Winning the Race of Space

Muhammad Waris

15th June 2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Conclusion
- Appendix

Executive Summary

- **Summary of methodologies**
  - Data Collection through API
  - Data Collection with Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analytics with Folium
  - Machine Learning Prediction
- **Summary of all results**
  - Exploratory Data Analysis Result
  - Interactive Analytics with Screenshots
  - Predictive Analytics Result

Introduction

- **Project background and context**

  Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch. This goal of the project is to create a machine learning pipeline to predict if the first stage will land successfully.

- **Problems you want to find answers**

  - What factors determine if the rocket will land successfully?
  - The interaction amongst various features that determine the success rate of a successful landing.
  - What operating conditions needs to be in place to ensure a successful landing program.

Methodology

Methodology

- Data collection methodology:
  - Data was collected using SpaceX API and web scraping from Wikipedia.
- Perform data wrangling
  - One-hot encoding was applied to categorical features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

- **The data was collected using different methods which are as follow:**
  - Data Collection was done by using get request to the SpaceX API

  - Next, we decoded the response content using .json() function call and converted it into a pandas dataframe using .json_normalize().

  - We then proceeded to clean the data, checked for missing values and fill in missing values where it was needed.

  - In addition, we performed web scraping on Wikipedia to obtain Falcon 9 launch records using BeautifulSoup.

  - Our objective was to extract the launch records from an HTML table, parse the table and convert it to a pandas dataframe for future analysis.

Data Collection – SpaceX API

- We used the get request to the SpaceX API to collect data, clean the requested data and did some basic data wrangling and formatting.
- The link to the notebook is
- https://github.com/Waris99/Final-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

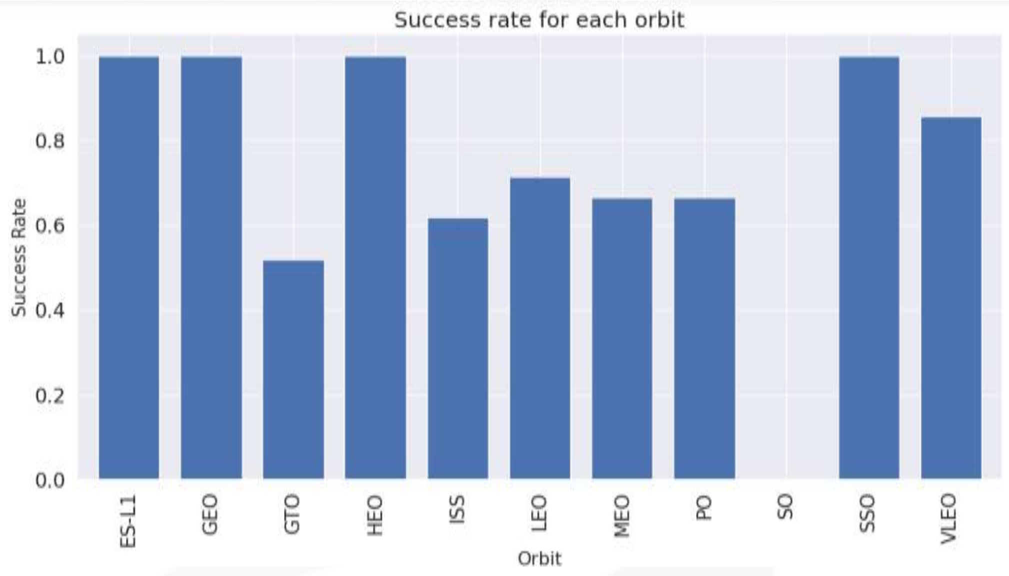
- We applied web scrapping to the records of Falcon 9 launch with BeautifulSoup
- We parsed the table and converted it into a pandas dataframe.
- The link to the notebook is
- https://github.com/Waris99/Final-Capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- We performed exploratory data analysis
- We calculated the number of launches at each site, and the number and occurrence of each orbits
- We created landing outcome label from outcome column and exported the results to csv.
- The link to the notebook is
- https://github.com/Waris99/Final-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- We explored the data by visualizing the relationship between flight number and launch Site, payload and launch site, success rate of each orbit type, flight number and orbit type, the launch success yearly trend.
- The link to the notebook is
- https://github.com/Waris99/Final-Capstone/blob/main/jupyter-labs-spacex-data%20visualization.ipynb



