IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

KARAN ARORA
DATE -
20-OCTOBER-2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - Data Visualization and SQL
  - Interactive visual analytics using Folium
  - Predictive analysis using classification models
- Summary of all results
  - Correlations variables
  - Build a predictive model

Introduction

- Project background and context
  - The commercial space age is here, companies are making space travel affordable for everyone. Different company providing suborbital space flights. The most successful is SpaceX. SpaceX's accomplishments include: sending spacecraft to the International Space Station. Starlink, a satellite internet constellation providing satellite Internet access.Sending manned missions to space. One reason SpaceX can do this is the rocket launches are relatively inexpensive. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. SpaceX's Falcon 9 launch like regular rockets.
- Problems you want to find answers
  - The aim of this project is to determine the price of each launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using web scraping.
- Perform data wrangling
  - Exploratory Data Analysis to find some patterns in the data and determine what would be the label for training supervised models.
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Exploratory Data Analysis using python library (Matplotlib, Seaborn) and using SQL , Feature Engineering
- Perform interactive visual analytics using Folium and Plotly Dash
  - Use Folium to view previously observed correlations
- Perform predictive analysis using classification models
  - Create a machine learning pipeline to predict if the first stage will land given the data from the preceding observations

Data Collection

Import Libraries and Define Auxiliary Functions

Libraries:	**Functions**:
Request	getBoosterVersion
Pandas	getLaunchSite
Numpy	getPayloadData
Datetime	getCoreData

Request and parse the SpaceX launch data using the GET request

Decode the response content as a Json using and turn it into a Pandas dataframe

Filter the dataframe to only include Falcon 9 launches

Remove the Falcon 1 launches keeping only the Falcon 9 launches.

Data Collection – SpaceX API

Request rocket launch data from SpaceX API using URL → **Decode the response content as a Json using and convert it into a pandas dataframe** → **Fixed information in correct columns (using previously defined functions)** → **Construct dataset using the data we have obtained**

GITHUB URL:-https://github.com/karanskyhigh/Data-Collection-API

Data Collection - Scraping

Request the Falcon9 Launch Wiki page from its URL

Extract all column/variable names from the HTML table header

Create a data frame by parsing the launch HTML tables

GITHUB URL: https://github.com/karanskyhigh/Webscraping

Data Wrangling

Dealing with Missing Values → Calculate the number of launches on each site → Calculate the number and occurrence of each orbit → Calculate the number and occurence of mission outcome per orbit type → Create a landing outcome label from Outcome column

GITHUB URL: https://github.com/karanskyhigh/Data_Wrangling

EDA with Data Visualization

- The graphs used are
  - Scatter Plot
  - Bar chart
  - Line chart

  because they are the ones that best highlight the relationships between the variables considered

  GITHUB URL: https://github.com/karanskyhigh/EDA_Visulization

EDA with SQL

- **SQL queries 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 the ground pad was achieved .
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
  - List of total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass.
  - List the failed landing_outcomes in drone ship,their booster versions, and launch site names for

GITHUB URL: https://github.com/karanskyhigh/EDA_SQL

Build an Interactive Map with Folium

- Summary of map objects:
    - Markers: Show a geo location from latitude and longitude data
    - Cluster: Show a group of markers
    - Circles: Show a single location
    - Lines: Show distance between two

