IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Waleed Abdulhadi
10/4/2023

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Commercial space travel is becoming more and more prominent and especially in the past decade, there is been an astounding advancement in space travel technology which made space travel more accessible and affordable.
- While this trend will only continue to rise in the next years and decades, it cannot be done without the proper knowledge and insight necessary to actualize it’s vision.
- One example of space travel successes is SpaceX with the falcon 9 rocket launches.
- Our project aims to understand the variables that affect the success or failure of the rockets’ stage one landing and then train a machine learning model to predict if SpaceX will reuse that rocket.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
    - The data was collected using the SpaceX REST API, the API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome, The SpaceX REST API endpoints, or URL, starts with api.spacexdata.com/v4/ and has different end points, We will be working with the endpoint api.spacexdata.com/v4/launches/past.
- Perform data wrangling:

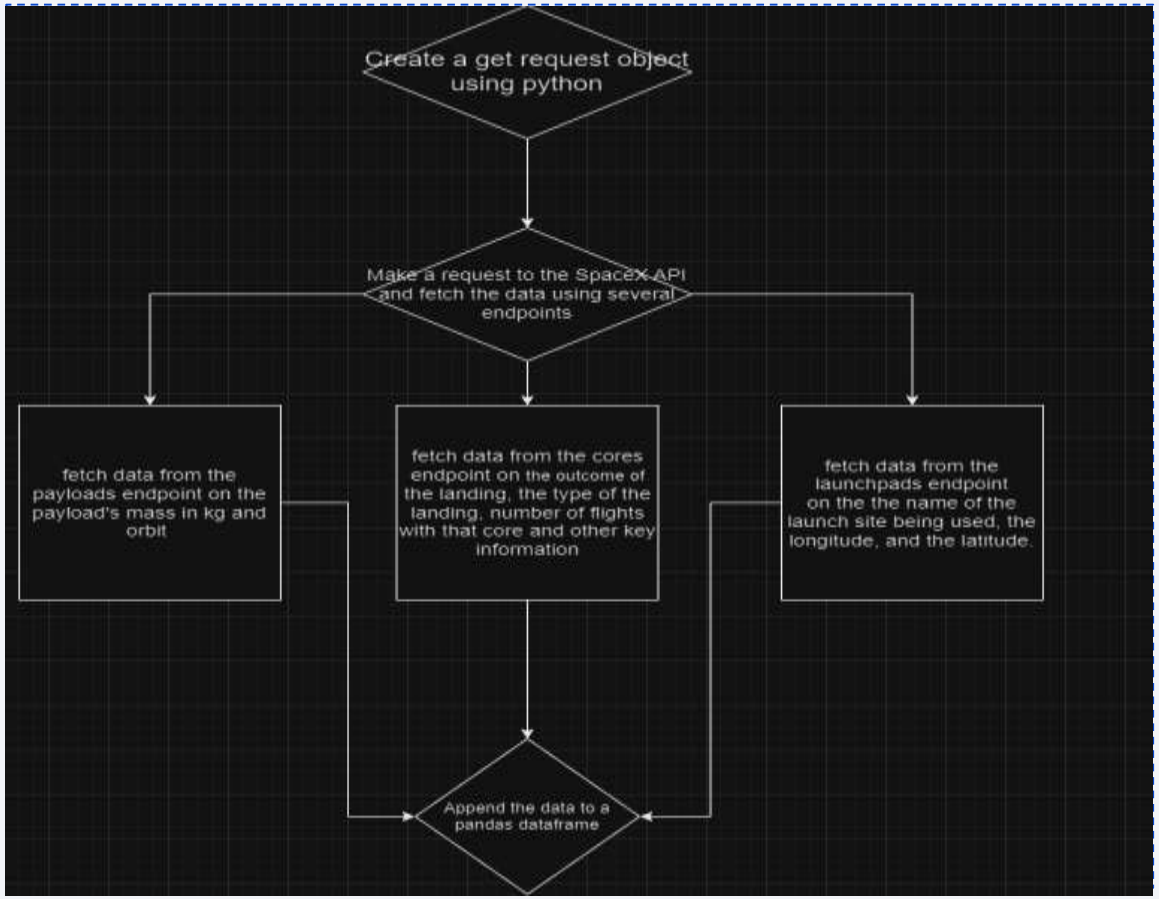
    in the dataset there were several cases where he booster did or did not land successfully. These outcomes were converted to training labels with 0 meaning the landing was unsuccessful and 1 meaning the landing was successful, and then using those labels to calculate the success rate.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

