IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

Anqi Zhang
8/1/2024

# Outline

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

# Executive Summary

- We created machine learning models to predict if the first stage of a Falcon 9 rocket launch will land.
- We gathered and prepared data from varying reliable sources, and identified the key features that influence landing success.
- Through hyperparameter tuning and model selection with cross-validation, we found the optimized model, which achieves 83% accuracy.

# Introduction

- Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars, about 100 million dollars cheaper than other providers.
- That is because Space X can reuse the first stage of the rocket.
- Therefore, we want to predict the cost of a launch by determining if the first stage will land with a machine learning pipeline.
- This information can be used if an alternate company wants to bid against space X for a rocket launch.

Section 1

# Methodology

# Data Collection

- The data sets were collected from the official SpaceX API and the Falcon 9 Wikipedia page.
- Features include the orbit type, the mass of the payload, the launch site, the launching pad, the longitude and latitude of the location, etc.
- The dependent variable to predict is the landing outcome. We focus on whether it succeeds or fails, while the original mission outcome variable also mentions if the rocket is landed on a ground pad, a drone ship, or exploded without landing.

# Data Wrangling

- We filtered the version of the booster to only include the Falcon 9.
- The missing value in the payload mass column is replaced by the mean value of the column.

# EDA with Data Visualization

- Flight Number vs. Launch Site
- Payload vs. Launch Site
- Success Rate vs. Orbit Type
- Flight Number vs. Orbit Type
- Payload vs. Orbit Type
- Launch Success Yearly Trend

# EDA with SQL

- The names of the unique launch sites in the space mission
- The 5 records where launch sites begin with “CCA”
- The total payload mass carried by boosters launched by NASA (CRS)
- The average payload mass carried by bosster version F9 v1.1
- The date when the first successful landing outcome in ground pad was achieved
- The names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- The total number of successful and failure mission outcomes
- The names of the booster versions which have carried the maximum payload mass
- Landing Outcomes Between 2010-06-04 and 2017-03-20

# Build an Interactive Map with Folium

- Marked and circled launch sites.
- Marked the distance between the launch site to the coast.

# Build a Dashboard with Plotly Dash

- Launch success count for all sites
- Launch success count for a specific site
- Launch success vs. Payload Mass

# Predictive Analysis (Classification)

- The features are the flight number, the payload mass, the flight number, the one-hot encoded orbits, launch sites, and landing pads, and more. In total, it has 83 dimensions.
- The dependent variable to predict is the class of landing success or failure.
- Data were split into the training set (80%) and the testing set (20%).
- Machine learning models to select were logistic regression, support vector machine, and decision tree.
- Their hyperparameters and parameters were fitted with grid search and cross validation.
- The accuracy in the test set and the confusion matrix were calculated.

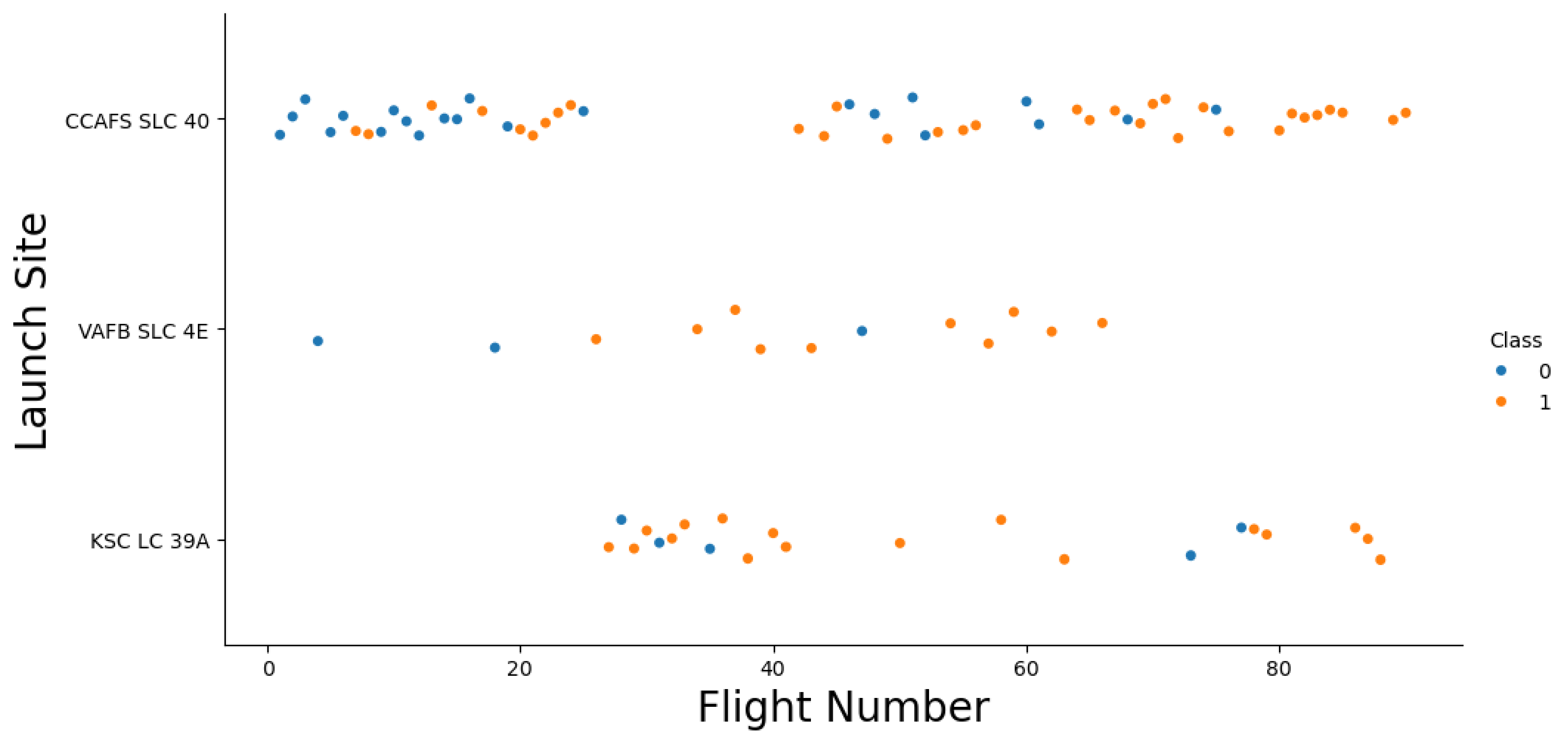
# Results

- For the VAFB-SLC  launch site there are no rockets  launched for  heavy payload mass.
- In the LEO orbit the Success appears related to the number of flights.
- There seems to be no relationship between flight number when in GTO orbit.
- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- The success rate since 2013 kept increasing till 2017.
- The optimal model is the logistic regression model with certain hyperparameters, resulting in an accuracy of 83%.

