Winning Space Race with Data Science

Yanina Colangelo
February 28th, 2026

Outline

- Executive Summary
- Introduction
- Data Collection & Wrangling
- Exploratory Data Analysis
- Predictive Modeling
- Results & Evaluation
- Conclusion
- Appendix

Executive Summary

- This project investigates the factors that influence Falcon 9 landing success using data analysis and machine learning techniques. Through data preprocessing, visualization, geospatial analysis, and predictive modeling, key performance patterns were identified.
- The results show a clear improvement in launch reliability over time, with certain launch sites and newer booster versions achieving higher success rates. The best-performing predictive models reached 94.4% accuracy, confirming the feasibility of forecasting landing outcomes based on historical mission data.

Introduction

- The objective of this project is to analyze historical Falcon 9 launch data and develop a classification model capable of predicting landing success based on mission characteristics.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

  Collect launch data using SpaceX REST API and web scrapping

- Perform data wrangling

  Clean, transform and structure the dataset

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

  Build, tune and evaluate classification model

Data Collection

Data Sources

- SpaceX REST API
- Wikipedia launch records (web scraping)

Collection Process

- Retrieve launch data using HTTP GET requests
- Convert JSON responses into structured DataFrames
- Extract additional launch details using BeautifulSoup
- Merge API and scraped data into a unified dataset

Data Collection – SpaceX API

Retrieve historical Falcon 9 launch data

Process Overview

- Send HTTP GET resquests to SpaceX REST API
- Retrieve launch data as JSON
- Normalize nested JSON data using pd.json_normalize()
- Select relevant features (flight number, payload mass, orbit, launch site, landing outcome, etc.)
- Convert into Pandas DataFrame

SpaceX REST API

↓

HTTP GET Request

↓

JSON Response

↓

Data Normalization

↓

Structured Pandas DataFrame

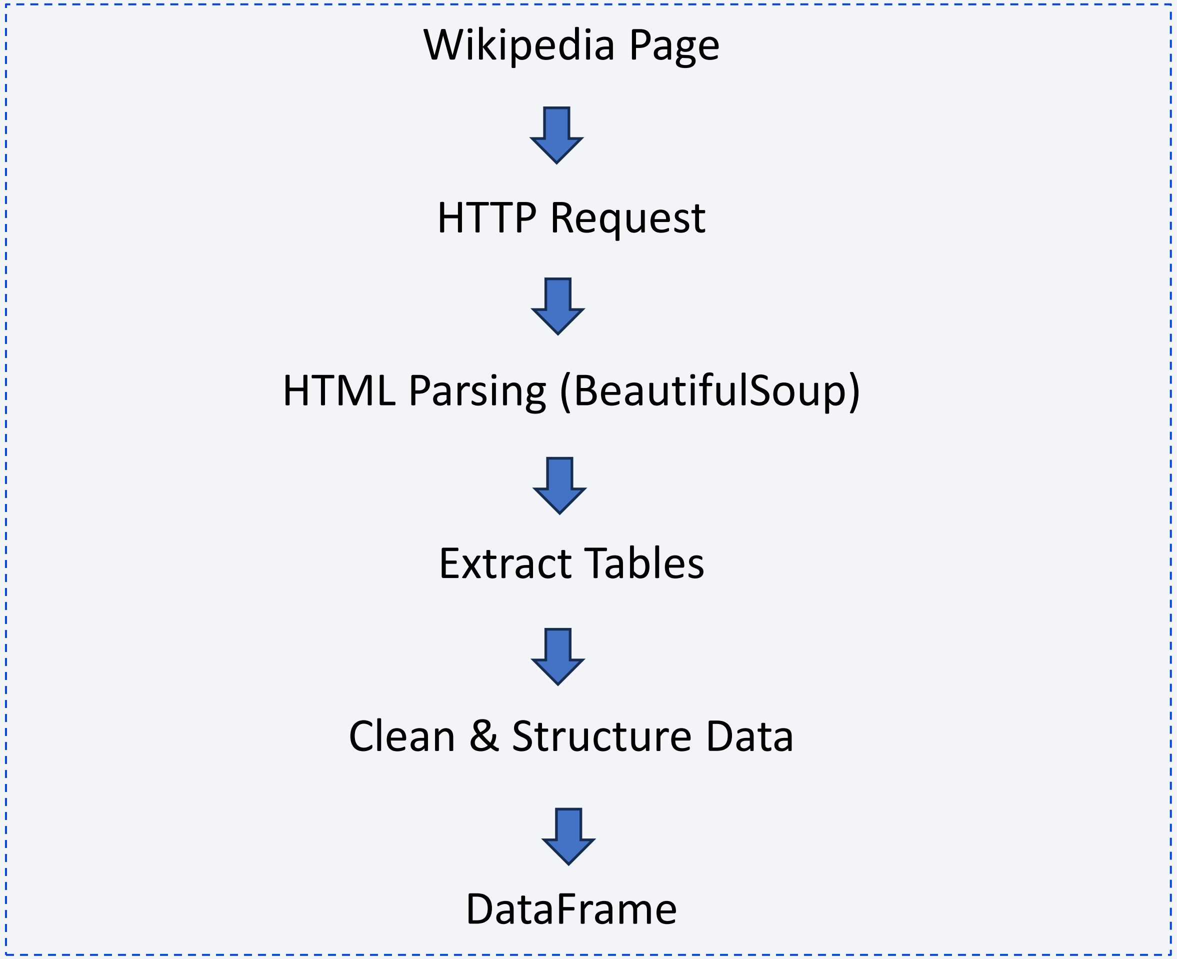
[GitHub Reference]

Data Collection - Scraping

Extract additional launch details not available directly in the API

Process Overview

- Extract additional launch details not available directly in the API
- Parse HTML using BeautifulSoup
- Locate launch tables
- Extract relevant launch records
- Clean and structure extracted data
- Convert into Pandas DataFrame



[GitHub Reference]

Data Wrangling

Objective:

- Prepare the dataset for analysis and modeling.

Process Overview

- Handle missing values
- Replace null payload mass with mean values
- Convert categorical variables into numerical format
- Create target variable (Landing Outcome → 0 / 1)
- Remove unnecessary columns
- Ensure consistent data types
- Store cleaned dataset for

GitHub Reference

EDA with Data Visualization

Objective:

- Identify patterns and relationships influencing landing success.

Charts and Purpose

- Flight Number vs Launch Site (Scatter Plot)

    → Analyze how launch frequency varies by site.

- Payload Mass vs Launch Site (Scatter Plot)

→ Examine payload distribution across different launch sites.

- Success Rate by Orbit Type (Bar Chart)

→ Compare landing success probability across orbit categories.

