IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Livia Gibertoni
08/09/2025

Outline

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

Executive Summary

- Summary of methodologies: Collected and cleaned SpaceX launch data, performed exploratory data analysis, engineered relevant features, applied data wrangling, and built/tuned classification models (Logistic Regression, SVM, Decision Tree, KNN) using GridSearchCV with cross-validation.
- Summary of all results: Decision Tree achieved the highest cross-validation accuracy (87.68%), while Logistic Regression and SVM had the best test accuracy (83.33%). KNN performed lowest among the tested models.

Introduction

- Project background and context: This project analyzes SpaceX launch data to identify factors influencing mission success and to predict whether a launch will be successful based on historical patterns.
- Problems you want to find answers: What factors most influence the success of a SpaceX launch? Can we accurately predict the outcome of future launches using machine learning models?

Section 1

Methodology

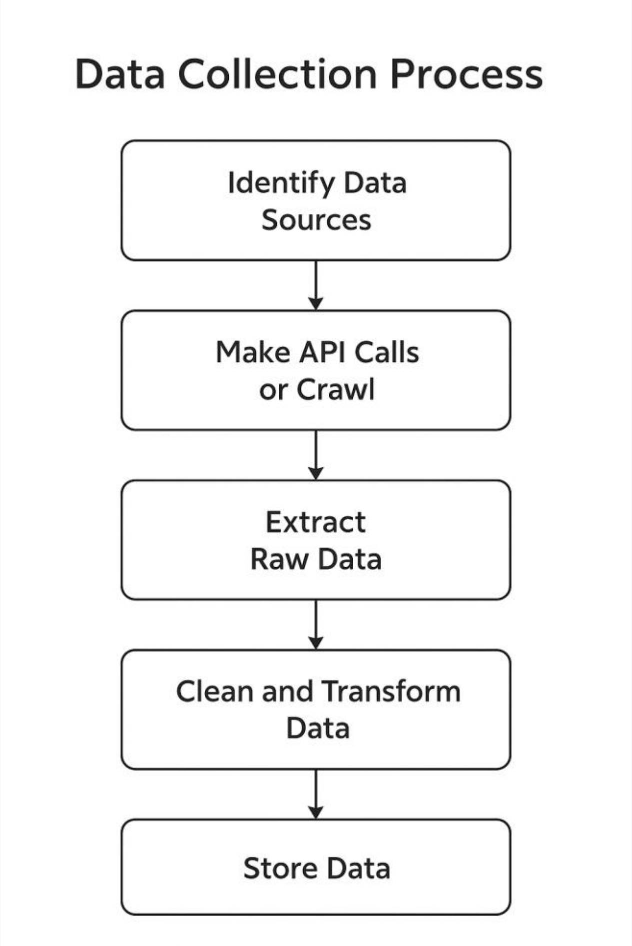
Methodology

Executive Summary

- Data collection methodology:
    - Data was obtained from SpaceX public launch records via web scraping and API, combined with the Kaggle “SpaceX Launch Records” dataset. Key features include launch site, payload mass, orbit, and launch outcome. Data was cleaned, standardized, and stored in CSV format.
- Perform data wrangling
    - Cleaned and transformed the dataset by handling missing values, encoding categorical data, and scaling numerical features to prepare for analysis.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Built multiple classification models (Logistic Regression, SVM, Decision Tree, KNN), tuned hyperparameters with GridSearchCV, and evaluated performance using cross-validation and test set accuracy.

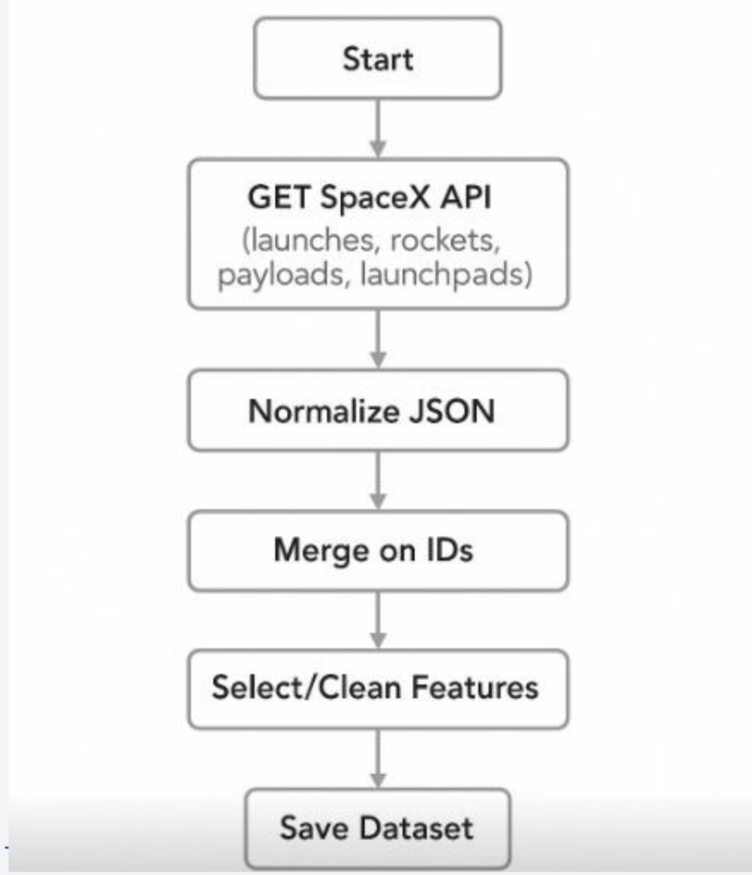
Data Collection

- Sources:
- SpaceX REST API: Launch details (date, site, payload, outcome)
- Web Scraping: SpaceX mission pages & tables for extra details
- Tools & Methods:
- REST calls with requests → JSON → Pandas DataFrame
- BeautifulSoup & pandas.read_html for HTML table extraction
- Rate limiting, User-Agent, robots.txt check for ethical scraping
- Data cleaning: remove whitespace, parse dates, normalize text/numbers
- Output:
- Unified, structured DataFrame saved to CSV with timestamp & source URLs



