IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- Summary of methodologies
  - Data Collection using API
  - Data Collection with web scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualisation
  - Interactive Visual Analytics with Folium
  - Machine learning Prediction
- Summary of all results

Introduction

- Capstone project to take the role of a data scientist
- I’ll built a machine learning model with the use of public information to predict the successful landing of SpaceX rockets

Section 1

Methodology

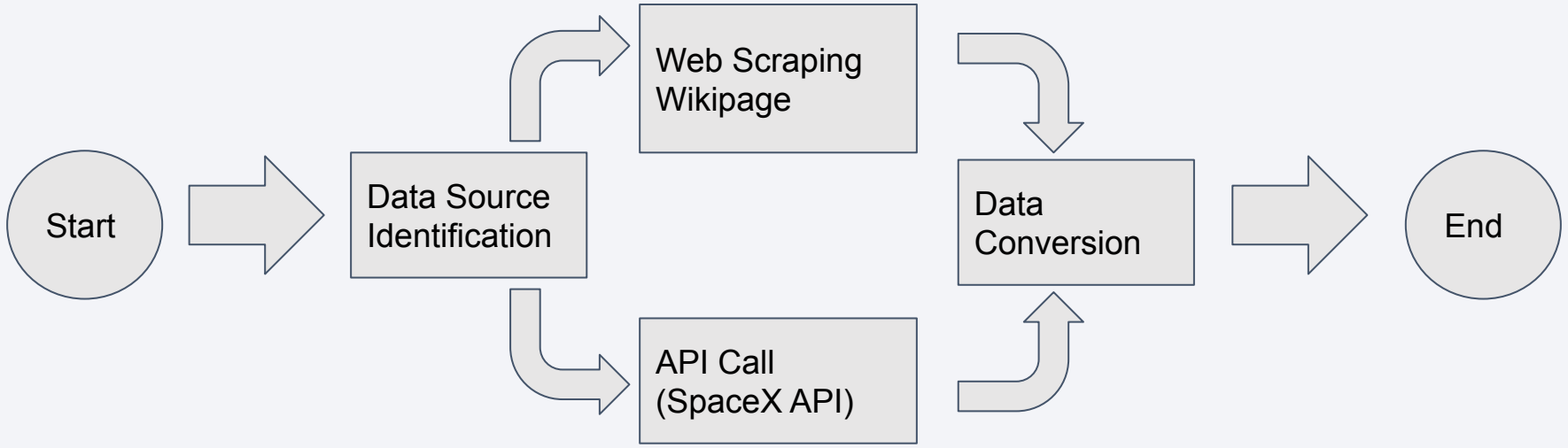
Methodology

Executive Summary

- Data collection methodology:
  - By scraping data from a static URL of a Wikipedia page that list all Falcon 9 and Falcon Heavy launches. Python and the request library is used to fetch the webpage content, and then used BeautifulSoup to parse the HTML content and extract the required information.
  - Additionally, another portion of the data is collected from a static JSON file using a an API Call. The JSON containing information about the rockets, payloads, launchpads, cores, flight number and data.
- Perform data wrangling
  - Calculated Numer of Launches on each site, calculated number of occurrences of each orbit, calculated the number of occurrence of mission outcome, and created a landing outcome label.
- 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

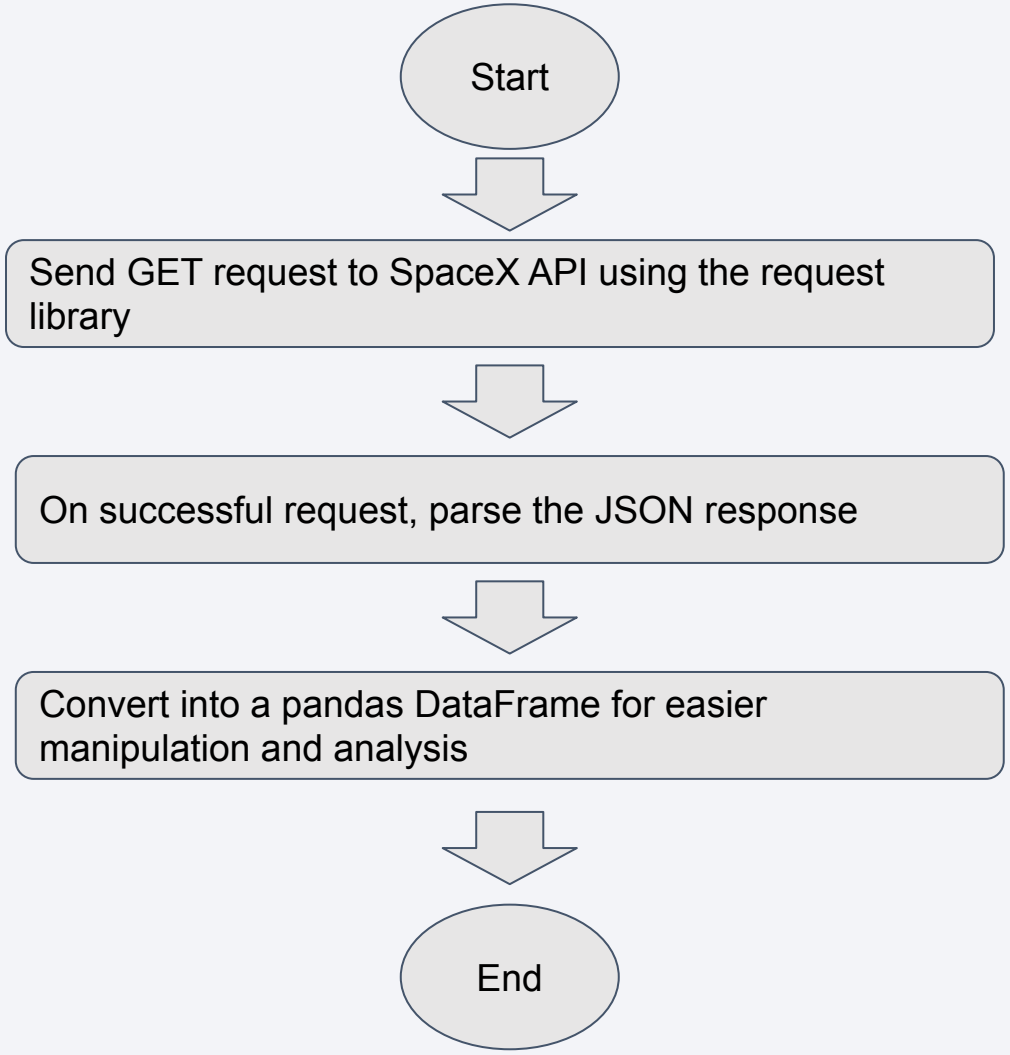
Data Collection

- Data collected by web scraping data from the List of Falcon 9 and Falcon Heavy launches Wikipage
- Data collected by HTTP request to SpaceX API



