IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Abdullahi mohamed Jeylani
24.11.2025

Outline

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

Executive Summary

Summary of Methodologies

- Data was collected using the SpaceX API and web scraping, followed by data wrangling to convert landing outcomes into binary classes (0 = fail, 1 = success). Exploratory Data Analysis (EDA) was performed using visualizations and SQL, including interactive mapping with Folium. Finally, several classification models—Logistic Regression, SVM, KNN, and Decision Tree—were trained and evaluated to predict first-stage landing success.

Summary of All Results

- EDA revealed clear improvements in landing success over time, consistent site-specific patterns, and varying success rates across orbit types. Among the predictive models tested, the **Decision Tree classifier achieved the highest accuracy** on the test dataset, making it the best-performing model for predicting Falcon 9 landing outcomes.

Introduction

In this capstone project, our goal is to predict whether the first stage of the **Falcon 9** rocket will successfully land. SpaceX advertises a launch cost of approximately **62 million USD**, significantly lower than other providers whose prices exceed **165 million USD**. This cost advantage is largely due to SpaceX's ability to **reuse the first-stage booster**.

Accurately predicting landing success allows us to estimate launch costs more reliably. Such insights are valuable for organizations that may wish to **compete with SpaceX** or bid on launch contracts, as they provide a data-driven understanding of expected mission outcomes and associated costs.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using Web Scraping and SpaceX API
- Perform data wrangling
  - We would like landing outcomes to be converted to Classes y. y. (either 0 or 1). 0 is a bad outcome, that is, the booster did not land. 1 is a good outcome, that is, the booster did land. The variable Y will represent the classification variable that represents the outcome of each launch.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium
- Perform predictive analysis using classification models
  - Logistic Regression, SVM, KNN, Decision Tree

Data Collection

Our dataset was collected via two methods:

- SpaceX API
- Web Scraping.

Data Collection – SpaceX API

```
data_falcon9
```

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
4	6	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	8	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	10	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	11	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	12	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	102	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	103	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	104	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	105	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	106	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

https://github.com/Jeylani-2526/SpaceX-coursera-DS/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

```
df.head()
```

	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	NaN	NaN	F9 v1.07B0003.18	NaN	4 June 2010	18:45
1	1	CCAFS	Dragon	0	LEO	NaN	NaN	F9 v1.07B0003.18	NaN	4 June 2010	18:45
2	2	CCAFS	Dragon	525 kg	LEO	NaN	NaN	F9 v1.07B0004.18	NaN	8 December 2010	15:43
3	3	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NaN	NaN	F9 v1.07B0005.18	NaN	22 May 2012	07:44
4	4	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NaN	NaN	F9 v1.07B0006.18	NaN	8 October 2012	00:35

https://github.com/Jeylani-2526/SpaceX-coursera-DS/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

We would like landing outcomes to be converted to Classes y. y. (either 0 or 1). 0 is a bad outcome, that is, the booster did not land. 1 is a good outcome, that is, the booster did land. The variable Y will represent the classification variable that represents the outcome of each launch.

EDA with Data Visualization

https://github.com/Jeylani-2526/SpaceX-coursera-DS/blob/main/
edadataviz.ipynb

EDA with SQL

https://github.com/Jeylani-2526/SpaceX-coursera-DS/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

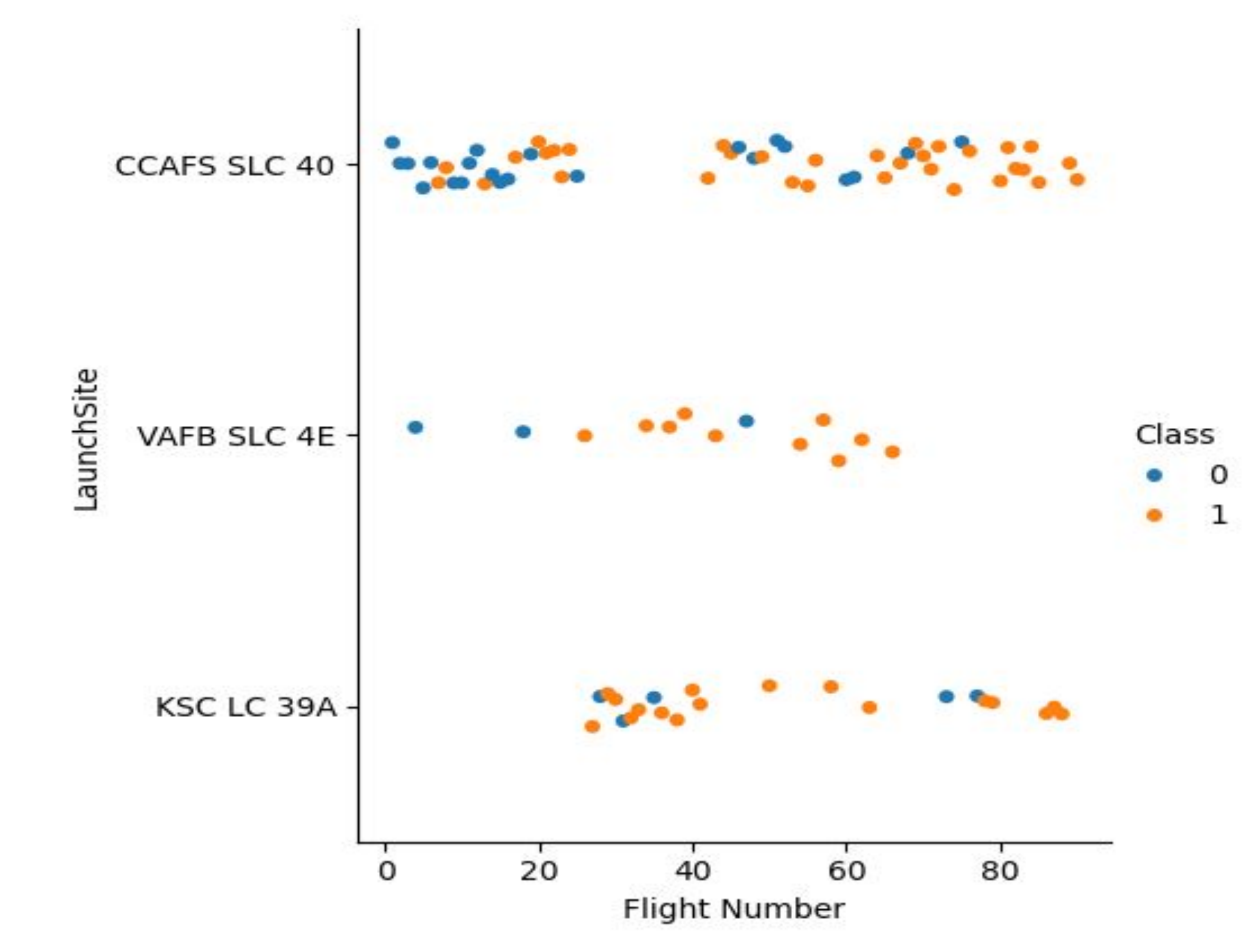
Predictive Analysis (Classification)

