IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

Abdul Ibrahim
12 September 2024

Abdul Ibrahim

# Outline

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

Abdul Ibrahim

# Executive Summary

**Summary of Methodologies:**

- Data was collected using two methods: the **SpaceX REST API** and **web scraping from Wikipedia**.
- Data wrangling was performed to clean and structure the data. Missing values were handled, and a landing outcome label was created for analysis.
- **Exploratory Data Analysis (EDA):** Visualizations and SQL queries were used to examine trends in the data
- **Interactive Visualization:** Folium and Plotly Dash were used to build interactive maps and dashboards.
- **Predictive Modeling:** Classification models (Logistic Regression, SVM, Decision Tree, KNN) were trained, tuned, and evaluated using GridSearchCV to predict landing success based on various features.

Abdul Ibrahim

# Executive Summary

**Summary of Results:**

- **EDA Results:** Showed relationships between flight number, payload, and orbit type with launch success.
- **Folium Maps:** Displayed interactive maps with launch site details and proximity to coastlines.
- **Plotly Dash:** Visualized launch outcomes using interactive charts.
- **Predictive Analysis:** All models achieved similar test accuracy (~83.33%), with Decision Tree selected as the best model due to the highest training accuracy

Abdul Ibrahim

# Introduction

Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch. This project aims to analyze SpaceX launch data to predict the likelihood of successful rocket landings.



Abdul Ibrahim

Section 1

# Methodology

# Methodology

## Executive Summary

- Data collection methodology:
  - The data was collected using the SPACEX REST API calls and web scraping from Wikipedia.
- Perform data wrangling
  - The missing values have been identified; landing outcome label is created.
- Perform exploratory data analysis (EDA) using visualization and SQL
  - We will conduct in-depth exploration of the dataset using various visualization techniques and SQL queries to uncover patterns, anomalies, and insights.

Abdul Ibrahim

# Methodology

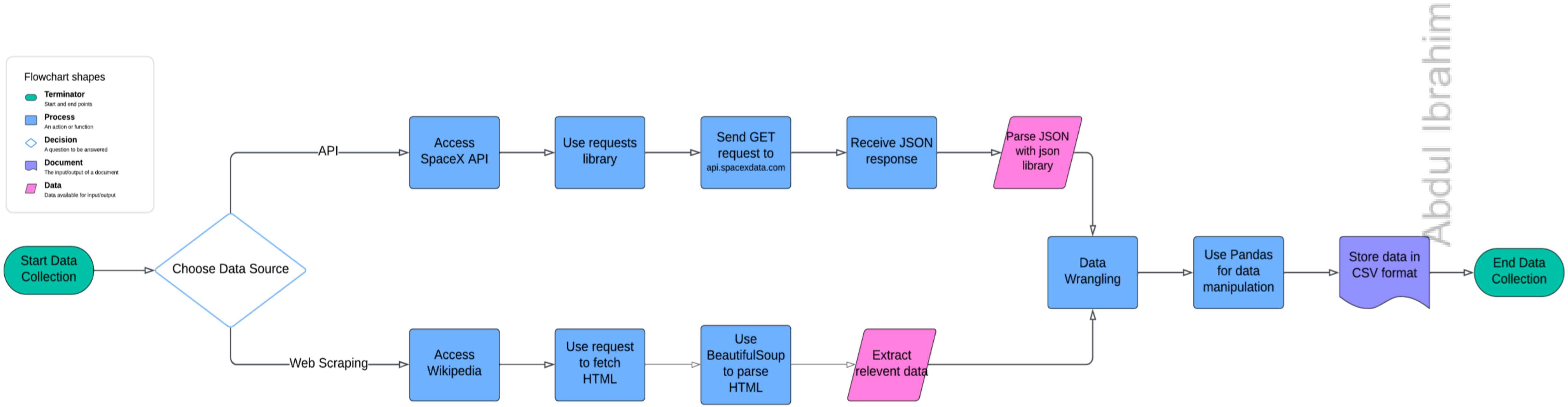
## Executive Summary

- Perform interactive visual analytics using Folium and Plotly Dash
  - Advanced interactive visualizations will be created using Folium for geospatial data and Plotly Dash for dynamic, web-based analytics dashboards.
- Perform predictive analysis using classification models
  - Model is build using training and testing data set, hyper parameters are tuned using gridsearchCV method and model is evaluated based on the accuracy.

Abdul Ibrahim

# Data Collection

- The data collection for this project was executed through two primary methodologies: programmatic access via the **SpaceX API** and **HTML web scrapping**.



# Data Collection – SpaceX API

1. Requesting Launch Data from SpaceX API
   API URL
   GET Request

2. Check Response Status
   Status Check

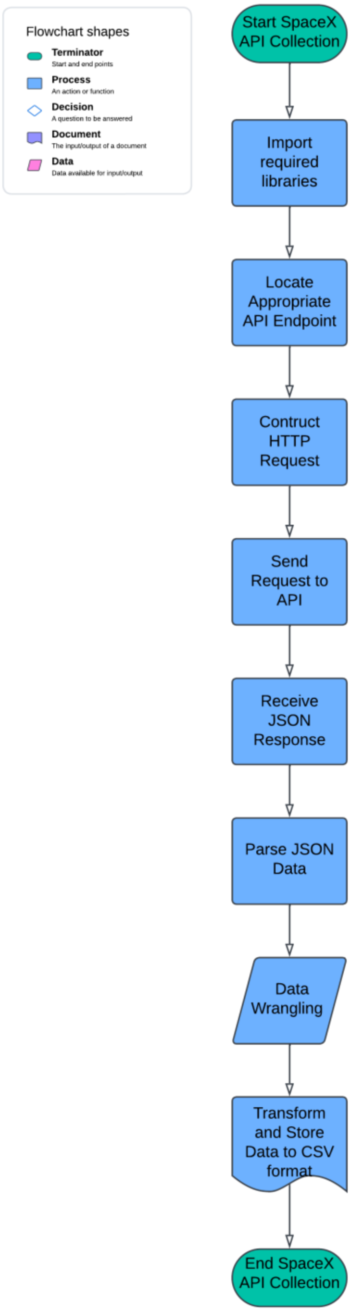
3. Parsing the Response into JSON
   Decode JSON Content

4. Normalize JSON Data into a Pandas DataFrame
   Flatten and Structure Data

5. Extracting Launch IDs and Related Data
   Extract Rocket, Payload, Launchpad, and Cores Data

[*GitHub URL of the completed SpaceX API calls notebook .*]()

(Click to follow link)




Abdul Ibrahim

# Data Collection - Scraping

1. Create HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response

2. Create BeautifulSoup object from the HTML response

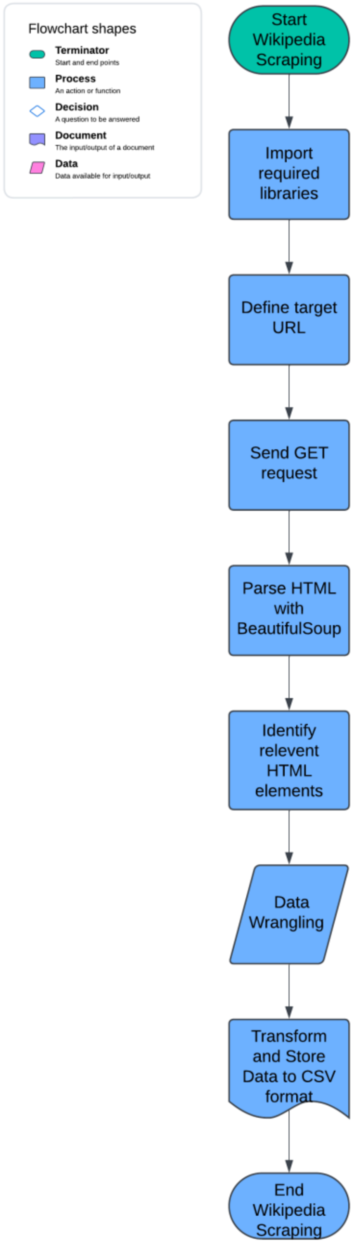
3. Create HTML Tables

4. Parse launch HTML Tables

5. Create a Data Frame

[*GitHub URL of the completed SpaceX Scraping calls notebook .*]

(Click to follow link)




Abdul Ibrahim

# Data Wrangling

1. Data Loading and Initial Exploration

- Loaded the dataset from a CSV file
- Checked data types of columns using df.types

2. Missing value analysis

- Calculated percentage of missing values in each column

3. Feature Engineering

- Created a binary classification label called ‘Class”
- 0=unsuccessful landing, 1=successful landing

4. Data Transformation

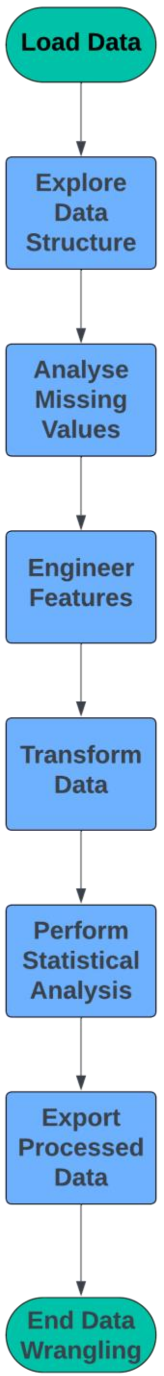
- Converted categorical ‘Outcome’ values to numeric ‘Class’ values

5. Statistical Analysis

- Calculated success rate of landing using mean of ‘Class’ column

*GitHub URL of the completed SpaceX  Data Wrangling calls notebook.*

(Click to follow link)




Abdul Ibrahim

# EDA with Data Visualization

Several charts were plotted as part of an exploratory data analysis

- PayloadMass vs FlightNumber **scatter plot** (with hue for Class)
  Purpose: To visualize how payload mass and flight number relate to landing success.

- FlightNumber vs LaunchSite **scatter plot** (with hue for Class)
  Purpose: To examine the relationship between flight number and launch sites, and how they relate to mission success.
  This helps identify any patters to certain launch sites and their payload masses.

