IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Gonzalo Ríos
Feb 1st, 2024

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Executive Summary

In this capstone, we will predict if the Falcon 9 first stage will land successfully. 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, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch. In this lab, you will collect and make sure the data is in the correct format from an API. The following is an example of a successful and launch.

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- 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

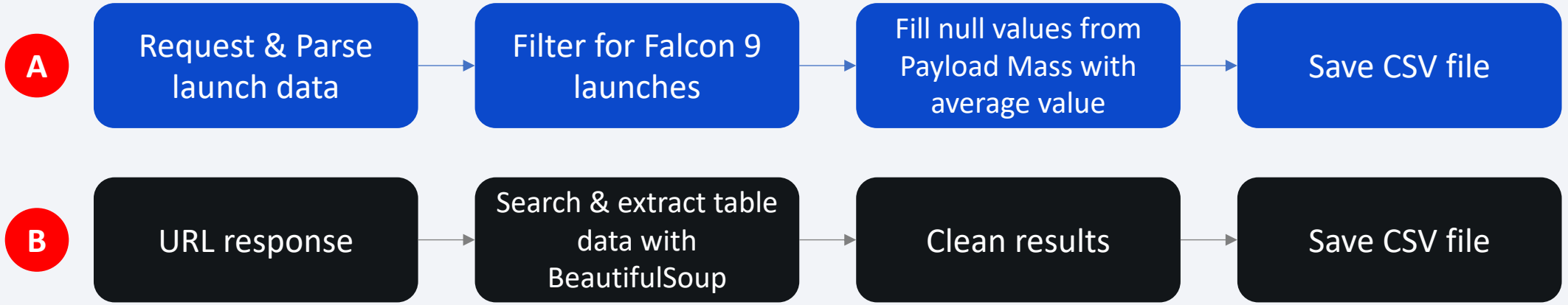
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX REST API
  - Web Scrapping (Wikipedia)
- Perform data wrangling
  - Null Values
  - Feature creation
- 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

- The data for this project was collected from two main sources:
  - SpaceX REST API **A**
  - Wikipedia page (Web Scrapping) **B**



Data Collection – SpaceX API

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

Place your flowchart of SpaceX API calls here

GitHub ref: https://github.com/GRiosG/IBM_DataScience_PC_Capstone

Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

Place your flowchart of web scraping here

GitHub ref: https://github.com/GRiosG/IBM_DataScience_PC_Capstone

Data Wrangling

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

GitHub ref: https://github.com/GRiosG/IBM_DataScience_PC_Capstone

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts

GitHub ref: https://github.com/GRiosG/IBM_DataScience_PC_Capstone

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed

GitHub ref: https://github.com/GRiosG/IBM_DataScience_PC_Capstone

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

GitHub ref: https://github.com/GRiosG/IBM_DataScience_PC_Capstone

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

GitHub ref: https://github.com/GRiosG/IBM_DataScience_PC_Capstone

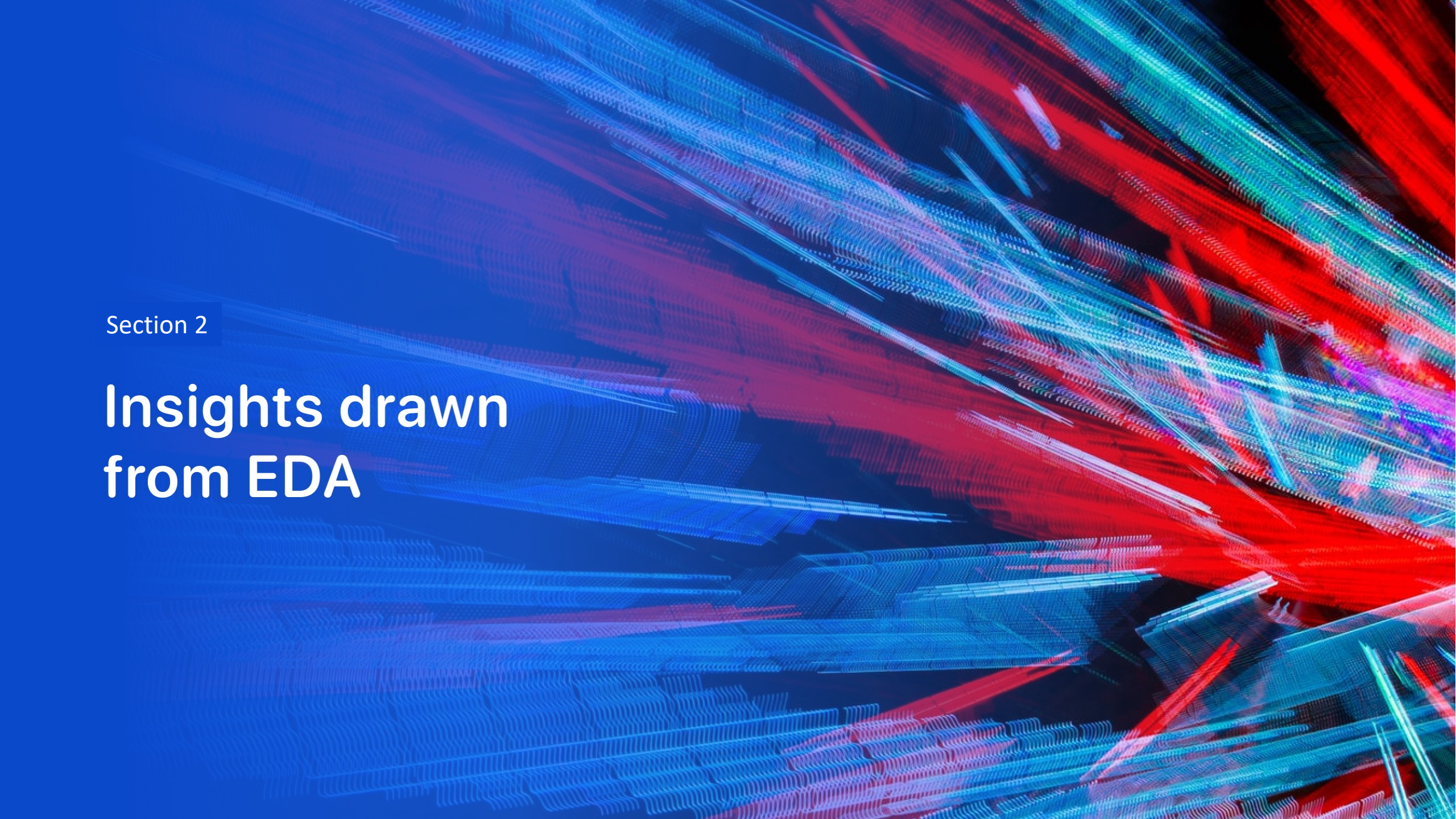
Predictive Analysis (Classification)

