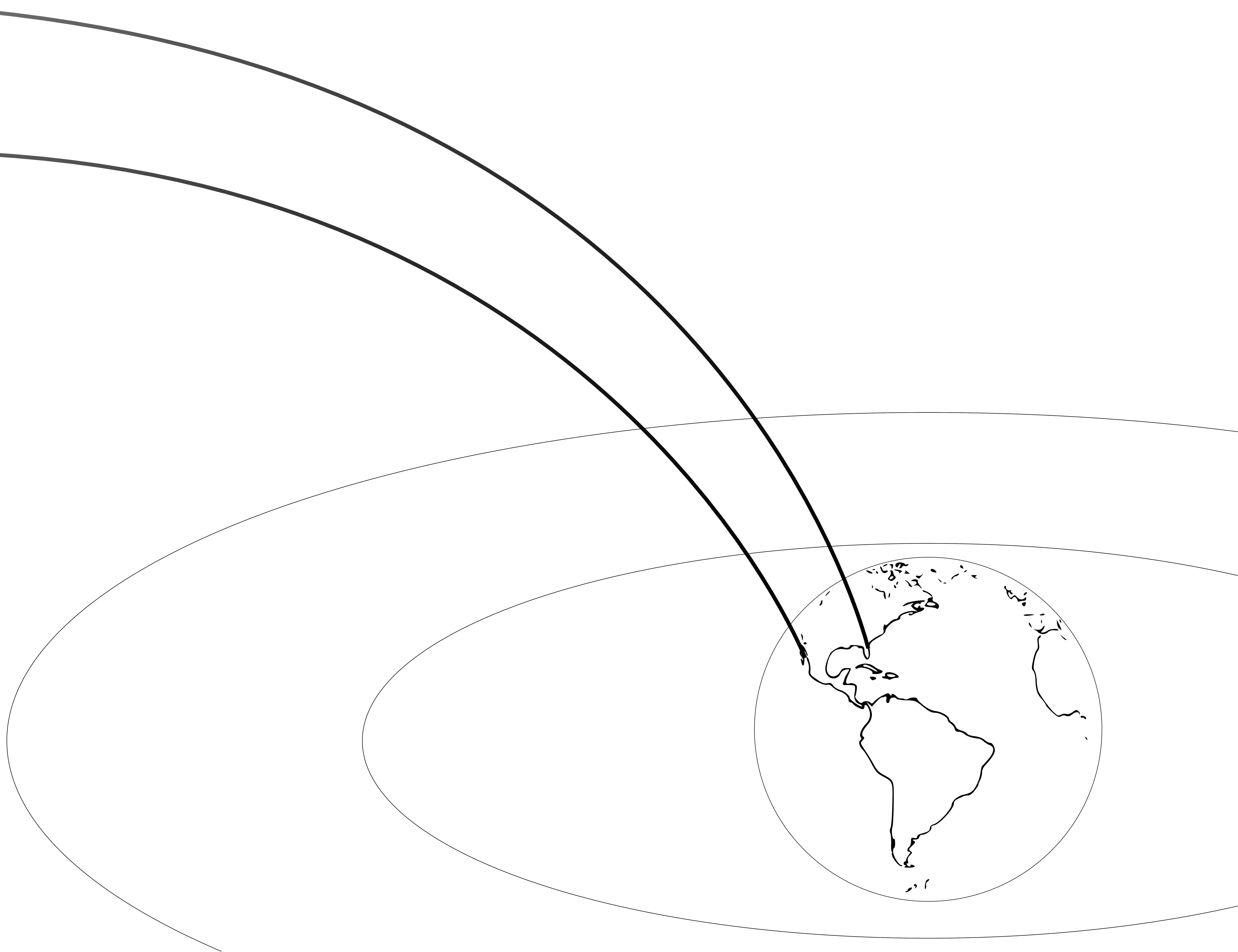
FALCON 9

First Stage Booster Success Rate Analysis



Dimitar Yosifov

20.01.2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
    Visualization - Charts
- Findings & Implications
- Conclusion
- Appendix

Executive Summary

Business Goals

- Data aggregation from available sources
- Data preparation for Analysis
- Data exploration and visualization
- Use of machine learning models for predictions

Analysis Results

- Present predictions for booster reuse

Introduction

SpaceX offers significantly lower cost-per-launch service compared to other providers.
The company is able to achieve low costs thanks to the innovative reuse of their first stage boosters.

In this analysis we will explore the different aspects of "Falcon 9" boosters, their success rate and predict if a booster will be reused.

Methodology

Data Collection

- SpaceX REST API and Web Scaping

Data Wrangling

- Missing values and Filtering

Exploratory Data Analysis

- SQL, Pandas and Matplotlib

Interactive Visual Analytics and Dashboard

- Folium and Dash

Predictive Analysis with Classification

- Classification Algorithms

Data Collection-API

Using Python libraries like "Requests" and "Pandas", we gather data from the SpaceX REST API and convert the JSON response to a DataFrame.

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1060	-80.603956	28.608058
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	13	B1058	-80.603956	28.608058
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1051	-80.603956	28.608058
92	89	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e534e7cc	5.0	12	B1060	-80.577366	28.561857
93	90	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca	5.0	8	B1062	-80.577366	28.561857

90 rows × 17 columns

Data Collection-Web

With the Python libraries "Requests" and "BeautifulSoup", we gather historical data from SpaceX's Wikipedia page and convert the HTML to a DataFrame.

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10
...	...	...	...	...	...	...	...	...	...	...	...
116	117	CCSFS	Starlink	15,600 kg	LEO	SpaceX	Success\n	F9 B5B1051.10	Success	9 May 2021	06:42
117	118	KSC	Starlink	~14,000 kg	LEO	SpaceX	Success\n	F9 B5B1058.8	Success	15 May 2021	22:56
118	119	CCSFS	Starlink	15,600 kg	LEO	SpaceX	Success\n	F9 B5B1063.2	Success	26 May 2021	18:59
119	120	KSC	SpaceX CRS-22	3,328 kg	LEO	NASA	Success\n	F9 B5B1067.1	Success	3 June 2021	17:29
120	121	CCSFS	SXM-8	7,000 kg	GTO	Sirius XM	Success\n	F9 B5	Success	6 June 2021	04:26

121 rows × 11 columns

Data Wrangling

Clean Data is essential for performing dependable analysis. Utilizing functions like "replace", "get_dummies" or list comprehensions, corrects missing values, encodes categorical data and transforms information into suitable state.

```
payload_mean = data_falcon9['PayloadMass'].mean()

data_falcon9['PayloadMass'].replace(np.nan, payload_mean, inplace=True)
data_falcon9.isnull().sum()
```

```
landing_class = [0 if x else 1 for x in df['Outcome'].isin(bad_outcomes)]
landing_class
```

Exploratory Analysis 1/2

As first exploratory step, we create a "Class" column in the dataframe depicting successful and unsuccessful landings of Falcon 9 boosters. We then calculate the average success rate.

Average Success Rate: 0.66

```
df["Class"].mean()
```

```
0.6666666666666666
```

Exploratory Analysis 2/2

SQL is a powerful tool which allows traversal of large datasets.

Additional information can be extracted with data queries.

NASA delivered total payload using Falcon 9: 45596 (kg)

Average payload by booster version F9 v1.1: 2928 (kg)

Display the names of the unique launch sites in the space mission

```
%sql SELECT DISTINCT(LAUNCH_SITE) FROM SPACEXTBL;
```

Done.

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

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

```
%%sql SELECT
    landing__outcome,
    booster_version,
    launch_site
FROM SPACEXTBL
WHERE YEAR(DATE) = 2015 AND landing__outcome = 'Failure (drone ship)';
```

Done.

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Visualization 1/7

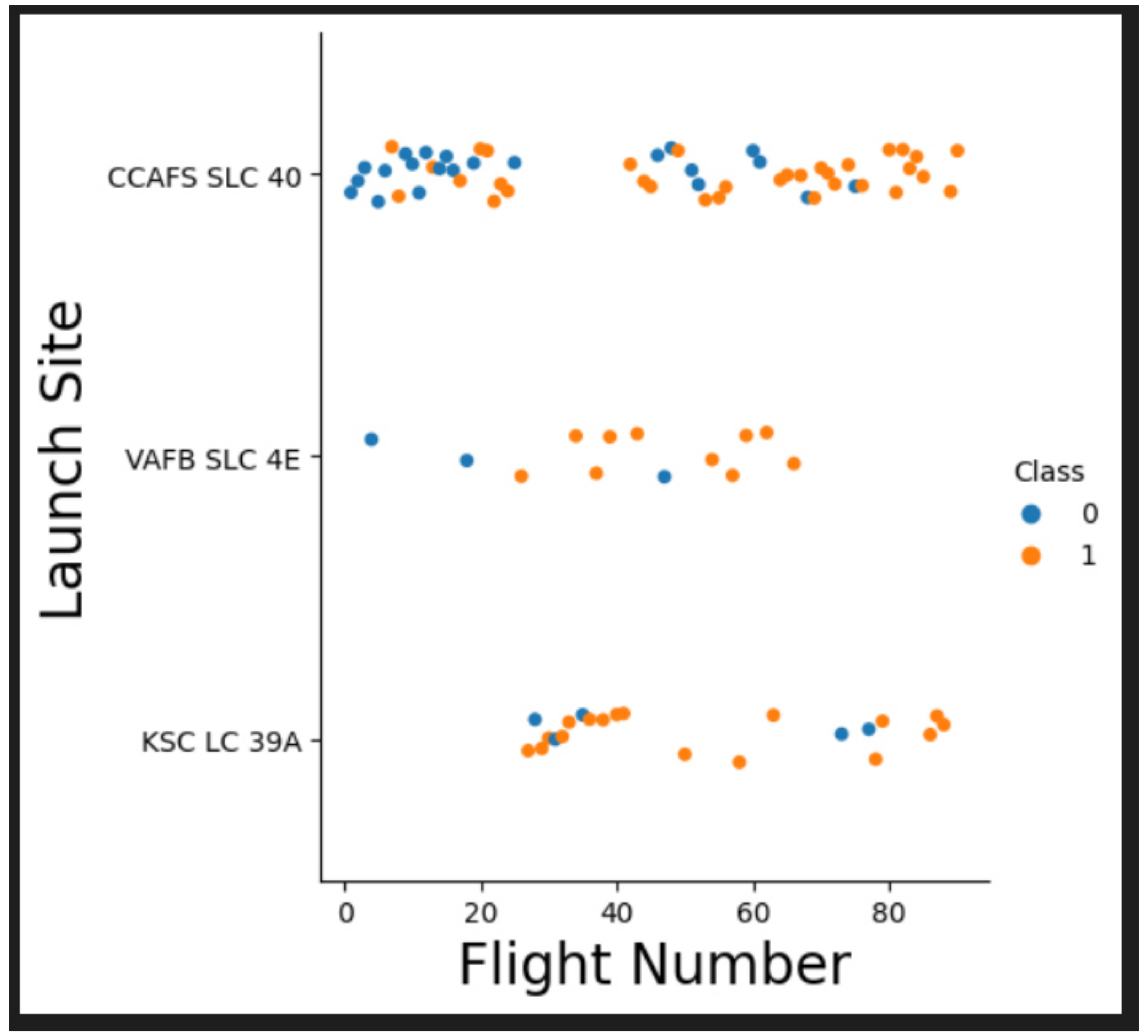
Data Visualization enables insight from complex datasets. Let's look at the relationship between Payload Mass and Number of Flights. It appears that greater payload mass has negative impact on success rate, as well as low flight numbers.

