FORECAST FOR THE LANDING RESULT OF FALCON9

MANUEL MARTIN
OCTOBER 2025

Outline

- Executive Summary
- Introduction
- Methodology
- Insights drawn from EDA
- Launch sites proximities analysis
- Interactive Dashboards
- Predictive Analysis (Classification)
- Conclusion

Executive Summary

SUMMARY OF METHODOLOGIES:

- Data collection.
- Data wrangling.
- Data exploring.
  - Significative request with SQL.
  - Data engineering with visualization
- Building interactive graphics for stakeholders.
- Predictive analysis with machine learning.

SUMMARY OF RESULTS:

- Exploratory data analysis result.
- Visualization analytics.
- Predictive analytics.

Introduction

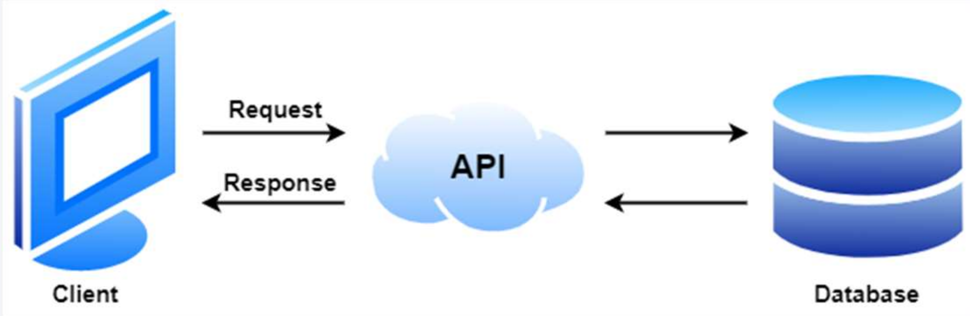
- Space X is an aerospace company dedicated to the design, manufacture, and launch of rockets and spacecraft, aiming to reduce costs. The primary method for reducing costs is the recovery of the rocket's launch vehicle, Stage 1, so it can be reused.
- The purpose of this paper is to analyze the information provided by the company on the various missions carried out and calculate the probability of success of new missions by implementing predictive machine learning models, thereby optimizing mission budgets.

Section 1: Methodology

Data Collection – SpaceX API

Steps:

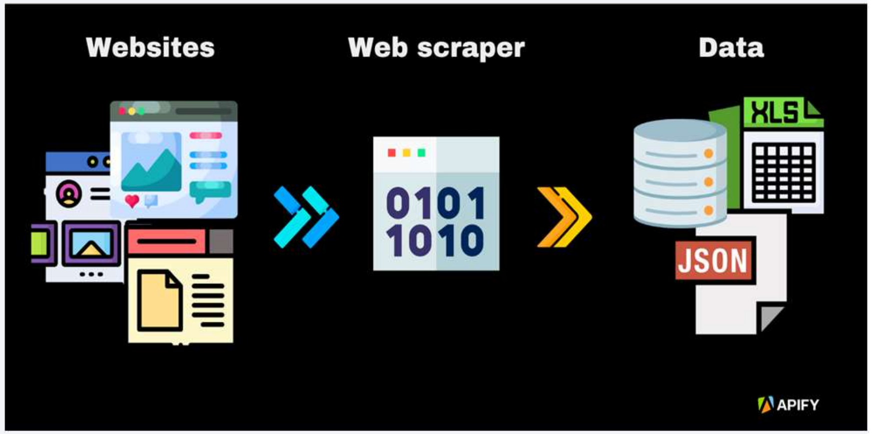
1. Request function to access the API and download information.
2. Decode to JSON with .Jason().
3. Covert to df wth .json_normalize().
4. Select the significant attributes
5. Obtain the needed information from the df in a list and then create another df.
6. Filter the df to contain only “Falcon 9 “lanches.
7. Replace missing values of the Payload mass the mean.
8. Export data to.csv file.



Data Collection - Scraping

Web Scraping Procedure:

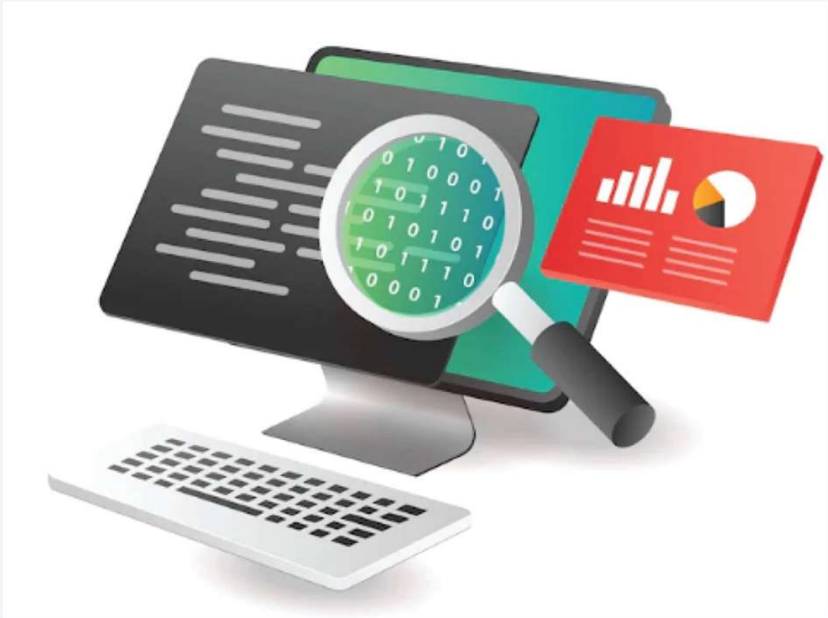
1. Utilize the get.request function to retrieve the Falcon9 Launch HTML as an HTTP response.
2. Apply the BeautifulSoup function to extract the text content from the HTML file.
3. Use the find_all() function to gather all "table" elements from the HTML document and select the desired table by specifying its index.
4. Retrieve the column headers from the table using find_all("th") and initialise an empty dictionary with these headers.
5. Populate the dictionary with data from the table and convert it into a dataframe using the pd.DataFrame() function.
6. Save the dataframe to a .csv file using the .to_csv() function.



Data Wrangling

During this stage, we will identify patterns and assign labels for the model's predictions:

1. Check for null values using .isnull() and examine the data types of various attributes with .dtype.
2. Use the .value_counts() function to determine the number of missions at each station, in each orbit, and the counts of different outcome types.
3. Create a list coding the outcomes, then add it to the data frame under the column “Class”.
4. Compute the outcome rate by applying the .mean() function.
5. Save the data frame as a .csv file using the .to_csv() function.



EDA with SQL

Different queries done using SQL:

1. Display different launch site names.
2. Display 5 records of launch sites that begin with"CCA".
3. Display the total load of all the missions done by NASA.
4. Display the average load of the booster model "F9 v1.1".
5. List the date of the first successful landing outcome of the ground-landing type.
6. List the booster model with success landing in drone ship and payload mass between 4000 and 6000 kg.
7. List the number of different types of outcomes.
8. List the booster versions that have transported the maximum load.
9. List all the booster models that have outcome failure on the drone ship during 2015, specifying the months.
10. List the number of missions of each type of outcome from 2010-06-04 to 2017-03-20

EDA with Data Visualization

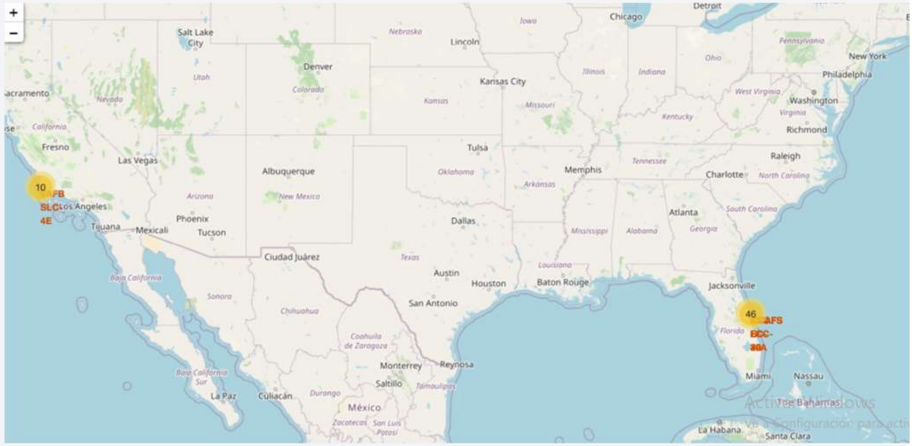
- In this part, the main question is which attributes are correlated with the land's success, so the scatter plot shows the correlation between two variables clearly. Important to know the result in each point using the parameters "hue" in the graphic.
- To know the rate of success of each orbit, the bar chart permits us to separate the different types of orbit and the rate for each one.
- Finally, the line chart lets us represent the evolution of the success rate with time.

