Predictive Pricing and First Stage Reusability for Space Y Rocket Company

Presentation Contents

- Executive Summary
- Introduction
- Methodology
- Results
- EDA with Visualization
- EDA with SQL
- Interactive Maps with Folium
- Plotly Dash Dashboard
- Predictive Analytics
- Conclusion

Executive Summary

In today's era, the commercial space industry is undergoing a transformation, making space travel more accessible and affordable for everyone. Companies like Virgin Galactic, Rocket Lab, Blue Origin, and, notably, SpaceX are pioneering innovations in space exploration and satellite deployment. SpaceX, in particular, has achieved remarkable success by sending spacecraft to the International Space Station, establishing the Starlink satellite internet constellation, and conducting manned missions to space.

Project Objectives:

Determine the pricing strategy for Space Y's rocket launches by analyzing SpaceX's pricing models.

Develop data visualization dashboards to provide insights and comparisons to the Space Y team.

Utilize machine learning techniques to predict the likelihood of SpaceX reusing the first stage.

Enable Space Y to make informed business decisions, enhance competition in the commercial space industry, and contribute to the democratization of space travel.

Introduction

One of SpaceX's key advantages is its cost-efficiency in rocket launches. SpaceX offers Falcon 9 rocket launches at a significantly lower cost than other providers, primarily because they can reuse the first stage of their rockets. This reusability factor plays a crucial role in determining the overall cost of a launch. This project aims to tackle two essential challenges:

Pricing Analysis: The project involves determining the price of each rocket launch, a critical factor for Space Y, a new rocket company aiming to compete with SpaceX.

First Stage Reusability Prediction: Instead of relying solely on traditional rocket science methods to predict the successful landing and reusability of the first stage, this project leverages machine learning and publicly available data to forecast whether SpaceX will reuse the first stage.

Space Y research strategy include gathering information about SpaceX, creating informative dashboards for your team, and developing a machine learning model to predict first stage reusability. This project enables Space Y to make informed decisions, compete effectively with SpaceX, and contribute to the ongoing transformation of the commercial space industry.

Methodology

1. Data Collection:

Gathered comprehensive data from various sources, including publicly available SpaceX records, historical launch data, and Falcon 9 performance reports.

Ensured data accuracy and completeness by conducting rigorous data validation and cleansing.

2. Exploratory Data Analysis (EDA):

Performed EDA to gain deep insights into SpaceX's launch history, pricing trends, and first-stage reusability patterns. Identified key features that might influence pricing and first stage reusability.

3. Data Preprocessing:

Prepared the data for machine learning by encoding categorical variables, handling missing values, and scaling numerical features. Ensured the dataset was ready for training and evaluation.

4. Machine Learning Model Selection:

Explored multiple machine learning algorithms, including logistic regression, decision trees, and random forests. Selected the most suitable model based on performance metrics and prediction accuracy.

Methodology

5. Model Training:
Trained the chosen machine learning model on a carefully curated dataset. Utilized advanced techniques to fine-tune the model for optimal performance.
6. Model Evaluation:
Evaluated the model's performance using robust metrics such as accuracy, precision, recall, and F1-score. Employed cross-validation to ensure model generalization.
7. Predictive Insights:
Leveraged the trained model to predict SpaceX's first-stage reusability for upcoming launches. Generated actionable insights to assist Space Y in pricing strategies and launch planning.
8. Dashboard Creation:
Developed informative and interactive dashboards to visualize pricing data, reusability predictions, and comparative analyses. Empowered Space Y's decision-makers with real-time insights.
9. Conclusion:
Summarized the methodology's key steps and highlighted its role in Space Y's competitive strategy. Emphasized the project's potential to contribute to the commercial space industry's growth and innovation.

Results

Exploratory Data Analysis

- Launch success has improved over time
- KSC LC-39A has the highest success rate among landing sites
- Orbits ES-L1, GEO, HEO and SSO have a 100% success rate

Visual Analytics

- Most launch sites are near the equator, and all are close to the coast
- Launch sites are far enough away from anything a failed launch can damage (city, highway, railway), while still close enough to bring people and material to support launch activities

