IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sid Ahmed Sahnoune
06/27/2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection - SpaceX API
  - Data Collection - Scraping
  - Data Wrangling
  - Exploratory Data Analysis - SQL
  - Exploratory Data Analysis - Data Visualization
  - Interactive Visual Analytics - Interactive maps
  - Interactive Data Visualization - Dashboard
  - Predictive Analysis
- Summary of all results
  - Exploratory Data Analysis results
  - Data visualizations
  - Predictive Analysis results

Introduction

- In the Space Race 2.0 that we are witnessing, keeping costs as low as possible is essential. One solution is reusable first stage, this implies the successful landing of this first stage. Otherwise, the launch cost would be much higher.
- One of the main problems of SPACEX is how to predict the outcome of a landing to successfully charge the right prices for its clients?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
    - Data was collected from SpaceX using its API and from Wikipedia using web scraping.
    - Perform data wrangling
    - The mean of the PayloadMass was used to replace the null values for this variable.
    - Landing outcome label was created from Outcome column.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - How to build, tune, evaluate classification models

Data Collection

The first dataset:

- Requested rocket launch data from SpaceX API using the following URL:
- https://api.spacexdata.com/v4/launches/past
- The response content was decoded as a Json and converted to a data frame.

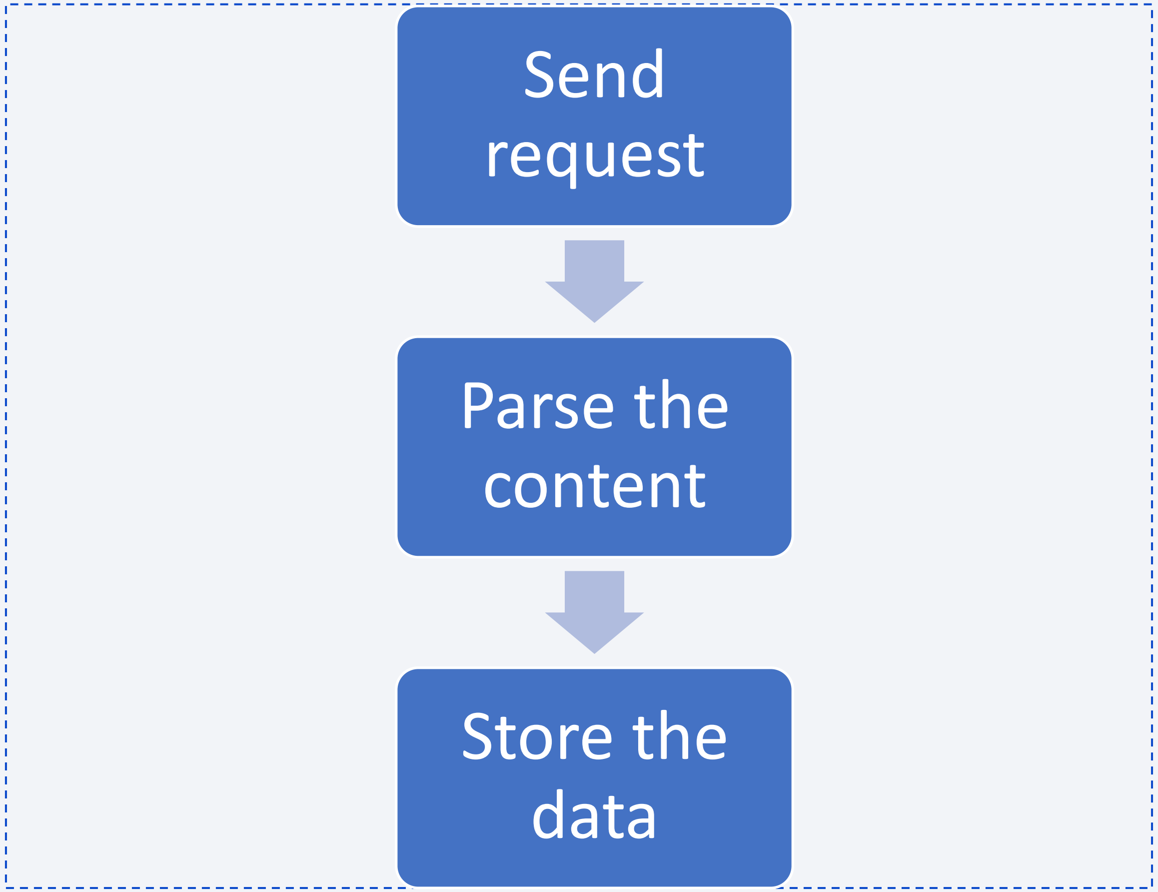
The second dataset:

- An HTTP GET method was performed to request the Falcon9 Launch HTML page, as an HTTP response.
- All column/variable names were extracted from the HTML table header.
- A data frame was created by parsing the launch HTML tables.

