IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Nicolas
February 2023

Outline

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

Executive Summary

- Summary of methodologies:
  - Data collection using SpaceX API and web scraping
  - Exploratory Data exploration (EDA) with Pandas and Folium
  - Machine Learning prediction to select the best prediction
- Summary of all results
  - Public Data Collection is possible
  - EDA gives us which features will allow us to have the best prediction of the launch outcome
  - Machine Learning Prediction gives us the best model to predict the outcome based on the previous features identified.

Introduction

- The objective is to determine if SpaceY can compete with SpaceX in term of revenue and success
- Problems you want to find answers:
  - Determine the cost for launches by predicting if we can reuse the first stage rocket from previous launches
  - Where is the best places to make launches

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data from Space X was available from 2 sources:
    - SpaceX API
    - Wikipedia Web Scraping
- Perform data wrangling
  - Collected Data has been enhanced by adding the feature launch_outcome based on other features available in the table.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - The data has been normalized and separated into training and test datasets. They have been evaluated by four different models through the accuracy of their results.

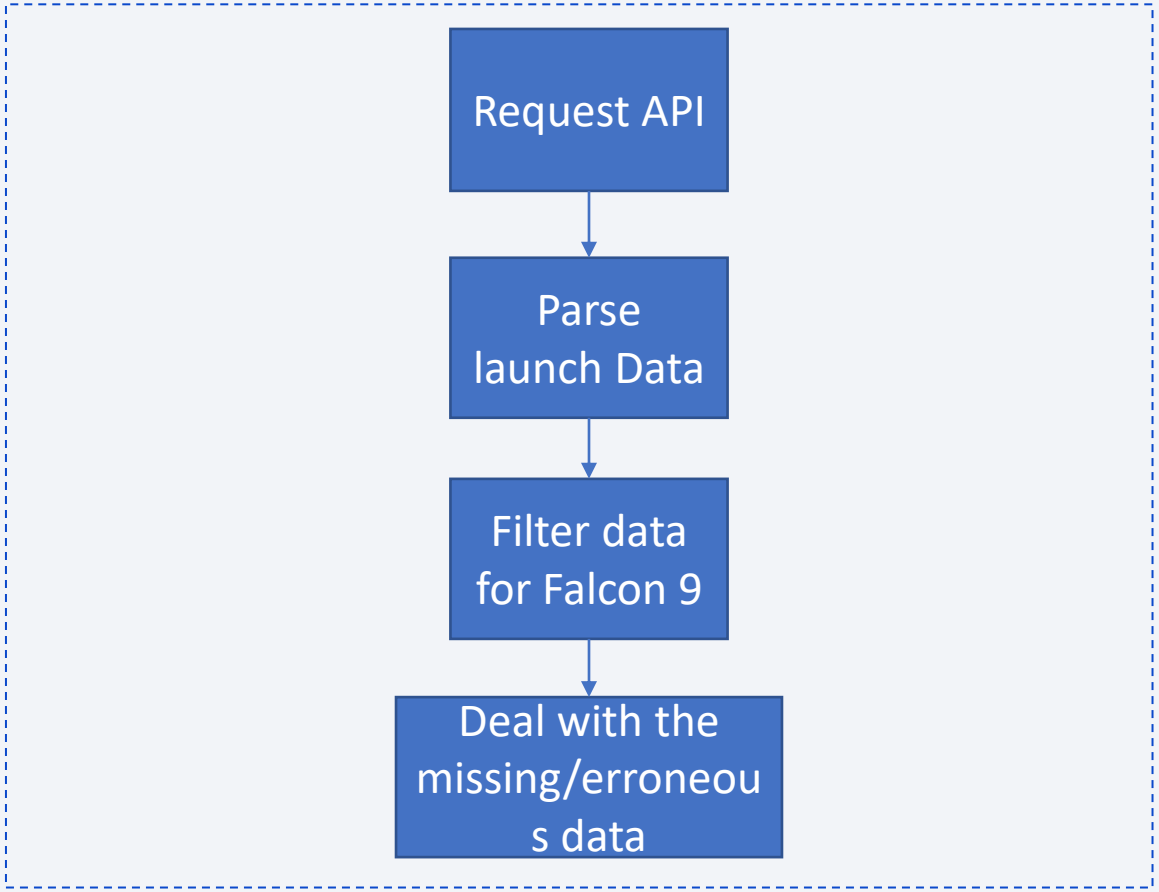
Data Collection

- Data sets were collected with this two techniques:
  - Space X API (https://api.spacexdata.com/v4/rockets/)
  - Wikipedia (Web scraping)

Data Collection – SpaceX API

- SpaceX has a public API serving data can be used in various way.
- The data is acquired through an HMTL request then filter and analyzed for missing/erroneous data

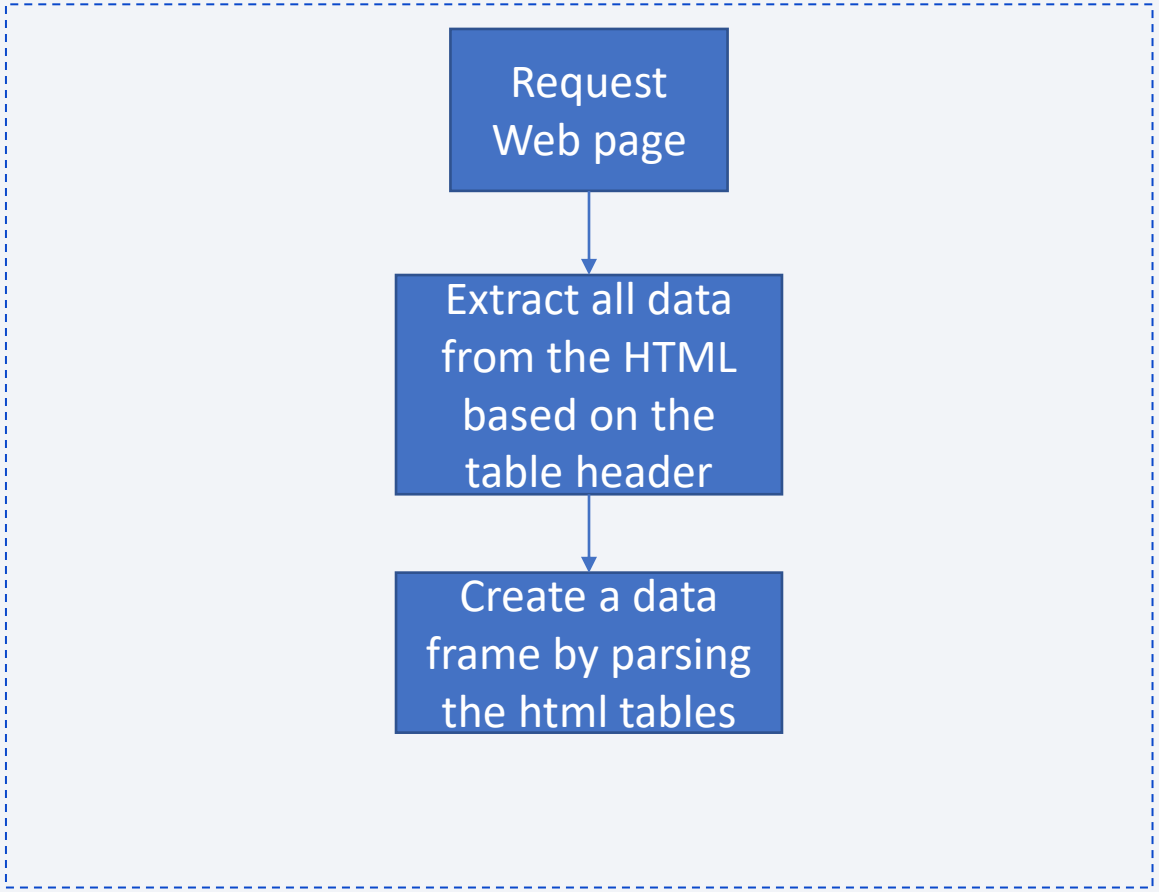
- Source Code:
  https://github.com/NicklerAWD/Coursera/blob/56f626e073a468d351dd68ee41c79d8510d36f90/CapStone/SpaceX_API.ipynb



Data Collection - Scraping

- Data is also available on Wikipedia
- Data is requested through a web call then scraped out of the response.

