IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Prashant S. Kale
16 March 2022

Outline

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

Executive Summary

Summary of methodologies

The Data Science project is for prediction of SPACEX rocket launch first stage Falcon 9 will land successfully. This is crucial information as SPACEX can reuse the Falcon 9 rocket for other launches and save millions of dollar cost.

Data collection for rocket launches will be from Open Source REST API and web scraping Falcon 9 launch records from Wikipedia page for the vital data about Rocket, Booster version, Core data, Launch site, Payload and outcome of the Launch.

The raw data collected will be wrangled using API and dealing with null values and dropping unnecessary data to get clear refined data.

This data will be useful for prediction whether the first stage rocket will be able to land successfully or not.

Executive Summary

Summary of all results

We will explore data visualization for finding correlation between different features of the Rocket launches by scatter plots of Flight number, Launch site, Payload mass and Bar chart for Orbit type and how these feature has direct relations with successful landings.

We will analyze Launch sites and proximity with Folium to mark the various Geographical Launch sites to choose the optimal launch site with more successful landings.

We will build Dashboard application with Python plotly Dash package to display the success percentage of landing on each site and overall success rate for all sites. It will also display scatter chart for success rate for payload mass on each launch site.

Finally we will build machine learning pipeline to predict whether the first stage of Falcon 9 will land successfully.

Introduction

- In this capstone, we will predict if the Falcon 9 first stage will land successfully.
- SpaceX cost for Falcon 9 first stage rocket is 62 million dollars while competitors cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- In our project we will consider various factors and parameters which may effect on outcome of Falcon 9 launch i.e if it will be successful or failure.
- Therefore if we can determine if launch will be successful, it is possible to predict the cost of the launch for the first stage rocket as we can reuse the first stage rocket for subsequent launches and save the cost of the launch.

Section 1

Methodology

Methodology

Executive Summary

Our methodology for the project will be first Business understanding and with analytic approach for correct data requirement, collection and understanding of the data.

After collection of required data, the data pre-processing will be done to remove unnecessary data fields, replacing null values with mean and convert categorical variables to dummy variables.

The final data set created will have numeric and Boolean values in all the fields.

We will further insights into the data by applying some basic statistical analysis and data visualization, you'll be able to see directly how variables might be related to each other.

Finally we will split the collected data into groups for training and test sets defined by categorical variables which will be used to build, evaluate, and refine predictive models for discovering the insights and predictions.

Data Collection – SpaceX API

Data collection for rocket launches will be from Open Source REST API

https://api.spacexdata.com/v4/ . there are different endpoints for API which capsules,cores,past launches.

The endpoint data can be collected in following steps:

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

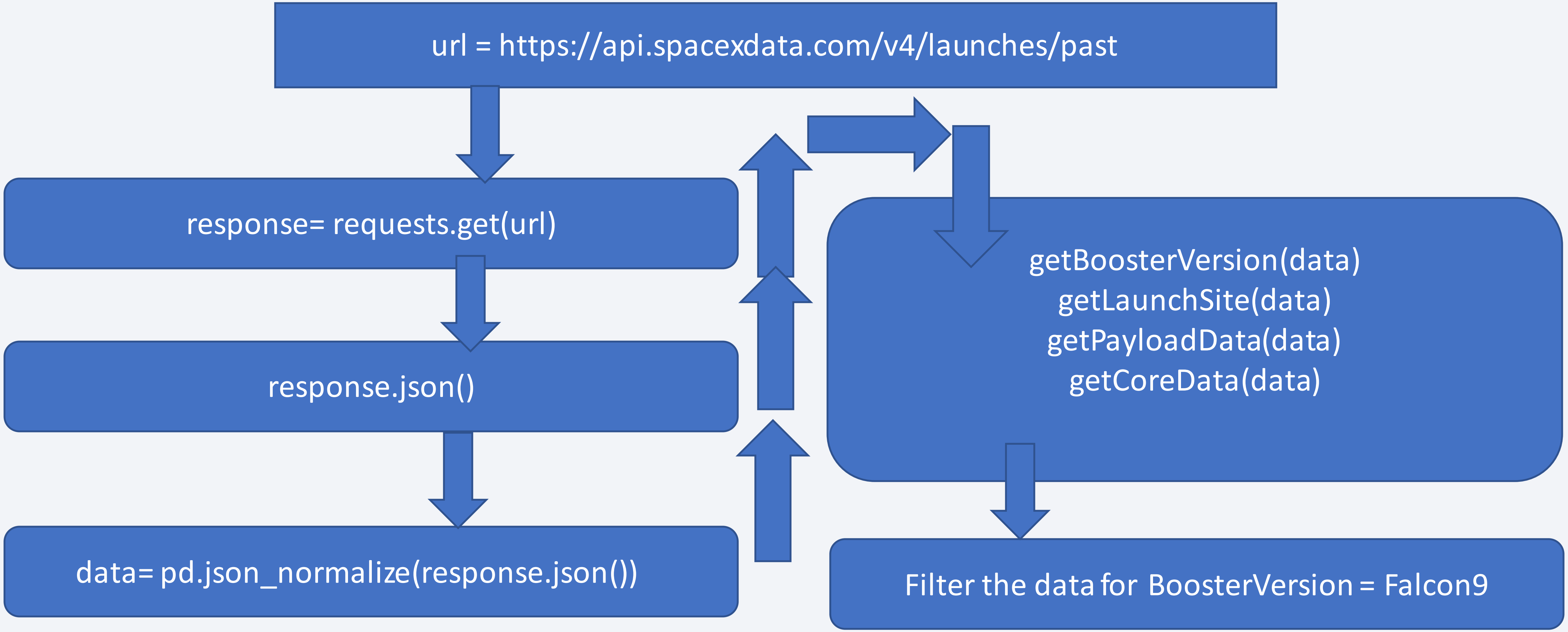
response= requests.get(url)

response.json()

finally json format will be converted to Pandas dataframe with following step.

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

Data Collection – API calls



Data Collection – SpaceX API – GitHub URL

Data Collection Notebook Link

Click on the above link or copy paste following URL in browser :
https://github.com/pskale/SPACEX/blob/master/jupyter-labs-spacex-data-collection-api.ipynb

Webscraping

Another source of data web scraping Falcon 9 launch records is from related Wikipedia pages by using Python BeautifulSoup package for extracting data from HTML tables.

It contains valuable information about Rocket, Booster version, Core data, Launch site, Payload and outcome of the Launch and convert it into Pandas dataframe for further visualization and analysis.

Web Scraping – SpaceX API – GitHub URL

[Web Scraping Notebook Link](https://github.com/pskale/SPACEX/blob/master/Spacex%20Falcon9%20Webscraping.ipynb)

Click on the above link or copy paste following URL in browser :
https://github.com/pskale/SPACEX/blob/master/Spacex%20Falcon9
%20Webscraping.ipynb

Data Wrangling

- For the field Rocket only Identification number is provided which is not enough so we will be using various functions to get endpoint information about Booster,Launchpad,Payload and core.
- Sampling Data

The data collected from API has Falcon 1 launches information along with Falcon 9 launches.

We will remove data for Falcon 1 and will keep data only for Falcon 9 launches.

- Dealing with Nulls

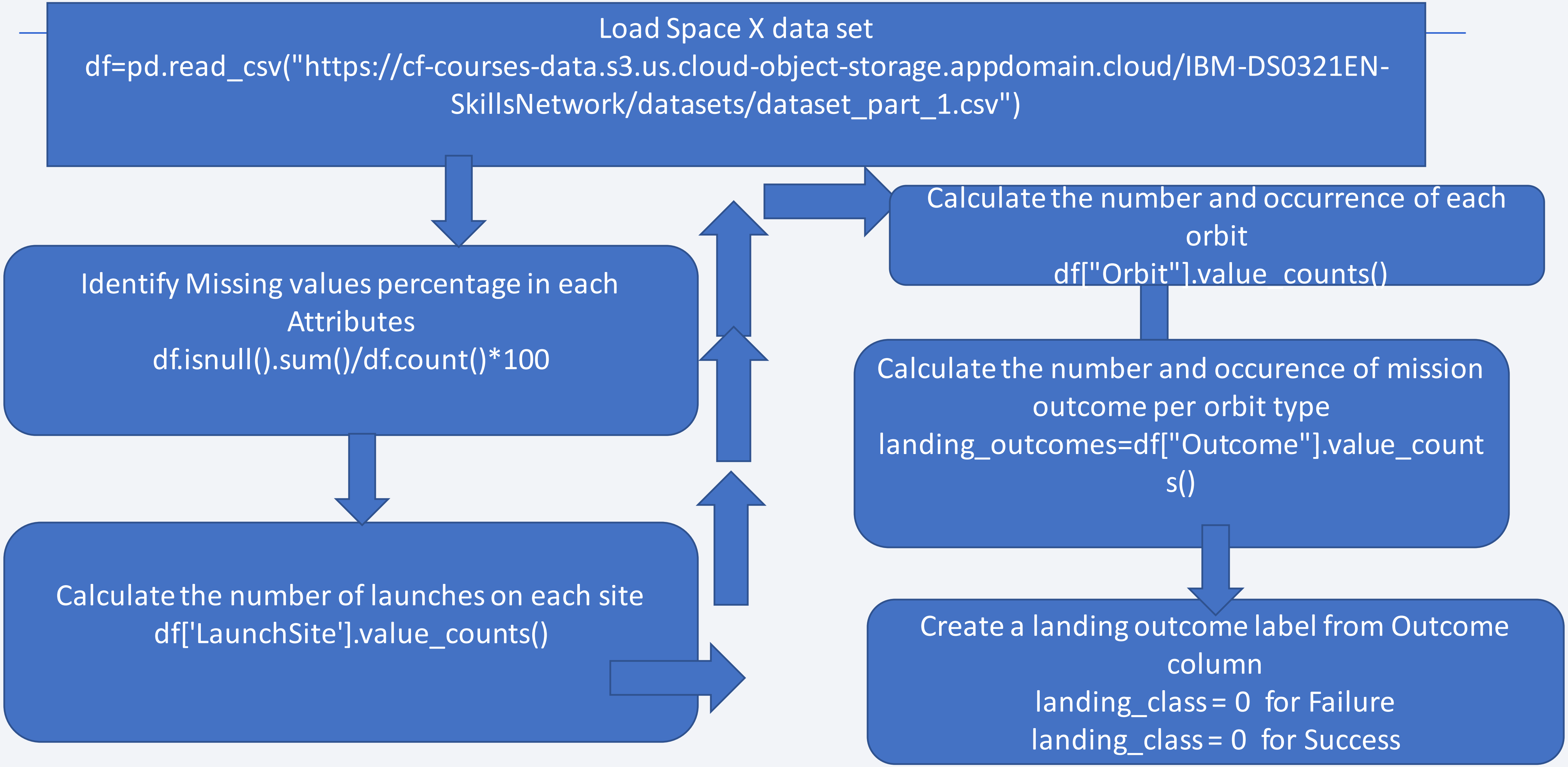
We will replace Null values in field Payloadmass with mean value of all data in payloadmass column.

