Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies
  We have some process like : Data collection methodology、data wrangling、exploratory data analysis (EDA) using visualization and SQL、interactive visual analytics using Folium and Plotly Dash、perform predictive analysis using classification models

- Summary of all results
  We find the method performs best is decision tree classifier.

Introduction

- Project background and context

  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.

- Problems you want to find answers

  I want to predict if the Falcon 9 first stage will land successfully.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

  Requesting rocket launch data from SpaceX API

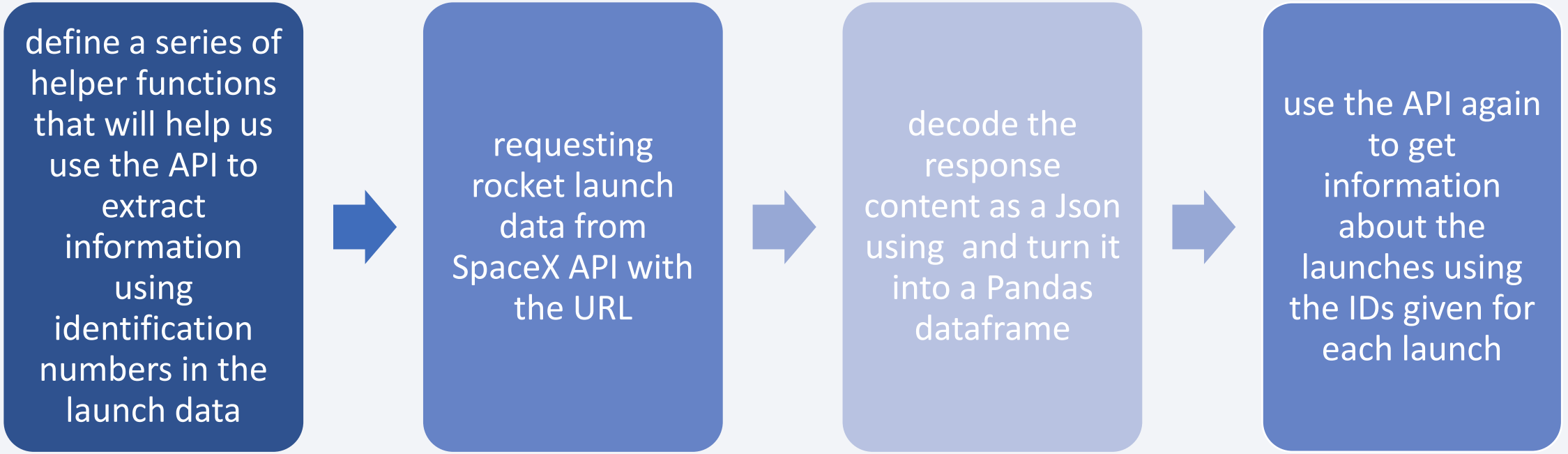
- Perform data wrangling

  Dealing with Missing Values- use the mean for the PayloadMass to replace null values

