IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Seth Farris
27-Aug-2024

Outline

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

Executive Summary

- Methodologies
  - Data collection API and Web Scraping
  - SQL
  - Folium
  - Plotly
  - Machine Learning
- Results
  - Flight/landing success
  - Proximity analyses
  - Dashboard

Introduction

- This is a final project for the Applied Data Science Capstone class. We used various data science methods to explore SpaceX rocket data
- We explored questions such as what influences flight success, where are the launch sites, and how do payloads affect success of launches.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected by downloading publicly available SpaceX rocket data
- Perform data wrangling
  - Data were process using machine learning, SQL, Python
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

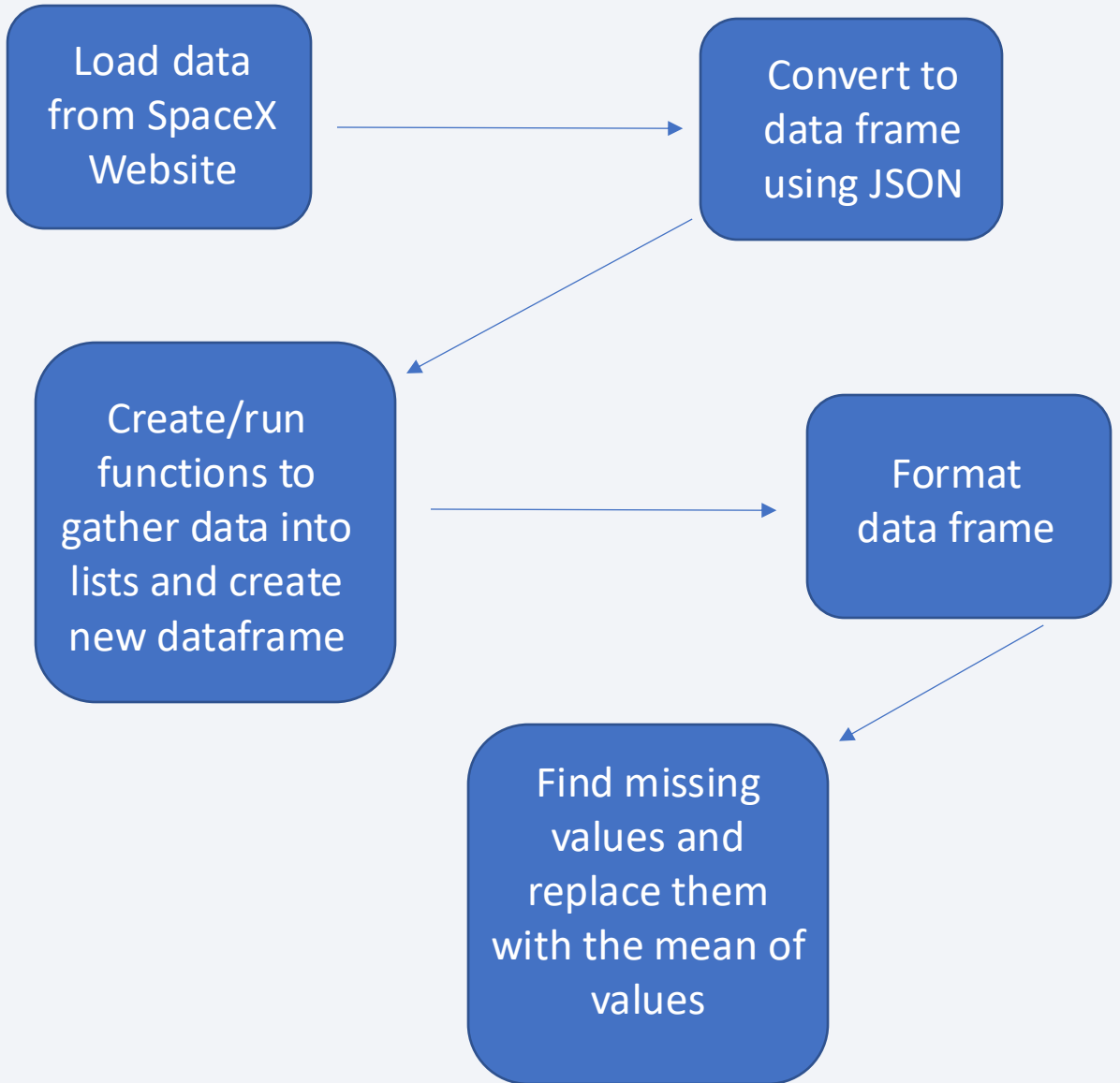
Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- First we loaded data from "https://api.spacexdata.com/v4/launches/past" then converted into a data frame before cleaning it up for analyses

