Winning Space Race with Data Science

Abdulrahman Bello
29th January 2025

Outline

EXECUTIVE SUMMARY

INTRODUCTION

METHODOLOGY

RESULTS

CONCLUSION

APPENDIX

Executive Summary

This project analyses SpaceX Falcon 9 launches using **data collection, exploratory analysis, visualization, interactive mapping, and predictive modeling** to identify factors influencing successful landings.

Key Findings:

VAFB SLC-4E has the highest launch success rate.

Payload Mass & Success: Higher payloads in **LEO, Polar, and ISS orbits** have higher success rates.

Orbit & Success: **ES-L1, GEO, HEO, and SSO** have the highest success rates, while **GTO** has the lowest.

Success Trend: The overall launch success rate has increased significantly since 2013

Introduction

SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.

if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch

Section 1

Methodology

Methodology

1. DATA COLLECTION
2. DATA WRANGLING
3. EXPLORATORY DATA ANALYSIS
4. INTERACTIVE VISUAL ANALYTICS USING FOLIUM AND PLOTLY DASH
5. PREDICTIVE ANALYSIS USING CLASSIFICATION MODEL

Data Collection

- Request to the SpaceX API(https://api.spacexdata.com/v4/launches/past)
- Clean the data
- Extract a Falcon 9 launch records HTML table from Wikipedia using BeautifulSoup

Data Collection – SpaceX API

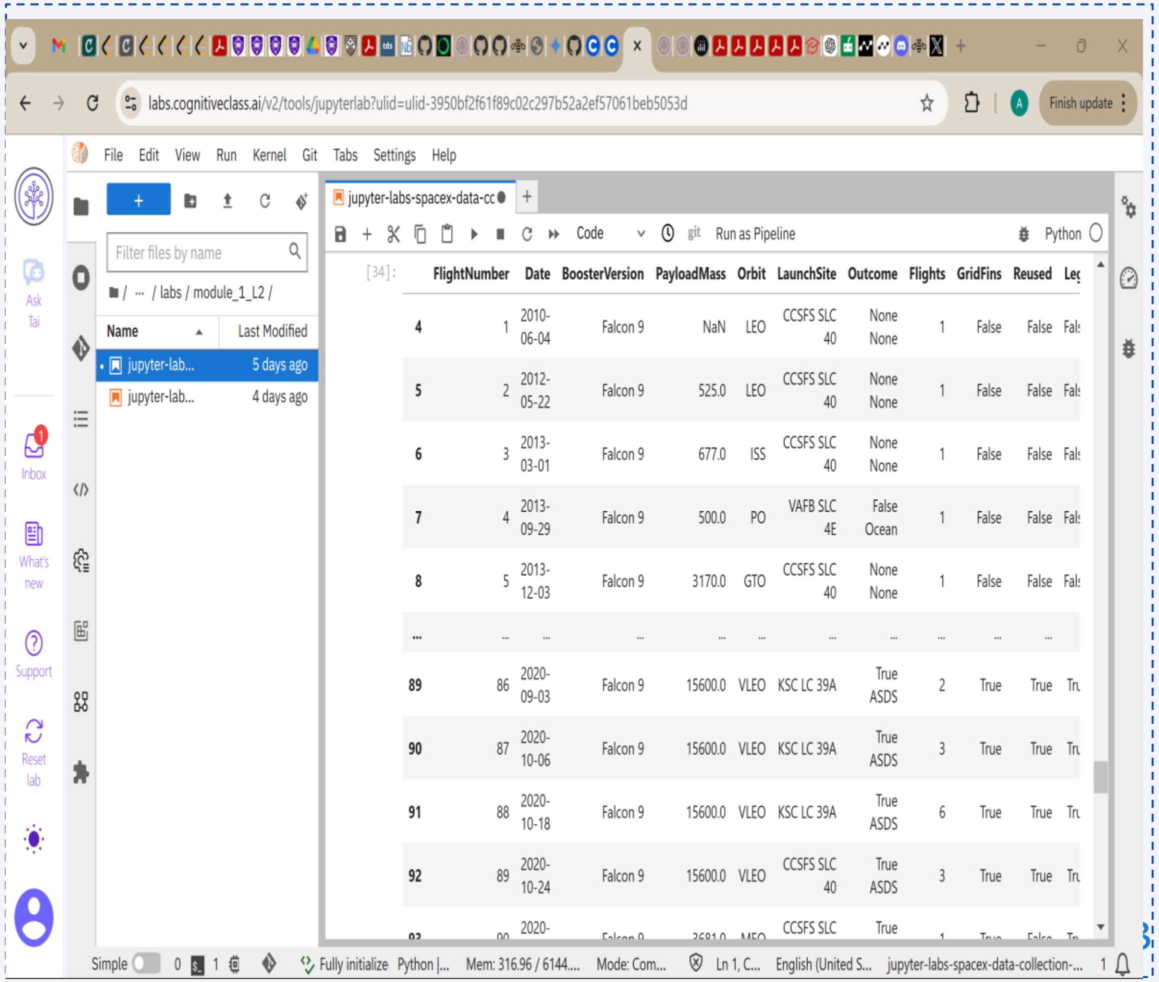
- SpaceX API

  (https://api.spacexdata.com/v4/launches/past)

- Request and parse the SpaceX launch data using the GET request

- Load into a dataframe

- Github url: https://github.com/Inuaabdul/Spacex_data/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- perform an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response.
- Create beautifulSoup object from the response.
- Extract all column/variable names from the HTML table header
- Create a data frame by parsing the HTML tables

Data Wrangling

- Exploratory Data Analysis was performed such as
    - Calculating the number of launches on each site
    - number and occurrence of each orbit
    - number occurence of mission outcome of the orbits
- Creation of a landing outcome label from Outcome column

EDA with Data Visualization

- Performed exploratory data analysis and feature engineering using and pandas and matplotlib to plot charts

Some of these charts include:

- **Categorical plot**: visualize the relationship between flight number and launch site
- **Scatter plot**: to show correlation between two independent variables
- **Bar chart**: to visually check if there are any relationship between success rate and orbit type.
- **Line plot**: show trends and changes over time

Github link: EDA with visualization.

EDA with SQL

Some of the queries carried out are to:

- Display the names of the unique launch sites in the space mission
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first succesful landing outcome in ground pad was acheived.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- Rank 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.

Github url: [EDA using sqlite]()

Interactive Map with Folium

- launch sites Maps, success/failed launches for each site and distance to its proximities.
- This was achieved using objects such as
  - ❑ Markers: to mark locations on the map
  - ❑ Circles: add a highlighted circle area with a text label on a specific coordinate
  - ❑ Marker cluster: simplify a map containing many markers having the same coordinate.
  - ❑ MousePosition: to get the coordinates of a point on the map as we hover
  - ❑ Polyline: to draw line between two locations
- GitHub URL :folium map

Dashboard with Plotly Dash

- Added a Launch Site Drop-down Input Component
- Added a callback function to render pie-chart based on selected site dropdown
- Added a Range Slider to Select Payload
- Added a callback function to render the scatter plot

These plots are added to show the total success launches and observe how payload may be correlated with mission outcomes for selected site(s).

