Rocket Launch Sucess Rate Forecast

https://www.nasa.gov/directorates/spacetech/feature/NASA_Tech_One_Step_Closer_to_Launch_on_Next_Falcon_Heavy

Veronica Ferreiros Lopez
Stundet in the Capstone Module of data Analysis

Executive Summary

- Forecast of the cost of sucessful rocket launches for Space Y owned by Allan Musk
- Gathering data about Space X and creating dashboards as Eploraatory data Analysis
- Exploratory Data Analysis to infer patterns, insights and relevant parameters
- Using Various Numerical Algorithms to predict successful rocket launch and first stage landing
- Potential for the delveopment of tools to asess risk and feasibilty of space development projects

Introduction

- Explain the nature of the analysis
- States the problem
- States the questions that are to be answered by performing the analysis

Introduction

- Companies are making space development afordable
- The minum cost to beat is 62m dollars offerd by spaceX
- Much of the savings comes from weather the first stage can be re-used
  - Predicting weather the first stage will land sucessfully allows for forecasting cost of the rocket launches

Introduction

- Are lauch sites is close proximity to railways,highways, coastlines or cities?
- Which site has the largest successful launches?
- Which site has the highest launch success rate?
- Which payload range(s) has the highest launch success rate?
- Which payload range(s) has the lowest launch success rate?
- Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest successful lanuch rate?

Methodology

- Data Sources
- Data Filetring
- Data Wrangling
- Exploratory Data Analysis
- Interactive Visual Analytics

Methodology: Data Sources

- Data obtainied using a get request to the SpaceX API.

	FlightNumber	PayloadMass	Flights	Block	ReusedCount	Longitude	Latitude
count	94.000000	88.000000	94.000000	90.000000	94.000000	94.000000	94.000000
mean	54.202128	5919.165341	1.755319	3.500000	2.478723	-75.553302	28.581782
std	30.589048	4909.689575	1.197544	1.595288	3.082133	53.391880	4.639981
min	1.000000	20.000000	1.000000	1.000000	0.000000	-120.610829	9.047721
25%	28.250000	2406.250000	1.000000	2.000000	0.000000	-80.603956	28.561857
50%	52.500000	4414.000000	1.000000	4.000000	1.000000	-80.577366	28.561857
75%	81.500000	9543.750000	2.000000	5.000000	4.000000	-80.577366	28.608058
max	106.000000	15600.000000	6.000000	5.000000	9.000000	167.743129	34.632093

	Date	BoosterVersion	Orbit	LaunchSite	Outcome	LandingPad	Serial
count	94	94	94	94	94	64	94
unique	94	2	11	4	8	5	57
top	2018-07-25	Falcon 9	GTO	CCSFS SLC 40	True ASDS	5e9e3032383ecb6bb234e7ca	B1049
freq	1	90	27	55	41	35	6

Methodology: Data Filetring

- After that the data was cleaned to only include Falcon9 launches and reset the Flight Number column.
- Next I checked how many rows have missing values.
  - PayloadMass have 5 missing values that we replaced with the average payload.
- Checking how many columns are categorical

Methodology: Data Filetring

Checking the number of launches for each site

```
df['LaunchSite'].value_counts()
```

```
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
Name: LaunchSite, dtype: int64
```

```
df['LaunchSite'].value_counts(normalize=True)
```

```
CCAFS SLC 40    0.611111
KSC LC 39A      0.244444
VAFB SLC 4E     0.144444
Name: LaunchSite, dtype: float64
```

Methodology: Data Filetring

- Checking the number and occurrence of each orbit

```
# Apply value_counts on Orbit column
df['Orbit'].value_counts()
```

```
GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
SO        1
GEO       1
ES-L1     1
HEO       1
Name: Orbit, dtype: int64
```

```
df['Orbit'].value_counts(normalize=True)
```

```
GTO      0.300000
ISS      0.233333
VLEO     0.155556
PO       0.100000
LEO      0.077778
SSO      0.055556
MEO      0.033333
SO       0.011111
GEO      0.011111
ES-L1    0.011111
HEO      0.011111
Name: Orbit, dtype: float64
```

Methodology: Data Filetring

- I also checked the number of seccessful landings

```
df['Outcome'].value_counts()
```

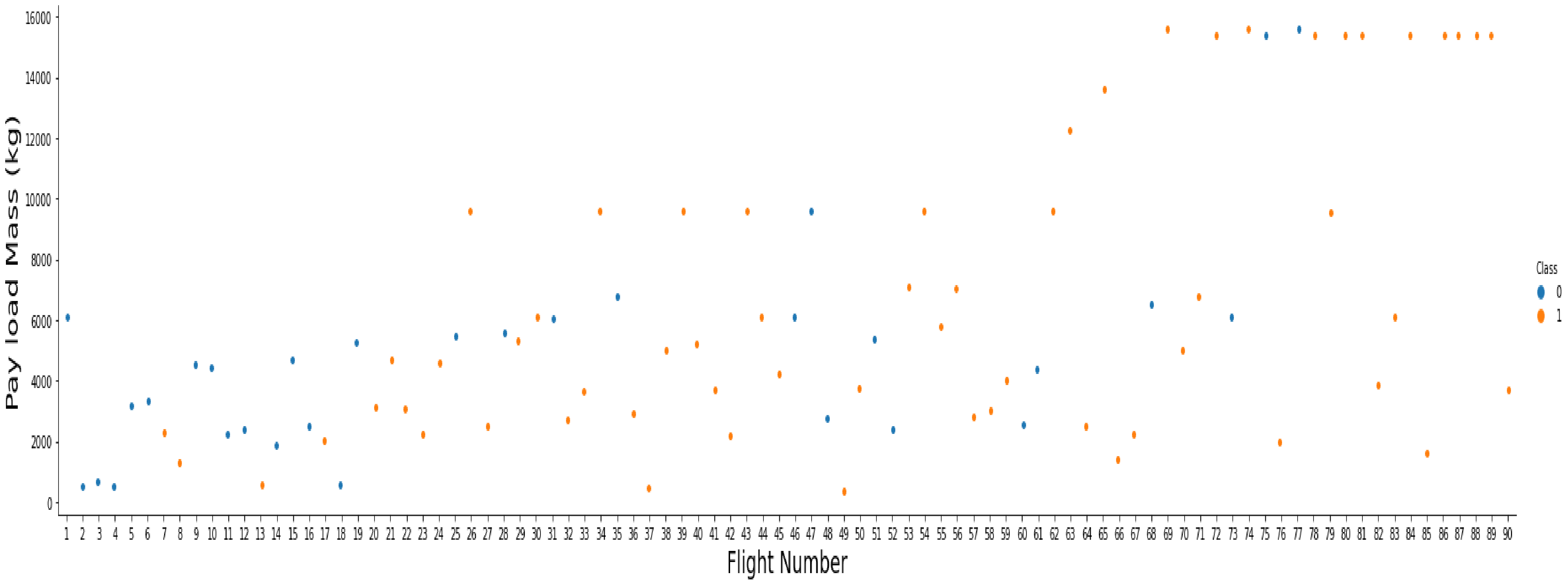
```
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
None ASDS       2
False Ocean     2
False RTLS      1
Name: Outcome, dtype: int64
```

Methodology: Data Wrangling

- I created an additional cloumn: “landing_class”
  - In this column, all successful landings have 1 as value
  - In this column, all unsuccessful landings have 0 as value

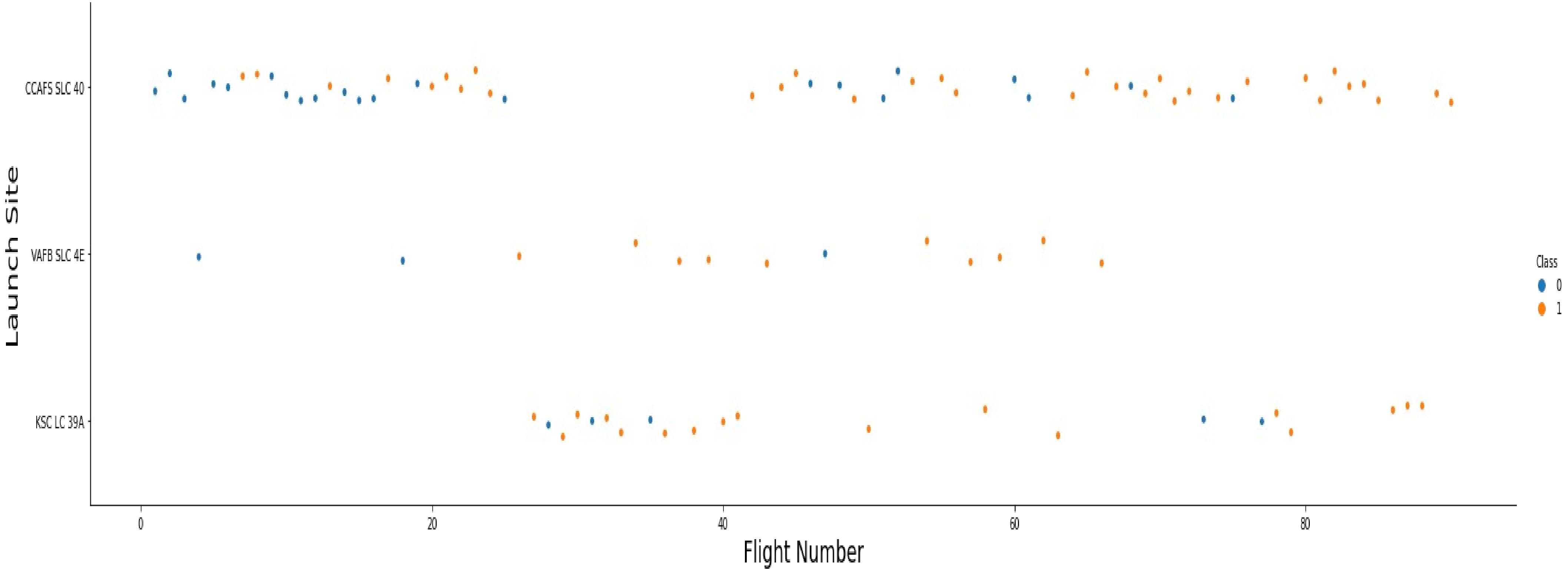
Methodology: Exploratory Data Analys

- Checking influence of FlightNumber and Payload variables.



