Commercial Space Launch Competition Study

Christian
Kanyankogote
June 18, 2024

Executive Summary

The study analyzed various factors, like the reusability of rockets' boosters, impacting the success and costs of commercial space launches by SpaceX. The study relied on following methodologies:

- **Data Collection:** Information on Falcon 9 rocket launches was collected from SpaceX's API using GET requests and from Wikipedia by Web scrapping with GET requests and BeautifulSoup.
- **Data Wrangling** was performed to capture and analyze relevant data, deal with missing values, create a landing outcome classification label and variable, and to prepare the data for analysis and modeling.
- **Exploratory Data Analysis** with visualization was used to assess relationships between different variables, and EDA with SQL to find patterns in launch and mission records.
- **Interactive Visualization** was created with interactive Folium maps to examine details about the launch sites used by SpaceX, and a Plotly Dashboard to analyze launch statistics.
- **Building, Training and Tuning Supervised Machine Learning Models** to predict booster landing outcomes (binary classification) using Logistic Regression, Support Vector Machine (SVM), Decision Tree and K-Nearest Neighbor (KNN).

Overall, the rate of successful booster landings increased over the period observed (from June 2010 to December 2020) as the number of rocket launches increased.

The four launch sites (CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, and VAFB SLC-4E) used by SpaceX were located in coastal areas, close to the equator line relative to the continental USA. Launches from the KSC LC-39A launch site had higher rate of successful booster landings compared to the other sites.

The Decision Tree model slightly outperformed in classification accuracy of predicted booster landing outcomes.

Table of Contents

Introduction

Study Background

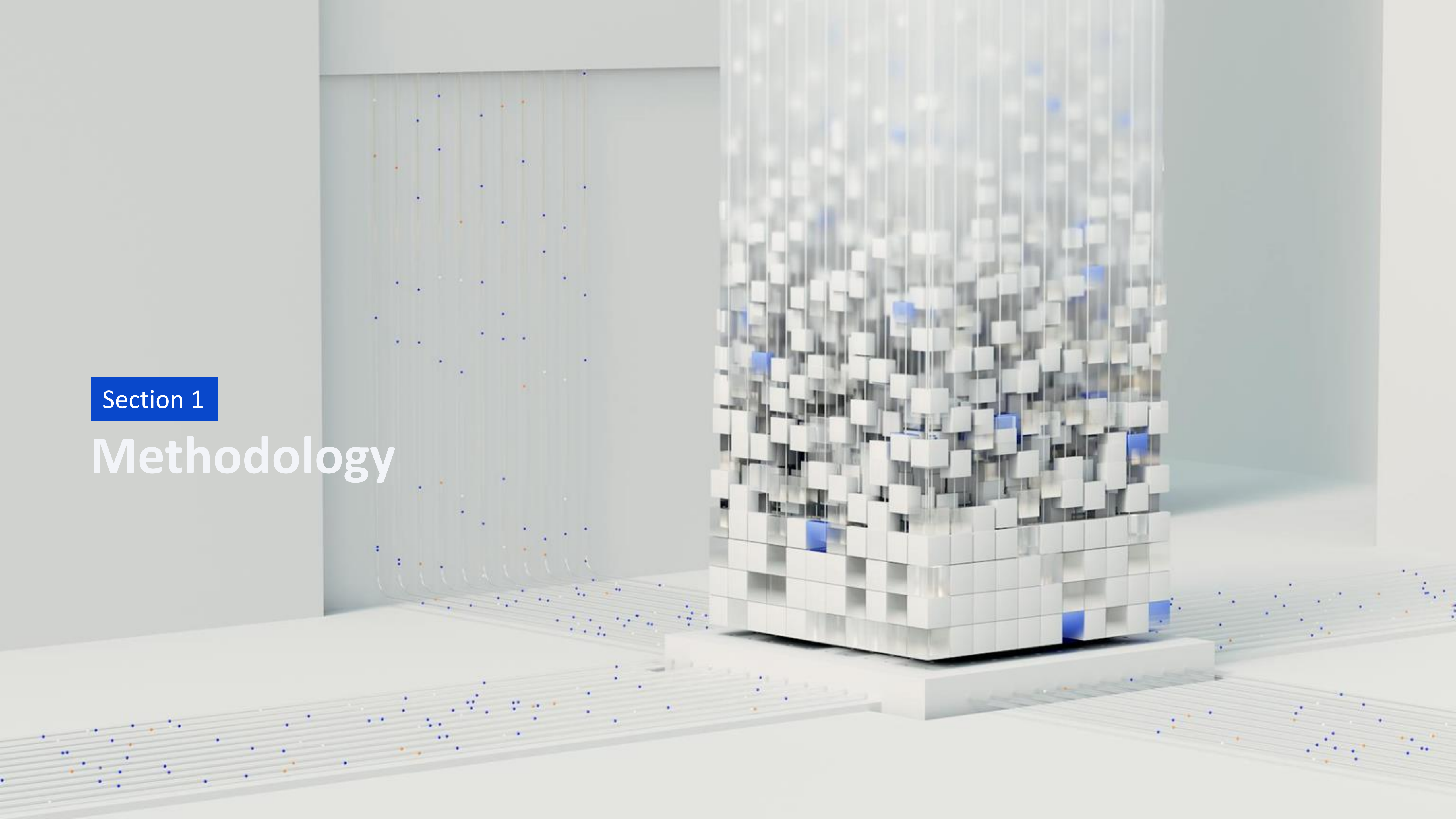
In the last twenty years, the Aerospace industry has experienced significant disruption due to the entry of new private companies who spurred innovation and competition in the commercial space launch services market. Since 2010, one private company, SpaceX, has rapidly seized a major share of the market because of their ability to build cost effective rockets, and to reuse rockets’ boosters (also called “first stage”). SpaceX offers launch services at significantly lower prices than its competitors. The study assesses opportunities for a new private company to compete with SpaceX by analysing various factors impacting the success of SpaceX’s Falcon 9 rocket launches, including mission specifications, booster landing and reusability.

The study attempts to answer the following questions:

- How do factors like launch frequency, launch site location, payload mass, and launch orbit affect booster landing outcomes?
- What unique insights can be gathered from SpaceX’s launch mission history and records?
- Which model between Logistic Regression, Support Vector Machine (SVM), Decision Tree or K-Nearest Neighbor (KNN) can achieve the best classification (binary) accuracy to predict booster landing outcome?

Section 1

Methodology



Methodology

Summary

- **Data Collection:** Information on Falcon 9 rocket launches was collected from SpaceX's API using GET requests and from Wikipedia by Web scrapping with GET requests and BeautifulSoup.
- **Data Wrangling** was performed to capture and analyze relevant data, deal with missing values, create a landing outcome classification label and variable, and to prepare the data for analysis and modeling.
- **Exploratory Data Analysis** with visualization was used to assess relationships between different variables, and EDA with SQL to find patterns in launch and mission records.
- **Interactive Visualization** was created with interactive Folium maps to examine details about the launch sites used by SpaceX, and a Plotly Dashboard to analyze launch statistics.
- **Supervised Machine Learning Models** were built, trained, and tuned to predict booster landing outcomes (binary classification) using Logistic Regression, Support Vector Machine (SVM), Decision Tree and K-Nearest Neighbor (KNN). All models were then evaluated to find the best performing model.

Data Collection – SpaceX API

Methodology

The study used the Requests, Pandas, NumPy, and Datetime libraries to collect data from SpaceX's API, and to clean and format it for further analysis.

The completed Data Collection SpaceX API notebook can be found on GitHub:

https://github.com/MonsieurVegas/IBM-Machine-Learning-Capstone/blob/main/Data%20Collection%20-%20SpaceX%20API_CK.ipynb

Process Flow Chart

1. Import Libraries and Define Auxiliary Functions
2. Use GET requests to request rocket launch data from the SpaceX API
3. Decode response as JSON using .json() & convert it to a dataframe using .json_normalize()
4. Create a dictionary launch_dict from the colums of the obtained data
5. Create Pandas dataframe from the dictionary launch_dict
6. Filter the dataframe, using the BoosterVersion column, to only include Falcon 9 launches
7. Calculate PayloadMass mean using .mean() & replace missing values in dataset with the mean
8. Export the dataframe data to a CSV file

Data Collection – Web Scraping

Methodology

The Requests, Pandas, and NumPy, and BeautifulSoup libraries were used to scrape Falcon 9 launch records from a Wikipedia web page.

The completed Data Collection Web scrapping notebook can be found on GitHub:

https://github.com/MonsieurVegas/IBM-Machine-Learning-Capstone/blob/main/Data%20Collection%20-%20Webscraping_CK.ipynb

Process Flow Chart

Import Libraries and Define Auxiliary Functions

↓

Use GET requests to request Falcon 9 launch data from from Wikipedia

↓

Create BeautifulSoup object from the HTML response

↓

Extract all column names from the HTML table header

↓

Create empty dictionary launch_dict with keys from extracted column names

↓

Parse data from HTML table rows with custom function & fill this data into launch_dict