- 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

GitHub ref: https://github.com/GRiosG/IBM_DataScience_PC_Capstone

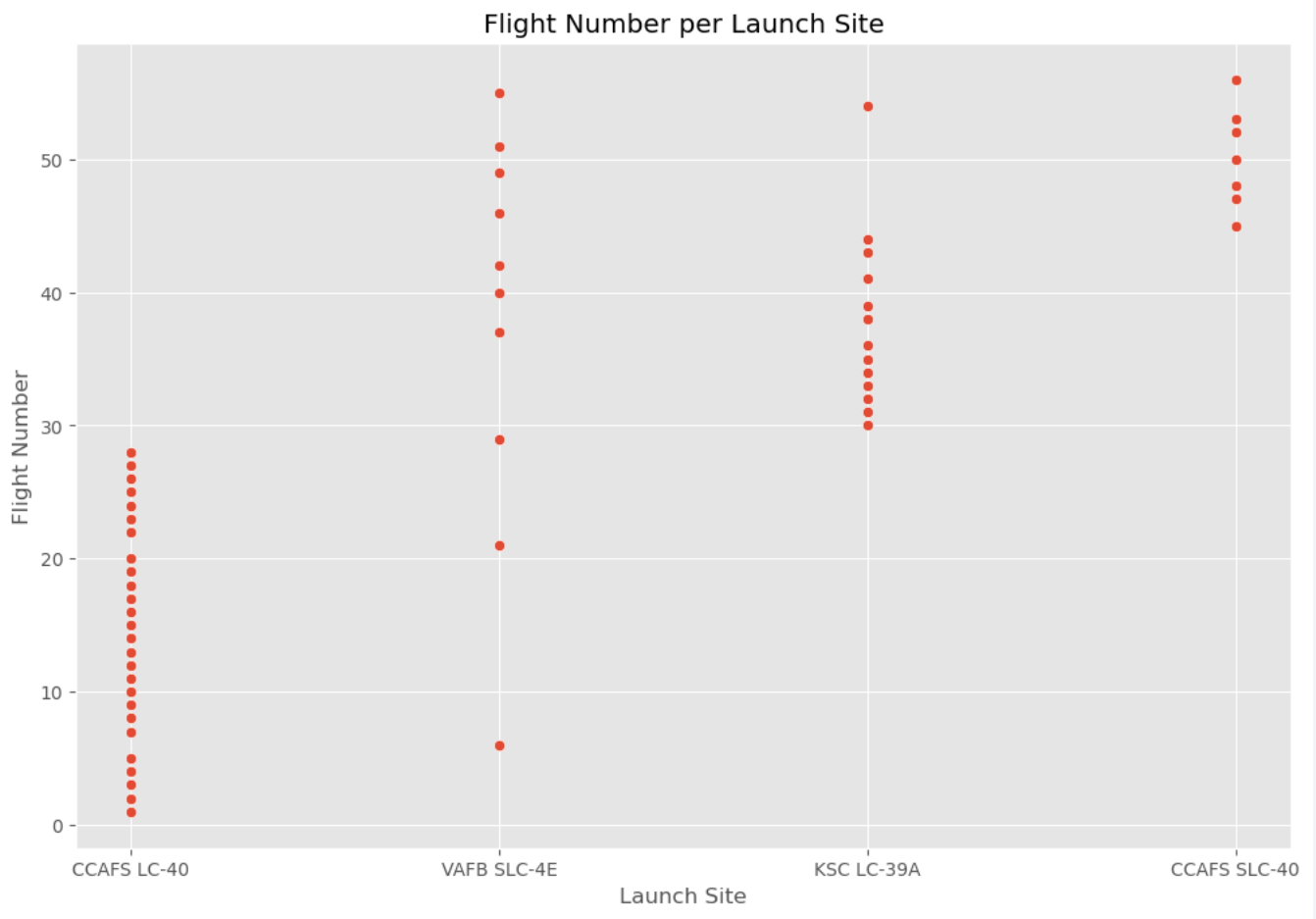
Results

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


Section 2
