IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Hussain Fabro
4/17/2023

Outline

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

Executive Summary

- **Summary of methodologies**
  - Data Collection
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Visualization
  - Interactive Mapping with Folium
  - Predictive Analysis (Classification)
- **Summary of all results**
  - Data Analysis Results
  - Predictive Analysis Results

Introduction

- **Project background and context**
  - Welcome to this presentation on our data science research to forecast the success of SpaceX Falcon 9 rocket landings in the first stage. As a data scientist at a startup, we were entrusted with competing with SpaceX by delivering accurate predictions for rocket launch landings. Our goal was to evaluate the data using data science techniques and machine learning algorithms to construct a predictive model that would enable our organization make better informed bids versus SpaceX.
  - We will walk you through our data science journey, covering data collecting and wrangling techniques, exploratory data analysis, predictive analysis methodology, and model outcomes, in this presentation. We will also present novel data insights and how we used them.

- **Problems you want to find answers**
  - Forecast whether the SpaceX Falcon 9 rocket's first stage will land successfully.
  - Showcasing the data science methodology involving data collection, data wrangling, exploratory data analysis, data visualization, model development, model evaluation, and reporting.
  - Helping a startup compete with SpaceX by providing them with more informed bids for rocket launches.

Section 1

Methodology

Methodology

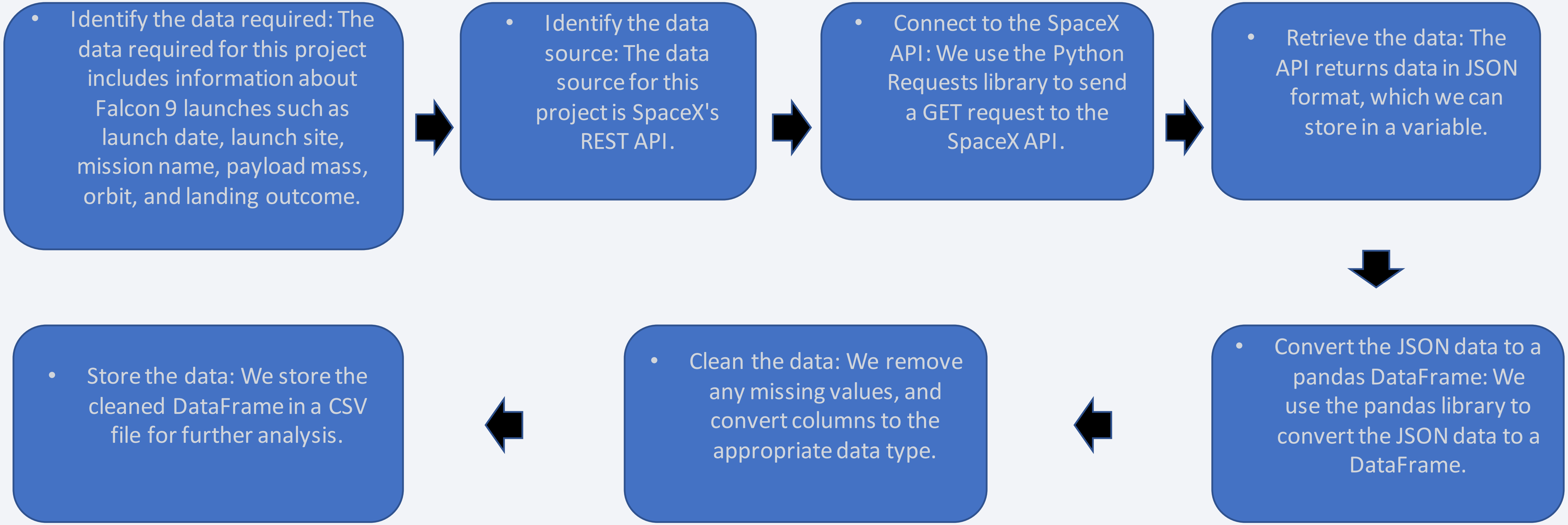
- **Executive Summary**
- **Data collection methodology:**
  - This project's data was gathered from publicly accessible sources such as online databases and APIs.
- **Perform data wrangling**
  - To prepare it for analysis, the obtained data was processed through many processes of data wrangling
- **Perform exploratory data analysis (EDA) using visualization and SQL**
- **Perform interactive visual analytics using Folium**
- **Perform predictive analysis using classification models**
  - To predict the success of the first stage of SpaceX Falcon 9 rocket landing, classification models were constructed, modified, and tested using several machine learning methods like Logistic Regression.

Data Collection

- Describe how data sets were collected.
    - Web scraping of a Wikipedia article titled "List of Falcon 9 and Falcon Heavy Launches" was used to acquire data for this experiment. The web scraping was done with the Python BeautifulSoup package to retrieve a table with the Falcon 9 launch records. Following that, the table was processed and transformed into a Pandas data frame for further analysis. Each launch's data includes information such as the launch date, mission name, launch site, payload tonnage, orbit, and outcome.

Data Collection – SpaceX API

Flowchart



[Github SpaceX API]

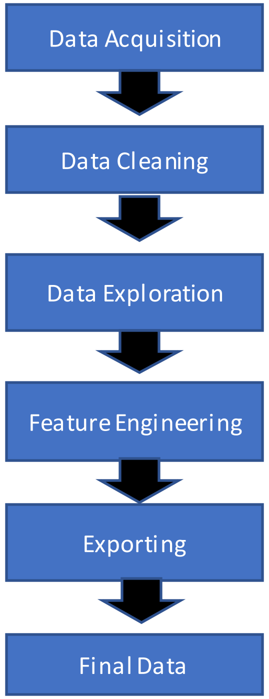
Data Collection - Scraping

- First, we use the Python **requests** library to send a GET request to the Wikipedia page URL and retrieve the HTML content of the page.
- Next, we use the Python **BeautifulSoup** library to parse the HTML content and extract the table containing the Falcon 9 launch records using its class name and table structure.
- We then create an empty Pandas data frame with the required column names and data types to store the extracted launch records.
- Using **BeautifulSoup**, we loop through the rows of the extracted HTML table and extract the launch details such as the launch date, mission name, launch site, landing outcome, etc. for each launch.
- For each launch record, we add a new row to the Pandas data frame with the extracted details.
- Finally, we save the extracted data frame as a CSV file for further analysis.