- 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

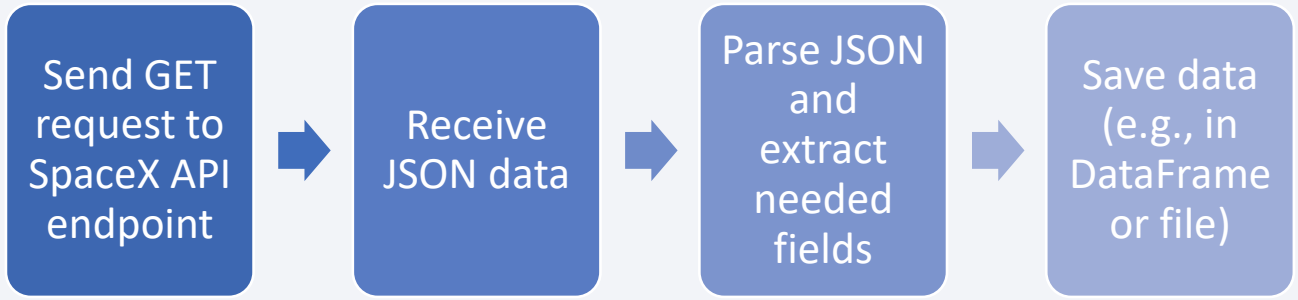
Data Collection

- Requesting rocket launch data from SpaceX API



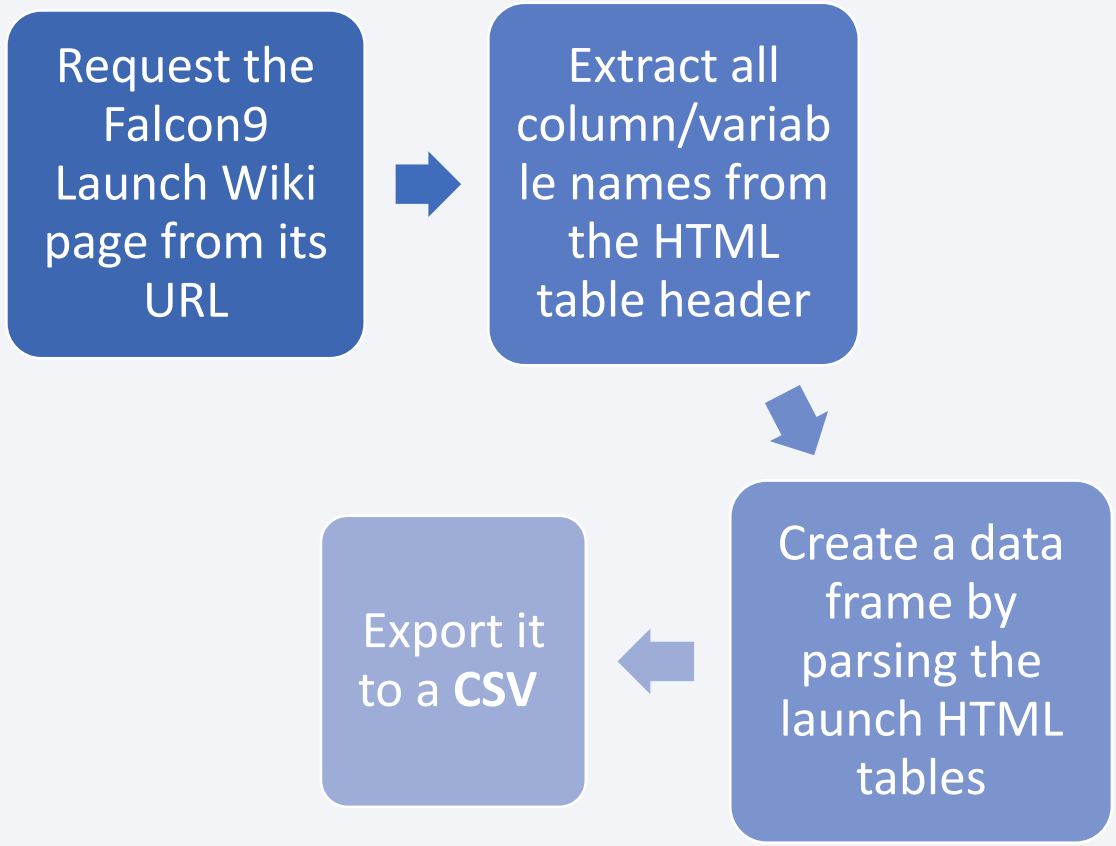
Data Collection – SpaceX API

- API Source: SpaceX REST API
- Tools Used: Python
- Objective: Collect SpaceX launch data (e.g., rocket name, launch date, success status)
- GitHub URL of the completed SpaceX API calls notebook: https://github.com/Zhan-Lin-06/Applied-Data-Science-Capstone/blob/9a91241a494877b261dfe2977eeefef1d403272d/jupyter-labs-spacex-data-collection-api.ipynb



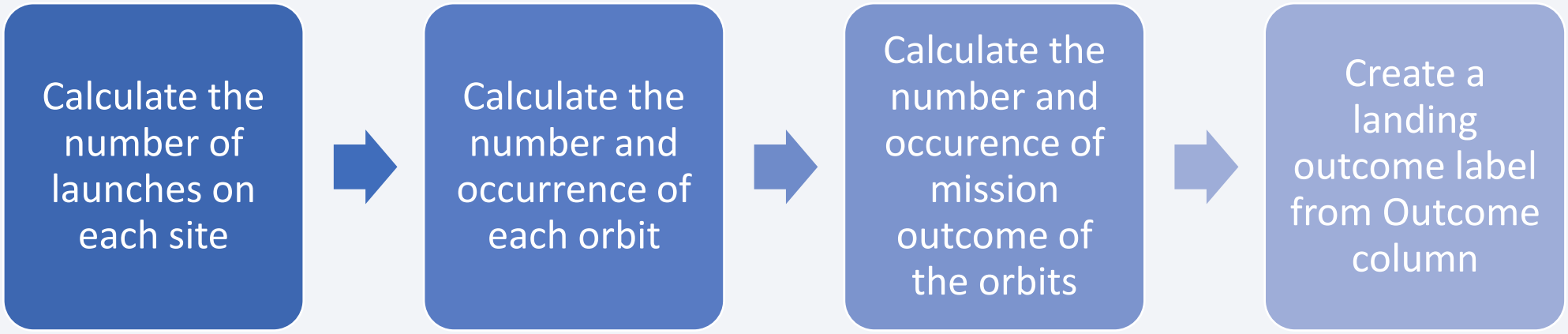
Data Collection - Scraping

- Objective: Collect Falcon 9 historical launch records from a Wikipedia
- Tools Used: Python,BeautifulSoup
- Target Website: Wikipedia page titled List of Falcon 9 and Falcon Heavy launches
- GitHub URL of the completed web scraping notebook: https://github.com/Zhan-Lin-06/Applied-Data-Science-Capstone/blob/9a91241a494877b261dfe2977eeefef1d403272d/jupyter-labs-webscraping.ipynb



Data Wrangling

- We mainly converted those outcomes into Training Labels with 1 means the booster successfully landed 0 means it was unsuccessful.
- GitHub URL of the completed data wrangling related notebooks : https://github.com/Zhan-Lin-06/Applied-Data-Science-Capstone/blob/9a91241a494877b261dfe2977eeefef1d403272d/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Some charts were plotted and why I used those charts :

1. Scatter point chart : Visualize the relationship between x and y
2. Bar chart : Visually check if there are any relationship between x and y
3. Line chart : Visualize the launch success yearly trend

- GitHub URL of the completed EDA with data visualization notebook : https://github.com/Zhan-Lin-06/Applied-Data-Science-Capstone/blob/9a91241a494877b261dfe2977eeefef1d403272d/edadataviz.ipynb

EDA with SQL

- The SQL queries I performed:

- select distinct Launch_Site from SPACEXTBL
- select * from SPACEXTBL where Launch_Site like 'CCA%' limit 5
- SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Customer='NASA (CRS)'
- select min(Date) as firstdate from SPACEXTBL where Landing_Outcome like '%Success%'
- SELECT * FROM SPACEXTBL WHERE Landing_Outcome='Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
- SELECT Mission_Outcome,COUNT(*) FROM SPACEXTBL GROUP BY Mission_Outcome
- SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ IN (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
- SELECT Landing_Outcome,COUNT(*) AS TIMES FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY TIMES DESC

- GitHub URL of my completed EDA with SQL notebook : https://github.com/Zhan-Lin-06/Applied-Data-Science-Capstone/blob/9a91241a494877b261dfe2977eeefef1d403272d/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:

  - Create and add folium.Circle and folium.Marker for each launch site on the site map

  - Draw a PolyLine between a launch site to the selected coastline point

- Explain why added those objects : From the color-labeled markers in marker clusters, we can easily identify which launch sites have relatively high success rates.

- GitHub URL of the completed interactive map with Folium map : https://github.com/Zhan-Lin-06/Applied-Data-Science-Capstone/blob/9a91241a494877b261dfe2977eeefef1d403272d/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)

- How to built, evaluated, improved, and found the best performing classification model :

1. Standardize the data in X then reassign it to the variable X using the transform provided below.
2. Use the function train_test_split to split the data X and Y into training and test data.
3. Create a GridSearchCV object.