EDA with SQL

- We loaded the SpaceX dataset into a PostgreSQL database without leaving the jupyter notebook.
- We applied EDA with SQL to get insight from the data. We applied different queries to find out for instance:
  - The names of unique launch sites in the space mission.
  - The total payload mass carried by boosters launched by NASA (CRS)
  - The average payload mass carried by booster version F9 v1.1
  - The total number of successful and failure mission outcomes
  - The failed landing outcomes in drone ship, their booster version and launch site names.
- The link to the notebook ishttps://github.com/Waris99/Final-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- We marked all launch sites, and added map objects such as markers, circles, lines to mark the success or failure of launches for each site on the folium map.
- We assigned the feature launch outcomes (failure or success) to class 0 and 1.i.e., 0 for failure, and 1 for success.
- We calculated the distances between a launch site to its proximities. We answered some question for instance:
  - Are launch sites near railways, highways and coastlines.
  - Do launch sites keep certain distance away from cities.
- The link to notebook is https://github.com/Waris99/Final-Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- We built an interactive dashboard with Plotly dash
- We plotted pie charts showing the total launches by a certain sites
- We plotted scatter graph showing the relationship with Outcome and Payload Mass (Kg) for the different booster version.
- The link to the notebook ishttps://github.com/Waris99/Final-Capstone/blob/main/app.py

Machine Learning: Predictive Analysis

- We loaded the data using numpy and pandas, transformed the data, split our data into training and testing.
- We built different machine learning models and tune different hyperparameters using GridSearchCV.
- We used accuracy as the metric for our model
- We found the best performing classification model.
- The link to the notebook is https://github.com/Waris99/Final-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction.ipynb

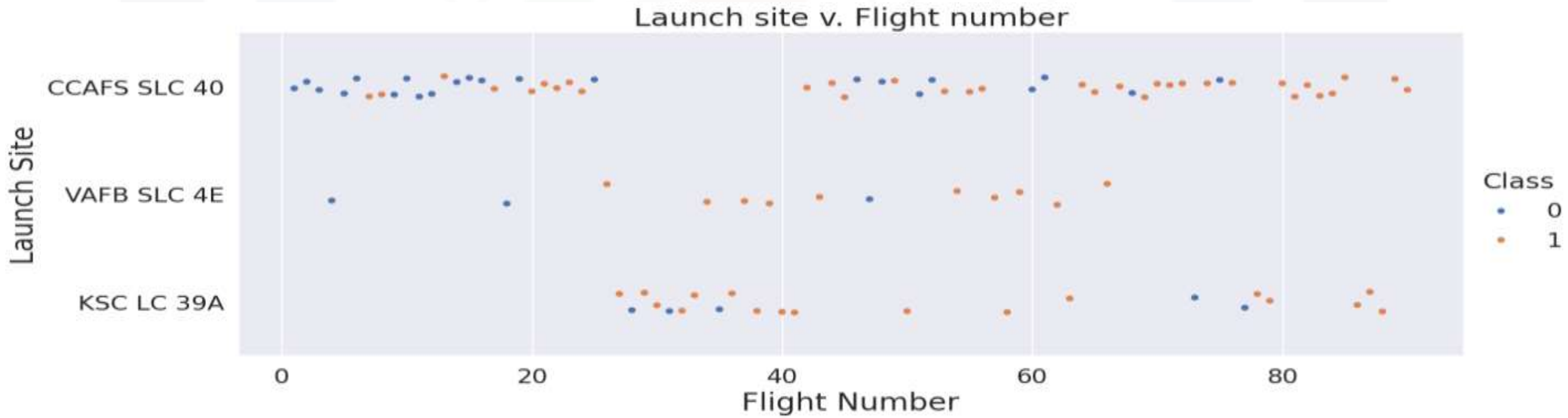
Results

- Exploratory data analysis results
- Demo of Interactive analytics in screenshots
- Predictive analysis results