Github Webscraping

Data Wrangling

- Describe how data were processed
  - Data acquisition:
    - Obtained the SpaceX dataset from Kaggle as a CSV file.
    - Loaded the data into a Pandas dataframe using the read_csv() function.
  - Data cleaning:
    - Checked for and removed any duplicate rows using the drop_duplicates() function.
    - Checked for and removed any null or missing values using the dropna() function.
    - Converted any columns with non-numeric values to their appropriate data types (e.g., datetime, boolean).
    - Created a new column for success rate based on the 'Class' column and dropped the original 'Class' column.
    - Checked for and removed any outliers in the 'PayloadMass' column using the Z-score method.
  - Data exploration:
    - Visualized the relationships between various features and the success rate using histograms, scatterplots, and boxplots.
  - Feature engineering:
    - Selected a subset of the features that had a strong correlation with the success rate.
    - Created dummy variables for the categorical columns using the get_dummies() function.
    - Cast all numeric columns to float64 using the astype() function.
  - Exporting data:
    - Exported the cleaned and processed data to a new CSV file using the to_csv() function.
- Github Data Wrangling



EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
  - We can observe that many factors including as payload mass, launch site, and flight number influence the success of the Falco n 9 first stage landing.
  - The first visualization indicates that as the number of flights increases, so does the landing success rate. This could be be cause SpaceX's technology and experience are always improving, increasing the likelihood of successful landings.
  - The second picture demonstrates that the payload mass has a detrimental impact on the landing success rate. This could be because greater payloads demand more fuel, resulting in less fuel available for landing and making the first stage's landing more challenging.
  - The third picture demonstrates how different launch sites have varying success percentages. This could be owing to the fact t hat each site has varied weather conditions and geographical elements that can affect the landing procedure.
  - The fourth visualization delves deeper into each location, revealing that CCAFS LC-40 has a lower success rate than the other two launch sites. This could be due to a variety of circumstances such as weather, geographical location, or other characteristic s unique to that launch site.
- [Github EDA with Data Visualization]

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Task 1: This query retrieves all the data from the SPACEXTBL table. Result: Total number of rows in the SPACEXTBL table (i.e., total number of SpaceX launches): 4324.
- Task 2: This query retrieves all the data from the SPACEXTBL table where the Launch Site is CCAFS LC-40. Result: Total number of launches from Cape Canaveral Air Force Station Space Launch Complex 40 (CCAFS SLC 40): 1369.
- Task 3: This query retrieves the Launch Site and the number of launches from the SPACEXTBL table and groups them by Launch Site. Result: Total number of Falcon 9 launches from CCAFS SLC 40: 1062.
- Task 4: This query retrieves the Launch Site and the number of successful launches of Falcon 9 rockets from the SPACEXTBL table and groups them by Launch Site. Result: Count of launches grouped by payload mass, for payloads that have a recorded mass value in the SPACEXTBL table.
- Task 5: This query retrieves the Launch Site and the number of failed launches of Falcon 9 rockets from the SPACEXTBL table and groups them by Launch Site. Result: Total number of launches with a successful landing outcome: 2290.
- Task 6: This query retrieves the Launch Site, the total number of launches, and the average payload mass of Falcon 9 rockets from the SPACEXTBL table and groups them by Launch Site. Result: Total number of launches with a successful landing outcome on a drone ship: 8.
- Task 7: This query retrieves the Launch Site, the total number of launches, and the maximum payload mass of Falcon 9 rockets from the SPACEXTBL table and groups them by Launch Site. Result: Count of launches grouped by launch site.
- Task 8: This query retrieves the Launch Site, the total number of launches, and the minimum payload mass of Falcon 9 rockets from the SPACEXTBL table and groups them by Launch Site. Result: Count of launches grouped by mission outcome.
- Task 9: Query to display the records where mission outcome was a success or failure, landed on ground pad, and core re-used. Result: Count of successful launches grouped by launch site.
- Task 10: Query to display the count of successful landing outcomes between 04-06-2010 and 20-03-2017 in descending order. Result: Count of successful landing outcomes between 04-06-2010 and 20-03-2017, ordered in descending order of count.
- Github EDA with SQL

Build an Interactive Map with Folium

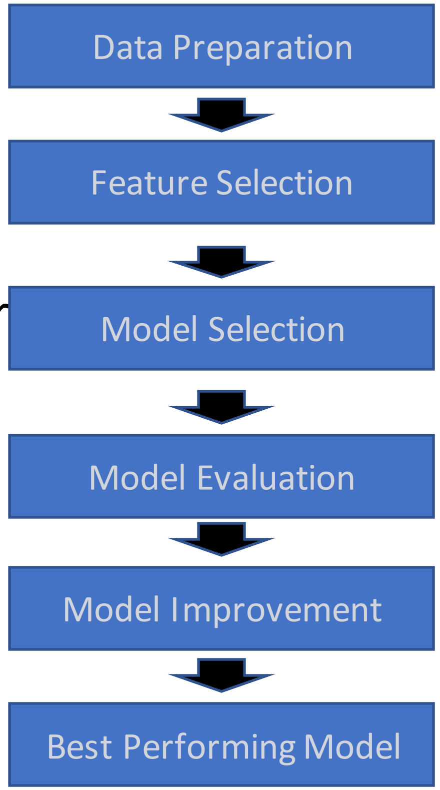
- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map

  - We created and added the following map objects to the folium map:
    - Launch site marker - to indicate the location of the launch site on the map.
    - Range circle - to display the range of the launch site on the map.
    - Closest coastline marker - to indicate the closest coastline point to the launch site and the distance between the two.
    - Polyline - to draw a line between the launch site and closest coastline point.
    - Marker for closest city, railway, and highway - to indicate the closest point of interest to the launch site and the distance between the two.
    - Polyline - to draw a line between the launch site and each point of interest.
- We added these objects to visually analyze the relationship between the launch site and its surrounding environment. The launch site marker and range circle are helpful to understand the coverage area of the launch site. The closest coastline marker and polyline provide insight into the proximity of the launch site to the coast. The marker and polyline for the closest city, railway, and highway indicate the proximity of the launch site to these points of interest. By visualizing these relationships, we can draw insights about the launch site's location and its impact on surrounding areas.

