SpaceY's Affordable Space Exploration

Fabio Odisio

12/11/2023

Contents

- P3 Executive Summary
- P4 Introduction
- P5 Methodology
- P15 Results
- P39 Conclusion
- P40 Appendix

Executive Summary

The research focuses on identifying the factors for a successful Falcon 9 first stage landing using data from its competitor, SpaceX. To determined these factors several methodologies were used:

- **Collect** data by using SpaceX REST API and web scraping techniques.
- **Wrangle** data to create success/fail outcome variable.
- **Explore** data by using data visualization techniques, considering factors such as: payload, flight number, launch site.
- **Analyze** data with the use of SQL, calculating statistics such as: total payload, payload range for successful launches, and total successful and failed outcomes.
- **Explore** launch sites success rate and proximity to significant land markers by implementing FOLIUM.
- **Build Models** to predict landing outcomes using logistic regression, decision tree, support vector machine (SVM), and K-nearest neighbor (KNN).

Results

Exploratory Data Analysis:

1. KSC LC-39A had the highest success rate among landing sites.
2. Orbits ES-L1, GEO, HEO, and SSO had maximum success rate of 100%.

Predictive Analytics:

1. Decision Tree model performed slightly better than other models, despite all models performing significantly well.

Introduction

- SpaceX has achieved significant milestones, including launching spacecraft to the International Space Station, establishing Starlink—a satellite internet constellation facilitating satellite Internet access—and conducting manned missions into space. A key factor enabling SpaceX's accomplishments is the cost-effectiveness of its rocket launches. While other providers may charge over 165 million dollars for each launch, SpaceX advertises Falcon 9 rocket launches on its website at a cost of 62 million dollars. Much of this cost savings is attributed to SpaceX's ability to reuse the first stage of the rocket. Consequently, by determining the successful landing of the first stage, one can ascertain the overall cost of a Falcon 9 launch. This determination is accomplished by the use of public data and machine leaning models by SpaceX or any other competing company such as SpaceY.

Section 1

Methodology

Methodology

- Executive Summary
- Data collection methodology:
  - Collected data using SpaceX REST API and web scraping techniques.
- Performed data wrangling
  - Filtered the data, dealt with missing values, and prepared the data for analysis and modeling.
- Performed exploratory data analysis (EDA) using visualization and SQL
- Performed interactive visual analytics using Folium and Plotly Dash
- Performed predictive analysis using classification models
  - Created models to predict landing outcomes using classification models. Tuned and evaluated models to find the best model and parameters.

Data Collection

API

1. Retrieved data from the SpaceX API
2. Decoded response using .json(), and transformed it into a data frame using .json_normalize().
3. Utilized custom functions to extract launch information from the SpaceX API,
4. Created dictionary from the data.
5. Created data frame from the dictionary.
6. Applied filtering to exclude all launches except those involving the Falcon 9.
7. Imputed missing values in the payload mass with its calculated mean.
8. Exported the processed data to a CSV file.

GitHub Data Collection – API

Data Collection

Web Scraping

1. Obtained data from Wikipedia, specifically Falcon 9 launch data.
2. Generated a BeautifulSoup object from the HTML response.
3. Extracted column names from the HTML table header.
4. Gathered data by parsing HTML tables.
5. Constructed a dictionary based on the collected data.
6. Formed a data frame using the created dictionary.
7. Exported the processed data to a CSV file.

GitHub Data Collection – Web Scraping

Data Wrangling

1. Performed EDA and determined data labels.
2. Calculated: launches for each site, occurrence of each orbit, and occurrence of mission outcome per orbit type.
3. Created binary landing outcome column.
4. Exported data to a csv file.

Landing Outcome

True Ocean: Mission outcome had a successful landing to a specific region of the ocean.

False Ocean: Mission outcome had an unsuccessful landing to a specific region of the ocean.

True RTLS: Mission outcome was successful landing to a ground pad.

False RTLS: Mission outcome was unsuccessful landing to a ground pad.

True ASDS: Mission outcome was successful landing to a drone ship.

False ASDS: Mission outcome was unsuccessful landing to a drone ship.

[GitHub Data Wrangling]

EDA with Data Visualization

- Created scatter charts with the following variables to better understand their relationships:
- Flight Number & Payload
- Flight Number & Launch Site
- Payload & Launch Site
- Flight Number & Orbit Type
- Payload & Orbit Type

Analysis

1. Analyzed the scatter plots to find significant variables for machine learning models.
2. Used bar charts to view relationships among the categories and measured values.

GitHub EDA with Data Visualization

EDA with SQL

Queries:

1. Display the names of the unique launch sites in the space mission
2. Display 5 records where launch sites begin with the string 'CCA'
3. Display the total payload mass carried by boosters launched by NASA (CRS)
4. Display average payload mass carried by booster version F9 v1.1
5. List the date when the first succesful landing outcome in ground pad was acheived.
6. List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
7. List the total number of successful and failure mission outcomes
8. List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
9. 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.
10. 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 EDA Using SQL]()

Interactive Map with Folium

Launch Site Markers

Added a blue circle at **NASA Johnson Space Center's** coordinate with a popup label showing its name.

Added red circles at all **launch sites** coordinates using a popup label showing its name.

Launch Outcome Markers

- Added markers of **successful** (green) and **unsuccessful** (red) launches at each launch site.

Launch Site Distance to Significant Landmarks

- Added lines to show the distance between launch site **CCAFS SLC-40** and significant landmarks such as: **nearest coastline, railway, highway, and city**.

