Winning Space Race with Data Science

PASALA TEJA SAIKRISHNA
26th MAY 2024

Outline

Outline

Executive Summary

- **Summary of Methodologies**

  1. Data Collection through SpaceX API
  2. Data Collection with Web Scraping from Wikipedia
  3. Data Wrangling
  4. Exploratory Data Analysis with SQL
  5. Exploratory Data Analysis with Data Visualization
  6. Interactive Visual Analytics with Folium
  7. Interactive Dashboard with Ploty Dash
  8. Machine Learning Prediction

- **Summary of all results**

  1. Exploratory Data Analysis result
  2. Interactive Visual Analytics in screenshots
  3. Interactive Dashboard in screenshots
  4. Predictive Analytics result

Introduction

- **Project background and context**

  SpaceX offers Falcon 9 launches at $62 million, a price significantly lower than competitors' rates, which start at $165 million. The key to SpaceX's pricing advantage is the reusability of the rocket's first stage. Accurately predicting the first stage's landing can help determine launch costs, providing crucial data for competitors aiming to outbid SpaceX.

Problems you want to find answers

  - How do variables such as payload mass, location of launch site and landing site, number of flights, rocket reusability and orbits affect the success of the first stage landing?
  - The rate of successful landings increase over the years?
  - The best algorithm that can be used for binary classification in this case?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was collected using a combination of the SpaceX API and web scraping techniques.
- Perform data wrangling
  - Identified missing values within the dataset and implemented strategies to address them
  - Applied one-hot encoding to convert categorical features into a numerical format
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Constructing, fine-tuning, and rigorously evaluating classification models to optimize performance and achieve the most accurate outcomes

Data Collection

The data was collected using two methods:

1. SpaceX API
   - Request to the SpaceX API and get a .json file
   - Decode the response content as a Json using .json() and turn it into a Pandas dataframe using .json_normalize()
   - We then cleaned the data, checked for missing values and fill in missing values where necessary.

2. Web Scrapping
   - Perform an HTTP GET method to request the Falcon9 Launch HTML Wikipedia page, as an HTTP response
   - Create a BeautifulSoup object from the HTML response
   - Extract the relevant tables from the page
   - Create a data frame by parsing the launch HTML tables

Data Collection – SpaceX API

1.Getting Response from API

```
[6]: spacex_url="https://api.spacexdata.com/v4/launches/past"

[7]: response = requests.get(spacex_url)
```

2.Converting Response json file to DataFrame

```
# Use json_normalize meethod to convert the json result into a dataframe
response = requests.get(spacex_url)
data = pd.json_normalize(response.json())
```

3.Custom functions to clean data

```
getBoosterVersion(data)
getLaunchSite(data)
getPayloadData(data)
getCoreData(data)
```

4. Constructing our dataset using the data

```
launch_dict = {'FlightNumber': list(data['flight_number']),
'Date': list(data['date']),
'BoosterVersion':BoosterVersion,
'PayloadMass':PayloadMass,
'Orbit':Orbit,
'LaunchSite':LaunchSite,
'Outcome':Outcome,
'Flights':Flights,
'GridFins':GridFins,
'Reused':Reused,
'Legs':Legs,
'LandingPad':LandingPad,
'Block':Block,
'ReusedCount':ReusedCount,
'Serial':Serial,
'Longitude': Longitude,
'Latitude': Latitude}
```

```
# Create a data from launch_dict
data = pd.DataFrame(launch_dict)
```

5. Filter Dataframe and export to .csv file

```
data_falcon9 = data[data['BoosterVersion']!='Falcon 1']
```

```
data_falcon9.to_csv('dataset_part_1.csv', index=False)
```

GitHub URL : SpaceX Data Collection API

Data Collection – Web Scraping

1.Requesting Falcon 9 launch data from Wikipedia

```
static_url = "https://en.wikipedia.org/w/index.php?title=List\
_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"

response = requests.get(static_url)
```

2.Creating a BeautifulSoup object from the HTML response

```
# Use BeautifulSoup() to create a BeautifulSoup object
#from a response text content
soup = BeautifulSoup(response.content,"html.parser")
```

3.Extracting all column names from the HTML table header (refer block 12 in notebook)

```
[12]: column_names = []

      # Apply find_all() function with `th` element on firs

      cols = first_launch_table.find_all('th')

      # Iterate each th element and apply the provided extr
      col_names = []
      for row in cols:
          cnam = extract_column_from_header(row)
          col_names.append(cnam)
```

4.Collecting the data by parsing HTML tables into a dictionary (refer block 15 in notebook)

```
[15]: extracted_row = 0
      #Extract each table
      for table_number,table in enumerate(soup.find_all('table',"wikitable
         # get table row
          for rows in table.find_all("tr"):
              #check to see if first table heading is as number correspondi
              if rows.th:
                  if rows.th.string:
                      flight_number=rows.th.string.strip()
                      flag=flight_number.isdigit()
              else:
                  flag=False
```

5.Creating a dataframe from the dictionary

```
df= pd.DataFrame({ key:pd.Series(value) for key, value in launch_dict.items() })
```

6.Exporting the data to CSV

```
df.to_csv('spacex_web_scraped.csv', index=False)
```


GitHub URL : SpaceX Data Collection with Web Scraping

Data Wrangling

- In Data wrangling, we create a binary landing outcome variable **"Class"** from the categorical landing outcome.
    - **True Ocean -** successfully landed to a specific region of the ocean
    - **False Ocean -** unsuccessfully landed to a specific region of the ocean.
    - **True RTLS -** successfully landed to a ground pad
    - **False RTLS -** unsuccessfully landed to a ground pad.
    - **True ASDS -** successfully landed to a drone ship
    - **False ASDS -** unsuccessfully landed to a drone ship.
    - **None ASDS** and **None None**, these represent a failure to land.
- We categorize the booster landing outcomes into a binary '**Class**' training label, where a '1' indicates a successful landing and a '0' denotes an unsuccessful attempt.

GitHub URL : SpaceX Data Wrangling

Perform exploratory Data Analysis and determine Training Labels

↓

Calculate the number of launches on each site

↓

Calculate the number and occurrence of each orbit

↓

Calculate the number and occurrence of mission outcome per orbit type

↓

Create a landing outcome label from Outcome column

↓

Exporting the data to CSV

EDA with Data Visualization

Charts were plotted:

1. Flight Number vs. Payload Mass (SP)
2. Flight Number vs. Launch Site (SP)
3. Payload Mass vs. Launch Site (SP)
4. Orbit Type vs. Success Rate (BC)
5. Flight Number vs. Orbit Type (SP)
6. Payload Mass vs Orbit Type (SP)
7. Success Rate Yearly Trend (LC)

- **Scatter Plots (SP):** Scatter plots are graphical representations that display the relationship between two quantitative variables. Each point on the plot corresponds to an observation in the dataset.
- **Bar Charts (BC):** Bar charts are used to compare quantities across different categories. By displaying bars of varying lengths, they visually represent the relationship and differences between discrete categories and associated values.
- **Line Charts (LC):** Line charts are ideal for illustrating trends and changes over time. By connecting data points with a continuous line, they help in identifying upward or downward trends.

GitHub URL : SpaceX EDA with Visualization

EDA with SQL

Summary of SQL Queries Performed:

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first succesful landing outcome in ground pad was acheived
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
- 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

GitHub URL : SpaceX EDA with SQL

Build an Interactive Map with Folium

Adding Markers:

- We’ve successfully added a marker for the **NASA Johnson Space Center**, complete with a circle, popup label, and text label, using its precise latitude and longitude coordinates. This serves as our initial reference point for the map.
- Building upon this, we’ve also incorporated markers for various launch sites. Each marker is designed to provide a comprehensive view of the site’s geographical context and its strategic location relative to the Equator and coastal lines. Here’s what we’ve included for each site:
    - **Circle**: A visual representation of the surrounding area, giving viewers a sense of scale and location.
    - **Popup Label**: An interactive element that offers detailed information about the launch site.
    - **Text Label**: A clear and concise label that identifies each launch site by name directly on the map.