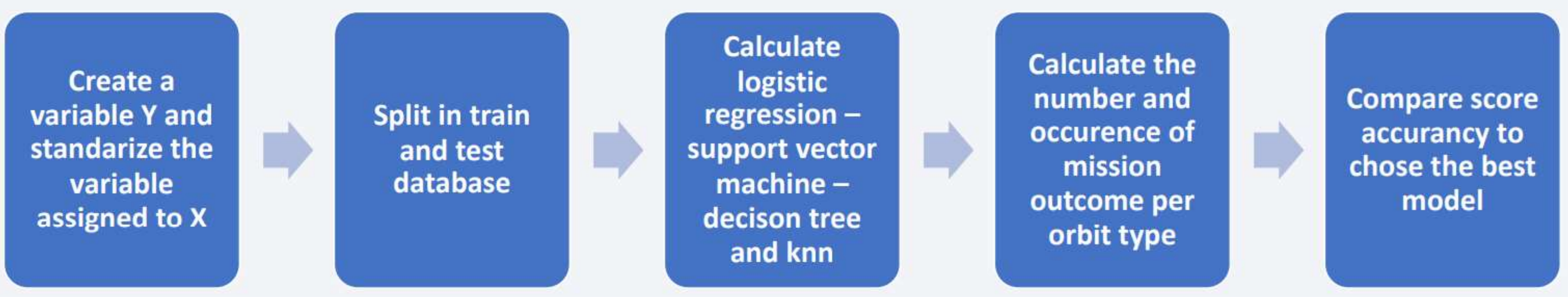
      I have added object to find some geographical patterns about launch site

GITHUB URL:https://github.com/karanskyhigh/Interactive-Visual-Analytics-with-Folium-lab

Build a Dashboard with Plotly Dash

- Summary of plots:
    - Bar: Show categories differences
    - Line: Reports time series changes
    - Pie: Shows the percentage of events
    - Tree: Shows complex relationship of variables in interactive way
    - Map: Shows variables of states on a map

Predictive Analysis (Classification)



GITHUB URL: https://github.com/karanskyhigh/Machine-Learning-Prediction-lab

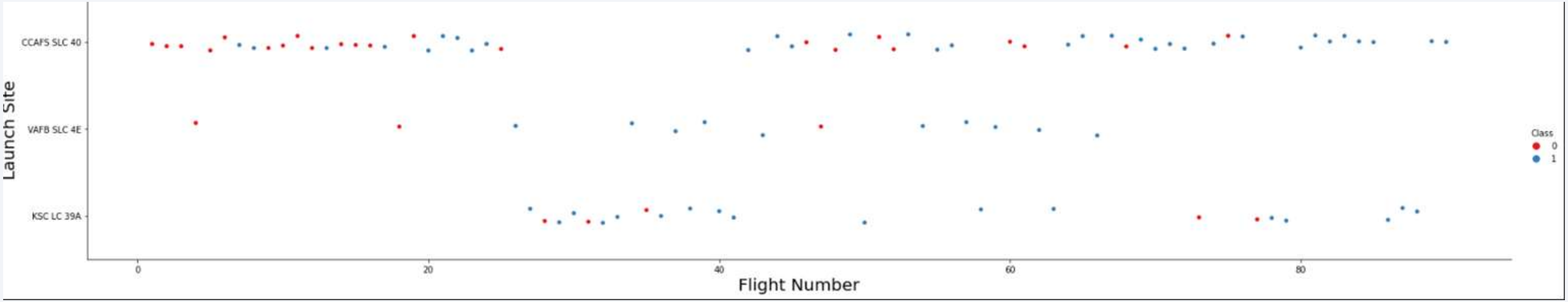
Results

- Exploratory data analysis results
    - Both API and web scraping are capable to collect Xspace data
- Interactive analytics demo in screenshots
    - EDA with SQL is effective for data filtering
    - EDA with interactive visualization provides informative information
    - Plotly Dash is powerful to show instant data change
- Predictive analysis results
    - Decision Tree Classifier Algorithm has the best accuracy of predicting.