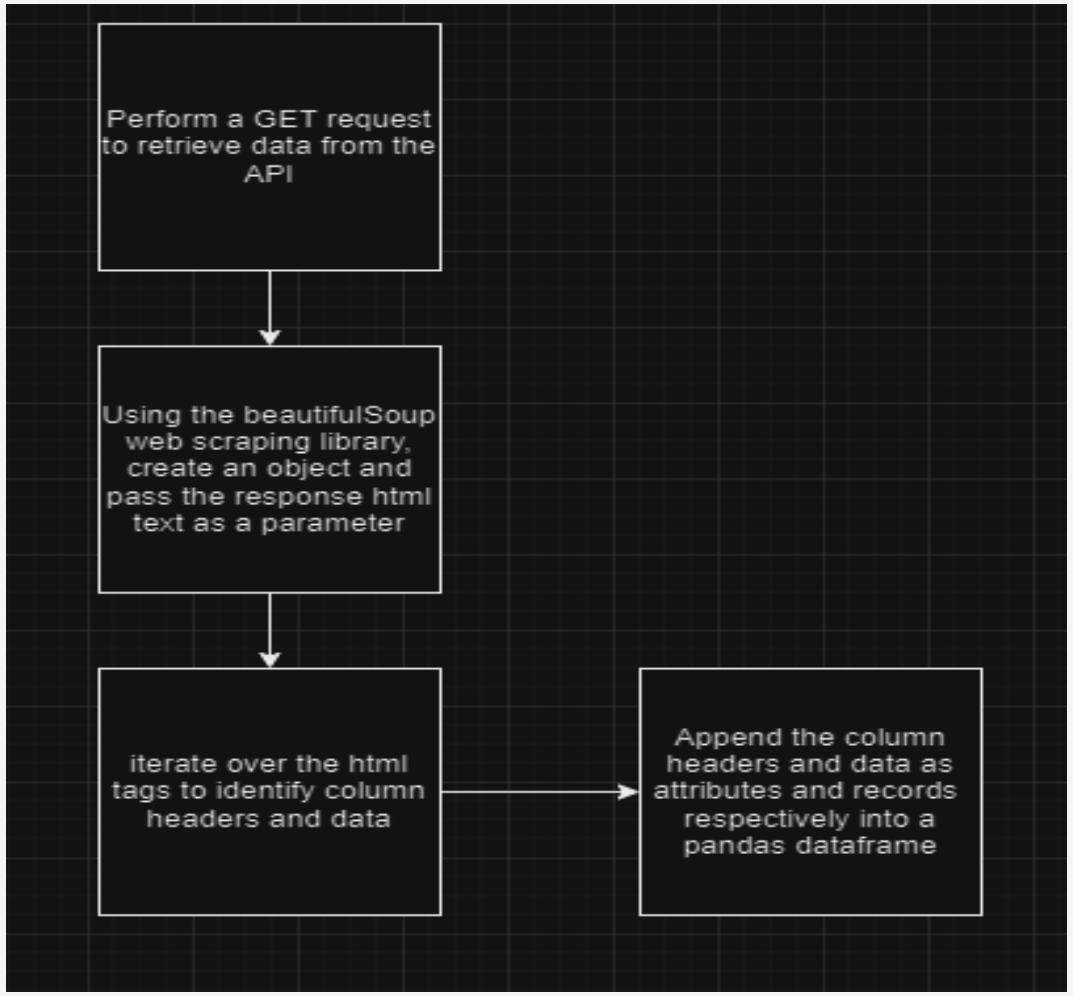
Data Collection – SpaceX API

- We used the python requests module to make a get request to the API and retrieving the data of past launches in JSON format and converting it to a pandas data frame
- [github link]



Data Collection - Scraping

- Using beautifulSoup, a python web scraping library, the html response was iterated over to find column headers and column data and extract them into a pandas dataframe.
- Github link



Data Wrangling

- In the dataset there were several cases where he booster did or did not land successfully. These outcomes were converted to training labels with 0 meaning the landing was unsuccessful and 1 meaning the landing was successful, and then using those labels to calculate the success rate.
- Then we dealt with the missing values of each column with continuous values by replacing them with the mean of the column.

- Github link

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

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

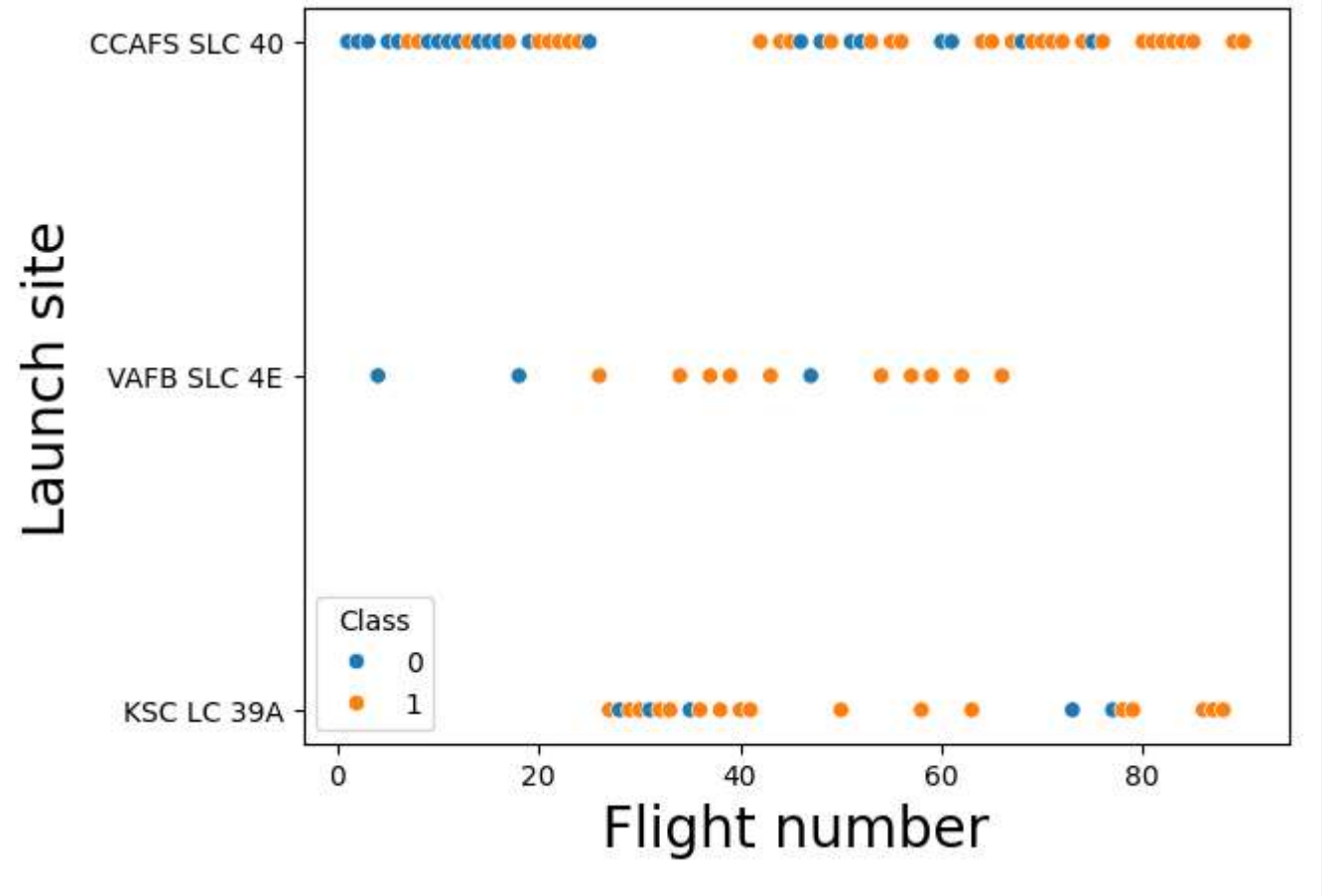
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

