Winning Space Race with Data Science

Shahamat Mehdi
15 September, 2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection through SpaceX API and web scraping of wikipedia page of SpaceX
  - Data wrangling
  - Exploratory data analysis
    - Analyzing the outcome by orbit type, payload mass and booster version
    - Visual analysis through charts and folium maps
  - Building dashboard using plotly dash
  - Machine learning analysis
- Summary of all results
  - Exploratory data analysis results
  - Interactive dashboard analysis results
  - Predictive analysis through machine learning results

Introduction

- Project background and context
  - SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars.
  - Other providers costs upward of 165 million dollars each.
  - The costs are saved due to the reuse of first stage of Falcon 9 rockets.
- Problems you want to find answers
  - We want to determine if the first stage will land successfully.
  - What parameters influence successful landing
  - Use these analysis to bid against SpaceX

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API
  - Web scrapping of SpaceX Wikipedia page
- Perform data wrangling
  - Replacing missing values by mean values for Payload Mass
  - One hot encoding for applying machine learning algorithms
  - Visualizing through Pandas data frames
- Perform exploratory data analysis (EDA) using visualization and SQL

Methodology

Executive Summary

- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Algorithms used
    - Logistic Regression
    - SVM
    - Decision Tree
    - KNN
  - Parameters tuned with grid Search
  - Best model chosen after testing on selected algorithms

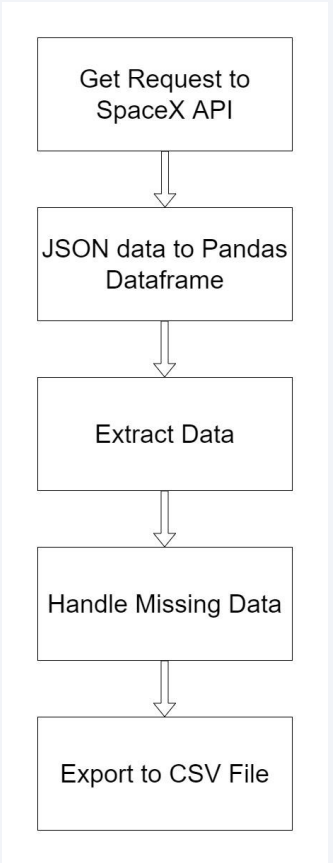
Data Collection

- SpaceX API is used to collect launch data
  - Core
  - Booster version
  - Launch site
  - Payload
- SpaceX Wikipedia page is used to collect Falcon 9 data
  - Used beautiful soup
  - Extract tables

Data Collection – SpaceX API

- Get request to SpaceX API interface
- Convert JSON data to pandas data frame
- Extract data
  - Core
  - Launch Site
  - Payload
  - Booster Version
- Filter data for Falcon 9 information
- GitHub URL of the SpaceX API calls notebook
  - https://github.com/userd4014/Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

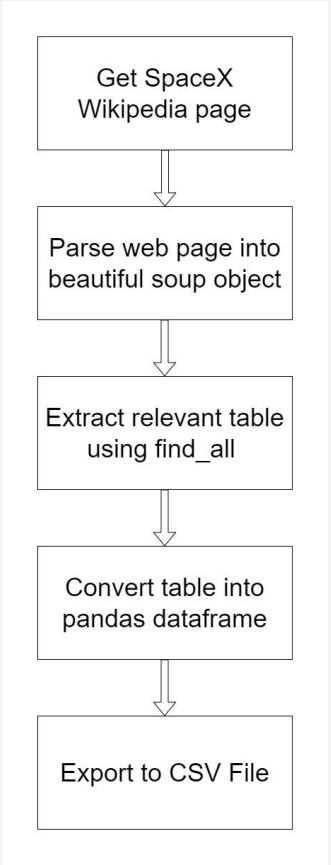
Flowchart



Data Collection - Scraping

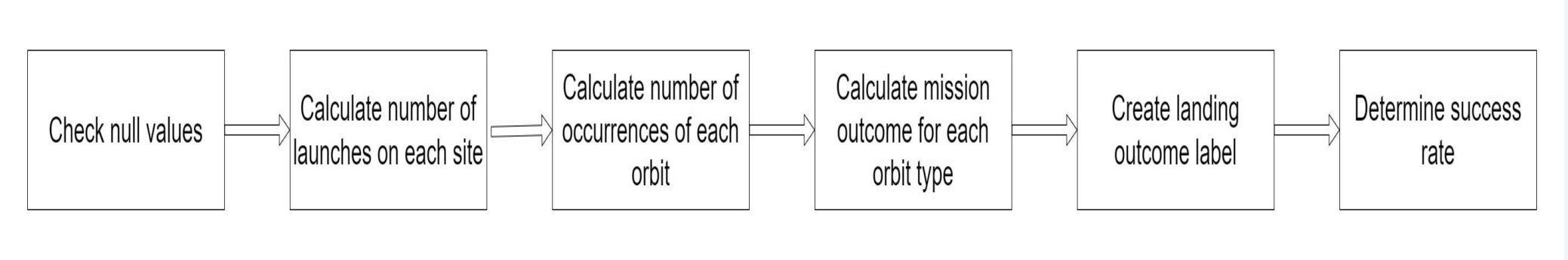
- Request to SpaceX Wikipedia page
- Parse response into beautiful soup object
- Extract tables using beautiful soup functions
- Store extracted table in pandas dataframe
- GitHub URL of the web scraping notebook
  - https://github.com/userd4014/Capstone/blob/main/jupyter-labs-webscraping.ipynb

Flowchart



Data Wrangling

- Dealing with missing values for payload using mean



- GitHub URL of data wrangling notebook
  - https://github.com/userd4014/Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Charts
  - Payload mass vs. Flight number vs. Success rate
  - Launch site vs. Flight number vs. Success rate
  - Launch site vs. Payload mass vs. Success rate
  - Orbit type vs. Success rate
  - Orbit type vs. Flight number vs. Success rate
  - Orbit type vs. Payload mass vs. Success rate
  - Success rate vs. Year
- GitHub URL of EDA with data visualization notebook
  - https://github.com/userd4014/Capstone/blob/main/EDA%20with%20Visualization%20lab.ipynb

EDA with SQL

- SQL queries
  - Unique launch sites
  - Display 5 launch sites that begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first successful landing outcome in ground pad was achieved
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List the total number of successful and failure mission outcomes
  - List the names of the booster versions which have carried the maximum payload mass
  - List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
  - Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010 06 04 and 2017 0 3 20, in descending order
- GitHub URL of completed EDA with SQL notebook
  - https://github.com/userd4014/Capstone/blob/main/EDA%20with%20SQL%20lab.ipynb

