IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

S. K.
3 August 2024

Space Y Corporation

Outline

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

Executive Summary

- Methodologies: Data analysis including…
  1. data collection
  2. data wrangling
  3. exploratory data analysis (EDA)
  4. visual analysis
  5. predictive analysis
- Results:
  - Space X saves costs if it can reuse the first stage
  - Possible in 66% of the cases (especially Orbit Type, Flight Number, Payload Mass and Launch Site affect the success)
  - Prediction possible with Decision Tree Classifier

Introduction

- Space X launches cost less than competitors
- Why? They reuse the first stage of a rocket launch if successfully recovered
- To be competitive the following question need to be answered:
  - What is the price of a given launch from Space X?
    - Underlying question:
      Will Space X reuse the first stage for a given launch?

Section 1

Methodology

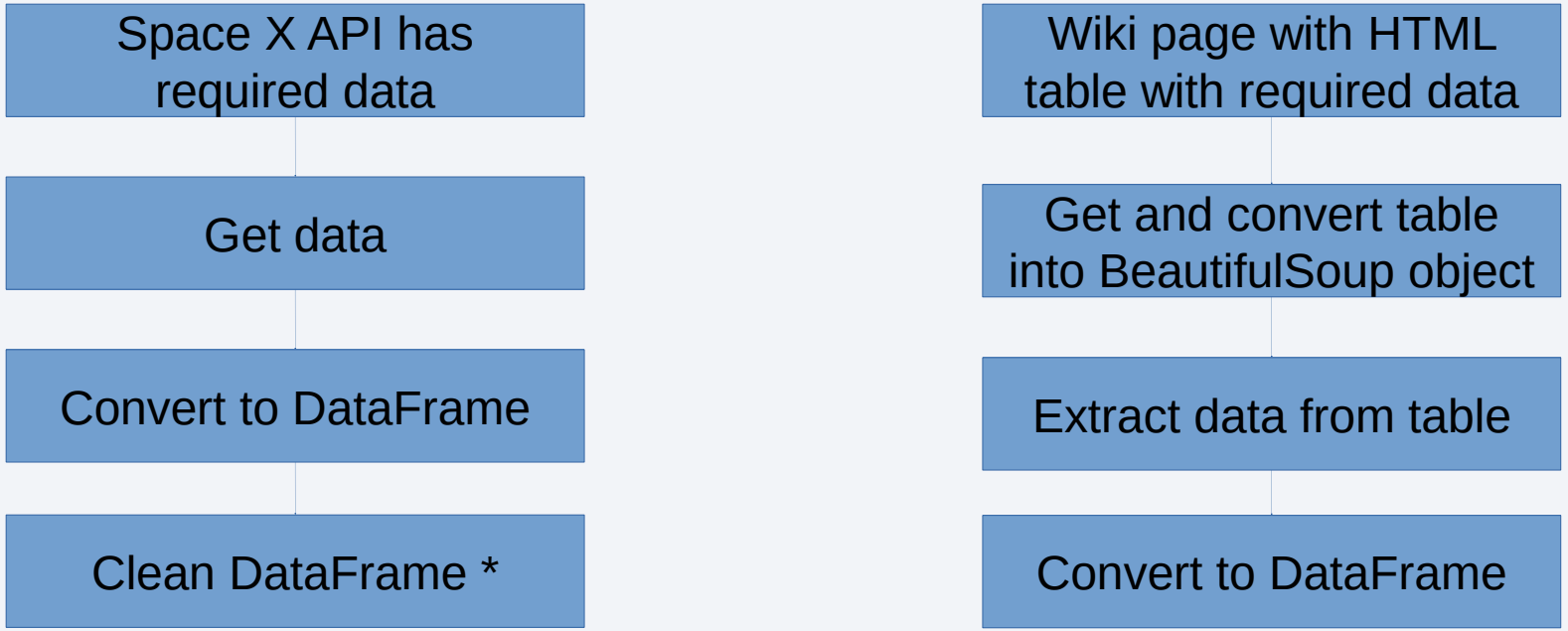
Methodology

Executive Summary

- Data collection methodology: Via Space X API and Wiki pages
- Perform data wrangling: Replace missing values, Extract labels for columns of type object if required for analysis
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models using sklearn

Data Collection

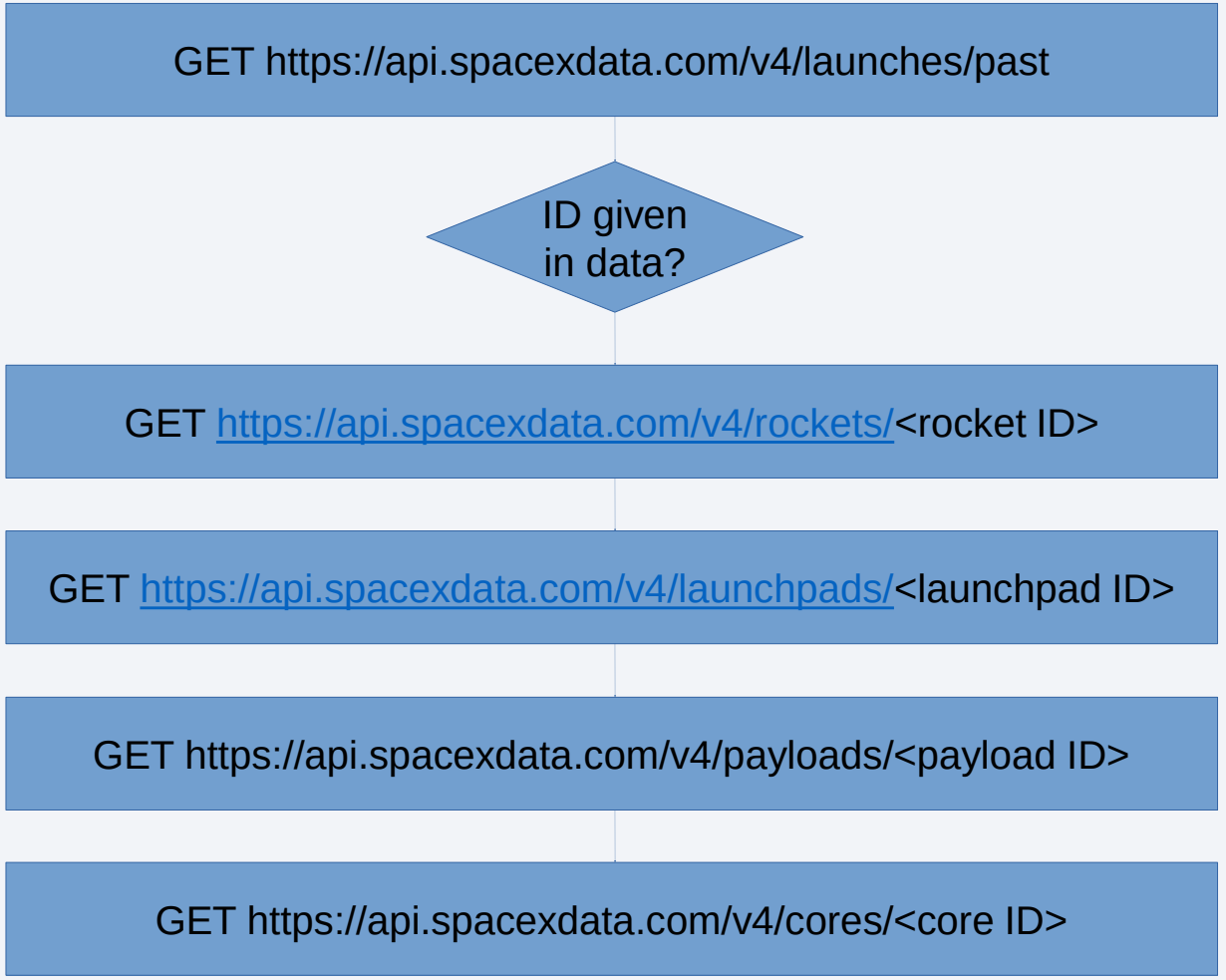
- Source: Space X API, Wiki pages
  Python libraries: requests, pandas, beautifulsoup4
- The collection process for each:



* Clean = if required: remove not required cells; ensure one value per cell; restrict time interval

