IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sarsha J S
June 10, 2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - EDA with SQL
  - EDA with visualization
  - Interactive visual analytics with folium
  - Interactive dashboard with Plotly Dash
  - Machine Learning prediction
- Summary of all results
  - EDA results
  - Interactive results
  - Prediction results

Introduction

- Project background and context

  The commercial space age is here, companies are making space travel affordable for everyone. Virgin Galactic is providing suborbital spaceflights. Rocket Lab is a small satellite provider. Blue Origin manufactures sub-orbital and orbital reusable rockets. Perhaps the most successful is SpaceX. SpaceX's accomplishments include: Sending spacecraft to the International Space Station. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.

- Problems you want to find answers

  Predict whether Falcon 9 first stage will land successfully or not.

Section 1

Methodology

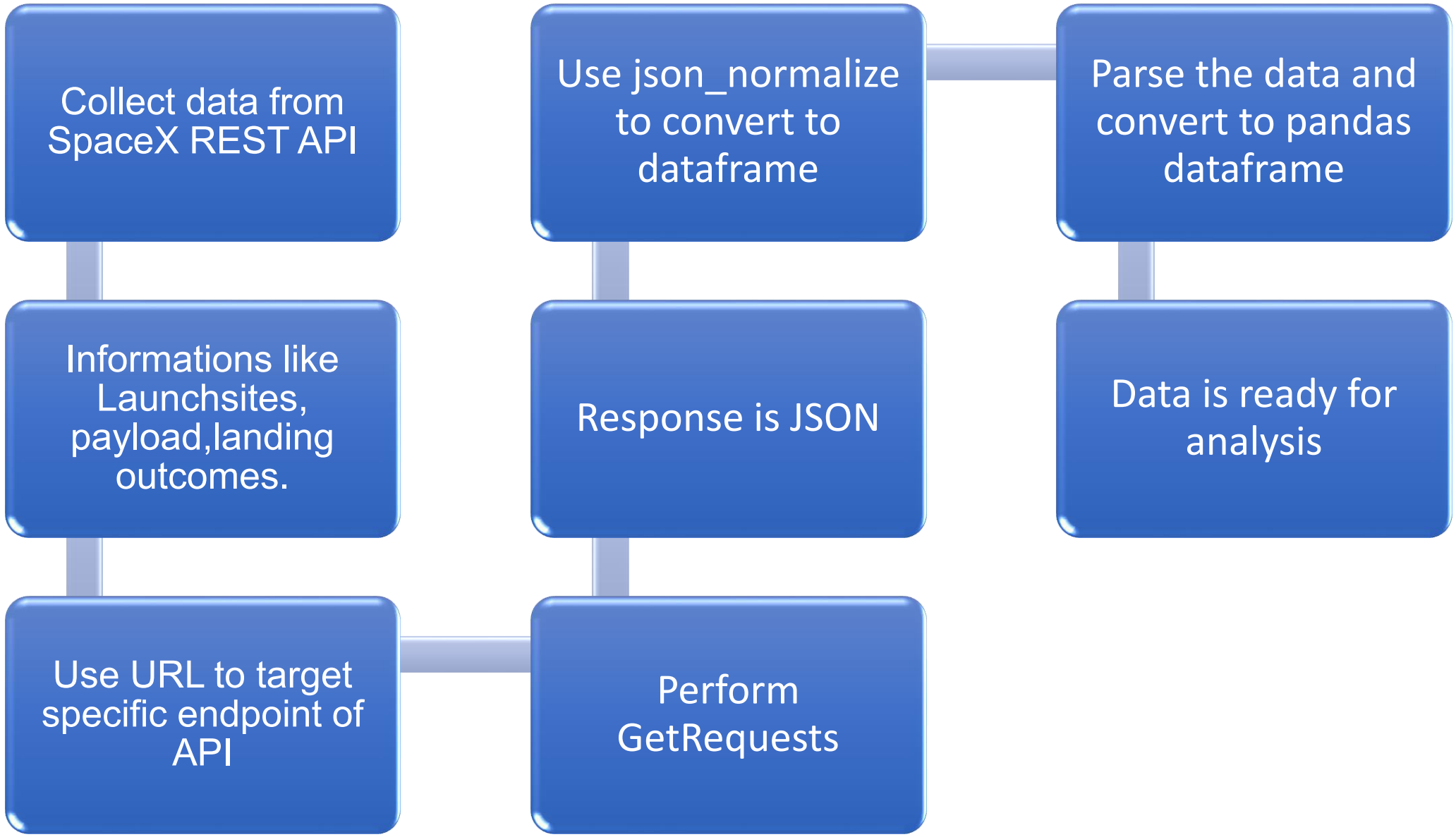
Methodology

Executive Summary

- Data collection methodology:
  - Data is collected using SpaceX rest API.
  - Web Scrapping from Wikipedia is done by BeautifulSoup.
- Perform data wrangling
  - Data cleaning to handle missing values, remove null values and removing unnecessary columns.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Logical regression, Support Vector Machines, Decision Tree, KNN algorithms have been used to train and test the models.

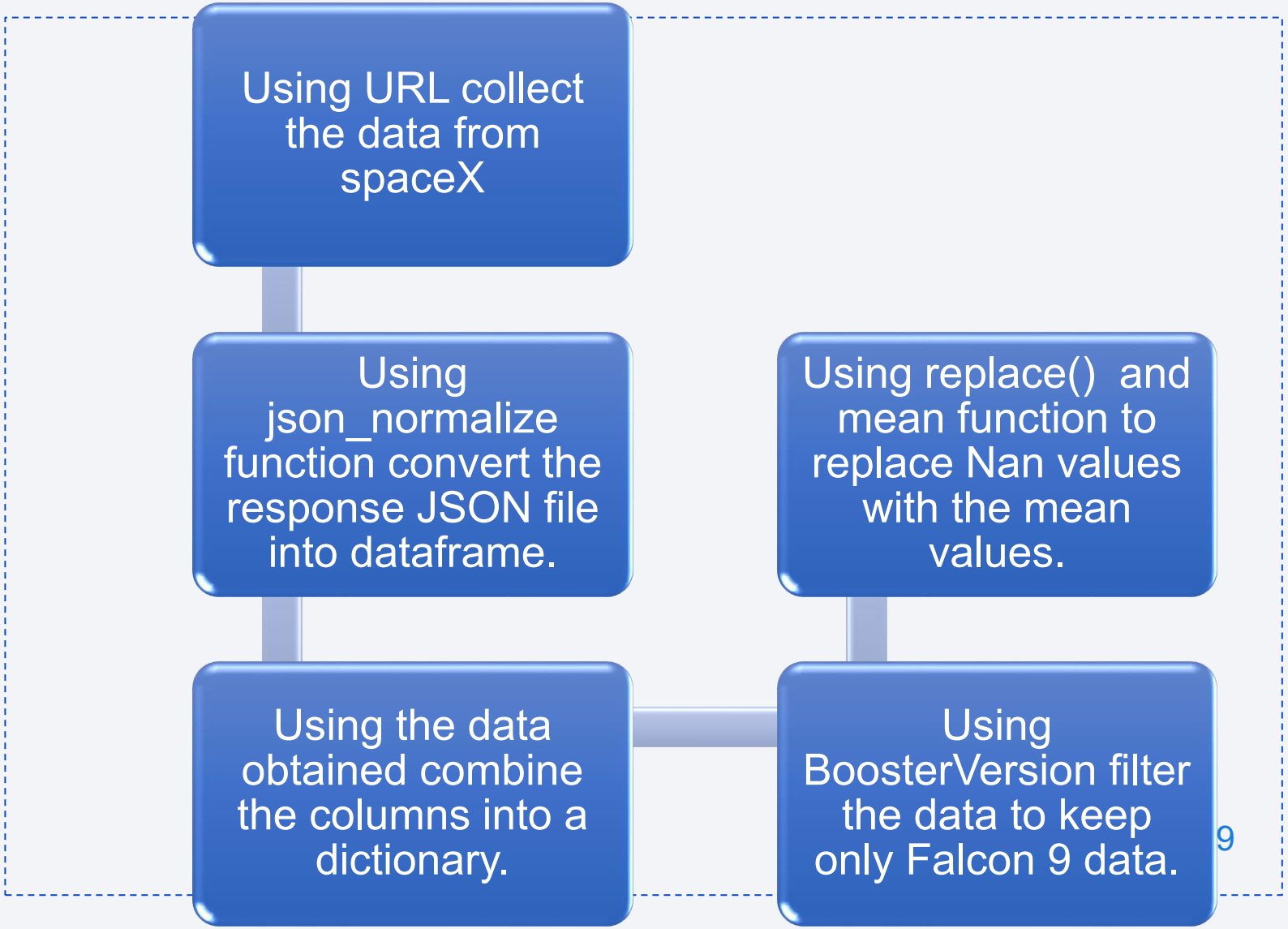
Data Collection

- Describe how data sets were collected.
  - Data sets were collected using SpaceX REST API.
  - This API will give data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications and landing outcomes.
  - Use the URL to target a specific endpoint of the API to get past launch data.
  - Perform a get requests using requests library to get the data from API. Response will be in JSON form.
  - Use json_normalize function to convert the JSON to dataframe.
  - Parse the data and convert it into pandas dataframe.
- You need to present your data collection process use key phrases and flowcharts