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

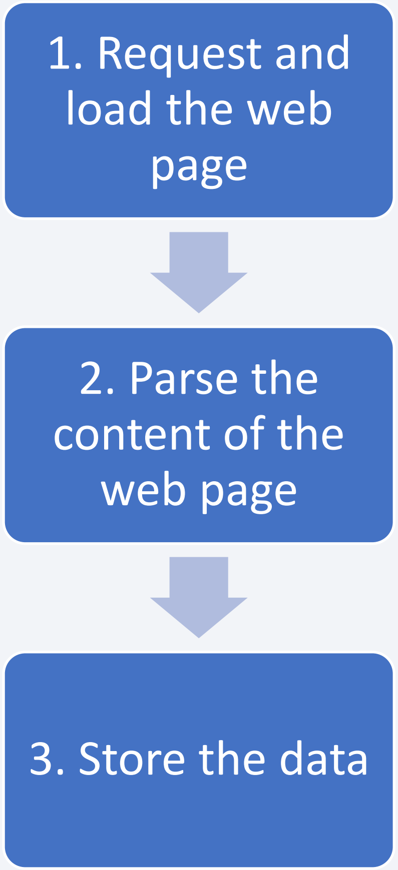
- https://github.com/Sid1941/Capstone-project-IBM-Data-Science/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- https://github.com/Sid1941/Capstone-project-IBM-Data-Science/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- The data was filtered to include only Falcon 9 launches.
- The mean of the PayloadMass was used to replace the null values for this variable.
- Created a landing outcome label from Outcome column.
- Created dummy variables to categorical columns.

- https://github.com/Sid1941/Capstone-project-IBM-Data-Science/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- https://github.com/Sid1941/Capstone-project-IBM-Data-Science/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- The names of the unique launch sites in the space mission.
- Total payload mass carried by boosters launched by NASA (CRS).
- The average payload mass carried by booster version F9 v1.
- The date when the first successful landing outcome in ground pad was achieved.
- Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000 KG.
- The total number of successful and failure mission outcomes.
- The names of the booster versions which have carried the maximum payload mass.
- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015.

- https://github.com/Sid1941/Capstone-project-IBM-Data-Science/blob/main/jupyter-labs-eda-sql-coursera.ipynb

Build an Interactive Map with Folium

We built maps and included Markers, circles, and lines to:

- Explain why you added those plots and interactions
- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities

- https://github.com/Sid1941/Capstone-project-IBM-Data-Science/blob/main/lab_jupyter_launch_site_location.ipynb
- For rendered maps, please visit this URL.

Build a Dashboard with Plotly Dash

An interactive dashboard was built using Plotly dash to display:

- A pie chart showing the success rate by site.
- A scatter plot showing the relationship of the outcome with the payload mass and the booster version.

- https://github.com/Sid1941/Capstone-project-IBM-Data-Science/blob/main/Dashboard.ipynb

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- The data was standardized and then split into a training set and a testing set.
- A logistic regression, a SVM, a decision tree and a KNN models were built and their hyperparameters were tuned using GridSearchCV.
- Accuracy was used to compare between the models, and it was found that the decision tree model performed better than the others.

- https://github.com/Sid1941/Capstone-project-IBM-Data-Science/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

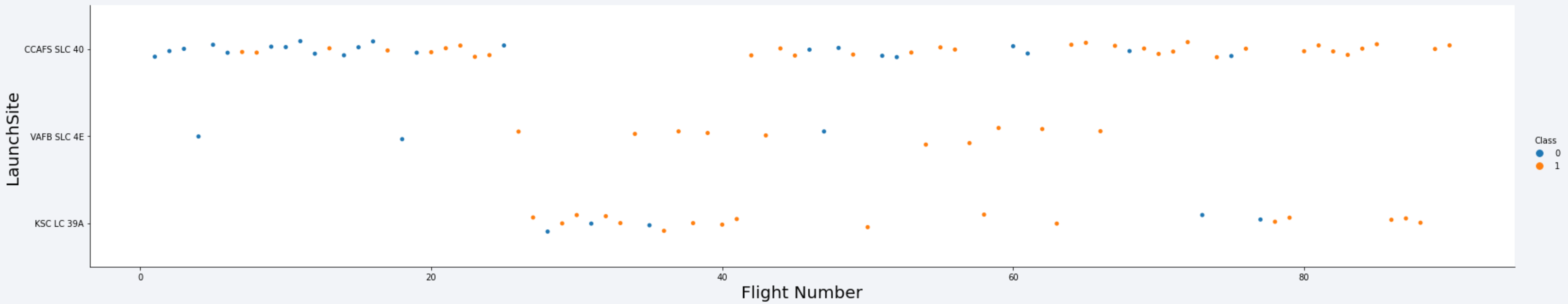
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

