IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rama Krishna Pradeep Maddi
07/06/2024

Outline

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

Executive Summary

- Summary of methodologies
  - Collected data using SpaceX REST API and Wikipedia web page scraping
  - Data Wrangling
  - Data Analysis
  - Insights drawn from EDA
  - Poximity Analysis
  - Predictive Analysis
- Decision Tree Model accuracy is better than other classification models

Introduction

- Space X 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 Space X can reuse the first stage.
- Problem is to determine if the first stage will land, we can determine the cost of a launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Collect data by using the SpaceX REST API calls
  - Use web scraping to collect Falcon 9 historical launch records from a Wikipedia page titled List of Falcon 9 and Falcon Heavy launches
- Perform data wrangling
  - Replace missing PayLoadMass with mean PayLoadMass
- 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

- Describe how data sets were collected.
  - Data sets were collected using SpaceX REST API Calls and Web Scrapping
- You need to present your data collection process use key phrases and flowcharts
  - Call REST API
  - Normalize the JSON response
  - Call REST APIs for each ID column to replace with actual values
  - Parse data and retrieve needed information like rocket, payload, launchpads and cores

Data Collection – SpaceX API

- Used SpaceX REST API to retrieve data: https://api.spacexdata.com/v4/launches/past

- GitHub URL of the completed SpaceX API calls notebook:

https://github.com/pmdatascience/testrepo/blob/main/jupyter-labs-spacex-data-collection-api%20(2).ipynb

Data Collection - Scraping

- Retrieved Falcon9 launch data from wiki page using Web Scraping: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- GitHub URL of the completed web scraping notebook: https://github.com/pmdatascience/testrepo/blob/main/jupyter-labs-webscraping%20(1).ipynb

Data Wrangling

- As part of Data Wrangling we mainly converted outcomes into Training Labels with 1 means booster successfully landed and 0 means unsuccessful
- Bad outcomes replaced the value with 0 and other outcomes with 1
- GitHub URL of data wrangling related notebooks: https://github.com/pmdatascience/testrepo/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Below charts were plotted
  - Categorical Plot (catplot): To Visualize the relationship between
    1. Flight Number and Launch Site
    2. Payload and Launch Site
    3. Flight Number and Orbit Type
    4. Payload and Orbit Type
  - Bar Plot: To Visualize the relationship between success rate of each orbit type
  - Line Plot: To Visualize the launch success yearly trend

- GitHub URL of completed EDA with data visualization notebook: https://github.com/pmdatascience/testrepo/blob/main/edadataviz.ipynb

EDA with SQL

- Below list of 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 date when the first succesful landing outcome in ground pad was acheived.
  - 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 landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.
- GitHub URL of completed EDA with SQL notebook: https://github.com/pmdatascience/testrepo/blob/main/jupyter-labs-eda-sql-coursera_sqllite%20(2).ipynb

Build an Interactive Map with Folium

- Below map objects such as markers, circles, lines, etc. Were created and added to a folium map
  - Added Circle Object for each launch site based on latitude and longitude
  - Mark the success or failed launches for each site
  - Calculate the distance between a launch site to its proximities (like City, Highway, Railway) and draw lines
- GitHub URL of your completed interactive map with Folium map: https://github.com/pmdatascience/testrepo/blob/main/lab_jupyter_launch_site_location%20(2).ipynb

