SPACEY

Final Project

Table of contents

3 - Executive summary
4 - Introduction
5 - Methodology
6 - Data Collection
7 - Data wrangling
8 - Exploratory Analysis with SQL
9 - EDA with Visualization
10 - Predictive Analysis Methodology
11 – Results
27 - Conclusion

Executive Summary

Methodologies

- Data collection
- Data wrangling
- Data visualization
- Exploratory data analysis with SQL
- Interactive map with Folium
- Dashboard
- Data analysis results
- Predictive analysis results

INTRODUCTION

SpaceY's accomplishments include:

- Sending spacecraft to the International Space Station.
- Starlink, a satellite internet constellation providing satellite Internet access.
- Sending manned missions to Space.

Objective:

- Determine the price of each launch.

Methodoly

- SpaceX Rest API
- Web Scrapping via Wikipedia
- Data wrangling
- Exploratory data analysis (graphs, charts, data bases)
- Plotting
- Interactive dashboars
- Predictive analysis
- Classification models

Data Collection

- Collect data using SpaceX REST API and web scraping techniques;
- Wrangle data – by filtering the data, handling missing values and applying one hot encoding – to prepare the data for analysis and modeling;
- Explore data via EDA with SQL and data visualization techniques;
- Visualize the data using Folium and Plotly Dash;
- Build Models to predict landing outcomes using classification models.
- Tune and evaluate models to find best model and parameters

Data wrangling

- Perform some *Exploratory Data Analysis (EDA)* to find some patterns in the data and determine what would be the label for training supervised models.
- Perform exploratory Data Analysis and determine:
    - Training Labels
    - Exploratory Data Analysis
    - Determine Training Labels

Exploratory Analysis with SQL

-Understand the Spacex DataSet;
- Load the dataset into the corresponding table in a Db2 database;
- Execute SQL queries to answer assignment questions;

EDA with Visualization

Objectives:

Perform EDA and featuring engineering using Pandas and Matplotlib:

- Exploratory Data Analysis;
- Preparing Data Feature Engineering;