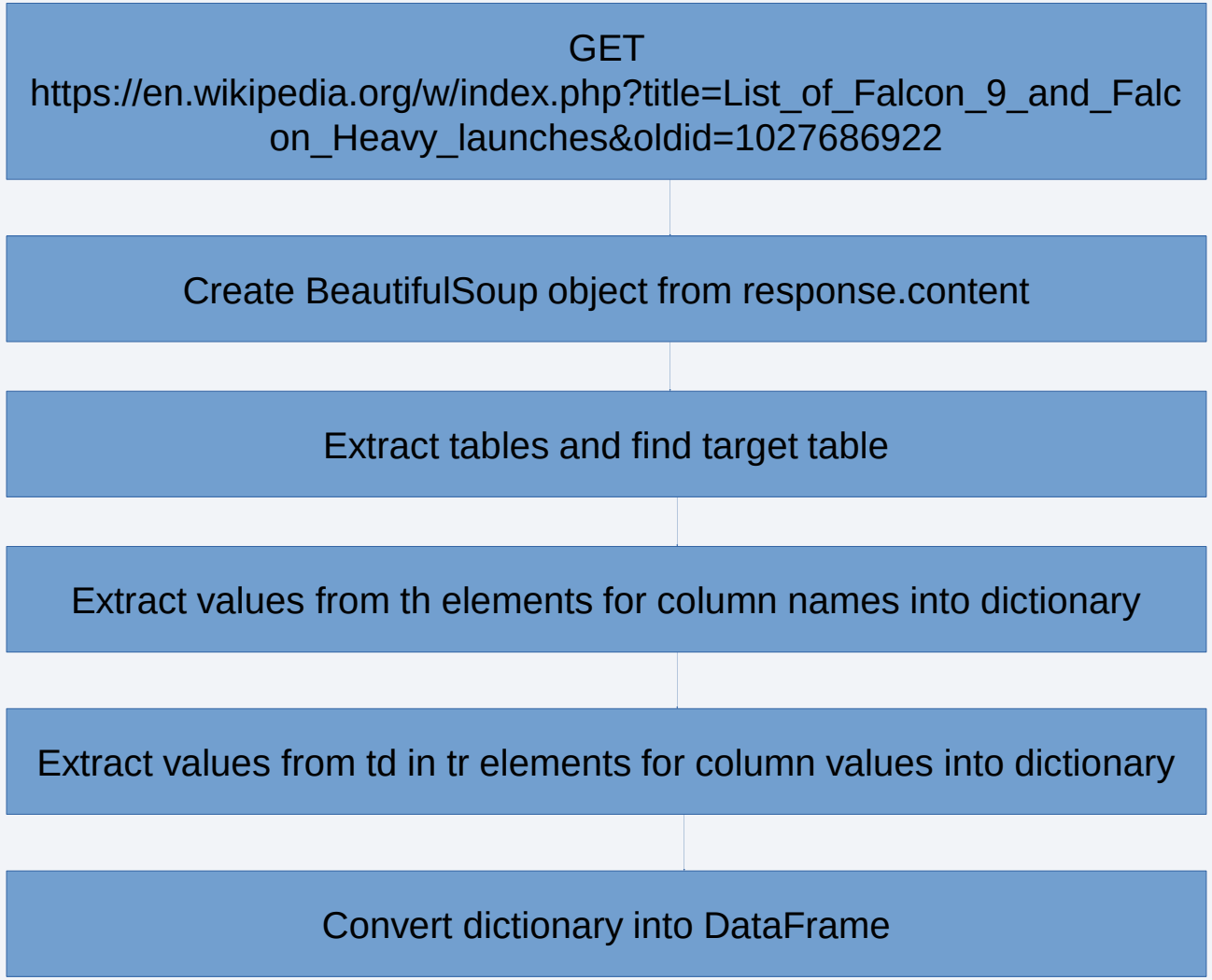
Data Collection – SpaceX API

- https://github.com/Sophia-attempto/coursera_ibm_data_science_capstone/blob/main/spacex_assignment/0_1_spacex_data-collection-api.ipynb



Data Collection - Scraping

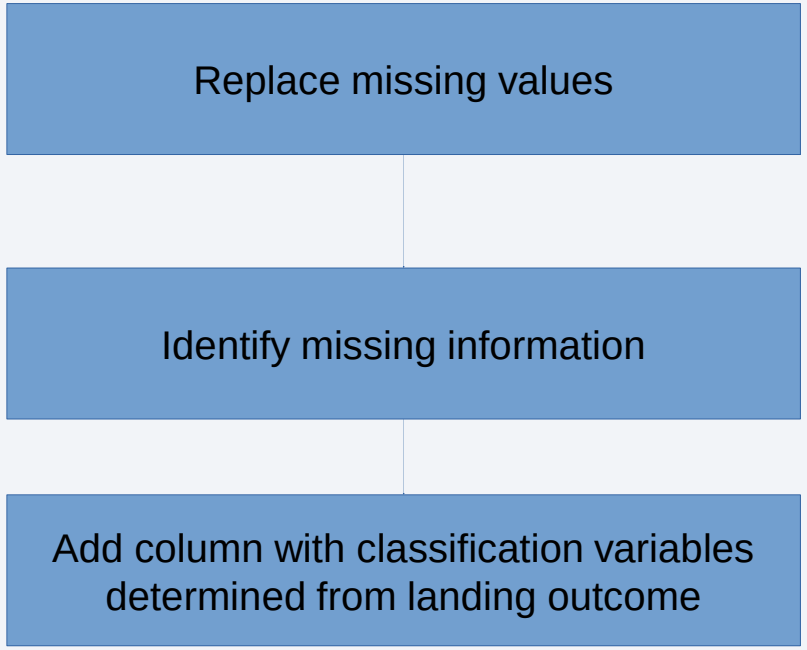
- https://github.com/Sophia-attempto/coursera_ibm_data_science_capstone/blob/main/spacex_assignment/0_2_spacex_data-collection-webscraping.ipynb



Data Wrangling

First insight: the success rate of Space X past recoveries of the first stage is round about 66%

https://github.com/Sophia-attempto/coursera_ibm_data_science_capstone/blob/main/spacex_assignment/1_spacex_data-wrangling.ipynb



Exploratory Data Analysis (EDA) with Data Visualization

EDA Goal: identify features in the data which have an affect on the Landing Outcome success rate

Plotted charts:

1. Five scatter point charts: how do features relate?
2. Bar chart: how do features relate?
3. Line chart: how is the yearly trend of Landing Outcome success?

https://github.com/Sophia-attempto/coursera_ibm_data_science_capstone/blob/main/spacex_assignment/2_spacex_eda-data-visualization.ipynb

EDA with SQL – Part 1

https://github.com/Sophia-attempto/coursera_ibm_data_science_capstone/blob/main/spacex_assignment/3_spacex_eda-sqllite.ipynb

- Select unique launch site names for an overview
- Select launch sites starting with CCA for an overview
- Select total payload mass of rockets launched by NASA (CRS)
- Select average payload mass of rockets with booster version F9 v1.1
- Select date when first Landing Outcome success was achieved
- Select booster names with Landing Outcome success on drone ship and Payload Mass greater than 4000 but less than 6000

EDA with SQL – Part 2

- Select total number of Mission Outcomes success and failure
- Select Booster Versions which have carried the max. Payload Mass
- Select Landing Outcome failure on drone ship in year 2015
- Select count of Landing Outcomes between 2010-06-04 and 2017-03-20

Build an Interactive Map with Folium

Goal: Identify factors for an optimal launch site location

Added map objects:

- Circles for space X launch site locations: to get an overview
- Red and green markers for Launch Outcomes:
  to see the success rate per launch site
- Lines: to analyze distances to proximities (railway, highway, city, equator line...)

https://github.com/Sophia-attempto/coursera_ibm_data_science_capstone/blob/main/spacex_assignment/4_spacex_folium_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

Added dashboard elements:

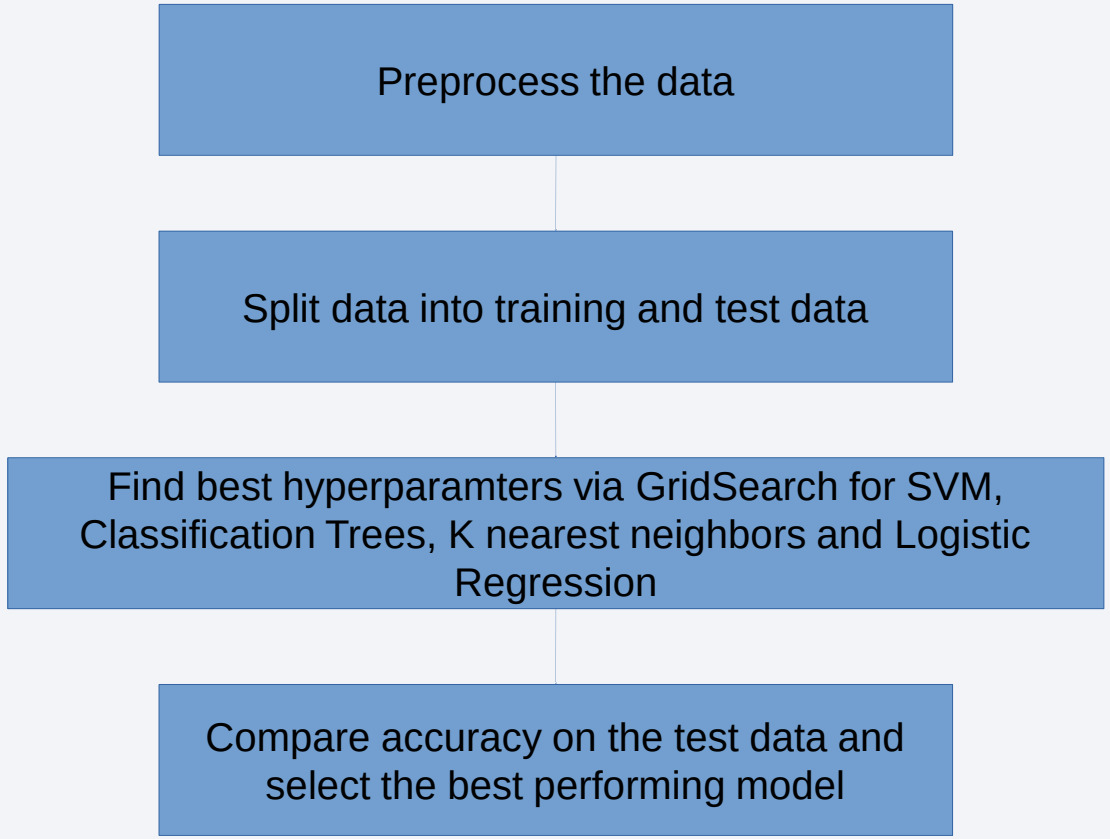
- Dropdown: to select **all** or a **specific** Launch Site
- Pie chart:
  - For all: to show the total successful launches count for all sites
  - For specific: to show the Success vs. Failed counts for the site
- Slider: to select Payload Mass range for the scatter chart
- Scatter chart: to show the correlation between Payload Mass and Launch success

https://github.com/Sophia-attempto/coursera_ibm_data_science_capstone/blob/main/spacex_assignment/5_spacex_dash_app.py

