Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - EDA with Data Visualization
  - Building an interactive map with Folium
  - Building a Dashboard with Plotly Dash
  - Predictive Analysis (Classification)
- Summary of all results
  - EDA Result
  - Interactive Analytics
  - Predictive Analysis

Introduction

- Project background and context
  - SpaceX advertises Falcon rocket launches on it’s website, with a cost of 62 Million dollars; other providers cost upward of 165 Million dollars, each much of the savings is because SpaceX can reuse the first stage.
- Problems you want to find answers
  - The project task is to predicting if the 1^{st} stage of the SpaceX Falcon rocket will land successfully.

Section 1

Methodology

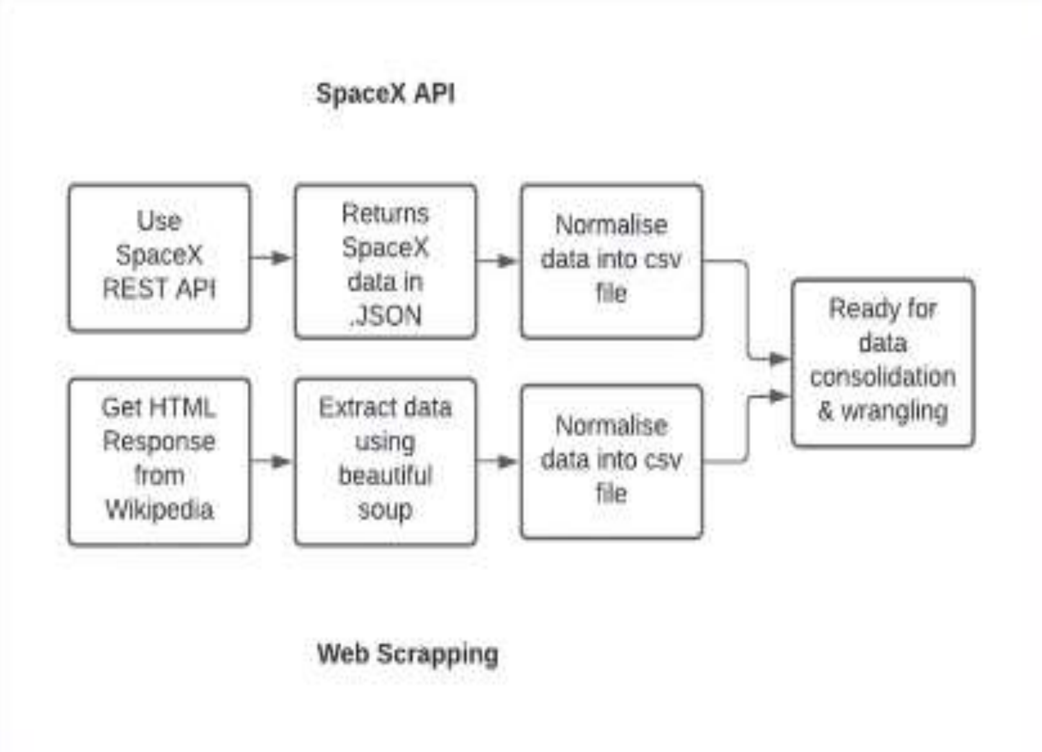
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX Rest API
- Perform data wrangling
  - One Hot Encoding data fields for machine learning and Data Cleaning of null values and irrelevant 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
  - LR, KNN, SVM, DT model have been built and evaluated for the best Classifier.

Data Collection

- The following dataset was collected:
  - SpaceX launch data that is gathered from the SpaceX REST API.
  - This API will give us data about launches, including information about the rocket used, payload delivered, launch specification, landing specifications and landing outcome.
  - The SpaceX REST API endpoints, or URL, starts with api.spacexdata.com/v4/.
  - Another popular data source for obtaining Falcon 9 Launch data is web scrapping Wikipedia using BeautifulSoup.



Data Collection – SpaceX API

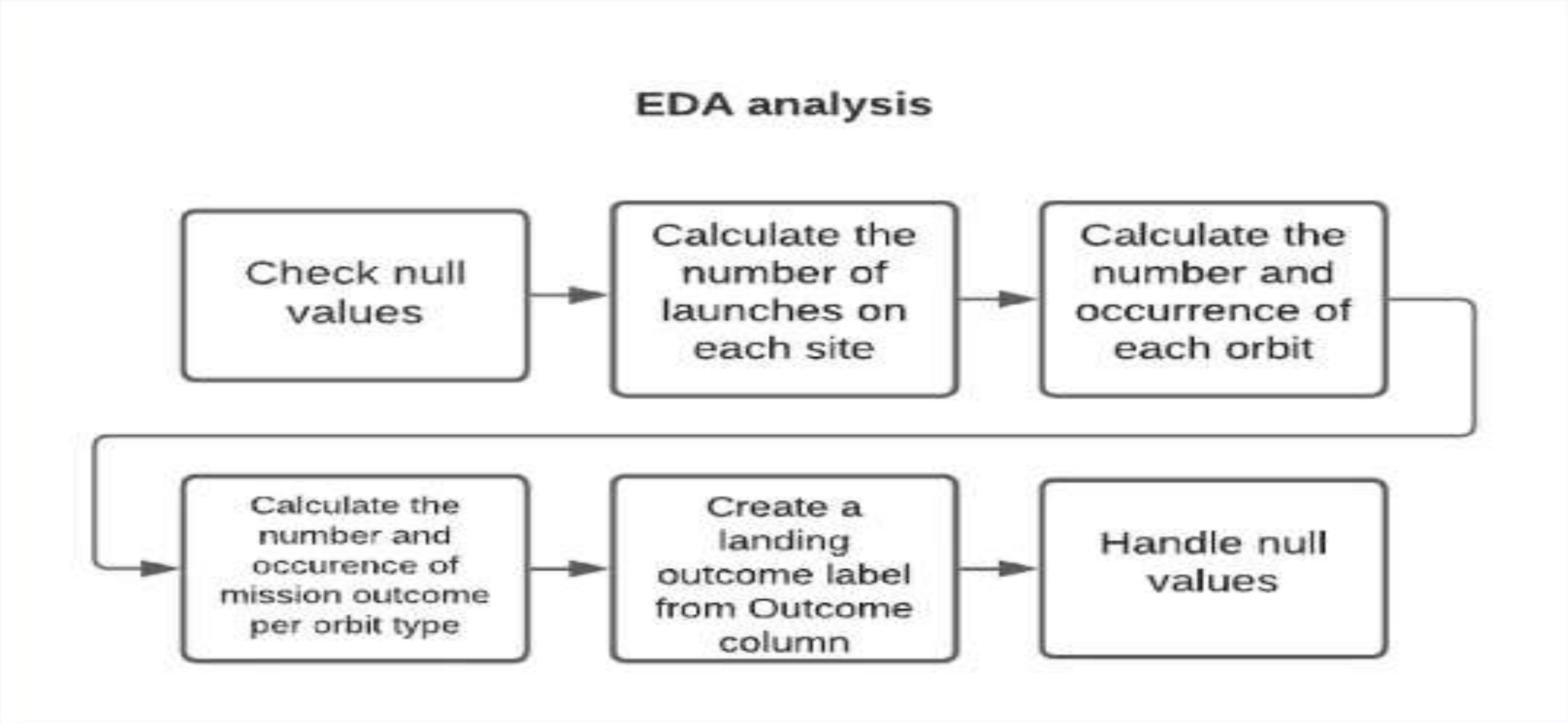
- Data collection with SpaceX Rest calls

```
[8]: spacex_url="https://api.spacexdata.com/v4/launches/past"
     response = requests.get(spacex_url)

[9]: static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API
     response.status_code

[9]: 200

[12]: # Use json_normalize meethod to convert the json result into a dataframe
      jlist = requests.get(static_json_url).json()
      df2 = pd.json_normalize(jlist)
      df2.head()
```