EDA with Data Visualization

- Flight Number vs Orbit Type (Scatter Plot)
→ Identify operational trends over time.
- Payload Mass vs Orbit Type (Scatter Plot)
→ Assess relationship between payload weight and orbit type.
- Yearly Launch Success Trend (Line Chart)
→ Observe improvement in landing success over time.

[GitHub Reference]()

EDA with SQL

SQL Questions Answered

- Retrieve the unique launch site names
- Filter launch sites that start with “CCA”
- Calculate total payload mass carried for NASA missions
- Compute average payload mass for booster version F9 v1.1
- Find the first successful ground landing date
- List boosters with successful drone ship landing and payload between 4000–6000 kg

EDA with SQL

- Count total successful vs failed mission outcomes
- Identify boosters that carried the maximum payload mass
- Extract 2015 launch records with failed drone ship landings (including booster version and launch site)
- Rank landing outcomes frequency between 2010-06-04 and 2017-03-20

[GitHub Reference]

Build an Interactive Map with Folium

Map Objects Created

- Launch site markers (Marker / Icon)

→ To display the geographic location of each SpaceX launch site.

- Color-coded landing outcome markers (Marker with colors)

→ To visually distinguish successful vs failed landings on the map.

Marker clusters (MarkerCluster) (si lo usaste)

→ To improve readability when multiple launches are close together.

Build an Interactive Map with Folium

Distance lines (PolyLine)

→ To connect a selected launch site with nearby points of interest.

Distance labels (Popup / Tooltip)

→ To display calculated distances (e.g., to coastline, railway, highway).

Circles / Radius overlays (Circle) (si lo usaste)

→ To highlight areas around the launch site and support proximity analysis.

GitHub Reference

Build a Dashboard with Plotly Dash

Dashboard Components

- Pie chart: Launch Success Count by Site

→ To compare total successful launches across all launch sites.

- Pie chart: Success Ratio for Selected Launch Site

→ To analyze success rate for a specific site.

- Scatter plot: Payload Mass vs Launch Outcome

→ To explore how payload mass and booster version relate to landing success.

Build a Dashboard with Plotly Dash

Interactions Added

- Launch Site Dropdown Filter
  → Allows users to focus analysis on one site and compare performance across sites.
- Payload Range Slider
  → Enables dynamic exploration of success/failure patterns for different payload ranges.
- Dynamic chart updates via callbacks
  → Ensures visualizations respond instantly to user selections for interactive analysis.

[GitHub Reference]

Predictive Analysis (Classification)

Model Development Process

- Split dataset into training and testing sets
- Apply feature scaling (StandardScaler)
- Train multiple classification models:
    - Logistic Regression
    - KNN
    - SVM
    - Decision Tree
- Perform hyperparameter tuning using GridSearchCV
- Evaluate models using cross-validation accuracy
- Compare model performance
- Select best-performing model

Predictive Analysis (Classification)

Model Selection

- Logistic Regression → 94.4% accuracy
- KNN → 94.4%
- SVM → 88.9%
- Decision Tree → 77.8%

Logistic Regression was selected due to:

- High accuracy
- Interpretability
- Stable generalization performance

GitHub Reference

Results

Exploratory Data Analysis (EDA)

- Launch success rate improved significantly over time
- Certain launch sites show higher landing success consistency
- Payload mass and orbit type influence landing outcome
- Operational experience (flight number progression) correlates with higher success rate
- Interactive analytics demo in screenshots

Interactive Analytics (Folium & Dash)

- Launch sites are strategically located near coastline and infrastructure
- Success and failure outcomes are visually distinguishable geographically
- Dashboard confirmed payload range and site selection affect landing probability

Predictive Analysis

- Logistic Regression achieved the highest performance (94.4% accuracy)
- KNN performed equally but was less interpretable
- SVM and Decision Tree showed lower accuracy
- Final model reliably predicts Falcon 9 landing outcomes

Section 2

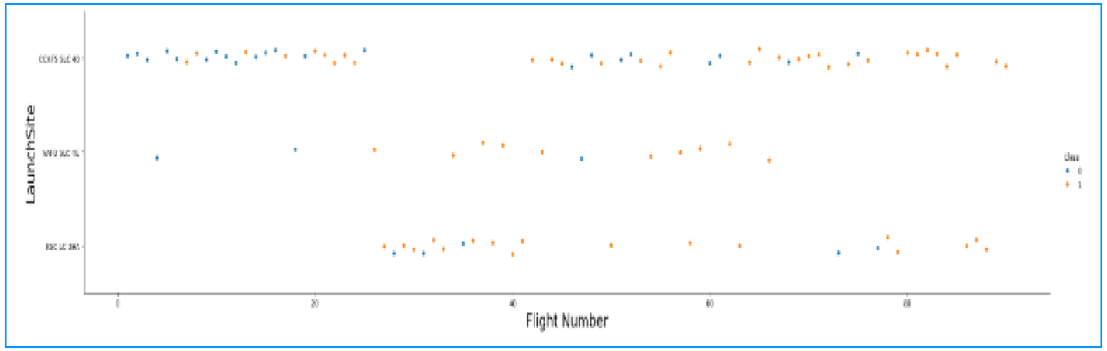
Insights drawn from EDA

Flight Number vs. Launch Site

- The scatter plot shows the distribution of launch missions across different launch sites over time.
- Each point represents a launch mission.
- Colors indicate landing outcome (0 = failure, 1 = success).
- Early missions show more failed landings across all sites.
- As flight numbers increase, successful landings become more frequent.
- CCAFS SLC-40 and KSC LC-39A show higher launch frequency and improved landing success in later missions.
- VAFB SLC-4E presents fewer missions with moderate success consistency.

```python
# Plot a scatter point chart with x axis to be Flight Number and y axis to be
sns.catplot(y="LaunchSite", x="FlightNumber", hue="Class", data=df, aspect =
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("LaunchSite",fontsize=20)
plt.show()
```

Python

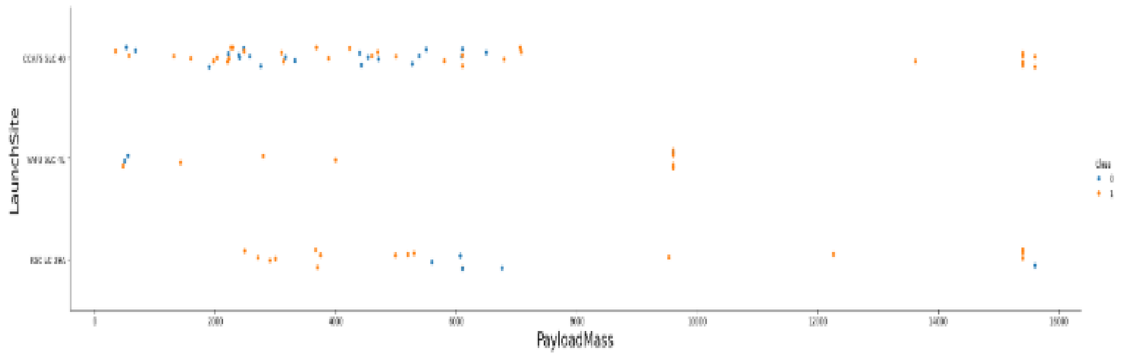


Payload vs. Launch Site

- The scatter plot shows the distribution of payload mass across different launch sites.
- Each point represents a launch mission.
- Colors indicate landing outcome (0 = failure, 1 = success).
- CCAFS SLC-40 and KSC LC-39A handle a wider range of payload masses.
- Higher payload missions (above ~10,000 kg) are mostly associated with successful landings.
- VAFB SLC-4E shows fewer launches with moderate payload ranges.

