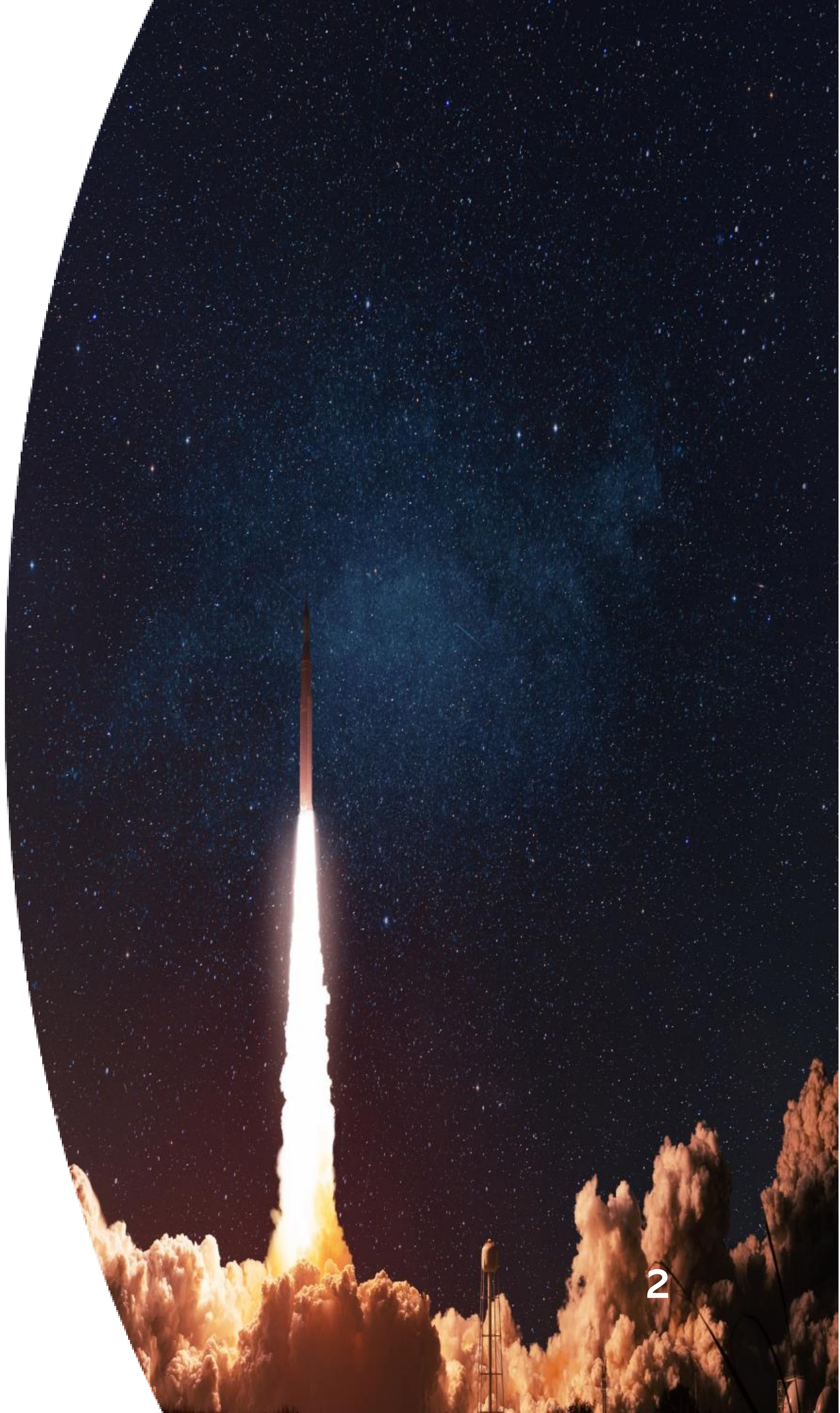
SPACE Y
Final Data Science Capstone Project

Thodoris Garefalakis
09/11/2024

Outline

- Executive Summary **P3**
- Introduction **P4**
- Methodology **P5**
- Results **P17**
    - EDA with Visualization **P19**
    - EDA with SQL **P25**
    - Interactive Maps with Folium **P31**
    - Plotly Dash Dashboard **P35**
    - Predictive Analytics **P39**
- Conclusion **P42**



Executive Summary

Summary of methodologies

- **Data collection** (including API integration and web scraping)
- **Data wrangling**
- **Exploratory Data Analysis** (using SQL)
- **Exploratory Data Analysis** (using Pandas and Matplotlib)
- **Interactive Visual Analysis** (using Folium)
- **Interactive Visual Analysis** (using Plotly Dash)
- **Development of predictive models** for successful landings prediction

Summary of all results

- **Exploratory Data Analysis** revealed that:
    - various relationship patterns among FlightNumber, PayloadMass, Launch Site, Orbit
    - orbits ES-L1, GEO, HEO, SSO demonstrate the highest (100%) success rate
    - success rate was increasing from 2013 till 2020
- **Interactive Visual Analysis** revealed that:
    - the majority of launch sites are close to the coast and near the equator
- **Predictive Analysis** revealed that:
    - Decision tree model outperformed the rest
    - overall, the models showed high predictive ability

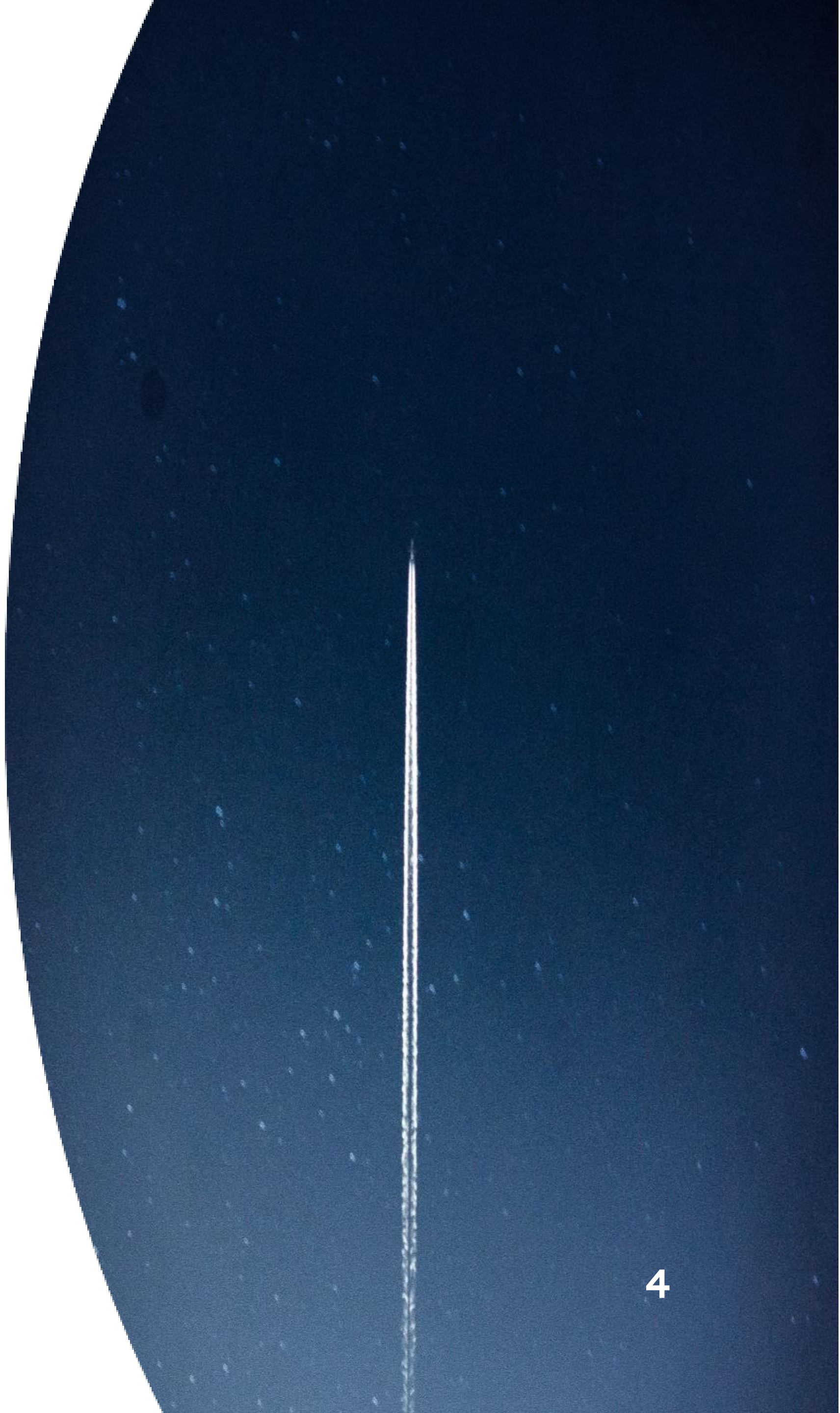
Introduction

Project Background

SpaceX advertises its Falcon 9 launches at $62 million per mission, a substantial savings compared to other providers who charge upwards of $165 million per launch. Much of this cost efficiency comes from SpaceX's ability to reuse the rocket's first stage. Accurately predicting whether this first stage will land successfully allows for more precise launch cost estimations, which is essential for companies, like SpaceY, that seek to competitively bid against SpaceX. Data for this analysis is gathered from available APIs and preprocessed to ensure quality and consistency. By investigating both successful and unsuccessful launch events, the project will identify key factors influencing landing outcomes, supporting more cost-effective and strategically planned space missions.

Project Objectives

- Analyze the influence of payload mass, launch site, number of flights, and orbit on landing success
- Investigate success trends over time
- Identify the most effective predictive model for landing outcomes



Methodology

Methodology

1. **Collect data** through SpaceX REST API and web scraping methods
2. Perform **data wrangling**, by managing missing data, filtering relevant information, and applying one-hot encoding to prepare it for further analysis
3. Perform **exploratory data analysis (EDA)** using visualization and SQL
4. Perform interactive **visual analytics** using Folium and Plotly Dash
5. Perform **predictive analysis** by building, refining, and evaluating classification models to ensure optimal accuracy

Data Collection

The data collection process aimed to compile detailed launch information for Falcon 9 and Falcon Heavy rockets, focusing on variables affecting landing success. Data was gathered from two primary sources: **the SpaceX API and Wikipedia**.

