IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Eemil Haapsaari
2025-02-13

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

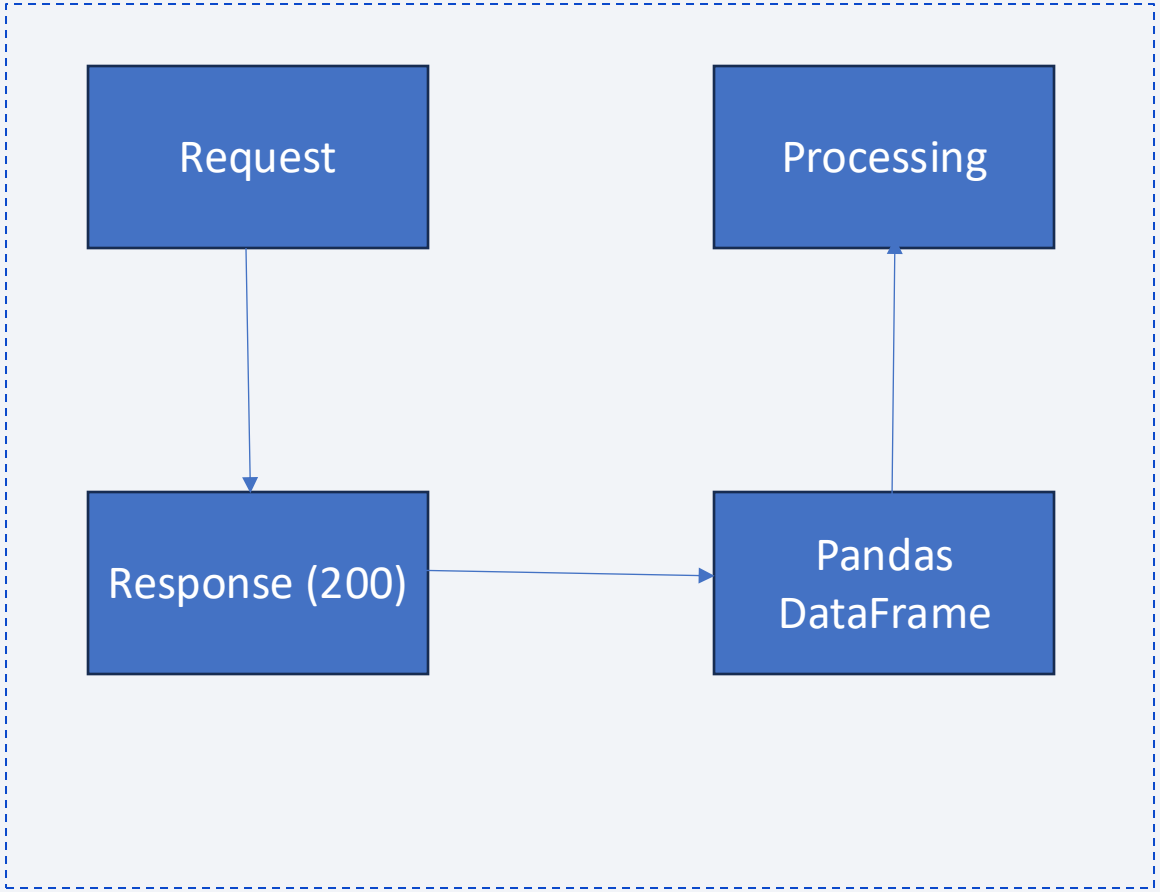
- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- SpaceX data was collected using
  - SpaceX rest api
  - Wikipedia (Web scraping)

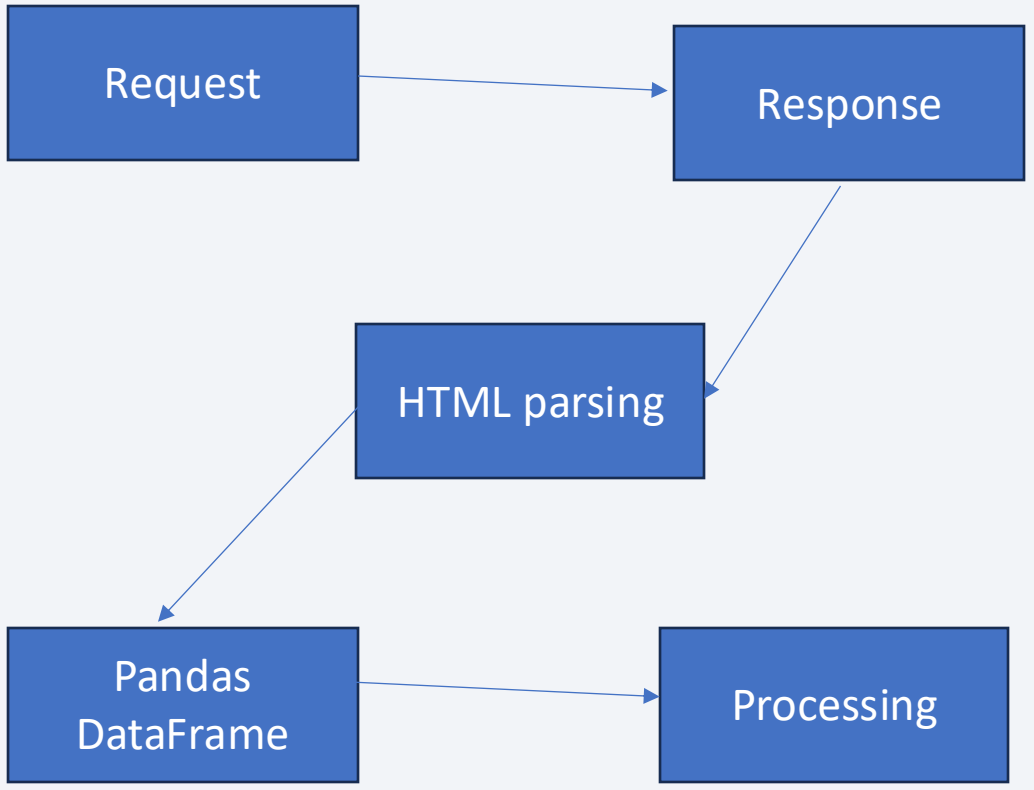
Data Collection – SpaceX API

- Data was requested from SPaceX RestAPI
- SpaceX rest api response was then loaded into pandas dataframe
- Pandas DataFrame was then analyzed and processed for the future usage
- https://github.com/haapsee/SpaceY/blob/main/JupyterLabs/jupyter-labs-spacex-data-collection-api-v2.ipynb



Data Collection - Scraping

- Web scraping data was requested from the Wikipedia
- Wikipedia page received as a response was then parsed using BeautifuloSoup
- Parsed data from HTML was loaded into pandas DataFrame and saved for the future use.
- https://github.com/haapsee/SpaceY/blob/main/JupyterLabs/jupyter-labs-webscraping.ipynb



Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts
- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

EDA with Data Visualization

- Scatter Point Plot was used for visualizing for how 2 different features affected Success
- Bar Chart was used to visualize success rate for different orbits
- Line chart was used to visualize how success rate has changed over the years.
- https://github.com/haapsee/SpaceY/blob/main/JupyterLabs/jupyter-labs-eda-dataviz-v2.ipynb

EDA with SQL

- SQL queries was used to find:
  - Unique values in columns (filtered and not-filtered)
  - Sum of column
  - Average of column
  - Minimum values
  - Counts
- https://github.com/haapsee/SpaceY/blob/main/JupyterLabs/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

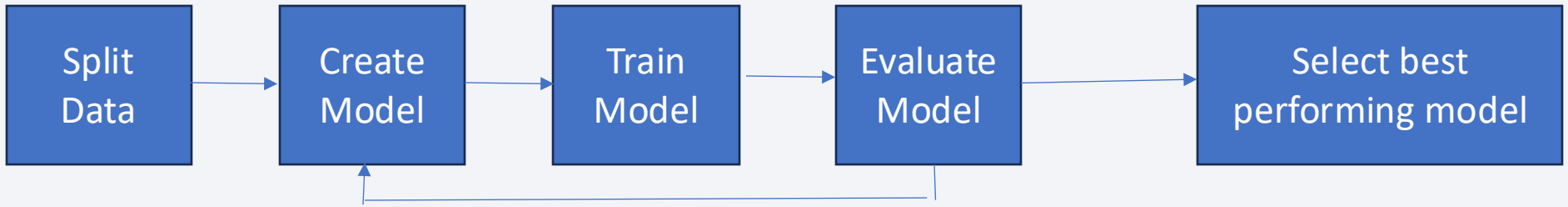
- Folium map object used are:
  - Markers for marking launches, grouped by Marker CLuster
  - Circles for marking launch site
  - Popup for showing launch site name
  - PolyLine for distance between objects
- https://github.com/haapsee/SpaceY/blob/main/JupyterLabs/lab-jupyter-launch-site-location-v2.ipynb

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Used Plots/Graphs are:
  - Pie Chart for success/failure rate visualization
  - Scatter plot to show the correlation between payload, booster version and launch success
- Graphs can be interacted with
  - Dropdown to select launch site
  - Slider to adjust payload (Scatter plot only)
- https://github.com/haapsee/SpaceY/blob/main/spacex_dash_app.py

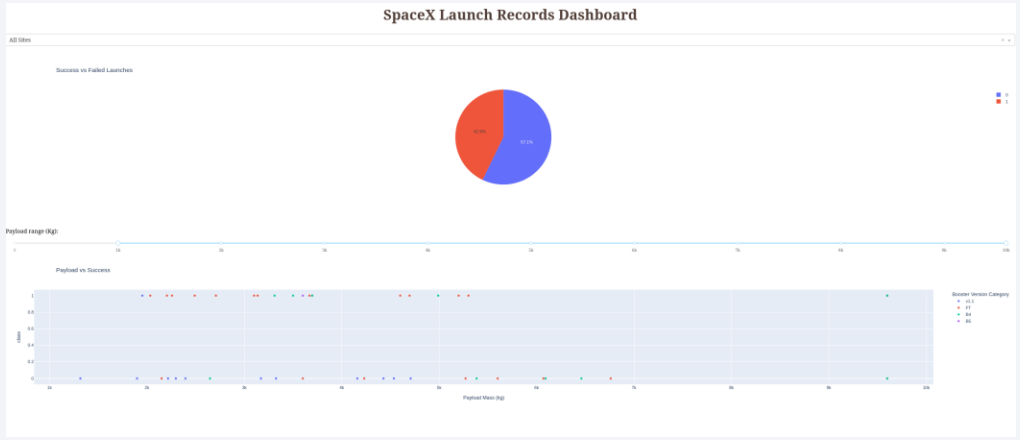
Predictive Analysis (Classification)

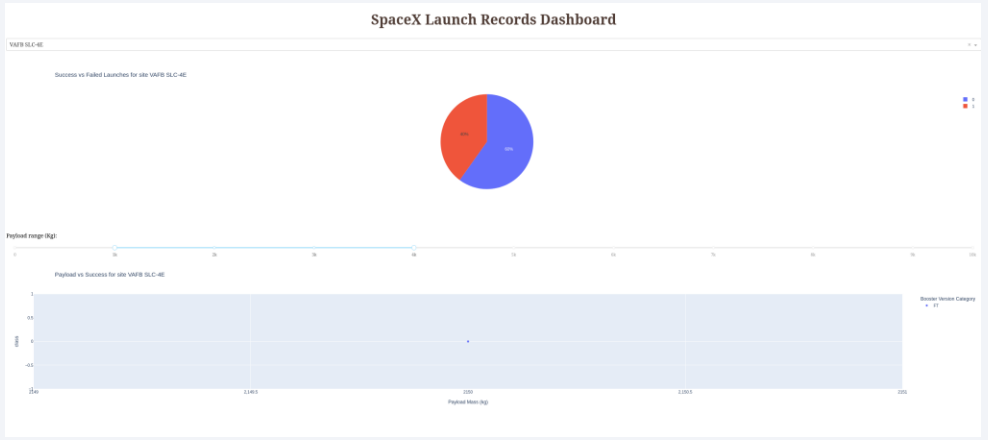
- Model was built and trained using Scikit Learn
- Data was split into test and training data
- Model was evaluated using test data
- Best performing model was found using confusion matrix, score function and best_score_ parameter
- https://github.com/haapsee/SpaceY/blob/main/JupyterLabs/SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb

