Winning Space Race with Data Science

Austin Yong
2022-11-03

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection using web scraping
  - Data Wrangling
  - Data Visualization
  - Machine Learning Prediction
- Summary of all results
  - Collected data from public sources and identified which features were best to predict success of launchings and used the ML prediction to show the best model

Introduction

- Project background and context
  - To evaluate the viability of the new company Space Y to compete with Space X
- Problems you want to find answers
  - How do we estimate the total cost for launches?
  - Where would be the best place to have launches?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Web Scraping
- Perform data wrangling
  - After collecting the data, I created a landing outcome label based on the outcome data after summarizing and analyzing the features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The data was normalized, divided into training and test data sets and evaluated by four different classification models

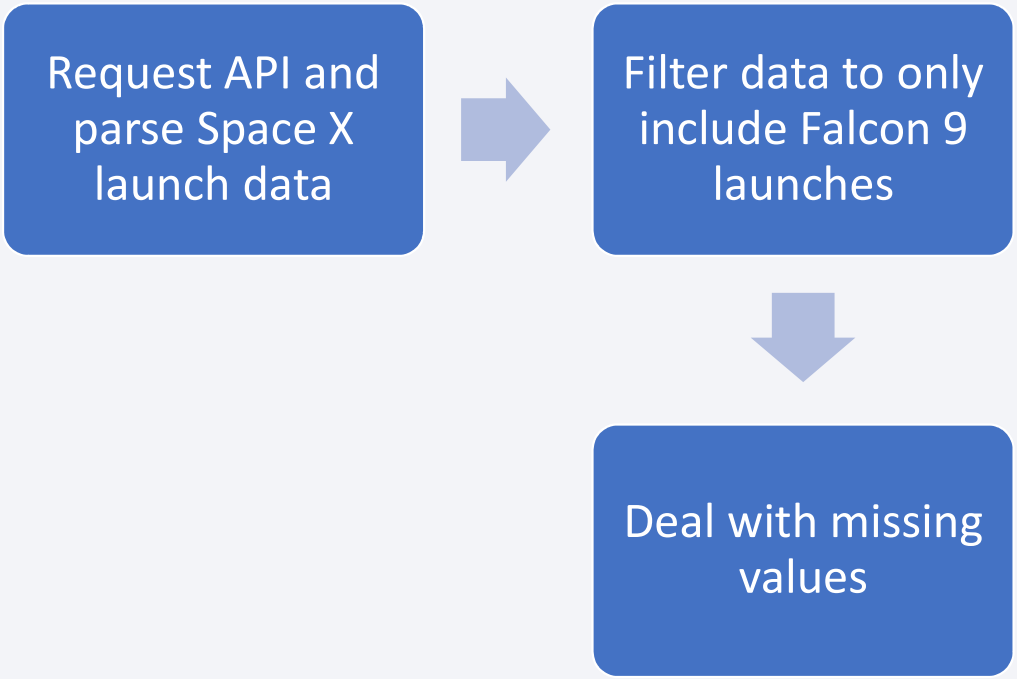
Data Collection

- Describe how data sets were collected.
  - Data sets were collected from Space X API and from Wikipedia
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

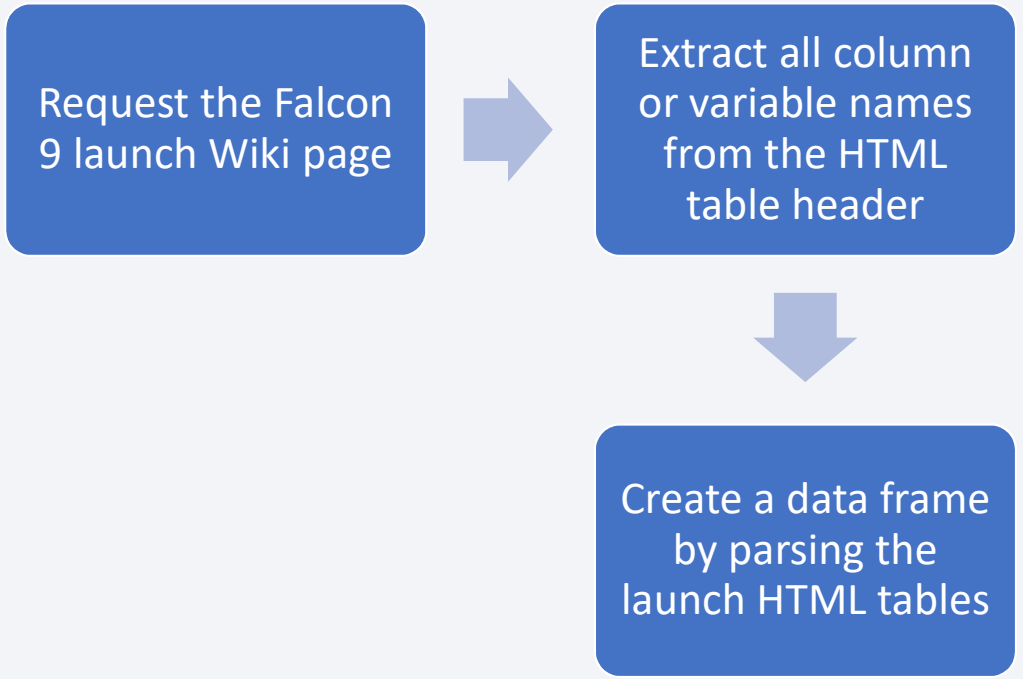
- https://github.com/astnyong/IBM-DS-Capstone-Project/blob/main/Data%20Collection.ipynb