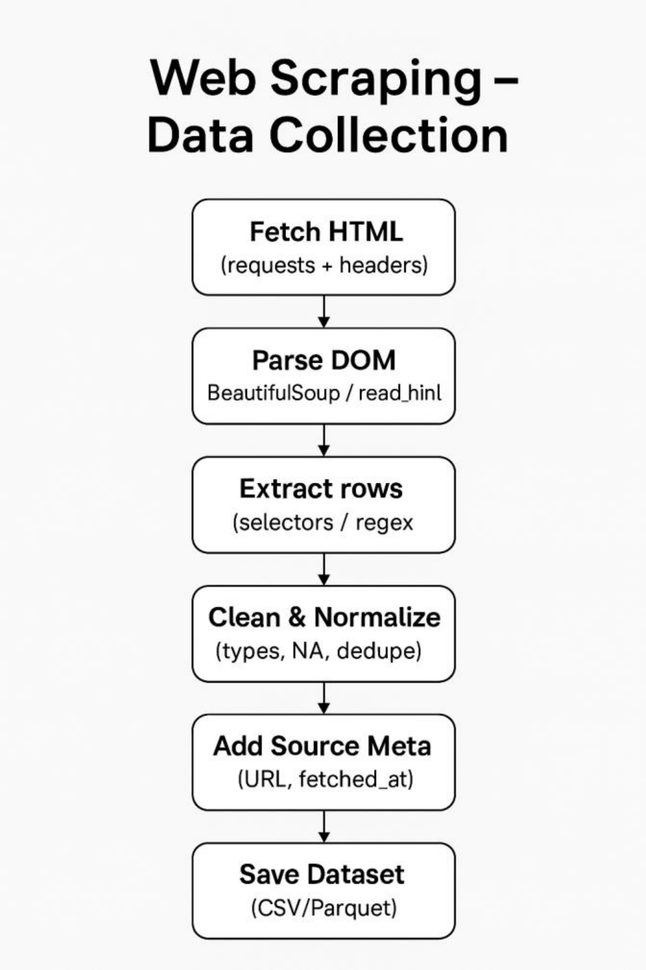
Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts
- **Source:** SpaceX public REST API (no auth), JSON responses
- **Base URL:** https://api.spacexdata.com/v4/
- **Endpoints used:** launches, rockets, payloads, launchpads, cores
- **Method:** GET; join on IDs returned in launches (e.g., rocket, payloads[], launchpad, cores[])
- **Process:** request → normalize JSON → merge by keys → select fields → export CSV/Parquet
- **Outputs:** features like launch site, date, payload mass, orbit, booster/serial, outcome (success flag)
- Capstone/jupyter-labs-spacex-data-collection-api.ipynb at main · liviagibertoni/Capstone



Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts
  - **Targets:** SpaceX launch pages + supporting tables (mission, date, site, payload, outcome)
  - **Tools:** requests, BeautifulSoup, pandas.read_html, optional time/random for polite delays
  - **Selectors:** HTML tags/attributes (e.g., table, tr, td, a[href], CSS classes)
  - **Politeness:** Check **robots.txt**, set **User-Agent**, add **rate limiting** (sleep/jitter), avoid heavy concurrency
  - **Cleaning:** strip whitespace, parse dates, convert numbers, handle missing values
  - **Outputs:** normalized DataFrame → **CSV/Parquet**; include timestamp and source URL for traceability
- [Capstone/jupyter-labs-webscraping.ipynb at main · liviagibertoni/Capstone]()



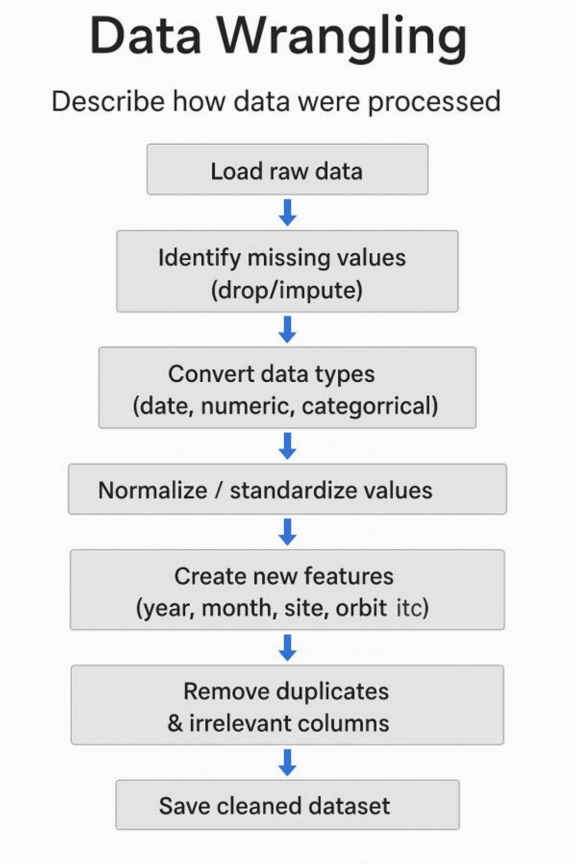
Data Wrangling

- **Data Wrangling**

- **Key steps:**

- **Load data from collected sources (API, web scraping, CSV files).**
- **Identify and handle missing values (drop or impute).**
- **Convert data types (dates, numeric, categorical).**
- **Normalize and standardize values where needed.**
- **Feature engineering (extract year, month, site name, orbit type).**
- **Remove duplicates and clean irrelevant columns.**
- **Save cleaned dataset for analysis.**

- [Capstone/labs-jupyter-spacex-Data wrangling.ipynb at main · liviagibertoni/Capstone]()



EDA with Data Visualization

- **Catplot to visualize the relationship between Flight Number and Payload – helps identify patterns between mission size and experience level.**
- **Catplot to visualize the relationship between Flight Number and Launch Site – reveals site-specific trends over time.**
- **Catplot to visualize the relationship between Payload and Launch Site – explores how payload capacity varies by site.**
- **Bar chart to visualize the success rate for each Orbit type – compares performance across mission profiles.**
- **Catplot to visualize the relationship between Flight Number and Orbit type – examines mission variety as launch experience grows.**
- **Catplot to visualize the relationship between Payload and Orbit type – analyzes payload distribution across orbit categories.**
- **Line chart to visualize the yearly launch success trend – highlights improvement or decline in performance over time.**

- [Capstone/edadataviz.ipynb at main · liviagibertoni/Capstone]()

EDA with SQL

EDA with SQL – Summary of Performed Queries

- **Created a new table SPACEXTABLE filtering out records with null dates.**
- **Retrieved all unique launch sites in the dataset.**
- **Displayed the first 5 records where the launch site starts with "CCA".**
- **Calculated the total payload mass carried by missions with NASA (CRS) as the customer.**
- **Found the average payload mass for booster version F9 v1.1.**
- **Retrieved the earliest date of a successful landing on a ground pad.**
- **Listed unique booster versions with successful drone ship landings carrying payloads between 4000 kg and 6000 kg.**
- **Counted the number of launches for each mission outcome category.**
- **Identified the booster version and payload mass of the heaviest launch.**
- **Extracted month, landing outcome, booster version, and launch site for failed drone ship landings in 2015.**
- **Counted and sorted landing outcomes for launches between June 4, 2010, and March 20, 2017.**

- Capstone/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · liviagibertoni/Capstone

Build an Interactive Map with Folium

Map objects added & why

- **Base map (folium.Map) – centered on NASA/launch-site coordinates with appropriate zoom_start; provides the canvas for all overlays.**
- **Circles / CircleMarkers (folium.Circle, folium.CircleMarker) – draw a visible radius around each launch site to highlight location context and density of nearby points.**
- **Site markers (folium.Marker) – pinpoint each launch pad; popups/tooltips show site name and coordinates for quick identification.**
- **Outcome markers in a cluster (MarkerCluster) – add one marker per launch and group them to avoid clutter; markers are colored by class (e.g., green = success, red = failure) so you can scan performance at a glance.**
- **Text/label markers (folium.features.DivIcon) – annotate the map with computed values (e.g., distances) directly on the tiles for quick reading without opening popups.**
- **Polylines (folium.PolyLine) – draw lines from a selected launch site to reference features (coastline/railway/highway/city) to visualize proximity and the direction of shortest distance used in the analysis.**

Why these objects?
They make the map both navigable (clustered markers, clean popups) and analytical (color-coded outcomes, labeled distances, and lines that reveal spatial relationships to infrastructure and coastline).