The SpaceX API provided features such as:

- **FlightNumber**
- **PayloadMass**
- **LaunchSite**
- **Outcome**
- **Orbit**

To supplement this, additional historical launch records were collected through Wikipedia web scraping, capturing details such as:

- **Date**
- **BoosterVersion**
- **Historical records of Falcon 9 and Falcon Heavy missions**

This combined dataset includes both structured API data and enhanced historical information from Wikipedia, offering a complete view for analysis

Data Collection - SpaceX API

Steps

01
API Request
Request to the SpaceX API to retrieve launch details

02
Parse API Response
Decode the JSON response from the API to access structured data

03
Extract Relevant Fields
Filter and select key information such as payload details, launch site, and mission outcomes

04
Convert to Dataframe
Organize the extracted data into a DataFrame format for easier handling and analysis

05
Handling Missing Values
from SpaceX API
(rocket launch data)

06
Export data
Save the finalized DataFrame to a CSV file for use in further analysis and modeling

GitHub URL

Data Collection - SpaceX Web Scraping

Steps

01 Request Data

Access the Wikipedia page titled "List of Falcon 9 and Falcon Heavy launches" to retrieve launch records

02 Parse with BeautifulSoup

Load the HTML response into BeautifulSoup to locate the specific table containing launch data

03 Extract Table Rows

Isolate rows and columns in the HTML table that hold relevant details, such as mission date, payload, and launch outcome

04 Structure with DataFrame

Organize the extracted information in a dictionary, then convert it into a DataFrame for ease of manipulation

05 Clean Data

Address any missing or inconsistent values within the DataFrame to ensure data quality

06 Export data

Save the structured data into a CSV file, ready for further analysis and modeling

GitHubURL

Data Wrangling

Exploratory Data Analysis (EDA): Examined dataset structure and calculated missing values

Categorical & Numerical Attributes Identification: Identified attribute types to determine suitable handling techniques

Label Creation: Converted landing outcomes to binary labels (1 for successful, 0 for unsuccessful)

Site and Orbit Analysis: Counted the number of launches from each site and analyzed the distribution of different orbit types (e.g., LEO, GTO, ISS) to identify frequently used facilities and understand the focus and frequency of target destinations across missions

Outcome Categorization: Categorized mission outcomes into “successful” and “unsuccessful” based on landing results, identifying unique patterns (e.g., “True ASDS” & “False RTLS”) to understand variations in mission success

Export to CSV File: Saved the processed dataset, including the new Class labels, to a CSV file for use in subsequent analyses and model training

GitHub URL

EDA with Data Visualization

Plotted Charts:

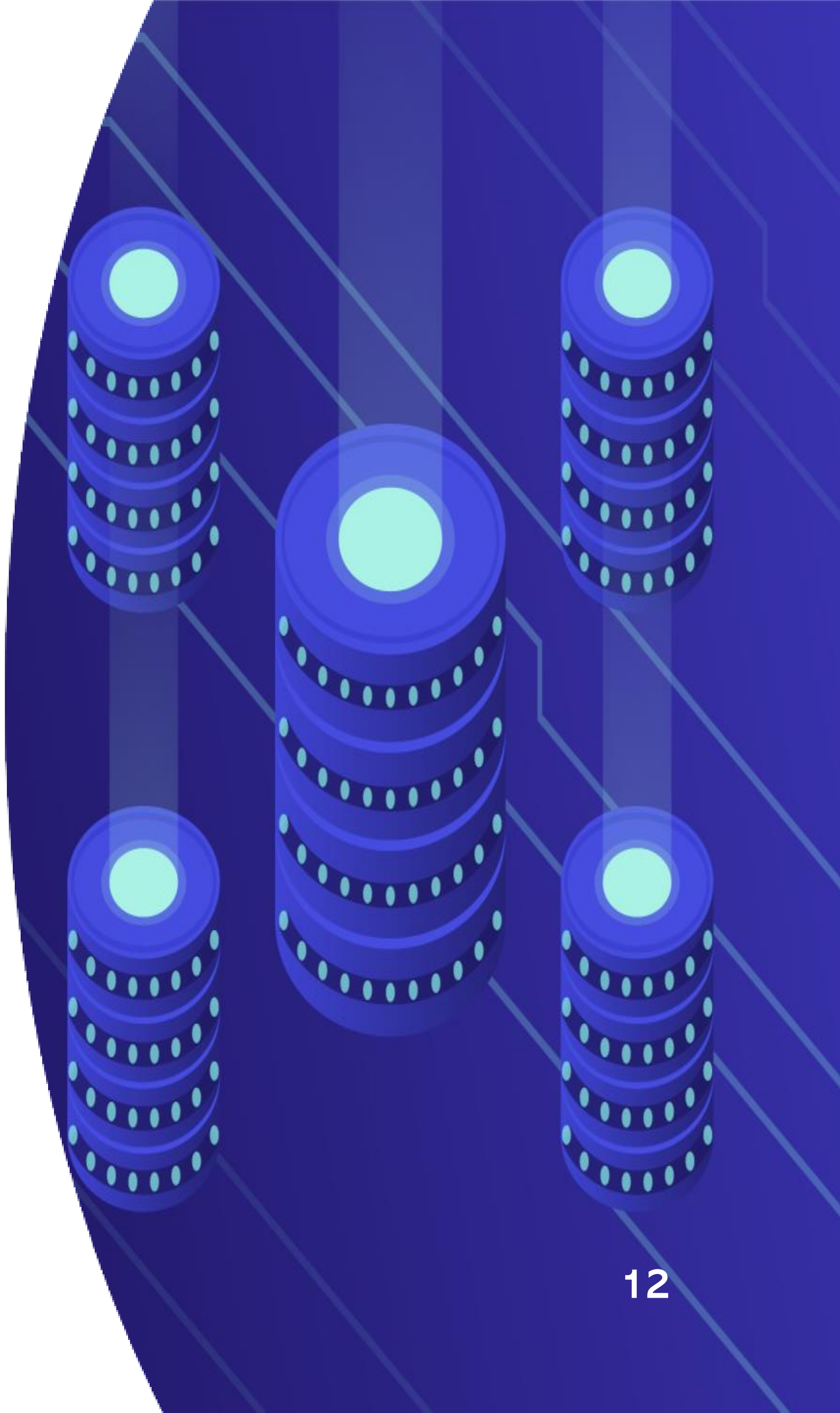
- **Payload Mass (kg) vs Flight Number**: Analyze how payload mass varies across the sequence of launches to identify trends in payload capacity over time
- **Lunch Site vs Flight Number**: Examine the order of launches at each site over time to identify trends in site usage and determine which sites have been used more frequently in SpaceX's launch history
- **Payload Mass (kg) vs Lunch Site**: Compare payload capacities at different launch sites, helping identify if certain sites handle heavier or lighter payloads
- **Success Rate by Orbit Type**: Evaluate the success rates for various orbit types (e.g., LEO, GEO) to understand how orbit choice may impact the likelihood of a successful landing
- **Flight Number vs Orbit Type**: Observe how different orbit types are distributed across launch sequences, indicating possible shifts in mission focus or orbit preferences over time
- **Payload Mass (kg) vs Orbit Type**: Investigate the payload mass associated with each orbit type to identify any relationships between orbit choice and payload capacity
- **Launch Success Over Time**: Track overall launch success trends across years, in order to identify specific patterns (periods of improvement etc)

GitHubURL

EDA with SQL

- **Retrieve Unique Launch Sites**: List all unique launch site names
- **Filter Records by Launch Site**: List records for launches at specific sites (e.g., starting with "CCA")
- **Total Payload by NASA**: Calculate total payload mass for NASA missions
- **Average Payload for Specific Booster Version:** Calculate the average payload mass for a specific booster version (e.g., F9 v1.1)
- **First Successful Ground Pad Landing**: Retrieve the date of the first successful ground pad landing
- **Successful Drone Ship Landings (with Specific Payload Range)**: List successful drone ship landings within a specific payload range (e.g., 4,000-6,000 kg)
- **Mission Success and Failures**: Count total successful and failed launches
- **Booster Versions with Maximum Payload:** Identify booster versions with the highest payload
- **Failed Drone Ship Landings in 2015**: List failed drone ship landings in 2015, with booster and site details
- **Landing Counts by Date Range**: Count landings in a specified date range for trend analysis

GitHubURL



Interactive Visuals with Folium

Launch Site Markers:

Markers placed at each launch site display the site name upon clicking, providing clear geographic reference points across all launch locations.

Success and Failure Markers

Color-coded markers (e.g., green for success, red for failure) distinguish the outcomes of each mission, allowing quick visual assessment of performance by site.

Proximity Lines to Infrastructure

Lines drawn from launch sites to nearby coastlines, cities, and railways show each site’s proximity to key infrastructure, which impacts safety and logistical planning.

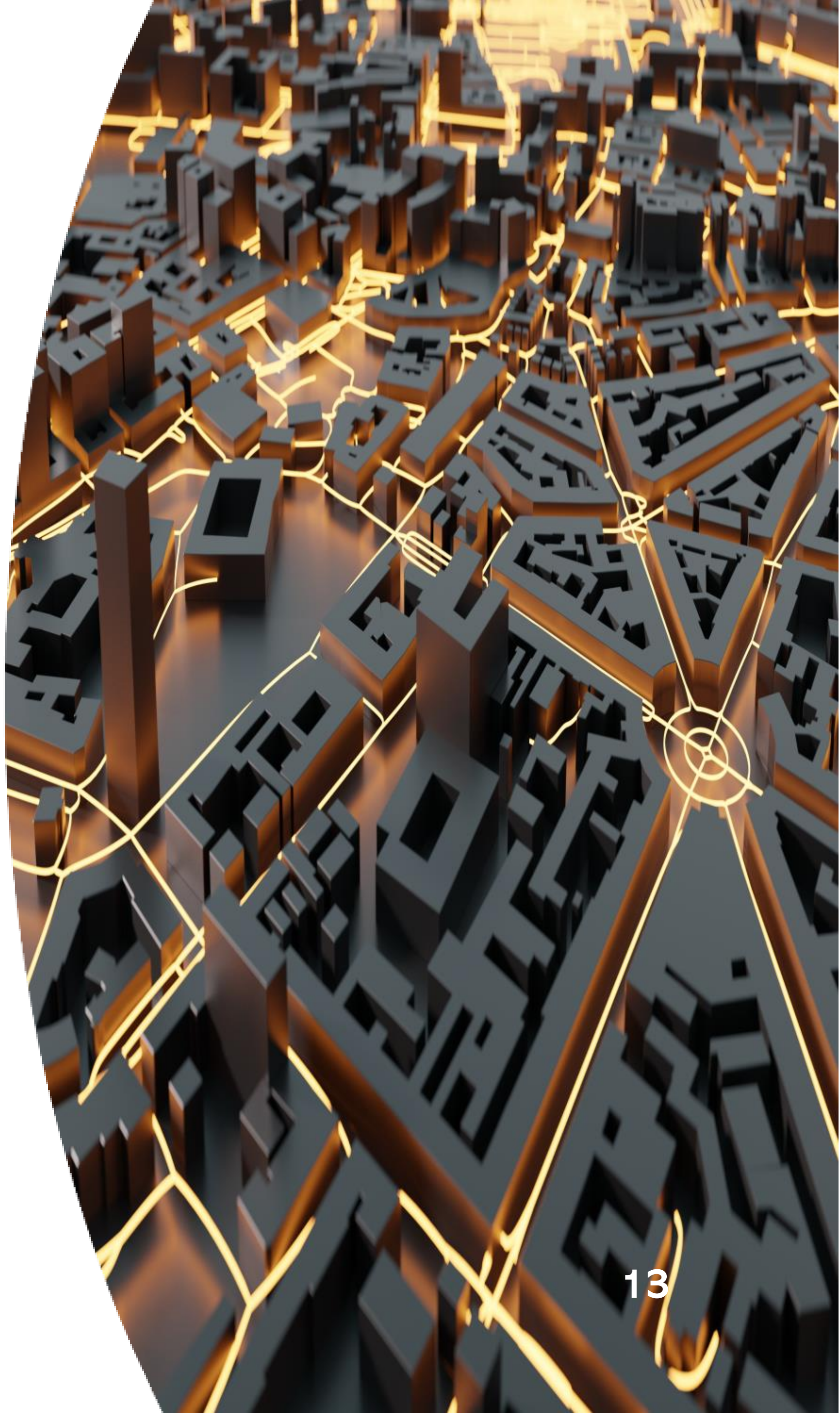
Launch Site Filter

Users can filter the map by individual launch sites to focus on specific locations, enabling a closer look at launch patterns and outcomes.

