IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<鄧巧婕>
<2025/3/10>

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward 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. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.
- Thus, the goal of this project is to predict if the Falcon 9 first stage will land successfully.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using the SpaceX API or web scraping
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium
- Perform predictive analysis using classification models
  - Create a machine learning pipeline to predict if the first stage will land
  - Find best Hyperparameter for KNN, SVM, Classification Trees and Logistic Regression

Data Collection

Data was collected using:

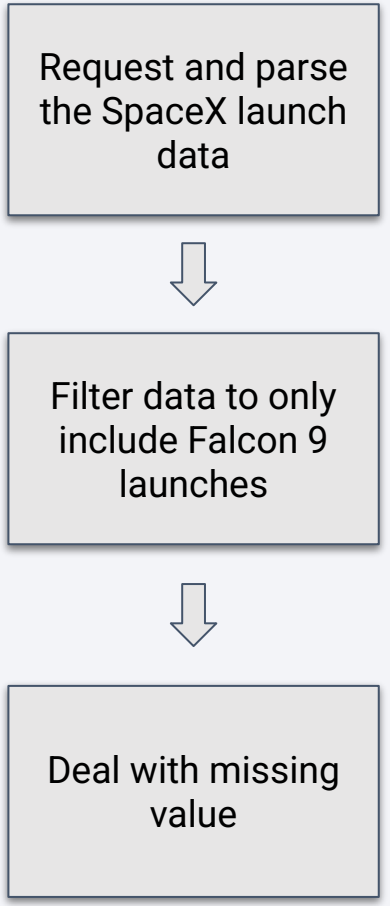
- SpaceX API (https://api.spacexdata.com/v4/rockets/)

- Web scraping (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)

Data Collection – SpaceX API

- We used the get request to the SpaceX API to collect data, clean the requested data and did some basic data wrangling and formatting.

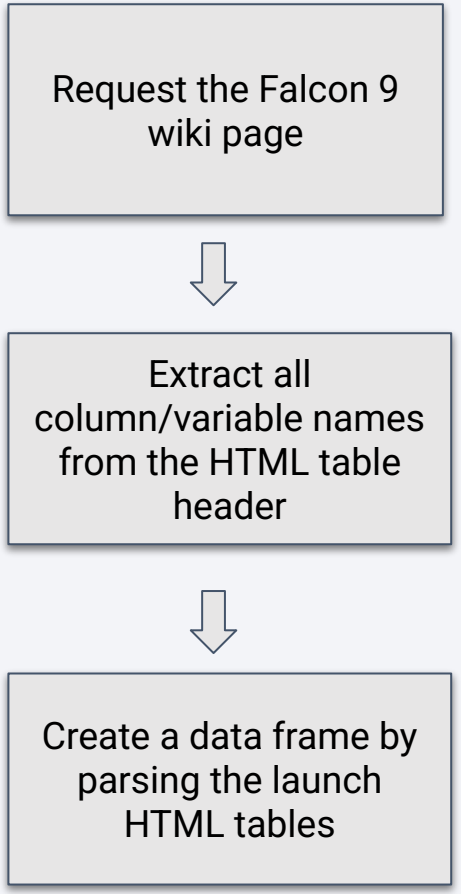
- GitHub URL https://github.com/christine1117/SpaceX/blob/main/spacex-data-collection-api.ipynb



Data Collection - Scraping

- Data could also be collected with web scraping from Wikipedia.

