IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Devan Scholefield
December 2023

Outline

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

Executive Summary

- **Summary of Methodologies**
  - Data collected from the SpaceX public API and Wikipedia
  - Data wrangling to extract launch outcome as the outcome variable
  - SQL, data visualization, and an interactive Folium map were used to discover insights about the dataset
  - A dashboard was built with Plotly Dash to display findings from the exploratory data analysis
  - Predictive analysis (classification using logistic regression, SVM, decision tree, and KNN) was used to see if launch outcomes could be predicted
- **Summary of all Results**
  - Launch sites were picked based on certain important factors
  - All four models perform relatively the same, with decision tree having the highest accuracy

Introduction

- **Project Background and Context**
  - SpaceX advertises the Falcon 9 rocket launches on their website at a cost of $62 million per launch. Other providers cost upwards of $165 million per launch. Much of these savings are due to the fact that the first stage of the Falcon 9 is reusable.
- **Problems you want to Find Answers to**
  - Hence, the project task is to predict if the first stage of the SpaceX Falcon 9 rocket will land successfully.

Section 1

Methodology

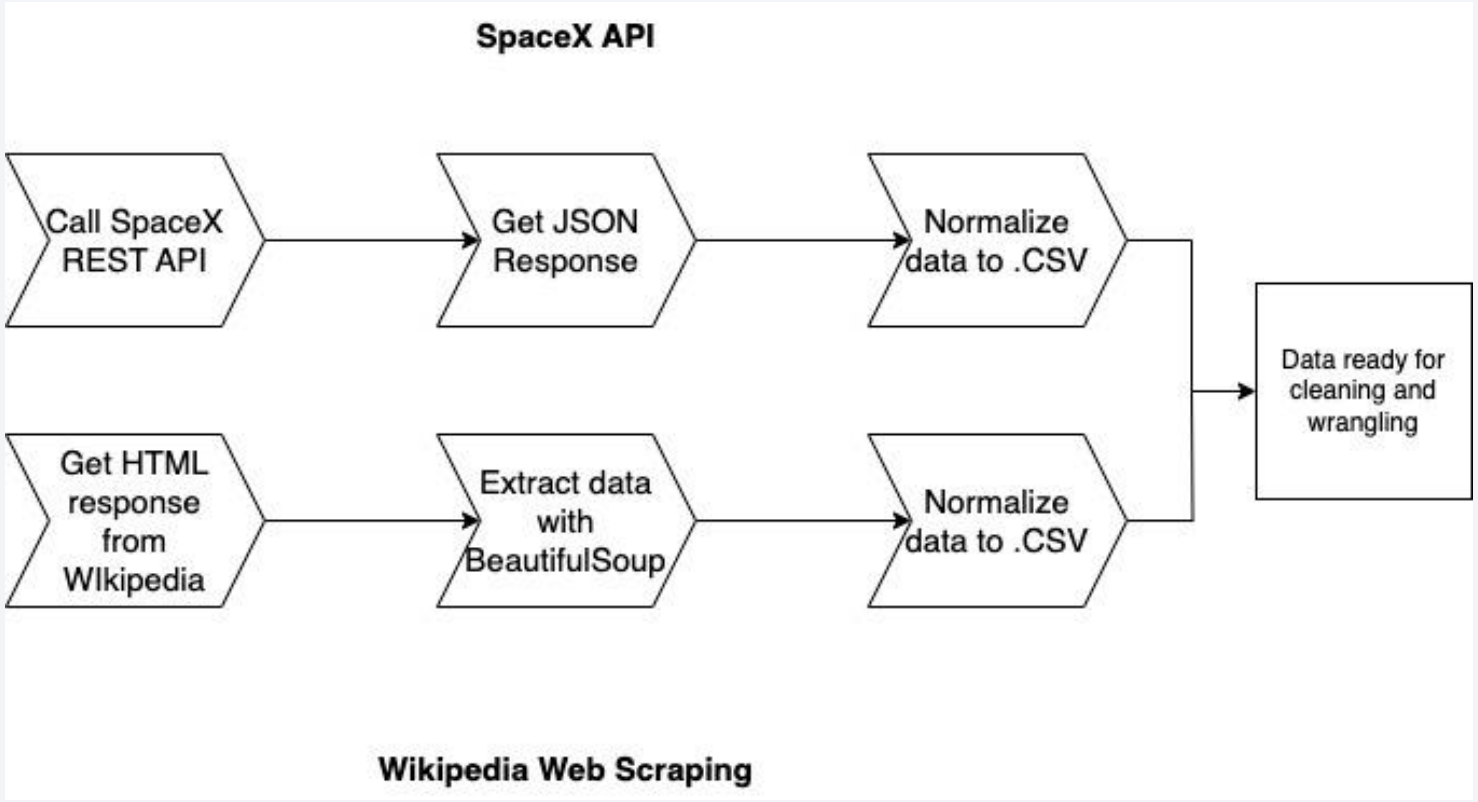
Methodology

Executive Summary

- Data collection methodology:
  - SpaceX Rest API
  - Wikipedia Web Scraping
- Data wrangling
  - One hot encoding to convert categorical data into a format for machine learning, removal or imputation of missing data, and removal of irrelevant columns
- Exploratory data analysis (EDA) using visualization and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis using classification models
  - Logistic Regression, KNN, SVM, and Decision Tree models were built and measured for the best classifier

Data Collection

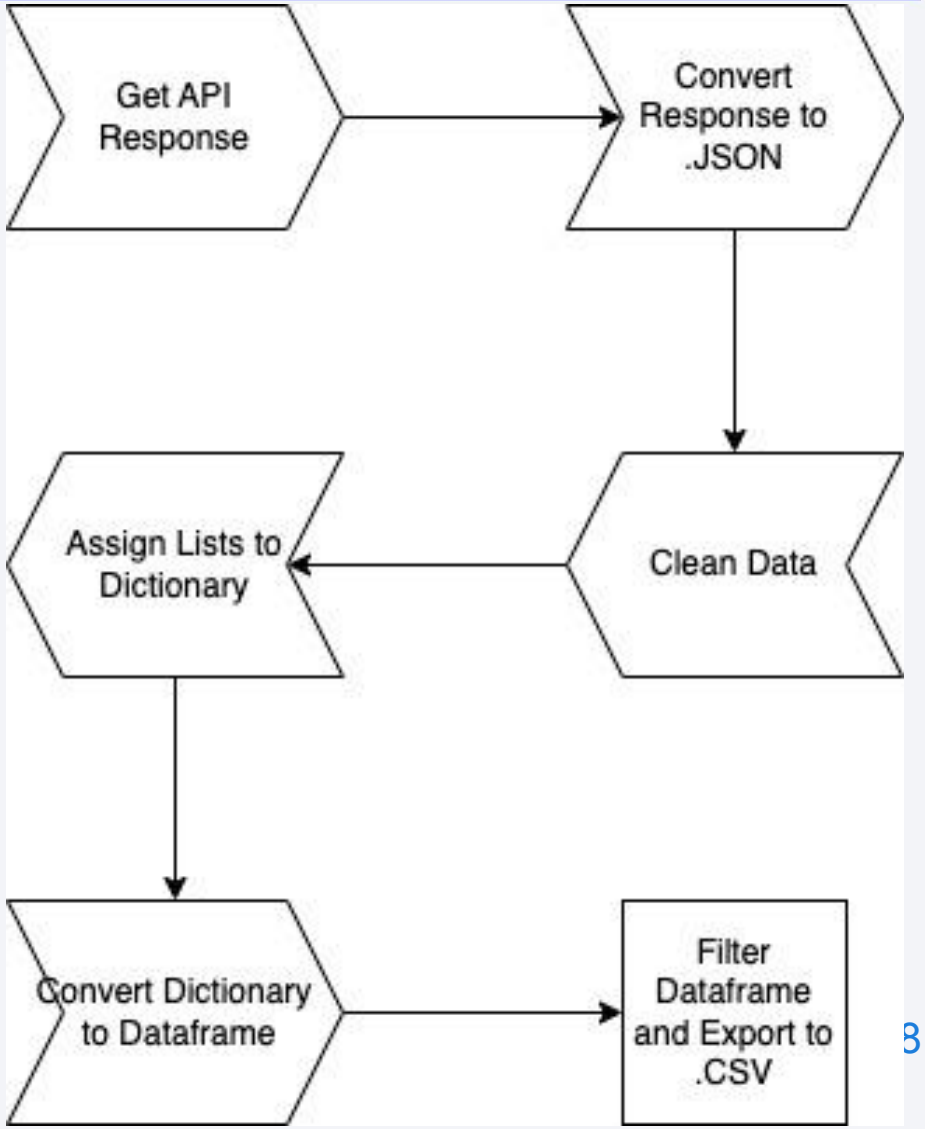
- Data was obtained from the SpaceX REST API and web scraped from Wikipedia



Data Collection – SpaceX API

- Data collection with SpaceX REST calls
- Once data is received from the API it can be converted to a Panda dataframe object for further use

