



IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Simon B.
13.10.2023



Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

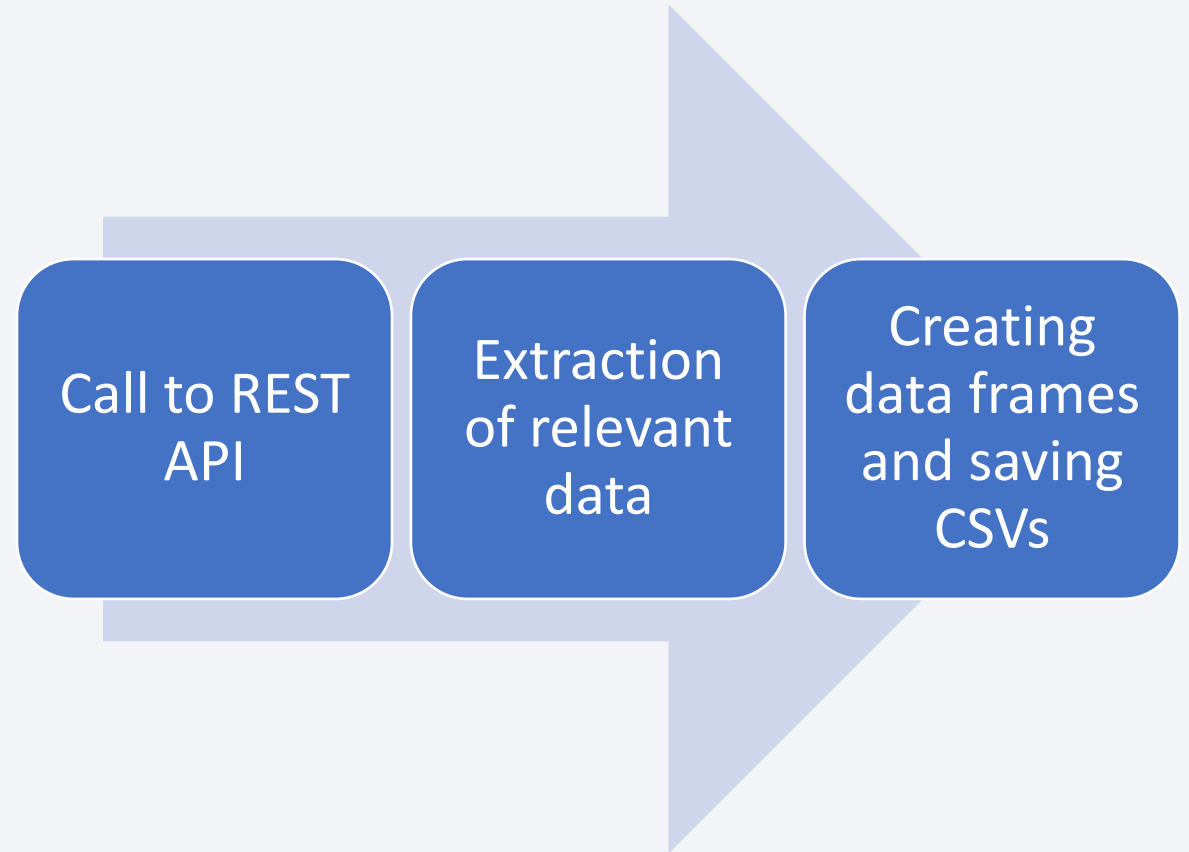
- Data collection methodology:
 - Describe how data was collected
- Perform data wrangling
 - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
 - How to build, tune, evaluate classification models

Data Collection

- Base Data was collected from SpaceX accessible online sources
- This was further enriched by web-scraping relevant Wikipedia sources

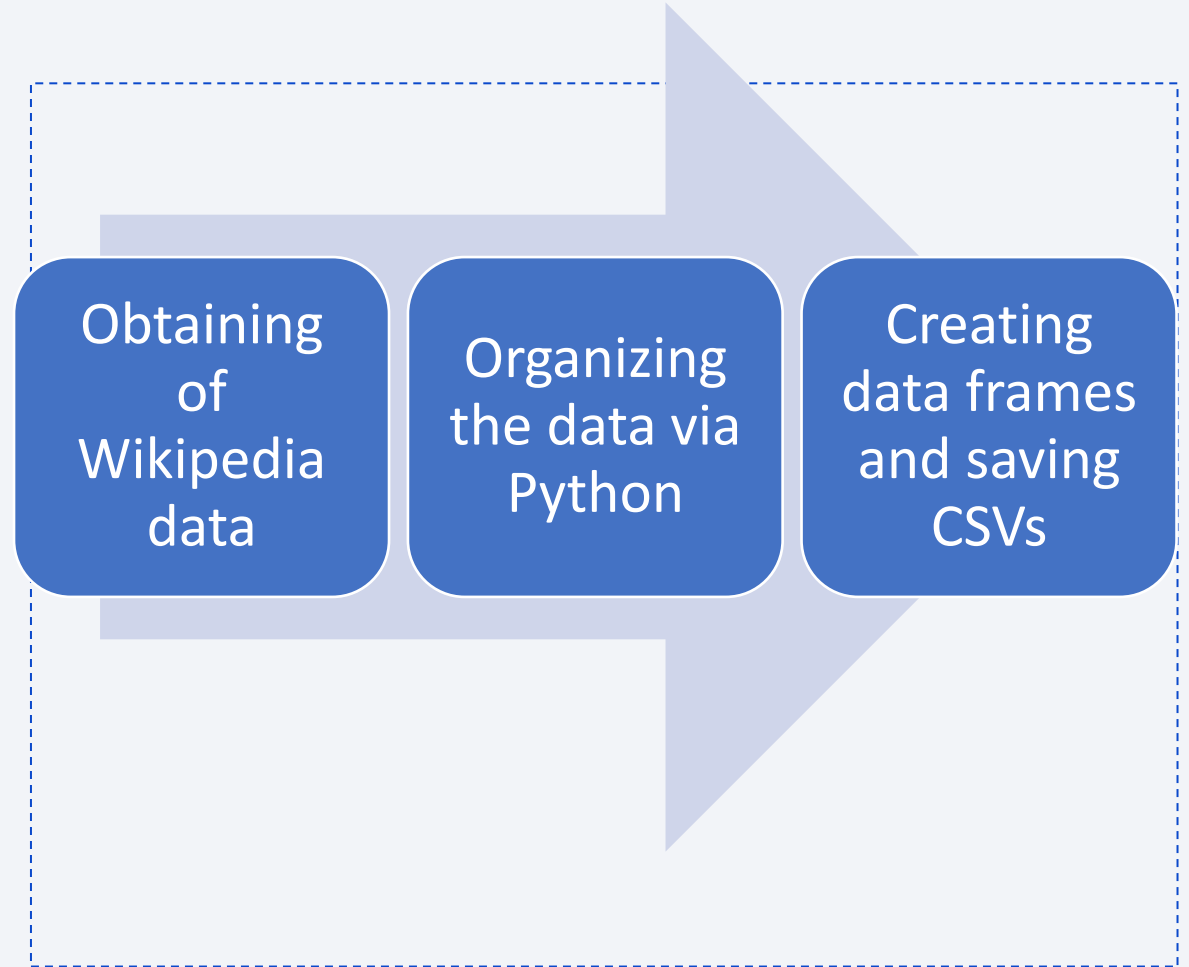
Data Collection – SpaceX API

- The REST API was contacted via http calls, specifically its “core” component. Data was then loaded into data frames using the Python library “Pandas”
- The Notebook can be found [here](#)



