Winning Space Race with Data Science

Silja George
11/08/2022

Outline

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

Executive Summary

Methodologies

- Data Collecting and Wrangling
- Data Analysis with SQL and Python
- Visualization
- Machine Learning Algorithms

Results

- Success Rate Prediction
- Visualize the Relationship between Different Attributions
- Launch Site Location Visualization
- Create Model for Landing Prediciton

Introduction

- SpaceX
- Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.
- Problems
- Attributions to launch success rate
- To predict if SpaceX will reuse the first stage

Section 1

Methodology

Methodology

Executive Summary

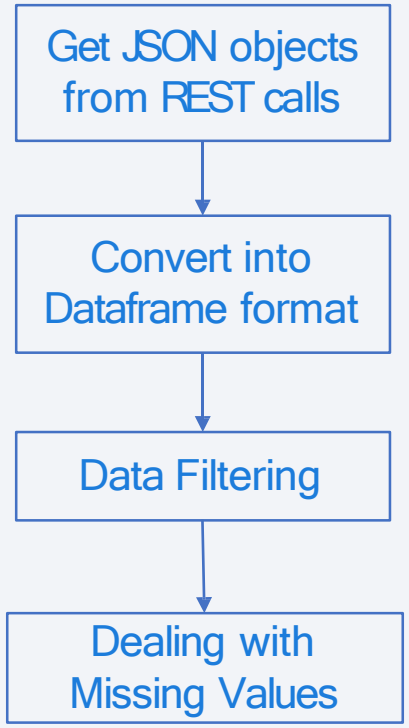
- Data collection methodology:
  - Request to the SpaceX API
- Perform data wrangling
  - Exploratory Data Analysis
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Find best Hyperparameter for SVM, Classification Trees and Logistic Regression

Data Collection

- Start requesting rocket launch data from SpaceX API with the following URL:

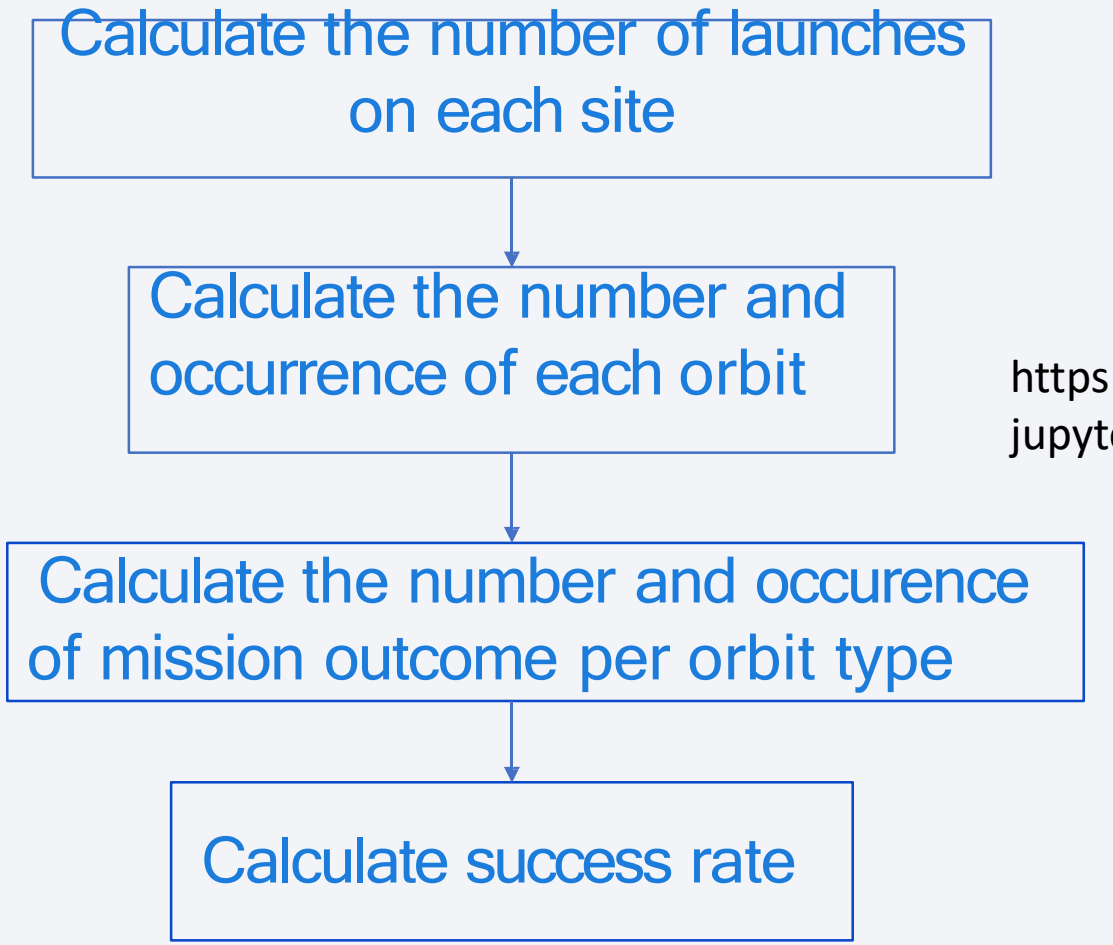
```
In [7]: spacex_url="https://api.spacexdata.com/v4/launches/past"

In [17]: response = requests.get(spacex_url)
```