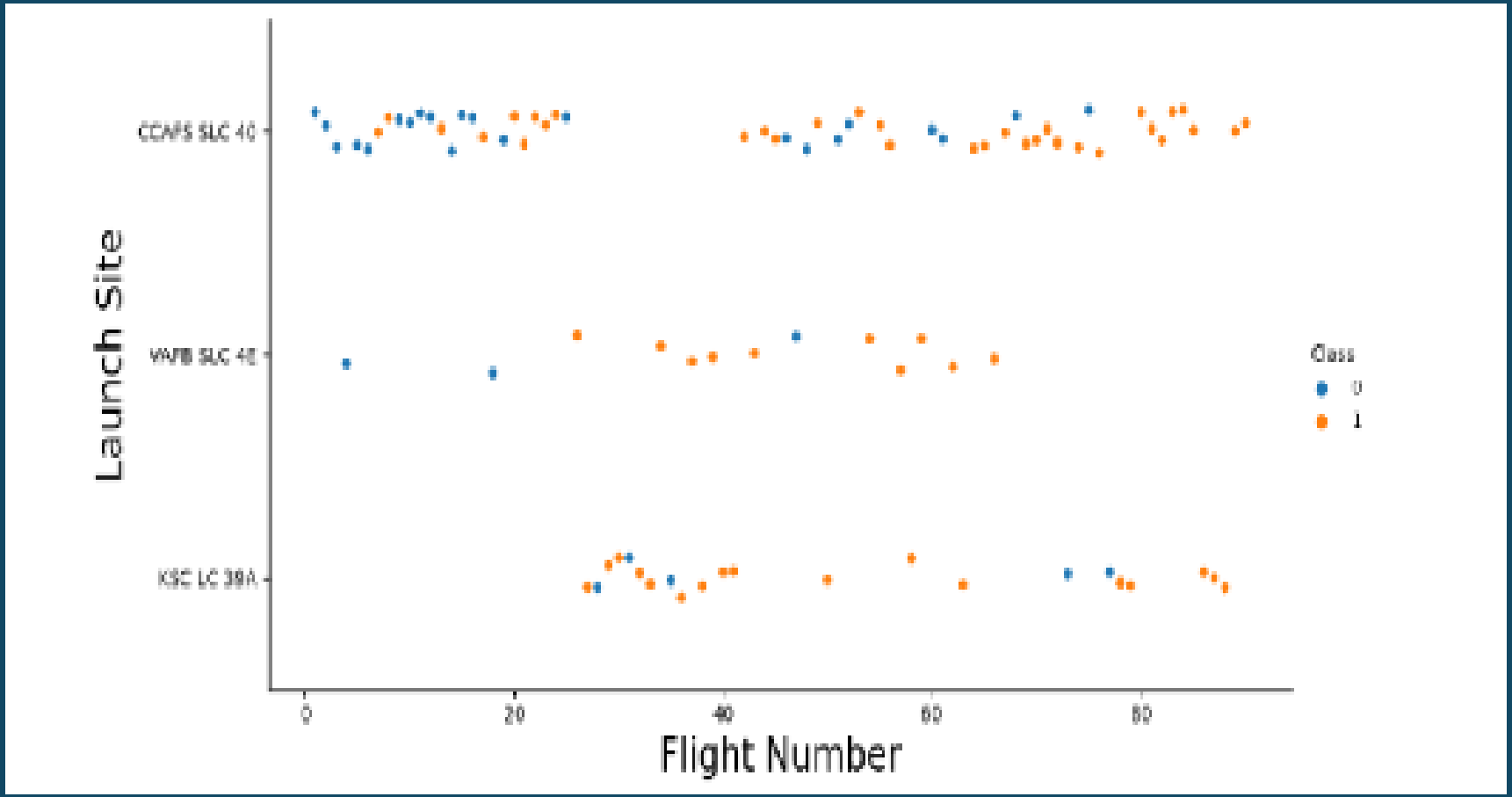
Predictive Analysis Methodology

- Build a machine learning pipeline to predict if the first stage of the Falcon 9 lands successfully.

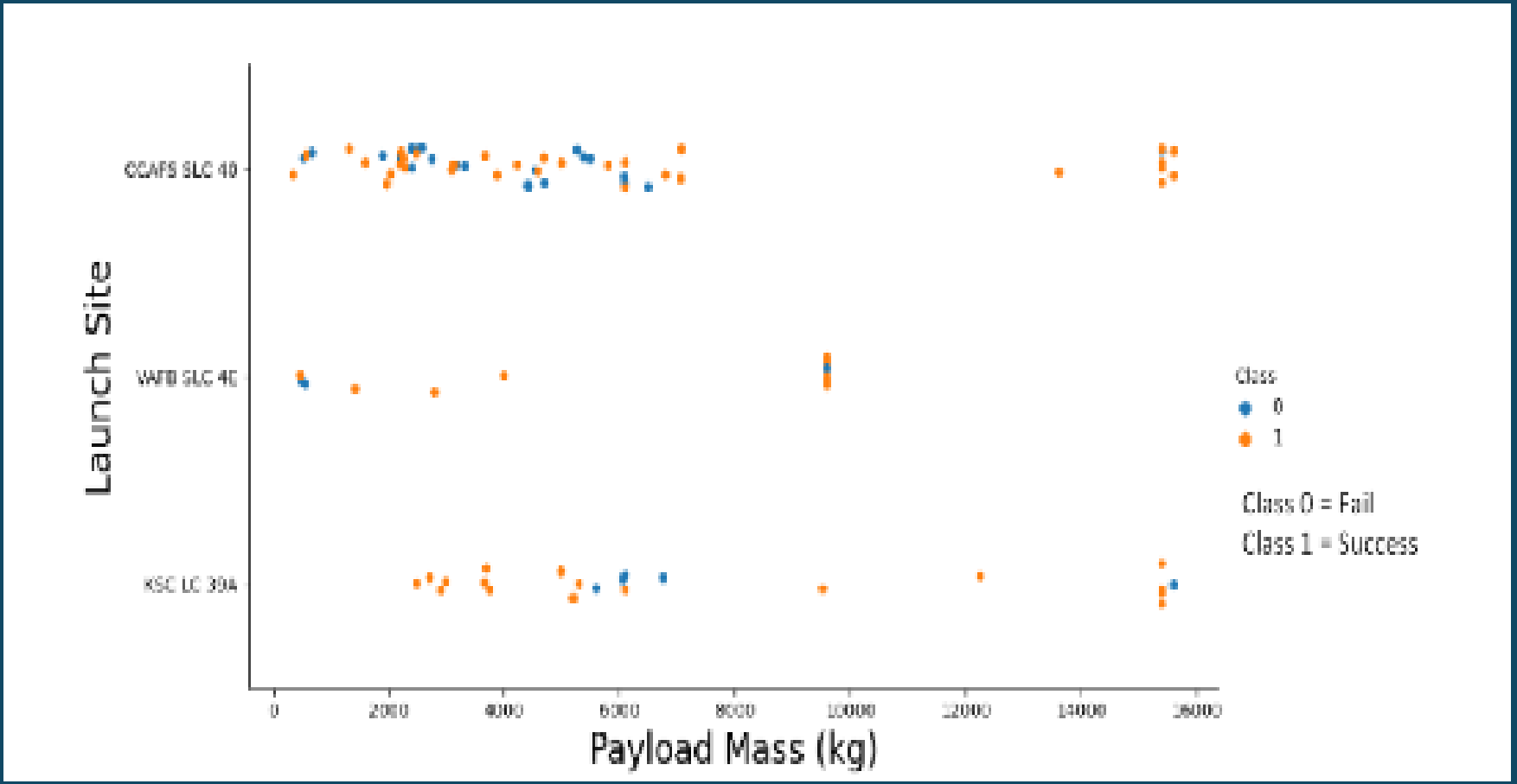
  - Preprocessing
  - Standardize data
  - Train-test-split
  - Train the model and perform Grid Search
    find the hyperparameters that allow a given algorithm to perform best
  - Test ***Logistic Regression, Support Vector machines, Decision Tree Classifier, and K-nearest neighbors***.