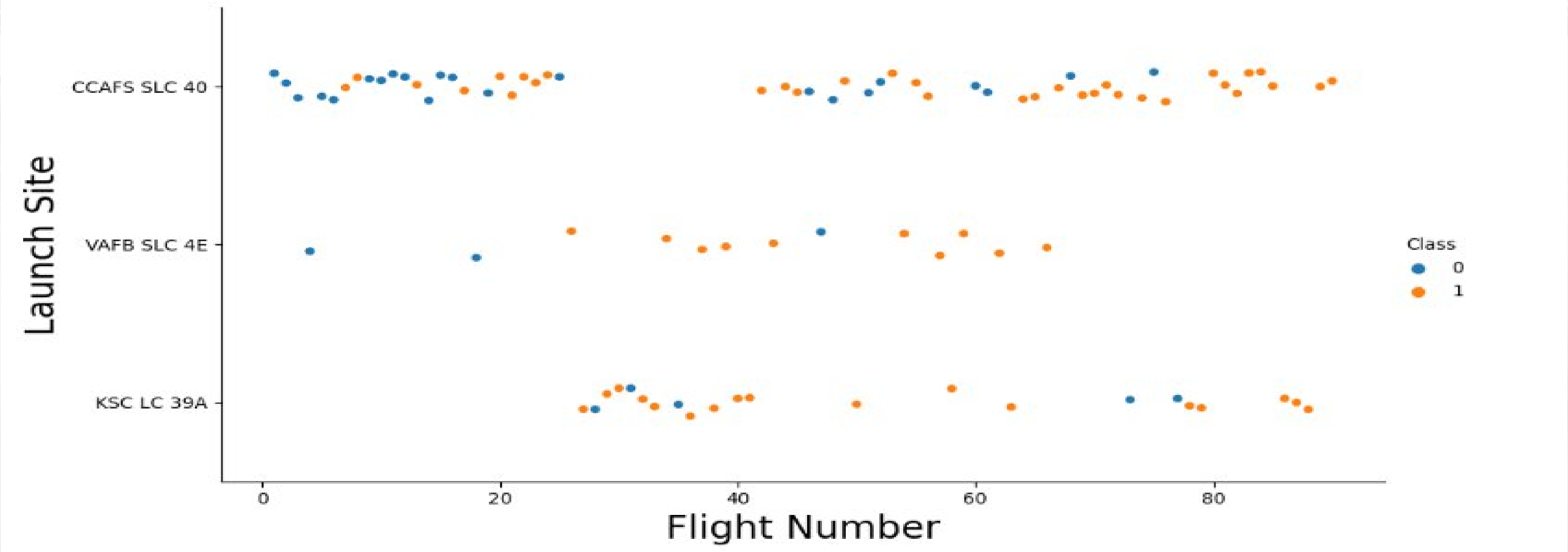
Predictive Analytics

- Decision Tree model is the best predictive model for the dataset

Flight Number vs. Launch Site

Exploratory Data Analysis

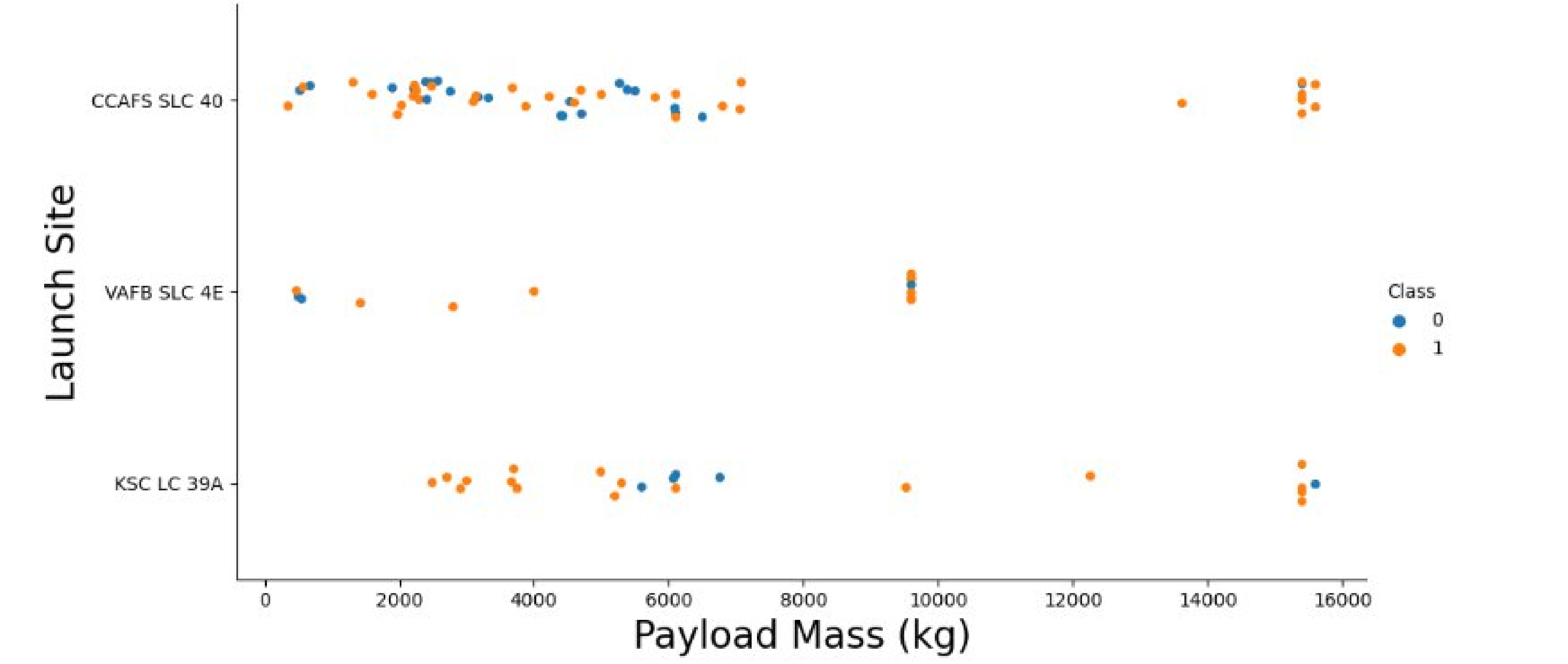
- Earlier flights had a lower success rate (blue = fail)
- Later flights had a higher success rate (orange = success)
- Around half of launches were from CCAFS SLC 40 launch site
- VAFB SLC 4E and KSC LC 39A have higher success rates
- We can infer that new launches have a higher success rate



