IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>Rohikian
<Date>08/17/2025

Outline

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

Executive Summary

- Summary of methodologies
  -Logistic Regression, SVM, Classification Trees, k nearest neighbors
- Summary of all results
  -The accuracies of four methods are all the same, 0.833.
  -Classification Trees is slightly higher classification accuracy for training data than the others.

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
  -Predict if the Falcon 9 first stage will land successfully.

Section 1

Methodology

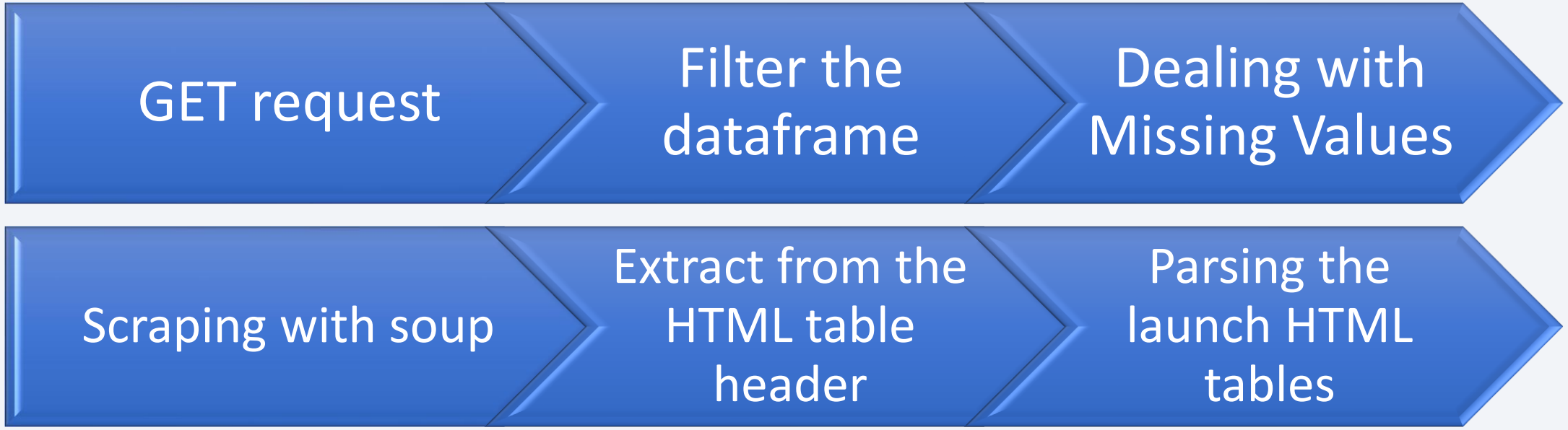
Methodology

Executive Summary

- Data collection methodology:
  - Request to the SpaceX, Extract a Falcon 9 launch records HTML table from Wikipedia
- Perform data wrangling
  - Filtering, Dealing with missing values, Parsing the HTML tags, 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
  - Standardize the data, Split into training data and test data, Fitting to the GridSearch
  - Find the method performs best using test data by comparing the accuracies and confusion matrixes.

Data Collection

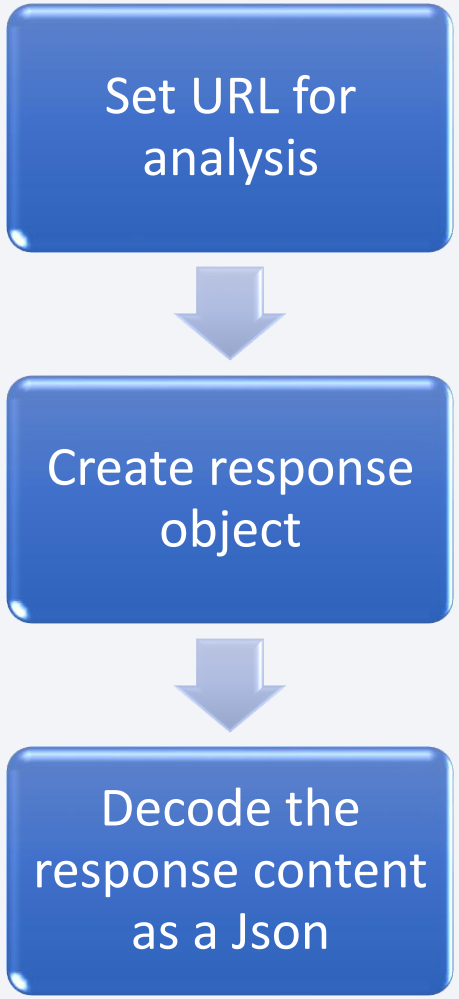
- Describe how data sets were collected.
  - Request to the SpaceX API and Web scrap Falcon 9 launch records
- You need to present your data collection process use key phrases and flowcharts



Data Collection – SpaceX API

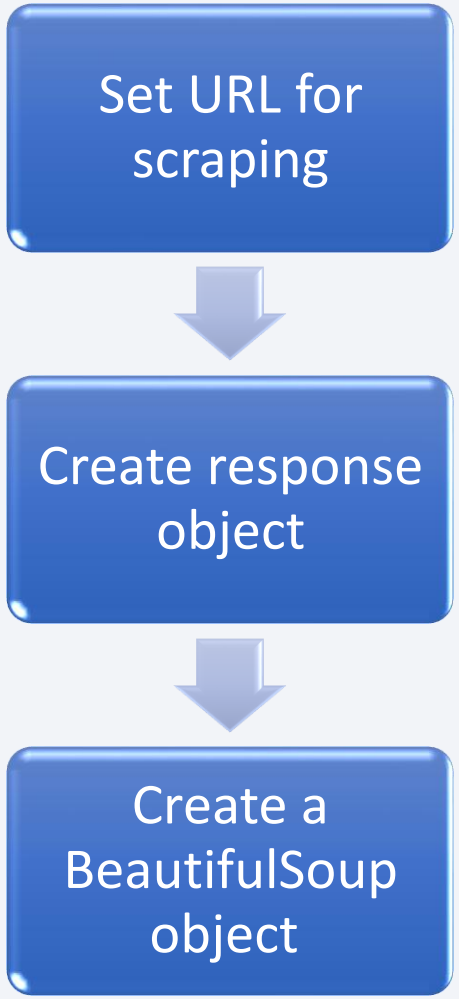
- Present your data collection with SpaceX REST calls using key phrases and flowcharts
  -Request and parse the SpaceX launch data using the GET request
  -We can see that the request was successfull with the 200 status response code by checking the 'status_code' attribute.

- 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/rohikian/Applied-Data-Science-Capstone.git



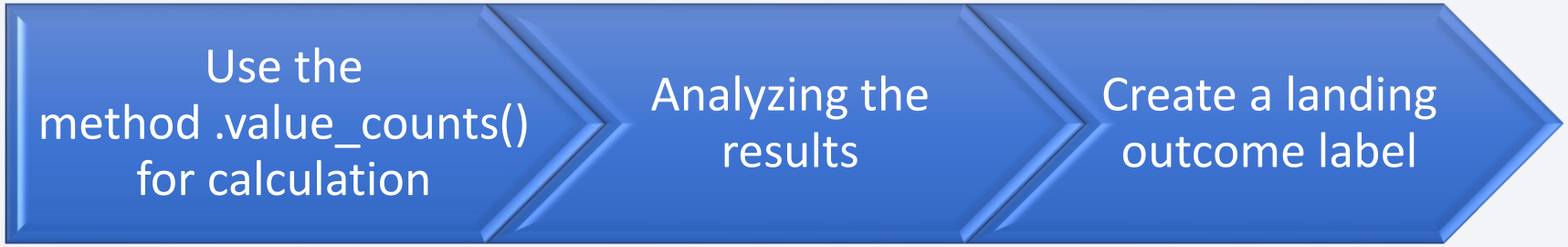
Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts
  -Request the Falcon9 Launch Wiki page from its URL
  -Create a BeautifulSoup object from the HTML response and extract necessary columns.
- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose
  - https://github.com/rohikian/Applied-Data-Science-Capstone.git