We will keep Null values in Landing pad as it may not be used for particular launches which will be formatted using one hot encoding.

Data Wrangling

- From the dataset , we will be using the fields like Outcome which is Success or Failure of the rocket landing on the launch pad..
- Orbits for the different orbits to which rocket launch will take payloads to particular orbit like LEO or GEO.which ranges for 1000 Km to 35768Km from the Earth.
- The outcome indicates if first stage successfully landed on earth or failed which will be converted to classes 0(Failure) or 1 (Success)

Data Wrangling



Data Wrangling– SpaceX API – GitHub URL

[Data Wrangling Notebook Link](https://github.com/pskale/SPACEX/blob/master/EDA-%20Data%20Wrangling.ipynb)

Click on the above link or copy paste following URL in browser :
https://github.com/pskale/SPACEX/blob/master/EDA-%20Data%20Wrangling.ipynb

EDA with Data Visualization

We will be doing Feature engineering for exploratory data analysis to determine Falcon 9 will land successfully. Some attributes like launch number can be considered as success rate has improved with time. Also different geographical launch sites has different success rate which can also be a feature. Payload mass feature also plays the important role as different payload mass has different success rate.

We will find all attributes, which are having correlation with successful landing.Categorical variables i.e. launch site, Orbit will be converted with one hot encoding.

EDA with Data Visualization- Scatter plots

Following Scatter plots were helpful in finding the correlations between various features.

1. Flight number V/s PayloadMass scatter plot shows that as the flight number increases, the first stage is more likely to land successfully. Also it shows more massive the payload, the less likely the first stage will return.

2. Flight number V/s LaunchSite scatter plot shows that Launch site CCAFS SLC 40 has maximum number of launches and also maximum successful launches.Higher flight number has more successful launches.

3. PayloadMass V/s LaunchSite scatter plot shows VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

4. FlightNumber V/s OrbitType scatter plot shows 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.

5. Payload V/s OrbitType scatter plot shows heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS

EDA with Data Visualization- Bar/Line Charts

1. Bar chart for success rate for each OrbitType shows that ES-L1,GEO,HEO and SSO orbit has very high success rate.

2. Line chart for yearly trend for launch success rate shows that sucess rate since 2013 kept increasing till 2020.

EDA with Data Visualization– SpaceX API – GitHub URL

[EDA with Data Visualization Notebook Link](https://github.com/pskale/SPACEX/blob/master/EDA%20with%20Visualization.ipynb)

Click on the above link or copy paste following URL in browser :
https://github.com/pskale/SPACEX/blob/master/EDA%20with%20Visualization.ipynb

EDA with SQL-Queries fired to explore the following data

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was achieved.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the   names of the booster_versions which have carried the maximum payload mass
- List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

EDA with SQL– SpaceX API – GitHub URL

EDA with SQL Notebook Link

Click on the above link or copy paste following URL in browser :
https://github.com/pskale/SPACEX/blob/master/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

We will build Interactive Visual Analytics to explore the data in interactive and real time which will enable user to find the visual pattern related to success rate more effectively.

We will analyze Launch sites and proximities with Folium to mark the various Geographical Launch sites with Latitude/Longitude data and pin the total successful launches from particular site on the interactive map.

This will help user to choose the optimal launch site with more successful landings.

Build an Interactive Map with Folium-GitHub URL

[Build an Interactive Map with Folium Notebook Link]

Click on the above link or copy paste following URL in browser :
https://github.com/pskale/SPACEX/blob/master/lab_jupyter_launch
_site_location1.ipynb

Build a Dashboard with Plotly Dash

We will build Dashboard application with Python plotly Dash package.

The Dashboard is built with dropdown list,range slider for interacting with Pie chart and scatter point chart .

It will display the success percentage of landing on each site and overall success rate for all sites.

It will also display scatter chart for success rate for payload mass on each launch site.

Build a Dashboard with Plotly Dash- GitHub URL

Build a Dashbord with Plotly Dash python program Link

Click on the above link or copy paste following URL in browser :
https://github.com/pskale/SPACEX/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

Once we have all the insights from the collected and refined data, we will build machine learning pipeline to predict whether the first stage of Falcon 9 will land successfully.

The first stage will be pre-processing to standardize the data and the split the data in Train and Test sets.

We will train different machine learning models and perform Grid search for optimal hyperparameter value for best Accuracy.

We will determine the model with highest accuracy with test data set.

We will test models for Logistic regression, Support vector machine, Decision Tree Classifier and K-nearest neighbors.

Finally we will check the confusion matrix for all the models and find the Accuracy for all the models to find the best model for prediction.

Predictive Analysis (Classification)

Load Space X data set

data=pd.read_csv("dataset_part_2.csv")
X=pd.read_csv("dataset_part_3.csv")

Create Numpy arrayfrom column Class in data
Standardize data in X
Y = data["Class"].to_numpy()
transform = preprocessing.StandardScaler()
transform.fit(X)

train_test_split to split the data X and Y into training and test data
X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.2,random_state=2)

Create Logistic Regression and train the model
Calculate the Accuracy
Check the confusion Matrix

Create Support Vector Machine(SVM)and train the model
Calculate the Accuracy
Check the confusion Matrix

Create Decision Tree Classifier
And train the model
Calculate the Accuracy
Check the confusion Matrix

Create K Nearest Neighbors(KNN)
and train the model
Calculate the Accuracy
Check the confusion Matrix

Predictive Analysis (Classification)- GitHub URL

Predictive Analysis(classification) Notebook Link

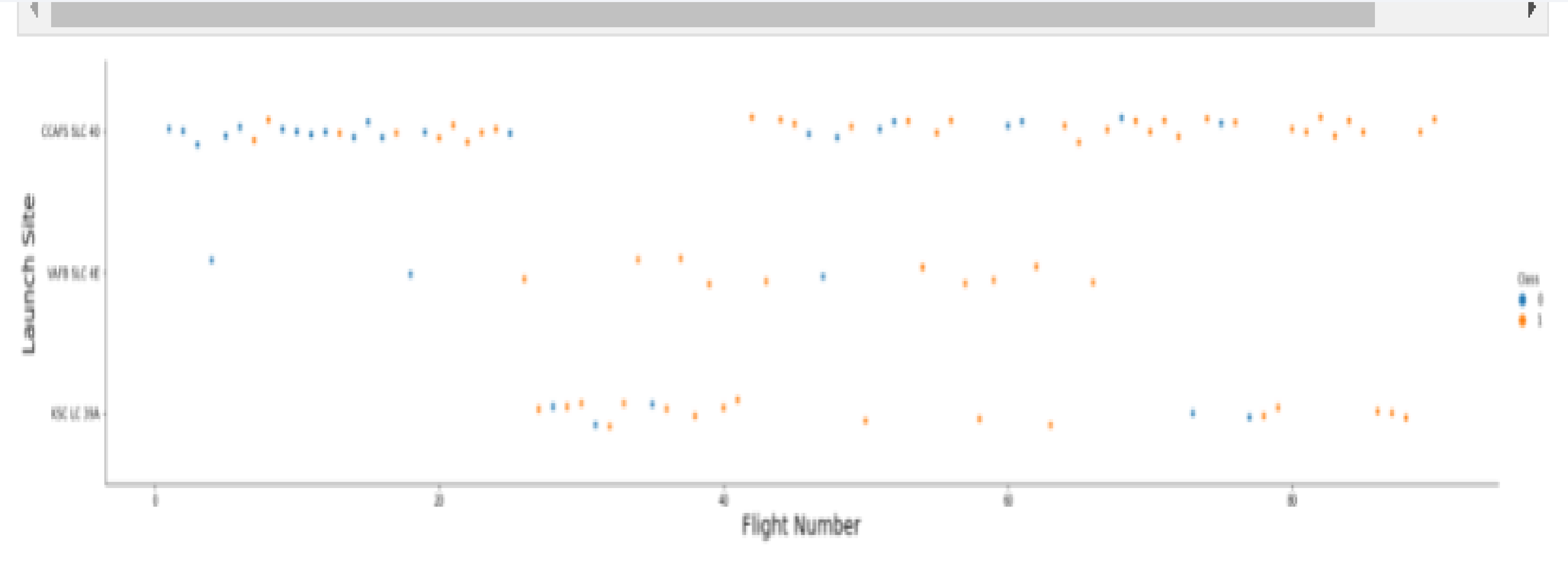
Click on the above link or copy paste following URL in browser :https://github.com/pskale/SPACEX/blob/master/Machine%20Learning%20Prediction2.ipynb

