IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- Summary of methodologies.
  - Data Collection
  - Data Wrangling
  - Exploratory Data Analysis (EDA) Using Data Visualization
  - Exploratory Data Analysis (EDA) Using SQL
  - Building an interactive Map using Folium
  - Building a Dashboard using Plotly
  - Performing Predictive Analysis
- Summary of all results
  - Exploratory Data Analysis Results
  - Interactive Analysis in Screen Shots
  - Predictive Analysis Results

- Project background and context
  - SpaceX is leading company in commercial Space Industry. SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each. If we can predict a successful landing of first stage , we can determine the cost of the lunch

- Problems you want to find answers
- Find the effects of variables such Launch Site, Payload Mass, Number of Flights , Orbits on the first stage successful landing

Introduction

Project background and context

SpaceX is leading company in commercial Space Industry. SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each. If we can predict a successful landing of first stage , we can determine the cost of the lunch

Problems you want to find answers

Find the effects of variables such Launch Site, Payload Mass, Number of Flights , Orbits on the first stage successful landing

Section 1

Methodology

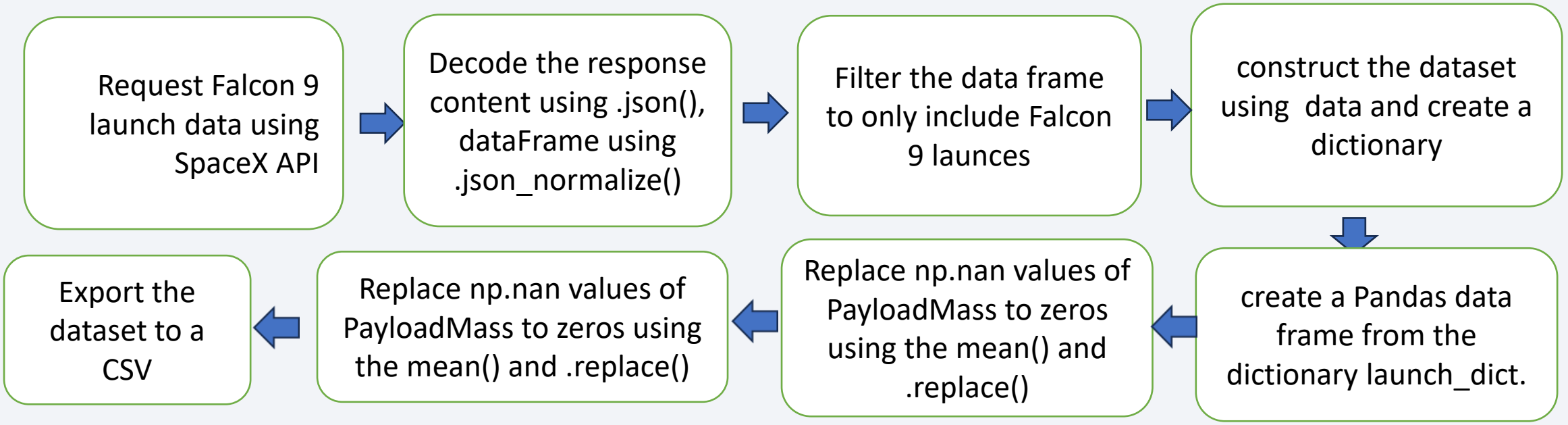
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX rest API
  - Data scraping from the web
- Perform data wrangling
  - Perform exploratory data analysis (EDA) using visualization and SQL
  - Perform interactive visual analytics using Folium and Plotly Dash
  - Perform predictive analysis using classification models
  - Use data analysis tools to load a dataset, clean it, and find out interesting insights from it

Data Collection

- Describe how data sets were collected.
  - Using Rest API to collect the data set
  - Using web scraping Wikipedia entry to collect Falcon 9 launch data
- Presentation of data collection process using key phrases and flowcharts

