# Falcon 9: Data Science Analysis

FRANCESCO COCCIRO

# Executive Summary

- This project focuses on predicting the landing success of the first stage of SpaceX's Falcon 9 rocket. By leveraging data analysis and Data Science models, the primary objective is to provide the competing startup with tools and insights to make more competitive bids for rocket launches. The outcomes of this project will enable the startup to make informed decisions based on the predictions of Falcon 9 landing success.
- The methodology involving data collection, data wrangling, exploratory data analysis, data visualization, model development, model evaluation, and reporting the results.

# Executive Summary

- Data Collection using SpaceX API and BS4
- Data visualization for data exploration
- Machine learning models to perform classification analysis
- Prediction models built off data collected via web scraping and SpaceX API
- Interactive dashboard creation used for exploratory data analysis

# Table of content

Welcome to the world of commercial space travel.

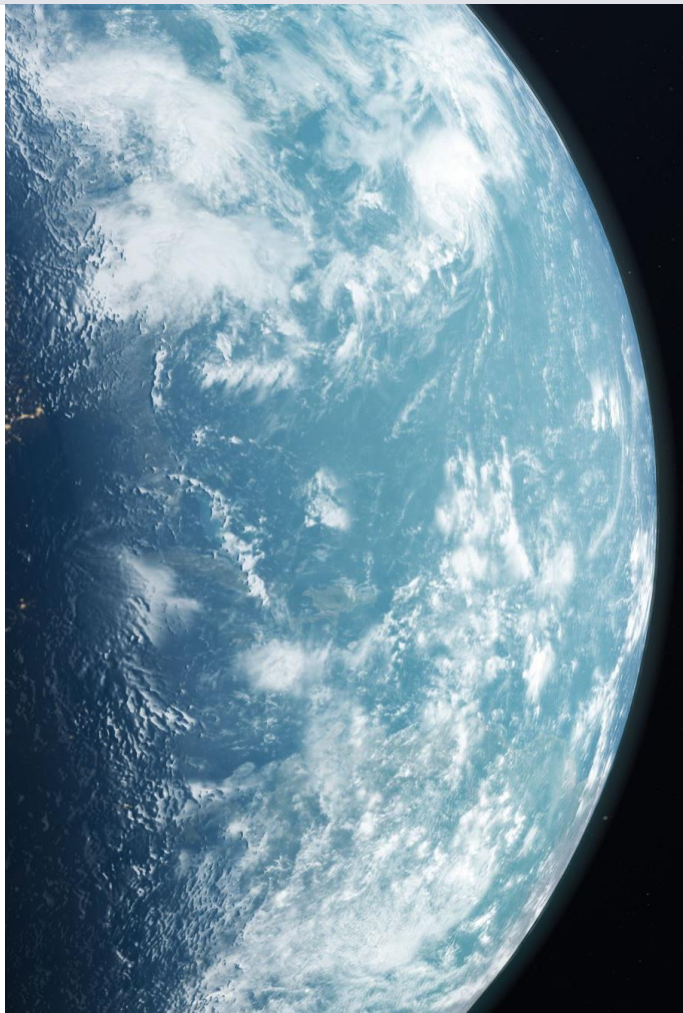
In an era where companies are revolutionizing space exploration and making it more accessible, SpaceX has emerged as a frontrunner in the industry. With remarkable achievements like sending spacecraft to the International Space Station, deploying the Starlink satellite internet constellation, and conducting manned missions to space, SpaceX has proven its capabilities.

One key factor behind SpaceX's success is the cost-effectiveness of its Falcon 9 rocket launches. While other providers charge significantly higher prices, SpaceX stands out by reusing the first stage of its rockets, resulting in substantial savings.

As data scientists working for the aspiring rocket company Space Y, our mission is to compete with SpaceX and establish ourselves in the industry.

Our task involves gathering information about SpaceX, creating insightful dashboards, and training machine learning models using public data.

Through this capstone project, we aim to forecast whether SpaceX will successfully reuse the first stage, ultimately determining the cost-effectiveness of our launches.