↓

Create a Pandas dataframe from the launch_dict dictionary

↓

Export the dataframe data to a CSV file

Data Wrangling

Methodology

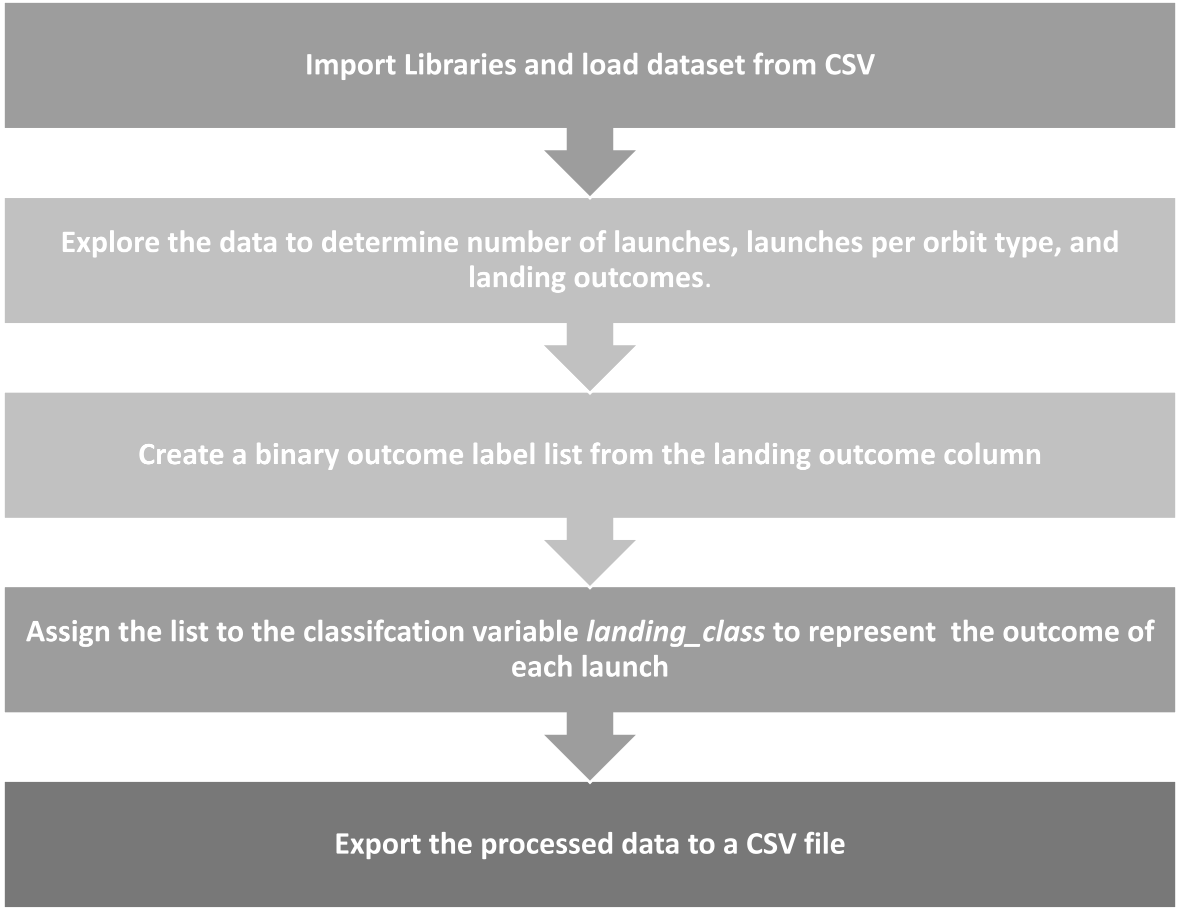
Pandas, and NumPy libraries were used to explore the data and to create a landing outcome variable which the study used as the classification variable.

The classification variable represents the outcome of each launch. If the value is zero, the booster did not land successfully; one means the booster landed successfully.

The completed Data Wrangling notebook can be found on GitHub:

- https://github.com/MonsieurVegas/IBM-Machine-Learning-Capstone/blob/main/Data%20Wrangling_CK.ipynb

Process Flow Chart



Exploratory Data Analysis with Visualization

Methodology

- The study used *matplotlib* and *seaborn* libraries to generate **scatter plots** to analyze how different variables like Launch Site, Payload mass, Flight Number, Orbit Type affect booster landing success rates, and to assess relationships between the variables.
- In addition, the study produced one **bar chart** to assess the relationship between Orbit Type and booster landing success rate, and a **line chart** to examine booster landing success rates during the observed period.
- The visual analytics helped prepare data feature engineering for machine learning prediction .

The completed EDA with Visualization notebook can be found on GitHub:

https://github.com/MonsieurVegas/IBM-Machine-Learning-Capstone/blob/main/EDA%20with%20Visualization_CK.ipynb

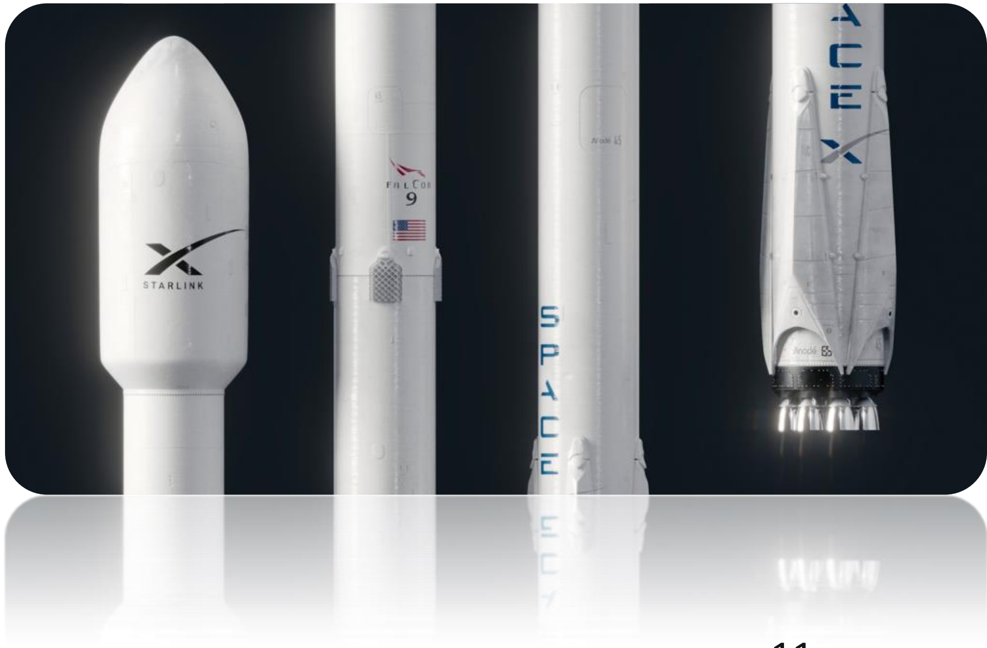
Exploratory Data Analysis with SQL

Methodology

- The SpaceX dataset was loaded into a table in a Db2 database, then various queries were executed using the *SQL extension* and *sqlite3* library to find information on launch sites, missions details such as booster versions, payloads, mission outcomes, and landing outcomes for selected time periods.

The completed EDA with SQL notebook can be found on GitHub:

https://github.com/MonsieurVegas/IBM-Machine-Learning-Capstone/blob/main/EDA%20with%20SQL_CK_sqllite.ipynb



Interactive Map with Folium

Methodology

- The study created an interactive Folium map to visually examine the relationship between booster landing outcomes and launch sites used by SpaceX.
- *folium.Circle* was added to the map to pinpoint the launch sites' location coordinates with highlighted circled bubbles and text labels of launch site names and number of rockets launched at each site.
- A *MarkerCluster* object and colored Folium markers were also added to assign colors (**green** or **red**) to launch outcomes based on class 1 (success) or 0 (failure) at each launch site.
- Lastly, colored *Folium PolyLines* were added to the map to show the distances between launch site CCAFS SLC-40 and its proximities like nearest coastline, railway, highway, and city

The completed EDA with SQL notebook can be found on GitHub:

https://github.com/MonsieurVegas/IBM-Machine-Learning-Capstone/blob/main/Launch%20Site%20Location%20Analysis%20with%20Folium_CK.jupyterlite.ipynb

Dashboard with Plotly Dash

Methodology

- The project created a Plotly Dashboard application using *callback functions* to visually analyze SpaceX Launch records in a **pie chart** and a **scatter point chart**.
- A Launch Site dropdown list *Input component* was added to allow users to examine the total count of booster landing success in a pie chart, broken down across all launch sites.
- When users select a specific site from the Launch Site dropdown list, the application renders a pie chart showing the ratio of successful and failed booster landings for that specific site.
- The Dash application also includes a scatter chart with a *dcc.RangeSlide*r enabling users to select different payload ranges to visualize the correlation between payload (kg), mission outcomes (class), and booster version.

The completed Plotly Dashboard Lab can be found on GitHub:

https://github.com/MonsieurVegas/IBM-Machine-Learning-Capstone/blob/main/SpaceX%20Launches%20Dashboard%20Plotly_CK.py

Predictive Analysis (Classification)

Methodology

The report used the Sklearn, Pandas, NumPy, Seaborn, and Matplotlib libraries to build, train, tune and evaluate Logistic Regression, Support Vector Machine (SVM), Decision Tree and K-Nearest Neighbor (KNN) Machine Learning models for predicting the classification of booster landing outcome.

The models were evaluated by calculating their accuracy on the test data and accuracy on validation data, and by plotting a Confusion Matrix for each model.

The completed Machine Learning Prediction Analysis notebook can be found on GitHub:

https://github.com/MonsieurVegas/IBM-Machine-Learning-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction%20Analysis_CK.jupyterlite.ipynb

Process Flow Chart

Import Libraries and Define Auxiliary Functions

↓

Create a NumPy array from the Class column of the wrangled data

↓

Standardize the data with StandardScaler, then fit and transform the data.

↓

Split the data using the function train_test_split

↓

Build each selected model and use the function GridSearchCV to find its best parameters

↓

Calculate each model’s accuracy on the validation data using the data attribute best_score_

↓

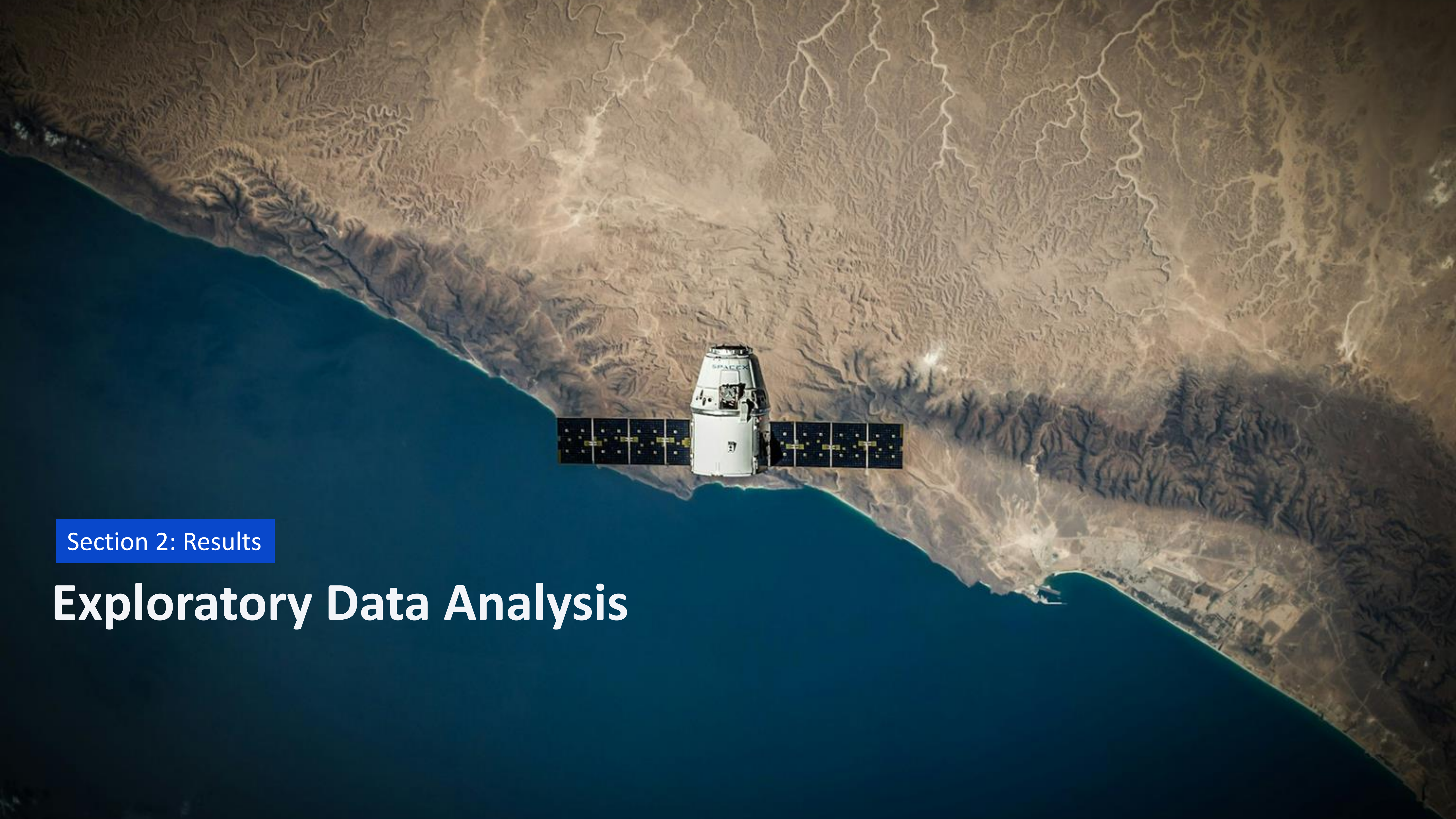
Calculate each model’s accuracy on the test data using .score()

