IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Hasan Al-Nimr
17/02/2023

Outline

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

Executive Summary

- Summary of methodologies
  - Collecting datasets using space x API and use web scraping
  - Exploratory data analysis and complete EDA with visualization.
  - Visualization for the data collected (Dashboards) with Folium
  - Develop Machine learning prediction of the success for the 1st stage.

- Summary of all results
  - Interactive Dashboards
  - Data Analysis Results
  - Prediction Model

Introduction

- Project background and context

we need to 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.

- Problems you want to find answers
  - What are the factors mostly results in success landing?
  - How to predict success of a given flight?

Section 1

Methodology

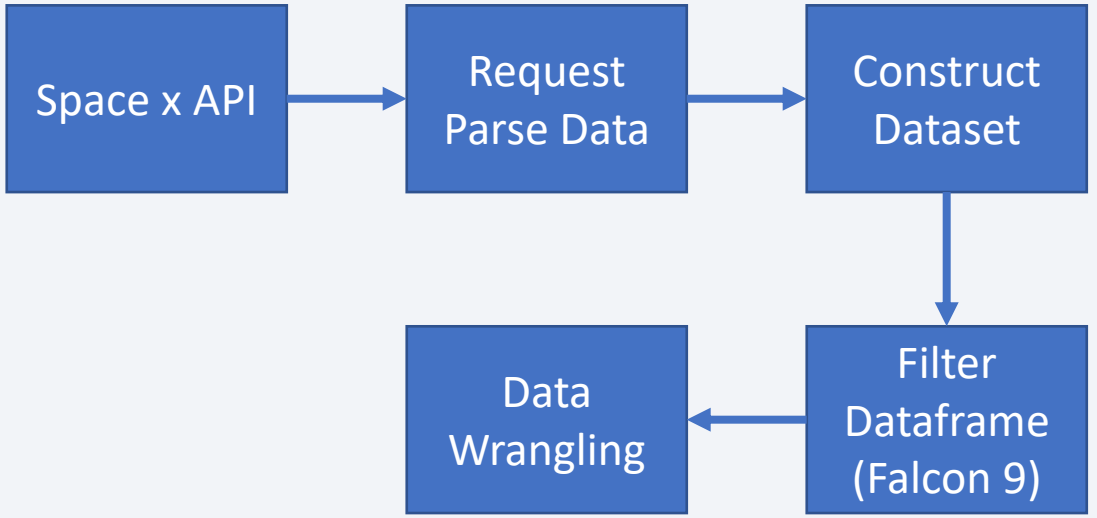
Methodology

Executive Summary

- Data collection methodology:
  - Space X public API datasets
  - WebScraping
- Perform data wrangling
  - We investigated the data targeted creating the needed dictionary for the columns and understanding the different terms within the data received.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Splitting the data into train and test data then determine the model with best accuracy (Logistic regression, support vector, decision tree classifier and K-nearest neighbor.

Data Collection

- How data sets were collected?
  - Collect the needed APIs
  - Get the Booster, Lunchpad, Payload and the Core (outcome of the landing, the type of the landing, number of flights with that core, whether gridfins were used, wheter the core is reused, wheter legs were used, the landing pad used, the block of the core which is a number used to seperate version of cores, the number of times this specific core has been reused, and the serial of the core info)
  - Append the results back to the data.
  - Deal with missing data by using the mean and replace it where needed (Replace Null Values)



