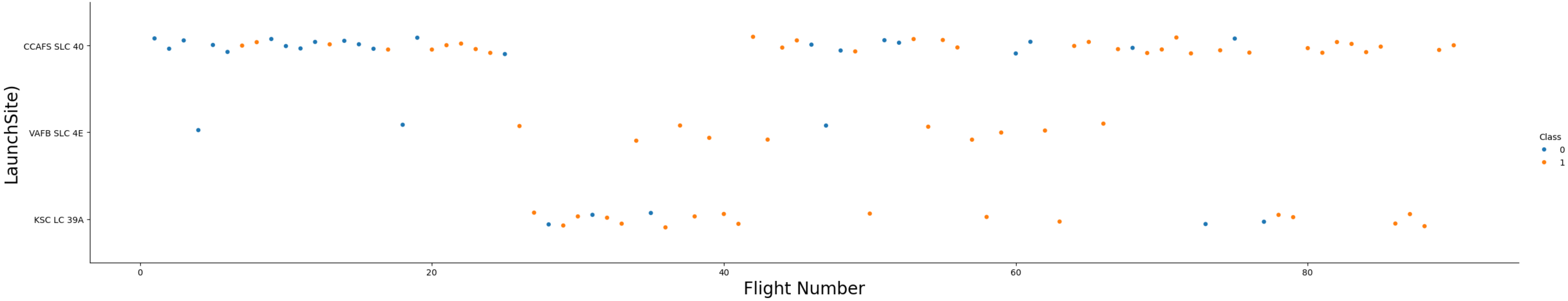
```
tree_cv.score(X_test, Y_test)
```

```
0.8333333333333334
```



Section 2

Insights drawn from EDA

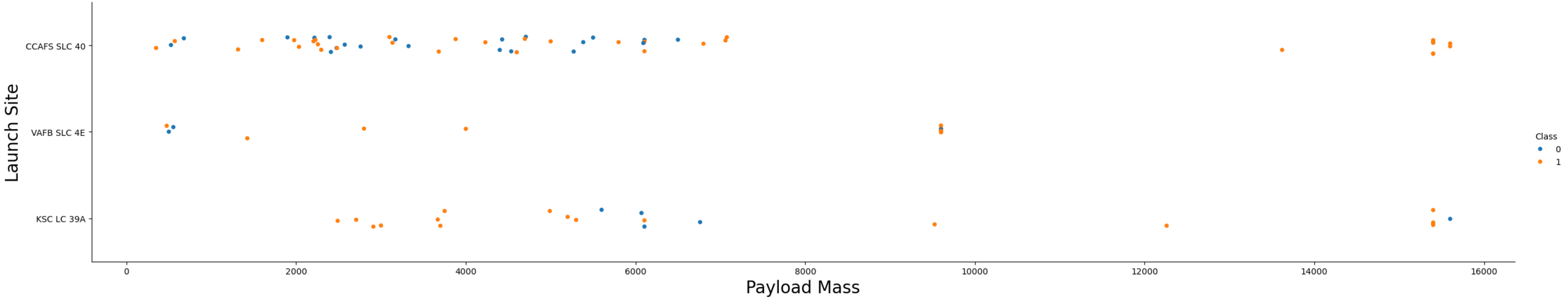


Flight Number vs. Launch Site

This scatterplot visualizes the distribution of SpaceX flights based on the flight number (x-axis) and the launch site (y-axis).

The plot uses different colored dots to indicate the success or failure of each launch, with blue representing unsuccessful launches (Class 0) and orange representing successful launches (Class 1).

The three launch sites depicted are CCAFS SLC 40, VAFB SLC 4E, and KSC LC 39A. The plot shows that launches have occurred from all three sites, with varying success rates distributed across the flight numbers.



