IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

HIEU NGUYEN
25 MAR 2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection using REST API, web scraping
  - Exploratory Data Analysis
    - Data wrangling
    - Data visualization
    - Interactive visual analytics
  - AI Machine Learning
- Summary of all results
  - Collect valuable data from internet
  - Enrich collected data
  - Predict key characteristics base on collected data

Introduction

- Project background and context: evaluate the opportunity for Space Y to against Space X
- Problems you want to find answers:
  - The optimal way to estimate the total cost for rocket launches
  - The optimal landside for launches

Section 1

Methodology

Methodology

Executive Summary

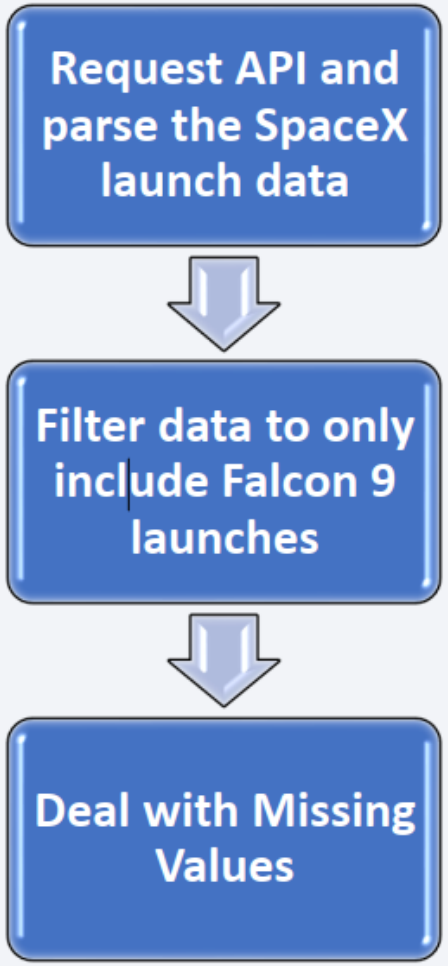
- Data collection methodology:
  - Space X API (https://api.spacexdata.com/v4/rockets/)
  - Web Crawling (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform data wrangling
  - Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features
- 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 that was collected until this step were normalized, divided in training and test data sets and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters

Data Collection

- Data sets were collected from
  - Space X API (https://api.spacexdata.com/v4/rockets/)
  - Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using web scraping technics.

Data Collection – SpaceX API

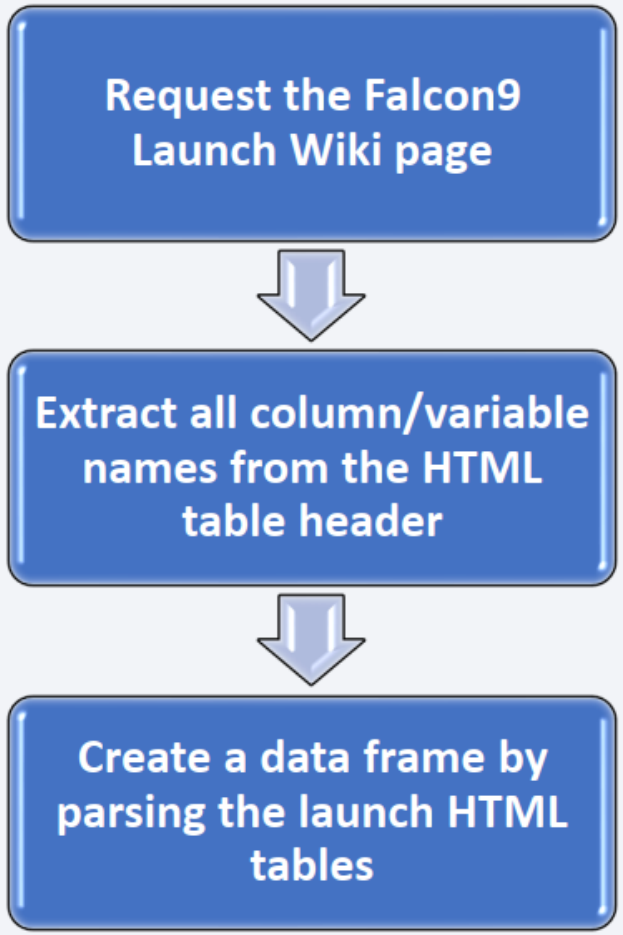
- SpaceX offers a public API from where data can be obtained and then used
- This API was used according to the flowchart beside and then data is persisted
- URL notebook: https://github.com/hieunt89/Applied-Data-Science-Capstone/blob/master/Data%20Collection%20API.ipynb



Data Collection - Scraping

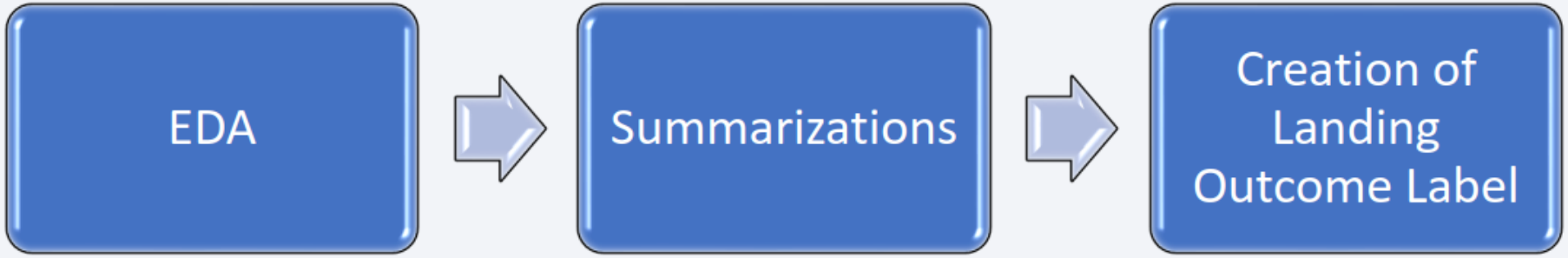
- Data from SpaceX launches can also be obtained from Wikipedia
- Data are downloaded from Wikipedia according to the flowchart and then persisted

- URL notebook: https://github.com/hieunt89/Applied-Data-Science-Capstone/blob/master/Data%20Collection%20API.ipynb