- Capstone/lab_jupyter_launch_site_location.ipynb at main · liviagibertoni/Capstone

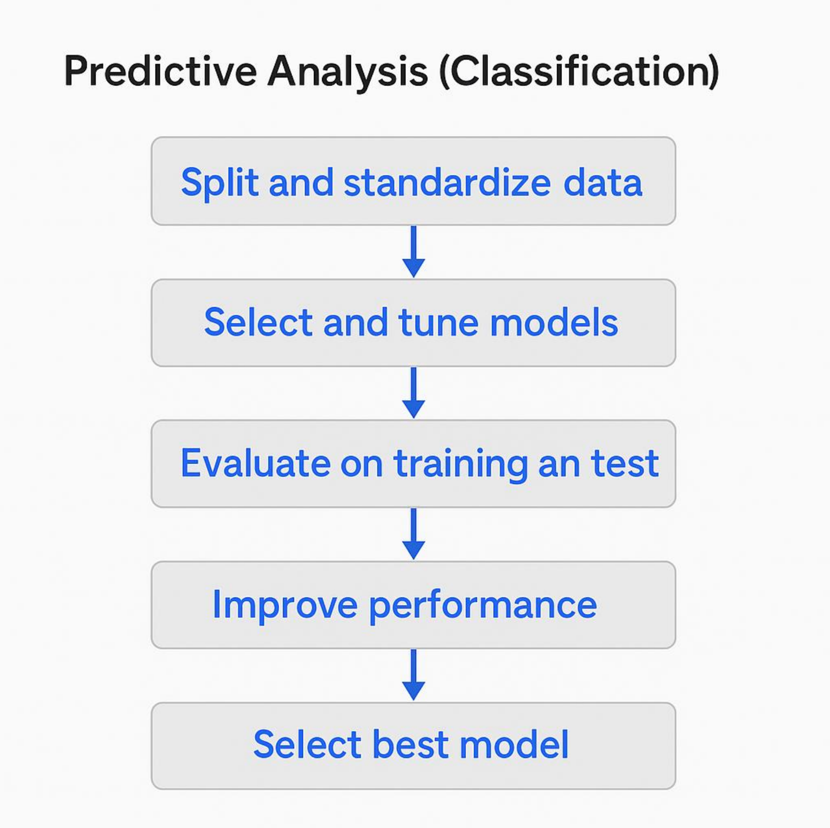
Build a Dashboard with Plotly Dash

Plots Created

- **Pie Chart:**
  - ***All Sites*: shows the distribution of successful launches by site.**
  - ***Specific Site*: compares successful vs. failed launches.**
- **Scatter Plot:**
  - **Relationship between payload mass and launch success.**
  - **Colors indicate booster version.**
- **Interactions Implemented**
- **Dropdown: selects launch site and updates all plots dynamically.**
- **Slider: filters scatter plot data by payload mass range.**
- **Why These Choices**
- **To allow an interactive and visual exploration of how launch site, payload, and booster type impact mission success.**
- Capstone/spacex-dash-app.py at main · liviagibertoni/Capstone

Predictive Analysis (Classification)

- **Process Overview:**
- **Data Preparation:**
  - **Split dataset into features (X) and target (Y).**
  - **Applied train_test_split (80/20 split, fixed random state for reproducibility).**
  - **Standardized features for better model performance.**
- **Model Selection & Tuning:**
  - **Tested Logistic Regression, SVM, Decision Tree, and KNN.**
  - **Used GridSearchCV (cv=10) to find optimal hyperparameters for each model.**
- **Evaluation:**
  - **Measured cross-validation accuracy and test accuracy.**
  - **Plotted confusion matrices for prediction analysis.**
- **Improvement:**
  - **Adjusted hyperparameter ranges to balance accuracy and runtime.**
  - **Compared models side-by-side to identify best performer.**
- **Best Model:**
  - **Logistic Regression achieved highest balanced performance on validation and test data.**
- [Capstone/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · liviagibertoni/Capstone]()



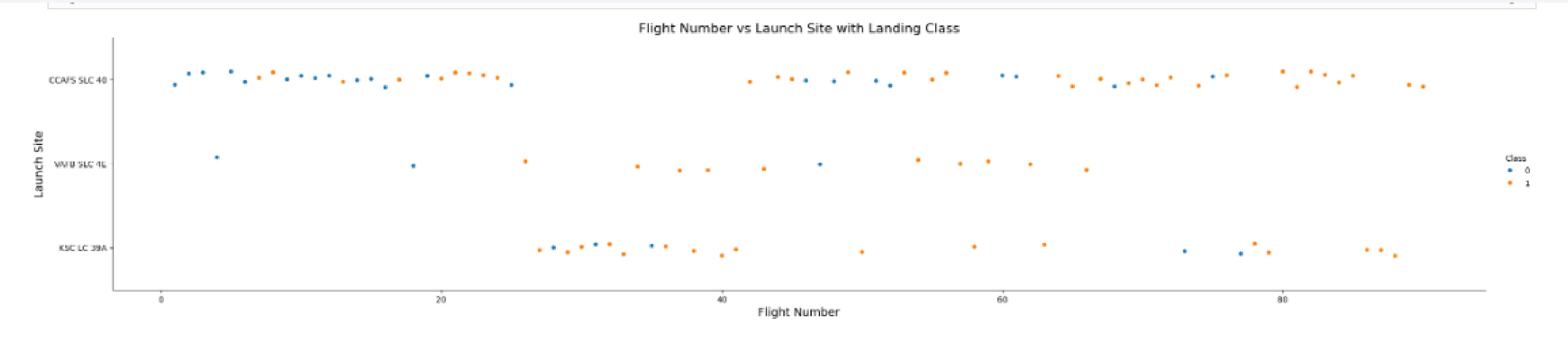
Results

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

Section 2

Insights drawn from EDA

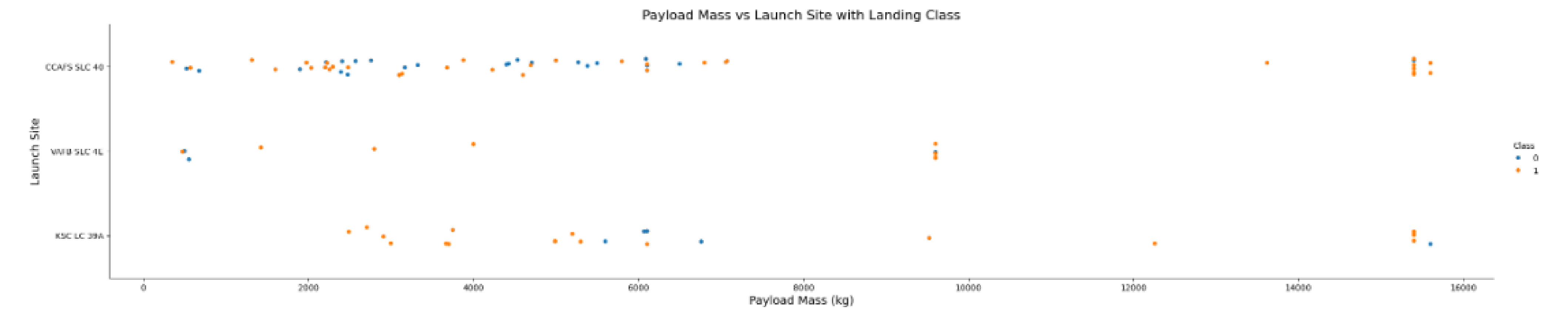
Flight Number vs. Launch Site



Scatter Plot – Flight Number vs. Launch Site

- **Goal: Show how launch outcomes vary by site over time.**
- **X-axis: Flight Number | Y-axis: Launch Site.**
- **Colors: Blue = Failure (0), Orange = Success (1).**
- **Key Insight: Later flights tend to have more successes, especially at certain sites, indicating performance improvements.**

