IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Soheyla Moghadam
https://github.com/SoheylaMoghadam
22/08/2023

Outline

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

Executive Summary

Summary of methodologies

- Data collection
- Data wrangling
- EDA with data visualization & SQL
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive analysis (Classification)

Summary of all results

- EDA results
- Interactive analytics
- Predictive analysis

Introduction

- Project background and context

SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million vs 165 million USD, because SpaceX can reuse the first stage.

- Problems you want to find answers

This project will Predict Successful Landing of SpaceX Falcon 9 First Stage by using Machine Learning Models.

Section 1

Methodology

Methodology

Executive Summary

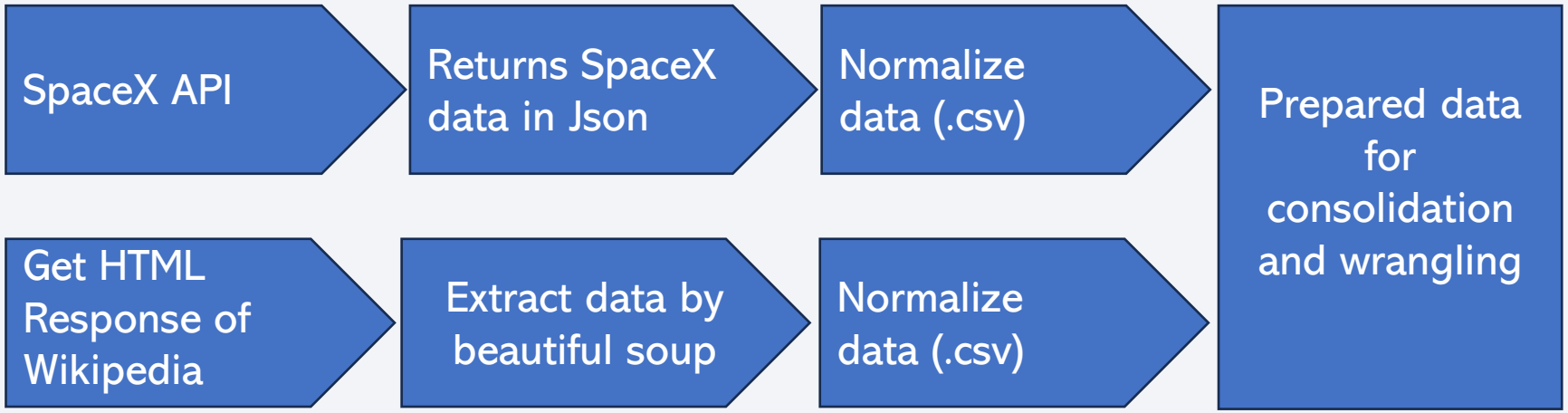
- Data collection methodology:
  - Combined dataset from the SpaceX public API and the SpaceX Wikipedia page
- Perform data wrangling
  - Cleaning data, handling missing values and encoding data for using machine learning models
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Tuned models using GridSearchCV and find the best classifier model

Data Collection

Data collection involves gathering relevant data API requests from Space X public API and web scraping data from a table in Space X's Wikipedia entry.

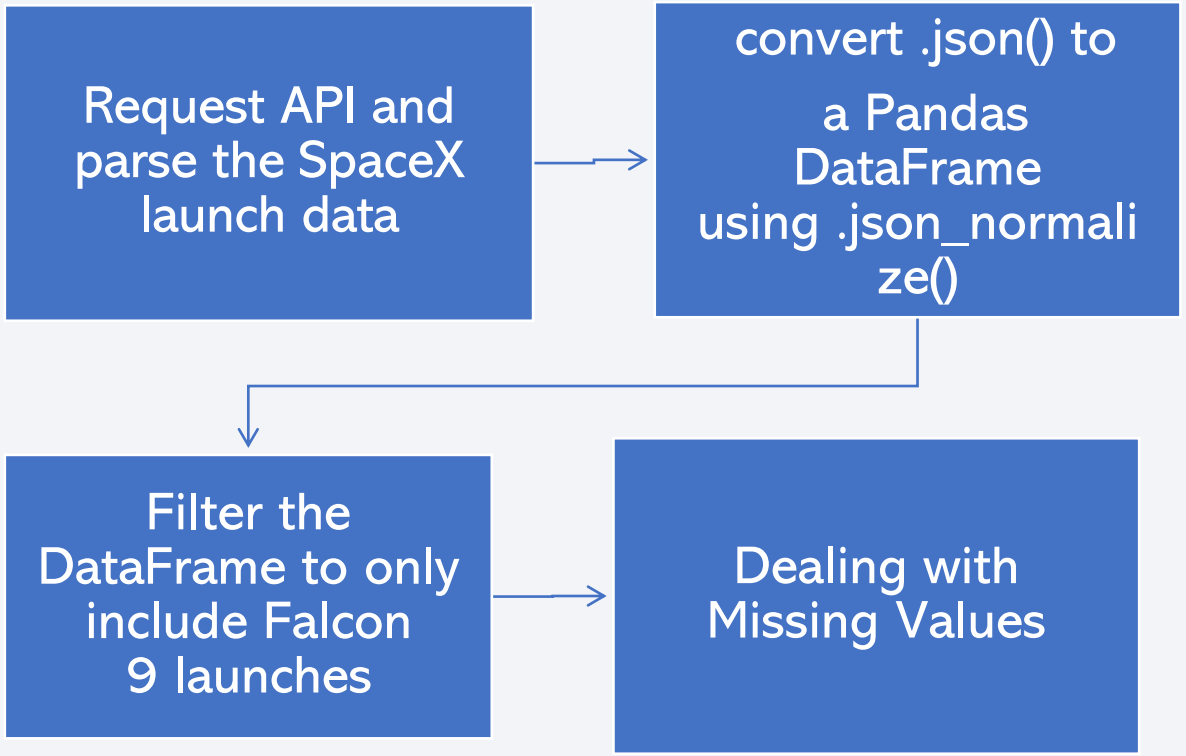
The datasets of SpaceX API include information such as rocket launches, mission details, payload information, and more.

The data columns of the Wikipedia webscraped dataset involve Flight No., Launch site, Payload, Payload Mass, Orbit, Customer, Launch outcome, Version, Booster, Booster landing, Date, and Time.



Data Collection – SpaceX API

Steps in the process of collecting SpaceX launch data using the SpaceX API:

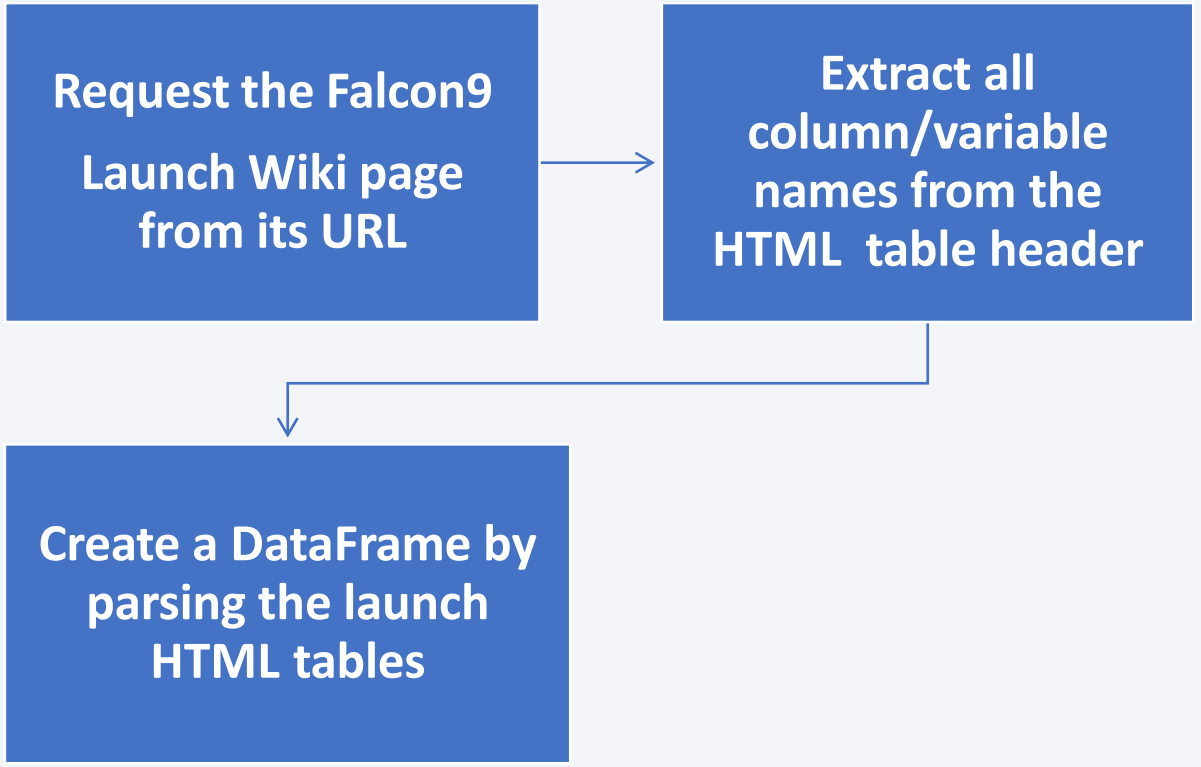


GitHub URL:

https://github.com/SoheylaMoghadam/Applied-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

Steps in the process of Web scrapping from Wikipedia:

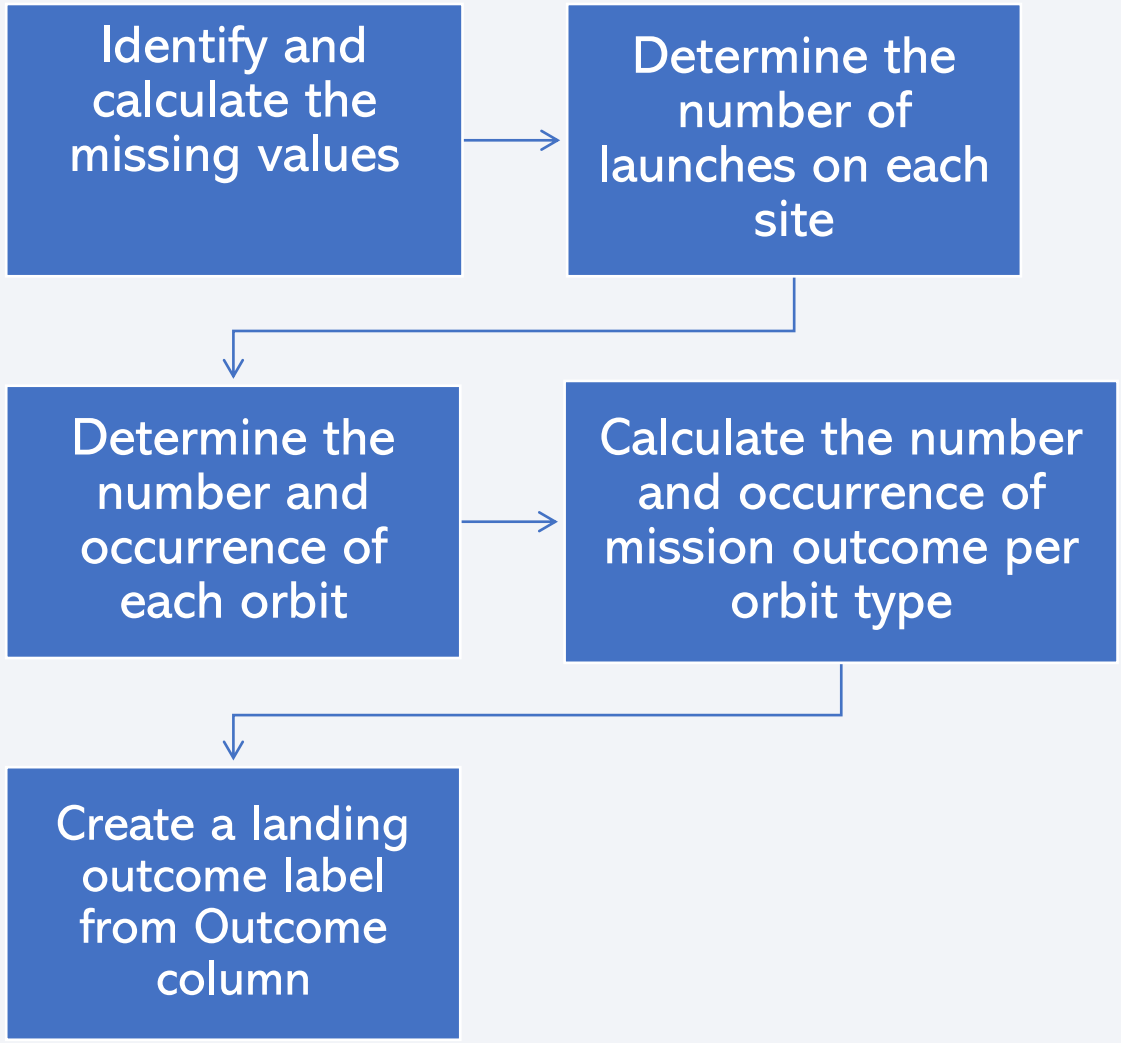


GitHub URL:

https://github.com/SoheylaMoghadam/Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

Steps of data wrangling:

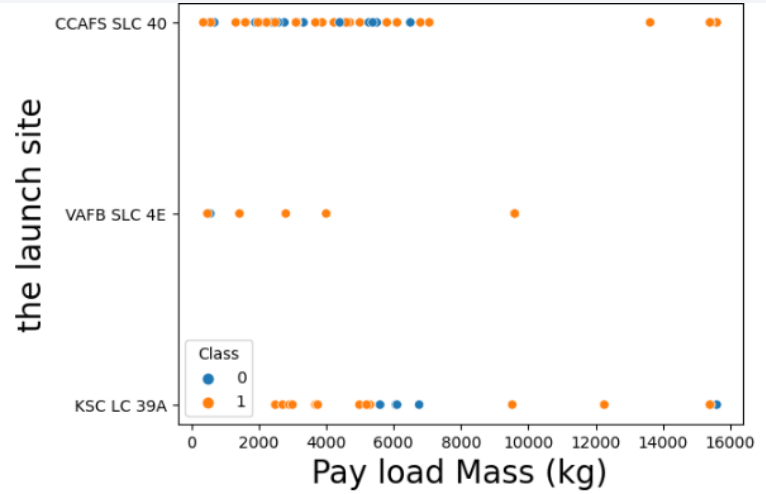


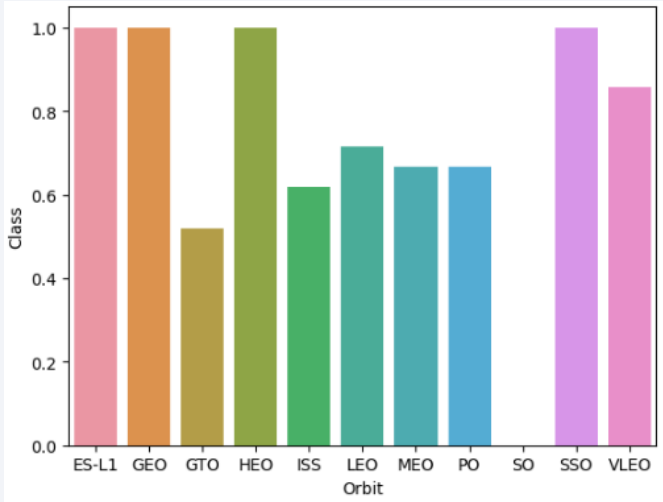
GitHub URL:

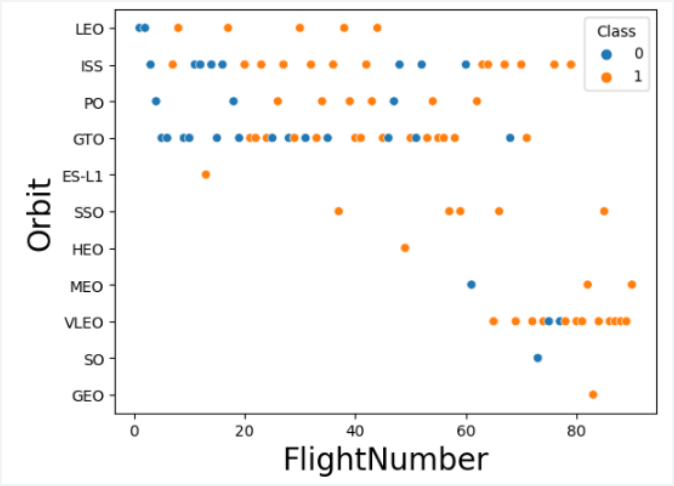
https://github.com/SoheylaMoghadam/Applied-Data-Science-Capstone/blob/main/labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

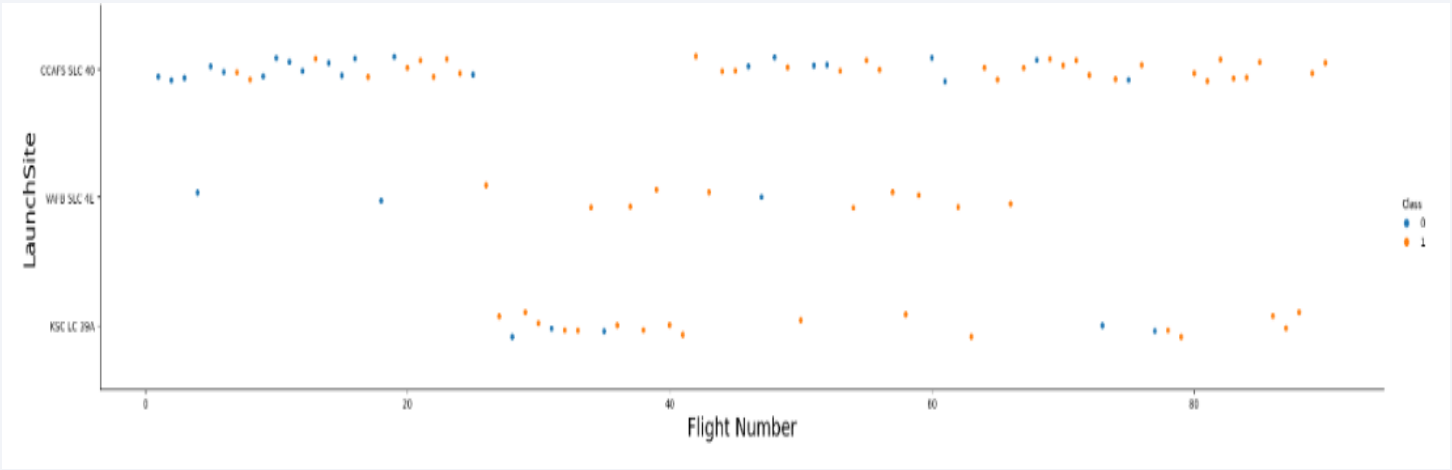
EDA with Data Visualization

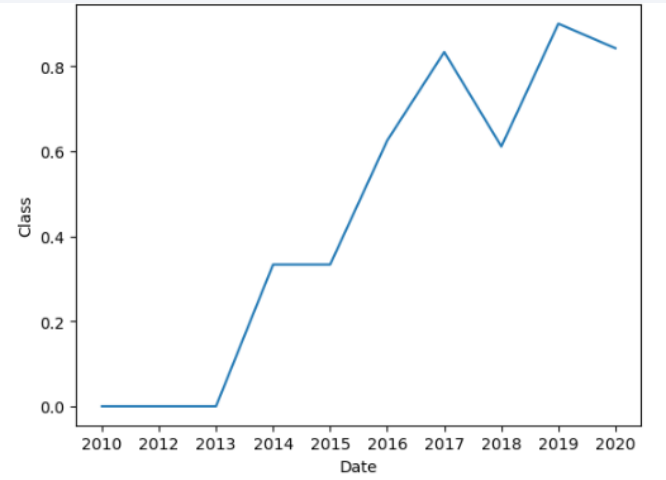
Exploratory Data Analysis performed on variables Flight Number, Payload Mass, Launch Site, Orbit, Class and Year.











