IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sathesh Muppudathi E

15/11/2023

Outline

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

Executive Summary

- Summary of methodologies
  - The data collection methodology, data wrangling, exploratory data analysis with SQL and visualization, interactive visual analytics, and predictive analysis
- Summary of all results
  - The results of Exploratory data analysis, Interactive analysis and Predictive analysis

Introduction

- The commercial space age is here, companies are making space travel affordable for everyone. The most successful commercial space agency is SpaceX. One reason SpaceX can do this is the rocket launches are relatively inexpensive.

- 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.

- Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Launch data collect using SpaceX REST API and Falcon 9 launch records collect using Beautifulsoup.
- Perform data wrangling
  - Calculating occurrence of orbits and Creating Landing outcome column.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Scikit-learn classification models tune using GridSearchCV and evaluate classification models using metrics.

Data Collection

- The data sets were collected using two methods. They are

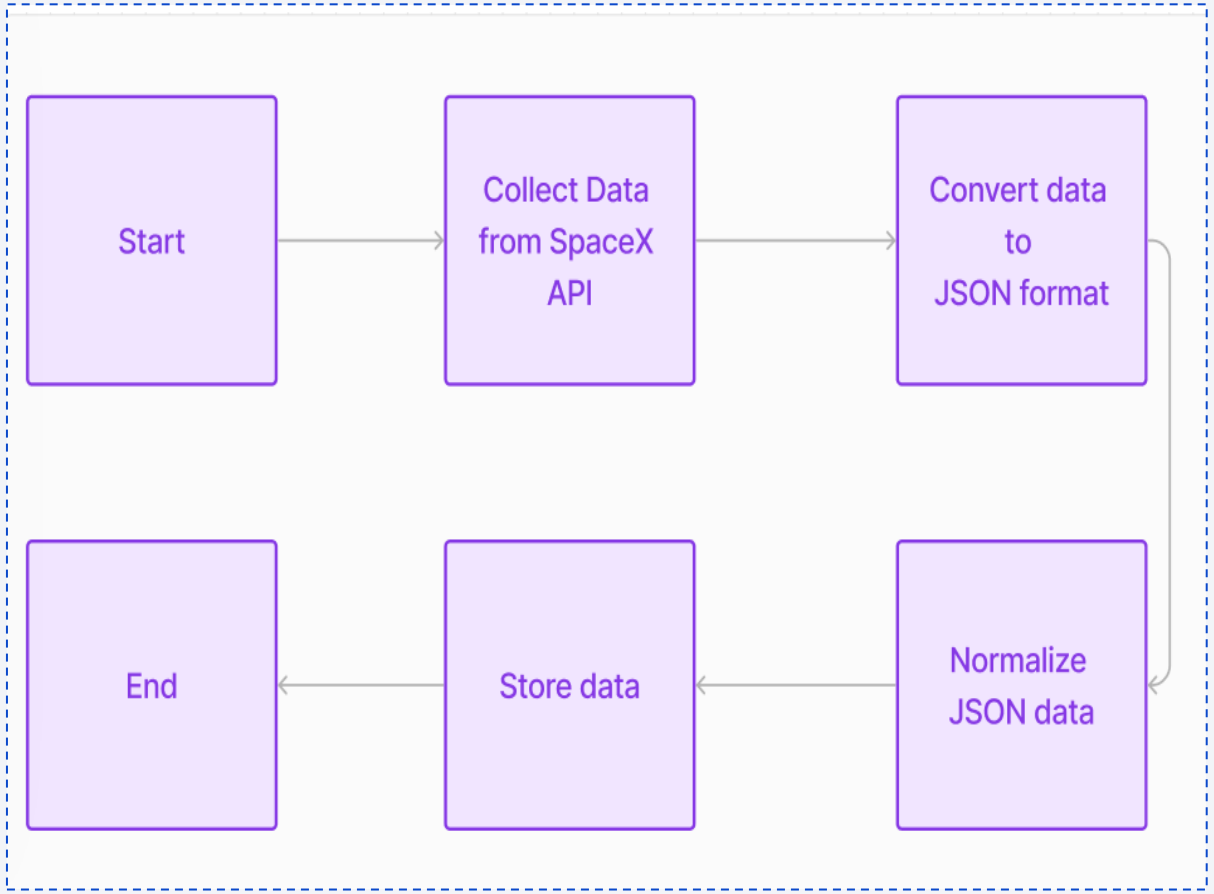
  1) SpaceX REST API used for Launch data collection

  2) Beautifulsoup Web Scraping used for Falcon 9 launch records collection

Data Collection – SpaceX API

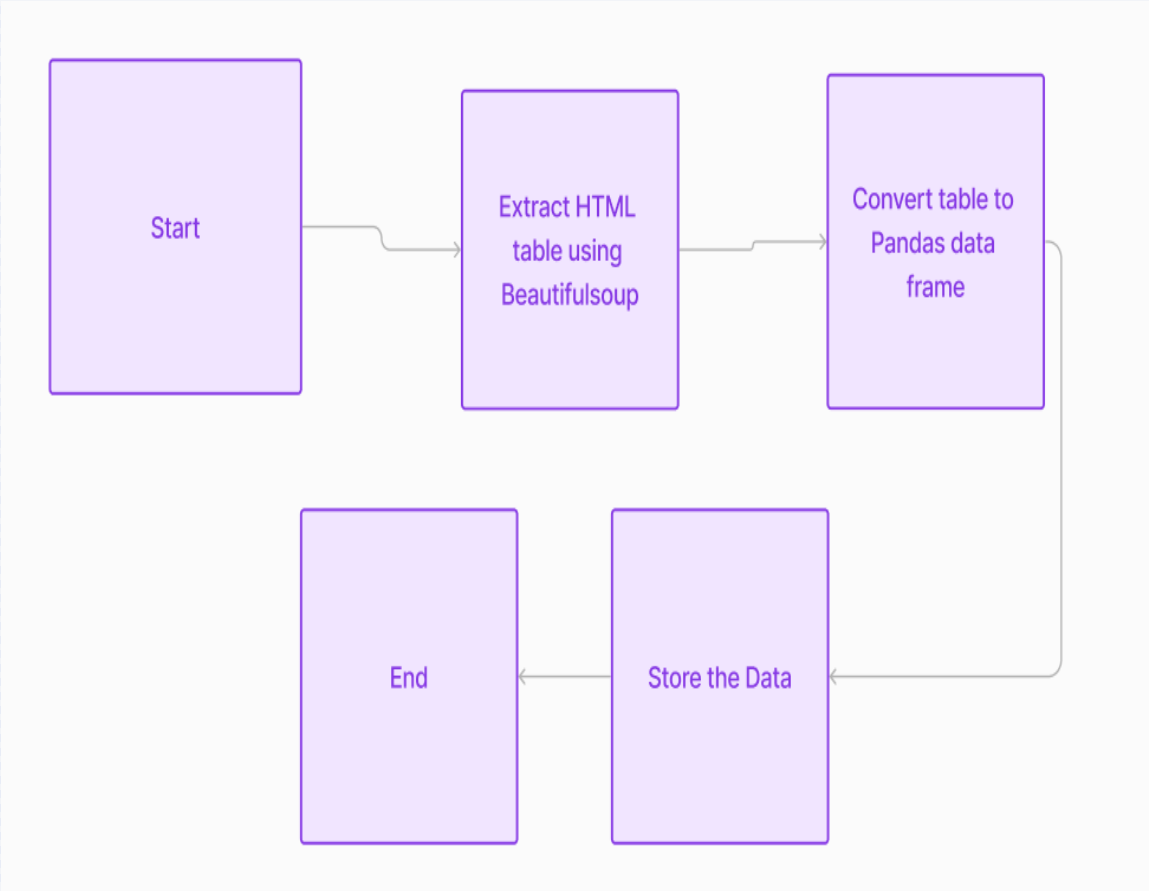
- Request and Collect data from SpaceX API in the form of Json data then json_normalize the data using pandas.
- Filtered the data for Falcon 9 and stored the data

- https://github.com/sathish-mass/SpaceX-Launch-Success-Classification/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



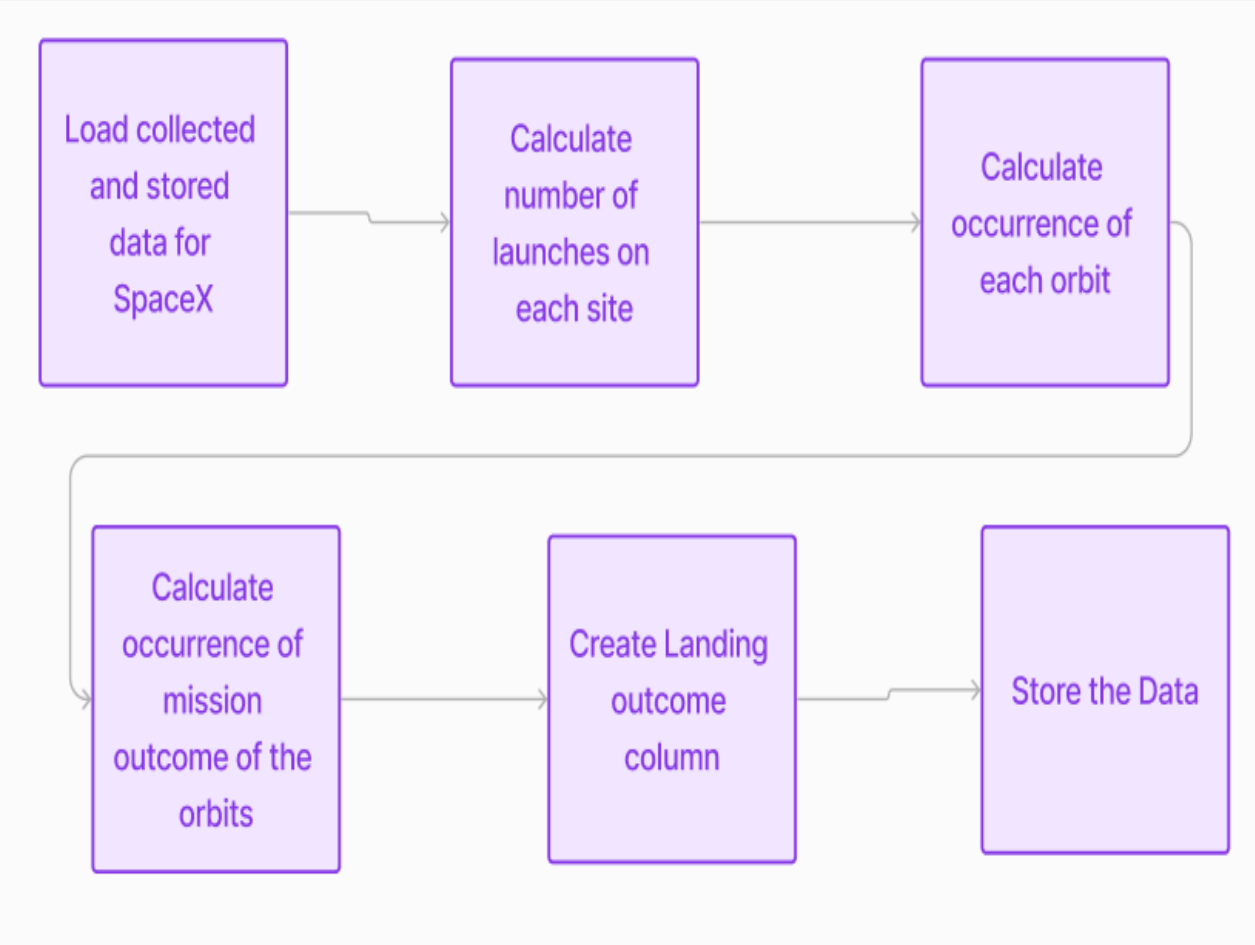
Data Collection - Scraping

- Using Beautifulsoup extract a Falcon 9 launch records HTML table from Wikipedia.
- Then Parse the table and convert it into a Pandas data frame
- https://github.com/sathish-mass/SpaceX-Launch-Success-Classification/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Loaded the collected and stored data for SpaceX
- Calculated the number of launches on each site, occurrence of each orbit, occurrence of mission outcome of the orbits and Created the Landing outcome column.
- https://github.com/sathish-mass/SpaceX-Launch-Success-Classification/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- In EDA with Data Visualization using matplotlib and seaborn has been visualized using four different plots.
- One is categorical plot for that show the relationship between a numerical and one or more categorical variables.
- It shown how the Flight Number (indicating the continuous launch attempts) and Payload variables would affect the launch outcome. It also shown the relationship between Flight Number and Launch Site would affect the launch outcome

EDA with Data Visualization

- Second is scatter plot used to shown the relationship between how the Payload and Launch Site, Flight Number and Orbit type, Payload and Orbit type would affect the launch outcome
- Third is Bar chart used to shown the success rate of each orbit type.
- Final one is Line plot used to shown the launch success yearly trend.