Insights Drawn From EDA

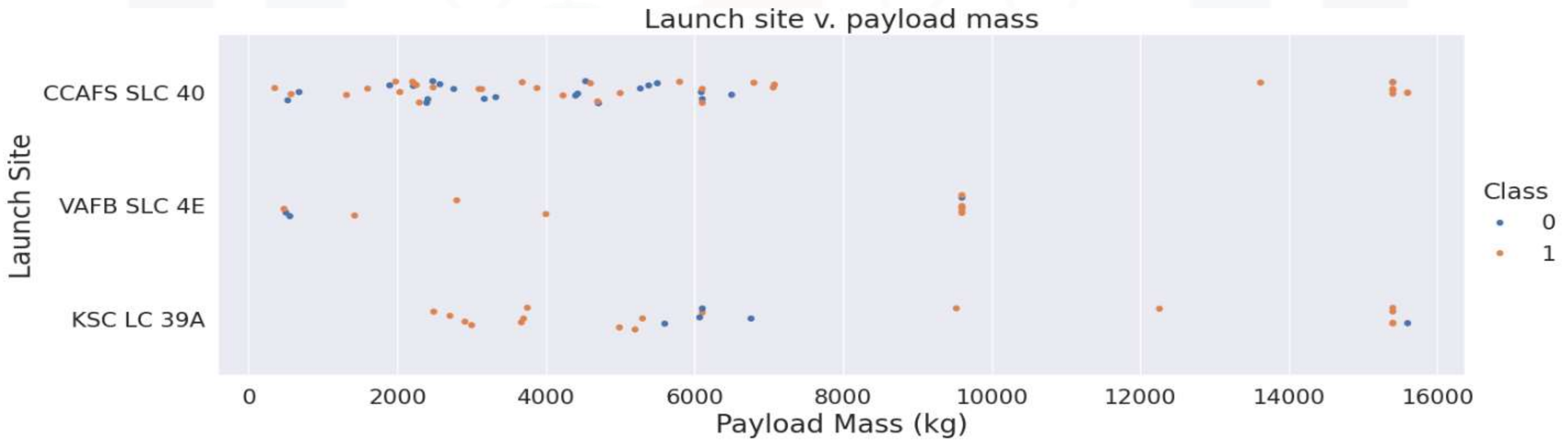
Flight Number vs. Launch Site

- From the plot, we found that the larger the flight amount at a launch site, the greater the success rate at a launch site



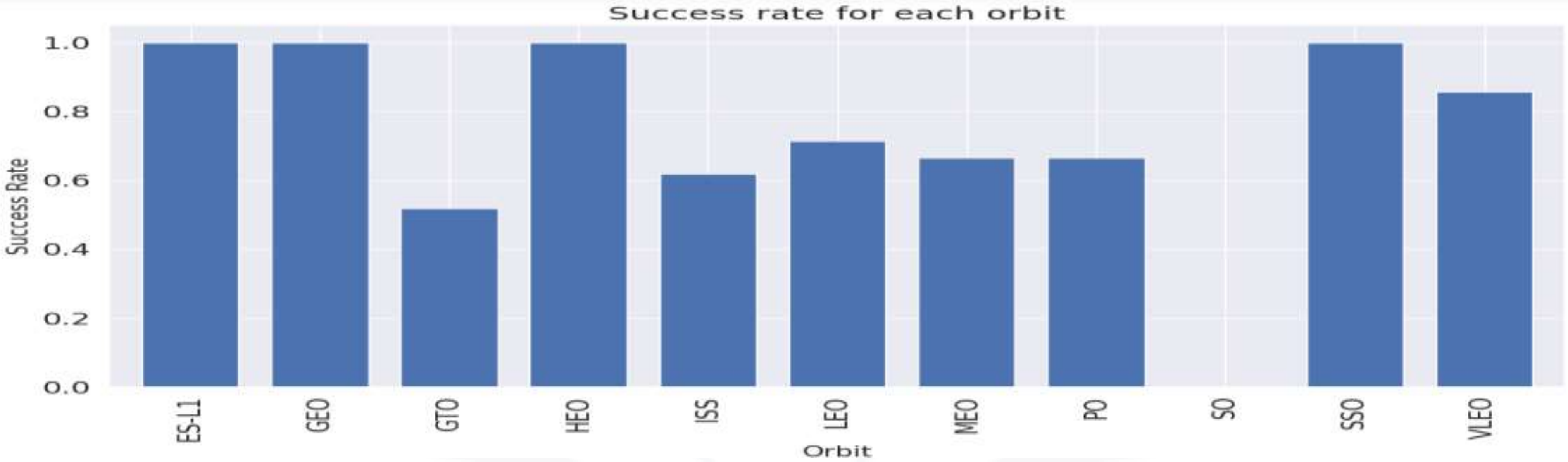
Payload vs. Launch Site

- The greater the payload mass for the site CCAFS SLC 40 the higher the success rate for the rocket