- Source Code
  https://github.com/NicklerAWD/Coursera/blob/56f626e073a468d351dd68ee41c79d8510d36f90/CapStone/SpaceX-webscraping.ipynb



Data Wrangling

- Some early analysis was performed on the dataset for searching some interesting features
- Some intermediate features has been calculated for improving the data analysis:
  - Launches per site
  - Occurrence of each orbits
  - Mission outcomes per orbit type
- At the end, the feature landing_outcome was created from the outcome column
- Source Code:
  https://github.com/NicklerAWD/Coursera/blob/56f626e073a468d351dd68ee41c79d8510d36f90/CapStone/Data_wrangling.ipynb

EDA with Data Visualization

- To quickly explore and find relationship between features, we can use scatterplots and barplots. For example: **Flight Number and Launch Site**
- Source Code
  https://github.com/NicklerAWD/Coursera/blob/56f626e073a468d351dd68ee41c79d8510d36f90/CapStone/EDA.ipynb

EDA with SQL

- The SQL queries you performed during the lab:
  - 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 acheived.
  - 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
- Source Code https://github.com/NicklerAWD/Coursera/blob/56f626e073a468d351dd68ee41c79d8510d36f90/CapStone/SQL_request.ipynb

Build an Interactive Map with Folium

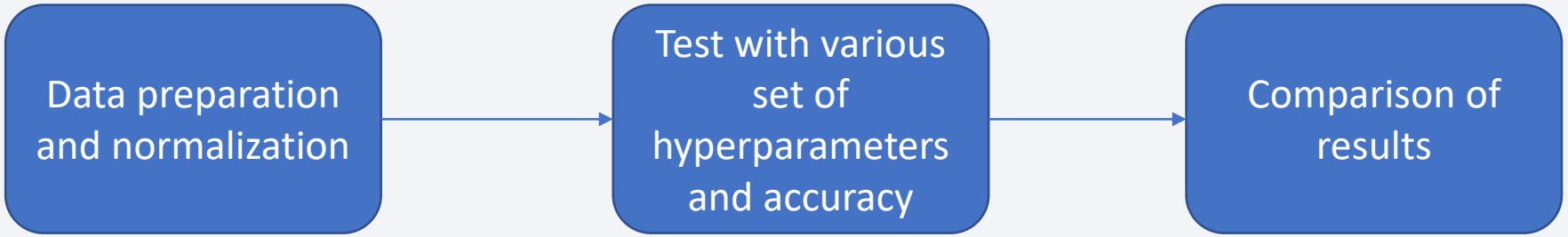
- Markers, circles, lines and marker cluster were used for marking several parameters:
  - Marker locates point of interests like launch sites
  - Circles indicates areas around specific coordinates
  - Marker cluster indicates groups of events in specific coordinate, like successful launches
  - Lines can be used to show distance between two coordinates
- Source Code <Folium>

Build a Dashboard with Plotly Dash

- Those graphs and plots has been used to visualize those parameters:
  - Percentage of launches by site
  - Payload range
- This combination helps us to:
  - Analyze the relation between payload and launch sites
  - Identify the best place to launch based on the payload.
- Source code
  https://github.com/NicklerAWD/Coursera/blob/56f626e073a468d351dd68ee41c79d8510d36f90/CapStone/folium.ipynb

Predictive Analysis (Classification)

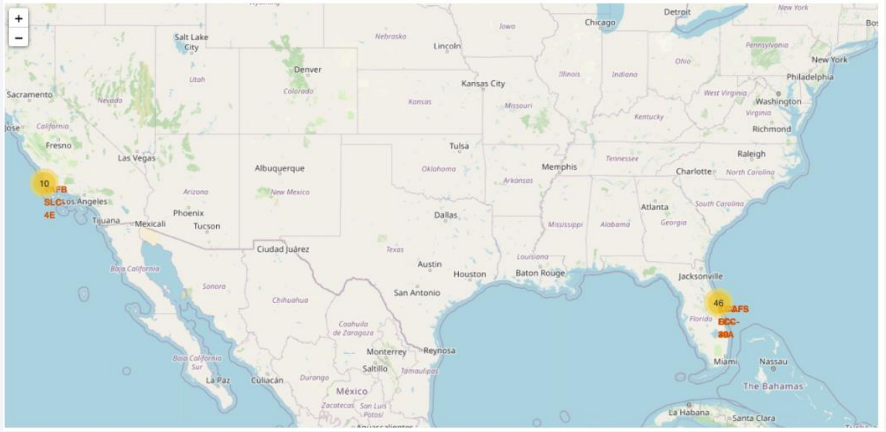
- We built 4 different classification models: Logistic regression, SVM, Decision tree and K-nearest neighbors. We also calculate their accuracy and compared to each other.

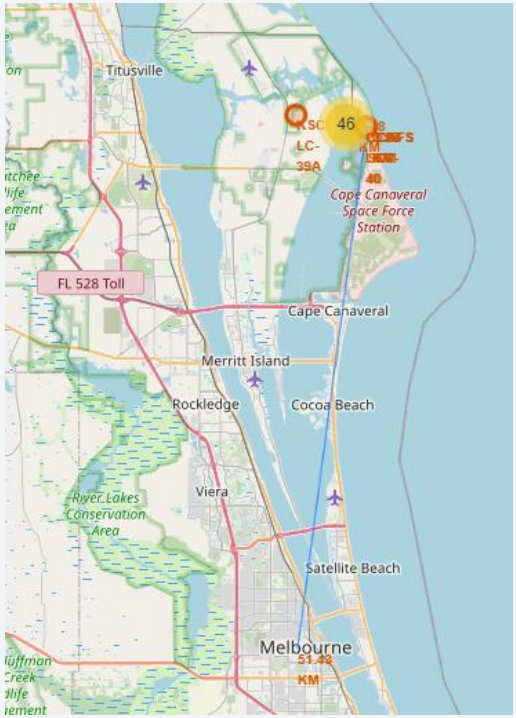


- Source Code:
  https://github.com/NicklerAWD/Coursera/blob/56f626e073a468d351dd68ee41c79d8510d36f90/CapStone/predictive.ipynb

Results

- Exploratory data analysis results
  - SpaceX uses 4 launch sites
  - The average payload of F9 is 2928kg
  - The first success landing outcome happened in 2015, five years after the first launch
- Through interactive analytics, we identify that launch sites are far from city centers, near the sea and have various logistic infrastructures.