- https://github.com/Jeylani-2526/SpaceX-coursera-DS

Section 2

Insights drawn from EDA

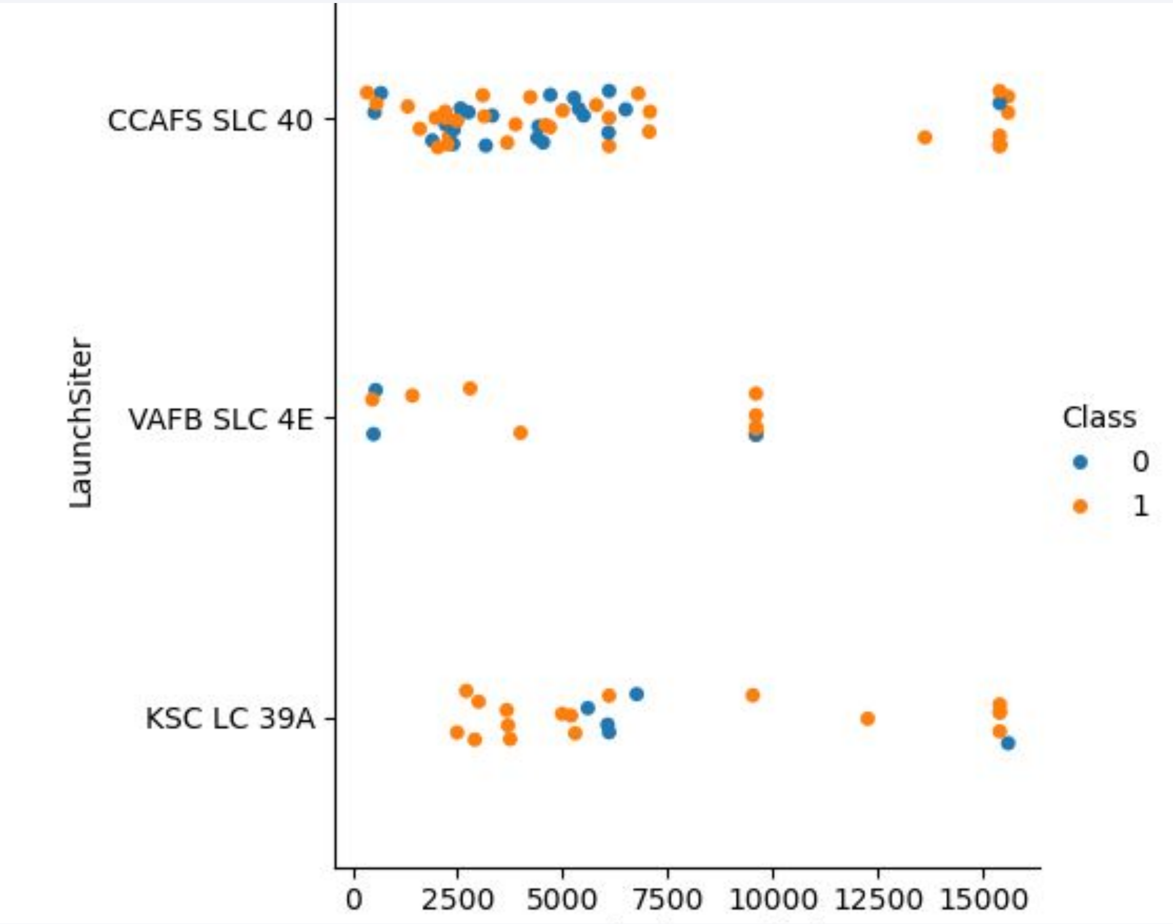
Flight Number vs. Launch Site



- CCAFS SLC 40 has the highest number of launches, showing a consistent mix of successful and unsuccessful outcomes.

