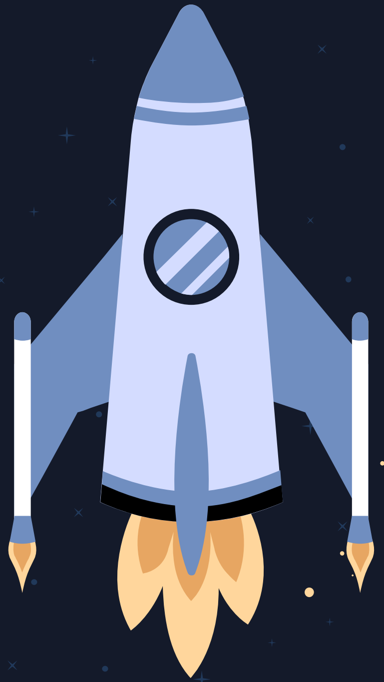
IBM Developer
SKILLS NETWORK

Wining Space Race with Data Science

Kevin Requena
16th September 2025

PRESENTATION CONTENTS

01 EXCECUTIVE SUMMARY
02 INTRODUCTION
03 METHODOLOGY
04 DATA COLLECTION
05 DATA WRANGLING
06 EDA WITH VISUALIZATION
07 EDA WITH SQL
08 MAP WITH FOLIUM
09 DASHBOARD WITH PLOTLY DASH
10 PREDICTIVE ANALYTICS
11 RESULT SUMMARY



EXECUTIVE SUMMARY

Exploratory Data Analysis:

Launch success has shown continuous improvement over time.The KSC LC-39A launch site has the highest success rate among all landing sites.Orbits ES-L1, GEO, HEO, and SSO have achieved a 100% success rate.

Visualization and Analytics:

Most launch sites are located near the equator, with all being close to the coast.

Predictive Analytics:

All models performed similarly on the test set, with the decision tree model slightly outperforming the others.

Data Collection Process:

- Data collection using SpaceX's REST API.
- Analyzing and understanding the structure of the collected data.
- Explaining the data through SQL and visualizations.
- Generating infographics using Folium.
- Developing an interactive dashboard with Plotly Dash.
- Implementing classification models using machine learning algorithms.
- Evaluating model performance and analyzing the results obtained.

INTRODUCTION

SpaceX, a leader in the space industry, is committed to making space travel more affordable for everyone. Its achievements include sending spacecraft to the International Space Station, launching a satellite constellation that provides global internet access, and conducting manned missions to space. A key factor in SpaceX's success is the relatively low cost of its rocket launches ($62 million per launch), made possible by the innovative reuse of the first stage of its Falcon 9 rocket. In contrast, other providers who are unable to reuse the first stage charge upwards of $165 million per launch.

By predicting whether the first stage of a rocket will successfully land, we can estimate the launch cost. To achieve this, we can leverage publicly available data and apply machine learning models to forecast whether SpaceX, or a competing provider, will be able to reuse the first stage.

Exploration:

- Impact of Payload Mass, Launch Site, Number of Flights, and Orbit Type on the success of first-stage landings.
- Trends in the Rate of Successful Landings over time.
- Identifying the Best Predictive Model for forecasting successful landings (binary classification).

METHODOLOGY

- **Collect Data** using the SpaceX REST API and web scraping techniques.
- **Data Wrangling:** Filter data, handle missing values, and apply one-hot encoding to prepare the **dataset for analysis and modeling.**
- **Exploratory Data Analysis (EDA):** Use SQL queries and data visualization techniques to explore the dataset.
- **Data Visualization:** Visualize insights with Folium and Plotly Dash for interactive and detailed representations.
- **Model Building:** Develop classification models to predict landing outcomes. Tune and evaluate models to identify the best-performing model and optimal parameters.

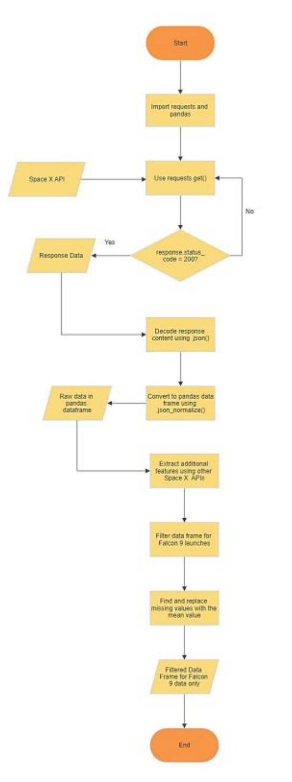
DATA COLLECTION –API

- **Library import:** The Python libraries requests, BeautifulSoup, and pandas were imported to enable data retrieval, parsing, and structuring.
- **Data request:** The Wikipedia source page was accessed using the requests.get() method.
- **HTML parsing:** The HTML response was parsed using the BeautifulSoup() constructor, generating a structured object representation of the web content.
- **Table extraction:** All HTML tables were extracted from the parsed object through the soup.find_all() method.
- Table selection: The third table in the sequence was selected for analysis, and its header (column names) was identified and retrieved.
- **Data extraction:** Each row of the selected table was iterated to extract the following attributes: Flight No., Launch site, Payload, Payload mass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, and Time.
- **Data structuring:** The extracted records were consolidated into a pandas DataFrame, providing a structured dataset suitable for subsequent processing and analysis."

DATA COLLECTION – SPACEX API

The flowchart illustrates the data collection workflow, which uses the Python requests library to interact with the SpaceX API and retrieve mission-related information in a structured manner. The acquired data is organized into a DataFrame to facilitate further processing and analysis. Finally, the dataset is exported as a CSV file using the df.to_csv() function, ensuring reproducibility and ease of access. The corresponding Jupyter Notebook, available in the Git repository linked below, documents the complete implementation of this process.

