Winning Space Race with Data Science

Corey Yang-Smith
May 29th 2023

Outline

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

Executive Summary

- Methodologies used to analyze data:
  - Data Collection: Web Scraping (beautiful soup) and SpaceX API (pandas/numpy)
  - Data Wrangling: Data pre-processing using pandas/numpy
  - Exploratory Data Analysis (EDA): Data Visualization and Interactive Dashboard using Folium & Plotly
  - Machine Learning: Prediction utilizing sci-kit learn

- Results Summary:
  - EDA, specifically dashboards, allowed for a concise and visual analysis of the data
  - Machine Learning Prediction determined the best model to predict characteristics

Introduction

- Background: The objective of this project was to explore the cost and success of space launches for SpaceX, considering a new competitor coming into the space
- Answers we want:
  - Estimated total cost for launches, as indicated by success rate of first stage rocket landings.
  - Best location for launches

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Space X API
  - Wikipedia Webscraping
- Perform data wrangling
  - Collected data was pre-processed to indicate landing outcomes by location in order to activate data visualization
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Predictive analysis was used in order to determine the success rate of new launches based on previous data.

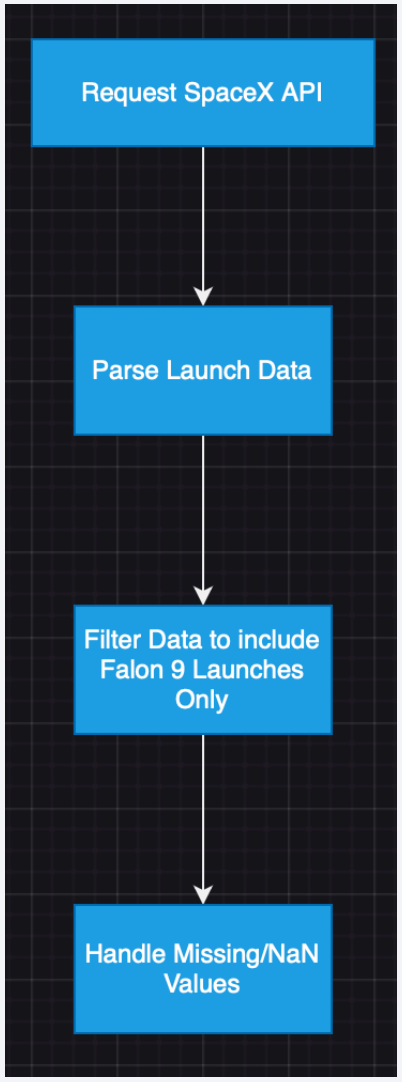
Data Collection

- Data sets were collected from the Space X API and Web Scraped from Wikipedia using beautiful soup.

Data Collection – SpaceX API

- Data from SpaceX acquired via public API

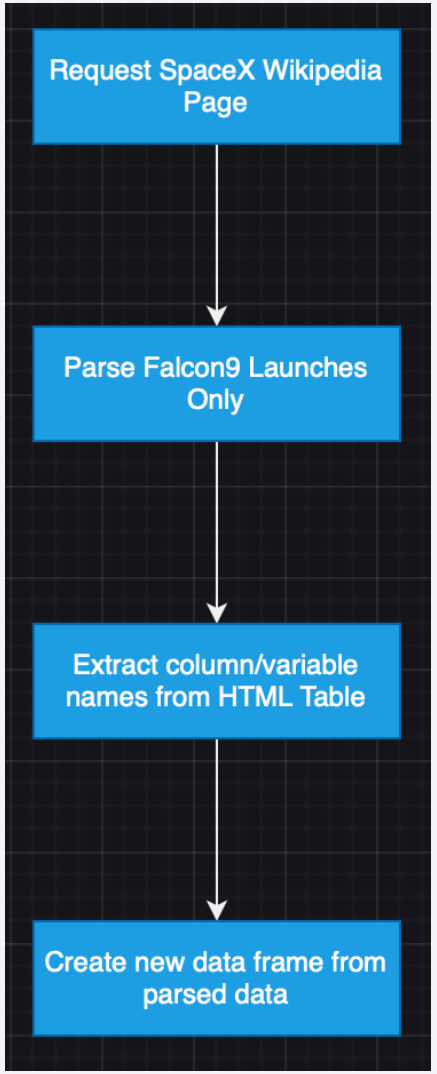
- Source Code: https://github.com/coreyyangsmith/ibm-data-science-captsone/blob/main/1.1-Data-Collection-API.ipynb



Data Collection - Scraping

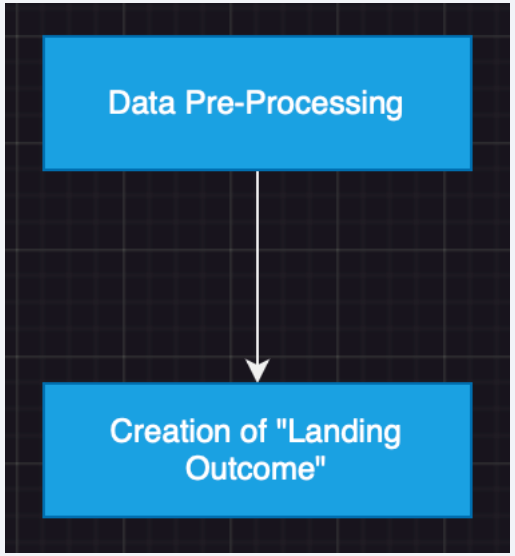
- Data from SpaceX Launches obtained from Wikipedia

- Source Code: https://github.com/coreyyangsmith/ibm-data-science-captsone/blob/main/1.2-Data-Collection-Web-Scraping.ipynb