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

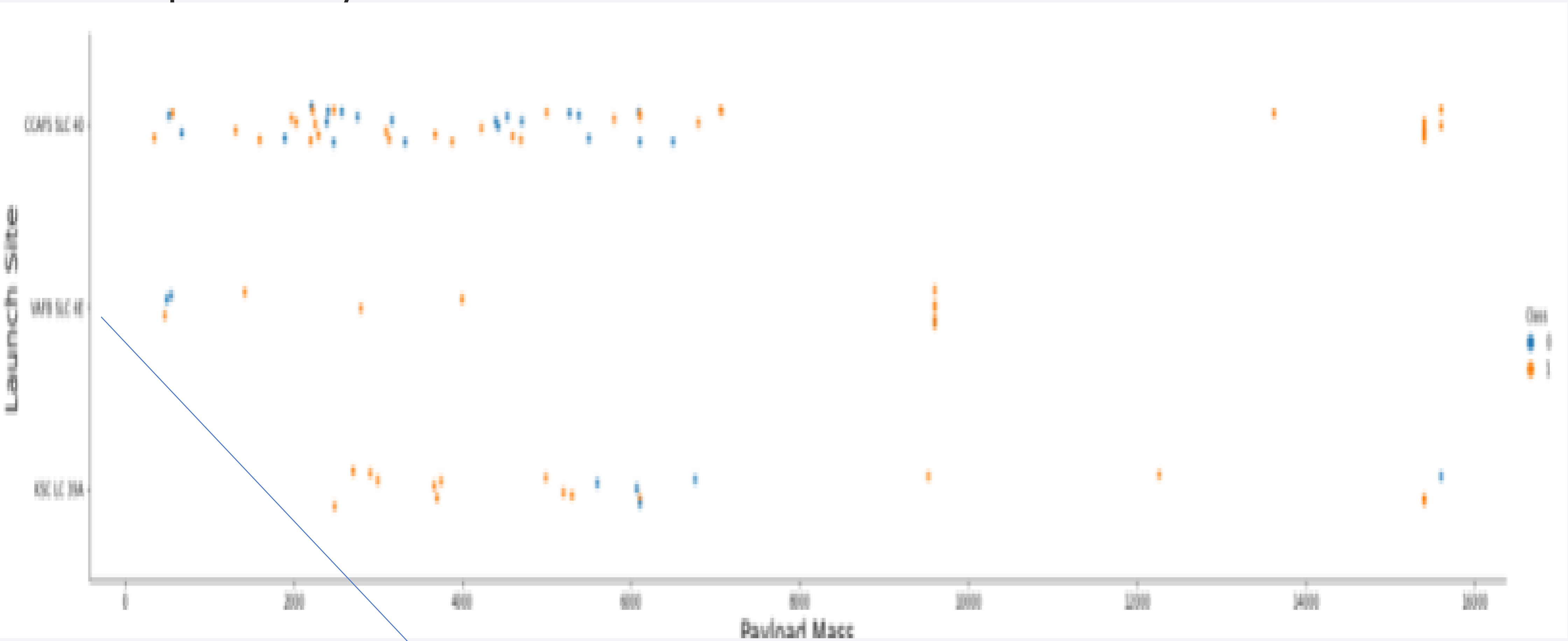
- Scatter plot of Flight Number vs. Launch Site



Launch site CCAFS SLC 40 has maximum number of launches and also maximum successful launches.
Higher flight number has more successful launches.

Payload vs. Launch Site

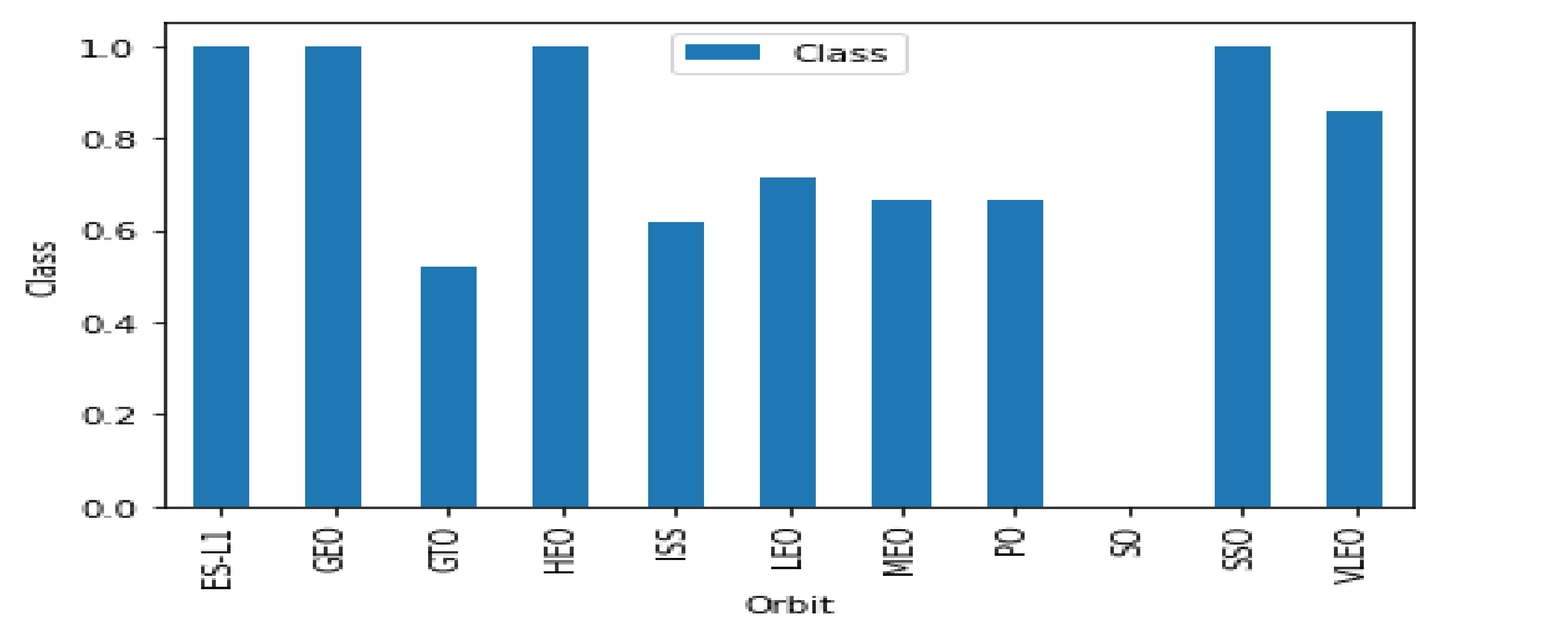
- Scatter plot of Payload vs. Launch Site



It is observed that for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type

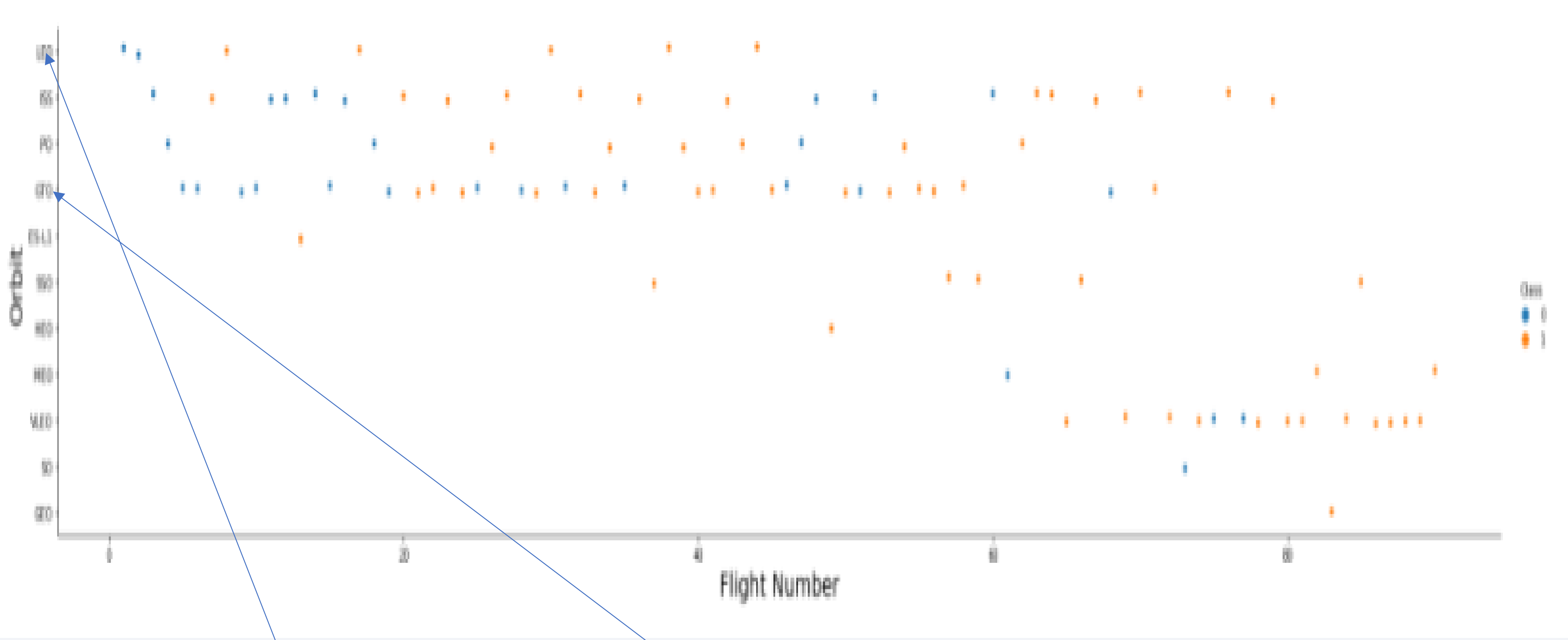
Bar chart for the success rate of each orbit type



It is observed that ES-L1,GEO,HEO and SSO orbit has very high success rate.

Flight Number vs. Orbit Type

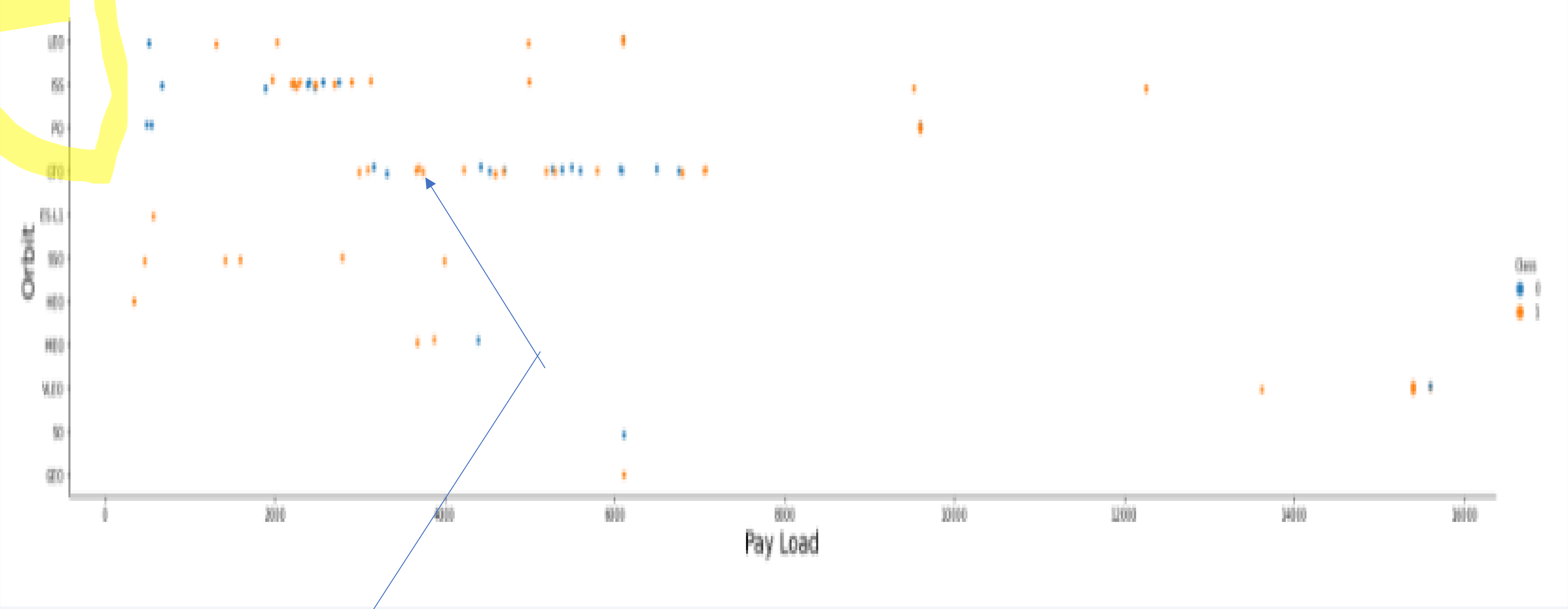
Scatter point of Flight number vs. Orbit type



It is observed that 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.