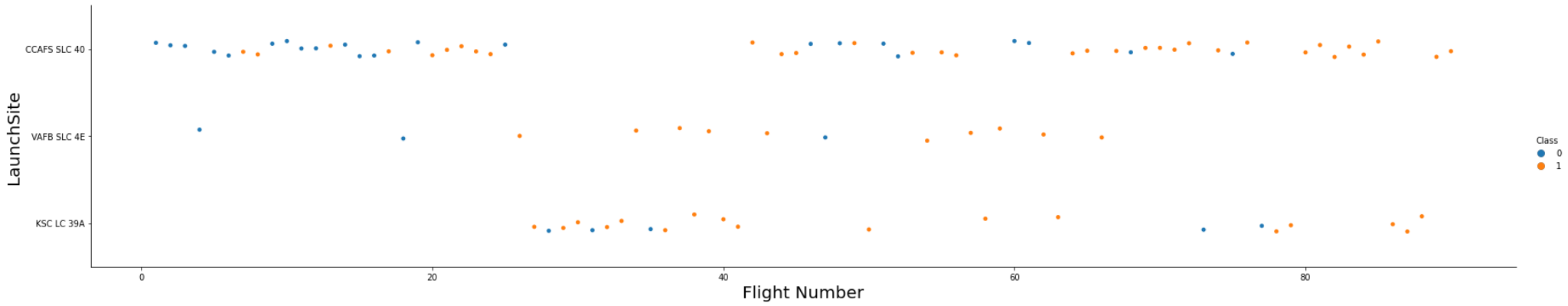
Data Wrangling

- Perform EDA on the dataset
- Summarize the launches per sites for each orbit with the outcomes.
- Adding outcome label
- Notebook URL: https://github.com/hieunt89/Applied-Data-Science-Capstone/blob/master/Data%20Wrangling.ipynb



EDA with Data Visualization

- Notebook URL: https://github.com/hieunt89/Applied-Data-Science-Capstone/blob/master/EDA%20with%20Data%20Visualization.ipynb



EDA with SQL

- Using bullet point format, summarize the SQL queries you 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 acheived.**
  - **List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000**

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
  - **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**
- URL notebook: https://github.com/hieunt89/Applied-Data-Science-Capstone/blob/master/EDA.ipynb

Build an Interactive Map with Folium

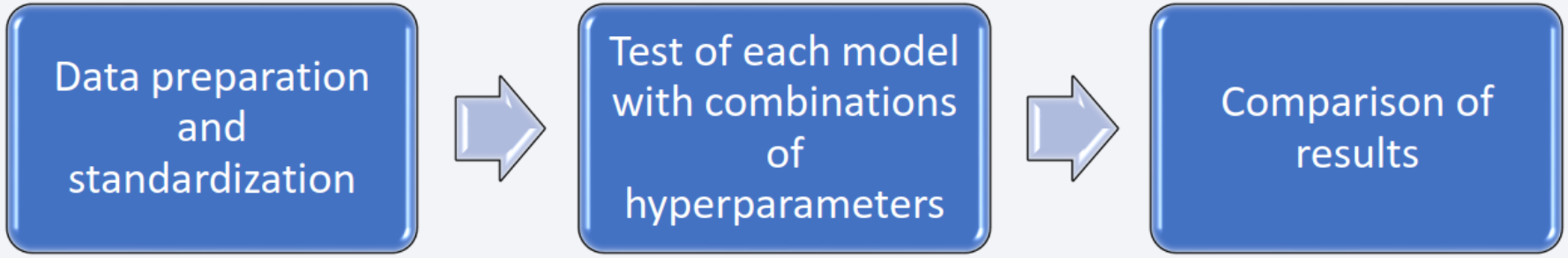
- Markers, circles, lines and marker clusters were used with Folium Maps
  - Markers indicate points like launch sites
  - Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center
  - Marker clusters indicates groups of events in each coordinate, like launches in a launch site
  - Lines are used to indicate distances between two coordinates.
- URL notebook: https://github.com/hieunt89/Applied-Data-Science-Capstone/blob/master/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize data
  - Percentage of launches by site
  - Payload range
- This combination allowed to quickly analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads
- URL notebook: https://github.com/hieunt89/Applied-Data-Science-Capstone/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

- Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbors
- URL notebook: https://github.com/hieunt89/Applied-Data-Science-Capstone/blob/master/Machine%20Learning%20Prediction.ipynb

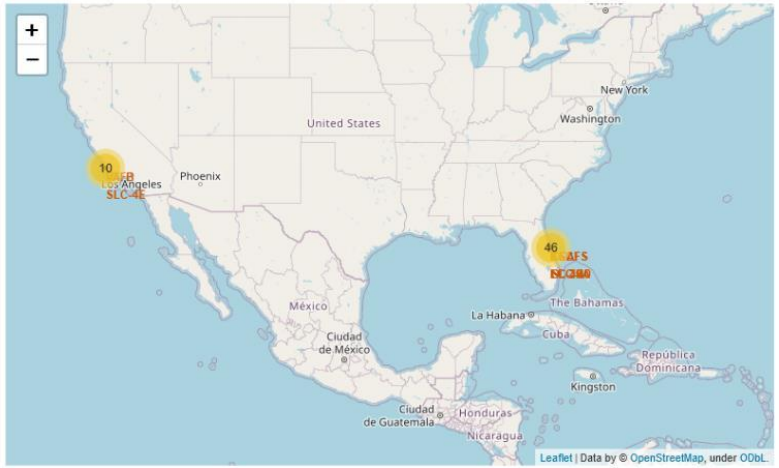


Results

- Exploratory data analysis results
  - Space X has 4 launch sites;
  - The first launches were done to Space X itself and NASA;
  - The average payload of F9 v1.1 booster is 2,928 kg;
  - The first success landing outcome happened in 2015 fiver year after the first launch;
  - Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;
  - Almost 100% of mission outcomes were successful;
  - Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;
  - The number of landing outcomes became as better as years passed.

Results

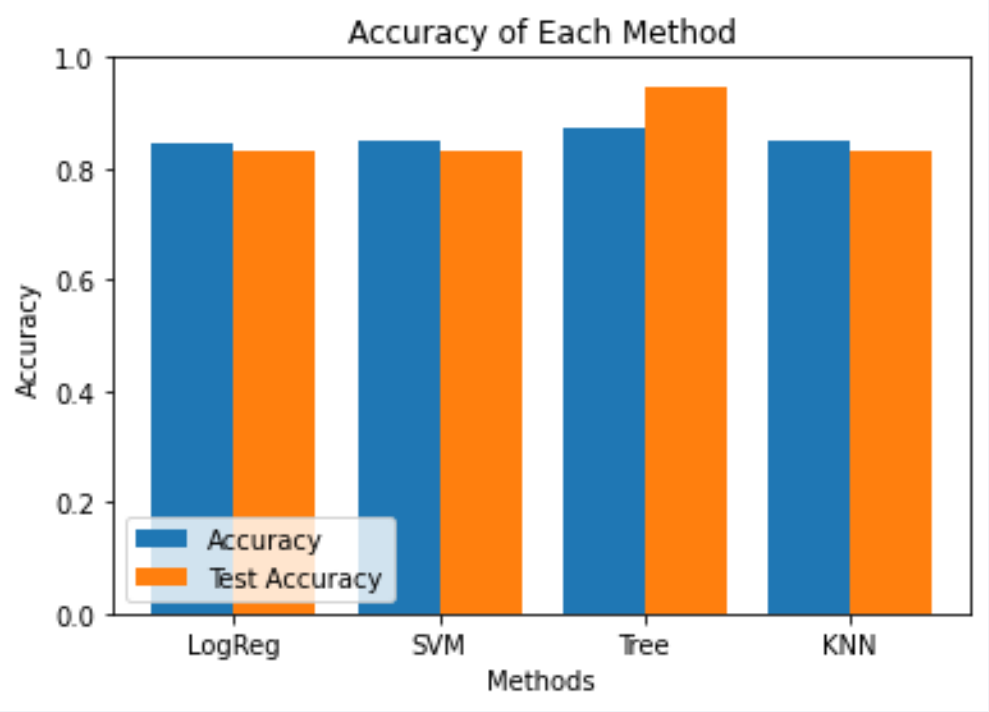
- Interactive analytics demo in screenshots

