IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Yago P
24-July-2025

Outline

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

Executive Summary

- Summary of methodologies
  - 1. Data Collection from API
  - 2. Webscraping from Wikipedia
  - 3. Data wragling
  - 4. Data Analysis with SQL
  - 5. Data Analysis with Pandas and Matplotlib
  - 6. Machine Learning Prediction
- Summary of all results
  - 1. Collected data from API and Wikipedia
  - 2. Exploratory Data Analysis outputs
  - 3. Predictive Findings

Introduction

- Project background and context
    - The objective of this project is to predict the success rate of Falcon 9 rocket landings. analyzing historical data, we can understand which factors influence landing success. To achieve this goal, we will collect data from various sources, such as the SpaceX API and Wikipedia, and develop a machine learning model to predict the likelihood of a successful landing. With this model, we hope to provide insights that can help optimize costs and improve decision-making for space launches.
- Problems you want to find answers:
    - Aspects that influence successful of SpaceX rockets landings
    - Rate of successful landings over the years
    - Which Machine learning model brings better outcomes

Section 1

Methodology

Methodology

Executive Summary

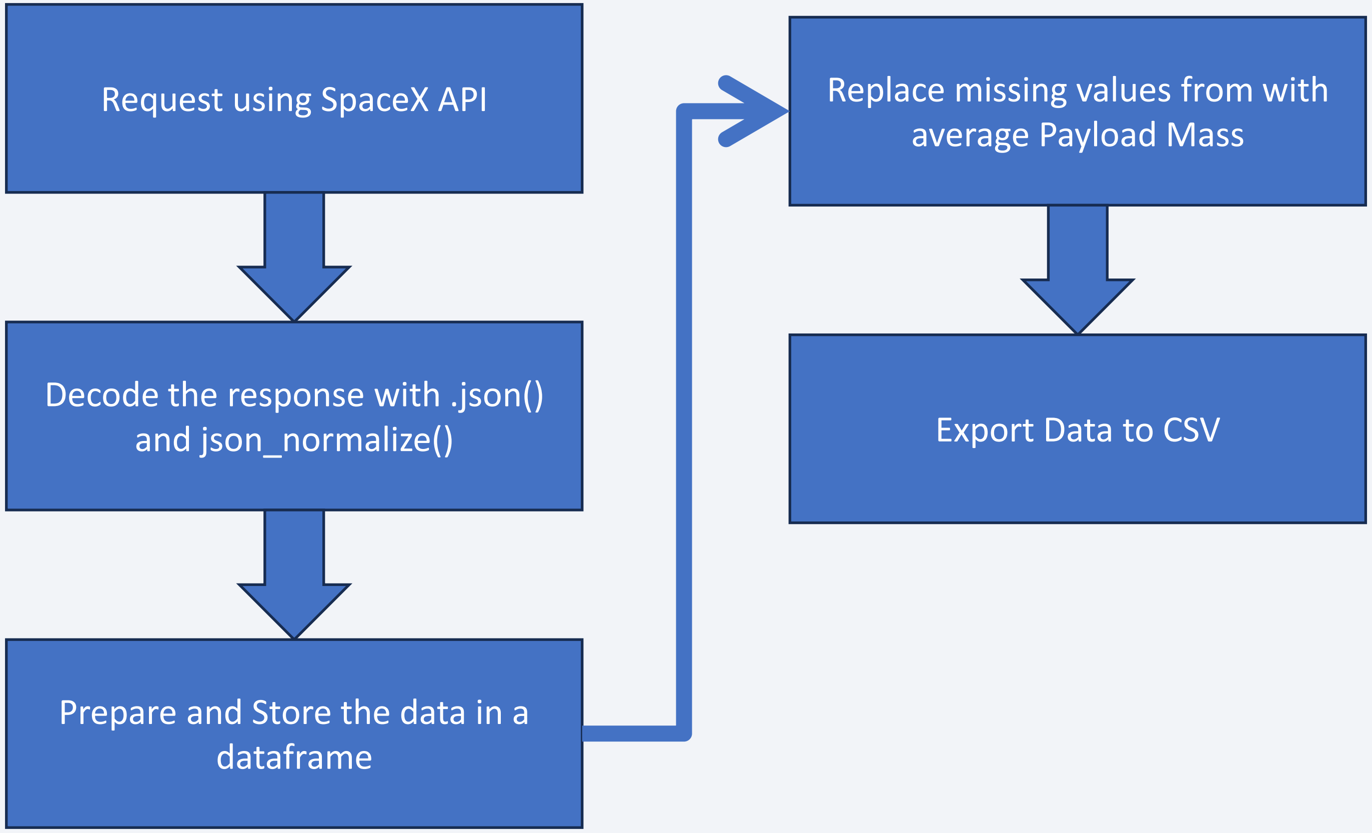
- Data collection methodology:
  - Using SpaceX Api and Webscraping from wikipedia
- Perform data wrangling
  - Filtering, dropping null values, and transforming the data in binary format
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

- Data was collected in two different ways:
  1. Requested data directly from the SpaceX API
  2. Web Scraping from Wikipedia