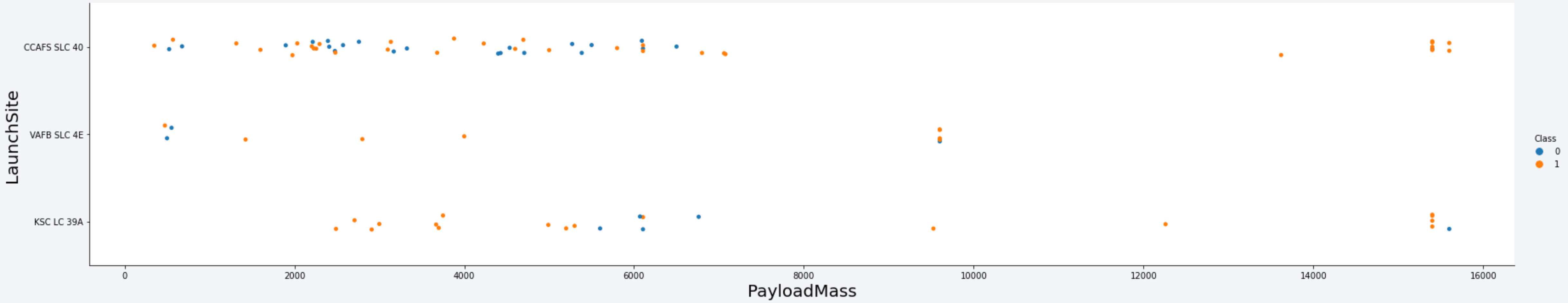
Flight Number vs. Launch Site

- The larger the flight number, the greater the success rate.



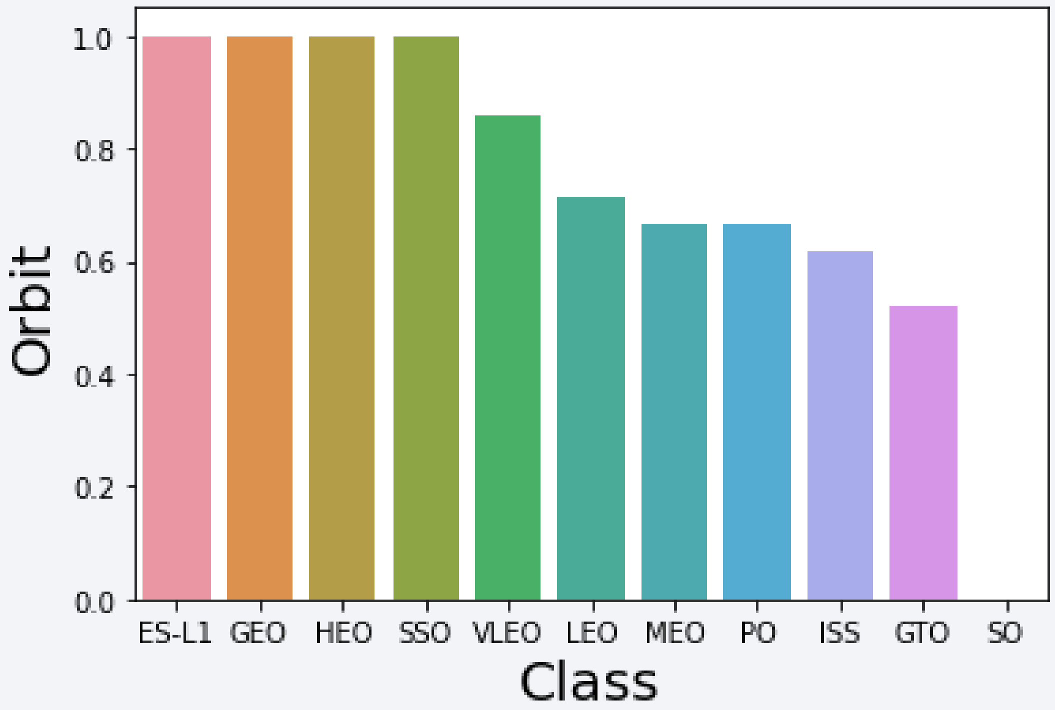
Payload vs. Launch Site

- The greater the payload mass, the greater the success rate for the launch site CCAFS SLC 4. Whereas the smaller the payload mass, the greater the success rate for the launch site KSC LC 39A.