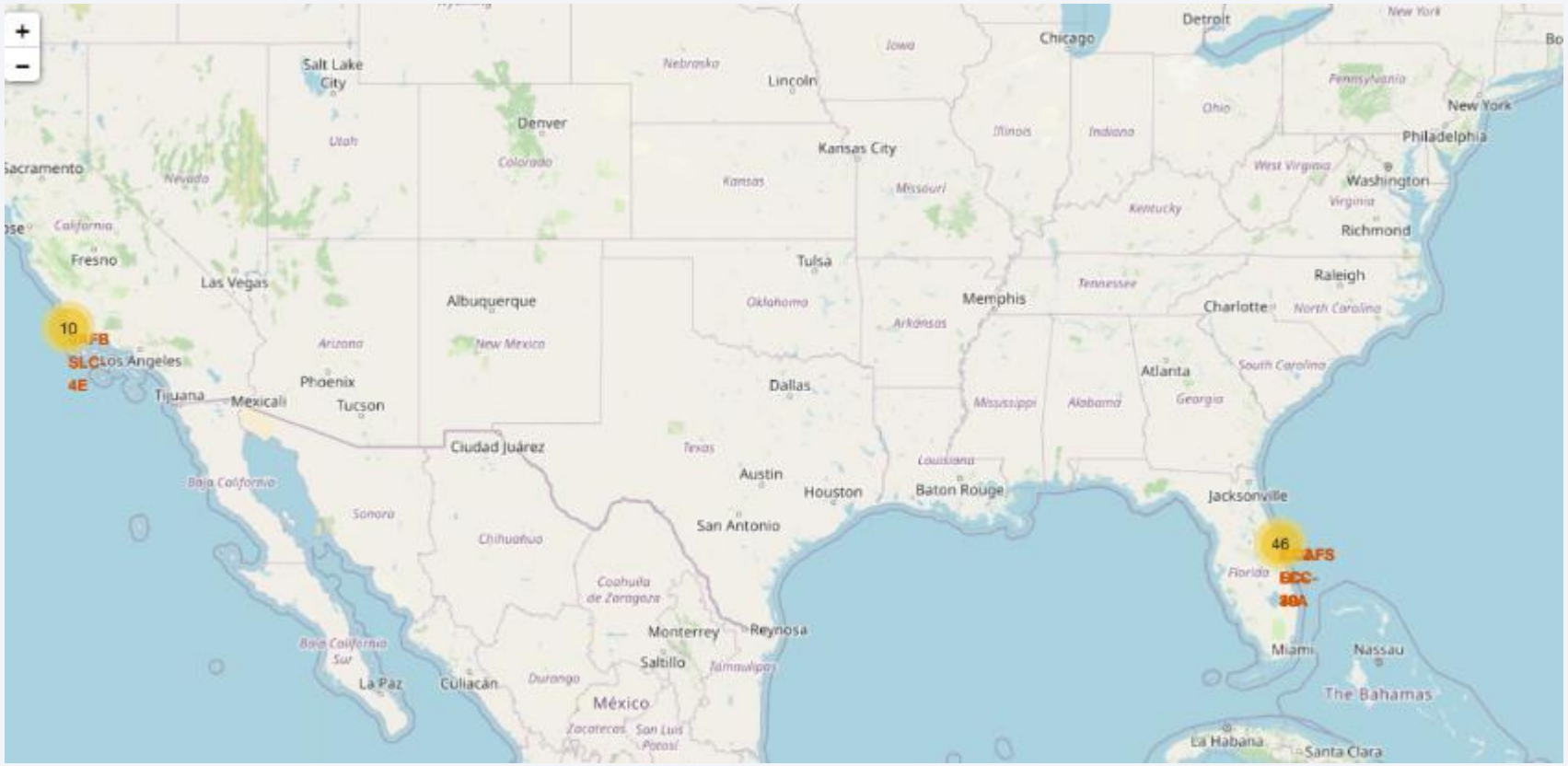
https://github.com/SoheylaMoghadam/Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Display the unique launch site names in the space mission.
- Show 5 records where launch sites start with the string 'KSC'.
- Display the total payload mass carried by boosters launched by NASA (CRS).
- Show the average payload mass carried by booster version F9 v1.1.
- List the date when successful landing outcomes were achieved on a drone ship.
- List the names of boosters that had success on a ground pad and had a payload mass greater than 4000 but less than 6000.
- List the total number of successful and failed mission outcomes.
- List the names of booster versions that carried the maximum payload mass.
- List the records displaying the month names, successful landing outcomes on ground pad booster versions, and launch site for the months in the year 2017.
- Rank the count of successful landing outcomes between the dates 2010-06-04 and 2017-03-20 in descending order.

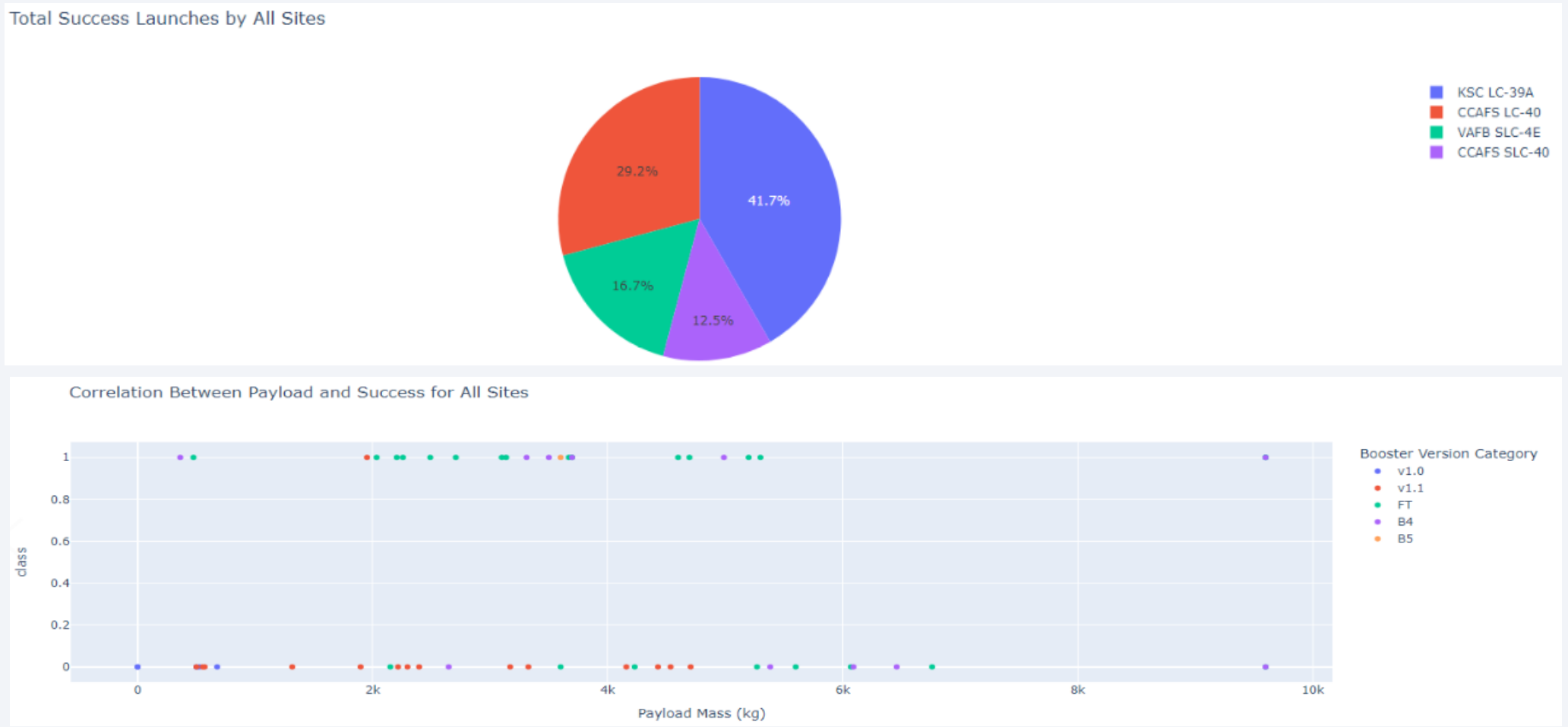
https://github.com/SoheylaMoghadam/Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium



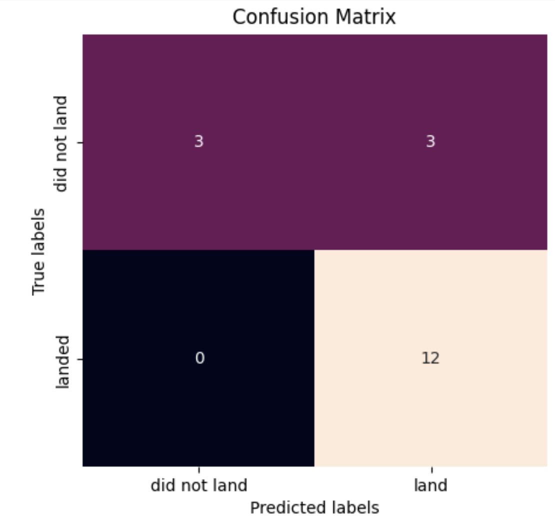
https://github.com/SoheylaMoghadam/Applied-Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

