IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Roy K

06-05-2025

Outline

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

Executive Summary

- To predict whether SpaceX Falcon 9 first stage will land successfully using Machine Learning Algorithms

- **Summary of methodologies**:

  Data Collection (through API and Web Scraping), Data Wrangling, EDA, ML prediction using various models to select the best model and parameters

- Falcon 9 first stage will land successfully and Decision Tree is the best ML algorithm for prediction.

Introduction

- With advancement in technologies, many companies like Virgin Galactic and SpaceX are trying to make space travel more accessible. However in case of failed experiments, the environmental and financial loss is huge.
- SpaceX's Falcon 9 rocket is highlighted for its cost-effectiveness. The first stage of the rocket can be reused.
- **Problems you want to find answers**
  - In this project we want to predict whether the Falcon 9's first stage landing successfully, which will help inform launch pricing.
  - Different ML models are used to find the most suitable model for prediction.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - API and Web Scraping
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Numpy, Pandas and SQL. Bar plot, scatter plots are used for visulaizaytion.
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Several ML models like KNN, SVM, Decision Trees with different set of parameters are used. The Best set of parameters are found out based on accuracy of the model

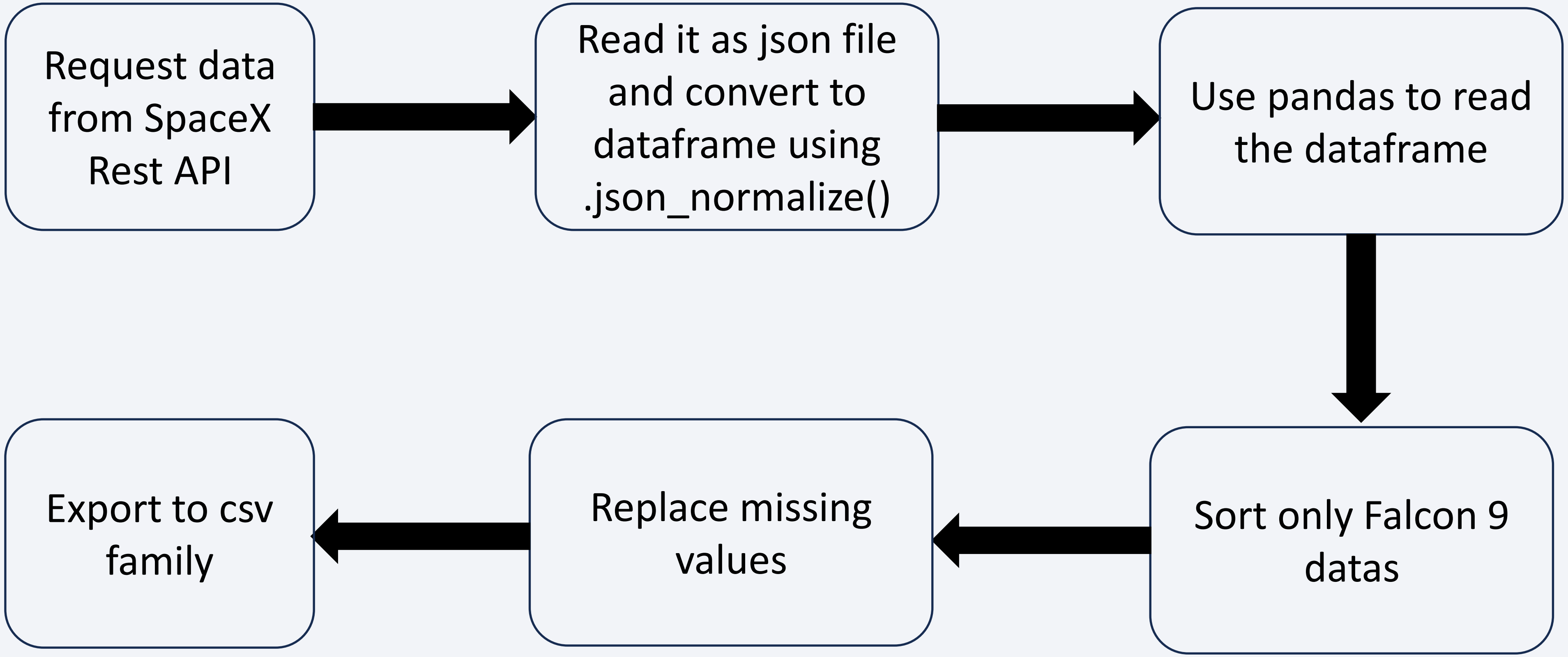
Data Collection

- Data is collected through API and Web scraping. In the following the data collection steps are described.

- Data is collected from Wikipedia using Web Scraping

Data Collection – SpaceX API

- Data is collected using SpaceX API

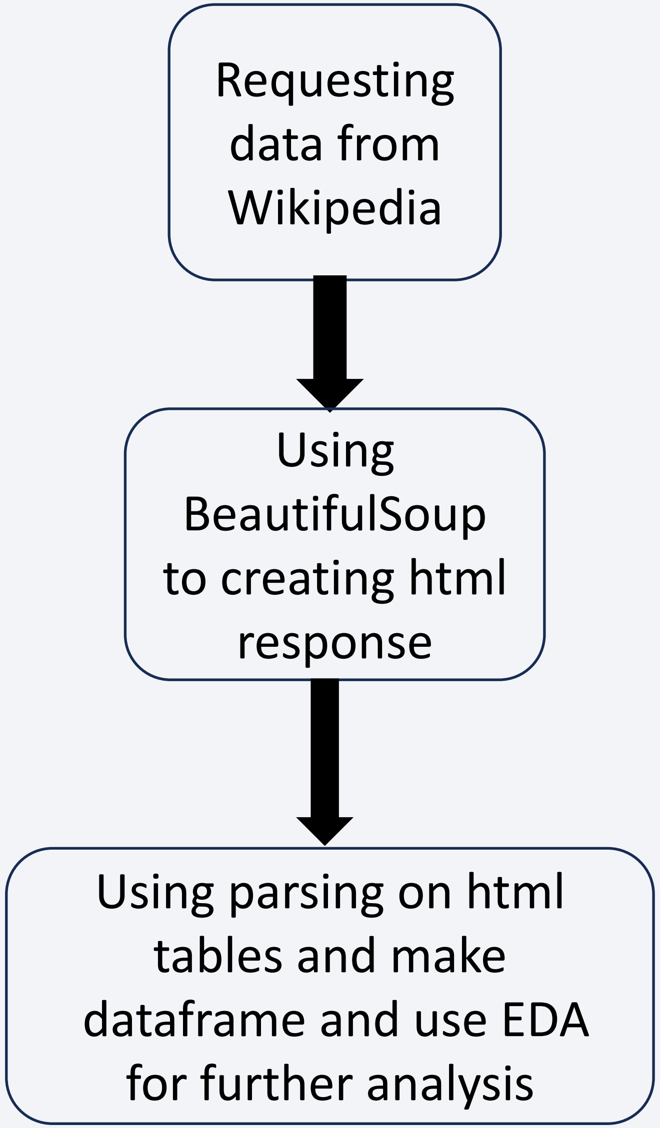


https://github.com/Git-create35/IBM-Applied-DS-Capstone/blob/main/jupyter-labs-spacex-data-collection-api%20(1).ipynb

Data Collection - Scraping

- Data is collected from Wikipedia via WebScraping.