Build a Dashboard with Plotly Dash

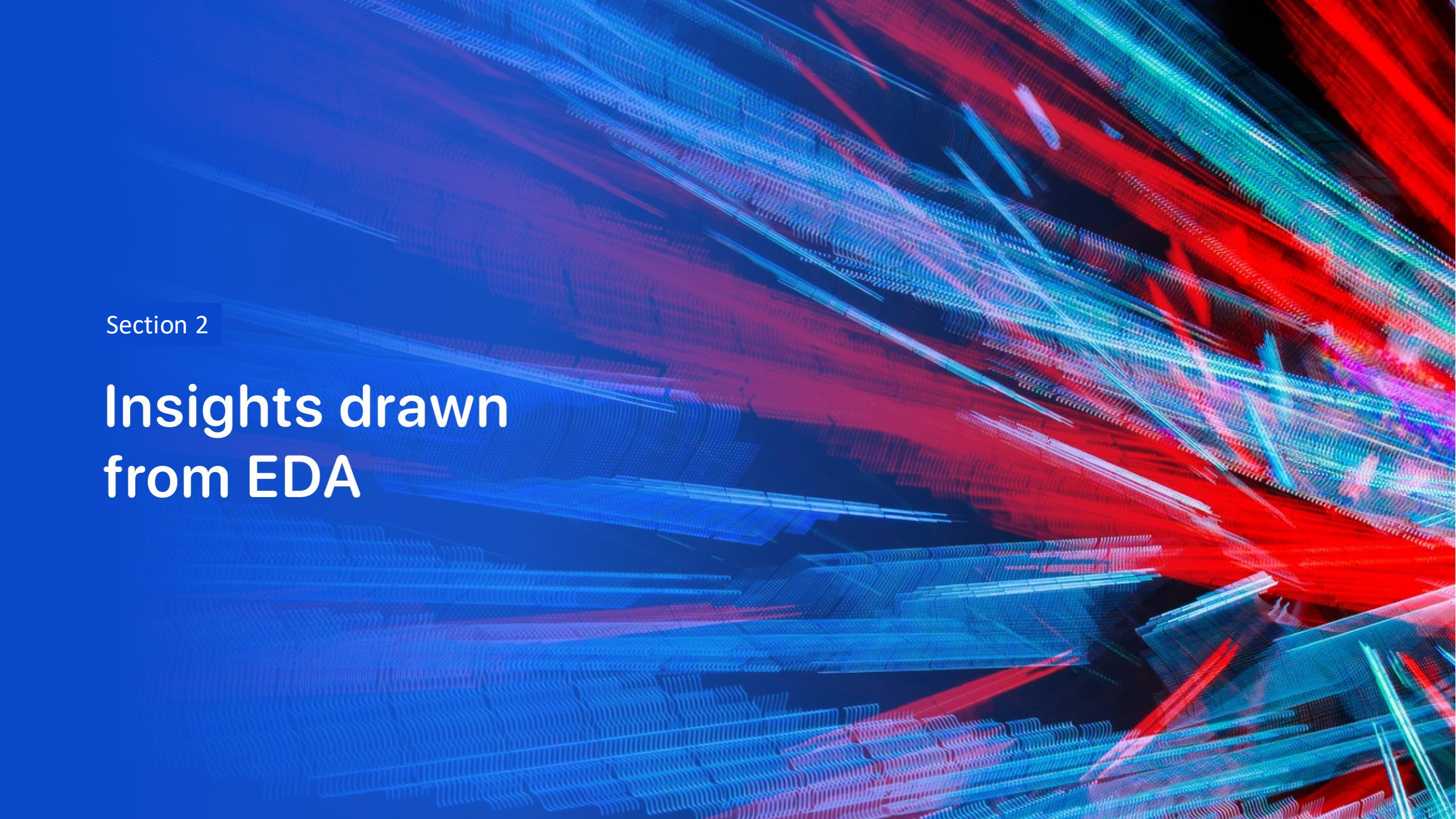
- Summarize what plots/graphs and interactions you have added to a dashboard
  - Created total success launch by site pie chart populates based on the site section in the dropdown
  - Created scatter plot to visualize the correlation between payload and class
- Explain why you added those plots and interactions
  - Helps the end user to check the success launches by site
- GitHub URL of your completed Plotly Dash lab: https://github.com/pmdatascience/testrepo/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

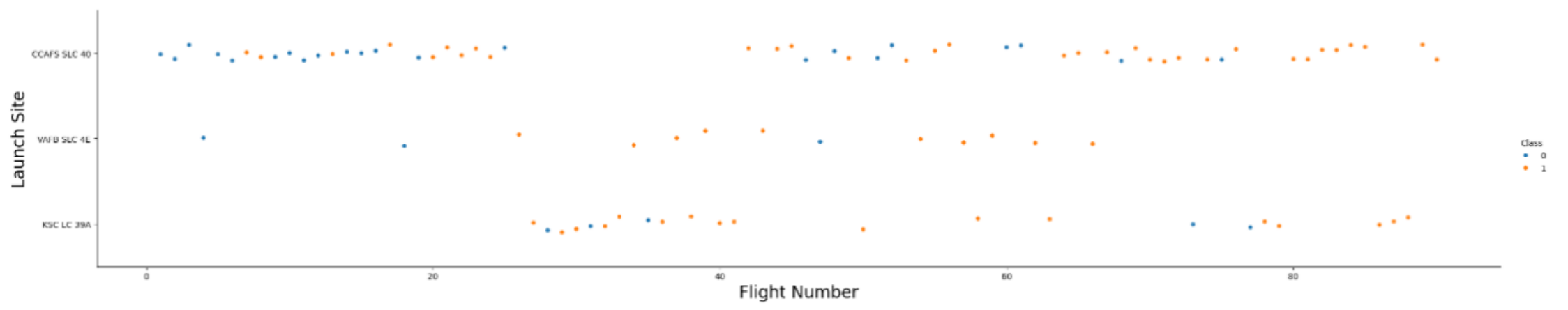
- Summarize how you built, evaluated, improved, and found the best performing classification model
  - Identified the best Hyperparameter for SVM, Classification Trees, and Logistic Regression
  - Calculated the method performance using the test data
  - Compare the results and Identify the best performing classification model
- GitHub URL of completed predictive analysis lab: https://github.com/pmdatascience/testrepo/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