- https://github.com/sathish-mass/SpaceX-Launch-Success-Classification/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- In EDA with SQL performed the following queries.
- Displayed the names of the unique launch sites in the space mission.
- Displayed 5 records where launch sites begin with the string 'CCA'.
- Displayed the total payload mass carried by boosters launched by NASA (CRS).
- Displayed average payload mass carried by booster version F9 v1.1 .
- Listed the date when the first successful landing outcome in ground pad was achieved.
- Listed the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

EDA with SQL

- Listed the total number of successful and failure mission outcomes.
- Used subquery to List the names of the booster versions which have carried the maximum payload mass.
- Listed the records which will display the month names, failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015.
- Ranked 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.
- https://github.com/sathish-mass/SpaceX-Launch-Success-Classification/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Folium is a powerful Python library that helps to create several types of Leaflet maps. It has map objects such as markers, circles, lines, etc.
- The Map object created for the Launch site coordinates. It shows the launch sites in real time maps. The Map object helps to add further to make markers, circles, lines, etc.
- The Folium circle helps to add a highlighted circle area with a text label on a specific coordinate. It helps to create circle for space centers and Launch sites.
- The Folium marker helps to add a markers at specific locations with appearance and popup. It helps to create a markers for Launch sites.

Build an Interactive Map with Folium

- The Folium Marker Cluster object helps to form a clusters for the specific coordinates. It helps to cluster the Launch sites based on success rate.
- The Mouse Position on the map helps to get coordinate for a mouse over a point on the map. It helps to find the coordinates of any places near the Launch sites.
- The Folium Polyline helps to draw a line between two Locations coordinates. It helps to draw a line between Launch sites and nearest locations.

- https://github.com/sathish-mass/SpaceX-Launch-Success-Classification/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Dash is Open Source Framework. Plotly stewards Python's leading data visualization and UI libraries.
- Dash Core Components (DCC) Dropdown is a component for rendering a user-expandable dropdown. It helps create a options for user to select from Launch sites to show the graphs about.
- Plotly pie chart is a circular statistical chart, which is divided into sectors to illustrate numerical proportion. It helps to visualize the relationship between Launch sites and outcome.

Build a Dashboard with Plotly Dash

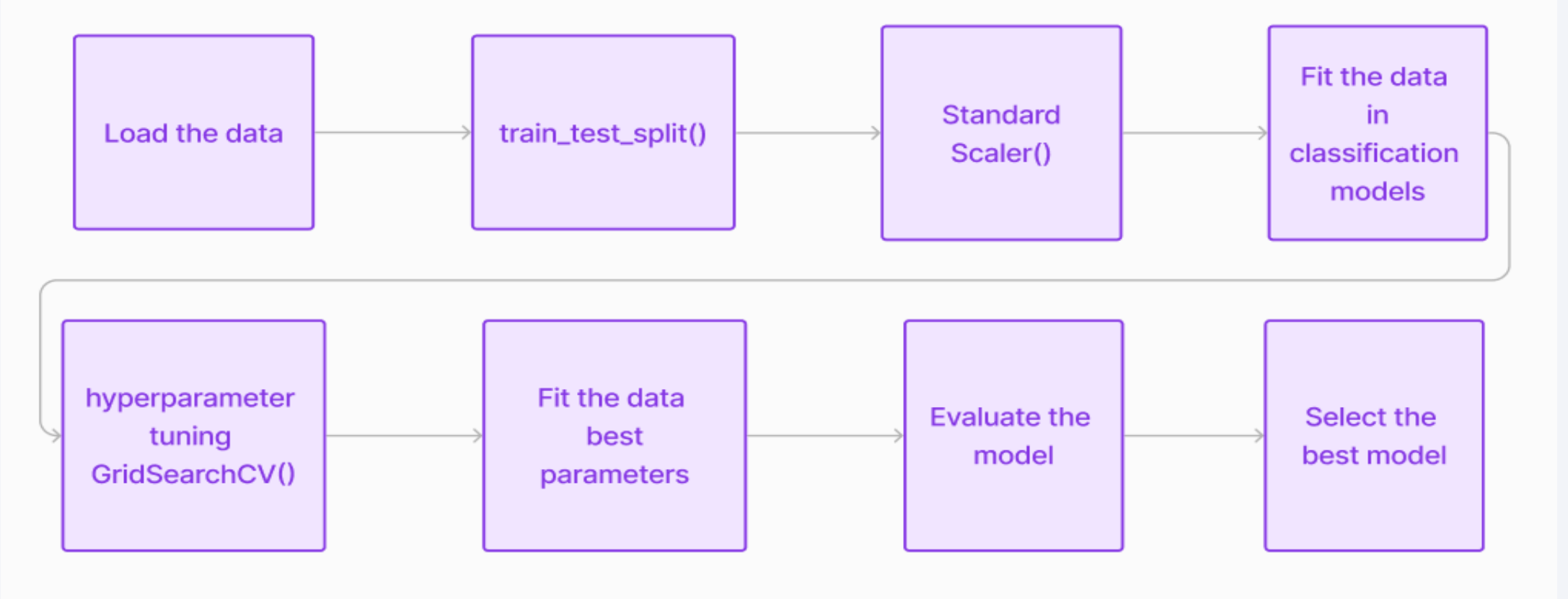
- DCC Range Slider is a component for rendering a range slider. Users interact with a DCC Range Slider by selecting areas on the rail or by dragging handles. It helps to render the ranges of Payload mass (Kg).
- Plotly Express scatter plot helps to plot the relationship between the Payload Mass and outcome based on Booster Versions.
- https://github.com/sathish-mass/SpaceX-Launch-Success-Classification/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Loaded the data after the EDA with data wrangling and Analysis, then split the data for to avoid overfitting using train_test_split(). After that standardize the data using StandardScaler().
- The standardized data ready to fit in the classification models such as Logistic Regression, SVM, KNN, Decision Tree Classifier with the help of Hyper parameter tuning using GridSearchCV() for finding the best parameters.
- The model evaluated using the confusion matrix and test data accuracy score.
- The best model selected from based on this confusion matrix and accuracy scores.
- https://github.com/sathish-mass/SpaceX-Launch-Success-Classification/blob/main/SpaceX_Machine_Learning_Prediction.ipynb

Predictive Analysis (Classification)

Flow Chart



Results

Exploratory data analysis results

- The flight number increases, the first stage is more likely to land successfully. The more massive the payload, the less likely the first stage will return.
- Different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.
- The VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).
- The orbit type ES-L1, GEO, HEO, SSO, and VLEO has high success rate. The orbit type SO has low success rate.

Results

Exploratory data analysis results

- 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.
- 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.
- The success rate since 2013 kept increasing till 2020.

Results

Exploratory data analysis results

- The names of the unique launch sites are CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, and CCAFS SLC-40.
- From Launch site CCAFS LC_40 major mission outcomes are success though still failed 2 times to land and 3 times no landing attempt has made. The mission major orbit type is LEO(ISS).
- The total payload mass carried by boosters launched by NASA (CRS) is 45596 kg.
- The average payload mass carried by booster version F9 v1.1 is 2928.4 Kg.
- The dates of the first successful landing outcome on ground pad is 22-12-2015

Results

Exploratory data analysis results

- The boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are F9 FT B1022, F9 FT B1026, F9 FT B1021.2, and F9 FT B1031.2
- The total number of successful and failure mission outcomes are 101 including 1 failure and 100 success outcomes.
- The maximum payload mass is 15600 Kg.
- The booster which have carried the maximum payload mass are 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

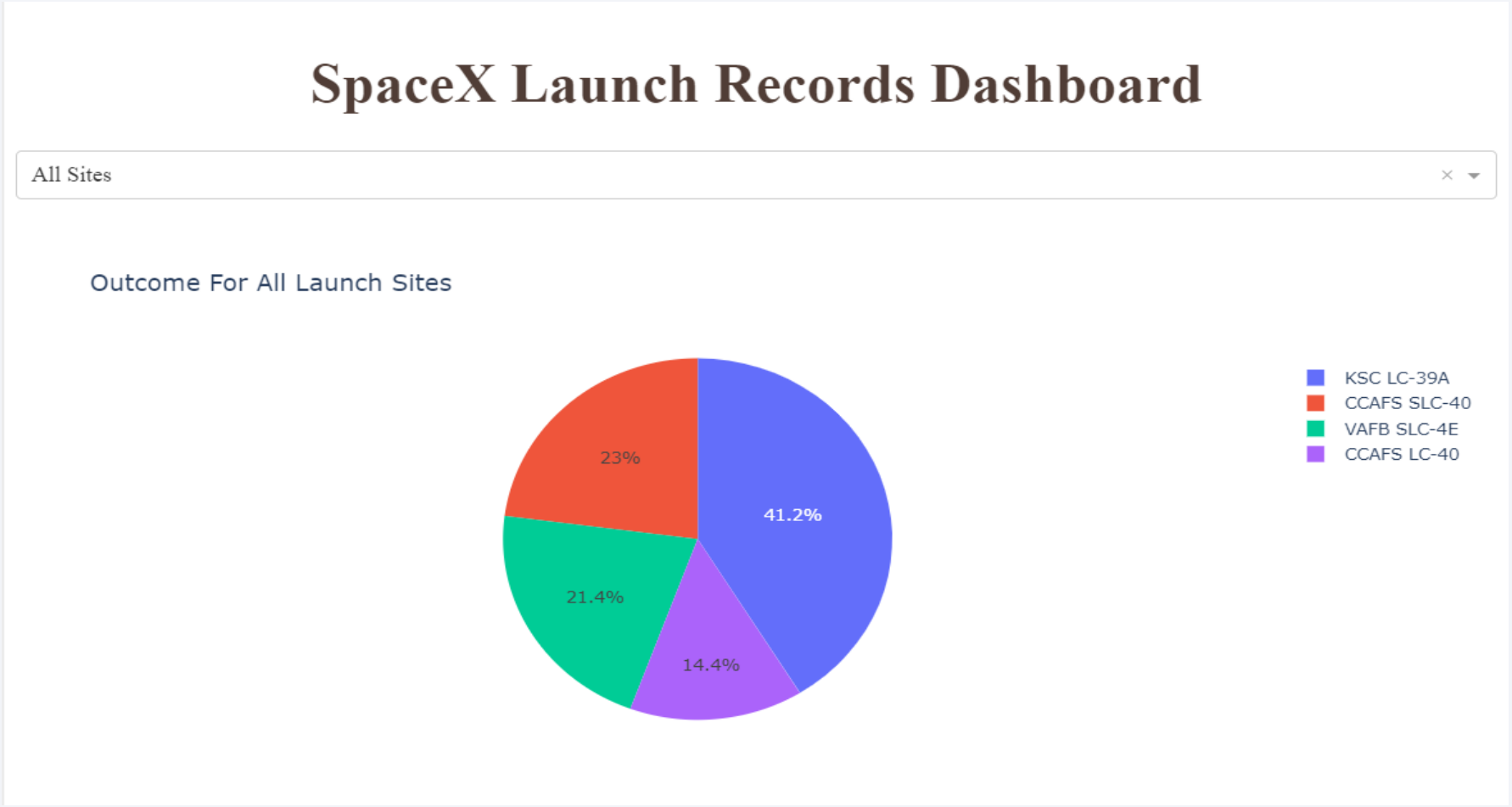
Results

Exploratory data analysis results

- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015 are F9 v1.1 B1012 Booster in Launch site CCAFS LC-40 and F9 v1.1 B1015 Booster in Launch site CCAFS LC-40 .
- 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 are Success (ground pad) with count of 5 and Failure (drone ship) with count of 5.