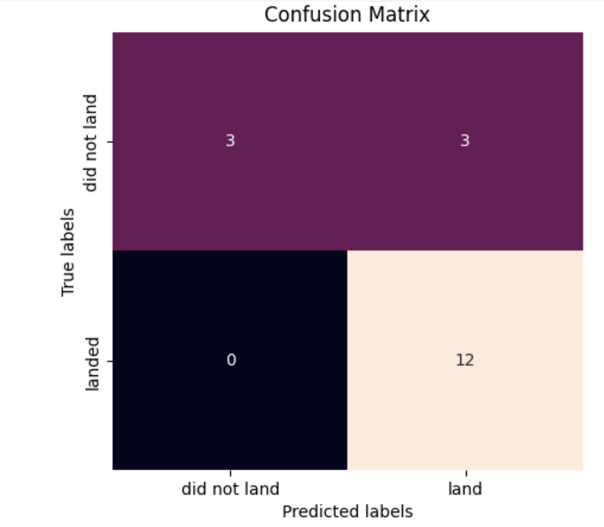
Build a Dashboard with Plotly Dash

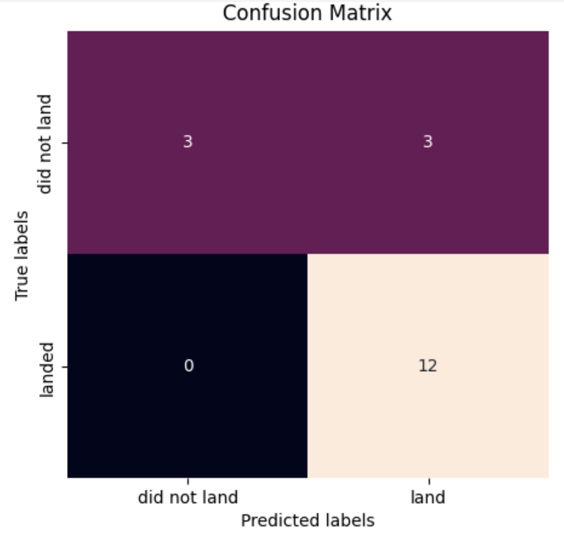


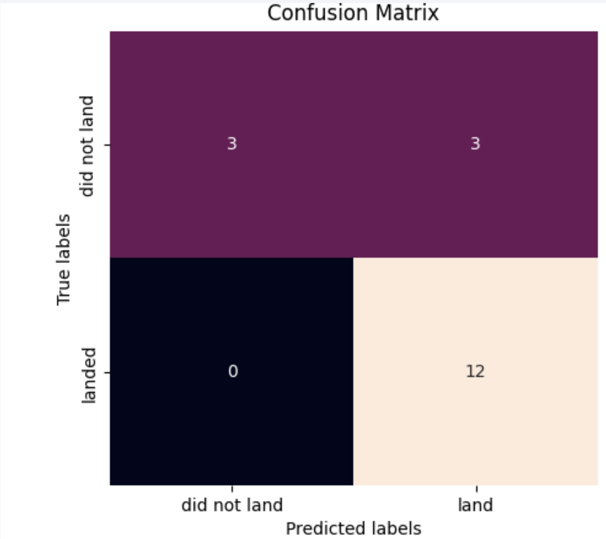
https://github.com/SoheylaMoghadam/Applied-Data-Science-Capstone/blob/main/spacex_dash_app.py

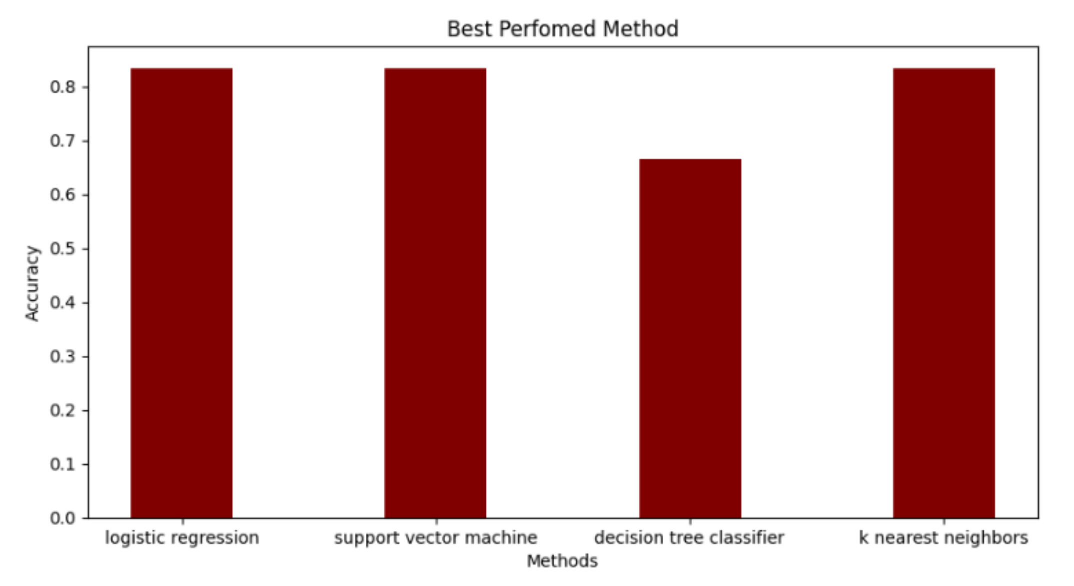
Predictive Analysis (Classification)







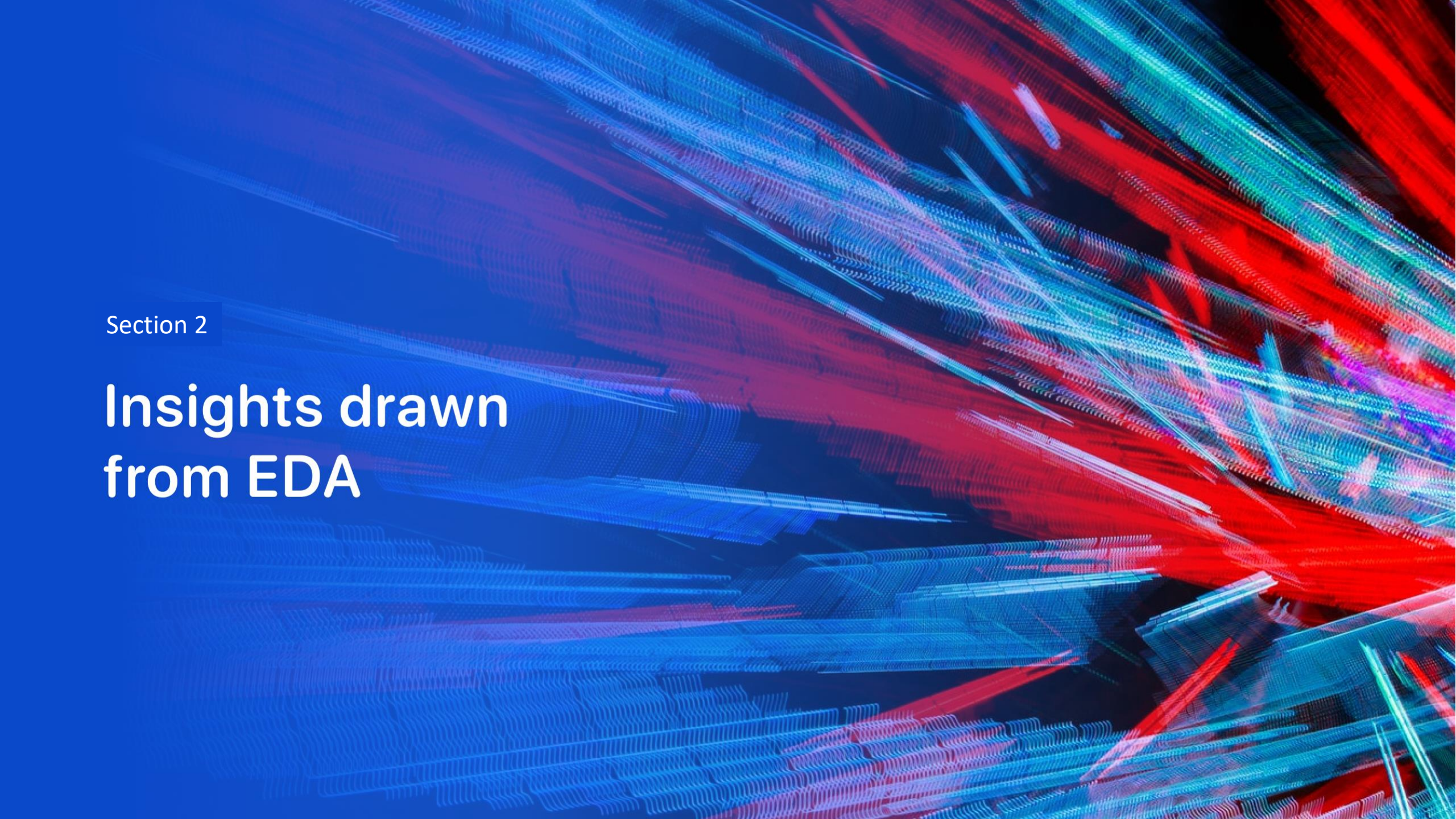




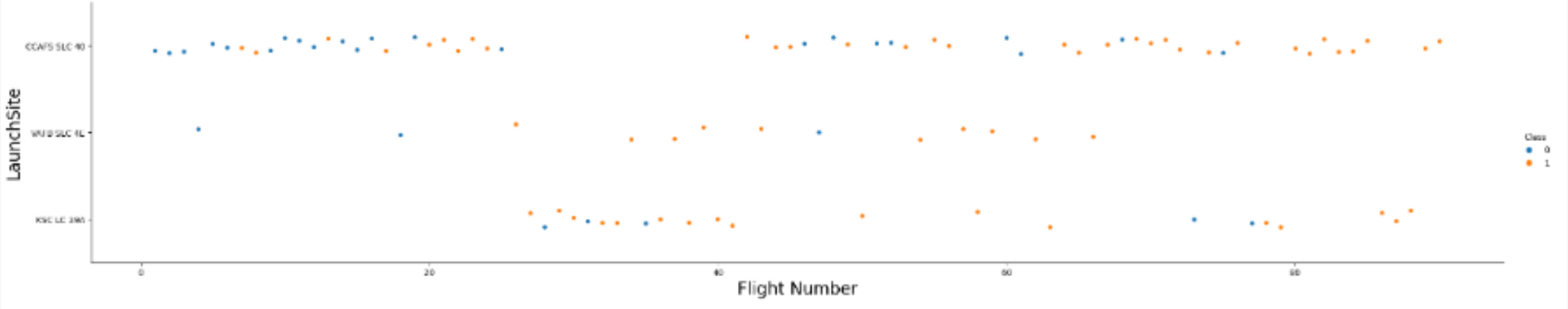
The models of SVM, KNN, and Logistic Regression demonstrated the highest level of accuracy, achieving a rate of 83.3%.

Results

- The SVM, KNN, and Logistic Regression models demonstrate superior prediction accuracy on this dataset.
- Lighter weighted payloads exhibit superior performance compared to heavier payloads.
- The success rates of SpaceX launches are expected to improve over time, with a direct correlation to the number of years they have been perfecting their launches.
- Among all the sites, KSC LC 39A has witnessed the highest number of successful launches.
- Orbits such as GEO, HEO, SSO, and ES L1 exhibit the highest success rates.

