KV

IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Karanvir Singh
16-05-2025

Outline

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

Executive Summary

- Summary of methodologies

SpaceX launch data gathered specifically from SpaceX REST API.

Wrangling data using an API, Sampling Data, and Dealing with Nulls.

- Summary of all results

Launch success rate since 2013 kept increasing till 2020.

With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS orbit.

Total Success launches by site is for KSC LC-39A

Introduction

- To predict if the Falcon 9 first stage will land successfully
- To determine if SpaceX will reuse the first stage
- Training a Machine Learning model and use public information to predict if SpaceX will reuse the first stage

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Making a get request to the SpaceX API
- Perform data wrangling
  - Dealing with Missing Values and focusing on only Falcon 9 launches
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Creating a column for the class and Standardizing the data
  - Splitting into training data and test data
  - Finding best Hyperparameter for SVM, Classification Trees and Logistic Regression

Data Collection

- SpaceX launch data gathered specifically from SpaceX REST API
- Targeting a specific endpoint of the API to get past launch data
- Another method used to obtain Falcon 9 launch data is using Web scraping
- By using the Python BeautifulSoup package to web scrape some HTML tables that contain valuable Falcon 9 launch records
- Parsing the data from those tables and convert them into a Pandas data frame for further visualization and analysis

Data Collection – SpaceX API

- SpaceX launch data is gathered from an API, specifically the SpaceX REST API.
- We worked with the endpoint api.spacexdata.com/v4/launches/past
- Using this URL to target a specific endpoint of the API to get past launch data.
- Performing a get request using the requests library to obtain the launch data, which we will use to get the data from the API.
- https://github.com/Karanvir399/SpaceX-Falcon-9-First-Stage-Landing-Prediction/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- Obtaining Falcon 9 Launch data using web scraping related Wiki pages
- Using the Python BeautifulSoup package to web scrape some HTML tables that contain valuable Falcon 9 launch records.
- Parsing the data from those tables and converting them into a Pandas data frame for further visualization and analysis.
- https://github.com/Karanvir399/SpaceX-Falcon-9-First-Stage-Landing-Prediction/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- Wrangling Data using an API, Sampling Data, and Dealing with Nulls to transform this raw data into a clean dataset which provides meaningful data
- Data to be stored in lists and to be used to create the dataset.
- Filtering/Sampling the data to remove Falcon 1 launches as we only need to analyze Falcon 9 launches.
- Dealing with null values in order to make the dataset viable for analysis
- https://github.com/Karanvir399/SpaceX-Falcon-9-First-Stage-Landing-Prediction/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Categorical plots: To visualize relationships, distributions, and comparisons across different categories.
- Scatter plot: To visualize the relationship between two numerical (continuous) variables.
- Bar Chart: To compare quantities or values across different categories.
- Line Chart: To show trends or changes over time (or another continuous variable).
- https://github.com/Karanvir399/SpaceX-Falcon-9-First-Stage-Landing-Prediction/blob/main/edadataviz.ipynb

EDA with SQL

- Listing unique launch sites
- Calculating total payload mass for NASA (CRS) missions
- Finding the earliest successful landing date
- Retrieving booster versions with successful drone ship landings carrying moderate payloads
- Categorizing mission outcomes and counting them
- Identifying booster and payload for the heaviest payload
- Counting landing outcomes over a time range
- https://github.com/Karanvir399/SpaceX-Falcon-9-First-Stage-Landing-Prediction/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Folium.Circle is used to add a highlighted circle area with a text label on a specific coordinate.
- Marker clusters are a good way to simplify a map containing many markers having the same coordinate.
- Lines is a standard way to visualize routes, paths, or connections between geographic coordinates.
- https://github.com/Karanvir399/SpaceX-Falcon-9-First-Stage-Landing-Prediction/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Added a Launch Site Drop-down Input Component
- Pie Chart to show total success launches by site
- Scatter plot showing correlation between payload and success for different sites
- Payload range slider to select the desired payloads
- https://github.com/Karanvir399/SpaceX-Falcon-9-First-Stage-Landing-Prediction/blob/main/spacex-dash-app_code.py

Predictive Analysis (Classification)

- Preprocessing to standardize our data
- Train_test_split, allowing us to split our data into training and testing data
- Training the model and performing Grid Search, allowing us to find the hyperparameters that allow a given algorithm to perform best
- Using the best hyperparameter values, determining the model with the best accuracy using the training data
- https://github.com/Karanvir399/SpaceX-Falcon-9-First-Stage-Landing-Prediction/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