- PayloadMass vs LaunchSite **scatter plot** (with hue for Class)
  Purpose: To observe relationship between launch sites and their payload masses

Abdul Ibrahim

# EDA with Data Visualization

- FlightNumber vs Orbit **scatter plot** (with hue for Class)
  Purpose: To examine relationship between flight number and orbit types

- Success Rate by Orbit Type **bar chart**
  Purpose: To visualize the success rate for each orbit type.
  This helps identify which orbits have higher success rates

- PayloadMass vs Orbit **scatter plot** (with hue for Class)
  Purpose: To reveal relationship between payload mass and orbit types

- Launch Success Yearly Trend **line chart**
  Purpose: To visualize the average launch success rate trend over years
  insight: Success rate increased from 2013 to 2020

*GitHub URL of the completed SpaceX Data Visualization calls notebook.*

(Click to follow link)

Abdul Ibrahim

# EDA with SQL

- Launch Site Analysis
  Identified unique launch sites and focused on sites beginning with ‘CCA’
  1. *SELECT with WHERE Launch_Site LIKE 'CCA%' and LIMIT 5*

- Payload Analysis
  Calculated total payload mass for NASA (CRS) missions and Determined average payload mass for F9 v1.1 boosters
  1. *SUM and WHERE clauses to filter for NASA CRS payloads*
  2. *AVG function with WHERE clause to filter for specific booster version*

- Landing Outcome
  Identified date of first successful ground pad landing and listed boosters with successful drone ship landings (payload 4000-6000 kg)
  1. *MIN function with WHERE clause to filter for successful ground landings*
  2. *SELECT DISTINCT with multiple WHERE conditions*

Abdul Ibrahim

# EDA with SQL

- Mission Success Metrics
  Quantified successful vs failed missions
  1. *SUM with CASE statements to count mission outcomes*

- Booster Performance
  Identified booster versions carrying maximum payload mass
  *1. Subquery to find max payload mass, then main query to get booster versions*

Abdul Ibrahim

# EDA with SQL

- Temporal Analysis
  Examined failed drone ship landings in 2015 and ranked landing outcomes from 2010-06-04 to 2017-03-20
  *1. Substr function to extract month and year from date, with WHERE conditions*
  *2. COUNT, GROUP BY, and ORDER BY to rank landing outcomes in a specific date range*

*GitHub URL of the completed SpaceX SQL calls notebook.*

(Click to follow link)

Abdul Ibrahim

# Build an Interactive Map with Folium

- **Folium Circle objects:** Created for each launch site
- **Folium Marker objects:** Created for each launch site and launch record
- **MarkerCluster object:** Added to group markers for launch records
- **MousePosition object:** Added to show coordinates on mouse hover
- **Folium Marker for coastline point:** Added to show distance from launch site
- **Folium Polyline:** Added to connect launch site to coastline point

*GitHub URL of the completed SpaceX Interactive Map With Folium calls notebook.*

(Click to follow link)

Abdul Ibrahim

# Build an Interactive Map with Folium

- **Circle and Marker objects for launch sites:** These were added to visually represent the locations of SpaceX launch sites on the map. The circles highlight the area around each site, while the markers provide precise location points
- **Marker for launch records:** These were added to show the outcomes of individual launches at each site. Green markers indicates successful launches, while red markers indicate failed launches. This allows for quick visual assessment of launch success rates at different sites
- **MarkerCluster:** This was used to group the many launch record markers, preventing visual clutter when zoomed out and allowing for easier exploration of dense areas of launches
- **MousePosition:** This was added to enable users to easily find coordinates of points of interest on the map, facilitating distance calculations and proximity analysis
- **Coastline marker and PolyLine:** These were added to demonstrate how to calculate and visualize the distance from a launch site to the nearest coastline. This helps in analyzing the proximity of launch sites to geographical features that might be relevant for launches

# Build a Dashboard with Plotly Dash

Summarize plots/graphs and interactions used:

- Dropdown menu for launch site selection
- Pie Chart displaying successful launches
- Range slider for payload mass selection
- Scatter plot showing payload mass vs launch success

*GitHub URL of the completed SpaceX Dashboard with Plotly Dash calls.*

(Click to follow link)

Abdul Ibrahim

# Build a Dashboard with Plotly Dash

- **Launch Site Dropdown:**
  Allows user to select specific launch sites or view data for all sites
  Enhances data exploration by enabling launches site-specific analysis

- **Success Rate Pie Chart:**
  Visualizes the proportion of successful launches for selected site(s)
  Provides a quick overview of launch success rates, aiding in site performance comparison

- **Payload Range Slider:**
  Enables users to filter data based on payload mass range
  Facilitates investigation of how payload mass impacts launch success

- **Payload vs Success Scatter Plot:**
  Illustrates the relationship between payload mass and launch success
  color-coded by booster version for additional insight
  Helps identify patterns or trends in launch success relative to payload and technology

Abdul Ibrahim

# Predictive Analysis (Classification)

**Key Phrases:**

1. **Data preparation**
   - Standardized the feature data
   - Split data into training and test sets (80% train, 20%test)

2. **Model Selection and Hyperparameter Tuning**
   - Implemented Logistic Regression, Support Vector Machine (VSM), Decision Tree and K-Neareast Neighbours (KNN) classification models
   - Used GridSearchCV with 10-fold cross-validation for each model
   - Defined hyperparameter grids for each model type

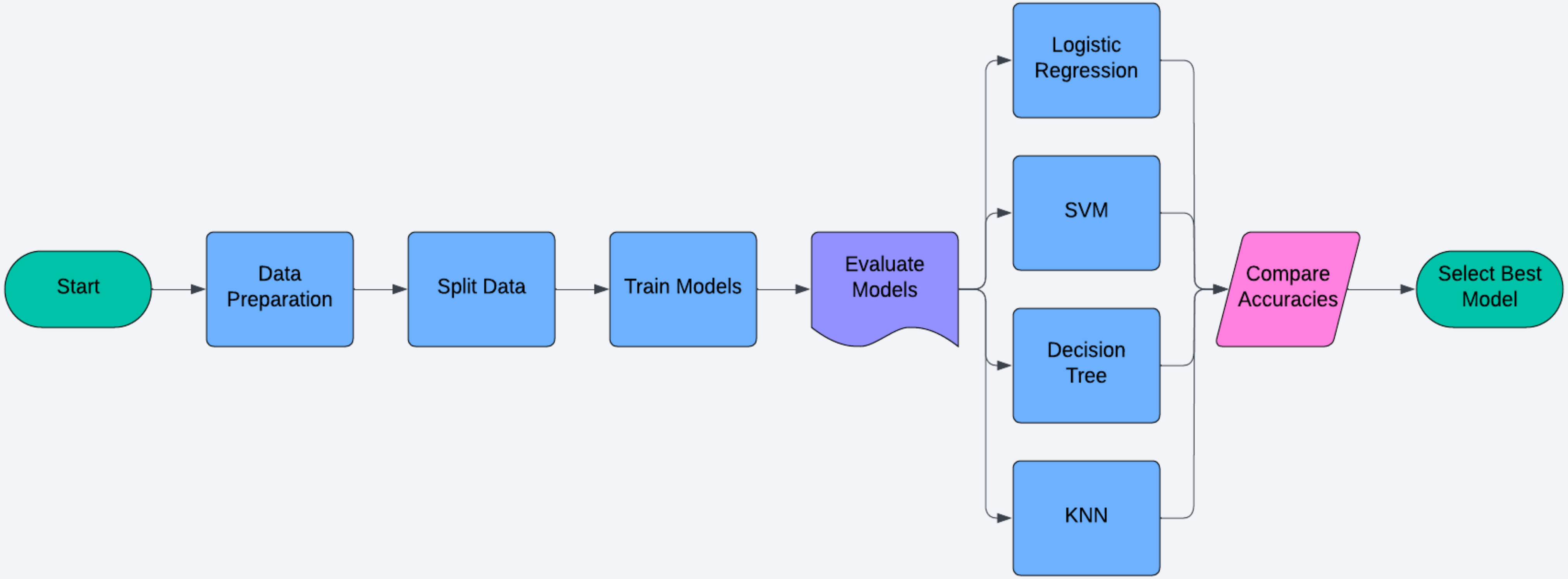
3. **Model Evaluation**
   - Assessed each model's performance using accuracy on the test set
   - Analyzed confusion matrices to understand error types

4. **Model Comparison**
   - Compared accuracies across all four models
   - Identified the best performing model based on test accuracy

Abdul Ibrahim

# Predictive Analysis (Classification)

## Flow Chart:




Abdul Ibrahim

# Predictive Analysis (Classification)

**Best Performing Classification Models:**

All models achieved the same accuracy of 0.8333 on the test set. However, Decision Tree was selected as the best-performing model based on its highest validation accuracy (87.5%)

*GitHub URL of the completed SpaceX Predictive Analysis calls.*

(Click to follow link)

Abdul Ibrahim

# Results

**Exploratory data analysis results:**

**Flight Number vs. Launch Site:**

- KSC LC-39A had the highest success rate, with more successful landings than other sites.
- CCAFS SLC-40 also showed a high success rate, particularly for heavier payloads.

**Payload Mass vs. Orbit Type:**

- LEO (Low Earth Orbit) had higher success rates for lower mass payloads.
- Heavier payloads in GTO (Geostationary Transfer Orbit) and Polar Orbit showed lower success rates.

**Success Rate vs. Orbit Type:**

- LEO launches had the highest overall success rates.
- Higher energy orbits, such as GTO, had lower success rates due to the complexity of the mission.

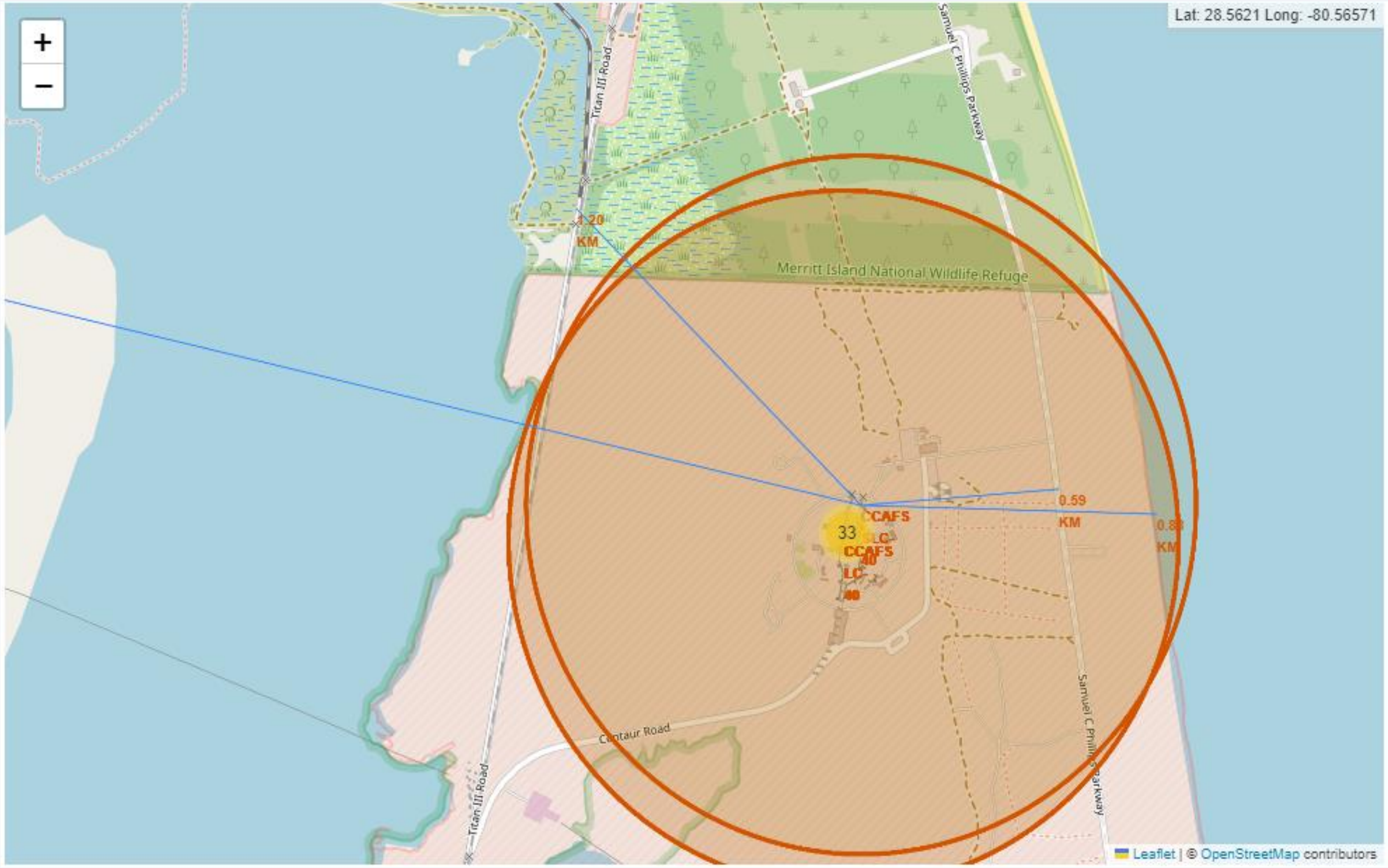
**Yearly Launch Success Trend:**

- The success rate improved significantly year-over-year, with more recent launches achieving near-perfect success.
- Early launches had more failures, but improvements in technology increased success rates in later years