Predictive Analysis (Classification)

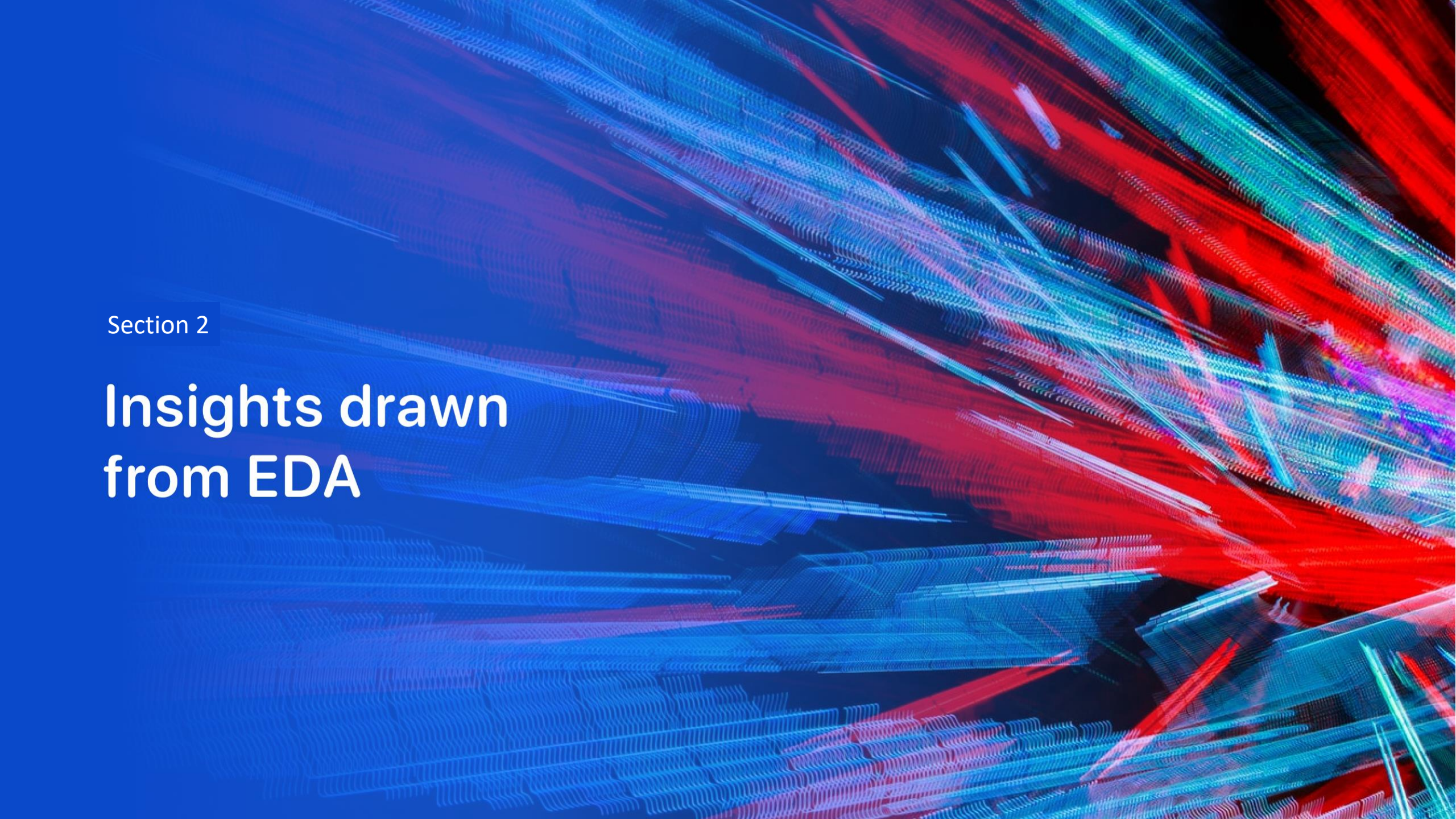
Goal: find best performing classification model

https://github.com/Sophia-attempto/coursera_ibm_data_science_capstone/blob/main/spacex_assignment/6_spacex_ML_prediction.ipynb



Results

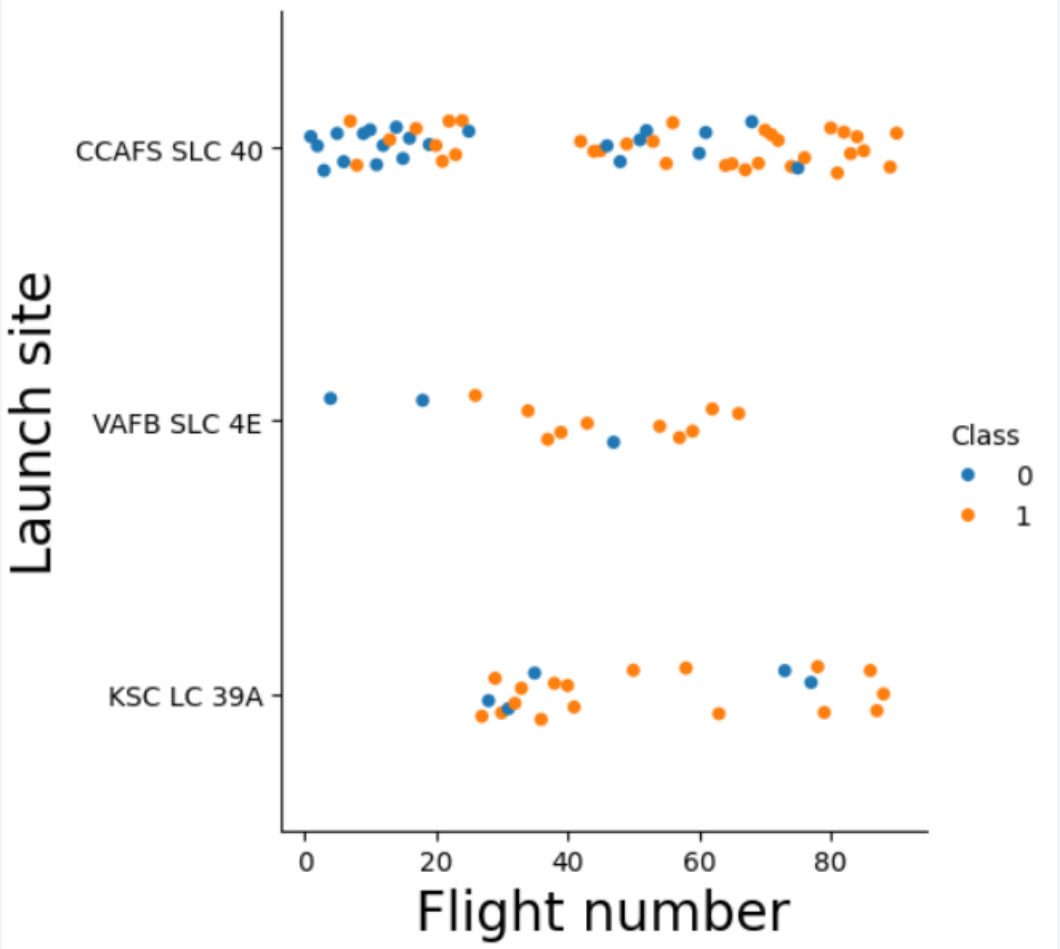
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

Insights:

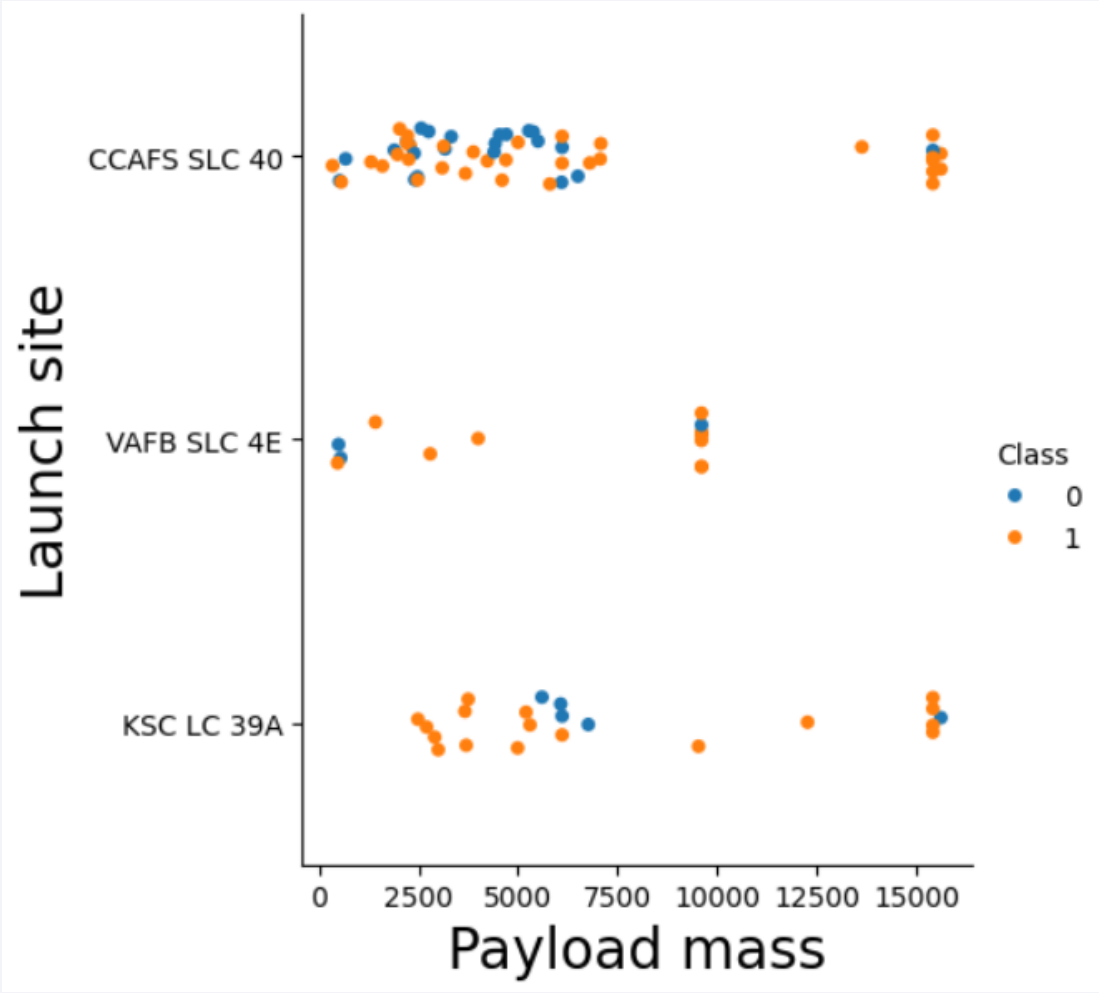
- the higher the Flight Number the more likely is a successful recovery
- CCAFS SLC 40 and KSC LC 39A both have higher flight numbers than VAFB SLC 4E