Visualization GitHub Link

Visualization 2/7

Let's explore the relationship between Launch Site and Flight Numbers.

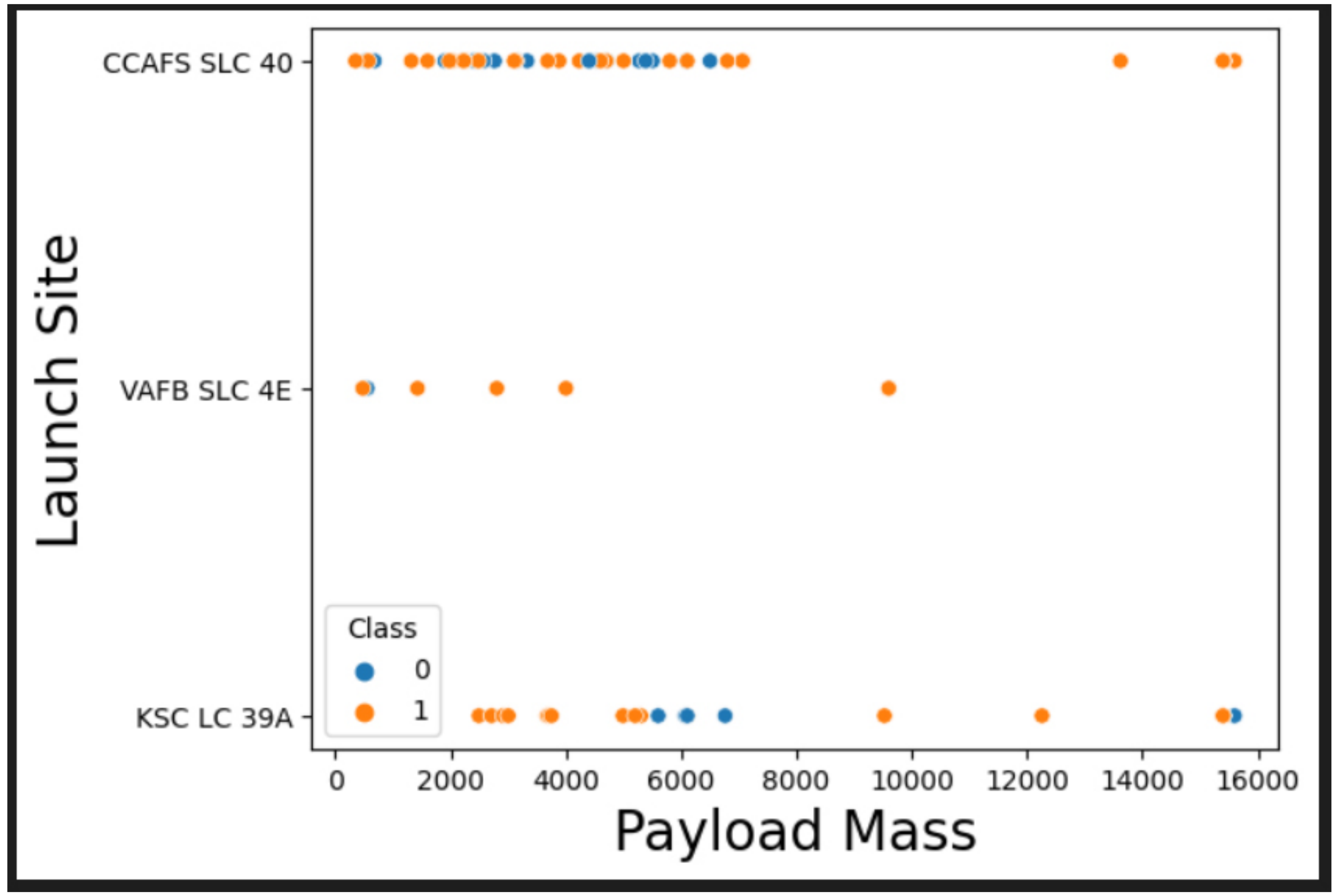
We observe greater frequency of flights from SLC 40, and a spread of successful outcomes for different Flight Numbers.



Visualization GitHub Link

Visualization 3/7

Observing the relationship between Launch Site and Payload Mass, we see that there are no payloads greater than 10000 (kg) served from SLC 4E, while SLC 40 has successful rates for maximum payload size.



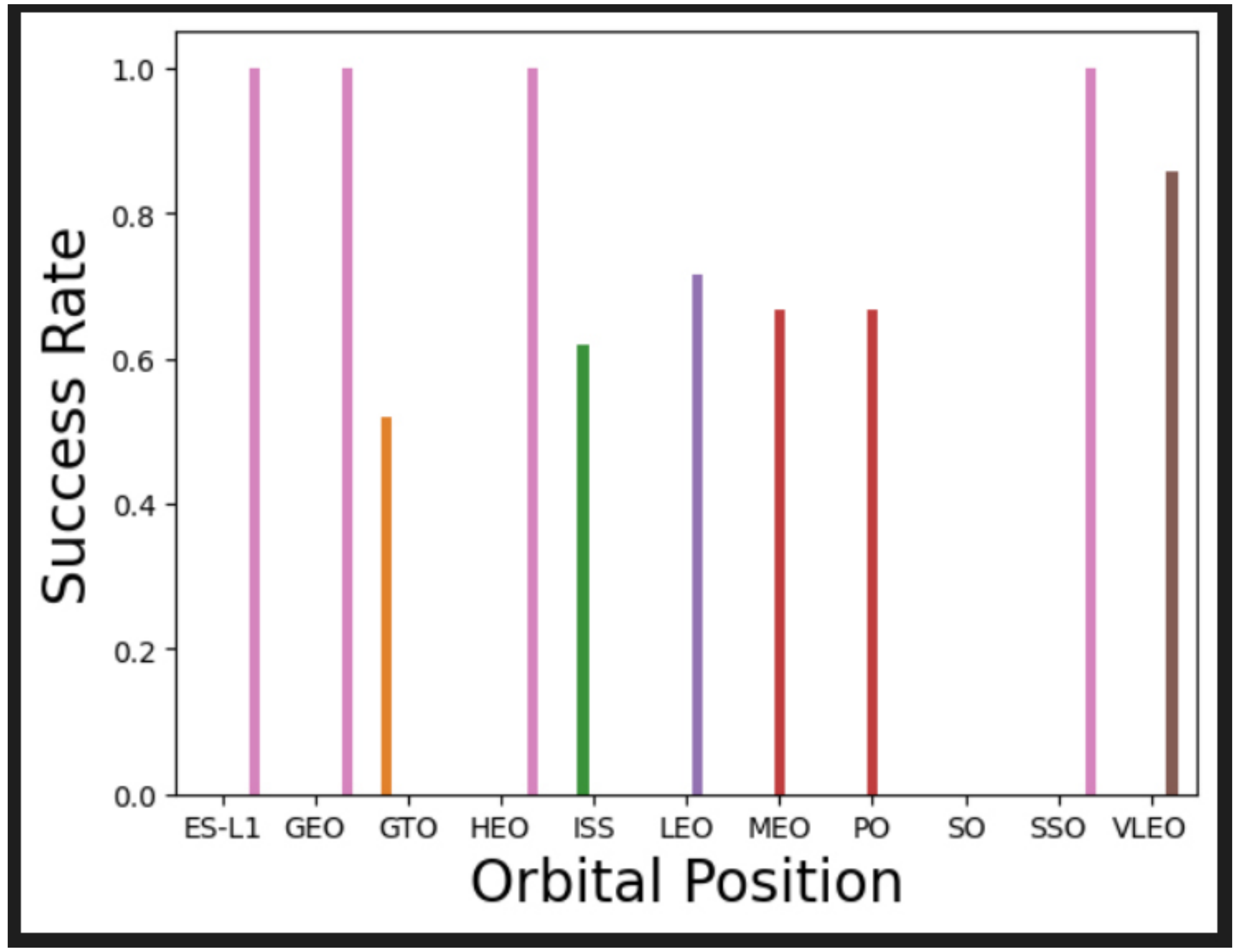
Visualization GitHub Link

Visualization 4/7

In the following graph, we observe the relationship between different Orbital positions and Success Rate.

Positions with High success rate: ES-L1, GEO, HEO, SSO, VLEO

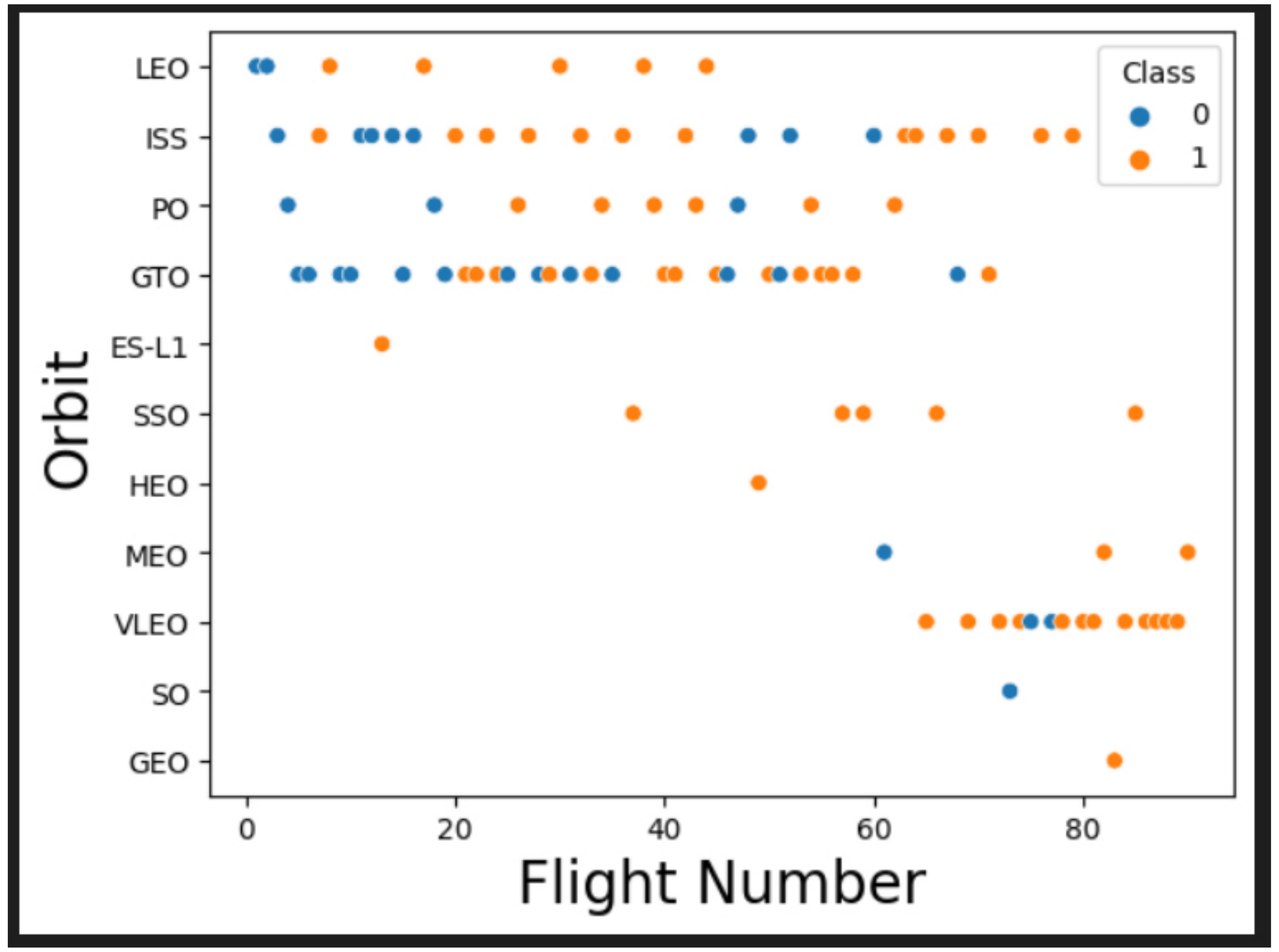
Positions with Medium success : GTO, ISS, LEO, MEO, PO