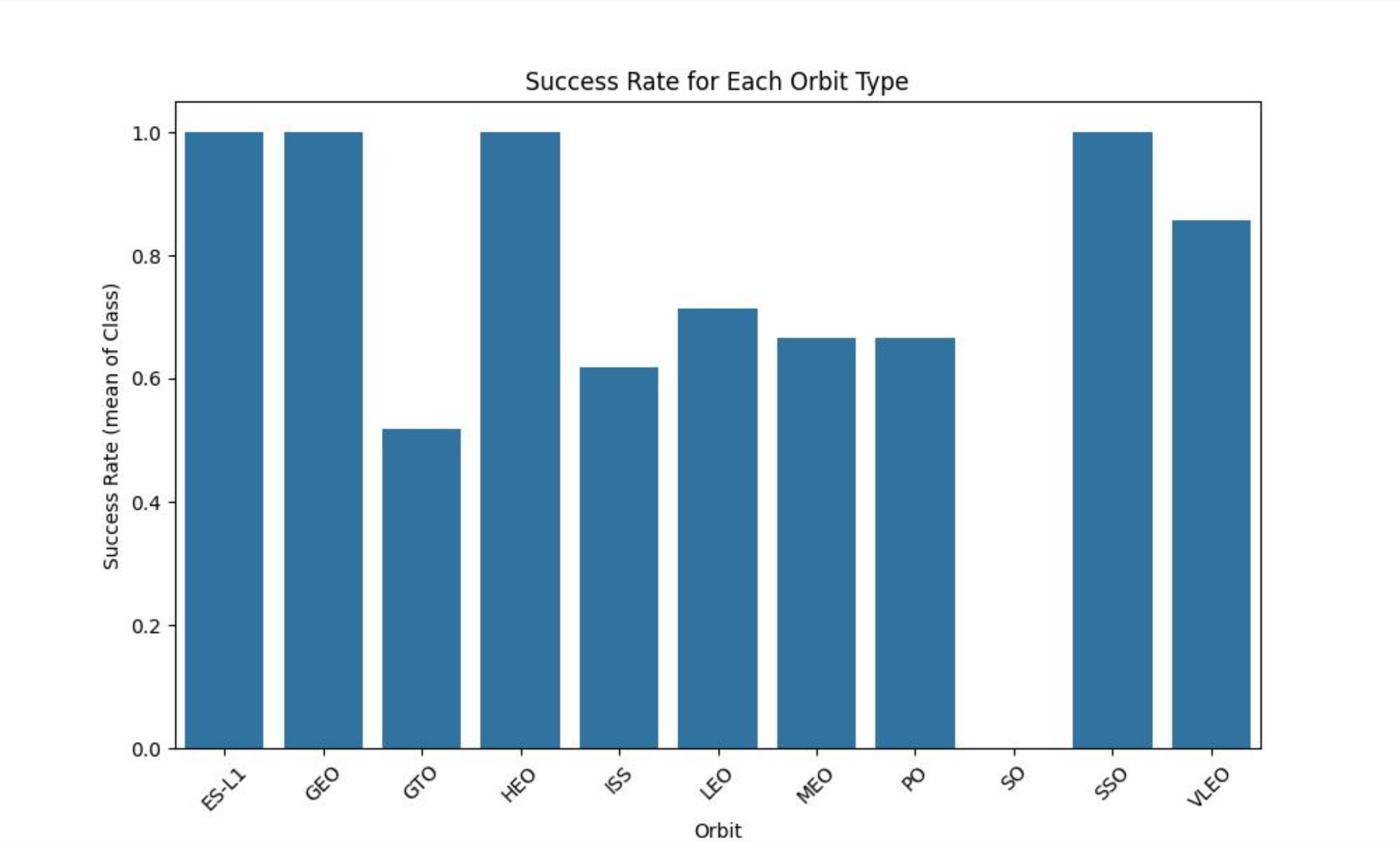
- VAFB SLC 4E and KSC LC 39A have fewer launches overall, but also show a spread of both success (1) and failure (0) across their flight numbers.