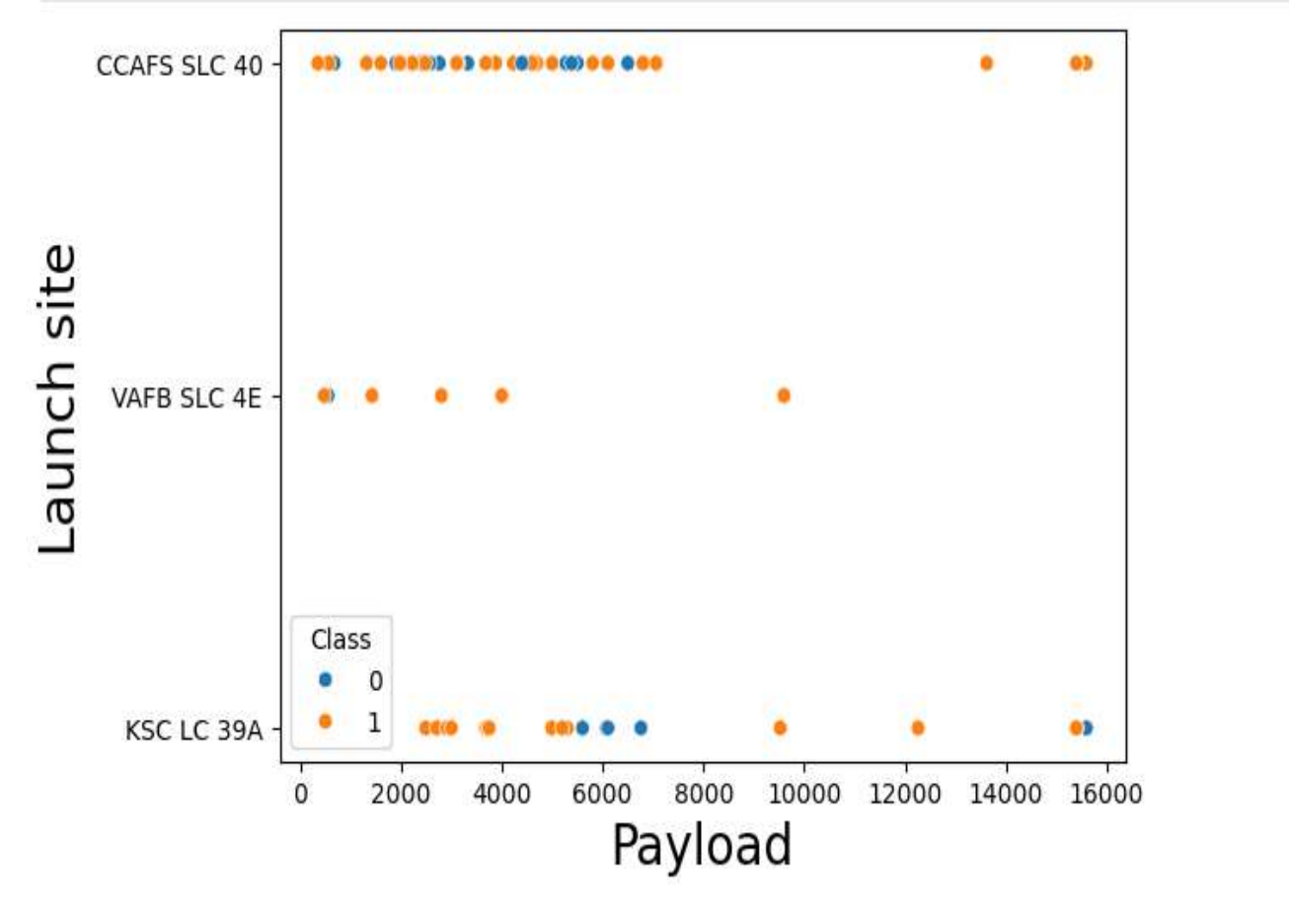
Flight Number vs. Launch Site

- Flight numbers between below 20 and above 40 mostly launched from launch site CCAFS SLC 40
- Flight numbers between 20 and 40 ,and around 80 mostly launched from site KSC LC 39A.
- Site VAFB SLC 4E has scattered values, ranging from 0 to around 70