Build an Interactive Map with Folium

- Map Objects
  - Edged Circles (radius 1000m): Space launch sites
  - Markers: for labeling all objects
  - Marker Cluster: for creating a bunch of markers around space launch sites to indicate success (green) or failure (red) of the landing of the rocket's first stage
  - Lines: Measure the distance between the launch site and the next coast or next city
- GitHub URL of completed interactive map with Folium map
  - https://github.com/userd4014/Capstone/blob/main/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

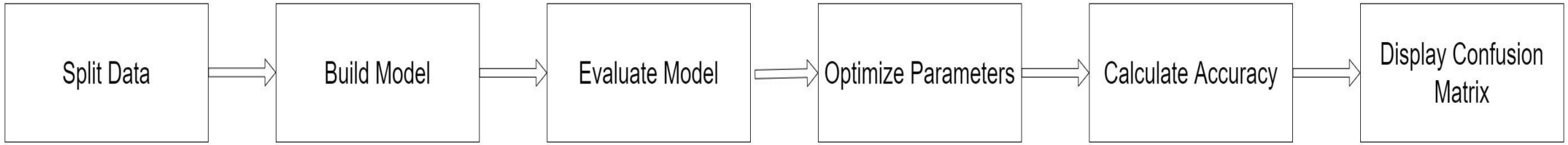
Build a Dashboard with Plotly Dash

- Pie Chart
  - Shows number of launches from each launch site as well as number of successful and failed launches from those sites
- Scatter Graph
  - Shows the relationship between the success of a launch (Outcome) and Payload (in kg) for different versions of boosters
- Input Elements
  - Dropdown list for the launch site (with option to select all)
  - Range Slider for selecting the payload mass
- GitHub URL of completed Plotly Dash lab
  - https://github.com/userd4014/Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Preprocessing
  - One-Hot-Encoding for Categorical Features
  - Split data into dependent/independent variables and train/test data
  - Scale Data with Standard Scaler
- Model Building for each Method
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K-Nearest Neighbor
- Optimization
  - Use Grid search for optimizing the models based on their hyperparameters
- Evaluation
  - Use Accuracy of Grid search for selecting the best parameter
  - Use Score to compare each classification method

Predictive Analysis (Classification)



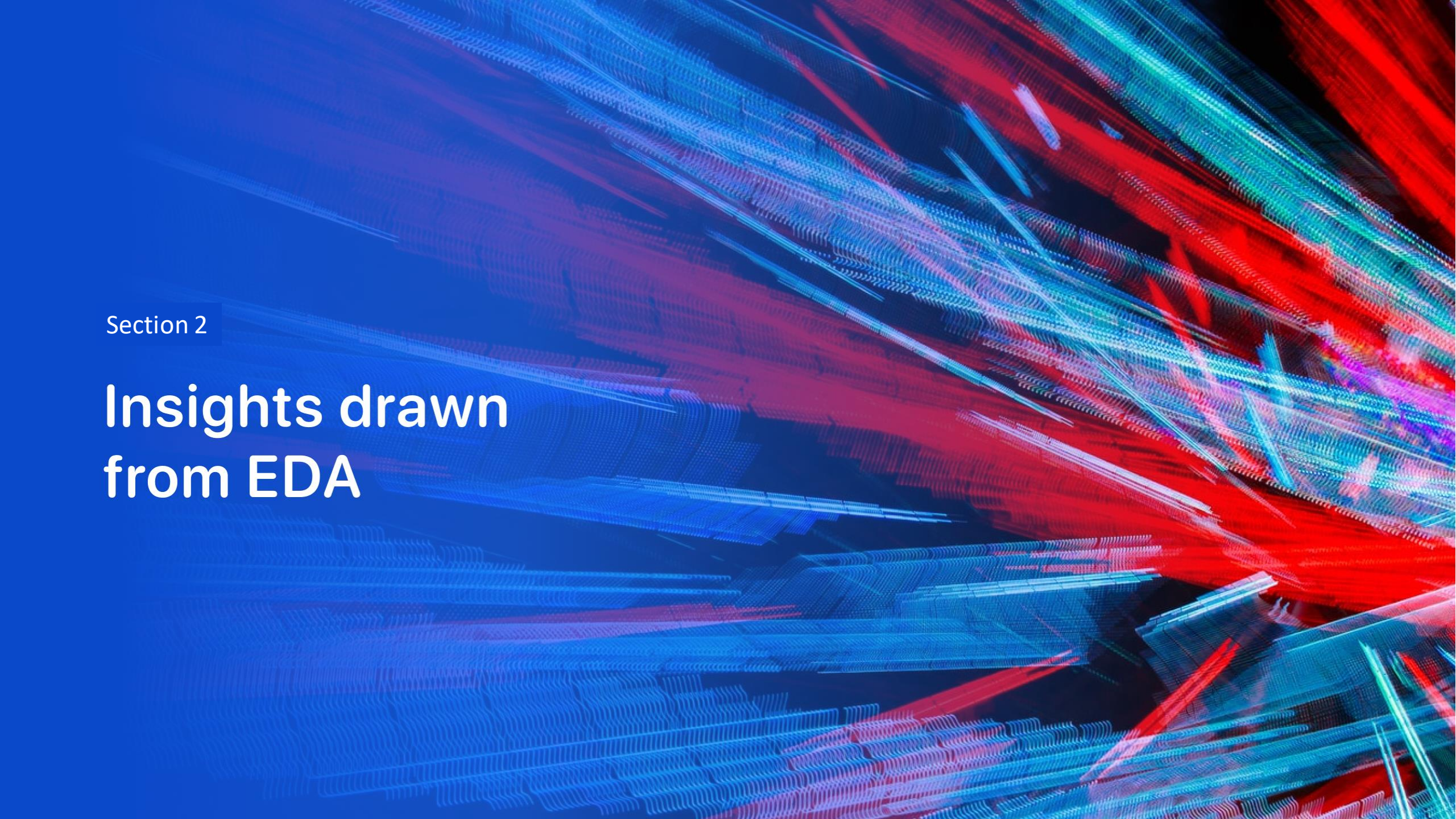
- GitHub URL of completed Predictive Analysis lab
  - https://github.com/userd4014/Capstone/blob/main/Machine%20Learning%20Prediction%20lab.ipynb

