IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Shreya Tanguturi
2022-10-10

Outline

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

Executive Summary

- We have used Data Collection, Exploratory Data Analysis and Machine Learning Prediction
- Using Exploratory Data Analysis, we were able to find the success rate of the launches, and using Machine Learning prediction we were able to predict the best model to be used for analyzing data.

Introduction

- Using the data from SpaceX, we wanted to predict the success rate of a competing company, SpaceY
- How successful are the landings of the Falcon 9 rocket launches.
- What are the safest launch sites

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected from https://api.spacexdata.com/v4/rockets/, https://api.spacexdata.com/v4/launchpads/, https://api.spacexdata.com/v4/payloads/
- Perform data wrangling
  - Data was filtered so that we only worked with Falcon 9 launches
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data was normalized and split into train and test sets, after which we employed 4 different classification models (KNN, SVM, Logistic Regression and Decision Tree), and the accuracy of each model was found.

Data Collection

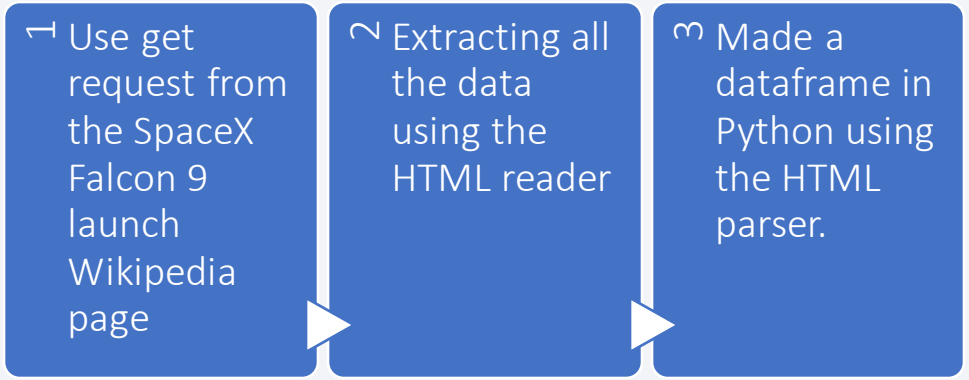
- Data sets were collected from the SpaceX API website (rockets, launchpads, and payloads), and from the SpaceX Wikipedia page.
  - https://api.spacexdata.com/v4/rockets/
  - https://api.spacexdata.com/v4/launchpads/
  - https://api.spacexdata.com/v4/payloads/
  - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Data Collection – SpaceX API

- The SpaceX website has API that is open to everyone for use. The API was requested and parsed through HTML.
- https://github.com/Shreyatt26/appliedddatasciencecapstone/blob/44df782fda618b70f1edc9f43889489436426cb8/SpaceX%20Data%20Collection.ipynb

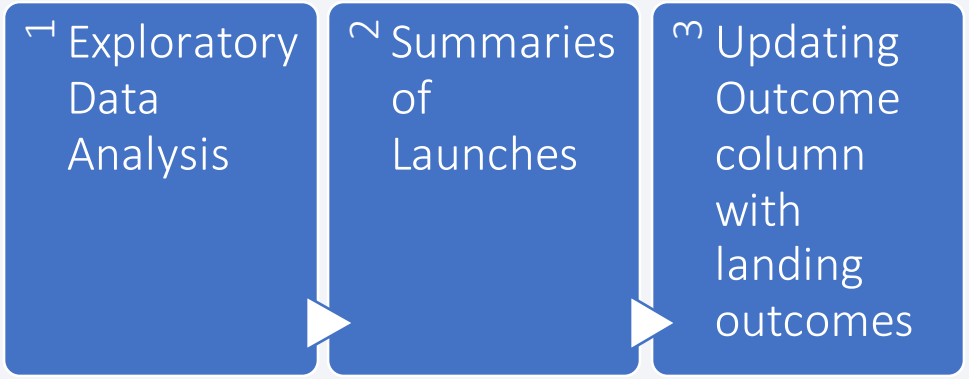
Data Collection - Scraping

- The SpaceX launch data is also available on their Wikipedia page, from which Webscraping techniques were used to obtain the data.
- https://github.com/Shreyatt26/applieddatasciencecapstone/blob/44df782fda618b70f1edc9f4388948943642 6cb8/SpaceX%20Webscraping.ipynb



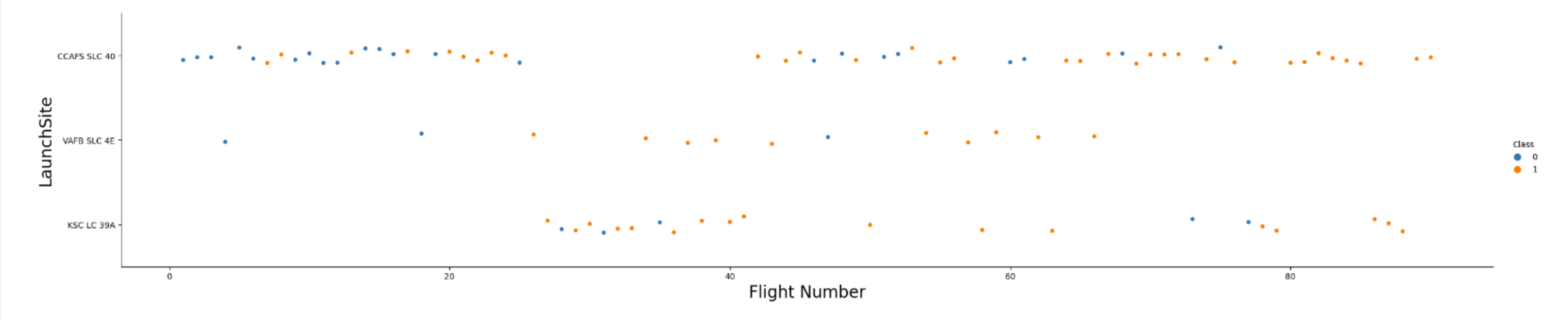
Data Wrangling

- The data was processed with Exploratory Data Analysis, where we looked at the launch summaries for each launch site and a landing outcome was calculated. The Outcome column was updated with the landing outcomes.
- https://github.com/Shreyatt26/applieddatasciencecapstone/blob/44df782fda618b70f1edc9f4388948943642 6cb8/SpaceX%20Data%20Wrangling.ipynb



EDA with Data Visualization

- The charts that were plotted were the Flight Number vs Launch Site, Payload vs Launch Site, Flight Number vs Orbit Type, and Payload vs Orbit Type.
- https://github.com/Shreyatt26/applieddatasciencecapstone/blob/44df782fda618b70f1edc9f4388948943 6426cb8/SpaceX%20EDA%20Visualization.ipynb



EDA with SQL

- SQL Queries that were performed
  - Selected the names of the unique launch sites
  - Displayed only records from launch sites that began with 'CCA'
  - Displayed the total payload mass carried by boosters launched by NASA (CRS)
  - Displayed average payload mass carried by booster version F9 v1.1
  - Listed the date when the first successful landing outcome in ground pad was achieved.
- https://github.com/Shreyatt26/applieddatasciencecapstone/blob/44df782fda618b70f1edc9f4388948943642 6cb8/SpaceX%20EDA%20with%20SQL.ipynb

EDA with SQL

- SQL Queries that were performed
  - Listed 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 records which will display the month names, failure landing_outcomes in drone ship, booster versions, launch_site for the months in year 2015.
  - Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.
- https://github.com/Shreyatt26/applieddatasciencecapstone/blob/44df782fda618b70f1edc9f4388948943642 6cb8/SpaceX%20EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