Data Collection - SpaceX API



https://github.com/Silja-GeorgePR/SpaceX/blob/main/jupyter-labs-spacex-data-collection-api_1.ipynb

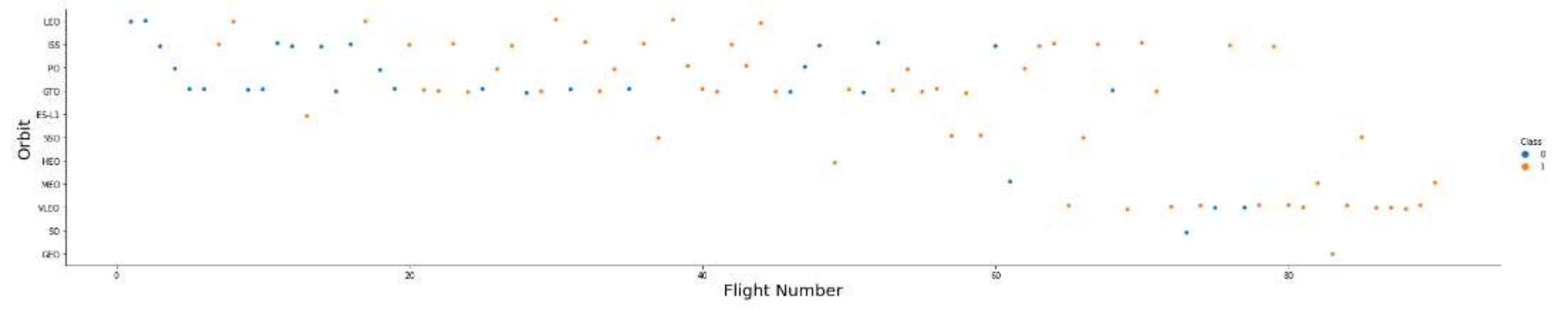
Data Collection - Wrangling



https://github.com/Silja-GeorgePR/SpaceX/blob/main/labs-jupyter-spacex-Data%20wrangling%20(1).ipynb

EDA with Data Visualization

- Exploratory Data Analysis
- Find Relationship between different elements through visualization



https://github.com/Silja-GeorgePR/SpaceX/blob/main/jupyter-labs-eda-dataviz%20(1).ipynb

