



DATA SCIENCE CAPSTONE PROJECT

● BUSINESS PRESENTATION

Phan Pham Quynh Hoa

19 Jun 2024

TABLE OF CONTENT

Data Science Capstone Project

01 ExecutiveSummary

02 Introduction

03 Methodology

04 Results

05 Conclusion

06 Appendix



EXECUTIVE SUMMARY



- Collected data from public SpaceX API and SpaceX Wikipedia page. Created labels column 'class' which classifies successful landings. Explored data using SQL, visualization, folium maps, and dashboards. Gathered relevant columns to be used as features. Changed all categorical variables to binary using one hot encoding. Standardized data and used GridSearchCV to find best parameters for machine learning models. Visualize accuracy score of all models.

EXECUTIVE SUMMARY

Four machine learning models were produced: Logistic Regression, Support Vector Machine, Decision Tree Classifier, and KNearestNeighbors. All produced similar results with accuracy rate of about 83.33%. All models over predicted successful landings. More data is needed for better model determination and accuracy.



A photograph showing two individuals from the side, focused on a document they are both holding and writing on. One person is wearing a black leather jacket and the other a white shirt. They appear to be in a collaborative environment, possibly a workspace or office.

INTRODUCTION

BACKGROUND

SpaceX is the most successful company of the commercial space industry, making space travel affordable. The company advertises 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 the first stage will land, we can determine the cost of a launch. Based on public information and machine learning models, we are going to predict if SpaceX will reuse the first stage.

PROBLEMS

01

Space Y tasks us to train a machinelearning model to predict successful Stage 1 recovery

02

What is the best Classificationalgorithm that can be used for theprediction of the reuseof a first stage?