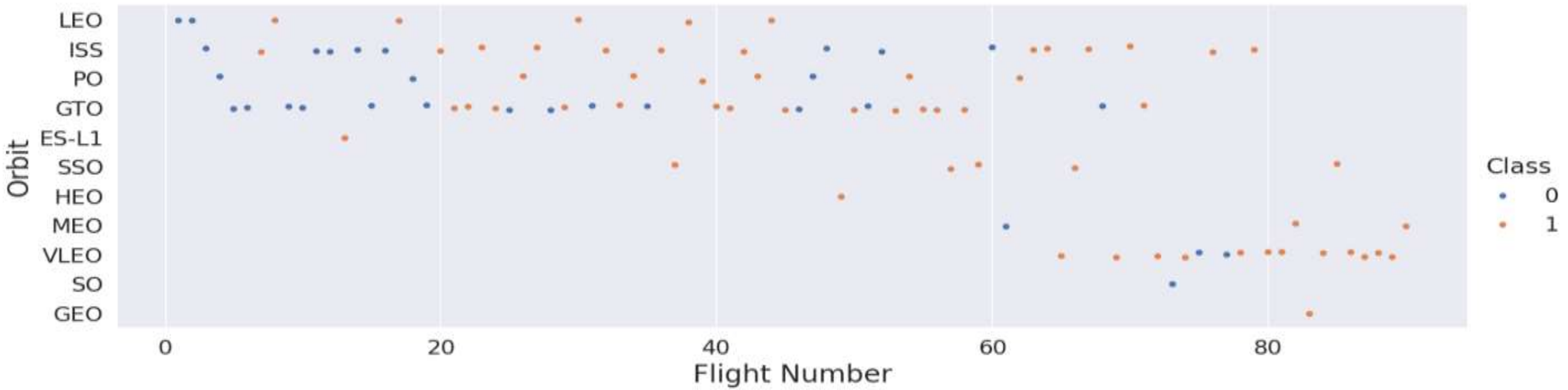
Success Rate vs. Orbit Type

- From the plot, we can see that ES-L1, GEO, HEO, SSO, VLEO had the most success rate.



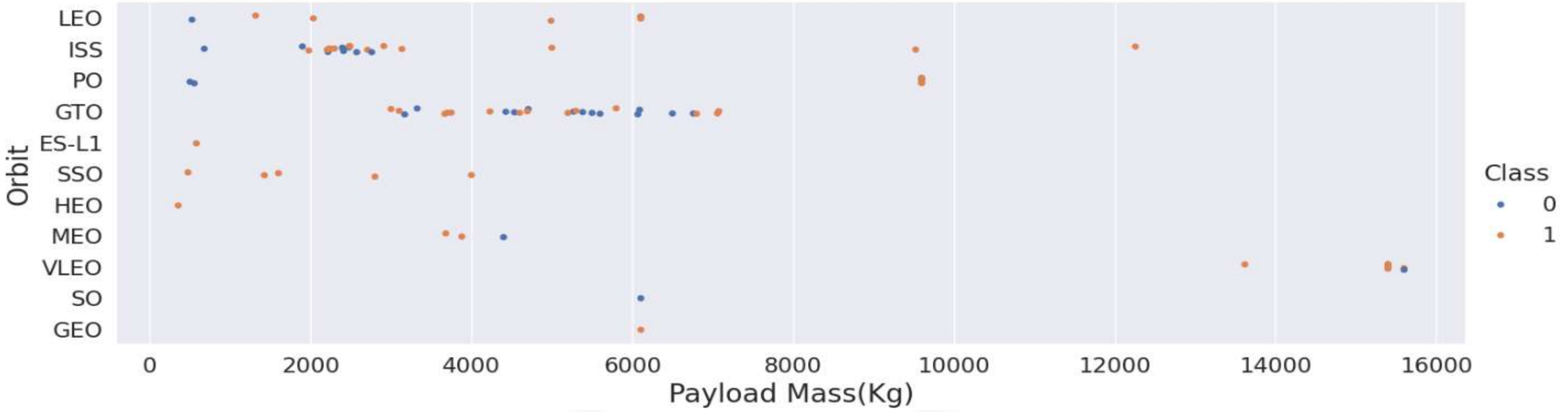
Flight Number vs. Orbit Type

- The plot below shows the Flight Number vs. Orbit type. We observe that in the LEO orbit, success is related to the number of flights whereas in the GTO orbit, there is no relationship between flight number and the orbit.