From a dashboard point of view, we want to be able to easily select different payload range and see if we can identify some visual patterns. a dropdown menu to let us select different launch sites.

GitHub URL: [spacex_dash_app]

Predictive Analysis (Classification)

Models such as logistic regression, decision tree, SVM, and KNN. They were trained and hyperparameters are selected using the function GridSearchCV . and used to predict. Evaluation metrics were used to check accuracy of each model

We split the data into training and testing data using the function train_test_split. Then we apply the different models and GridSearchCV which Allows us to test parameters of classification algorithms and find the best one. The accuracy score is calculated and a confusion matrix is plotted

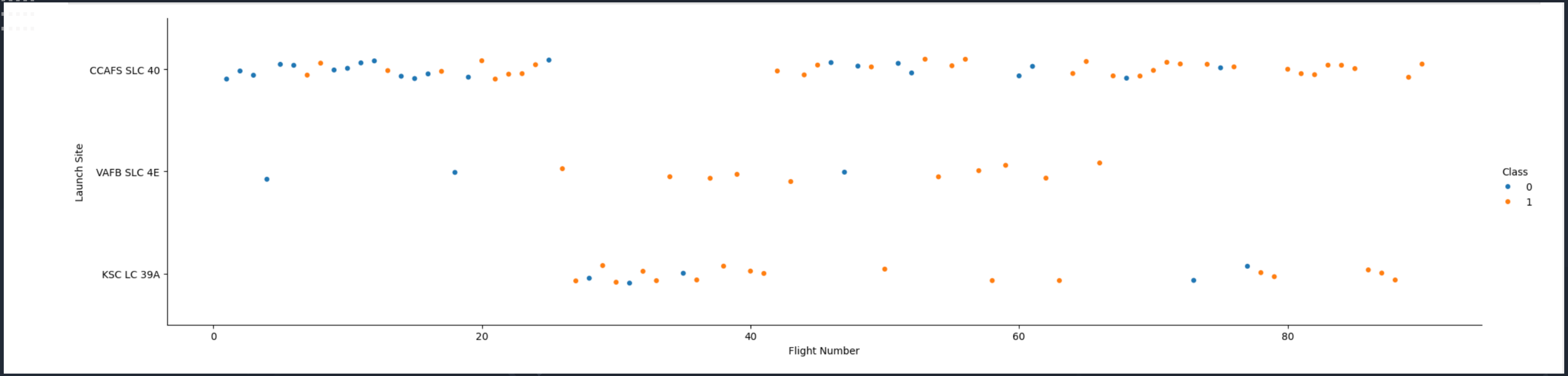
GitHub URL: SpaceX_MachineLearningPrediction

RESULTS

model	accuracy	tuned hpyerparameters
KNN	0.94	{'algorithm': 'auto', 'n_neighbors': 5, 'p': 1}
DecisionTree	0.89	{'criterion': 'entropy', 'max_depth': 4, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 2, 'splitter': 'best'}
Logistic Regression	0.94	{'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
SVM	0.94	{'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}

