Winning Space Race with Data Science

Billy Chow
2022-5-22

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
    - Web Scraping
    - SpaceX API
  - Exploratory Data Analysis (EDA)
    - Data Wrangling
    - Data Visualization
    - Interactive Visual analytics (Dashboard)
  - Machine Learning Prediction
- Summary of all results
  - Data collected from public sources
  - EDA identified the best predictors for success rate of rocket launches
  - Machine Learning Prediction determined the best model for prediction of successful rocket launches

Introduction

- Background and Context
  - New company SpaceY would like to evaluate the viability of competing with existing company SpaceX in the production of reusable rockets
- Key Questions
  - Under what conditions are reusable rockets feasible?
  - What is the best method for predicting successful rocket landings
  - Where is the best location or launch site for rocket launches?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected via
    - SpaceX API (https://api.spacexdata.com/v4/)
    - Web Scraping

      (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)
- Perform data wrangling
  - The data was processed to feature a classification variable for successful and unsuccessful landing outcomes. Using this variable we can easily determine the success rate of SpaceX launches.

Methodology (Continued)

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - For the predictive analysis, processed data is first split in to training and testing sets. Using the training set, each of the four different classification models are fit and tuned to the best parameters. Finally, each model is evaluated based on accuracy score with the testing set.

Data Collection

- Data sets were collected through SpaceX's API (https://api.spacexdata.com/v4/) and using web scraping techniques on the Falcon 9 and Falcon Heavy Launch Wikipedia article (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches).

Data Collection – SpaceX API

- Data is collected from SpaceX's public API where information about the rocket boosters, payloads, launchpads, and cores are obtained.

URL: https://github.com/bichow/Coursera-Data-Science-Capstone/blob/master/Data%20Collection%20API.ipynb

Data Collection - Scraping

- Additional data is collected using BeautifulSoup to web scrape the Falcon 9 Launch Wikipedia page.

Request wiki page using an HTTP GET method and assign the HTML response to a BeautifulSoup object

Find all tables on the wiki page and collect all relevant columns

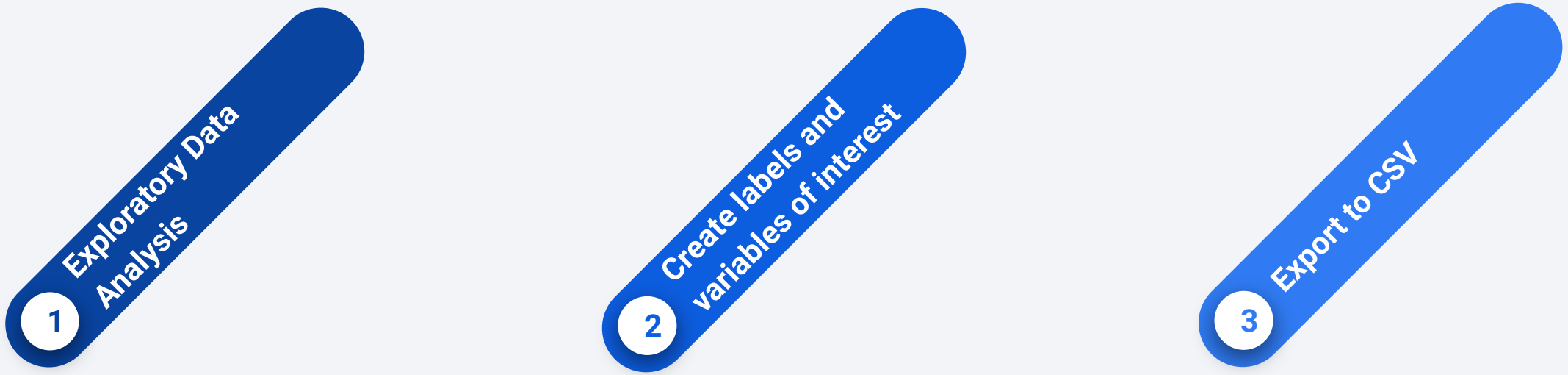
Create a dataframe by parsing the HTML tables

Export dataset to CSV

URL: https://github.com/bichow/Coursera-Data-Science-Capstone/blob/master/Data%20Collection%20Web%20Scraping.ipynb

Data Wrangling

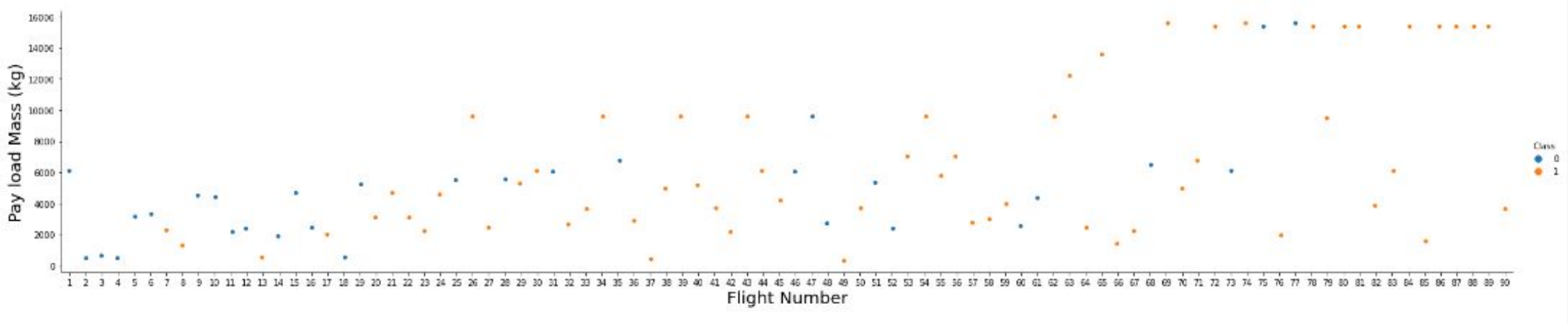
- Since SpaceY is interested in predicting successful rocket launches, the data was processed to feature a classification variable for successful and unsuccessful landing outcomes. Using this variable we can easily determine the success rate of SpaceX launches.



URL: https://github.com/bichow/Coursera-Data-Science-Capstone/blob/master/Data%20Wrangling.ipynb

EDA with Data Visualization

- As part of the exploratory data analysis, the data was visualized using scatterplots, barplots, and line plots to visually check for relationships between variables such as Payload mass and Flight Number.



As an example, we can observe that successive flights have had higher rates of launch success and feature greater payload masses.

URL: https://github.com/bichow/Coursera-Data-Science-Capstone/blob/master/EDA%20with%20Visualization.ipynb

EDA with SQL

SQL queries performed

- Names of the unique launch sites in the space mission
- 5 records where launches sites begin with ‘CCA’
- Total payload mass launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1
- Date of the first successful landing on a ground pad
- Names of boosters with successful landings on drone ship with payload mass greater than 4000kg
- Total number of successful and failed missions
- Names of booster versions that have carried the maximum payload mass
- List of failed landing outcomes on drone ship with booster versions, launch site names, and in the year 2015
- Ranked count of landing outcomes between 2010-06-04 and 2017-03-20 in descending order

URL: https://github.com/bichow/Coursera-Data-Science-Capstone/blob/master/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- In addition to charts and queries, an interactive map was constructed with Folium to visualize SpaceX rocket launches on a map. The interactive map included:
  - Markers to indicate the launch sites and Marker Clusters to show successful/failed launches for each site
  - Circles to highlight the areas surrounding each launch site
  - Lines used to locate and find the distance between launch sites and key locations

URL: https://github.com/bichow/Coursera-Data-Science-Capstone/blob/master/Interactive%20Visual%20Analytics%20With%20Folium.ipynb

Build a Dashboard with Plotly Dash

- An interactive dashboard was built using Plotly Dash to visualize details about launch sites, launch success rates at each launch site, and payload masses.
  - The percentages of total successful launches by site are visualized by pie chart
  - The percentage of successful launches for each site are also visualized by pie chart
  - Payload mass vs. success rate is visualized by scatterplot utilizing a slider for payload mass.
- With the dashboard, it is easy to discern the launch site as well as the payload mass where the highest rate of successful launches occurs.

