IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Yashdeep Singh Dadiala
22/06/25

Outline

Executive Summary

Summary of Methodologies

- Data Collection: Web scraping and CSV extraction of SpaceX launch data.
- Data Wrangling: Cleaned and merged data using pandas; normalized columns and handled missing values.
- Exploratory Data Analysis:
  - SQL queries via SQLite to uncover patterns.
  - Visual analysis with matplotlib, plotly, and Folium.
- Interactive Tools:
  - Created a Plotly Dash dashboard with dropdowns, pie charts, sliders, and scatter plots.
  - Geospatial Analysis with Folium for mapping launch sites and calculating distances.
- Predictive Modeling:
  - Built a Decision Tree Classifier using GridSearchCV for hyperparameter tuning.
  - Evaluated using accuracy score and confusion matrix.

Summary of All Results

- Highest Success Rate Site: Kennedy Space Center (KSC LC-39A).
- Payload Insight: Range of 2000–6000 kg had highest launch success.
- Best Booster Version: FT and B5 had best success record.
- SQL Analysis: Identified first ground pad landing and booster performance with payload ranges.
- Folium Analysis: Launch sites are close to coasts, railways, and highways — supporting efficient logistics.
- Model Accuracy:
  - Train: ~85.9%
  - Test: ~72.2%

Introduction

Project Background and Context

SpaceX has transformed the aerospace industry by developing reusable rockets, significantly reducing launch costs. With an increasing number of launches and innovations, analyzing historical launch data allows us to understand performance patterns and success factors. This project applies data science methods to explore, visualize, and model SpaceX launch data for actionable insights and predictive capabilities.

Problems You Want to Find Answers For

- Which launch sites have the highest success rates?
- How does payload mass influence launch outcomes?
- What booster versions are associated with the most successful launches?
- Are launch sites optimally located with respect to geographical features?
- Can we predict launch success using key mission features?

Section 1

Methodology

Methodology

Executive Summary

- **Data Collection Methodology:**
  We collected SpaceX launch data using a combination of web scraping and public datasets, consolidating launch details, payloads, and booster versions into a unified dataset.
- **Data Wrangling:**
  The data was cleaned, merged, and transformed using pandas and SQL to handle missing values, standardize columns, and prepare new features like success flags and location identifiers.
- **Exploratory Data Analysis (EDA):**
  EDA was conducted using SQL queries and visual tools to uncover patterns in launch outcomes, payload distributions, and booster performance across different launch sites.
- **Interactive Visual Analytics:**
  We developed interactive visualizations using Folium maps for geographic insights and a Plotly Dash dashboard for dynamic analysis of site performance and payload success rates.
- **Predictive Analysis Using Classification Models:**
  Machine learning models were built and tuned to predict launch success, with Decision Tree achieving the best performance at 86% cross-validated accuracy and 72% test accuracy.

Data Collection

SpaceX launch data was collected via REST API calls to the official SpaceX API. The primary endpoint provided JSON data on historical launches, which was then normalized into a structured DataFrame. To enrich this dataset, additional API calls were made using ID fields to extract booster version, payload mass, orbit type, launch site coordinates, and core specifications. Data was filtered to include only Falcon 9 missions, with payload mass cleaned using mean imputation for missing values. The final curated dataset was exported for subsequent analysis.

Data Collection – SpaceX API

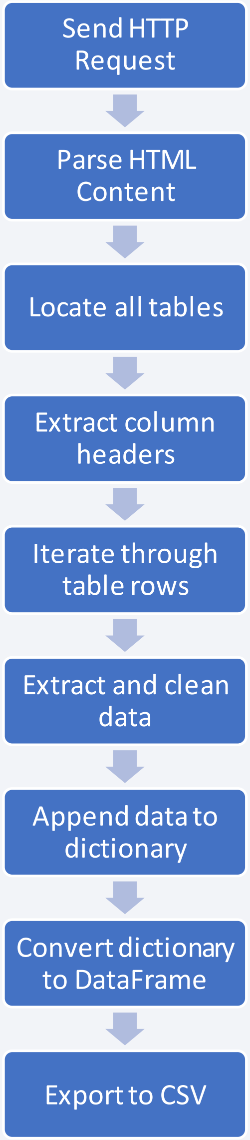
- We used SpaceX's public REST API to collect launch data. Key fields such as rocket ID, launchpad, payloads, and cores were retrieved via GET requests. Auxiliary API calls extracted additional attributes like booster version, landing outcome, and coordinates. The final data was compiled into a structured DataFrame and filtered to include only Falcon 9 launches.
- Link to Github

Data Collection - Scraping

Key Phrases – Web Scraping Process

- **URL Access:** Requested HTML snapshot from Wikipedia using requests.get().
- **HTML Parsing:** Parsed content with BeautifulSoup.
- **Table Extraction:** Located and isolated target launch table.
- **Header Processing:** Cleaned and extracted column names.
- **Row Iteration:** Parsed each row for relevant launch data fields.
- **Data Cleaning:** Applied helper functions to extract and clean values.
- **Data Structuring:** Stored records in a dictionary then converted to a DataFrame.
- **Export:** Saved dataset as CSV for further analysis.

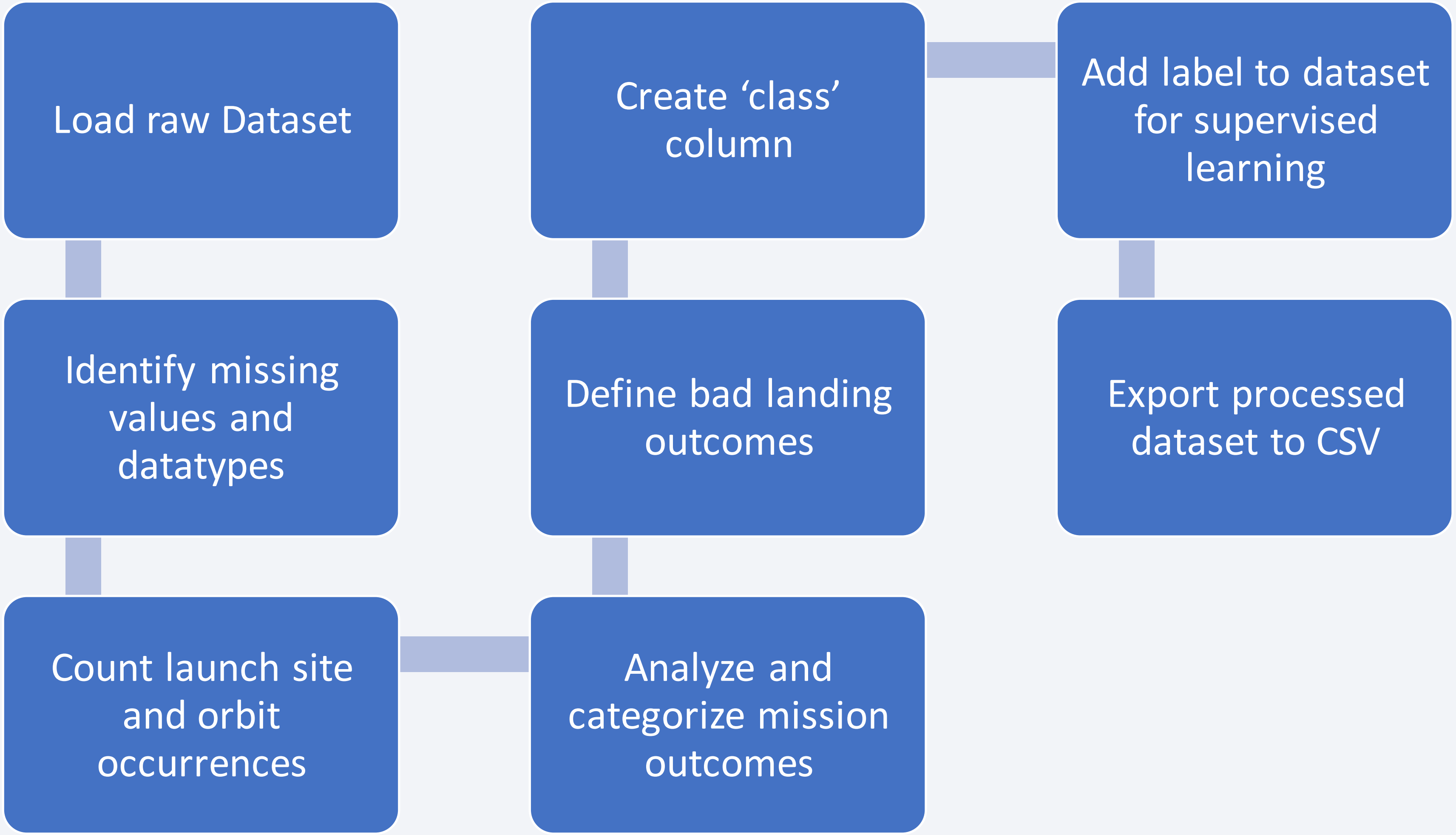
[GitHub Link]



Data Wrangling

Data Cleaning: Removed rows with multiple payloads/cores and handled null values. Column Normalization: Extracted nested fields (e.g., payloads, cores, rockets) using API responses. Date Formatting: Converted UTC timestamp to clean date-only format. Filtering: Removed Falcon 1 entries to isolate Falcon 9 launches. Payload Mass Imputation: Filled missing payload values with the column mean. Final Dataset: Combined all features into a structured DataFrame for analysis.