Collect data from SpaceX REST API
Informations like Launchsites, payload,landing outcomes.
Use URL to target specific endpoint of API
Perform GetRequests
Response is JSON
Use json_normalize to convert to dataframe
Parse the data and convert to pandas dataframe
Data is ready for analysis

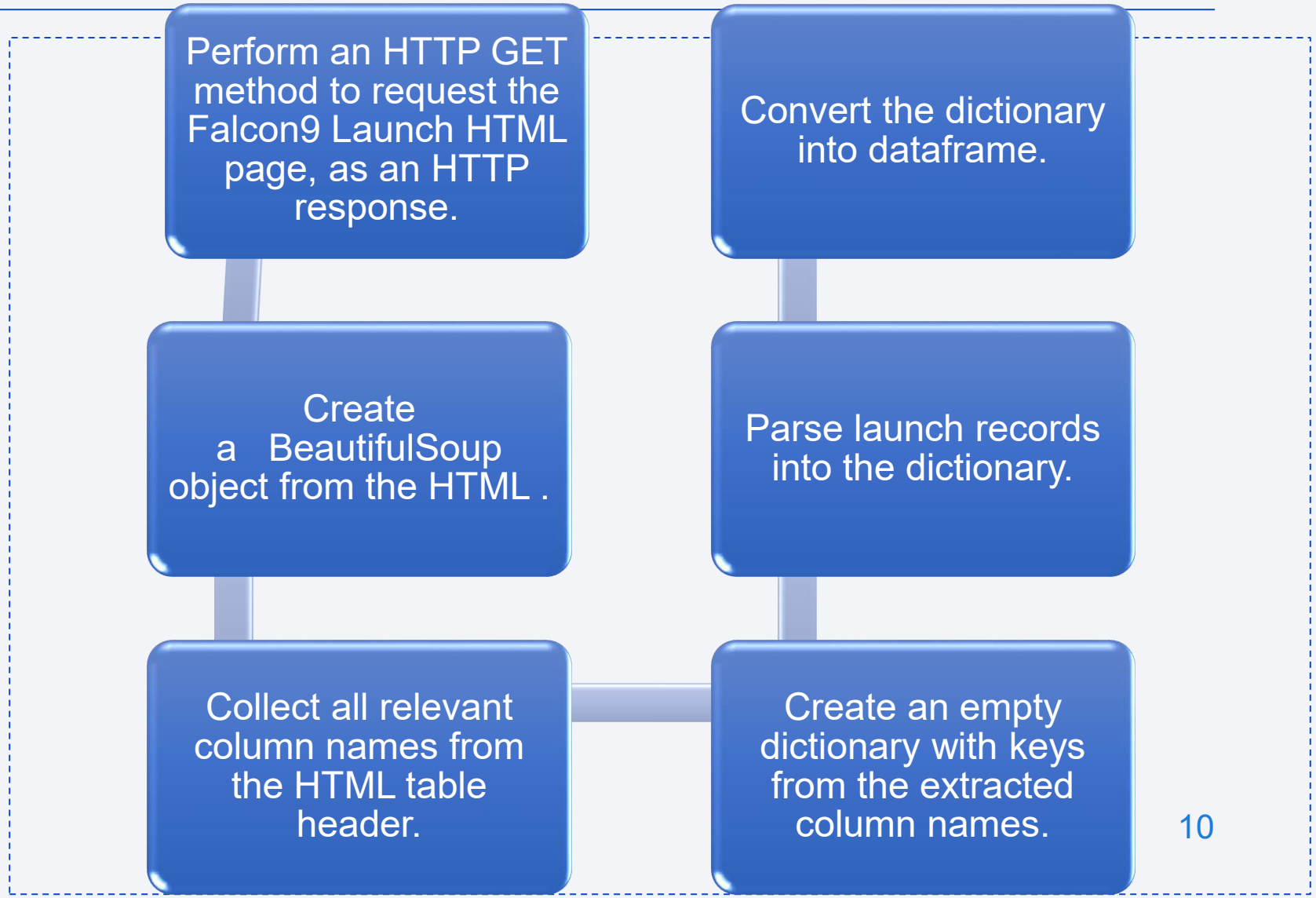
Data Collection – SpaceX API

Https://github.com/SarshaJS18/myrepo/blob/main/Week_1_Data_Collection_using_API.ipynb



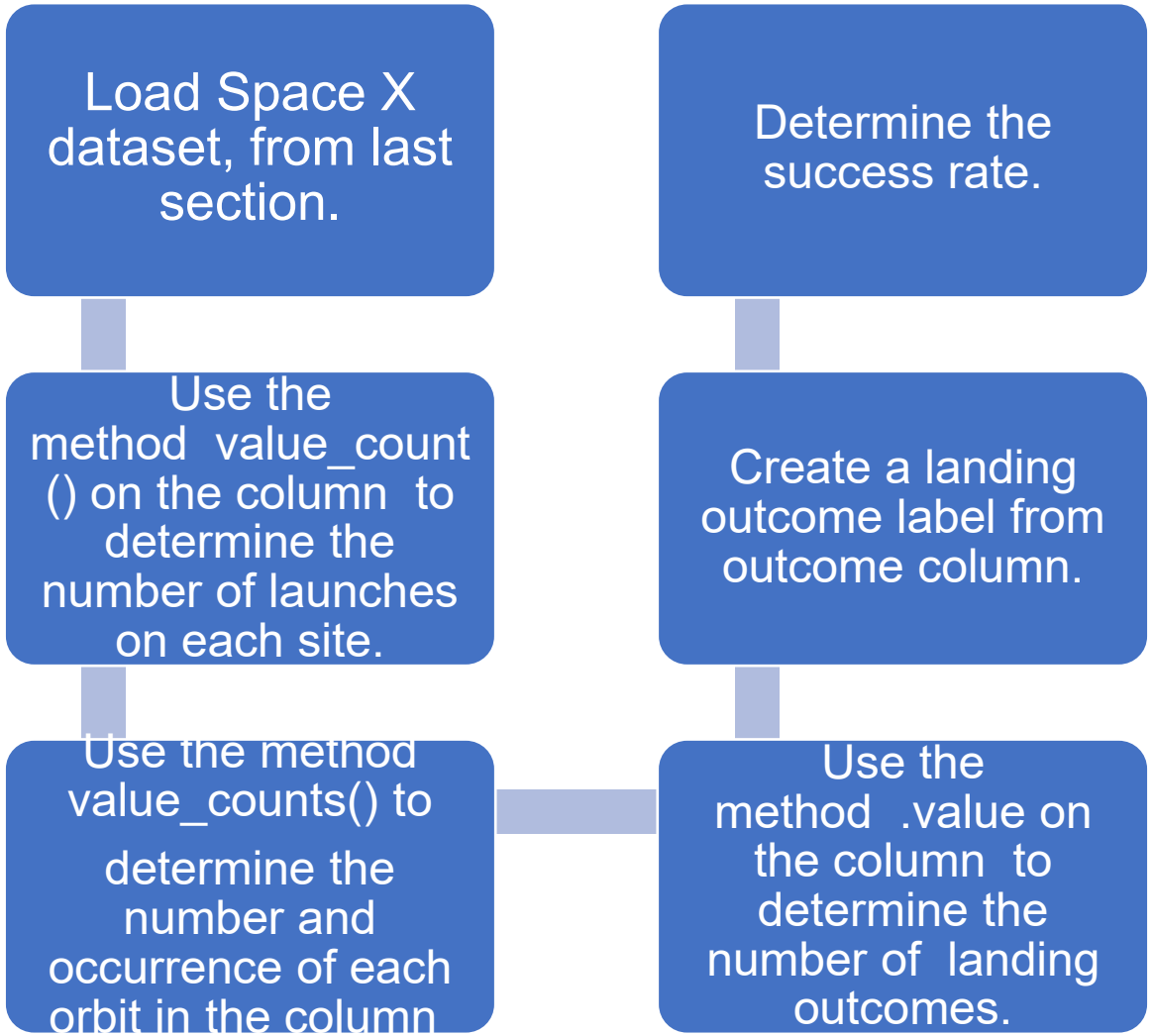
Data Collection - Scraping

GitHub URL:
https://github.com/SarshaJS18/myrepo/blob/main/Week1_webscraping.ipynb



Data Wrangling

https://github.com/SarshaJS18/myrepo/blob/main/week1_data_wrangling.ipynb



EDA with Data Visualization

The charts used are:

- Scatter point chart- for visualizing the relationship between variables, such as the relationship between payload mass and launch sites.
- Bar chart- for visualizing the relationship between success rate and orbit type.
- Line chart- for visualizing the success rate.

