SpaceX Falcon 9 first stage landing prediction

LIDIA DANILOV,

23.03.2025

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

Overview

- The SpaceX Falcon 9 launch system is a partially reusable rocket designed to deliver payloads into orbit while enabling cost reduction through booster recovery. Predicting the first stage landing location and conditions is crucial for mission planning, safety, and operational efficiency.

Objective

- The primary objective of Falcon 9 first stage landing prediction is to accurately estimate the booster’s touchdown location, taking into account various factors such as launch trajectory, atmospheric conditions, propulsion parameters, and landing site constraints. Achieving precise predictions enhances the success rate of recovery missions and improves the reusability of boosters.

INTRODUCTION

- This capstone project focuses on predicting the successful landing of the Falcon 9 rocket's first stage, a critical factor that directly influences the cost efficiency of SpaceX launches. This is framed as a binary classification problem, where the target variable is whether the first stage landed successfully (1) or failed to land (0).
- The outcome of this project will not only enhance cost estimation accuracy but also support strategic decision-making in the commercial spaceflight industry.
- Historical launch data from SpaceX were used, which includes information such as payload mass, orbit type, rocket version, landing type, launch site, and weather conditions. The data was sourced from public SpaceX records and supplemented with external data from NASA and other open-source APIs where applicable.

METHODOLOGY

1. Data collection, using SpaceX API
2. Data cleaning, using:
   - Data Wrangling
   - Web scraping
3. Exploratory data analysis (EDA), using:
   - SQL
   - Pandas and Matplotlib
4. Data visualization, using:
   - Folium
   - Dash
5. Machine learning prediction

1. Data collection using SpaceX API

- To gain a deeper understanding of SpaceX's Falcon 9 launch history, the **SpaceX API** (https://api.spacexdata.com/v4/rockets/) was used to retrieve detailed information about all SpaceX rockets.
- The API provides extensive data on various SpaceX rockets, including Falcon 1, Falcon 9, Falcon Heavy, and Starship. However, this analysis focused specifically on **Falcon 9** launches:

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1060	-80.603956	28.608058
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	13	B1058	-80.603956	28.608058
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1051	-80.603956	28.608058
92	89	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e534e7cc	5.0	12	B1060	-80.577366	28.561857
93	90	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca	5.0	8	B1062	-80.577366	28.561857

90 rows × 17 columns

2. Data cleaning

- Data Wrangling - Every missing value in the data is replaced with the mean of the column it belongs to.
- Web scrap from: List_of_Falcon_9_and_Falcon_Heavy_launches
- The picture below shows the first few rows of the data:

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10

3. Exploratory Data Analysis - SQL

- Diving into SpaceX's historical launch data, I needed to uncover key insights about launch sites, mission success rates, and the most powerful boosters. I used SQL for efficiently extracting and analyzing this critical information.
- The first step was to understand the launch sites. Using SQL, I queried the database to list all the unique launch sites used by SpaceX. This helped me identify their geographical distribution and frequency of use.
- The picture below ilustrates the unique launch sites name:

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

3. Exploratory Data Analysis (1) - SQL

- With a clear view of the launch sites, the attention turned to mission outcomes. The total number of successful and failed missions is 101.
- Next step was to identify the booster versions that carried the maximum payload mass. To achieve this, I wrote an SQL query to retrieve the names of the boosters that transported the heaviest payloads.
- The query revealed the most powerful booster versions ever used by SpaceX, showcasing their evolving engineering capabilities, like it is shown in the image from the side:

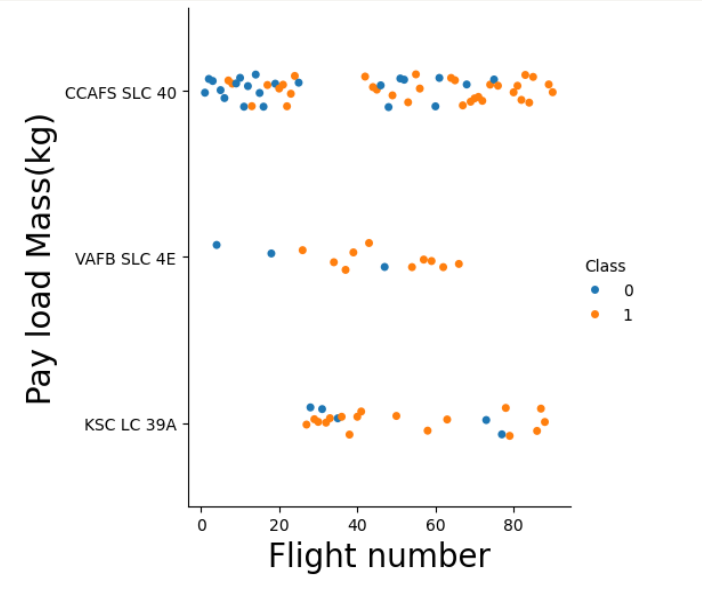
Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

3. Exploratory Data Analysis (2) - Pandas and Matplotlib

- Using **Pandas** for data manipulation and **Matplotlib** for visualization, some patterns that influence launch success were revealed.