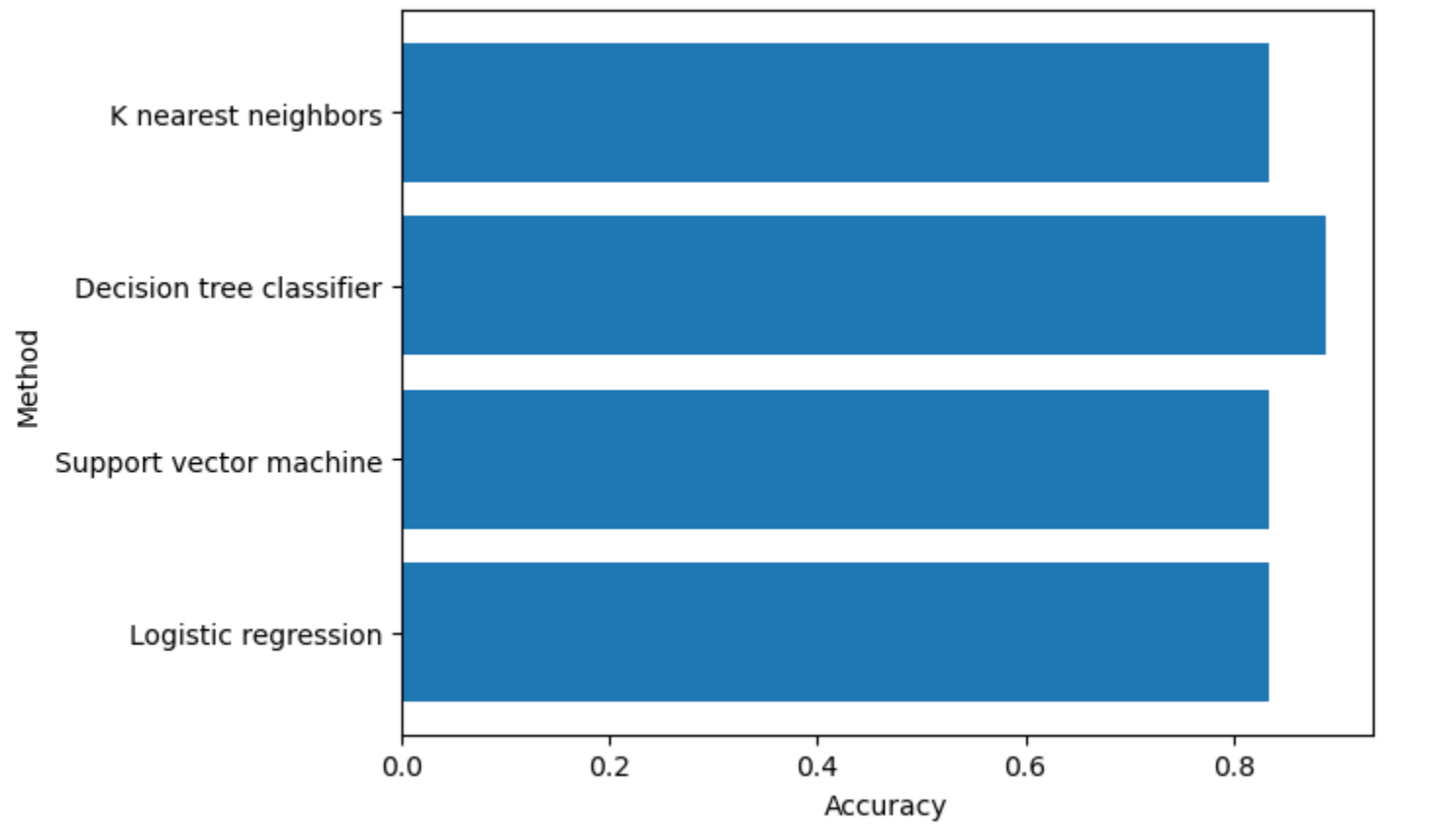


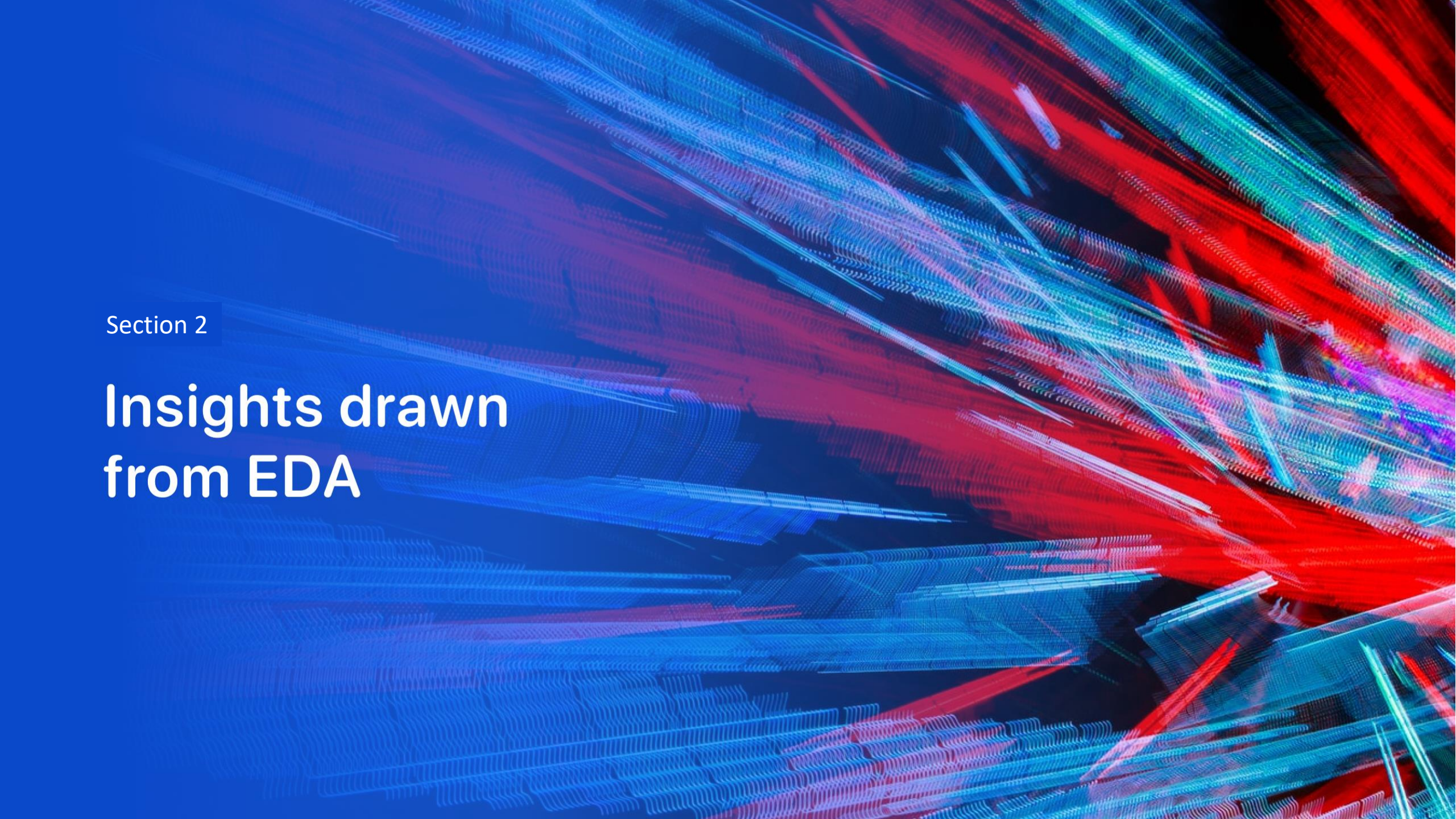


- Predictive analysis results

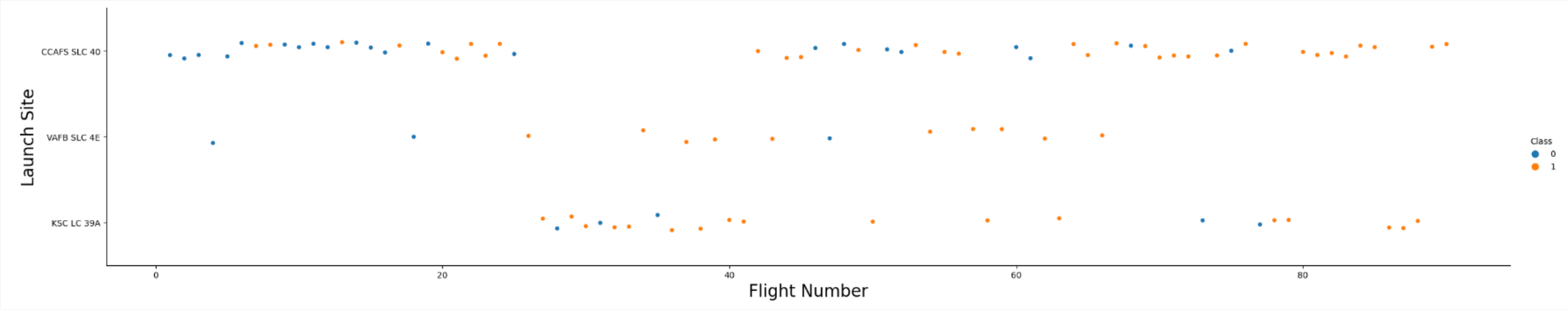
Results

- Through the predictive analysis, we demonstrate that the decision tree classifier is the best model to predict successful landing. The classifier has a 10% better accuracy than the other classifier.



