IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies
- Data Collection
- Data Wrangling
- EDA with Data Visualization
- EDA with SQL
- An Interactive Map Using Folium
- Creating Plotly Dashboard
- Classification Tasks

Introduction

- **Project background and context:**

  This project aims to look at the possibility of the landing of Falcon successfully using multiple factor to make that determination

- **Problems you want to find answers:**

  The main problem to find its answer is the factors that will increase the possibilities of successful landing

Section 1

Methodology

Methodology

-Data collected using Space X API and Scrapping the Web.

-Perform data wrangling to clean and transform the data to utilize in finding the answer.

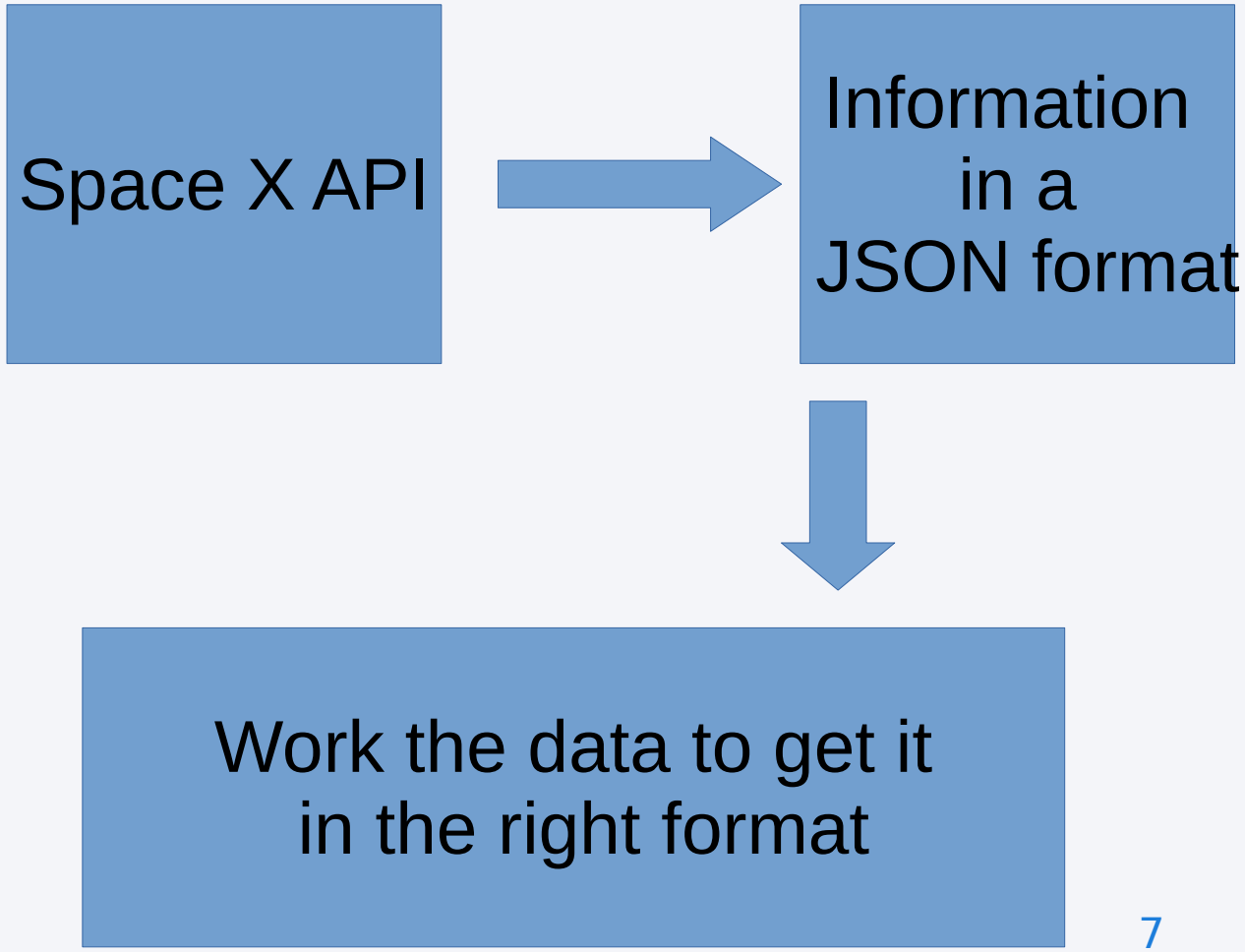
-Perform exploratory data analysis (EDA) using visualization and SQL by using different plots to see patterns in the data.

-Perform interactive visual analytics using Folium and Plotly Dash to mark locations that need for the task.

-Perform predictive analysis using classification models to determine the right classification model.