GitHub URL : SpaceX_Interactive Visual Analytics with Folium

Build an Interactive Map with Folium

- **Color-Coded Success and Failure:**
  - To distinguish success rates, we've color-coded the markers:
    - **Green**: Represents successful launches.
    - **Red**: Indicates failed launches.
  - By analyzing these markers, viewers can identify which launch sites have relatively high success rates.
- **Marker Clusters for Launch Site:**
  - To maintain visual clarity, we've implemented **Marker Clusters**. Even when multiple markers are close together, the map remains organized.
- **Distance Visualization to Key Infrastructure:**
  - We've taken it a step further by adding colored lines to represent distances between each launch site and nearby infrastructure:
    - **Railways**: Highlighting transportation connectivity.
    - **Highways**: Indicating accessibility.
    - **Coastlines**: Emphasizing proximity to water bodies.
    - **Closest Cities**: Showing urban centers in the vicinity.

GitHub URL [: SpaceX_Interactive Visual Analytics with Folium]

Build a Dashboard with Plotly Dash

- Dropdown List of Launch Sites
  - Implemented an interactive dropdown menu to facilitate the selection of a Launch Site, enhancing the user interface for streamlined operations.
- Creation of Pie Chart
  - Integrated a dynamic pie chart to display the aggregate count of successful launches across all sites, and a comparative analysis of Success versus Failure rates for individual Launch Sites upon selection
- Payload Mass Range Slider
  - Enhanced the data analysis interface with an interactive slider for precise payload range selection, optimizing user engagement and data exploration
- Payload Mass vs. Success Rate Scatter Chart
  - Implemented a scatter chart to effectively illustrate the correlation between payload mass and launch success rates across various booster versions, providing a clear visual analysis of the dataset

GitHub URL : SpaceX Interactive Dashboard with Ploty Dash

Predictive Analysis (Classification)

1. Create a NumPy array from the column 'Class' in dataframe
2. Standardize and transform the data in X then reassign it to self
3. Use the function train_test_split to split the data X and Y into training and test data
4. Create a GridSearchCV object with cv = 10 to find the best parameters for each algorithm
5. Applying GridSearchCV on Logistic Reg, SVM, Decision Tree and KNN models
6. Calculating the accuracy on the test data using the method .score() for all models
7. Examining the confusion matrix for all models
8. Finding the method that performs best by examining the accuracy

GitHub URL : SpaceX Machine Learning Prediction

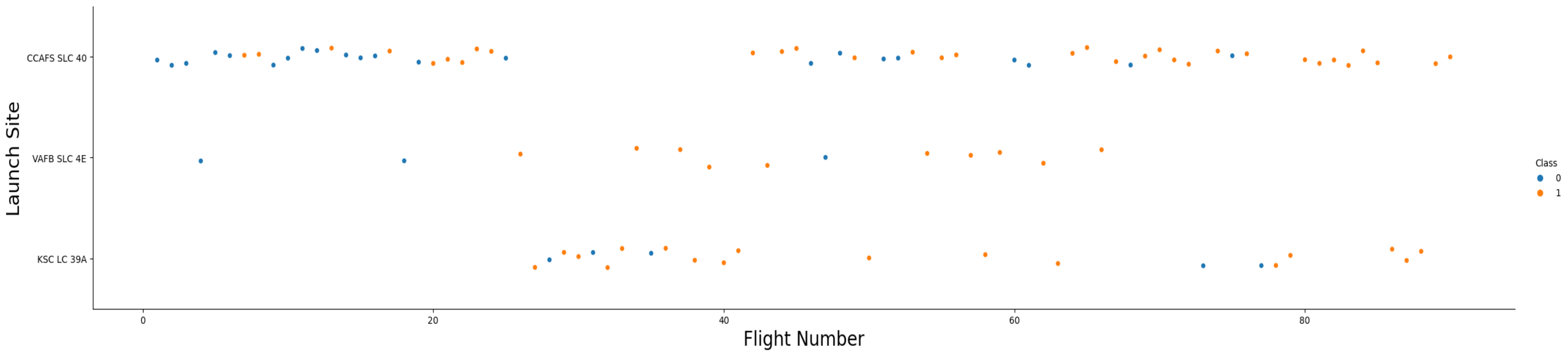
Results

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

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site



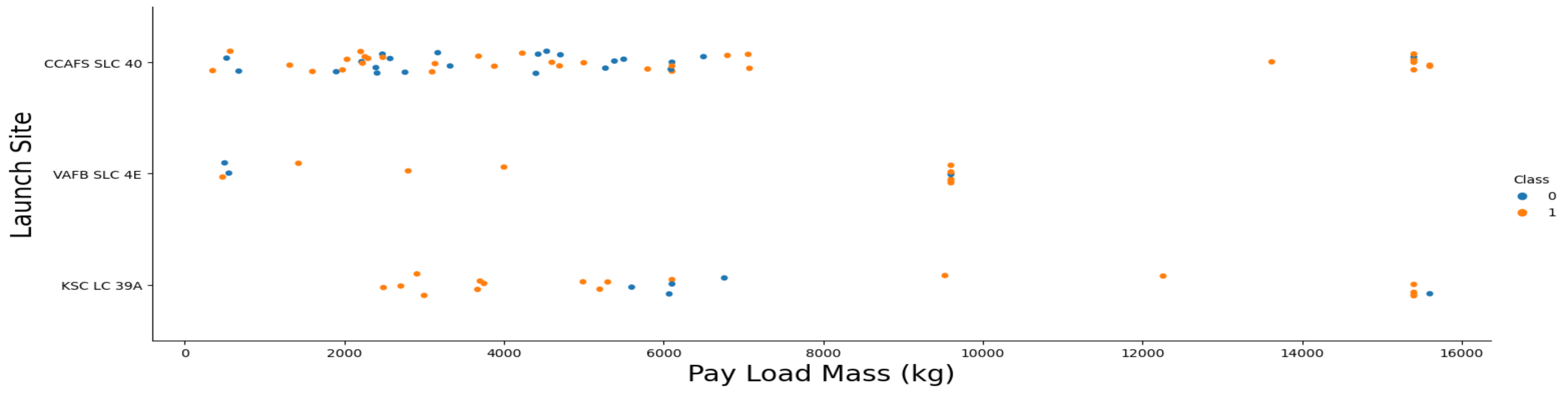
- **Initial Launch Outcomes:** There was a clear trend observed in the early stages of the space program where the earliest flights encountered failures. However, as the program matured and technology advanced, the most recent flights have been consistently successful.

- **Launch Site Distribution:** The CCAFS SLC 40 launch site has been a pivotal location for space endeavors, accounting for approximately 50% of all launches. This site has played a significant role in the history and development of space exploration.

- **Success Rates by Site:** When comparing launch sites, VAFB SLC 4E and KSC LC 39A stand out with notably higher success rates. These sites have demonstrated a reliable track record, contributing to the overall success of the space program.