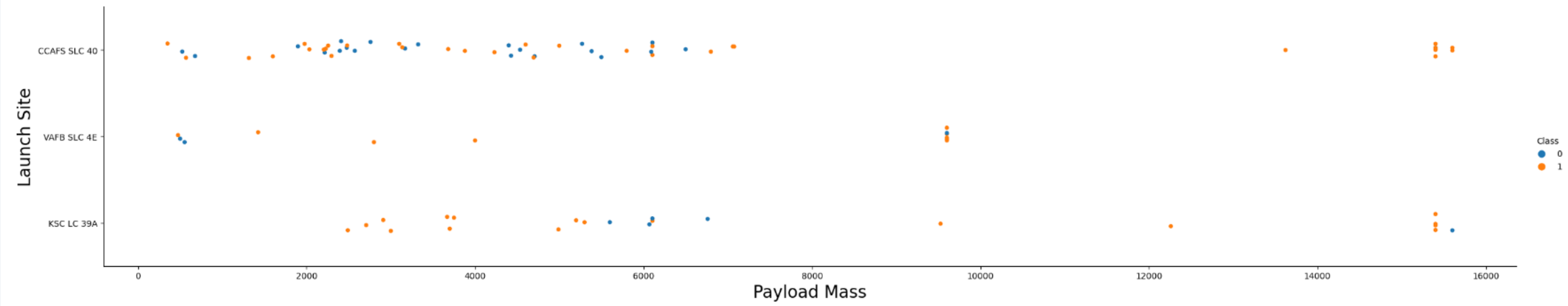

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- The plot above shows us the CCAPS SLC 40 is the best launch site overall with the most successful launches.
- Also, we can see the CCAPS SLC 40 is the preferred launch site for SpaceX
- We can conclude the success rate across all the sites improved over time.

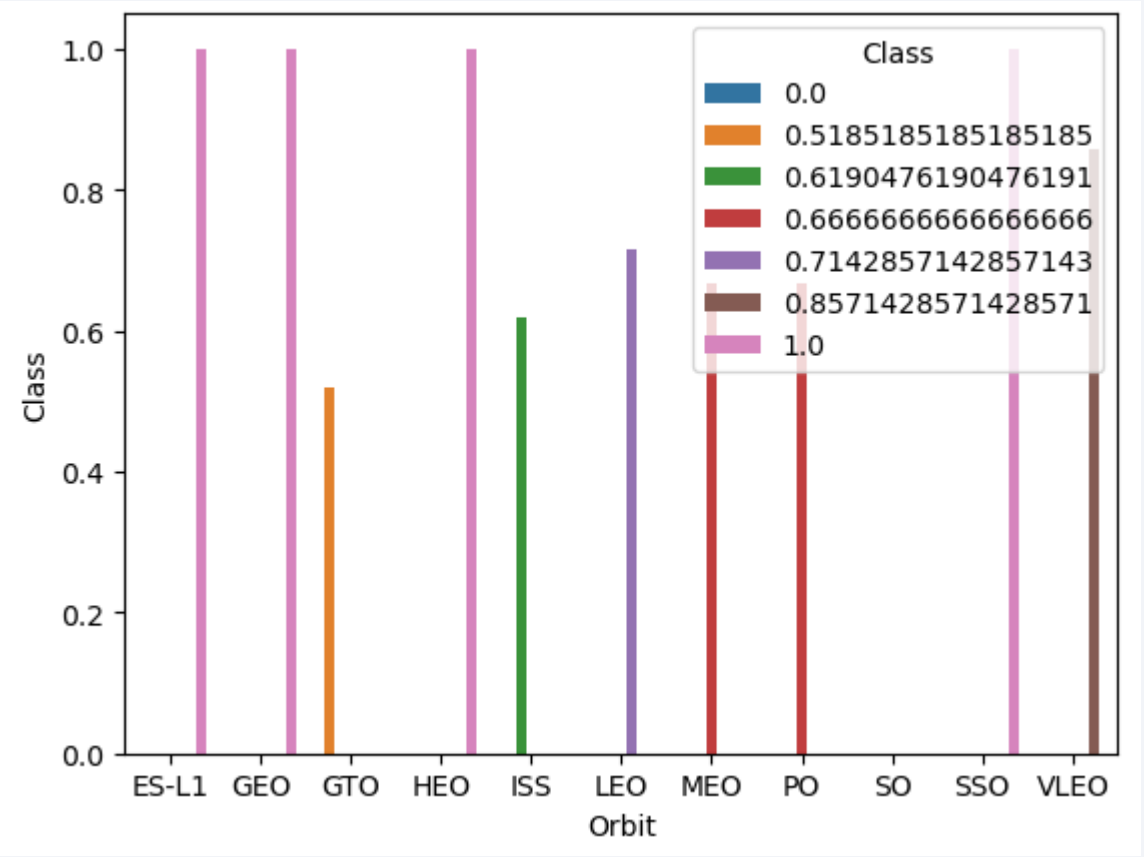
Payload vs. Launch Site



- Any payloads over 9000kg have a good success rate.
- VSFB SLC 4E doesn't have any payload over 9000kg
- Payloads over 12000kg only launch from CCAFS SLC 40 and KSC LC 39A with a near perfect success rate

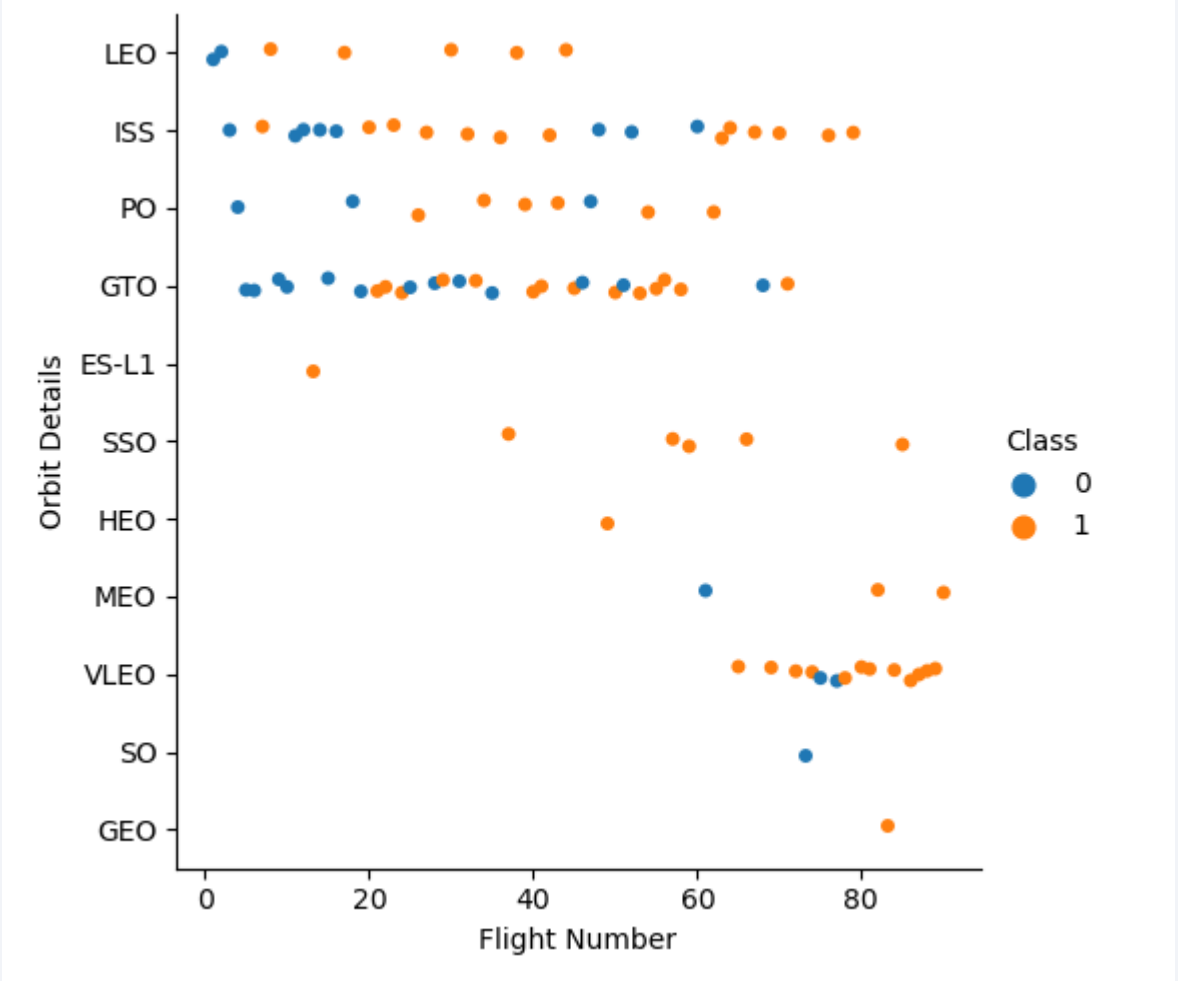
Success Rate vs. Orbit Type

- The biggest success rate are for the orbits: ES-L1, GEO, HEO and SSO
- The lowest success rate are SO (or no mission targeting this orbit)