Abdul Ibrahim

# Results

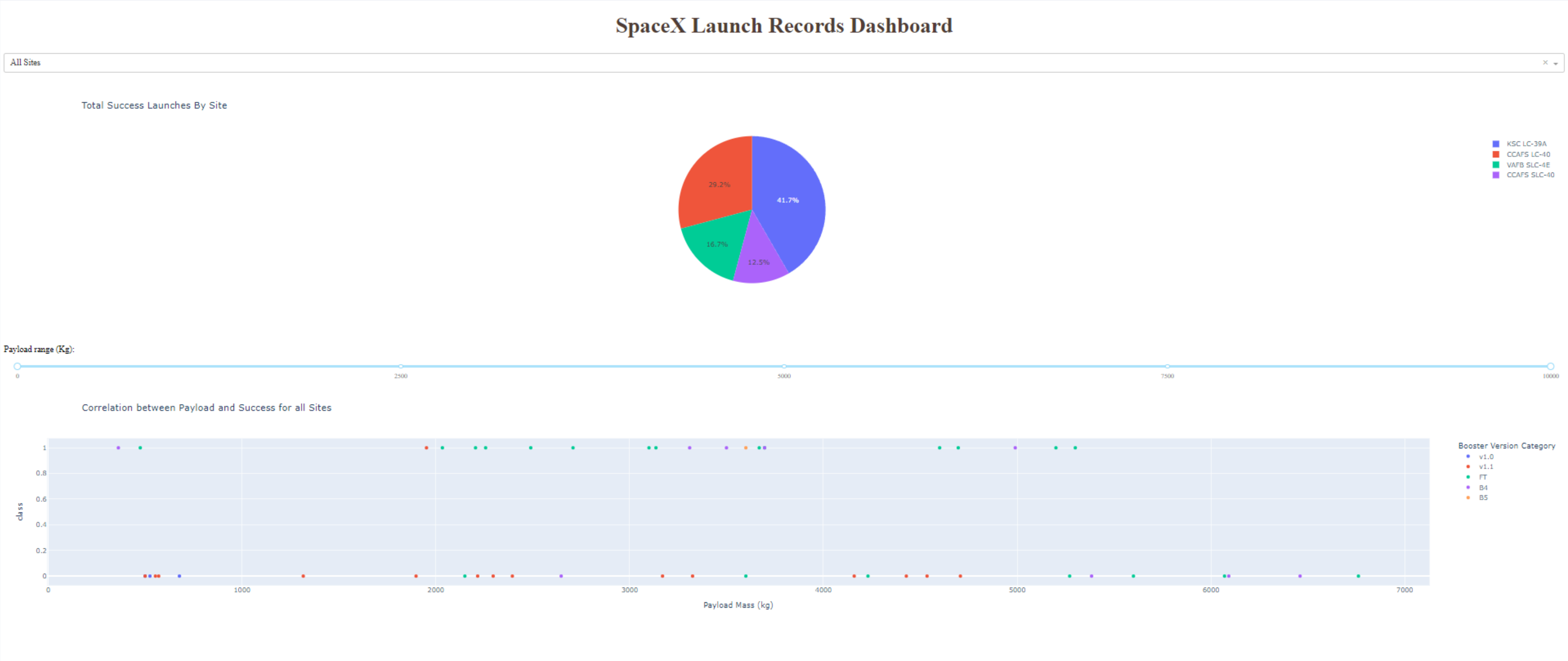
## Interactive analytics with folium:




Abdul Ibrahim

# Results

## Interactive analytics Dashboard:




Abdul Ibrahim

# Results

**Predictive analysis results:**

**1. Logistic Regression**

a) Best Parameters:

- C:0.01
- Penalty: 12
- Solver: ibfgs

b) Cross-Validation Accuracy: 0.8462

c) Test Accuracy: 0.8333

d) Confusion Matrix:

- True positive:12
- False positive:3
- False Negatives and True Negatives: Not explicit stated, but can be inferred to be 3 total

Abdul Ibrahim

# Results

**Predictive analysis results:**

**2. Support Vector Machine (SVM)**

a) Best Parameters:

- C:1.0
- Gamma: 0.03162
- Kernal: sigmoid

b) Cross-Validation Accuracy: 0.8482

c) Test Accuracy: 0.8333

d) Confusion Matrix: Not explicitly stated, but accuracy matches Logistic Regression

Abdul Ibrahim

# Results

**Predictive analysis results:**

**3. Decision Tree**

a) Best Parameters:

- Criterion: entropy
- Max_depth: 18
- Max_features: sqrt
- Min_sample_leaf: 4
- Min_sample_split: 2
- Splitter: random

b) Cross-Validation Accuracy: 0.8607

c) Test Accuracy: 0.8333

d) Confusion Matrix: Not explicitly provided, but accuracy matches other models

Abdul Ibrahim

# Results

**Predictive analysis results:**

**4. K-Neareast Neighbors (KNN)**
a) Best Parameters:
 - Algorithm: auto
 - N_neighbors: 10
 - P: 1
b) Cross-Validation Accuracy: 0.8482
c) Test Accuracy: 0.8333
d) Confusion Matrix: Not explicitly provided, but accuracy matches other models

Abdul Ibrahim

# Results

**Predictive analysis results:**

**Overall Comparison**

1. All four models achieved the same test accuracy of 0.8333 (approximately 83.33%).
2. The Decision Tree model had the highest cross-validation accuracy (86.07%), followed by SVM and KNN (both 84.82%), and then Logistic Regression (84.64%).
3. Despite having the lowest cross-validation accuracy, Decision tree was highlighted as the best performing method, due to its validation accuracy.

