Space X Falcon 9 Rocket Analysis

Hardik Seju

05-18-2022

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- References

EXECUTIVE SUMMARY

- This project initially focused on collecting data for Falcon 9 booster rocket through REST API and web scrapping.
- Space X falcon 9 launch analysis among different launching sites and satellite deployment orbits. Maximum number of the launches were from Cape Carnival site and GTO orbit accounted for maximum deployment.
- Machine learning models created to predict the successful landing of Falcon 9 Rockets where Decision algorithm showed the highest accuracy.
- KSC LC launch site showed the highest success rate while Orbit of payload deployment and payload mass were found to have good correlation with the success of the mission.

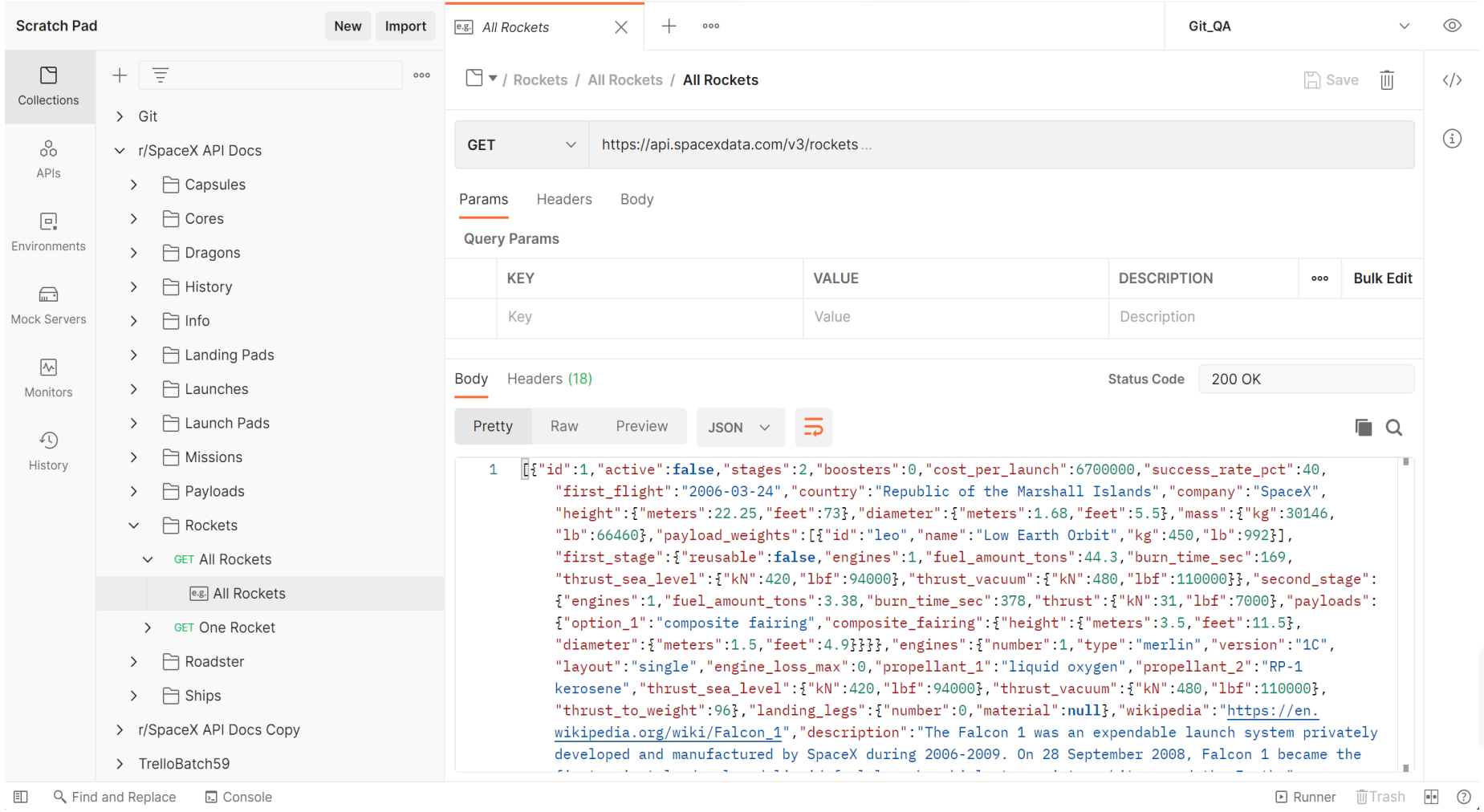
INTRODUCTION

- In age of new space era, many space companies want to reduce the cost of rocket launches along with successful missions.
- In this project, we have analyzed the data from Space X falcon 9 rocket launches and aimed to predict the cost of each launch.
- Additionally, this project also aims to find meaningful insights as to how many successful launches have occurred, their launch and landing sites and other such relevant information.

METHODOLOGY

- Data Collection through REST API from spacex API website https://api.spacexdata.com and web scrapping was carried out from https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- For the purpose of finding the cost of launches, we needed amount of rocket launches data atleast of 5 years. Only falcon 9 was considered for the purpose of study and related fields like date of launch, location of launch, payload mass, booster version, orbit of deployment, etc. were taken into consideration.
- Data Wrangling and data cleaning
- Exploratory Data analysis to gain insights on spacex data