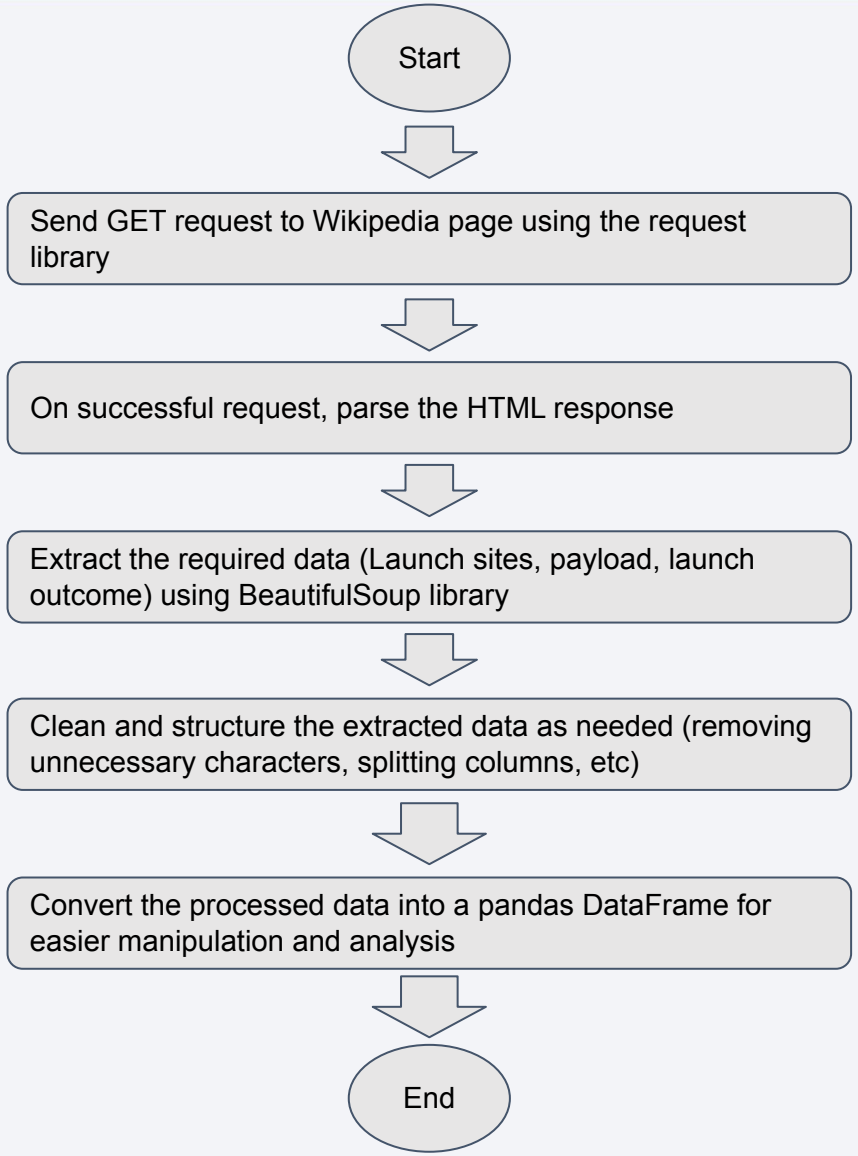
Data Collection – SpaceX API

- Flowchart overview of the data collection process through SpaceX API. Full notebook can be seen in URL below.
- Github URL:
  - https://github.com/georn/ibm-applied-data-science-project/blob/master/DataCollection/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Web Scraping

- Flowchart overview of the data collection process through webscraping the Wikipage. Full notebook can be seen in URL below.
- Github URL:
  - https://github.com/georn/ibm-applied-data-science-project/blob/master/DataCollection/jupyter-labs-webscraping.ipynb



Data Wrangling

- Calculates the number on each site
- Calculates the number of occurrences of each orbit
- Calculate the number and occurrence of mission outcome per orbit type
- Create a landing outcome label from Outcome column
- Github URL
    - https://github.com/georn/ibm-applied-data-science-project/blob/master/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

EDA with Data Visualization

- Charts:
  - Scatter Plot: Flight Number vs Payload Mass, Flight Number vs Launch Sites, Payfload vs Launch Sites, Payload vs Orbit Type
  - Bar Chart: Launch Success rate of each orbits
  - Line Plot: Launch Success rate vs Dates

- Github URL
  - https://github.com/georn/ibm-applied-data-science-project/blob/master/Visualisations/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Displays the names of the unique launch sites
- Display 5 records where launch sites begin with ‘CCA’
- Display the total payload mass carried by boosters launched by NASA
- Display average payload mass carried by booster F9 v1.1
- List the date when the first successful landing outcome in ground was achieved
- List of names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List of total number of successful and failure mission outcomes
- List the names of the booster versions which have carried the maximum payload mass.
- List the failed landing outcomes in drone ship and launch sites names for in year 2015
- Rank the counting of landing outcomes

- Github URL:
  - https://github.com/georn/ibm-applied-data-science-project/blob/master/dataWrangling/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Folium markers were implemented to pinpoint SpaceX launch sites and their surrounding notable landmarks, which include railways, highways, cities, and coastlines.

  Polylines were utilized to establish connections between the launch sites and their closest landmarks.

- A red coloration signifies instances of rocket launch failures.
- Successes in rocket launches are represented by the color green.

- Github URL
  - https://github.com/georn/ibm-applied-data-science-project/blob/master/VisualAnalyticsFolium/lab_jupyter_launch_site_location.ipynb

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)

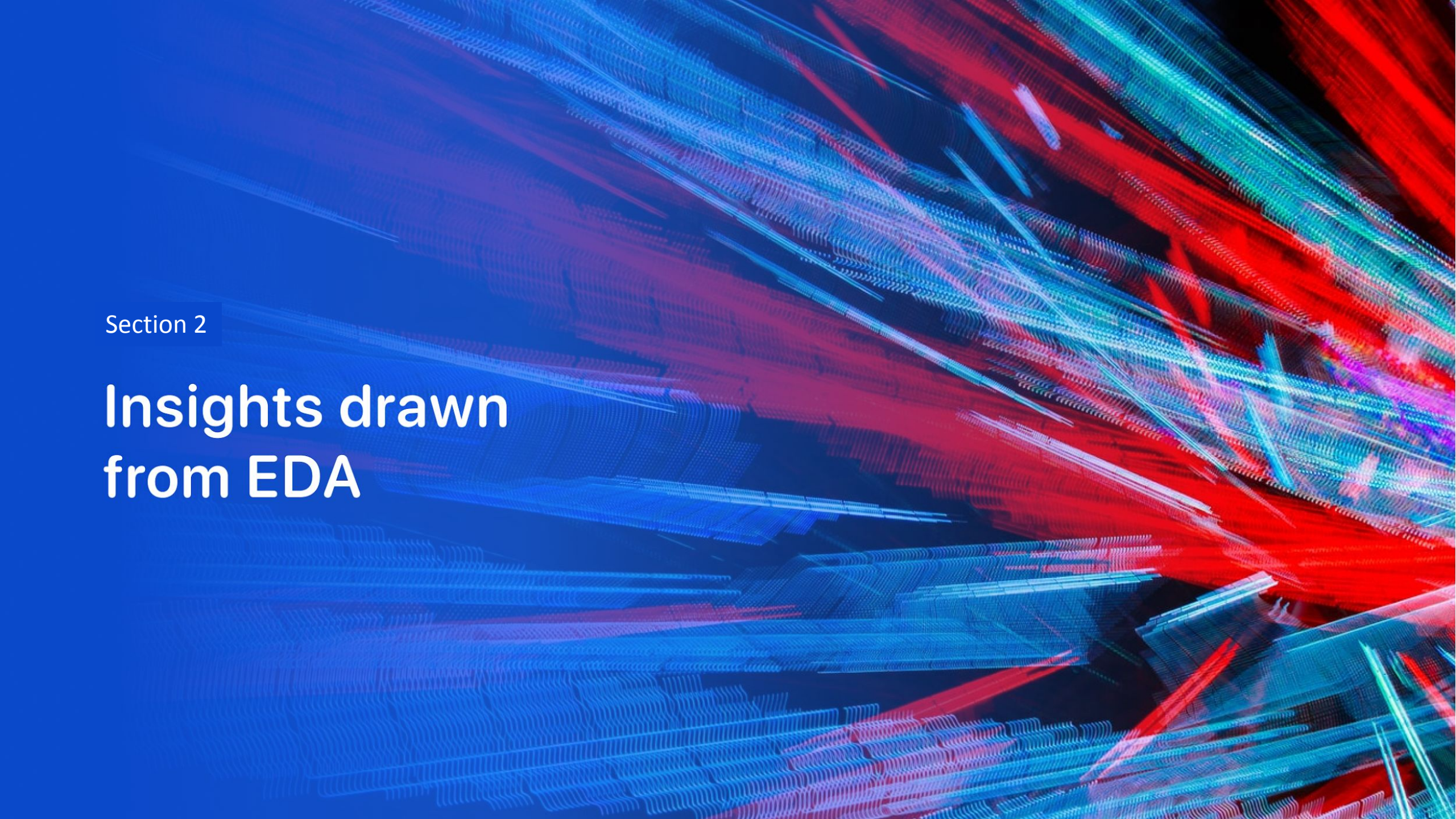
- Using Scikit-learn creating a machine learning pipeline to predict if the stage will land given data
- Github URL
  - https://github.com/georn/ibm-applied-data-science-project/blob/master/PredictiveAnalysis/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