Data Wrangling

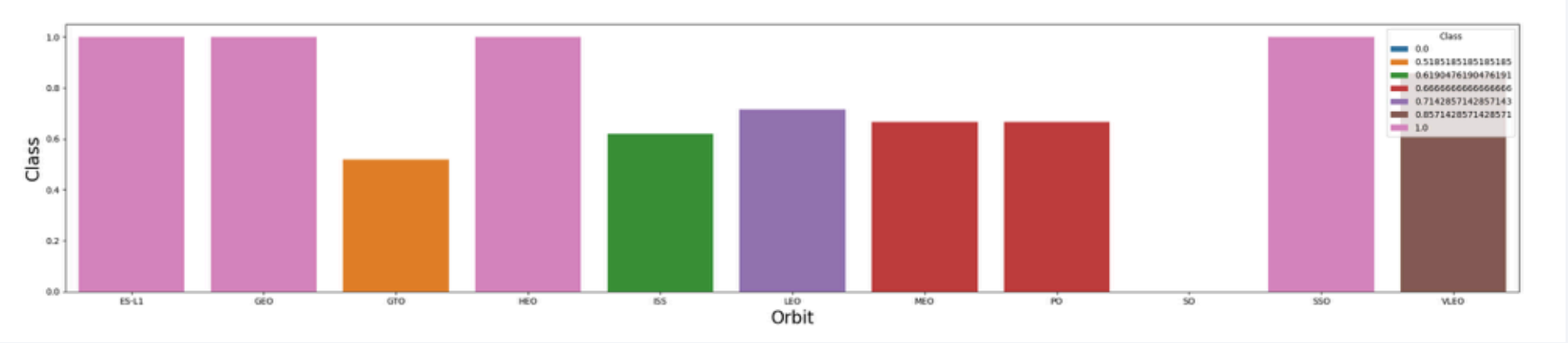
- Exploratory Data Analysis (EDA) performed on the data set.



- Source Code: https://github.com/coreyyangsmith/ibm-data-science-captsone/blob/main/1.3-Data-Wrangling.ipynb

EDA with Data Visualization

- Scatter plots and bar plots were generated to visualize the relationship between a number of features. Mainly, we wanted to observe how the success rate was influence by a number of factors: payload mass, flight number, launch site, orbit, and more.



- Source Code: https://github.com/coreyyangsmith/ibm-data-science-captsone/blob/main/2.2-EDA-with-Visualization.ipynb

EDA with SQL

- SQL Queries Performed:
  - Successful landing with a ground pad
  - Successful drone ships with 4000 < payload mass < 6000
  - Total successful vs unsuccessful mission outcomes
  - Boosters with maximum payload mass
  - Successful landing outcomes between set dates
- Source Code: https://github.com/coreyyangsmith/ibm-data-science-captsone/blob/main/2.1-EDA-with-SQL.ipynb

Build an Interactive Map with Folium

- Folium map was created with radius circle in order to indicate launch locations
- Various markers added in marker groups at launch locations to visualize the success rates at those launch sites
- Lines used to indicate distance between two coordinates
- Source Code: https://github.com/coreyyangsmith/ibm-data-science-captsone/blob/main/3.1-Visual-Analytics-with-Folium.ipynb

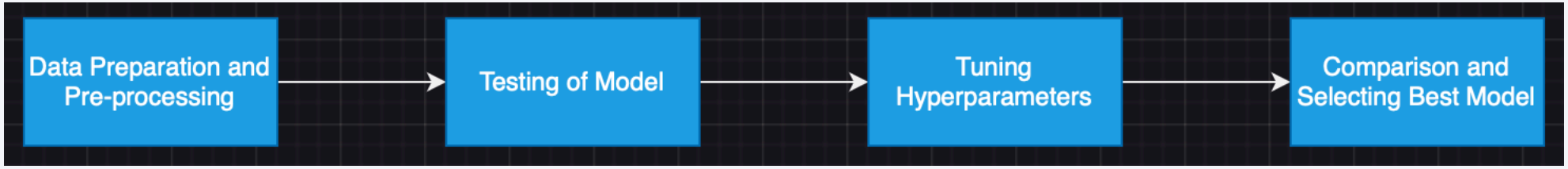
Build a Dashboard with Plotly Dash

- The following graphs and plots were created to visualize data:
  - Pie Chart - percentage of successful launches by Site
  - Scatter Plot - success by payload mass
- These interactive plots were added in order to better visualize the success at each site, and how payload mass may have affected their success rates.

- Source Code: https://github.com/coreyyangsmith/ibm-data-science-captsone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Four classification models were compared
  - Logistic Regression
  - Support Vector Machine (SVM)
  - Decision Tree Classification
  - K-Nearest-Neighours (KNN)



  - Source Code: https://github.com/coreyyangsmith/ibm-data-science-captsone/blob/main/4.1-Machine-Learning.ipynb

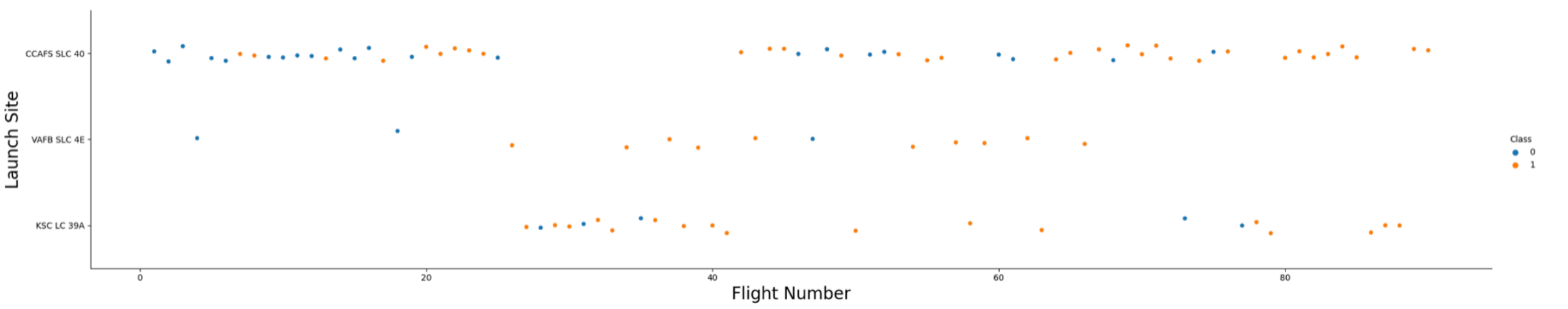
Results

- Exploratory Data Analysis Results:
  - Space X launched at 4 different sites.
  - Most mission outcomes were successful.
  - The first successful landing outcome happened in 2015
  - The number of successful landing outcomes increased as years passed.