Abdul Ibrahim

Section 2

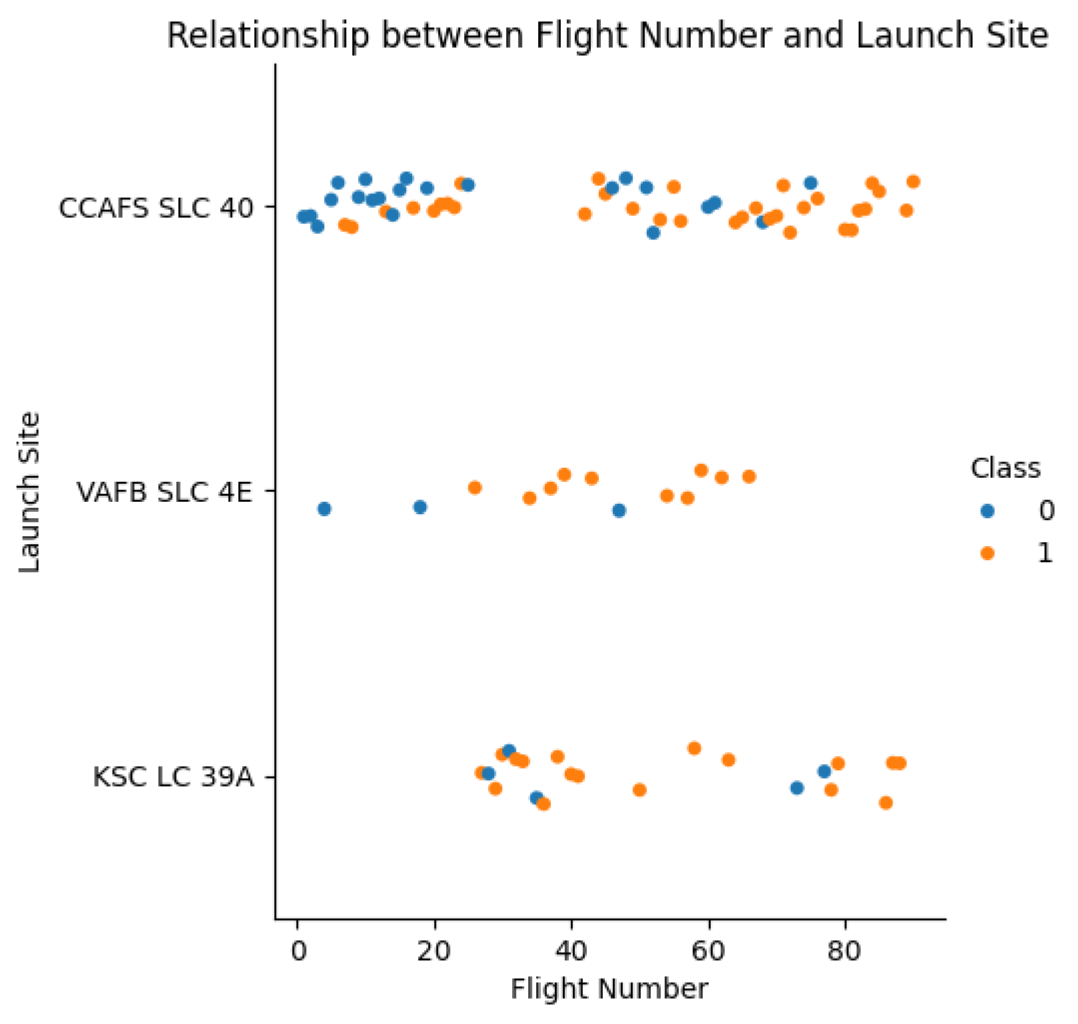
# Insights drawn from EDA

# Flight Number vs. Launch Site

1. Different launch sites have varying numbers of launches

2. Some sites seem to have more successful launches in later flight numbers

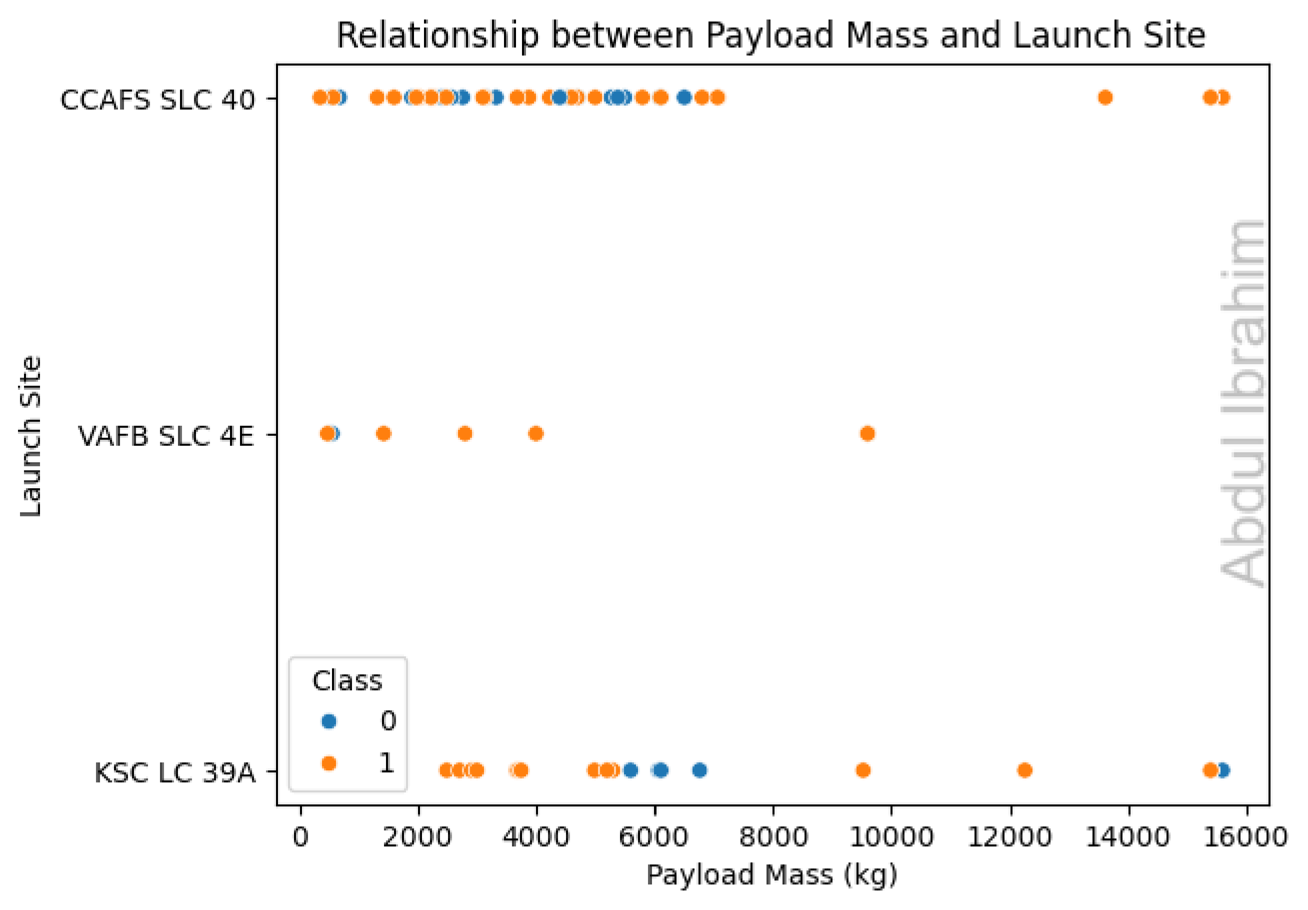
3. This could indicate site-specific improvements or learning curves




Abdul Ibrahim

# Payload vs. Launch Site

1. Different launch sites handle varying ranges of payload masses
2. VAFB-SLC launch site doesn’t have launches for payloads over 10,000 kg
3. Some sites seem to have more successful launches with certain payload ranges

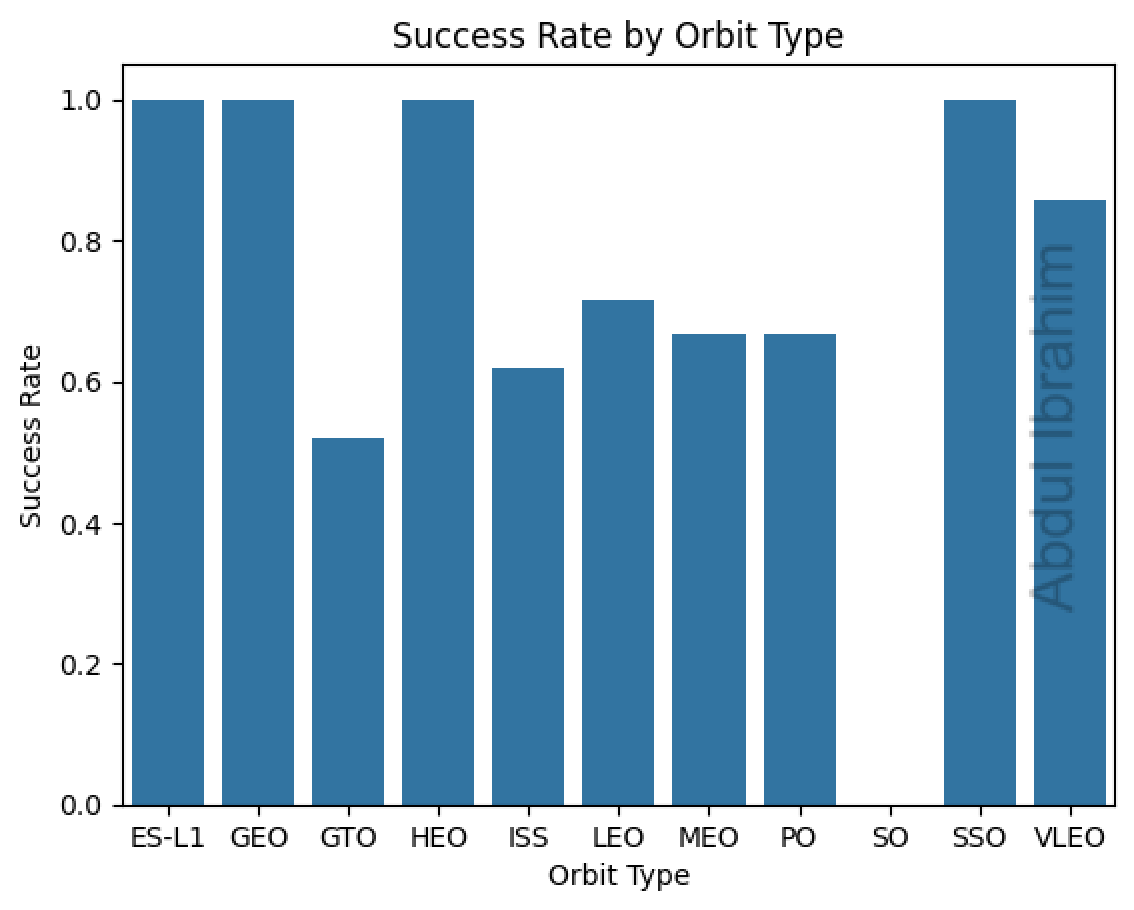


# Success Rate vs. Orbit Type

1. Some orbits (e.g. SSO, HEO) have higher success rates than others

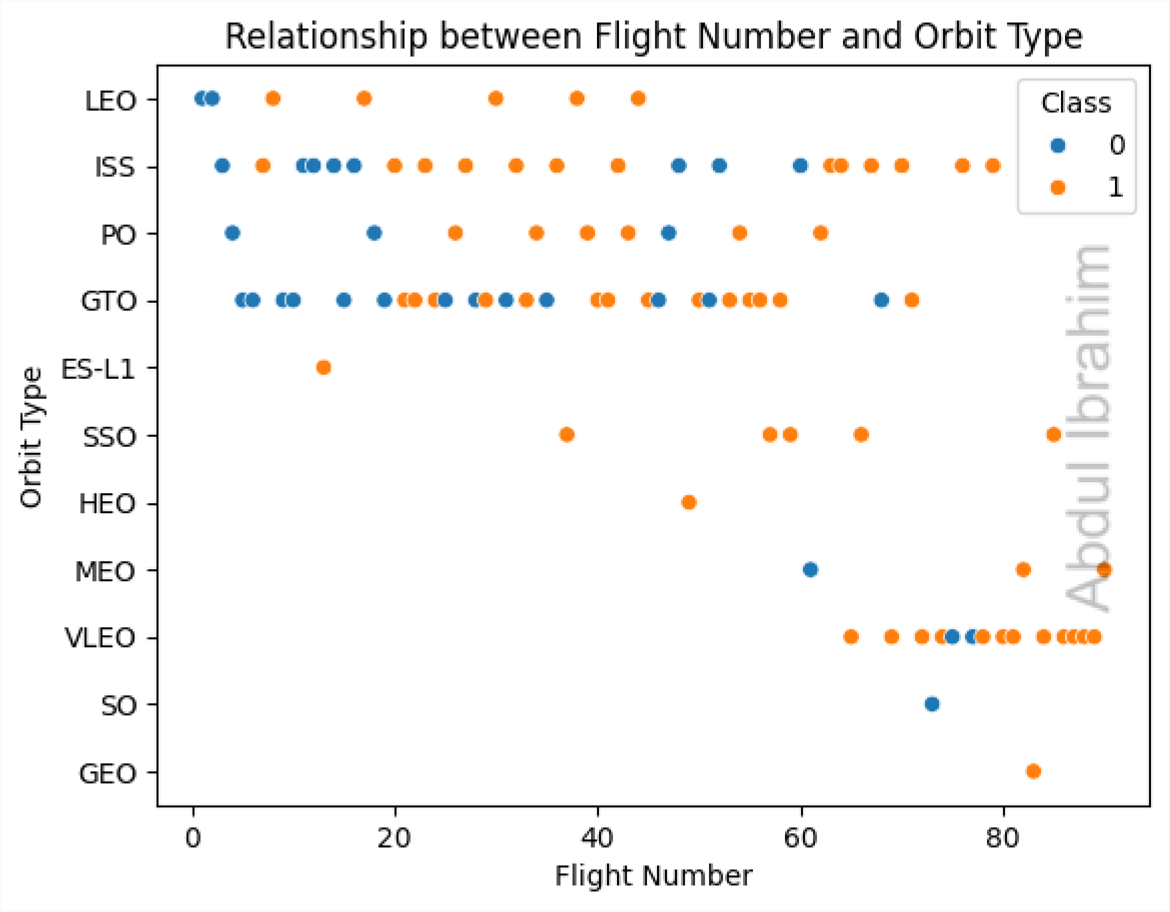
2. GTO seems to have a lower success rate

