SPACEX FALCON 9 ANALYSIS AND PREDICTION

A capstone project performed with Data Science methodology

Author : Saad Sadaf

GitHub Repo link: https://github.com/saad0-0hehe/SpaceX-Capstone-Project-

EXECUTIVE SUMMARY

- Falcon 9 rocket reusability lowers launch costs.
- We used SpaceX launch data for **Exploratory Data Analysis (EDA)**, **SQL queries**, **interactive dashboards**, and **machine learning classification models**.
- Goal: Predict whether Falcon 9's first stage will land successfully.
- Achieved 84% **accuracy** with the KNN model

AGENDA

1. Introduction
2. Objectives
3. Dataset Overview
4. Methodology and Tools
5. Result and Analysis
    -Dashboard and visuals
    -Machine Learning Models
6. Discussion
7. Conclusion

INTRODUCTION

- SpaceX has significantly reduced the cost of space travel by reusing rockets.
- Goal: Predict whether Falcon 9's first stage will land successfully.
- Importance:
- Helps competitors and stakeholders understand cost reduction.
- Provides insights into SpaceX's business model.

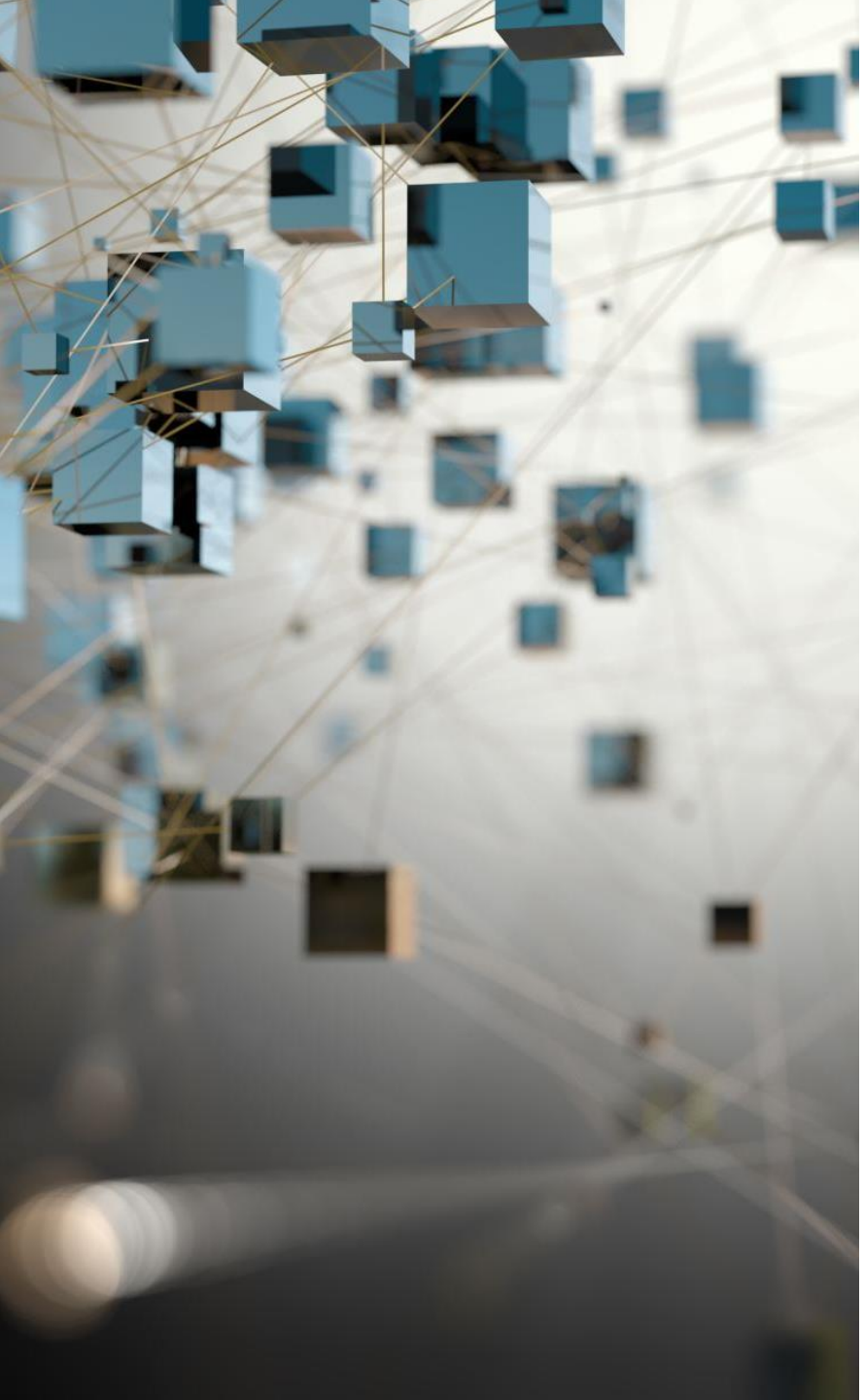
PROJECT OBJECTIVE

Analyze SpaceX Falcon 9 Launch Data set

Predict first-stage landing success

Explore different factors and geographical locations

Increase success rate and make cost effective decisions.



DATASET GATHERING AND FEATURES

- Source: SpaceX Launch Data (CSV).
- Key Features:
- Launch Site
- Payload Mass (kg)
- Orbit
- Booster Version
- Outcome (Success/Failure)

DATA WRANGLING

- Data collected from **SpaceX REST API** and **Wikipedia**
- Stored in **CSV format** for analysis
- Wrangling steps:
- Removed irrelevant / duplicate columns
- Handled missing values
- Encoded categorical variables (e.g., landing outcomes)
- Created new features such as `Class` (success/failure of landing)

76]:

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	La
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6t
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6t
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6t
92	89	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9
93	90	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6t