Data Collection - Scraping

- Data was scraped from Wikipedia via the Python library “beautifulsoup”, relevant data was loaded into data frames and stored as CSV
- The notebook can be found [here](#)



Data Wrangling

- Data was cleaned and data types were assigned
- Mission and landing outcomes were calculated based on relevant parameters such as orbit type and launch site
- Notebook can be found [here](#)

EDA with Data Visualization

- Outcomes were plotted against potential predictive variables such as flight number and date (to track progression), payload, and orbit type
- Notebook with full set of visuals can be found [here](#)

EDA with SQL

- SQL was used to create summarized overviews such as success counts, payload averages, and others. More detailed information on certain booster types was also obtained that way.
- Notebook can be found [here](#)

Build an Interactive Map with Folium

- The Python library “Folium” was used to create an interactive map showing launch sites, missions and their outcomes, as well as distances to certain points of interest
- This enabled a better understanding how launch site locations are chosen and how well they perform as starting points for missions.
- Notebook can be found [here](#)

Build a Dashboard with Plotly Dash

- A dashboard with insights into success rates based on launch site and payload was created
- This allows a quick interactive overview
- Code file can be found [here](#)

Predictive Analysis (Classification)

- Mission outcome prediction models were created and tested via grid search
- Predictive variables were chosen and prepared
 - Normalizing numerical variables
 - One-hot encoding categorical variables
- Notebook can be found [here](#)

Results

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

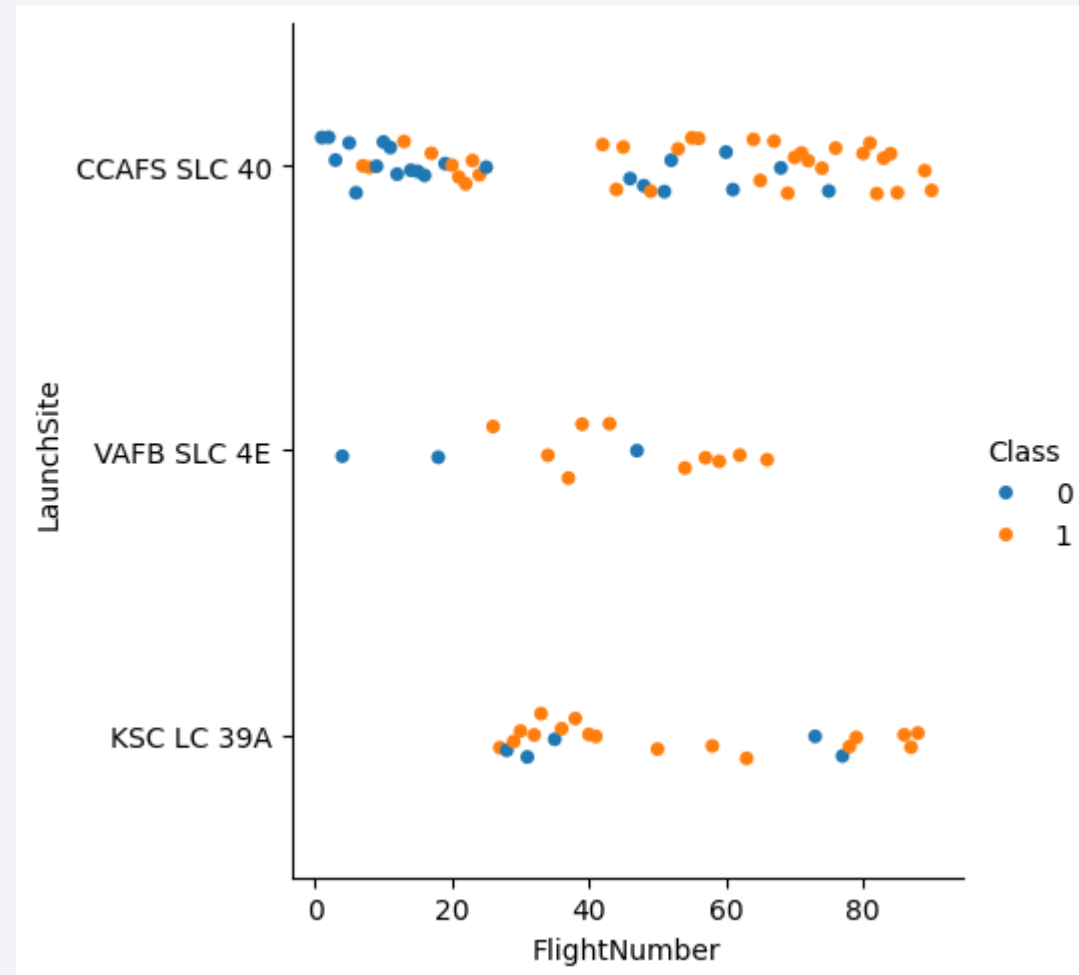
The background of the slide is an abstract composition. It features a dark blue base color. Overlaid on this are numerous diagonal streaks in shades of red and cyan. A faint, light blue grid pattern is also visible, particularly in the lower-left quadrant. The overall effect is dynamic and technological.