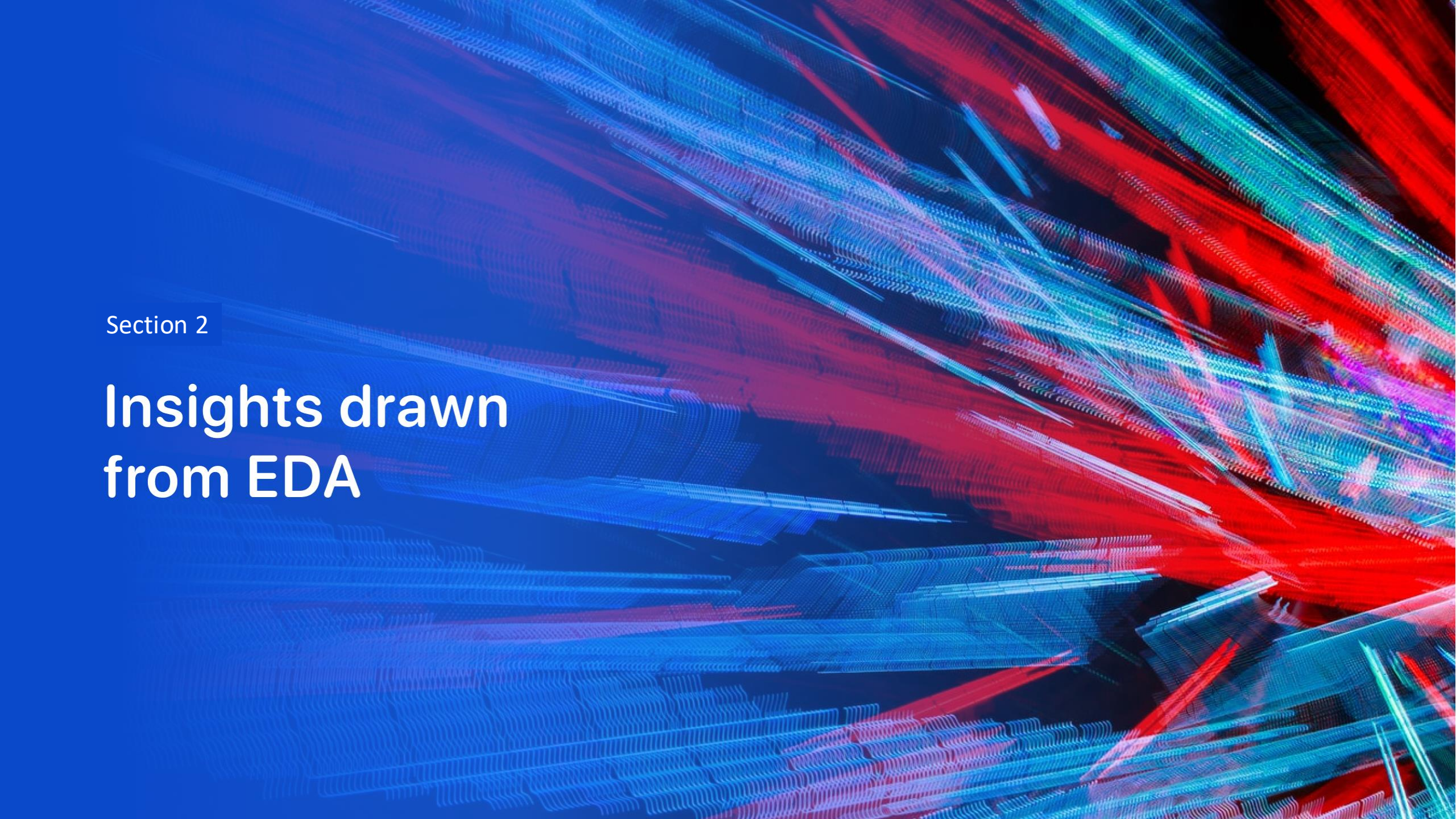


Results

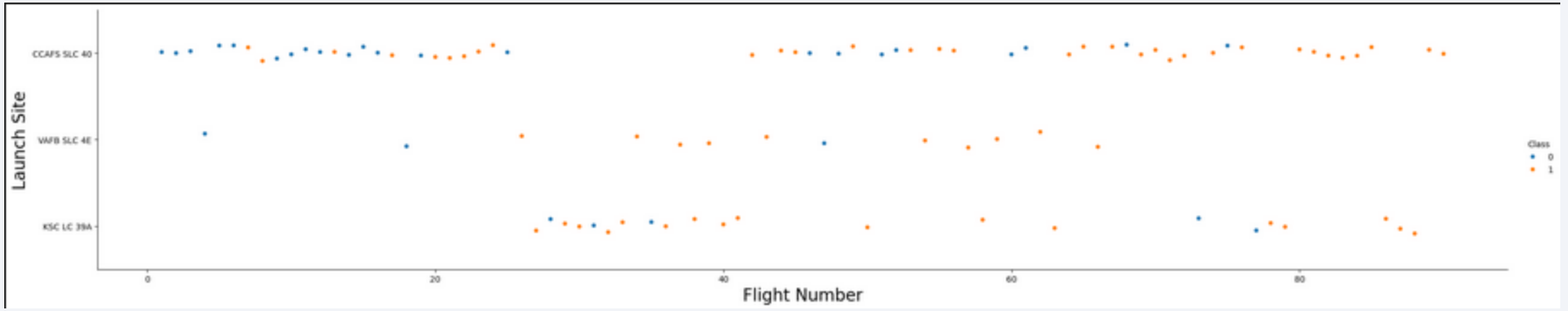
- Exploratory data analysis results tells that success rate has improved over the years
- Predictive analysis results tells that DecisionTree would be the best fit for analysing coming up success





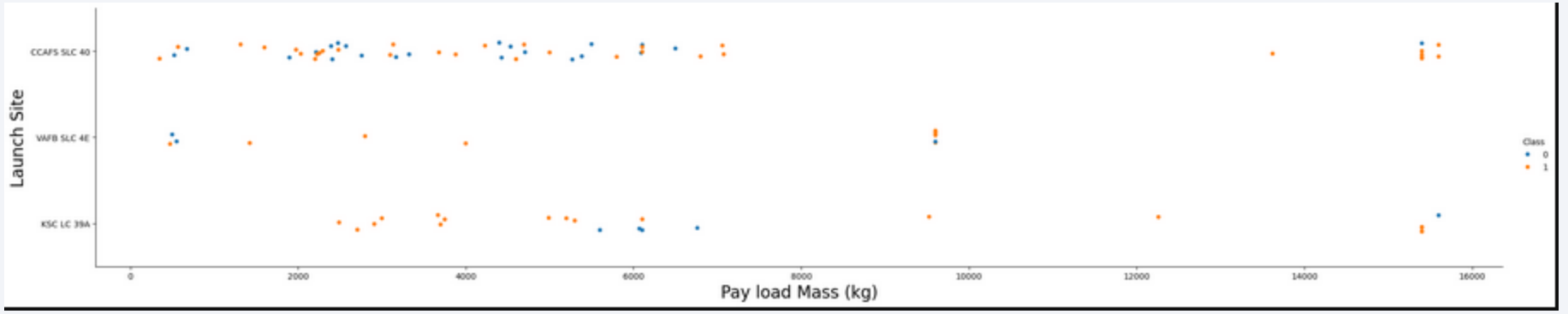

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- Scatter plot of Success rate between flight number and launch site
- Launch site on y-axis
- Flight number on x-axis
- Blue = Failure
- Orange = Success