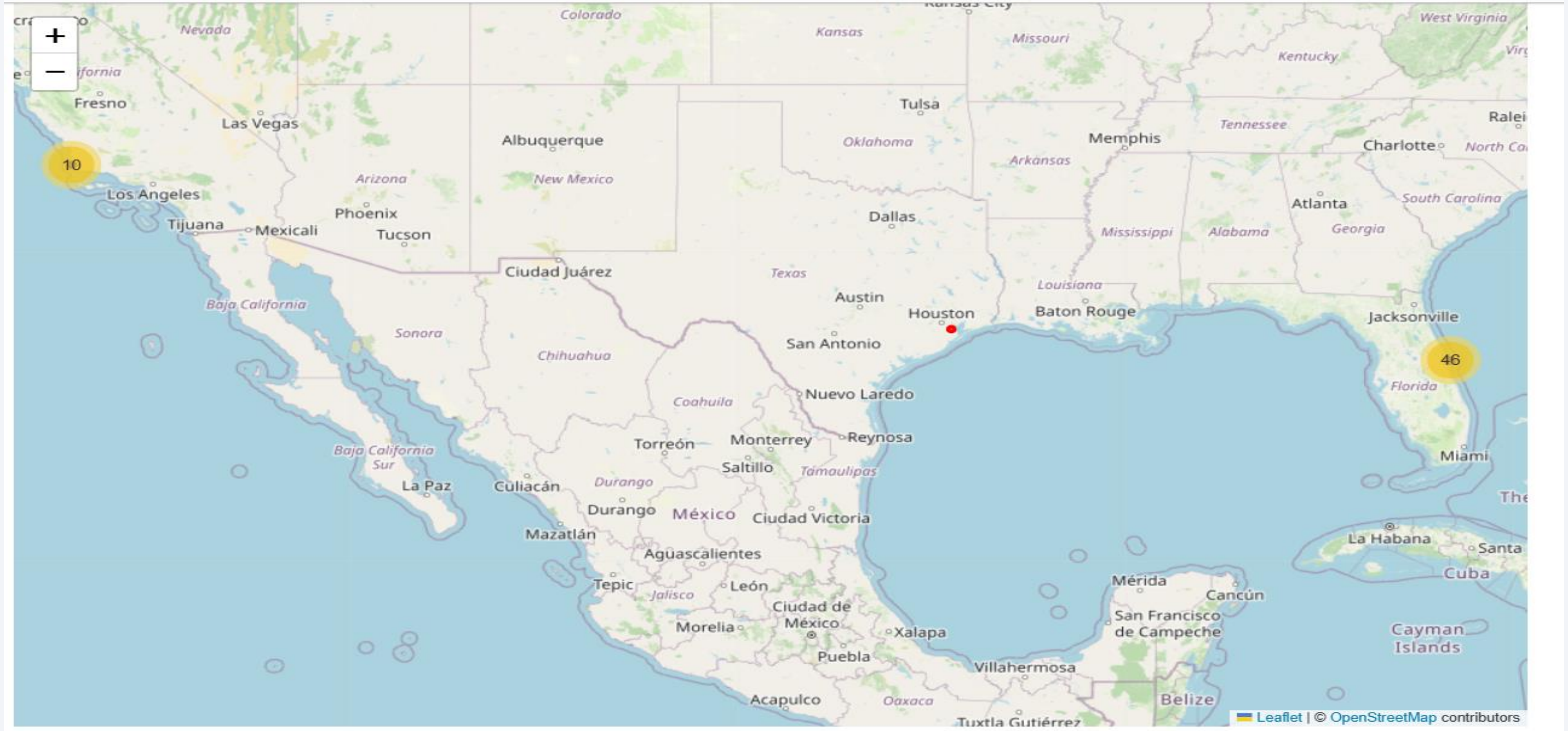
Results

- Exploratory data analysis results:

1. For the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).
2. Orbit type ES-L1, GEO, HEO, and SSO have the highest success rates.
3. With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
4. Launch sucess rate since 2013 kept increasing till 2020

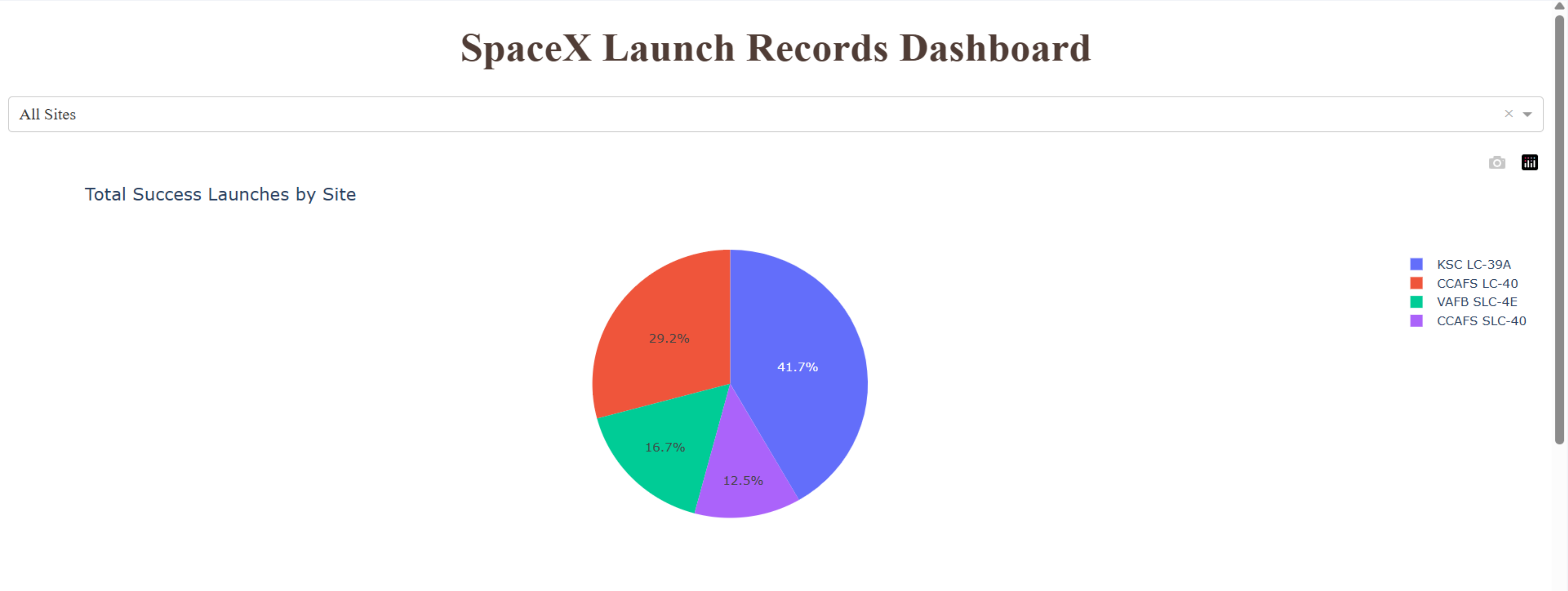
Results

- Interactive analytics demo in screenshots



Results

- Interactive analytics demo in screenshots



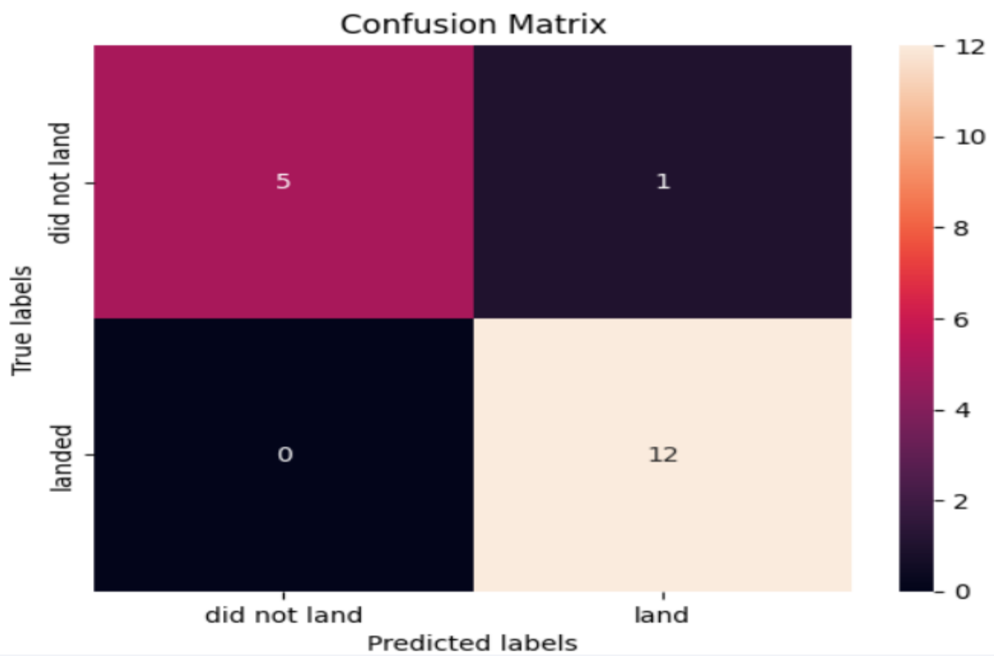
Results

- Predictive analysis results

In [22]:
```
test_accuracy = svm_cv.score(X_test, Y_test)
print("test accuracy", test_accuracy)
```
```
test accuracy 0.9444444444444444
```