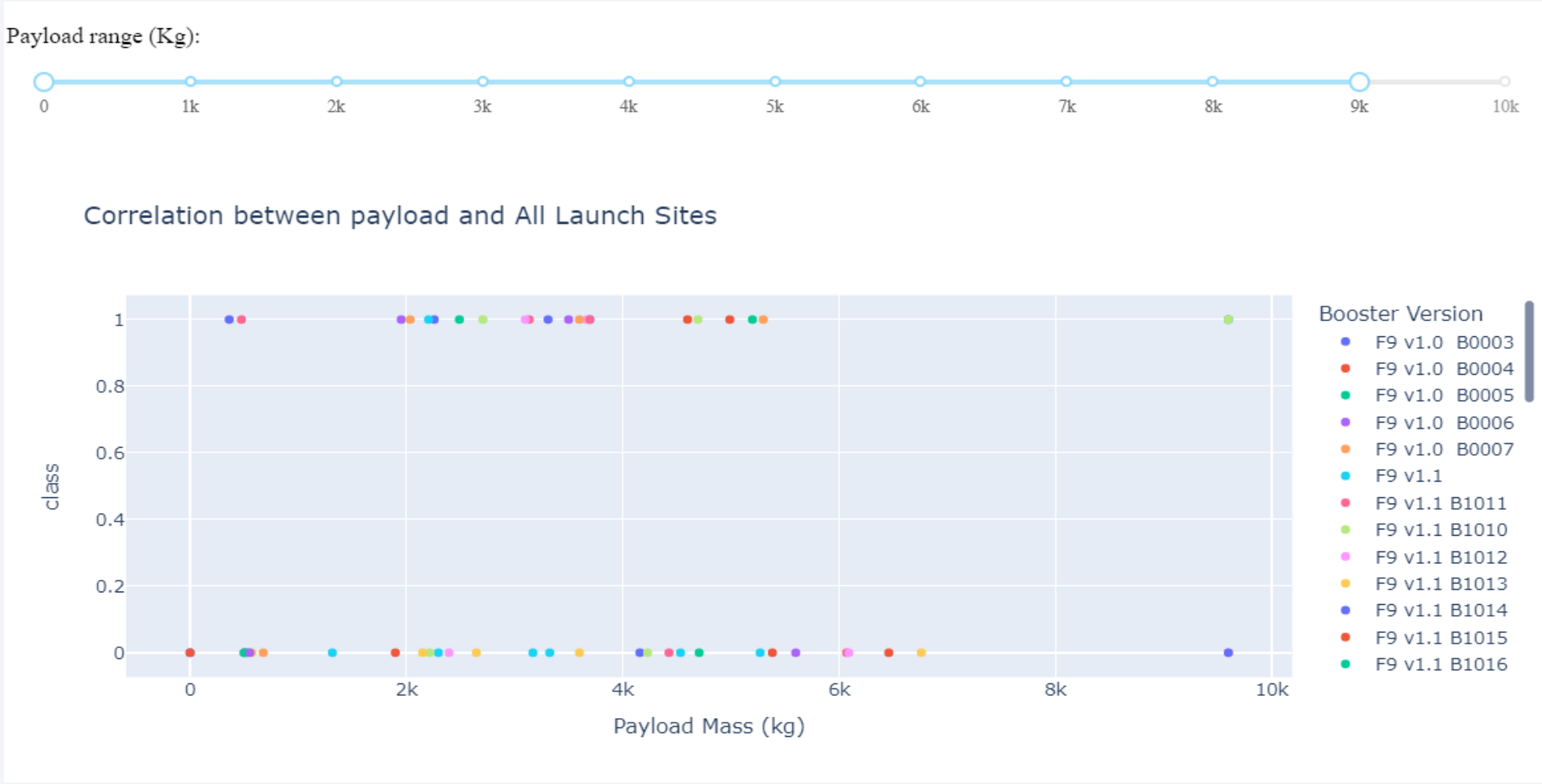
Results

Interactive analytics demo



Results

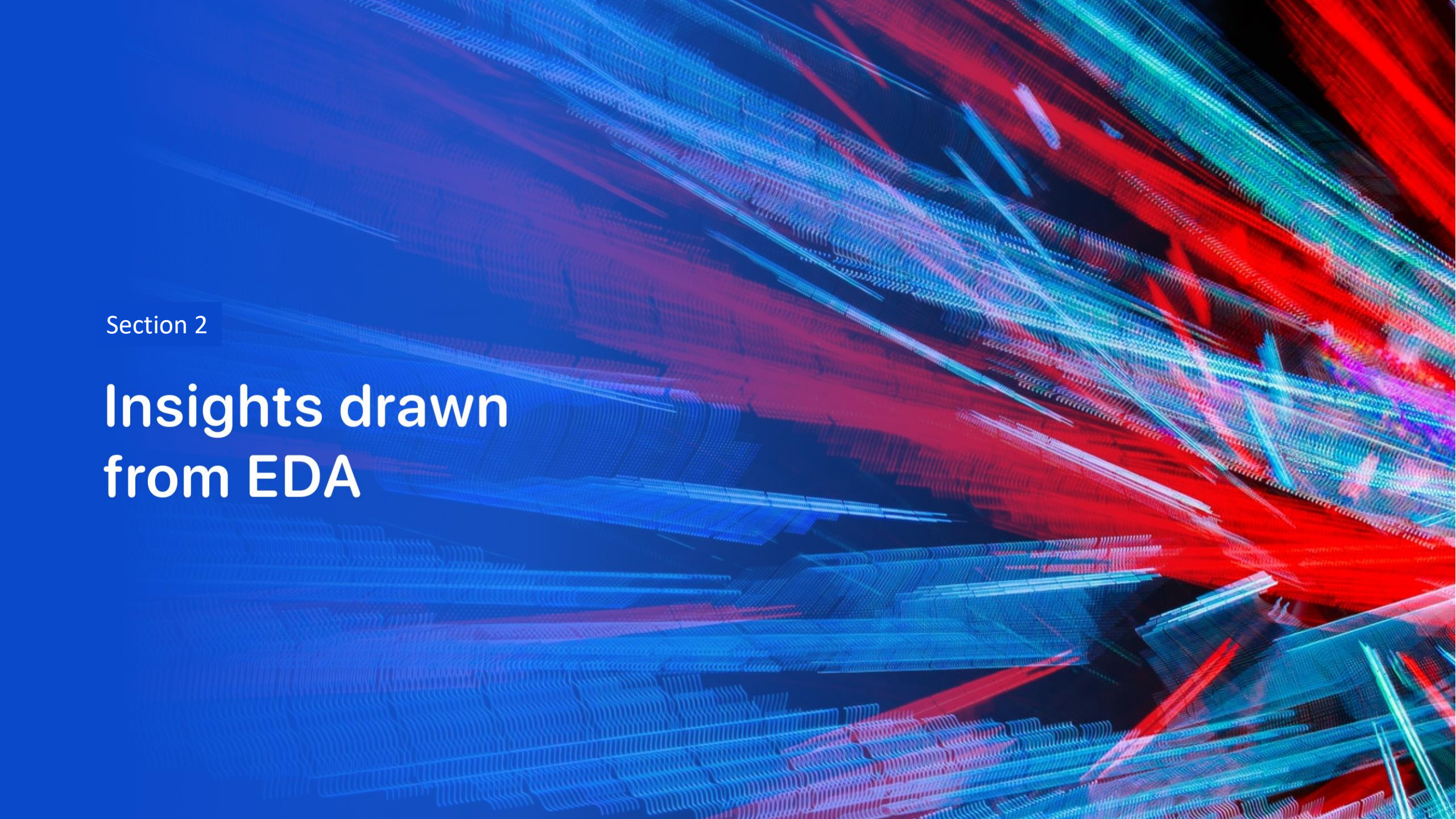
Interactive analytics demo



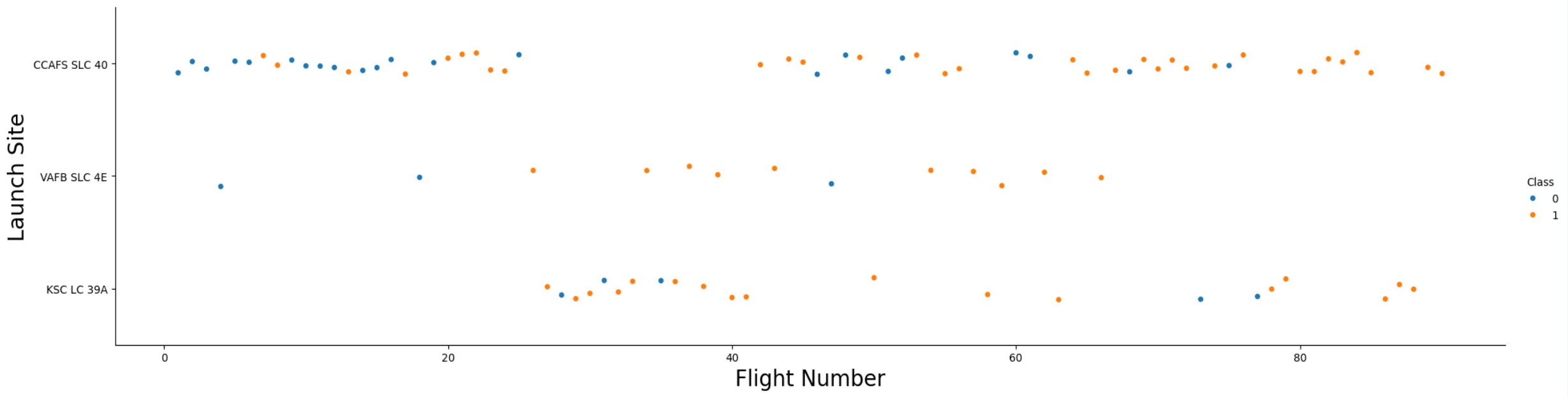
Results

Predictive analysis results

- The tree model performs very well with the data for classification. Decision Tree has the highest accuracy for classification with 88.89% .