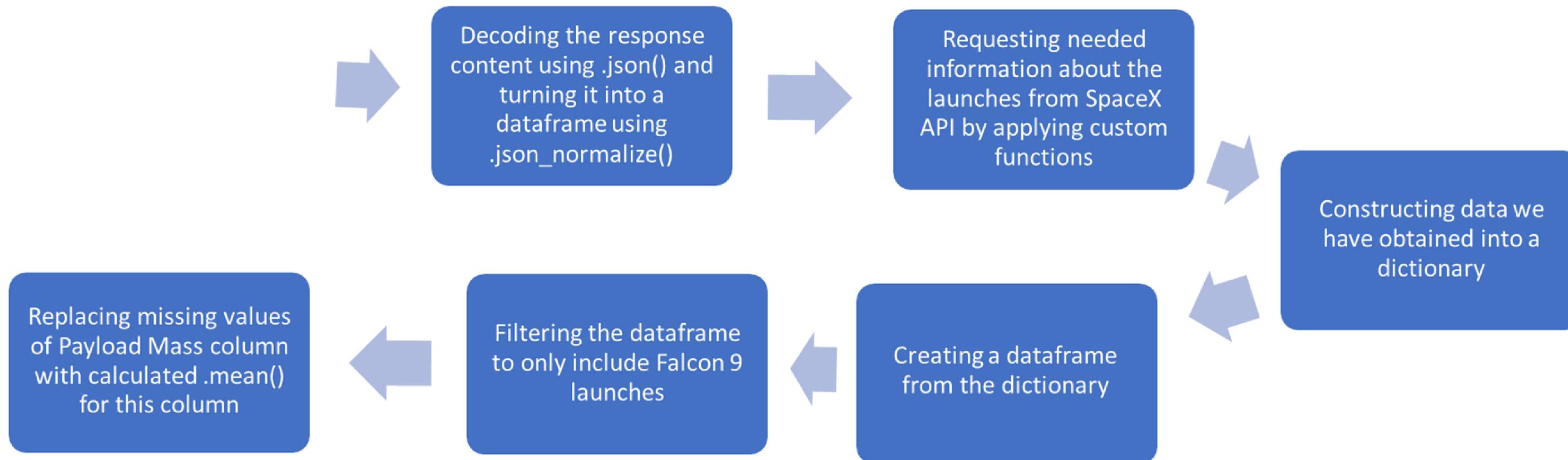
METHODOLOGY



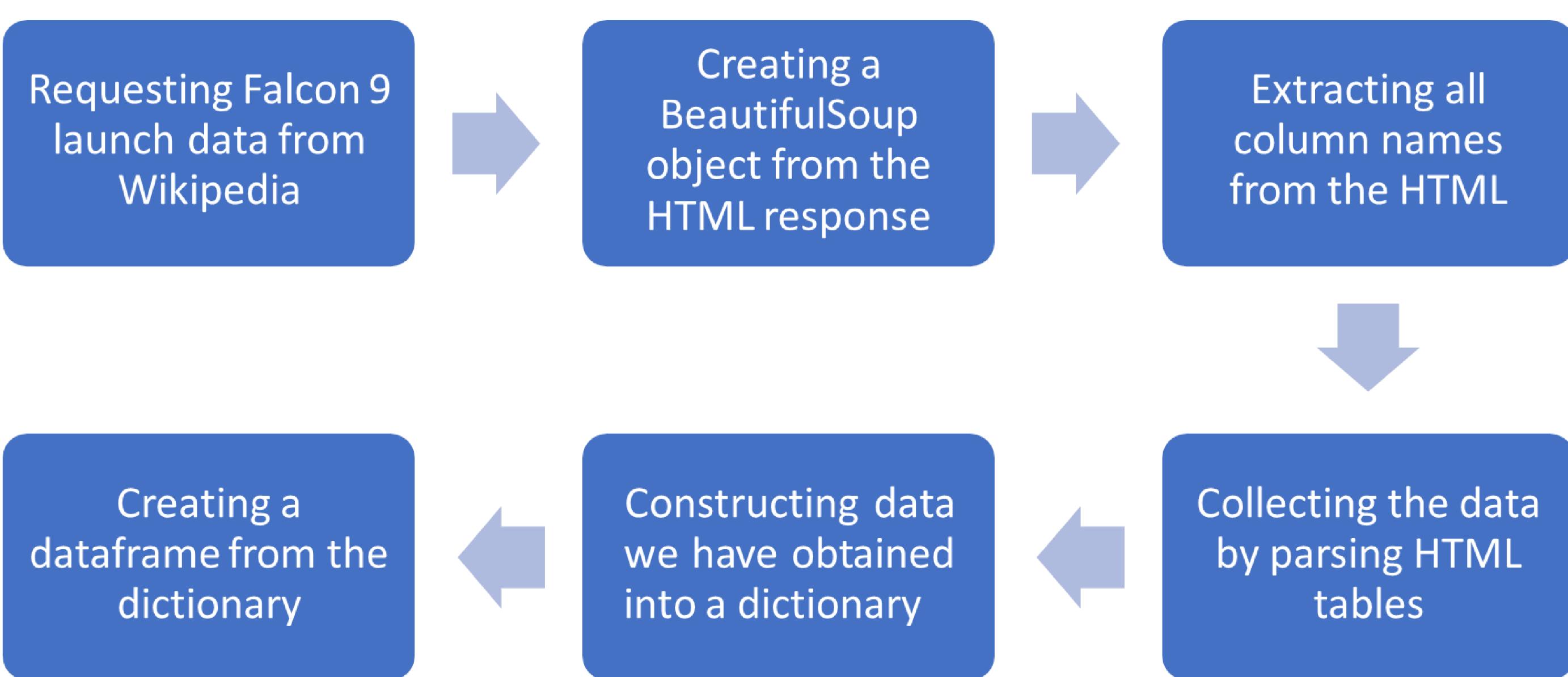
- Data collection methodology: Combined data from SpaceX public API and SpaceX Wikipedia page
- Perform data wrangling: Classifying true landings as successful and unsuccessful otherwise
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models • Tuned models using GridSearchCV



DATA COLECTION- SPACEXAPI



DATA COLECTION- WEBSRAPING



DATA WRANGLING



- In the data set, there are several different cases where the booster did not land successfully. Sometimes a landing was attempted but failed due to an accident; for example, True Ocean means the mission outcome was successfully landed to a specific region of the ocean while False Ocean means the mission outcome was unsuccessfully landed to a specific region of the ocean and others.
- We mainly convert those outcomes into Training Labels with “1” means the booster successfully landed, “0” means it was unsuccessful.