URL: https://github.com/bichow/Coursera-Data-Science-Capstone/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

- For the purpose of predictive analysis on launch success, four classification models were created. The methods used are as follows:
  - Logistic regression
  - Support vector machine
  - Decision tree
  - K-nearest neighbors

Data preparation — Standardize data and perform train/test split on the dataset

Train and test models — Train and test each model to find best hyperparameters

Compare results — Compare training and testing accuracy across methods

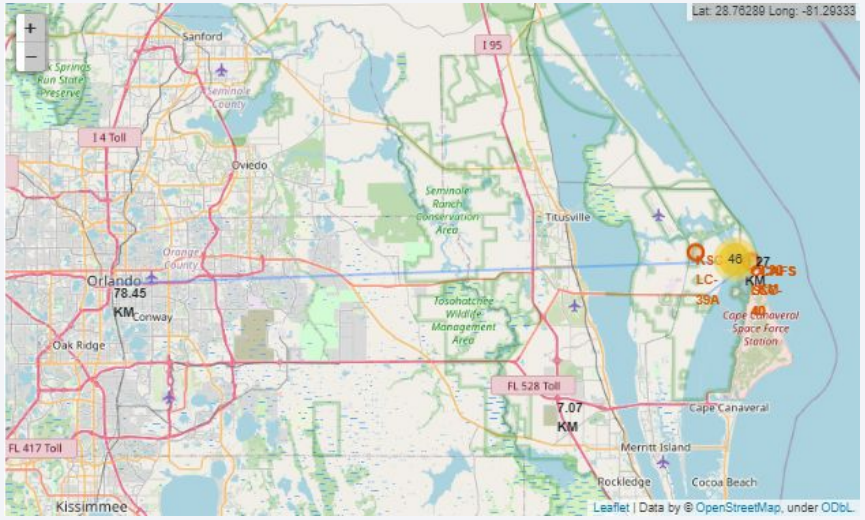
URL: https://github.com/bichow/Coursera-Data-Science-Capstone/blob/master/Machine%20Learning%20Prediction.ipynb

Results

- Exploratory data analysis results
  - Spacex uses 4 unique launch sites:
    - CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, and VAFB SLC-4E
    - KSC LC-39A sees the highest percentage of total successful launches
  - The F9 v1.1 booster carries and average payload mass of 2928 kg
  - The first successful landing on ground pad occurred on 2015-12-22
  - There have been 4 successful landings on drone ship with payload masses greater than 6000 kg
  - There have been 101 missions
    - 100 successful, 1 failure
  - Successive launches have been more successful and have carried greater payload masses

Results

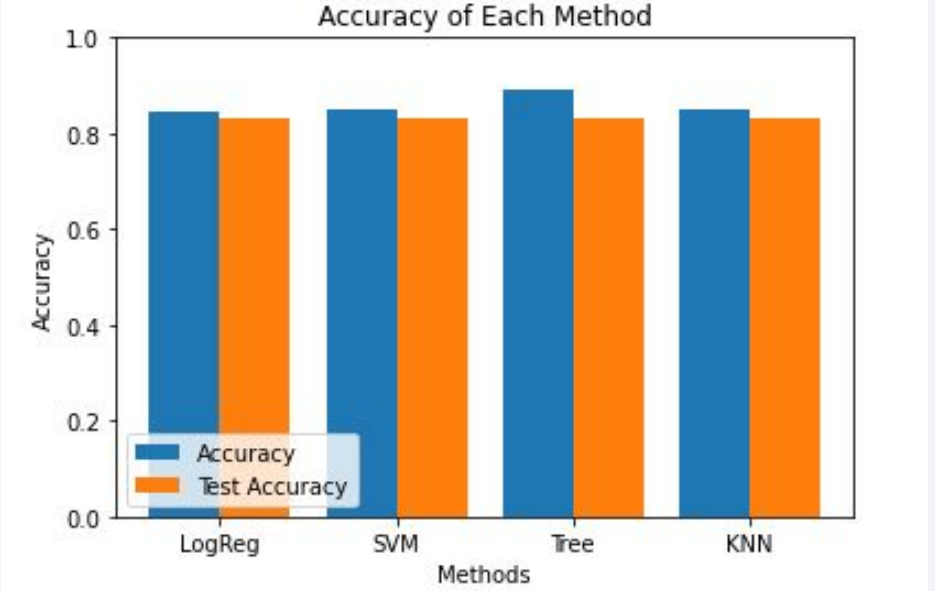
- Interactive analytics results
  - SpaceX is an American company so naturally launch sites are located on the coastal U.S.
  - Additionally, launch sites are predominantly as close to both the equator as well as infrastructure as possible
  - The majority of launches occur on the east coast



Results

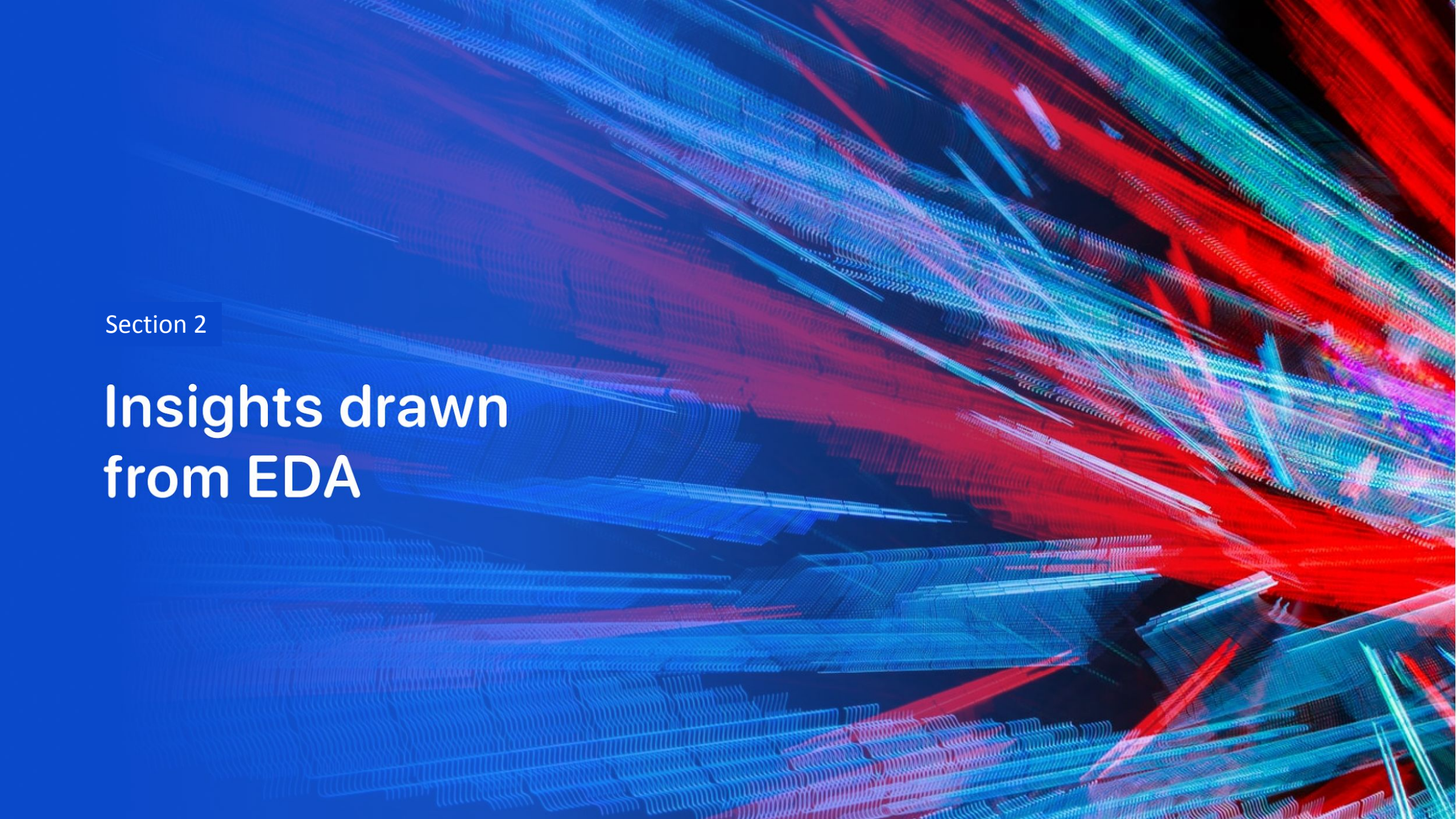
- Predictive analysis results
  - The decision tree classifier model shows the highest accuracy at 89 percent however in terms of test accuracy all classification models seem to be the same at 83 percent
  - The decision tree classifier model should be the best model of the models constructed

Model	Accuracy	TestAccuracy
LogReg	0.84643	0.83333
SVM	0.84821	0.83333
Tree	0.88929	0.83333
KNN	0.84821	0.83333