GitHub Link



EDA with Data Visualization

- **Bar Chart of Launches per Site**
  Used to visualize the distribution of SpaceX launches across different sites and identify the most active launch location.
- **Pie Chart of Orbit Types**
  Displayed the proportion of launches targeting each orbit, helping assess mission diversity.
- **Histogram of Payload Mass**
  Illustrated the distribution of payload weights to detect outliers and understand common payload sizes.
- **Outcome Frequency Bar Plot**
  Showed the counts of various landing outcomes (e.g., True ASDS, False Ocean), critical for labeling success and failure.
- **Heatmap of Feature Correlation**
  Revealed how numerical features relate to each other and to the success label, aiding feature selection for modeling.

GitHub Link

EDA with SQL

- Extracted unique launch sites using DISTINCT for site overview.
- Filtered records to launch sites beginning with 'CCA' for pattern inspection.
- Calculated total payload mass delivered by NASA (CRS) missions.
- Computed average payload mass for F9 v1.1 boosters.
- Found the earliest date of a successful ground pad landing.
- Identified boosters with successful drone ship landings and payloads between 4000–6000 kg.
- Tallied successful vs. failed mission outcomes.
- Retrieved boosters with the maximum payload mass using a subquery.
- Filtered records with failure on drone ships in 2015, displaying month, site, and booster details.
- Ranked landing outcomes by count between 2010-06-04 and 2017-03-20 using ORDER BY.

GitHub Link

Build an Interactive Map with Folium

Summary of Map Objects in Folium Map

- Added Marker objects at each launch site to show exact coordinates and site names.
- Added Circle objects around markers to highlight the launch site zones visually.
- Drew Lines between launch sites and landing zones to indicate flight paths.
- Used Popup and Tooltip features on markers to show launch outcomes and booster versions for quick reference.

Reason for Adding These Objects

- Markers helped visually identify and differentiate SpaceX launch sites.
- Circles emphasized spatial context (e.g., ground pad vs. drone ship areas).
- Lines illustrated trajectory possibilities and landing routes.
- Popups/Tooltips improved interactivity and presented extra info without clutter.

GitHub Link

Build a Dashboard with Plotly Dash

Summary of Dashboard Plots and Interactions

- Launch Success Pie Chart shows proportion of successful vs. failed landings across all sites.
- Payload vs. Success Scatter Plot visualizes how payload mass influences landing success, with color indicating booster version.
- Launch Site Dropdown allows users to filter charts by specific launch site.
- Payload Range Slider lets users adjust the payload range to explore relationships interactively.

Why These Elements Were Added

- The pie chart provides a quick visual summary of overall mission outcomes.
- The scatter plot helps reveal correlations between payload weight and landing success.
- Dropdown and slider interactivity allows focused exploration by launch site and payload range, improving analytical insights.

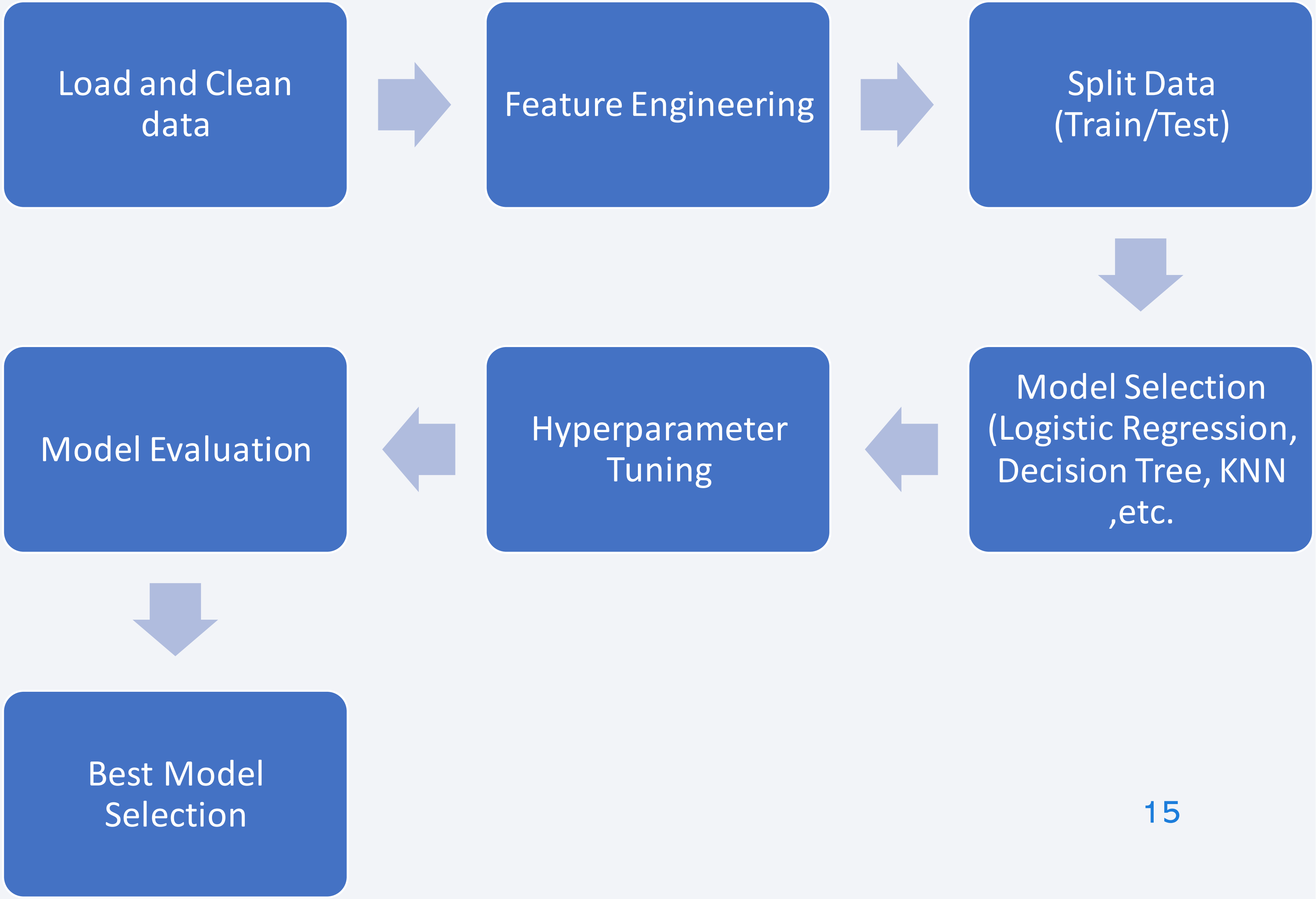
[GitHub Link]

Predictive Analysis (Classification)

Key Points

- Data was preprocessed and labeled for binary classification (success/failure).Multiple models were trained and evaluated.
- The best-performing model was selected based on accuracy and generalization.
- Grid search was used for parameter tuning.
- Results were validated with metrics like precision, recall, and F1 score.

[GitHub Link]



Results

Exploratory Data Analysis Results

- Identified most frequent launch sites, with KSC LC 39A having the highest launch count.
- Determined orbit distribution, showing LEO and GTO as the most common.
- Calculated landing success rate; approximately 60% of landings were successful.
- Created binary classification labels from mission outcomes for model training.

Interactive Analytics Demo in Screenshots

- Displayed launch locations using Folium maps with markers and labels.
- Differentiated successful and failed landings visually.
- Provided geographic insights into launch and landing patterns.