Payload vs. Launch Site

Exploratory Data Analysis

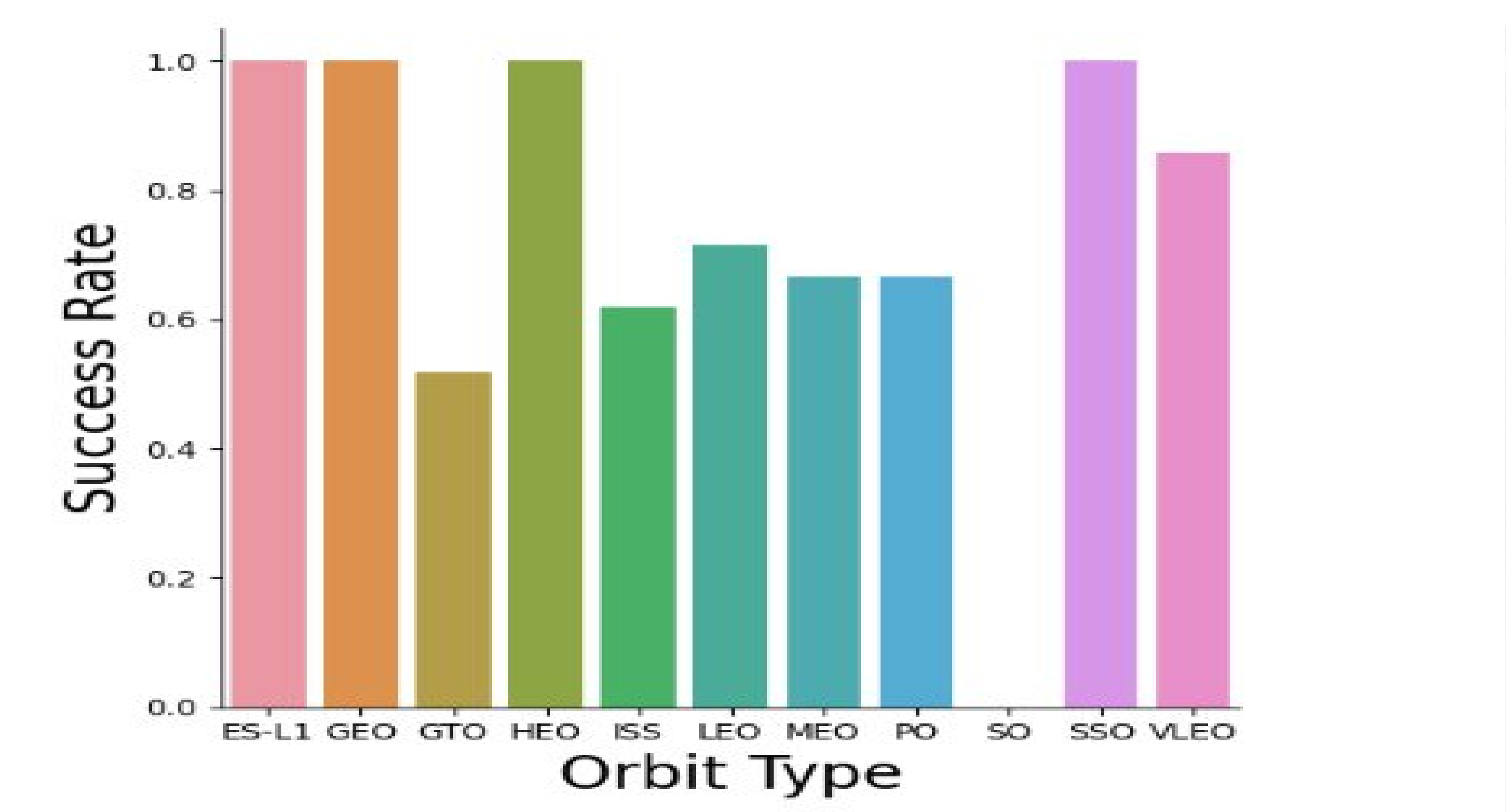
- Typically, the higher the payload mass (kg), the higher the success rate
- Most launces with a payload greater than 7,000 kg were successful
- KSC LC 39A has a 100% success rate for launches less than 5,500 kg
- VAFB SKC 4E has not launched anything greater than ~10,000 kg