3. This could inform mission planning and risk assessment for different orbit types



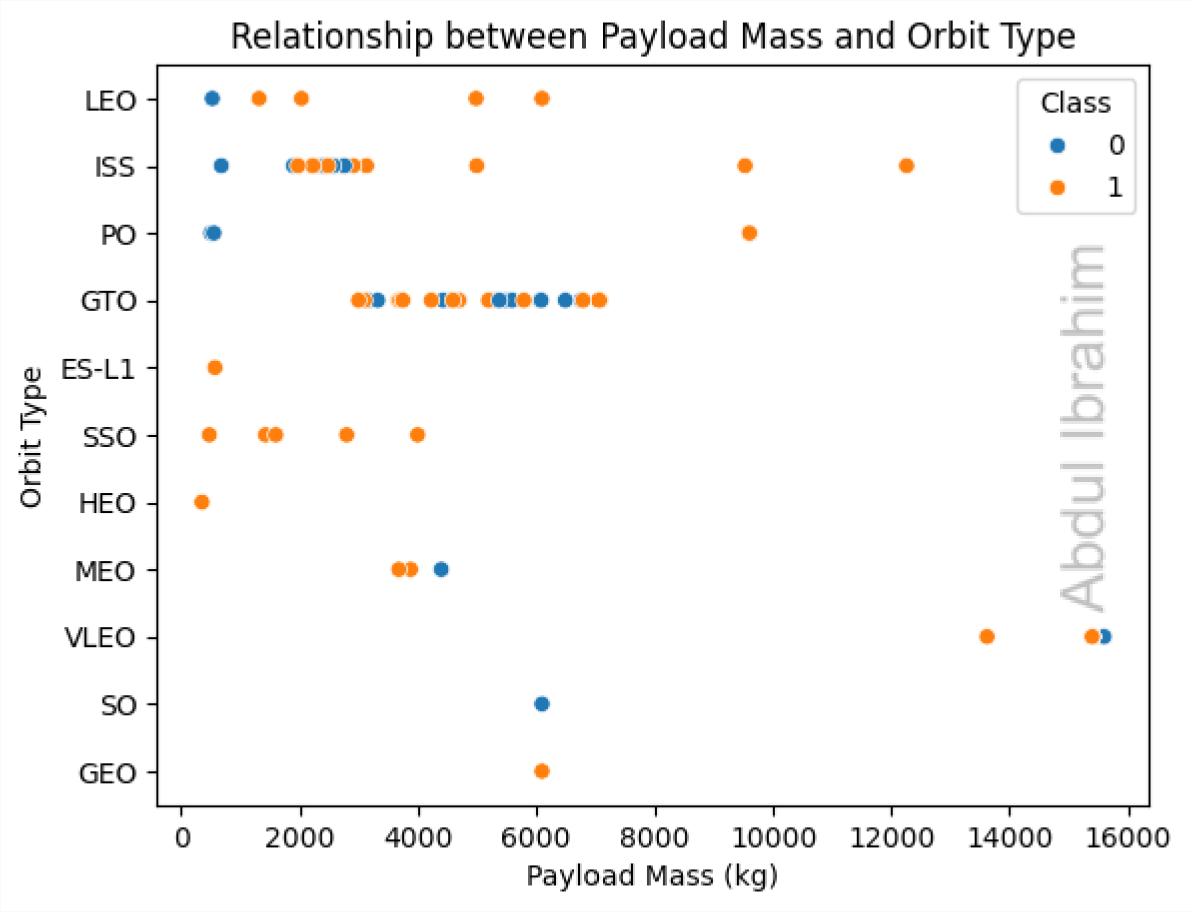
# Flight Number vs. Orbit Type

1. Certain orbits (e.g. LEO) show improved success rates in later flights
2. Other orbits (e.g. GTO) don't show a clear relationship between flight number and success
3. This suggest that some orbit types benefited more from experiences and technological improvements.



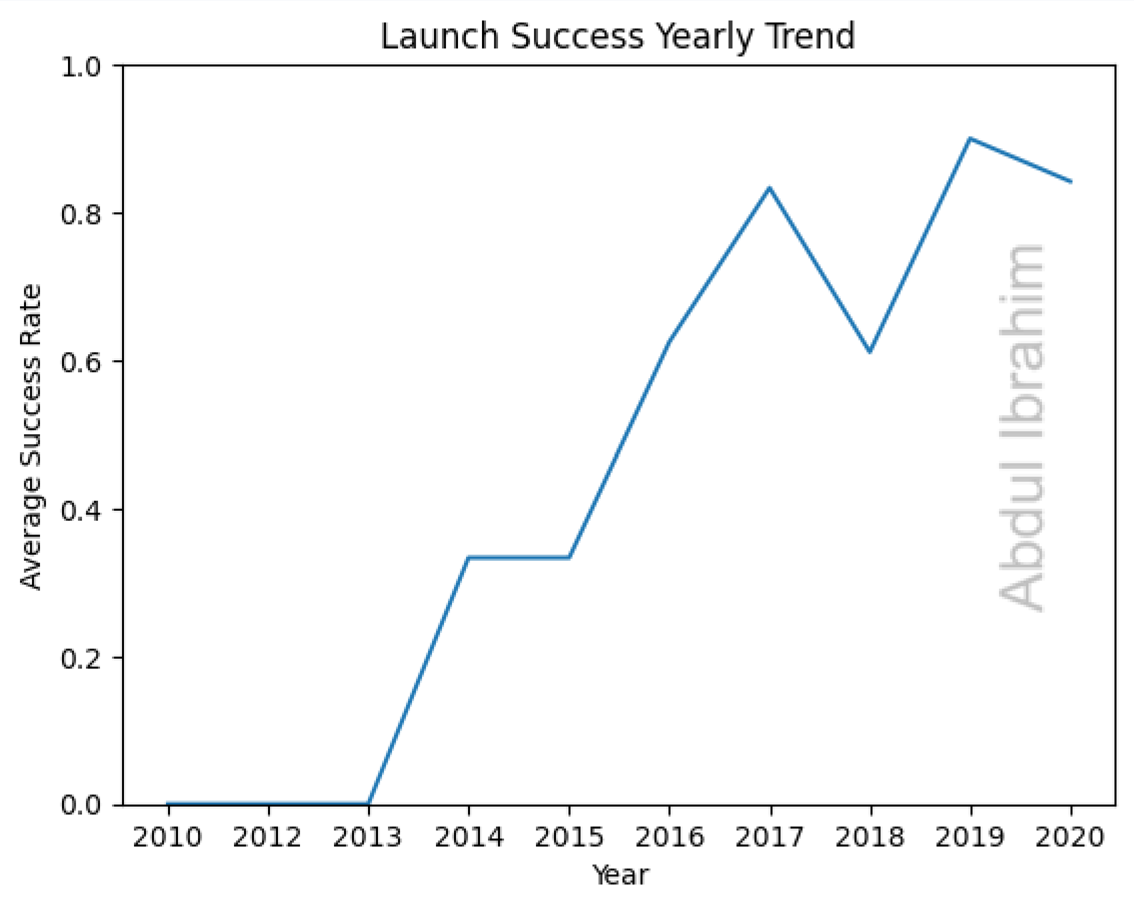
# Payload vs. Orbit Type

1. Different orbit types have varying ranges of payload mass
2. For Polar, LEO and ISS orbits, there's a higher success rate with heavier payloads
3. GTO orbits show mixed results across payload masses



# Launch Success Yearly Trend

1. There's a clear upward trend in success rates from 2013 to 2020
2. The success rate increased significantly, reaching near 100% by 2020
3. This demonstrates SpaceX's improving technology and experience over time



# All Launch Site Names

Unique launch sites as follows:

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

The sql query returns the unique launch site names. The keyword '**DISTINCT**' is used to get the unique values in that column.

```
%%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 |

Abdul Ibrahim

# Launch Site Names Begin with '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 |
|---|---|---|---|---|---|---|---|---|---|
| 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 |

Abdul Ibrahim

- The query returned 5 records, all from the CCAFS LC-40 launch sites.
- This shows that many of SpaceX's early launches (2010-2013) were conducted from CCA Launch Complex 40. These launches included test flights and early NASA Commercial Resupply Services (CRS) missions.

# Total Payload Mass

- Total payload carried by boosters from NASA is 60,268 kg
- This significant payload mass demonstrates the substantial amount of cargo SpaceX has transported for NASA's Commercial Resupply Services program to the International Space Station.

```
%%sql

SELECT SUM(PAYLOAD_MASS__KG_) AS Total_Payload_Mass
FROM SPACEXTBL
WHERE Payload LIKE '%CRS%' AND Customer LIKE 'NASA%';
```

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

| Total_Payload_Mass |
|---|
| 60268 |

Abdul Ibrahim

# Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 is 2,928.4 kg
- This shows that the Falcon 9 v1.1 booster version had an average payload capacity of nearly 3 metric tons, which was a considerable improvement over earlier versions.

Abdul Ibrahim

```
%%sql

SELECT AVG(PAYLOAD_MASS__KG_) AS Average_Payload_Mass
FROM SPACEXTBL
WHERE Booster_Version = 'F9 v1.1';
```

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

| Average_Payload_Mass |
|---|
| 2928.4 |

# First Successful Ground Landing Date

- Dates of the first successful landing outcome on ground pad is on 2015-12-22
- This date marks a significant milestone for SpaceX, as it represents their first successful ground landing of a Falcon 9 first stage, a key achievement in their reusability program.

```
%%sql

SELECT DISTINCT Landing_Outcome
FROM SPACEXTBL
```

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

| Landing_Outcome |
|---|
| Failure (parachute) |
| No attempt |
| Uncontrolled (ocean) |
| Controlled (ocean) |
| Failure (drone ship) |
| Precluded (drone ship) |
| Success (ground pad) |
| Success (drone ship) |
| Success |
| Failure |
| No attempt |

```
%%sql

SELECT MIN(Date) AS First_Successful_Ground_Landing
FROM SPACEXTBL
WHERE Landing_Outcome = 'Success (ground pad)';
```

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

| First_Successful_Ground_Landing |
|---|
| 2015-12-22 |

Abdul Ibrahim

# Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
  a) F9 FT B1022
  b) F9 FT B1026
  c) F9 FT B1021.2
  d) F9 FT B1031.2
- These four boosters successfully landed on drone ships while carrying payloads between 4-6 metric tons, showcasing SpaceX's ability to recover boosters after launching medium-heavy payloads.

```
%%sql

SELECT DISTINCT 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 |

Abdul Ibrahim

# Total Number of Successful and Failure Mission Outcomes

- 98 successful missions, 0 failed missions
- This result indicates an impressive 100% success rate for the missions in the dataset, highlighting SpaceX's reliability.

```
%%sql

SELECT
    SUM(CASE WHEN Mission_Outcome = 'Success' THEN 1 ELSE 0 END) AS successful_missions,
    SUM(CASE WHEN Mission_Outcome = 'Failure' THEN 1 ELSE 0 END) AS failed_missions
FROM SPACEXTBL;
```

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

| successful_missions | failed_missions |
|---|---|
| 98 | 0 |

Abdul Ibrahim

# Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass:
  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, and F9 B5 B1049.7
- The Falcon 9 Block 5 version appears to be capable of carrying the heaviest payloads, which aligns with it being the final and most advanced iteration of the Falcon 9.

```
%%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 |

Abdul Ibrahim

# 2015 Launch Records

- Two failures in January and April 2015, both using F9 v1.1 boosters from CCAFS LC-40
- This shows that SpaceX experienced some challenges with drone ship landings in early 2015, which was still early in their attempts to land boosters at sea.