1. Flight Numbers and Launch Sites: Identifying Trends

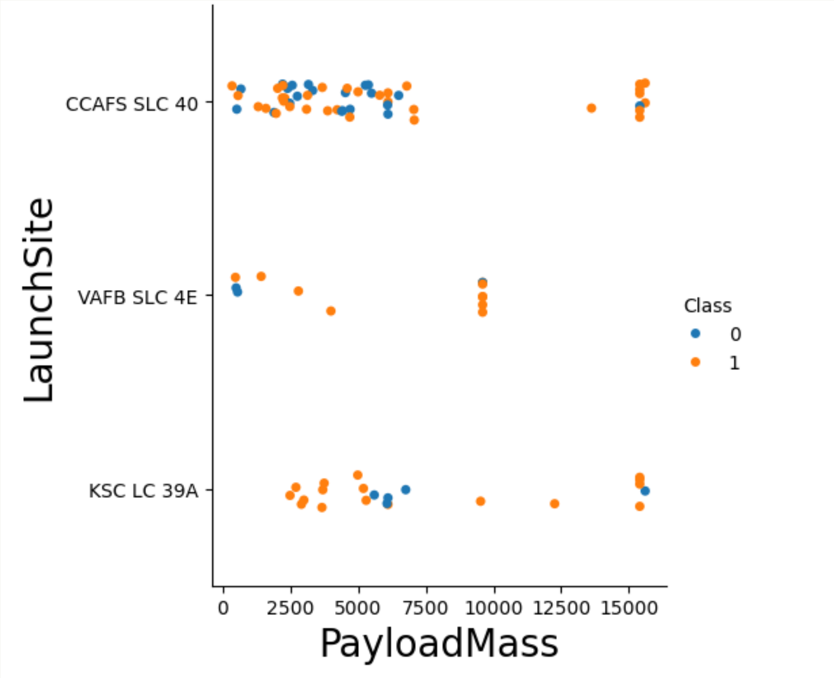
- The first visualization examined the relationship between **Flight Number and Launch Site**. Plotting this helps us understand which sites handled the most launches and how frequently they were used. I observed that CCAFS SLC 40 and KSC LC 39A had the highest flight numbers:



3. Exploratory Data Analysis (3) - Pandas and Matplotlib

2. Payload Mass vs. Launch Site: Analyzing Capacity

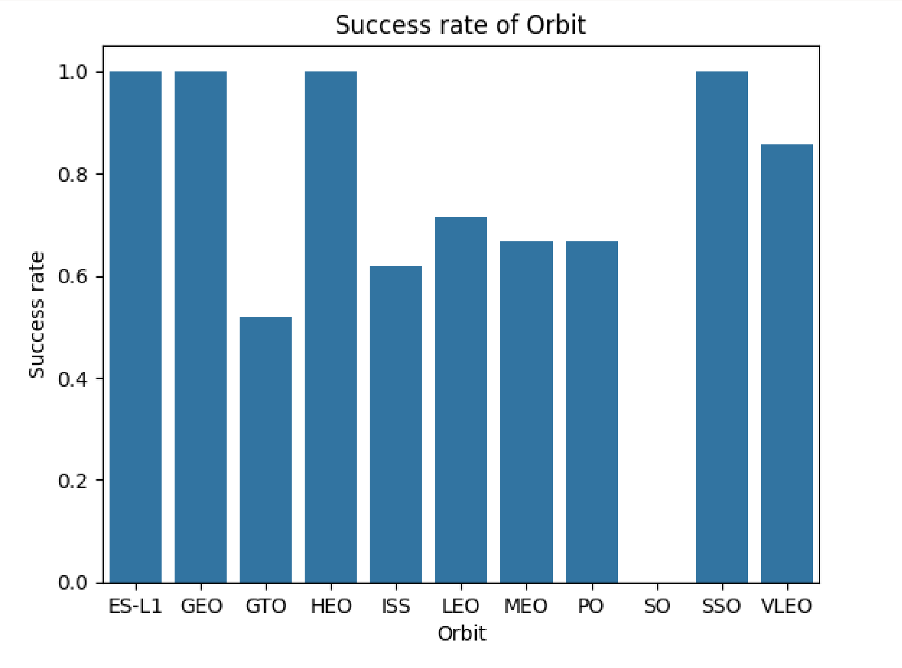
- Next, I explored the **Payload Mass for each Launch Site** to determine how different locations handled varying payload capacities. The visualization revealed that some sites specialized in handling heavier payloads, talking again about CCAFS SLC 40 and KSC LC 39A:



3. Exploratory Data Analysis (4) - Pandas and Matplotlib

3. Success Rate Across Orbit Types

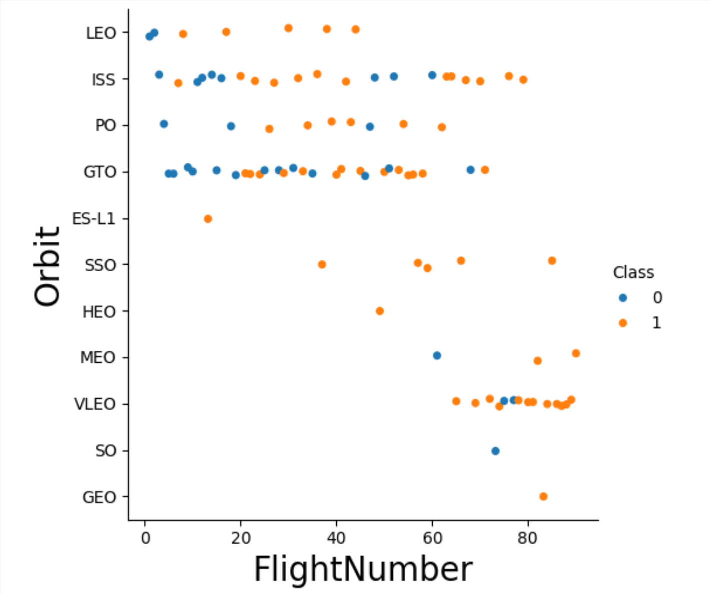
- The results showed that SSO, HEO, GEO and ES-L1 orbits had the highest success rates, whereas GTO orbit presented slightly lower:



3. Exploratory Data Analysis (5) - Pandas and Matplotlib

4. Flight Number vs. Orbit Type: Tracking Evolution

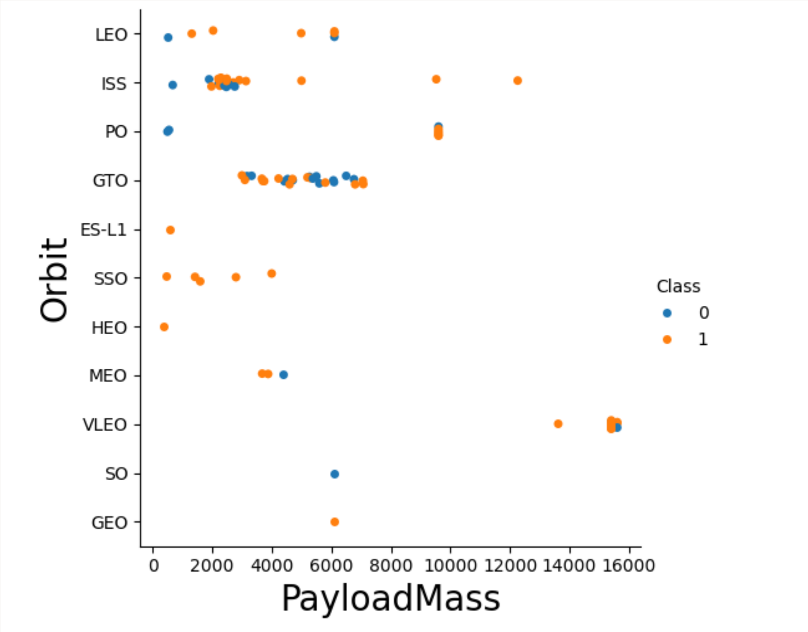
- Understanding how **Flight Number correlates with Orbit Type** provided valuable insights into SpaceX's progression. It can be observed that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success:



3. Exploratory Data Analysis (6) - Pandas and Matplotlib

5. Payload Mass vs. Orbit Type: Examining Performance

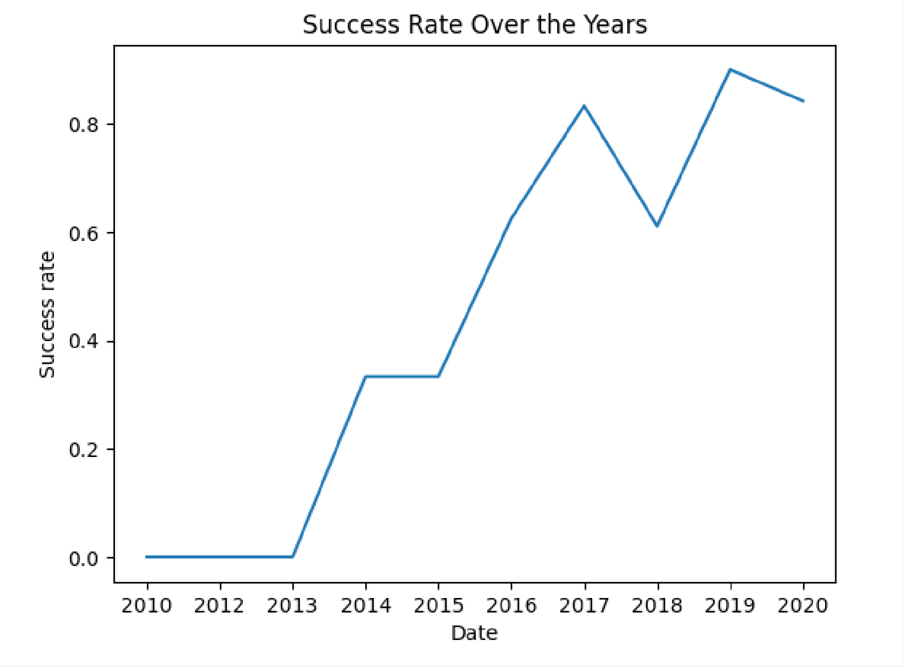
- By plotting **Payload Mass against Orbit Type**, I observed that with heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS. However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.



3. Exploratory Data Analysis (7) - Pandas and Matplotlib

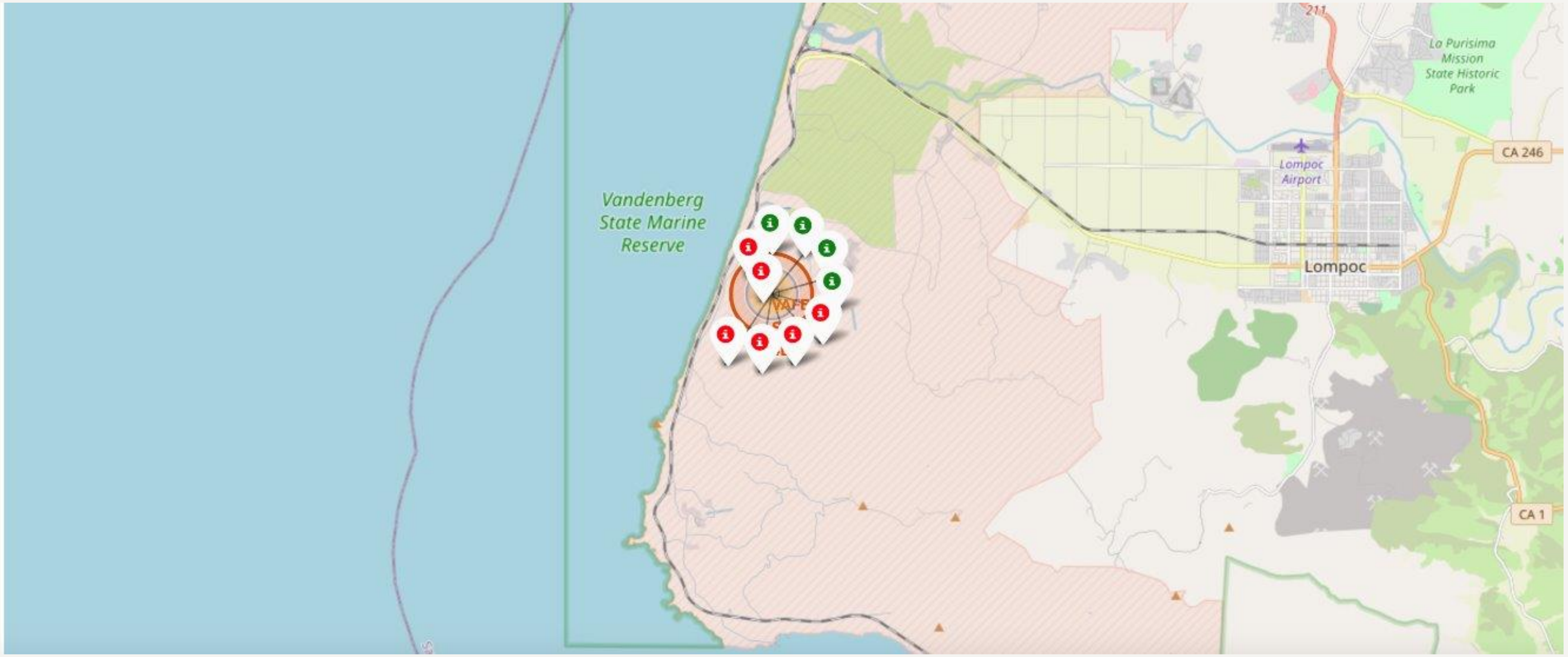
6. Yearly Launch Success Trends: Measuring Progress