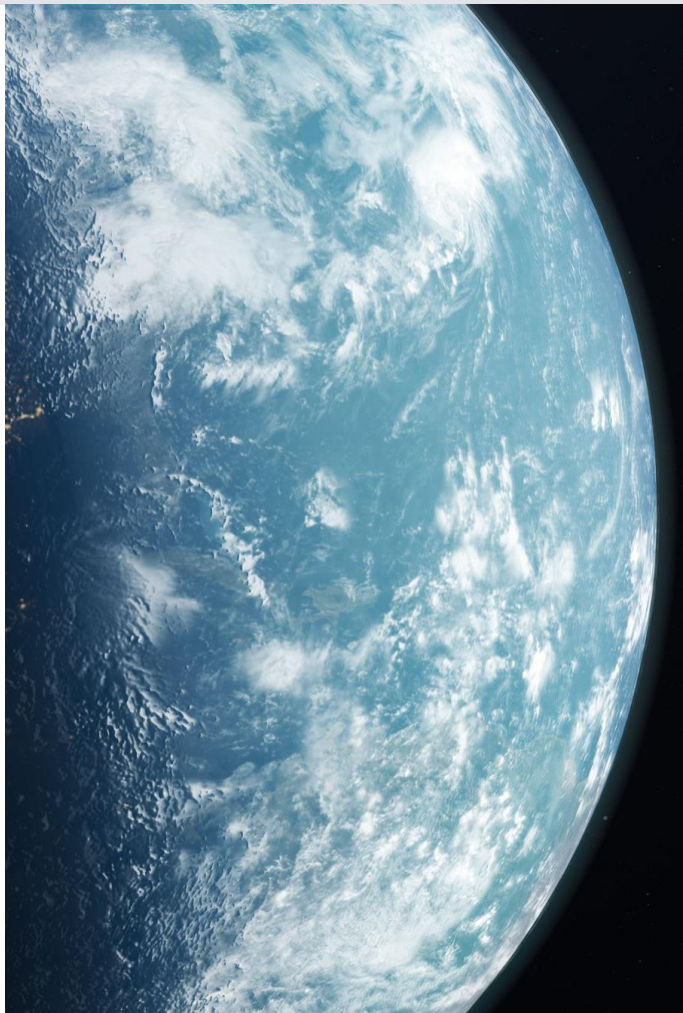
Data collection:

- Web scraping: our data was collected using the BS4 (beautiful soup 4) library
- API: using the SpaceX API we were able to get direct data from SpaceX

```
url="https://api.spacexdata.com/v4/launches/past"

response =requests.get(url)

response.json()
```