Section 2

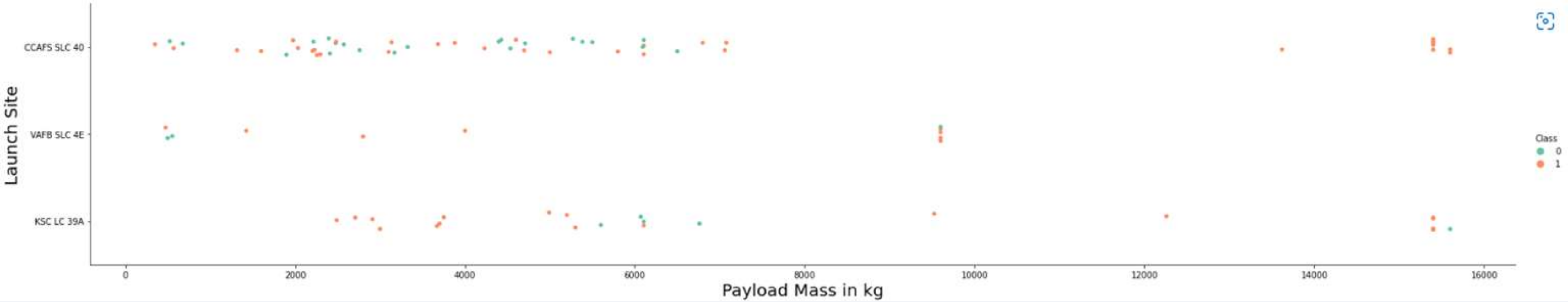
Insights drawn from EDA

Flight Number vs. Launch Site



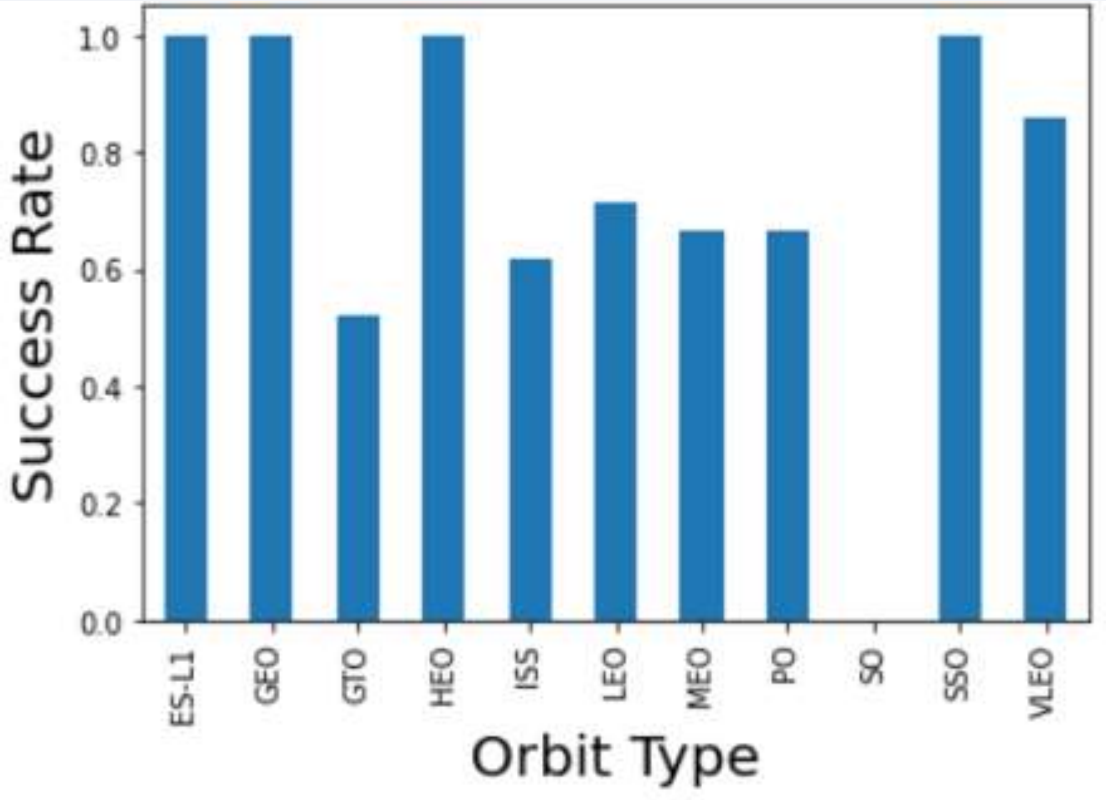
- For the CCFS SLC 40 category there seems to be a higher concentration than class 1 for flights with Flight Number high.