- Finally, I visualized **launch success trends over the years**, revealing an impressive growth trajectory. It can be observed that the sucess rate since 2013 kept increasing untill 2020.



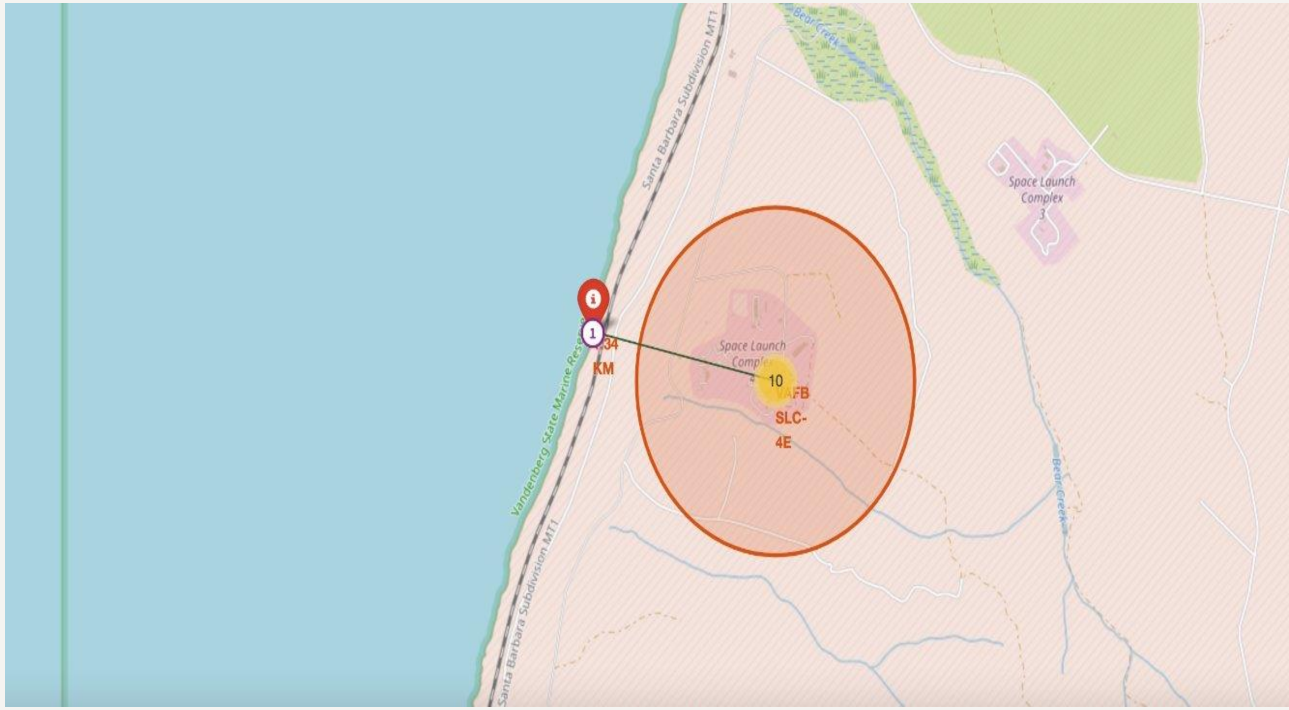
4. Data visualization - Folium

- Using **Folium**, I created an **interactive map** that displayed every Falcon 9 launch site, with **color-coded markers** indicating mission outcomes—**green for success and red for failure:**