↓

Plot and examine the Confusion Matrix for each model

↓

Compare the accuracy of all models to identify the best performing model



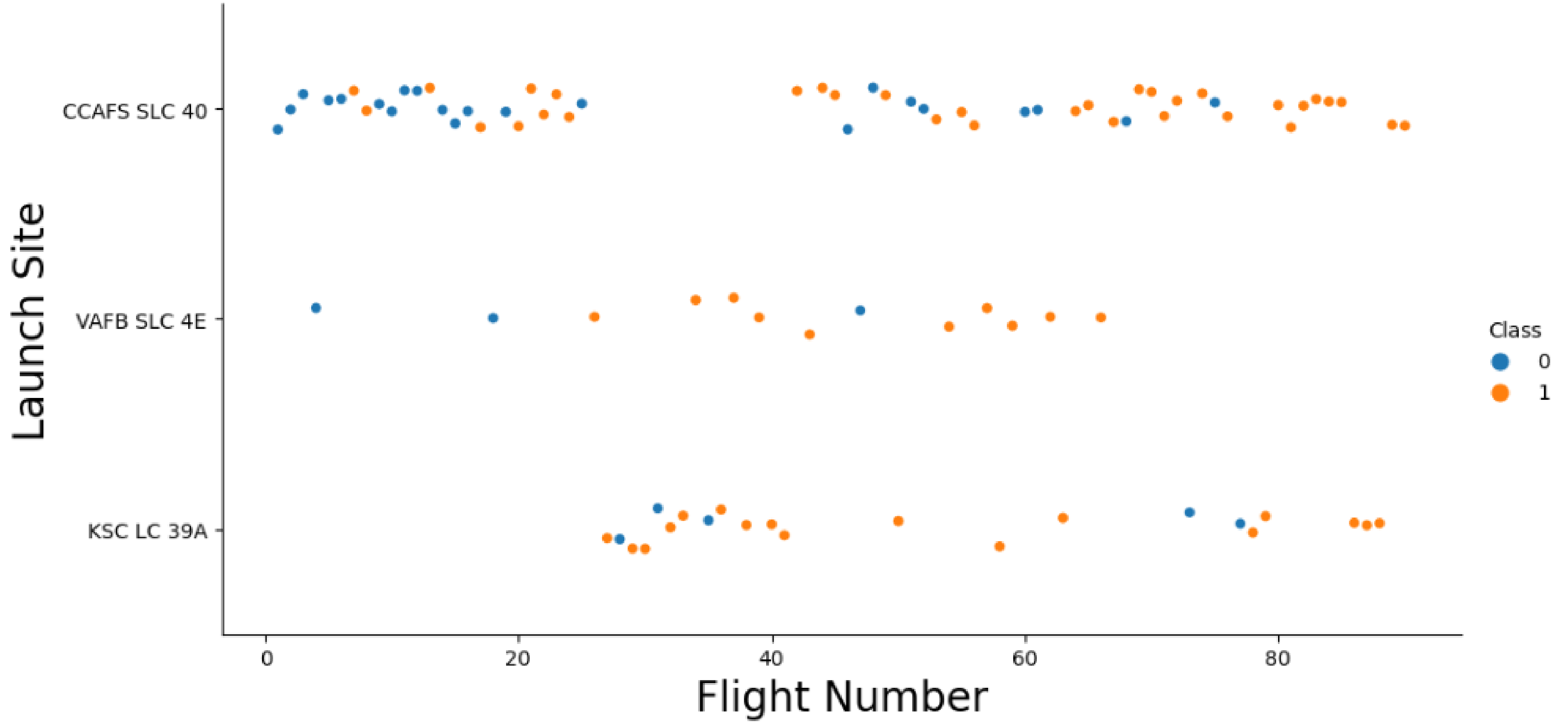
Section 2: Results

Exploratory Data Analysis

Flight Number vs. Launch Site

Exploratory Data Analysis with Visualization

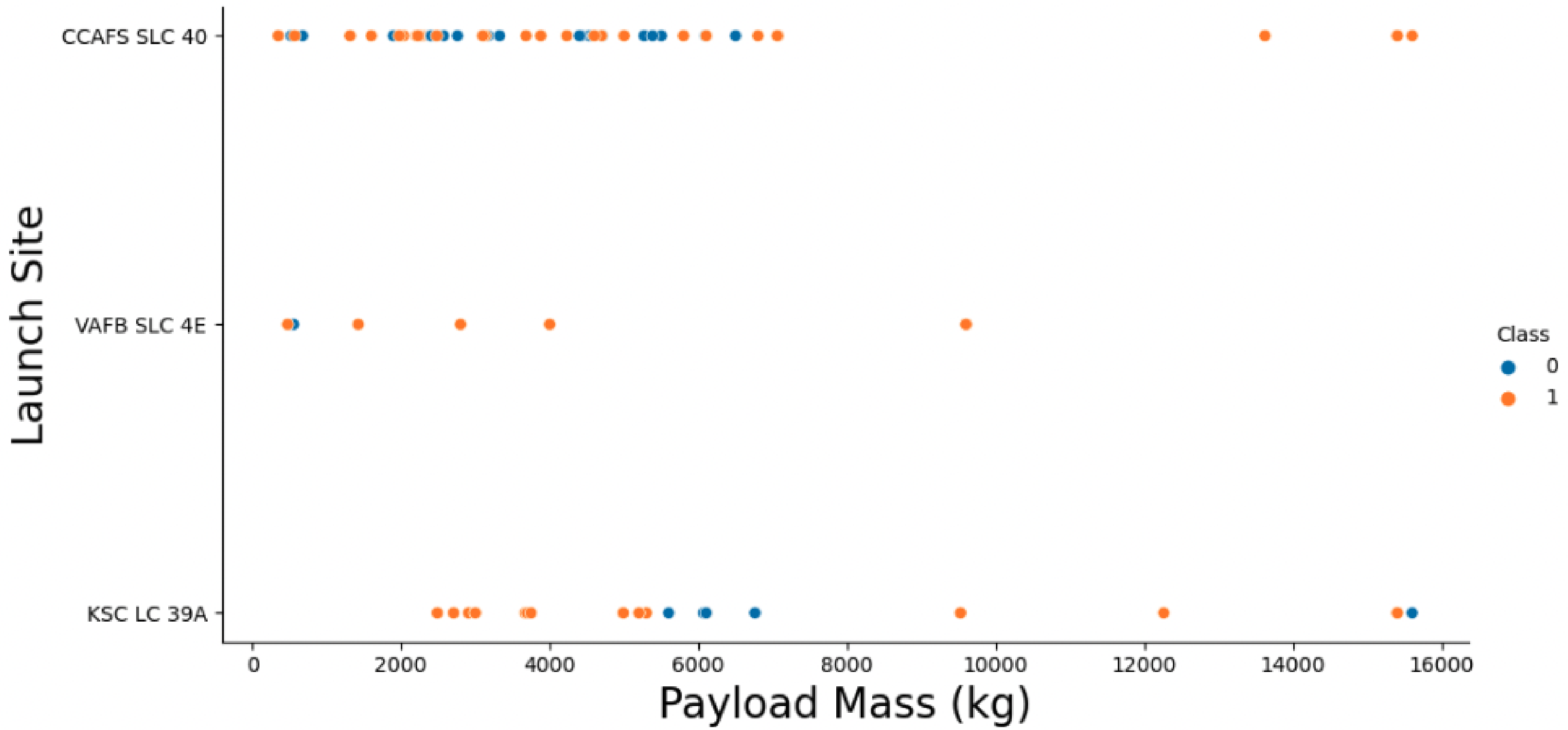
- Initial flights had low booster landing success rates(Class 0 indicates landing failure, Class 1 indicates success). The success rate increased as the number of flights increased.
- Roughly half of the total launches were from the **CCAFS SLC 40** launch site. Launches from the **KSC LC 39A and VAFB SLC 4E** sites had higher landing success rates than CCAFS SLC 40. This preliminary analysis indicates a correlation between launch site and booster landing outcome.