Payload vs. Launch Site



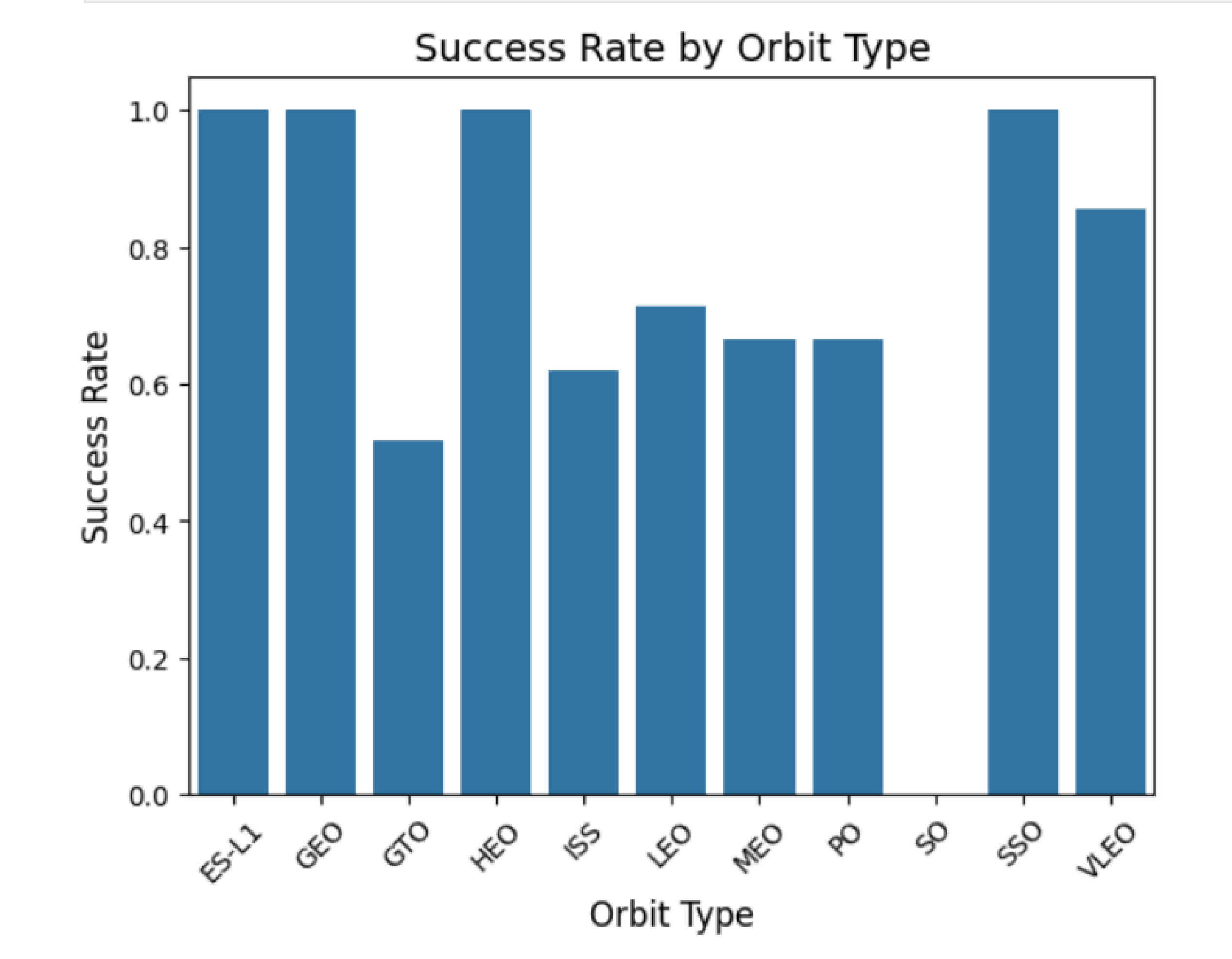
Scatter Plot – Payload vs. Launch Site

- **Goal:** Analyze the relationship between payload mass and landing success across launch sites.
- **X-axis:** Payload Mass (kg) | **Y-axis:** Launch Site.
- **Colors:** Blue = Failure (0), Orange = Success (1).
- **Key Insight:** Successful landings occur across various payload ranges, but some sites show higher success rates in specific payload intervals.

Success Rate vs. Orbit Type

Bar Chart – Success Rate by Orbit Type

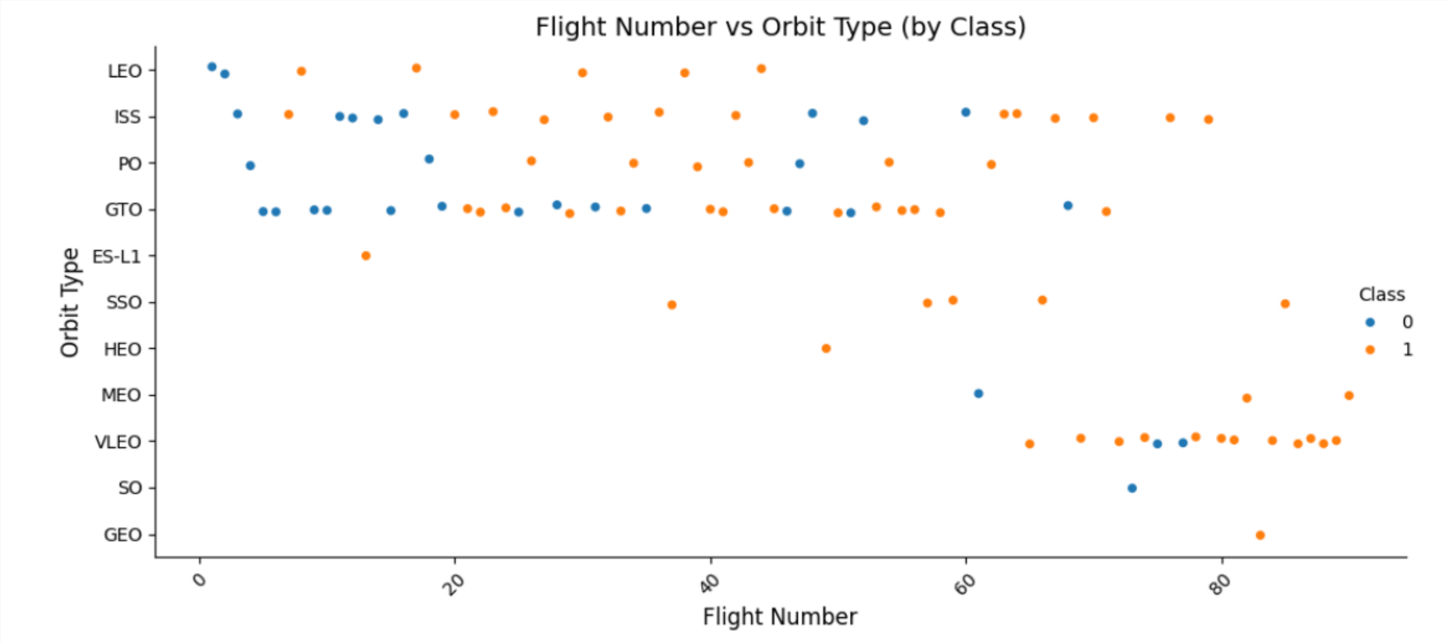
- **Goal:** Compare landing success rates across different orbit types.
- **X-axis:** Orbit Type | **Y-axis:** Success Rate (0 to 1).
- **Key Insight:** Orbits such as **ES-L1, GEO, HEO, and SSO** show the highest success rates (near or at 100%), while **GTO** has a noticeably lower success rate (~50%).
- **Usefulness:** Helps identify which mission profiles are most reliably executed, guiding operational planning and risk assessment