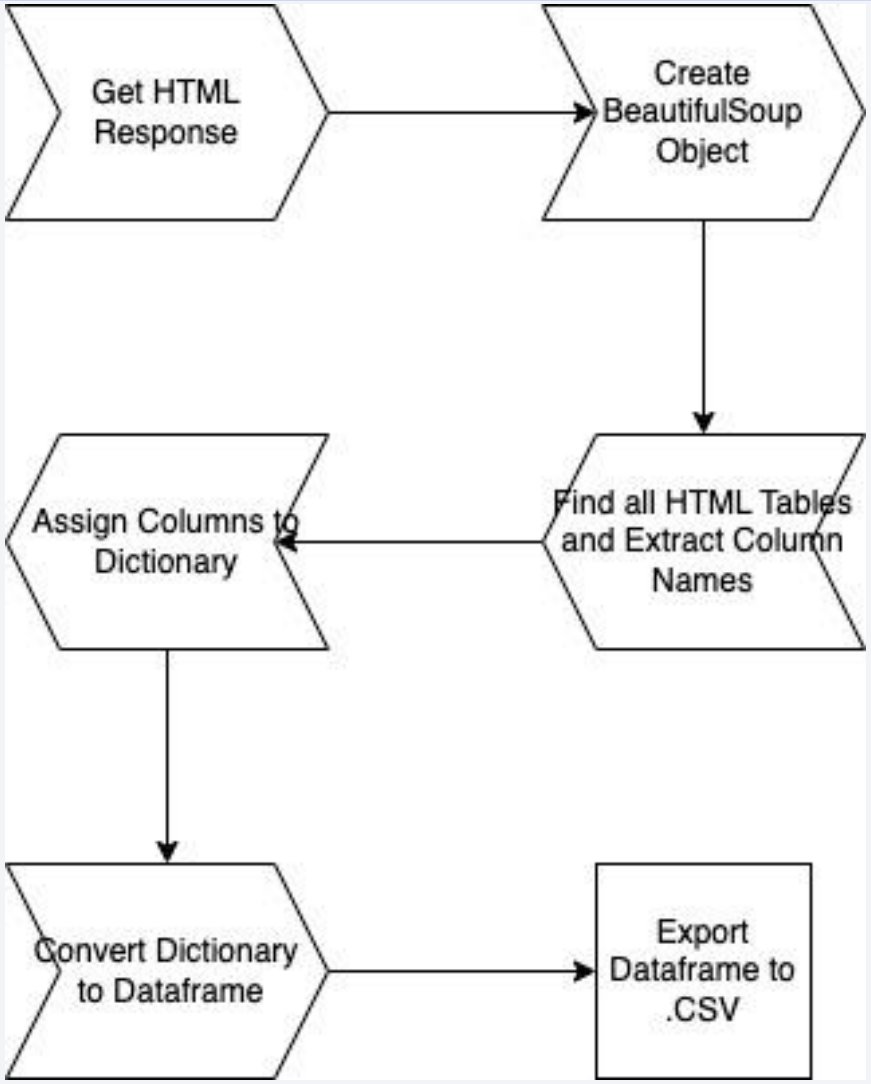
- GitHub URL



Data Collection - Web Scraping

- Data collection with web scraping from Wikipedia
- Tables were scraped and converted to a Pandas dataframe object for further use

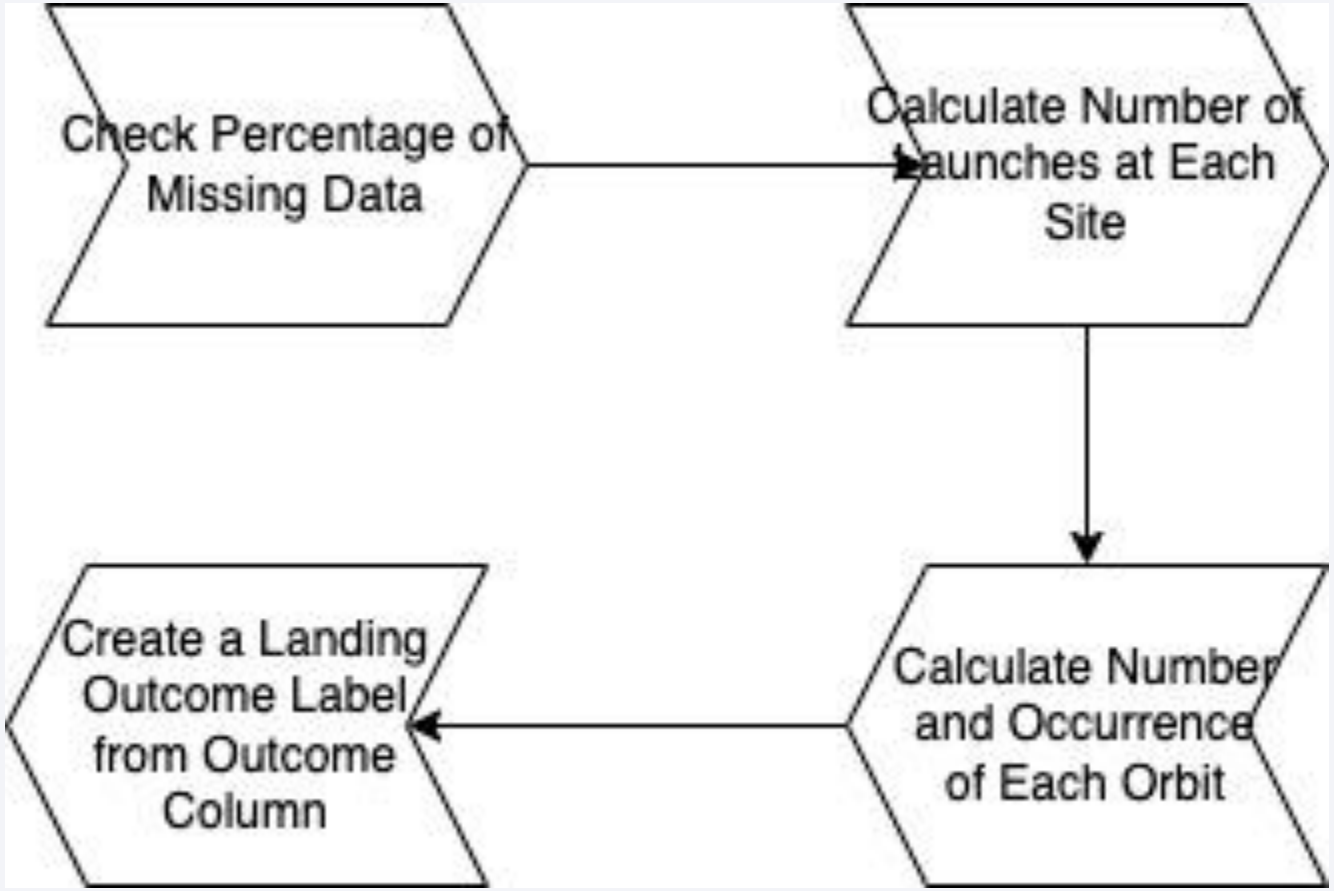
- GitHub URL



Data Wrangling

- Data was cleaned to deal with missing values
- Mission outcomes were converted to a binary classification with 1 being successful and 0 being failure

- GitHub URL



EDA with Data Visualization

- Scatter plot of Payload Mass (kg) vs Flight Number, Outcome overlayed
  - Visualize effect of mass and flight number on outcome
- Scatter plot of Launch Site vs Flight Number, Outcome overlayed
  - Visualize site usage over time and outcome per site
- Scatter plot of Launch Site vs Payload Mass (kg), Outcome overlayed
  - Visualize relationship between launch site and payload mass
- Bar chart of Success Rate vs Orbit Type
  - Visualize if there is a relationship between the orbit type and success
- Scatter plot of Orbit Type vs Flight Number, Outcome overlayed
  - Visualize if there is a relationship between orbit type and flight number
- Scatter plot of Orbit Type vs Payload Mass (kg)
  - Visualize if there is a relationship between orbit type and payload mass
- Line chart of Success Rate vs Year
  - Visualize the yearly trend for success rate since 2010

- GitHub URL

EDA with SQL

- Display names of unique launch sites
- Display 5 records with launch site starting with “CCA”
- Display total payload mass carried by booster launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List date when first successful landing outcome in ground pad was achieved
- List names of boosters with 4000<payload mass<6000 (kg) and drone ship success
- List total success and failures of missions
- List names of boosters versions which have carried the maximum payload mass
- List month names, failure landing outcomes in drone ship, booster versions, and launch site for 2015
- Rank count of landing outcomes between 2010-06-04 and 2017-03-20

- GitHub URL

Build an Interactive Map with Folium

- All launch sites were marked on the map with a circle
- Successes and failures were marked for each launch site with green and red icons respectively
- Lines were added to see the proximity of launch sites to points of interest
    - Points of interest included cities, coastline, highways, and railways
- GitHub URL