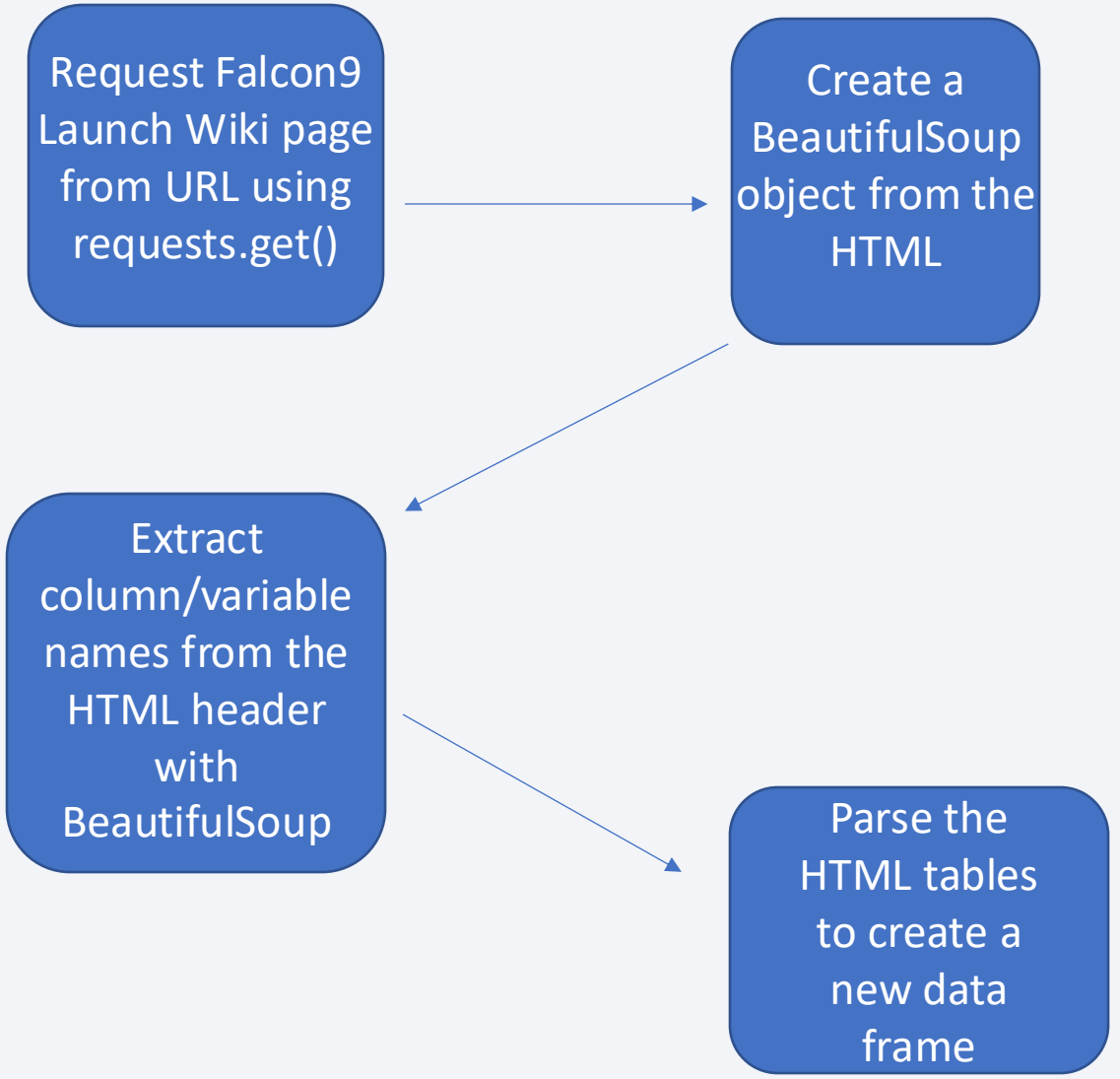
- GitHub URL: https://github.com/sethcfarris/AppliedDataScienceCapstone/blob/1073977da40656771c9308ae33e3f24b4eb4e9bb/1_jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