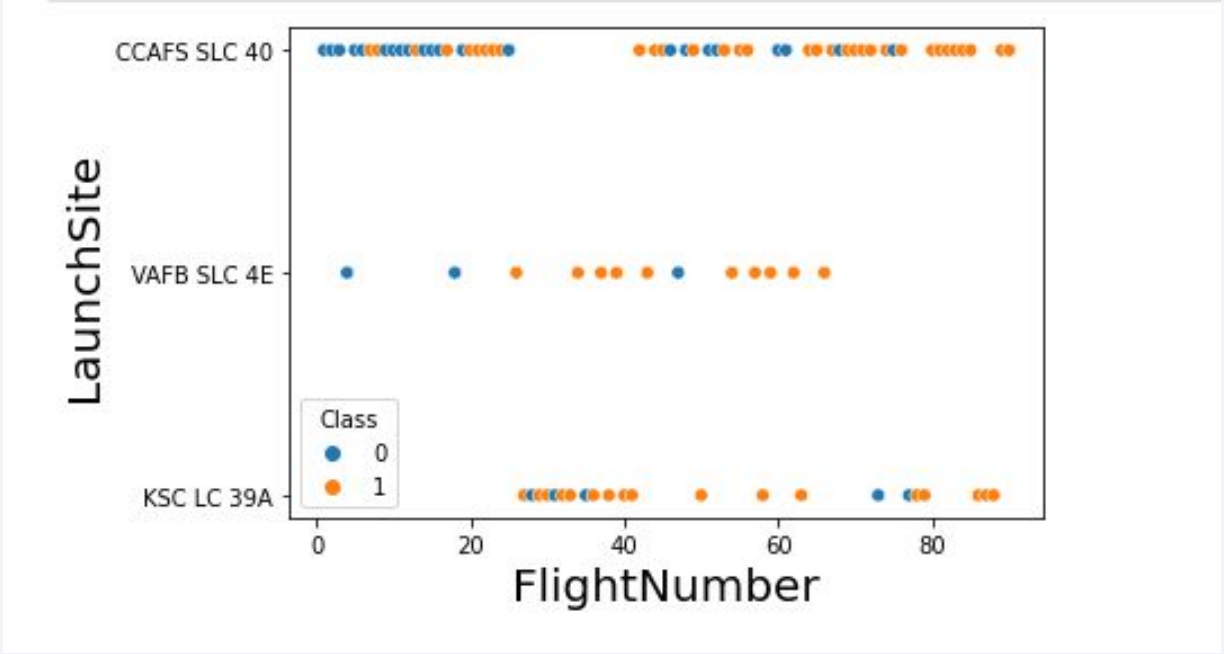
Section 2

Insights drawn from EDA



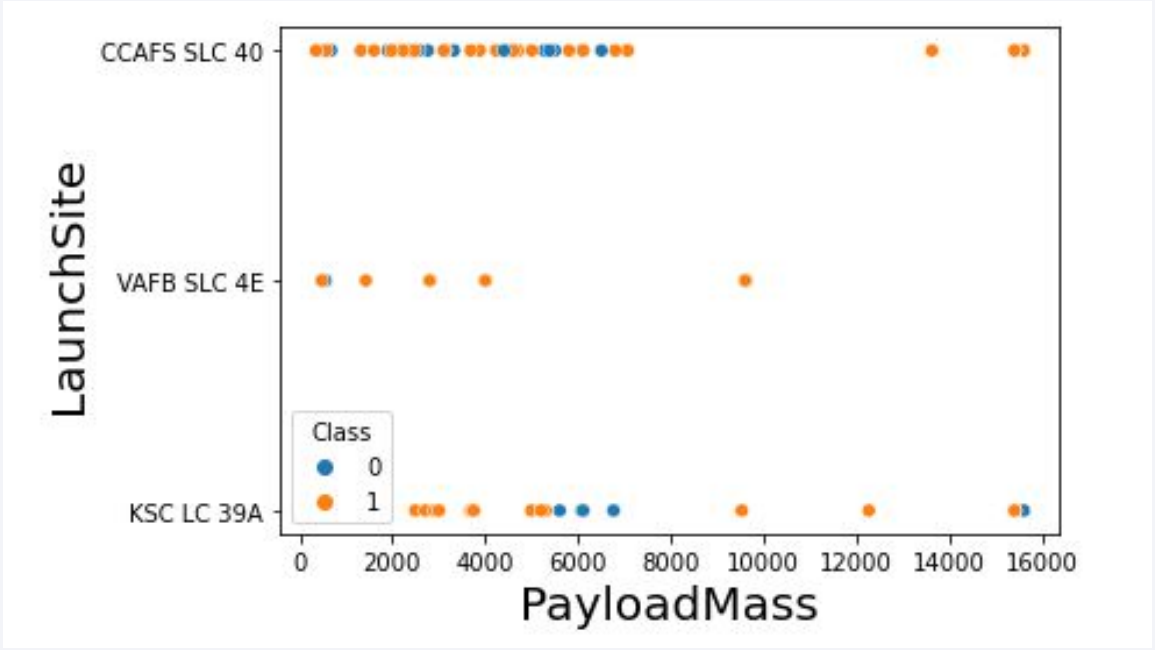
Flight Number vs. Launch Site

- From the scatterplot, recent launches have been more successful than early launches
- SpaceX launches the most rockets from launch site CCAFS SLC 40.
- SpaceX has stopped launching from VAFB SLC 4E as of recent



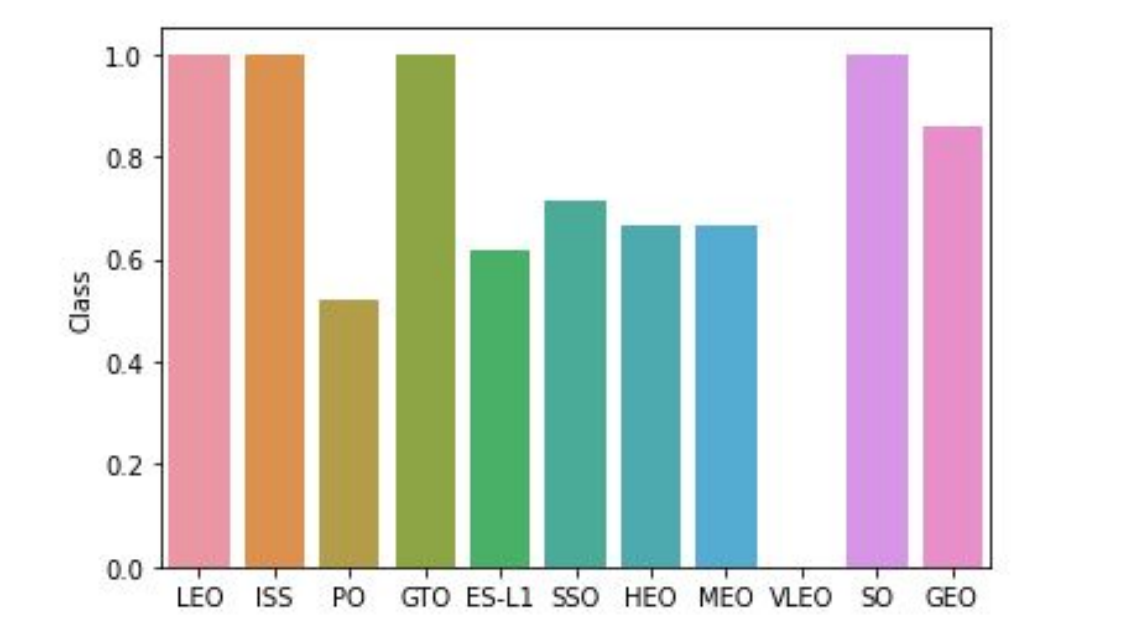
Payload vs. Launch Site

- Most launches have payloads under 8000 kg
- Launches over 8000 kg are mostly successful
- Rocket launches from launch site VAFB SLC 4E are generally successful but the launch site doesn’t appear to support payloads greater than 10000 kg



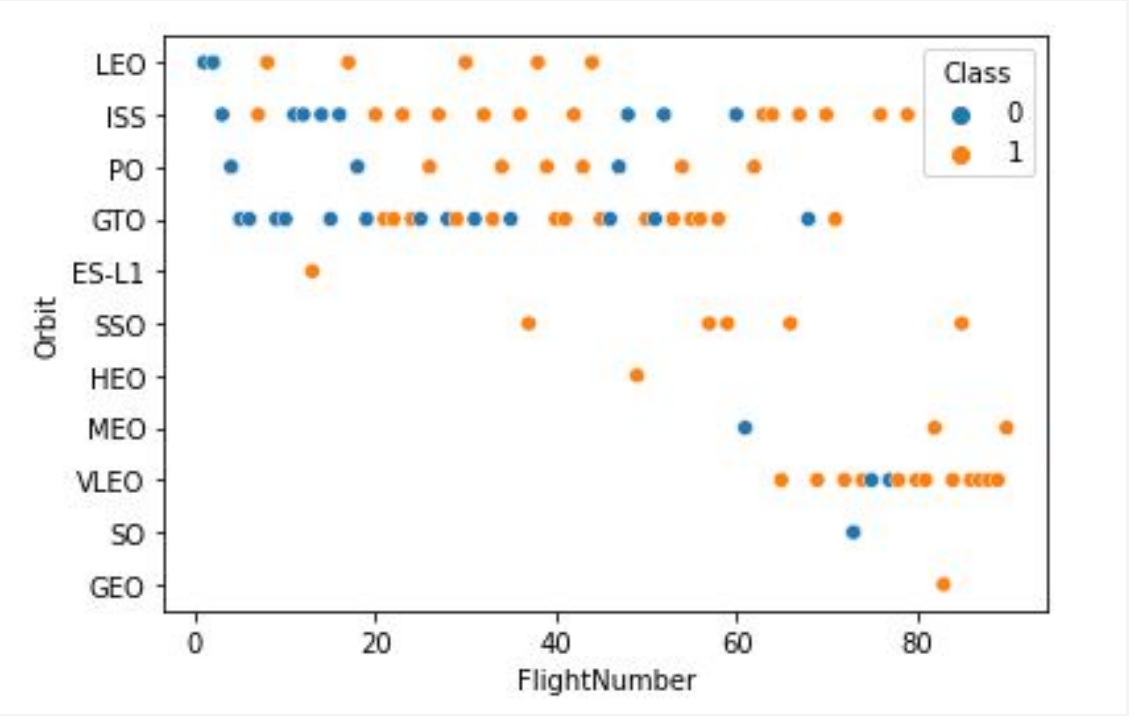
Success Rate vs. Orbit Type

- Successful orbit types:
  - LEO: low Earth orbit
  - ISS: International Space Station
  - GTO: geosynchronous orbit
  - SO: sun-synchronous orbit
- Less successful orbit types (desc):
  - GEO: circular geosynchronous orbit
  - SSO
  - HEO: high eccentricity elliptic orbit
  - MEO
  - ES-L1
  - PO
  - VLEO: very low Earth orbits