Results

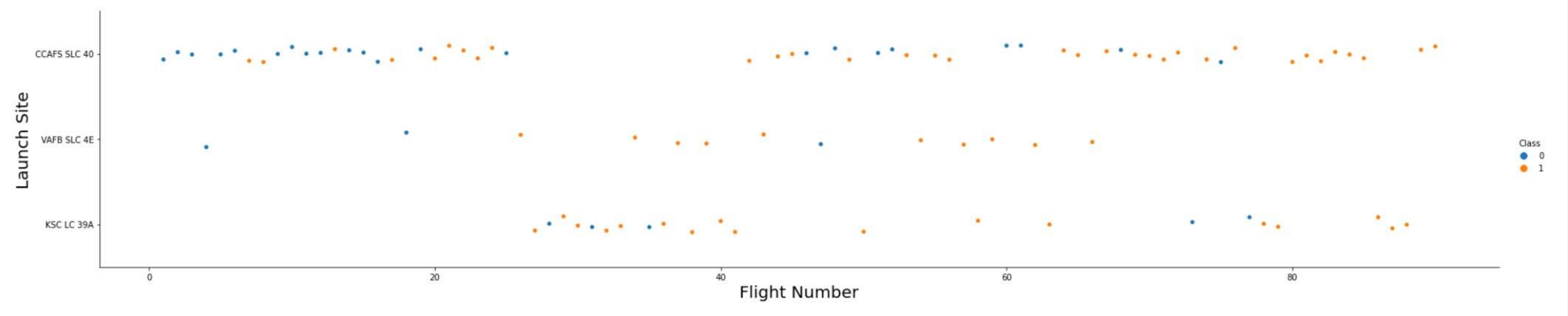
- Exploratory data analysis results
  - Launch success rate increases over time
  - Higher success rate for higher orbits
- Interactive analytics demo in screenshots
  - Higher success rate for higher payload mass
  - Low success rate for booster versions v1.0, v1.1
  - Higher success rate for booster versions FT, B4, B5
  - Higher success rate for Kennedy Space Center
- Predictive analysis results
  - Best prediction results with Logistic Regression and Support Vector Machine

Section 2

Insights drawn from EDA

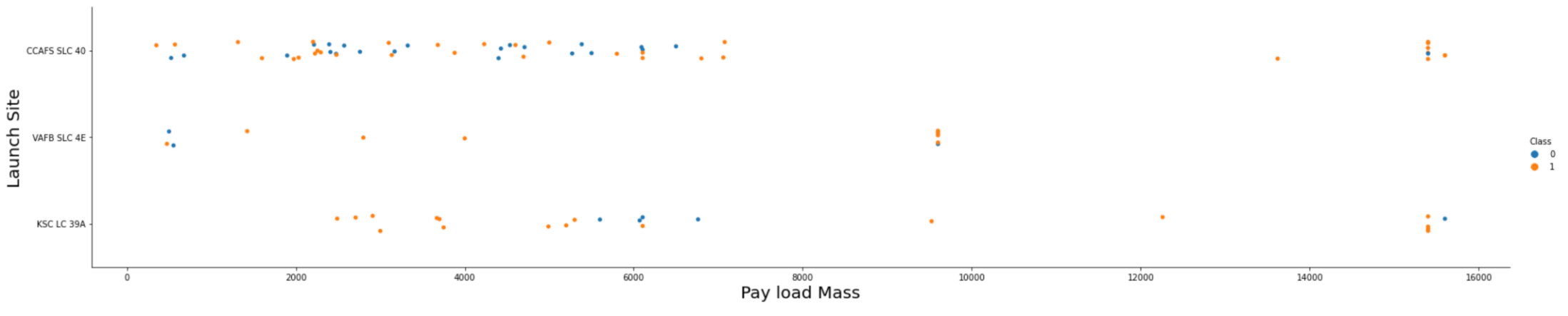


Flight Number vs. Launch Site



- The launches from CCAFS SLC 40 are more than launches from other sites

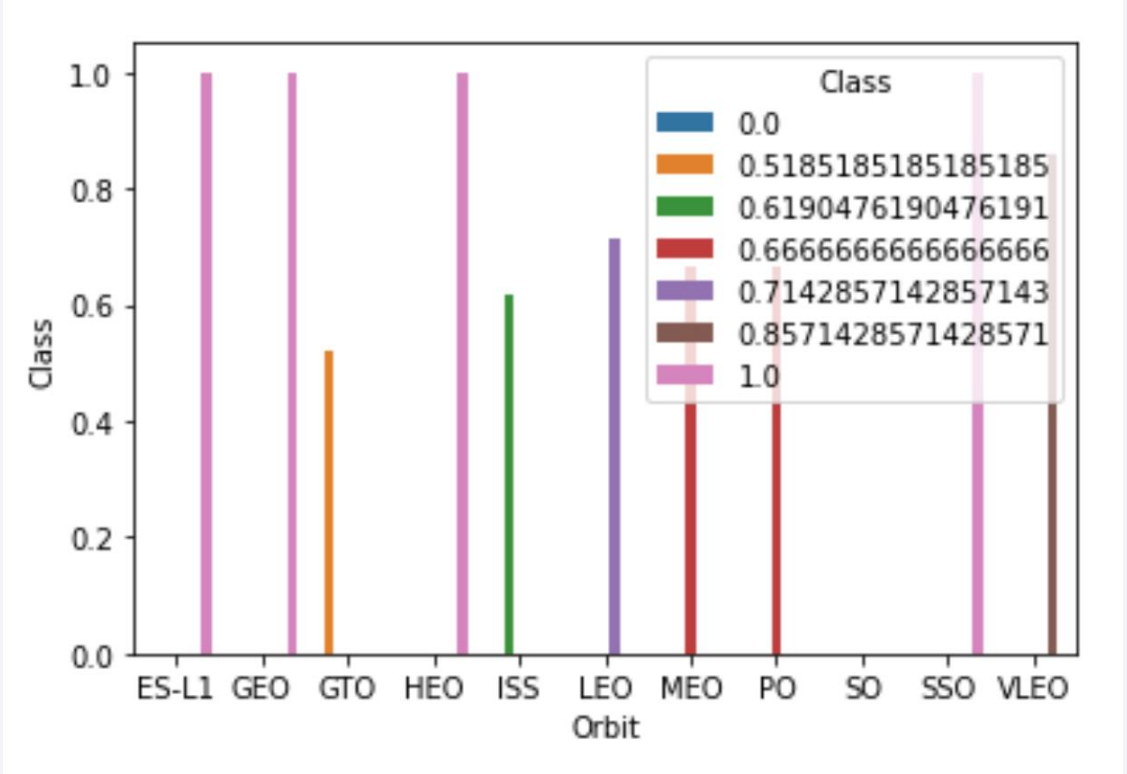
Payload vs. Launch Site



- Most payloads with lower mass are launched from CCAFS SLC 40
- There are no launches from VAFB SLC 4E with mass greater than 10000