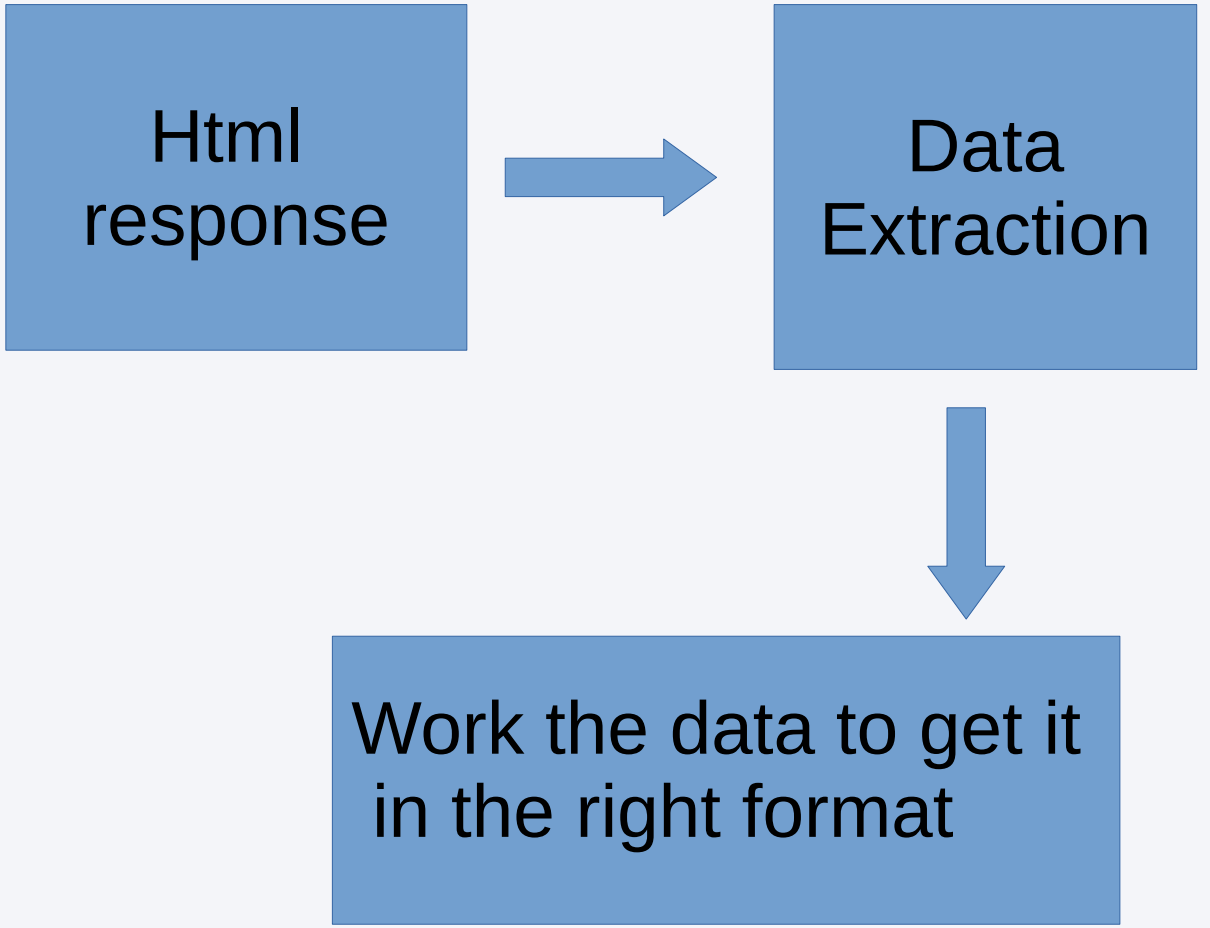
Data Collection – Space X API

- Using the Space X API to get the data from the API



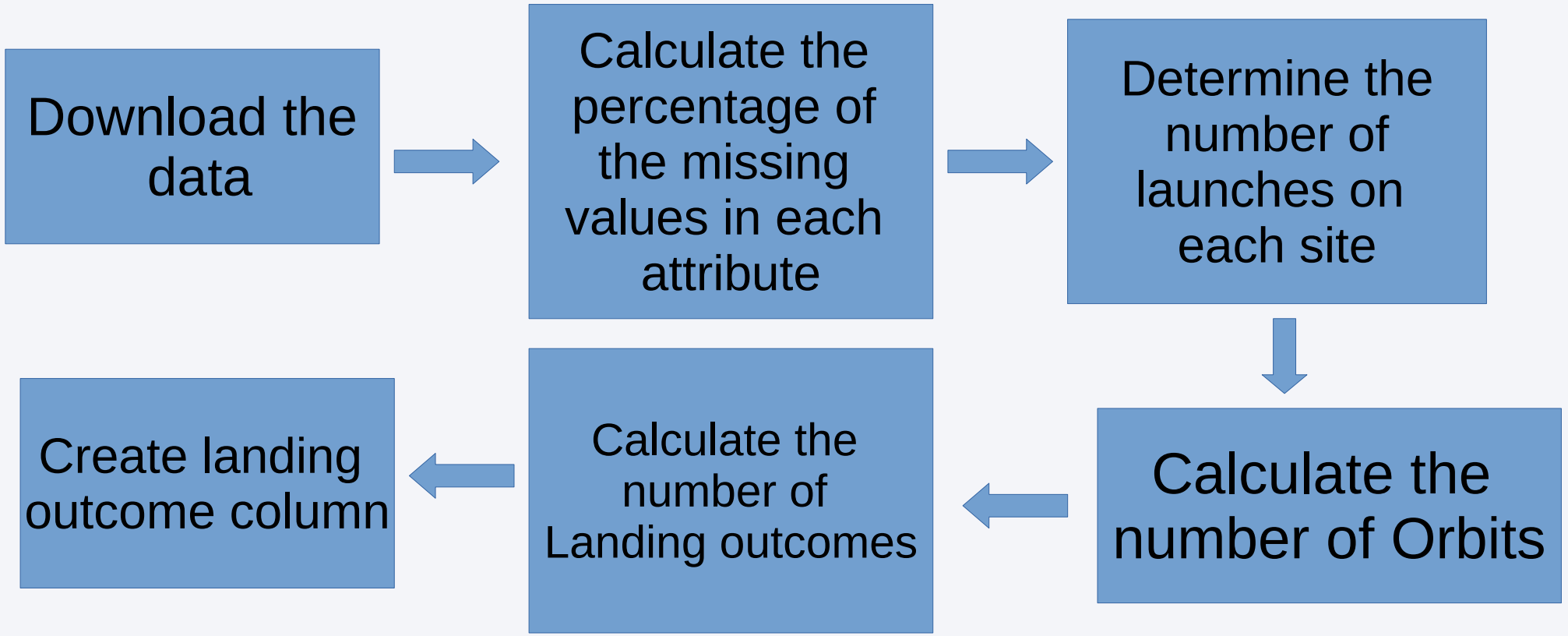
Data Collection - Scraping

- Web Scraping is a method to get data in html file by using request function to get a response which contains the data in an unstructured way, for which libraries like BS4 are built for.



Data Wrangling

GitHub URL of the completed data wrangling process notebook:



EDA with Data Visualization

GitHub URL of the completed EDA with data visualization notebook:

https://github.com/Moha02/Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb

Summary of the charts being used to plot

- Category plot for flight number versus launch site and overlay the outcome of the launch to observe the correlation between the launch site and flight number for successful launch.

- Category plot for payload mass versus launch site and overlay the outcome of the launch to observe the correlation between the launch site and pay load mass for successful launch

- Bar chart for success rate of every orbit to find the higher successful launch

- Scatter plot for the flight number and every orbit to observe the correlation between the orbit and the flight number for successful launch.

- Scatter chart show the payload mass and every orbit to observe the correlation between the orbit and pay load mass for successful launch.

- Line chart for the yearly successful launch trend to see the yearly trend of successful launch

EDA with SQL

GitHub URL of the completed EDA with SQL notebook

https://github.com/Moha02/Capstone/blob/main/eda-sql.ipynb

Summary of the SQL queries performed

- Display the names of the unique launch sites in the space mission

- 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

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

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

EDA with SQL

Summary of the SQL queries performed

- List the total number of successful and failure mission outcome

- List the names of the booster_versions which have carried the maximum payload mass.

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

Build an Interactive Map with Folium

GitHub URL of the completed Interactive Map with Folium notebook

https://github.com/Moha02/Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Summary of the created map objects added to a folium map

- A Circle map object to add a highlighted circle area on a specific coordinate for each launch site.
- A mark map object to specify the text label on a specific coordinate for each launch site.
- A cluster of mark map object with an icon to specify the success and failure of the launch.
- A mark map object to specify the distance measurement to the proximities such as city, railway, highway and coastline
- A PolyLine map object to draw a line from the launch site to the proximities

Build a Dashboard with Plotly Dash

GitHub URL of the completed Dashboard with Plotly notebook

https://github.com/Moha02/Capstone/blob/main/spacex_dash_app.py

Summary of the plots/graphs and interactions added to the dashboard

A Circle

Predictive Analysis (Classification)

GitHub URL of the completed Dashboard with Plotly notebook

https://github.com/Moha02/Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Summary of how you built, evaluated, improved, and found the best performing classification model

- The process begins by determining the X and Y values in the data.

- Afterward, is important to standardize and split the data to be used in the models.

- Apply the models by using the data

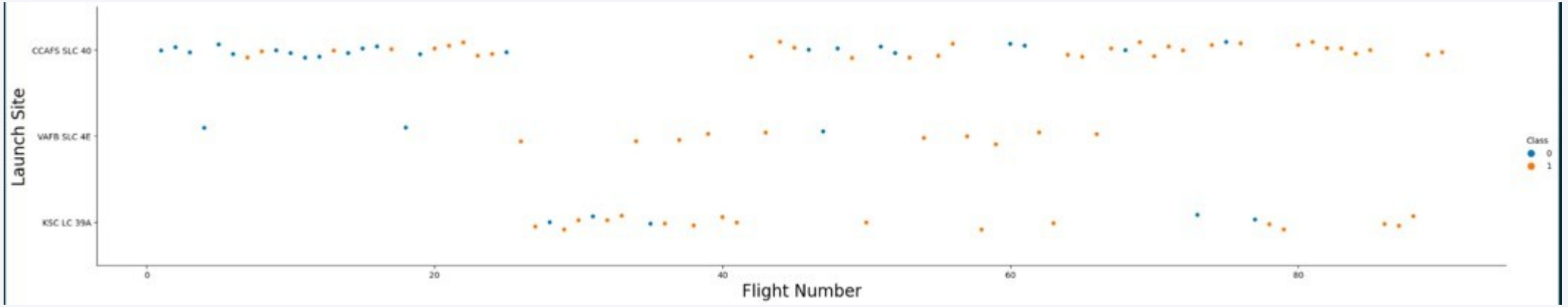
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site

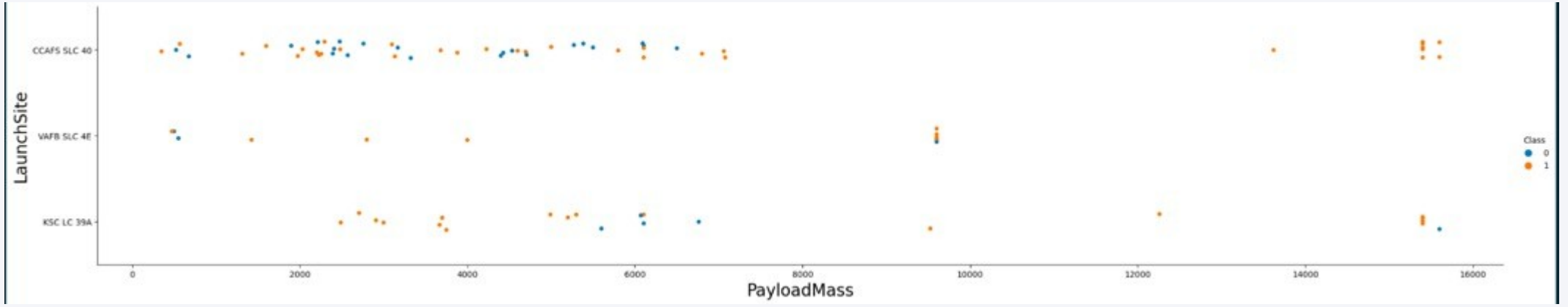
- Show the screenshot of the scatter plot with explanations