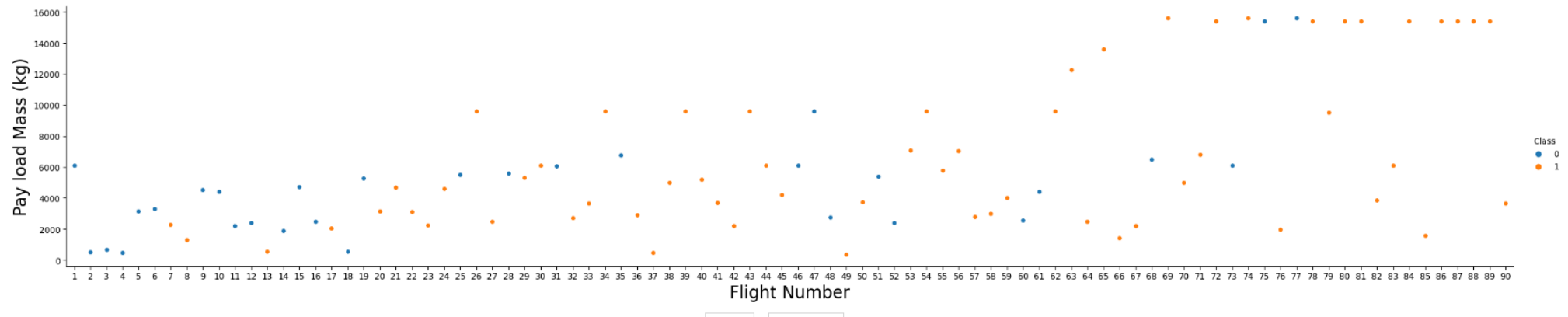


# Wrangling Data using an API

```
data = pd.json_normalize(response.json())
```

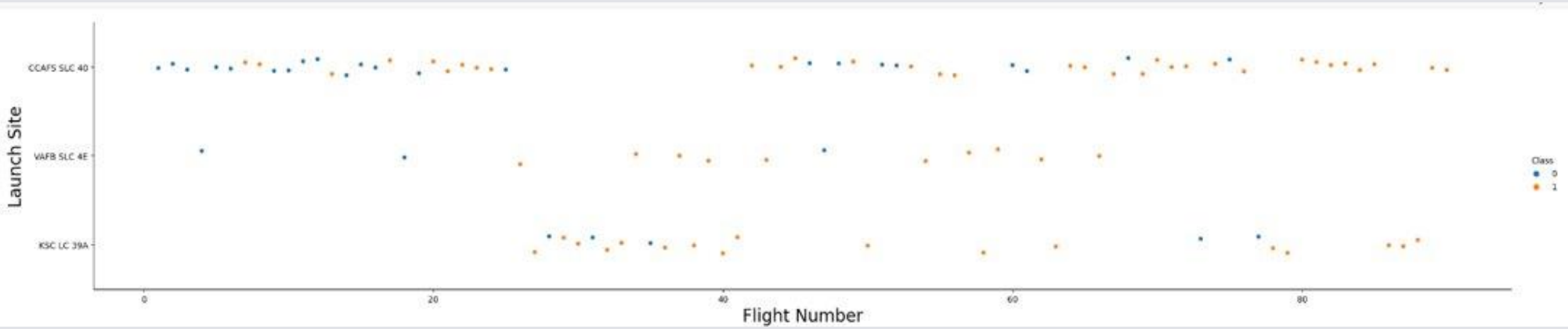
| | static_fire_date_utc | static_fire_date_unix | tbd | net | window | rocket | success | details | crew | ships | capsules | payloads | launchpad | auto_update | failures | flight_number | name | date_utc | date_unix |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 2006-03-17T00:00:00.000Z | 1.142554e+09 | False | False | 0.0 | 5eb0d095eda69955f709d1eb | False | Engine failure at 33 seconds and loss of vehicle | [] | [] | [] | [5eb0e4b5b6c3bb0006eeb1e1] | 5e9e4502f5090995de566f86 | True | [{'time': 33, 'altitude': None, 'reason': 'merlin engine failure'}] | 1 | FalconSat | 2006-03-24T22:30:00.000Z | 1143239400 |
| 1 | None | NaN | False | False | 0.0 | 5eb0d095eda69955f709d1eb | False | Successful first stage burn and transition to second stage, maximum altitude 289 km, Premature engine shutdown at T+7 min 30 s, Failed to reach orbit, Failed to recover first stage | [] | [] | [] | [5eb0e4b6b6c3bb0006eeb1e2] | 5e9e4502f5090995de566f86 | True | [{'time': 301, 'altitude': 289, 'reason': 'harmonic oscillation leading to premature engine shutdown'}] | 2 | DemoSat | 2007-03-21T01:10:00.000Z | 1174439400 |
| 2 | None | NaN | False | False | 0.0 | 5eb0d095eda69955f709d1eb | False | Residual stage 1 thrust led to collision between stage 1 and stage 2 | [] | [] | [] | [5eb0e4bdb6c3bb0006eeb1e3, 5eb0e4bdb6c3bb0006eeb1e4] | 5e9e4502f5090995de566f86 | True | [{'time': 140, 'altitude': 35, 'reason': 'residual stage-1 thrust led to collision between stage 1 and stage 2'}] | 3 | Trailblazer | 2008-09-03T03:34:00.000Z | 1217734440 |
| 3 | 2008-09-20T00:00:00.000Z | 1.221869e+09 | False | False | 0.0 | 5eb0d095eda69955f709d1eb | True | Ratsat was carried to orbit on the first successful orbital launch of any privately funded and developed, liquid-propelled carrier | [] | [] | [] | [5eb0e4b7b6c3bb0006eeb1e5] | 5e9e4502f5090995de566f86 | True | [] | 4 | RatSat | 2008-09-28T23:15:00.000Z | 1222643700 |