Payload Range Slider

A slider allows filtering by payload mass, showing how payload size may correlate with landing success at various sites.

GitHub URL



Interactive Visuals with Plotly Dash

Dashboard Features and Interactions

Launch Site Selector:

Dropdown Menu: Allows users to choose a specific launch site or view data for all sites combined. This feature provides focused insights into performance by location

Total Success Launches by Site:

Pie Chart: Displays the proportion of successful and unsuccessful launches at each site, offering a quick visual of site-specific performance

Payload Range Filter:

Slider: Enables users to filter launches by payload mass, helping to identify how payload size correlates with launch success across different sites

Payload vs Success by Booster Version:

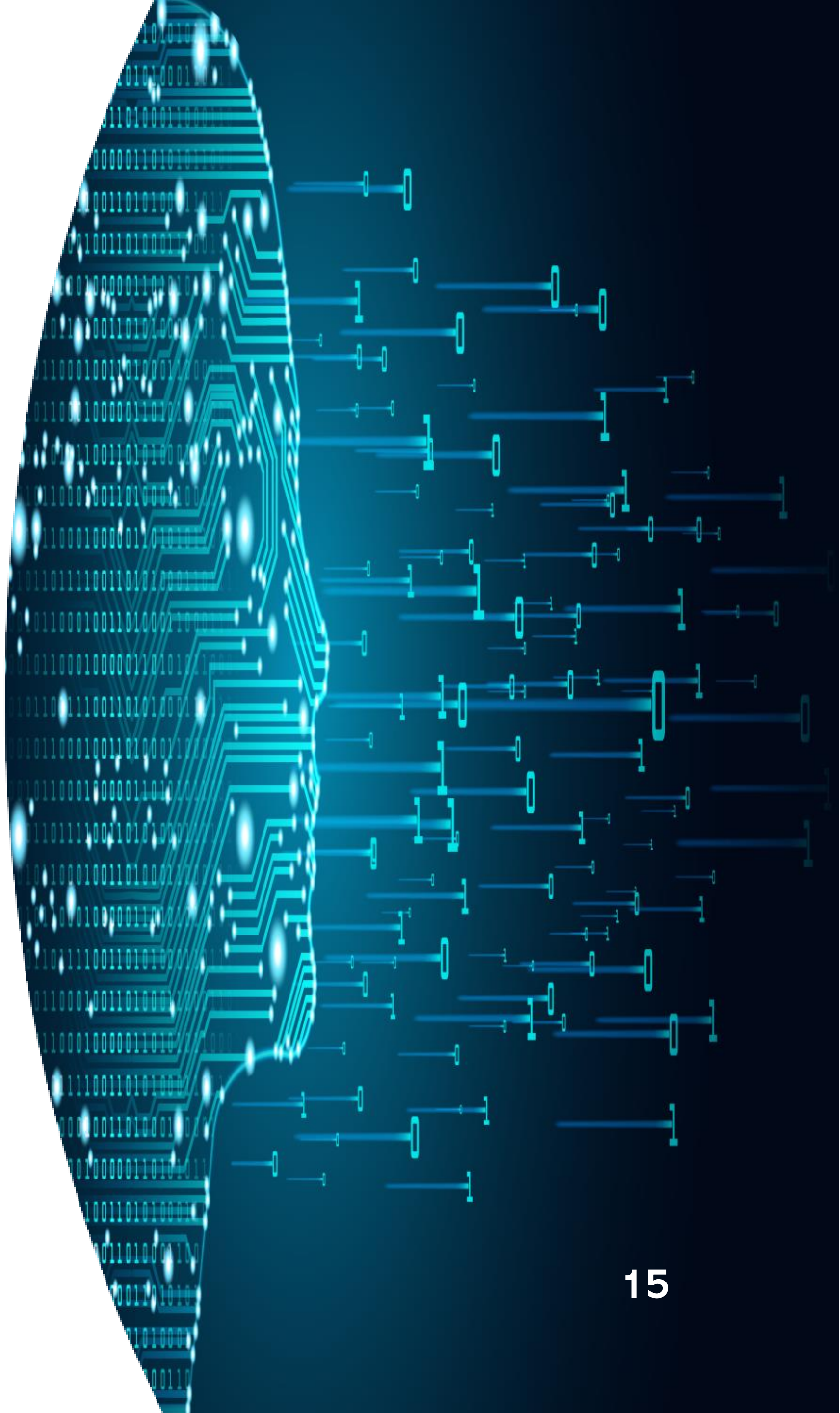
Scatter Plot: Shows the relationship between payload mass and landing success, categorized by booster version. This helps to assess how payload and booster type influence success rates

GitHubURL

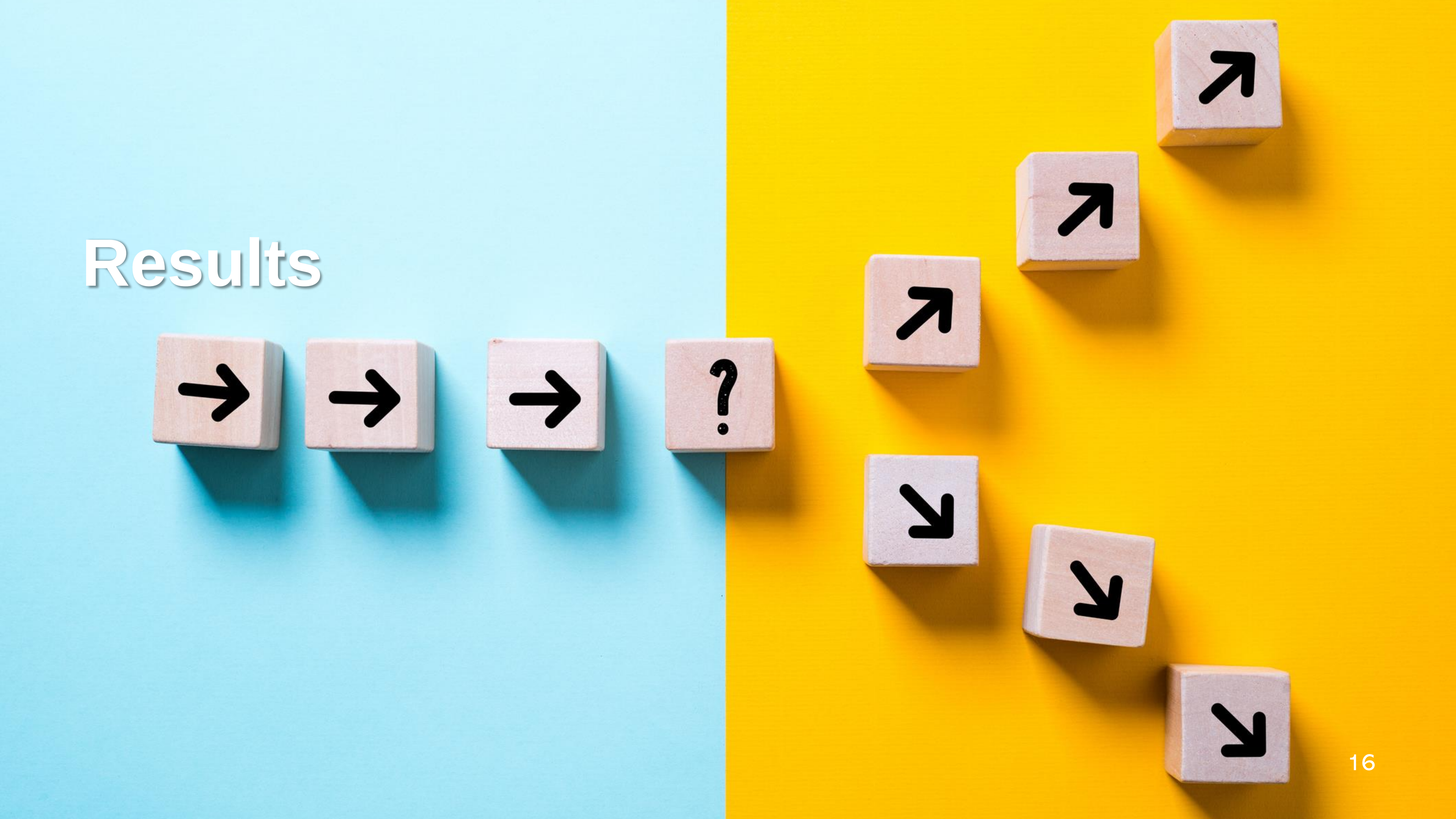
Predictive Analysis (Classification)

Summary of the Prediction Analysis

- **Prepare Data:** Extracted Class column values as a target array
- **Preprocessing Data:** Standardized the features using StandardScaler
- **Split Data:** Divided data into training and test sets with train_test_split
- **Hyperparameter Tuning:** Initialized GridSearchCV with 10-fold cross-validation to tune parameters for various algorithms, including Logistic Regression, SVM, Decision Tree, and K-Nearest Neighbors
- **Evaluate Performance:** Calculated accuracy on test data for each model and analyzed confusion matrices to assess classification effectiveness.
- **Optimize and Select Best Model:** Compared models based on Jaccard_Score, F1_Score, and Accuracy, choosing the highest-performing model as the final classifier