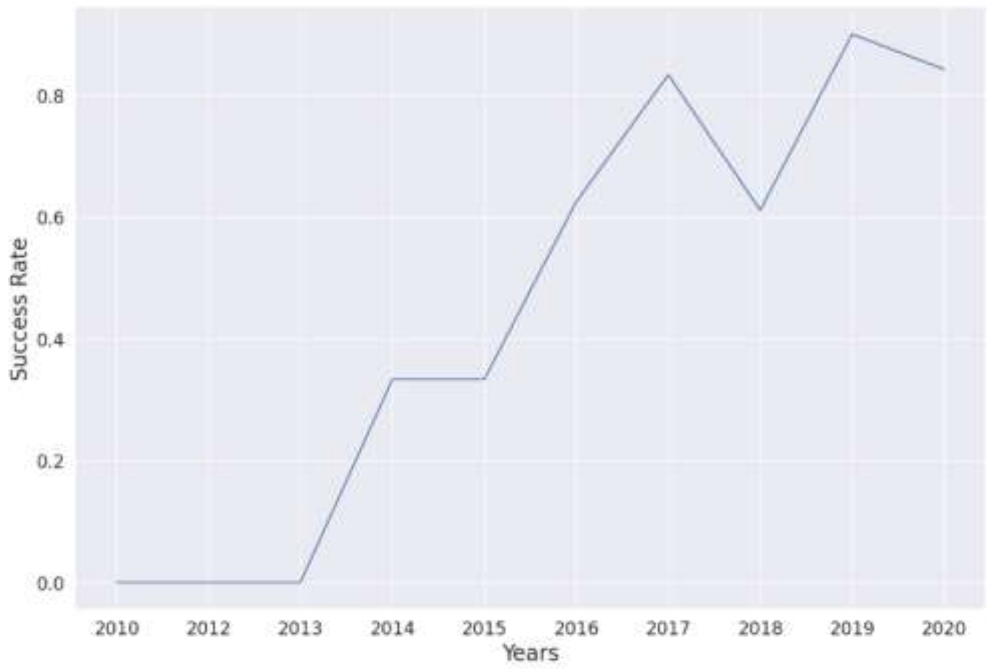
Payload vs. Orbit Type

- We can observe that with heavy payloads, the successful landing are more for PO, LEO and ISS orbits.



Launch Success Yearly Trend

- From the plot, we can observe that success rate since 2013 kept on increasing till 2020.



All Launch Site Names

- We used the **Distinct** key word to show the Unique launch sites from SpaceX Data.

```
%sql SELECT DISTINCT(LAUNCH_SITE) FROM SPACEXTBL
```

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

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

Launch Site Names Begin with 'CCA'

- We used the above query to display the 5 records of launch sites begin with 'CCA'

```
%sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5
```

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

- We calculated the total payload mass carried by the boosters from NASA as 45596 using the below query

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)'
```

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

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by the F9 v1.1 was 2928.4 as calculated.

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1'
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- We observed that the dates of the first successful landing on ground pad was 22nd December 2015

```
%sql SELECT MIN(DATE) as FIRST_SUCCESFUL_LANDING FROM SPACEXTBL WHERE LANDING_OUTCOME = 'Success (ground pad)'
```

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

FIRST_SUCCESFUL_LANDING
2015-12-22

Successful Drone Ship Landing with Payload Mass Between 4000 and 6000

- We used the WHERE clause to filter the desired result for boosters which have successfully landed on drone ship and applied the **AND** condition to determine successful landing with

```
%sql SELECT PAYLOAD FROM SPACEXTBL WHERE LANDING_OUTCOME = 'Success (drone ship)' and PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
```

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

Payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

- We used wildcard like ‘%’ to filter the result for **WHERE** Mission Outcomes was a success or failure

```
%sql SELECT MISSION_OUTCOME, COUNT(*) FROM SPACEXTBL GROUP BY MISSION_OUTCOME
```

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

Mission_Outcome	COUNT(*)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- We determine the booster that have carried the maximum payload using a subquery in the **WHERE** clause and the **MAX()** function