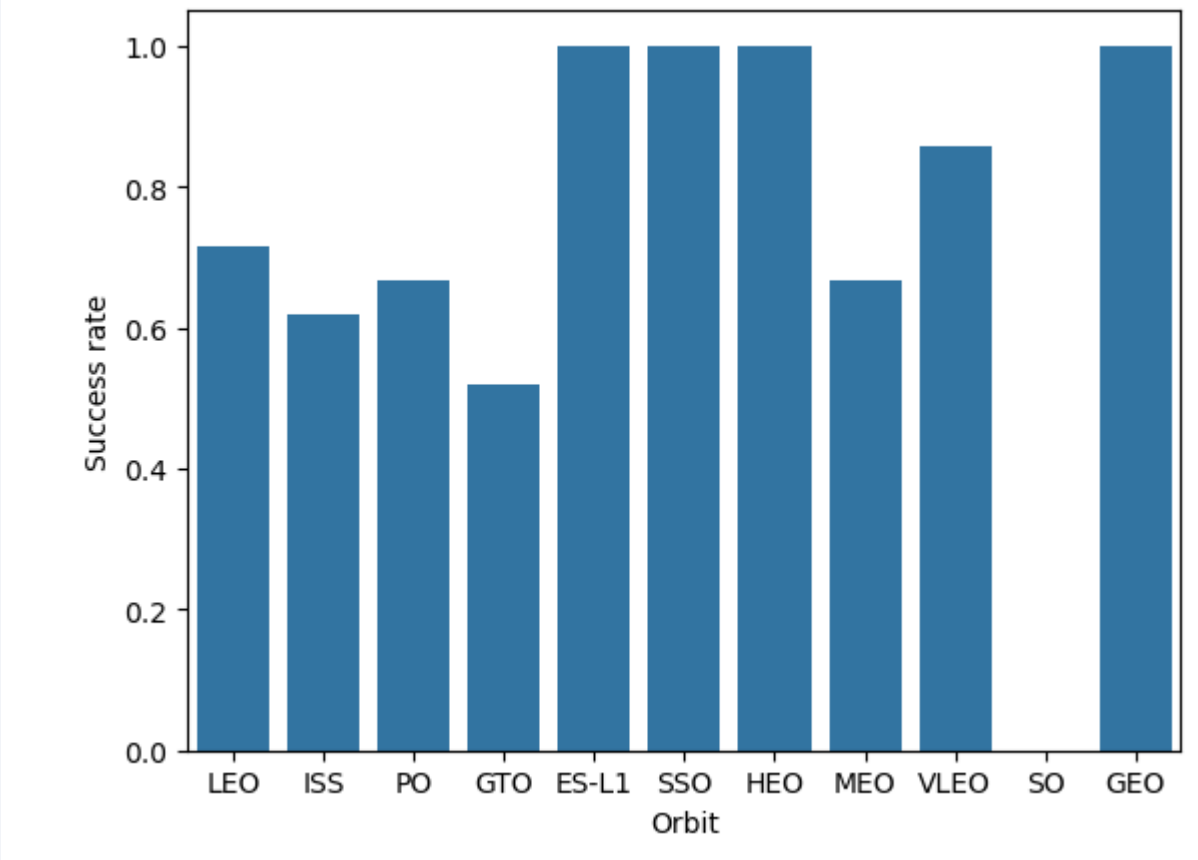
Payload vs. Launch Site

- Site CCASFS SLC 40 launched rockets which mostly had a payload mass ranging from just below 2000 to 8000 kg
- Site KSC LC 39A launched rockets which mostly had a payload mass ranging 2000 to 4000 kg and from 5000 to about 7000 kg
- Site VAFB SLC 4E didn’t launch a lot of rockets, the few that it did mostly ranged in mass from around 500 to 5000 kg.



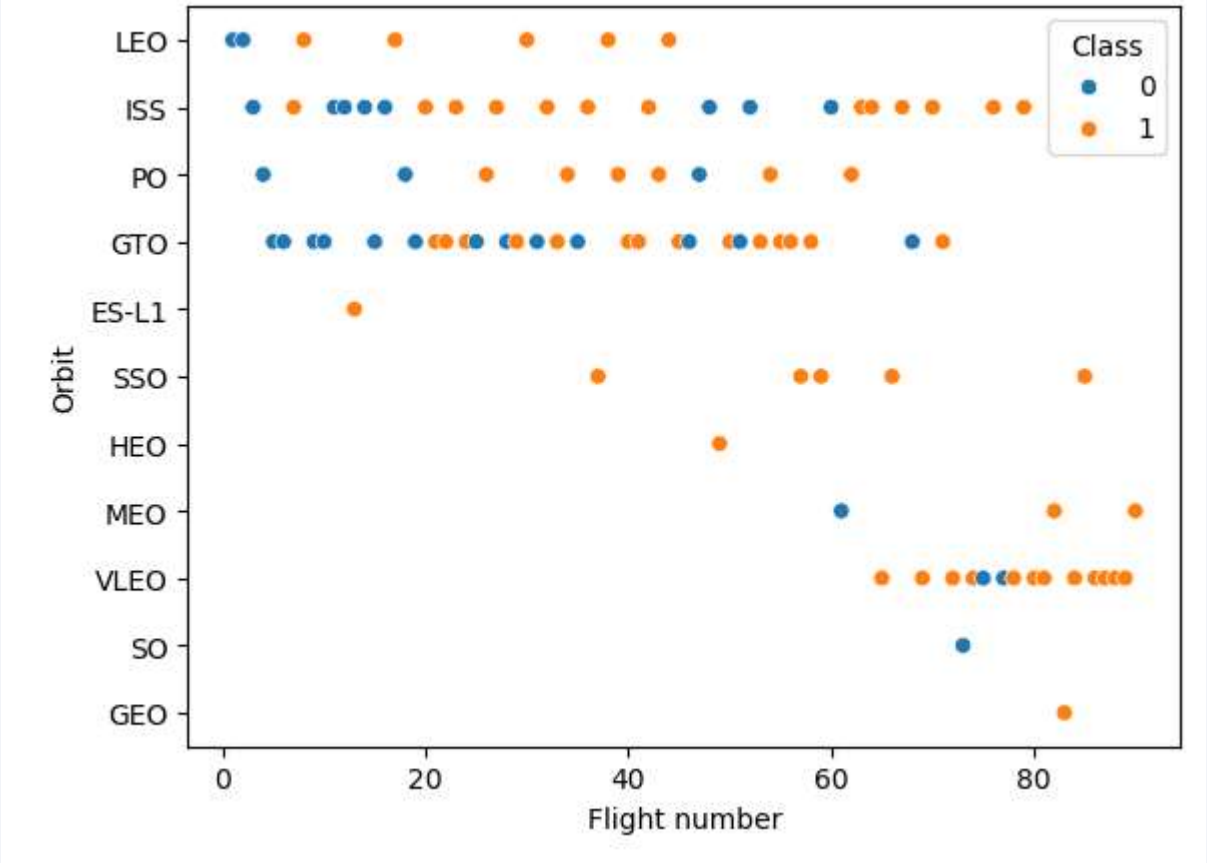
Success Rate vs. Orbit Type

- Target orbits with highest mission success rates are GEO,HEO,SSO and ES-L1
- Target orbit with lowest mission success rate is GTO