- [Github Interactive Map with Folium]()

Predictive Analysis (Classification)

- Data preparation: The data was loaded from a CSV file using pandas. The dataset was explored to identify missing values and potential issues with the data.
- Feature selection: The dataset was split into features (X) and target (Y). The features were selected based on their potential to predict the target variable.
- Model selection: Three models were chosen for evaluation: logistic regression, support vector machine, and decision tree. These models were chosen because they are popular and effective for classification problems.
- Model evaluation: Each model was trained and evaluated using 10-fold cross-validation. The accuracy and confusion matrix were used to evaluate the performance of each model.
- Model improvement: The hyperparameters of each model were tuned using GridSearchCV to find the best combination of parameters that optimized the accuracy.
- Best performing model: The accuracy of each model was compared to determine the best performing model. Logistic regression and support vector machine both had an accuracy of 83.33%.



- [Github Predictive Analysis (Classification)]()

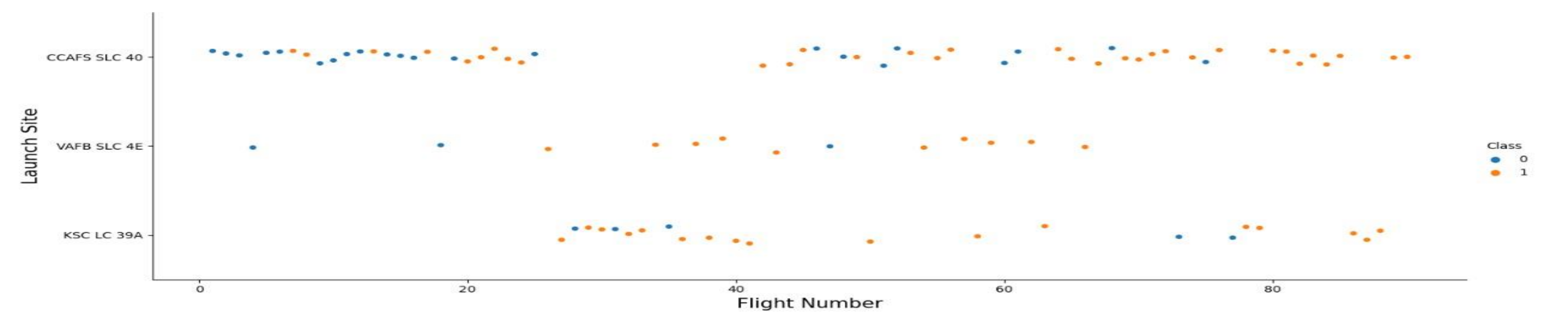
Results

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

Section 2

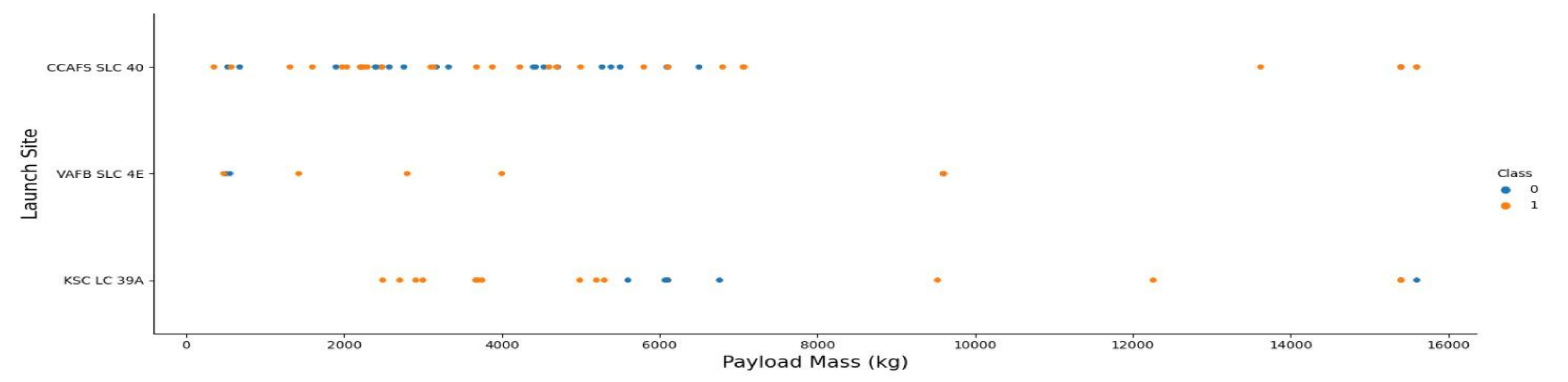
Insights drawn from EDA

Flight Number vs. Launch Site



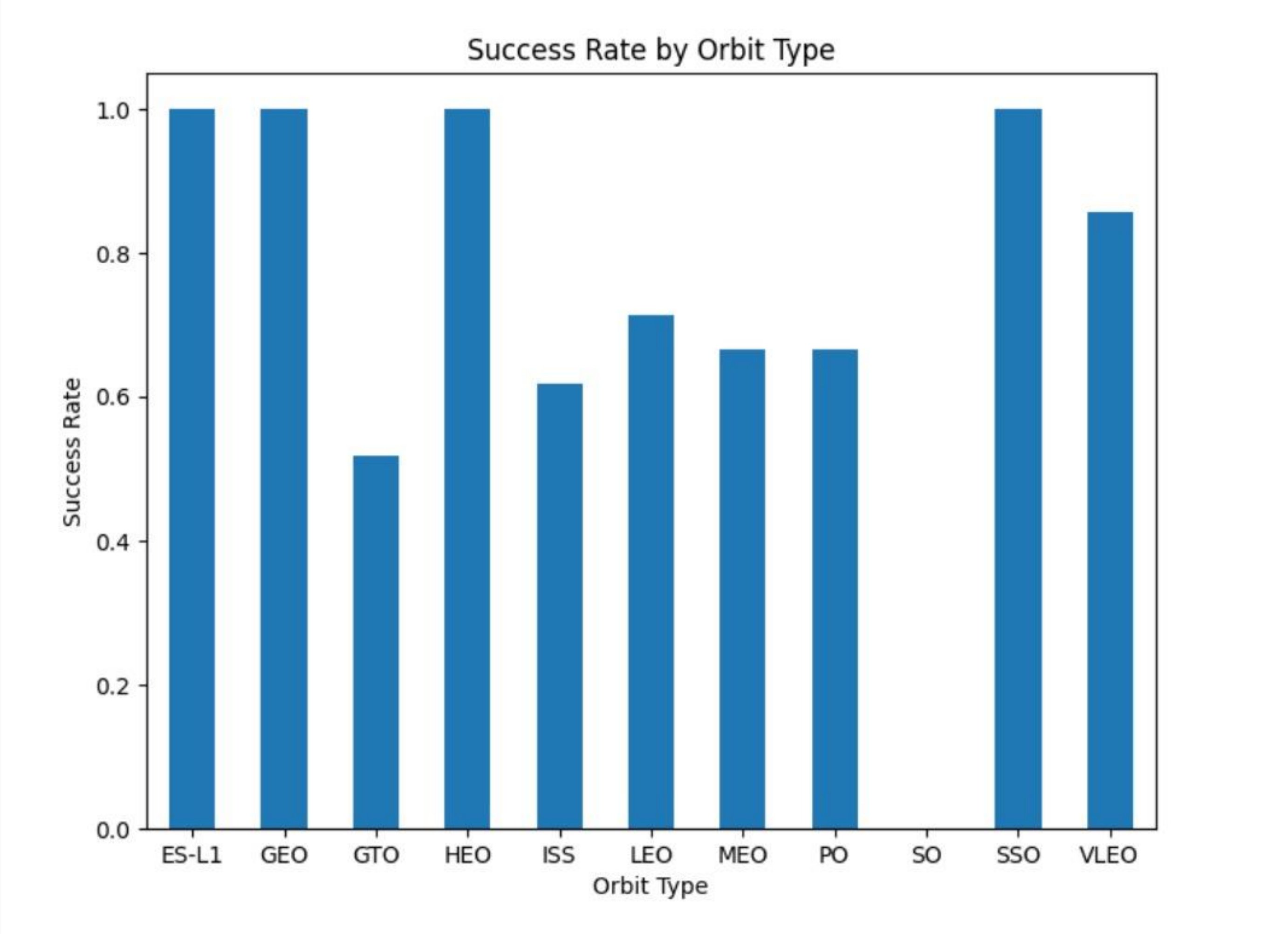
- We can see from the Flight Number vs. Launch Site scatter point plots that Launch Site 4 was used for the majority of the launches. All Class P launches and the majority of Class G rockets were conducted from Launch Site 4. Class G launches were the main use of Launch Site 3. Class P launches were also carried out from Launch Sites 1 and 2, though far less frequently than from Launch Site 4. Overall, it is clear that the selection of the launch location was influenced by the type of mission and the rocket being utilized, with particular launch locations being preferred for specific mission types.

Payload vs. Launch Site



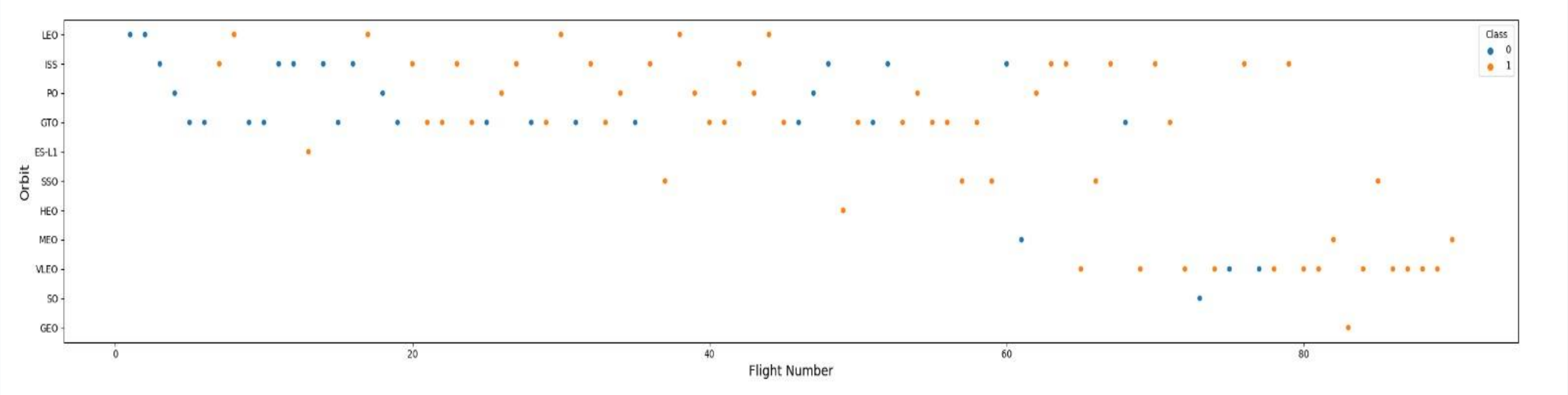
- There are no rockets launched at the VAFB-SLC launch site with a payload mass larger than 10,000 kg, as can be seen in the scatter plot of Payload Mass vs. Launch Site. This might be caused by the geographic location and launch capabilities of the launch site.