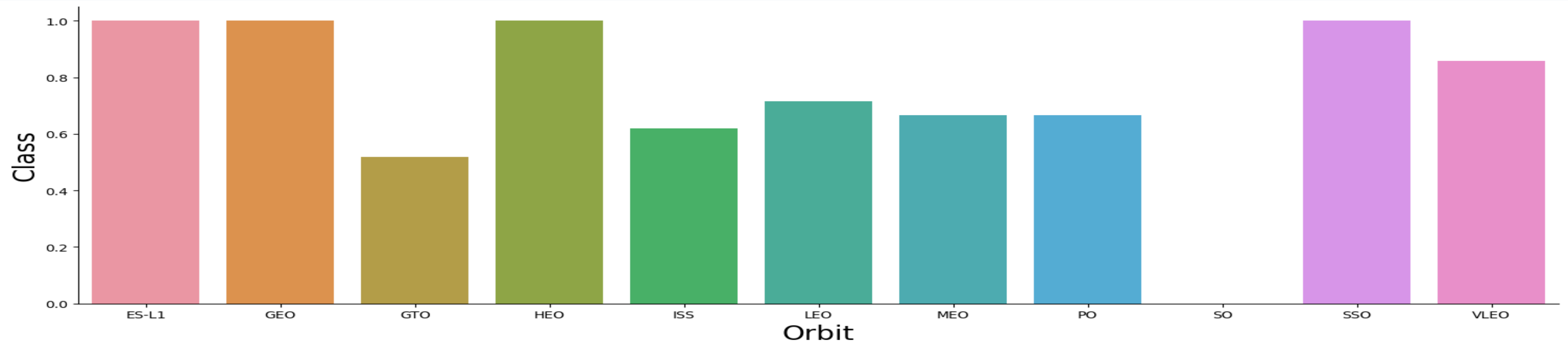
- **Progressive Improvement:** Analyzing the data trends, it's reasonable to infer that each subsequent launch is more likely to succeed than the previous one. This progressive improvement can be attributed to the accumulation of knowledge, experience, and technological advancements in aerospace engineering

Payload vs. Launch Site



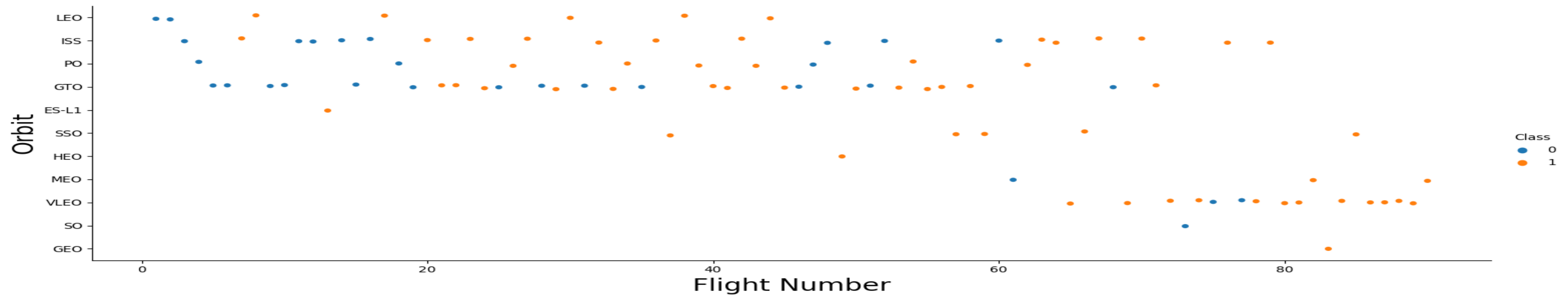
- Payload Mass vs. Success Rate:
  - There is a clear trend: higher payload mass tends to correlate with a higher success rate for rocket launches.
  - Most launches with a payload mass exceeding 7000 kg were successful.
- KSC LC 39A: For payload mass under 5500 kg, it boasts an impressive 100% success rate

Success Rate vs. Orbit Type



- **Orbits with 100% Success Rate:** ES-L1, GEO, HEO, SSO. These orbits have demonstrated a flawless track record in mission success.
- **Orbits with 0% Success Rate: SO** (Suborbital)
- **Orbits with Variable Success Rates (50% - 85%):**GTO, ISS, LEO, MEO, PO. These orbits show a moderate level of success, reflecting a mix of mission complexities and varying degrees of technical challenges

Flight Number vs. Orbit Type



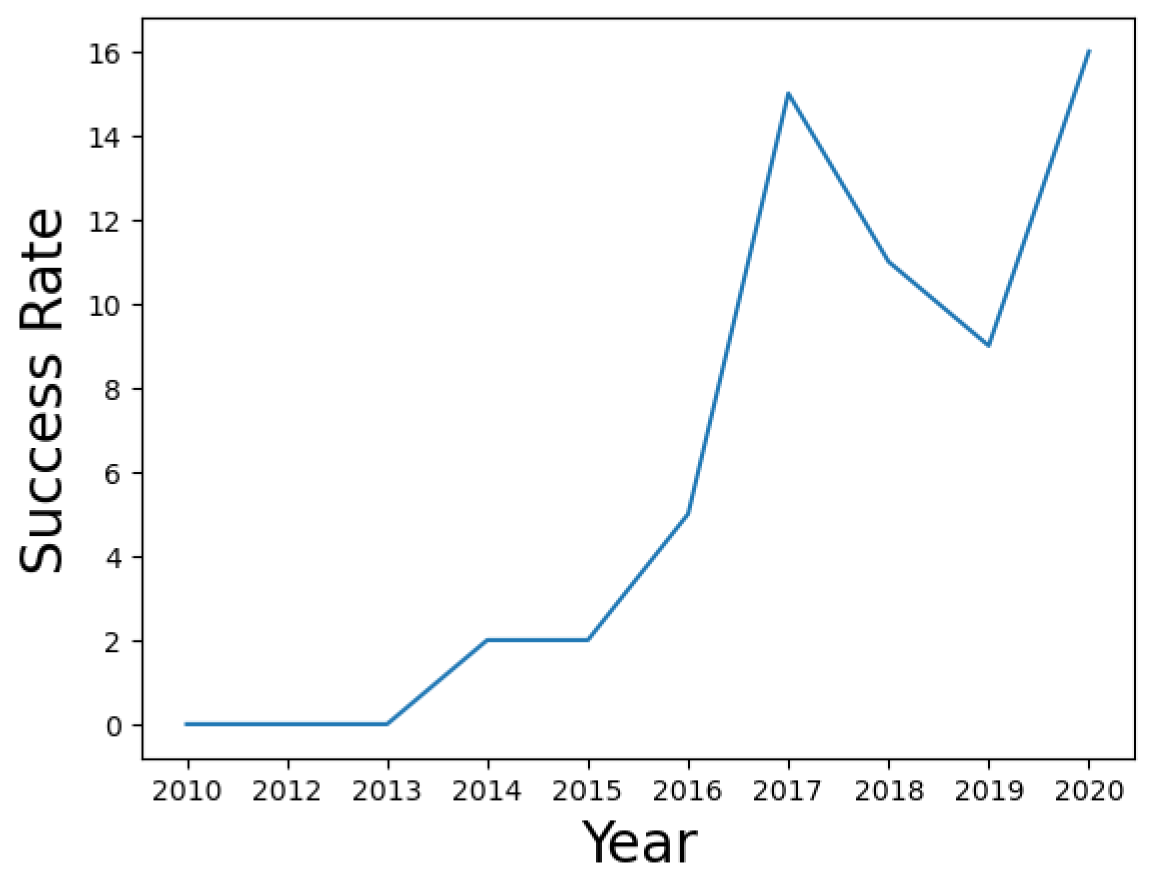
- **LEO - Success Correlation:**
  - In LEO, there is a positive correlation between the number of flights and mission success.
  - This suggests that frequent flights may contribute to higher success rates, possibly due to improved operational experience and refined technology over time.
- **GTO - Lack of Correlation:**
  - Contrarily, in the GTO orbit, the number of flights does not appear to have a discernible impact on mission success.
  - This indicates that factors other than flight frequency, such as technical complexities or mission-specific challenges, play a more significant role in determining success in GTO missions