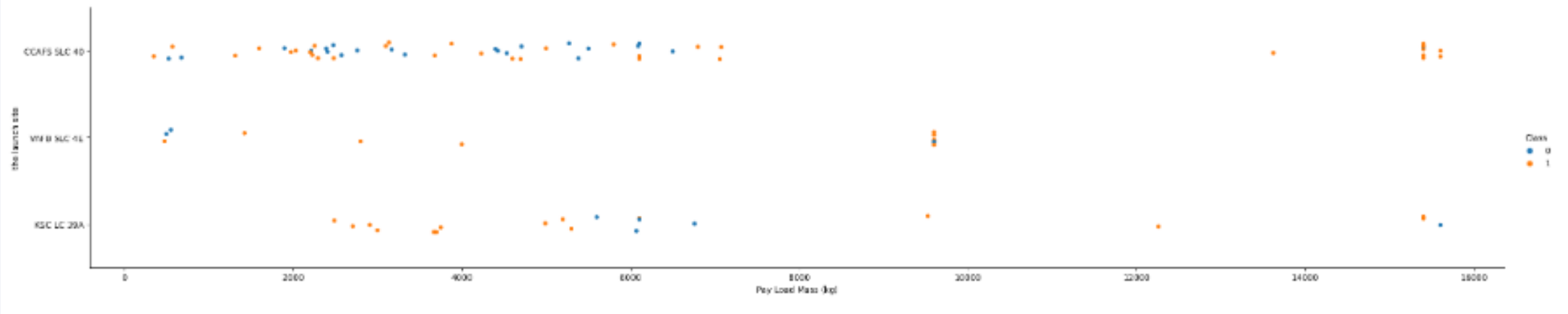

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



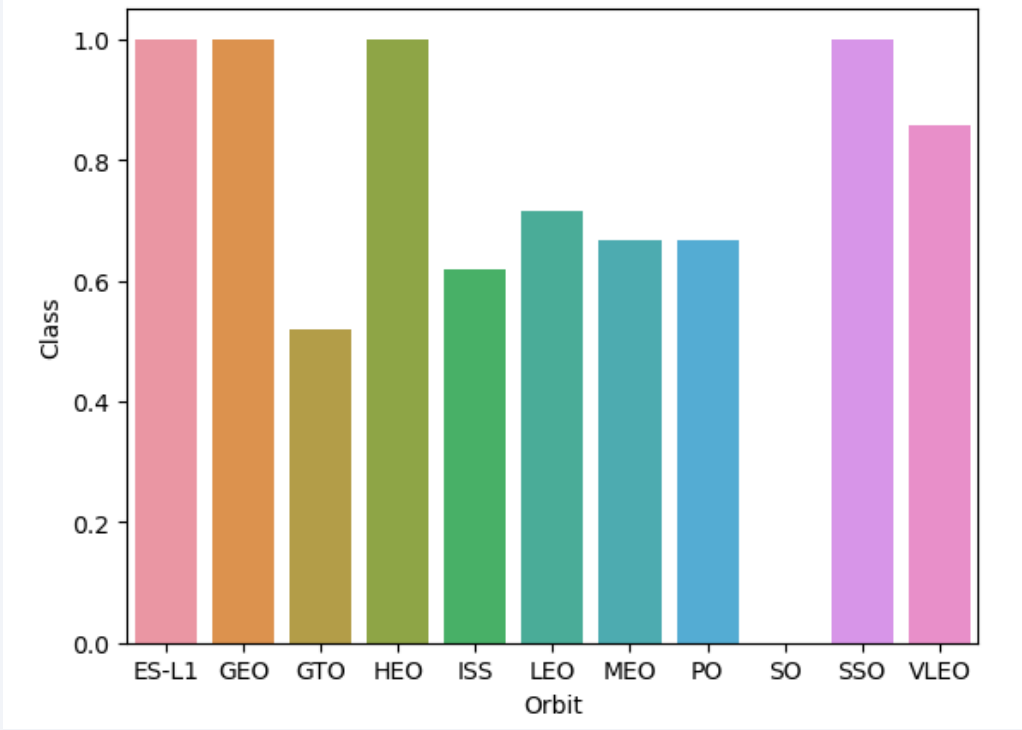
- Launches originating from CCAFS SLC 40 exceed the number of launches from alternative sites by a substantial margin

Payload vs. Launch Site



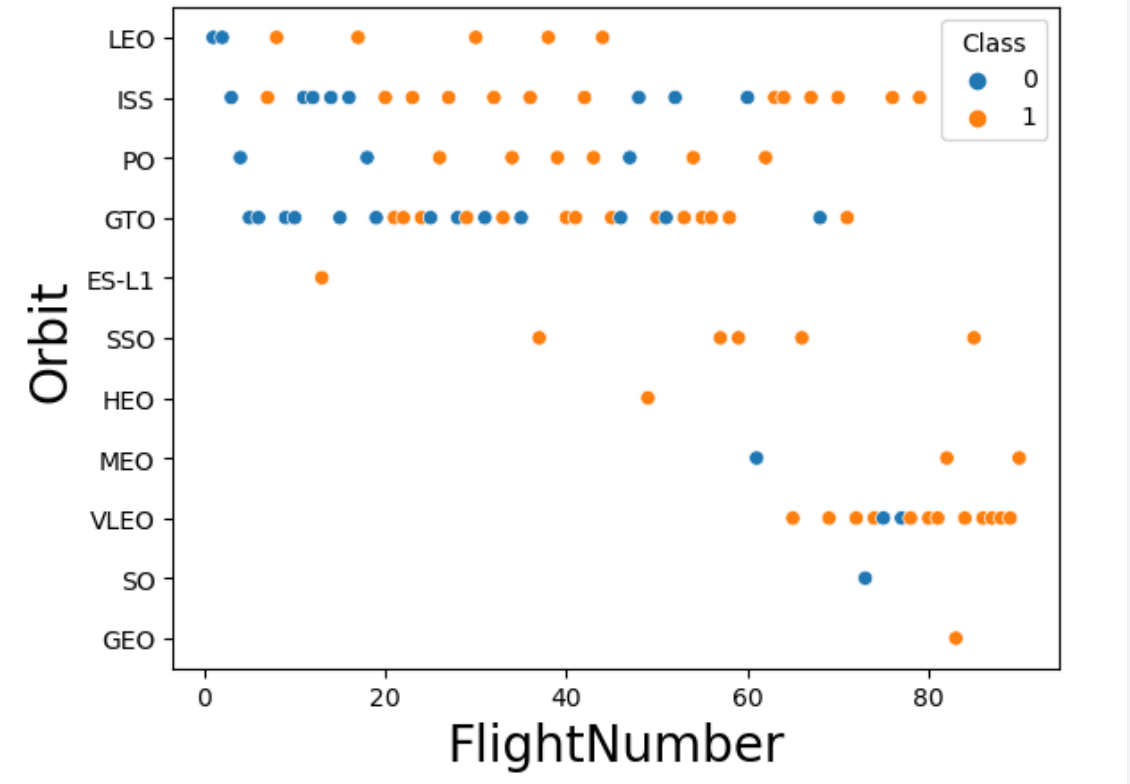
- CCAFS SLC 40 has witnessed the launch of a significant number of payloads that possess comparatively lower masses.

Success Rate vs. Orbit Type

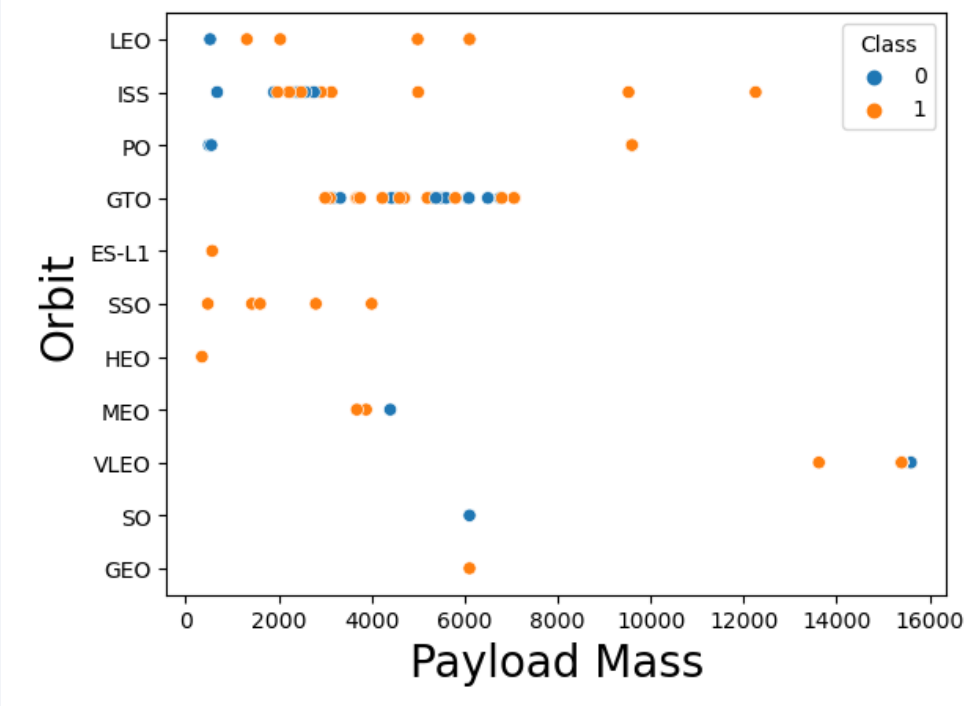


- ES-L1, GEO, HEO, and SSO orbit types demonstrate exceptionally high success rates.