GitHub Map using Folium

Build a Dashboard with Plotly Dash

Dropdown List with Launch Sites

Allows user to select all launch sites or specific ones.

Pie Chart of Successful Launches

Allows user to view successful and unsuccessful launches as a percent of the total.

Slider of Payload Mass Range

Allows user to select payload range mass.

Scatter Chart of Payload Mass & Success Rate by Booster Version

Allows user to see the relationship between Payload and Launch Success.

GitHub Dashboard With Plotly Dash

Predictive Analysis (Classification)

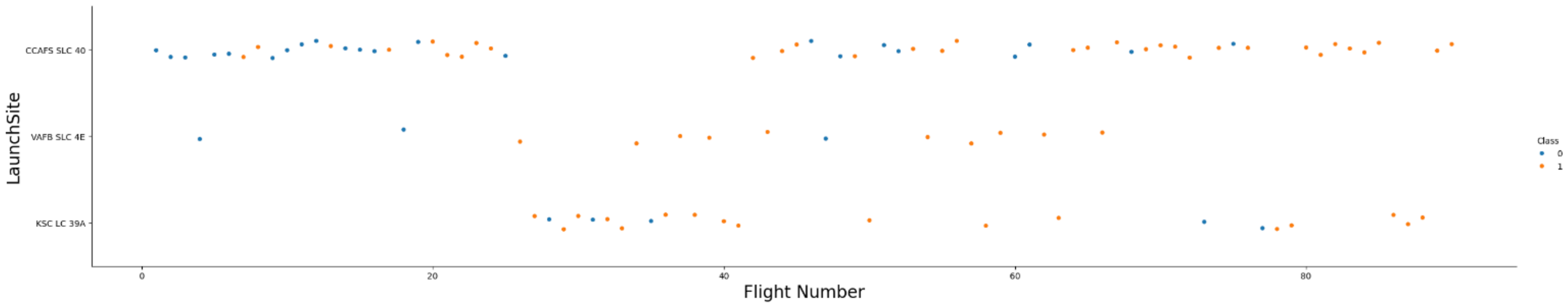
1. Generated a Numpy array from the "class" column
2. Standardized the data using StandardScaler by fitting and transforming the dataset.
3. Split the data using train_test_split.
4. Established a GridSearchCV object with cv=100 for parameter optimization.
5. Applied GridSearchCV on various algorithms, including logistic regression, SVM, decision tree, and KNN.
6. Evaluated the accuracy on the test data using .score() for all models.
7. Examined the confusion matrix for each model.
8. Determined the best model based on Jaccard_score, f1_score, and Accuracy.

[GitHub Predictive Analysis]

Results

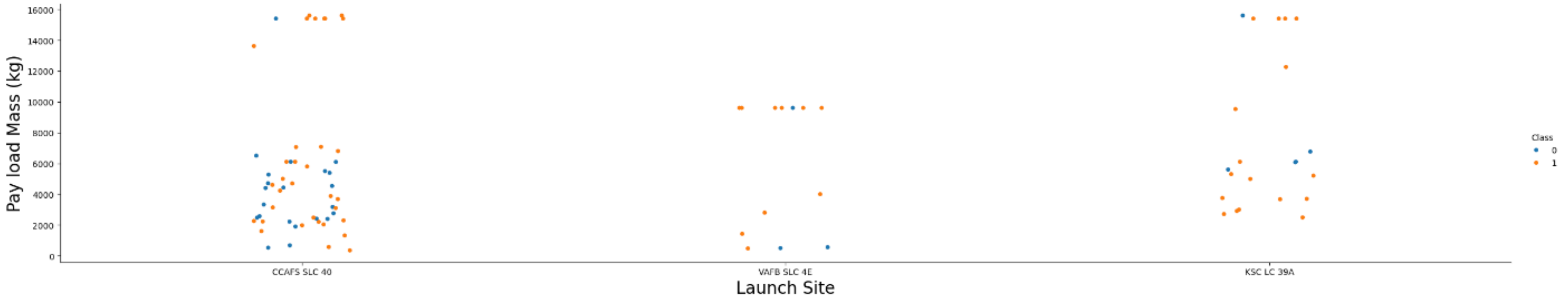
Flight Number vs. Launch Site

- Earlier flights had a lower success rate (blue = fail).
- Later flights had a higher success rate (orange = success).
- Roughly half of launches were from **CCASFS SLC 40** Launch site.
- **VAFB SLC 4E** and **KSC LC 39A** had higher success rates.