Payload vs. Launch Site

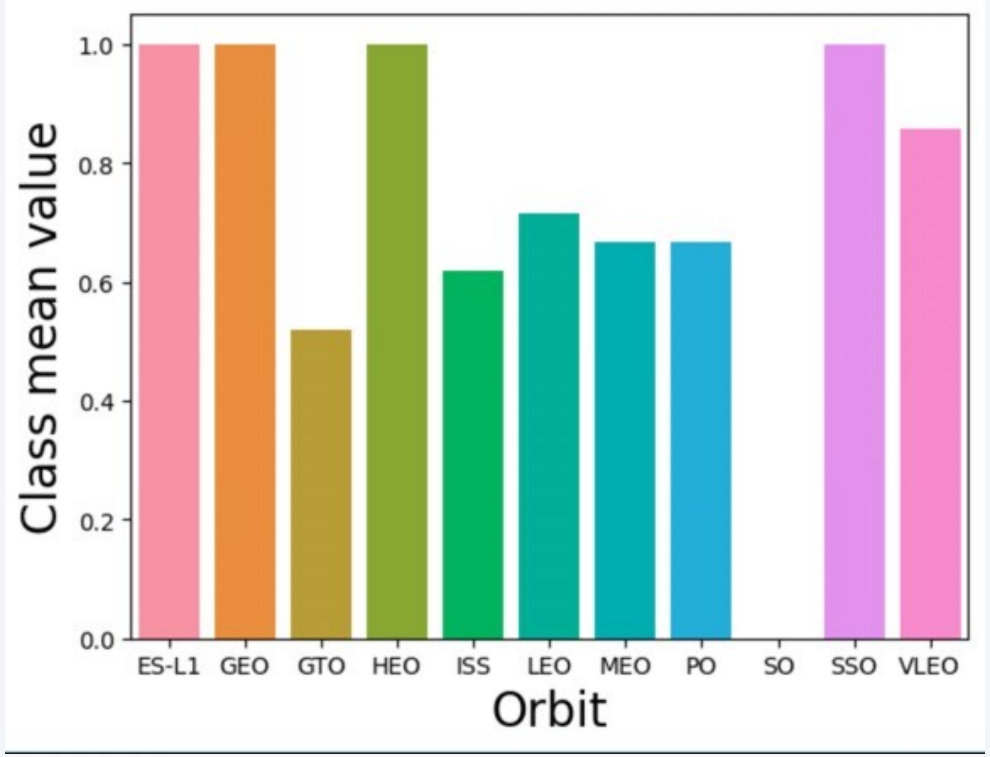
- Show a scatter plot of Payload vs. Launch Site

- Show the screenshot of the scatter plot with explanations



Success Rate vs. Orbit Type

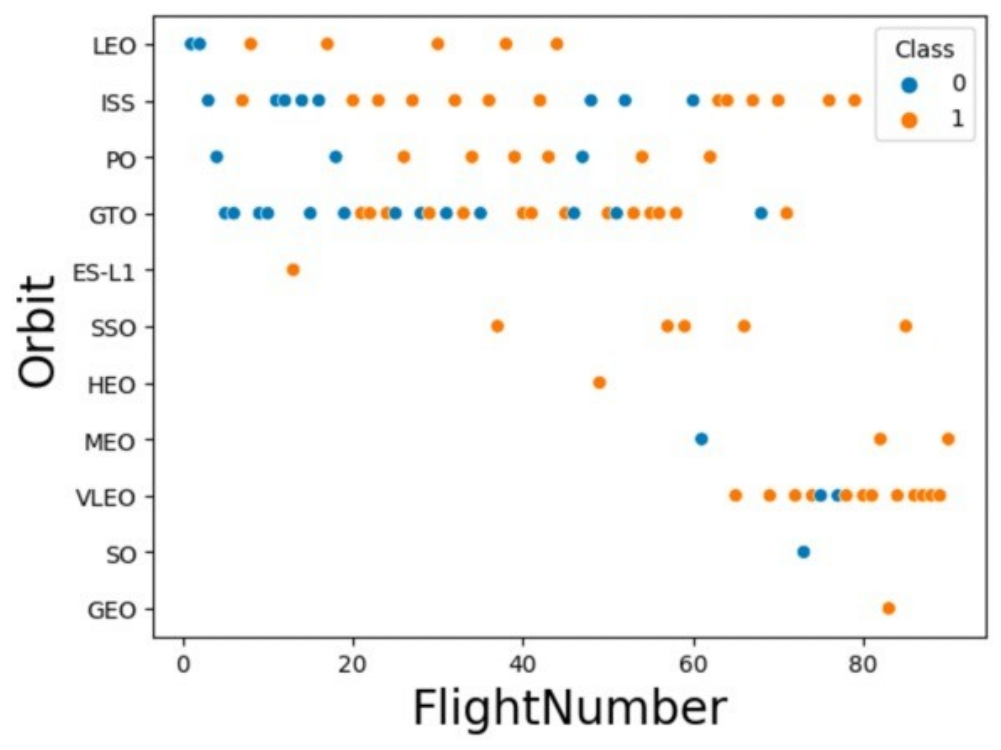
- Show a bar chart for the success rate of each orbit type
- Four Orbits have 100% success rate.



Flight Number vs. Orbit Type

•Show a scatter point of Flight number vs. Orbit type

The higher the flight number, the higher the successful probability in each orbit unlike orbit GTO and SO.

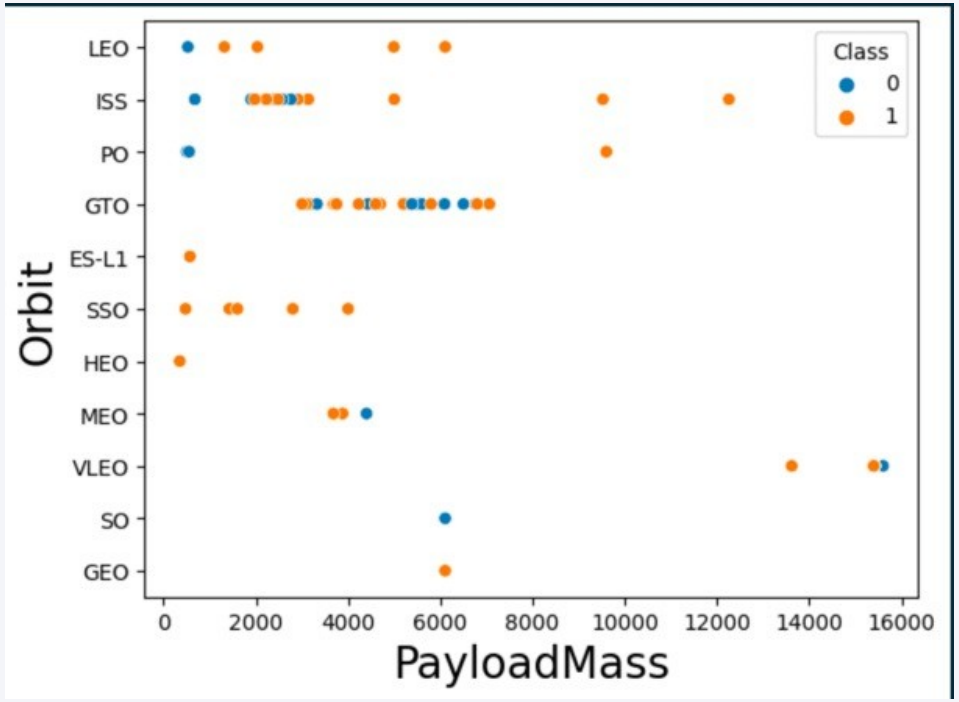


Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type

- With higher payloads. the successful landing or positive landing rate are more for orbit PO, LEO,and ISS.

For orbit GTO, we cannot distinguish this well as both positive landing rate and unsuccessful landing occurred randomly when payload mass is increased.