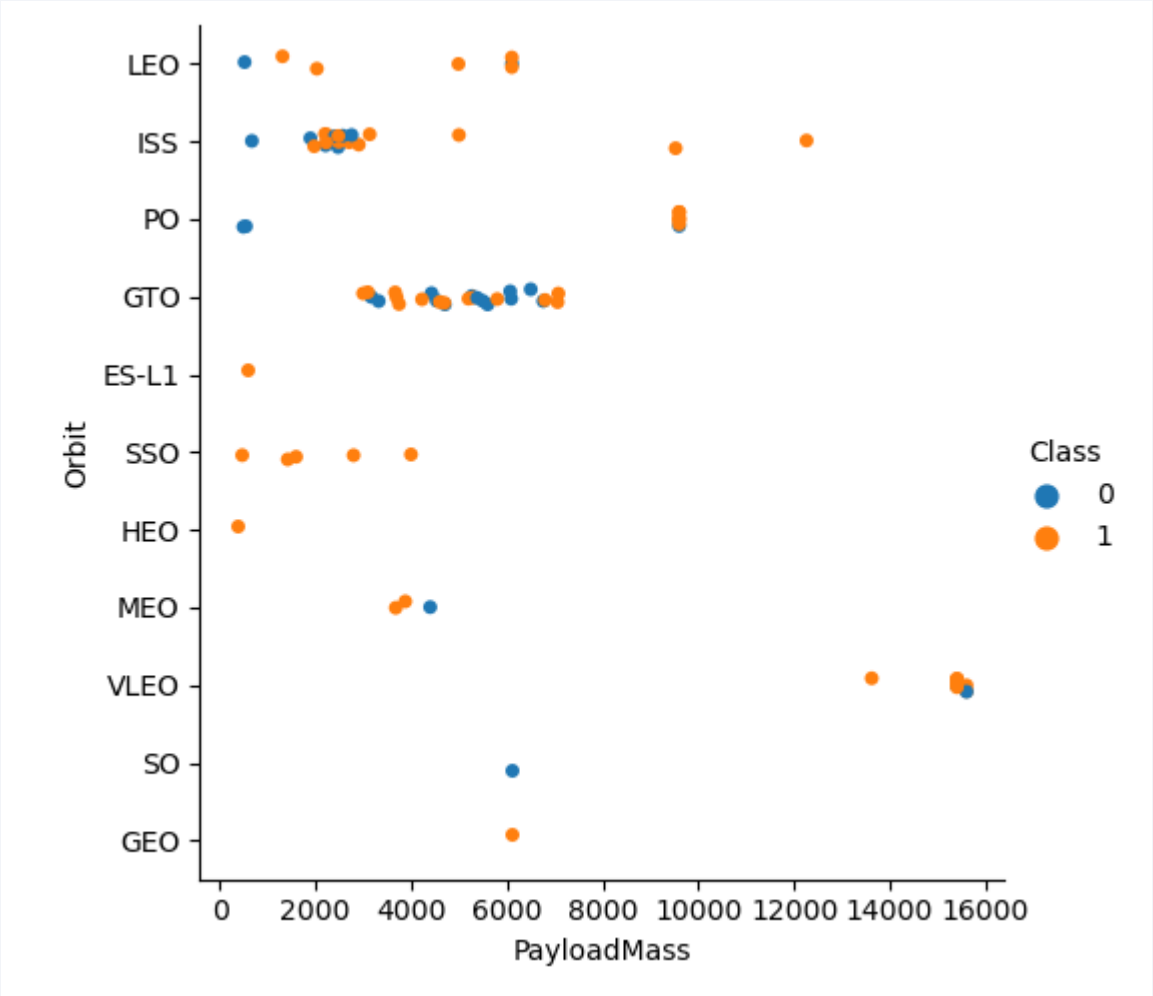
Flight Number vs. Orbit Type

- The success rate increased overtime.
- VLEO presents an increase in frequency to the point to be the only target orbit. (new business opportunity?)



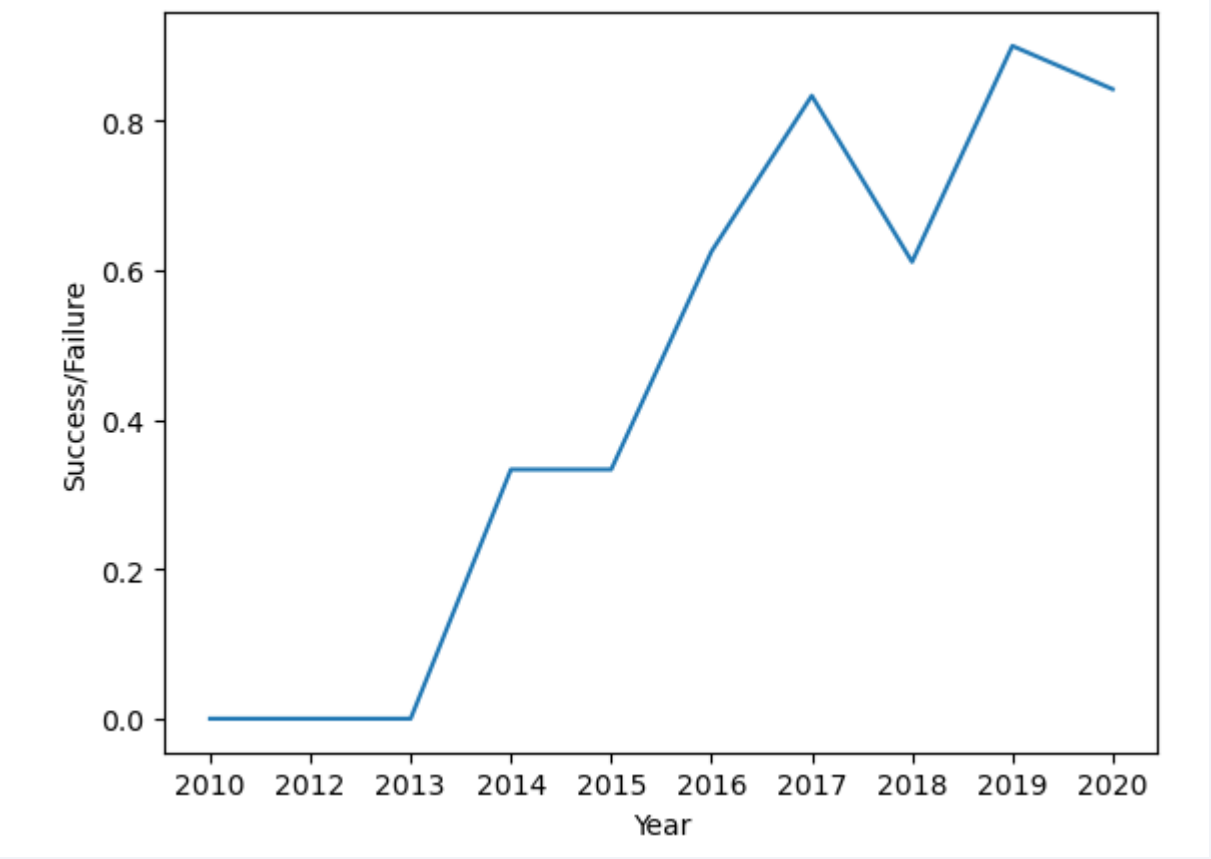
Payload vs. Orbit Type

- The VLEO has the heaviest payload, and the HEO has the lightest. We can see the orbit type can limit the maximum payload to be taken
- ISS orbit has the widest range and a good success rate.



Launch Success Yearly Trend

- The success rate since 2013 kept increasing till 2020
- Based on SpaceX history, the first three years on this graph shows the period of technology improvement and adjustements



All Launch Site Names

- Here the four launch site names:

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

- We are selecting the unique occurrence of the launch_site column from the database

Launch Site Names Begin with 'CCA'

- Here the 5 records from the launch sites beginning with ‘CCA’:

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

- We query the database with the condition the launch site name start with ‘CCA’ and limit the results list to only the first 5.

Total Payload Mass

- Total payload carried by boosters from NASA is: 45596kg
- The total payload is calculated by summing all the payload whose customer codes contains ‘NASA (CRS)’

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928kg
- This result is obtained by filtering the data by booster version and calculating the average payload mass.

First Successful Ground Landing Date

- The first successful landing outcome on ground pad was on the 22/12/2015
- We obtained this date by filtering the data by successful landing on the ground pad and looking for the minimum value for date.