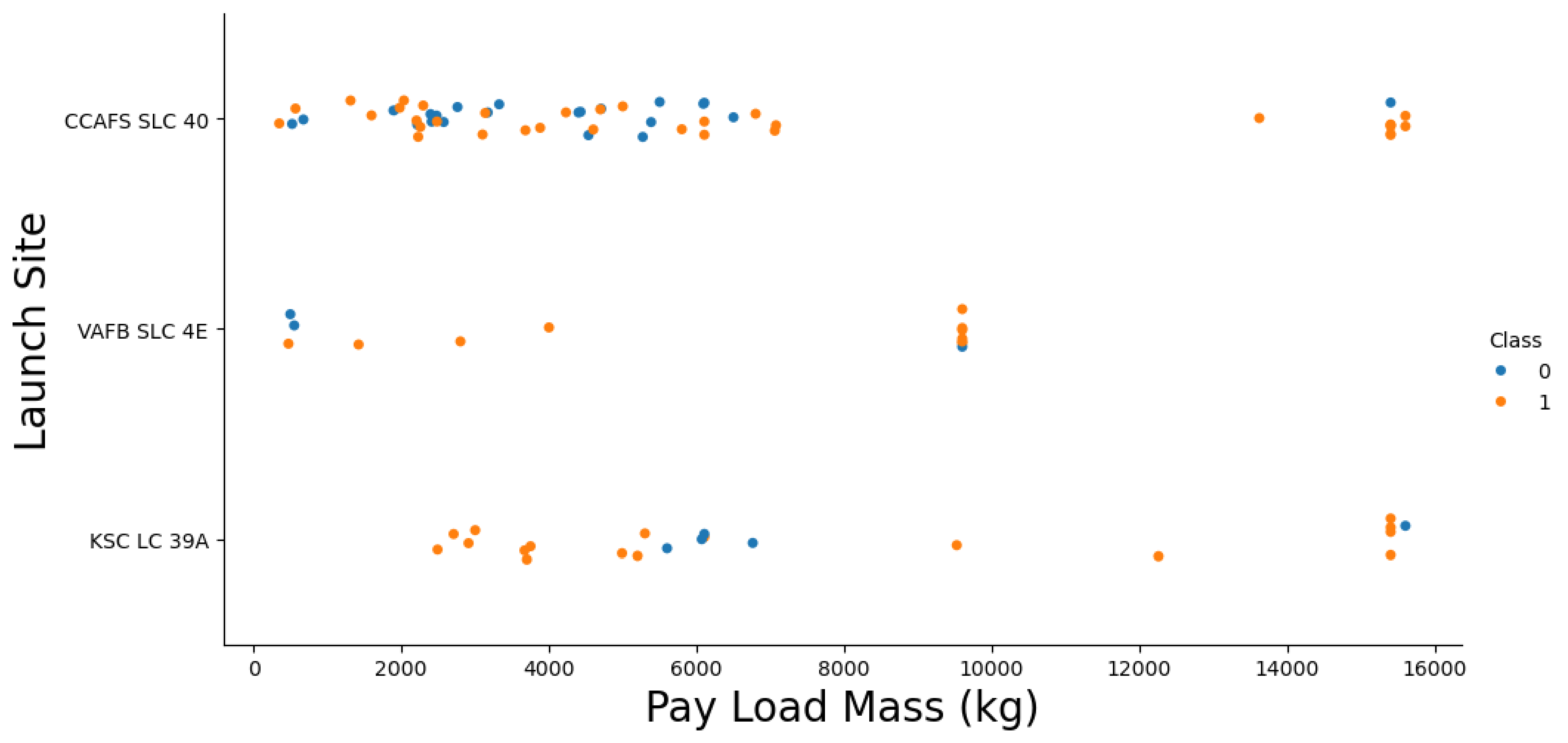
Section 2

# Insights drawn from EDA

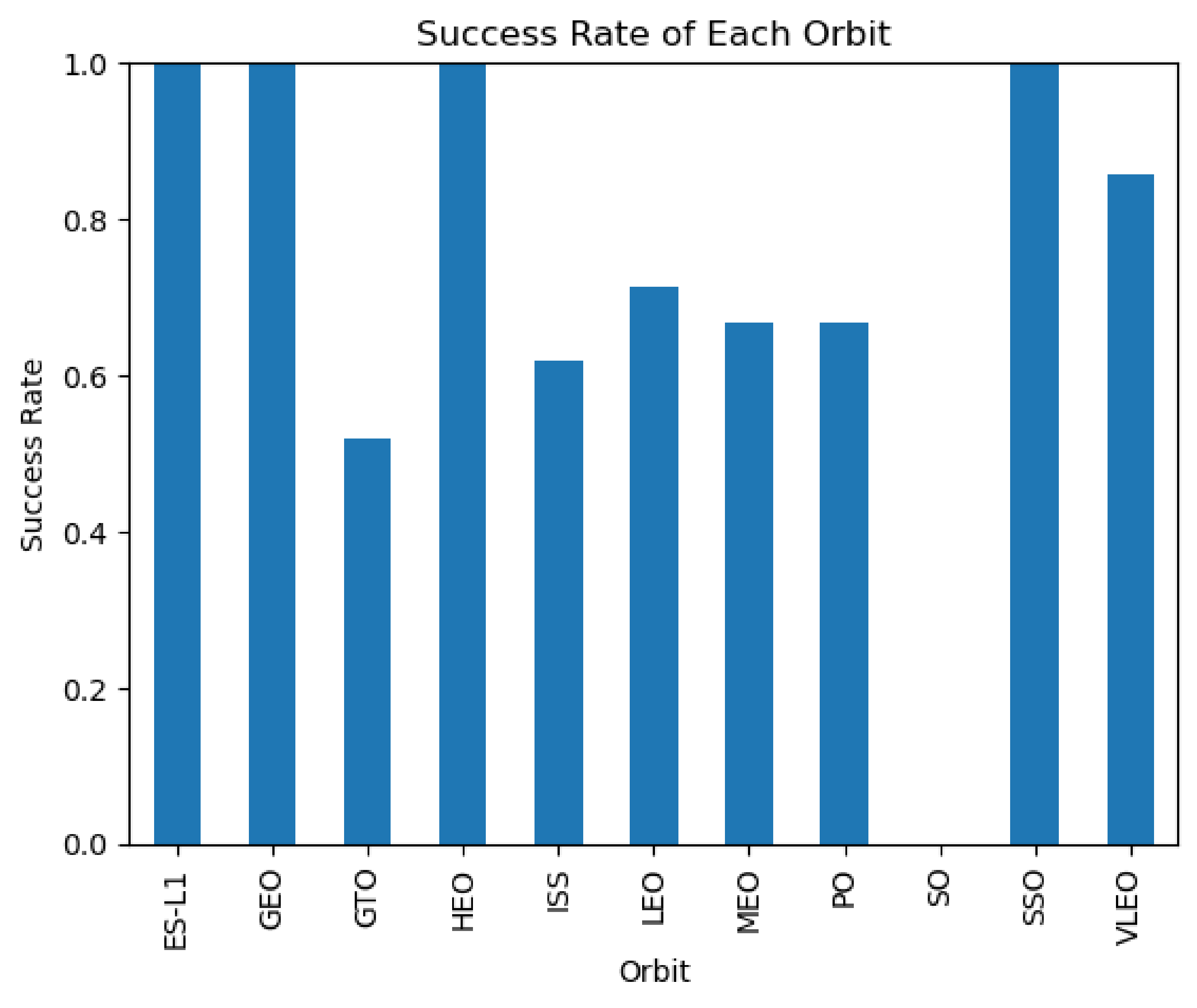
# Flight Number vs. Launch Site



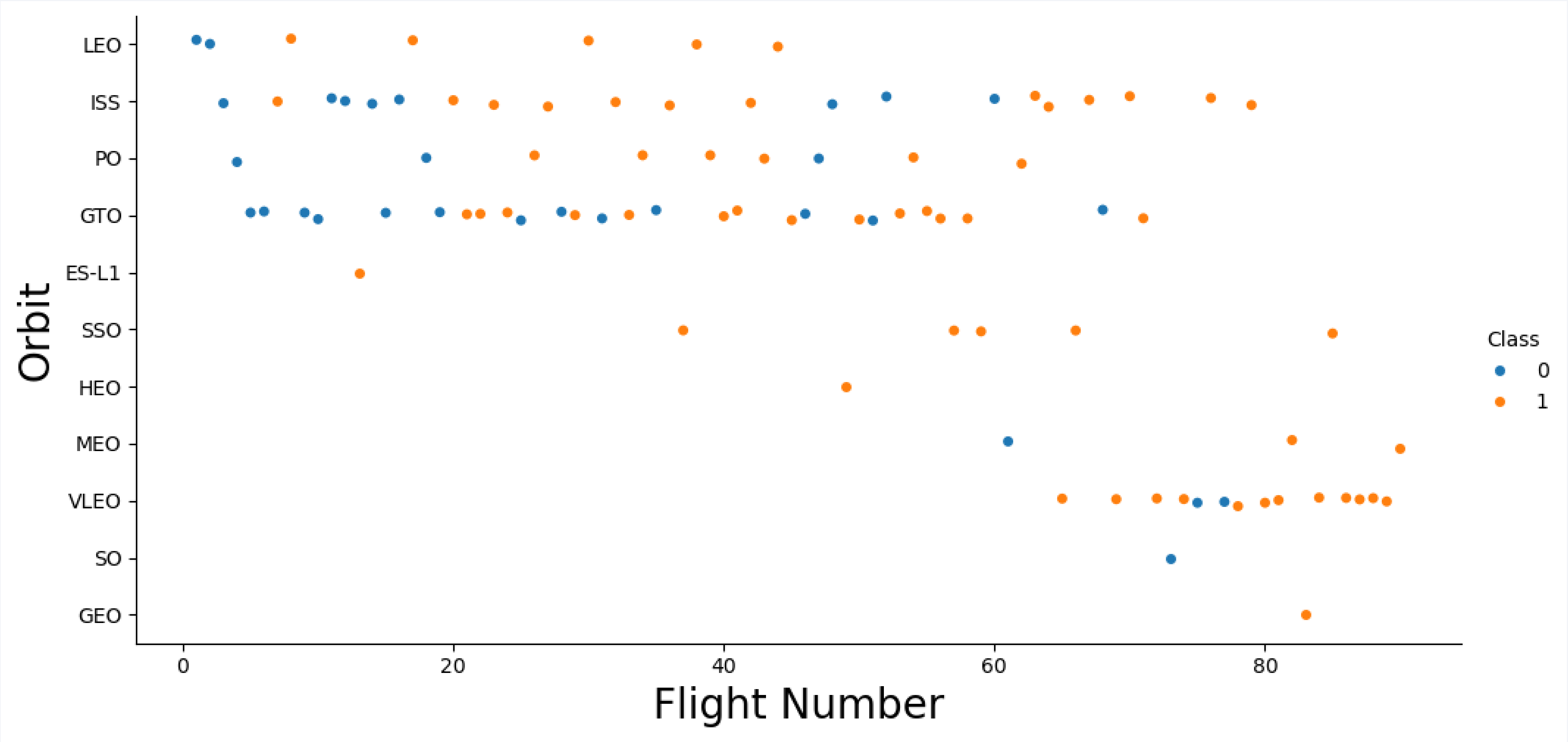
# Payload vs. Launch Site



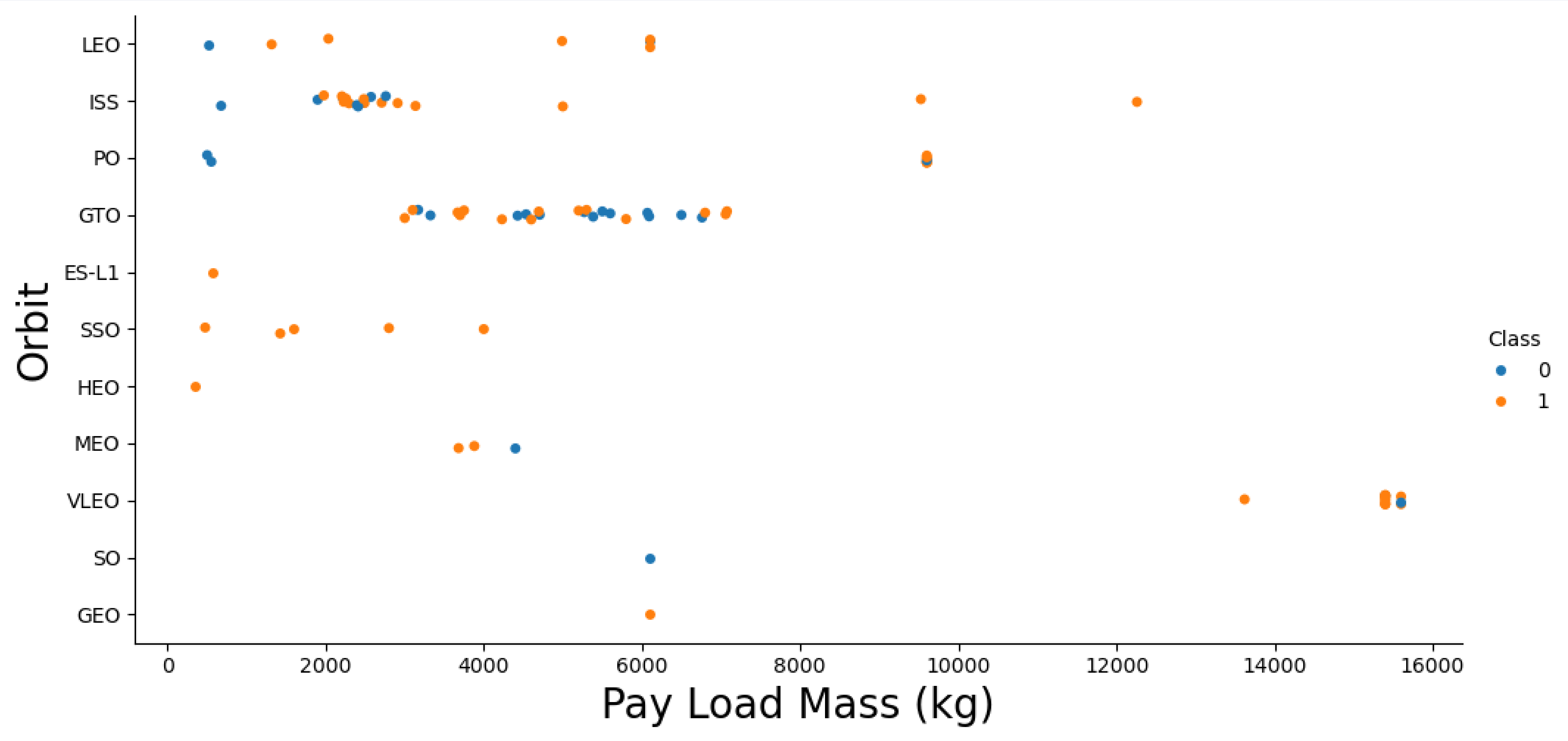
# Success Rate vs. Orbit Type



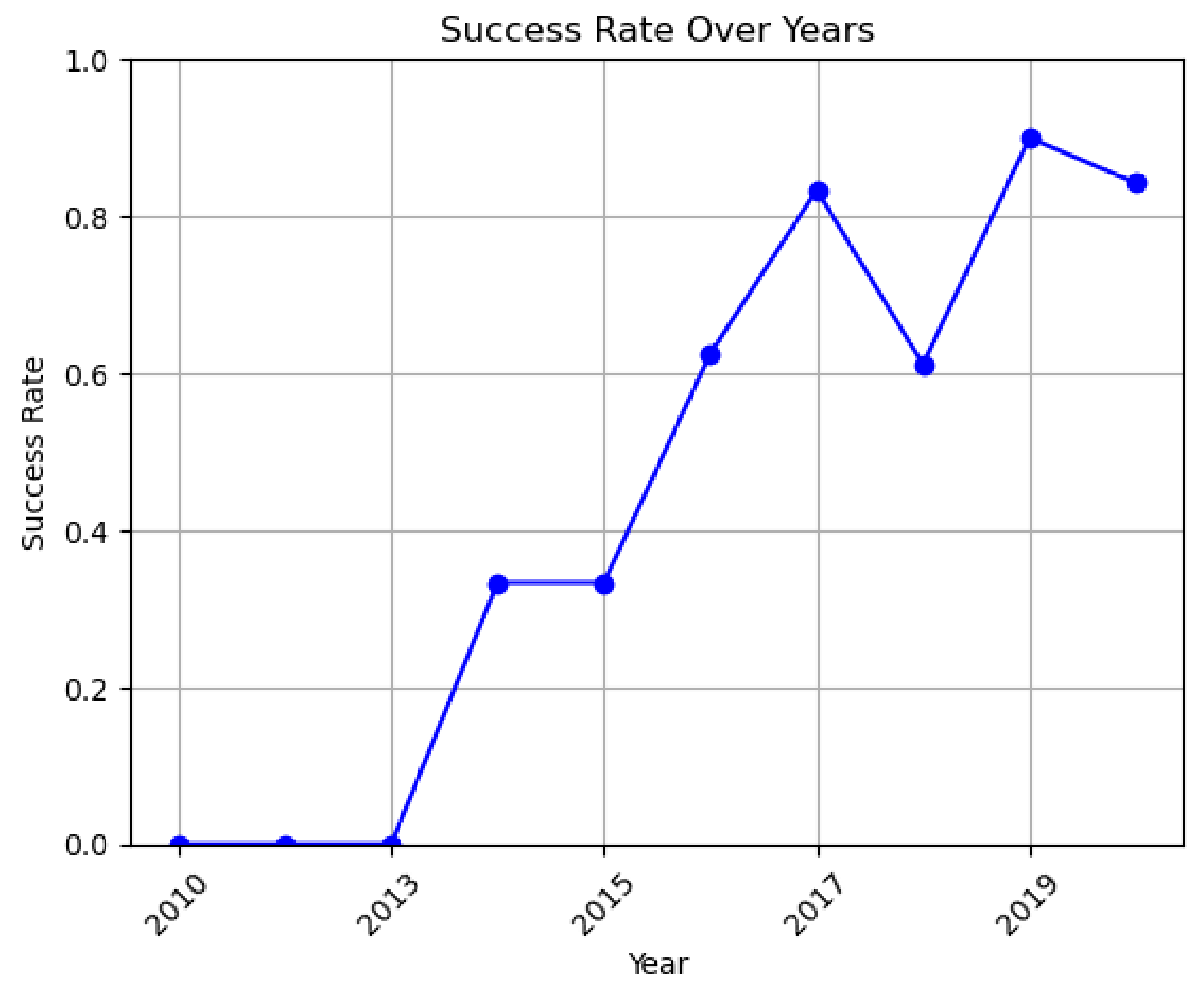
# Flight Number vs. Orbit Type



# Payload vs. Orbit Type



# Launch Success Yearly Trend



# All Launch Site Names

- CCAFS SLC 40: Cape Canaveral Air Force Station Space Launch Complex 40
- VAFB SLC 4E: Vandenberg Air Force Base Space Launch Complex 4E
- KSC LC 39A: Kennedy Space Center Launch Complex 39A