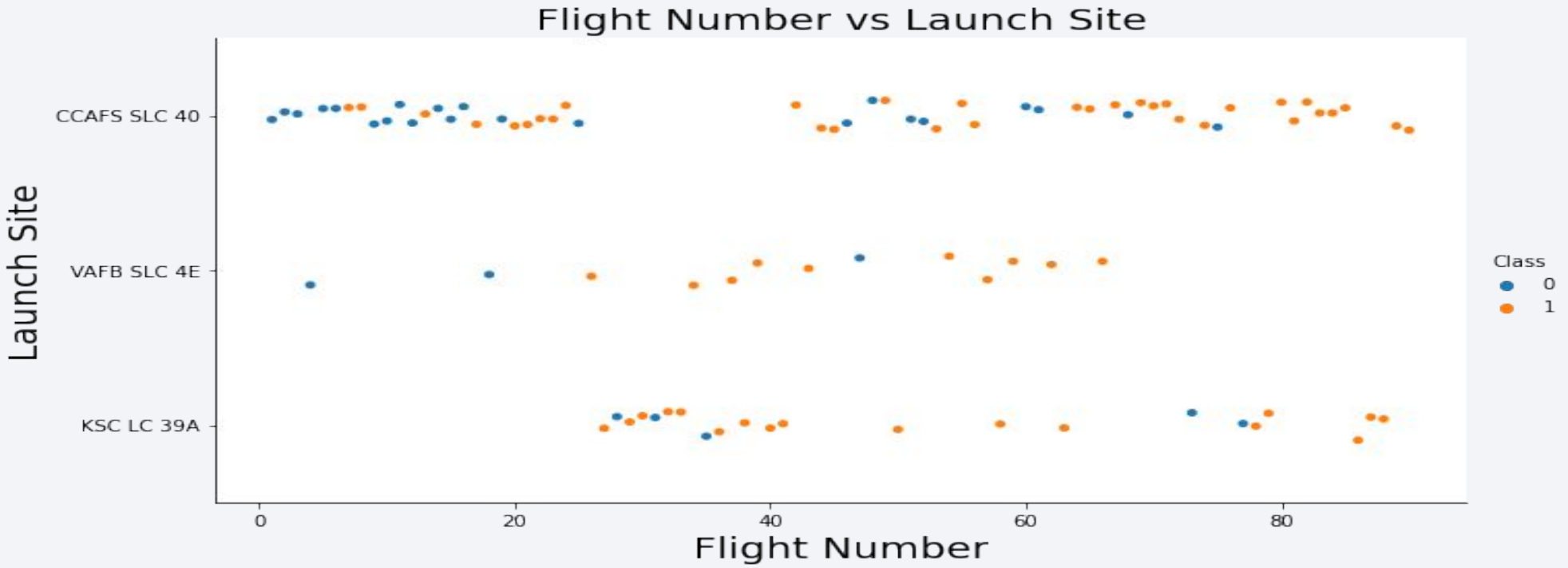
- It was shown that successful landing outcomes had increased since 2015
- Launch sites located near the coast line and away from populated areas to easier test the rockets
- Near highways and railways to help transportation of equipments and rocket pieces
- Machine learning able to predict the succes of rockets with an accuracy score of 83.33%

Section 2

Insights drawn from EDA

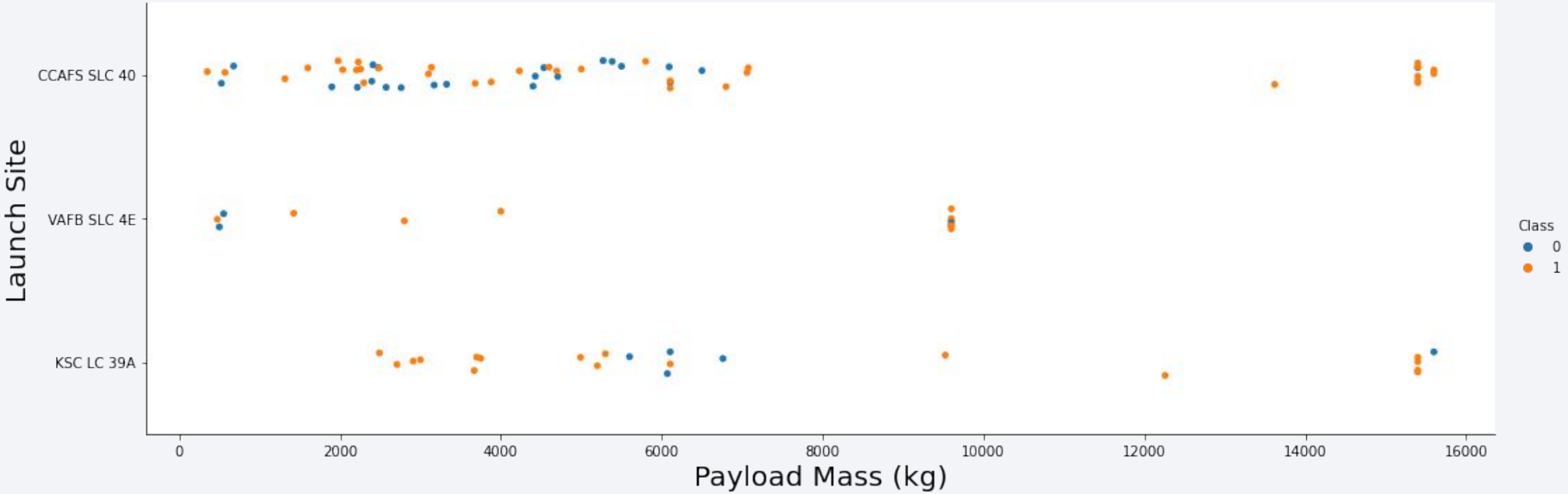


Flight Number vs. Launch Site



It seems that the number of successful landings grew in correlation with the increase in flight numbers. Notably, CCAFS SLC 40 was the launch site with the highest number of landings.