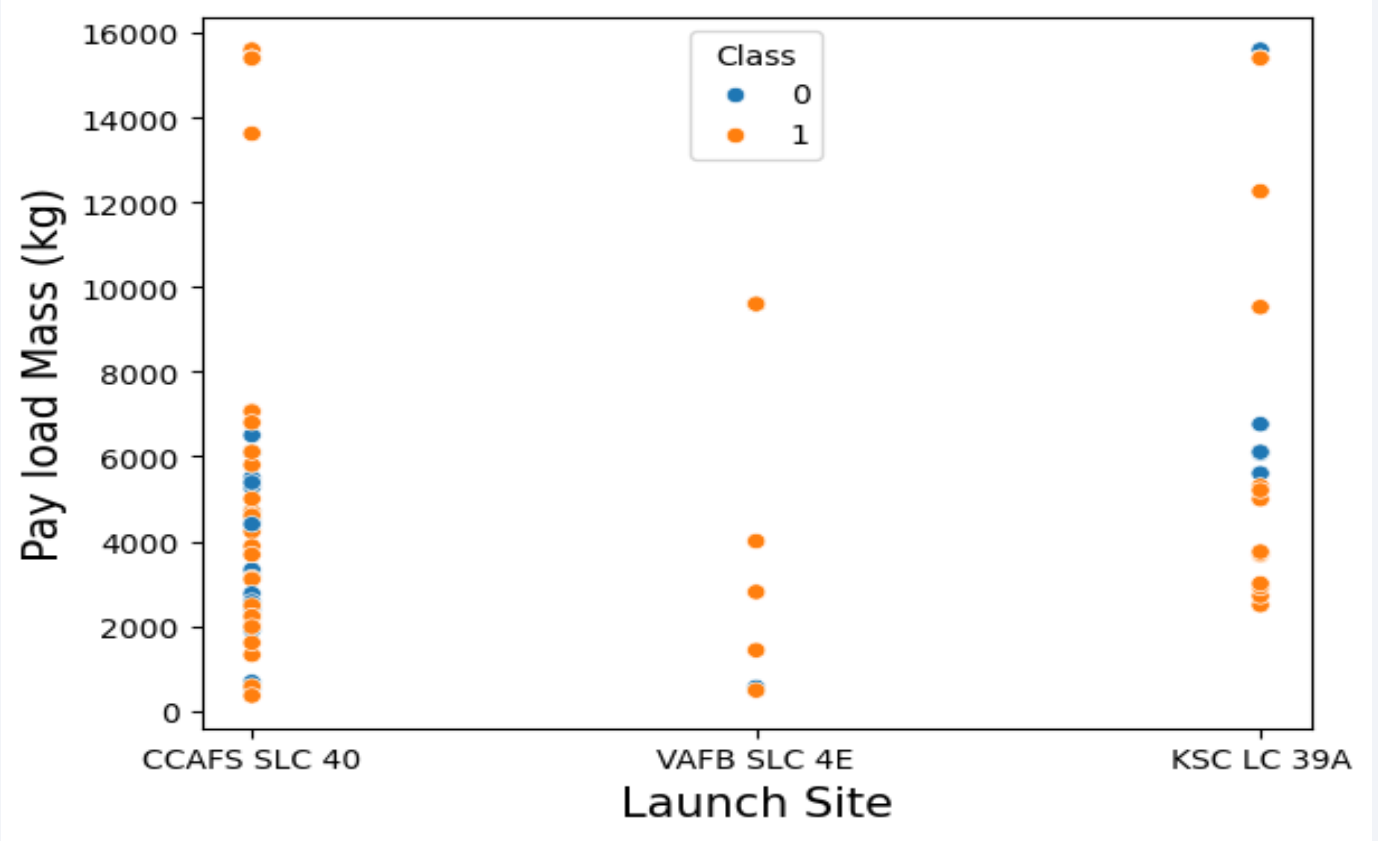

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



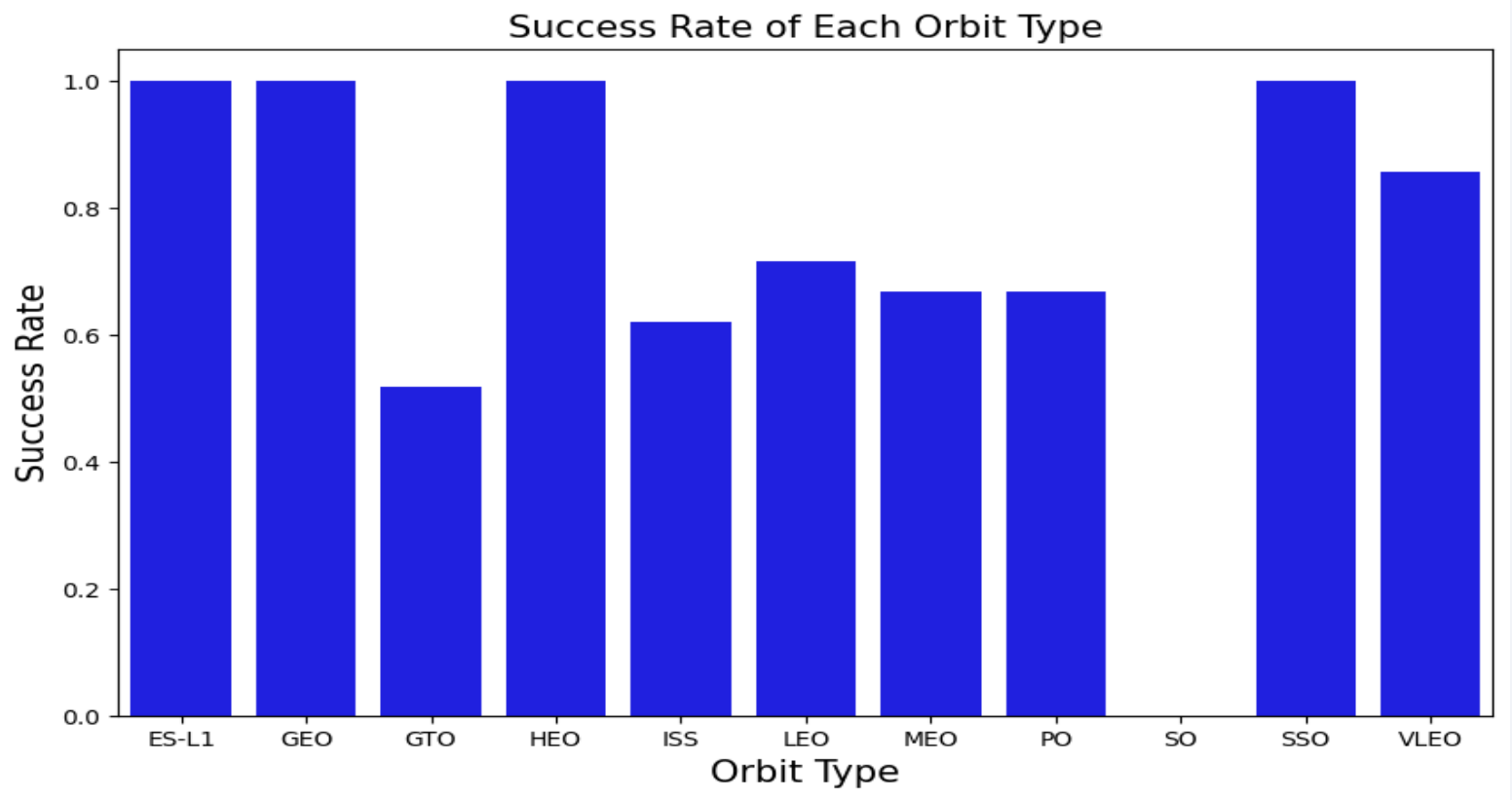
- The different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%

Payload vs. Launch Site



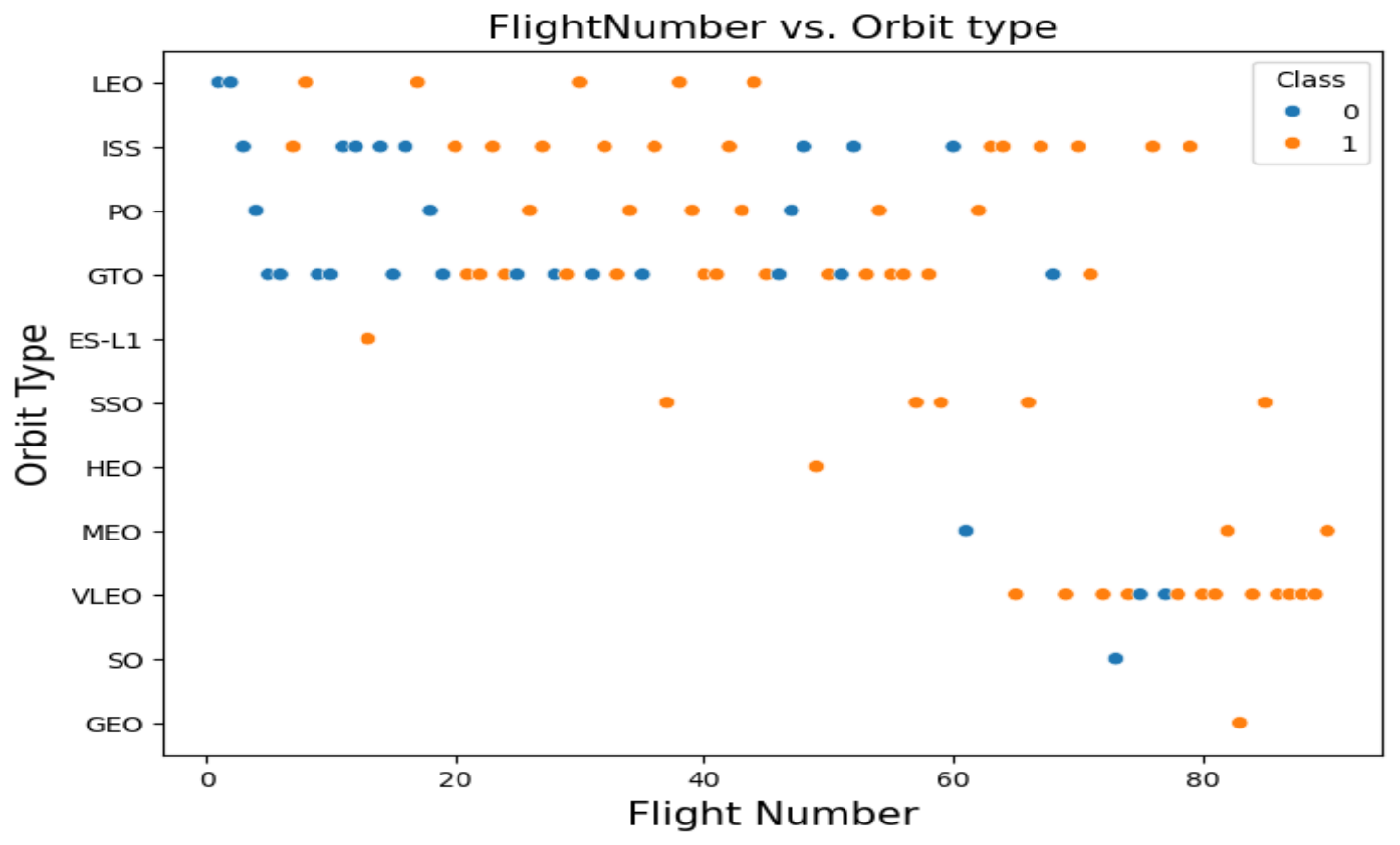
- The VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).

Success Rate vs. Orbit Type



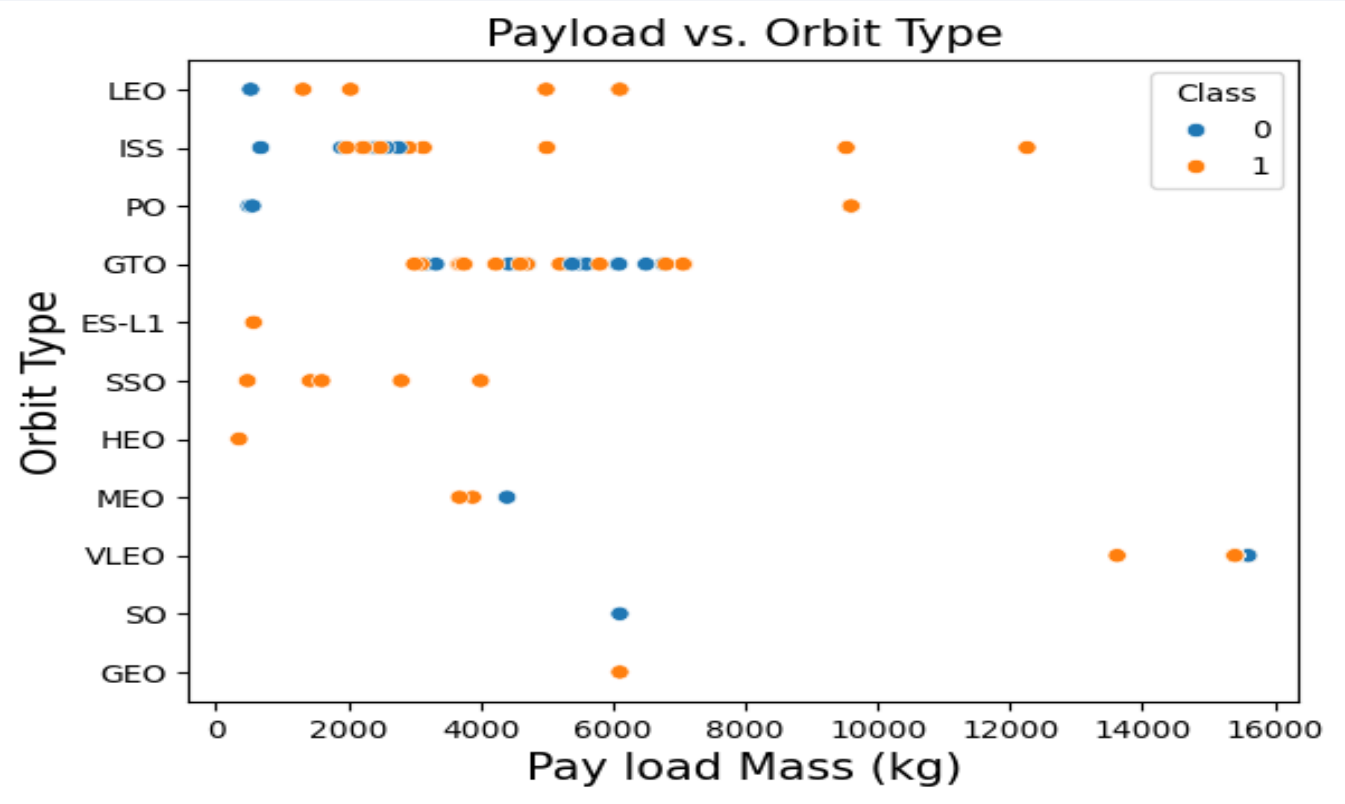
- The orbit type ES-L1, GEO, HEO, SSO, and VLEO has high success rate. The orbit type SO has low success rate.

