IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sandeep Kumar Palo
03/07/24

Outline

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

Executive Summary

•**Objective**: Predict the probability of SpaceX Falcon 9 rocket boosters successfully landing back at the launch pad.
•**Background**: SpaceX revolutionizes space travel with reusable rocket technology, significantly reducing launch costs.

•**Methodology**:
•Data collection via APIs and web scraping.
•Data wrangling to clean and preprocess launch data.
•Exploratory Data Analysis (EDA) using visualization tools and SQL queries.
•Interactive visualization with Folium and Plotly Dash to analyze launch site dynamics.
•Predictive analysis using machine learning models to forecast landing outcomes.
•**Predictors**: Factors include launch site attributes, payload orbit, mass, booster version, and landing pad location.
•**Benefit**: Enhance mission success rates, optimize operational strategies, and further reduce launch costs for commercial space missions.

Introduction

SpaceX, renowned for its pioneering efforts in aerospace, has significantly lowered satellite launch costs by reusing Falcon 9 rocket boosters. This project focuses on leveraging historical data from Falcon 9 launches to predict the probability of booster landing success. Key factors influencing this prediction include launch site characteristics, payload details, booster version, and landing pad location. By applying machine learning models and interactive visualizations with tools like Folium and Plotly Dash, we aim to enhance predictive accuracy. Insights gained will aid SpaceX in optimizing mission planning, improving operational efficiency, and continuing to disrupt the space launch market with cost-effective innovations.

Section 1

Methodology

Methodology

Data Collection Methodology

Utilized API and web scraping techniques for data retrieval.

Data Wrangling

Extracted, loaded, and transformed data to prepare for analysis.

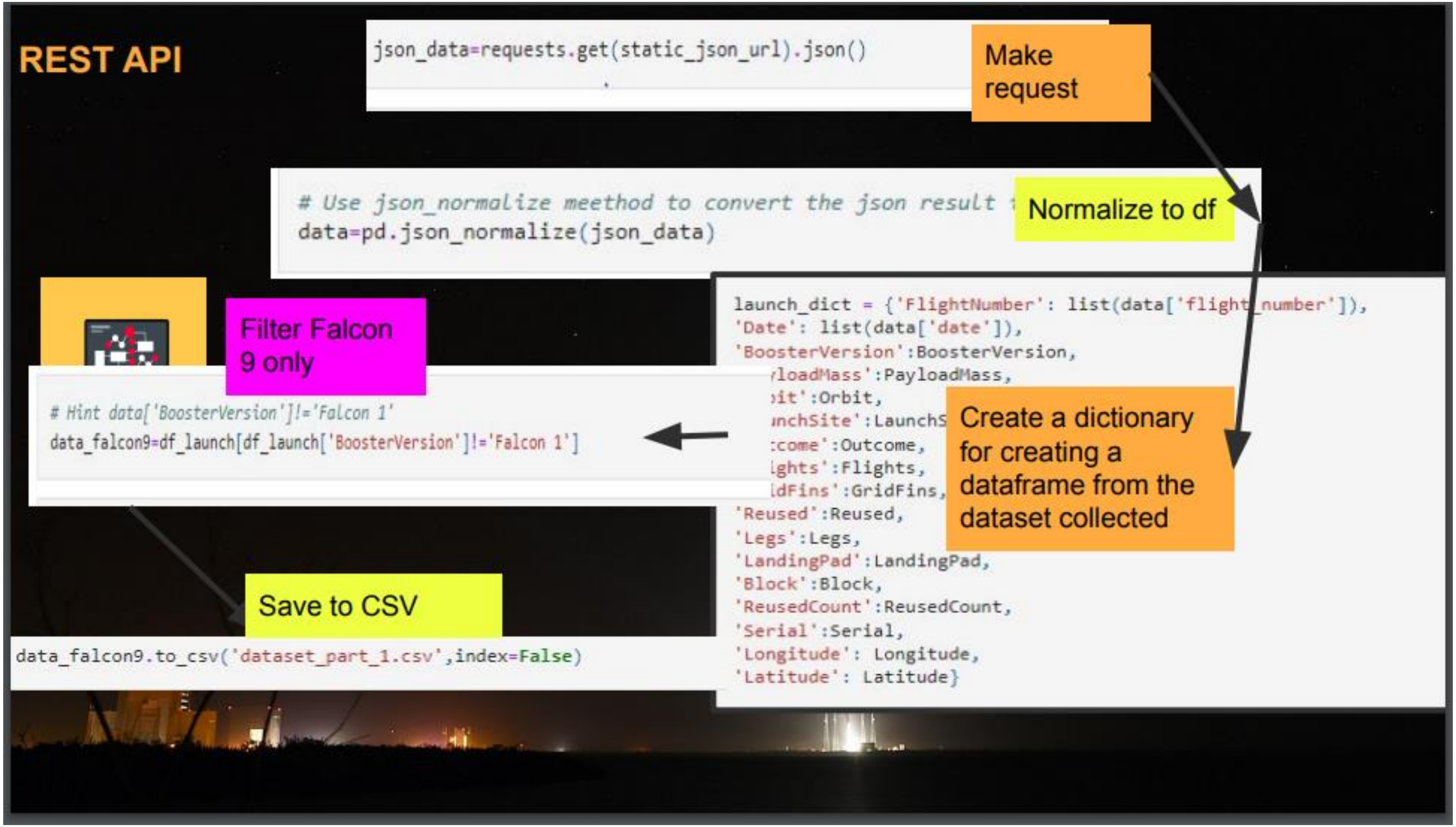
Exploratory Data Analysis (EDA)

