IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sayeed Hassan
February 10, 2025

Outline

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

Executive Summary

- Summary of methodologies
- Data collection methodology: API and web scrapping
- Perform data wrangling: Data Transformation, , Encoding categorical features, Data formatting.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

- Summary of all results
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Introduction

- Project background and context
- 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:

  Space X Falcon 9 First Stage Landing Prediction

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - API and Web Scraping
- Perform data wrangling
  - Data transformation, Encoding categorical features, Data formating
- 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 features, Split data into training and testing, Tune Hyperparameter with Grid Search CV, find best parameters, find accuracy score and choose best model.

Data Collection

- API requests to spacex website:
  . Make get requests to spacex API

- Web Scraping:
  Web scrap Falcon 9 launch records with `BeautifulSoup`:
  . Extract Falcon 9 launch records HTML table from Wikipedia.
  . Parse the table and convert into a Pandas data frame.

Data Collection – SpaceX API

1. Get Booster Version data
   Get Luanch Sites data
   Get Payloads data
   Get Cores data
   Get Past Launches data

2. https://github.com/me-in-tech/DataScienceCapstone/blob/4c64e3df9c57342e1b0383947f7e5c7c399835f3/1.%20jupyter-labs-spacex-data-collection-api.ipynb

https://api.spacexdata.com/v4/rockets/

https://api.spacexdata.com/v4/launchpads/

https://api.spacexdata.com/v4/payloads/

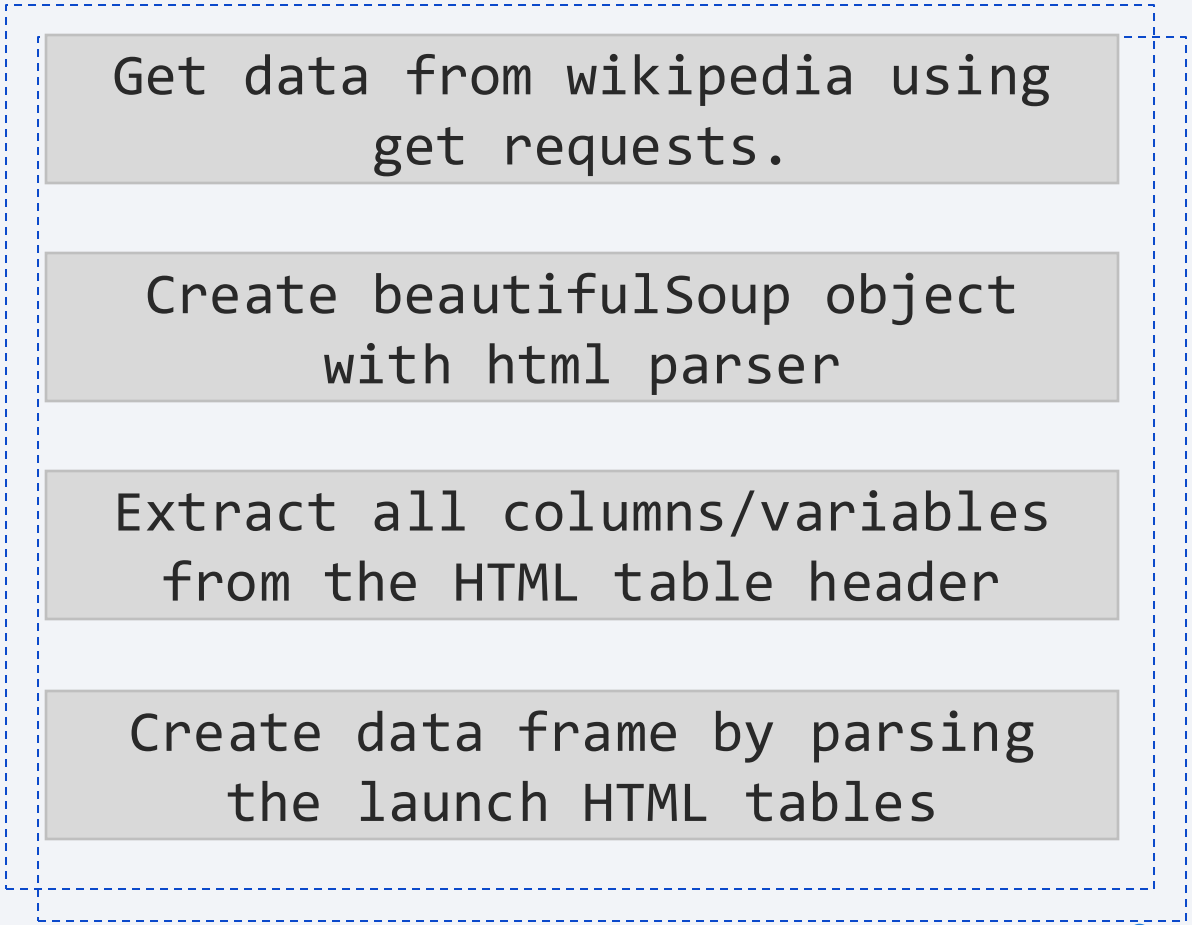
https://api.spacexdata.com/v4/cores/

https://api.spacexdata.com/v4/launches/past

Data Collection - Scraping

- Use BeautifulSoup, extract HTML tables, Parse the tables.

- https://github.com/me-in-tech/DataScienceCapstone/blob/4c64e3df9c57342e1b0383947f7e5c7c399835f3/2.%20jupyter-labs-spacex%20data%20collection%20webscraping.ipynb



9

Data Wrangling

- Calculate the number of launches on each site

  Calculate the number of occurrence of each orbit

  Calculate the number and occurrence of mission outcome of the orbits

  Create landing outcome label from Outcome column.

- https://github.com/me-in-tech/DataScienceCapstone/blob/4c64e3df9c57342e1b0383947f7e5c7c399835f3/3.%20labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Plots prepared for feature engineering.
  . Categorical Plot for relationship between flight number and launch site.
  . Categorical Plot for relationship between payload mass and launch site.
  . Bar Plot for relationship between success rate of each orbit type.
  . Categorical Plot for relationship between flightnumber and orbit type.
  . Categorical Plot for relationship between payload and mass and orbit type.
  . Line Plot for Launch success yearly trend.
  After, select the features which will be used in success prediction.
  Create dummy variables (one hot encoding) of categorical features.

- https://github.com/me-in-tech/DataScienceCapstone/blob/4c64e3df9c57342e1b0383947f7e5c7c399835f3/5.%20eda%20data%20visualization.ipynb

EDA with SQL

- Make sqlite3 connection with the database.
  Make a cursor through the connection to perform querring.
  Connect the database using SQL magic within jupyter notebook

- https://github.com/me-in-tech/DataScienceCapstone/blob/4c64e3df9c57342e1b0383947f7e5c7c399835f3/4.%20jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Circle object to mark the location of launch site.
  Marker object to put the name of the site.
  Marker Clusters to show success/failure numbers at the site.
- https://github.com/me-in-tech/DataScienceCapstone/blob/4c64e3df9c57342e1b0383947f7e5c7c399835f3/7.%20lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Pie chart to show the proportion of success and failure of each site launches. Scatter plot to show the correlation between payload and launch success.

- https://github.com/me-in-tech/DataScienceCapstone/blob/4c64e3df9c57342e1b0383947f7e5c7c399835f3/8.%20spacex_dash_app.py

Predictive Analysis (Classification)