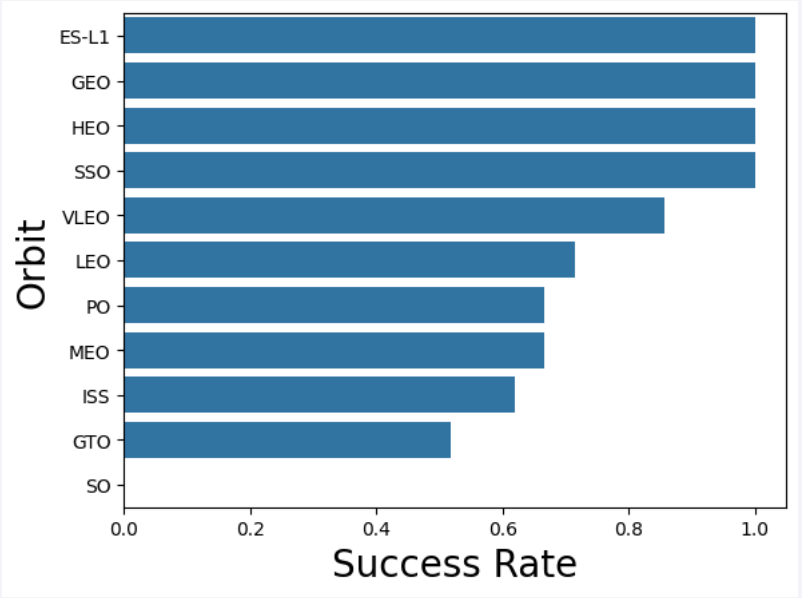
Payload vs. Launch Site



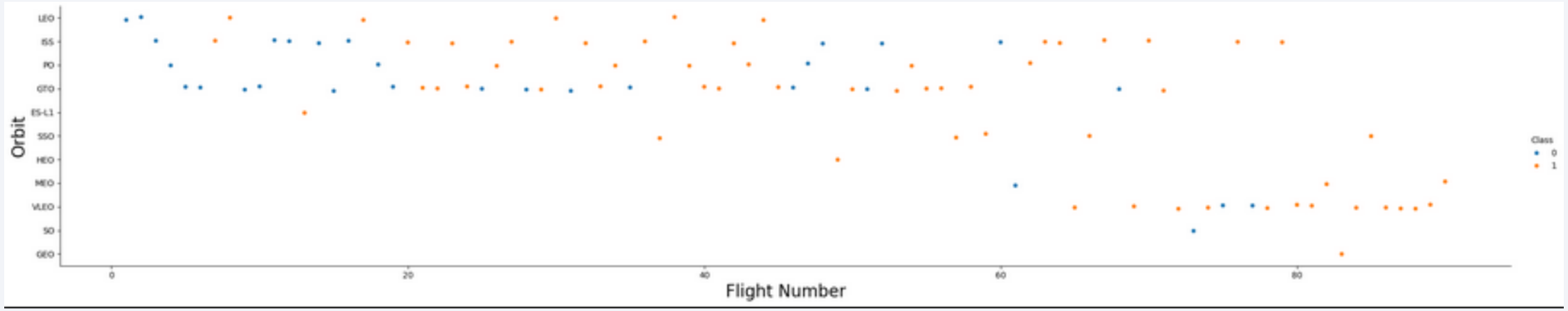
- Scatter plot of Success rate between payload and launch site
- Launch site on y-axis
- Payload on x-axis
- Blue = Failure
- Orange = Success

Success Rate vs. Orbit Type

- Bar plot of Success rate between orbit types
- Orbit type on y-axis
- Success rate on x-axis

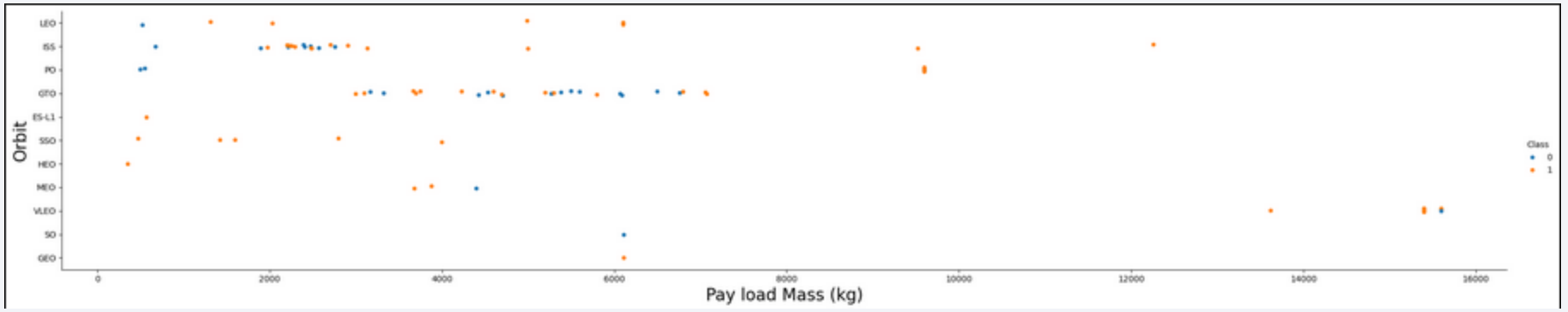


Flight Number vs. Orbit Type



- Scatter plot of Success rate between orbit and flight number
- orbit on y-axis
- Flight number on x-axis
- Blue = Failure
- Orange = Success