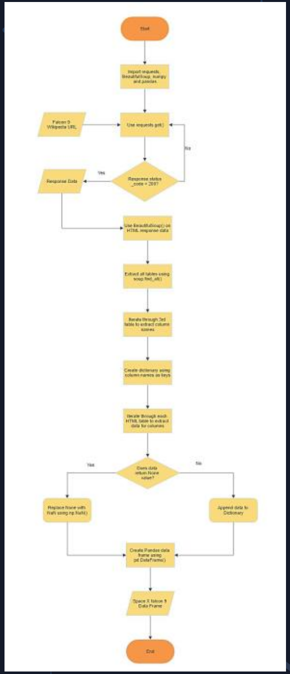
https://github.com/Keliu89/Capstone_Project_SpaceX-s/blob/main/Module%201%20-%20Capstone%20Introduction%20and%20Understanding%20the%20Datasets/2.%20Hands-on%20Lab.%20Complete%20the%20Data%20Collection%20API%20Lab.ipynb



DATA COLLECTION –WEB SCRAPING

- Request Falcon 9 Launch Data from Wikipedia.
- Create a BeautifulSoup Object from the HTML response for parsing.
- Extract Column Names from the HTML table header for accurate data mapping.
- Parse HTML Tables to collect the launch data.
- Create a Dictionary from the parsed data.
- Convert the Dictionary to a DataFrame for structured analysis.
- Export the DataFrame to a CSV file for further processing.

https://github.com/Keliu89/Capstone_Project_SpaceX-s/blob/main/Module%201%20-%20Capstone%20Introduction%20and%20Understanding%20the%20Datasets/3.%20Hands-on%20Lab.%20Complete%20the%20Data%20Collection%20with%20Web%20Scraping%20lab.ipynb



DATA WRANGLING

- Objectives of Exploratory Data Analysis (EDA):
- **Launches per Site:** Analyze the number of launches from each launch site.
- **Occurrence of Different Orbits:** Investigate the frequency of different orbit types in the dataset.
- **Number of Launches per Orbit Type:** Count the number of launches for each orbit type.
- **Label Encoding of the Target Column:** Apply label encoding to transform categorical target data for modeling purposes.

https://github.com/Keliu89/Capstone_Project_SpaceX-s/blob/main/Module%201%20-%20Capstone%20Introduction%20and%20Understanding%20the%20Datasets/4.%20Hands-on%20Lab.%20Data%20Wrangling.ipynb

Landing Outcome:

- **TrueOcean:** The mission outcome indicates a successful landing in a specific region of the ocean.
- **False Ocean:** Represented an unsuccessful landing in a specific ocean region.
- **True RTLS:** Indicates a successful landing on a ground pad.
- **False RTLS:** Represents an unsuccessful landing on a ground pad.
- **True ASDS:** Denotes a successful landing on a drone ship.
- **False ASDS:** Represents an unsuccessful landing on a drone ship.
- **Outcome Encoding:** Outcomes were converted into 1 for successful landings and 0 for unsuccessful landings.

EDA WITH VISUALIZATION

Charts:

- **Flight Number vs. Payload:** Visualize the relationship between the flight number and payload mass.
- **Flight Number vs. Launch Site:** Analyze how the launch site is distributed across different flight numbers.
- **Payload Mass (kg) vs. Launch Site:** Explore how payload mass varies across different launch sites.
- **Payload Mass (kg) vs. Orbit Type:** Examine the distribution of payload mass across various orbit types.

Analysis:

- **Scatter Plots:** Use scatter plots to identify potential relationships between variables. These relationships may be useful for machine learning models if they exist.
- **Bar Charts:** Use bar charts to compare discrete categories, showing how each category relates to a measured value. Bar charts help visualize the relationships and distribution across different categories.

https://github.com/Keliu89/Capstone_Project_SpaceX-s/blob/main/Module%202%20-%20Exploratory%20Data%20Analysis%20(EDA)/6.%20Hands-on%20Lab.%20Complete%20the%20EDA%20with%20Visualization.ipynb

EDA WITH SQL

Queries:

- **Unique Launch Sites:** Retrieve the names of all unique launch sites.
- **Launch Sites Beginning with 'CCA':** Get 5 records where the launch site name begins with 'CCA'.
- **Total Payload Mass by NASA (CRS):** Calculate the total payload mass carried by boosters launched by NASA under the CRS program.
- **Average Payload Mass for F9 v1.1:** Find the average payload mass carried by the F9 v1.1 booster version.
- **Date of First Successful Ground Pad Landing:** Retrieve the date of the first successful landing on a ground pad.
- **Boosters with Successful Drone Ship Landings (4,000-6,000 kg payload):** List boosters that successfully landed on a drone ship, with a payload mass between 4,000 and 6,000 kg.
- **Successful and Failed Missions:** Count the total number of successful and failed missions.
- **Boosters with Maximum Payload Capacity:** Identify the names of booster versions that carried the maximum payload.
- **Failed Drone Ship Landings (2015):** Retrieve the failed landing outcomes on drone ships, along with their booster versions and launch sites for the months in 2015.
- **Landing Outcomes (2010-06-04 to 2017-03-20):** Count the landing outcomes (successful/failed) between June 4, 2010, and March 20, 2017, in descending order.

https://github.com/Keliu89/Capstone_Project_SpaceX-s/blob/main/Module%202%20-%20Exploratory%20Data%20Analysis%20(EDA)/5.%20Hands-on%20Lab.%20Complete%20the%20EDA%20with%20SQL.ipynb

MAP WITH FOLIUM

Markers Indicating Launch Sites:

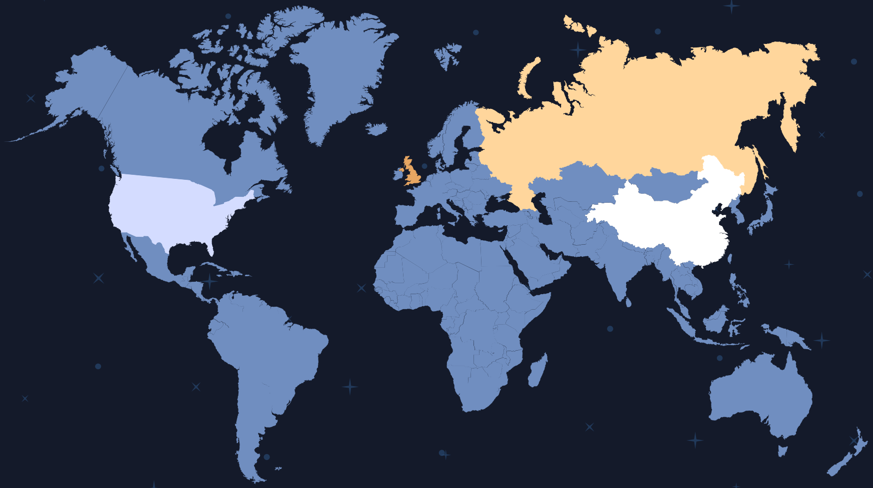
- Added a **blue circle** at the NASA Johnson Space Center's coordinates, with a popup label showing its name, using its latitude and longitude.
- Added **red circles** at all launch site coordinates, with popup labels showing the launch site names, also using their latitude and longitude coordinates.

Colored Markers of Launch Outcomes:

- Added **colored markers** to represent launch outcomes:
- **Green** markers for successful launches.
- **Red** markers for unsuccessful launches.
- This visualization helps identify launch sites with high success rates.