Visualization GitHub Link

Visualization 5/7

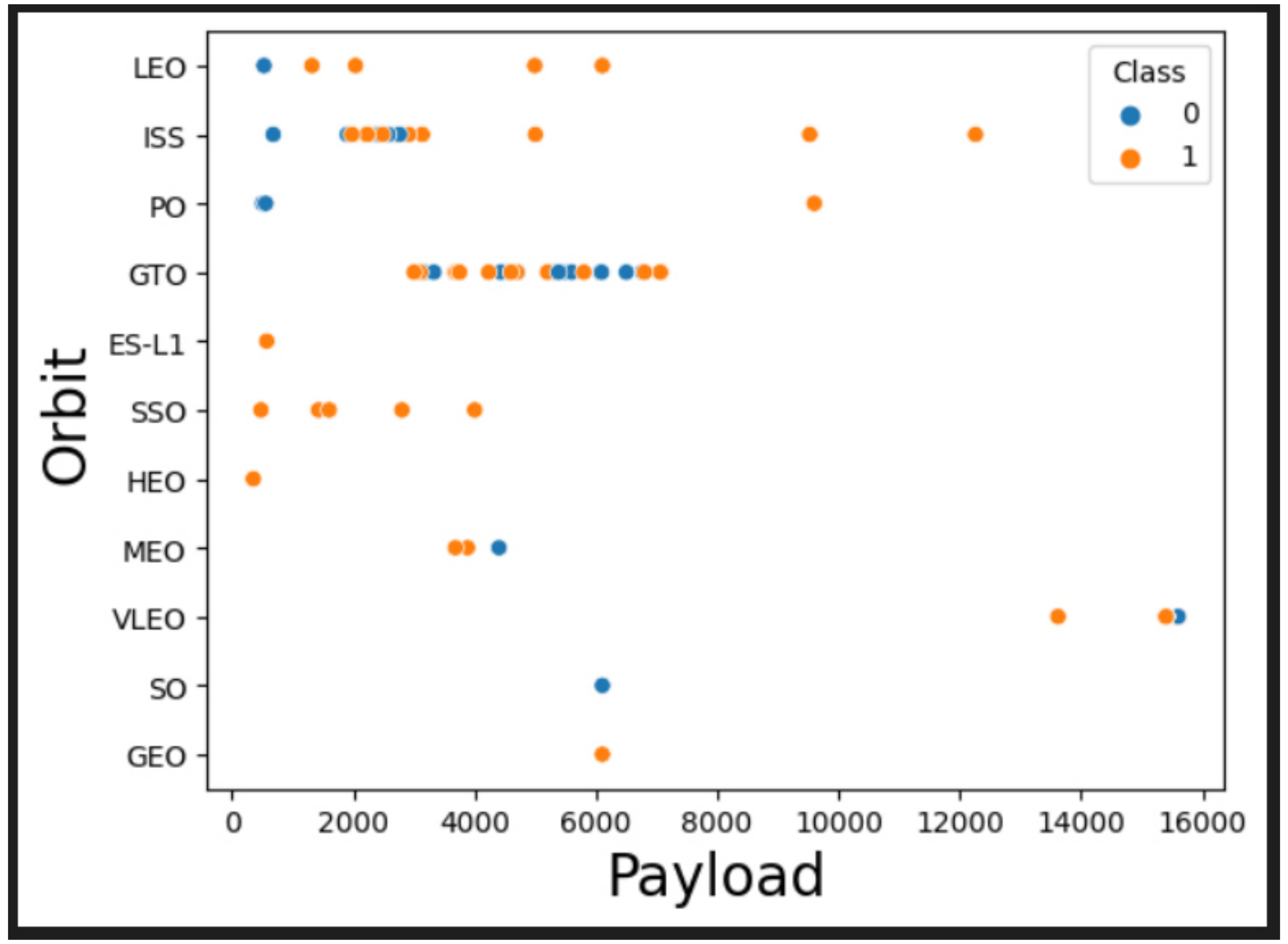
With the following visualization, we compare the relationship between Flight Number and Orbit type.
It appears that the success of some Orbit types is related to the flight numbers, while there's no relation in others.



Visualization GitHub Link

Visualization 6/7

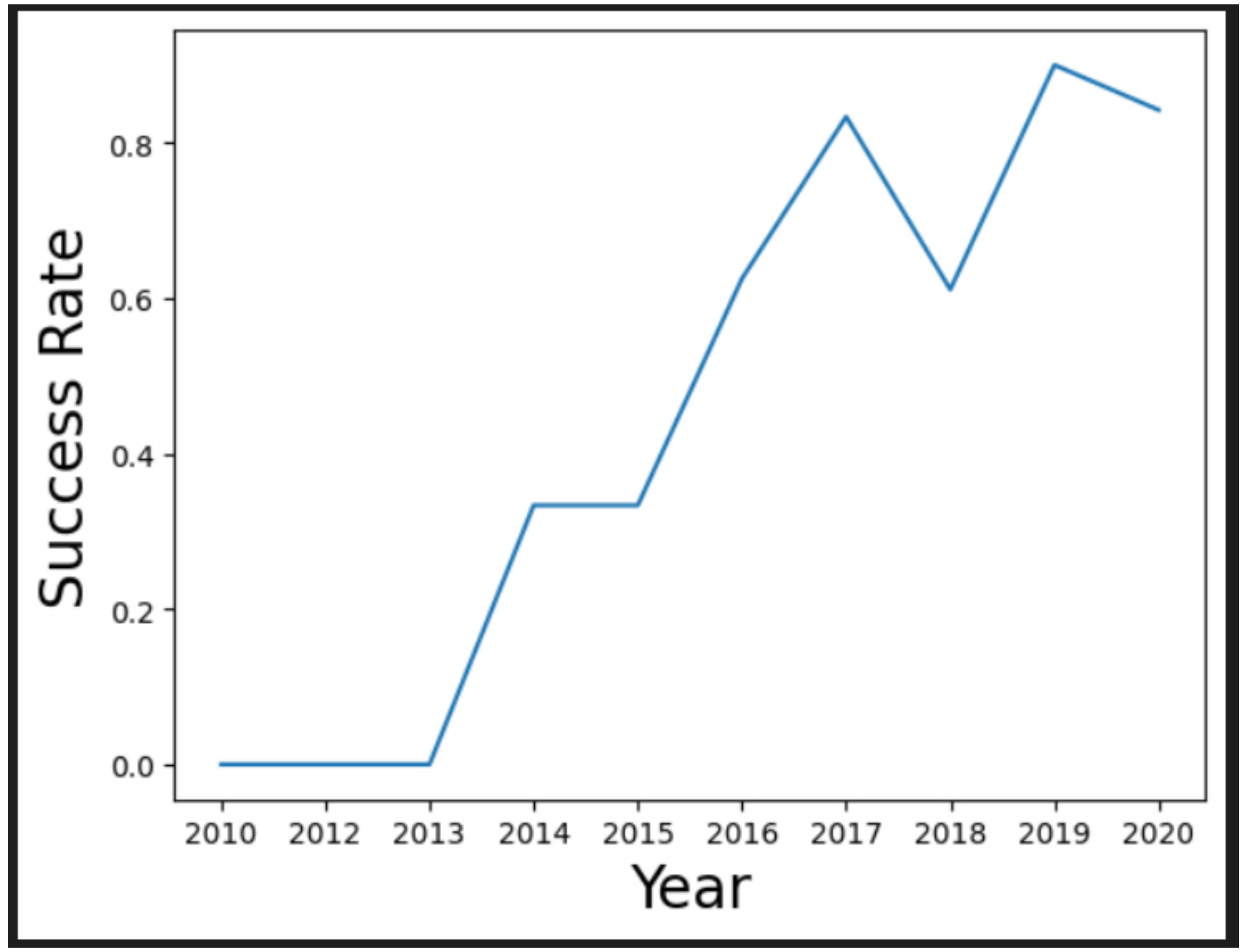
Let's explore the relationship between Orbit type and Payload size. It appears that successful outcomes with heavy payloads are more common for Polar, LEO and ISS destinations.



Visualization GitHub Link

Visualization 7/7

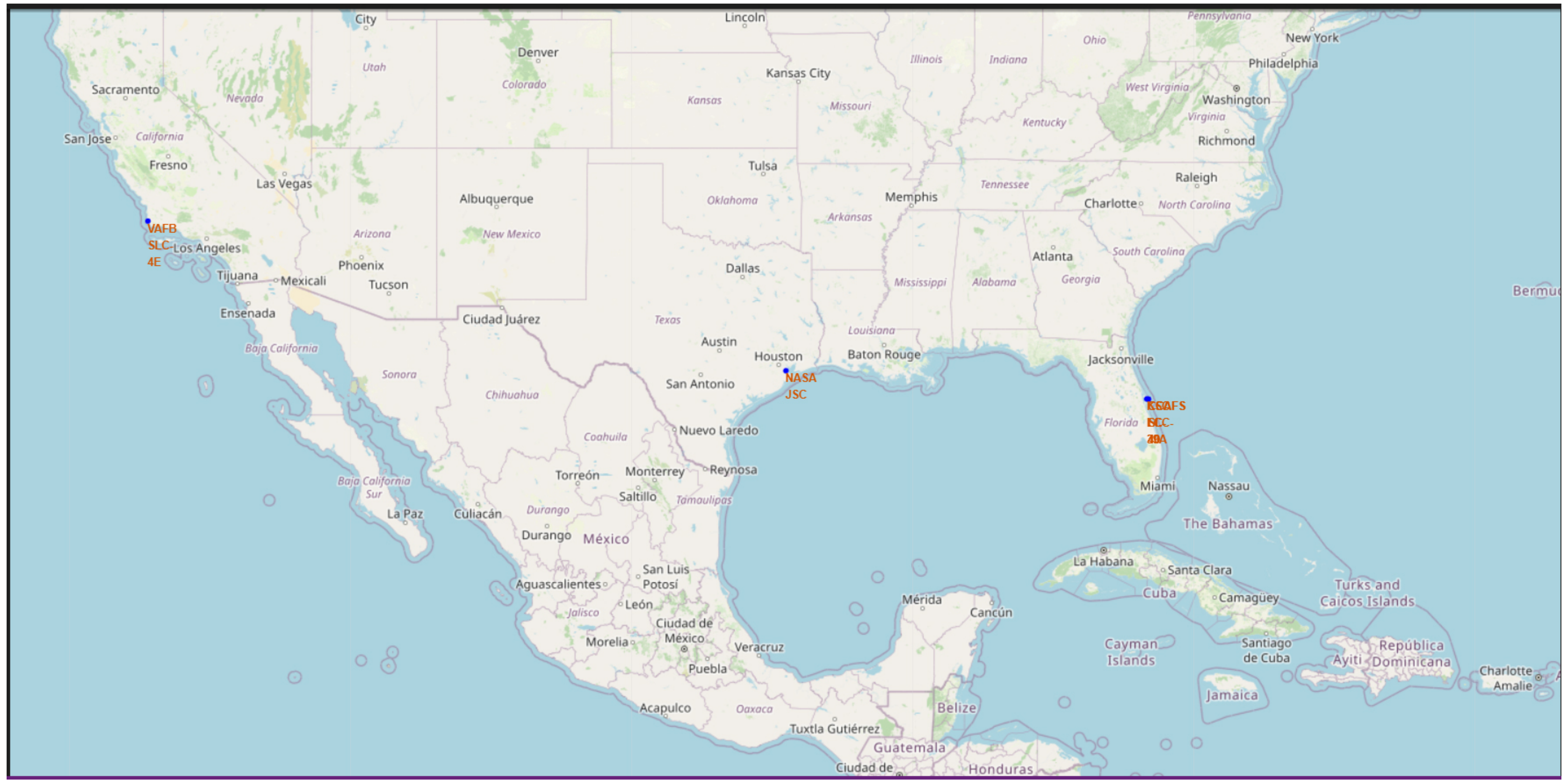
Finally, let's observe the trends for successful outcomes through the years 2010-2020.