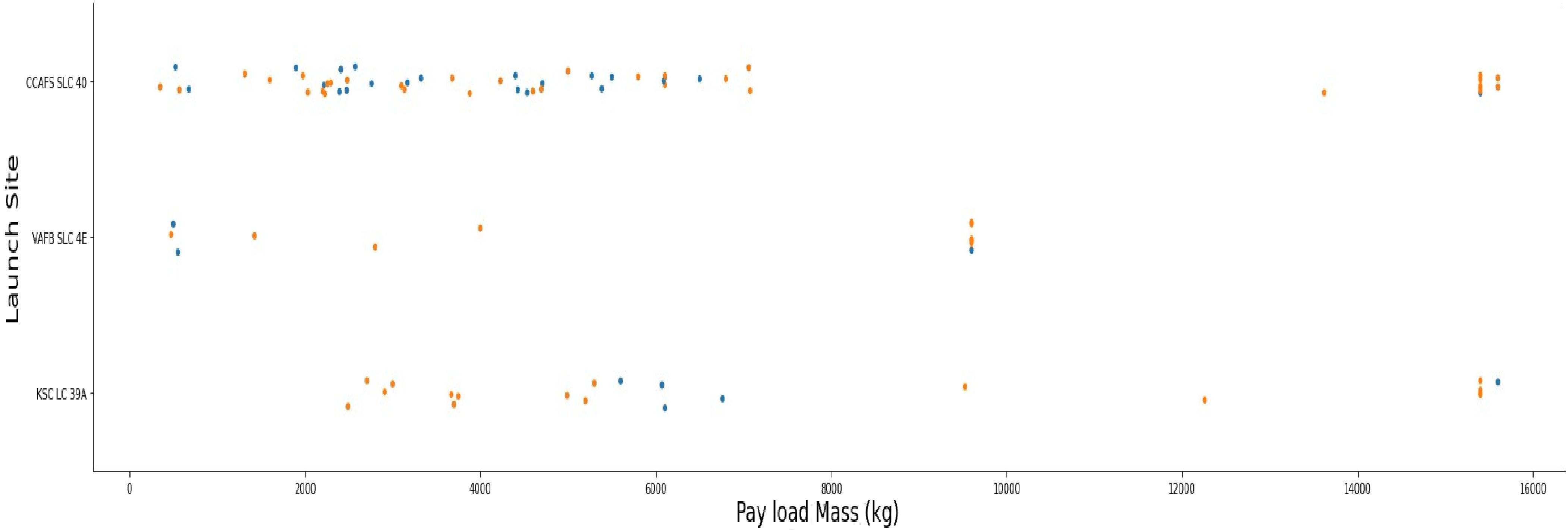
Methodology: Exploratory Data Analys

- Checking influence of PayloadMass and LaunchSite variables.



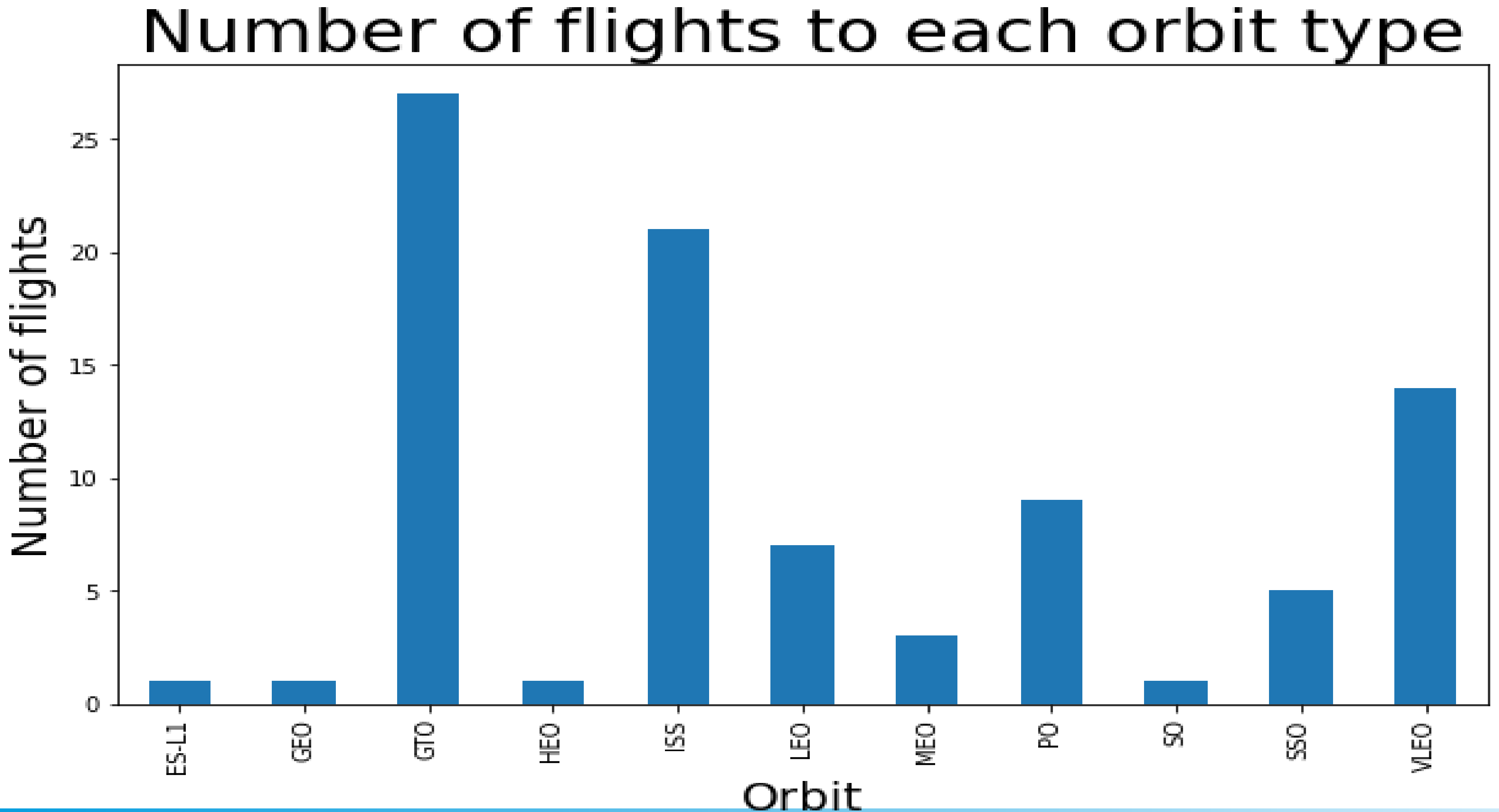
Methodology: Exploratory Data Analys

- Checking influence of PayloadMass and LaunchSite variables.



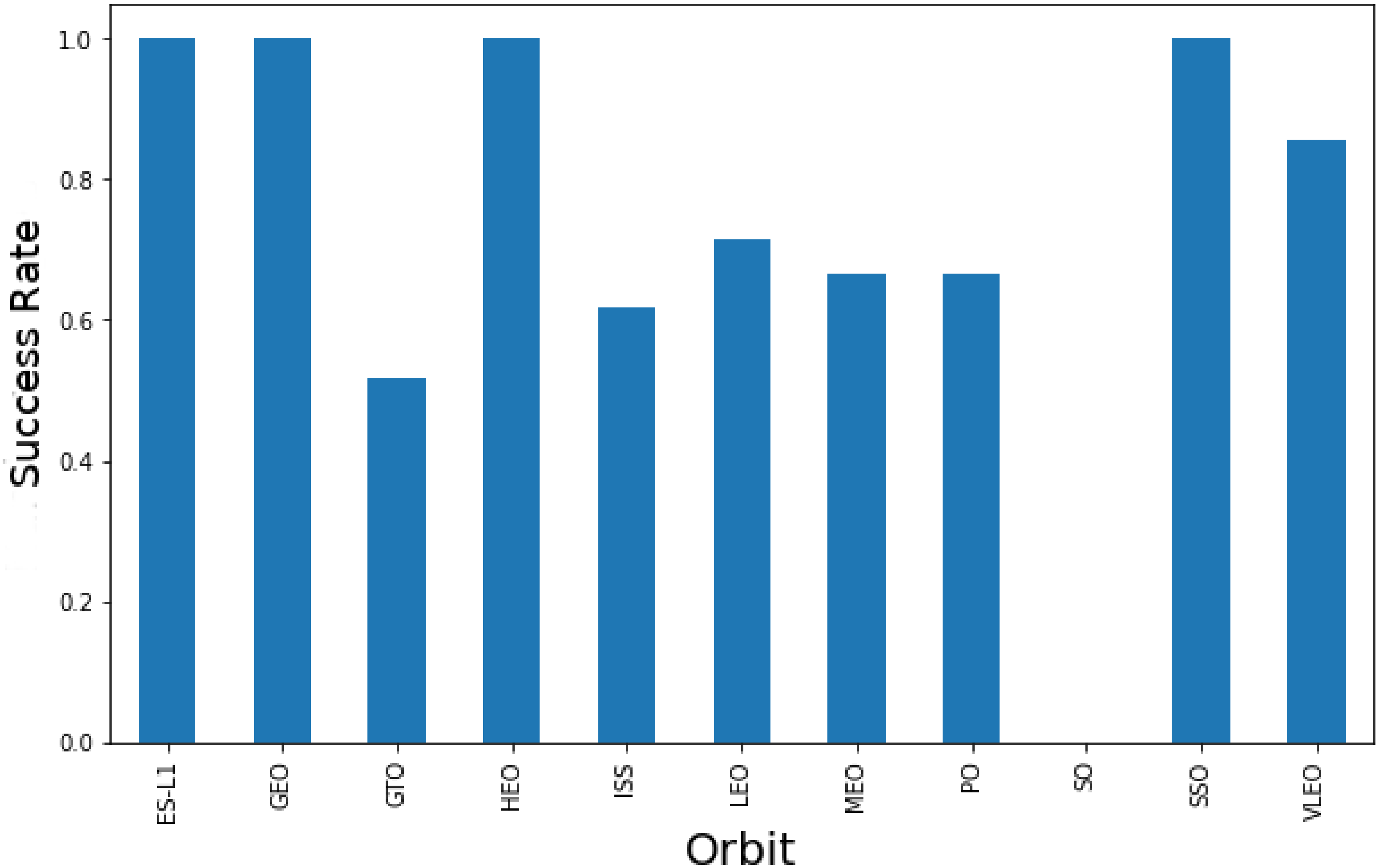
Methodology: Exploratory Data Analys

- Checking influence of Orbit and Successful Launches.



