IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
12.11.2023

Outline

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

Executive Summary

- Summary of methodologies
  - Exploratory Data Analysis (EDA) with SQL and data visualitaion
  - Data collection via the execution of API's and web scraping
  - Data wrangling
  - Building dashboard via Plotly Dash
  - Use of Machine Learning
- Summary of all results
  - Finding the best features to predict the success of launching via EDA and the use of machine learning algorithms

Introduction

- Project background and context
  - Working
- Problems you want to find answers
  - Main features that have an influence on a successful landing

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Space X API
  - Web scrapping
- Perform data wrangling
  - Using one hot encoding
- 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

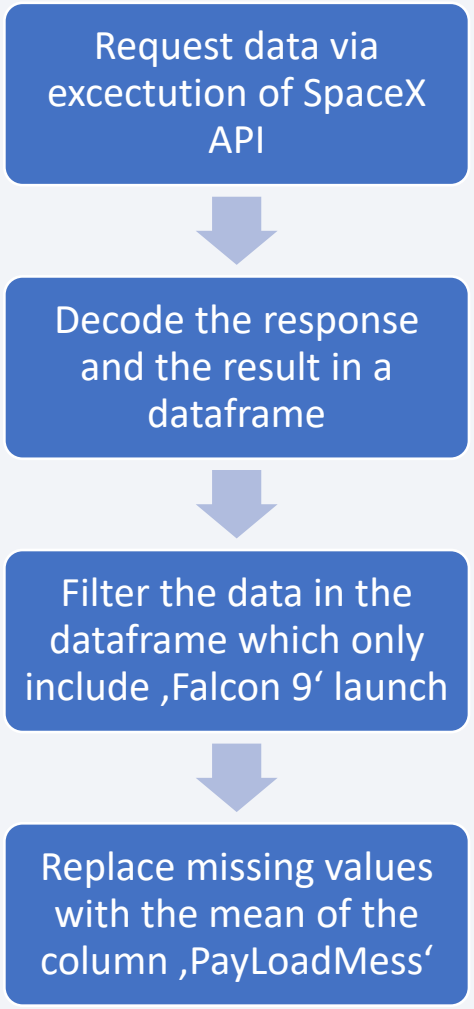
Data Collection

- Describe how data sets were collected.
  - The data sets were generated by running the SpaceX API and further information from Wikipedia through web scraping.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

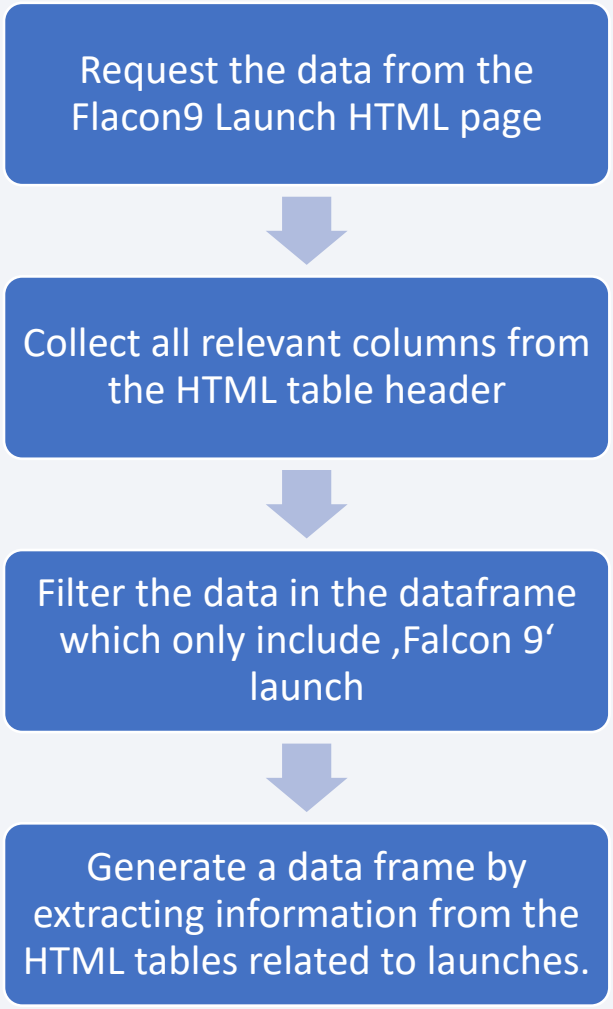
- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose
- https://github.com/enes-13-bot/coursera_ibm_applied_data_science_capstone_enes/blob/8cf6fd70b4978c110ad6192c1ac03ad83e7f3f38/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