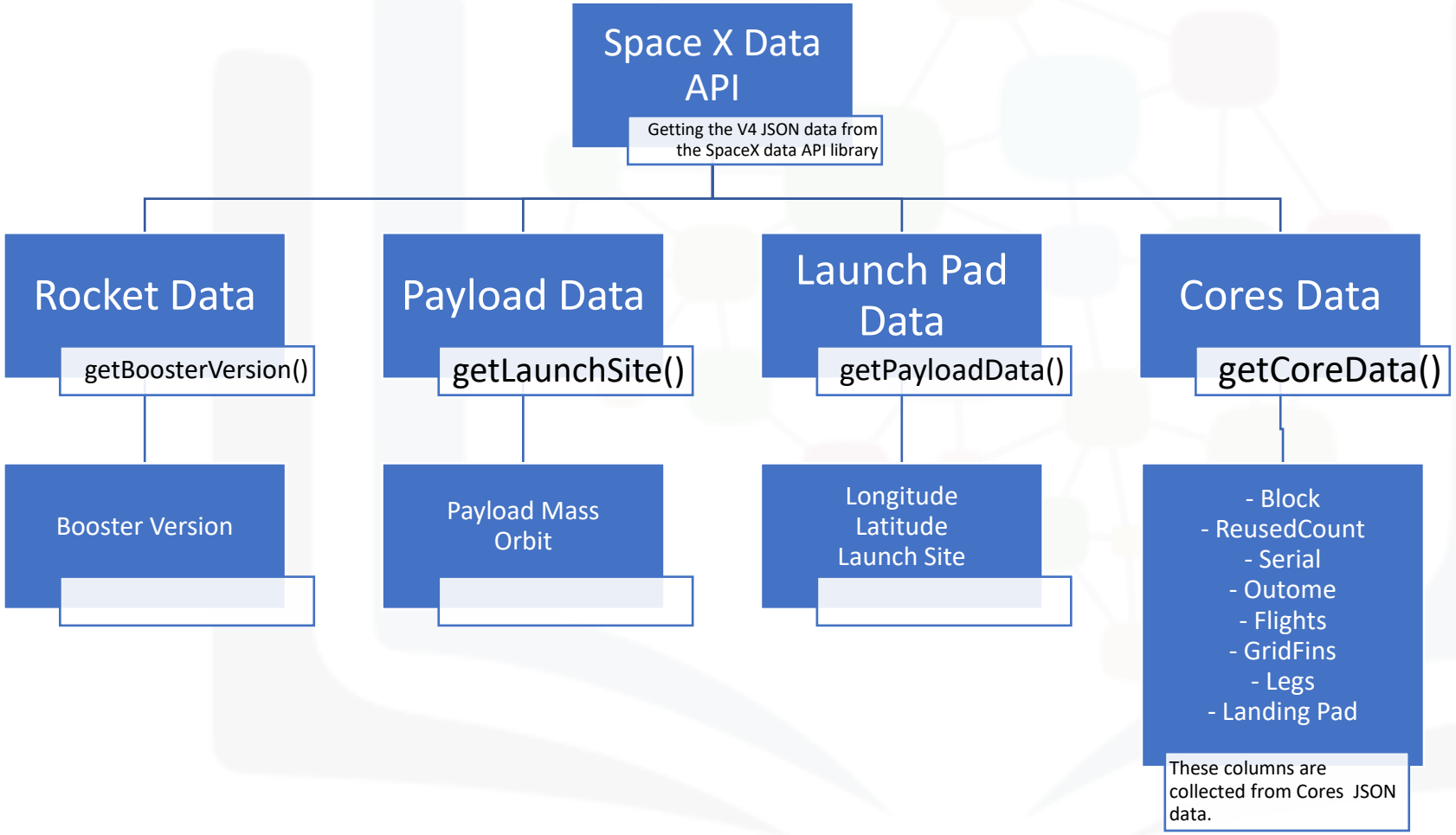
POSTMAN Access to SPACEX API



IBM Developer

SKILLS NETWORK

Extracting Data Using REST API



ID's for rocket, payload, launchpad and cores are fetched from SPACEX API V4 Json file and using these IDs, more data is then extracted from each of these sub data.

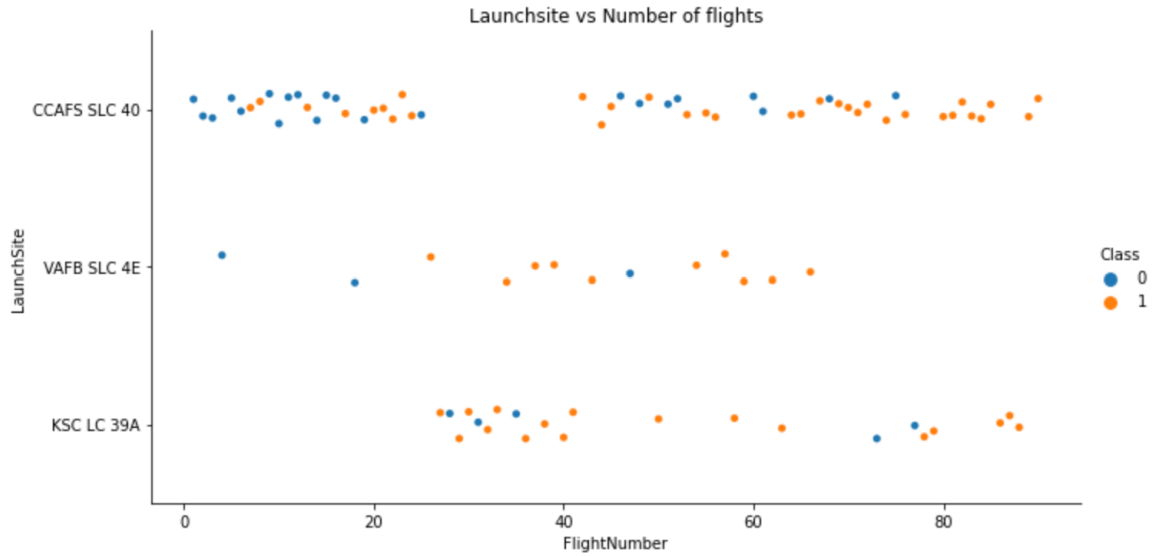
Extracting Data Using Web Scraping

1. Wikipedia Space X Rocket Data
2. Extracting all Column Names from HTML table header
3. Parsing the data to create a Dataframe