Build an Interactive Map with Folium

This section examines the significance of the locations and features surrounding the launch site, utilizing the Python Folium library for analysis.

Summary of the Folium-related tasks:

1. Initialize a map centered on specific coordinates using folium.Map.
2. Highlight a location on the map with labels using folium.Circle and folium.map.Marker.
3. Add markers indicating the type of outcome for each launch site with MarkerCluster().
4. Measure distances from the launch site to various important locations using the MousePosition() function.
5. Connect these key locations around the launch site by drawing polylines.



Build a Dashboard with Plotly Dash

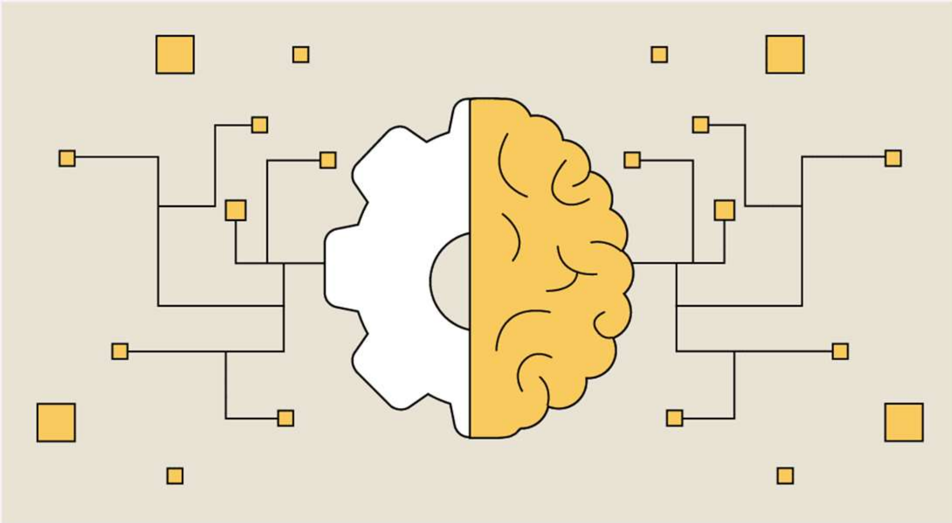
When using Plotly Dash, two different chart types have been chosen:

- The first chart is a pie chart that shows the success rate for each launch site, paired with a dropdown menu to select a particular site.
- The second chart is a scatter plot that depicts the correlation between payload and success rate, including a slider to modify the payload range.
- Both charts are interactive, as demonstrated.
- This method effectively meets the stakeholders' requirements for data exploration and deeper understanding.

Predictive Analysis (Classification)

In this section, several models were implemented to identify the one that most accurately reflects reality. The process involved:

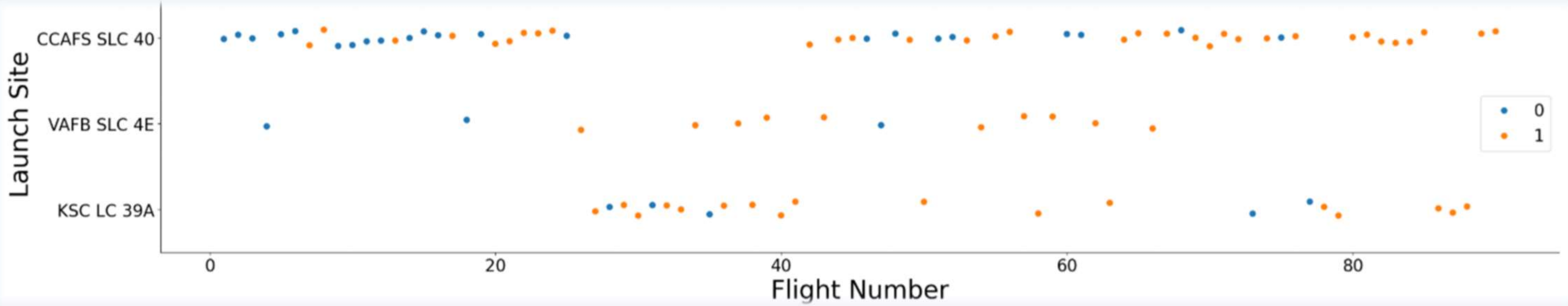
1. Splitting the data into training and testing sets using the train_test_split function.
2. Applying various models to the data and tuning their parameters with the GridSearchCV function to find the best fit.
3. Using the .predict() function to generate predictions and comparing them with actual values to evaluate model accuracy.
4. Comparing the performance metrics of the different models to determine which best aligns with real-world data.



Section 2
Insights drawn from EDA

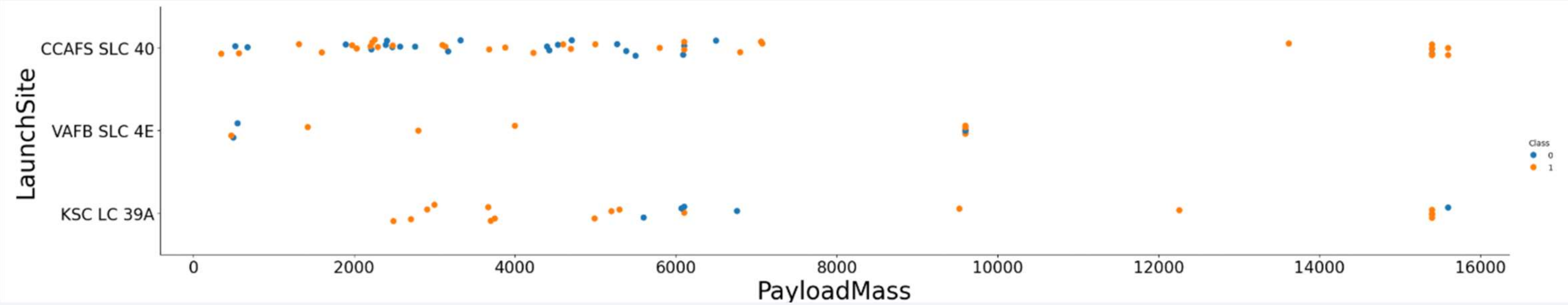
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Flight Number vs. Launch Site