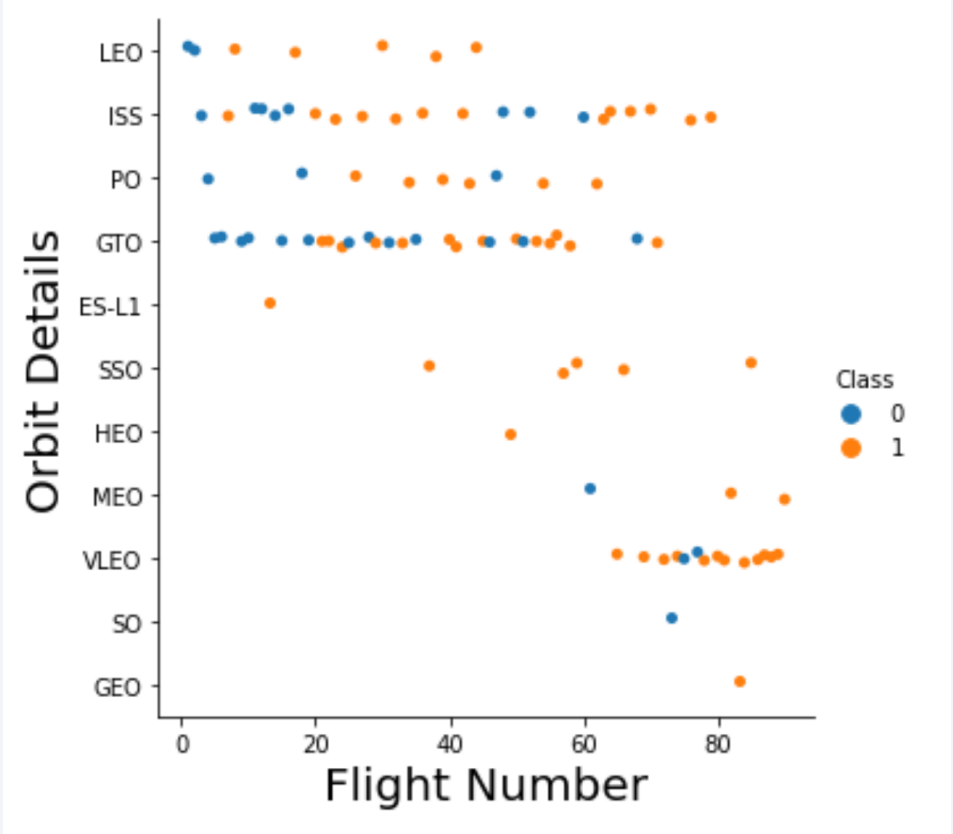
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO have 100% Success Rate



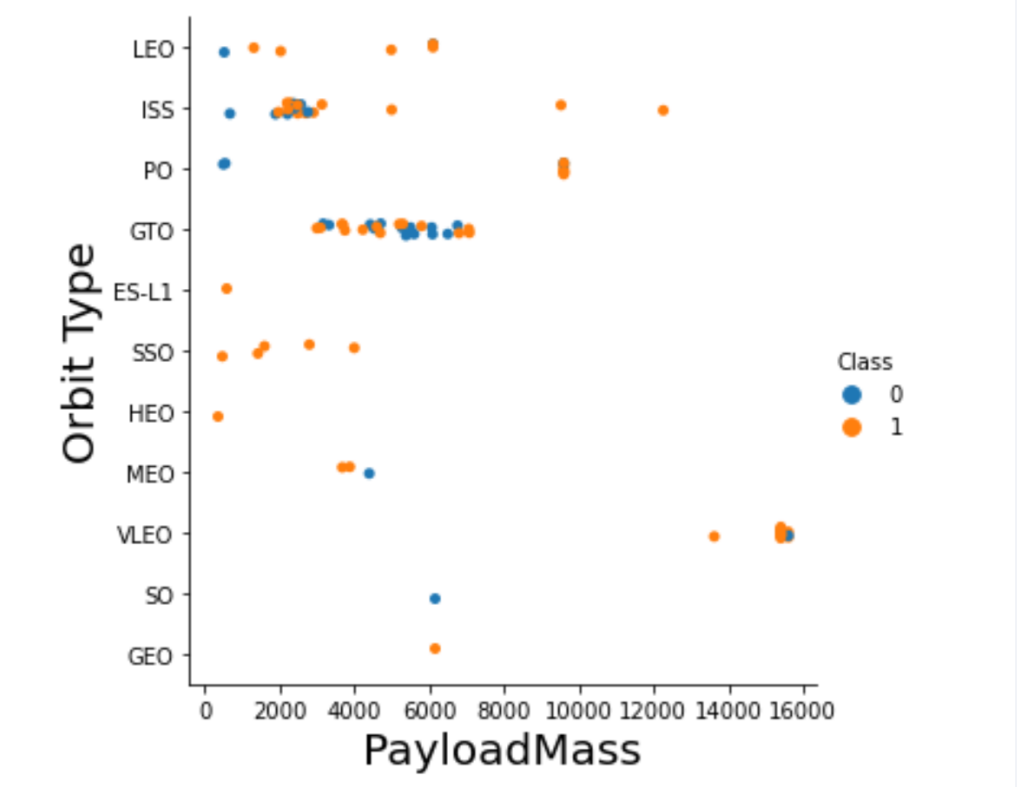
Flight Number vs. Orbit Type

- The predominant orbit types have changed over time
- Success rate has increased over time for all orbit types