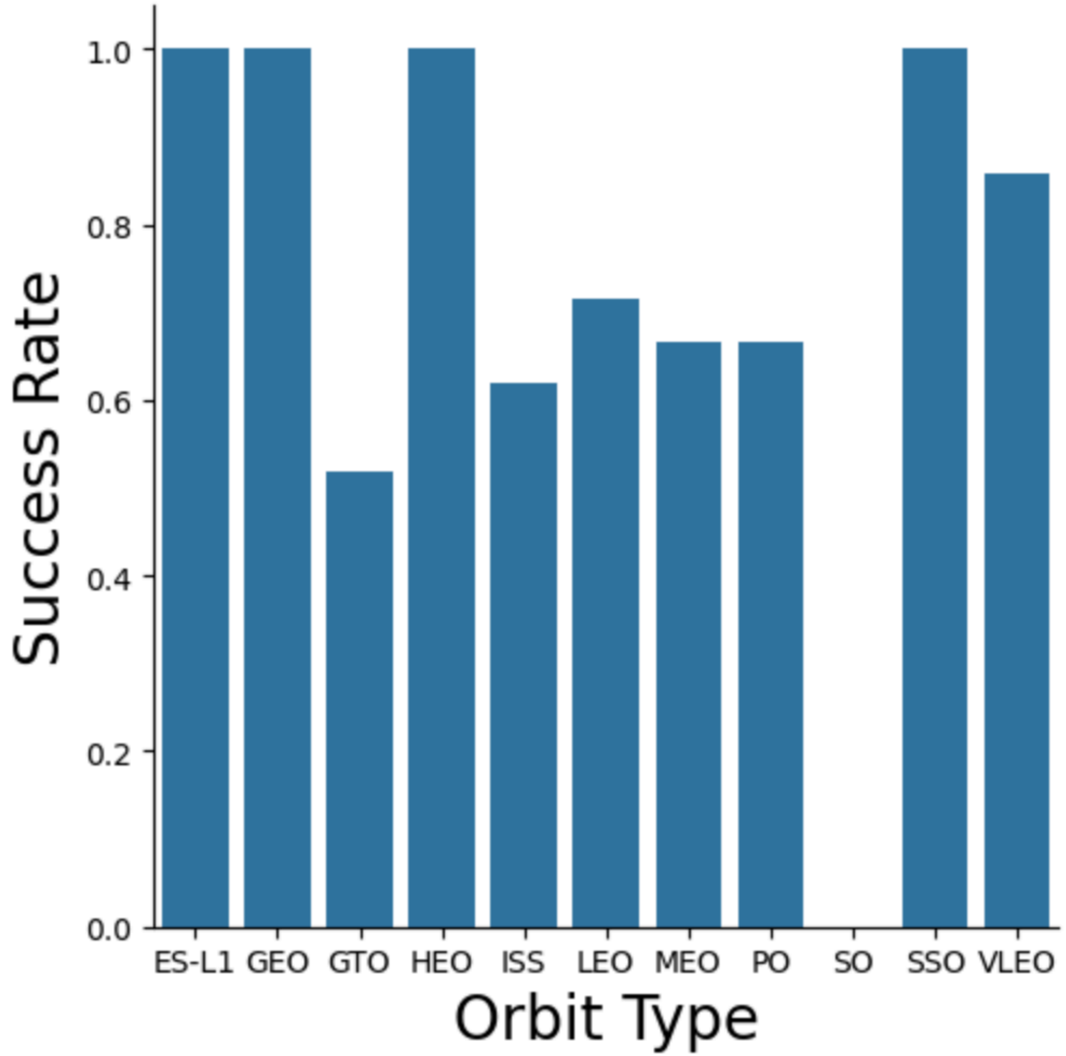


Payload vs. Launch Site

- The higher the payload mass, the higher the success rate.
- Most launches with a payload of 8,000 kg or more were successful.
- **KSC LC 39A** has a 100% success rate for launches less than 5,000 kg.
- **VAFB SKC 4E** has not launched anything heavier than 10,000 kg.