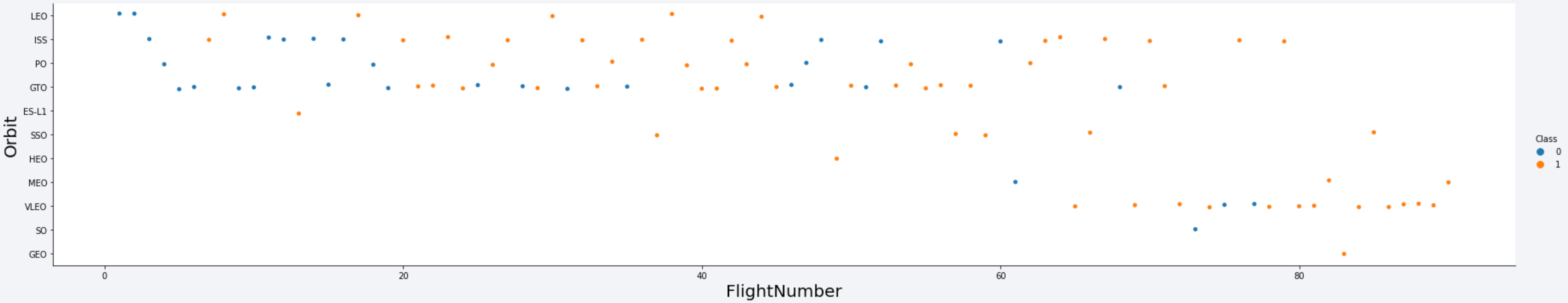
Success Rate vs. Orbit Type

- The orbits with the highest success rates are ES-L1, GEO, HEO, and SSO had the highest success rates.



Flight Number vs. Orbit Type

- For the VLEO, MEO, and LEO orbits, the largest the flight number, the highest the success rate.

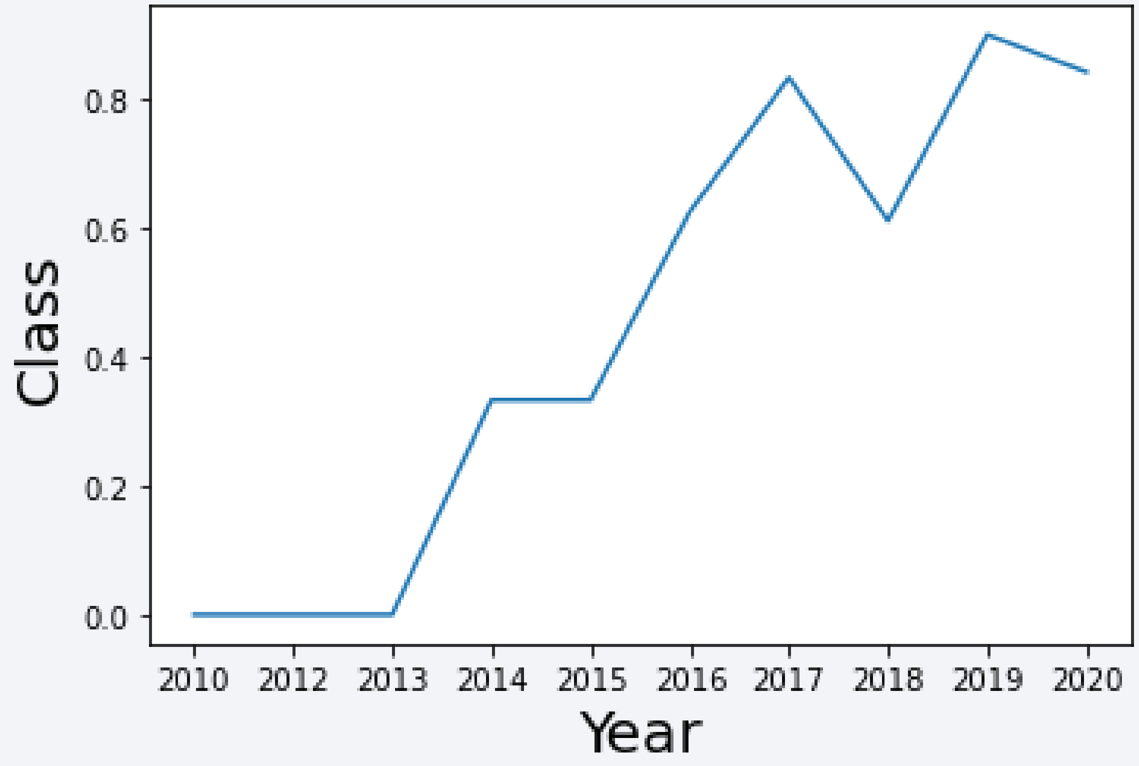


Payload vs. Orbit Type

- For the orbits ISS and LEO, the largest the payload mass, the highest the success rate.

Launch Success Yearly Trend

- The success rate chart exhibits an upward trend from 2013 to 2020.



All Launch Site Names

- Used a select distinct statement to return all the launch sites used by SPACEX.
- Query:

```
%sql SELECT DISTINCT launch_site FROM SPACEX;
```

- Result:

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

Launch Site Names Begin with 'CCA'

- Found 5 records where launch sites begin with `CCA` using a condition “WHERE”, the logical operator “LIKE”, percent literal to match the pattern and the “LIMIT” clause.