GitHubURL


Results

Results

Exploratory Data Analysis Results

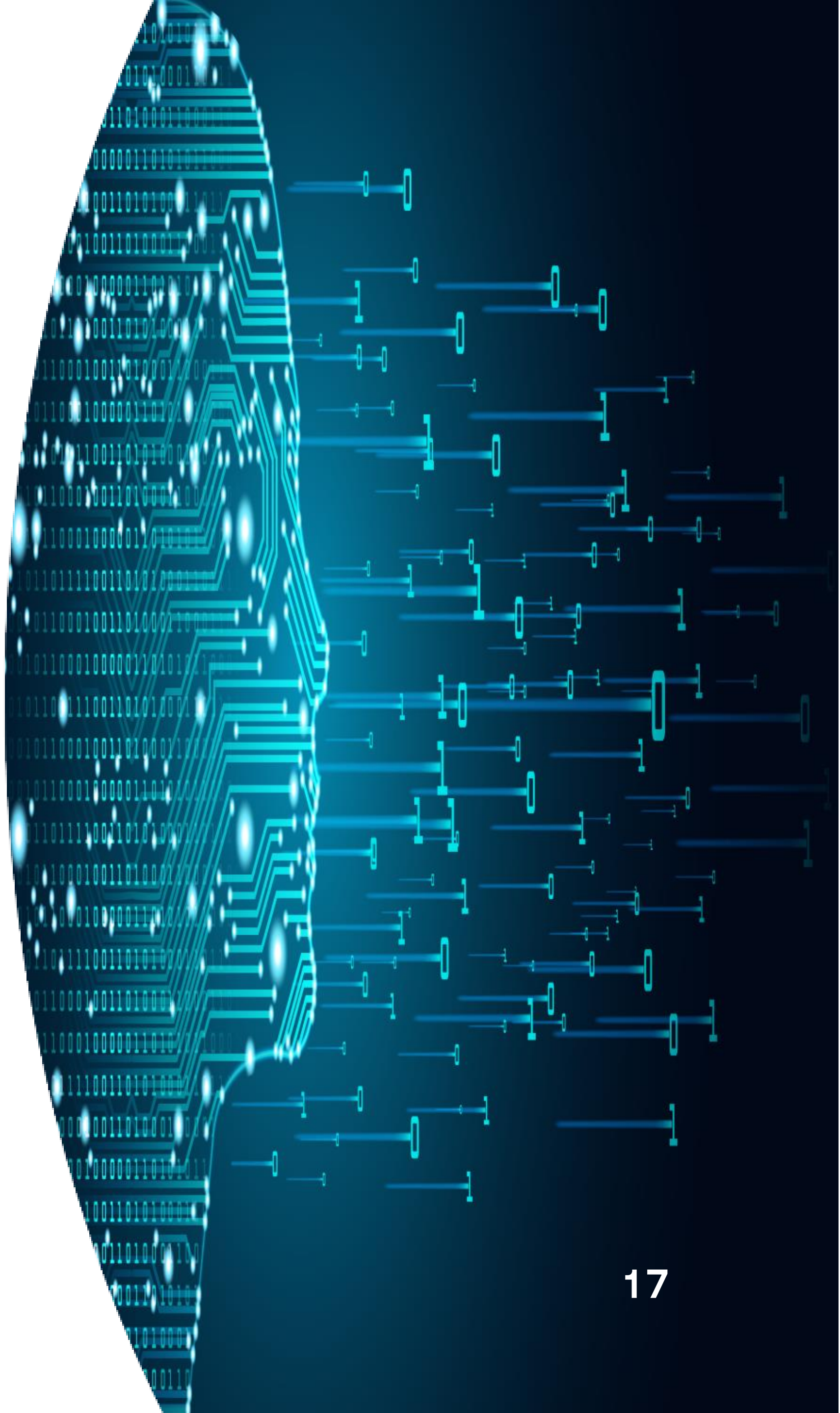
- Key insights showed increased success rates over time,
- Specific launch sites achieved notably high success rates. KSC LC 39A demonstrated exceptional performance
- Orbits ES L1, GEO, HEO and SSO exhibited a 100% success rate

Interactive Visual Analytics

- The majority of launch sites are near coastlines and the equator but are distant from cities, highways, and railways. This placement minimizes risks to populated areas and leverages Earth's rotation for launch efficiency
- Higher payload masses tend to correlate with increased landing success, particularly for specific booster versions

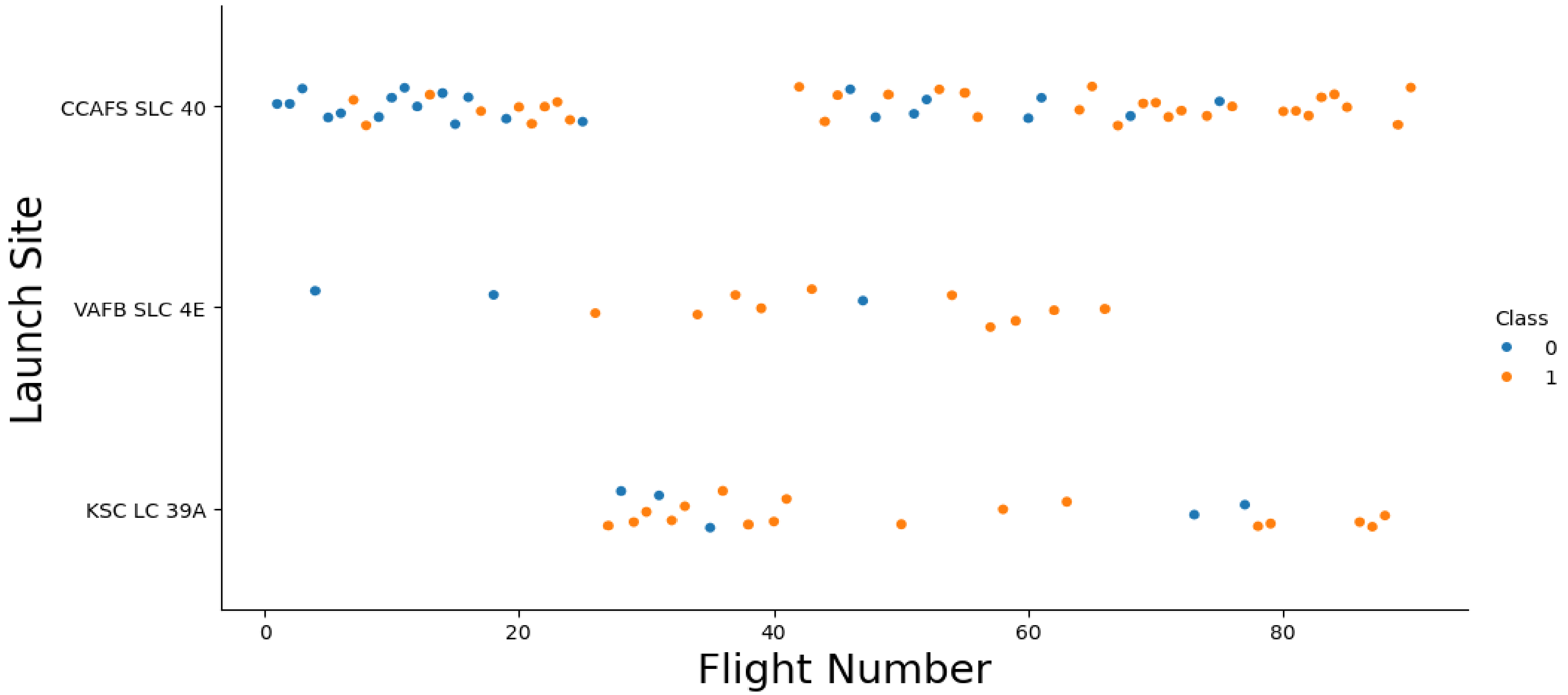
Predictive Analysis Results

- Decision Tree classifier emerged as the best-performing model, achieving high accuracy, Jaccard Score, and F1 Score, making it a reliable tool for predicting landing success based on launch parameters



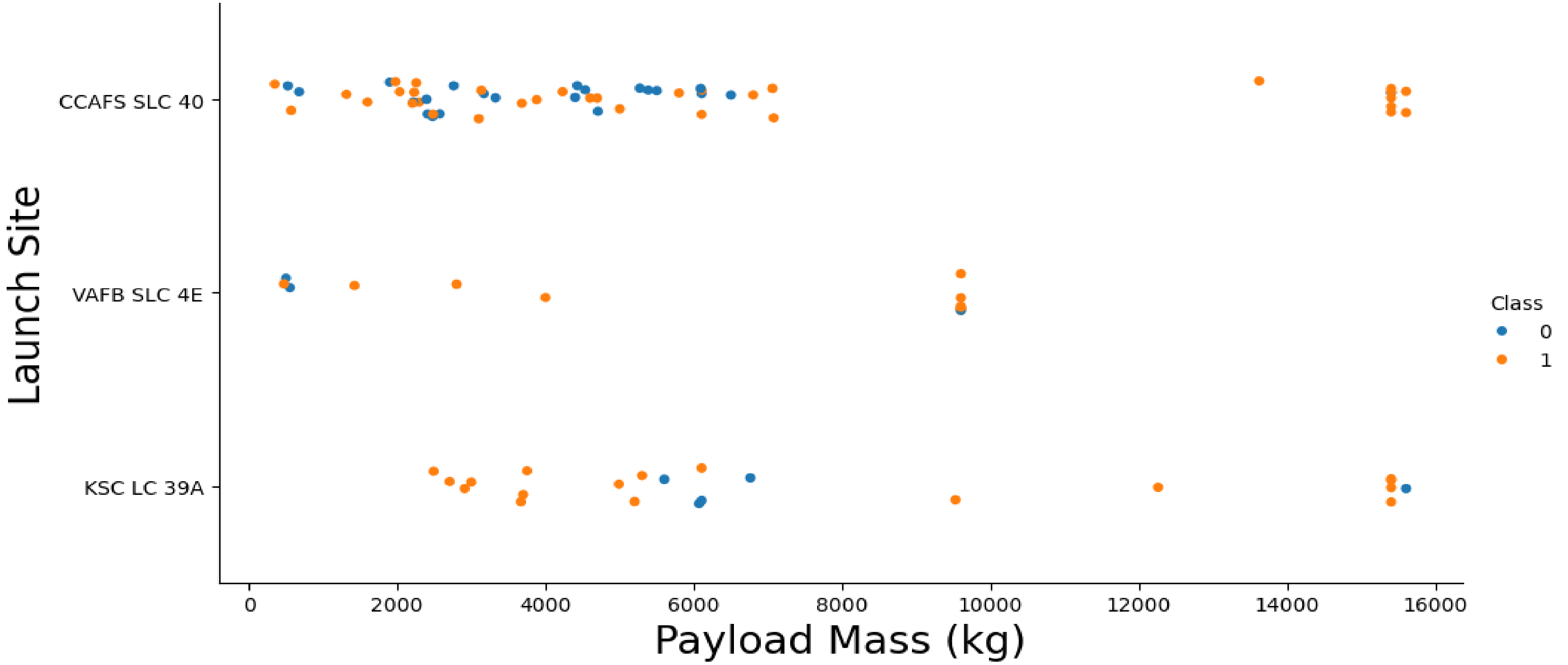
Flight Number vs Launch Site

- **Launch Site Activity**: CCAFS SLC 40 and KSC LC 39A are the most frequently used launch sites, while VAFB SLC 4E has fewer launches
- **Success Rates by Site**: KSC LC 39A shows a higher success rate, with more orange dots (1 = success) than blue (0 = fail)
- **Improvement Over Time**: Higher flight numbers generally show more orange (success), indicating improved landing success over time



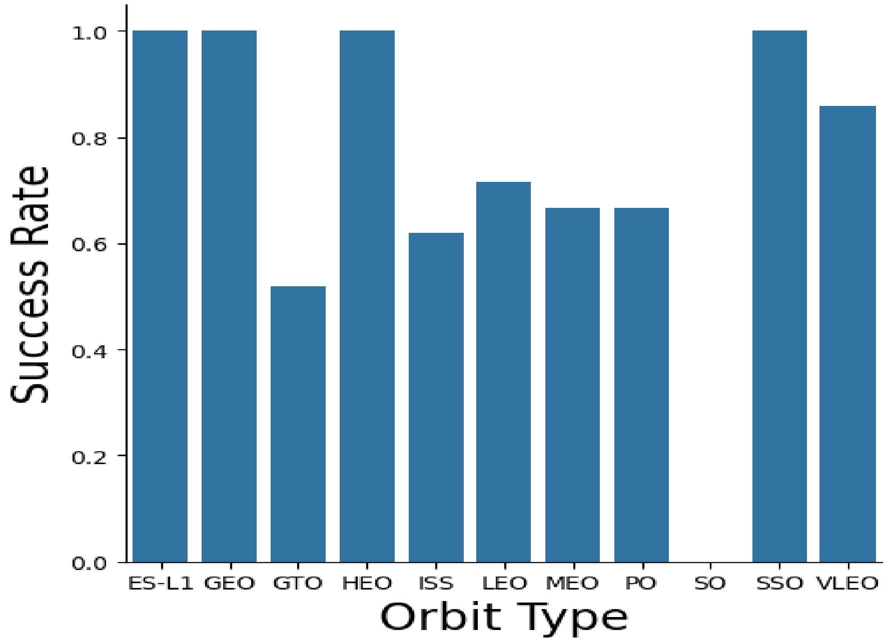
Payload vs Launch Site

- **Launch Site Patterns:** CCAFS SLC 40 and KSC LC 39A handle a wide range of payload masses, while VAFB SLC 4E launches mostly lighter payloads
- **Success by Payload Mass:** Higher payloads (above 10,000 kg) tend to have more successful landings (orange = 1) than failures (blue = 0)
- **Site-Specific Success Trends:** KSC LC 39A shows a higher proportion of successful landings (orange) across a variety of payload masses, suggesting a favorable launch site for consistent landing success