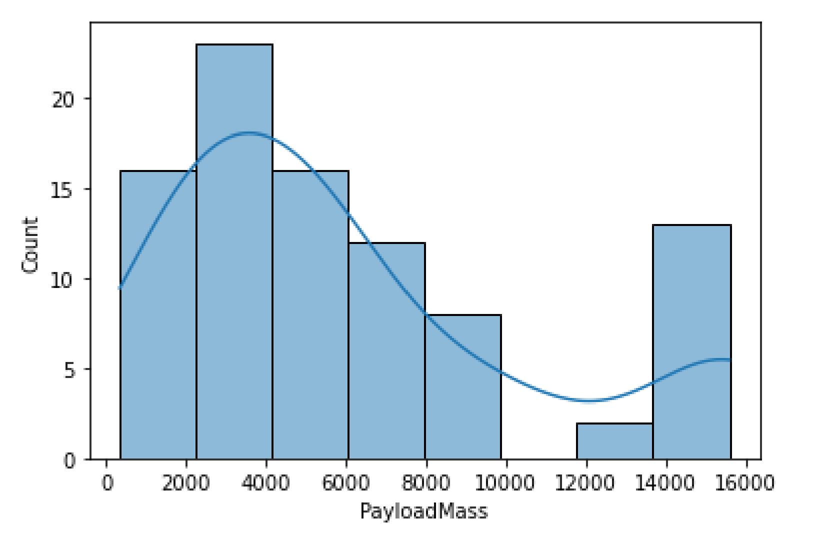
[12]:

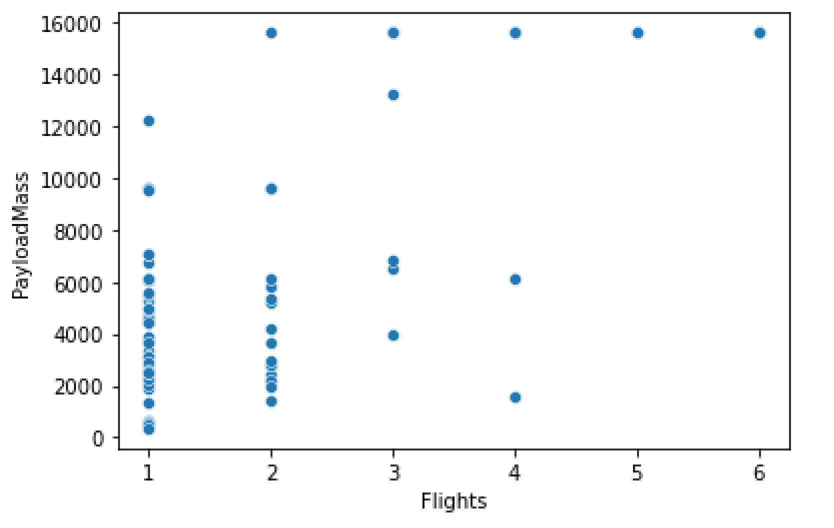
	static_fire_date_utc	static_fire_date_unix	tbd	net	window	rocket	success	details	crew	ships	capsules	
0	2006-03-17T00:00:00.000Z	1.142554e+09	False	False	0.0	5e9d0d95eda69955f709d1eb	False	Engine failure at 33 seconds and loss of vehicle	[]	[]	[]	[5eb0e4b