RESULTS

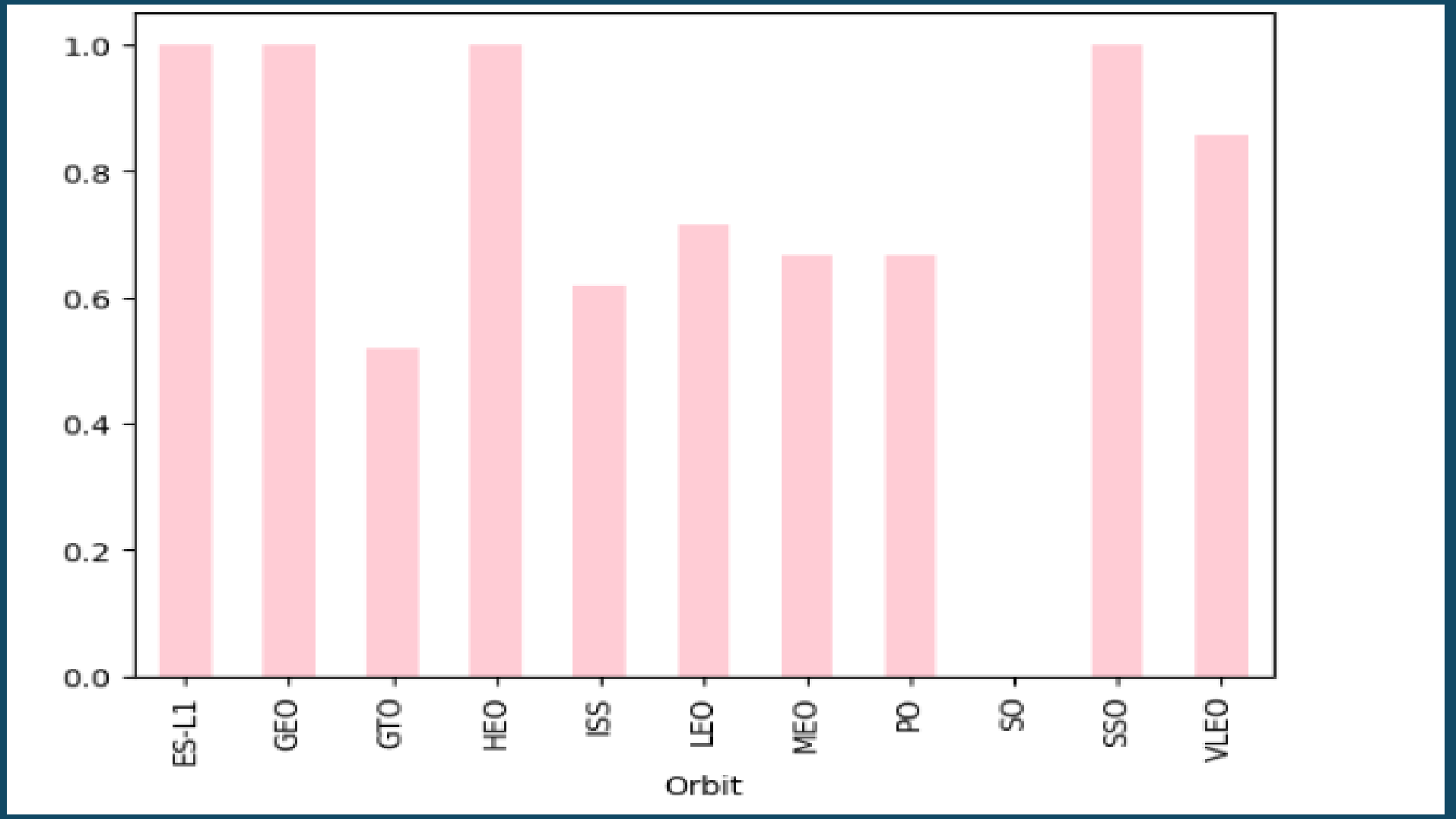
Flight number vs Launch site



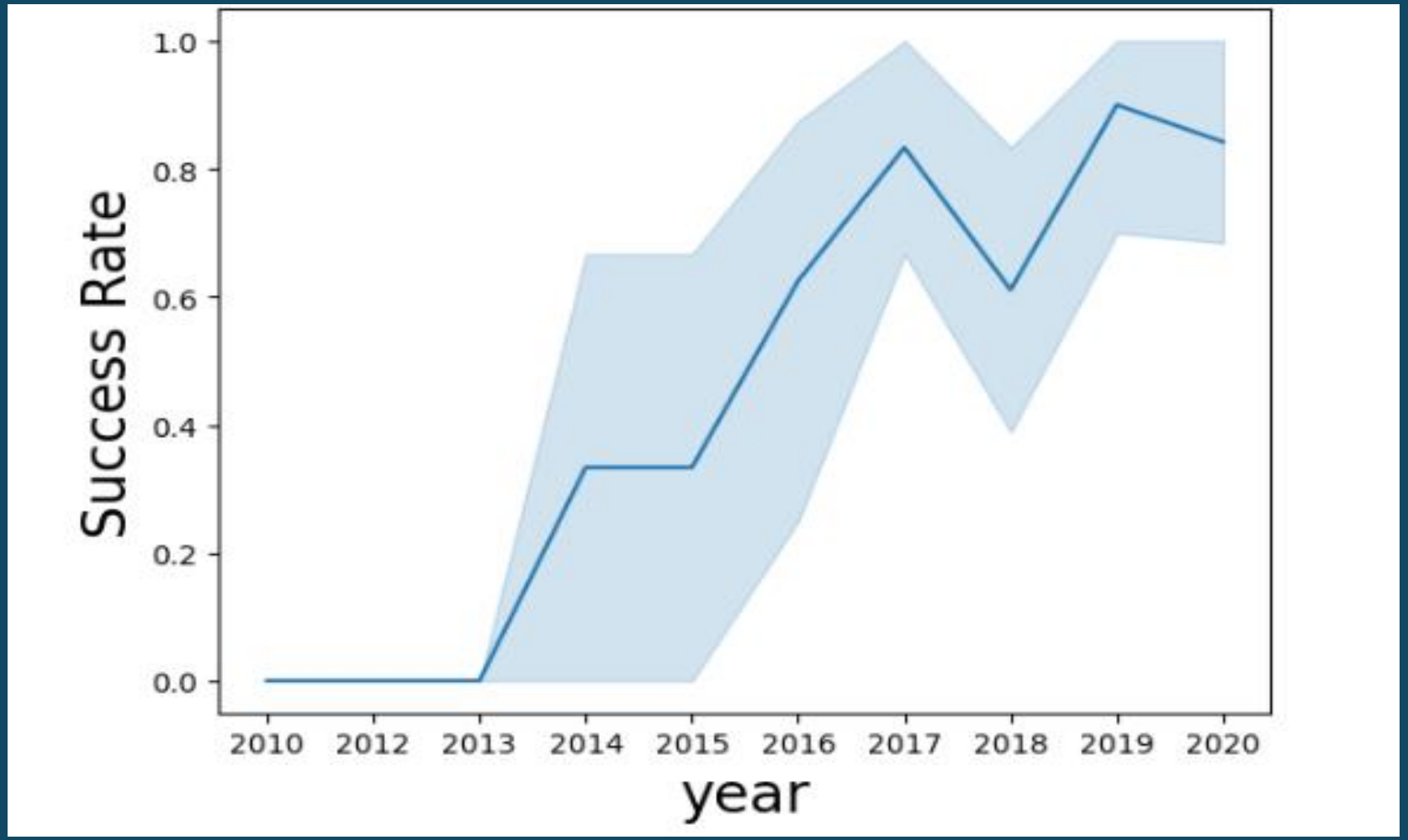
Payload vs Launch site



Success rate by orbit



Success rate



Launch Sites Locations Analysis with Folium

Launch Sites Locations Analysis with Folium

Launch Sites Locations Analysis with Folium

SQL Queries

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) \
    FROM SPACEXTBL \
    WHERE CUSTOMER = 'NASA (CRS)';
```