Payload vs. Launch Site



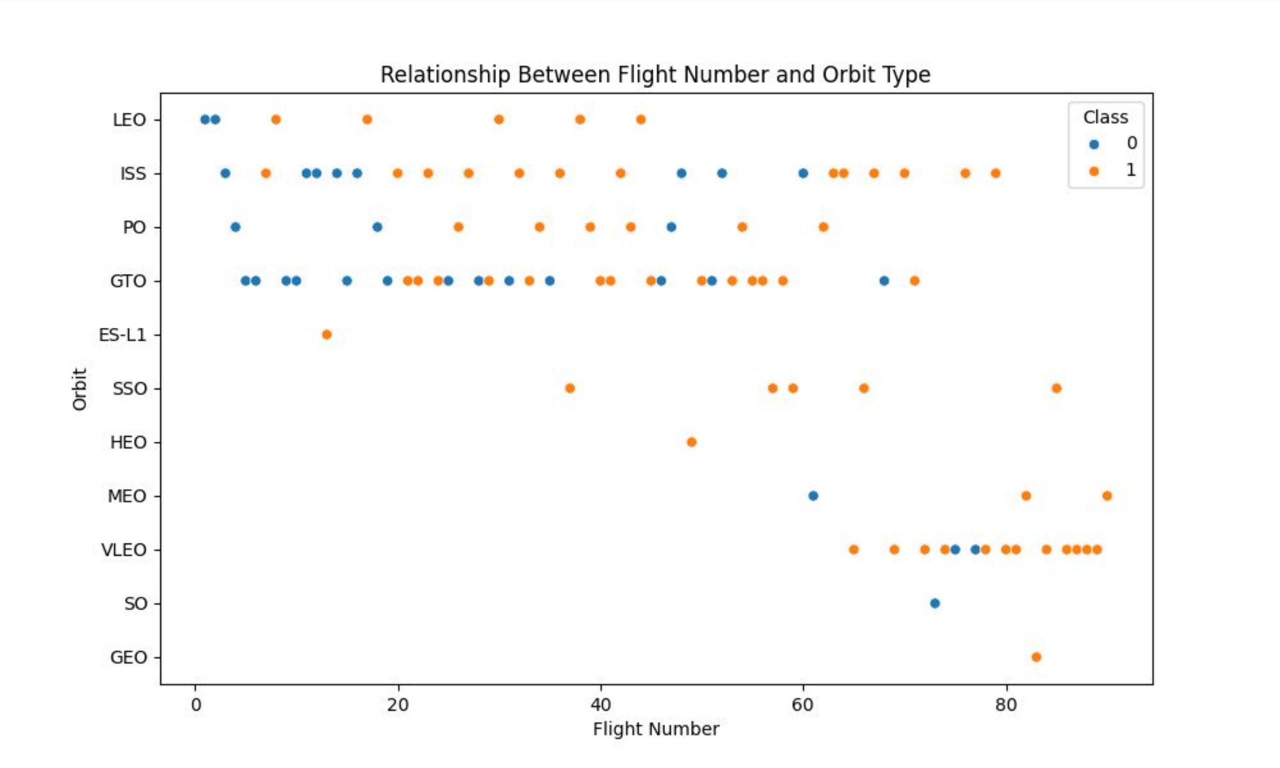
- CCAFS SLC 40 handles a wide range of payload masses and shows a balanced mix of successful (1) and unsuccessful (0) landings across all payload sizes.
- VAFB SLC 4E and KSC LC 39A launch fewer missions, but their outcomes also appear distributed independently of payload mass, showing no clear success-to-payload relationship.

Success Rate vs. Orbit Type



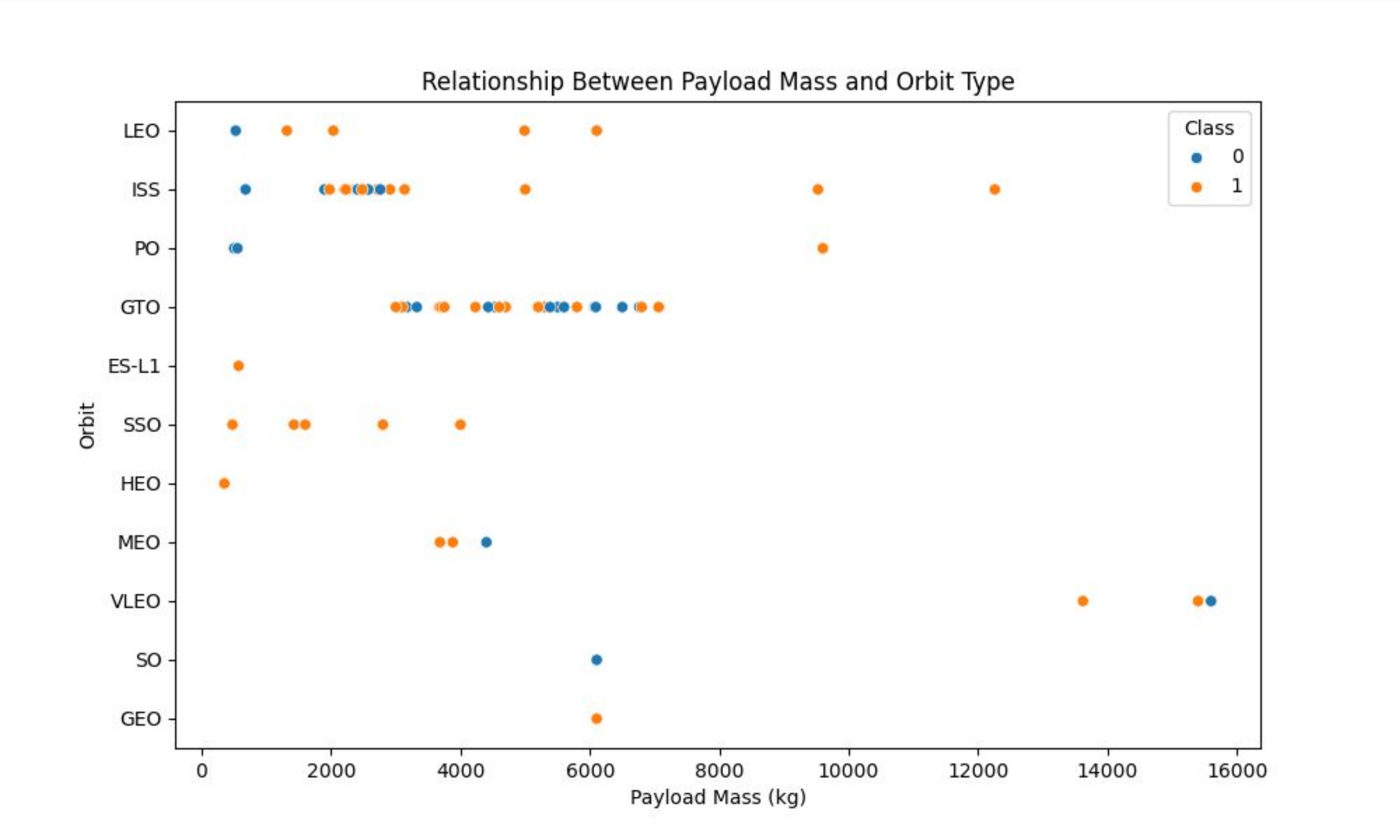
- Most orbit types show high landing success, with ES-L1, GEO, HEO, and SSO achieving perfect or near-perfect performance.