Build a Dashboard with Plotly Dash

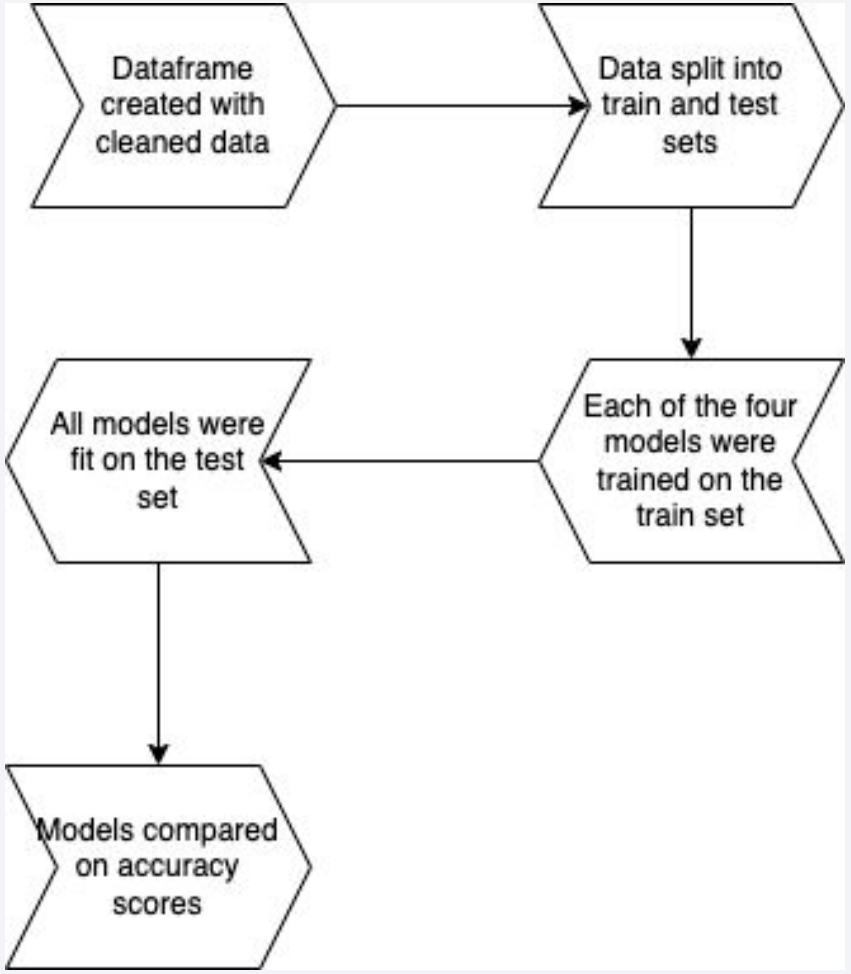
- Pie chart of number of launches per site
  - Each launch site selection contains a pie chart of success vs failures for that specific launch site
- Scatterplot of launch outcome vs payload mass (kg) with launch site overlaid
  - The plot will only display outcomes from the selected launch site, or “all sites”
  - A slider is given to adjust the displayed payload range

- GitHub URL

Predictive Analysis (Classification)

- Data was standardized and partitioned into train and test sets
- The following models were built using the sklearn library and fit on the train set
    - Logistic regression, SVM, decision tree, and KNN
- Model fit was evaluated using the score method on the test set
- A confusion matrix was used to evaluate each model for performance regarding false positives and false negatives