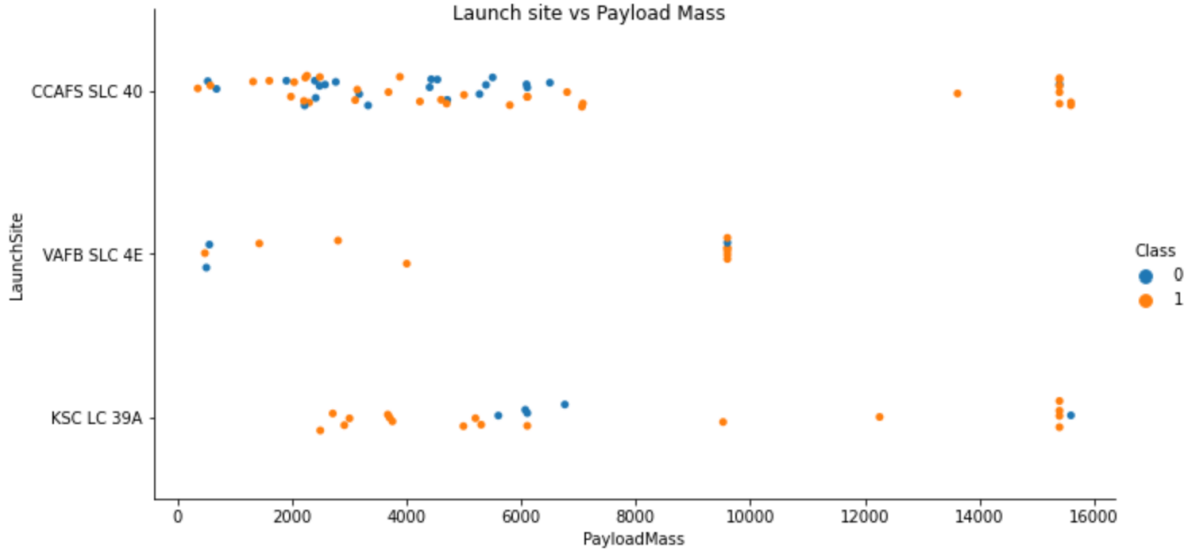
Using “requests” and “Beautiful Soup” libraries for python, the data from table available in spaceX Wikipedia is extracted to be parsed in to a padas Dataframe.

EDA

Most of the successful flights for all of the three-launch site happened at greater many launches or flight number. It also seems the CCAFS has the most amount of launches than any other site.

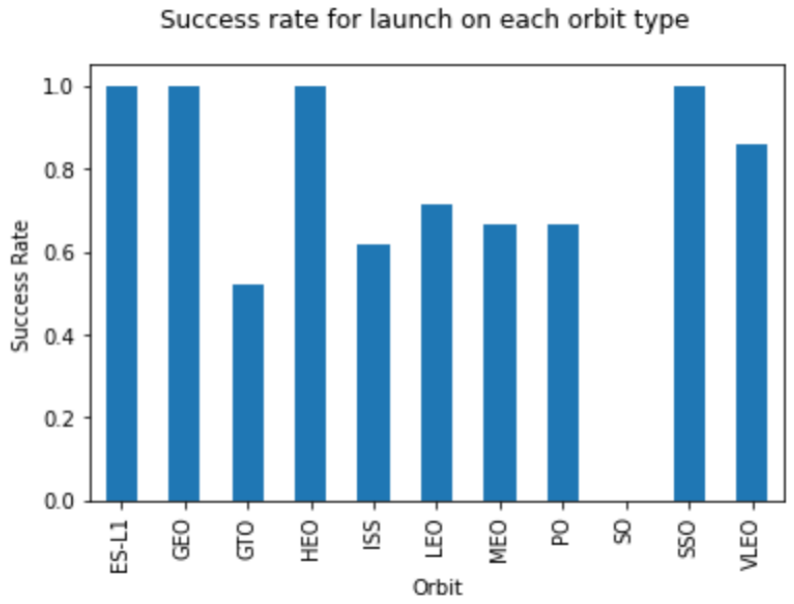


Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavy payload mass(greater than 10000). Additionally, highest payload mass has been distributed between CCAFS and KSC launch sites having high success ratio.

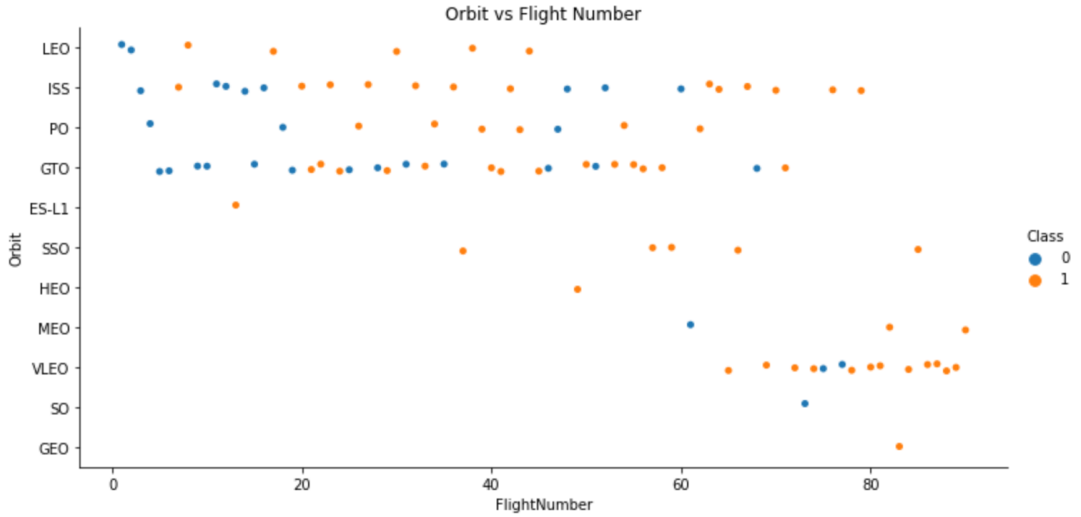


EDA – continued..

It appears that ES-L1, GEO, HEO and SSO are the orbits having the highest success rate of launch.