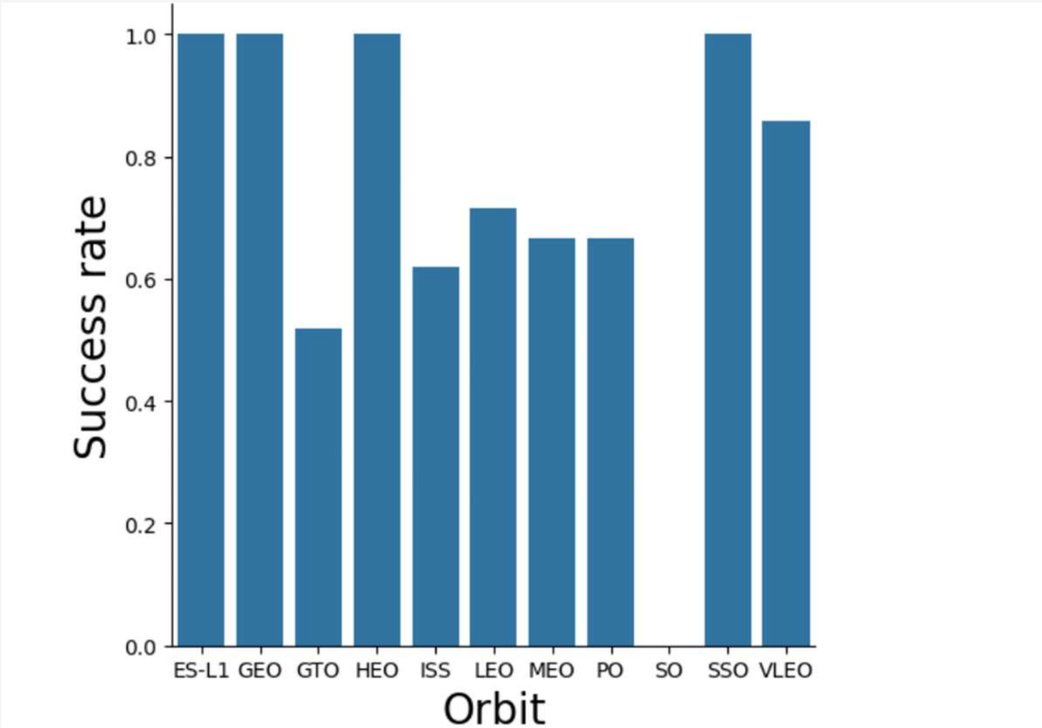
- The first mission was launched from CCAFS SLC-40. The launch site remained unchanged at VAFB SLC-4E until the fourth mission. Around mission 25, operations began at KSC LC-39.
- The number of missions at each launch site is as follows: 55 at CCAFS SLC-40, 13 at KSC LC-39, and 22 at VAFB SLC-4E, totaling 90 flights.
- Over time, there is a noticeable improvement in the success rate of first-stage landings.

Payload vs. Launch Site



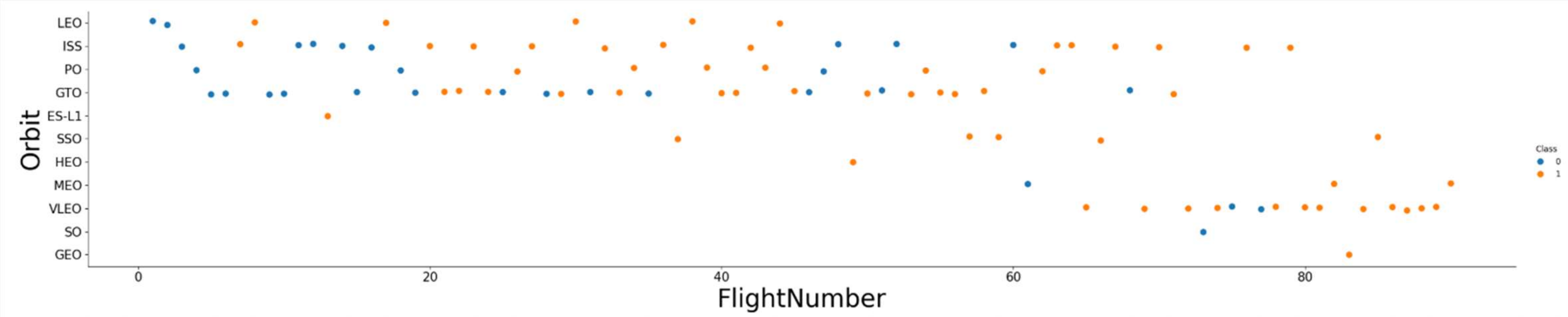
- The plot shows that the usual payload range at the three launch locations is between 0 and 7000 kg.
- At VAFB SLC-4E, the maximum payload is below 10,000 kg, while at the other sites, it nears 16,000 kg.
- Concerning success rates in relation to payload weight, heavier payloads seem to be linked with higher success; however, this finding should be interpreted carefully, since other factors like flight timing could also affect the results.

Success Rate vs. Orbit Type



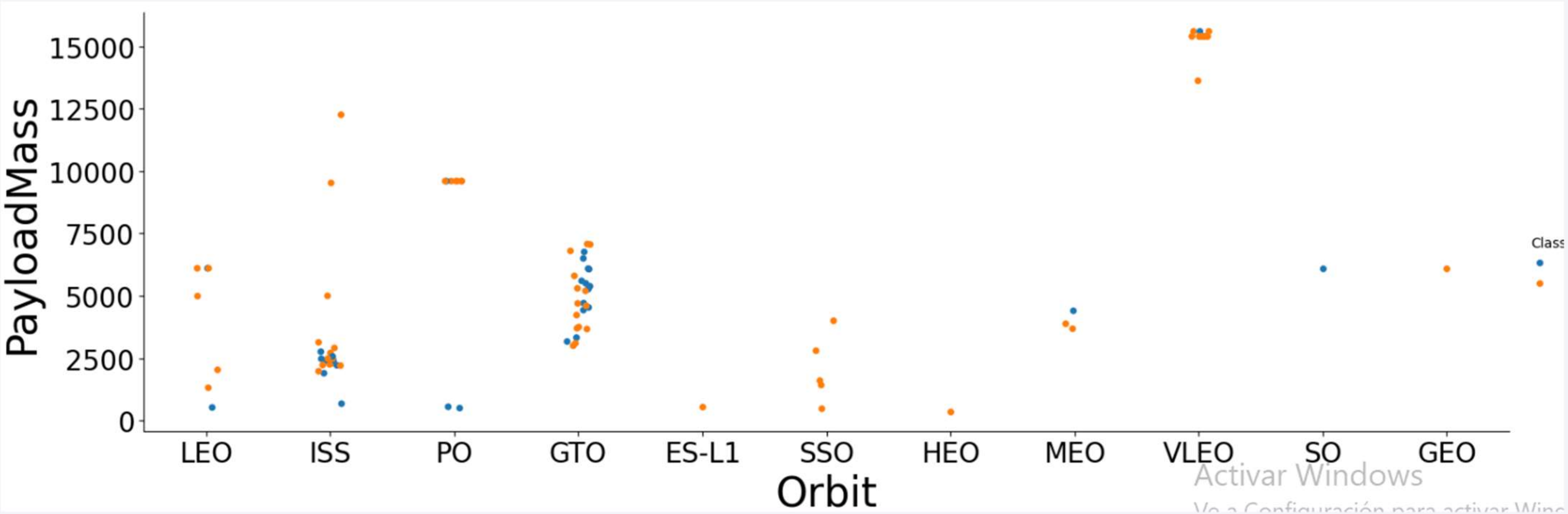
- There is a zero percent success rate in SO orbit, but since there has only been one mission there, this holds little significance.
- In other orbits, the success rate is at least 50%.
- Four orbits—ES-L1, GEO, HEO, and SSO—have a 100% success rate; however, only SSO's results are meaningful because it has had five missions, while the others had just one each.
- LEO, ISS, PO, GTO, and VLEO each have more than ten missions, making their results significant. LEO and VLEO boast success rates above 70%, highlighting the importance of selecting the right orbit for a mission.