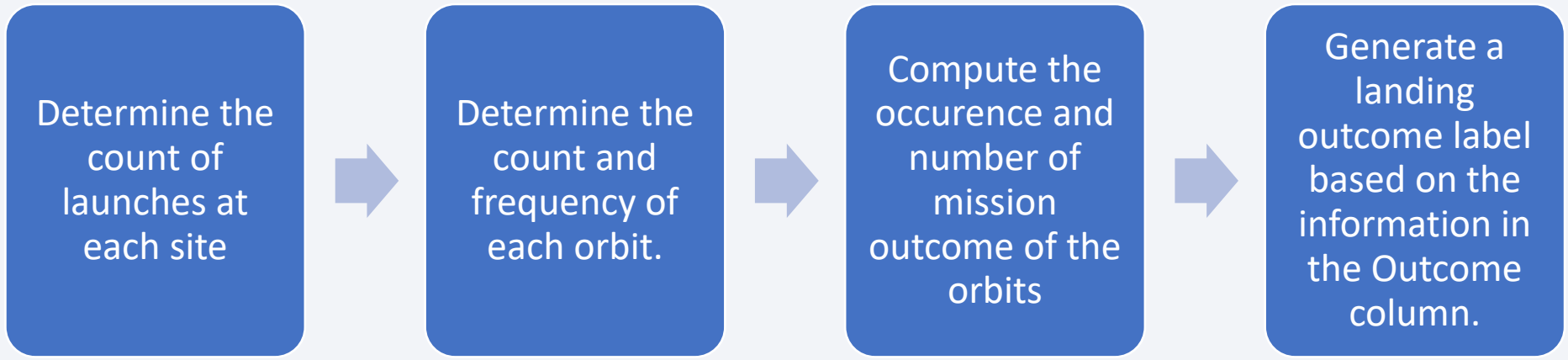
- Present your web scraping process using key phrases and flowcharts

- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose
- https://github.com/enes-13-bot/coursera_ibm_applied_data_science_capstone_enes/blob/8cf6fd70b4978c110ad6192c1ac03ad83e7f3f38/jupyter-labs-webscraping.ipynb



Data Wrangling

- Describe how data were processed



- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose
- https://github.com/enes-13-bot/coursera_ibm_applied_data_science_capstone_enes/blob/8cf6fd70b4978c110ad6192c1ac03ad83e7f3f38/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
  - Scatterplots and barplots were used to display the relationship between different pair of columns/feature
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose
- https://github.com/enes-13-bot/coursera_ibm_applied_data_science_capstone_enes/blob/8cf6fd70b4978c110ad6192c1ac03ad83e7f3f38/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
  - Show 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 succesful 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
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
  - 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/enes-13-bot/coursera_ibm_applied_data_science_capstone_enes/blob/8cf6fd70b4978c110ad6192c1ac03ad83e7f3f38/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
  - Folium maps annotate Launch Sites, both successful and unsuccessful landings, and illustrate proximity to key features such as Railway, Highway, Coast, and City. This visualization aids in comprehending the rationale behind the selection of launch sites and also depicts the spatial distribution of successful landings.

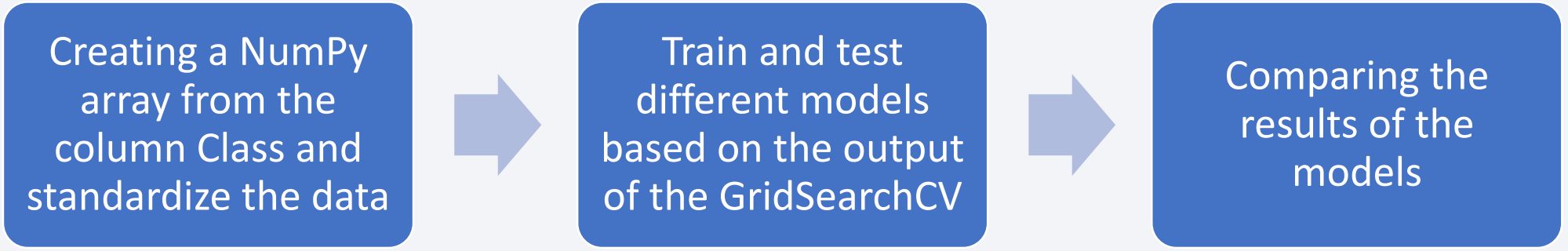
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose
- https://github.com/enes-13-bot/coursera_ibm_applied_data_science_capstone_enes/blob/8cf6fd70b4978c110ad6192c1ac03ad83e7f3f38/lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- 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)