- Github URL: https://github.com/SarshaJS18/myrepo/blob/main/Week2_lab2_EDA_with_Visualization.ipynb

EDA with SQL

GithubURL:https://github.com/SarshaJS18/myrepo/blob/main/Week_2_lab1_EDA_withSQL.ipynb

- SQL queries performed:
  - Display 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 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 successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

Build an Interactive Map with Folium

- The map objects created and added to a folium map are:
  - Circle – to add a highlighted circle area with a text label on a specific coordinate.
  - Marker – to show a specific location.
  - Line – to show multiple connected points to represent boundaries, routes etc.
  - Marker_cluster – to show a map containing many markers having the same coordinate.
- GitHub URL: https://github.com/SarshaJS18/myrepo/blob/main/Week3_lab1_Interactive_visual_analytic_with_folium.ipynb

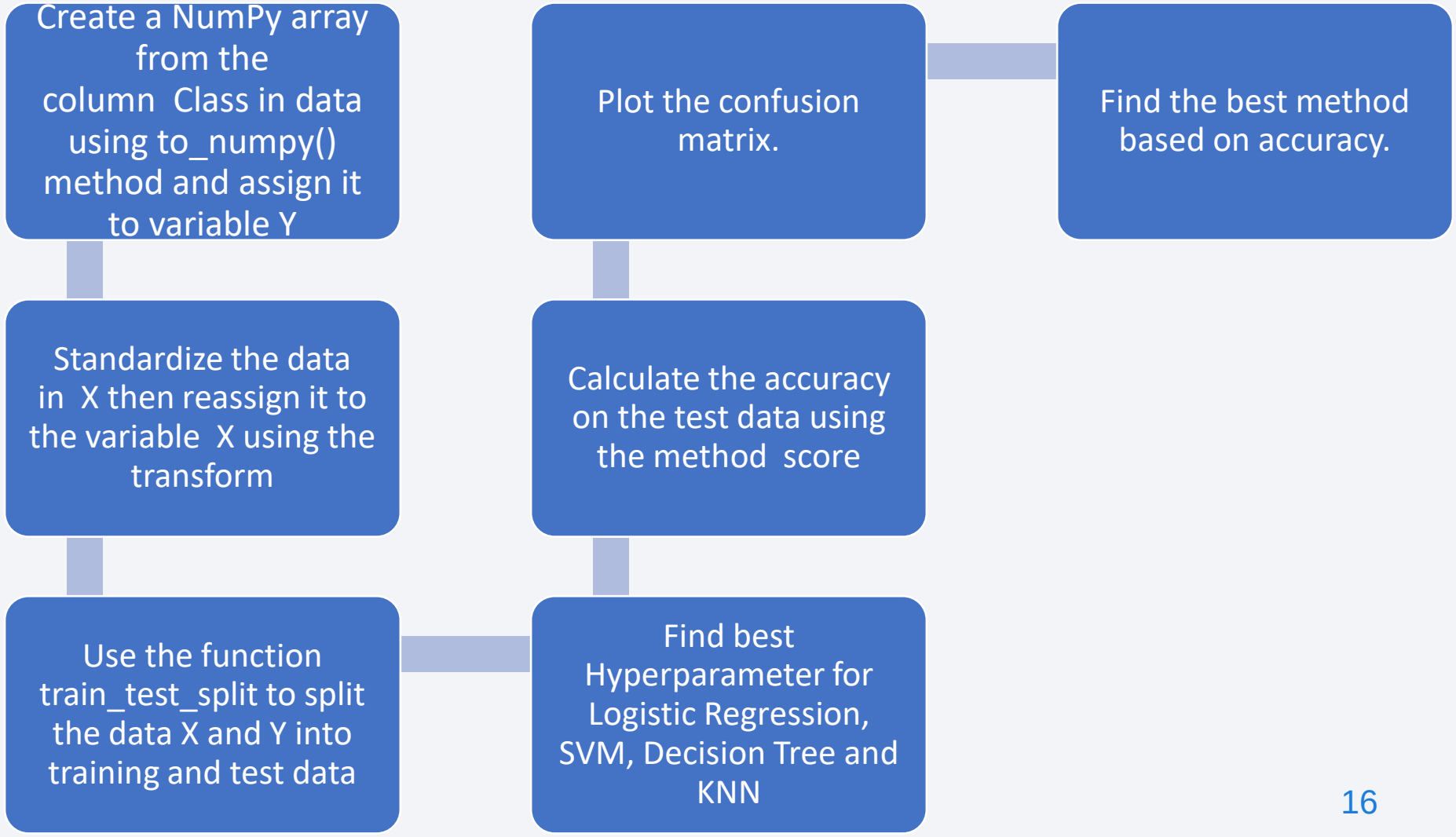
Build a Dashboard with Plotly Dash

- Plots/graphs and interactions added to a dashboard are:
  - Scatter plot – visualizing the relationship between two continuous variables.
  - Line chart - visualizing changes in data over continuous or discrete intervals.
  - Pie chart - visualizing percentages or proportions.

- GitHub URL: https://github.com/SarshaJS18/myrepo/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

Github URL:
https://github.com/SarshaJS18/myrepo/blob/main/Week4_lab1_Machine_learning_predictive.ipynb



Results

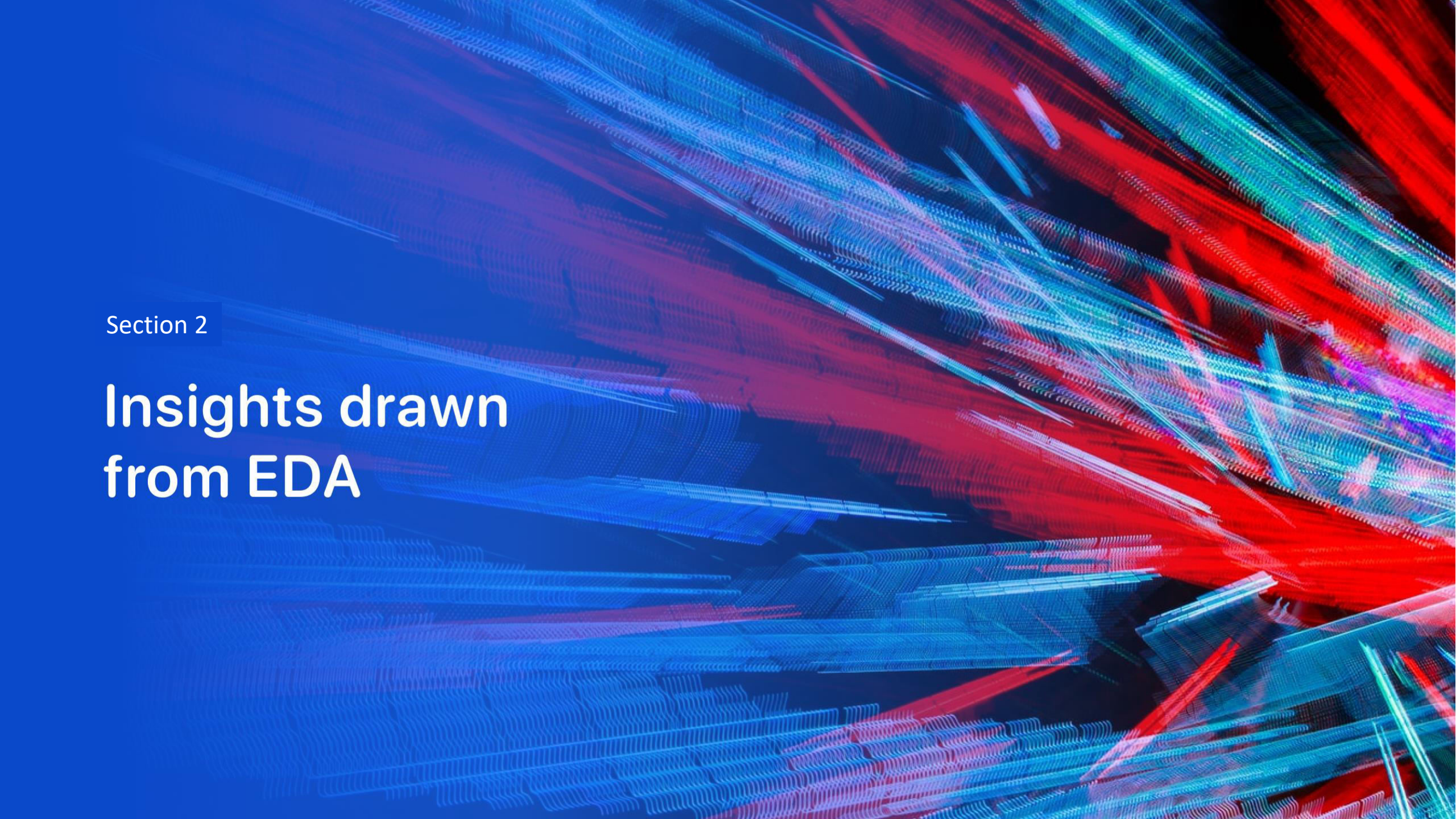
- Exploratory data analysis results

  Data can be collected using REST API and web scrapping from Wikipedia.