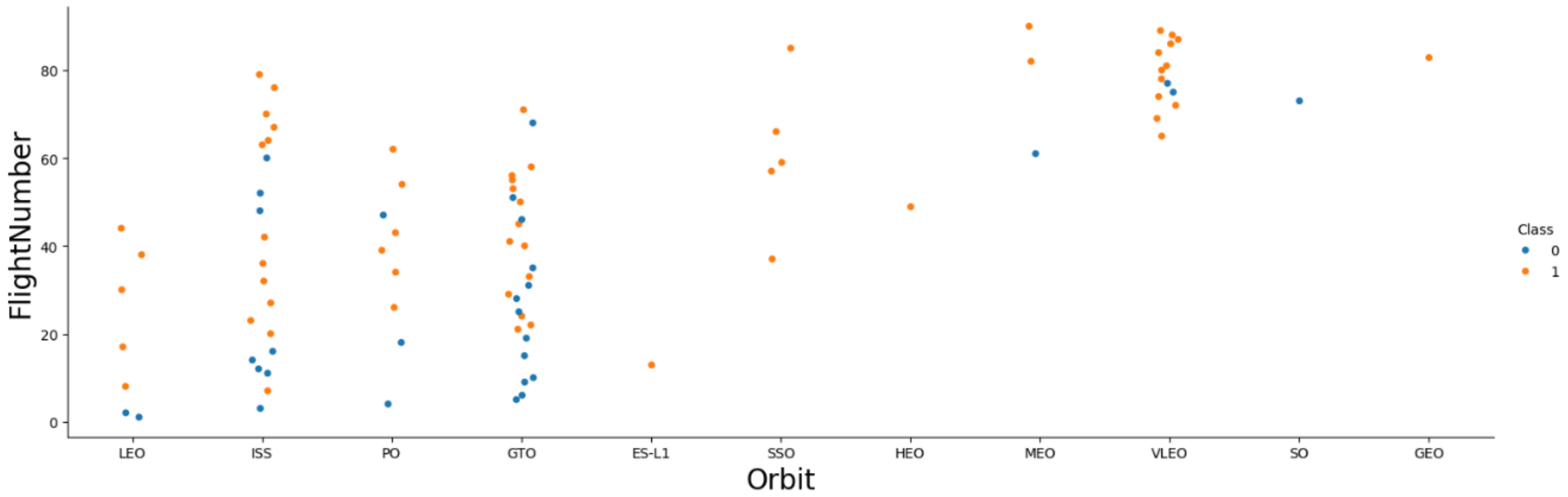


Success Rate vs. Orbit Type



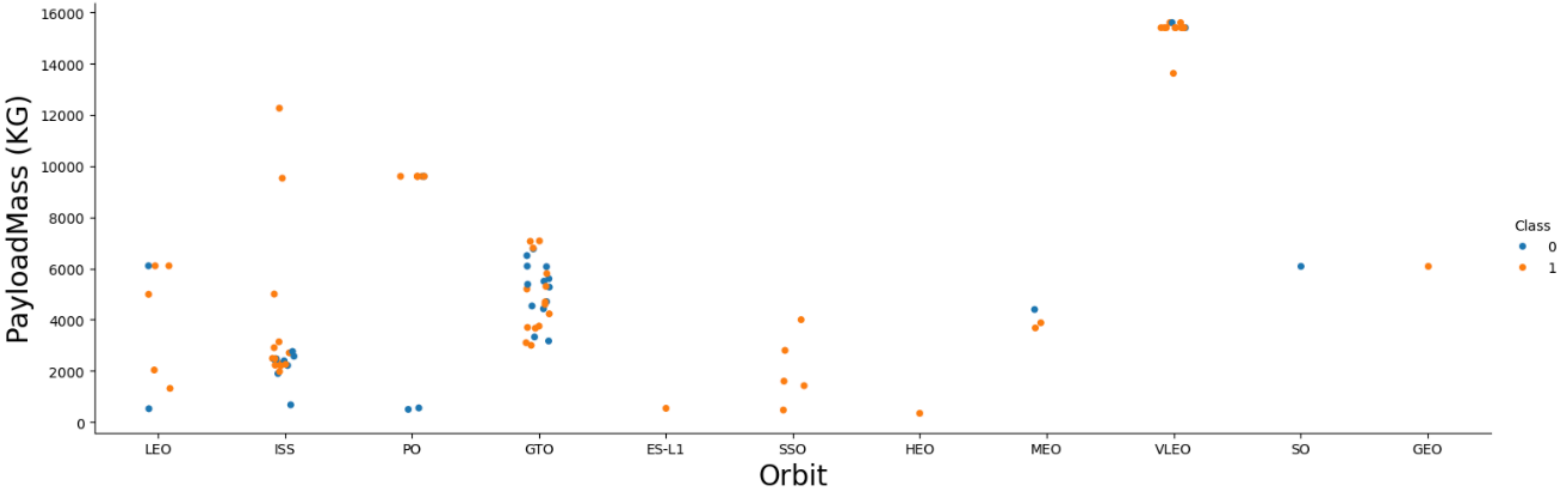
Flight Number vs. Orbit Type

- The success rate typically increases with the number of flights for each orbit.
- This relationship is highly apparent for the LEO orbit.
- The GTO orbit, however, does not follow the trend.