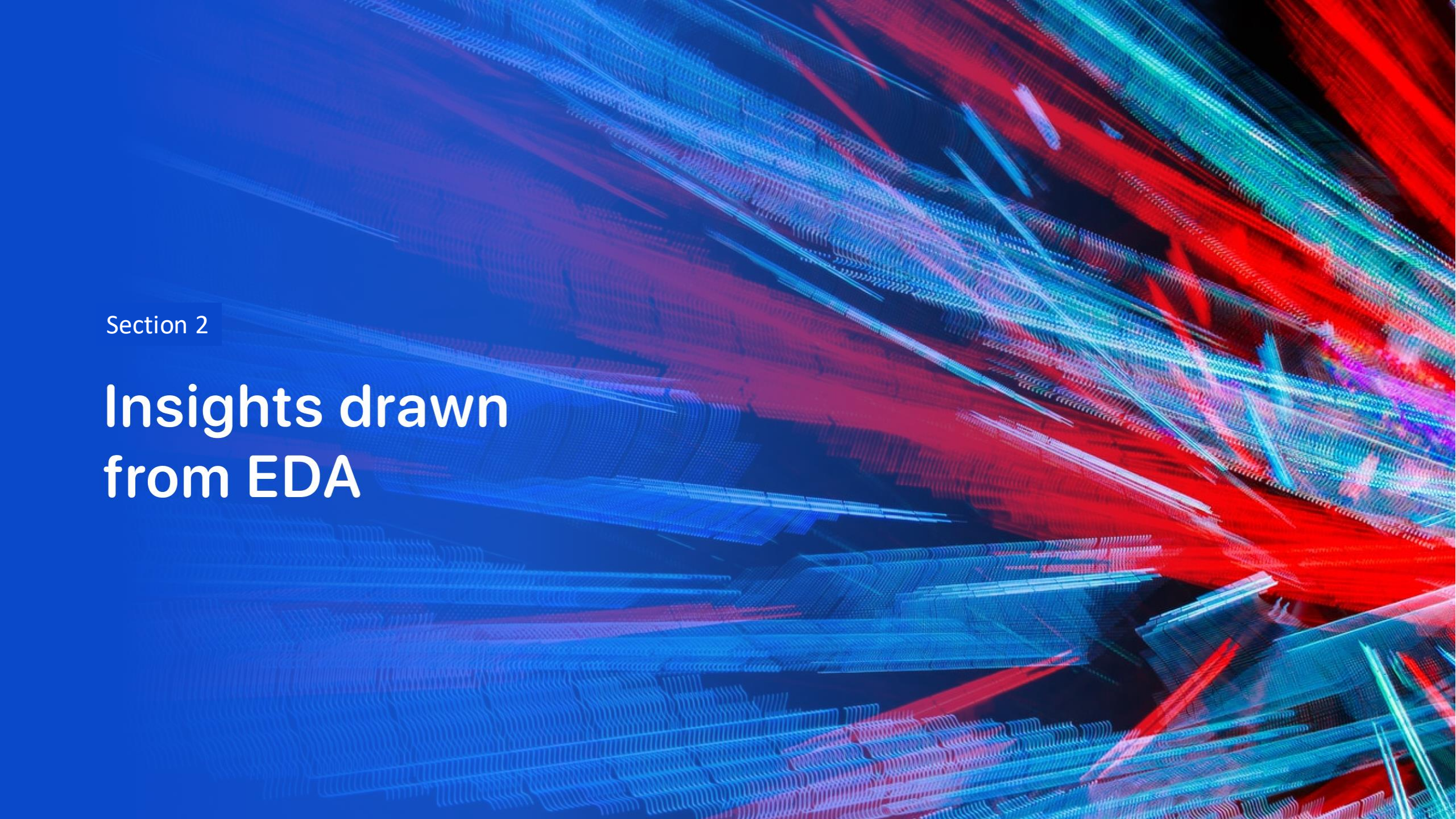
- create a column for the class
- Standardize the data
- Split into training data and test data
- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
- Find the method performs best using test data

- https://github.com/me-in-tech/DataScienceCapstone/blob/4c64e3df9c57342e1b0383947f7e5c7c399835f3/6.%20SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

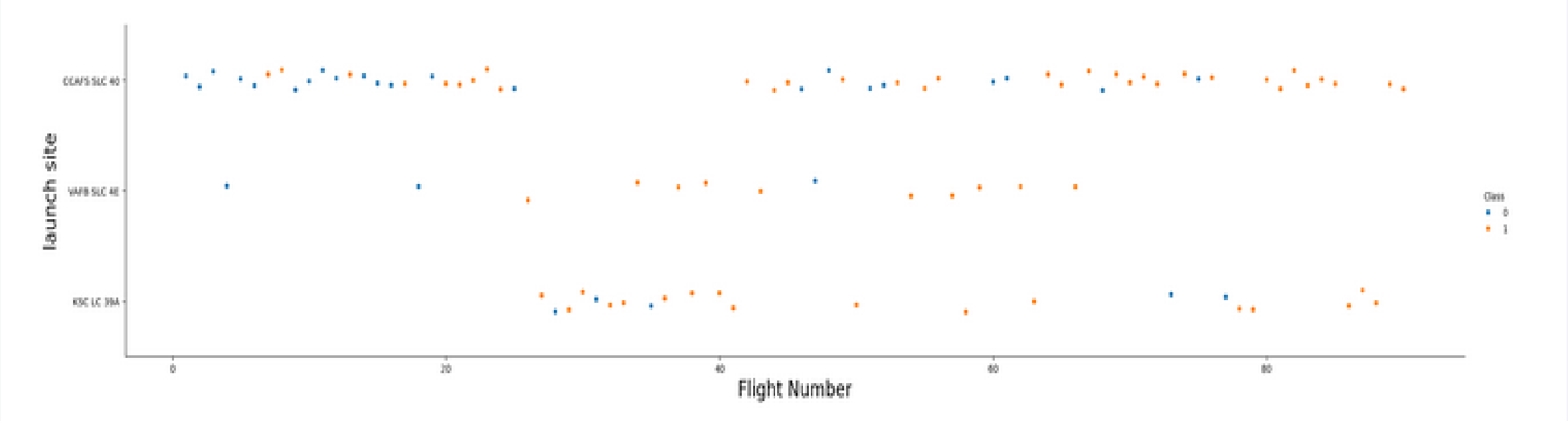
Results

- Features selected are FlightNumber, PayloadMass, Orbit, Launch Site, Flights, GridFins Reused Legs, Landingpad, Block ResuedCount, Serial.

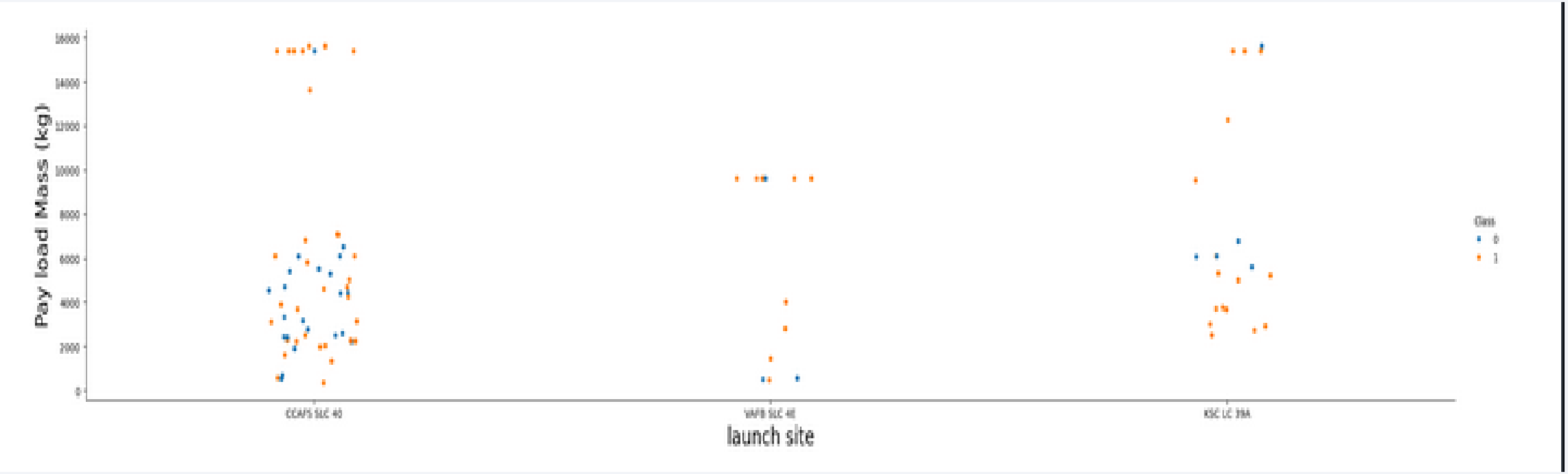
- Decision Tree classifier performed best with training data accuracy of 86.25% and test data accuracy 83.33%