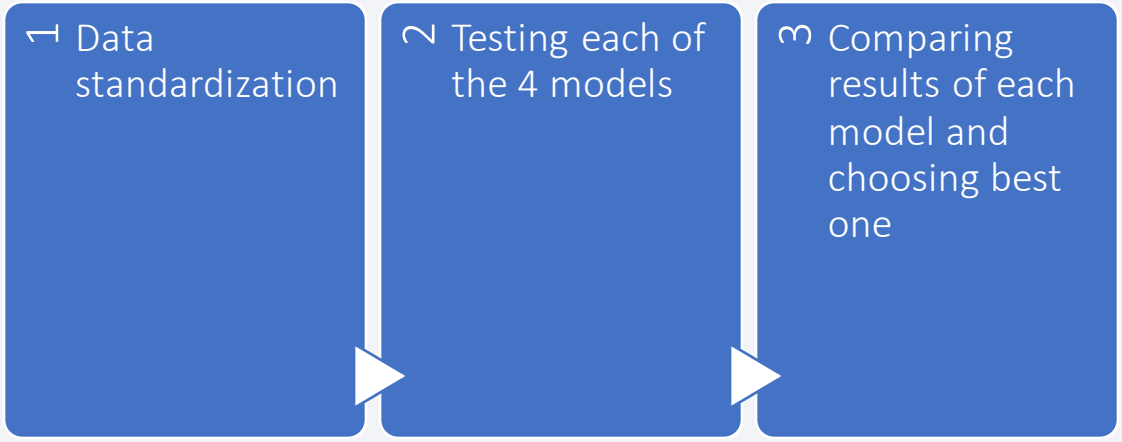
- A Folium Map object was created with an initial map marker at the NASA Johnson Space Center in Houston, Texas.
- Using the center map marker, we created a folium Circle object to highlight the area of the initial location
- All four launch and landing sites were marked with a map marker
- Used lines to show distances between a launch site and its proximities
- https://github.com/Shreyatt26/applieddatasciencecapstone/blob/44df782fda618b70f1edc9f4388948943642 6cb8/SpaceX%20Folium%20Launch%20Sites.ipynb

Build a Dashboard with Plotly Dash

- The payload range and percentage of launches by site were used to visualize the data in pie charts
- Doing so, we were able to find the relation between the payload and launch site, which corresponds to the success of each launch site. This way, we can determine which launch site is the best to launch future rockets at.
- https://github.com/Shreyatt26/applieddatasciencecapstone/blob/9ffa826edc8886935802572f8c782463605652f5/space_dash_app.py

Predictive Analysis (Classification)

- We used four different types of classification models
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K-Nearest Neighbours
- https://github.com/Shreyatt26/applieddatasciencecapstone/blob/9ffa826edc8886935802572f8c782463605652f5/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

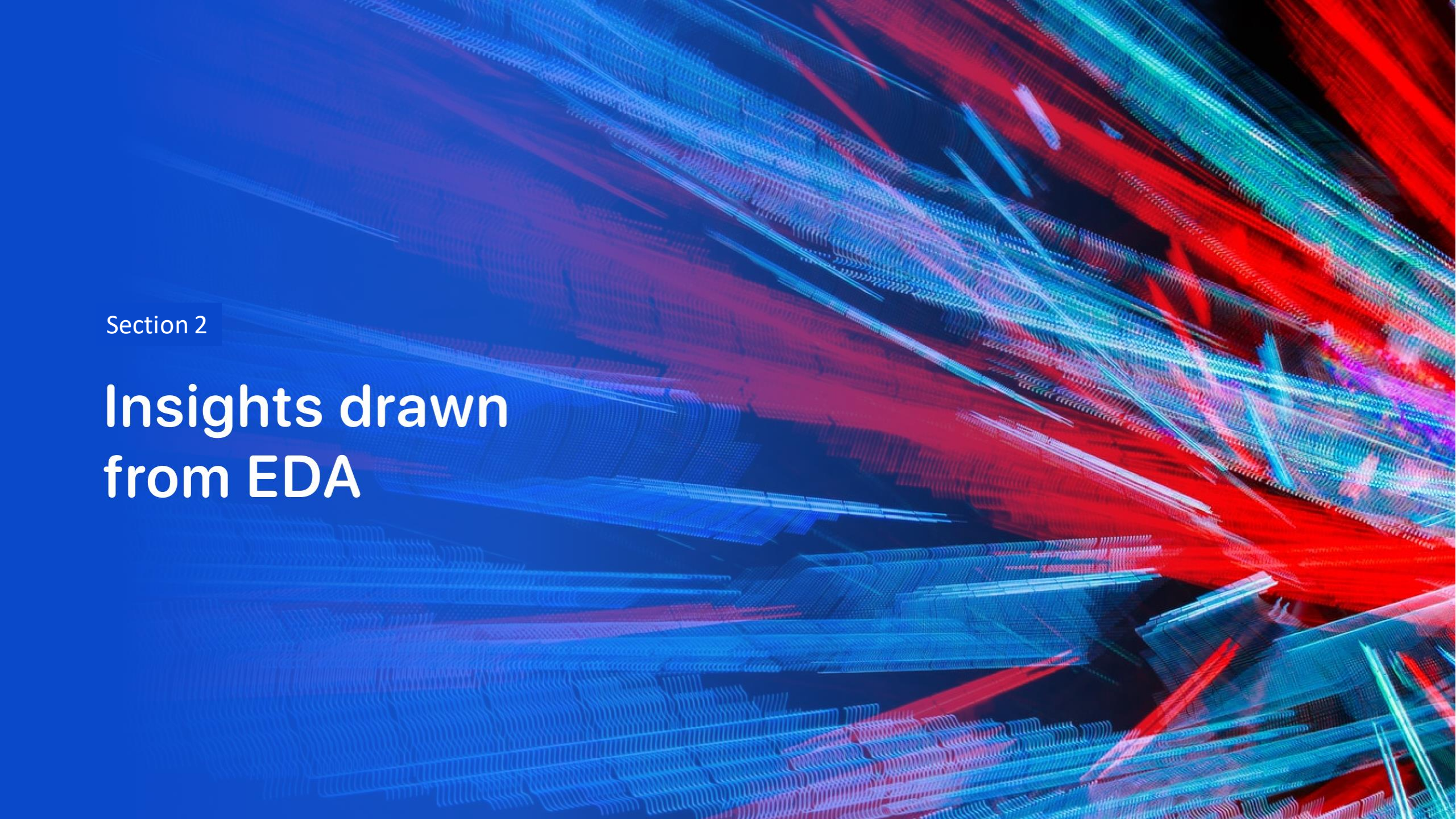


Results

- Exploratory data analysis results
  - Space X has 4 launch sites
  - Falcon 9 booster versions are very successful
  - Landing Outcomes improve as the years pass
- Interactive analytics demo in screenshots
  - Using Folium maps and map markers, we were able to identify the safest Launch Sites due to their proximity to highways, major roads and cities etc.
- Predictive analysis results
  - We found that the Decision Tree classifier model was the most accurate with 88% accuracy.