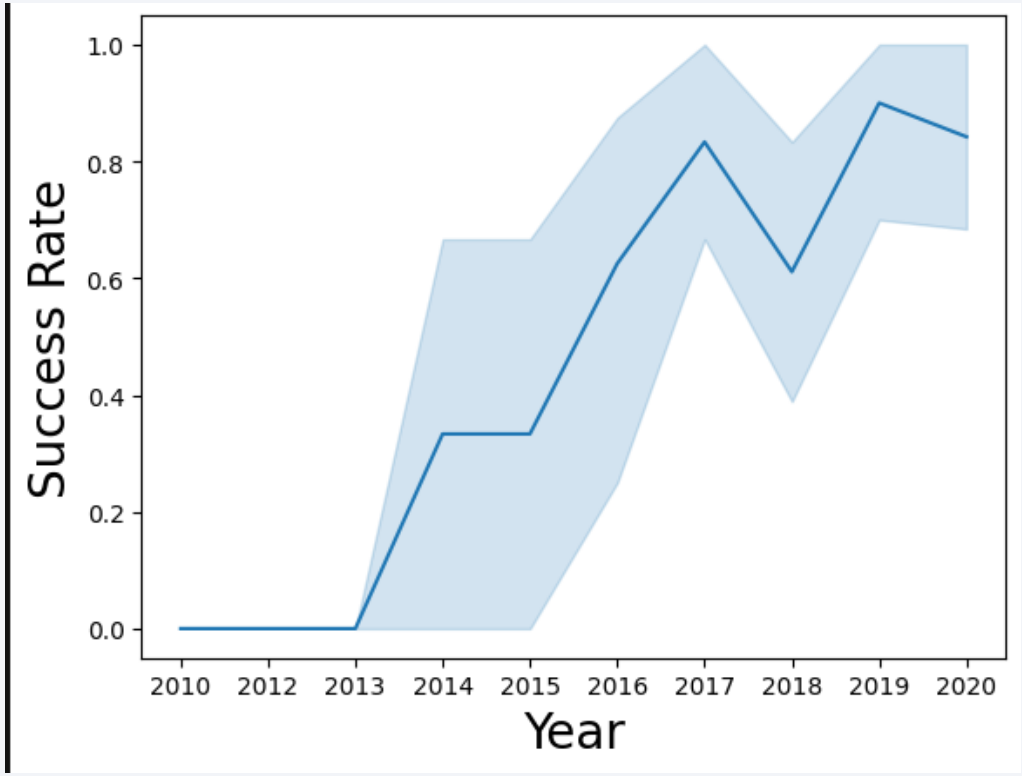
Payload vs. Orbit Type



- Scatter plot of Success rate between orbit and payload
- orbit on y-axis
- payload on x-axis
- Blue = Failure
- Orange = Success

Launch Success Yearly Trend

- Line chart of success rate development over the years
- Success rate (0-1) on y-axis
- Year on x-axis
- Line = Average
- Area = Variation



All Launch Site Names

- All Launch site names found in the dataset
- 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`

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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	

Total Payload Mass

- Calculated the total payload carried by boosters from NASA
- 45596kg

Average Payload Mass by F9 v1.1

- Calculated the average payload mass carried by booster version F9 v1.1
- 2928.4kg

First Successful Ground Landing Date

- The first successful landing outcome on ground pad
- 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List of the names of the boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

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

Boosters Carried Maximum Payload

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

F9 B5 B1048.4

F9 B5 B1049.4

F9 B5 B1051.3

F9 B5 B1056.4

F9 B5 B1048.5

F9 B5 B1051.4

F9 B5 B1049.5

F9 B5 B1060.2

F9 B5 B1058.3

F9 B5 B1051.6

F9 B5 B1060.3

F9 B5 B1049.7

2015 Launch Records

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

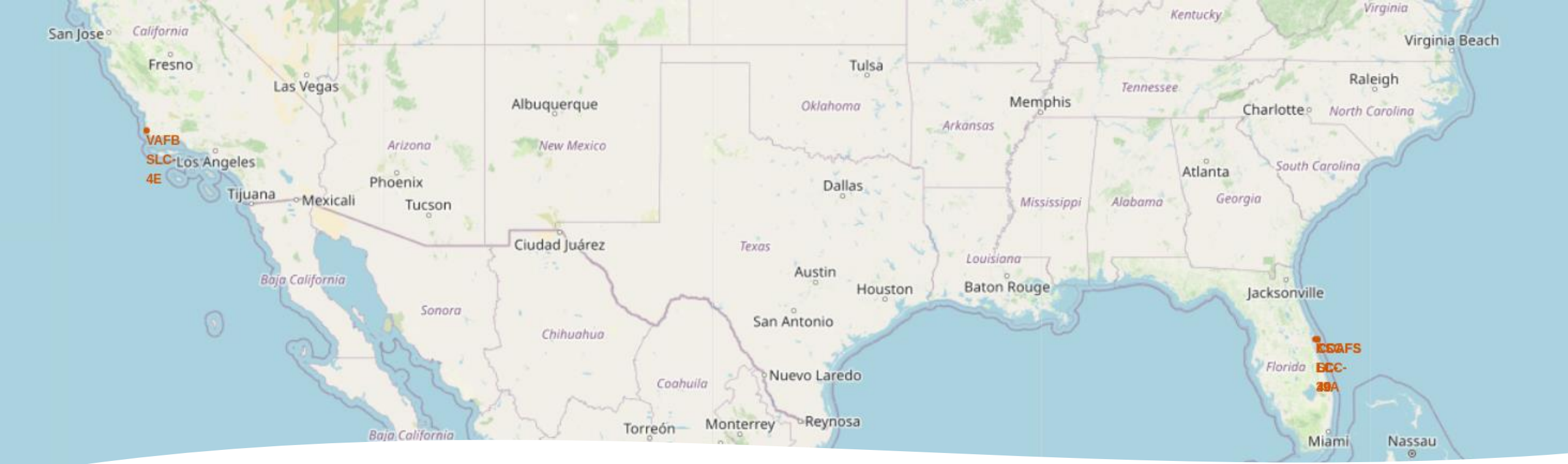
Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- 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
Success	38
No attempt	21
Success (drone ship)	14
Success (ground pad)	9
Failure (drone ship)	5
Controlled (ocean)	5
Failure	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1
No attempt	1