Distances Between a Launch Site and Proximities:

- Added **colored lines** to show the distance between the CCAFS SLC-40 launch site and its proximity to the nearest coastline, railway, highway, and city.



https://github.com/Keliu89/Capstone_Project_SpaceX-s/blob/main/Module%203%20-%20Interactive%20Visual%20Analytics%20and%20Dashboard/7.%20Hands-on%20Lab.%20Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

BUILD A DASHBOARD WITH PLOTLY DASH

Dropdown Menu for Launch Sites

- Enable users to select either all launch sites or a specific one.

Pie Chart Displaying Successful Launches

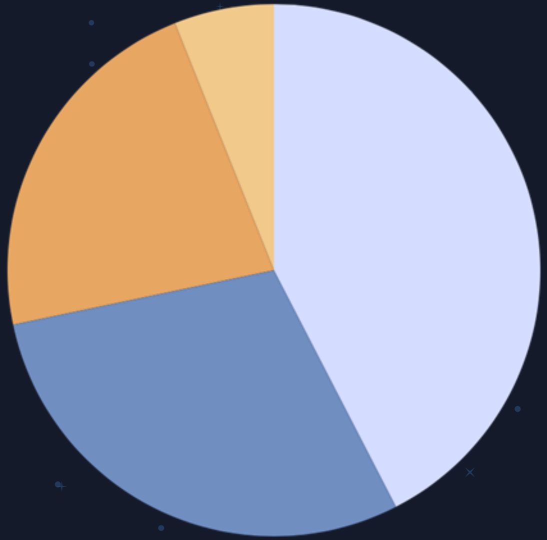
- Allow users to view the percentage of successful and unsuccessful launches relative to the total.

Payload Mass Range Slider

- Let users choose a specific payload mass range.

Scatter Plot: Payload Mass vs. Success Rate by Booster Version

- Provide users with insights into the correlation between payload mass and launch success across different booster versions.



https://github.com/Keliu89/Capstone_Project_SpaceX-s/blob/main/Module%203%20-%20Interactive%20Visual%20Analytics%20and%20Dashboard/8.%20Hands-on%20Lab.%20Build%20an%20Interactive%20Dashboard%20with%20Ploty%20Dash.ipynb

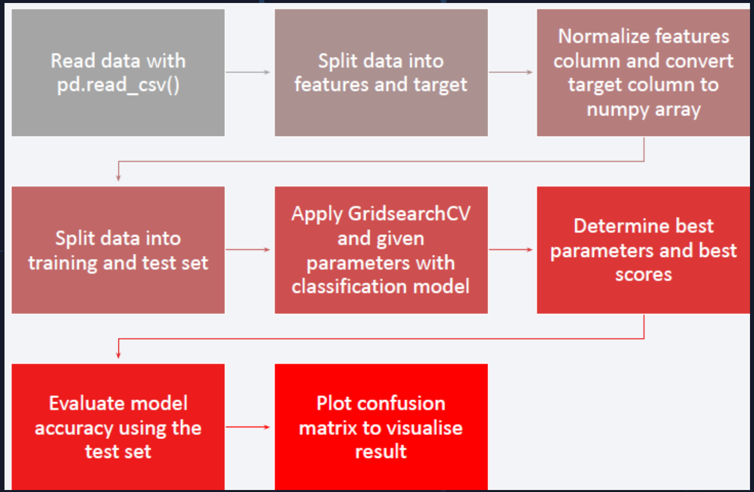
PREDICTIVE ANALYTICS

- ✓ Convert the "Class" column into a NumPy array.
- ✓ Standardize the data using StandardScaler, fitting and transforming the data.
- ✓ Split the dataset into training and testing subsets using train_test_split.
- ✓ Set up a GridSearchCV object with cv=10 for parameter tuning.
- ✓ Apply GridSearchCV to various algorithms, including:
- ✓ Logistic Regression (LogisticRegression())
- ✓ Support Vector Machine (SVC())
- ✓ Decision Tree (DecisionTreeClassifier())
- ✓ K-Nearest Neighbors (KNeighborsClassifier())
- ✓ Calculate the accuracy of each model on the test data using .score().
- ✓ Evaluate each model's performance using the confusion matrix.
- ✓ Identify the best model by comparing the Jaccard Score, F1 Score, and Accuracy.

PREDICTIVE ANALYTICS (Classification)

- ✓ The Figure shows a flow chart of building and evaluating the classification models
- ✓ The URL below is the GITrepository containing the Jupyter Notebook

https://github.com/Keliu89/Capstone_Project_SpaceX-s/blob/main/Module%204%20-%20Predictive%20Analysis%20(Classification)/9.%20Hands-on%20Lab.%20Complete%20the%20Machine%20Learning%20Prediction%20lab.ipynb



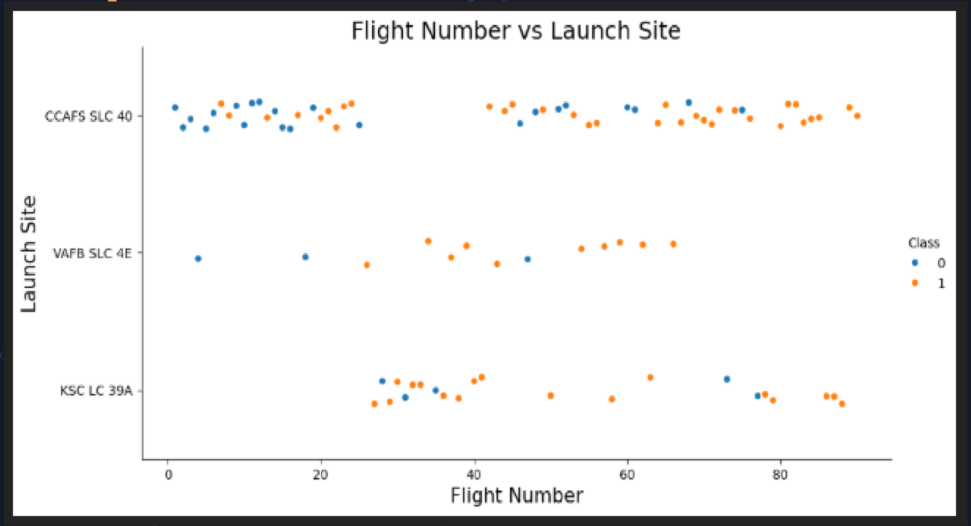
Flow chart of Model development and Evaluation process

RESULTS

RESULTS SUMMARY

- Launch success rates show a consistent upward trend.
- **KSC LC-39A** records the highest success rate among launch sites.
- Missions to **ES-L1, GEO, HEO, and SSO** achieved **100% success.**
- **Visual Analytics**
- Most launch sites are **near the equator** and **close to coastlines**.
- Sites are located to **minimize risk to populated areas** while ensuring **logistical accessibility.**
- **Predictive Analytics**
- **Decision Tree** identified as the most accurate predictive model for the dataset.