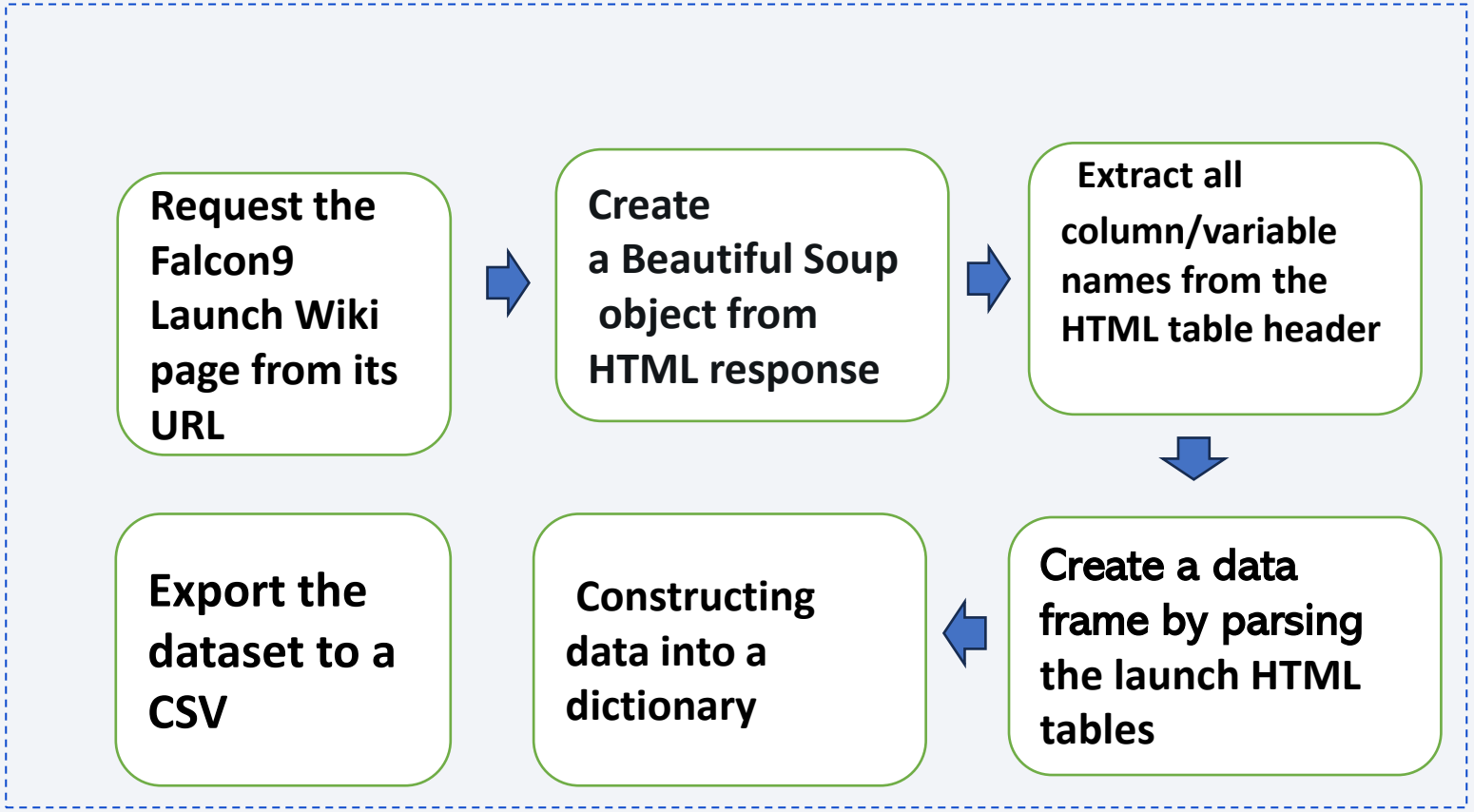


Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

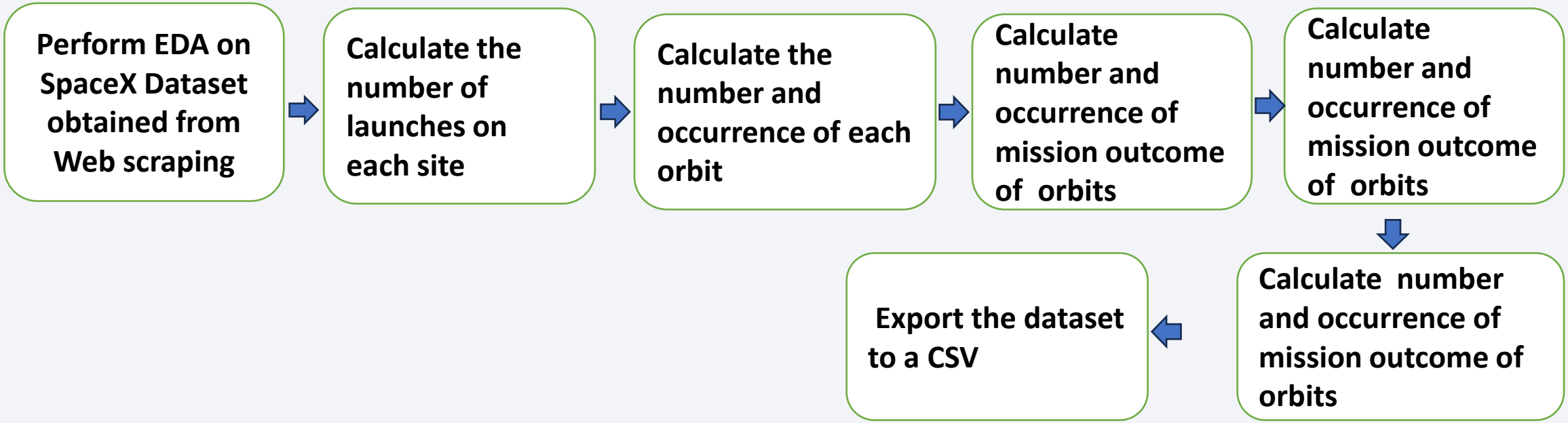
- https://github.com/sahmed1956/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb



flowchart of web scraping

Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts



- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose
- https://github.com/sahmed1956/IBM-Applied-Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling%20(1).ipynb

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
  - Scatter plots to show the relationships between variables such as : Fight Number vs. Payload Mass, LaunchSite vs. Flight Number, LaunchSite vs. Payload Mass, Flight Number vs. Orbit, Bar plot for displaying Success Rate vs. Orbit .
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose
  - https://github.com/sahmed1956/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- Using bullet point format, summarize the SQL queries you 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 achieved.**
  - **List the names of the boosters which have success in drone ship and have payload mass > 4000 but < 6000.**

EDA with SQL, continues

- **List the total number of successful and failure mission outcomes**
- List the names of the booster_versions which have carried the maximum payload mass.
- **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.**

EDA with SQL, continues

Using bullet point format, summarize the SQL queries you performed

List the total number of successful and failure mission outcomes

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

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.

- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose
- https://github.com/sahmed1956/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
  - Circles, Popup labels, text labels of NASA Space Center and all other Launch Sites were created and added

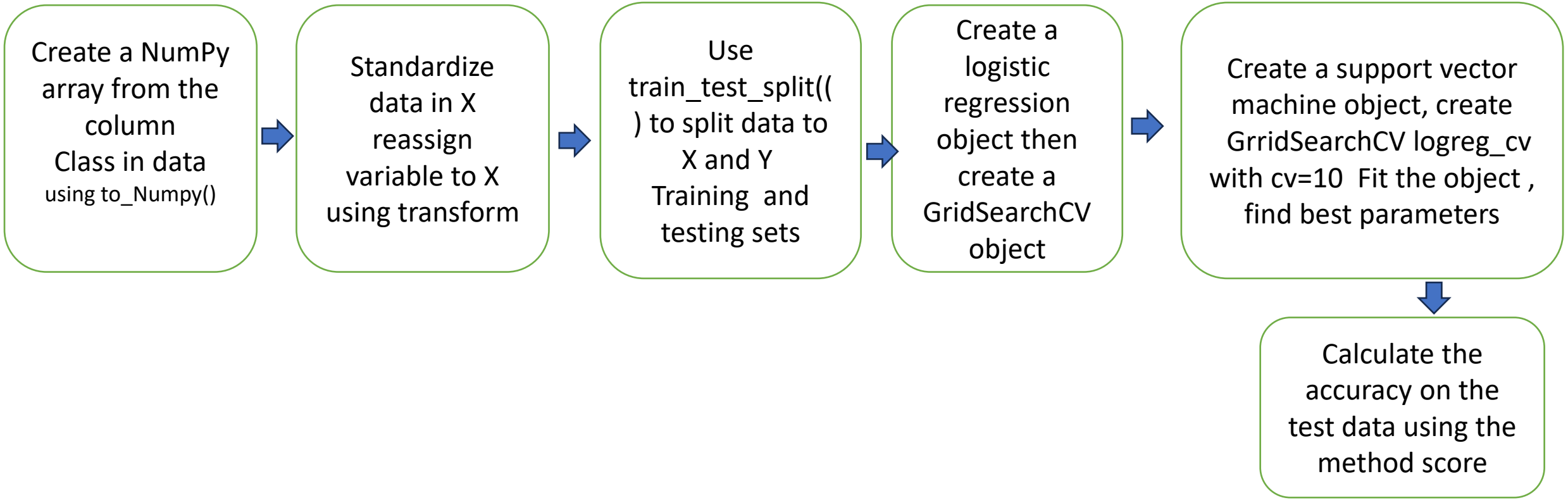
- Explain why you added those objects
  - To take Measurements to use in Machine Learning analysis and predictions

- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose
- https://github.com/sahmed1956/IBM-Applied-Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
  - Pie chart Plot showing Success rate of launch sites.
  - Range slider to select Payload Mass.
  - Scatter Plot to measure Success rate vs. Pay load Mass.
  - 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
- https://github.com/sahmed1956/IBM-Applied-Data-Science-Capstone/blob/main/spacex_dash_app%20(1).py