# The first 5 records where launch sites begin with “CCA”

('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

- 45596 Kg

# Average Payload Mass by F9 v1.1

- 2928.4 Kg

# First Successful Ground Landing Date

- 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')

# Total Number of Successful and Failure Mission Outcomes

- Total Success Mission Outcomes: 98
- Total Failure Mission Outcomes: 1

# 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 ')

# 2015 Launch Records

```
(Month, Landing Outome)
('01', 'Failure (drone ship)'),
('01', 'Failure (drone ship)'),
('02', 'Failure'),
('03', 'Failure (drone ship)'),
('03', 'Failure'),
('04', 'Failure (drone ship)'),
('06', 'Failure (parachute)'),
('06', 'Failure (drone ship)'),
('12', 'Failure (parachute)'),
('12', 'Failure')
```

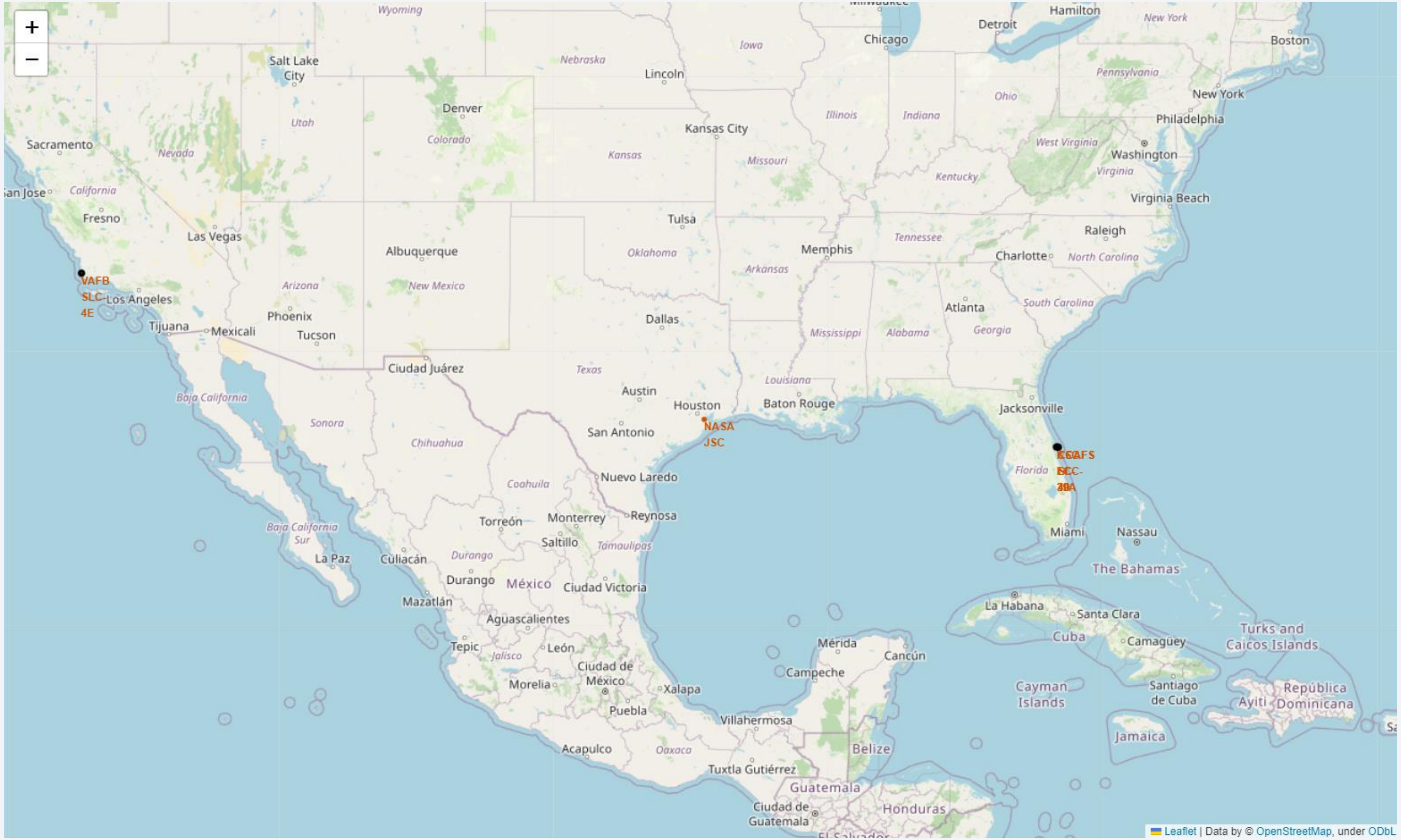
# Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

```
('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)
```