FLIGHT NUMBER VS. LAUNCH SITE



Insights

Early flights show **lower success rates** (blue = fail).

~50% of launches originated from **CCAFS SLC-40.**

Recent flights demonstrate **higher success rates** (orange = success).

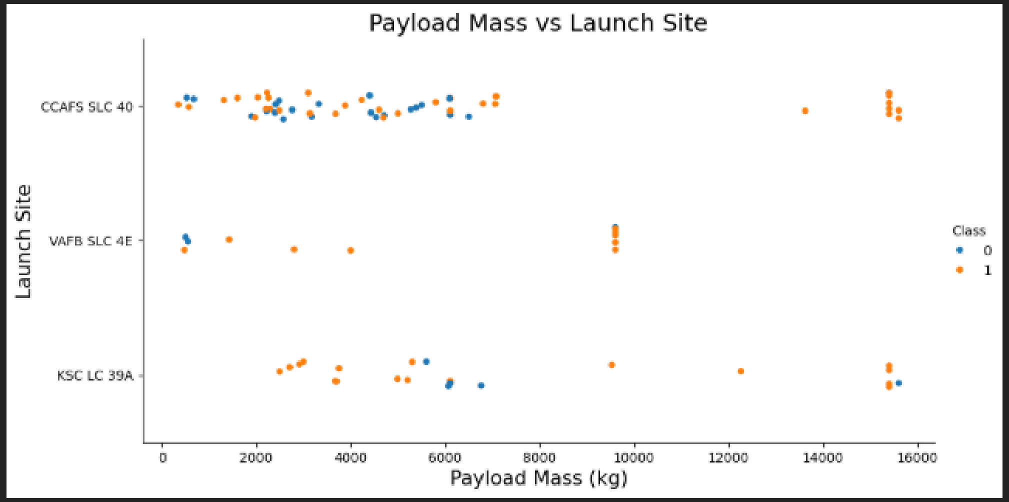
VAFB SLC-4E and **KSC LC-39A** record **higher success rates.**

- The **early flights from CCAFS SLC 40** had more failures, but over time this site achieved a high success rate.
- **VAFB SLC 4E** remained more consistent, with few failures in its history.
- **KSC LC 39A** became consolidated in later stages, associated with more important missions, showing a predominantly successful trend.

PAYLOAD MASS VS. LAUNCH SITE

Insights

- **CCAFS SLC 40** has been the most frequently used site, especially for medium payload launches, with mixed results in the early years.
- **KSC LC 39A** is associated with **critical missions and heavier payloads**, showing a high success rate.
- **VAFB SLC 4E** has played a smaller role but has shown solid performance with intermediate payloads.



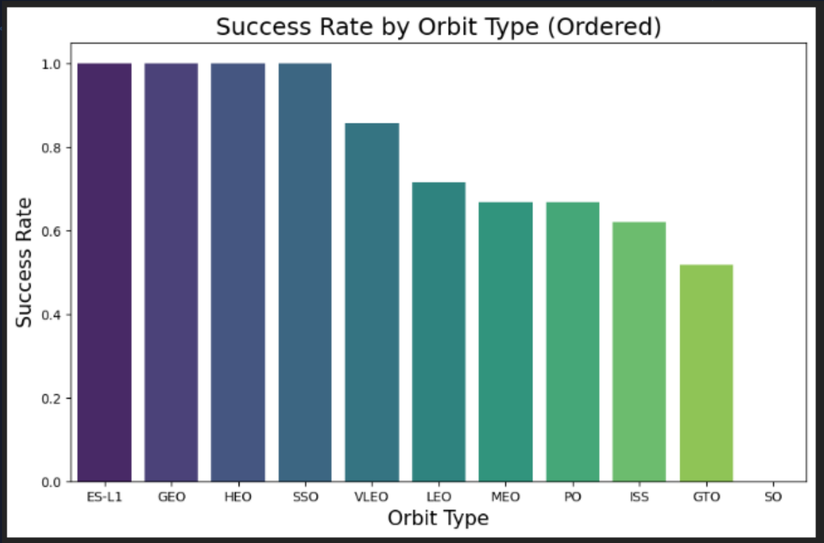
- 👉 Overall, the chart suggests that as SpaceX consolidated its operations, heavier and more complex launches from **KSC LC 39A** and **CCAFS SLC 40** achieved higher success rates.

SUCCESS RATE BY ORBIT TYPE

Insights

100%: ES-L1, GEO, HEO, SSO

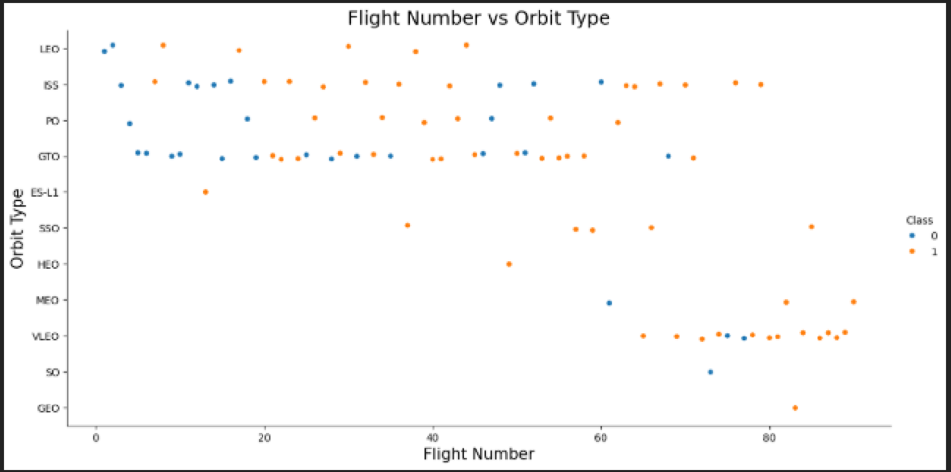
50–80%: GTO, ISS, LEO, MEO, PO



0%: SO

- SpaceX has achieved high success rates in less frequent or mission-specific orbits (ES-L1, GEO, HEO, SSO).
- The most commonly used and demanding orbits (ISS and GTO) show lower rates, reflecting both their technical complexity and the fact that many tests and initial failures occurred in these missions.
- Overall, the trend shows how SpaceX consolidated reliability across most orbits, while still facing challenges in more difficult missions.