Employed SQL for data exploration and visualization.

Interactive Visual Analytics

Implemented Folium for geospatial data visualization.
Developed interactive dashboards using Plotly Dash.

Predictive Analysis:Utilized classification models to predict booster landing success.
Implemented model building, tuning, and evaluation processes.


REST API
json_data=requests.get(static_json_url).json()
Make request
# Use json_normalize meethod to convert the json result
data=pd.json_normalize(json_data)
Normalize to df
launch_dict = {'FlightNumber': list(data['flight_number']),
'Date': list(data['date']),
'BoosterVersion':BoosterVersion,
loadMass':PayloadMass,
it':Orbit,
nchSite':LaunchS
come':Outcome,
ghts':Flights,
dFins':GridFins,
'Reused':Reused,
'Legs':Legs,
'LandingPad':LandingPad,
'Block':Block,
'ReusedCount':ReusedCount,
'Serial':Serial,
'Longitude': Longitude,
'Latitude': Latitude}
Create a dictionary for creating a dataframe from the dataset collected
Filter Falcon 9 only
# Hint data['BoosterVersion']!='Falcon 1'
data_falcon9=df_launch[df_launch['BoosterVersion']!='Falcon 1']
Save to CSV
data_falcon9.to_csv('dataset_part_1.csv',index=False)

```
# use requests.get() method with the provided static_url
# assign the response to a object
response=requests.get(static_url)
```

Get content of the wiki

Loop Through and add column names

```
column_names = []
temp = soup.find_all('th')
for x in range(len(temp)):
    try:

        name = extract_column_from_header(temp[x])
        if (name is not None and len(name) > 0):
            column_names.append(name)
    except:
        pass
```

Create a Dataframe for the important columns

Loop through the request content and extract data

```
df.to_csv('spacex_web_scraped.csv', index=False)
```

Save to cvs

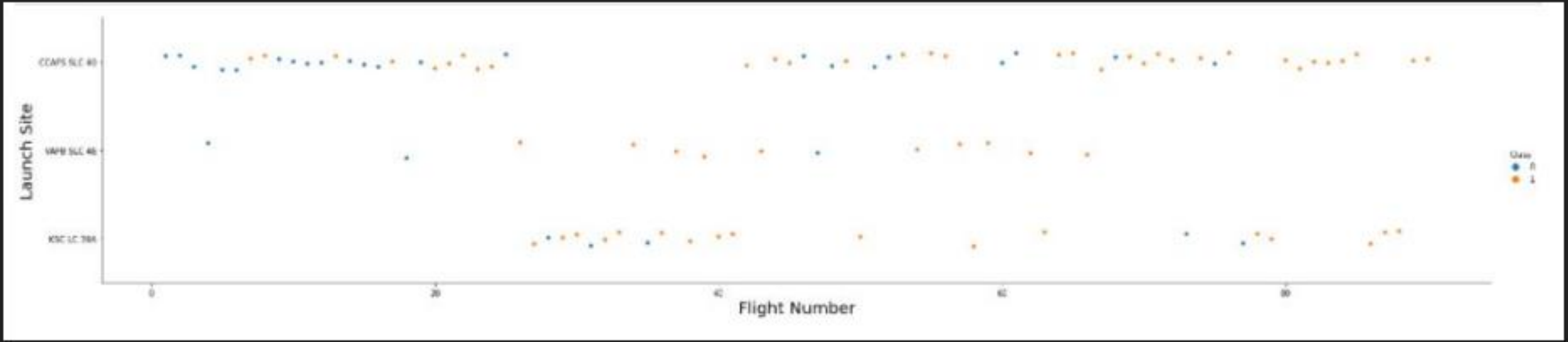
DATA Wrangling

- In the data set, there are several different cases where the booster did not land successfully. Sometimes a landing was attempted but failed due to an accident; for example, True Ocean means the mission outcome was successfully landed to a specific region of the ocean while False Ocean means the mission outcome was unsuccessfully landed to a specific region of the ocean. True RTLS means the mission outcome was successfully landed to a ground pad False RTLS means the mission outcome was unsuccessfully landed to a ground pad.True ASDS means the mission outcome was successfully landed on a drone ship False ASDS means the mission outcome was unsuccessfully landed on a drone ship.