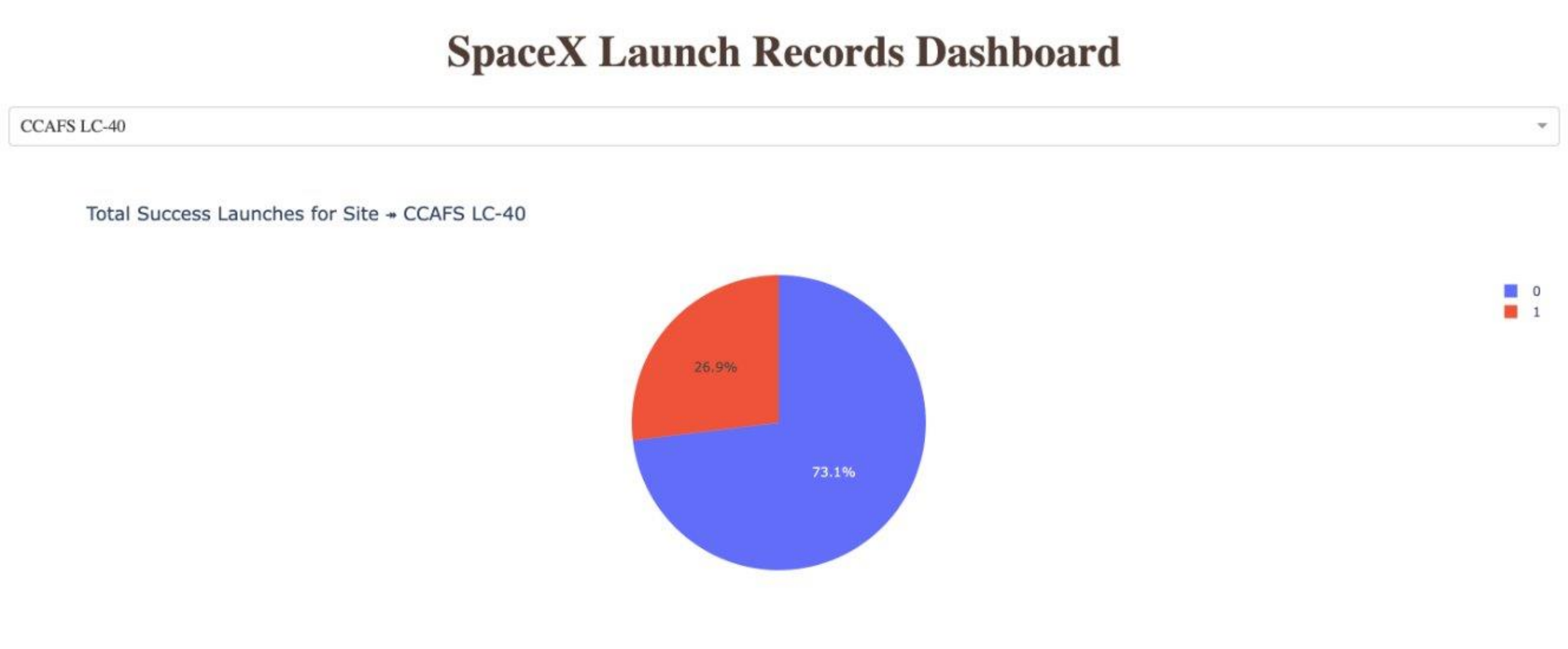
4.Data visualization (1) - Folium

- The distances between a launch site to its proximities such as the nearest city, railway, or highway
- The picture below shows the distance between the VAFB SLC-4E launch site and the nearest coastline:



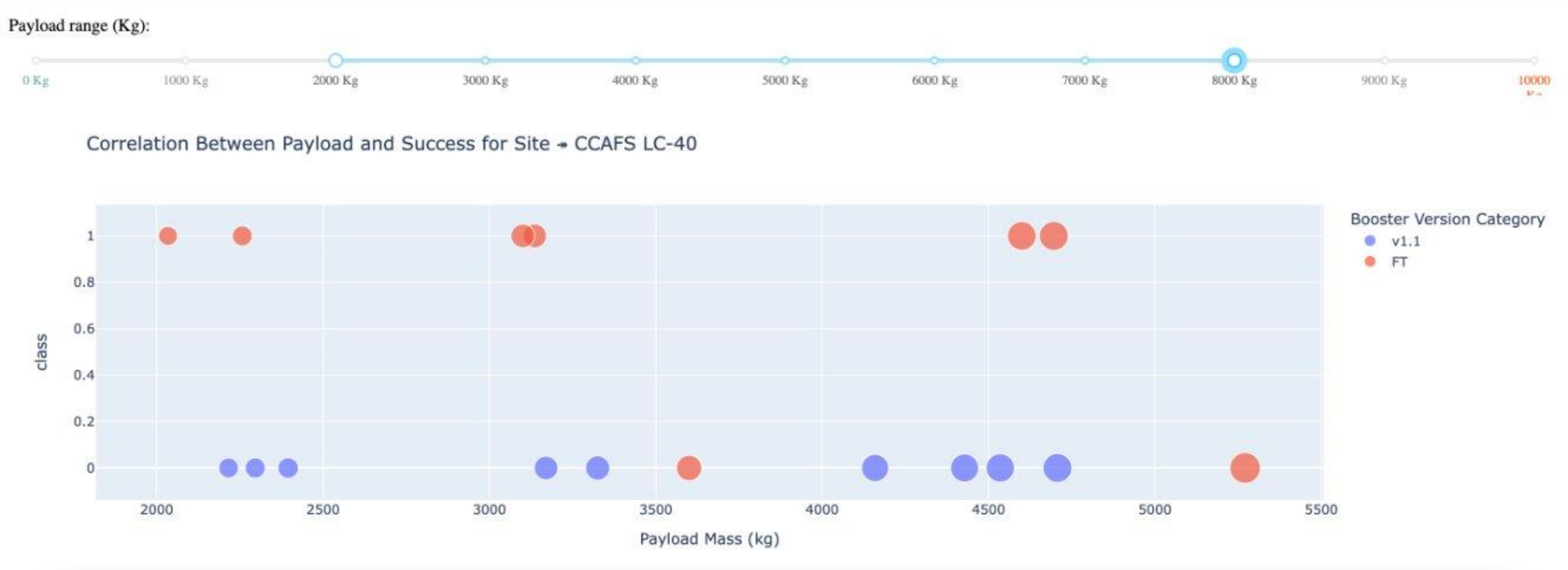
4.Data visualization (3) - Dash

- The picture below shows a pie chart when launch site CCAFS LC-40 is chosen.
- 0 represents failed launches while 1 represents successful launches. We can see that 73.1% of launches done at CCAFS LC-40 are failed launches.