Data Collection - Scraping

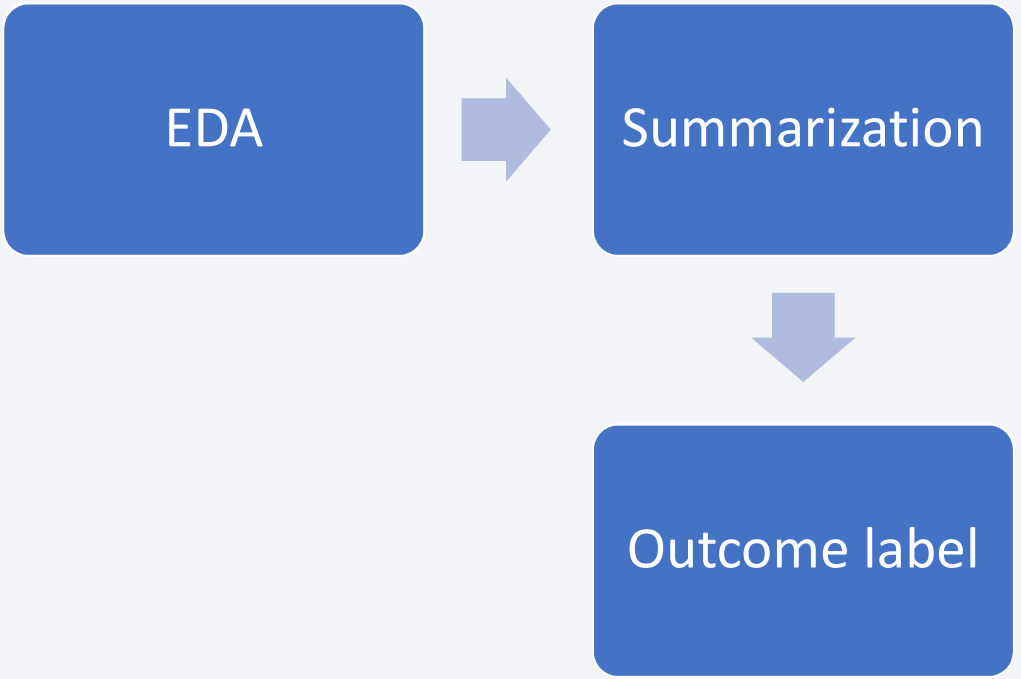
- Present your web scraping process using key phrases and flowcharts

- https://github.com/astnyong/IBM-DS-Capstone-Project/blob/main/Data%20Collection%20w%20Web%20Scraping.ipynb



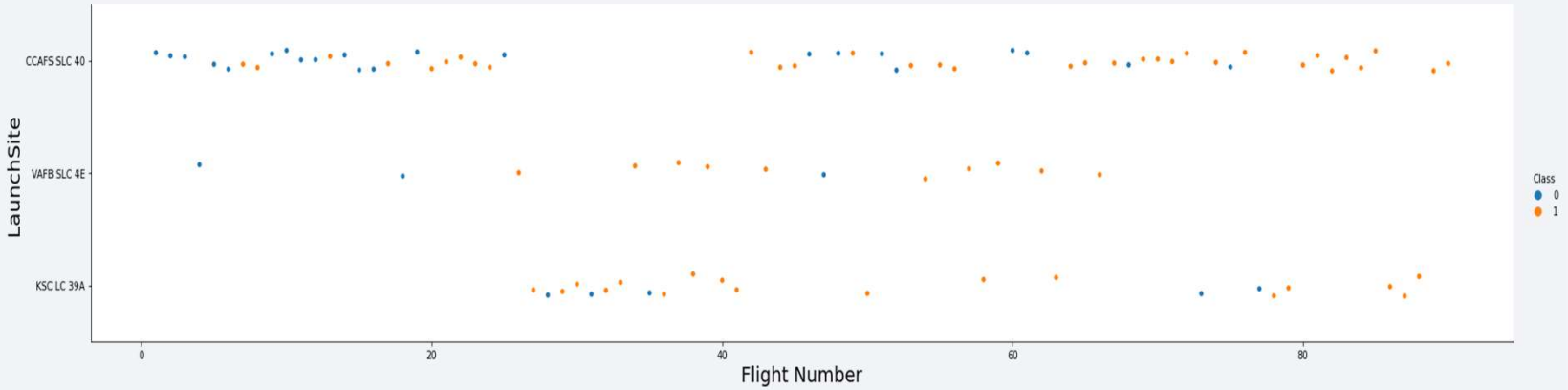
Data Wrangling

- Describe how data were processed
  - Exploratory data analysis was performed on the data set, then summarization, and finally the outcome label was created
- https://github.com/astnyong/IBM-DS-Capstone-Project/blob/main/Final%20Project%20-%20Data%20Wrangling.ipynb



EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
  - Scatterplots and barplots were used to visualize the relationship between pair of features:
    - Payload mass & flight number; launch site & flight number; launch site & payload mass; orbit & flight number; payload & orbit
    - https://github.com/astnyong/IBM-DS-Capstone-Project/blob/main/EDA%20w%20Data%20Visualization.ipynb



EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
  - Names of the unique launch sites in the space mission
  - Top 5 launch sites whose name begin with the string 'CCA'
  - Total payload mass carried by boosters launched by NASA(CRS)
  - Average payload mass carried by booster version F9 v1.1
  - Date when the first successful landing outcome in ground pad was achieved
  - Names of boosters which have success and payload mass between 4000 and 6000 kg
  - Total number of successful and failure mission outcomes
  - Names of the booster versions which have carried the maximum payload mass
  - Failed landing outcomes in drone ship, their booster versions, and launch site names for in
  - year 2015Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20
- https://github.com/astnyong/IBM-DS-Capstone-Project/blob/main/EDA%20w%20SQL.ipynb