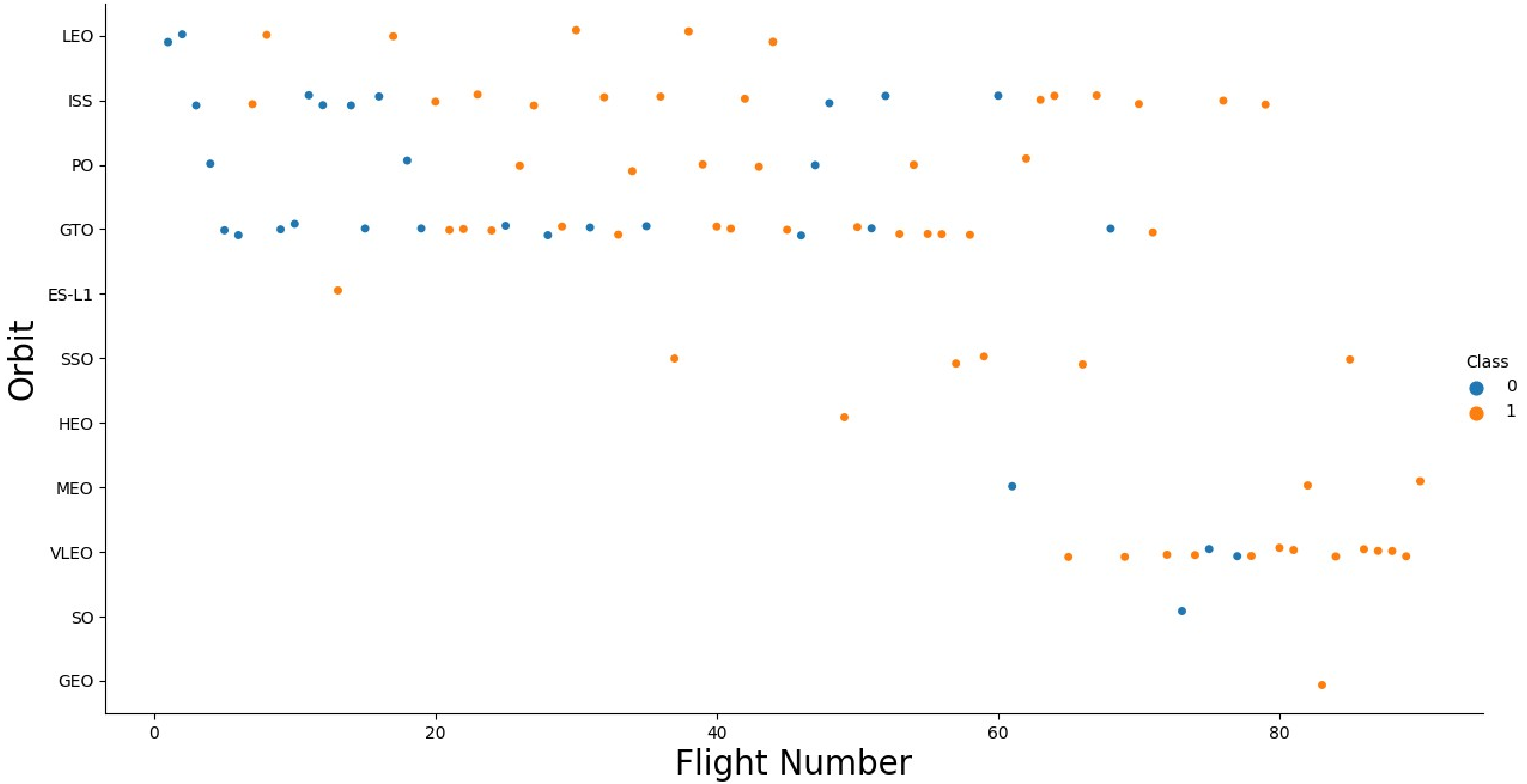
Methodology: Exploratory Data Analys

- Checking influence of Orbit and Success Rate.