Section 2

Insights drawn from EDA

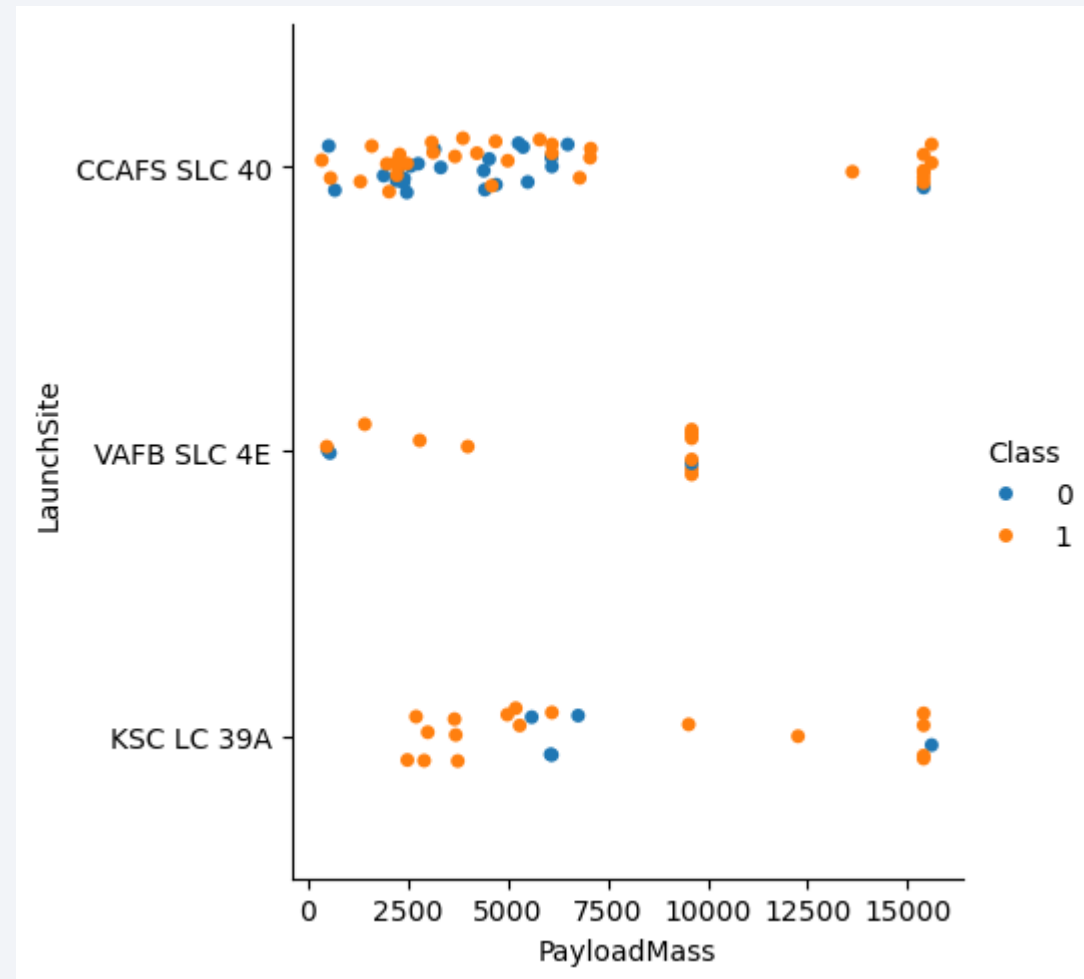
Flight Number vs. Launch Site

- Flights over time can be seen broken down by the three launch sites
- Red dots indicate successful missions, blue dots failures



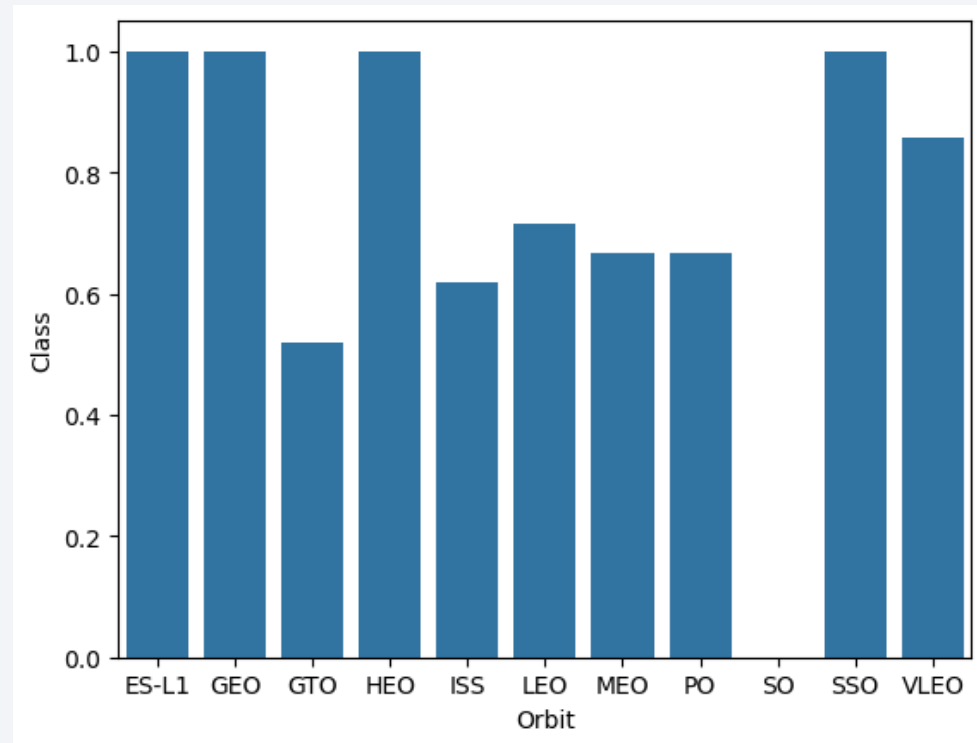
Payload vs. Launch Site

- Payload massed can be seen broken down by the three launch sites
- Red dots indicate successful missions, blue dots failures