Payload vs. Launch Site

Insights:

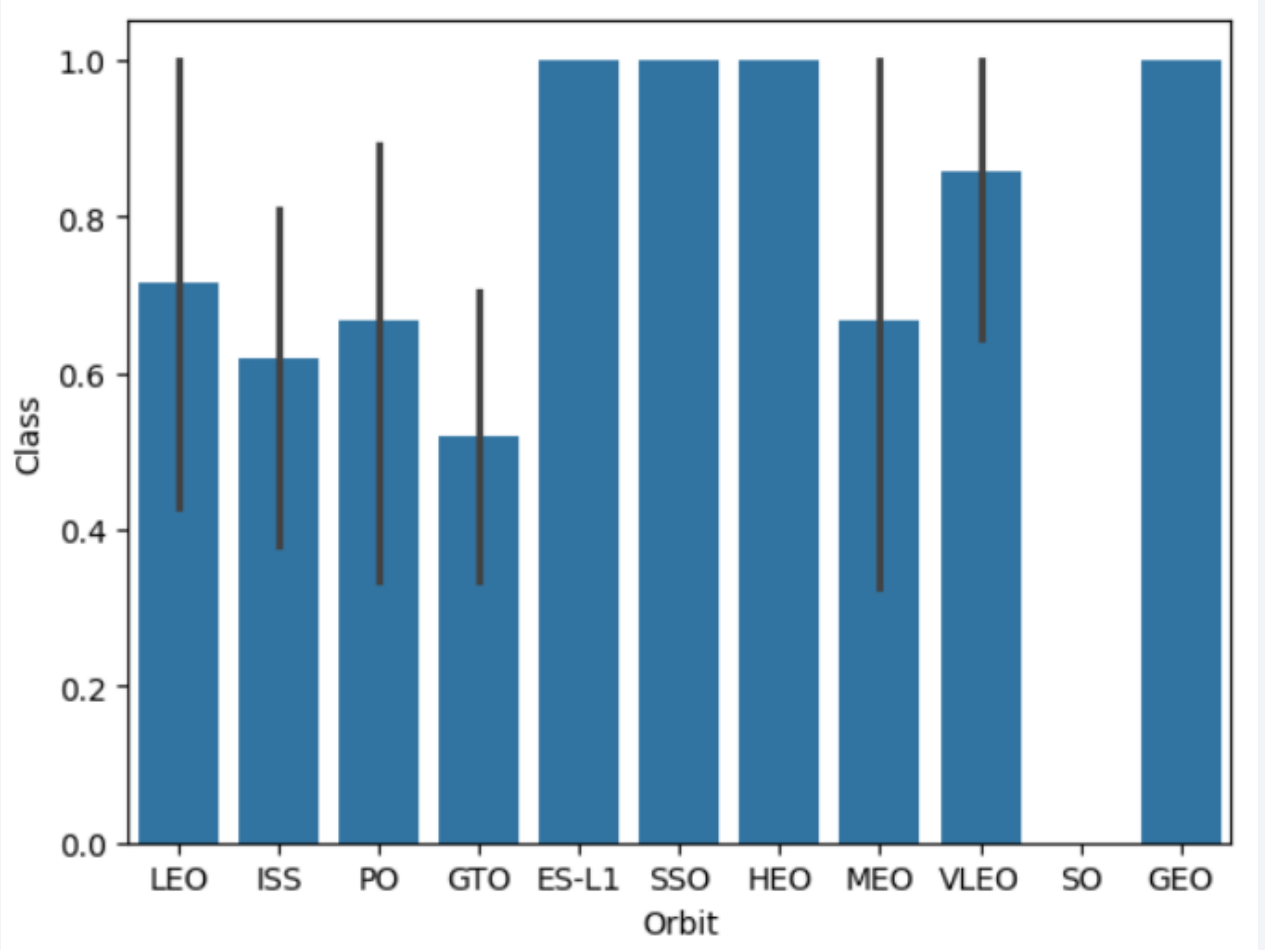
- the more massive the Payload the less likely is a successful recovery
- VAFB SLC 4E has no rockets launched for heavypayload mass (greater than 10000)



Success Rate vs. Orbit Type

Insight:

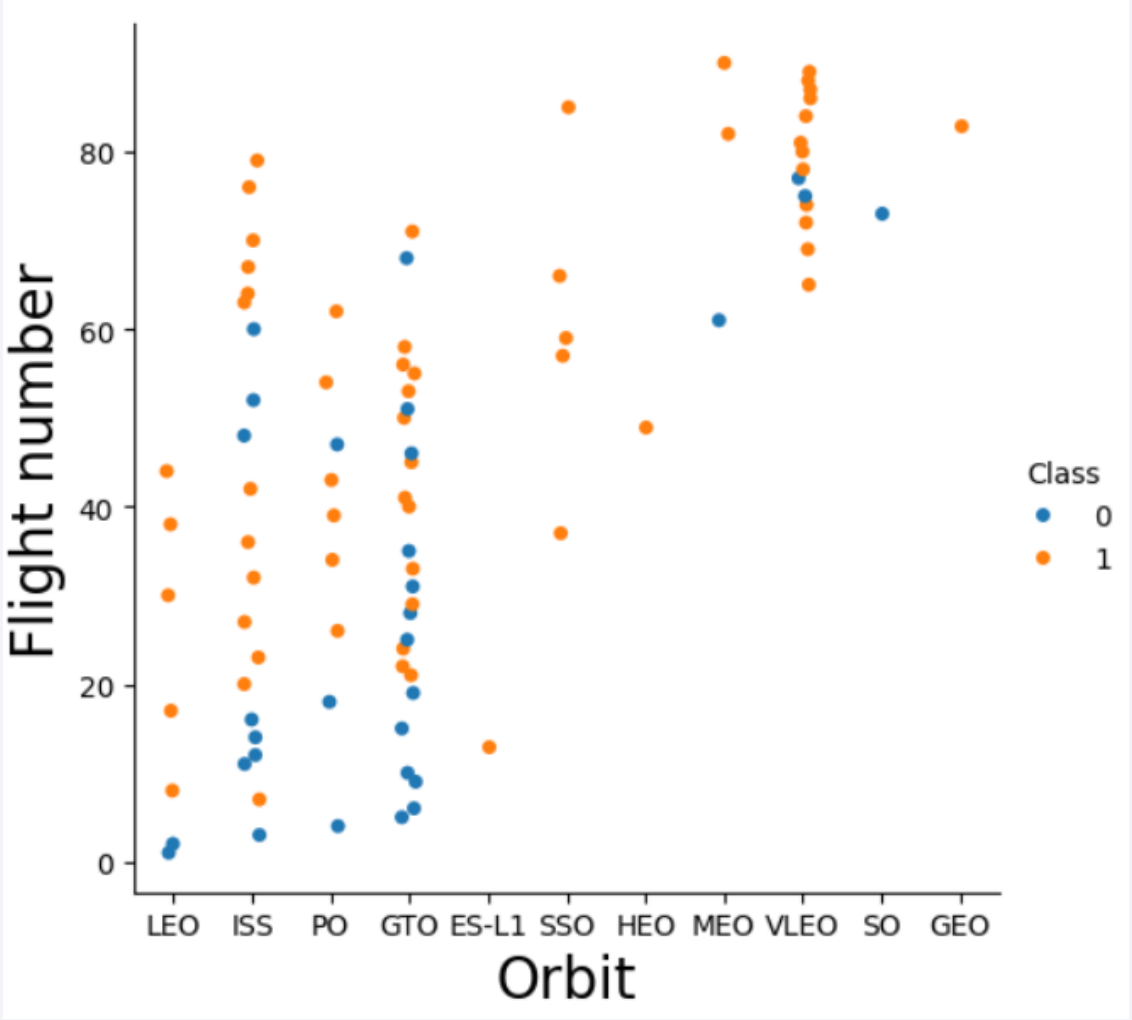
Orbit Types with high success rate: ES-L1, SSO, HEO, GEO



Flight Number vs. Orbit Type

Insights:

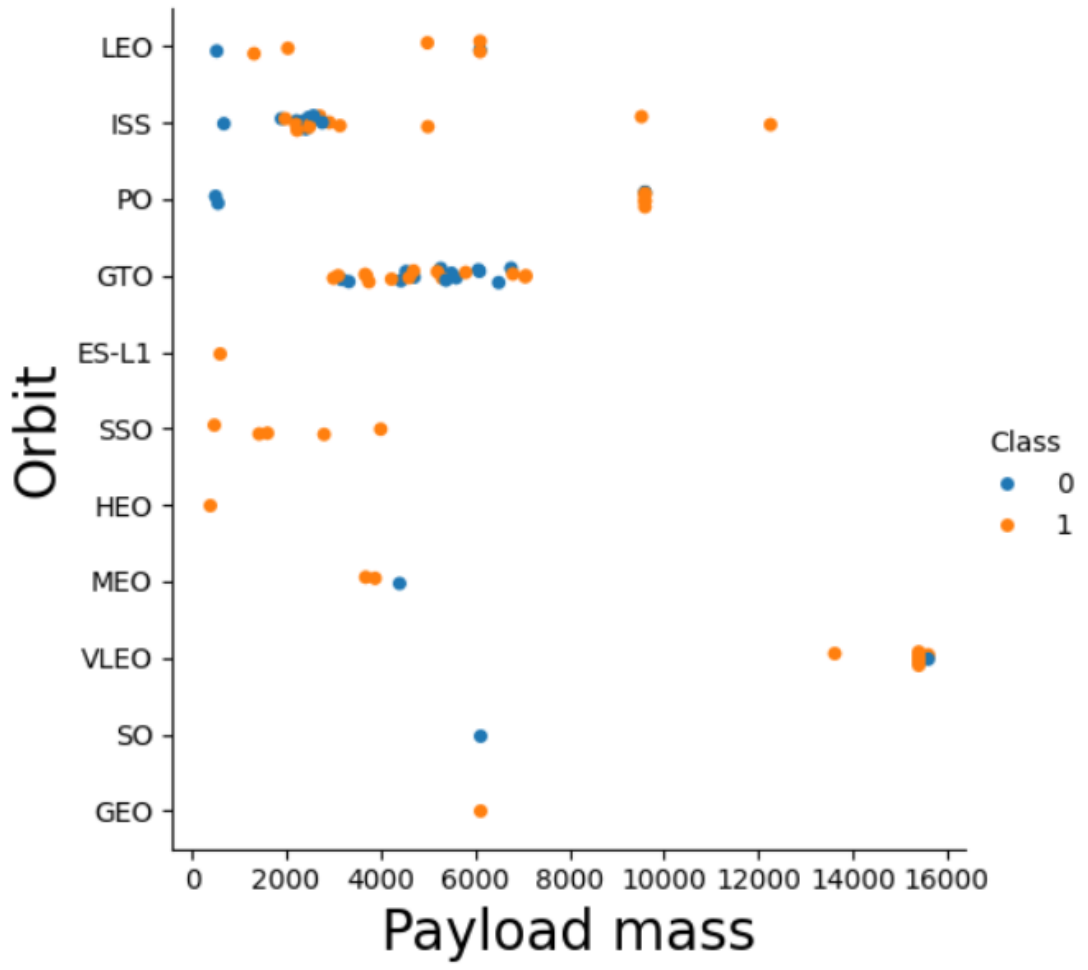
- LEO: the higher the Flight Number the more likely the success
- GTO: no relationship



Payload vs. Orbit Type

Insights:

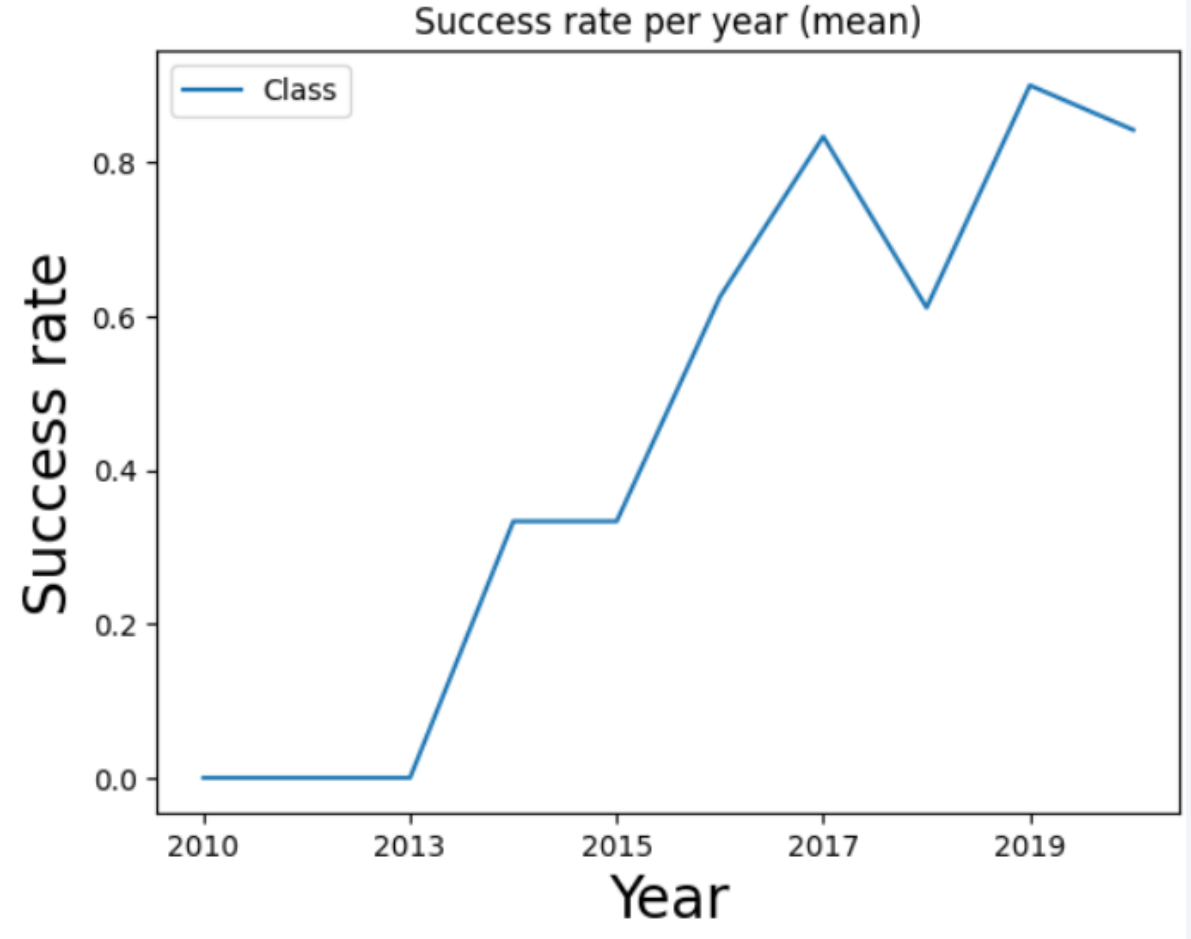
- Polar, LEO and ISS: with heavy payloads more successes
- GTO: no relationship



Launch Success Yearly Trend

Insights:

Since 2013 success rate kept increasing till 2017



All Launch Site Names

Unique Launch Site Names:

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

Launch Site Names Begin with 'CCA'

First 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
0	2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	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
3	2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	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

Total payload carried by boosters launched by NASA (CRS):

```
PAYLOAD_MASS__KG_    45596
dtype: int64
```

Average Payload Mass by F9 v1.1

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

```
PAYLOAD_MASS__KG_    2928.4
dtype: float64
```

First Successful Ground Landing Date

Date of the first successful landing outcome on ground pad:

min(Date)
2018-07-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Names of boosters which have successfully landed on drone ship with Payload Mass greater than 4000 but less than 6000

Payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

3 failure vs. 98 successful mission outcomes:

count("Mission_Outcome")
3
98

Boosters Carried Maximum Payload

Names of the booster which have carried the maximum Payload Mass:

Booster_Version
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

Failed Landing Outcome in drone ship, booster versions, and launch site names per month in year 2015

Month in 2015	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

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(*)
Controlled (ocean)	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	10
Precluded (drone ship)	1
Success (drone ship)	5
Success (ground pad)	3
Uncontrolled (ocean)	2