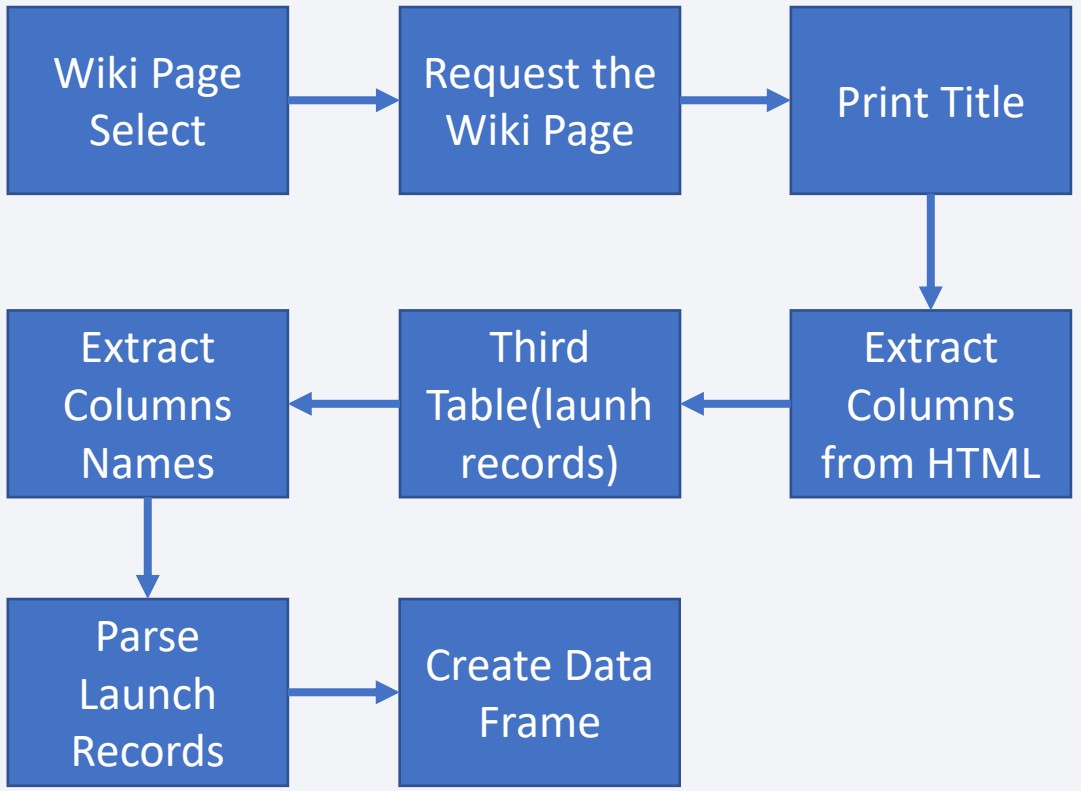
Data Collection – SpaceX API

- https://github.com/jupyter123321/Space-X-ipynp-Files/blob/main/Python%20Space%20X%20Model%20AI(2)%20-%20API%20Calls.ipynb

Place your flowchart of SpaceX API calls here

Data Collection - Scraping

- Identify the pages we are targeting for Webscraping.
- Request the Falcon9 Page
- Extract the columns names from the table.
- Create Data by Parsing the Launch HTML tables
- https://github.com/jupyter12332 1/Space-X-ipynp-Files/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- After data being collected from the different APIs we must deal with the missing values in the landingpad column which include some none values.
- We will calculate the mean for the payload mass using the .mean function and replace it where we have the null values in the data.

- https://github.com/jupyter123321/Space-X-ipynp-Files/blob/main/Python%20Space%20X%20Model%20AI(2)%20-%20API%20Calls.ipynb

EDA with Data Visualization

- Flight Number Vs. PayloadMass (We see as flight number increase the first stage is more likel;y land successful the pay load mass is important it seems the more massive the payload the less likely first stage will return)
- Flight Number vs. Launches
- Payloadmass vs. Launches
- Orbit vs. Success Rate
- Flight Number vs. Orbit
- Payload vs. Orbit

- https://github.com/jupyter123321/Space-X-ipynp-Files/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

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

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
- https://github.com/jupyter123321/Space-X-ipynp-Files/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

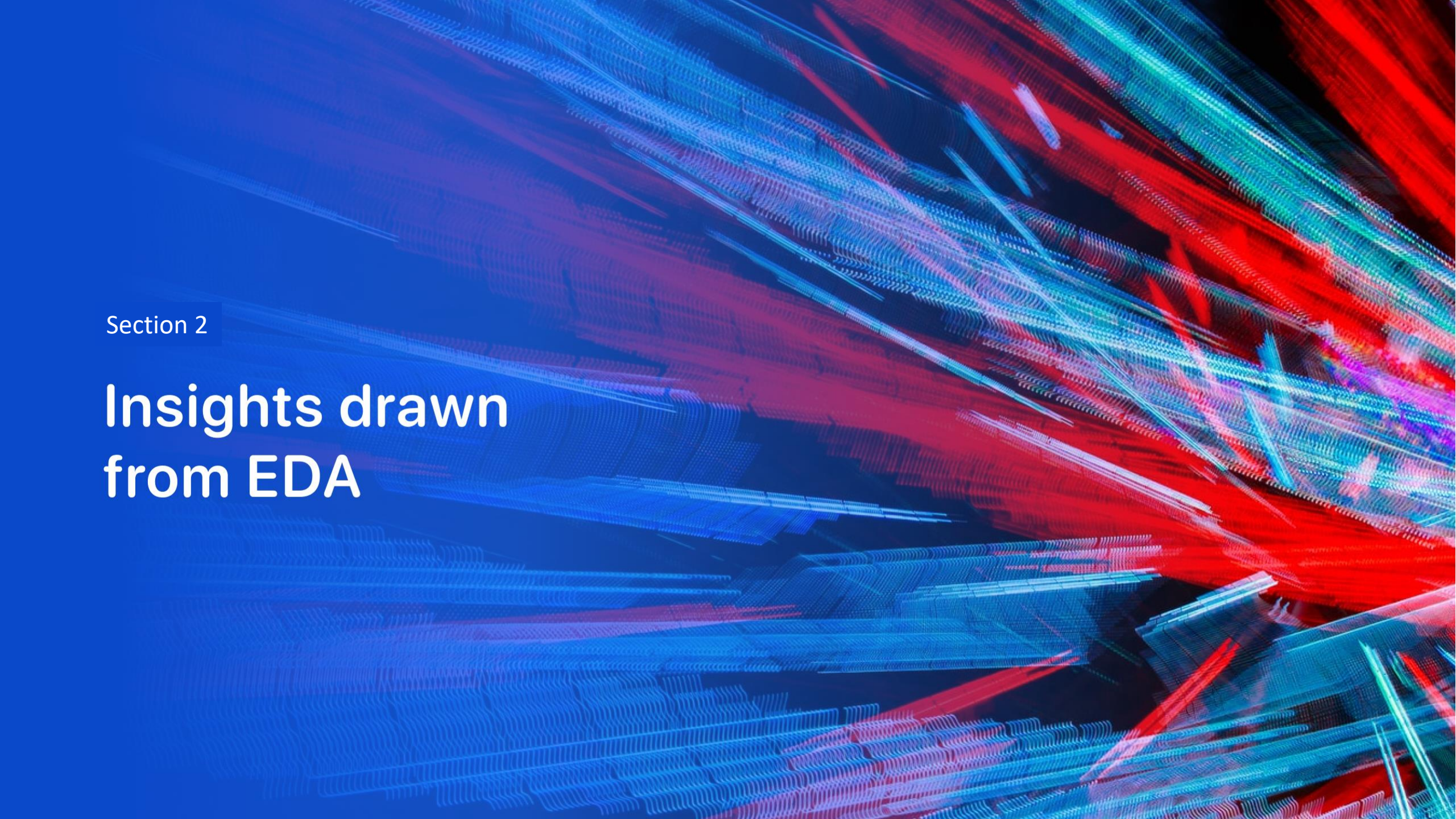
- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