Payload vs. Orbit Type

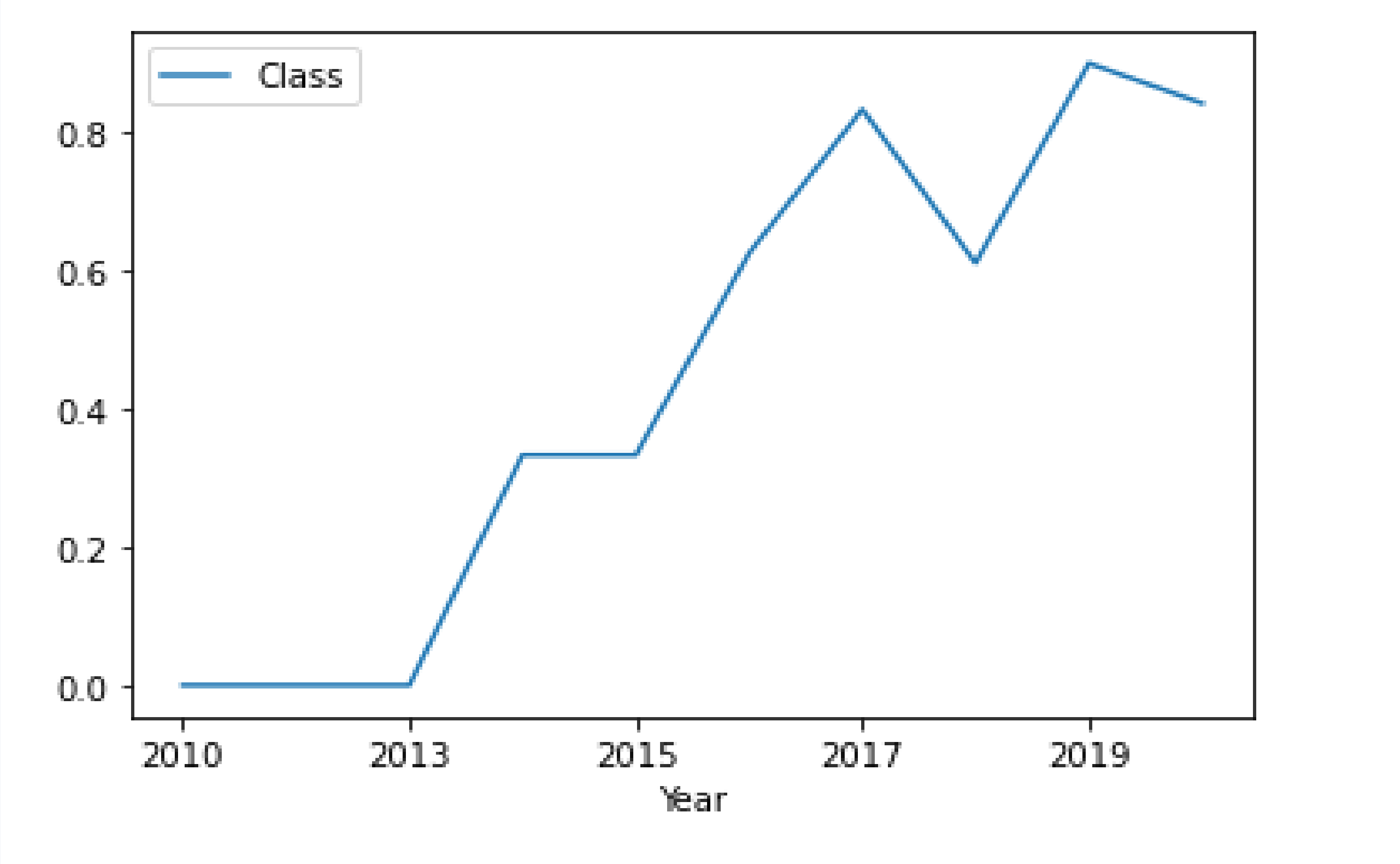
Scatter point of payload vs. orbit type



It is observed that With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS. However for GTO it can not be distinguished.

Launch Success Yearly Trend

Line chart of yearly average success rate



It is observed that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

Names of the unique launch sites

```
select  distinct LAUNCH_SITE from SPACEXDATASET
```

```
 * ibm_db_sa://ddd40799:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:
31505/bludb
Done.
```

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

There are four distinct Launch sites for space shuttle launches with Falcon 9 rocket.

Launch Site Names Begin with 'CCA'

5 records where launch sites begin with `CCA`

```
select * from SPACEXDATASET where LAUNCH_SITE like 'CCA%' LIMIT 5
```

```
 * ibm_db_sa://ddd40799:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/bludb
Done.
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	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
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

It is observed that for all 5 records the outcome was failure or No attempt

Total Payload Mass

Total payload carried by boosters from NASA

```
sql
lect customer,SUM(payload_mass__kg_) Total_Payload from SPACEXDATASET group by customer having customer like 'NAS
```

```
 * ibm_db_sa://ddd40799:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:
31505/bludb
Done.
```

customer	total_payload
NASA (CRS)	45596

It is observed that Total Payload carried by booster from NASA is 45596 Kg.

Average Payload Mass by F9 v1.1

Average payload mass carried by booster version F9 v1.1

```
%%sql
select AVG(payload_mass__kg_) AVERAGE_PAYLOAD_MASS from SPACEXDATASET where booster_version='F9 v1.1'
```

```
 * ibm_db_sa://ddd40799:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdoma
31505/bludb
Done.
```

average_payload_mass
2928

The Average Payload Mass carried by booster version F9 v1.1 is 2928 Kg

First Successful Ground Landing Date

Date of the first successful landing outcome on ground pad

```
%%sql
select min(date) First_succesful_LandingOutcome   from SPACEXDATASET where UPPER(landing__outcome) like '%SUCCES
```

```
 * ibm_db_sa://ddd40799:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/bludb
Done.
```

first_succesful_landingoutcome
2015-12-22

First successful ground landing date was on 22 December 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%%sql
select booster_version,landing__outcome,payload_mass__kg_ from SPACEXDATASET
where UPPER(landing__outcome) like '%SUCCESS%DRONE%' and payload_mass__kg_ > 4000 and payload_mass__kg_ <6000
```

```
 * ibm_db_sa://ddd40799:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/bludb
Done.
```

booster_version	landing__outcome	payload_mass__kg_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

List of Boosters which successfully landed for payload mass between 4000 and 6000

Total Number of Successful and Failure Mission Outcomes

Total number of successful and failure mission outcomes

```
%%sql
select mission_outcome,count(mission_outcome) Total_Mission_Outcomes from SPACEXDATASET group by mission_outcome
```

```
 * ibm_db_sa://ddd40799:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/bludb
Done.
```

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

It is observed that more than 99% of missions are successful.

Boosters Carried Maximum Payload

Names of the booster which have carried the maximum payload mass

```
%%sql
select booster_version,payload_mass__kg_ from SPACEXDATASET
where payload_mass__kg_ = (select MAX(payload_mass__kg_) from SPACEXDATASET )
```

```
 * ibm_db_sa://ddd40799:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1c
31505/bludb
Done.
```

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

Shows various Booster versions with maximum payload mass of 15600Kg

2015 Launch Records

List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%%sql
select DATE,landing__outcome,booster_version,launch_site from SPACEXDATASET
where UPPER(landing__outcome) like '%FAIL%DRONE%' and year(DATE) = 2015
```

```
 * ibm_db_sa://ddd40799:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/bludb
Done.
```

DATE	landing__outcome	booster_version	launch_site
2015-01-10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Shows list of failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

```
%%sql
select landing__outcome,count(landing__outcome) COUNT from SPACEXDATASET
where DATE between '2010-06-04' and '2017-03-20' group by landing__outcome
order by count(landing__outcome) desc
```