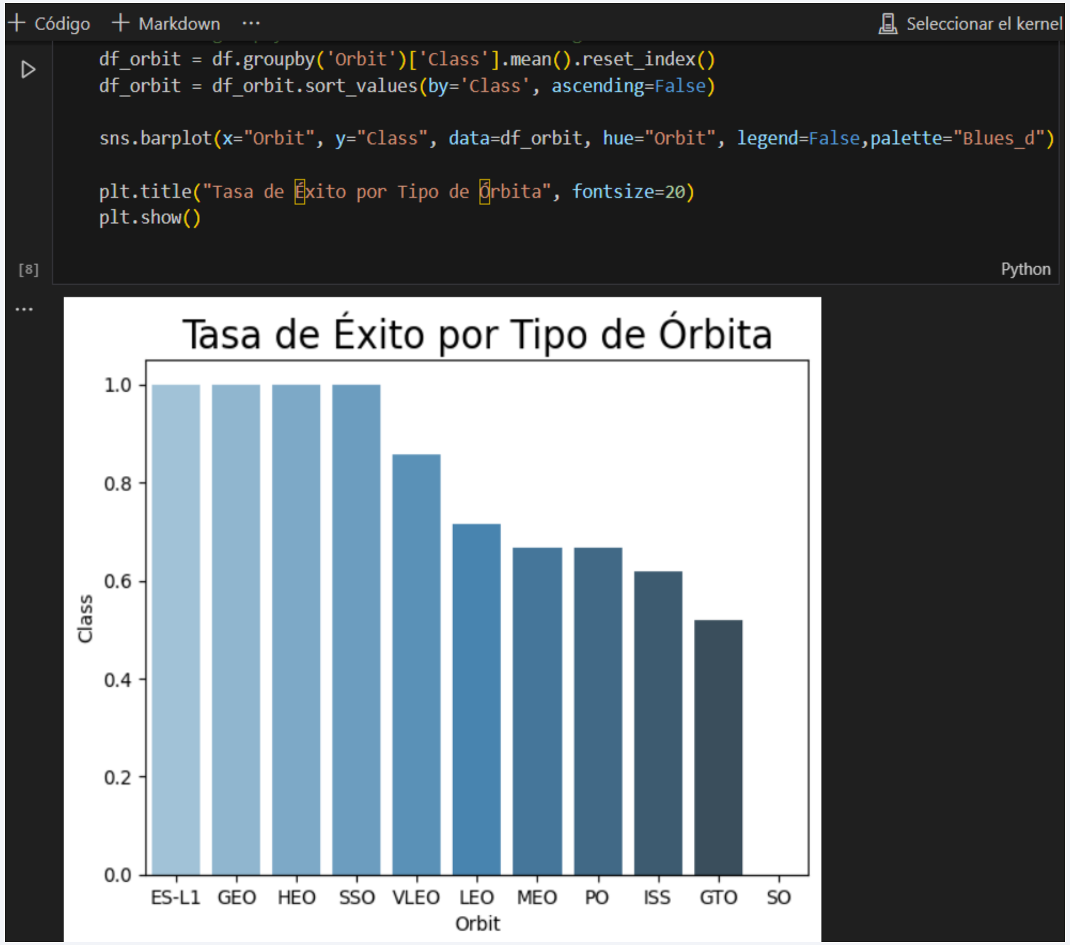
```python
# Plot a scatter point chart with x axis to be Pay Load Mass (kg) and y axis
sns.catplot(y="LaunchSite", x="PayloadMass", hue="Class", data=df, aspect = 
plt.xlabel("PayloadMass",fontsize=20)
plt.ylabel("LaunchSite",fontsize=20)
plt.show()
```

Python



Success Rate vs. Orbit Type

- The bar chart shows the average landing success rate for each orbit type.
- ES-L1, GEO, HEO, and SSO exhibit the highest success rates (close to 100%).
- VLEO and LEO show moderately high success consistency.
- GTO and SO present lower landing success rates compared to other orbit types.
- This suggests that mission profile and orbit complexity influence landing outcomes.