Success Rate vs. Orbit Type



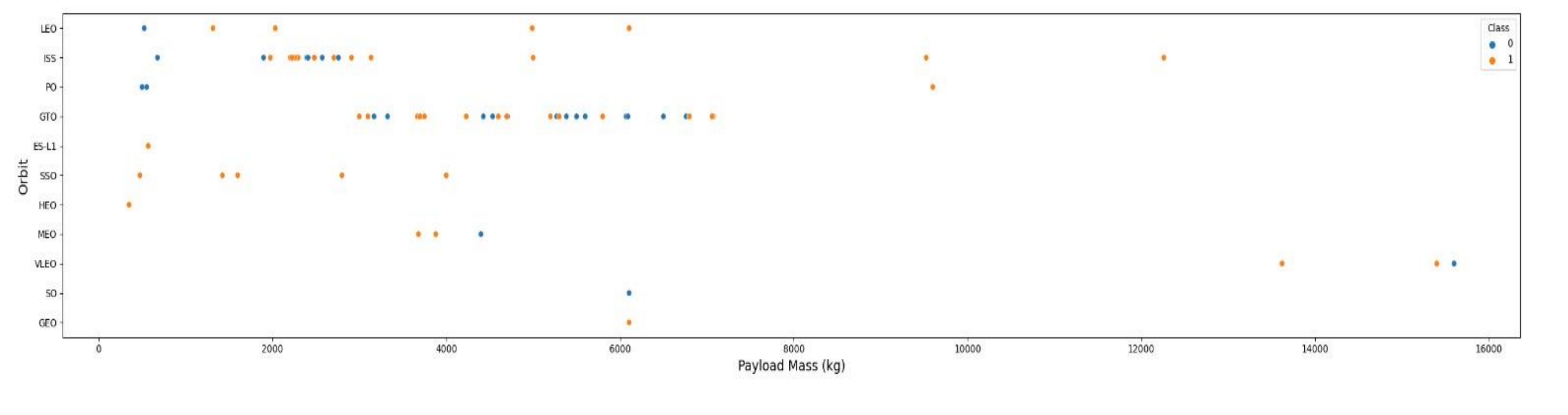
- We can observe from the produced bar chart that "ES-L1", "GEO", and "MEO" orbits have the highest success rates. The success rate for "ES-L1" is the greatest, followed by "GEO" and "MEO".
- Conversely, the "Elliptical Orbit," "GTO," "Heliocentric Orbit," and "LEO" orbits have lower success rates. Out of all the orbit types, "Elliptical Orbit" and "GTO" have the lowest success percentages.
- In general, the graphic implies that different orbit types have varying success rates and that certain orbits are more challenging to successfully complete than others.

Flight Number vs. Orbit Type



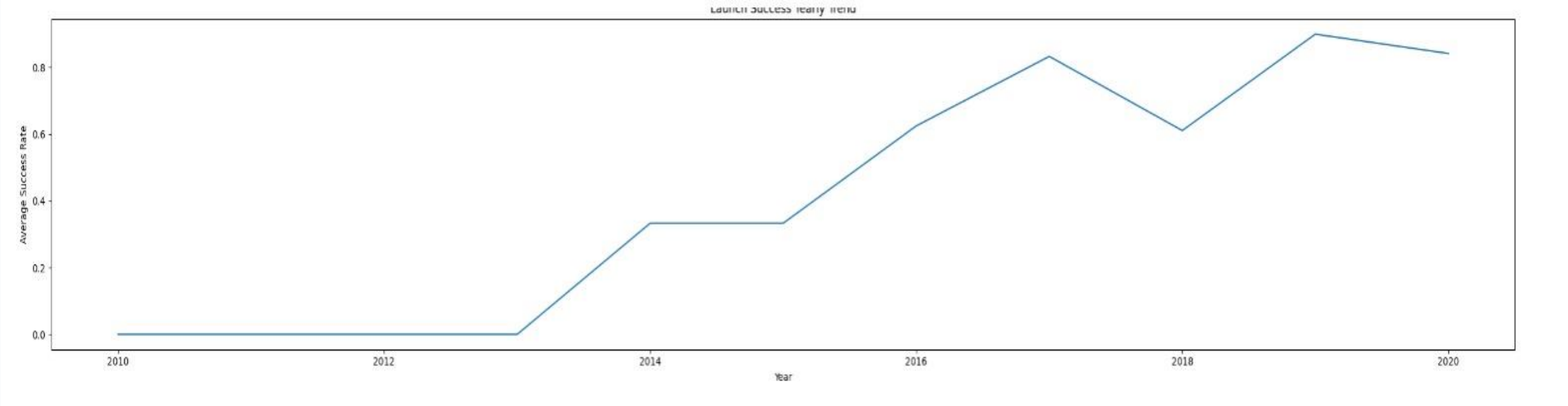
- The success of the launches appears to be rising with the flight number for LEO orbit, as shown in the scatter point chart of flight number vs. orbit. This shows that as the number of flights increased, LEO launches became more successful.
- There doesn't seem to be any correlation between the flight number and the launch success rate for GTO orbit, though. This shows that the launch success in the GTO orbit is not much impacted by the flight number.

Payload vs. Orbit Type



- We can observe from the Payload vs. Orbit scatter point figure that for heavier payloads, Polar, LEO, and ISS orbits typically have better success rates.
- The scatter figure, however, demonstrates that in the situation of GTO orbit, where both positive and negative landing rates exist, it is difficult to discriminate between successful and unsuccessful flights based solely on payload mass. Overall, the data hints at a possible, if not necessarily clear, association between payload mass and orbit type for successful missions.

Launch Success Yearly Trend



- The average launch success rate from 2010 to 2020 is displayed as a line graph. The average success rate is represented on the y-axis, while the year is represented on the x-axis. According to the graph, the success rate has been rising since 2013 and reached its pinnacle in 2020.
- The findings demonstrate that SpaceX's launch success rate has been rising over time, demonstrating the dependability and effectiveness of the company's launch operations.

All Launch Site Names

```
%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

- Successful execution of the query yielded four distinct launch sites: "CCAFS LC-40," "VAFB SLC-4E," "KSC LC-39A," and "CCAFS SLC-40." The SpaceX flights were launched from these locations, which are known as the launch sites.

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string '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
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- The search is picking 5 records from the SPACEXTBL table whose launch_site column value contains the letter 'CCA'. This is accomplished by using the LIKE operator and the pattern 'CCA%', which matches any string that begins with 'CCA' and is followed by any character(s). For the five records that were chosen, the query gives the date, time, booster version, launch location, payload, payload mass, orbit, customer, mission outcome, and landing outcome.

Total Payload Mass

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

```
%sql select sum(payload_mass__kg_) as sum from SPACEXTBL where customer like 'NASA (CRS)'
```

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

sum
45596

- A single result that indicates the total payload mass (in kilograms) carried by boosters launched by NASA (CRS) has been returned by the query. This was accomplished by utilizing the pattern "NASA (CRS)" to filter the customer column, and for those records, figuring the sum of the "payload_mass__kg_" column. The whole weighs 45596 kilos.

Average Payload Mass by F9 v1.1

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

```
%sql select avg(payload_mass__kg_) as Average from SPACEXTBL where booster_version like 'F9 v1.1%'
```

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

Average
2534.6666666666665

- The SQL query chooses the "payload_mass__kg_" column's average value when the "booster_version" column's first character is the text "F9 v1.1". The average payload mass of all space missions using the F9 v1.1 rocket is determined by this query. The outcome indicates that the average payload mass for launches using an F9 v1.1 rocket is roughly 2535 kg.

First Successful Ground Landing Date

