Winning Space Race with Data Science

Gurdeep Gurnasinghani

October 29th 2021

Outline

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

Executive Summary

- Methodologies Used:
  - Data Collection
  - Data Wrangling
  - EDA Data Visualization
  - EDA with SQL
  - Interactive Map with Folium
  - Predictive Analysis(Classification)
- Summary of all results
  - EDA results
  - Predictive Analysis results

Introduction

- Project background and context

Space X wants to determine the success rate of the falcon 9 rockets based on various variables and conditions.

- Problems you want to find answers:

What variables determine if the rocket will land successfully?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using SpaceX data API using Requests package
- Perform data wrangling
  - Data was processed using pandas and requests library from the SpaceX API
- 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

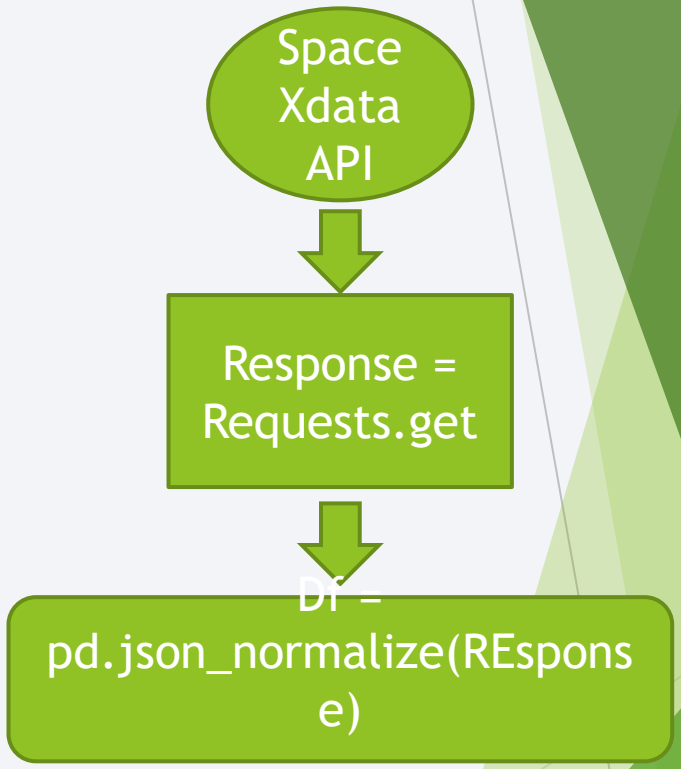
Data Collection

- Describe how data sets were collected.
  - SpaceX Rest API using Requests library
  - Web scraping with Beautiful Soup library