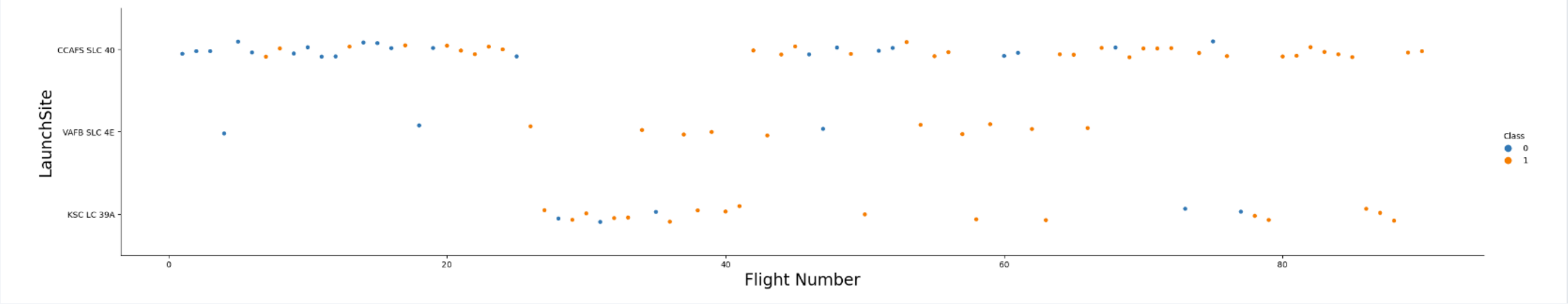
Section 2

Insights drawn from EDA



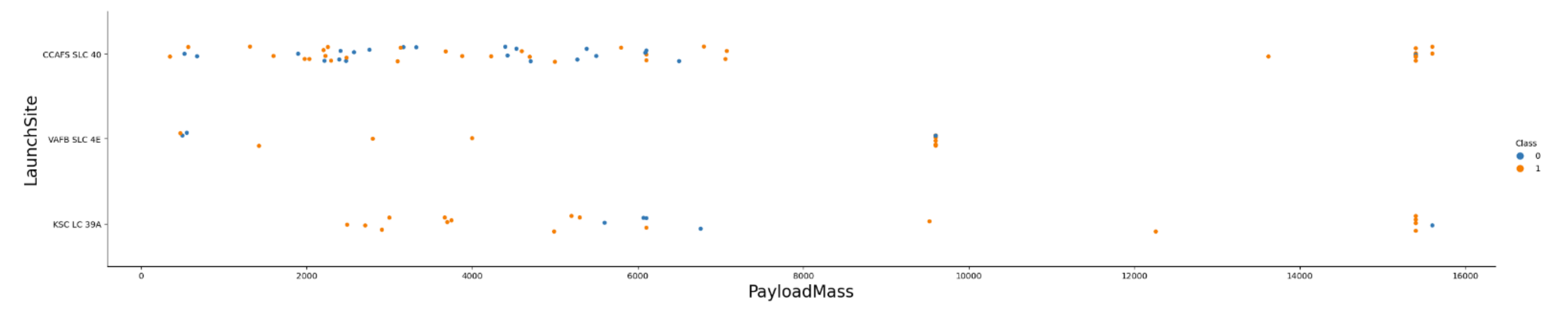
Flight Number vs. Launch Site

- Scatter-plot of Flight Number vs. Launch Site



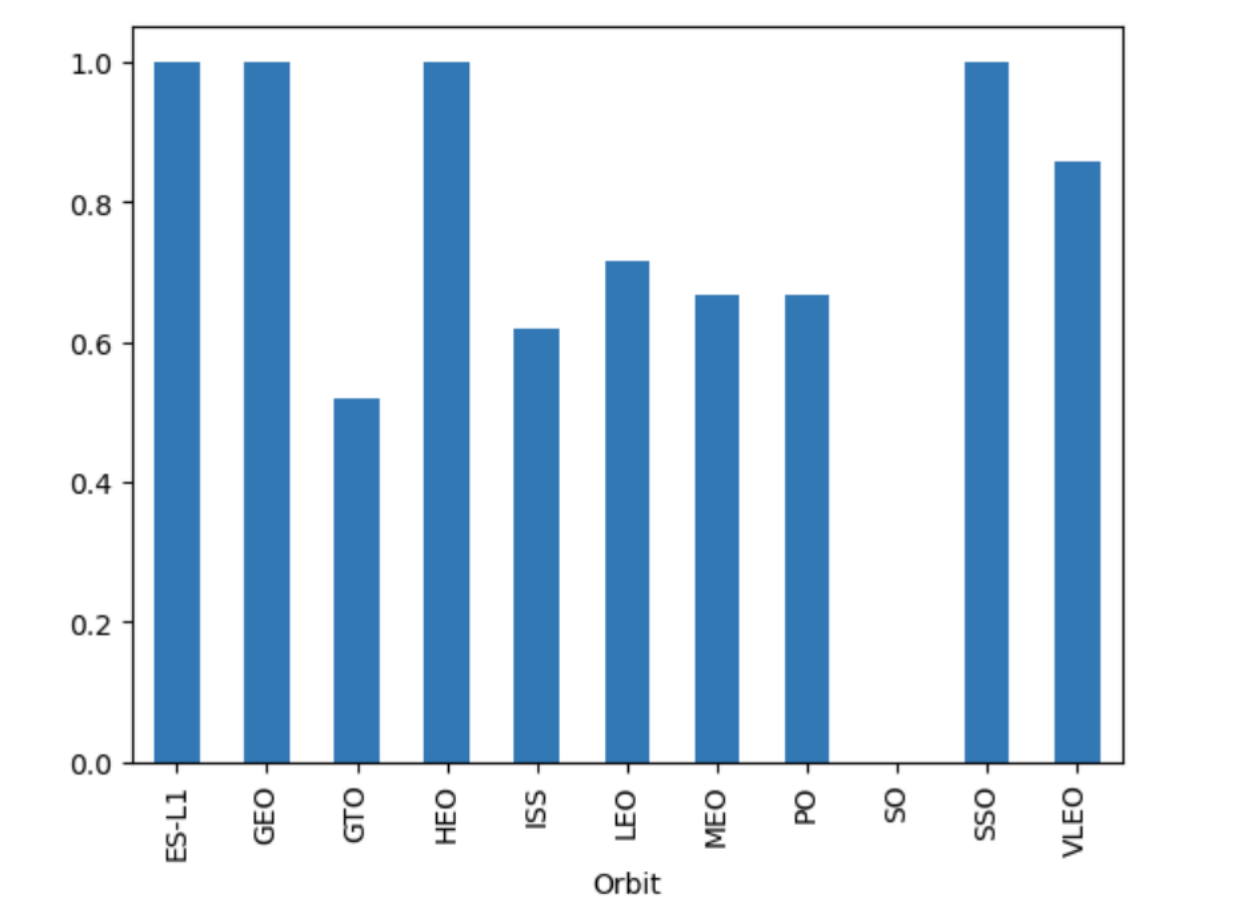
Payload vs. Launch Site

- Scatter-plot of Payload vs. Launch Site



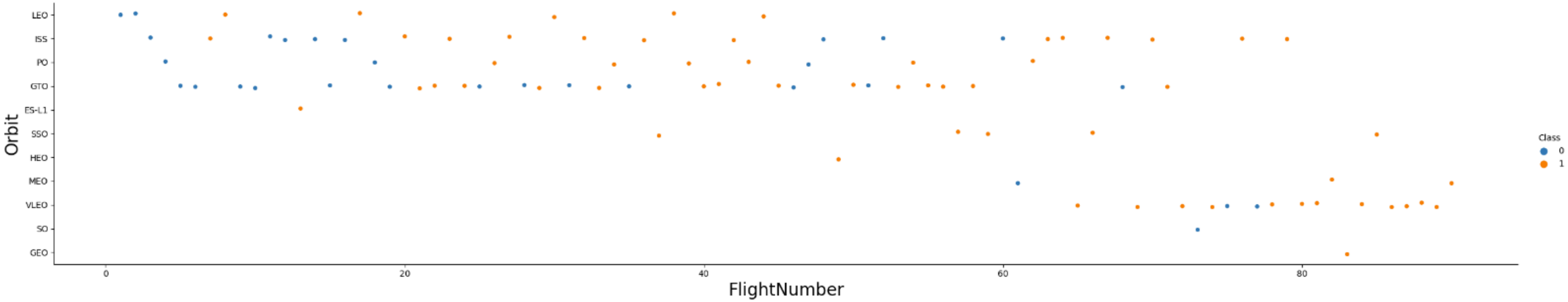
Success Rate vs. Orbit Type

- Orbits ES-L1, GEO, HEO, and SSO had the highest success rates.