Build an Interactive Map with Folium

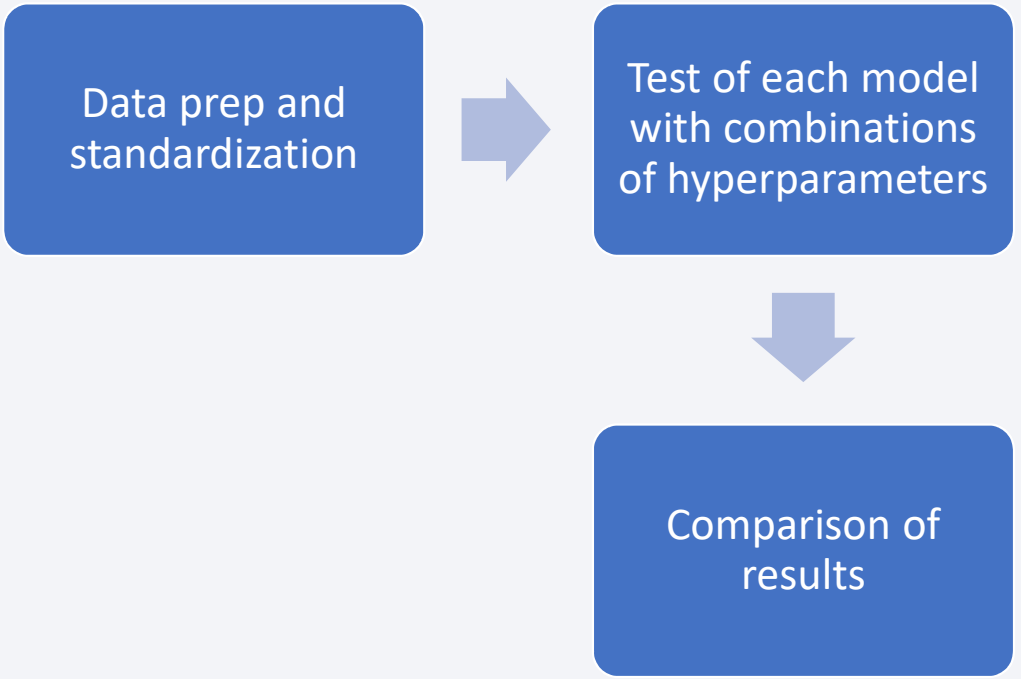
- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
  - Markers indicate points like launch sites;
  - Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space
  - Center;
  - Marker clusters indicates groups of events in each coordinate, like launches in a launch site; and
  - Lines are used to indicate distances between two coordinates
- https://github.com/astnyong/IBM-DS-Capstone-Project/blob/main/Interactive%20Visual%20Analytics%20w%20Folium.ipynb

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
  - Percentage of launches by site
  - Payload range
- This combination allowed me to analyze the relation between payloads and launch sites to help identify the best place to launch
- https://github.com/astnyong/IBM-DS-Capstone-Project/blob/main/spacex%20plotly%20dash.py

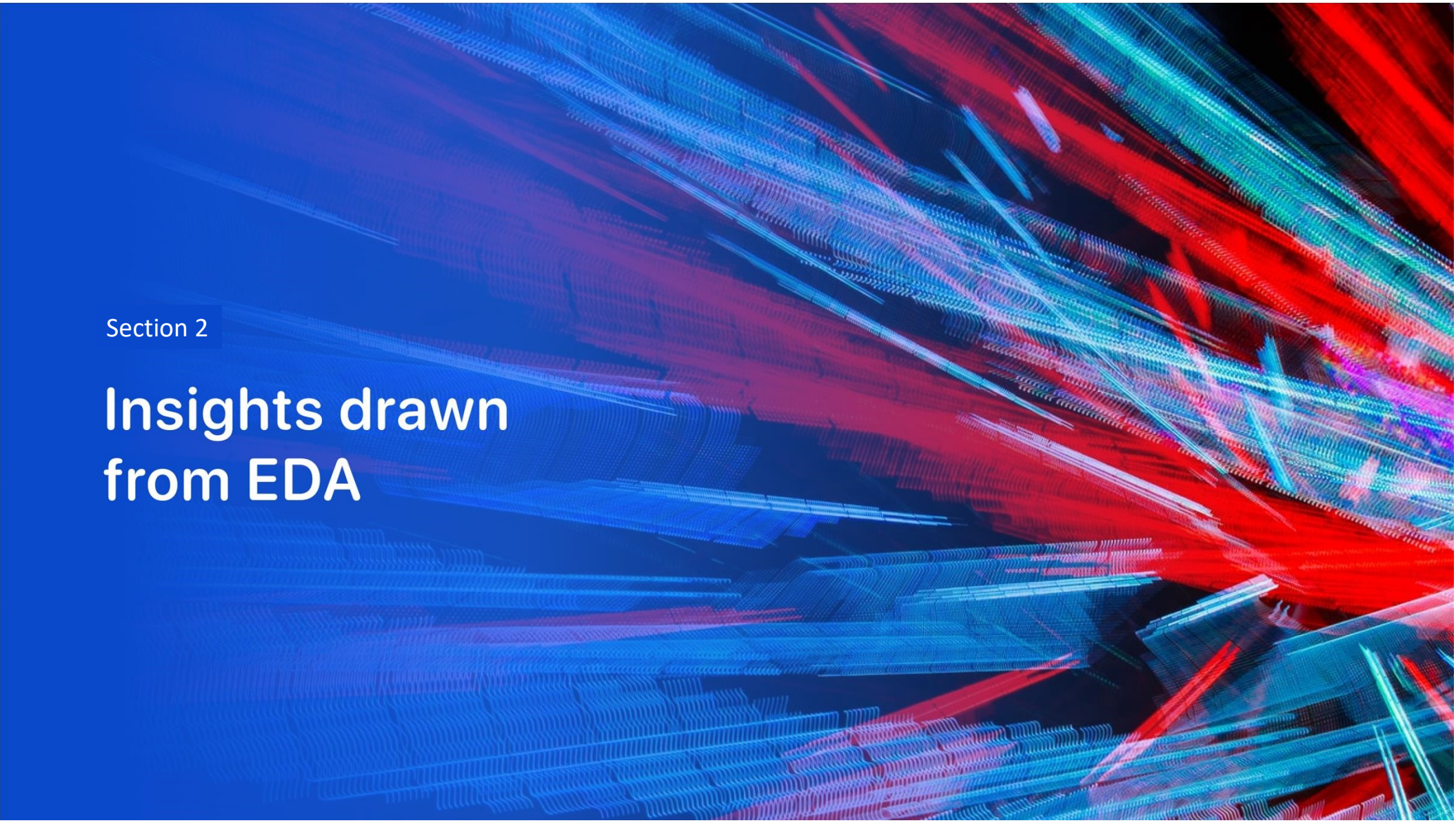
Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
    - Four classification models were compared: logistic regression, decision tree, k nearest neighbors, and support vector machine
- https://github.com/astnyong/IBM-DS-Capstone-Project/blob/main/ML%20Prediction.ipynb



