IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Danish Ezwan
August 9, 2025

Outline

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

Executive Summary

- **Summary of methodologies**

- Data Collection API
- Data Collection with Web Scraping
- Data Wrangling
- EDA with SQL
- EDA with Visualization
- Interactive Visual Analytics with Folium
- Interactive Dashboard with Plotly Dash
- Machine Learning Prediction

- **Summary of all results**

- EDA results
- Interactive Analysis results
- Machine Learning Prediction results

Introduction

- **Project background and context**

- SpaceX 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 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 you want to find answers**

- What factors affect the success of the first stage landing?

- Which machine learning model accurately predicts the success of the first stage landing?

Section 1

Methodology

Methodology

Executive Summary

- **Data collection methodology:**
  - Data was collected from the SpaceX REST API which contains information such as payload mass, launch site, outcome, and orbit
- **Perform data wrangling**
  - Data was filtered for the most important entries and missing values were handled
- **Perform exploratory data analysis (EDA) using visualization and SQL**
- **Perform interactive visual analytics using Folium and Plotly Dash**
- **Perform predictive analysis using classification models**
  - Logistic regression, support vector machine, decision tree, and k nearest neighbors

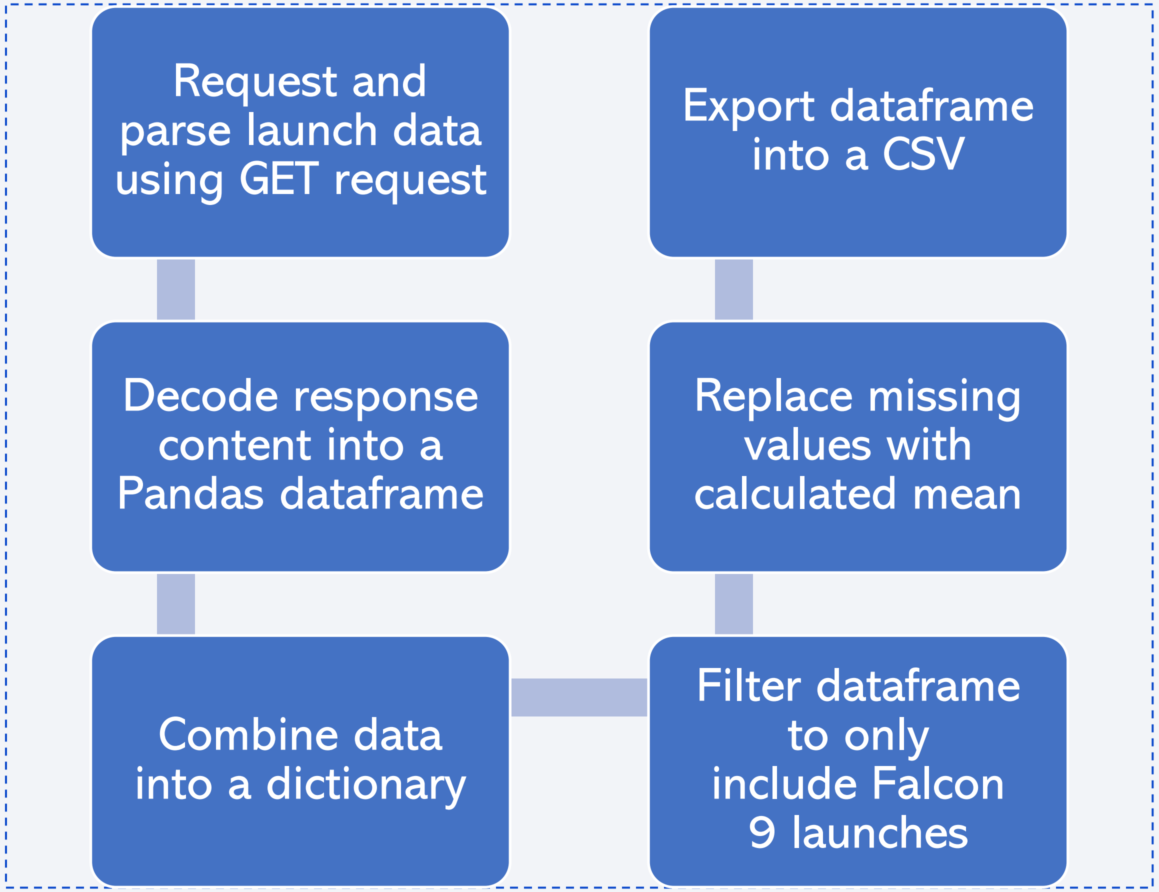
Data Collection

SpaceX API

Scraping

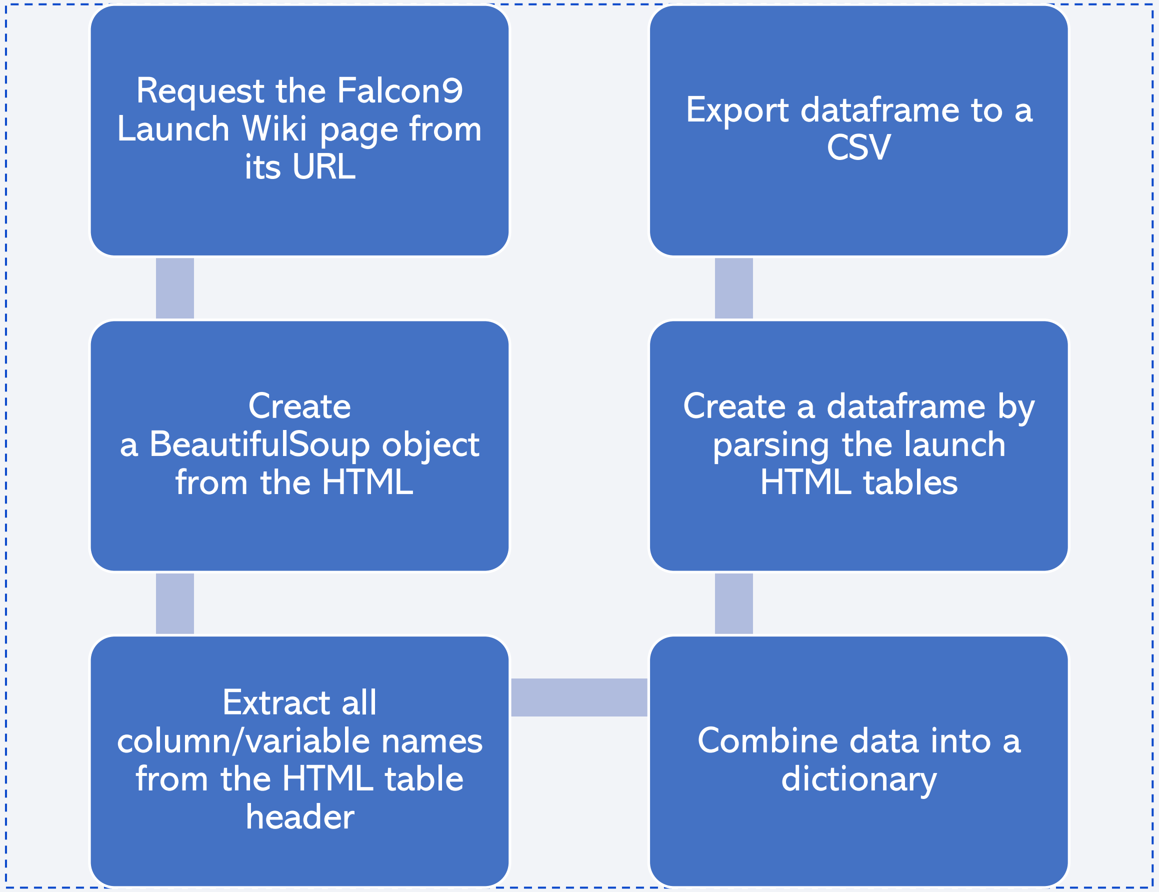
Data Collection – SpaceX API

GitHub URL:
https://github.com/muhddanishaiman/ds-ml-capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

GitHub URL: https://github.com/muhddanishaiman/ds-ml-capstone/blob/main/jupyter-labs-webscraping%20(1).ipynb



Data Wrangling

GitHub URL: https://github.com/muhddanishaiman/ds-ml-capstone/blob/main/labs-jupyter-spacex-data%20wrangling_jupyterlite.ipynb

Identify and calculate the percentage of the missing values in each attribute → Calculate the number of launches on each site → Calculate the number and occurrence of each orbit → Calculate the number and occurences of mission outcome per orbit type → Create a landing outcome label from outcome column → Export data to a CSV

EDA with Data Visualization

- Scatter plots were made to analyze the relationship between the following variables:

    - Flight number and launch site

    - Payload mass and launch site

    - Flight number and orbit type

    - Payload mass and orbit type

- A bar graph of success rate of each orbit type was made to distinguish which orbit type was most successful

EDA with Data Visualization

- A line graph of success rate by year was made to analyze how the success rate changed overtime
- GitHub URL: https://github.com/muhddanishaiman/ds-ml-capstone/blob/main/edadataviz.ipynb

EDA with SQL

- Display unique launch sites
- Display launch sites that start with certain letters
- Calculate and display total and average payload mass
- List and display dates and booster versions of successful landings
- Count and display successful and failed mission outcomes
- Filter and display records based on certain constraints
- GitHub URL: https://github.com/muhddanishaiman/ds-ml-capstone/blob/main/jupyter-labs-eda-sql-edx-sqllite-v2.ipynb

Build an Interactive Map with Folium

- Circle objects highlight launch site proximity to determine how big a launch site operates
- Marker objects highlights launch outcomes to determine if they were successful or unsuccessful
- PolyLine objects highlight the distance between proximities such as highways and coastline to determine a minimum distance from these proximities
- GitHub URL: https://github.com/muhddanishaiman/ds-ml-capstone/blob/main/lab_jupyter_launch_site_location%20(1).ipynb

Build a Dashboard with Plotly Dash

- Pie chart used to determine the distribution of total successful launches for each launch site to determine which is most favorable
- Scatter plot used to determine the relationship between payload mass and launch outcome to determine the range of mass that are most successf
- Range slider used to set a range for payload mass to highlight specific values in the pie chart and scatter plot
- GitHub URL: https://github.com/muhddanishaiman/ds-ml-capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

GitHub URL: https://github.com/muhddanishaiman/ds-ml-capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

1. Split the data into training and test data
2. Create a GridSearchCV object with cv = 10 to find the best parameters
3. Apply GridSearchCV on logistic regression, SVM, decision tree, and KNN
4. Examine the confusion matrix for all models
5. Determine the best performing model by observing each models' best score

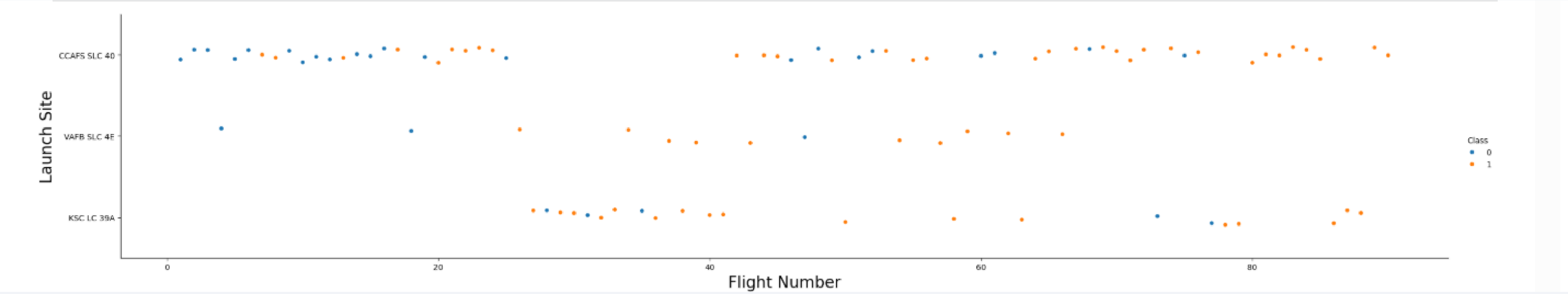
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