# Exploratory data analysis




*Distribution of flight vs payload mass*

# Exploratory data analysis




*Distribution of flight vs launch sites*

# Exploratory data analysis (SQL)

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

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

# total payload mass carried by boosters launched by NASA (CRS)

45596.0

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

2928.4

# the date when the first succesful landing outcome in ground pad was achieved

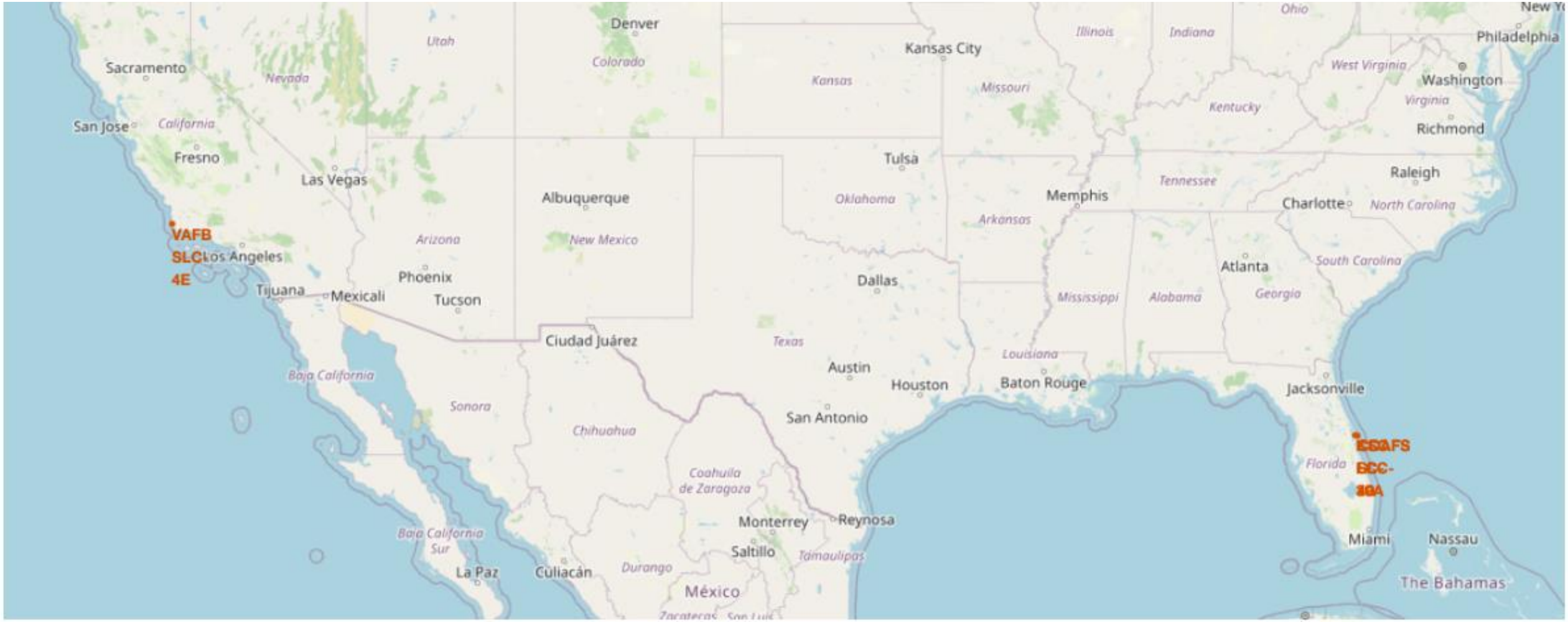
01/08/2018

# Exploratory data analysis (SQL)

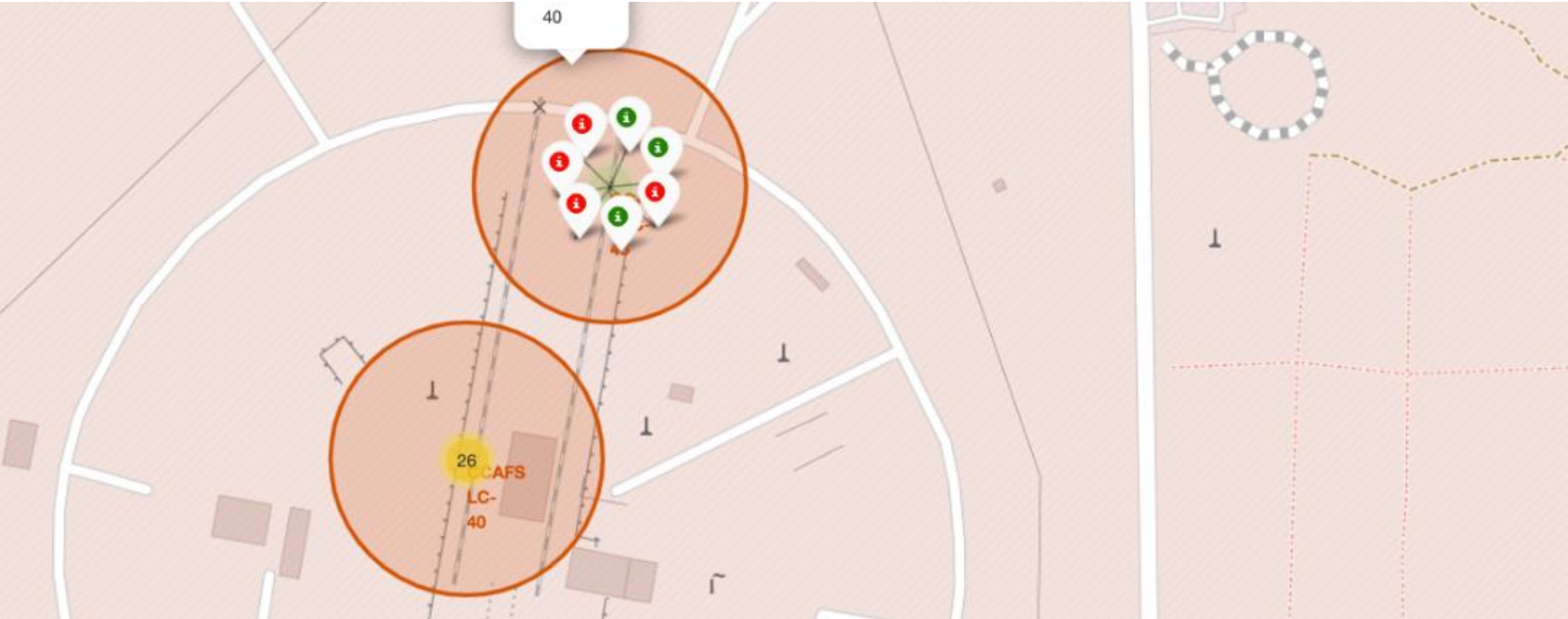
# the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

('October', 'Failure (drone ship)', 'F9 v1.1 B1012', 'CCAFS LC-40') ('November', 'Controlled (ocean)', 'F9 v1.1 B1013', 'CCAFS LC-40') ('February', 'No attempt', 'F9 v1.1 B1014', 'CCAFS LC-40') ('April', 'Failure (drone ship)', 'F9 v1.1 B1015', 'CCAFS LC-40') ('April', 'No attempt', 'F9 v1.1 B1016', 'CCAFS LC-40') ('June', 'Precluded (drone ship)', 'F9 v1.1 B1018', 'CCAFS LC-40') ('December', 'Success (ground pad)', 'F9 FT B1019', 'CCAFS LC-40')