FLIGHT NUMBER VS. ORBIT



Insights

Success rate rises with higher flight numbers across most orbits.

Trend is **most evident in LEO.**

GTO does not follow this pattern.

The chart reflects how SpaceX went from a **high failure rate in its early flights** to achieving **consistent success in later missions**.
This evolution highlights the company's **technological maturity** and the refinement of its operations, with increasingly frequent successes in **critical and complex orbits**.

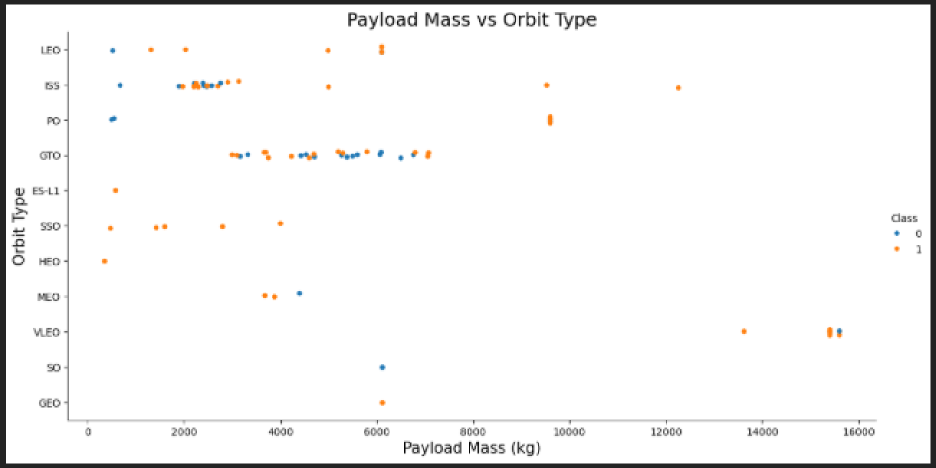
PAYLOAD VS. ORBIT

Insights

Heavy payloads perform best in **LEO, ISS, and PO** orbits.

Trend is most evident in LEO.

GTO does not follow this pattern.



•SpaceX achieved greater consistency in heavier payloads and critical missions (GTO, LEO, ISS).
•The initial failures were concentrated in lower-payload missions, reflecting the experimental phase.
•Over time, the success rate stabilized and expanded to more demanding orbits, demonstrating technological maturity in launches of all types.

LAUNCH SUCCESS OVER TIME

Success rate **improved** during **2013–2017** and **2018–2019.**

Success rate **declined** during **2017–2018** and **2019–2020.**

Overall upward trend since 2013.

Data confirms **a steady increase in success rates up to 2020.**

The graph shows how SpaceX went from a period of initial failures (2010–2013) to achieving high reliability (from 2017 onward). The trend highlights the company's technological and operational maturity, ultimately establishing itself as a leader in reusable space launches.

ALL LAUNCH SITE NAMES

Launch Site Names

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

Explanation

The query retrieves the distinct launch site names from SPACEXTBL.

```
# Consulta
query_task_1 = """
SELECT DISTINCT Launch_Site
FROM SPACEXTBL;
"""
df = pd.read_sql(query_task_1, con)
print(df)
```

	Launch_Site
0	CCAFS LC-40
1	VAFB SLC-4E
2	KSC LC-39A
3	CCAFS SLC-40

The dataset shows three main launch sites, with a minor naming variation for Cape Canaveral (LC vs. SLC).

LAUNCH SITE NAMES BEGIN WITH 'KSC'

Explanation

The query retrieves the first five records from the table SPACEXTBL where the launch site begins with "KSC". This pattern corresponds to launches carried out at the Kennedy Space Center (KSC), specifically at Launch Complex 39A.

```
query_task_2 = """
SELECT *
FROM SPACEXTBL
WHERE "Launch_Site" LIKE 'KSC%'
LIMIT 5
"""
df = pd.read_sql(query_task_2, con)
df.head(5)
```

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landin
0	2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success
1	2017-03-16	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attem
2	2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success
3	2017-05-01	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success
4	2017-05-15	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attem

All five retrieved launches occurred at **KSC LC-39A**, confirming it as one of the primary sites for SpaceX missions. The payloads include both commercial and governmental missions (e.g., *NASA CRS*, *EchoStar*, *SES*, *NRO*, *Inmarsat*), with payload masses ranging from 2,490 kg to 8,070 kg.

TOTAL AND AVERAGE PAYLOAD MASS

Total Payload Mass

48,213 kg carried by boosters launched under NASA CRS.

```
query_task_3 = """
SELECT SUM(PAYLOAD_MASS__KG_) AS Total_Payload_Mass_KG
FROM SPACEXTBL
WHERE Customer LIKE '%NASA (CRS)%';
"""
df_3 = pd.read_sql(query_task_3, con)
df_3.head()
```

	Total_Payload_Mass_KG
0	48213

Average Payload Mass

2,928.4 kg average carried by F9 v1.1 boosters.

```
query_task_4 = """
SELECT AVG(PAYLOAD_MASS__KG_) AS Average_Payload_Mass_KG
FROM SPACEXTBL
WHERE Booster_Version = 'F9 v1.1';
"""
df_4 = pd.read_sql(query_task_4, con)
df_4.head()
```

	Average_Payload_Mass_KG
0	2928.4

FIRST SUCCESSFUL GROUND LANDING DATE

Explanation

The query retrieves the earliest (minimum) date in which the landing outcome was recorded as a **successful drone ship landing**.

The first successful Falcon 9 landing on a drone ship occurred on **April 8, 2016**, marking a milestone in SpaceX's reusable rocket program.

```
query_task_5 = """
SELECT MIN(Date) AS First_Successful_Drone_Ship_Landing
FROM SPACEXTBL
WHERE "Landing_Outcome" = 'Success (drone ship)';
"""

df_5 = pd.read_sql(query_task_5, con)
df_5.head()
```

	First_Successful_Drone_Ship_Landing
0	2016-04-08

SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD BETWEEN 4000 AND 6000

Explanation

This query retrieves the distinct booster versions that successfully landed on the **ground pad** with payload masses between **4,000 kg and 6,000 kg**.