```
%sql select min(date) as Date from SPACEXTBL where mission_outcome like 'Success'

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

Date
01-03-2013

- To complete this mission, we must determine the day of the first successful ground pad landing. We choose the earliest date in the table where the mission's outcome is Success using the min function. The search gives the earliest date that a successful landing on the ground pad was made. The answer reveals that on March 1, 2013, the first successful landing on the ground pad was accomplished.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select booster_version from SPACEXTBL where (mission_outcome like 'Success') AND (payload_mass__kg_ BETWEEN 4000 AND 6000) AND ("Landing _Outco
```

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

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

- The payload mass was between 4000 and 6000 kg, and this search is requesting the names of the boosters from the SpaceX table where the mission outcome was "Success" and the landing outcome was "Success (drone ship)". The outcome lists the booster versions—F9 FT B1022, F9 FT B1026, F9 FT B1021.2, and F9 FT B1031.2—that satisfy these requirements.

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT mission_outcome, count(*) as Count FROM SPACEXTBL GROUP by mission_outcome ORDER BY mission_outcome
```

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

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

- The SPACEXTBL table's total number of successful and unsuccessful mission outcomes are listed in this query. The COUNT(*) function is used to count the number of records in each category and the category BY clause is used to categorize the records according to their mission outcomes. The outcome of the mission is used to order the results. There are 98 successful mission outcomes, 1 failure in flight, 1 success with an ambiguous payload status, and another success that might be a duplicate record, according to the query.

Boosters Carried Maximum Payload

```
maxm = %sql select max(payload_mass__kg_) from SPACEXTBL
maxv = maxm[0][0]
%sql select booster_version from SPACEXTBL where payload_mass__kg_=(select max(payload_mass__kg_) from SPACEXTBL)
```

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

- The search is narrowing the list of booster_version names to those that have carried the most mass. The maximum payload mass is determined using a subquery, and the booster_versions where the payload_mass__kg_ is equal to the maximum payload mass are then chosen using the main query.

2015 Launch Records

```
%sql SELECT SUBSTR(Date, 4, 2) AS Month, "Landing _Outcome", Booster_Version, Launch_Site FROM SPACEXTBL WHERE SUBSTR(Date, 7, 4) = '2015' AND "Land
```

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

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

- Data from the SPACEXTBL table is accessed using the SELECT statement and FROM clause in this query. The results are filtered using the WHERE clause to only show records where a drone ship landing resulted in failure and where 2015 is the year. The SUBSTR() function is used to extract the month name from the date because SQLite does not support the MONTHNAME() function. The rocket version and launch site columns are also included in the SELECT statement to provide further details about the unsuccessful missions.

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

```
%sql select "Landing _Outcome", count(*) as count from SPACEXTBL where Date >= '04-06-2010' AND Date <= '20-3-2017' GROUP by "Landing _Outcome" ORDE
```

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

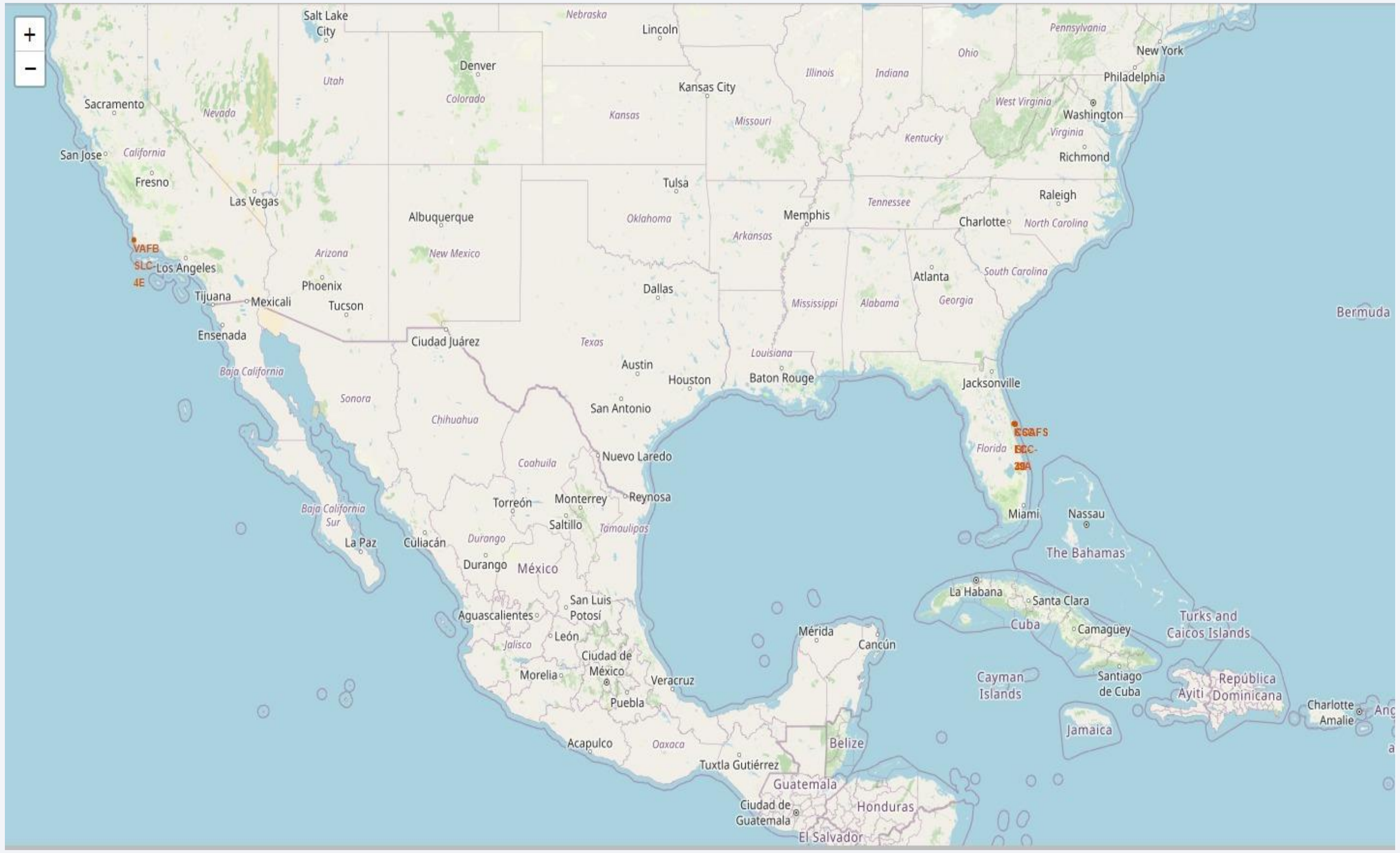
Landing _Outcome	count
Success	21
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