```
%%sql

SELECT
    substr(Date, 6, 2) AS month,
    Landing_Outcome,
    Booster_Version,
    Launch_Site
FROM SPACEXTBL
WHERE substr(Date, 1, 4) = '2015'
  AND Landing_Outcome LIKE 'Failure (drone ship)%';
```

```
 * 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 |

Abdul Ibrahim

# 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
  -Failure(drone ship): 5
  -Success(ground pad): 3

- This indicates that in the early years of SpaceX's landing attempts, they had more failures on drone ships than successes on ground pads. This is expected, as landing at sea is generally more challenging, and these represent some of the first attempts at booster recovery.

```
%%sql

SELECT
    Landing_Outcome,
    COUNT(*) AS count
FROM SPACEXTBL
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
  AND (Landing_Outcome LIKE 'Failure (drone ship)%' OR Landing_Outcome = 'Success (ground pad)')
GROUP BY Landing_Outcome
ORDER BY count DESC;
```

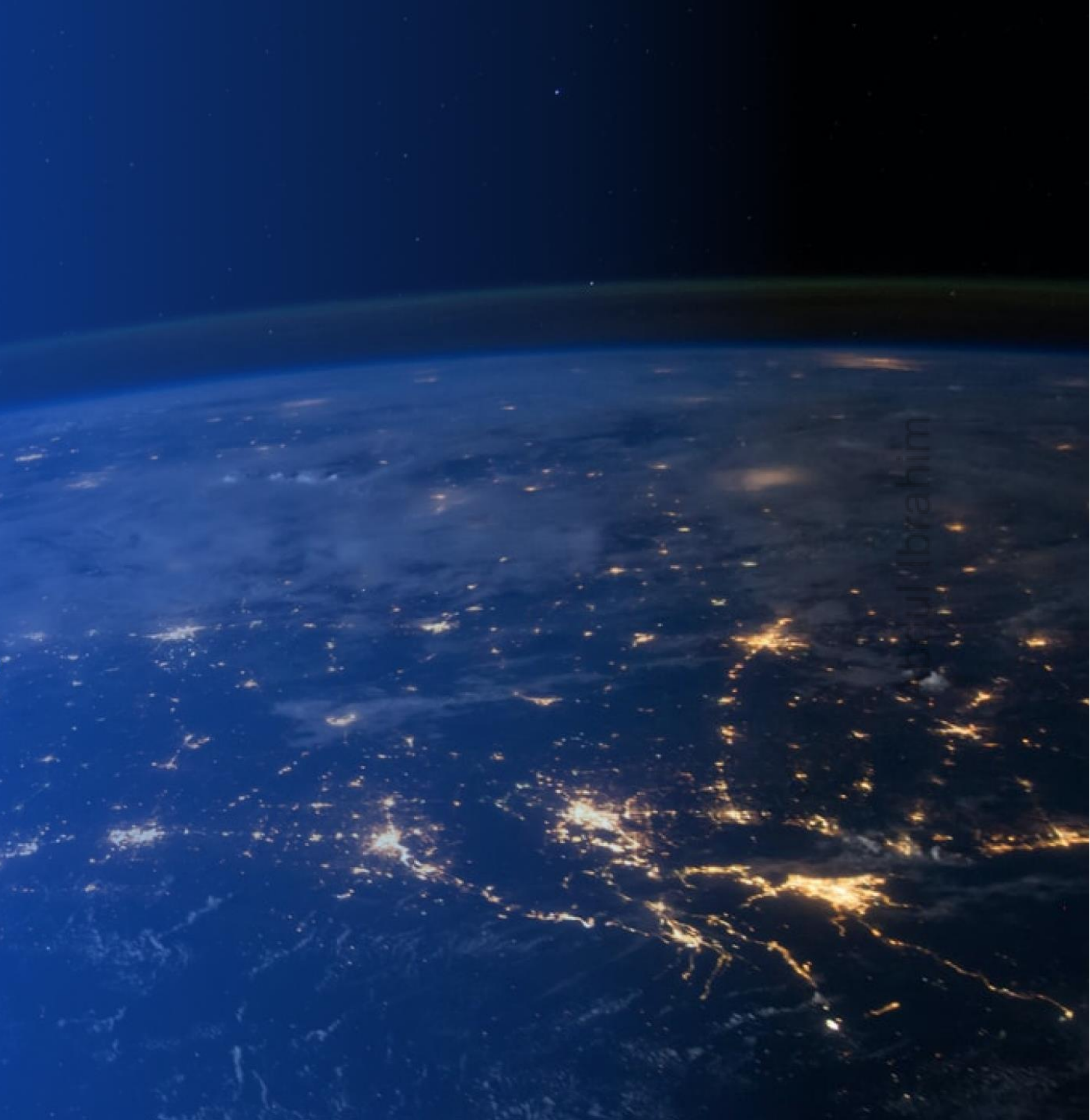
```
 * sqlite:///my_data1.db
Done.
```

| Landing_Outcome | count |
|---|---|
| Failure (drone ship) | 5 |
| Success (ground pad) | 3 |

Abdul Ibrahim

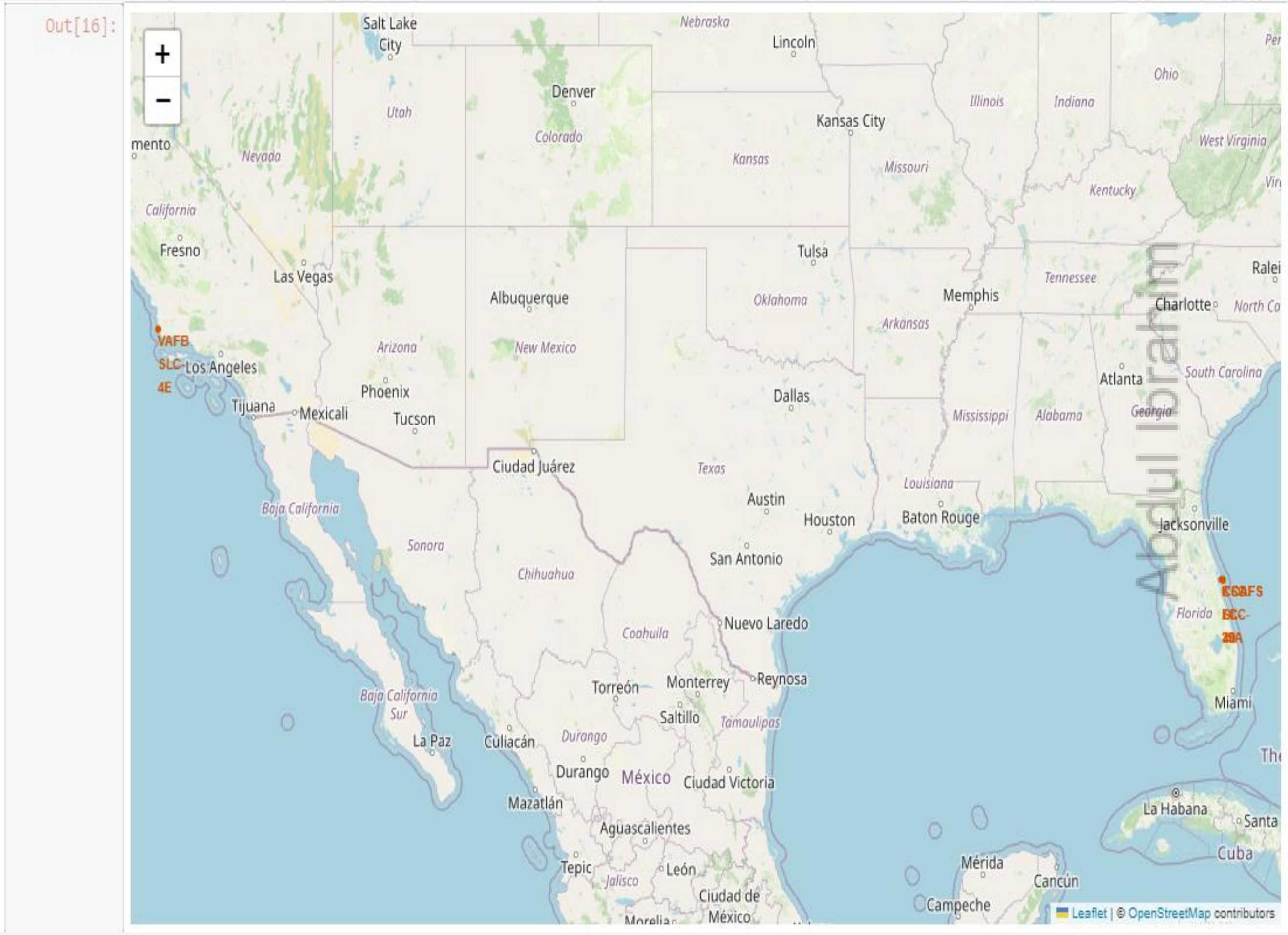
Section 3

# Launch Sites Proximities Analysis



# SpaceX Launch Sites Overview Map

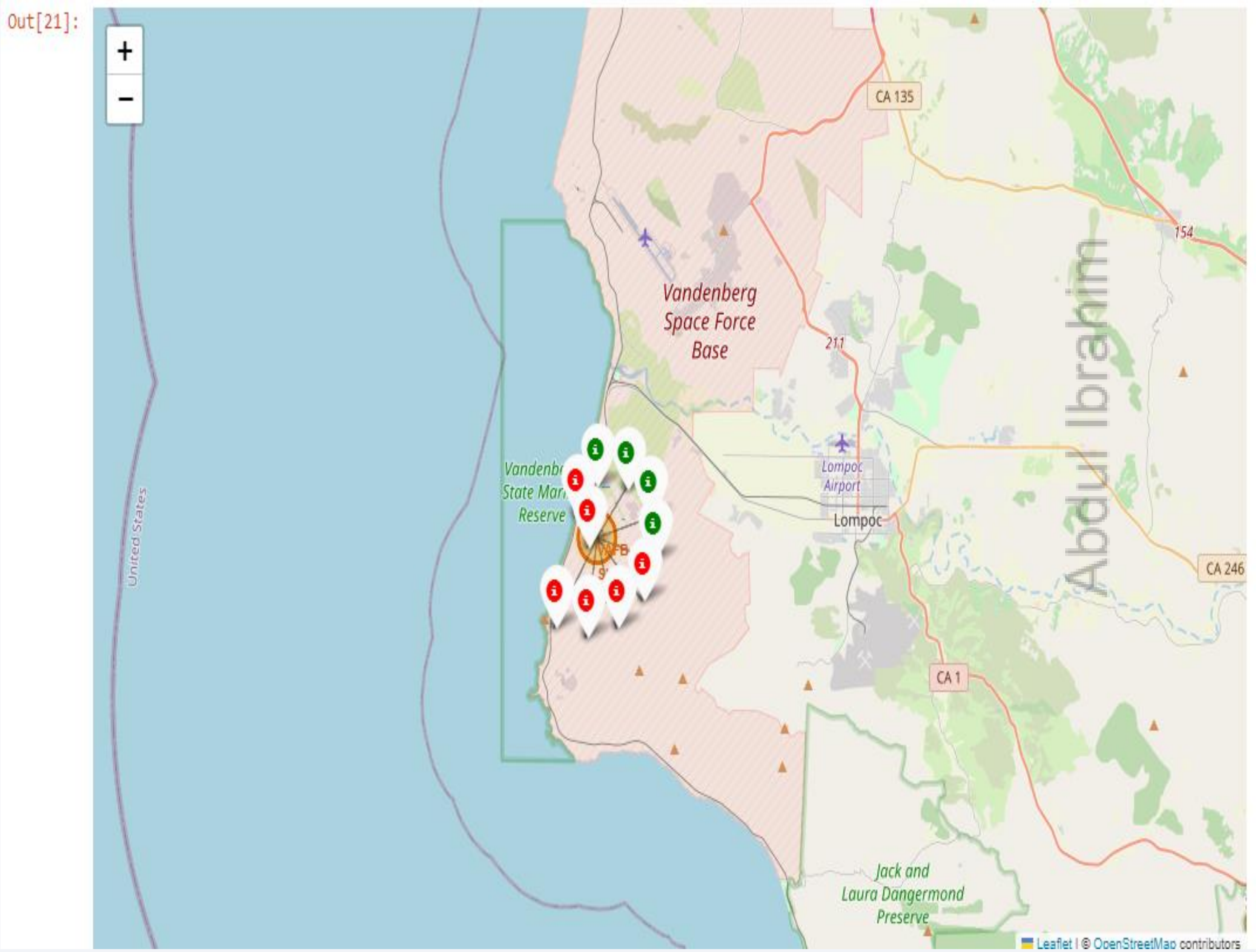
- This Map shows the location of all SpaceX launch sites marked with circles and labels
- It provides a broad overview of the geographical distribution of launch sites
- **Key Findings:**
  a) All launch sites appear to be in close proximity to the coast, suggesting a preference for coastal locations
  b)The launch sites are distributed across different parts of the United States, likely for strategic reasons