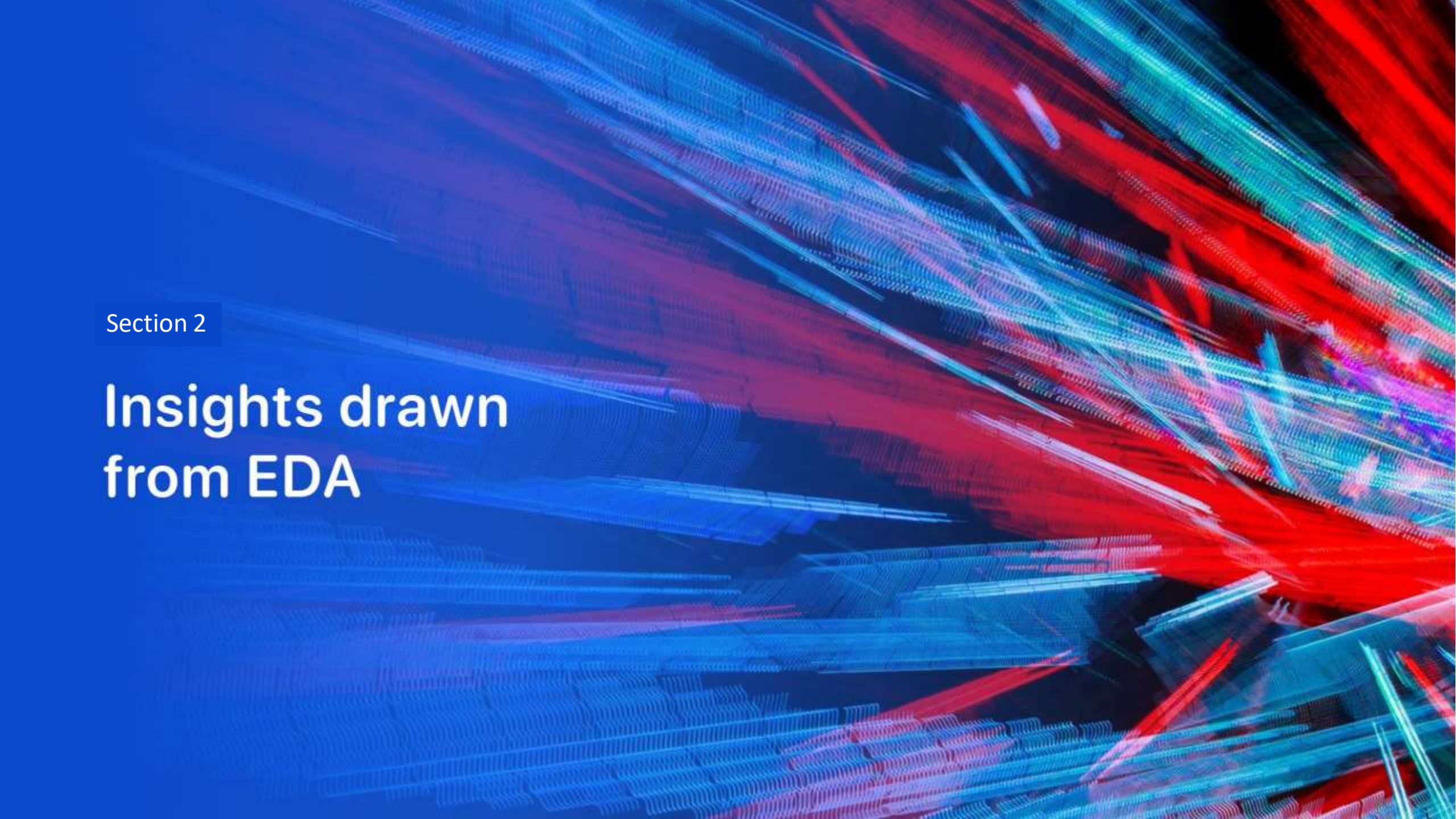
Predictive Analysis (Classification)

- Several kinds of Algorithms employed: KNN, Decision Tree, Logistics Regression And SVM.
- GridSearchCV is used to find out the best hyperparameters of the mentioned models.