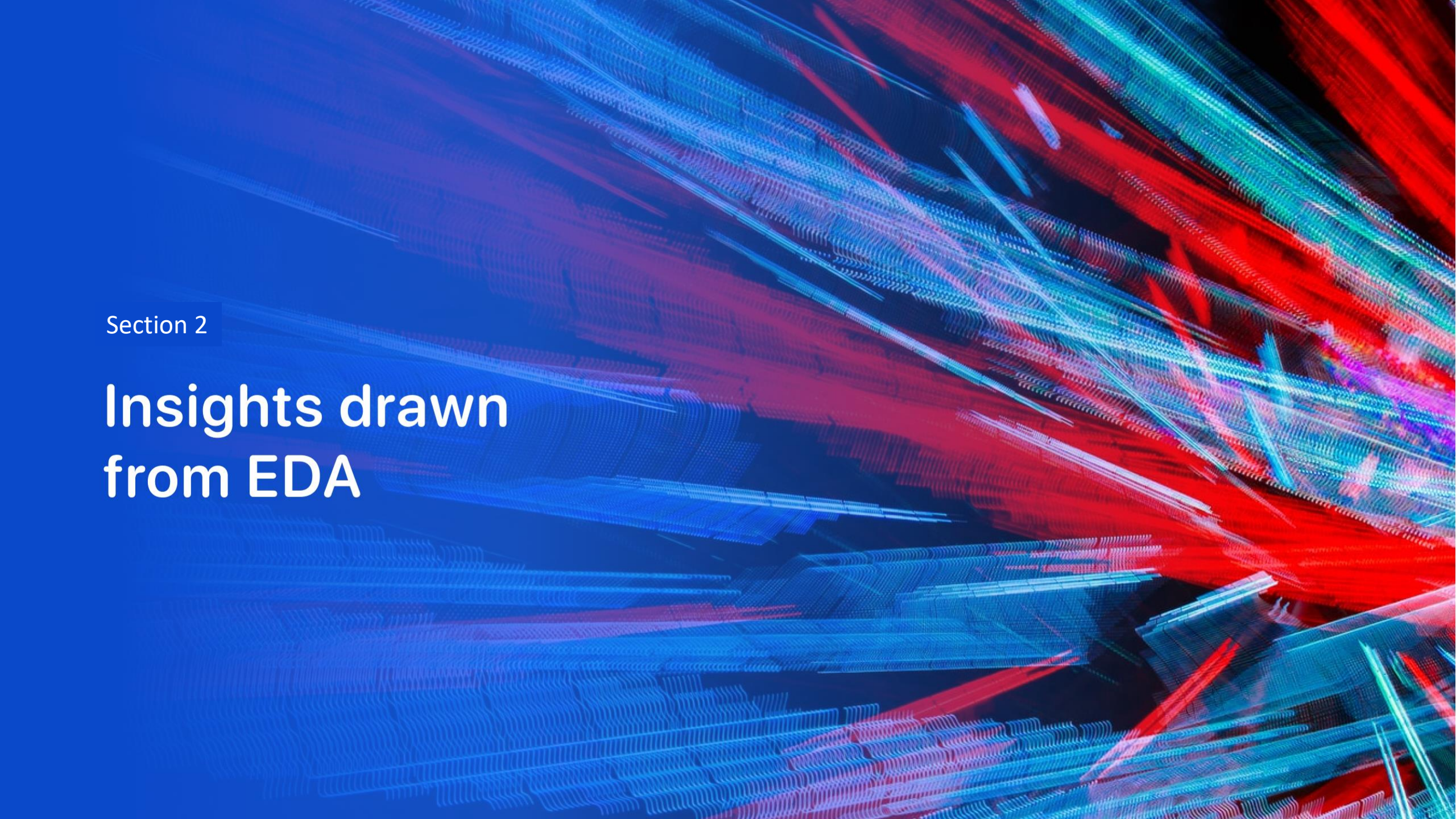
classification model



- 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

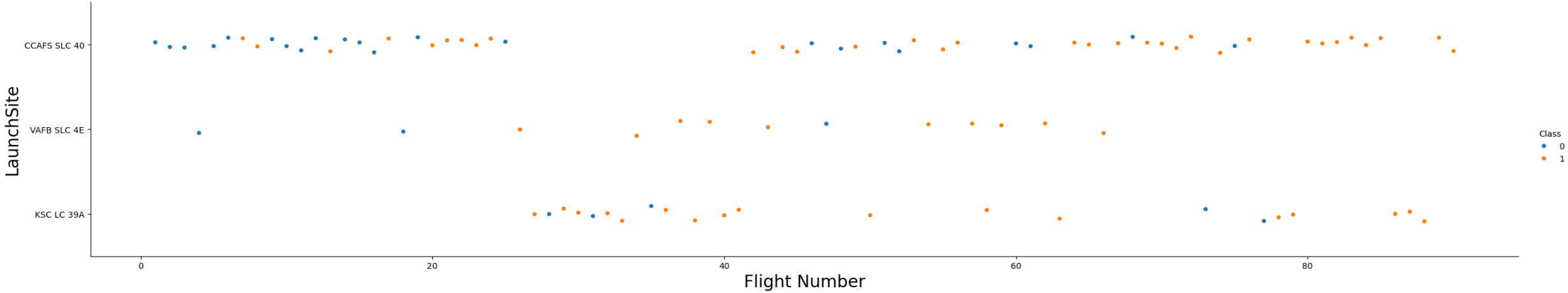
Results

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

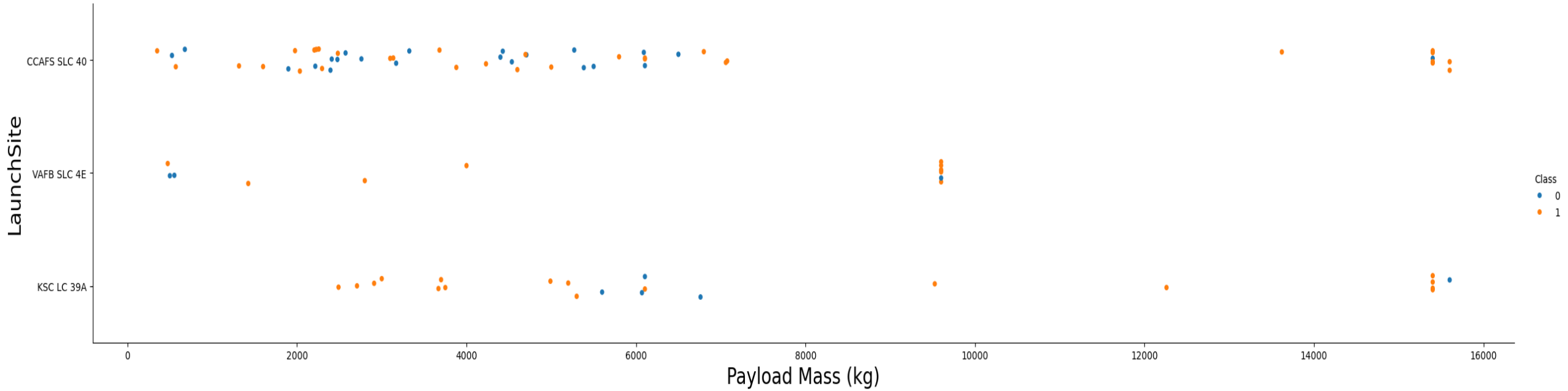

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

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



- Launch Site Number CCAFS LC-40 is the moat used and the most successful.
- Site Number CCAFS SLC-40 is second in success and followed by KSC LC-39A.
- Success rate improved over time.