Insights drawn from EDA

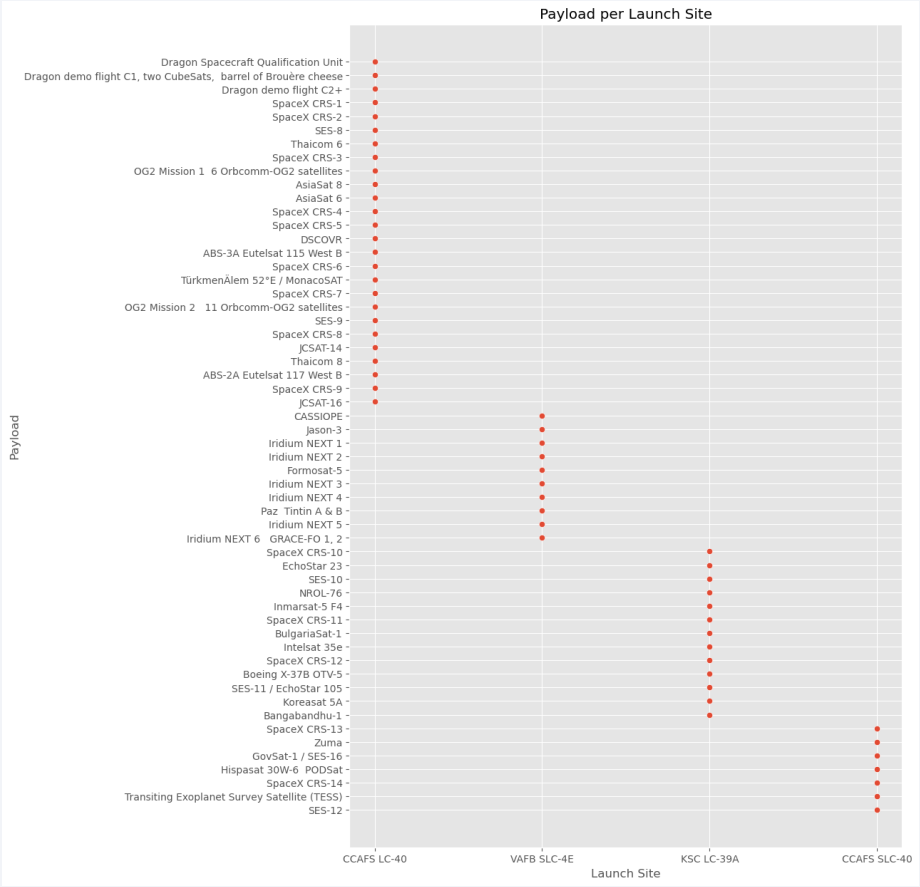
Flight Number vs. Launch Site

- Show the screenshot of the scatter plot with explanations



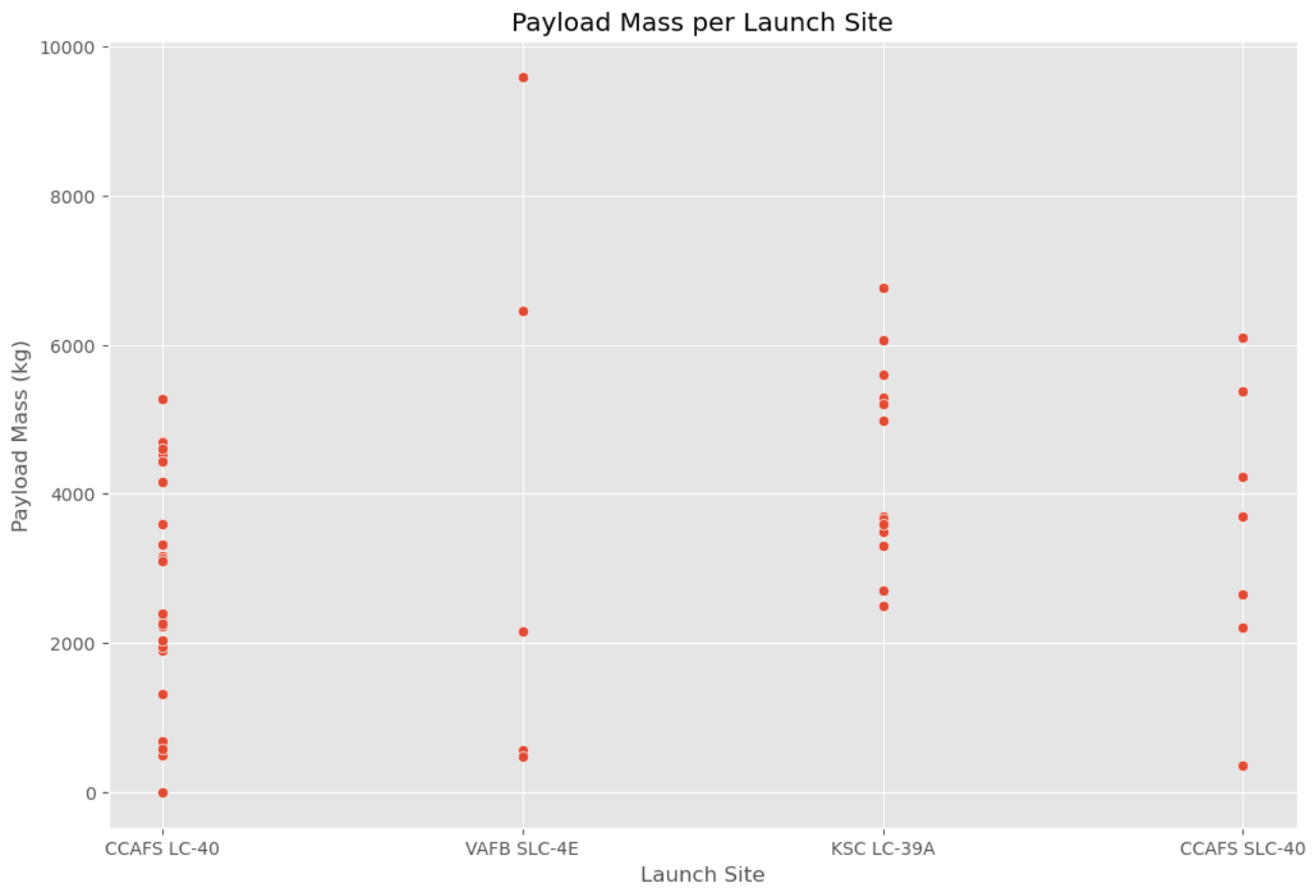
Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site