Payload vs. Launch Site

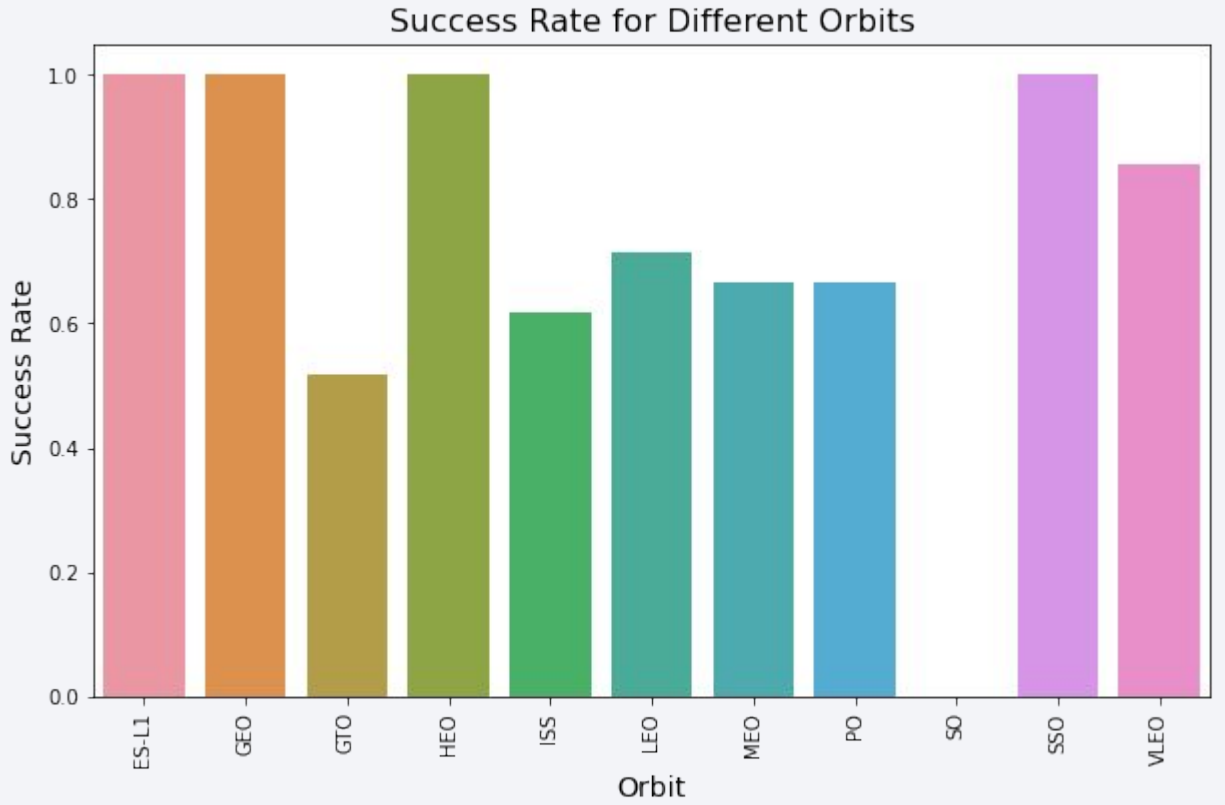


Upon examining the scatter point chart, it becomes apparent that for the VAFB-SLC launch site, there haven't been any rocket launches carrying a heavy payload mass exceeding 10,000 kg.

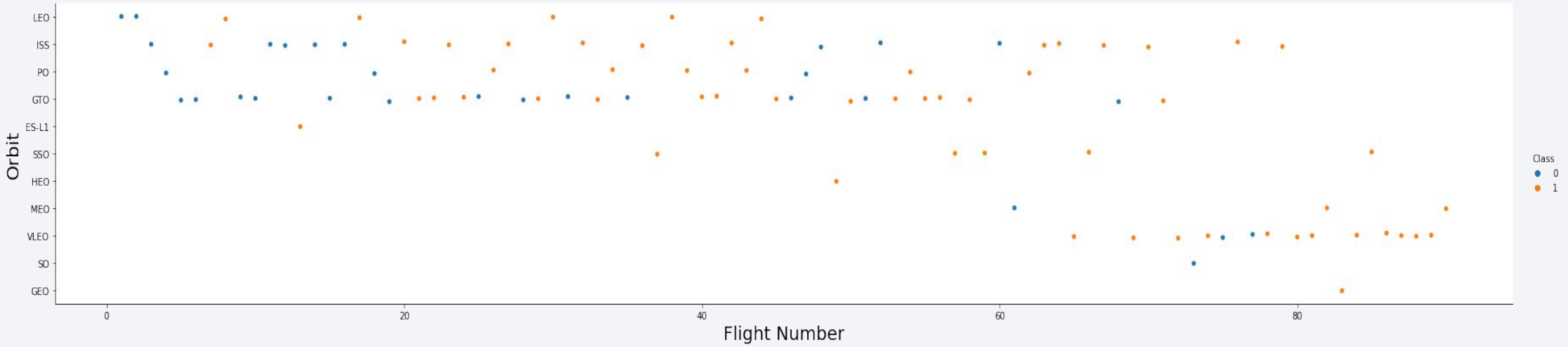
Success Rate vs. Orbit Type

The highest success rate Orbits are

- ES-L1
- GEO
- SSO
- HEO

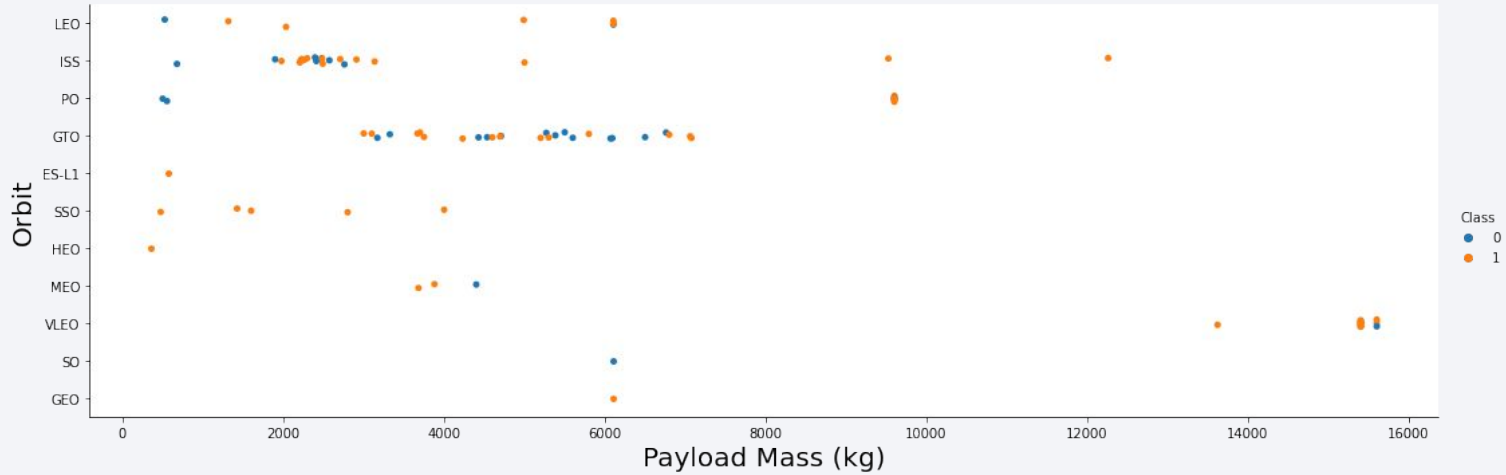


Flight Number vs. Orbit Type



It is evident that for the LEO orbit, success seems to correlate with the number of flights. However, no such relationship is apparent between flight number and success rate for GTO orbit.