Three boosters — **B1032.1, B1040.1, and B1043.1** — meet the specified conditions, highlighting their operational success in medium payload launches.

```
query_task_6 = """
SELECT DISTINCT Booster_Version
FROM SPACEXTBL
WHERE "Landing_Outcome" = 'Success (ground pad)'
  AND PAYLOAD_MASS__KG_ > 4000
  AND PAYLOAD_MASS__KG_ < 6000;
"""

df_6 = pd.read_sql(query_task_6, con)
df_6.head()
```

	Booster_Version
0	F9 FT B1032.1
1	F9 B4 B1040.1
2	F9 B4 B1043.1

TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

Query Output

Failure (in flight): 1Success: 98Success (payload status unclear): 1Success (with some note): 1

Totaling

Successful Missions:98+1+1=100
Failed Missions:1

Totaling

Successful Missions:98+1+1=100
Failed Missions:1


```
query_task_7 = """
SELECT Mission_Outcome, COUNT(*) AS total
FROM SPACEXTBL
GROUP BY Mission_Outcome;
"""

df_7 = pd.read_sql(query_task_7, con)
df_7.head()
```

	Mission_Outcome	total
0	Failure (in flight)	1
1	Success	98
2	Success	1
3	Success (payload status unclear)	1

So, out of all missions, 100 were successful (including unclear status but marked as success) and 1 mission failed.

BOOSTERS

Booster_versions that have carried the maximum payload

```
query_task_8 = """
SELECT DISTINCT Booster_Version
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (
    SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEXTBL
);
"""

df_8 = pd.read_sql(query_task_8, con)
df_8.head()
```

	Booster_Version
0	F9 B5 B1048.4
1	F9 B5 B1049.4
2	F9 B5 B1051.3
3	F9 B5 B1056.4
4	F9 B5 B1048.5

- The boosters that carried the **maximum payload mass** in the dataset all belong to the **Falcon 9 Block 5 (F9 B5) version.**
- This shows that the **technological improvements** in the Block 5 version allowed for **greater payload capacity** compared to earlier versions (F9 FT, F9 B4, etc.).
- We also see that the same booster (e.g., **B1048**) appears multiple times (**B1048.4 and B1048.5**), which demonstrates **successful rocket reusability**—a key factor in SpaceX's strategy to reduce launch costs.

- The largest payloads were carried by reused Block 5 boosters, highlighting their success in terms of capacity, reliability, and reusability

2017 LAUNCH RECORDS

- All successful ground pad landings in 2017 took place at Kennedy Space Center Launch Complex 39A (KSC LC-39A).
- During that year, we can also observe the transition from the Falcon 9 Full Thrust (F9 FT) boosters to the Falcon 9 Block 4 (F9 B4) version.

👉 This highlights two key points:

- SpaceX consistently achieved reliable ground pad landings throughout 2017.
- The company demonstrated technological progress by upgrading its boosters while maintaining mission success. 🚀

```
query_task_9 = """
SELECT substr(Date,6,2) AS Month,
       Booster_Version,
       Launch_Site,
       "Landing_Outcome"
FROM SPACEXTBL
WHERE substr(Date,1,4) = '2017'
  AND "Landing_Outcome" = 'Success (ground pad)';
"""

df_9 = pd.read_sql(query_task_9, con)
df_9.head()
```

	Month	Booster_Version	Launch_Site	Landing_Outcome
0	02	F9 FT B1031.1	KSC LC-39A	Success (ground pad)
1	05	F9 FT B1032.1	KSC LC-39A	Success (ground pad)
2	06	F9 FT B1035.1	KSC LC-39A	Success (ground pad)
3	08	F9 B4 B1039.1	KSC LC-39A	Success (ground pad)
4	09	F9 B4 B1040.1	KSC LC-39A	Success (ground pad)

COUNT OF SUCCESSFUL LANDINGS

The ranking of landing outcomes by frequency is:

- No attempt → 10
- Success (drone ship) → 5
- Failure (drone ship) → 5
- Success (ground pad) →3
- Controlled (ocean) → 3

- In the early years (2010–2017), most launches had **no landing attempt**.
- Attempts on drone ships were **risky**, with an equal split between **successes (5)** and **failures (5)**, showing SpaceX was still refining the technology.
- **Ground pad landings (3 successes)** started to become viable during this period.
- Some rockets were deliberately landed in the **ocean (3 controlled ocean landings)**, usually as expendable missions.
- 👉 Overall, this time frame reflects **SpaceX's experimental phase**, gradually moving from no attempts to consistent successes on both ground pads and drone ships.

```
query_task_10 = """
SELECT "Landing_Outcome",
       COUNT(*) AS outcome_count
FROM SPACEXTBL
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY "Landing_Outcome"
ORDER BY outcome_count DESC;
"""

df_10 = pd.read_sql(query_task_10, con)
df_10.head()
```

	Landing_Outcome	outcome_count
0	No attempt	10
1	Success (drone ship)	5
2	Failure (drone ship)	5
3	Success (ground pad)	3
4	Controlled (ocean)	3

LAUNCH SITE ANALYSIS

LAUNCH SITES

With Markers

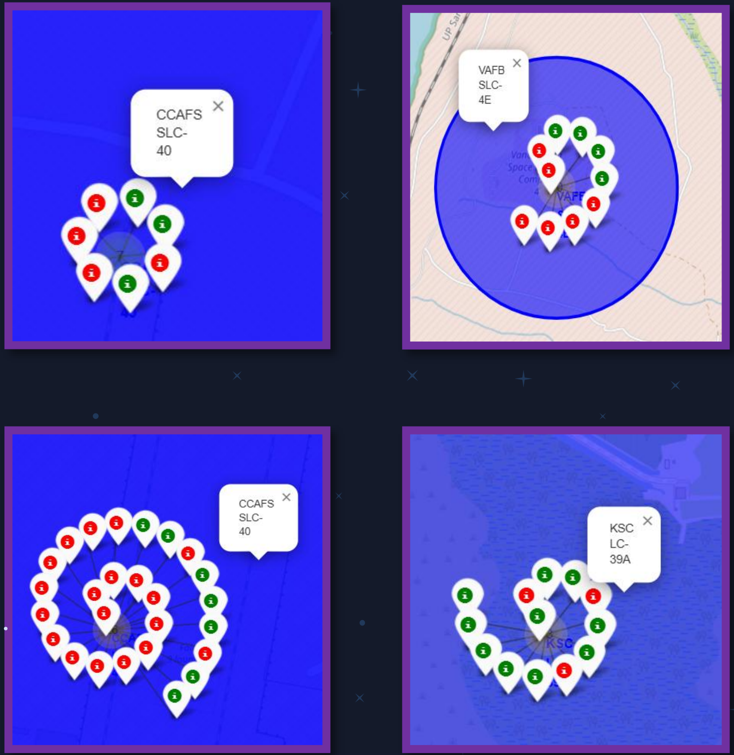
Near the Equator: The closer a launch site is to the equator, the easier it becomes to reach an equatorial orbit, as Earth's rotation provides additional momentum for a prograde trajectory.