Payload vs. Launch Site

Exploratory Data Analysis with Visualization

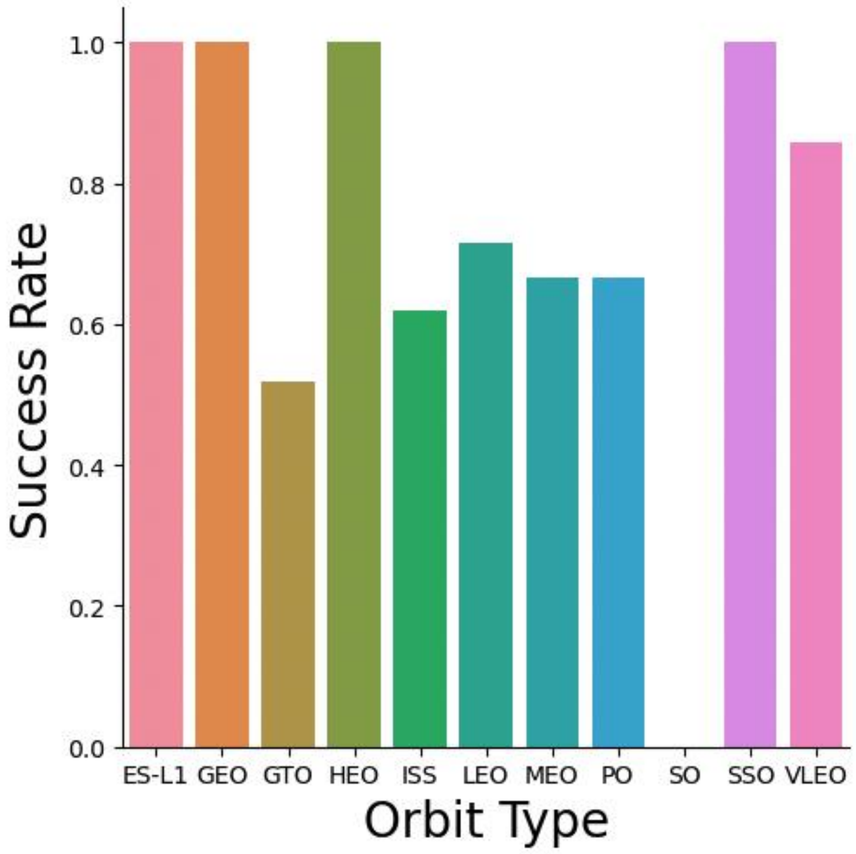
- Launches from **CCAFS SLC-40 and KSC LC-39A** sites with payloads greater than 8,000 kg had higher booster landing success rates compared to smaller payloads.
- The **VAFB SKC 4E** site had not launched any boosters with payload mass greater than 10,000 kg.



Launch Success Rate vs. Orbit Type

Exploratory Data Analysis with Visualization

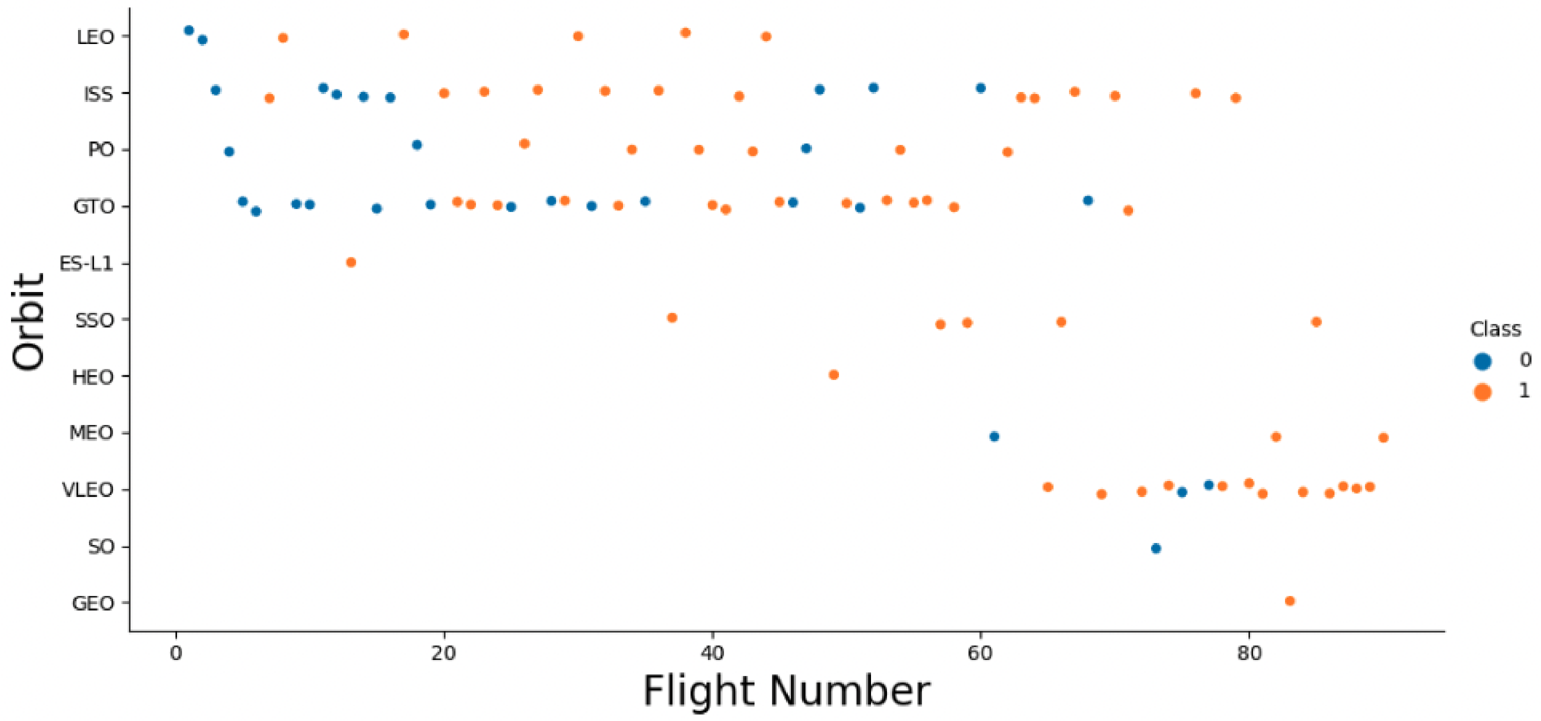
- Launches into **ES-L1, GEO, HEO and SSO** orbits achieved the highest landing success (100% Success Rate).
- Launches into **LEO and VLEO** orbits attained above average to strong success rates (70%-85%).
- Launches into **GTO, and ISS** orbits achieved middling booster landing success (50% to 60% Success rate) , despite there being a greater number of flights into these two orbits compared to other orbits.



Flight Number vs. Orbit Type

Exploratory Data Analysis with Visualization

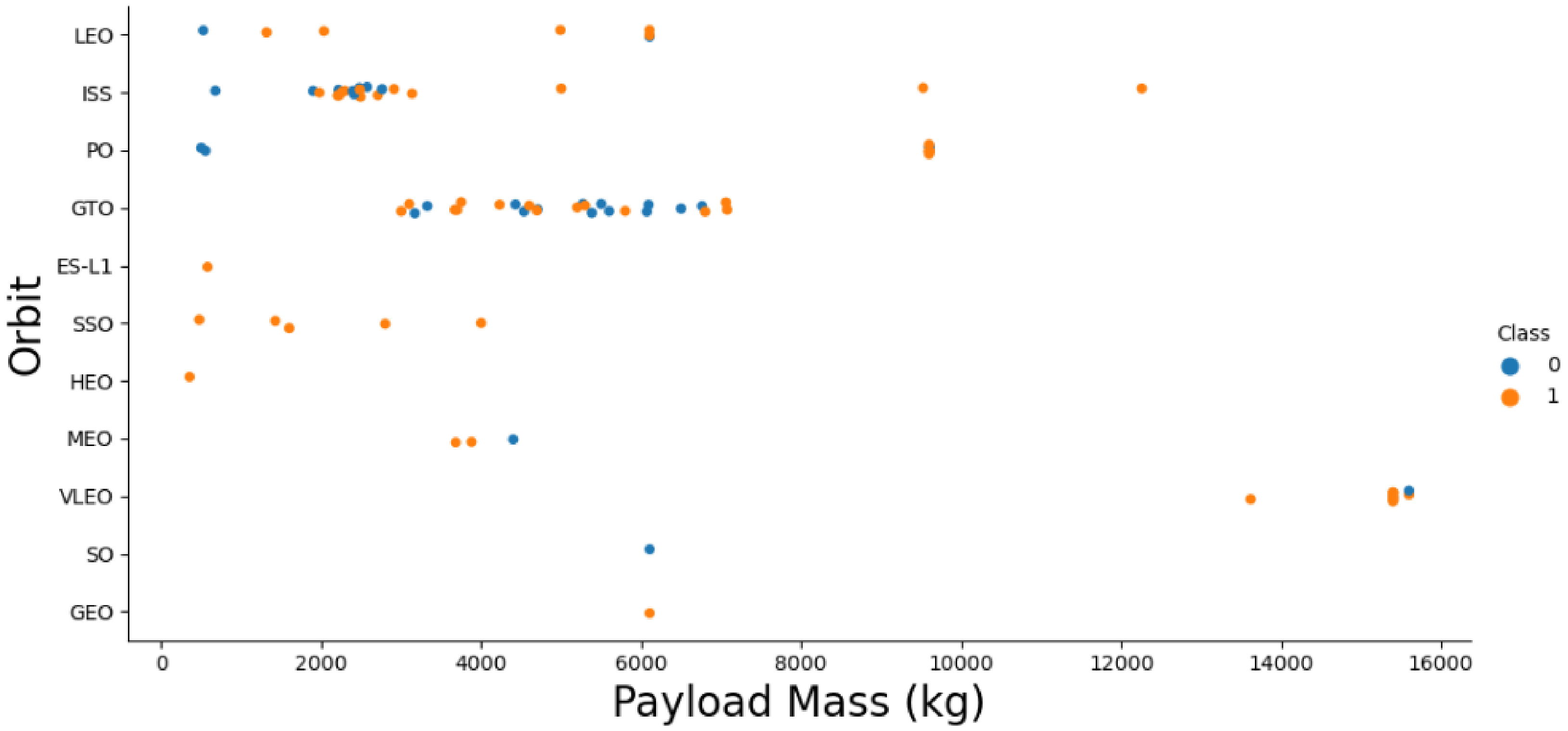
- The LEO orbit shows a trend of increasing booster landing success as the number of flights increases;
- However, there does not seem to be a similarly strong correlation between the number of flights and landing success for launches into GTO orbit.