EDA WITH DATA VISUALIZATION



Charts plotted:

- Flight Number vs. Payload Mass.
- Flight Number vs. Launch Site.
- Payload Mass vs. Launch Site.
- Orbit Type vs. Success Rate .
- Flight Number vs. Orbit Type.
- Payload Mass vs OrbitType.
- Success Rate Yearly Trend.



ChartType and theirUse Cases:

Scatter plots show the relationship between variables. If a relationship exists, they could be used in machine learning model.

Bar charts show comparisons among discrete categories. The goal is to show the relationship between the specific categories being compared and a measured value.

Line charts show trends in data over time (time series).

EDA WITH SQL

PerformedSQLqueries:

- Displaying the names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with the string 'CCA'
- Displaying the total payload mass carried by boosters launched by NASA(CRS)
- Displaying average payload mass carried by boosters version F9 v1.1
- Listing the date when the first successful landing outcome in ground pad was achieved
- Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listing the total number of successful and failure mission outcomes
- Listing the names of the booster versions which have carried the maximum payload mass
- Listing the failed landing outcomes in drone ship, their booster versions and launch site names for the months in year 2015
- Ranking 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

BUILD AN INTERACTIVE MAP WITH FOLIUM

Markers of all Launch Sites

Added Marker with Circle, Popup Label and Text Label of NASA Johnson Space Center using its latitude and longitude coordinates as a start location.

- Added Markers with Circle, Popup Label and Text Label of all Launch Sites using their latitude and longitude coordinates to show their geographical locations and proximity to Equator and coasts.

Coloured Markers of the launch outcomes for each Launch Site:

Added coloured Markers of success (Green) and failed (Red) launches using Marker Cluster to identify which launch sites have relatively high success rates.

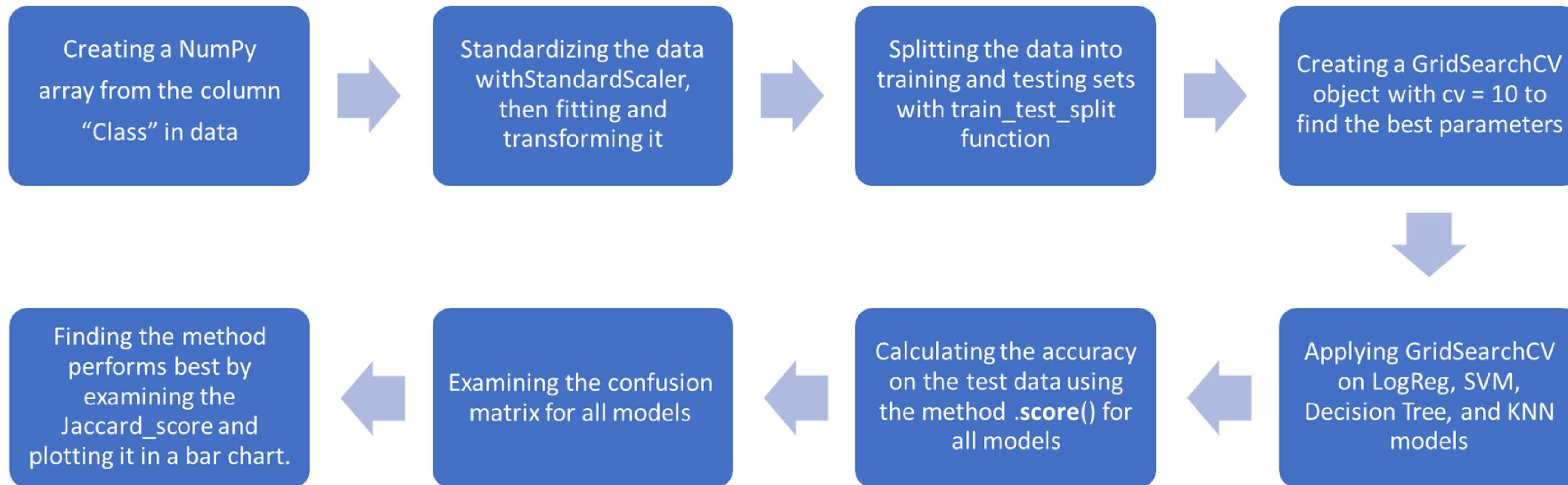
Distances between a Launch Site to its proximities:

Added coloured Lines to show distances between the Launch Site KSC LC-39A (as an example) and its proximities like Railway, Highway, Coastline and Closest City.