```
 * ibm_db_sa://ddd40799:***@ea286ace-86c7-4d5b-8580-3fbfa46b1c66.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31505/bludb
Done.
```

landing__outcome	COUNT
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

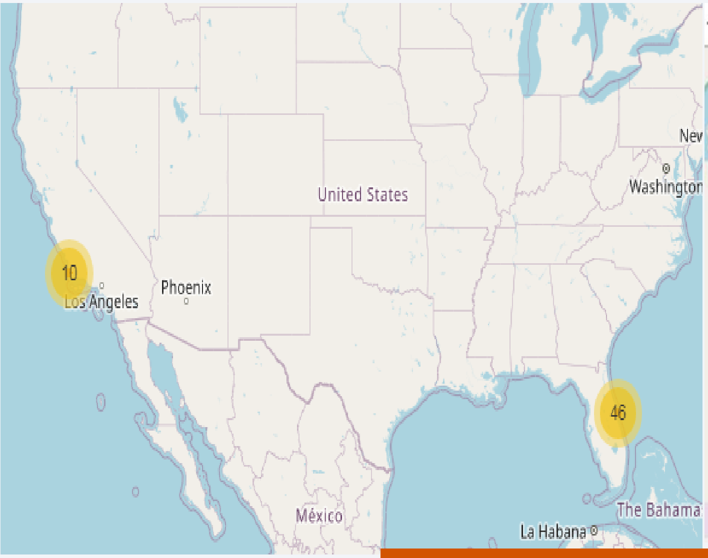
All Launch Sites location Markers on the Global Map

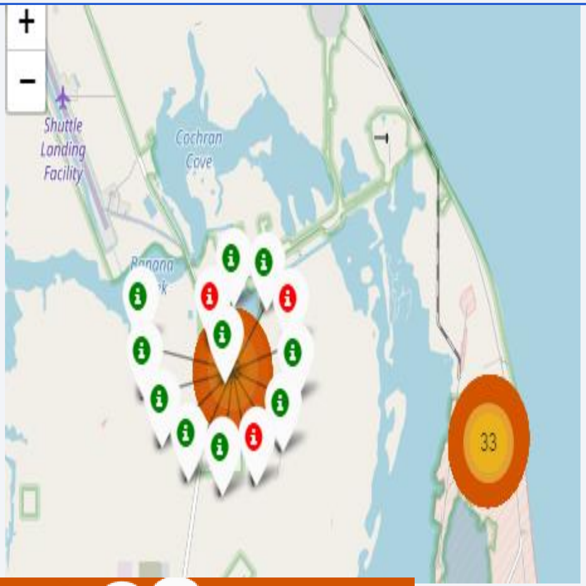
	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610746

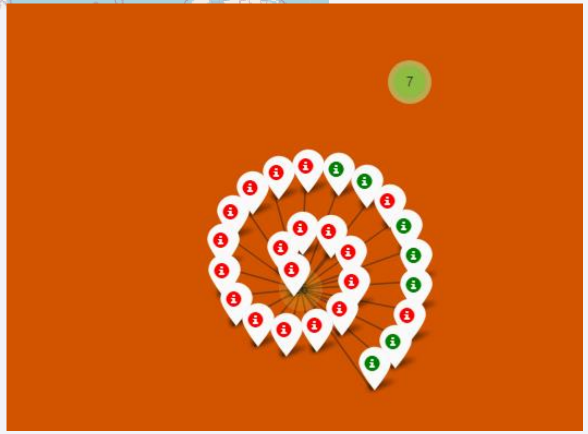
Launch Outcomes at all sites on Global map with color markers

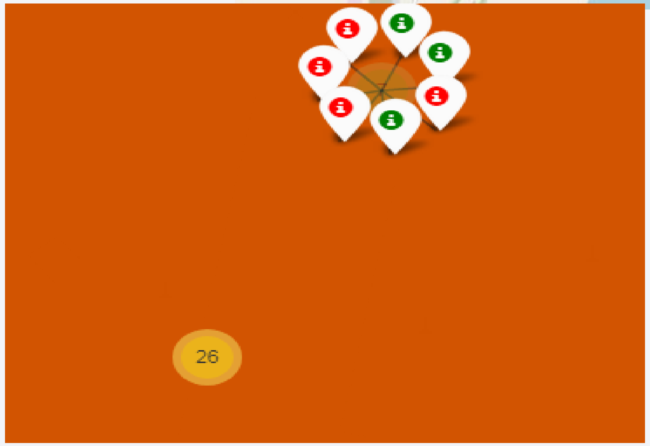
We can observe the Launch Outcomes at the launch site with Green for Success and Red for failure

Launch Outcomes at all sites on Global map with color markers



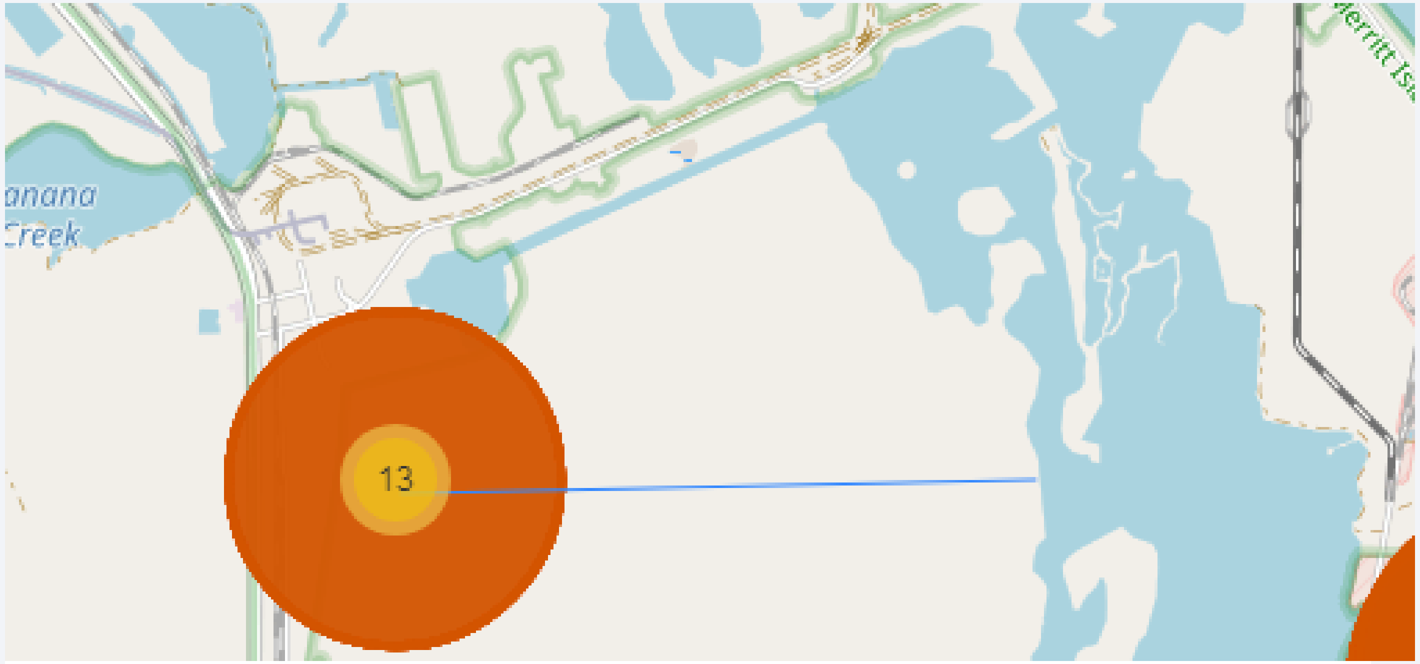






We can see the pattern of Success(Green) and Failure(Red) at launch sites at various locations in map with counts showing total number of outcomes at particular launch site

Proximities at Launch site