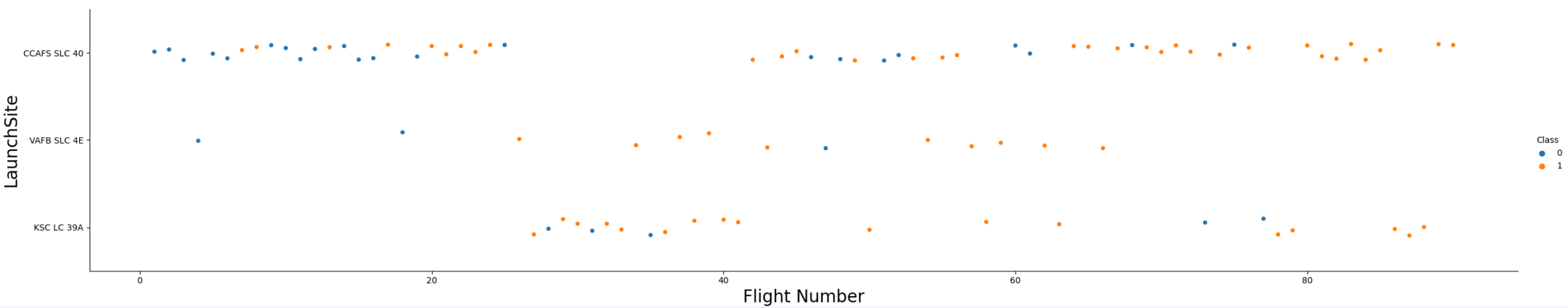
- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