- GitHub URL

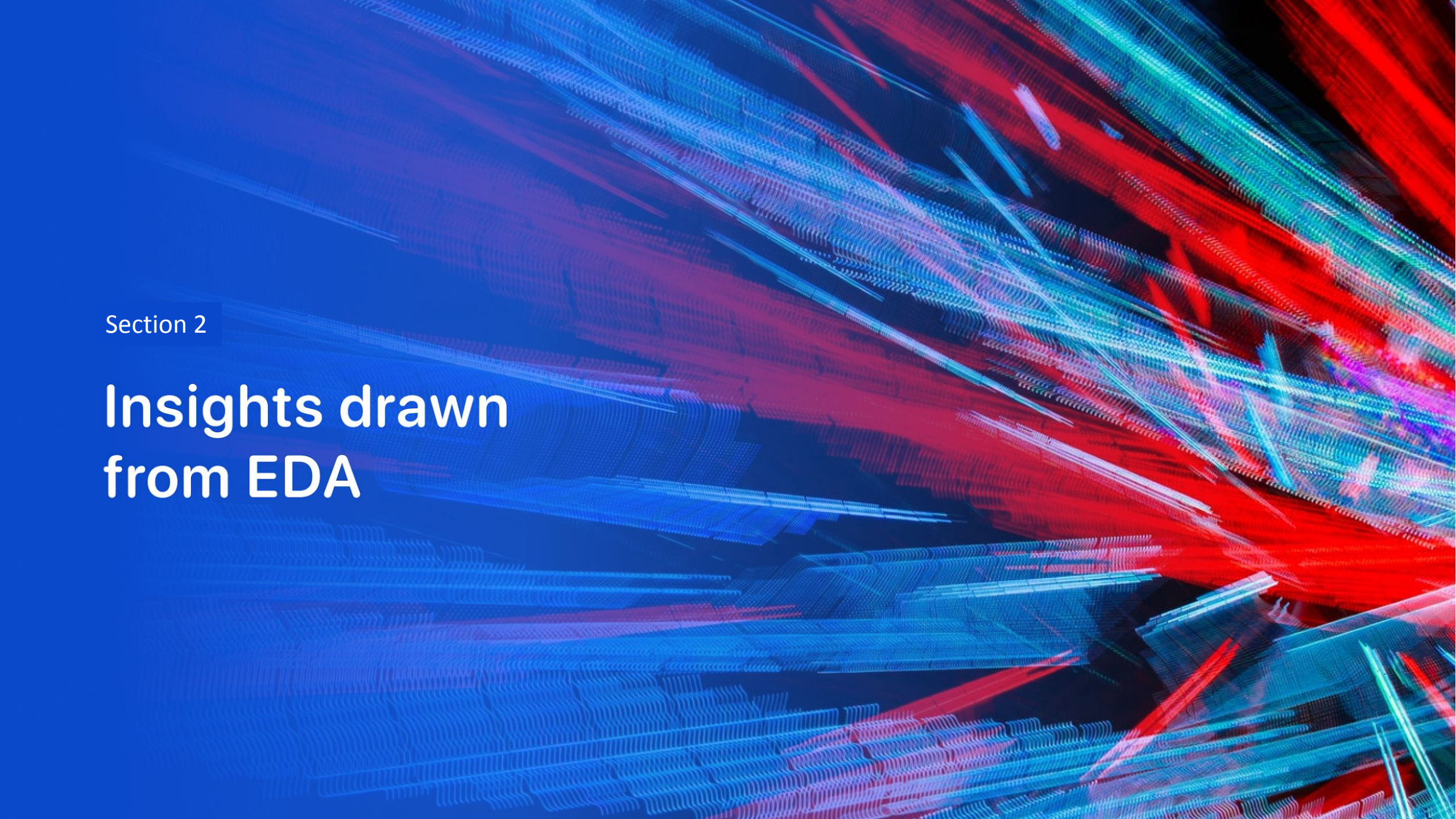


Results

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

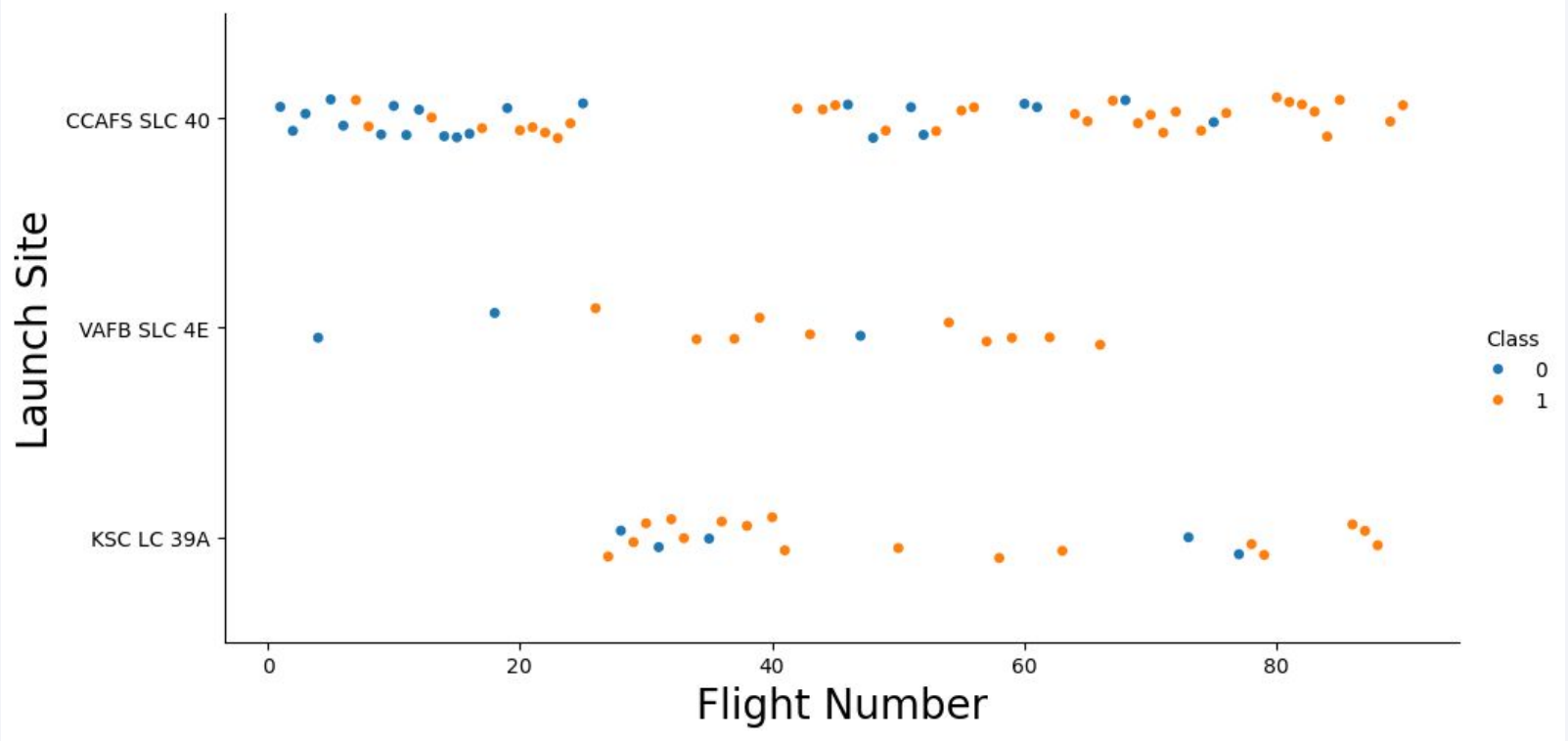
Section 2

Insights drawn from EDA



Flight Number vs. Launch Site

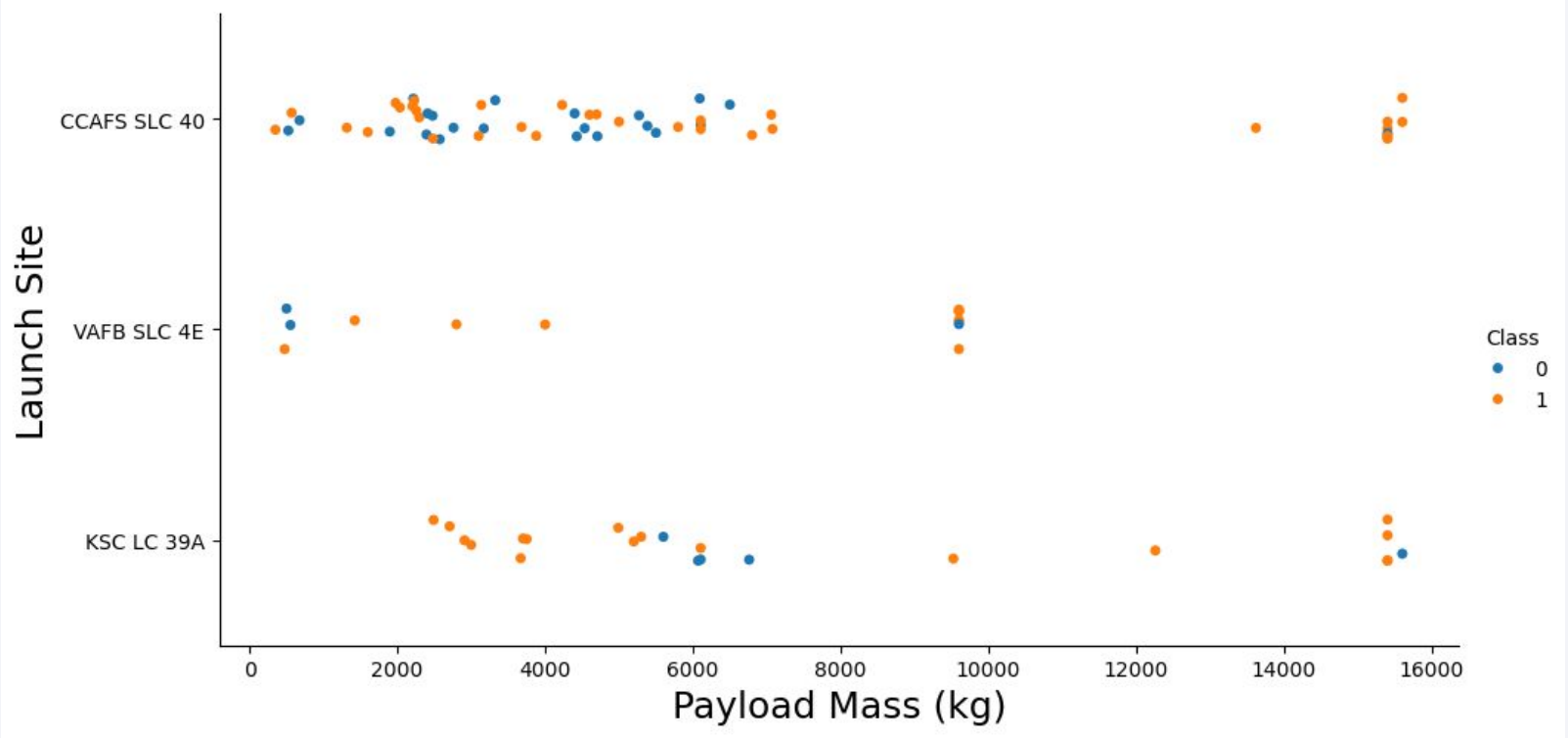
- Success rate varies by launch site
- It appears as time goes on (flight number increases) successful landings become more common



- Falcon 9 first stage successful landings are indicated by orange circles, while failed landings are indicated by blue circles

Payload vs. Launch Site

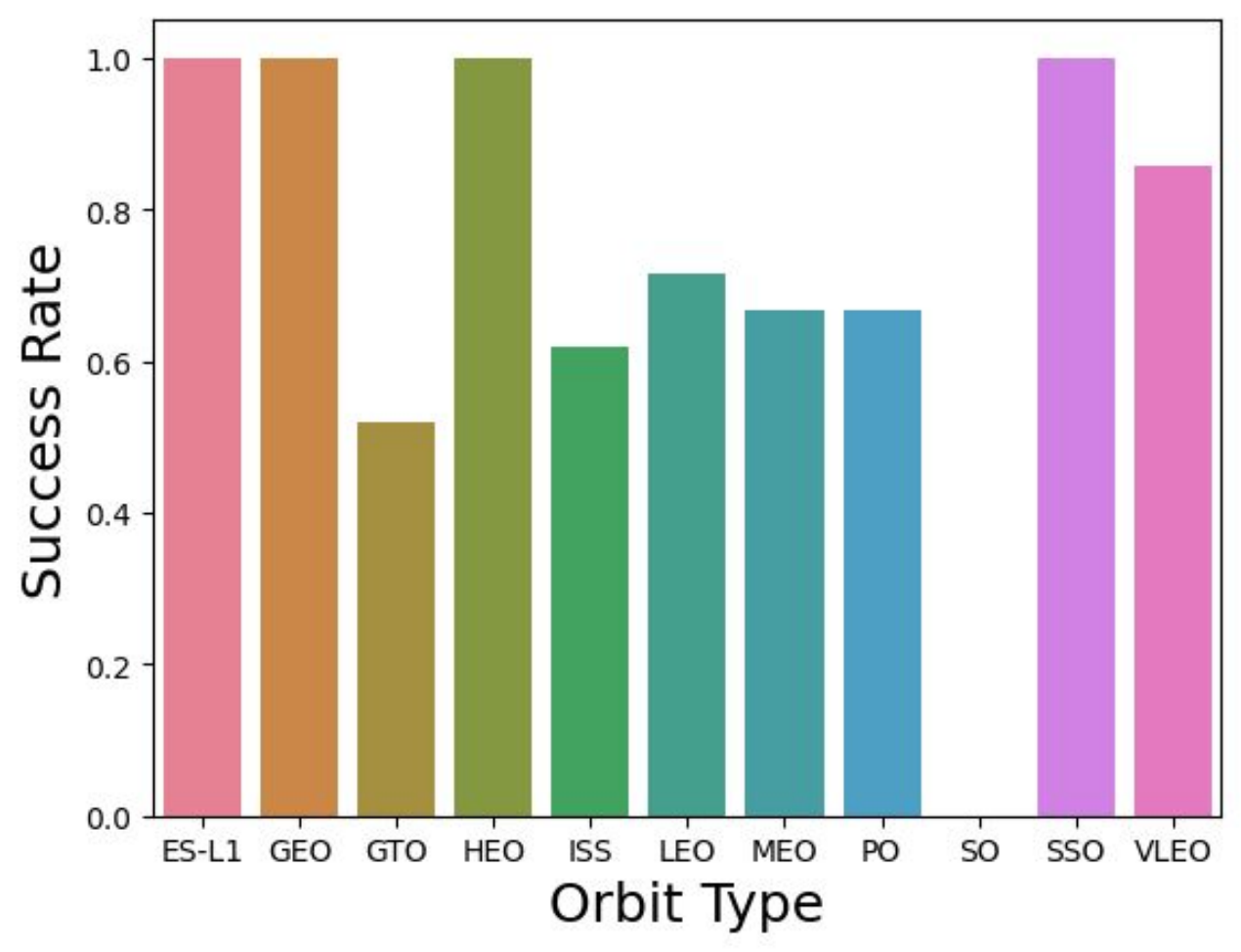
- Payload mass does not seem to strongly correlate with the outcome of Falcon 9 first stage landings



- Falcon 9 first stage successful landings are indicated by orange circles, while failed landings are indicated by blue circles