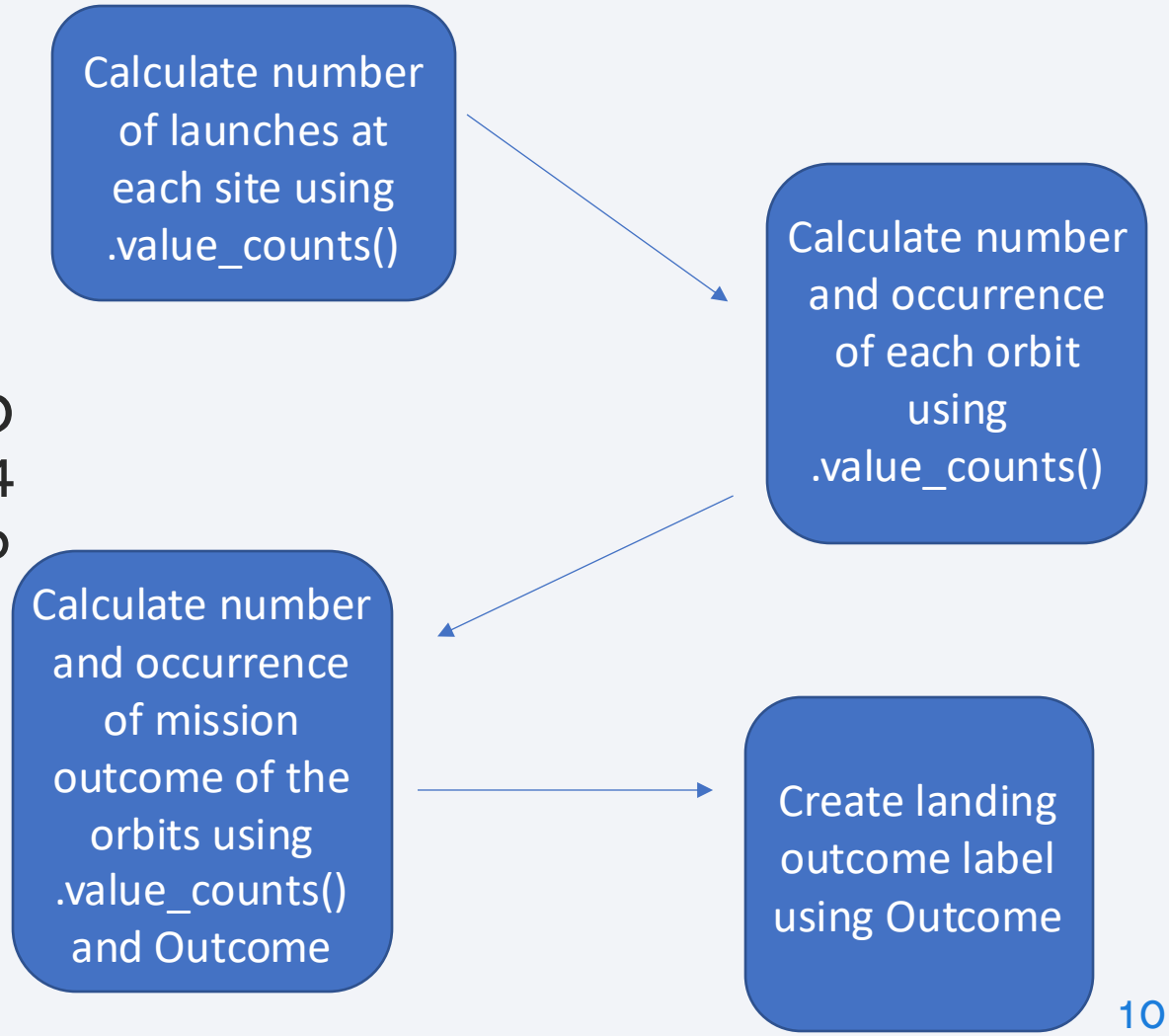
- We web scraped Wikipedia with BeautifulSoup for Falcon 9 Launch records

- GitHub URL: https://github.com/sethcfarris/AppliedDataScienceCapstone/blob/1073977da40656771c9308ae33e3f24b4eb4e9bb/2_jupyter-labs-webscraping.ipynb