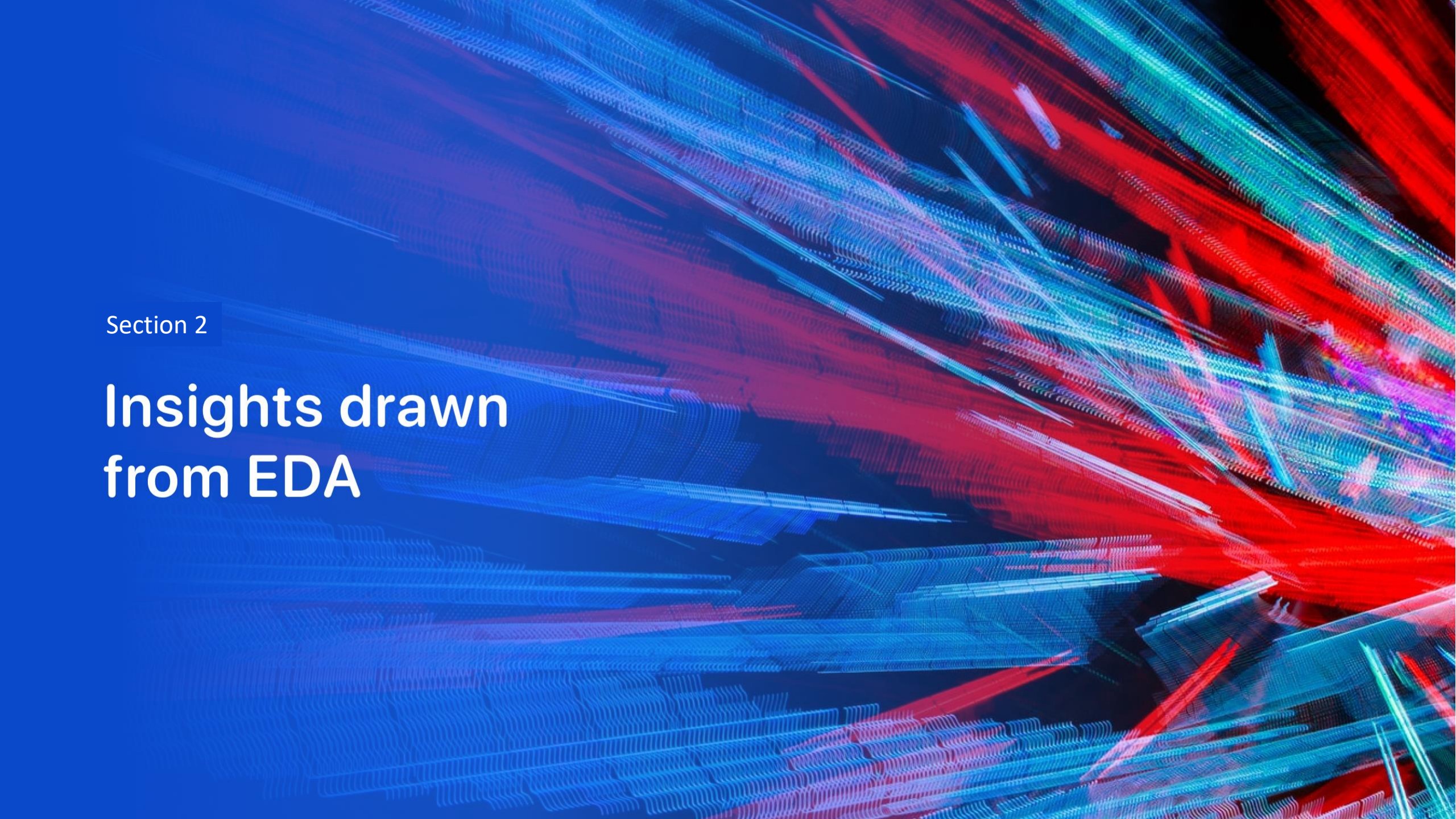


- Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around. Most launches happens at east cost launch sites

Results

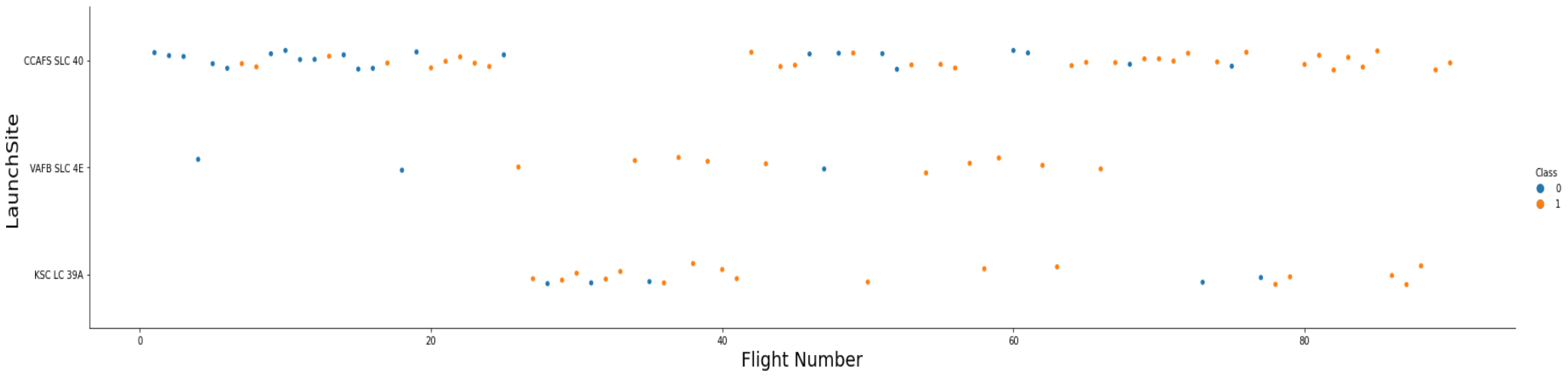
- Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%.




Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

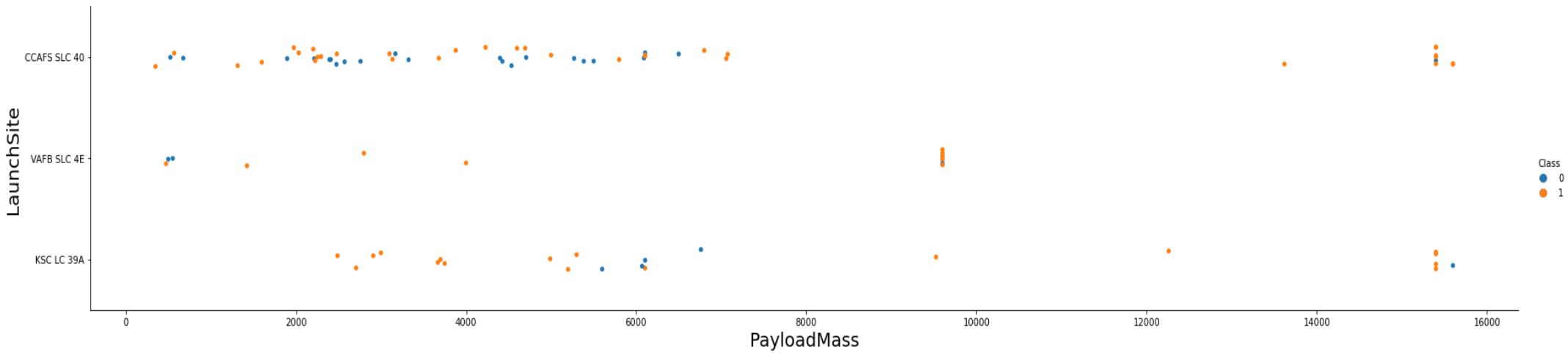
- Show a scatter plot of Flight Number vs. Launch Site



- The best launch site is CCAF5 SLC 40,
- The most successful launch site is VAFB SLC 4E
- The success rate improved over time

Payload vs. Launch Site

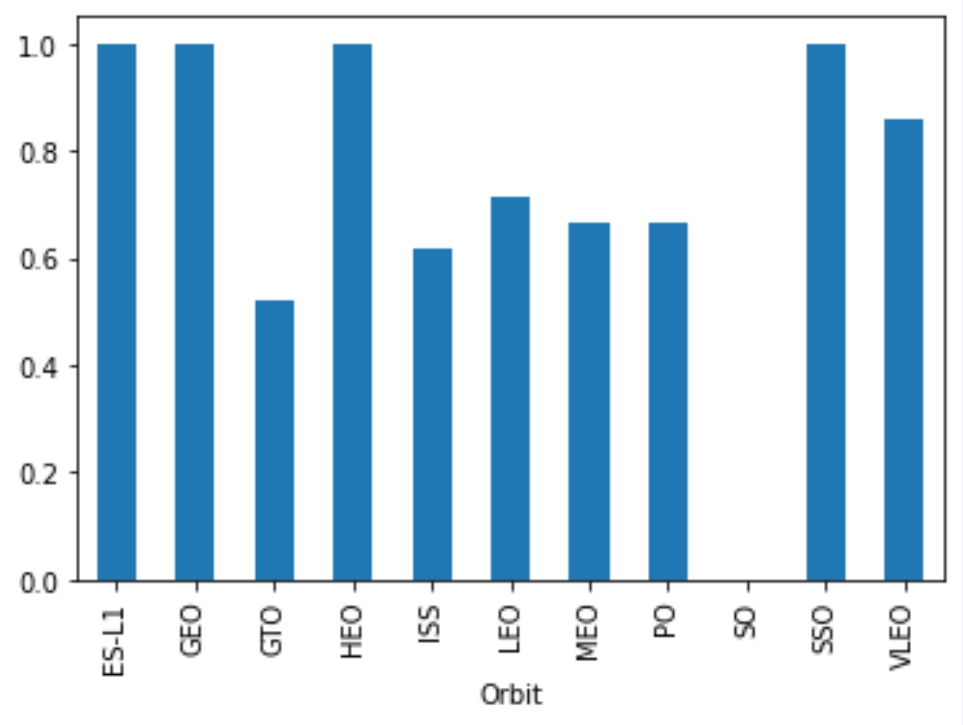
- Show a scatter plot of Payload vs. Launch Site



- Payloads over 9,000kg have the highest success rate.
- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC39A launch sites.

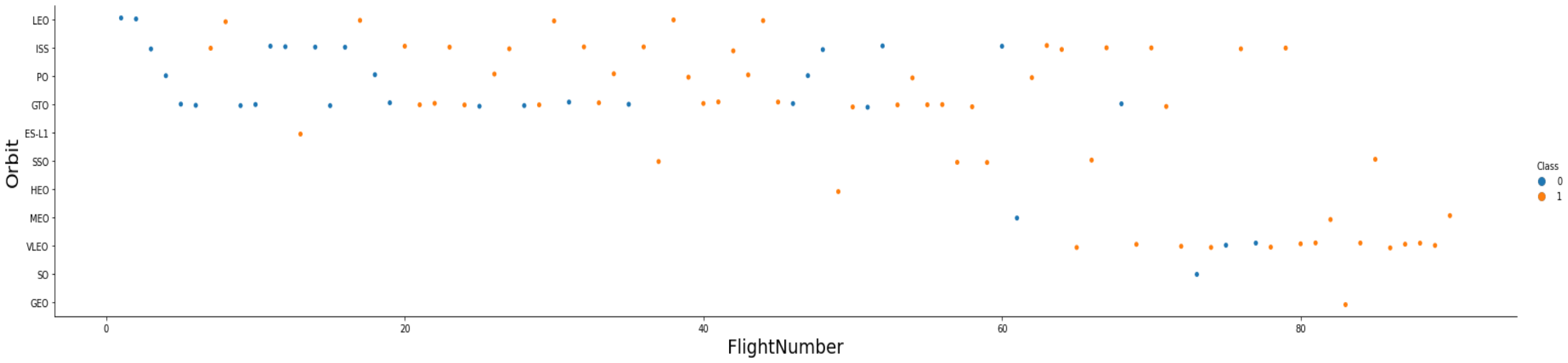
Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type
- The biggest success rates happens to orbits:
  - ES-L1
  - GEO
  - HEO
  - SSO
- Followed by:
  - VLEO (above 80%)
  - LFO (above 70%)



Flight Number vs. Orbit Type

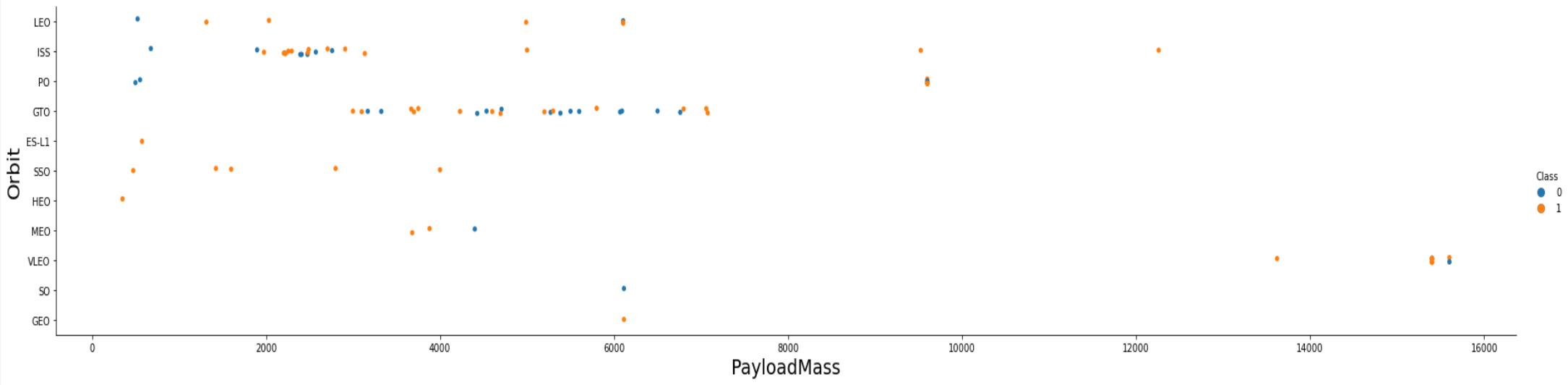
- Show a scatter point of Flight number vs. Orbit type