- GitHub URL of the completed predictive analysis lab : https://github.com/Zhan-Lin-06/Applied-Data-Science-Capstone/blob/9a91241a494877b261dfe2977eeefef1d403272d/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

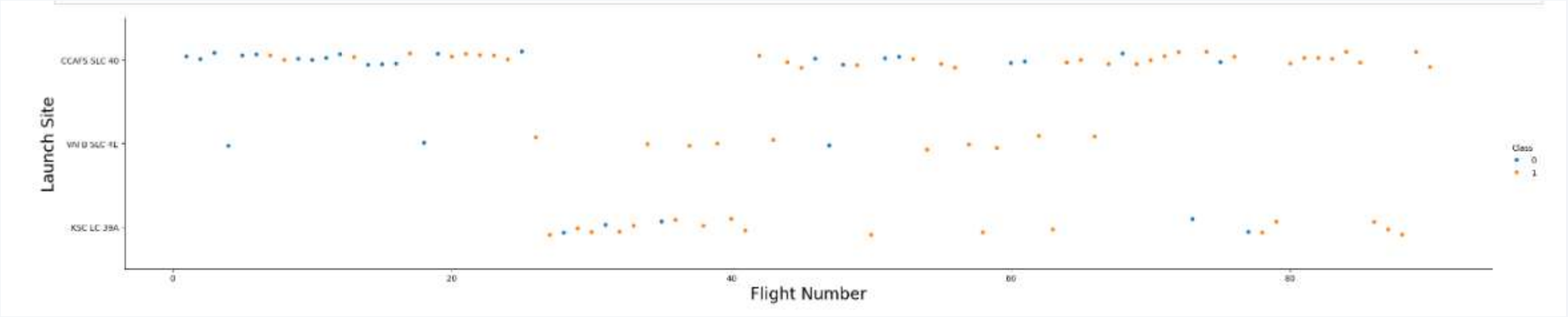
- Exploratory data analysis results :
  - For GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.
  - In the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.
  - For the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).
- Interactive analytics demo in screenshots
- Predictive analysis results : Find the method performs best is decision tree classifier.