- GitHub URL https://github.com/christine1117/SpaceX/blob/main/webscraping.ipynb

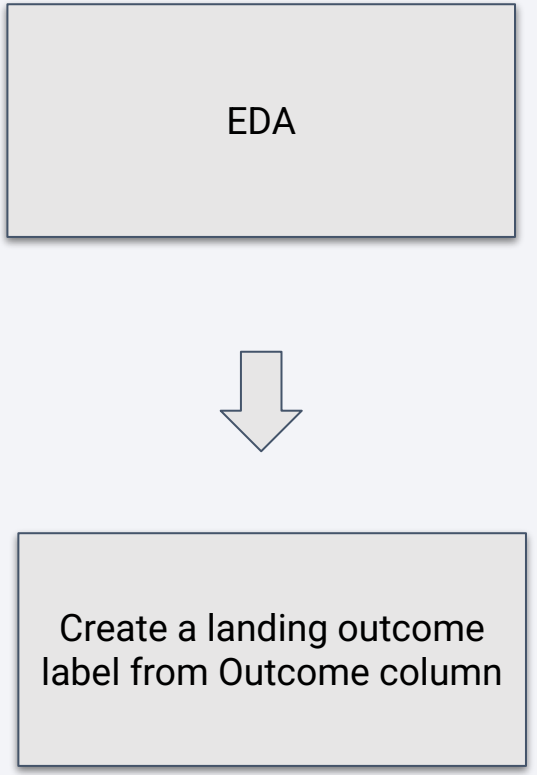


Data Wrangling

- Calculate the number of launches on each site, the number and occurrence of each orbit, and the number and occurence of mission outcome of the orbits

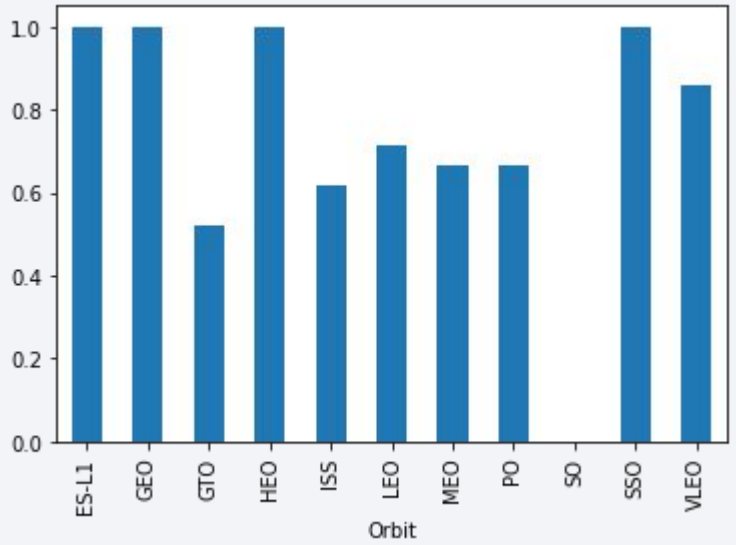
- Create a landing outcome. 0 for bad outcome, 1 otherwise.

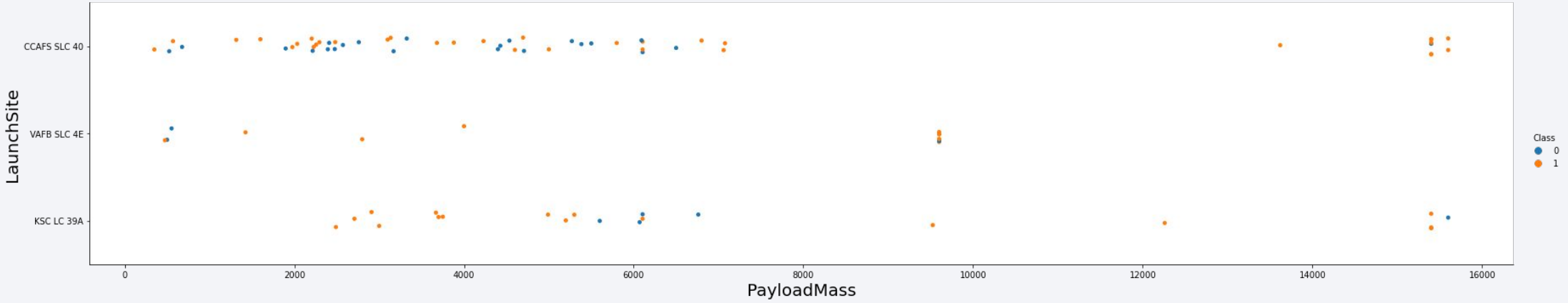
- GitHub URL https://github.com/christine1117/SpaceX/blob/main/spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Scatter plots and bar plots were used to visualize the relationship between two pairs of features.