Payload vs. Launch Site

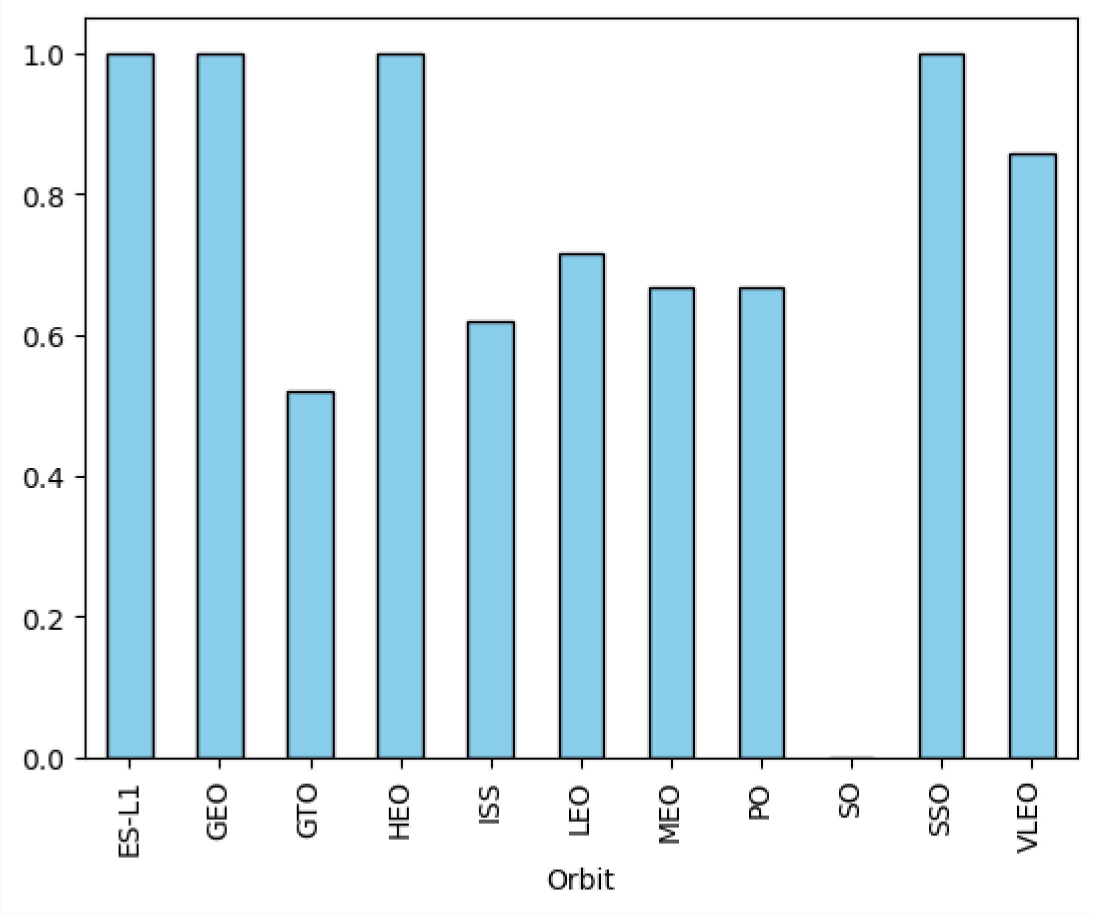
- CCAFS SLC 40 and KSC LC 39A have the widest range of payload masses, with many successful launches (orange dots) across various payloads.
- VAFB SLC 4E shows fewer launches, primarily successful ones at lower payload masses.
- The success rate appears generally consistent, with more successful launches (orange) present across different payload masses at each site.
- The highest payload masses (around 14,000–16,000 kg) have been launched successfully from CCAFS SLC 40 and KSC LC 39A.

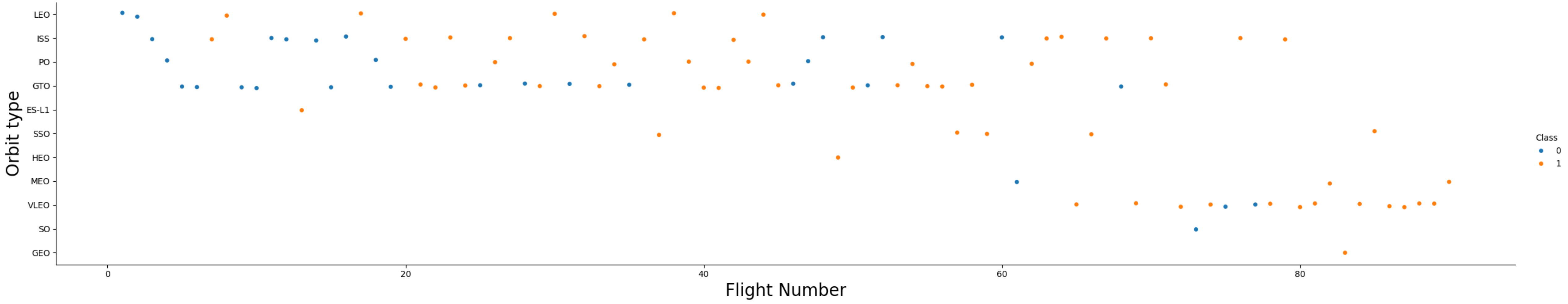
Success Rate vs. Orbit Type

This bar chart illustrates the success rate of launches for different orbit types. The y-axis represents the success rate, ranging from 0 (failure) to 1 (success), while the x-axis lists various orbit types.

- The orbits ES-L1, GEO, HEO, and SSO have a perfect success rate of 1, indicating that all launches targeting these orbits were successful.
- Orbits like VLEO also show a high success rate, close to 1.
- GTO (Geostationary Transfer Orbit) has the lowest success rate among the orbits shown.
- ISS, LEO, MEO, PO, and SO orbits have moderate success rates, varying between approximately 0.6 and 0.8.

This suggests that the choice of orbit type may impact the likelihood of launch success, with some orbits showing consistently better results than others.





Flight Number vs. Orbit Type

Blue dots (Class 0) represent unsuccessful outcomes.

Orange dots (Class 1) represent successful outcomes.

1. **Low Earth Orbit (LEO)** and **Geostationary Transfer Orbit (GTO)** are the most frequently targeted orbits.
2. Success rates (indicated by the prevalence of orange dots) appear higher in certain orbits, particularly LEO and ISS, while orbits like GTO have more variation in outcomes.
3. As flight numbers increase (rightward on the x-axis), the rate of successful flights generally appears to improve, indicating possible advancements or increased reliability over time.
4. Less common orbits, like **SSO** and **VLEO**, show mixed results with fewer data points, making it hard to conclude specific trends for these orbits.

This suggests that over time, the mission success rates have generally improved, especially in certain common orbit types.