Section 3

Launch Sites Proximities Analysis

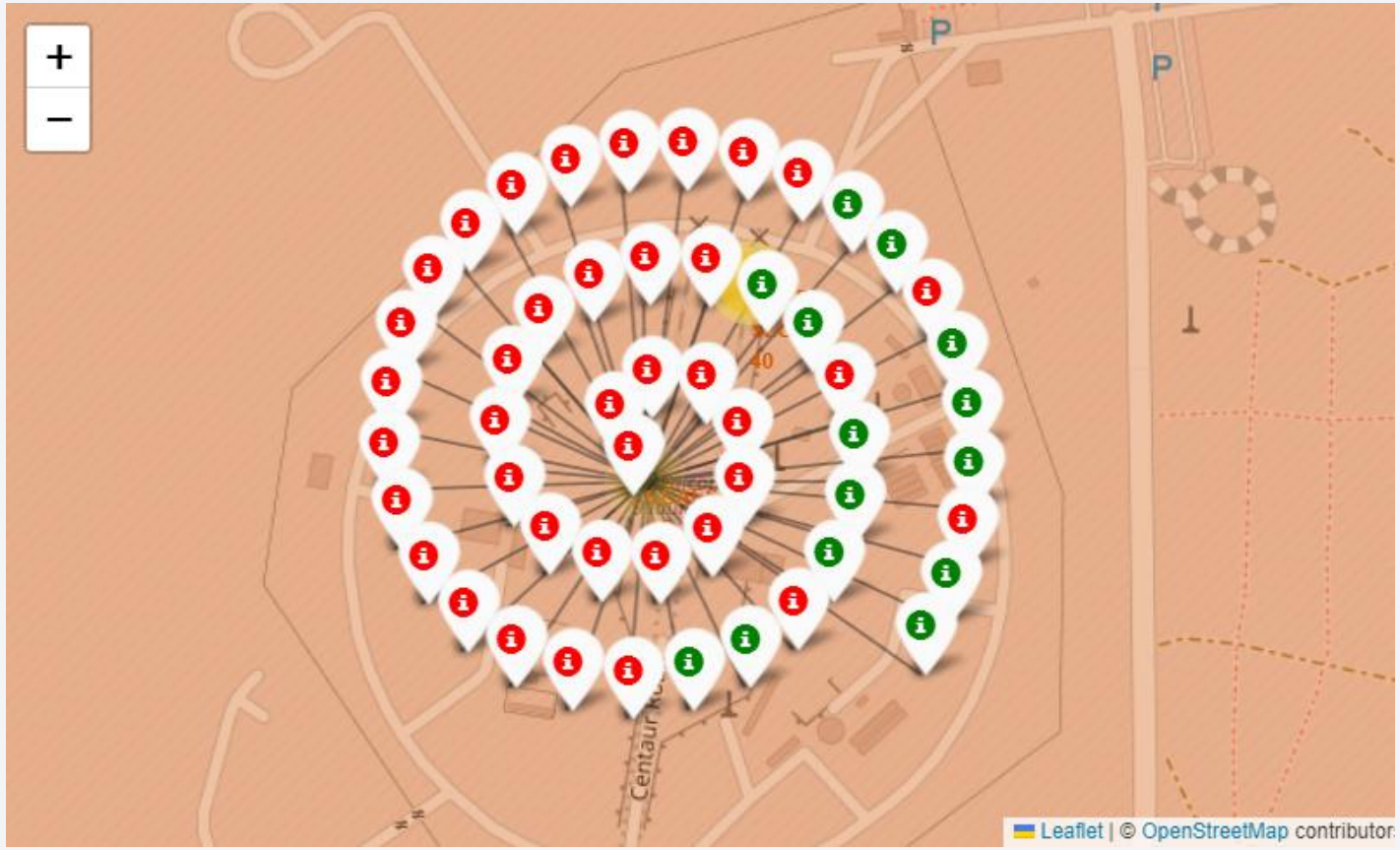
Space X Launch Site locations

Insights:

- All launch sites are in proximity to the Equator line
- All launch sites are in proximity to the coast

Launch Outcomes for Launch Sites

Example for Launch Site CCAFS SLC-40:



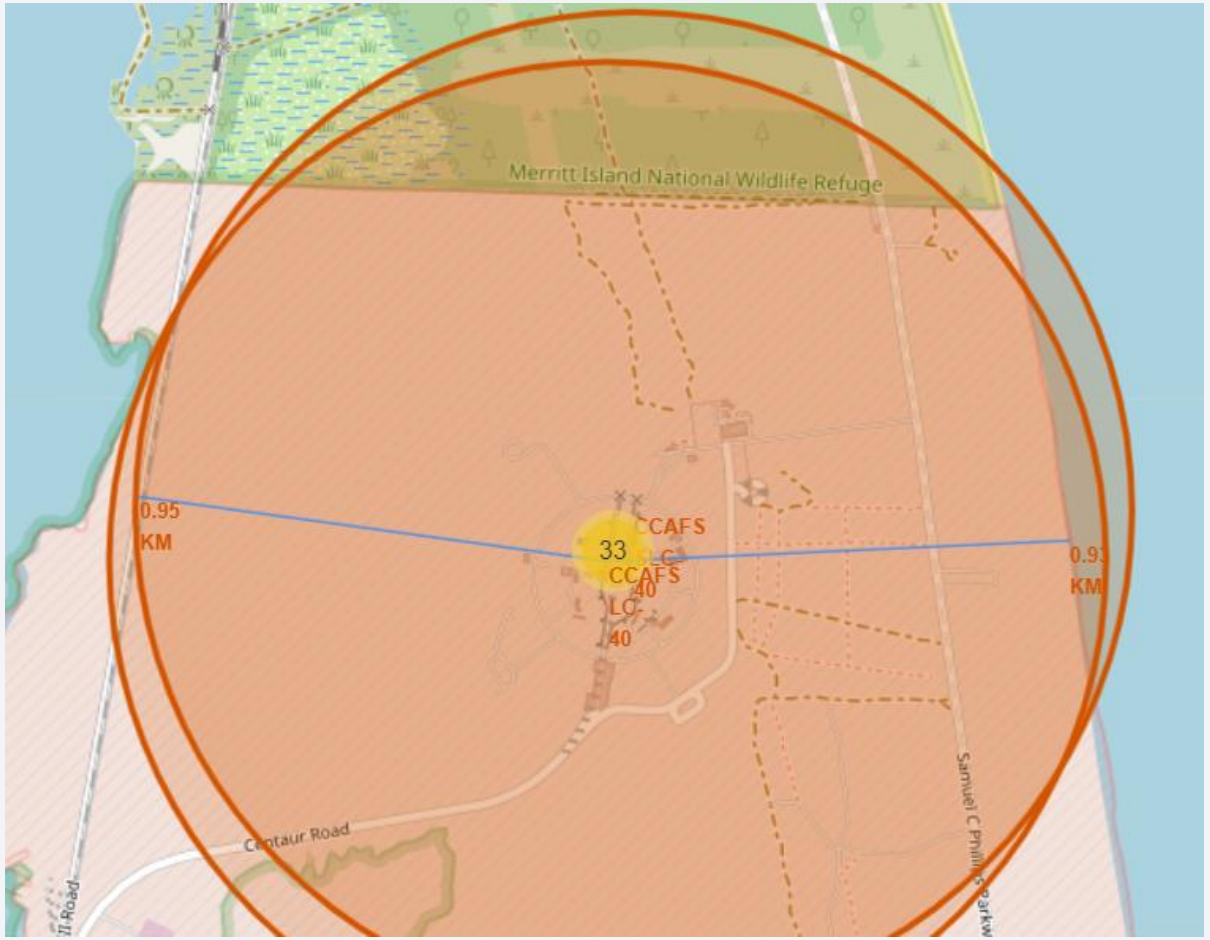
Insights:

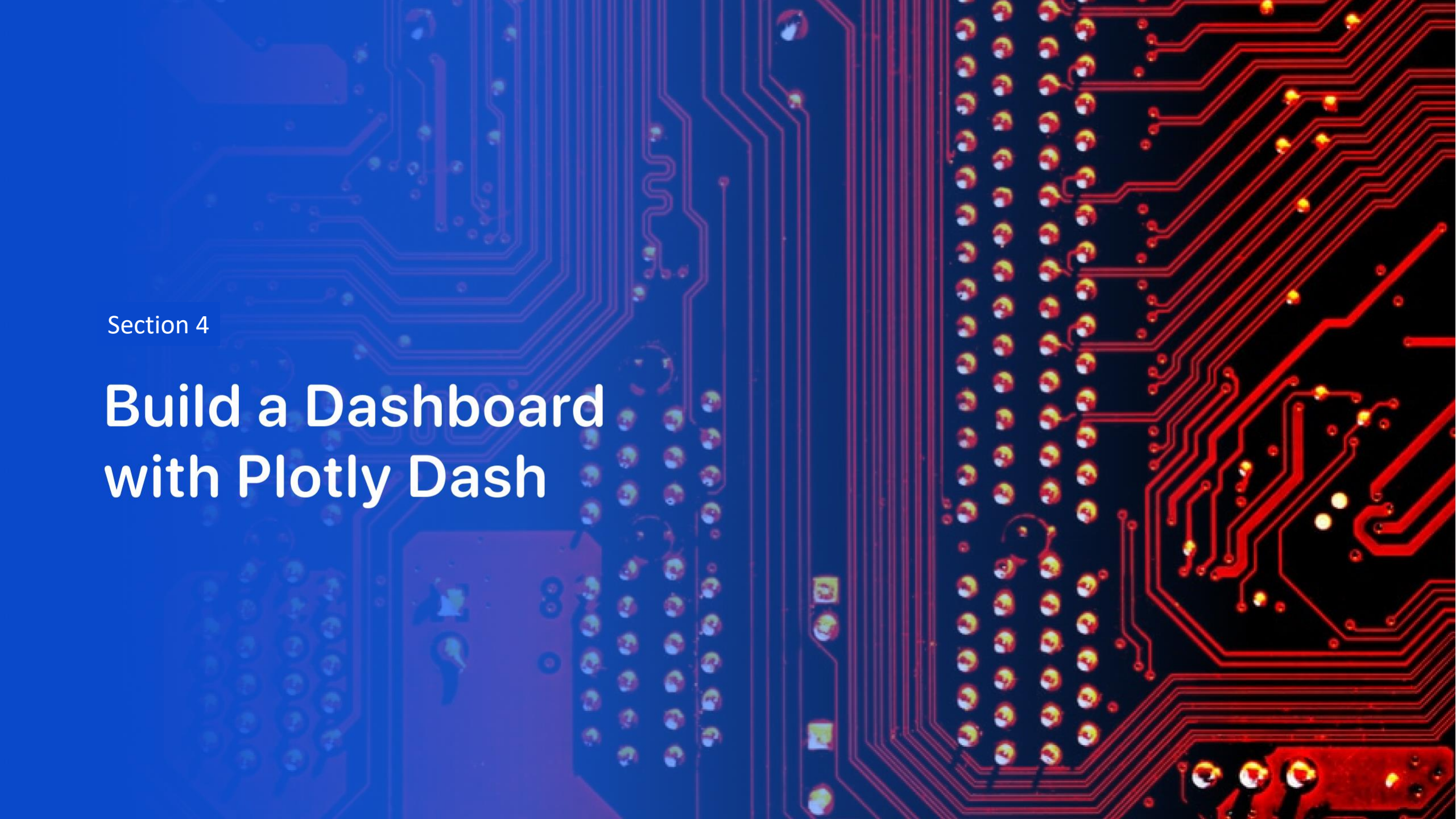
- CCAFS SLC-40 has the most launches
- KSC LC-39A has the most successful launches