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

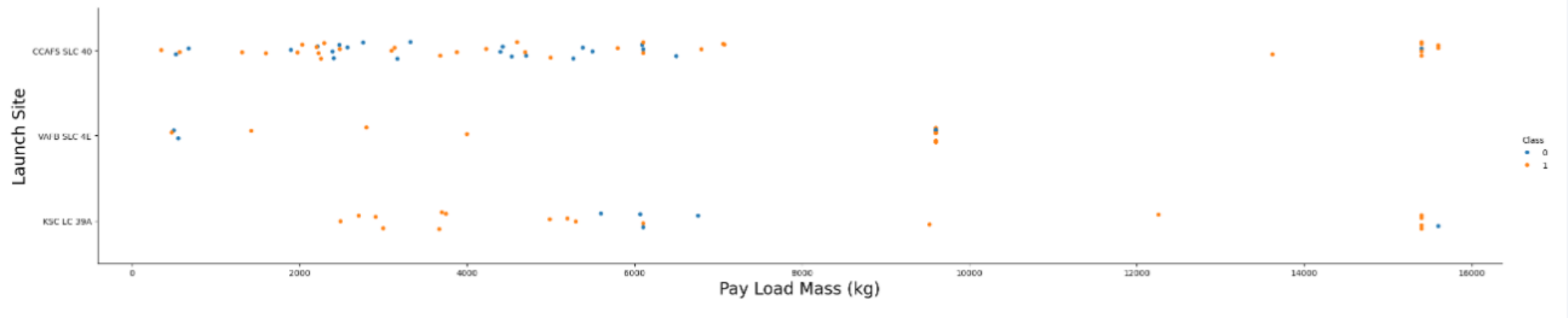

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- Explanation
  - More Flights launched from site CCAFS SLC 40
  - Launch Sites "VAFB SLC 4E" and "KSC LC 39A" have more successrate