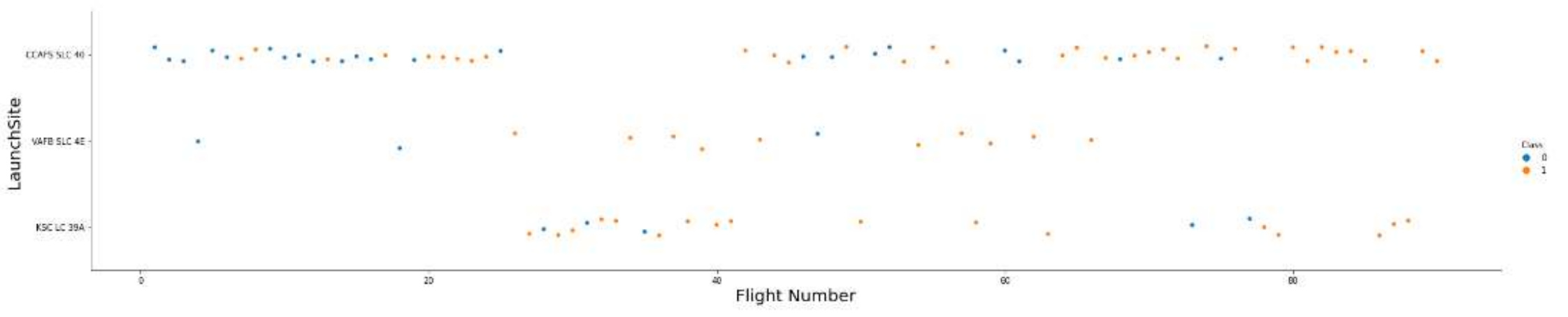
https://github.com/Silja-GeorgePR/SpaceX/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Section 2

Insights drawn from EDA



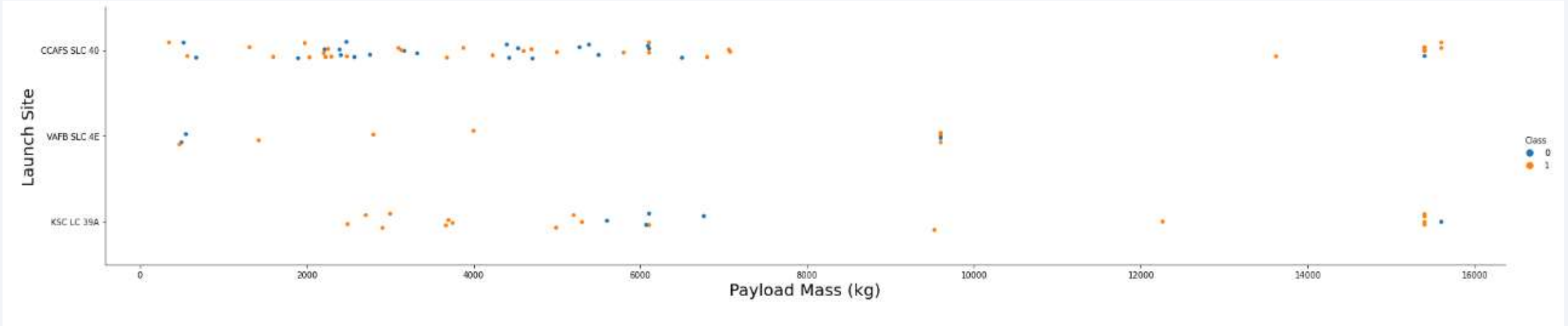
Flight Number vs. Launch Site



It is clearly shown in the scatter plot that most launches took place at CCAFS SLC 40, while the least took place at VAFB SLC 4E.