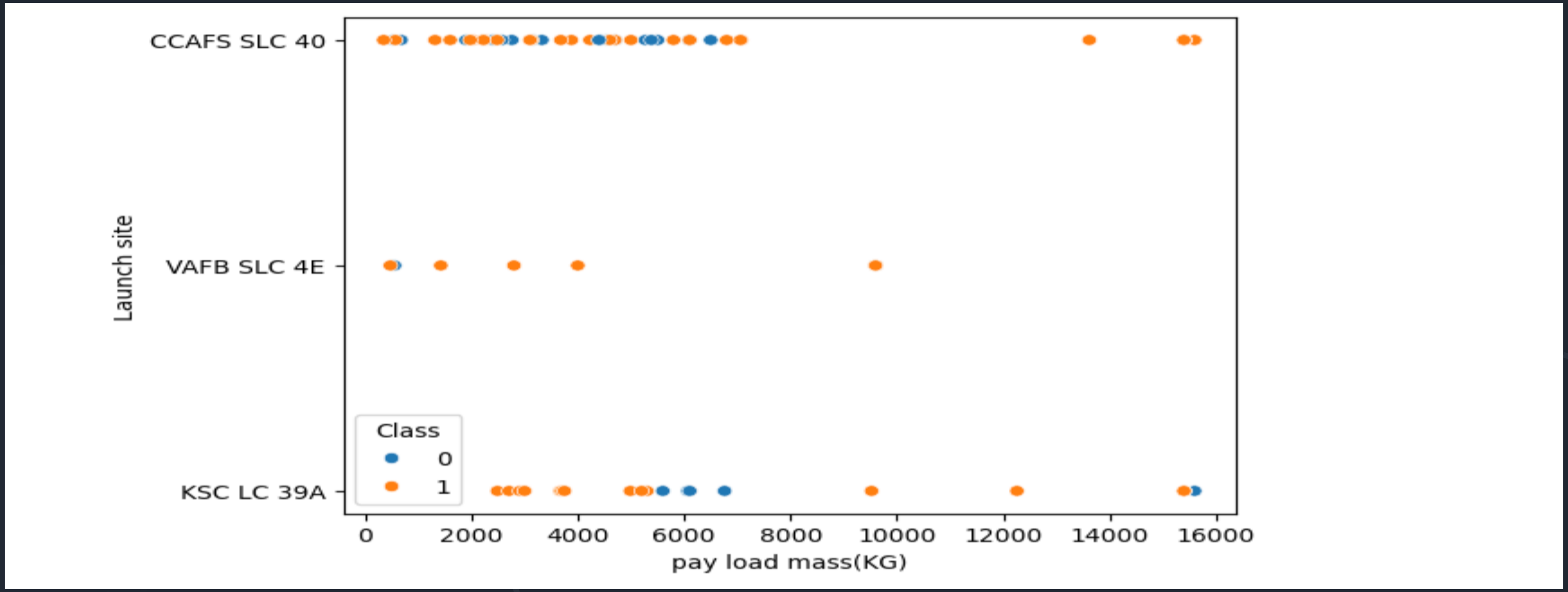
Section 2

Insights drawn from EDA



Flight Number vs. Launch Site

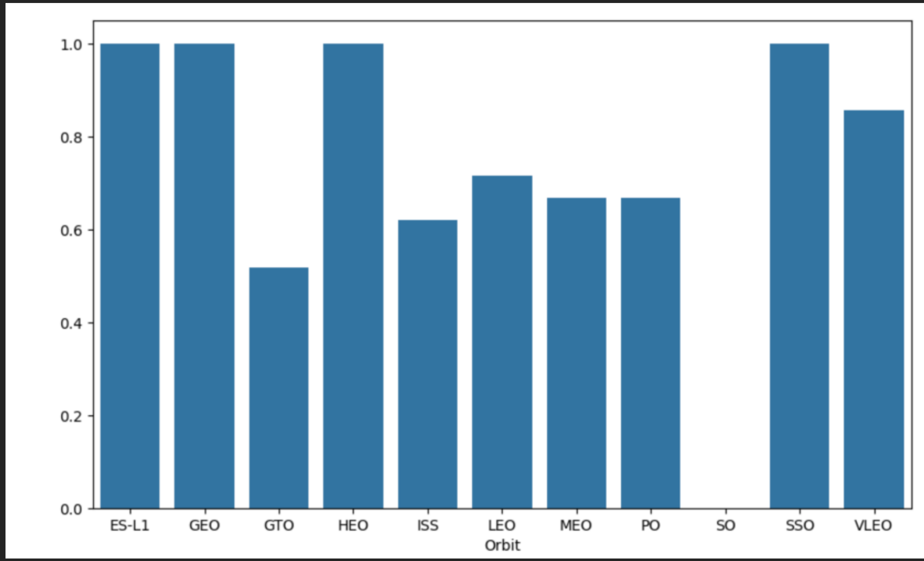
- VAFB SLC 4E has the highest launch success rate
- With more flight number, the success launch in each site increases



Payload vs. Launch Site

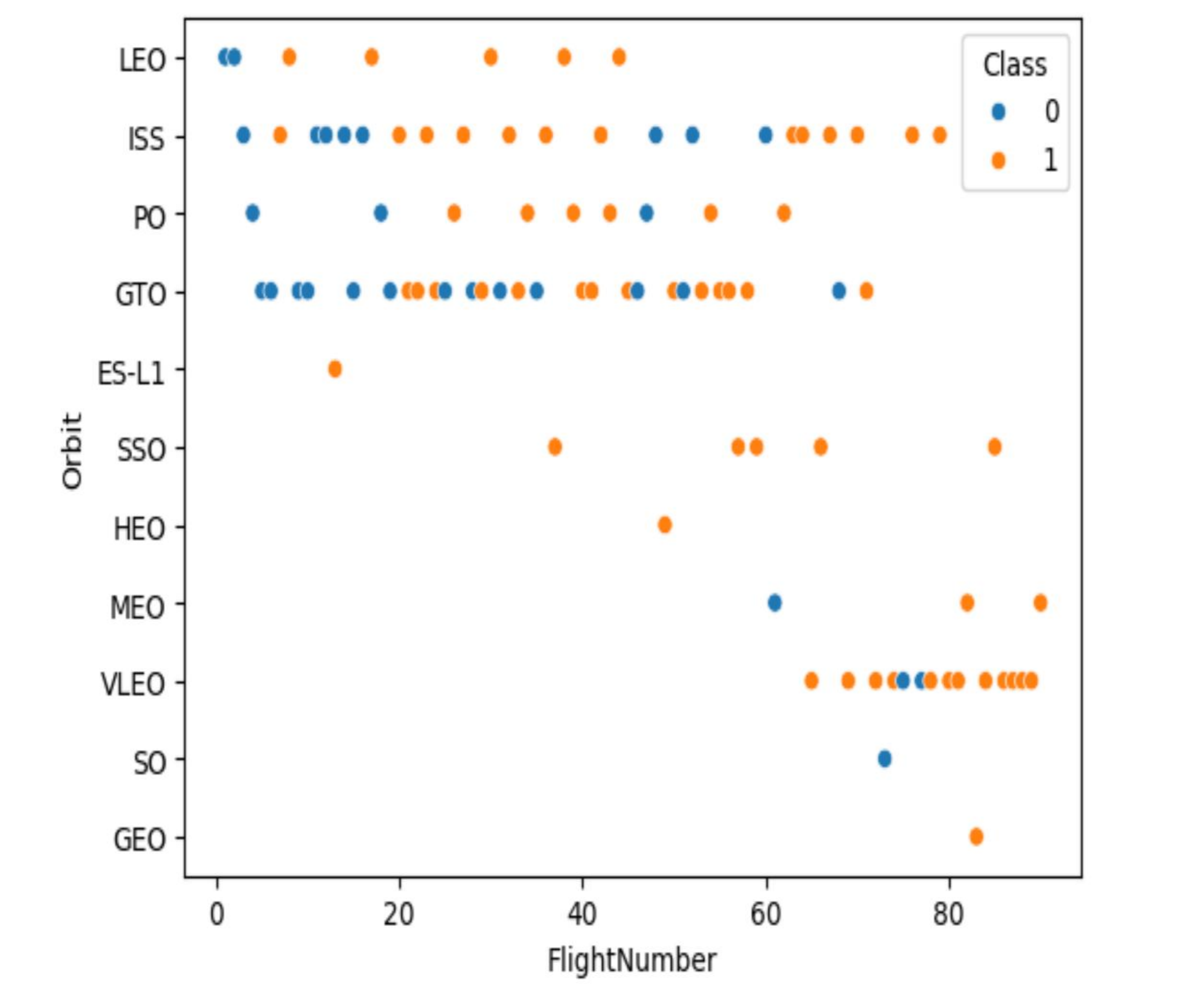
If you observe payload mass vs. Launch site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type



ES-L1,GEO,HEO and SSO have the highest success rates with GTO having the lowest.