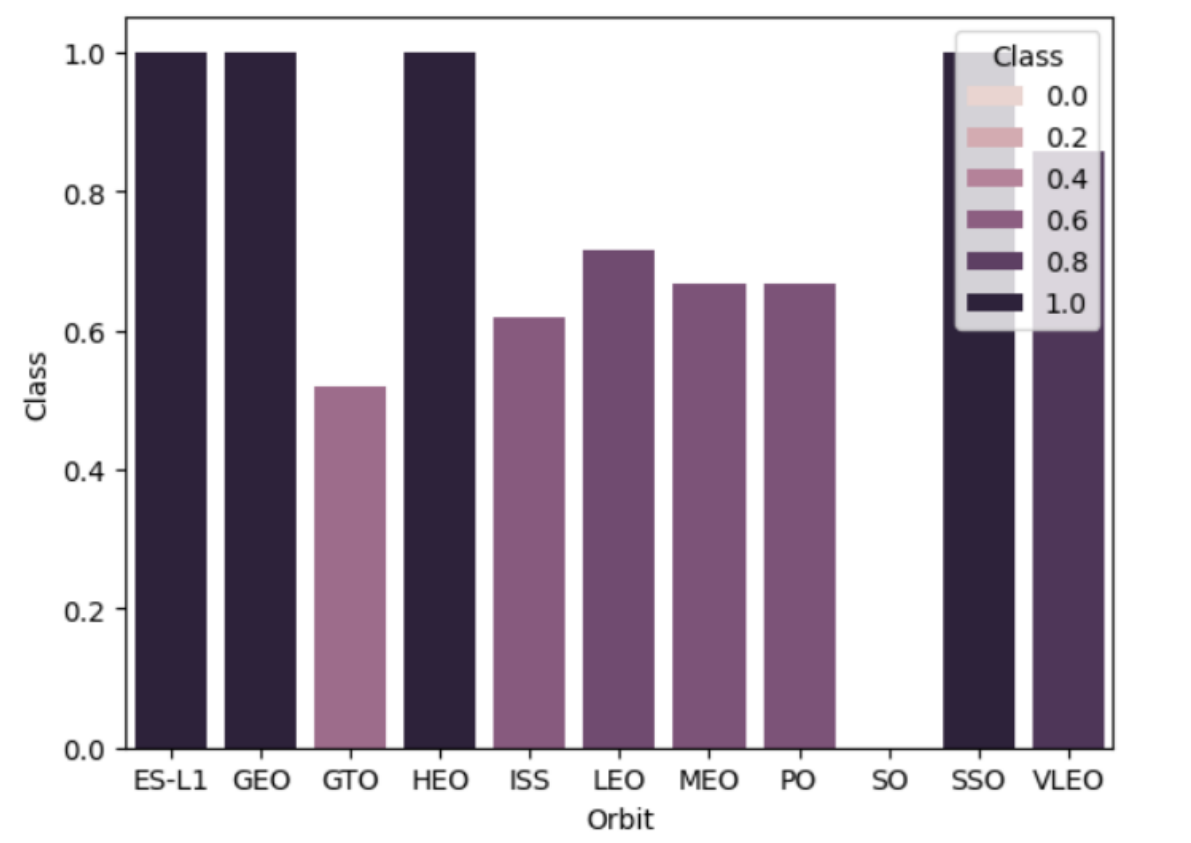
Payload vs. Launch Site



- Explanation

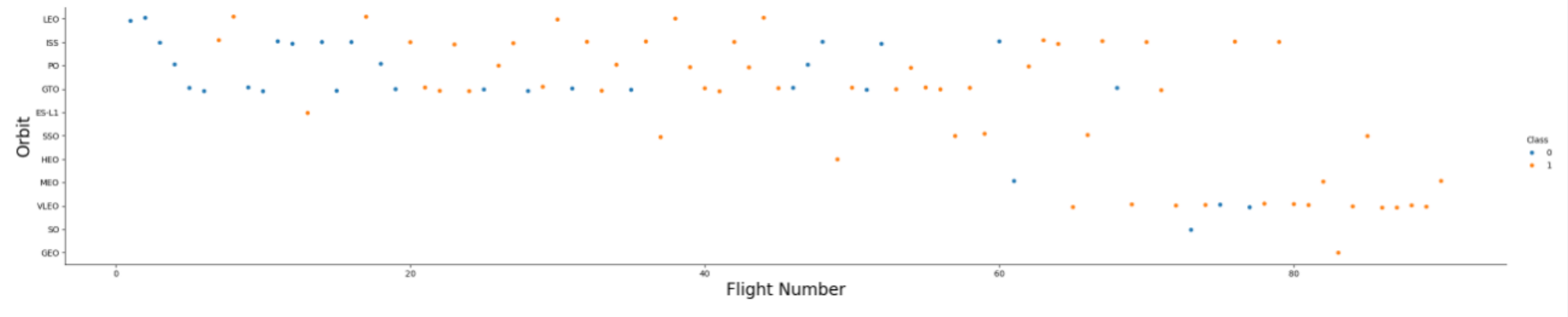
Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type



- Explanation
  - ES-L1, GEO, HEO, SSO orbit types have 100% success rate
  - SO, GTO orbit types have low success rate