4.Data visualization (4) - Dash

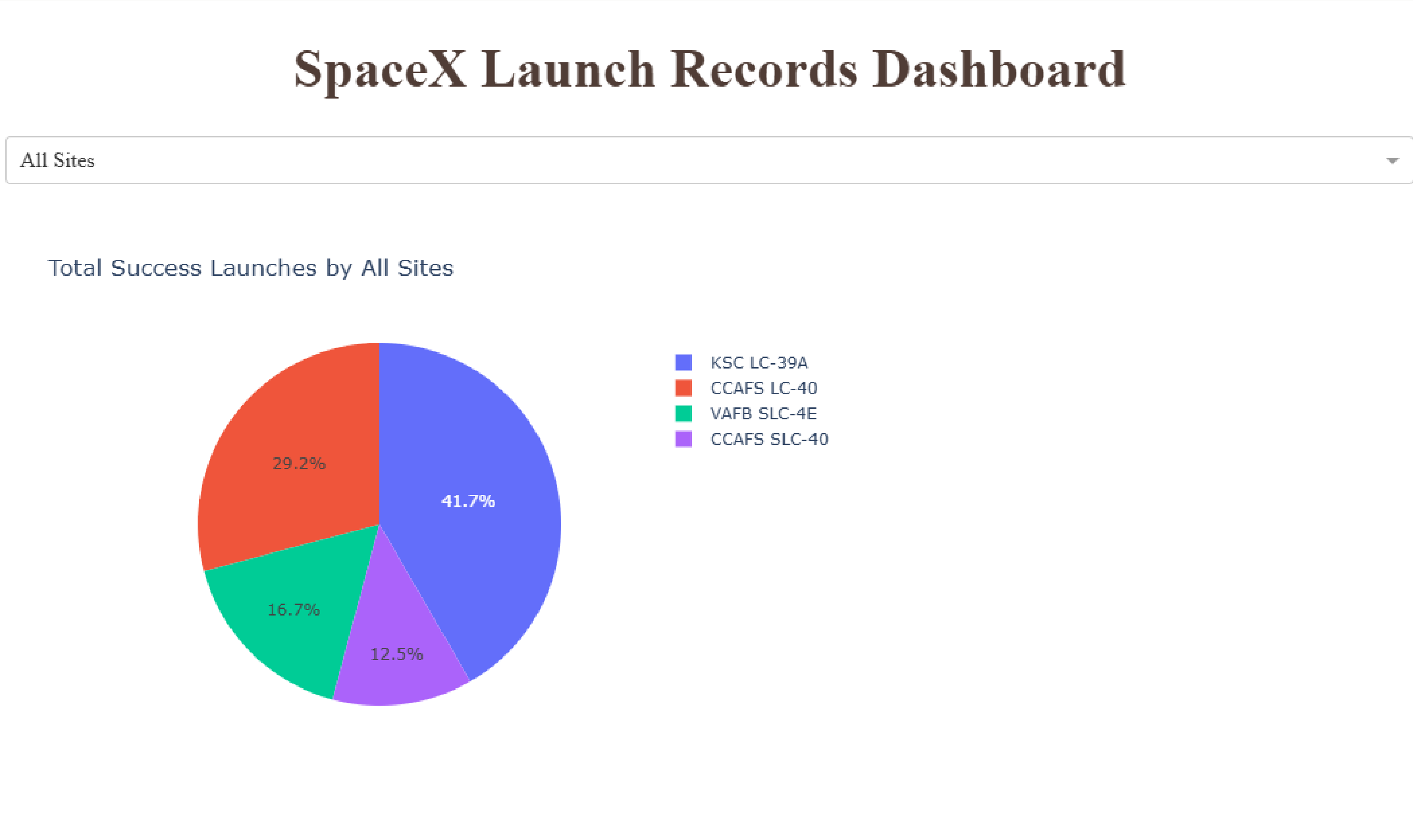
- The picture below shows a scatterplot when the payload mass range is set to be from 2000kg to 8000kg.
- Class 0 represents failed launches while class 1 represents successful launches.