```
# find coordinate of the closet coastline
# e.g.,: Lat: 28.56367  Lon: -80.57163
distance_coastline = calculate_distance(28.5724,-80.6465,28.5732, -80.6072)
print(distance_coastline)
```

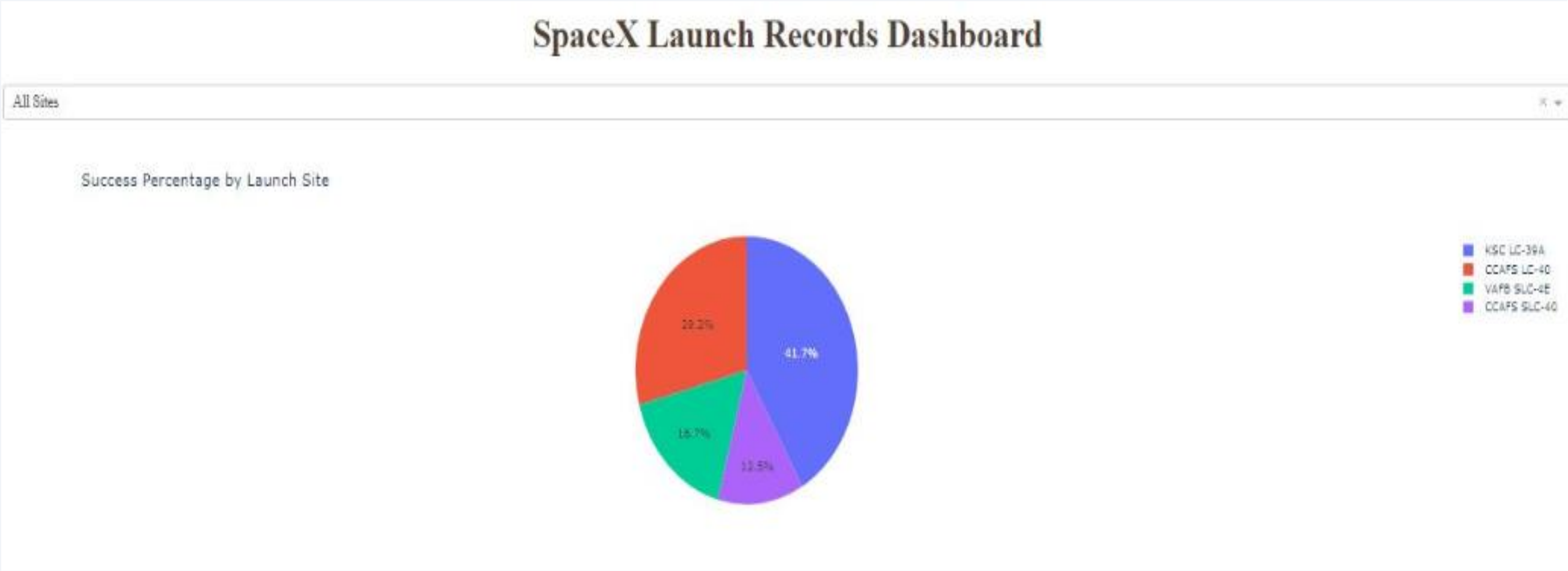
```
3.8399795496084446
```

Shows closest coastline from launch centre at 3.83 Km

Section 4

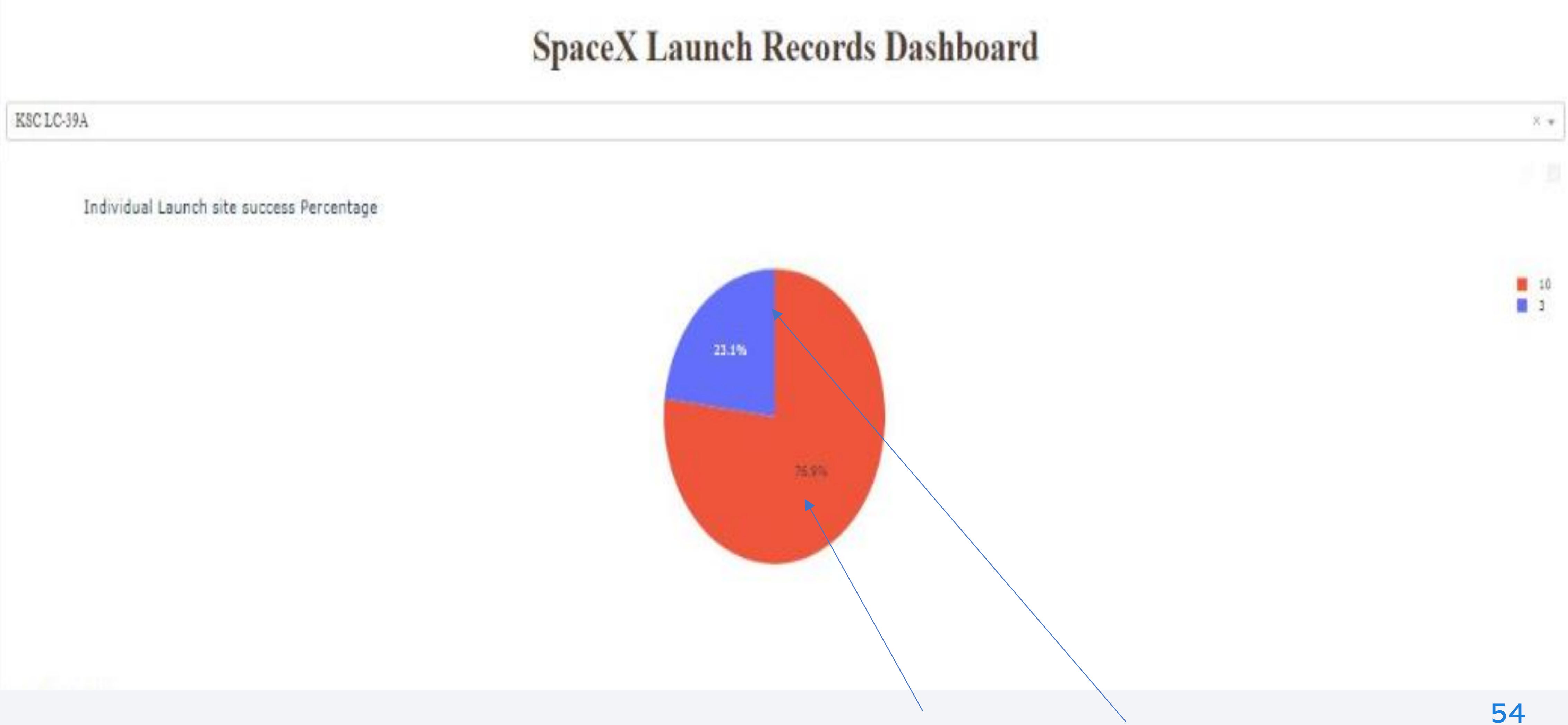
Build a Dashboard with Plotly Dash

Success Percentage by Launch site



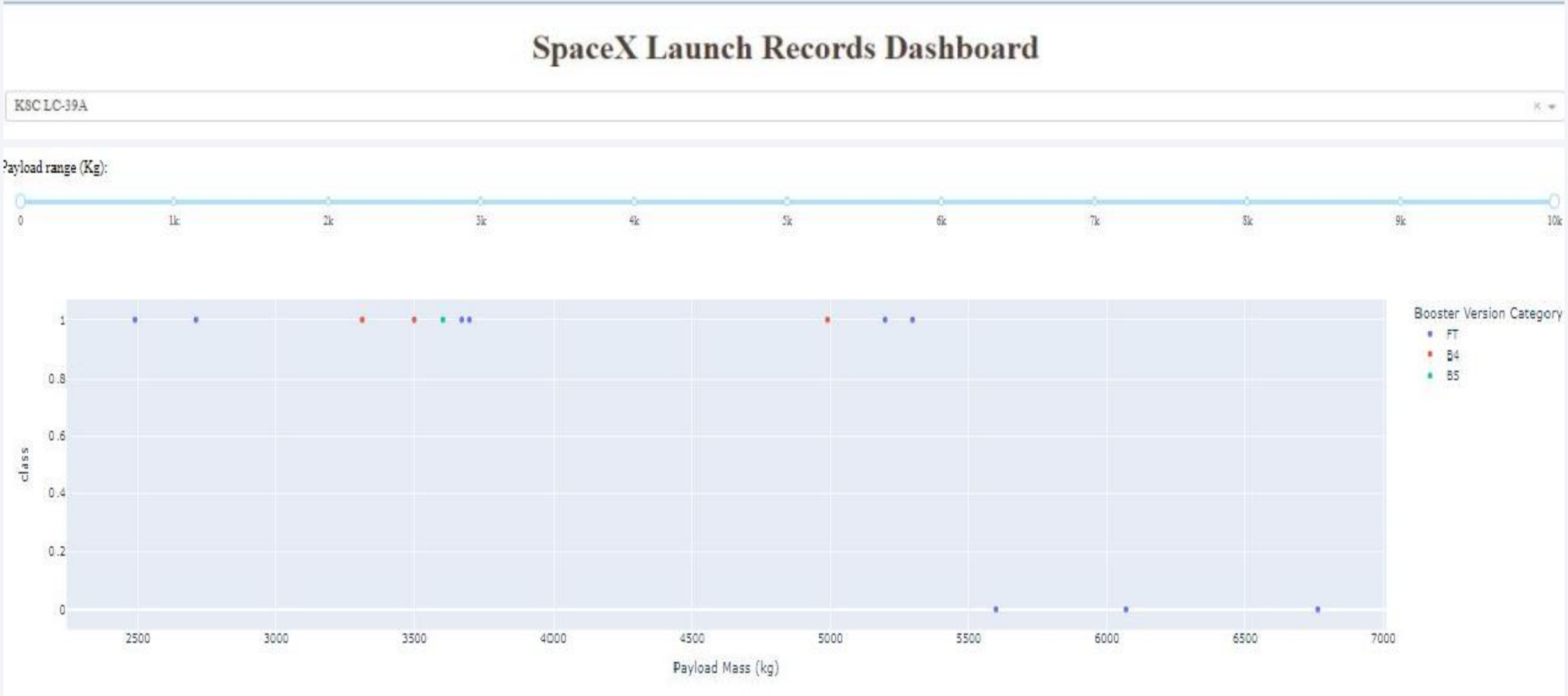
It is observed from the Pie chart that the site KSC-LC-39A has most successful landing record while CCAFS- SLC40 has least successful landings.

Launch site KSCLC-39A launch success ratio chart



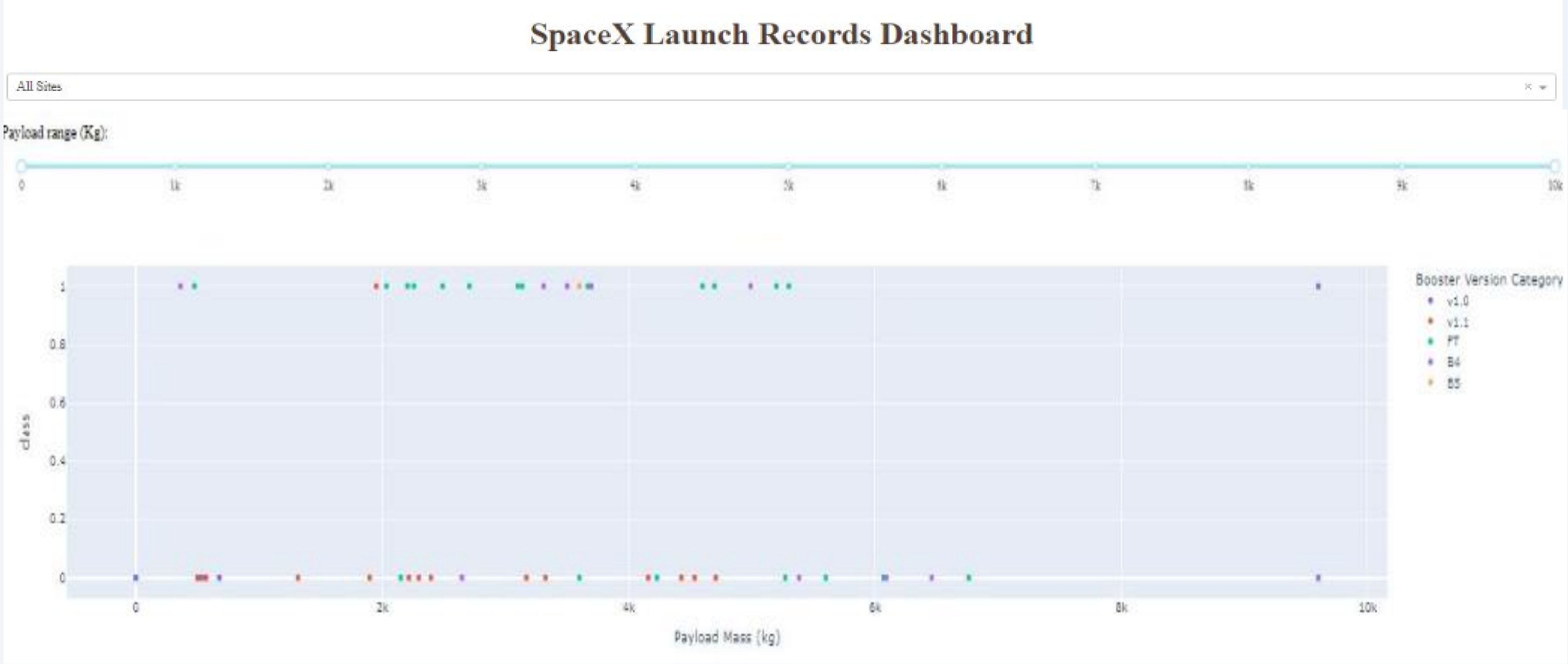
The site KSC LC-39A has highest success ratio of 76.9 % with 10 success(RED) and 3 failures landings(BLUE)

Payload v/s Launch outcome scatter plot-KSC LC-39A



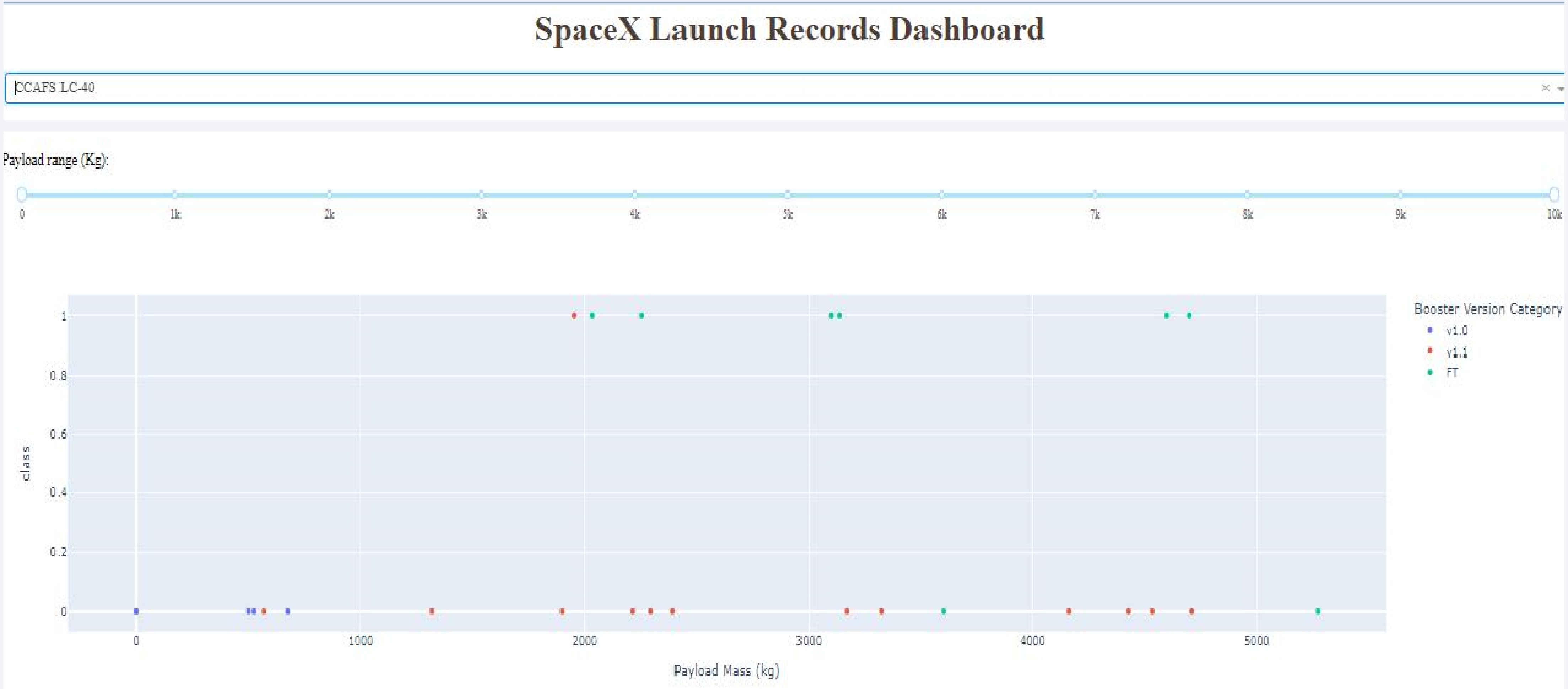
For site KSC LC-39A lower payload mass around 2500 to 5500 has more success rate and FT booster version has more successful outcomes than B4 and B5