```
 * sqlite:///my_data1.db
Done.
```

SUM(PAYLOAD_MASS__KG_)
45596

Display average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) \
    FROM SPACEXTBL \
    WHERE BOOSTER_VERSION = 'F9 v1.1';
```

```
 * sqlite:///my_data1.db
Done.
```

AVG(PAYLOAD_MASS__KG_)
2928.4

List the total number of successful and failure mission outcomes

```
%sql SELECT MISSION_OUTCOME, COUNT(*) as total_number \
FROM SPACEXTBL \
GROUP BY MISSION_OUTCOME;
```

```
 * sqlite:///my_data1.db
Done.
```

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

SQL Queries

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%sql SELECT BOOSTER_VERSION \
FROM SPACEXTBL \
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

```
 * sqlite:///my_data1.db
Done.
```

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

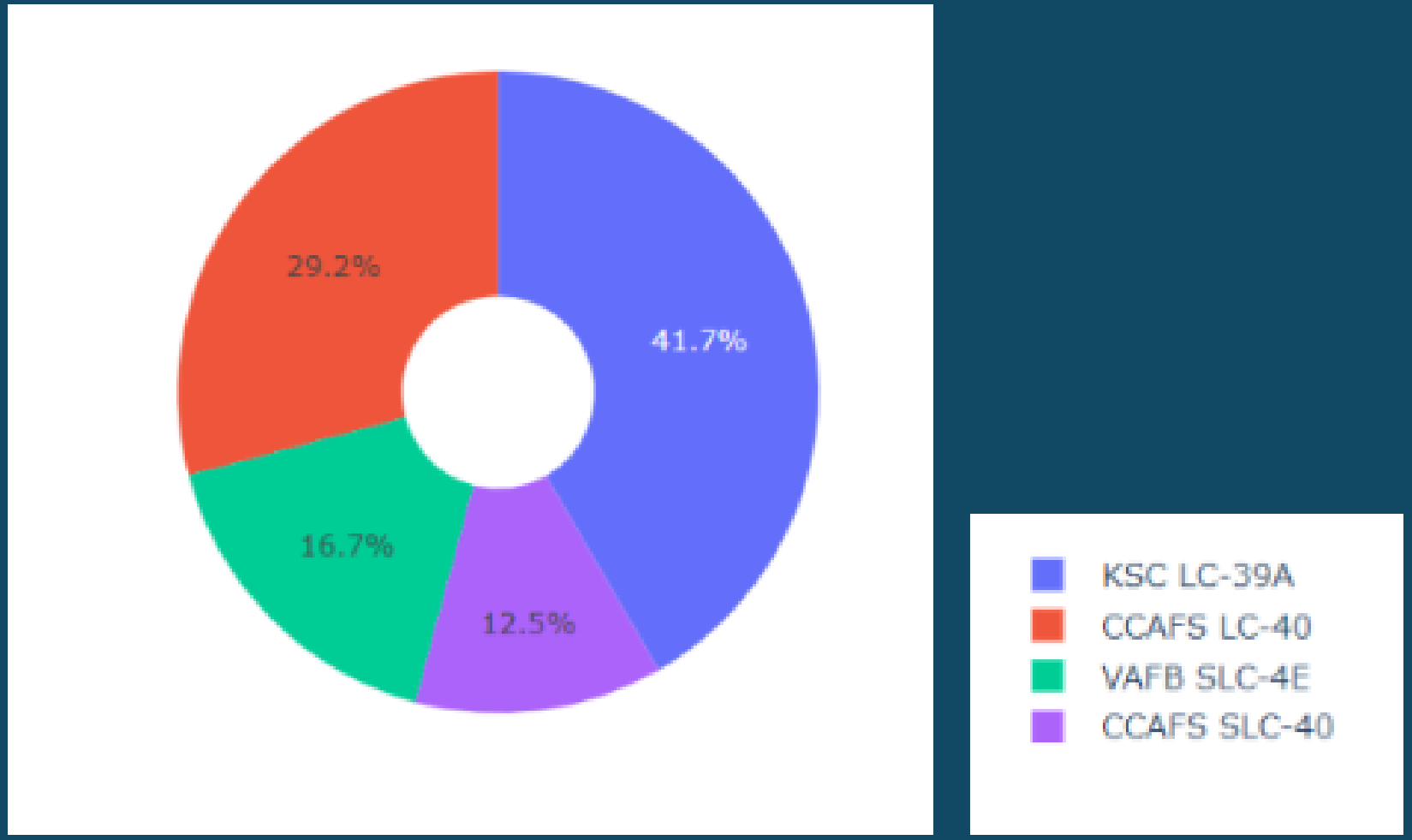
SQL Queries

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.