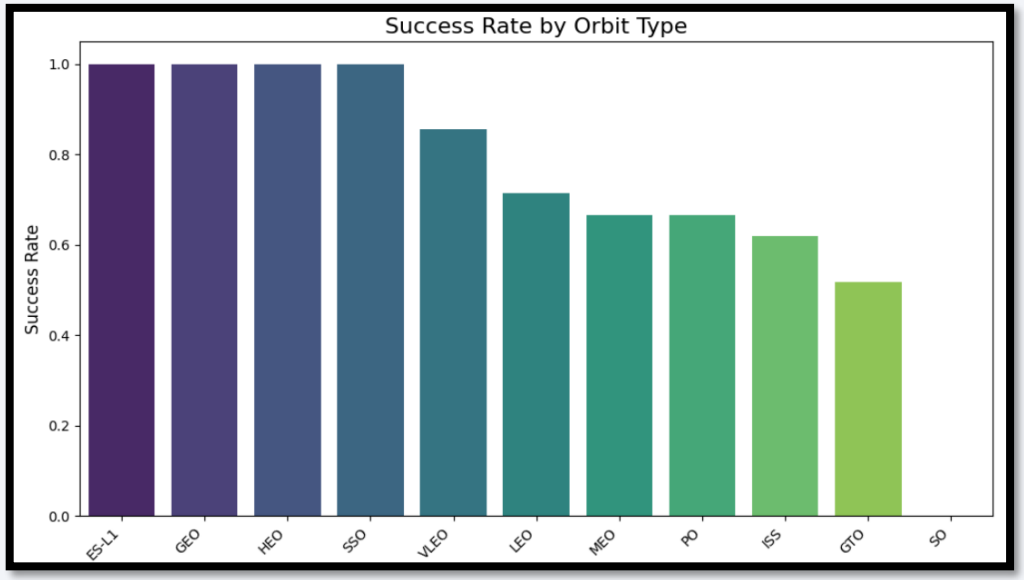
Data Wrangling

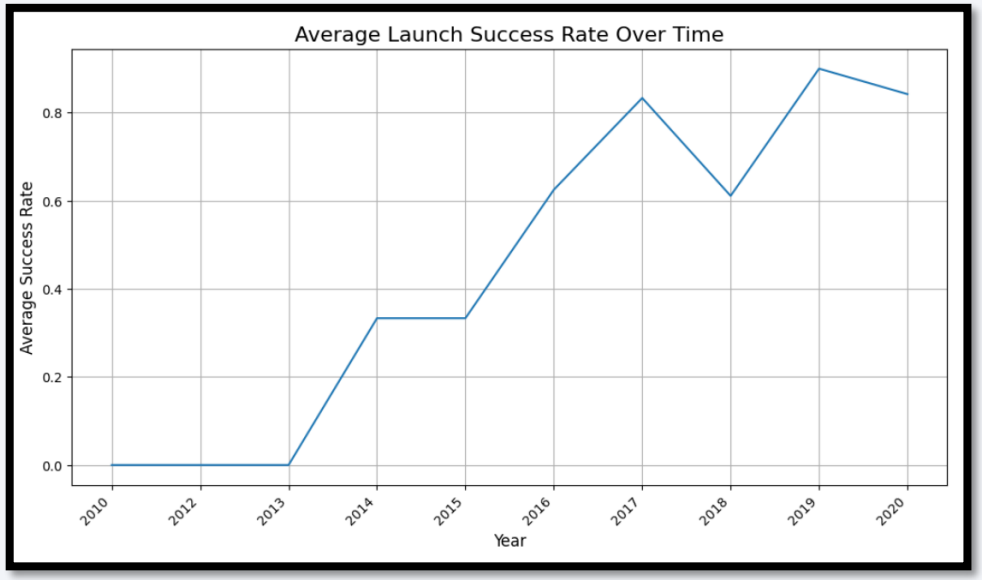
- Describe how data were processed
  -Calculate the number of launches on each site
  -Calculate the number and occurrence of each orbit
  -Calculate the number and occurence of mission outcome of the orbits
  -Create a landing outcome label from Outcome column
- You need to present your data wrangling 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/rohikian/Applied-Data-Science-Capstone.git



EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
  -Those charts indicate the relationship between success rate and other variables.
  -I can select relevant features by using these charts.
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose
  - https://github.com/rohikian/Applied-Data-Science-Capstone.git





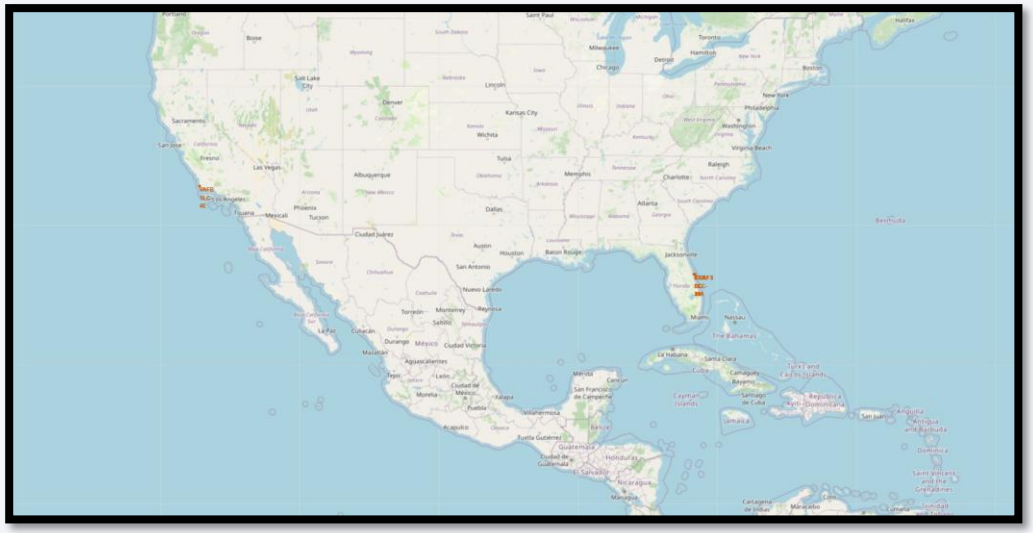
EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose
  - https://github.com/rohikian/Applied-Data-Science-Capstone.git

- The names of the unique launch sites in the space mission are CCAFS LC-40,VAFB SLC-4E,KSC LC-39A,CCAFS SLC-40.
- Launch sites begin with the string 'CCA' include F9 v1.0 B0003, F9 v1.0 B0004, F9 v1.0 B0005, F9 v1.0 B0006, F9 v1.0 B0007.
- The total payload mass carried by boosters launched by NASA (CRS) is 45596.
- The average payload mass carried by booster version F9 v1.1 is 2928.4
- The first successful landing outcome in ground pad was 2015-12-22.
- The names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000 are F9 FT B1022,F9 FT B1026,F9 FT B1021.2,F9 FT B1031.2.
- Mission outcomes are broken down into 1 Failure (in flight), 99 Success and 1 Success (payload status unclear),
- There are 12 booster versions that have carried the maximum payload mass
- In 2015, there were 2 launches with Failure (drone ship).
- Between the date 2010-06-04 and 2017-03-20, the number of Success (drone ship) is the highest.

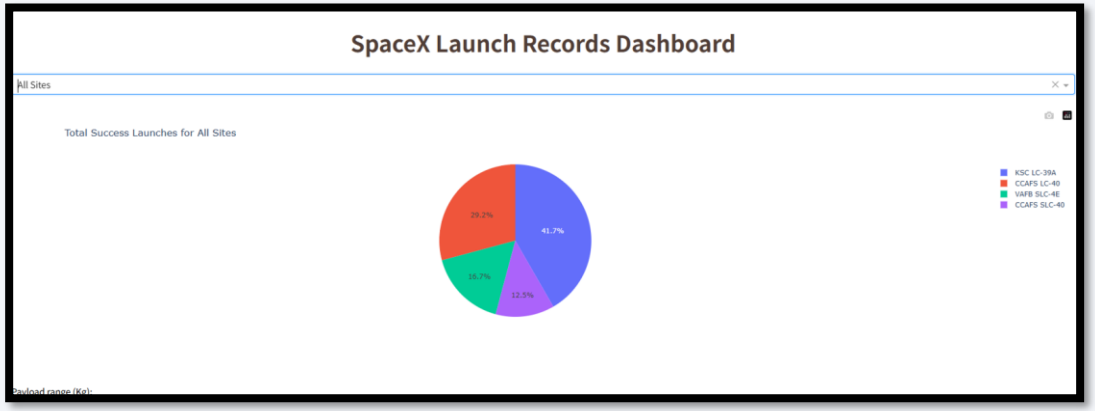
Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
  -I added markers, circles, lines to a folium map.
- Explain why you added those objects
  -Those objects make it easier to understand where the launch sites are, which sites have high success rates and distances to the proximities.
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose
  - https://github.com/rohikian/Applied-Data-Science-Capstone.git



Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
  -I added dropdown menu and Range Slider to a dashboard.
- Explain why you added those plots and interactions
  -These features enable users to understand the results interactively and analyze the real-time data.
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose
  - https://github.com/rohikian/Applied-Data-Science-Capstone.git