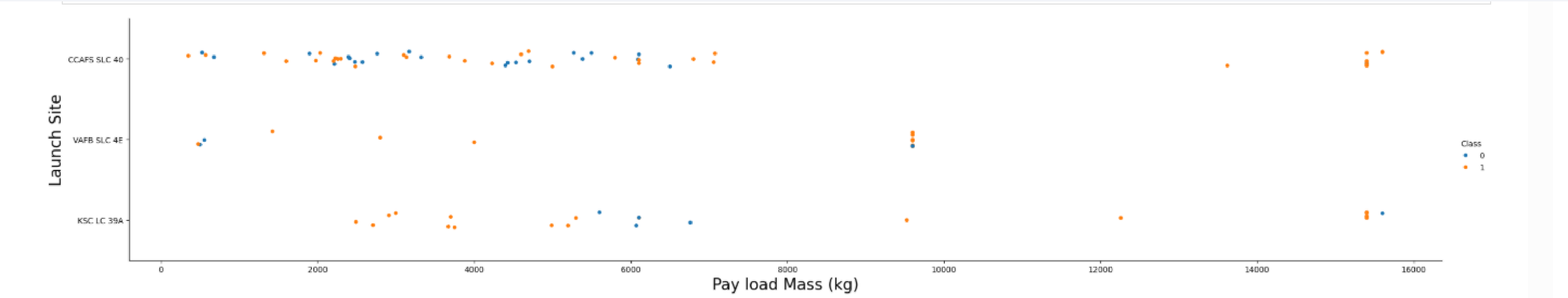
Insights drawn from EDA

Flight Number vs. Launch Site



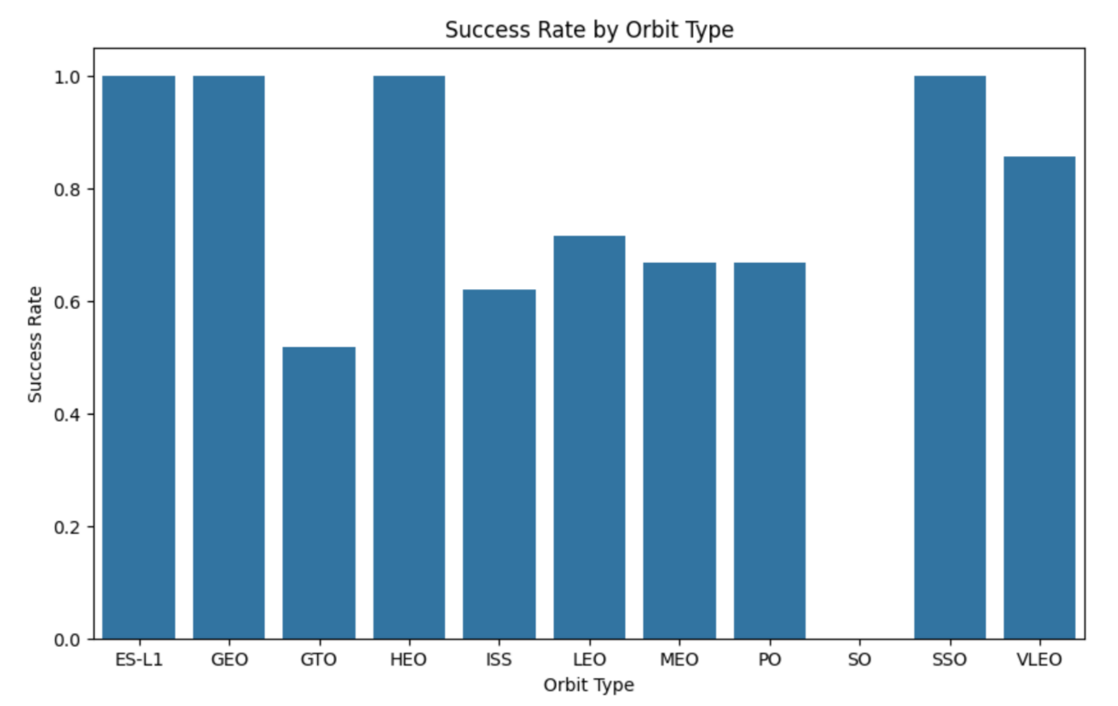
CCAFS SLC-40 has a mix of successful and unsuccessful launches. VAFB SLC-4E and KSC LC-39A also have a mix of successful and unsuccessful launches, but the unsuccessful ones are more frequent

Payload vs. Launch Site



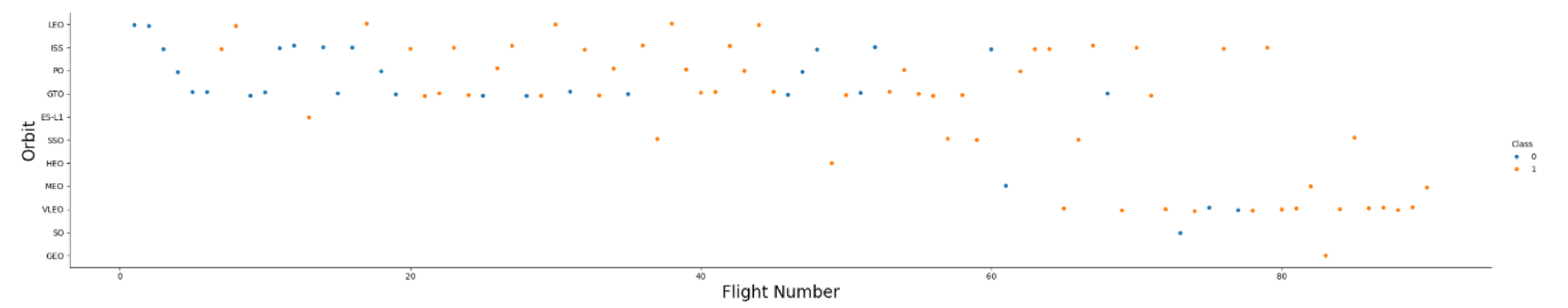
Most of the launches occur with a payload mass of less than 8,000 kg