Payload vs. Orbit Type



For heavier payloads, the rate of successful or positive landings is higher for Polar LEO and ISS. Conversely, for GTO, it's harder to discern a pattern as both successful and unsuccessful missions are quite common.

Launch Success Yearly Trend

It is shown that the success rate has significantly increased from 2013 to 2020

All Launch Site Names

- All Launch Site Names
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

Launch Site Names Begin with 'CCA'

Task 2

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

In 12

```
query = "SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%' LIMIT 5"
records = pd.read_sql(query, con)
print(records)
```

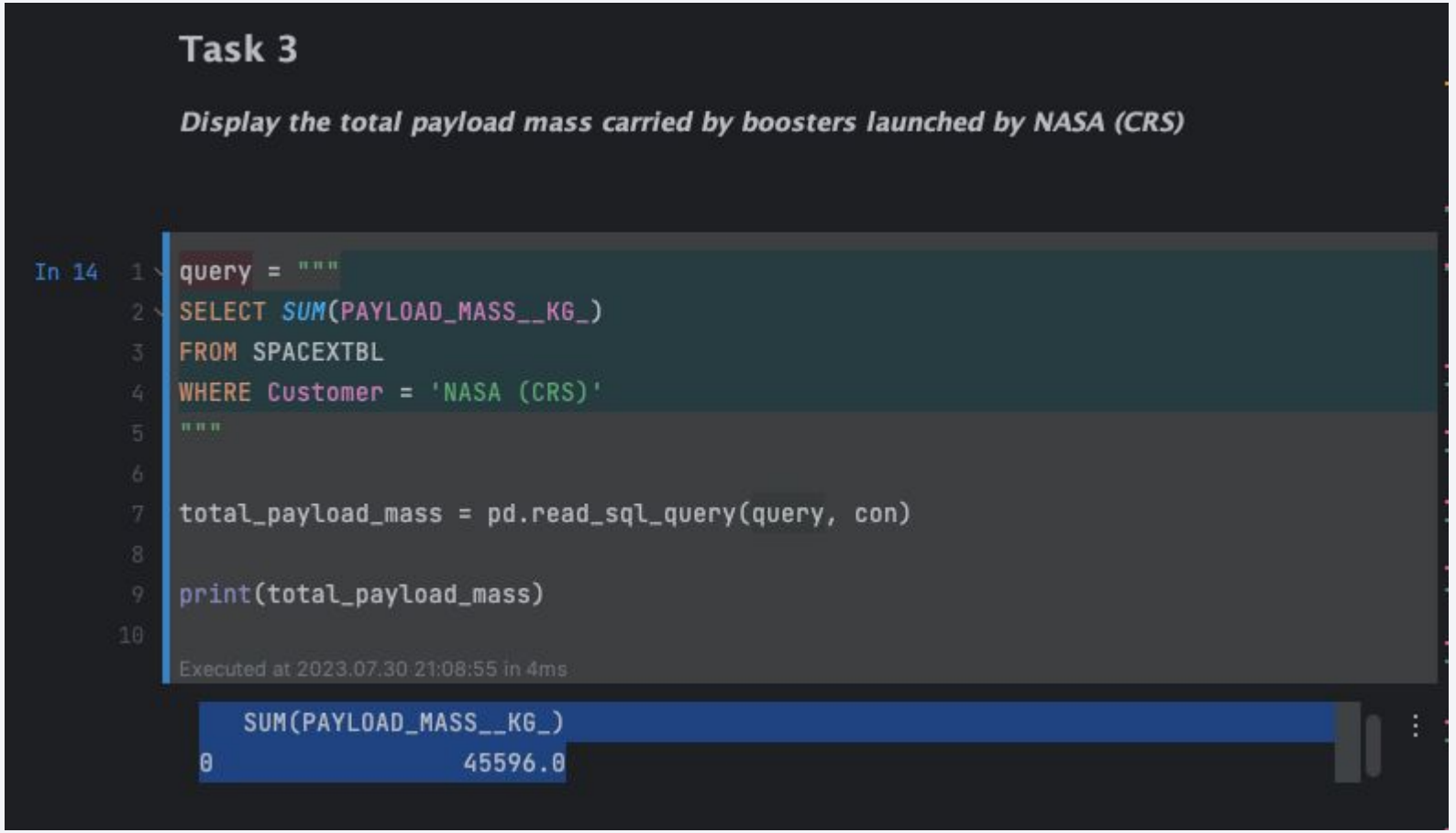
Executed at 2023.07.19 21:10:20 in 96ms

```
         Date Time (UTC) Booster_Version  Launch_Site  \
0  06/04/2010   18:45:00  F9 v1.0  B0003  CCAFS LC-40
1  12/08/2010   15:43:00  F9 v1.0  B0004  CCAFS LC-40
2  22/05/2012    7:44:00  F9 v1.0  B0005  CCAFS LC-40
3  10/08/2012    0:35:00  F9 v1.0  B0006  CCAFS LC-40
4  03/01/2013   15:10:00  F9 v1.0  B0007  CCAFS LC-40
```

It is shown the 5 records where launch sites begin with letters 'CCA'

Total Payload Mass

It is shown that the total mass carried by Boosters launched by NASA



Average Payload Mass by F9 v1.1

It is shown the average payload mass carried by a F9 v1.1 was 2928.4

Task 4

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

In 16

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

avg_payload_mass = pd.read_sql_query(query, con)

print(avg_payload_mass)
```

Executed at 2023.07.30 21:10:56 in 4ms

	AVG(PAYLOAD_MASS__KG_)
0	2928.4

First Successful Ground Landing Date

From the result it says 01/08/2018

Task 5

List the date when the first successful landing outcome in ground pad was acheived.

Hint:Use min function

In 16

```
query = """
SELECT MIN("Date")
FROM SPACEXTBL
WHERE "Landing_Outcome" = 'Success (ground pad)'
"""

first_successful_landing_date = pd.read_sql_query(query, con)

print(first_successful_landing_date)
```

Executed at 2023.07.19 21:14:13 in 5ms

```
   MIN("Date")
0  01/08/2018
```

Successful Drone Ship Landing with Payload between 4000 and 6000

The only 4 boosters with payload mass between 4000kg and 6000kg

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

In 17

```
query = """
SELECT "Booster_Version"
FROM SPACEXTBL
WHERE "Landing_Outcome" = 'Success (drone ship)'
AND "Payload_Mass__kg_" BETWEEN 4000 AND 6000
"""

boosters = pd.read_sql_query(query, con)

print(boosters)

```

Executed at 2023.07.19 21:14:48 in 5ms

```
  Booster_Version
0     F9 FT B1022
1     F9 FT B1026
2  F9 FT  B1021.2
3  F9 FT  B1031.2
```

Total Number of Successful and Failure Mission Outcomes

It is shown the total number of successful and failure mission outcomes

Task 7

List the total number of successful and failure mission outcomes

In 18