Exploratory Data Analysis

- Through EDA on the data from APi and Wiki,we will find some insights on :
- Flight number & Launch Sites-Visualizing the launch from every site
- Payload & Launch Sites-Payload launch from sites
- Success rate & Orbit type-Success rate compared to the orbit type
- Flight number & Orbit Type -Type of orbit for each launch
- Payload & Orbit type -Payload and the orbit .
- Trend of success rate-Trend of the success rate over the years

Flight number & Launch Sites



From the Visualization we can concluded that:

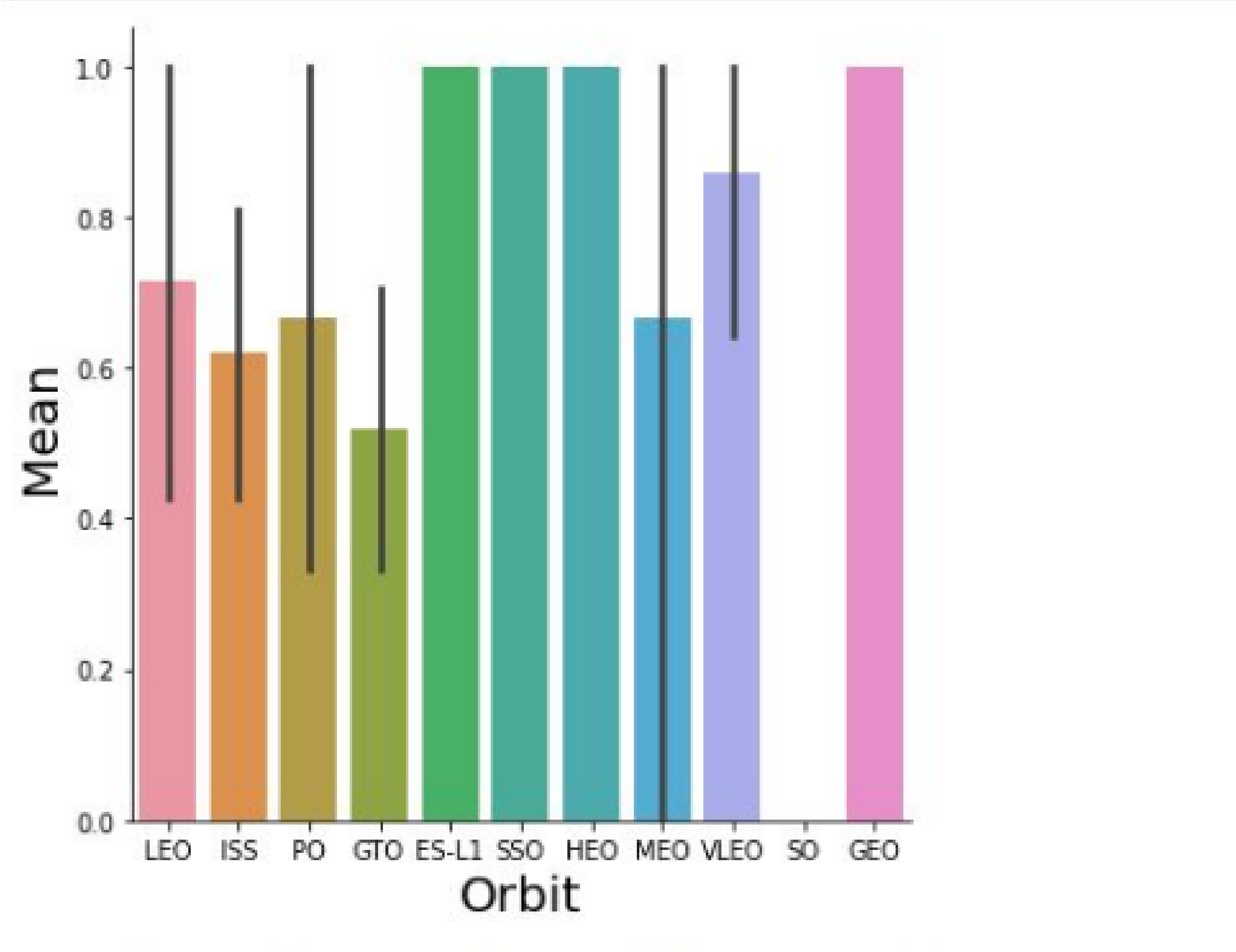
- Earlier flights lauch were from CCAFS-SLC-40 site ,Followed by KSC-LC-39A
- Most Launches are Launched from CCAFS-SLC-40
- Fewer Launches from **VAFB SLC 4E site**