Payload vs. Orbit Type

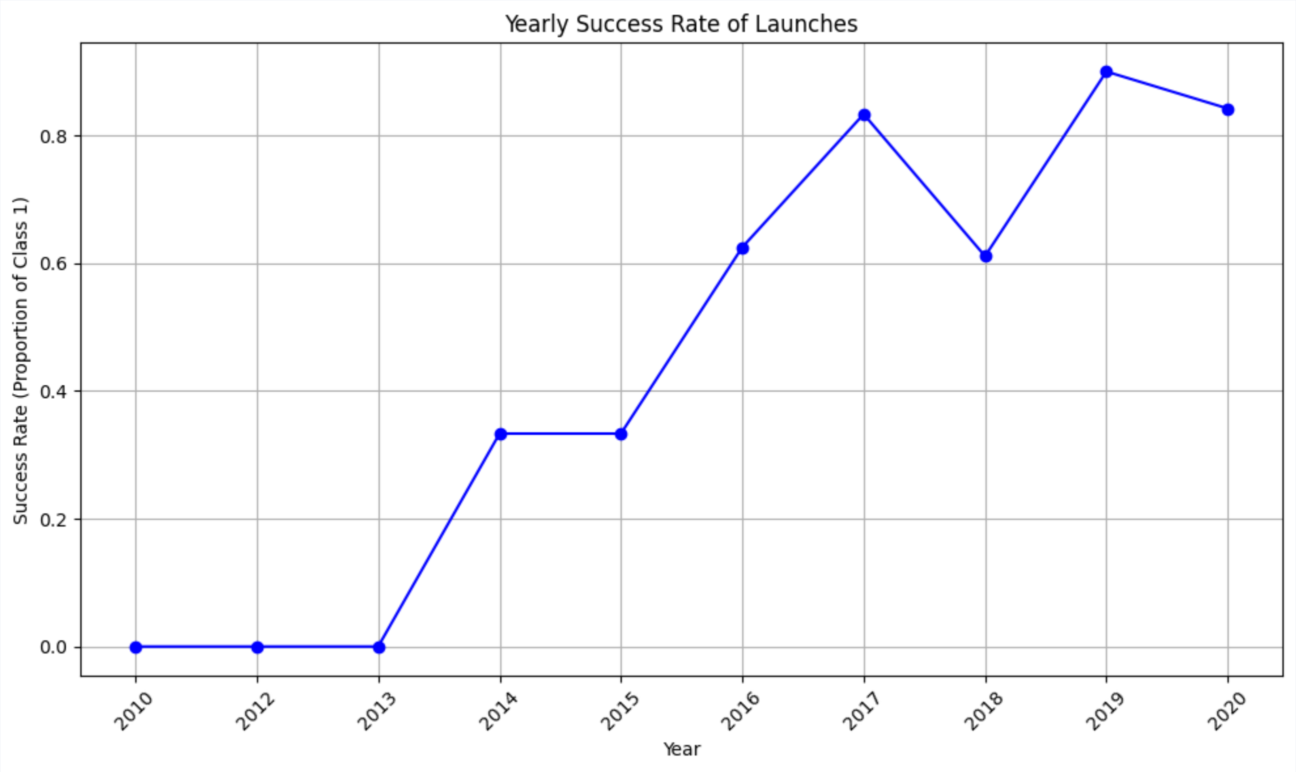
1. **Lower Payload Mass** (under 5000 kg):
    1. There are a large number of both successful and unsuccessful flights, particularly in **LEO**, **ISS**, and **GTO** orbits.
    2. Success appears more consistent in **LEO** and **ISS** compared to GTO at these lower payload masses.
2. **Medium Payload Mass** (5000–10,000 kg):
    1. Flights to **GTO** are common in this payload range and show mixed results, although success is more frequent.
    2. LEO also sees some flights in this range, with mostly successful outcomes.
3. **Higher Payload Mass** (above 10,000 kg):
    1. Only a few flights are in this range, with destinations in **GTO** and **GEO**.
    2. Most high-mass payloads show successful outcomes, suggesting reliable performance for heavy payloads, particularly for these orbits.

Overall, successful outcomes appear more frequent in certain orbits, like LEO and ISS, across a range of payload masses, while higher payloads generally have better success rates in GTO and GEO orbits.

Launch Success Yearly Trend

1. **Early Years (2010-2013)**:
   1. The success rate was initially low, with no successful launches from 2010 to 2012.
   2. In 2013, the success rate started to improve slightly.
2. **Significant Improvement (2014-2017)**:
   1. There was a steady increase in success rates starting in 2014, with a sharp rise by 2016, reaching a success rate above 0.6.
   2. By 2017, the success rate peaked near 0.9, showing substantial reliability improvements.
3. **Fluctuation and Stabilization (2018-2020)**:
   1. In 2018, there was a slight dip in success rate, though it remained high.
   2. Success rates improved again in 2019, reaching the highest level around 0.9.
   3. In 2020, there was a minor decrease, but the rate remained above 0.8, indicating stable, high success rates.

Overall, the plot reveals a strong upward trend in launch success rates over the years, with significant improvements after 2013, leading to consistently high success rates from 2017 onward.



All Launch Site Names

The data contains the following Space X launch facilities: Cape Canaveral Space Launch Complex, Vandenberg Air Force Base Space Launch Complex VAFB SLC-4E , Kennedy Space Center Launch Complex KSC LC-39A

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

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

- Here are 5 records where launch sites begin with `CCA`.