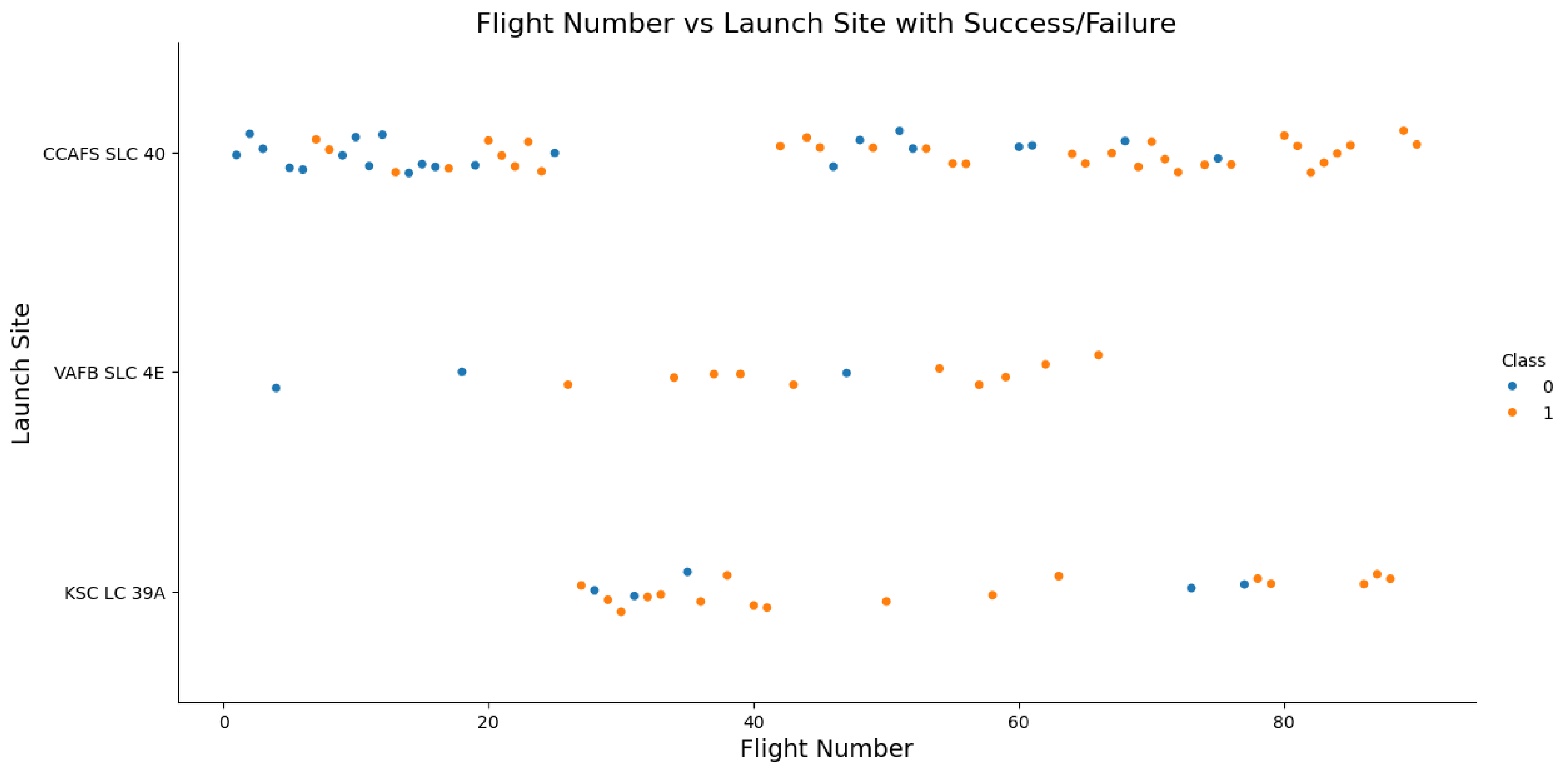
Predictive Analysis Results

- Built and evaluated multiple classification models.
- Random Forest classifier achieved the best performance based on F1 score.
- Feature importance revealed that Payload Mass and Launch Site were key predictors.
- Final model capable of predicting landing outcomes with strong accuracy.

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site



Now try to explain the patterns you found in the Flight Number vs. Launch Site scatter point plots.

Launch Site Distribution:

CCAFS SLC 40 has the highest number of launches, spread throughout the timeline (early to late flight numbers).

KSC LC 39A appears mainly in mid-to-late flight numbers.

VAFB SLC 4E has relatively fewer launches, scattered across the timeline.

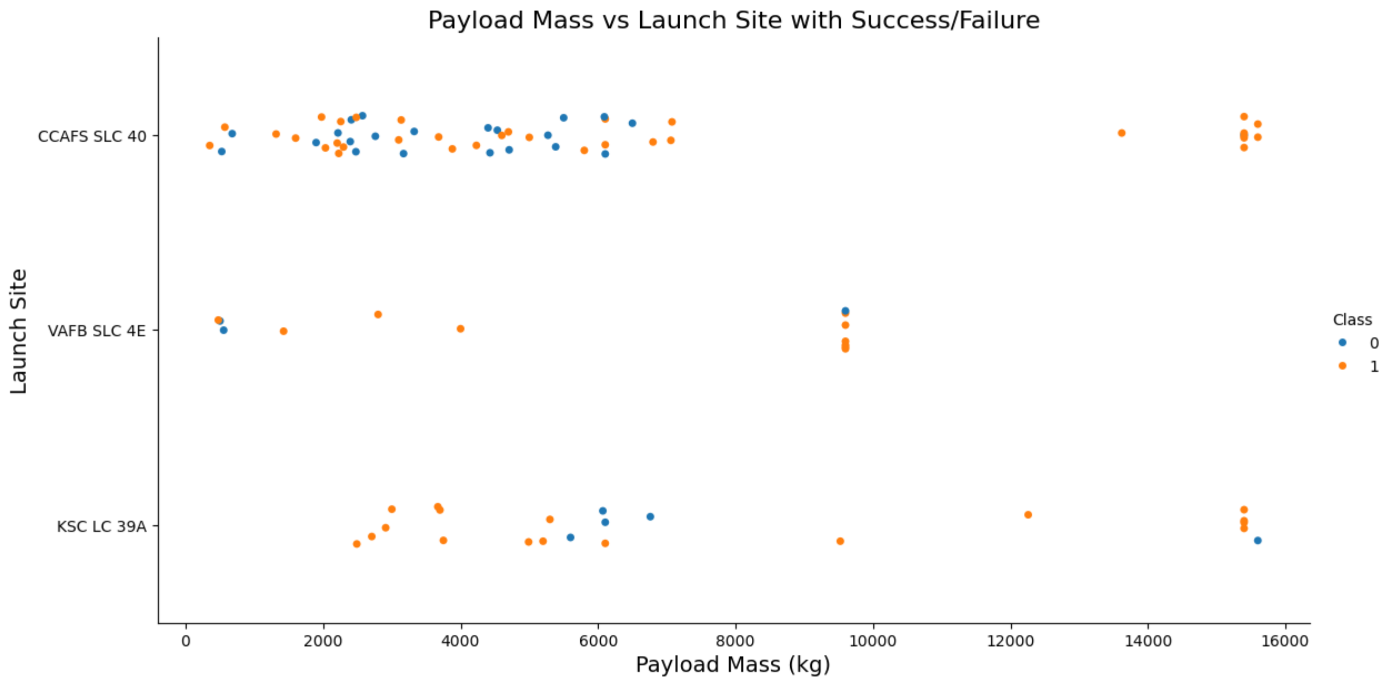
Outcome Trends:

Early flights across all sites have more failures (blue points), especially on CCAFS SLC 40, which likely served as the testing ground.

As flight numbers increase, successes (orange) become dominant, suggesting improvement in reliability over time.

KSC LC 39A shows a strong pattern of successful launches, indicating it may be used more for established, lower-risk missions.

Payload vs. Launch Site



Now if you observe Payload Mass Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Key Patterns and Insights:

1. CCAFS SLC 40:

This site has the most launches across a wide payload range (from near 0 up to ~15,000 kg). It shows a mix of successes and failures across all payload sizes, but successes are more common as payload mass increases — especially beyond 6,000 kg. This suggests early payload tests or missions with lower mass were riskier or experimental.

2. KSC LC 39A:

Most launches from this site have higher payloads (3,000 to 15,000 kg). Nearly all of them are successful, indicating it's likely used for well-established, high-confidence missions. A small number of failures occurred, but they are minimal compared to successes.

3. VAFB SLC 4E:

Very few launches from this site, scattered across different payloads. Appears to have a moderate mix of success and failure, but with limited data it's hard to derive strong trends.

Success Rate vs. Orbit Type

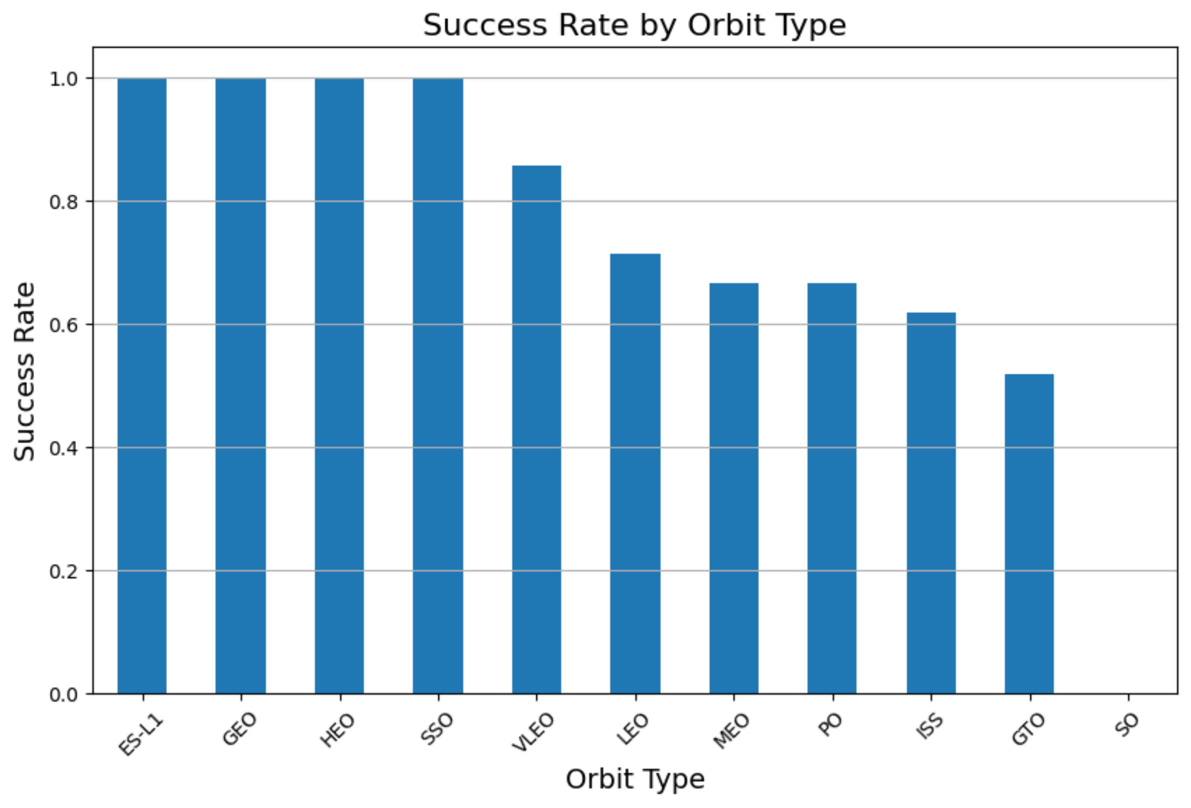
Analyze the plotted bar chart to identify which orbits have the highest success rates.
💡