- Show the screenshot of the scatter plot with explanations



Payload vs. Launch Site (KG?)

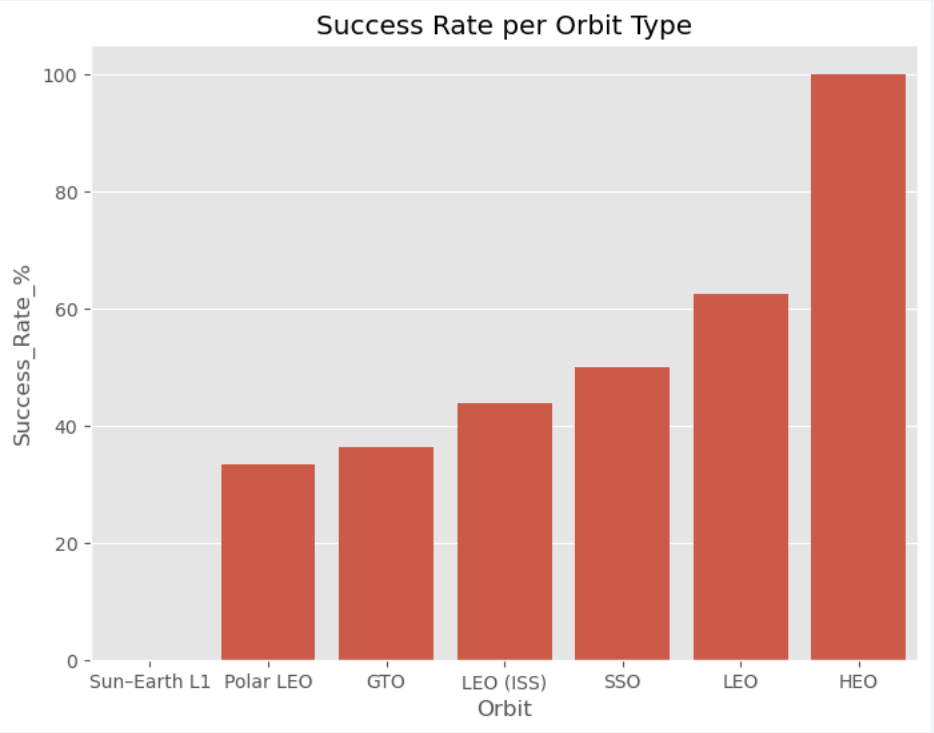
- Show a scatter plot of Payload vs. Launch Site
- Show the screenshot of the scatter plot with explanations