Total Payload Mass

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Customer == 'NASA (CRS)'
```

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

SUM(PAYLOAD_MASS__KG_)
45596

Here is the total payload carried by boosters from NASA. It is 45596

Average Payload Mass by F9 v1.1

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Booster_Version == 'F9 v1.1'
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

Here is the average payload mass carried by booster version F9 v1.1. Its 2928.4

First Successful Ground Landing Date

The date of the first successful landing outcome on ground pad was 22 Dec 2015

```
%%sql
SELECT MIN(DATE) FROM SPACEXTBL
    WHERE Landing_Outcome == 'Success (ground pad)'
    AND Mission_Outcome == 'Success'
```

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

MIN(DATE)
2015-12-22

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 are

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

```
%%sql
SELECT Booster_Version FROM SPACEXTBL
    WHERE Mission_Outcome == 'Success'
    AND 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

The calculated total number of successful and failure mission outcomes are as follows

```
%%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

Boosters Carried Maximum Payload

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

```
%%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

The list the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015 is empty because there are n records that fit that criteria

```
%%sql
SELECT CASE substr(Date, 4, 2)
    WHEN '01' THEN 'January'
    WHEN '02' THEN 'February'
    WHEN '03' THEN 'March'
    WHEN '04' THEN 'April'
    WHEN '05' THEN 'May'
    WHEN '06' THEN 'June'
    WHEN '07' THEN 'July'
    WHEN '08' THEN 'August'
    WHEN '09' THEN 'September'
    WHEN '10' THEN 'October'
    WHEN '11' THEN 'November'
    WHEN '12' THEN 'December'
    END AS month_name, Landing_Outcome, Booster_Version, Launch_Site
    FROM SPACEXTBL WHERE substr(Date, 7, 4) = '2015' AND Landing_Outcome = 'Failure (drone ship)';
```