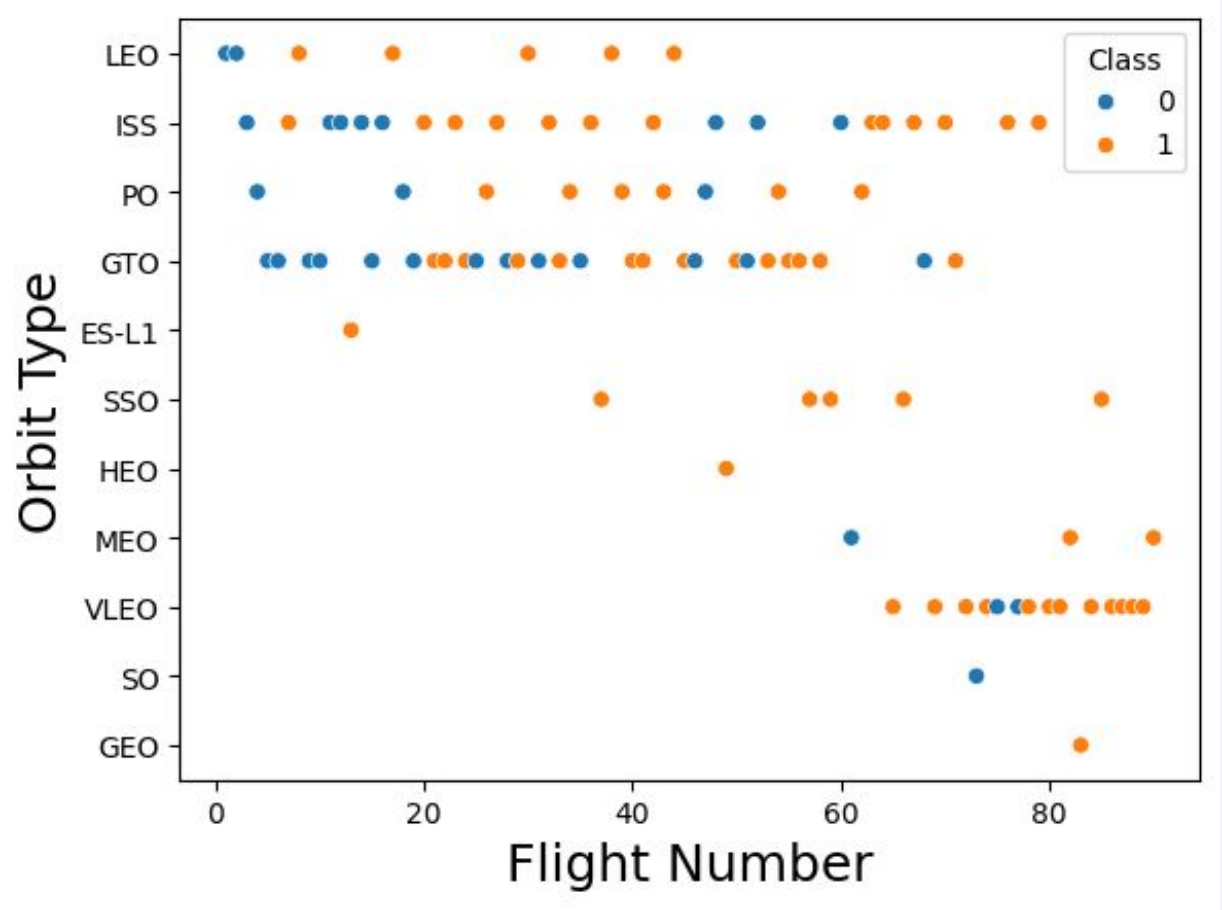
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, and SSO orbits have no failed first stage landing
- SO orbits have no successful first stage landings



Flight Number vs. Orbit Type

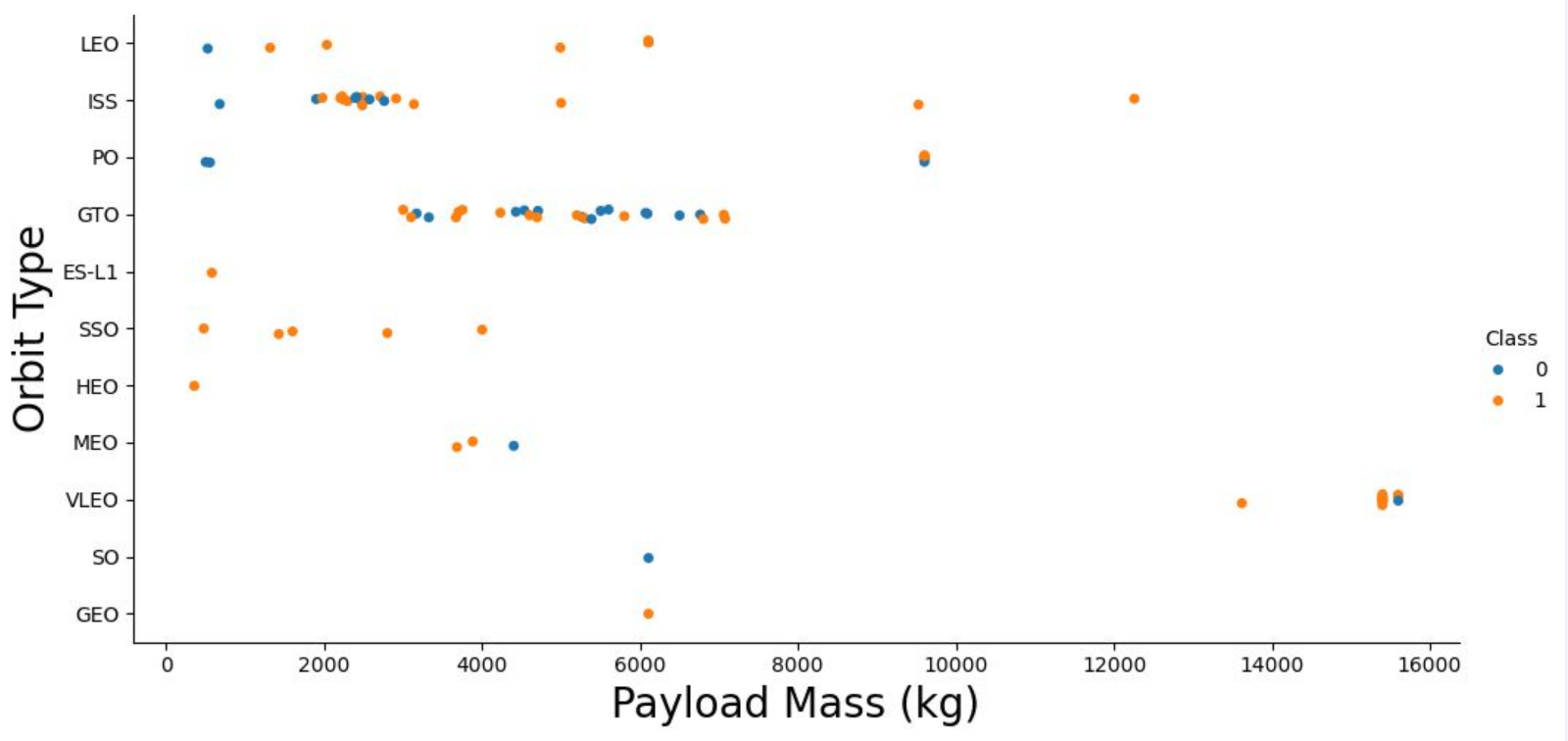
- It appears as time goes on (flight number increases) successful landings become more common



- Falcon 9 first stage successful landings are indicated by orange circles, while failed landings are indicated by blue circles

Payload vs. Orbit Type

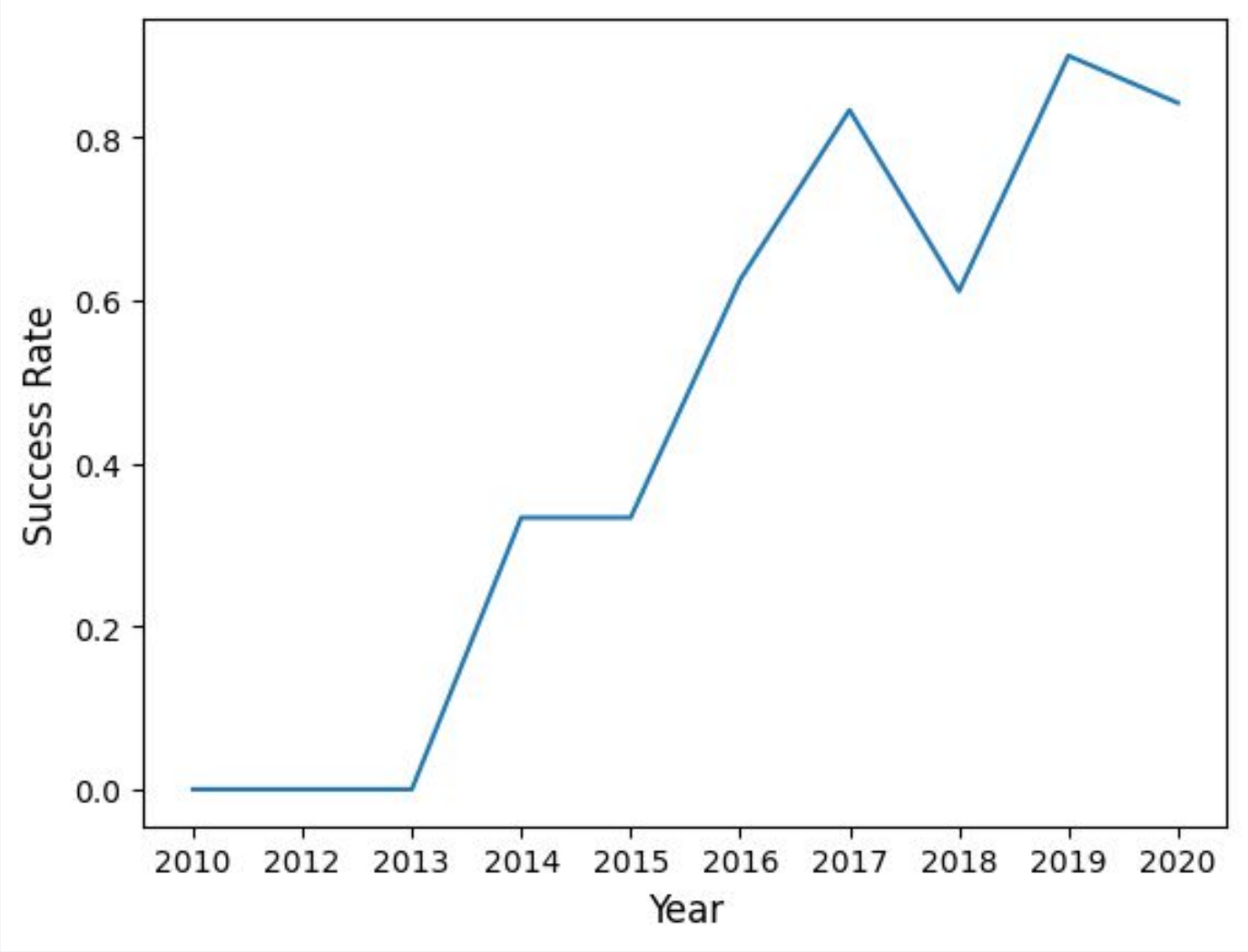
- Some orbit types have higher success rates that others
- Payload mass appears to have no correlation with success rate