Flight Number vs. Orbit Type

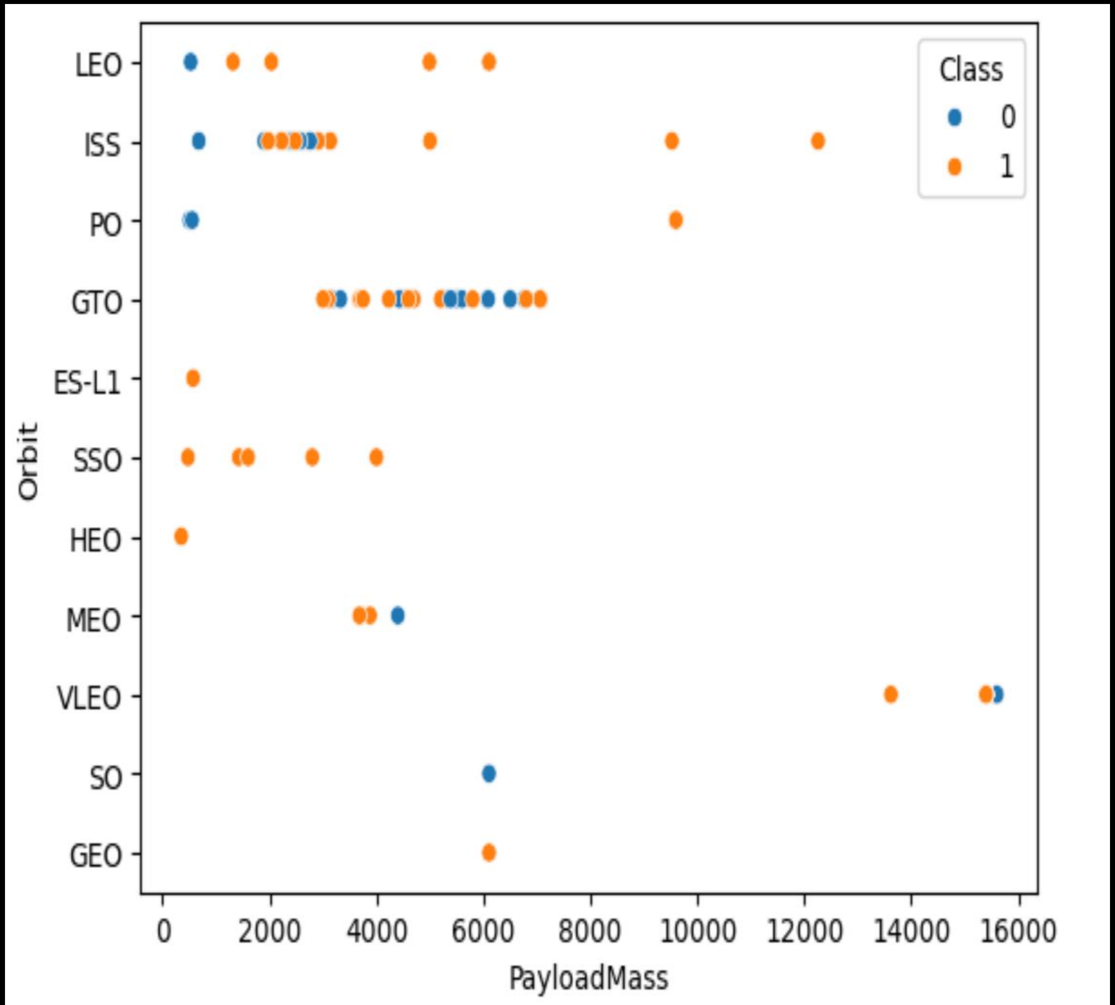


Observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

Payload vs. Orbit Type

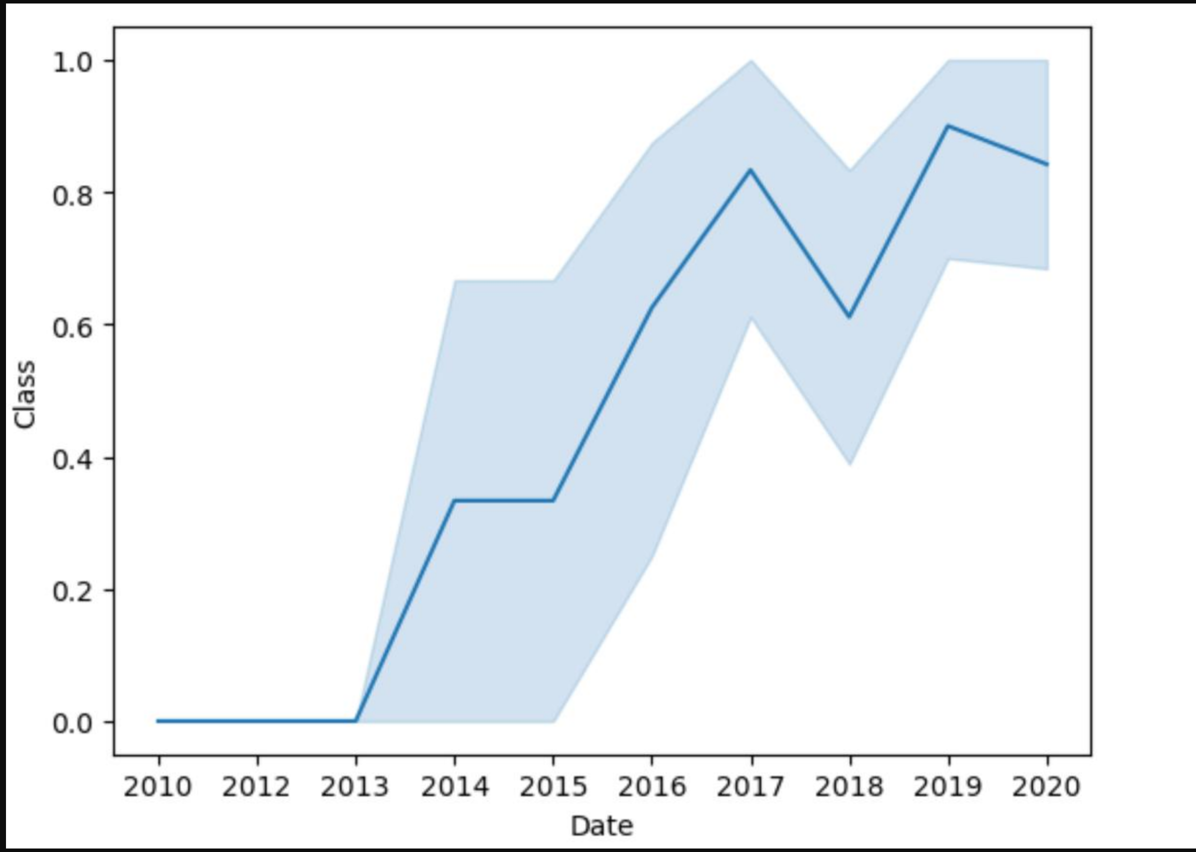
With heavy payloads the successful landing or positive landing rate are more for polar, leo and ISS.

However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.



Launch Success Yearly Trend

You can observe that the success rate since 2013 kept increasing with a slight pullback in 2018 and kept rising till 2020



All Launch Site Names

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

Launch Site Names Begin with 'CCA'

Launch sites that begin with ‘CCA’ from the table. The result was limited to the first five occurrence.

- CCAFS LC-40
- CCAFS LC-40
- CCAFS LC-40
- CCAFS LC-40
- CCAFS LC-40

Total Payload Mass

The total payload carried by boosters from NASA is
45596

Average Payload Mass by F9 v1.1

the average payload mass carried by booster version F9 v1.1 is **2928.4**

First Successful Ground Landing Date

The date of the first successful landing outcome on ground pad is **22^{nd} December, 2015**

Successful Drone Ship Landing with Payload between 4000 and 6000

- F9 FT B1022
- F9 FT B1026
- F9 FT B1029.1
- F9 FT B1021.2
- F9 FT B1036.1
- F9 B4 B1041.1
- F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