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

month_name	Landing_Outcome	Booster_Version	Launch_Site

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

The ranked 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.
```

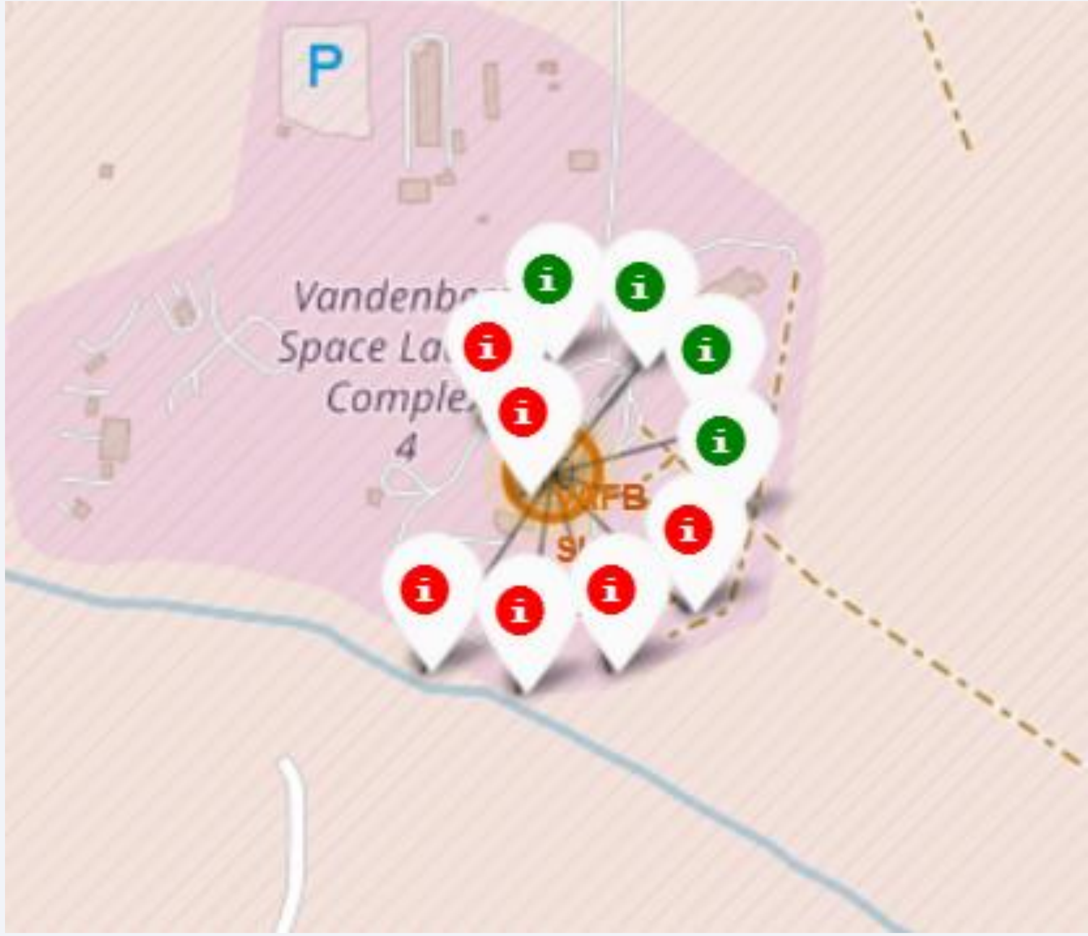
Landing_Outcome	outcome_count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

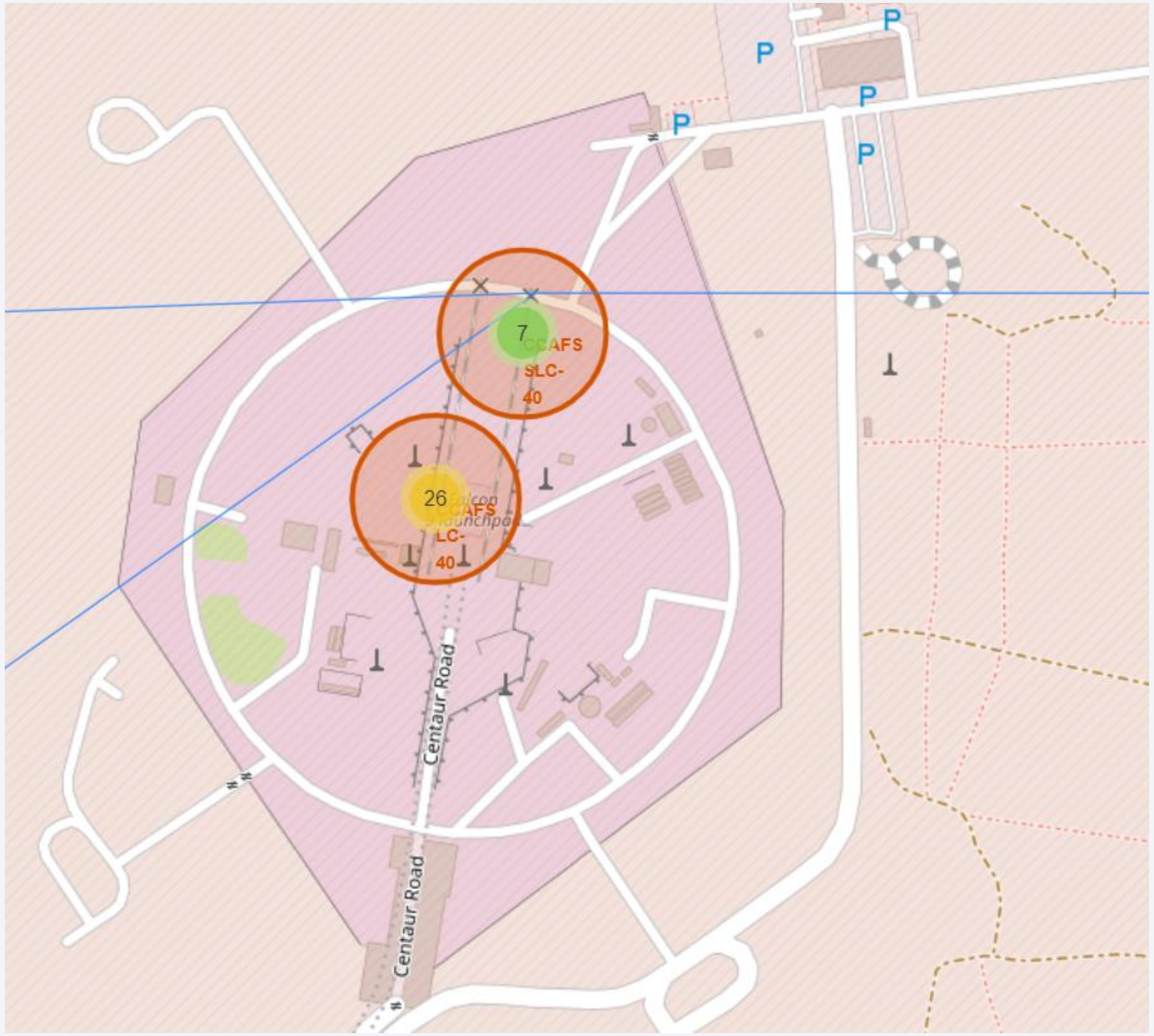
Section 3

Launch Sites Proximities Analysis

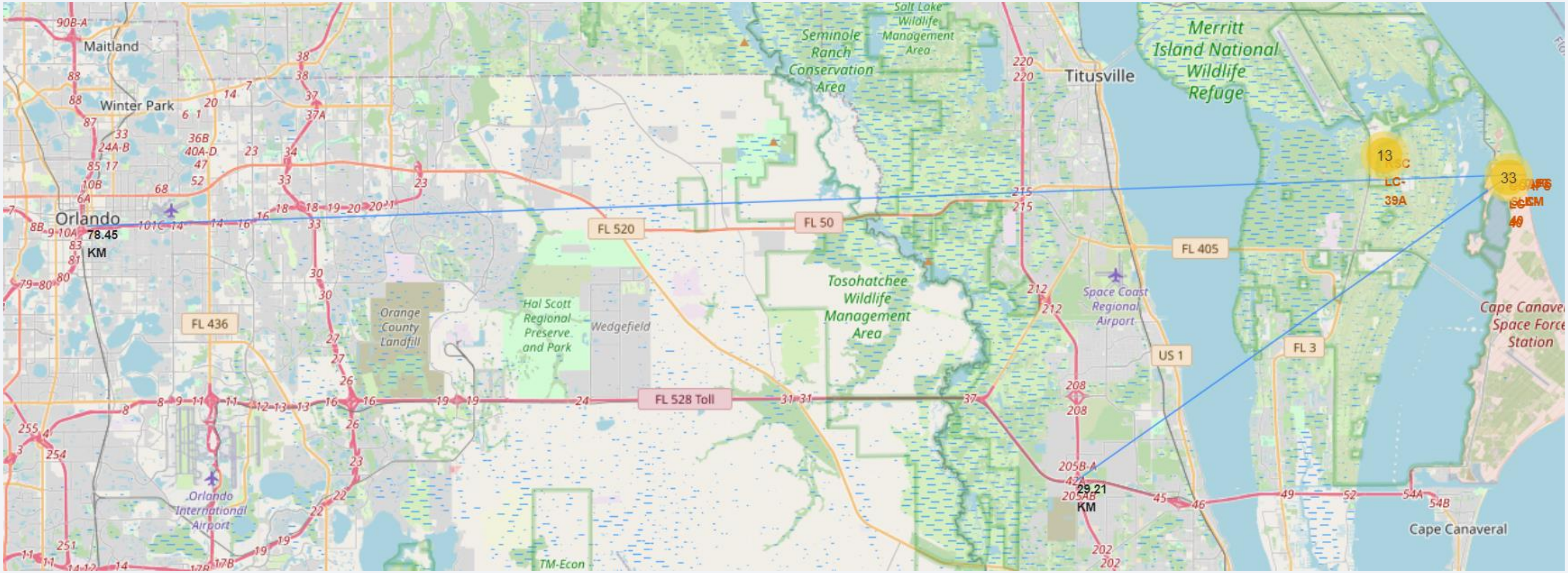