Payload vs. Orbit Type

Exploratory Data Analysis with Visualization

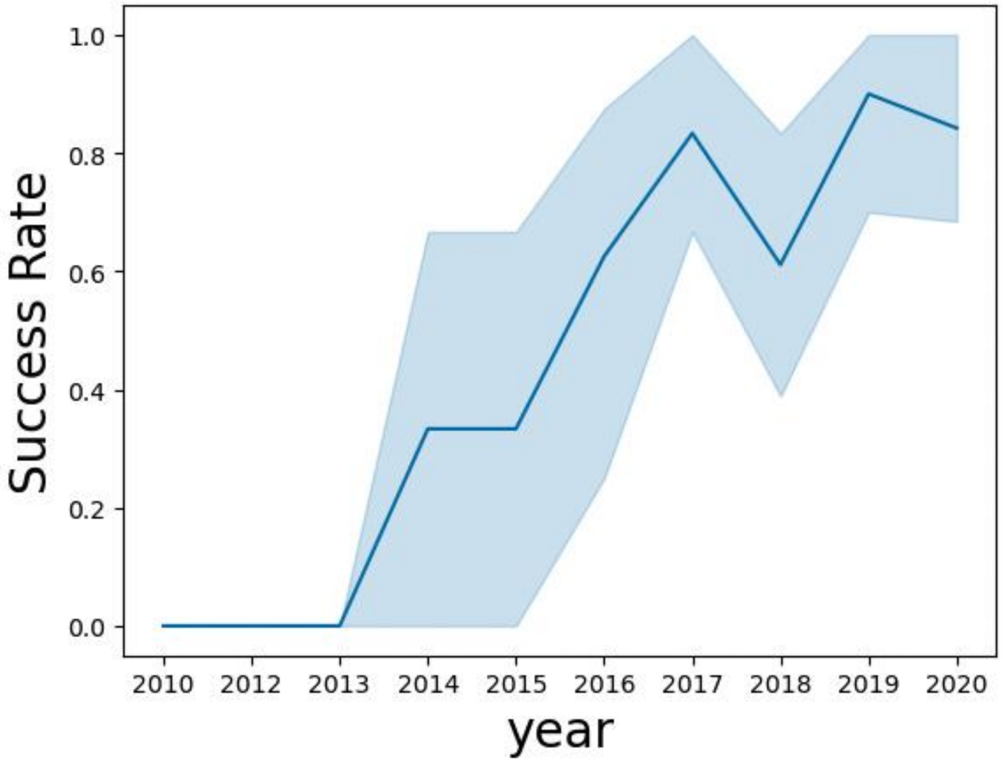
- Boosters carrying payload masses heavier than 2,000 kg launched into SSO, PO, LEO and VLEO orbits had high successful landing rates .
- However, boosters carrying these payloads into GTO orbit have mixed landing results (success vs. unsuccessful).



Launch Success Yearly Trend

Exploratory Data Analysis with Visualization

- The booster landing success rate steadily increased from 2013 until 2017 (stable in 2014).
- The success rate sharply decreased from 2017 to 2018, but rebounded in 2019.
- Overall, the landing success rate rose over time during the period from 2010 to 2020.



Launch Sites Information

Exploratory Data Analysis with SQL

Launch Site Names

- In its mission history SpaceX has used launch sites **CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, and VAFB SLC-4E**
- The SpaceX dataset was loaded into a table in a Db2 database, one query executed *SELECT DISTINCT* command to find the unique names from the database.

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

```
%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

5 Records with Launch Site 'KSC'

- Another query searched for 5 records where launch sites' names start with "KSC" by executing *SELECT* command and filtering for string "KSC" with *LIKE*. In addition to setting the *LIMIT* parameter to only include 5 records in the query.

Display 5 records where launch sites begin with the string 'KSC'

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

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-03-16	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-05-01	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
2017-05-15	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt

SpaceX Mission History & Highlights

Exploratory Data Analysis with SQL

- The total payload carried by boosters for NASA is **45,596 kg**

- The query executed *SELECT*, and *SUM* commands to aggregate the payload mass, along with filtering for Customer "NASA (CRS)" from the SQL database.

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

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

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

SUM(PAYLOAD_MASS__KG_)
45596

- The average payload mass carried by booster version F9 v1.1 is **2,928 Kg**

- The query executed *SELECT*, and *AVG* commands to average the payload mass, along with filtering for Booster Version "F9 v1.1" from the SQL database.

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

```
%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

SpaceX Mission History & Highlights

Exploratory Data Analysis with SQL

- The first successful landing in Ground pad took place on **December 22, 2015**.
- Executed *SELECT* command, *min(Date)* operation, and *WHER*E command to choose the first date for the given landing outcome.

```
%sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)';

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

min(DATE)
2015-12-22

- The Boosters **F9 FT** versions (**B1022, B1026, B1021.2, and B1031.2**) have all successfully landed on drone ship and with payload mass between 4,000 Kg and 6,000 Kg
- Executed *SELECT* and *WHERE* commands, and </> operation to choose boosters carrying specified payload mass for the given landing outcome.

```
%sql SELECT BOOSTER_VERSION from SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)' \
and  PAYLOAD_MASS__KG_  >4000 and PAYLOAD_MASS__KG_ <6000;

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

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

Summary of Mission Outcomes

Exploratory Data Analysis with SQL

- The total number of successful mission outcomes is 99. There was 1 failed mission (Failure in Flight) and 1 mission success with payload status unclear.
- Executed *SELECT, COUNT, and GROUP BY* to select, tally and group the number of mission outcomes observed.

List the total number of successful and failure mission outcomes

```
%sql SELECT MISSION_OUTCOME, COUNT(*) as total_number \
from SPACEXTBL GROUP BY MISSION_OUTCOME;
```

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

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

Maximum Payloads Carried

Exploratory Data Analysis with SQL

- The booster versions listed below have carried the maximum payload mass within the time period observed.
- Executed *SELECT* and *WHERE* commands, and max operator to query boosters which carried maximum payloads.

```
%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

Selected Landing Records in Year 2017

Exploratory Data Analysis with SQL

- The records listed in the SQL query below display the months, dates, booster versions, and launch sites of successful landings in ground pad during the year 2017.
- Executed *SELECT* and *WHERE* commands, and *substr(Date,,)*operators to query boosters, launch sites, and 2017 dates information for the specified landing outcome.

```
%sql SELECT substr(Date,6,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, [Landing_Outcome] \
FROM SPACEXTBL \
where [Landing_Outcome] = 'Success (ground pad)' and substr(Date,0,5)='2017';
```

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

month	Date	Booster_Version	Launch_Site	Landing_Outcome
02	2017-02-19	F9 FT B1031.1	KSC LC-39A	Success (ground pad)
05	2017-05-01	F9 FT B1032.1	KSC LC-39A	Success (ground pad)
06	2017-06-03	F9 FT B1035.1	KSC LC-39A	Success (ground pad)
08	2017-08-14	F9 B4 B1039.1	KSC LC-39A	Success (ground pad)
09	2017-09-07	F9 B4 B1040.1	KSC LC-39A	Success (ground pad)
12	2017-12-15	F9 FT B1035.2	CCAFS SLC-40	Success (ground pad)

Landing Outcomes from 2010-06-04 to 2017-03-20

Exploratory Data Analysis with SQL

- The SQL query below ranks the count of landing outcomes between the dates 2010-06-04 and 2017-03-20, in descending order.
- Executed SELECT, COUNT, GROUP BY, and ORDER BY (DESC) commands to select, tally, group, and rank by the number of landing outcomes, along with WHERE command to filter for dates in the specified range.