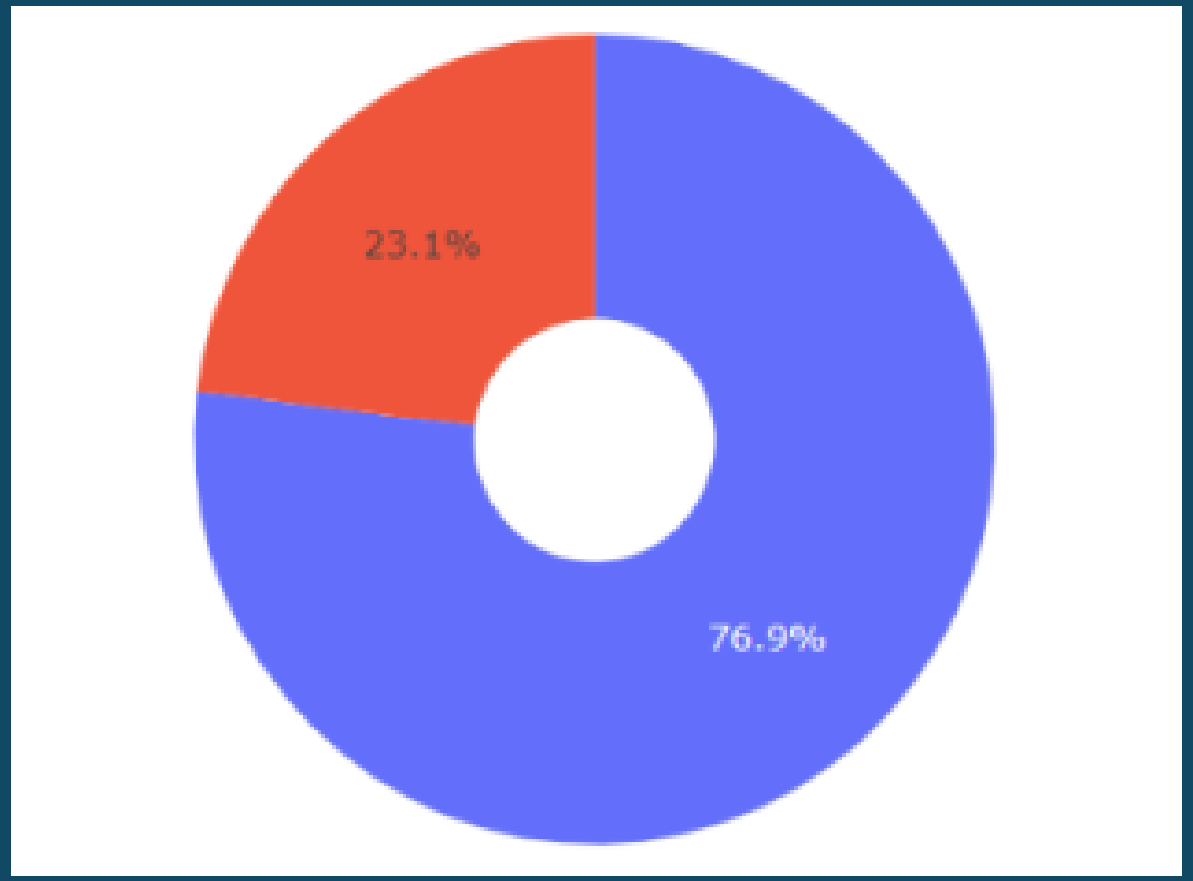
Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
%sql SELECT substr(Date,4,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, [Landing _Outcome] \
FROM SPACEXTBL \
where [Landing _Outcome] = 'Failure (drone ship)' and substr(Date,7,4)='2015';
```