# SpaceX Launch Success and Failure Map

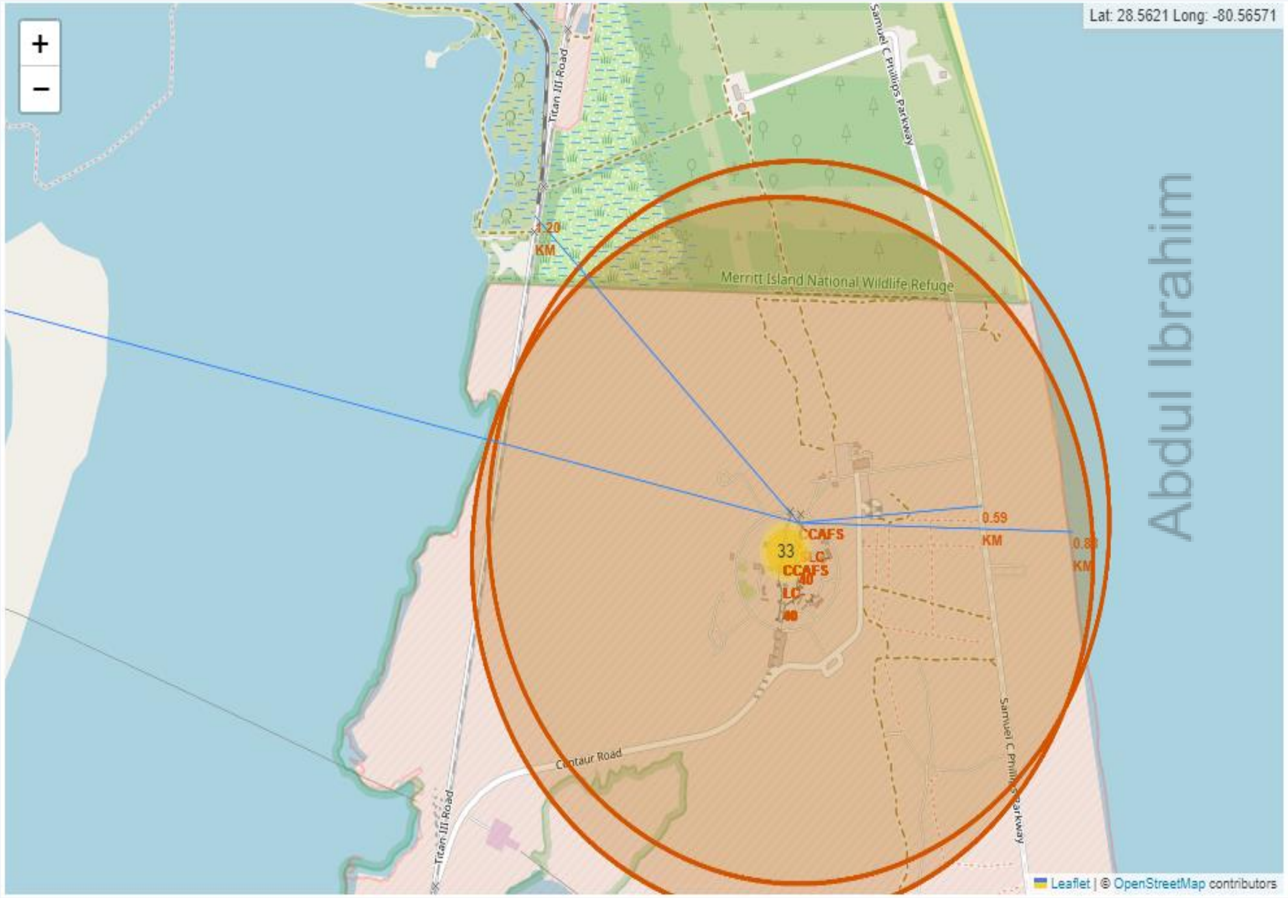
- This map uses a **MarkerCluster** to show all launch attempts, with green markers for successful launches and red markers for failed launches

- **Key Findings:**
  a)The clustering of markers allows for easy identification of high activity launch sites
  b)This map helps identify which launch sites have relatively higher success rates based on the predominance of green markers



# Launch Site Proximity Analysis Map

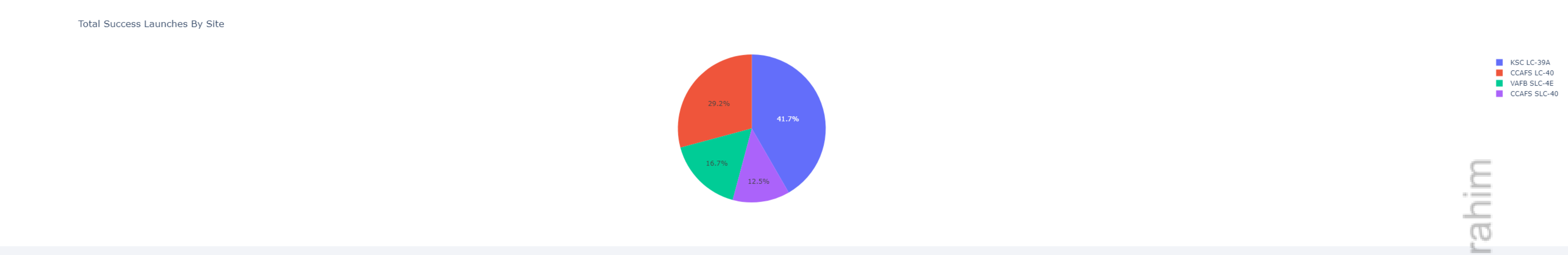
- This map focuses on analyzing the proximity of a specific launch site (likely Kennedy Space Center) to nearby features such as coastline, railways, highways and cities
- A distance marker showing the calculated distance to the nearest coastline point
- **Key findings:**
  a) This proximity to the coast is likely strategic for launch trajectories and safety considerations



Section 4

# Build a Dashboard with Plotly Dash

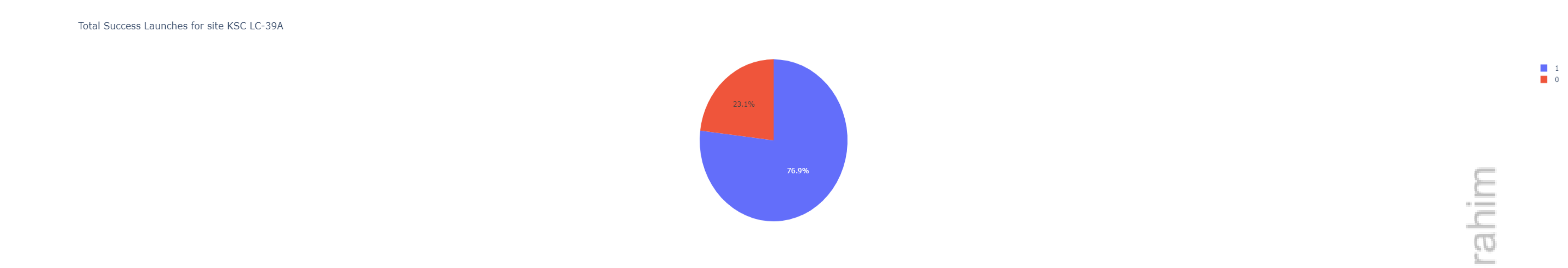
# Total Successful Launches by Site




- The pie chart shows the distribution of successful launches across different launch sites
- KSC LC-39A has the highest percentage of successful launches at 41.7%
- CCAFS SLC-40 follows closely with 25.7% of successful launches
- VAFB SLC-4E accounts for 10.7% of successful launches
- CCAFS LC-40 has the lowest percentage at 13.5%

Abdul Ibrahim

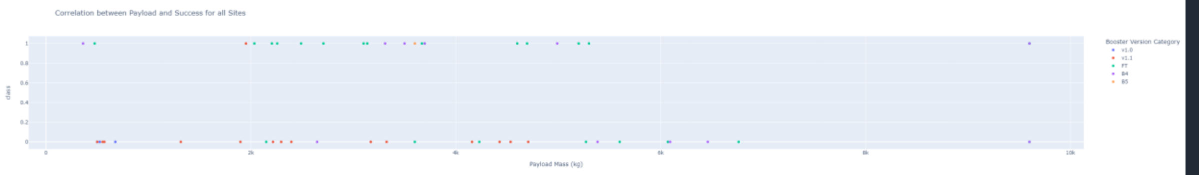
# Success Rate for KSC LC-39A Launch Site

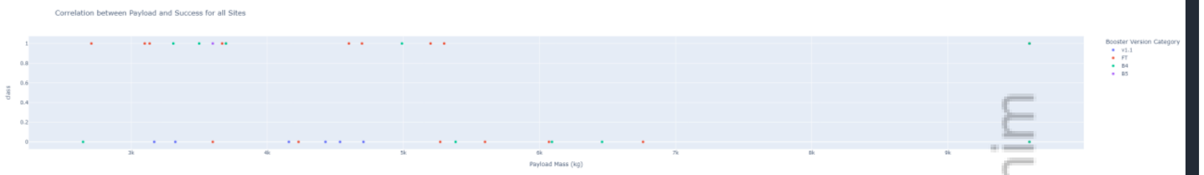


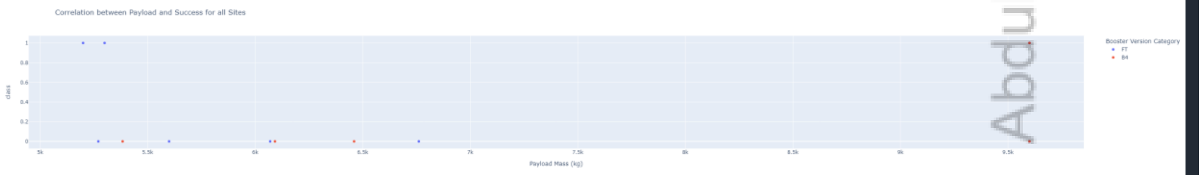

Abdul Ibrahim

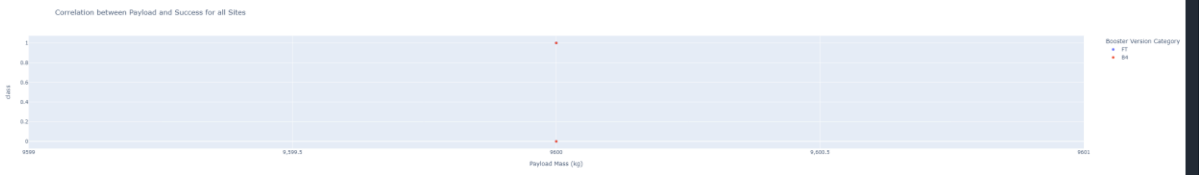
- The pie chart shows the success rate specifically for the KSC LC-39A launch site
- There are two categories: successful launches (1) and failed launches (0)
- The success rate for KSC LC-39A is impressively high at 76.9% and Only 23.1% of launches from this site have failed
- This chart highlights that KSC LC-39A is the most reliable launch site for SpaceX, with over three-quarters of its launches being successful.