Section 2

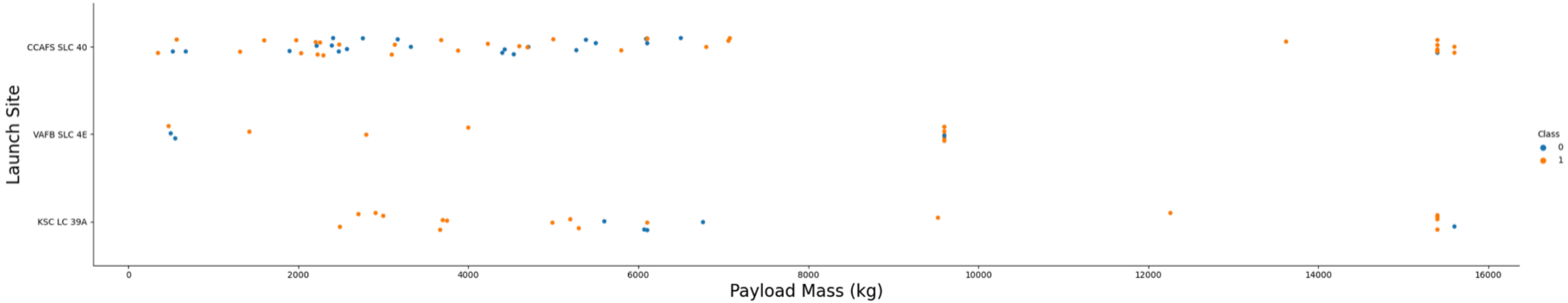
Insights drawn from EDA

Flight Number vs. Launch Site



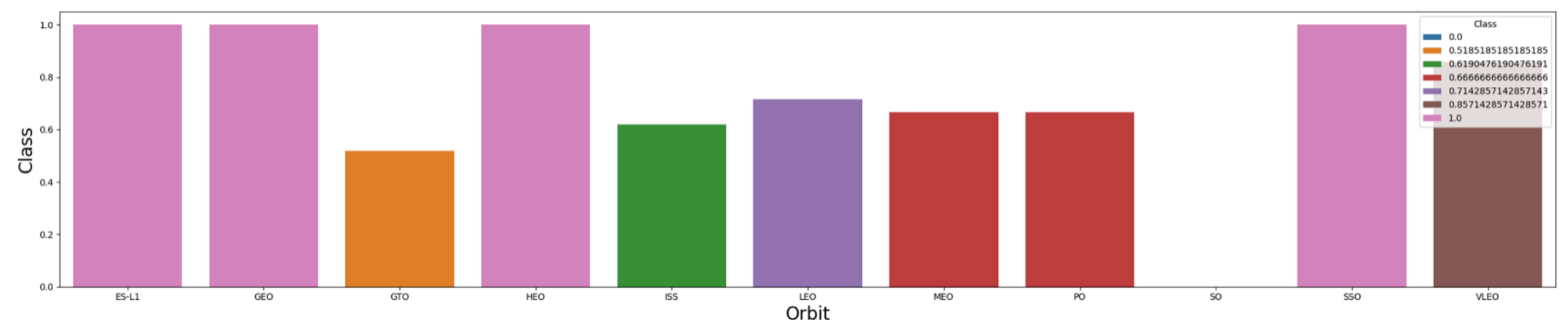
- According to the plot, CCAFS SLC 40 was the most successful of recent launches.
- We see that over time, successful launches increase

Payload vs. Launch Site



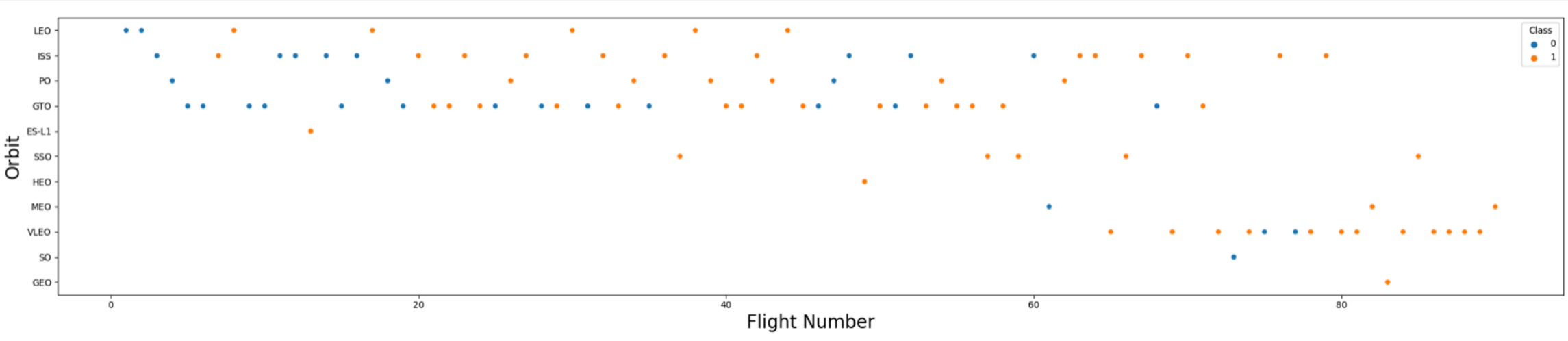
- Heavier payloads tend to be more successful