Payload vs. Launch Site

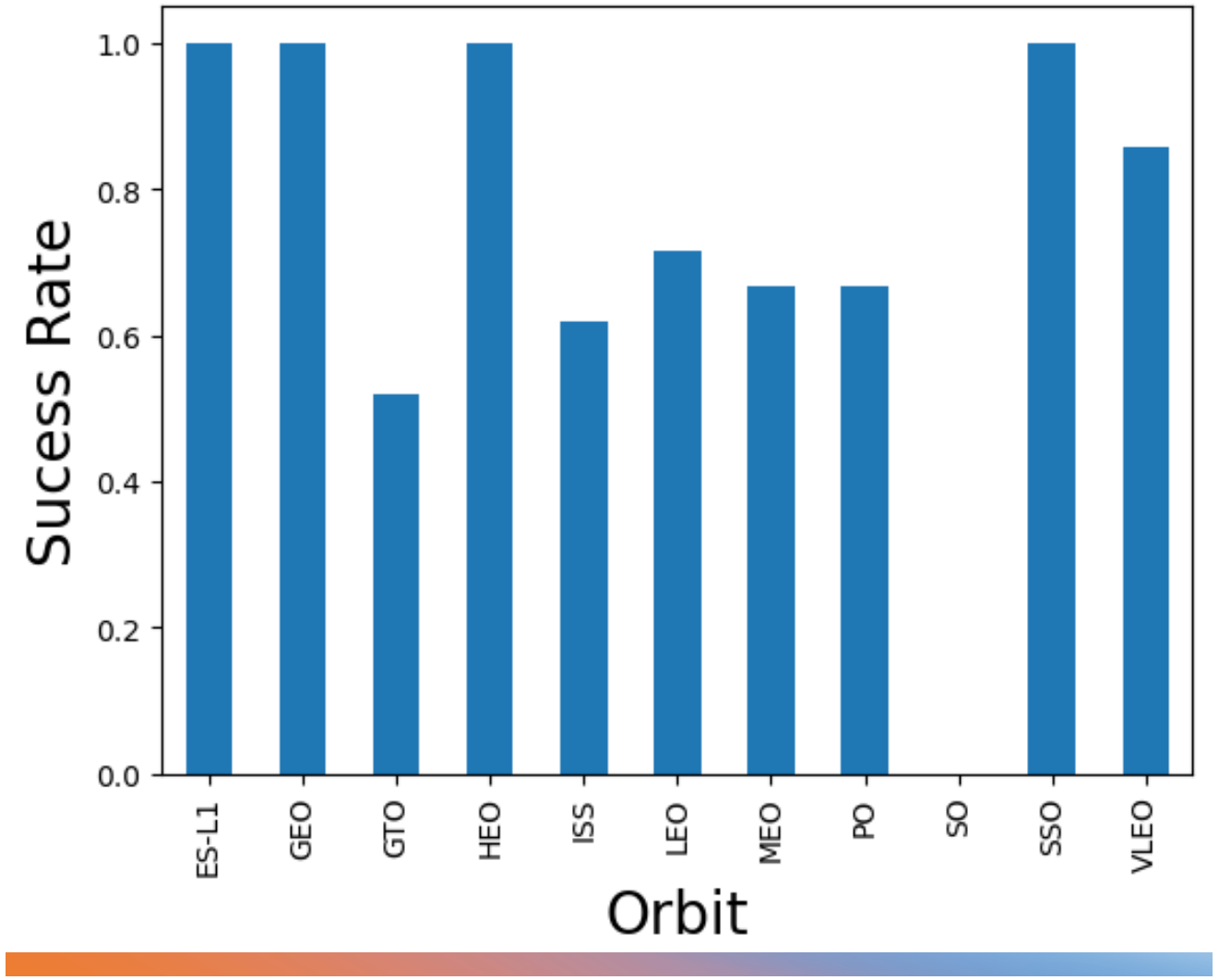
- Show a scatter plot of Payload vs. Launch Site
- Show the screenshot of the scatter plot with explanations
  - All sites enjoy higher success rate when pay load mass is high
  - Site CCAFS LC-40 can hand pay loads over 12000 Kg.

Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type
- Show the screenshot of the scatter plot with explanations
- Highest success are in Orbits:

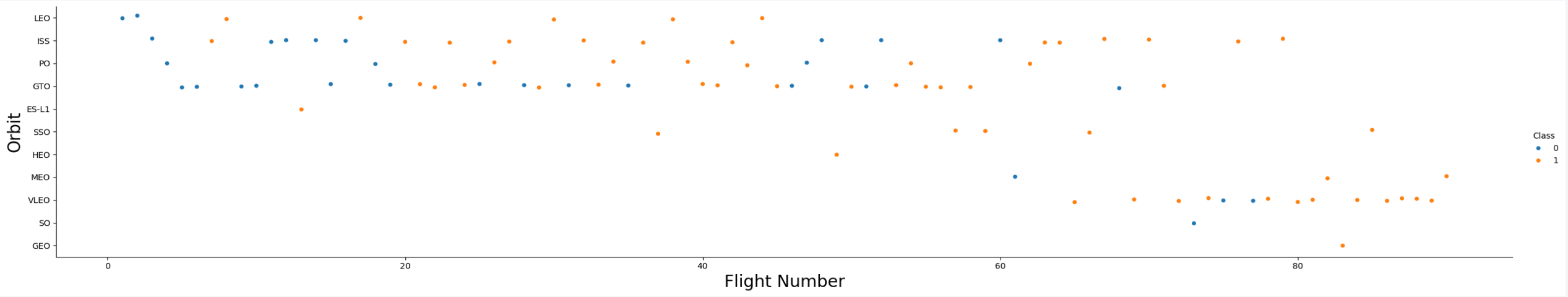
  ES-L1, GEO, HEO, SSO.

  FOLLOWED BY LEO, MEO, PO, ISS, GTO



Flight Number vs. Orbit Type

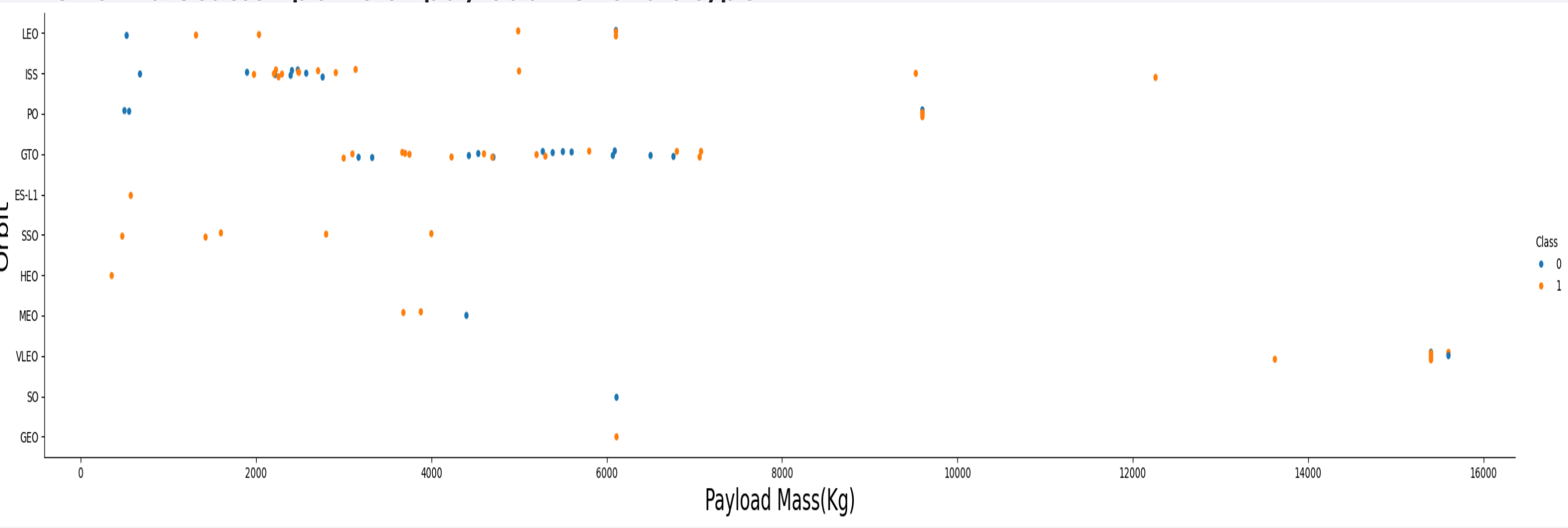
- Show a scatter point of Flight numbers vs. Orbit type



- Show the screenshot of the scatter plot with explanations
  - Success rate increases with increase of Flight numbers

Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type



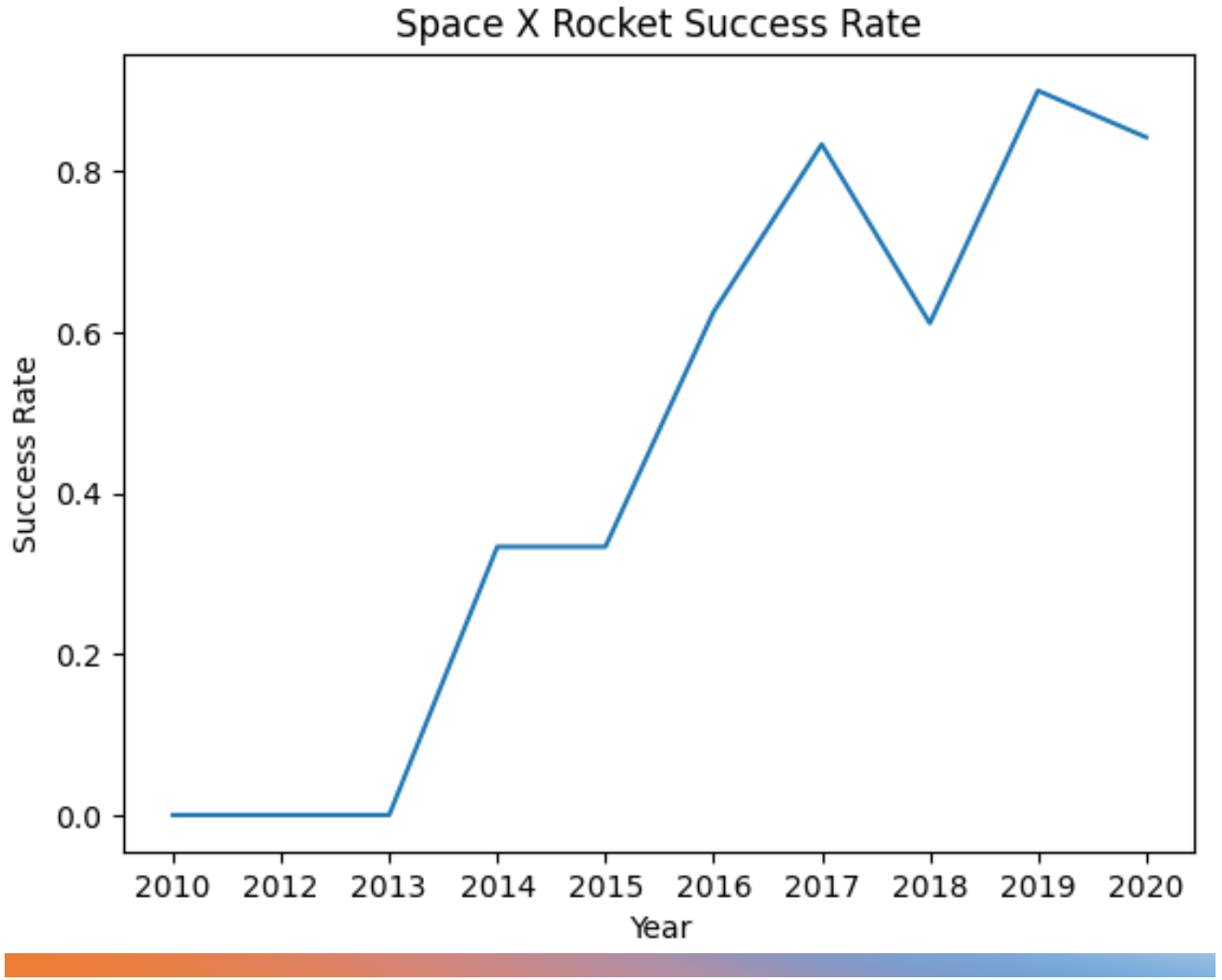
- ISS has a good rate of success with increasing payload
- GTO success rate is around 50 percent.

Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations

  - The success rate keep increasing over the years



All Launch Site Names

- Find the names of the unique launch sites

- Present your query result with a short explanation here
- %sql SELECT DISTINCT(LAUNCH_SITE)\ FROM SPACEXTBL
- **This query names of the unique launch sites in the space mission**

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`

- Present your query result with a short explanation here
- %sql SELECT LAUNCH_SITE from SPACEXTBL where (LAUNCH_SITE) LIKE 'CCA%' LIMIT 5;
- This query display **5 records where launch sites begin with the string 'CCA'**

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

TOTAL_PAYLOAD_MASS
45596

- Present your query result with a short explanation here
- %sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD_MASS FROM SPACEXTBL \
- WHERE CUSTOMER = 'NASA (CRS)';
- This query displays total payload mass carried

Average Payload Mass by F9 v1.1

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

Average **Payload mass**

6138.287128712871

- Present your query result with a short explanation here
- %sql select avg(PAYLOAD_MASS__KG_) as payload Mass from SPACEXTBL;

First Successful Ground Landing Date

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

- Present your query result with a short explanation here
- %sql select min(DATE) from SPACEXTBL;

min(DATE)
2010-06-04

%sq
%sql SELECT BOOSTER_VERSION FROM SPACEXTBL \ WHERE (Landing_Outcome = 'Success (drone ship)') AND (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000);

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

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

- Present your query result with a short explanation here

%sql SELECT BOOSTER_VERSION FROM SPACEXTBL \ WHERE (Landing_Outcome = 'Success (drone ship)') \
AND (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000);

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

missionoutcomes
1
98
1
1

- Present your query result with a short explanation here
- %sql select count(MISSION_OUTCOME) as missionoutcomes from SPACEXTBL\ GROUP BY MISSION_OUTCOME;

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- resent your query result with a short explanation here
- %sql select BOOSTER_VERSION as booster version from SPACEXTBL \
- where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from SPACEXTBL)

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

month	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- Present your query result with a short explanation here

```
%sql SELECT substr(Date, 6,2) as month, substr(Date,0,5) as Date, BOOSTER_VERSION, LAUNCH_SITE, Landing_Outcome\ FROM SPACEXTBL WHERE
(LANDING_OUTCOME = 'Failure (drone ship)') AND (substr(Date,0,5)='2015');
* sqlite:///my_data1.db Done.
Out[137]:
```