Visualization GitHub Link

Interactive Visualization 1/3

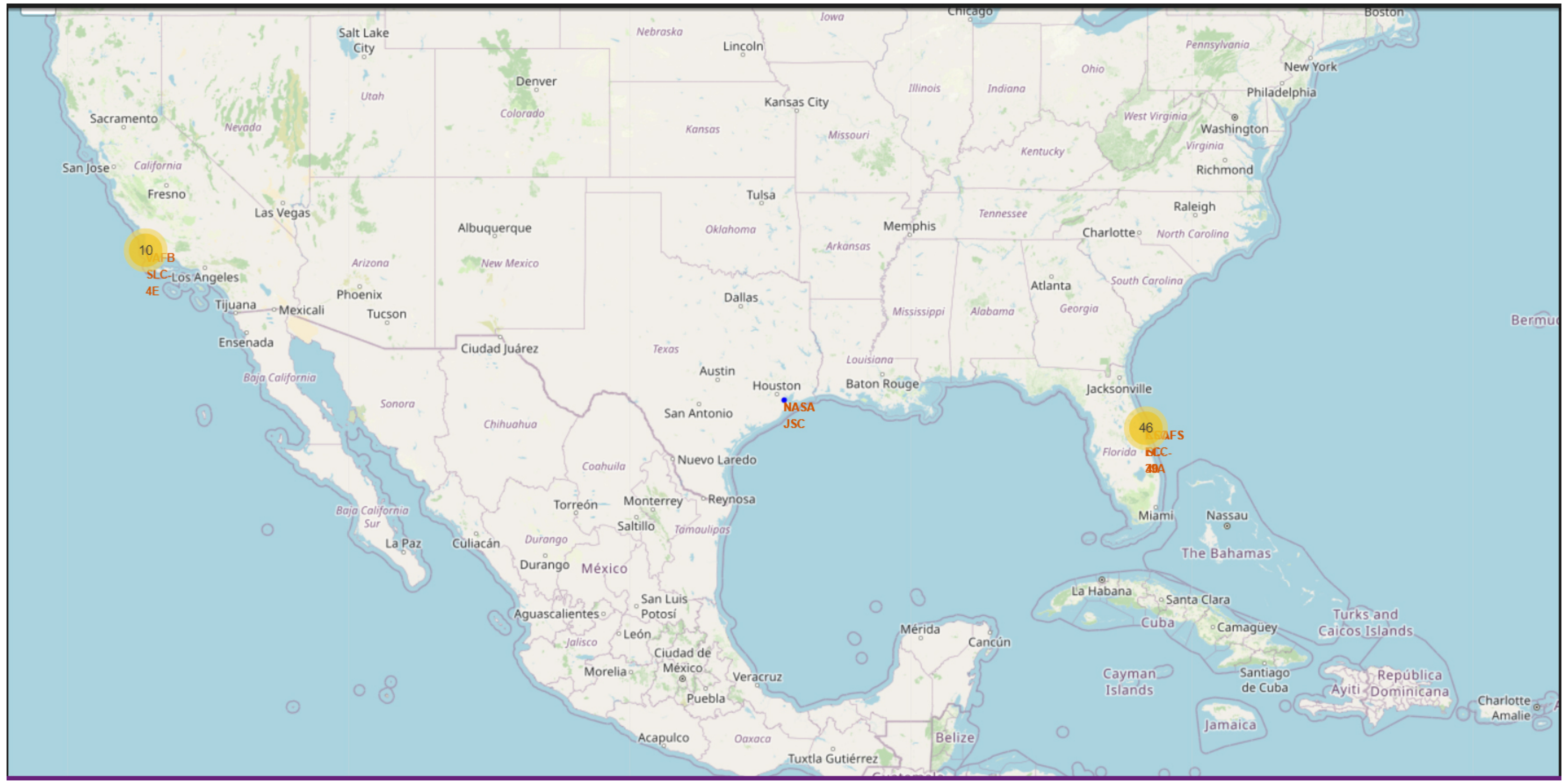
By using interactive visualization libraries like Folium, we can understand relationships of distance and location.



Visualization GitHub Link

Interactive Visualization 2/3

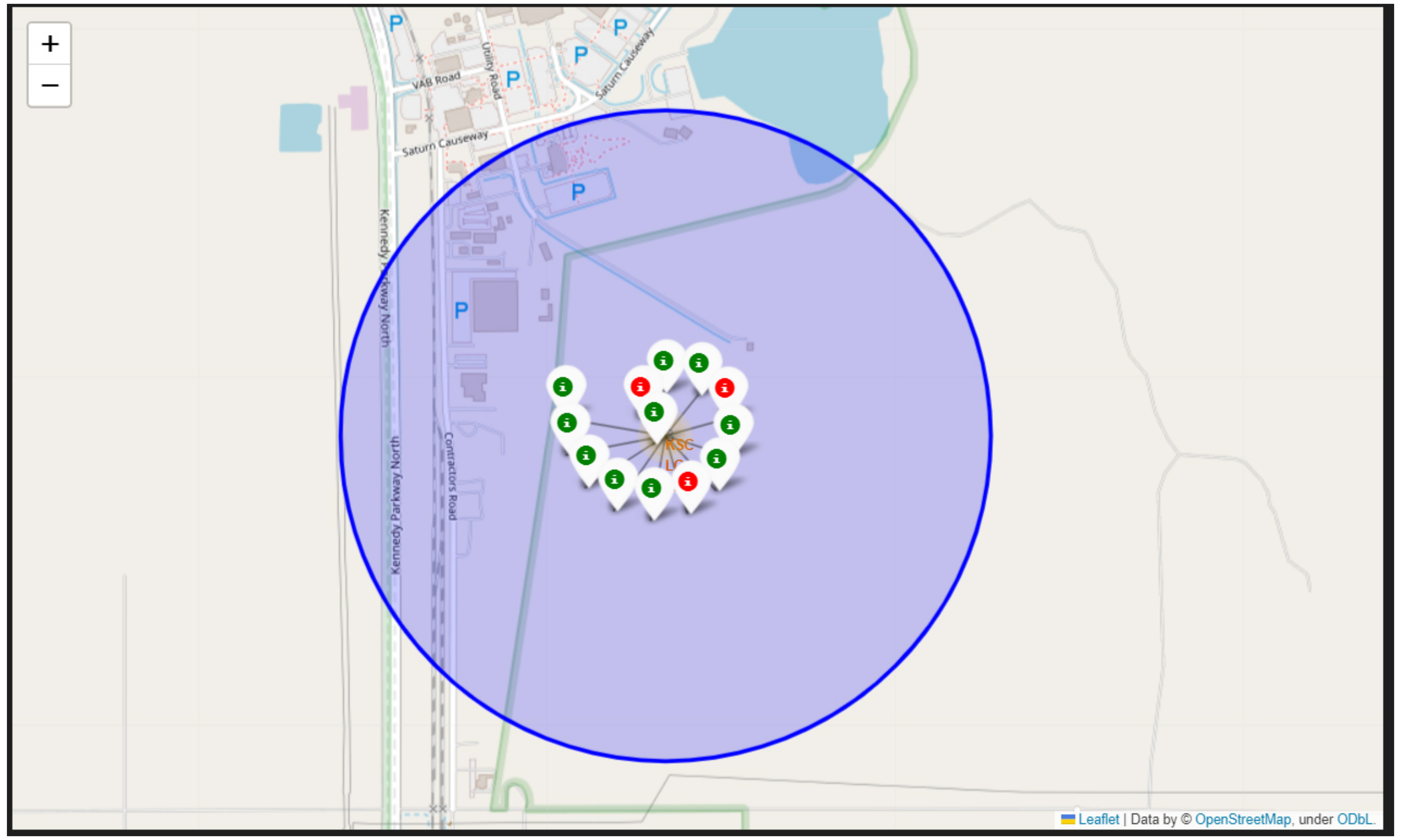
On this representation, we clearly see that there's high preference for launches being performed at CCAFS.