Success Rate vs. Orbit Type



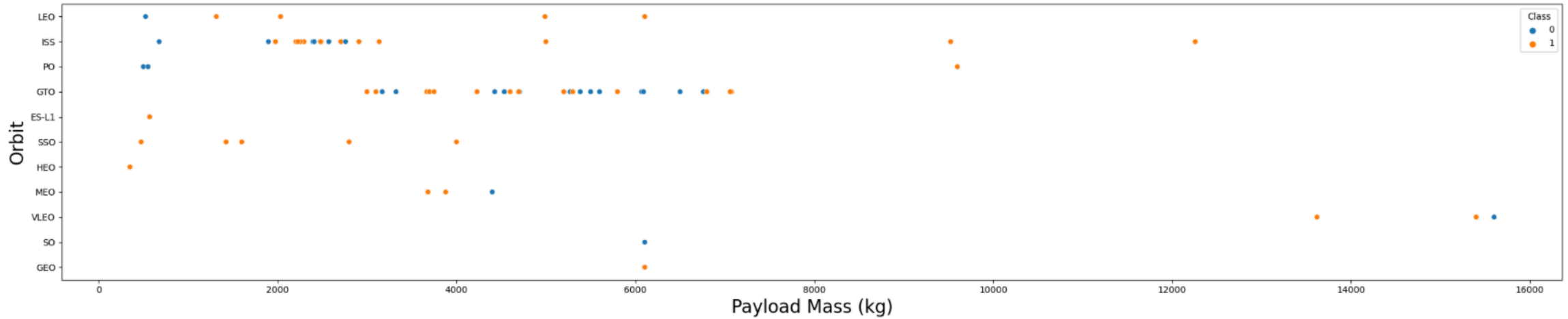
- HEO, SSO, ES-L1, and GEO orbit are the most successful.
- SO has not seen any successful launches.

Flight Number vs. Orbit Type



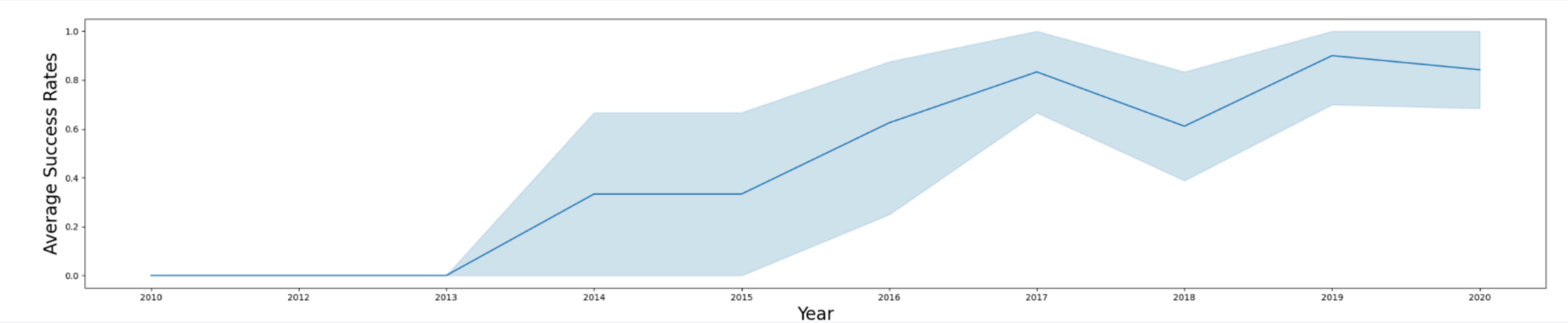
- Successful launches increase over time
- SO only had one flight (unsuccessful)

Payload vs. Orbit Type



- ISS has a wide range of payload mass launches and success
- GTO has good payload mass distribution as well

Launch Success Yearly Trend



- Average success increases steadily over time

All Launch Site Names

- According to the data, there are 4 unique launch sites

Launch Site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Lan
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Fai
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Fai
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	

- We can see the first 5 results of the CCAFS LC-40 Launch Site

Total Payload Mass

SUM("PAYLOAD_MASS__KG_")
99980.0

- The total mass in KG launched by NASA

Average Payload Mass by F9 v1.1

AVG("PAYLOAD_MASS__KG_")
6138.287128712871

- The average payload mass by the F9 v1.1

First Successful Ground Landing Date

MIN("DATE")
01/07/2020

- First successful landing outcome

Successful Drone Ship Landing with Payload between 4000 and 6000

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing
05/01/2017	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300.0	LEO	NRO	Success	Succe
09/07/2017	14:00:00	F9 B4 B1040.1	KSC LC-39A	Boeing X-37B OTV-5	4990.0	LEO	U.S. Air Force	Success	Succe
01/08/2018	1:00:00	F9 B4 B1043.1	CCAFS SLC-40	Zuma	5000.0	LEO	Northrop Grumman	Success (payload status unclear)	Succe

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

COUNT("Landing_Outcome")
10

- 10 Failures out of 90 Entries

Boosters Carried Maximum Payload

- Listed boosters with 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

2015 Launch Records

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
Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40