- Falcon 9 first stage successful landings are indicated by orange circles, while failed landings are indicated by blue circles

Launch Success Yearly Trend

- Success rate has dramatically increased since 2010



All Launch Site Names

- There are four unique launch sites in the dataset

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

Launch Site Names Begin with 'CCA'

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

- Five launch sites that begin with “CCA”

Total Payload Mass

Total_Payload_Mass
45596

- The total payload mass carried by boosters launched by NASA is 45,596 kg

Average Payload Mass by F9 v1.1

Average_Payload_Mass
2928.4

- The average payload mass carried by the F9 v1.1 is 2,928.4 kg

First Successful Ground Landing Date

MIN(Date)
2015-12-22

- The date of the first successful landing was December 22nd, 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

- Four boosters have successfully landed on the drone ship with a payload mass between 4000 kg and 6000 kg

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

Success	Failure
61	40

- There were 61 successful missions and 40 failed missions

Boosters Carried Maximum Payload

- Twelve different boosters have carried the maximum payload of 15,600 kg

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

month	Booster_Version	Launch_Site	Landing_Outcome
10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- There were two failure landing outcomes in 2015

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

- Between 2010-06-04 and 2017-03-20 the most common landing outcome was “No attempt”

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

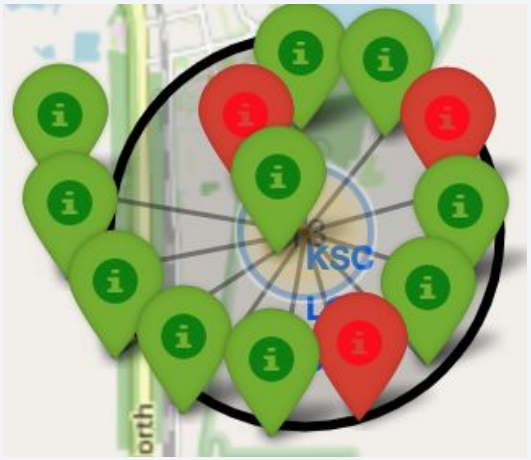

Section 3
Launch Sites
Proximities Analysis

Falcon 9 Launch Site Locations

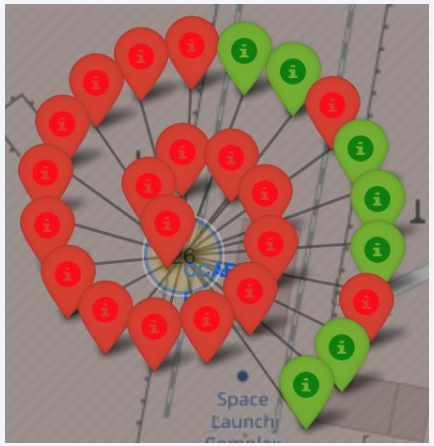
- There are 4 marked Falcon 9 launch locations, 3 in Florida and 1 in California
- VAFB SLC-4E (California)
- KSC LC-39A (Florida)
- CCAFS LC-40 (Florida)
- CCAFS SLC-40 (Florida)