Payload & Launch Sites

From the Visualization we can concluded that:

- VAFB SLC 4E has Low Payload launches
- CCAFS SLC 40 has more Higher Payload Launches and Low Payload Lauches .

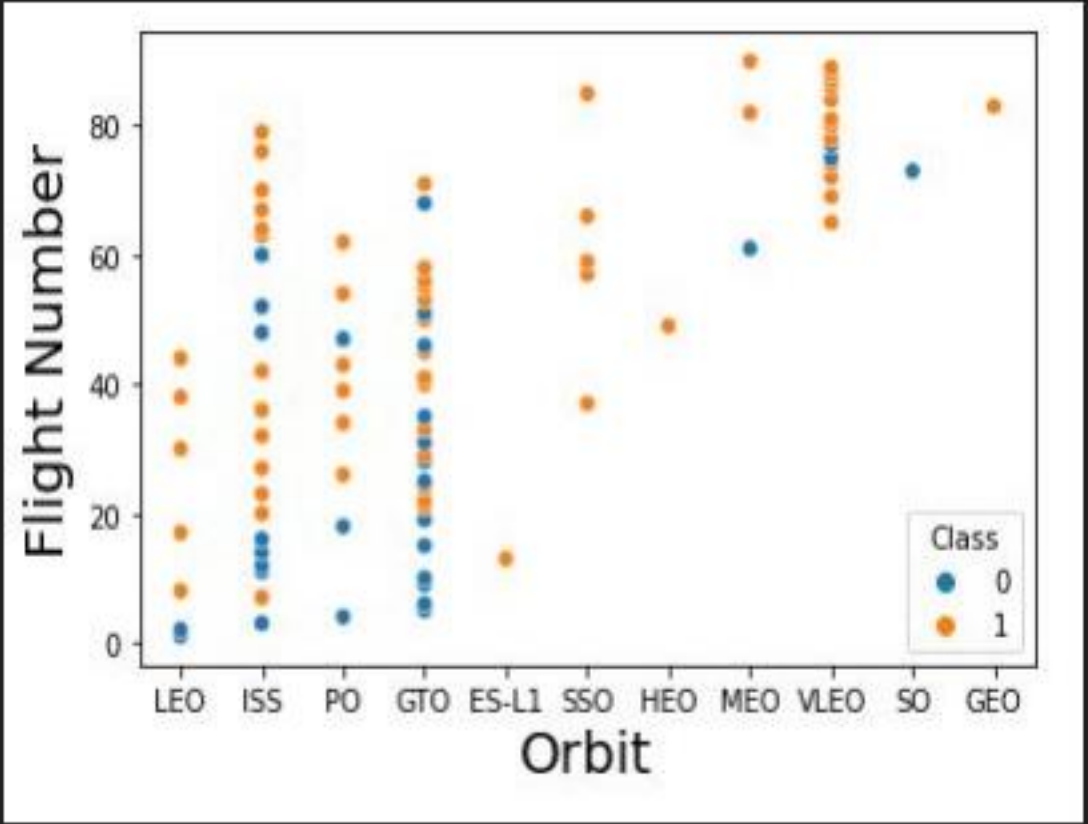
Orbit Success



From the Visualization we can concluded that:

- GEO,HEO & ES-L1,SS) have high success rate .