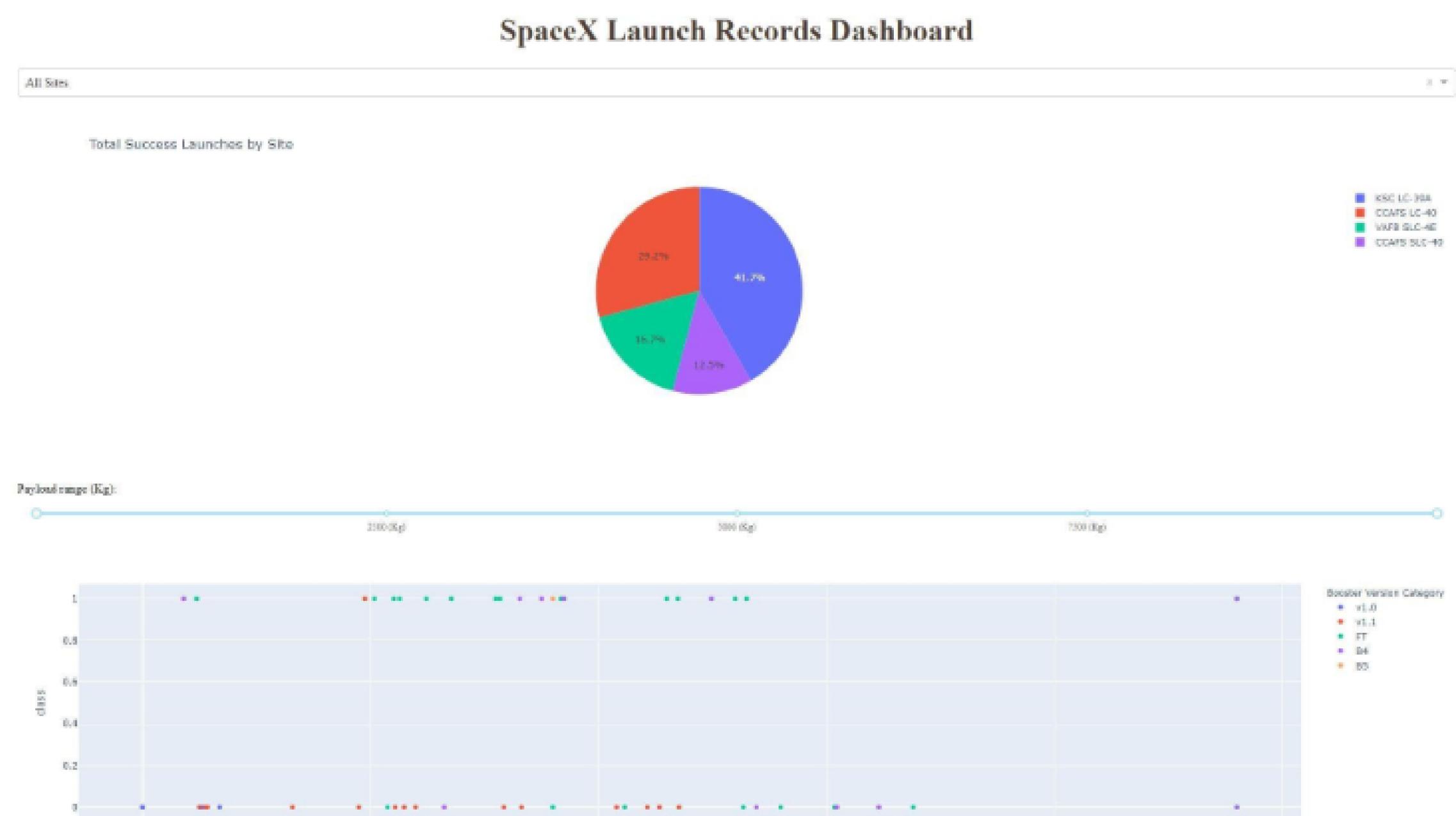
BUILD A DASHBOARD WITH PLOTLY DASH

Dashboard includes a pie chart and a scatter plot. Pie chart can be selected to show distribution of successful landings across all launchsites and can be selected to show individual launch site success rates. Scatter plot takes two inputs: All sites or individual site and payloadmass on a slider between 0 and 10000 kg. The pie chart is used to visualize launch site success rate. The scatter plot can help us see how success varies across launch sites, payloadmass, and booster version category.