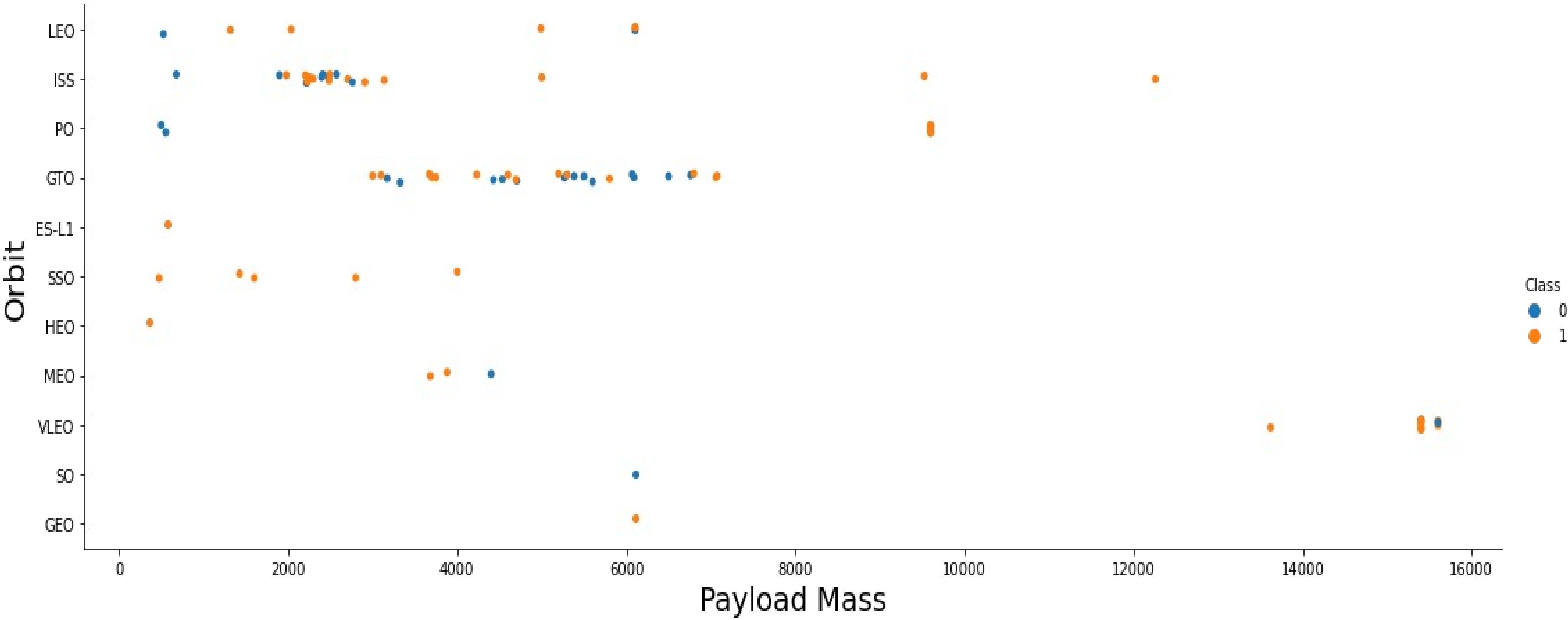
Methodology: Exploratory Data Analys

- Checking relationship between Orbit and Flight Number



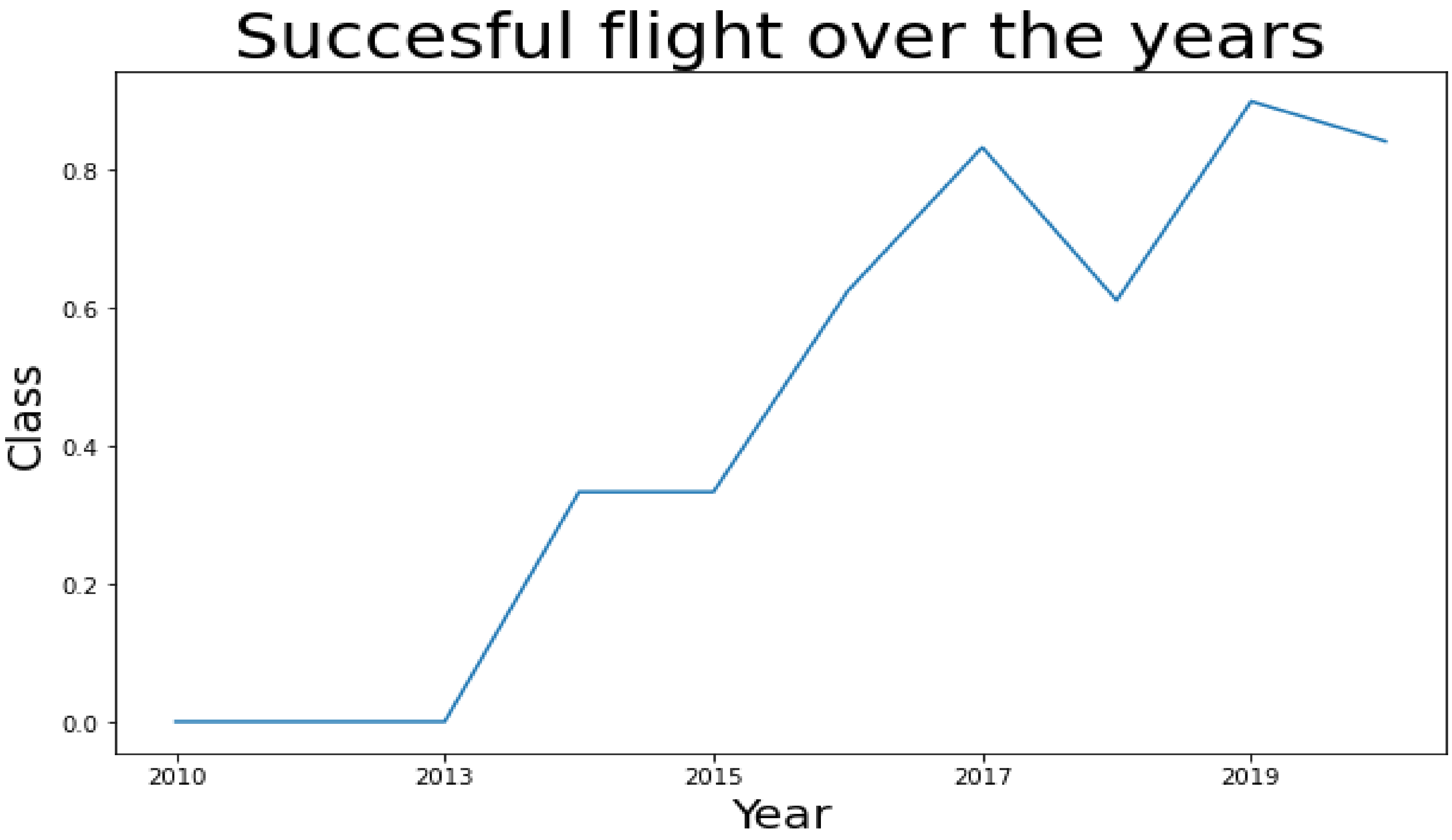
Methodology: Exploratory Data Analys

- Checking relationship between Orbit and PayloadMass.



Methodology: Exploratory Data Analys

- Success Rate over the years.



Methodology: Exploratory Data Analys

- We see that as the flight number increases the first stage is more likely to land successfully.
- The payload mass is also important. It seems that the more massive the payload, the less likely the palylad will return.
- We see that different launch sites have different success rates:
  - CCAFS LC-40 has a success rate of 60%
  - KSC LC-39 has a success rate of 77%
  - VAFB SLC 4Ehas a success rate of 77%

Methodology: Exploratory Data Analys

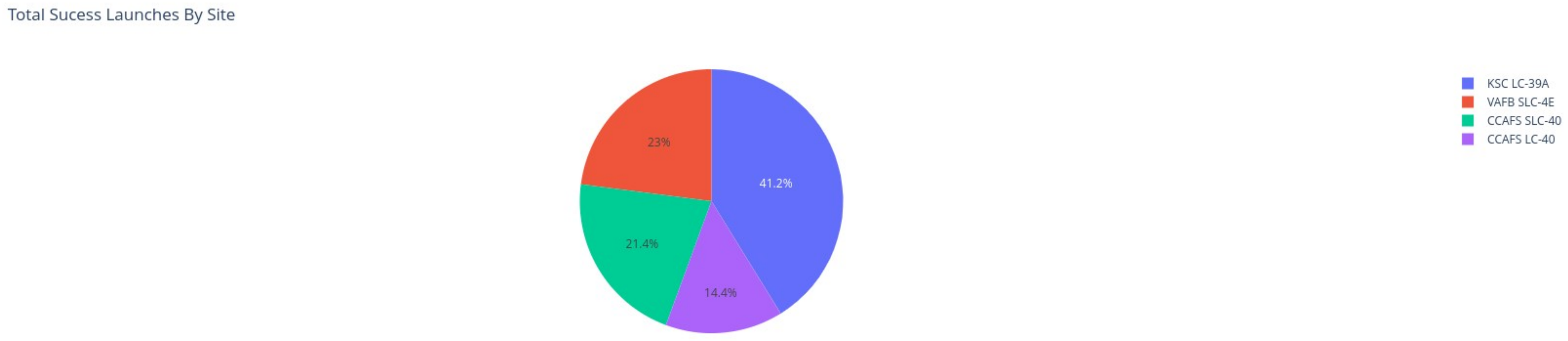
- In site CCA appears to be more lauches, probalby bacuse is more effecint if you want to achieve a certain orbit.
- The other 2 sites appear to be back up lauching pads of lauching pads for testing sinde the number of flights are more spread out, save for the case of site KSCLC, 4 beacuse many consecutive flights happened here between flight numbers 30 and 40 while none happened at site CCA, thus this my be a back p site

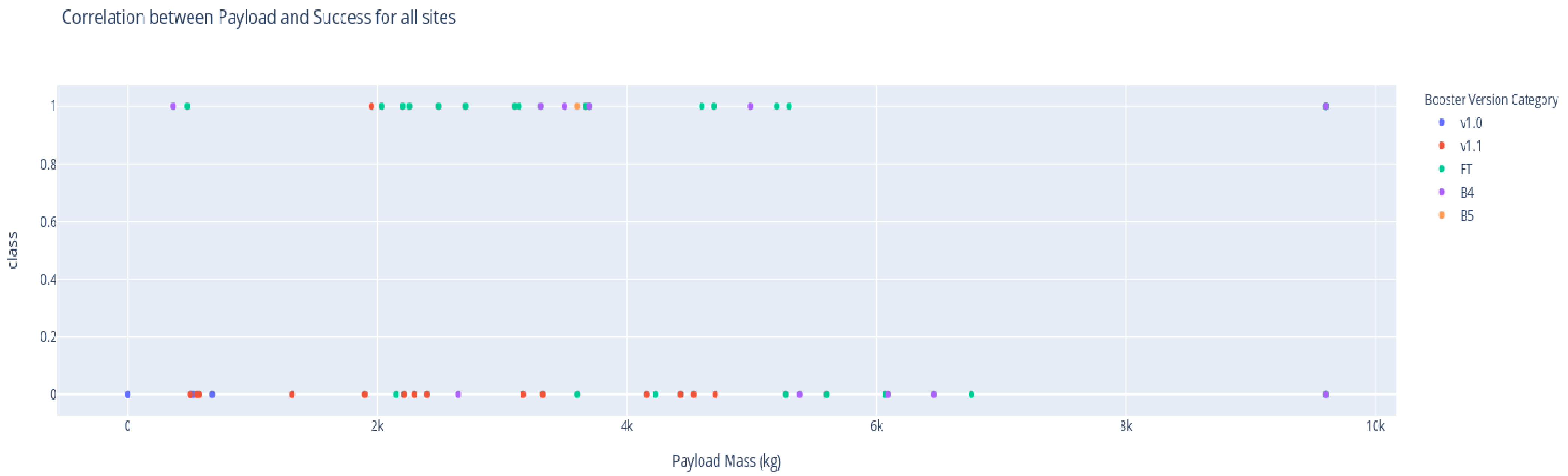
Methodology: Exploratory Data Analys

- The following orbits have 100% succes rates, but they only have veru few flights ES-L1, GEO, HEO, SSO
- VLEO is the only orbit with sucess rate of more than 80% with more than 10 flights