Section 2
Insights drawn from EDA

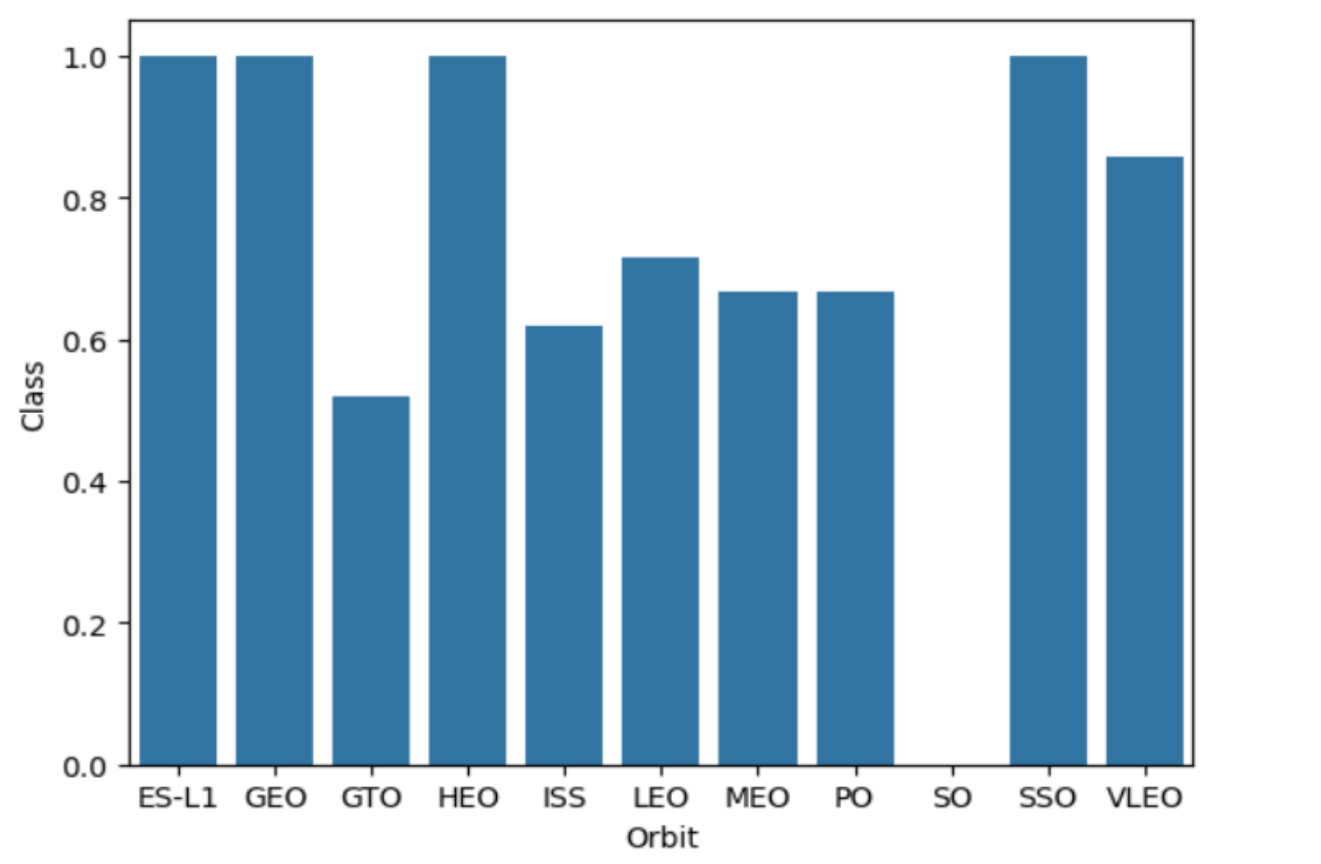
Flight Number vs. Launch Site



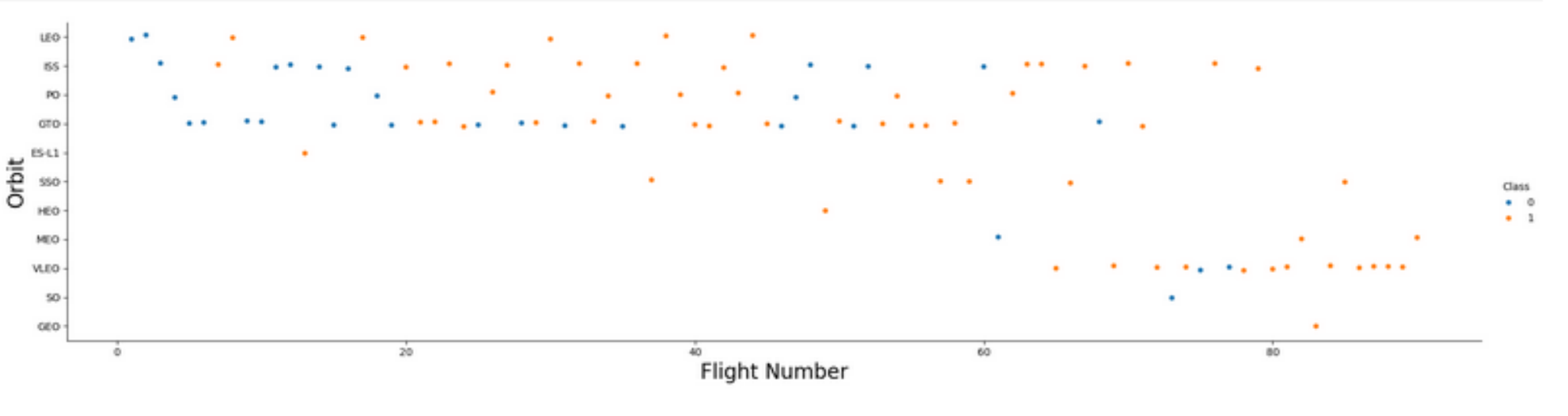
Payload vs. Launch Site



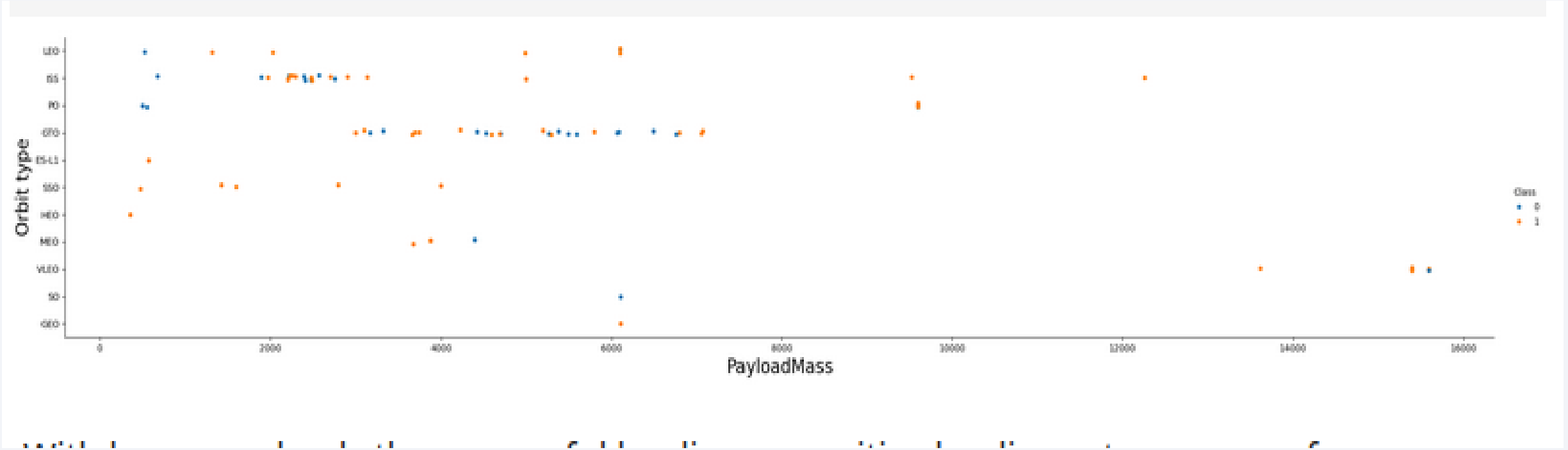
Success Rate vs. Orbit Type



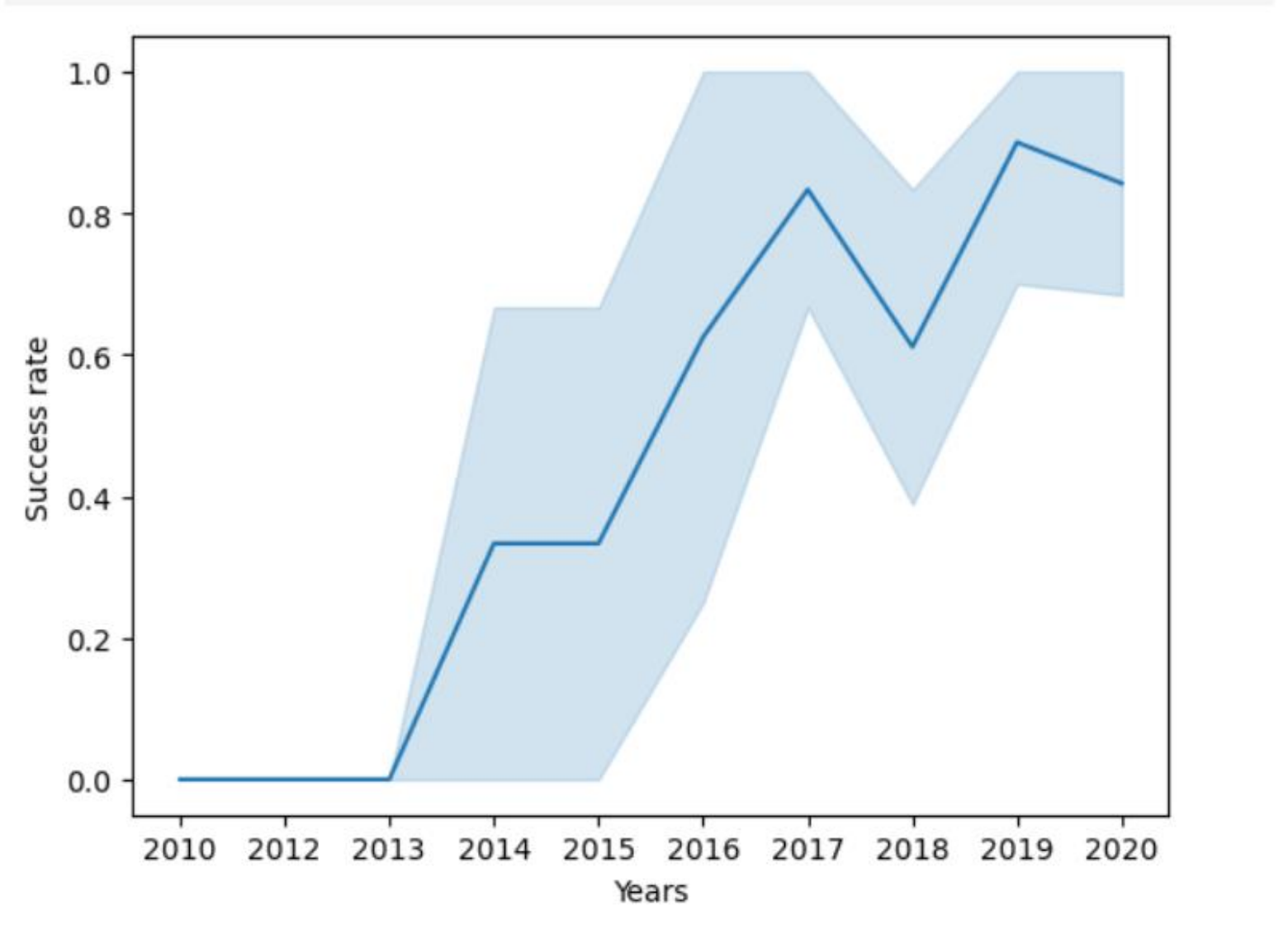
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

- Names of the unique launch sites

  CCAFS LC-40

  VAFB SLC-4E

  KSC LC-39A

  CCAFS SLC-40

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

  CCAFS LC-40

  CCAFS LC-40

  CCAFS LC-40

  CCAFS LC-40

  CCAFS LC-40

- `%sql SELECT Launch_Site FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;`

Total Payload Mass

- Total payload carried by boosters from NASA:

  45596

- ```
  %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE
  Customer LIKE 'NASA (CRS%)';
  ```
  ```

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1
  2534.6666666666665

- %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version LIKE 'F9 v1.1%';

First Successful Ground Landing Date

- Dates of the first successful landing outcome on ground pad:
  2010-06-04

- `%sql SELECT MIN(Date) FROM SPACEXTABLE WHERE Mission_Outcome LIKE 'Success';`