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

Boosters Carried Maximum Payload

- List of the booster which have carried the maximum payload mass

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

Month	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

- List of failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

- Ranking 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

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

Section 3

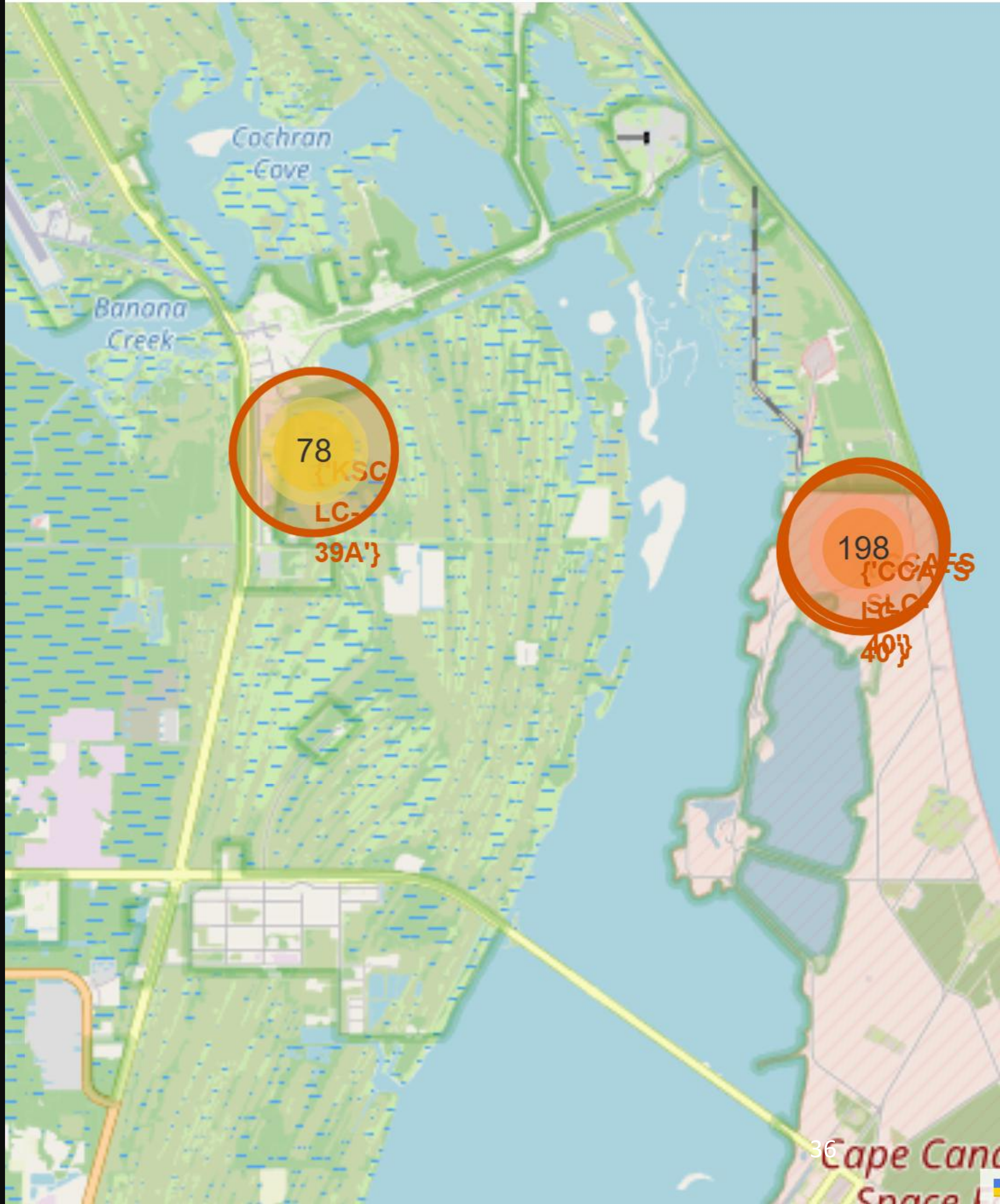
Launch Sites Proximities Analysis

LAUNCH SITES MAP

As you can see, the launch sites have been marked on the map using their respective coordinates.

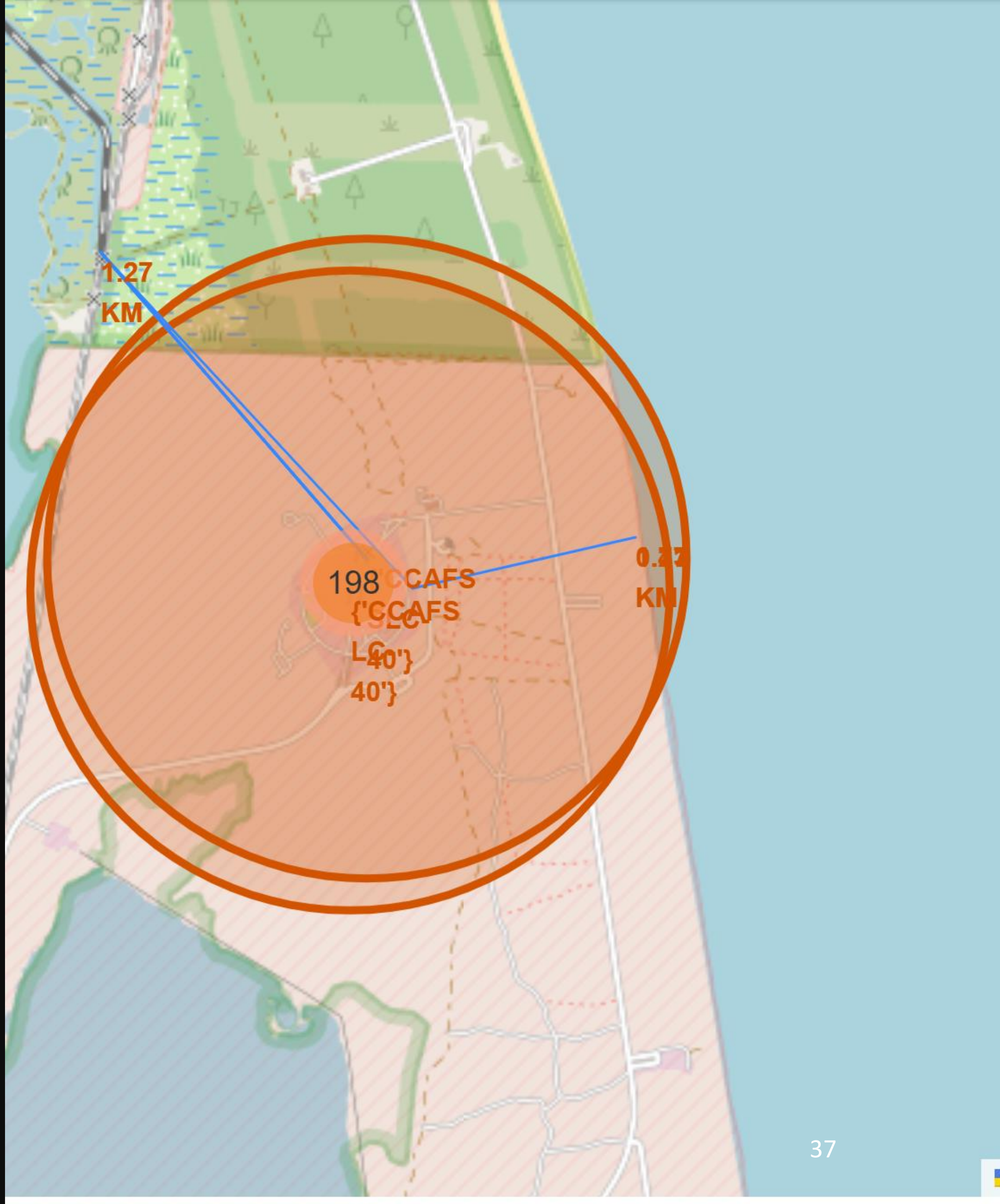
COLOR-LABELED MARKERS

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



LAUNCH SITE DISTANCE TO PROXIMITIES

- The distance from CCAFS –slc40 to the coastline is **0.9km**
- The distance to rail line is **1.27km**