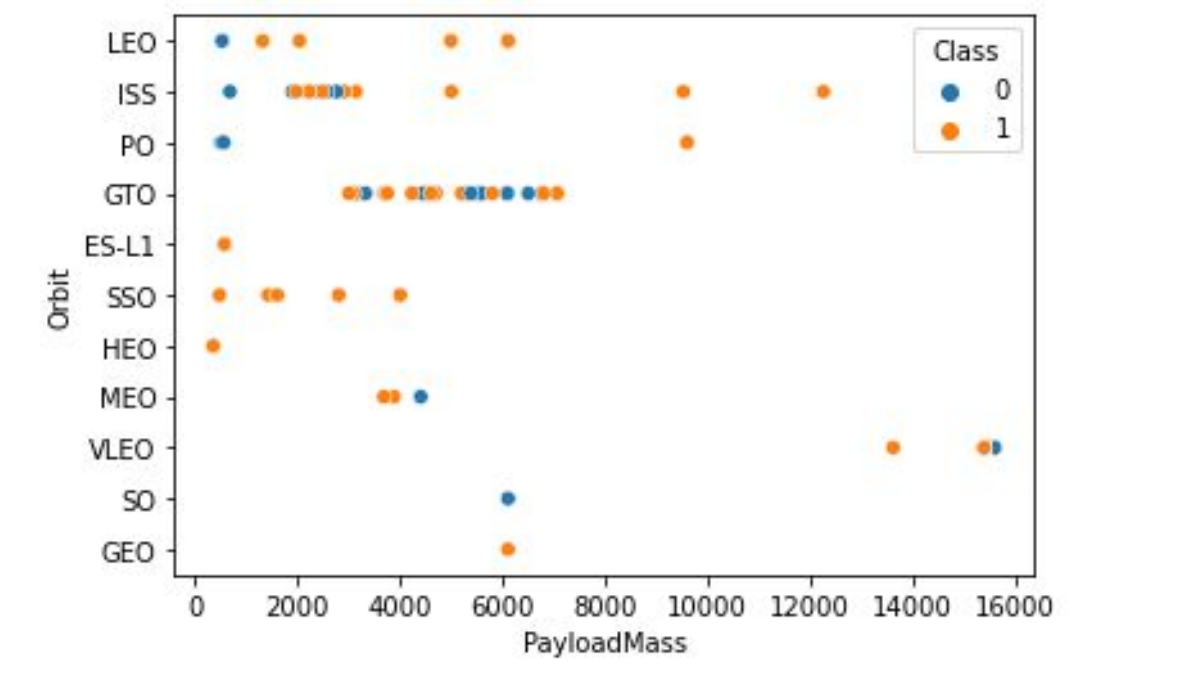
Flight Number vs. Orbit Type

- Recent launches have mostly been launched into VLEO and have been mostly successful
- Launches into LEO seem to be discontinued



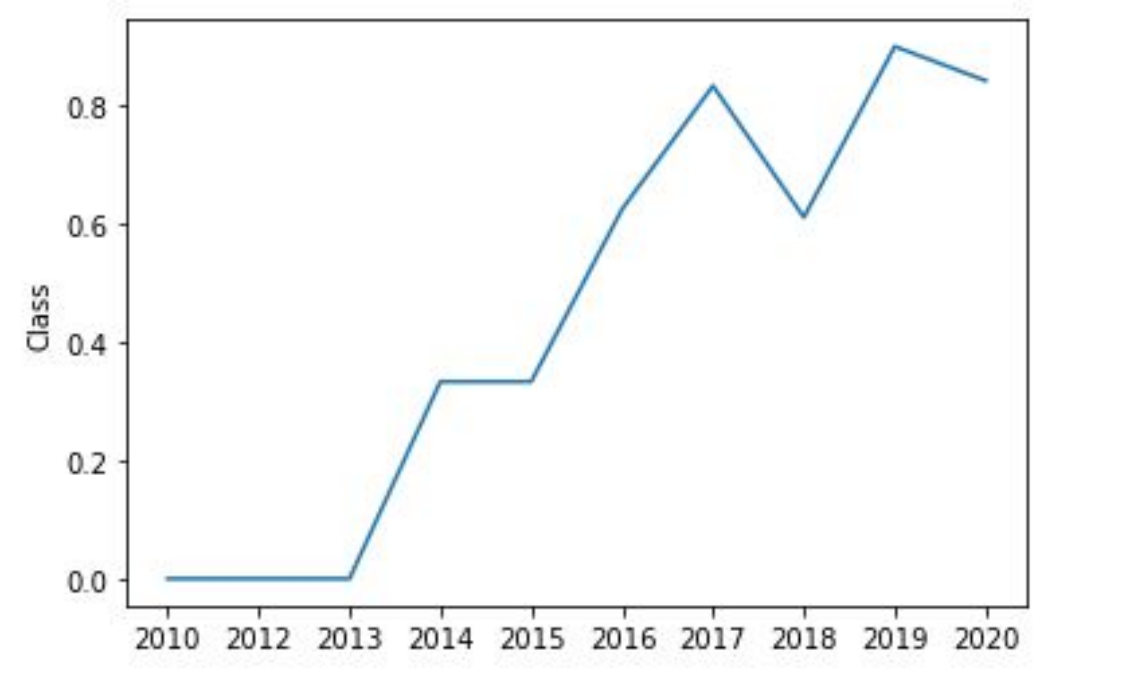
Payload vs. Orbit Type

- There are fewer launches into orbit with payload mass over 8000 kg
- Heavier payloads (over 12000 kg) are launched into very low Earth orbit
- Payloads under 8000 kg are mostly launched into GTO



Launch Success Yearly Trend

- SpaceX started seeing successful landing outcomes with rockets in 2013
- There is a rapid increase in success rate 2013 to 2014
- Generally, SpaceX launches have been seeing increasing success over the years



All Launch Site Names

- There are four launch sites:
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E
- CCAFS LC-40 and CCAFS SLC-40 are next to each other
- VAFB SLC-4E is the only launch site on the West coast

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	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)	NARA (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

Launch Site Names Begin with 'CCA'

- The previous table shows 5 records where launch sites begin with `CCA`, denoting a launch site at Cape Canaveral
- The records show that early launches at CCAFS failed to land or did not attempt to land

Total Payload Mass

- The total payload carried by boosters from NASA is 45596 kg
- The total payload is the sum of all payloads from the launches whose customer correspond to ‘CRS’ or NASA

Title
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928 kg
- The average payload mass is calculated by filtering for the booster version and then finding the average of payload mass

Title
2928

First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad was December 22, 2015
- The date of the first successful landing outcome on ground pad is found by filtering for successful landing outcomes on ground pad and finding the minimum value for the date

Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

Booster Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes
  - 1 Failure (in flight)
  - 99 Successes
  - 1 Success (payload status unclear)

Mission Outcomes	Count
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass:
  - 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

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

- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015 are as follows:

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

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

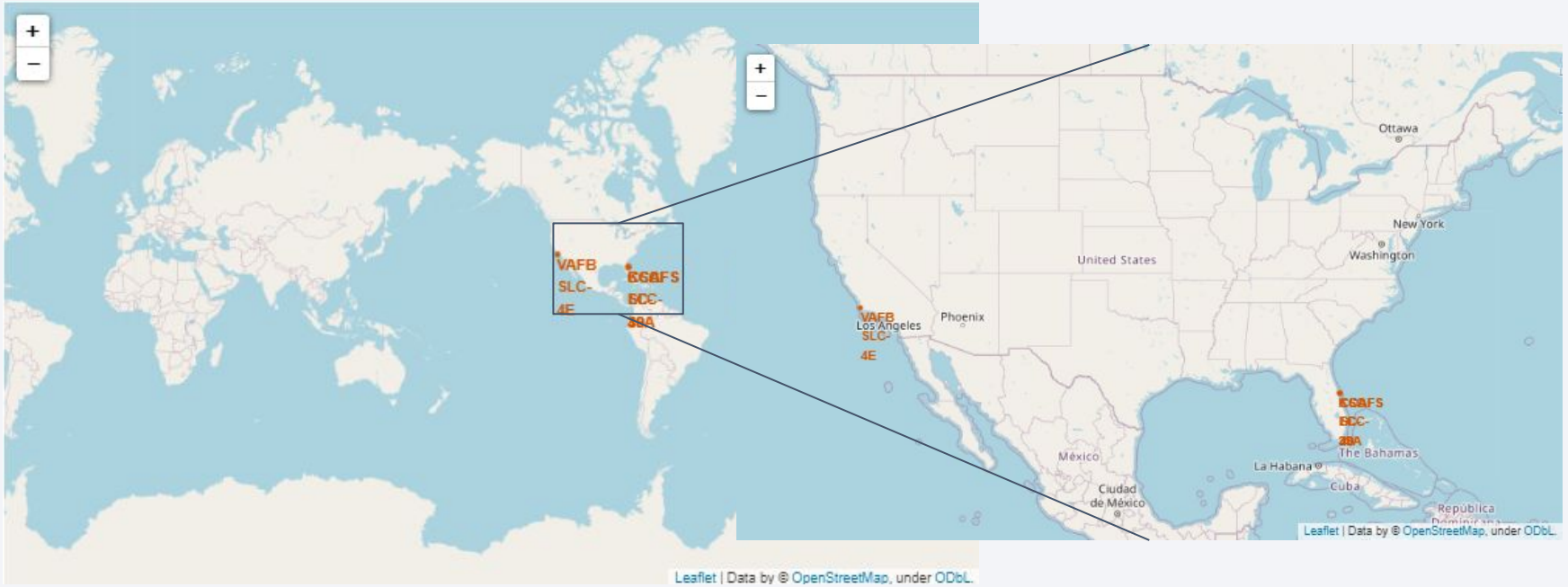
- The count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order
- The most frequent landing outcome is “No attempt”
- The least frequent landing outcome is “Precluded (drone ship)