Query:

%sql SELECT * FROM SPACEX WHERE launch_site LIKE 'CCA%' LIMIT 5;

Result:

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

Total Payload Mass

- Calculated the total payload carried by boosters from NASA using the SUM function, a condition “WHERE”, the logical operator “LIKE”, percent literal to match the pattern and the “LIMIT” clause.

- Query:

%sql SELECT SUM(payload_mass__kg_) AS payload_nasa FROM SPACEX WHERE customer LIKE 'NASA%';

- Result:

payload_nasa
99980

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Query:
- %sql SELECT AVG(payload_mass__kg_) AS payload_F9 FROM SPACEX WHERE booster_version = 'F9 v1.1’;

- Result:

payload_f9
2534

First Successful Ground Landing Date

- Found the date of the first successful landing outcome on ground pad using the “MIN” function and a condition.
- Query:

%sql SELECT MIN (date) AS first_successful_landing_date FROM SPACEX WHERE landing__outcome LIKE('Success (ground pad)’);

Result:

first_successful_landing_date
2015-12-22

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 using “WHERE” clause and “AND” operators.
- Query :

%sql SELECT booster_version FROM SPACEX WHERE payload_mass__kg_<6000 AND payload_mass__kg_>4000 AND landing__outcome LIKE ('Success (drone ship)’);

- Result:

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

Total Number of Successful and Failure Mission Outcomes

- Calculated the total number of successful and failure mission outcomes using the “COUNT” function and a “GROUP BY” clause.
- Query:
- %sql SELECT Mission_Outcome, COUNT(Mission_Outcome) AS COUNT FROM SPACEX GROUP BY Mission_Outcome ORDER BY COUNT DESC;
- Result:

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

Boosters Carried Maximum Payload

- Listed the names of the booster which have carried the maximum payload mass using a "WHERE" clause and a subquery.
- Query:
- %sql SELECT booster_version, payload_mass__kg_ FROM SPACEX WHERE payload_mass__kg_=(SELECT MAX(payload_mass__kg_) FROM SPACEX);

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

2015 Launch Records

- Listed the failed landing_outcomes in drone ship, their booster versions, and launch site names in year 2015 using a “WHERE” clause and the function “YEAR”.

- Query:

%sql SELECT booster_version, launch_site FROM SPACEX WHERE landing__outcome =('Failure (drone ship)') AND YEAR(DATE)=(2015);

- Result:

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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
- Query:

```
%sql SELECT landing__outcome, COUNT (landing__outcome) AS COUNT FROM SPACEX WHERE DATE >= '2010-06-04' and DATE <='2017-03-20' GROUP BY landing__outcome ORDER BY COUNT DESC;
```

- Result:

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

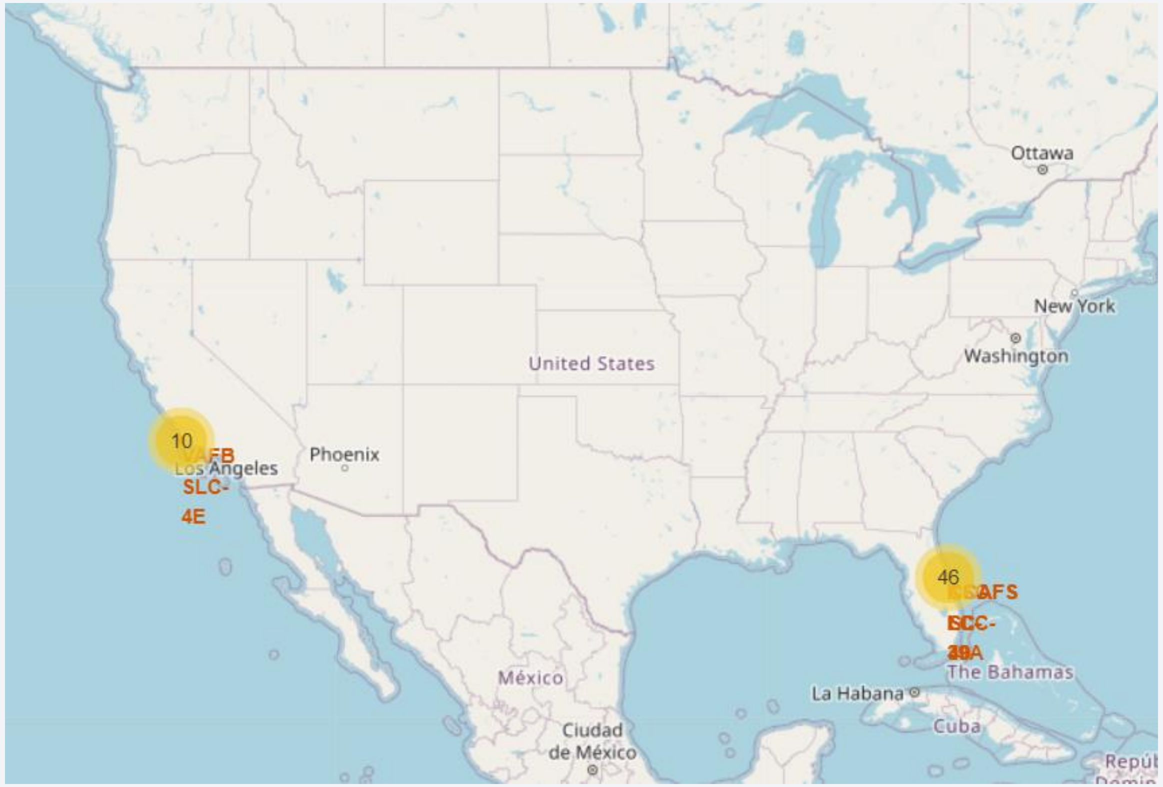
Launch Sites Proximities Analysis

SPACEX Launch Sites Locations

- SPACEX has two launches its rockets from the East Coast and from the west coast of the U.S.

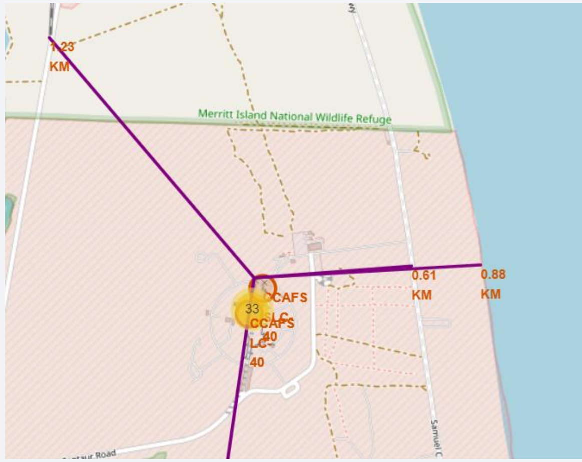
Number of Launches by Location

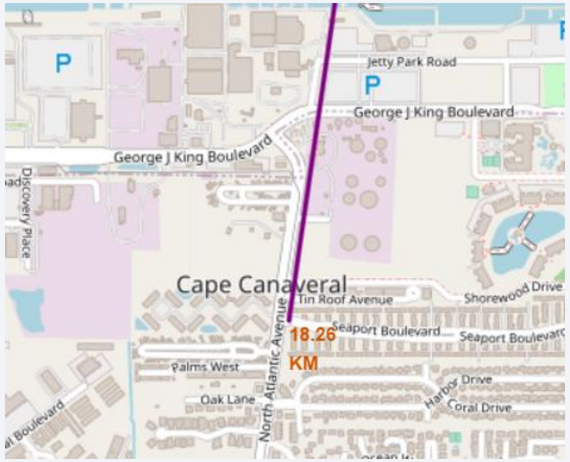
- Most of SPACEX launches were performed from the East Coast of the U.S.



Launch site proximity

- The launch site is close to the shore.
- The launch site is close to a railway.
- The launch site is close to a parkway.
- The launch site is 18 KM far away from the City of Cape Canaveral.