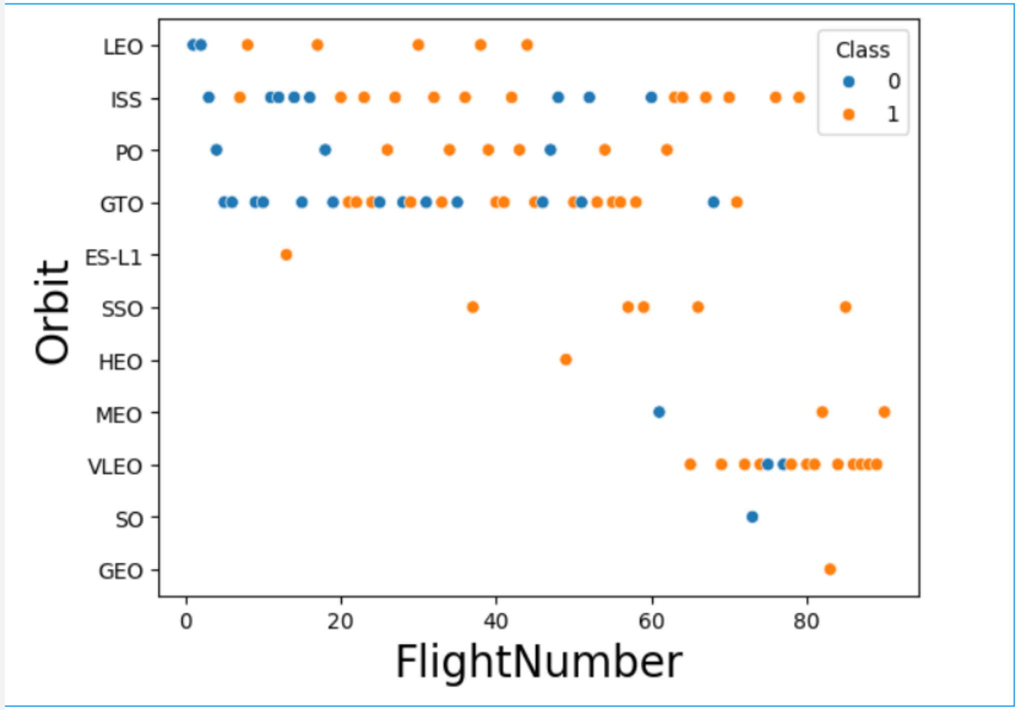


Flight Number vs. Orbit Type

- The scatter plot displays launch missions by orbit type over time.
- Each point represents a launch mission.
- Colors indicate landing outcome (0 = failure, 1 = success).
- Early missions show a higher proportion of failed landings across multiple orbit types.
- As flight numbers increase, successful landings become more frequent.
- VLEO and ISS orbits exhibit a strong concentration of successful landings in later missions.
- GTO missions show more variability, suggesting higher operational complexity.

```python
# Plot a scatter point chart with x axis to be FlightNumber and y axis to be the
sns.scatterplot(y="Orbit", x="FlightNumber", hue="Class", data=df)
plt.xlabel("FlightNumber",fontsize=20)
plt.ylabel("Orbit",fontsize=20)
plt.show()
```

Python



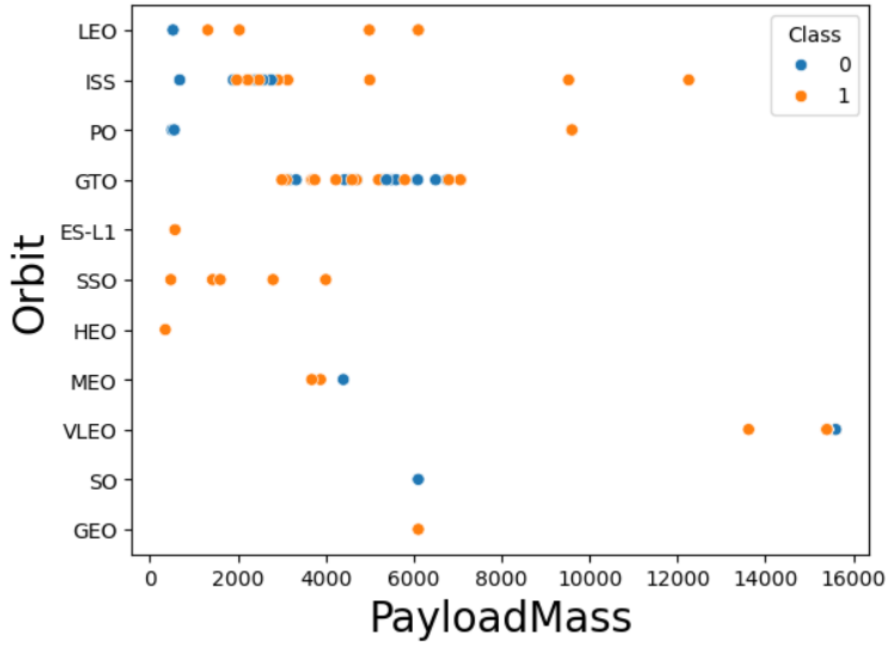
Payload vs. Orbit Type

- The scatter plot shows the distribution of payload mass across different orbit types.
- Each point represents a launch mission.
- Colors indicate landing outcome (0 = failure, 1 = success).
- Moderate payload ranges (approximately 3000–6000 kg) show a high concentration of successful landings, especially in GTO and ISS missions.
- Extremely high payload missions (above 13,000 kg), mainly in VLEO, also demonstrate strong landing success in later flights.
- Some orbit types (e.g., GTO) show mixed outcomes, suggesting increased mission complexity.

```python
# Plot a scatter point chart with x axis to be Payload Mass and y axis to be
sns.scatterplot(y="Orbit", x="PayloadMass", hue="Class", data=df)
plt.xlabel("PayloadMass",fontsize=20)
plt.ylabel("Orbit",fontsize=20)
plt.show()
```

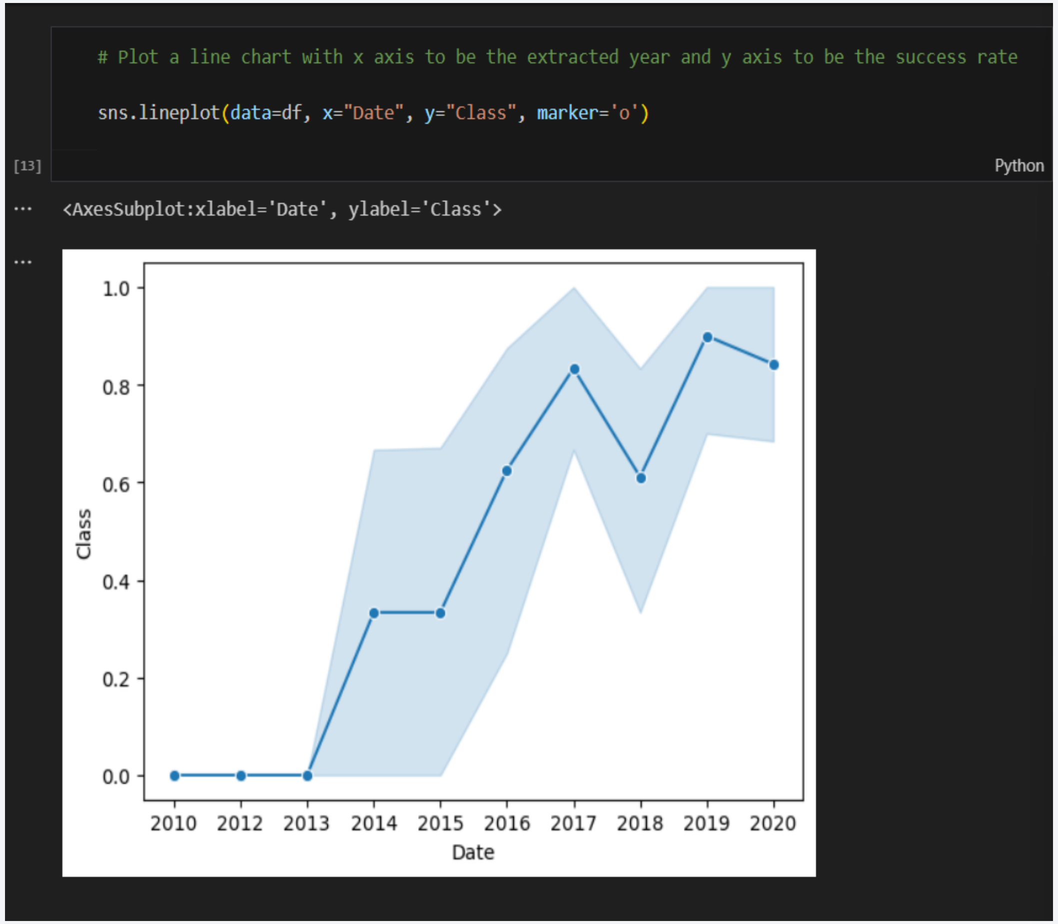
[11] Python

...