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are shown in the table booster_version.
- We obtained this result by selecting the distinct booster version after applying a filter on the payload mass value

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes is shown on the table below

failed	success
10	61

- For getting this result, we counted the mission outcome as success if it contained the word success and failed if the mission outcome contained the word failed then display in this table format.

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass are displayed in the table on the right
- First, we found the max payload in the dataset, then we get all the booster version which carried the max payload.

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

- The failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015 are shown in the table below:

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

- We obtained this two occurrence by filtering the data by:
  - date in 2015
  - the landing outcome was failed on the drone ship.

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

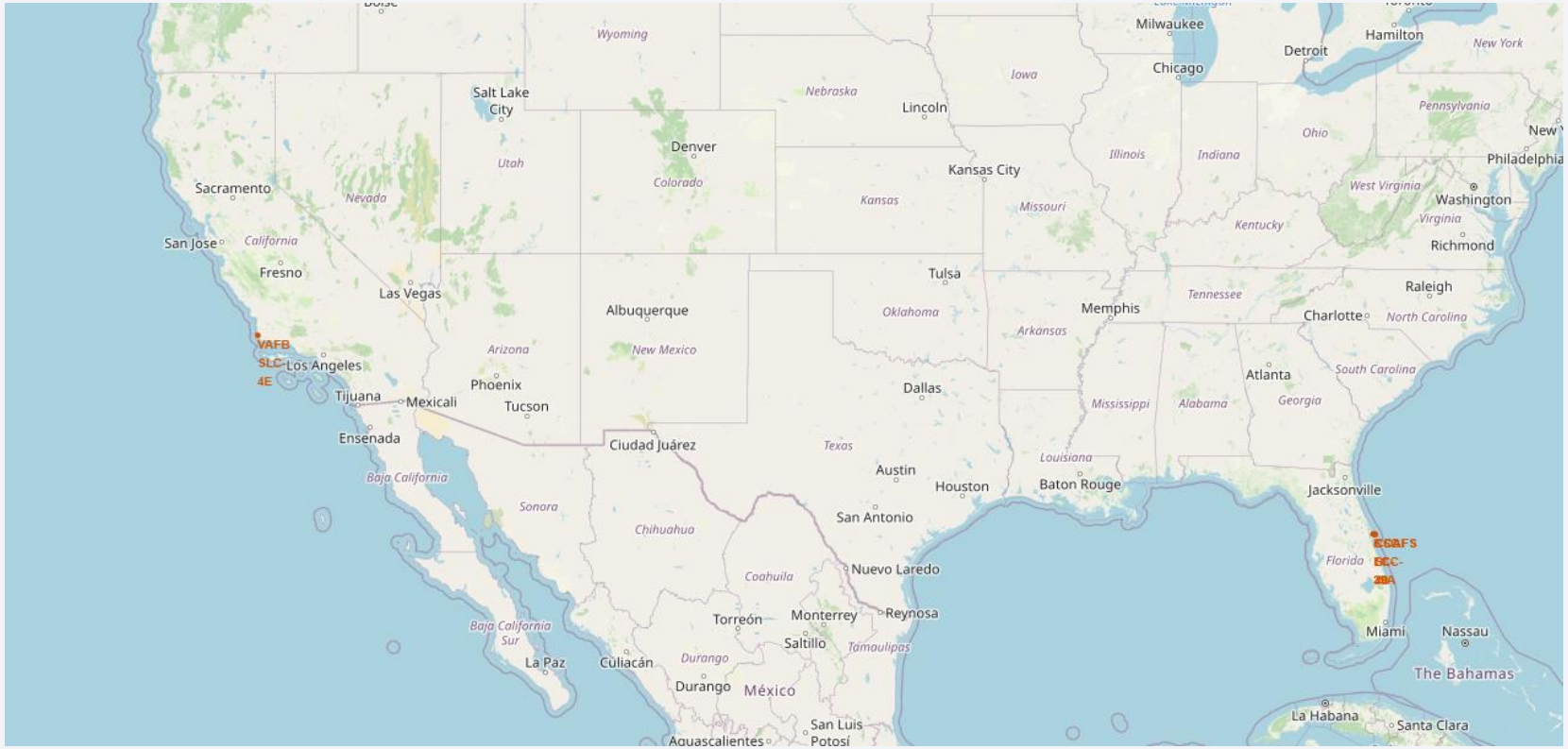
- The count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order are shown in the table
- The landing outcome shall be taken in account during the prediction and further analysis.

Landing outcome	occurence
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

Section 3

Launch Sites Proximities Analysis

All launch sites



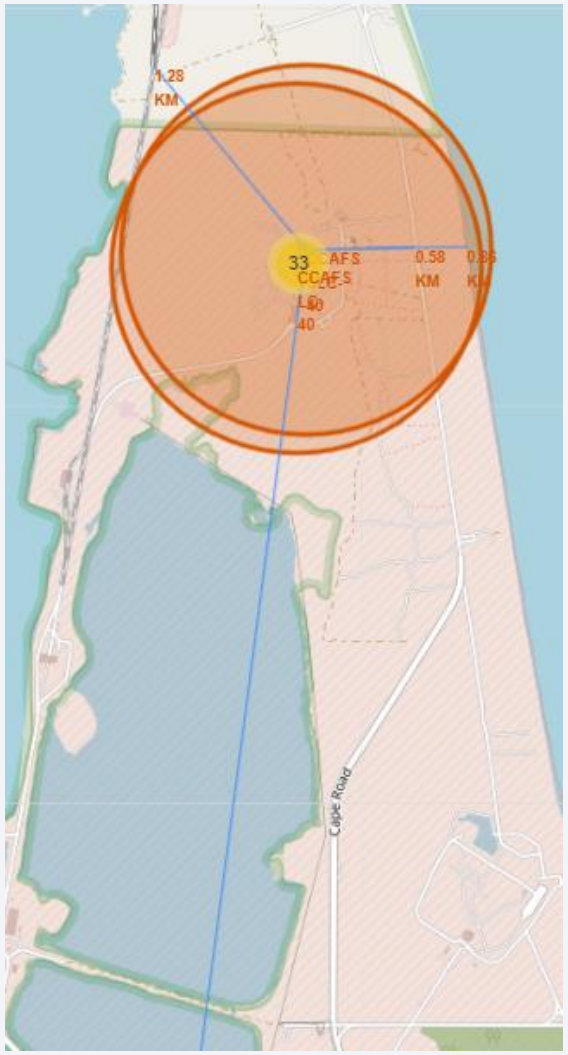
- The launch sites are located near the sea, with good logistic infrastructure like railroad.