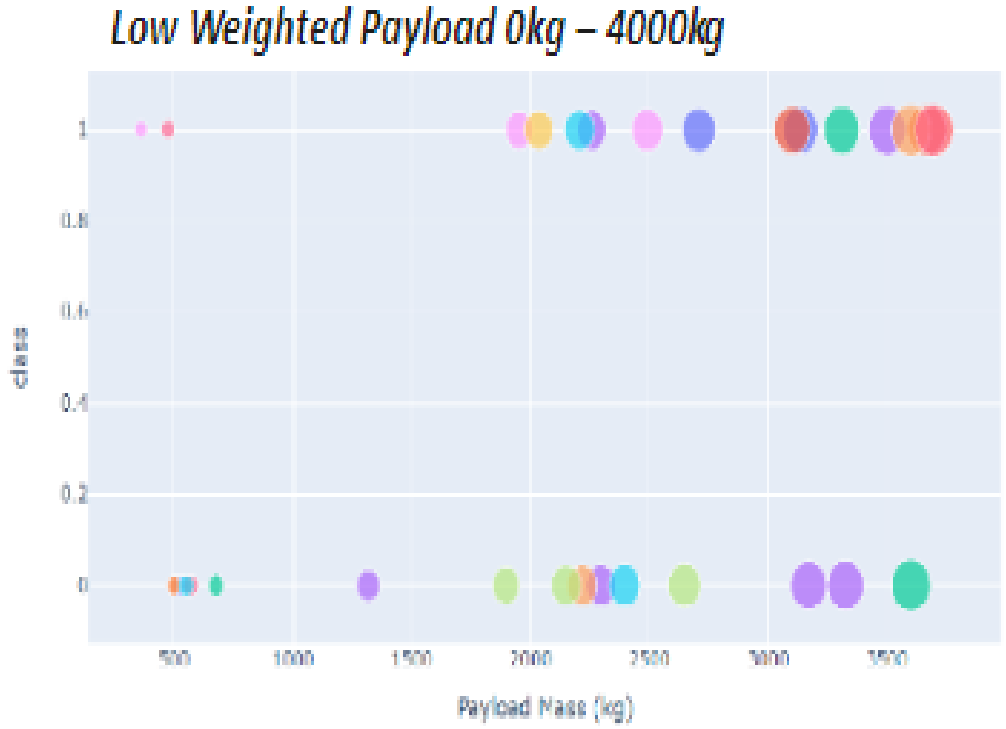
Plotly Dash dashboard

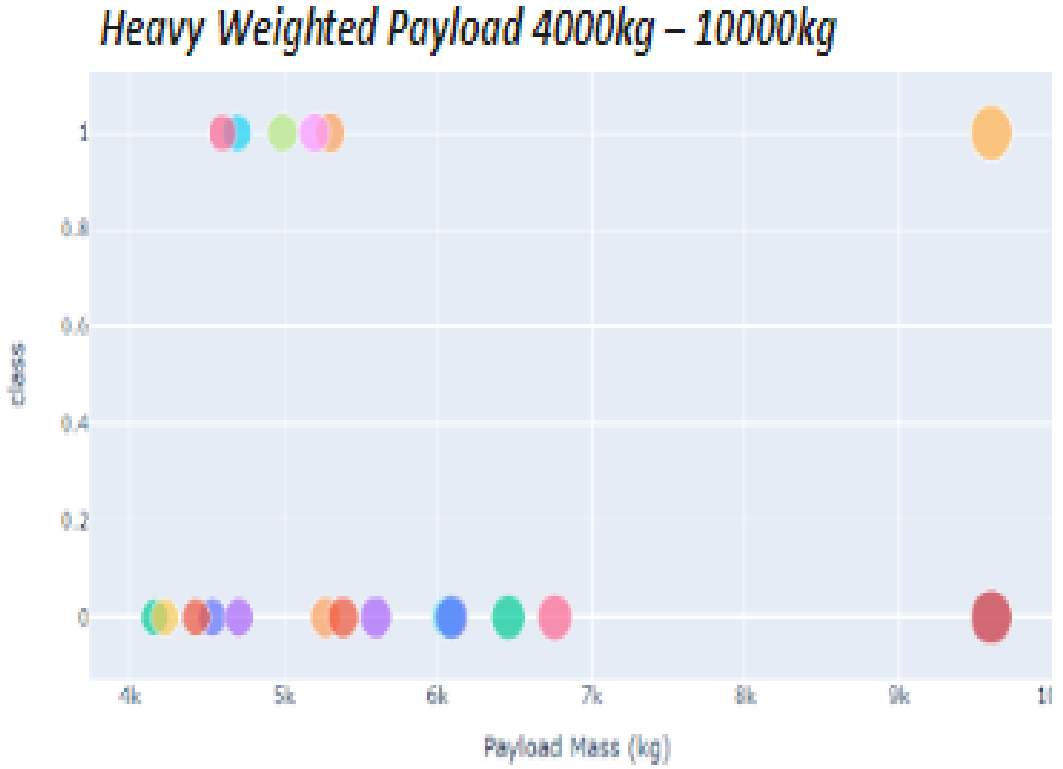


Plotly Dash dashboard



Plotly Dash dashboard

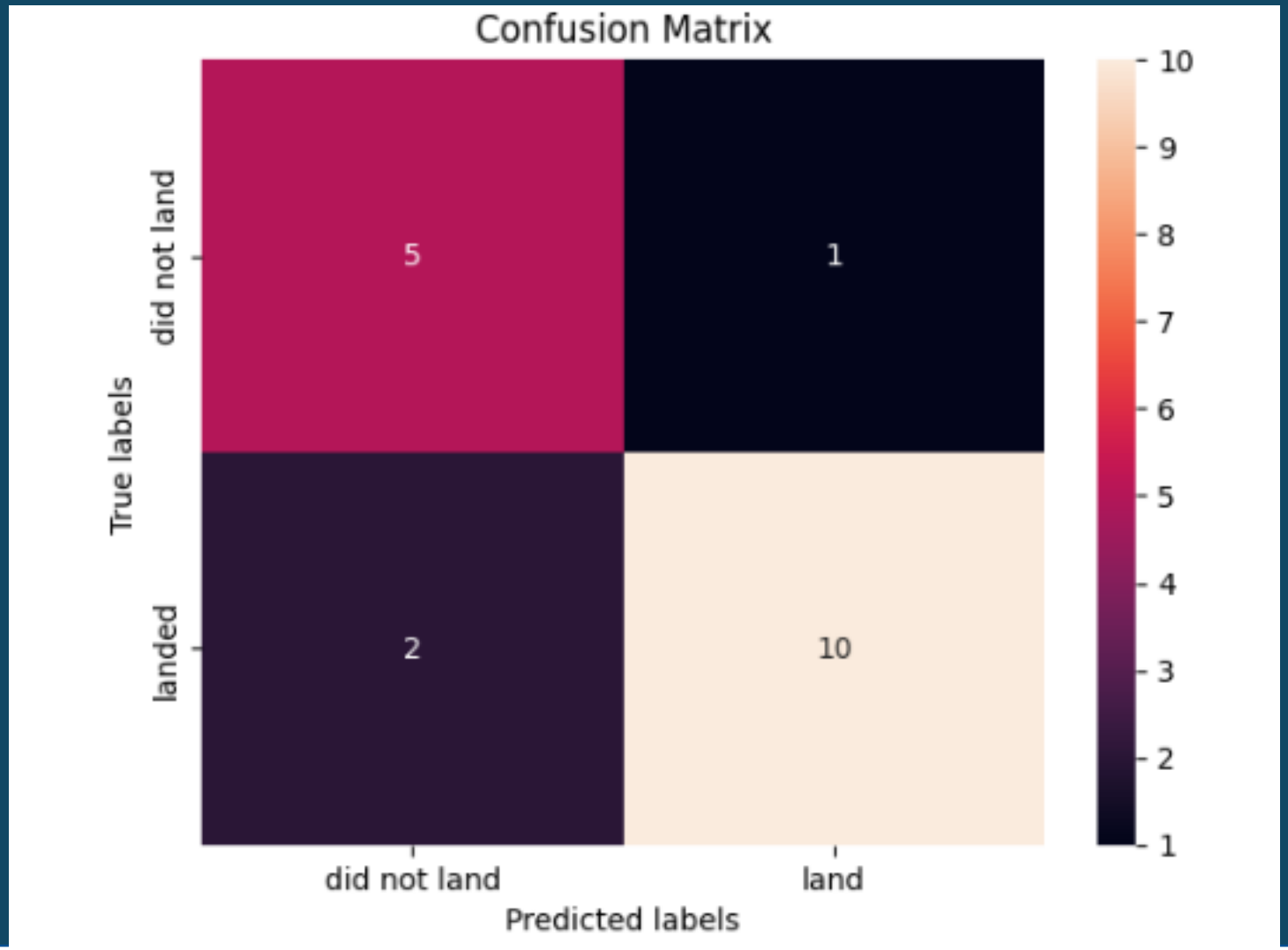




Predictive analysis (Classification)

Algorithm	Accuracy
KNN	0.6535
Tree	0.6678
Logistic regression	0.6678

Predictive analysis (Classification)



Conclusion

Based on the present data set, the algorithm that fits better the data is the Tree Classifier Algorithm. In addition, it is possible to infer that load weighted payload are better than the heavier ones. Over time, the rate of success has been increasing sharply.