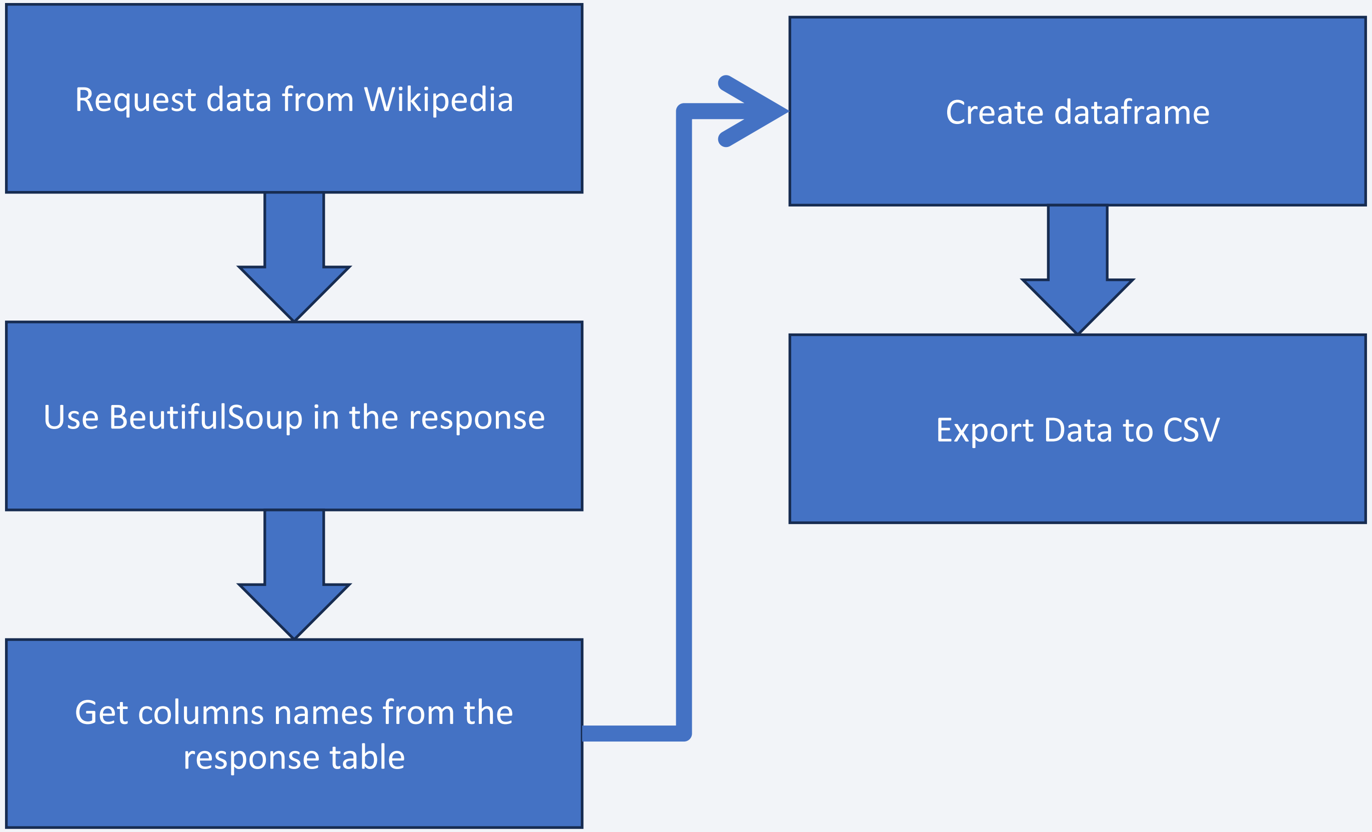
Data Collection – SpaceX API

- GitHub: [Data Collection]



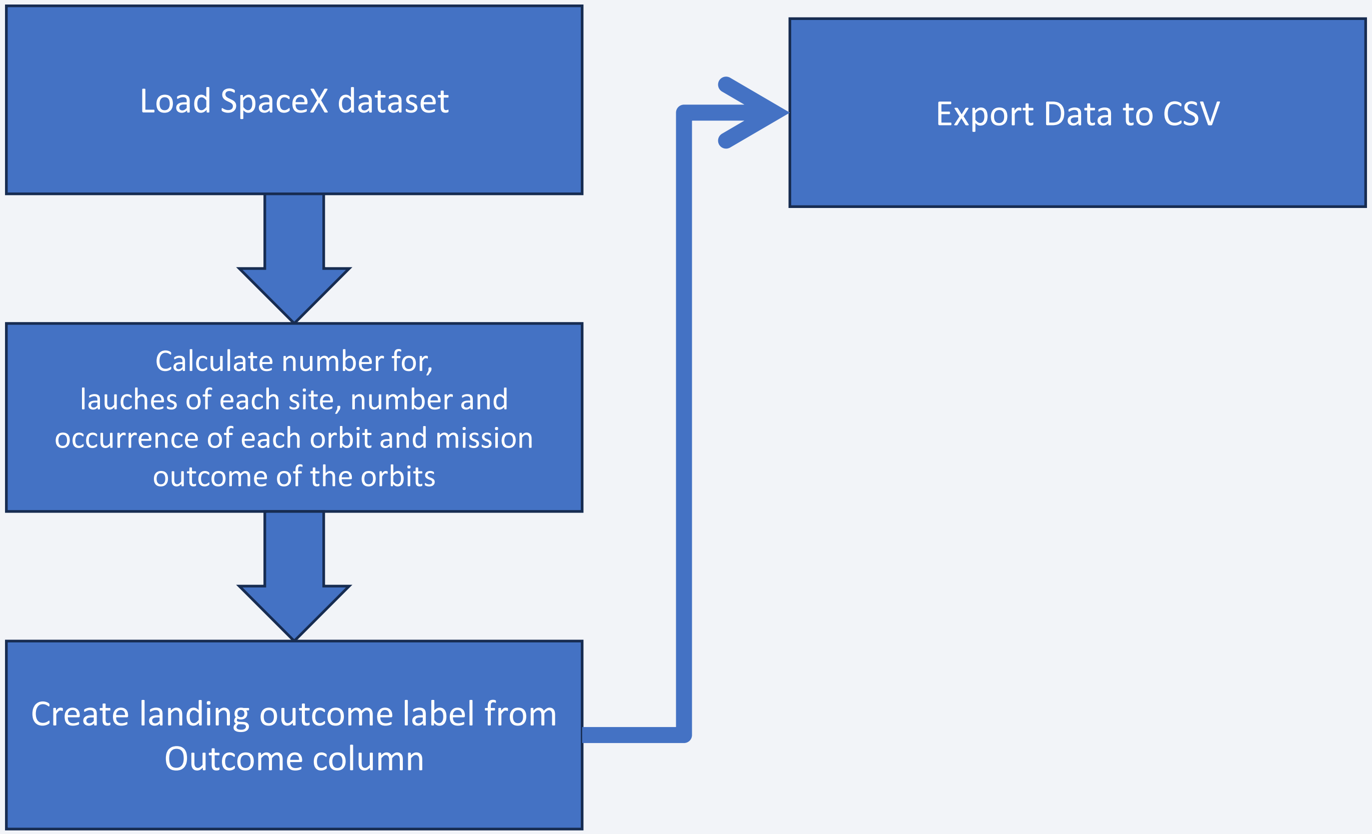
Data Collection - Scraping

- GitHub: Webscraping



Data Wrangling

- GitHub: [Data Wrangling]



EDA with SQL

- SQL queries
  - Display the names of the unique launch sites in the space mission
  - Display 5 records from 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 achieved
  - 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 all the Booster_Versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function
  - 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: [EDA SQL]

EDA with Data Visualization

- Charts
  - Flight Number and Launch Site
  - Flight Number and Payload
  - Payload Mass and Launch Site
  - Payload Mass and Orbit type

- Using bar charts to see comparison and scatter plots to see relationship between the variables.