Success Rate by Orbit

Exploratory Data Analysis

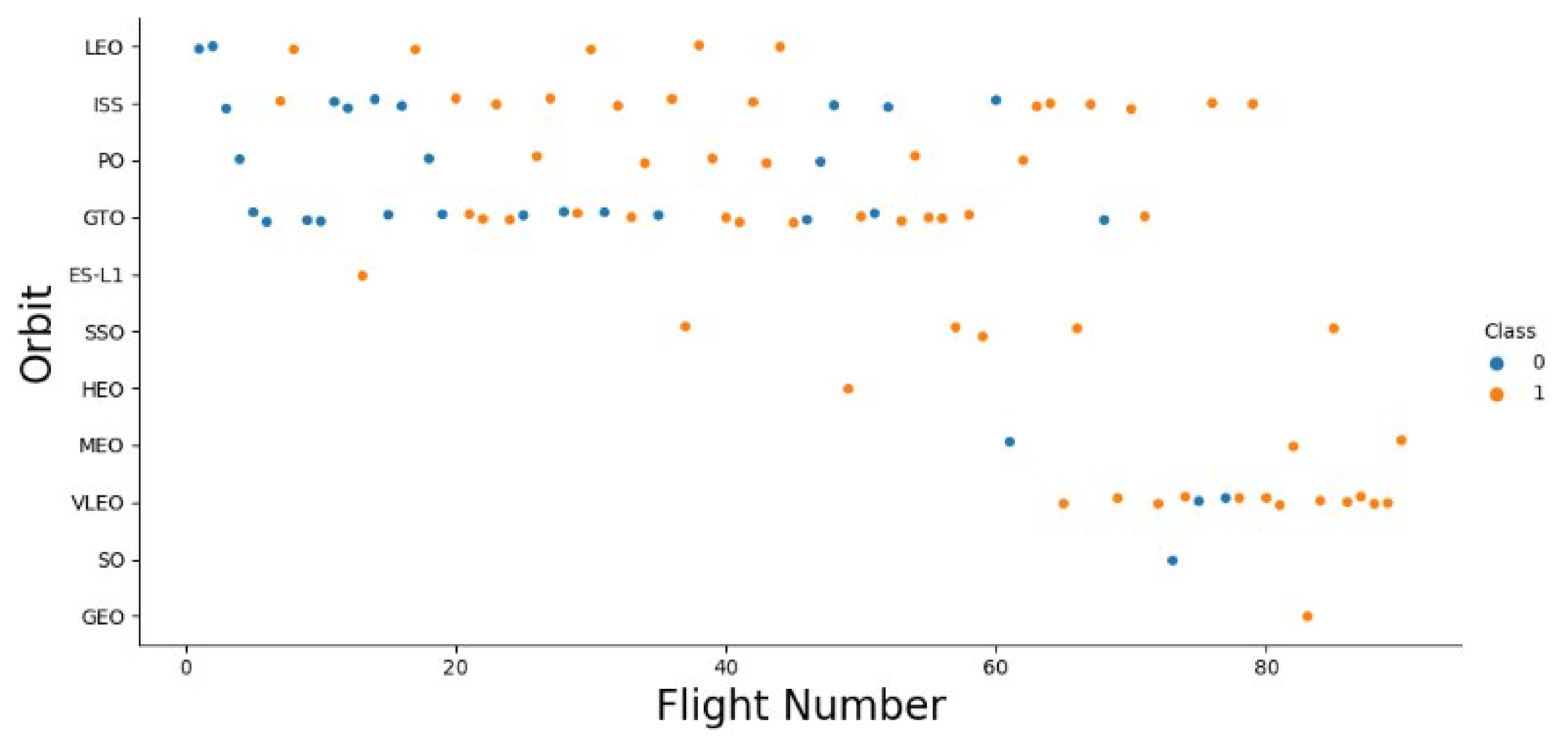
- 100% Success Rate: ES-L1, GEO, HEO and SSO
- 50%-80% Success Rate: GTO, ISS, LEO, MEO, PO
- 0% Success Rate: SO



Flight Number vs Orbit

Exploratory Data Analysis

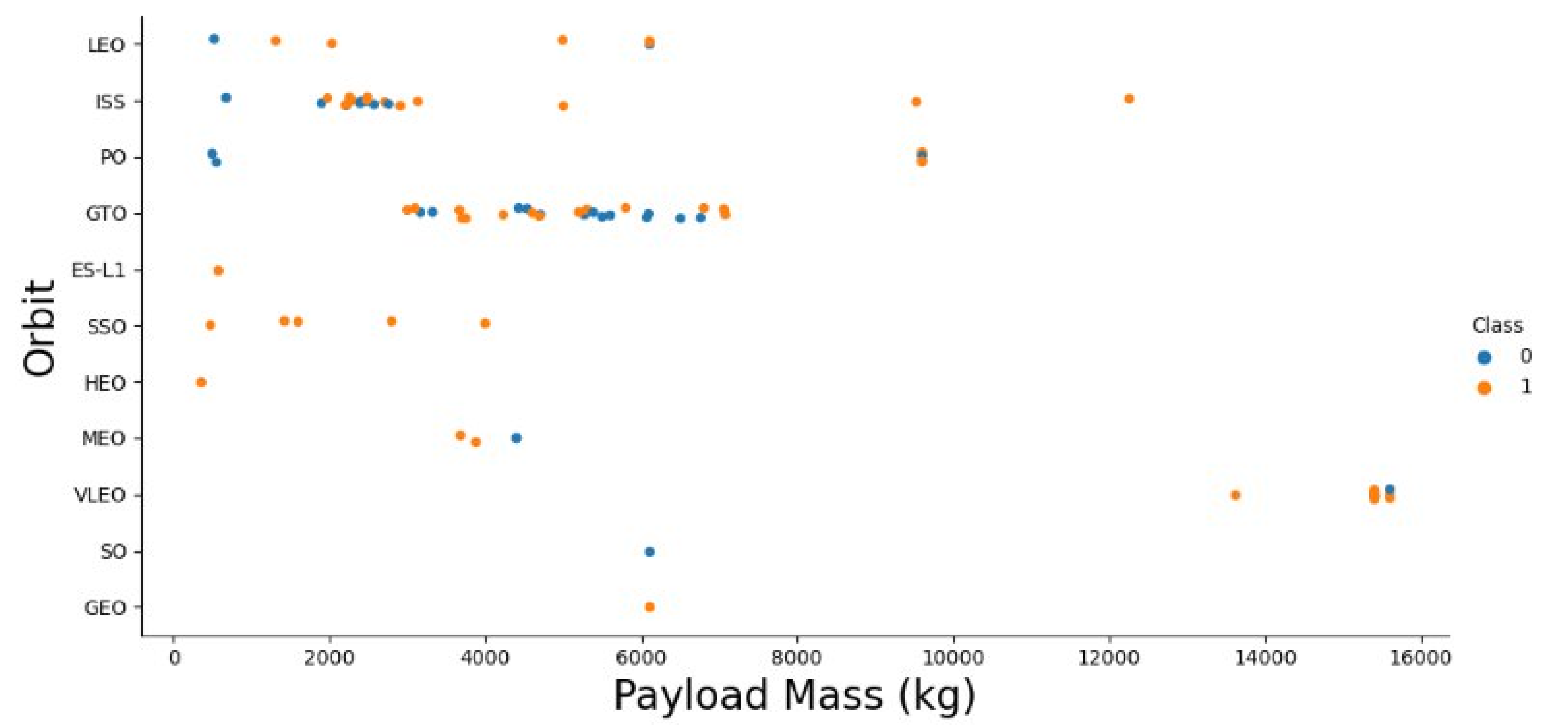
- The success rate typically increases with the number of flights for each orbit
- This relationship is highly apparent for the LEO orbit
- The GTO orbit, however, does not follow this trend