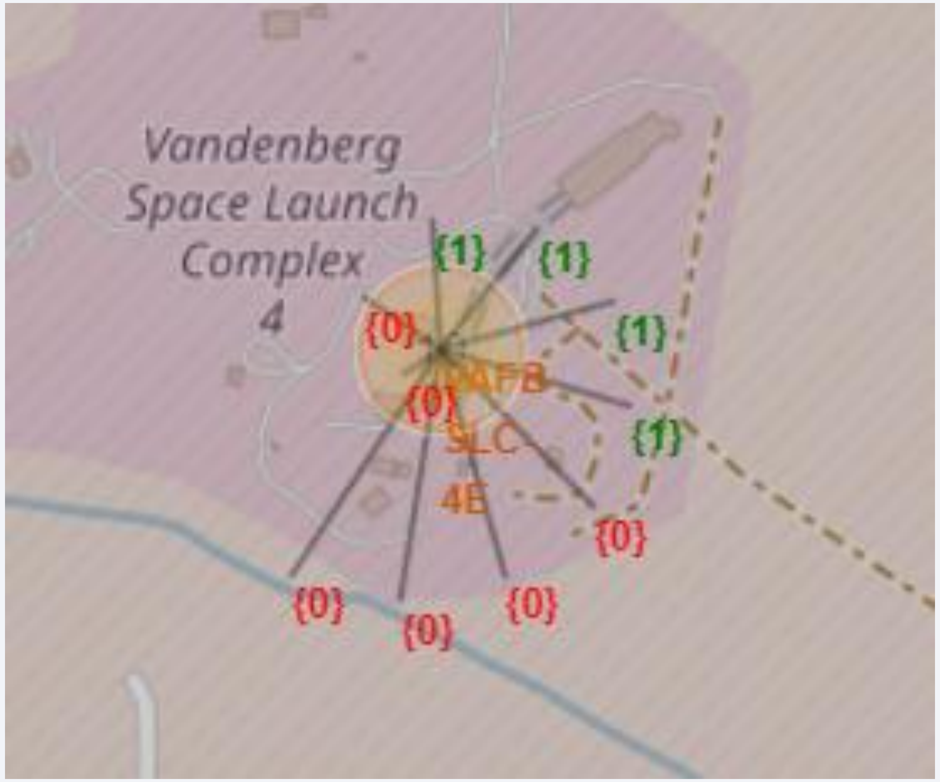
Section 3

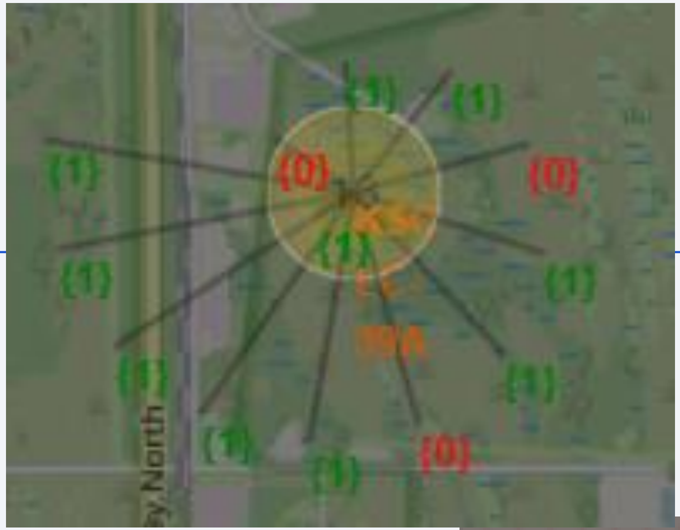
# Launch Sites Proximities Analysis

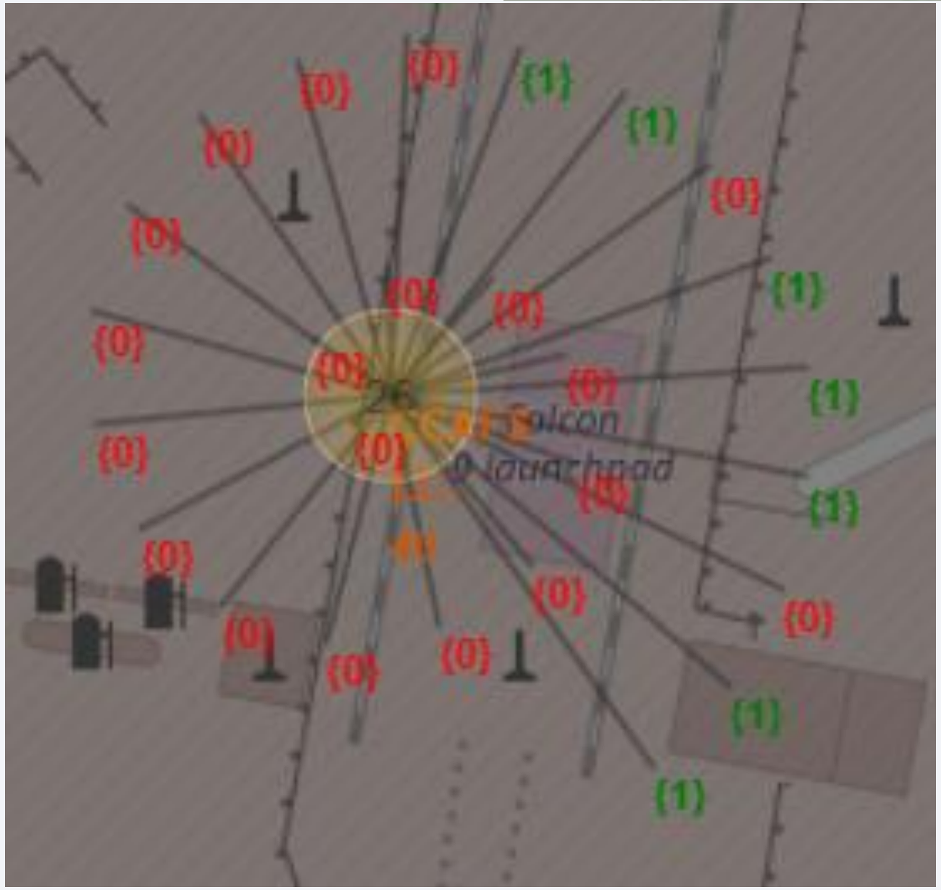
# All launch sites



# Success and failure at each site



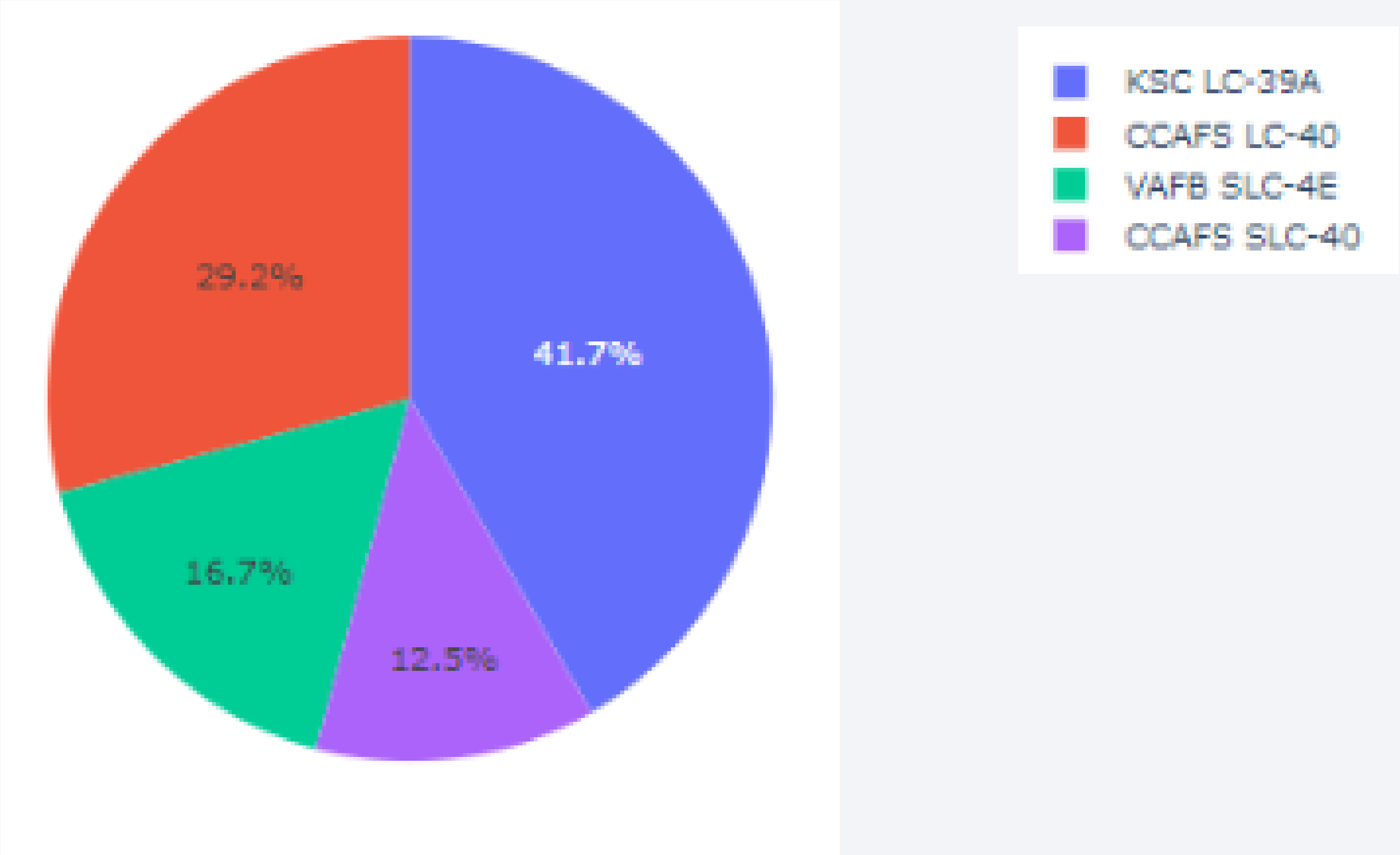


# Distance from the site to the shore

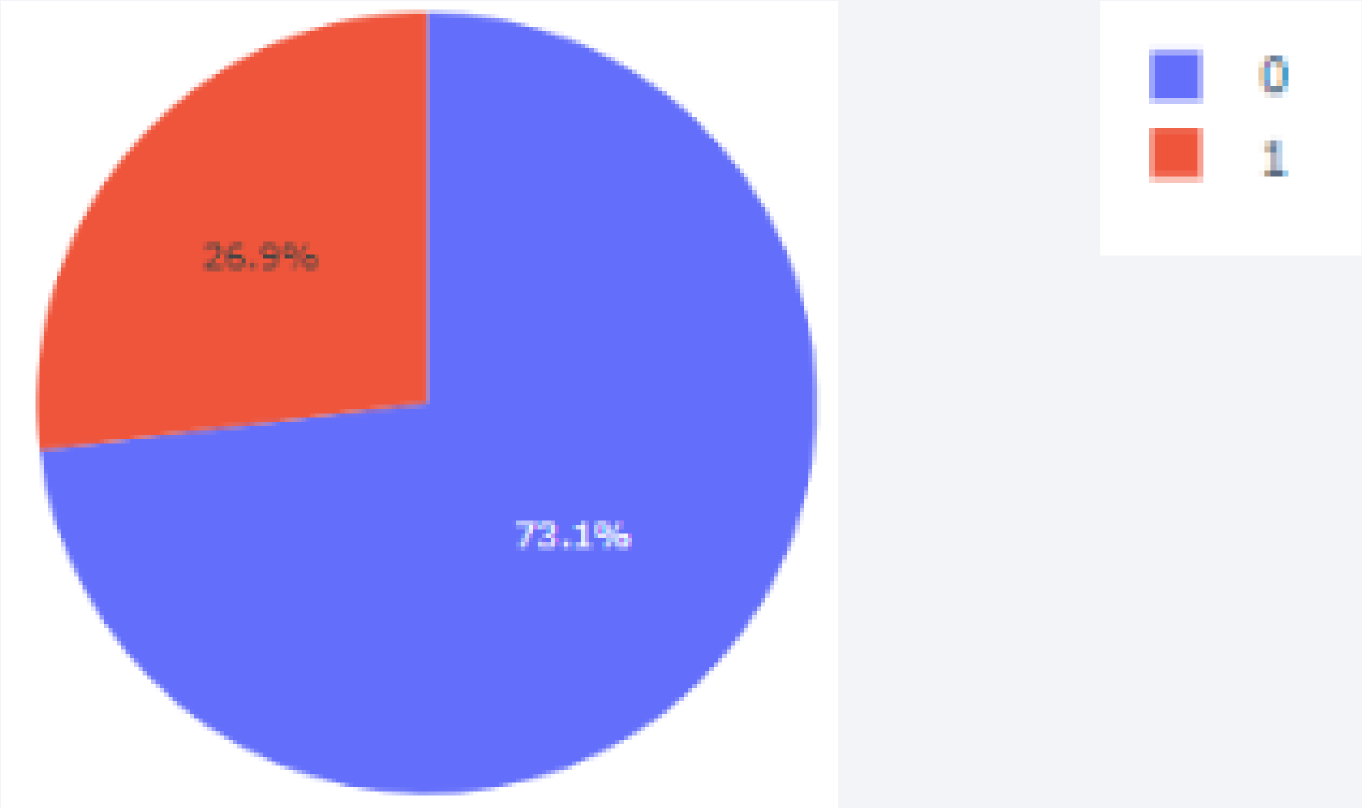
Section 4

# Build a Dashboard with Plotly Dash

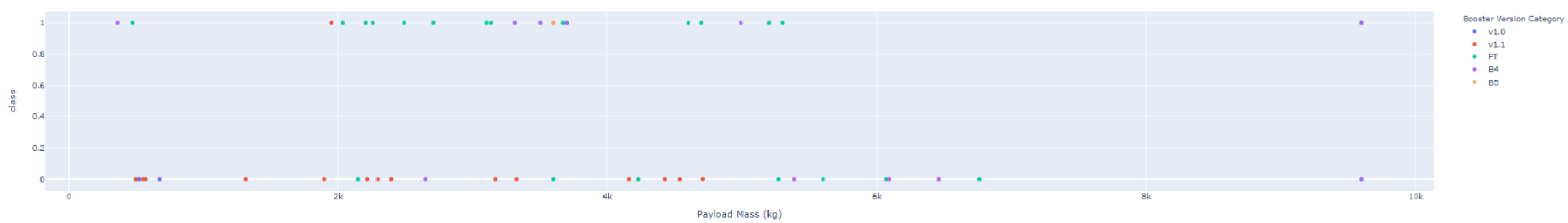