Launch Success Yearly Trend

- The line chart shows the yearly average landing success rate.
- Early years (2010–2013) exhibit low or zero landing success.
- A clear upward trend begins around 2014.From 2016 onward, the success rate increases significantly.
- By 2019–2020, landing success stabilizes at a high level (above 80–90%).



All Launch Site Names

```
[11]: %sql SELECT DISTINCT Launch_Site FROM SPACEXTBL;
 * sqlite:///my_data1.db
Done.
```

[11]:

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

- There are four distinct launch sites used for Falcon 9 missions.
- CCAFS SLC-40 has the highest number of launches.
- KSC LC-39A and VAFB SLC-4E are also significant operational sites.
- This confirms that launch activity is concentrated in a limited number of strategic locations.

Launch Site Names Begin with 'CCA'

```
[12]: %sql SELECT * FROM SPACEXTBL where Launch_Site LIKE 'CCA%' LIMIT 5;
 * sqlite:///my_data1.db
Done.
```

[12]:

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

- SQL Query Result
- The query returned 5 launch records where the launch site begins with “CCA”.
- All retrieved records correspond to:
- CCAFS LC-40These missions include early Falcon 9 launches between 2010 and 2013.
- The prefix “CCA” corresponds to Cape Canaveral launch facilities.
- Early Falcon 9 missions were primarily conducted from CCAFS LC-40.
- This confirms that Cape Canaveral was a key operational site during the initial development phase of Falcon 9.

Total Payload Mass

```
[13]: %sql SELECT SUM("PAYLOAD_MASS__KG_") as Masa_Total FROM SPACEXTBL WHERE Customer LIKE '%NASA (CRS)%';

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

[13]:

Masa_Total
48213

SQL Query Result

- The total payload mass carried for NASA (CRS) missions is: **48,213 kg**
- This value represents the cumulative payload mass delivered for NASA Commercial Resupply Services (CRS) missions.
- NASA missions account for a significant portion of Falcon 9 operational activity.
- The large cumulative payload demonstrates Falcon 9's reliability for high-value cargo missions.
- This reinforces NASA's continued partnership with SpaceX.

The substantial total payload mass reflects Falcon 9's operational capability and trust from major institutional customers.

Average Payload Mass by F9 v1.1

```
[13]: %sql SELECT AVG("PAYLOAD_MASS__KG_") as Masa_AVG FROM SPACEXTBL WHERE Booster_Version = 'F9 v1.1';

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

[13]:

Masa_AVG
2928.4

SQL Query Result

- The average payload mass carried by booster version F9 v1.1 is:**2,928.4 kg**
- Booster version F9 v1.1 carried an average payload of approximately 2.9 metric tons.
- This reflects the earlier operational phase of Falcon 9.Compared to later versions, payload capacity was more limited.
- The evolution of booster versions contributed to improved performance and mission capability.

First Successful Ground Landing Date

```
[15]: %sql SELECT *,MIN("Date") as Primer_Aterrizaje_Extitoso FROM SPACEXTBL WHERE Landing_Outcome LIKE '%Success%';
```

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

[15]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome	Primer_Aterrizaje_Extitoso
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)	2015-12-22

SQL Query Result

The first successful Falcon 9 landing on a ground pad occurred on: **December 22, 2015**

Mission details:

- Booster Version: F9 FT B1019
- Launch Site: CCAFS LC-40
- Landing Outcome: Success (ground pad)

- This date marks a major milestone in SpaceX's reusability program.
- The successful ground landing demonstrated the feasibility of first-stage recovery.
- This breakthrough contributed to the significant improvement in landing success rates observed in later years.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
[16]: %sql SELECT Booster_Version, Landing_Outcome, PAYLOAD_MASS__KG_ FROM SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;

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

[16]:

Booster_Version	Landing_Outcome	PAYLOAD_MASS__KG_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

SQL Query Result

The following booster versions successfully landed on a drone ship with payload mass between 4000 and 6000 kg:

- F9 FT B1022 → 4696 kg
- F9 FT B1026 → 4600 kg
- F9 FT B1021.2 → 5300 kg
- F9 FT B1031.2 → 5200 kg

- These missions demonstrate successful drone ship recoveries within a moderate payload range.
- The 4000–6000 kg range appears operationally stable for drone ship landings.
- This reinforces the relationship observed in EDA between payload mass and landing success probability.
- Booster reuse (e.g., B1021.2, B1031.2) highlights improved reusability performance.

Total Number of Successful and Failure Mission Outcomes

SQL Query Result

Mission outcomes distribution:

- Success → 99 missions
- Failure (in flight) → 1 mission
- Success (payload status unclear) → 1 mission

```
[18]: %sql SELECT Mission_Outcome, COUNT(*) AS Total FROM SPACEXTBL GROUP BY Mission_Outcome;

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

[18]:

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

- The overwhelming majority of missions were successful
- .Only one in-flight failure is recorded in the dataset.
- This reflects Falcon 9's strong operational reliability.
- High mission success consistency supports the increasing landing success trend observed in EDA.

Boosters Carried Maximum Payload

SQL Query Result

- The following booster versions carried the maximum payload mass recorded in the dataset:
  - F9 B5 B1048.4F9 B5 B1049.4F9 B5 B1051.3F9 B5 B1056.4F9 B5 B1048.5F9 B5 B1051.4F9 B5 B1049.5F9 B5 B1060.2F9 B5 B1058.3F9 B5 B1051.6F9 B5 B1060.3F9 B5 B1049.7
- All boosters carrying the maximum payload belong to the Falcon 9 Block 5 (F9 B5) version.
- Block 5 represents the most advanced and powerful iteration of Falcon 9.
- Increased payload capacity aligns with later missions and higher landing success rates.
- This confirms technological progression across booster versions.