Section 3
Launch Sites
Proximities Analysis

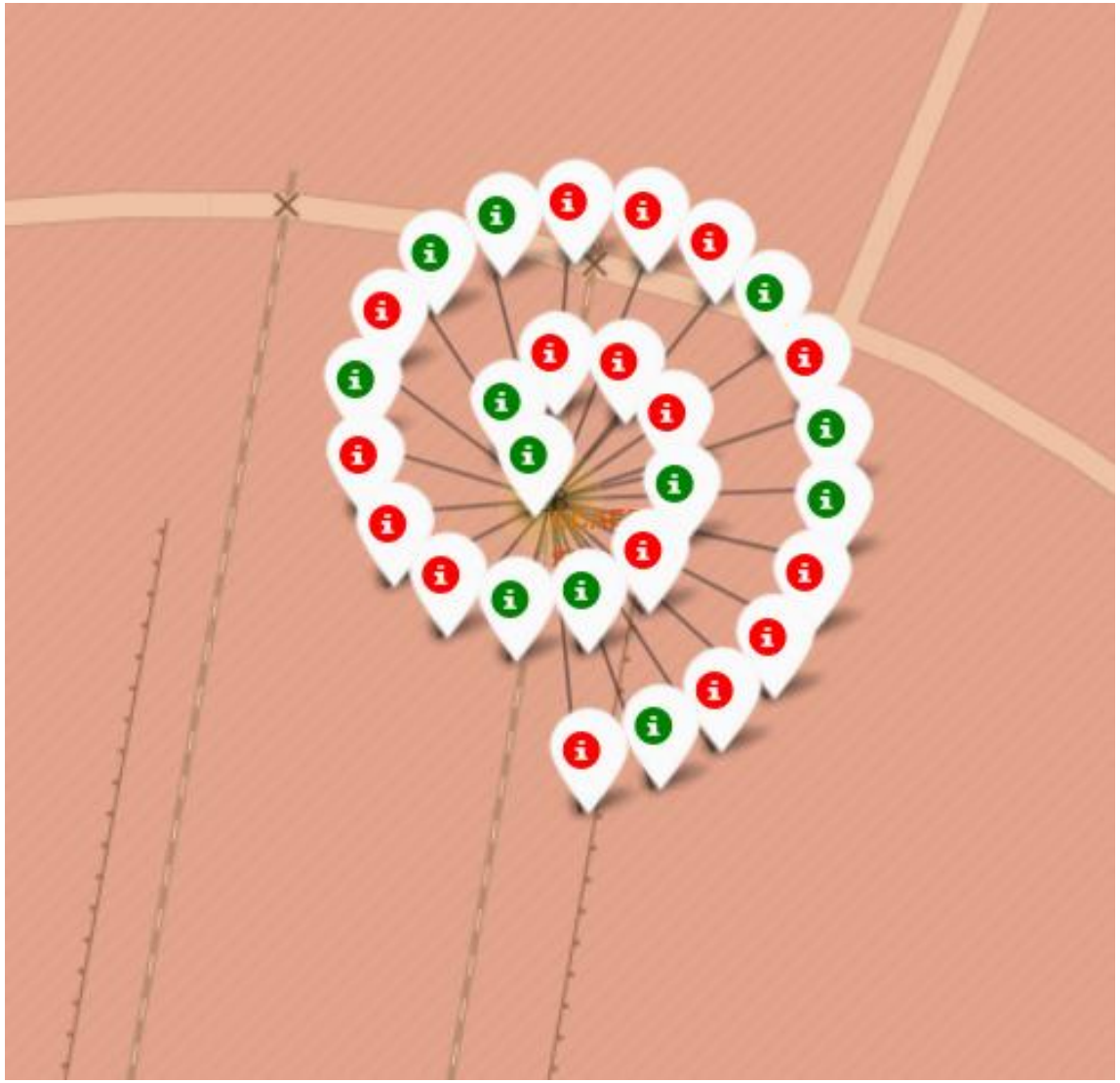


Launch site locations

- Launch site locations marked with red

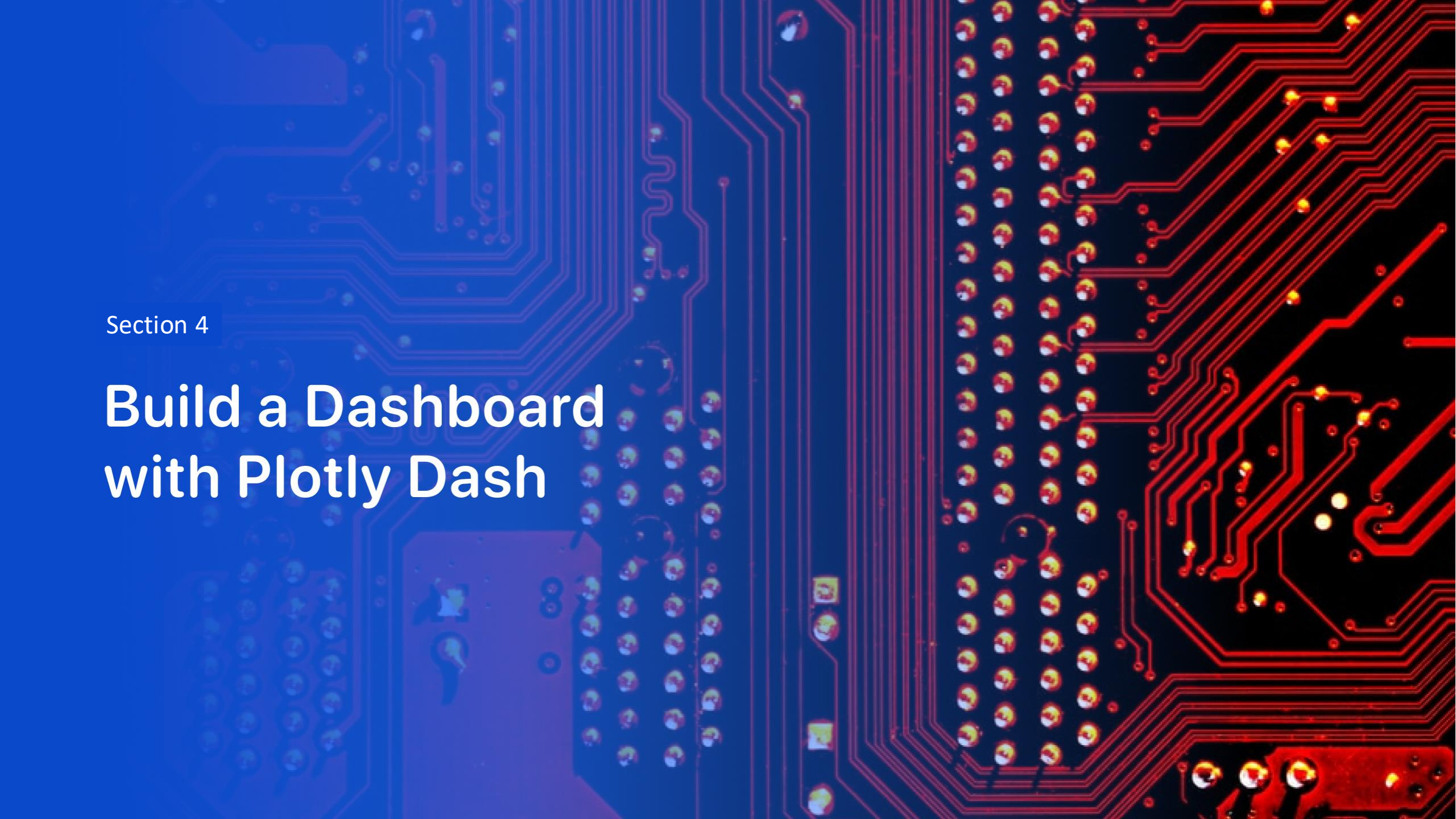
Launch outcomes

- Launch outcomes labeled with red and green markers
- Red = Failure
- Green = Success