Data Collection – SpaceX API

- Data collection with SpaceX REST calls using key phrases and flowcharts

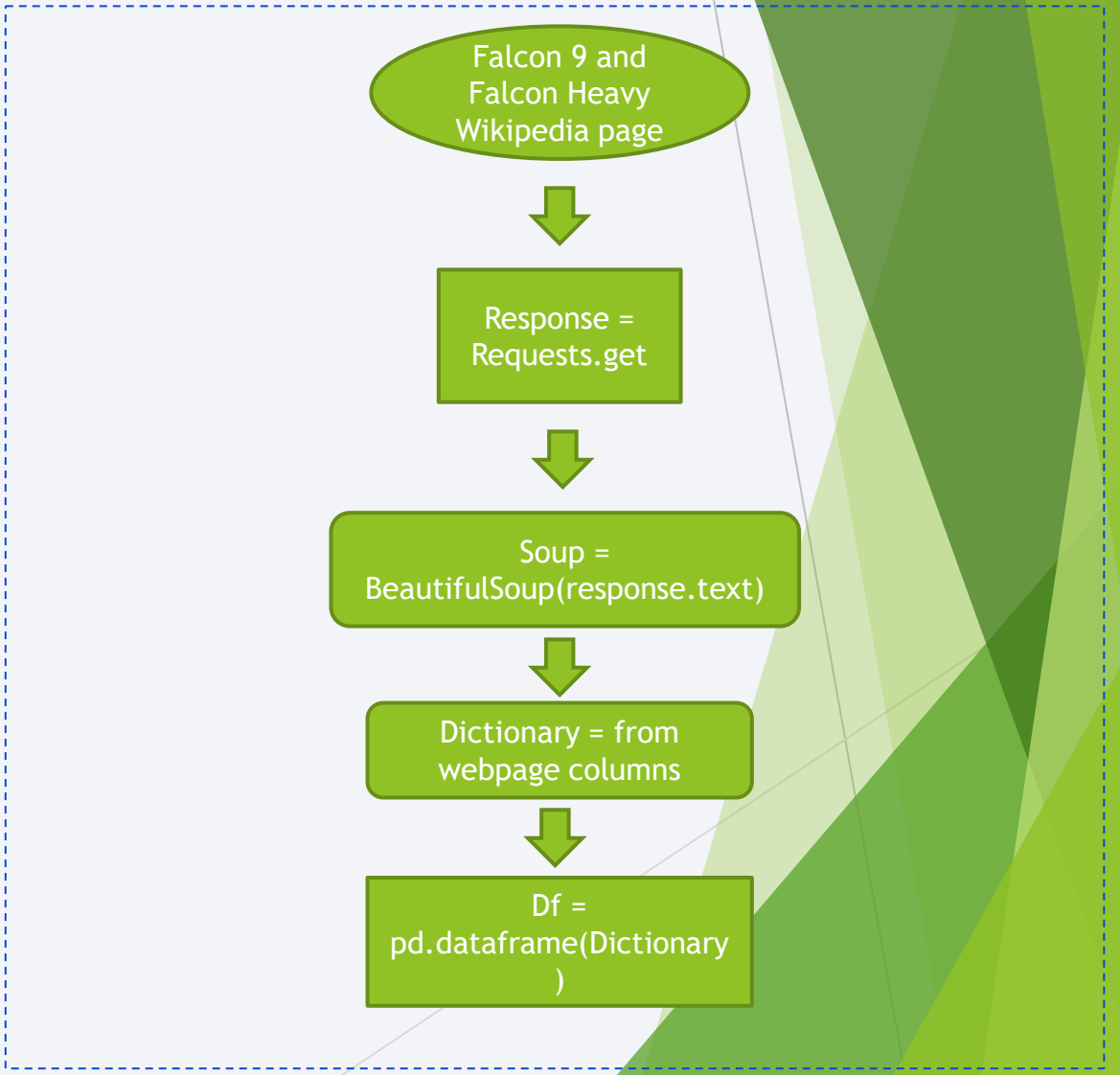
- Add the GitHub URL: https://github.com/GurdeepGurna/IBM_DataScience_Capstone/blob/master/Data%20Collection%20API%20Lab.ipynb



Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- GitHub URL: https://github.com/GurdeepGurna/IBM_DataScience_Capstone/blob/master/Data%20Collection%20with%20Web%20Scraping%20lab.ipynb



Data Wrangling

- Describe how data were processed:
  - Data was loaded into a pandas dataframe
  - Missing values in columns were identified
  - Column types were identified numerical and categorical
  - Number of launches per site was calculated
  - Number and occurrences of each orbit was calculated
  - Number and occurrences of mission outcome per orbit type was calculated
  - Landing outcome column was created and added to dataframe
- GitHub URL: https://github.com/GurdeepGurna/IBM_DataScience_Capstone/blob/master/Data%20Wrangling.ipynb

EDA with Data Visualization

- Charts used:
  - relationship between Flight Number and Launch Site
  - relationship between Payload and Launch Site
  - relationship between success rate of each orbit type
  - relationship between FlightNumber and Orbit type
  - relationship between Payload and Orbit type
  - launch success yearly trend
- GitHub URL: https://github.com/GurdeepGurna/IBM_DataScience_Capstone/blob/master/EDA%20with%20Visualization.ipynb

EDA with SQL

- SQL Queries Performed:
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites 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. Use a subquery
  - 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 2017-03-20, in descending order
- Add the GitHub URL: https://github.com/GurdeepGurna/IBM_DataScience_Capstone/blob/main/EDA%20with%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
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