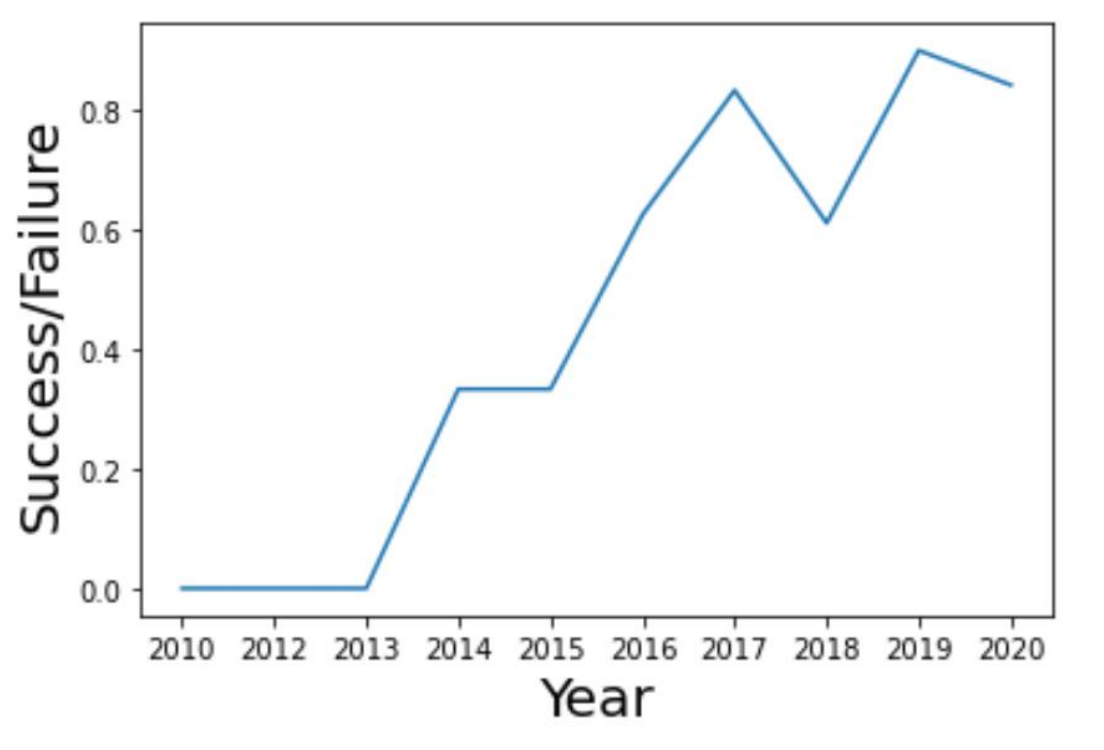
Payload vs. Orbit Type

- LEO, ISS, PO, SSO have direct correlation to Success by Payload Mass
- SSO has high Success in lower Payload Mass category
- LEO, ISS, PO have higher Success in higher Payload Mass category



Launch Success Yearly Trend

- Launch success has increased over the years



All Launch Site Names

- CCA: Cape Canaveral Launch Center
- KSC: Kennedy Space Center
- VAFB: Vandenburg Air Force Base

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

Launch Site Names Begin with 'CCA'

```
%%sql
Select * from SPACEXDATASET
where launch_site like 'CCA%'
limit 5
```

```
 * ibm_db_sa://hyz17199:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.
```

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

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

total_mass	customer
45596	NASA (CRS)

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

average_mass	booster_version
2928	F9 v1.1

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

first_successful_landing_on_ground_pad
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

mission_outcome	COUNT
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

landing__outcome	booster_version	YEAR
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

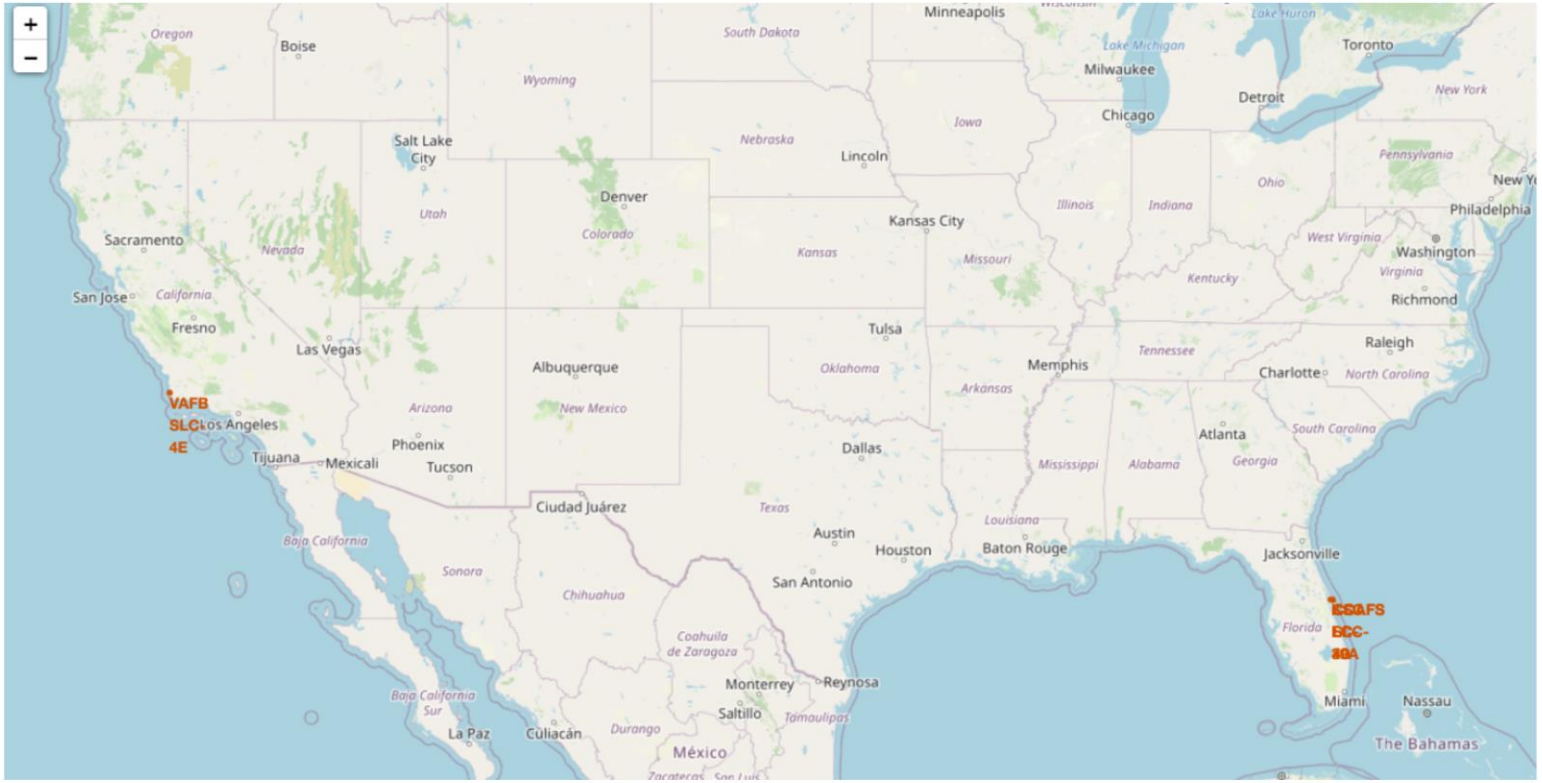
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