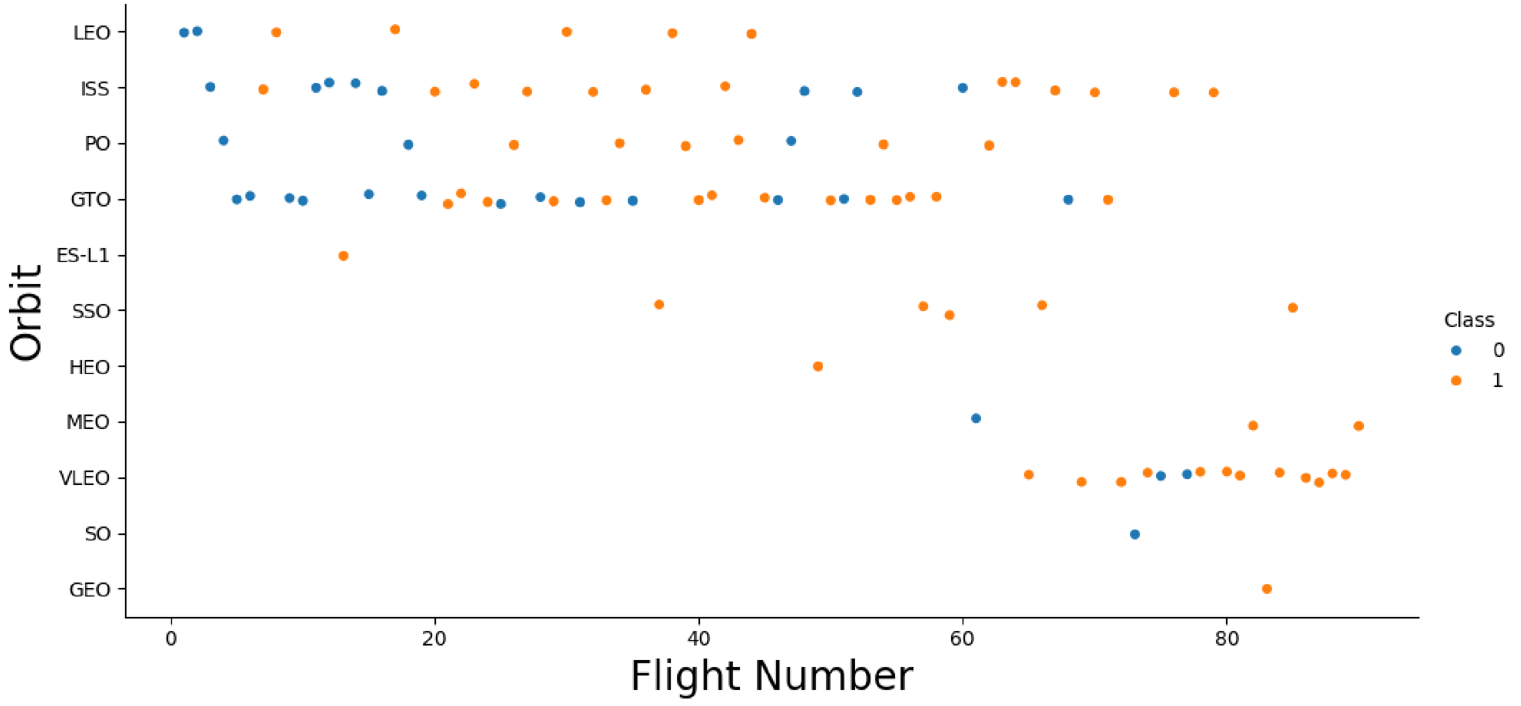
Success Rate vs Orbit Type

- **High Success Orbits:** Orbits like ES-L1, GEO, HEO, and SSO show near 100% success rates, indicating reliable landing outcomes for these mission types
- **Low or Zero Success Orbits:** Orbits like GTO and SO have significantly lower success rates, with SO showing 0% success, which may indicate these missions face higher risks or technical challenges
- **General Success Trends:** Most orbit types still maintain a success rate above 60%, demonstrating SpaceX's overall consistency in landing success across various mission profiles



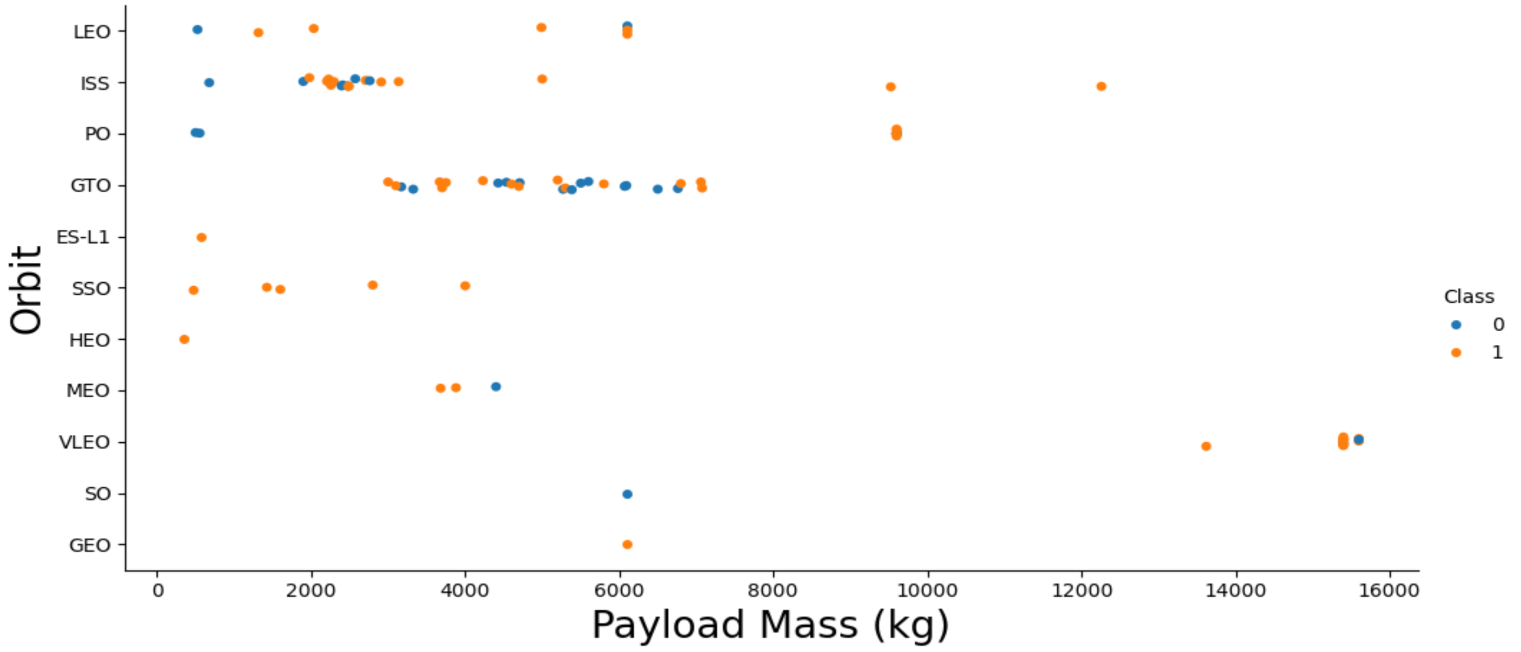
Flight Number vs Orbit Type

- **Time-Based Success:** Success rates (orange = 1) improve with higher flight numbers, reflecting advancements over time
- **Orbit-Specific Patterns:** 100% success for SSO and GEO; 100% failures for SO
- **Common Orbits:** LEO, ISS, PO, and GTO - widely used orbits - show a mix of successes and failures, with more recent launches generally achieving better outcomes



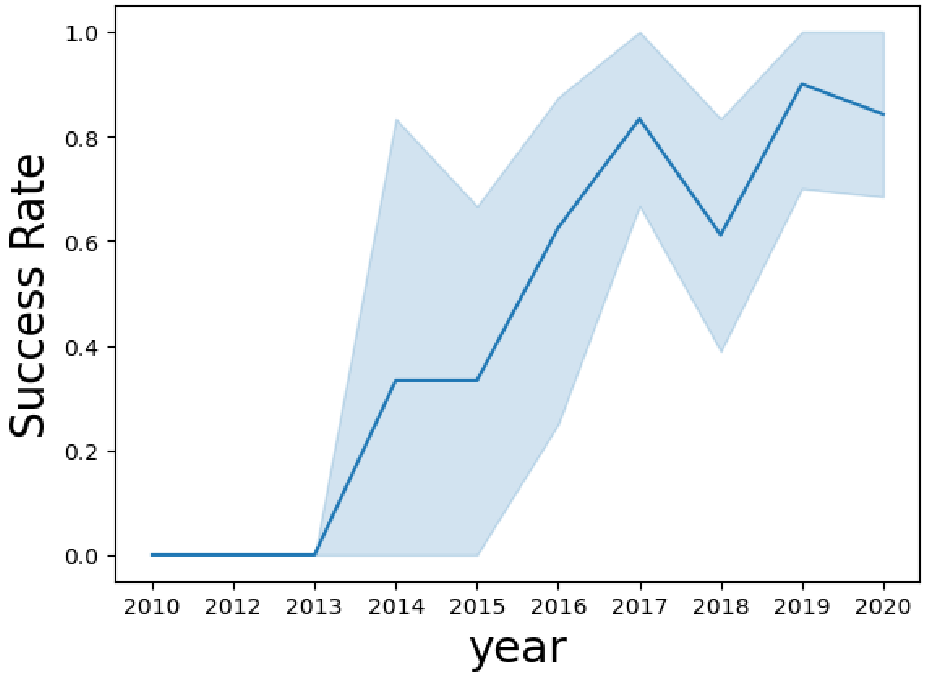
Payload vs Orbit Type

- **Heavy Payload Success:** Orbits like Polar, LEO, and ISS show a higher success rate for heavier payloads
- **Mixed Outcomes for GTO:** For GTO orbit, both successful and unsuccessful landings are observed across different payload masses, making it challenging to establish a clear success trend
- **General Trends Across Orbits:** Other orbits, such as GEO and ES-L1, primarily carry lighter payloads and generally exhibit successful landings, while VLEO often handles heavier payloads with a mix of outcomes



Launch Success Yearly Trend

- **Rising Success Rate:** The success rate shows a clear upward trend from 2013 to 2020, reflecting SpaceX's advancements in landing technology and reliability
- **Notable Improvements:** Significant improvements are visible around 2015 and 2017, where the success rate climbs steeply
- **Consistency in Recent Years:** From 2018 onward, the success rate remains relatively high and stable, generally staying above 70%



All Launch Site Names

The query retrieves distinct launch site names from the dataset. This identifies the unique locations used for SpaceX launches, highlighting the primary sites where operations are conducted