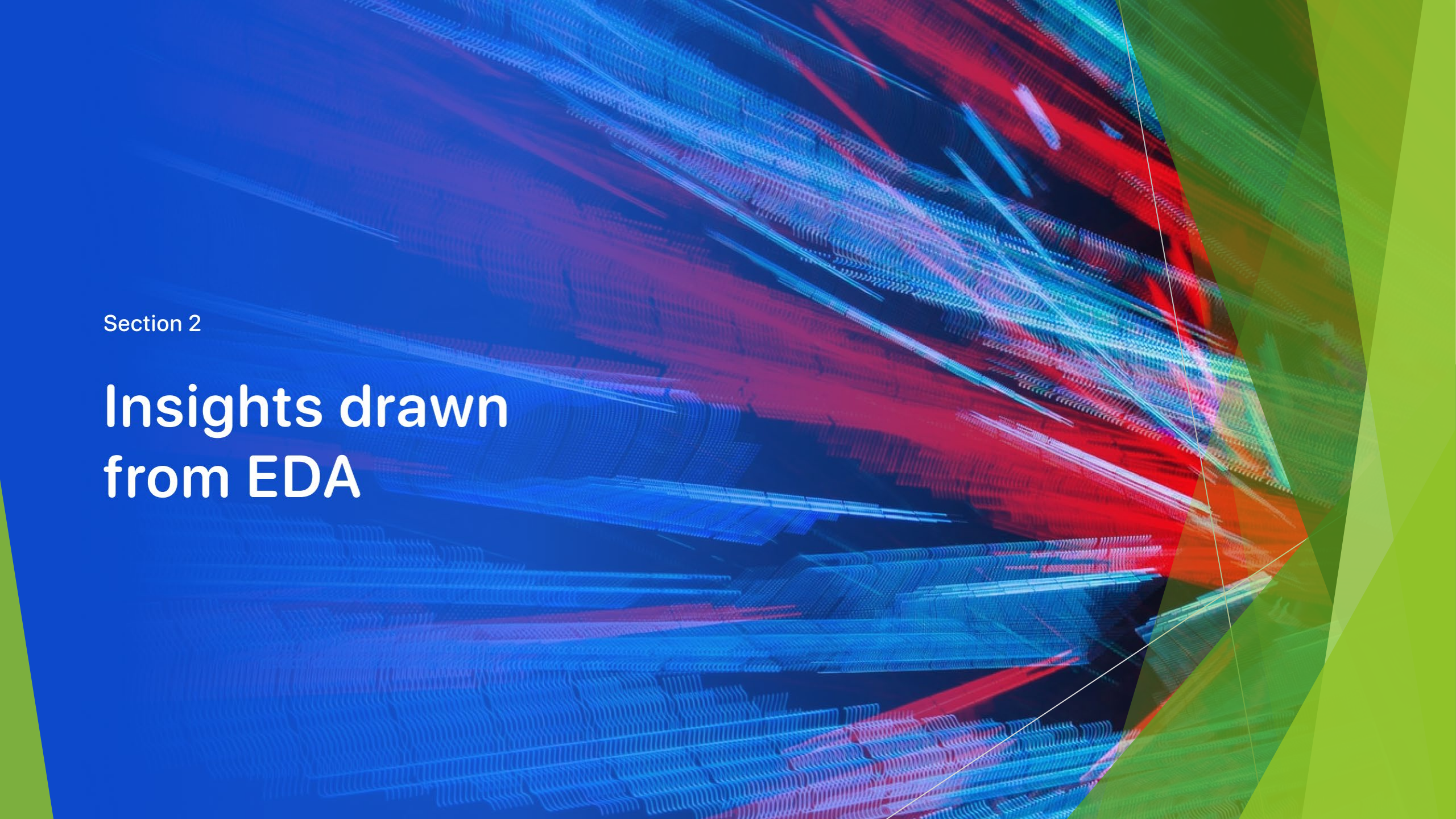
- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

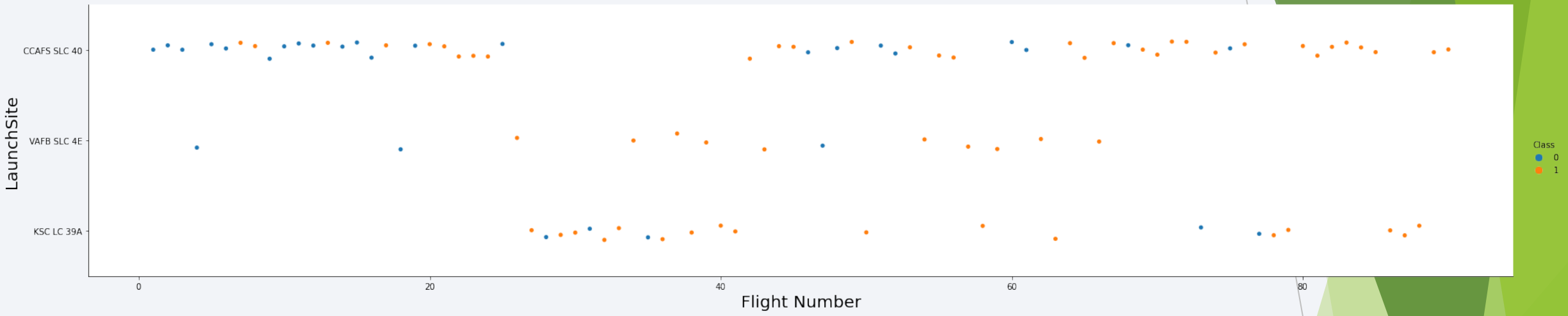
- Data was processed using Pandas and Numpy
- Split into train and test data sets
- Machine learning model used:
  - knn_cv
  - tree_cv
  - svm_cv
  - logreg_cv
- Accuracy of models accessed
- GitHub URL: https://github.com/GurdeepGurna/IBM_DataScience_Capstone/blob/master/Machine%20Learning%20Prediction.ipynb