Map Markers of Successful/Failed Landings

• VAFB SLC-4E



• KSC LC-39A



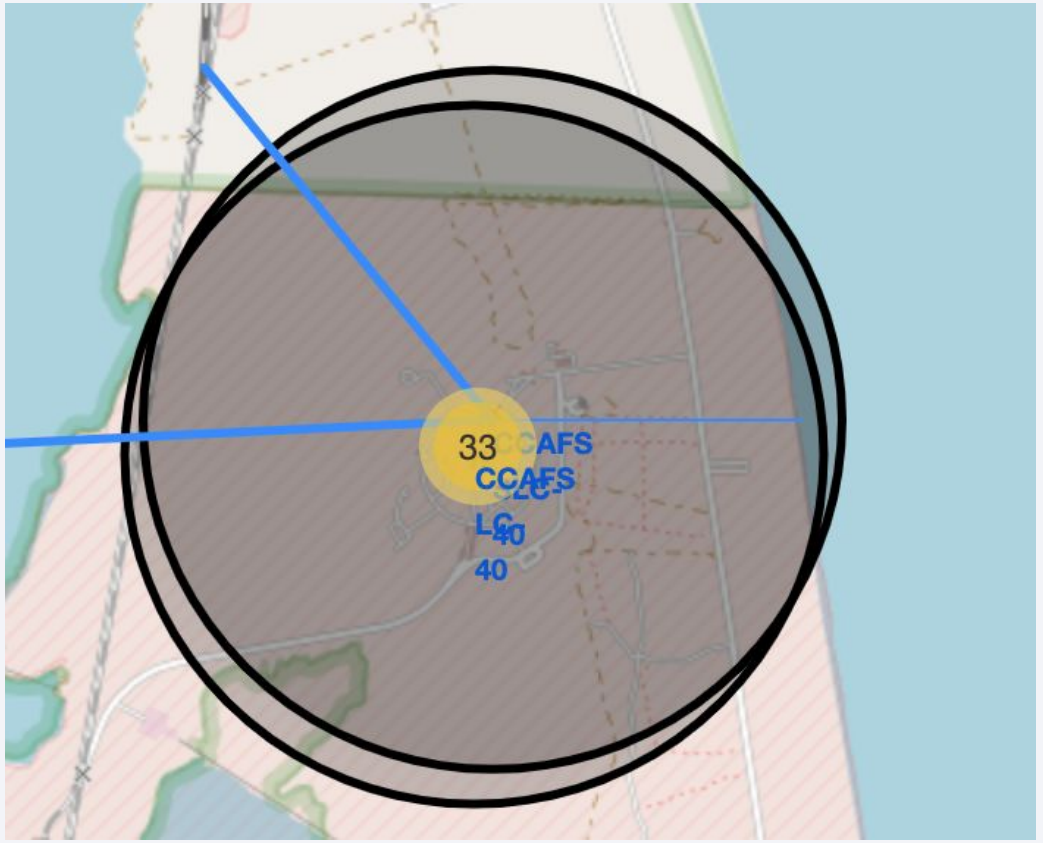
• CCAFS LC-40

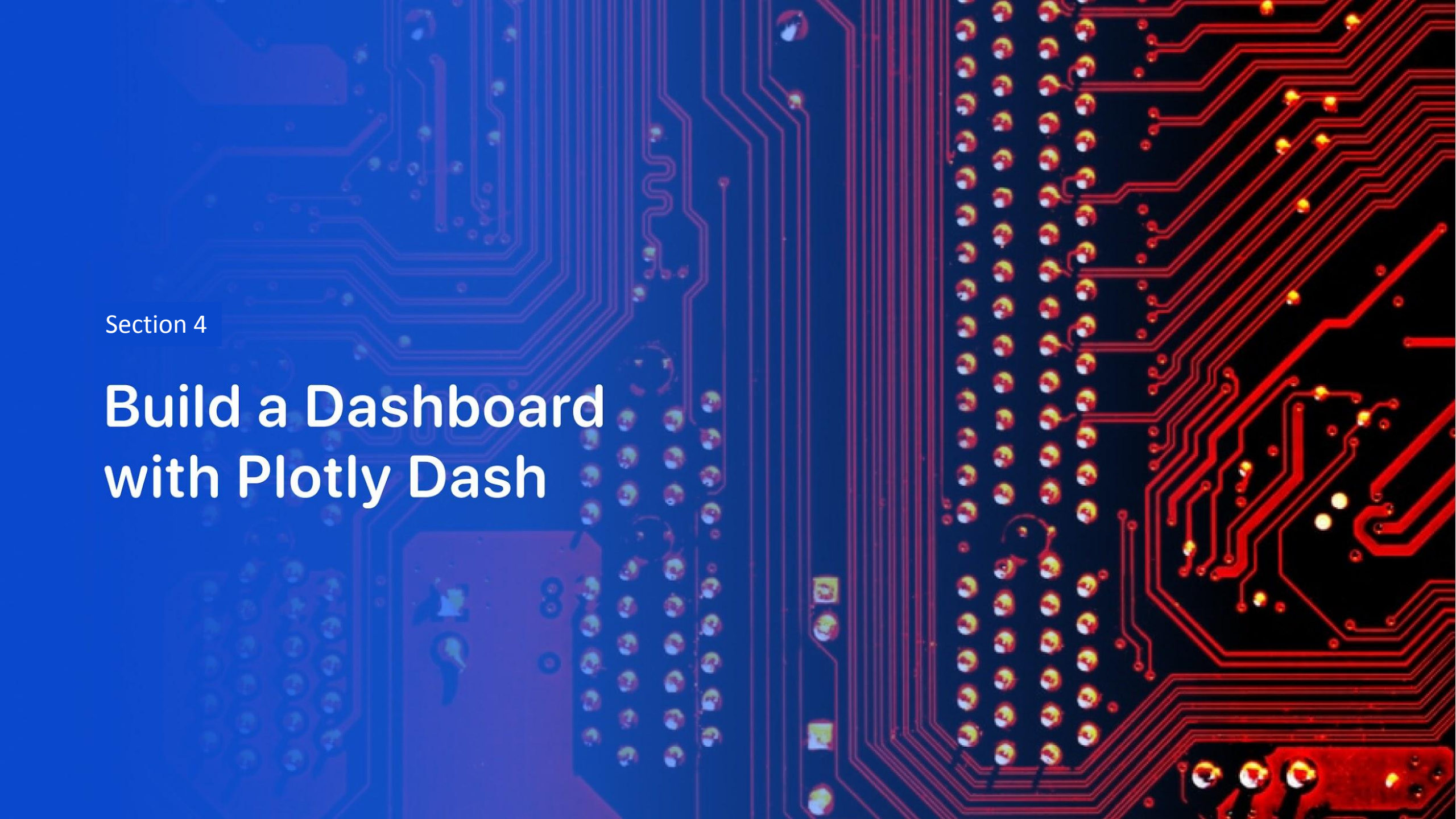
• CCAFS SLC-40

• The markers display successful and failed launches for each launch site

Distance from Launch Site to Proximities

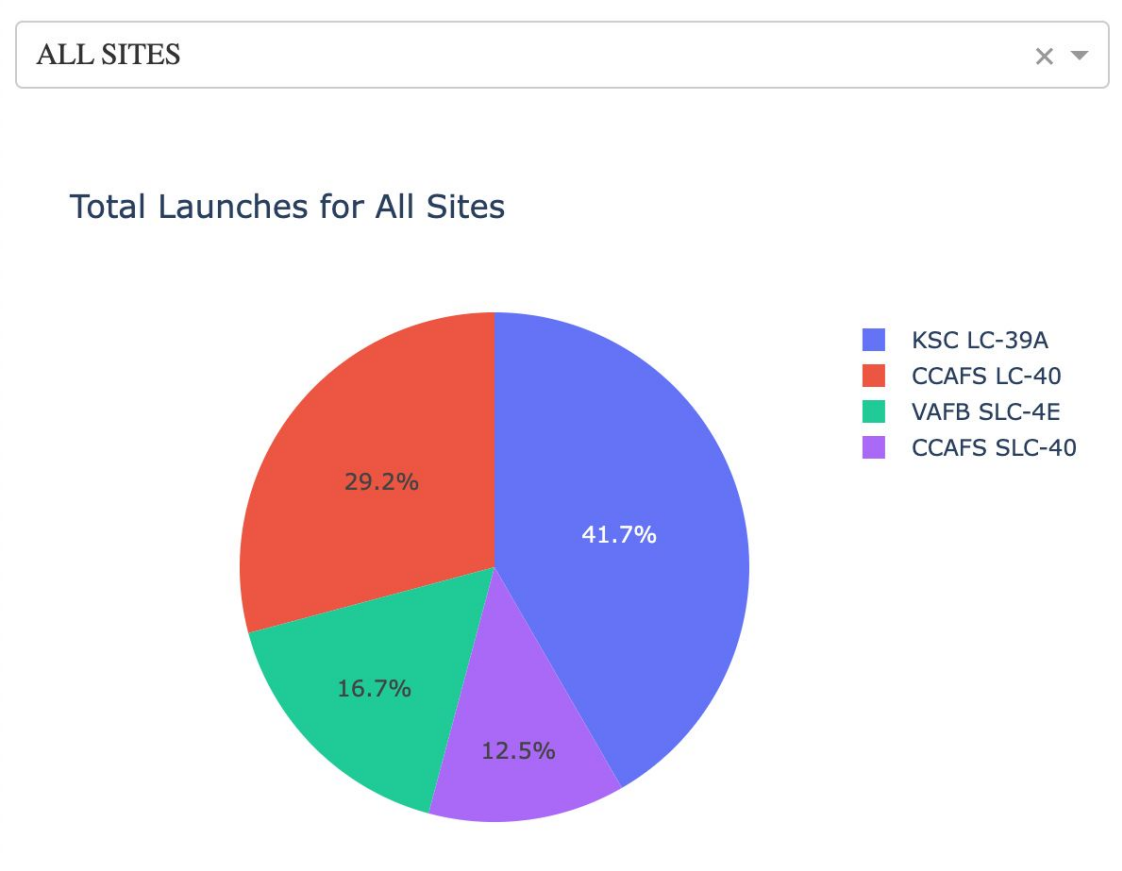
- Launch sites appear to be in close proximity to the equator and the coast
  - Requires less fuel due to Earth's rotational speed aiding the thrust of the rocket.
  - Minimizes debris causing injury or harm
- Launch sites are relatively close to highways and railways in order for quick transport of materials




Section 4
Build a Dashboard with Plotly Dash

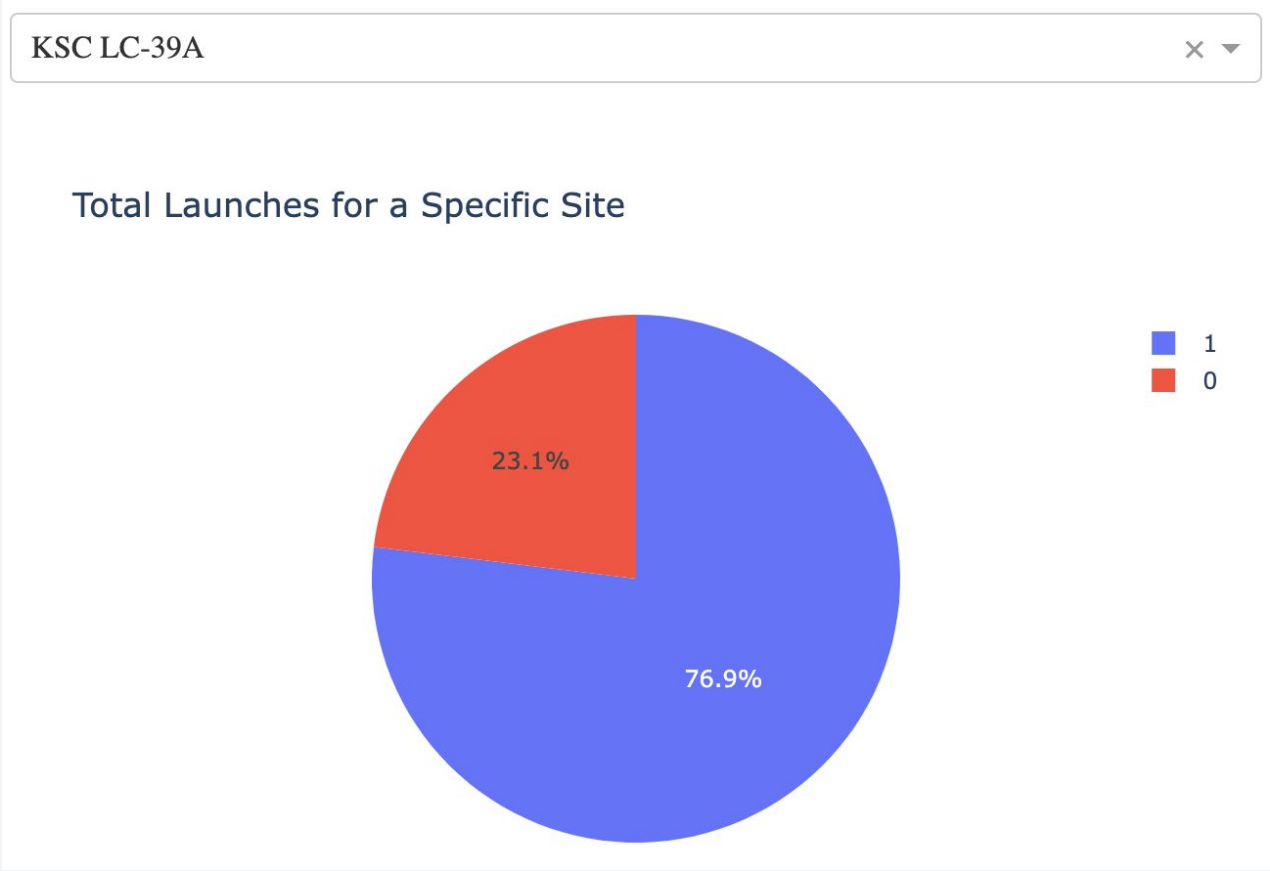
Successful Launches By Site

- With “ALL SITES” selected the pie chart displays the distribution of successful launches by site
- KSC LC-39A had the greatest proportion of successful launches

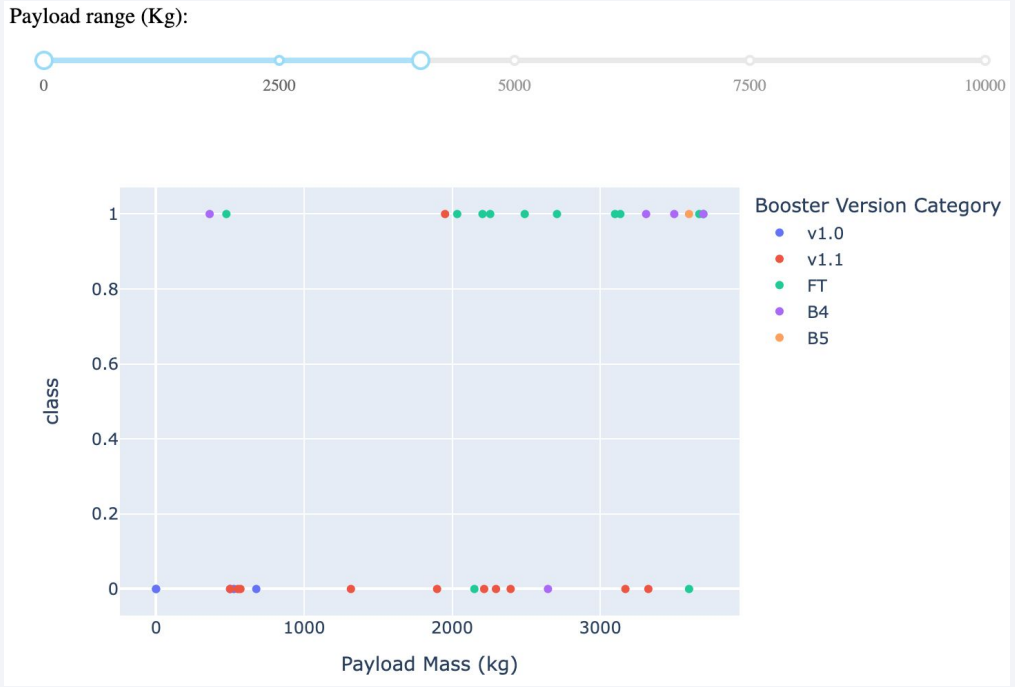


Launch Site with Highest Success Ratio

- Successful launches are indicated by the blue portion of the pie chart, while failed launches are indicted by the red portion
- The launch site KSC LC-39A had the highest success ratio with 76.9%