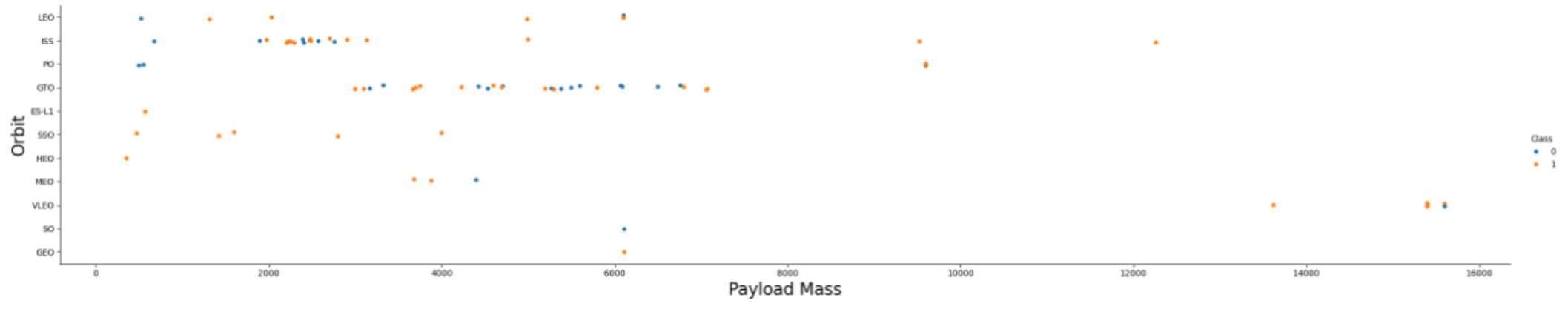
Flight Number vs. Orbit Type



- Explanation

You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type

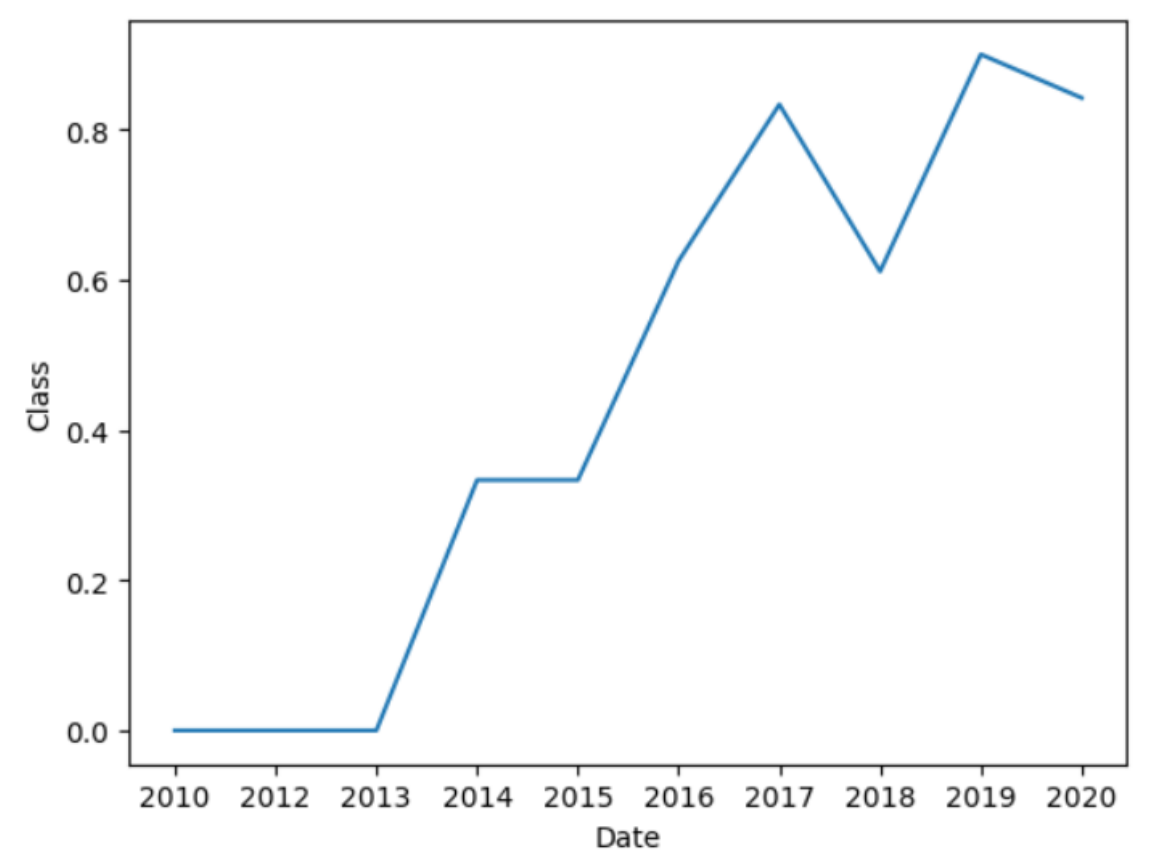


- Explanations

With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend



- Explanation
  - Most recent years have very high success rate

All Launch Site Names

- Below are the list of all Launch Sites

[12]:

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- Below are the five records where launch sites begin with `CCA`

[15]:

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

- Total 45,596 Kgs payload carried by boosters from NASA

```
[18]:  sum(PAYLOAD_MASS__KG_)
       45596
```

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 is approx 2535 kgs

```
[22]: avg(PAYLOAD_MASS__KG_)
      2534.6666666666665
```

First Successful Ground Landing Date