- Plot indicates success rate improve over time for every orbits

Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type

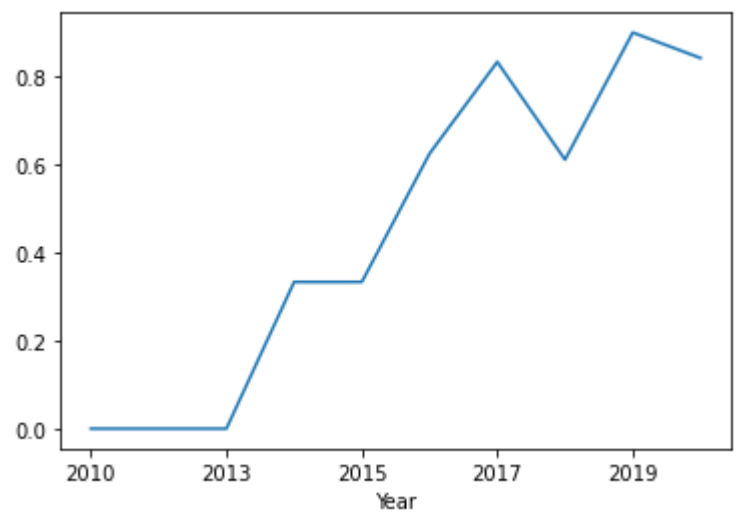


- ISS orbit has the widest range of payload
- ISS orbit has a high success rate
- There is no relation between payload and success rate

Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Success rate start pluming from 2013



All Launch Site Names

- Find the names of the unique launch sites
- Result is distinctive selection of launch sites

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`
- Present your query result with a short explanation here

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	**CCA**FS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	**CCA**FS 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	**CCA**FS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	**CCA**FS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	**CCA**FS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attemp

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Total payload calculated above, by summing all payloads whose codes contain ‘CRS’, which corresponds to NASA

Total Payload (kg)
111.268

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Filtering data by the booster version above and calculating the average payload mass we obtained the value of 2,928 kg

Avg Payload (kg)
2.928

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Filtering data by successful landing outcome on ground pad and getting the minimum value for date which is 12/22/2015.

Min Date
2015-12-22

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

- Result is distinctive selection of booster version

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

Mission Outcome	Occurrences
Success	99
Success (payload status unclear)	1
Failure (in flight)	1

- Summary mission outcome by Success and Failure

Boosters Carried Maximum Payload

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

Booster Version
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

- 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

- Field landing_outcomes in drones shop are two records above

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

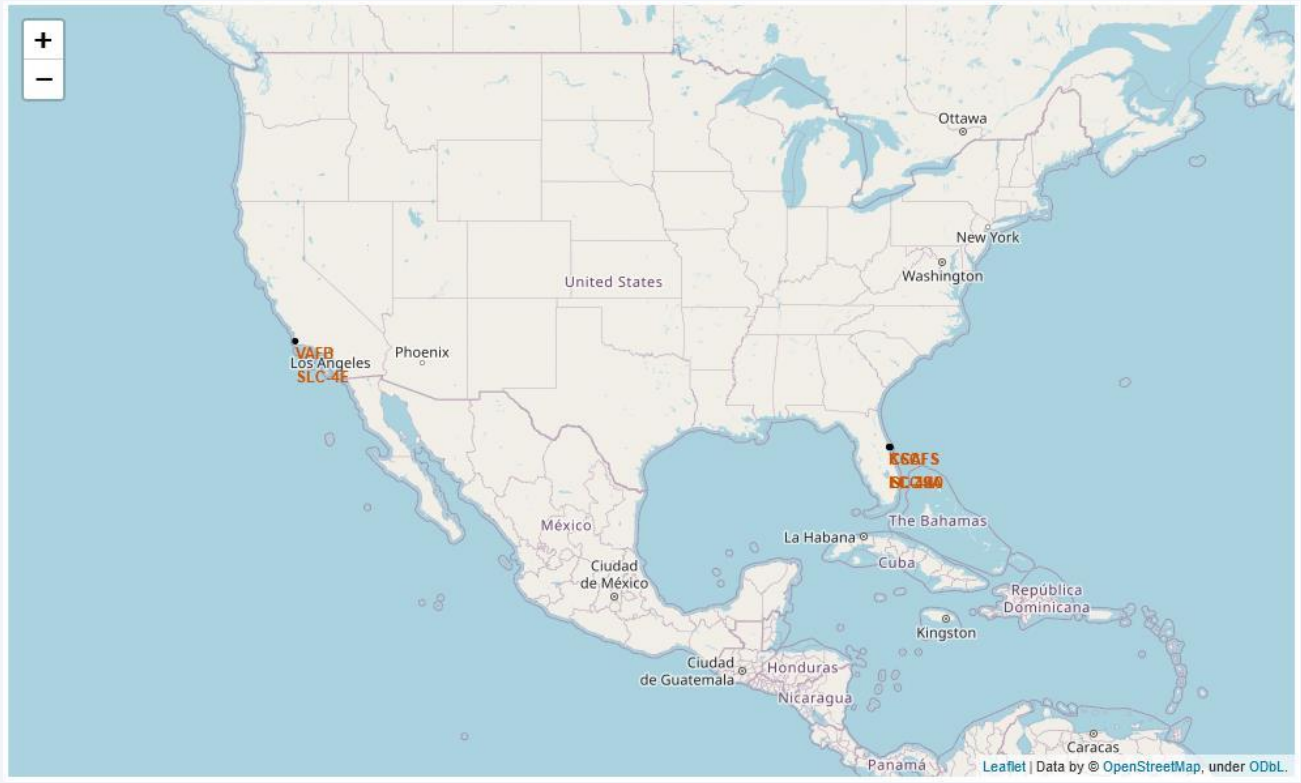
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