<Folium Map Screenshot 1>

<Folium Map Screenshot 2>





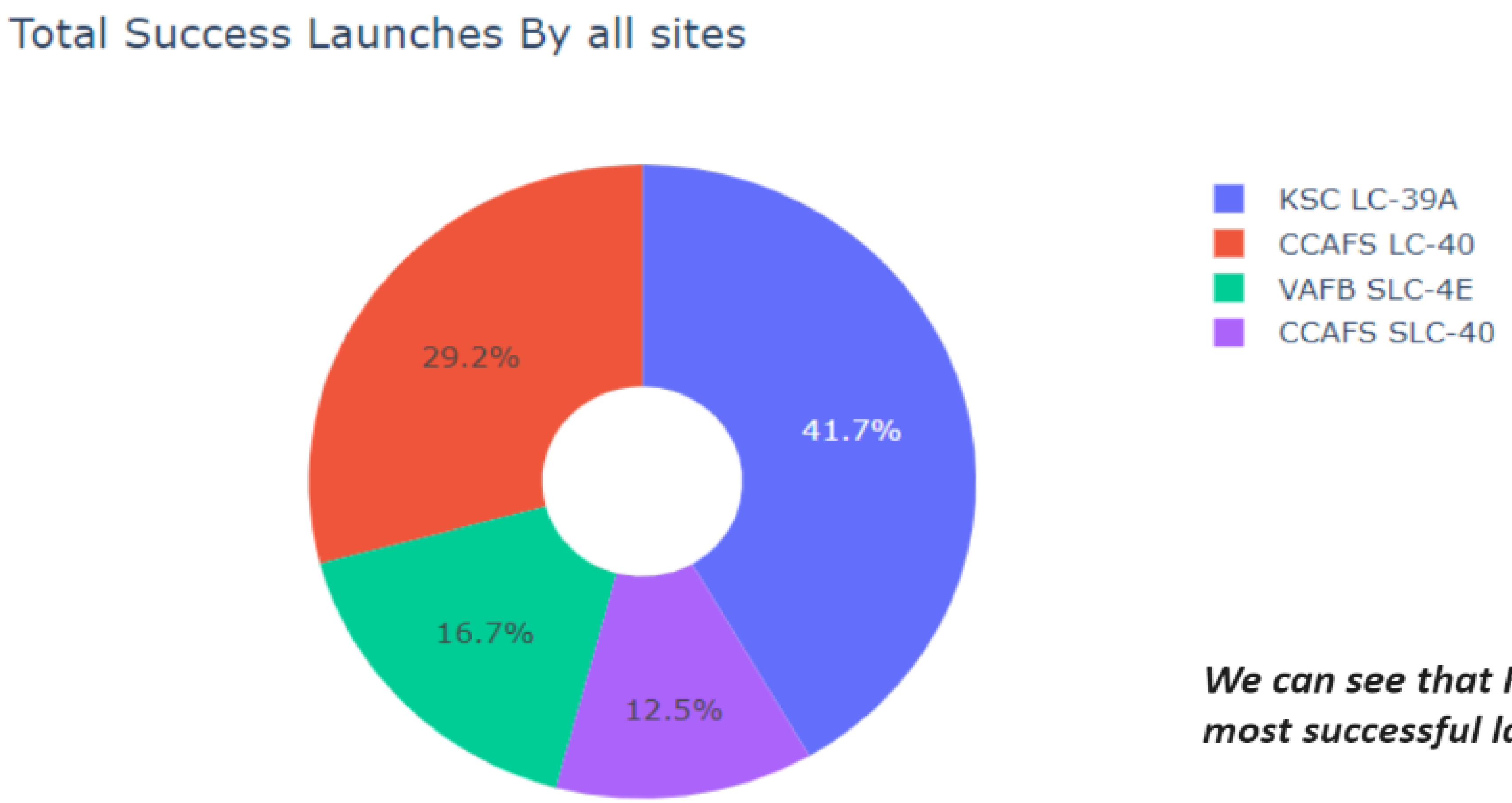
<Folium Map Screenshot 3>



Section 4

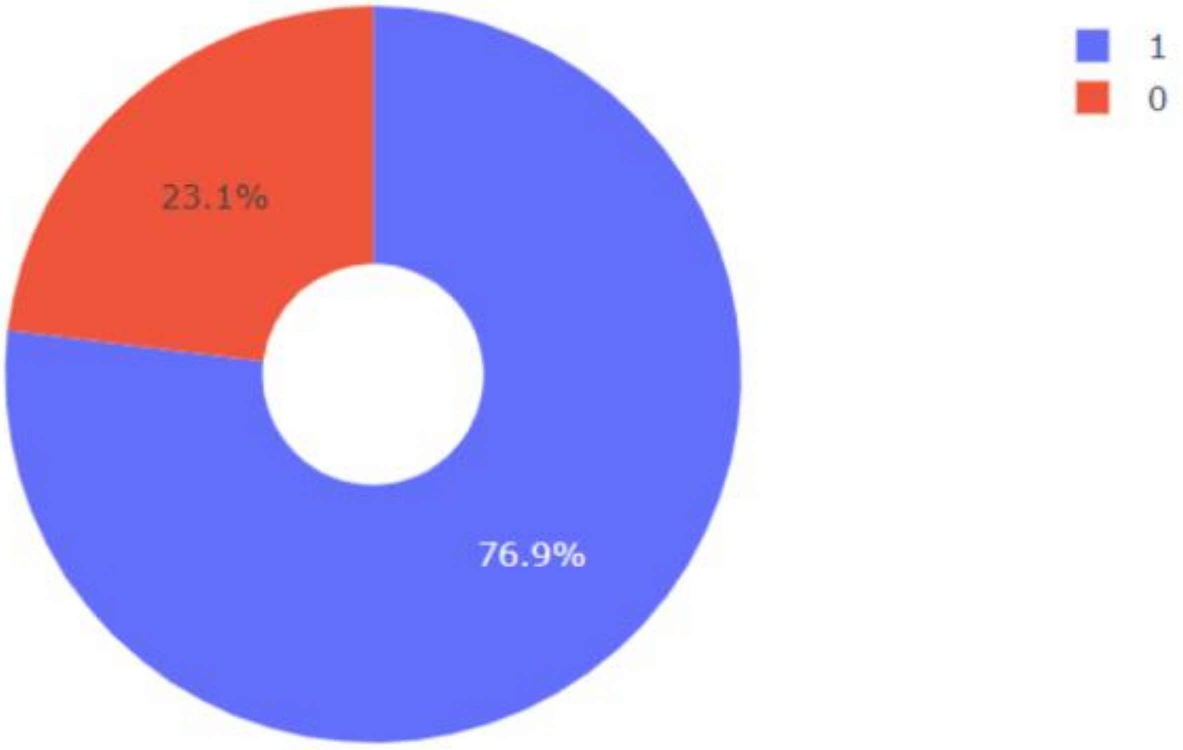
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>



We can see that KSC LC-39A had the most successful launches from all the sites

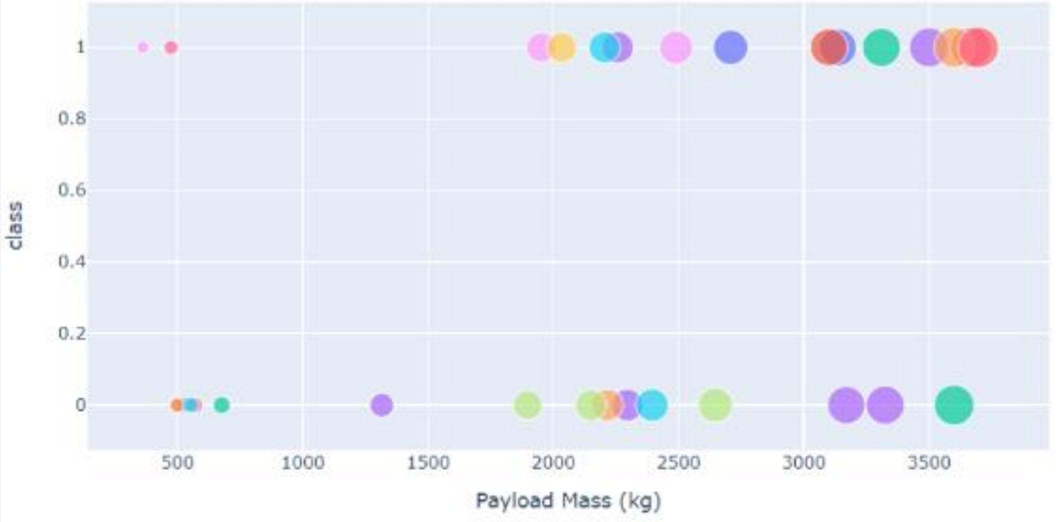
<Dashboard Screenshot 2>



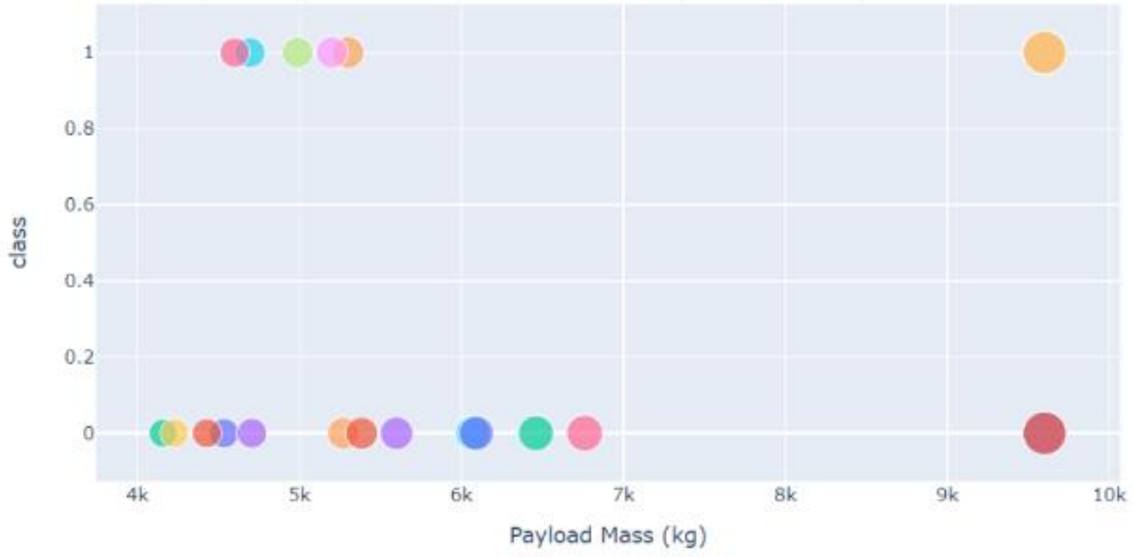
KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