```
%sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
```

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

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

2015 Launch Records

- We used a combinations of the **WHERE** clause, **LIKE, NAD, and BETWEEN** conditions to filter for failure landing outcomes in drone ship, their booster versions, and launch site names for year 2015

```
%sql SELECT SUBSTR(Date,6,2) as MONTH, DATE,BOOSTER_VERSION, LAUNCH_SITE, LANDING_OUTCOME FROM SPACEXTBL
WHERE LANDING_OUTCOME = 'Failure (drone ship)' AND SUBSTR(Date,0,5)='2015'
```

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

MONTH	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

- We selected Landing Outcomes and the **COUNT** of landing outcomes for the data and used **WHERE** clause to filter the result for landing outcomes **BETWEEN** 2010-06-04 to 2010-03-20
- We applied the **GROUP BY** clause to group the landing outcomes and the

```
%sql SELECT LANDING_OUTCOME, COUNT(*) as COUNT_LANDING_OUTCOMES FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING_OUTCOME ORDER BY COUNT_LANDING_OUTCOMES DESC
```

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

Landing_Outcome	COUNT_LANDING_OUTCOMES
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

Launch Sites Proximities Analysis

All launch Sites Global Map Maker

- We ca see that the launch sites are in the United States of America coasts Florida and California

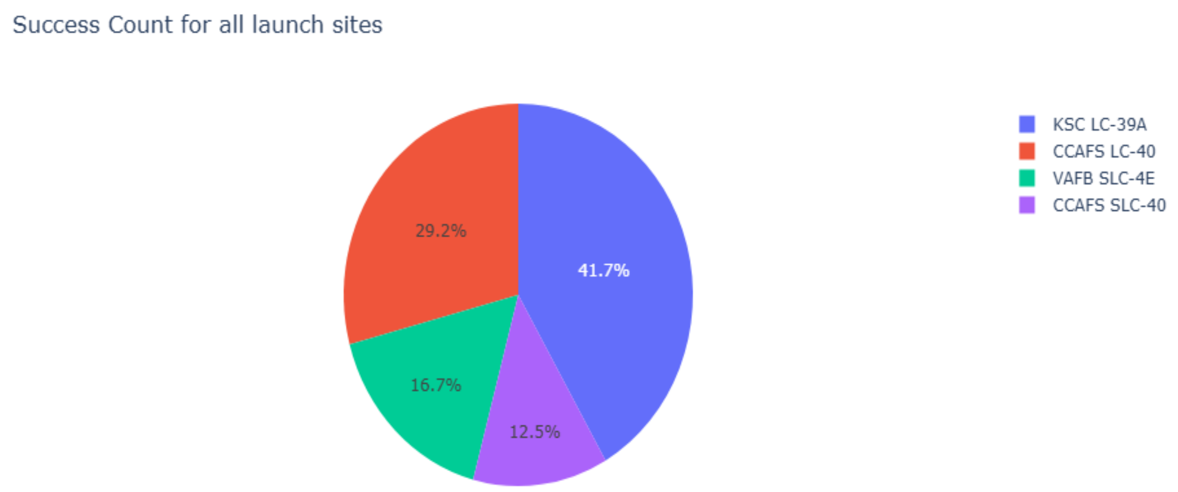
Launch Site Distance to Landmarks

- Are launch sited in close proximity to railways? NO
- Are launch sites in close proximity to highways? No
- Are launch sites in close proximity to coastline? No
- Do launch sites keep certain distance away from cities? Yes

Build a Dashboard with Plotly Dash

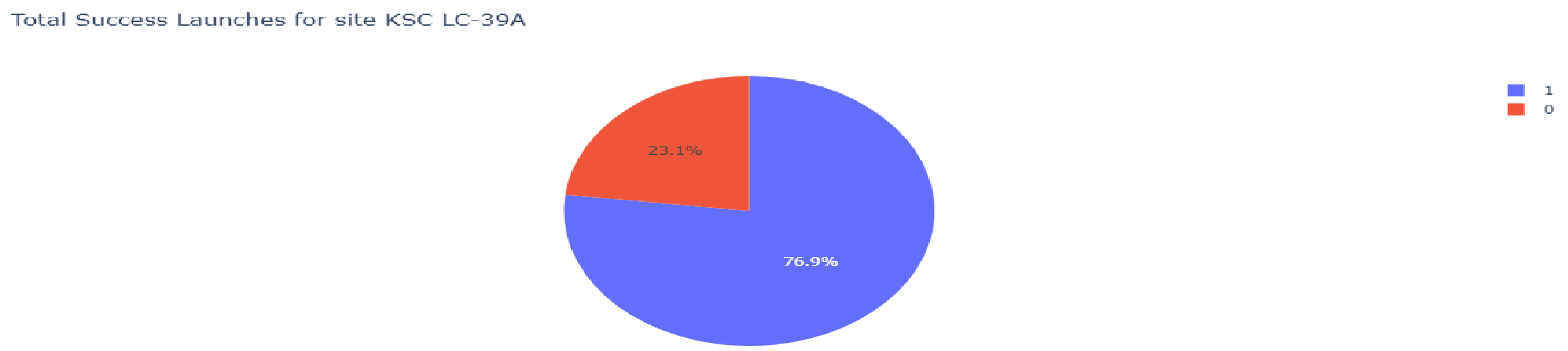
Pie chart showing success percentage achieved by each launch site

- Total Success Launches by all sites.
- The link to notebook is
- https://github.com/Waris99/Final-Capstone/blob/main/app.py



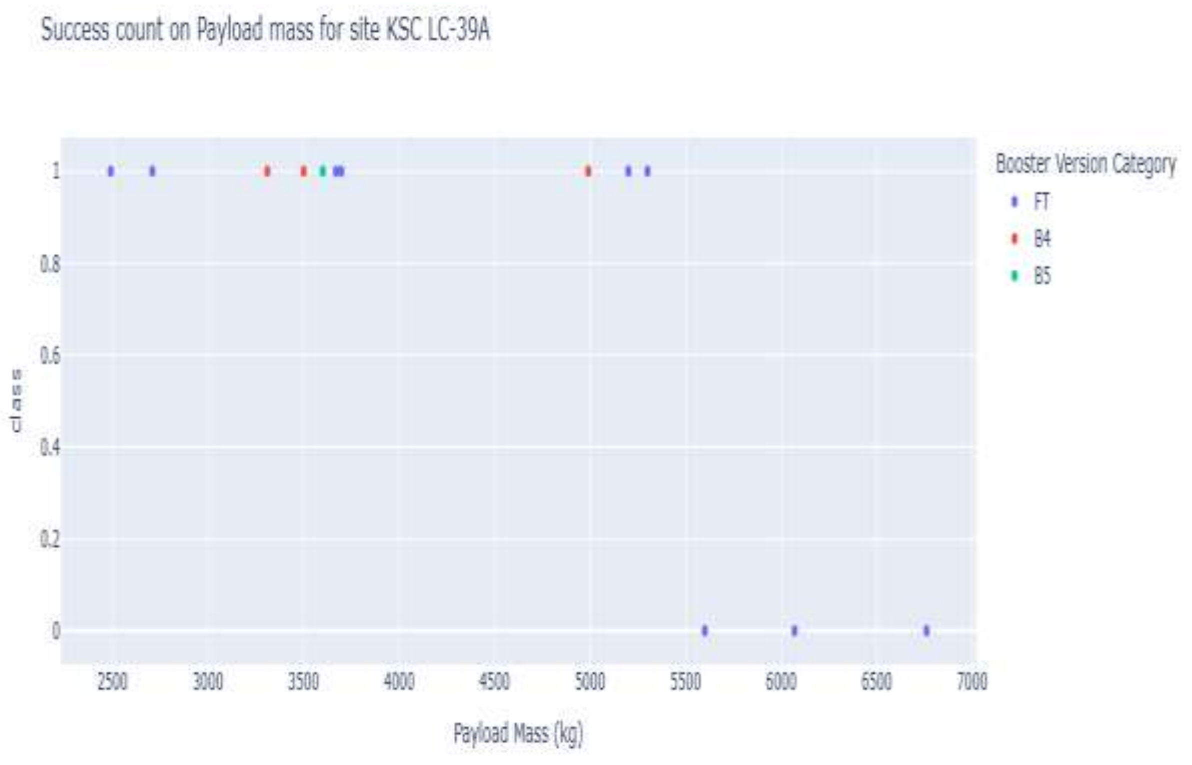
Pie chart showing Launch site with highest launch success ratio

- KSC LC-39A has the highest success ratio of 76.9% while getting a 23.1% failure rate



Scatter plot of Payload vs Launch Outcome for all sites, with different payload selected in the range slider

- The link to notebook is
- https://github.com/Waris99/Final-Capstone/blob/main/app.py



IBM Developer SKILLS NETWORK

Predictive Analysis

Classification Accuracy