Results

- Exploratory data analysis results
  - Space Xuses 4 different launch sites;
  - The first launches were done to Space Xitself and NASA;
  - The average payload of F9 v1.1 booster is 2,928 kg;
  - The first success landing outcome happened in 2015 fiver year after the first launch;
  - Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;
  - Almost 100% of mission outcomes were successful;
  - Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;
  - The number of landing outcomes became as better as years passed.

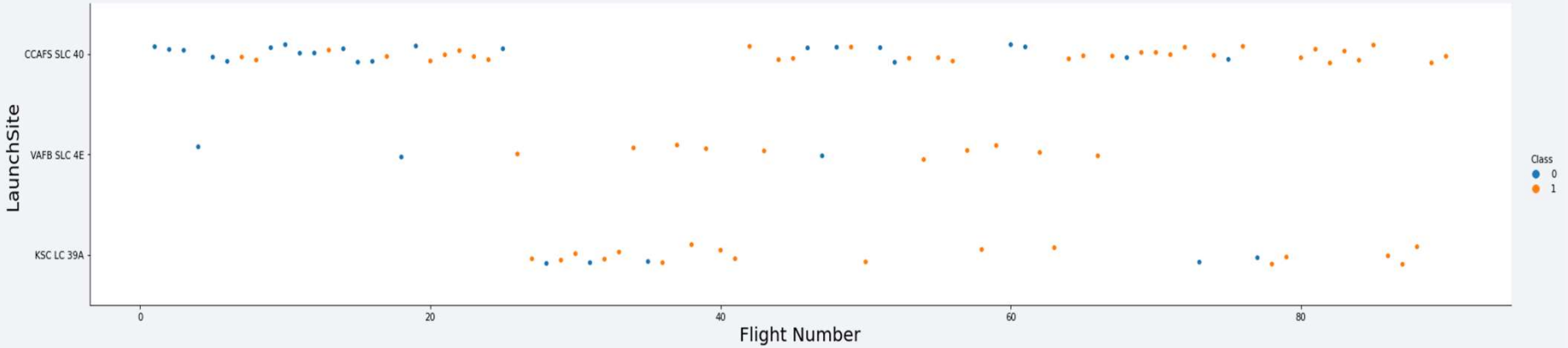


Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

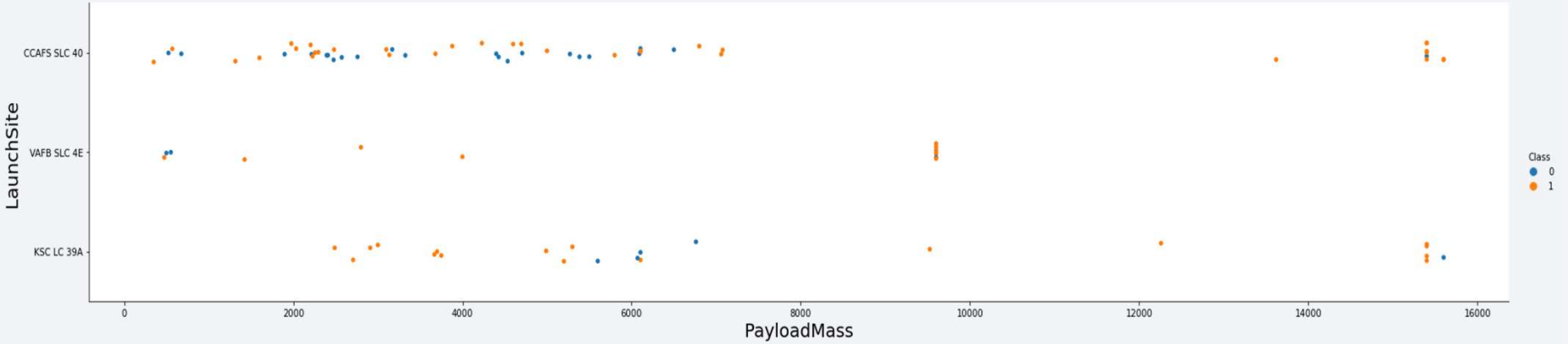
- Show a scatter plot of Flight Number vs. Launch Site



- According to the plot above, it's possible to verify that the best launch site nowadays is CCAF5 SLC40, where most of recent launcheswere successful;
- In second place VAFBSLC4E and third place KSCLC 39A;
- It's also possible to see that the general success rate improved overtime.