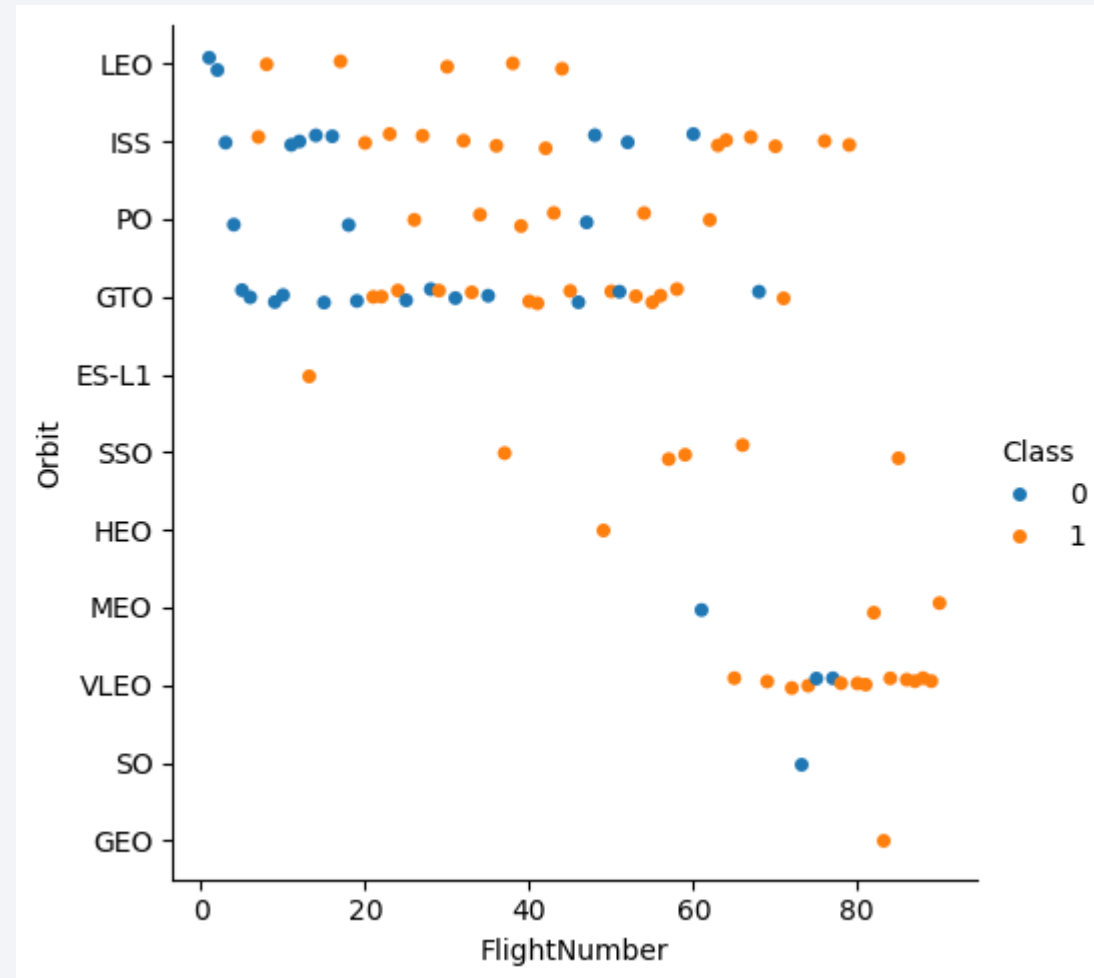
Success Rate vs. Orbit Type

- This bar chart shows the mission success rates for different orbit types



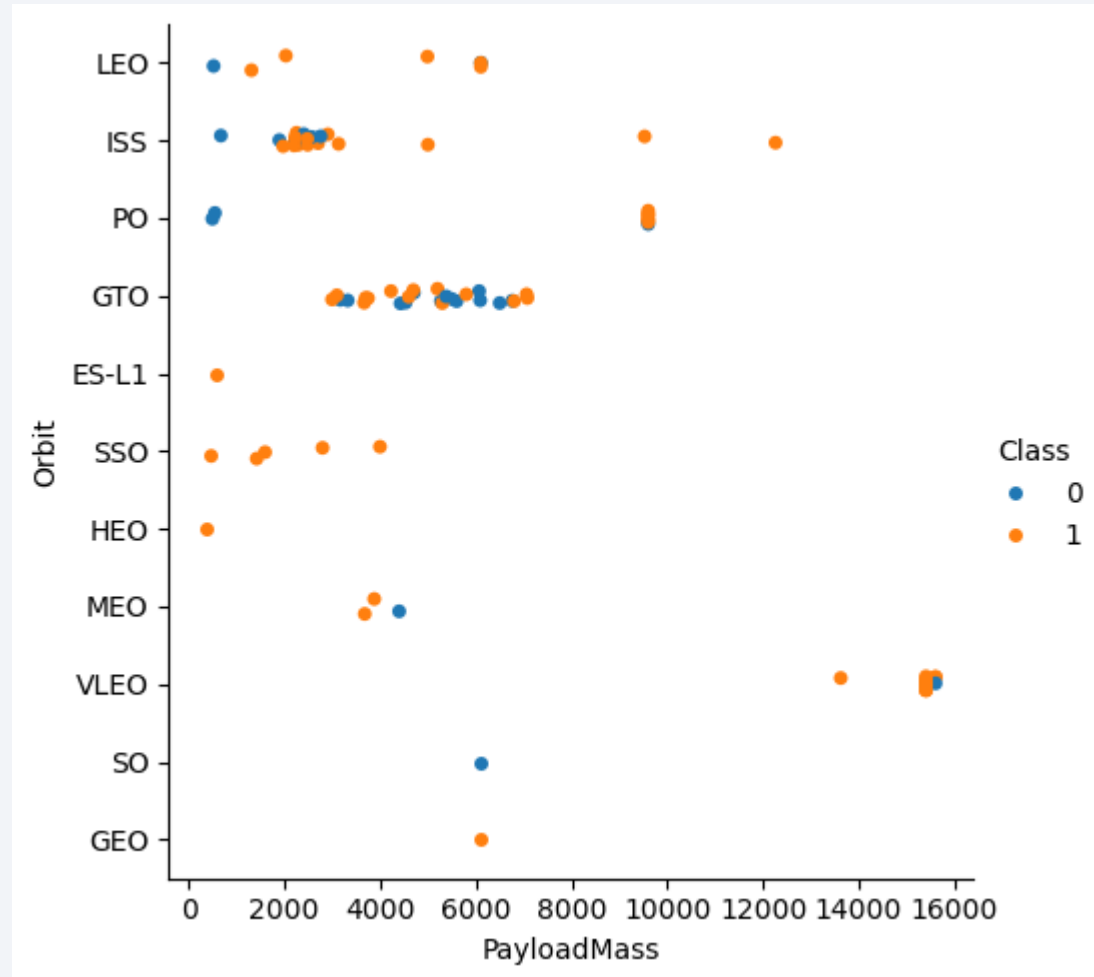
Flight Number vs. Orbit Type

- This plot shows the orbit types by flight number, giving an overview over when missions started going to those orbits
- Red dots indicate successful missions, blue dots failures



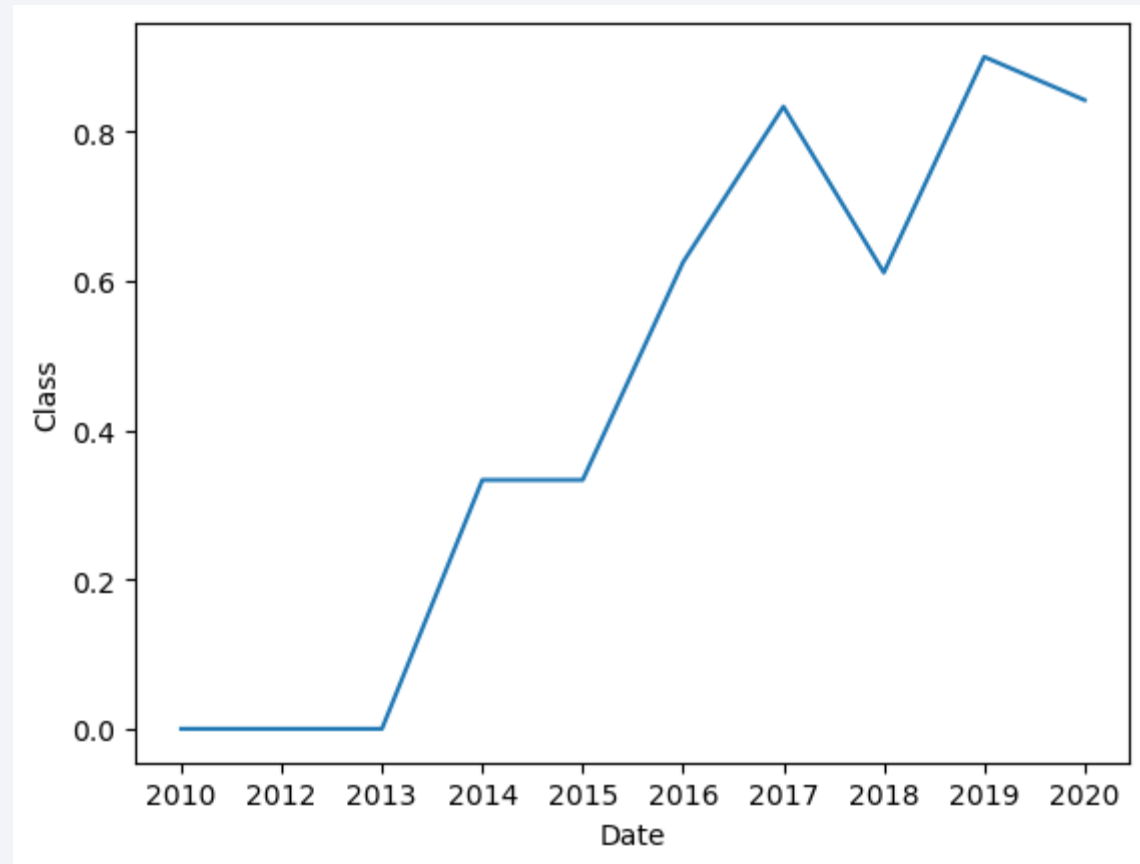
Payload vs. Orbit Type

- The plot shows payload masses broken down by orbit types
- Red dots indicate successful missions, blue dots failures



Launch Success Yearly Trend

- The graph shows mission success rate over time



All Launch Site Names

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

Launch Site Names Begin with 'CCA'

```
In [14]: %sql select * from SPACEXTABLE where Launch_Site like "CCA%" limit 5
```

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

Out[14]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

```
In [22]: %sql select SUM(PAYLOAD_MASS_KG_) from SPACEXTABLE where Customer = "NASA (CRS)"
```

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

```
Done.
```

```
Out[22]: SUM(PAYLOAD_MASS_KG_)  
         45596
```