Flight Number vs. Orbit Type

Scatter Plot – Flight Number vs. Orbit Type (by Class)

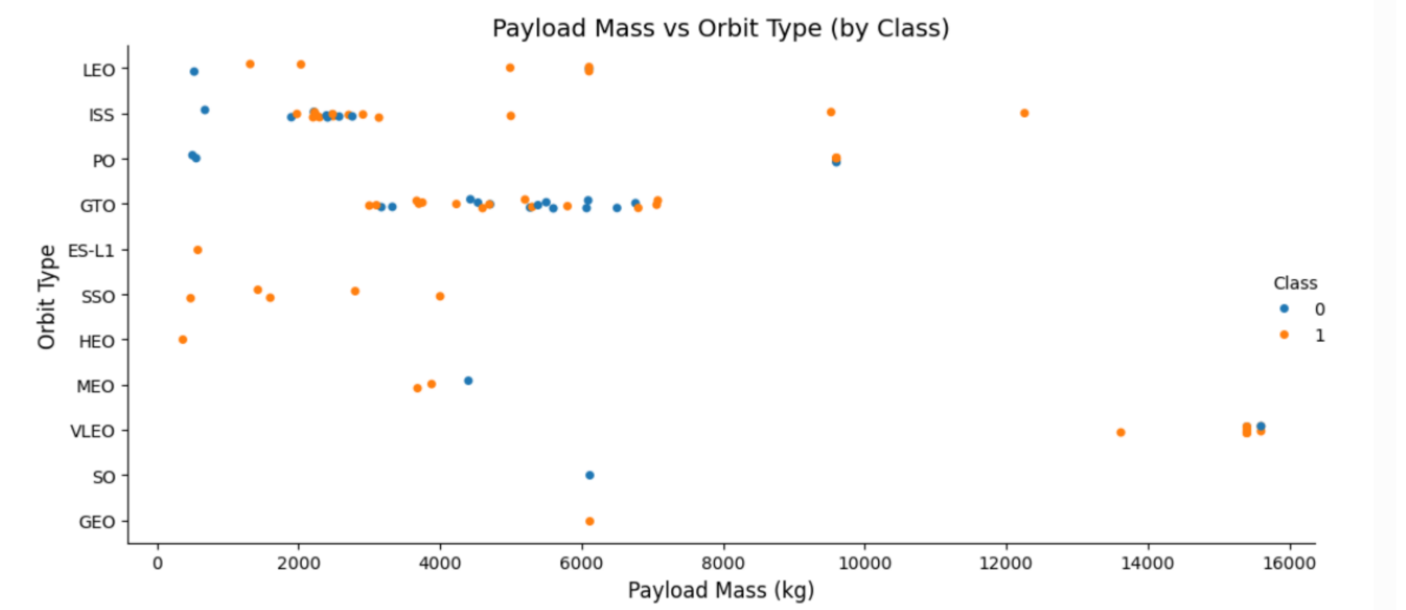
- **Goal:** Examine how mission outcomes vary by orbit type and flight number.
- **X-axis:** Flight Number | **Y-axis:** Orbit Type.
- **Color coding:**
    - **Orange (1):** Successful landings.
    - **Blue (0):** Unsuccessful landings.
- **Key Insight:** Success rates tend to improve with higher flight numbers for most orbit types, suggesting experience and operational refinements play a role.
- **Usefulness:** Highlights which orbits have more consistent success over time and informs mission planning.



Payload vs. Orbit Type

Scatter Plot – Payload Mass vs. Orbit Type (by Class)

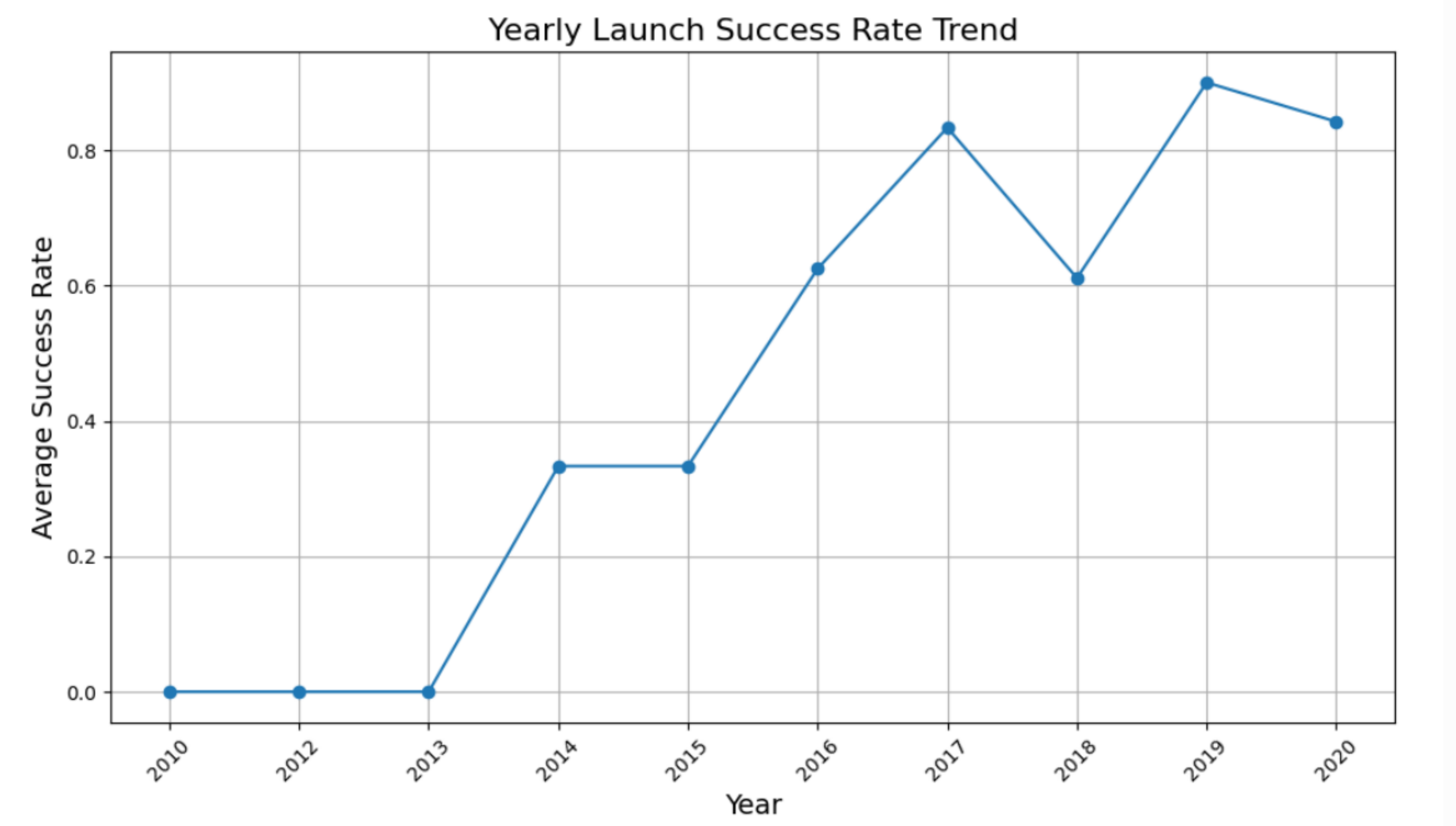
- **Goal:** Analyze the relationship between payload mass and orbit type, along with landing success.
- **X-axis:** Payload Mass (kg) | **Y-axis:** Orbit Type.
- **Color coding:**
    - **Orange (1):** Successful landings.
    - **Blue (0):** Unsuccessful landings.
- **Key Insight:**
    - Most payloads are below 6,000 kg across orbit types, with certain orbits like GTO handling heavier payloads.
    - Some heavier payload missions still achieved successful landings, indicating capability for high-mass deliveries.
- **Usefulness:** Helps evaluate performance limits for different orbits and identify trends in mission success based on payload size.