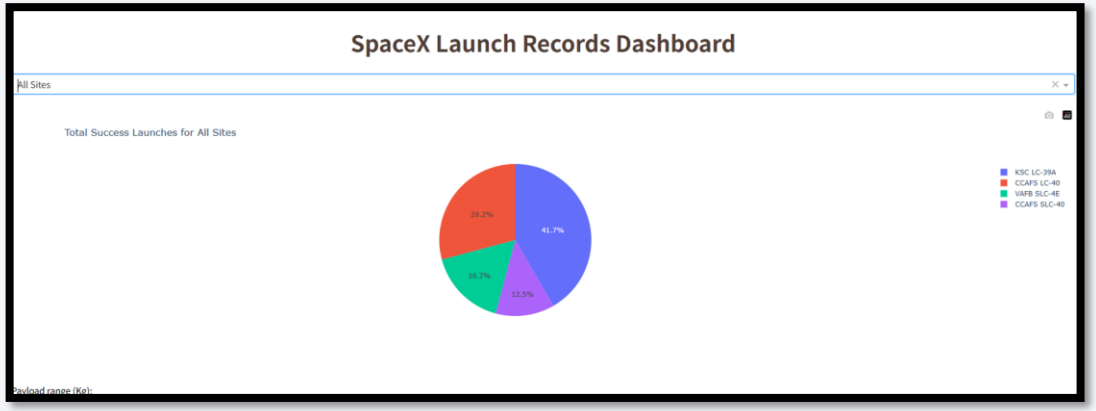


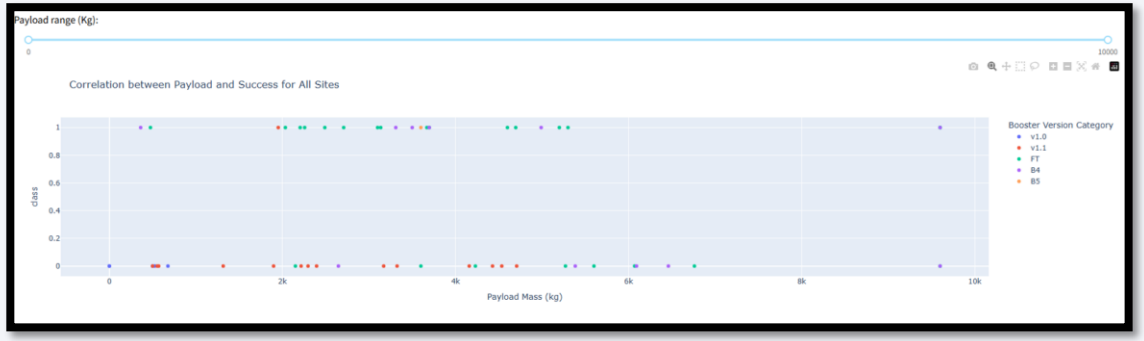
Predictive Analysis (Classification)

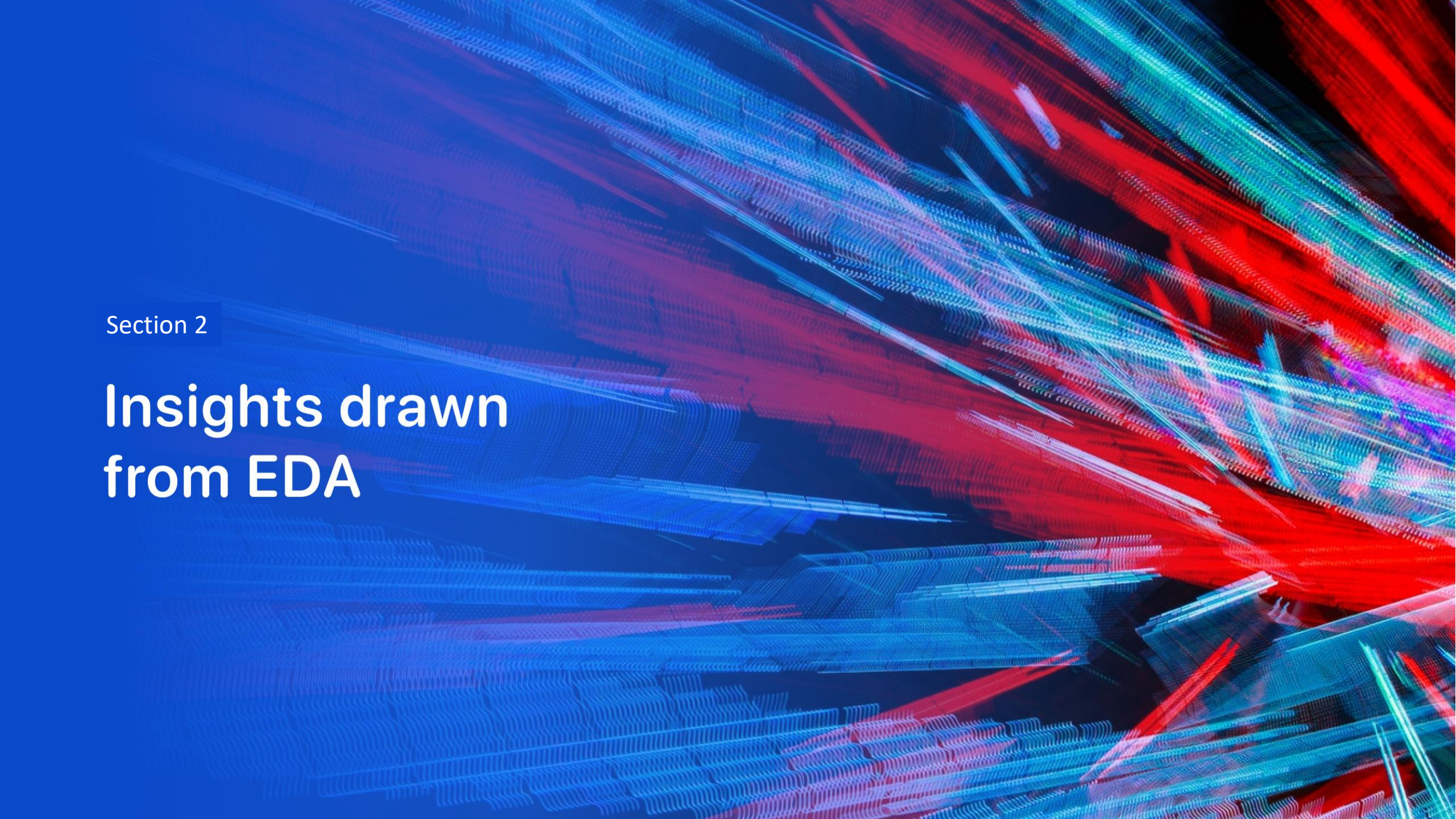
- Summarize how you built, evaluated, improved, and found the best performing classification model
  -Standardize the data, Split into training data and test data, Fitting to the GridSearch
  -Find the method performs best using test data by comparing the accuracies and confusion matrixes.
- You need 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
  - https://github.com/rohikian/Applied-Data-Science-Capstone.git

Results

- Exploratory data analysis results
  -The relevant features are ['FlightNumber', 'PayloadMass', 'Orbit', 'LaunchSite', 'Flights', 'GridFins', 'Reused', 'Legs', 'LandingPad', 'Block', 'ReusedCount', 'Serial'].
- Interactive analytics demo in screenshots
  -See below
- Predictive analysis results
  - The accuracies of four methods are all the same, 0.833.
  - Any method of this lab could be used to predict the chance of landing.