- https://github.com/Git-create35/IBM-Applied-DS-Capstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- The dataset includes cases of rocket booster landings with outcomes labeled as success or failure.
- A “True” label indicates a successful landing, while “False” indicates a failure.
- These outcomes are converted into labels: 1 for success and 0 for failure.
- Landing outcomes are labeled as the ocean, a ground pad (RTLS), or a drone ship (ASDS).

https://github.com/Git-create35/IBM-Applied-DS-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling%20(1).ipynb

Calculating number of launches on each site

↓

Calculate the number and occurrence of each targetted orbit

↓

Calculate the number of landing outcomes

↓

Create landing outcome label

EDA with Data Visualization

- relationship between Flight Number and Launch Site
- relationship between Payload Mass and Launch Site
- relationship between success rate of each orbit type
- relationship between Flight Number and Orbit type
- relationship between Payload Mass and Orbit type
- Visualize the launch success yearly trend

https://github.com/Git-create35/IBM-Applied-DS-Capstone/blob/main/edadataviz%20(1).ipynb

EDA with SQL

- Connect to the dataset
- Display names of the unique launch sites
- 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
- Find the date when the first successful landing outcome in ground pad was achieved.
- List the names of the boosters having success in drone ship and have payload mass greater than 4000 but less than 6000
- List all the booster versions which have the maximum payload mass.
- Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order.

https://github.com/Git-create35/IBM-Applied-DS-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite%20(2).ipynb

Build an Interactive Map with Folium

- Create map object with NASA Johnson Space Center at Houston, Texas as centre location. Used Folium circle marker with text.
- Add marker to each launch site using latitude and longitude.
- Mark the success/failed launches for each site on the map
- Use mouse position marker to calculate the distance between the coastline point and the launch site.

https://github.com/Git-create35/IBM-Applied-DS-Capstone/blob/main/lab_jupyter_launch_site_location%20(1).ipynb

Build a Dashboard with Plotly Dash

- Added a dropdown list for Launch Site selection.
- Added a pie chart to show the total successful launches count for all sites and the success vs. failure for specific launch site.
- Added a slider to select Payload range.
- Added a scatter chart to show the correlation between Payload Mass and Success rate