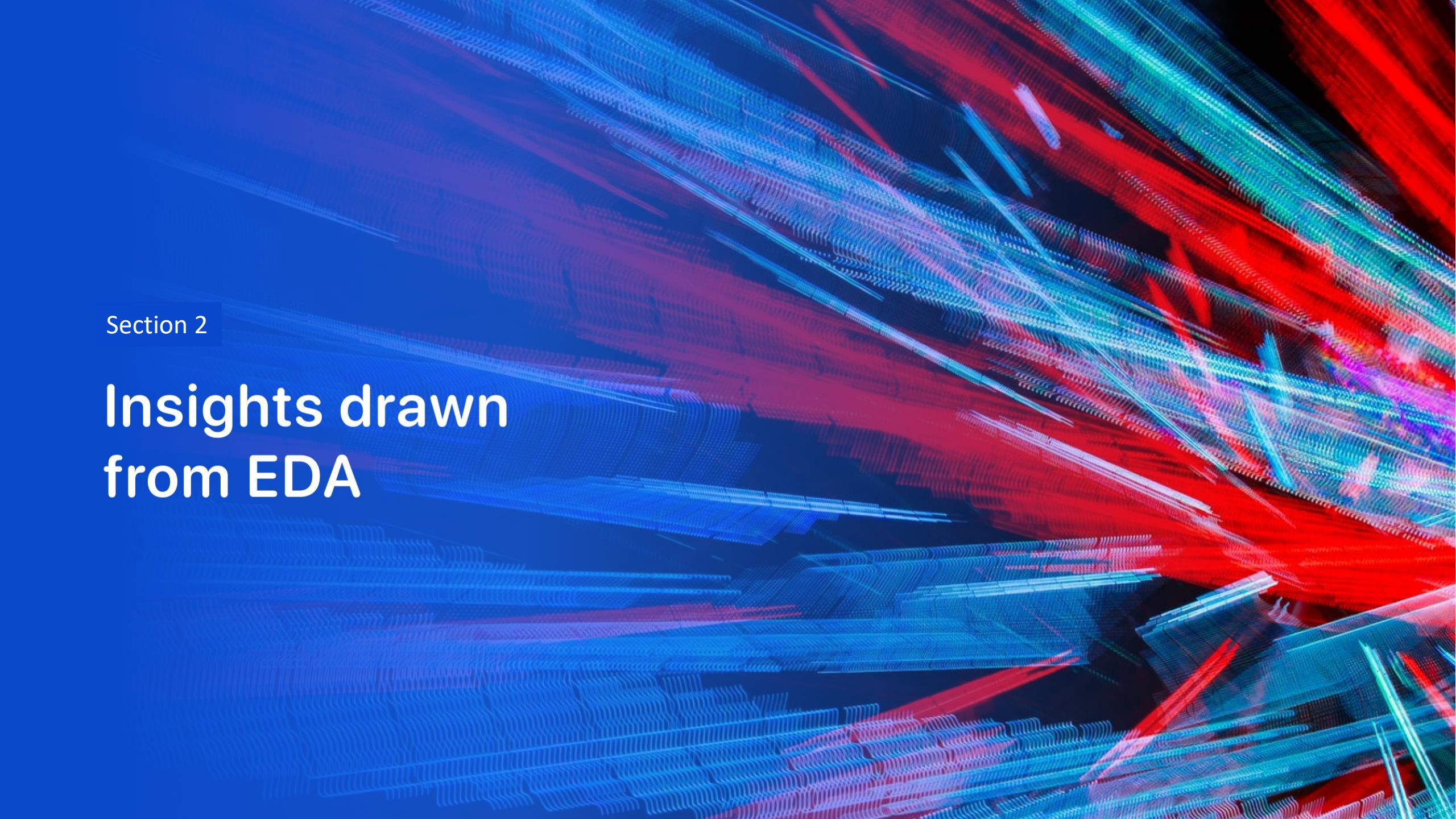
- Summarize how you built, evaluated, improved, and found the best performing classification model



- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose
- https://github.com/enes-13-bot/coursera_ibm_applied_data_science_capstone_enes/blob/8cf6fd70b4978c110ad6192c1ac03ad83e7f3f38/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