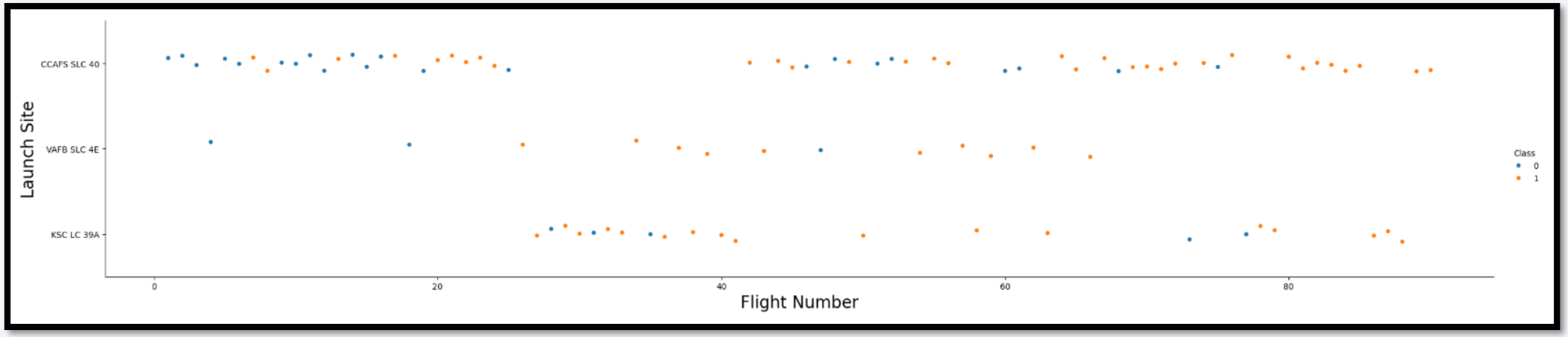



Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

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

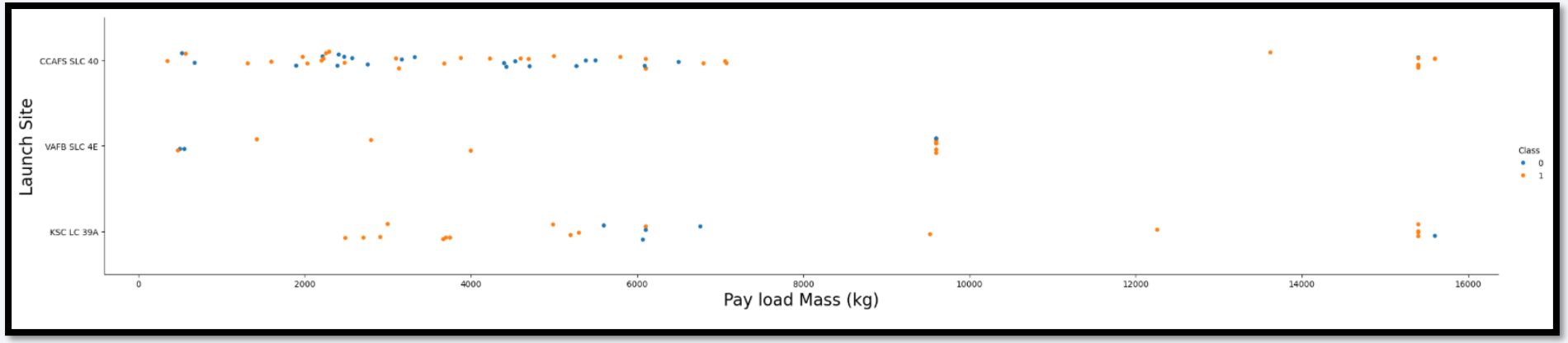
  As the flight number increases, the first stage is more likely to land successfully, regardless of launch sites.



Payload vs. Launch Site

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

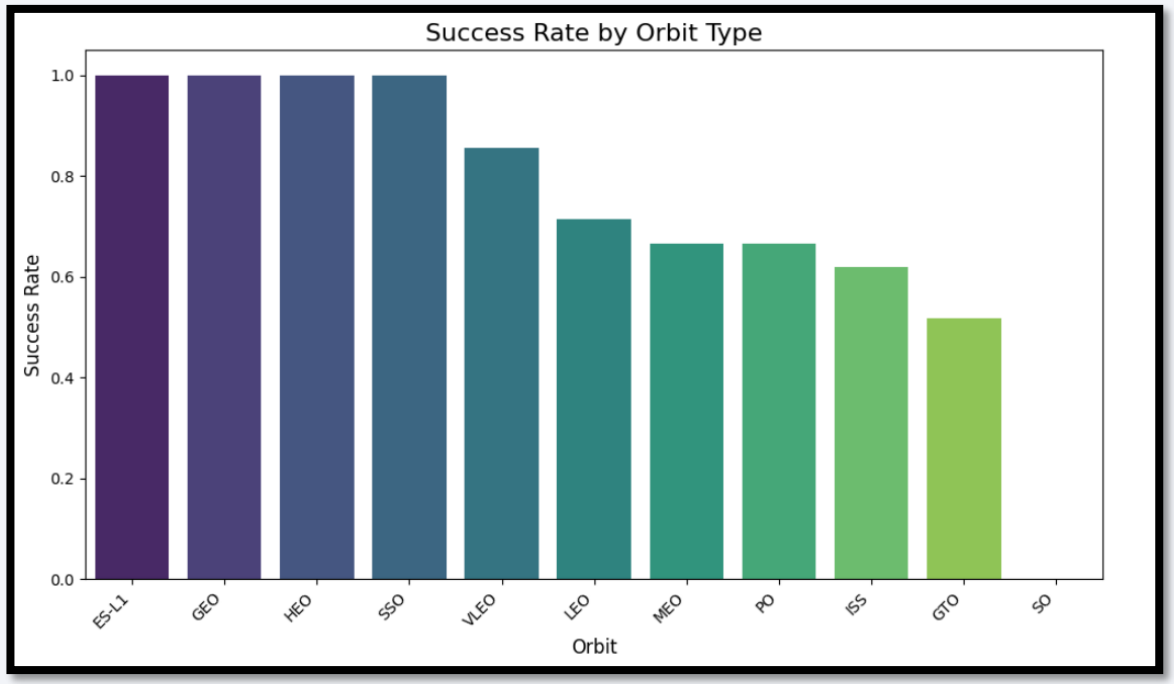
  There seems to be no significant relationship between Payload and Launch site.



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

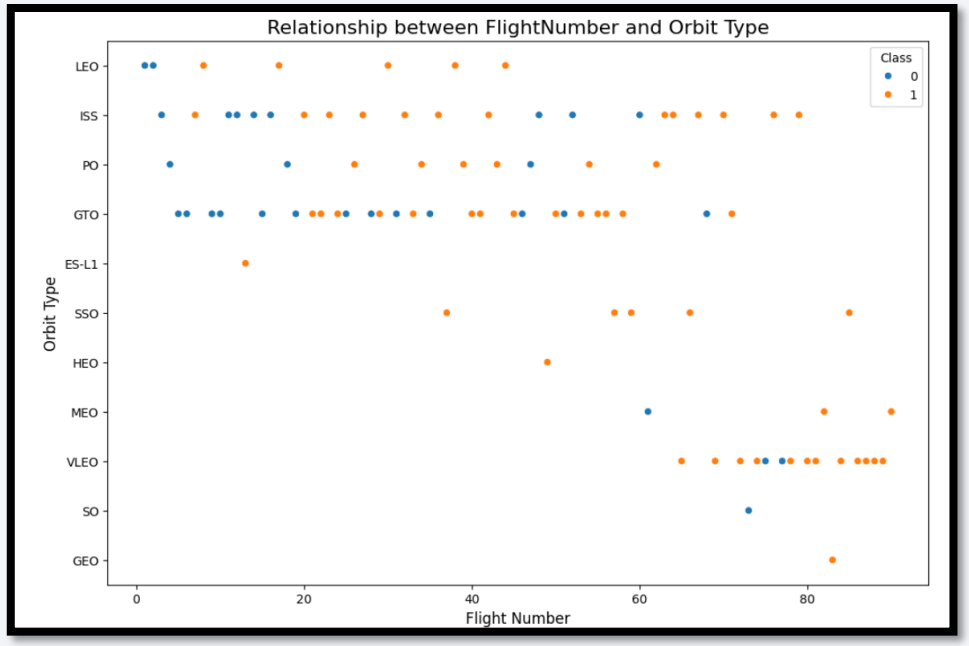
ES-L1, GEO, HEO and SSO have the highest success rates.



Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type
- Show the screenshot of the scatter plot with explanations

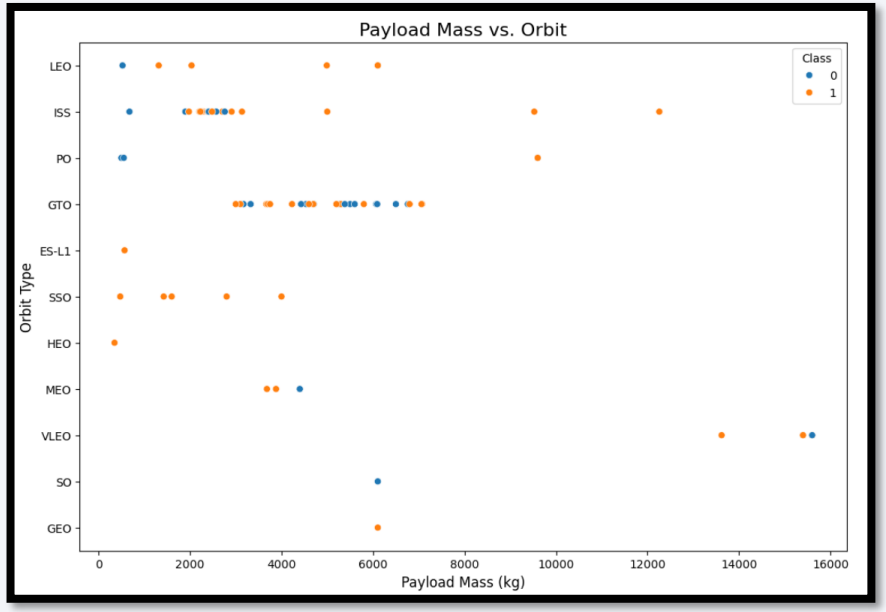
  The results are confusing.
  In the LEO orbit, success seems to be related to the number of flights.
  Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.