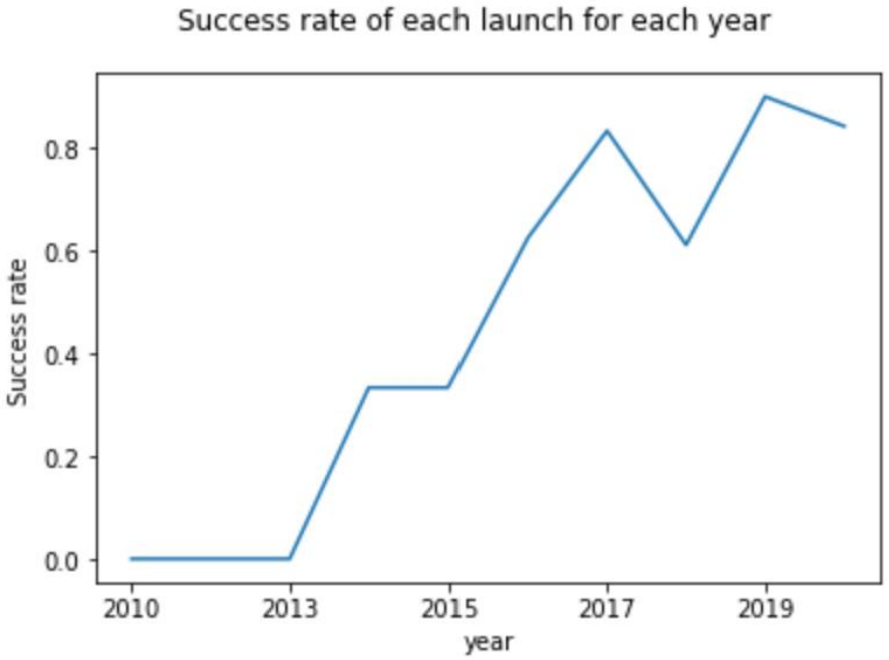


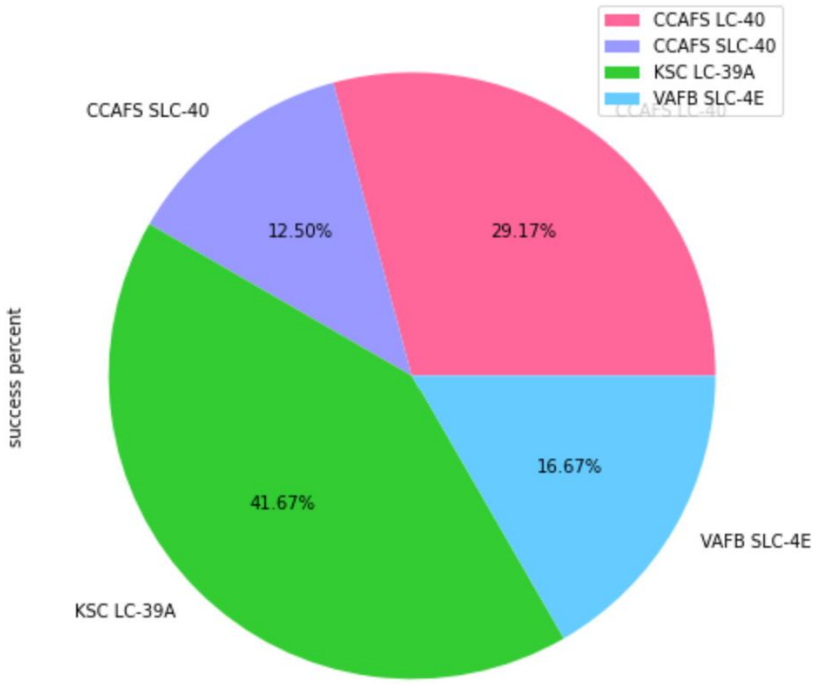
You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit. As the number of flights increased the success ratio is also increased among all orbits however most being in the lower earth orbit below SSO.



Data Visualization Insight

The success rate of the launches has a steep rise after 2013 with a slight dip in 2018.





KSC LC has the most the successful rocket launches from its site, while least successful being CCAFS.

EDA USING SQL

IBM DB2 has been used to connect to IBM database and querying the data using SQL.

Results of few questions asked:

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

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

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

1
45596

- List the total number of successful and failure mission outcomes

mission_outcome	COUNT
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

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

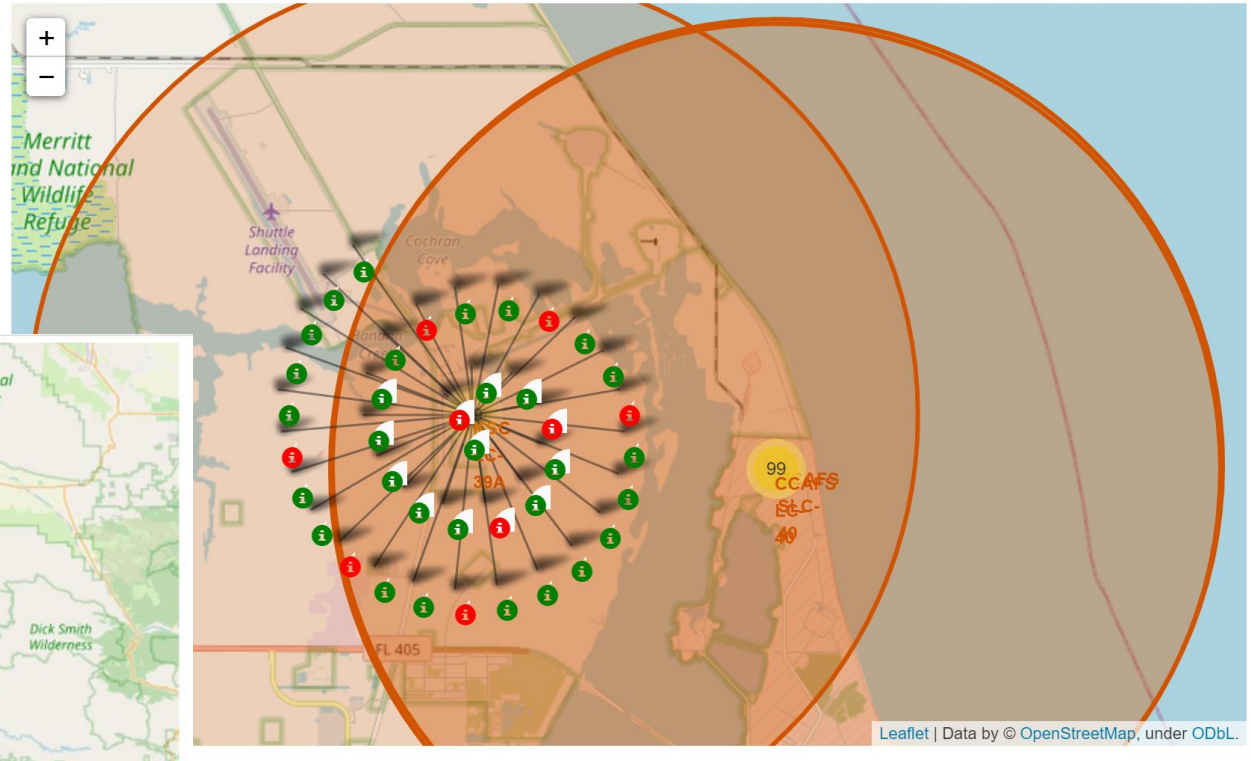
1
2928

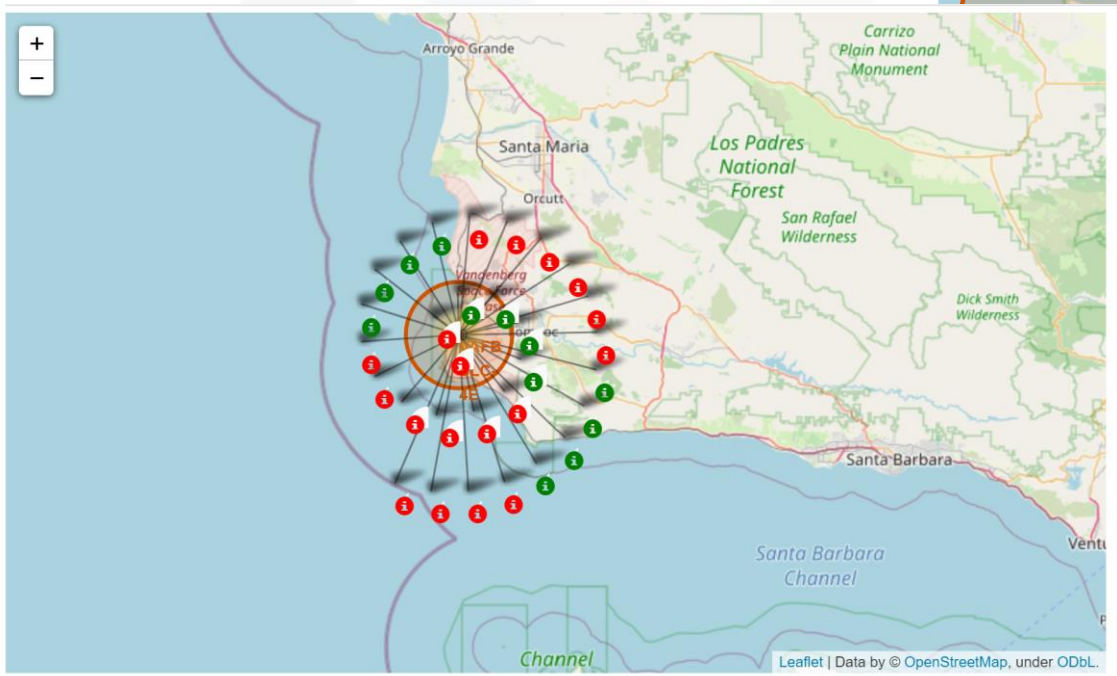
- List the date when the first successful landing outcome in ground pad was achieved.

1
2015-12-22

RESULTS USING FOLIUM

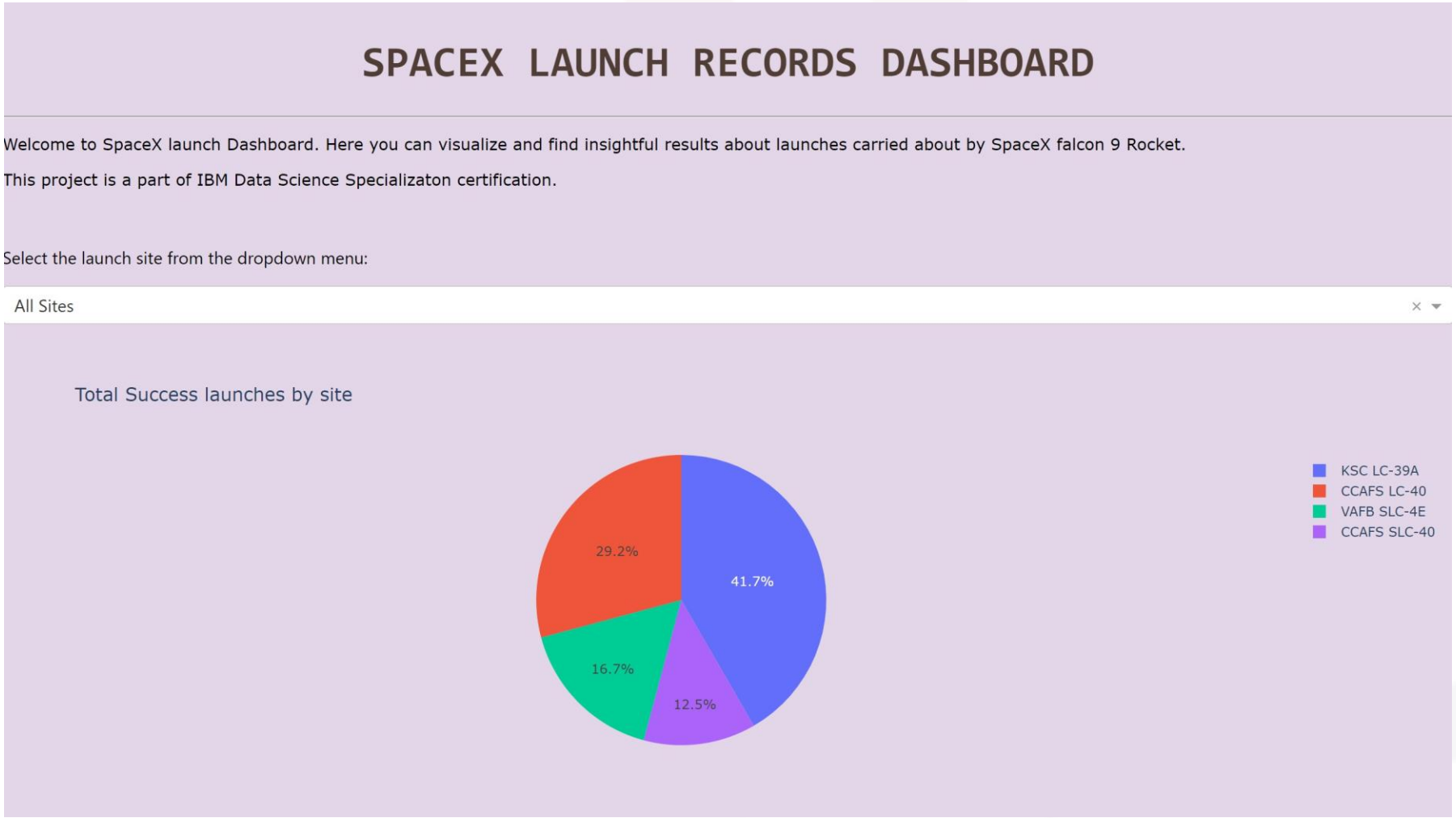
- The map on right shows 39 launches from the site KSC LC with green markers being the successful one and red being unsuccessful.