Flight Number vs. Orbit Type



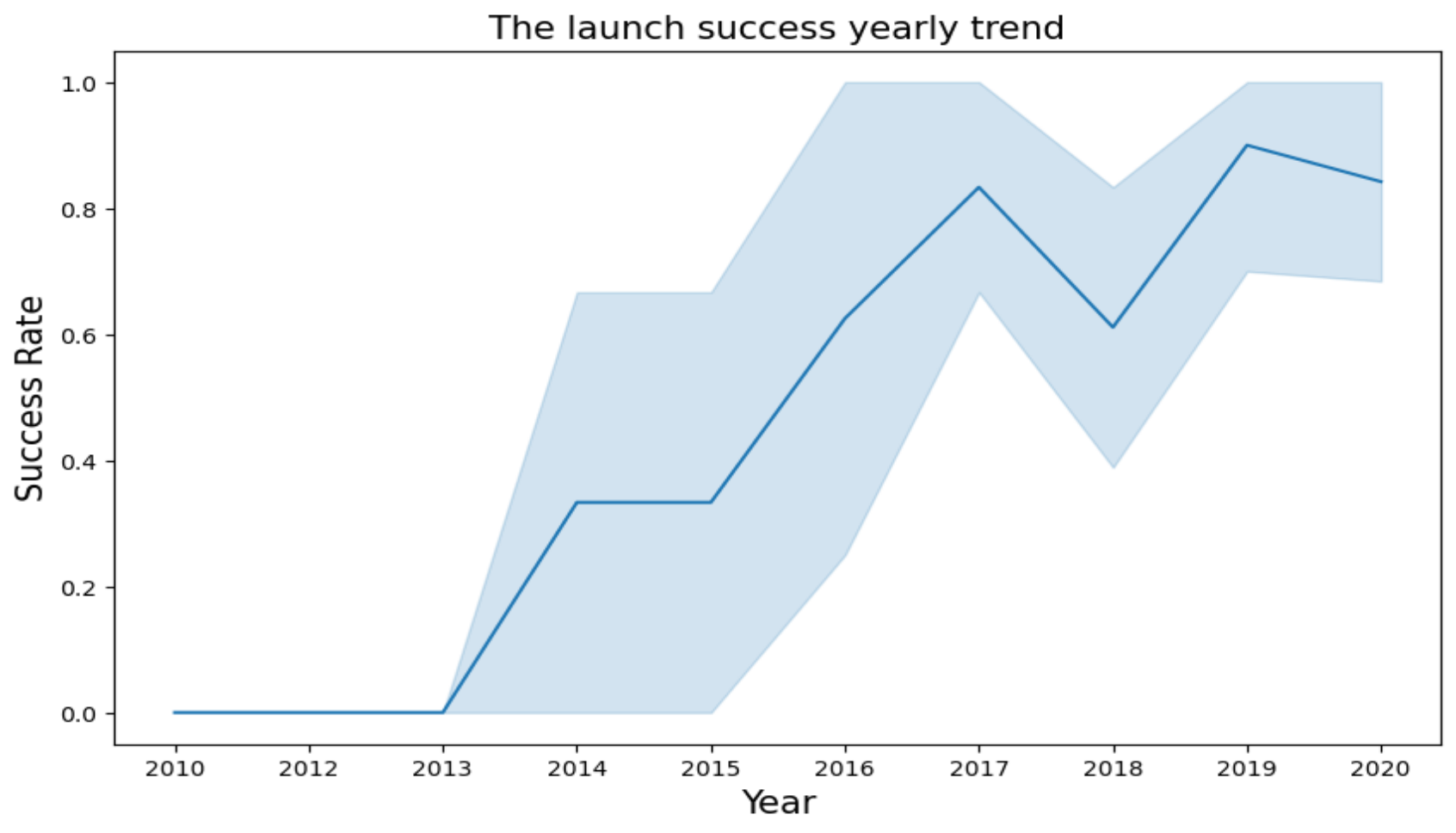
- The LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



- 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



- The success rate since 2013 kept increasing till 2020.

All Launch Site Names

- The names of the unique launch sites are
    1. CCAFS LC-40
    2. VAFB SLC-4E
    3. KSC LC-39A
    4. CCAFS SLC-40

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

- From the launch sites these four sites are major launches happening here.

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
6/4/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
12/8/2010	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)
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
10/8/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
3/1/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- From Launch site CCAFS LC_40 major mission outcomes are success though still failed 2 times to land and 3 times no landing attempt has made. The mission major orbit type is LEO(ISS).

Total Payload Mass

sum(PAYLOAD_MASS__KG_)
45596

- The total payload mass carried by boosters launched by NASA (CRS) is 45596 kg.

Average Payload Mass by F9 v1.1

avg(PAYLOAD_MASS__KG_)
2928.4

- The average payload mass carried by booster version F9 v1.1 is 2928.4 Kg.

First Successful Ground Landing Date

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
22/12/2015	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

- The dates of the first successful landing outcome on ground pad is 22-12-2015

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster_Version	Landing_Outcome	PAYLOAD_MASS__KG_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

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

1. F9 FT B1022
2. F9 FT B1026
3. F9 FT B1021.2
4. F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

Total	Failure	Success
101	1	100

Mission_Outcome
Success
Failure (in flight)
Success (payload status unclear)
Success

- The total number of successful and failure mission outcomes are 101 including 1 failure and 100 success outcomes.

Boosters Carried Maximum Payload

- The maximum payload mass is 15600 Kg.
- The booster which have carried the maximum payload mass are

1. F9 B5 B1048.4
2. F9 B5 B1049.4
3. F9 B5 B1051.3
4. F9 B5 B1056.4
5. F9 B5 B1048.5
6. F9 B5 B1051.4
7. F9 B5 B1049.5
8. F9 B5 B1060.2
9. F9 B5 B1058.3
10. F9 B5 B1051.6
11. F9 B5 B1060.3
12. F9 B5 B1049.7

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

Date	month	Landing_Outcome	Booster_Version	Launch_Site
1/10/2015	10/	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
14/04/2015	/04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015 are

1. F9 v1.1 B1012 Booster in Launch site CCAFS LC-40
2. F9 v1.1 B1015 Booster in Launch site CCAFS LC-40

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

Landing_Outcome	count
Success (ground pad)	5
Failure (drone ship)	5

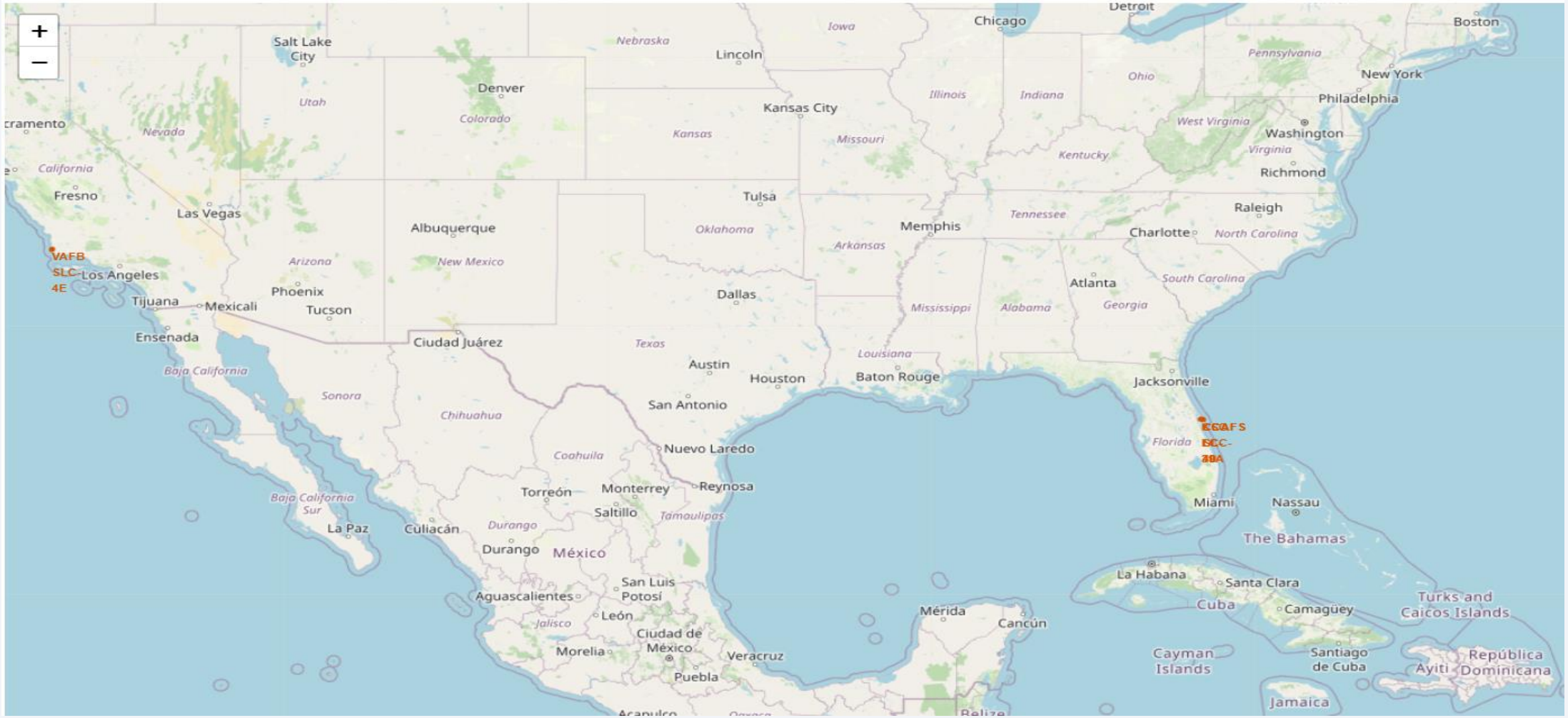
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order are