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

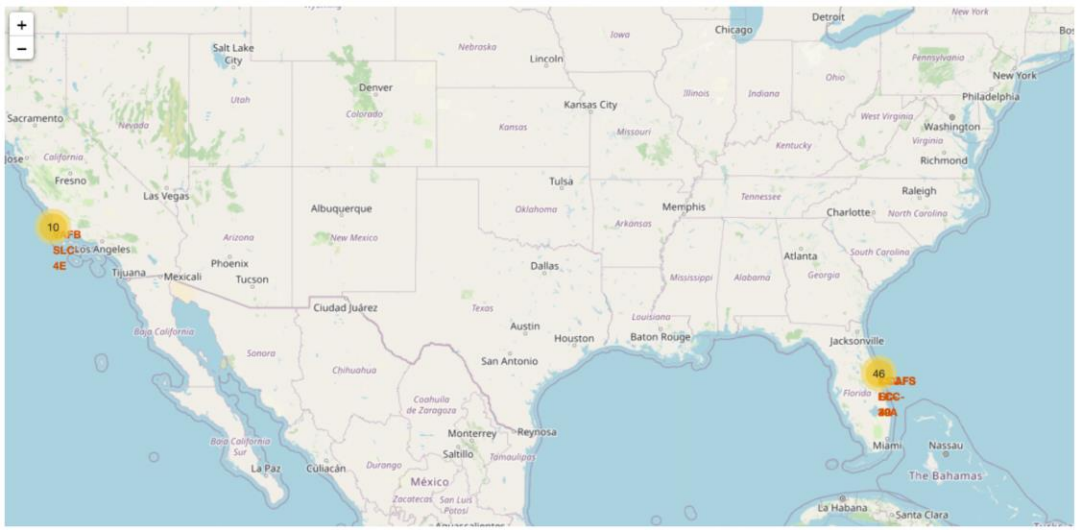

Section 3
Launch Sites
Proximities Analysis

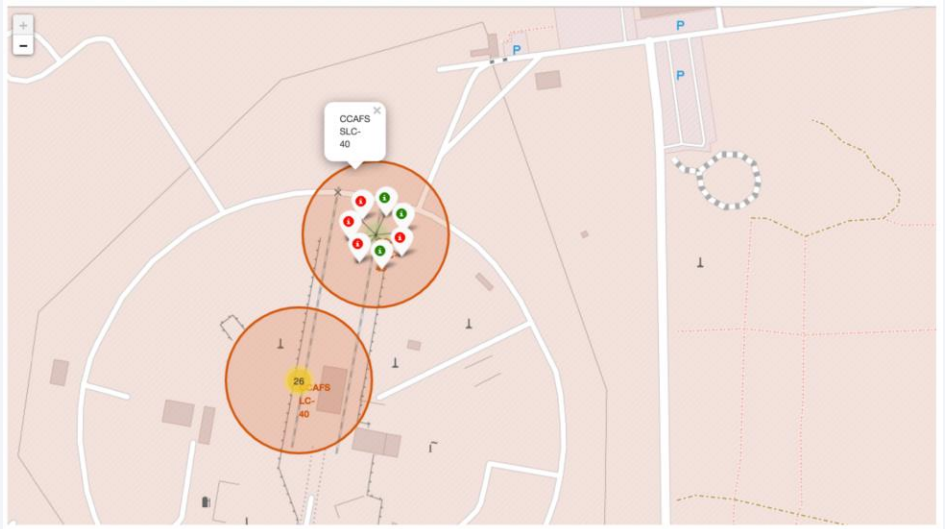
Launch Sites Map

- The launch sites are in Florida on the USA east coast and in California on the USA west coast
- Launch sites are near the southernmost U.S. mainland area