- 2015 Launch Records

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

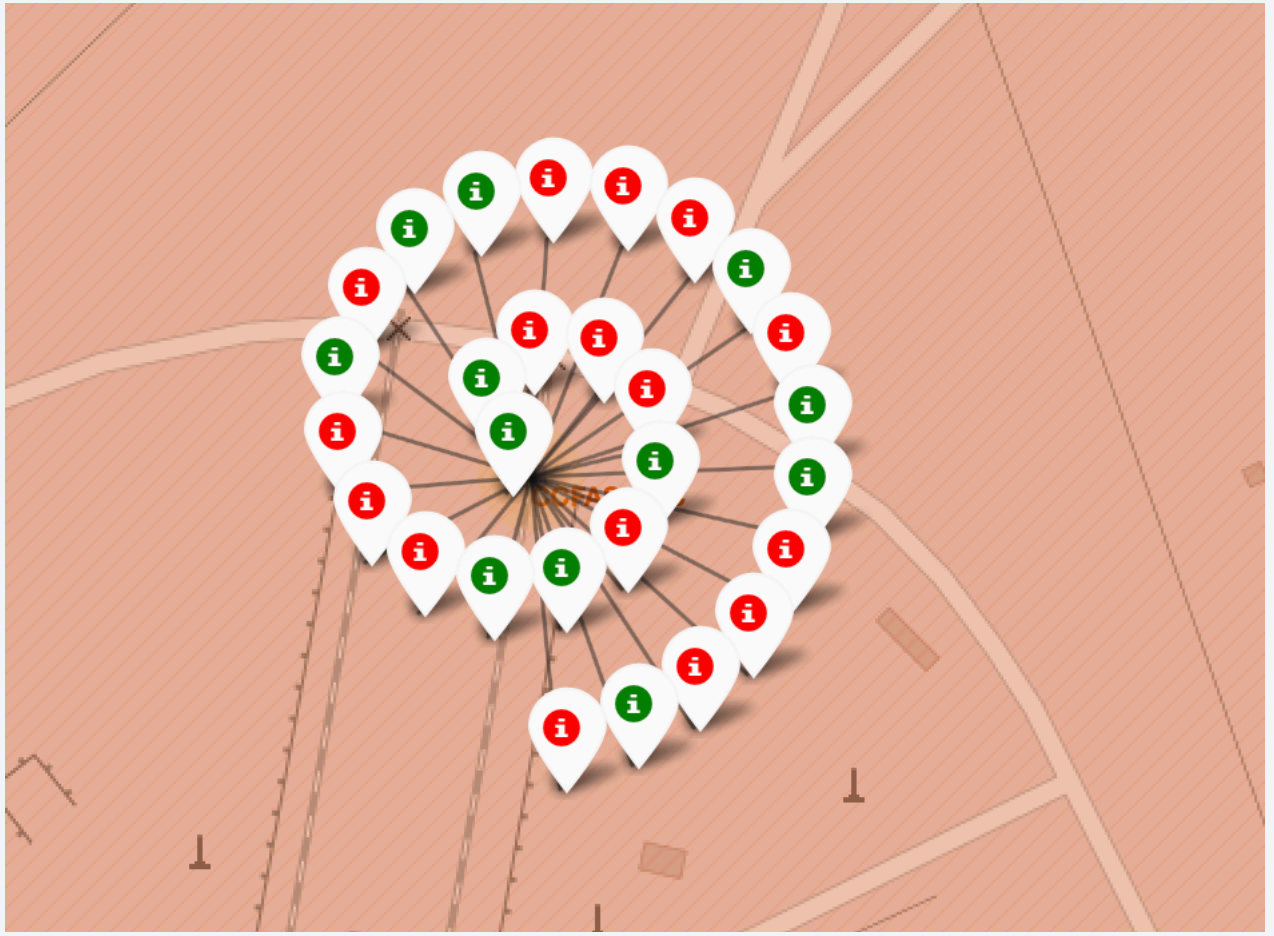
Landing_Outcome	Date
Success (drone ship)	04/08/2016
Success (drone ship)	05/06/2016
Success (ground pad)	18/07/2016
Success (drone ship)	14/08/2016
Success (drone ship)	14/01/2017
Success (ground pad)	19/02/2017
Success (ground pad)	05/01/2017