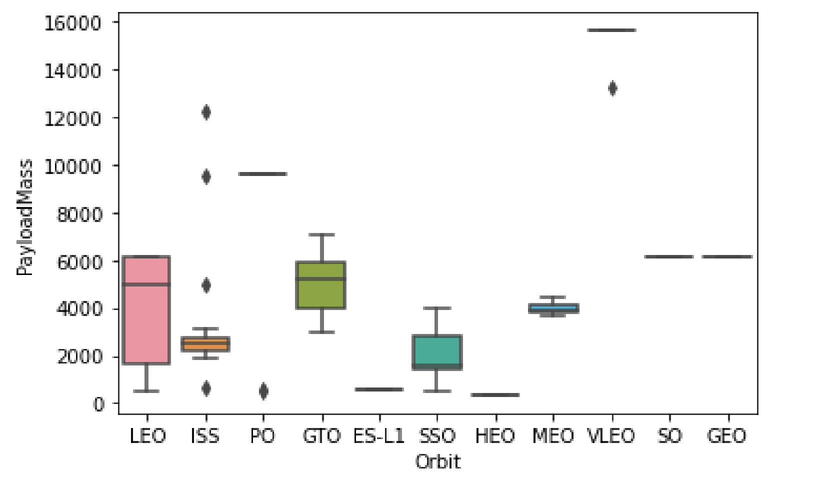
Data Wrangling

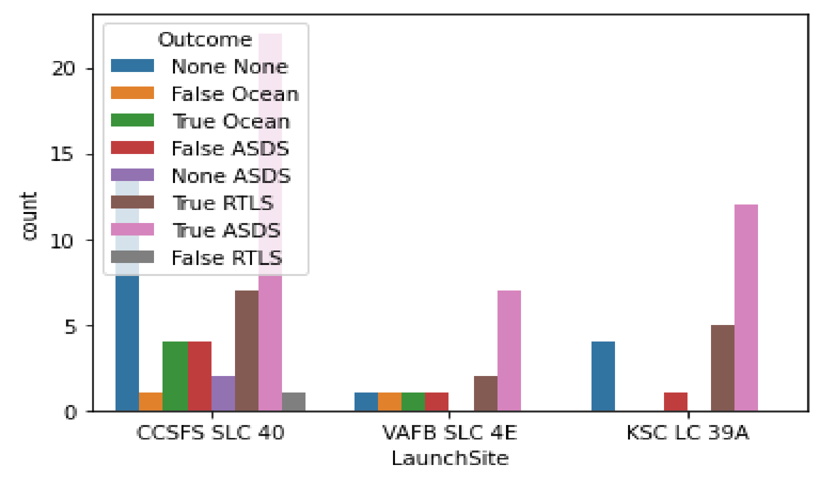


EDA with Data Visualization





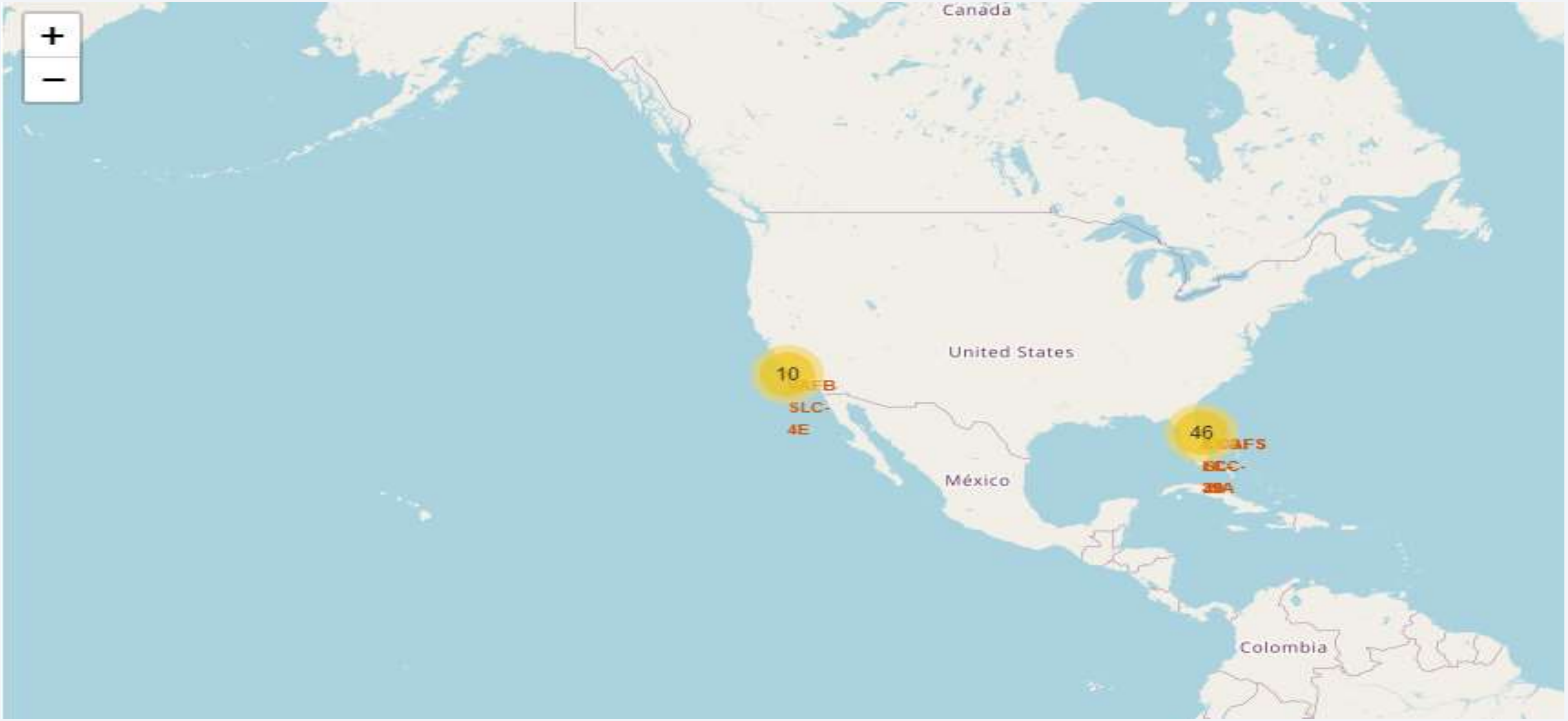




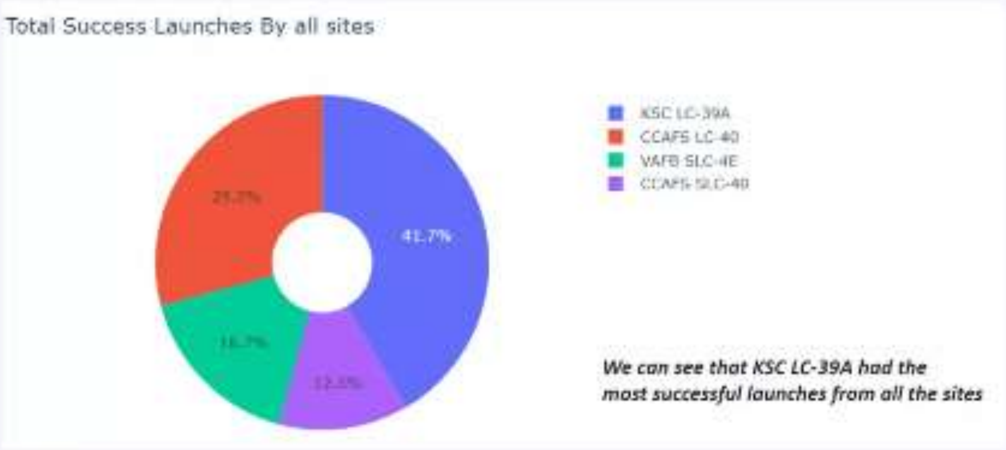
EDA with SQL

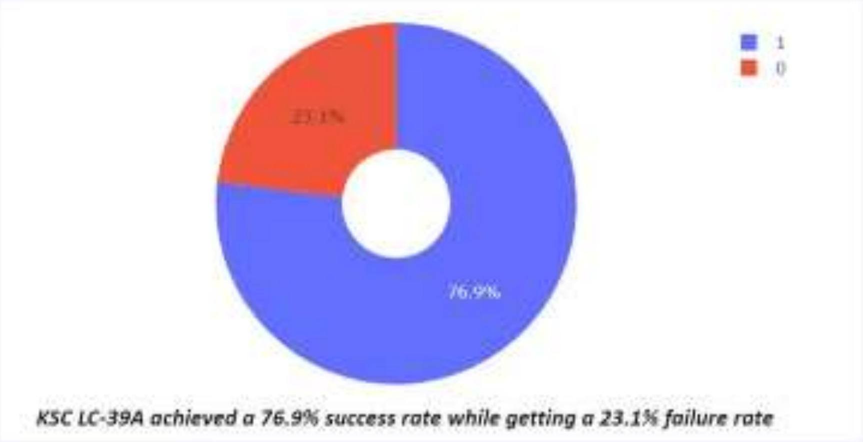
- SQL queries performed include:
  - Displaying the names of the unique launch sites in the space mission
  - Displaying 5 records where launch sites begin with the string 'KSC'
  - Displaying the total payload mass carried by boosters launched by NASA (CRS)
  - Displaying average payload mass carried by booster version F9 v1.1
  - Listing the date where the successful landing outcome in drone ship was achieved.
  - Listing the names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000
  - Listing the total number of successful and failure mission outcomes
  - Listing the names of the booster versions which have carried the maximum payload mass.
  - Listing the records which will display the month names, successful landing outcomes in ground pad, booster
  - versions, launch site for the months in year 2017
  - Ranking the count of successful landing outcomes between the date 2010 06 04 and 2017 03 20 in descending order.

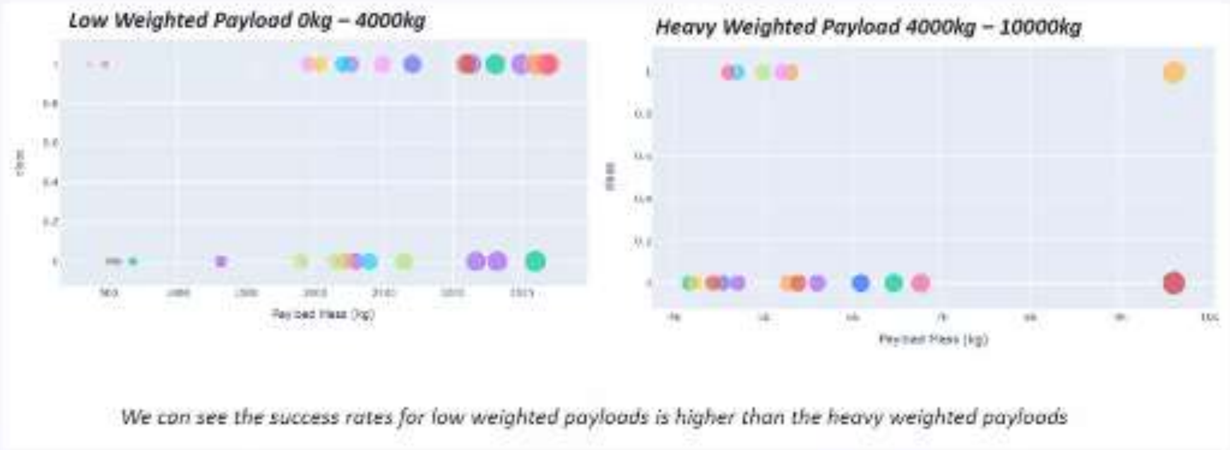
Build an Interactive Map with Folium



Build a Dashboard with Plotly Dash

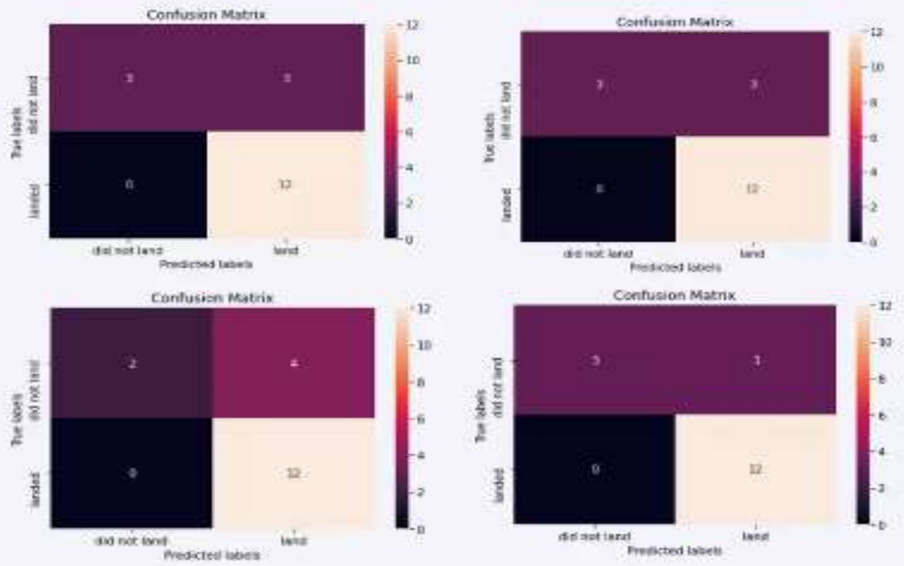






Predictive Analysis (Classification)

- The SVM, KNN, and Logistic Regression model achieved the highest accuracy at 83.3%, while the SVM performs the best in terms of Area Under the Curve at 0.958.