Average Payload Mass by F9 v1.1

Task 4

Display average payload mass carried by booster version F9 v1.1

```
In [23]: %sql select AVG(PAYLOAD_MASS_KG_) from SPACEXTABLE where Booster_Version like "F9 v1.1 %"
```

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

```
Out[23]: AVG(PAYLOAD_MASS_KG_)  
          2337.8
```

First Successful Ground Landing Date

```
In [28]: %sql select MIN(Date) from SPACEXTABLE where Landing_Outcome = "Success (ground pad)"
```

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

```
Done.
```

```
Out[28]: MIN(Date)
```

```
2015-12-22
```

Successful Drone Ship Landing with Payload between 4000 and 6000

```
In [30]: %sql select distinct(Booster_Version) from SPACEXTABLE where PAYLOAD_MASS_KG_ > 4000 and PAYLOAD_MASS_KG_ < 6000 and Land:  
* sqlite:///my_data1.db  
Done.
```

```
Out[30]: Booster_Version
```

F9 FT B1022

F9 FT B1026

F9 FT B1021.2

F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

```
In [41]: %sql select count(Mission_Outcome) as Successes from SPACEXTABLE where Mission_Outcome = "Success" union select count(Mission_Outcome) as Failures from SPACEXTABLE where Mission_Outcome = "Failure"
```

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

```
Done.
```

```
Out[41]: Successes
```

```
1
```

```
98
```

Boosters Carried Maximum Payload

```
In [42]: %sql select Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) from SPACEXTABLE)
```

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

```
Out[42]: 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

In [74]:

```
%%sql
select
    substr(Date,0,5) as year_,
    substr(Date, 6,2) as month,
    Booster_Version,
    Launch_Site,
    count(Mission_Outcome) as launches,
    sum(case when Landing_Outcome = "Failure (drone ship)" then 1 else 0 end) as landing_fails_droneship