We can plot the confusion matrix

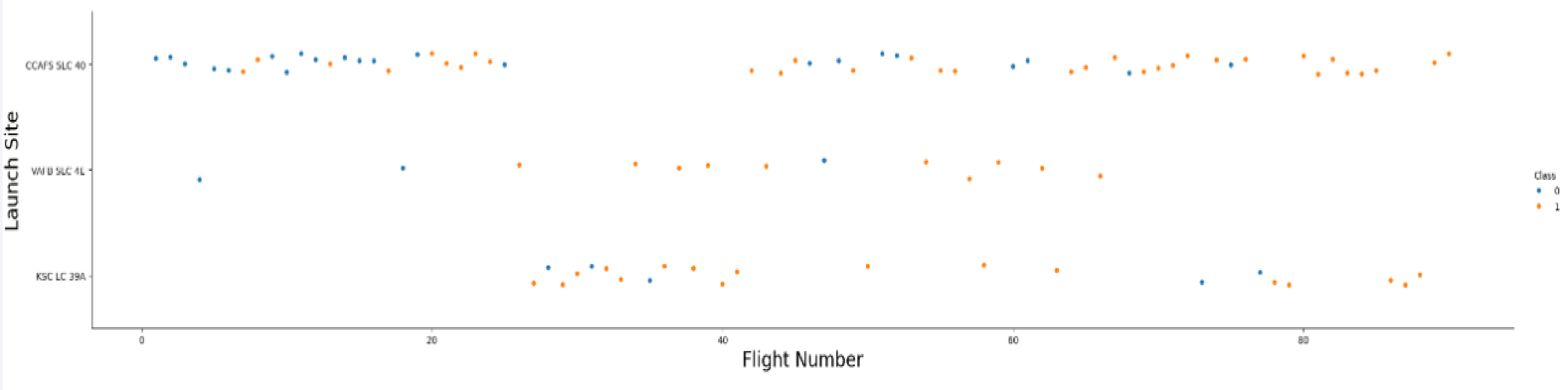
In [23]:
```
yhat=svm_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



Section 2

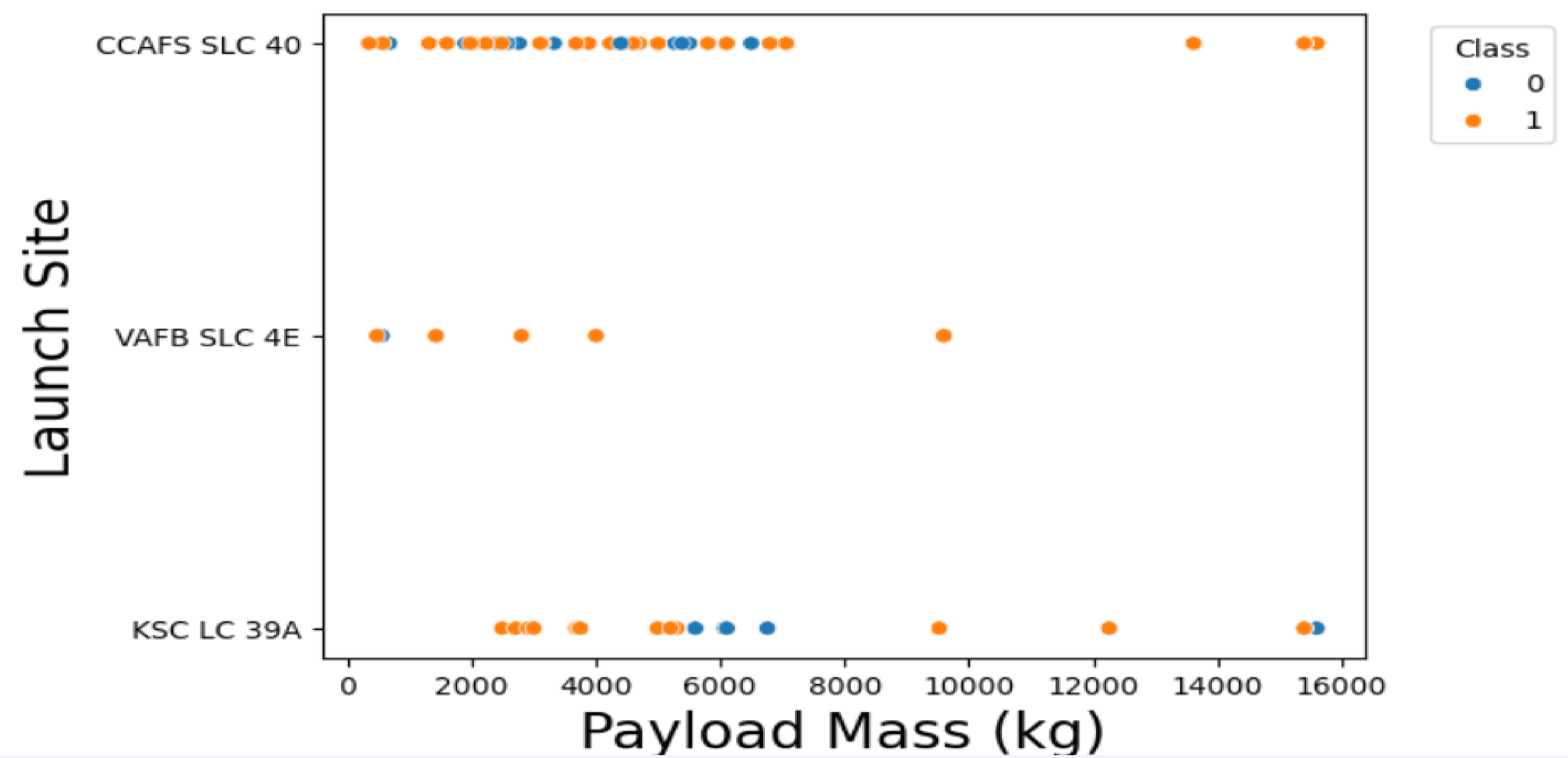
Insights drawn from EDA

Flight Number vs. Launch Site



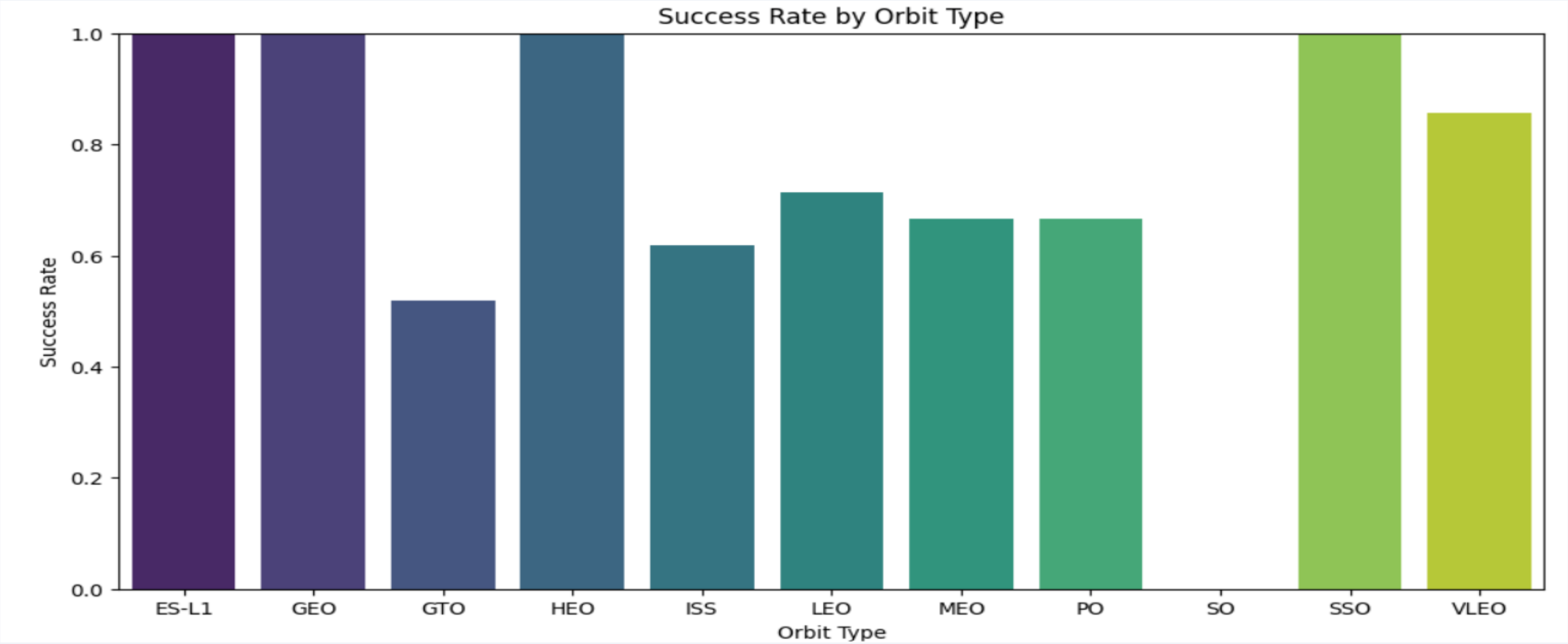
- This Categorical plot shows that VAFB SLC 4E and KSC LC 39A have a fairly good success rate.
- KSC LC 39A and CCAFS SLC 40 site have been used for recent launches.

Payload vs. Launch Site



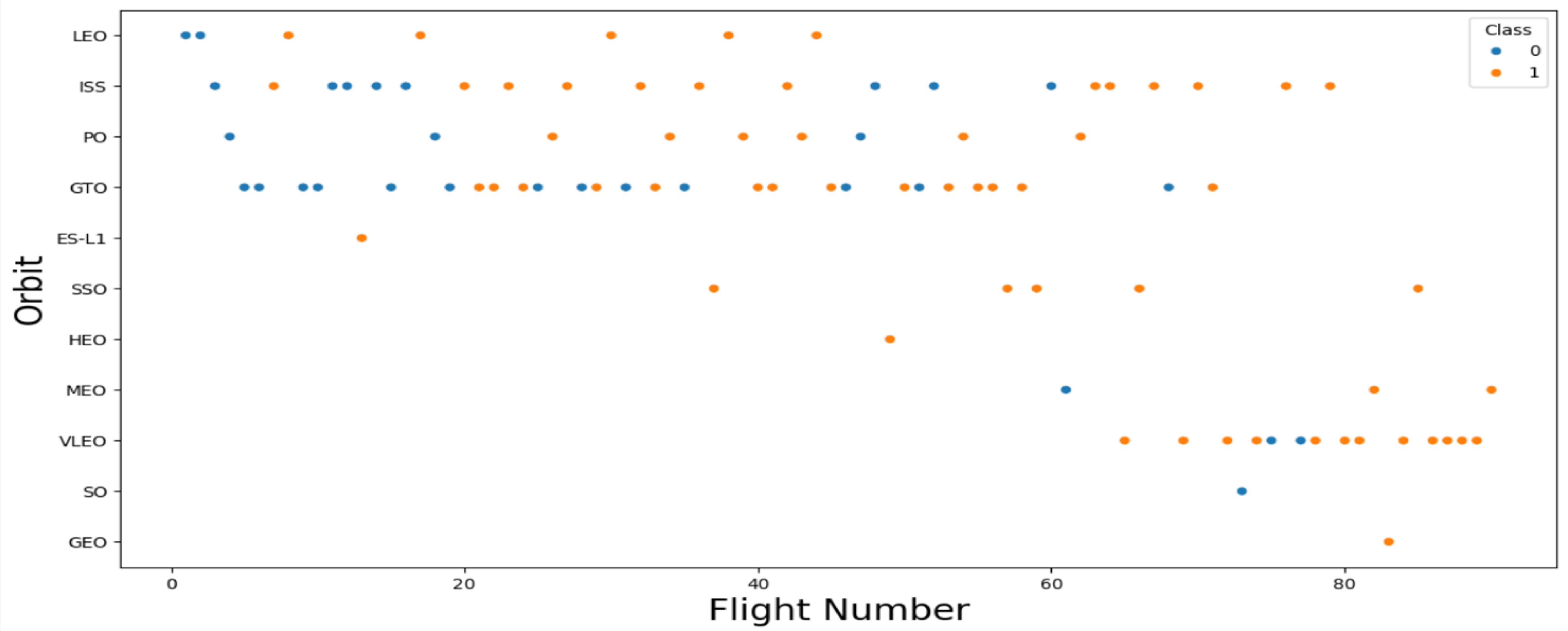
- This scatter plot shows that rocket launches with a payload above 8000 kg have a very good rate of success, i.e., 87.5%