1. Success (ground pad) with count of 5
2. Failure (drone ship) with count of 5

Section 3

Launch Sites Proximities Analysis

Folium Map for SpaceX Launch Sites

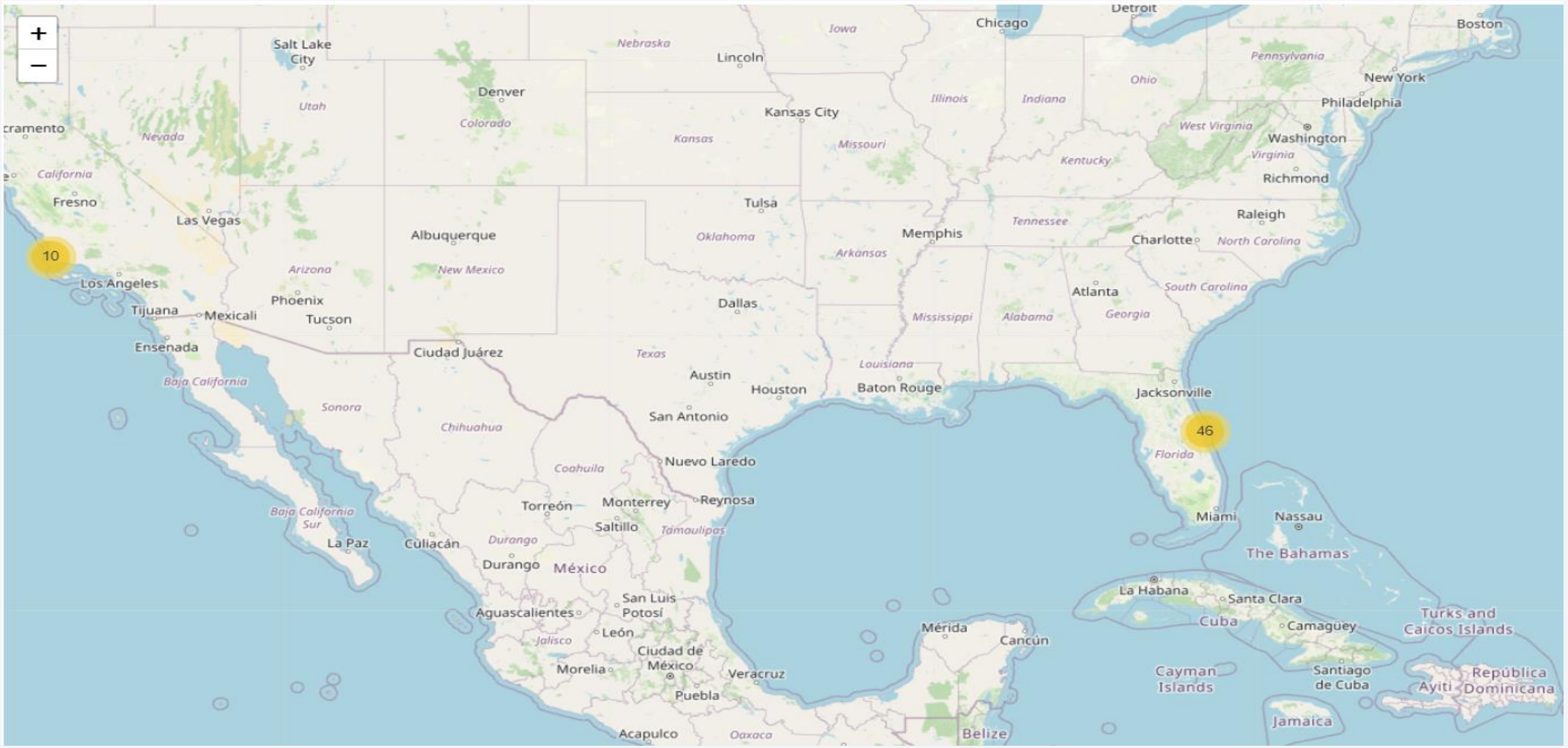


Folium Map for SpaceX Launch Sites

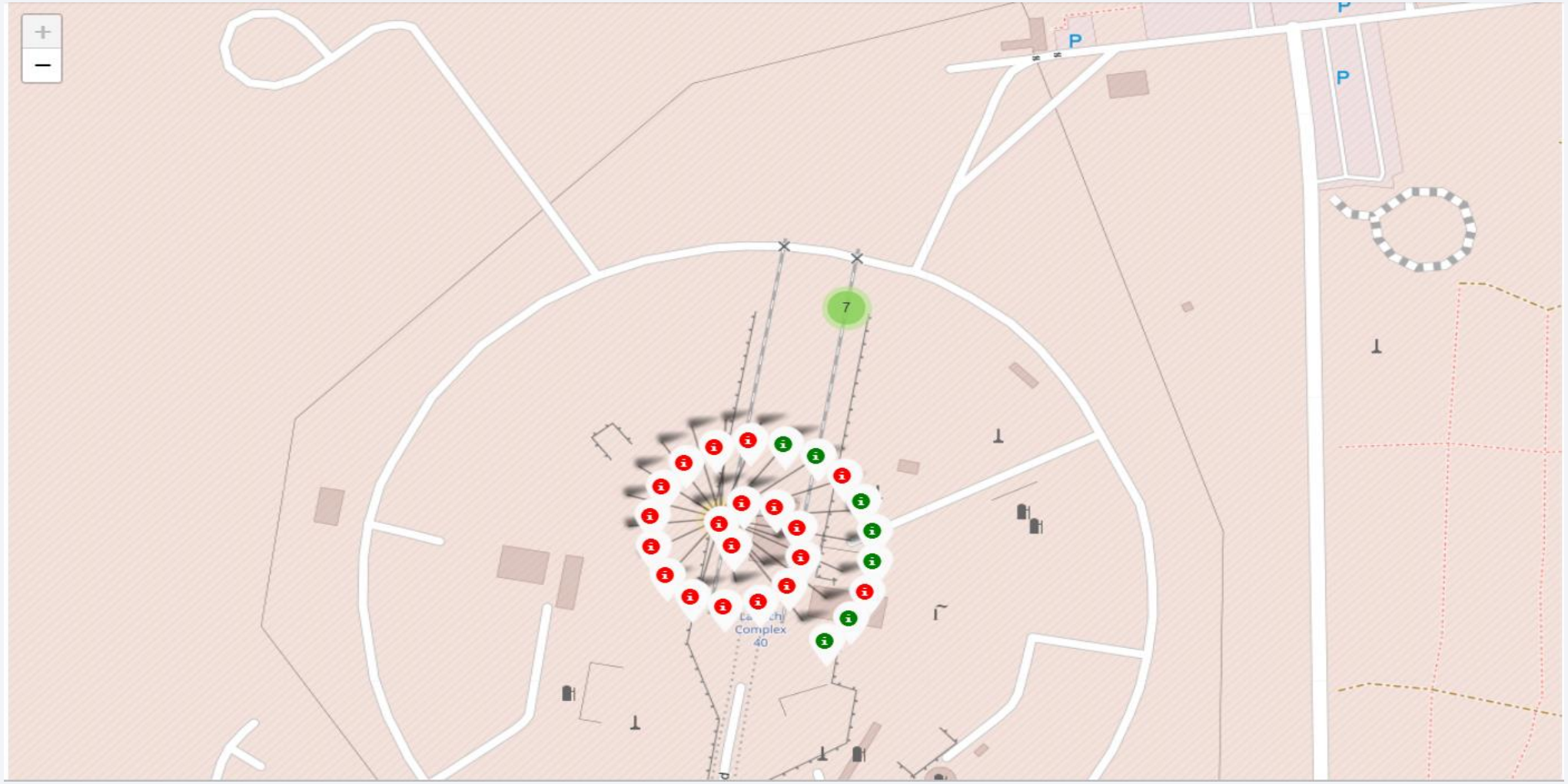
- The Launch sites are all close to the coastal area and 3 out of 4 located near to each other. The closest Launch sites are

1. CCAFS LC-40
2. CCAFS SLC-40
3. KSC LC-39A

Folium Map for SpaceX Launch Outcomes



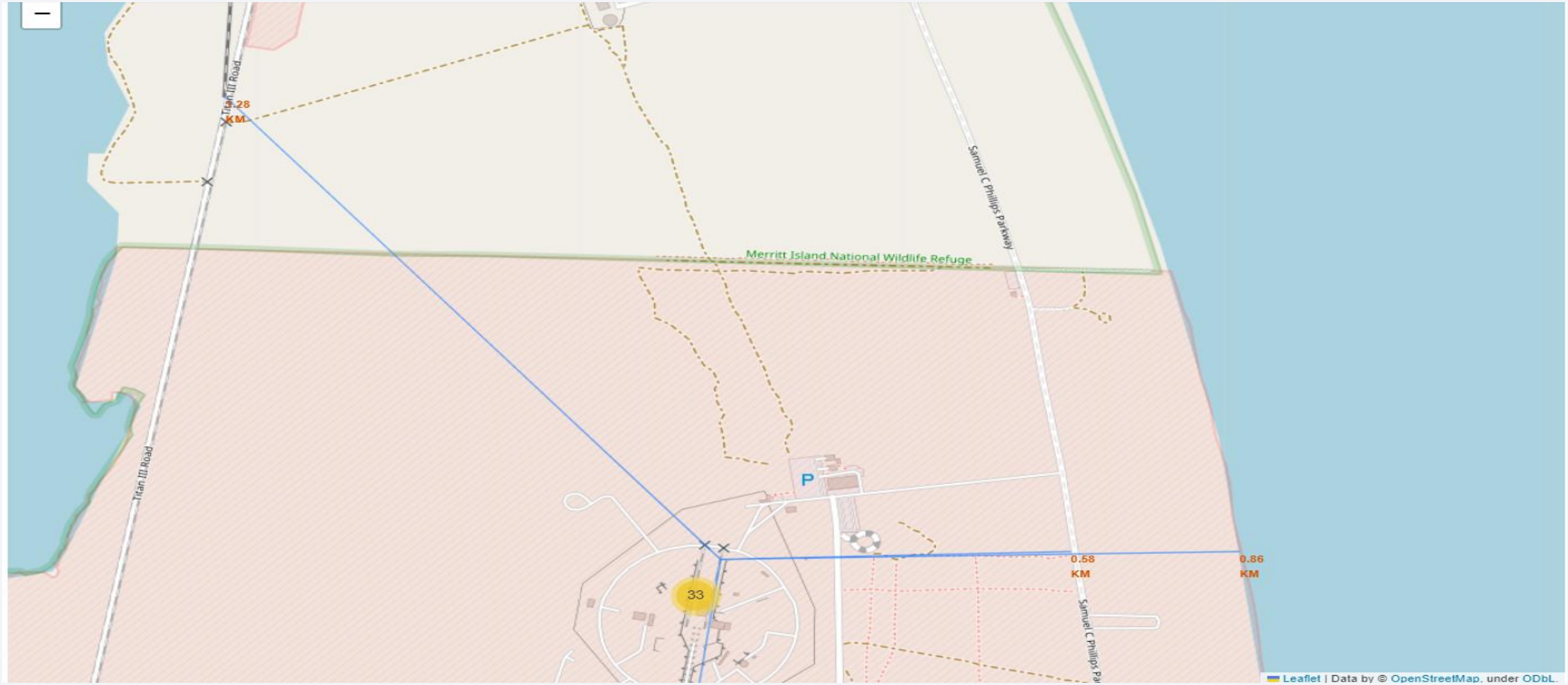
Folium Map for SpaceX Launch Outcomes



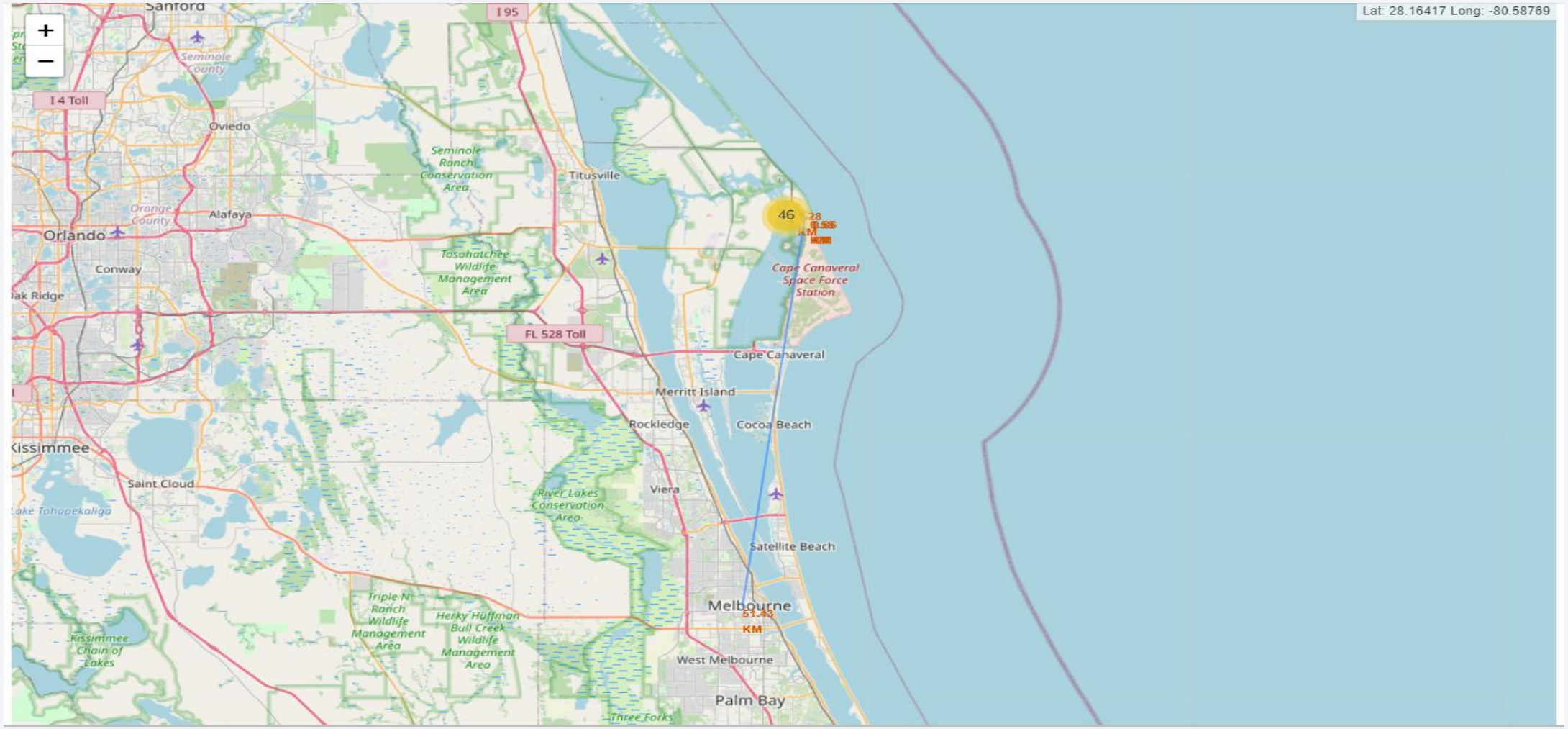
Folium Map for SpaceX Launch Outcomes

- The Launch site KSC LC-39A has a highest successful outcomes.
- The Launch site CCAFS LC-40 has a most number of Launches. It also has the most not successful outcomes with count of 19.