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:

  F9 FT B1022

  F9 FT B1026

  F9 FT B1021.2

  F9 FT B1031.2

- `%%sql SELECT Booster_Version`
- `FROM SPACEXTABLE`
- `WHERE Landing_Outcome LIKE "Success (drone ship)"`
- `  AND PAYLOAD_MASS__KG_ > 4000`
- `  AND PAYLOAD_MASS__KG_ < 6000;`

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes:

Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	21
No attempt	1
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2

- `%sql SELECT DISTINCT Landing_Outcome, COUNT(Landing_Outcome) FROM SPACEXTABLE GROUP BY Landing_Outcome;`

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

  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

- `%sql SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ IN (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);`

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

  F9 B5 B1051.4

  F9 B5 B1049.5

  F9 B5 B1060.2

  F9 B5 B1058.3

  F9 B5 B1051.6

- `%sql SELECT Booster_Version FROM SPACEXTABLE WHERE Date LIKE '2015%' AND Landing_Outcome LIKE 'Failure(drone ship)';`

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

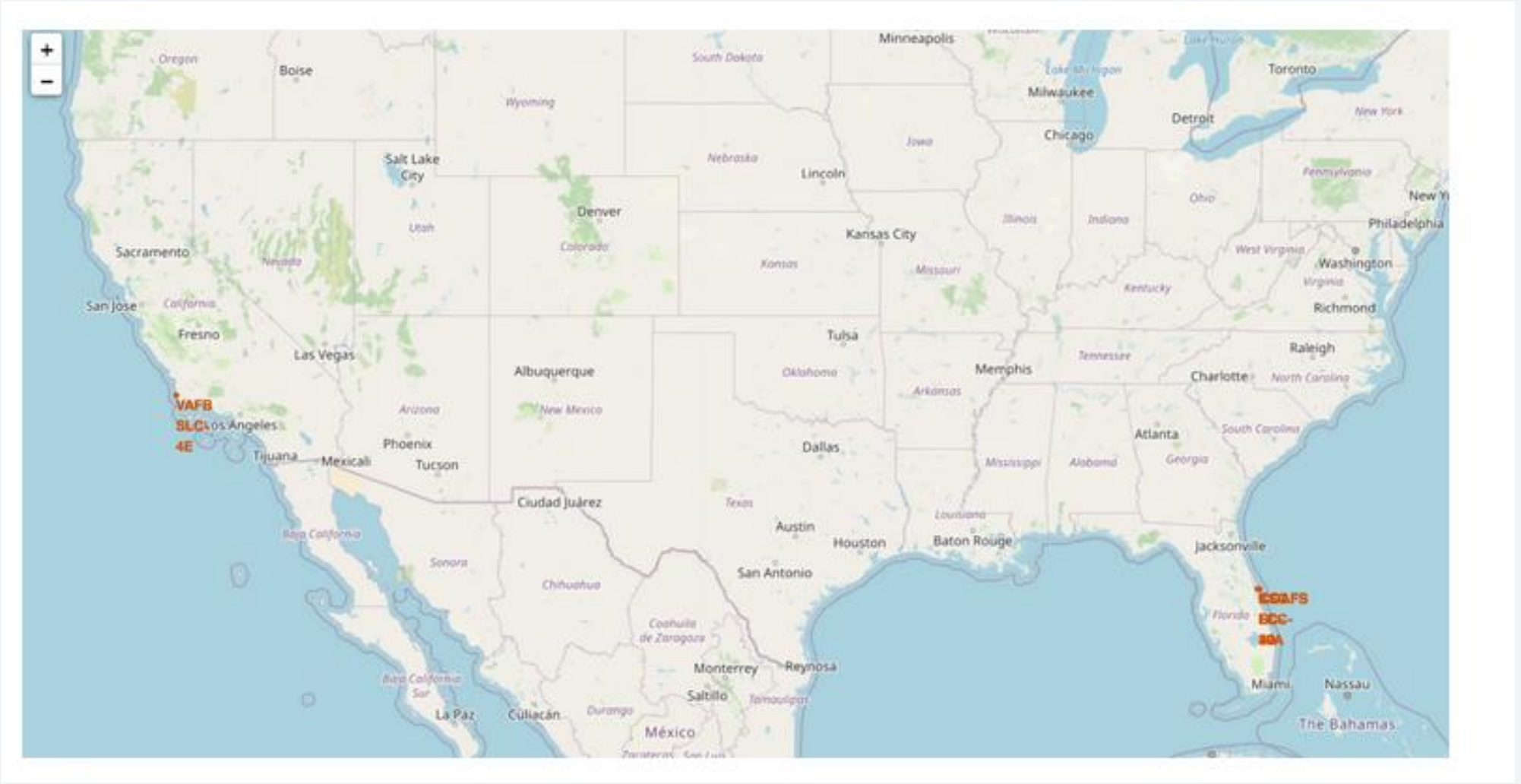
Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