Success Rate vs. Orbit Type

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

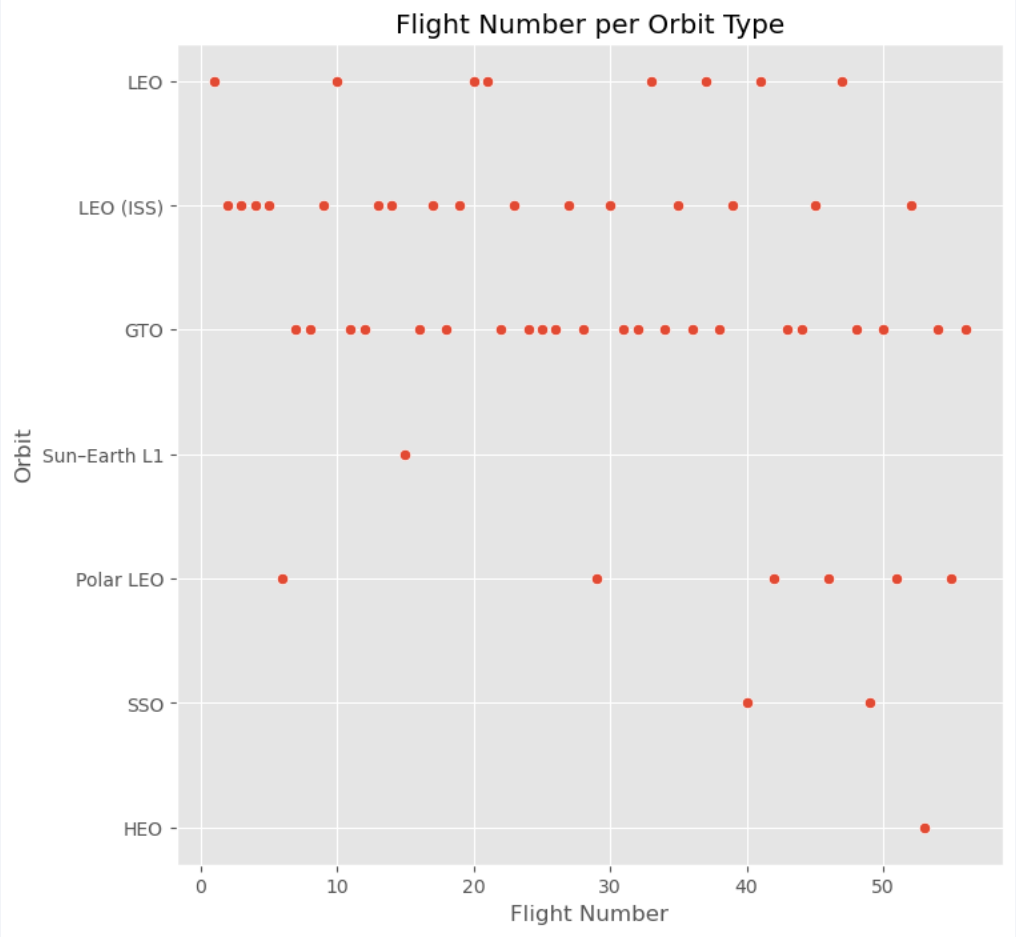
- Show the screenshot of the scatter plot with explanations



Flight Number vs. Orbit Type

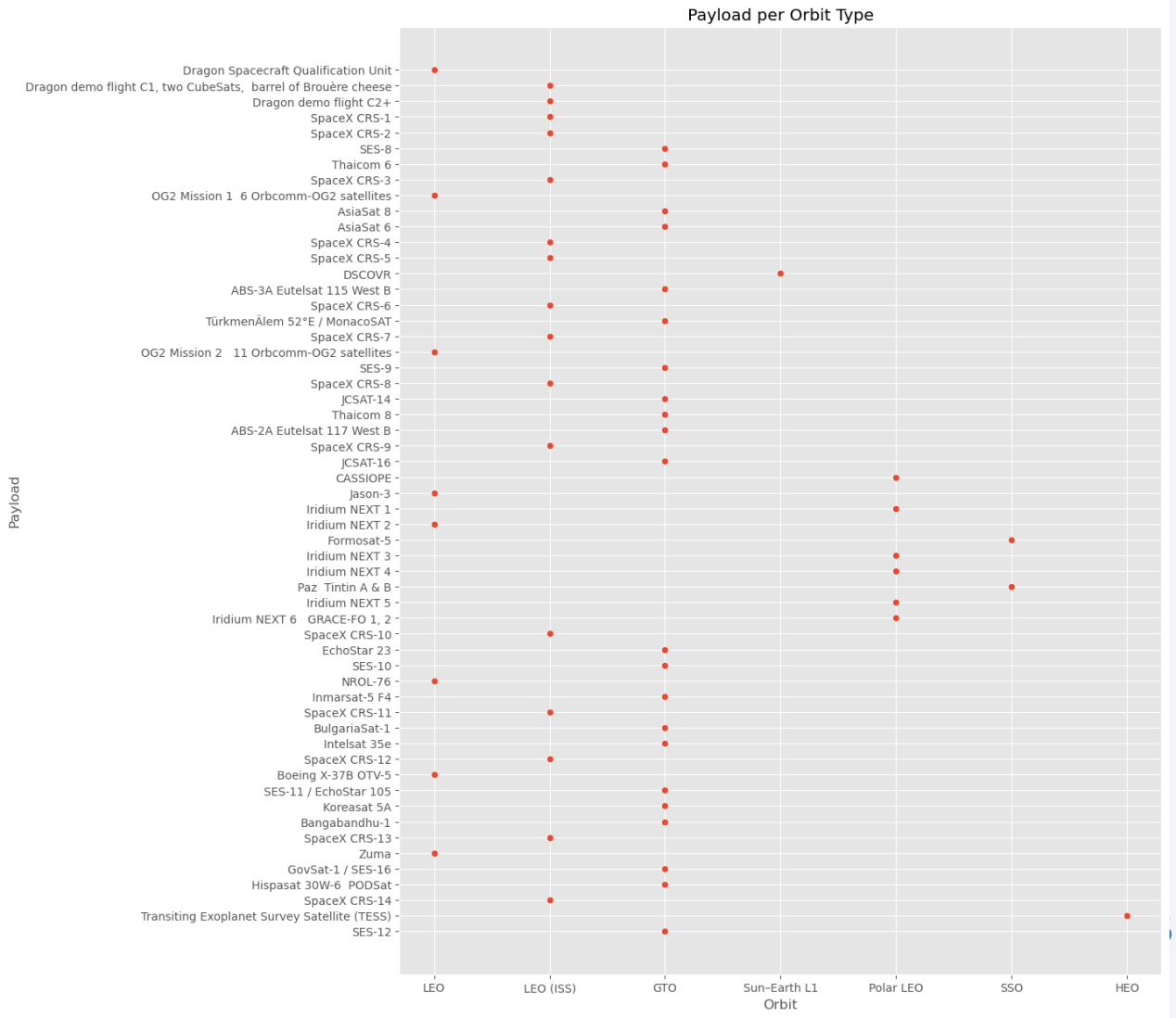
- Show a scatter point of Flight number vs. Orbit type