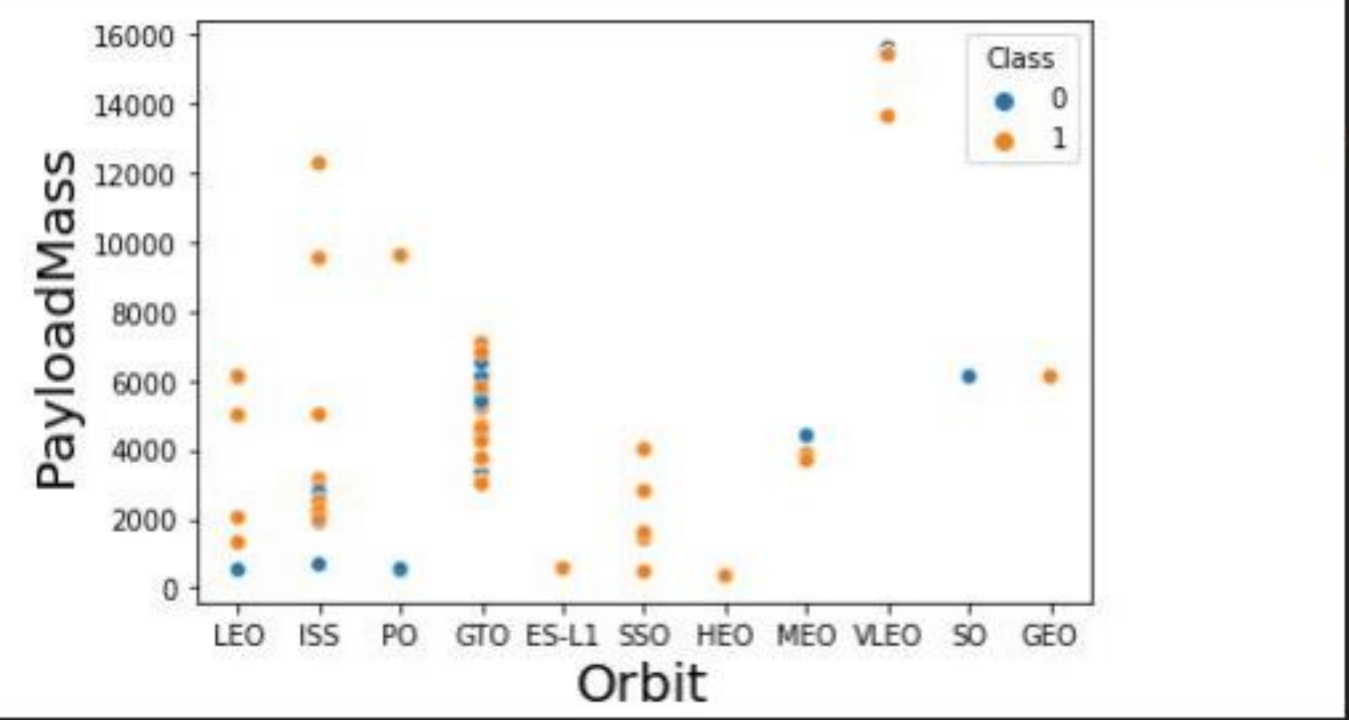
Flight No & Orbit



From the Visualization we can concluded that:

- Most Flight are to ISS,PO,GTO and VLEO
- MOST fails are for ISS,GTO
- SSO & VLEO has high success rate .