Payload vs. Launch Site



- In this case there does not appear to be strong correlations

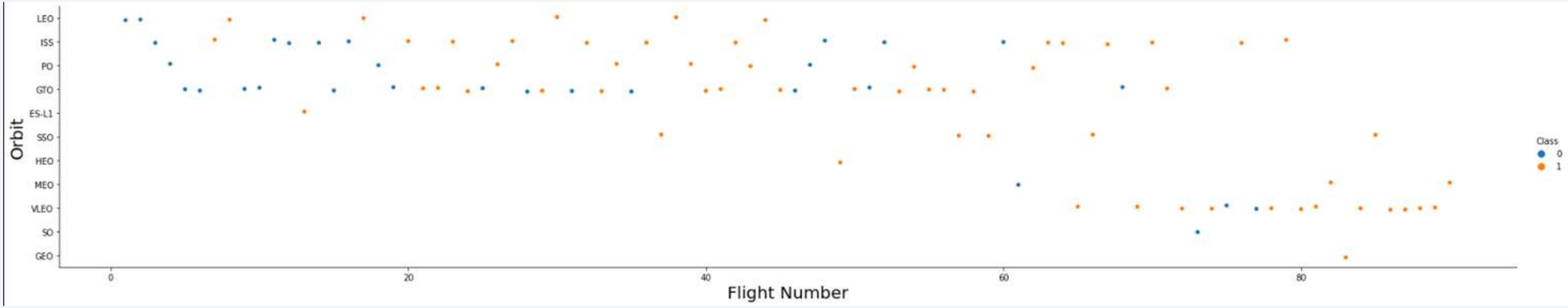
Success Rate vs. Orbit Type



- We can see that the orbits with the highest success rate are:SSO, HEO, GEO, ES-L1

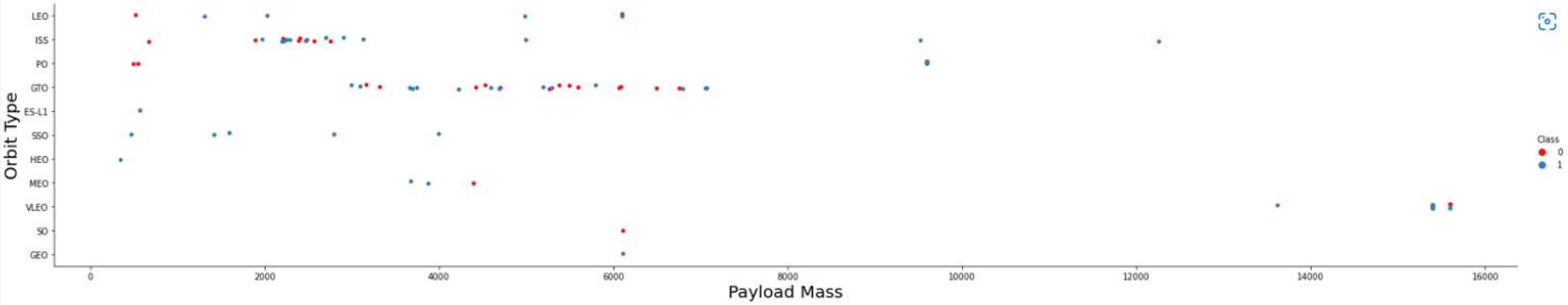
While the GTO Orbit it is the one with lowest rate