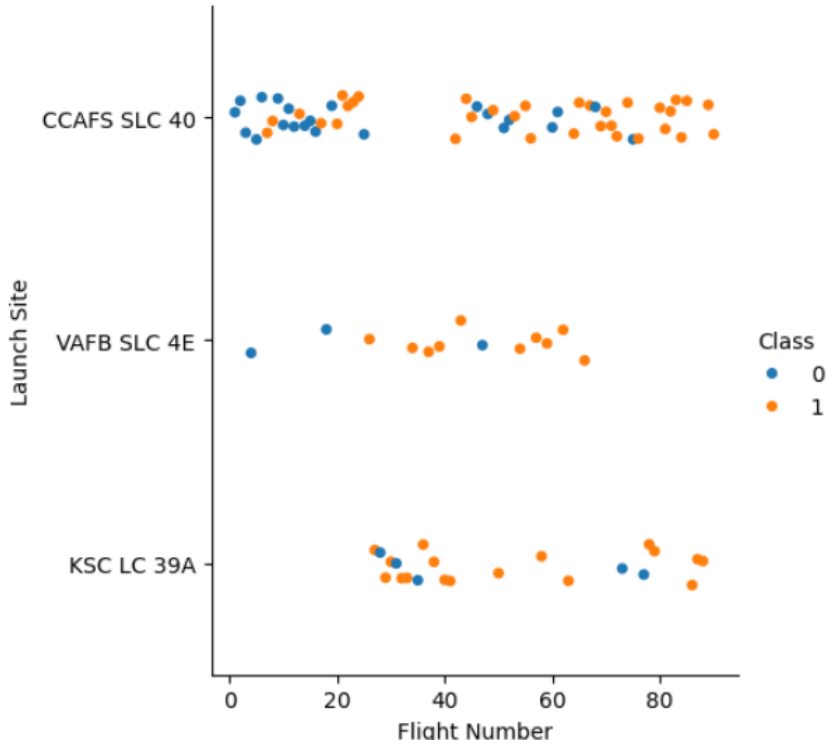
90 rows × 17 columns

API + Wiki → Pandas → Cleaned Dataset

METHODOLOGY AND TOOLS

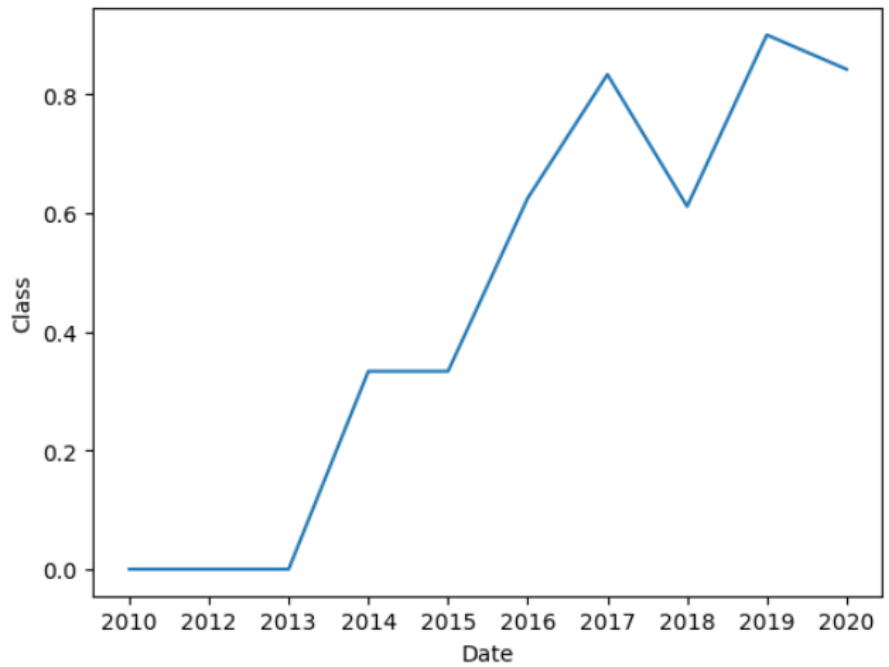
- **Data Cleaning & Preprocessing:** Pandas, NumPy
- **Visualization:** Folium, Plotly, Dash
- **Machine Learning:** Logistic Regression, SVM, GridSearchCV
- **Evaluation:** Accuracy, Confusion Matrix

EDA & INTERACTIVE VISUAL ANALYTICS



- Explored dataset using:
- **Descriptive statistics** (payload range, orbit types)
- **SQL queries** to summarize launch outcomes
- Built interactive visualizations using:
- **Folium** for geospatial mapping of launch sites
- **Plotly** for payload vs. success scatterplots & pie charts
- **Dash dashboard** for interactive analytics