Data Wrangling

- Using .value_counts() and Outcome, we calculated number of occurences and outcomes of landings
- GitHub URL: https://github.com/sethcfarris/AppliedDataScienceCapstone/blob/1073977da40656771c9308ae33e3f24b4eb4e9bb/3_labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Launch Site vs Flight Number scatter point chart
  - Easy way to visualize the relationship between the number of successful and unsuccessful flights at each launch site
  - CCAFS SLC 40 had the most overall launches
  - KSC LC 39A seems to have the highest ratio of successful flights to unsuccessful flights
- Payload Mass vs Launch Site scatter point chart
  - Payload seemed to have more of an affect than launch site
  - Fewer flights over 10000 kg, but most of those were successful
- Success rate of each orbit bar chart
  - Easily displays average success rate for each orbit
- GitHub URL: https://github.com/sethcfarris/AppliedDataScienceCapstone/blob/1073977da40656771c9308ae33e3f24b4eb4e9bb/5_edadataviz.ipynb

EDA with SQL

- Use “select” to
  - display the names of each unique launch site used for the space mission
  - Display 5 where launch sites begin with letters “CCA”
  - Find and display the sum of all payload mass carried by boosters launched by NASA
  - Find and display the sum of all payload mass carried by boosters launched by booster version F9 v1.1
  - List the booster version of the successful drone ship landings and have a payload mass greater than 4000 but less than 6000
- Use min function to find the date of the first successful landing
- https://github.com/sethcfarris/AppliedDataScienceCapstone/blob/1073977da40656771c9308ae33e3f24b4eb4e9bb/4_jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

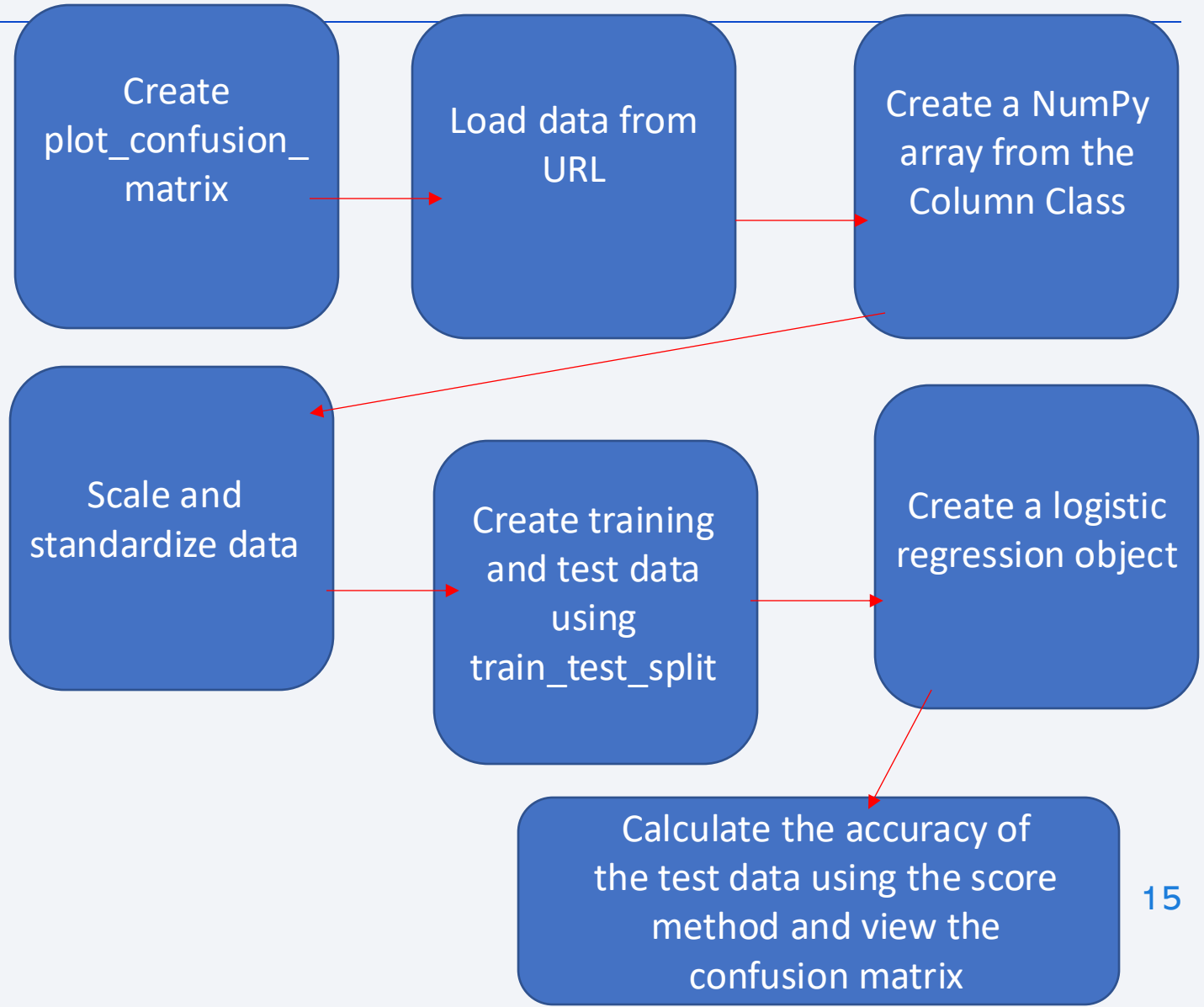
- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Circles
  - Highlighted circle around a specific area on the map. This makes it easier to find locations on the map from a distance
- Markers
  - A mark on top of a location on a map. Makes it easier to find locations on map when zoomed in
- Cluster
  - Combines circles/markers into one point when zoomed out. Keeps the map looking clean.
- Mouse_Position
  - Displays the coordinates of where your mouse is on the map. Makes it easier to find locations.
- https://github.com/sethcfarris/AppliedDataScienceCapstone/blob/1073977da40656771c9308ae33e3f24b4eb4e9bb/6_lab_jupyter_launch_site_location_folium.ipynb