Payload vs. Orbit Type

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

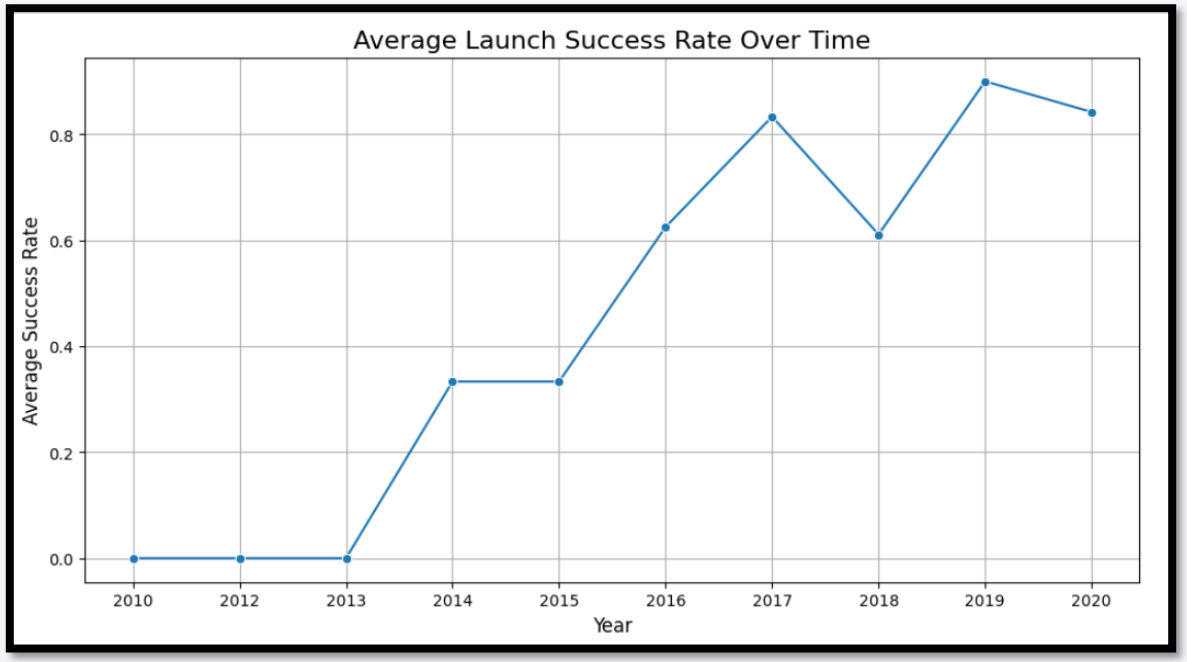
  The results are confusing.
  With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
  However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.



Launch Success Yearly Trend

- Show a line chart of yearly average success rate
- Show the screenshot of the scatter plot with explanations

  Obviously, the success rate increases as the time goes by.



All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

There are four sites.

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

  Launch sites begin with the string 'CCA' include F9 v1.0 B0003, F9 v1.0 B0004, F9 v1.0 B0005, F9 v1.0 B0006, F9 v1.0 B0007.

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

The result is 45596.

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

  The result is 2928.4.

```
%sql SELECT AVG(Payload_Mass__kg_) FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1';
 * sqlite:///my_data1.db
Done.
AVG(Payload_Mass__kg_)
2928.4
```

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

The date is December 22th, 2015.

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
- Present your query result with a short explanation here

  The names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000 are F9 FT B1022,F9 FT B1026,F9 FT B1021.2,F9 FT B1031.2.

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

  Mission outcomes are broken down into 1 Failure (in flight), 99 Success and 1 Success (payload status unclear).

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

  There are 12 booster versions that 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

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 2015, there were 2 launches with Failure (drone ship).

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
- Present your query result with a short explanation here

  Between the date 2010-06-04 and 2017-03-20, the number of Success (drone ship) is the highest.

Landing_Outcome	count_outcome
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

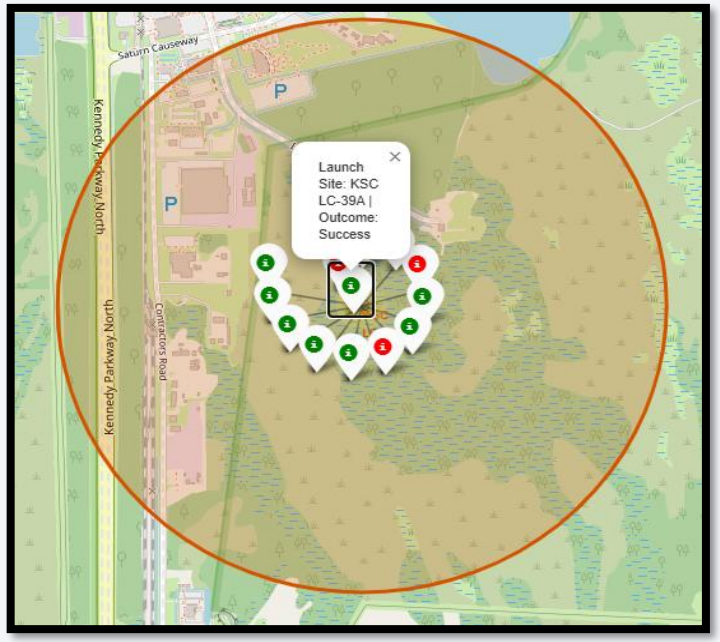
Launch Sites Proximities Analysis

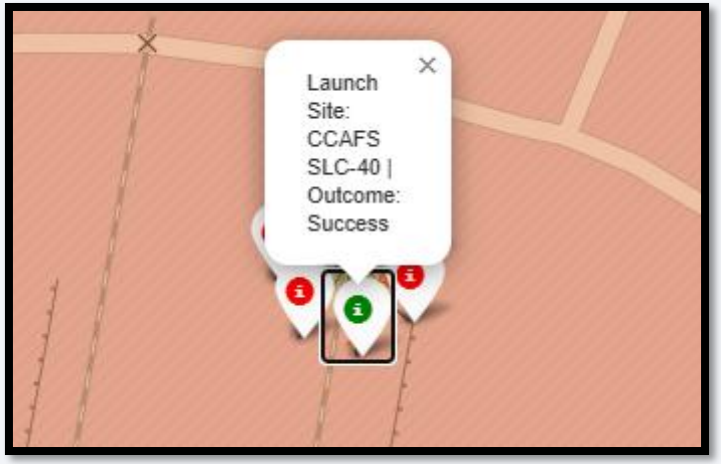
All launch sites

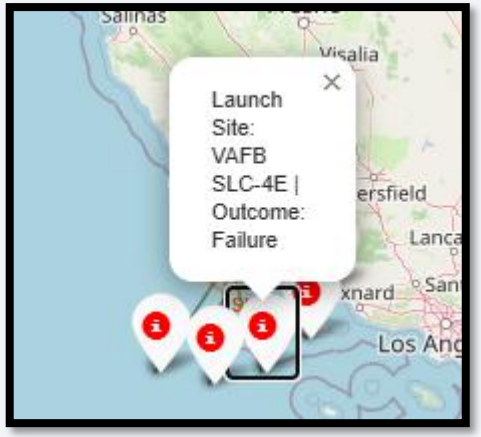
- All launch sites are located near the coast. Also, they are far away from cities.

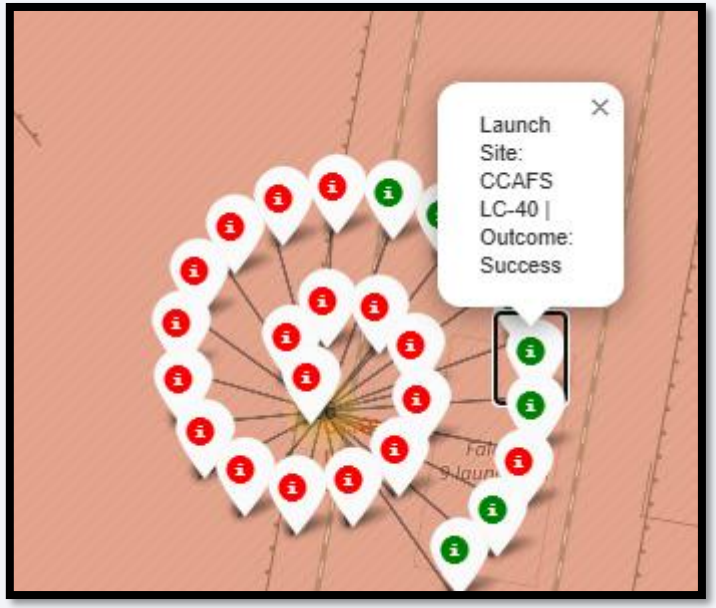
The success/failed launches for each site

- KSC LC-39A has relatively high success rate.