Flight Number vs. Orbit Type



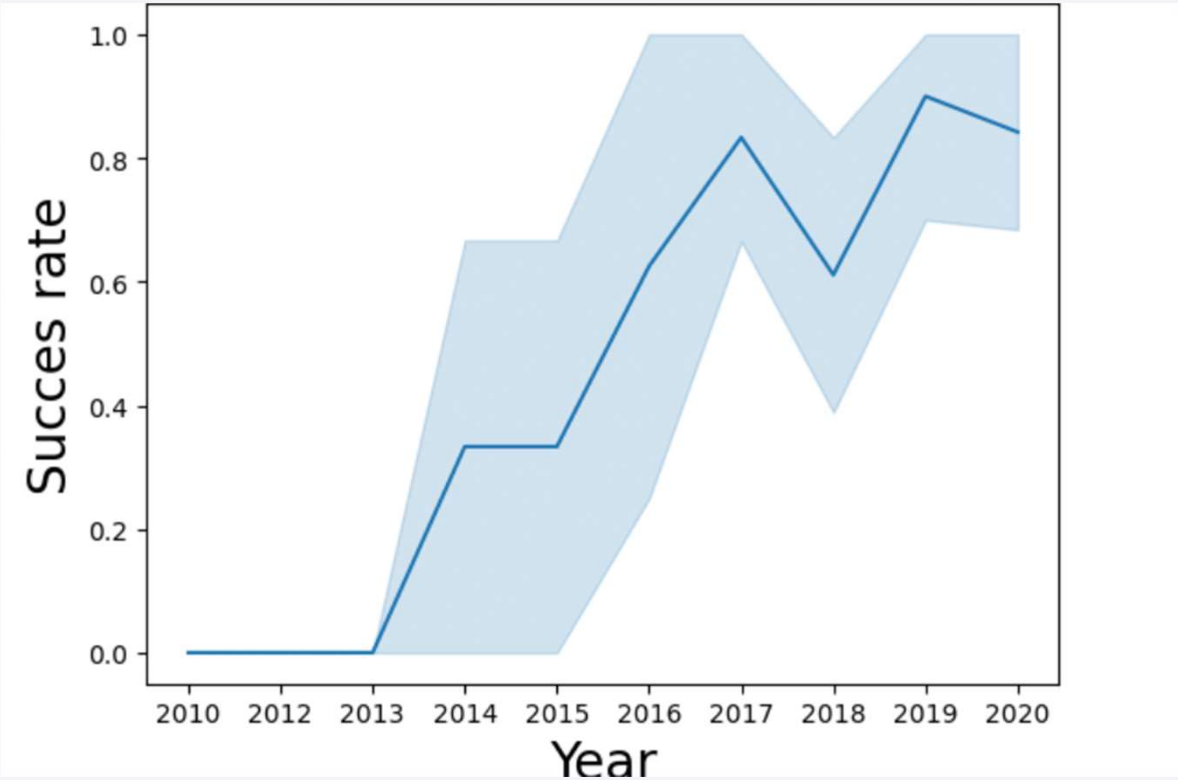
- In the first 55 missions, rockets were mainly sent to LEO, ISS, PO, and GTO orbits, with just three exceptions. Since that time, the range of target orbits has expanded, with VLEO now hosting the most missions, followed by ISS.
- Additionally, a clear trend seen in other charts is that the success rate tends to rise as the number of missions grows.

Payload vs. Orbit Type



- Missions transporting payloads over 10,000 kg have been successfully sent to the ISS, PO, and VLEO.
- Most missions carry payloads under 7,500 kg.
- There are no missions with payloads between 7,500 kg and 10,000 kg.

Launch Success Yearly Trend



- Between 2010 and 2013, no first-stage landings were successful.
- The success rate steadily improved over the following years, reaching approximately 80% by 2016 and staying consistent afterward.
- In 2018, the success rate declined to 60%.
- Therefore, it is evident that the success rate of first-stage landings has generally improved over time.

All Launch Site Names

```
%sql SELECT DISTINCT Launch_Site     FROM SPACEXTABLE
```

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

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

- The SQL query obtains unique launch site names involved in SpaceX missions.
- By applying the "SELECT DISTINCT" statement to the "Launch_Site" column in the "SPACEXTABLE" dataframe, it generates this output.

Launch Site Names Begin with 'CCA'

```
%sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5
```

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

- The “SELECT *” statement fetches all columns from the data frame “SPACEXTABLE,” applying the condition set by the “WHERE” clause. This condition filters rows where the “Launch-site” column's values begin with “CCA,” utilizing the “LIKE ‘CCA%’” command.

Total Payload Mass

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) as total_payload_mass FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)'
```

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

total_payload_mass
45596

- This query consists of three distinct clauses: “SELECT” to specify the columns to retrieve, “FROM” to indicate the data frame, and “WHERE” to set the condition for filtering. Within the “SELECT” clause, the SUM() function is used to total all charges associated with NASA, as defined by the condition in the “WHERE” clause.
- Consequently, the output represents the combined total of all payloads from NASA missions.

Average Payload Mass by F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) as Average_payload_F9 FROM SPACEXTABLE WHERE Booster_Version LIKE 'F9 v1.1'
```

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

Average_payload_F9
2928.4

- This query aims to calculate the average payload for missions using the booster model "F9 v1.1."
- To obtain this result, the average function is applied to the payload mass field within the "SELECT" statement, with the booster type specified in the "WHERE" clause.

First Successful Ground Landing Date

```
%sql SELECT Date FROM SPACEXTABLE WHERE Landing_Outcome = "Success"  ORDER BY Date LIMIT 1
```

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

Date
2018-07-22

- This query aims to obtain the date of the first successful landing.
- To do so, the date column is selected in the “SELECT” statement, while the “WHERE” clause filters for successful landings. The results are ordered by date with the ORDER BY clause, and the LIMIT command is used to return only the earliest record from the sorted list.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT  Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome LIKE "Success (drone ship)" AND  PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
```

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

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

- Here, since there are two conditions, both should be included within the "WHERE" clause.
- In the "SELECT" statement, the booster attribute's type has been clearly defined.

Total Number of Successful and Failure Mission Outcomes

```
%sql  SELECT DISTINCT(Mission_Outcome), COUNT(*)as Event_Number  FROM SPACEXTABLE GROUP BY Mission_Outcome
```

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

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

- To find the count of unique occurrences of an attribute, the “DISTINCT” keyword is used within the “SELECT” statement along with the COUNT function applied to the desired column.
- Additionally, the “GROUP BY” clause is essential to define the various categories of the chosen attribute.

Boosters Carried Maximum Payload

```
%sql SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE) ORDER BY Date
```

```
 * 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

- This query aims to identify the various booster models that have carried the highest payload charge.
- To achieve this, a subquery is included in the "WHERE" clause to select the maximum value within the payload mass attribute.

2015 Launch Records

```
%sql SELECT SUBSTR(Date, 6, 2) as Month, Booster_Version, launch_site  FROM SPACEXTABLE
WHERE Landing_Outcome = "Failure (drone ship)" AND Date LIKE "2015%"
```

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

Month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

- The query requests the months, launch locations, and booster models associated with unsuccessful drop ship landings in 2015.
- The main difficulty in this query lies in extracting the month from the date, which is achieved using the SUBSTR() function.

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

```
%sql SELECT DISTINCT(Landing_Outcome), COUNT(*) as Number_events FROM SPACEXTABLE WHERE Date BETWEEN "2010-06-04" and "2017-03-20"
    GROUP BY Landing_Outcome ORDER BY (COUNT(*)) desc
```