Results

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


Section 2
Insights drawn from EDA

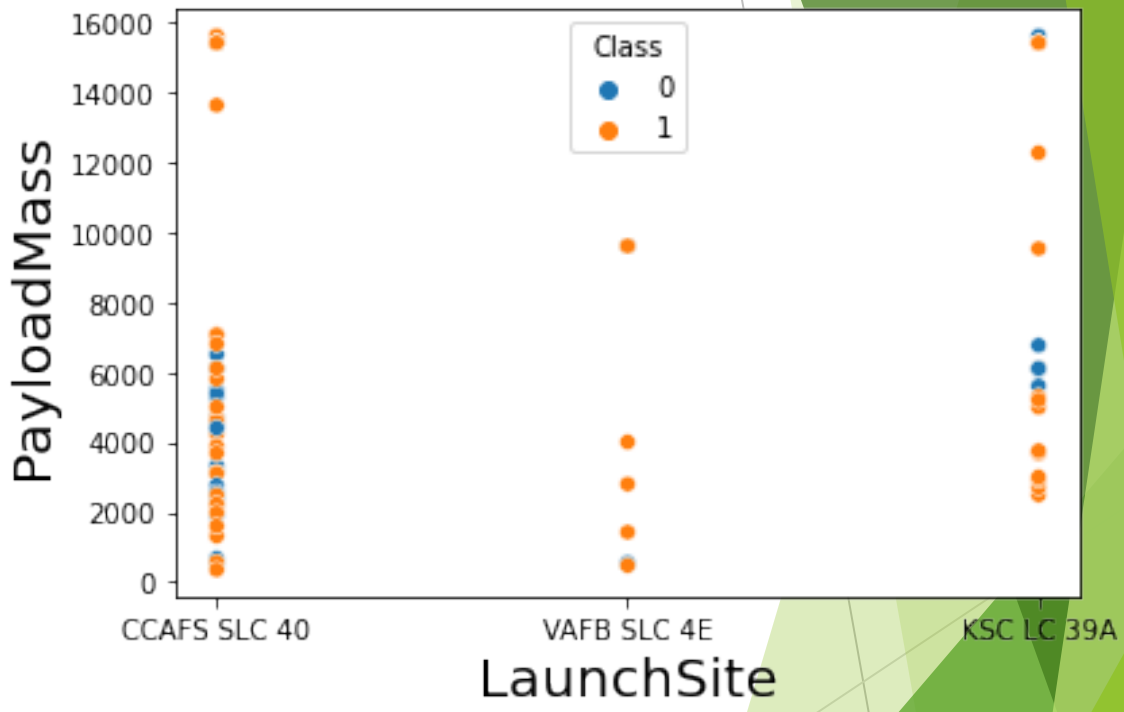
Flight Number vs. Launch Site



- Above the relationship between the launch sites and flight numbers can be seen
- The launch site has CCAFS SLC 40 has the most flights and the highest rate of success