Only too failed landing_outcomes in drone ship

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

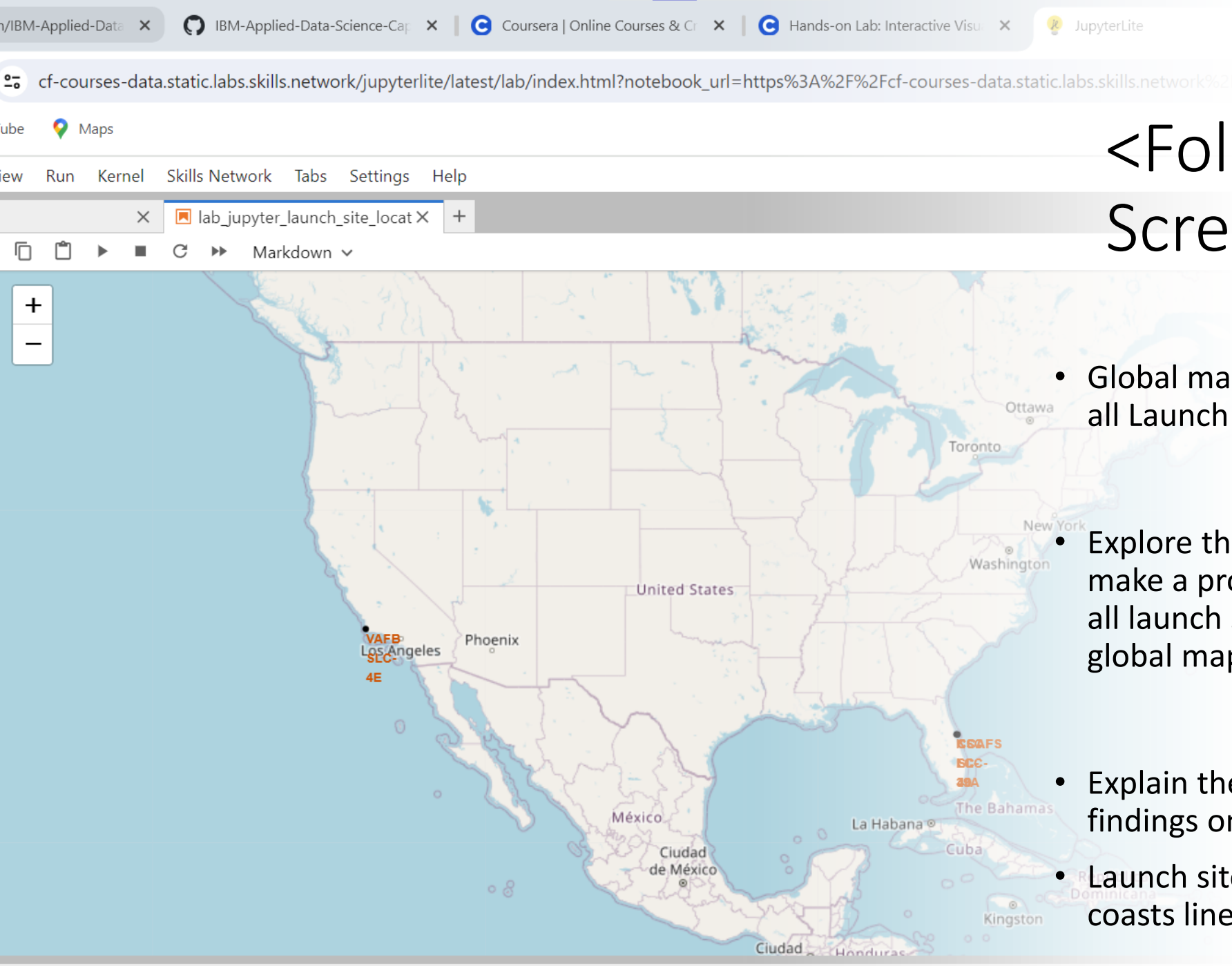
- Present your query result with a short explanation here
- %sql SELECT LANDING_OUTCOME FROM SPACEXTBL\
- WHERE DATE BETWEEN '2010-06-04' \
- AND 2017-03-20' ORDER BY DATE DESC;

Landing Outcome
No attempt
Success (ground pad)
Success (drone ship)
Success (drone ship)
Success (ground pad)
Failure (drone ship)
Success (drone ship)
Success (drone ship)
Success (drone ship)
Failure (drone ship)
Failure (drone ship)
Success (ground pad)
Precluded (drone ship)
No attempt
Failure (drone ship)