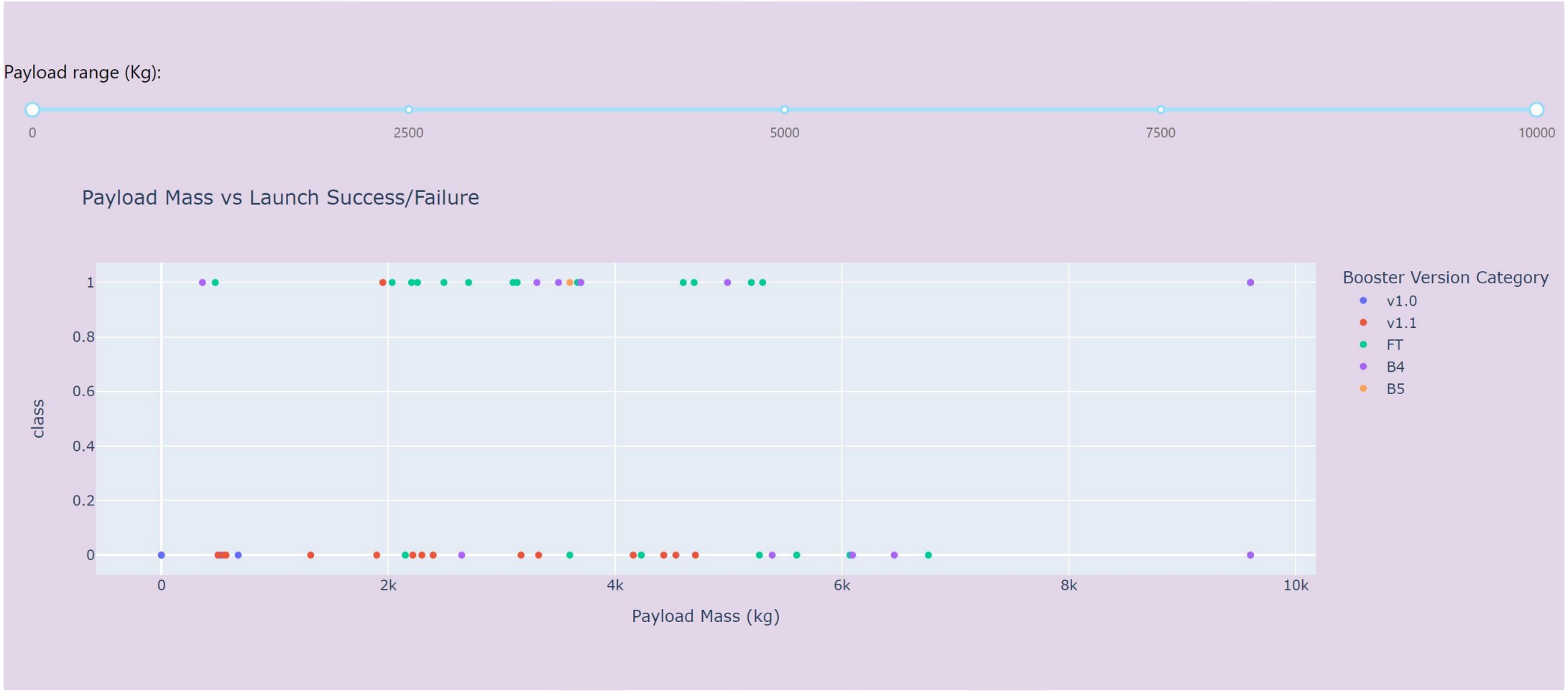


- The map on left shows 30 launches from the site VFAB with green markers being the successful one and red being unsuccessful.