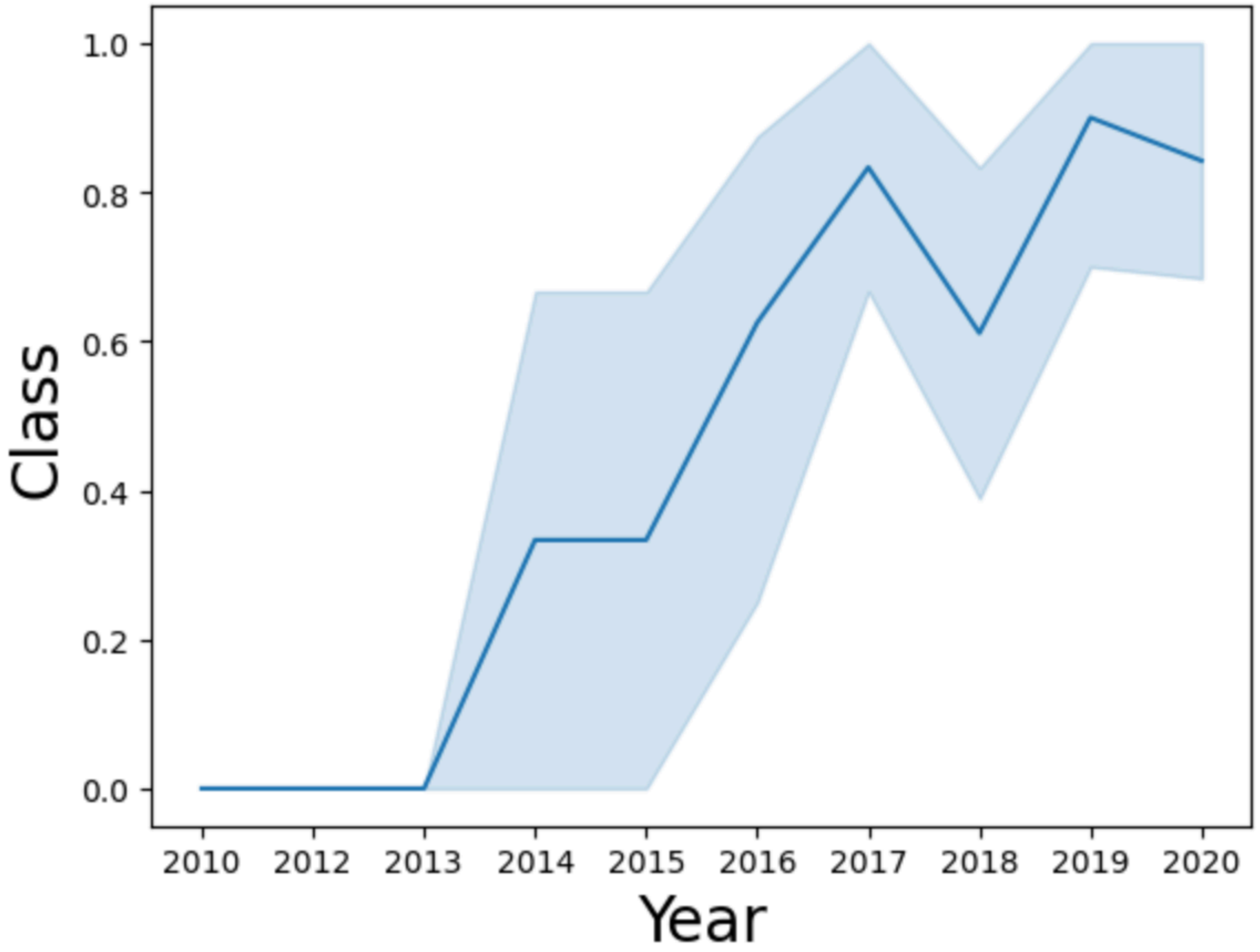
Payload vs. Orbit Type

- LEO, ISS, and PO orbits deal better with heavy payloads.
- The GTO orbit has overall mixed success when analyzing by payload.



Launch Success Yearly Trend

- From 2017-2018 and from 2019-2020 the success rate declined.
- From 2013-2017 and 2018-2019 the success rate increased.
- 2019 had the highest success rate from years 2010 to 2020.



Launch Site Informations

Launch Site Names:

CCAFS LC-40

CCAFS SLC-40

KSC LC-39A

VAFB SLC-4E

```
%sql select distinct(LAUNCH_SITE) from SPACEXTBL
```

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

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

Records with Launch Site Starting with CCA

```
%sql select * from SPACEXTBL where LAUNCH_SITE like "CCA%" limit 5
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO
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)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)

Payload Mass

Total Payload Mass: 45,596 kg

Average Payload Mass: 2,928 kg

sum(PAYLOAD_MASS__KG_)
45596

avg(PAYLOAD_MASS__KG_)
2928.4

Landing & Mission

1st successful landing in ground pad: 12/22/2015

min(DATE)
2015-12-22

Total Number of Successful and Failed Mission Outcomes

count(MISSION_OUTCOME)
98

Booster Drone Ship Landing

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

Boosters

Carrying Max Payload

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

Failed Landings on Drone Ship for 2015

Showing month, date, booster version, launch site and landing outcome

```
%sql SELECT substr(Date,6,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, [Landin
FROM SPACEXTBL \
where [Landing_Outcome] = 'Failure (drone ship)' and substr(Date,0,5)='2015';
```

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

month	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

Count of Succesful Landings

Count of landing outcomes between 06/04/2010 and 03/20/2017 in descending order.