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

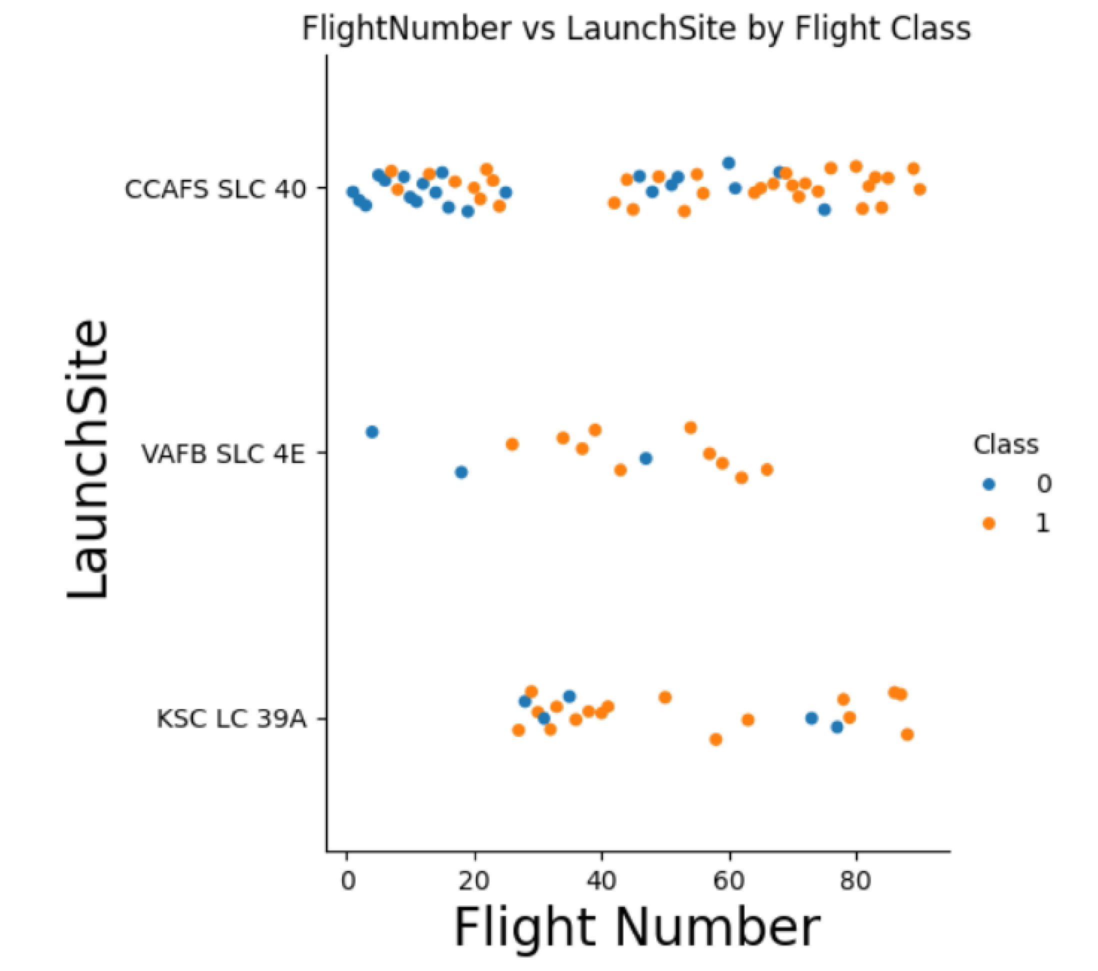
Results

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

Section 2

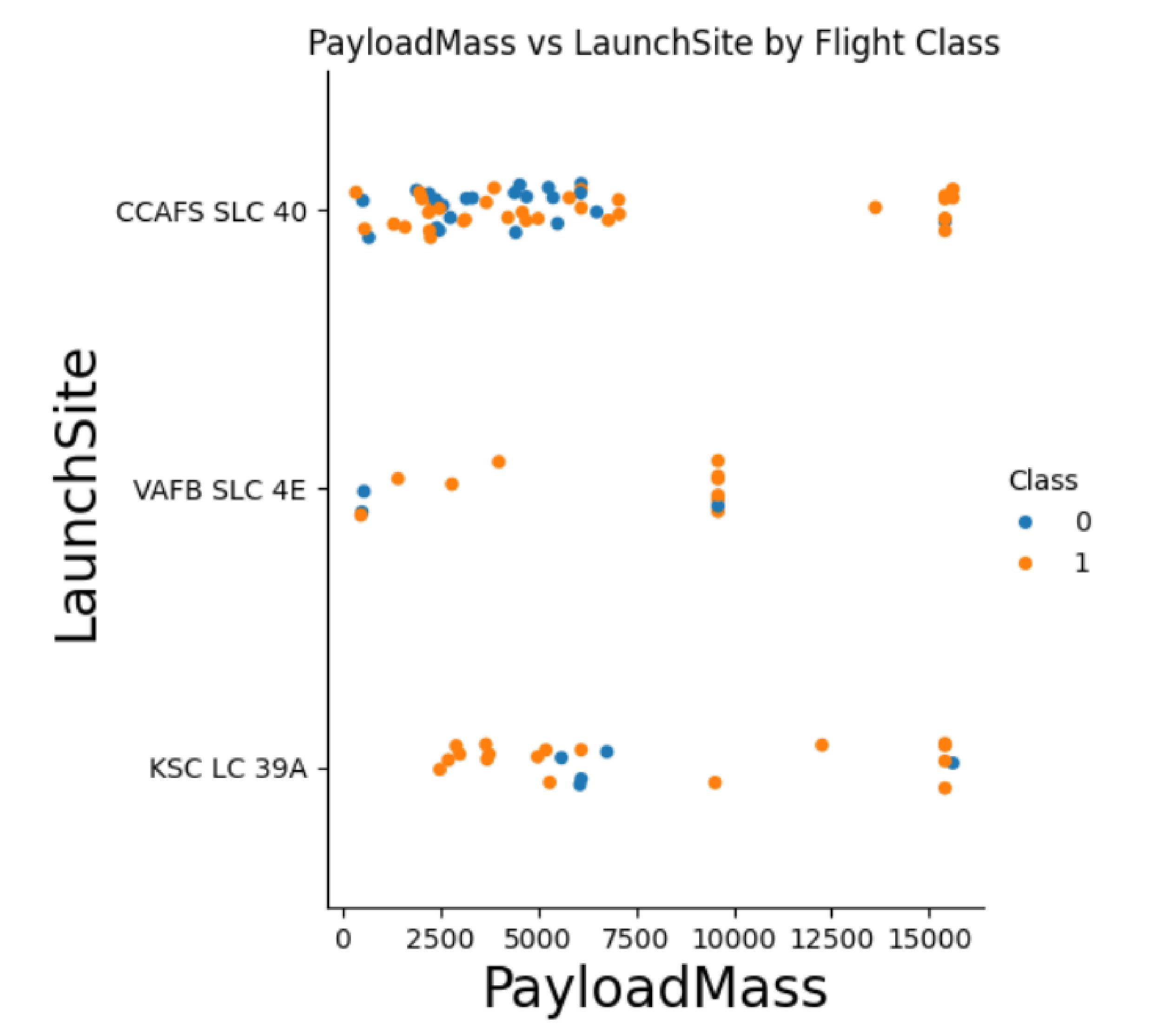
Insights drawn from EDA

Flight Number vs. Launch Site



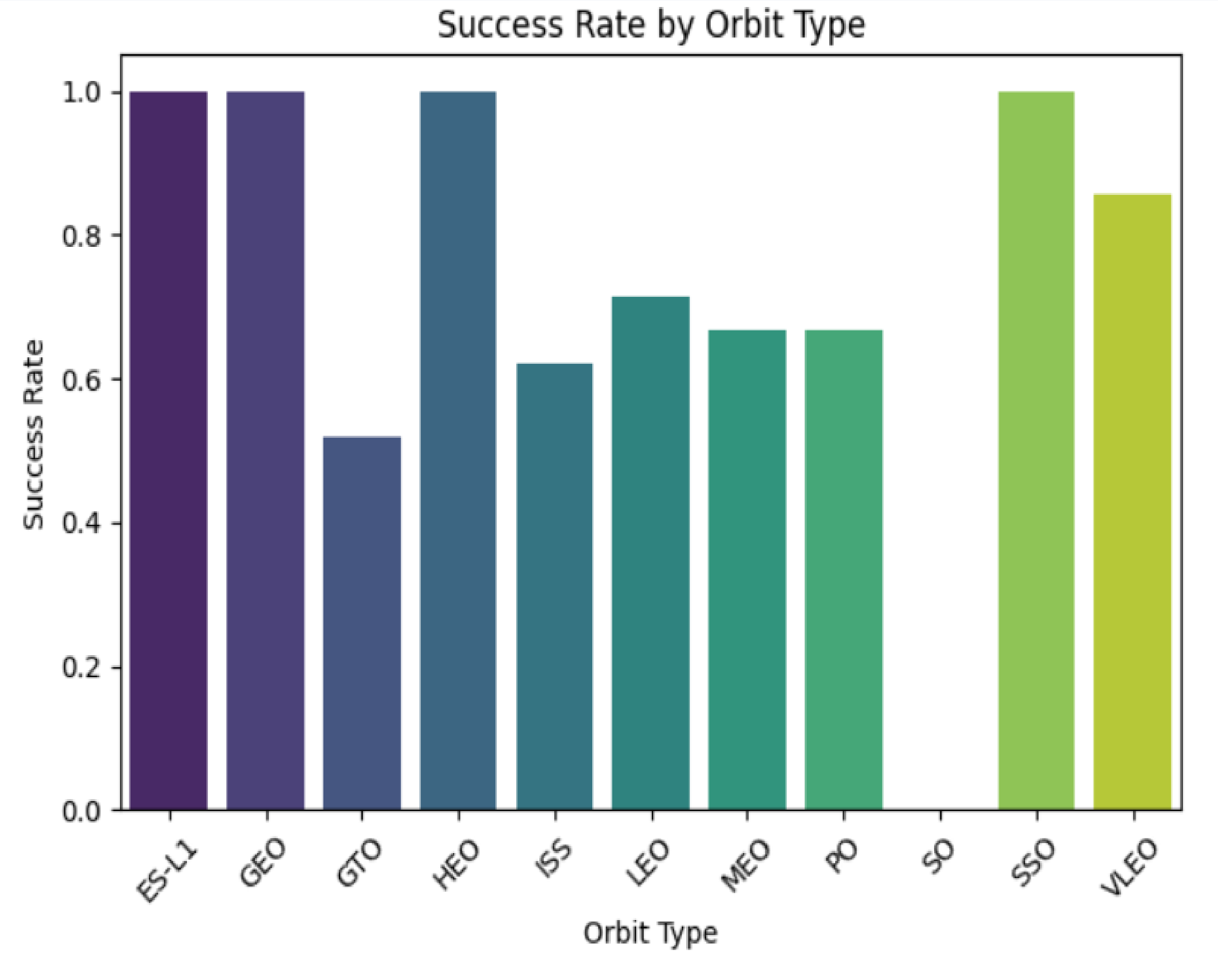
1. Earlier flights mostly resulted in failure.
2. CCAFS SLC 40 launch site has about a half of all launches.
3. VAFB SLC 4E and KSC LC 39A have higher success rates.