EDA with SQL

- The following SQL queries were performed
  - the names of the unique launch sites in the space mission
  - 5 records where launch sites begin with the string 'CCA'
  - total payload mass carried by boosters launched by NASA (CRS)
  - average payload mass carried by booster version F9 v1.1
  - date when the first successful landing outcome in ground pad was achieved
  - the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - the total number of successful and failure mission outcomes
  - the names of the booster_versions which have carried the maximum payload mass.
  - the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
  - 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

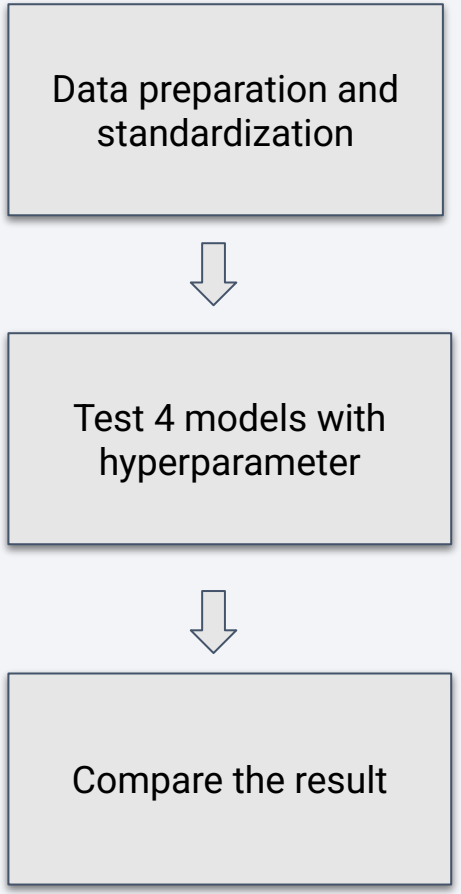
Build an Interactive Map with Folium

- Circle highlighted circle area with a text label on a specific coordinate
- Marker clusters simplify a map containing many markers having the same coordinate.
- Lines calculate the distances between two coordinates
- GitHub URL https://github.com/christine1117/SpaceX/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Predictive Analysis (Classification)

- Compare the result of four model with test accuracy: logistic regression, svm, decision tree, KNN

- GitHub URL https://github.com/christine1117/SpaceX/blob/main/Machine%20Learning%20Prediction_Part_5.ipynb

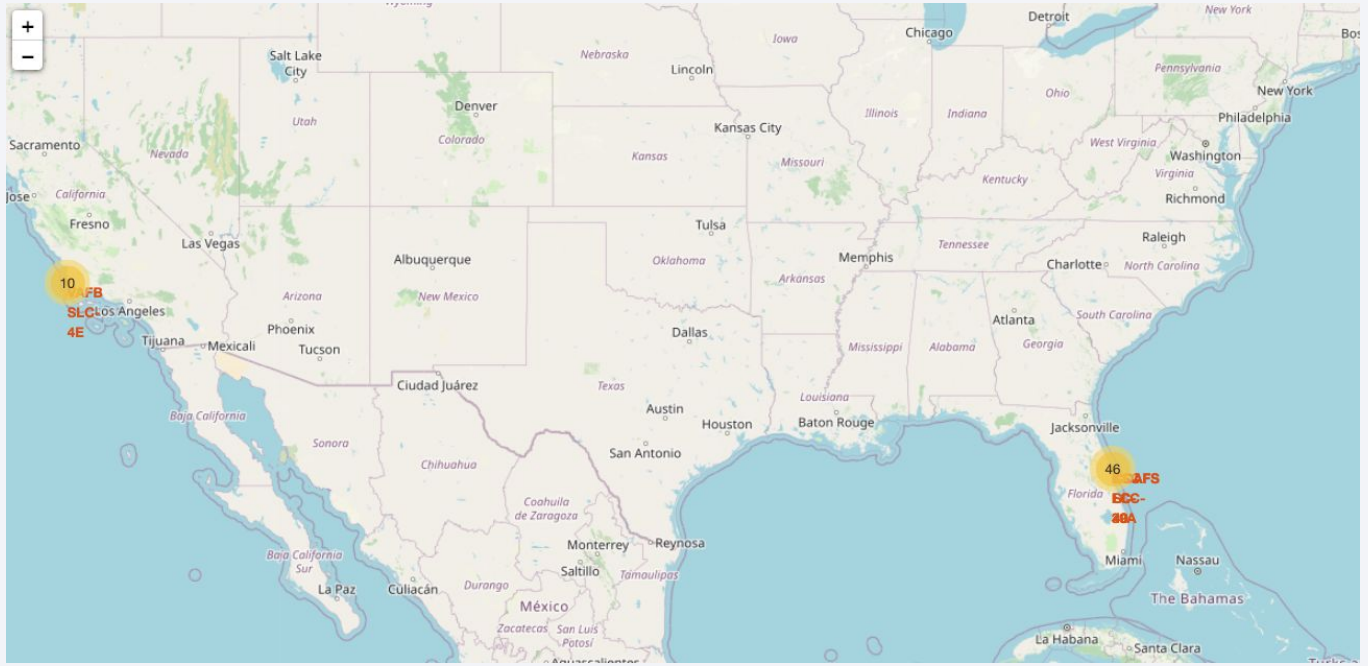


Results

- SpaceX used 4 different launch sites
- The rocket was launched from CCAFSLC-40 mostly by SpaceX 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 FT 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