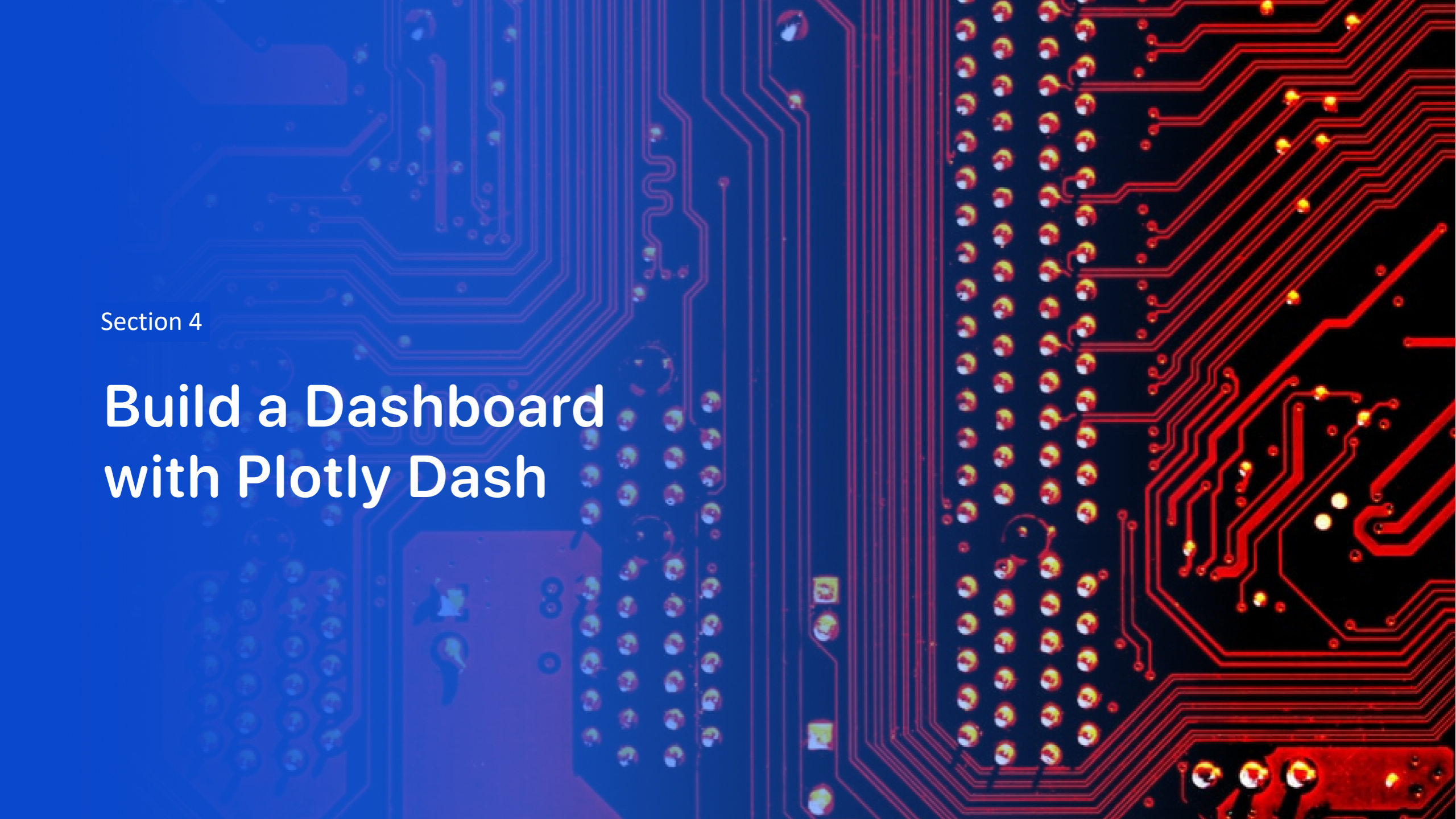
Section 3

Launch Sites Proximities Analysis

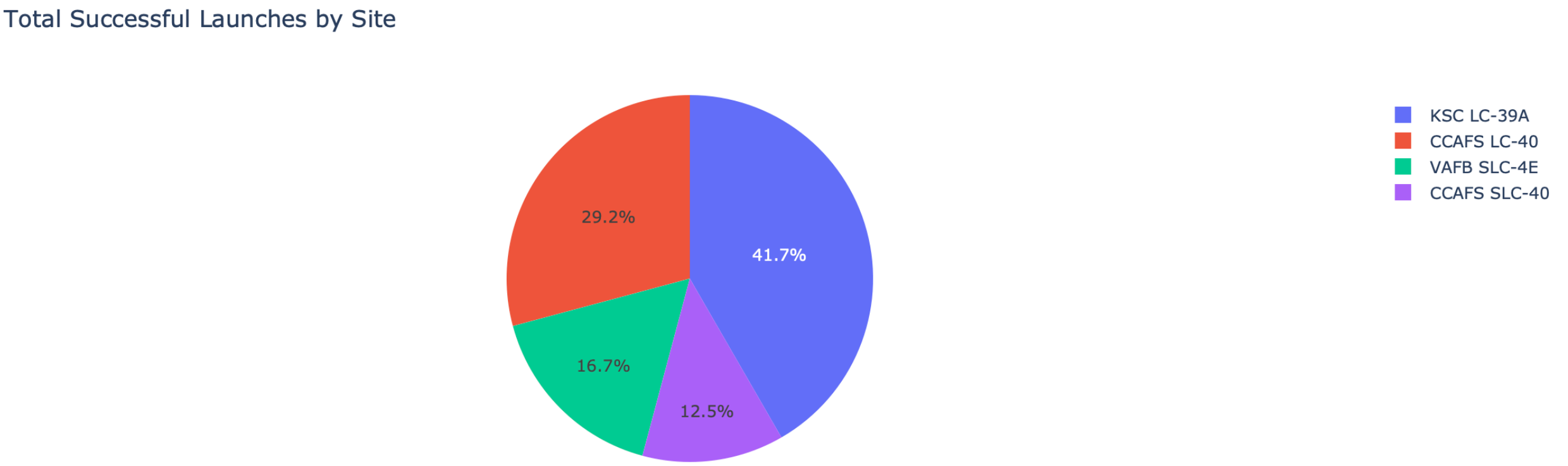
All Launch Sites

Launch Outcome by Sites

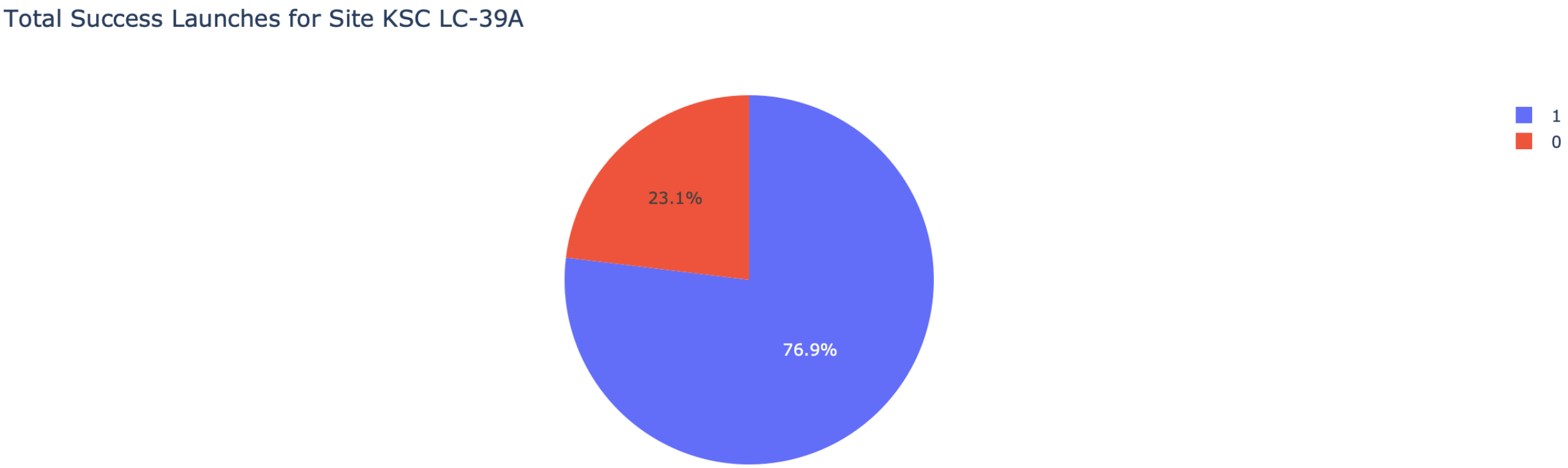