Payload vs. Launch Site



1. Higher payload mass mostly results in success
2. VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type

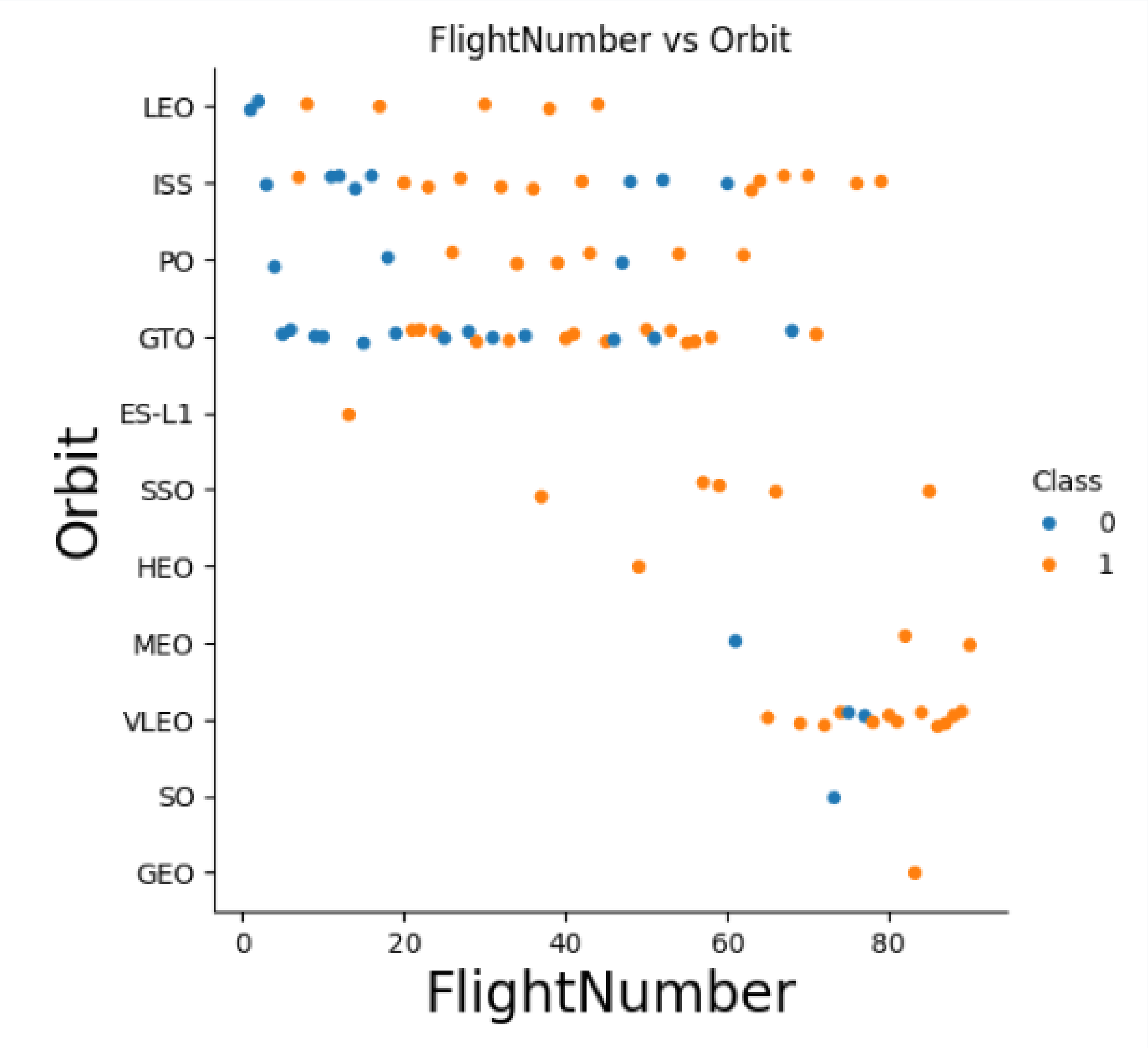


1. ESL1, GEO, HEO, SSO has almost 100% sucess rate.

2. SO has no success.

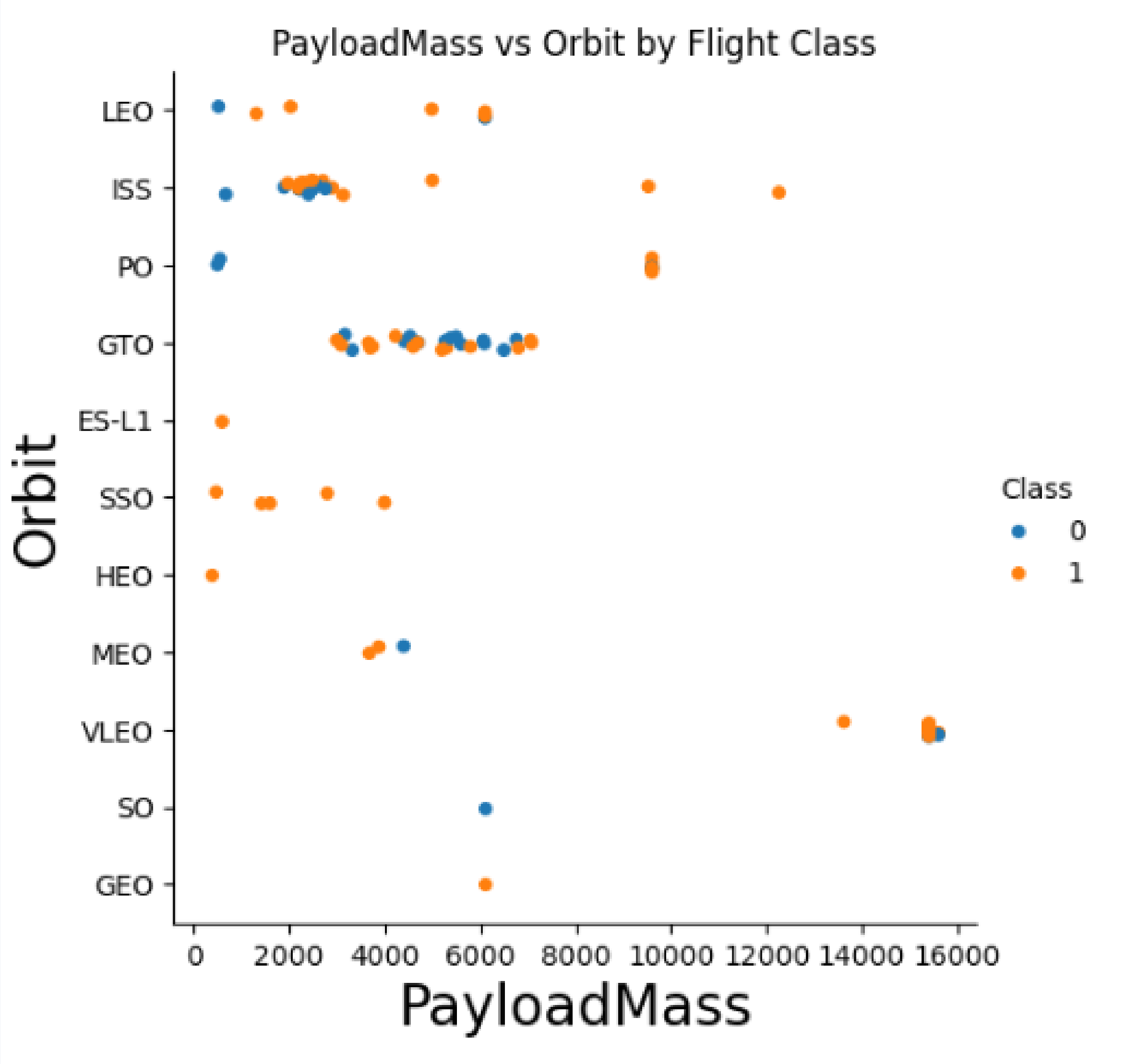
3. GTO, ISS, LEO, MEO, PO, VLEO also has success rate between 50-70%