- GitHub: [EDA Data Viz]

Build an Interactive Map with Folium

- Mark all launch sites on a map
  - First create a marker of NASA Johnson space center and then create markers for all the launches sites
- Mark the success/failed launches for each site on the map
  - Create Green markers for successful launches and red markers for failed launches
- Calculate the distances between a launch site to its proximities
  - Closest to the hailway
  - Closest to the railway
  - Closest to the city
  - Closest to the Coast

- GitHub: Launch Sites Locations Analysis with Folium

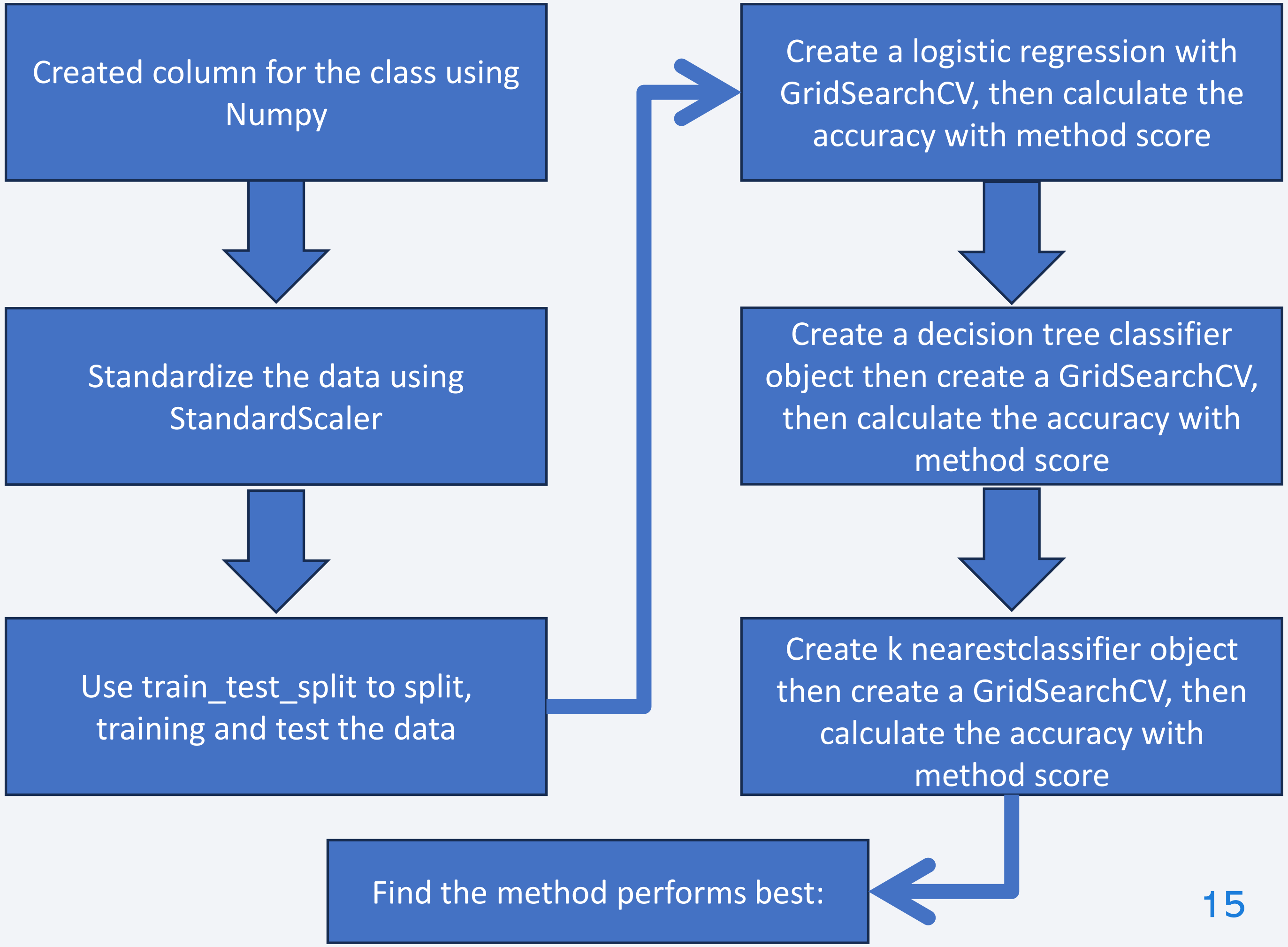
Build a Dashboard with Plotly Dash

- In the dashboard, we first create a dropdown containing all available launch sites. Selecting a specific launch site updates the information displayed in the charts.
- The dashboard includes a pie chart showing the success rate of launches, and a scatter plot to analyze the correlation between payload and launch success. A slider is also used to adjust the payload range.

- GitHub: Dash Plotly

Predictive Analysis (Classification)

- GitHub: [Machine Learning Prediction]