```
%sql SELECT LANDING_OUTCOME, COUNT(*) as count_outcome \
FROM SPACEXTBL \
WHERE DATE between '2010-06-04' and '2017-03-20' GROUP BY LANDING_OUTCOME \
ORDER BY count_outcome DESC;
```

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

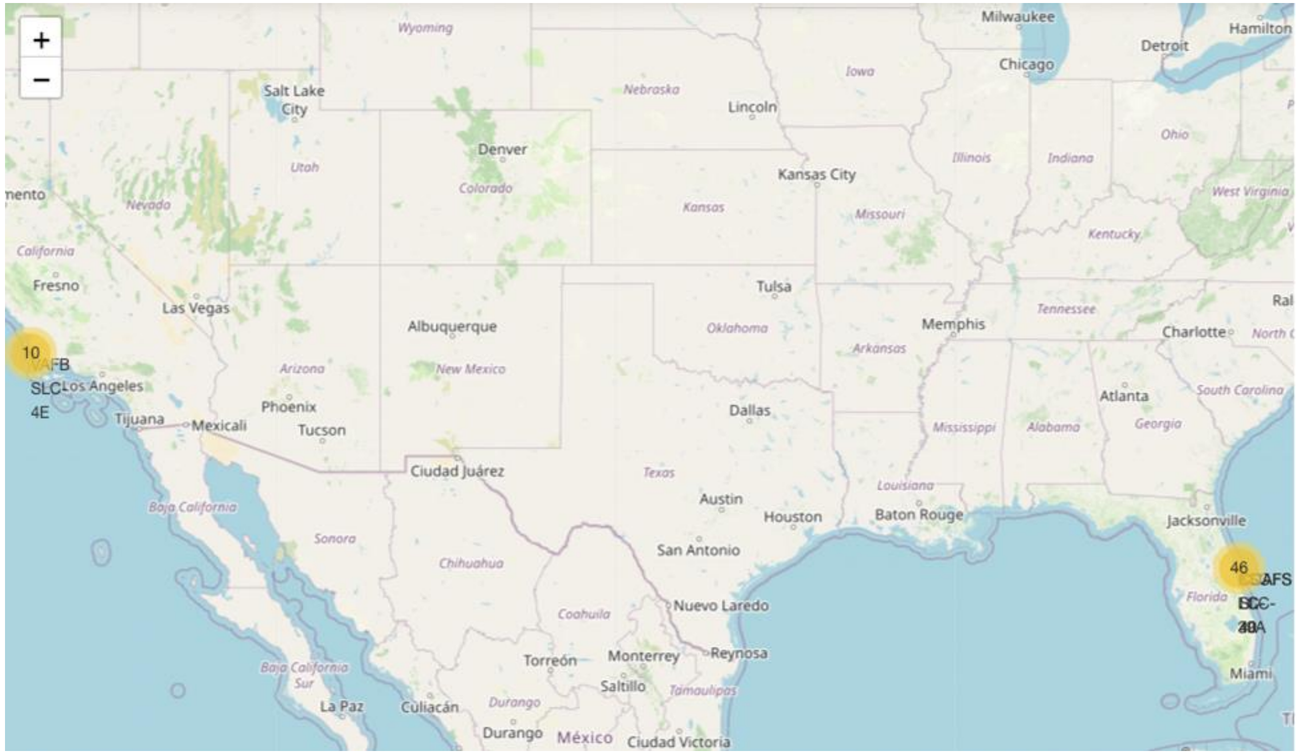
Landing_Outcome	count_outcome
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3: Results

Launch Sites Analysis with Folium

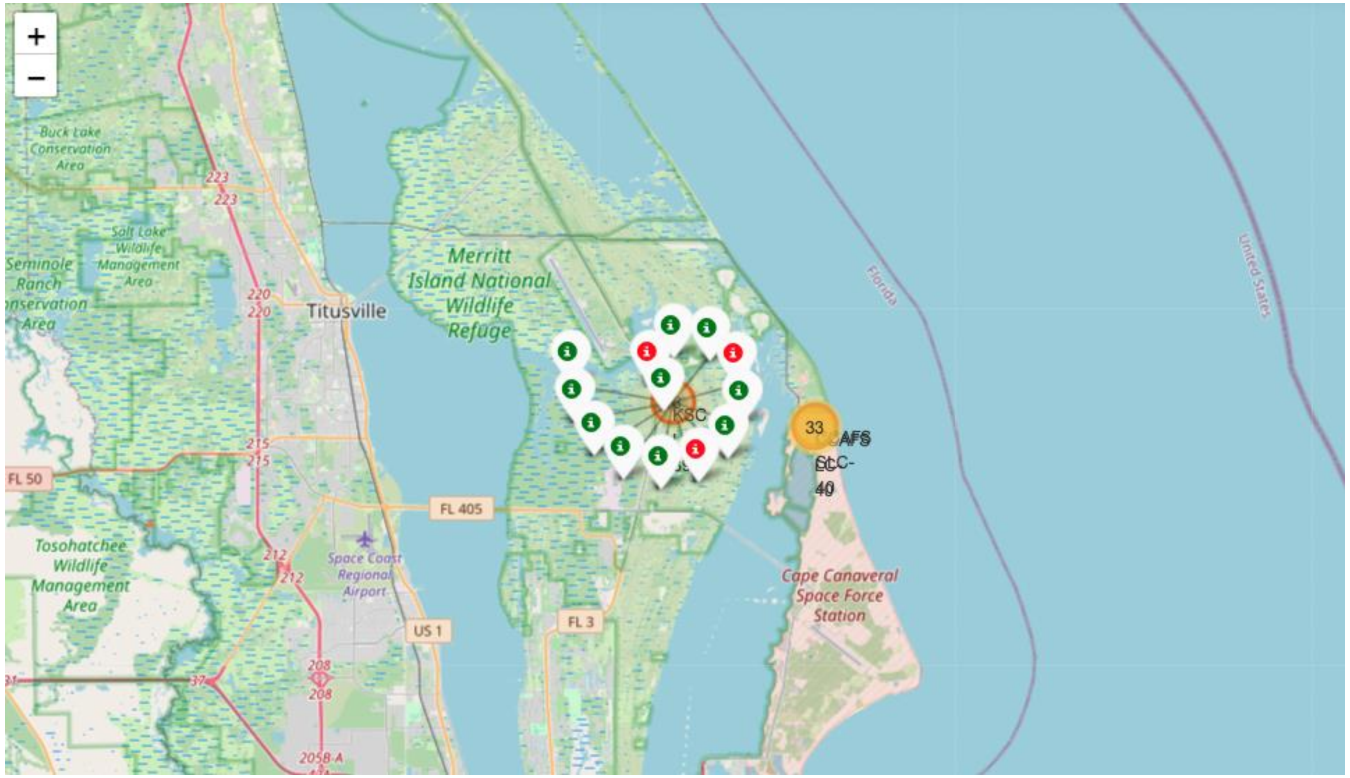
Map of Launch Sites Used by SpaceX

- Three launch sites (CCAFS LC-40, CCAFS SLC-40, KSC LC-39A) are located **on the coast of Florida**. One launch site (VAFB SLC-4E) is located **on the coast of California**.
- All sites are close to the equator line relative to the continental USA. The **Florida sites are closer to the equator than the California site.** Proximity to the equator helps to reduce fuel and energy costs for Rocket launches due to the additional boost from the Earth's rotation.



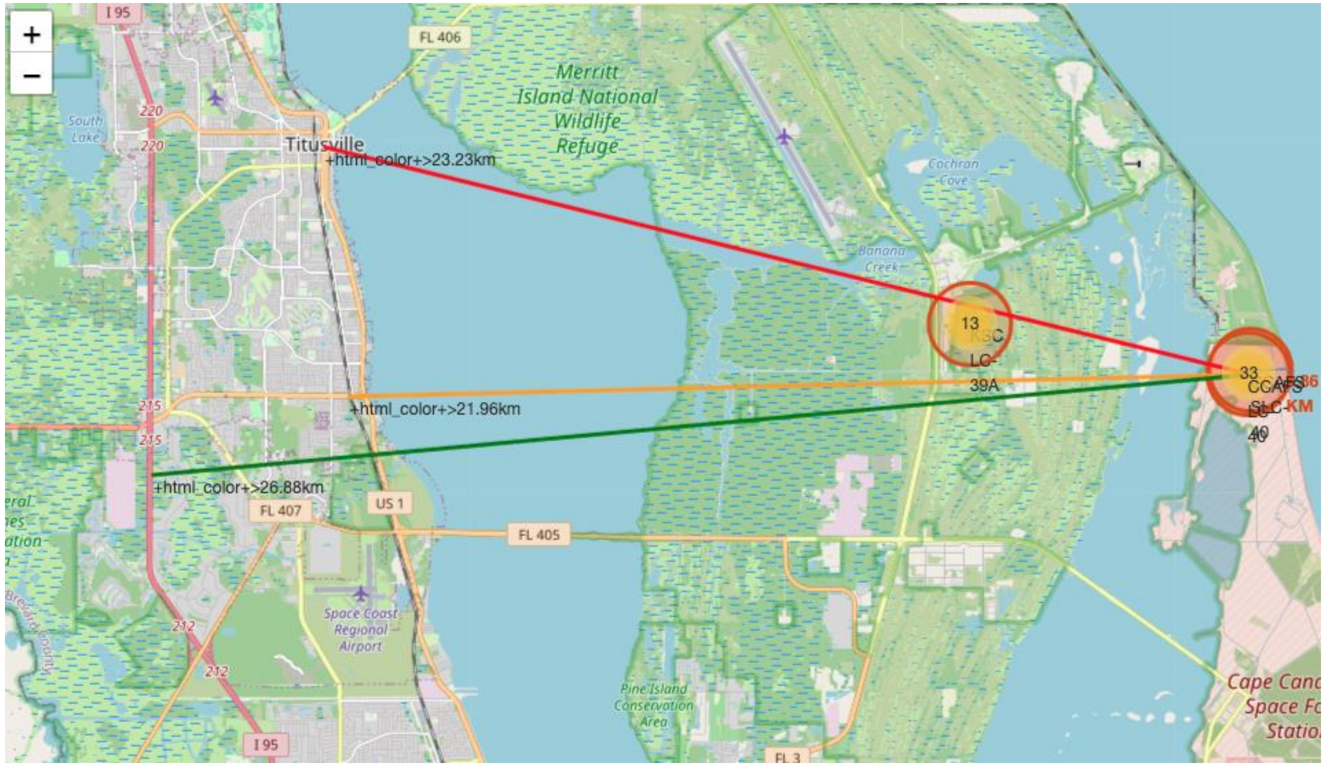
Launch Outcomes at KSC LC-39A site

- **KSC LC-39A** had the highest landing success rate **(76.92%)** among all launch sites for the period observed.
- The **green** markers on the map indicate successful landings, while the **red** markers indicate failed landings.



Distance to Proximities of CCAFS SLC-40 site

- The selected **CCAFS SLC-40** site **is 0.86 km** from the nearest coastline, and **23.23 km away** from the nearest city. The distances are an important safety measure to minimize the risk of damage to habited areas from debris in the event of a rocket launch failure.
- This site is **21.96 km** from the nearest railway and **26.88 km** from the nearest highway, this allows for the transportation of people and materials in support of launch activities while maintaining safety and security around the launch site.