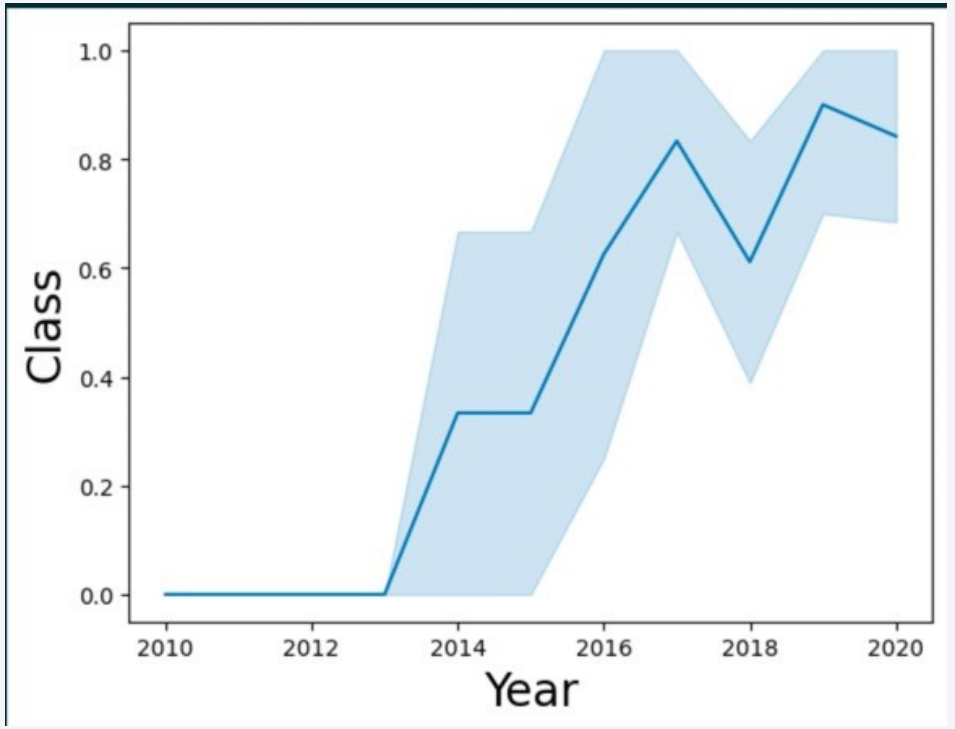


Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- The successful launch rate since 2013 kept increasing till 2017.

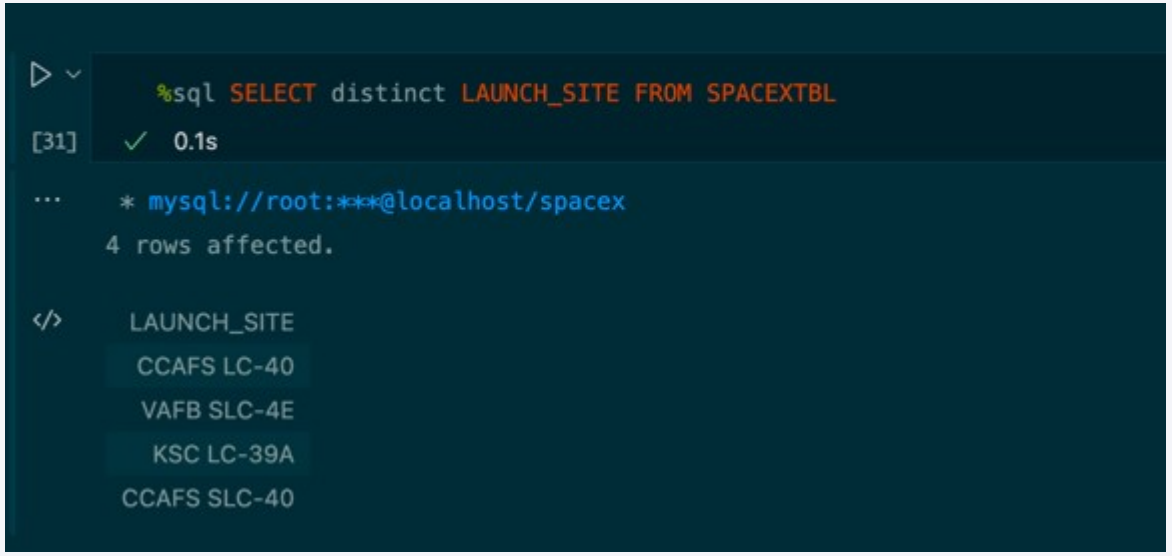
- The successful launch rate is up and down during 2018 and 2020.



All Launch Site Names

Find the names of the unique launch sites

There are four unique launch sites which are CCAFSLC-40, VAFB SLC-4E, KSC LC-39A and CCAFS SLC-40



Launch Site Names Begin with 'CCA'

Find 5 records where launch sites begin with `CCA`

All 5 records listed are from the Launch sit CCAFSLC-40

```
%sql SELECT * FROM SPACEXTBL where LAUNCH_SITE Like '%CCA%' limit 5
```

[43] ✓ 0.2s

```
* mysql://root:***@localhost/spacex
5 rows affected.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAY
4/6/10	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	
8/12/10	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	
22-05-2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	
8/10/12	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	
1/3/13	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	

Total Payload Mass

Calculate the total payload carried by boosters from NASA

- The total payload mass is 45596 KG

```
%sql select sum(PAYLOAD_MASS__KG_) as total1 from SPACEXTBL where CUSTOMER='NASA (CRS)'
[94]  ✓ 0.6s
...   * mysql://root:***@localhost/spacex
      1 rows affected.
</>     total1
        45596
```

Average Payload Mass by F9 v1.1

Calculate the average payload mass carried by booster version F9 v1.1

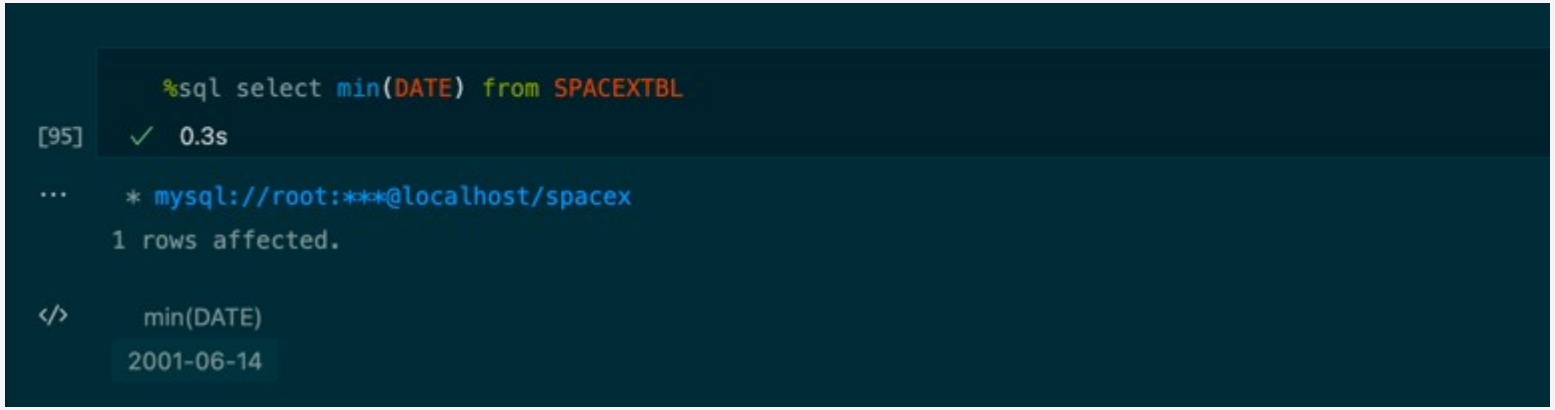
The average payload mass carried by booster version F9 v1.1 is 14642 KG