Payload vs Orbit

Exploratory Data Analysis

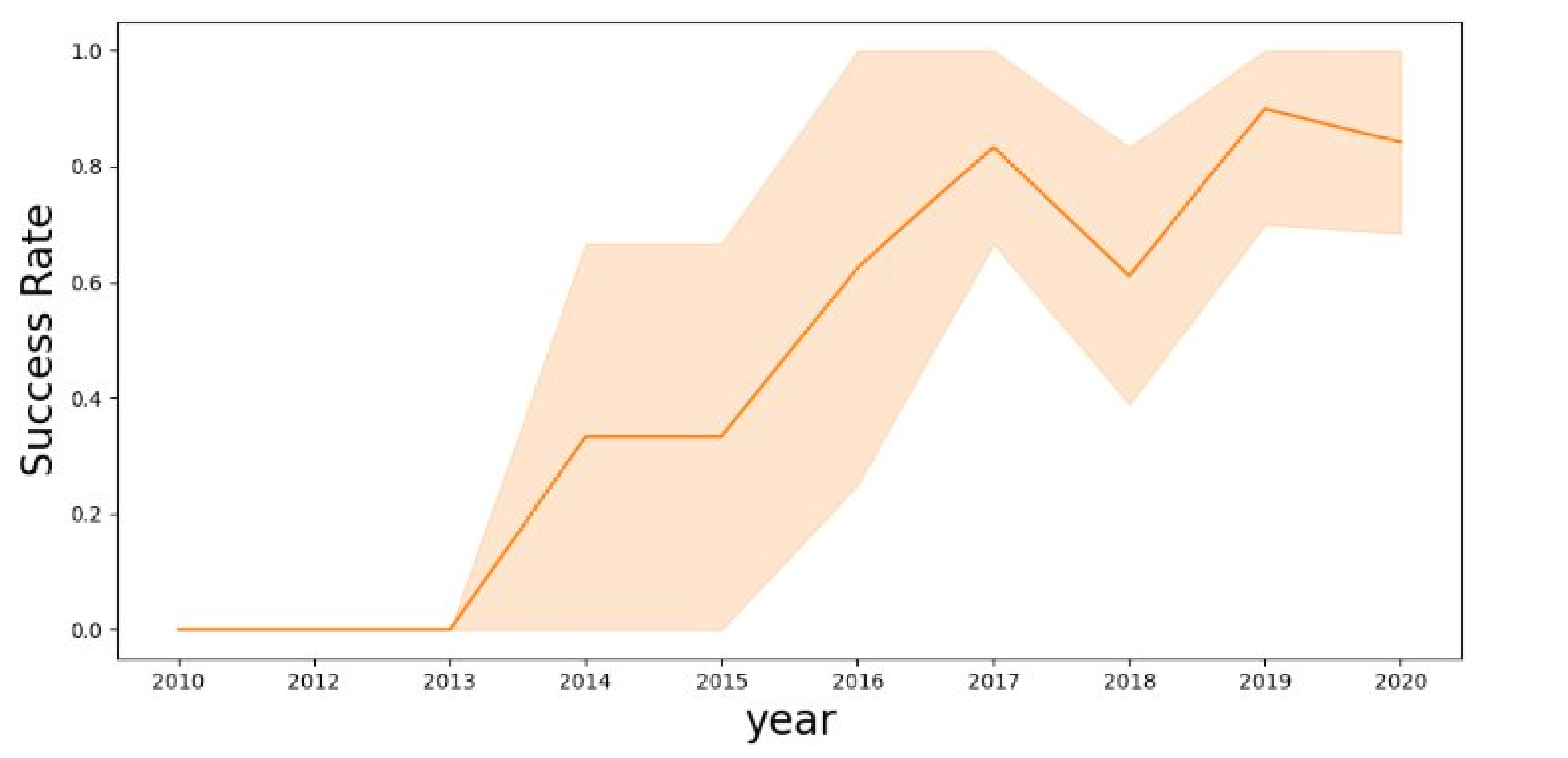
- Heavy payloads are better with LEO, ISS and PO orbits
- The GTO orbit has mixed success with heavier payloads



Launch Success over Time

Exploratory Data Analysis

- The success rate improved from 2013-2017 and 2018-2019
- The success rate decreased from 2017-2018 and from 2019-2020
- Overall, the success rate has improved since 2013



EDA with SQL

1st Successful Landing in Ground PadBooster
Drone Ship Landing:
12/22/2015

Booster mass greater than 4,000 but less than 6,000
- JSCAT-14,
- JSCAT-16,
- SES-10,
- SES-11 / EchoStar 105

Boosters
Carrying Max 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

EDA with SQL

Total Number of Successful and Failed Mission Outcomes

- 1 Failure in Flight
- 99 Success
- 1 Success (payload status unclear)

Ranked Descending

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

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

Landing _Outcome	count_outcomes
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

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

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

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

Launch Sites

• Near Equator: the closer the launch site to the equator, the easier it is tolaunch to equatorial orbit, and the more help you get from Earth's rotation for a prograde orbit. Rockets launched from sites near the equator get an additional natural boost - due to the rotational speed of earth - that helps save the cost of putting in extra fuel and boosters.