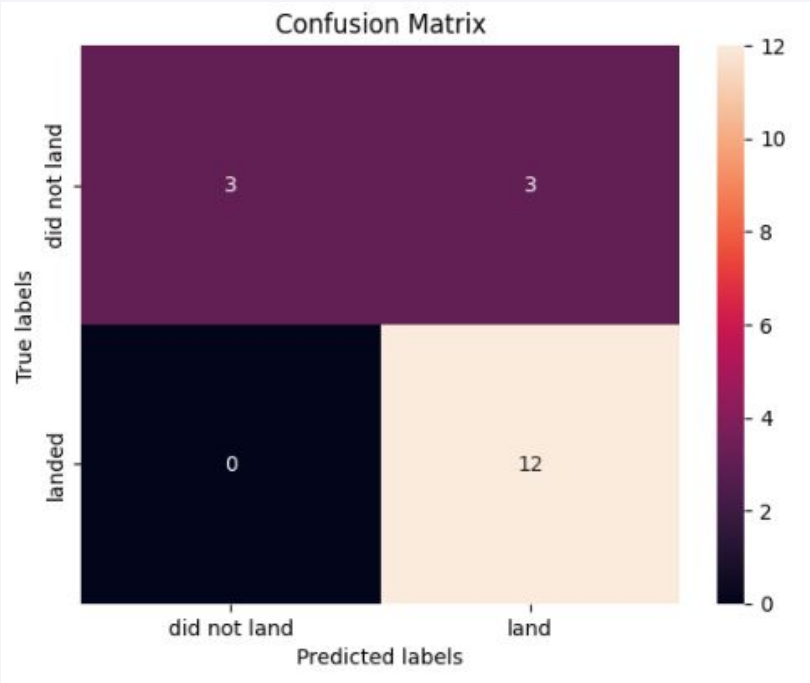
- Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea.



Results

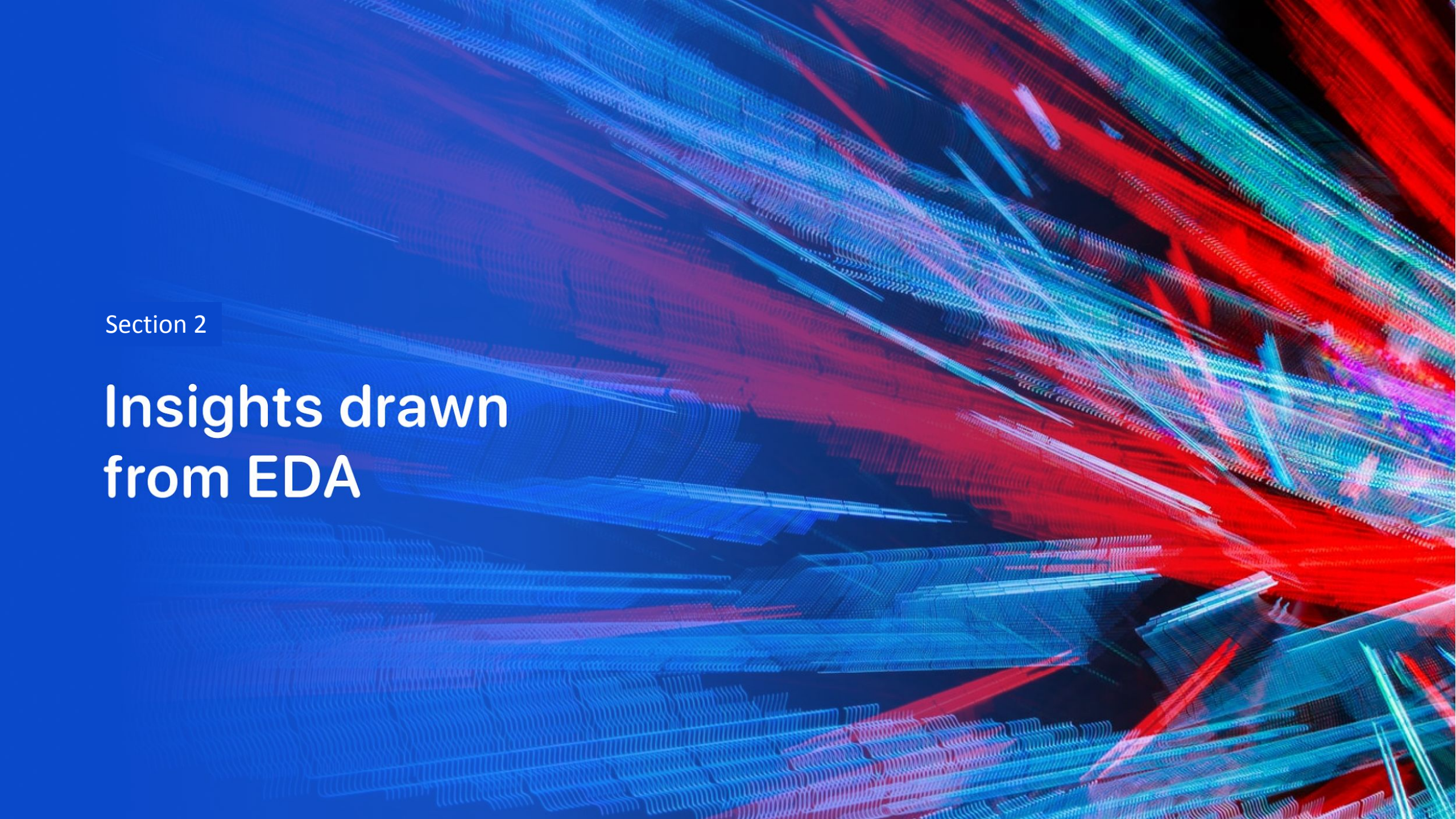
- Predictive analysis results showed that all models presented same accuracy.