Methodology: Interactive Analytics

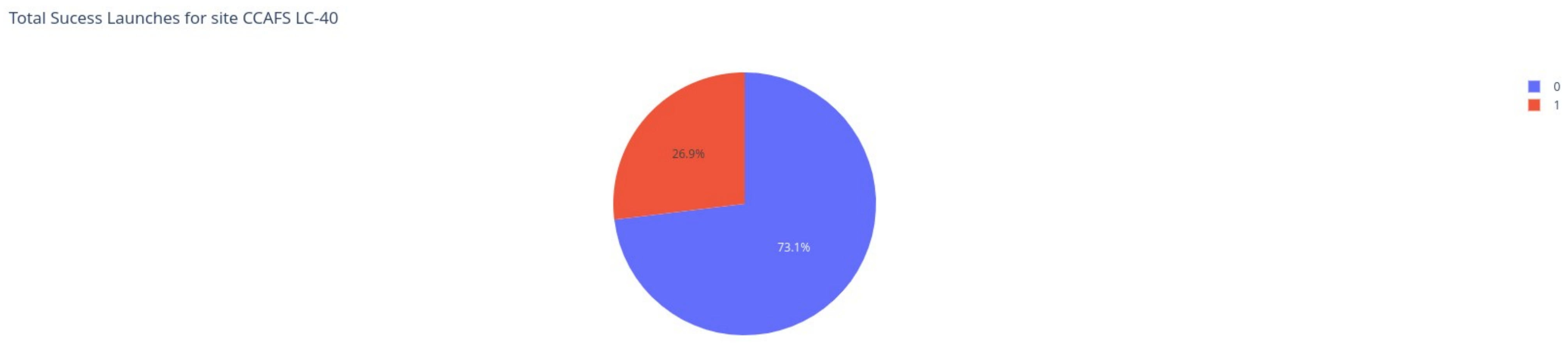
- Success rate for all sites





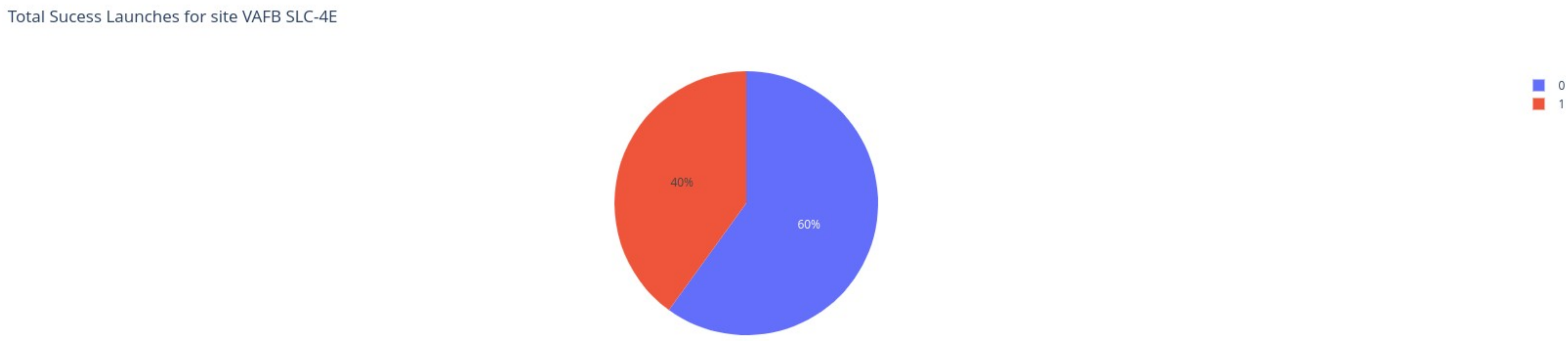
Methodology: Interactive Analytics

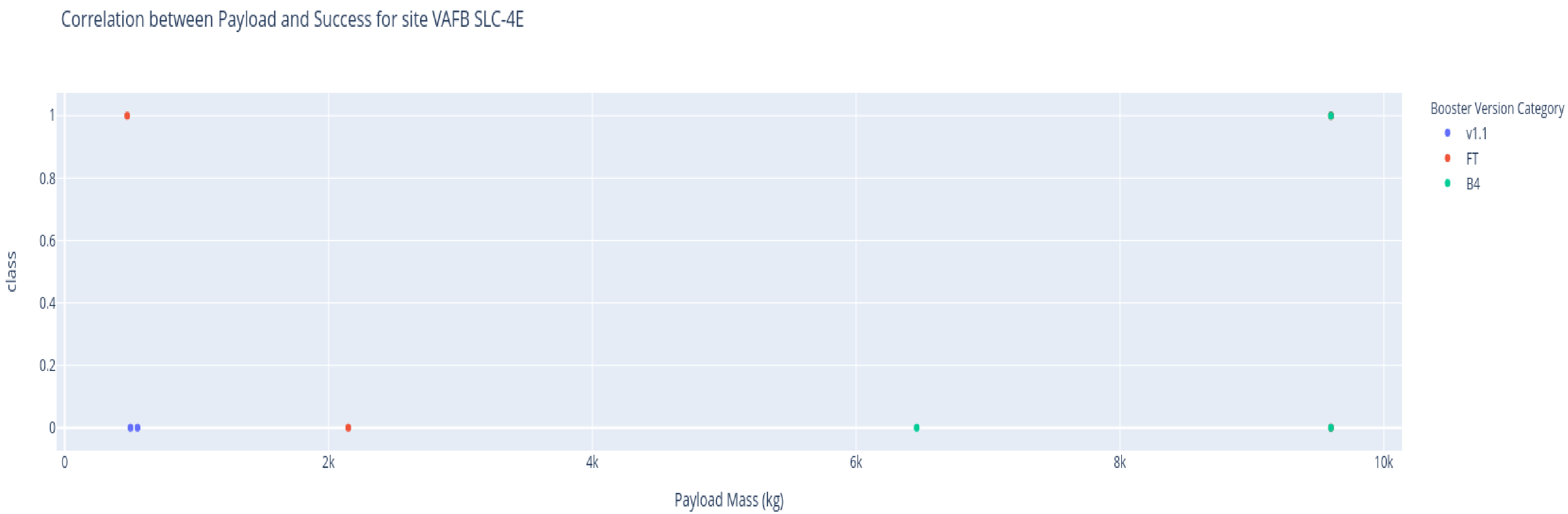
- Success rate for site CCAFS LC-40



Methodology: Interactive Analytics

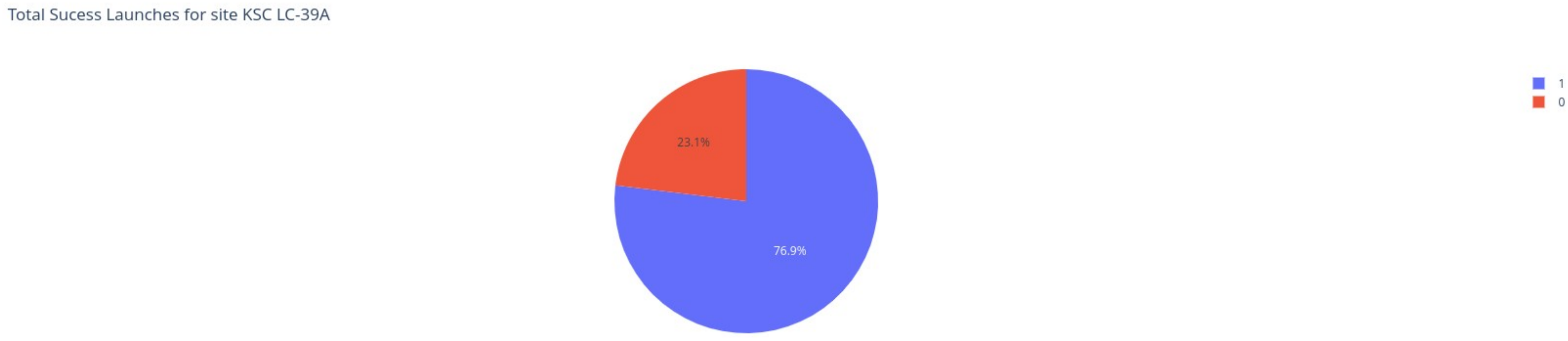
- Success rate for site VAFB SLC-4E

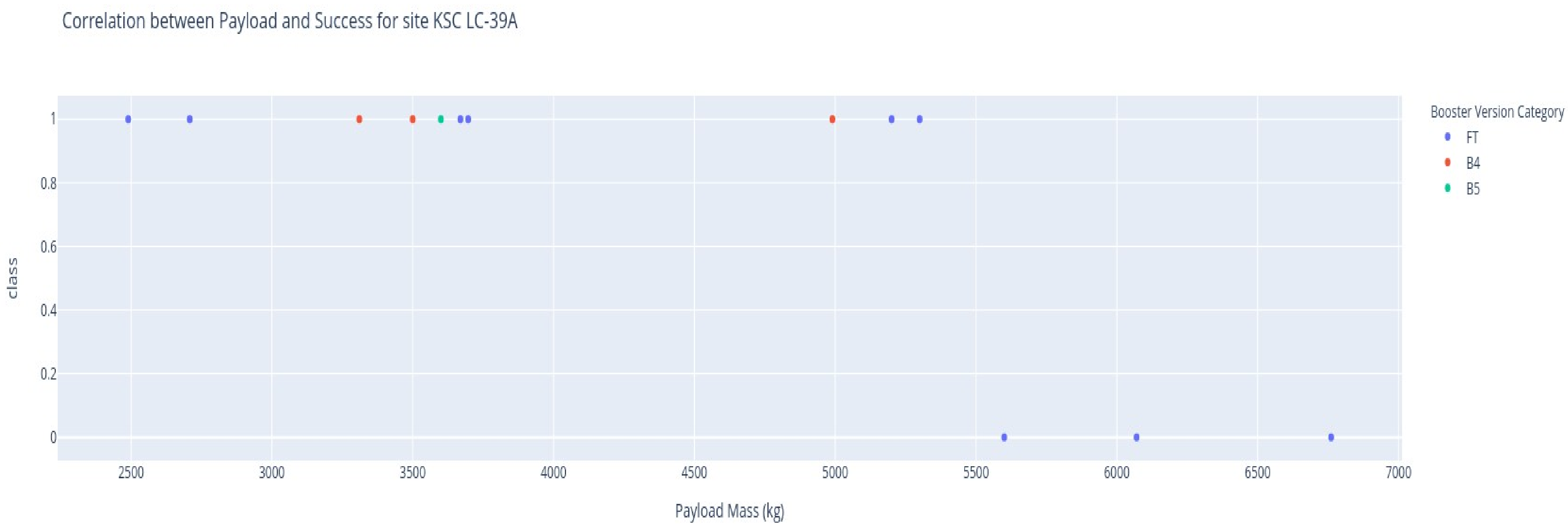




Methodology: Interactive Analytics

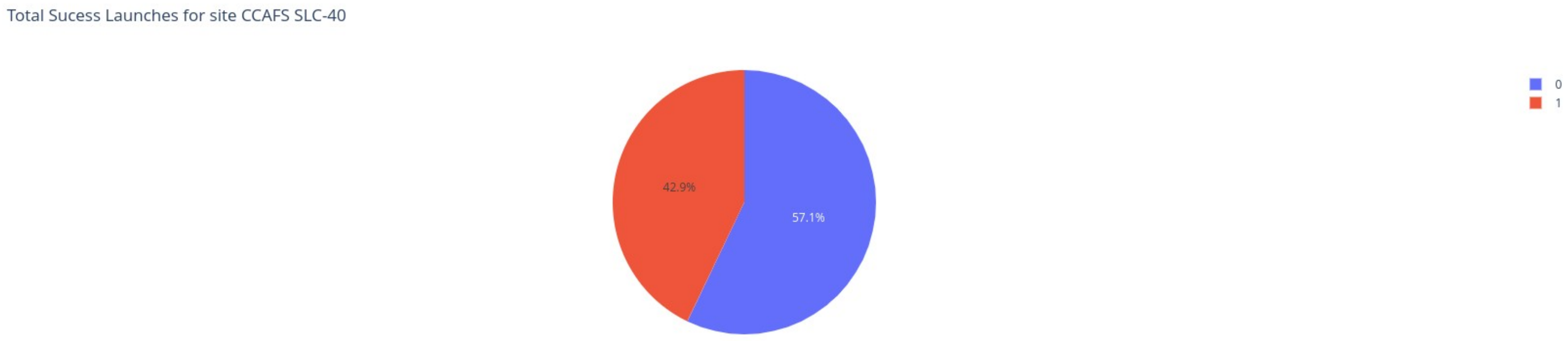
- Success rate for site KSC LC-39A

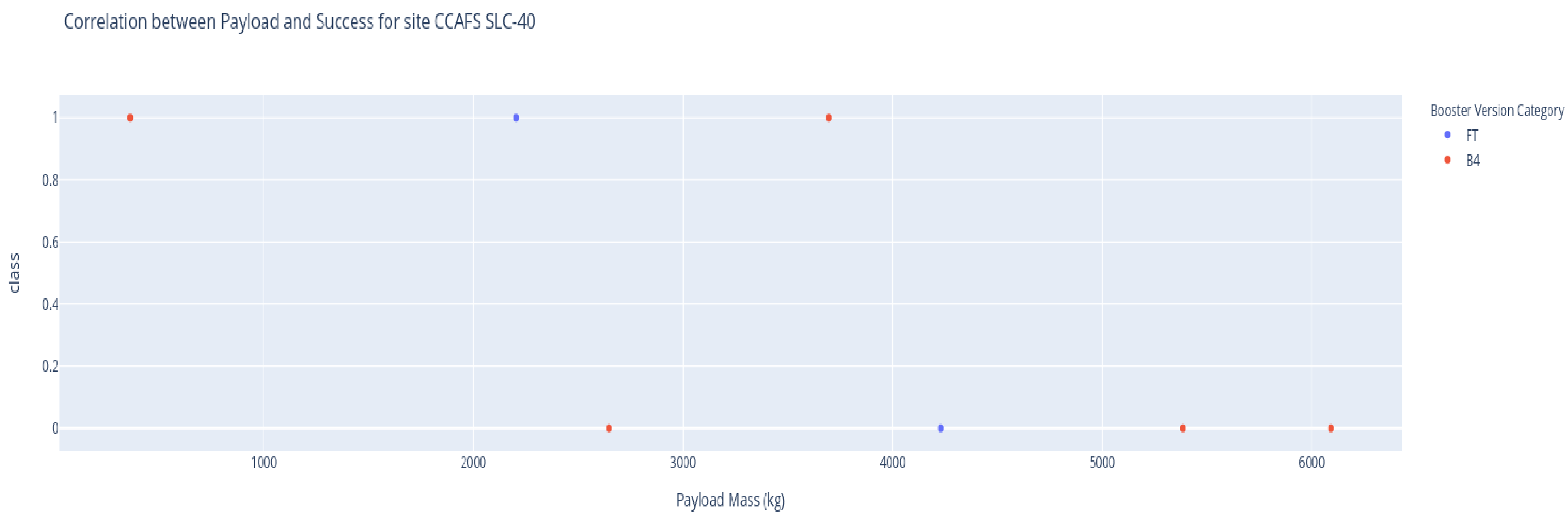




Methodology: Interactive Analytics

- Success rate for site CCAFS SLC-40





Methodology: Interactive Analytics

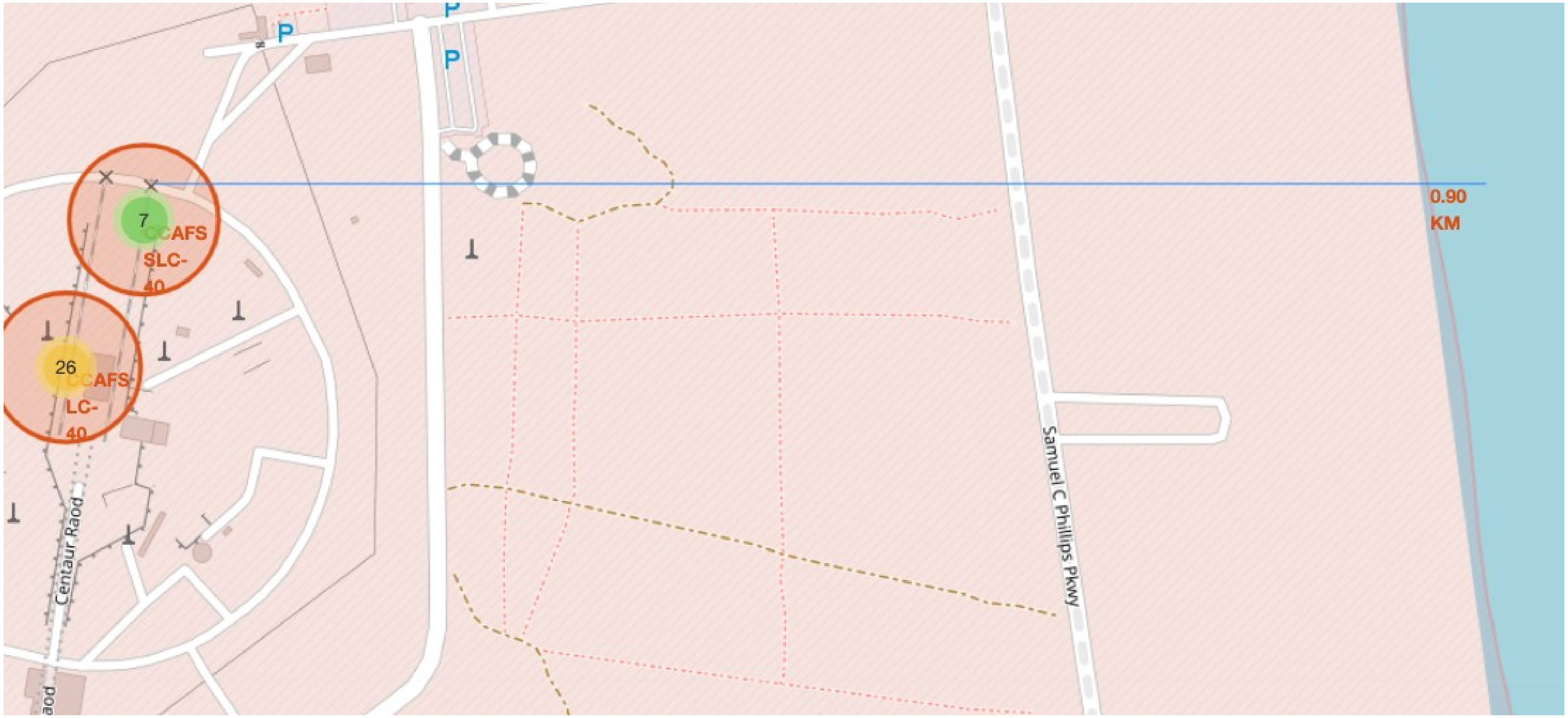
- **KSC LC-39A has the largest successful launches: 10**
- **KSC LC-39A has the highest launch success rate: 76%**
- **The payload ranges with the highest launch success rate are:**
  - **All sites: 2034 – 5300 kg**
  - **KSC LC-39A: 2490 - 5300 kg**
  - **CCAFS LC-40: 2205 - 3696 kg**

Methodology: Interactive Analytics

- **The payload range(s) with the lowest launch success rate are:**
  - **CCAFS: 3325 – 4535 kg**
  - **VAFB SLC-4E: 500 – 9600kg**
  - **KSC LC-39A: 5500 – 6791 kg**
  - **CCAFS LC-40: 4230 – 6092 kg**
- **The F9 Booster versions with the highest launch success rate are:**
  - **F9 B4 with a 100% rate and 1 successful launch**
  - **F9 FT with a 66% rate and 16 successful launches**

Methodology: Interactive Analytics

- **Launch sites are far from cities and close to coaslines that allows for wasy transportaion on rocket boosters.**



Results

- Explain how the data was organized
- Explain how the data was analyzed
- Charts and grapsh to substantiate the results
- Interpretation of data:
  - Call attention to more complex or crucial findings
  - Detail explanation to the audience
  - Convey the answers to the problem stated in the introduction

Results: data organizacion

- Forecasting successful launches with algorithms
- Extraction of Features
  - Numeric Features: FlightNumber,PayloadMass,Flights,Gridfins,Reused,Legs,Block and Resused Count.
  - The oneHotEncoding was appieId to the categorical data: Orbits, LaunchSite, LandingPad and Serial
  - Cheking for null values
- Data standarization
  - Casting data as float64