Flight Number vs. Orbit Type



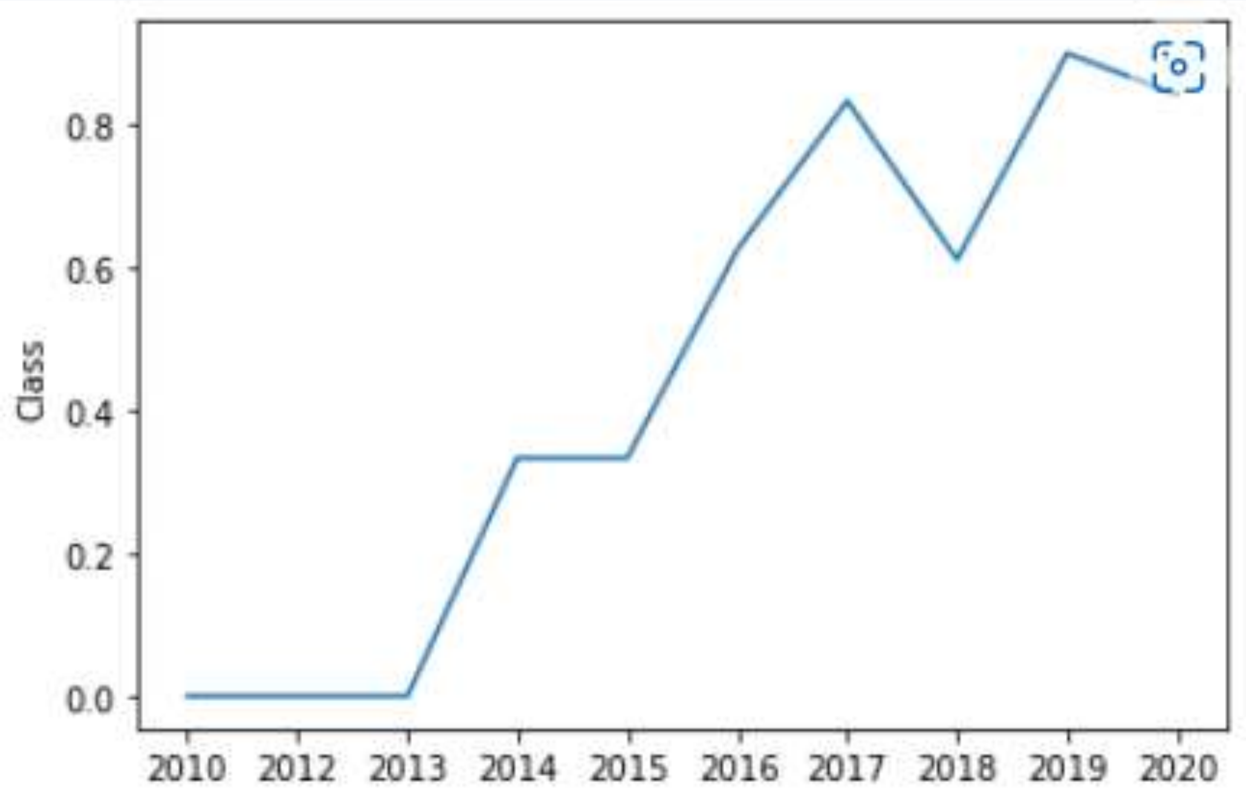
- You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand , there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



- You should observe that Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.

Launch Success Yearly Trend



- You can see that the success rate since 2013 kept increasing till 2020.

All Launch Site Names

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

- This four launch sites are present in the database.

Launch Site Names Begin with 'CCA'

- Records where launch sites begin with `CCA`

DATE	time_utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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
2012-08-10	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- The total payload carried by boosters from NASA is 45596.

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928

First Successful Ground Landing Date

- First successful landing outcome on ground pad is 2015-12-22

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

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum 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

- This is the names of the booster which have carried the maximum payload mass

2015 Launch Records

landing_outcome	launch_site	booster_version
Failure (drone ship)	CCAFS LC-40	F9 v1.1 B1012
Failure (drone ship)	CCAFS LC-40	F9 v1.1 B1015

- There are two failed landing_outcomes in drone ship, in 2015, and they have the same launch site.