```
[18]: %sql SELECT DISTINCT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);

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

[18]:

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

2015 Launch Records

```
[19]: %sql SELECT substr(Date, 6, 2) AS Mes, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTBL WHERE Date LIKE '2015%' AND Landing_Outcome = 'Failure (drone ship)';

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

[19]:

Mes	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

- **SQL Query Result:** Failed drone ship landings in 2015:
  - January (01)Booster: F9 v1.1 B1012 Launch Site: CCAFS
  - April (04) Booster: F9 v1.1 B1015 Launch Site: CCAFS LC-40
- Two drone ship landing failures occurred in 2015.
- Both missions were launched from CCAFS LC-40.
- These failures occurred before the first successful ground landing in December 2015.
- This period represents the experimental phase of landing recovery.

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

```
[20]: %sql SELECT Landing_Outcome, COUNT(*) AS Conteo_Total FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY Conteo_Total DESC;

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

[20]:

Landing_Outcome	Conteo_Total
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

- Early missions frequently had no landing attempt, reflecting initial non-reusability phases.
- Drone ship successes and failures appear equally during the experimental stage.
- Ground pad successes begin appearing as recovery technology matures.
- Ocean landings and parachute failures reflect early recovery strategies.
- The data illustrates the transition from experimental recovery attempts to successful controlled landings

Section 3

Launch Sites Proximities Analysis

Global Distribution of SpaceX Launch Sites

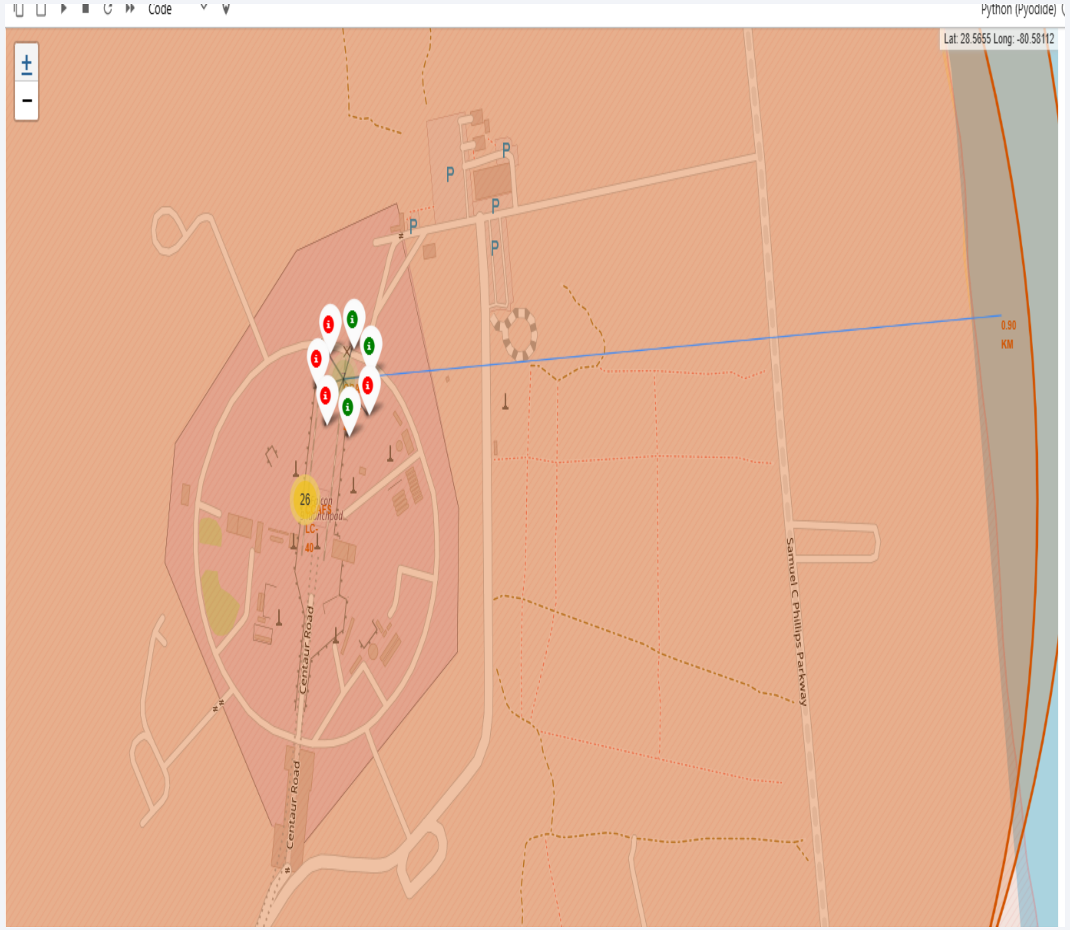
- The map displays the geographic location of all Falcon 9 launch sites.
- Four major launch sites are visible:
  - CCAFS LC-40 (Florida, USA)
  - CCAFS SLC-40 (Florida, USA)
  - KSC LC-39A (Florida, USA)
  - VAFB SLC-4E (California, USA)
- Launch sites are strategically positioned along coastlines.
- Coastal locations minimize risk to populated areas during launch and recovery operations.
- The geographic clustering in the United States reflects centralized launch operations.

Landing Outcomes by Launch Site

- The map displays launch sites with landing outcomes represented by color-coded markers.
- Each marker corresponds to a Falcon 9 launch mission.
- Different colors distinguish between:
  - Successful landings
  - Failed landings
  - No landing attempts
- Florida sites (CCAFS SLC-40 and KSC LC-39A) show a higher concentration of successful landings in later missions.
- VAFB SLC-4E presents fewer missions but consistent outcomes.

Proximity Analysis of Launch Site (Geospatial Distance Measurement)

- The map zooms into a selected launch site to analyze nearby infrastructure.
- The launch pad is highlighted at the center.
- Colored markers indicate landing outcomes.
- A distance line (PolyLine) is drawn from the launch site to the coastline.
- The calculated distance (0.90 km) is displayed directly on the map.
- Circular overlays highlight the safety buffer zone around the launch facility.

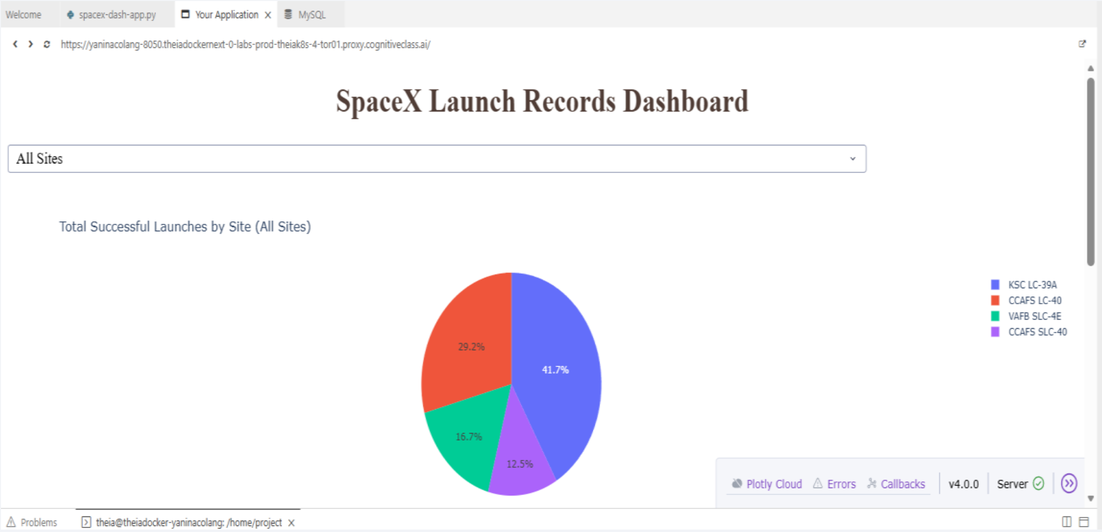


Section 4

Build a Dashboard with Plotly Dash

Launch Success Distribution by Site (All Sites Overview)