Launch Success Yearly Trend

Line Chart – Yearly Launch Success Rate Trend

- **Goal: Show how SpaceX's average launch success rate evolved over the years.**
- **X-axis: Year (2010–2020) | Y-axis: Average Success Rate.**
- **Trend:**
  - **2010–2013: 0% success rate.**
  - **2014–2015: Modest improvement (~33%).**
  - **2016–2017: Significant growth, peaking at ~83% in 2017.**
  - **2018: Slight dip (~61%).**
  - **2019–2020: Recovery and stabilization above 85%.**
- **Key Insight:** Reflects SpaceX's learning curve, with rapid improvements after 2014 and consistently high success from 2017 onward**.**



All Launch Site Names

Result:

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

- **Explanation:**
The query retrieves all unique launch site names from the SPACEXTABLE,
all locations used for SpaceX launches in the dataset.

```
In [10]: %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;

 * sqlite:///my_data1.db
Done.
Out[10]:
```

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

Launch Site Names Begin with 'CCA'

In [11]:

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

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

Out[11]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Result (first 5 matching records):

- **2010-06-04 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | LEO | SpaceX | Success | Failure (parachute)**
- **2010-12-08 | CCAFS LC-40 | Dragon demo flight C1 | LEO (ISS) | NASA (COTS), NRO | Success | Failure (parachute)**
- **2012-05-22 | CCAFS LC-40 | Dragon demo flight C2 | LEO (ISS) | NASA (COTS) | Success | No attempt**
- **2012-10-08 | CCAFS LC-40 | SpaceX CRS-1 | LEO (ISS) | NASA (CRS) | Success | No attempt**
- **2013-03-01 | CCAFS LC-40 | SpaceX CRS-2 | LEO (ISS) | NASA (CRS) | Success | No attempt**

Explanation:
The query filters launch site names starting with "CCA" (Cape Canaveral sites) and limits the output to the first five matches. This shows early missions from CCAFS LC-40, including demo flights and CRS missions, with mixed landing outcomes

Total Payload Mass

Result:

- **Total Payload Mass:** 48,213 kg

Explanation:
This query calculates the sum of payload mass for all launches where the customer is **NASA (CRS)**. The total payload delivered by SpaceX boosters for NASA's Commercial Resupply Services missions amounts to **48.2 metric tons**.

```
In [12]:  %sql SELECT SUM("PAYLOAD_MASS__KG_") AS Total_Payload_Mass FROM SPACEXTABLE WHERE "Customer" LIKE '%NASA (CRS)%';

 * sqlite:///my_data1.db
Done.
Out[12]:
```

Total_Payload_Mass
48213

Average Payload Mass by F9 v1.1

Result:

- **Average Payload Mass:** 2,928.4 kg

Explanation:
This query computes the **average payload mass** for all launches using the **F9 v1.1** booster version. On average, each F9 v1.1 launch carried approximately **2.93 metric tons** of payload into orbit.

Display average payload mass carried by booster version F9 v1.1

In [13]:
```
%sql SELECT AVG("PAYLOAD_MASS__KG_") AS Average_Payload_Mass FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
```

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

Out[13]:

Average_Payload_Mass
2928.4

First Successful Ground Landing Date

Result:

- **First Successful Ground Landing Date: 2015-12-22**

Explanation:
This query finds the earliest date when a SpaceX launch achieved a **successful landing on a ground pad**.
The result shows that the first successful ground landing took place on **December 22, 2015**, marking a major milestone in SpaceX’s reusable rocket program.

```
In [14]: %sql SELECT MIN("Date") AS First_Successful_Ground_Landing FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground pad)'

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

Out[14]:

First_Successful_Ground_Landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Result:

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

```
In [15]: %sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__
 * sqlite:///my_data1.db
Done.
```

Out[15]:

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

Explanation:
This query retrieves all **unique booster versions** that:

- Successfully landed on a **drone ship**.
- Carried a **payload mass between 4000 kg and 6000 kg**.
- The results show four booster versions that meet both conditions, indicating they were capable of medium-heavy payload launches while also achieving successful sea landings.

Total Number of Successful and Failure Mission Outcomes

```
In [16]: %sql SELECT "Mission_Outcome", COUNT(*) AS Total FROM SPACEXTABLE GROUP BY "Mission_Outcome";

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

Out[16]:

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

Result:

- **Failure (in flight):** 1
- **Success:** 98
- **Success (duplicate entry):** 1
- **Success (payload status unclear):** 1

Explanation:
From the grouped counts:

- **Total Successes** = 98 + 1 + 1 = **100**
- **Total Failures** = 1
- This shows that the vast majority of missions were successful (100 out of 101), with only **1 recorded failure in flight**. The extra “Success” row appears to be a duplicate label or a slightly different entry in the dataset.

Boosters Carried Maximum Payload

Result:
All the boosters listed below carried the **maximum payload mass** of **15,600 kg**:

- 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

In [17]:
```
%sql SELECT "Booster_Version", "PAYLOAD_MASS__KG_" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__K
```
```
 * sqlite:///my_data1.db
Done.
```
Out[17]:

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

Explanation:
This query identifies all booster versions whose payload mass matches the maximum found in the dataset. Here, **15,600 kg** is the highest payload mass recorded, and multiple boosters achieved it across different missions.