# Launch success count for all sites



# Launch success at launch site CCAFS LC-40



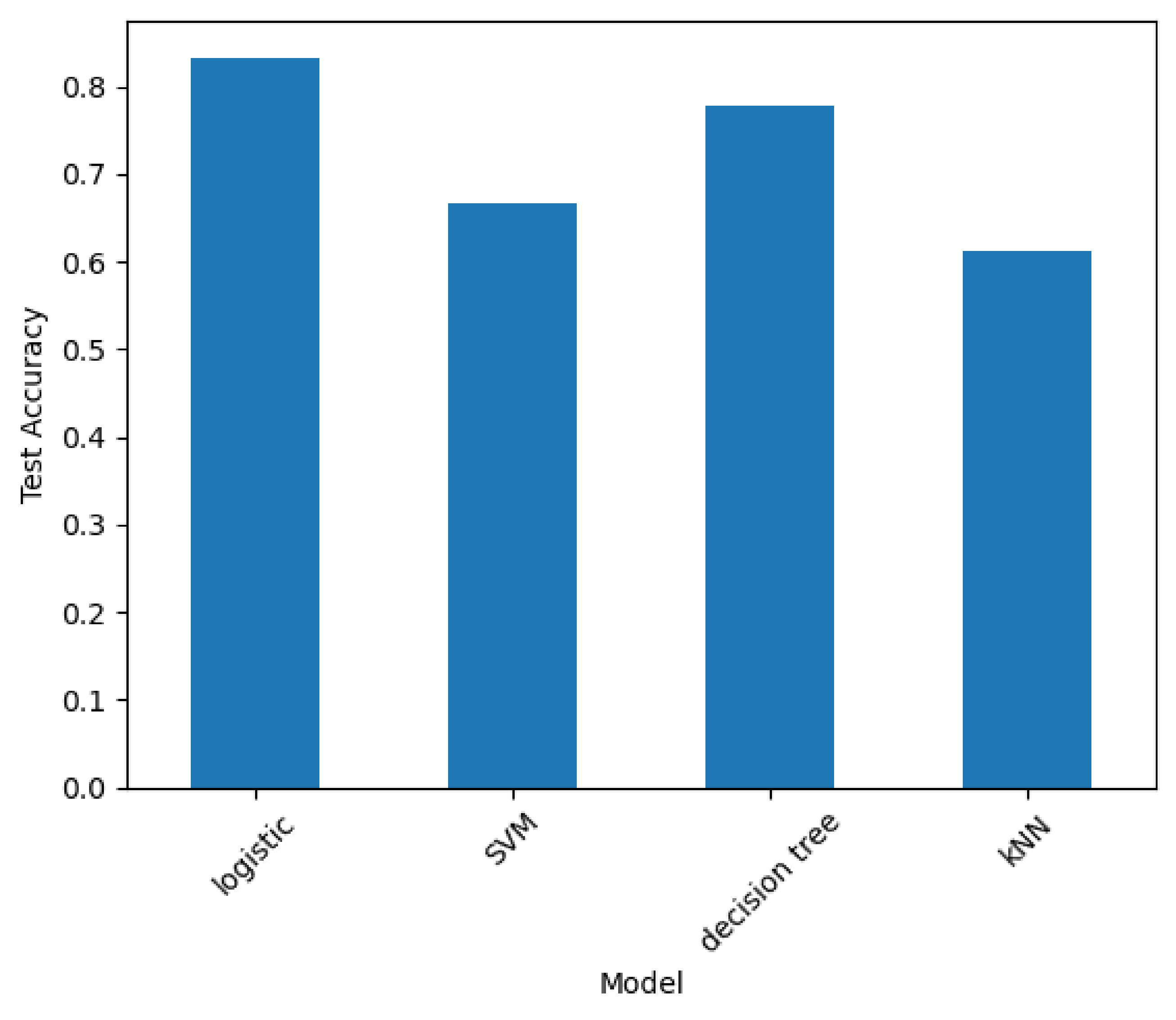
# Payload vs. Launch Outcome



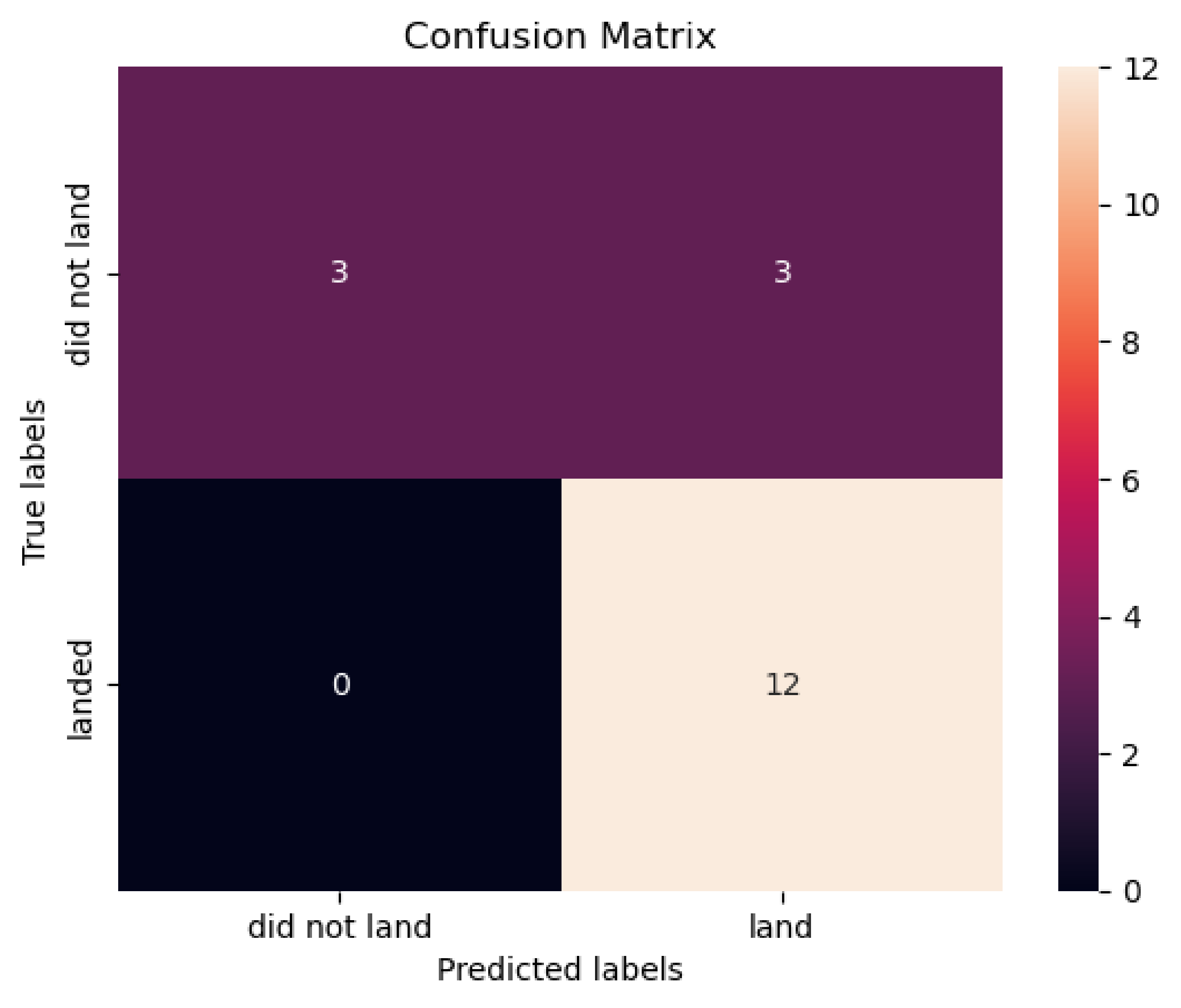
Section 5

# Predictive Analysis (Classification)

# Classification Accuracy



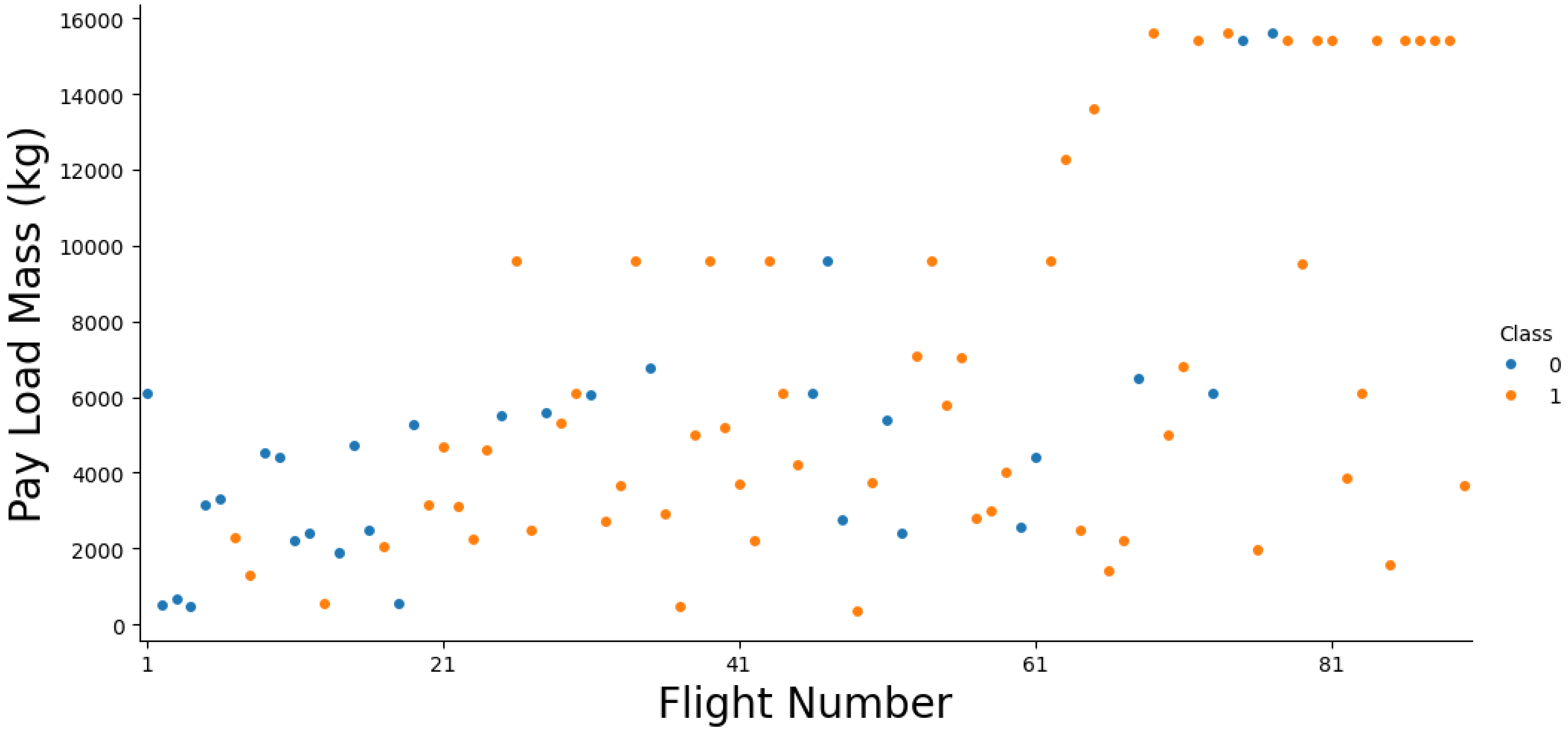
# Confusion Matrix



# Conclusions

1. SpaceX has increasingly successful landing since 2013. The insights from this data analysis will help alternate companies who wants to bid against space X for a rocket launch.
2. All launch sites are close to the coast and the Equator line.
3. For the VAFB-SLC launch site there are no rockets launched for heavy payload mass. In the LEO orbit the Success appears related to the number of flights. With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
4. We have developed a logistic regression machine learning model that predicts the success or failure of Falcon 9 rocket landing with an accuracy of 83%.

