IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Franklin Chukwuemeka

30th October, 2024

Outline

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

Executive Summary

- **Summary of methodologies:**

    Data Collection, Data Wrangling,

    Exploratory Analysis, Visual Analytics,

    Predicitve Analysis, Model Evaluation

- **Summary of all results**

    The best performing model (Decision Tree) can distinguish between the different classes. The major problem is false positives.

Introduction

- We want to explore SpaceX Falcon 9 launches
- The problem is training a machine learning model to predict/classify success or failure a Falcon 9 launch

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
    - Using an API and also using web scraping of a Wikipedia page
- Perform data wrangling
    - We kept the related features, used one-hot encoding to convert the categorical features to numeric values, replaced the missing data with mean, and we normalized the 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
    - We used logistic regression, support vector machine, decision tree, and K nearest neighbors. We used GridSearch to tune the models. Finally, we used models score to evaluate and choose the best model.

Data Collection

- Used SpaceX REST API to collect data from SpaceX
- Also used web scraping of a Wikipedia page to collect data

Data Collection - SpaceX API

- SpaceX REST API -> JSON -> Pandas Data Frame

- https://github.com/franklinchuks/IBM-DATA-SCIENCE-LABS/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- Wikipedia url -> Beautiful Soup -> Pandas Data Frame

- https://github.com/franklinchuks/IBM-DATA-SCIENCE-LABS/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- Created a target Class column of 1 if outcome is success and 0 otherwise

- https://github.com/franklinchuks/IBM-DATA-SCIENCE-LABS/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Scatter plots of Flight Number vs Pay Load vs Launch Site
- Bar chart of Success Rate of each Orbit Type
- Scatter plots of Flight Number vs Orbit Type
- Line plot of Year vs Success Rate
- https://github.com/franklinchuks/IBM-DATA-SCIENCE-LABS/blob/main/spacex_dash_app.py

EDA with SQL

- names of the unique launch sites
- the total payload mass carried by boosters launched by NASA (CRS)
- average payload mass carried by booster version F9 v1.1
- the total number of successful and failure mission outcomes
- names of the booster_versions which have carried the maximum payload
- Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20

Build an Interactive Map with Folium

- Map centerd at NASA location
- Added launch site locations to the map
- Added cluster of success and failure for each launch site
- Added closest proximities (like coastline) and their distance from launch sites
- https://github.com/franklinchuks/IBM-DATA-SCIENCE-LABS/blob/main/launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- A drop down menu to choose the launch site
- A range slider to choose the range of the payload mass
- Pie chart of success rate per launch site
- Pie chart of success / failure for each launch site
- Scatter plot payload mass vs success class colored by launch site
- https://github.com/franklinchuks/IBM-DATA-SCIENCE-LABS/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- 4 classification machine learning methods: logistic regression, support vector machine, decision tree, and K nearest neighbors
- Used GridSearch to tune each models with different parameters
- Used models' score to evaluate models and choose the best algorithm with best tuned parametes
- https://github.com/franklinchuks/IBM-DATA-SCIENCE-LABS/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