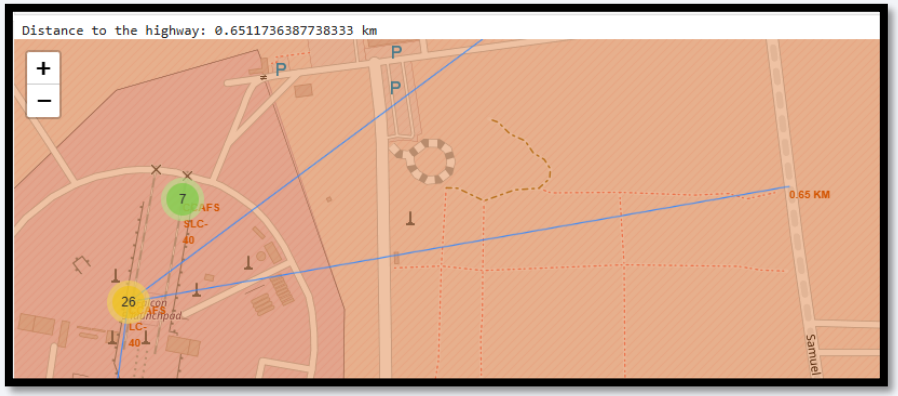


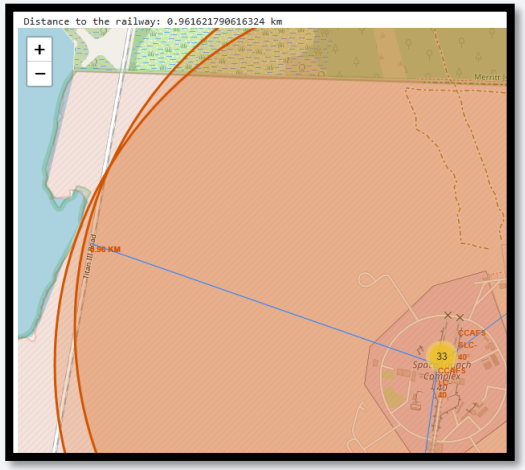


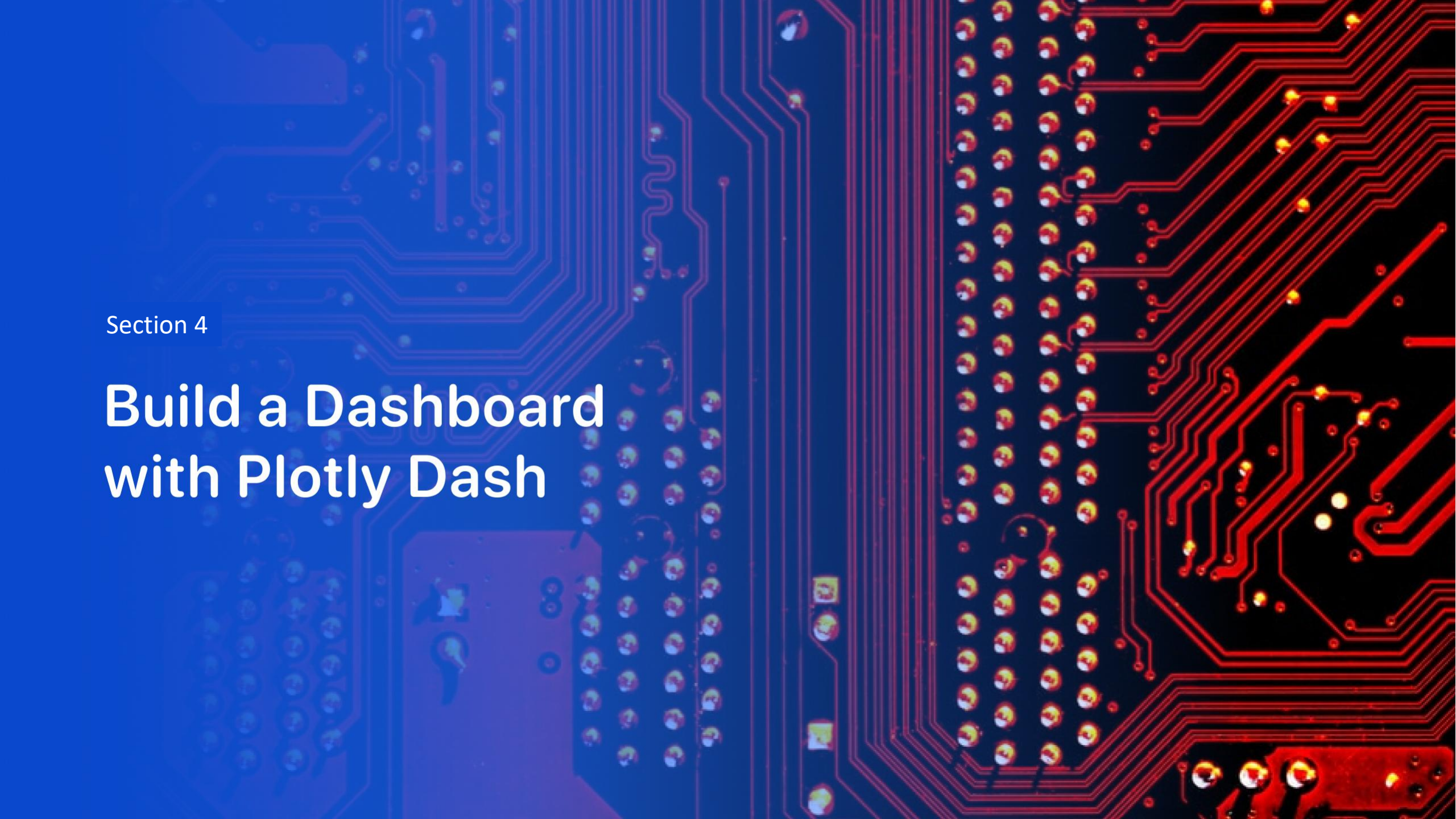


The distances between a launch site to its proximities

- The launch site are close to places including railway, highway, coastline. But it is far away from a city.