Results: Data Analysis

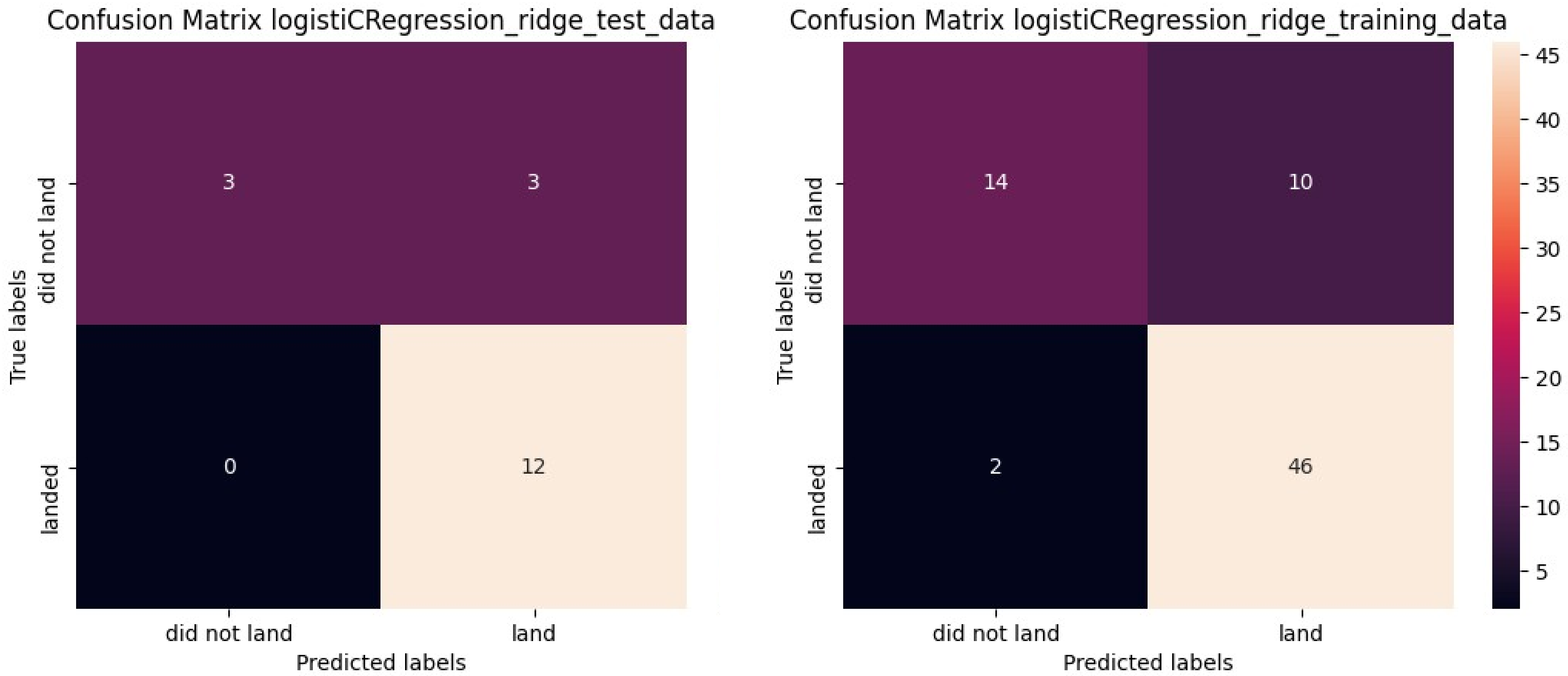
- Split into training data and test data
  - test_size=0.2 & random_state=0
- Find best Hyperparameter for:
  - Classification Trees
  - Clustering
  - Logistic Regression
  - Support Vector Machine
- Find the method performs best using test data.
  - Criteria: Least square error, Mean square error & Confussio Matrix

Results: Graphs for Logistic Regression

- Best Hiperparameters
  - {'C': 1, 'penalty': 'l2', 'solver': 'lbfgs'}
  - Accuracy with testing data : 0,82
- Training Data
  - MSE: 0,78
  - Accuracy, R^2: 0.83
- Testing Data
  - MSE: 0,83
  - Accuracy, R^2: 0.83
  - Succesful Prediction Rate: 83%

Results: Graphs for Logistic Regression

- Examining the confusion matrix, we see that the major problem is false positives

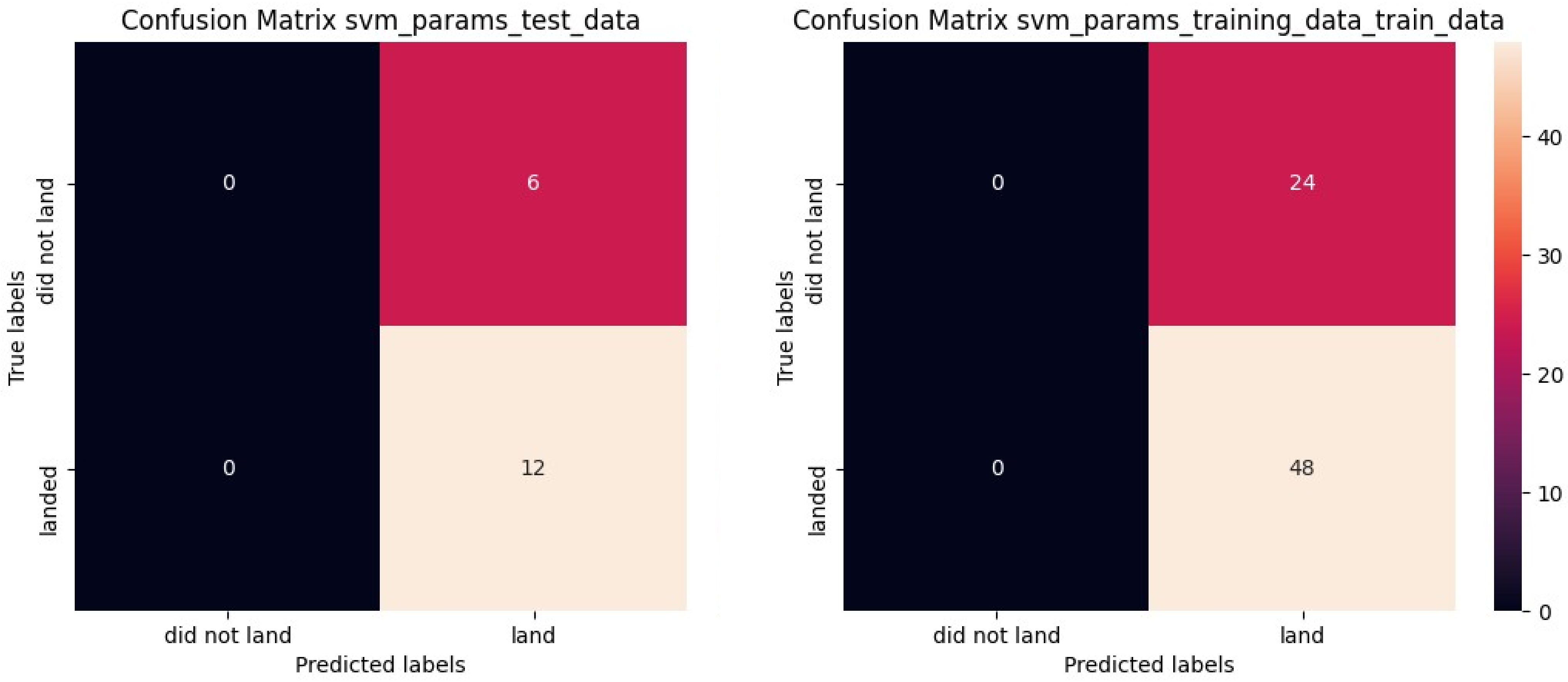


Results: Support Vector Machine

- Best Hiperparameters
  - {'C': 1.0, 'gamma': 0.001, 'kernel': 'rbf'}
  - Accuracy with testing data : 0,66
- Training Data
  - MSE: 0,33
  - Accuracy, R^2: 0,66
- Testing Data
  - MSE: 0,33
  - Accuracy, R^2: 0,66
  - Succesful Prediction Rate: 66%

Results: Support Vector Machine

- This is the worst results obtained so far.