DASHBOARD

<[testrepo/spacex-dash-app.py at main · LidiaVlaicu/testrepo]>

DASHBOARD TAB 1



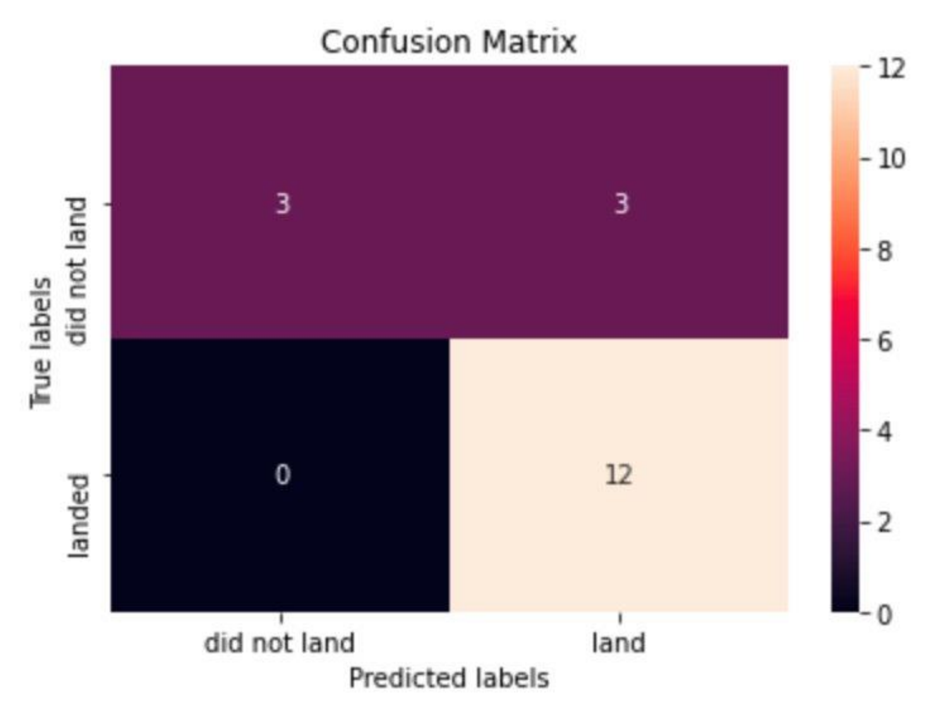
DASHBOARD TAB 2

5. Machine learning prediction

- The main goal is to use a **machine learning model** to predict whether the **first stage of the Falcon 9 rocket would land successfully**. To ensure accuracy, I **split the dataset into training and test sets**, allowing the evaluation of the model performance effectively.
- To optimize predictions, I experimented with different **classification algorithms**, including **Support Vector Machines (SVM), Classification Trees, and Logistic Regression**. By tuning hyperparameters, I identified the best-performing model, refining its ability to distinguish between **successful and failed landings**.
- This machine learning approach provided **data-driven insights into Falcon 9's landing patterns**, paving the way for more reliable and cost-effective spaceflight operations.

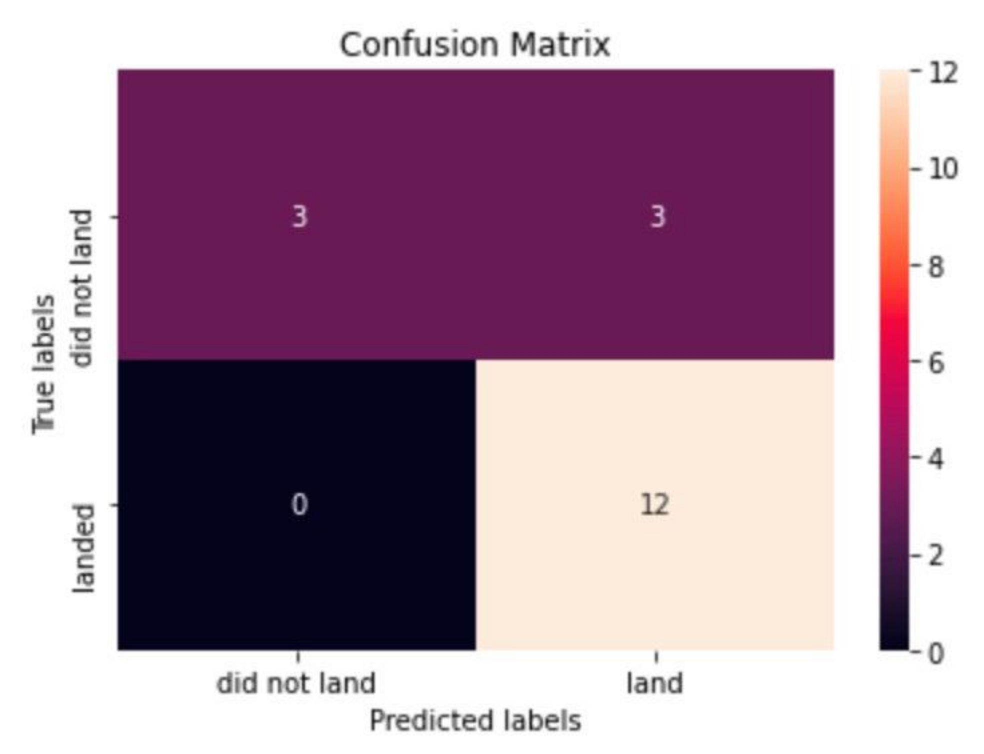
5. Machine learning prediction (1) - Logistic regression

- GridSearchCV best score: 0.8464285714285713
- Accuracy score on test set: 0.8333333333333334
- Confusion matrix:



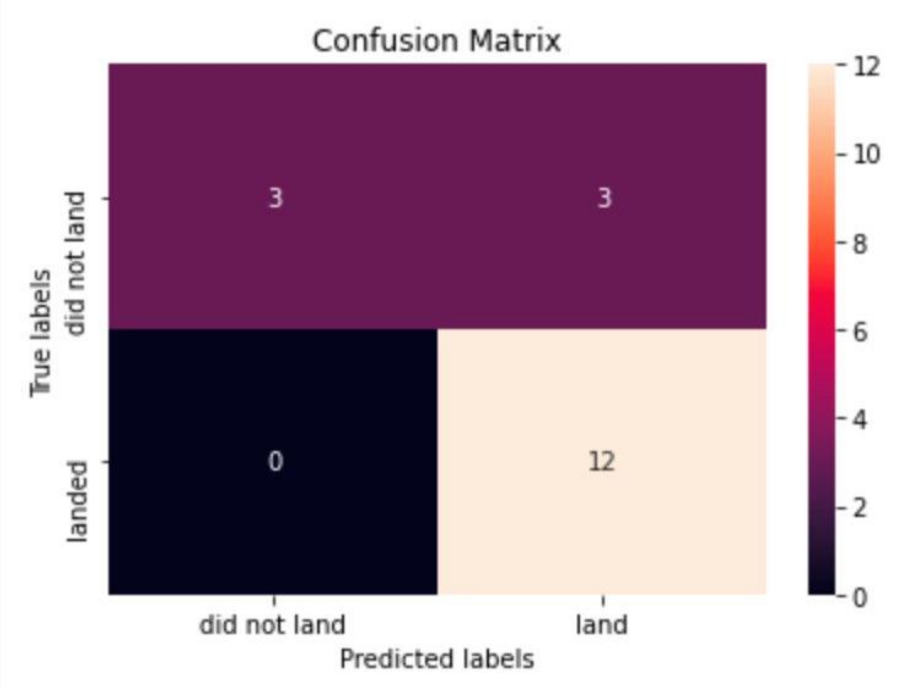
5. Machine learning prediction (2) - SVM