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


Section 3
Launch Sites
Proximities Analysis

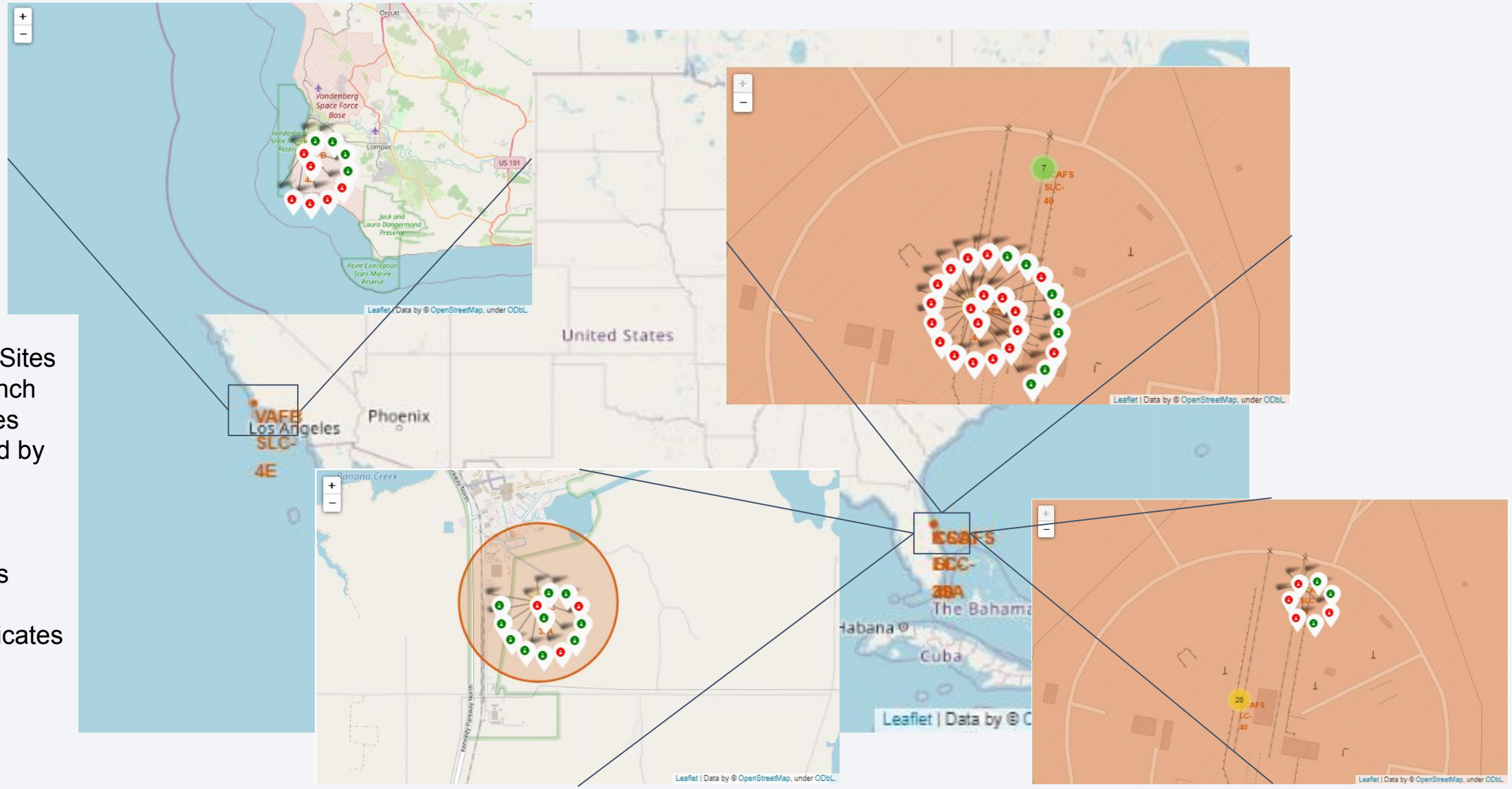
Launch Site on a Global Map

- All launch sites are on the southern coastal U.S. with the majority being on the East coast



Launch Sites and Launch Outcome Clusters

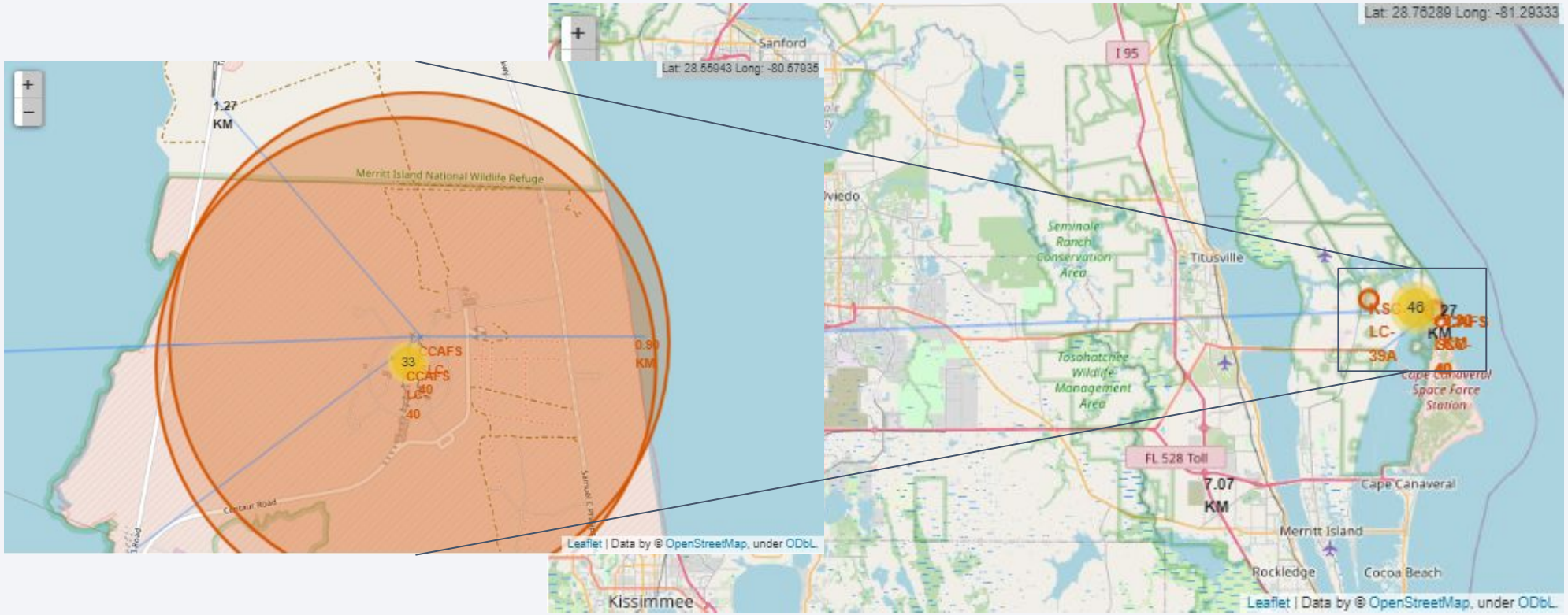
- Launch Sites with launch outcomes indicated by marker clusters
- Green indicates success
- Red indicates failure