# Correlation between Payload Mass and Launch Success




Abdul Ibrahim

# Correlation between Payload Mass and Launch Success

- The scatter plot show the relationship between payload mass (x-axis) and launch outcome (y-axis, where 1 indicates success and 0 indicates failure)
- Different booster version are represented with different colors and the payload range can be adjusted using the slider at the top
- **Key findings:**
  a) No clear correlation between payload mass and launch success. Successful launches occur across various payload ranges
  b) Most launches cluster in the 2,000 kg to 8,000 kg range and Booster versions FT and B5 appear most frequently and seem to have high success rates
  c) Heavier payloads (>9,000 kg) show a slightly lower success rate, but there are fewer data points in this range.
  d) Mid-range payloads (2,000-8,000 kg) have the highest concentration of successful launches, possibly due to more frequent missions and optimized systems for these masses.
- In conclusion, while payload mass doesn't strongly predict launch success, certain booster versions (particularly FT and B5) appear more reliable. The mid-range payloads have the highest number of successful launches, likely due to more missions in this range and well-tuned systems for these masses.

Abdul Ibrahim

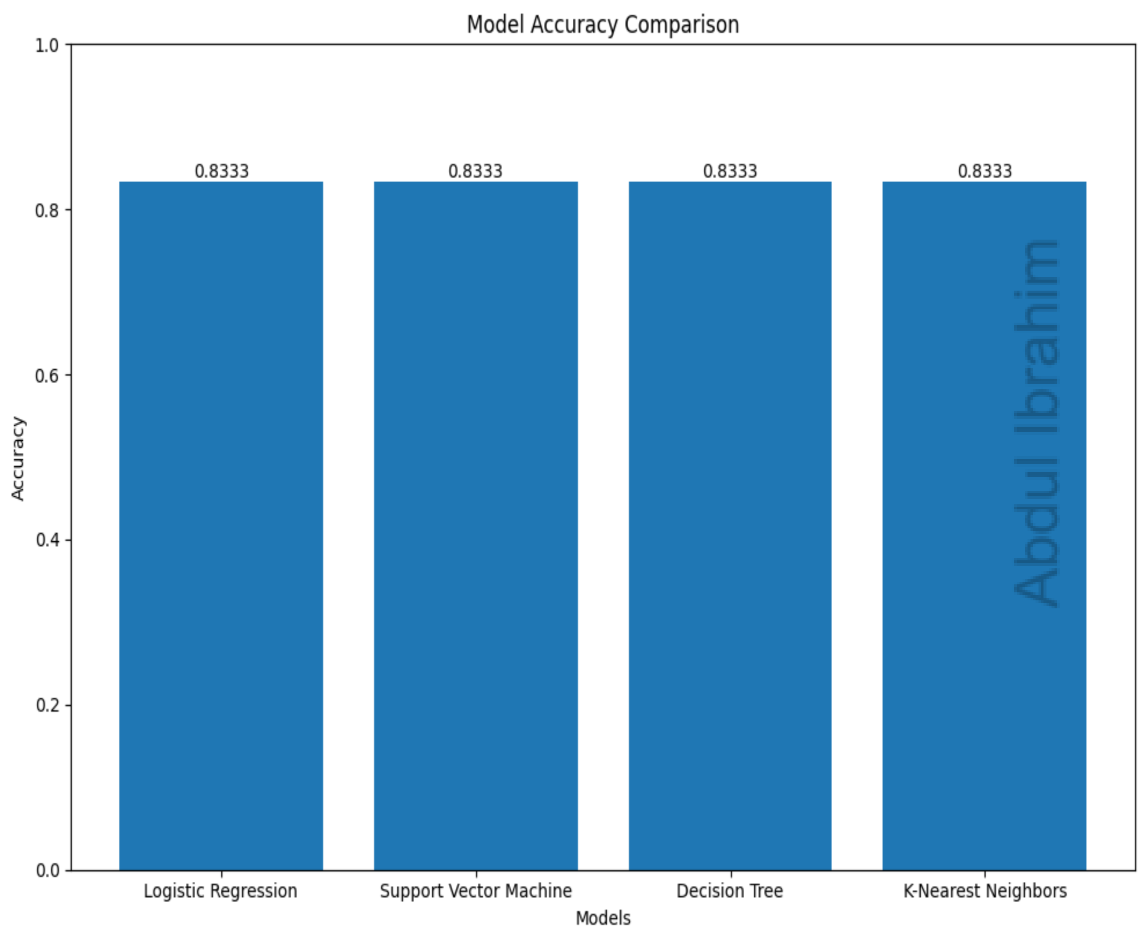
Section 5

# Predictive Analysis (Classification)

Abdul Ibrahim

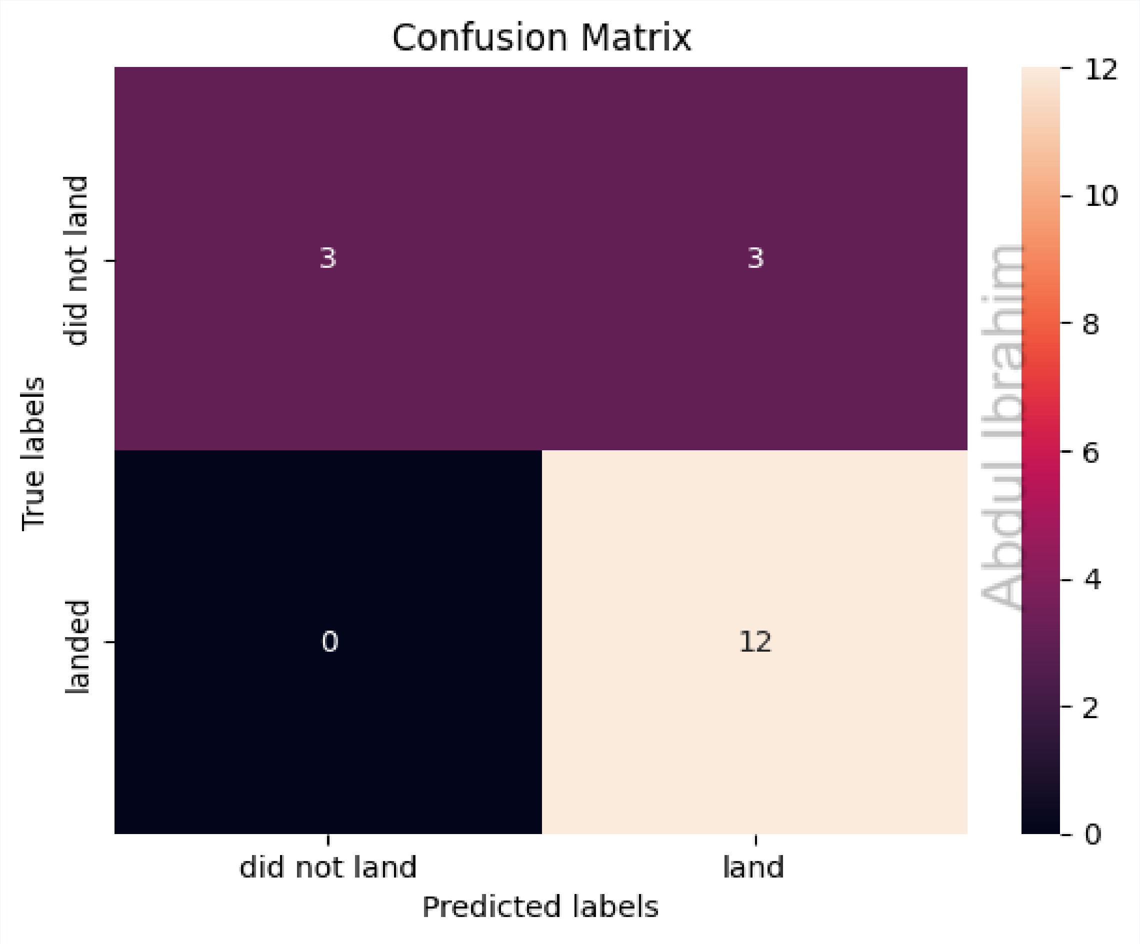
# Classification Accuracy

- The bar chart shows the accuracy for all models (Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest Neighbors) have the same classification accuracy of 0.8333 (83.33%). This is an interesting result, as it suggests that all models performed equally well on the test set.
- For the sake of the implicitly, Logistic Regression will be chosen as model with highest classification accuracy



# Confusion Matrix

- The confusion matrix for Decision Tree model shows **True Negatives (TN): 12**, **False Positives (FP): 3**, **False Negatives (FN): 0** and **True Positives (TP): 3.**
- The model correctly predicted 12 cases where the first stage did not land.
- It correctly predicted 3 cases where the first stage did land
- There were 3 false positives, where the model predicted a landing, but it actually didn't land.
- There were no false negatives, meaning the model didn't miss any successful landings.



# Conclusions

- All four classification models (Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest Neighbors) achieved the same accuracy of 83.33% on the test set. This suggests that the problem may be equally well-suited to various classification approaches.
- The Decision Tree model showed good performance in distinguishing between classes. It correctly identified all successful landings and most of the unsuccessful ones (and potentially the others) while having the highest validation across others model.
- False Positives: The main issue appears to be false positives (3 cases), where the model predicted a successful landing when it actually failed. This could be costly in real-world applications, as it might lead to overestimating the success rate of landings.
- No False Negatives: The model correctly identified all successful landings, which is a positive aspect. It means the model isn't missing any opportunities to recover and reuse the first stage.
- Limited Test Set: Given the small test set size (18 samples), these results should be interpreted with caution. A larger test set would provide more reliable estimates of the models' performance.

Abdul Ibrahim

# Conclusions

- In the early years of SpaceX's landing attempts, they had more failures on drone ships than successes on ground pads. This is expected, as landing at sea is generally more challenging, and these represent some of the first attempts at booster recovery.
- The success rate increased significantly, reaching near 100% by 2020. This demonstrates SpaceX’s improving technology and experience over time

Abdul Ibrahim

# Appendix

The Python code snippets, SQL queries, charts, Notebook outputs, or data sets that used in this project provided in the GitHub repository link (Click to follow) serves as the appendix. This code can be easily modified or expanded to include additional analyses or visualizations as needed.

Abdul Ibrahim

## Thank you!

Abdul Ibrahim