Sites and Launch outcomes

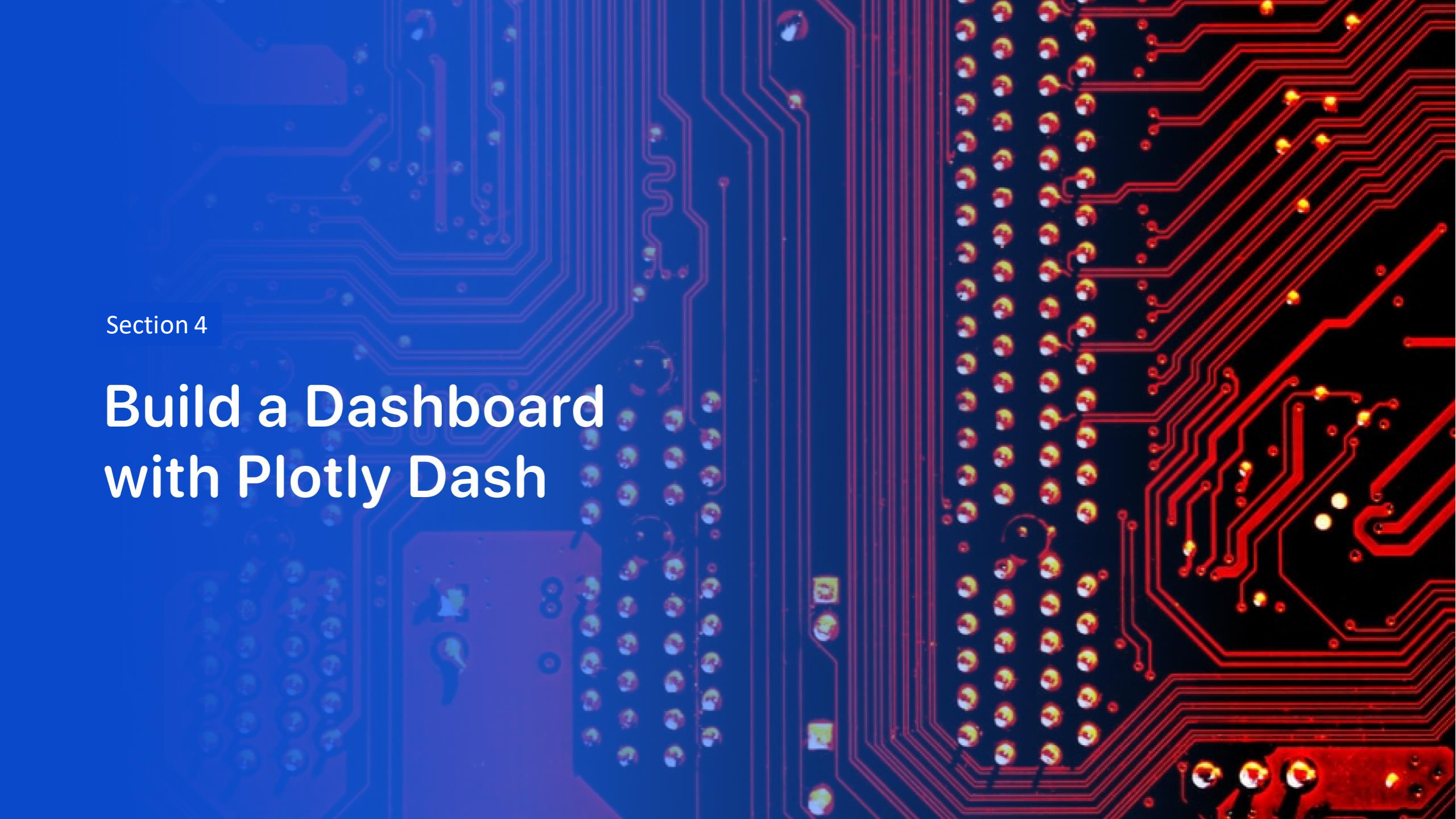




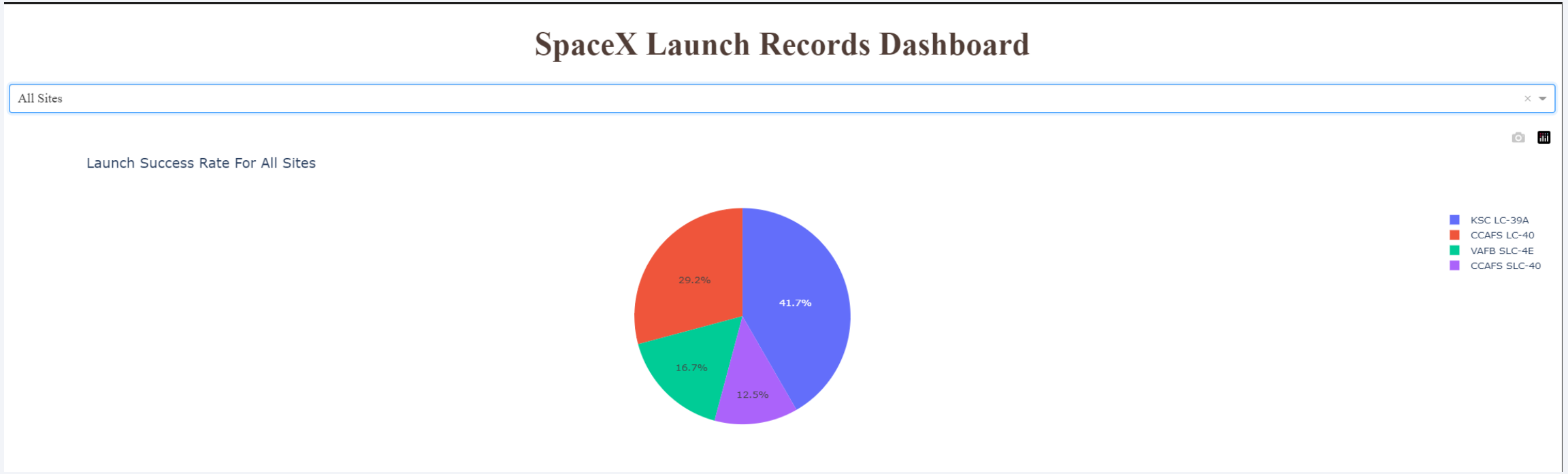
- Green colored are the successful launches
- red colored are the failed launches

Launch Site Proximities

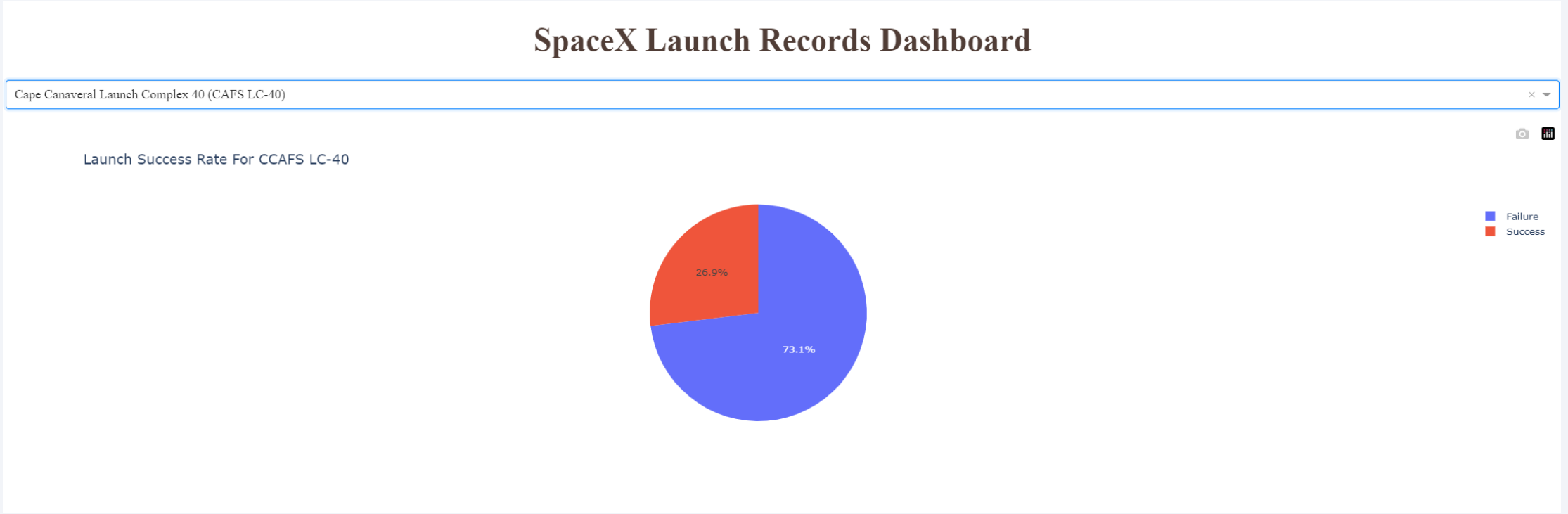
- Distance of launch site from east coast