```
%%sql
SELECT Landing_Outcome, COUNT(*) AS Outcome_Count
FROM SPACEXTABLE
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
```

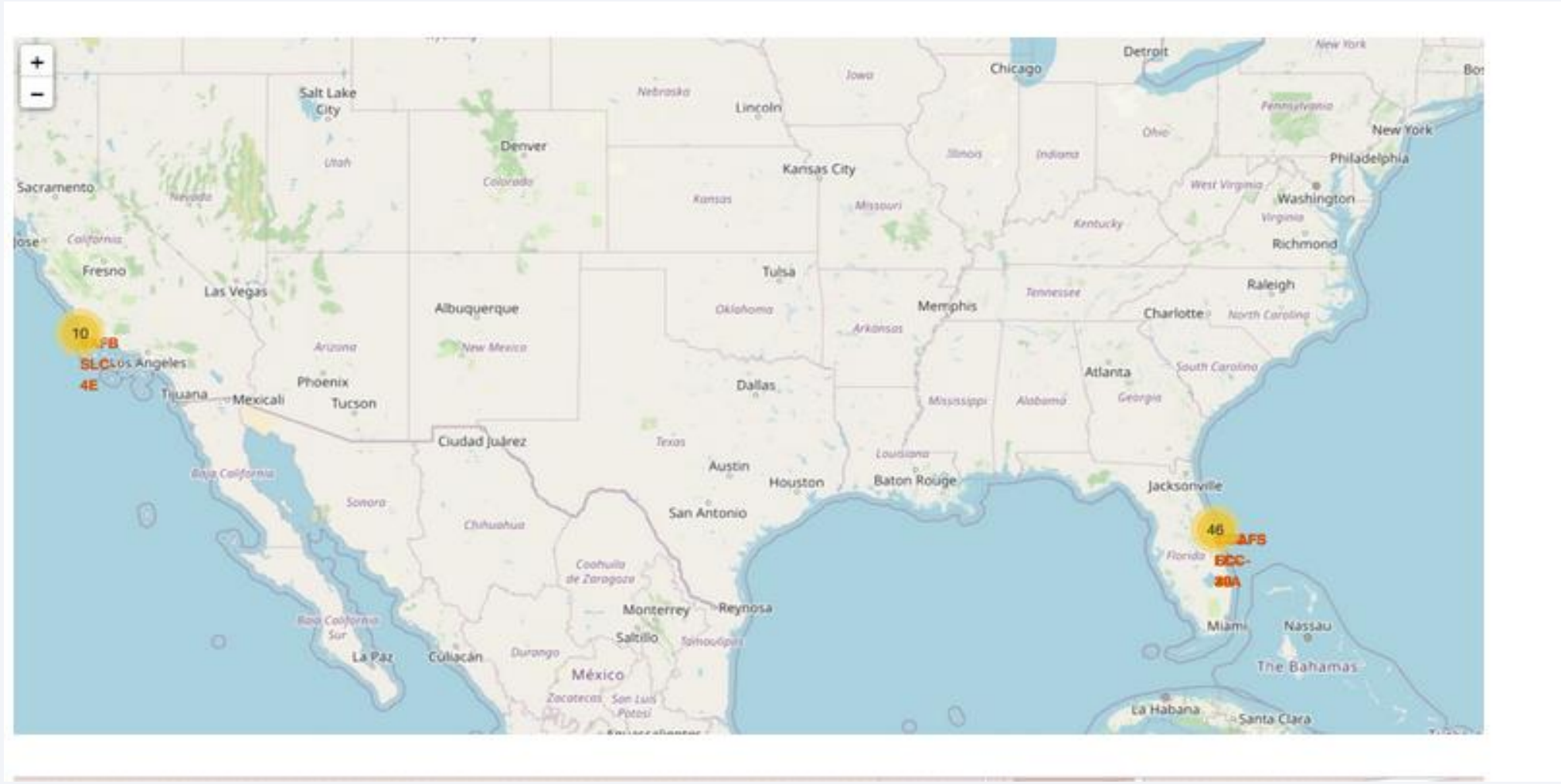
Section 3

Launch Sites Proximities Analysis

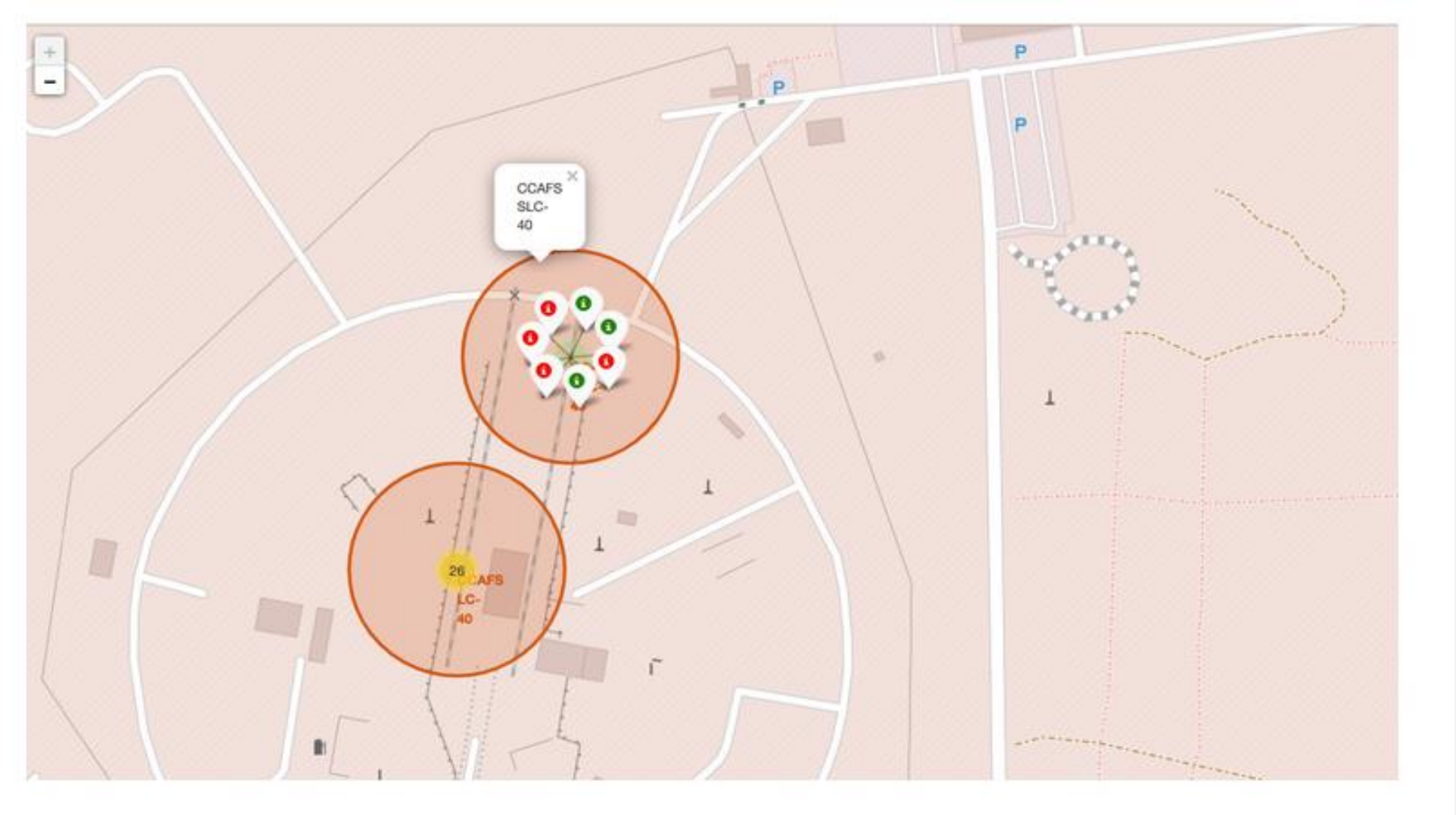
All launch sites marking on a map

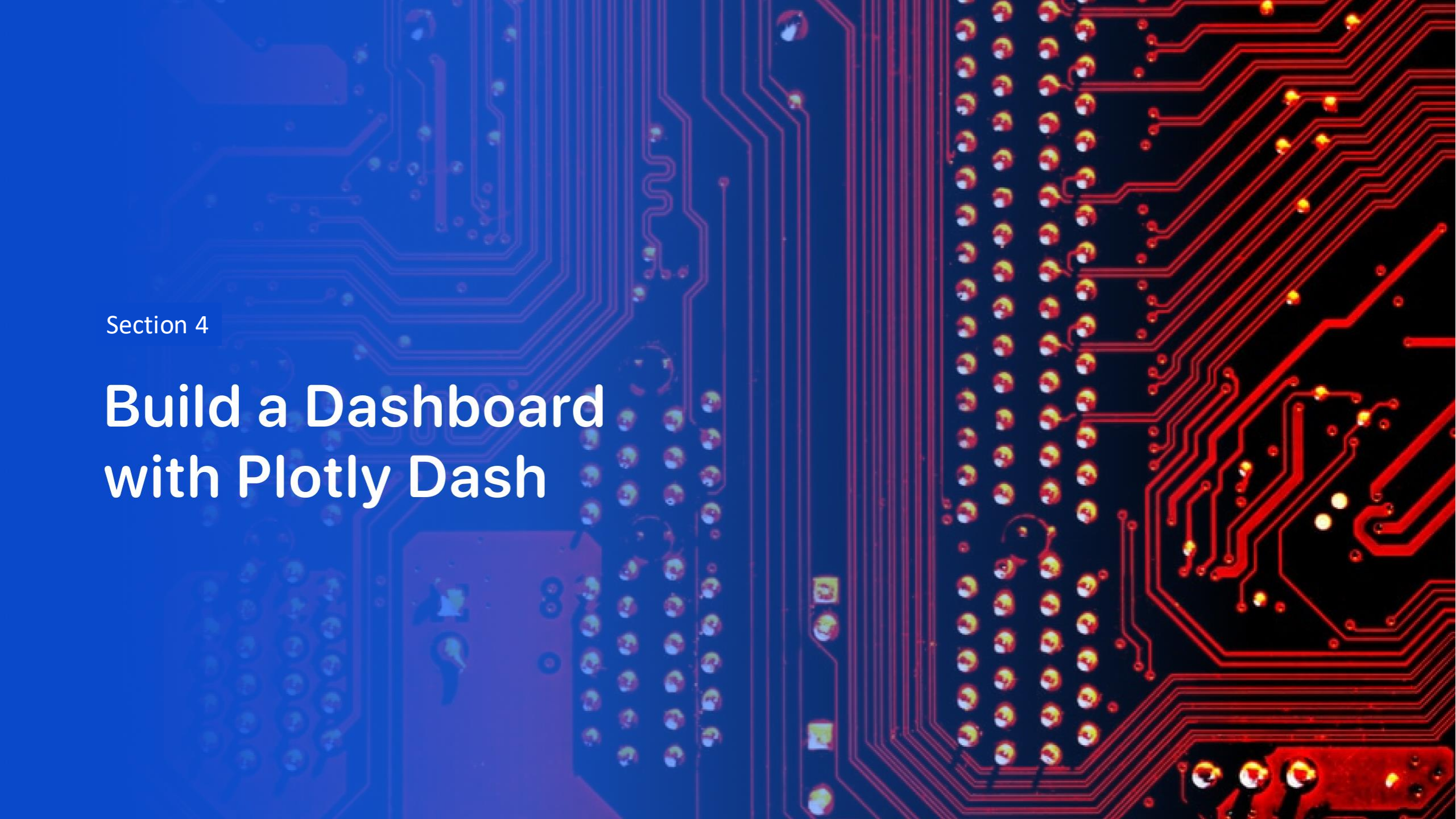


Add Circle object to the marked sites

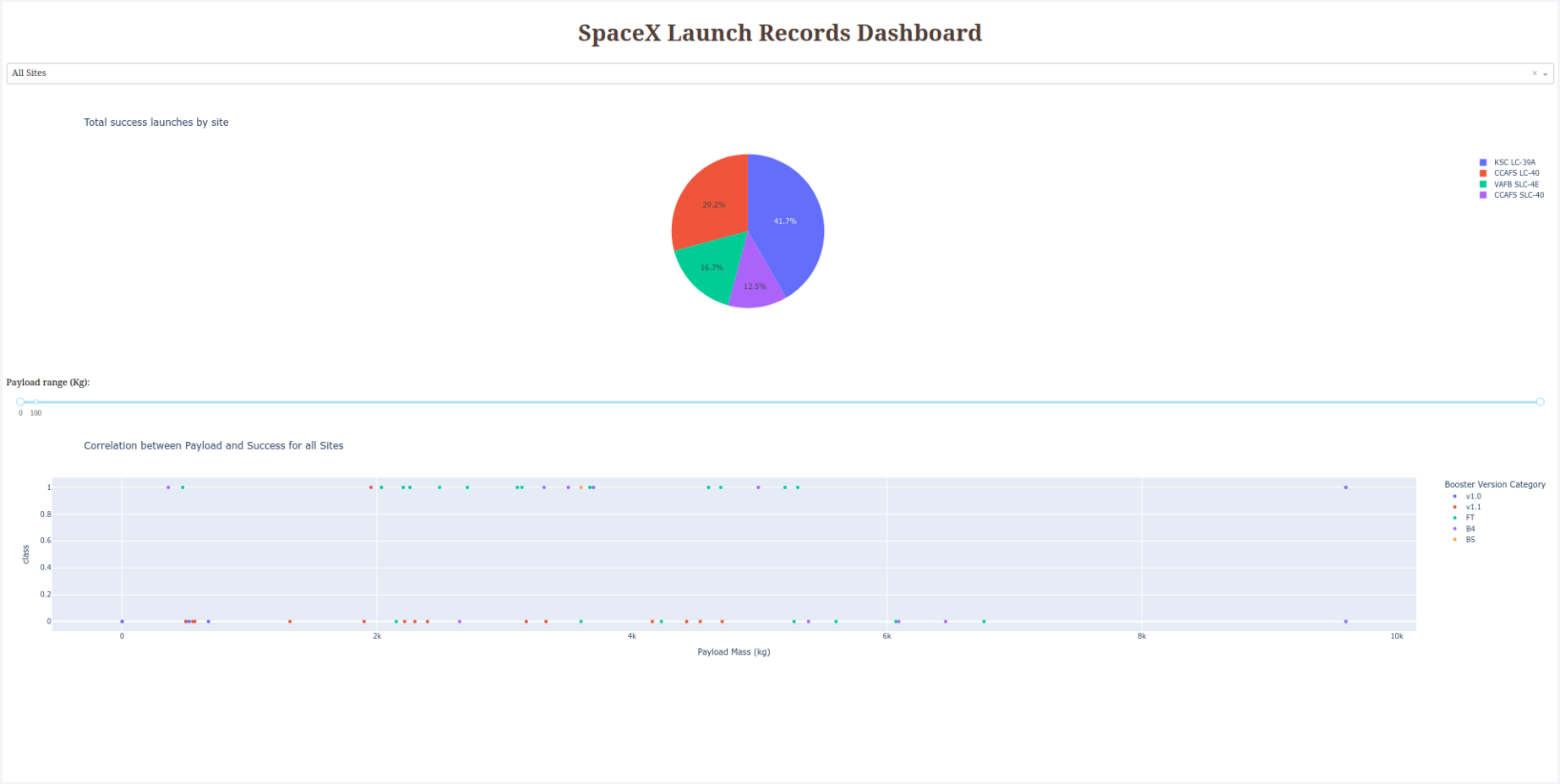


Add Success/Failure markers to each site