```
%sql select sum(PAYLOAD_MASS__KG_) as total2 from SPACEXTBL where BOOSTER_VERSION='F9 v1.1'
[45]  ✓ 0.6s
...   * mysql://root:***@localhost/spacex
      1 rows affected.
</>   total2
      14642
```

First Successful Ground Landing Date

Find the dates of the first successful landing outcome on ground pad.

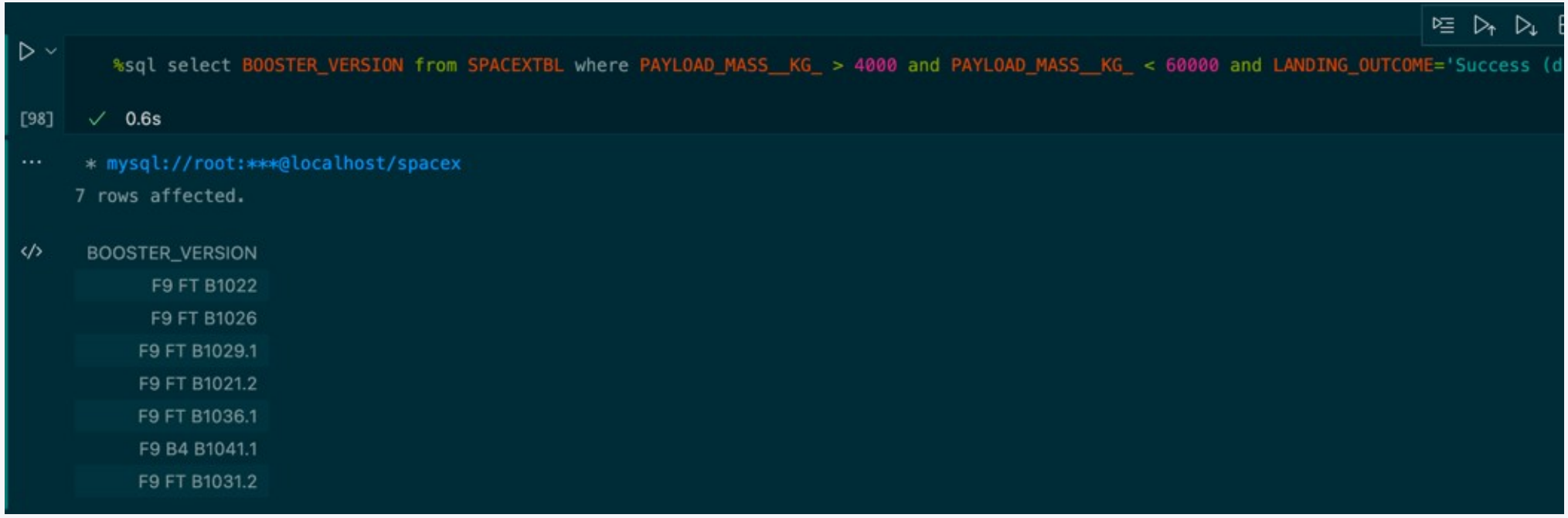
2001-06-14 is the date of the first successful landing.



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

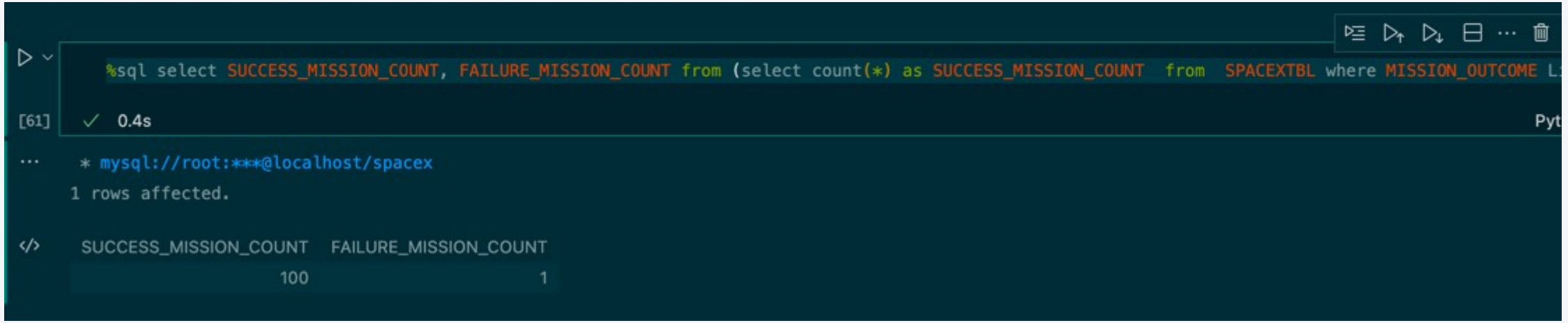
The higher the flight are 6 booster version for the conditions above



Total Number of Successful and Failure Mission Outcomes

Calculate the total number of successful and failure mission outcomes

There are 100 Successful mission and 1 failure mission



Boosters Carried Maximum Payload

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

There are 12 Booster Version carried the same maximum payload mass

```
%sql select * from (select Booster_Version, MAX(PAYLOAD_MASS__KG_) AS MAX_PAYLOAD_MASS  FROM SPACEXTBL GROUP BY Booster_Version  ORDER BY MAX_PAYLO
```

[104] ✓ 0.8s Python