- First successful landing outcome on ground pad was achieved on Dec 22, 2015

min(date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Below are the list of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

[27]:

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Below are total number of successful and failure mission outcomes

[18]:

Mission_Outcome	count(*)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Below are the list of the boosters which have carried the maximum payload mass

[30]:

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

[19]:

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

[25]:

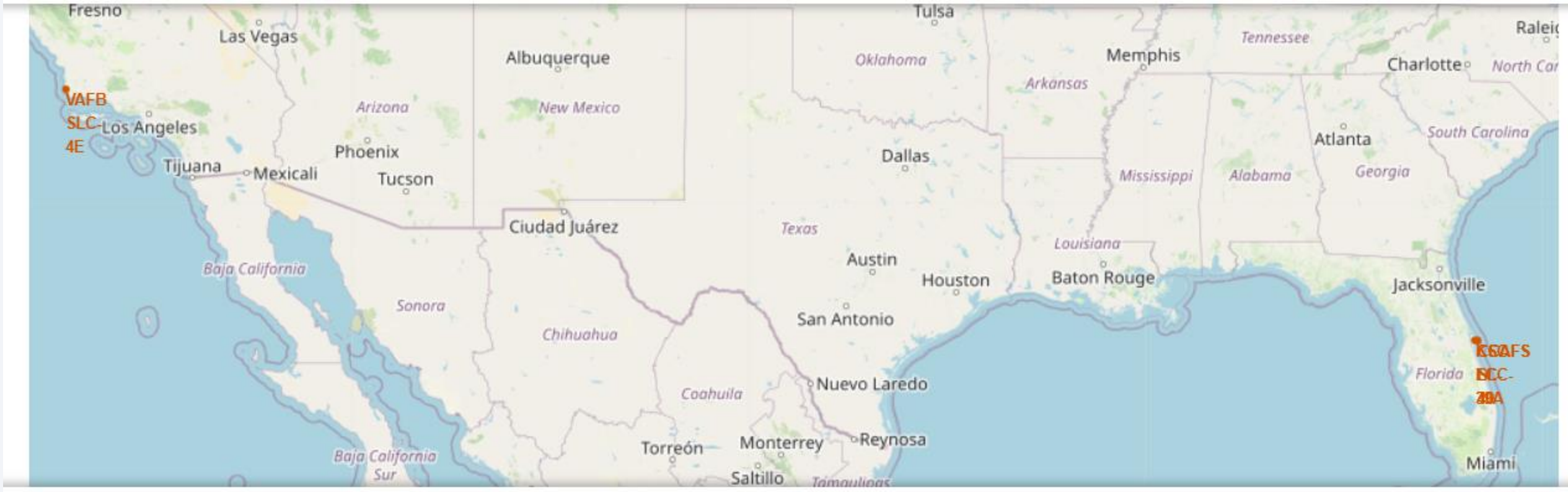
Landing_Outcome	count(*)	rank() OVER (order by count(*) desc)
No attempt	10	1
Success (drone ship)	5	2
Failure (drone ship)	5	2
Success (ground pad)	3	4
Controlled (ocean)	3	4
Uncontrolled (ocean)	2	6
Failure (parachute)	2	6
Precluded (drone ship)	1	8