Flight Number vs. Orbit Type

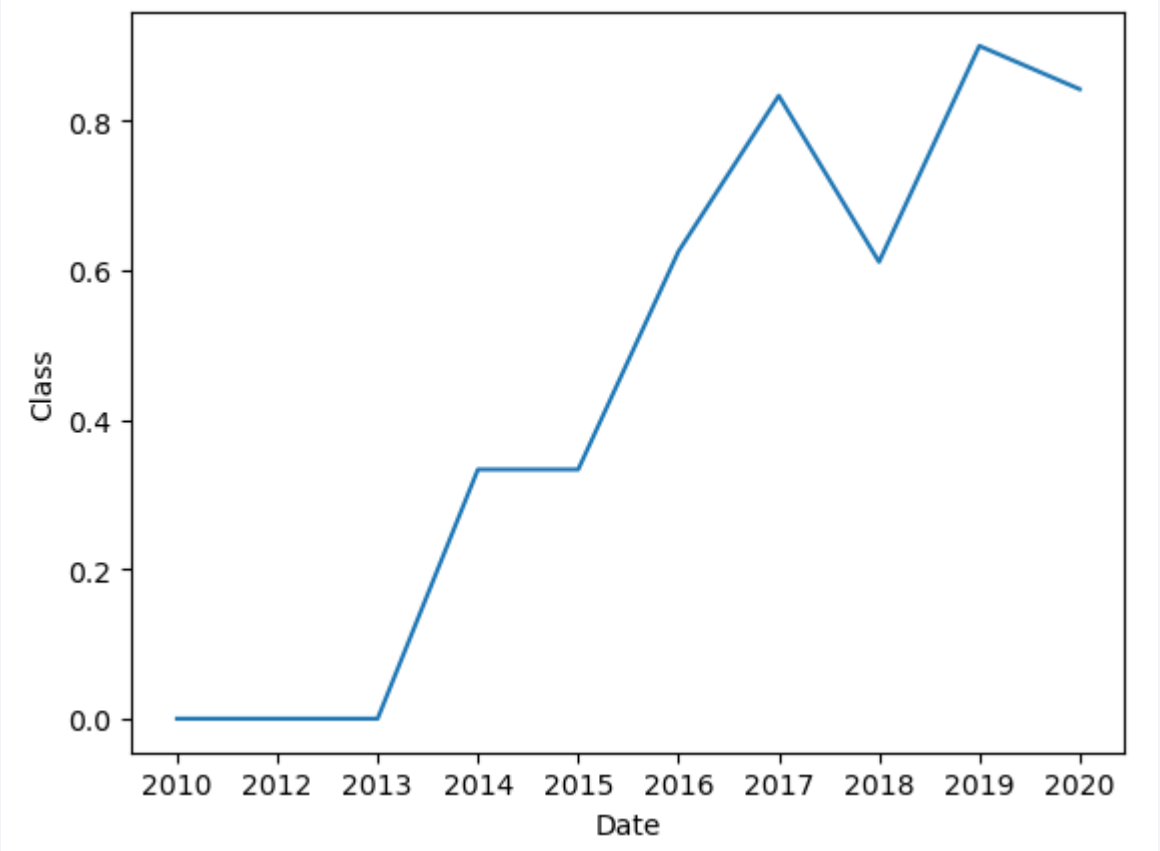


Payload vs. Orbit Type



- The orbit of the International Space Station (ISS) not only supports a wide range of payload sizes but also demonstrates a remarkable success rate.

Launch Success Yearly Trend



- The success rate of launches began to rise in 2013 and continued to increase steadily until 2020.

All Launch Site Names

Display the names of the unique launch sites in the space mission

In [44]:
```
%sql select distinct "launch_site" from SPACEXTBL
```

```
 * sqlite:///my_data1.db
Done.
```

Out[44]:

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

Launch Site Names Begin with 'CCA'

Out[45]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

Display the total payload mass carried by boosters launched by NASA (CRS)

In [46]:
```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where customer = 'NASA (CRS)'
```

```
 * sqlite:///my_data1.db
Done.
```

Out[46]:

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

Display average payload mass carried by booster version F9 v1.1

In [47]:
```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where "Booster_Version" = "F9 v1.1"
```

```
 * sqlite:///my_data1.db
Done.
```

Out[47]:

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad

```
In [48]: %sql select min(DATE) from SPACEXTBL where "Landing_Outcome" = 'Success (ground pad)'
 * sqlite:///my_data1.db
Done.
```

Out[48]:

min(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

In [49]:
```
%sql select booster_version from SPACEXTBL where "Landing_Outcome"='Success (drone ship)' and PAYLOAD_MASS__KG_ between 400(
```

```
 * sqlite:///my_data1.db
Done.
```

Out[49]:

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

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

In [50]:
```
%sql select Mission_Outcome, count(*) from SPACEXTBL group by Mission_Outcome
```

```
 * sqlite:///my_data1.db
Done.
```

Out[50]:

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

Boosters Carried Maximum Payload

- The given list shows the names of the booster versions that have achieved the highest payload masses

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

- %sql select substr(Date, 4, 2) as month, "Landing_Outcome", Booster_Version , Launch_Site from SPACEXTBL where "Landing_Outcome" = 'Failure (drone ship)' and substr(Date,7,4)='2015'

- This query returns the Month, Landing Outcome, Booster Version, Payload Mass (kg), and Launch site of 2015 launches where stage 1 failed to land on a drone ship.

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

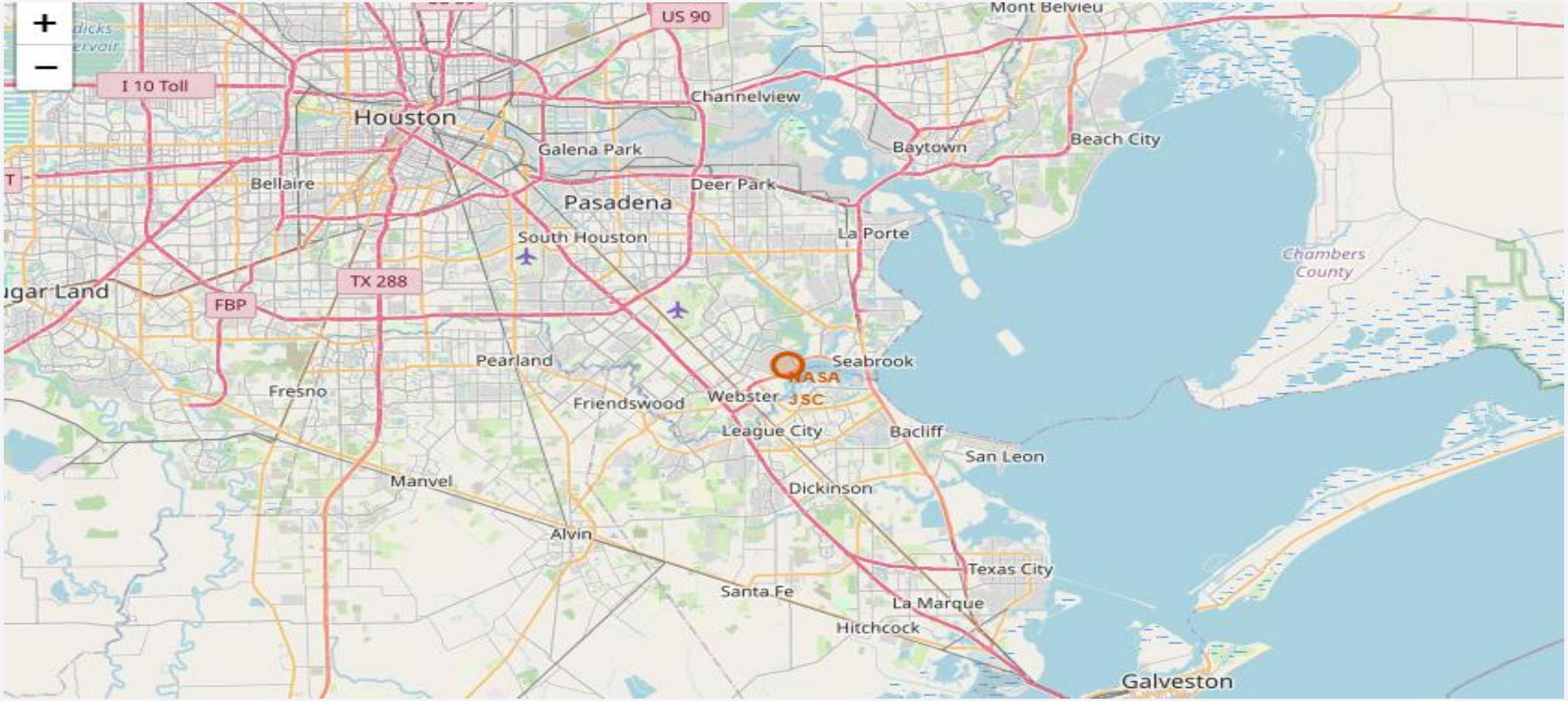
- %sql SELECT "LANDING__OUTCOME", COUNT(*) AS COUNT_LAUNCHES FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING__OUTCOME ORDER BY COUNT_LAUNCHES DESC;

- This query returns a list of successful landings and between 2010-06-04 and 2017-03-20 inclusively.