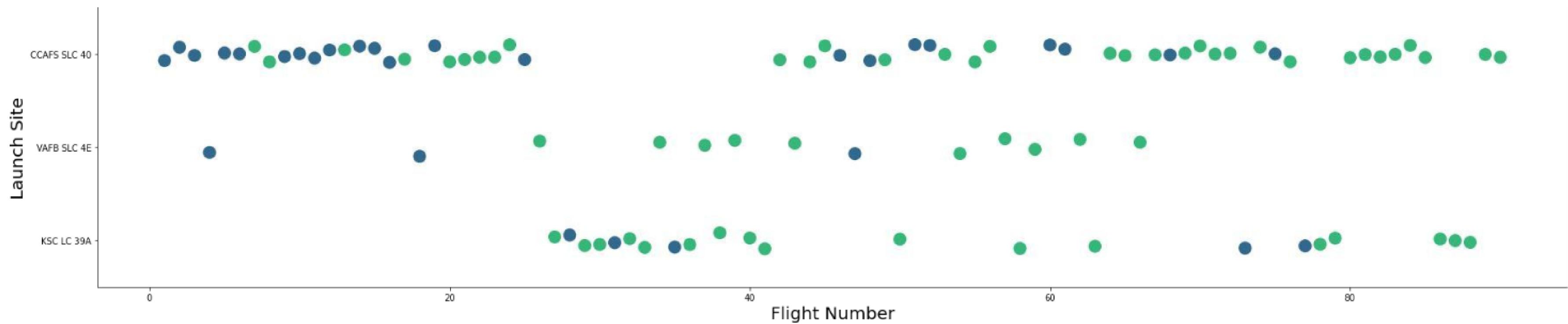
PREDICTIVE ANALYSIS (CLASSIFICATION)