Scatterplot of Payload vs Launch Outcome



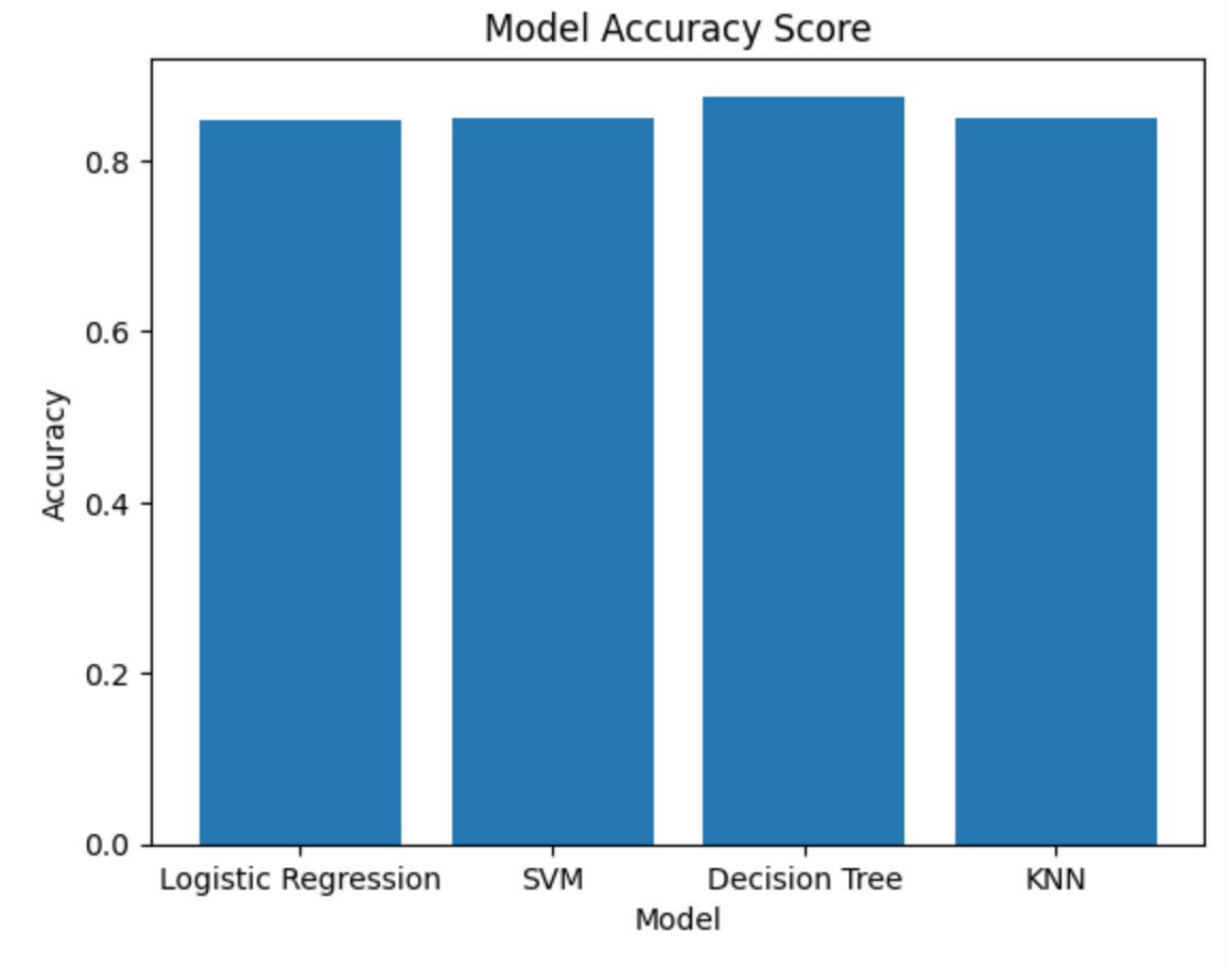
- Two different scatterplots are presented with different ranges of payload mass
- The “FT” booster category has the highest success rate

Section 5

Predictive Analysis (Classification)

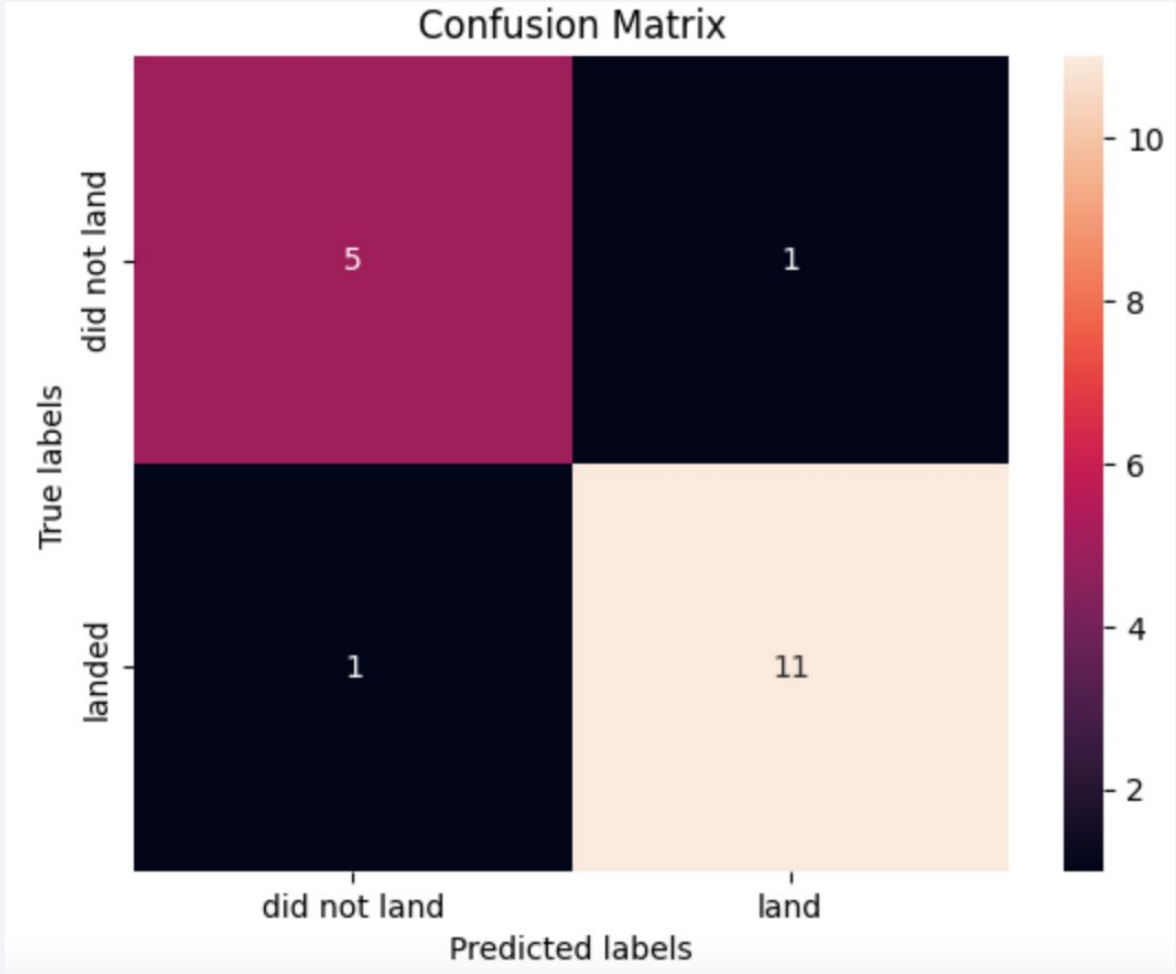
Classification Accuracy

- The Decision Tree Classifier performed slightly better with an accuracy of 0.875
  - All other models had the same accuracy of 0.848



Confusion Matrix

- Confusion matrix for the Decision Tree Classifier
- We observe 1 false negative in the bottom left corner and one false positive in the top right corner
- We observe 5 true negatives in the top left corner and 11 true positives in the bottom right corner



Conclusions

- SpaceX has improved their success rate as more and more launches have been made
- Different orbits have achieved differing success rates with ES-L1, GEO, HEO, and SSO orbits all having a 100% success rate
- KSC LC-39A had the greatest number of successful launches and the highest success rate out of all four launch sites
- The Decision Tree Classifier performed the best on the data with an accuracy score of 0.875
    - Thus we can conclude that machine learning models can indeed be used to accurately predict the success of Falcon 9 launches

Appendix

- All code and outputs can be found in this GitHub link

Thank you!