Flight Number vs. Orbit Type



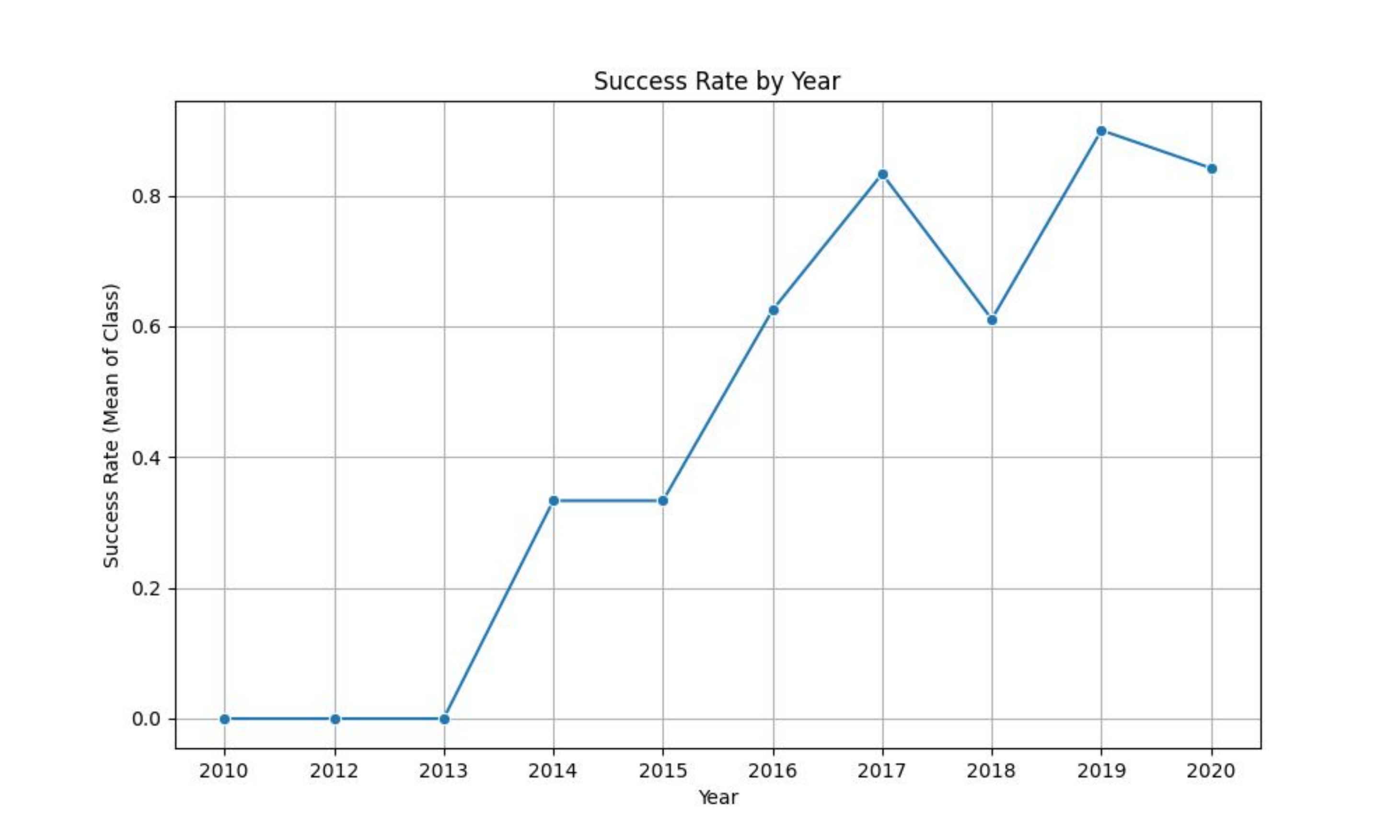
- Higher flight numbers show more consistent successful landings across most orbit types, especially in later missions.

Payload vs. Orbit Type



- Heavier payloads appear across many orbit types, but landing success remains fairly consistent regardless of payload mass.

Launch Success Yearly Trend



- Landing success improves sharply over the years, rising from repeated failures early on to consistently high success rates after 2016.