2015 Launch Records

Result:

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

Explanation:
This query filters the dataset to only show records from 2015 where the landing outcome was a failure on a drone ship. It returns the month of launch, the booster version, and the launch site for each case. In 2015, there were two such failures: one in January with booster B1012 and one in April with booster B1015, both launched from CCAFS LC-40.

```
In [18]: %sql SELECT substr("Date", 6, 2) AS Month, "Landing_Outcome", "Booster_Version", "Launch_Site" FROM SPACEXTABLE WHERE "Landi

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

Out[18]:

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

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

Results:

- **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

```
In [19]: %sql SELECT "Landing_Outcome", COUNT(*) AS Outcome_Count FROM SPACEXTABLE WHERE Date >= '2010-06-04' AND Date <= '2017-03-20
 * sqlite:///my_data1.db
Done.
Out[19]:
```

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

Explanation:
The query counts each type of landing outcome between **2010-06-04** and **2017-03-20**, ranking them in descending order by frequency. "No attempt" was the most common, occurring 10 times, followed by equal counts of successes and failures on drone ships.

Section 3

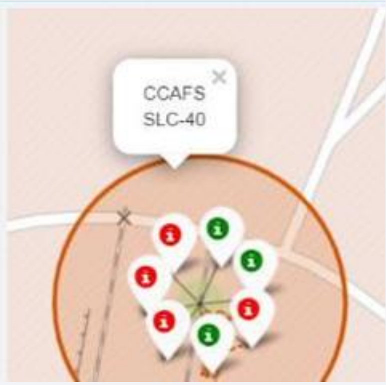
Launch Sites Proximities Analysis

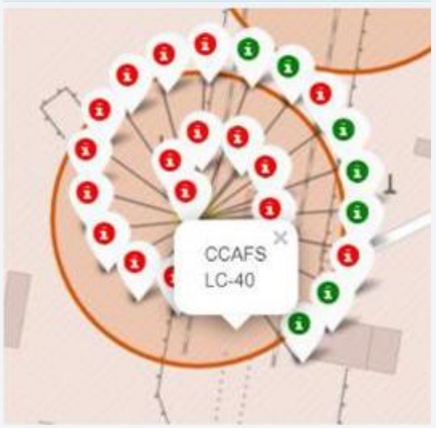
All Launch Sites

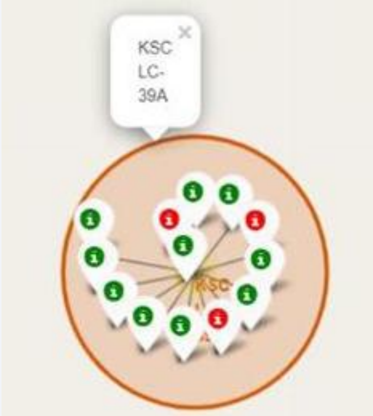
- **SpaceX Launch Sites – Map Overview**
- **Markers:** Show main SpaceX launch site locations.
- **Sites:**
  - **VAFB SLC-4E** (California)
  - **CCAFS SLC-40** (Florida)
  - **KSC LC-39A** (Florida)

Success / Failed launches

- **Launch Outcomes by Site**
- **Left map**: Shows clusters of launches at each site (numbers = total launches).
- **Right map**: Green markers = successful launches, Red markers = failed launches.
- Green markers = successful launches
  Red markers = failed launches







Launch Site Proximities

- **Key point:**
  The distances illustrate that launch sites are strategically located near transportation routes for accessibility, but also far enough from populated areas to ensure safety during launches.

Section 4

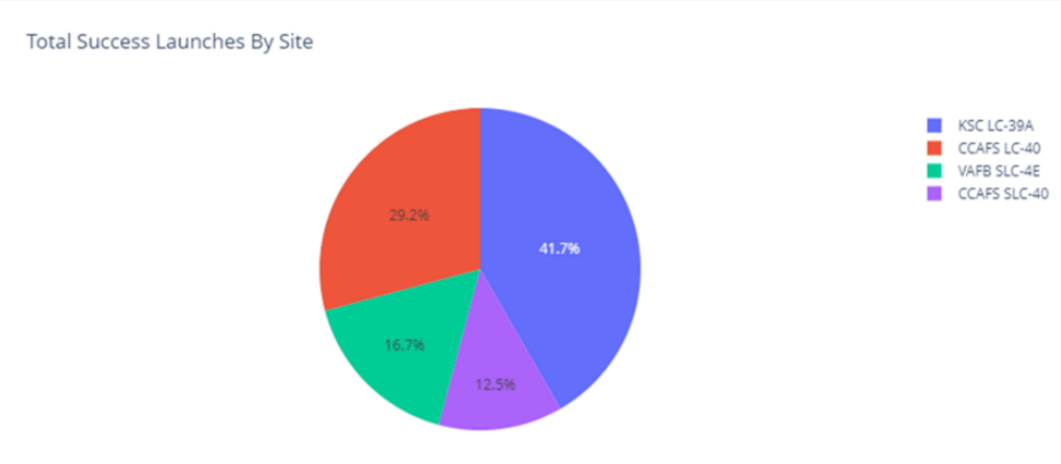
Build a Dashboard with Plotly Dash

Success Launches by site

This pie chart shows the percentage of **successful launches** for each SpaceX launch site:

- **KSC LC-39A** – 41.7% (highest success rate)
- **CCAFS LC-40** – 29.2%
- **VAFB SLC-4E** – 16.7%
- **CCAFS SLC-40** – 12.5%

Key finding:
The Kennedy Space Center LC-39A leads in successful launches, indicating its major role in SpaceX operations compared to other sites.

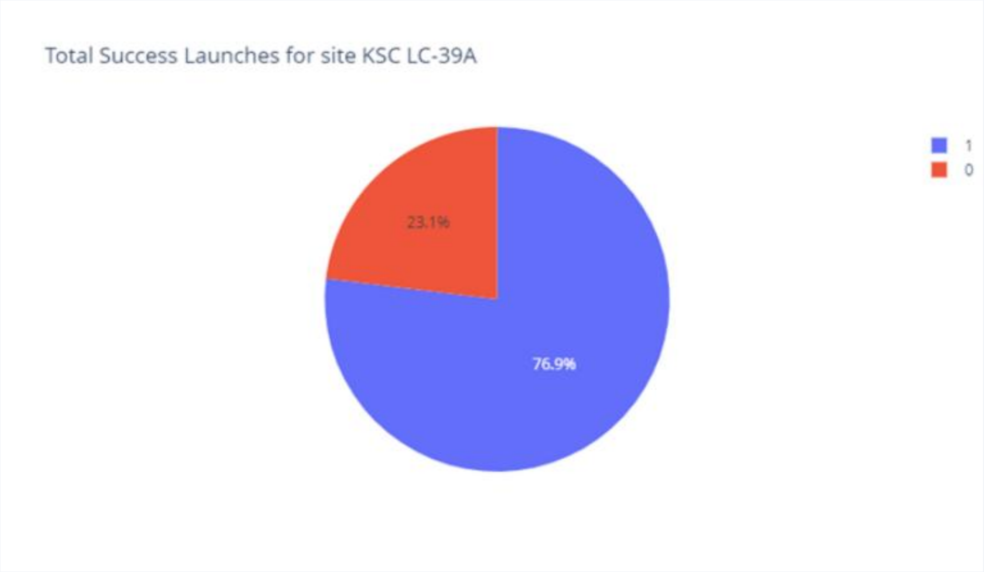


Highest success rate

This pie chart shows the **launch success ratio** for the site **KSC LC-39A**, which has the **highest success rate** among all launch sites.

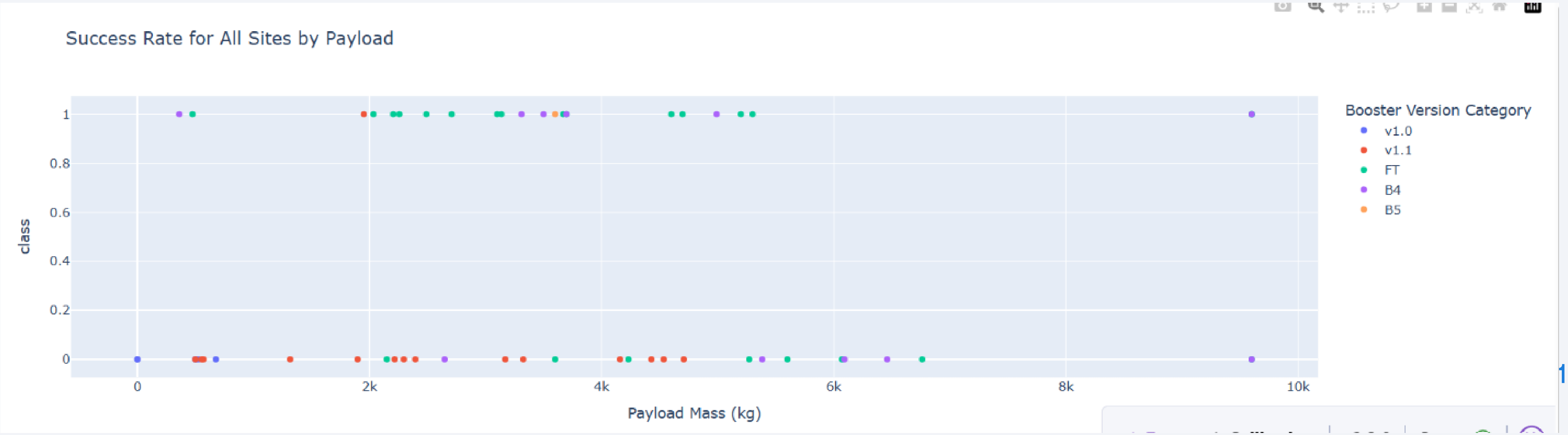
- **Blue (76.9%)** → Successful launches
- **Red (23.1%)** → Failed launches

Key finding:
KSC LC-39A demonstrates strong performance, with **over three-quarters of launches being successful**, confirming it as SpaceX's most reliable launch site in the dataset.



Payload Mass vs. Launch Success by Booster Version

- **Each dot represents a launch, showing success (1) or failure (0).**
- **Colors indicate different booster version categories.**
- **Most launches, regardless of payload mass, were successful (class = 1).**
- **Booster versions FT and B5 show high success rates across various payload ranges.**
- **Very heavy payloads (>8000 kg) also achieved success, though with fewer launches.**



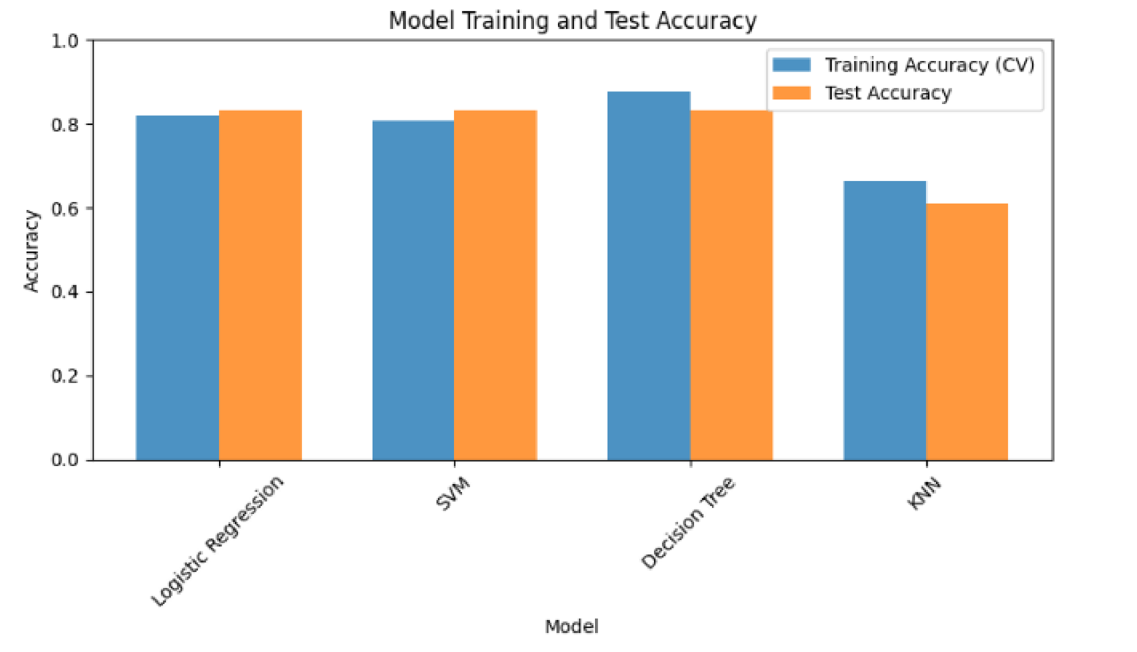
Section 5

Predictive Analysis (Classification)

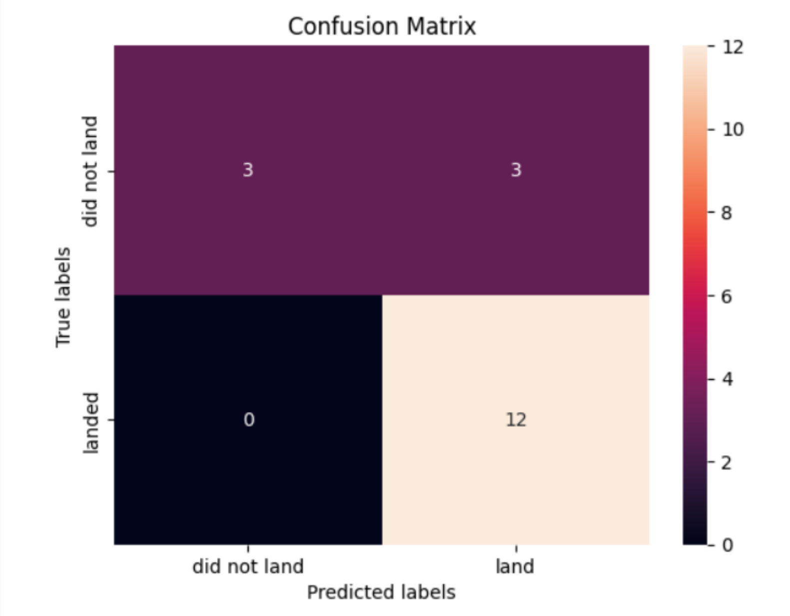
Classification Accuracy

Highest Accuracy:

- **Test Accuracy:** Logistic Regression, SVM, and Decision Tree all have the highest (0.8333).
- **Training Accuracy:** Decision Tree (0.8768).



Confusion Matrix



Confusion Matrix

- **True Positives (TP):** 12 — Correctly predicted "landed."
- **True Negatives (TN):** 3 — Correctly predicted "did not land."
- **False Positives (FP):** 3 — Predicted "landed" when it actually "did not land."
- **False Negatives (FN):** 0 — No missed predictions for "landed."

Key Findings:

- The model correctly classified **15 out of 18 launches** (≈ 83.3% accuracy).
- Zero false negatives means the model always detected when a landing actually occurred.
- Most errors came from false positives, where it predicted a landing that did not happen.

Conclusions

- Launch sites with higher flight volumes generally achieve higher success rates.
- Launch success rates steadily increased from 2013 to 2020.
- The orbits ES-L1, GEO, HEO, SSO, and VLEO achieved the highest success rates.
- KSC LC-39A recorded the highest number of successful launches among all sites.
- Payload mass has some influence on success, but strong performance was observed across multiple booster versions.
- Among the models tested, the Decision Tree Classifier delivered the best predictive performance, showing high accuracy and consistent classification results.

Thank you!