Section 4
Build a Dashboard with Plotly Dash

Total Successful Launches (All Sites)

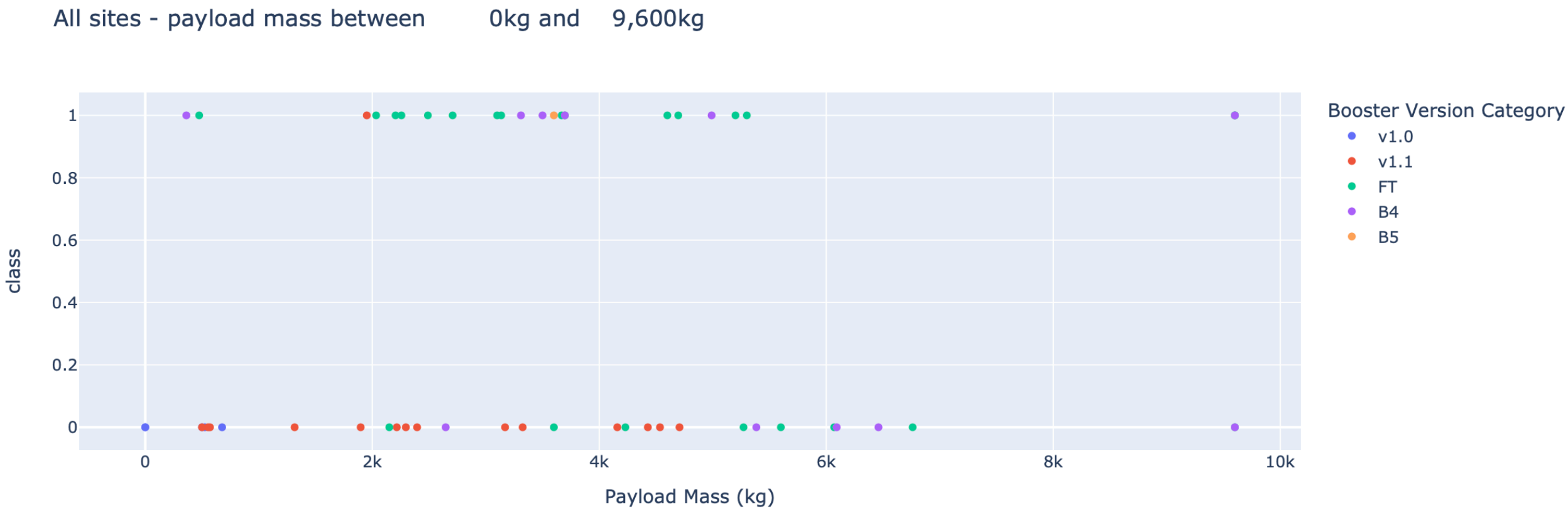


Total Successful Launches (KSC LC-39A)



Payload Range

Payload range (Kg):