```
%sql select LANDING_OUTCOME , count(*) as COUNT_OUTCOMES from SPACEXTBL \
where DATE between '2010-06-04' and '2017-03-20' group by (LANDING_OUTCOME) order
```

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

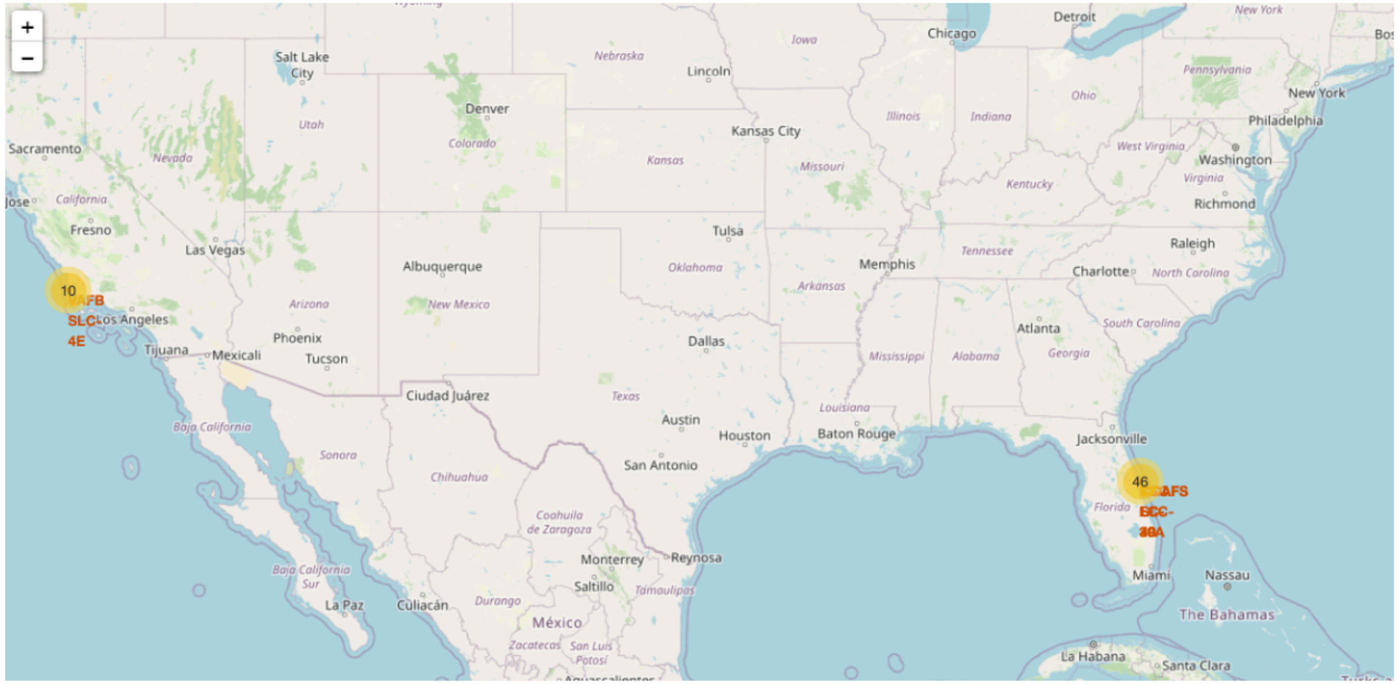
Landing_Outcome	COUNT_OUTCOMES
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

Launch Sites with Markers

- We can observe that the launch sites are near the equator, this makes the launch easier due to earth's rotation.

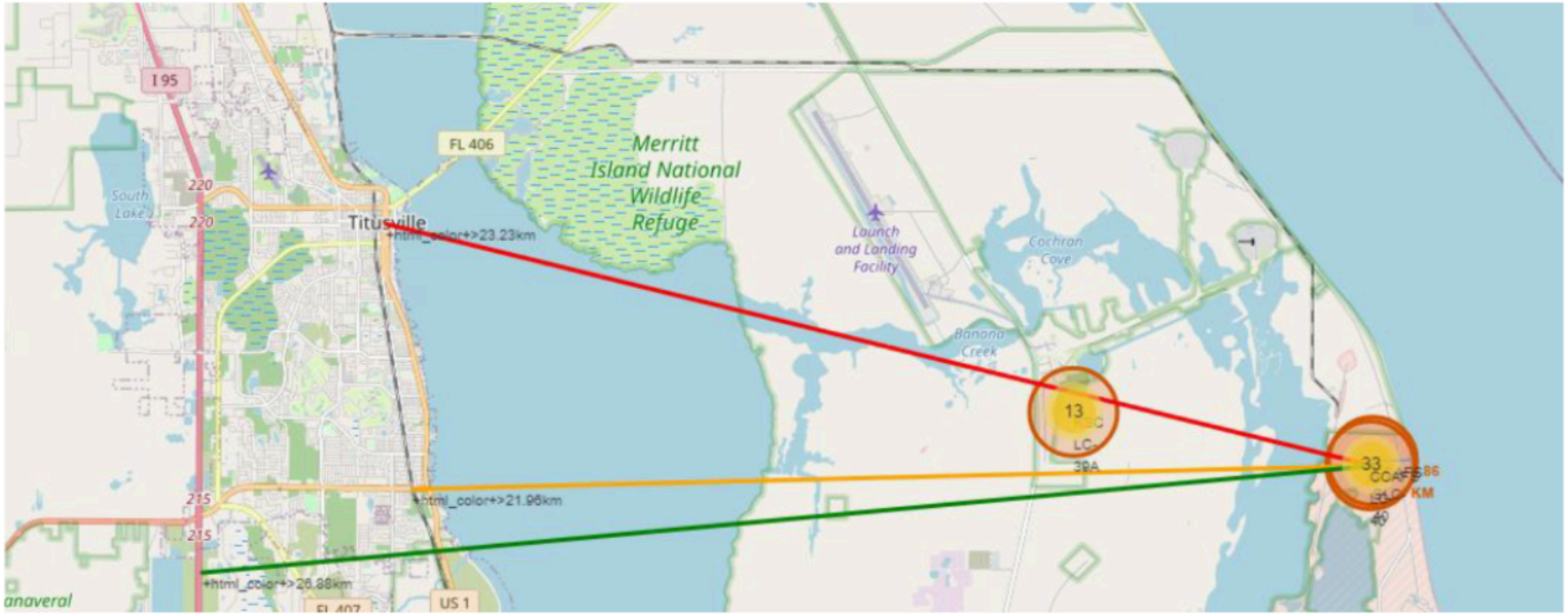


Launch Outcomes at Each Launch Site

- Green markers for successful launches.
- Red markers for unsuccessful launches.

CCAFS SLC-40 Distance to Landmarks