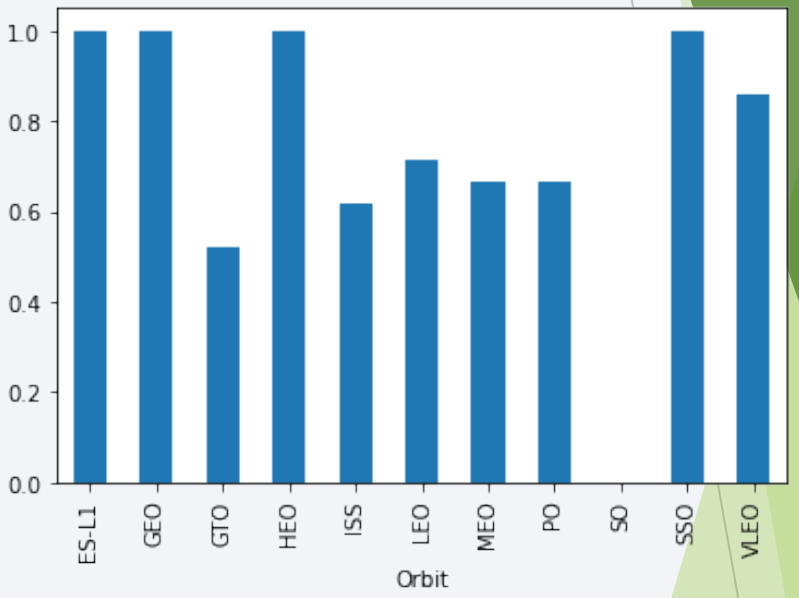
Payload vs. Launch Site

- Majority of Payload Masses for CCAFS SLC 40 are between 1000 and 7000
- VAFB SLC has the least number of launches