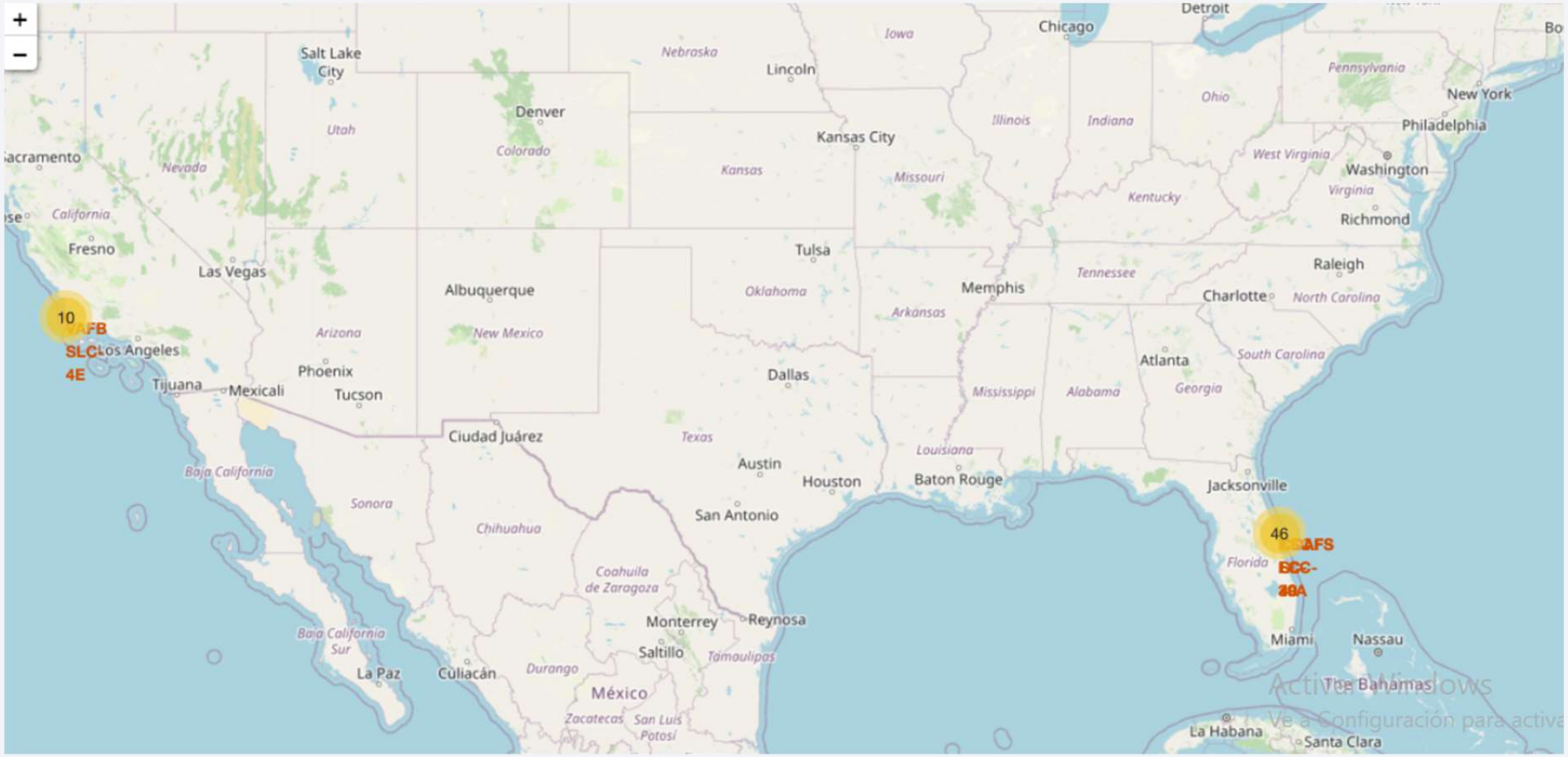
```
 * sqlite:///my_data1.db
Done.
```

Landing_Outcome	Number_events
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

- The function “DISTINCT” was applied within the “SELECT” statement along with the “COUNT” function to determine the number of unique outcome types occurring on those dates.
- It is essential to use the “GROUP BY” clause to define the target of the “COUNT” function.

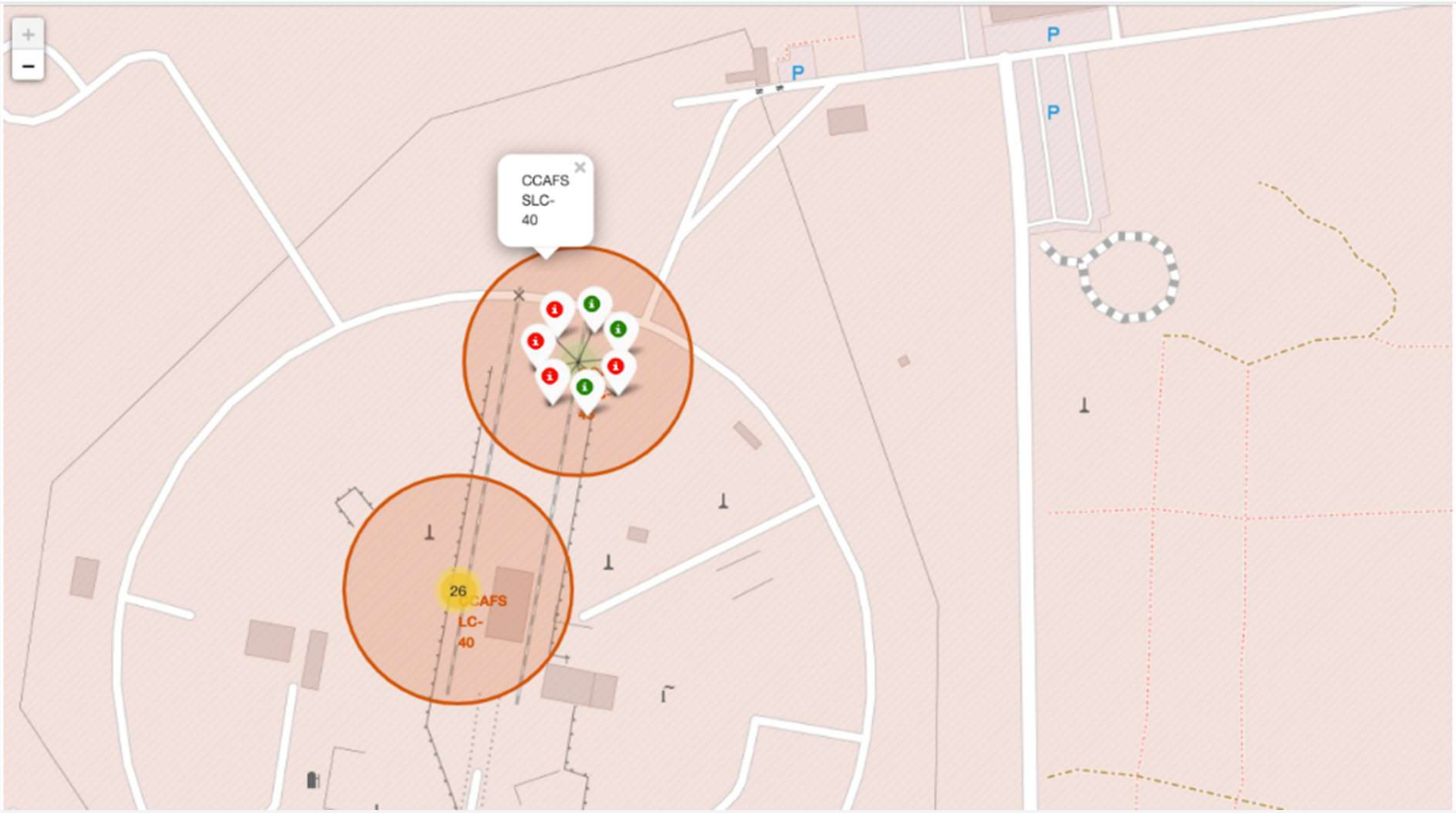
Section 3
Launch sites proximities analysis