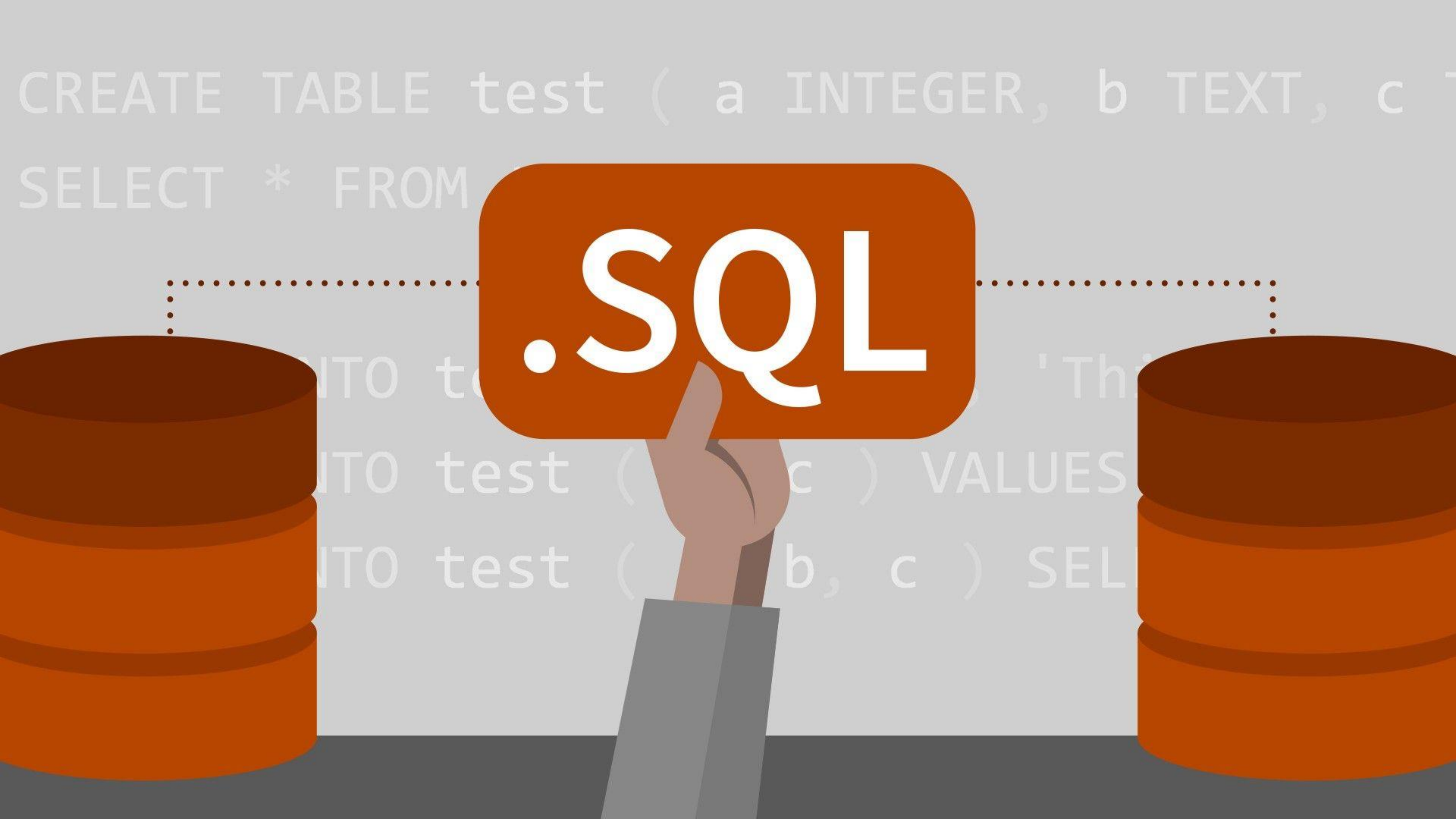
Payload vs. Orbit Type

- **GTO (Geostationary Transfer Orbit):** There seems to be no observable correlation between the payload mass and the success rate for missions to GTO. This suggests that success in this orbit may depend more on other factors, such as technical specifications or mission design.
- **ISS (International Space Station) Orbit:** The ISS orbit accommodates a diverse range of payload masses and maintains a high success rate. This versatility reflects the extensive experience and robust infrastructure supporting ISS missions.
- **SO (Suborbital) and GEO (Geostationary Earth Orbit):** Missions to these orbits are less frequent. The limited number of launches could be due to specific mission requirements or the complexity associated with these orbits.

Launch Success Yearly Trend

- **Trend in Success Rate:**
    - The success rate for launches began to climb in 2013 and continued to rise through 2020. This upward trend reflects a period of significant advancements and optimizations in space technology.
- **Initial Years of Development:**
    - The initial three years appear to have been a critical phase of adjustment and technological refinement. This period likely involved iterative improvements and learning from early challenges, contributing to the later success.




CREATE TABLE test (a INTEGER, b TEXT, c
SELECT * FROM
.SQL

All Launch Site Names

- Names of the unique launch sites in the space mission

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

- Unique Launch Site Selection:
  - The dataset is refined by extracting distinct “launch_site” values, ensuring that each launch site is represented only once

Launch Site Names Begin with 'CCA'

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

```
[11]: %%sql
select * FROM SPACEXTABLE WHERE "Launch_Site" LIKE "CCA%" LIMIT 5

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

[11]:

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

- This query is designed to efficiently fetch the first 5 records from the launch_data table where the launch_site values begin with ‘CCA’. The LIKE operator is used to match the specified pattern, and the LIMIT clause restricts the output to a manageable number of records.

Total Payload Mass

- The total payload carried by boosters from NASA

```
[12]: %%sql
SELECT  sum("PAYLOAD_MASS__KG_") as "Total Payload Mass" FROM SPACEXTABLE
WHERE "Customer" = "NASA (CRS)"

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

[12]:

Total Payload Mass
45596

- Total Payload Mass for NASA (CRS) Missions:
  - The cumulative payload mass transported by boosters on NASA's Commercial Resupply Services (CRS) missions can be displayed using a SQL query using sum function

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1

```
[26]: %%sql
SELECT  CAST(AVG("PAYLOAD_MASS__KG_") AS INT) as "Average Payload Mass(KG)" FROM SPACEXTABLE
WHERE Booster_Version Like "%F9 v1.1%"

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

[26]:

Average Payload Mass(KG)
2534

- Average Payload Mass for F9 v1.1:
  - The average payload mass transported by the F9 v1.1 booster is approximately 2534 kg. This value represents the typical payload capacity of this specific version of the Falcon 9 rocket.

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad

```
[28]: %%sql
select min(Date) as "First Successful Landing" FROM SPACEXTABLE WHERE "Landing_Outcome" LIKE "%ground pad%"

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

[28]:

First Successful Landing
2015-12-22

- Historic First Successful Landing:
  - On December 22, 2015, SpaceX achieved a groundbreaking milestone by successfully landing the first stage of its Falcon 9 rocket vertically. This achievement marked the inaugural instance of a rocket booster returning to Earth intact during an orbital launch

Successful Drone Ship Landing with Payload between 4000 and 6000

- Based on the filters provided, the following four booster versions satisfy the conditions:

```
[15]: %%sql
select Booster_Version as Boosters FROM SPACEXTABLE
WHERE "PAYLOAD_MASS__KG_">4000 and "PAYLOAD_MASS__KG_"<6000 and "Landing_Outcome" =="Success (drone ship)"

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

[15]:

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

- List of names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes

```
%%sql
select Mission_Outcome , count(*) AS "Total Number" FROM SPACEXTABLE
GROUP BY Mission_Outcome like "%Success%"
```

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

Mission_Outcome	Total Number
Failure (in flight)	1
Success	100

- This query will:
  - Select records from SPACEXTABLE.
  - Filter the results to only include rows where Mission_Outcome contains the string "Success".
  - Count the total number of these filtered results.
  - Group the results by Mission_Outcome

Boosters Carried Maximum Payload

- This query will return the booster versions that have carried the heaviest payload, according to the SPACEXTABLE Table