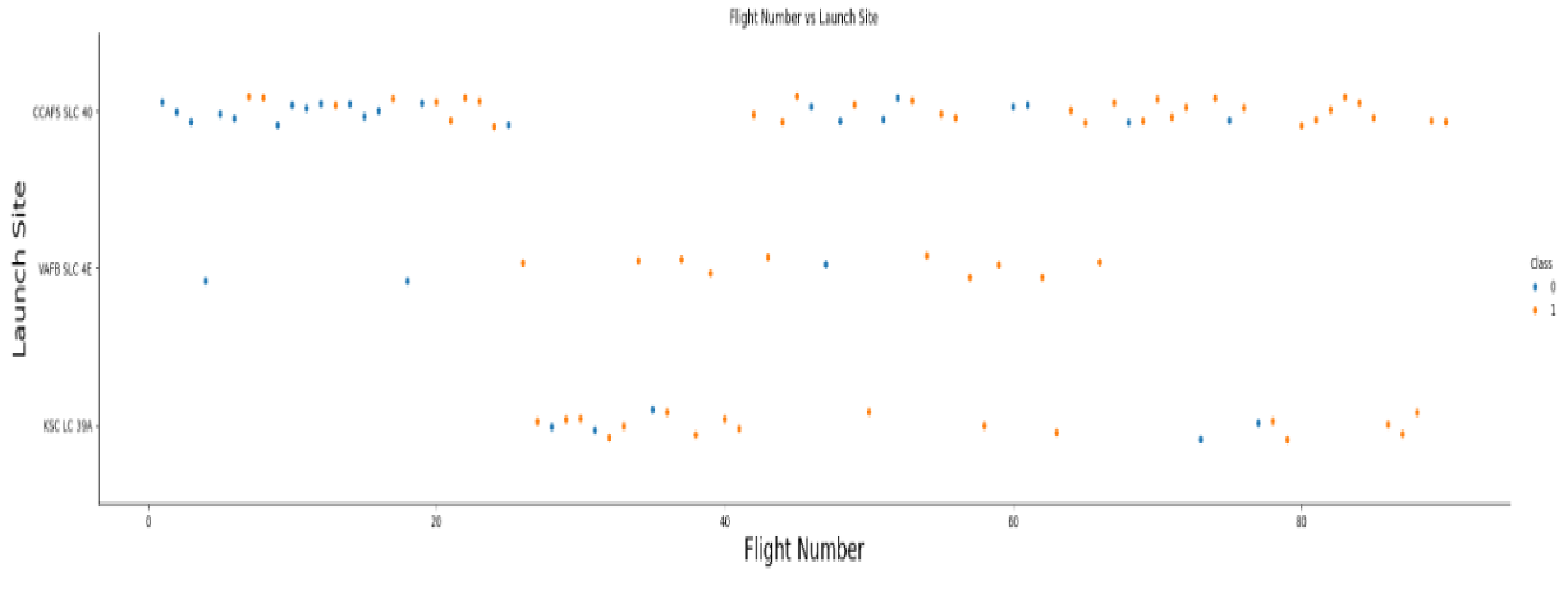
Results

- Exploratory data analysis results:
- The most successful orbit types are: ES-L1, GEO, HEO, SSO, VLEO
- The most successful launch sites are KSC LC-39A and CCAFS LC-40
- The success rate since 2013 kept increasing till 2020
- Predictive analysis results:
- The best performing model (Decision Tree) can distinguish between the different classes with high accuracy. The major problem is false positives.