Results

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

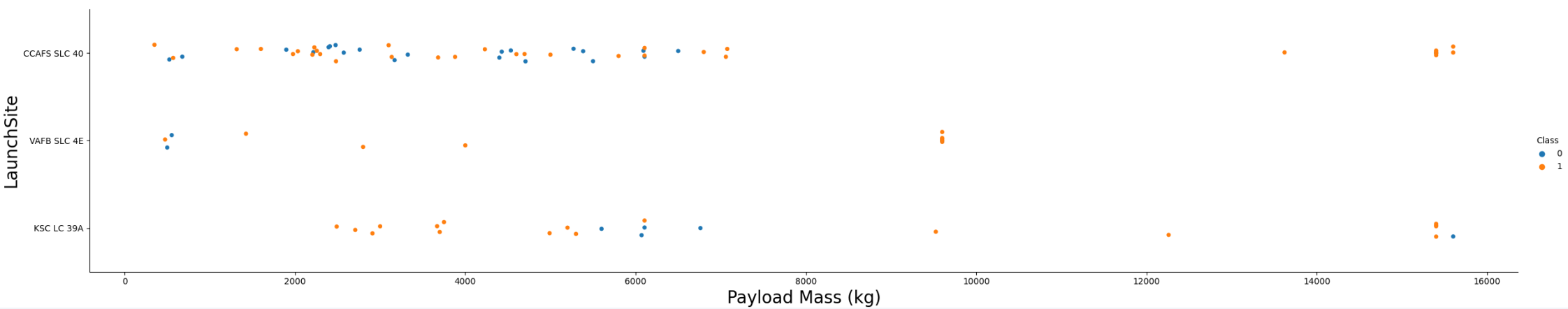

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



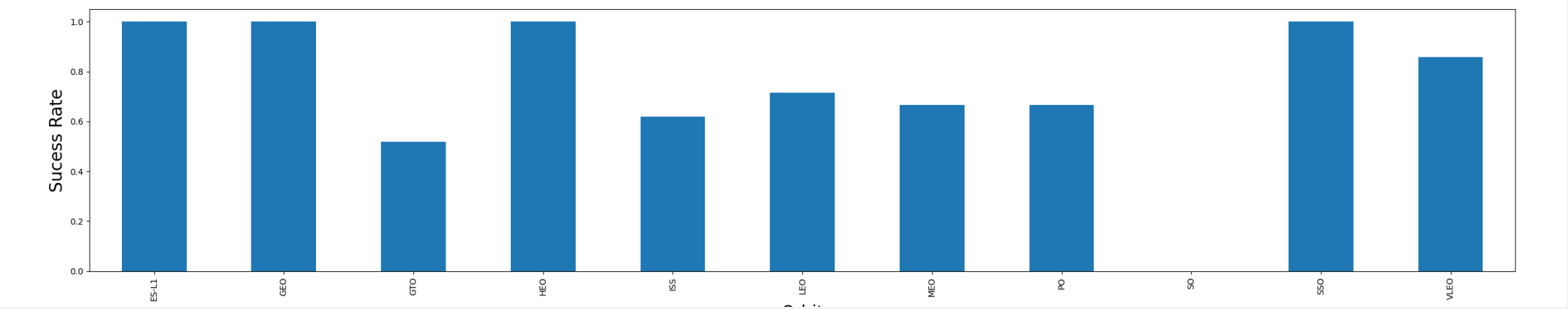
- The sequence of the flight number with the success rate showing increments in the success of stage 1 with the increased number of tests.

Payload vs. Launch Site



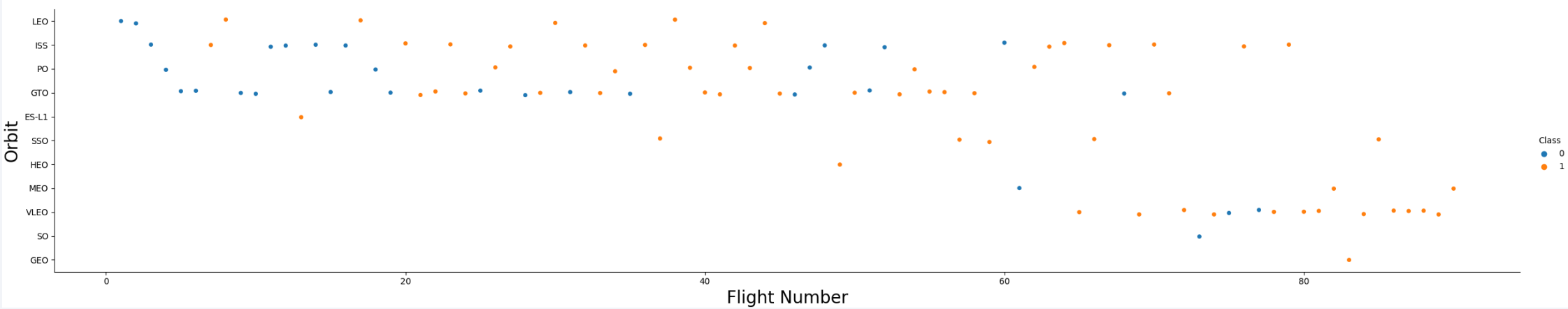
- As payload mass increase the success rate of stage 1 increase on the sites.

Success Rate vs. Orbit Type



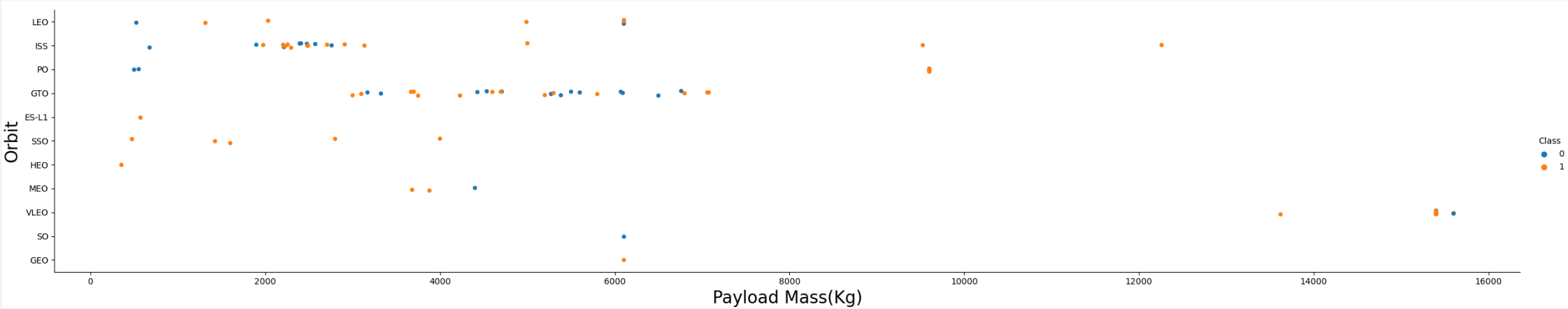
- Show the screenshot of the scatter plot with explanations