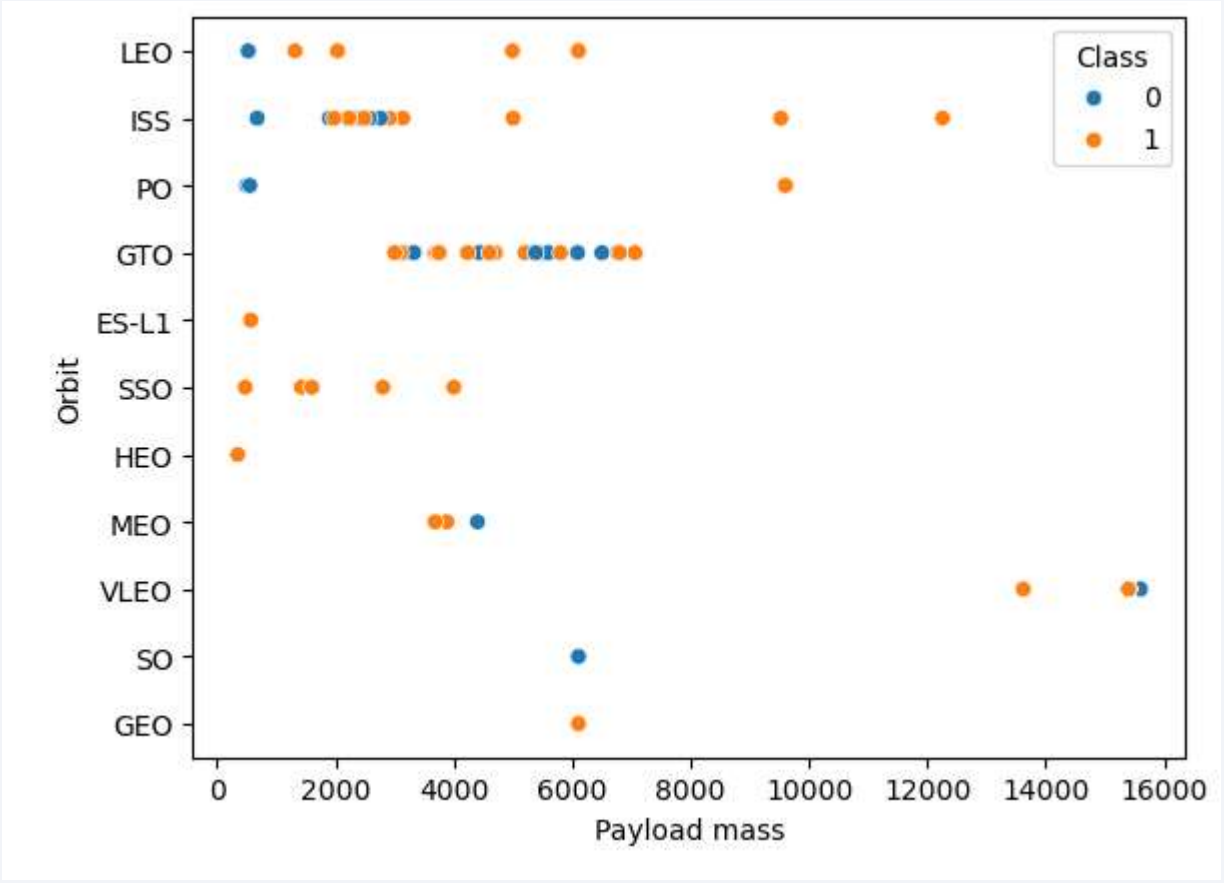
Flight Number vs. Orbit Type

- Most flight numbers ranging from 0 to 70 were in the highest 4 orbits
- While many of flight numbers between 60 and above 80 were in orbit VLEO.



Payload vs. Orbit Type

Flights with payload masses below 2000 were in the highest 2 orbits, while payload masses between 3000 and 8000 were mostly in orbit GTO.



Launch Success Yearly Trend

- Success rate was at it’s lowest before 2013, then it started rapidly rising, reaching a peak between 2019 and 2020.

All Launch Site Names

- Find the names of the unique launch sites
- This query returned all launch sites between 2010 and 2020

```
%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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Returned 5 launches which launch site begin with CCA

```
%sql SELECT "Launch_Site" FROM SPACEXTBL WHERE "Launch_Site" LIKE "CCA%" LIMIT 5;
```

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

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Returned the sum of payload mass for boosters from all the records that have the customer attribute “NASA (CRS)”

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") as sum FROM SPACEXTBL WHERE "customer" = "NASA (CRS)" group by "Customer";
```

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

sum
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Returned the average of payload mass of boosters from NASA which have the “Booster_Version” attribute as “F9 v1.1”.

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") as average FROM SPACEXTBL WHERE "Booster_Version" = "F9 v1.1" group by "Booster_Versio
```

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

average
2928.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Grouped the records by landings on ground pad which were successful and retrieved the record with the minimum date.