```
%%sql
select Booster_Version as Boosters FROM SPACEXTABLE
WHERE "PAYLOAD_MASS__KG_" == (SELECT  MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE)
```

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

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

2015 Launch Records

- List of failed landing_outcomes in drone ship, their booster versions, and launch site names in year 2015

```
%%sql
select substr(Date, 6,2) as "Month Name","Landing_Outcome" as "Landing Outcome",
Booster_Version as Boosters,"Launch_Site" as "Launch Site" FROM SPACEXTABLE
where substr(Date,0,5) =='2015' and "Landing_Outcome" =="Failure (drone ship)"
```

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

Month Name	Landing Outcome	Boosters	Launch Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

In 2015, there were two failed landing outcomes on a drone ship:

- Booster Version: F9 v1.1 B1012
- Launch Site: Cape Canaveral1
- Booster Version: F9 v1.1 B1015
- Launch Site: Cape Canaveral

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
- This query will group the landing outcomes, count the number of occurrences for each outcome, and then order the results in descending order based on the count. You'll get a ranked list of outcomes, from the most frequent to the least frequent within the specified date range.

```
%%sql
select "Landing_Outcome", COUNT(Landing_Outcome) as "Landing Outcome Count" FROM SPACEXTABLE
WHERE "Date" BETWEEN '2010-06-04' and '2017-03-20'
GROUP BY "Landing_Outcome"
ORDER BY COUNT(Landing_Outcome) DESC
```