- The classification accuracy of all models are as follow

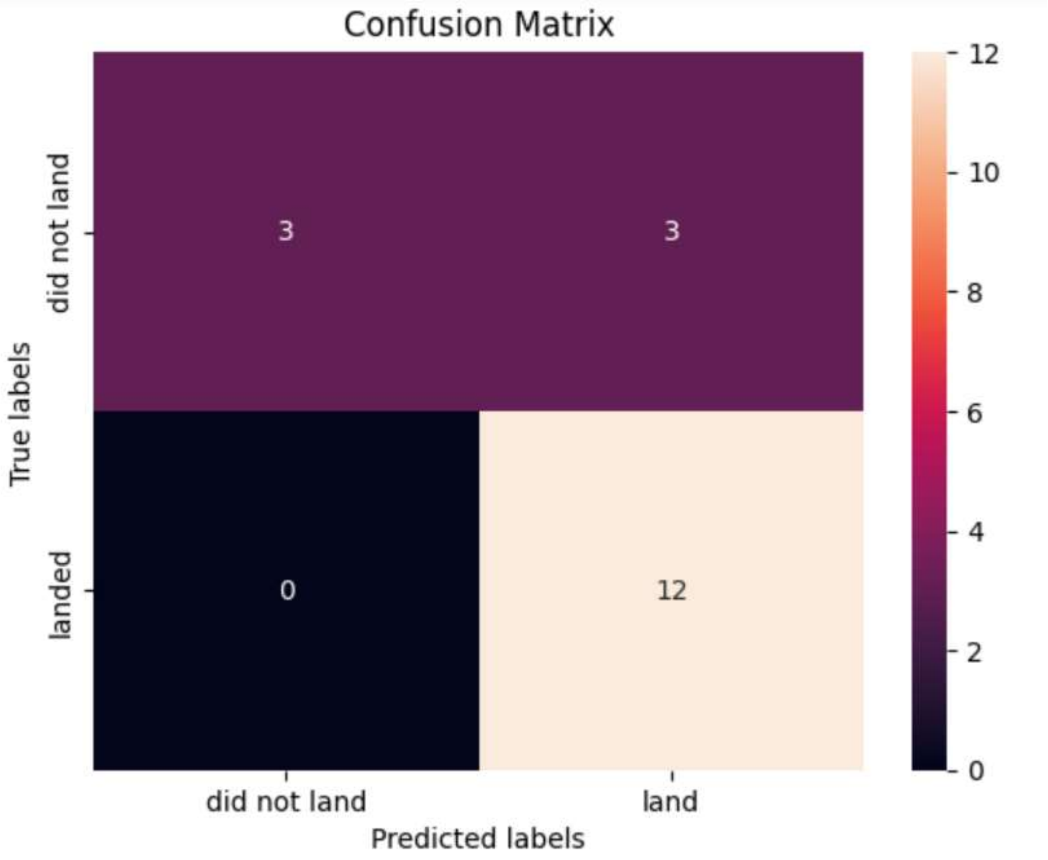
```
report = pd.DataFrame({'Method' : ['Test Data Accuracy']})

knn_accuracy=knn_cv.score(X_test, Y_test)
Decision_tree_accuracy=tree_cv.score(X_test, Y_test)
SVM_accuracy=svm_cv.score(X_test, Y_test)
Logistic_Regression=logreg_cv.score(X_test, Y_test)

report['Logistic_Reg'] = [Logistic_Regression]
report['SVM'] = [SVM_accuracy]
report['Decision Tree'] = [Decision_tree_accuracy]
report['KNN'] = [knn_accuracy]
report.transpose()
```

]:

	0
Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.833333
KNN	0.833333

Confusion Matrix

- The confusion matrix for the decision tree classifier indicates that it effectively differentiates between the various classes. However, a significant issue is the occurrence of false positives, where the classifier incorrectly labels unsuccessful landings as successful



Conclusions

From all the analysis, we can conclude that:

- The larger the flight at a launch site, greater the success rate at a launch site.
- Launch success rate started to increase in 2013 till 2020
- Orbits ES-L1, GEO, HEO, SSO, VLEO had most success rate
- KSC LC-39A had most successful launches of any site

THANK YOU!