Flight Number vs. Orbit Type



- in the LEO orbit, success seems to be related to the number of flights.
- in the GTO orbit, there appears to be no relationship between flight number and success.
- SSO, HEO, VLEO, MEO have higher flight number for success

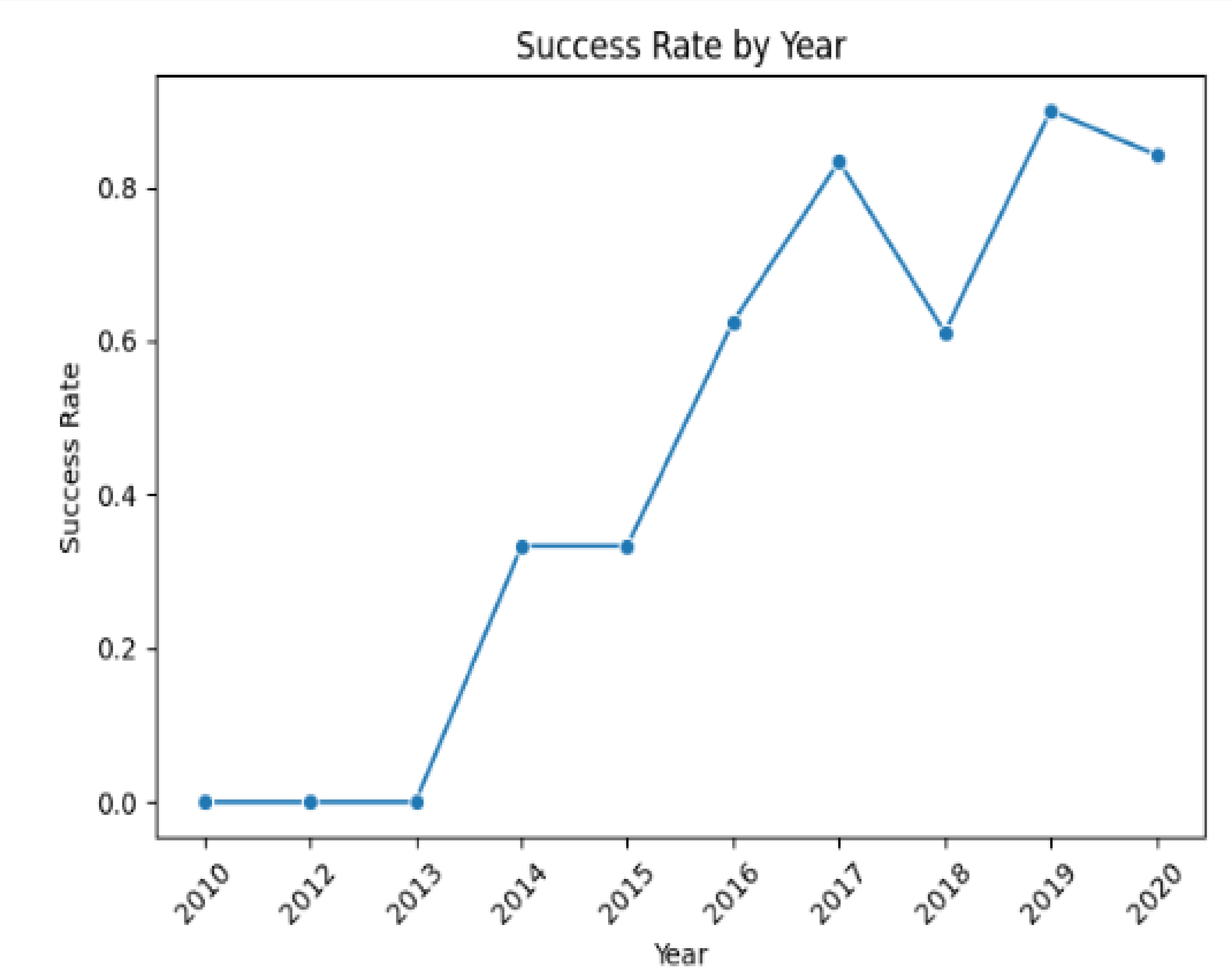
Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend



- the sucess rate since 2013 kept increasing till 2020

EDA with SQL

All Launch Site Names

```
%sql SELECT DISTINCT Launch_Site FROM my_data
```

```
 * 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'

```
%sql SELECT * FROM my_data 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

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

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

SUM(PAYLOAD_MASS__KG_)
45596

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

Average Payload Mass by F9 v1.1

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

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

AVG(PAYLOAD_MASS__KG_)
2928.4

- Displaying the average payload mass carried by F9 v1.1

First Successful Ground Landing Date

```
%sql SELECT MIN("Date") AS first_successful_landing_date FROM "SPACEXTBL" WHERE "Mission_Outcome" = 'Success' AND "Landing_Outcome" = 'Success
```

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

first_successful_landing_date
2015-12-22

- The first successful ground landing date is displayed

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT "Booster_Version" FROM my_data WHERE "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000 AND Landing_Outcome = 'Success (dr
```

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

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

Only four booster version is there whose payload mass between 4000 and 6000 and outcome is success

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT Mission_Outcome, COUNT(*) as total_number FROM my_data Group By "Mission_Outcome"
```

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

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

- 99 success, 1 failure and 1 payload status unclear.