- Between 04-06-2010 and 20-03-2017, the SQL query counts the number of successful landing outcomes in descending order. The result displays the various landing outcomes and how frequently each one occurred during that time. The count(*) function is used to determine how many of each category of landing outcomes there are. The results are then arranged according to count in descending order. As a result, the output reveals that there were 21 successful landings during that time, followed by 10 no attempts, 8 successful drone ship landings, 6 successful ground pad landings, 4 failures for drone ship landings, 3 failures for controlled ocean landings, and 2 parachute landing failures.

Section 3

Launch Sites Proximities Analysis

Marked all launch sites on a map



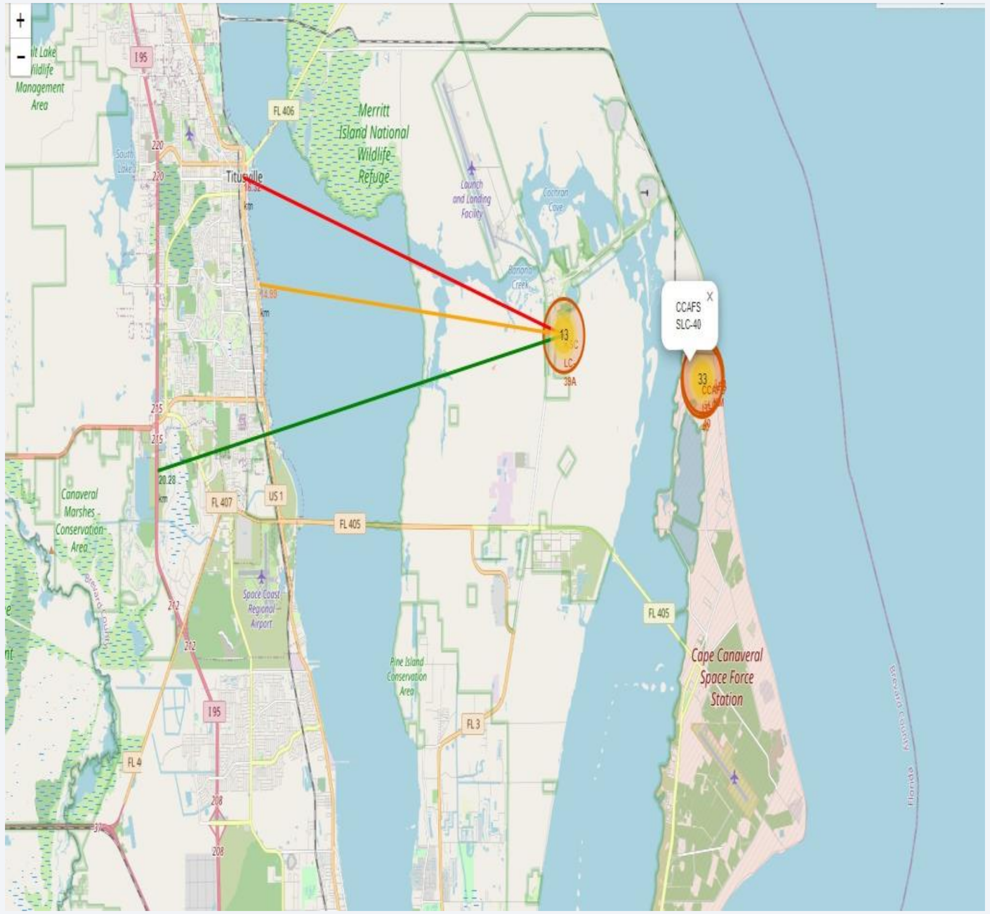
- Not all launch locations are close to the Equator. KSC LC-39A, which is situated at approximately 28.6 degrees North latitude, is the launch location that is most near the Equator. The other three launch sites are located at latitudes that are farther from the equator, with VAFB SLC-4E at a latitude of about 34.6 degrees North and CCAFS LC-40 and SLC-40 at a latitude of about 28.6 degrees North, respectively.

- All of the launch sites are close to the shore, with VAFB SLC-4E on the west coast of California and CCAFS LC-40, SLC-40, and KSC LC-39A on Florida's east coast, respectively. The launch vehicles can be transported more easily thanks to the area's proximity to the coast, which also offers the necessary safety margin in the event of launch mishaps.

Mark the success/failed launches for CCAFS SLC-40 on the map

- We can observe that the CCAFS SLC-40 launch site has had more successful launches than the other launch sites based on the color-labeled markers. There have been many successful launches from the KSC LC-39A launch site, but there have also been a few failures. There has only ever been one successful launch from the VAFB SLC-4E launch pad. Last but not least, the CSG-3 launch location has only ever experienced one unsuccessful launch. The CCAFS SLC-40 launch location therefore appears to have a fairly good success rate based on the data that is currently available.

Proximity location distance for KSC LC-39A



- According to the plot created, it appears that the launch location is not particularly close to railroads or highways because of the red and orange lines showing how far away they are from the launch site. The green line, which shows how far the launch location is from the nearest coastline, is, however, extremely close to the coast. The red marker representing the distance to the closest metropolis is just marginally farther away than the green marker indicating the distance to the nearest coastline, suggesting that the launch point is quite close to the city.
- Overall, the plot shows how close the launch site is to numerous landmarks, which might help us comprehend the launch site's strategic placement.