Payload vs. Launch Site

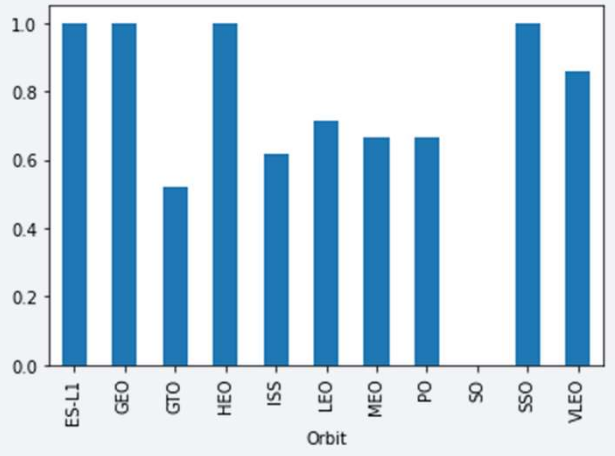
- Show a scatter plot of Payload vs. Launch Site



- Payloads over 9,000kg (about the weight of a school bus) have excellent success rate
- Payloads over 12,000kg seems to be possible only on CCAFSSLC40 and KSC LC 39A launch sites

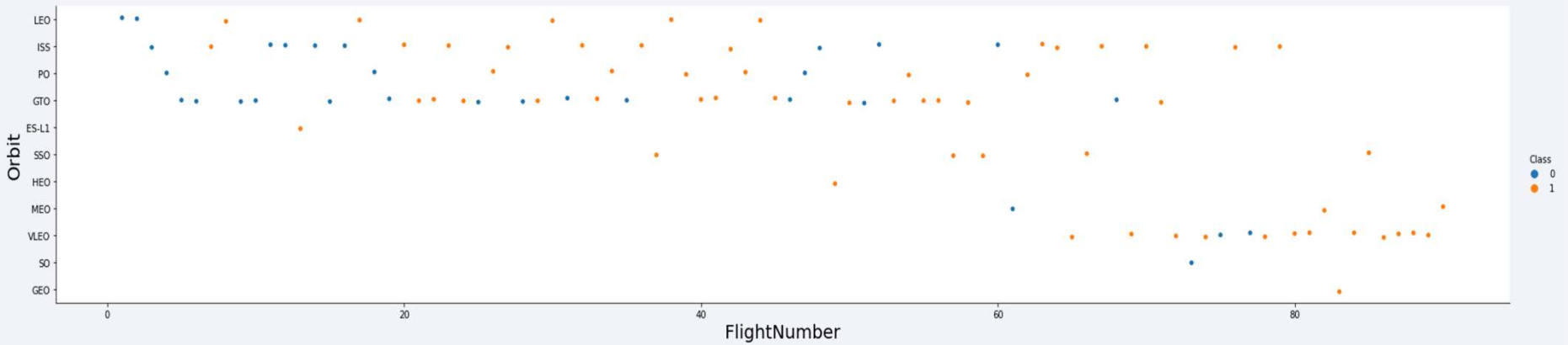
Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type
  - The biggest success rates happens to orbits:
    - ES-L1
    - GEO
    - HEO
    - SSO



Flight Number vs. Orbit Type

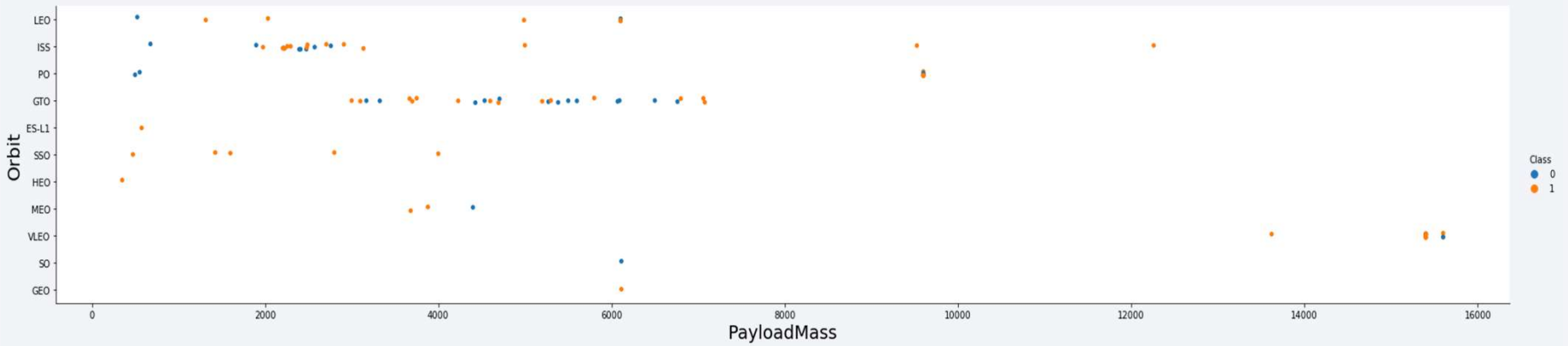
- Show a scatter point of Flight number vs. Orbit type



- Success rate improved over time to all orbits

Payload vs. Orbit Type

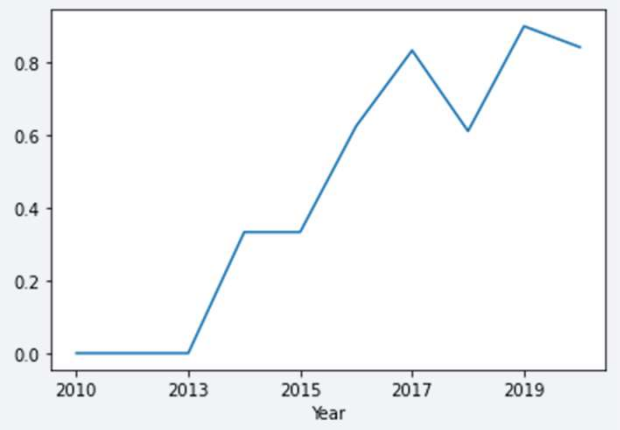
- Show a scatter point of payload vs. orbit type



- There is no relationship between payload and success rate to orbit
- There are few launches to the orbits SO and GEO

Launch Success Yearly Trend

- Show a line chart of yearly average success rate
  - Success rate increased in 2013 until 2020
  - The first 3 years were most likely research and development



All Launch Site Names

- Find the names of the unique launch sites
  - They are obtained by selecting unique occurrences of “launch_site” values from the dataset.