Section 2

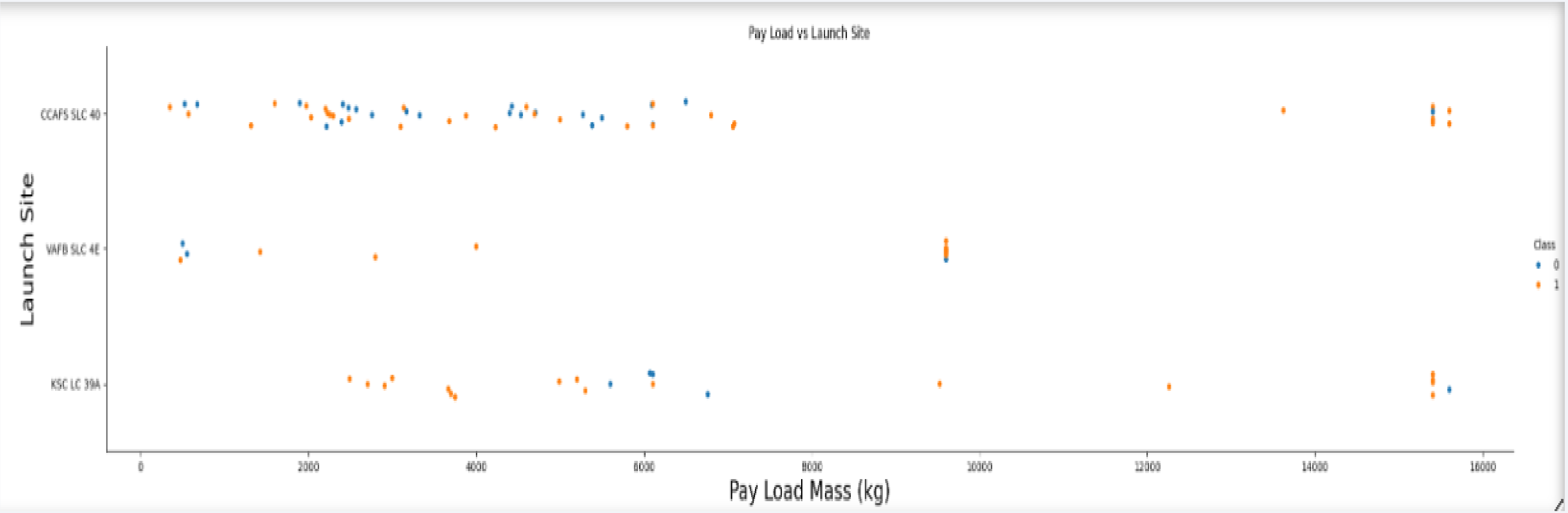
Insights drawn from EDA

Flight Number vs. Launch Site



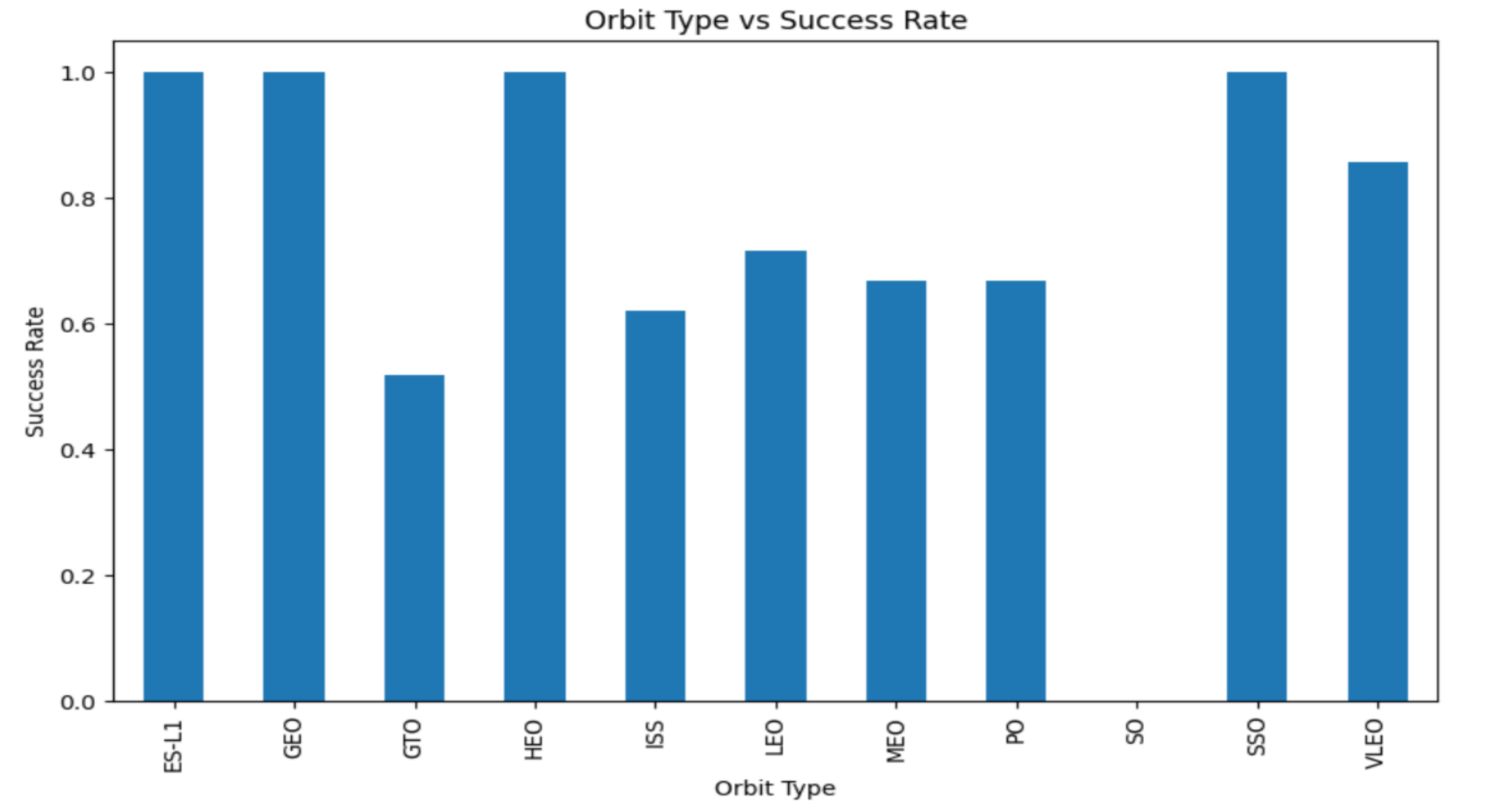
- The site CCAFS SLC 40 has the most number flights