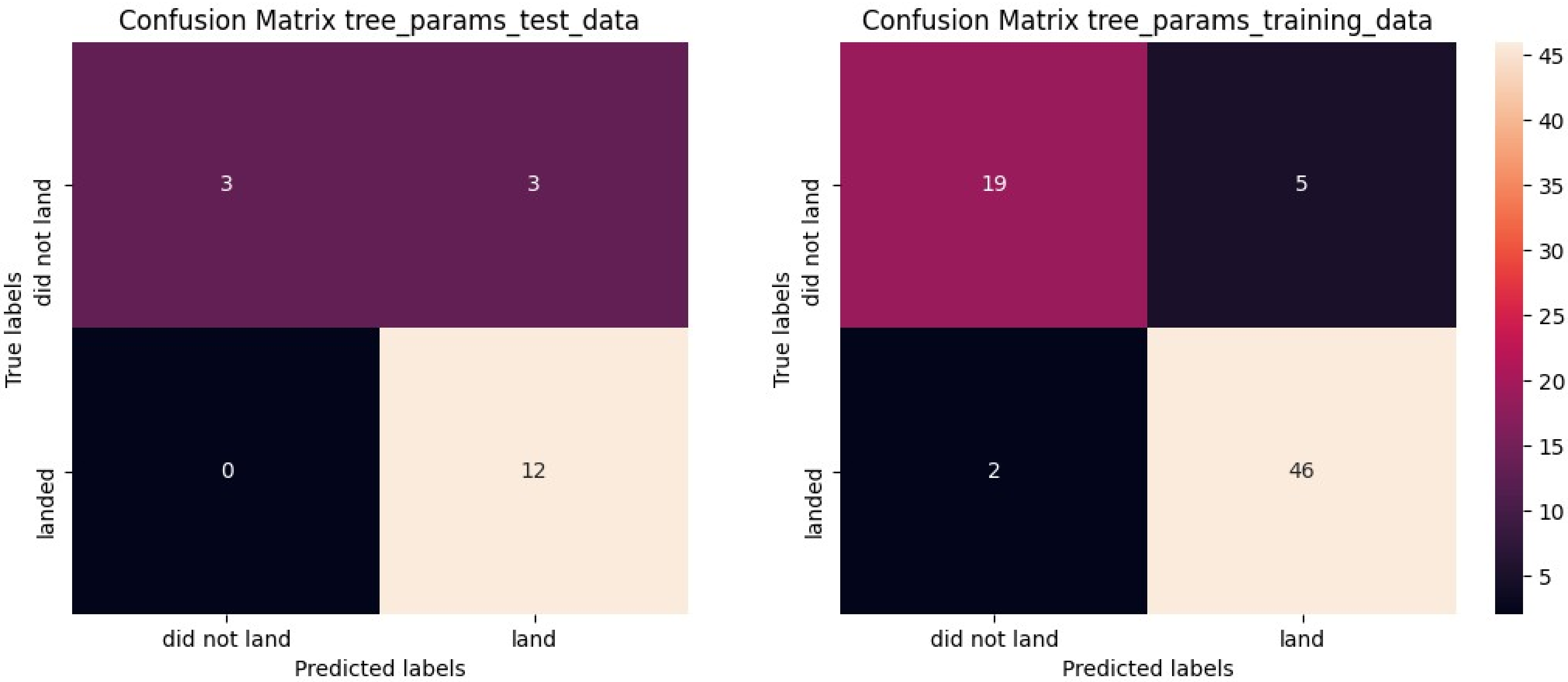


Results: Classification Trees

- Best Hiperparameters
  - {'criterion': 'gini', 'max_depth': 8, 'max_features': 'auto', 'min_samples_leaf': 1, 'min_samples_split': 5, 'splitter': 'random'}.
  - Accuracy with testing data : 0,87
- Training Data
  - MSE: 0,10
  - Accuracy, R^2: 0,90
- Testing Data
  - MSE: 0,17
  - Accuracy, R^2: 0.90
  - Succesful Prediction Rate: 83%

Results: Classification Trees

- Issues with false positives and negatives

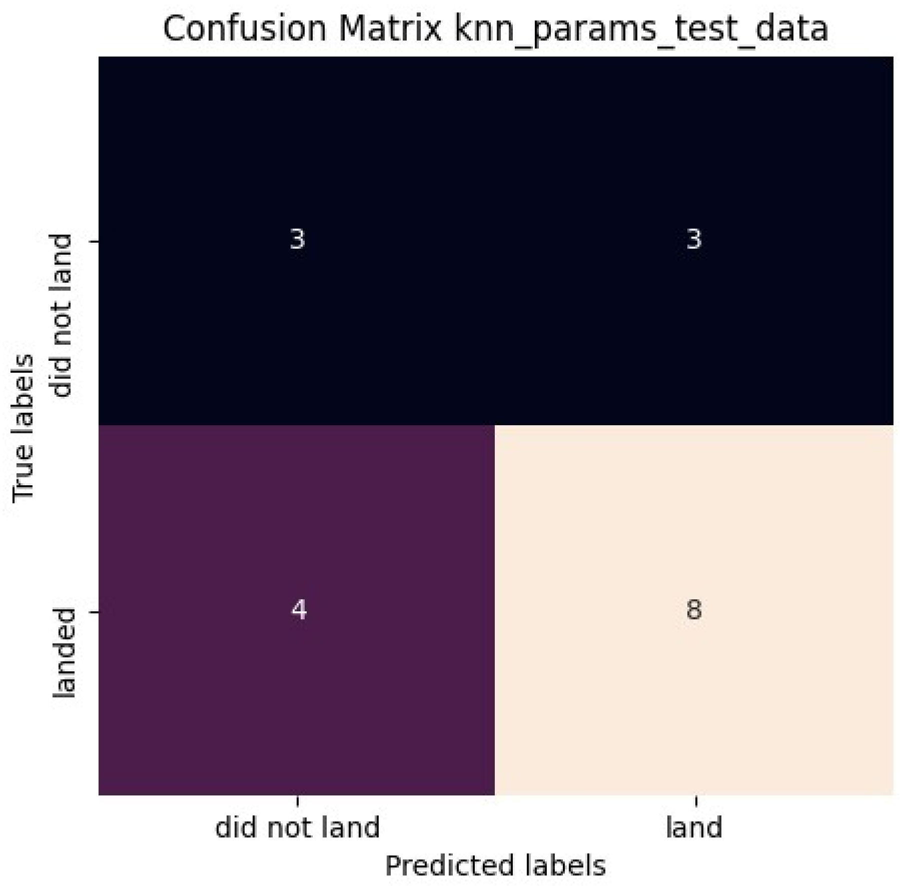


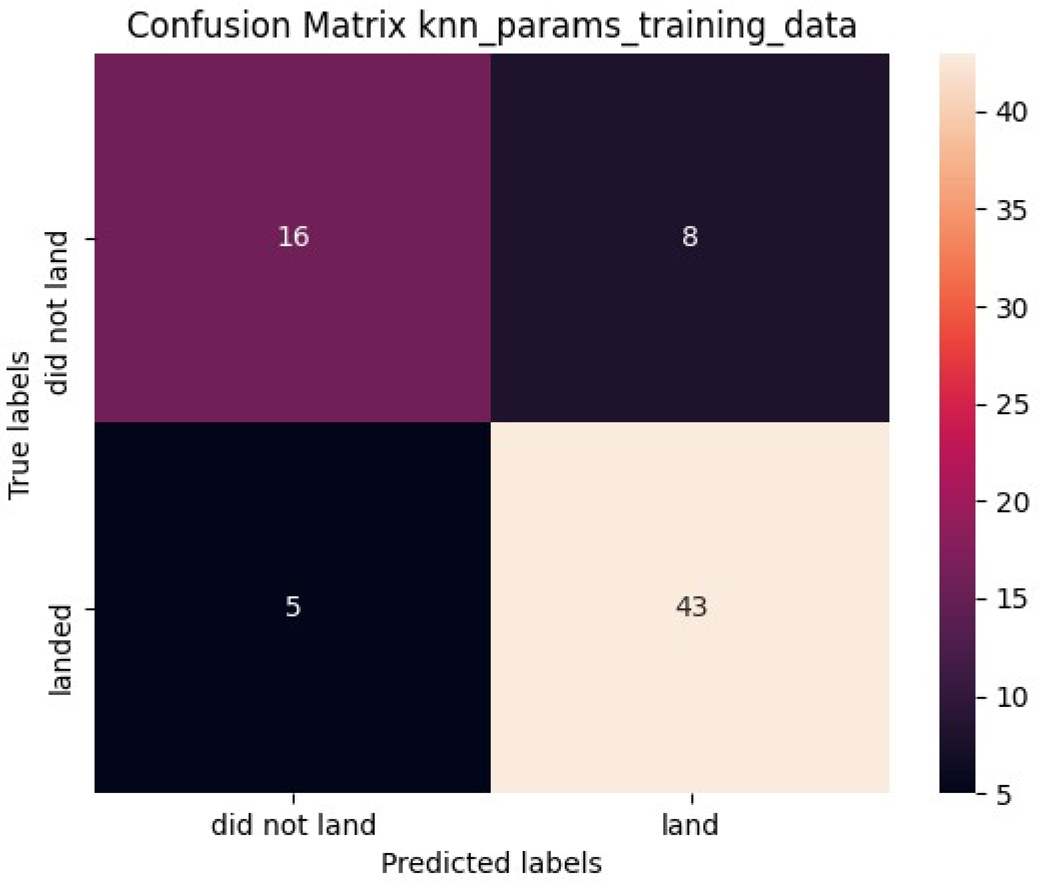
Results: K-Nearest Neighbors

- Best Hiperparameters
  - {'algorithm': 'auto', 'n_neighbors': 3, 'p': 1}
  - Accuracy with testing data : 0,66
- Training Data
  - MSE: 0,18
  - Accuracy, R^2: 0.61
- Testing Data
  - MSE: 0,39
  - Accuracy, R^2: 0.61
  - Succesful Prediction Rate: 61%

Results: K-Nearest Neighbors

- Issues with false positives and negatives





Results: Interpretation

- The best prediction model were the Logistic Regression and Classification Trees.
- We can predict a successful landing if we have data about: Orbit, Payload, Flight Number, Lauch Site, Fins, Grid and Landing Pad.
  - Call attention to more complex or crucial findings
  - Detail explanatio to the audience
  - Convey the answers to the problem stated in the introduction

Discussion of Findings and Implications

- We can predict a successful landing if we have data about: Orbit, Payload, Flight Number, Lauch Site, Fins, Grid and Landing Pad.
- Thus if we can guarantee a successful landing it is feasible to assume that the first stage can be re-used.
  - This alloes for forecasting that the cost will be around 62 million dollars.

Conclusion

- Findings:
  - We can predict the successfull landing of the first stage using Logistic Regression and Suppot Vector Machines.
  - We have a tool to predict if the First Satge will be re-used.
  - This tool has the potential to help forecasting economic feasibility of space development projects
- Further work:
  - Impact of thruster to imporve manuvrability

Annex: url adress for the Capstone labs

- Week 1:
  - https://github.com/ferreir3/Coursera_Capstone/blob/main/week1_capstone_w1_spcaeX_api_dataSampling_filtering_webScrapping_beautifulSoup.ipynb
  - https://github.com/ferreir3/Coursera_Capstone/blob/main/week1_capstone_w1_dataerangling.ipynb
  - https://github.com/ferreir3/Coursera_Capstone/blob/main/week1_capstone_w1_spcaeX_api_dataSampling_filtering.ipynb
- Week 2:
  - https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/7de51eba-1f77-486d-b0f1-f3682d894a9e/view?projectid=7f1b7d4b-1d6a-42ad-878e-b0b063b7aea8&context=cpdaas
- Week 3:
  - https://github.com/ferreir3/Coursera_Capstone/blob/main/week3_Capstone_w3_findingPatterns_folium.ipynb
  - https://github.com/ferreir3/Coursera_Capstone/blob/main/week3_spacex_dash_app.py
- Week 4:
  - https://github.com/ferreir3/Coursera_Capstone/blob/main/week4_SpaceX_Machine%20Learning%20Prediction_Part_5_completed.ipynb