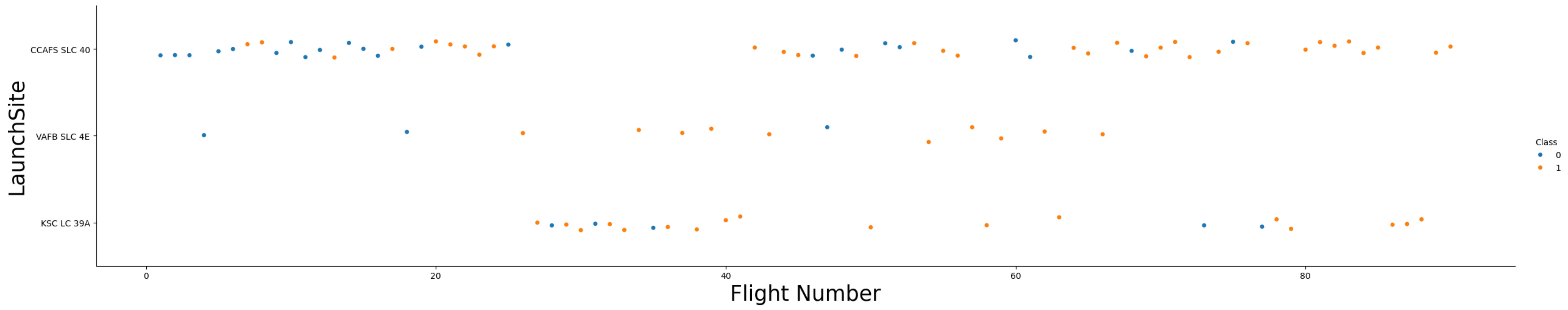
Results

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

Section 2

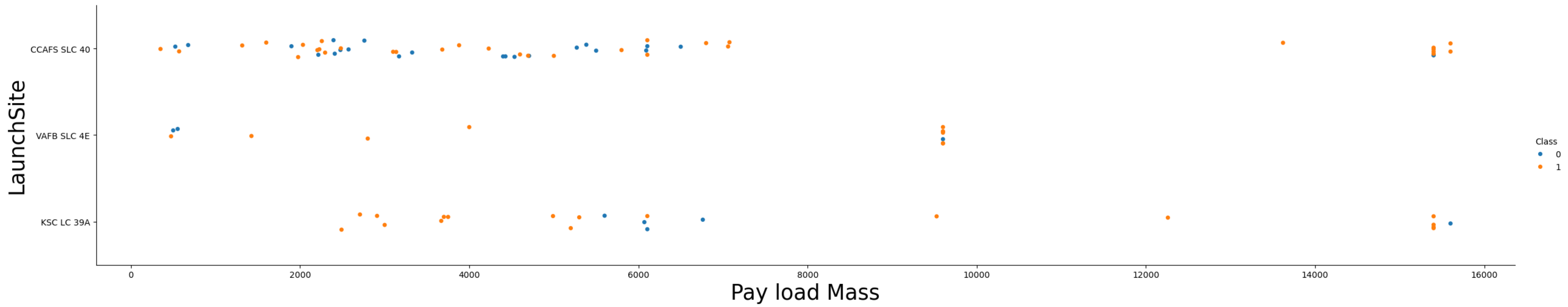
Insights drawn from EDA

Flight Number vs. Launch Site



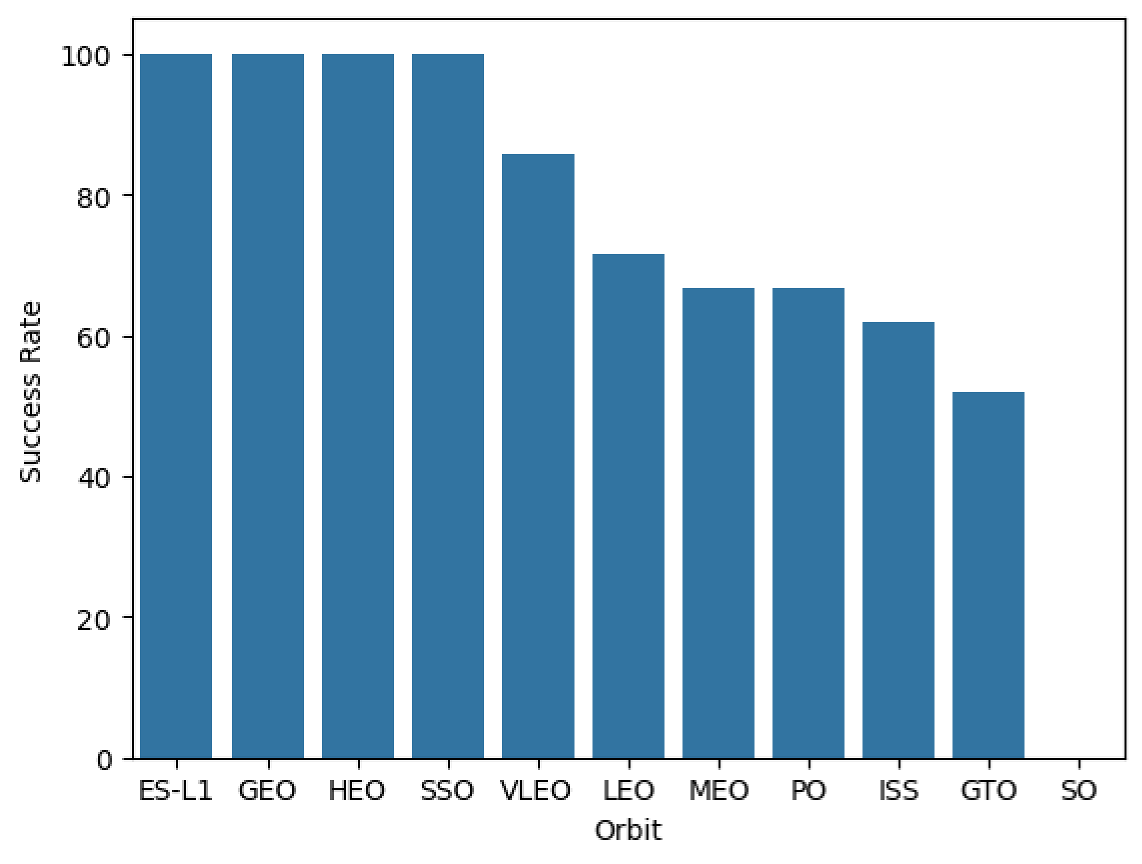
- We can see that with fewer flights, the success rate is very low, but as the number of flights increases, the success rate begins to improve.

Payload vs. Launch Site



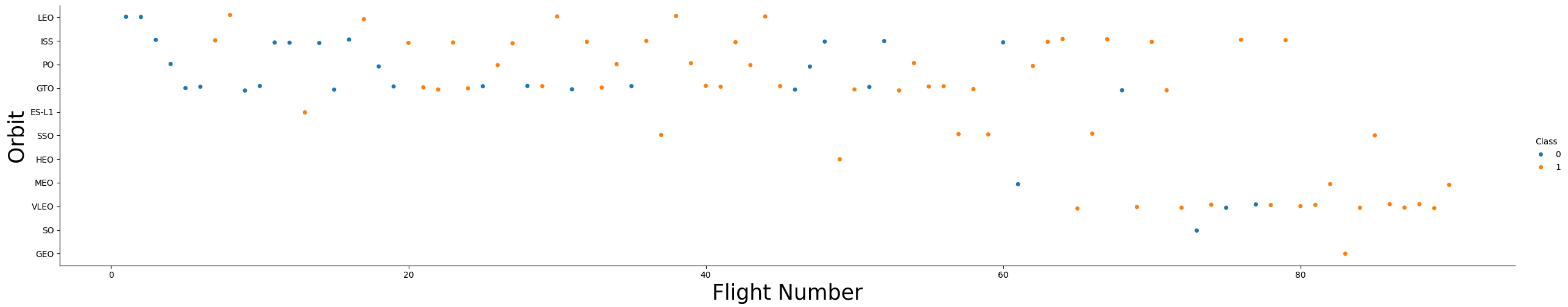
- We can see that heavier payloads present a better success rate. However, KSA LC-39A has a 100% success rate with payloads of less than approximately 5,500kg.

Success Rate vs. Orbit Type



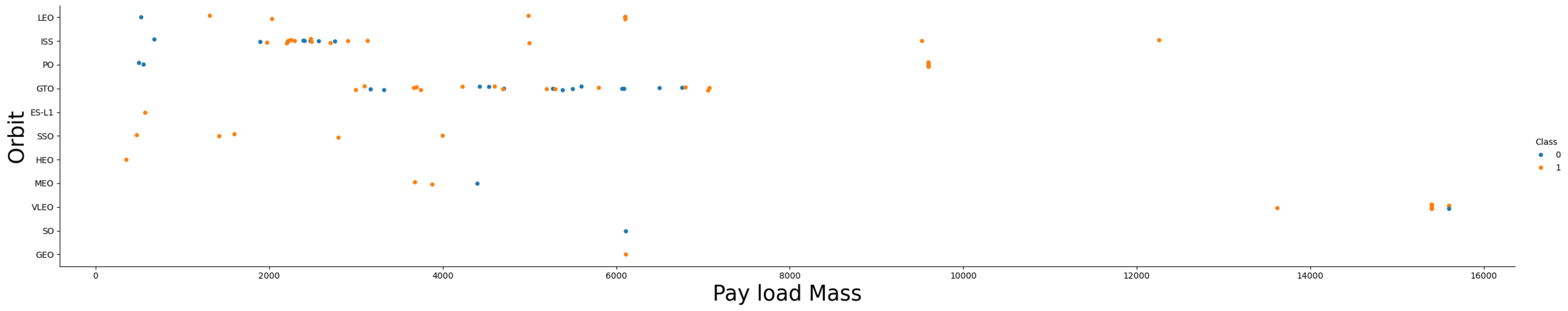
- This bar chart shows that ES-L1, GEO, HEO, and SSO have a 100% success rate.