```
%sql SELECT MIN("DATE") FROM SPACEXTBL WHERE "Landing_Outcome" = "Success (ground pad)";
```

```
 * 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

```
%sql SELECT "Booster_Version" FROM SPACEXTBL
where "Landing_Outcome" = "Success (drone ship)" and 4000 < "PAYLOAD_MASS__KG_" < 6000;
```

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

Booster_Version
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Listed distinct mission outcomes and the count of each one

```
%sql SELECT "Mission_Outcome" COUNT("Mission_Outcome") from SPACEXTBL GROUP BY "Mission_Outcome";
```

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

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

Boosters Carried Maximum Payload

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

- Used a nested query to retrieve the max payload mass then used that value to retrieve the booster version which had the highest 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

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Used substr to get month names from ‘Date’

```
%%sql
SELECT SUBSTR(Date, 4, 1) AS Month_name, Landing_Outcome, Booster_Version
FROM SPACEXTBL WHERE SUBSTR(Date, 0, 5) = '2015'
AND Landing_Outcome = 'Failure (drone ship)';
```

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

Month_name	Landing_Outcome	Booster_Version
5	Failure (drone ship)	F9 v1.1 B1012
5	Failure (drone ship)	F9 v1.1 B1015

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

Grouped the data by landing outcome and retrieved the landing outcome and count value for each outcome

```
%%sql
SELECT "Landing_Outcome", COUNT("Landing_Outcome")
from SPACEXTBL where "Date"
between '2010-06-04' and '2017-03-20' group by "Landing_Outcome" order by COUNT("Landing_Outcome") desc;
```