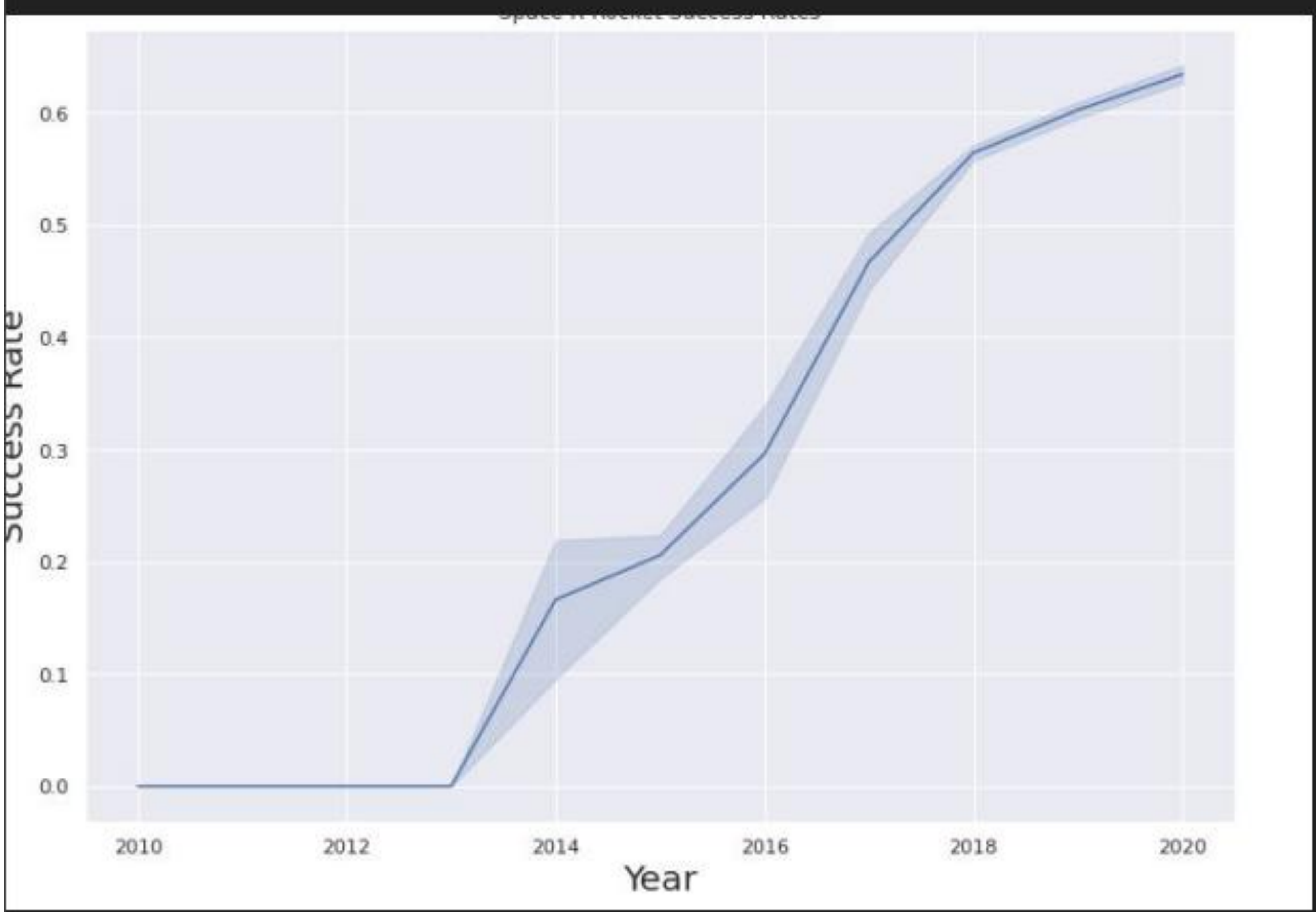
Payload & Orbit



From the Visualization we can concluded that:

- Higher Payload are to the VLEO
- Least Payload are for HEO,ISS,PO,ES-L1
- GTO has average payload size .

Success Rate Trend



The rate of success of the launches increase over time since to the data collected from the previous fails and success launches .

EDA with SQL

- Exploratory Data Analysis on the follow criteria:
- Unique Sites
- Max Payload
- Average Payload
- Day when First Success Landing
- Success and Failures count
- Boosters With Max Payload

EDA with SQL

- For the categories above we find that :
- Sites that SpaceX operates in are:

  CCAFS LC-40,CCAFS SLC-40,KSC LC-39A,VAFB SLC-4E
- Max Payload:48213
- Average Payload for all Launches: 2928 kgs
- First Success Landing was Made on:06/05/2016
- Booster Version that carry over 4000 kg and 6000 Kg : F9 FT B1020,F9 FT B1022,F9 FT B1026,F9 FT B1021.2,F9 FT B1031.2

INTERACTIVE MAP WITH FOLIUM

- Visualization of the launches for every site and every launch in a Interactive Map
- Visualization of:
- ❖ Launch Sites
- ❖ Visualize the launches on the map base on Fail or Success

Visualize the Launches on Map

Data Set Contained 3 Separate Launch Sites that are displayed on the picture on the left .

This gives us insights to the launches success and failures .

PREDICTIVE ANALYSIS

- Through Models,tuned for best performance we go the insights on the probability if a launch being success or a failure .

- Models used include:
- ❑ KNeighboursClassfier
- ❑ Decision Tree
- ❑ Logistic Regression
- ❑ Support Vector Machine

Best Model Prediction

After Analyzing all the Models,the KNN was the best Model with accuracy of 77% and best Score of 87%

```
parameters = {'n_neighbors': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
              'algorithm': ['auto', 'ball_tree', 'kd_tree', 'brute'],
              'p': [1,2]}

KNN = KNeighborsClassifier()
gscv=GridSearchCV(KNN,parameters,scoring="accuracy",cv=10)
KNN_cv=gscv.fit(X_train,y_train)
```