Model	Accuracy	TestAccuracy
LogReg	0.84643	0.83333
SVM	0.84821	0.83333
Tree	0.8625	0.83333
KNN	0.84821	0.83333



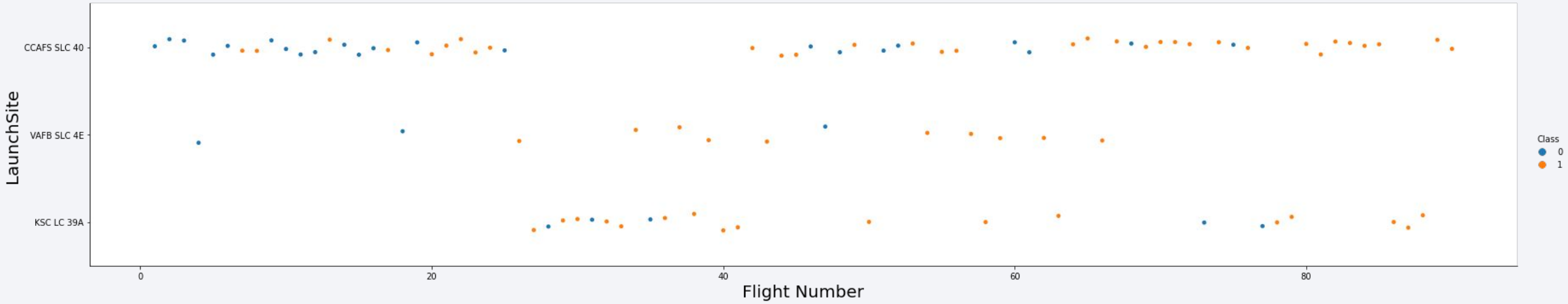
Section 2

Insights drawn from EDA



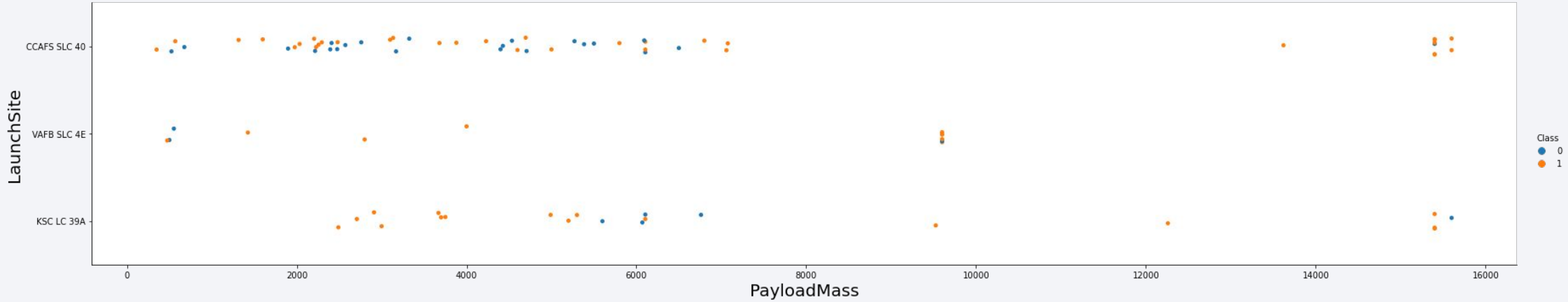
Flight Number vs. Launch Site

- CCAFS SLC 40 was the launch site that had the most flight number
- Successful outcomes increases over time