Proximities to Launch Sites

Insights:

- Near the launch sites are the coast, railways and highways
- Cities are not in proximity

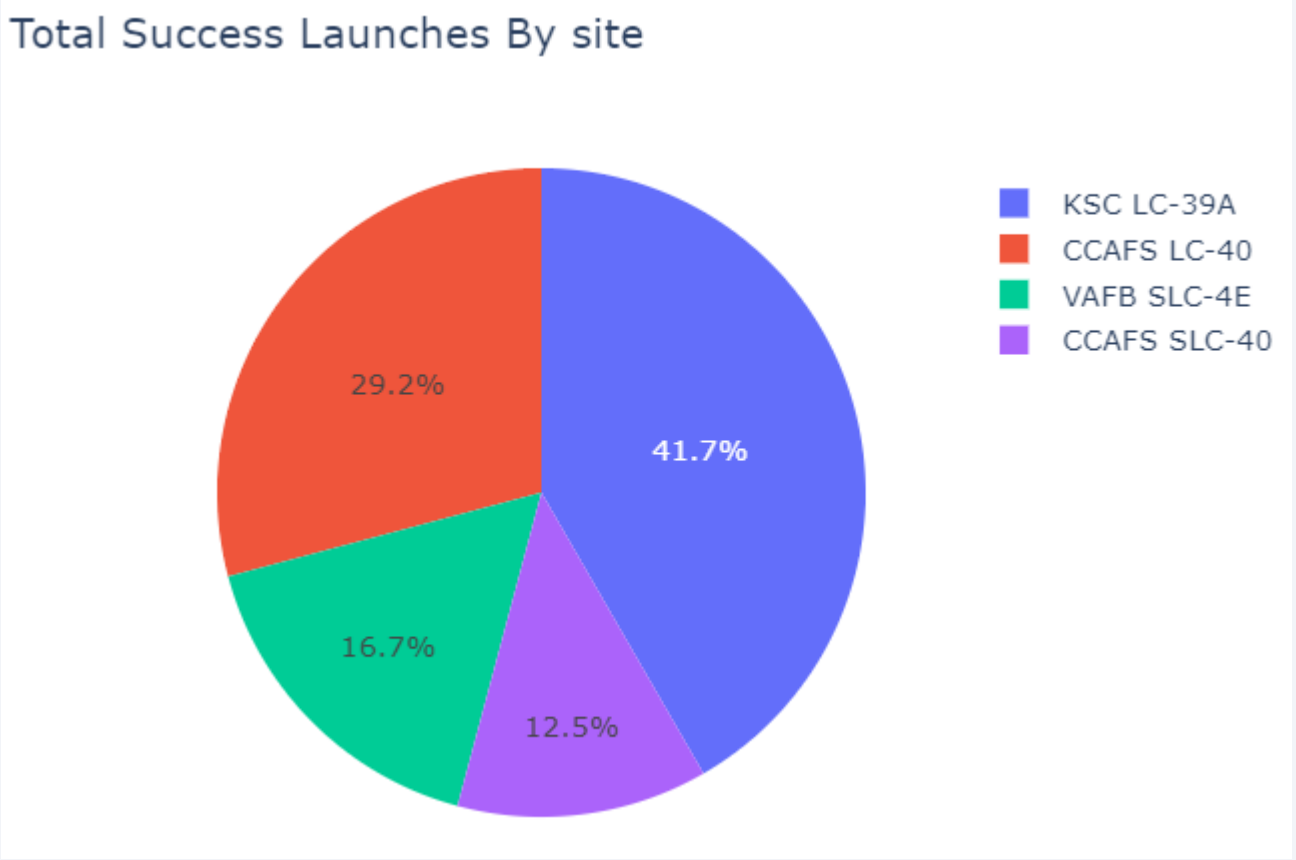



Section 4
Build a Dashboard with Plotly Dash

Total success launch ratio for all sites

Insights:

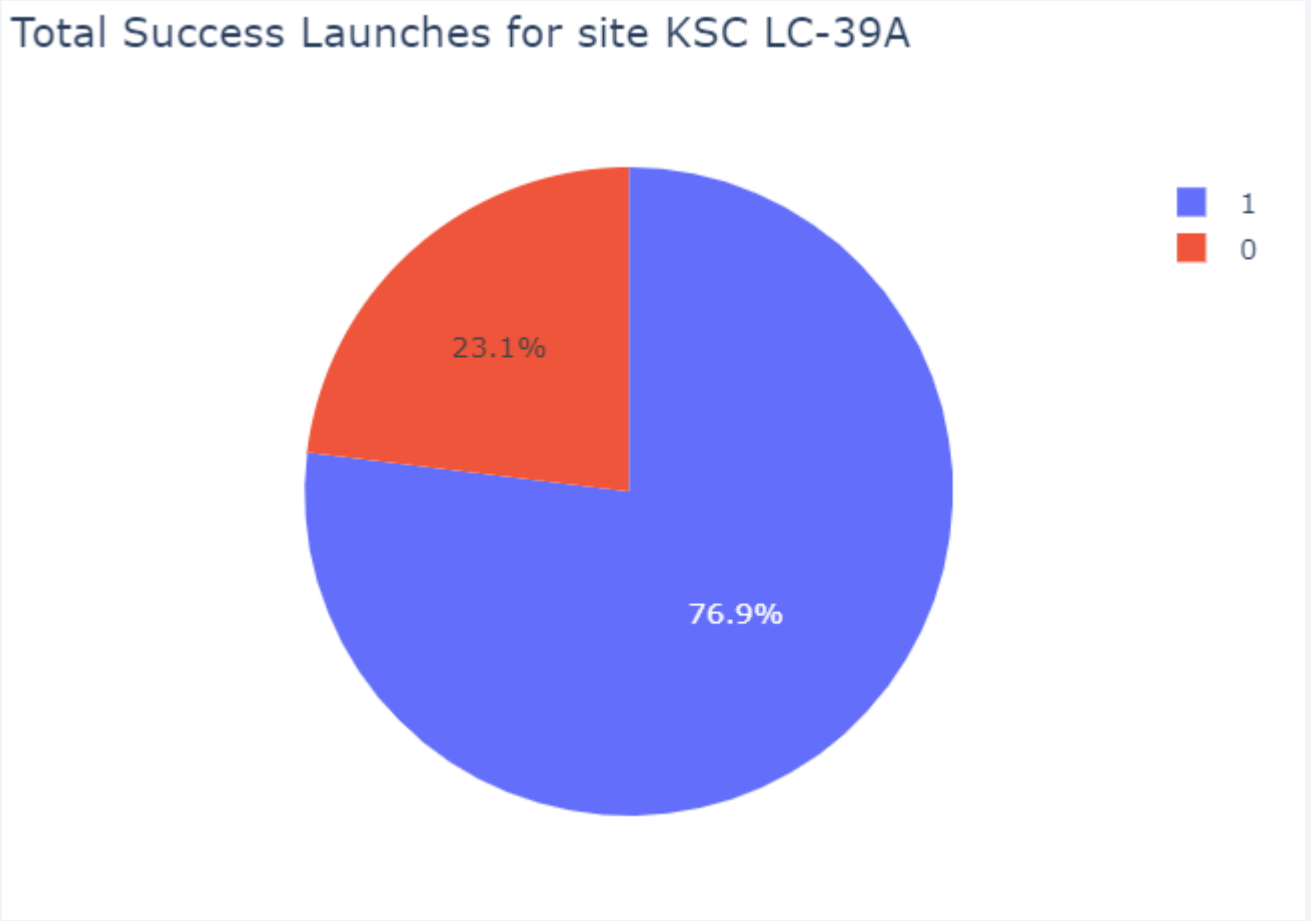
- KSC LC-39A has the highest launch success ratio
- CCAFS SLC-40 has the lowest launch success ratio



Total success launch ratio for KSC LC-39A

Insight:

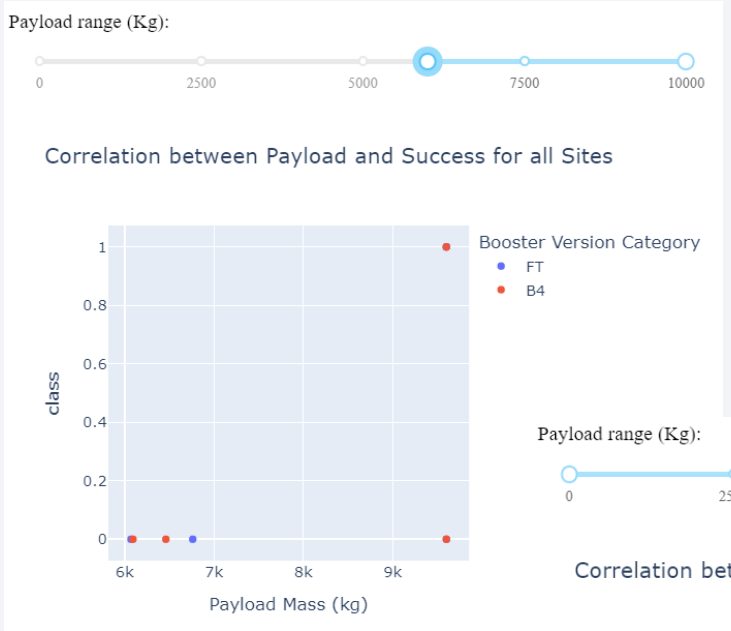
- 76.9% successful launches

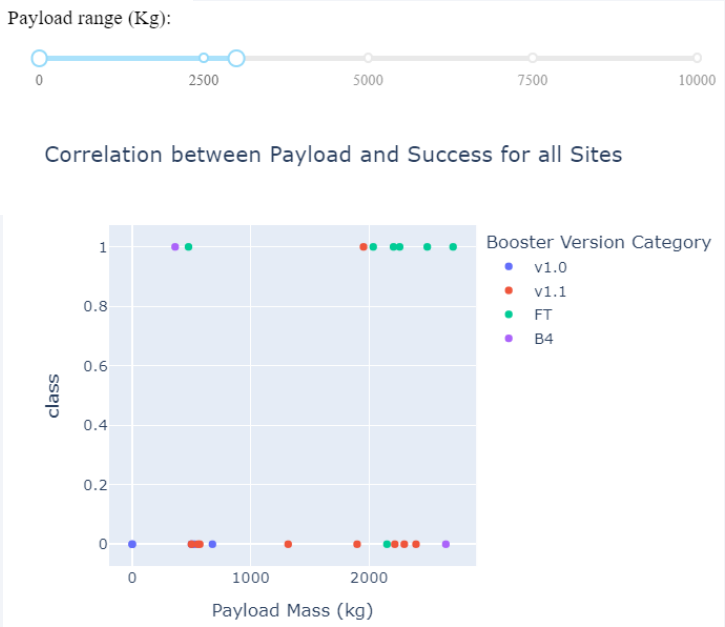


Payload vs. Launch Outcome for all Launch Sites

Insights:

- Booster version F9 B4 with payload mass 9600 has the largest launch success rate
- Booster version F9 B4 with payload mass 362 has the lowest launch success rate



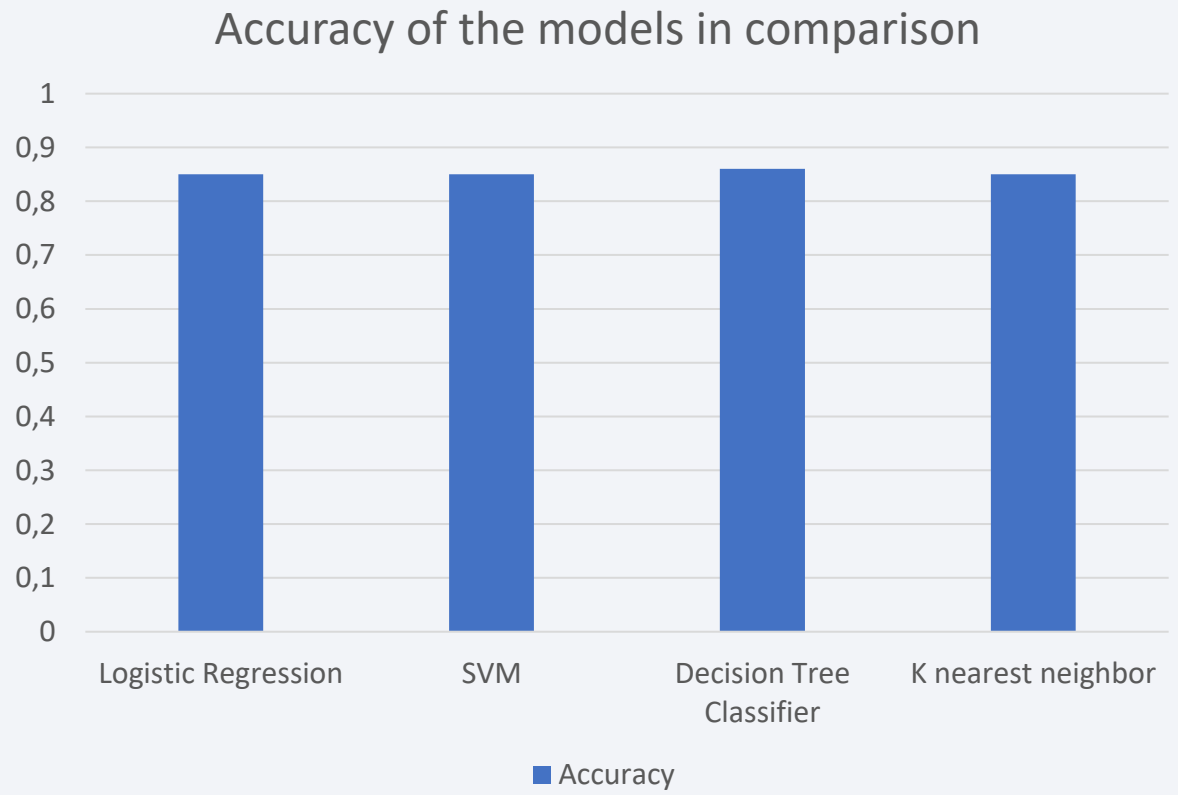


Section 5

Predictive Analysis (Classification)

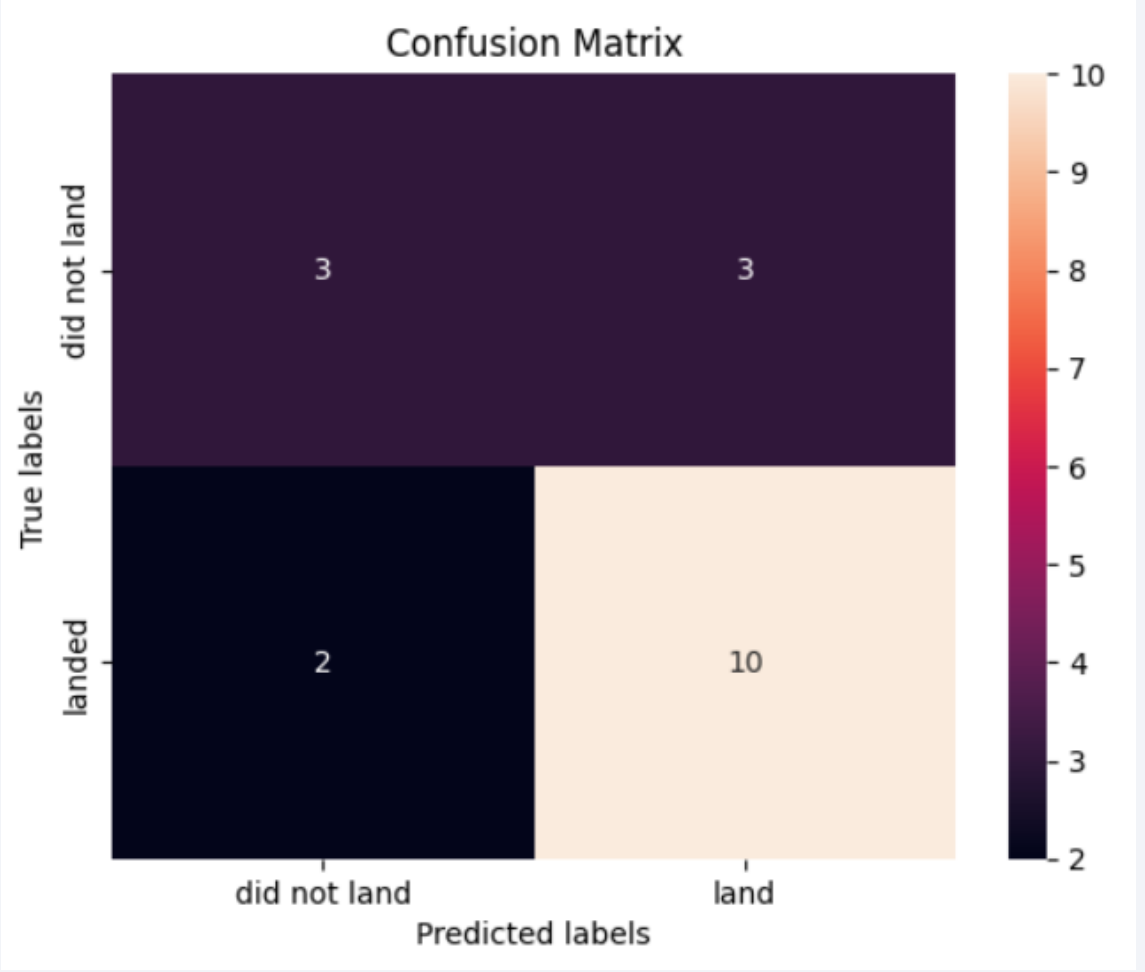
Classification Accuracy

- The decision tree classifier model has the highest accuracy



Confusion Matrix of the best performing model

- The decision tree classifier performed best with an accuracy of 86%
- Some false positives and false negatives are given



Conclusions

Will Space X reuse the first stage for a given launch?

- In 66% of the cases
- Factors that hint at a reuse:
  - Orbit Type: ES-L1, SSO, HEO, GEO
  - Higher Flight Number
  - Less massive Payload Mass
  - Launch Site: KSC LC-39A
- Best model for prediction: Decision Tree Classifier

Appendix

Relevant assets are included in the GitHub repository: GitHub Repo Link

Possible follow up research can look into:

- Why has the Launch Site KSC LC-39A the most successful launches?
- Why does Space X have little success rate for the orbit types GTO, PO, ISS, LEO, MEO?

Thank you!