- GridSearchCV best score: 0.8482142857142856
- Accuracy score on test set: 0.8333333333333334
- Confusion matrix:



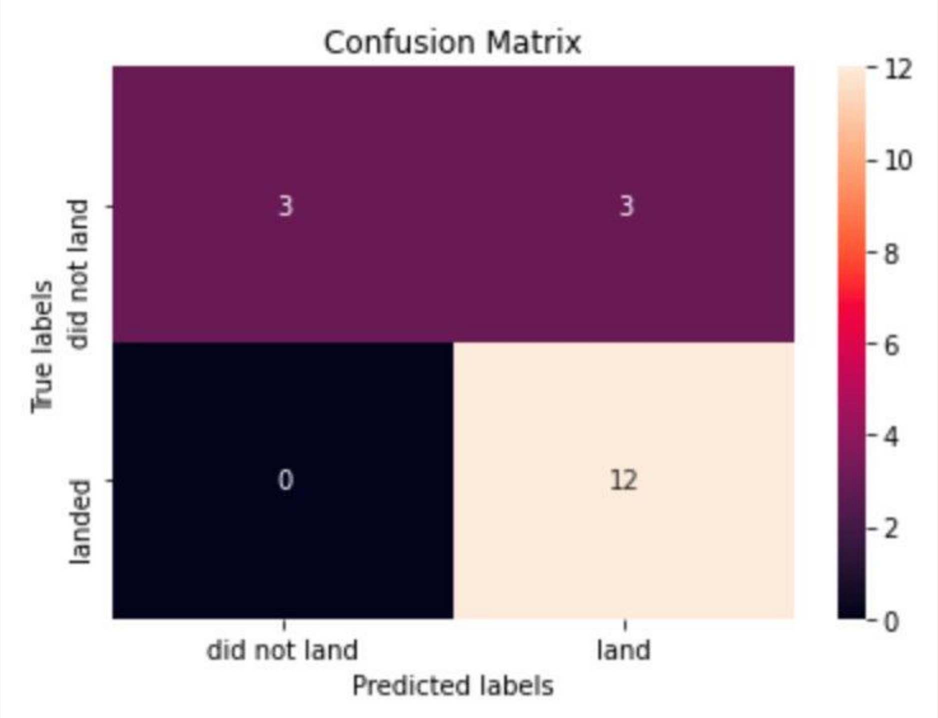
5. Machine learning prediction (3) - Decision tree

- GridSearchCV best score: 0.8892857142857142
- Accuracy score on test set: 0.8333333333333334
- Confusion matrix:



5. Machine learning prediction (4) - KNN

- GridSearchCV best score: 0.8482142857142858
- Accuracy score on test set: 0.8333333333333334
- Confusion matrix:



RESULTS (1)

- After training four different machine learning models to predict the **first-stage landing success of Falcon 9**, I was eager to compare their performance. Initially, all models produced **identical accuracy scores** and confusion matrices on the test set, making direct comparisons difficult.
- To break the tie, I turned to **GridSearchCV best scores**, a metric that captures each model's optimized performance. The results revealed a clear ranking:

1️⃣ **Decision Tree** – **Best performer** with a GridSearchCV score of 0.889
2️⃣ **K-Nearest Neighbors (KNN)** – 0.848
3️⃣ **Support Vector Machine (SVM)** – **0.848** (slightly trailing KNN)
4️⃣ **Logistic Regression** – **0.846**

RESULTS (2)

- With this ranking, it became evident that the **Decision Tree model** was the most effective in predicting Falcon 9 landings, offering the highest optimized score. However, since KNN and SVM performed closely, additional fine-tuning could potentially improve their results.
- This analysis reinforced the importance of **hyperparameter tuning and model selection** in machine learning, proving that accuracy alone isn't always the best measure of performance. By leveraging GridSearchCV, I was able to **differentiate models and select the most reliable approach** for predicting Falcon 9's landing success.

DISCUSSION

- As I explored the **data visualization** section, intriguing patterns emerged. Certain features seemed to correlate with the **success of Falcon 9 landings**, though not in uniform ways. For instance, missions carrying **heavy payloads** to **Polar, LEO, and ISS orbits** demonstrated a **higher success rate**. However, for **GTO missions**, the distinction wasn’t as clear—both successful and failed landings were present, making it harder to draw a definitive conclusion.
- This raised an important question: **How exactly do these features impact the final mission outcome?** While correlations were evident, the complexity of rocket launches made it difficult to pinpoint precise cause-and-effect relationships. Many factors—such as weather conditions, launch site, and booster reuse—interact in ways that aren’t immediately obvious through visualization alone.

CONCLUSION

- In this project, our goal was to predict whether the first stage of a **Falcon 9 launch would land successfully**, providing valuable insights into the cost and efficiency of future missions. We explored how various **features**—such as **payload mass** and **orbit type**—influence the **mission outcome**, and how these relationships can be leveraged for more accurate predictions.
- By applying several **machine learning algorithms**, we were able to train models on historical **Falcon 9 launch data**, learning the underlying patterns that determine landing success. After evaluating the performance of different models, we found that the **decision tree algorithm** outperformed the others, providing the most reliable predictive capability.
- This analysis not only enhances our understanding of **Falcon 9 launch dynamics** but also offers a **data-driven approach** to improve launch planning, reduce costs, and increase the success rate of future space missions.

APPENDIX

- All launch sites on a map