Payload v/s Launch outcome scatter plot- All sites



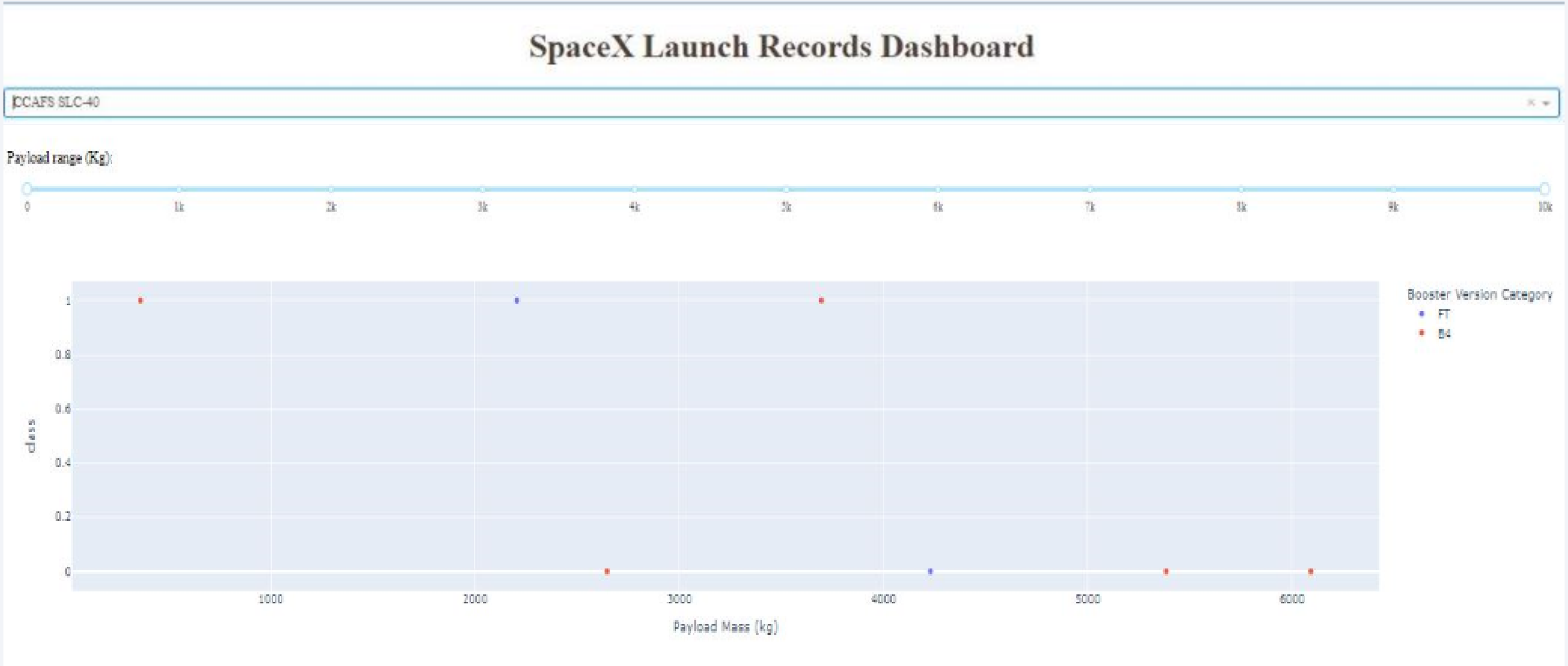
For all sites, across payload mass from 1K till 6 K FT booster version has more successful outcomes while V1.0 is successful in higher payload mass of 10K across all lauch sites.

Payload v/s Launch outcome scatter plot-CCAFS LC-40



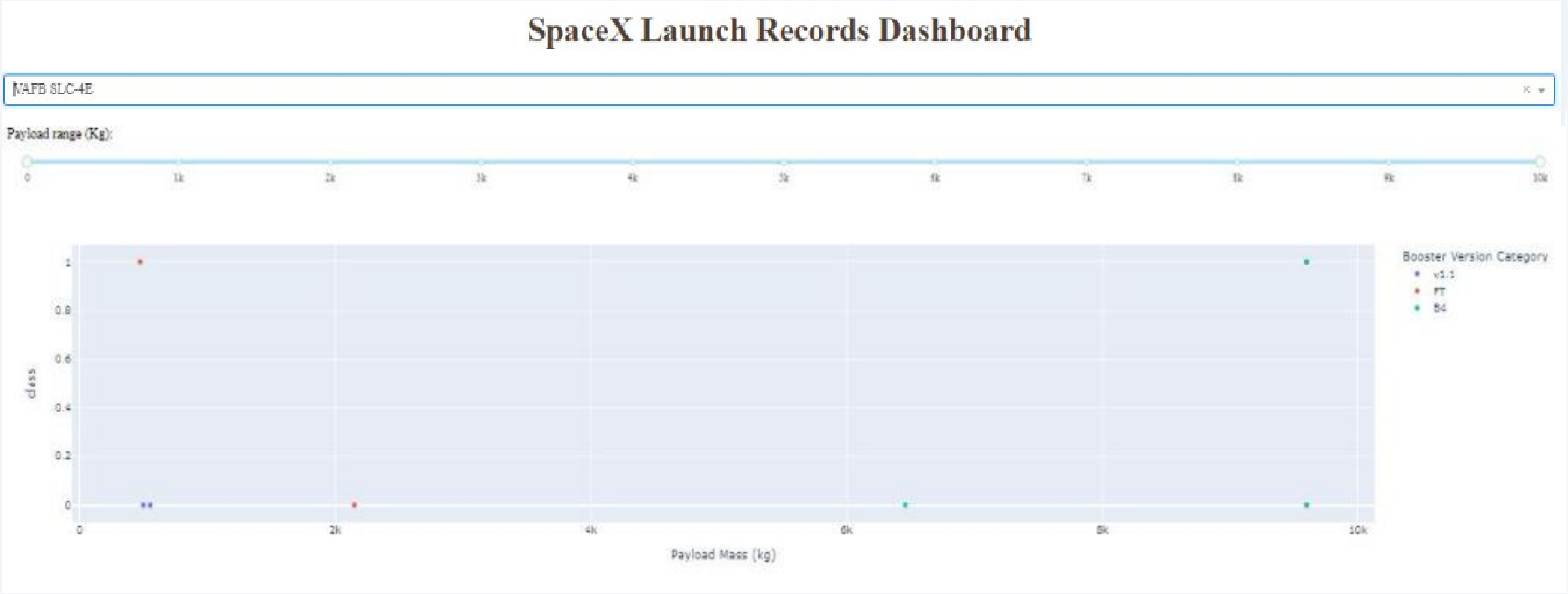
For site CCAFS LC-40, payload mass around 2000 to 5000 has more success rate and FT booster version has more successful outcomes than v1.0 and v1.1

Payload v/s Launch outcome scatter plot-CCAFS SLC-40



For site CCAFS SLC-40, lower payload mass below 4000 has more success and B4 booster version has more success rate than FT.

Payload v/s Launch outcome scatter plot-VABF SLC-4E



For site VABF SLC-4E higher payload mass around 9K has better success with B4 booster version and lower payload mass around 1K has success with FT booster version.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

Accuracy of the model Logistic Regression : 0.8333