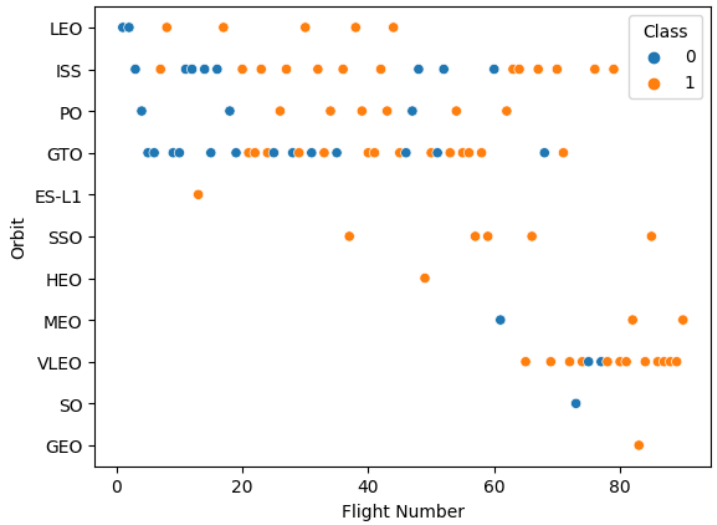
# Folium interactive Map

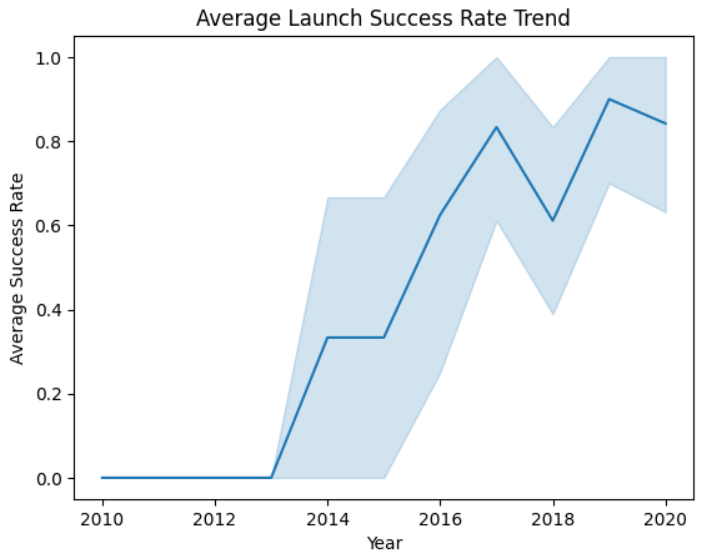


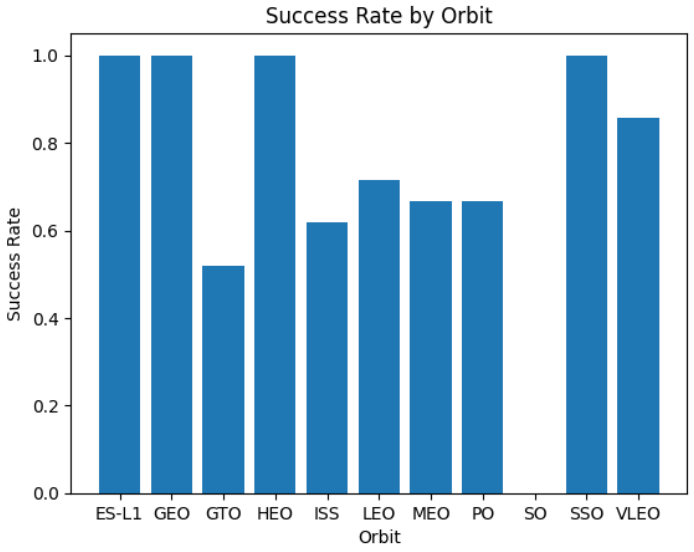
# Folium interactive Map



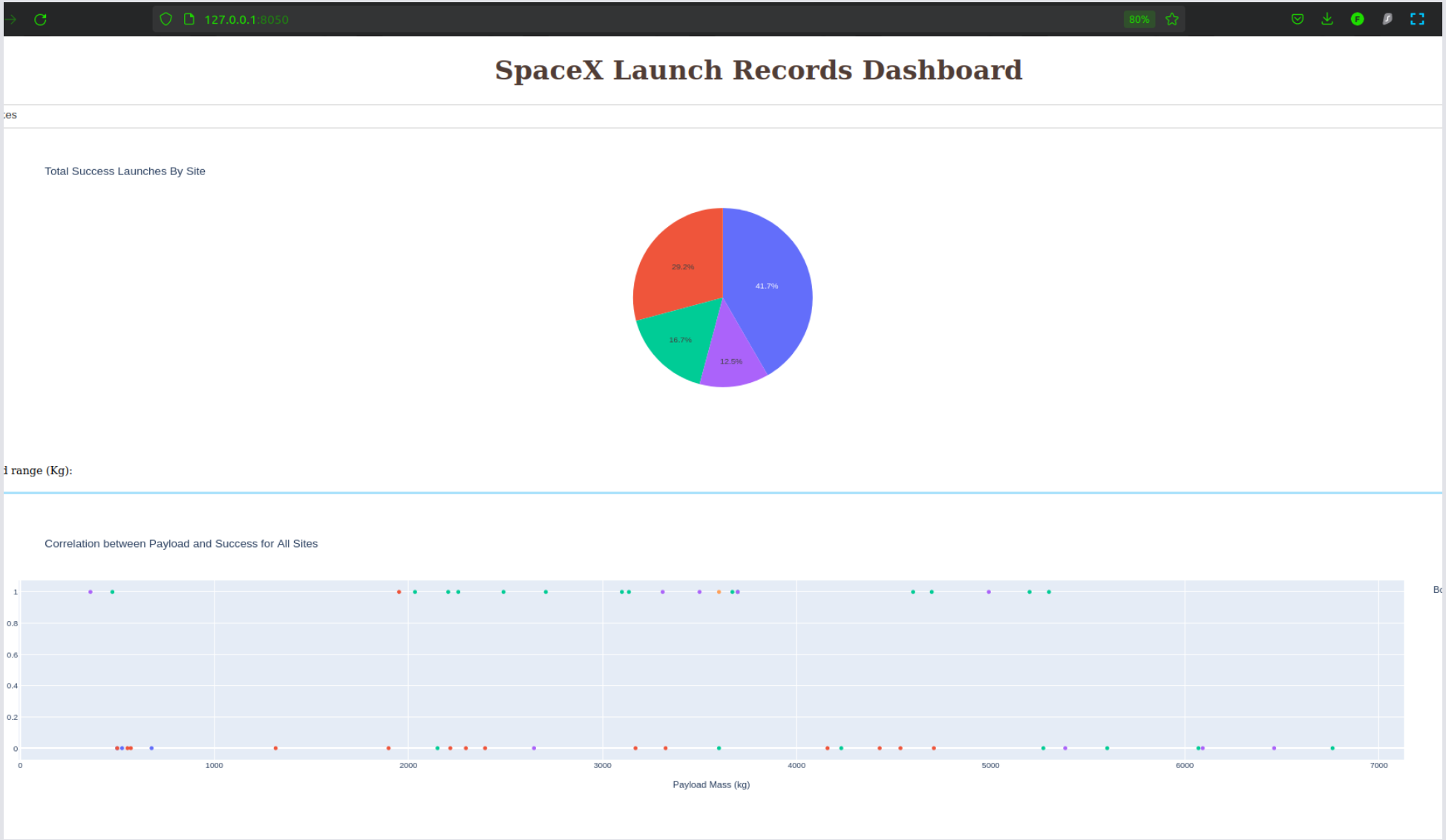
# Data Visualization



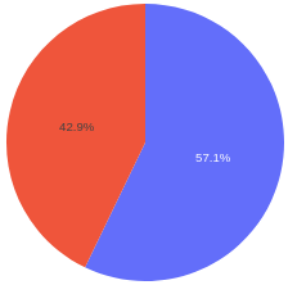
# **Dashbord** (Plotly/Dash)



# SpaceX Launch Records