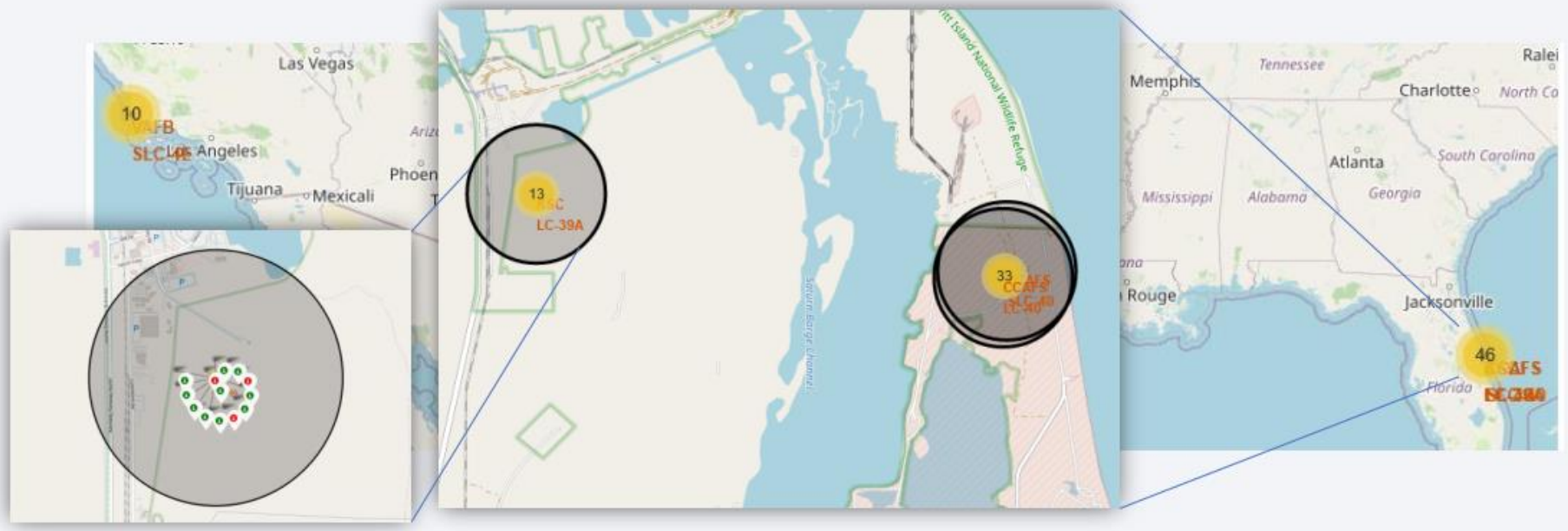
Launch sites

- All launches are located near the coast



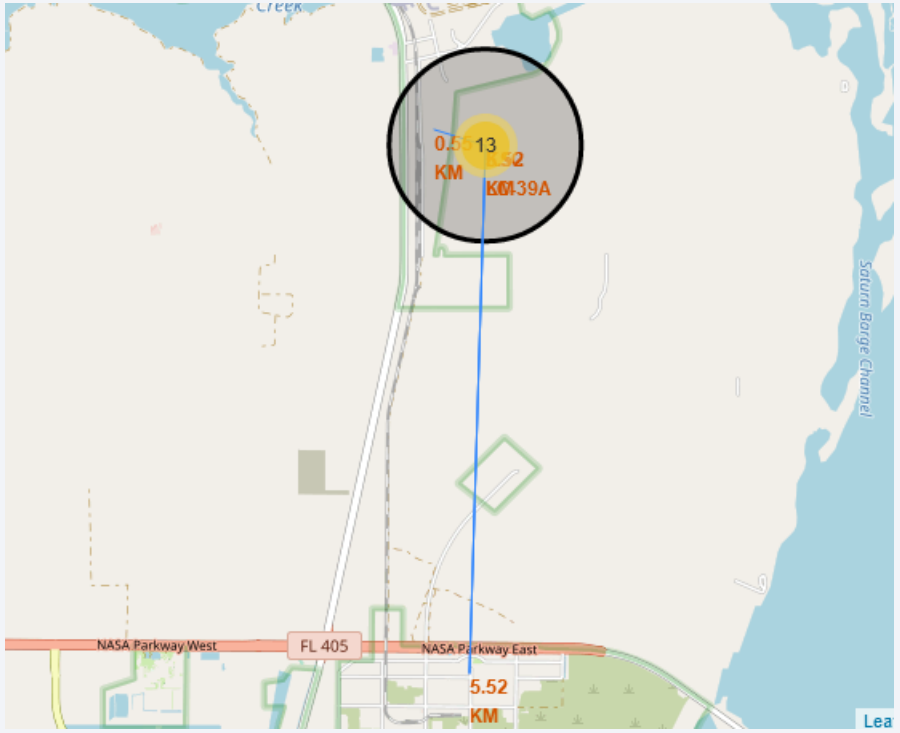
Launch site's outcome

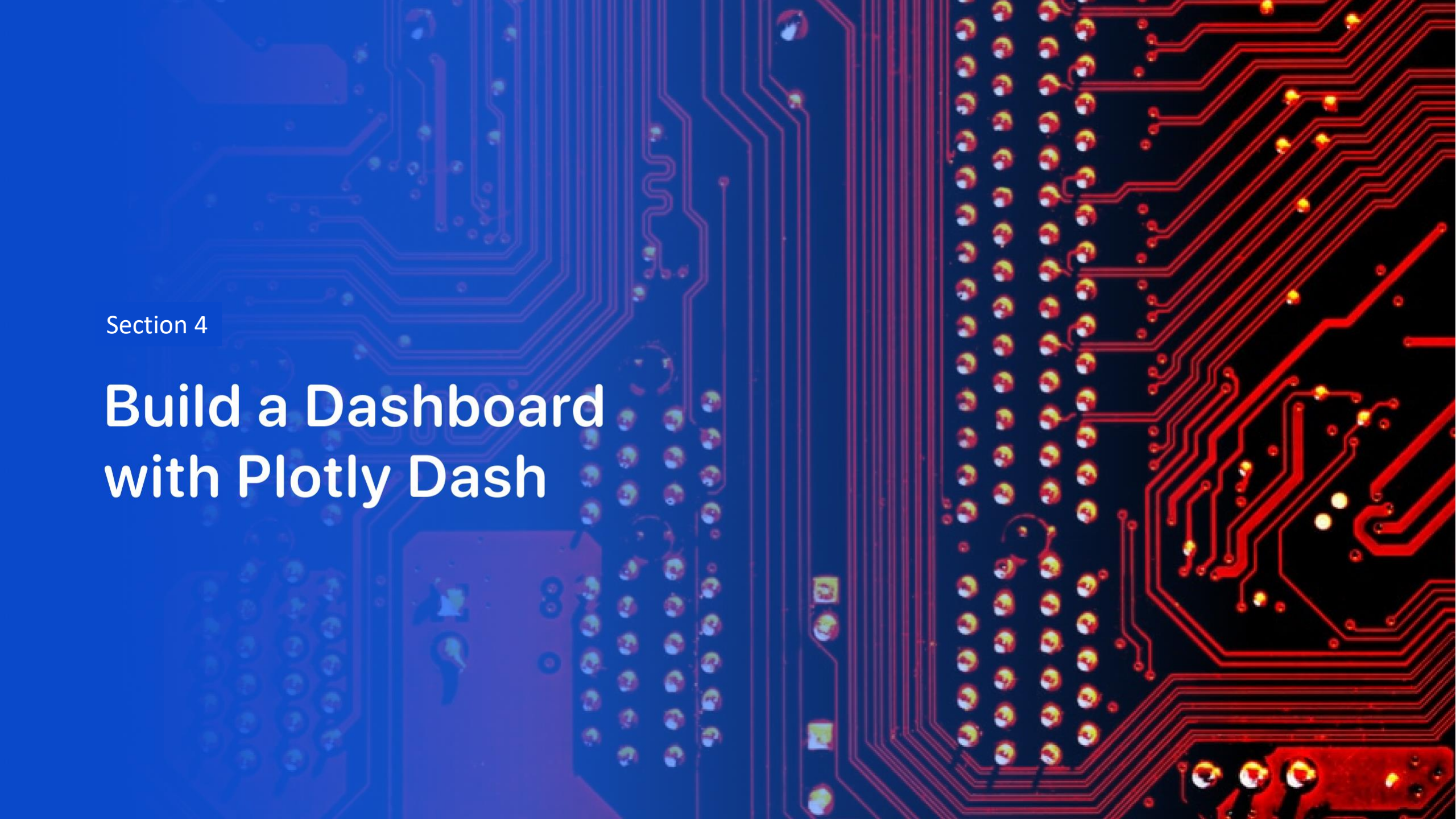
- KSC LC-39A launch site with successful and failure outcomes