- Show the screenshot of the scatter plot with explanations



Payload vs. Orbit Type

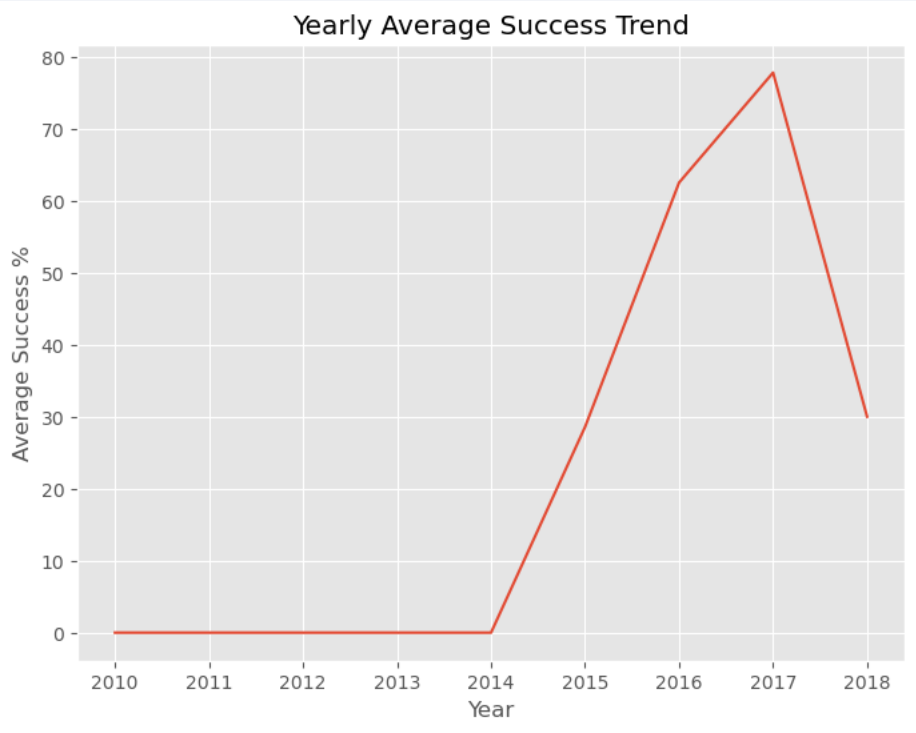
- Show a scatter point of payload vs. orbit type
- Show the screenshot of the scatter plot with explanations



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

```
In 8   1  %sql select distinct Launch_Site from SPACEXTABLE

           * sqlite:///my_data1.db
          Done.

Out 8      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

```
In 9  1  %sql select * from SPACEXTABLE where Launch_Site like 'CCA%' limit 5

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

Out 9

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

```
In 10  1  %sql select sum(PAYLOAD_MASS__KG_) from SPACEXTABLE where Customer is 'NASA (CRS)'

           * sqlite:///my_data1.db
          Done.
Out 10     sum(PAYLOAD_MASS__KG_)
           45596
```

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

```
In 11  1  %sql select avg(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version like 'F9 v1.1%'

           * sqlite:///my_data1.db
          Done.

Out 11     avg(PAYLOAD_MASS__KG_)
           2534.6666666666665
```

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

```
In 12   1  %sql select Date from SPACEXTABLE where Landing_Outcome is "Success (ground pad)" limit 1
        2
```

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

Out 12

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

```
In 13  1  %sql select Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000 and Landing_Outcome is "Success (drone ship)"

   * sqlite:///my_data1.db
  Done.

Out 13   Booster_Version
         F9 FT B1022
         F9 FT B1026
         F9 FT B1021.2
         F9 FT B1031.2
```

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Present your query result with a short explanation here

```
In 14  1  %sql select Mission_Outcome, count() from SPACEXTABLE group by Mission_Outcome

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

Out 14

Mission_Outcome	count()
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
- Present your query result with a short explanation here