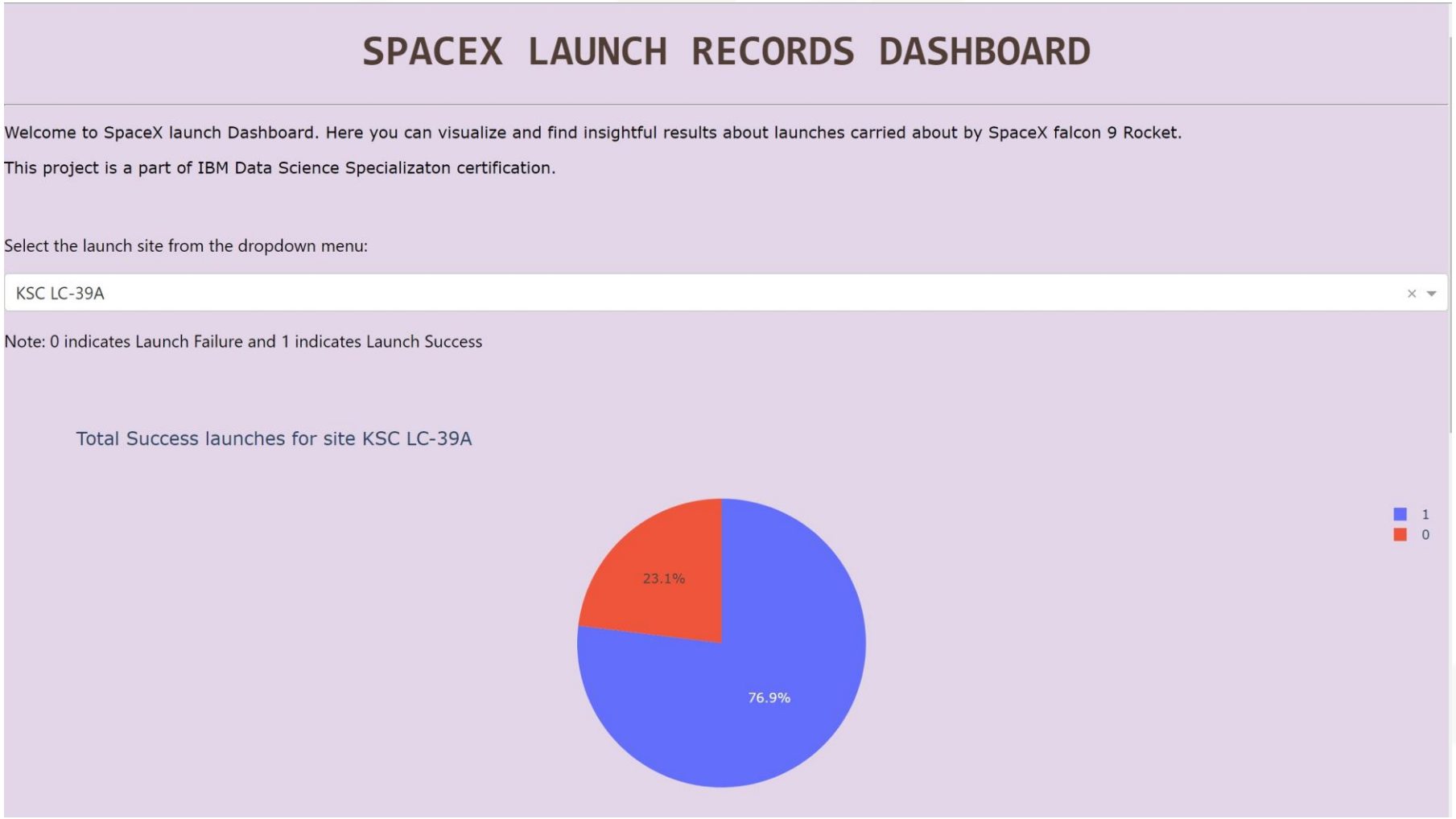
DASHBOARD – All Sites selected



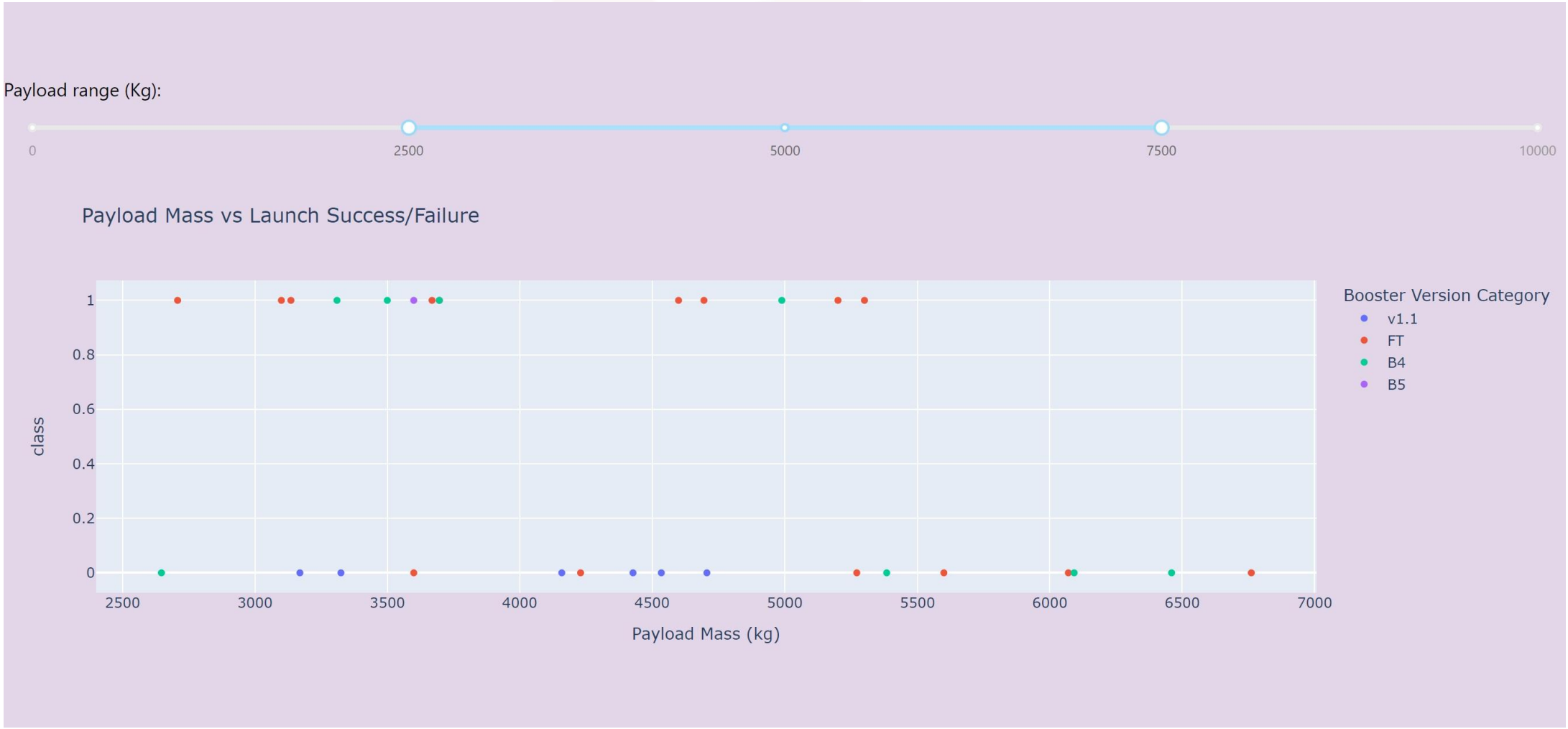
DASHBOARD – All sites selected



DASHBOARD for KSC LC



DASHBOARD – Limited Payload Range



Results on Dashboard

Following are some insights from Dashboarding:

- KSC LC has the largest number of successful launches.
- Again, KSC LC 39A has highest success rate among all sites.
- Rockets carrying payload between 2000-4000 kg appears to have highest success rate while rockets having payload higher than 6000 kg has lowest success rate.
- Booster Version FT is most successful to launch the payload in range 2k-4k kg.

PREDICTIVE ANALYSIS METHOD:

Predictive modeling is carried on the dataset to determine whether the rocket will land successfully or not. For this purpose, following steps are performed as a part of predictive analysis:

1. Load the data and find the list of dependent features as well as predictive variable, that will be Class in our case.
2. Standardize the data for higher accuracy of the model by scaling the values in same range for all columns. This step is carried out after One hot Encoding for categorical data like Orbit of satellite deployment.
3. Transformed data is then divided into training and testing set.
4. Selecting best set of hyper-parameters using Sklearn's GridSerachCV for all four algorithms chosen for prediction.