<Dashboard Screenshot 3>

Low Weighted Payload 0kg – 4000kg



Heavy Weighted Payload 4000kg – 10000kg



We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

Section 5

Predictive Analysis (Classification)

Classification Accuracy

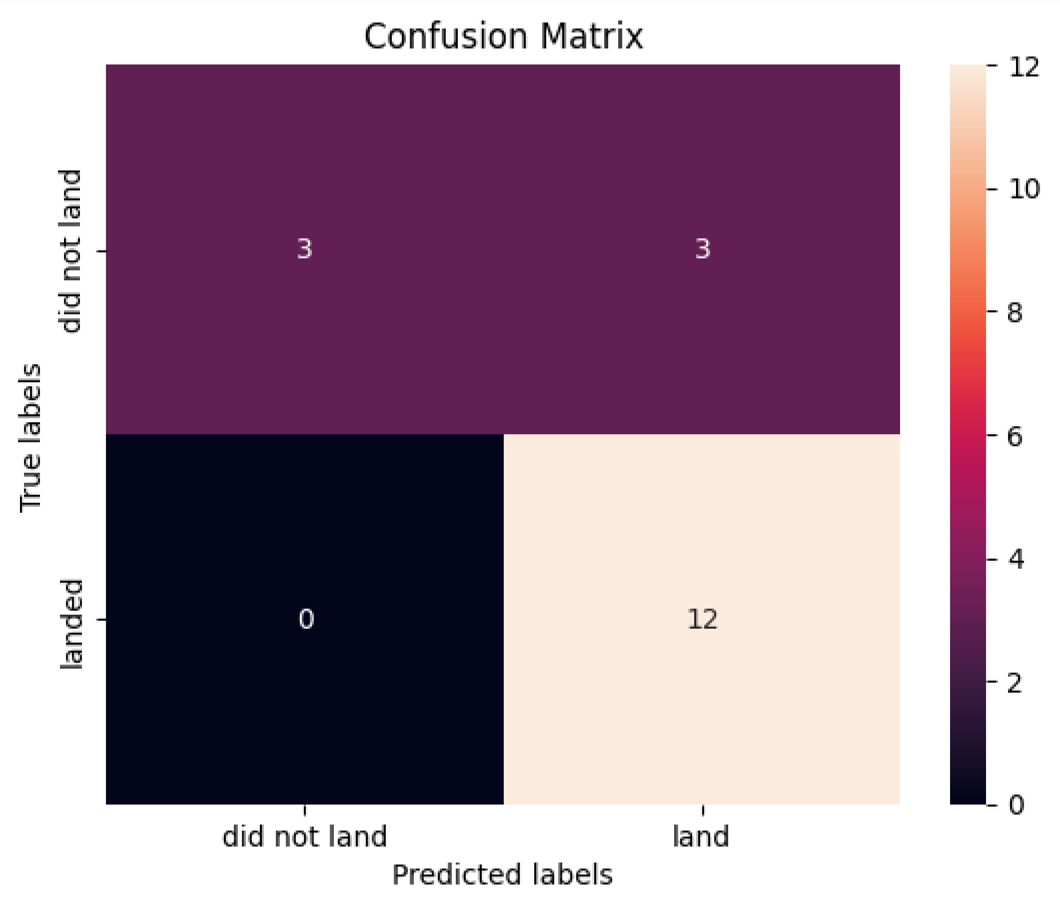
```
models = {'KNeighbors':knn_cv.best_score_,
              'DecisionTree':tree_cv.best_score_,
              'LogisticRegression':logreg_cv.best_score_,
              'SupportVector': svm_cv.best_score_}