```
query = """
SELECT "Mission_Outcome", COUNT(*) as Number_of_Missions
FROM SPACEXTBL
GROUP BY "Mission_Outcome"
"""

mission_outcomes = pd.read_sql_query(query, con)

print(mission_outcomes)
```

Executed at 2023.07.19 21:15:33 in 3ms

	Mission_Outcome	Number_of_Missions
0	None	898
1	Failure (in flight)	1
2	Success	98
3	Success	1
4	Success (payload status unclear)	1

Boosters Carried Maximum Payload

It is shown that 12 boosters have carried the maximum payload mass.

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

In 19

```
query = """
SELECT "Booster_Version"
FROM SPACEXTBL
WHERE "Payload_Mass__kg_" = (SELECT MAX("Payload_Mass__kg_") FROM SPACEXTBL)
"""

booster_versions = pd.read_sql_query(query, con)

print(booster_versions)
```

Executed at 2023.07.19 21:15:46 in 9ms

```
   Booster_Version
0    F9 B5 B1048.4
1    F9 B5 B1049.4
2    F9 B5 B1051.3
3    F9 B5 B1056.4
4    F9 B5 B1048.5
5    F9 B5 B1051.4
6    F9 B5 B1049.5
7   F9 B5 B1060.2
8   F9 B5 B1058.3
9    F9 B5 B1051.6
10   F9 B5 B1060.3
11  F9 B5 B1049.7
```

2015 Launch Records

Two boosters F9 v1.1 B1012 and F9 v1.1 01015 from CCAFS LC-40 failed to land at 2015

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.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.

In 18

```
query = """
SELECT *
FROM SPACEXTBL
WHERE strftime('%Y', Date) = '2015'
UNION
SELECT *
FROM SPACEXTBL
WHERE "Landing_Outcome" LIKE 'Failure (drone ship)%'
"""