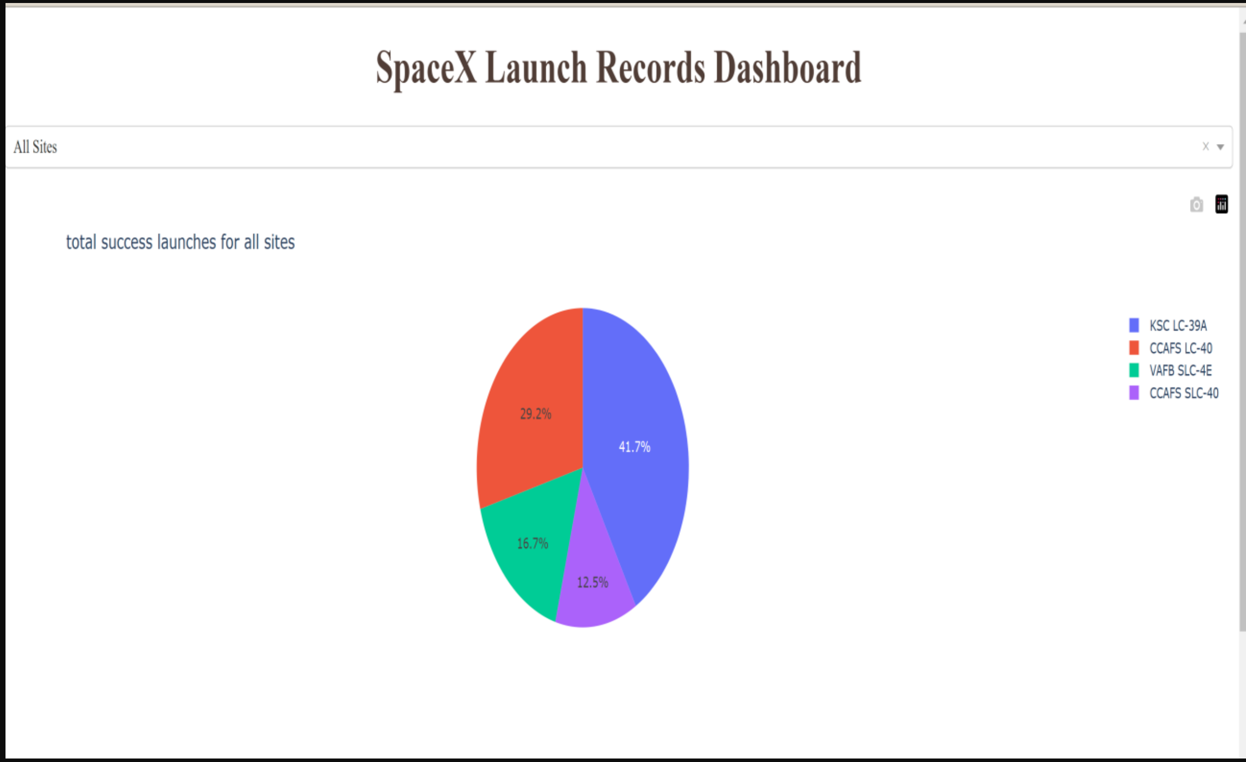


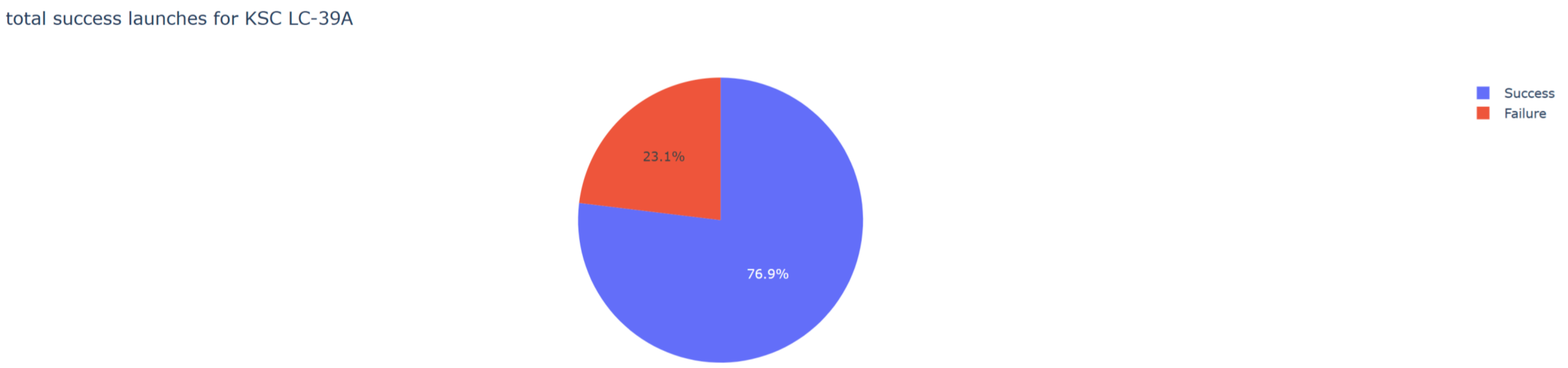
Section 4

Build a Dashboard with Plotly Dash

Dashboard PIE-CHART

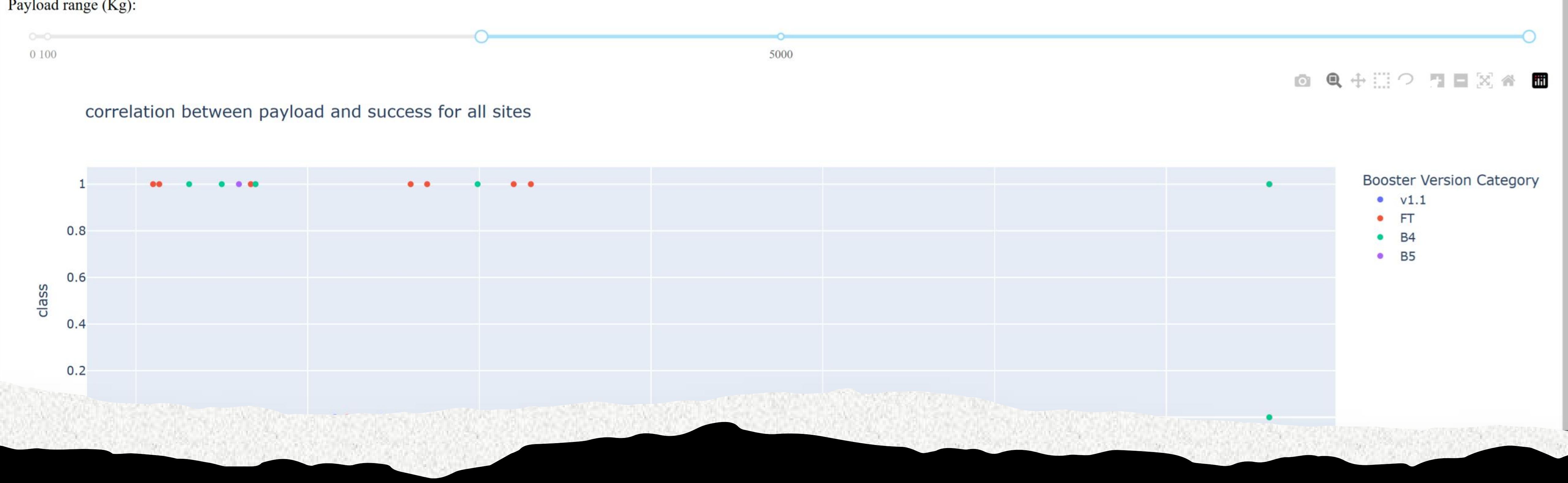
The image attached by the side is a launch success count for all sites, in a pie chart.





PIE-CHART FOR LAUNCH SITE WITH HIGHEST SUCCESS RATE