Folium Map for SpaceX Launch site close proximity



Folium Map for SpaceX Launch site close proximity



Folium Map for SpaceX Launch site close proximity

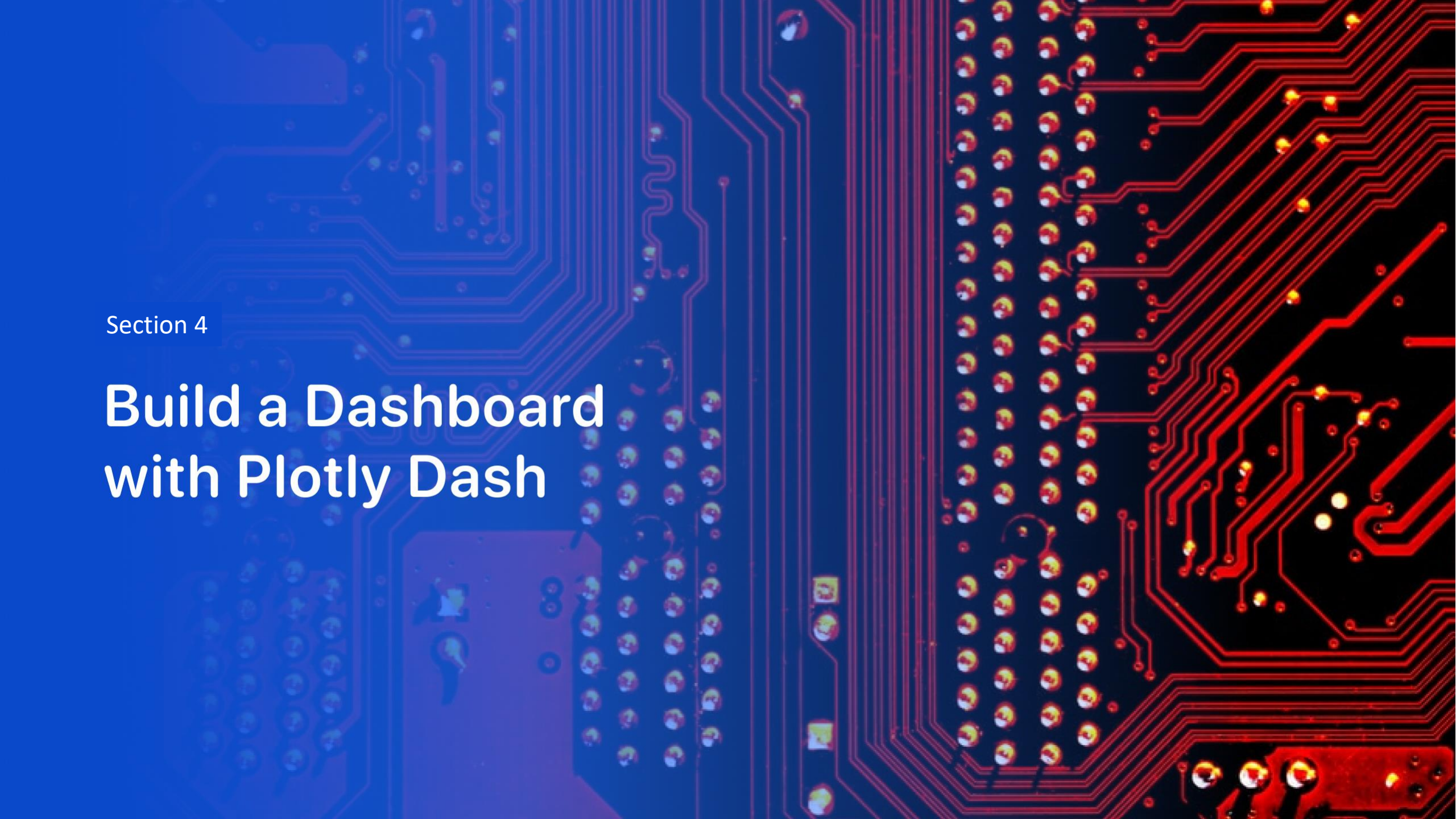
- From CCAFS SLC-40 launch site to its proximities such as
    - railway in 0.58 Km,
    - highway in 1.28 Km,
    - city in 51.43 Km and
    - coastline in 0.86 Km
- Launch sites are in close proximity to coastline so they can fly over the ocean during launch, for at least two safety reasons

    (1) crew has option to abort launch and attempt water landing

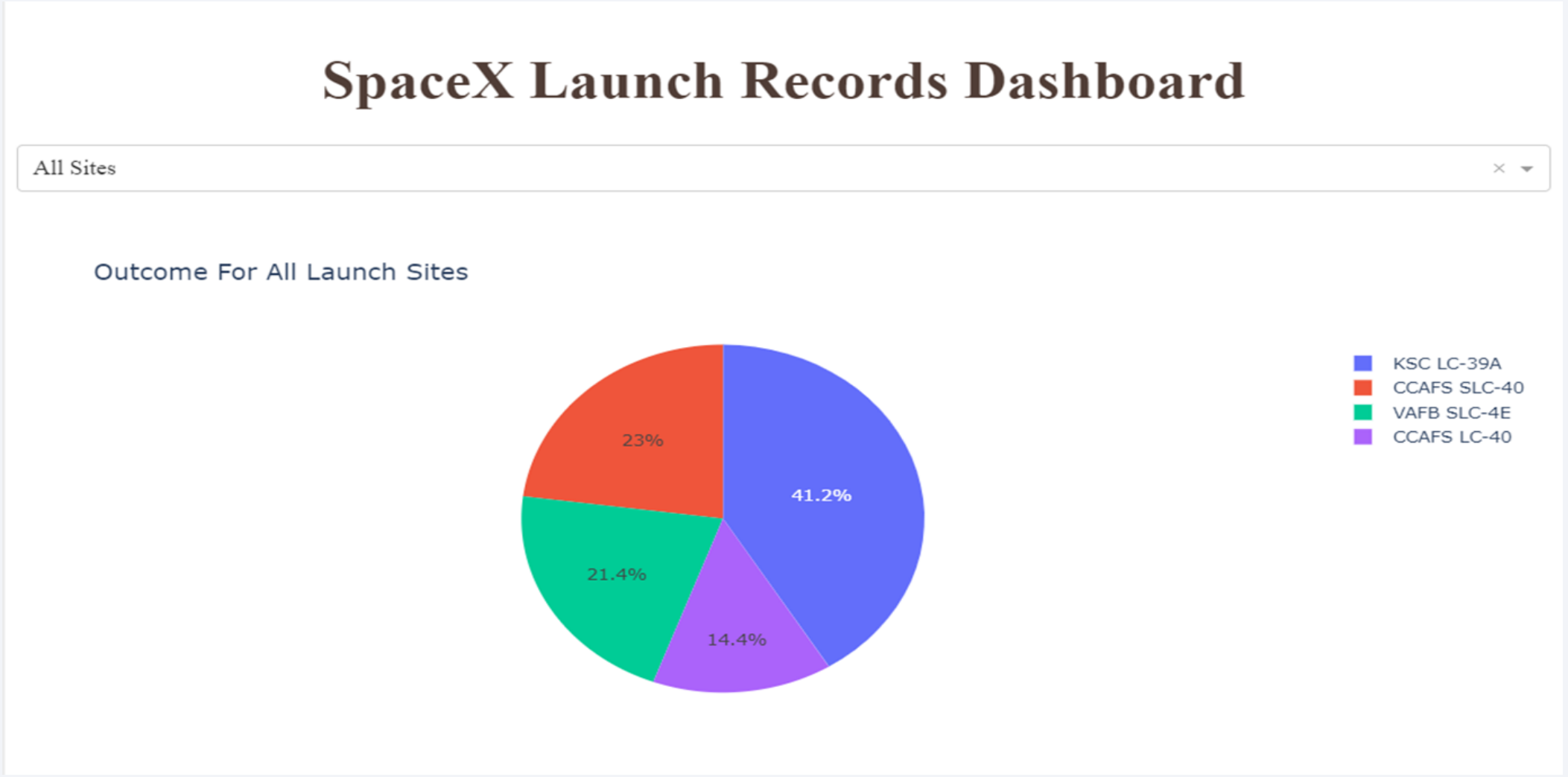
    (2) minimize people and property at risk from falling debris.

Folium Map for SpaceX Launch site close proximity

- Launch sites are in close proximity to highways, which allows for easily transport required people and property.
- Launch sites are in close proximity to railways, which allows transport for heavy cargo.
- Launch sites are not in close proximity to cities, which minimizes danger to population dense areas.


Section 4
Build a Dashboard with Plotly Dash

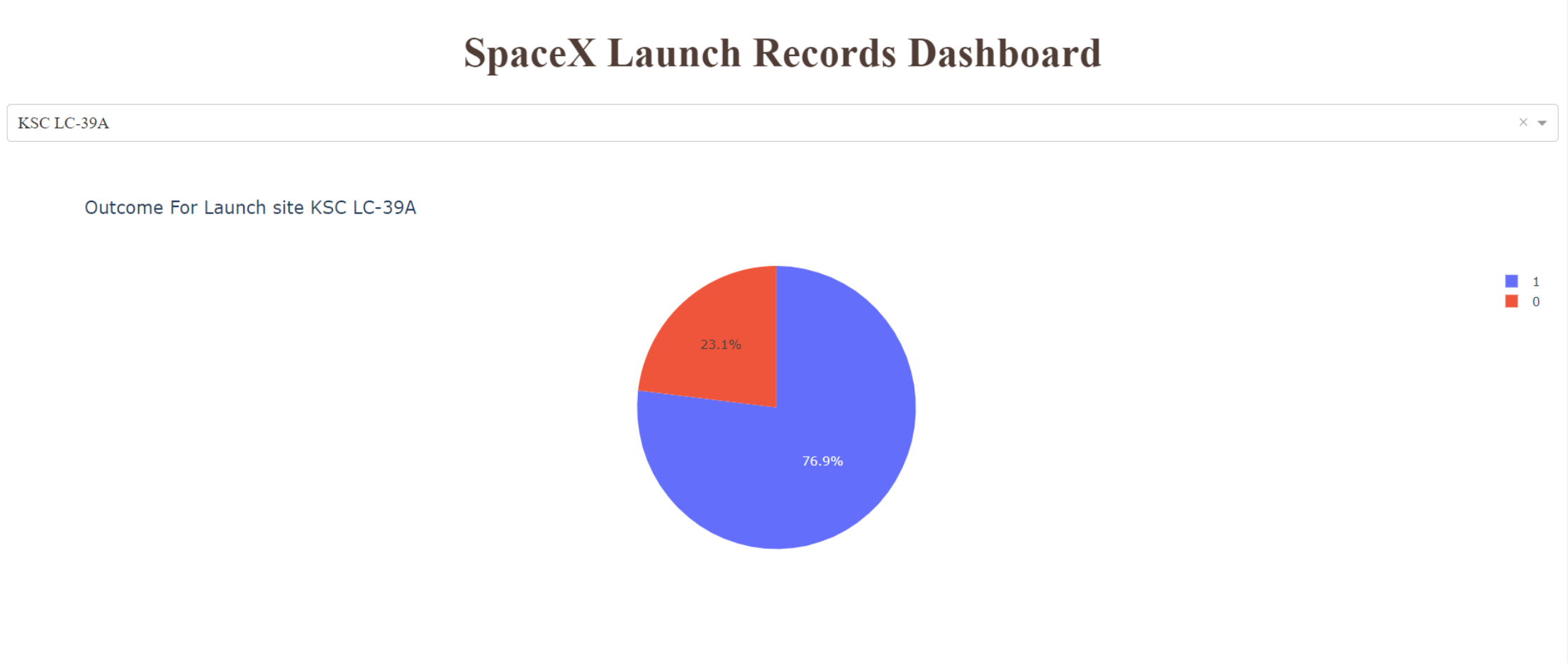
Dashboard for SpaceX Launch Outcomes Data



Dashboard for SpaceX Launch Outcomes Data

- The Launch sites successful outcomes ranks are
    1. KSC LC-39A with 41.2%
    2. CCAFS SLC-40 with 23%
    3. VAFB SLC-4E with 21.4%
    4. CCAFS LC-40 with 14.4%