Flight Number vs. Orbit Type



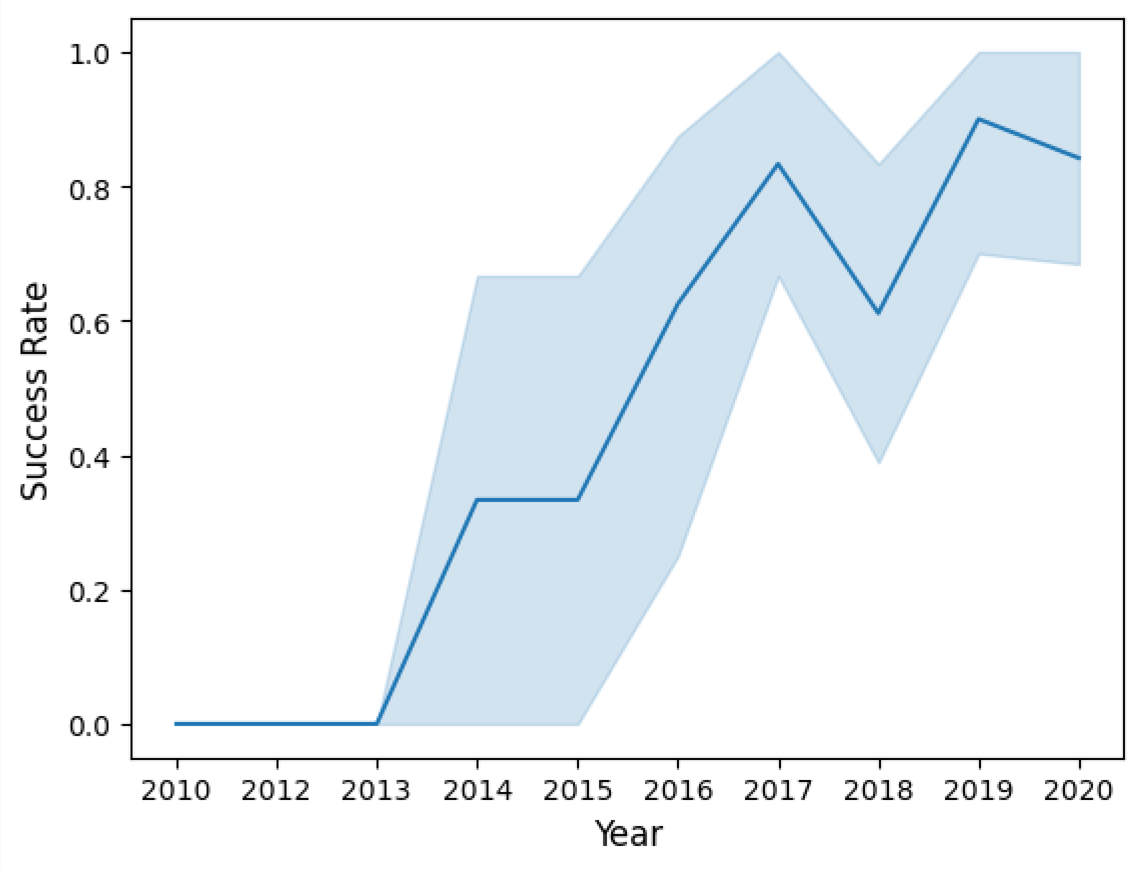
- With this plot was possible determinate the Leo performed better than other orbits after some flights, the number of flights influences in the success rate, is not possible to determinate the relationship in GTO.

Payload vs. Orbit Type



- With this plot, it was possible to determine that LEO performed better than other orbits after a certain number of flights. The number of flights influences the success rate. However, it is not possible to determine a clear relationship for GTO.

Launch Success Yearly Trend



- We can observe an increase in the success rate after 2013, when looking at the data over the years.

All Launch Site Names

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

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

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

- The names of the unique launch sites

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

- 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

- The total payload carried by boosters from NASA

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

- The average payload mass carried by booster version F9 v1.1

First Successful Ground Landing Date

```
%sql select min(DATE) from SPACEXTBL where Landing_Outcome = "Success (ground pad)"
```

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

min(DATE)
2015-12-22

- The dates of the first successful landing outcome on ground pad

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select Booster_Version from SPACEXTBL where 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

- List of the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

```
%sql select count(Mission_Outcome) from SPACEXTBL where Mission_Outcome = 'Success' or Mission_Outcome = 'Failure (in flight)'
```

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

count(Mission_Outcome)
99

- The total number of successful and failure mission outcomes

Boosters Carried Maximum Payload

```
%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

- List of the names of the booster which have carried the maximum payload mass

2015 Launch Records

```
%sql select substr(Date, 6,2) as month, substr(Date, 1,4) as Year , Booster_Version, Launch_site from SPACEXTBL \
    where Landing_Outcome like 'Failure (drone ship)' and Year = '2015'
```

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

month	Year	Booster_Version	Launch_Site
01	2015	F9 v1.1 B1012	CCAFS LC-40
04	2015	F9 v1.1 B1015	CCAFS LC-40

- List of the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