Launch Site Names:

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

In [61]:

```
%sql ibm_db_sa://yyy33800:dwNKg8J3L0IBd6CP@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databas

%sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL;
```

Out[61]:

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

Launch Site Names Begin with 'CCA'

The query filters launch records to display only those where the launch site name starts with 'CCA'. This result highlights launches at Cape Canaveral (CCAFS), providing details on payload, customer, and outcomes for each launch

In [45]:
```
%sql SELECT * \
    FROM SPACEXTBL \
    WHERE LAUNCH_SITE LIKE'CCA%' LIMIT 5;
```

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

Out[45]:

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

Payload Mass

Total Payload Mass

This query calculates the total payload mass carried by SpaceX boosters specifically for NASA (CRS) missions. The result shows that SpaceX has launched a combined payload of 45,596 kg for NASA, highlighting the significant payload capacity utilized in these missions

In [46]:
```
%sql SELECT SUM(PAYLOAD_MASS__KG_) \
    FROM SPACEXTBL \
    WHERE CUSTOMER = 'NASA (CRS)';
```

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

Out[46]:

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

This query calculates the average payload mass carried by the F9 v1.1 booster version. The result, 2,928.4 kg, provides insight into the typical payload capacity handled by this specific booster model

In [47]:
```
%sql SELECT AVG(PAYLOAD_MASS__KG_) \
    FROM SPACEXTBL \
    WHERE BOOSTER_VERSION = 'F9 v1.1';
```

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

Out[47]:

AVG(PAYLOAD_MASS__KG_)
2928.4

Successful Landings & Mission Outcomes

First Successful Ground Landing Date

This query identifies the date of the first successful ground landing by SpaceX. The result, December 22, 2015, marks a significant milestone in SpaceX's landing achievements

```
In [55]: %sql SELECT MIN(Date) \
         FROM SPACEXTBL \
         WHERE Landing_Outcome = 'Success (ground pad)';

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

Out[55]:

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

This query lists payloads that successfully landed on a drone ship and had a mass between 4,000 and 6,000 kg

```
In [56]: %sql SELECT Payload \
         FROM SPACEXTBL \
         WHERE Landing_Outcome = 'Success (drone ship)' \
         AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;

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

Out[56]:

Payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

This query counts the total number of missions categorized by outcome. The results show a high success rate with 98 confirmed successful missions, one in-flight failure, and two other successful outcomes with different status notes

```
In [57]: %sql SELECT MISSION_OUTCOME, COUNT(*) as total_number \
         FROM SPACEXTBL \
         GROUP BY MISSION_OUTCOME;

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

Out[57]:

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

Boosters Carried Maximum Payload

This query retrieves the names of boosters that have carried the maximum payload mass. The result shows multiple F9 B5 booster versions that successfully launched the heaviest payloads, indicating their high capacity and frequent use for heavy-lift missions

In [58]:
```
%sql SELECT BOOSTER_VERSION \
FROM SPACEXTBL \
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

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

Out[58]:

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

Landing Outcome Counts and Rankings

2015 Launch Records

This query retrieves records of failed drone ship landings in 2015, displaying the booster versions and launch sites associated with each failed landing. Both failures in 2015 occurred at CCAFS LC-40 with the F9 v1.1 booster version

```
In [59]: %sql SELECT strftime('%m', Date) AS month, Date, BOOSTER_VERSION, LAUNCH_SITE, [Landing_Outcome] \
FROM SPACEXTBL \
WHERE [Landing_Outcome] = 'Failure (drone ship)' AND strftime('%Y', Date) = '2015';

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

Out[59]:

month	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

This analysis ranks the different landing outcomes within the specified date range. "No attempt" is the most frequent outcome, followed by an equal number of successes and failures on drone ships. This ranking provides insight into the frequency of each landing type during SpaceX’s early launch history

```
In [60]: %sql SELECT [Landing_Outcome], count(*) as count_outcomes \
FROM SPACEXTBL \
WHERE DATE between '2010-06-04' and '2017-03-20' group by [Landing_Outcome] order by count_outcomes DESC;

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

Out[60]:

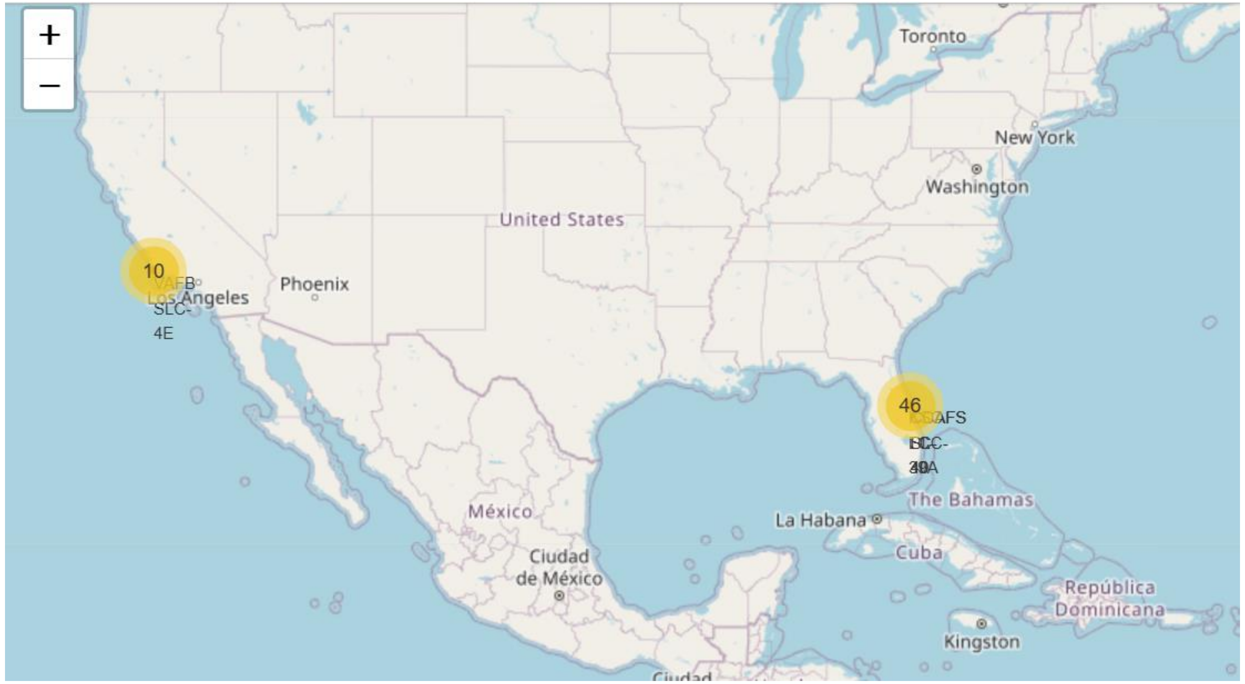
Landing_Outcome	count_outcomes
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Launch Site Proximities Analysis

Launch Sites Overview on Folium Map

Key Insights:

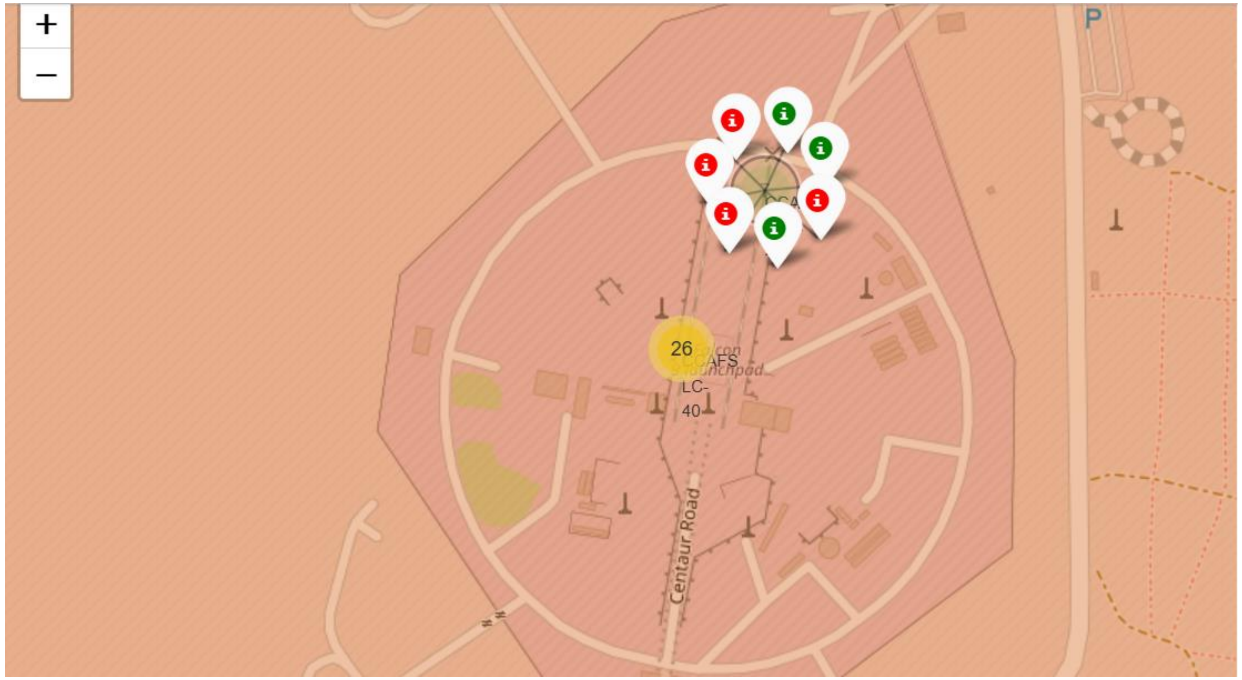
- East Coast has the highest number of launches (46), while West Coast has a smaller number of launches (10)
- Launch sites on the East and West Coasts provide access to a variety of orbital inclinations, making SpaceX's operations flexible and efficient
- Coastal locations allow for safe recovery zones over the ocean, minimizing risks to populated areas



Launch Site Outcomes on Folium Map

Key Insights:

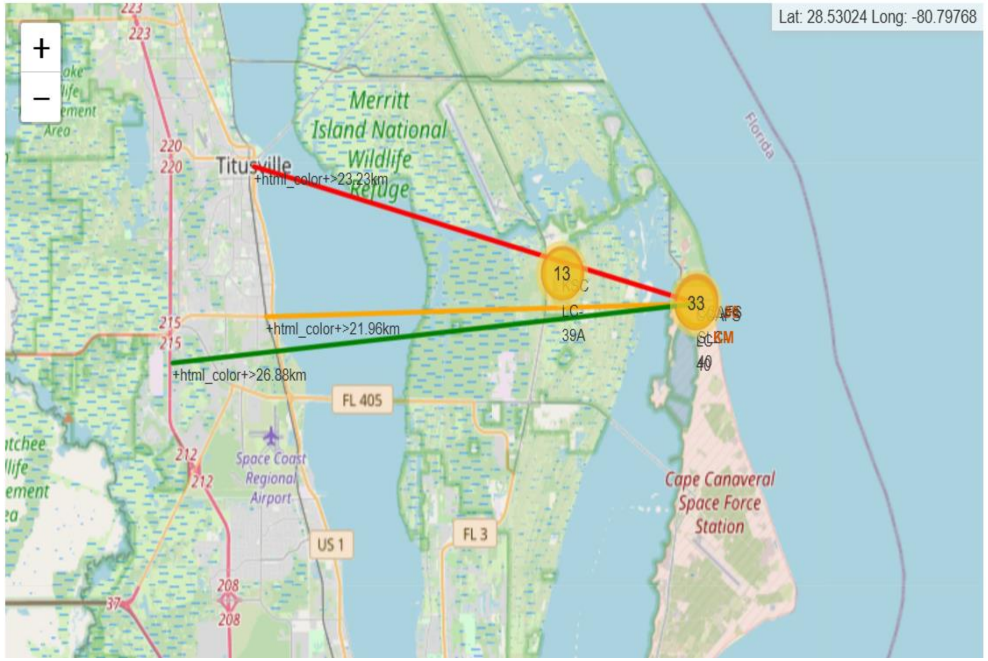
- Color-coding allows for quick assessment of success rates, highlighting operational progress at this location over time
- Each marker represents a specific launch attempt, where green markers show successful launches and red markers indicate failures
- The map reveals that CCAFS LC-40 launch site, has approx. 43% success rate (3/7 successful launches)