Orbits with the Highest Success Rates (100%)
The following orbits have a perfect success rate (1.0):

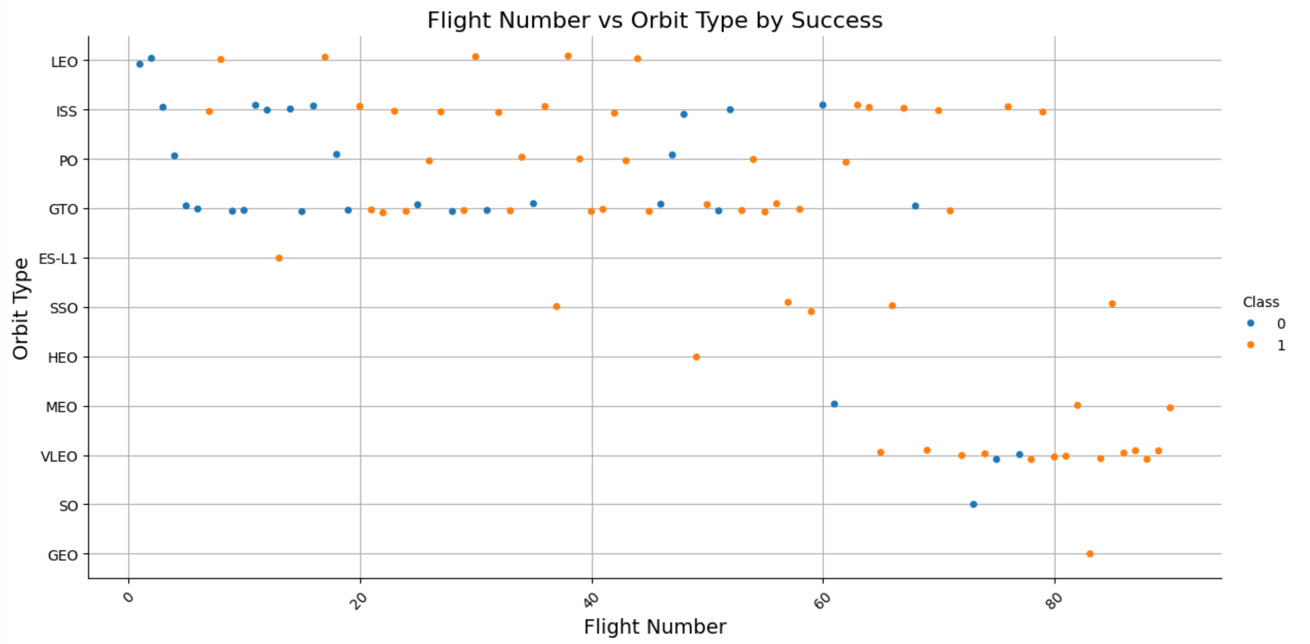
ES-L1 (Earth-Sun Lagrange Point 1)
GEO (Geostationary Earth Orbit)
HEO (Highly Elliptical Orbit)
SSO (Sun-Synchronous Orbit)
These orbits had no recorded failures in the dataset, indicating either:
Very few, but highly controlled missions or a mature and reliable launch process for these orbit types

Moderate Success Rates
VLEO (Very Low Earth Orbit): approximately 85%
LEO (Low Earth Orbit): approximately 71%
MEO (Medium Earth Orbit) and PO (Polar Orbit): approximately 67%
These orbits are more frequently used, especially LEO, and naturally carry more operational variability.

Lowest Success Rates
ISS (International Space Station orbit): approximately 62%
GTO (Geostationary Transfer Orbit): approximately 52%
SO (Solar Orbit): 0% success rate



Flight Number vs. Orbit Type



You can observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

LEO (Low Earth Orbit): Most frequently targeted orbit. A mix of early and later flights. Later flights (higher flight numbers) have mostly successful outcomes, indicating improved reliability over time.

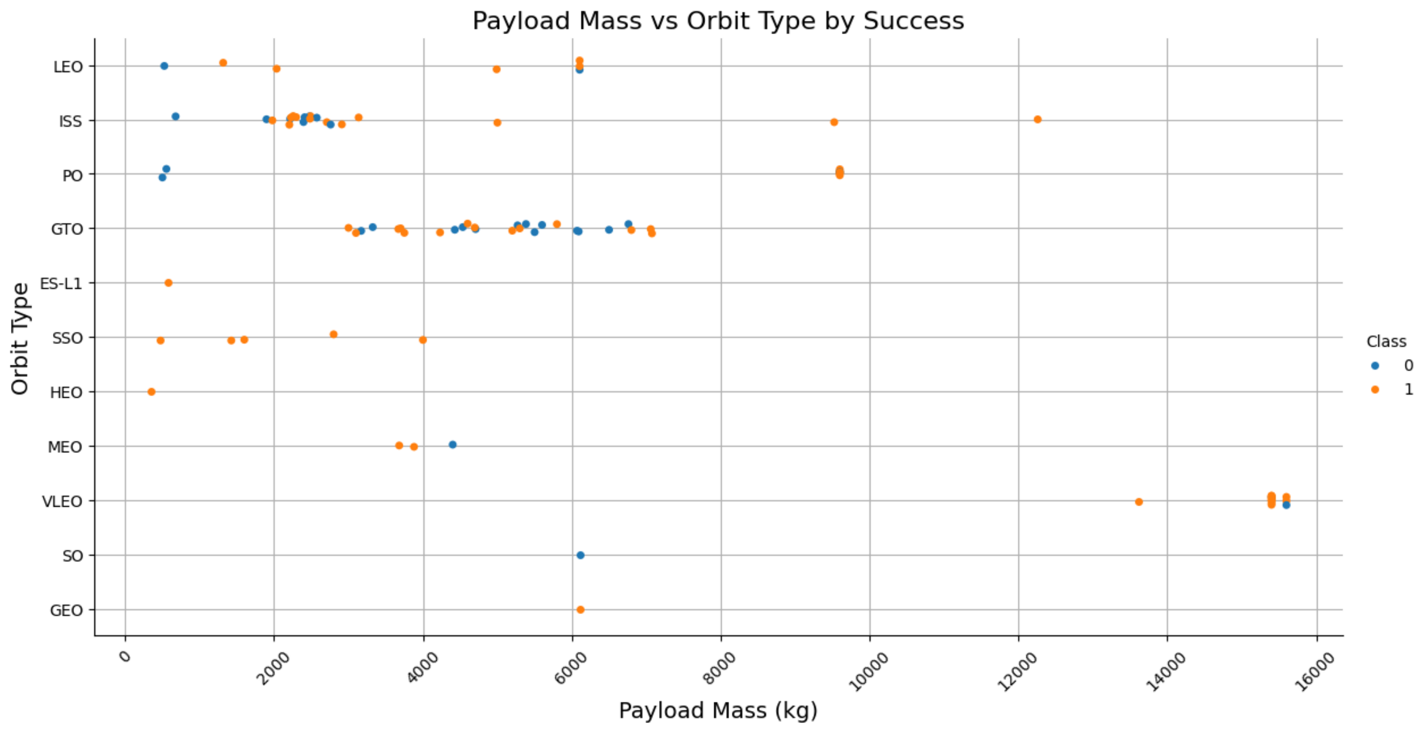
VLEO (Very Low Earth Orbit): Concentrated in the later flight numbers. Majority of launches here are successful. Shows confidence in this orbit in more recent missions.

GTO (Geostationary Transfer Orbit): Spread across mid-range flight numbers. Noticeable number of failures in early GTO missions. Improvement over time can be inferred as later GTO launches are more successful.

ISS (International Space Station): Mostly successful, even from earlier flights. Suggests high mission reliability due to established protocols.

SSO, GEO, HEO, MEO, PO, ES-L1, SO: Fewer missions overall. Nearly all are successful, which may reflect either better mission planning or fewer but more critical launches.

Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

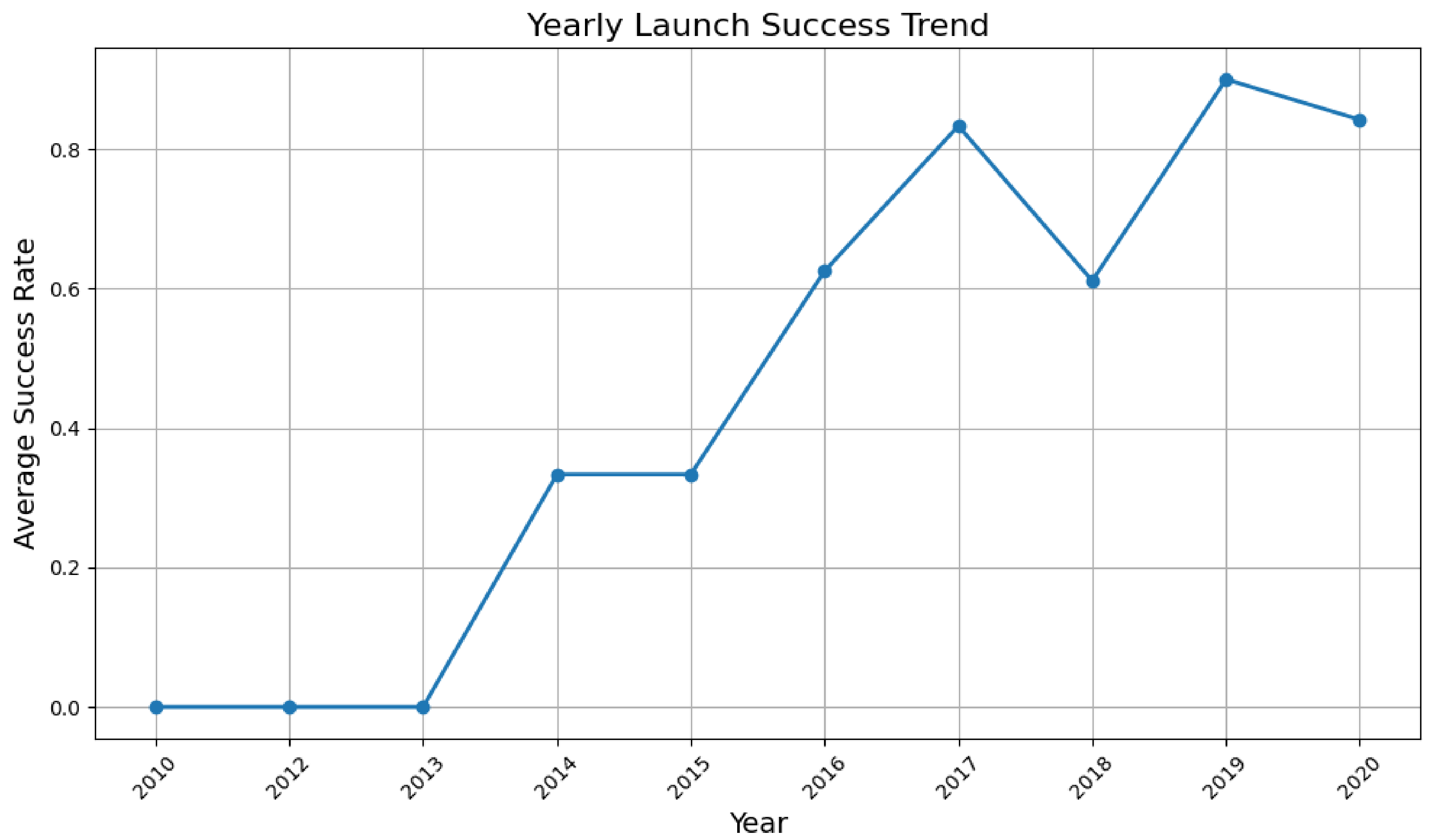
However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Higher Payload Mass with Positive Outcomes: For LEO, ISS, and Polar Orbits (PO), missions with larger payloads tend to result in successful landings more consistently. This suggests that these orbits are better optimized for heavy payload deliveries and that SpaceX has achieved strong reliability here.

GTO (Geostationary Transfer Orbit): Despite frequent usage, the success and failure outcomes are mixed across all payload sizes. Indicates that GTO missions carry higher technical risk, likely due to the altitude and complexity of the transfer orbit.

VLEO and SSO: Appear toward the lighter-to-mid payload range and also demonstrate a high success rate, though with fewer data points.

Launch Success Yearly Trend



you can observe that the sucess rate since 2013 kept increasing till 2020

Yearly Launch Success Trend Analysis From 2010 to 2013, the success rate remained at 0, indicating early development and testing phases. Beginning 2014, SpaceX started achieving successful missions, with a consistent upward trend in success rates. The highest average success rate was recorded in 2019, nearing 90%, reflecting significant operational improvements. Although there's a slight dip in 2020, the overall trajectory from 2013 to 2020 demonstrates clear growth in reliability and performance.

All Launch Site Names

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

This query is used to retrieve all unique launch site names from the SPACEXTABLE. The DISTINCT keyword ensures that duplicate entries are filtered out, so each launch site appears only once in the result.

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

Launch Site Names Begin with 'CCA'

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

- This query retrieves up to 5 rows from the SPACEXTABLE where the Launch_Site name starts with 'CCA'.
- %%sql: Indicates that the query is executed in a Jupyter notebook cell using a SQL magic command.
- SELECT *: Selects all columns from the matching rows.
- WHERE "Launch_Site" LIKE 'CCA%': Filters the results to only include rows where the Launch_Site column begins with 'CCA'. The % symbol is a wildcard matching any sequence of characters.
- LIMIT 5: Returns only the first 5 matching rows.

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

Total Payload Mass

```
%%sql
SELECT SUM("PAYLOAD_MASS__KG_") AS TotalPayloadMass
FROM SPACEXTABLE
WHERE "Customer" = 'NASA (CRS)';
```

- SUM("PAYLOAD_MASS__KG_"): Adds up all the payload masses.
- AS TotalPayloadMass: Names the result column TotalPayloadMass.
- WHERE "Customer" = 'NASA (CRS)': Filters records to only include those where the customer is NASA's Commercial Resupply Services (CRS) program.

TotalPayloadMass
45596

Average Payload Mass by F9 v1.1

```
%%sql
SELECT AVG("PAYLOAD_MASS__KG_") AS AvgPayloadMass
FROM SPACEXTABLE
WHERE "Booster_Version" = 'F9 v1.1';
```

- AVG("PAYLOAD_MASS__KG_"): Computes the mean of the payload masses.
- AS AvgPayloadMass: Labels the result as AvgPayloadMass.
- WHERE "Booster_Version" = 'F9 v1.1': Filters the data to only include records using the F9 v1.1 booster.

AvgPayloadMass
2928.4

First Successful Ground Landing Date

```
%%sql

SELECT MIN("Date") AS FirstSuccessfulGroundPadLanding
FROM SPACEXTABLE
WHERE "Landing_Outcome" = 'Success (ground pad)';
```

- MIN("Date"): Gets the earliest (minimum) launch date.
- AS FirstSuccessfulGroundPadLanding: Names the result for clarity.
- WHERE "Landing_Outcome" = 'Success (ground pad)': Filters to only include successful ground pad landings.

FirstSuccessfulGroundPadLanding
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql

SELECT DISTINCT Booster_Version
FROM SPACEXTABLE
WHERE Landing_Outcome LIKE '%Success (drone ship)%'
  AND PAYLOAD_MASS__KG_ > 4000
  AND PAYLOAD_MASS__KG_ < 6000;
```

- Successfully landed on a drone ship (Landing_Outcome LIKE '%Success (drone ship)%’)
- Carried a payload mass between 4000 and 6000 kg

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

Total Number of Successful and Failure Mission Outcomes

```
%%sql

SELECT Mission_Outcome, COUNT(*) AS OutcomeCount
FROM SPACEXTABLE
GROUP BY Mission_Outcome;
```

- It helps summarize the distribution of different mission results (e.g., success, failure) in the dataset.

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

Boosters Carried Maximum Payload

```
%%sql

SELECT Booster_Version, PAYLOAD_MASS__KG_
FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ = (
    SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEXTABLE
);
```

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

- It identifies the record with the maximum payload mass in the dataset.

2015 Launch Records

```
%%sql

SELECT
    substr(Date, 6, 2) AS Month,
    Landing_Outcome,
    Booster_Version,
    Launch_Site
FROM SPACEXTABLE
WHERE substr(Date, 0, 5) = '2015'
  AND Landing_Outcome LIKE '%Failure%'
  AND Landing_Outcome LIKE '%drone ship%';
```

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

- It filters records from the year 2015 where the "Landing_Outcome" includes both 'failure' and 'drone ship'

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

- It groups records by "Landing_Outcome", counts occurrences for each, and orders them from most to least frequent (DESC).