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

Landing_Outcome	Landing Outcome Count
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

Launch Sites Proximities Analysis

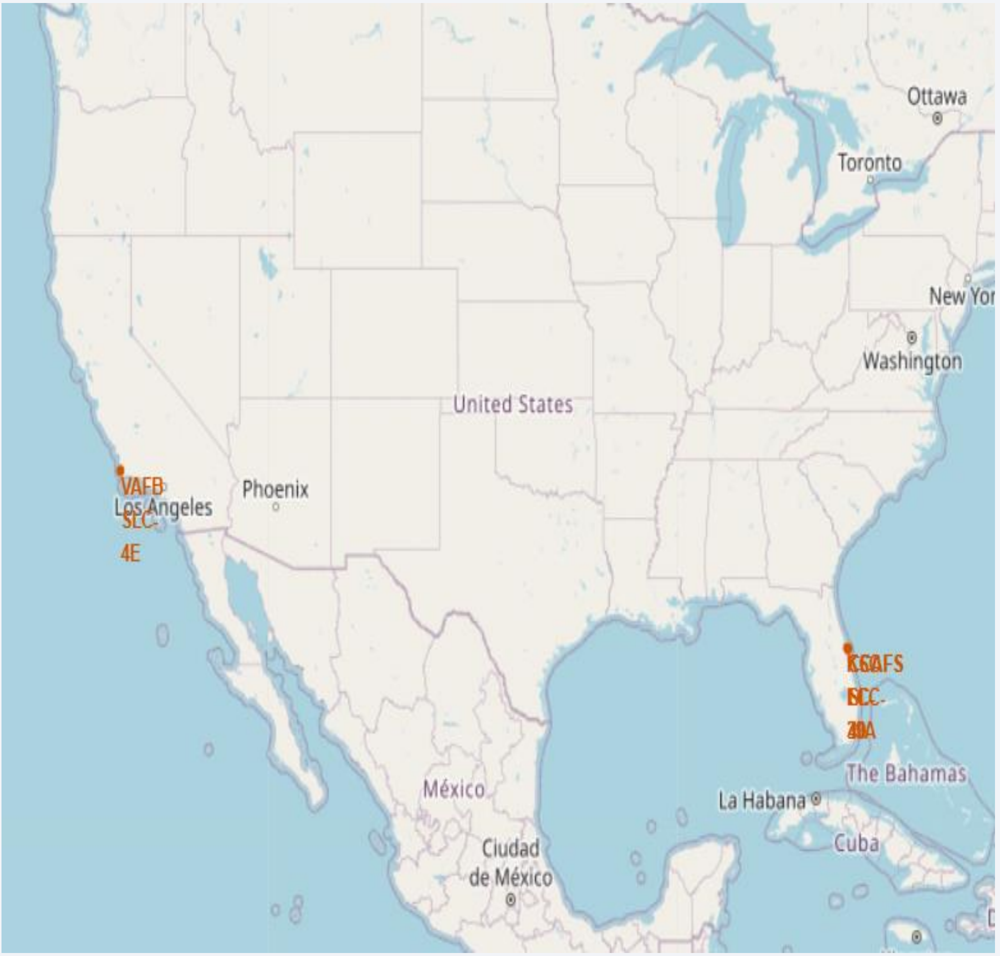
All Launch Sites Location Markers on a global map

Proximity to the Equator line:

- Launching rockets from near the Equator provides an additional boost due to the rotational speed of Earth.
- Launching eastward from the Equator ensures that any debris or malfunctioning stages fall into the ocean
- Proximity to the Equator leads to a substantial increase in payload capacity.

Very Close Proximity To The Coast:

- Coastal locations allow for launches over the ocean, reducing the risk of debris falling over populated areas in case of a launch failure.
- Being near the coast provides more options for the rocket's flight path.
- Proximity to the coast facilitates easier transportation by sea, which is often more practical and cost-effective than land transport.
- Launches can have significant environmental impacts, such as noise and pollution. Coastal sites help mitigate these effects by directing them over the ocean.



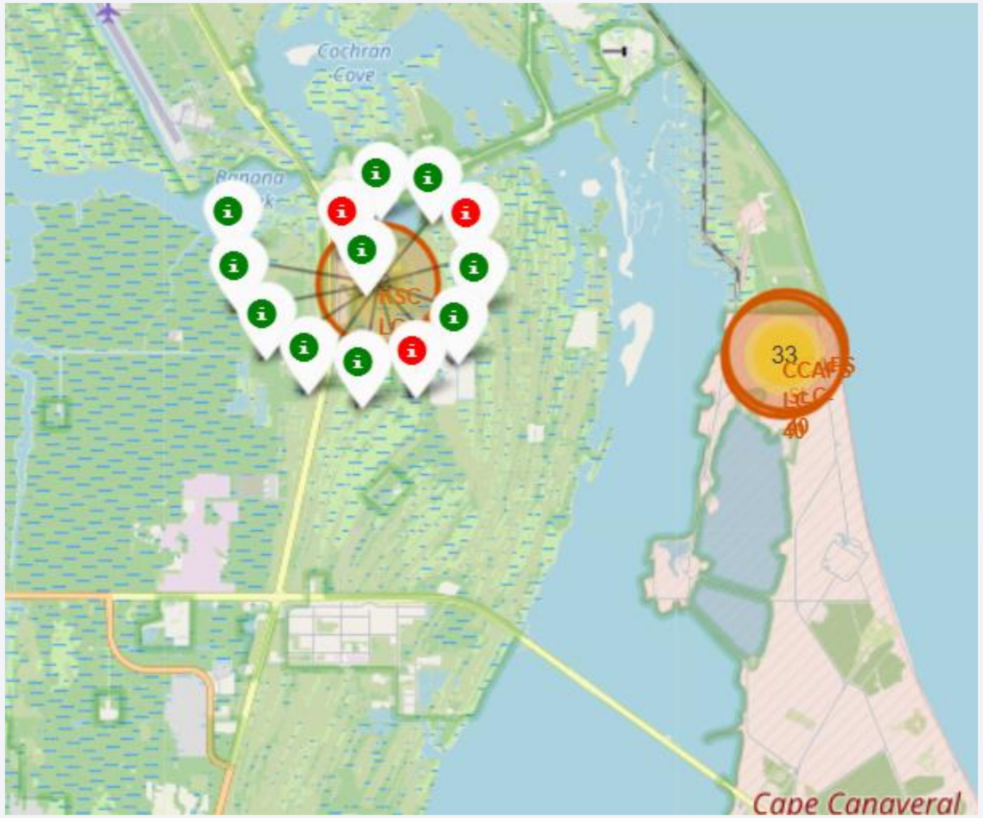
Color Labeled Launch Records

- Color-labeled markers to identify launch sites based on their success rates:

  Green Marker: Represents successful launches.

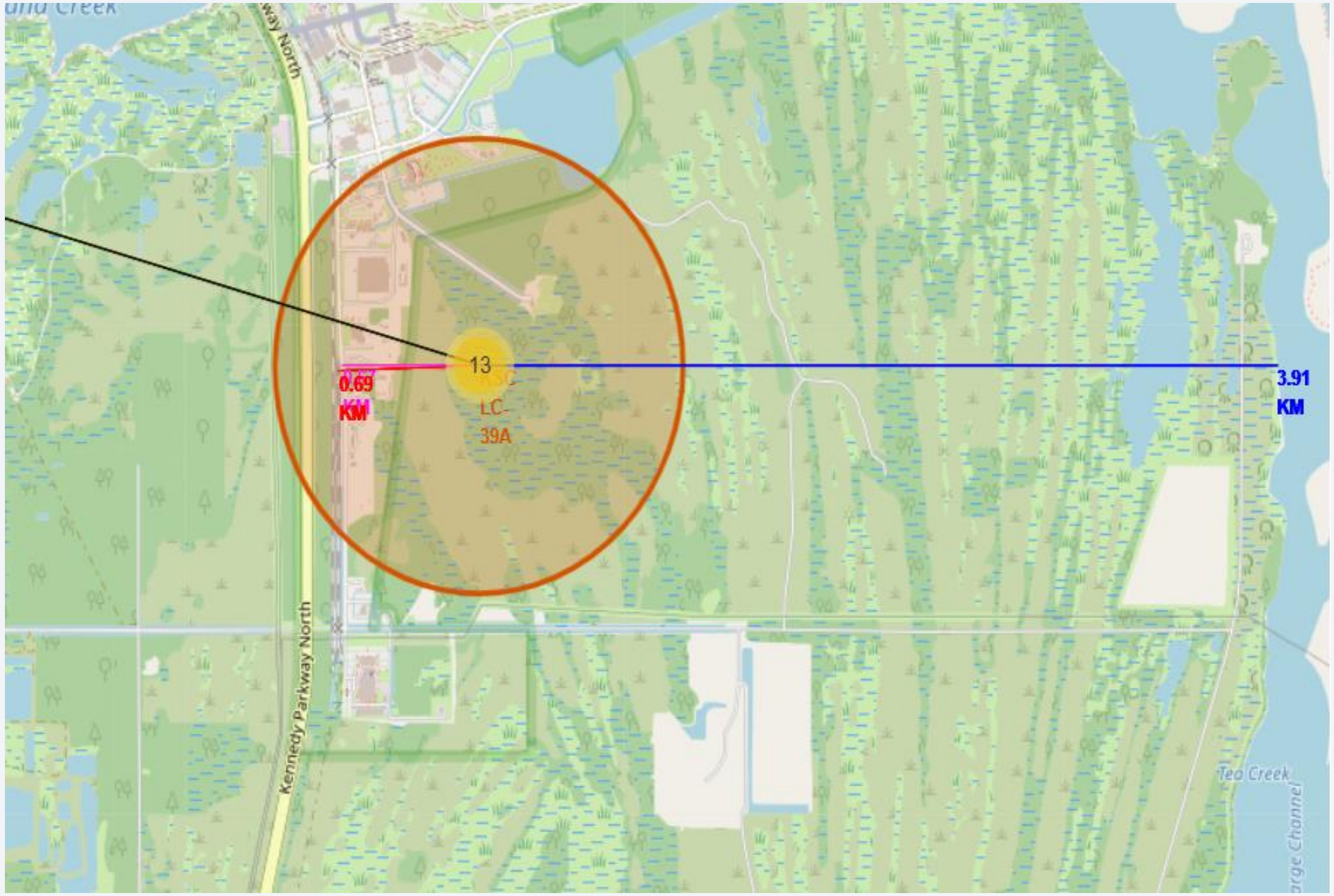
  Red Marker: Represents failed launches.

- SpaceX's Falcon 9 and Falcon Heavy rockets frequently lift off from LC-39A. These rockets have achieved an impressive success rate, delivering payloads to orbit, launching satellites, and supporting scientific missions