Success Rate vs. Orbit Type



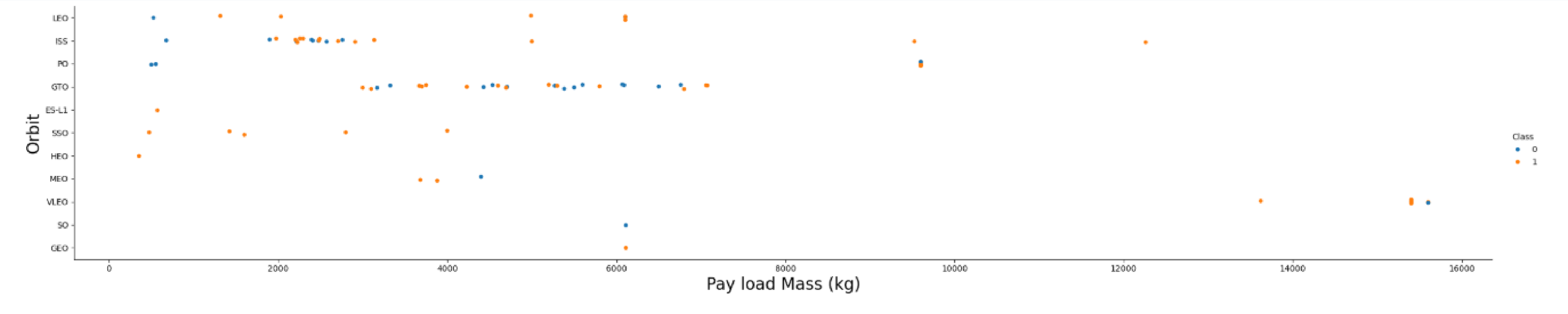
ES-L1, GEO, HEO, and SSO orbit types are the most successful with a success rate of 1.0. SO orbit type are the lest successful with a success rate of 0. Other orbit types have a success rate of at least 0.5

Flight Number vs. Orbit Type



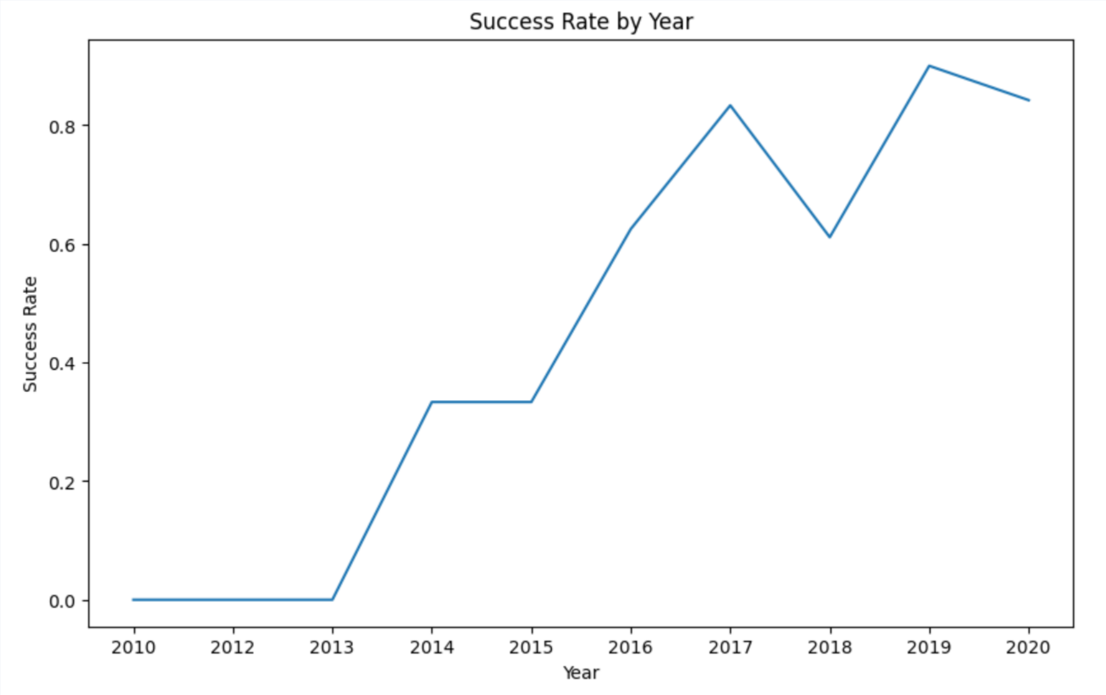
The more recent flights have lower chances of failures across various orbit types

Payload vs. Orbit Type



Payload masses below 8,000 kg have the highest frequency of successful and unsuccessful launches across various orbit types

Launch Success Yearly Trend



The launch success rate increases as the year goes by despite a dip in 2018

All Launch Site Names

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

Launch Site Names Begin with 'KSC'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-03-16	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-05-01	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
2017-05-15	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt

Total Payload Mass

sum(payload_mass__kg_)
45596

Average Payload Mass by F9 v1.1

avg(payload_mass__kg_)
2928.4

First Successful Ground Landing Date

min(date)
2016-04-08

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster_Version
F9 FT B1032.1
F9 B4 B1040.1
F9 B4 B1043.1

Total Number of Successful and Failure Mission Outcomes

Mission_Outcome	Counts
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

Booster_Version
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

2017 Launch Records

month_name	Landing_Outcome	Booster_Version	Launch_Site
February	Success (ground pad)	F9 FT B1031.1	KSC LC-39A
May	Success (ground pad)	F9 FT B1032.1	KSC LC-39A
June	Success (ground pad)	F9 FT B1035.1	KSC LC-39A
August	Success (ground pad)	F9 B4 B1039.1	KSC LC-39A
September	Success (ground pad)	F9 B4 B1040.1	KSC LC-39A
December	Success (ground pad)	F9 FT B1035.2	CCAFS SLC-40