Launch Outcomes

At Each Launch Site

- Outcomes:
- Green markers for successful launches
- Red markers for unsuccessful launches
- Launch site CCAFS SLC-40 has a 3/7 success rate (42.9%)

Distance to Proximities

CCAFS SLC-40

- 0.86 km from nearest coastline
- 21.96 km from nearest railway
- 23.23 km from nearest city
- 26.88 km from nearest highway

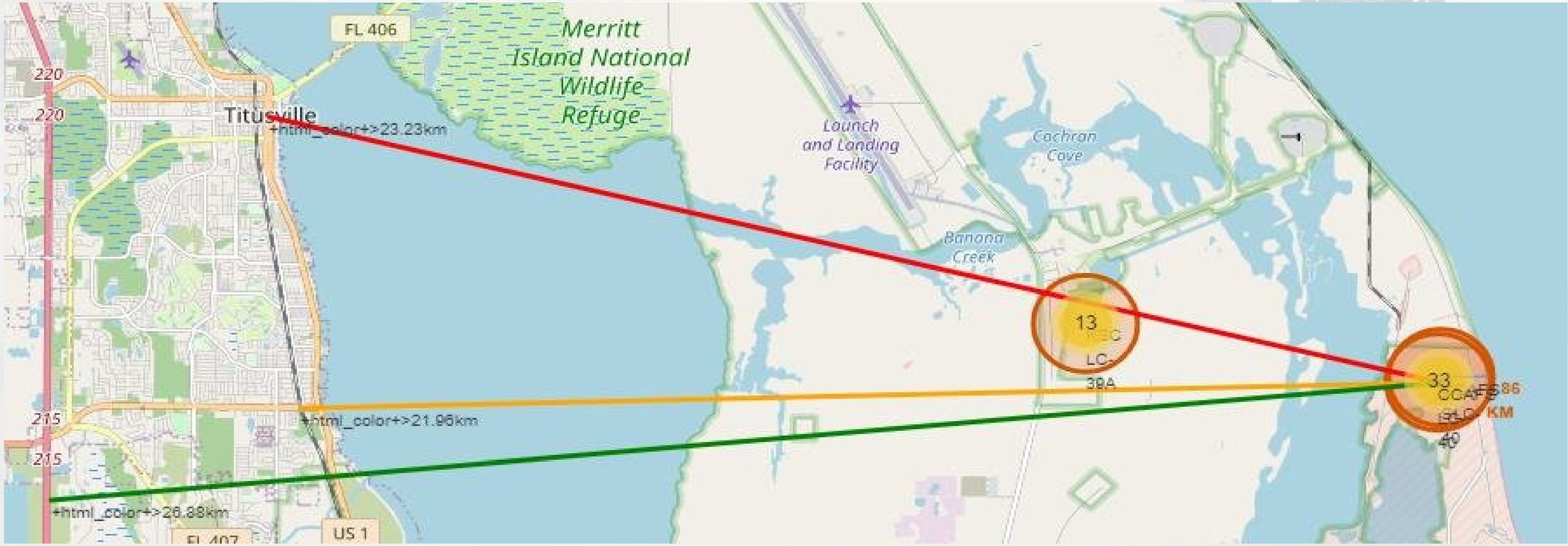
Distance to Proximities

CCAFS SLC-40

• Coasts: help ensure that spent stages dropped along the launch path or failed launches don't fall on people or property.

• Safety / Security: needs to be an exclusion zone around the launch site to keep unauthorized people away and keep people safe.

• Transportation/Infrastructure and Cities: need to be away from anything a failed launch can damage, but still close enough to roads/rails/docks to be able to bring people and material to or from it in support of launch activities.