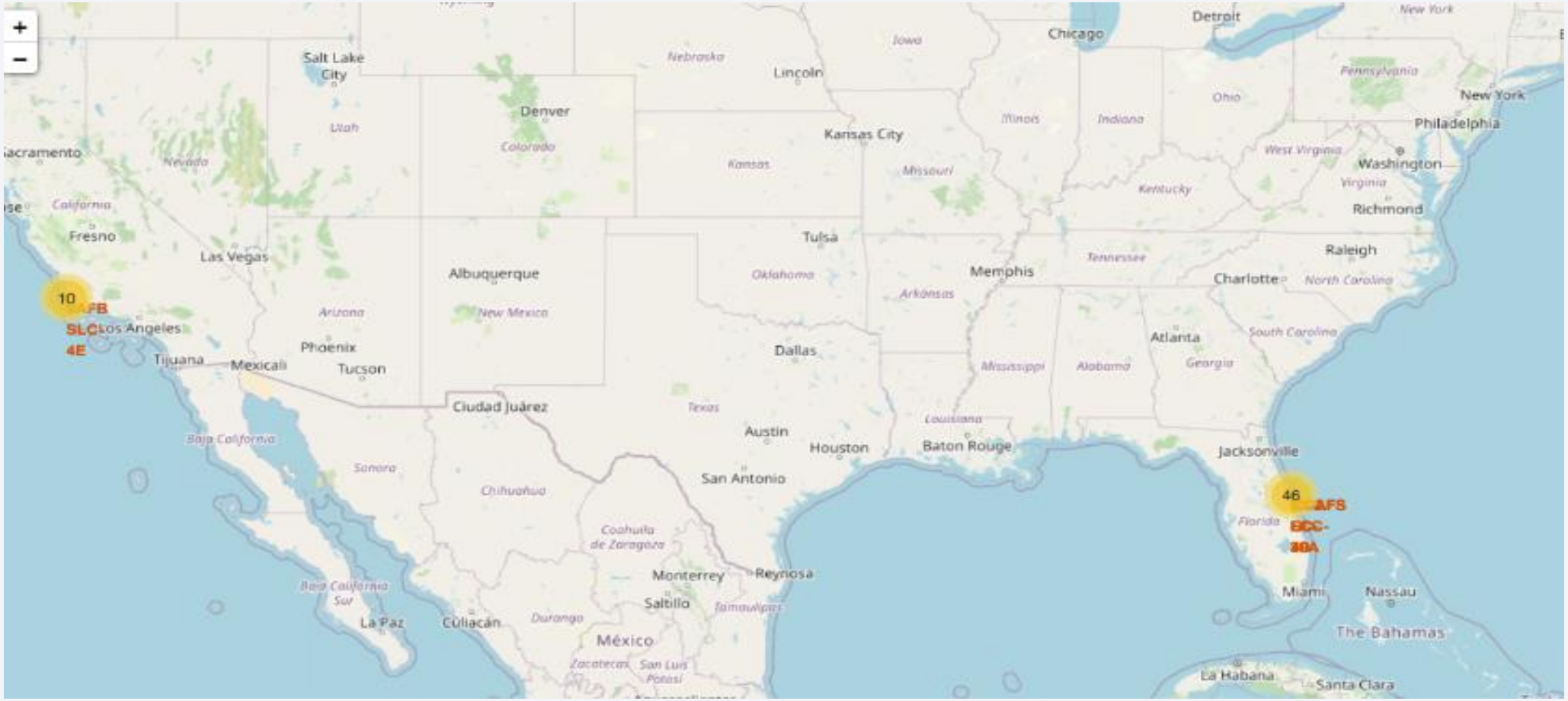
Section 3

Launch Sites Proximities Analysis

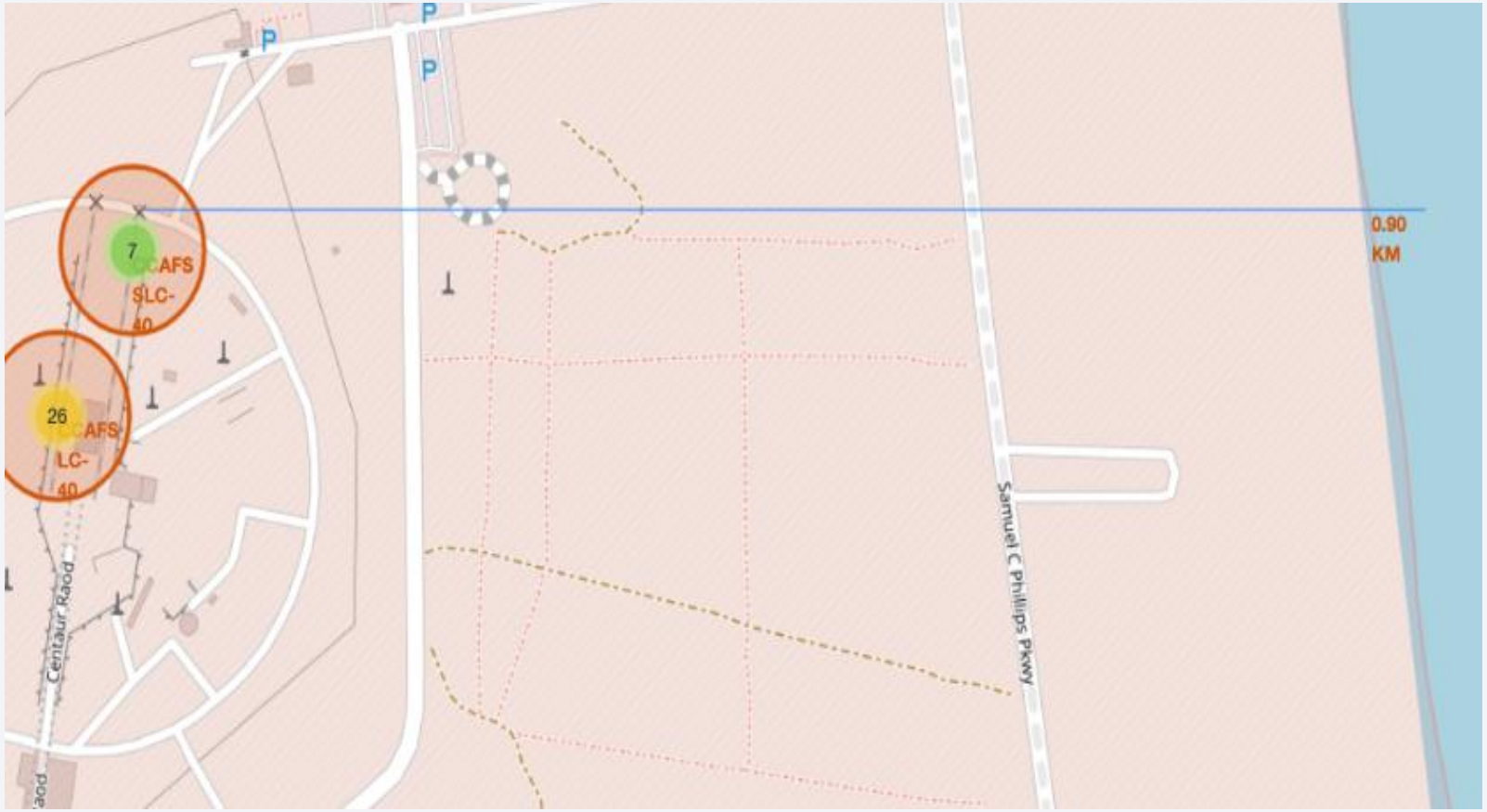
Folium Map Screenshot 1

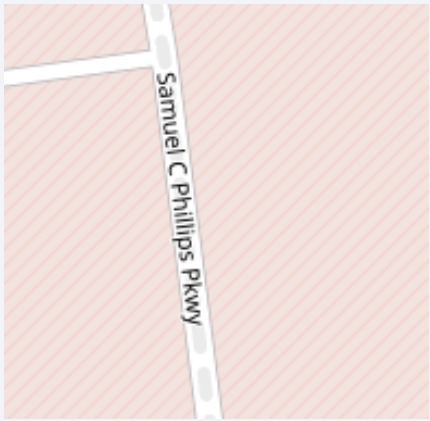


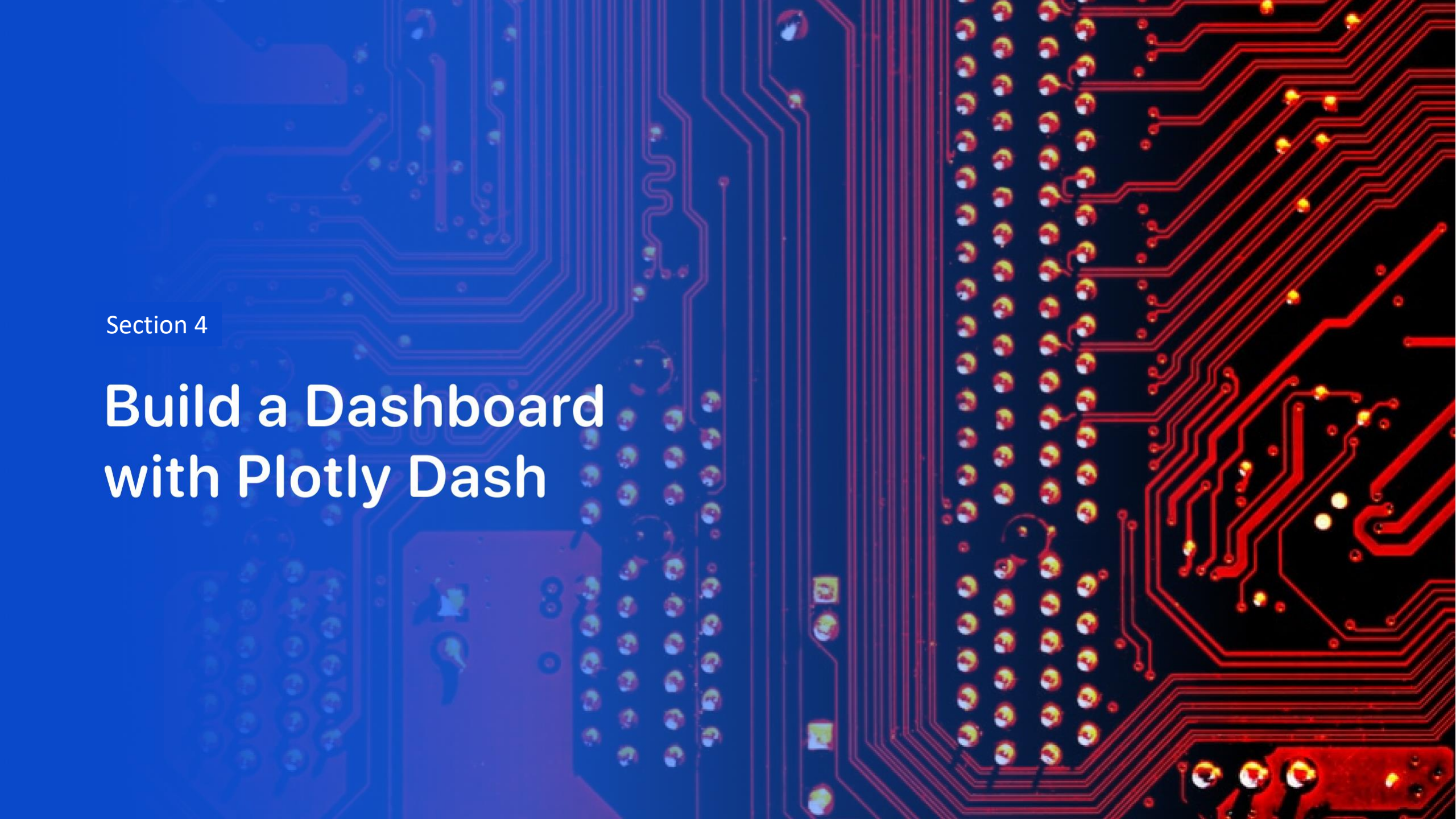
Folium Map Screenshot 2



Folium Map Screenshot 3






Section 4
Build a Dashboard with Plotly Dash

SpaceX Launch Records Dashboard

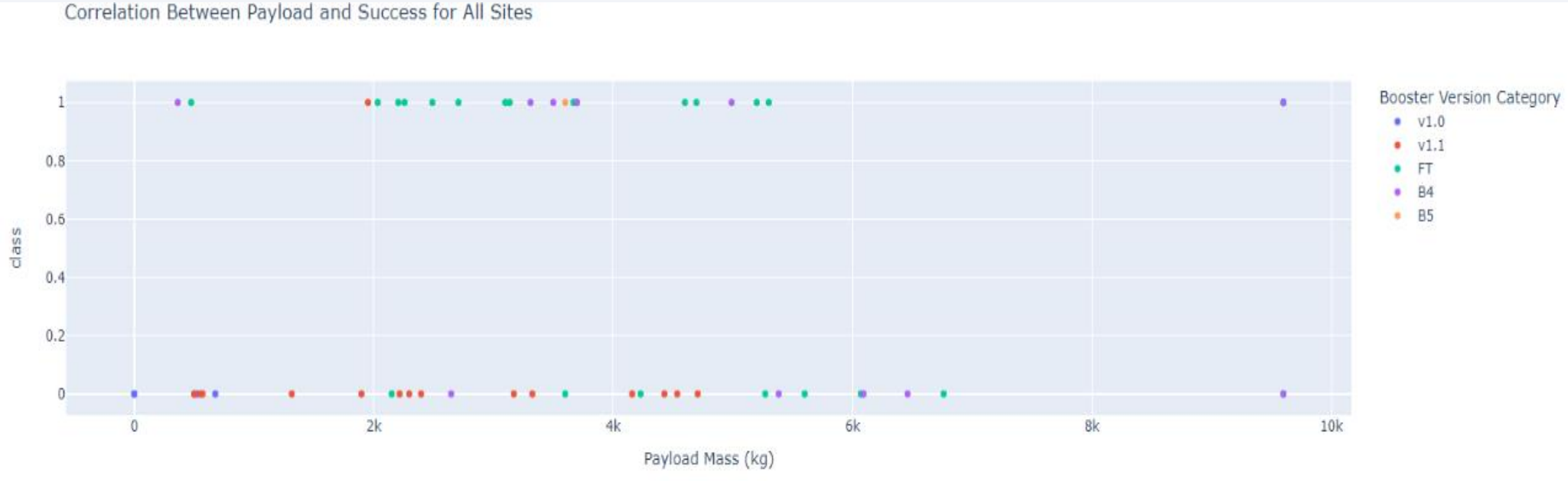
Total success launches by all sites

SpaceX Launch Records Dashboard

Total success launches by site: KSC LC-39A

SpaceX Launch Records Dashboard

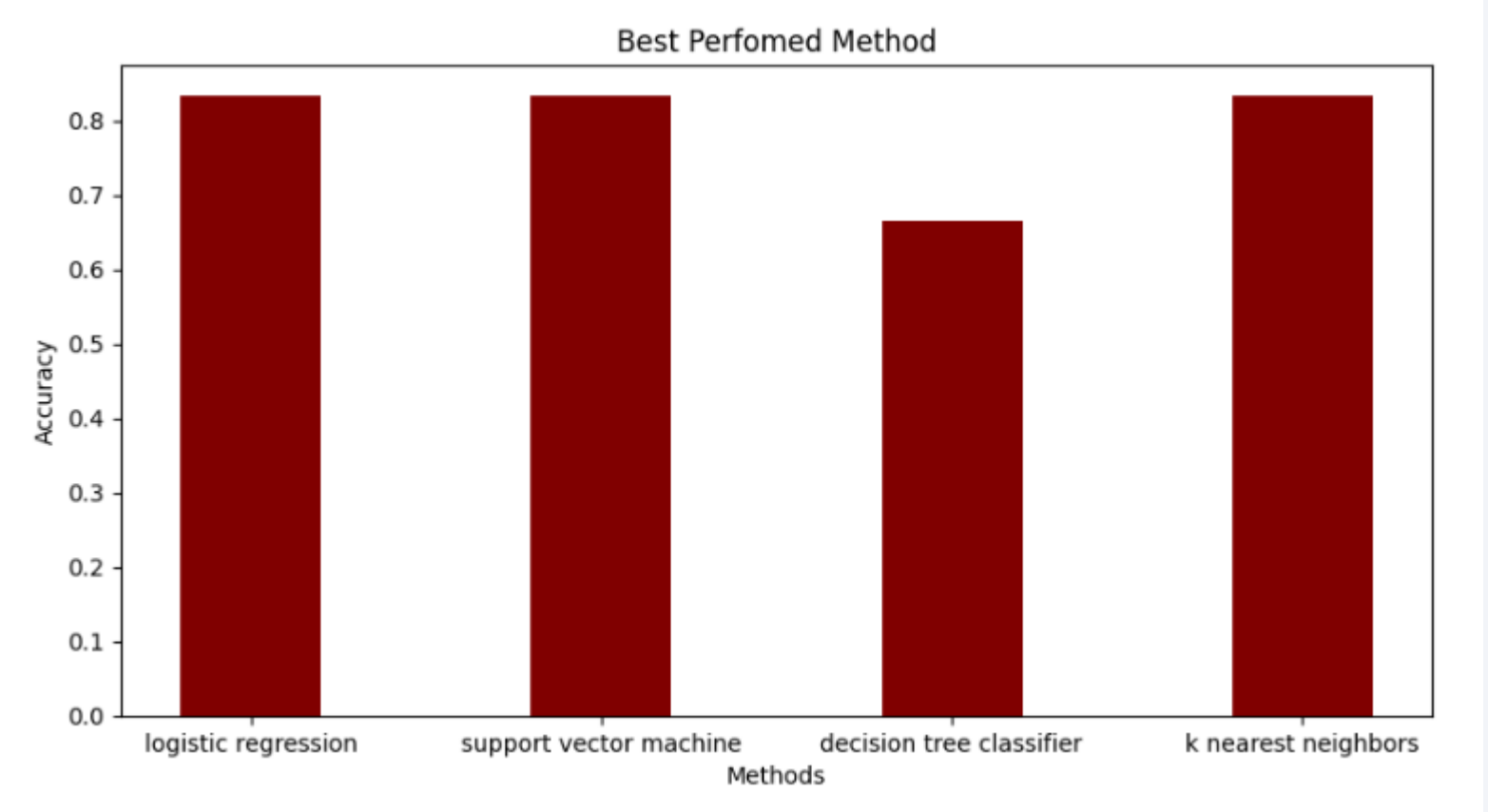
The given scatter plot shows the correlation between payload and launch success



Section 5

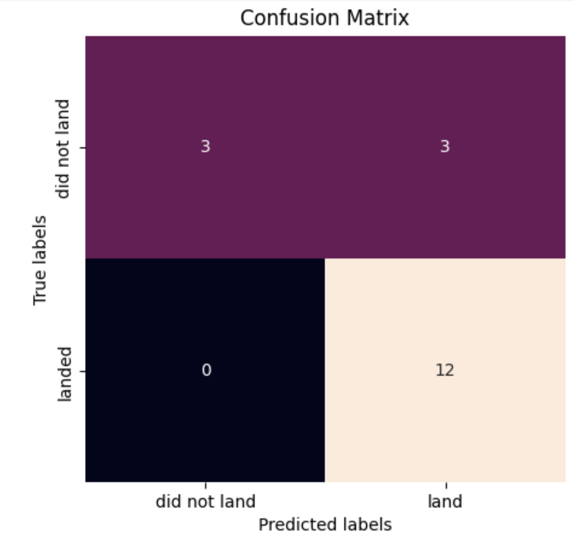