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

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

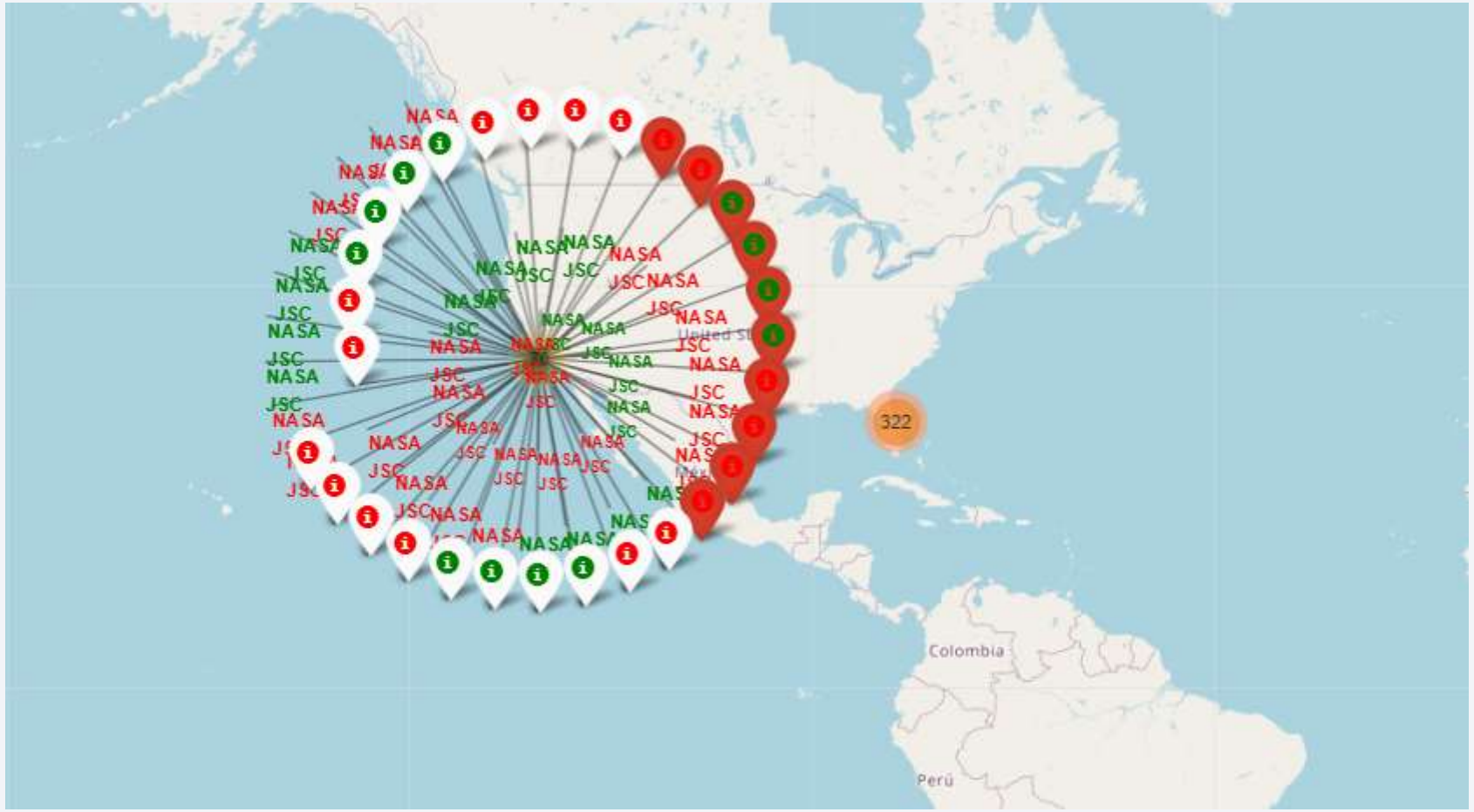
Section 3

Launch Sites Proximities Analysis

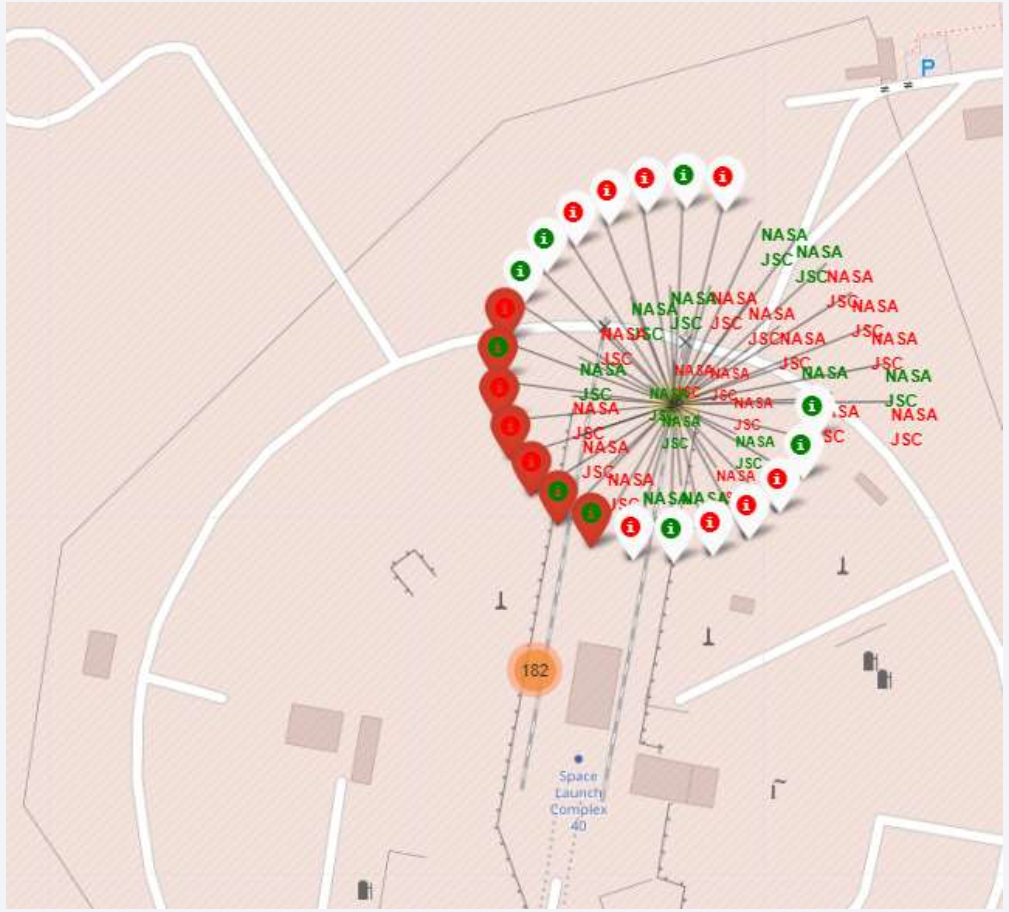
<Folium Map Screenshot 1>

<Folium Map Screenshot 2>

- Launch sites map

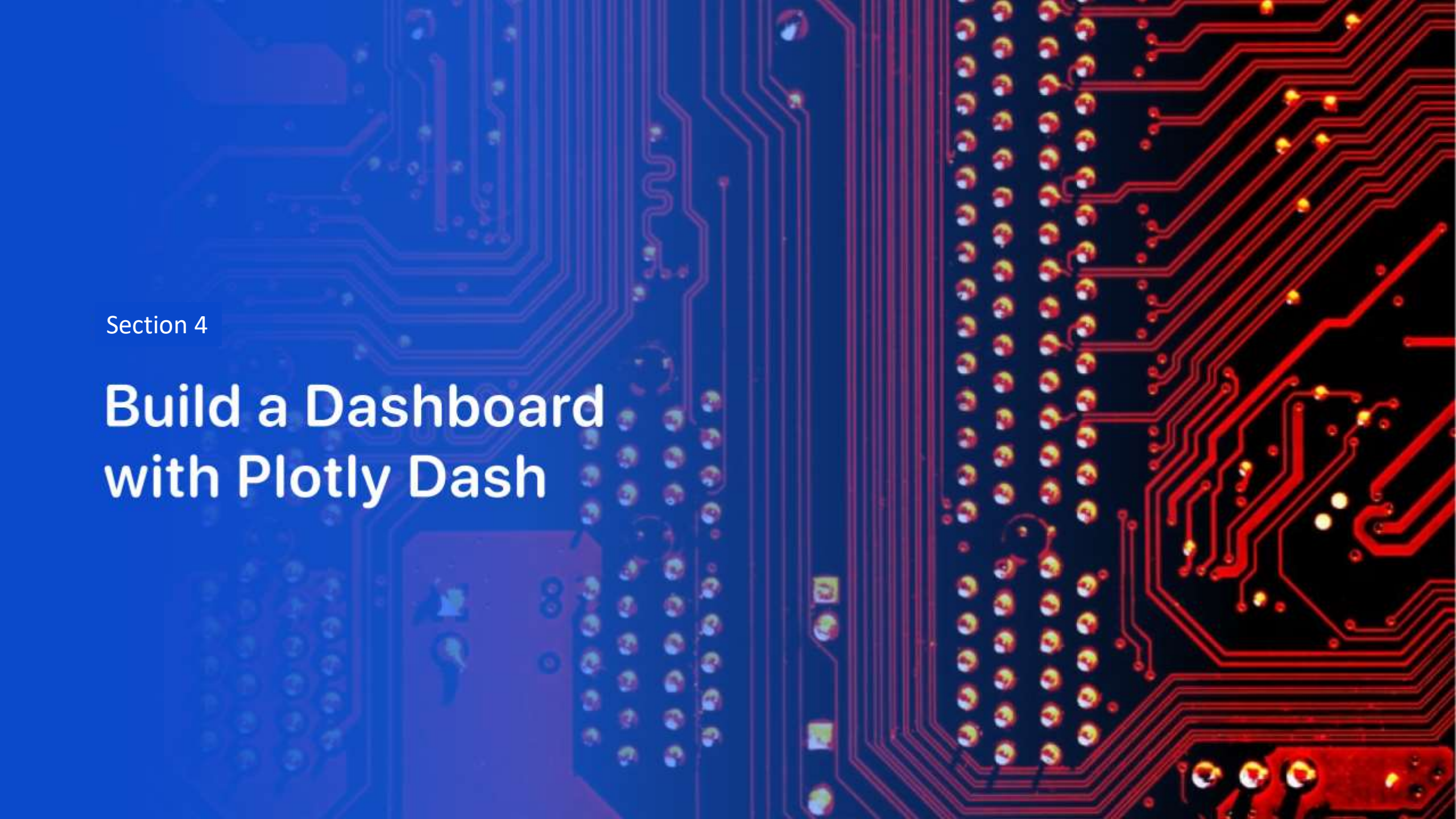


<Folium Map Screenshot 3>



Section 4

Build a Dashboard with Plotly Dash



<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

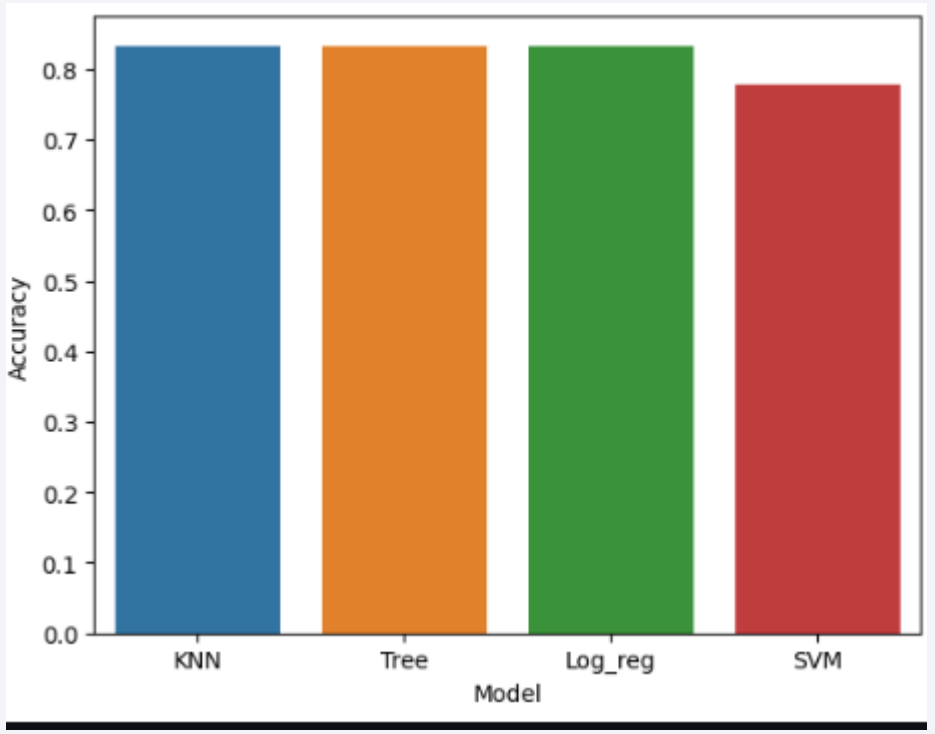
- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

Predictive Analysis (Classification)

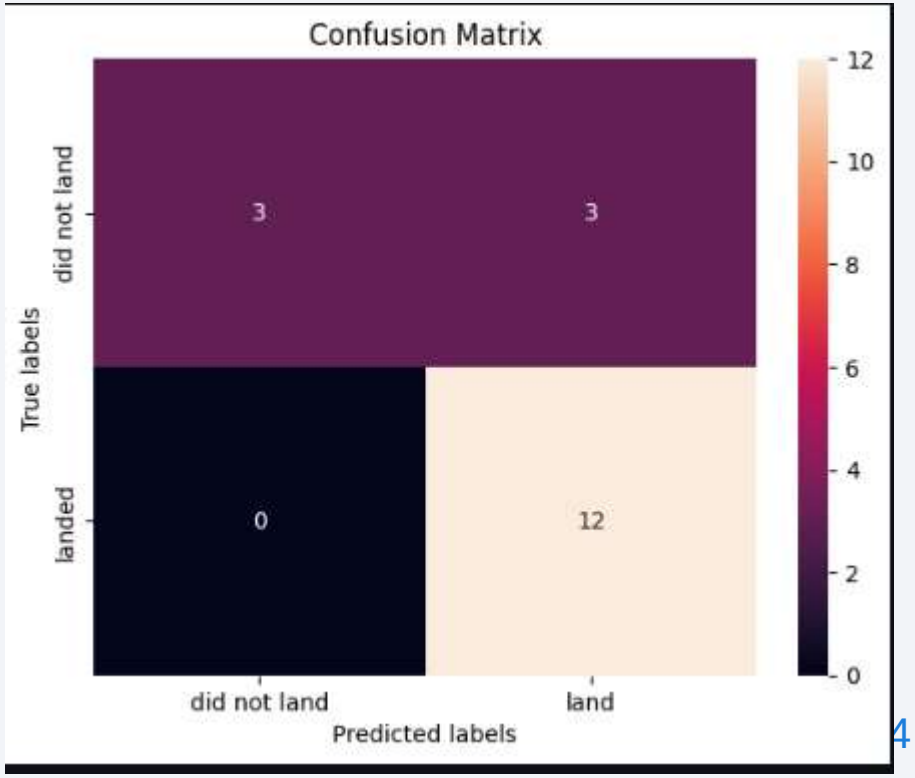
Classification Accuracy

- KNN,Decision tree classifier and logistic regression are all very similar in accuracy of 83.3%
- While SVM has an accuracy of 77.7%



Confusion Matrix

- KNN ,Logistic regression and decision tree are all tied in accuracy
- Having a true to false value ratio of 5:1



Conclusions

- Point 1
- Point 2
- Point 3
- Point 4
- …

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!