Launch site location



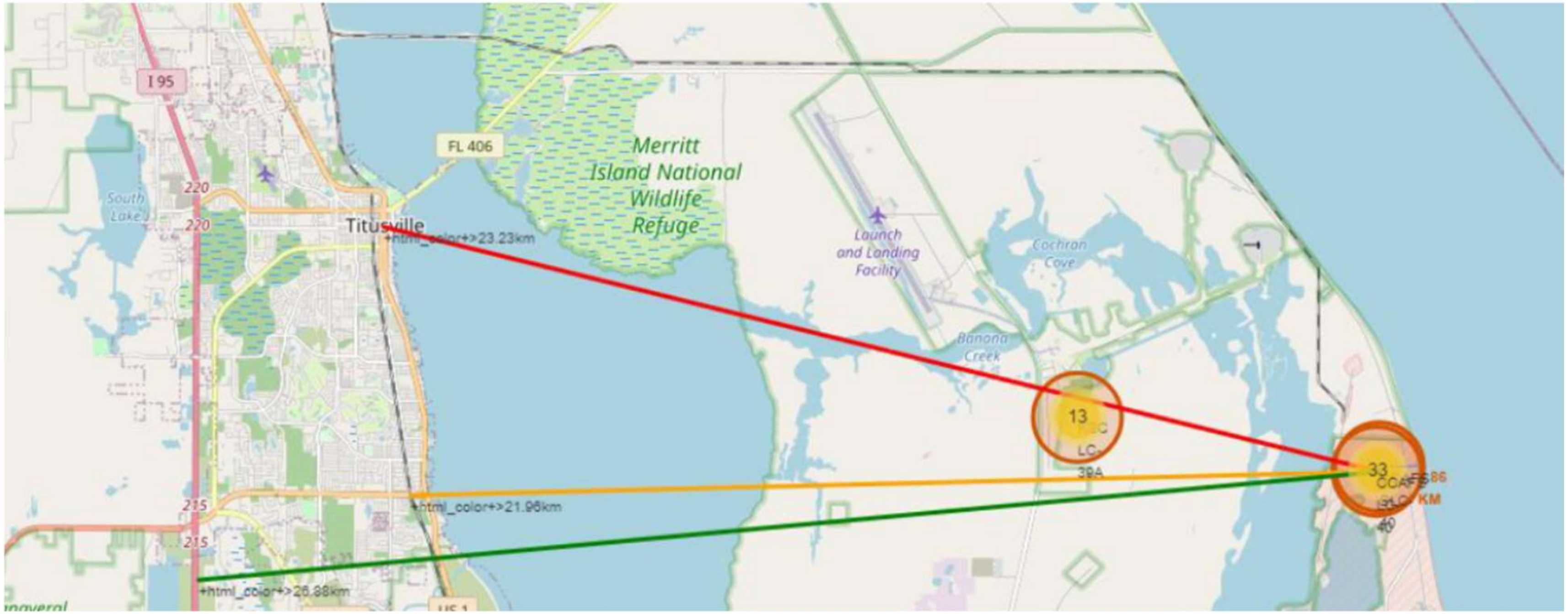
- This map displays the position of SpaceX's launch facility in the United States.
- It was generated using the folium.Map function with specific geographic coordinates.
- Locations are marked with folium.Circle, and folium.map.Marker is employed to add text labels to these points.

Launch sites outcomes on the map



- Label with the results of the landing outcomes of every mission at each launch site.
- The function used is add_child(marker_cluster).

Characteristics of the area near the launch zone

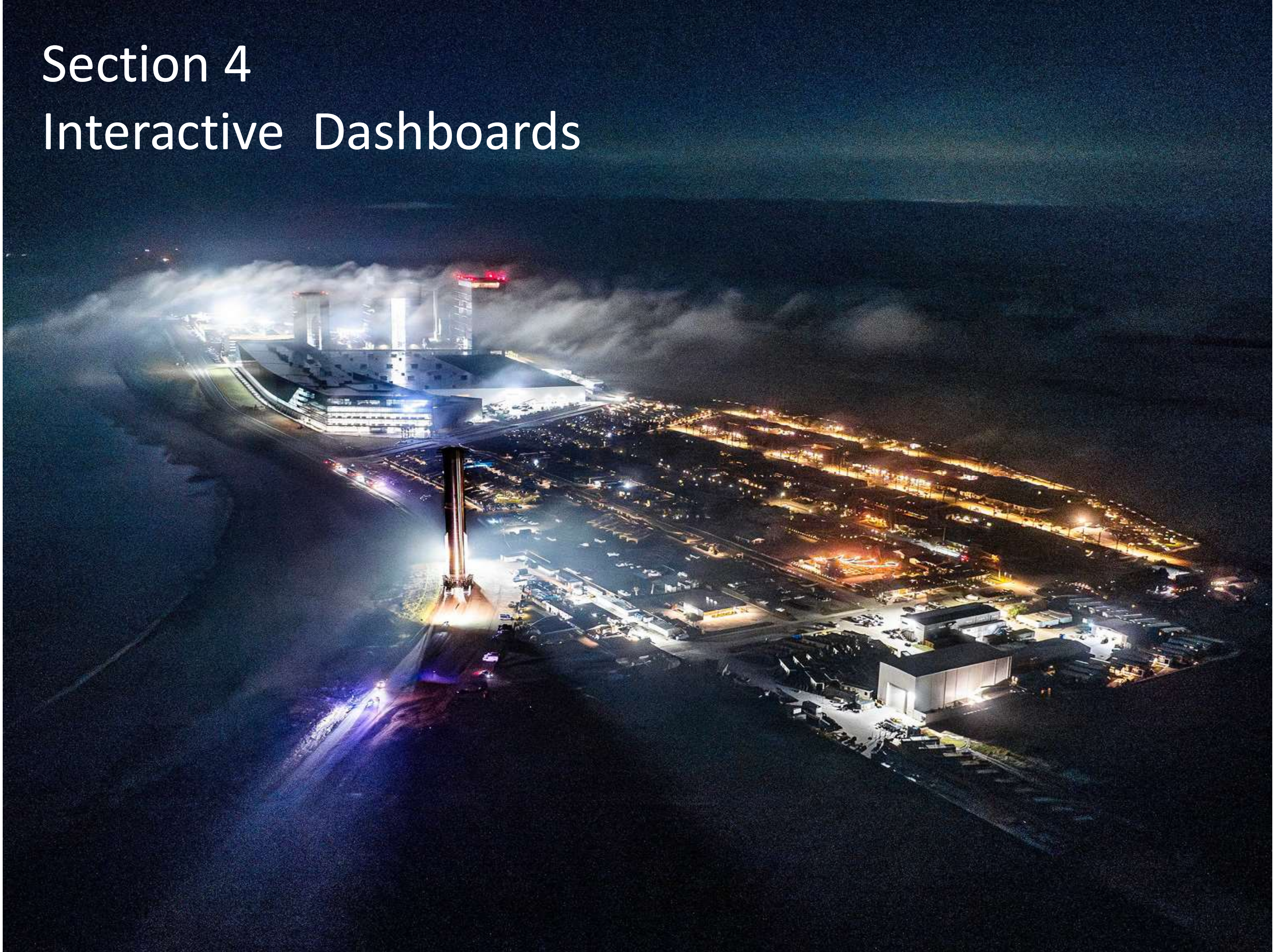


- The goal is to find a location distant enough from urban centers to prevent any impact on them in case of an accident, while remaining close enough to infrastructure to support the routine operations at the launch site.