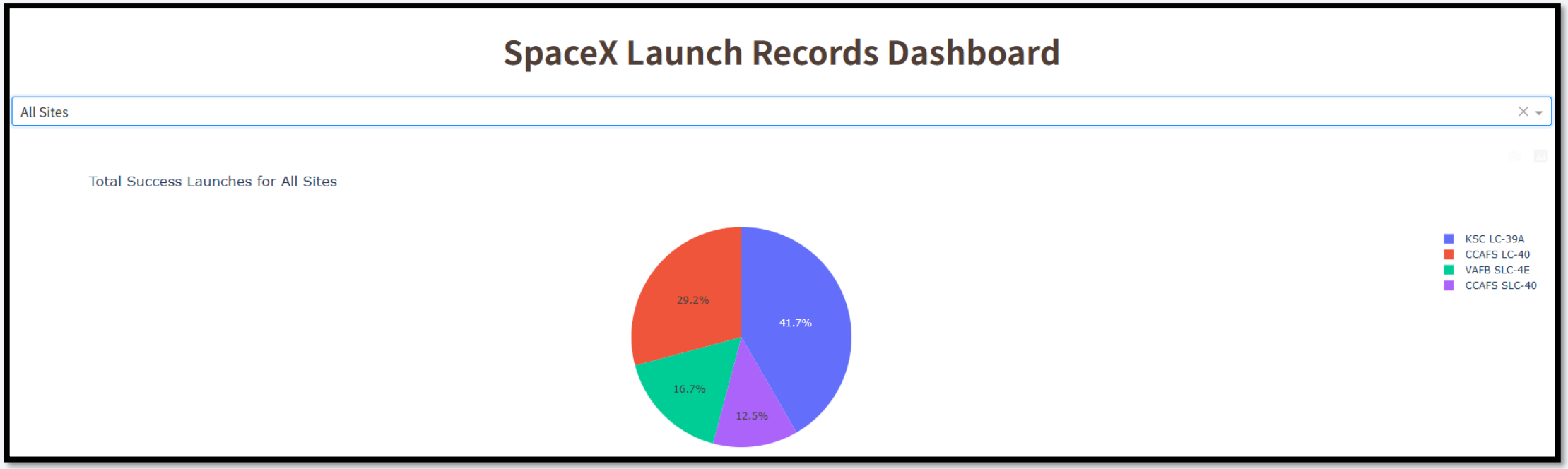





Section 4
Build a Dashboard with Plotly Dash

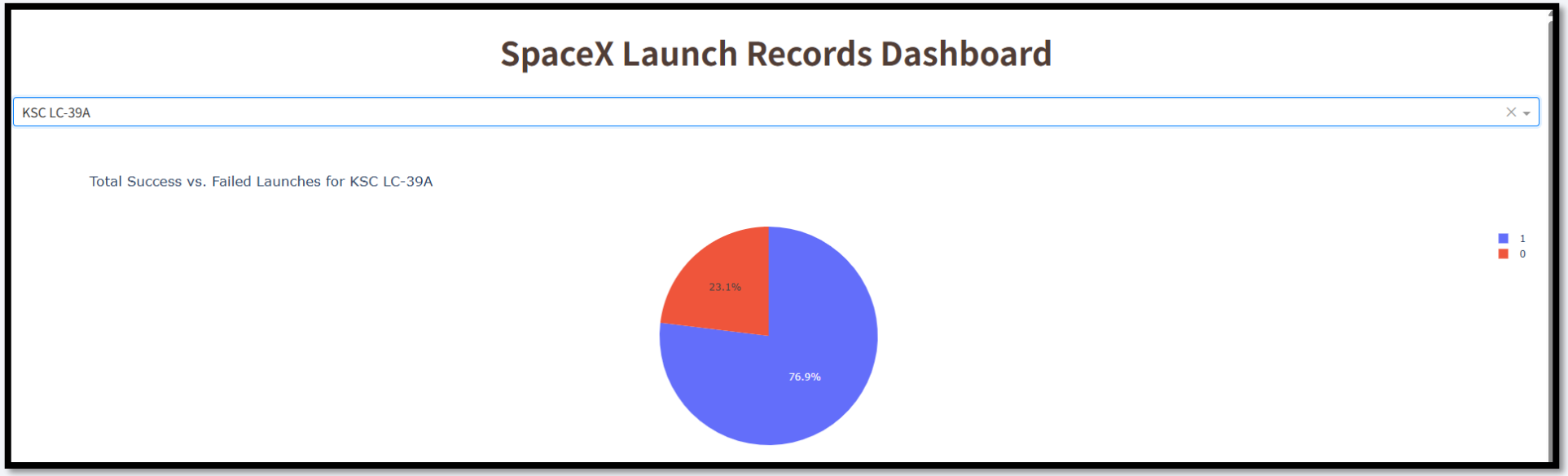
Launch success count for all sites

- KSC LC-39A accounts for 41.7% of total success launches.



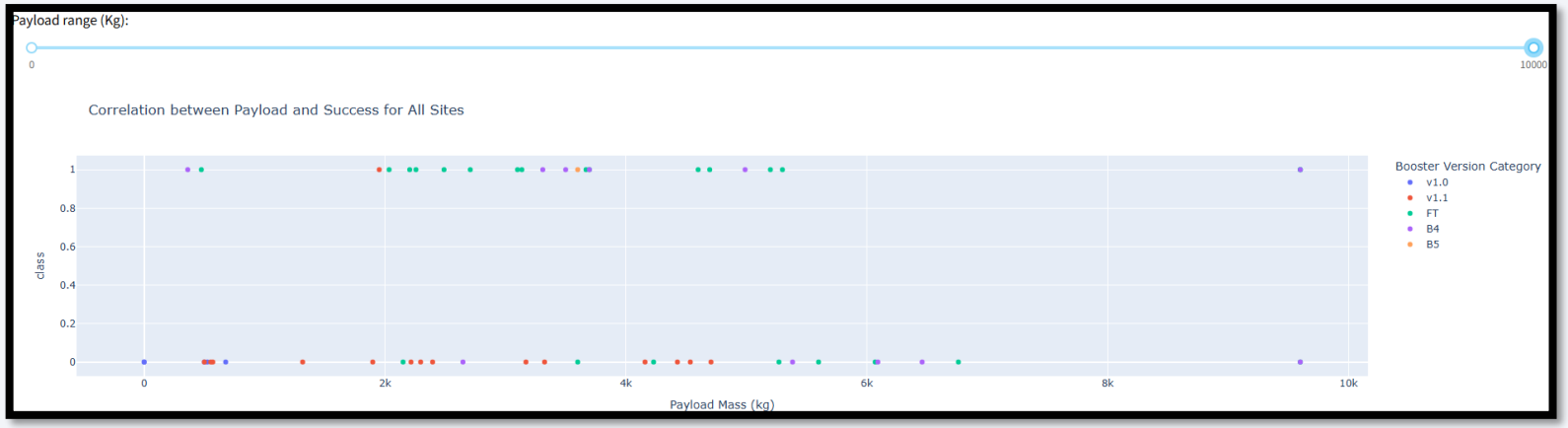
The launch site with highest launch success ratio

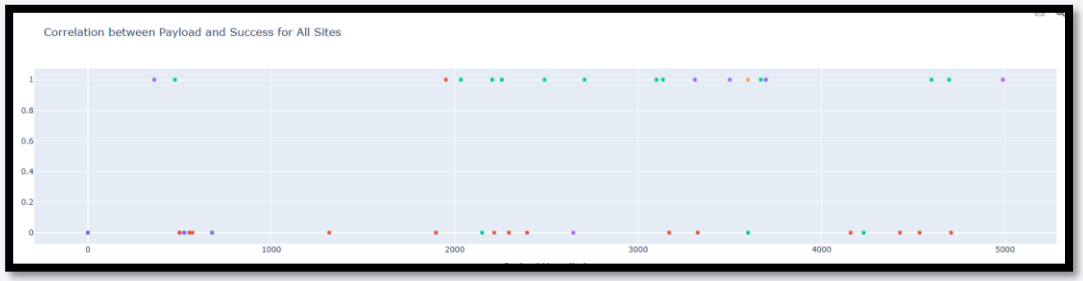
- KSC LC-39A has the highest launch success ratio. The ratio is larger than three quarters.

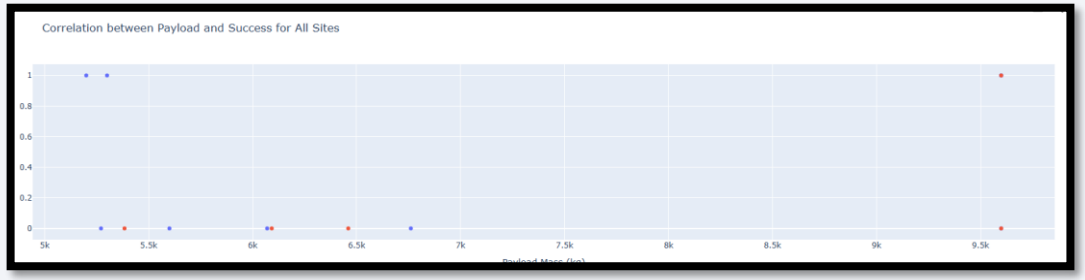


Payload vs. Launch Outcome scatter plot for all sites

- More than 5k, there are only two types of booster. On the other hand, less than 5k, there are five types of booster version.
- FT seems to have higher success rate.