Flight Number vs. Orbit Type



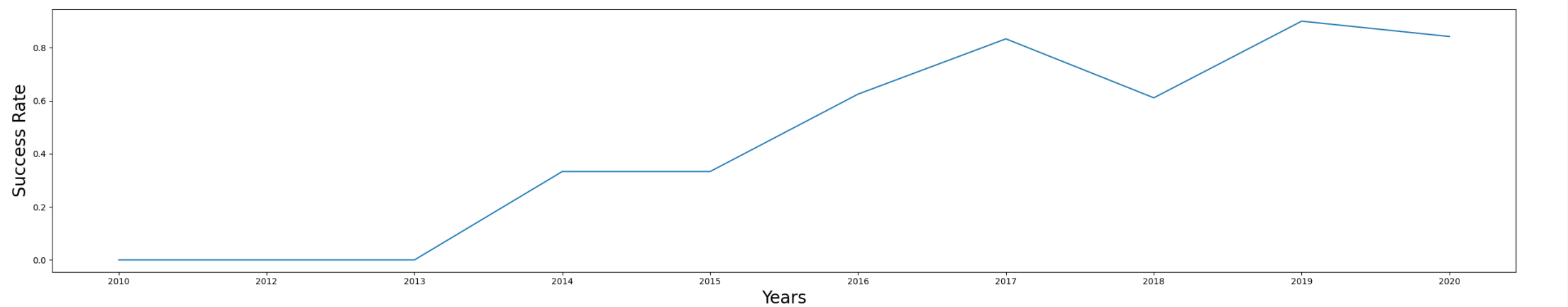
- Show the screenshot of the scatter plot with explanations

Payload vs. Orbit Type



- Show the screenshot of the scatter plot with explanations

Launch Success Yearly Trend



- 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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here

Total Payload Mass

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

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

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

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

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

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

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

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

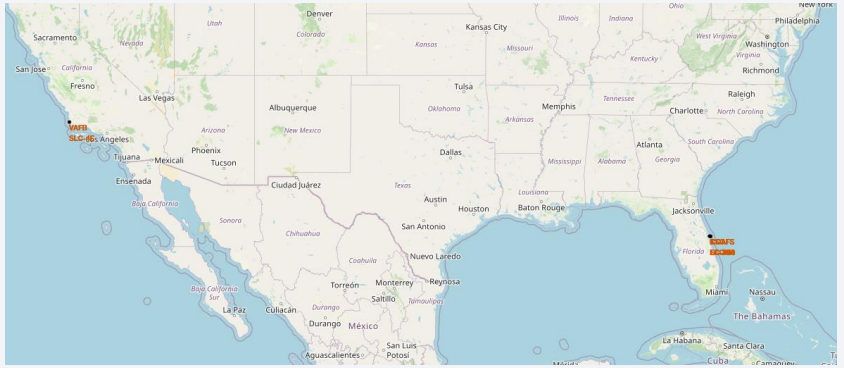
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

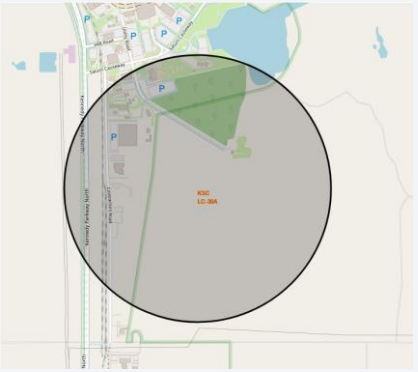
- Present your query result with a short explanation here


Section 3
Launch Sites
Proximities Analysis

Space-X Lunch Sites

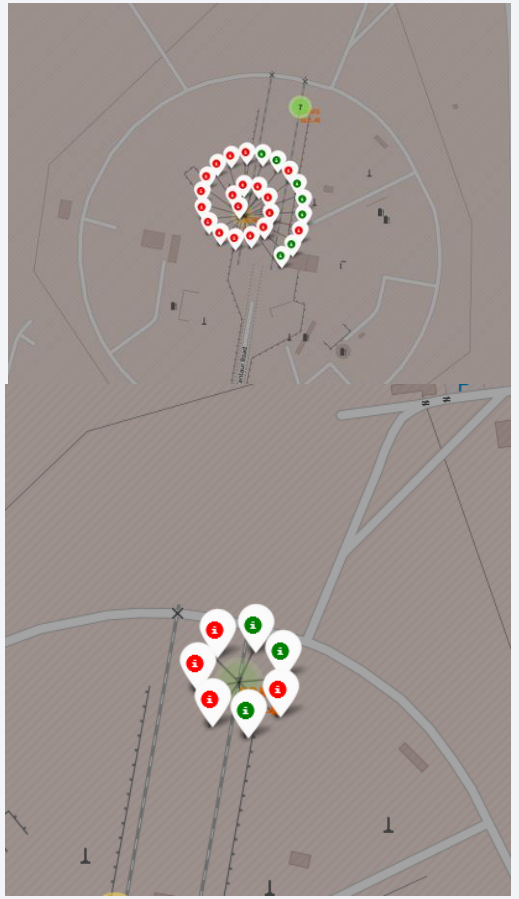
- Sites are close to the coast side of USA.