```
%%sql

SELECT
    Landing_Outcome,
    COUNT(*) AS OutcomeCount
FROM SPACEXTABLE
WHERE Date >= '2010-06-04' AND Date <= '2017-03-20'
GROUP BY Landing_Outcome
ORDER BY OutcomeCount DESC;
```

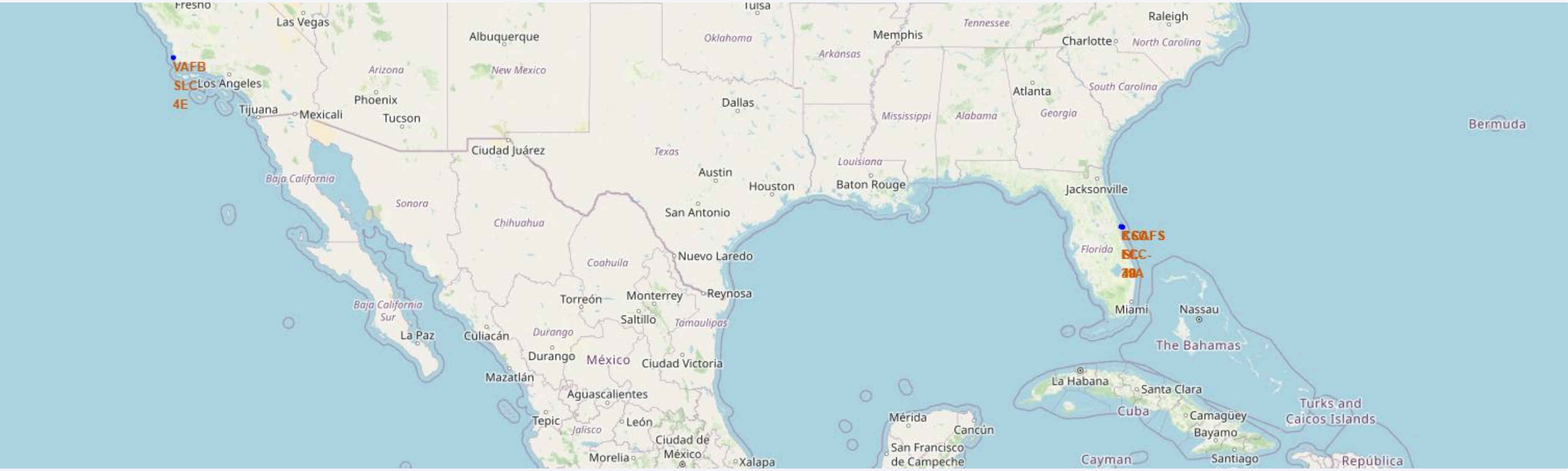
Landing_Outcome	OutcomeCount
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

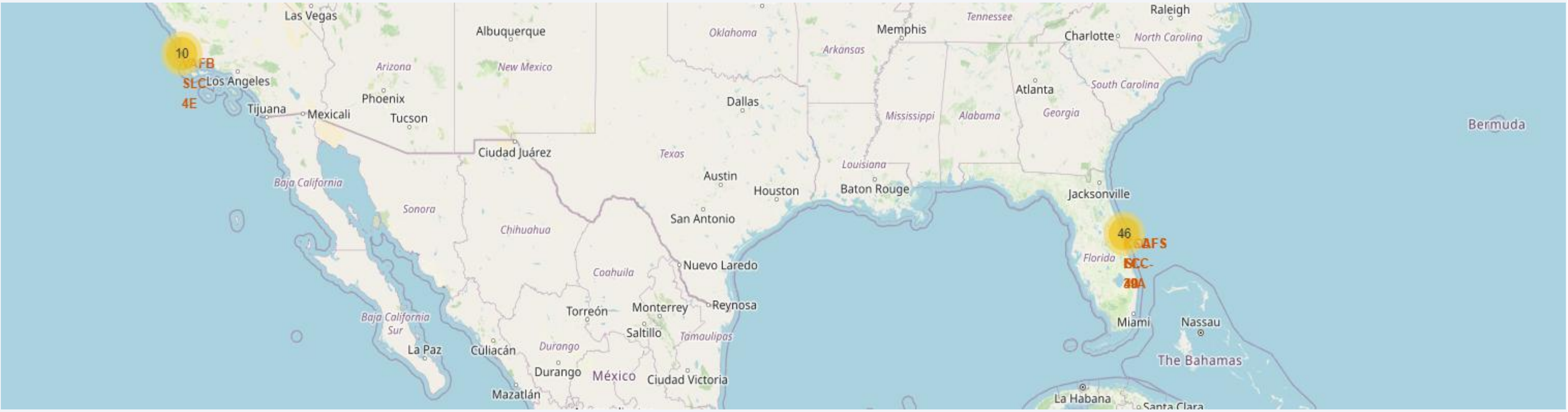
Global Launch Sites Visualization Using Folium

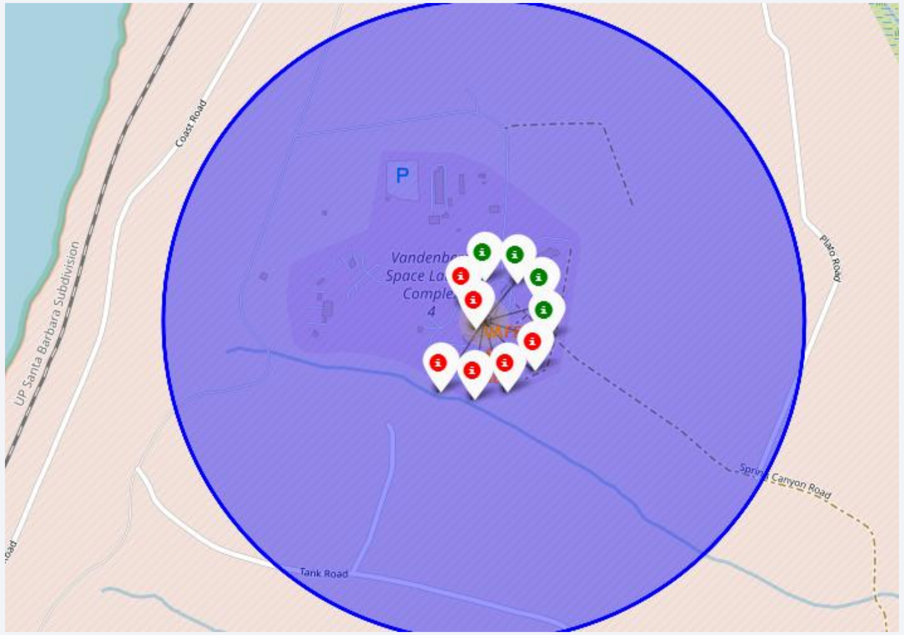
- The Folium map displays the geographic positions of all unique SpaceX launch sites.
- Each site is marked with a location pin that helps visually identify spatial distribution.
- Launch sites shown may include:
- CCAFS LC-40 (Cape Canaveral)
- VAFB SLC-4E (California)
- KSC LC-39A (Kennedy Space Center)
- This map allows viewers to understand launch coverage across the U.S. and helps assess the proximity of launch infrastructure to oceans and populated regions.
- Such maps are useful for logistics, planning future missions, or analyzing launch site performance geographically.



Global Launch Site Locations

- This map shows the geographical locations of all launch sites used in the dataset. The markers highlight three key launch sites:
- CCAFS LC-40
- KSC LC-39A
- VAFB SLC-4E
- Clusters of launches are visible in Florida and California, which are major hubs for U.S. space missions. The marker count represents the frequency of launches from each site.





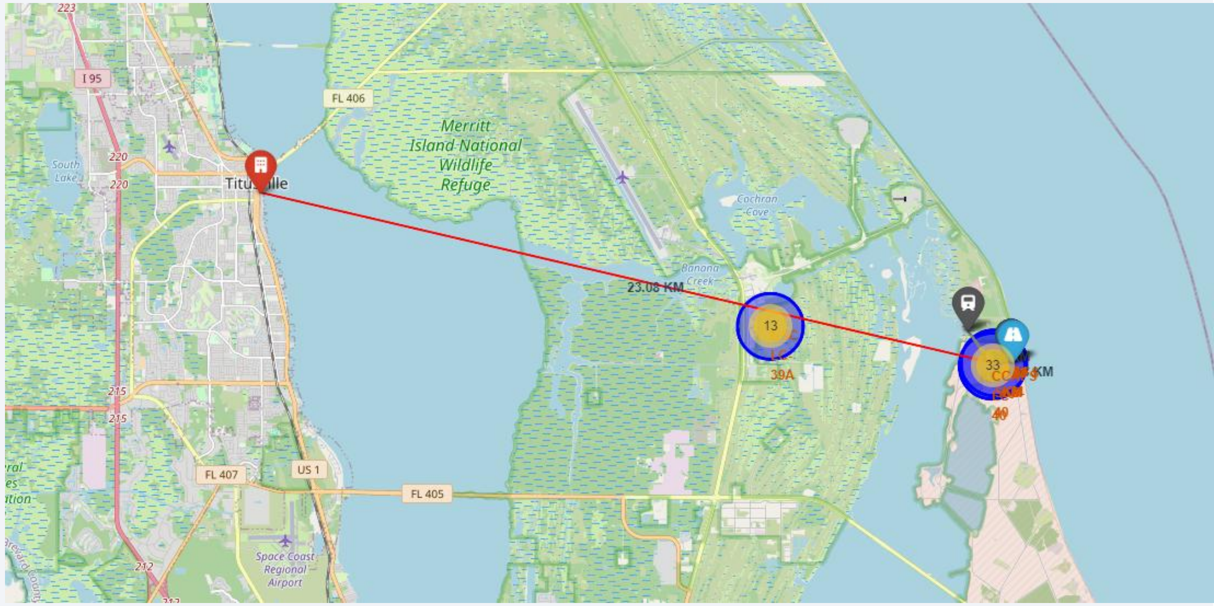
Launch Site Proximity Analysis Using Folium Map

Key Elements:

- Markers: Each marker identifies a SpaceX launch site, with clustering to indicate activity level.
- Distance Line: A geodesic line visually shows proximity to urban support infrastructure.
- Color-coded Labels: These help distinguish launch sites and enhance readability.

Findings:

- The KSC LC-39A site is strategically located near transport infrastructure and coastlines, crucial for logistics and safety.
- The map demonstrates how launch infrastructure is tightly integrated with geographic and civic planning.

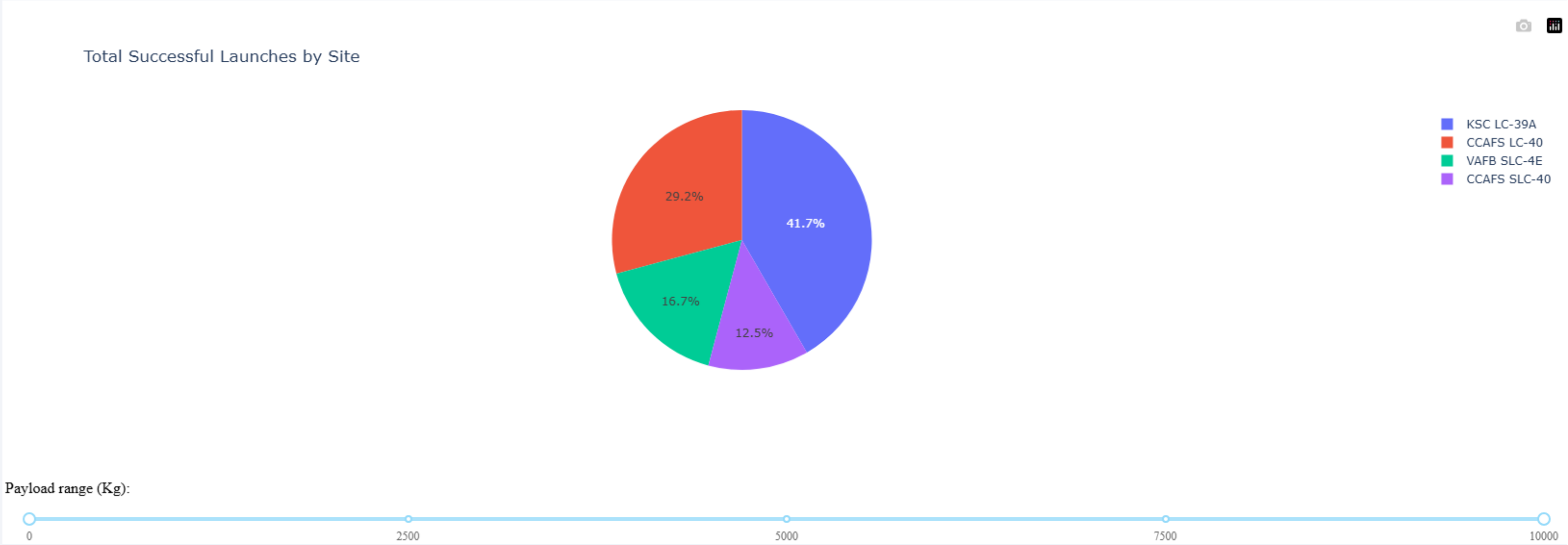


Section 4

Build a Dashboard with Plotly Dash

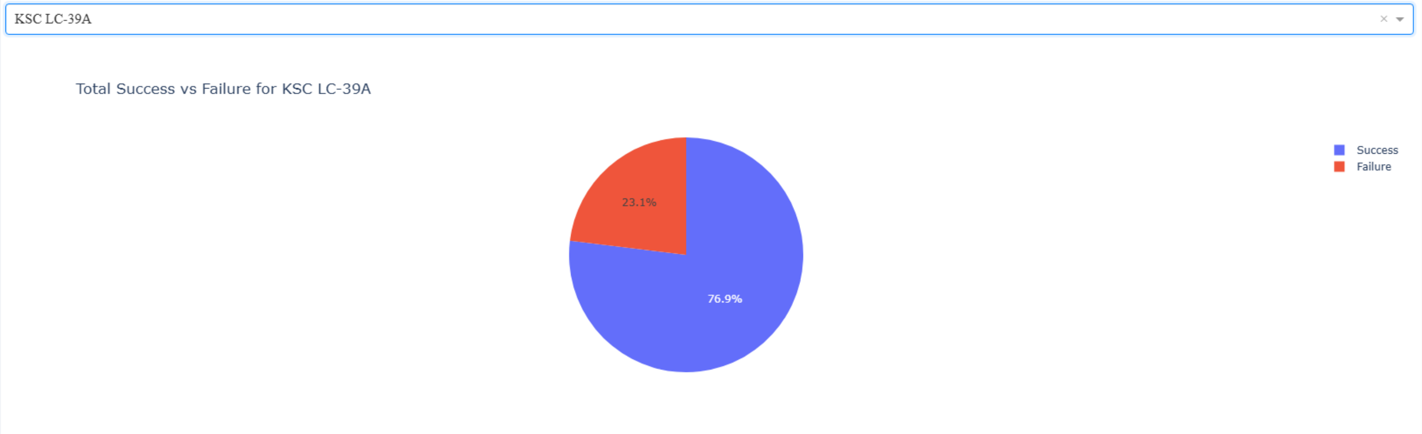
Launch Success Distribution by Site