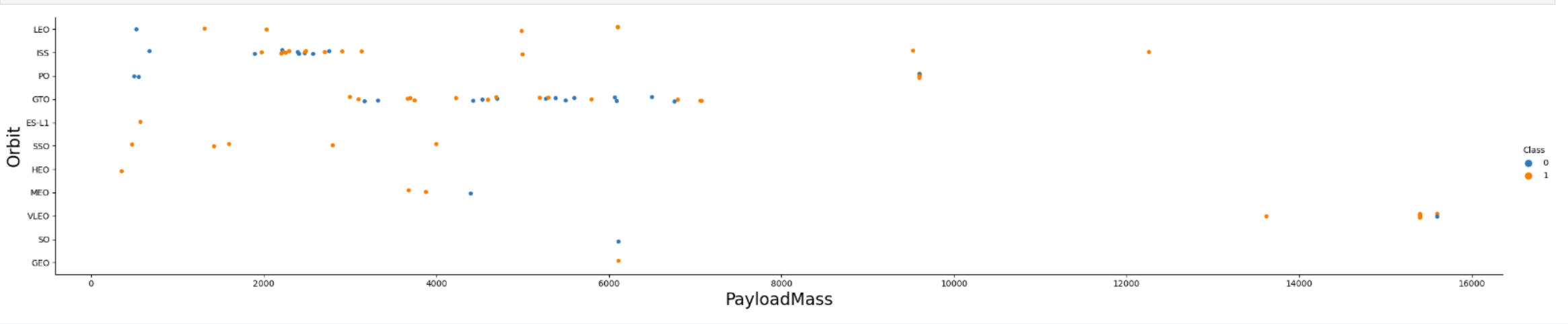
Flight Number vs. Orbit Type

- Scatter-plot of Flight number vs. Orbit type



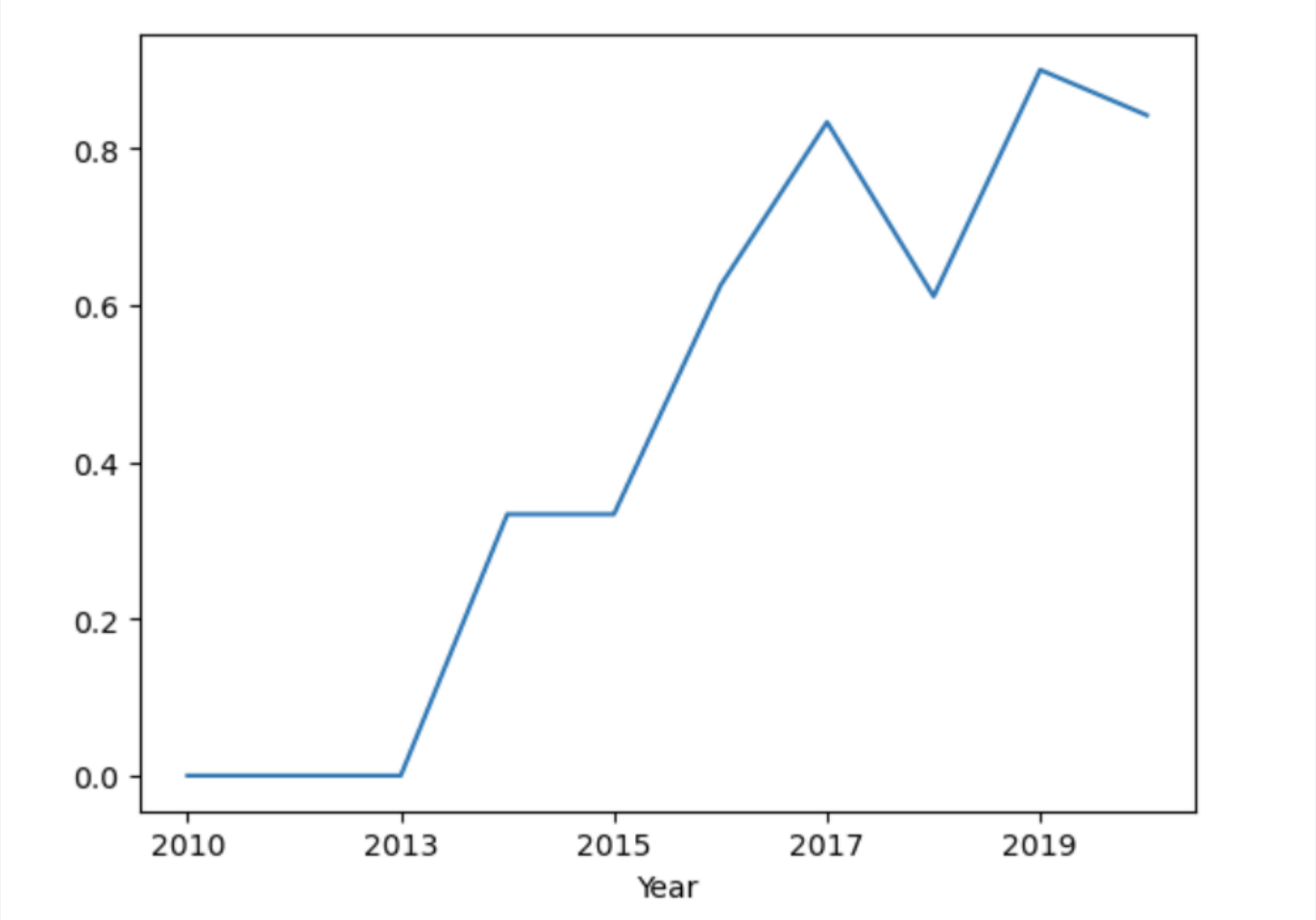
Payload vs. Orbit Type

- Scatter-plot of payload vs. orbit type



Launch Success Yearly Trend

- The Success Rate tends to increase linearly, with one drop at around 2018, after which is continues to increase



All Launch Site Names

- The four unique launch sites are:
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

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

Launch Site Names Begin with 'CCA'

- 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
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Total payload carried by boosters from NASA

TOTAL_PAYLOAD
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1

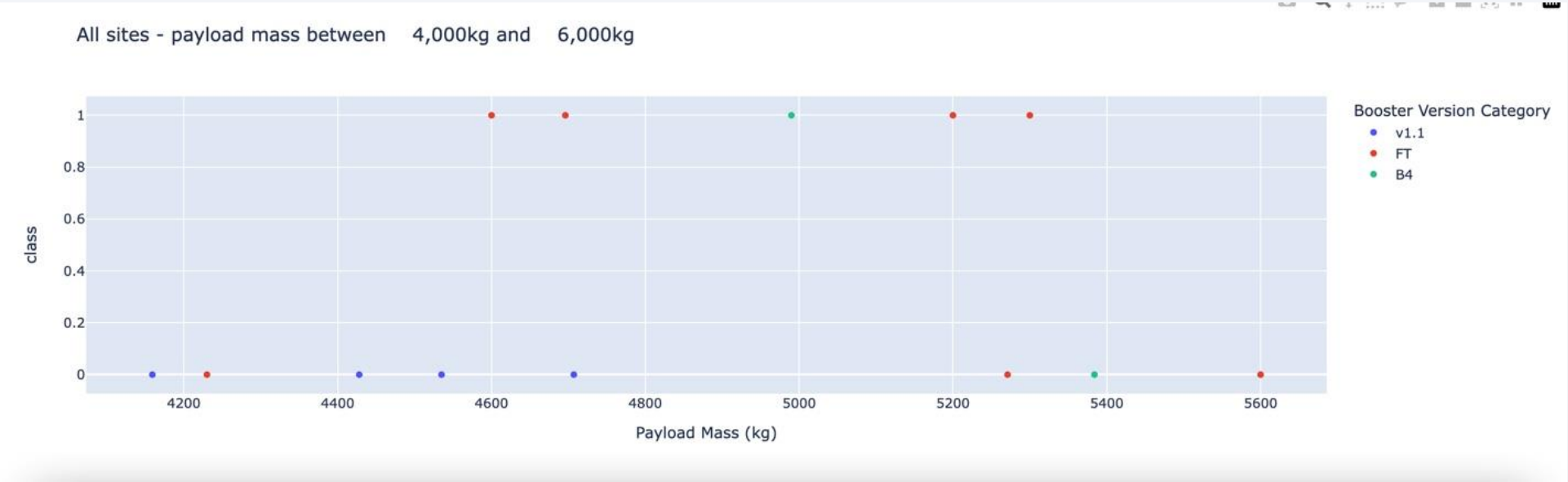
AVERAGE
2534.6666666666665

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad
  - January 3rd, 2013

Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
  - V1.1
  - FT
  - B4



Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes
  - Total success – 99
  - Total failure - 1

Mission_Outcome	COUNT
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- These are the boosters that carried the maximum payload

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

- Failed landing_outcomes in drone ship, their booster versions, and launch site names for the year 2015
- This shows that in January and April of 2015, there were failed landing outcomes.