Rockets launched from these locations benefit from the planet's rotational speed, which gives them a natural boost and reduces the need for extra fuel or boosters.

LAUNCH OUTCOMES

The figure shows the launch outcomes for various launch sites;

- Top left: CCAFS SLC-40
- Top right: VAFB SLC-4E
- Bottom left: CCAFS LC-40
- Bottom Right: KSC LC-39A

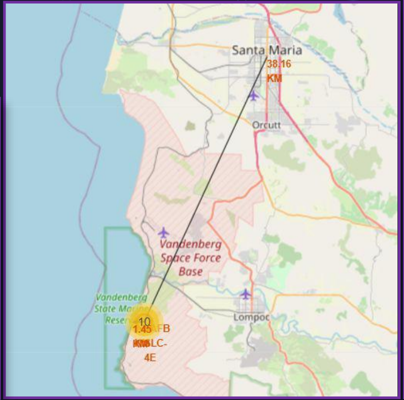


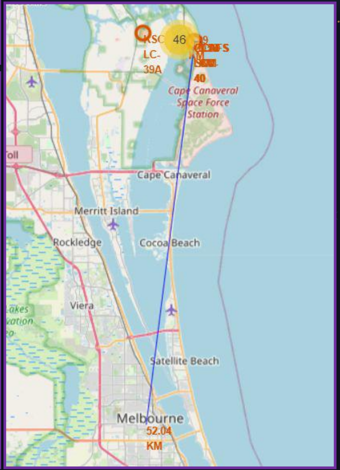
Launch site **KSC LC-39A** has a **10/13 success rate** (76.9%)

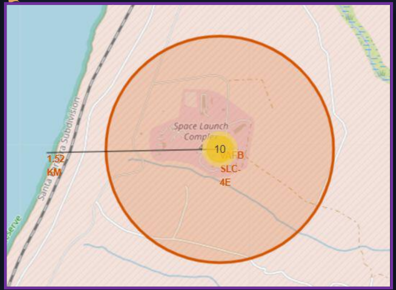
❖ Green markers: successful launches.
❖ Red markers: failed launches.

DISTANCE TO PROXIMITIES

- The figure demonstrates that the launch sites are situated in close proximity to the coastline, with distances of approximately 0.95 km from Cape Canaveral Space Force Station (CCAFS) Launch Complex 40 (SLC-40) and 1.52 km from Vandenberg Space Force Base (VAFB) Launch Complex 4E (SLC-4E).
- In contrast, certain railway lines and highways are not located as near to the launch facilities.
- Furthermore, it is evident that these launch sites are positioned at significant distances from major urban centers. Specifically, VAFB SLC-4E is located approximately 38.16 km from the nearest city, Santa Maria, while CCAFS SLC-40 is about 56.04 km from Melbourne, Florida.







DASHBOARD WITH PLOTLY

PIE CHART OF LAUNCH SUCCESS FOR ALL SITES

As illustrated in Figure 1, the Kennedy Space Center Launch Complex 39A (KSC LC-39A) demonstrates the highest success rate, contributing approximately 41.7% to the overall distribution of successful launches across the analyzed sites.

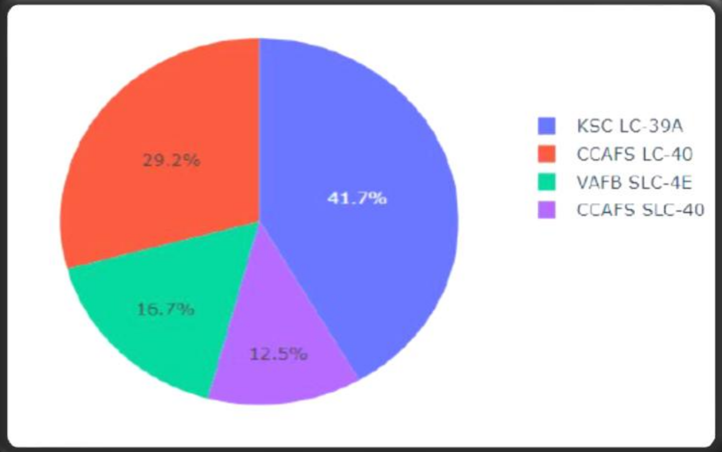


Figure 1. Pie showing the Success rate of all Launch sites

LAUNCH SUCCESS

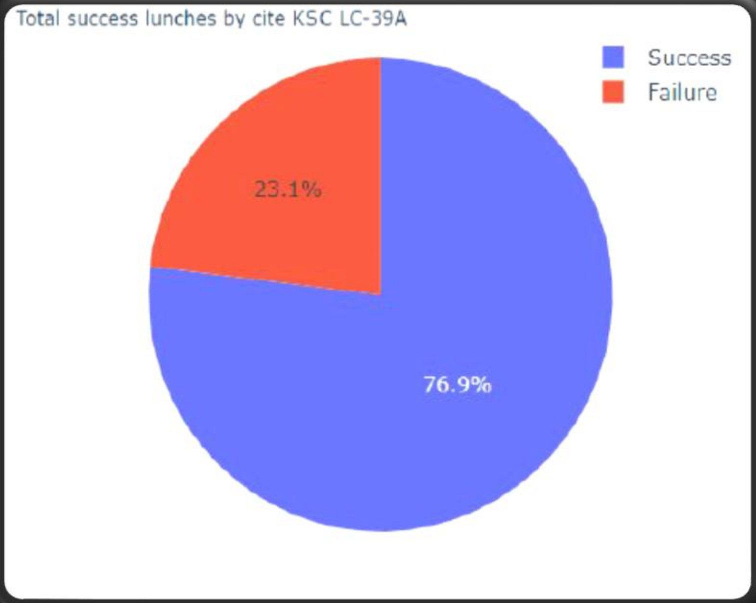


Figure 2. Pie showing the launch site with the highest success ratio

KSC LC-39A

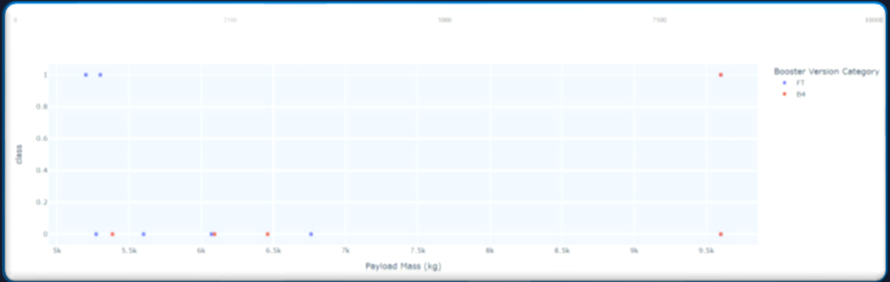
Figure 2 shows that KSC LC-39A achieves the highest success rate, approximately 76.9%, which is greater than that observed at the other launch sites.