All Launch Site Names

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

Launch Site Names Begin with 'CCA'

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

Total Payload Mass

Total Payload = 45596

- This is the total payload carried by boosters launched by **NASA (CRS)**

Average Payload Mass by F9 v1.1

Average Payload = 2928.4

- This is the average payload carried by booster version **F9v1.1**

First Successful Ground Landing Date

The first successful landing outcome in ground pad was achieved on 2015-12-22.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

- these are boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

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

total number of successful and failure mission outcome

Boosters Carried Maximum Payload

- 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

booster versions that have carried the maximum payload mass

2015 Launch Records

Month	Landing_Outcome	Booster_Version	Launch_Site	Month
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	01
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	04

- records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

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

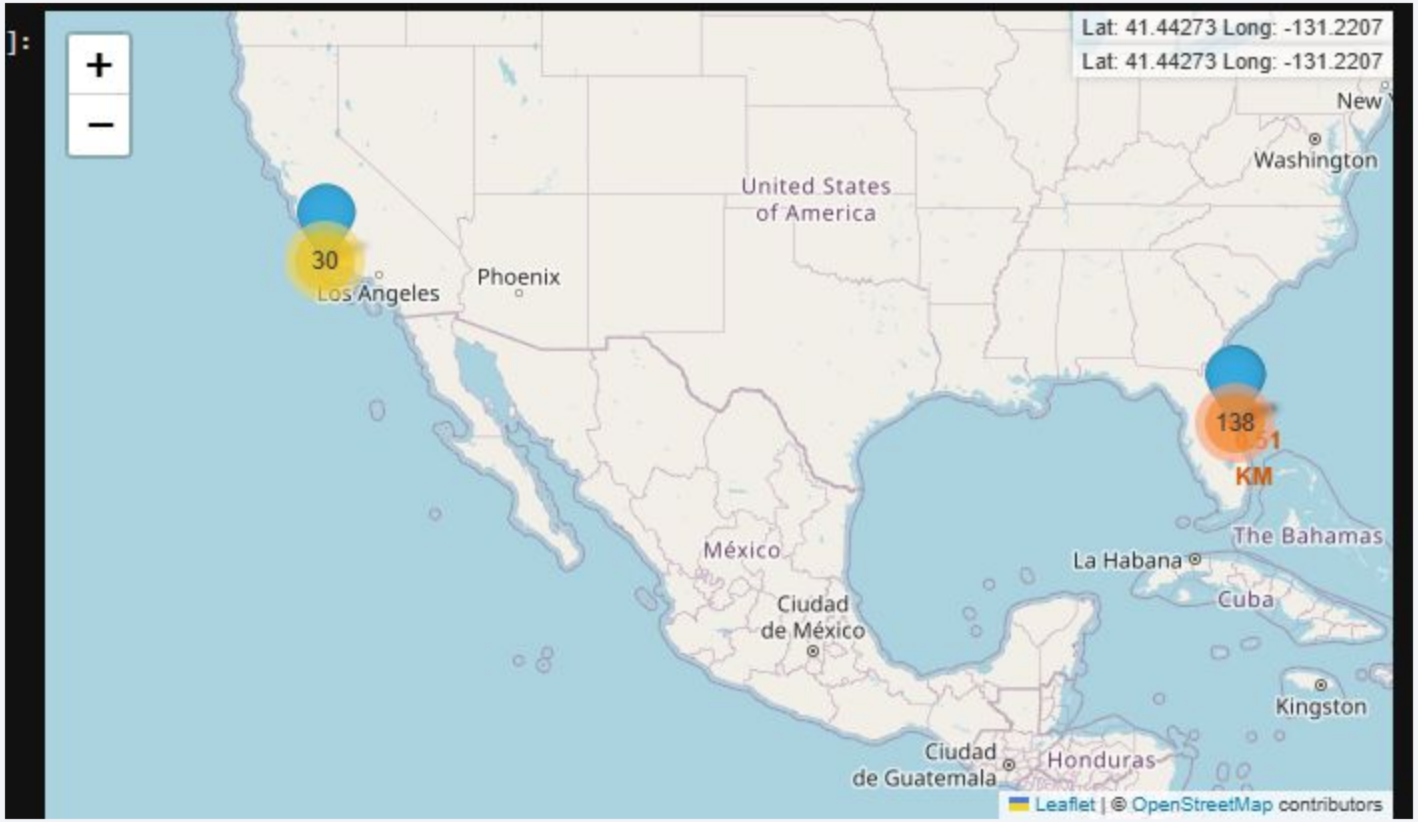
Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