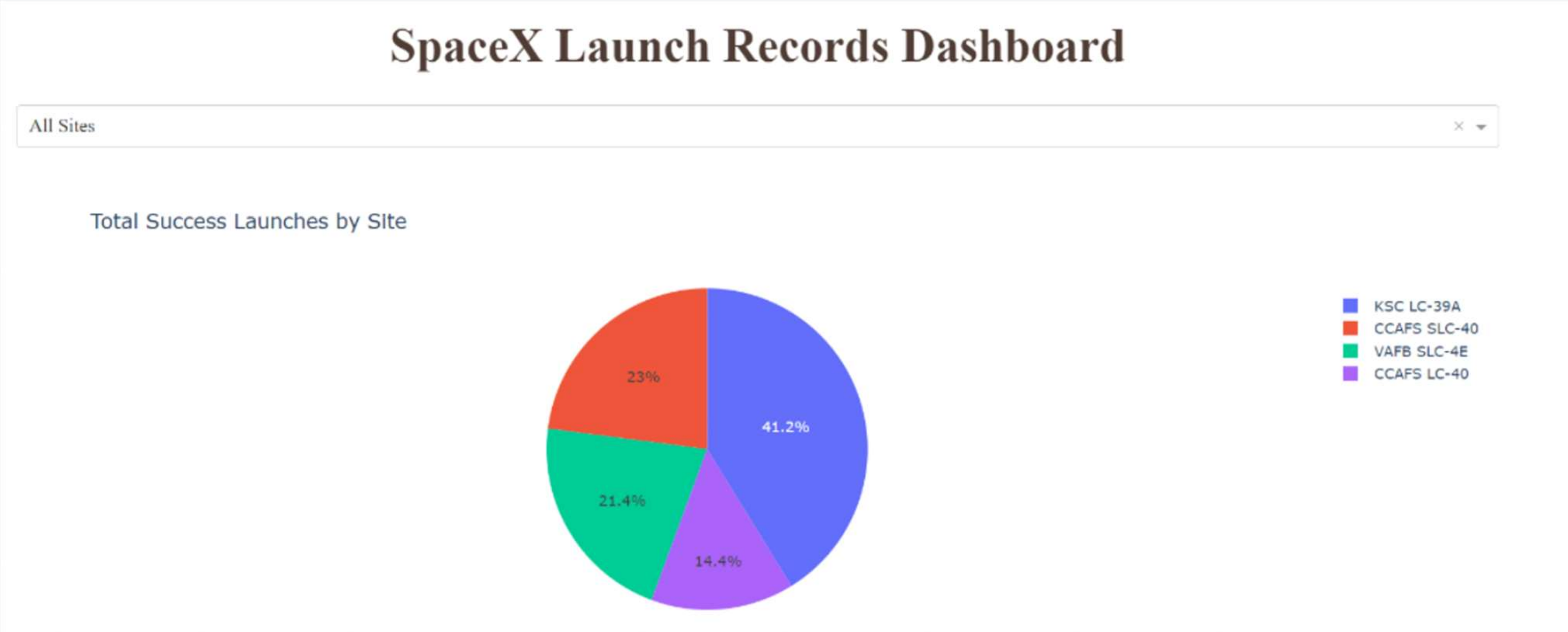
CCAFS SLC_40 (Distance to Major Infrastructure):

- Located 0.86 km from the closest coastline.
- Approximately 21.95 km away from the nearest railway.
- Around 23.23 km from the closest city.
- Situated 26.88 km from the nearest highway.

Section 4
Interactive Dashboards

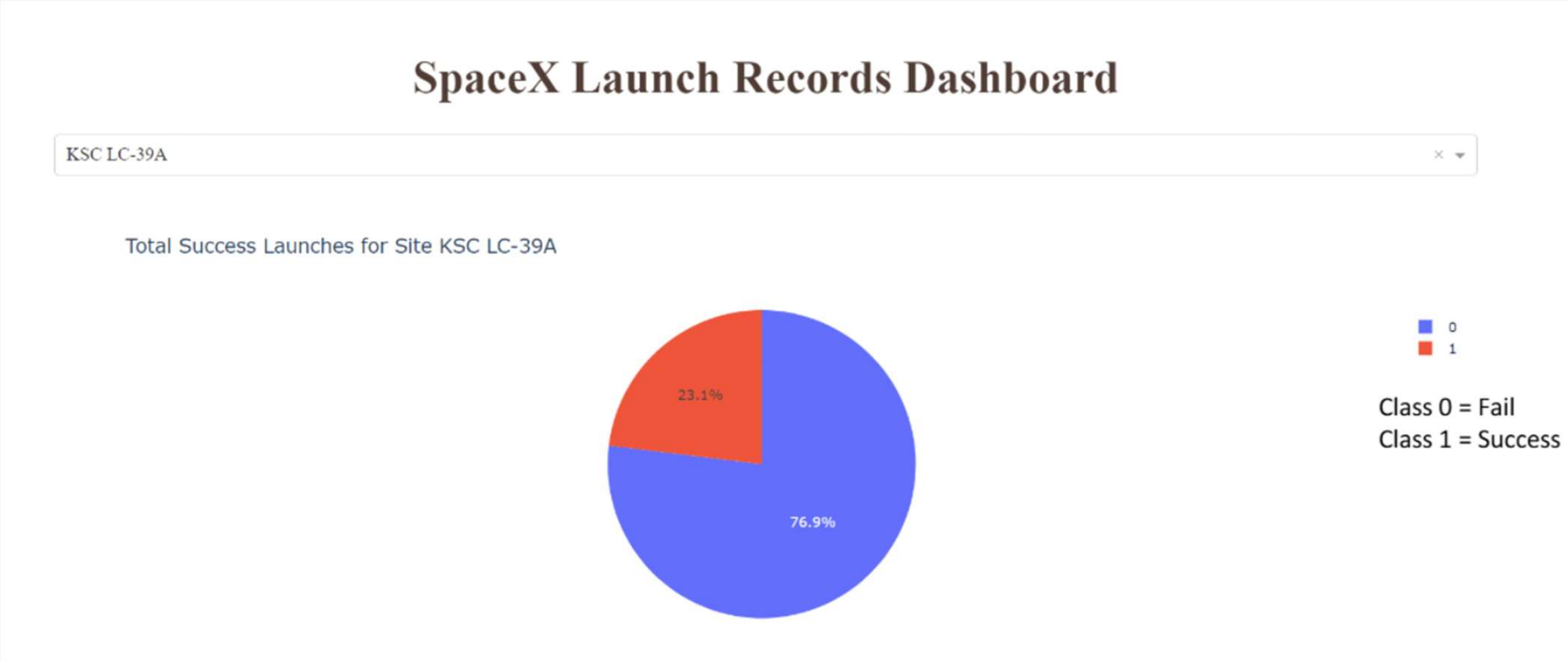


Launch success count for all sites



- This chart shows the percentage of successful launches at each site relative to the total number of successes. However, this data alone does not reveal which site has the highest success rate or which is the most successful overall.

Launch site with highest launch success ratio



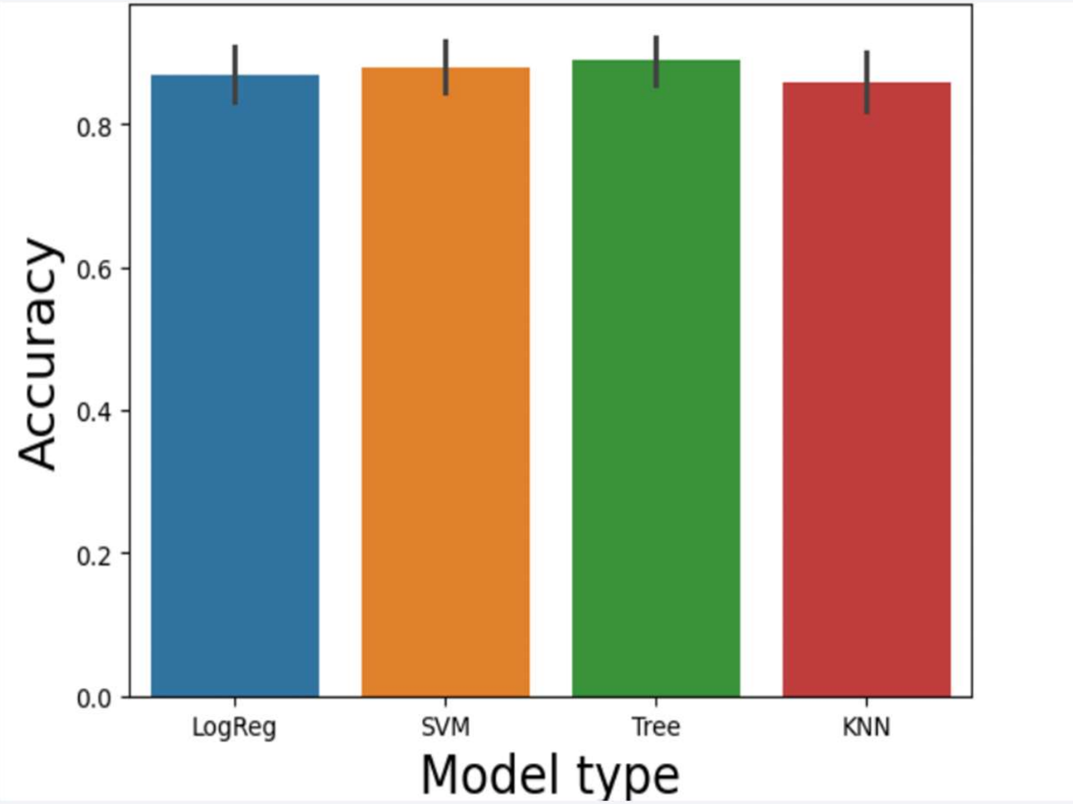
- The KSC LC-39 launch site holds the record for the highest launch success rate, making its features valuable for understanding what makes an ideal launch site.
- With this knowledge, selecting the most reliable location for a new mission should become more straightforward.