Dashboard

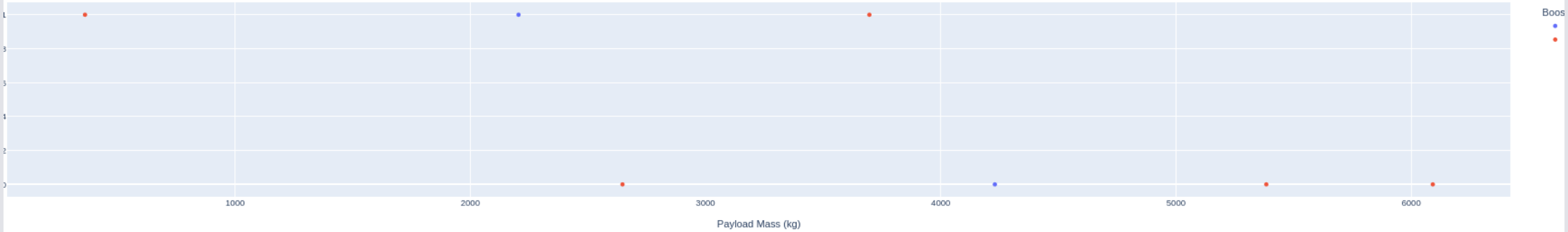
SLC-40

Total Success Launches By Site




ange (Kg):

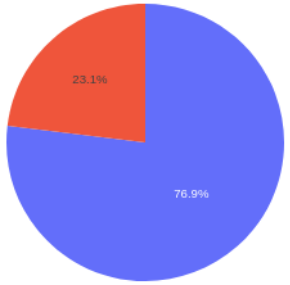
Correlation between Payload and Success for CCAFS SLC-40



# SpaceX Launch Records Dashboard

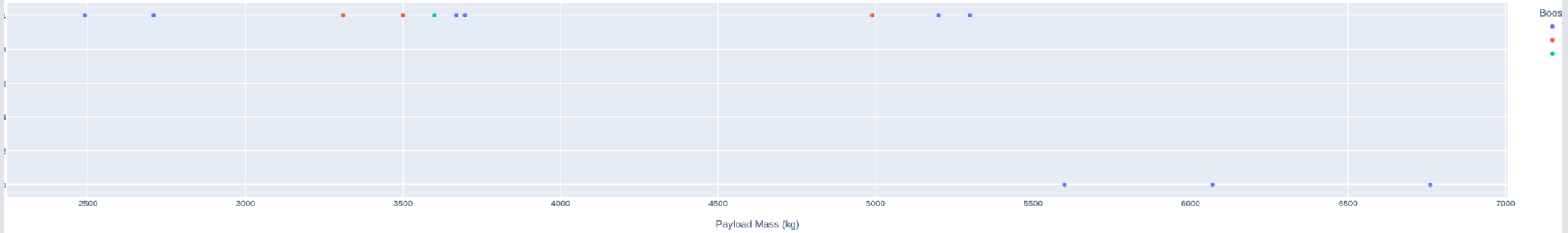
-39A

Total Success Launches By Site




ange (Kg):