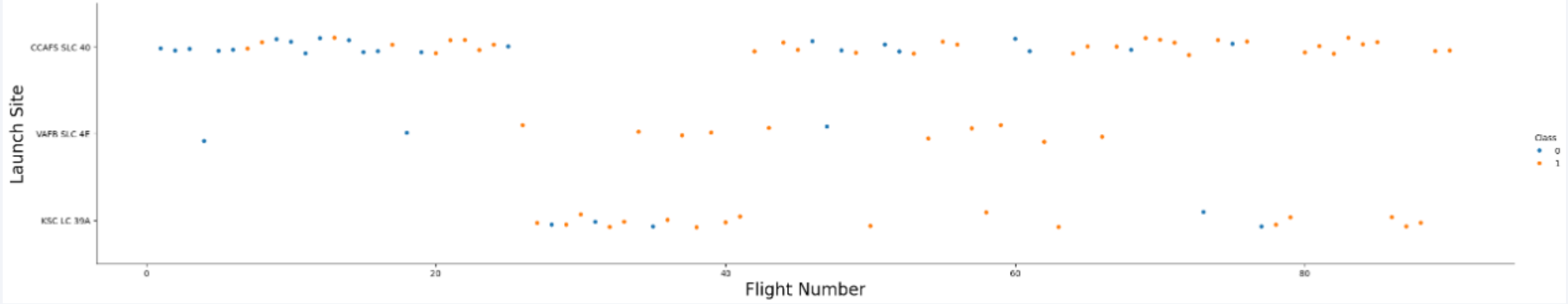
Results

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


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

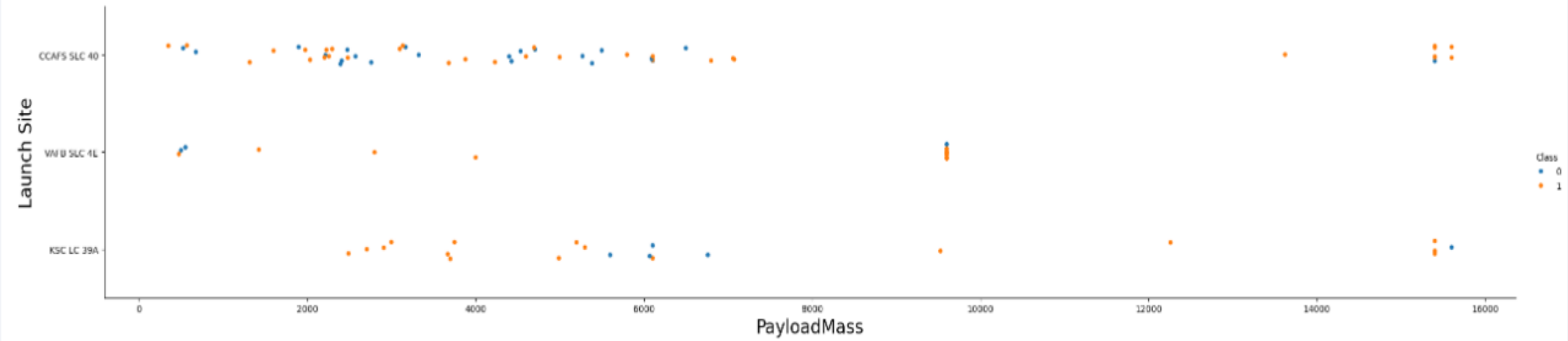
- Show a scatter plot of Flight Number vs. Launch Site



- Show the screenshot of the scatter plot with explanations
  - Success rate grow over time

Payload vs. Launch Site

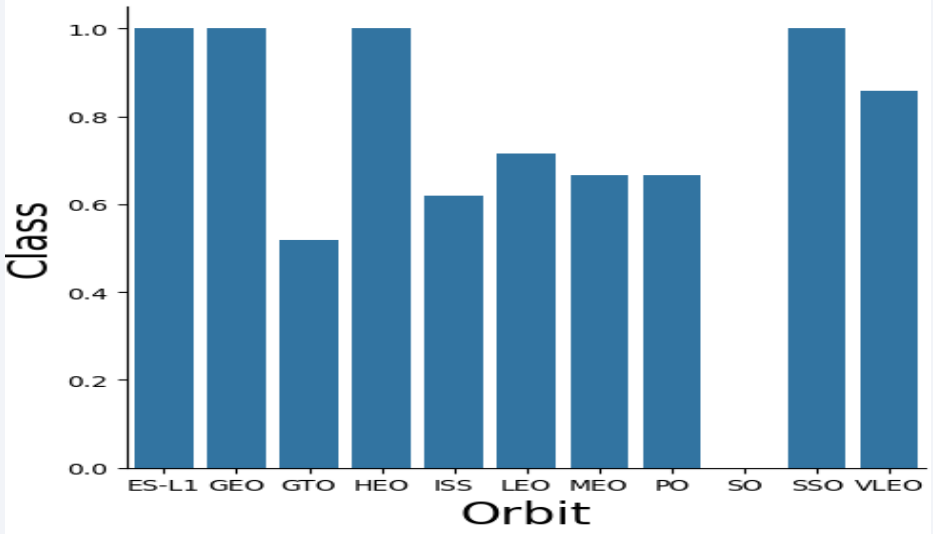
- Show a scatter plot of Payload vs. Launch Site



- Show the screenshot of the scatter plot with explanations
  - Payload Mass over 8000 kg have a good success rate

Success Rate vs. Orbit Type

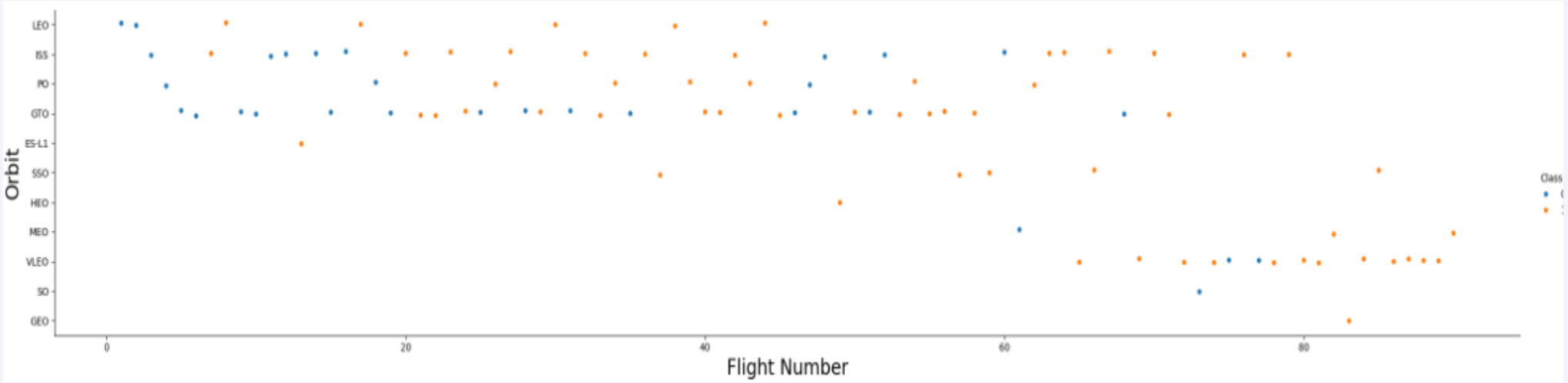
- Show a bar chart for the success rate of each orbit type