```
parameters ={"C":[0.01,0.1,1],'penalty':['l2'], 'solver':['lbfgs']}# l1 lasso l2 ridge
lr=LogisticRegression()
logreg_cv=GridSearchCV(lr,parameters,cv=10,iid=None)
logreg_cv.fit(X_train,Y_train)
```

Calculate the accuracy on the test data using the method `score` :

```
BestLR=logreg_cv.best_estimator_
print("accuracy :",BestLR.score(X_test,Y_test))
```

```
accuracy : 0.8333333333333334
```

Accuracy of the model Support Vector Machine(SVM) : 0.8333

```
svm_cv = svm.SVC(kernel='linear')
```

```
#svm_cv=GridSearchCV(svm,parameters,cv=10,iid=None)
svm_cv.fit(X_train,Y_train)
```

```
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma='auto_deprecated',
  kernel='linear', max_iter=-1, probability=False, random_state=None,
  shrinking=True, tol=0.001, verbose=False)
```

```
yhat = svm_cv.predict(X_test)
print("Accuracy:",metrics.accuracy_score(Y_test, yhat))
```

```
('Accuracy:', 0.8333333333333334)
```

Classification Accuracy

Accuracy of the model Decision Tree : 0.7222

```
tree_cv=GridSearchCV(tree,parameters,cv=10,iid=None)
tree_cv.fit(X_train,Y_train)
```

```
BestTree=tree_cv.best_estimator_
print("accuracy :",BestTree.score(X_test,Y_test))
```

```
('accuracy :', 0.7222222222222222)
```

Click to add text

Accuracy of the model K Nearest Neighbors(KNN) : 0.5555

```
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()
```

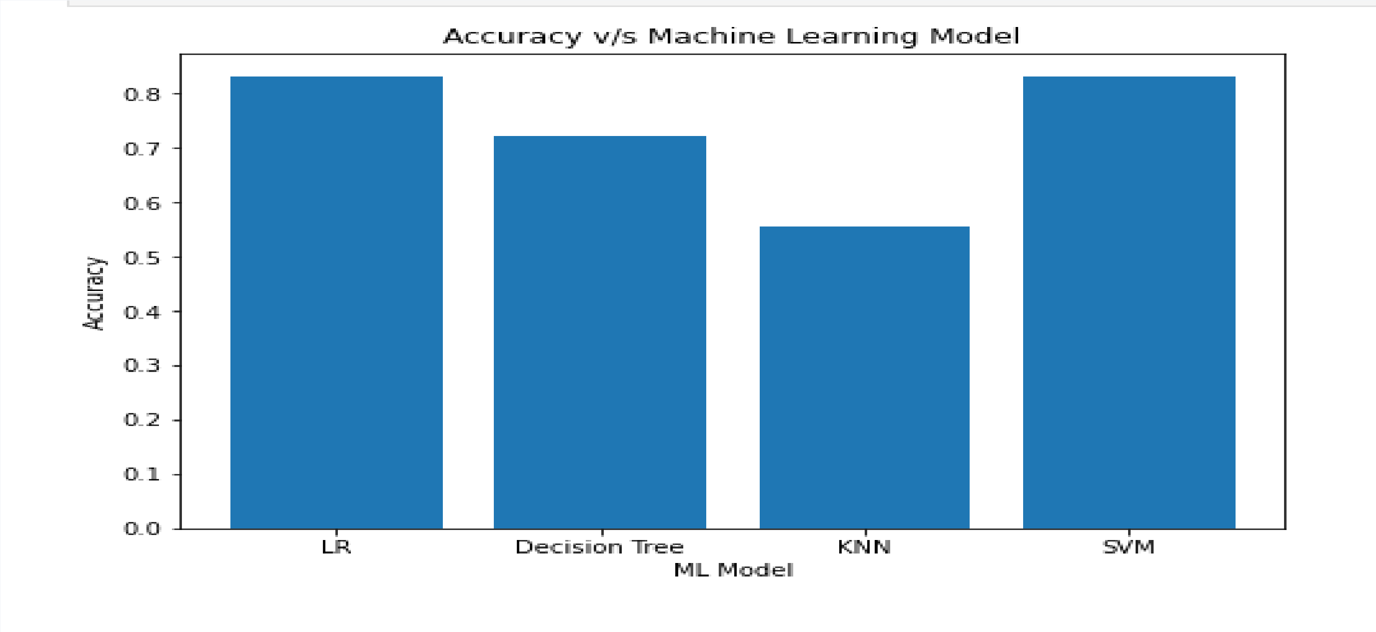
```
knn_cv=GridSearchCV(KNN,parameters,cv=10,iid=None)
knn_cv.fit(X_train,Y_train)
```

```
BestKNN=knn_cv.best_estimator_
print("accuracy :",BestKNN.score(X_test,Y_test))
```

```
('accuracy :', 0.5555555555555556)
```

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart



- It is evident from the bar chart that Linear regression/SVM model has highest accuracy over Decision Tree,KNN models.

Confusion Matrix

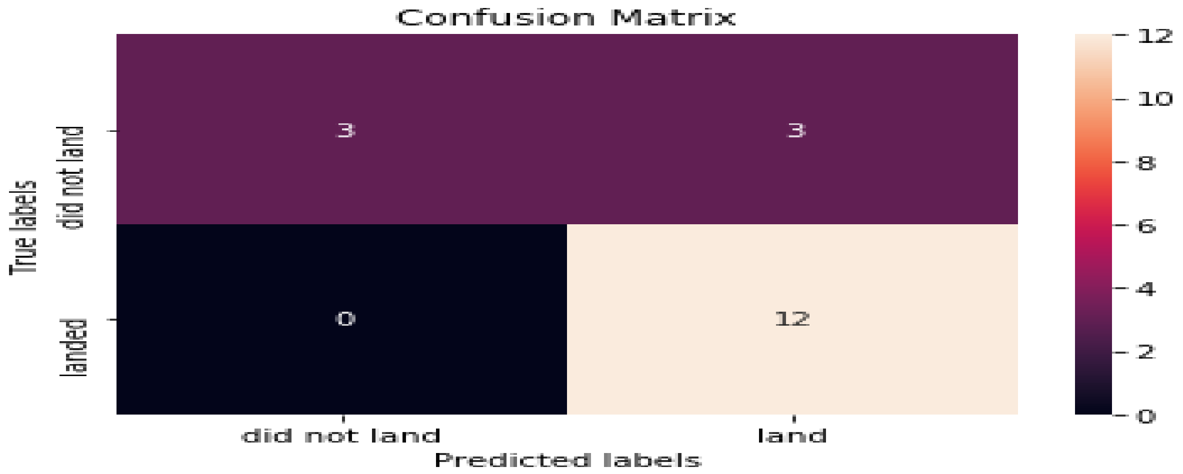
- Confusion matrix of the best performing Logistic Regression model

```
[24]: BestLR=logreg_cv.best_estimator_
      print("accuracy :",BestLR.score(X_test,Y_test))

      ('accuracy :', 0.8333333333333334)
```

Lets look at the confusion matrix:

```
[25]: yhat=logreg_cv.predict(X_test)
      plot_confusion_matrix(Y_test,yhat)
```



Confusion matrix shows 12 True positive and 6 true negatives which matches the predicted labels while only 3 shows false positives and 0 false negatives which is very good result.

Confusion Matrix

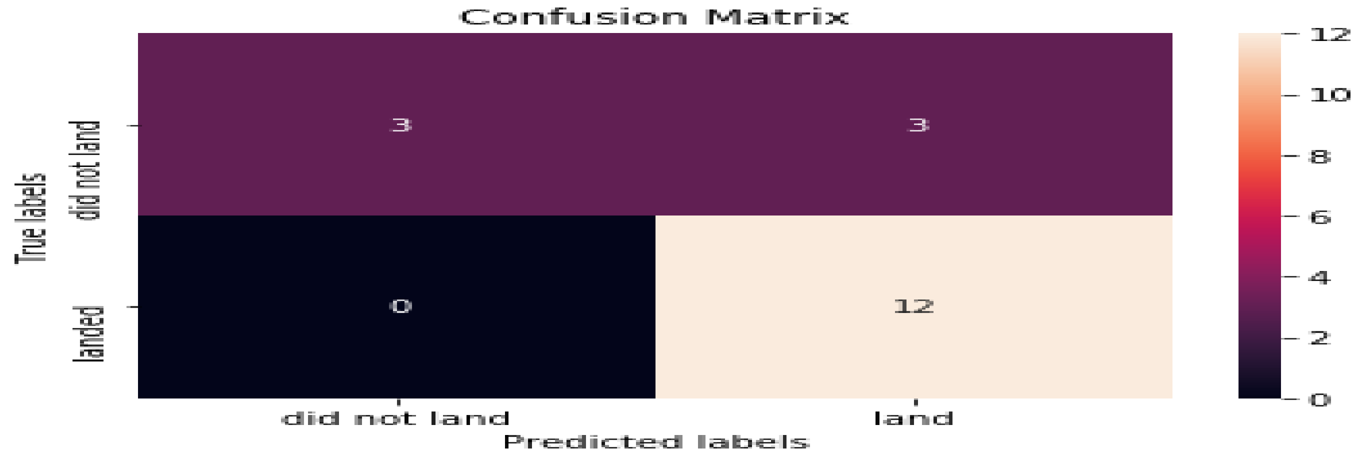
- Confusion matrix of the best performing SVM model

```
from sklearn import metrics
yhat = svm_cv.predict(X_test)
print("Accuracy:",metrics.accuracy_score(Y_test, yhat))
```

```
('Accuracy:', 0.8333333333333334)
```

We can plot the confusion matrix

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



Confusion matrix shows 12 True positive and 3 true negatives which matches the predicted labels while only 3 shows false positives and 0 false negatives which is very good result.

Conclusions

For the Data Science project for prediction of landing outcomes of Falcon 9 rocket for various launches ,we can conclude following points:

- There are numerous factors like Rocket Booster version ,Payload Mass,Orbit,Launch site which are crucial in prediction of successful landing of Falcon 9 after launch.
- ES-L1,GEO,HEO and SSO orbit has very high success rate of landing.
- The sucess rate of landings since 2013 kept increasing till 2020.
- Launch site KSC-LC-39A has most successful landing record while CCAFS- SLC40 has least successful landings.
- Booster Rocket version FT has most successful landings.
- The Machine learning model Logistic Regression has most accuracy for prediction of landing over Decision Tree,SVM and KNN models.

Appendix

Following is GitHub path for project Notebooks, Datasets

https://github.com/pskale/SPACEX/tree/master

Thank you!