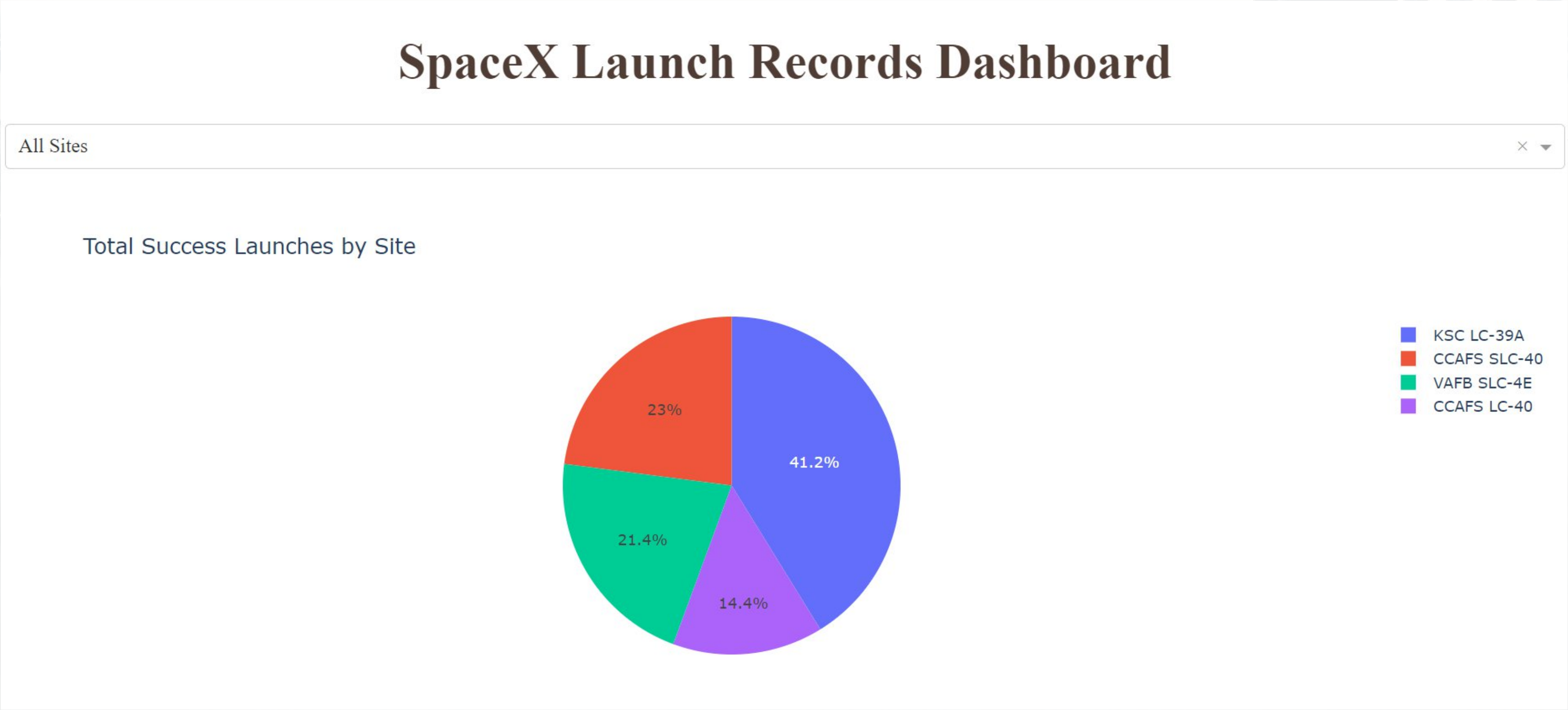


Dashboard

Launch Success by Site

Success as Percent of Total

• KSC LC-39A has the most successful launches amongst launch sites (41.2%)

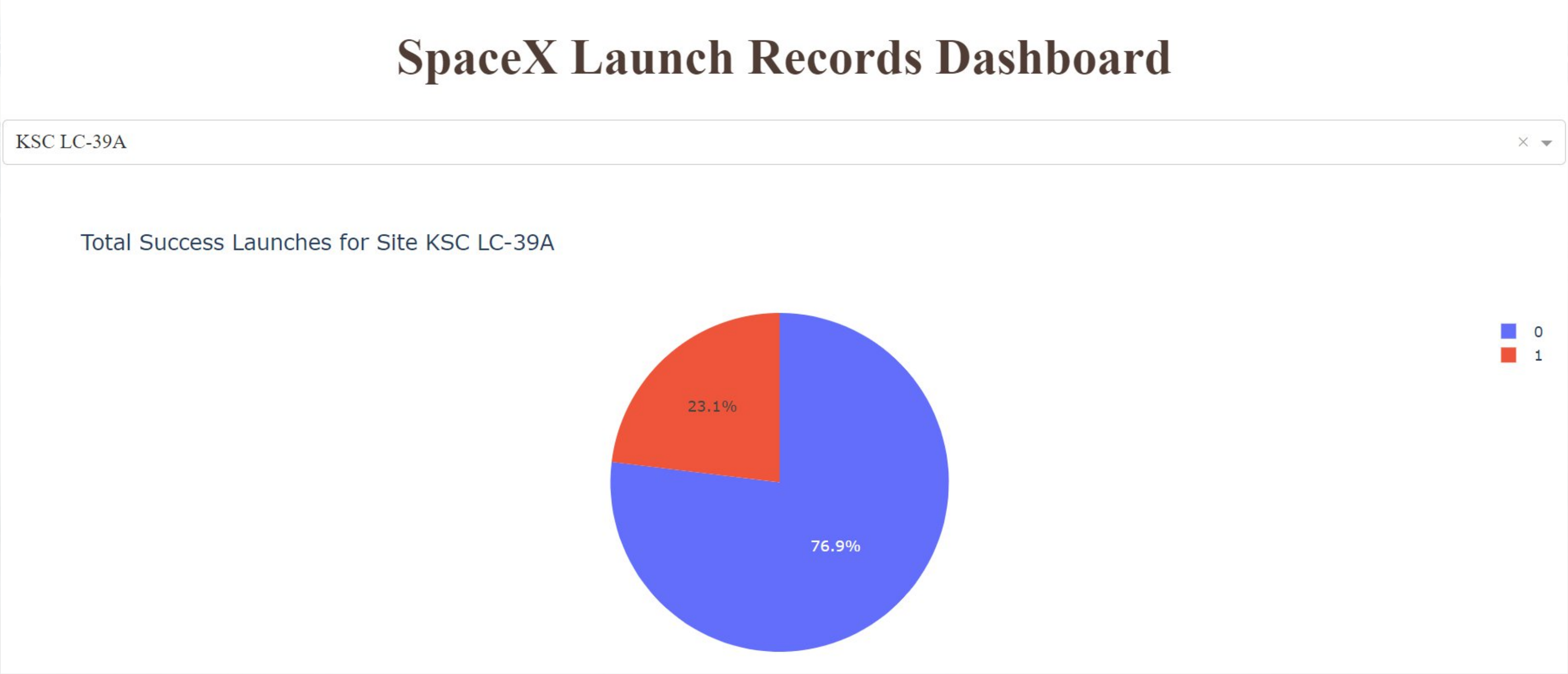


Dashboard

Launch Success (KSC LC-29A)