- The picture above Show the screenshot of the piechart for the launch site with highest launch success ratio(KSC LC-39A)
- Success rate: 76.9%
- Failure rate: 23.1%



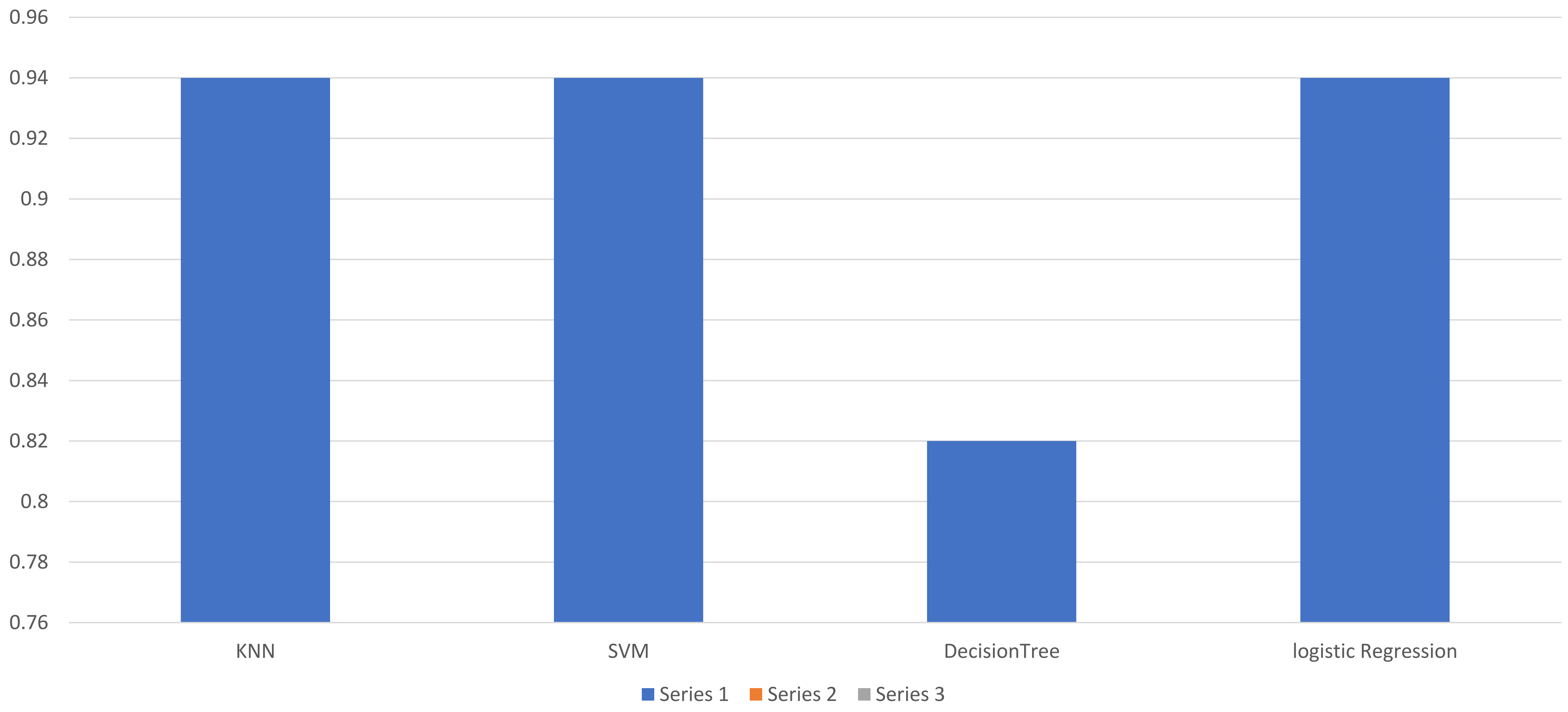
Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider

- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- The range slider allows for selection of a specific range of payload