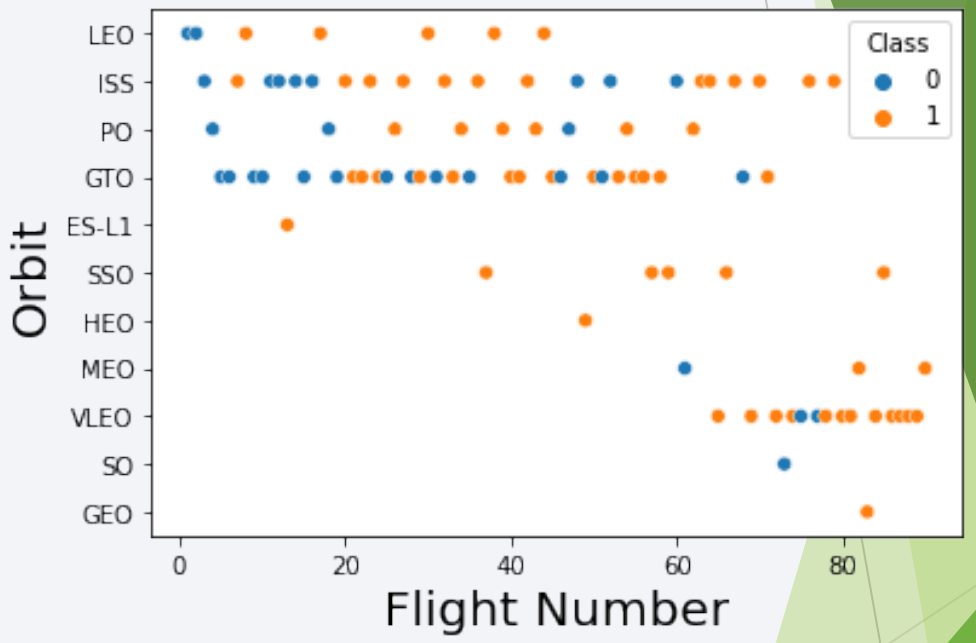
Success Rate vs. Orbit Type

- 4 Orbit types(ES-L1, GEO, HEO, SSO) have a 100% success rate



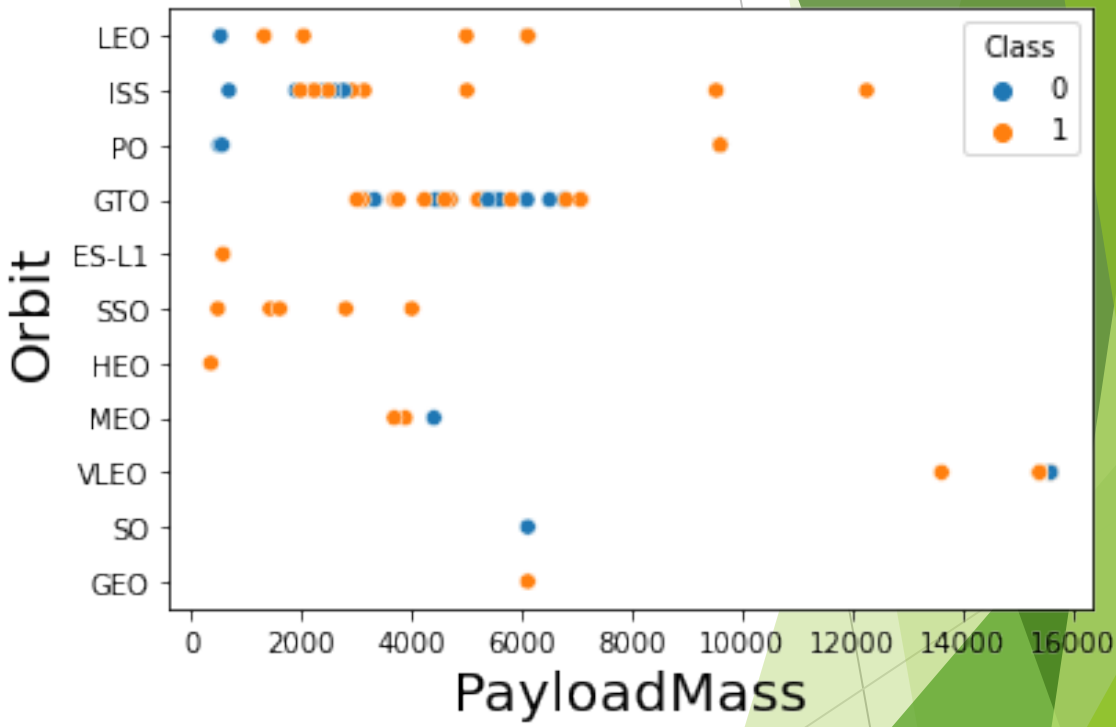
Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type
- Show the screenshot of the scatter plot with explanations



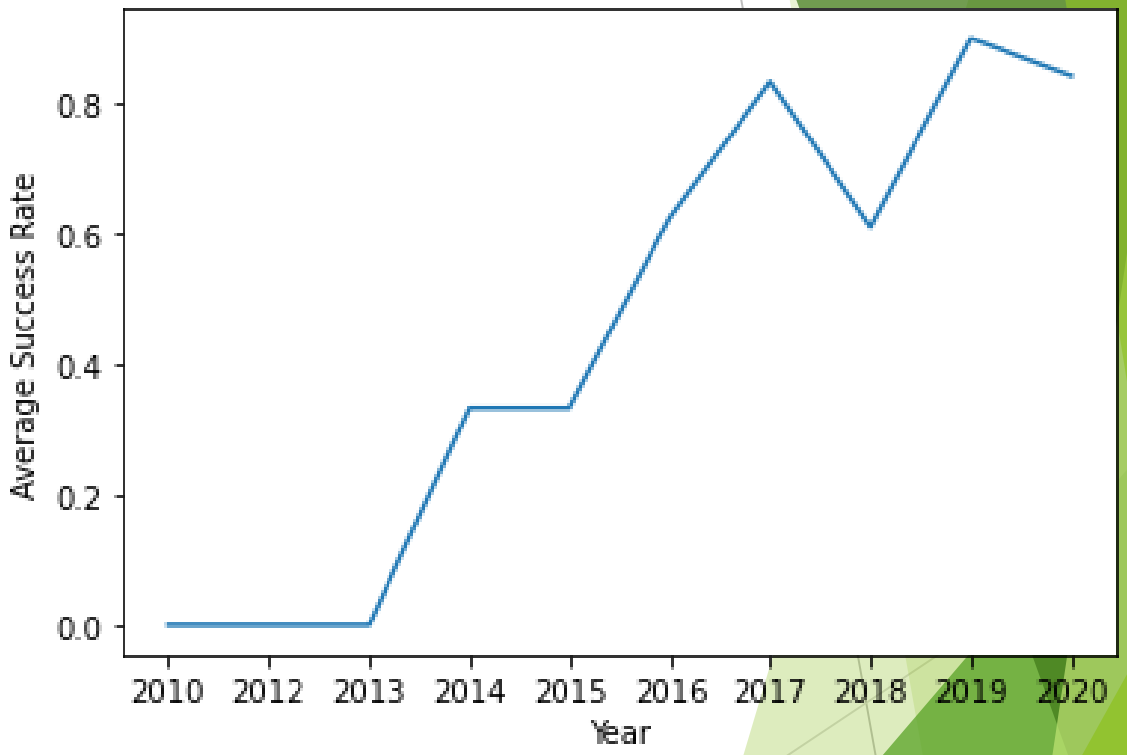
Payload vs. Orbit Type

- Highest concentration of Payload Mass and Orbit can be seen for ISS between 2000 to 4000 payload mass and for GTO 3000 to 7000