MONTH	Landing _Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

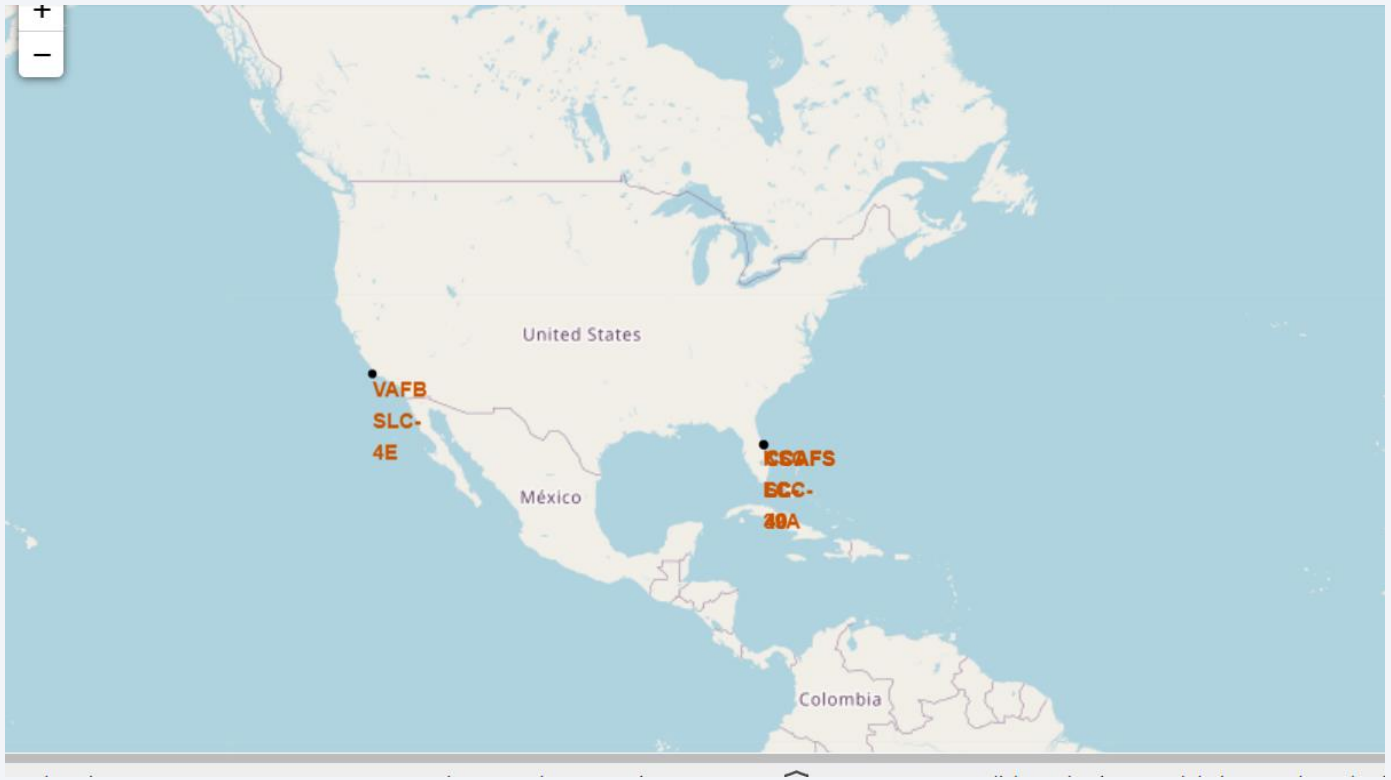
- Ranking of landing outcomes between the date 2010-06-04 and 2017-03-20 in descending order.

Landing _Outcome	QTY
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1


Section 3
Launch Sites
Proximities Analysis

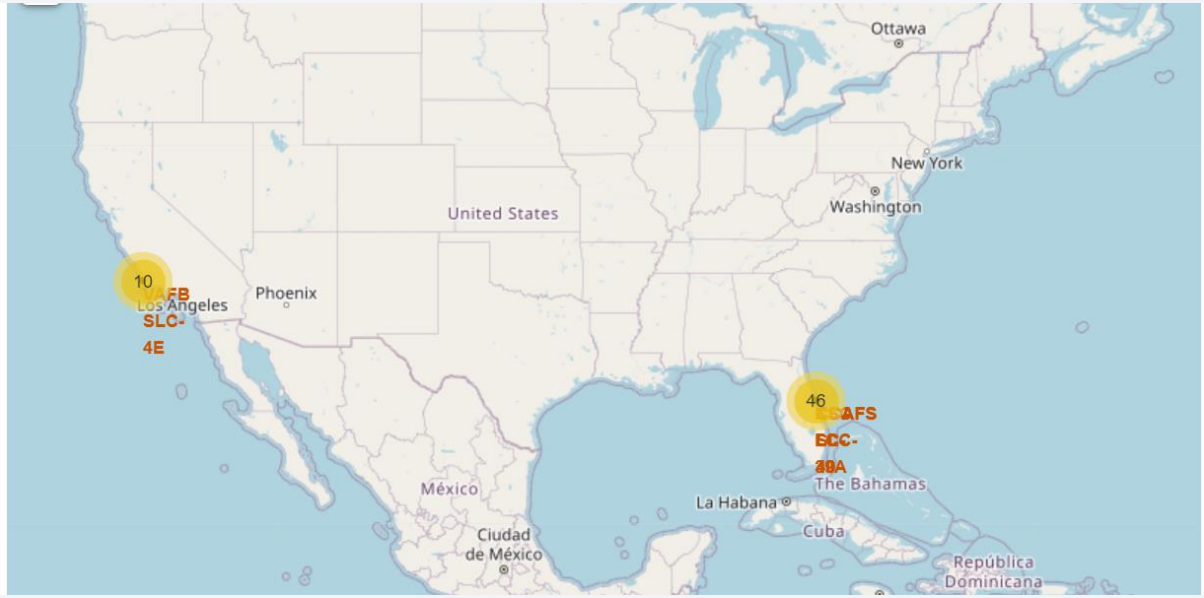
Launch Site Locations

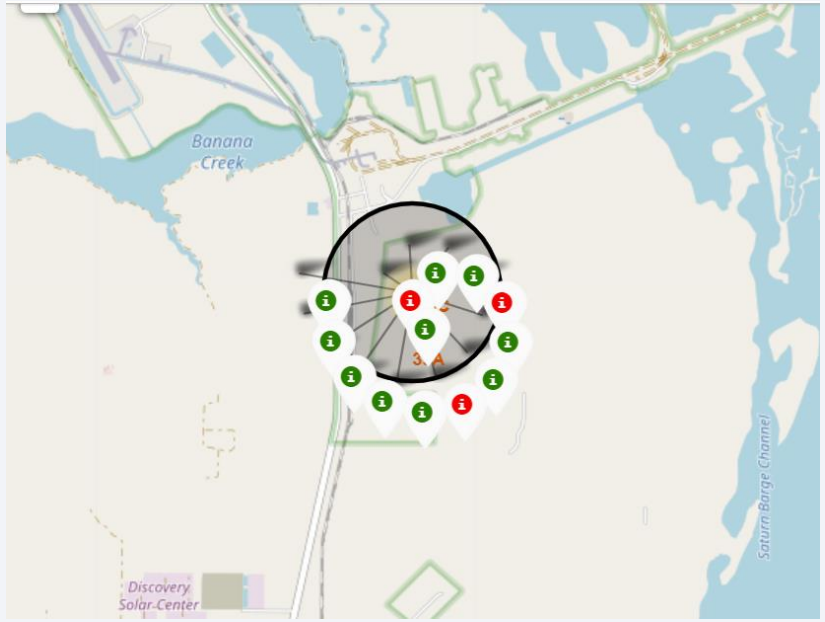
- The Folium Map shows the launch sites, VAFS SLC-4E, CCAFS LC-40 and CCAFS SLC-40