records = pd.read_sql_query(query, con)
print(records)
```

Executed at 2023.07.30 21:22:06 in 4ms

	Date	Time (UTC)	Booster_Version	Launch_Site
0	01/10/2015	9:47:00	F9 v1.1 B1012	CCAFS LC-40
1	03/04/2016	23:35:00	F9 FT B1020	CCAFS LC-40
2	14/04/2015	20:10:00	F9 v1.1 B1015	CCAFS LC-40
3	15/06/2016	14:29:00	F9 FT B1024	CCAFS LC-40
4	17/01/2016	18:42:00	F9 v1.1 B1017	VAFB SLC-4E

	Payload	PAYLOAD_MASS__KG_	Orbit
0	SpaceX CRS-5	2395.0	LEO (ISS)
1	SES-9	5271.0	GTO
2	SpaceX CRS-6	1898.0	LEO (ISS)

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

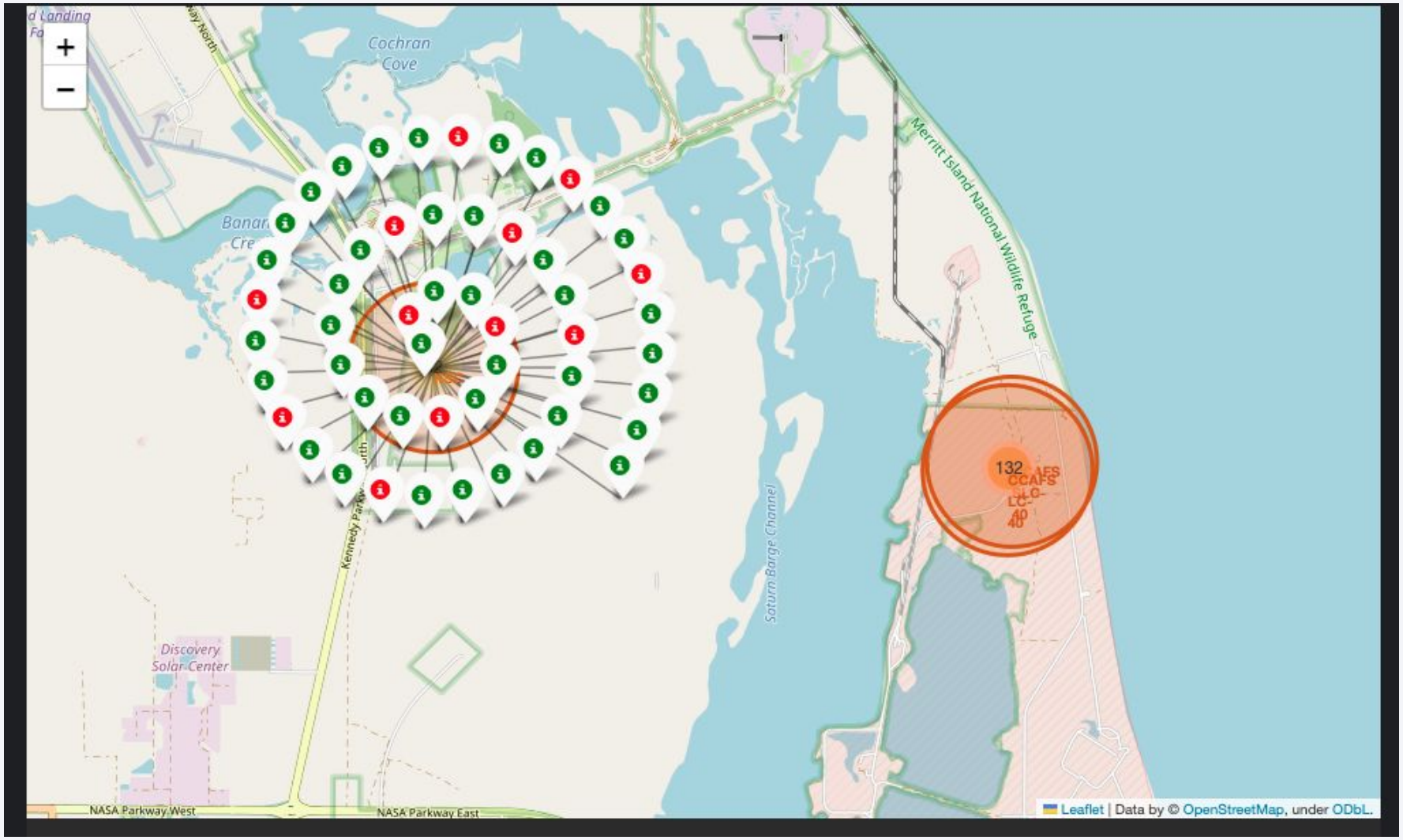

Section 3
Launch Sites
Proximities Analysis

Launch sites Locations

Each launch site is located in close proximity to the coast and couple thousand kilometers away from the equator.

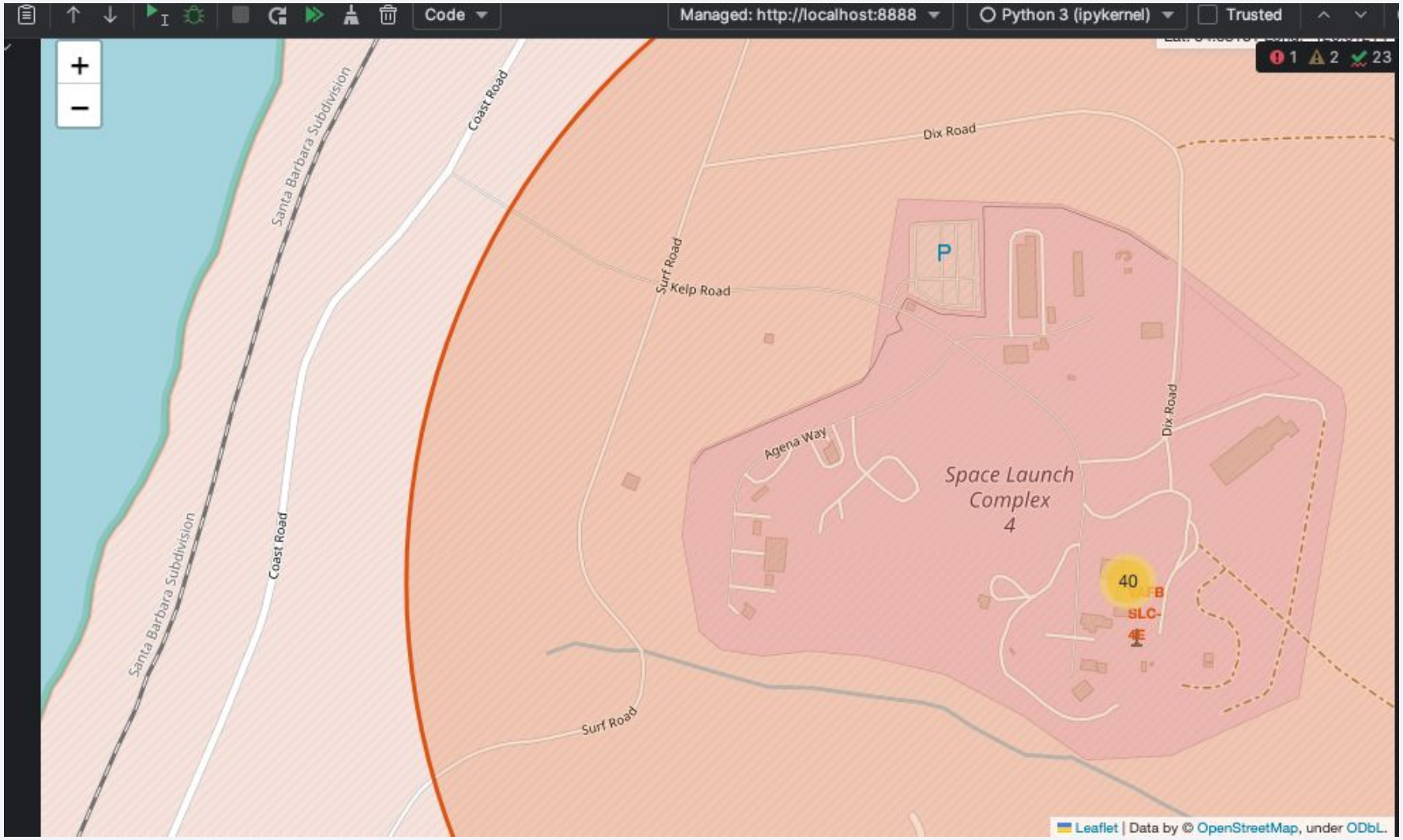
Success Rate of Rocket Launches

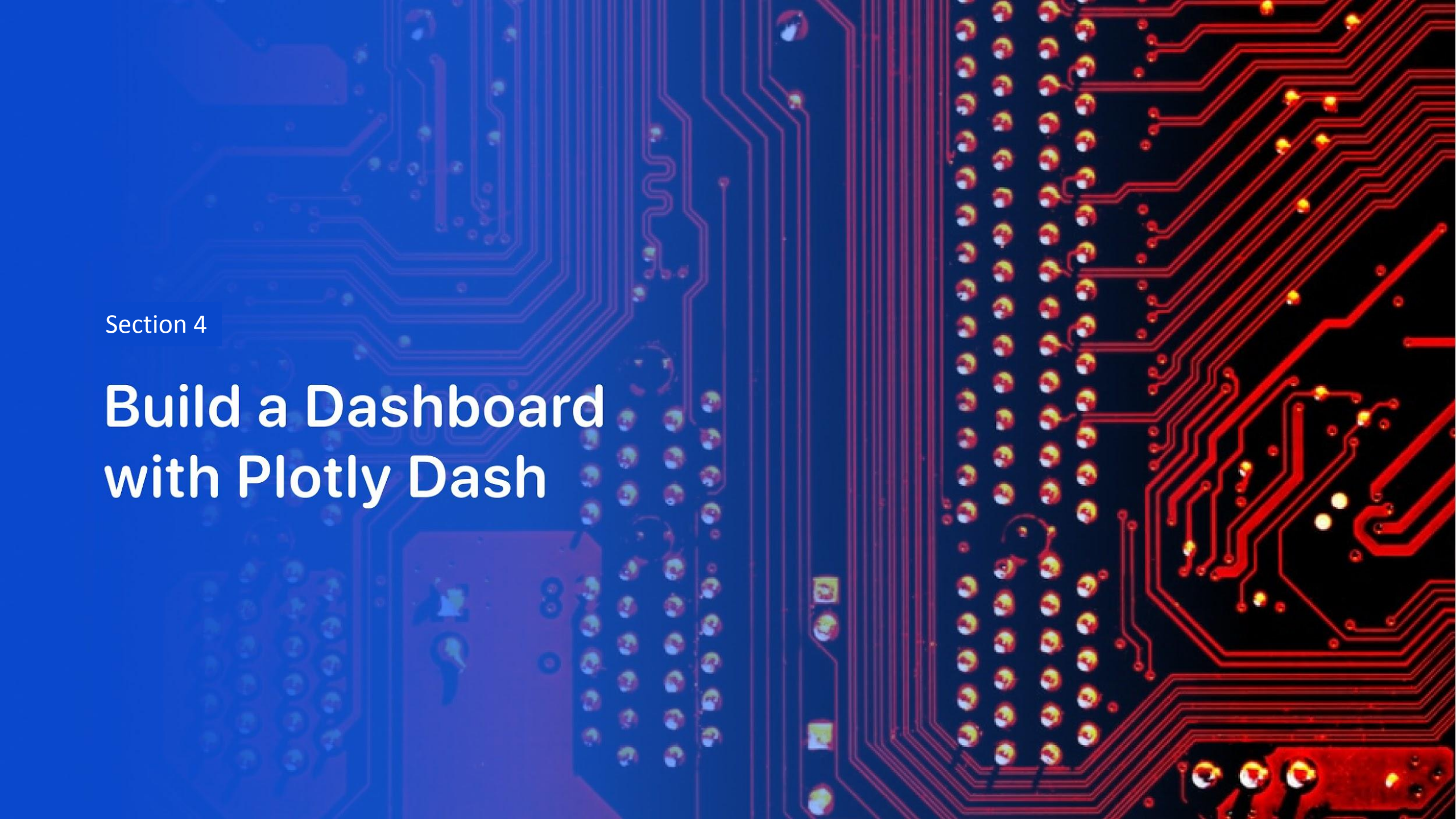
The green markers denote successful rocket launches, while the red markers indicate failed launches.



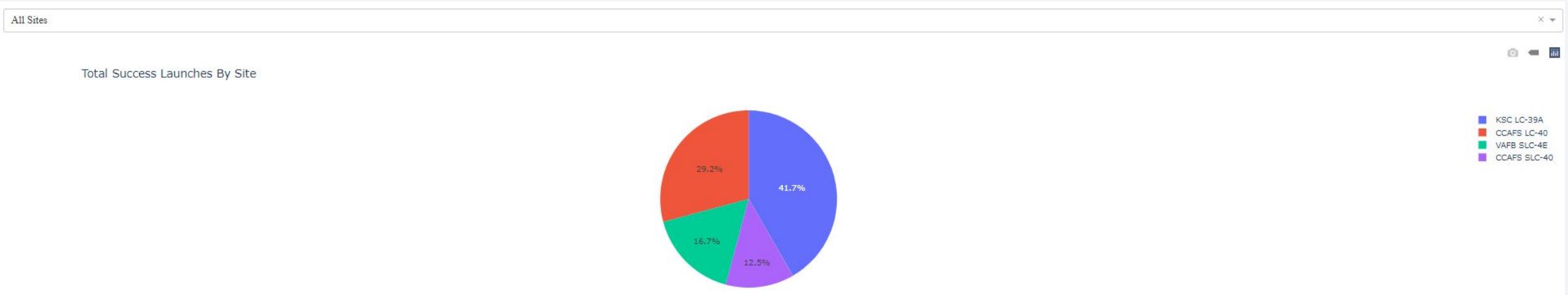
Surrounding Landmarks

It seems that launch sites are typically situated far from urban areas, with close proximity to railways and highways. They are also usually located near the coastline.