Success as Percent of Total

- KSC LC-39A has the highest success rate amongst launch sites (76.9%)
- 10 successful launches and 3 failed launches



Dashboard

Payload Mass and Success

By Booster Version

- Payloads between 2,000 kg and 5,000 kg have the highest success rate
- 1 indicating successful outcome and 0 indicating an unsuccessful outcome

Predictive Analytics

Classification

Accuracy

• All the models performed at about the same level and had the same scores and accuracy. This is likely due to the small dataset. The Decision Tree model slightly outperformed the rest when looking at .best_score_

• .best_score_ is the average of all cv folds for a single combination of the parameters

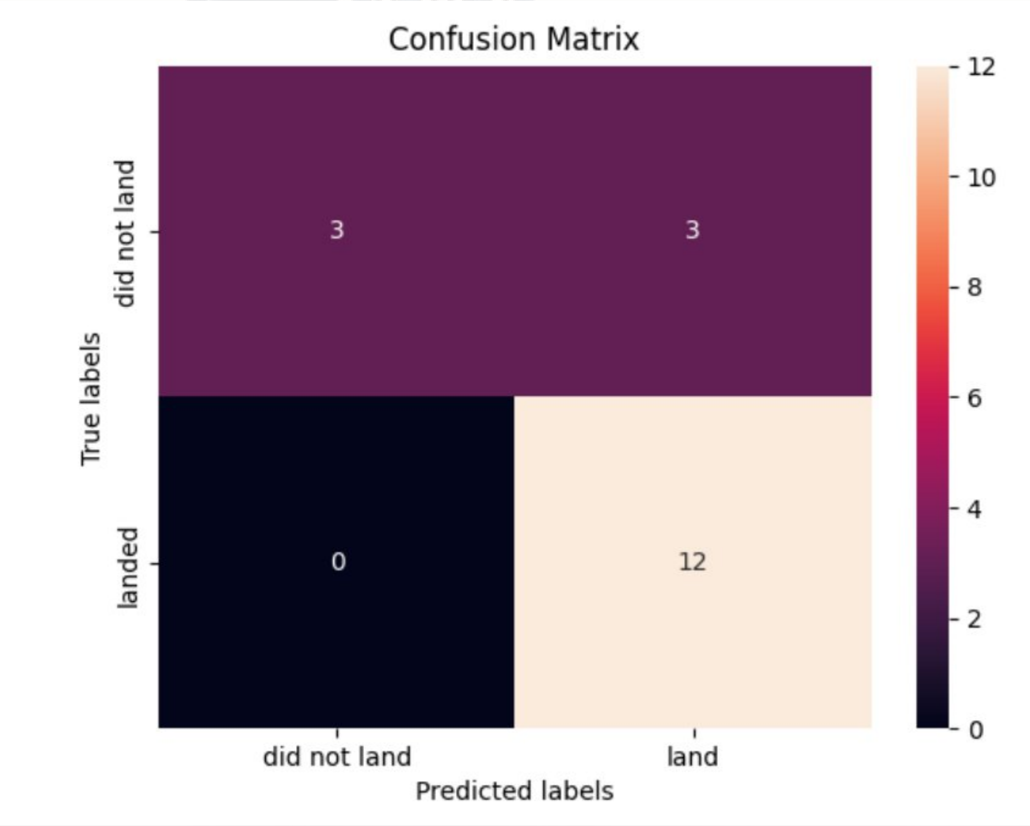
	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

Predictive Analytics

Confusion Matrices

Performance Summary

- A confusion matrix summarizes the performance of a classification algorithm
- All the confusion matrices were identical
- The fact that there are false positives (Type 1 error) is not good
- Confusion Matrix Outputs:
- 12 True positive
- 3 True negative
- 3 False positive
- 0 False Negative
- Precision = TP / (TP + FP)
- 12 / 15 = .80
- Recall = TP / (TP + FN)
- 12 / 12 = 1
- F1 Score = 2 * (Precision * Recall) / (Precision + Recall)
- 2 * (.8 * 1) / (.8 + 1) = .89
- Accuracy = (TP + TN) / (TP + TN + FP + FN) = .833



Conclusion

Research

• Model Performance: The models performed similarly on the test set with the decision tree model slightly outperforming

• Equator: Most of the launch sites are near the equator for an additional natural boost - due to the rotational speed of earth - which helps save the cost of putting in extra fuel and boosters

• Coast: All the launch sites are close to the coast

• Launch Success: Increases over time

• KSC LC-39A: Has the highest success rate among launch sites. Has a 100% success rate for launches less than 5,500 kg

• Orbits: ES-L1, GEO, HEO, and SSO have a 100% success rate

• Payload Mass: Across all launch sites, the higher the payload mass (kg), the higher the success rate

Conclusion

Things to Consider

• Dataset: A larger dataset will help build on the predictive analytics results to help understand if the findings can be generalizable to a larger data set

• Feature Analysis / PCA: Additional feature analysis or principal component analysis should be conducted to see if it can help improve accuracy

• XGBoost: Is a powerful model which was not utilized in this study. It would be interesting to see if it outperforms the other classification models