- the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

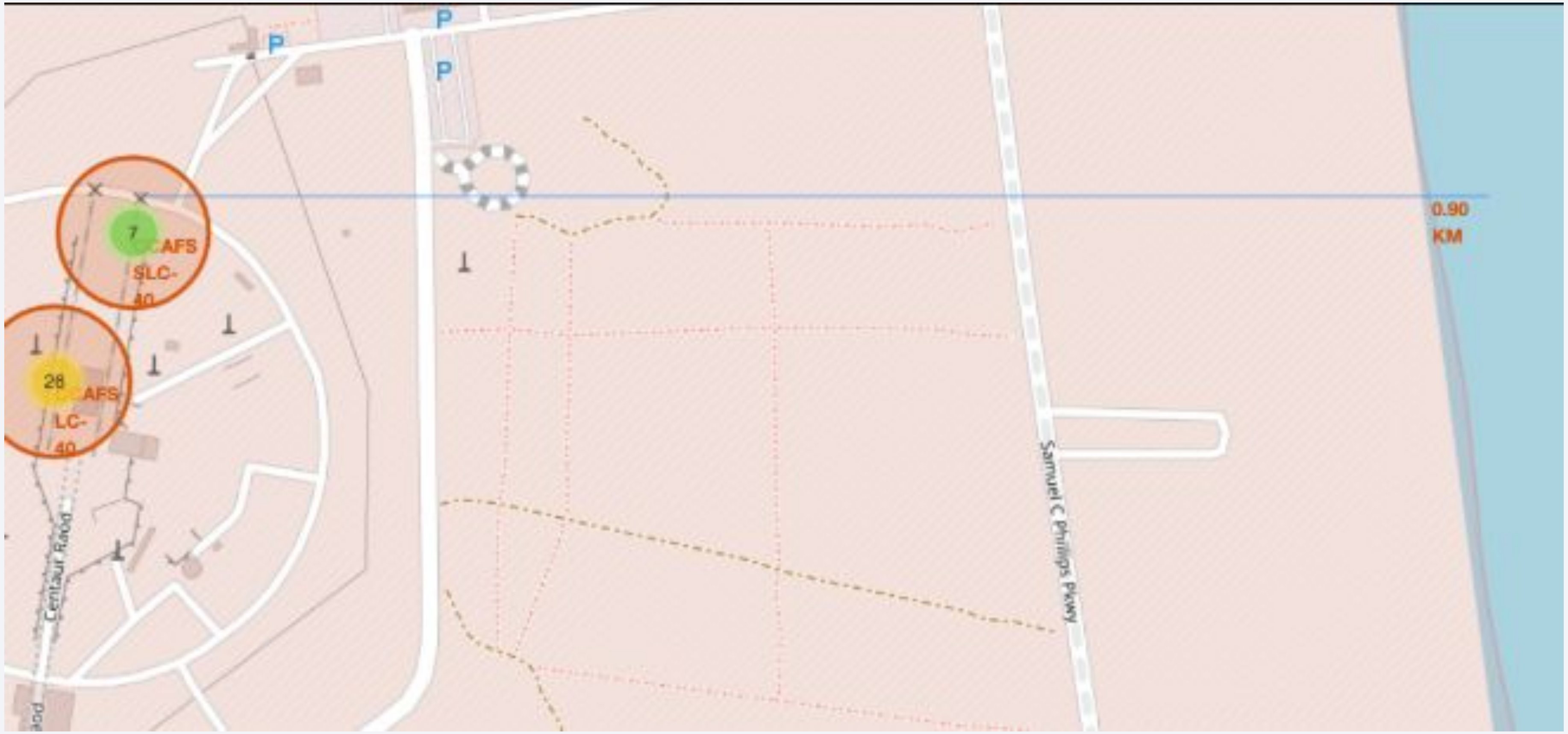
Section 3

Launch Sites Proximities Analysis

all launch sites



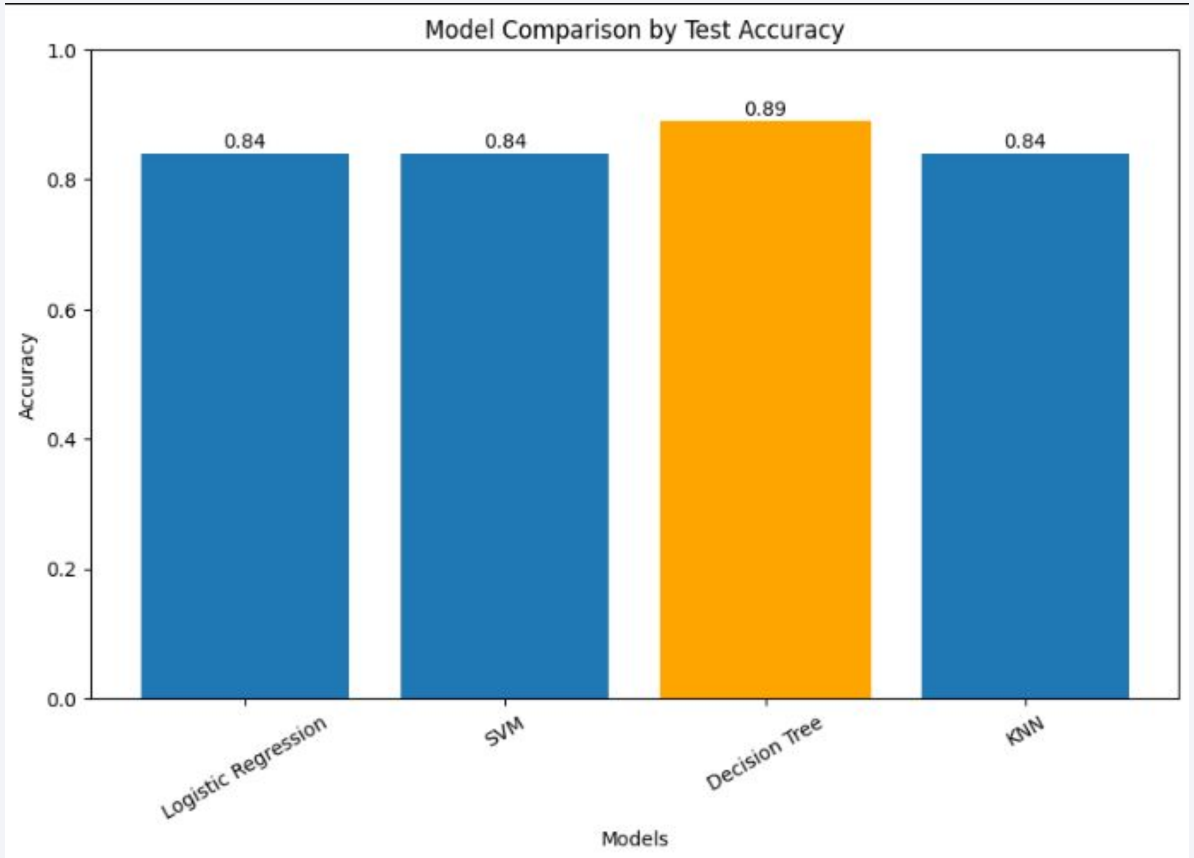
Distance to coastline



Section 5

Predictive Analysis (Classification)