PREDICTIVE ANALYSIS METHOD:

```
def plot_confusion_matrix(y,y_predict):
    "this function plots the confusion matrix"
    from sklearn.metrics import confusion_matrix

    cm = confusion_matrix(y, y_predict)
    ax= plt.subplot()
    sns.heatmap(cm, annot=True, ax = ax); #annot=True to annotate cells
    ax.set_xlabel('Predicted labels')
    ax.set_ylabel('True labels')
    ax.set_title('Confusion Matrix');
    ax.xaxis.set_ticklabels(['did not land', 'land']); ax.yaxis.set_ticklabels(['did not land', 'landed'])
```

Confusion Matrix function declaration.

```
X_train, X_test, Y_train, Y_test
```

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=2)
```

we can see we only have 18 test samples.

```
Y_test.shape
```

```
(18,)
```

Splitting data into training and testing set.

Finding the set of hyper parameters

```
parameters ={'C':[0.01,0.1,1],
             'penalty':['l2'],
             'solver':['lbfgs']}
```

```
parameters_log ={"C":[0.01,0.1,1],'penalty':['l2'], 'solver':['lbfgs']}# l1 lasso l2 ridge
lr=LogisticRegression()
logreg_cv = GridSearchCV(lr, parameters_log, cv=10)
logreg_cv.fit(X_train, Y_train)
```

```
GridSearchCV(cv=10, estimator=LogisticRegression(),
             param_grid={'C': [0.01, 0.1, 1], 'penalty': ['l2'],
                         'solver': ['lbfgs']})
```

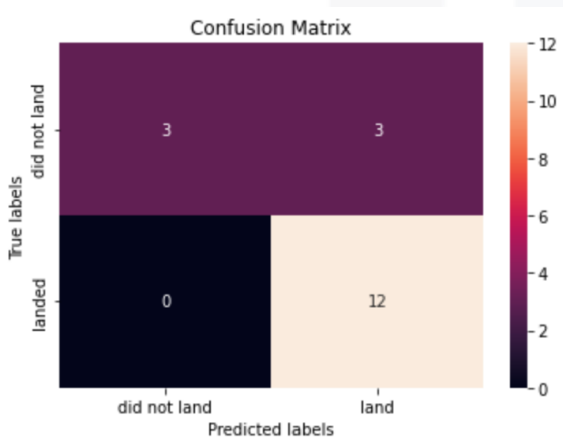
We output the `GridSearchCV` object for logistic regression. We display the best parameters using the data attribute `best_params_` and the accuracy on the validation data using the data attribute `best_score_`.

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

```
tuned hpyerparameters :(best parameters)  {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
accuracy : 0.8464285714285713
```

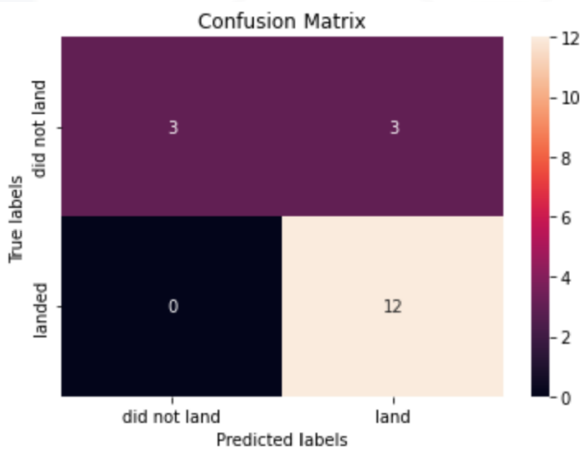
PREDICTIVE ANALYSIS Results:

Logistic Regression



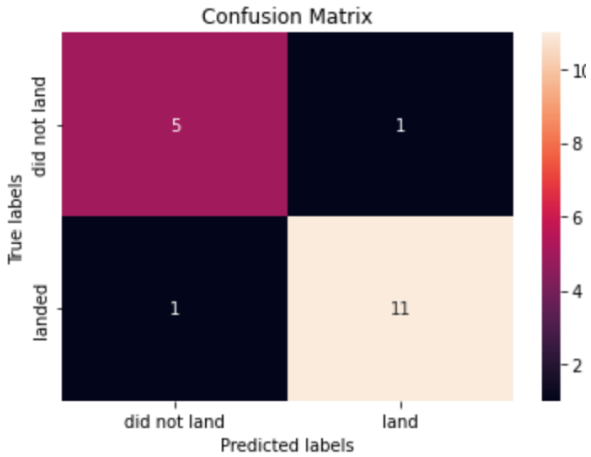
Accuracy: 83.33 %

SVM Classifier



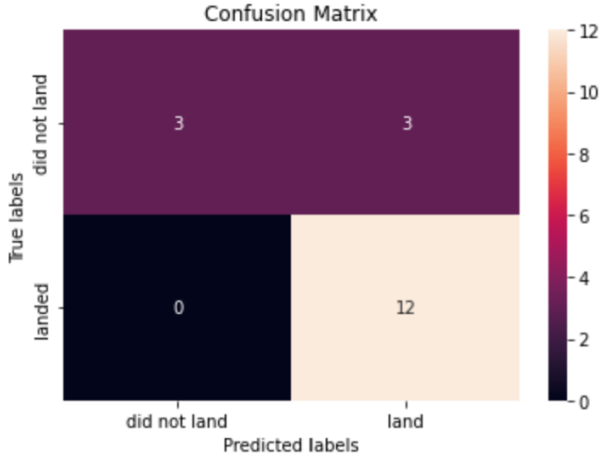
Accuracy: 83.33 %

Decision Trees



Accuracy: 88.88 %

KNN classifier



Accuracy: 83.33 %

Only Decision Tree classifier has lower False Positive and False Negative giving the highest accuracy among all models. Hence **Decision Tree Classifier** is best the model for Predictive Analysis.

DISCUSSION

- The raw data is continuously added on Wikipedia, and it would be interesting to see the prediction accuracy with more amount of data. The more the data, the more accuracy on findings and concrete conclusions.

CONCLUSION

- Space X falcon 9 launches had 66.66% successful landings which were distributed across 3 launch sites and the deployment of satellites distributed in 11 different orbits.
- Falcon 9 rockets launched from KSC will be more successful than other launch sites and any payload mass between 2k-4k will have higher chances of successful mission.
- Landing success depends on many features like the physical working attributes of a rocket, orbit of payload deployment, etc.
- All the sites are close to important transportation needs and near the sea cost. VAFB site has the closest access to NASA railroad network.

APPENDIX

Correlation Matrix to see which parameter is closely dependent or related to Class (Output label).