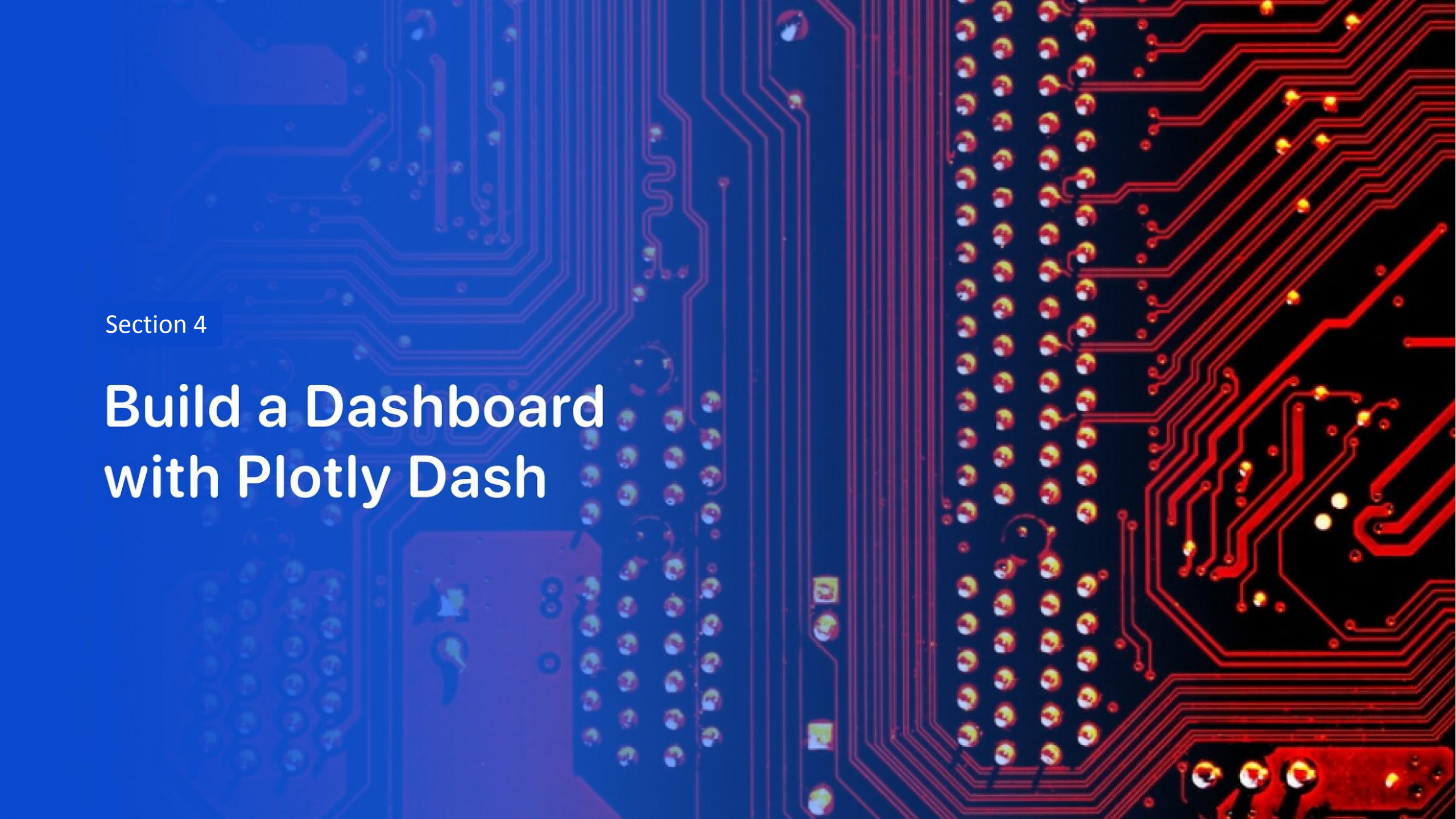


Launch Sites and Key Locations

- Launch sites with lines drawn to nearby key locations

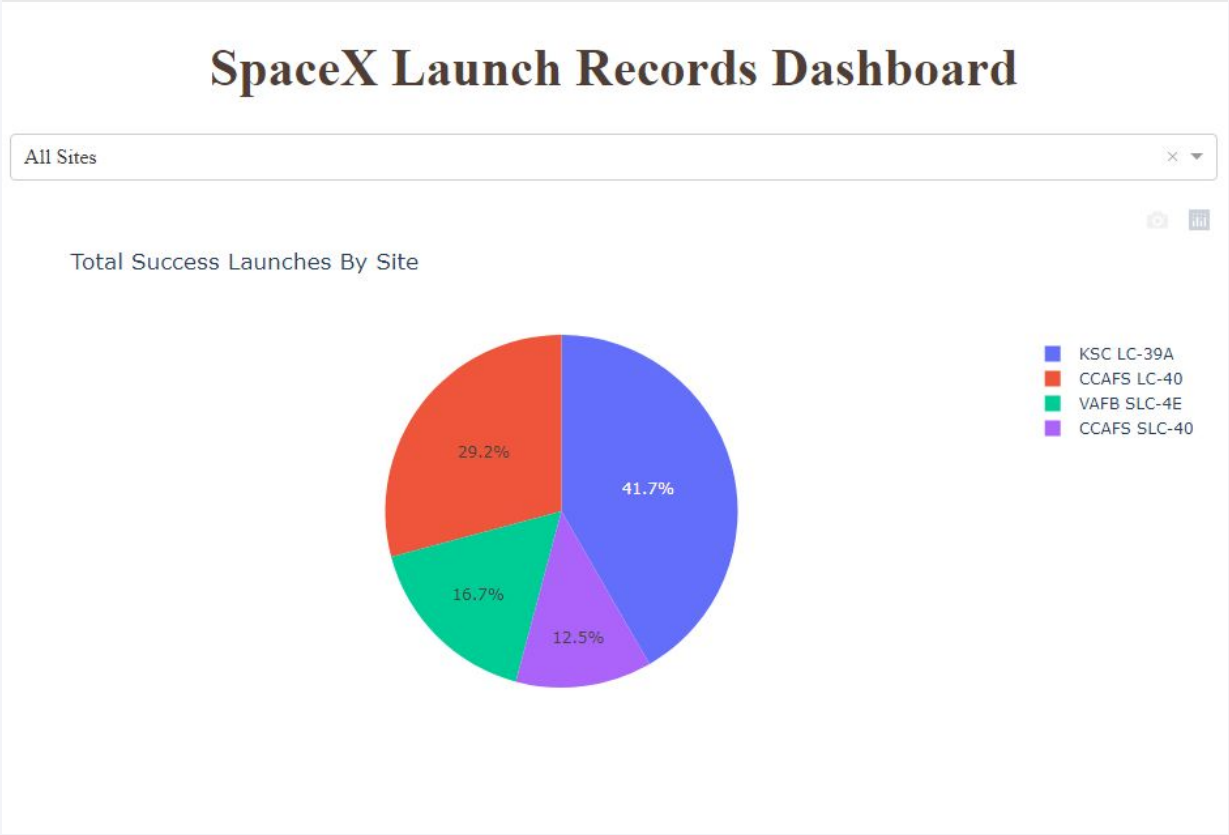
Distance to nearest:
Coastline: 0.90km
Railway: 1.27km
Highway: 7.07km
City: 78.45km