- Interactive analytics demo in screenshots

- Predictive analysis results

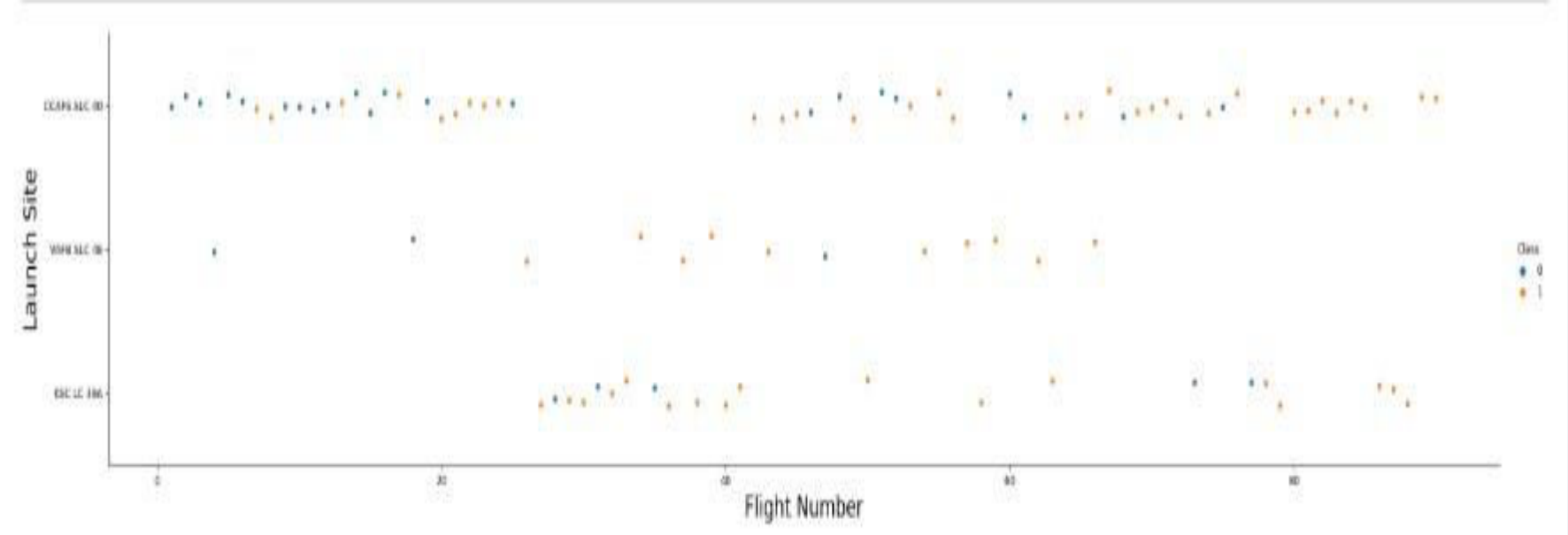
  Decision Tree method performs best using test data.


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

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

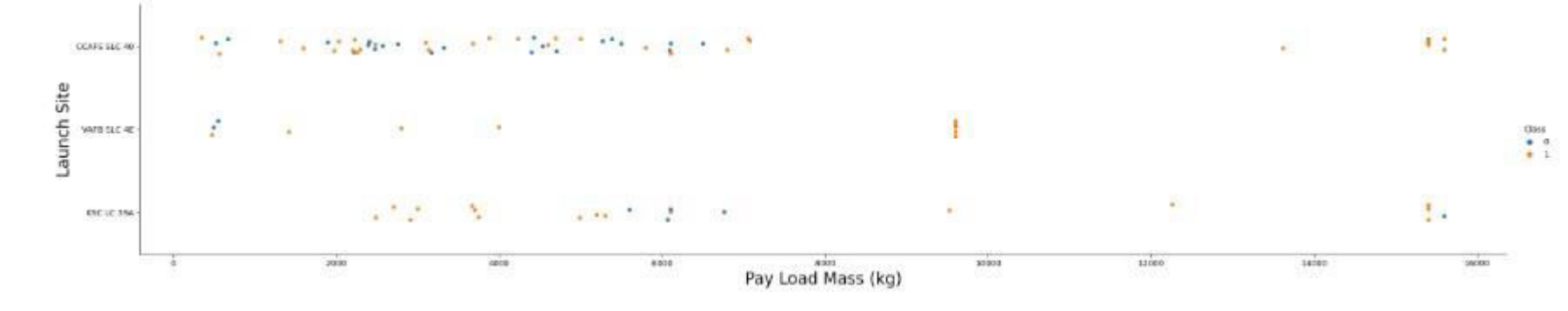
  - The below screenshot shows the relationship between flight Number and launch site.



Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site

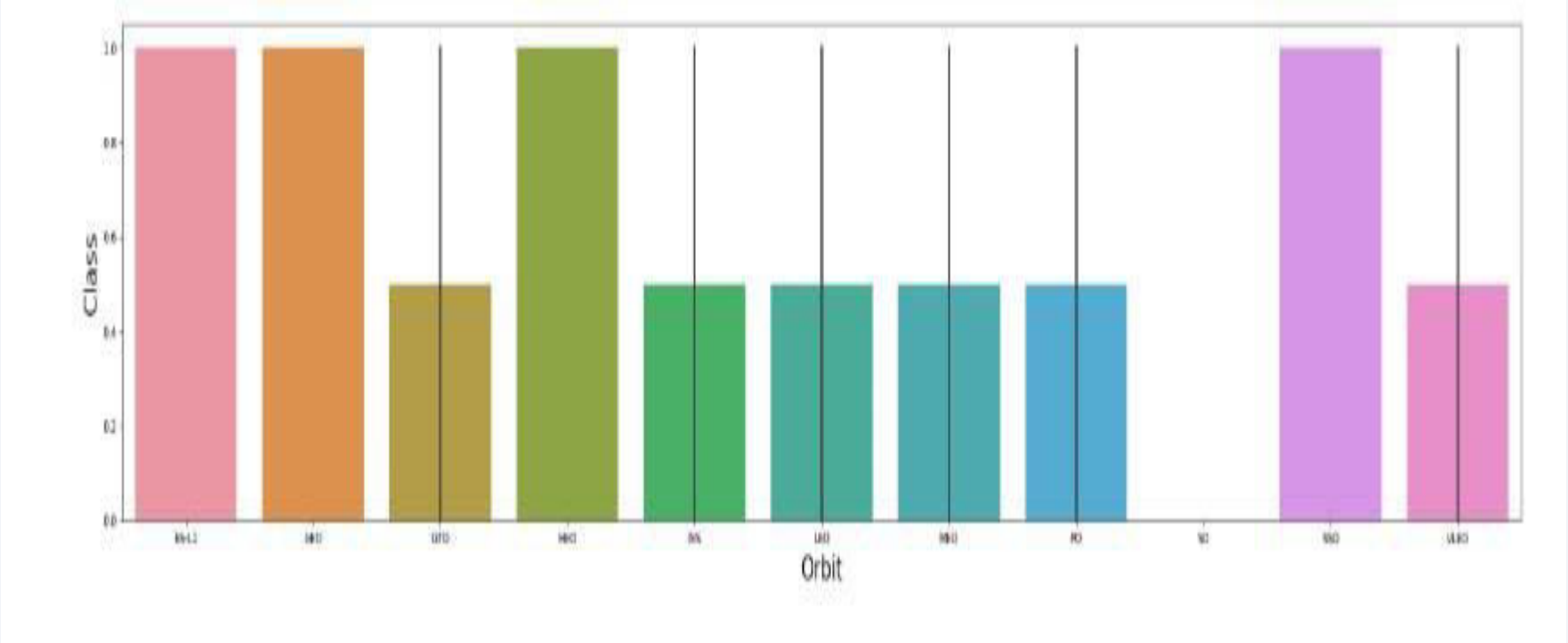
  - The screenshot below shows the relationship between Payload and Launch sites.