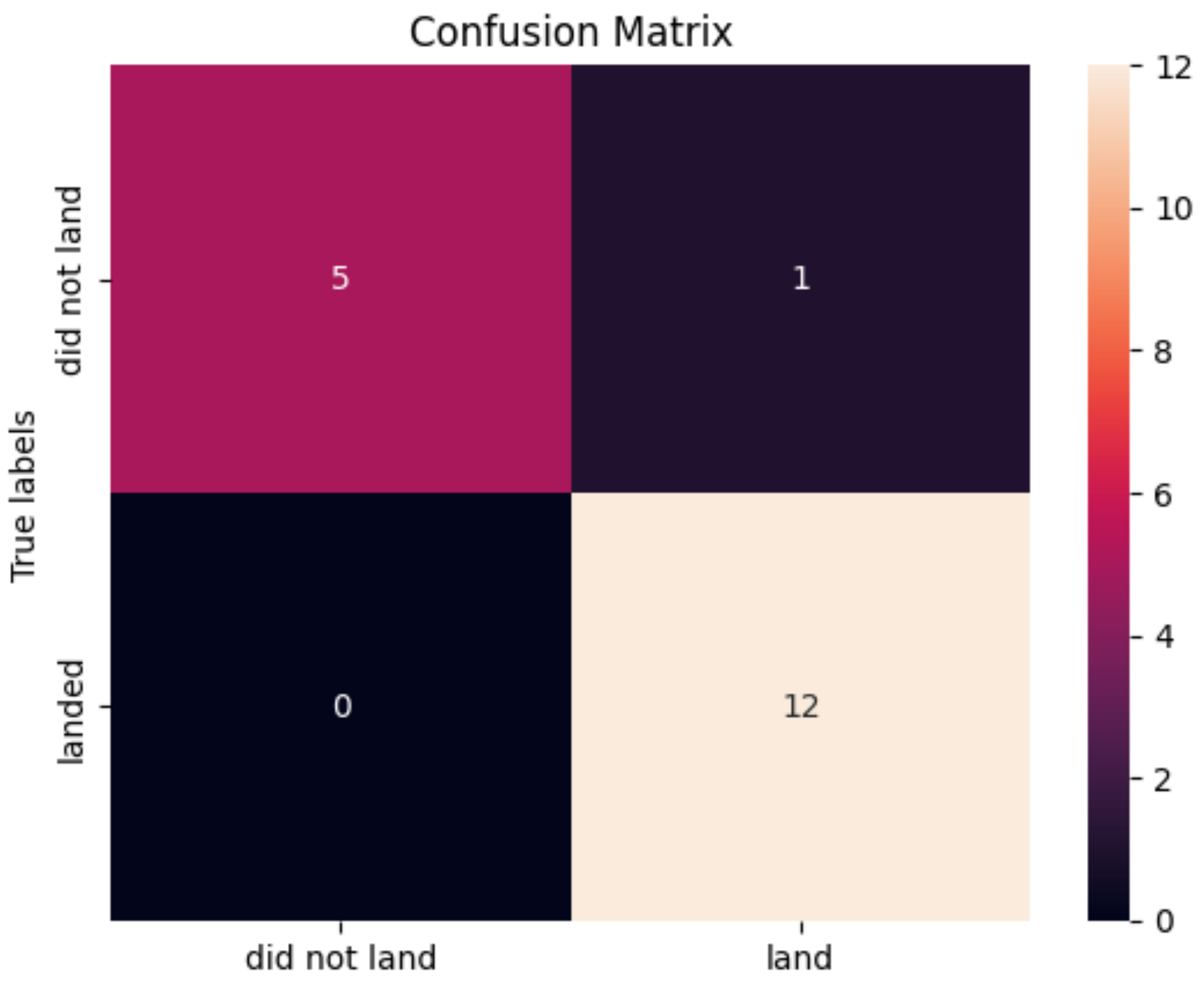
Section 5

Predictive Analysis (Classification)

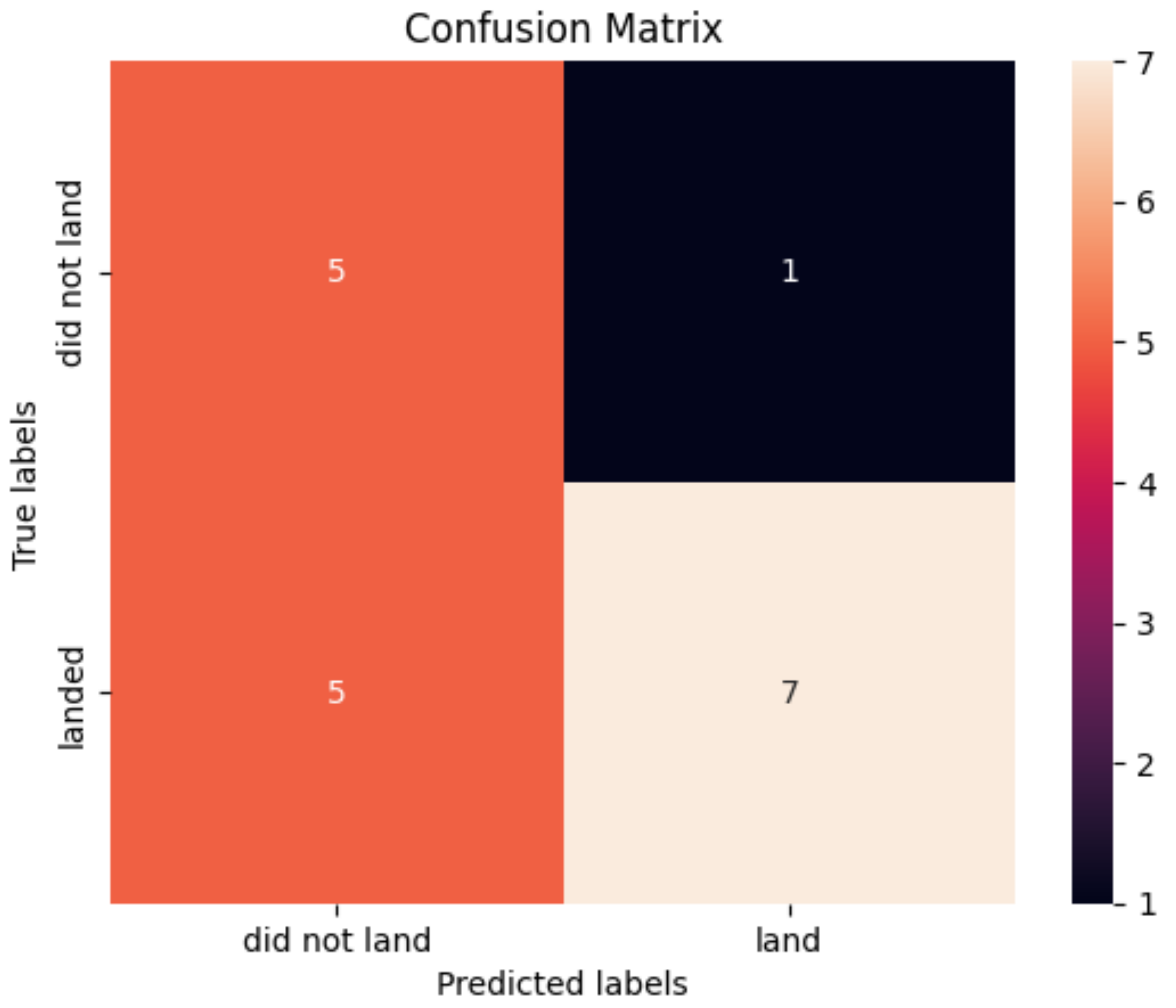
CLASSIFICATION ACCURACY



Confusion Matrix for KNN,SVM and logistic Regression



Confusion Matrix for Decision tree



Conclusions

- **Summary of Key Insights:**
- Success rates vary by **launch site, orbit type, and payload mass**.
- **Flight number impacts success** in some orbits but not others.
- **GIS mapping** highlights the proximity of launch sites to railways and coastlines.
- **Interactive dashboards** allow for deeper exploration of launch outcomes.

- 📈 **Model Performance:**
- **Logistic Regression = Best accuracy**
- **Decision Tree = least accuracy**
- 🚀 **Future Work:**
- Test **Random Forest & Neural Networks** for better accuracy.
- **Factor in weather conditions** for improved predictions.
- **Optimize payload-orbit combinations** to boost success rates

Thank you!