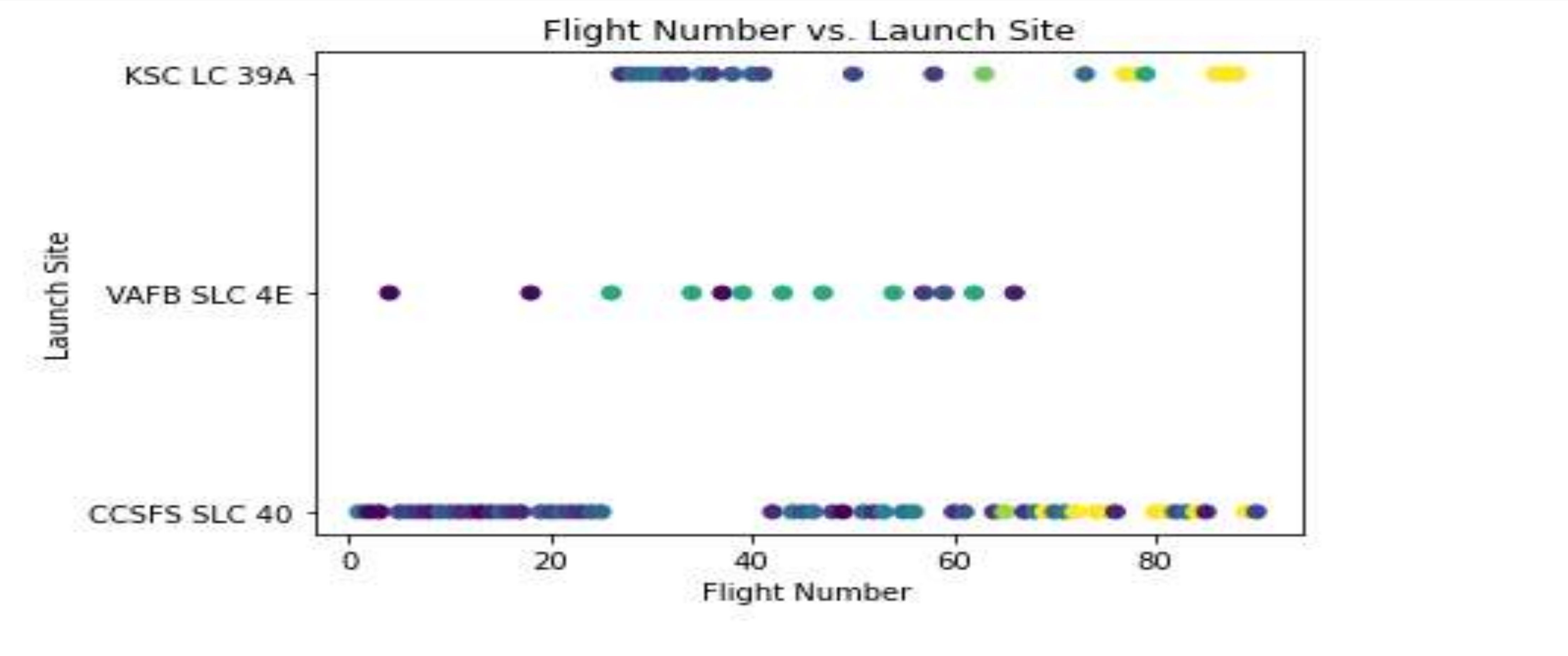
Results

- The SVM, KNN, and Logistic Regression models are the best in terms of
- prediction accuracy for this dataset.
- Low weighted payloads perform better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional time in years they
- will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Orbit GEO,HEO,SSO,ES L1 has the best Success Rate.