- Show the screenshot of the scatter plot with explanations
  - ES-L1, GEO, HEO and SSO have the highest success rate

Flight Number vs. Orbit Type

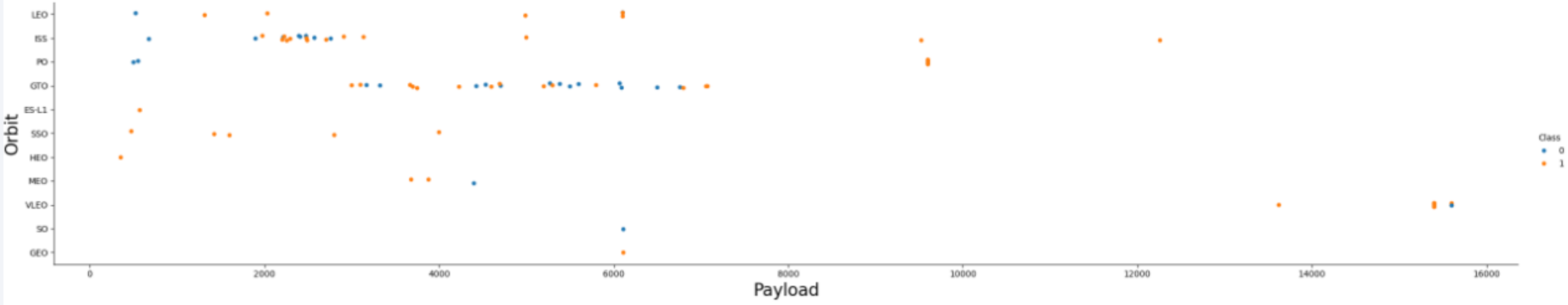
- Show a scatter point of Flight number vs. Orbit type



- Show the screenshot of the scatter plot with explanations
  - Success rate grow over time

Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type



- Show the screenshot of the scatter plot with explanations
  - ISS orbit has a good success rate and the widest range of payload

Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations

All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here
  - Use the command ‘distinct’ to detect the unique launch sites

```
[10]: %sql select distinct Launch_Site from SPACEXTABLE
 * sqlite:///my_data1.db
Done.
```

[10]:

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`

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

- Present your query result with a short explanation here
  - Use the command ‘like’ to filter only records which begin with ‘CCA’

```
%sql Select * from SPACEXTABLE where Launch_Site like 'CCA%'
```

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

sum(PAYLOAD_MASS__KG_)	Customer
7759	ABS Eutelsat
8963	AsiaSat
3669	Bulsatcom
3000	CONAE
3130	CONAE, PlanetIQ, SpaceX

- Present your query result with a short explanation here

```
%sql Select sum(PAYLOAD_MASS__KG_),customer  from SPACEXTABLE group by  customer
```

  - Use a aggregation to calculate the total payload

Average Payload Mass by F9 v1.1

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

avg(PAYLOAD_MASS__KG_)
2928.4

- Present your query result with a short explanation here

```
%sql Select avg(PAYLOAD_MASS__KG_)  from SPACEXTABLE where Booster_Version = 'F9 v1.1'
```

  - Use a aggregation(avg) to calculate the average payload mass carried by booster version F9 v1.1

First Successful Ground Landing Date

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

min(Date)
2018-03-12

- Present your query result with a short explanation here

```
%sql Select min(Date)  from SPACEXTABLE where Landing_Outcome = 'Success'
 * sqlite:///my_data1.db
```

  - Use a aggregation(min) to find the dates of the first successful landing outcome on ground pad

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 SPACEXTABLE where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000
```

  - Use the commands ‘filter’ and ‘between’ to list the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

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

- Present your query result with a short explanation here

```
%sql Select count(Mission_Outcome),Mission_Outcome from SPACEXTABLE group by Mission_Outcome
```

Boosters Carried Maximum Payload

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

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

- Present your query result with a short explanation here

```
select booster_version from SPACEXTABLE where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXTABLE);
```

2015 Launch Records

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

Booster_Version	Launch_Site	month	Date	Landing_Outcome
F9 v1.1 B1012	CCAFS LC-40	10	2015-10-01	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	04	2015-04-14	Failure (drone ship)

- Present your query result with a short explanation here

```
%%sql
select
         booster_version
      ,launch_site
      , substr(Date,6,2) as month
      , date
      , landing_outcome
from SPACEXTABLE
    where landing_outcome = 'Failure (drone ship)' and substr(date,0,5)='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