```
 * mysql://root:***@localhost/spacex
12 rows affected.
```

Booster_Version	MAX_PAYLOAD_MASS
F9 B5 B1048.4	15600
F9 B5 B1049.7	15600
F9 B5 B1048.5	15600
F9 B5 B1060.3	15600
F9 B5 B1051.6	15600
F9 B5 B1058.3	15600
F9 B5 B1051.4	15600
F9 B5 B1056.4	15600
F9 B5 B1060.2	15600
F9 B5 B1051.3	15600
F9 B5 B1049.5	15600
F9 B5 B1049.4	15600

2015 Launch Records

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

There are 1 failed landing_outcomes in drone ship in year 2015

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

There are 3 Failed launches of drone ship and 2 successful ground pad

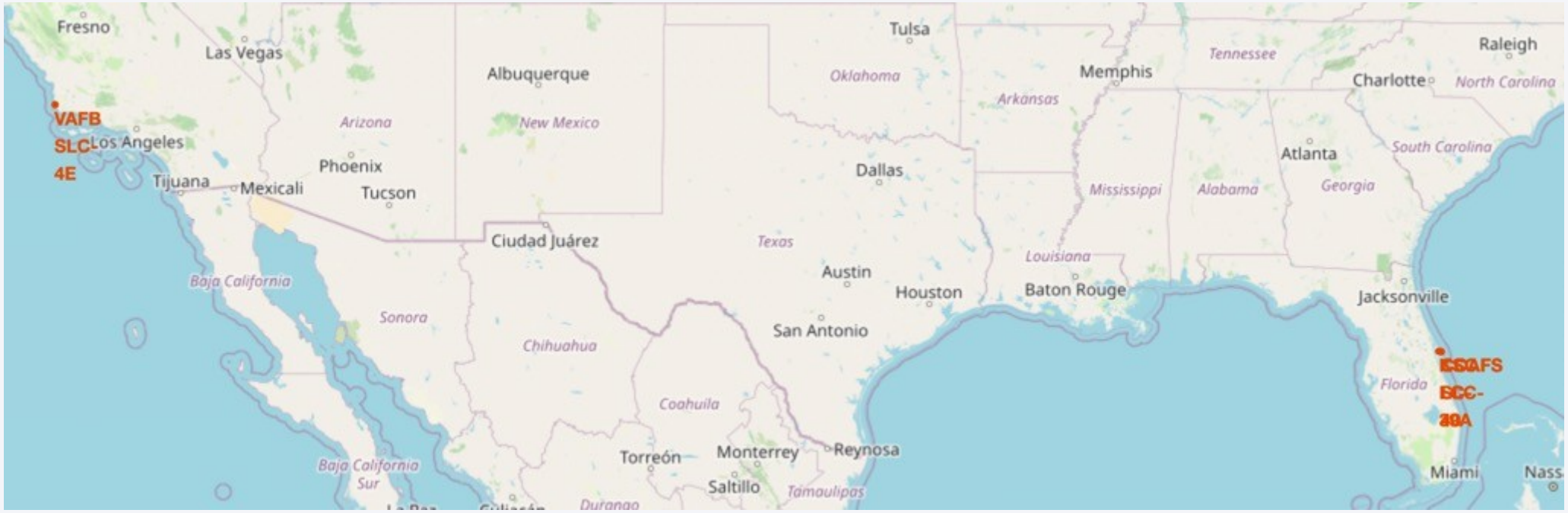
```
    as count from SPACEXTBL  where (LANDING_OUTCOME='Failure (drone ship)' or LANDING_OUTCOME='Success (ground pad)') and Date BETWEEN '2010-06-04' AND '
[93] ✓ 0.4s                                                                                                                                    Python
 * mysql://root:***@localhost/spacex
2 rows affected.
```

LANDING_OUTCOME	count
Failure (drone ship)	3
Success (ground pad)	2


Section 3
Launch Sites
Proximities Analysis