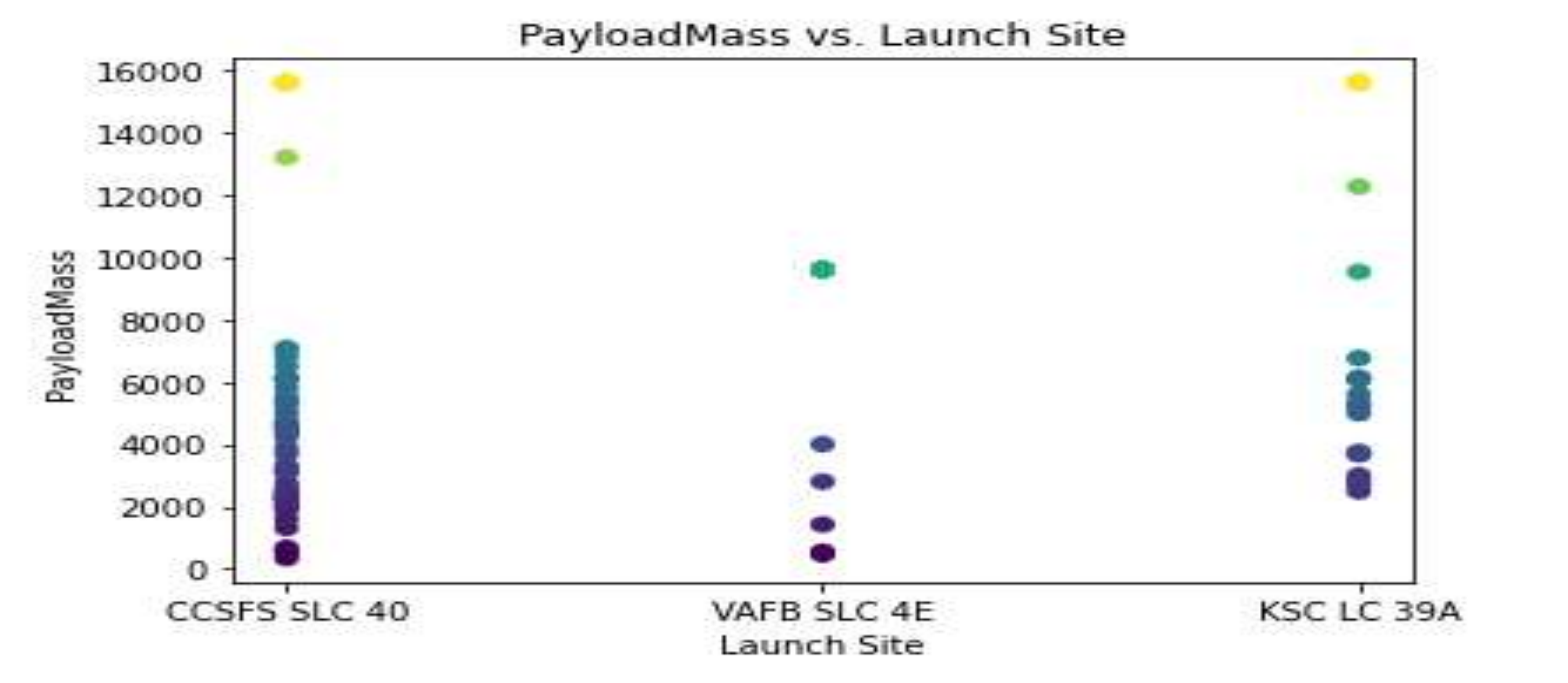
Section 2

Insights drawn from EDA

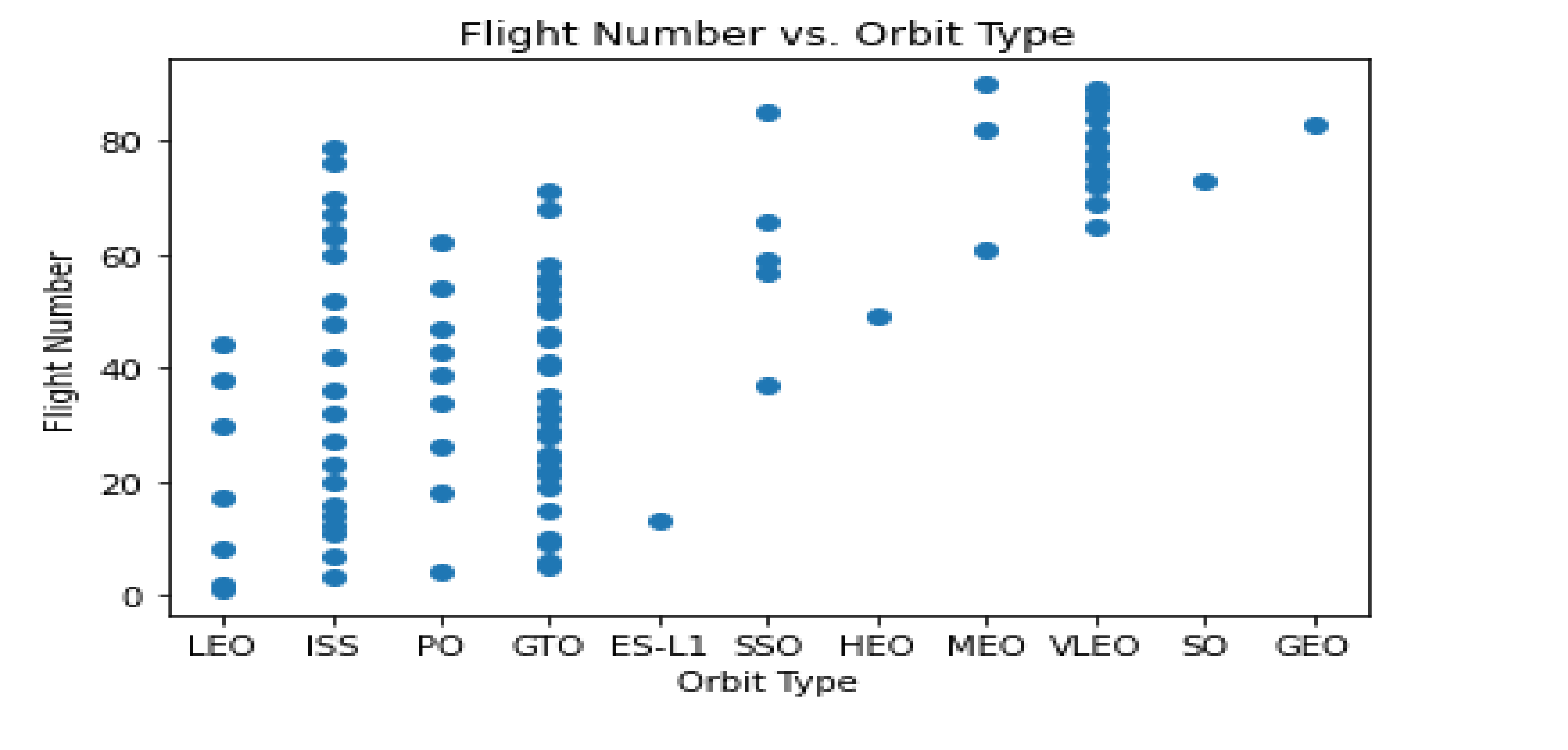
Flight Number vs. Launch Site



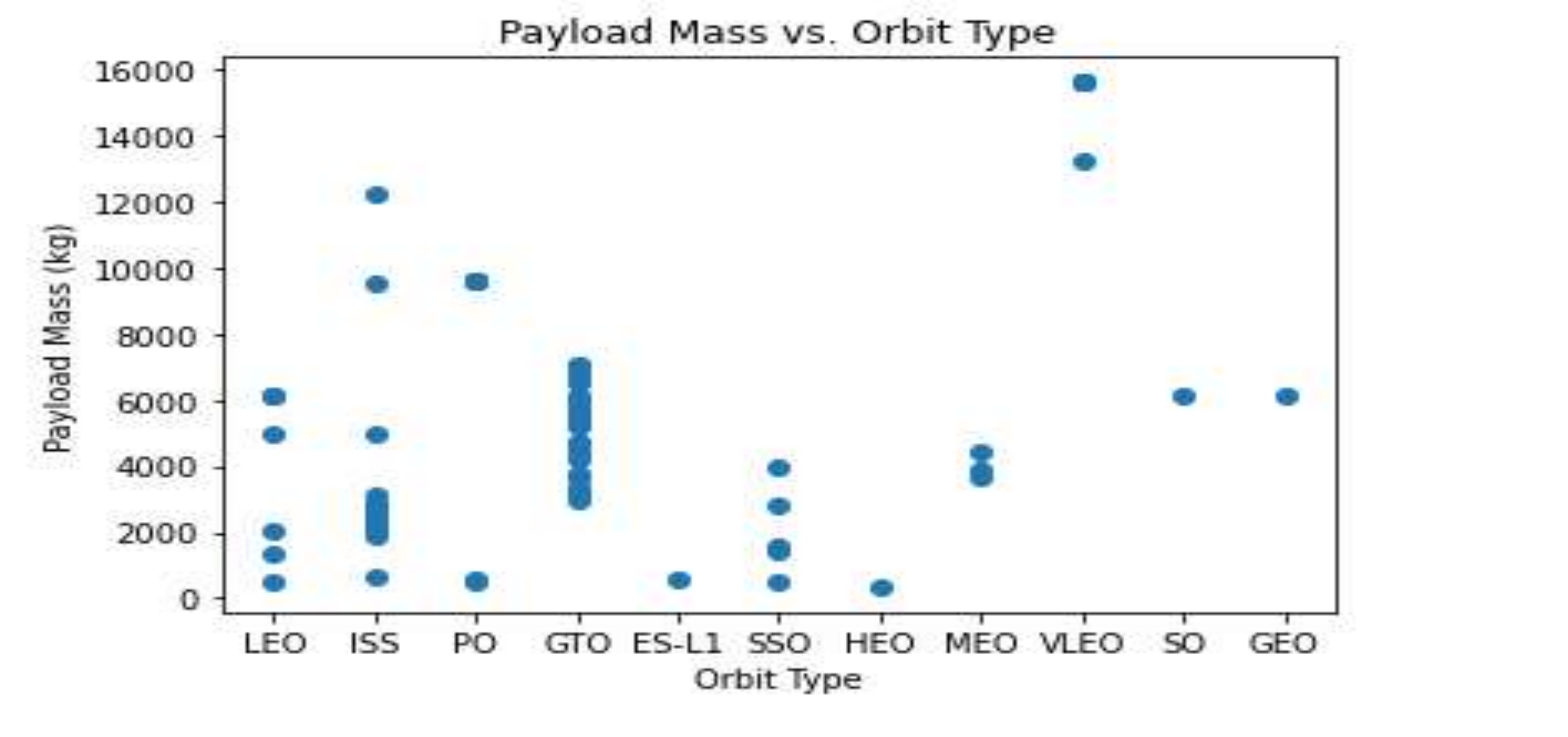
Payload vs. Launch Site



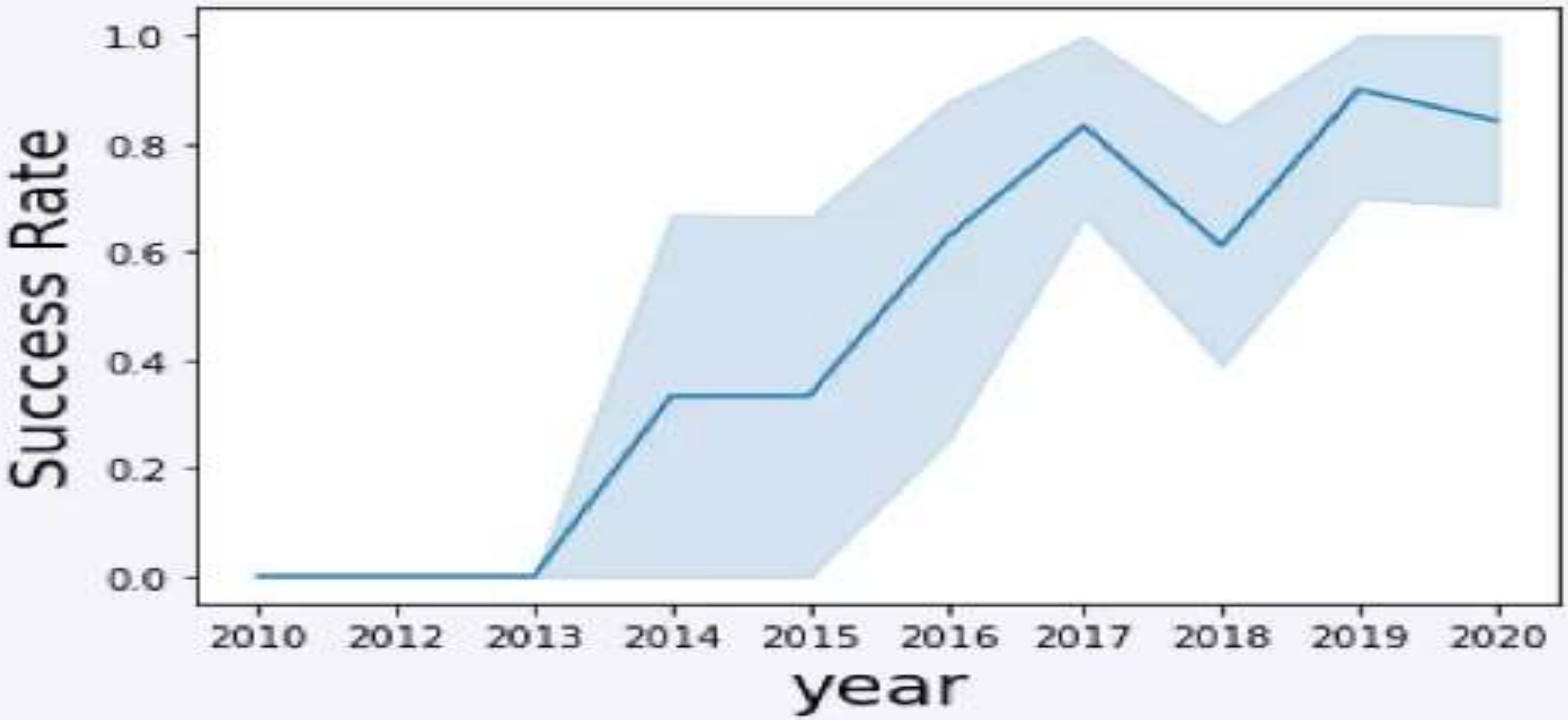
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