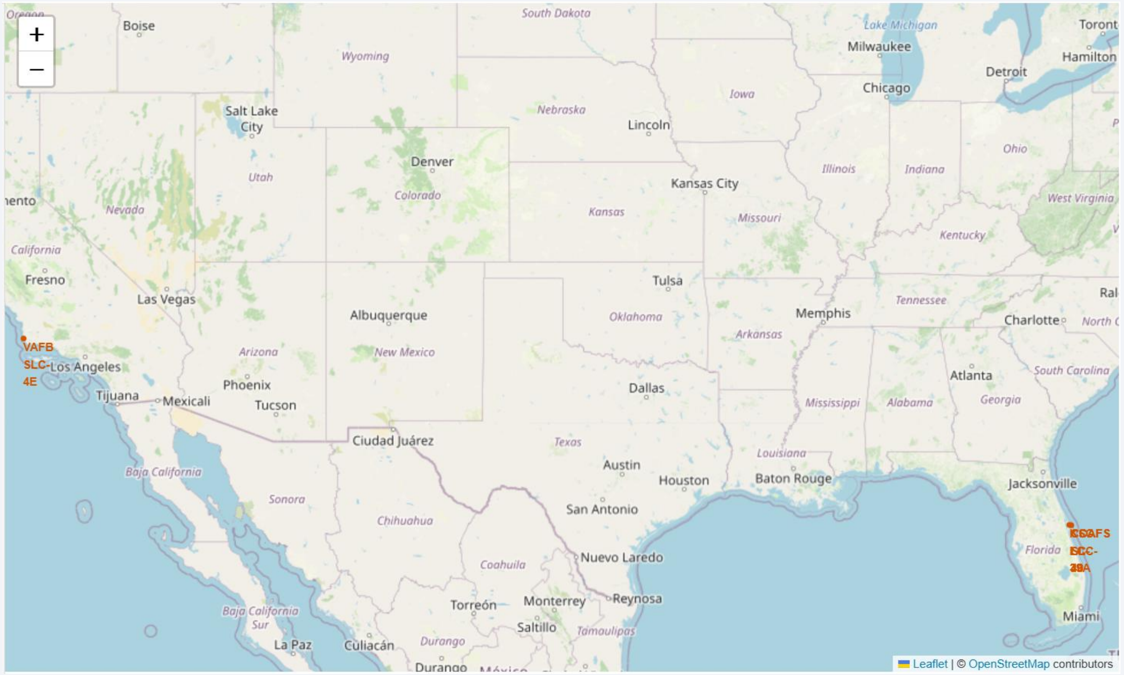
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Landing_Outcome	Counts
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