Build a Dashboard with Plotly Dash

- Pie chart showing Total Success Launches by Sites
- Payload range
- Success count scatter plot of payload mass
- You can pick which site or all sites you would like to view information about. This makes it easier to digest multiple sources of information at once.
- https://github.com/sethcfarris/AppliedDataScienceCapstone/blob/1073977da40656771c9308ae33e3f24b4eb4e9bb/SpaceXDashboard_all.png

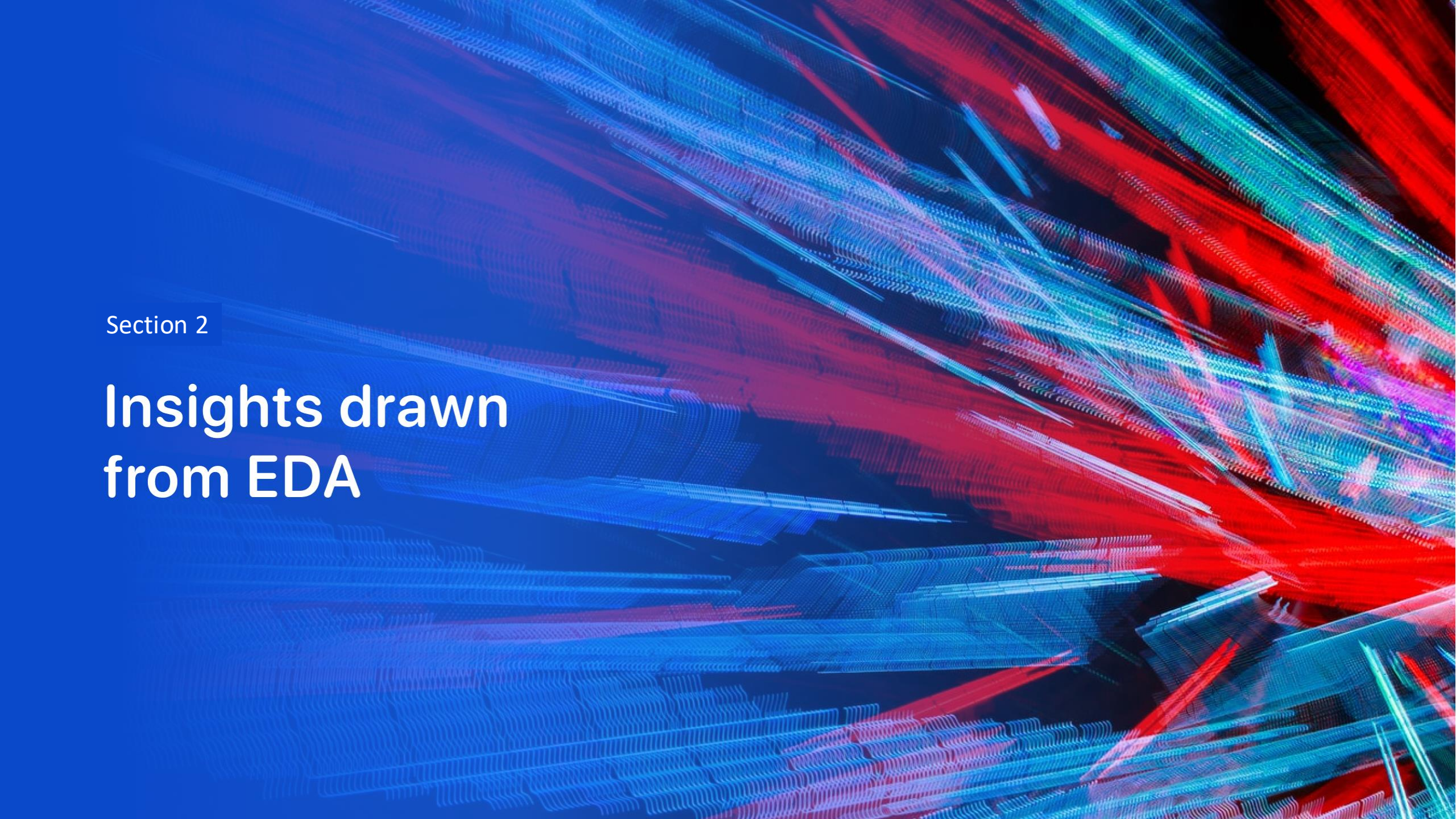
Predictive Analysis (Classification)