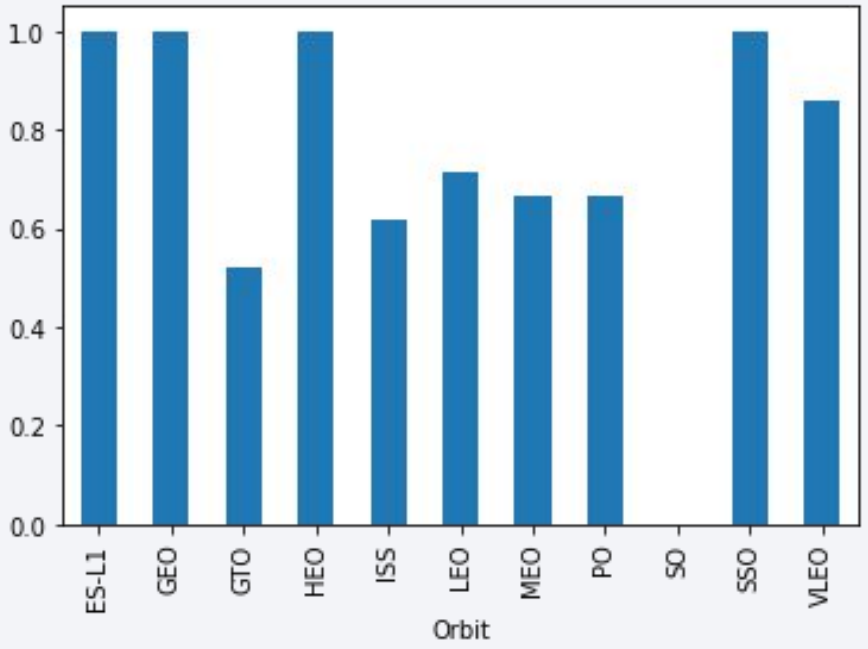
Payload vs. Launch Site

- The greater payload mass, the higher the likelihood of a successful landing outcomes
- The payload mass range was mostly between 200 and 7000



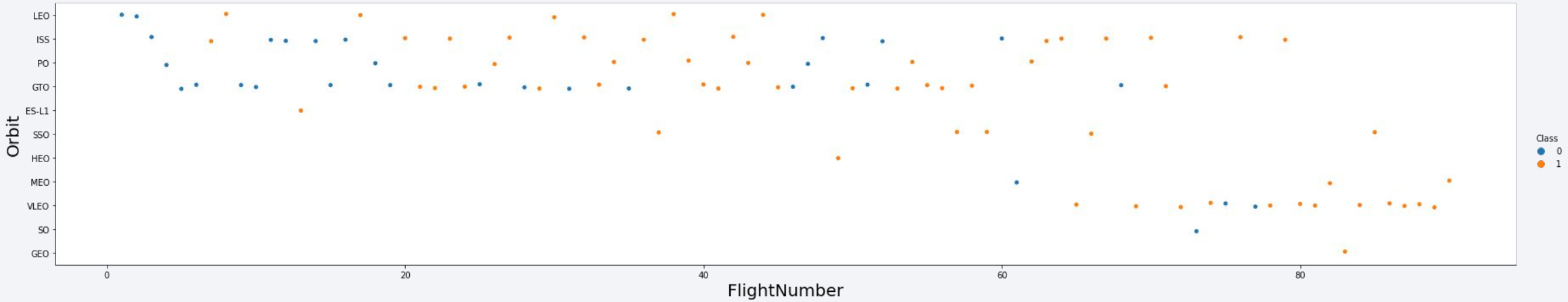
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO had a 100% success rate, whereas SO failed entirely.