Payload vs. Launch Site



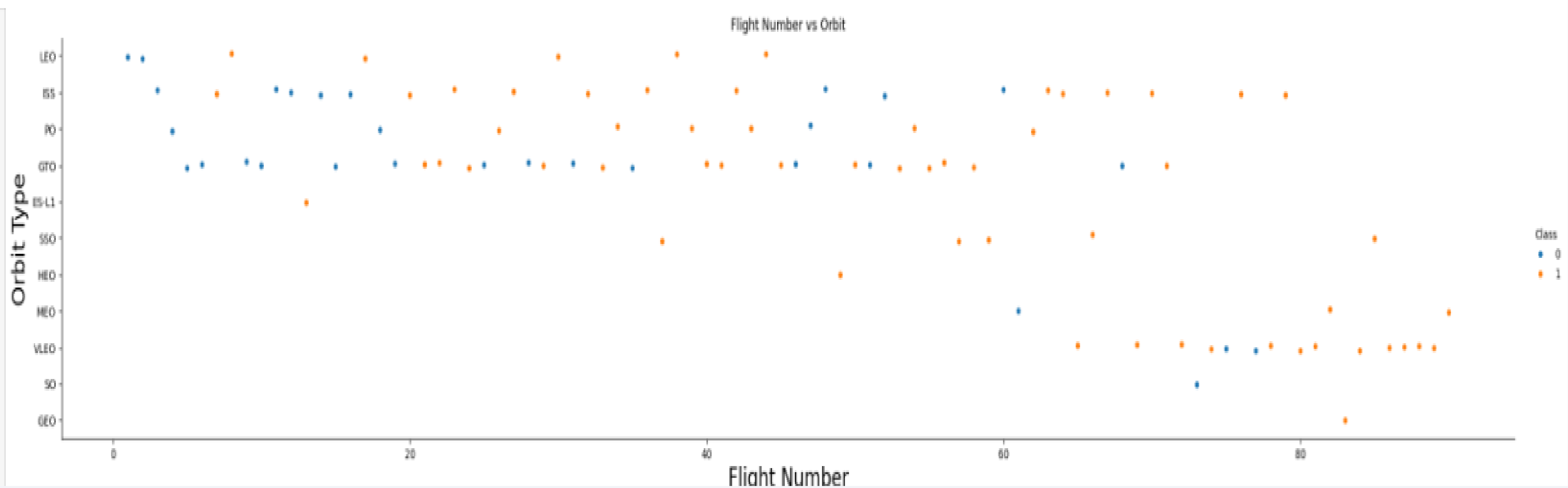
- There is no very heavy payload at the site VAFB SLC 4E
- There is no very light payload at the site KSC LC 39A