Visualization GitHub Link

Interactive Visualization 3/3

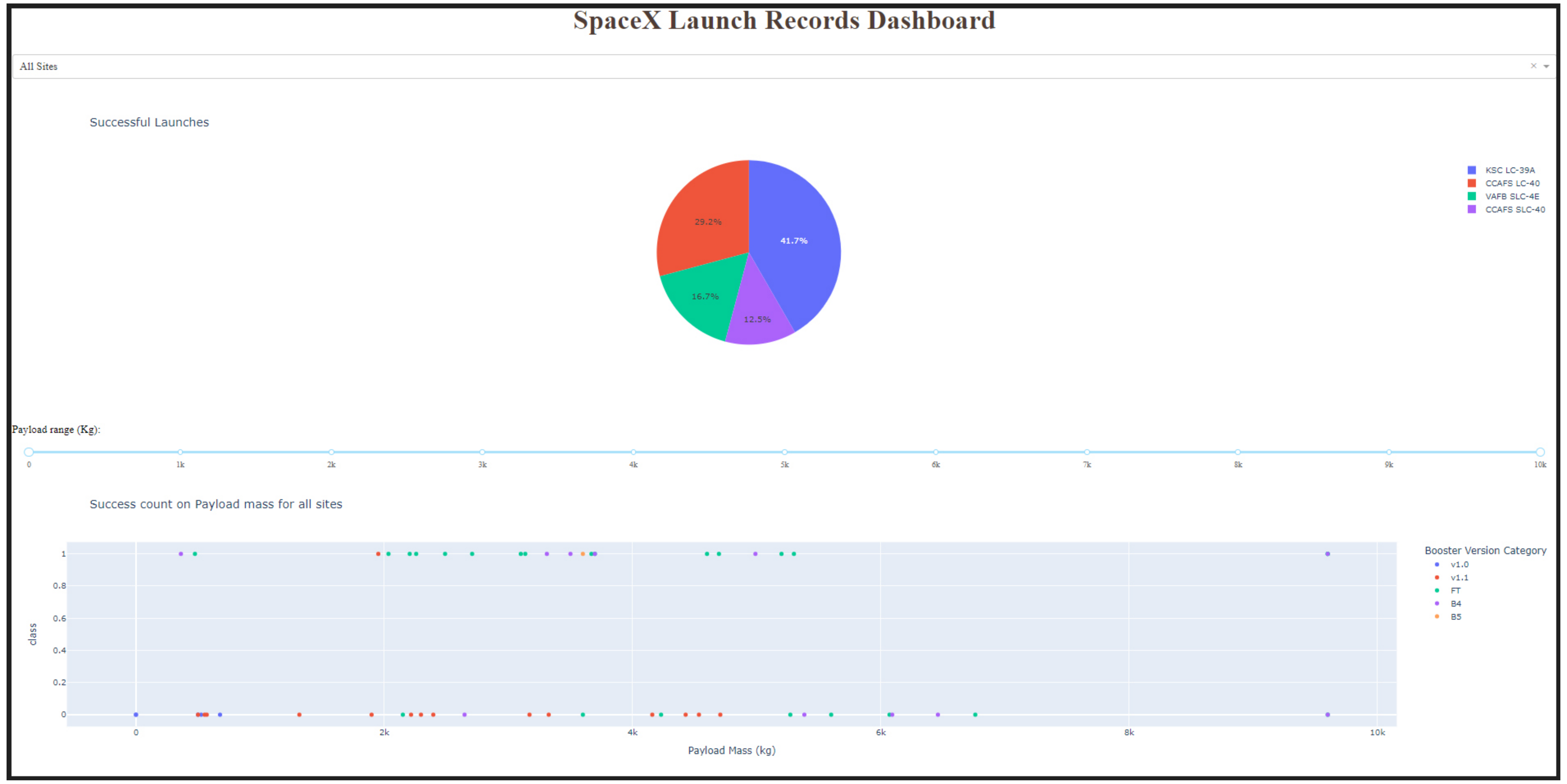
Launch sites tend to keep distance from populated areas, while being close to infrastructure like railways.



Visualization GitHub Link

Interactive Visualization 1/4

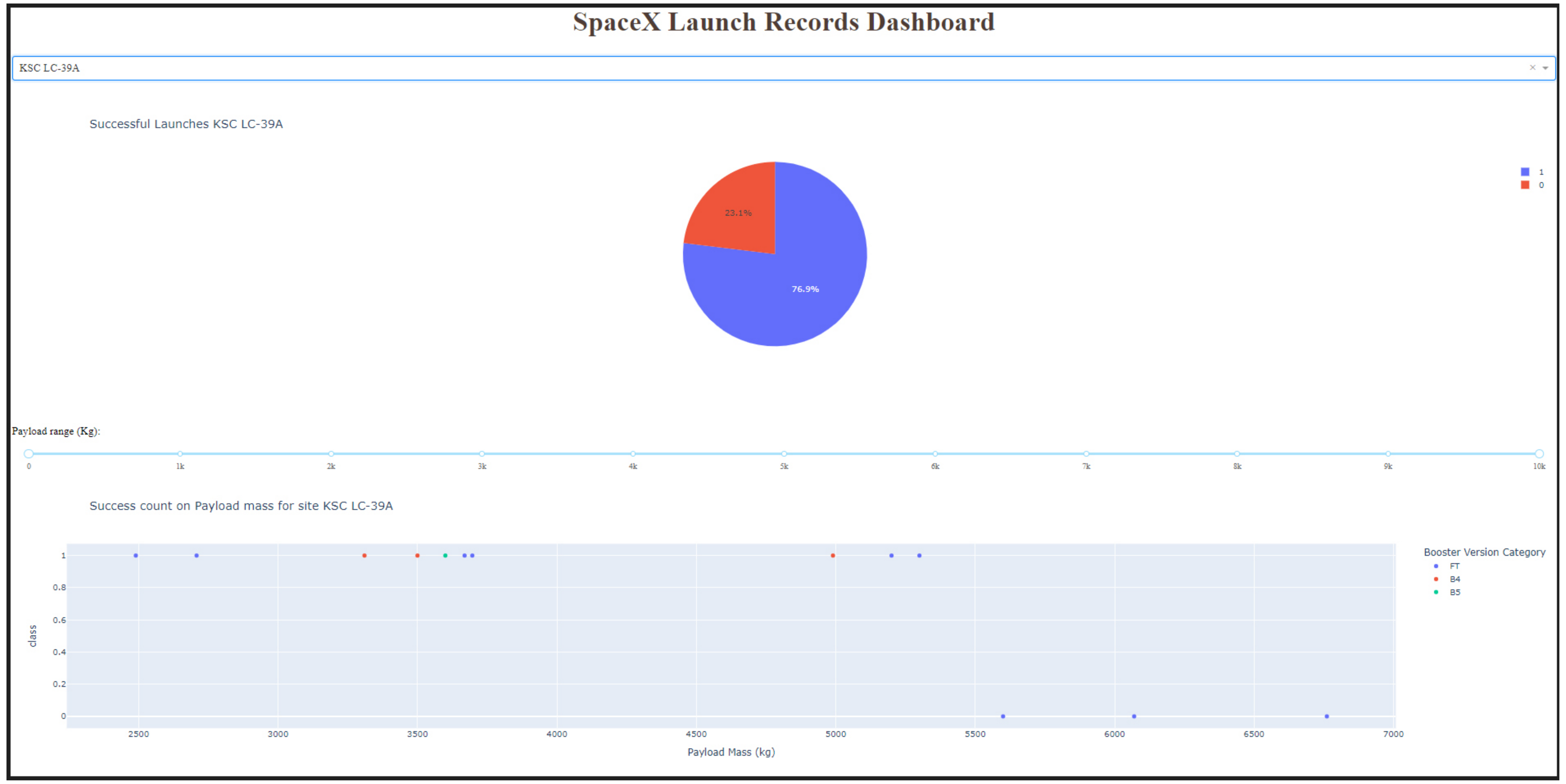
Interactive Dashboards further enhance our ability to segment different aspects of data and focus on specific combinations of observations. In this summery of successful launches for all sites, we observe that KSC has the highest success rate.



Visualization GitHub Link

Interactive Visualization 2/4

Taking a closer look, launches at KSC have 76.9% success rate in booster reusability.



Visualization GitHub Link

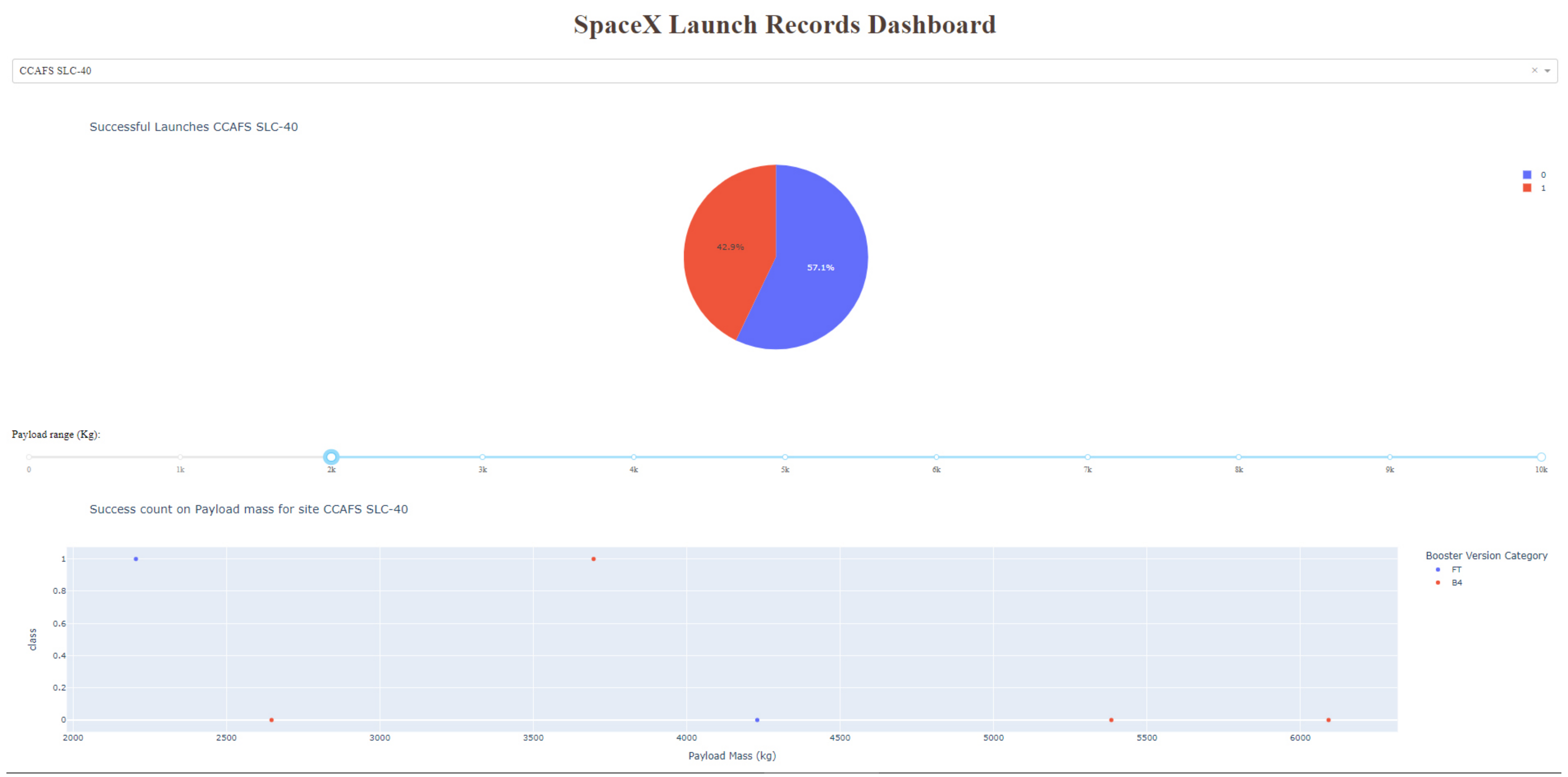
Interactive Visualization 3/4

Using the Scatterplot representation, we observe that although negative outcomes are present, they are proportionally acceptable.

Visualization GitHub Link

Interactive Visualization 4/4

Another comparison shows SLC 40 with 42.9% successful launches and Kennedy Space Center with 76.9% success rate.



Visualization GitHub Link

Predictive Analysis 1/4

We train four models with Grid Search, for optimal parameters.

Logistic Regression:

84% accuracy on training data.

83% on test data.