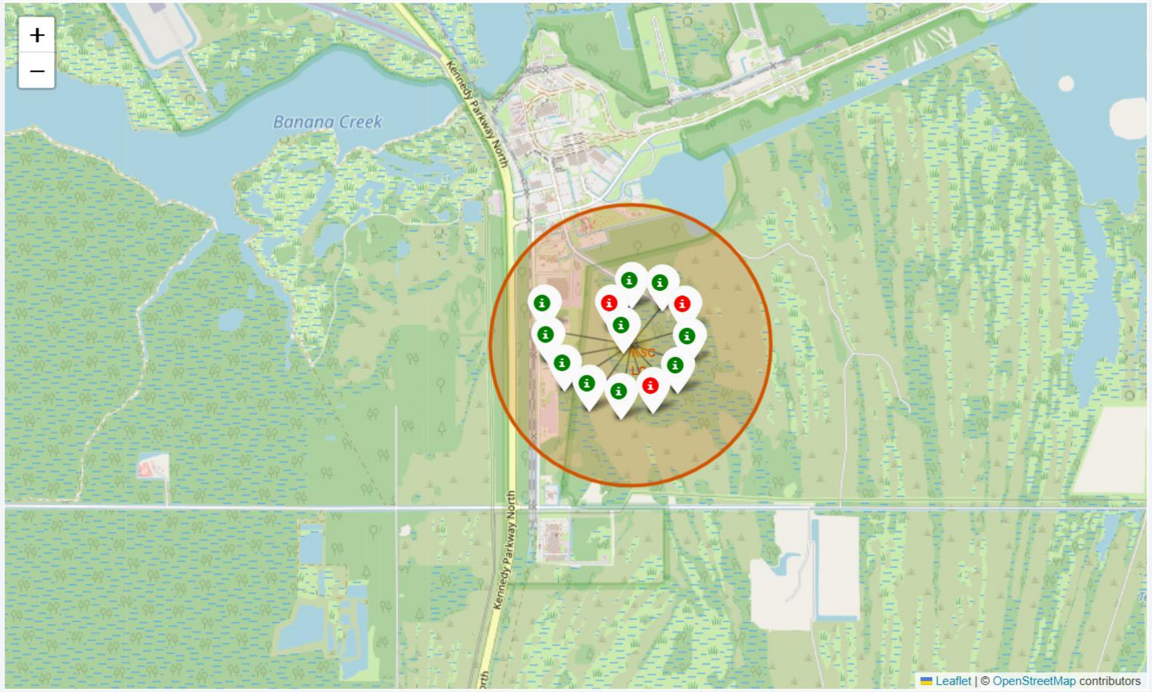
Launch Sites Proximities Analysis

Launch Site Locations



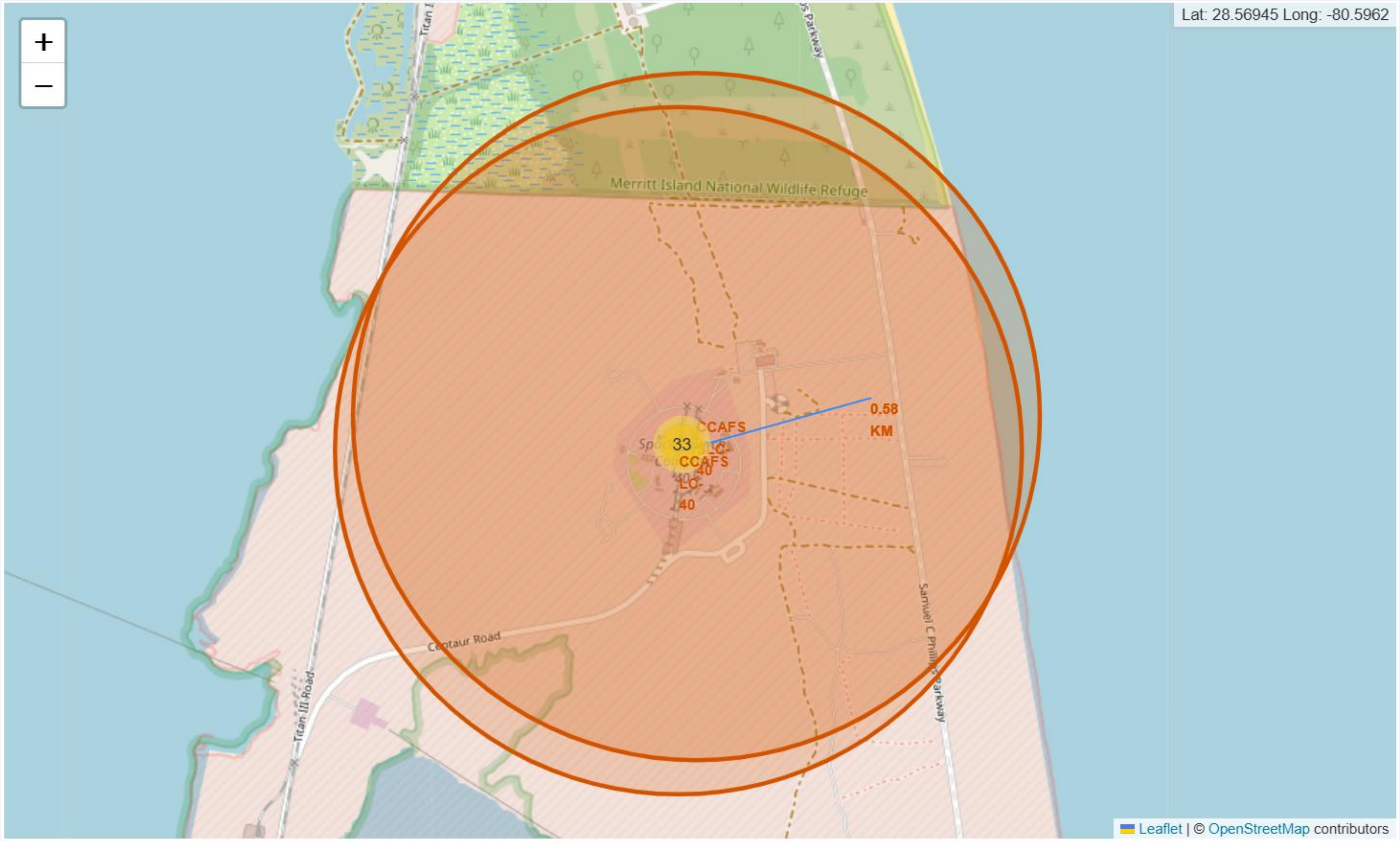
Not all launch sites are in close proximity with the Equator line. All launch sites are in close proximity to a coast.

Color-labeled Launch Outcome



A green marker indicates a successful launch while a red marker indicates an unsuccessful launch. KSC LC-39A has the highest success rate