Success Rate vs. Orbit Type

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

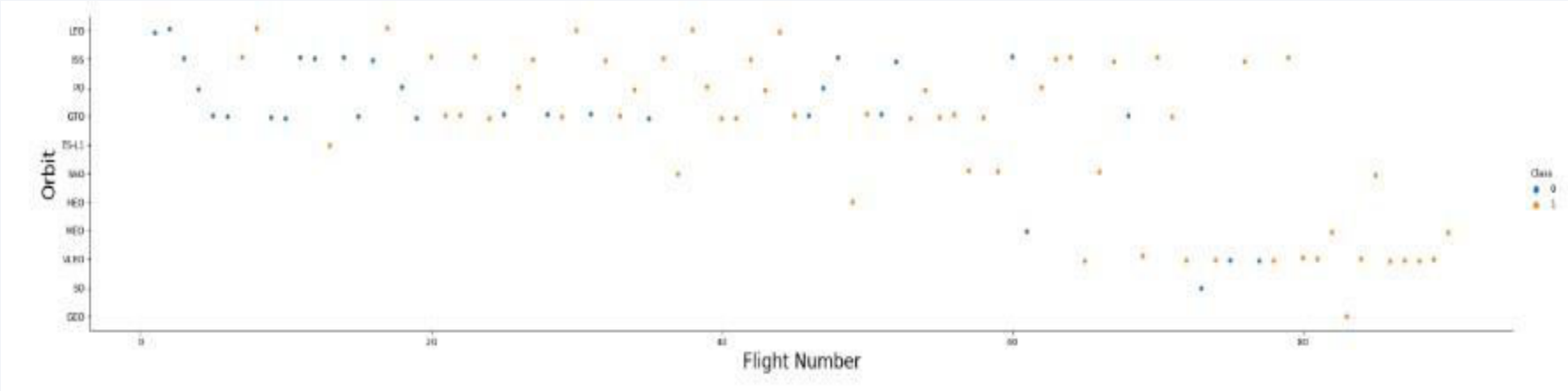
  - The below screenshot shows the orbit SSO, HEO, GEO, ES_L1 have high success and the orbit GTO has low success among all of the orbits.



Flight Number vs. Orbit Type

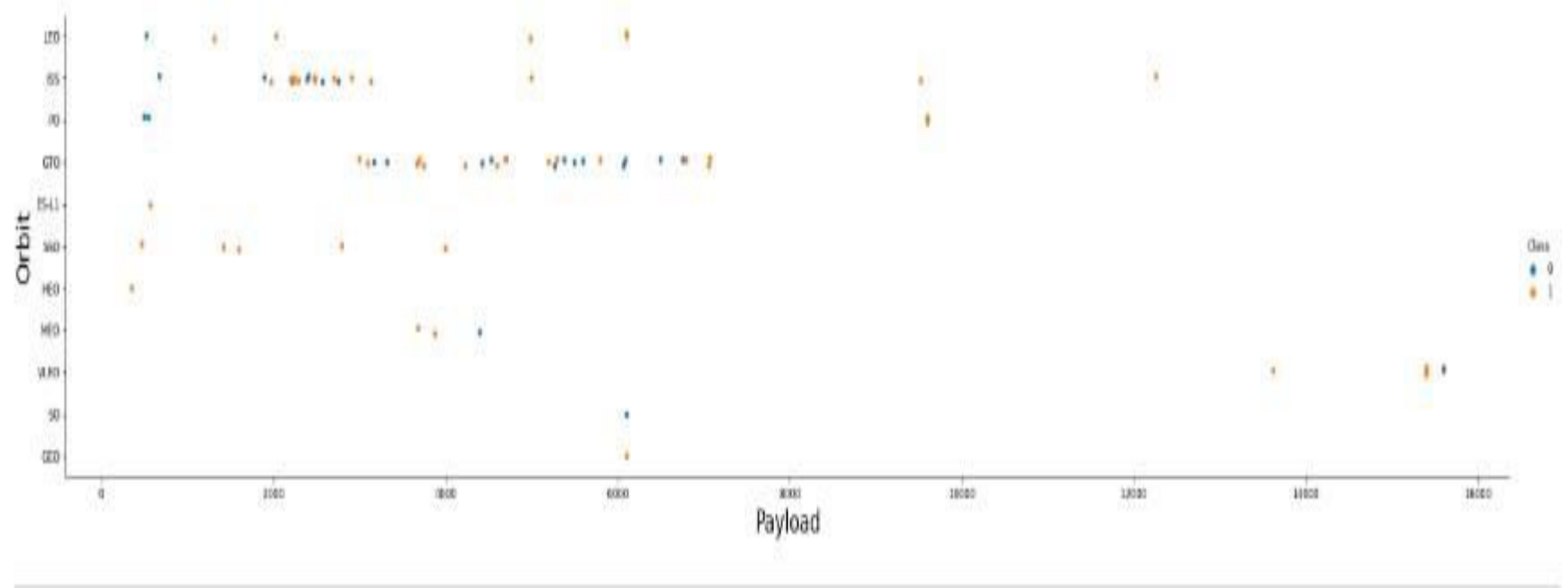
- Show a scatter point of Flight number vs. Orbit type

  - The below screenshot shows the success occur in LEO orbit and not in GTO orbit.



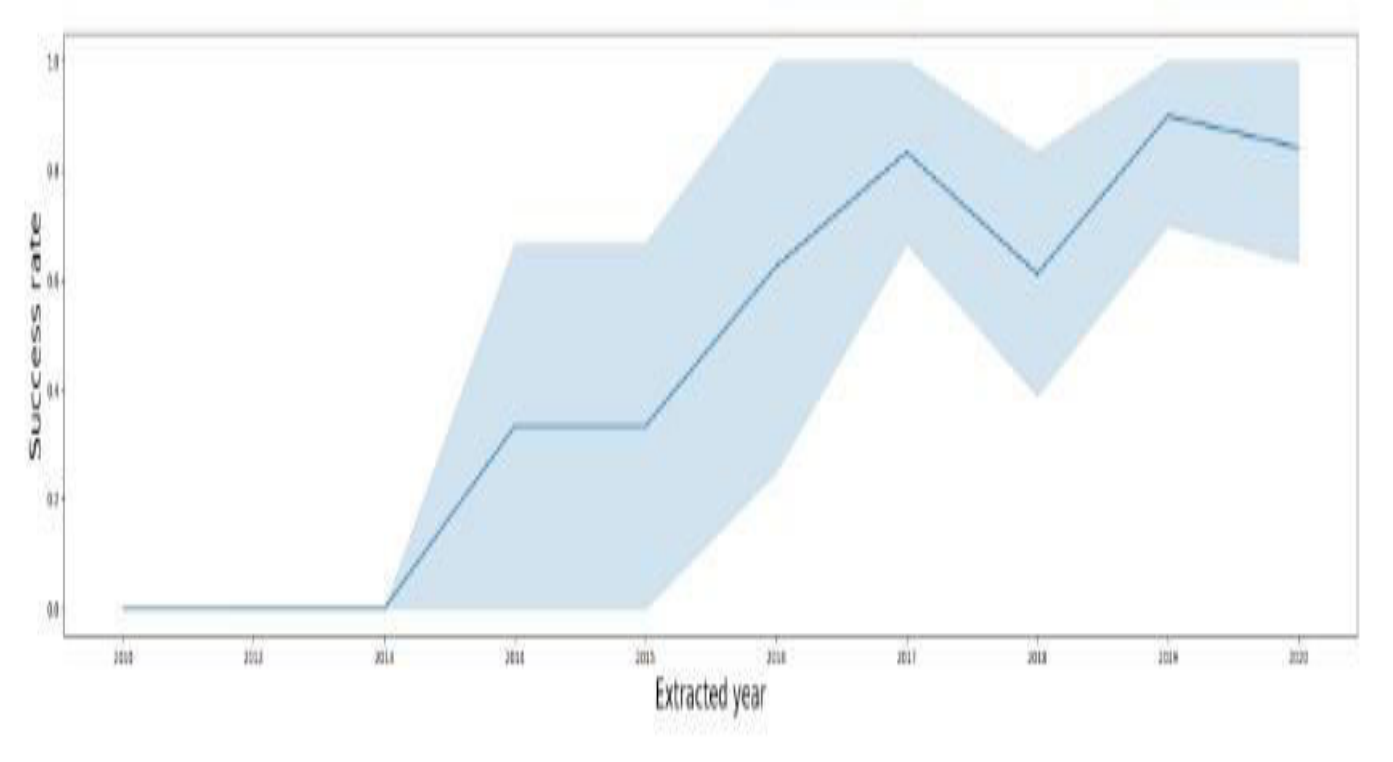
Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type
  - The below screenshot shows that heavy payloads have positive response LEO orbit and negative response on GTO orbits.



Launch Success Yearly Trend

- Show a line chart of yearly average success rate
  - The below screenshot shows success is steadily increasing from the year 2013.



All Launch Site Names

- the names of the unique launch sites are shown in the screenshot.