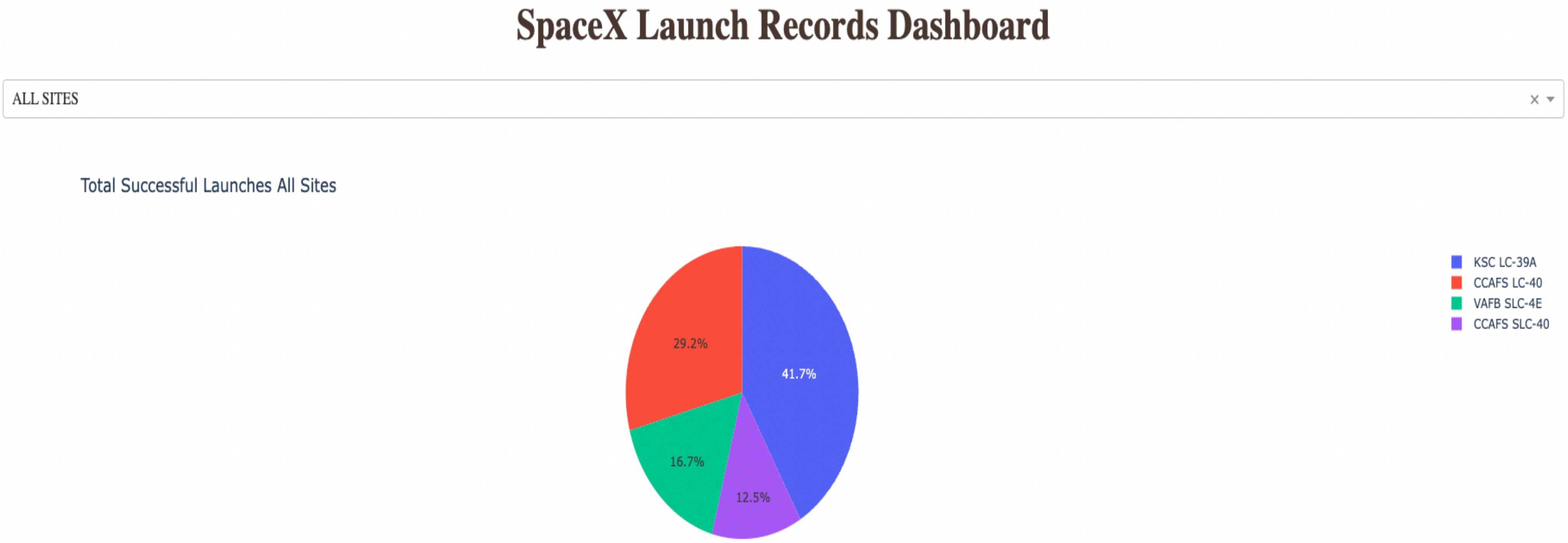


Section 4: Results

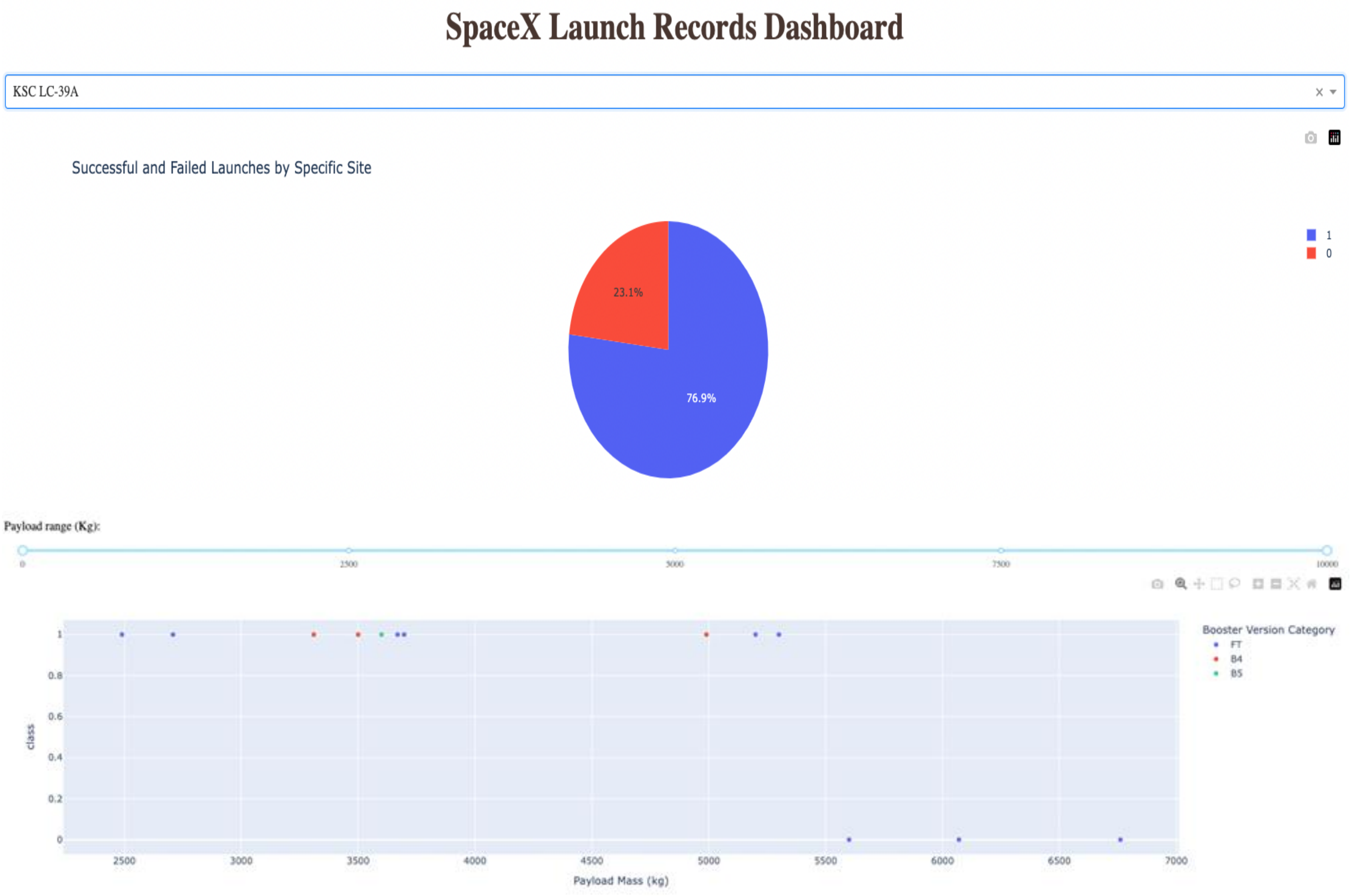
Launch Records Dashboard with Plotly

Launch Records for All Sites

- **KSC LC-39A** had the highest number of successful booster landing outcomes amongst all launch sites for the selected period.



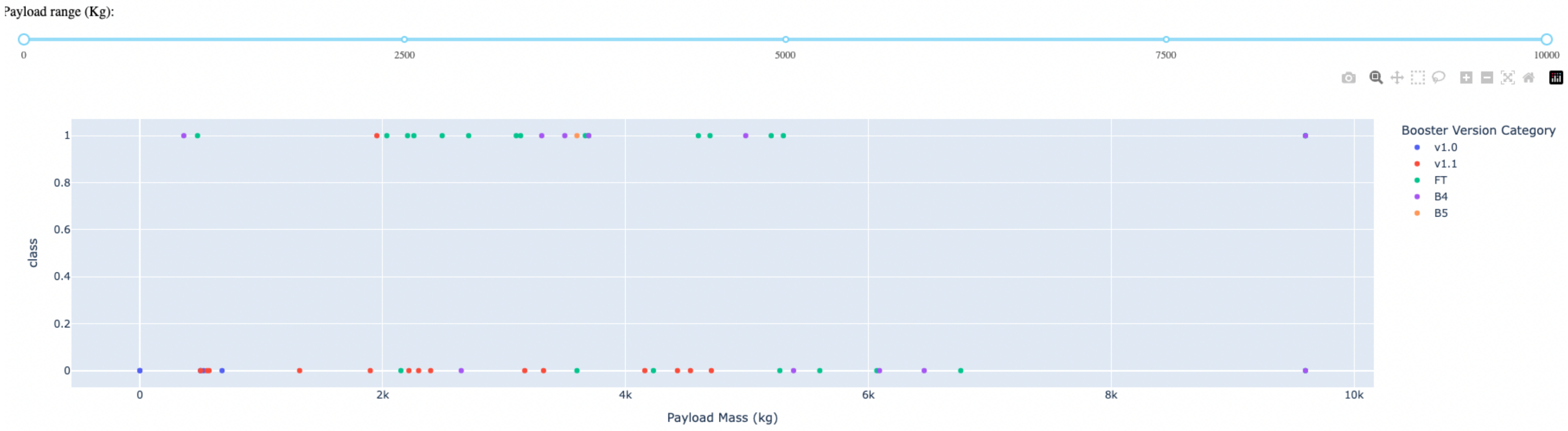
Launch Records for KSC LC-39A Site



- The selected **KSC LC-39A** site's landing success rate was **76.9%**.
- At the **KSC LC-39A site,** boosters carrying payloads masses ranging from **2,000 kg to 5,500 kg** all landed successfully.

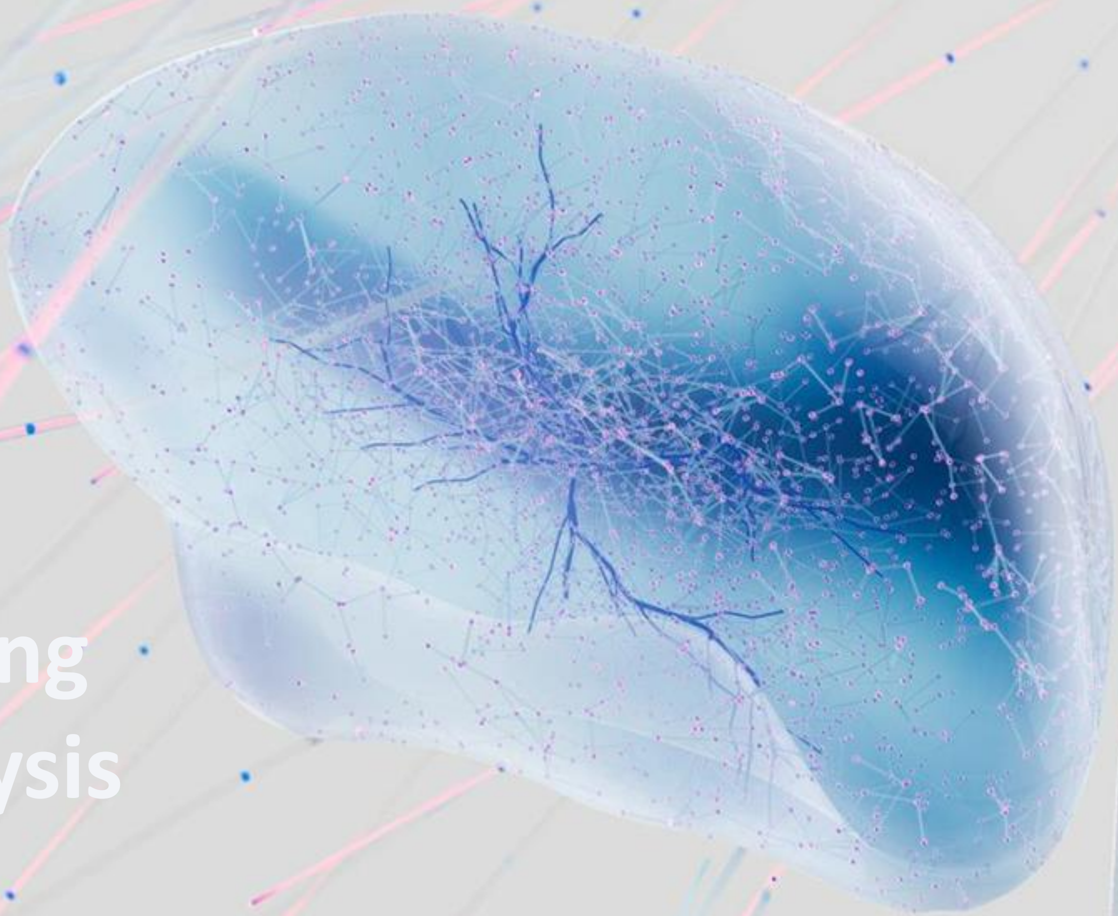
Launch Records by Payload and Booster Version

- Across all sites, boosters carrying payloads ranging from **2,000 kg to 5,500 kg** had higher landing success rates.
- While boosters with payloads between **6,000 kg to 9,000 kg** had the lowest success rate.
- Overall, **F9 Booster version FT** had the highest success rate out of all booster versions.



Section 5: Results

Machine Learning Predictive Analysis



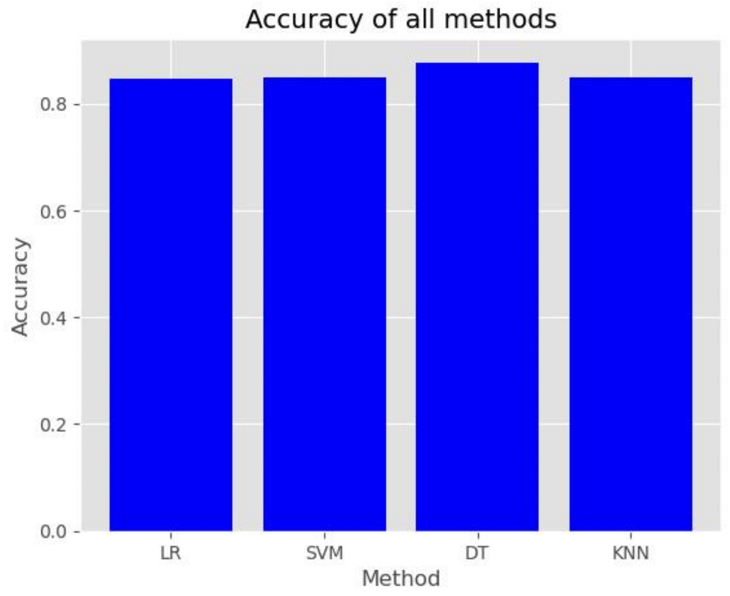
Machine Learning Classification Accuracy

- The study used four different supervised machine learning algorithms; K-Nearest Neighbors, Support Vector Machine, Logistic Regression, and Decision Tree to predict the outcome (binary class) of the booster landing.
- The analysis trained each model and found its best hyper parameters by using the *GridSearchCV* function to determine the highest classification accuracy of each model.
- The **Decision Tree model (best score 87.68%)** slightly outperformed the other models.