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

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

- We can see that the most numerous are <<No attempt>>,followed by failure/success (drone ship)

Section 3

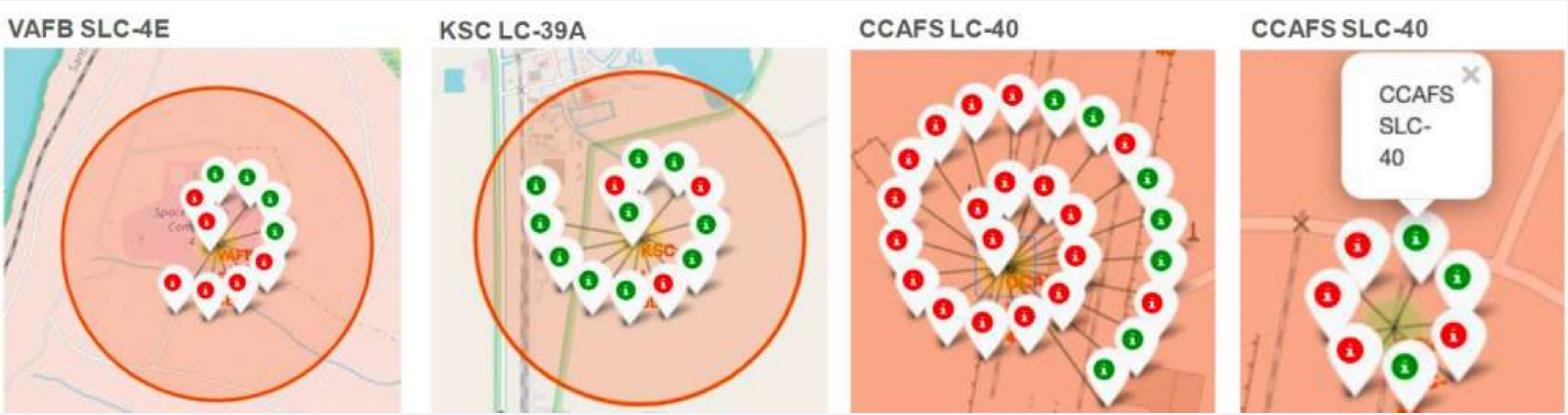
Launch Sites Proximities Analysis

Launch Site

- Both ground and sea surface sites are necessary also south areas maybe a proper areas.
- The transportation base chosen maybe important

Success of launch

- Color icons is impressive way to show the rate of success. For example, in KSC there is the best rate of success.



Distance

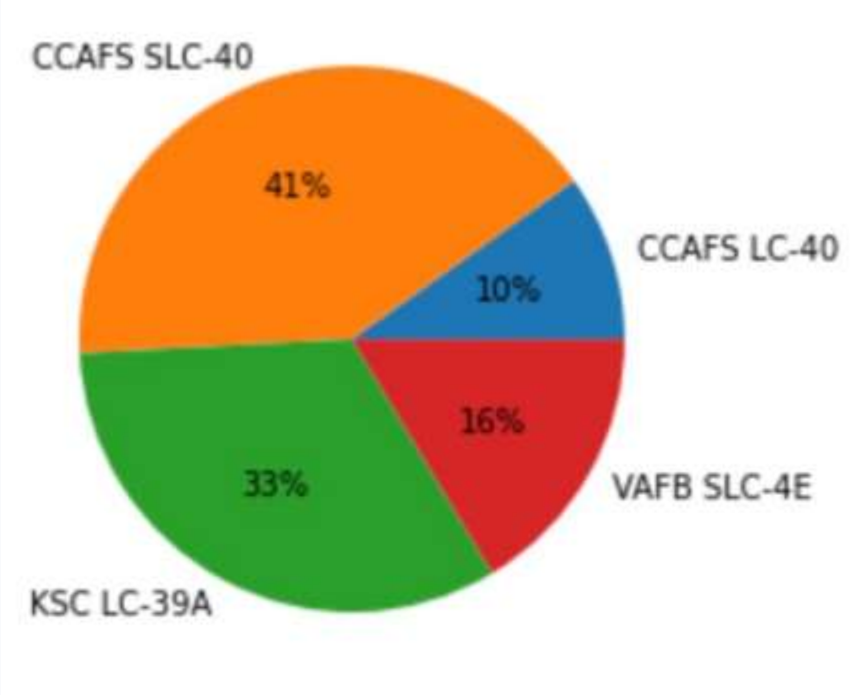
- This map shows the distance from points of interest. In this case the distance from railway, coast or port may be importance