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

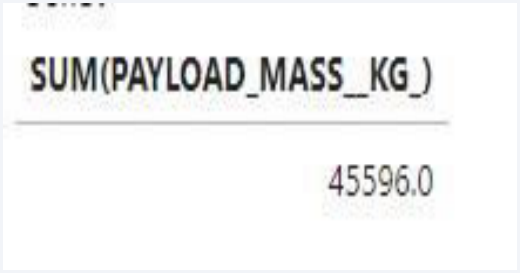
Launch Site Names Begin with 'CCA'

- 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_
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (p
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (p
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	N
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	N

Total Payload Mass

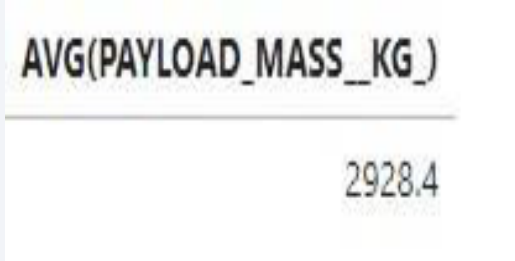
-the below screenshot shows the total payload carried by boosters from NASA is 45596.0



SUM(PAYLOAD_MASS__KG_)
45596.0

Average Payload Mass by F9 v1.1

- The below screenshot shows the average payload mass carried by booster version F9 v1.1 is 2928.4



First Successful Ground Landing Date

- The below screenshot shows the dates of the first successful landing outcome on ground pad is August 1,2018

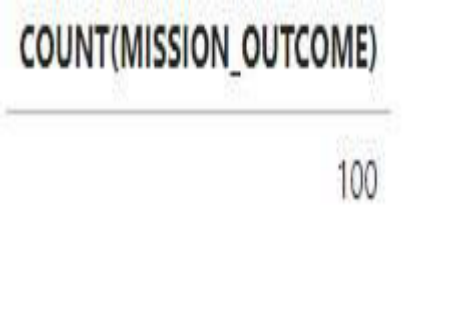
Successful Drone Ship Landing with Payload between 4000 and 6000

- The below screenshot shows 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

Total Number of Successful and Failure Mission Outcomes

- The below screenshot shows the total number of successful and failure mission outcomes is 100



Boosters Carried Maximum Payload

The below screenshot shows 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

2015 Launch Records

- The below screenshot shows the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

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

- The below screenshot shows the 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

Date	Landing_Outcome	COUNT	RANK
08/07/2018	Success	20	1
04/08/2016	Success (drone ship)	8	2
18/07/2016	Success (ground pad)	7	3


Section 3
Launch Sites
Proximities Analysis

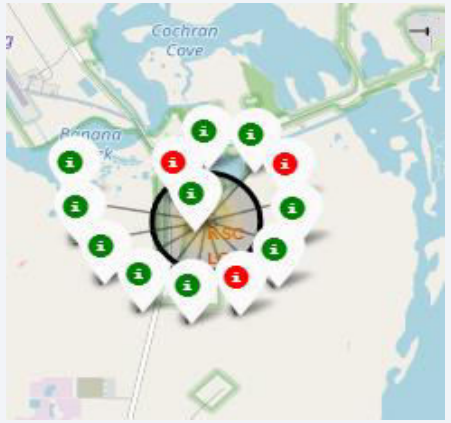
Launch sites

-The below screenshot shows the location of all launch sites in red label.

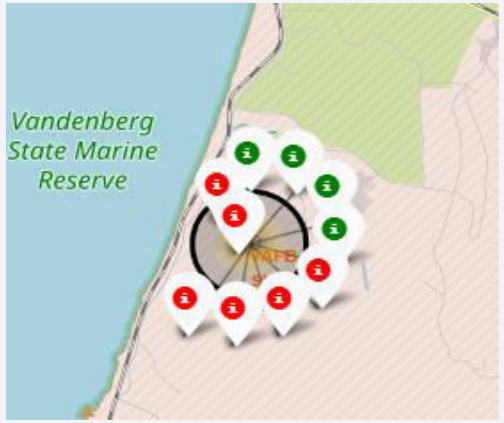
Launch Success

- The below screenshot shows the launch sites success in green icon and failure in red icon.
- From the screenshots we can conclude that KSC LC-39A has moss success among all of them