Section 2

Insights drawn from EDA

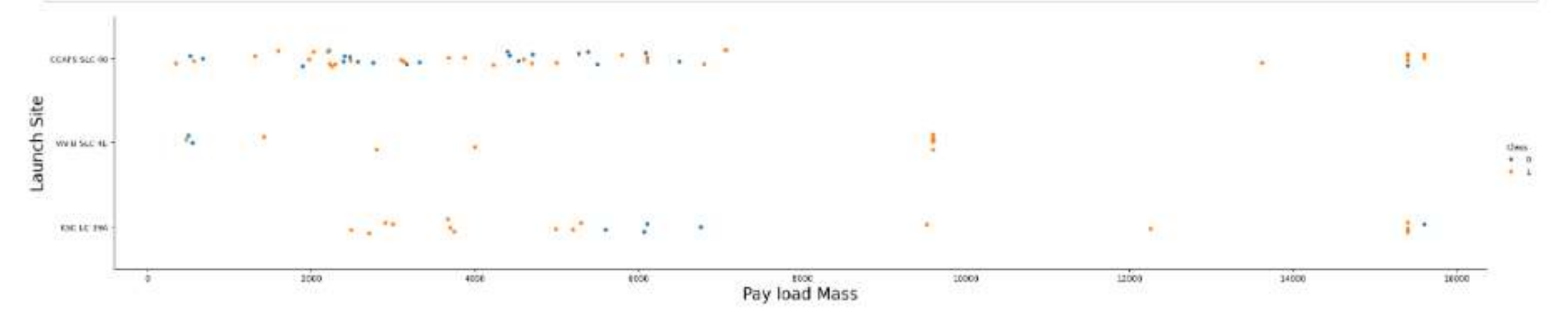
Flight Number vs. Launch Site

Scatter plot of Flight Number vs. Launch Site



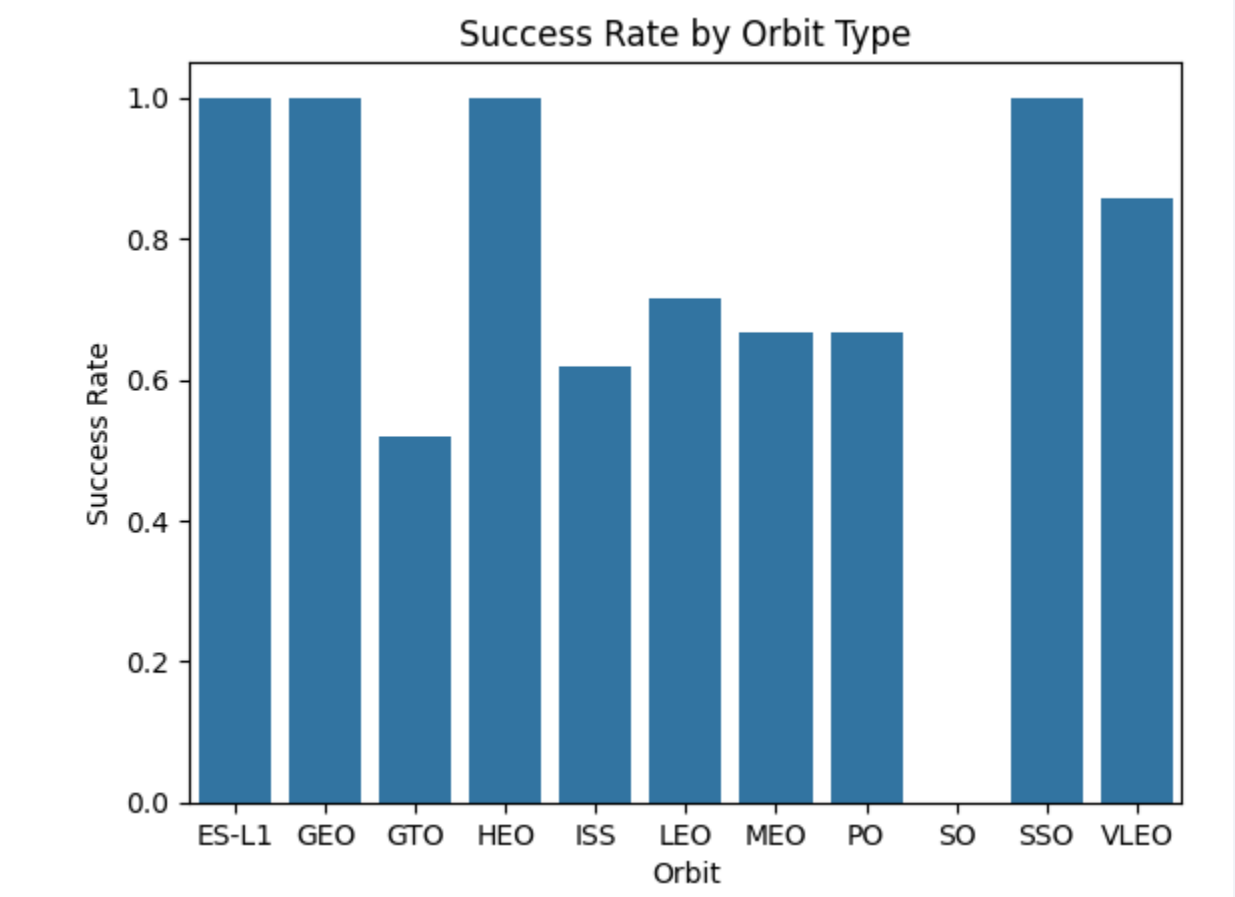
Payload vs. Launch Site

Scatter plot of Payload vs. Launch Site



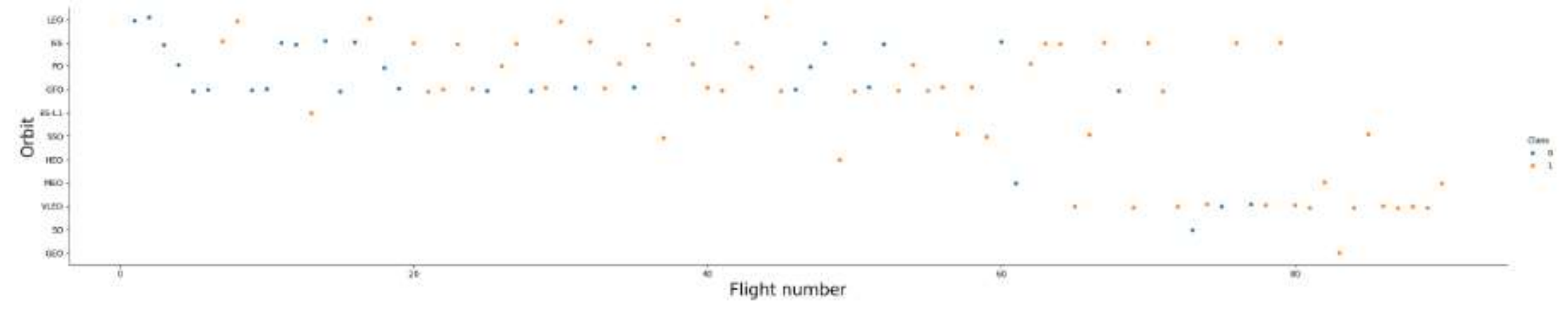
Success Rate vs. Orbit Type

Bar chart for the success rate of each orbit type



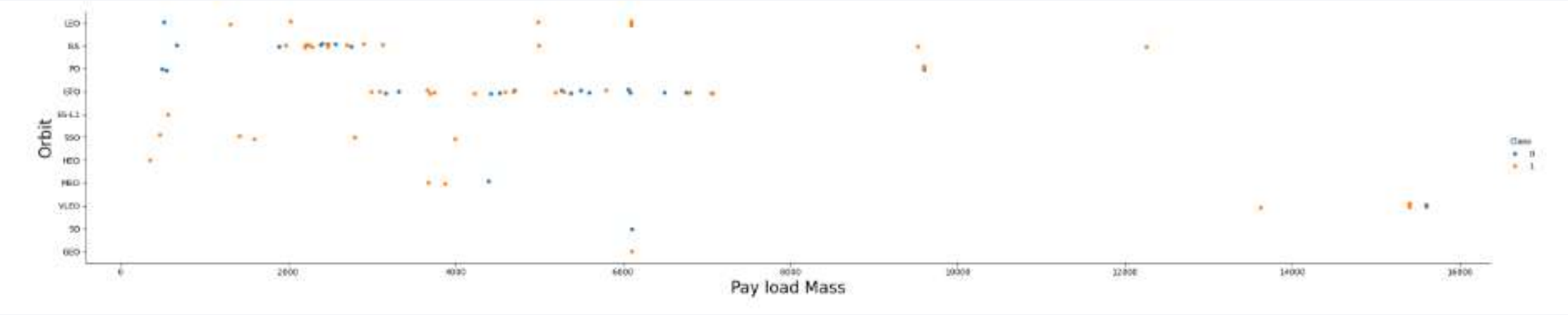
Flight Number vs. Orbit Type

Scatter point of Flight number
vs. Orbit type



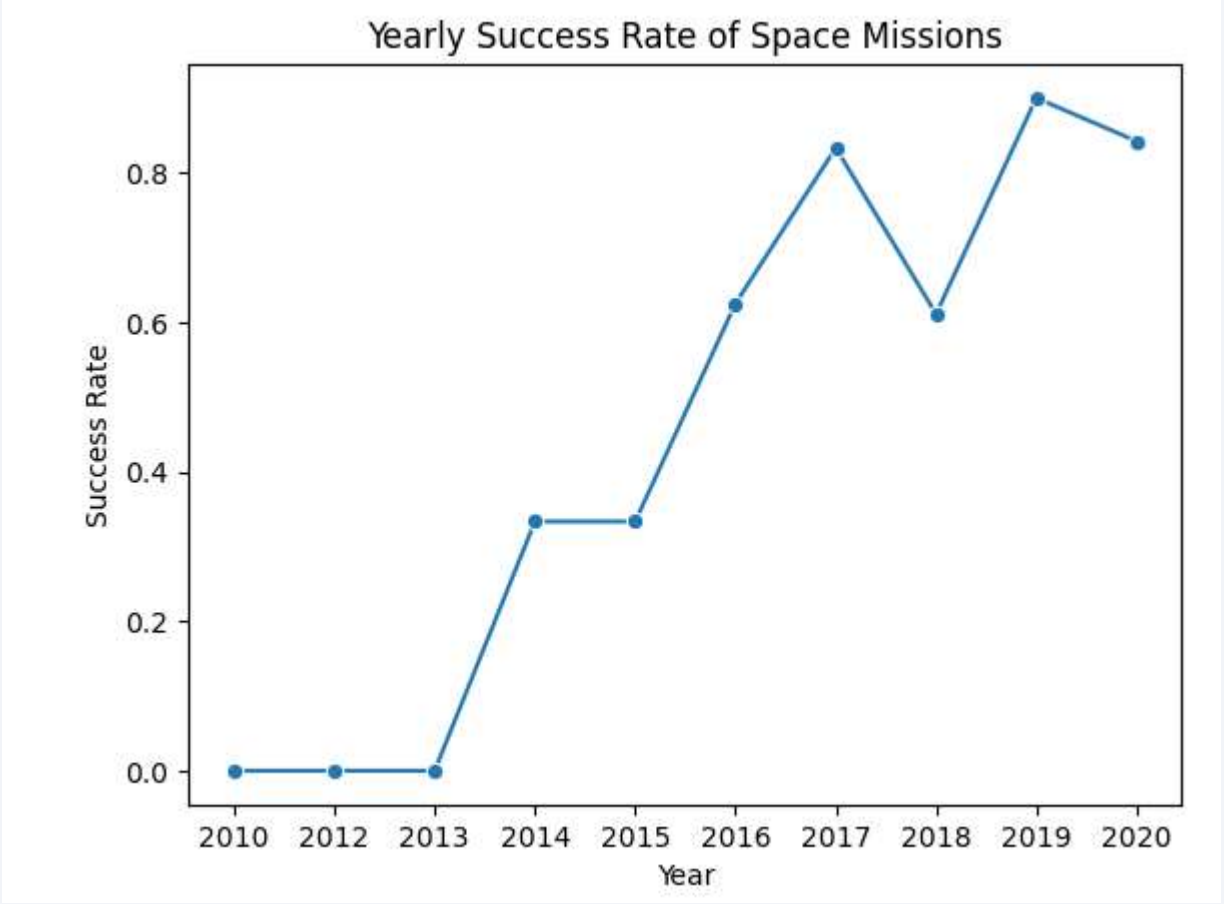
Payload vs. Orbit Type

Scatter point of payload vs.
orbit type



Launch Success Yearly Trend

Line chart of yearly average success rate



All Launch Site Names

- Find the names of the unique launch sites
  %sql select distinct Launch_Site from SPACEXTBL

- Present your query result with a short explanation here
  CCAFS LC-40、VAFB SLC-4E、KSC LC-39A、CCAFS SLC-40
  It displayed the names of the unique launch sites in the space mission

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
  %sql select * from SPACEXTBL where Launch_Site like 'CCA%' limit 5
- Present your query result with a short explanation here

done.

Out[12]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outc
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (paracl
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 (paracl
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No atte
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No atte
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No atte

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
  %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Customer='NASA (CRS)'
- Present your query result with a short explanation here