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites 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 attemp

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

Total Payload (kg)
111.268

Average Payload Mass by F9 v1.1

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

Avg Payload (kg)
2.928

First Successful Ground Landing Date

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

Min Date
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	Occurrences
Success	99
Success (payload status unclear)	1
Failure (in flight)	1

Boosters Carried Maximum Payload

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

Booster Version
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

Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

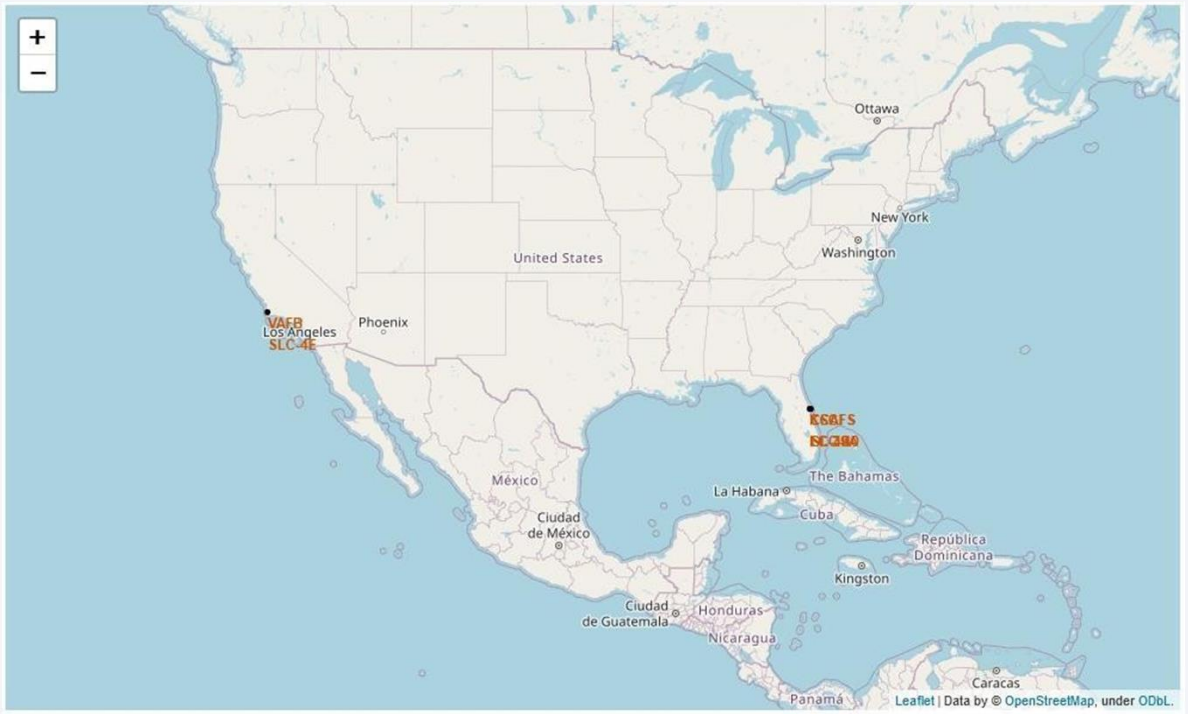
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Landing Outcome	Occurrences
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