Launch Success Yearly Trend

- Average success rate increases year over year despite a slight dip in 2018



All Launch Site Names

- Find the names of the unique launch sites:
- Query: select distinct launch_site from SPACEXTBL
- Explanation: using select distinct to get unique values for the column launch_site

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`

- Query: select * from SPACEXTBL where launch_site like 'CCA%' limit 5

- Explanation: querying uses where statement to filter launch sites that start with CCA and returns only 5 records using limit clause

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_payload_nasa_crs
45596

- Query: select sum(payload_mass__kg_) as total_payload_NASA_CRS from SPACEXTBL where customer = 'NASA (CRS)'
- Explanation: query aggregates the total of the payload column where customer is 'NASA (CRS)'

Average Payload Mass by F9 v1.1

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

avg_payload
2534

- Query: select avg(payload_mass__kg_) as avg_payload from SPACEXTBL where booster_version like 'F9 v1.1%'

- Explanation: aggregating the average of payload for booster version F9 v1.1

First Successful Ground Landing Date

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

first_successful_landing_date
2015-12-22

- Query: select min(date) as first_successful_landing_date from SPACEXTBL where landing__outcome = 'Success (ground pad)'

- Explanation: using the min function the date of the first successful landing on the ground pad is foung

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

- Query: select booster_version from SPACEXTBL where landing__outcome ='Success (drone ship)' and payload_mass__kg_ between 4000 and 6000

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

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

- Query: select mission_outcome, count(DATE) as total_success from SPACEXTBL group by mission_outcome
- Explanation: aggregating counts by mission outcome

Boosters Carried Maximum Payload

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

- Query: select booster_version, payload_mass__kg_ from SPACEXTBL where payload_mass__kg_ in (select max(payload_mass__kg_) from SPACEXTBL)
- Explanation: using subquery to find booster version that carried max payload mass

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

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

- Query: select booster_version, launch_site from SPACEXTBL where landing__outcome = 'Failure (drone ship)' and year(DATE) = 2015
- Explanation: selecting booster version and launch sites filtered on date in 2015 and outcome failure in drone ship

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

Query:

```
select landing__outcome, count(date) as count_

from SPACEXTBL

where date between '2010-06-04' and '2017-03-20'

group by landing__outcome
order by count(date) desc
```

Section 4

Launch Sites Proximities Analysis

<Folium Map Screenshot 1>

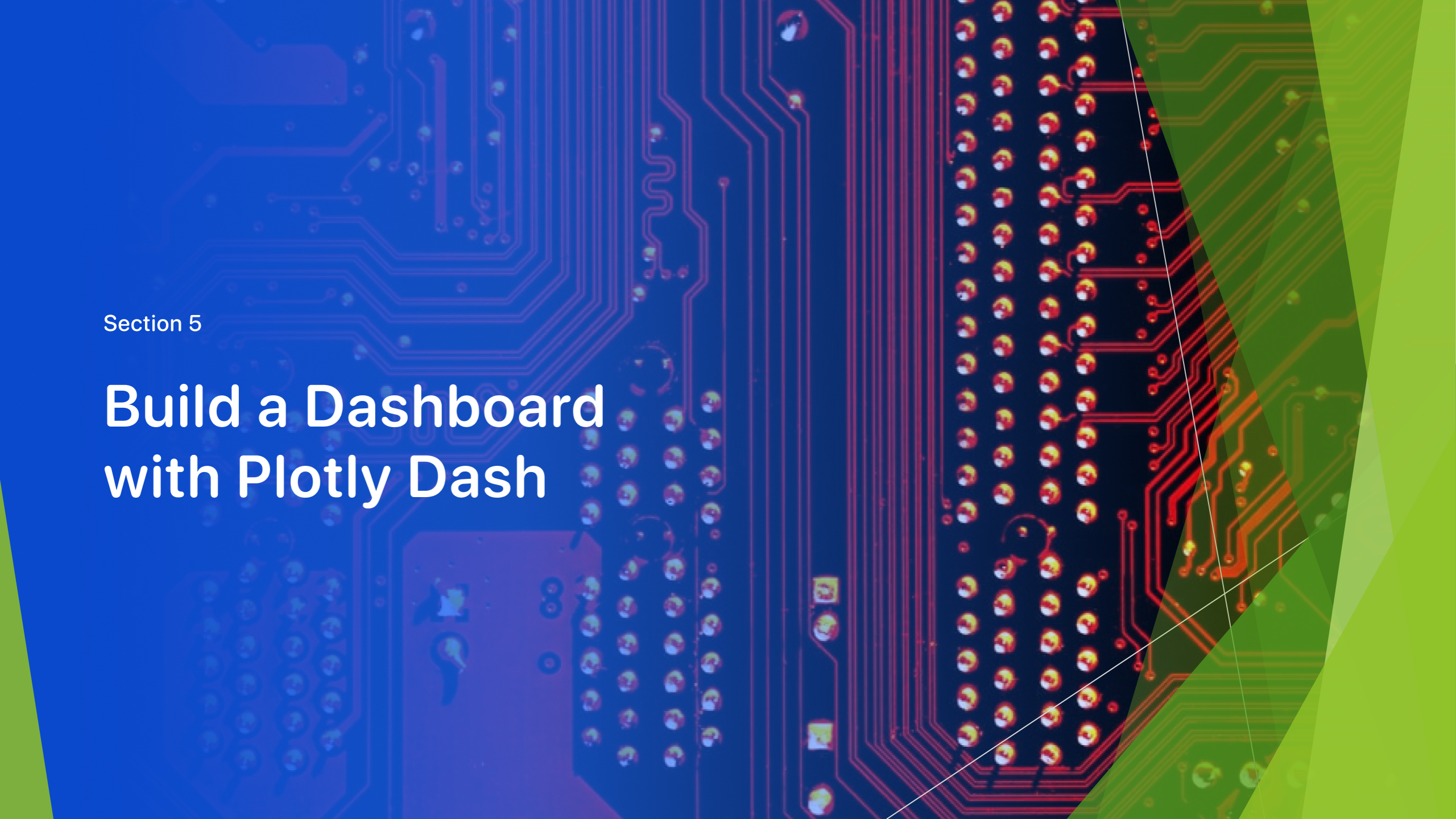
- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 2>