Payload vs. Launch Outcome

- This interactive chart allows you to adjust the payload mass and observe how it affects the launch results.
- Each point represents a unique mission, with the chart displaying the booster type alongside the launch success, too.

Section 5
Predictive Analysis (Classification)

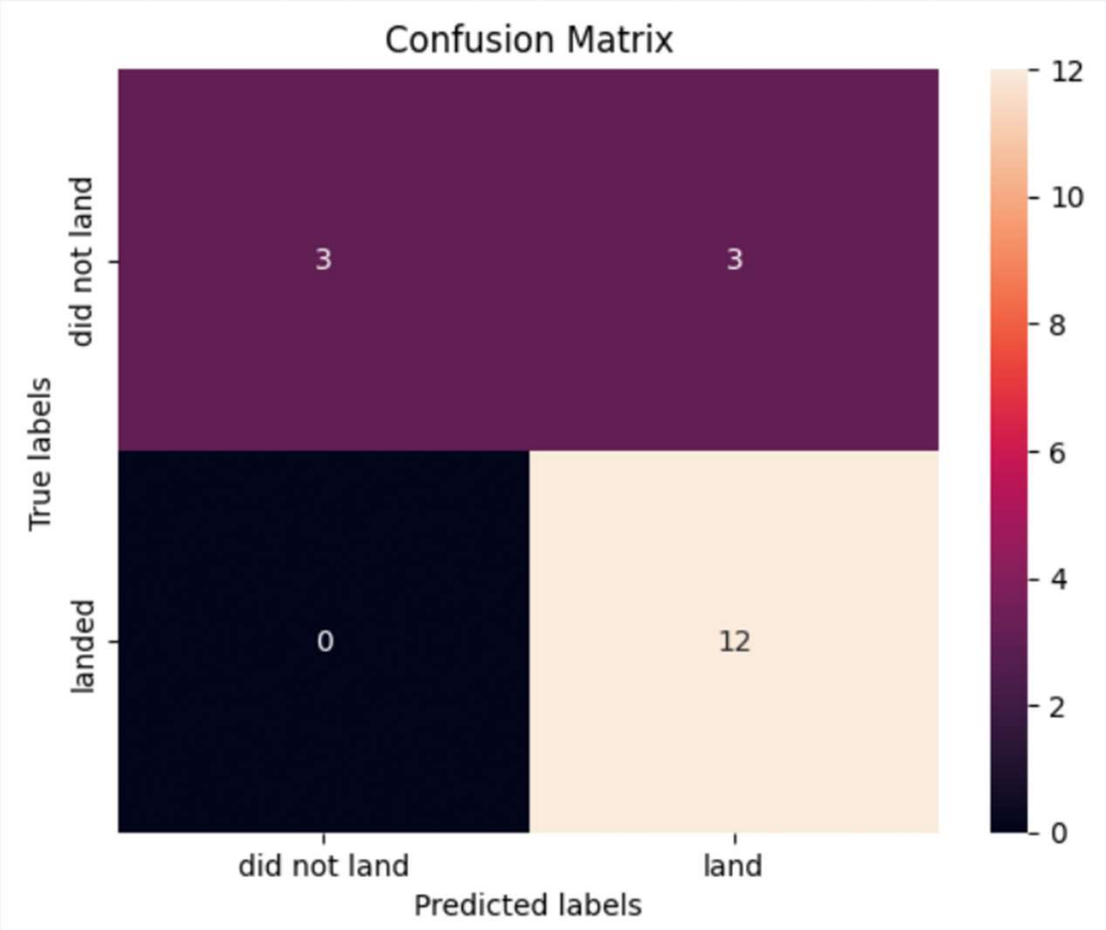
Classification Accuracy



	LogReg	SVM	Tree	KNN
Jaccard_Score	0.833333	0.845070	0.855072	0.819444
F1_Score	0.909091	0.916031	0.921875	0.900763
Accuracy	0.866667	0.877778	0.888889	0.855556

- This chart allows for comparison of the Accuracy levels among the various models.
- From the data, the Decision Tree classifier stands out with the highest accuracy, achieving a value of 0.8888.

Confusion Matrix



- This is the confusion matrix of the Decision Tree classifier.
- This chart indicates strong performance for the landed prediction: all 12 predictions were accurate.
- On the other hand, the non-landed prediction is less accurate than previously observed. Out of 6 cases were actually non-landed, 3 were predicted as landed.
- In summary, there have been instances of false positives.

Section 6
Conclusions

Conclusions

- *PROJECT GOAL*: To improve budget planning for space missions by applying prediction models that assess the likelihood of a successful landing for the Stage 1 propulsion module in each mission.
- *MODEL VARIABLES*: launch site, rocket model, payload mass, mission orbit, grid fin presence, reuse status, number of landing legs, landing pad, block number, reuse count, series, longitude, and latitude.
- *MAIN INSIGHTS ON VARIABLE FEATURES*:

1. KSC LC39-A is the launch site with the highest success rate.
2. Payload masses between 2000 and 5000 kg achieve the best success rates.
3. Orbits exceeding a 70% success rate include Low Earth Orbit (LEO) and Very Low Earth Orbit (VLEO).
4. Every mission employed the Falcon 9 booster.

Conclusions

- *OVERALL CONCLUSIONS*:

1. Success rates have increased over time.
2. Selecting a launch site requires choosing locations near the ocean, distant from urban centers, and with strong infrastructure support.
3. Proximity to the equator is crucial because the higher rotational velocity of the Earth aids in overcoming gravity during launch.

- *PREDICTIONS WITH MODELS:*

1. The Decision Tree model achieved the highest accuracy, reaching 0.88.
2. The confusion matrix reveals difficulties in identifying non-landed missions, with 6 missions that were actually non-landed, 3 of which were predicted as landed.

- *FINAL RECOMMENDATIONS:*

1. Budgeting should consider a price range that varies based on whether the First Stage module is recovered, factoring in not just the outcome of a single mission but also the anticipated number of flights for each reusable propulsion module.

Appendices

- All the specific work done, saved on GitHub:

1. https://github.com/Manu-free/Capstone_IBM/blob/main/jupyter-labs-spacex-data-collection-api.ipynb
2. https://github.com/Manu-free/Capstone_IBM/blob/main/jupyter-labs-webscraping%20(1).ipynb
3. https://github.com/Manu-free/Capstone_IBM/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb
4. https://github.com/Manu-free/Capstone_IBM/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb
5. https://github.com/Manu-free/Capstone_IBM/blob/main/edadataviz.ipynb
6. https://github.com/Manu-free/Capstone_IBM/blob/main/06_SpaceX_Interactive_Visual_Analytics_Folium.ipynb
7. https://github.com/Manu-free/Capstone_IBM/blob/main/07_SpaceX_Interactive_Visual_Analytics_Plotly.py
8. https://github.com/Manu-free/Capstone_IBM/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Thank you!