Launch Outcomes

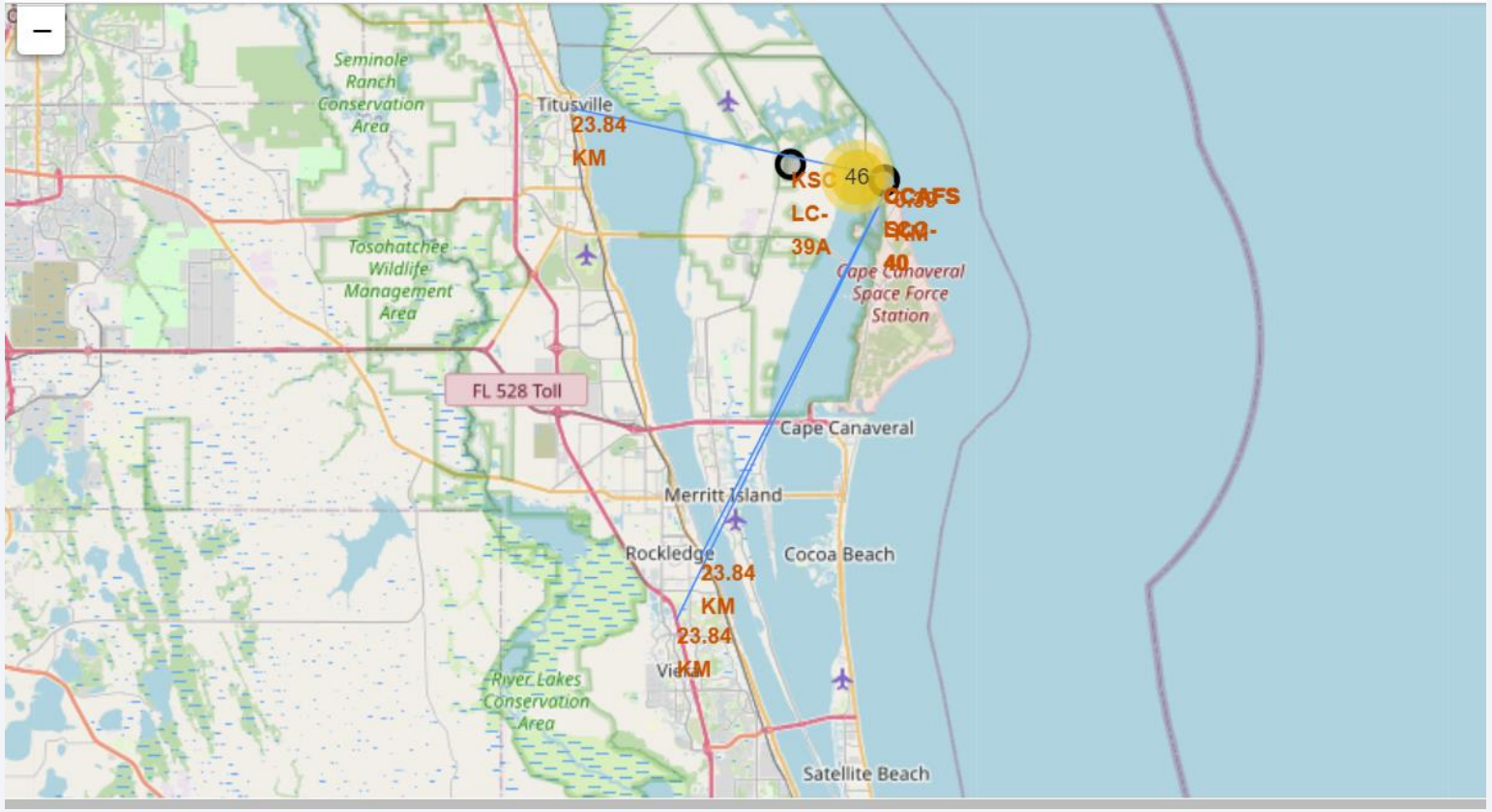
- KSC LC-39A is the most successful launch site

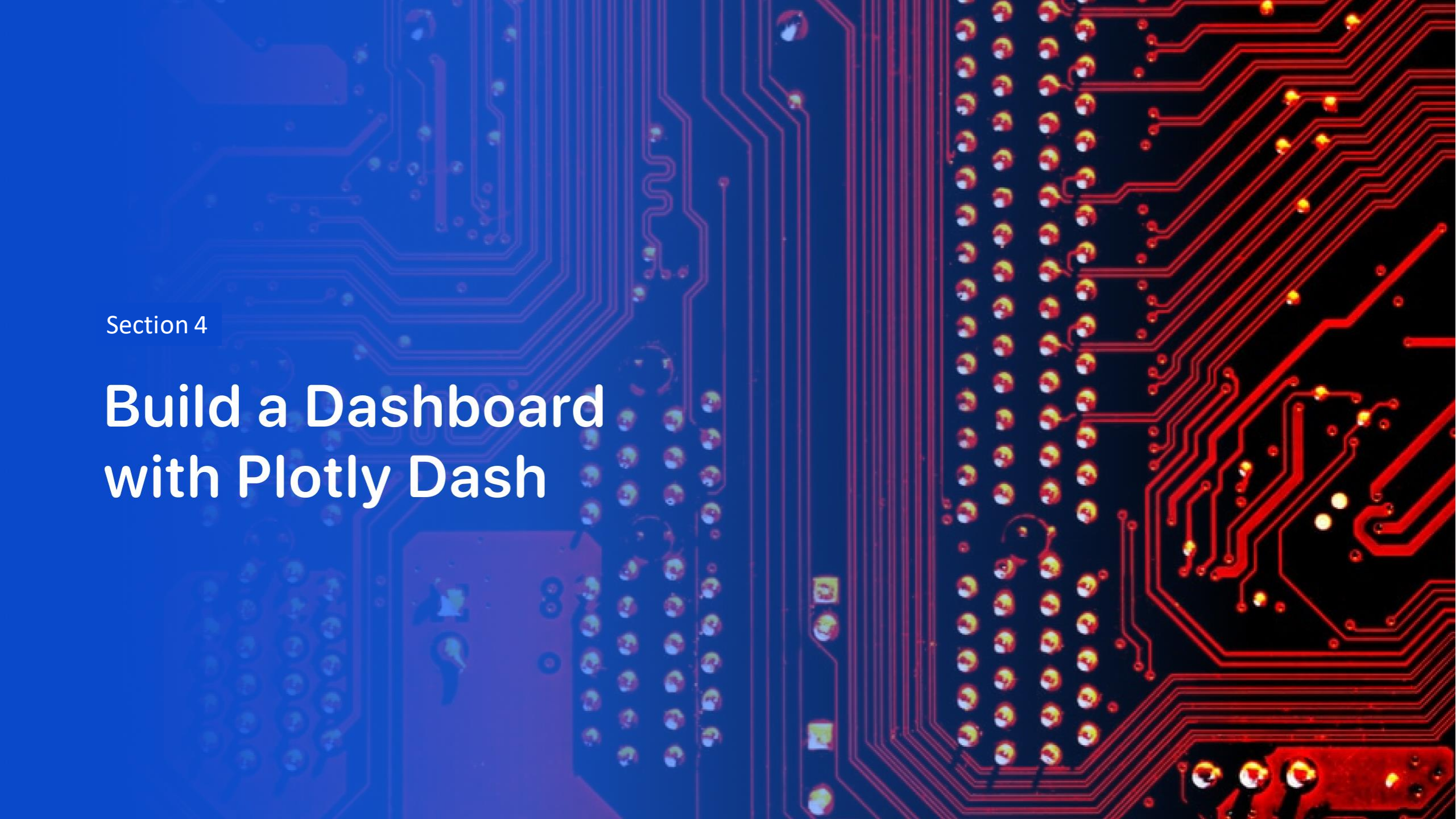




Launch Site Safety

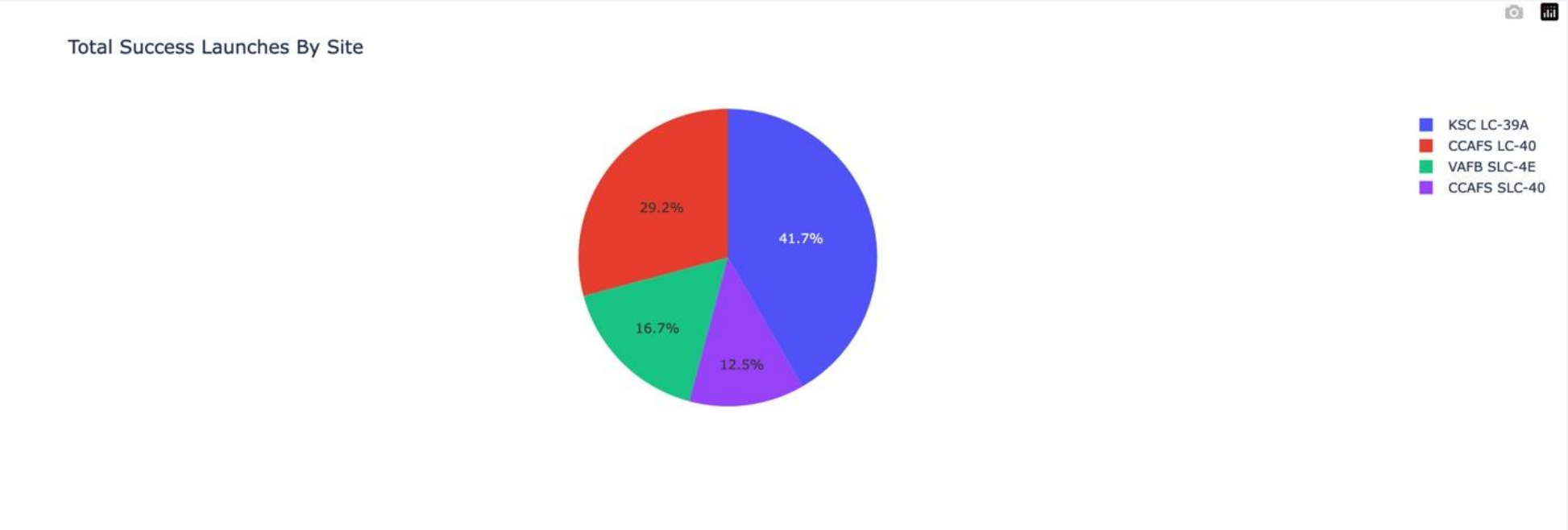
- CCAFS SLC-40 is not the safest site as it is not in close proximity to Railroads, Cities and Highways.