  It displayed the total payload mass carried by boosters launched by NASA (CRS)

```
In [32]: %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Customer='NASA (CRS)'
 * sqlite:///my_data1.db
Done.
Out[32]: SUM(PAYLOAD_MASS__KG_)
                          45596
```

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
  %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Booster_Version LIKE '%F9 v1.1%'
- Present your query result with a short explanation here
  It displayed average payload mass carried by booster version F9 v1.1

```
 * sqlite:///my_data1.db
Done.
Out[37]: AVG(PAYLOAD_MASS__KG_)
         2534.6666666666665
```

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
  %sql select min(Date) as firstdate from SPACEXTBL where Landing_Outcome like '%Success%'
- Present your query result with a short explanation here
  It list the date when the first succesful landing outcome in ground pad was acheived.

```
 * sqlite:///my_dat
Done.
Out[23]:  firstdate
          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
  %sql SELECT * FROM SPACEXTBL WHERE Landing_Outcome='Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
- Present your query result with a short explanation here

Out[39]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcom
2016-05-06	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (dron ship
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (dron ship
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (dron ship
2017-10-11	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (dron ship

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
  %sql SELECT Mission_Outcome,COUNT(*) FROM SPACEXTBL GROUP BY Mission_Outcome
- Present your query result with a short explanation here
  It list the total number of successful and failure mission outcomes