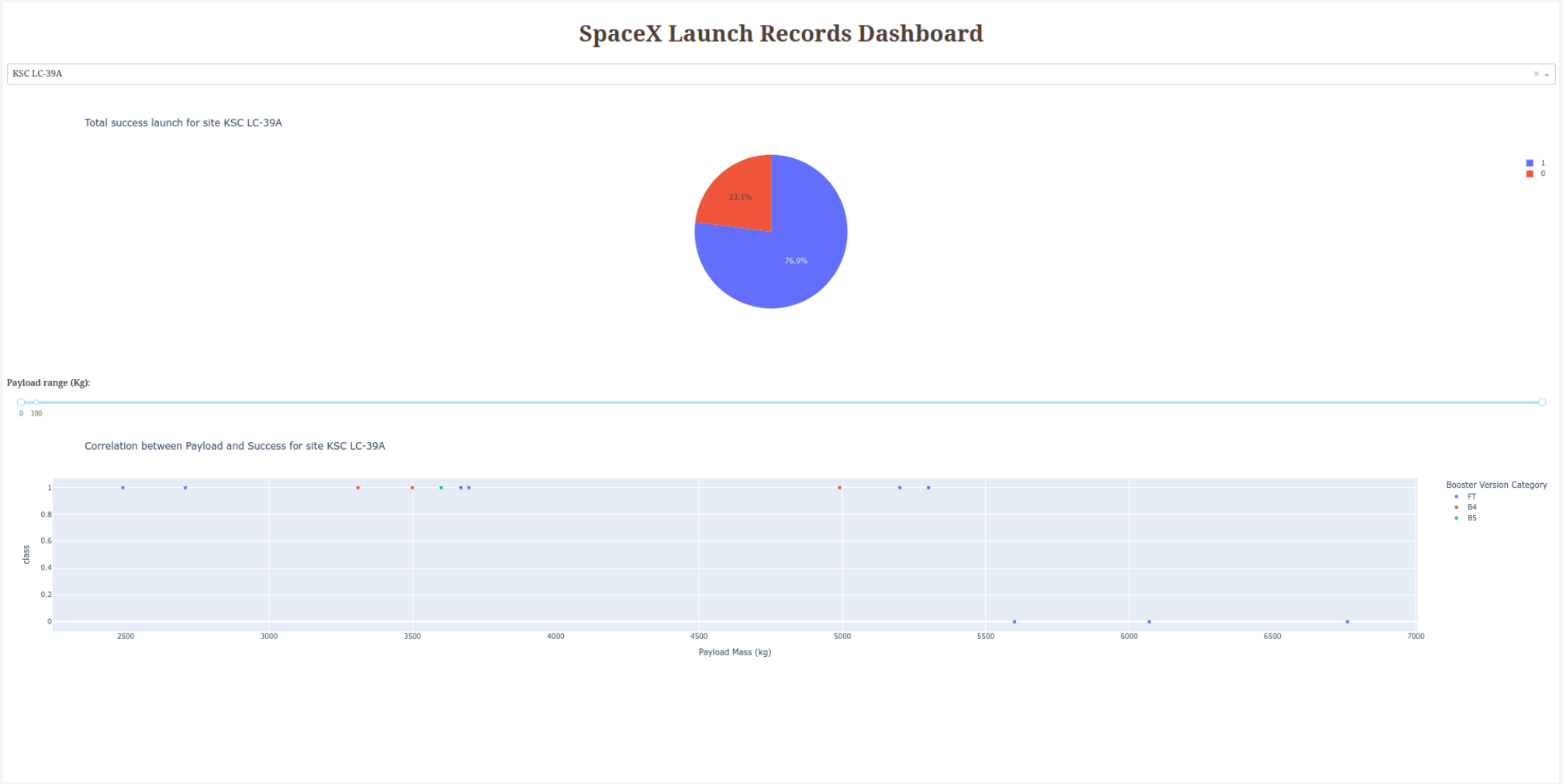



Section 4
Build a Dashboard with Plotly Dash

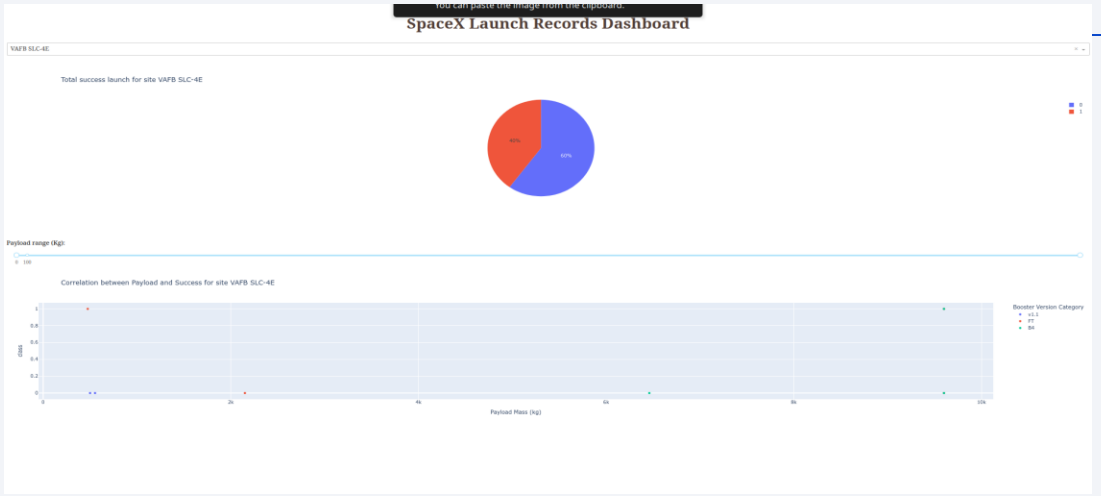
Total success launches by sites

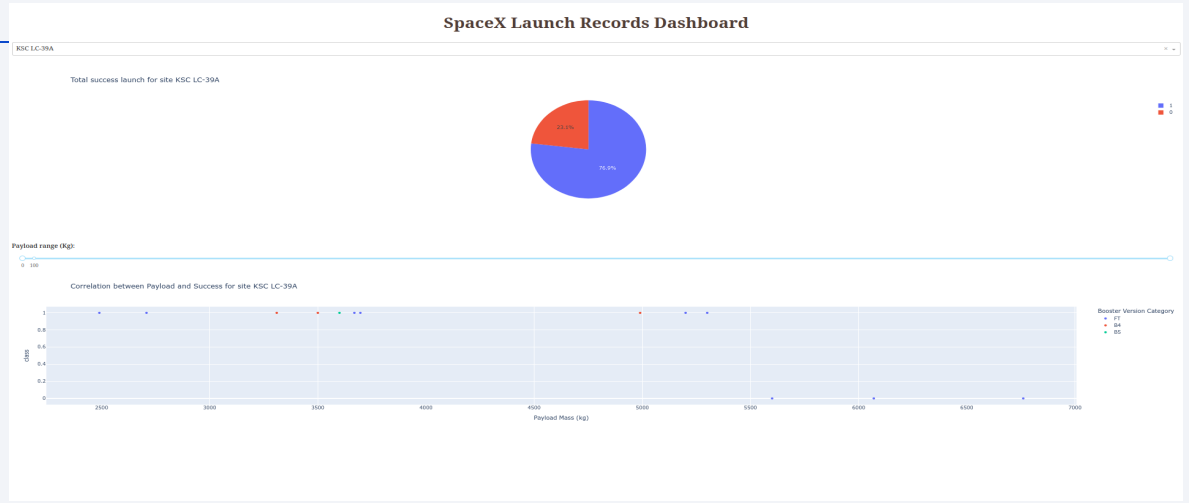


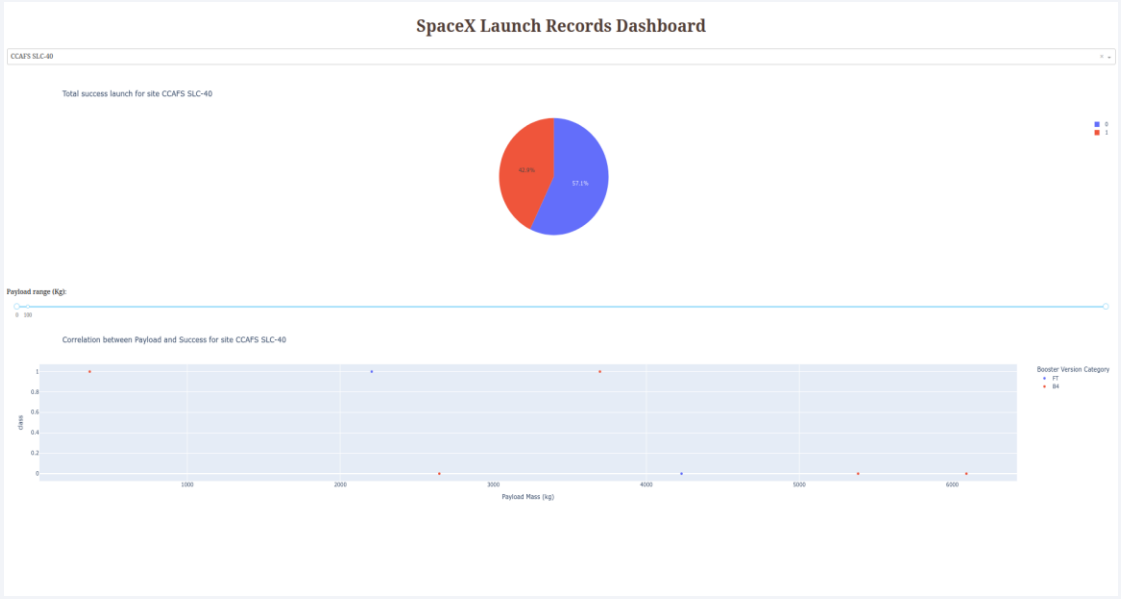
Highest launch success rate: KSC LC-39A

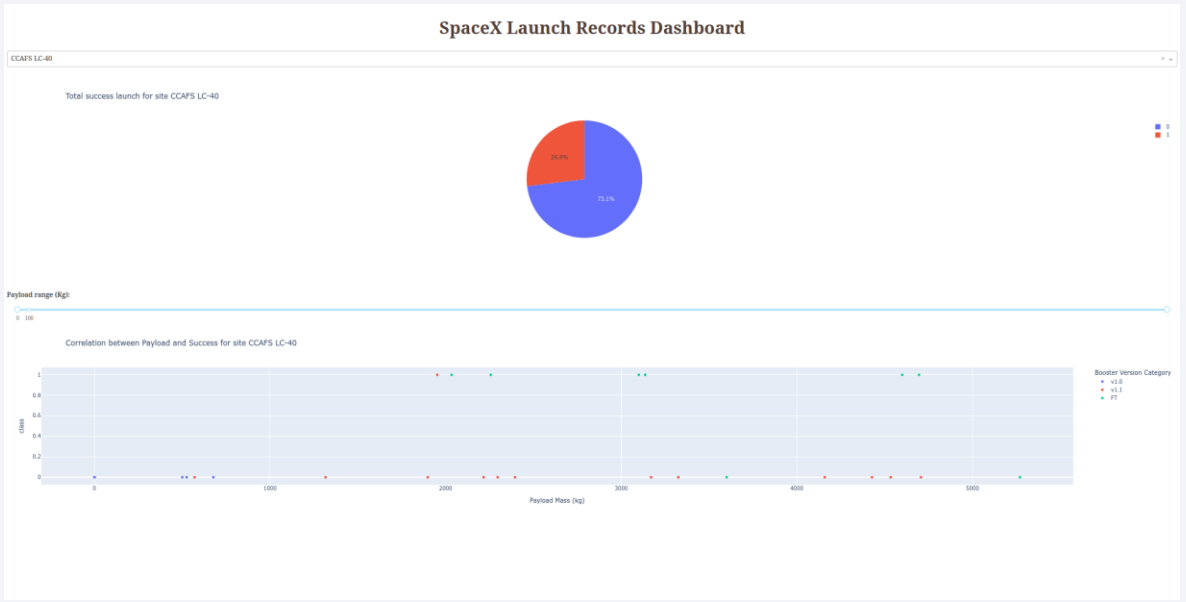


Pie and Scatter charts of all sites







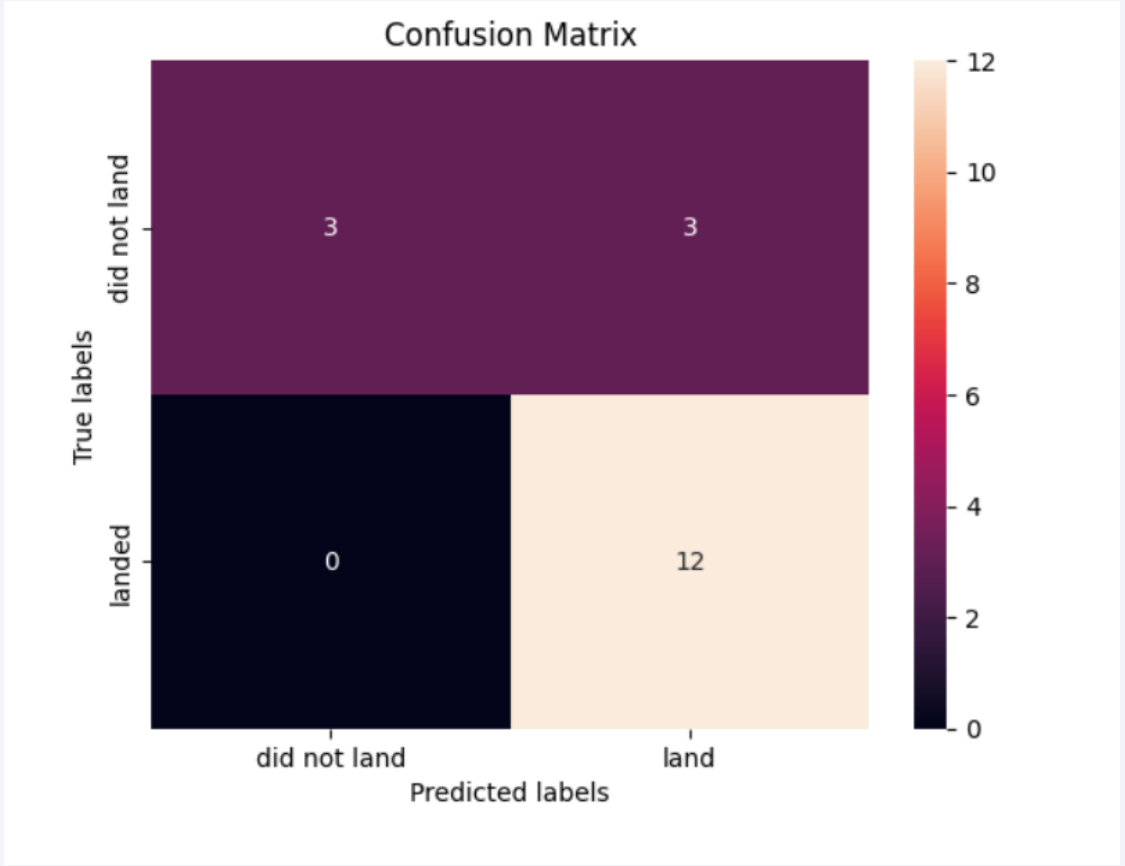


Section 5