- Success Rate(Class) vs Date Growth



EDA WITH SQL

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

AVG_Payload_Mass
2928.4

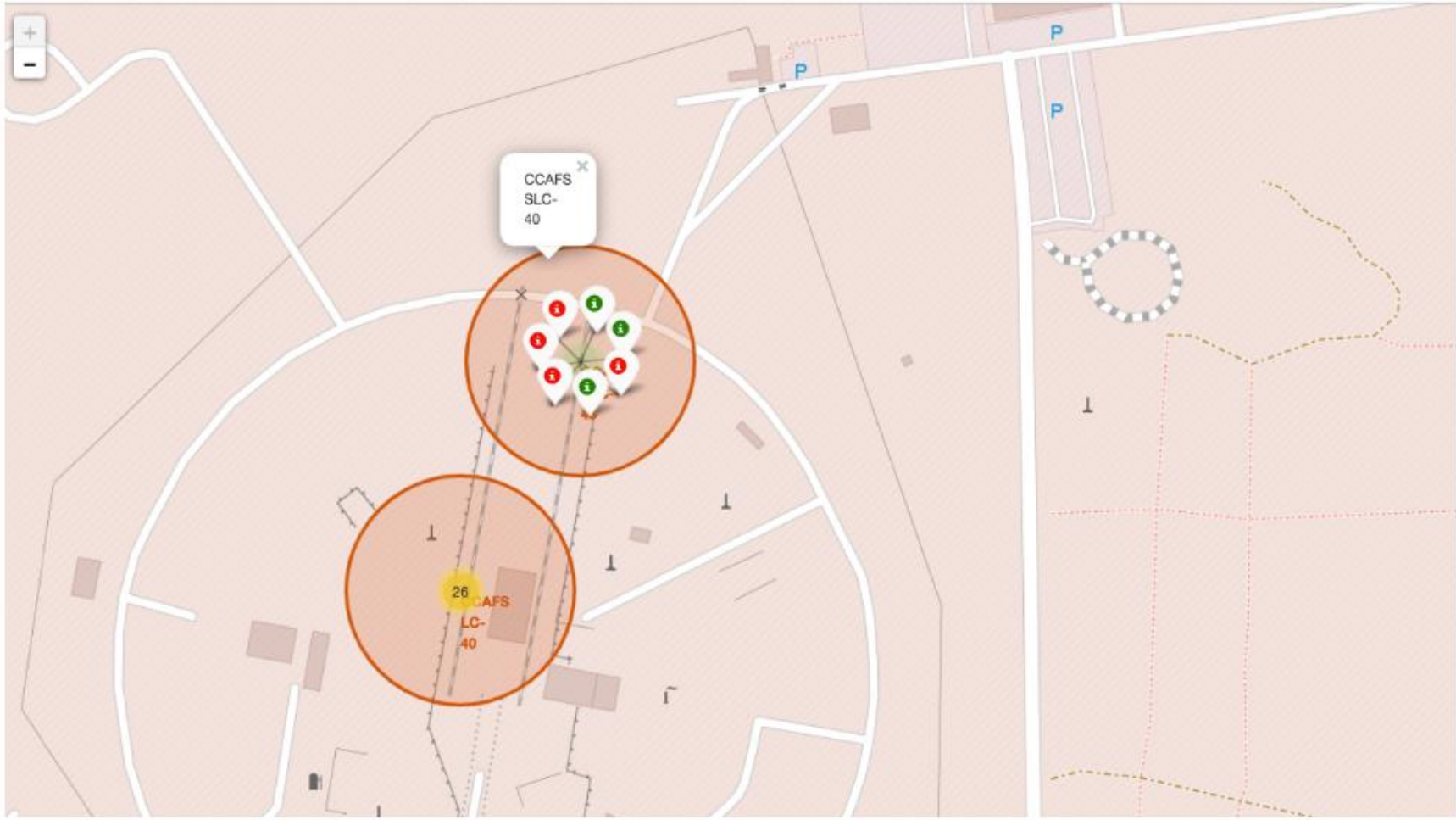
Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