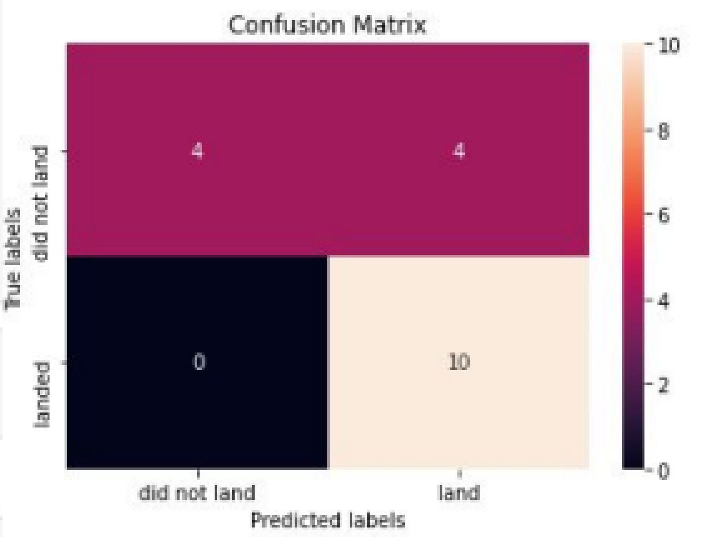
```
print("Accuracy",KNN_cv.score(X_test,y_test))
```

```
Accuracy 0.7777777777777778
```

```
print("tuned hpyerparameters :(best parameters) ",KNN_cv.best_params_)
print("accuracy :",KNN_cv.best_score_)
```

```
tuned hpyerparameters :(best parameters)  {'algorithm': 'auto', 'n_neighbors': 4, 'p': 1}
accuracy : 0.8767857142857143
```

```
yhat = KNN_cv.predict(X_test)
plot_confusion_matrix(y_test,yhat)
```



True Positives

INTERACTIVE WITH DASH

Visualization of the Launches from Site in Dashboard

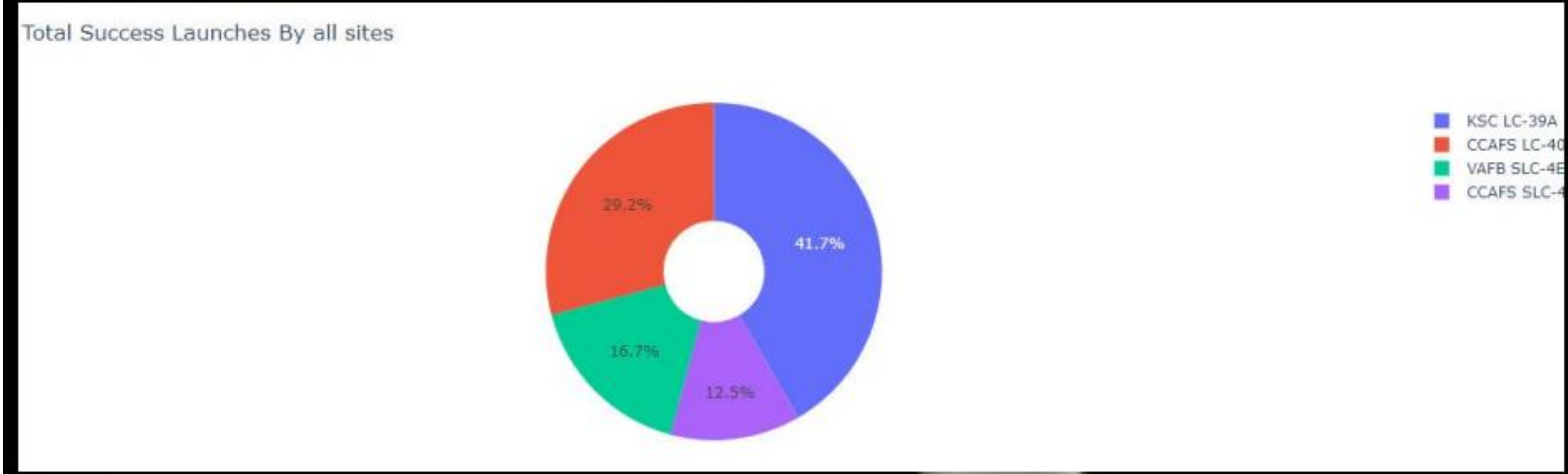
Visualization of:

- ❖ Success Launch Launch Sites
- ❖ Visualize payload from different sites with rangeSlider for interacting with the plot .