More, we can discover that Launch Site KSC LC 39A has the highest success rate.

Payload vs. Launch Site

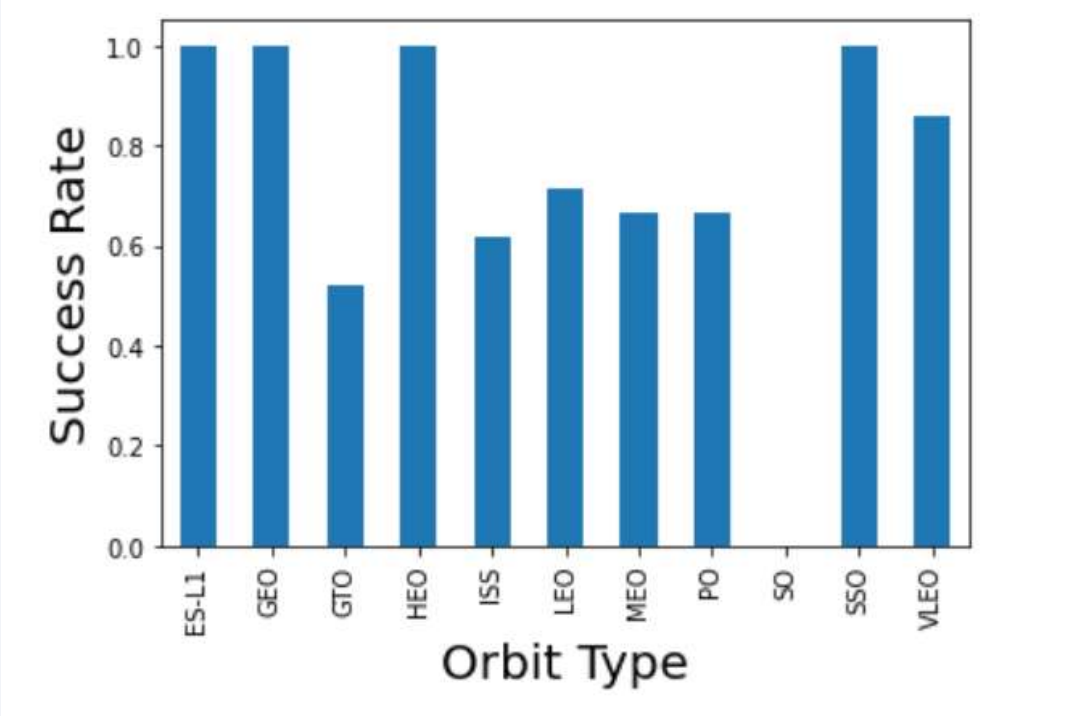


Payloads that approach MAX(Payload) tended to launch from CCAFS SLC 40 & KSC LC 39A

Payloads less than 8000kg tended to fail at a higher risk when launched from CCAFS SLC

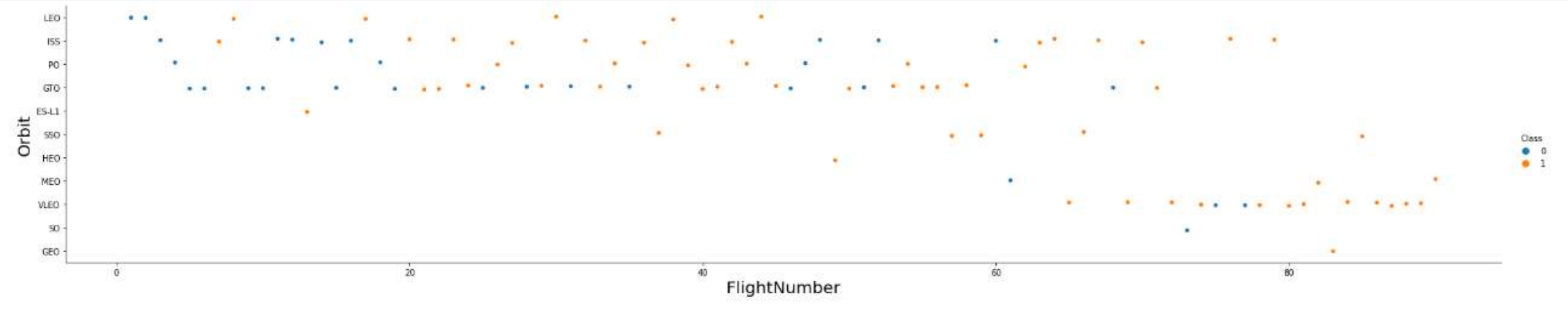
Sites used with less failure-tolerant payloads