Outcome	Total
Failure	1
Success	100

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

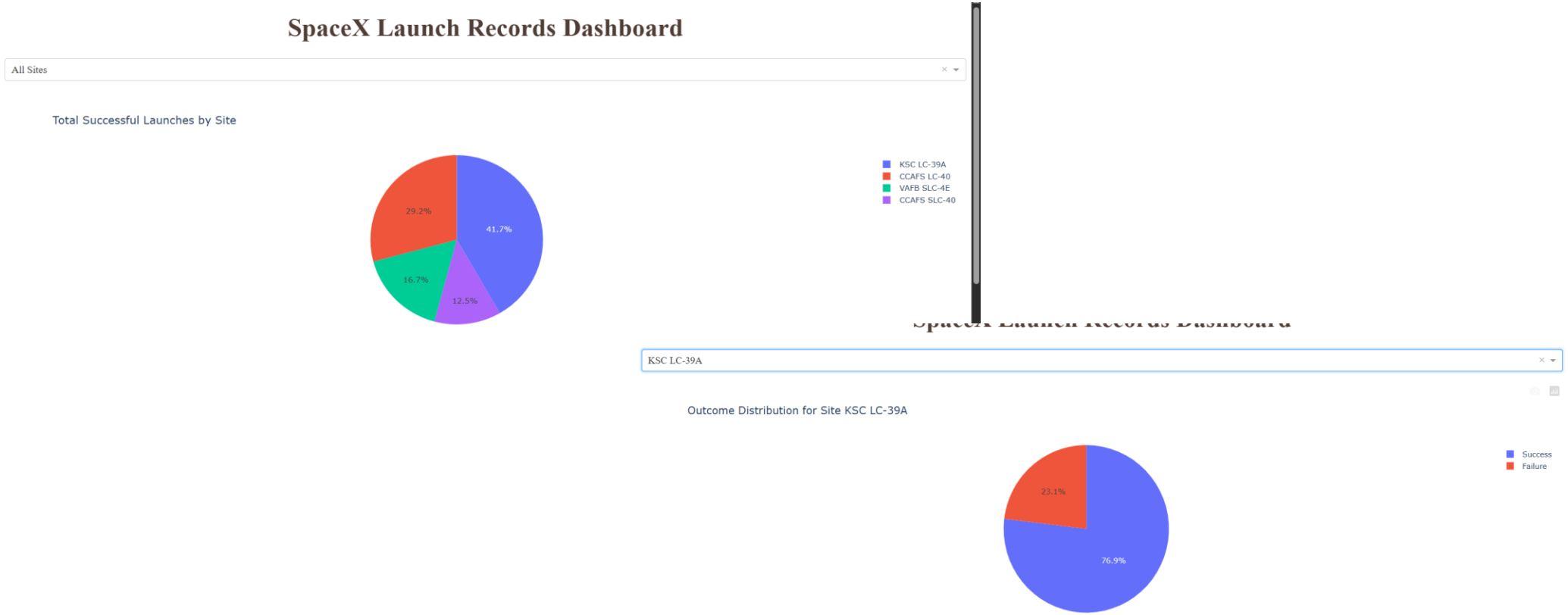
GEOSPATIAL ANALYSIS

- Launch sites are near the equator and coastlines.
- Sites are also close to highways and railways for logistics.
- Safe distance maintained from populated cities.

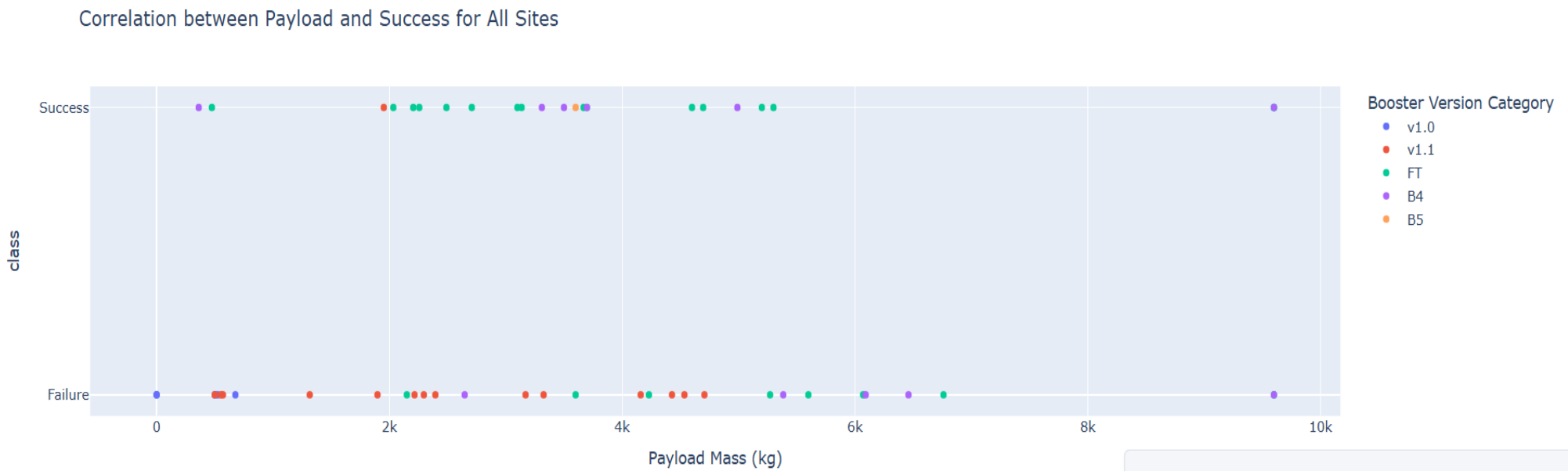


From the color-labeled markers in marker clusters, you should be able to easily identify which launch sites have relatively high success rates.