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

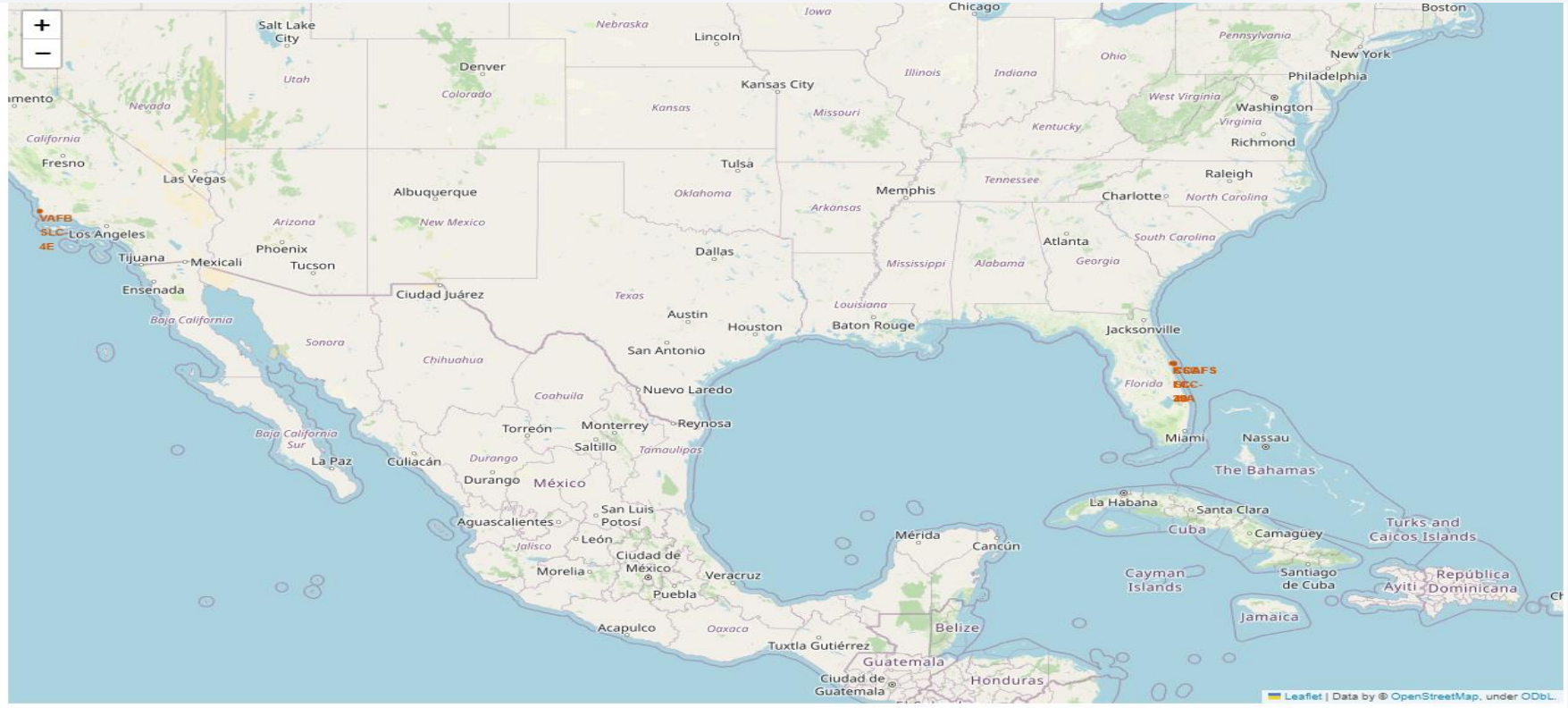
- Present your query result with a short explanation here

```
%%sql
SELECT landing_outcome , count(landing_outcome) FROM SPACEXTABLE
WHERE Date > '2010-06-04' AND Date < '2017-03-20'
GROUP BY landing_outcome
ORDER BY COUNT(landing_outcome) DESC;
```