Distances Between A Launch Site To Its Proximities

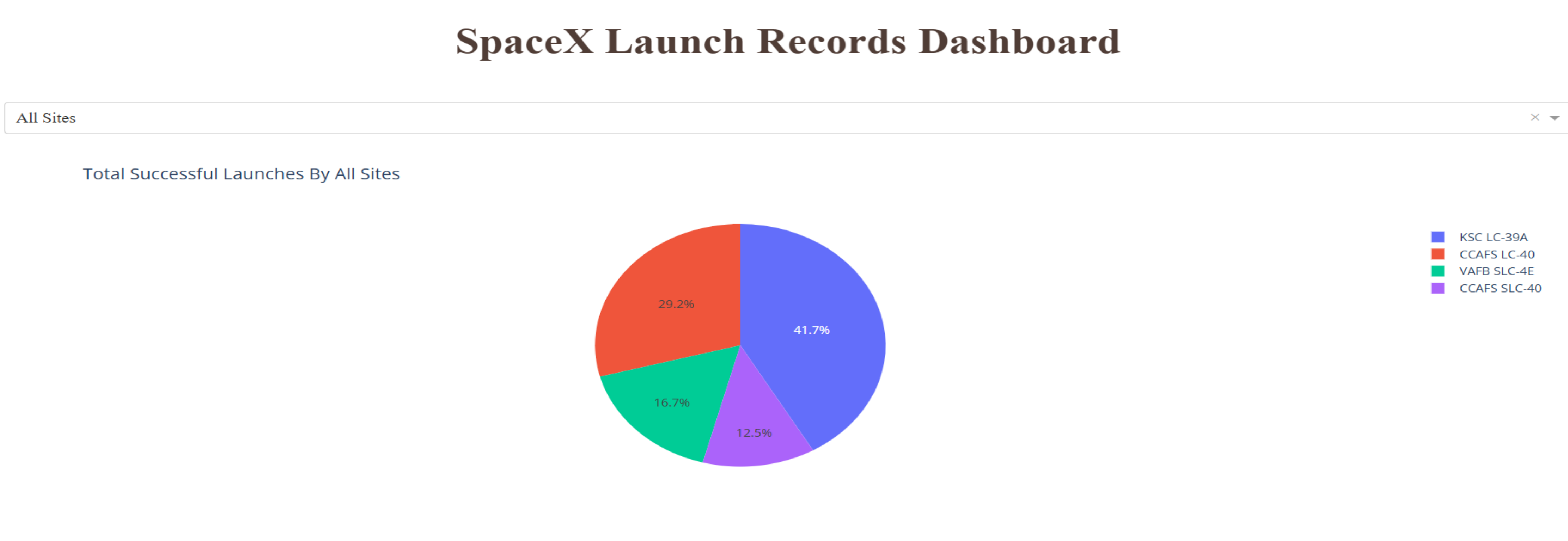
- **Launch Site KSC LC-39A Close Proximity**
  - Railway line – 0.69 KM
  - Highway – 0.61 KM
  - Coast – 3.91KM
  - Nearest City – 16.31 KM (Titusville)
- Launch Site is closer to transportation and farther to city due to a failed rocket, can indeed cover distances of 15-20 kilometers in just a matter of seconds



Section 4

Build a Dashboard with Plotly Dash

Successful Launches of All Launch Sites

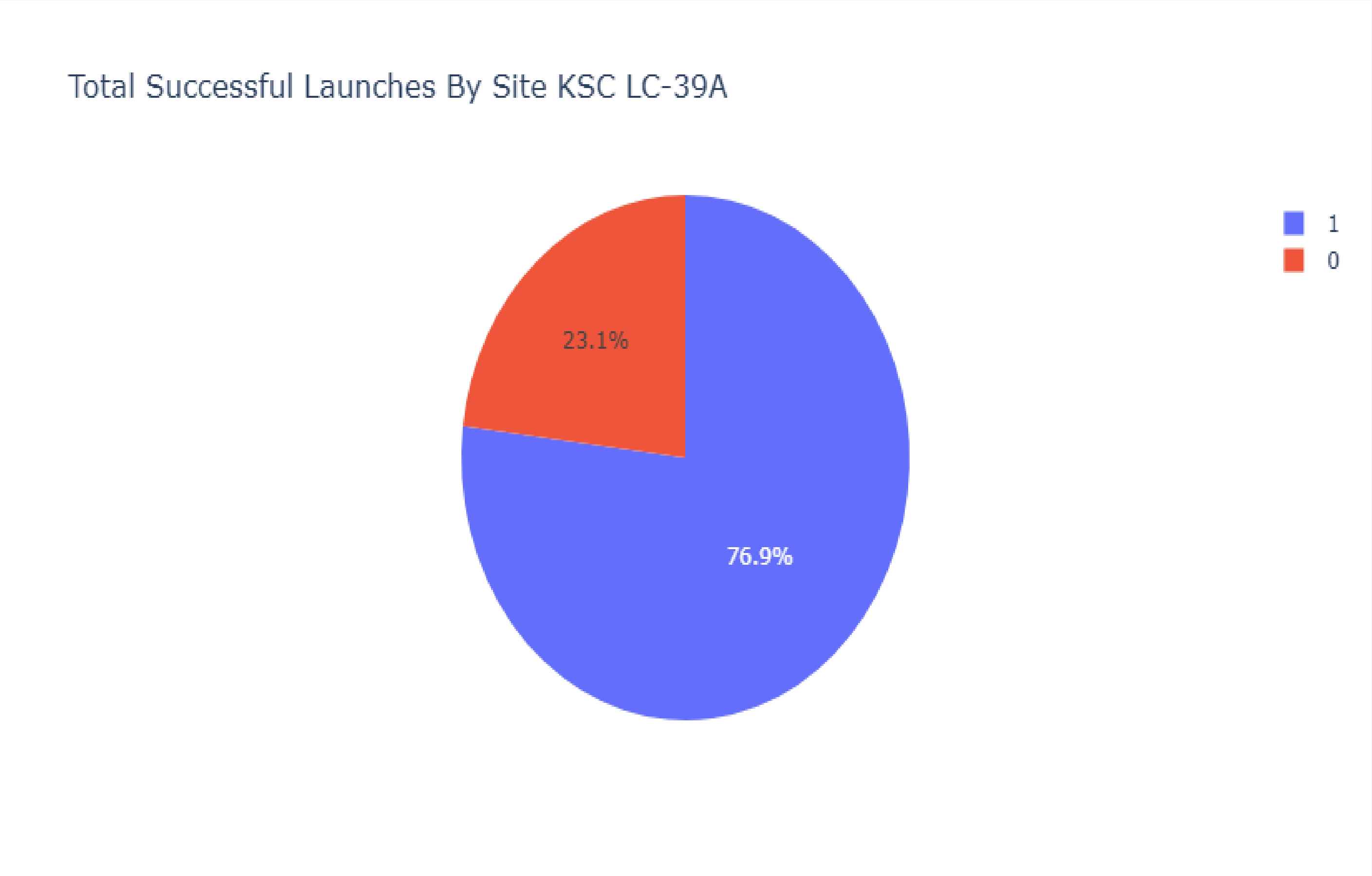


- The chart unmistakably highlights that KSC LC-39A stands out with the most successful launches of all launch sites

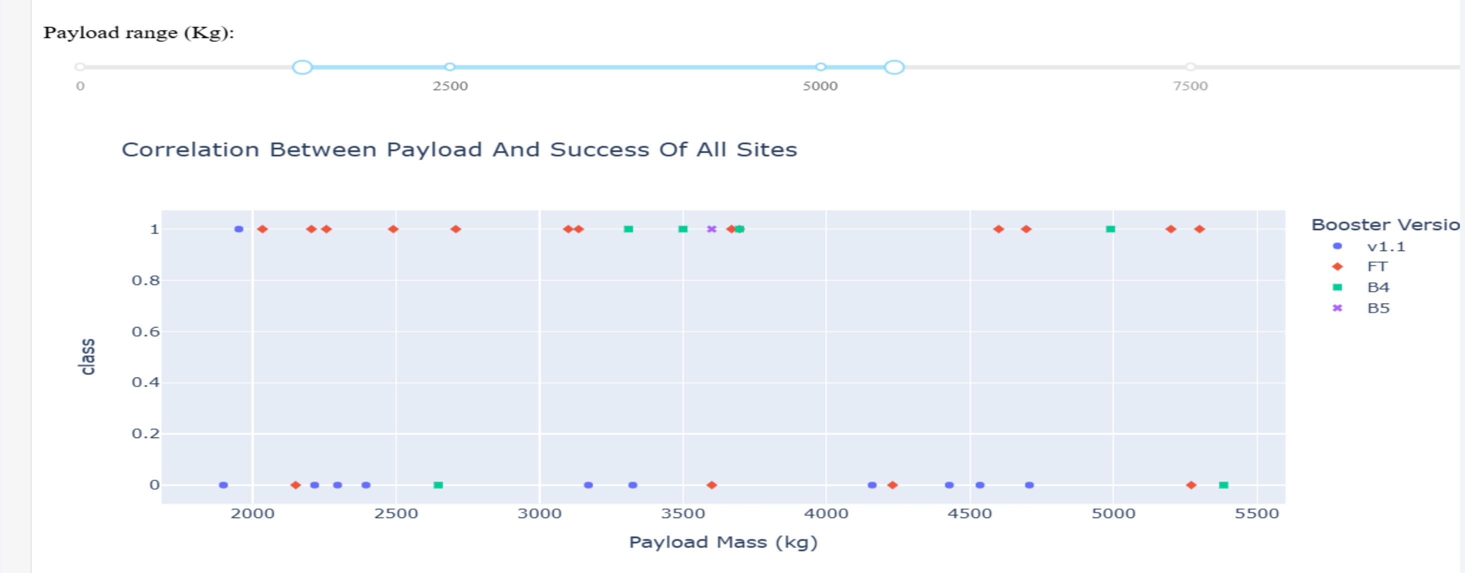
Launch Site With Highest Launch Success Ratio

- Kennedy Space Center Launch Complex 39A (KSC LC-39A) boasts an impressive launch success rate of 76.9%. Let's break down the numbers:

- Successful Landings: 10
- Failed Landings: 3



Payload Mass vs. Launch Outcome for all sites



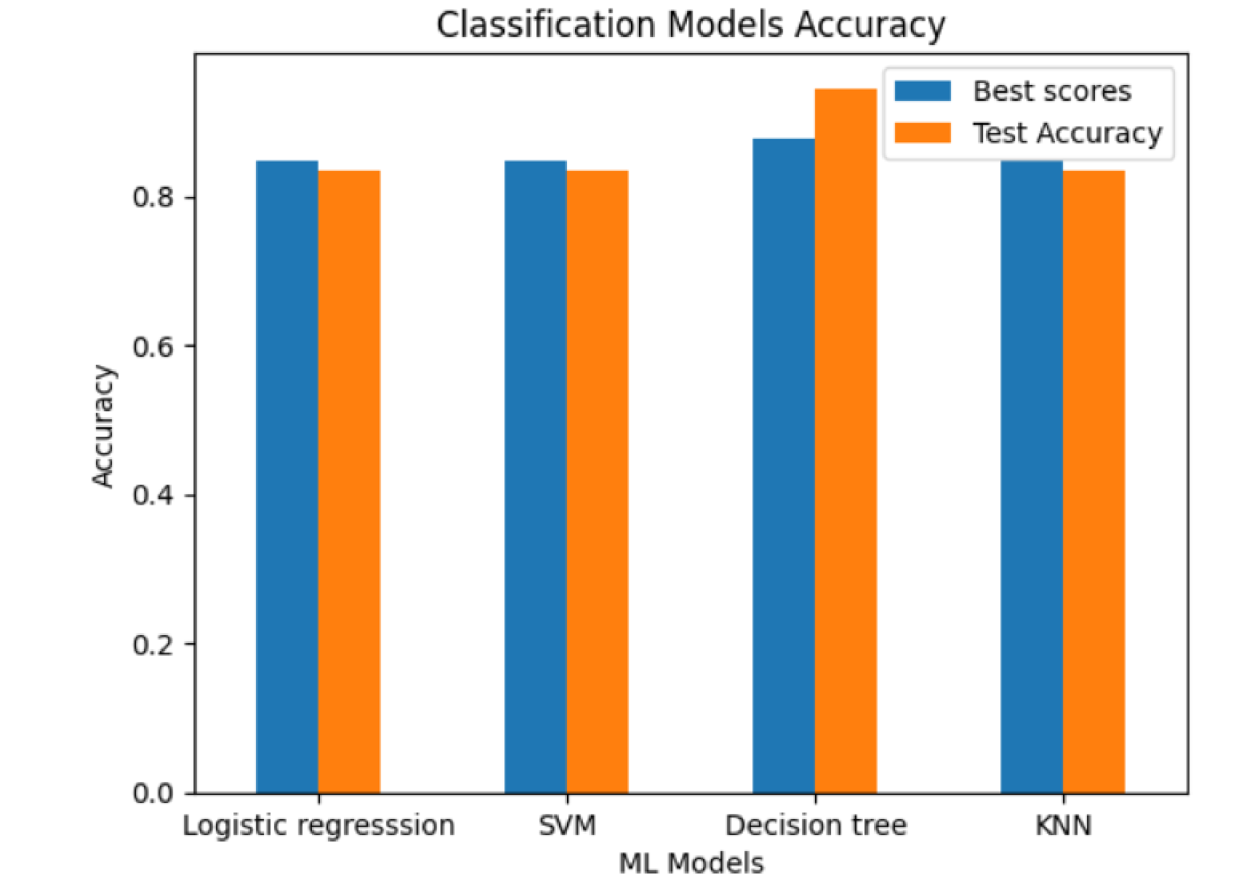
- Payloads under 6,000kg combined with Falcon 9 Full Thrust (FT) boosters indeed form a highly successful pairing for space missions
- Payloads between 2000 and 5500 kg have the highest success rate suggests a sweet spot in payload capacity for launch vehicles.

Section 5

Predictive Analysis (Classification)

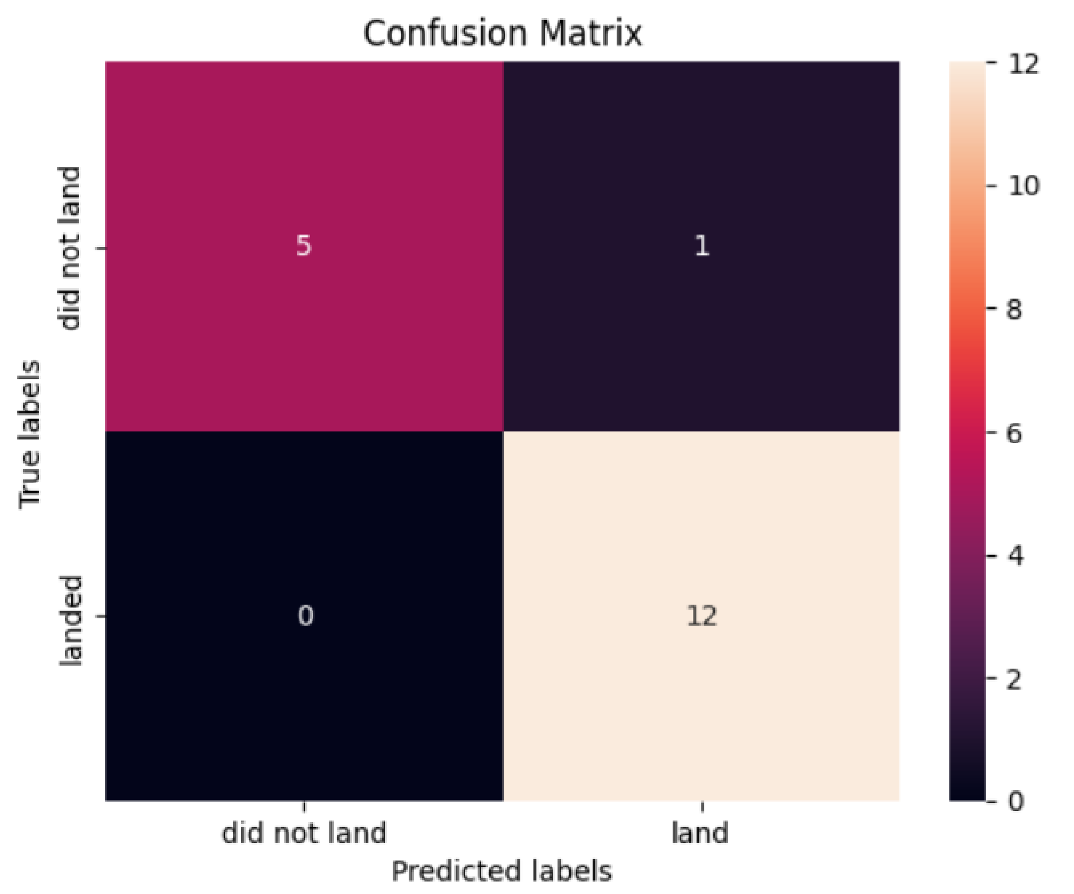
Classification Accuracy

- Four classification models were tested, and their accuracies are plotted beside.
- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 94%.