Boosters Carried Maximum Payload

```
%sql SELECT "Booster_Version" FROM my_data WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM my_data)
```

```
 * 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

- Listing the names of the booster versions which have carried the maximum payload mass

2015 Launch Records

```
%sql SELECT strftime('%m', "Date") AS "Month","Landing_Outcome","Booster_Version","Launch_Site" FROM my_data WHERE strftime('%Y', "Date") = '2015' AND "Landing_Outcome" = 'F
```

```
 * 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

- Two failures in January and April of 2015.

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

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

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

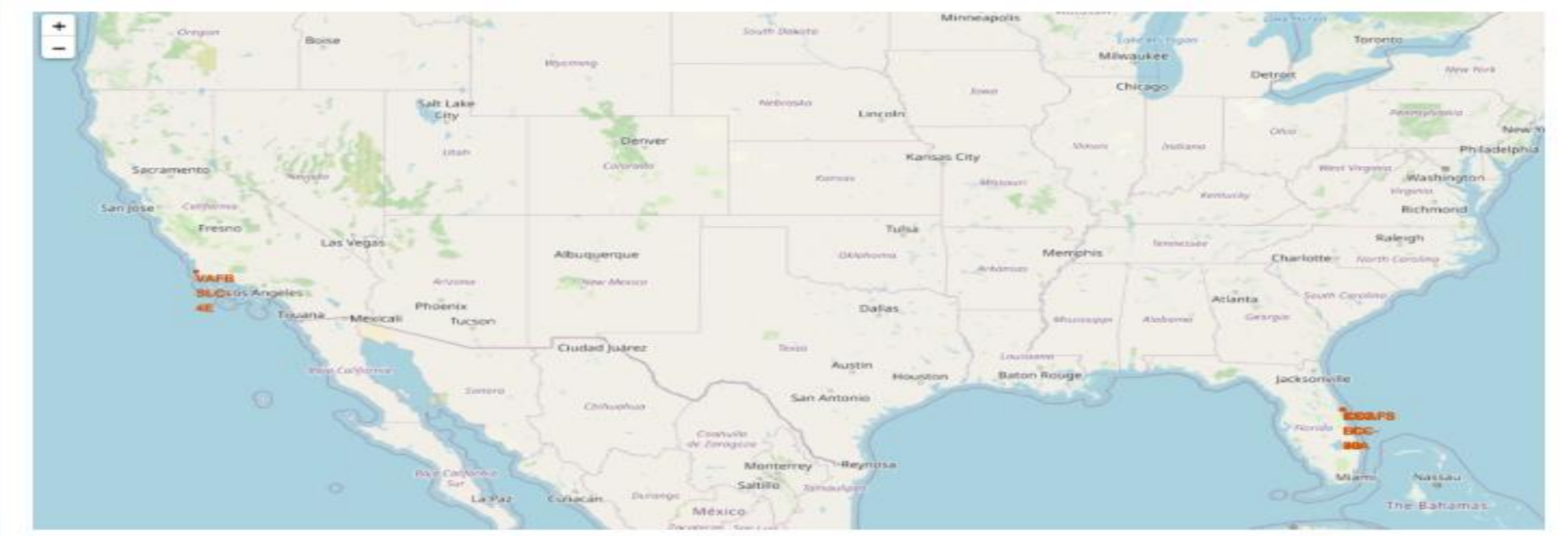
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

- Ranking 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

Section 3

Launch Sites Proximities Analysis

<Folium Map Screenshot 1>

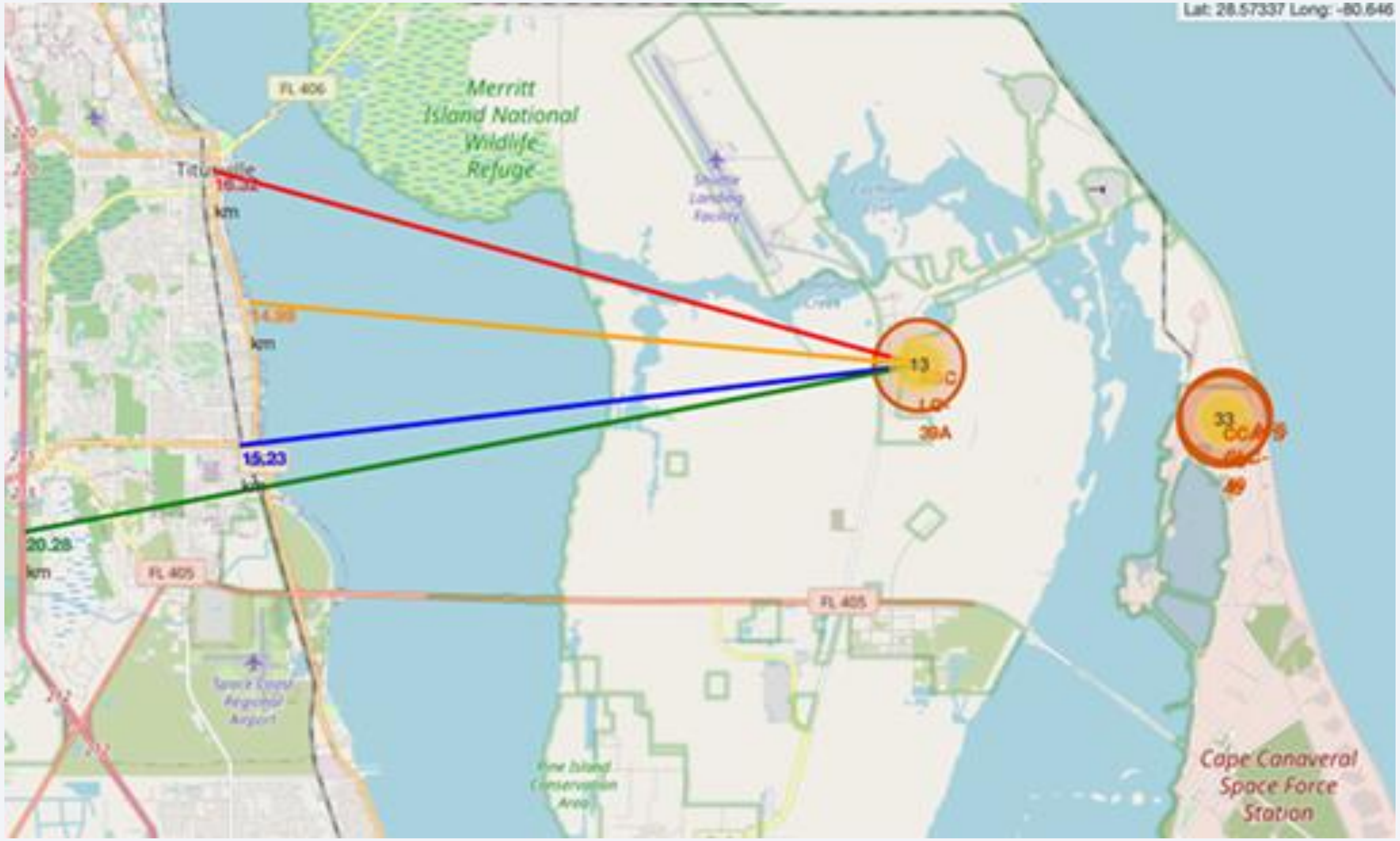


- Most of Launch sites are in proximity to the Equator line.
- All launch sites are in very close proximity to the coast