- %sql select distinct (LaunchSite) from spaceX

```
[54]: unique_launch_sites = df_spaceX['LaunchSite'].unique()
      print(unique_launch_sites)

      ['CCSFS SLC 40' 'VAFB SLC 4E' 'KSC LC 39A']
```

Launch Site Names Begin with 'CC'

- %sql select * from spaceX where LaunchSite 'CC%' limit 5

]:

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	Reused
0	1	2010-06-04	Falcon 9	6123.547647	LEO	CCSFS SLC 40	None None	1	False	False	False	NaN	1.0	
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCSFS SLC 40	None None	1	False	False	False	NaN	1.0	
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCSFS SLC 40	None None	1	False	False	False	NaN	1.0	
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCSFS SLC 40	None None	1	False	False	False	NaN	1.0	
5	6	2014-01-06	Falcon 9	3325.000000	GTO	CCSFS SLC 40	None None	1	False	False	False	NaN	1.0	

Total Payload Mass

- %sql select sum(PayloadMass) from spaceX where customer = ‘NASA’

```
Total payload carried by boosters from NASA: 168179.10 kg
```

Average Payload Mass by F9 v1.1

- %sql select avg(PayloadMass) from spaceX where BOOSTER_VERSION=‘F9 v1.1’

```
[61]: # Filter the dataframe to only include rows where BoosterVersion is F9 v1.1
df_f9_v1_1 = df_spaceX[df_spaceX['BoosterVersion'] == 'F9 v1.1']

# Calculate the mean of the PayloadMass column
avg_payload_mass = df_f9_v1_1['PayloadMass'].mean()

print("Average payload mass carried by F9 v1.1 boosters:", avg_payload_mass, "kg")

Average payload mass carried by F9 v1.1 boosters: nan kg
```

First Successful Ground Landing Date

- %sql select min(DATE) from spaceX where Landing_outcome = ‘Success(ground pad)’

Successful Drone Ship Landing with Payload between 4000 and 6000

- %sql select BOOSTER_VERSION from spaceX where Landing_Outcome = ‘Success(drone ship)’ and PayloadMass>4000 and PayloadMass < 6000

Total Number of Successful and Failure Mission Outcomes

- %sql select count(MISSION_OUTCOME) from spaceX where MISSION_Outcome = ‘Success’ or MISSION_Outcome = ‘Failure (in flight)’

Boosters Carried Maximum Payload

- %sql select BOOSTER_VERSION from spaceX where PayloadMass =(select max(PayloadMass) from spaceX)

2015 Launch Records

- %sql select * from spaceX where Landing_Outcome like ‘Success%’ and (DATE between ‘2015-01-01’ and ‘2015-12-31’) order by date desc

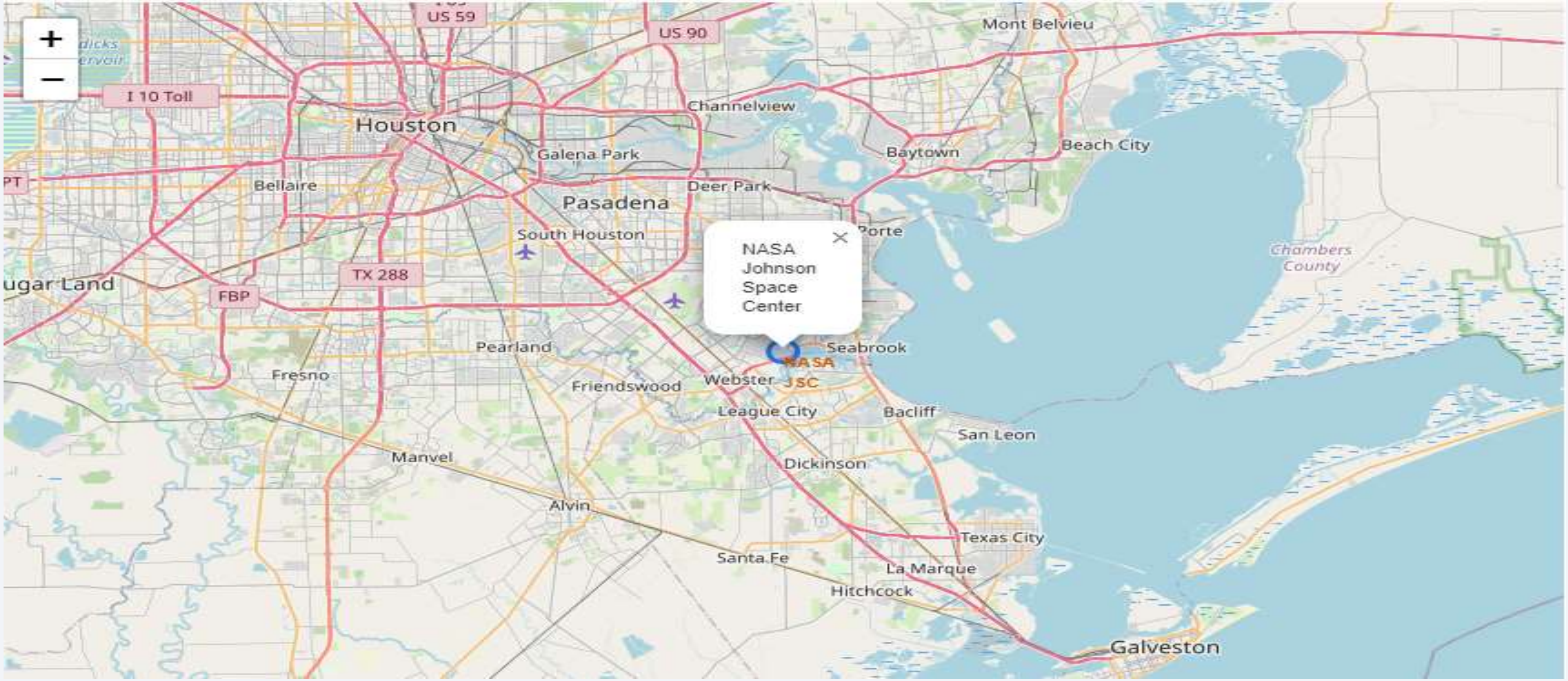
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- %sql select * from spaceX where Landing_Outcome like ‘Success%’ and (DATE between ‘2010-06-04’ and ‘2017-03-20’) order by date desc