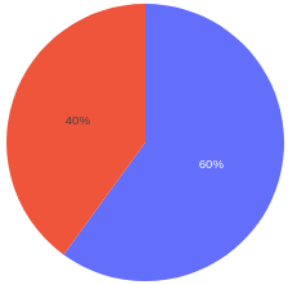
Correlation between Payload and Success for KSC LC-39A



# SpaceX Launch Records Dashboard

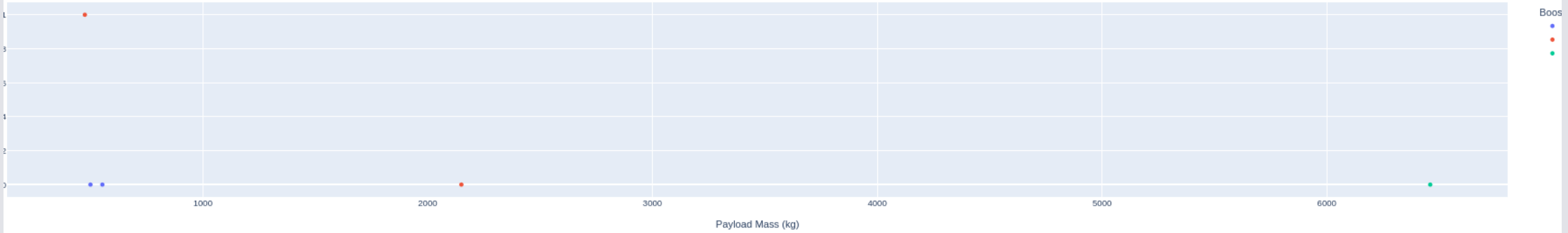
LC-4E

Total Success Launches By Site




range (Kg):

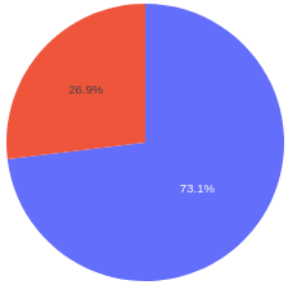
Correlation between Payload and Success for VAFB SLC-4E



# SpaceX Launch Records Dashboard

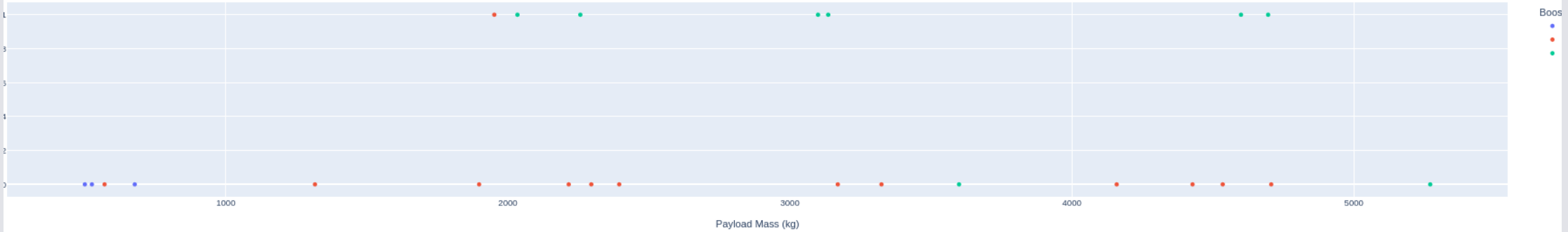
LC-40

Total Success Launches By Site



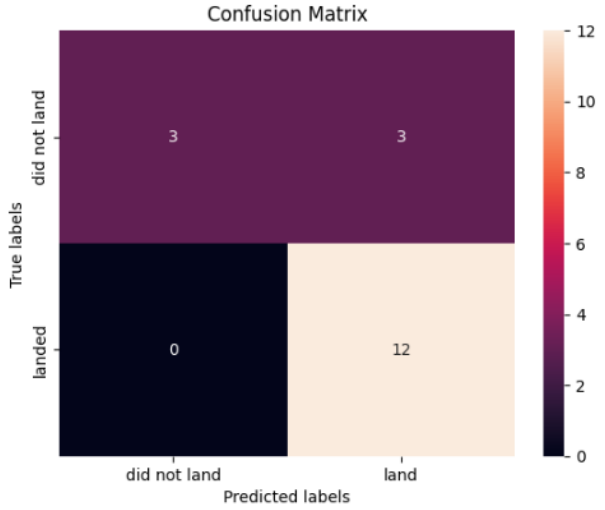

ange (Kg):