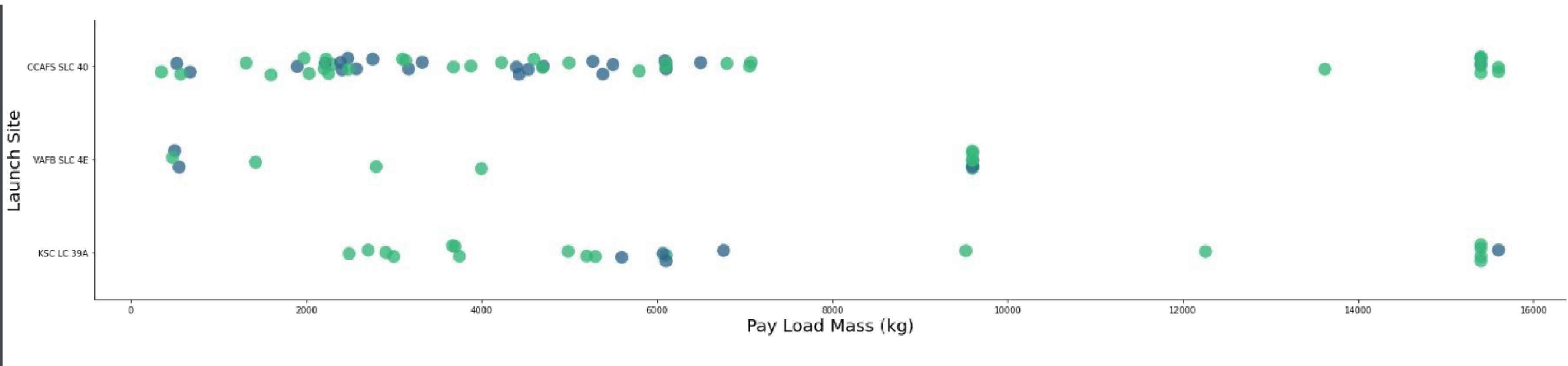
RESULTS



FLIGHT NUMBER VS. LAUNCH SITE



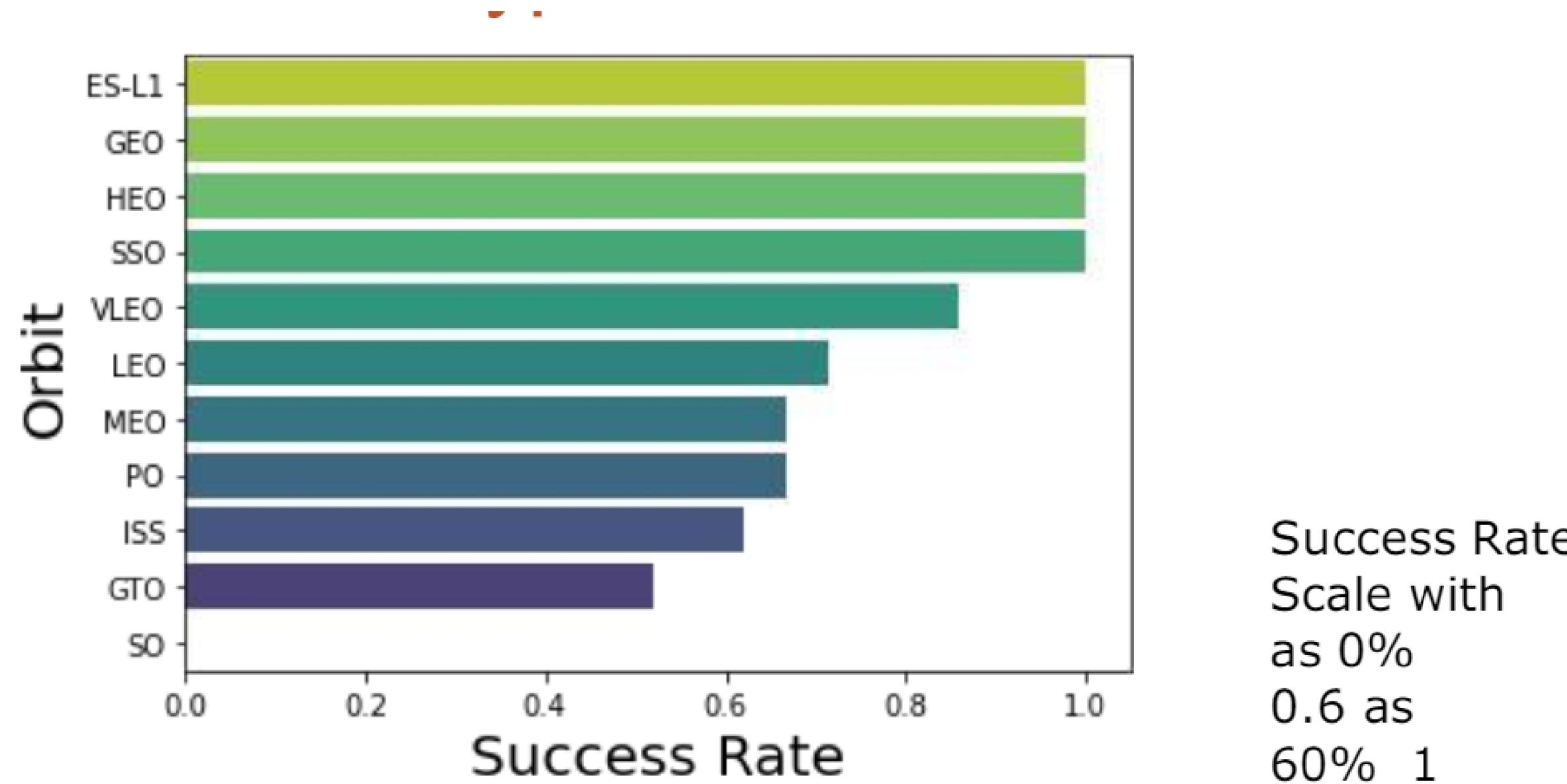
PAYLOAD VS. LAUNCH SITE



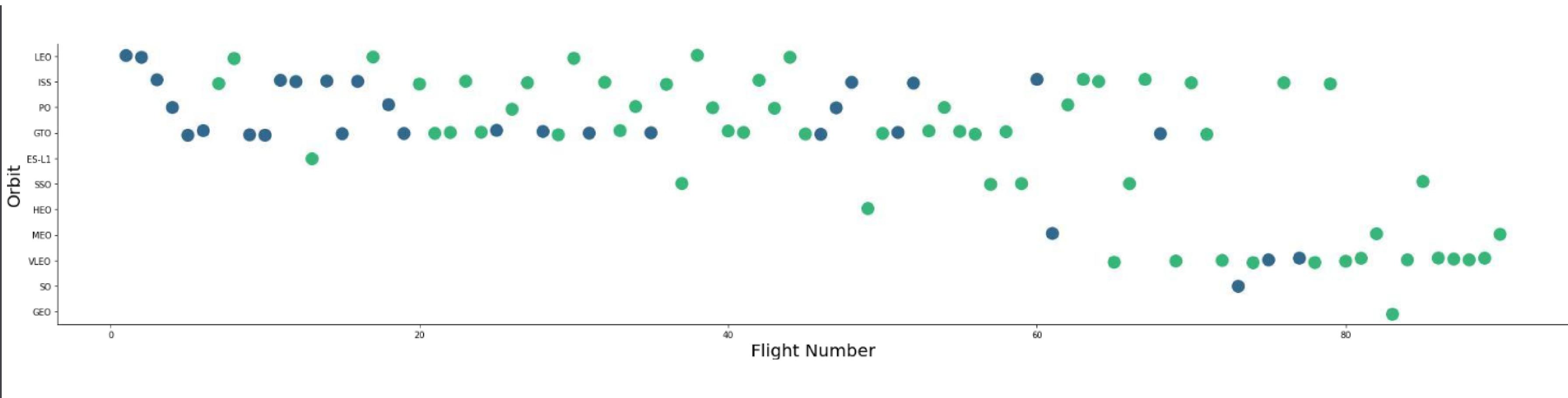
Explanation:

- For every launch site the higher the payload mass, the higher the success rate.
- Most of the launches with payload mass over 7000kg were successful.
- VAFB SLC 4E has the highest success rate among the launch sites.