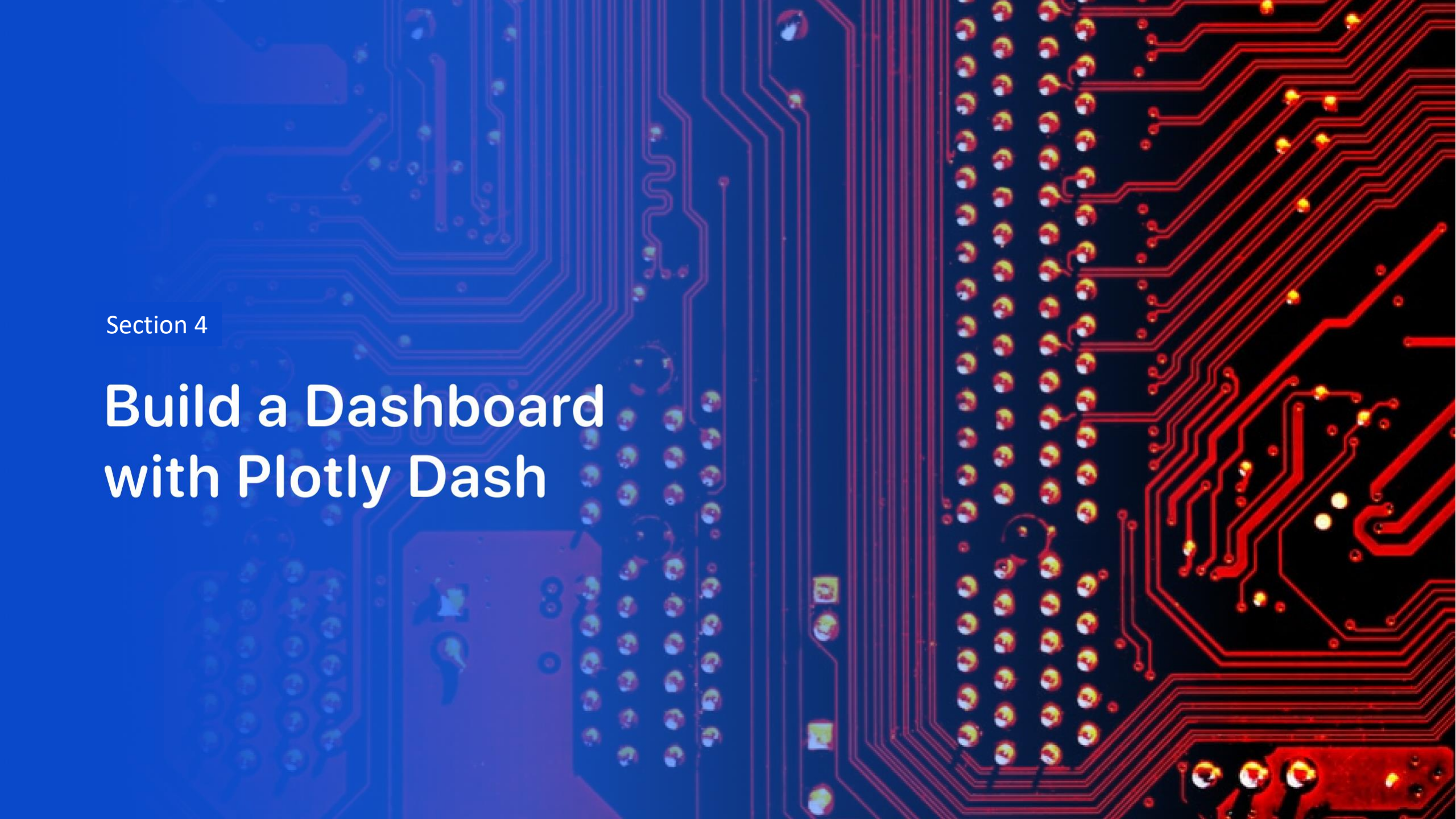
Launch outcomes by site

- The marker cluster is done by launch site
- Expanded, we get more information about the outcome by launch. Red shows failed outcomes and green successful ones.

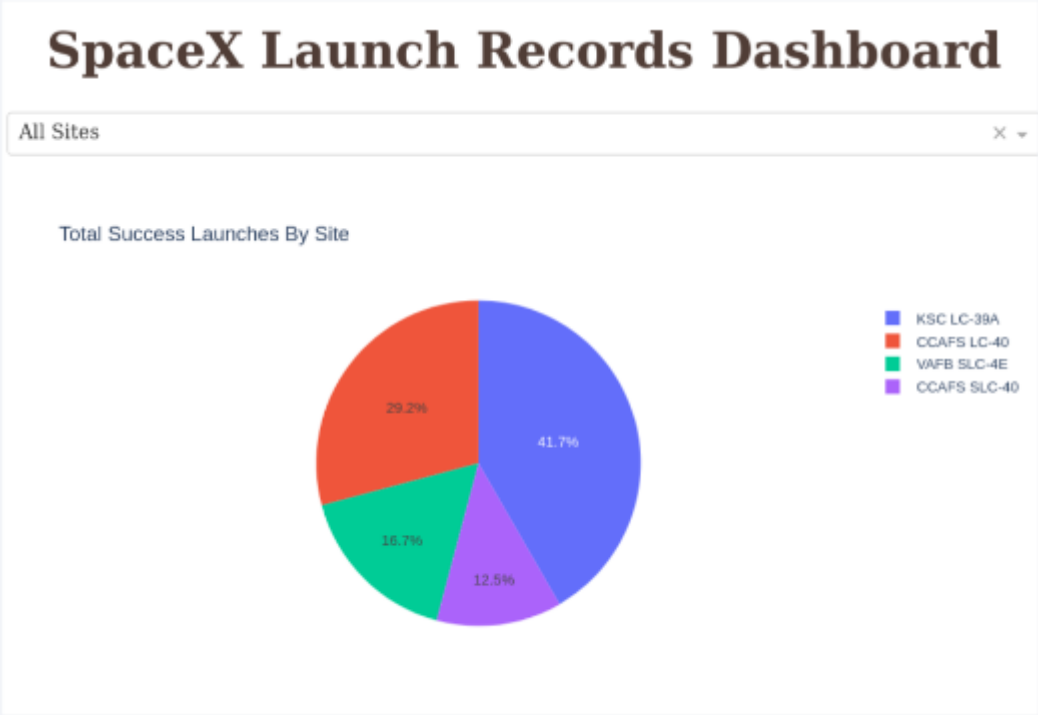
Logistic and Safety

- Launch site CCAFS-LC 40 has good logistic access like roads and railroad
- The site is also far from cities.