Success Rate vs. Orbit Type



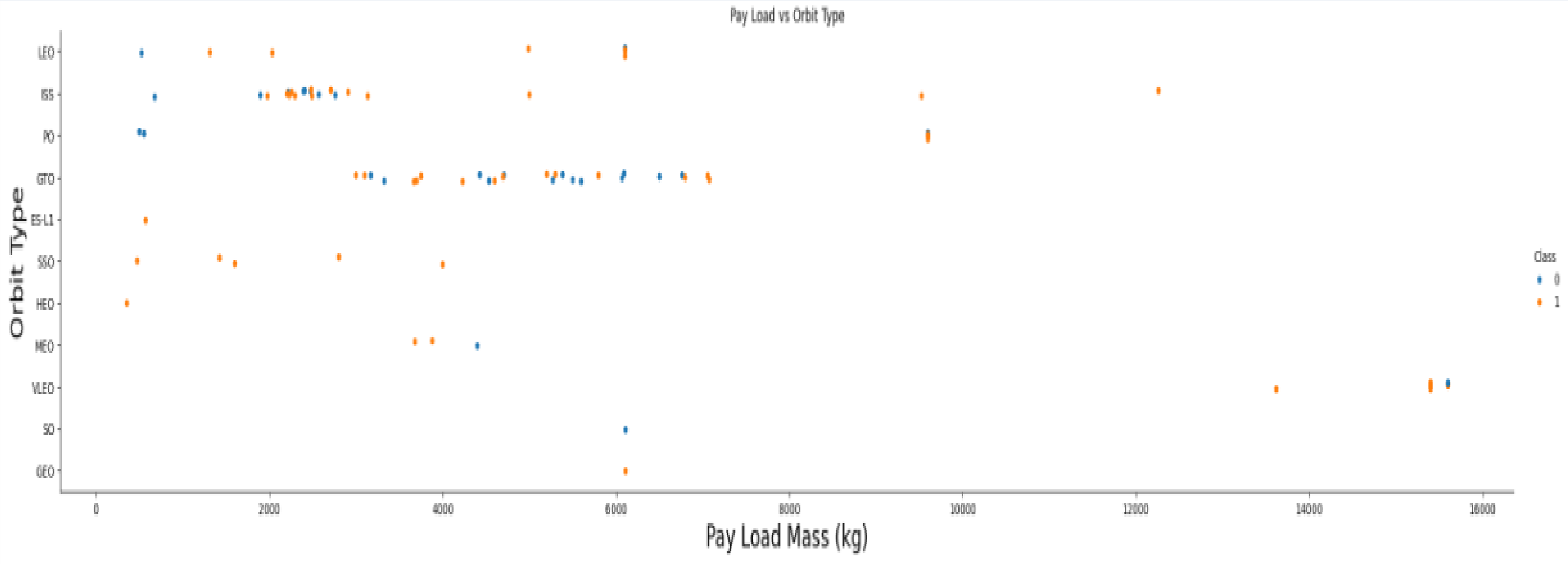
- The most successful orbit types are: ES-L1, GEO, HEO, SSO, VLEO

Flight Number vs. Orbit Type



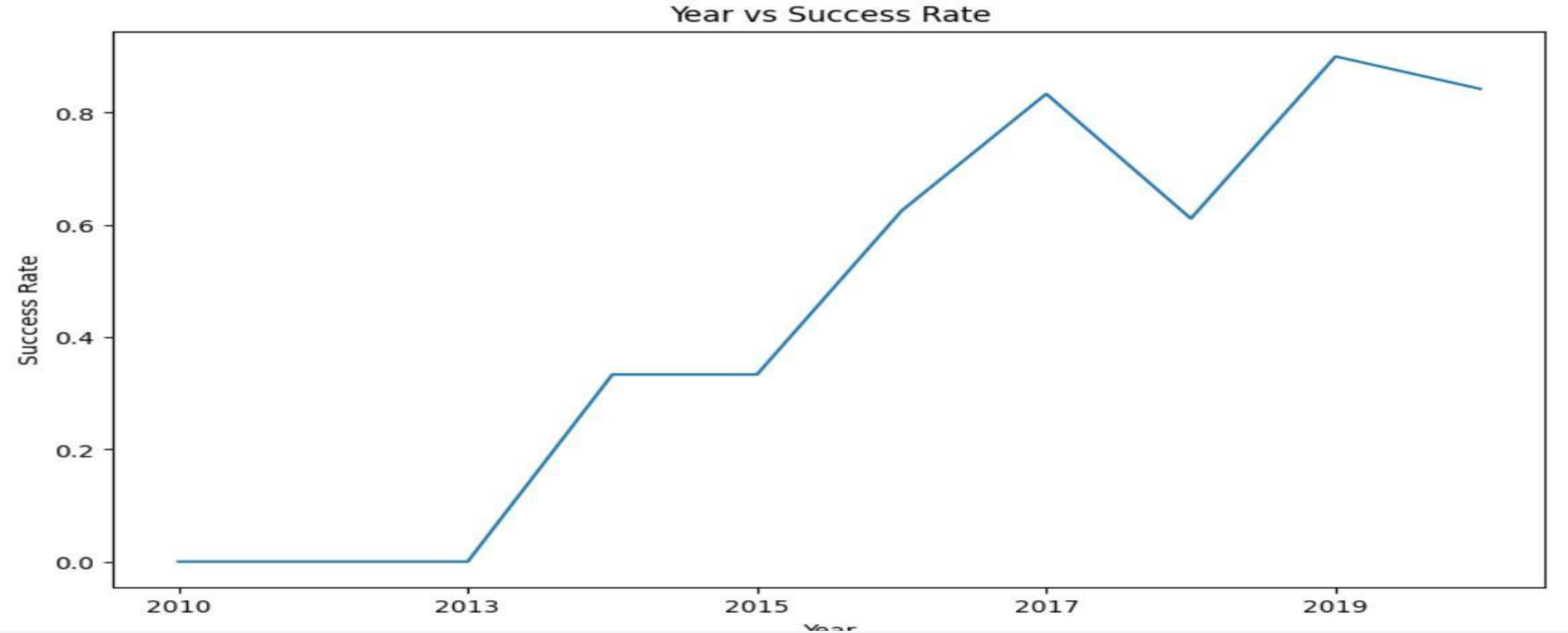
- The most commonly used orbit types are ISS and GTO
- The success of LEO orbit appears related to the number of flights
- there seems to be no relationship between flight number when in GTO orbit