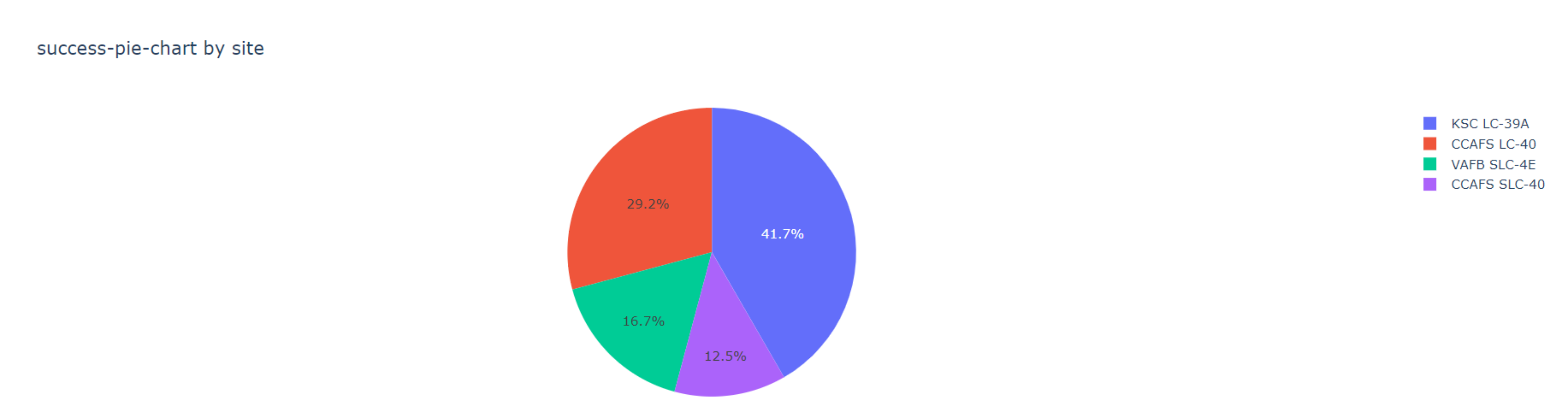


Section 4

Build a Dashboard with Plotly Dash

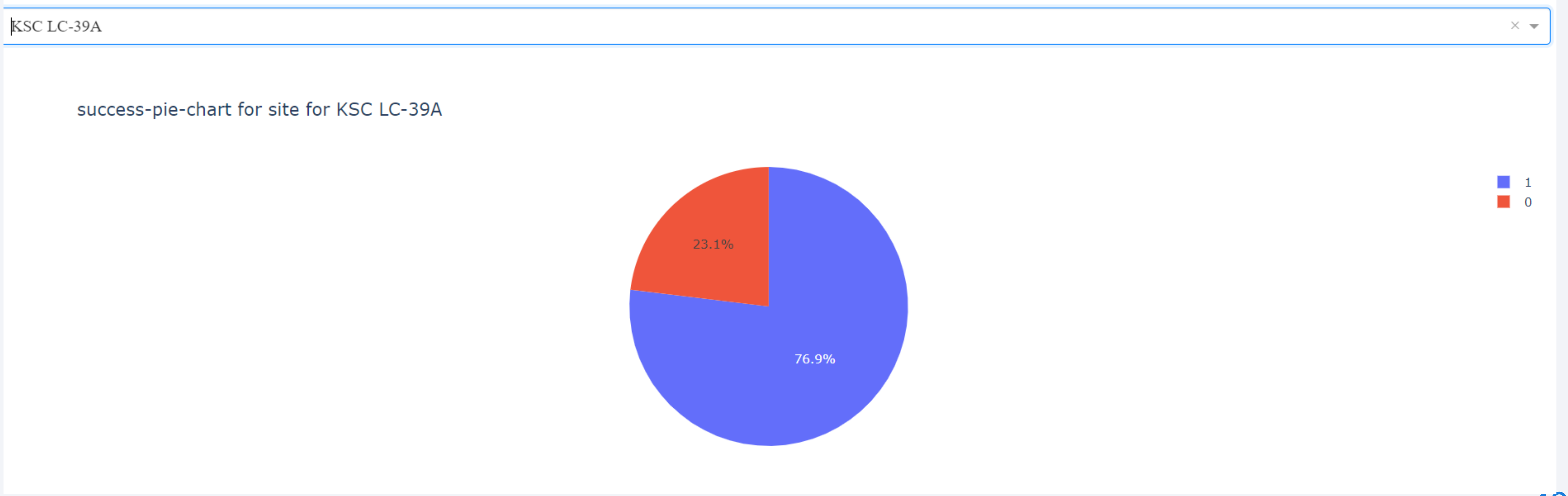
Launch Success by Launch Site

- 41.7% of successful launches were performed from launch site KSC LC-39A.



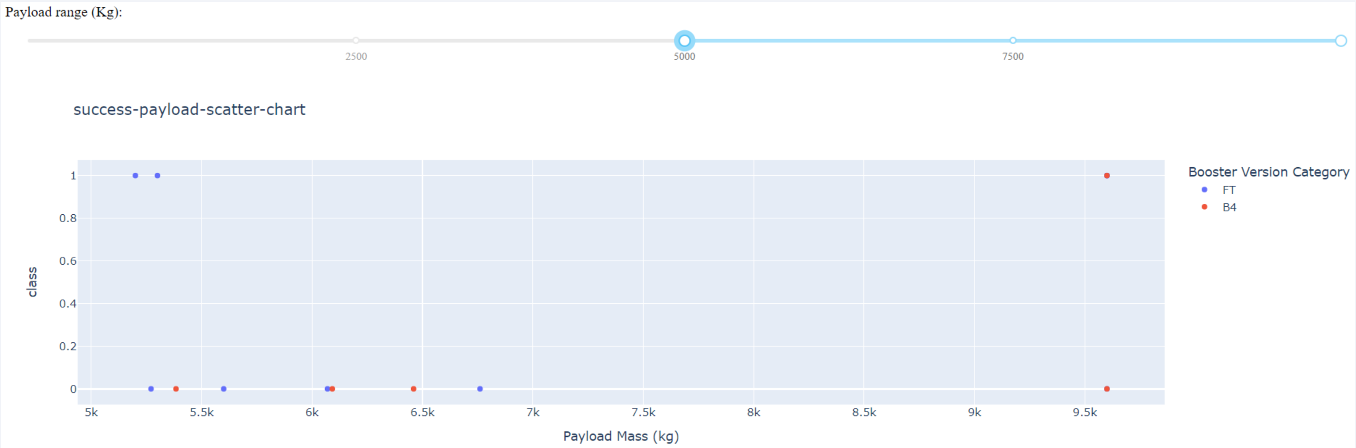
Success Rate for Launch Site KSC LC-39A

- The success rate of launch site KSC LC-39A is 76.9%.