<Folium Map Screenshot 2>

- Green markers show successful launch
- Red markers show failure

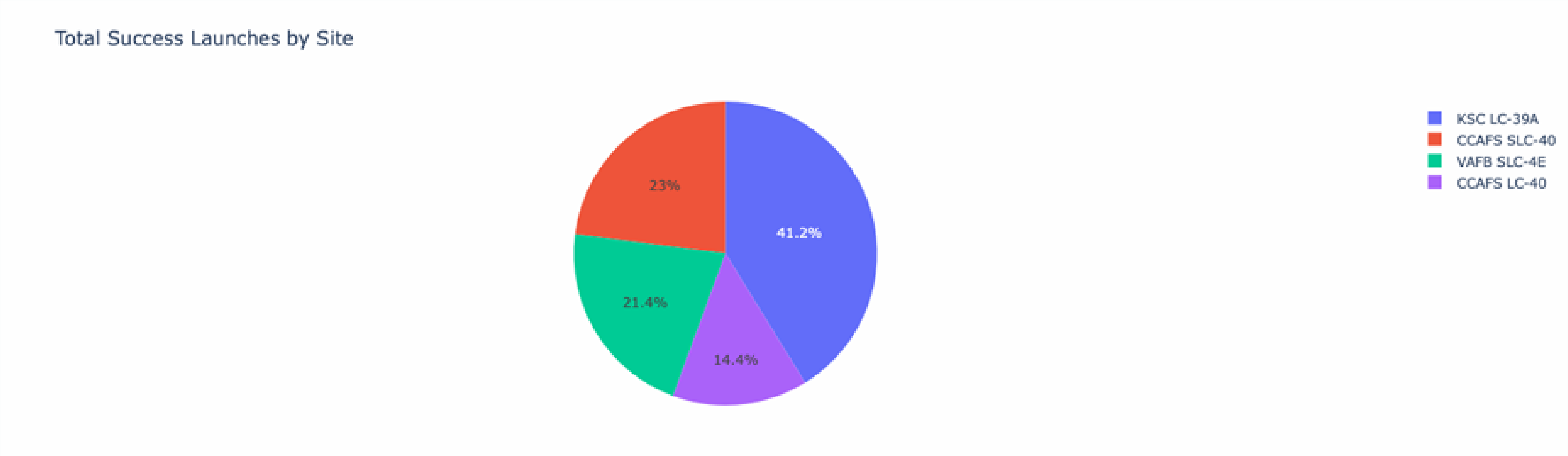
<Folium Map Screenshot 3>



Section 4

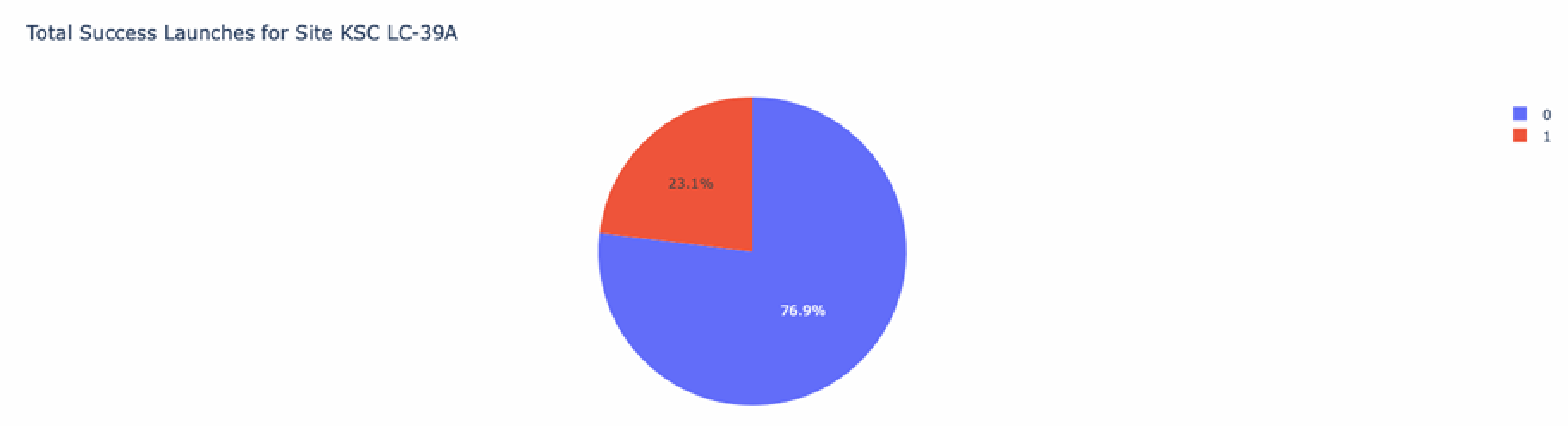
Build a Dashboard with Plotly Dash

Total success launch by site



- KSC LC 39 A has the highest success rate among all sites

Total success for Site KSC LC 39A



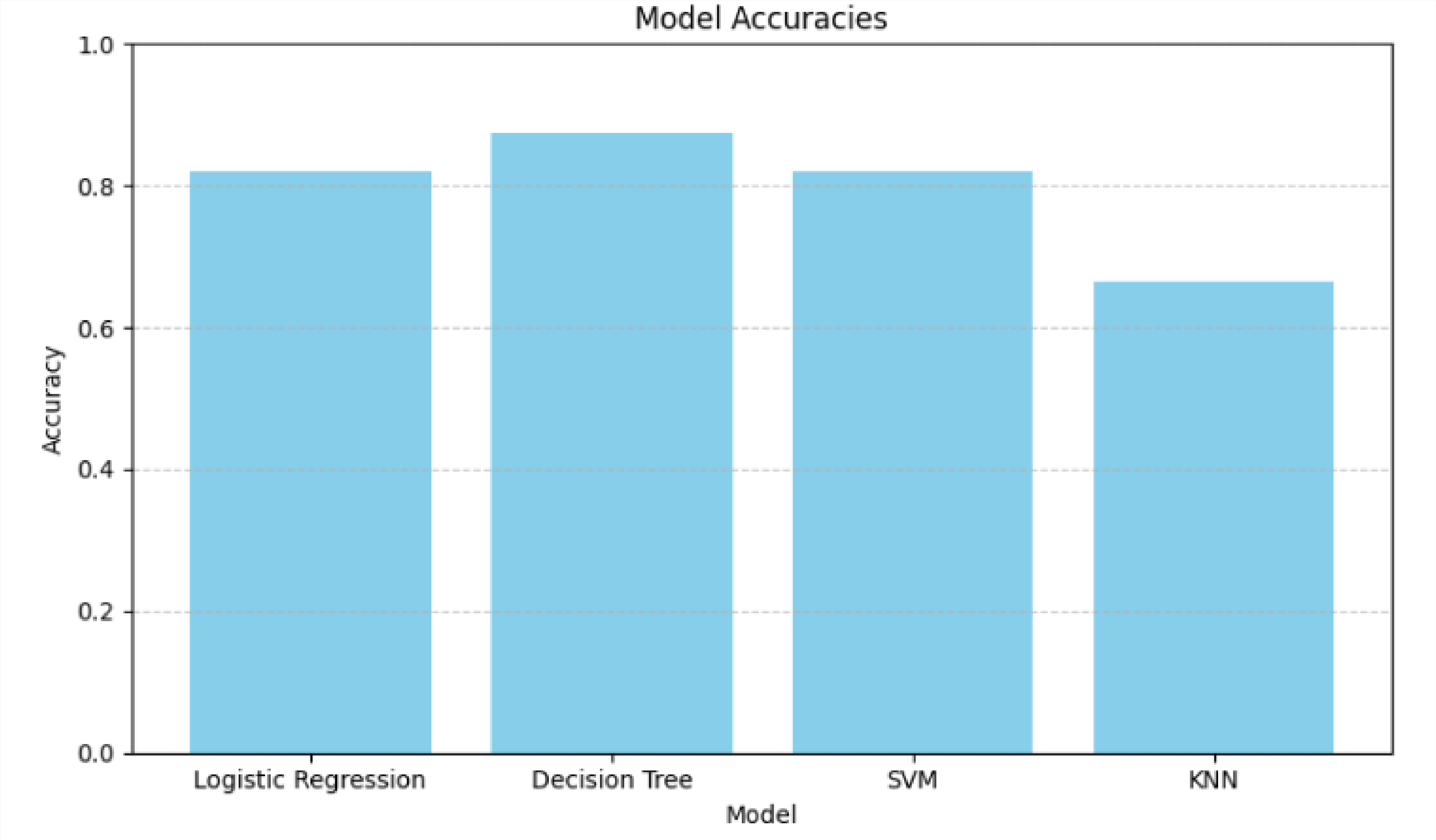
- KSC LC 39 A has a success rate of 77%

Correlation between payload m and success rate of sites

Section 5

Predictive Analysis (Classification)