Success Rate vs. Orbit Type



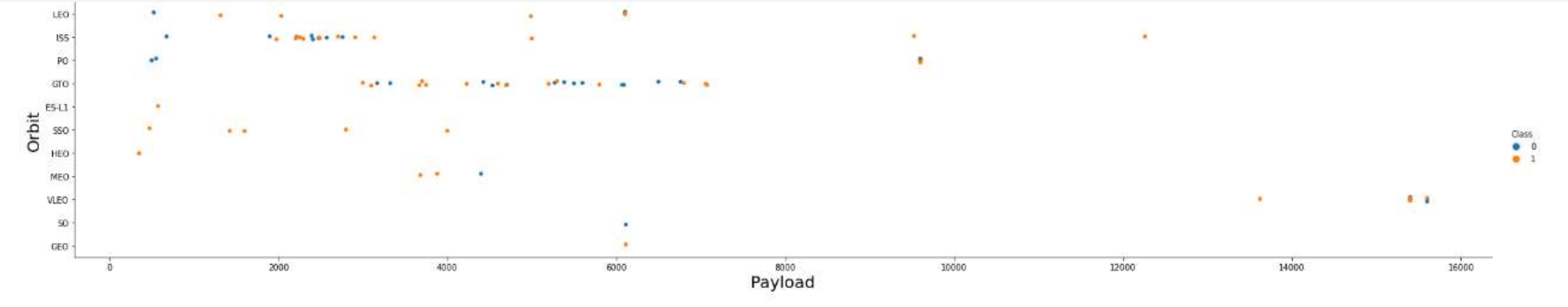
Plain to see that Orbit Type GTO had the lowest Success rate except SO which has never been successfully launched

Flight Number vs. Orbit Type



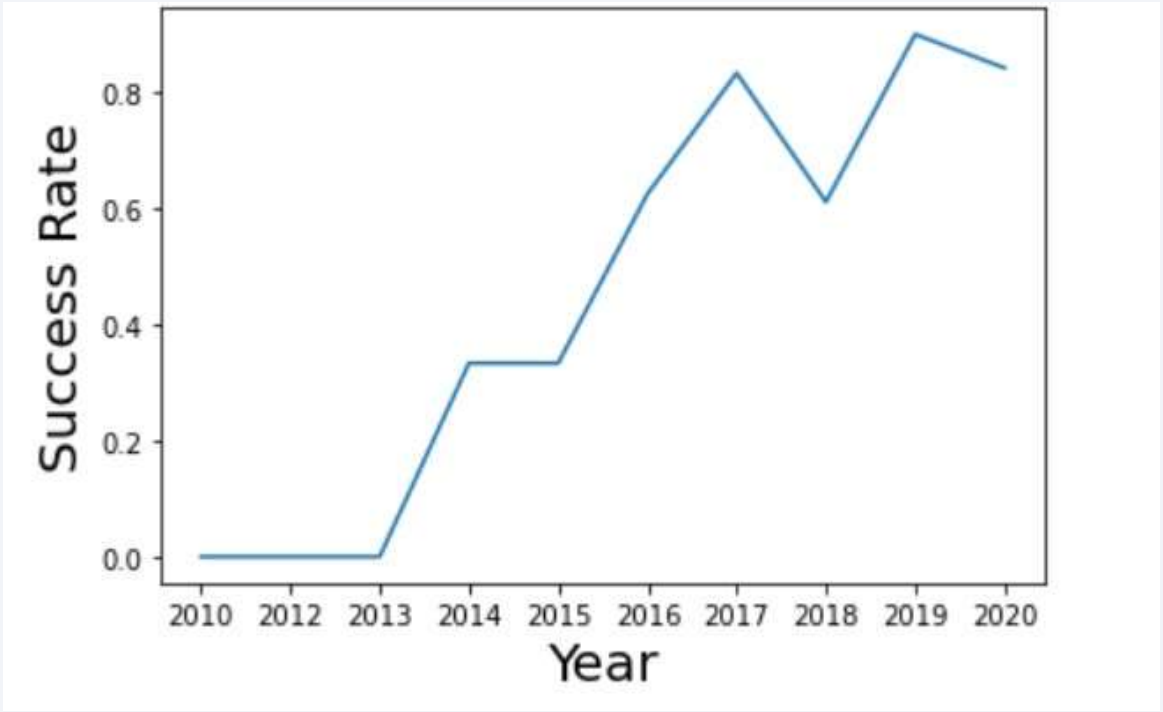
In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.