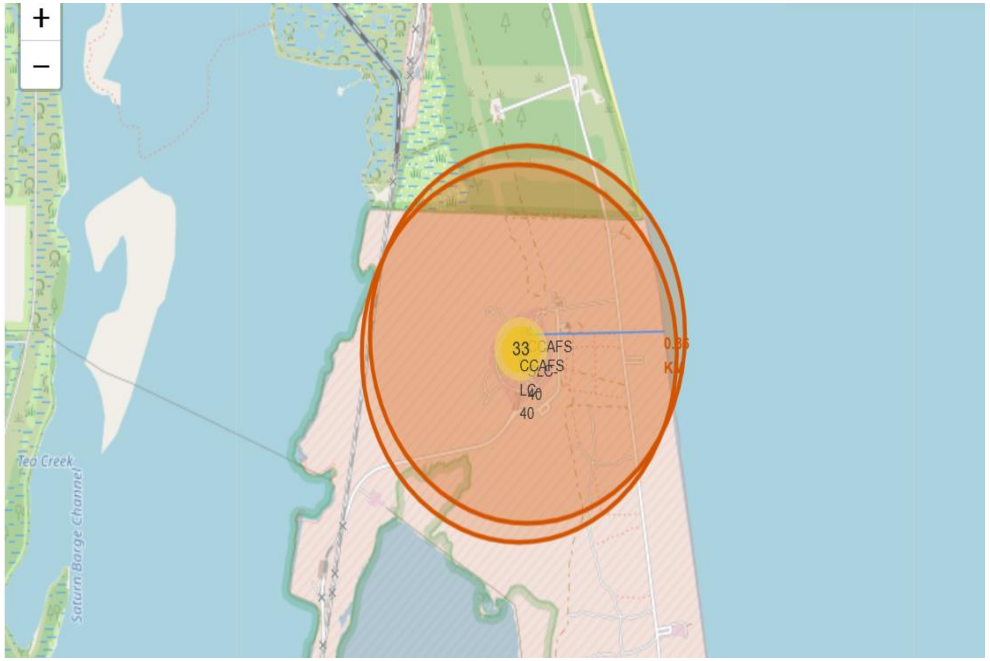


Proximity Analysis of Launch Sites

Key Insights (for CCAFS SLC-40 launch site):

- **Blue Line**: Distance to the coastline (0.86 km), allowing for safe over-water launch and landing trajectories
- **Red Line**: Distance to the nearest city (23.23 km), showing urban separation for safety
- **Orange Line**: Distance to the closest railway (21.96 km), which may facilitate logistical operations
- **Green Line**: Distance to the nearest highway (26.88 km), providing access while maintaining a safe boundary



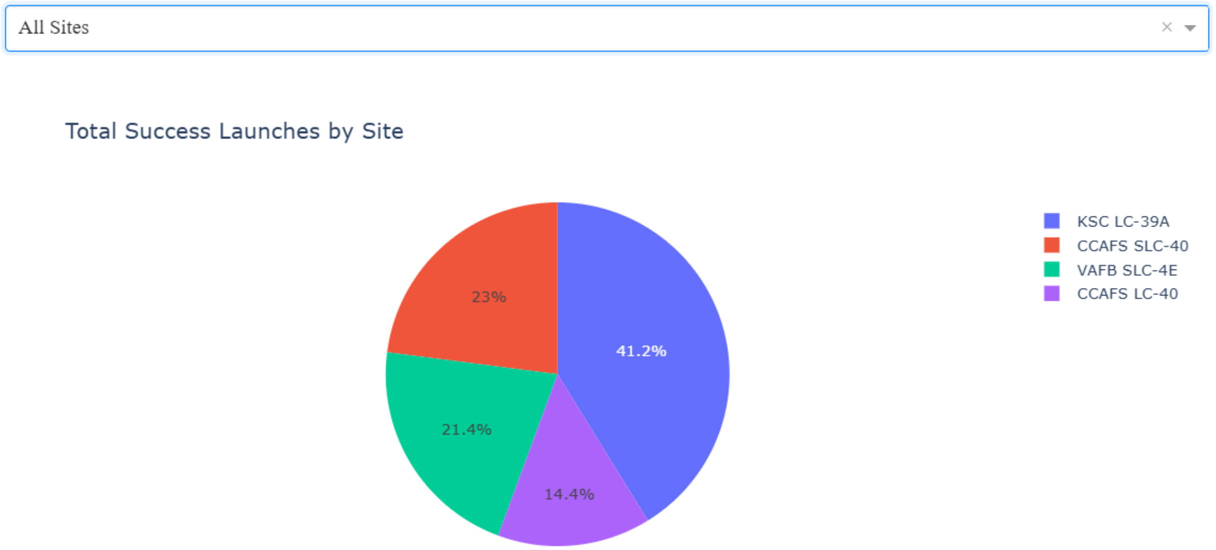


Build a Dashboard With Plotly Dash

Launch Success Distribution by Site

Key Insights:

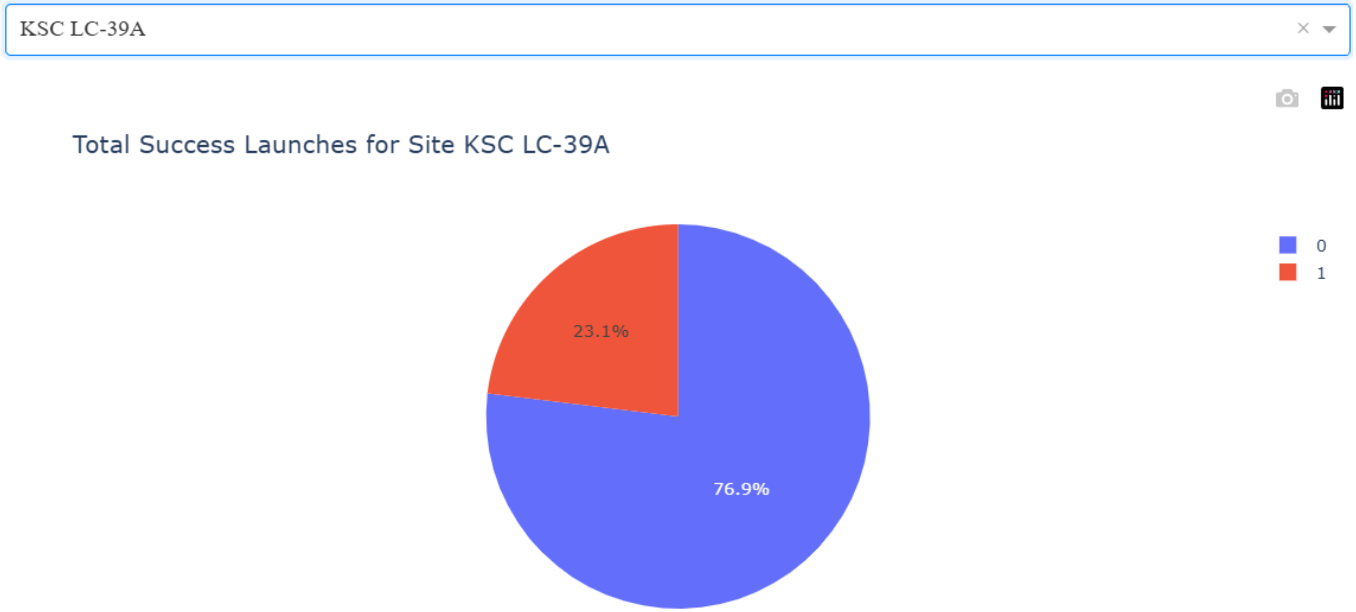
- **KSC LC-39A** has the highest success rate, representing 41.2% of all successful launches, indicating its role as a major launch site.
- **CCAFS SLC-40** and **VAFB SLC-4E** follow with 23% and 21.4% respectively, showing significant activity and success.
- **CCAFS LC-40** has the lowest success share at 14.4%, possibly reflecting fewer launches or a less critical role in SpaceX's operations.
- The diversity in success rates across sites provides insight into site-specific usage and operational focus within SpaceX's launch network



Success Ratio for KSC LC-39A Launch Site

Key Insights:

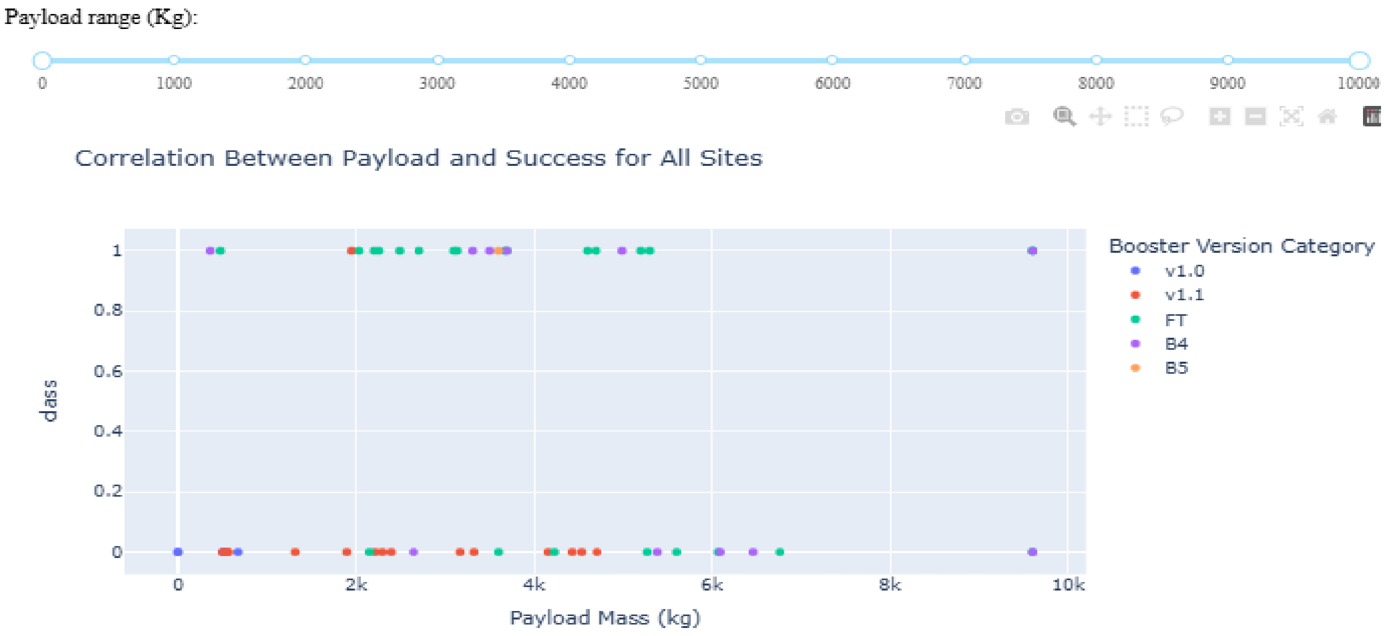
- **76.9%** success rate (in blue) indicates a high reliability in successful launches from this site, making it a critical asset in SpaceX's operations
- **23.1%** failure rate (in red) provides insight into the challenges faced at KSC LC-39A but still highlights a strong success trend compared to other sites
- This success ratio emphasizes KSC LC-39A's role in contributing to SpaceX's achievements in achieving consistent and reliable launches