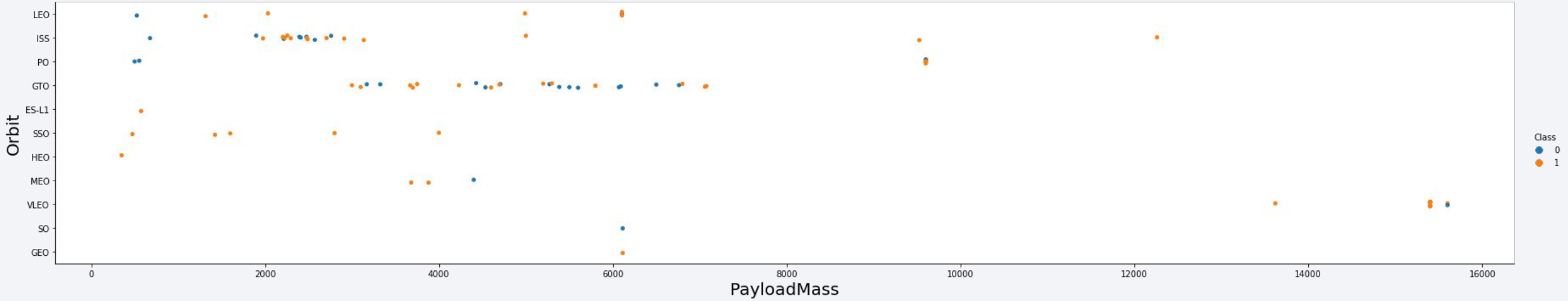
Flight Number vs. Orbit Type

- Most of the flights were concentrated around LEO, ISS, GTO, and VLEO
- VLEO is potential due to the highly successful outcome in recent years



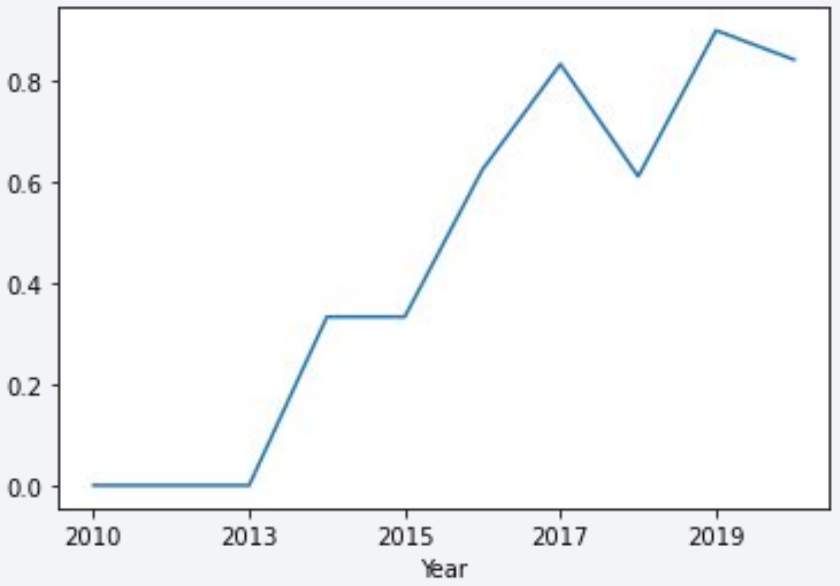
Payload vs. Orbit Type

- The flight with a VLEO orbit type had the highest payload mass
- ISS has the widest payload mass range



Launch Success Yearly Trend

- Between 2010 and 2013, the launches failed entirely. Over time, the success rate gradually increased, reaching 0.8 in 2020.



All Launch Site Names

- SpaceX used 4 different launch site

Launch Site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

Launch Site Names Begin with 'CCA'

- The rocket was launched from CCAFSLC-40 mostly by SpaceX and NASA.

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

- The average payload of F9 v1.1 booster is 111,268 kg

Total Payload (kg)
111.268

Average Payload Mass by F9 v1.1

- The average payload of F9 v1.1 booster is 2,928 kg

Avg Payload (kg)
2.928

First Successful Ground Landing Date

- The first success landing outcome happened in 2015 fiver year after the first launch

Min Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Many Falcon 9 FT booster versions were successful at landing in drone ships having payload above the average.