```python
models = {'KNeighbors':knn_cv.best_score_,
              'DecisionTree':tree_cv.best_score_,
              'LogisticRegression':logreg_cv.best_score_,
              'SupportVector': svm_cv.best_score_}

bestalgorithm = max(models, key=models.get)
print('Best model is', bestalgorithm,'with a score of', models[bestalgorithm])
if bestalgorithm == 'DecisionTree':
    print('Best params is :', tree_cv.best_params_)
if bestalgorithm == 'KNeighbors':
    print('Best params is :', knn_cv.best_params_)
if bestalgorithm == 'LogisticRegression':
    print('Best params is :', logreg_cv.best_params_)
if bestalgorithm == 'SupportVector':
    print('Best params is :', svm_cv.best_params_)
```

```
Best model is DecisionTree with a score of 0.8767857142857143
Best params is : {'criterion': 'gini', 'max_depth': 18, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples
_split': 2, 'splitter': 'best'}
```



Confusion Matrix

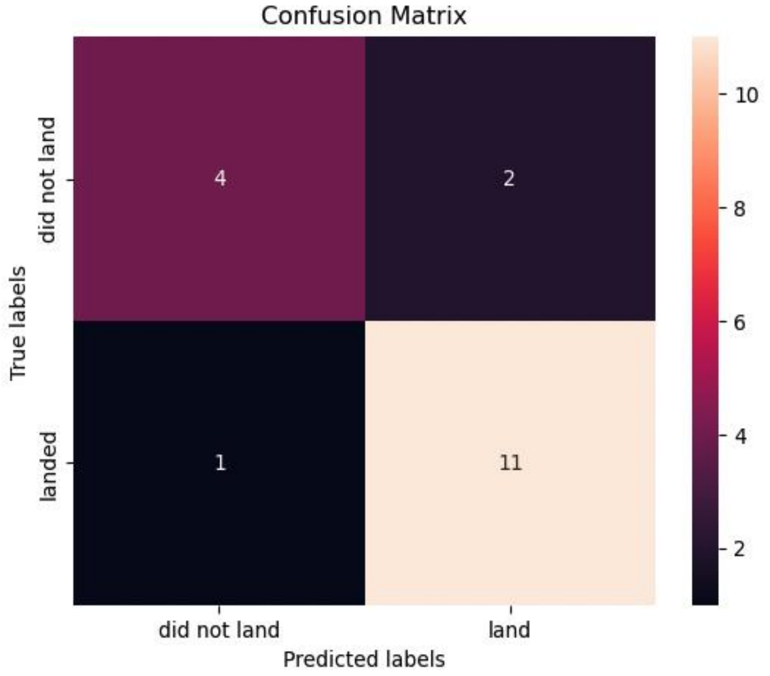
- The Confusion Matrix for Decision Tree evaluates the model's ability to correctly predict or separate the classes.
- Confusion Matrix Outputs: **11 True positives** , **4 True negative**, **2 False positives**, **1 False Negative.**
- False Positives are a problem
- **Precision** is a measure of the accuracy, provided that a class label has been predicted.

Precision = TP / (TP +FP) = **0.846153846**

- **Recall** is the true positive rate.

Recall = TP / (TP + FN) = **0.916666667**

```
yhat3 = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat3)
```



Confusion Matrix

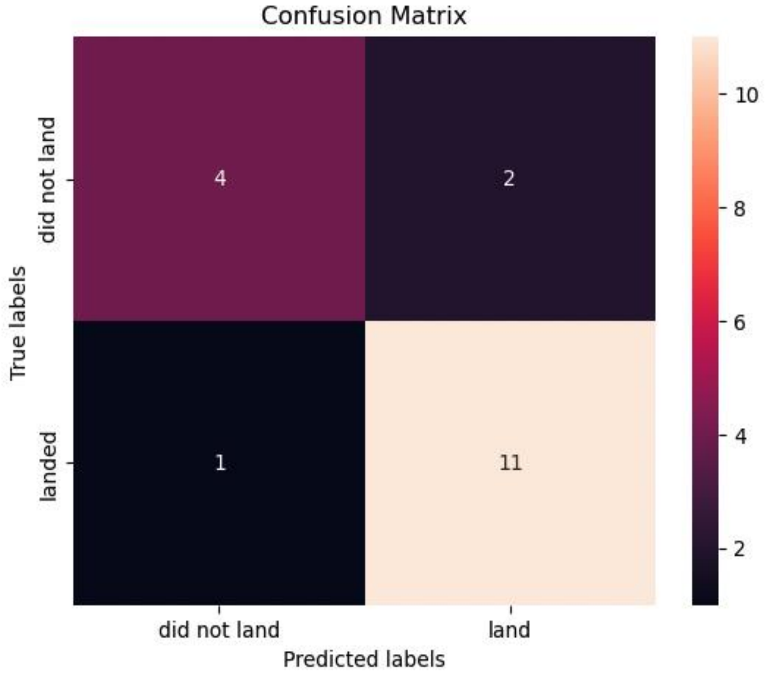
- The **F1 score** is the harmonic average of the precision and recall, where an F1 score reaches its best value at 1 and its worst at 0.

F1 Score = 2 * (Precision * Recall) / (Precision + Recall) = **0.8800**

- **Accuracy** compares the actual values in the test set with the values predicted by the model.

Accuracy = (TP + TN) / (TP + TN + FP + FN) = **0.833333333**

```
yhat3 = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat3)
```



Conclusion

The study found that the Decision Tree model slightly performed better in its classification accuracy of predicted booster landing outcomes compared to the Logistic Regression, Support Vector Machine, and K-Nearest Neighbor models.

Exploring the SpaceX data set showed that the rate of successful booster landing increased as the number of launches increased during the period observed.

All launch sites (CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, and VAFB SLC-4E) used by SpaceX were located in coastal areas, close to the equator line relative to the continental USA.

Launches from the KSC LC-39A launch site had higher rate of successful booster landings compared to the other sites.

The study's analysis indicates a correlation between launch site and landing outcome. Additionally, a launch site's location impacts launch mission specifications like payload and intended launch orbits – because some location are better/worse suited for launches into certain orbits. For example, the CCAFS sites are not suitable for polar orbit launches.

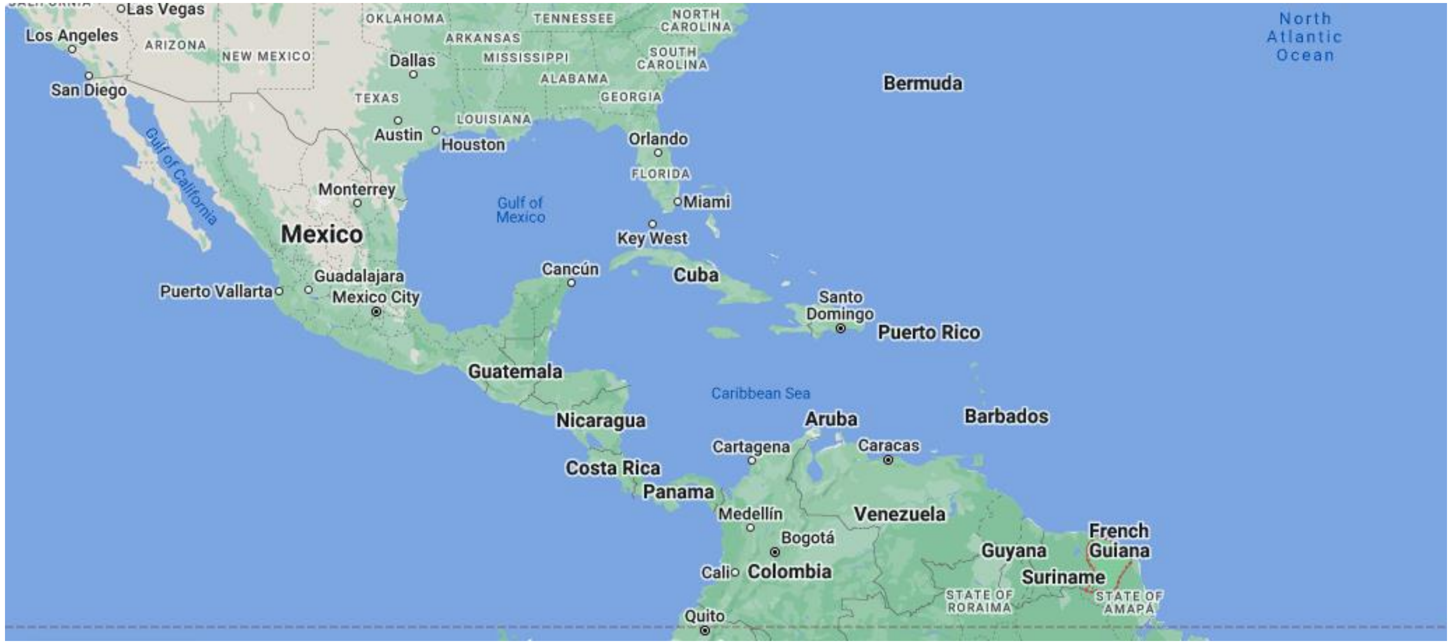
It should be noted, that the launch sites used by SpaceX are owned by the United States Space Force. In 2014, SpaceX signed a 20-year exclusive lease agreement with NASA for the KSC LC-39A site. The study found that SpaceX had carried the largest total payload for NASA compared to its other customers during the period.

Conclusion

Discussion

The major launch site used by the European Space Agency is the Centre Spatial Guyanais (CSG) located in Kourou, French Guiana. The Guiana spaceport has a notable advantage over US sites due to its near equatorial position, 5 degrees north of the equator line. CSG's location is more ideal for launches into geostationary orbit.

Following the invasion of Ukraine by Russia in 2022, commercial launches of Soyuz-2 Russian rockets were suspended at CSG. As a result, OneWeb, a major customer using Soyuz launches, switched to SpaceX. Arianespace, a French Company, and the primary user of the CSG spaceport, is currently developing its Ariane 6 reusable rocket, for which it received subsidies from European government agencies, to compete with SpaceX's Falcon 9.



Conclusion

Discussion

Geopolitical alliances will play a significant role in shaping the evolution of the commercial space industry by influencing access to private and public funding, access to launch facilities and to segments of customers. In addition, geopolitics will impact the development and manufacturing of new rockets.

More research should be done to include SpaceX's launch data from the subsequent five years after the 2010-2020 period. This study should be repeated in five-year intervals thereafter.

Future iteration of this study would benefit from a larger data set, and should examine other predictive models like Gradient Boosting Classifier and XGBoost to compare their performance with the selected models in the initial study.