```
%sql select Landing_Outcome, count(*) as outcomes from SPACEXTBL where Date Between '2010-06-04' and '2017-03-20' group by Landing_Outcome order by outcomes desc;
```

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

Landing_Outcome	outcomes
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

- 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 orde

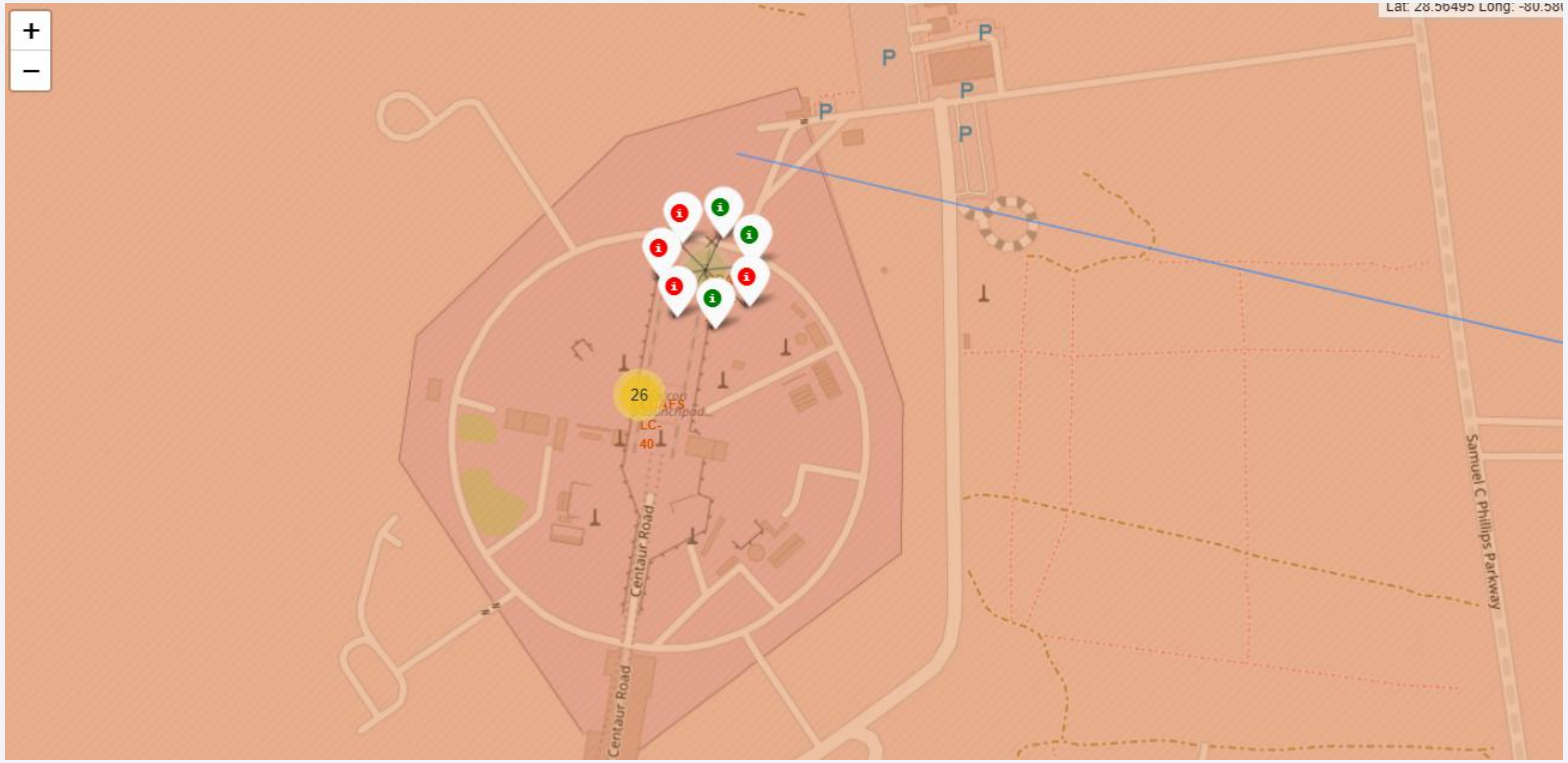
Section 3

Launch Sites Proximities Analysis

All launch Sites

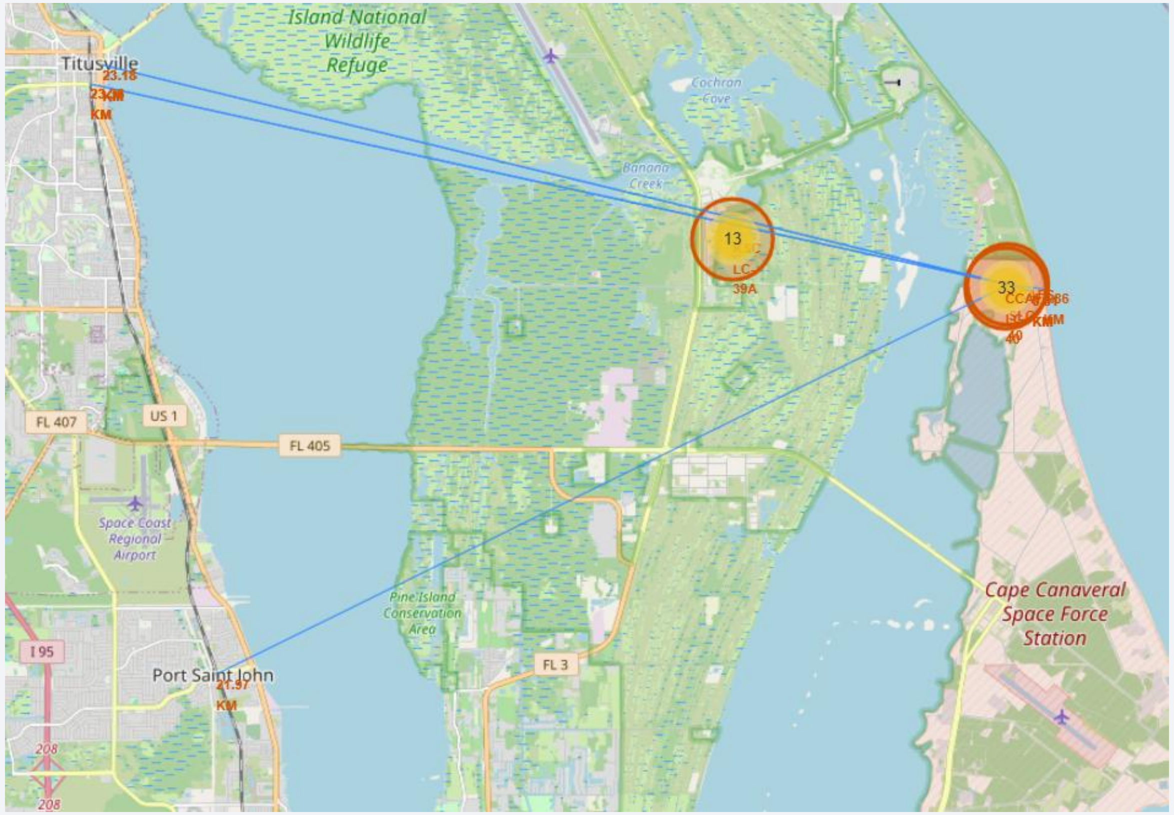
- Here we can see the launch sites are located closer to the coasts

Launch outcomes



- Comparison between successful launches (Green) and failed (Red)