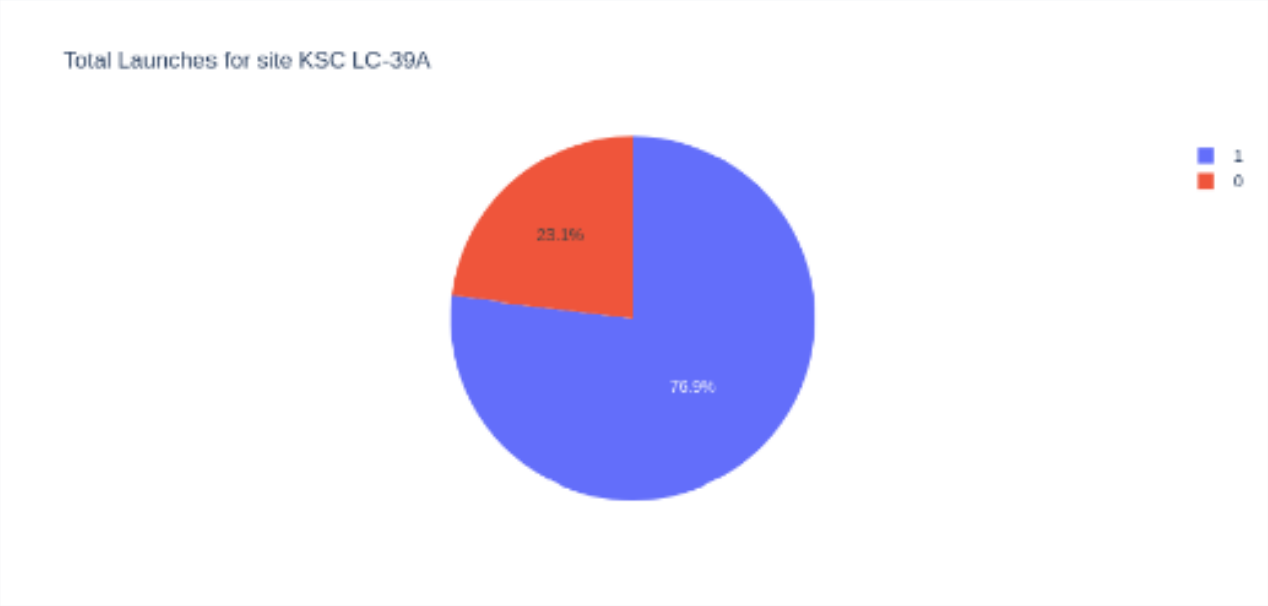

Section 4
Build a Dashboard with Plotly Dash

Successful Launches by site



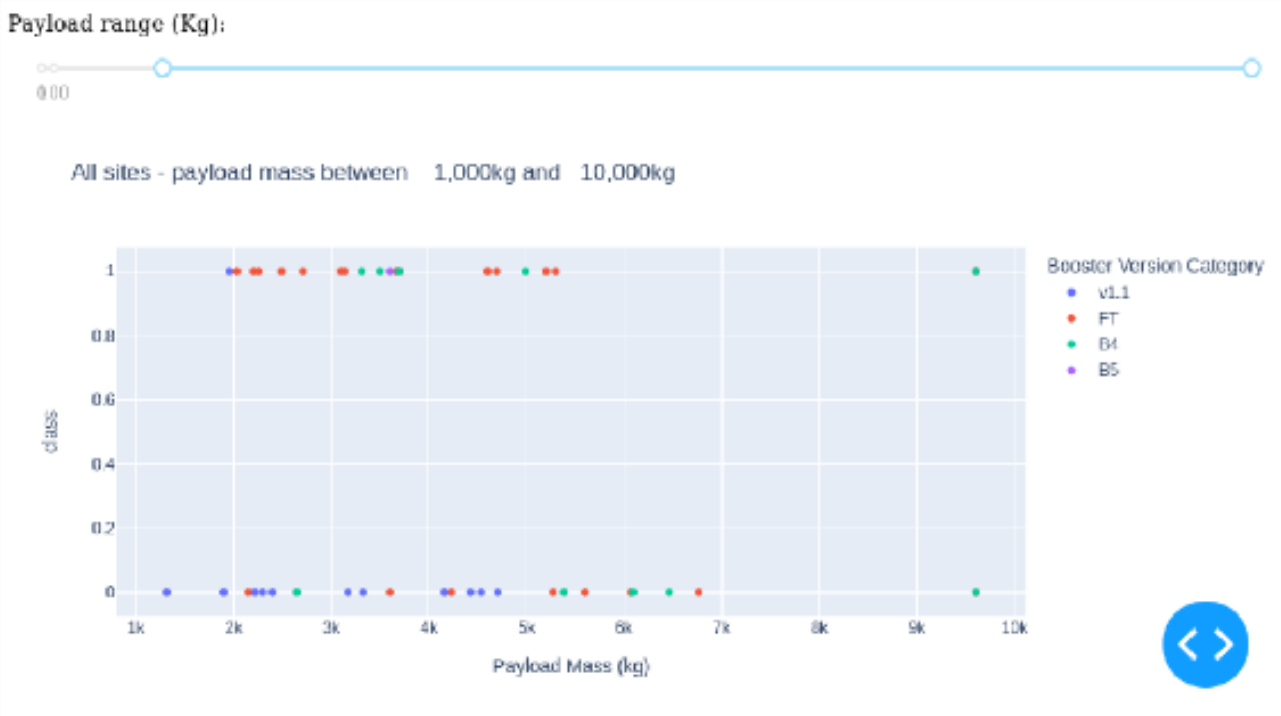
- The most successful site is the KSC LC-39A. We cannot determine which site is the most successful.

Launch success rate for KSC LC-39A



- This site has a success rate of 76.9%.

Payload vs Launch Outcome



- The FT Booster Version seems to have a good success rate when the payload is below 6000kg.

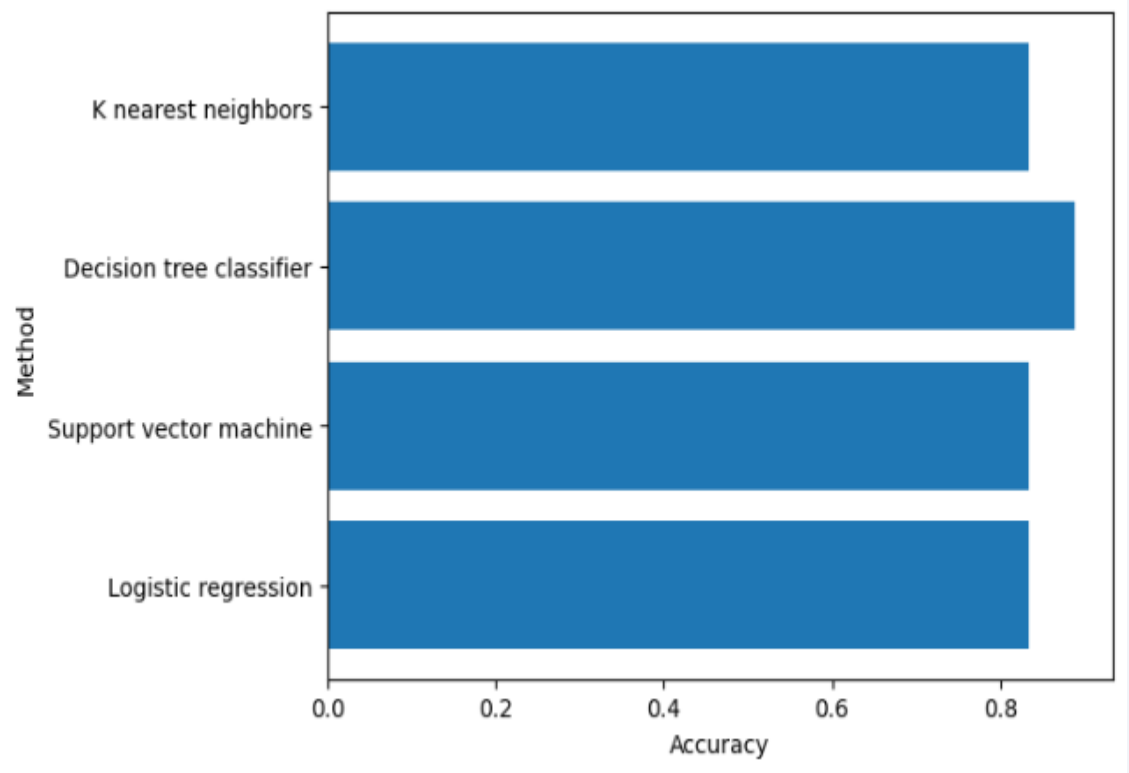
Section 5

Predictive Analysis (Classification)

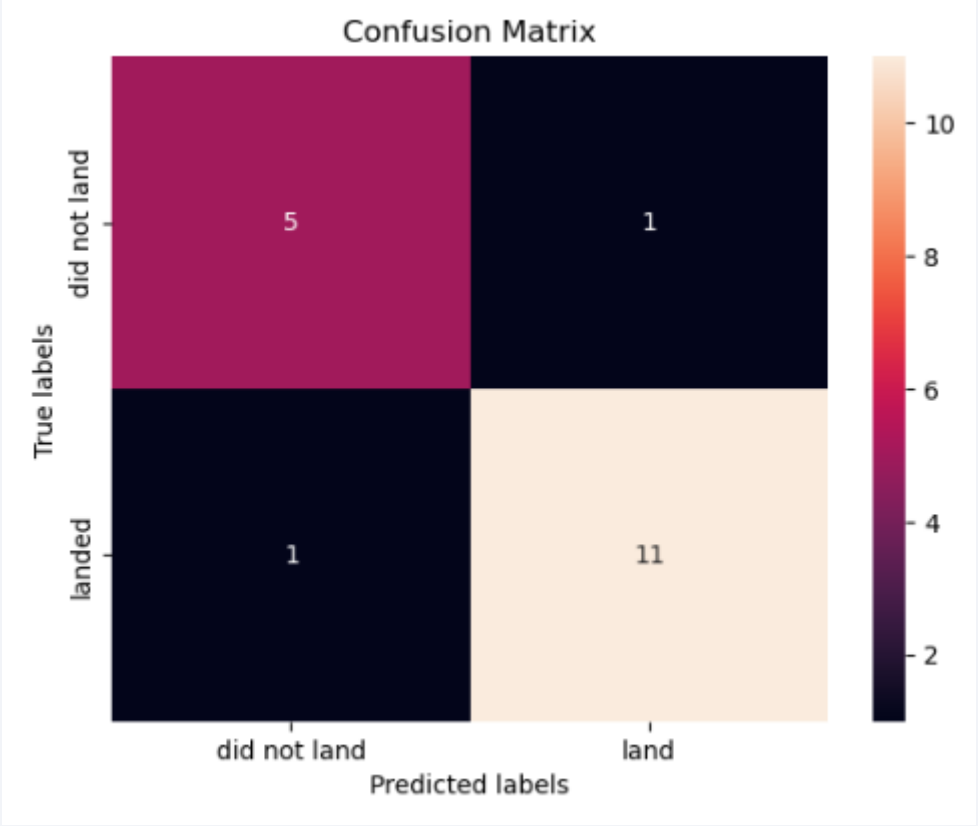
Classification Accuracy

- We tested 4 classification models and their accuracies has been plotted on the figure →

- The decision tree classifier is the model with the best accuracy (88.8%)



Confusion Matrix



Conclusions

- We analyzed different data sources for getting the best picture of Space X business
- We know the best launch site is: KSC LC-39A.
- The FT booster version is the most successful booster when the payload is below 7000kg
- We also noticed an improvement of the landing outcomes, which might suggest an evolution of rockets and processes refinements.
- Decision Tree Classifier can be used to refine the business model by predicting properly the outcomes.

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!