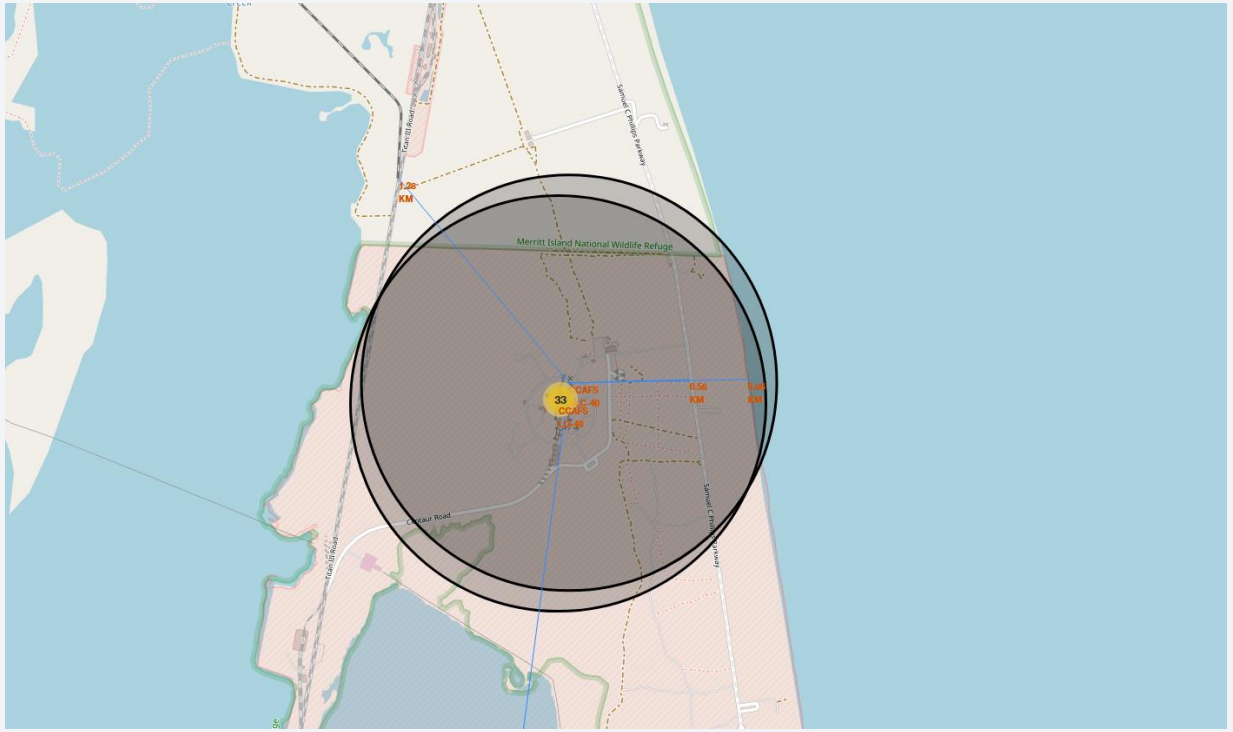


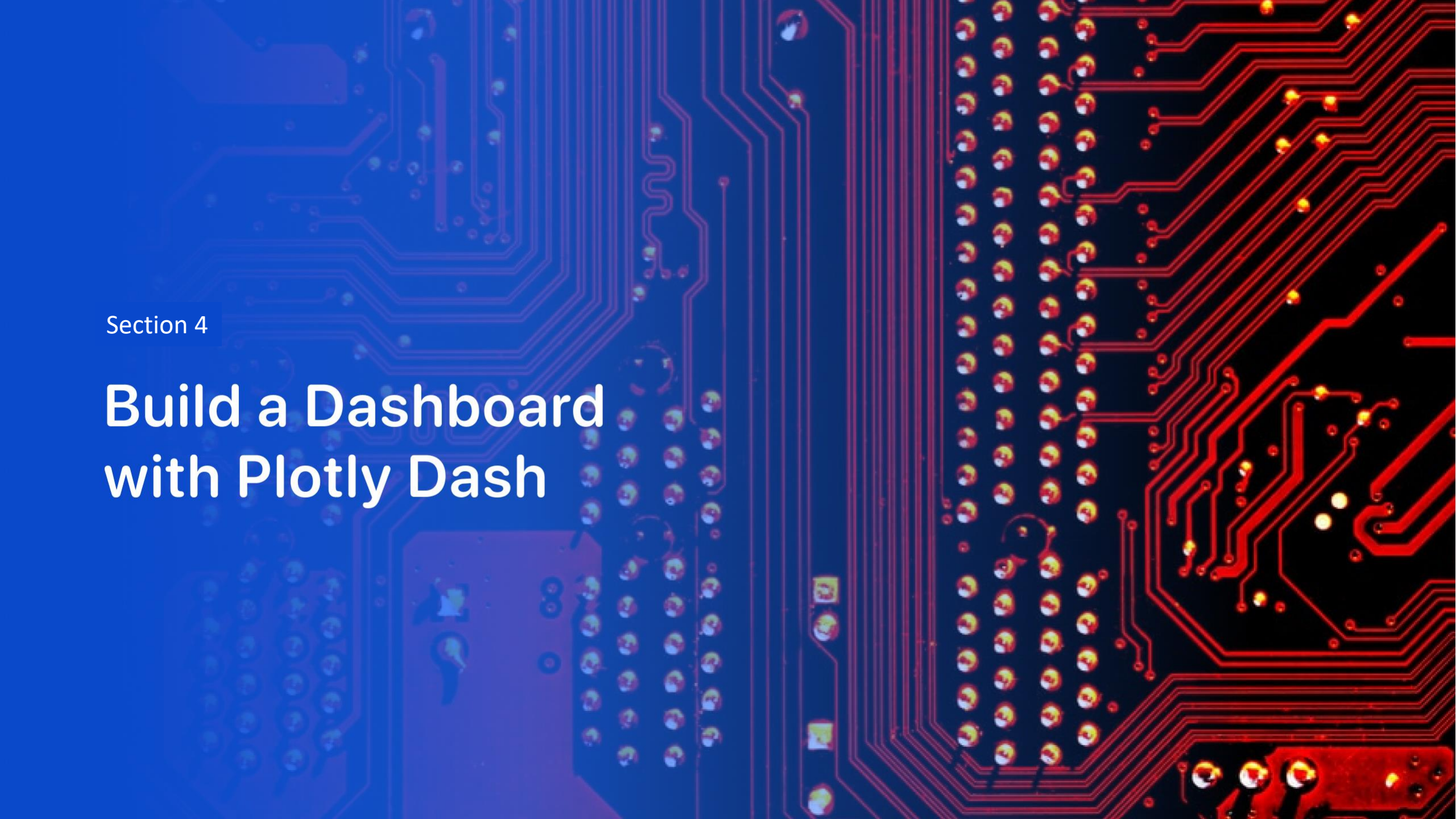
Space X Success Ratios \ Sites

- Green markers indictae a success return
- Red markers indicate a failure return
- KSC LC-39A has higher success rates.



Lunch Sites Proximities