- Replace <Folium map screenshot 2> title with an appropriate title
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed
- Explain the important elements and findings on the screenshot


Section 5
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 6

Predictive Analysis (Classification)

Classification Accuracy

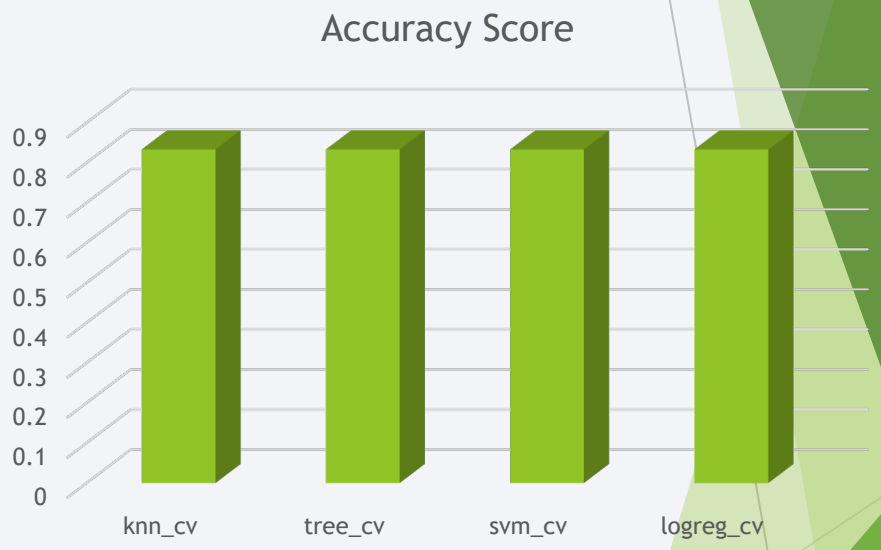
- Visualize the built model accuracy for all built classification models, in a bar chart
- All m… …racy scor…

```
In [29]: knn_cv.score(X_test, Y_test)
Out[29]: 0.8333333333333334
```

```
In [24]: tree_cv.score(X_test, Y_test)
Out[24]: 0.8333333333333334
```

```
In [19]: svm_cv.score(X_test, Y_test)
Out[19]: 0.8333333333333334
```

```
In [14]: logreg_cv.score(X_test, Y_test)
Out[14]: 0.8333333333333334
```



Confusion Matrix

Conclusions

- Any classification model can be used to predict the outcomes of this dataset
- Success rate of launches increases with time meaning that the more Space X continues to launch these rockets the more successful they will be
- The orbits ES-L1, GEO, HEO, SSO have a 100% success rate

Appendix

Thank you!