- From a launch map: Multiple launch sites are concentrated in Florida, such as CCAFS LC-40 and KSC LC-39A, showing it's a central hub for missions.
- From a pie chart: KSC LC-39A contributes to the largest portion of successful launches, implying reliability or preference.
- From a distance map: The launch site's proximity (e.g., 23.08 km) to infrastructure like cities or highways is important for logistics and safety.
- From a mission outcome chart: Most frequent outcomes indicate dominant success patterns or highlight operational improvements over time.



Launch Outcome Distribution at KSC LC-39A

- Selected Site: The pie chart focuses on Kennedy Space Center Launch Complex 39A (KSC LC-39A), one of the major SpaceX launch facilities.
- Success Rate: Approximately 76.9% of launches from this site were successful, indicating a strong performance record.
- Failure Rate: Around 23.1% of launches failed, which could inform operational reviews or be tied to earlier phases of SpaceX's development.
- Insight: While the success rate is relatively high, this site has seen a higher failure ratio compared to other top-performing launch complexes. This could be influenced by experimental missions or early launches in SpaceX's timeline.



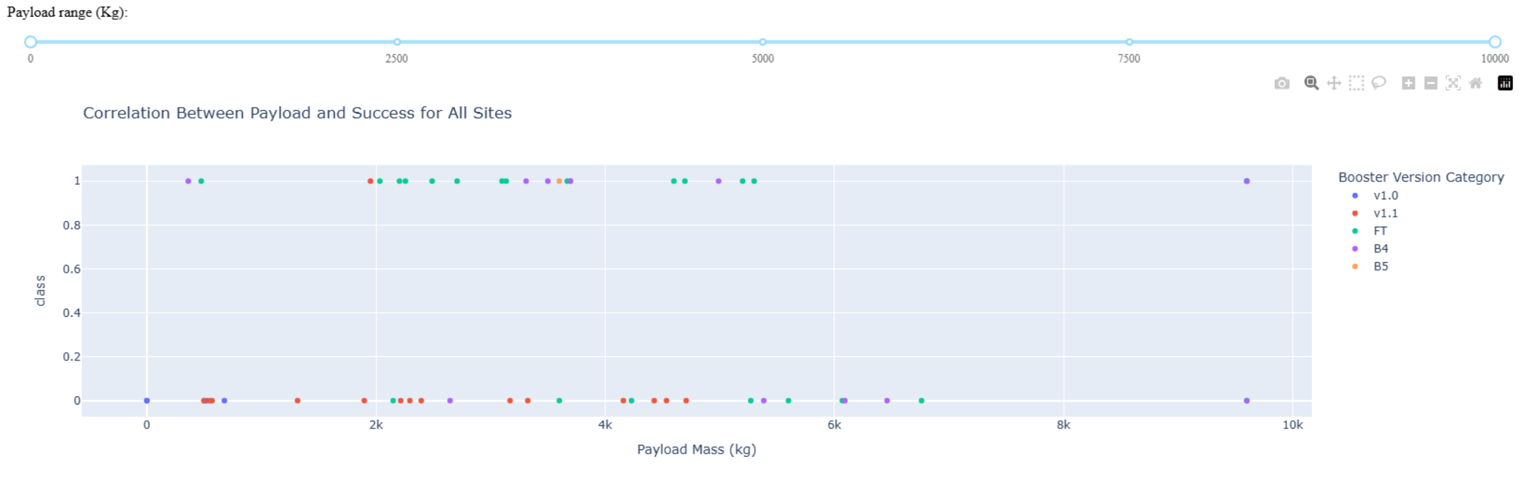
Payload vs Launch Outcome Insights

Scatter Plot Description:

- This chart shows the correlation between payload mass and launch success (1 = Success, 0 = Failure), color-coded by booster version categories such as v1.0, v1.1, FT, B4, and B5.
- Range Slider Use: The top slider lets users filter payload ranges (0–10,000 kg), helping identify performance trends within specific payload limits.

Findings:

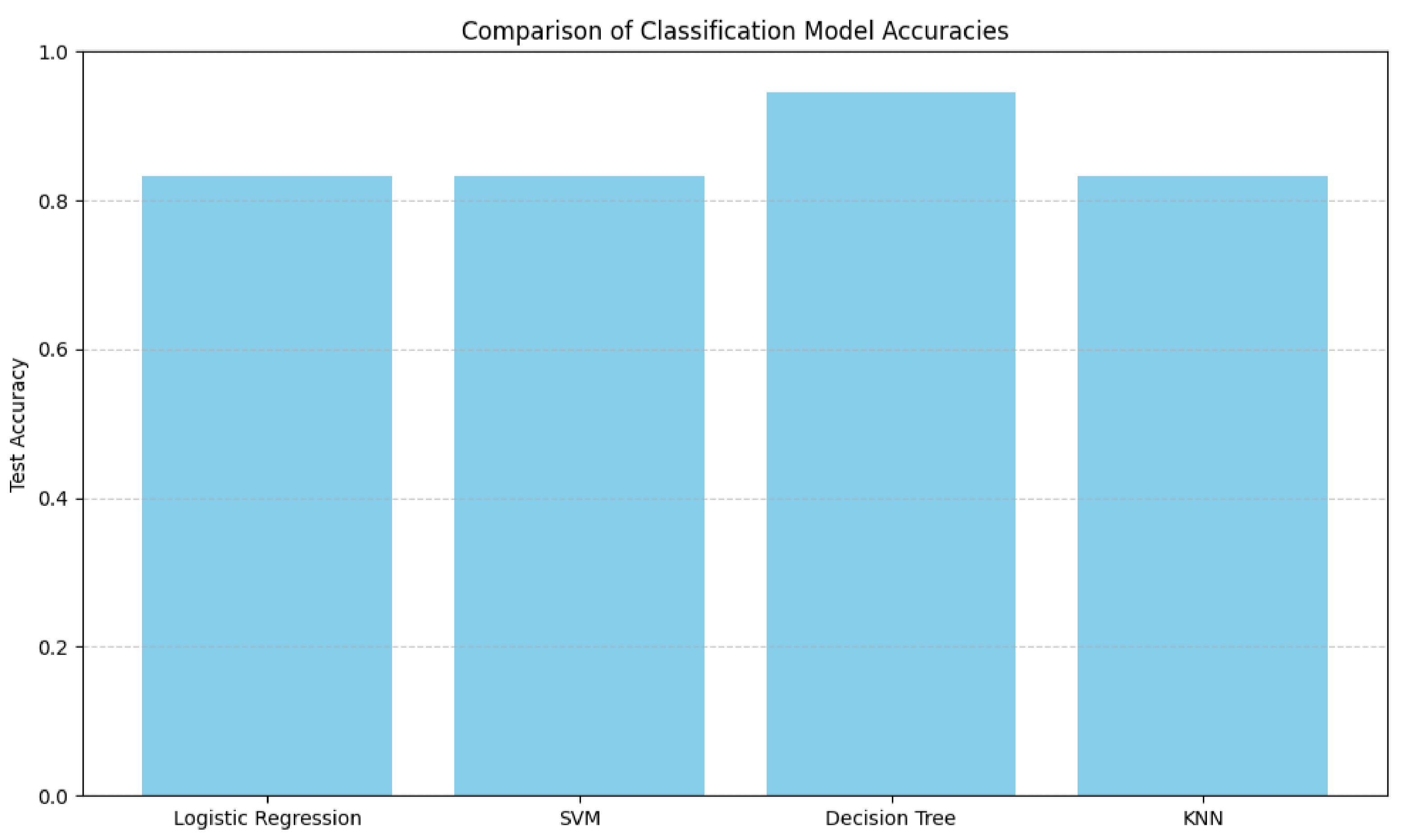
- Most successful launches cluster around payloads between 2000–6000 kg.
- Newer booster versions like FT and B5 generally show higher success rates compared to earlier versions (v1.0 and v1.1).
- The scatter indicates that payload mass does not strongly predict success alone, but combined with booster version data, clear reliability patterns emerge.



Section 5

Predictive Analysis (Classification)

Classification Accuracy



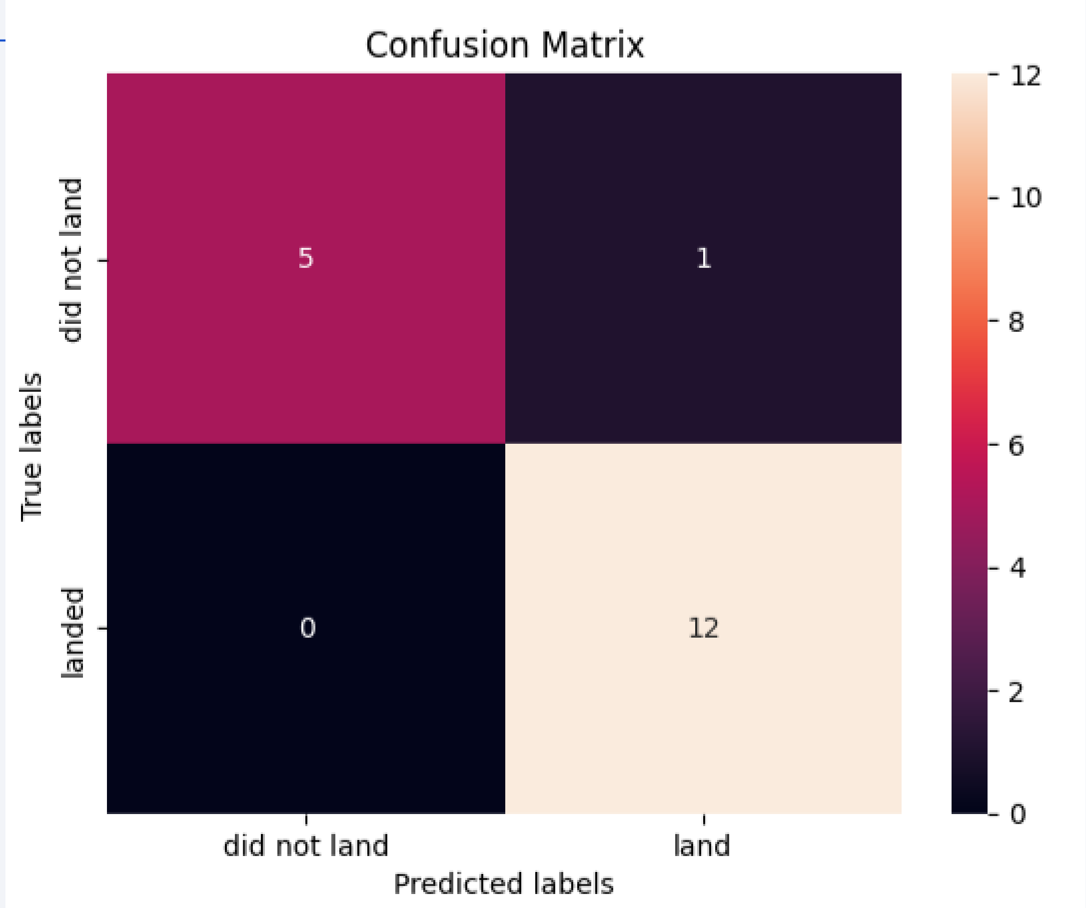