Section 4
Build a Dashboard with Plotly Dash

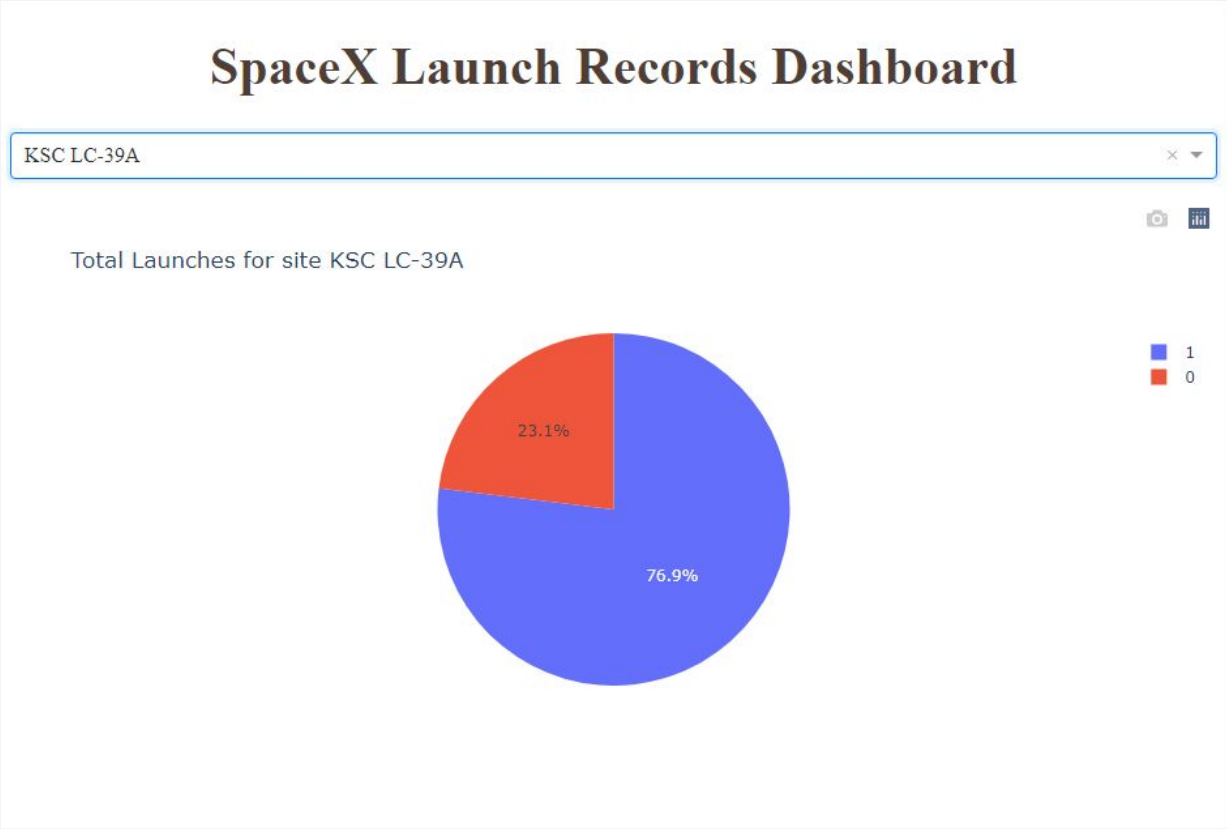
Total Successful Launches by Site

- The greatest percentage of successful launches were launched from KSC LC-39A



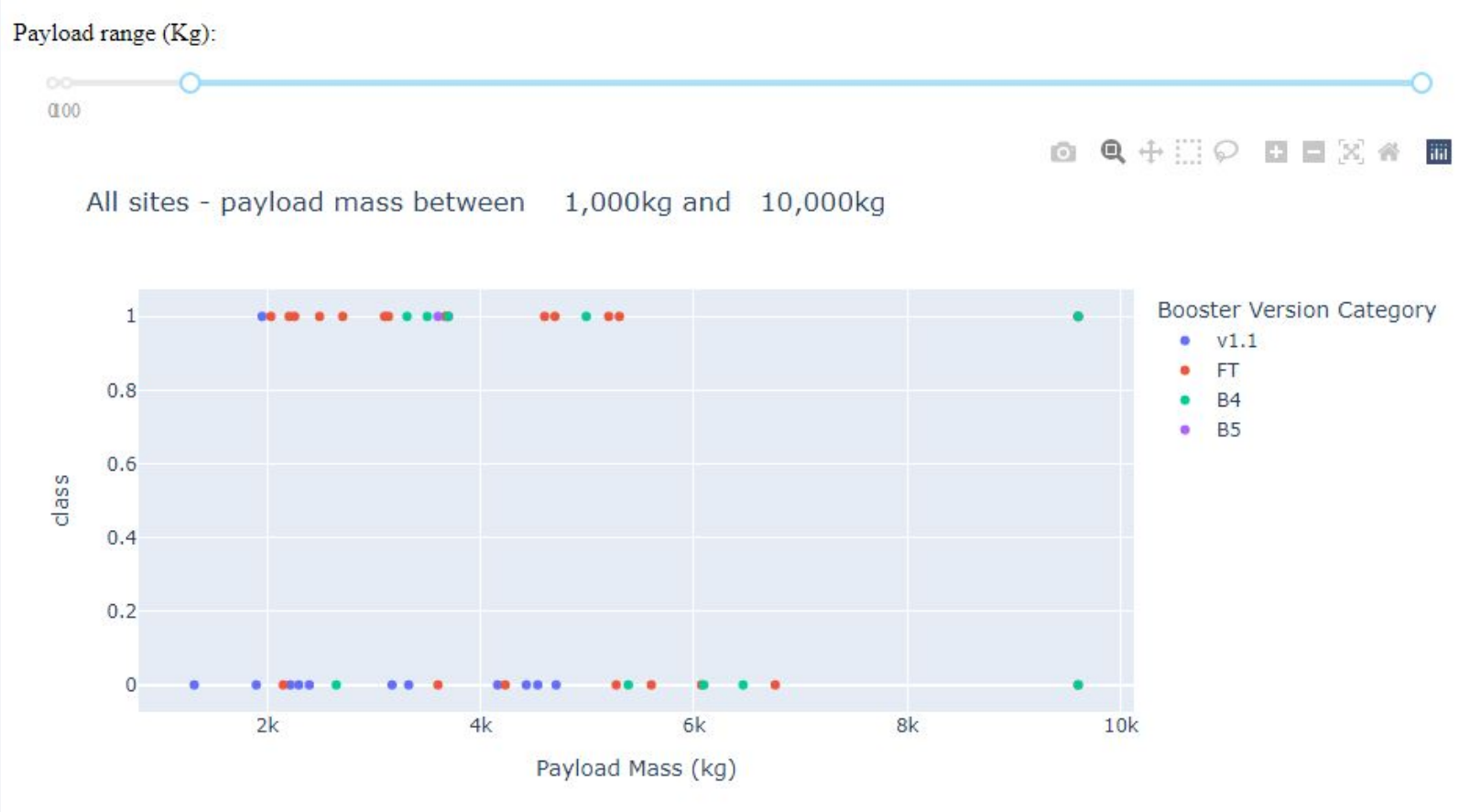
Launch Success Rate at KSC LC-39A

- Launches at launch site KSC LC-39A are 76.9% successful



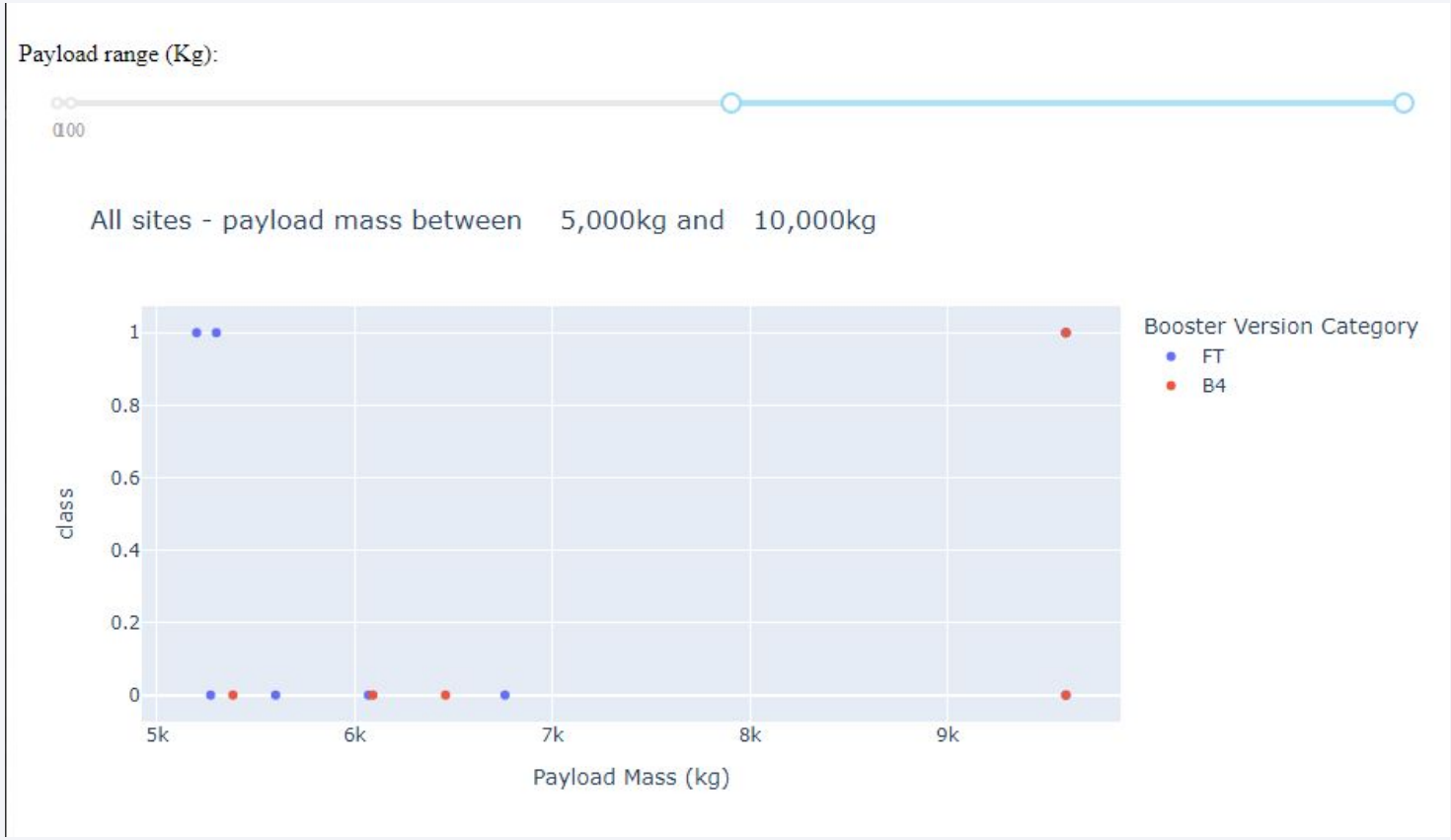
Full Range Payload Mass vs. Success Rate