Launch Success Yearly Trend



The sucess rate since 2013 kept increasing till 2020

All Launch Site Names

Launch Site	Lat	Long
CCAFS LC-40	28.562302	-80.577356
CCAFS SLC-40	28.563197	-80.576820
KSC LC-39A	28.573255	-80.646895
VAFB SLC-4E	34.632834	-120.610746

The latitude(Lat) and longitude(Long) of the 4 Launch sites are presented above.

Launch Site Names Begin with 'CCA'

CCAFS SLC-40	28.563197	-80.576820	1
CCAFS SLC-40	28.563197	-80.576820	1
CCAFS SLC-40	28.563197	-80.576820	0
CCAFS SLC-40	28.563197	-80.576820	0
CCAFS SLC-40	28.563197	-80.576820	0
CCAFS SLC-40	28.563197	-80.576820	1
CCAFS SLC-40	28.563197	-80.576820	0

Section 3

Launch Sites Proximities Analysis

Distance between a launch site and proximities

- Launch sites are in close proximity to highways, which allows for easily transport required people and property.
- Launch sites are in close proximity to railways, which allows transport for heavy cargo.
- Launch sites are not in close proximity to cities, which minimizes danger to population dense areas.

Successful/Failed Launch Map

•Launch sites are in close proximity to coastline so they can fly over the ocean during launch, for at least two safety reasons