Section 4
Build a Dashboard with Plotly Dash

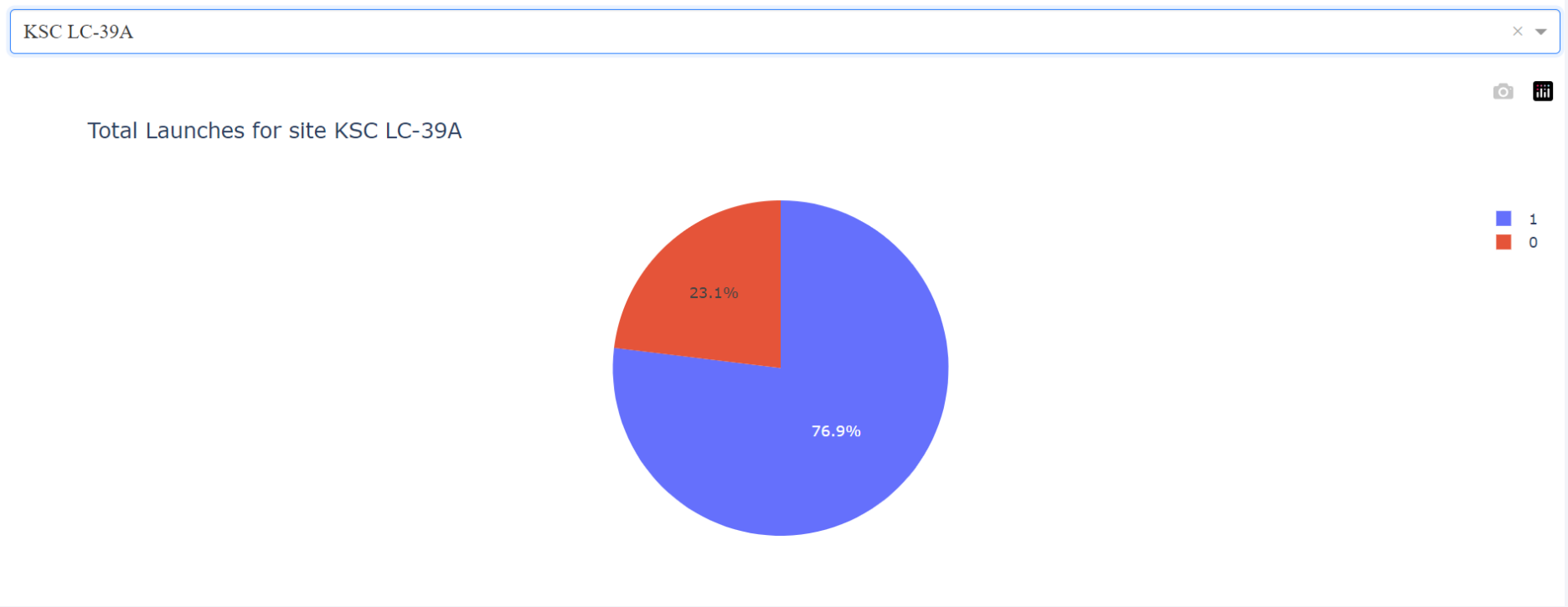
Launch Success Count

- The Launch Success Pie Chart shows that KSC LC-39A is the most successful launch site, and CCAFS SLC-40 is the least successful launch site.

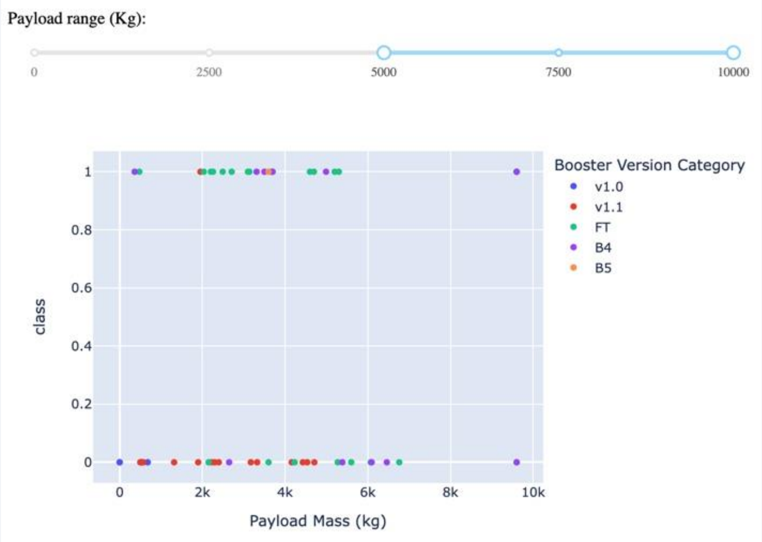


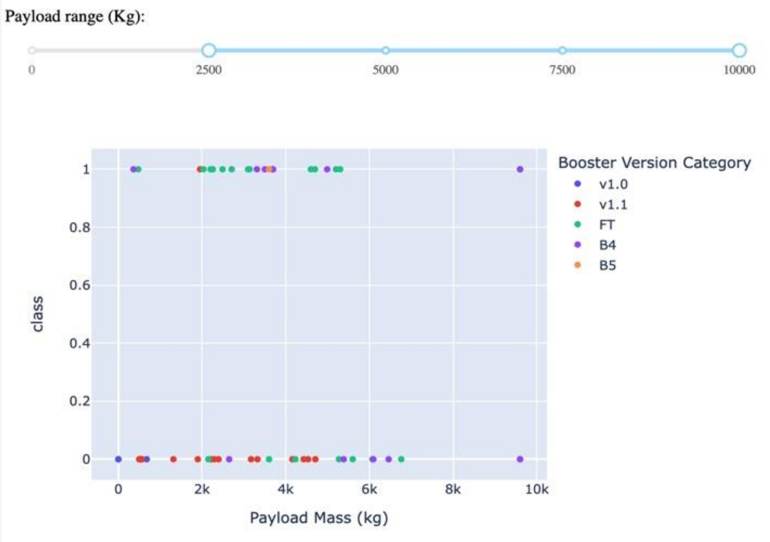
Highest Launch Success Ratio

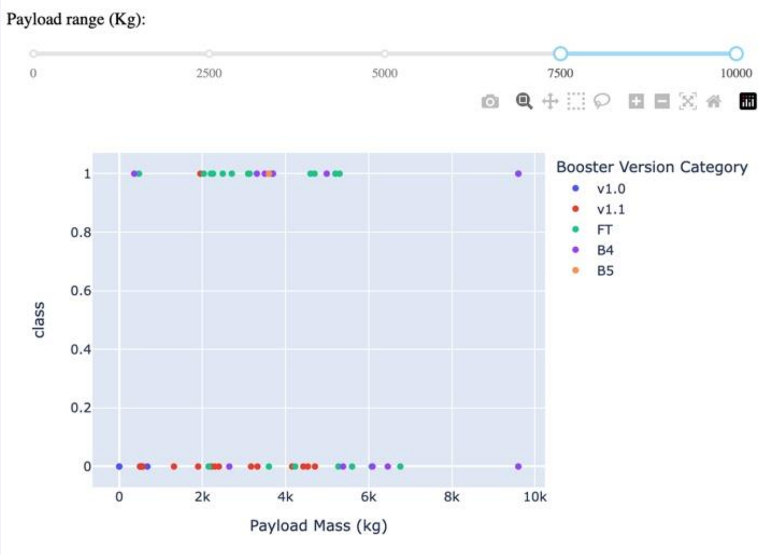
- The highest launch success ratio is for launch site KSC LC-39A
  - 76.9% of the launches are successful