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

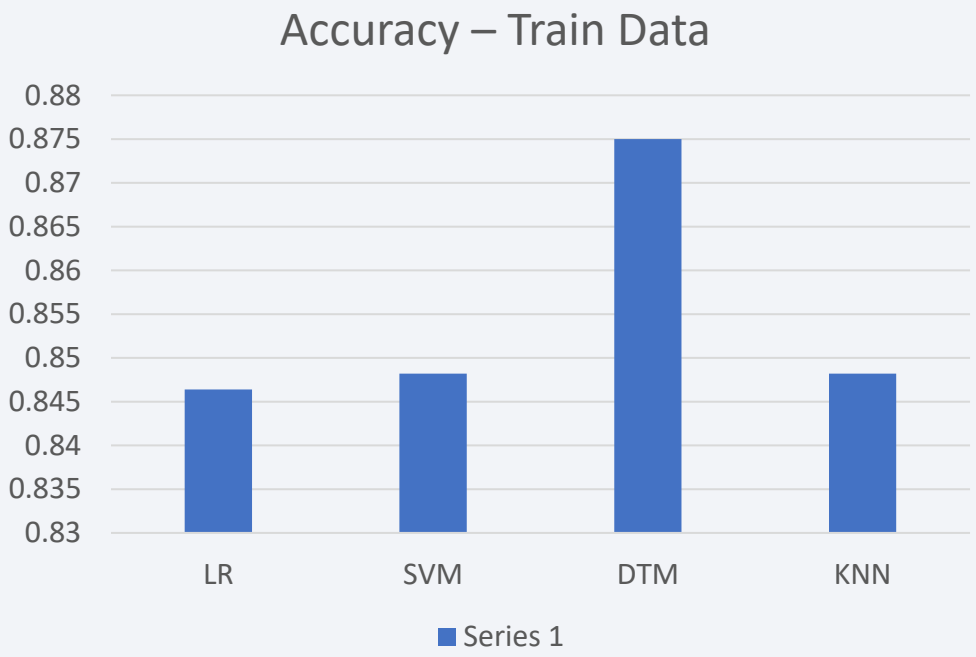
<Dashboard Screenshot 3>

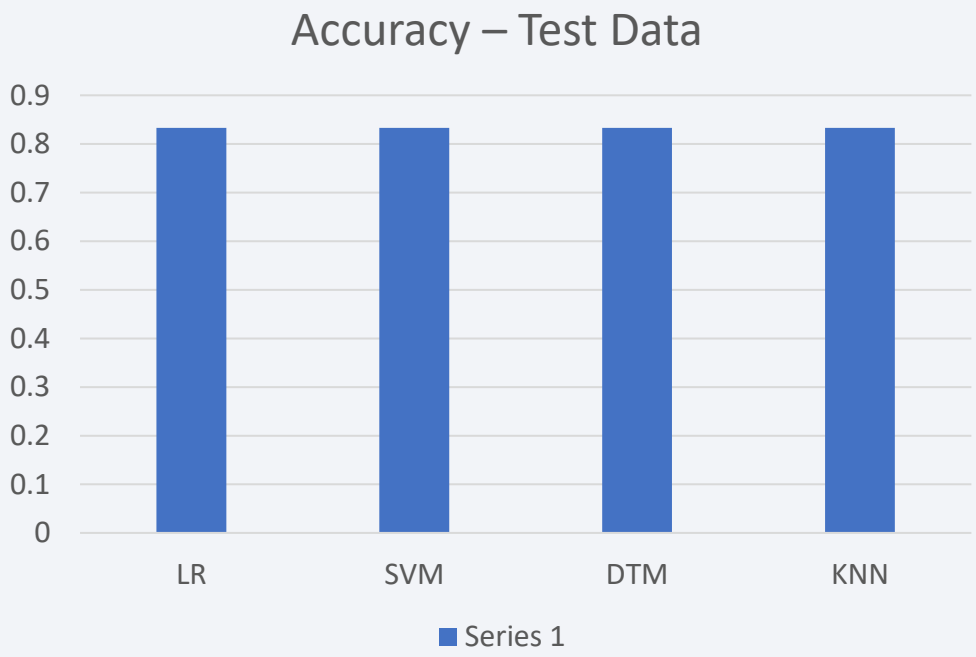
- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version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