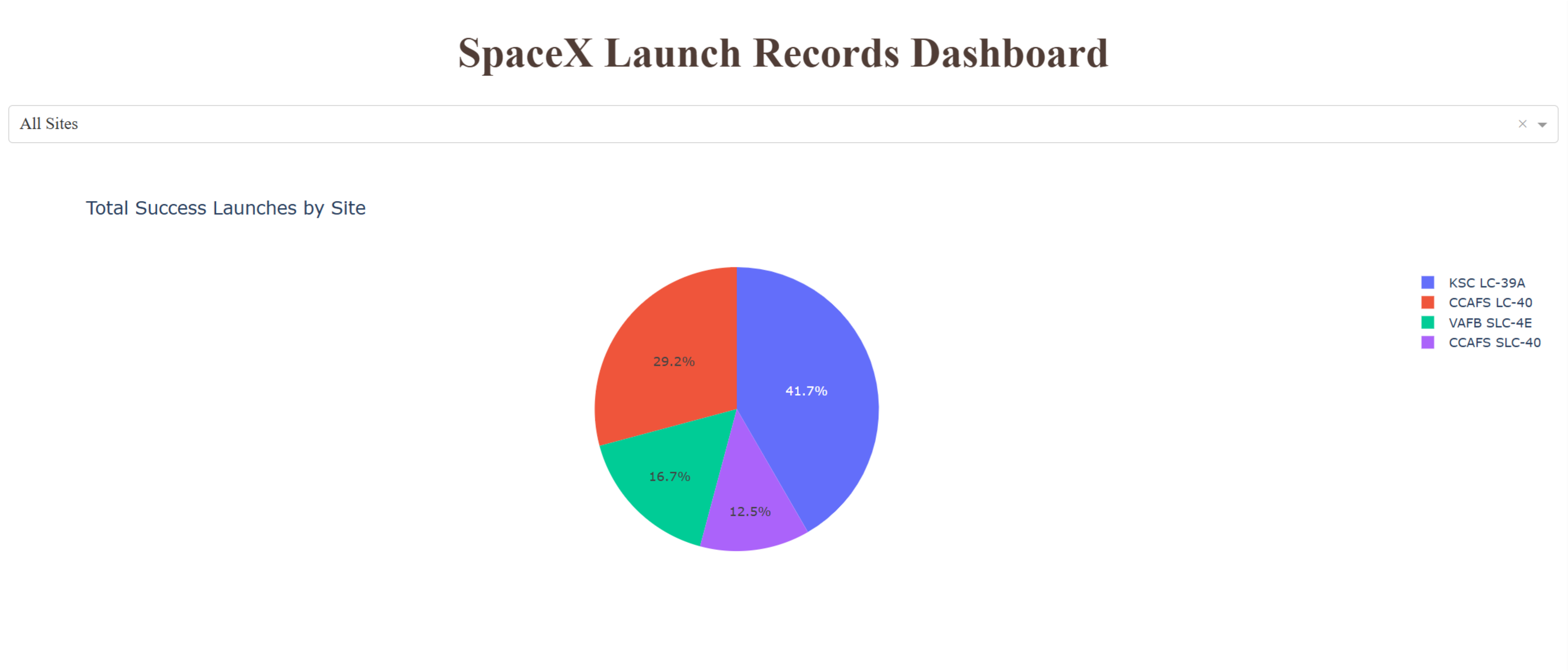
Distance from Coastline



Section 4

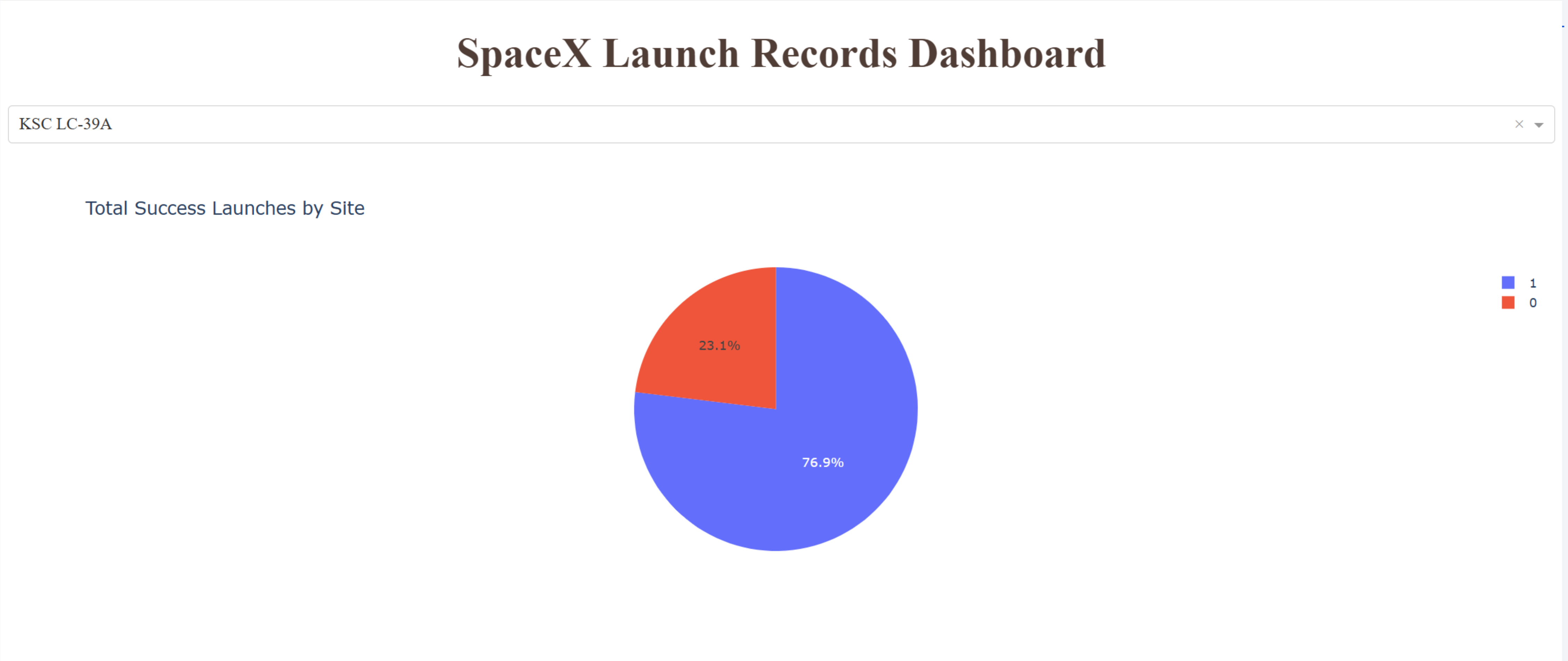
Build a Dashboard with Plotly Dash

Total Success Launches by Site



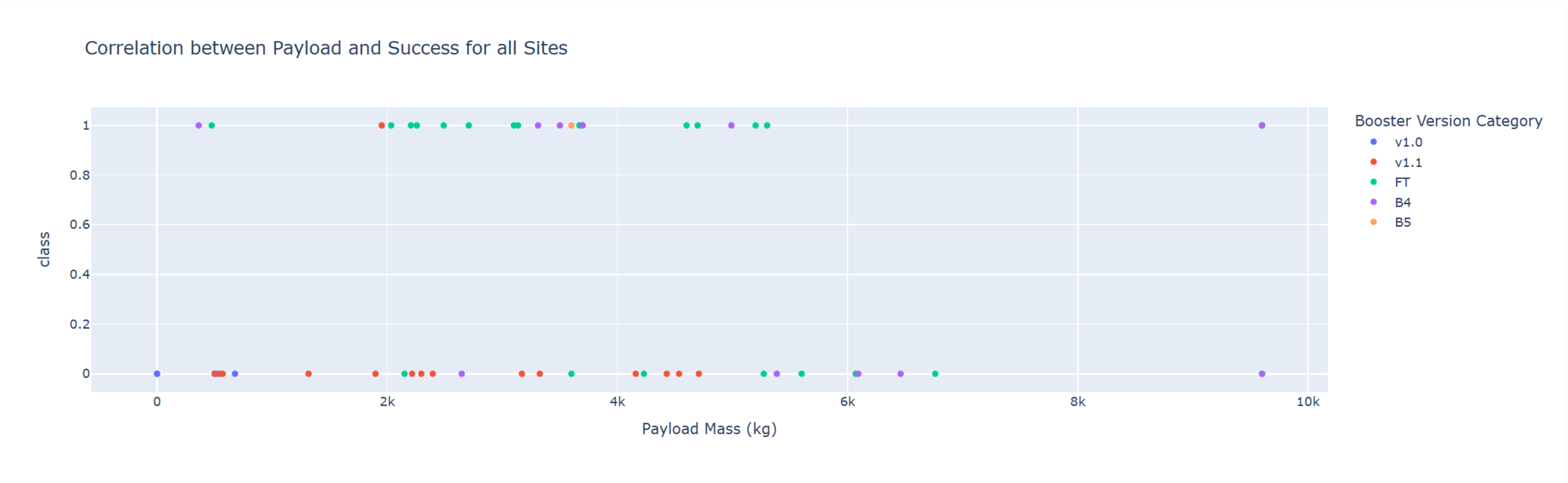
KSC LC-39A has the highest total success launches by site. This makes it the most favorable launch site

Launch Site with Highest Success Ratio



KSC LC-39A has the highest success ratio. This makes it the most successful launch site