Correlation between Payload and Success for CCAFS LC-40



# Predictive Analysis

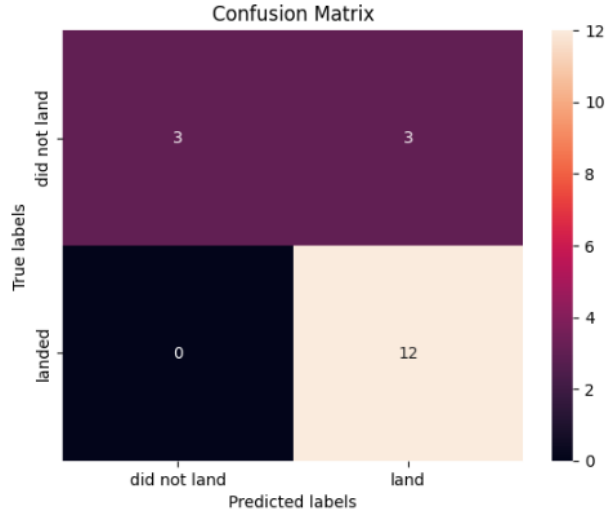
```
[15]: logreg_cv.score(X_test, Y_test)
[15]: 0.8333333333333334
[16]: yhat=logreg_cv.predict(X_test)
      plot_confusion_matrix(Y_test,yhat)
```





```
X_train, X_test, Y_train, Y_test
[10]: X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=2)
[11]: Y_test.shape
[11]: (18,)
[12]: parameters ={'C':[0.01,0.1,1],
                  'penalty':['l2'],
                  'solver':['lbfgs']}
*[13]: parameters = {"C":[0.01,0.1,1],'penalty':['l2'], 'solver':['lbfgs']}
      lr = LogisticRegression()
      logreg_cv = GridSearchCV(lr, parameters, cv=10)
      logreg_cv.fit(X_train, Y_train)
```

```
yhat = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```




```
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()
```

```
knn_cv = GridSearchCV(KNN, parameters, cv=10)
knn_cv.fit(X_train, Y_train)
```

```
GridSearchCV
estimator: KNeighborsClassifier
KNeighborsClassifier
```

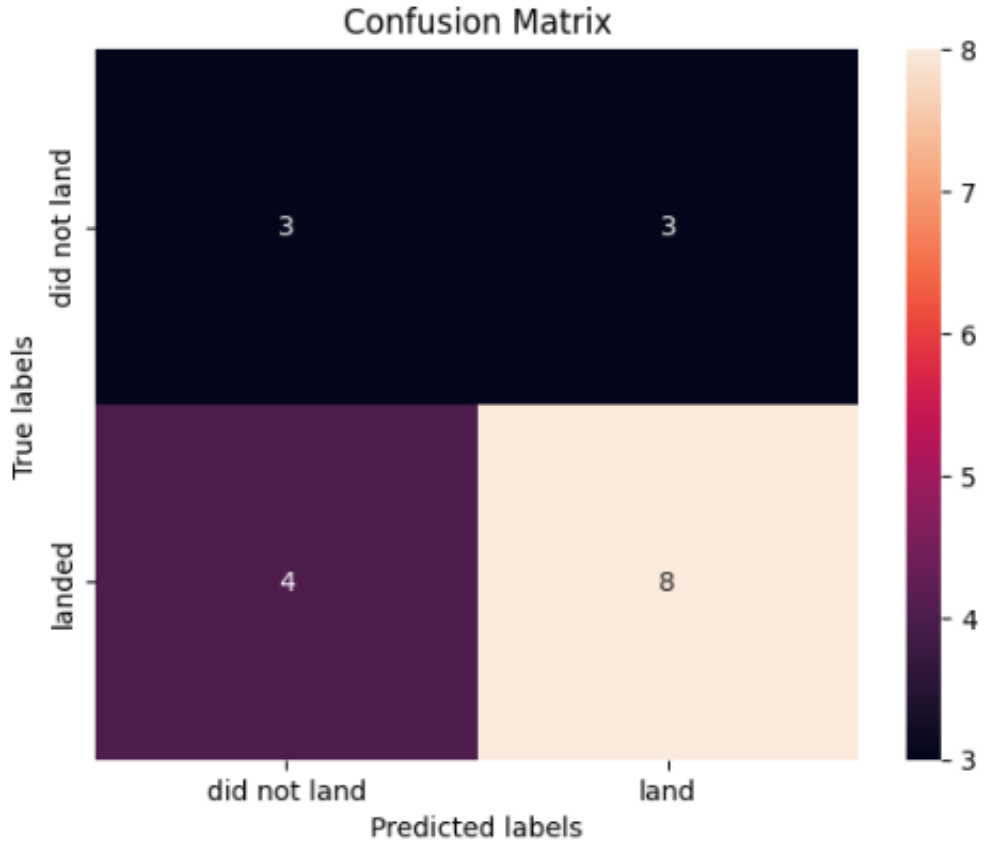
```
print("tuned hpyerparameters :(best parameters) ",knn_cv.best_params_)
print("accuracy :",knn_cv.best_score_)
```

```
tuned hpyerparameters :(best parameters)  {'algorithm': 'auto', 'n_neighbors': 3, 'p': 1}
accuracy : 0.6642857142857143
```

[25]:
```
knn_cv.score(X_test, Y_test)
```

[25]:
```
0.6111111111111112
```

[26]:
```
yhat = knn_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



# Results

- Falcon 9 Boosters proved to be the most successful
- Successful launches have increased over the years
- SpaceX uses 4 launch sites
- Launch with a heavier payload generally performed better

# Thanks