Section 5

Predictive Analysis (Classification)

Classification Accuracy

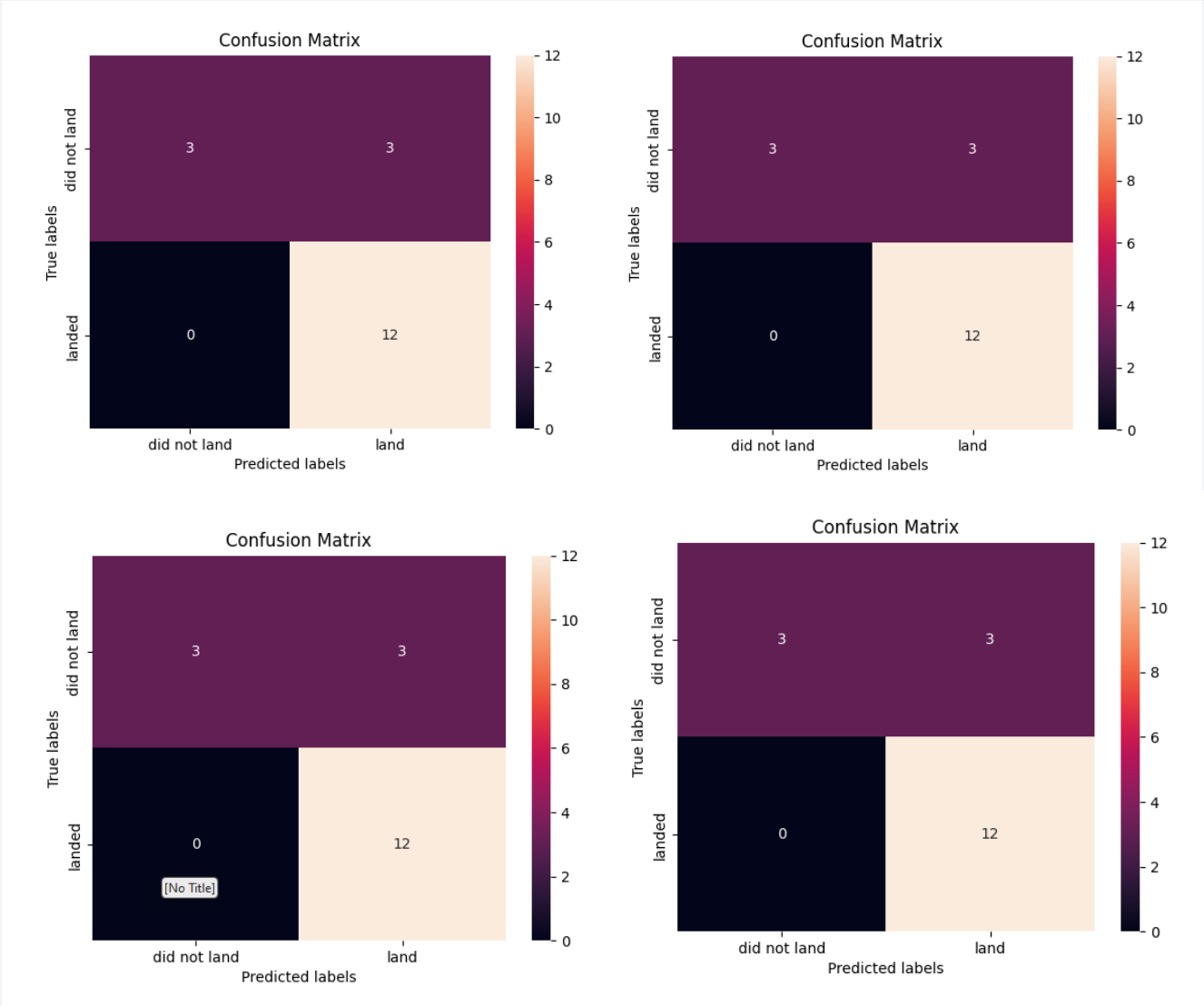




- Classification models perform same on test data however the decision tree model perform best on the train data (0.875 accuracy)

Confusion Matrix

- All models performs the same on the confusion matrix with accuracy of 0.833



Conclusions

- Different Factors predicts the success of the returning for stage 1 of falcon 9.
  - Payload Mass
  - Lunch Site
  - Orbit
- Descion Tree performs the best on the train data hence will be selected as the prediction model. Although other models performs the same on the test data.
- Highest payload mass has more success rate with increasing of launch times.
- ES11, GEO, HEO and SEO have the highest success rates.

Appendix

- Access to all files
  - https://github.com/jupyter123321/Space-X-ipynp-Files/tree/main

Thank you!