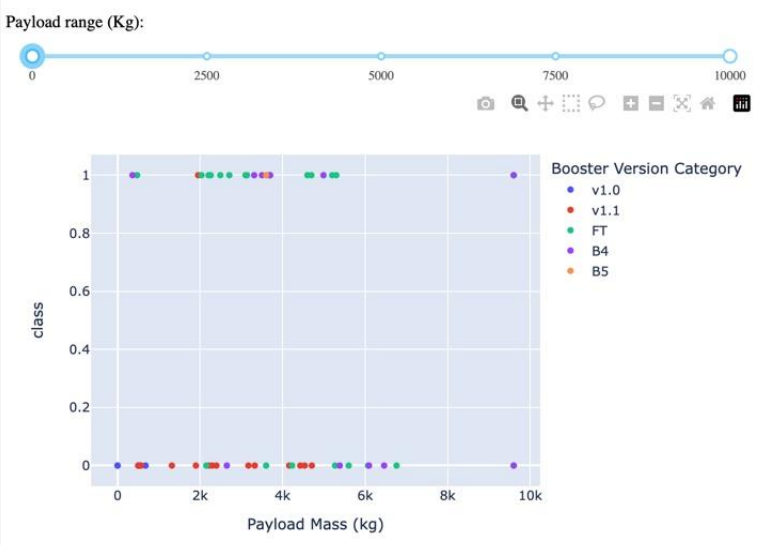


Payload vs Launch







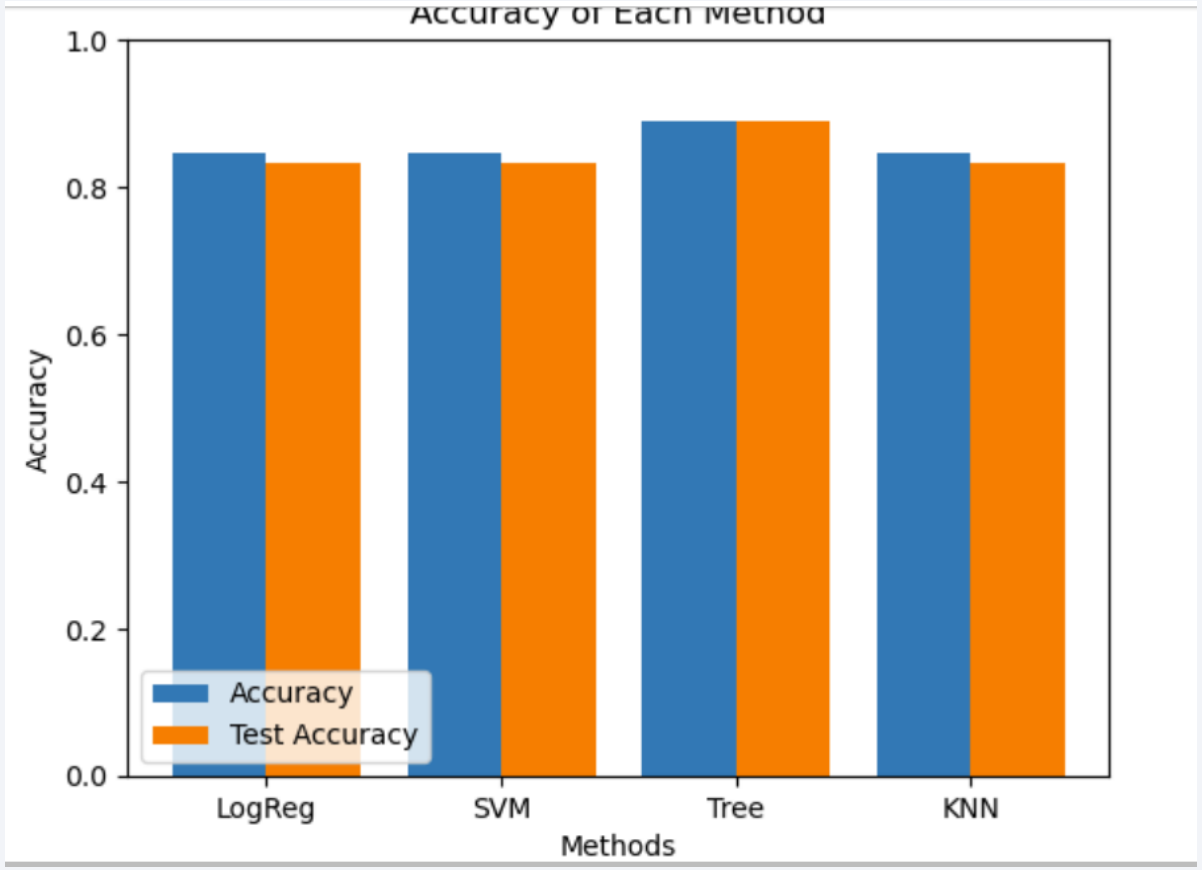


Section 5

Predictive Analysis (Classification)

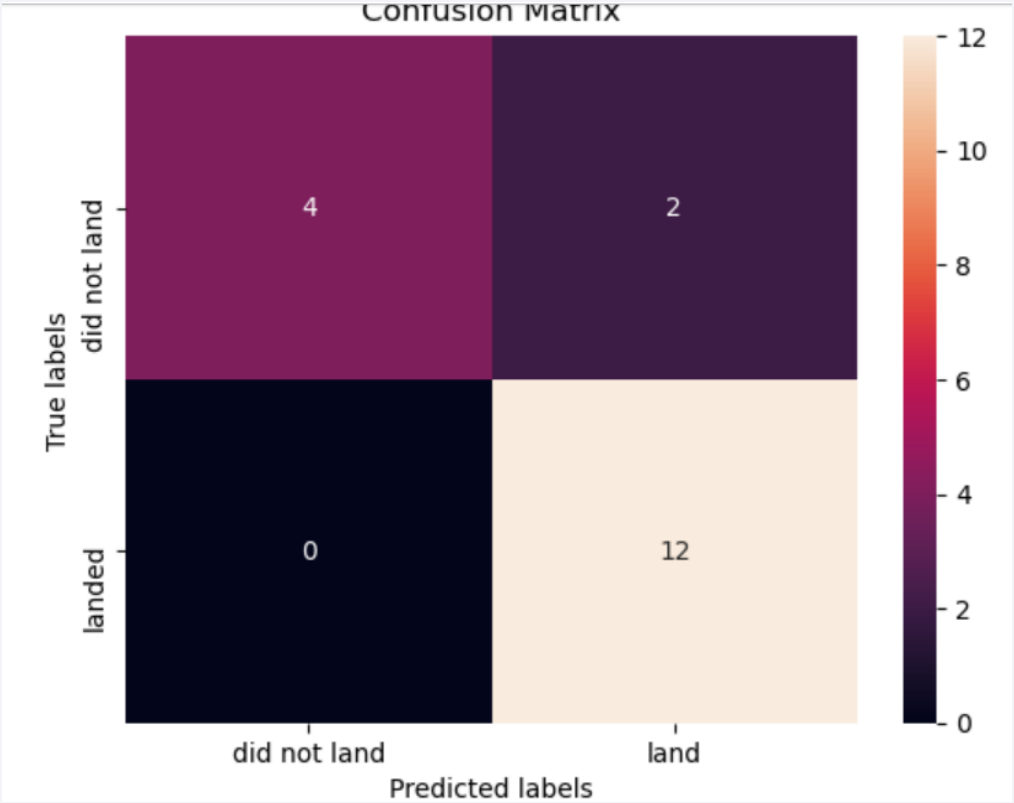
Classification Accuracy

- Looking at the bar chart, it is evident that the Decision Tree has the highest accuracy since the Test value and Actual values are close. In reference to the Machine Learning Lab, we also found that the Tree Model has an 88% accuracy.



Confusion Matrix

- The Decision Tree Model was the most accurate, with 88% accuracy.



Conclusions

- KSC LC-39A is the most successful launch site
- Decision Tree Classifier is the most accurate model with 88% accuracy

Appendix

- All labs were done in Jupyter Notebook
  - https://github.com/Shreyatt26/applieddatasciencecapstone

Thank you!