Section 3

Launch Sites Proximities Analysis

Map of all launch sites

- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map

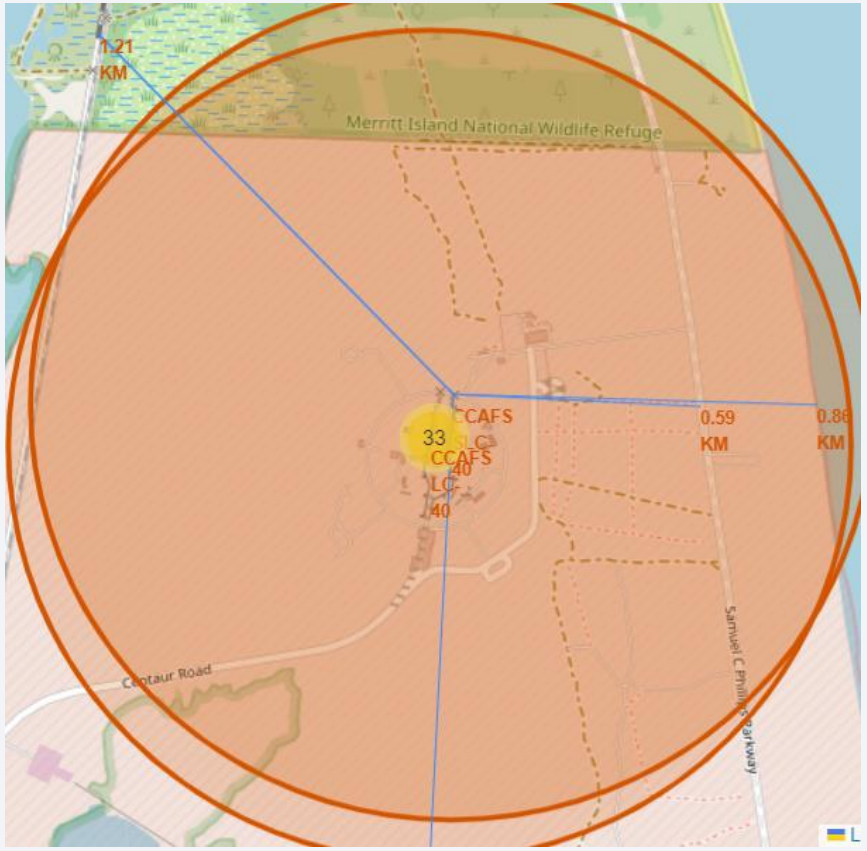


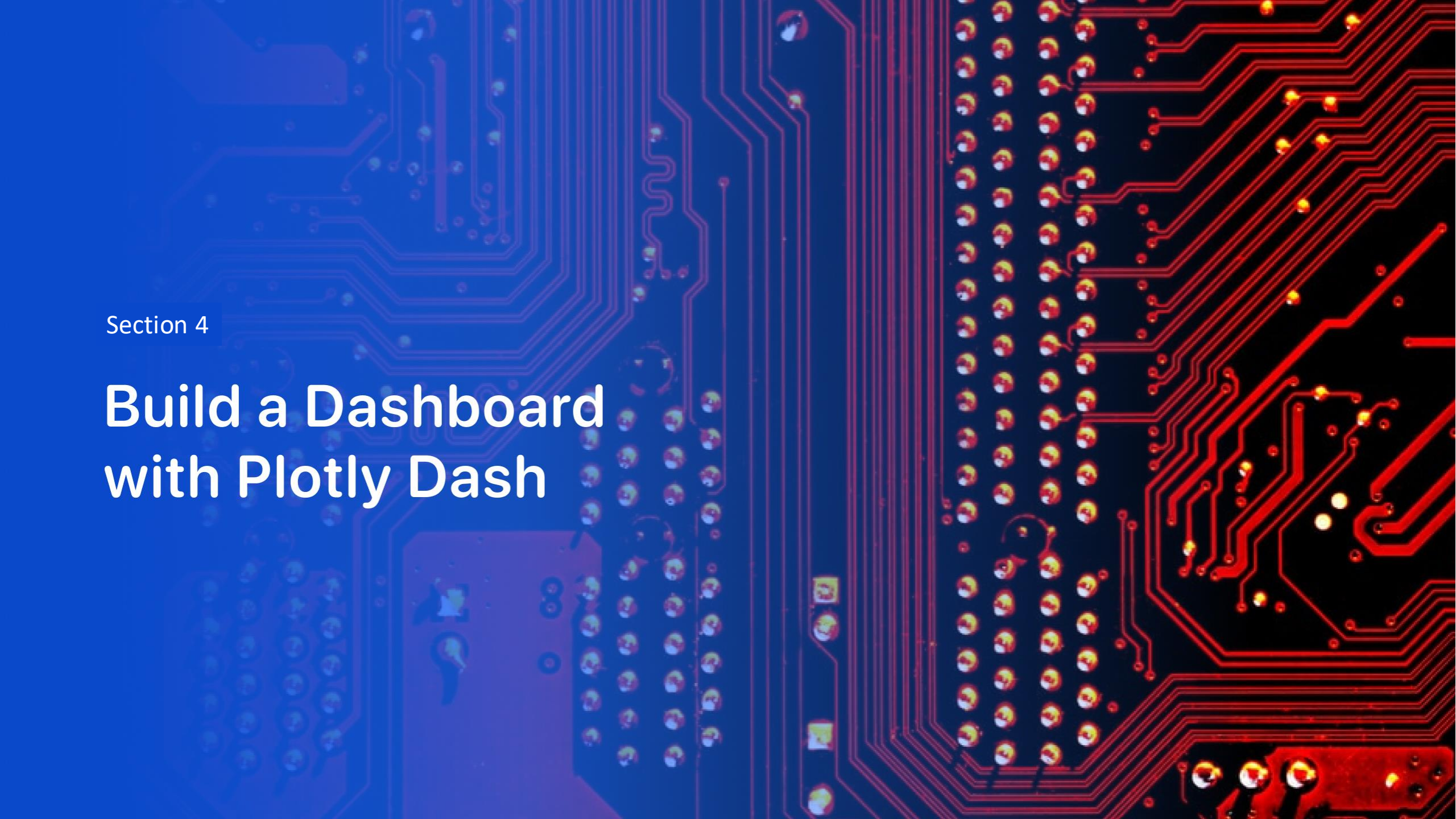
Launch Outcomes

- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Refer the map for color coded launch outcomes
  - Red - Failure
  - Green - Success