from SPACEXTABLE
where substr(Date,0,5)='2015'
group by year_, month, Launch_Site, Booster_Version
```

* sqlite:///my_data1.db

Done.

Out[74]:

year_	month	Booster_Version	Launch_Site	launches	landing_fails_droneship
-------	-------	-----------------	-------------	----------	-------------------------

2015	02	F9 v1.1 B1014	CCAFS LC-40	1	0
2015	04	F9 v1.1 B1015	CCAFS LC-40	1	1
2015	04	F9 v1.1 B1016	CCAFS LC-40	1	0
2015	06	F9 v1.1 B1018	CCAFS LC-40	1	0
2015	10	F9 v1.1 B1012	CCAFS LC-40	1	1
2015	11	F9 v1.1 B1013	CCAFS LC-40	1	0
2015	12	F9 FT B1019	CCAFS LC-40	1	0

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

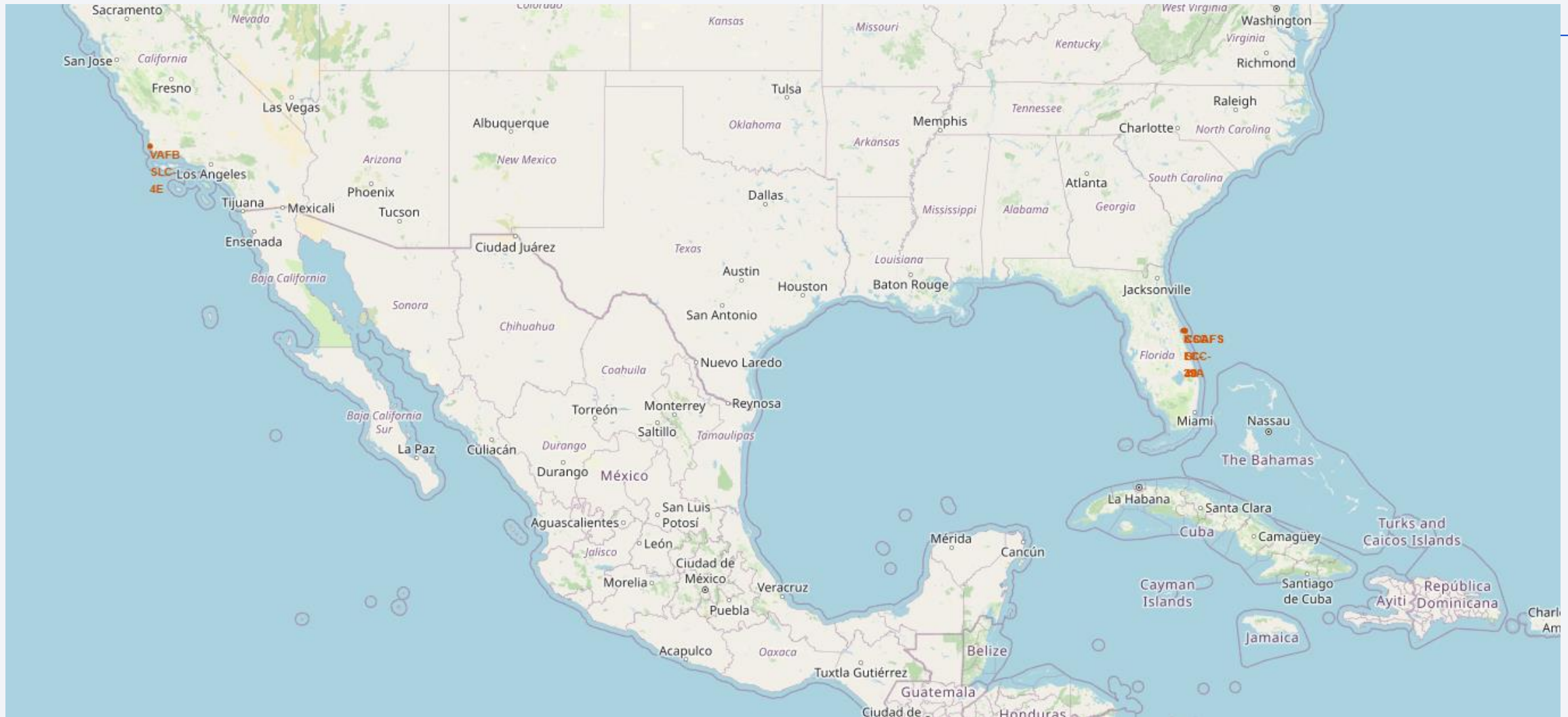
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order
- Present your query result with a short explanation here

A satellite view of Earth from space, showing the curvature of the planet and city lights at night. The background is a deep blue gradient.

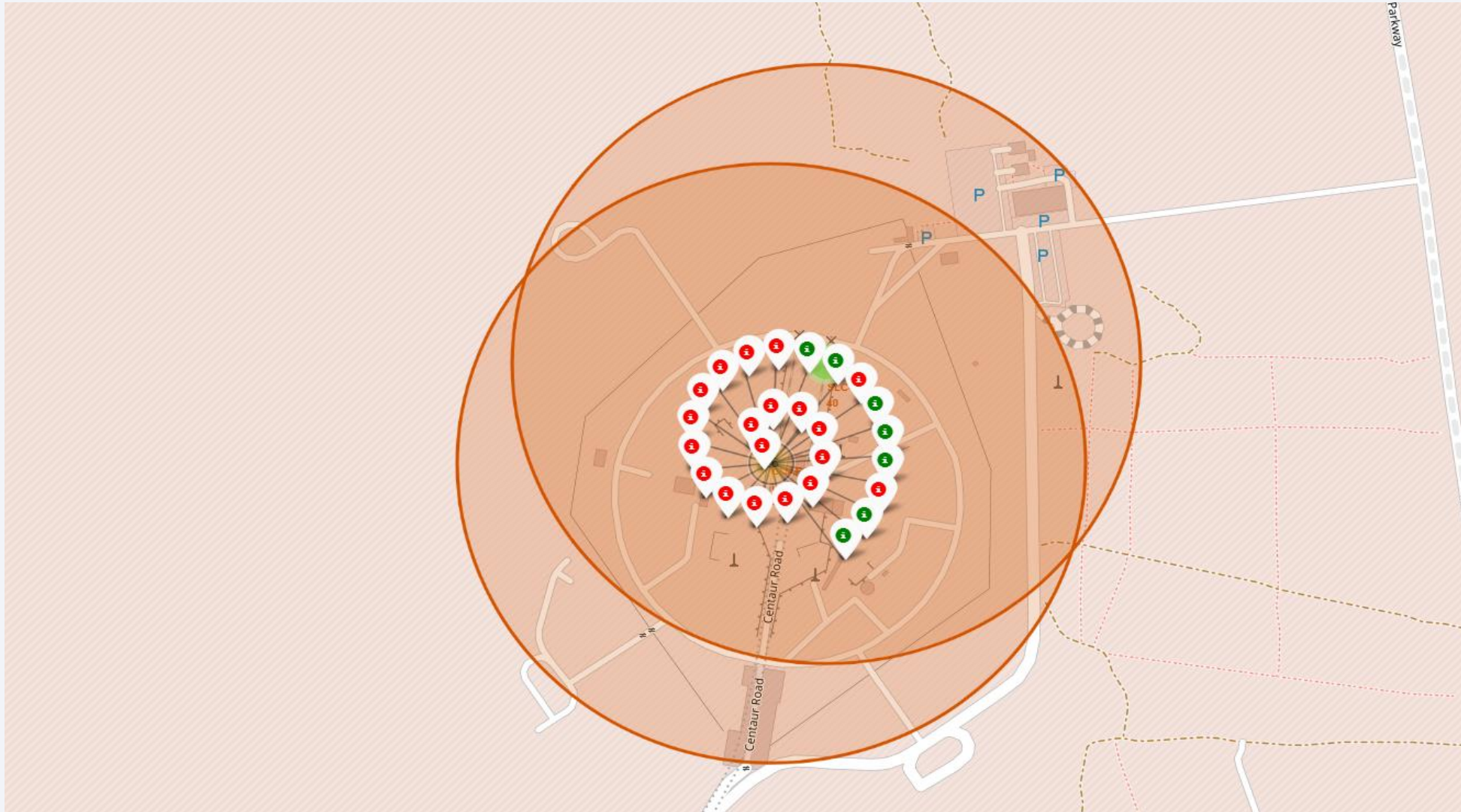
Section 3

Launch Sites Proximities Analysis

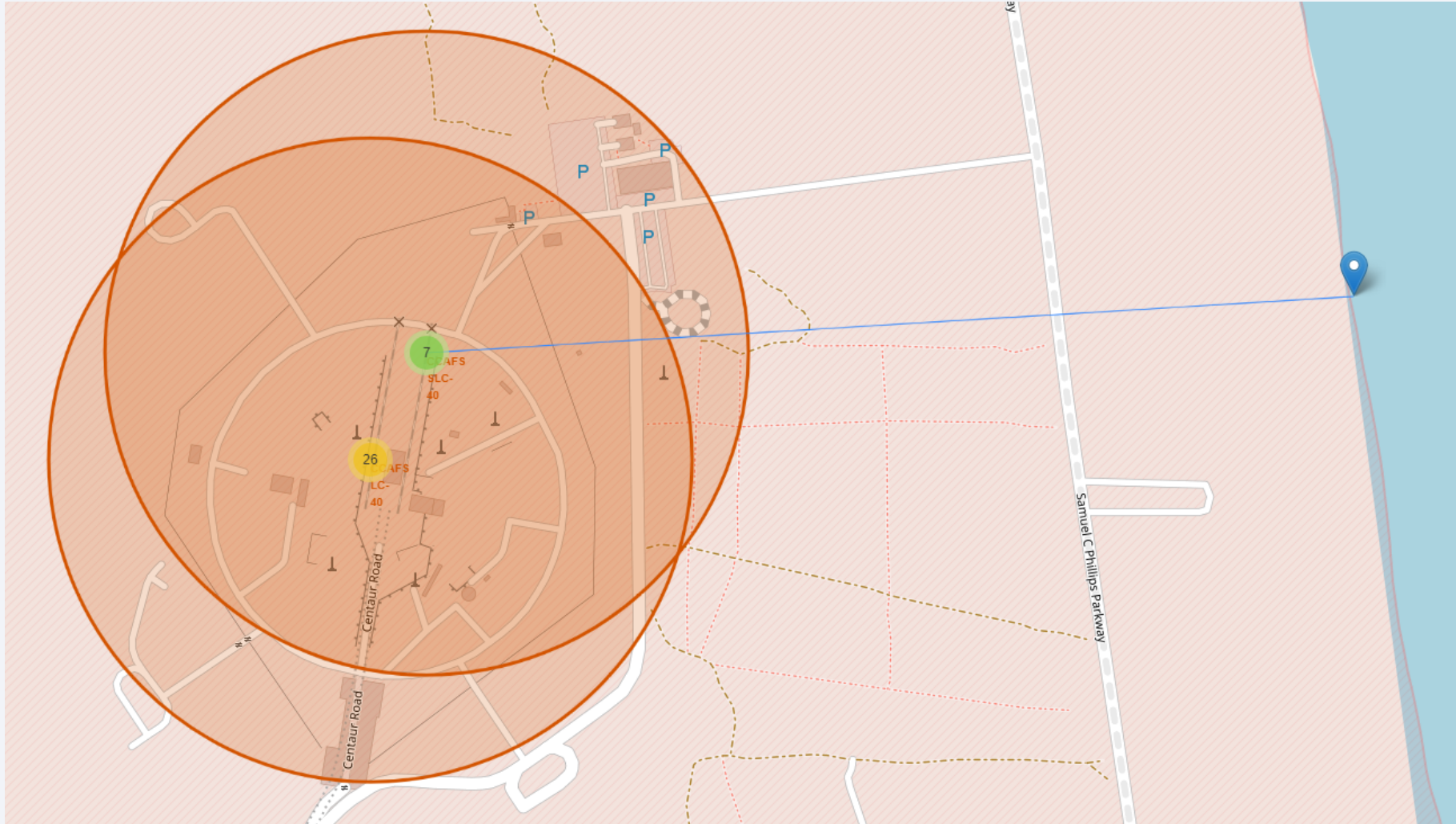
<Folium Map Screenshot 1>



<Folium Map Screenshot 2>



<Folium Map Screenshot 3>



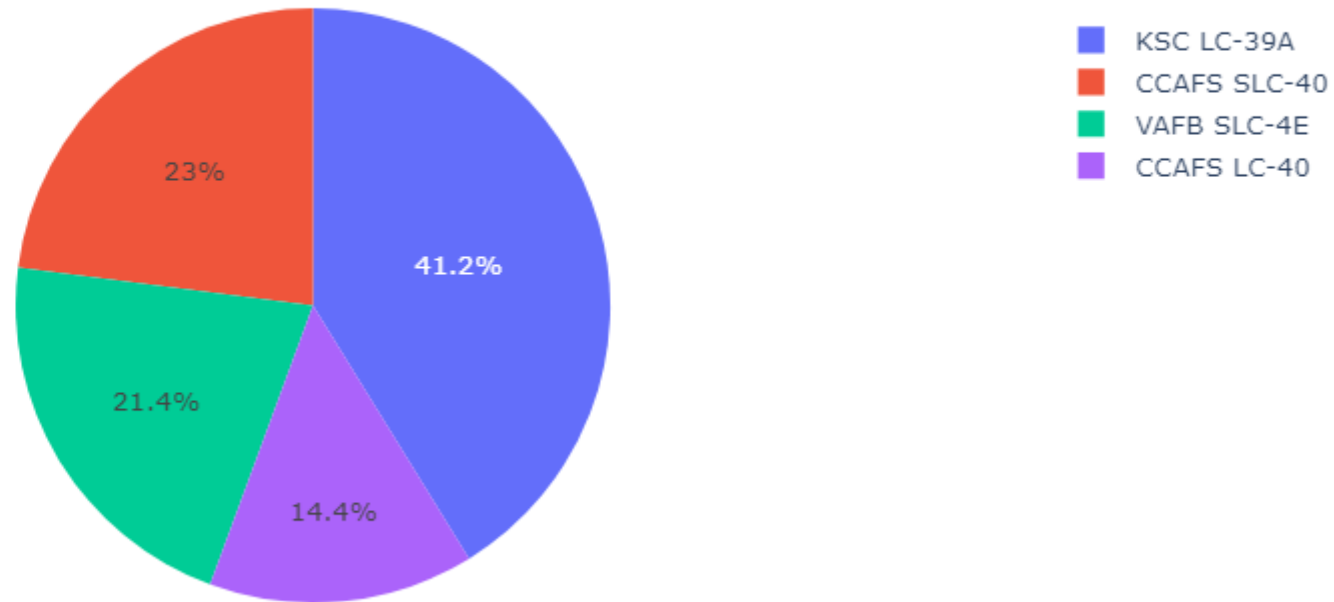


Section 4

Build a Dashboard with Plotly Dash

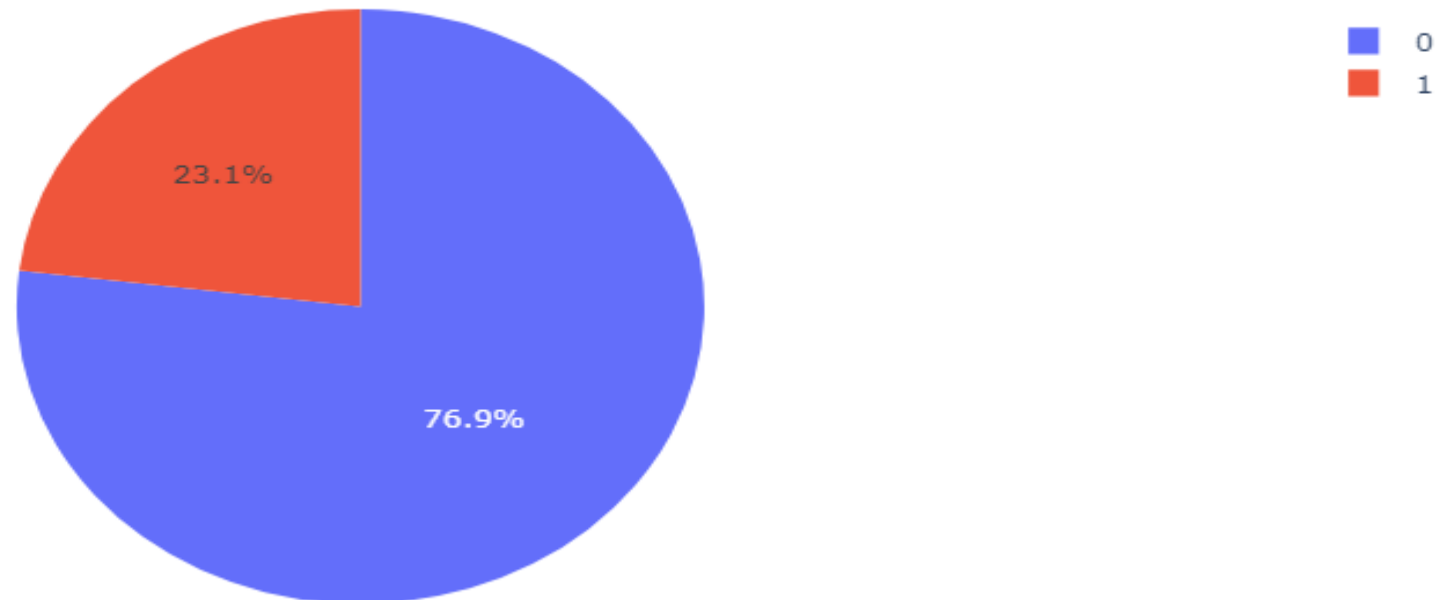