Section 4

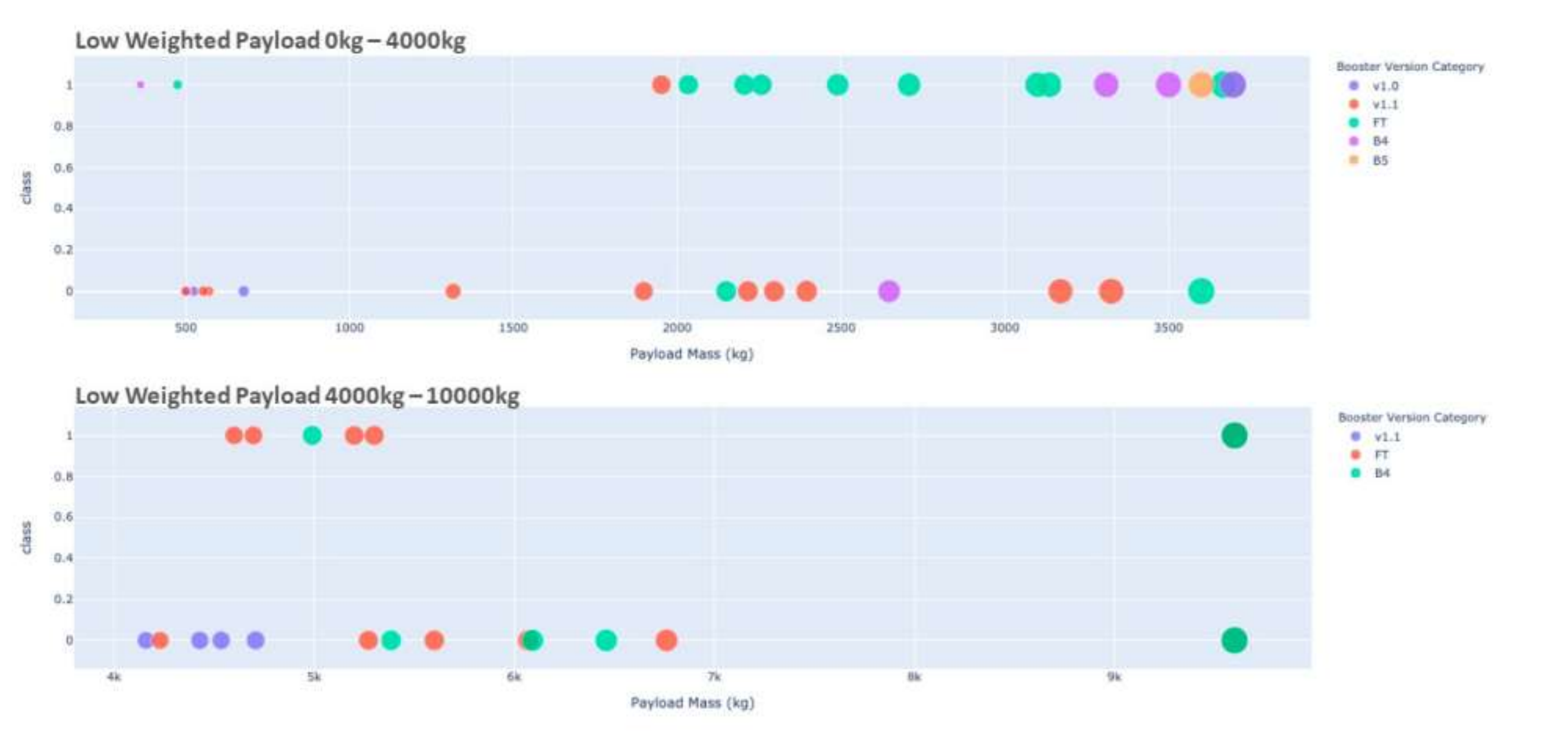
Build a Dashboard with Plotly Dash

The most successful launch site

- We can see that CCAFS SLC-40 had the most successful launches from all the sites.