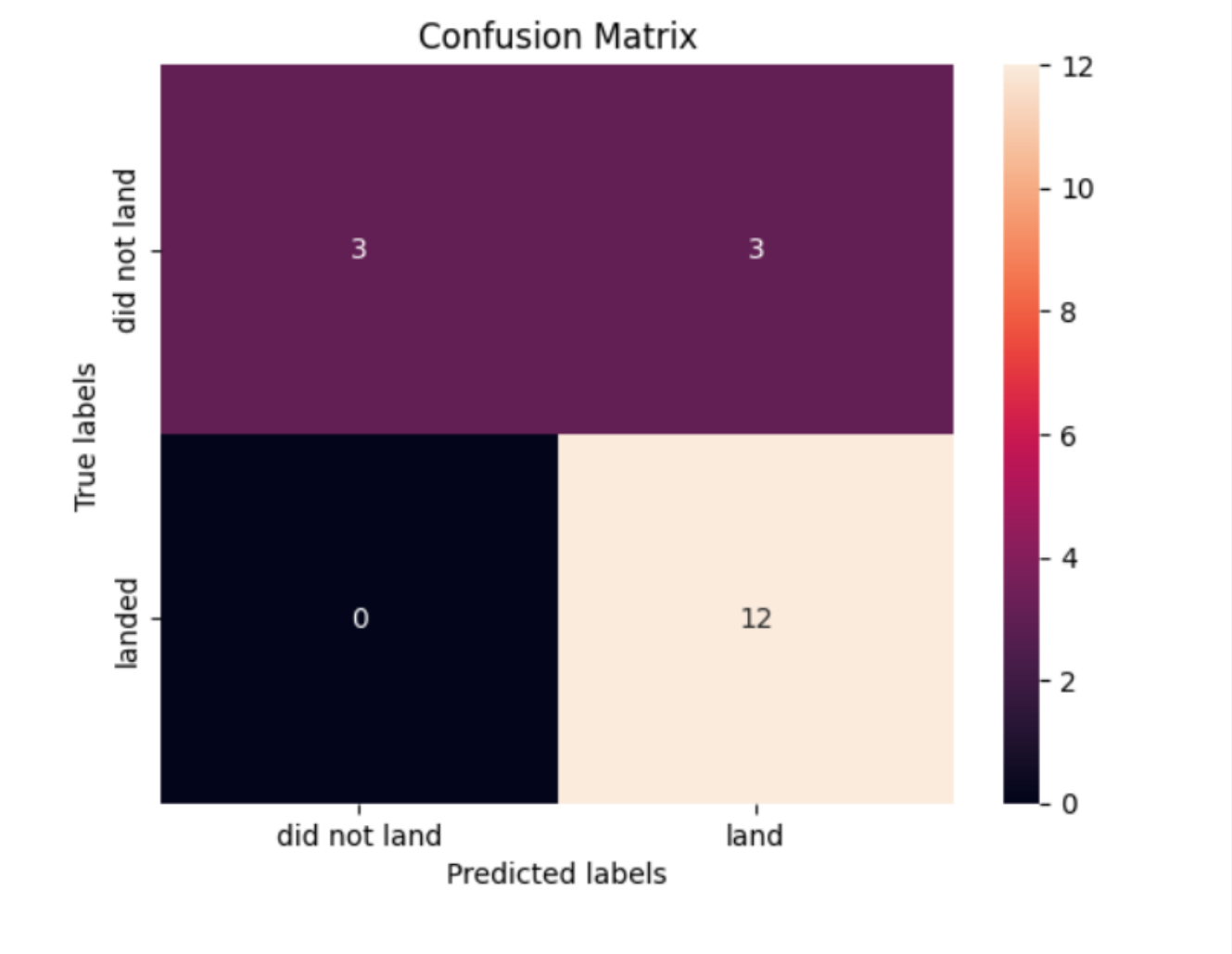
Predictive Analysis (Classification)

Classification Accuracy

- Decision tree classifier performed best with training data accuracy of 86.25% and test data accuracy 83.33%



Confusion Matrix



Conclusions

- Data collection was done using API and webscrapping.
- Performed Data wrangling on the acquired data.
- Performed EDA with SQL and Visualization and predictive features were selected.
- Different model were fitted to the data get the best accuracy and acquired best parameters of each model.
- Decision Tree Classifier performed best.
- Launch sites were marked using folium , with success/failure markings on each site.

Appendix

- SpaceX launch data.

```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json'
```

- List of Falcon 9 and Falcon Heavy launches wikipage updated on 9th June 2021.

```
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
```

Thank you!