Predictive Analysis (Classification)

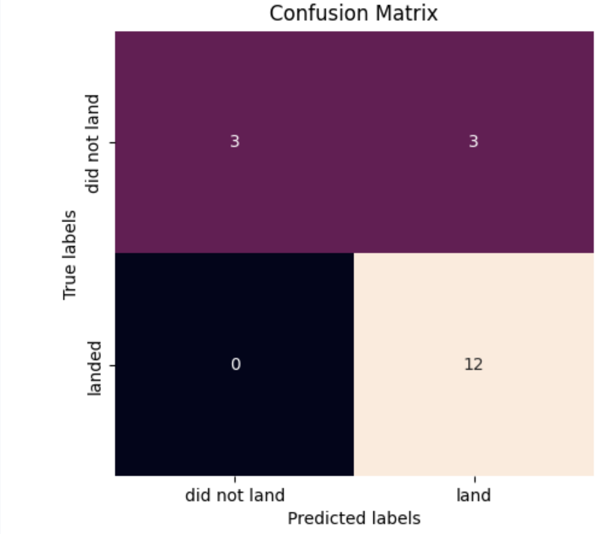
Classification Accuracy

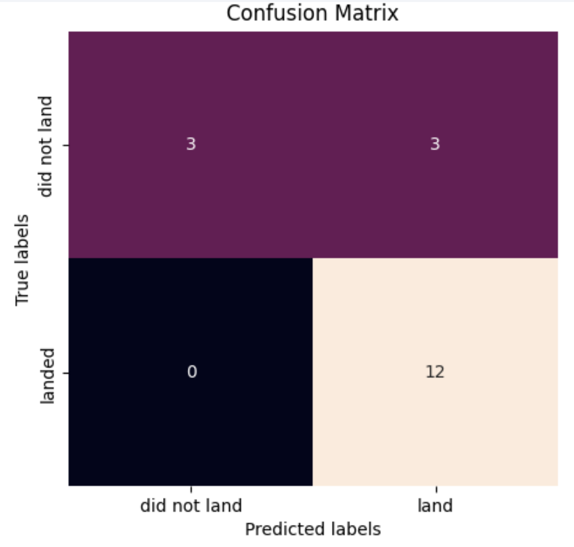


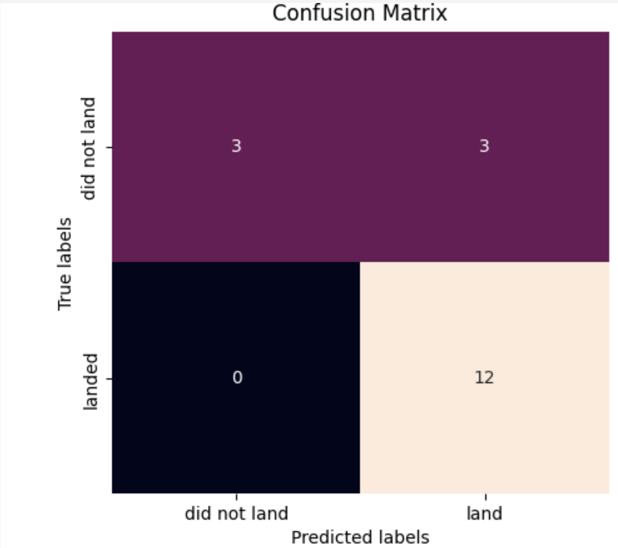
SVM, KNN and Logistic Regression have the highest classification accuracy

Confusion Matrix









Conclusions

- The SVM, KNN, and Logistic Regression models exhibit better prediction accuracy compared to Decision Tree model on this dataset.
- Payloads with lower weights exhibit better performance compared to heavier payloads.
- The success rates of SpaceX launches increase proportionally with the number of years dedicated to perfecting the launches.
- Among all the launch sites, KSC LC 39A stands out with the highest number of successful launches.

Thank you!