Section 3

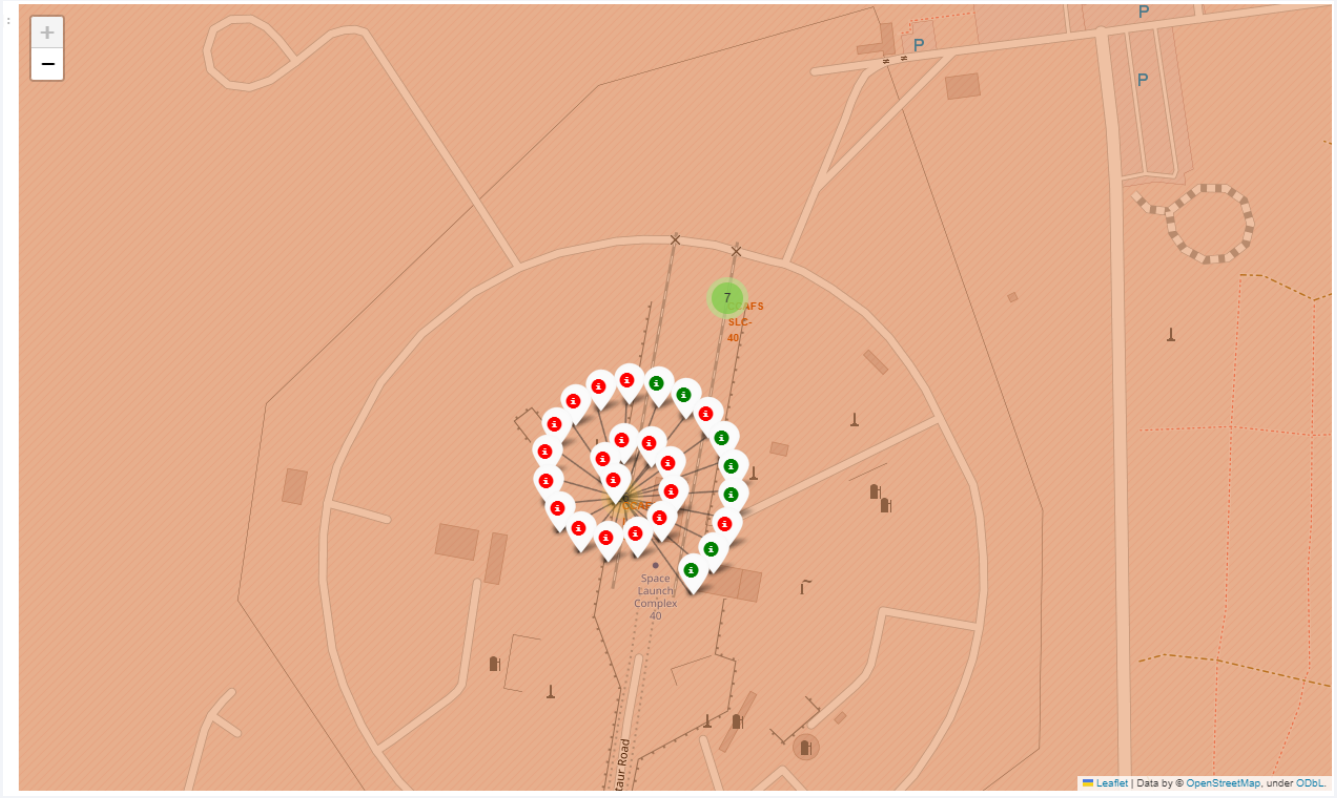
Launch Sites Proximities Analysis

Launch Sites on the map



- Explain the important elements and findings on the screenshot
  - All displayed launch sites are close to the sea

Launch Markers



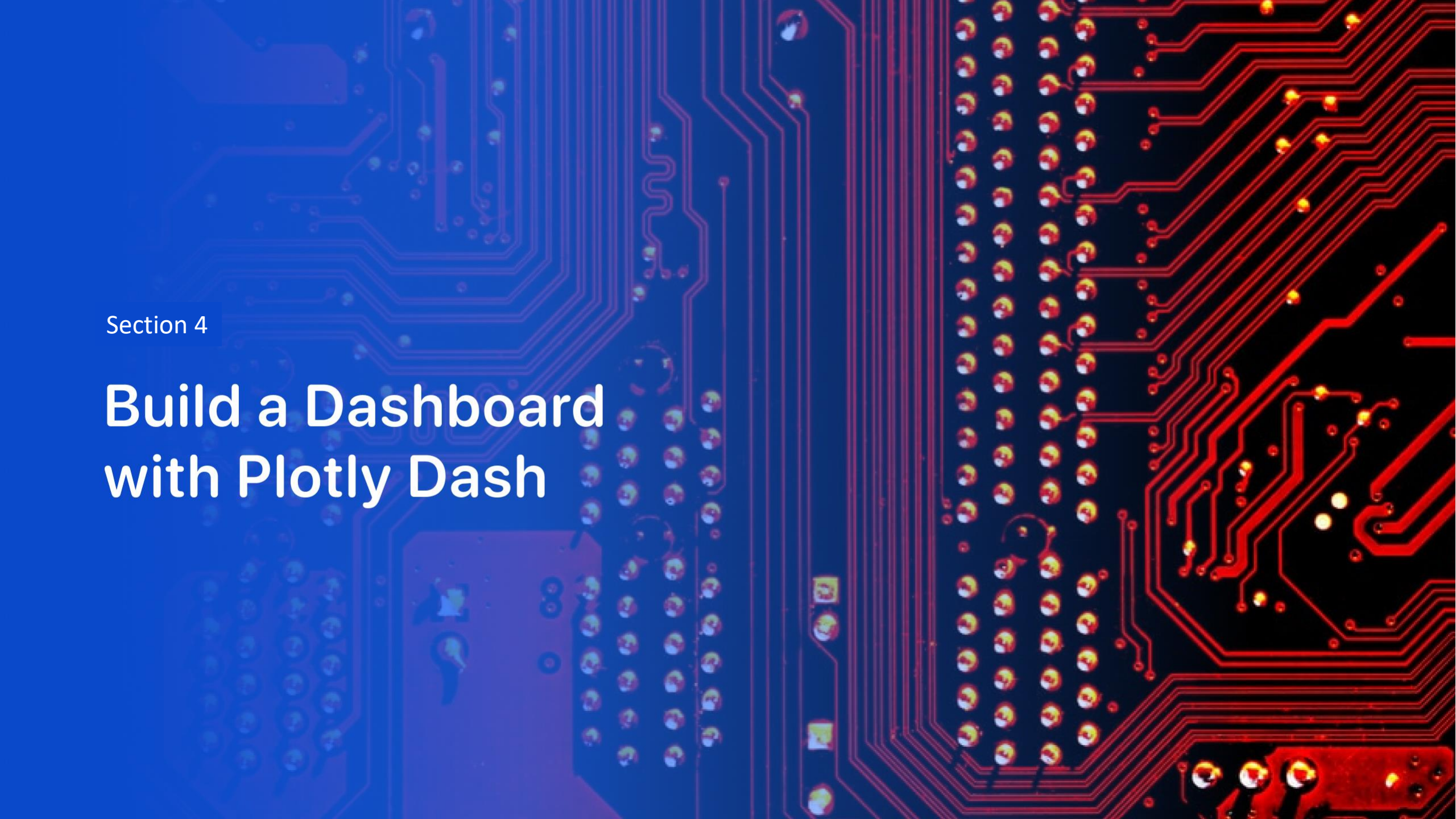
- Explain the important elements and findings on the screenshot
  - Marker which indicates if a launch was successful (green marker) or if the launch failed (red marker)