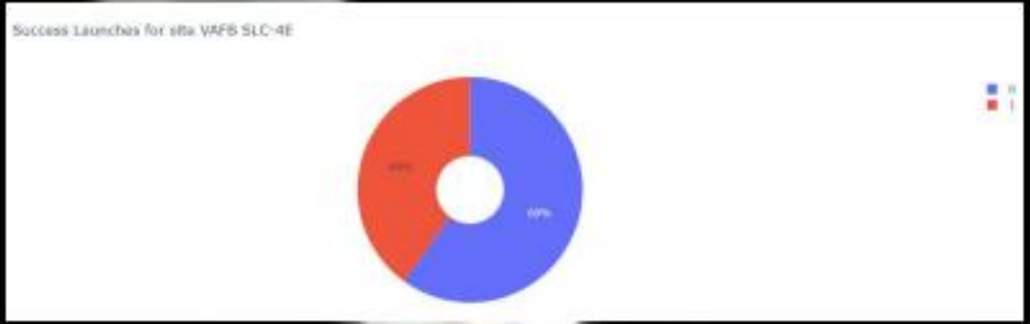
INTERACTIVE WITH DASH

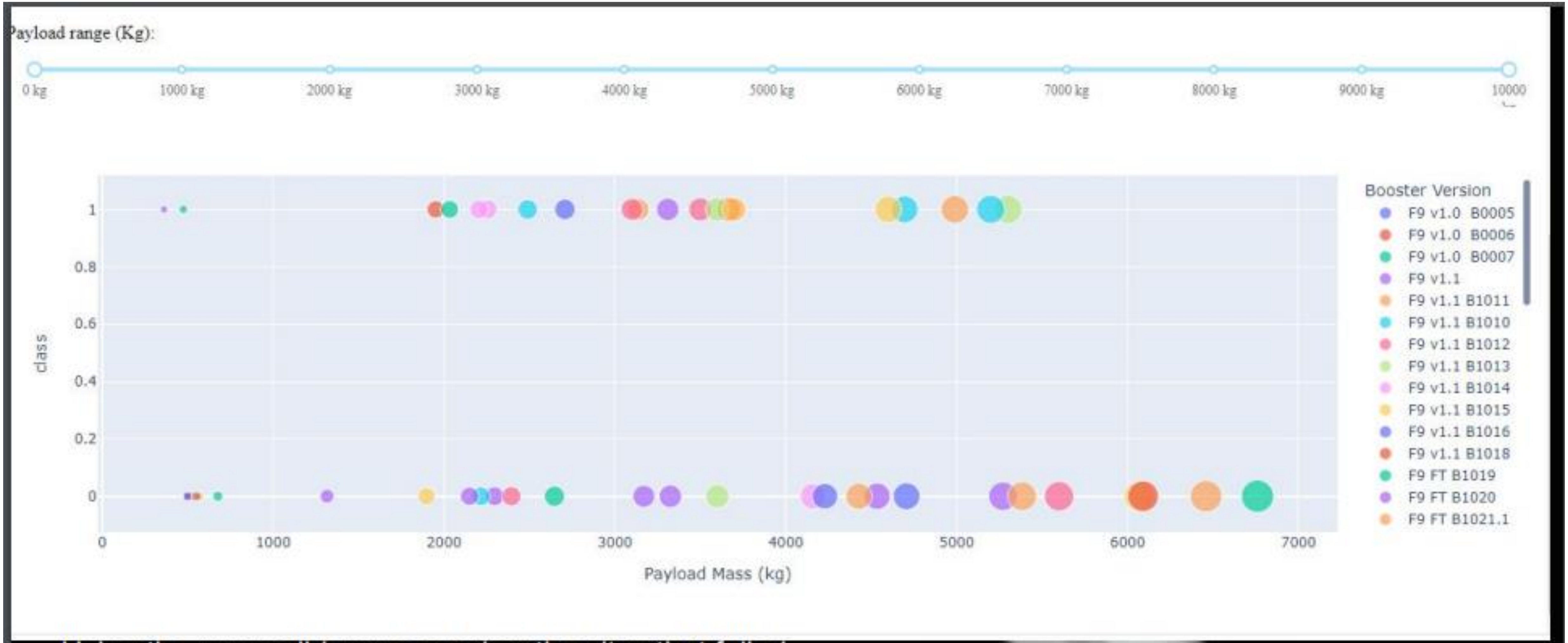
- Visualization of the Launches from Site in Dashboard

- Visualization of:
- ❖ Success Launch Launch Sites
- ❖ Visualize payload from different sites with rangeSlider for interacting with the plot .



Observation is that KSC has more launches compared to other sites .
Using the drop down on the dashboard it's possible to view single site launches





Using the range slider we can view the sites that failed and succeed for each booster version and the Payload they were carrying .

Conclusion

- In conclusion, this project demonstrated the effectiveness of leveraging data science techniques to predict the probability of Falcon 9 booster landing success for SpaceX. Through comprehensive data collection via API and web scraping, rigorous data wrangling, and insightful exploratory data analysis using SQL and visualization, we gained valuable insights into the factors influencing booster landings. Interactive visual analytics with Folium and Plotly Dash provided intuitive ways to explore spatial and operational dynamics. Finally, employing predictive analysis using classification models allowed us to build, tune, and evaluate models, contributing to enhanced decision-making in mission planning and operational optimization for SpaceX. This project underscores the power of data-driven approaches in the aerospace industry, paving the way for continued innovation and cost-effectiveness in space exploration.