Success Rate vs. Orbit Type



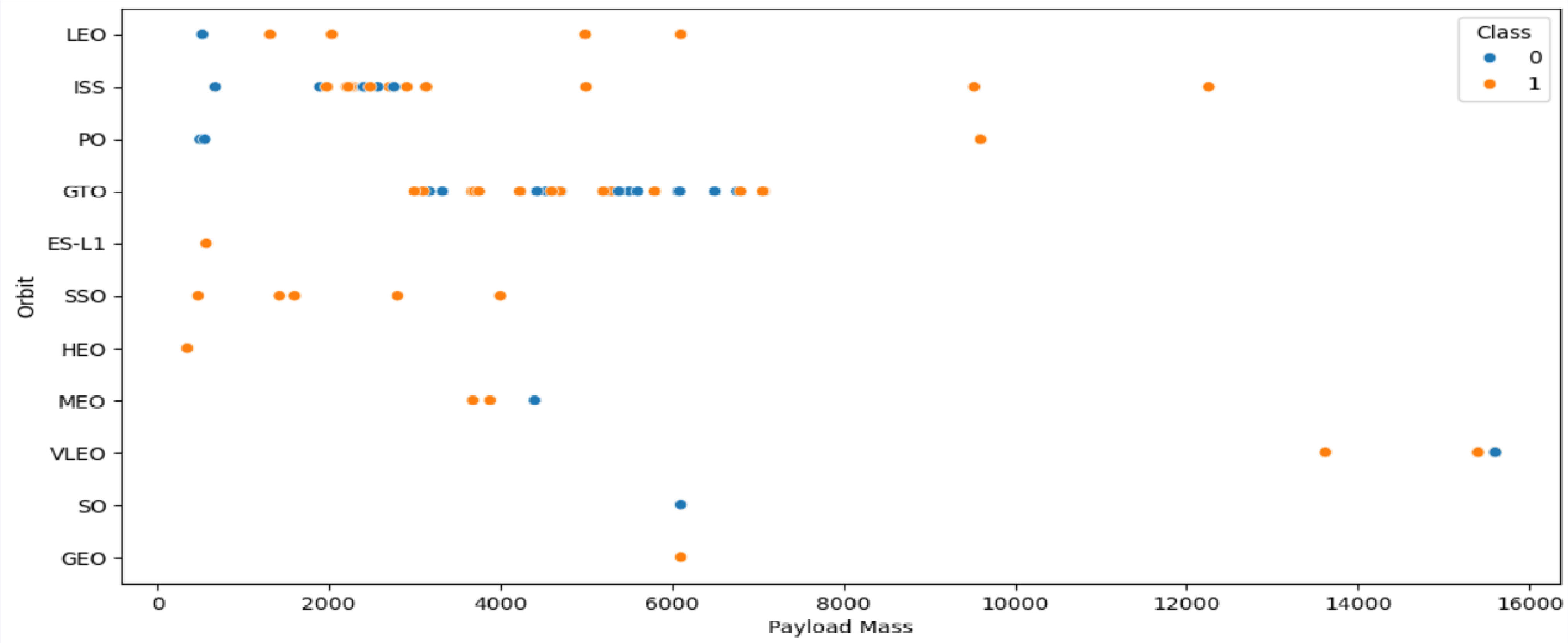
- By taking number of launches in account, launches to orbit VLEO have been quite successful and most have been recent launches.
- Launches to orbit SSO have been 100% success but the sample size is small.

Flight Number vs. Orbit Type



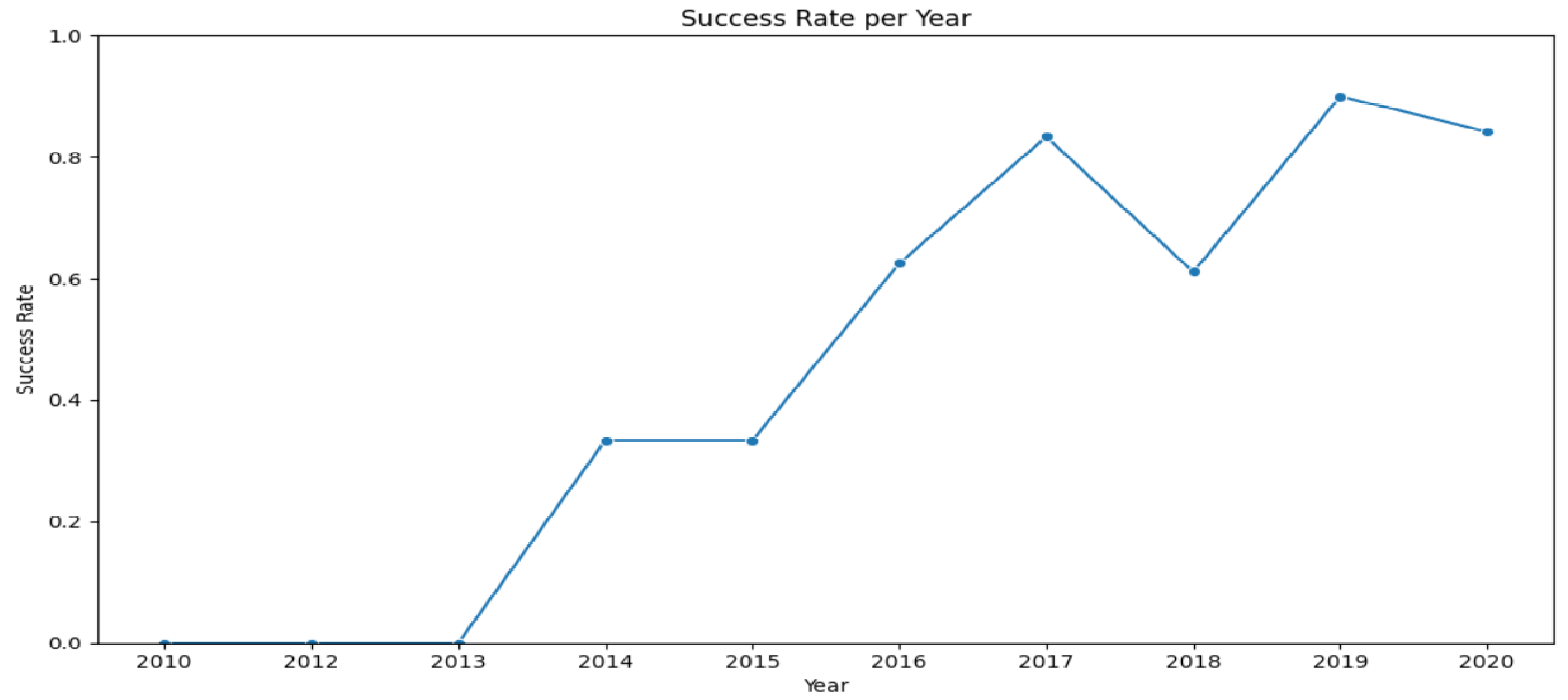
- Out of last 30 odd launches, most of the launches were made to Orbit ISS and VLEO
- Consistent launches were made to Orbit ISS and GTO in the past

Payload vs. Orbit Type



- Payload above 8000 kg have only been sent to orbit ISS, PO, and VLEO
- Although sample size is not big, but still, we can make an inference that success rate is around 83% when payload is above 8000 kg.

Launch Success Yearly Trend



- Launch success rate since 2013 kept increasing till 2020.

All Launch Site Names

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

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

Launch Site Names Begin with 'CCA'

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

Total Payload Mass

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

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Customer" LIKE 'NASA (CRS)';
```

```
 * sqlite:///my_data1.db
Done.
```