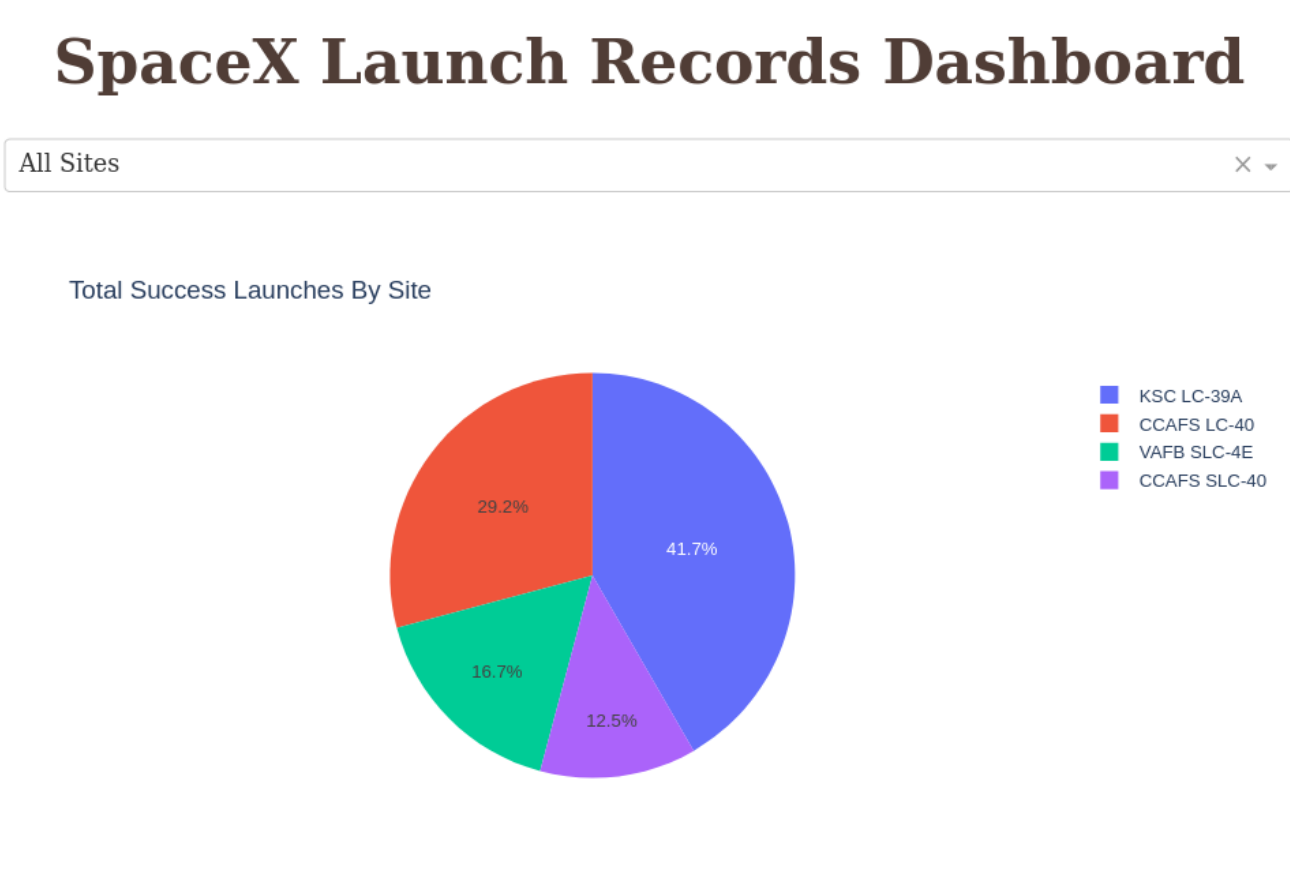
Launch site logistic

- Launch site KSC LC-39A.