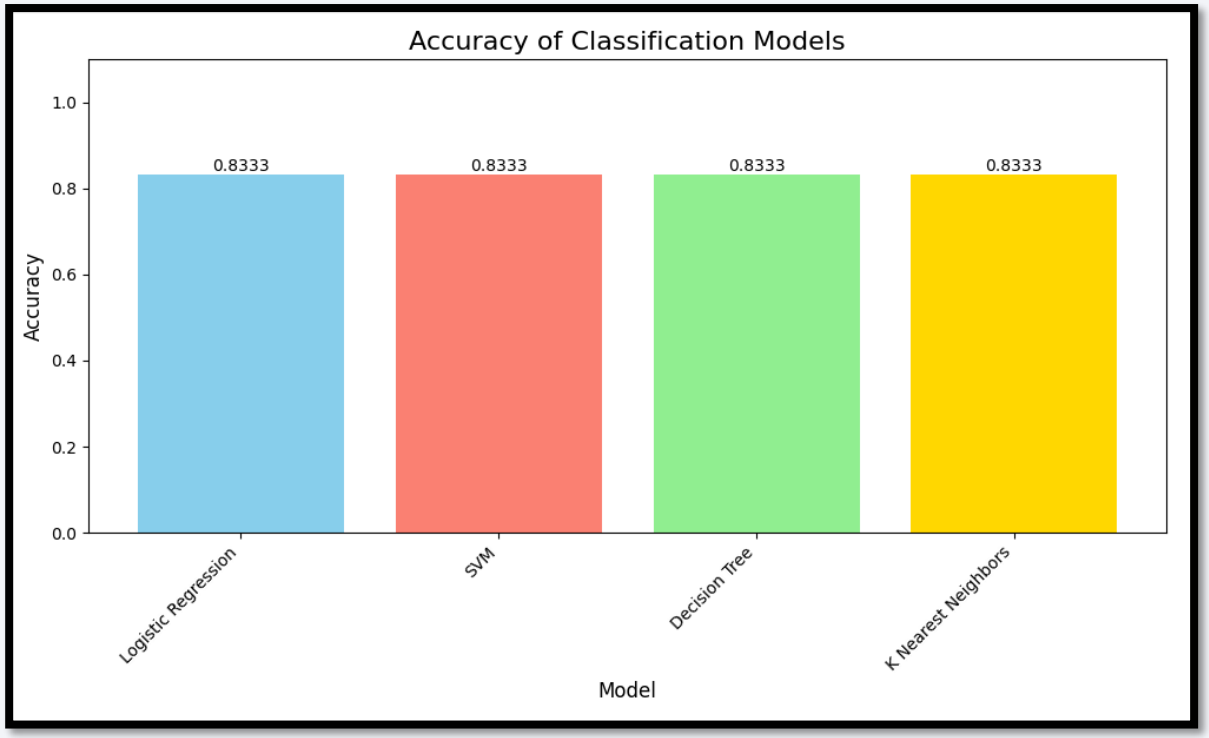


Section 5

Predictive Analysis (Classification)

Classification Accuracy

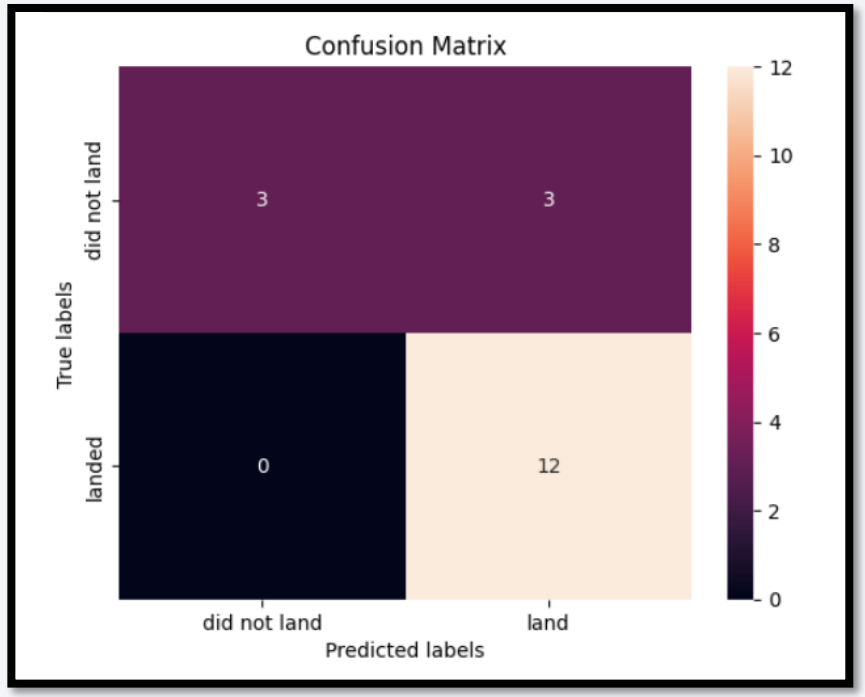
- Visualize the built model accuracy for all built classification models, in a bar chart
- Find which model has the highest classification accuracy
  -The accuracies of four methods are all the same, 0.833. However, the accuracy for train data are different(See appendix).



Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation
  -Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the problem is false positives.
  -True Postive - 12 (True label is landed, Predicted label is also landed)

  -False Postive - 3 (True label is not landed, Predicted label is landed)

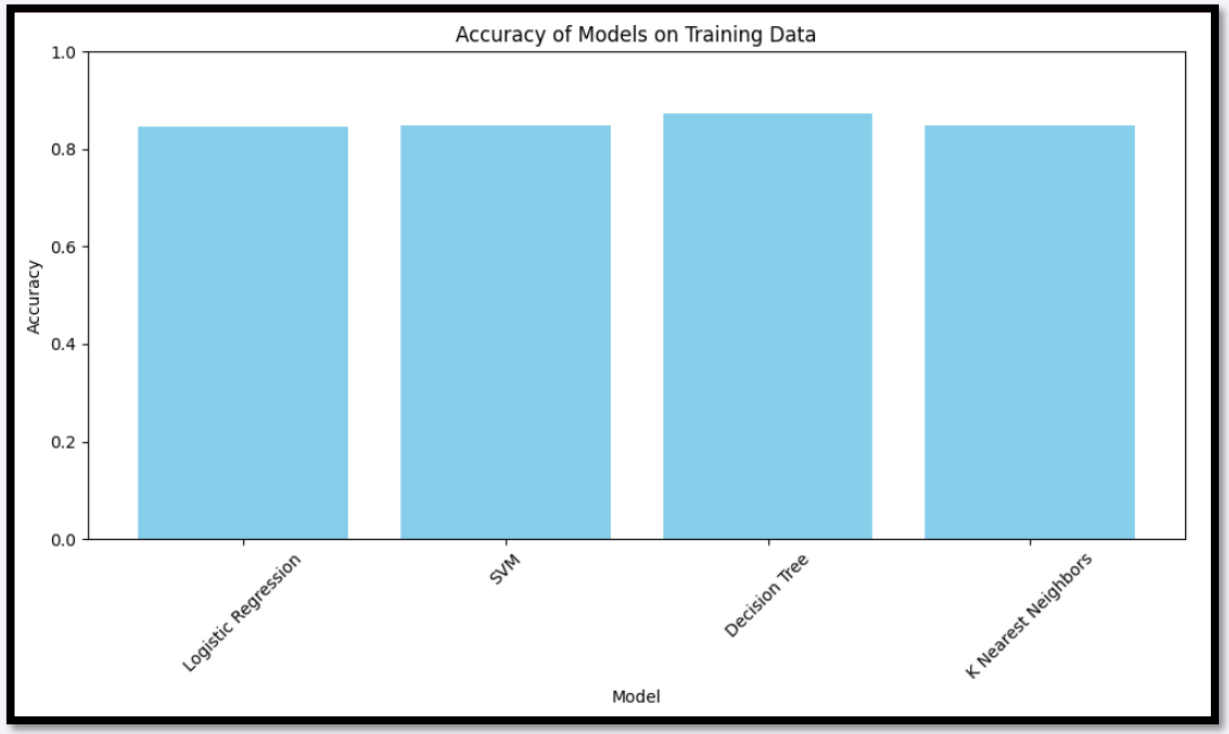


Conclusions

- Point 1
  - The accuracies of four methods are all the same, 0.833.
- Point 2
  -Any method of this lab could be used to predict the chance of landing, as they all have high accuracies.
- Point 3
  -However, if you want to increase the probability with even slight difference, Classification Trees might be appropriate to use.

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
  -Accuracy of training data(Refer to pp43)



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
-Classification report of four models(Refer to pp43)

```
Classification Report for Logistic Regression:
              precision    recall  f1-score   support

           0       1.00      0.50      0.67         6
           1       0.80      1.00      0.89        12

    accuracy                           0.83        18
   macro avg       0.90      0.75      0.78        18
weighted avg       0.87      0.83      0.81        18

Classification Report for SVM:
              precision    recall  f1-score   support

           0       1.00      0.50      0.67         6
           1       0.80      1.00      0.89        12

    accuracy                           0.83        18
   macro avg       0.90      0.75      0.78        18
weighted avg       0.87      0.83      0.81        18

Classification Report for Decision Tree:
              precision    recall  f1-score   support

           0       1.00      0.50      0.67         6
           1       0.80      1.00      0.89        12

    accuracy                           0.83        18
   macro avg       0.90      0.75      0.78        18
weighted avg       0.87      0.83      0.81        18

Classification Report for K Nearest Neighbors:
              precision    recall  f1-score   support

           0       1.00      0.50      0.67         6
           1       0.80      1.00      0.89        12

    accuracy                           0.83        18
   macro avg       0.90      0.75      0.78        18
weighted avg       0.87      0.83      0.81        18
```

Thank you!