Section 3

Launch Sites Proximities Analysis

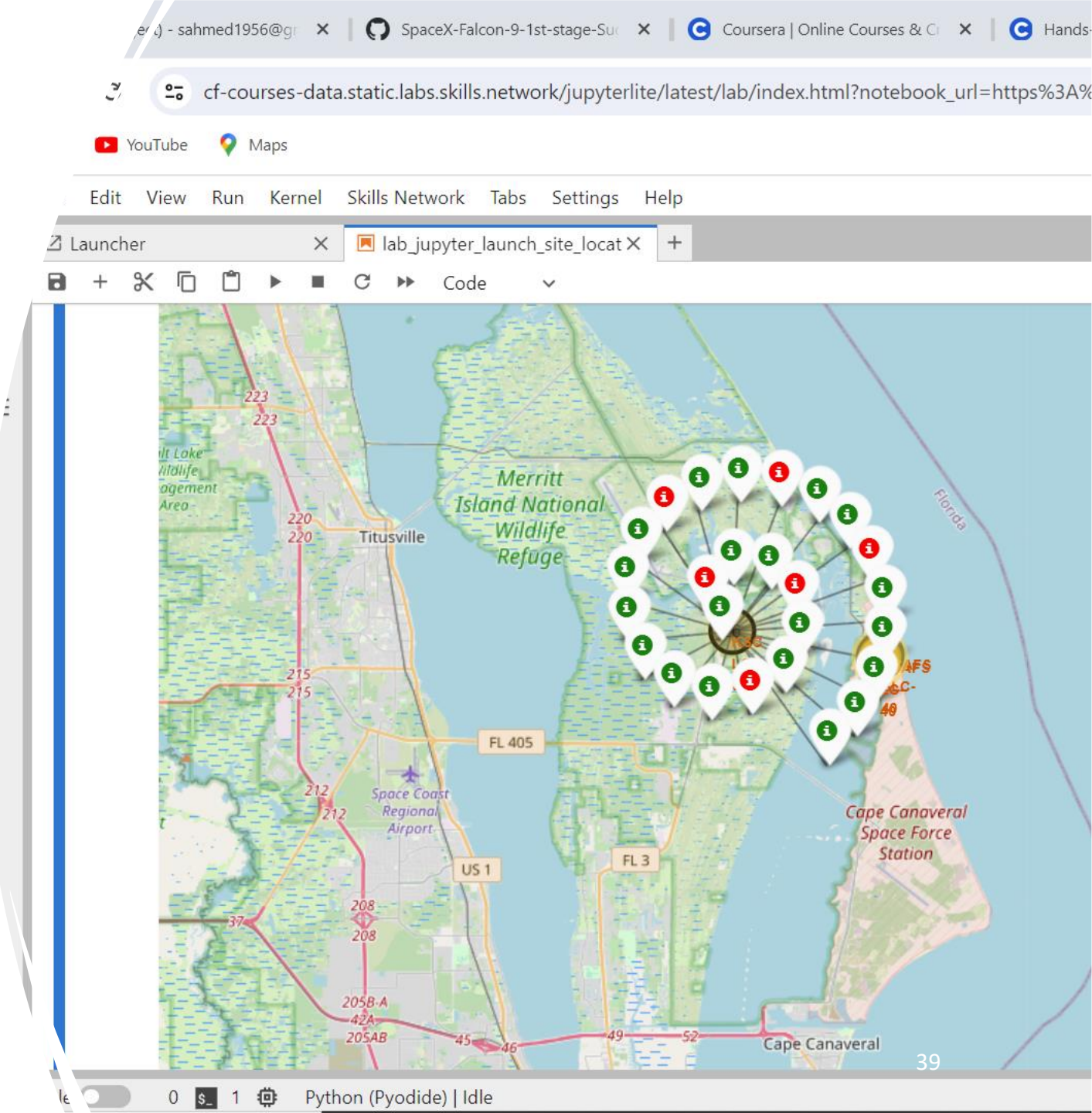
<Folium Map Screenshot 1>



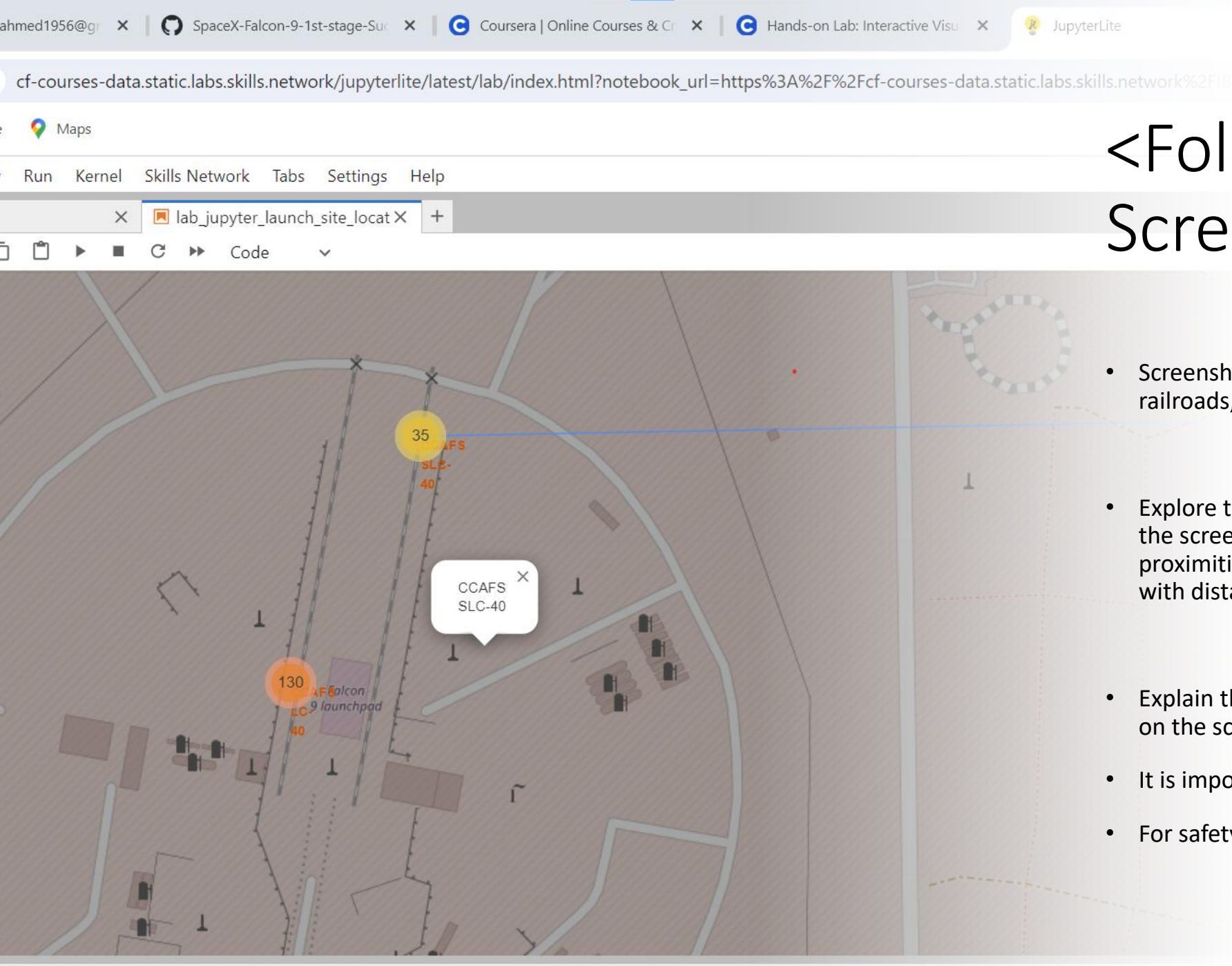
- Global map representing markers for all Launch Sites locations.
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot
- Launch sites are in proximity to the coasts lines, the Equator

<Folium Map Screenshot 2>

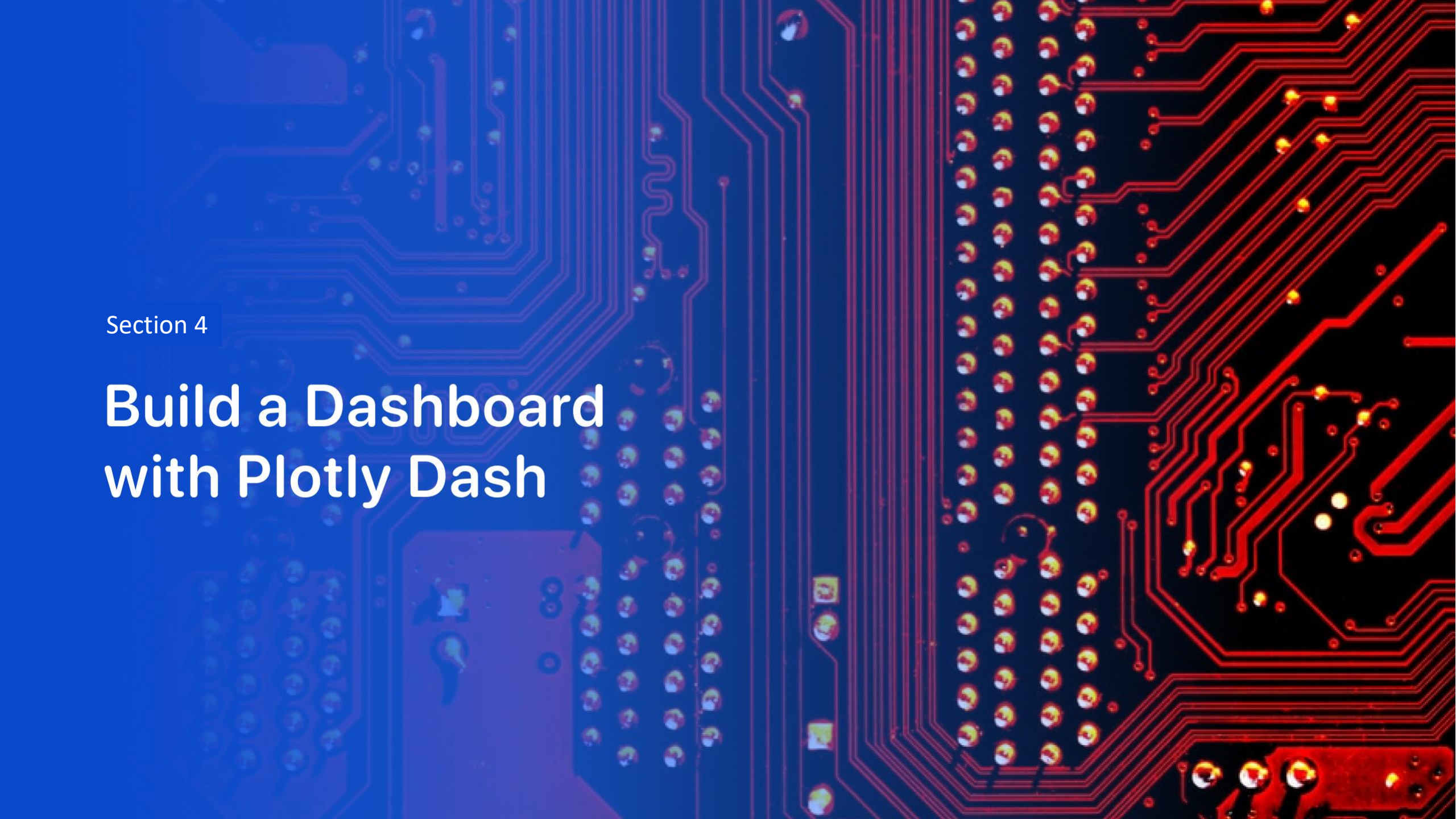
- Color map for launch sites
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Explain the important elements and findings on the screenshot
- launch site explore its proximity railway, highway, coastline.