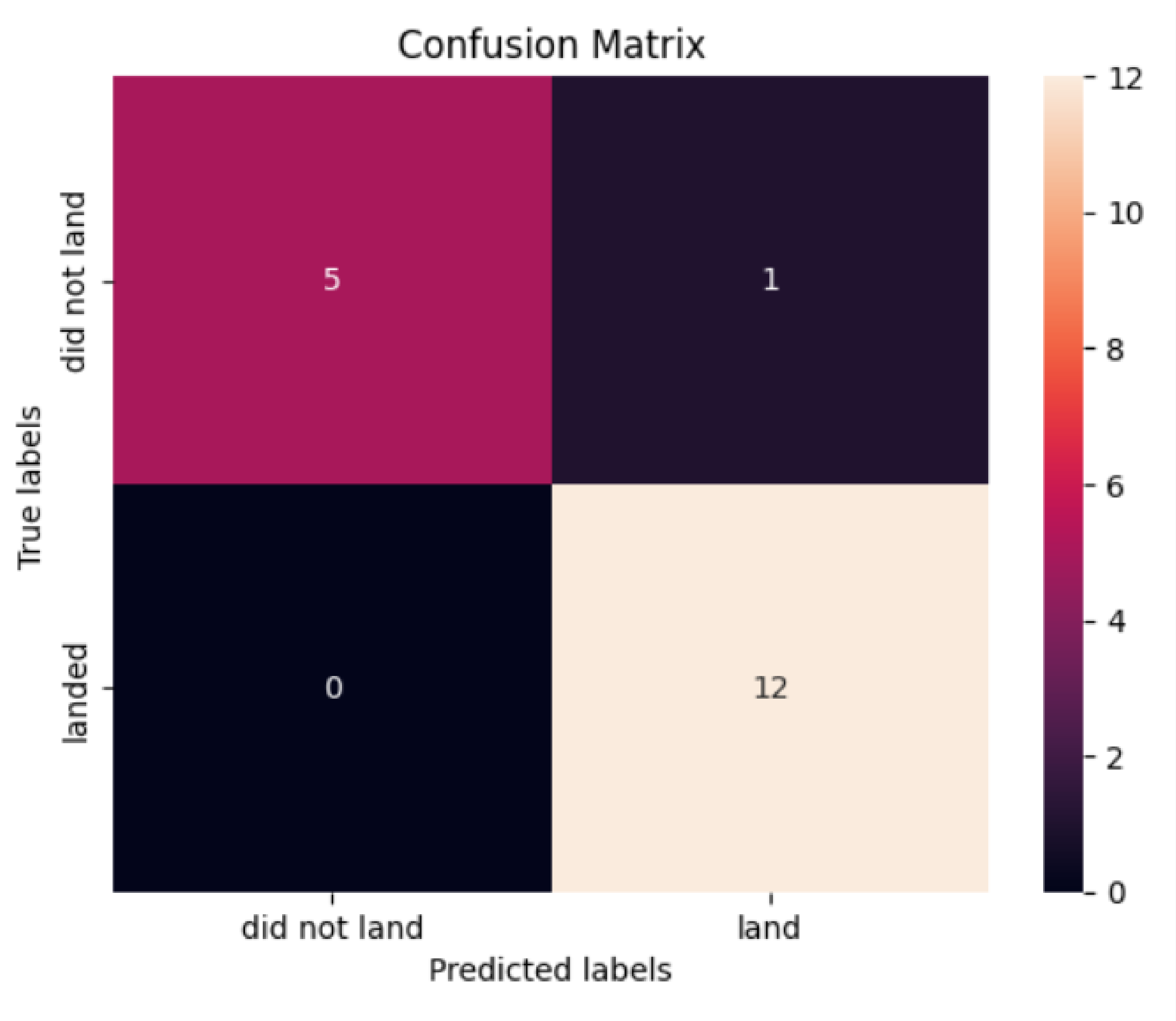
Classification Accuracy



- Decision tree is the best classifier

Confusion Matrix

- Confusion matrix for the Decision Tree classifier.
- There is only 1 misclassification



Conclusions

- Higher payload mass results in successes
- ESL1, SSO, GEO, HEO orbits have 100% success rate
- SO has 0% success
- SSO, HEO, VLEO, MEO have higher flight number for success
- SSO has higher success rate with low payload mass
- Success rate has increased over the years.
- Decision tree classifier is best performing with very few misclassification.

Thank you!