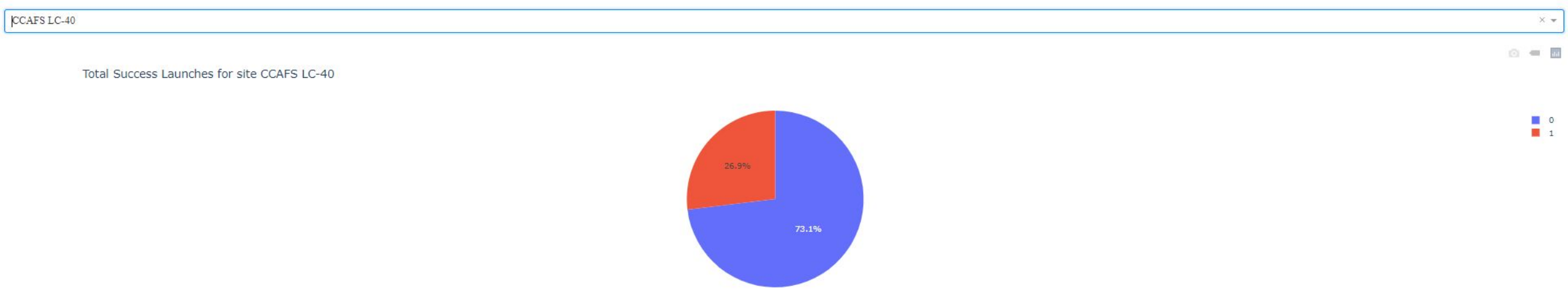

Section 4
Build a Dashboard with Plotly Dash

Successful Launches by Site



The plot illustrates that Site KSC LC-39A not only has the most successful launches but also boasts the highest rate of successful launches.

Total Successful Launches for Site CCAFS LC-40



The launch site CCAFS LC-40 has demonstrated a success rate among all SpaceX launch sites, with 73.1% of its total launches being successful.

Payload vs Launch Sucess for all Sites

The highest success rate is observed when the payload falls within the range of 2000kg to 4000kg.

Section 5

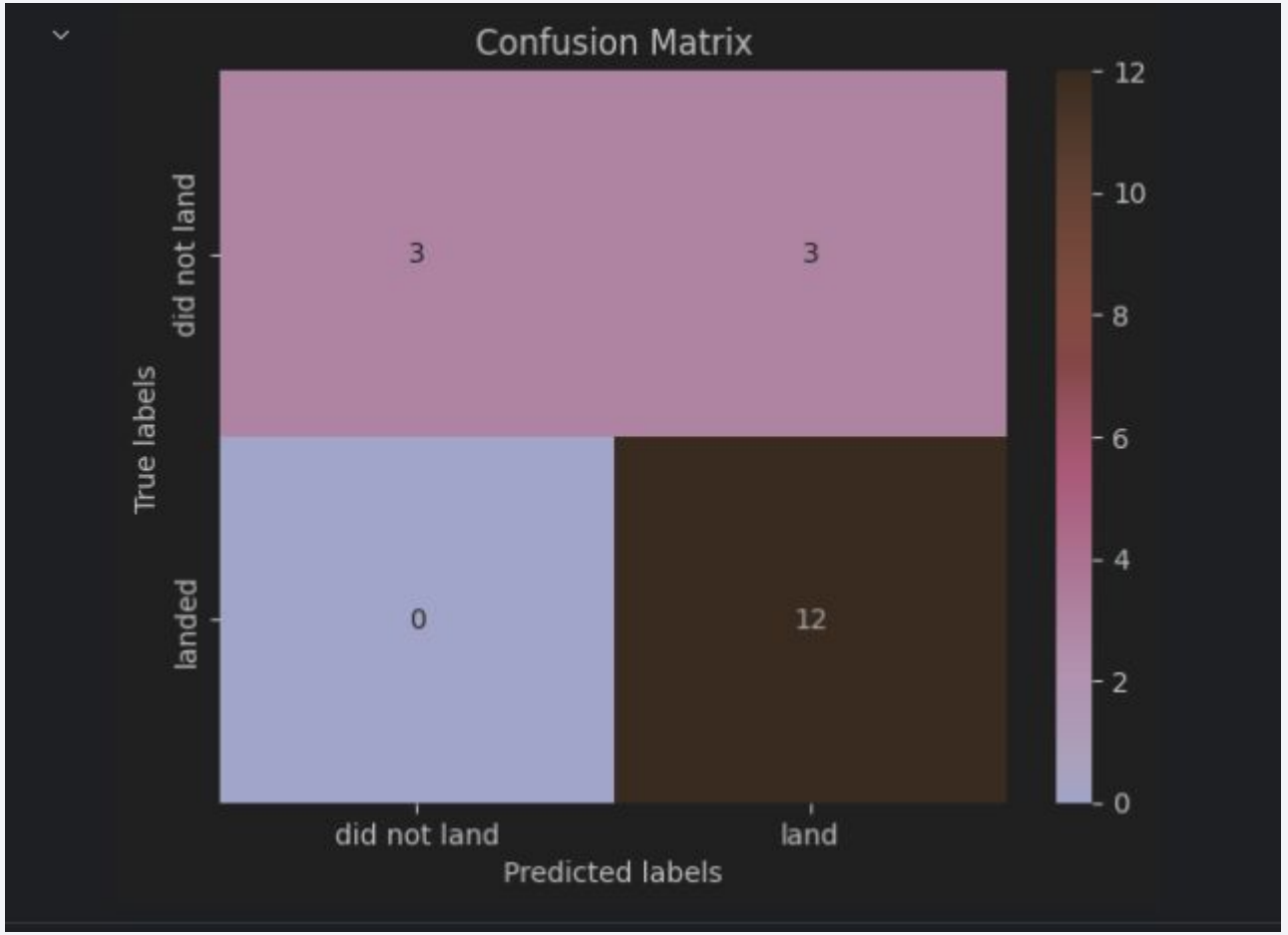
Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy

Confusion Matrix



Conclusions

To compete effectively with SpaceX, it’s crucial to understand their strategies for success:

- SpaceX strategically locates their launch sites near coastlines and away from populated areas, allowing them to conduct rocket landing tests with minimal disruption.
- Among all the launch sites, KSC LC 39A has demonstrated the highest launch success rate.
- Since 2015, SpaceX has seen a significant increase in landing success rates. Moreover, a clear trend shows that the success rate improves with an increase in flight numbers.

Thank you!