- https://github.com/sethcfarris/AppliedDataScienceCapstone/blob/1073977da40656771c9308ae33e3f24b4eb4e9bb/8_SpaceX_Machine%20Learning%20Prediction.ipynb



Results

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


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

- Relationship between Flight Number and Launch Site where 0 equals unsuccessful and 1 equals successful

Payload vs. Launch Site

- Payload vs. Launch site where 0 equals unsuccessful and 1 equals successful

Success Rate vs. Orbit Type

- Bar chart displaying which orbits have the highest success rates where 0 = unsuccessful and 1 = successful

Flight Number vs. Orbit Type

- Examining the relationship between flight number and orbit type where 0 = unsuccessful and 1 = successful

Payload vs. Orbit Type

- Scatter point examining relationship between pay load mass and orbit with successful landing where 0 = unsuccessful and 1 = successful

Launch Success Yearly Trend

- Line chart showing the trend in rate of success for each year

All Launch Site Names

Launch Site Names Begin with 'CCA'

Total Payload Mass

Average Payload Mass by F9 v1.1

First Successful Ground Landing Date

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

- Present your query result with a short explanation here

Total Number of Successful and Failure Mission Outcomes

Boosters Carried Maximum Payload

2015 Launch Records

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

Section 3

Launch Sites Proximities Analysis

Location of SpaceX Launch Sites

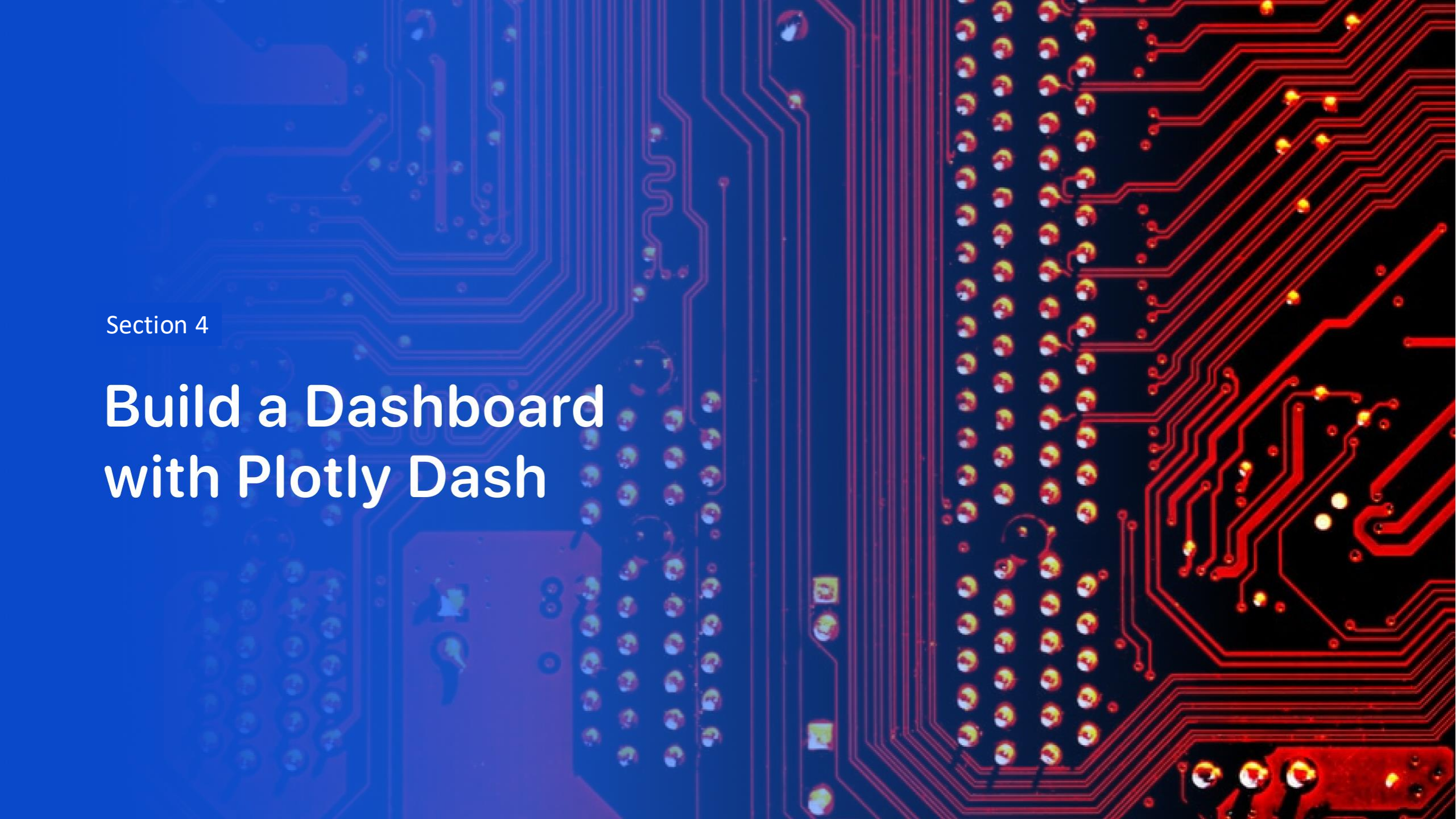