Payload vs. Orbit Type



- The orbit type VLEO is only used for the heaviest payload mass
- With heavy payloads the success rate are more for Polar,LEO, ISS, and VLEO

Launch Success Yearly Trend



- The success rate since 2013 kept increasing till 2020

All Launch Site Names

- The names of the unique launch sites (there are 4):

Launch_Site
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`:

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attemp
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attemp
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS 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 Mass by NASA(CRS)
45596

Average Payload Mass by F9 v1.1

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

Average Payload Mass by F9 v1.1
2534.6666666666665

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad:

first succesful landing outcome in ground pad
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 B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes:

Mission_Outcome	count(*)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- The 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

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

Month of 2015	Landing_Outcome	Booster_Version	Launch_Site
Jan	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Apr	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
No attempt	21
Success (drone ship)	14
Success (ground pad)	9
Failure (drone ship)	5
Controlled (ocean)	5
Uncontrolled (ocean)	2
Precluded (drone ship)	1

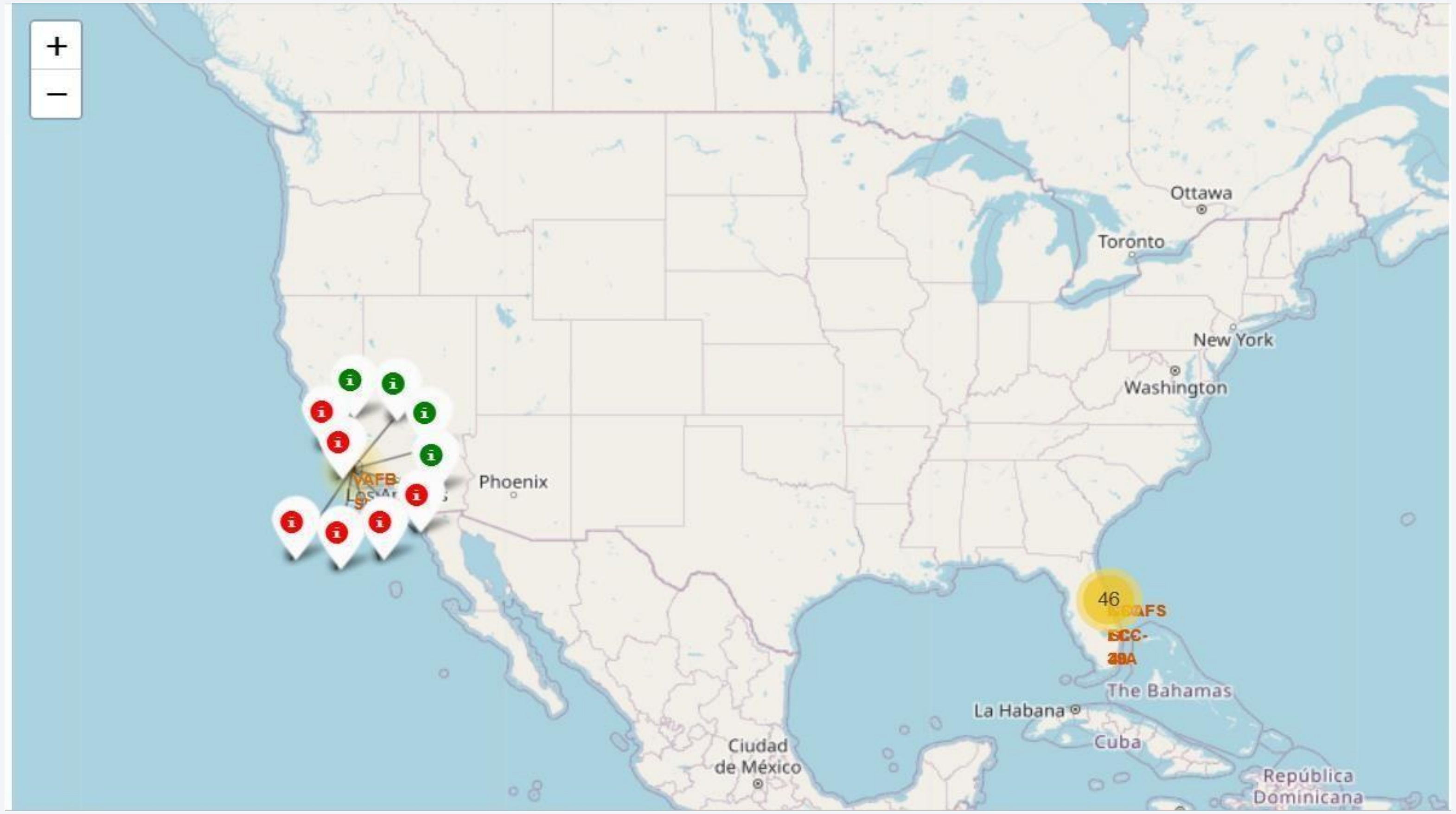
Section 3

Launch Sites Proximities Analysis

Launch Site Locations on Map

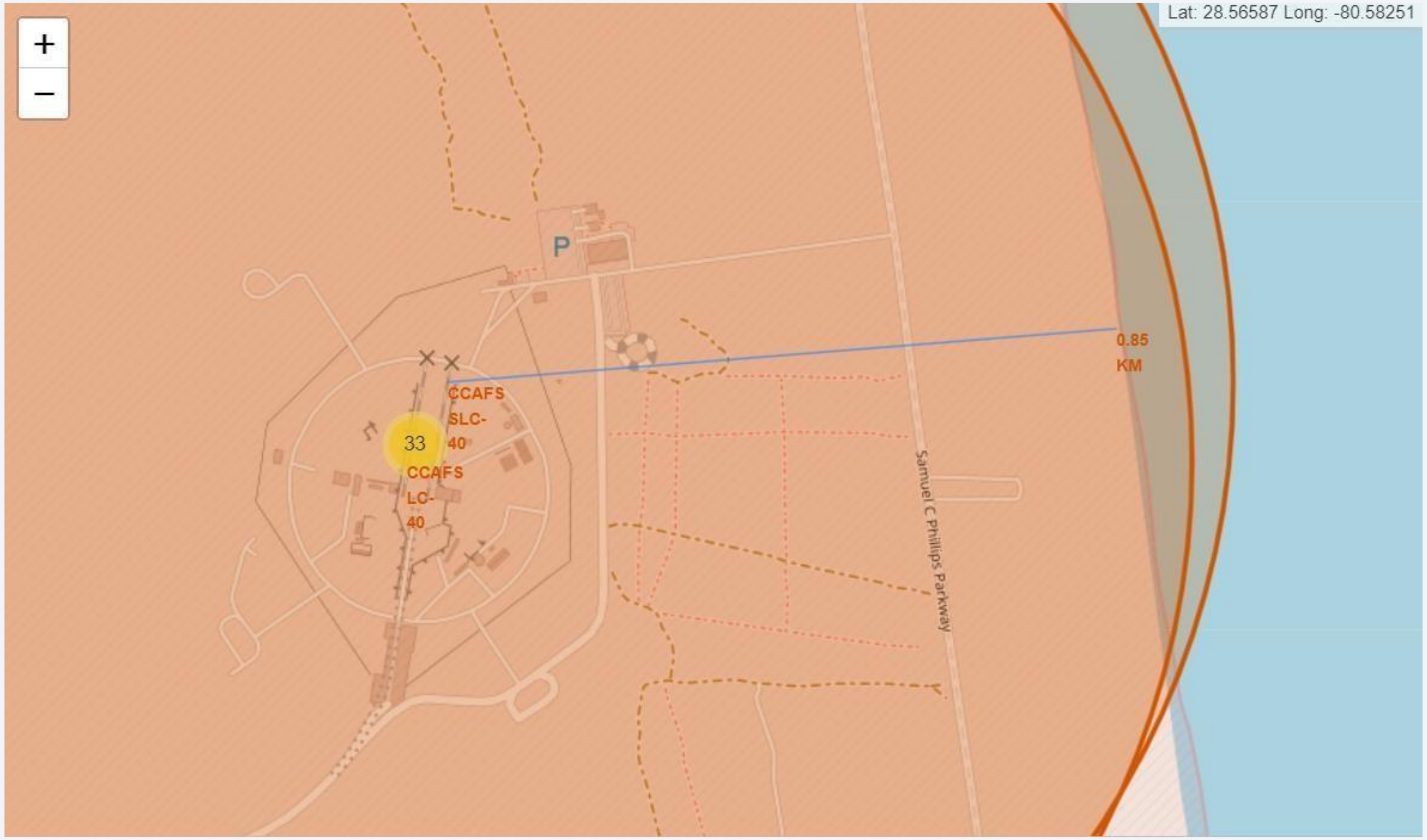
- 3 launch sites are located close to each other in Florida
- 1 launch site is loacated in Santa Maria