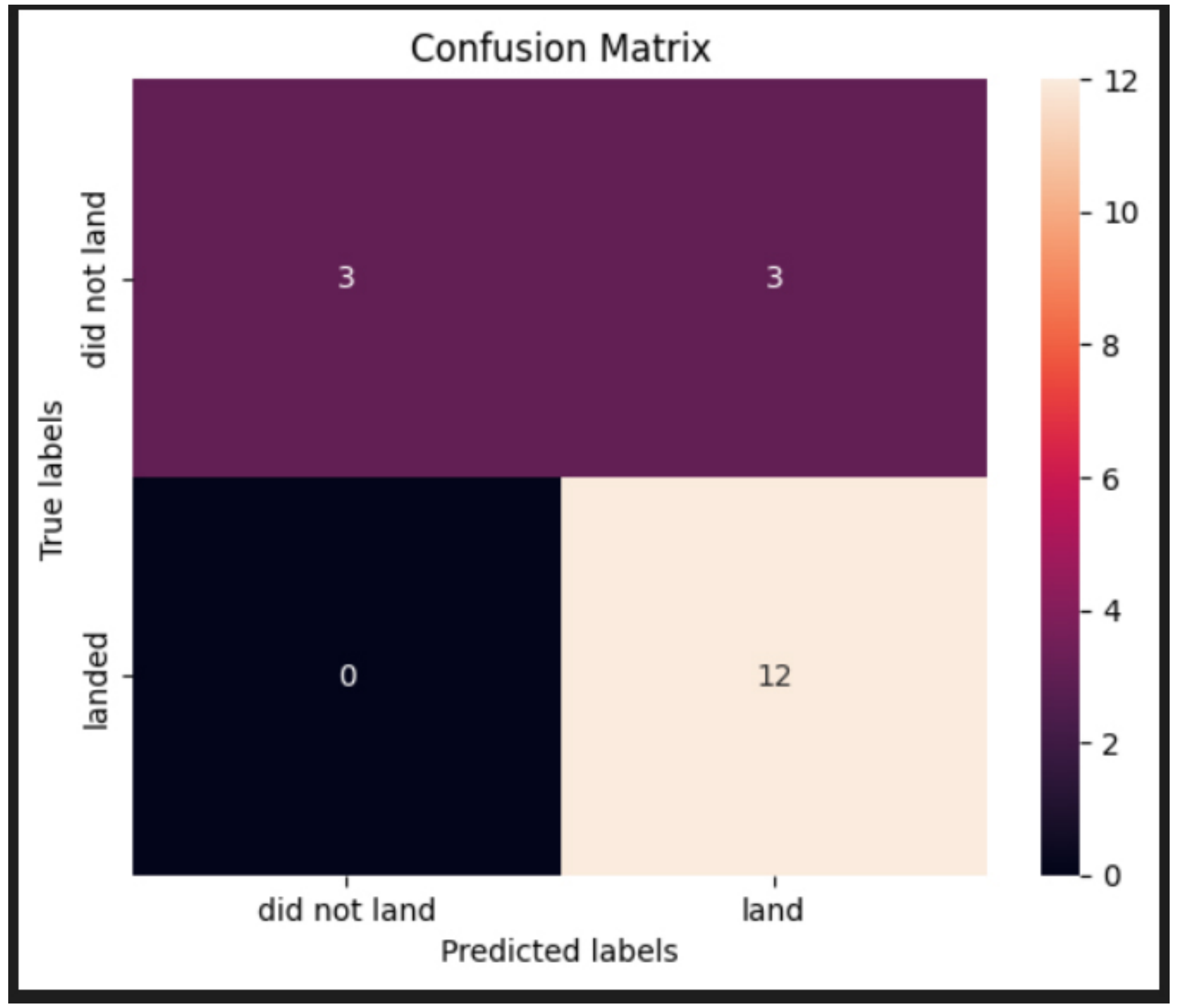
```
parameters ={"C":[0.01,0.1,1],'penalty':['l2'], 'solver':['lbfgs']}
lr=LogisticRegression()

logreg_cv = GridSearchCV(lr, parameters, cv=10)
logreg_cv.fit(X_train, Y_train)
```

```
print("tuned hpyerparameters :(best parameters) ",logreg_cv.best_params_)
print("accuracy :",logreg_cv.best_score_)
```

```
tuned hpyerparameters :(best parameters)  {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
accuracy : 0.8464285714285713
```

```
logreg_cv.score(X_test, Y_test)
```

```
0.8333333333333334
```



Predictive Analysis 2/4

Support Vector Machine:
84% accuracy on training data.
83% on test data.

```
parameters = {'kernel':('linear', 'rbf','poly','rbf', 'sigmoid'),
              'C': np.logspace(-3, 3, 5),
              'gamma':np.logspace(-3, 3, 5)}
svm = SVC()
```

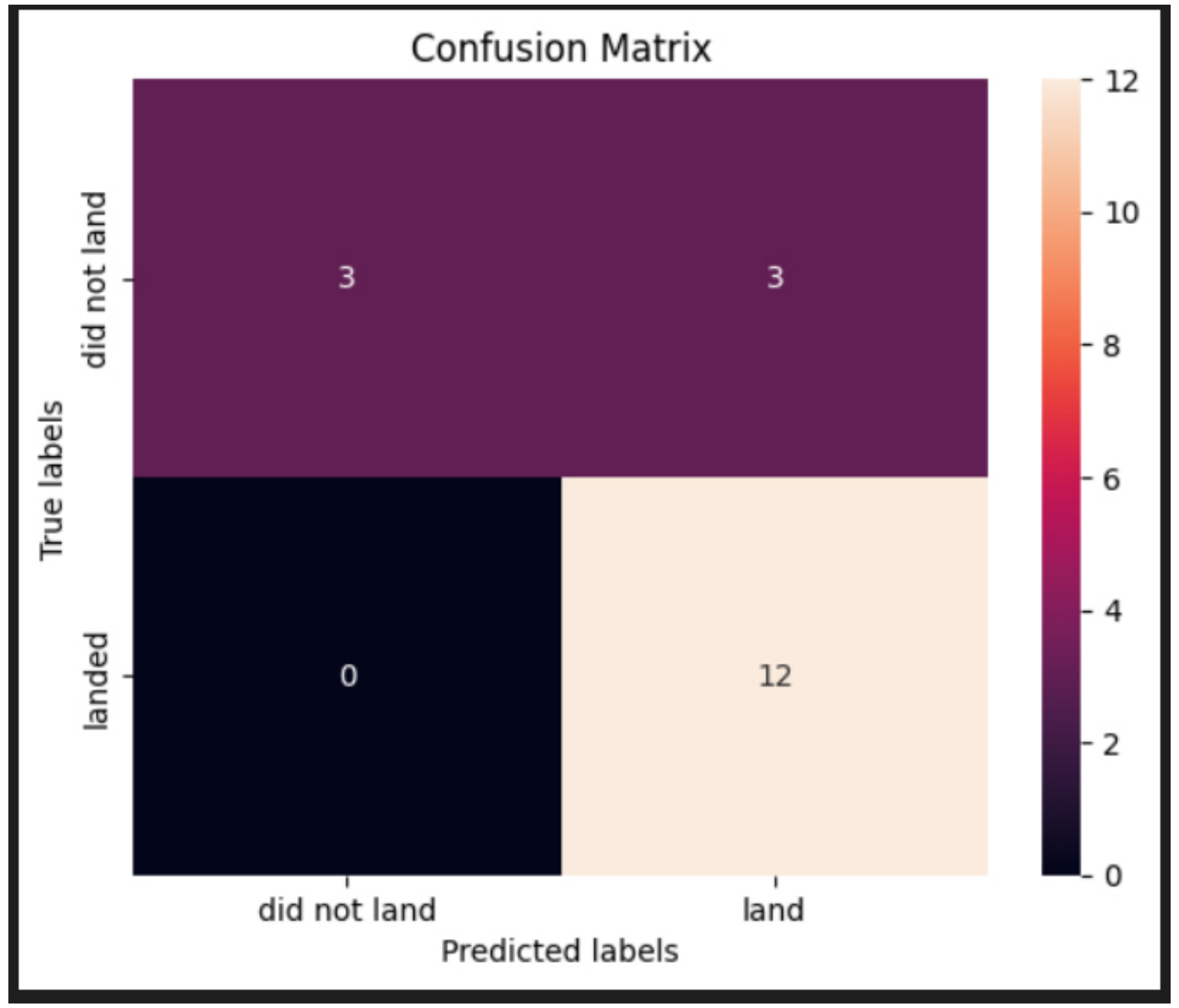
```
svm_cv = GridSearchCV(svm, parameters, cv=10)
svm_cv.fit(X_train, Y_train)
```

```
print("tuned hpyerparameters :(best parameters) ",svm_cv.best_params_)
print("accuracy :",svm_cv.best_score_)
```

```
tuned hpyerparameters :(best parameters)  {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
accuracy : 0.8482142857142856
```

```
svm_cv.score(X_test, Y_test)
```

```
0.8333333333333334
```



Predictive Analysis 3/4

Decision Trees:

87% accuracy on training data.

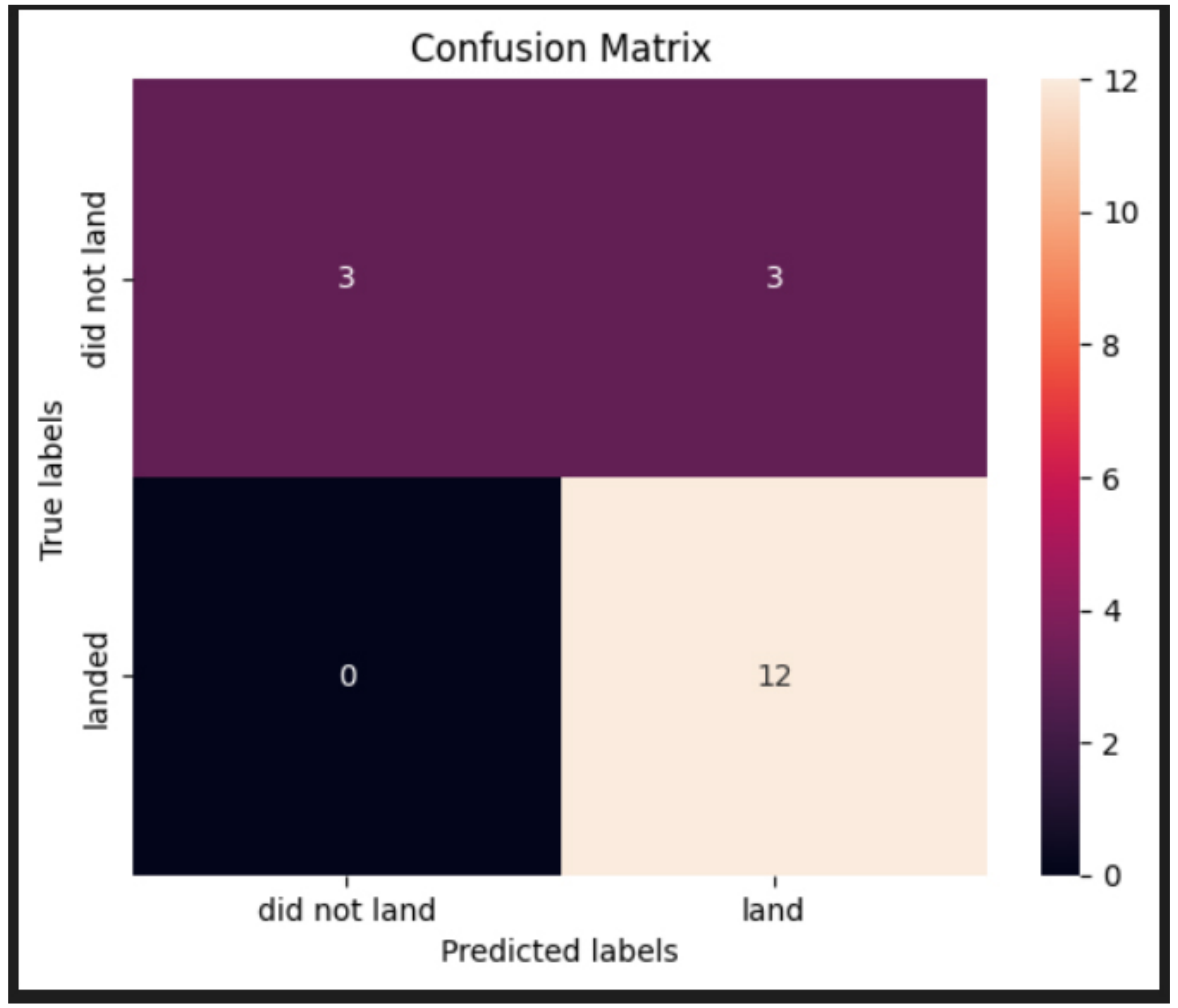
83% on test data.

```
parameters = {'criterion': ['gini', 'entropy'],
     'splitter': ['best', 'random'],
     'max_depth': [2*n for n in range(1,10)],
     'max_features': ['auto', 'sqrt'],
     'min_samples_leaf': [1, 2, 4],
     'min_samples_split': [2, 5, 10]}

tree = DecisionTreeClassifier()

tree_cv = GridSearchCV(tree, parameters, cv=10)
tree_cv.fit(X_train, Y_train)

print("tuned hpyerparameters :(best parameters) ",tree_cv.best_params_)
print("accuracy :",tree_cv.best_score_)
```

tuned hpyerparameters :(best parameters) {'criterion': 'gini', 'max_depth': 6, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 2, 'splitter': 'random'}
accuracy : 0.8767857142857143

```
tree_cv.score(X_test, Y_test)
```