- Locations of launch sites on each US coast
- Circle marker showing location with site label names
- Markers are clustered. Because we are zoomed out on the map, all of the sites on each coast appear to be in the same spot.

Success Rates of Launches at KSC

- Success rates for this launch site are clustered. When selected, the split so that they can be examined individually
- Red = Unsuccessful
- Green = Successful

Distance with Line on Map

- This shows the launch site with a line and distance to the closest coast line


Section 4
Build a Dashboard with Plotly Dash

Dashboard Pie Chart

- Pie chart showing total percent of successful launches from all sites.

Most Successful Launch Site

- This pie chart shows that almost 77% of the launches at KSC LC-39A were successful while 23% were unsuccessful

Payload Range Slider

- By moving the Payload range slider, the success count scatter plot below changes accordingly

Section 5

Predictive Analysis (Classification)

Classification Accuracy

Logistic regression has the best score

Confusion Matrix

- Confusion matrix for the logistic regression method

Conclusions

- Scatterplots are useful to show a lot of information/relationships at once but can be hard to find specific values
- Launches from KSC have the highest chance of being successful
- Launch Successes have been trending upwards as time moves on
- Payload and Launch Site seem to have the most affect on success
- Logistic Regression has the best predictive analysis score

Appendix

- In order to code in the CloudIDE shell environment, you must remember to run “python3”

Thank you!