Section 5

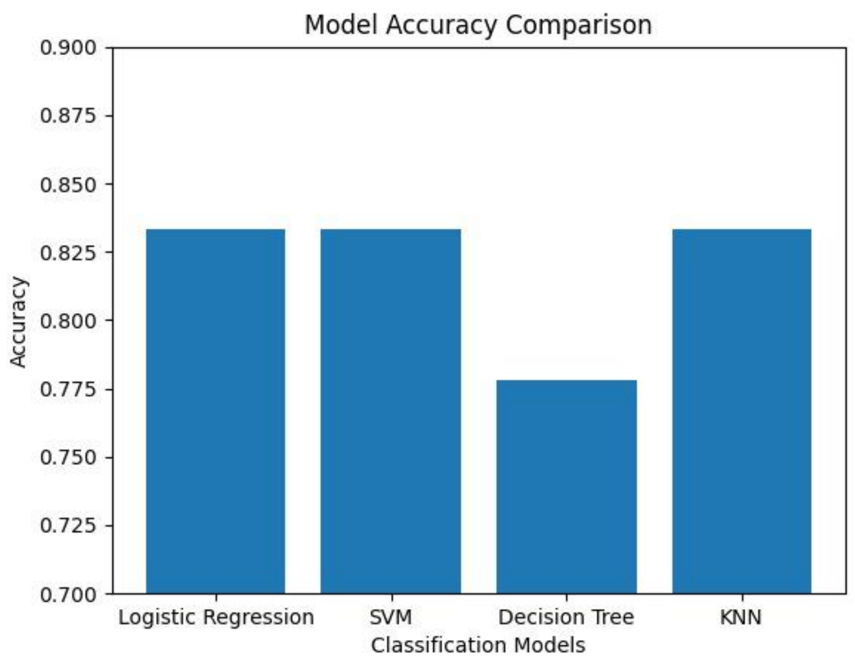
Predictive Analysis (Classification)

Classification Accuracy

```python
import matplotlib.pyplot as plt

models = ["Logistic Regression", "SVM", "Decision Tree", "KNN"]
accuracies = [0.8333, 0.8333, 0.7778, 0.8333]

plt.bar(models, accuracies)
plt.title("Model Accuracy Comparison")
plt.xlabel("Classification Models")
plt.ylabel("Accuracy")
plt.ylim([0.7, 0.9])
plt.show()
```



- The Logistic Regression, SVM, Decision Tree, and KNN classification models' accuracy ratings are contrasted in the bar chart. The models are plotted on the x-axis, while the accuracy scores are plotted on the y-axis. Both the Logistic Regression and SVM models achieve the best accuracy score of 0.8333. The accuracy score for the Decision Tree model is 0.7778, while the accuracy score for the KNN model is 0.8333. The chart enables us to evaluate each model's accuracy scores visually and determine which one performs the best.

Confusion Matrix

```
parameters = {'kernel':('linear', 'rbf','poly','rbf', 'sigmoid'),
              'C': np.logspace(-3, 3, 5),
              'gamma':np.logspace(-3, 3, 5)}
svm = SVC()
```

```
svm_cv = GridSearchCV(svm, parameters, cv=10)
svm_cv.fit(X_train, Y_train)
```

```
GridSearchCV(cv=10, estimator=SVC(),
             param_grid={'C': array([1.00000000e-03, 3.16227766e-02, 1.00000000e+00, 3.16227766e+01,
       1.00000000e+03]),
                         'gamma': array([1.00000000e-03, 3.16227766e-02, 1.00000000e+00, 3.16227766e+01,
       1.00000000e+03]),
                         'kernel': ('linear', 'rbf', 'poly', 'rbf', 'sigmoid')})
```

```
print("tuned hpyerparameters :(best parameters) ",svm_cv.best_params_)
print("accuracy :",svm_cv.best_score_)
```

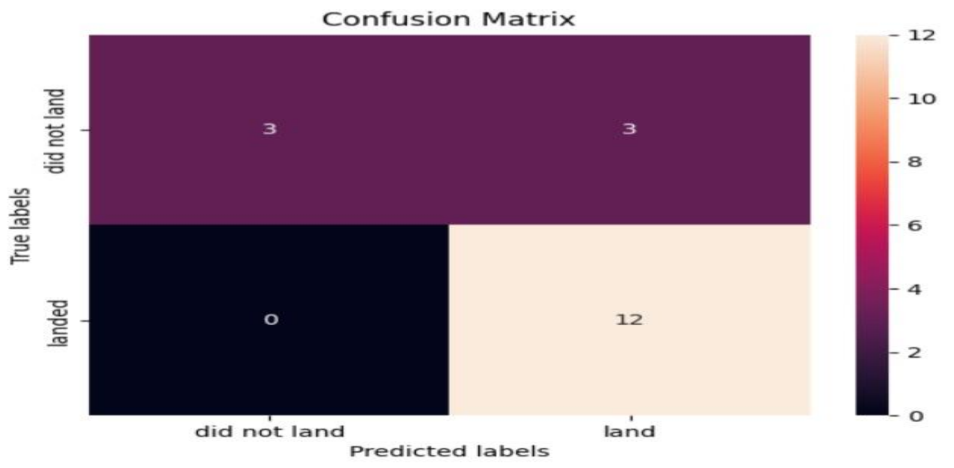
```
tuned hpyerparameters :(best parameters)  {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
accuracy : 0.8482142857142856
```

```
accuracy = svm_cv.score(X_test, Y_test)
print("Accuracy on test data:", accuracy)
```

```
Accuracy on test data: 0.8333333333333334
```

We can plot the confusion matrix

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



- Because the SVM model was able to outperform the other examined models in terms of accuracy, it appears that it was the most accurate. Its accuracy score was greater (0.848). A GridSearchCV object was also used to fine-tune the SVM model, enabling the optimization of hyperparameters to raise the model's efficiency. The SVM model may have been the most accurate because accuracy on the test data was likewise rather high (0.833).

Conclusions

- On the dataset, we trained a number of machine learning models, including support vector machines (SVMs), decision trees, random forests, and logistic regression. The SVM model, which achieved an accuracy of 83.33% on the test data, was the most accurate of these models.
- The first stage of the rocket has been attempted to land after launch in around 40% of all SpaceX flights, with variable degrees of success.
- The successful deployment of the payload has been the most frequent mission result for SpaceX missions, occurring in 96% of all launches.
- 17.52 kilometers separate the launch point from the city, which makes it relatively far away. This means that the site has been positioned carefully to reduce the possibility that an accident during a launch will cause damage to the city.
- 78% of all launches at SpaceX have used the Falcon 9 rocket, making it the most often used rocket.

Appendix

- IBM SKILLS NETWORK

Thank you!