<Dashboard Screenshot 1>

Total Success Launches by Site



<Dashboard Screenshot 2>

Total Success Launches for Site KSC LC-39A

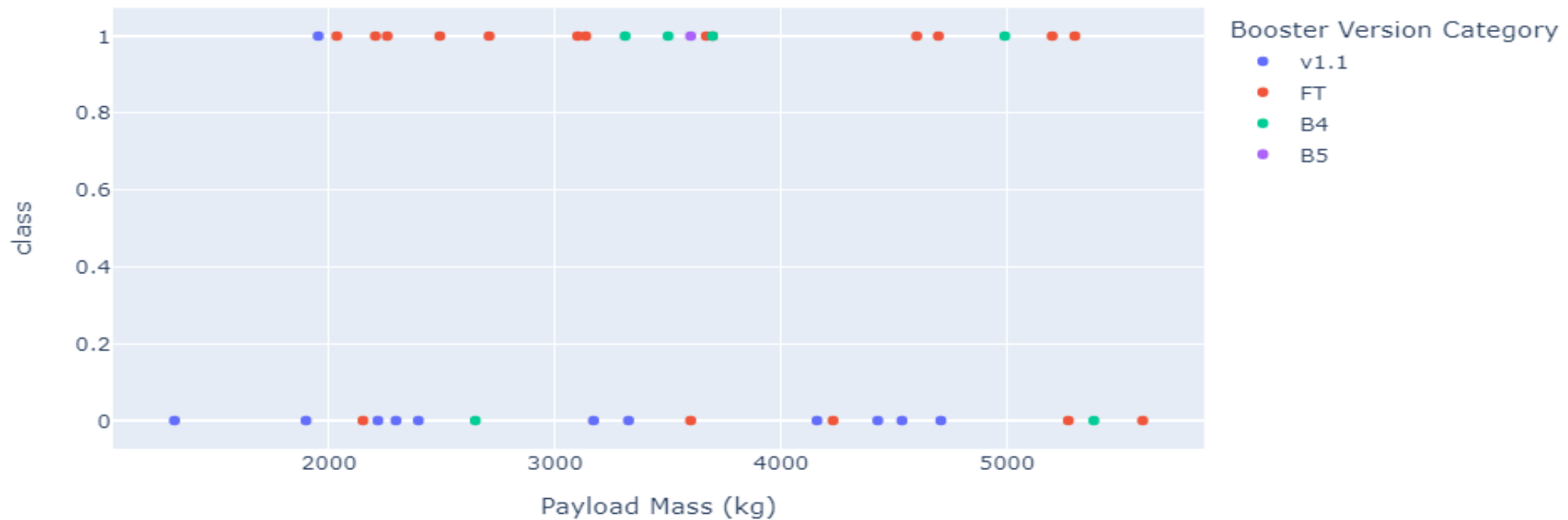


<Dashboard Screenshot 3>

Payload range (Kg):



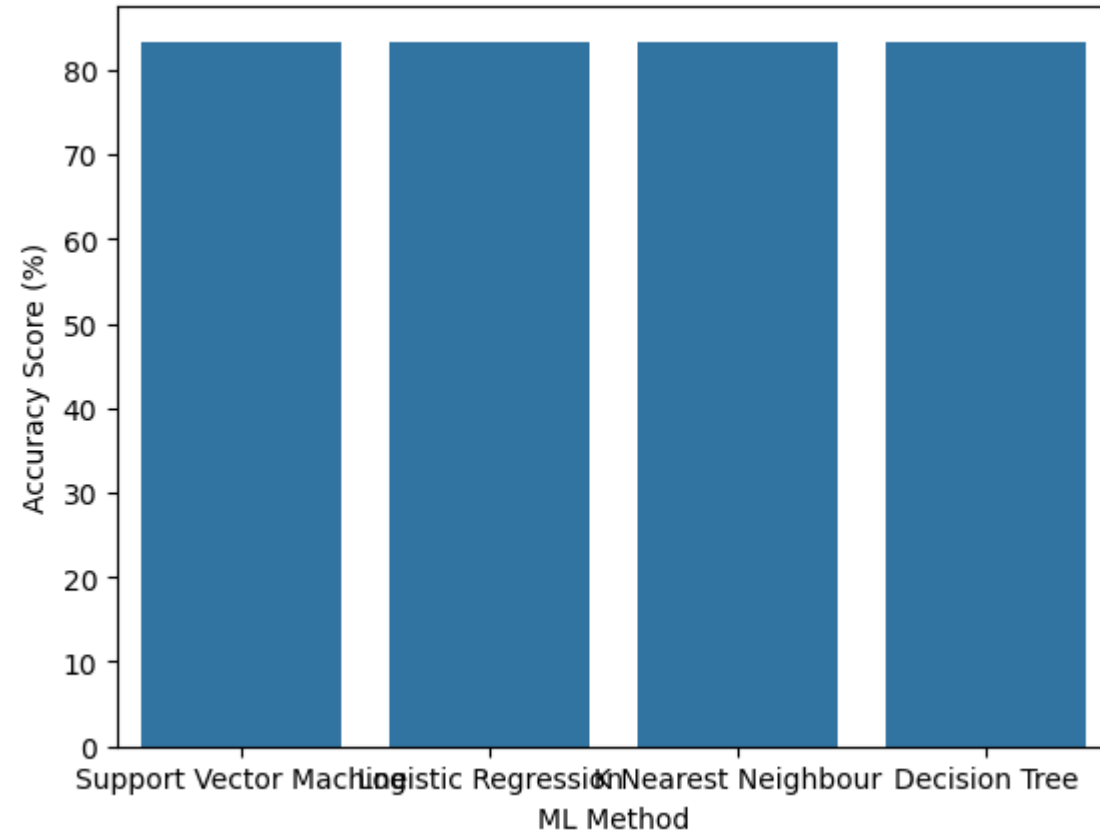
Correlation Between Payload and Success for All Sites



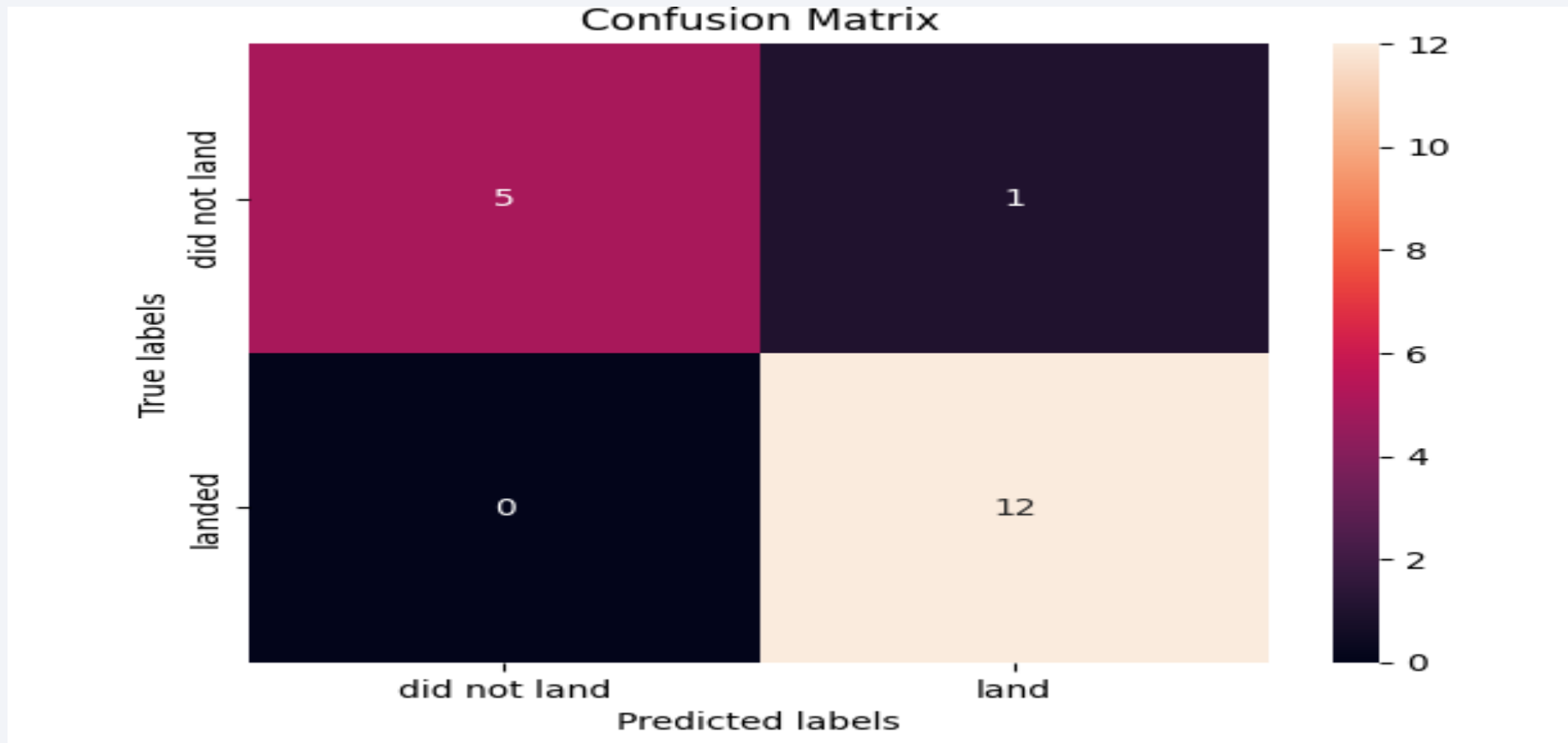
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- Data showed interesting insights into SpaceX progress over time and success factors
- A model was built with decent success to create launch outcomes

Appendix

- [Link to full repository](#)

Thank you!