```
Logistic Regression Test Accuracy: 0.8333333333333334
SVM Test Accuracy: 0.8333333333333334
Decision Tree Test Accuracy: 0.9444444444444444
KNN Test Accuracy: 0.8333333333333334
```

Confusion Matrix

- True Positives (Landed, predicted as landed): 12
- True Negatives (Did not land, predicted as did not land): 5
- False Positives (Predicted landed, but didn't): 1
- False Negatives: 0

Key Insights:

- The Decision Tree classifier showed the highest accuracy.
- The confusion matrix confirms reliable predictions, with only 1 misclassified case out of 18.
- No False Negatives: All successful landings were correctly predicted, which is crucial for cost-saving estimation in SpaceX's business context.
- This model may be particularly valuable for operational decisions where minimizing missed successful landings is a priority.



Conclusions

- Performed classification to predict Falcon 9 first-stage landing success using historical mission data
- Built and evaluated four models: Logistic Regression, SVM, Decision Tree, and KNNDecision Tree model showed the highest test accuracy of 94.4%, outperforming others
- Confusion matrix for Decision Tree confirmed strong predictive performance with only 1 false positive
- Key influencing factors included payload mass, booster version, and launch site
- Predictive modeling can assist stakeholders in estimating launch cost savings and risk

Thank you!