Section 4
Build a Dashboard with Plotly Dash

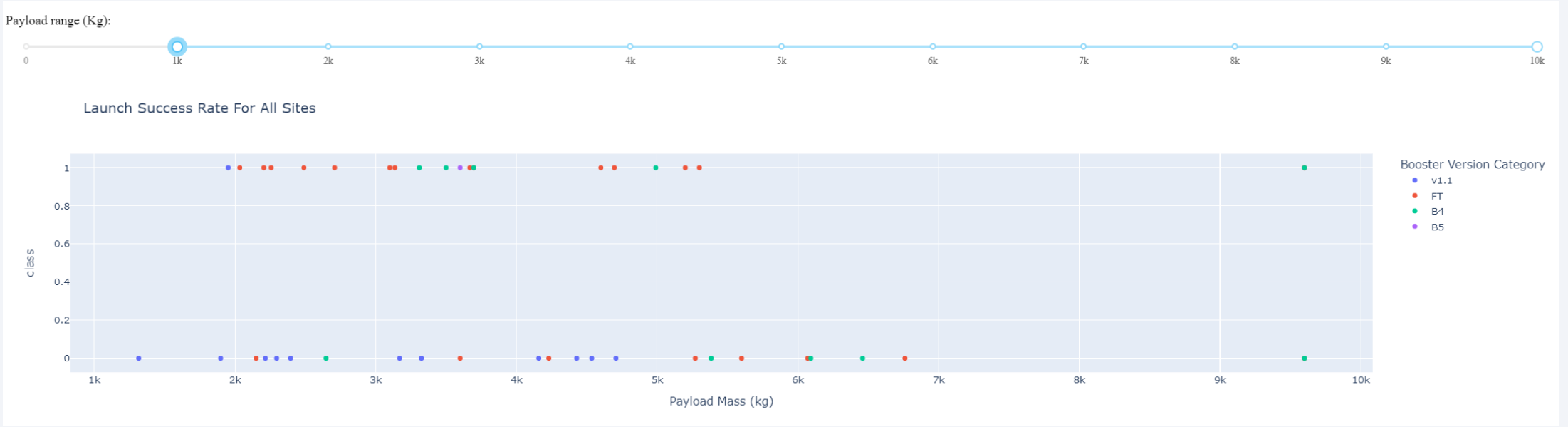
Launch Records for all sites



Launch Records for CAFS LC 40



Success by Payload and Booster Version

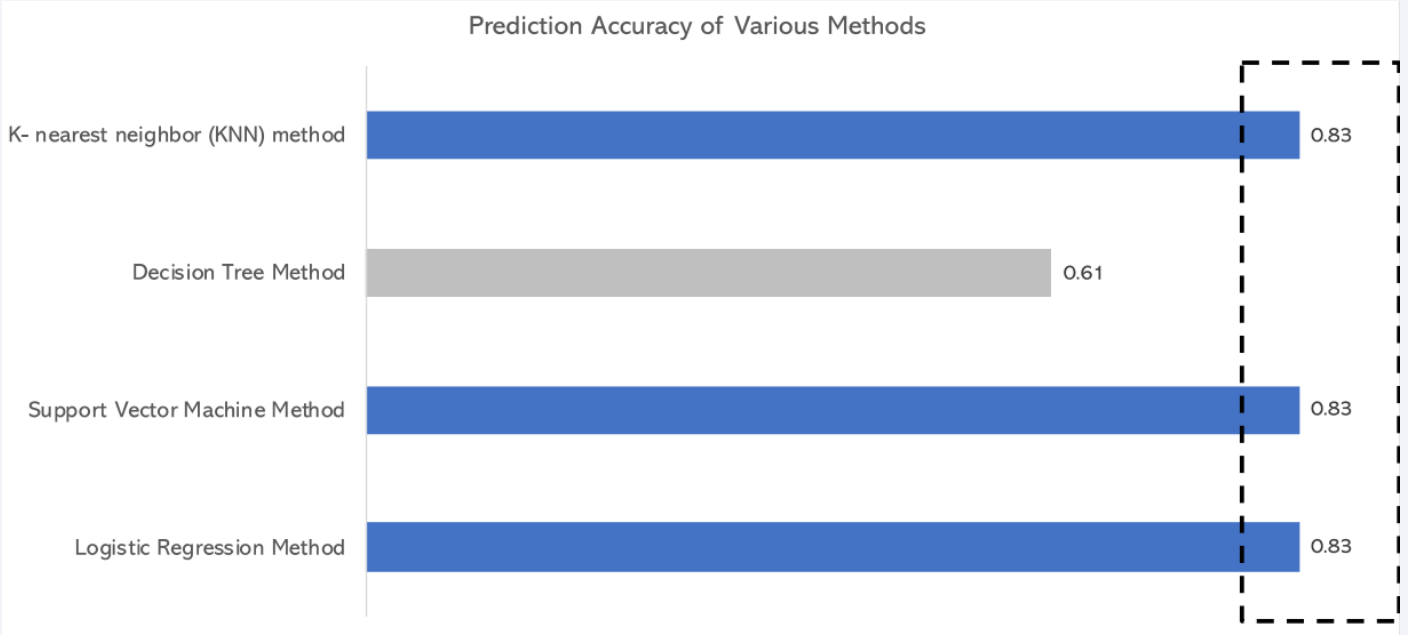


Section 5

Predictive Analysis (Classification)

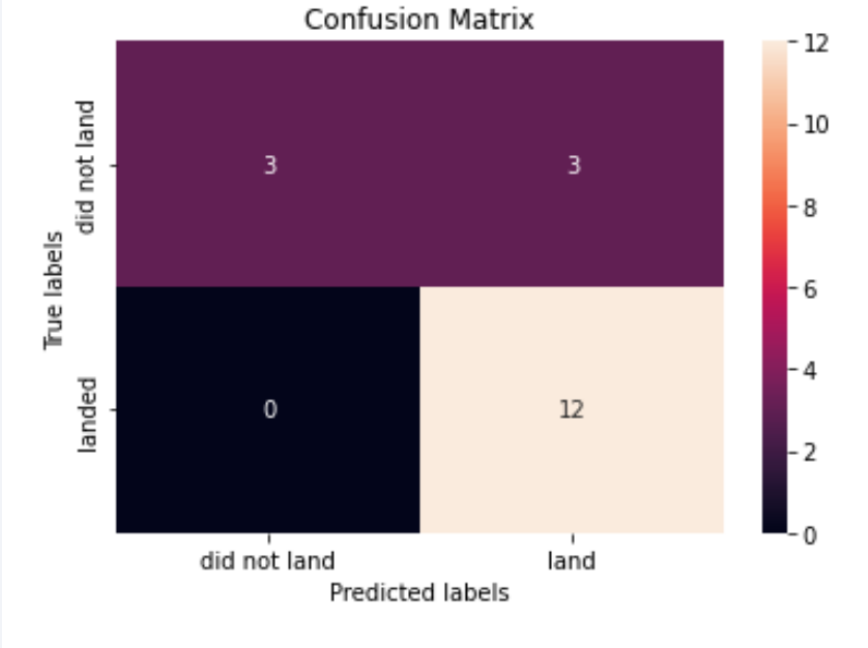
Classification Accuracy

- KNN, Support Vector and Logistic Regression Methods have high accuracy



Confusion Matrix

- True Positives: 12
- True Negatives: 3
- False Positives: 3
- False Negatives: 0



Conclusions

- KNN, Logistic Regression and SVM are the best classifier models for this dataset
- None of the models had false negatives
- All models had at least one false positive
- The lower payload launches have higher success rate than heavier payloads
- Site KSC LC-39A has the most successful launches from all sites
- F9 Booster versions v1.0, v1.1, FT, B4, B5 have the highest launch success rates
- The SpaceX launches have been continuously getting better from year 2013 to 2020

Thank you!