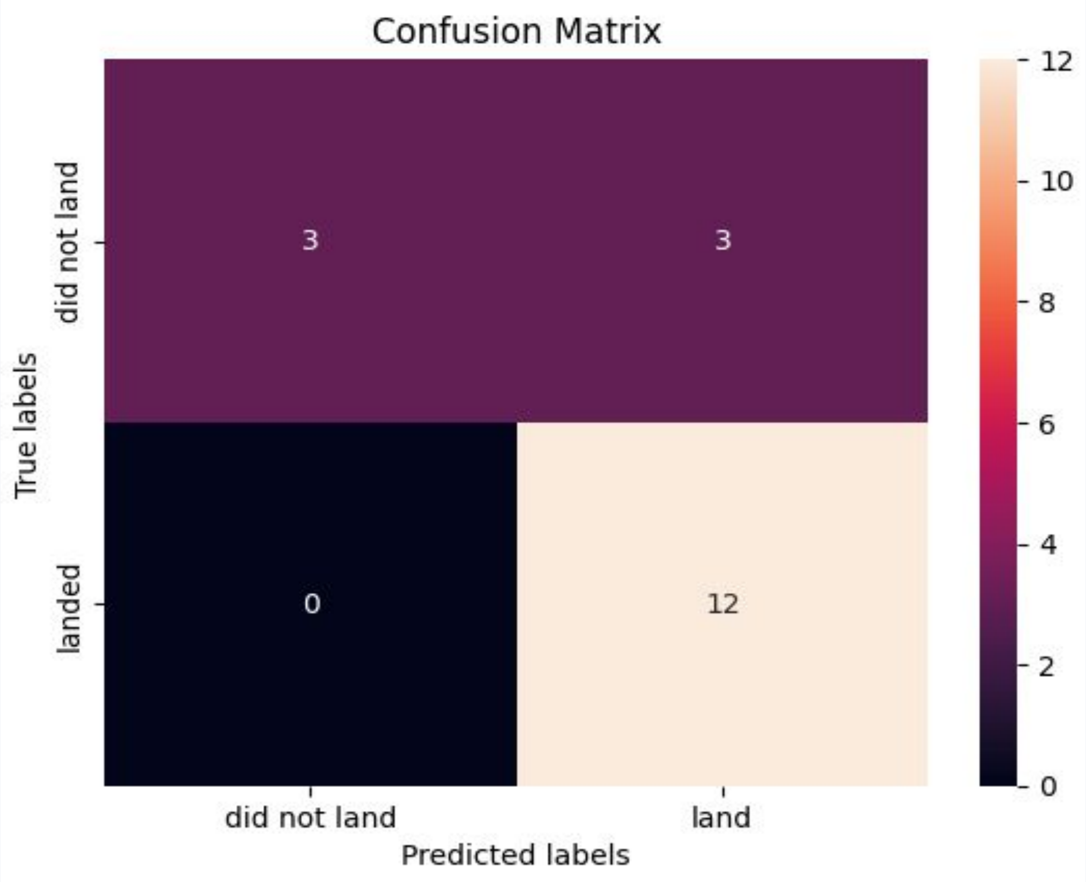
Classification Accuracy



We evaluated four classification models to predict whether a Falcon 9 first stage would successfully land. Among all models tested — Logistic Regression, SVM, KNN, and Decision Tree — the **Decision Tree classifier achieved the highest performance**, with an accuracy of **0.89** on the test dataset.

This indicates that the model correctly predicted landing outcomes for **89% of the launches**, making it the most effective approach for this problem.

Confusion Matrix



The confusion matrix illustrates how well the Decision Tree model distinguishes between successful and failed landings. Most predictions fall along the diagonal, showing that the classifier accurately identified both **successful (1)** and **unsuccessful (0)** outcomes.
Only a small number of misclassifications occurred, primarily due to overlapping feature patterns between some failure and success cases.
Overall, the matrix confirms strong predictive performance and balanced classification capability.

Conclusions

- SpaceX landing success has significantly improved over time, especially after 2016.
- Launch sites such as CCAFS SLC-40 and KSC LC-39A show consistent mission activity and performance patterns.
- Orbit type, payload mass, and flight number all influence landing outcomes, but no single factor dominates on its own.
- The Decision Tree model achieved the highest accuracy (0.89), making it the most reliable classifier for predicting Falcon 9 landing success.
- Predictive insights from this project can support organizations planning to compete with SpaceX or estimate launch costs more effectively.
- The analysis demonstrates how data collection, EDA, and machine-learning models can be combined to solve real-world aerospace problems.

Appendix

This section includes relevant project assets such as:

- Python code used for data collection (API + web scraping)
- Data wrangling and feature preprocessing scripts
- SQL queries executed during exploratory analysis
- Visualizations generated during EDA
- Interactive Folium map outputs
- Jupyter Notebook links from the GitHub repository

Thank you!