```
Done.
```

Out[40]:

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
  %sql SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ IN (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
- Present your query result with a short explanation here
  It list all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.

Out[41]:

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
  %sql SELECT substr(Date, 6,2) AS MONTH ,Landing_Outcome,Booster_Version,Launch_Site FROM SPACEXTBL WHERE Landing_Outcome='Failure (drone ship)' AND substr(Date,0,5)='2015'
- Present your query result with a short explanation here
  It 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.

Done.

Out[42]:

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

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
  %sql SELECT Landing_Outcome,COUNT(*) AS TIMES FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY TIMES DESC
- Present your query result with a short explanation here : 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.

Out[48]:

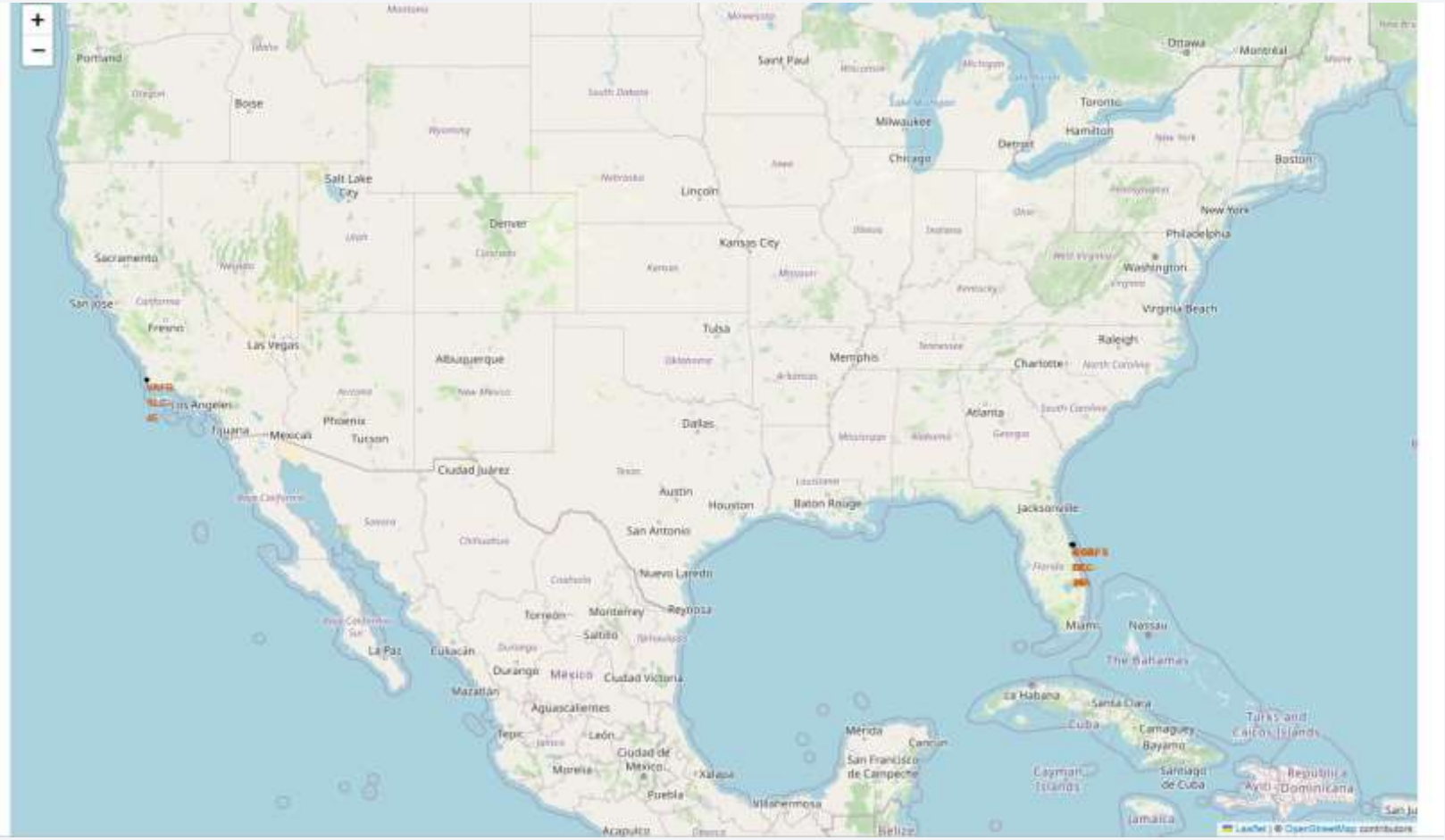
Landing_Outcome	TIMES
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

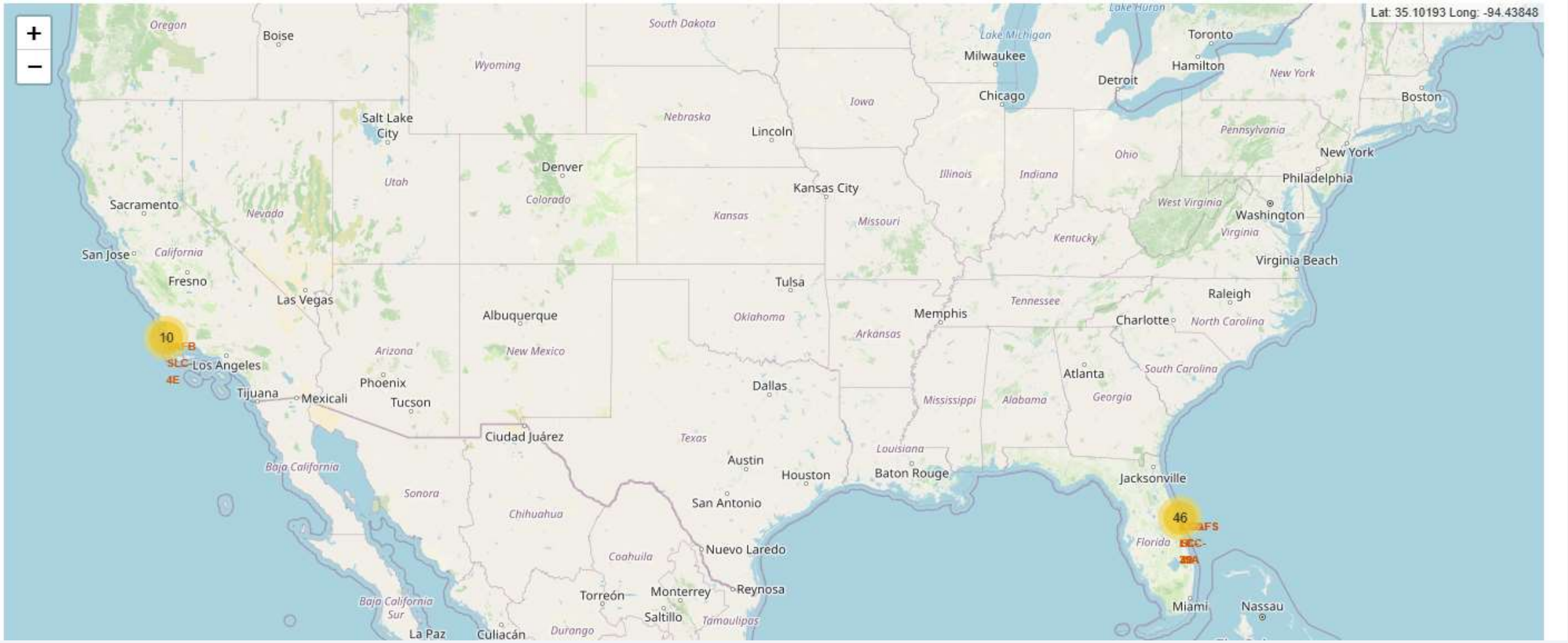
<Launch Site Folium Map >

There are two launch sites show on the map.



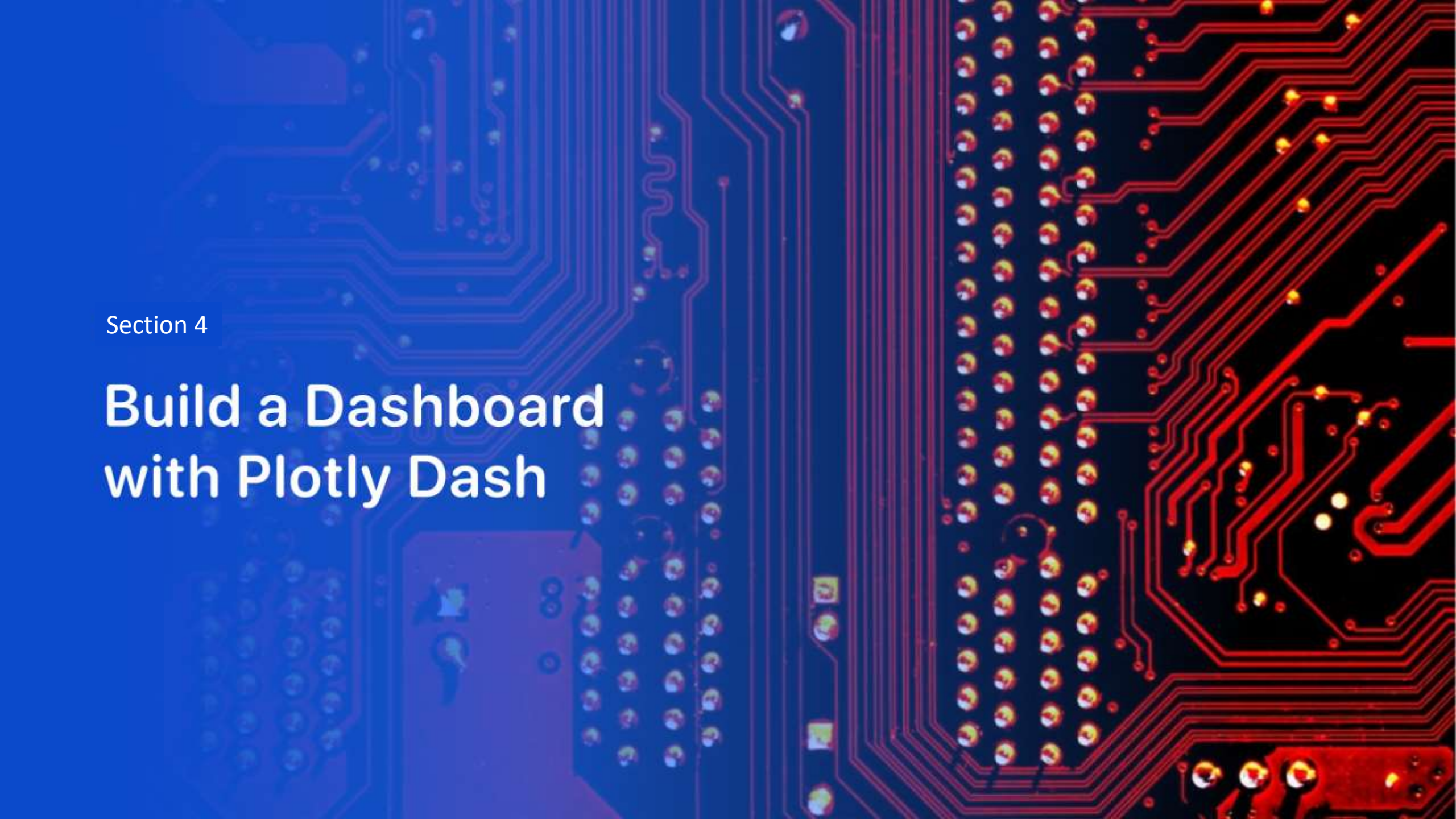
<Folium Map with coordinate>

- We can get coordinate (Lat, Long) for a mouse over on the map



<Folium map with some information >

- The folium map show a selected launch site to its proximities such as railway, highway, coastline, with distance calculated.