Proximities

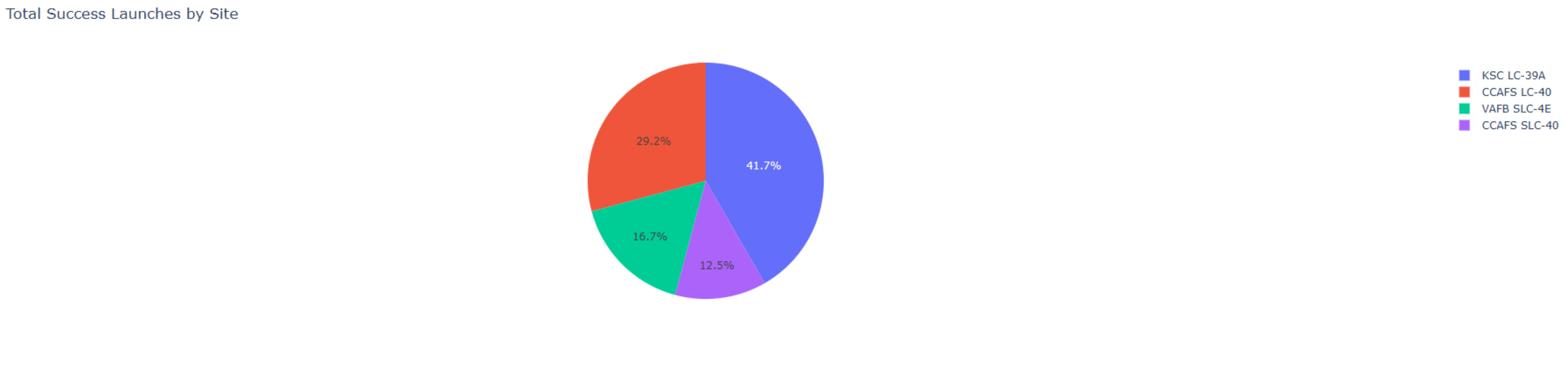


- We can see and the determinate the distance between coast, highway, railway and cities.

Section 4

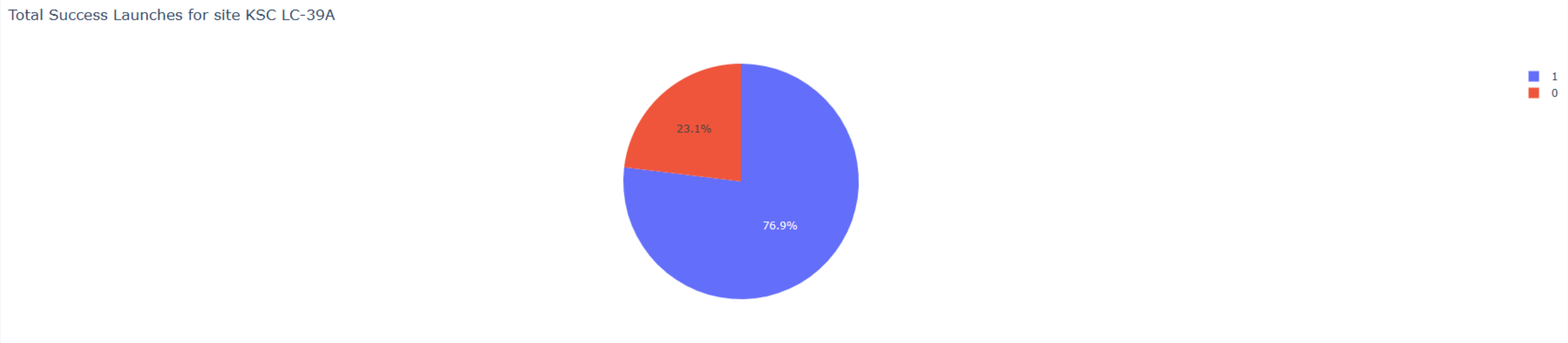
Build a Dashboard with Plotly Dash

All sites Piechart



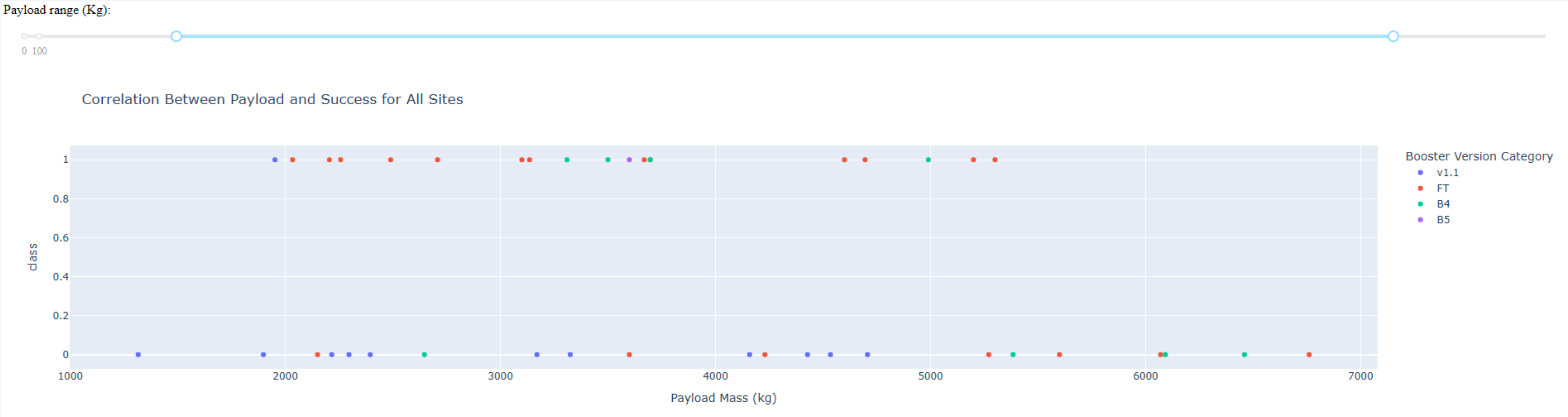
- Among all the launch sites, KSC LC-39A has the most successful rate

Pie chart for site KSC LC-39A



- KSC LC-39A has the best success rate with 76.9%

Scatter Chart for All Sites

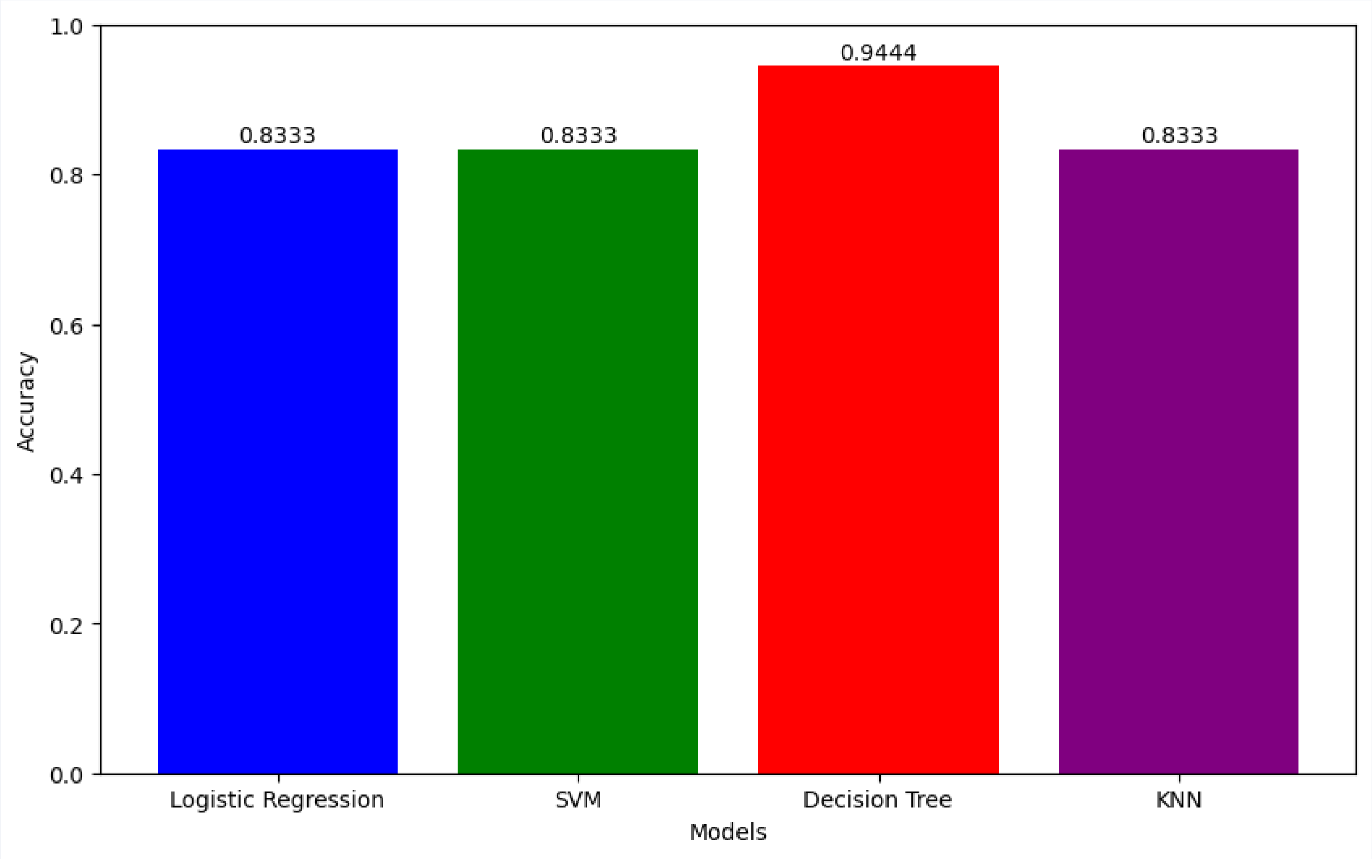


- Payloads between 2000kgs and 5000kgs have the best success rate

Section 5

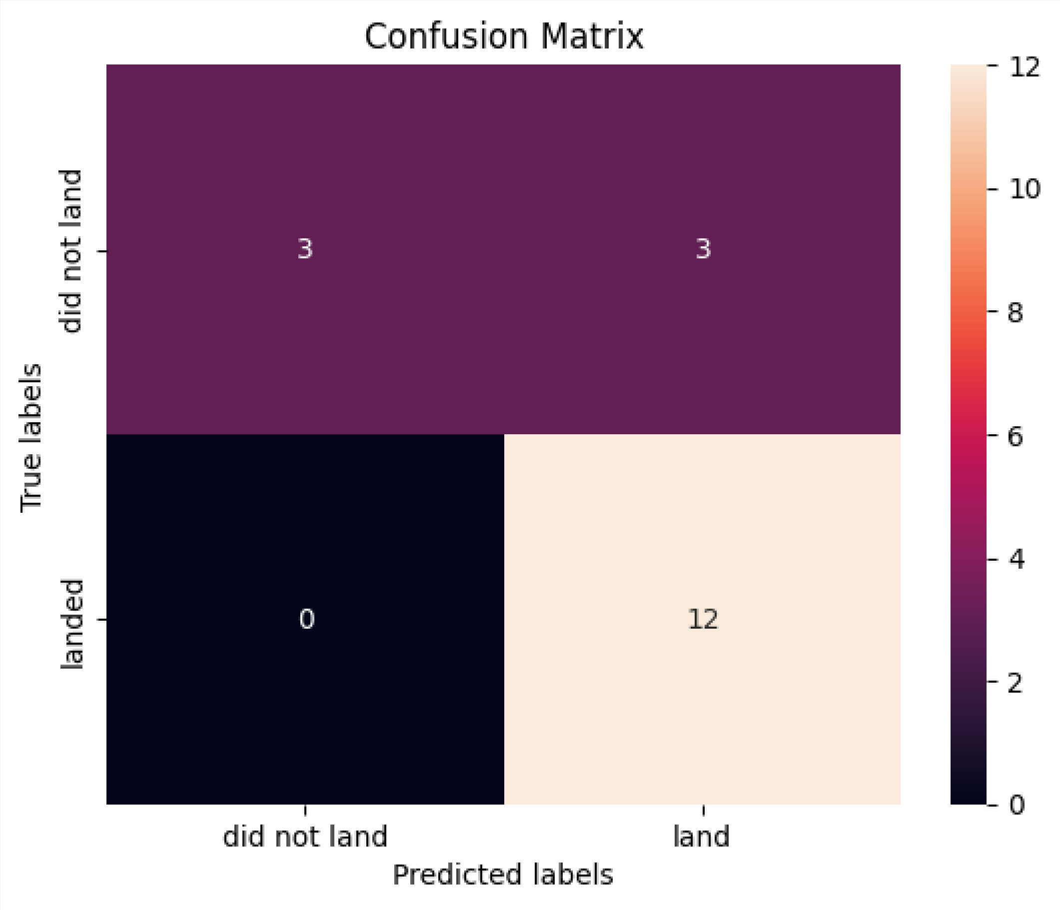
Predictive Analysis (Classification)

Classification Accuracy



In the tests, all algorithms have very similar results, but the decision tree classifier has better accuracy.

Confusion Matrix



Although the decision tree give us the best accuracy, all the confusion matrices are the same

Conclusions

- Point 2
- Orbits ES-L1, SSO, HEO and GEO, has 100% successful rate
- The KSC LC-39A Lauch Site has the most successful landing
- The best classifier algorithm, for this dataset, was The Tree classifier algorithm

Appendix

- All graphs and code used is this presentation can be found in the GitHub: GitHub - YagoPassos - Applied Data Science Capstone

Thank you!