PLOTLY DASH APP



Payload range (Kg):



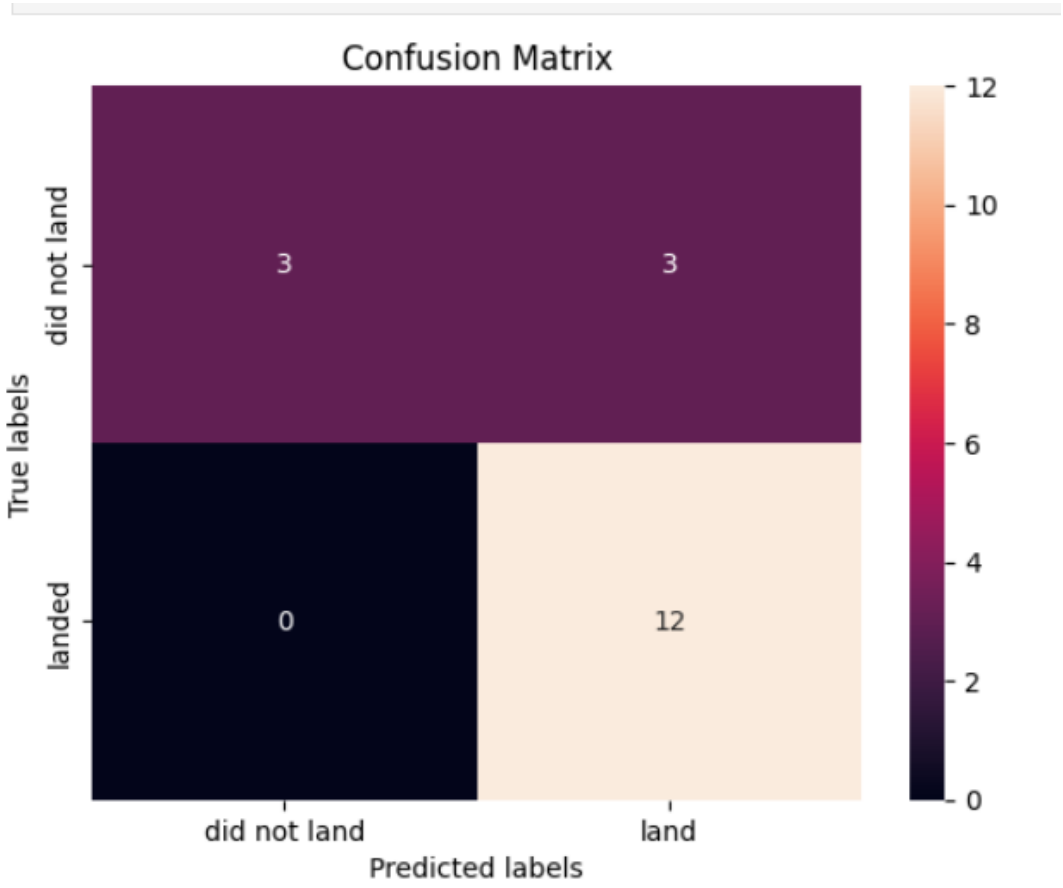
PREDICTIVE ANALYSIS METHODOLOGY

- Machine Learning models applied:
- Logistic Regression
- Support Vector Machine (SVM)
- Decision Tree Classifier
- K-Nearest Neighbors (KNN)

- Used **GridSearchCV (cv=10)** for hyperparameter tuning
- Evaluated models with **accuracy, confusion matrix, classification report**

SUPPORT VECTOR MACHINE

ACCURACY = 84%



KNN CLASSIFIER

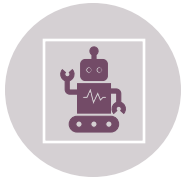
ACCURACY =83%

CONCLUSION

Launch outcomes influenced by **payload mass, booster version, and site location**

Most sites located near **coastlines for safety** and **close to infrastructure**

Machine learning models successfully **predicted launch success with high accuracy**



Interactive tools (Folium & Dash) made results **more accessible & insightful**

Thank you