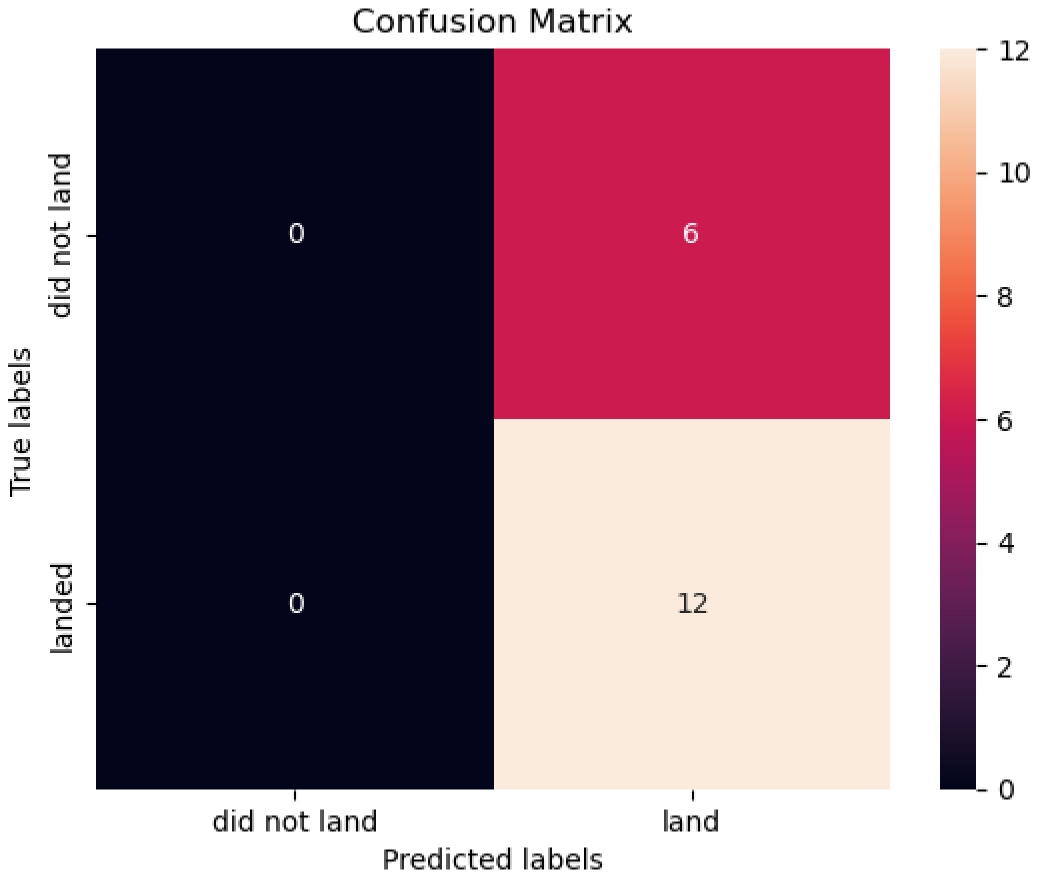
# Appendix



Pay Load Mass (kg)
16000
14000
12000
10000
8000
6000
4000
2000
0
1
21
41
61
81
Flight Number
Class
0
1

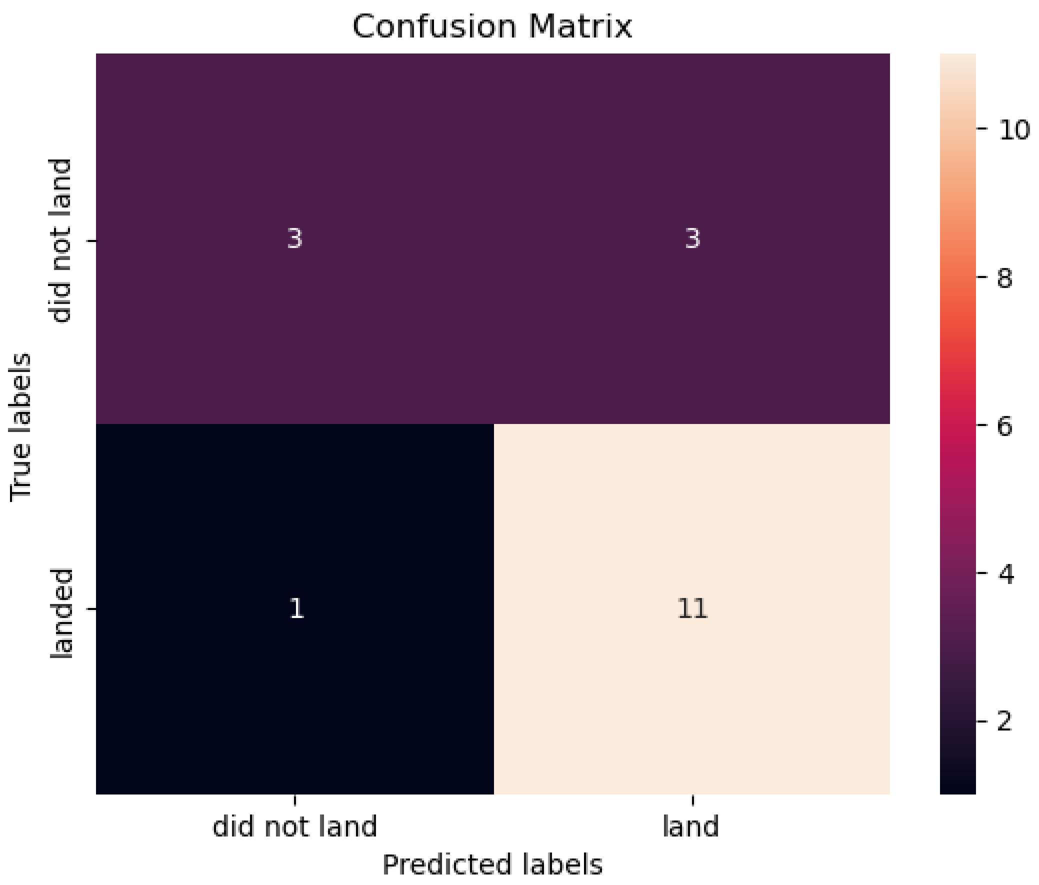
# Appendix

## Confusion matrix: SVM



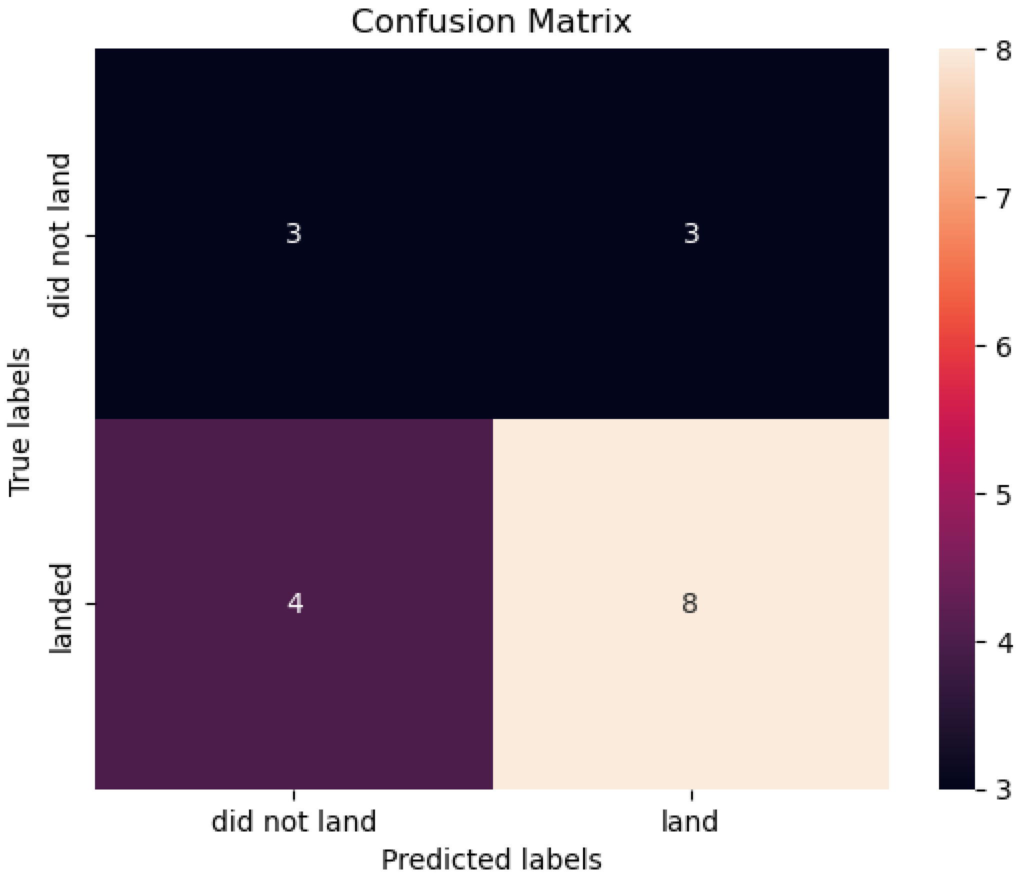
# Appendix

## Confusion matrix: Decision Tree



# Appendix

## Confusion matrix: k-NN



Thank you!