KSC LC-39A

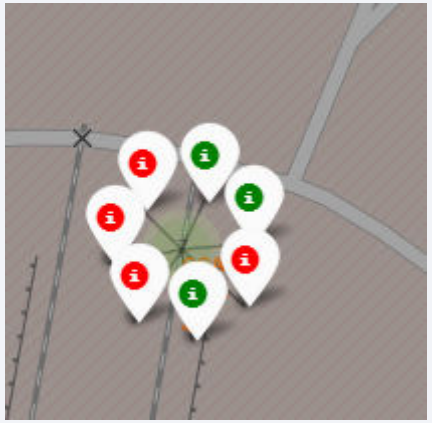


VAFB SLC-4E



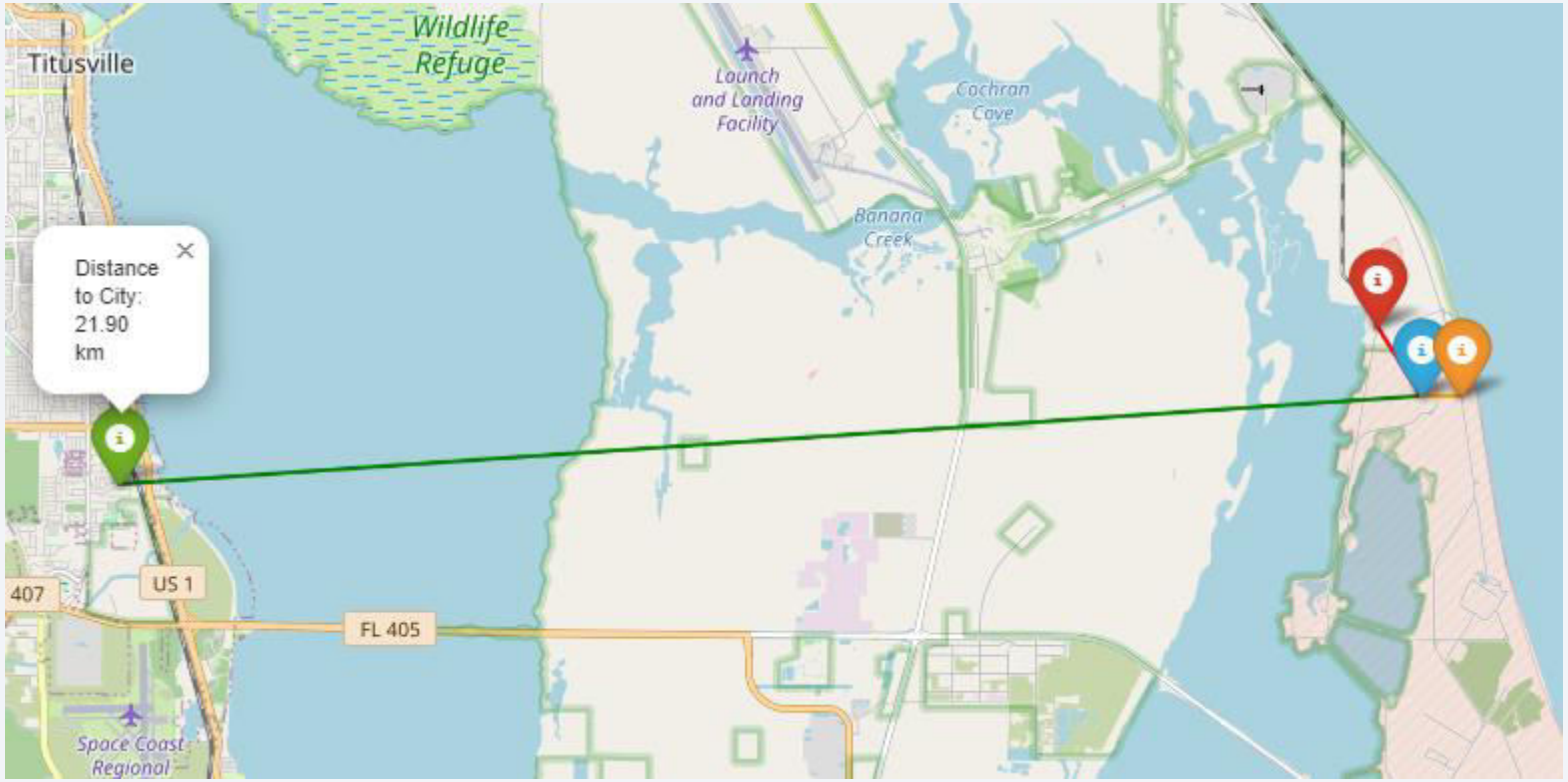
CCAFSLC-40

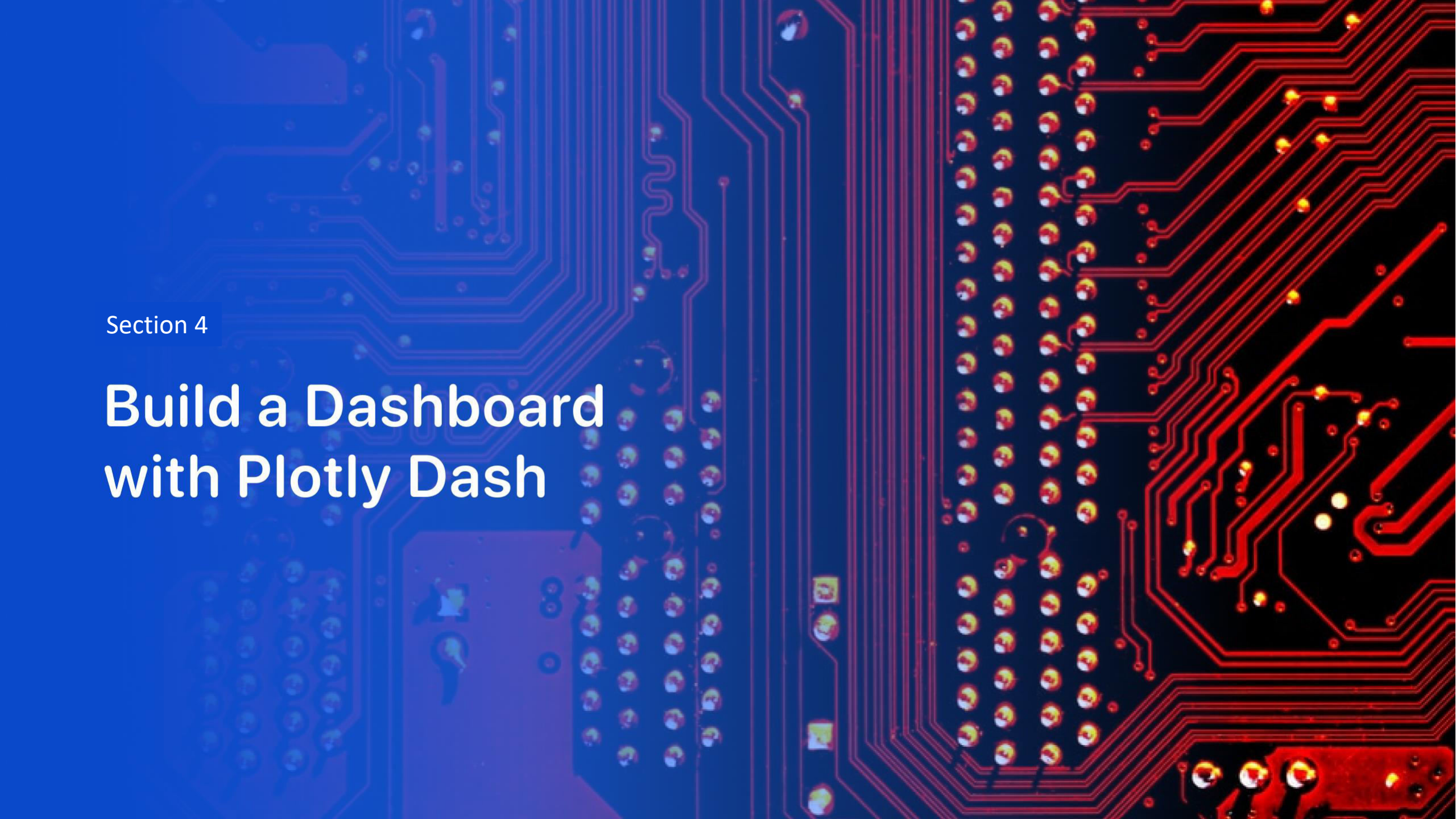
CCAFS SLC-40



Distance from Launch site

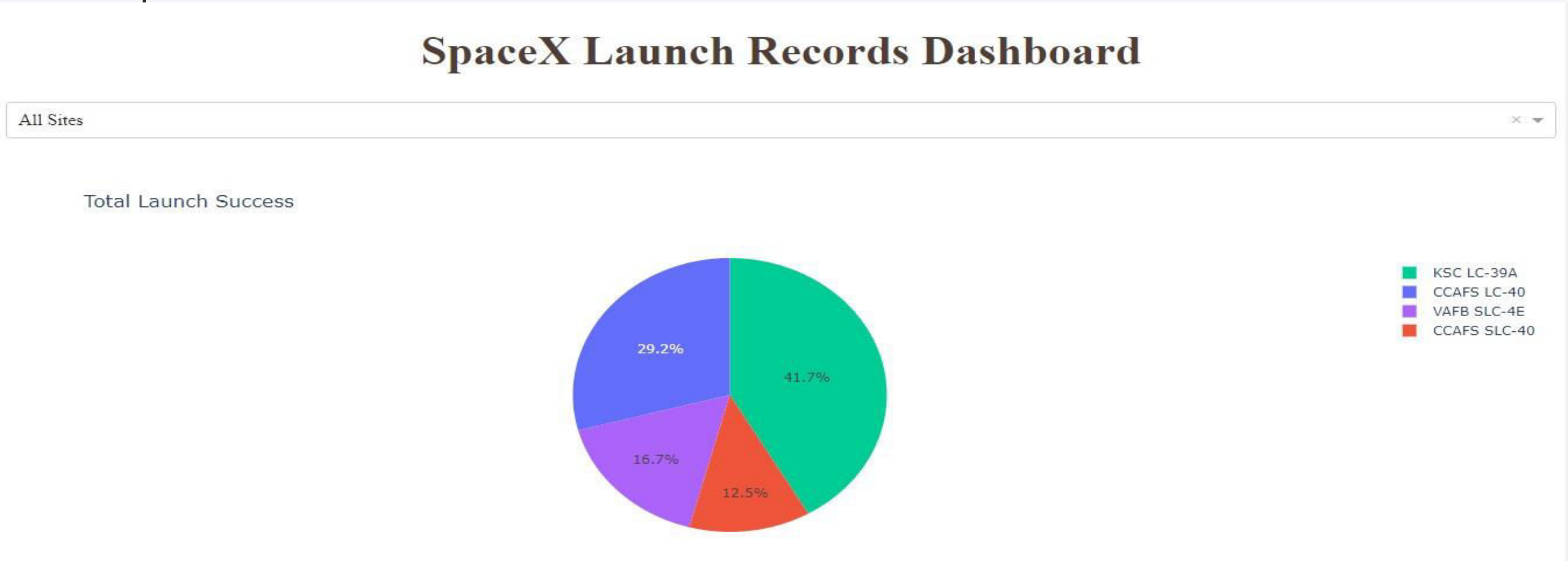
- The below screen shot shows launch site in blue icon, distance to city in green icon, distance to railroad in red icon and distance to highway road in orange icon.




Section 4
Build a Dashboard with Plotly Dash

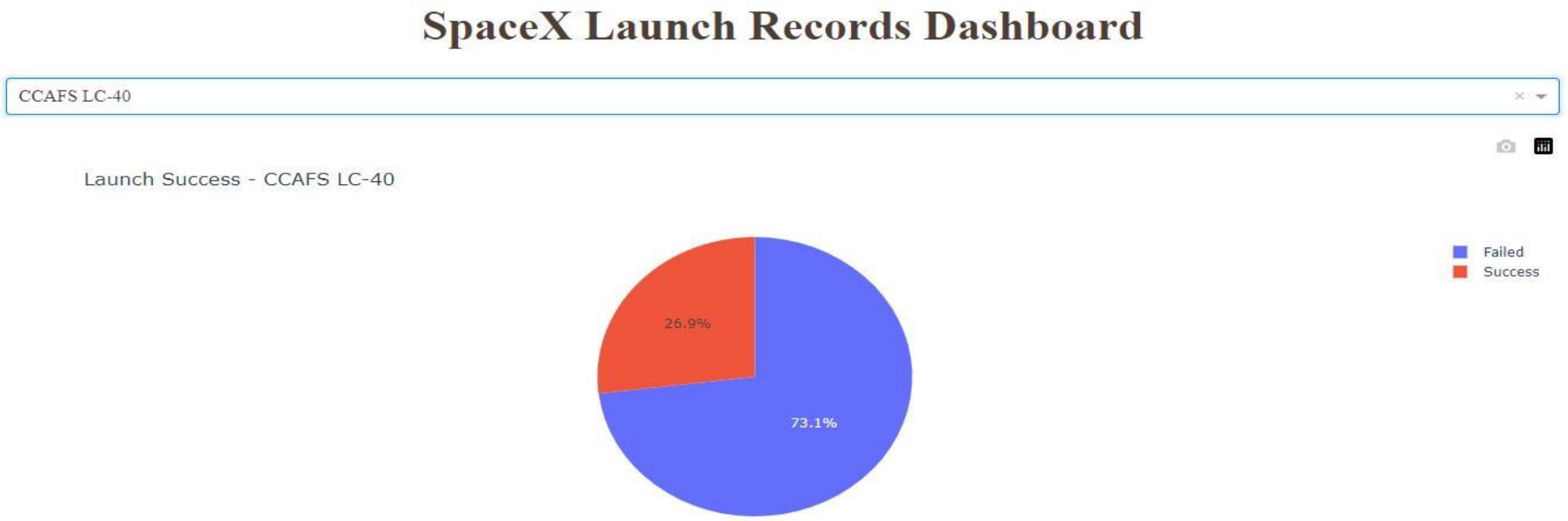
Launch success count

- The below screenshot shows the launch success count for all sites, in a piechart