- The dashboard displays a pie chart representing the total number of successful launches across all launch sites.
- The dropdown selector is set to “All Sites,” allowing a global comparison.
- Each slice represents a launch site.
- The percentage values show the proportion of total successful launches per site.

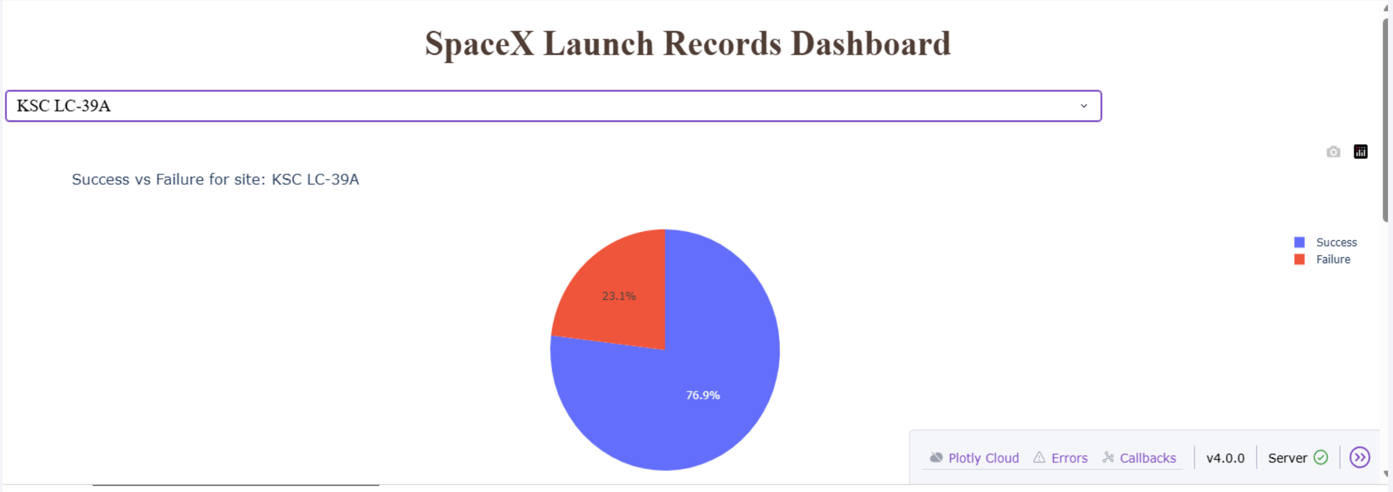


What this tells us:

- Florida-based launch sites dominate successful missions.
- KSC LC-39A is the strongest-performing site.
- Success distribution is not uniform across sites.

Launch Site with Highest Success Ratio (KSC LC-39A)

- KSC LC-39A has the highest launch success ratio among all sites.
- More than three-quarters of missions launched from this site were successful.
- The large dominance of the blue section visually confirms operational reliability.
- This reinforces earlier findings from EDA and success rate analysis.
- Explain the important elements and findings on the screenshot

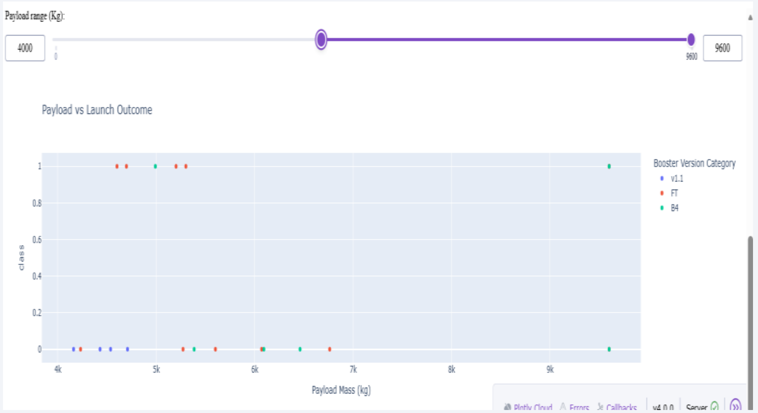


- KSC LC-39A demonstrates the strongest operational performance, making it the most reliable launch site in the dataset.

Payload vs Launch Outcome (All Sites with Range Slider)

Overall Interpretation

- Payload mass alone does not fully determine launch success.
- However, certain payload ranges combined with newer booster versions show significantly higher success probability.
- Booster version appears to be a strong influencing factor.
- While payload mass influences mission complexity, booster version maturity plays a stronger role in determining launch success.