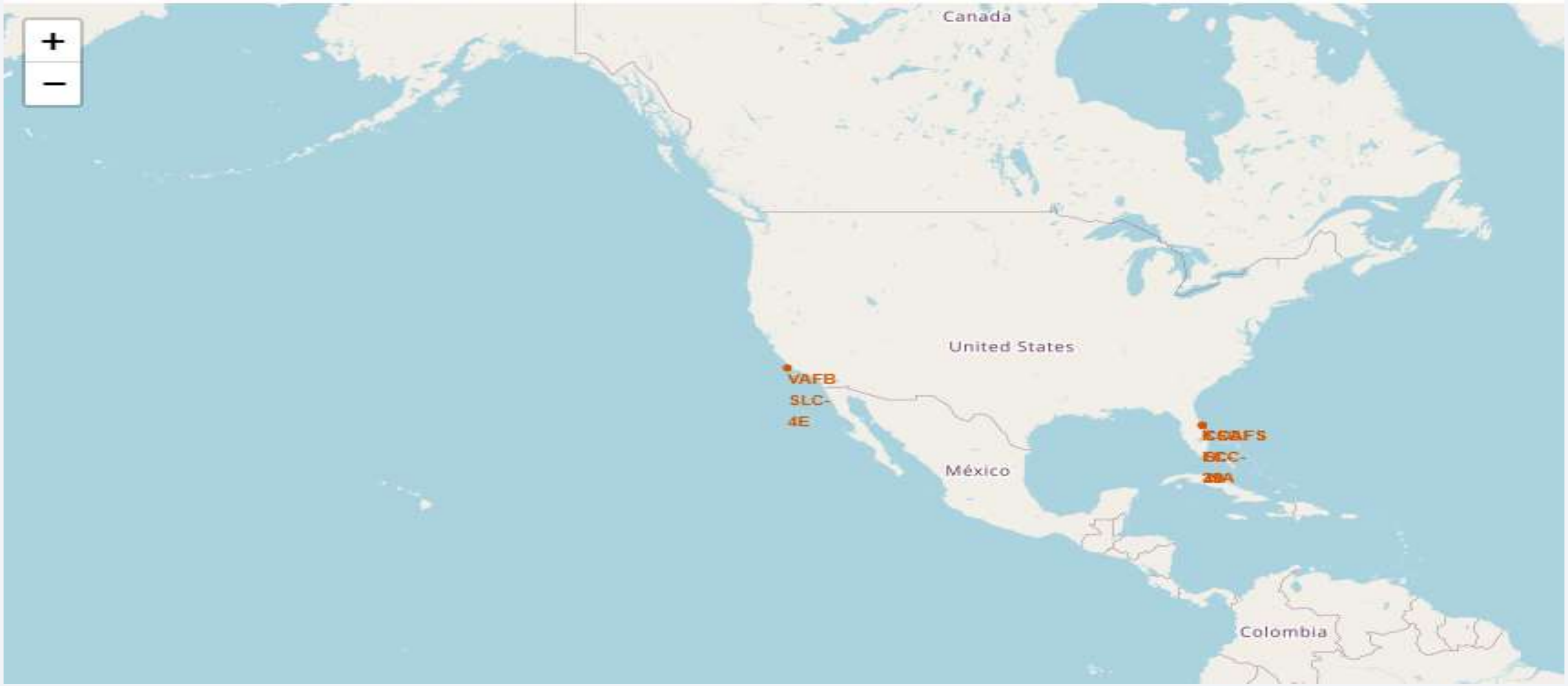
Section 3

Launch Sites Proximities Analysis

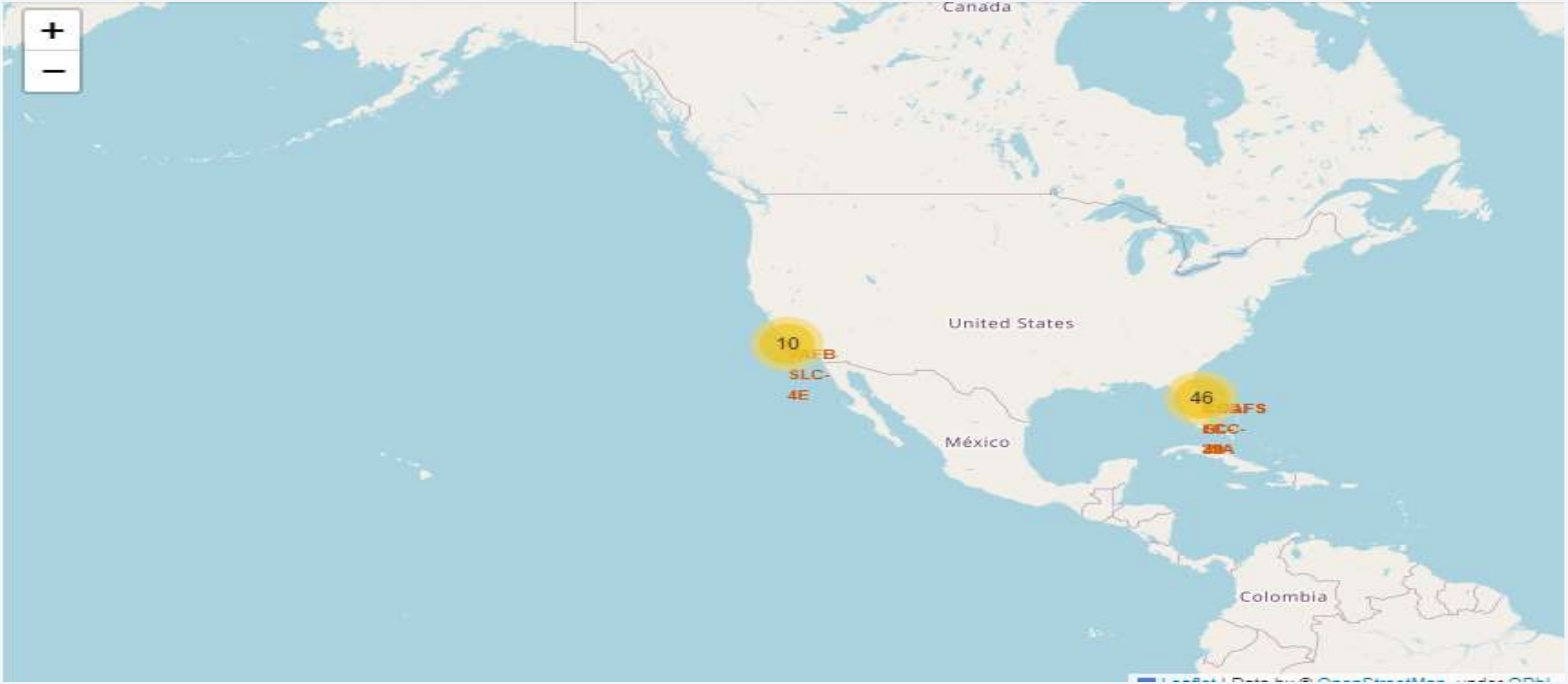
All launch sites marked on a map



Success / failed launches marked on the map



Distance between a launch site to its proximities

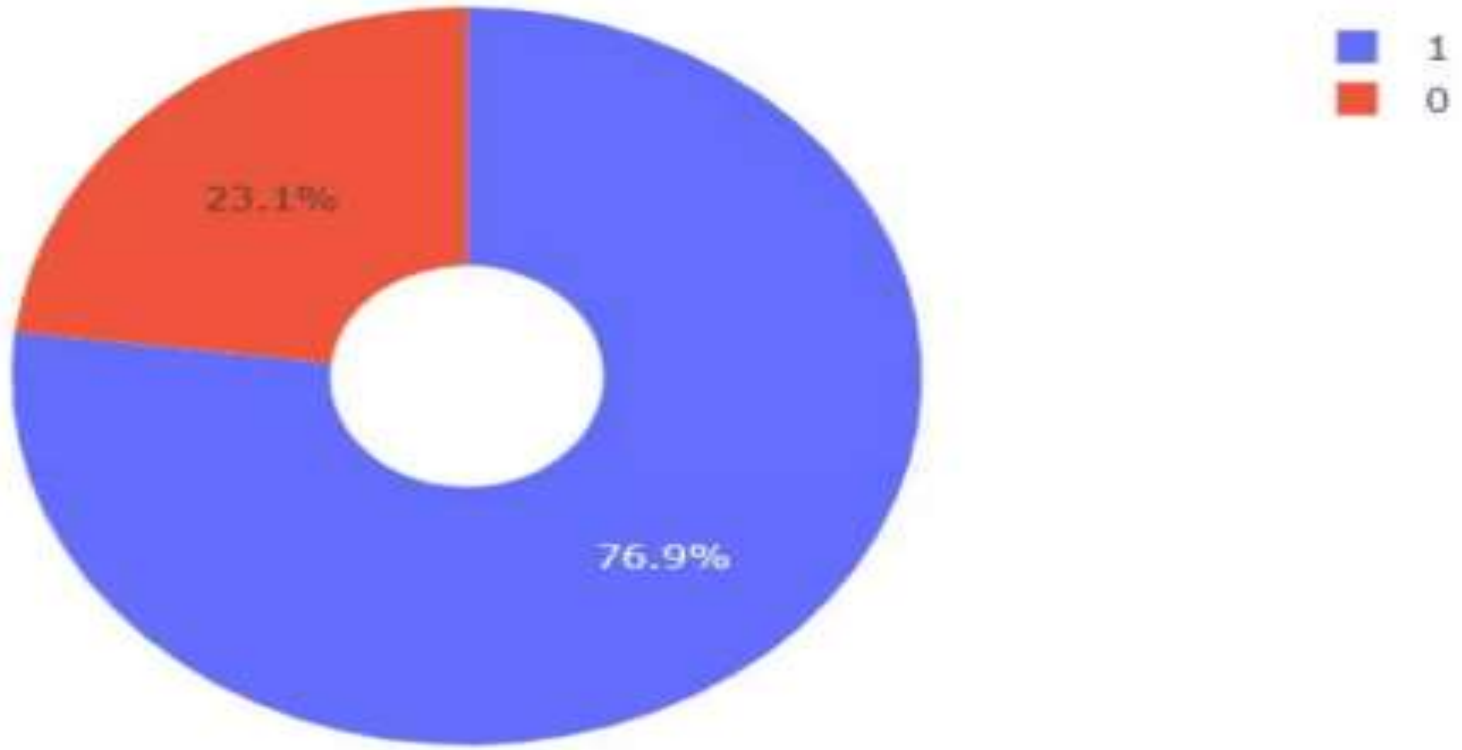


Section 4

Build a Dashboard with Plotly Dash

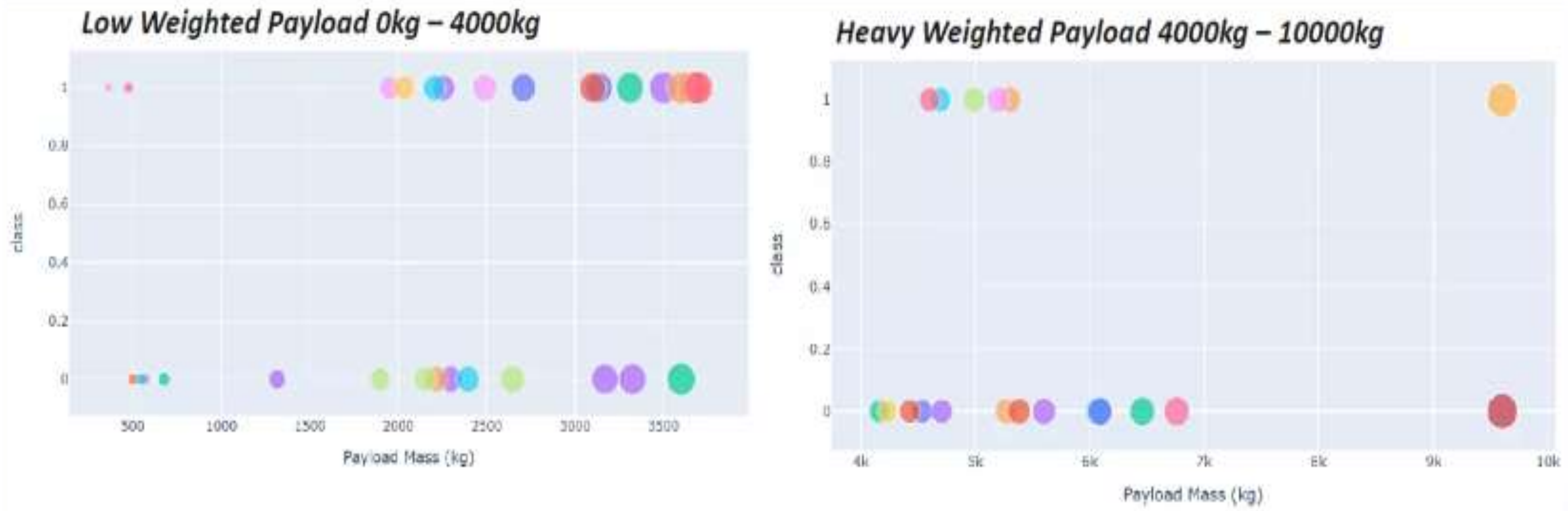
Total Success Launch by all sites

Success Rate by site



KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Payload vs Launch Outcome

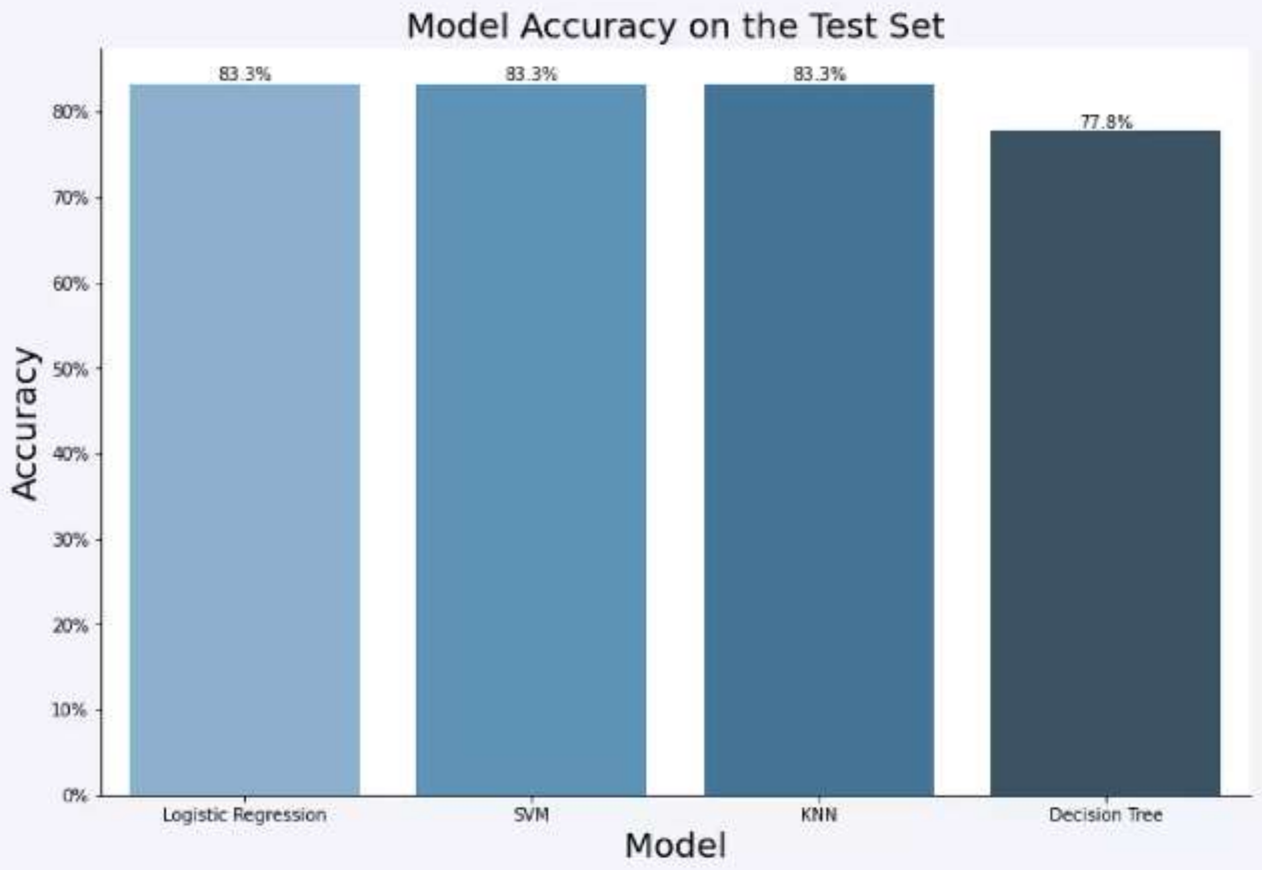


We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

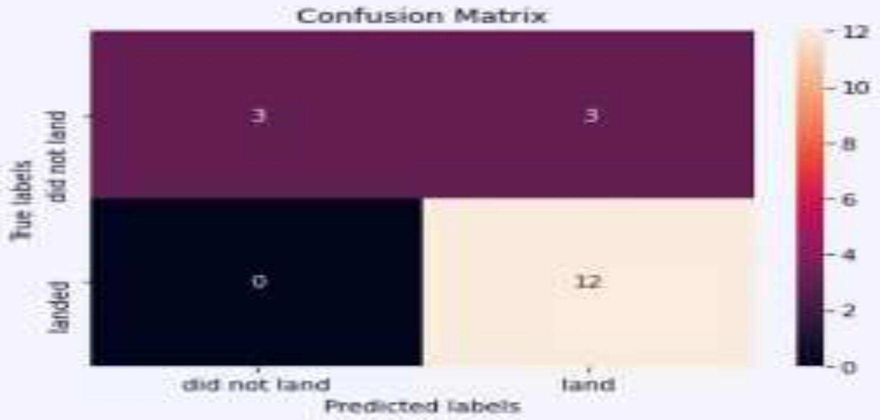
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.
- Low weighted payloads perform better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional time in years they will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Orbit GEO,HEO,SSO,ES L1 has the best Success Rate.

Thank you!