Payload and Success launch



Section 5

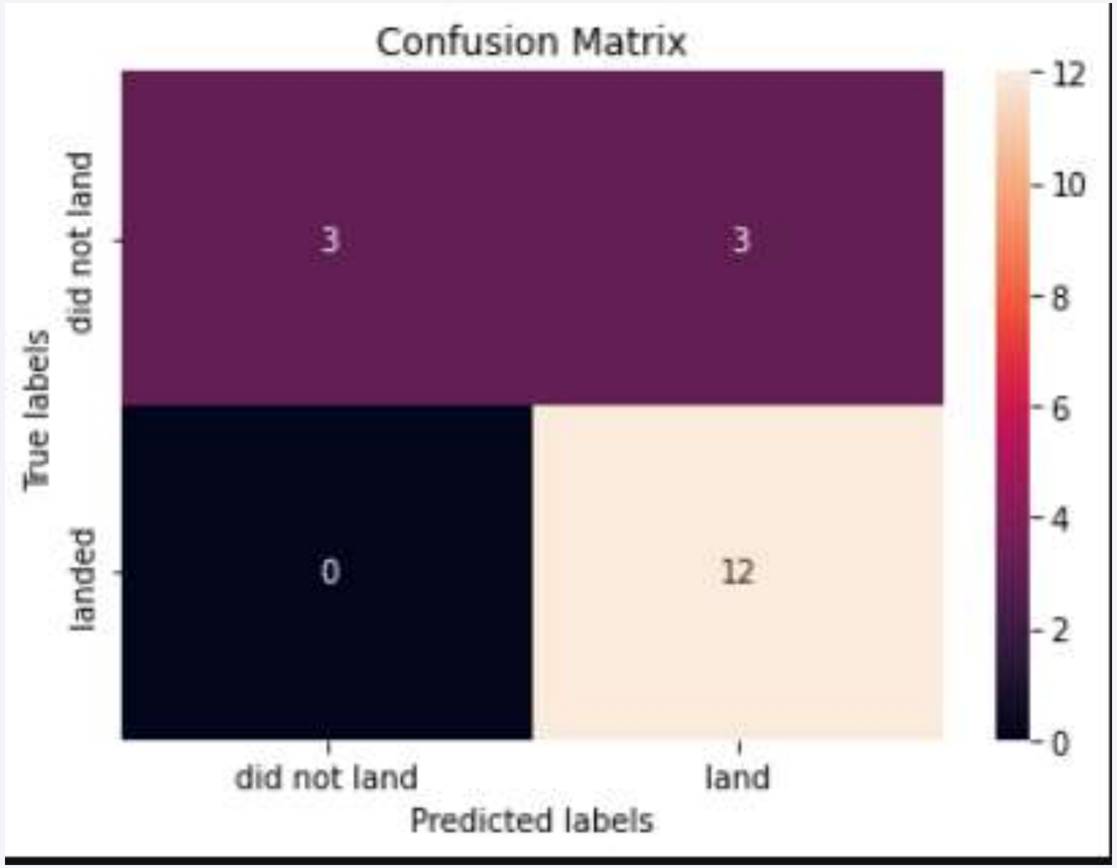
Predictive Analysis (Classification)

Classification Accuracy

- As you can see our accuracy is extremely close, but the best is the “Decision Tree” with a score of 0.89

	Algorithm	Score
0	KNN	0.833333
1	Decistion Tree	0.888889
2	SVM	0.833333
3	LogisticRegression	0.833333

Confusion Matrix



Conclusions

- Orbits ES-L1,GEO,HEO,SSO has highest Success rates.
- Success rates for spaceX launches has been increasing relatively with time and it looks like soon they will reach the required target
- Decision Tree Classifier Algorithm is the best for Machine Learning Model for provided dataset

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!