10 successful launches and 3 failed launches

PAYLOAD VS LAUNCH OUTCOME FOR ALL SITES

- From the figures below, Booster version FT has the highest success rate with its payload mass of about between 700kg to 5,500kg.
- It is also shown that rockets with payload mass above 5,500kg have a lower success rate, which means the heavier the payload, the slimmer the chance of a successful outcome.

Scatter plot showing the booster versions with different payload mass for all launch sites



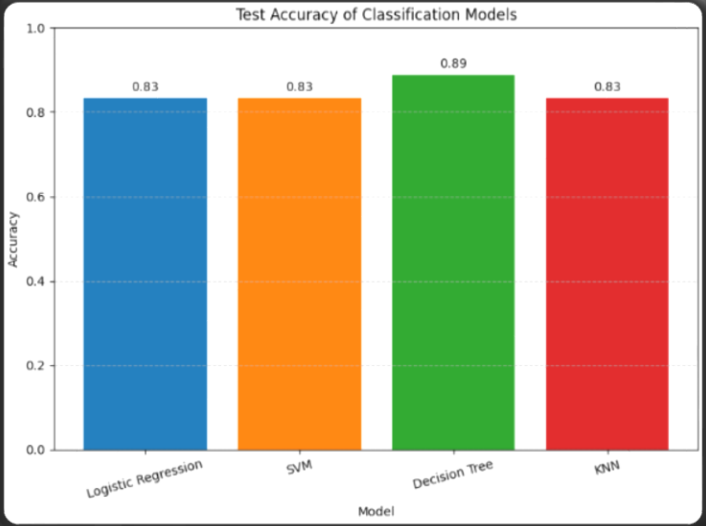
Scatter plot showing the booster versions of different payload mass greater than 5,500kg.

DESCRIPTIVE ANALYSIS (CLASSIFICATION)

CLASSIFICATION ACCURACY

1. Visualization of Model Accuracies

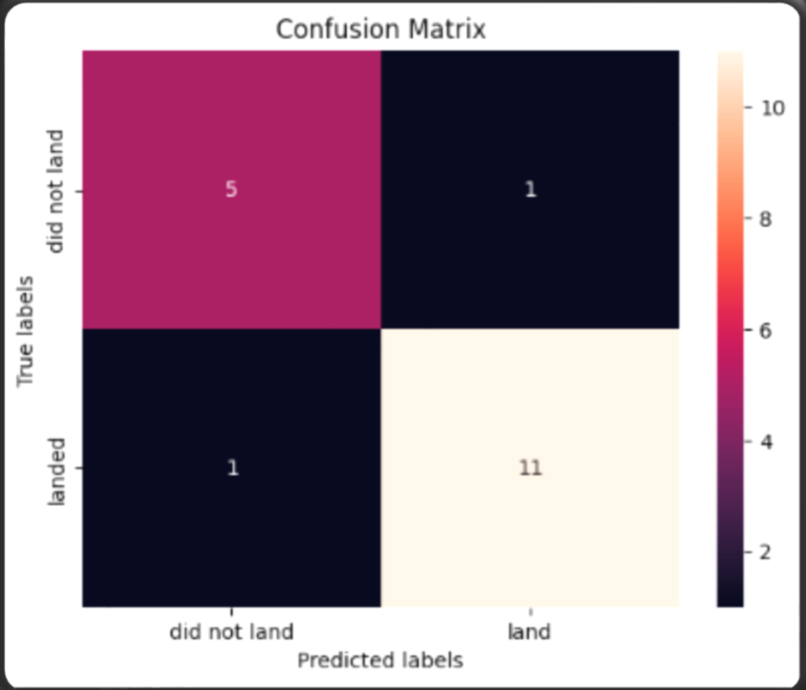
- The classification model performance has been visualized using a bar chart, where the test accuracy of each model is plotted. This visualization provides a clear comparison of the models' predictive performance.



2. Best Performing Model

Among the evaluated models, the **Decision Tree** classifier demonstrates the highest test accuracy, achieving **0.89**. This indicates that, under the given dataset and evaluation setup, the Decision Tree outperforms Logistic Regression, SVM, and KNN in terms of classification accuracy.

CONFUSION MATRIX



Interpretation

- True Negatives (TN = 5): The model correctly predicted 5 cases where the event did not land.
- False Positives (FP = 1): The model incorrectly predicted "landed" in 1 case where it did not actually land.
- False Negatives (FN = 1): The model incorrectly predicted "did not land" in 1 case where it actually landed.
- True Positives (TP = 11): The model correctly predicted 11 cases where the event landed.

CONCLUSIONS

Conclusions

- A higher number of flights at a given launch site is associated with a higher success rate.
- Launch success rates have shown a consistent increase from 2013 to 2020.
- Missions targeting orbits such as ES-L1, GEO, HEO, SSO, and VLEO achieved the highest success rates.
- KSC LC-39A recorded the largest number of successful launches among all sites.
- Among the evaluated algorithms, the Decision Tree classifier demonstrated the best performance for predicting landing outcomes.

Key Determinants of Mission Success

- **Payload Mass:** Missions with smaller payloads tended to achieve higher success rates; however, for KSC LC-39A, launches with payloads below 5,500 kg achieved a 100% success rate.
- **Orbit Type:** Launches to ES-L1, GEO, HEO, and SSO exhibited consistently high success rates, reaching 100% in the dataset analyzed.
- **Launch Site Location:** Most sites are situated in coastal areas to facilitate retrieval and recovery operations, while also being located away from densely populated regions to minimize risk in the event of failure.
- **Temporal Trend:** In recent years, launch outcomes have improved, with later flights demonstrating higher success probabilities.
- **The confusion matrix** confirms that the Decision Tree model not only achieves the highest accuracy overall (0.89) but also demonstrates strong recall and precision, making it a robust choice for this classification task.

Observed Patterns

- **Geographical Advantage:** Many launch sites are positioned near the equator, leveraging Earth’s rotational velocity to reduce fuel and booster requirements.
- Coastal Positioning: All analyzed sites are located close to coastlines, reinforcing their strategic placement for safety and recovery.
- **Success Over Time:** Success rates have increased steadily, reflecting technological and operational improvements.

Thanks!