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>

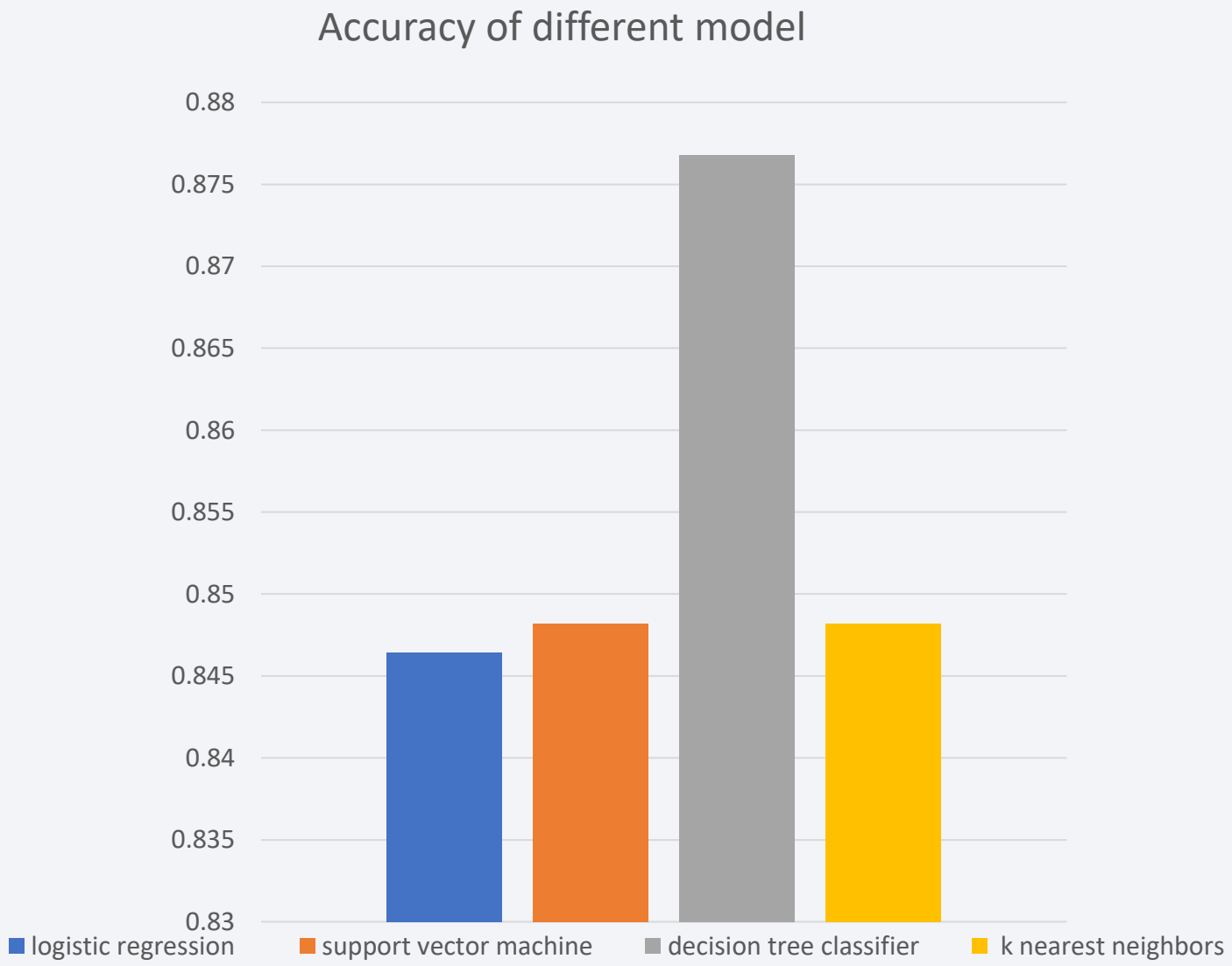
- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version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

Section 5

Predictive Analysis (Classification)

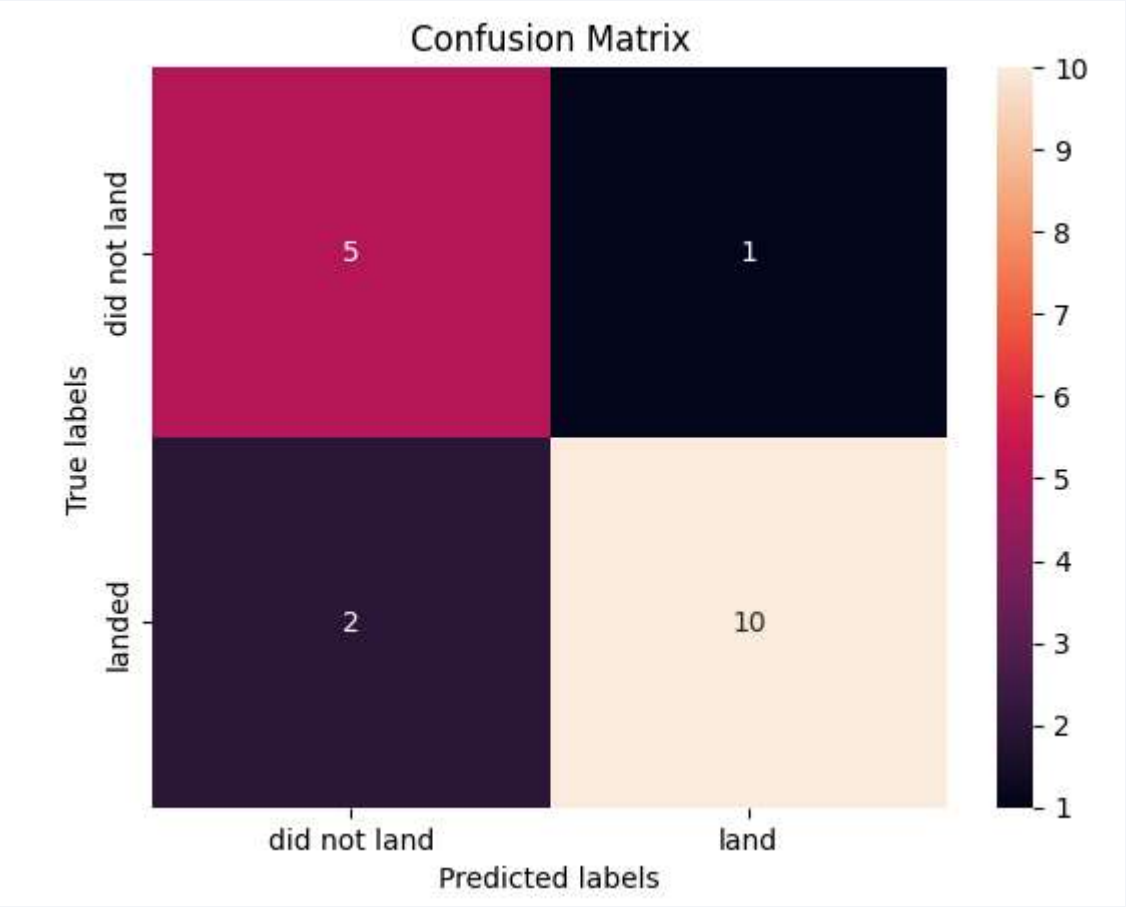
Classification Accuracy

- Decision tree classifier model has the highest classification accuracy



Confusion Matrix

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



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

```
import piplite
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%sql sqlite:///my_data1.db
import folium
spacex_df = spacex_df[['Launch Site', 'Lat', 'Long', 'class']]
```

Thank you!