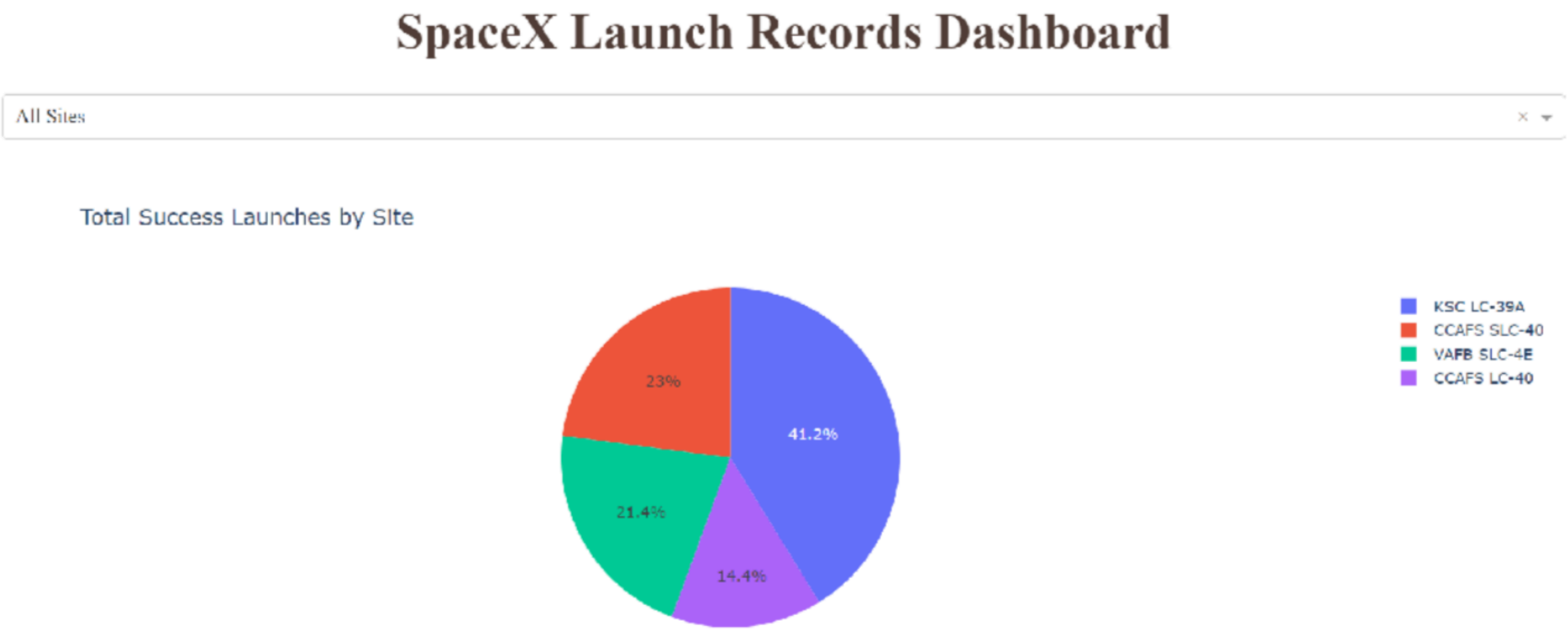
- .86 km from nearest coastline.
- 21.96 km from nearest railway.
- 23.23 km from nearest city.
- 26.88 km from nearest highway.



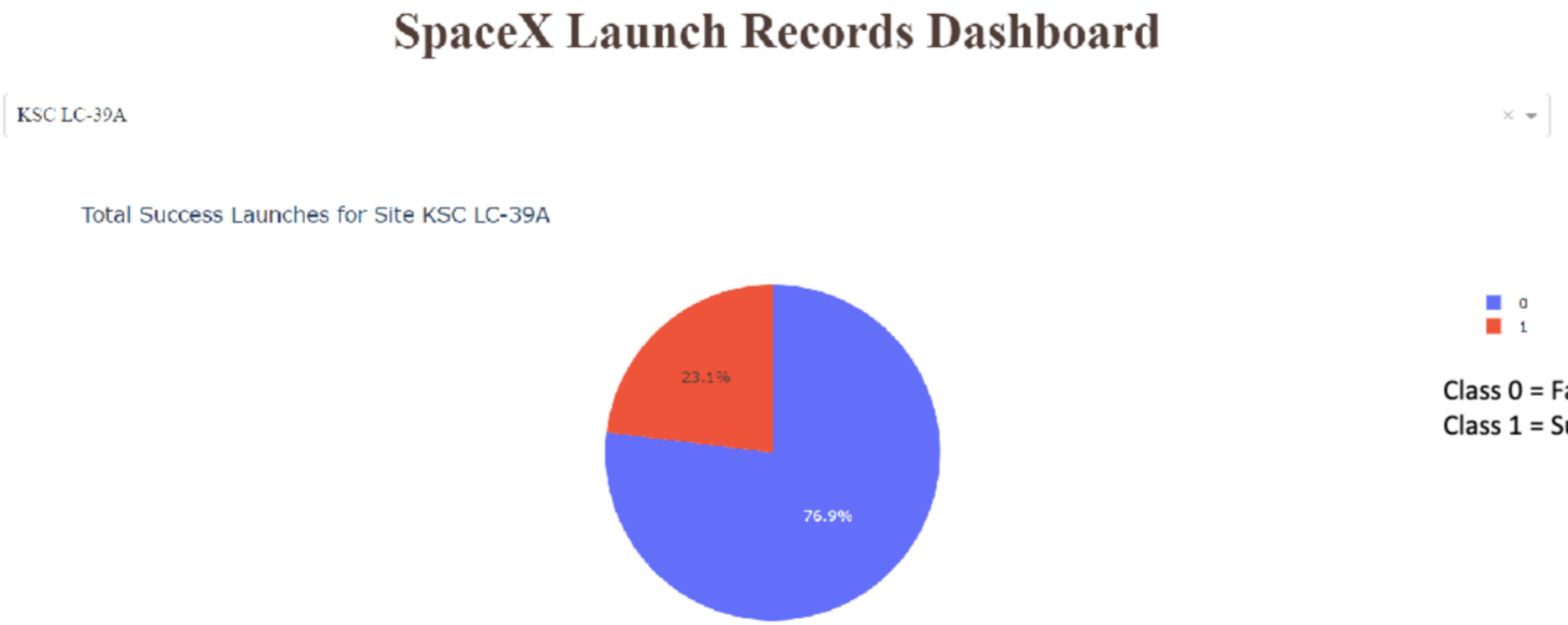
Section 4

Build a Dashboard with Plotly Dash

Launch Success by Site (success as percent of total)



KSC LC-29A Launch Success



Class 0 = Fail
Class 1 = Success

Payload Mass and Success

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- All the models performed at about the same level and had same scores and accuracy. The decision tree model slightly outperformed the other models.

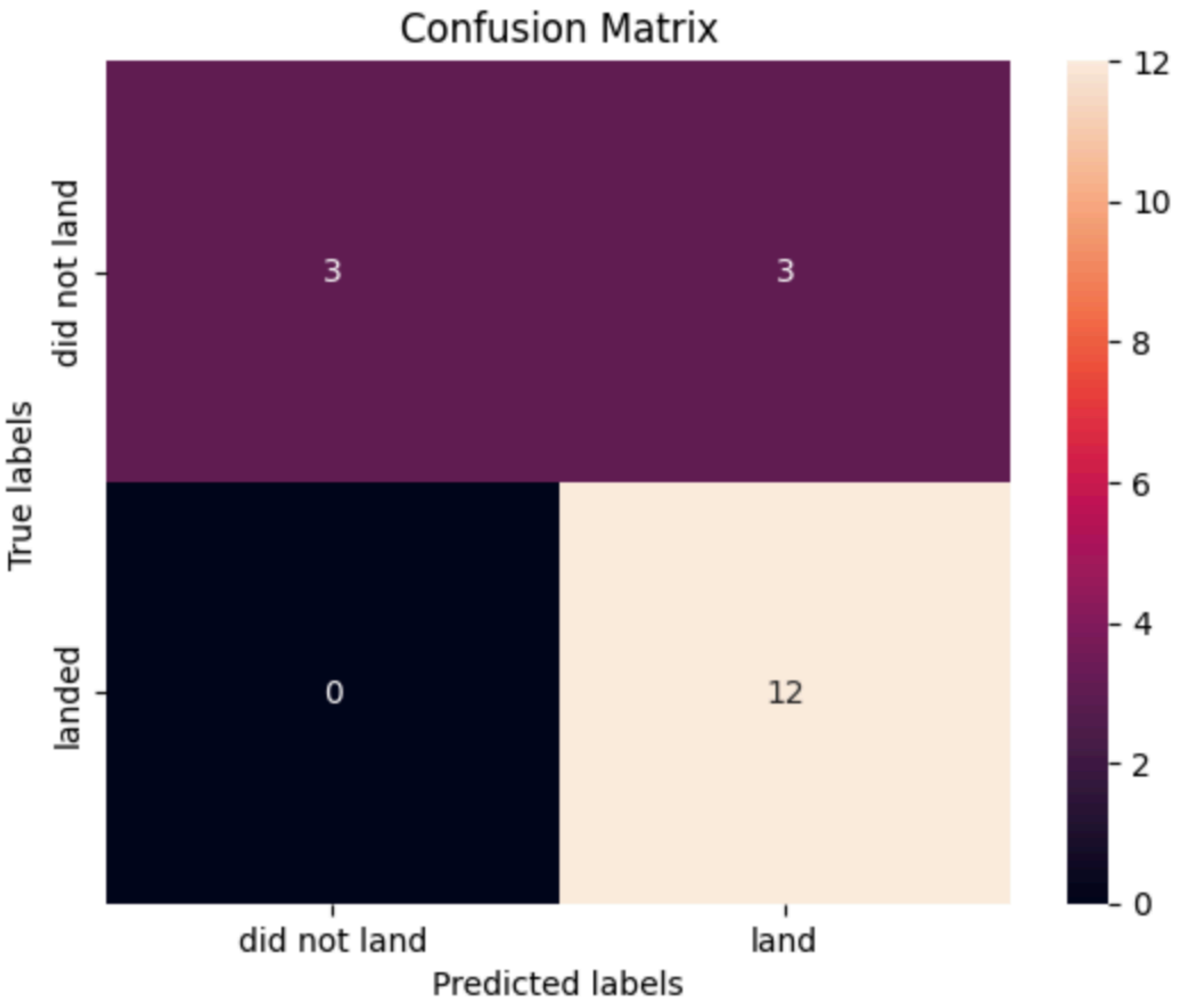
In [49]:
```
from sklearn.metrics import jaccard_score, f1_score