- This plot shows launch success rate across payload masses greater than 1000 kg



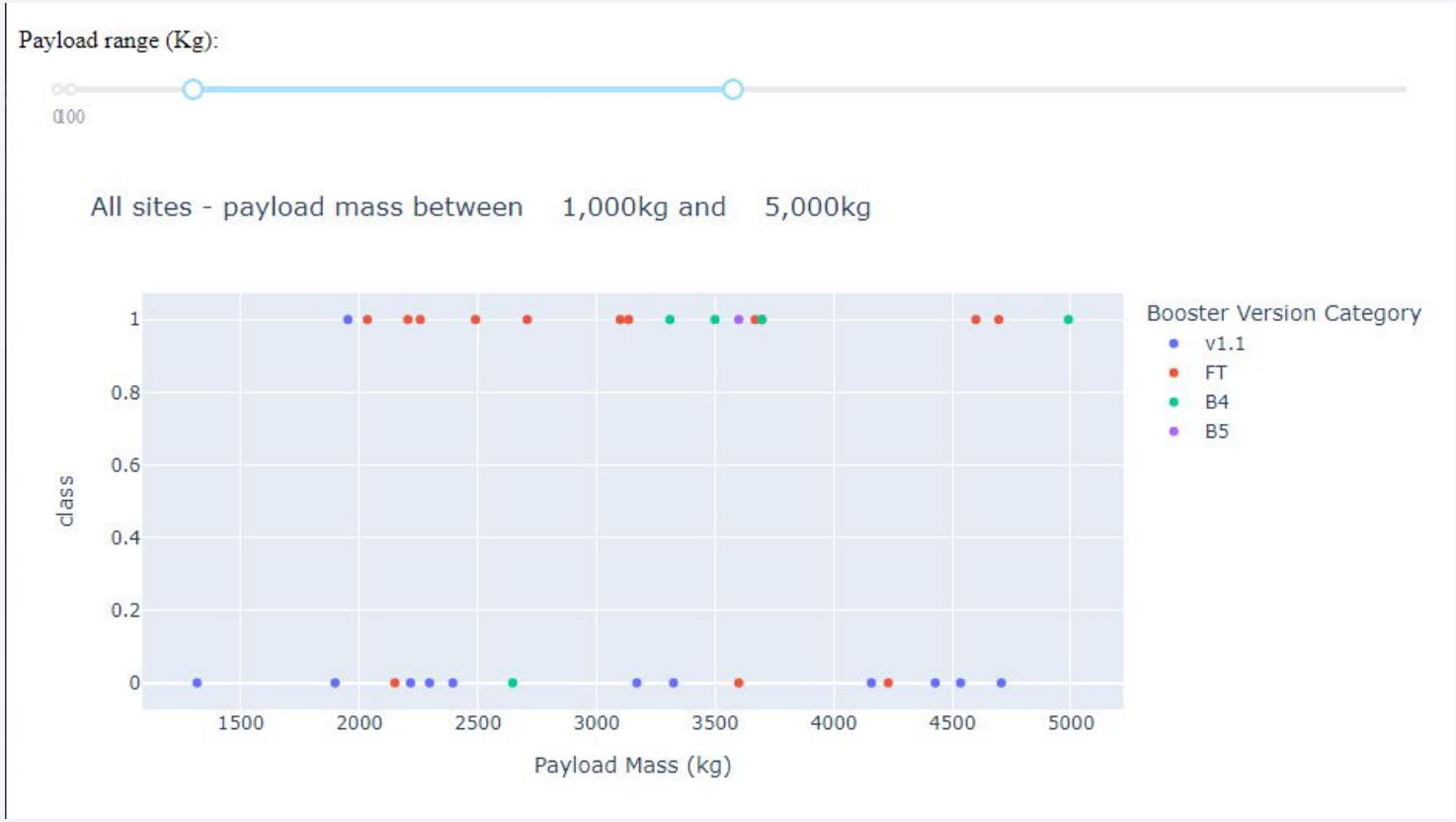
Payload Mass (>5000kg) vs. Success Rate

- The majority of rockets with payload mass greater than 5000 kg are unsuccessful



Payload Mass (1000-5000kg) vs. Success Rate

- In contrast, rockets with payload mass between 1000 and 5000 kg see a greater rate of success



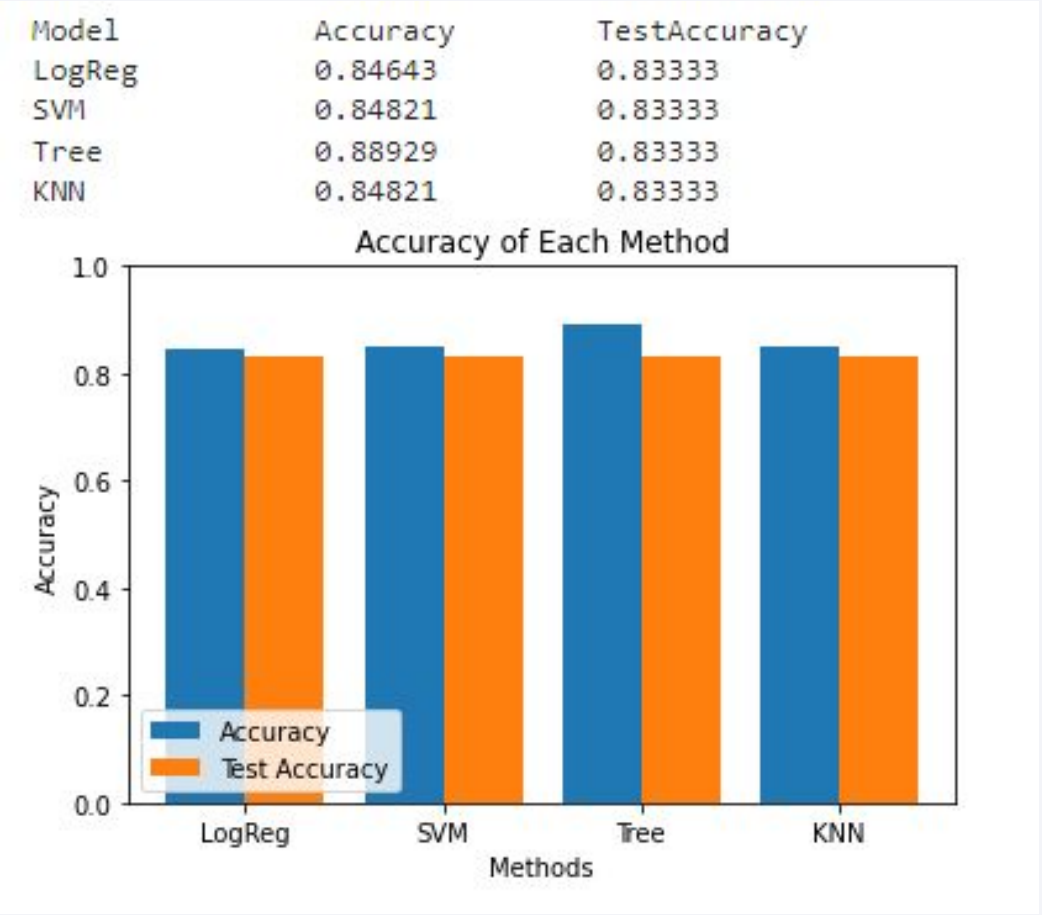
Section 5

Predictive Analysis (Classification)

Classification Accuracy

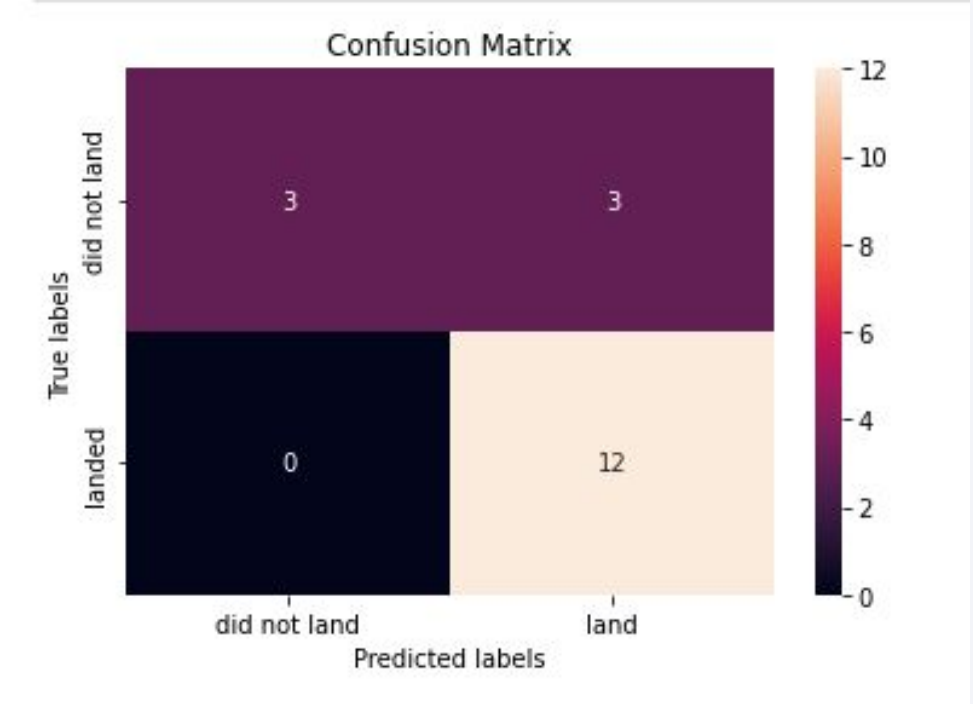
- The bar chart visualizes the model accuracies for each of the classification models
- All models have the same test accuracy
- The decision tree classification model may be the best due to having the highest training accuracy

Model	Accuracy	TestAccuracy
LogReg	0.84643	0.83333
SVM	0.84821	0.83333
Tree	0.88929	0.83333
KNN	0.84821	0.83333



Confusion Matrix

- The confusion matrix of the decision tree classification model shows that the model is relatively accurate
- It should be noted that the model error has a tendency towards a false positive (predicting a landing success when the rocket did not land)



Conclusions

- SpaceX launches most frequently from Cape Canaveral but launches from KSC LC-39A account for the greatest percentage of successful launches
- Based on most recent launches, launches into VLEO (very low Earth orbit) are more successful
- Rocket success rate was very poor at the start but successive flights have increased launch success rate considerably
- Rocket launches with payload masses between 1000 and 5000 kg are more successful than rockets with heavier (5000kg+) payload masses
- The decision tree classifier model can be used to predict landing outcomes although it may have a tendency towards false positives

Thank you!