Launch Sites Location Analysis

All launch site's location.

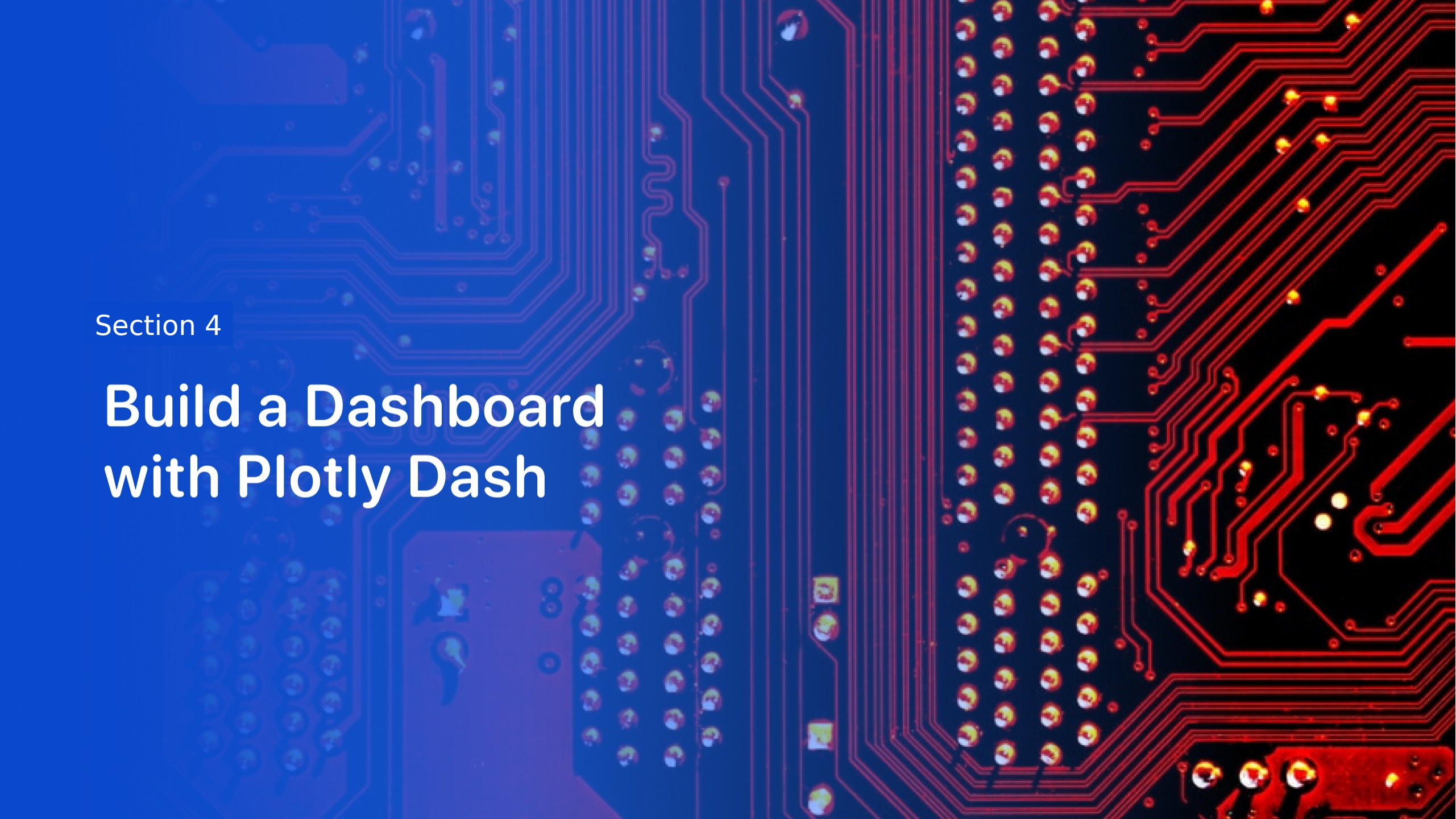


Success and Failure Launch Statistics on Launch Site

Launch site location with total count, success and failure launch count

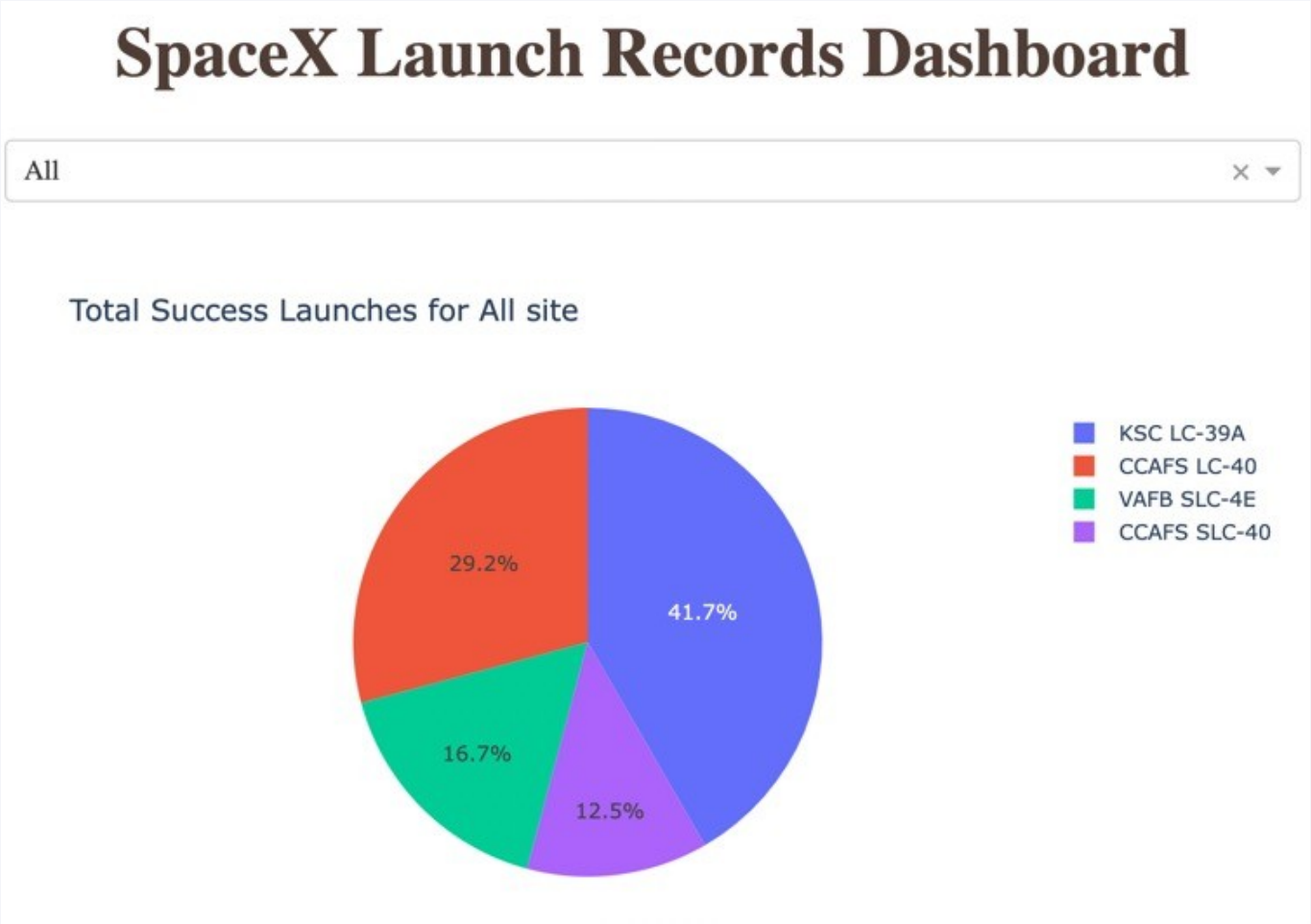
Launch site and its proximities

The distance from CCAFS SLC-40 Launch site to coastline is 0.84 KM


Section 4
Build a Dashboard with Plotly Dash

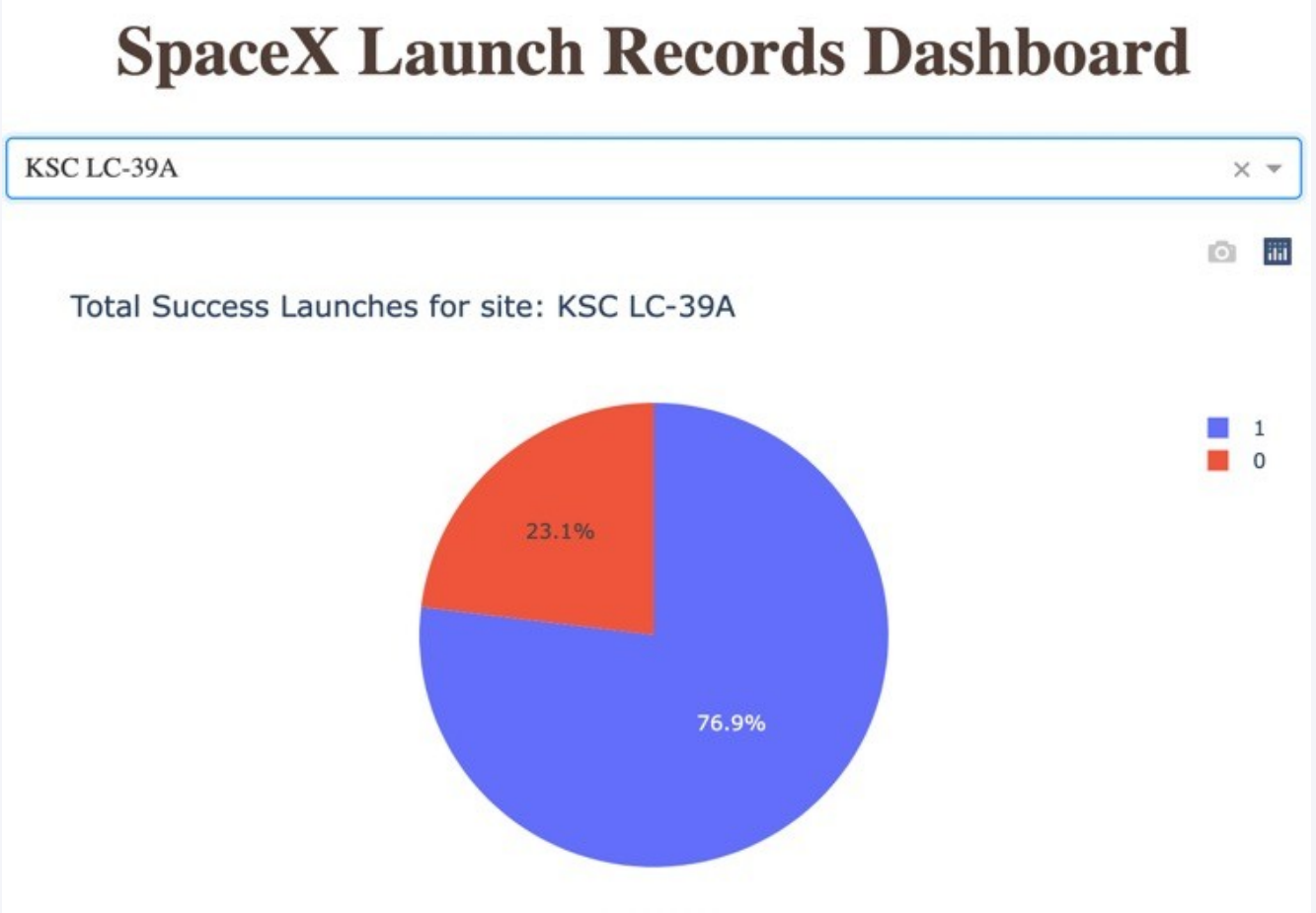
Success Launches for All SpaceX Launch Sites

- The screenshot of launch success count for all sites in a pie chart.
- Explanation of the important elements and findings on the screenshot
  - #Launch site KSC LC-39A has the highest number of successful launch
  - #Launch site CCAFS SLC-40 has the lowest number of successful launch



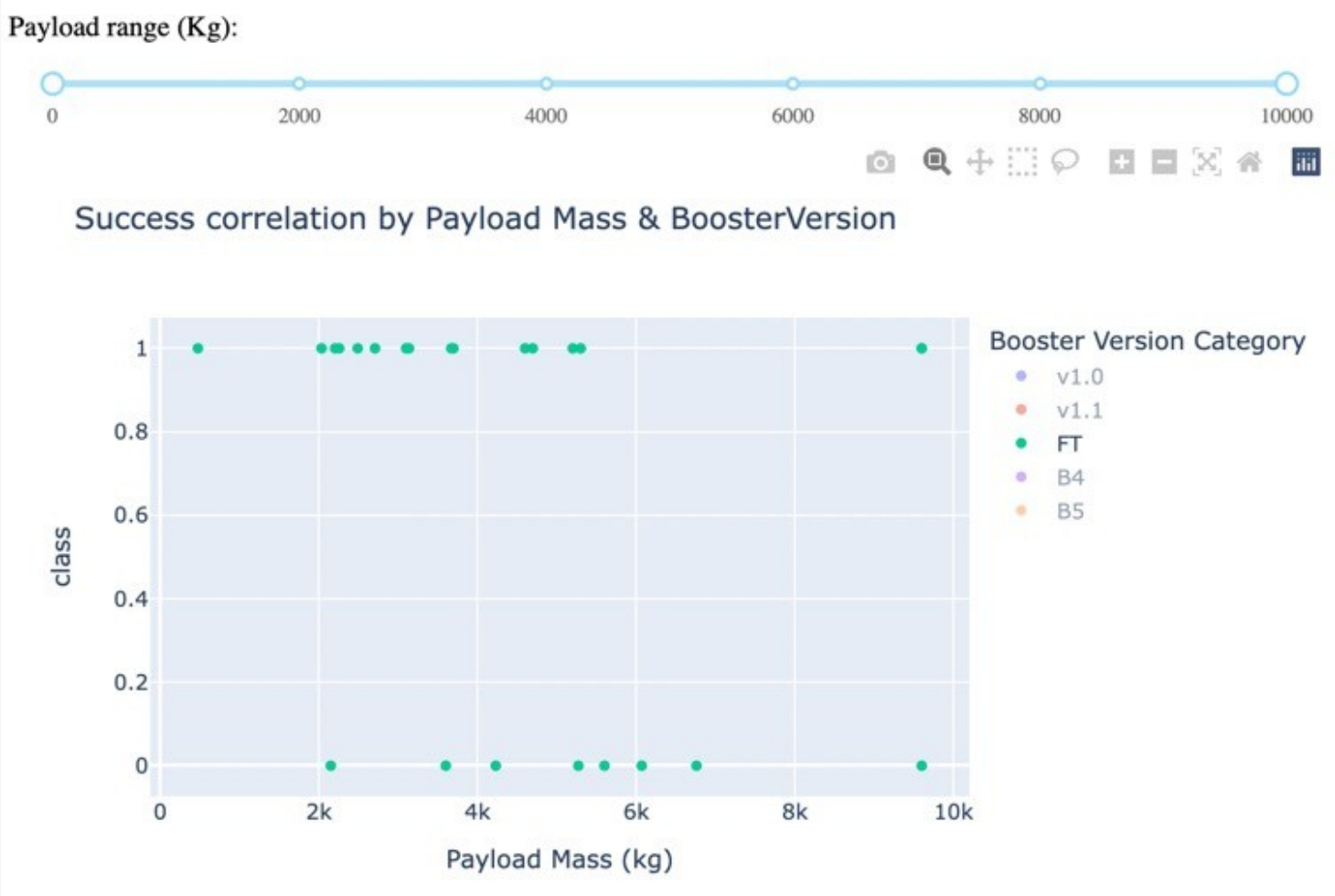
Highest Launch Success Ratio

- The screenshot of the highest launch success ratio (Site KSC LC-39A) in a pie chart.
    - Launch site KSC LC-39A has 76.9% of successful launch and 23.1% for failure



Payload vs. Launch Outcome Correlation