Confusion Matrix

- Show the confusion matrix of the Decision Tree Classifier
- This model has more true outcome than the fale one
- The sum of TN & Tp is 17
- Accuracy = (TP + TN) / (TP + TN + FP + FN)
- Accuracy = 17/18 = 94%



Conclusions

- Different data sources were analyzed, refining conclusions along the process
- The strategic placement of most launch sites near the Equator and their close proximity to the coast play pivotal roles in space exploration
- KSC LC-39A stands out as the pinnacle of launch success among all sites with a success rate of 77%, it surpasses other launch sites in terms of reliability and efficiency
- The success rate of space launches has indeed been on an upward trajectory.
- The orbits ES-L1, GEO, HEO, and SSO have achieved a remarkable 100% success rate in terms of landing outcomes
- Payload mass plays a crucial role in determining mission success.
- While the majority of missions have historically achieved success, the precision in landing outcomes has notably improved over time. This evolution is attributed to advancements in rocket technology, enhanced understanding of aerodynamics, and the development of sophisticated control systems.
- The Decision Tree Classifier serves as a powerful tool for predicting successful landings and optimizing profits in the aerospace industry.

Appendix

- Folium file can be viewed in nbviewer by pasting link
- "close.csv" contains Latitude and Longitude of close proximities like city, railway, highway.
- Dashboard application done in Jupyter notebook

Thank you!