```
In 20  1  # Using a subquery
       2  %sql select Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)

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

Out 20

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

- Present your query result with a short explanation here

```
In 26  1 %sql SELECT substr(Date, 6, 2) AS Month,Booster_Version,Launch_Site,Landing_Outcome FROM SPACEXTABLE WHERE substr(Date, 0, 5) = '2015'AND Landing_Outcome = 'Failure (drone ship)'

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

Out 26

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

```
In 27  1 %sql SELECT Landing_Outcome, COUNT(*) AS Count_Outcomes FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY Count_Outcomes DESC;
       2
        * sqlite:///my_data1.db
       Done.
```

Out 27

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


Section 3
Launch Sites
Proximities Analysis

<Folium Map Screenshot 1>

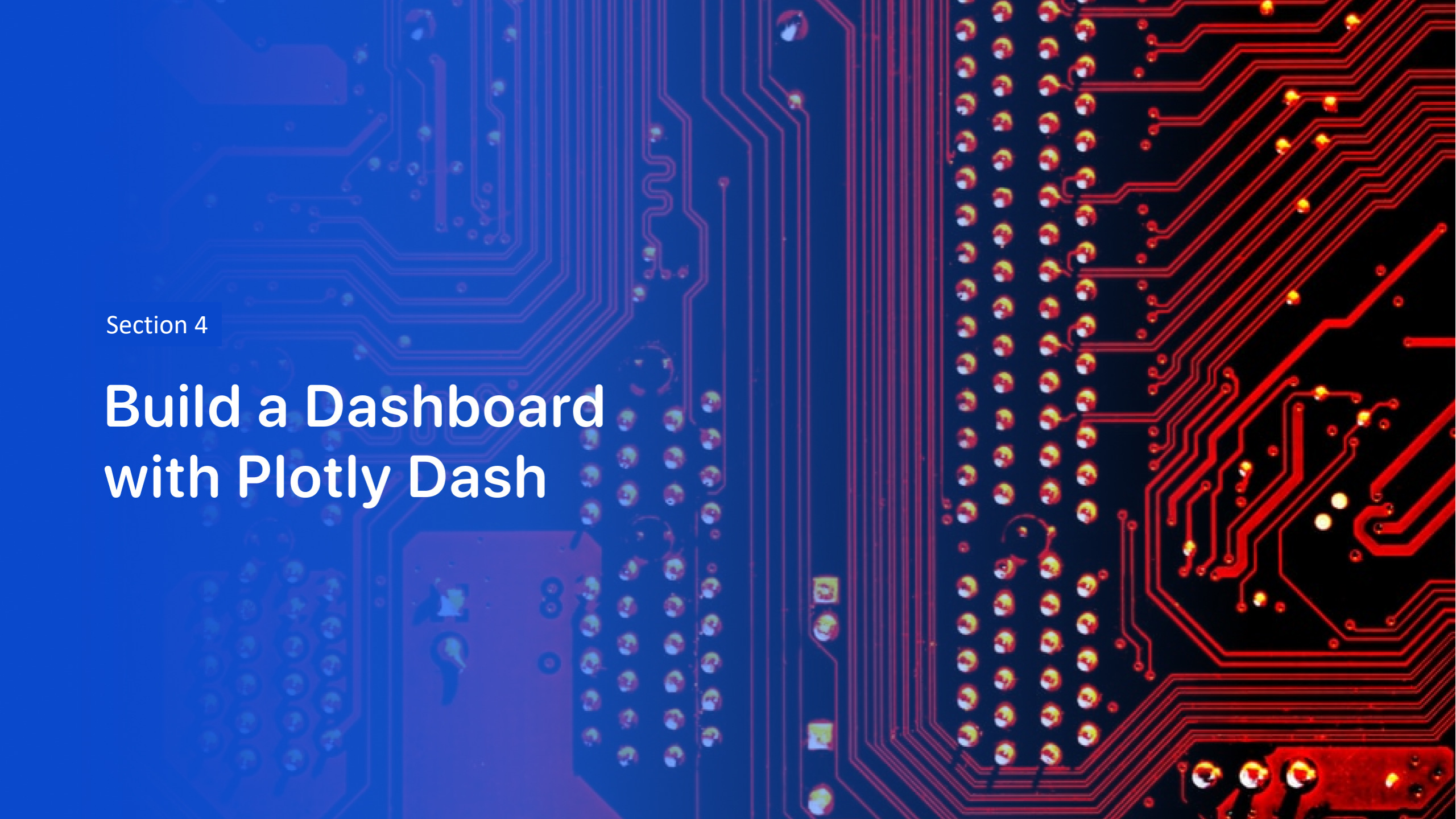
- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 2>

- Replace <Folium map screenshot 2> title with an appropriate title
- 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

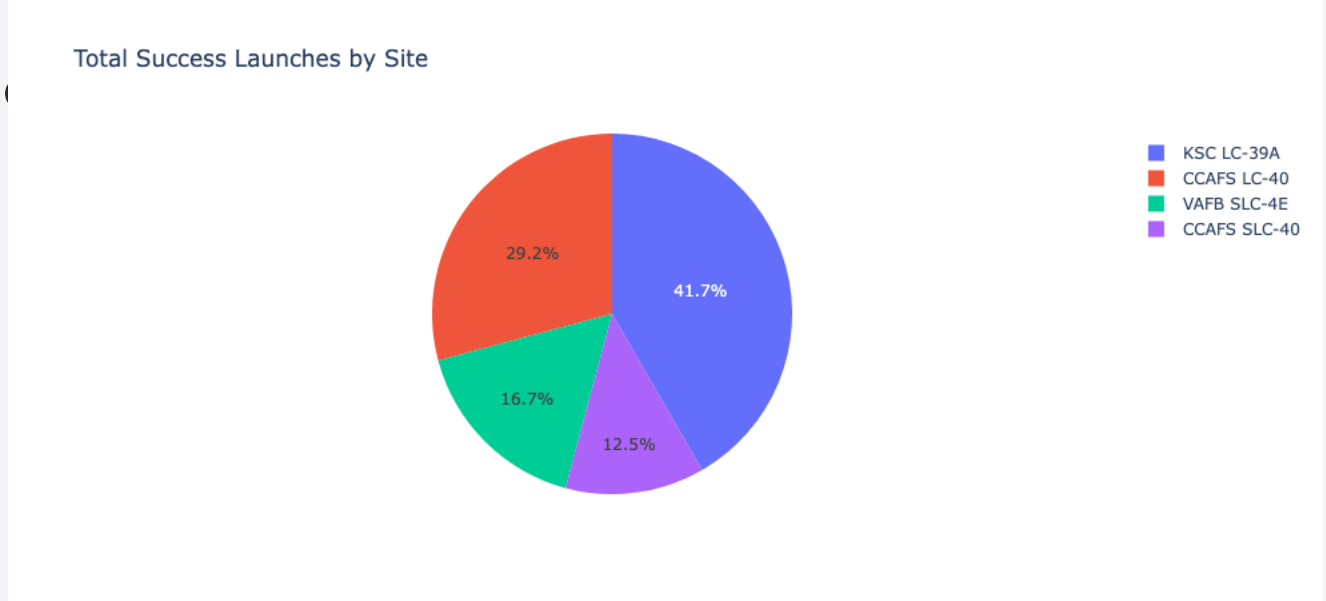
<Folium Map Screenshot 3>

- 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

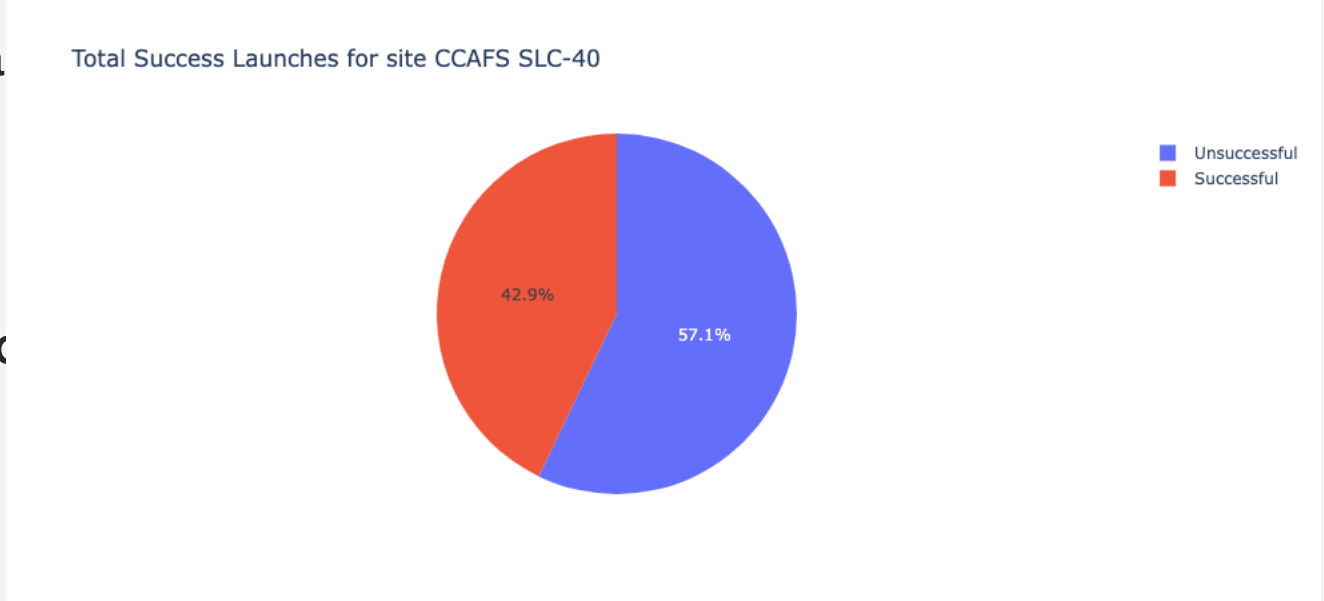
Space X Launch Records Dashboard

- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements an



<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piecha ratio
- Explain the important elements anc



<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

Confusion Matrix

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

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!