Correlation Between Payload Mass and Launch Success for All Sites

Key Insights:

- **Lighter payloads** (under 4,000 kg) generally show a higher success rate, especially for booster versions FT, B4, and B5
- **Heavy payloads** (above 6,000 kg) have mixed results, indicating that higher payloads might pose more challenges for successful laucnhes
- Booster Versions FT and B5 appear more reliable across different payload ranges, suggesting advancements in booster technology contribute to higher success rates



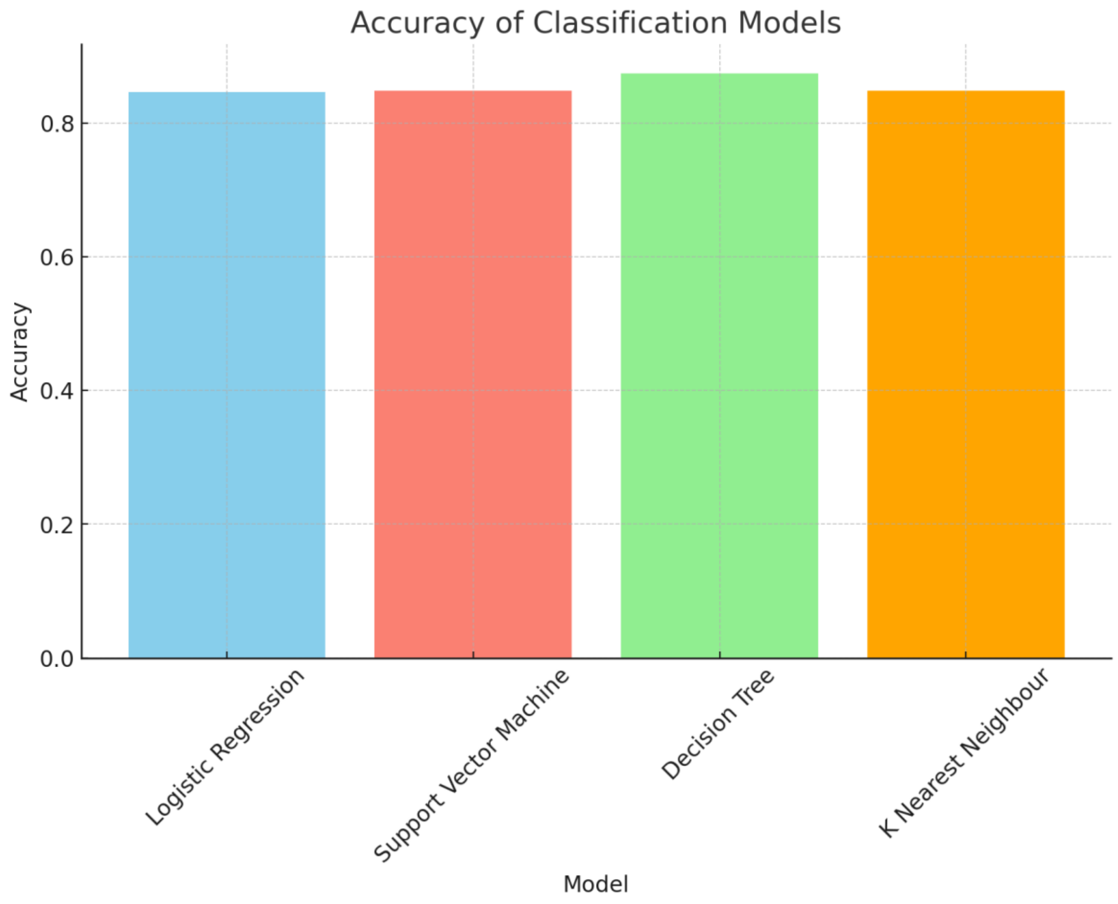
Predictive Analysis (Classification)

Classification Accuracy

Key Insights:

- All models performed relatively well, with accuracy rates above 84%, making them all viable for classification
- Decision Tree stands out as the best model due to its slightly higher accuracy (87.7%) and the fact that its optimal hyperparameters could be fine-tuned for even better results

```
In [40]: # Best scores for each model
logreg_score = logreg_cv.best_score_
svm_score = svm_cv.best_score_
tree_score = tree_cv.best_score_
knn_score = knn_cv.best_score_

# Best parameters for each model
logreg_params = logreg_cv.best_params_
svm_params = svm_cv.best_params_
tree_params = tree_cv.best_params_
knn_params = knn_cv.best_params_

# Create a dictionary to store scores and parameters for each model
model_scores = {
    'Logistic Regression': {'score': logreg_score, 'params': logreg_params},
    'Support Vector Machine': {'score': svm_score, 'params': svm_params},
    'Decision Tree': {'score': tree_score, 'params': tree_params},
    'K Nearest Neighbour': {'score': knn_score, 'params': knn_params}
}

# Find the model with the highest score
best_model = max(model_scores, key=lambda x: model_scores[x]['score'])
best_score = model_scores[best_model]['score']
best_params = model_scores[best_model]['params']

# Display the best model, its score, and optimal hyperparameters
print(f"The best model is: {best_model}")
print(f"Best Score: {best_score}")
print(f"Optimal Hyperparameters: {best_params}")

The best model is: Decision Tree
Best Score: 0.8767857142857143
Optimal Hyperparameters: {'criterion': 'gini', 'max_depth': 18, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_spl
it': 5, 'splitter': 'random'}
```



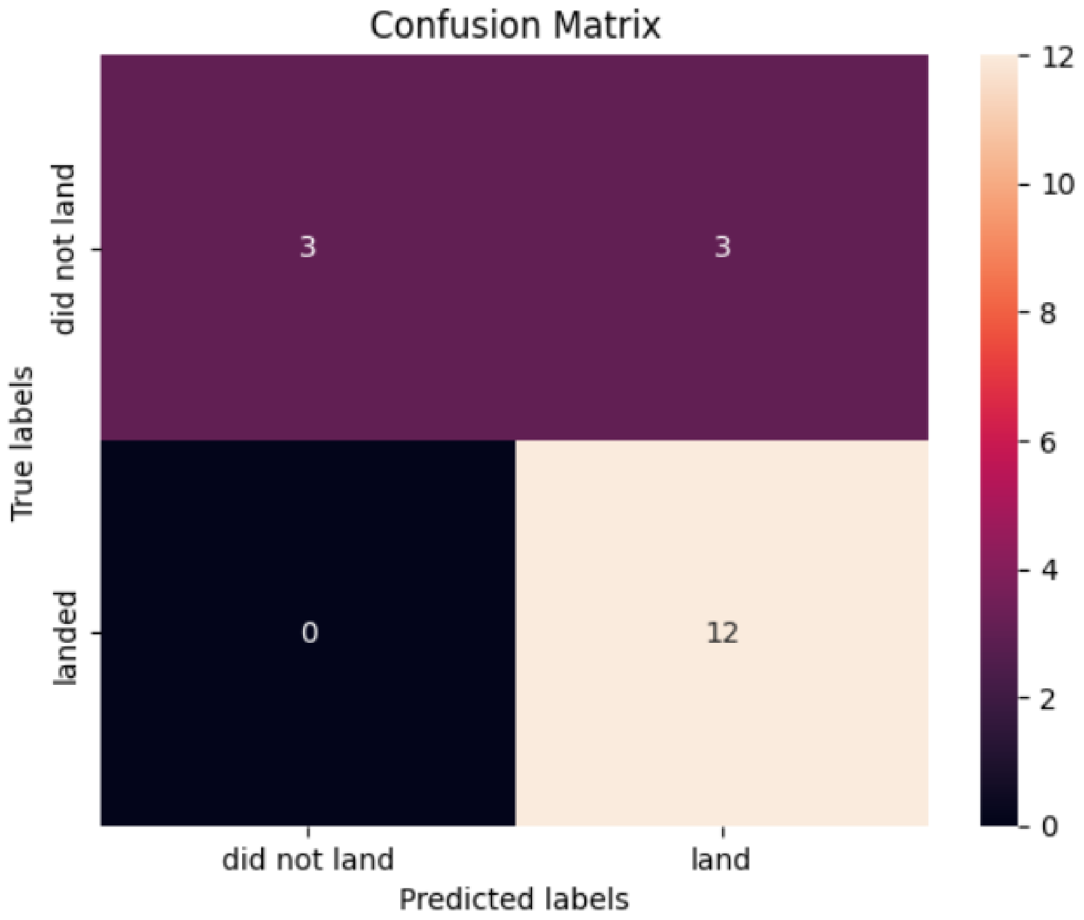
Confusion Matrix

- **False Positives (3)**: Incorrectly predicted “land” when the actual outcome was “did not land”

The confusion matrix provides insights into the performance of the Decision Tree model in predicting landing outcomes:

- **True Positives (12)**: Predicted “land” correctly when the actual outcome was “land”
- **True Negatives (3)**: Predicted “did not land” correctly when the actual outcome was “did not land”
- **False Negatives (0)**: There were no instances where the model predicted “did not land” while the actual outcome was “land”

This matrix indicates a **high accuracy** in correctly identifying successful landings, with a few misclassifications in predicting unsuccessful landings. The absence of false negatives demonstrates the model's strong performance in identifying all actual landings correctly



Conclusions

Improved Landing Success Over Time

- Success rates showed a clear upward trend from 2013 to 2020, driven by advancements in SpaceX's landing technology and reliability

Strategic Launch Site Utilization

- KSC LC-39A had the highest success rate, highlighting its strategic importance. Coastal launch sites near the equator, such as Cape Canaveral, provide safe recovery zones over the ocean and benefit from Earth's rotation, optimizing launch efficiency

Orbit and Mission Trends

- Orbits like ES-L1, GEO, HEO, and SSO achieved a 100% success rate, showing reliability for specific mission types. Other orbits, like SO, faced greater challenges, impacting mission success variability

Map-Based Insights on Launch Sites

- Interactive map analysis revealed that launch sites are strategically located near coastlines but away from populated areas, minimizing risks and optimizing logistics. Color-coded markers on the map illustrated success and failure rates at each site, allowing a quick visual assessment of site-specific performance over time

Predictive Model Performance

- The Decision Tree classifier was the top-performing model, accurately predicting outcomes and enabling data-driven decisions for future missions


THANK YOU!

SPACE Y
Final Data Science Capstone Project

Thodoris Garefalakis
09/11/2024