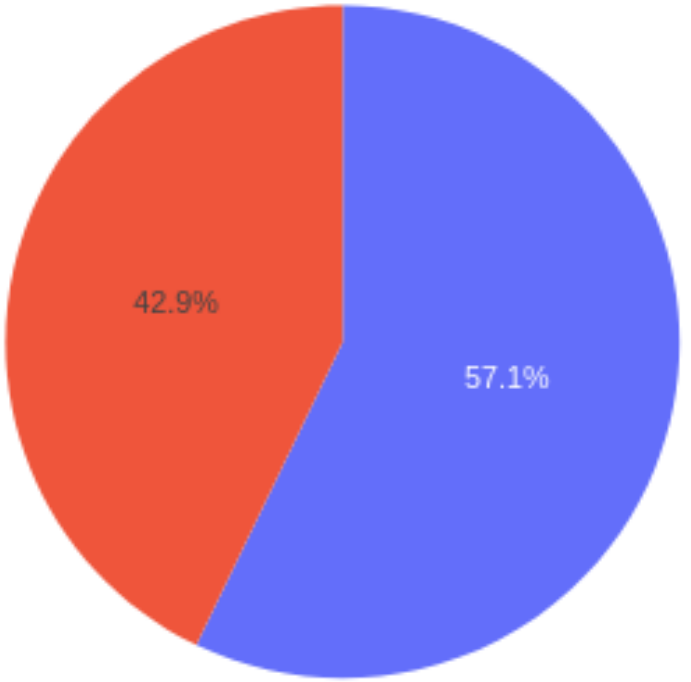
Distance on map

Distance from CCAFS SLC 40 to coast line


Section 4
Build a Dashboard with Plotly Dash

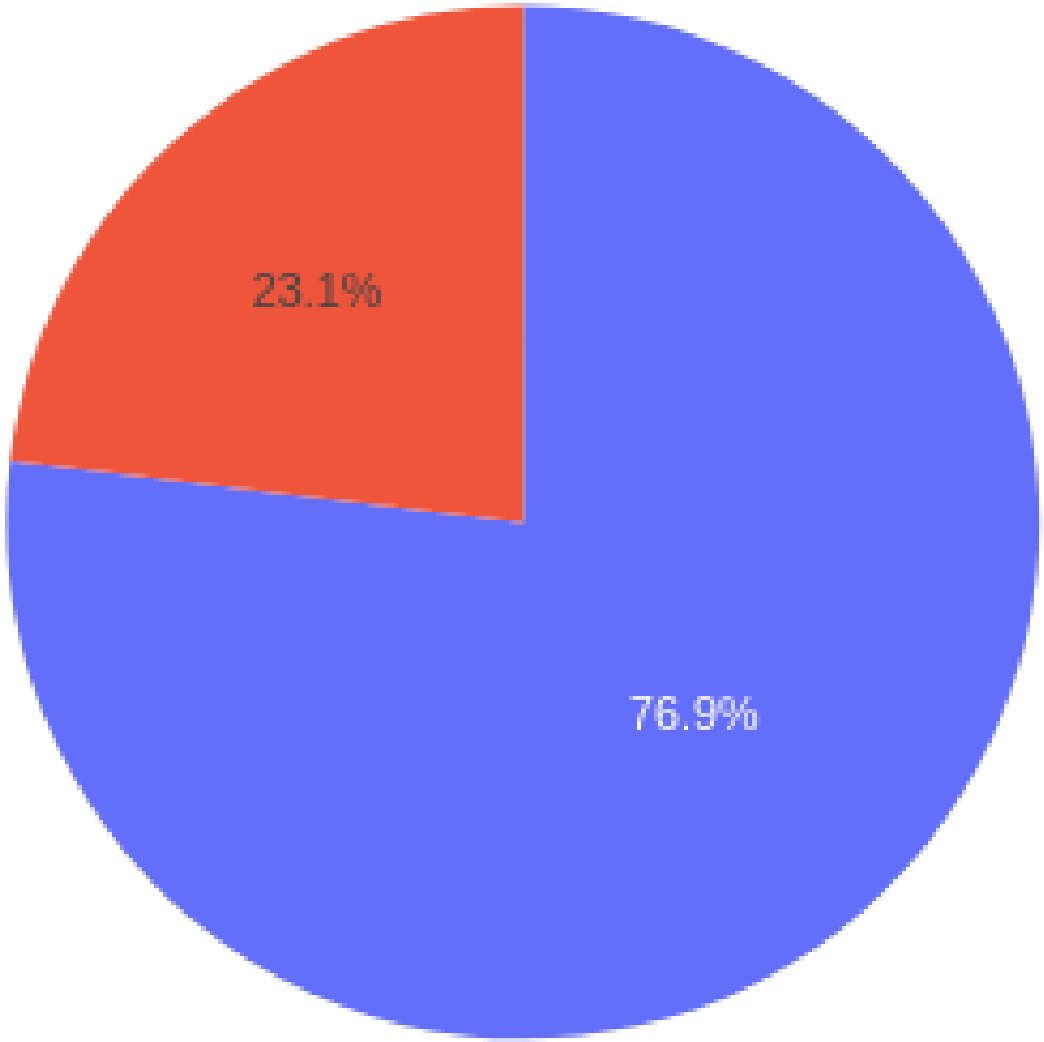
Pie chart success/Failure

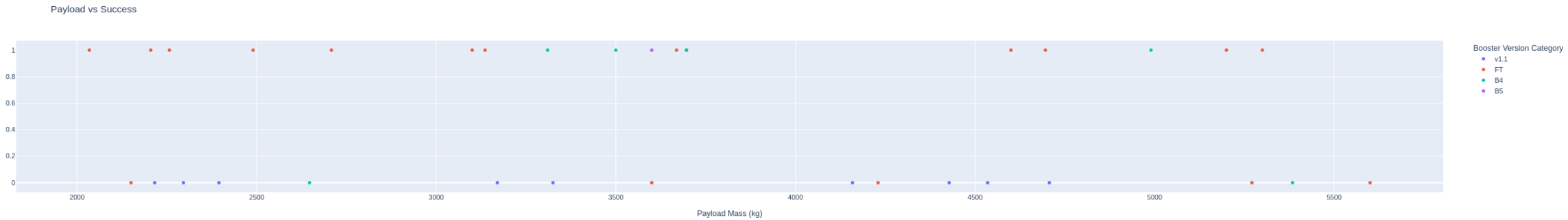
- On pie chart you can see success failure rate on all launch sites.
- Red = Failure
- Blue = Success