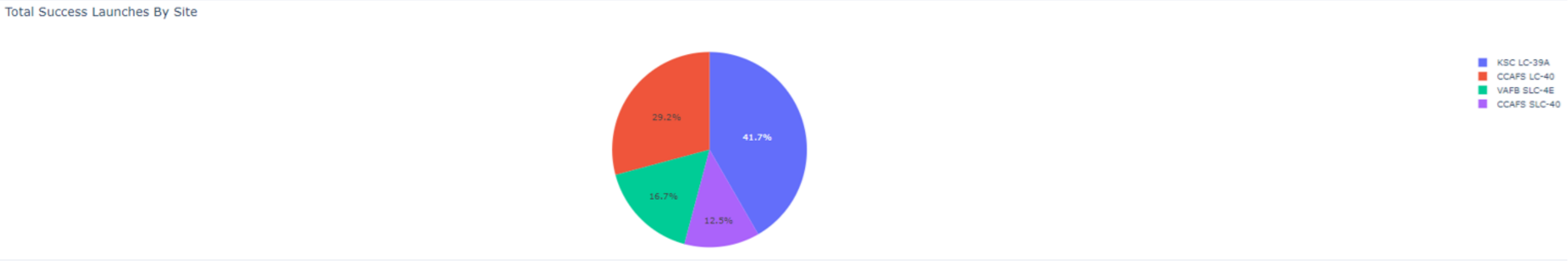
Proximity Map of Launch Site

- Refer map for a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed



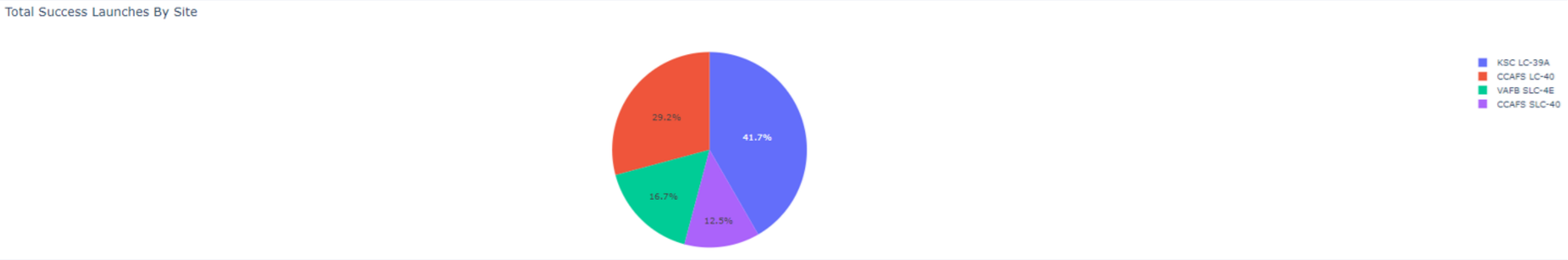

Section 4
Build a Dashboard with Plotly Dash

Total Success Launches by Site



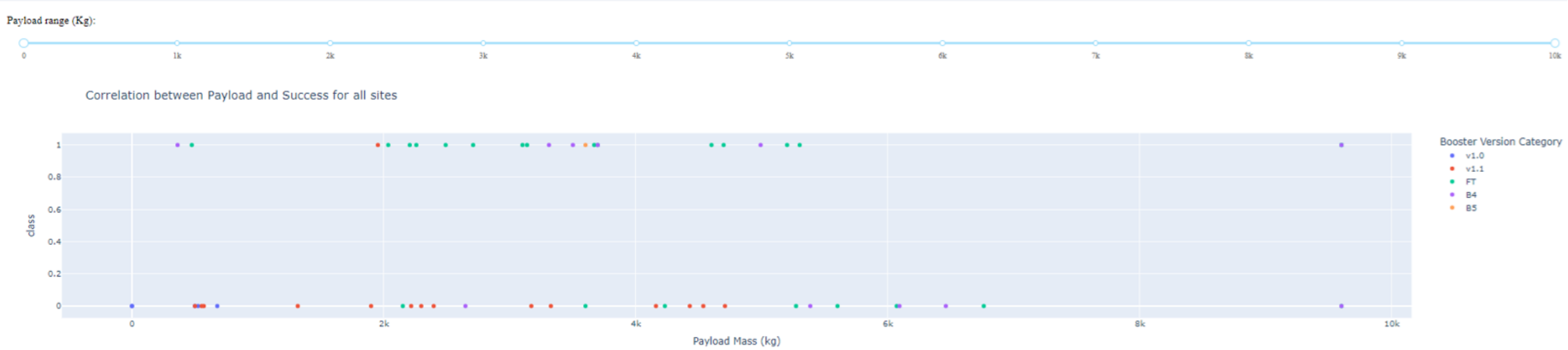
- Site "CCAFS SLC-40" has very low success rate

Launch Site with highest Success Rate



- Launch Site "KSC LC-39A" has highest success rate of 41.7%

<Dashboard Screenshot 3>



- 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.

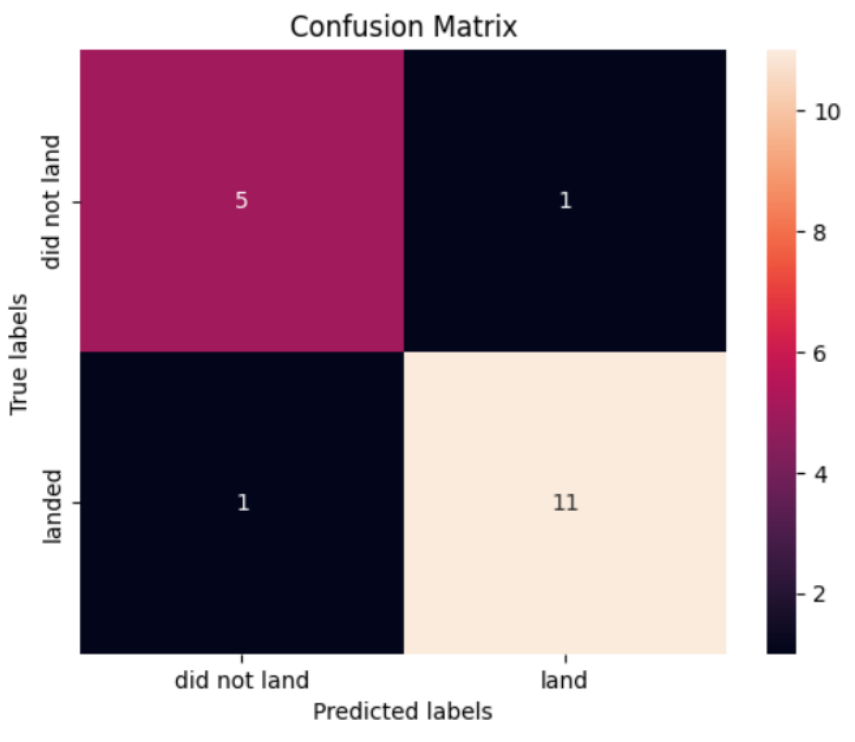
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Decision Tree classification modes has highest accuracy and prediction rate

Confusion Matrix



- From the above model confusion matrix it is identified that False Positive and False Negative are very low.

Conclusions

- Decision Tree Model predicts is good when compared to the other classification models
- Decision Tree Model prediction rate is 88.93%

Appendix

- Github: https://github.com/pmdatascience/testrepo

Thank you!