0.8333333333333334



Predictive Analysis 4/4

K-Nearest Neighbor:

84% accuracy on training data.

83% on test data.

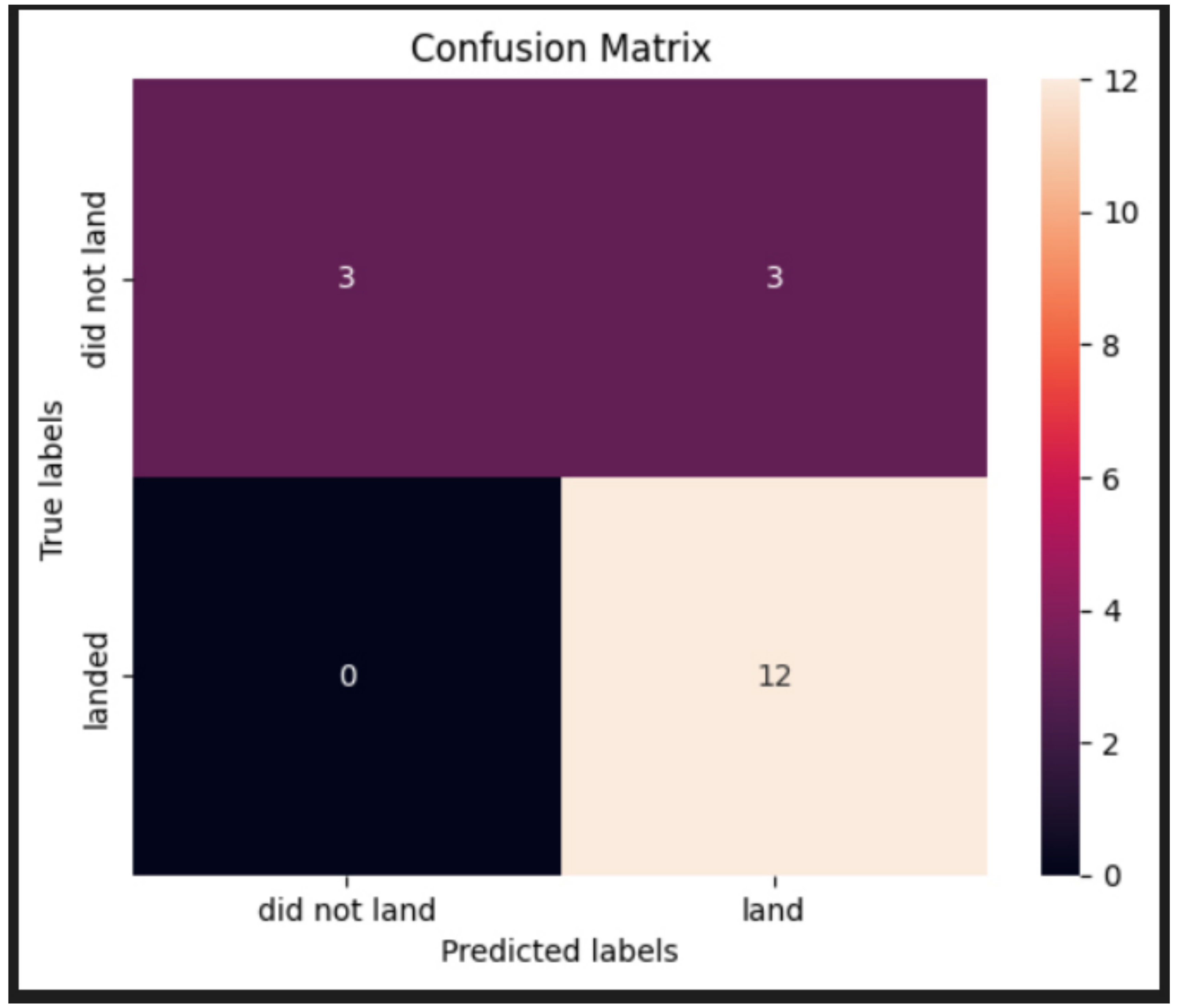
```
parameters = {'n_neighbors': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
              'algorithm': ['auto', 'ball_tree', 'kd_tree', 'brute'],
              'p': [1,2]}

KNN = KNeighborsClassifier()
```

```
knn_cv = GridSearchCV(KNN, parameters, cv=10)
knn_cv.fit(X_train, Y_train)

print("tuned hpyerparameters :(best parameters) ",knn_cv.best_params_)
print("accuracy :",knn_cv.best_score_)
```

```
tuned hpyerparameters :(best parameters)  {'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}
accuracy : 0.8482142857142858
```

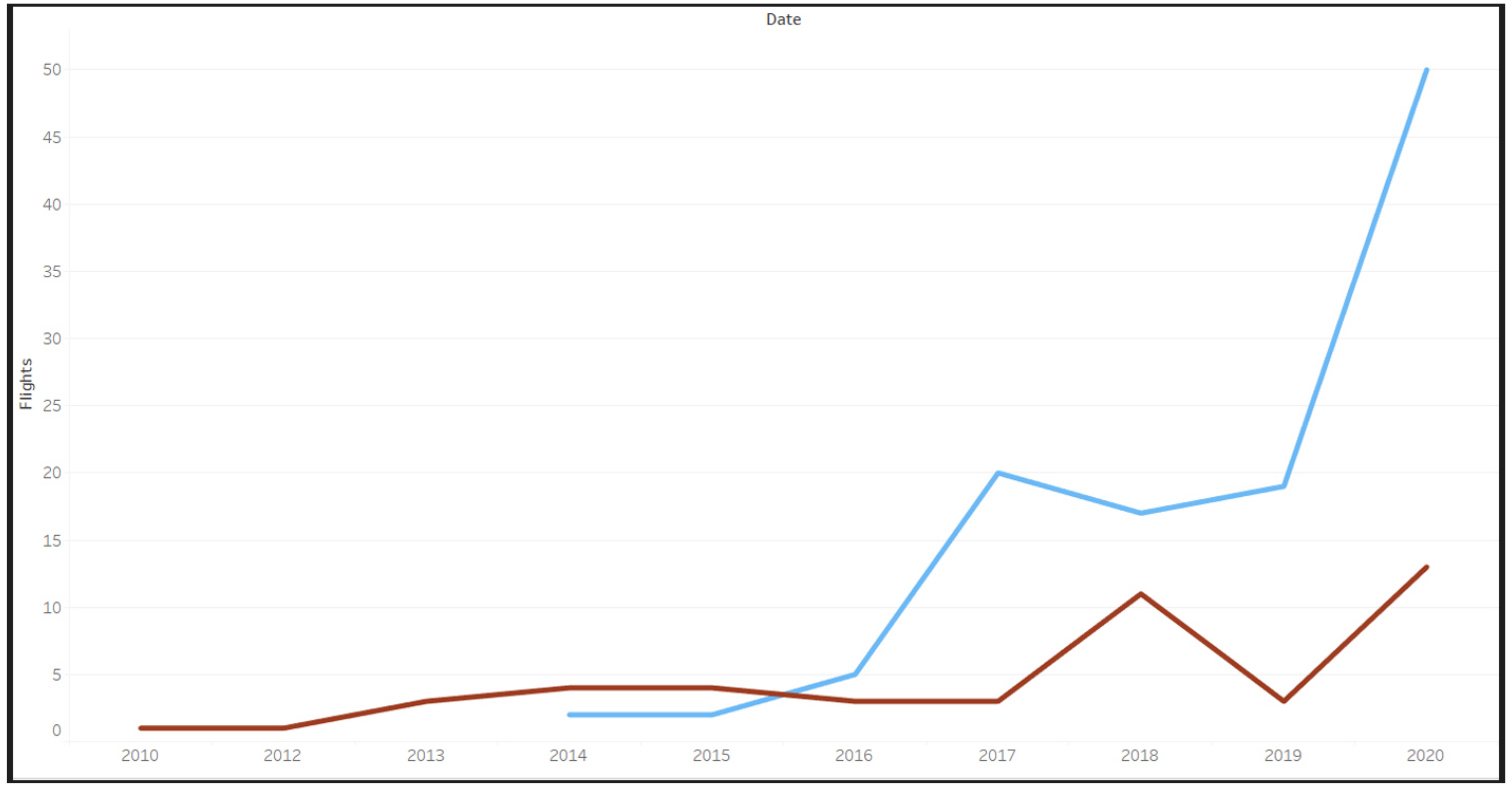
```
knn_cv.score(X_test, Y_test)
```

```
0.8333333333333334
```