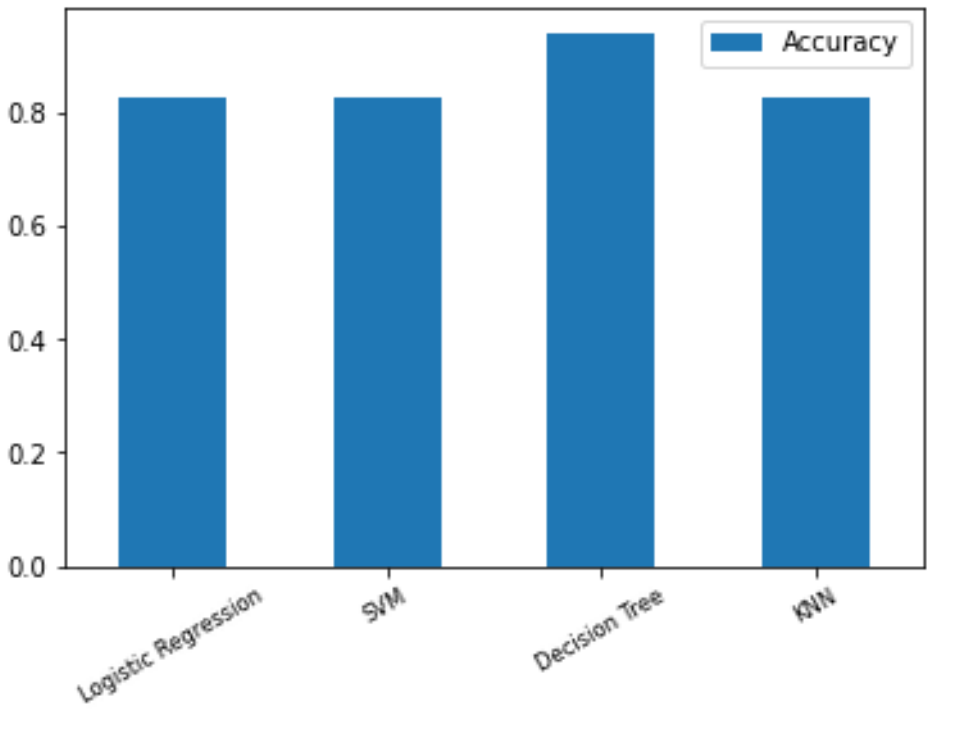
(1) crew has option to abort launch and attempt water landing

(2)minimize people and property at risk from falling debris.

Section 4

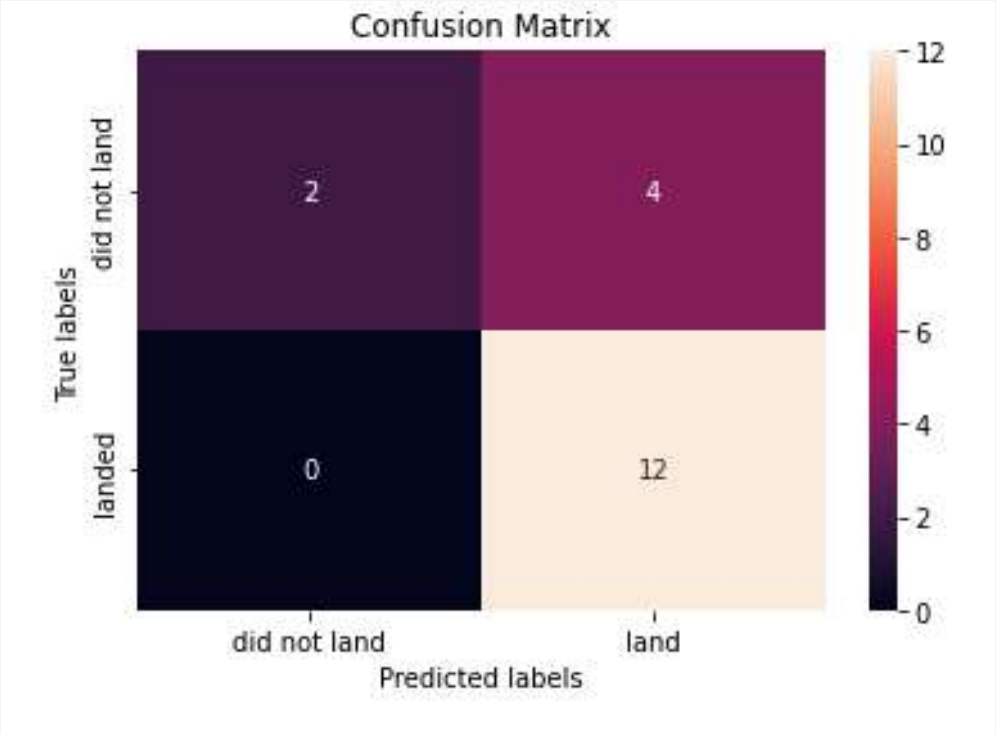
Predictive Analysis (Classification)

Classification Accuracy



Decision Tree achieved the best performance.

Confusion Matrix



Thank you!