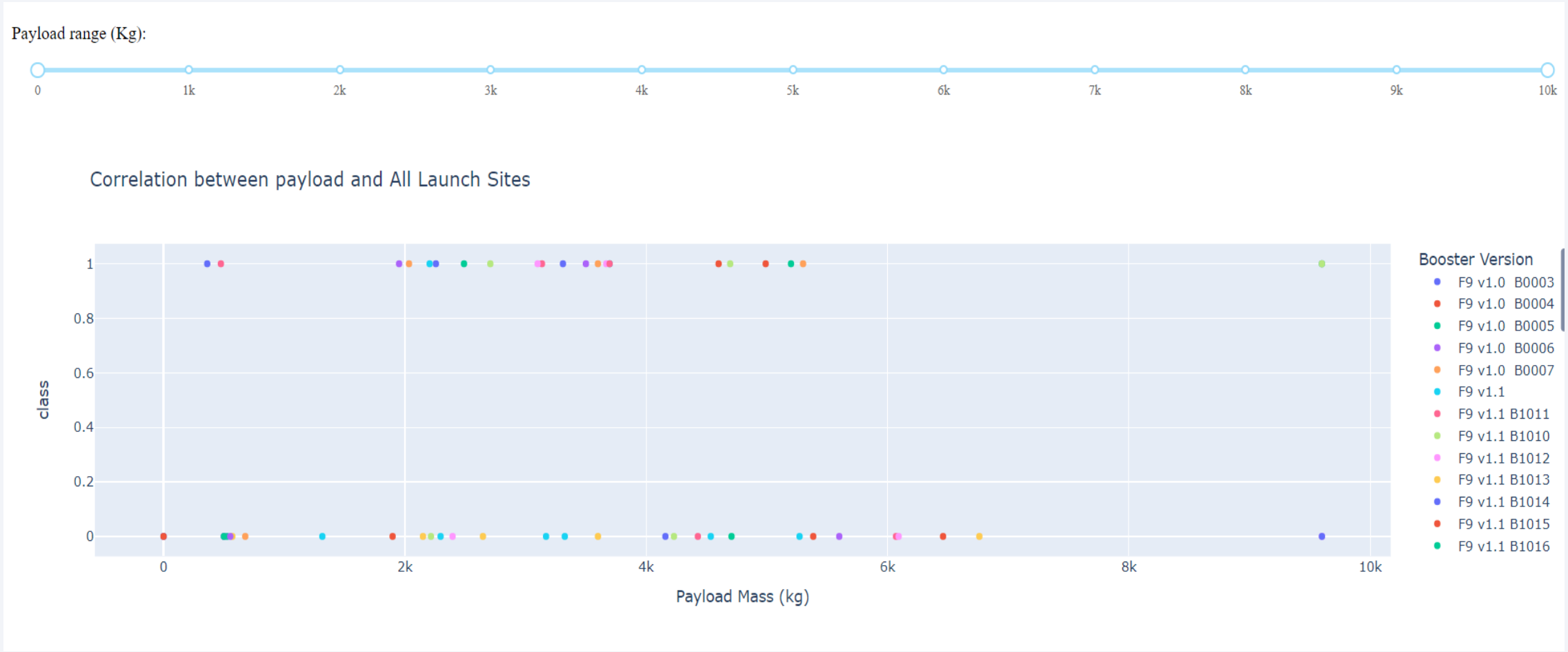
Dashboard for Highest Launch Site Success



Dashboard for Highest Launch Site Success

- The Launch site with highest launch success ratio is KSC LC-39A
- The Launch site KSC LC-39A has a success ratio of 76.9% and failure ratio of 23.1%

Dashboard for Payload and Outcome Relationship



Dashboard for Payload and Outcome Relationship

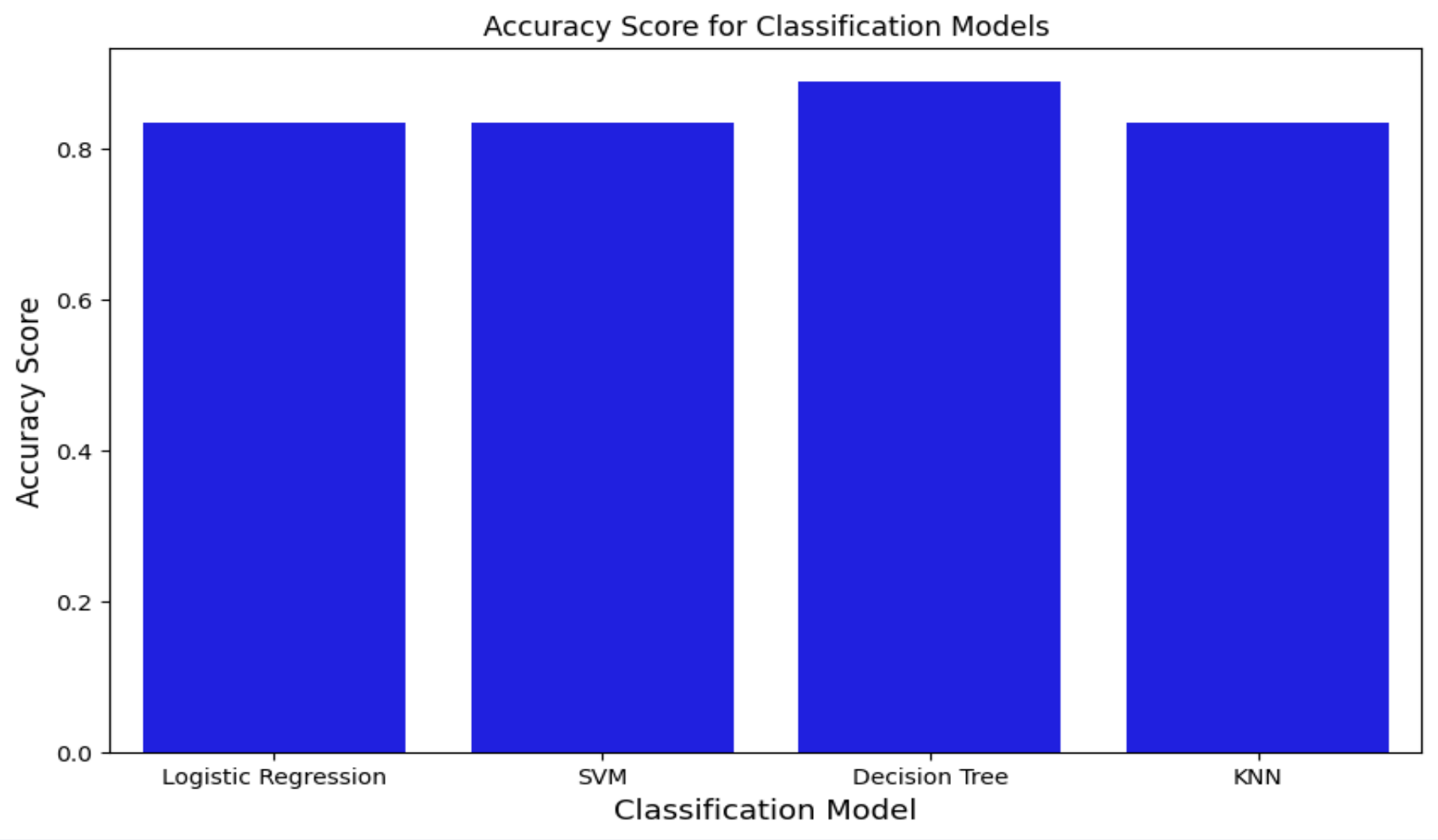
Dashboard for Payload and Outcome Relationship

- The payload mass range from 1900 Kg to 5400 Kg has the largest success rate
- The different Booster version success rate is without any much difference.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

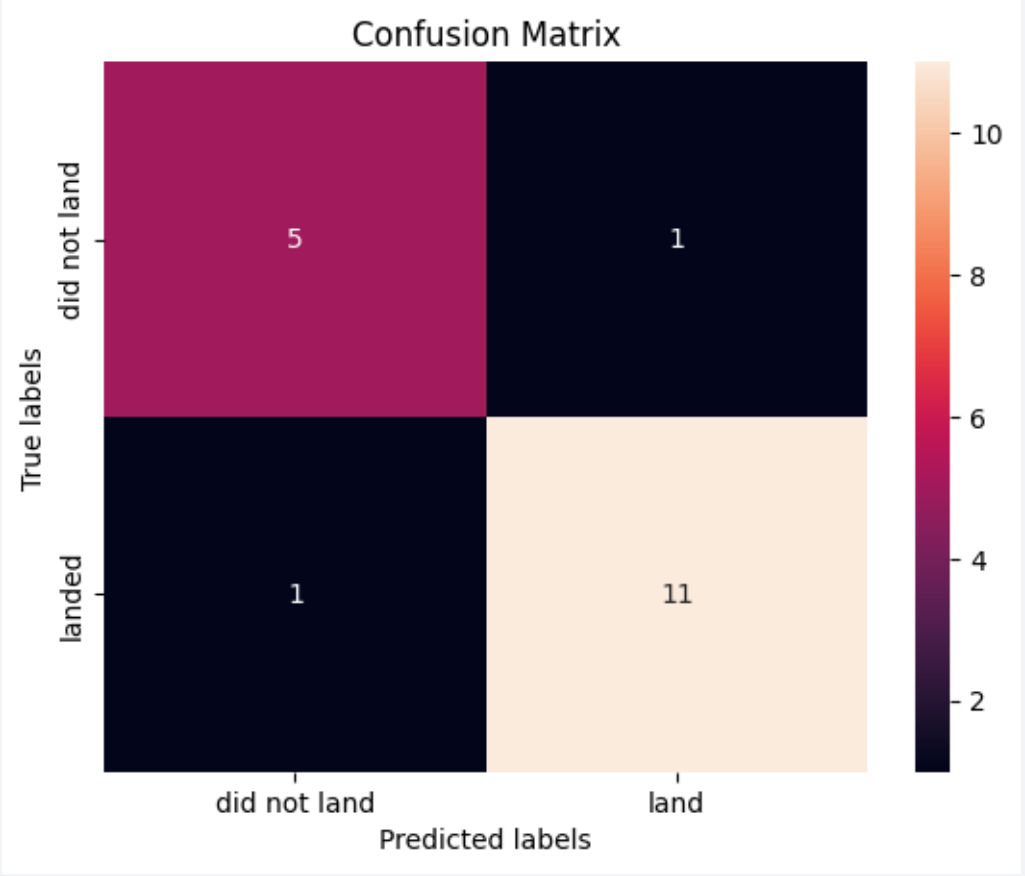


Classification Accuracy

- Decision Tree classification model has the highest accuracy with 88.89% .
- Decision Tree classification model best parameters are
    - 'criterion': 'gini',
    - 'max_depth': 6,
    - 'max_features': 'sqrt',
    - 'min_samples_leaf': 4,
    - 'min_samples_split': 10,
    - 'splitter': 'best'

Confusion Matrix

- The Decision Tree classification model predicted correctly the 11 outcomes as landed and 5 outcomes as did not land.
- It also wrongly predicted the 1 outcome as did not landed as landed and 1 outcome as landed as did not landed.



Conclusions

- SpaceX Launch data has 4 major Launch sites. The Launch site KSC LC-39A has a success ratio of 76.9% and failure ratio of 23.1%
- The payload mass range from 1900 Kg to 5400 Kg has the largest success rate
- The Launch sites are close to coastal area
- The success rate since 2013 kept increasing till 2020
- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS Orbit type.
- The Decision Tree classification model has the highest accuracy with 88.89% for predicting Launching outcomes .

Appendix

Python code snippet

```
# Import required libraries
import pandas as pd
import dash
import dash_html_components as html
import dash_core_components as dcc
from dash.dependencies import Input, Output
import plotly.express as px

# Read the airline data into pandas dataframe
spacex_df = pd.read_csv("spacex_launch_dash.csv")
max_payload = spacex_df['Payload Mass (kg)'].max()
min_payload = spacex_df['Payload Mass (kg)'].min()

# Create a dash application
app = dash.Dash(__name__)

# Create an app layout
app.layout = html.Div(children=[html.H1('SpaceX Launch Records Dashboard',
                                        style={'textAlign': 'center', 'color': '#503D36',
                                               'font-size': 40}),
                                # TASK 1: Add a dropdown list to enable Launch Site selection
                                # The default select value is for ALL sites
                                  dcc.Dropdown(id='site-dropdown',
                                               options=[
                                                   {'label': 'All Sites', 'value': 'ALL'},
                                                   {'label': 'CCAFS LC-40', 'value': 'CCAFS LC-40'},
                                                   {'label': 'CCAFS SLC-40', 'value': 'CCAFS SLC-40'},
                                                   {'label': 'KSC LC-39A', 'value': 'KSC LC-39A'},
                                                   {'label': 'VAFB SLC-4E', 'value': 'VAFB SLC-4E'},
                                               ],
                                               value='ALL',
                                               placeholder="place select here",
                                               searchable=True
                                               ),
                                html.Br(),
```

Thank you!