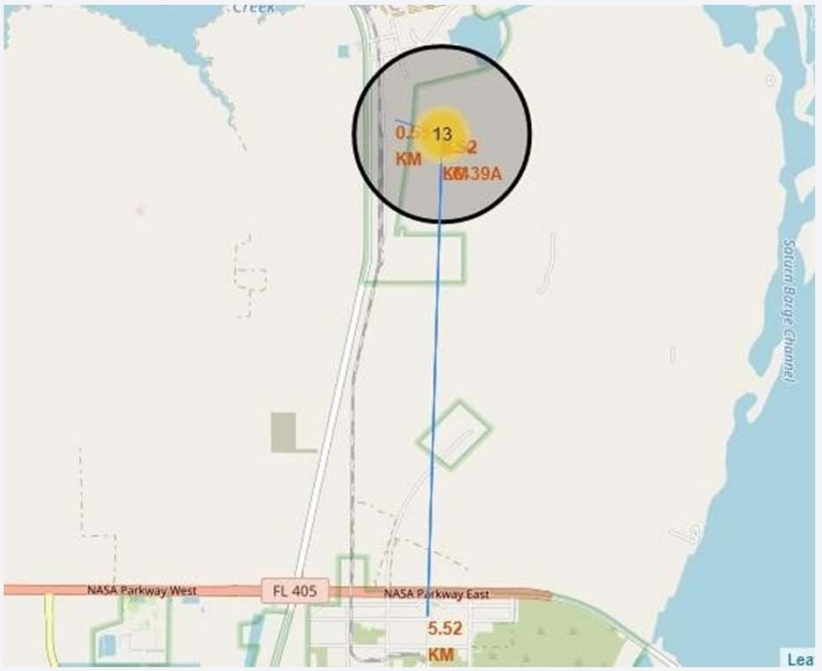
Launch Sites Proximities Analysis

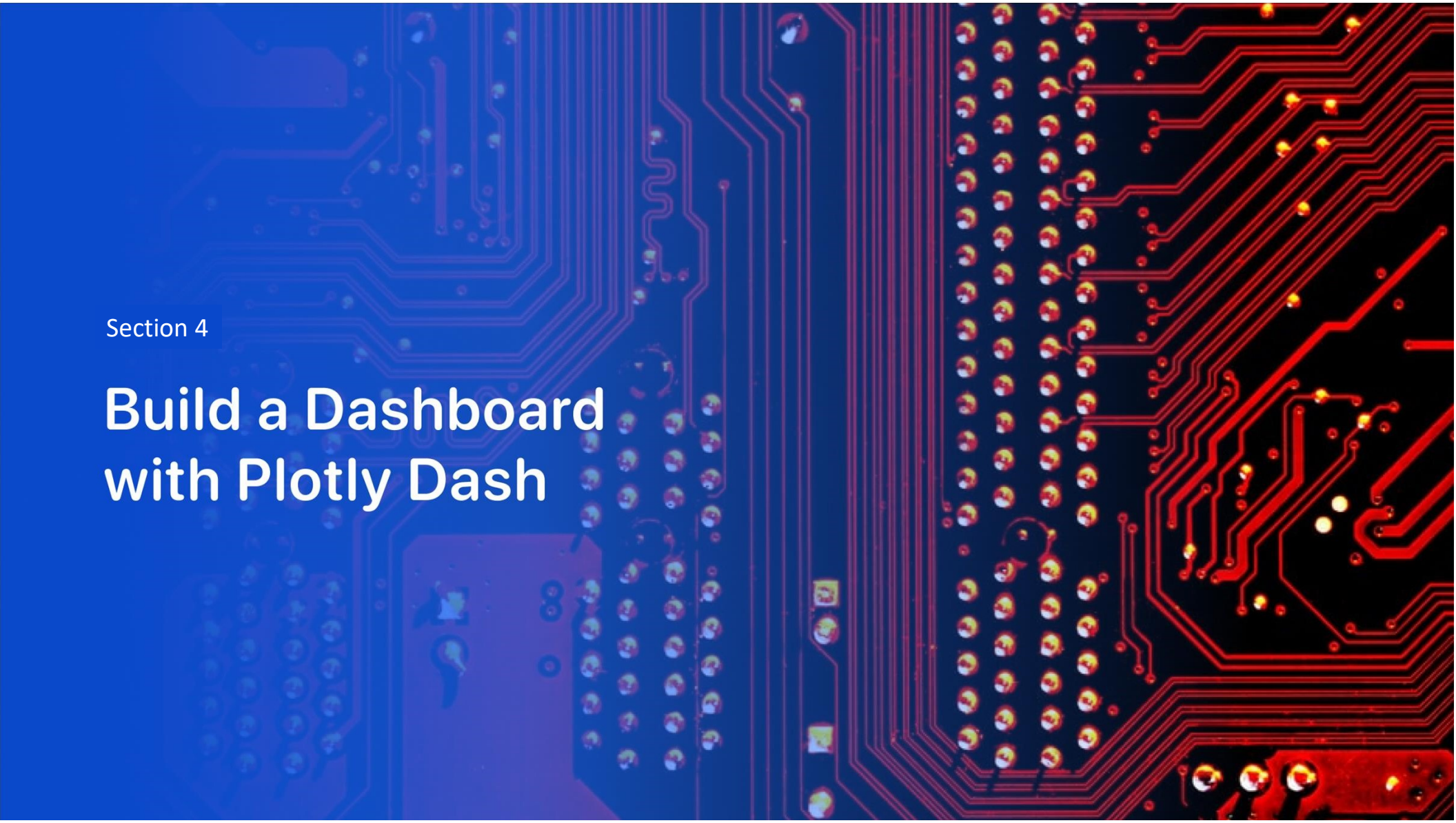
All Launch Sites



Launch Outcomes by Site

Logistics and Safety

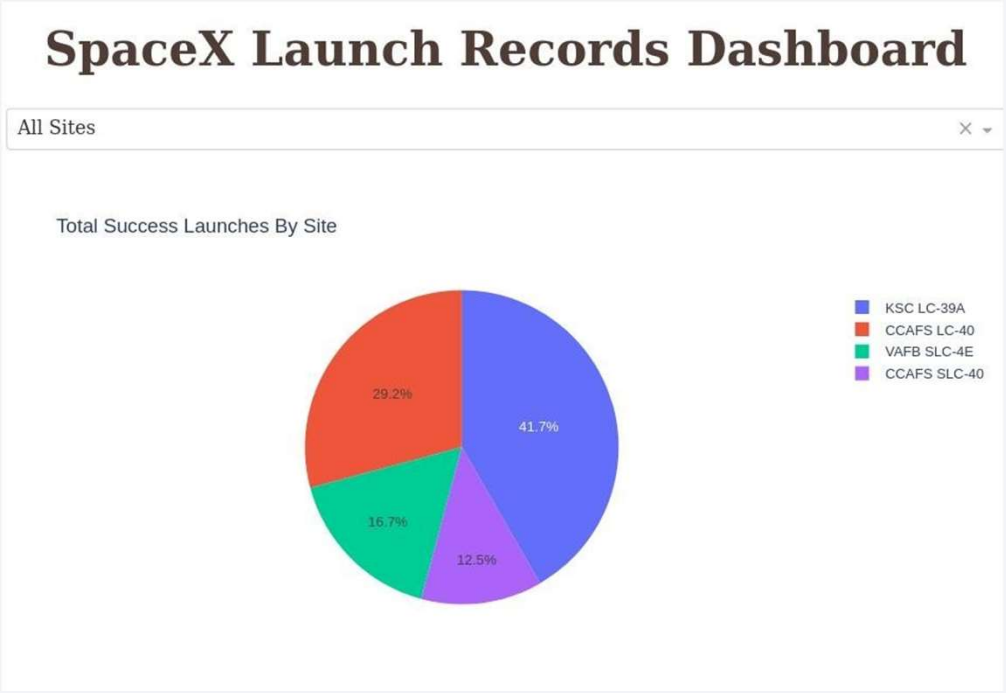




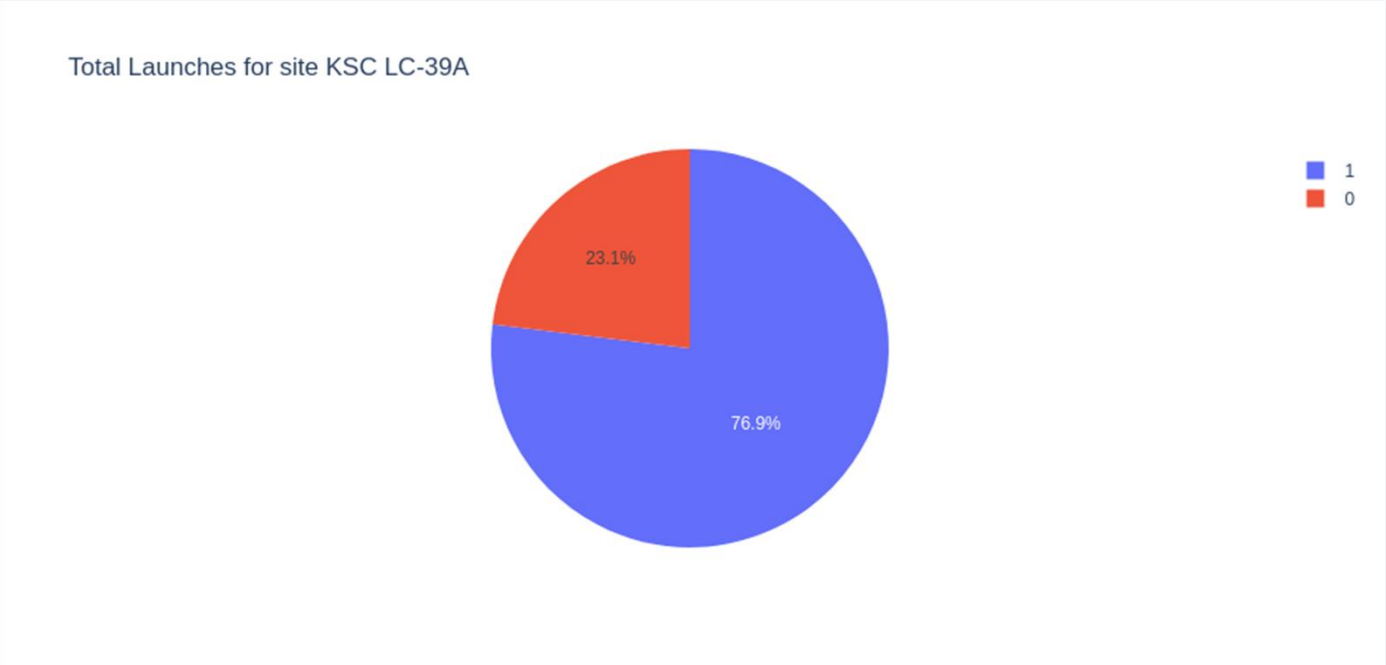
Section 4

Build a Dashboard with Plotly Dash

Successful Launches by Site



Launch Success Ratio for KSCLC-39A



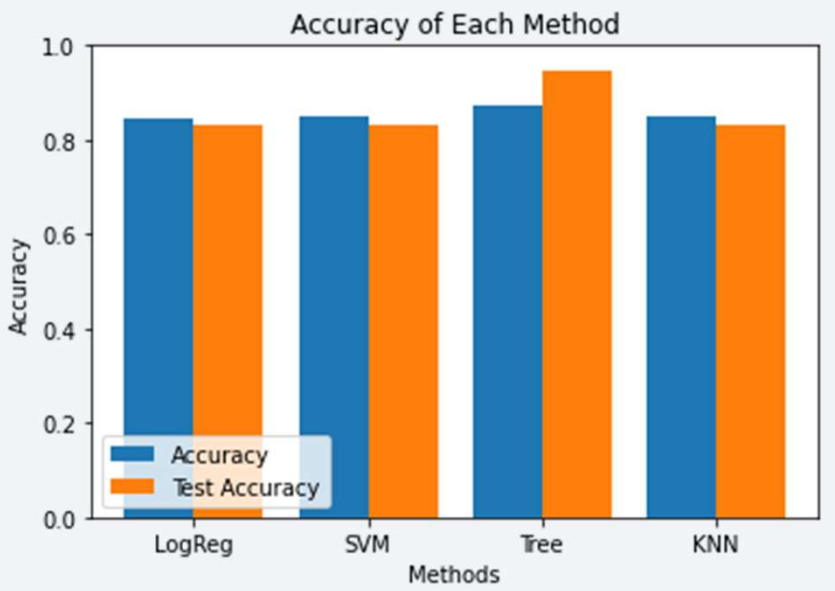
Payload vs. Launch Outcome

Section 5

Predictive Analysis (Classification)

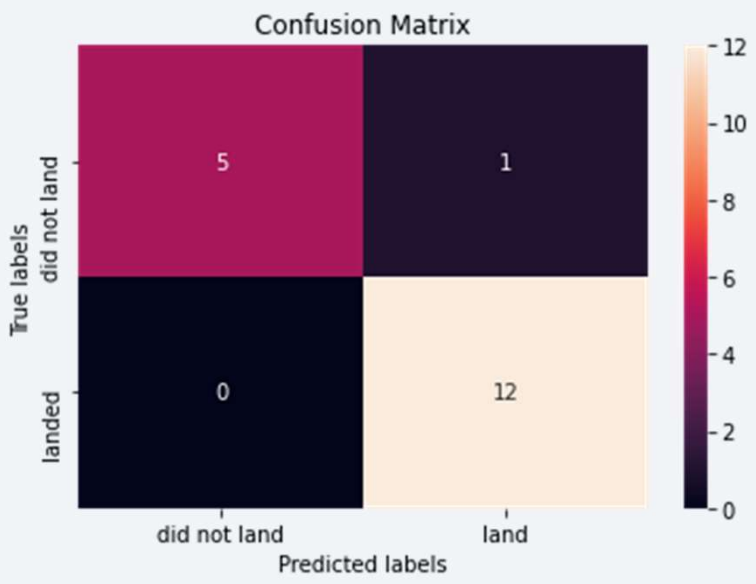
Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart
  - The model with the highest classification accuracy is the decision tree classifier



Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation



Conclusions

- Different data sources were analyzed, refining conclusions along the
- process;
- The best launch site is KSCLC-39A;
- Launches above 7,000kg are less risky;
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets;
- Decision Tree Classifier can be used to predict successful landings and
- increase profits.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!