# Examining the scores from Test sets
jaccard_scores = [
                  jaccard_score(Y_test, logreg_yhat, average='binary'),
                  jaccard_score(Y_test, svm_yhat, average='binary'),
                  jaccard_score(Y_test, tree_yhat, average='binary'),
                  jaccard_score(Y_test, knn_yhat, average='binary'),
                 ]

f1_scores = [
             f1_score(Y_test, logreg_yhat, average='binary'),
             f1_score(Y_test, svm_yhat, average='binary'),
             f1_score(Y_test, tree_yhat, average='binary'),
             f1_score(Y_test, knn_yhat, average='binary'),
            ]

accuracy = [logreg_score, svm_cv_score, tree_cv_score, knn_cv_score]

scores_test = pd.DataFrame(np.array([jaccard_scores, f1_scores, accuracy]), index
scores_test
```

Out[49]:

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

Confusion Matrix

- All the confusion matrices were identical.
- We can observe there are 3 false positives, which is not a good sign.
- Precision = .80
- Recall = 1
- F1 Score = .89
- Accuracy = .833



Conclusions

The Models performed similarly on the tests, with decision tree performing a little better than the other models.

Most of the Launches were near the equator for a better success rate of launches.

All the launch sites are on the coast.

Launch success increases over time.

The highest success rate among launch sites was seen in KSC LC-39A, which has a success rate of 100% for launches under 5,500 kg.

Orbits ES-L1, GEO, HEO, and SSO have 100% success rates.

The higher the Payload mass, the higher the success rate.

Thank you!