<Folium Map Screenshot 2>



- The most successful launch site is KSC LC-39A

<Folium Map Screenshot 3>

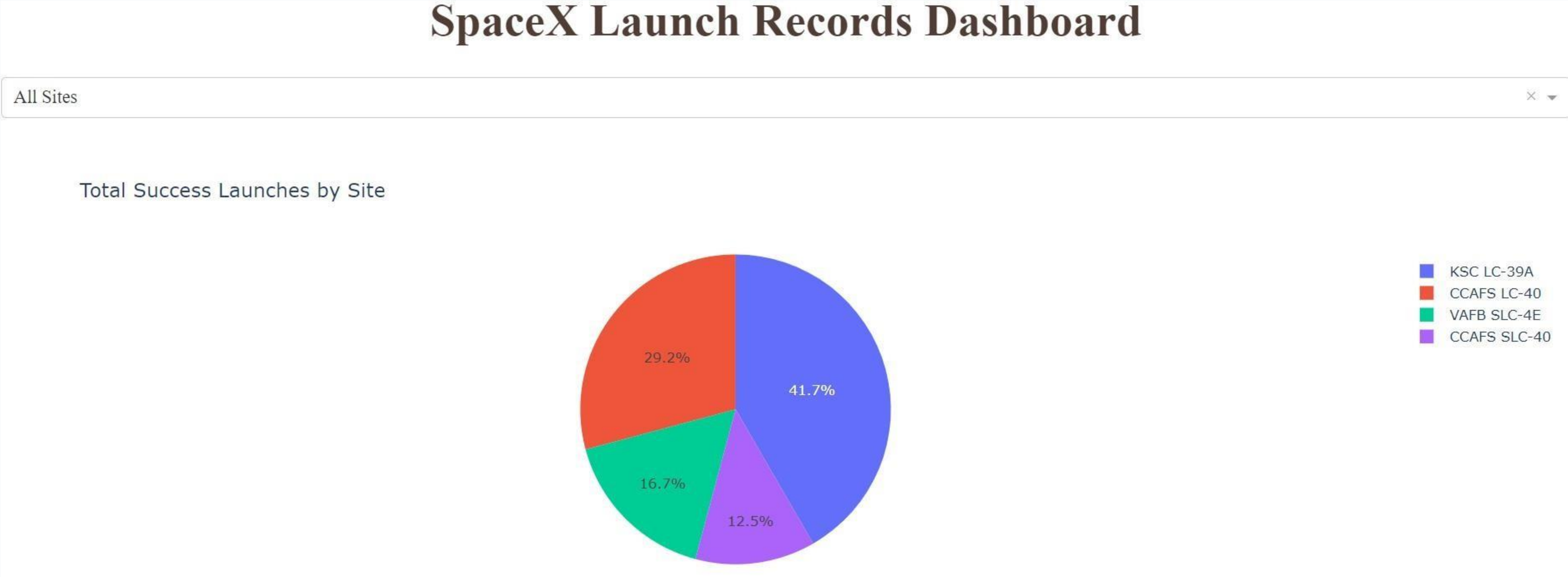


- The closed coastline to the launch site CCAFS SLC-40 is 850 meters away

Section 4

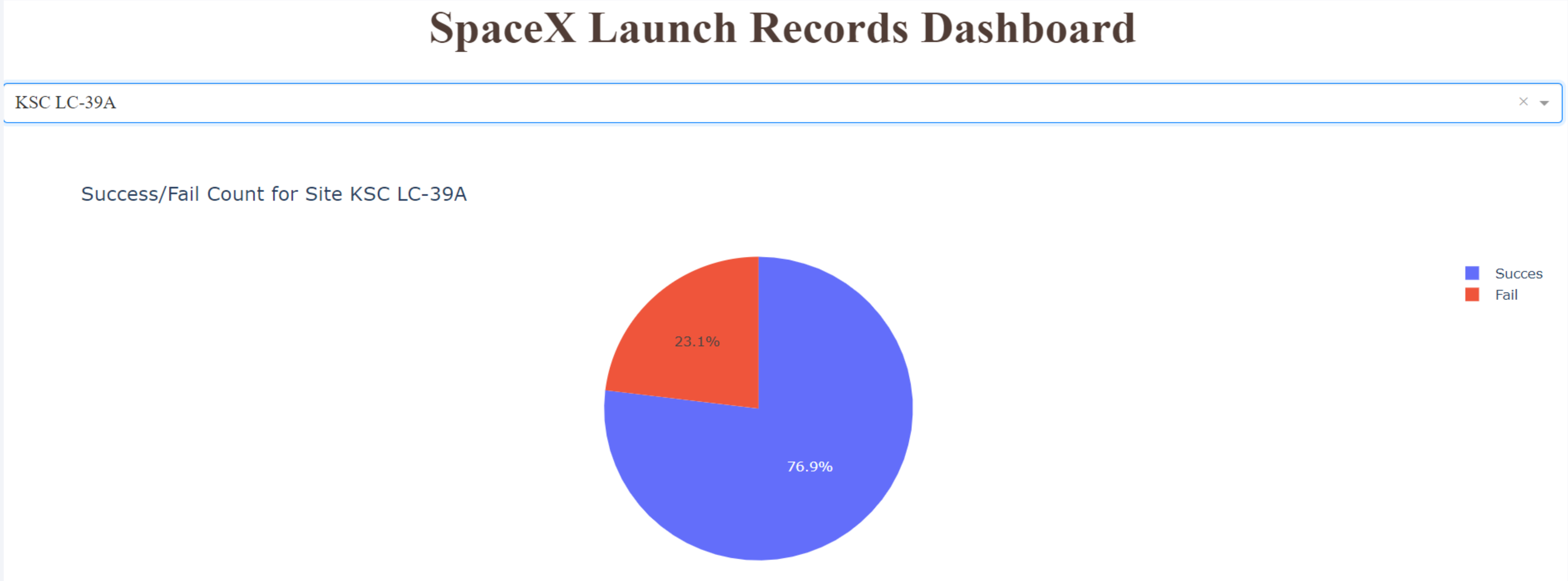
Build a Dashboard with Plotly Dash

Lauch Site Success Count



- The most successful site is KSC LC-39A
- The Second most successful site is CCAFS LC-40