- Explanation of the important elements and findings on the screenshot
- The booster version FT has the largest success rate



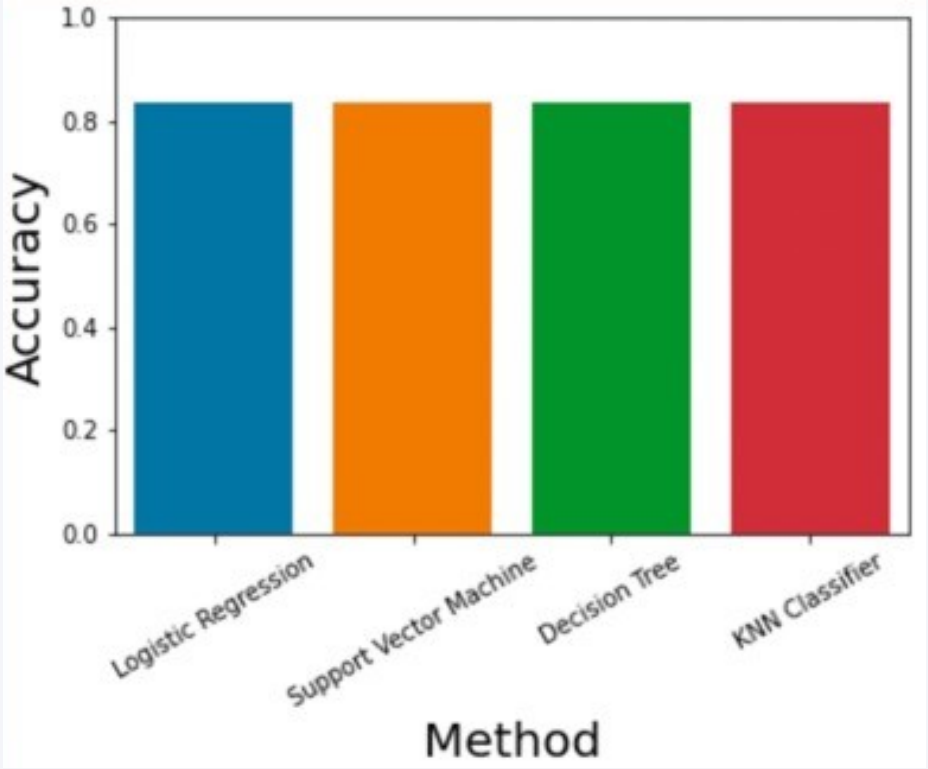
Section 5

Predictive Analysis (Classification)

Classification Accuracy

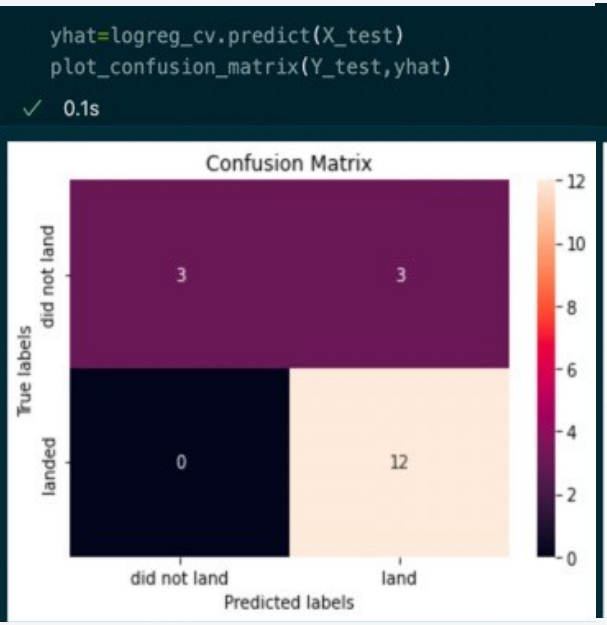
- Visualize the built model accuracy for all built classification models, in a bar chart

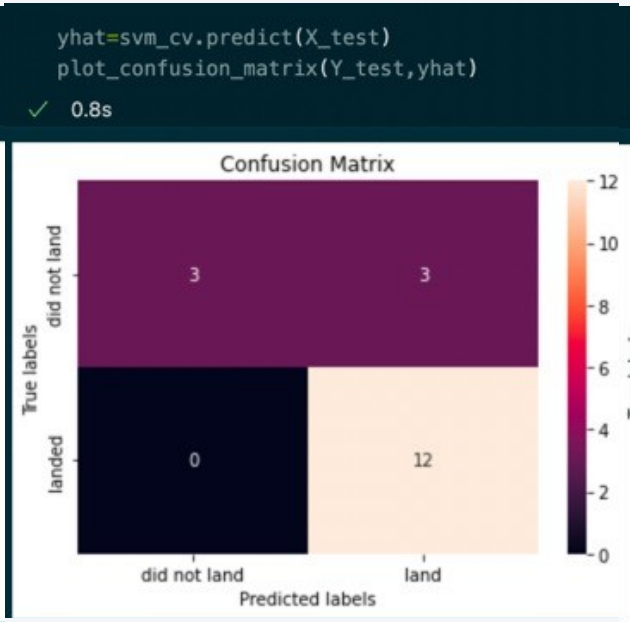
- All models have the same accuracy

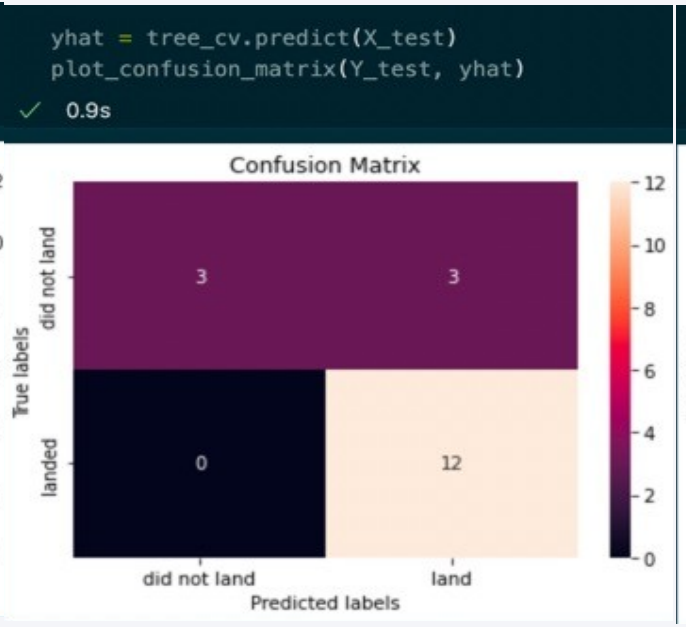


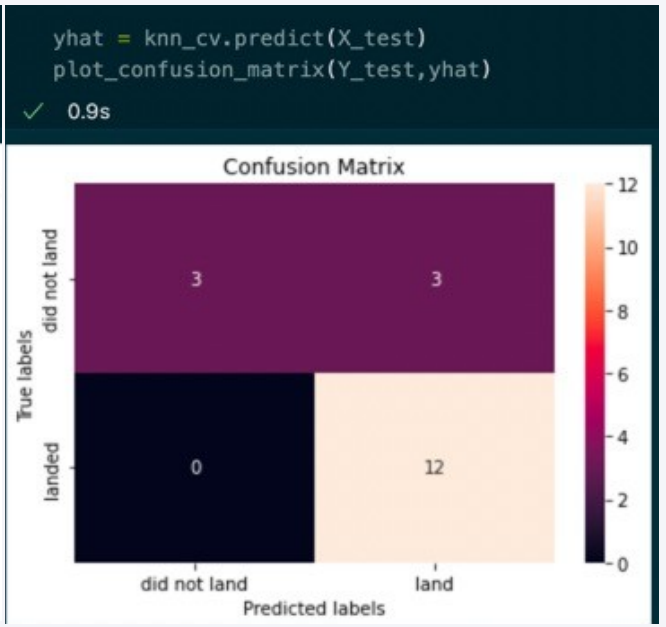
Confusion Matrix

- All models have the same accuracy and the Confusion Matrix elements are the same









Conclusions

- The prediction accuracy is the same for all models probably because the test data size are small is very small.

- The decision tree is the most suitable algorithm for this task.

- KSC LC-39A records the most site for successful launches.

- Heavier payloads may play a role in failed launches.

- GEO, SSO, ES-L1, HEO Orbits are the most successful orbits.

Appendix

- dataset_part_1.csv
- dataset_part_2.csv
- dataset_part_3.csv
- spacex_launch_dash.csv
- spacex_launch_geo.csv

Thank you!