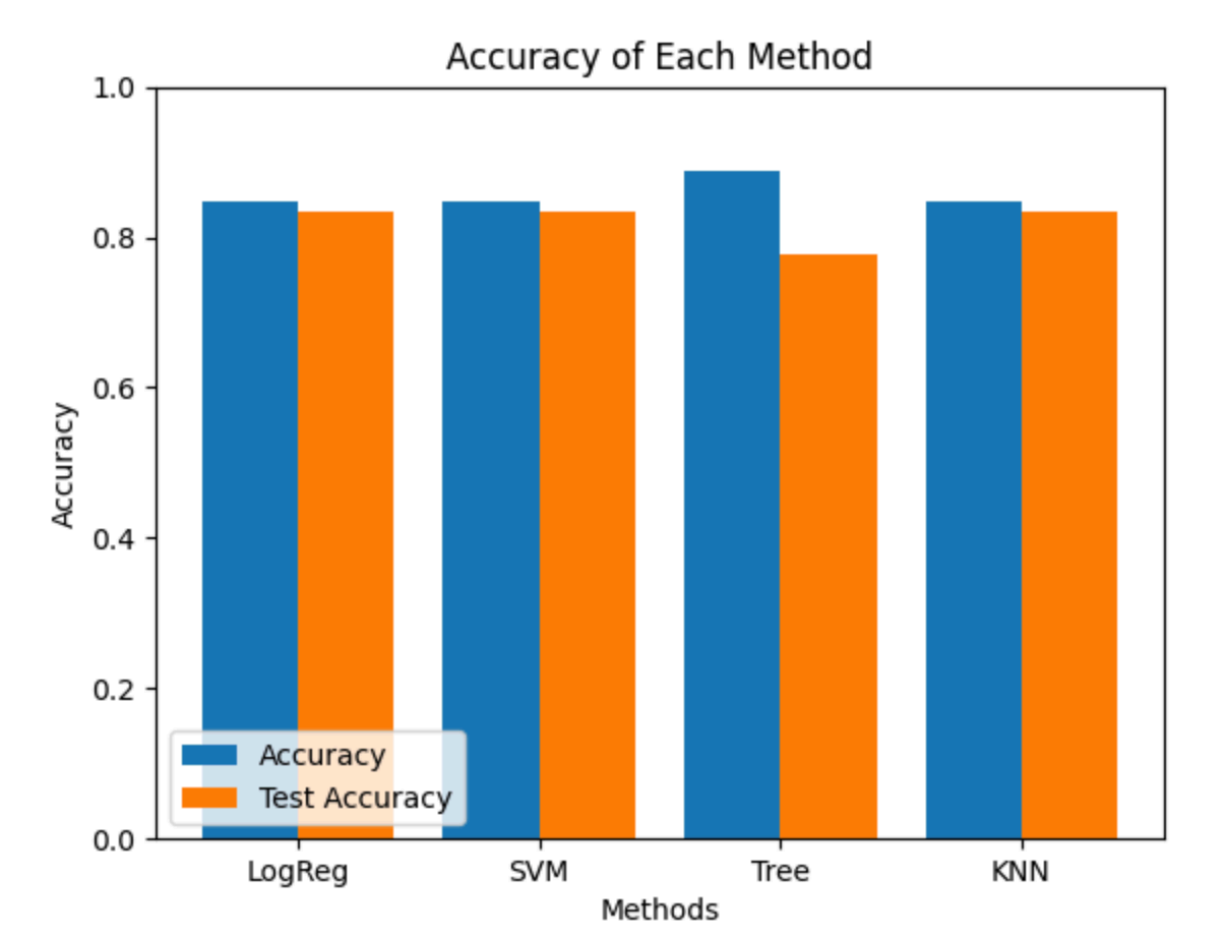


Section 5

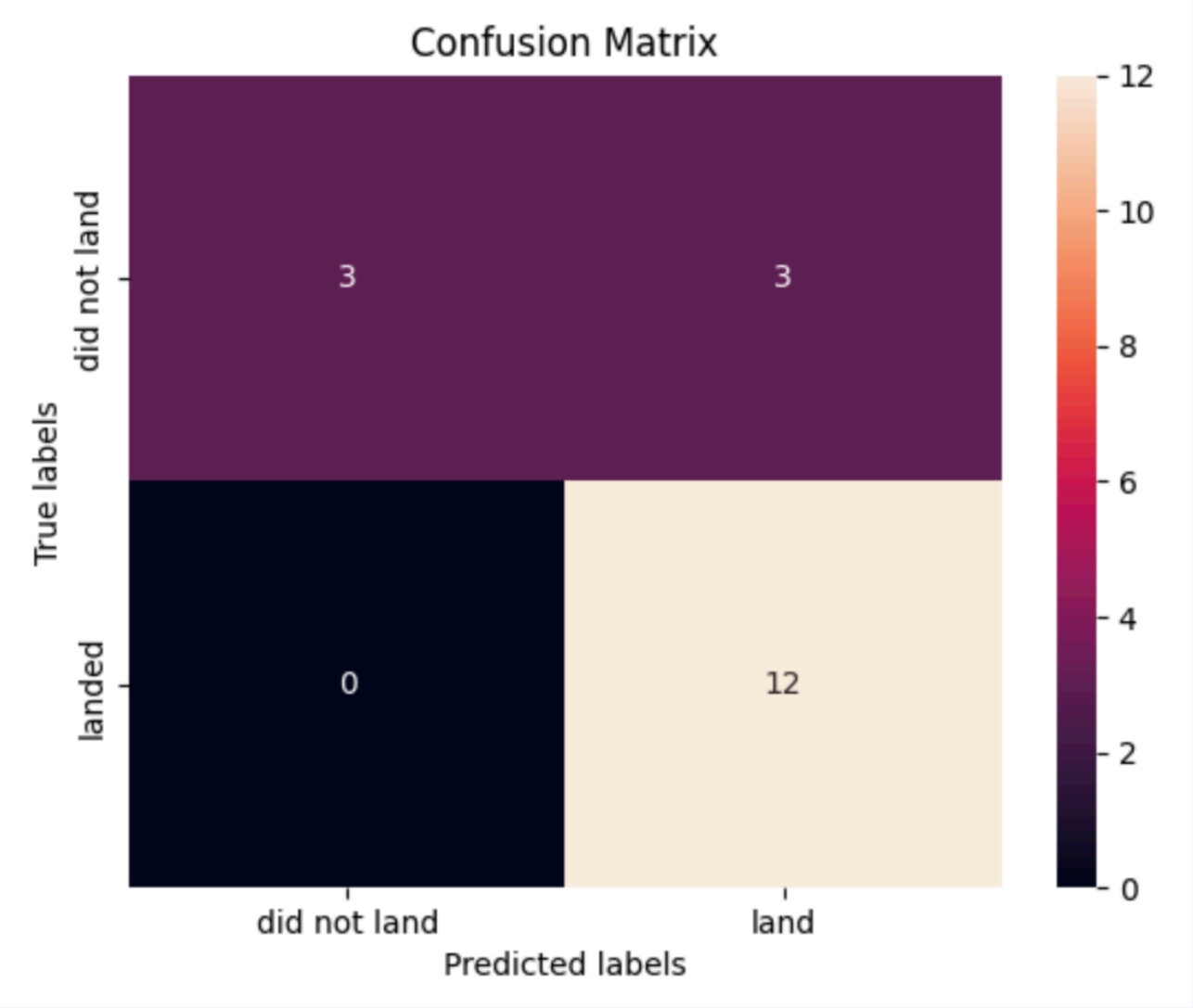
Predictive Analysis (Classification)

Classification Accuracy

- KNN Had the highest testing accuracy



Confusion Matrix (KNN)



Conclusions

- Different data sources were analyzed
- Best launch site: KSC LC-39A
- Launched over 7000kg were less risky
- We see improvement over time
- KNN has the best prediction for successful landings

Thank you!