Highest success rate

- Highest success rate found on landing site KSC LC-39A





Payload vs success

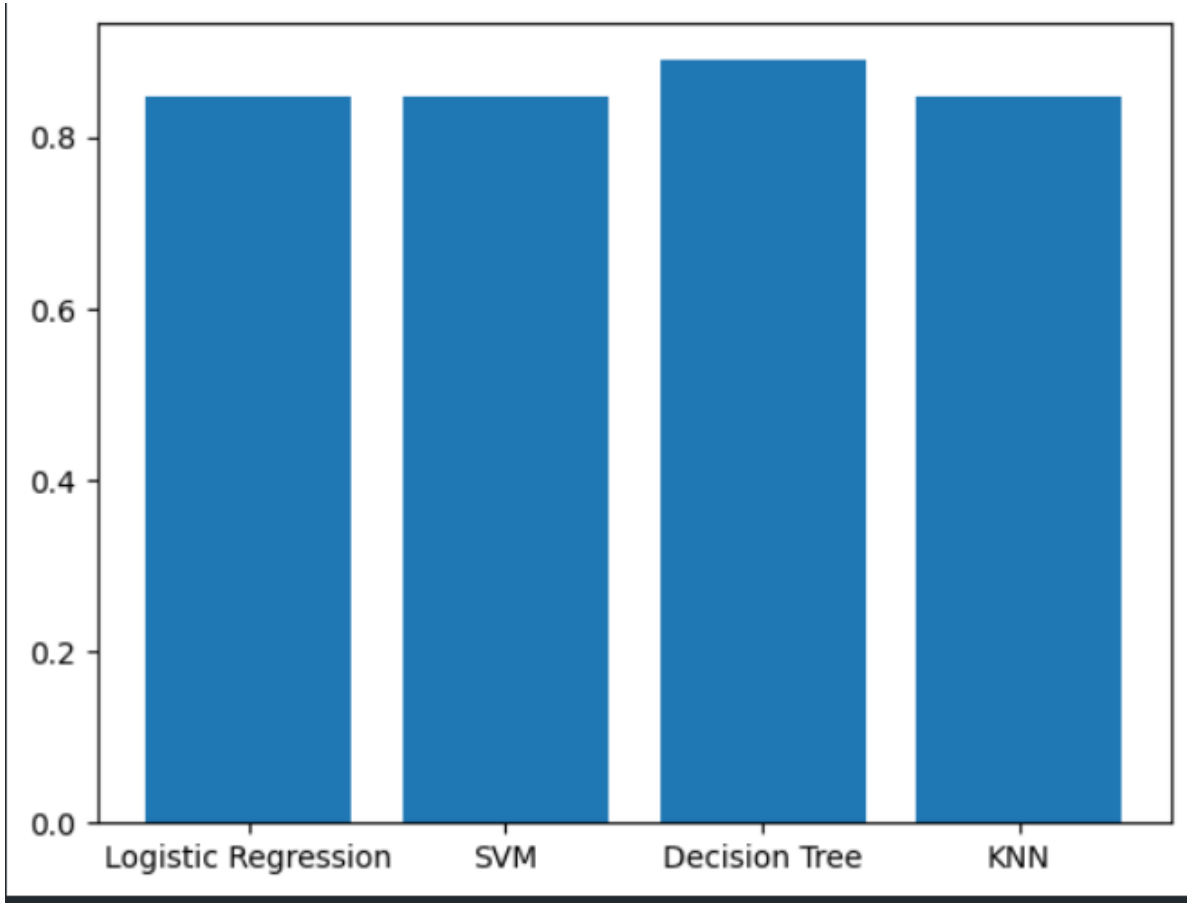
- On the plot you can see different booster versions with payload mass between 2000 and 6000
- Color indicates Booster version
- Bottom = Failure, top = Success
- Mass on x-axis

Section 5

Predictive Analysis (Classification)

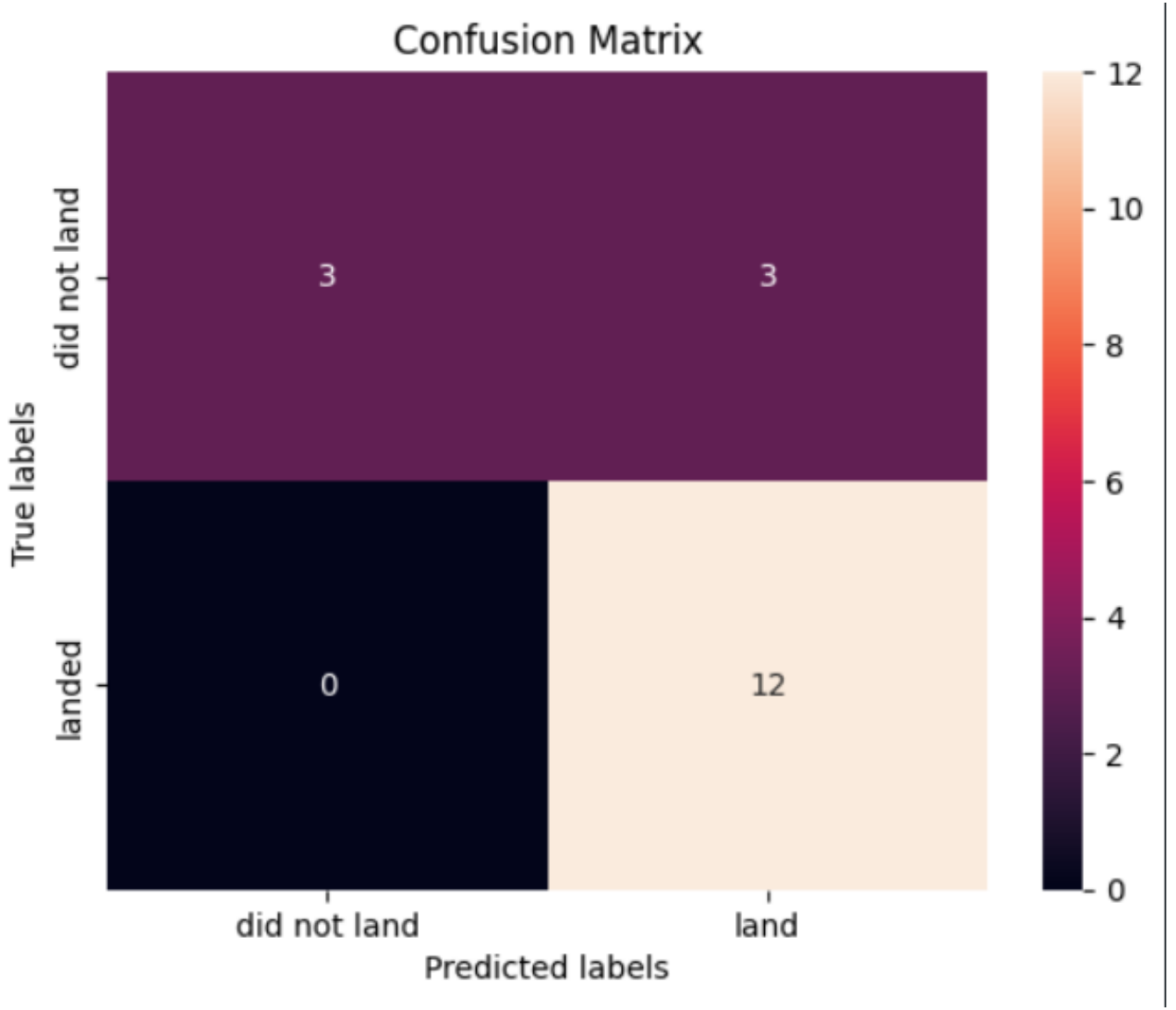
Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart
- Best accuracy model: Decision Tree



Confusion Matrix

Best performing model got 12 true positives and 3 false positives.



Conclusions

- Success more common when time passes by
- Launch site, Payload, orbit target and other parameters may affect success
- Mission success can be predicted from given data (not with 100% accuracy)

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!