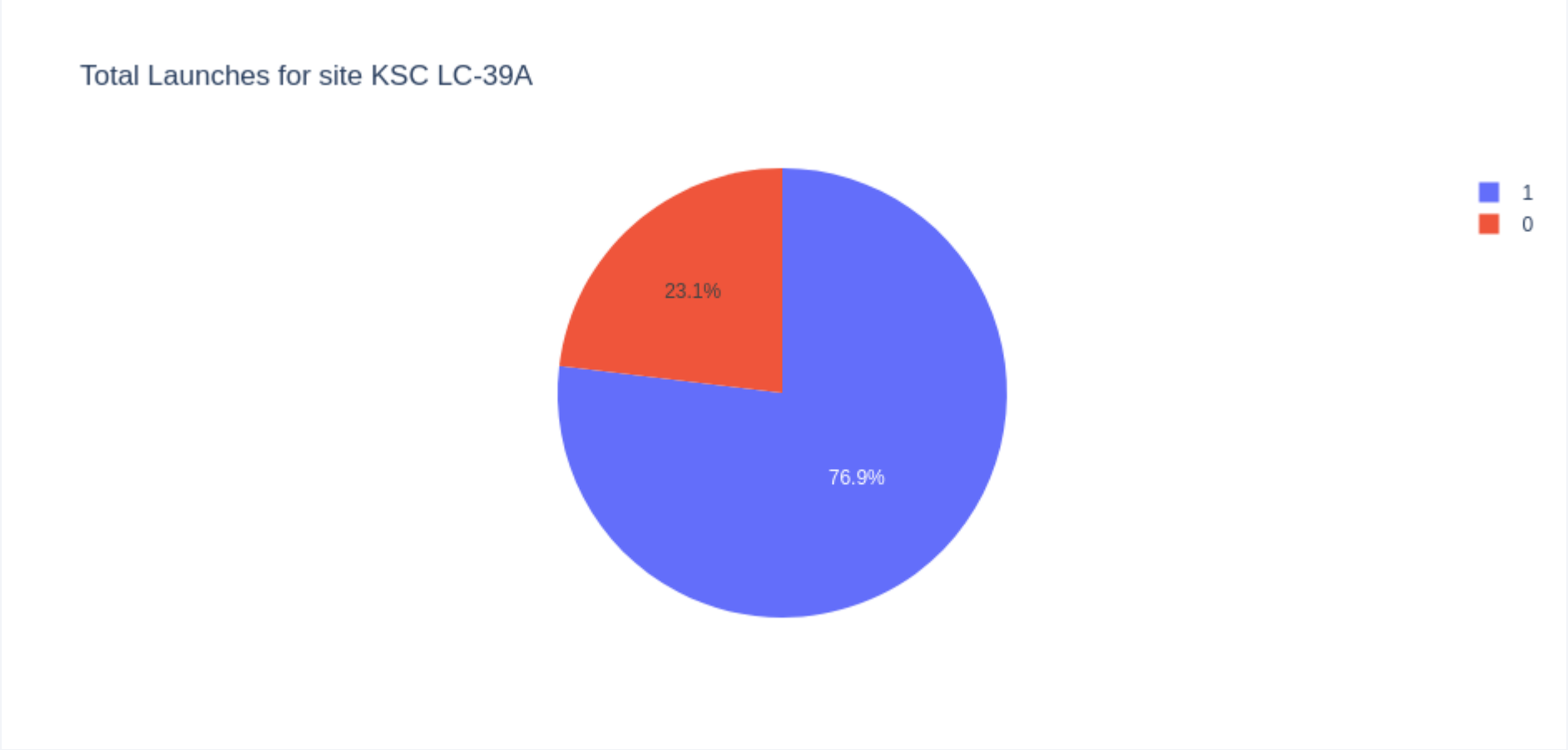



Section 4
Build a Dashboard with Plotly Dash

Successful launches by site



Launch success rate of KSC LC 39A



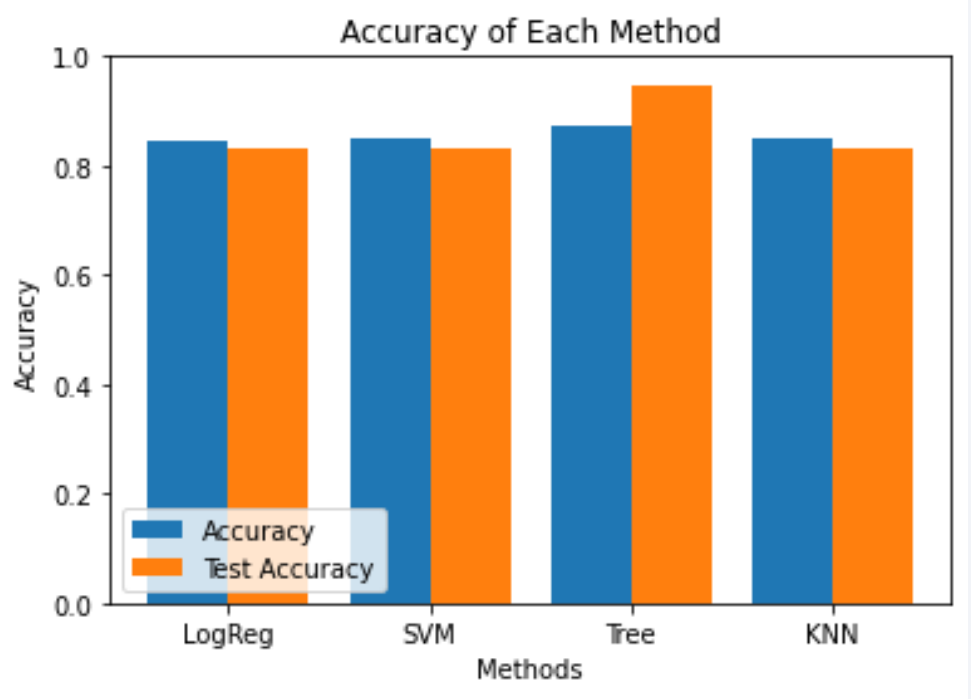
Payload vs outcome

Section 5

Predictive Analysis (Classification)

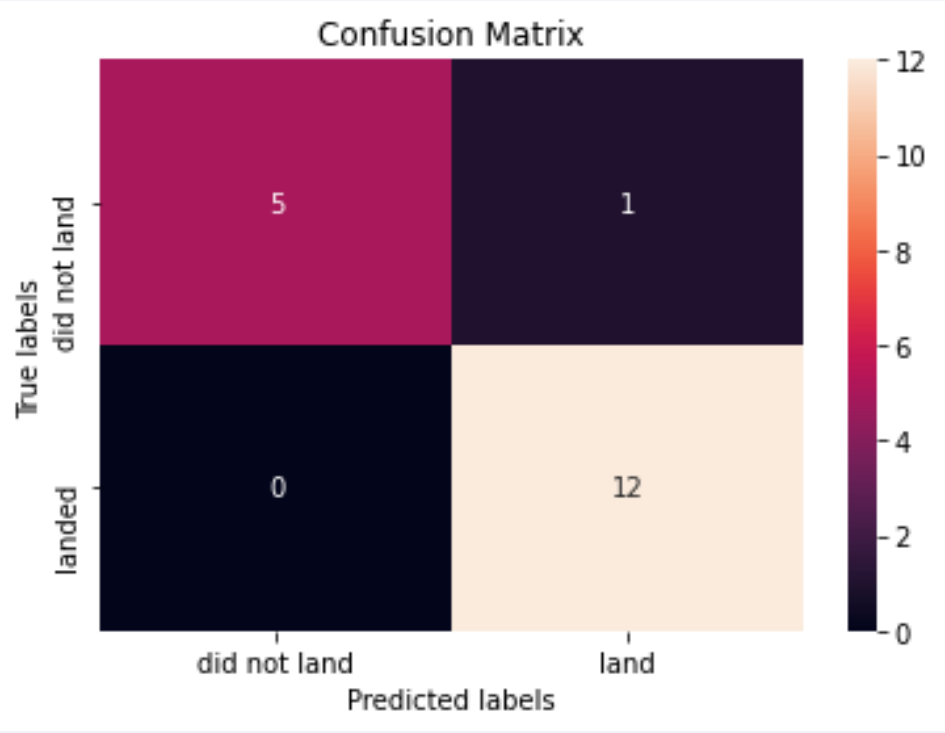
Classification Accuracy

- Four classification models were tested
- Decision Tree Classifier is the best model, which has accuracy rate over than 85%



Confusion Matrix

- Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones



Conclusions

- The best launch site is KSC LC-39A;
- Launches above 7,000kg are less risky;
- Decision Tree Classifier can be used to predict successful landings and increase profits.
- Successful outcome ratio keep improving over the time

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!