Payload vs Launch Outcome

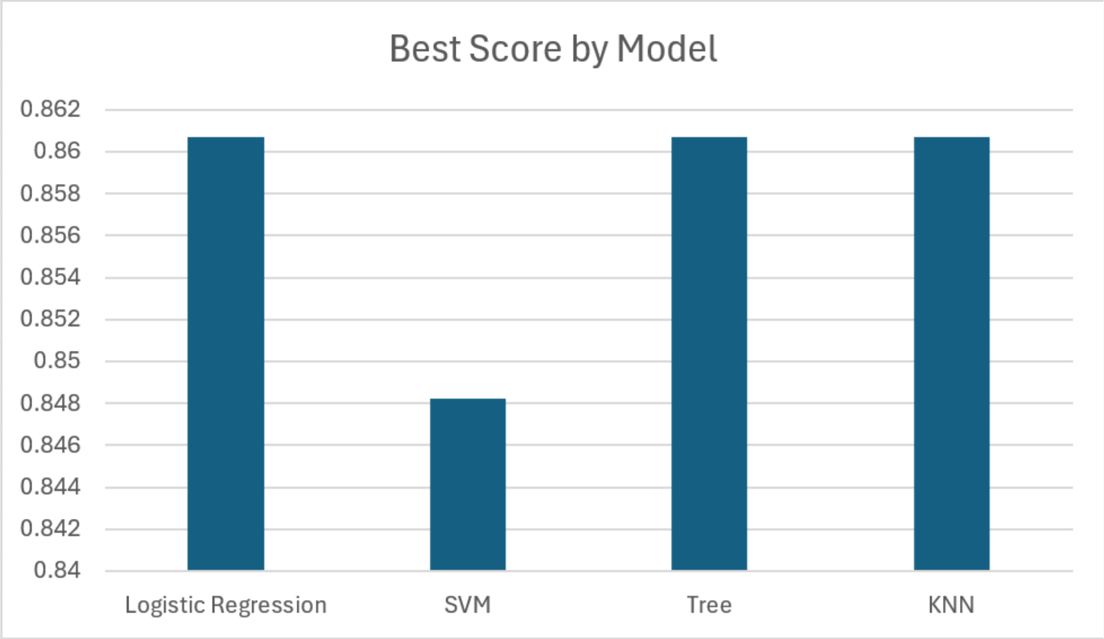


FT has the highest frequency of success across various payload masses while v1.1 has the highest frequency of failure across various payload masses

Section 5

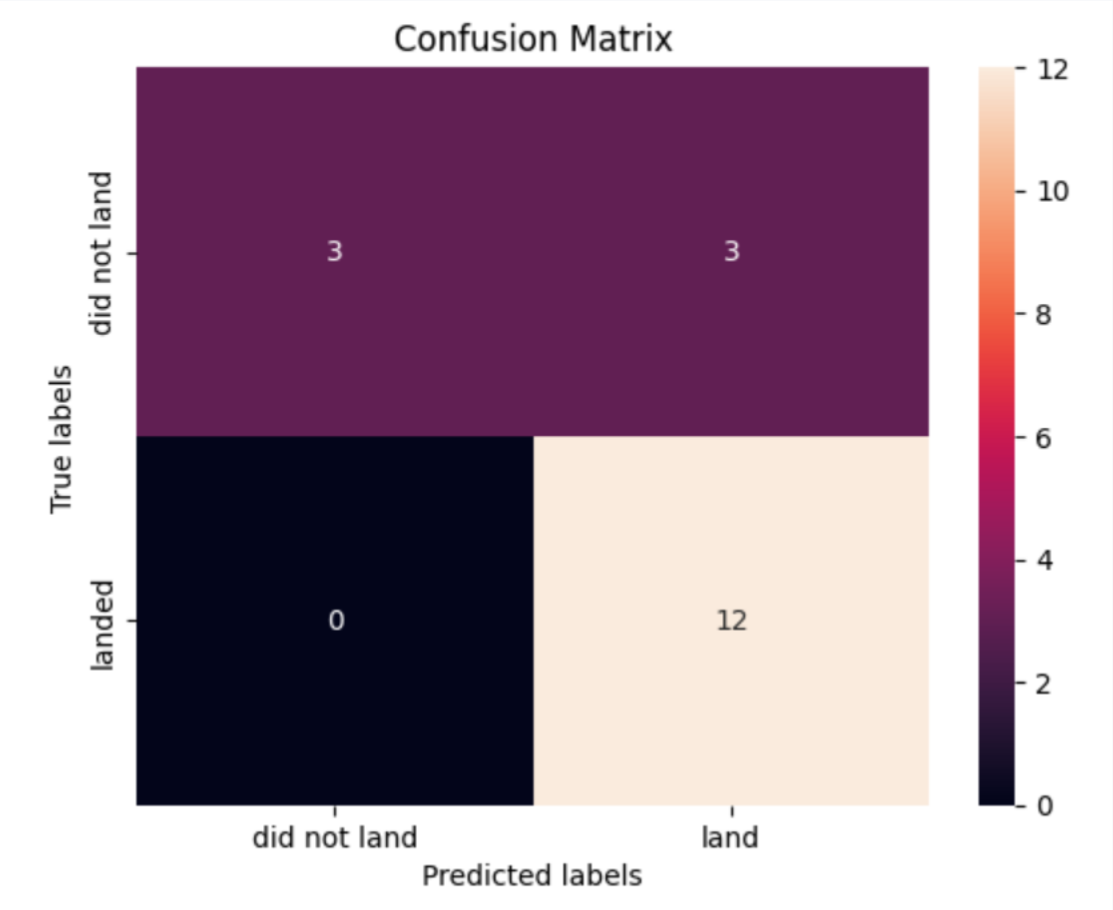
Predictive Analysis (Classification)

Classification Accuracy



Logistic regression, decision tree, and k nearest neighbors have the highest classification accuracy compared to support vector machine

Confusion Matrix



The models correctly predict most of the successful landings when the rocket indeed successfully landed

Conclusions

- Launch site KSC LC-39A has the highest amount of successful launches
- Logistic regression, decision tree, and k nearest neighbors are most accurate in predicting if a launch will be successful
- Lower payload masses have higher tendencies to have successful launches
- Launch success rate increases as the year goes by
- Booster version FT has the highest amount of successful launches
- Orbit types ES-L1, GEO, HEO, and SSO have the highest amount of successful launches

Thank you!