<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title

- 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


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>

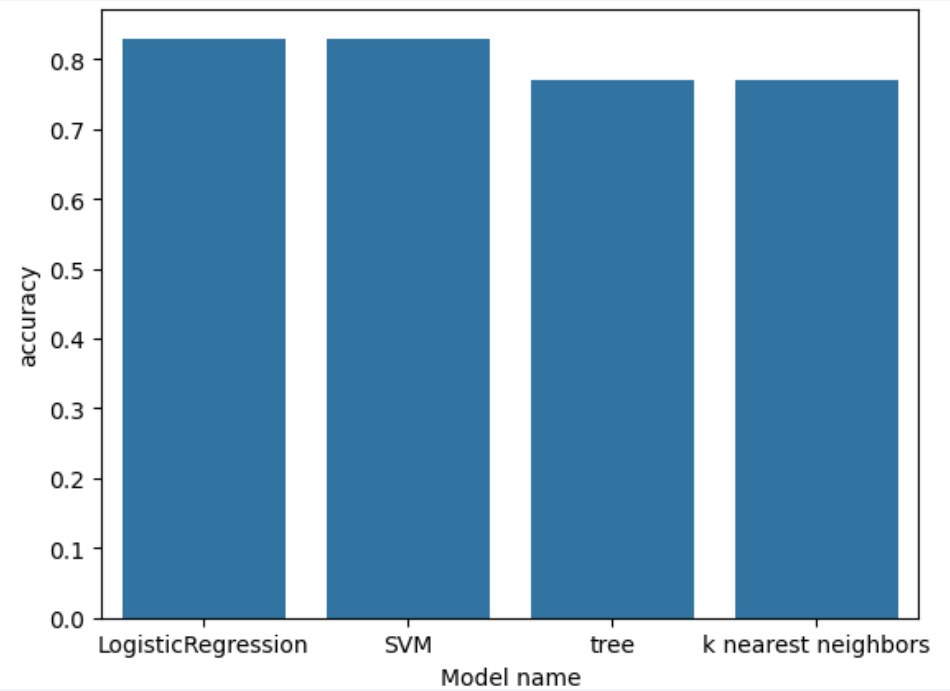
- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version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

Section 5

Predictive Analysis (Classification)

Classification Accuracy

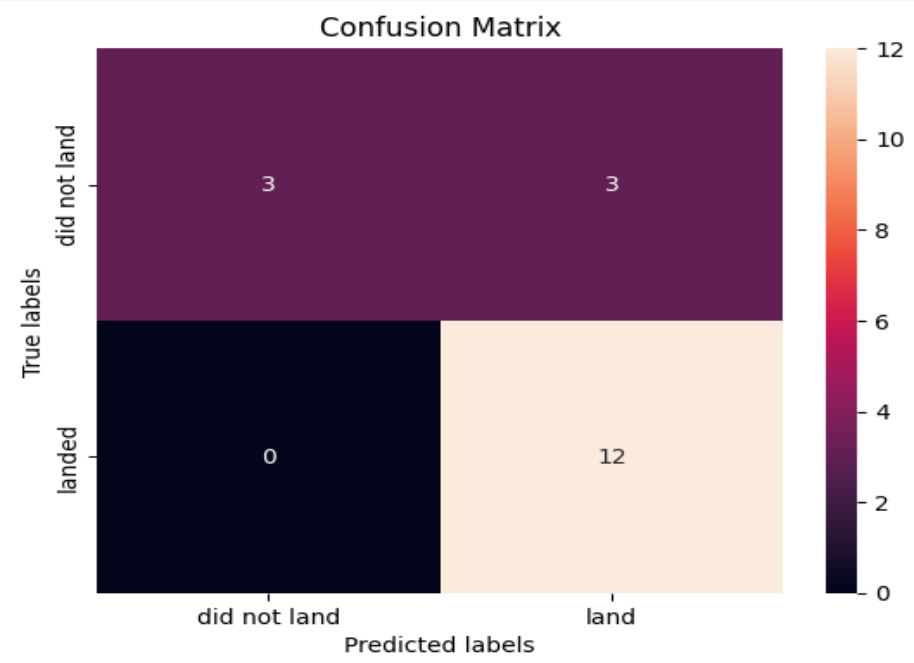
- Visualize the built model accuracy for all built classification models, in a bar chart



- Find which model has the highest classification accuracy
  - LogisticRegression
  - SVM

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation
- SVM confusion matrix



- The models erroneously forecasted 3 successful landings when the actual outcome was unsuccessful landings, indicating false positives.

Conclusions

- Launch success increase over time
- The orbits SSO, GEO, HEO and ES-L1 have a very high success rate

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!