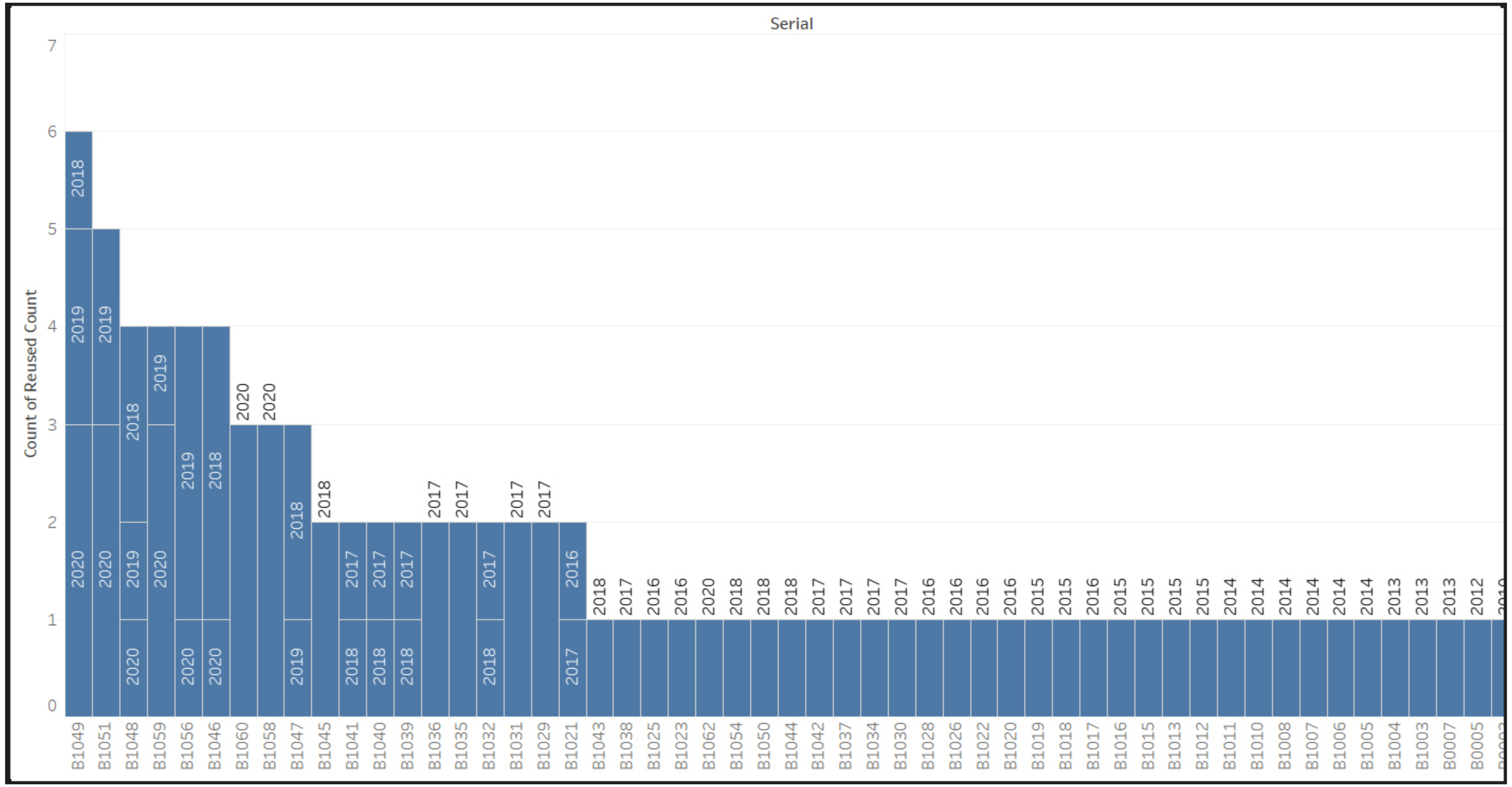
Results 1/4

From 2010 to 2020, Falcon 9 reusability rate has improved drastically, while significantly increasing the number of flights.
2020: 50 successful outcomes, 13 unsuccessful
2018: 17 successful outcomes, 11 unsuccessful



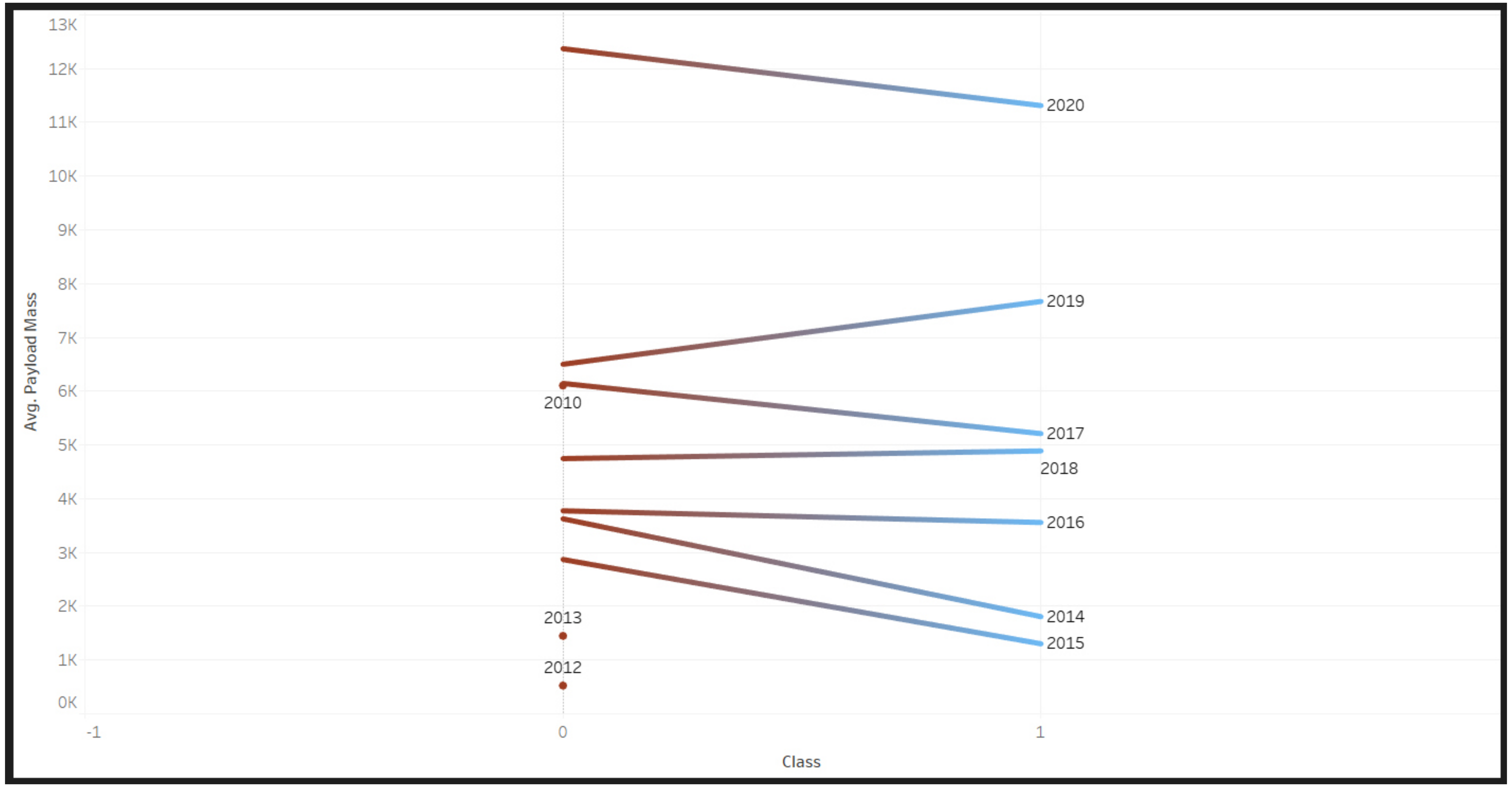
Results 2/4

Reuse rate has also improved with booster B1049 reaching 6 launches in 2020, while current records are 15 launches in 2022.



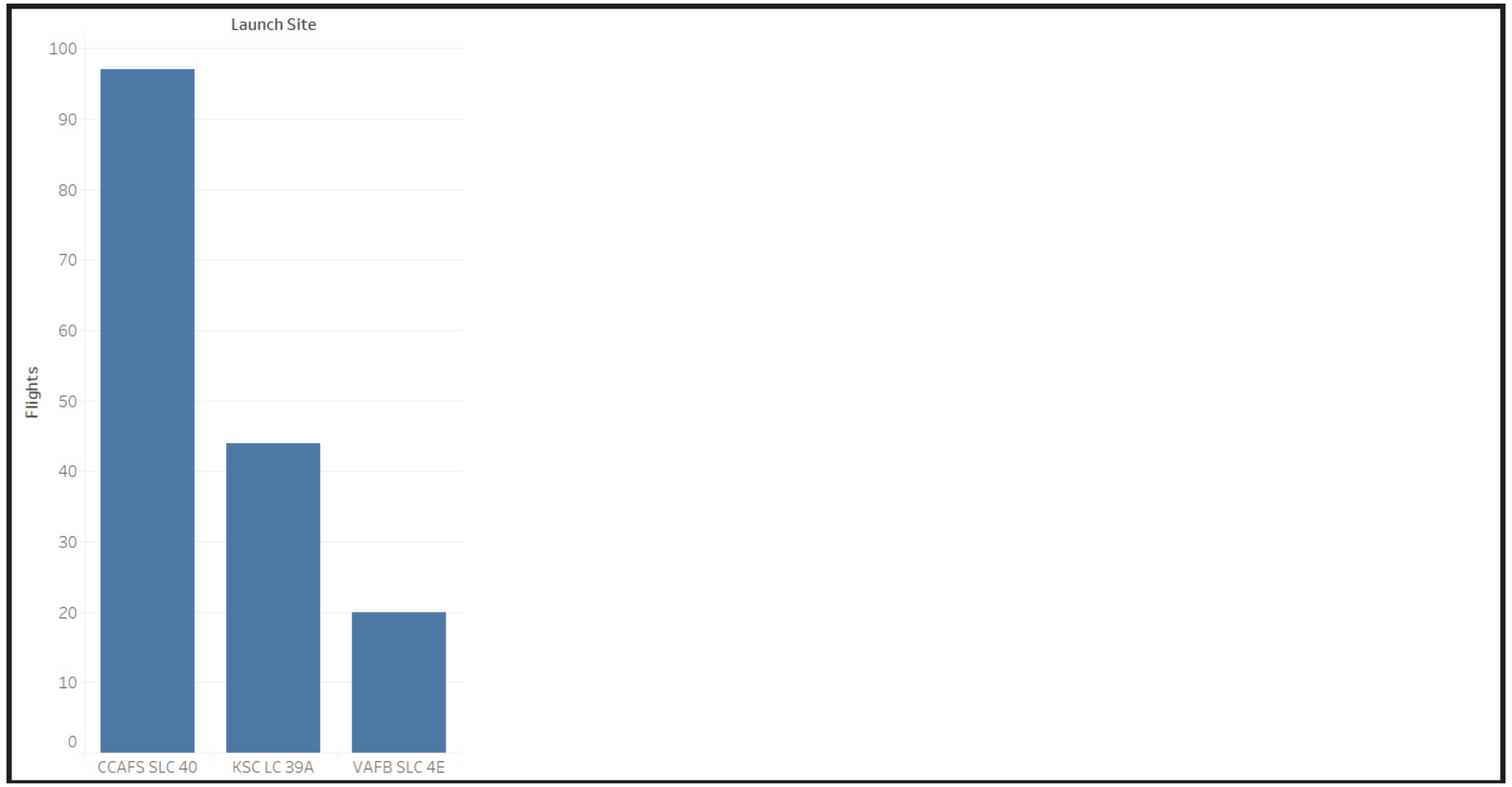
Results 3/4

Trends for successful reuse with payload size are also improving.



Results 4/4

Although the trained models show high accuracy of 83%, the also high amount of false positive predictions indicate imbalance in the sampling data. This is clearly visible as the number of flights from CCAFS is more than KSC and VAFB combined.



Conclusion

The overall trend for Falcon 9 booster reuse is strong and positive. In 2022 there are multiple boosters with more than 14 launches.

Although accuracy of Machine Learning models is high, the extreme bias in sample representation from one of the launch sites severely impacts their realibility for dependable predictions.

Finally, due to the successful implementation of booster reusability and the competitive advantages it provides, SpaceX is likely to continue its development and utilization.

Appendix

Sources:

- SpaceX REST API
- Wikipedia