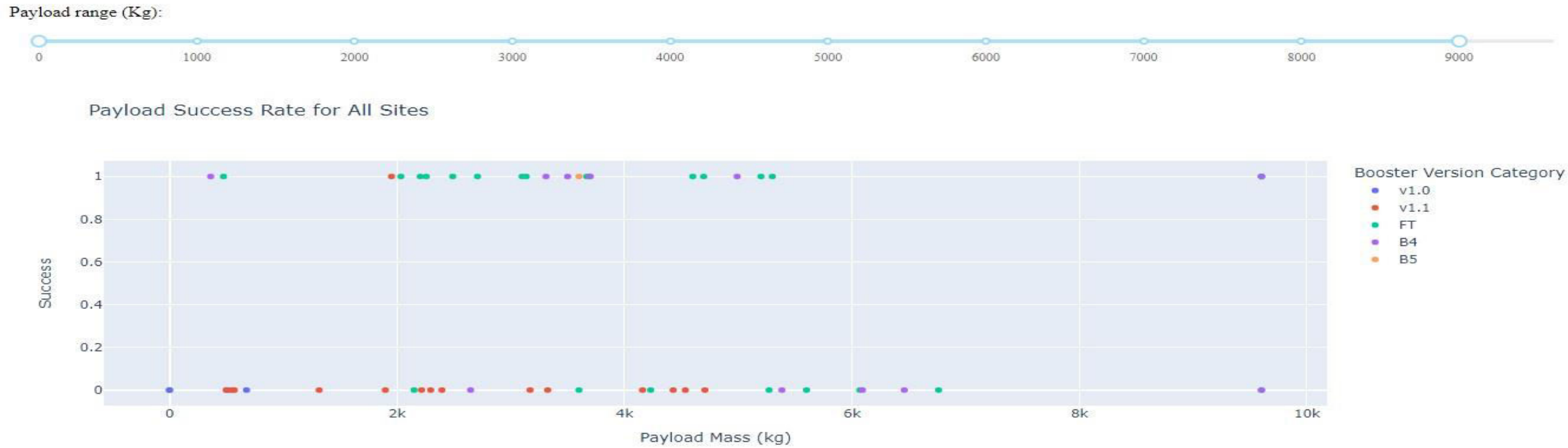
Lunch site with highest success ratio

-The below screenshot shows the piechart for the launch site with highest launch success ratio



Payload vs Success

-The below screenshots shows the Payload vs. Launch Outcome scatter plot for all sites

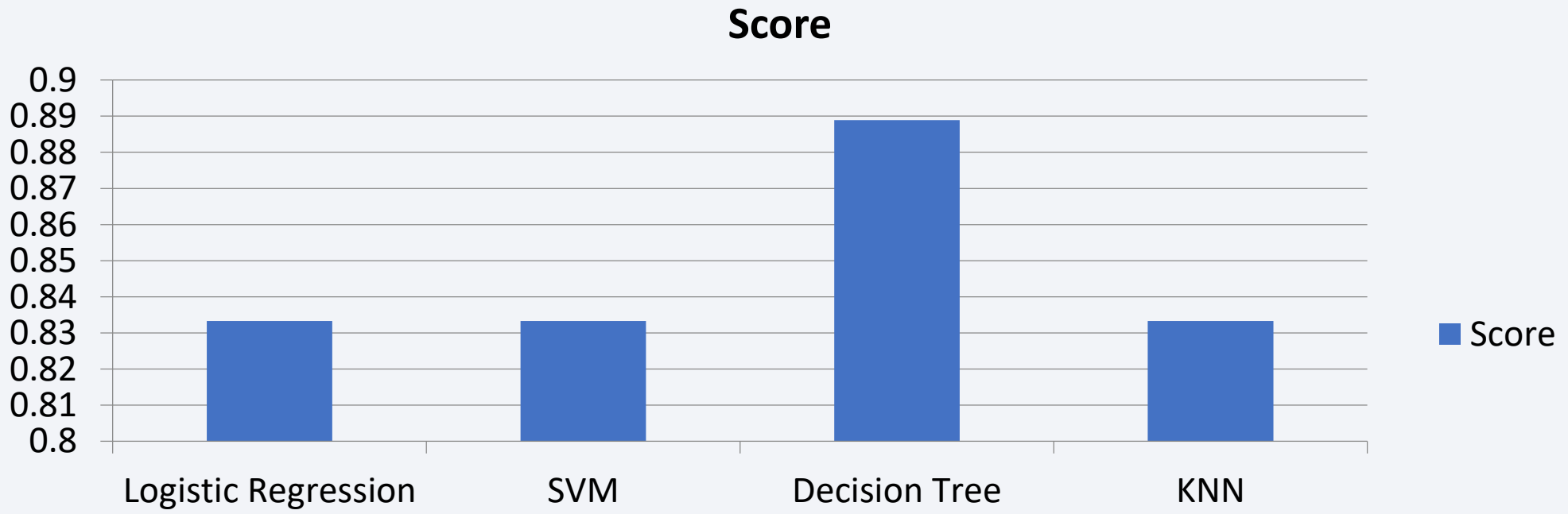


Section 5

Predictive Analysis (Classification)

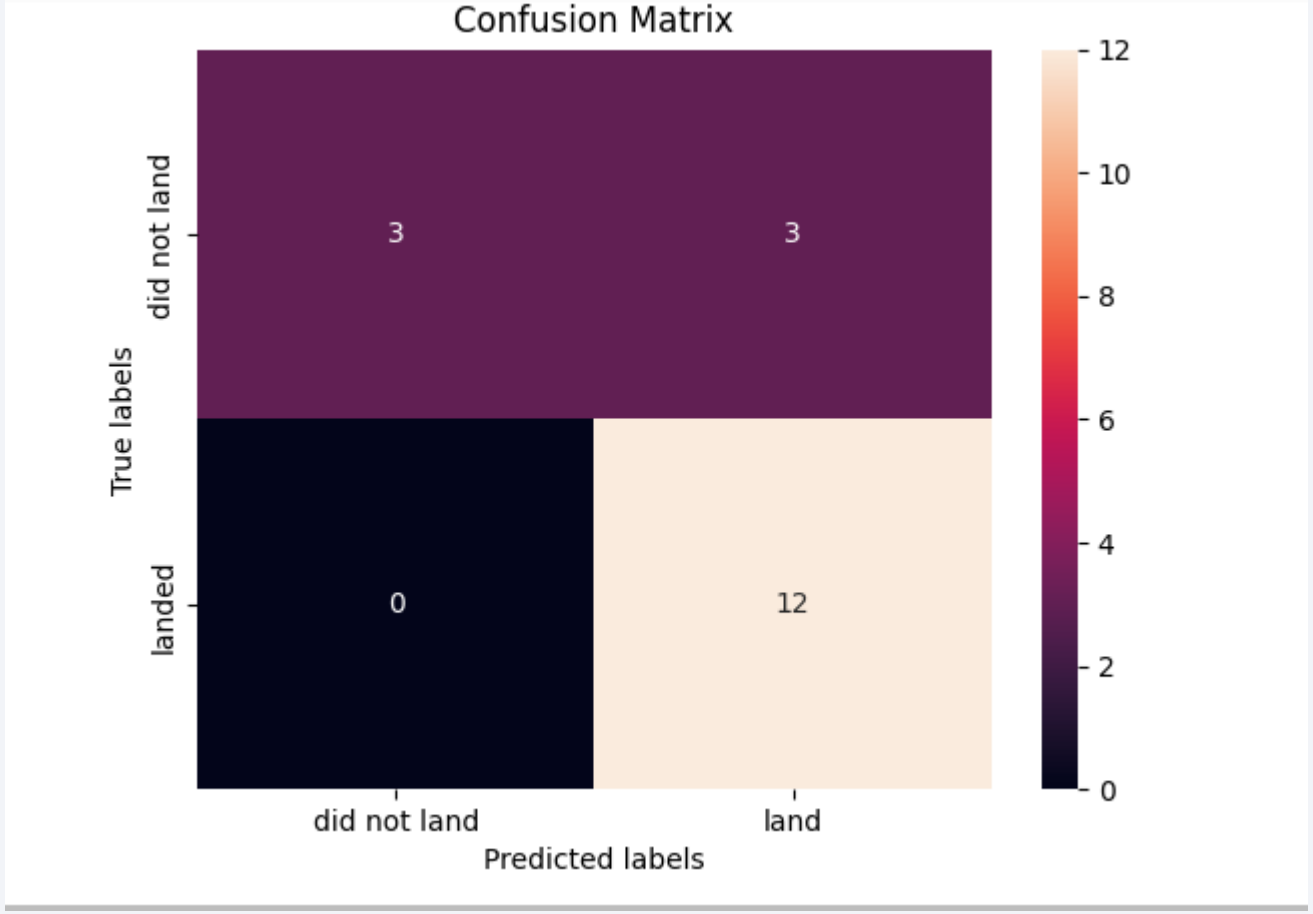
Classification Accuracy

From the bar chart we can conclude that Decision Tree has the highest classification accuracy.



Confusion Matrix

-The below screenshot shows the confusion matrix of the best performing model “Decision Tree”



Conclusions

- Data from SpaceX can be collected by both REST API and web scrapping.
- Data visualization is performed by charts and plot like scatter plot, bar chart and line graphs.
- SQL queries are performed.
- Dashboard is created using pie chart, line chart and scatter plots.
- Machine learning model is built and data are trained and tested.
- We can conclude that Decision Tree Algorithm is the method based on accuracy.

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!