Success Rate by Booster Version and Payload Mass

- Booster version categories FT and B4 have higher success rates with payload mass less than 5000KG than with payload mass of over 5000KG.

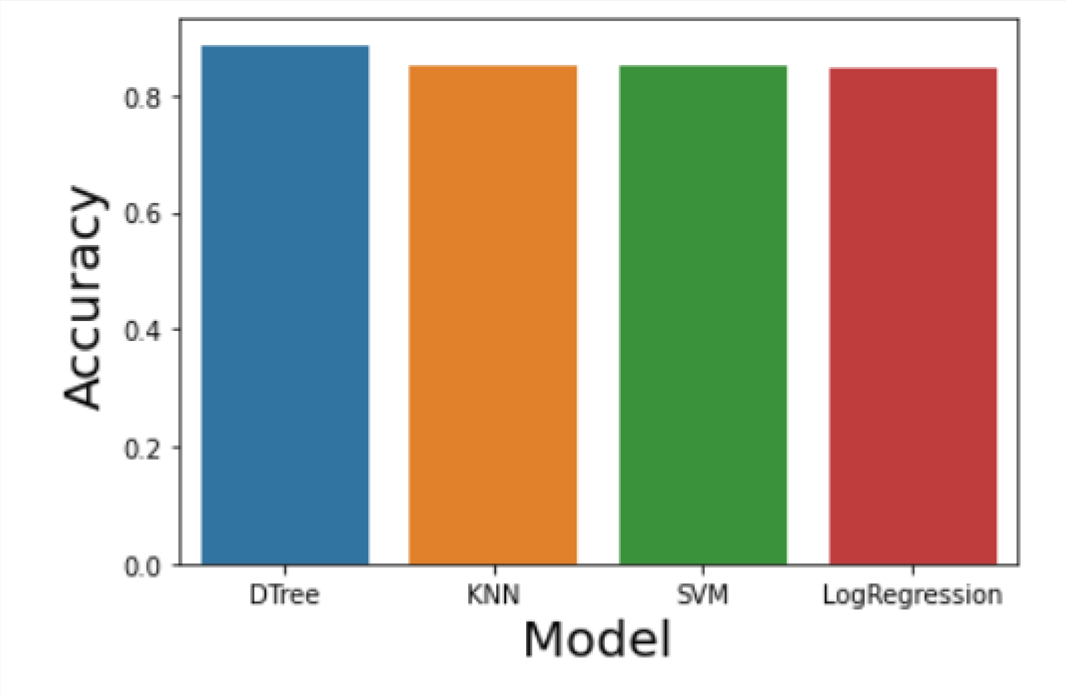


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The decision tree model has the highest classification accuracy.

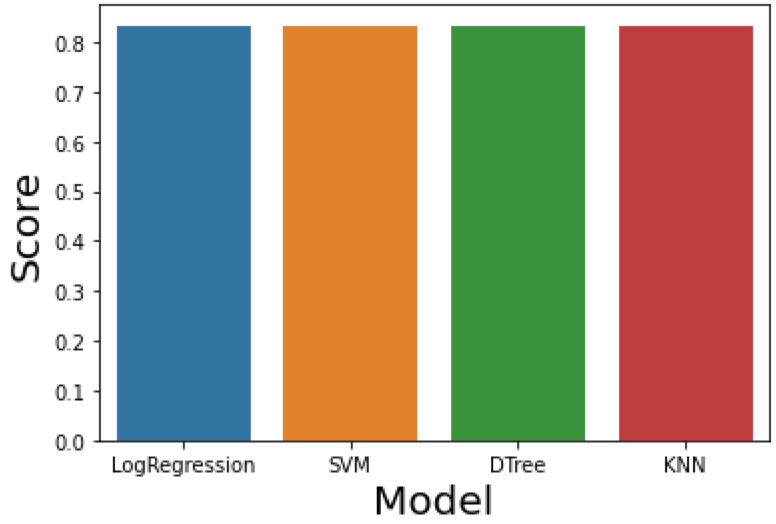


Confusion Matrix

- The decision tree incorrectly predicted 3 unsuccessful landings as being successful (false positives).

Prediction Power

- All considered models have the same predictive power.



Conclusions

- The larger the flight number, the greater the success rate.
- The orbits ES-L1, GEO, HEO, and SSO are the orbits with the highest success rates.
- The success rate chart exhibits an upward trend from 2013 to 2020.
- 41.7% of successful launches were performed from launch site KSC LC-39A.
- Booster version categories FT and B4 have higher success rates with payload mass less than 5000KG than with payload mass of over 5000KG.
- The decision tree model has a highest classification accuracy compared to the logistic regression, SVM, and KNN models.

Thank you!