bestalgorithm = max(models, key=models.get)
print('Best model is', bestalgorithm,'with a score of', models[bestalgorithm])
if bestalgorithm == 'DecisionTree':
    print('Best params is :', tree_cv.best_params_)
if bestalgorithm == 'KNeighbors':
    print('Best params is :', knn_cv.best_params_)
if bestalgorithm == 'LogisticRegression':
    print('Best params is :', logreg_cv.best_params_)
if bestalgorithm == 'SupportVector':
    print('Best params is :', svm_cv.best_params_)
```

```
Best model is DecisionTree with a score of 0.8732142857142856
Best params is : {'criterion': 'gini', 'max_depth': 6, 'max_features': 'auto', 'min_samples_leaf': 2, 'min_samples_split': 5, 'splitter': 'random'}
```

- The decision tree classifier is the model with the highest classification accuracy

Confusion Matrix

- The confusion matrix for the decision tree classifier shows that the classifier can distinguish between the different classes. The major problem is the false positives .i.e., unsuccessful landing marked as successful landing by the classifier.



Conclusions

We can conclude that:

- The larger the flight amount at a launch site, the greater the success rate at a launch site.
- Launch success rate started to increase in 2013 till 2020.
- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.
- KSC LC-39A had the most successful launches of any sites.
- The Decision tree classifier is the best machine learning algorithm for this task.

Thank you!