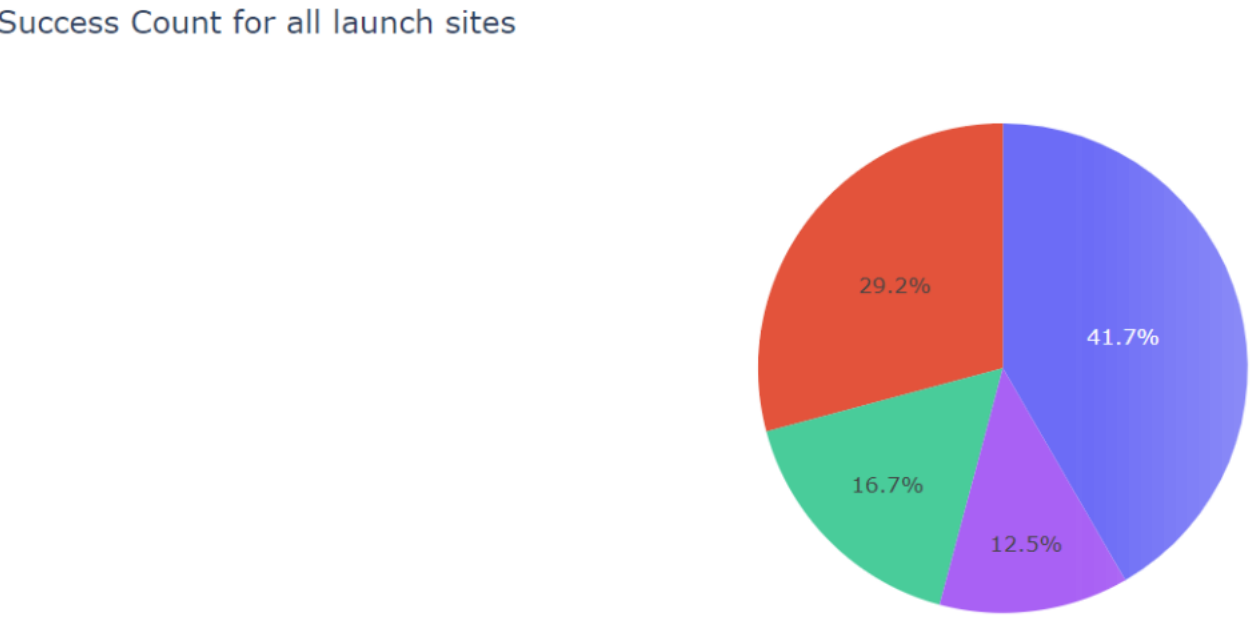
<Folium Map Screenshot 3>



- Screenshot showing Launch Site proximity to railroads, Highways
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed
- Explain the important elements and findings on the screenshot
- It is import to keep a distance from population
- For safety reasons


Section 4
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 2>



- The Launch Site with highest launch success
- Screenshot of a Pie chart for success ratio for all the launch sites
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- 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.

- <Dashboard Screenshot 3>

Section 5

Predictive Analysis (Classification)

Classification Accuracy

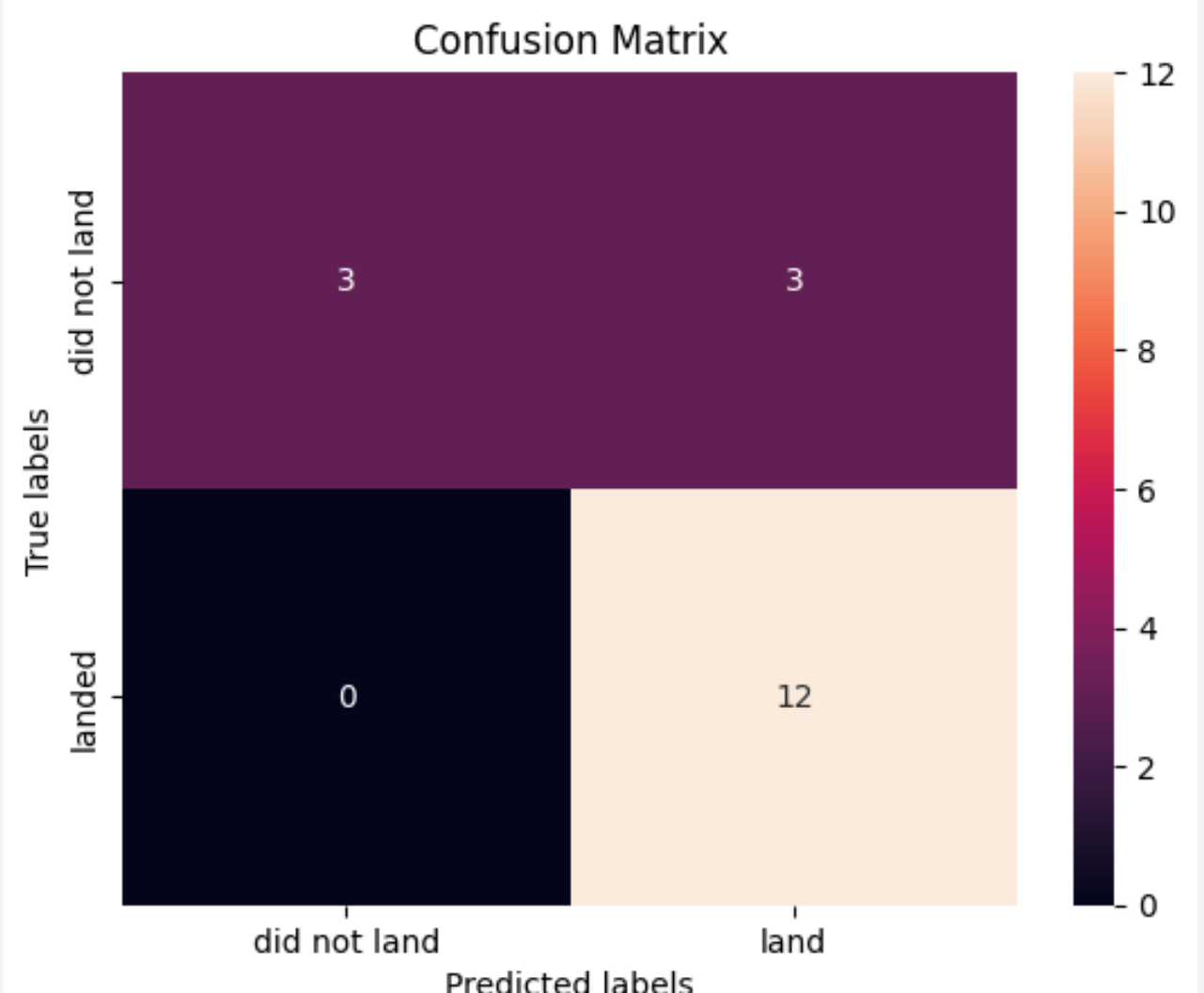
- Visualize the built model accuracy for all built classification models, in a bar chart

- which model has the highest classification accuracy.
- k-Nearest Neighbors (k-NN),

Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.833333
KNN	0.833333

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation
- All confusion matrices are the same
- All models performed almost the same.



Conclusions

- **Launch Site** CCAFS LC-40 success rate of 60% , whereas VAFB SLC-4E, KSC LC-39A, and launch Sites CCAFS SLC-40 success rate 77%
- s numbers increases so the success rate
- No loads over 1000 kg launched at launch site VAFB SLC-4E
- Highest success are in Orbits:

  ES-L1, GEO, HEO, SSO FOLLOWED BY LEO, MEO, PO, ISS, GTO

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

- Thanks to All of you Educators.

Thank you!