SUM("PAYLOAD_MASS__KG_")
45596

Average Payload Mass by F9 v1.1

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

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Booster_Version" LIKE 'F9 v1.1';
```

```
 * sqlite:///my_data1.db
Done.
```

AVG("PAYLOAD_MASS__KG_")
2928.4

First Successful Ground Landing Date

```
%sql SELECT MIN(Date) FROM SPACEXTABLE WHERE "Landing_Outcome" LIKE 'Success';
```

```
 * sqlite:///my_data1.db
Done.
```

MIN(Date)
2018-07-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

%sql SELECT CASE WHEN "Mission_Outcome" LIKE 'Success%' THEN 'Success' WHEN "Mission_Outcome" LIKE 'Failure%' THEN 'Failure' ELSE 'OTHER' END AS Outcome_Category, COUNT(*) AS Total FROM SPACEXTABLE GROUP BY Outcome_Category;

Outcome_Category	Total
Failure	1
Success	100

Boosters Carried Maximum Payload

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

2015 Launch Records

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

Month	Landing_Outcome	Launch_Site	Booster_Version
01	Failure (drone ship)	CCAFS LC-40	F9 v1.1 B1012
04	Failure (drone ship)	CCAFS LC-40	F9 v1.1 B1015

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

- Ranking 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

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


Section 3
Launch Sites Proximities Analysis

Location markers on a global map

Yellow circles represent Marker clusters simplifying a map containing many markers having the same coordinate.

Color-labeled launch outcomes on the map

Red color represents unsuccessful and Green color represents successful launch

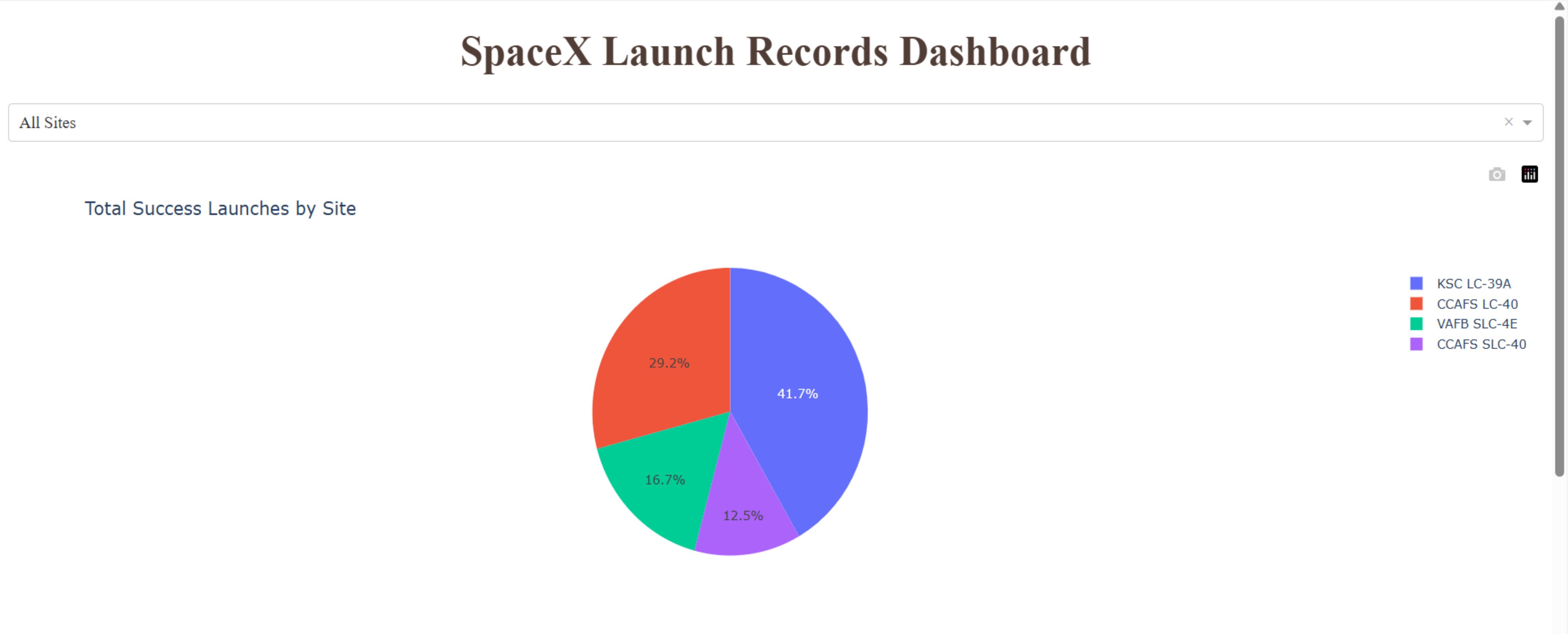
Distance between coastline point and launch site

Purple color line represents distance between the closest coastline point and the launch site

Section 4

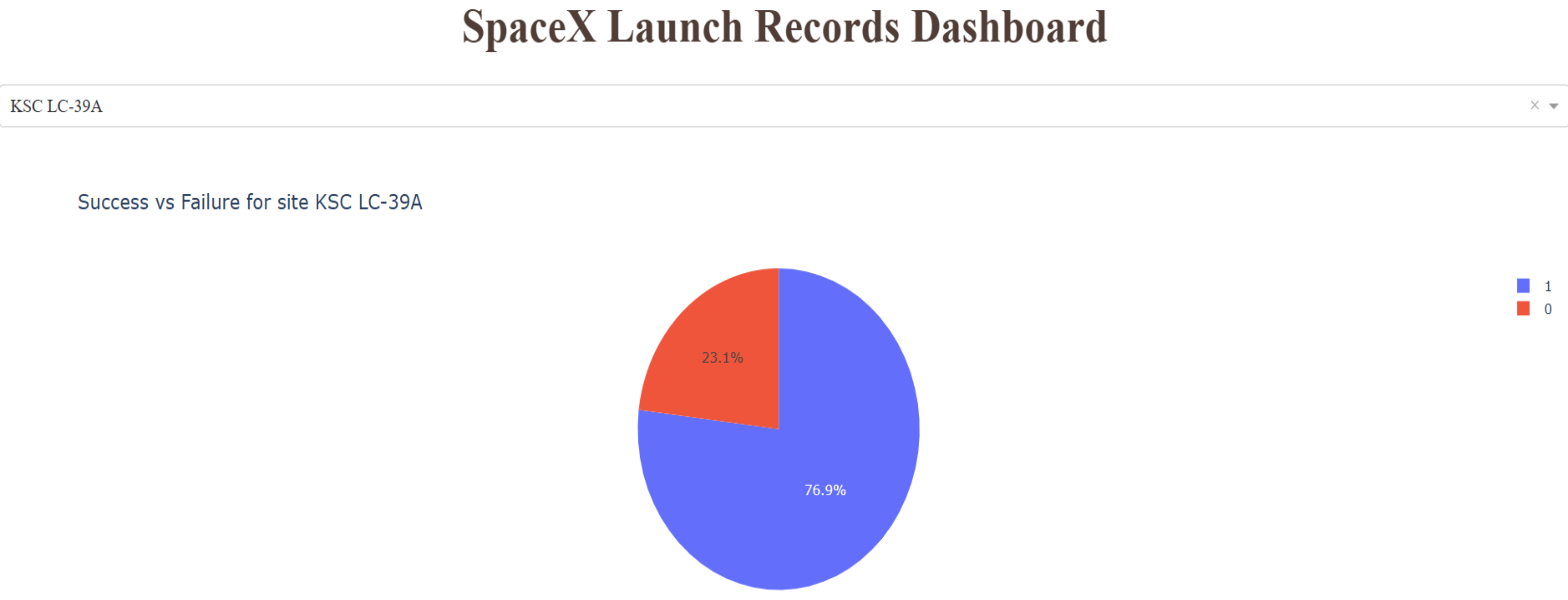
Build a Dashboard with Plotly Dash

Launch success count for all sites



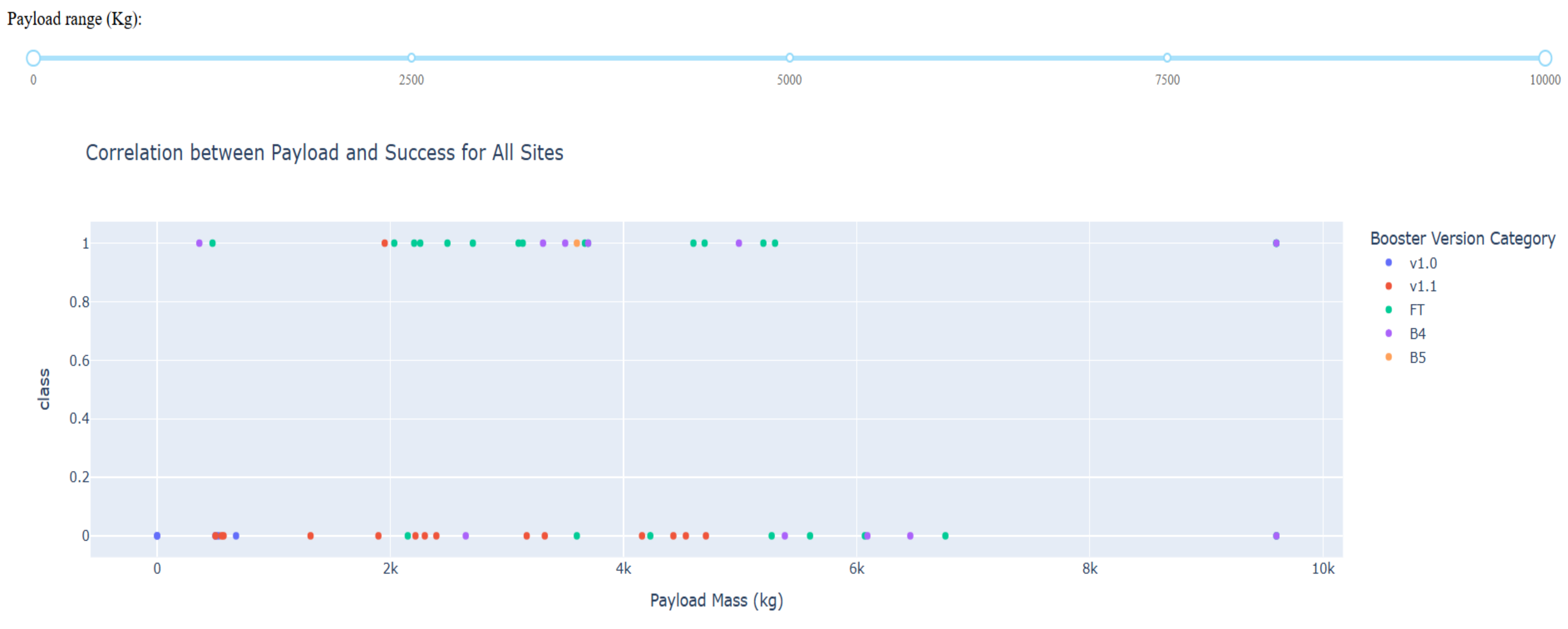
Nearly half of the successful launches have been made from KSC LC-39A

Launch site with highest launch success ratio



10 successful and 3 unsuccessful launches have been made from KSC LC-39A

Correlation between Payload and Success for All Sites



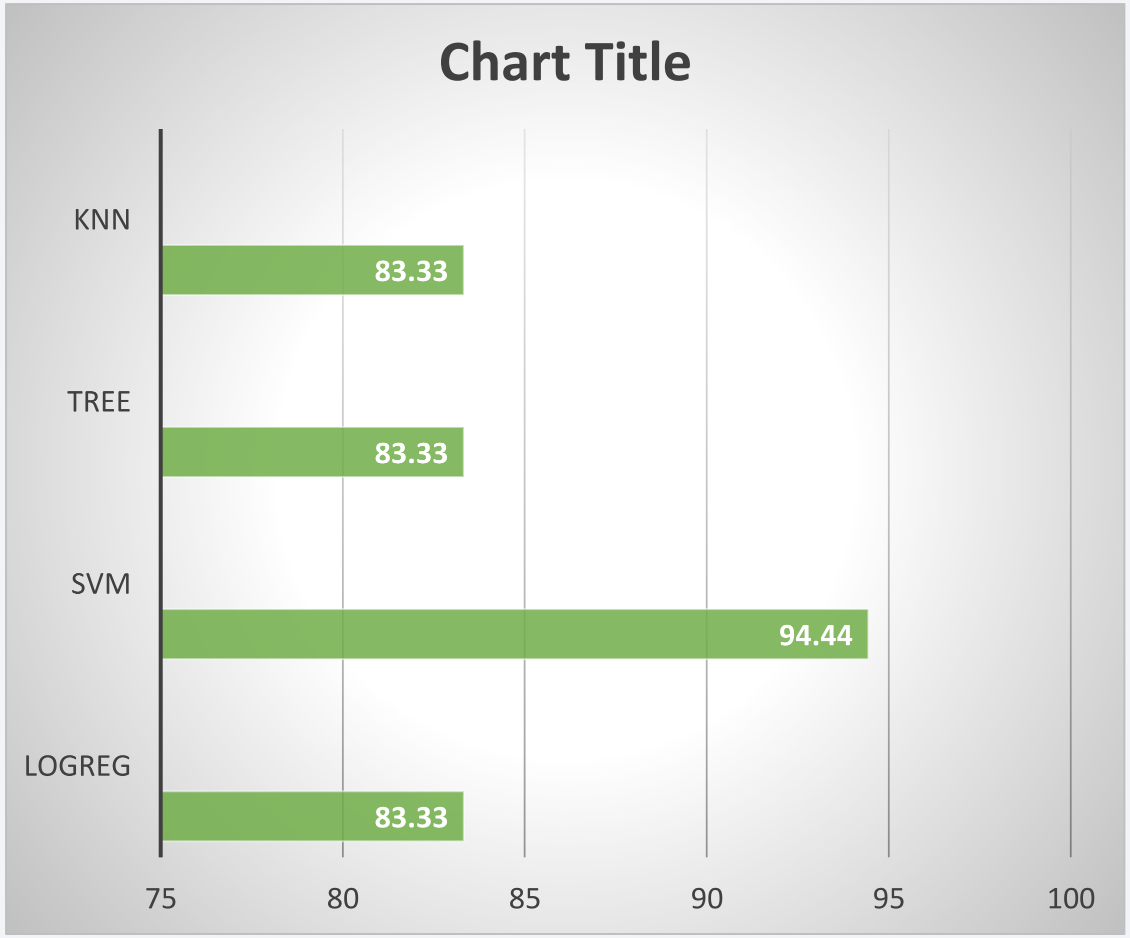
- In terms of number of launches, most of the successful launches have come when carried payload was between 2k and 4k

Section 5