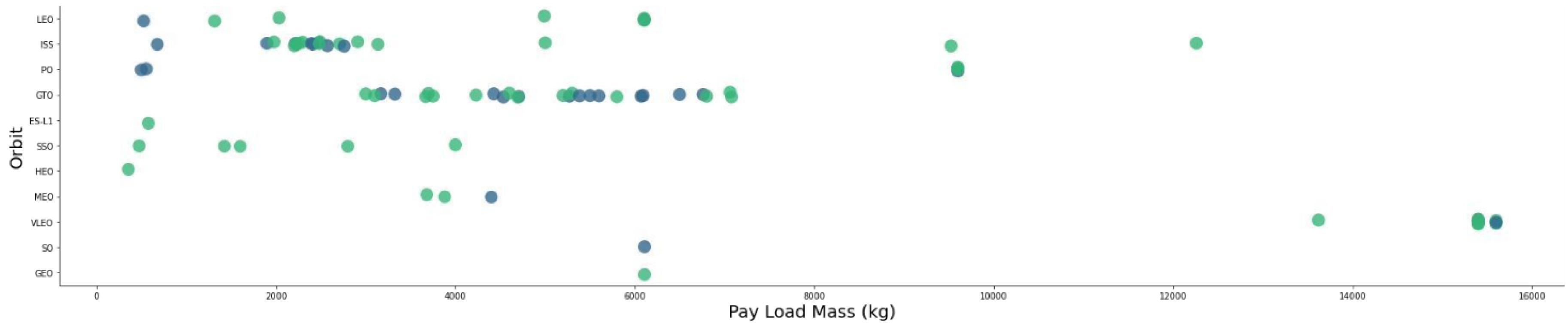
SUCCESS RATE VS. ORBIT TYPE



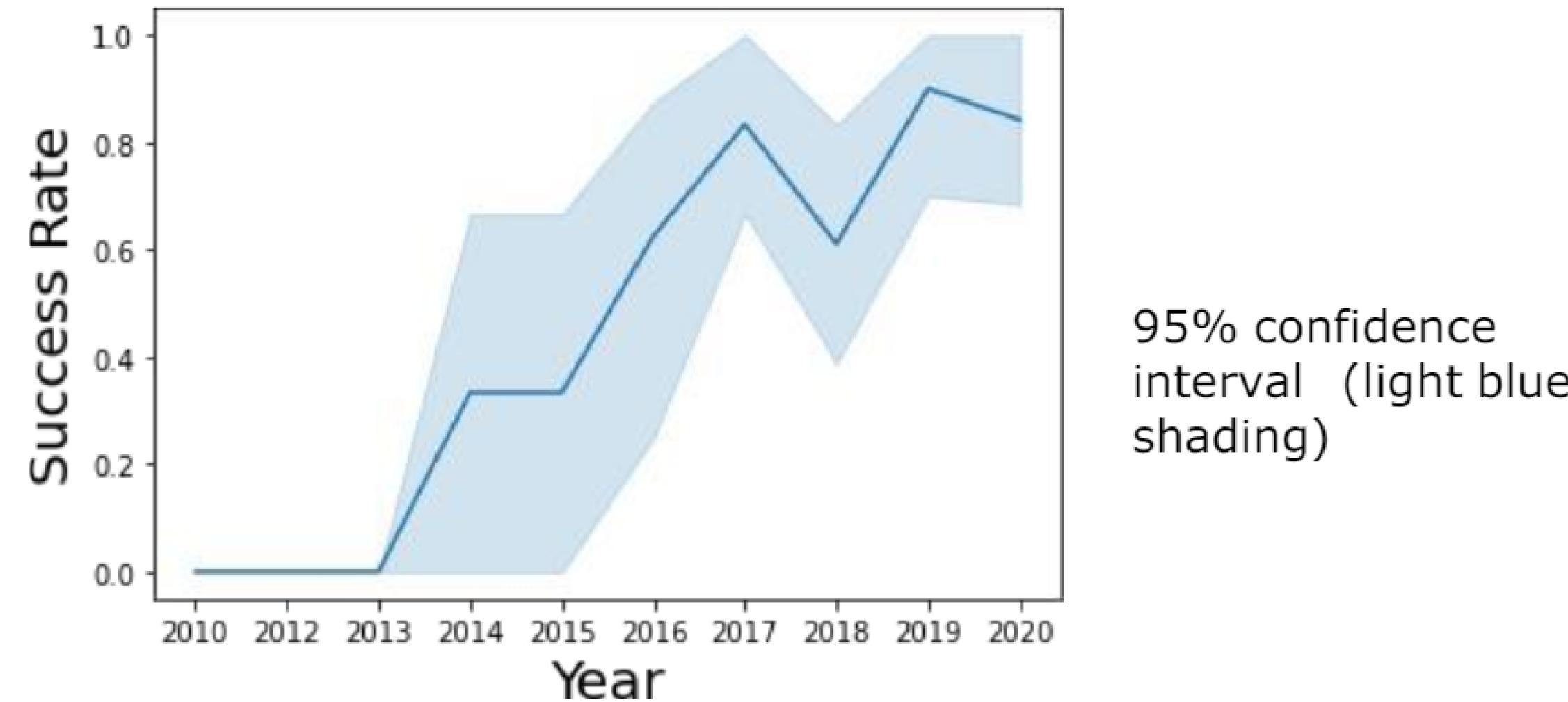
FLIGHT NUMBER VS. ORBIT TYPE



PAYLOAD VS. ORBIT TYPE



LAUNCH SUCCESS YEARLY TREND



ALL LAUNCH SITE NAMES

In [14]:

```
%sql select distinct Launch_Site from SPACEXTABLE
```

* sqlite:///my_data1.db

Done.

Out[14]:

Launch_Site

CCAFS LC-40

VAFB SLC-4E

KSC LC-39A

CCAFS SLC-40



LAUNCH SITE NAMES BEGINNING WITH `CCA`

In [9]:

```
%sql select * from SPACEXTABLE where Launch_Site like 'CCA%' Limit 5
```

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

```
Done.
```

Out[9]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt



THANK YOU

19 Jun 2024