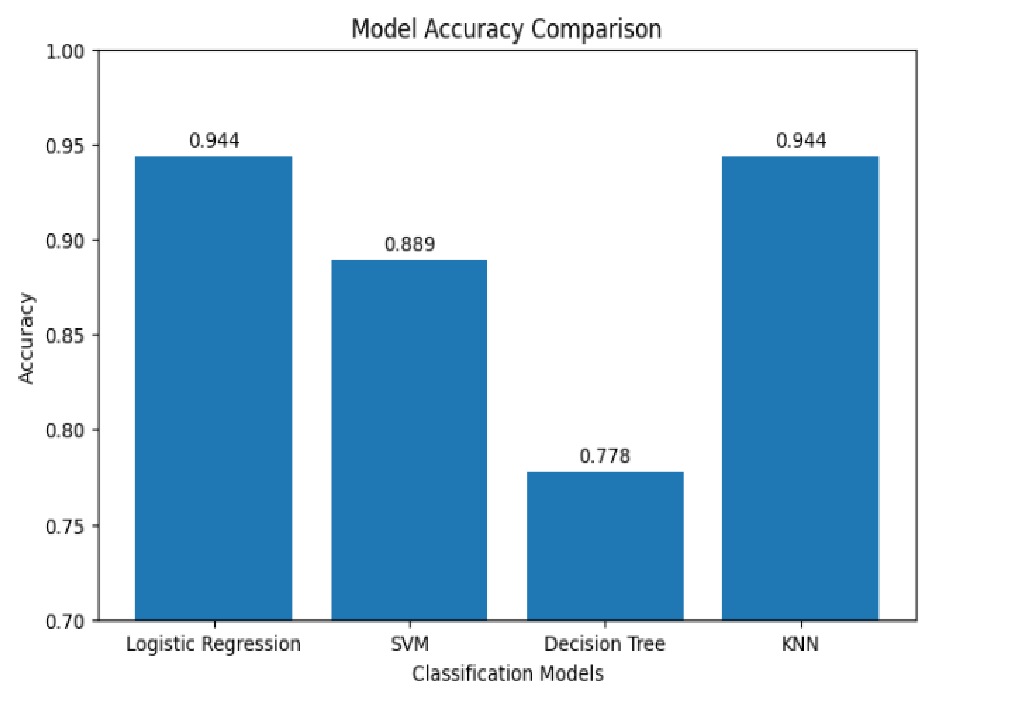
Section 5

Predictive Analysis (Classification)

Classification Model Accuracy Comparison

Key Findings

- Logistic Regression → 94.4% accuracy
- KNN → 94.4% accuracy
- SVM → 88.9% accuracy
- Decision Tree → 77.8% accuracy

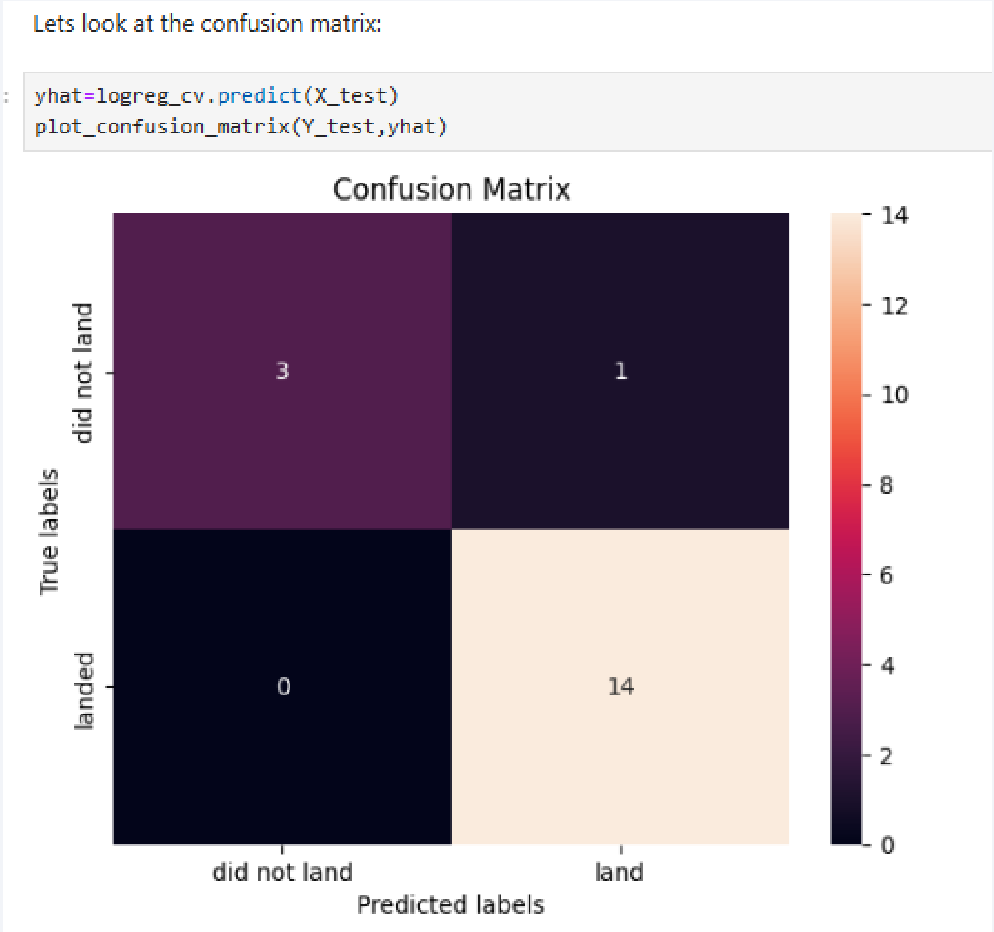


Analytical Interpretation

- Logistic Regression performed extremely well, suggesting that the decision boundary between success and failure may be close to linear.
- KNN also performed strongly, indicating that local feature similarity plays an important role.
- Decision Tree showed lower performance, possibly due to overfitting or limited generalization.

Confusion Matrix

- **Model Performance** Test Accuracy: **94.4%**
- **Best Model**: Logistic Regression (tie with KNN, but shown here)
- The model shows extremely strong performance in predicting successful landings.
- Zero false negatives means it did not miss any successful landing.
- Only one misclassification occurred.
- The model demonstrates excellent generalization on test data.



Conclusions

1. Falcon 9 landing success has significantly improved over time, showing clear technological and operational advancement.
2. Launch site plays an important role in mission success, with KSC LC-39A demonstrating the highest success ratio.
3. Booster version maturity strongly influences landing outcomes, with newer versions (B4, B5) achieving higher success rates.
4. Logistic Regression and KNN achieved the highest classification accuracy (94.4%), demonstrating strong predictive capability for landing success.

The combination of exploratory analysis, geospatial visualization, and machine learning modeling provides a robust framework for predicting Falcon 9 landing outcomes.

Appendix

Python Code Snippets

- Data cleaning and preprocessing
- Feature engineering (Orbit encoding, Launch Site encoding)
- Train-test split
- Model training using:
  - Logistic Regression
  - SVM
  - Decision Tree
  - KNN
- Hyperparameter tuning with GridSearchCV

SQL Queries

- Key analytical queries performed:
- Total launches by site
- NASA total payload mass
- First successful landing date
- Boosters with maximum payload
- Ranking landing outcomes

Appendix

Data Visualization Assets

- Scatter plots (Payload vs Orbit)Success rate by orbit (Bar chart)Launch success yearly trend
- Folium interactive maps
- Plotly Dash dashboard
- Model accuracy comparison bar chartConfusion matrix

Dataset Information

- SpaceX Launch Records dataset
- Features used:
  - Flight Number
  - Payload Mass
  - Orbit
  - Launch Site
  - Booster Version
  - Landing Outcome

Interactive Components

- Plotly Dash dropdown filter by site
- Payload range slider
- Folium distance measurement (coastline proximity)Folium interactive maps

Appendix

SQL Queries

- Key analytical queries performed:
- Total launches by site
- NASA total payload mass
- First successful landing date
- Boosters with maximum payload
- Ranking landing outcomes

Thank you!