Success Ratio of KSC LC-39A



- The most successful launch site has success ratio of 77%

Payload vs. Launch Outcome

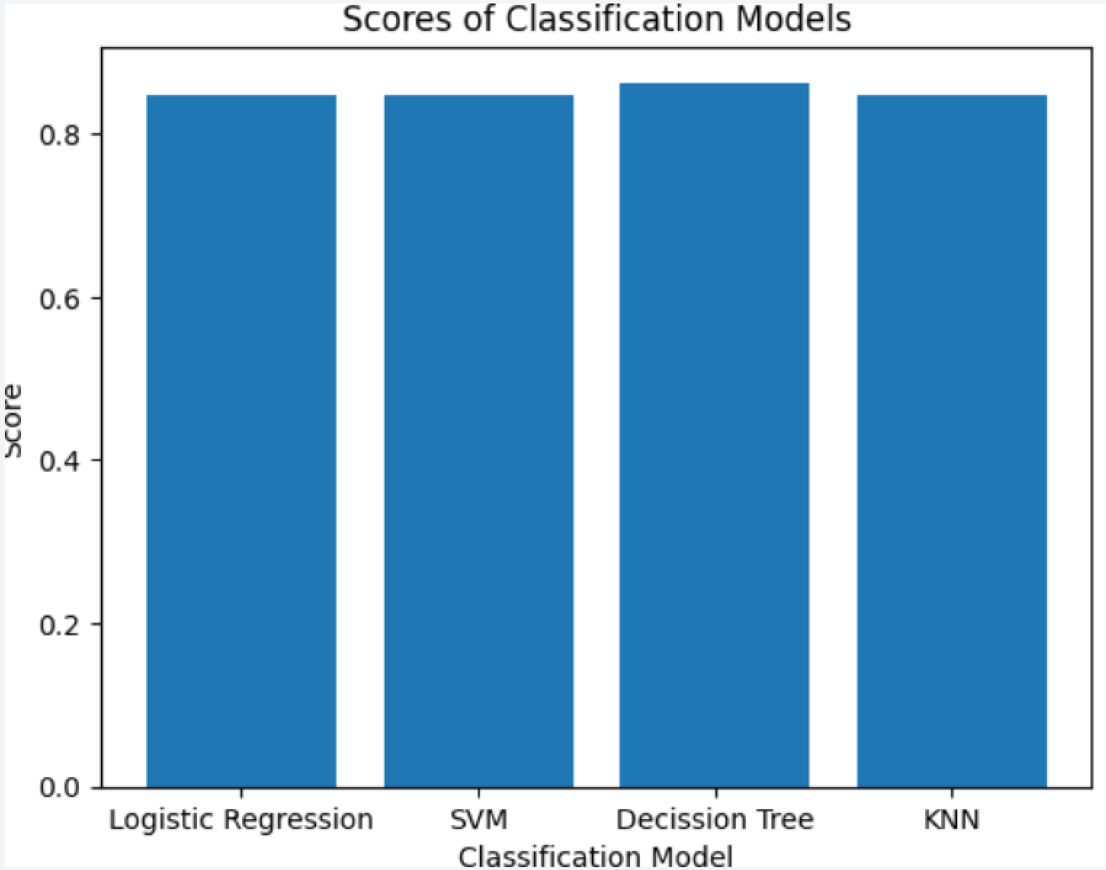
- Payload vs. Launch Outcome scatter plot for all sites, with payload between 1000 and 7000 kg

Section 5

Predictive Analysis (Classification)

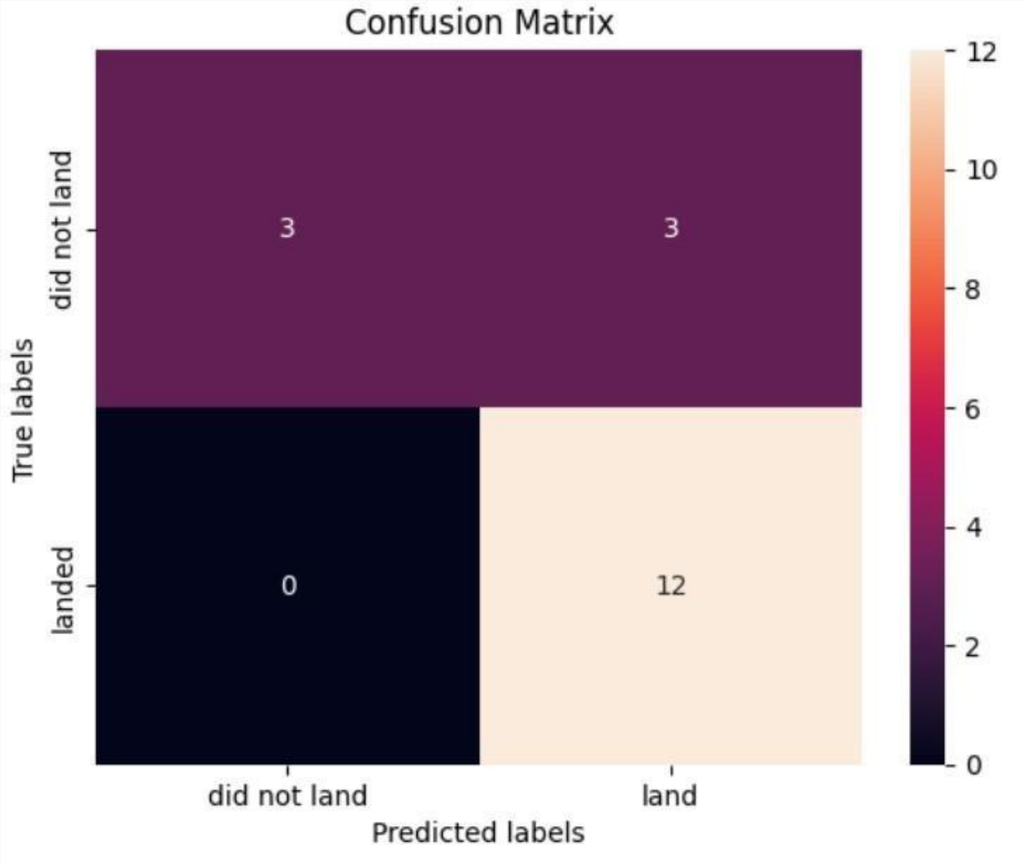
Classification Accuracy

- All classification models perform the same on test data but Decision Tree has the highest classification accuracy in train data:



Confusion Matrix

- The confusion matrix of the best performing model (Decision Tree) :
- The model can distinguish between the different classes. The major problem is false positives.



Conclusions

- The most successful orbit types are: ES-L1, GEO, HEO, SSO, VLEO
- The most successful launch sites are KSC LC-39A and CCAFS LC-40
- The most successful launch site (KSC LC-39A) has success ratio of 77%
- The success rate since 2013 kept increasing till 2020
- The site CCAFS SLC 40 has the most number flights
- The most commonly used orbit types are ISS and GTO
- All 4 launch sites are close to coastline
- The best performing model (Decision Tree) can distinguish between the different classes with high accuracy. The major problem is false positives.

Appendix

- You can find all everything about this project including the Python codes, SQL querries, and data in the following GitHub repository:
- https://github.com/franklinchuks/IBM-DATA-SCIENCE-LABS

Thank you!