Predictive Analysis (Classification)

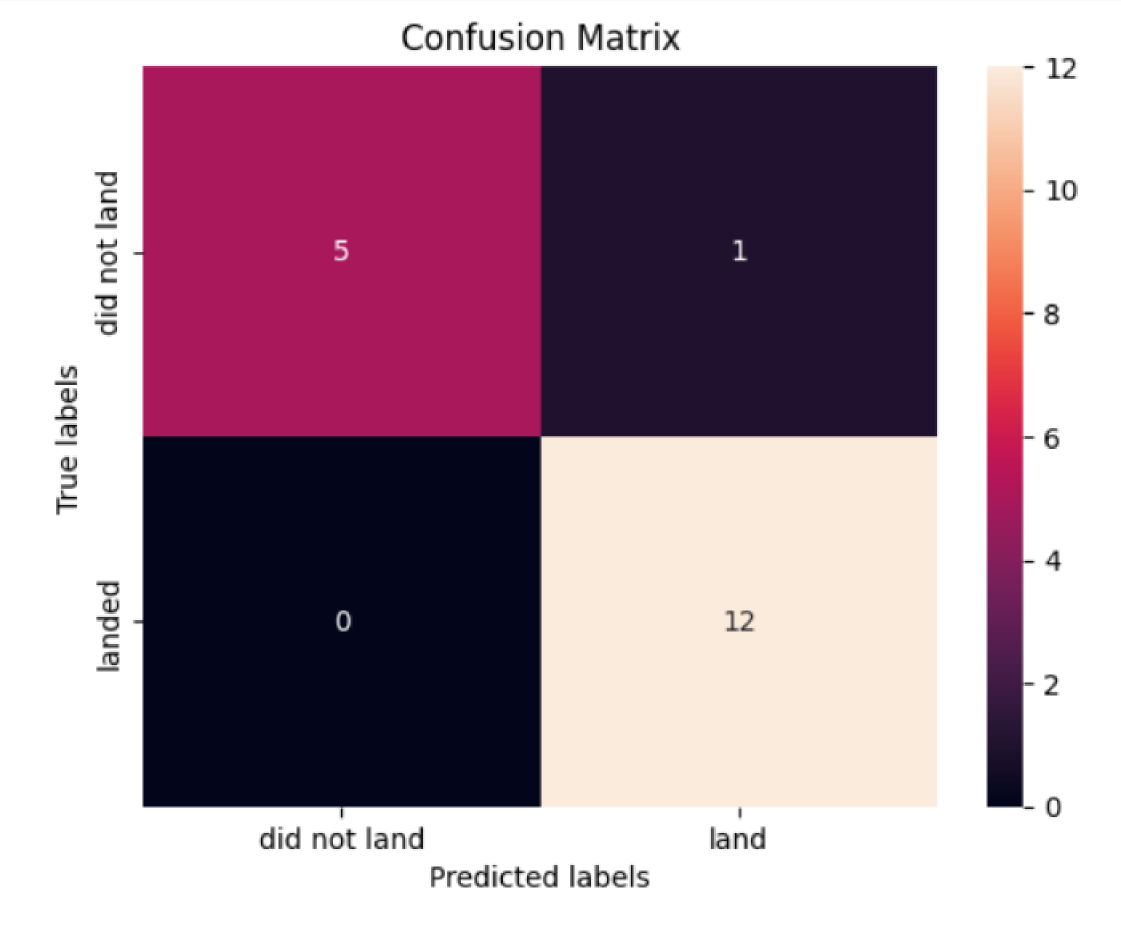
Classification Accuracy

- Model accuracy for all built classification models
- SVM model has the highest classification accuracy



Confusion Matrix

- Confusion matrix of the best performing model which is Support Vector Machine model in this case.
- It has a test accuracy of 94.44%
- True Positive - 12 (True label is landed, Predicted label is also landed)
- False Positive - 1 (True label is not landed, Predicted label is landed)



Conclusions

- KSC LC 39A and CCAFS SLC 40 site have been used for recent launches.
- Past record has shown that Rocket launches with a payload above 8000 kg have a very good rate of success, i.e., 87.5%
- By taking number of launches in account, launches to orbit VLEO have been quite successful and most of them have been recent launches.
- Out of last 30 odd launches, most of the launches were made to Orbit ISS and VLEO
- Nearly half of the successful launches have been made from KSC LC-39A site

Appendix

- SpaceX Dash App Code:

https://github.com/Karanvir399/SpaceX-Falcon-9-First-Stage-Landing-Prediction/blob/main/spacex-dash-app_code.py

Thank you!