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

Total Number of Successful and Failure Mission Outcomes

- Almost 100% of mission outcomes were successful

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

Boosters Carried Maximum Payload

- Many F9 B5 booster version 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

- Two booster versions failed at landing in drone ships in 2015
  - F9 v1.1 B1012
  - F9 v1.1 B1015

Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

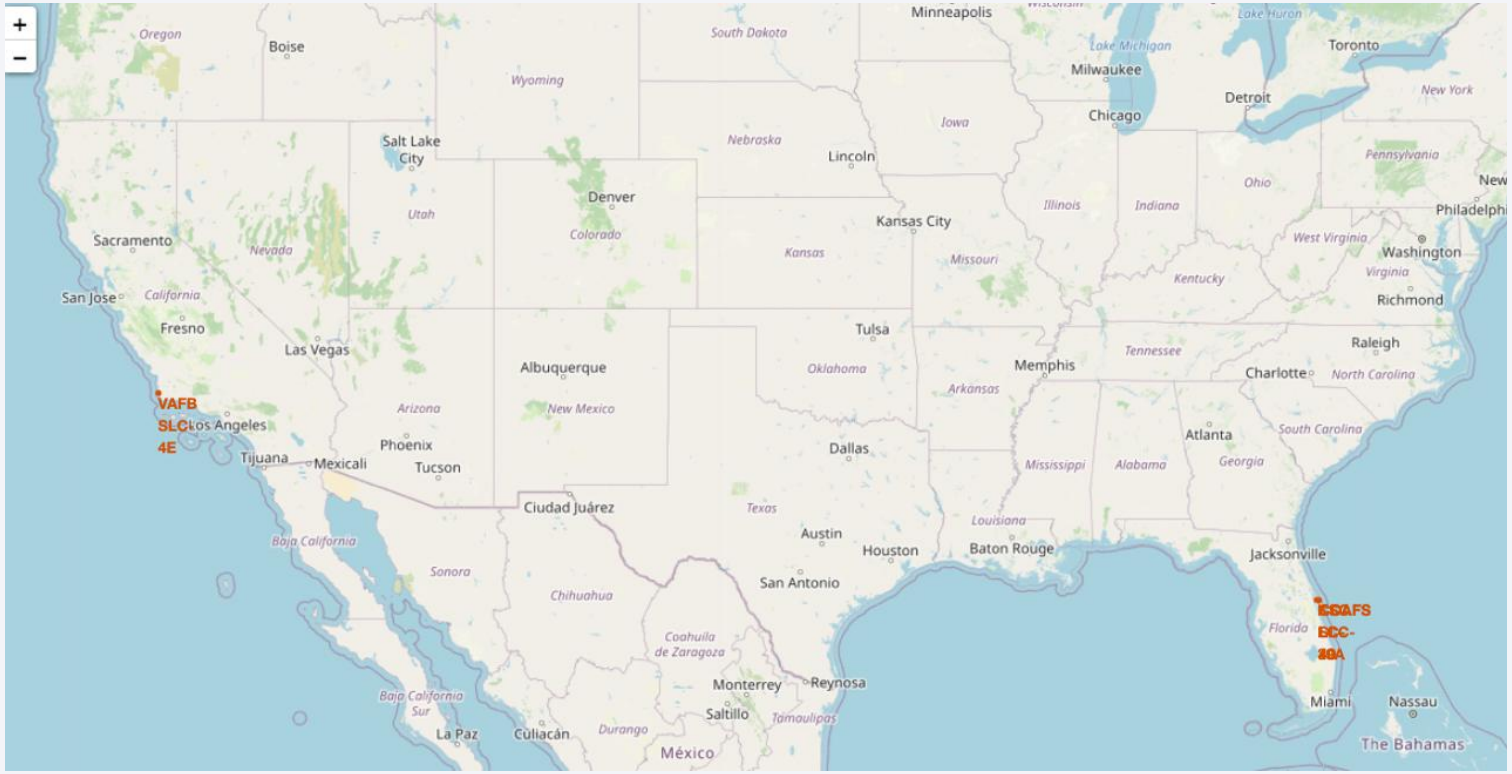
- No attempt had occurred the most within 7 years

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

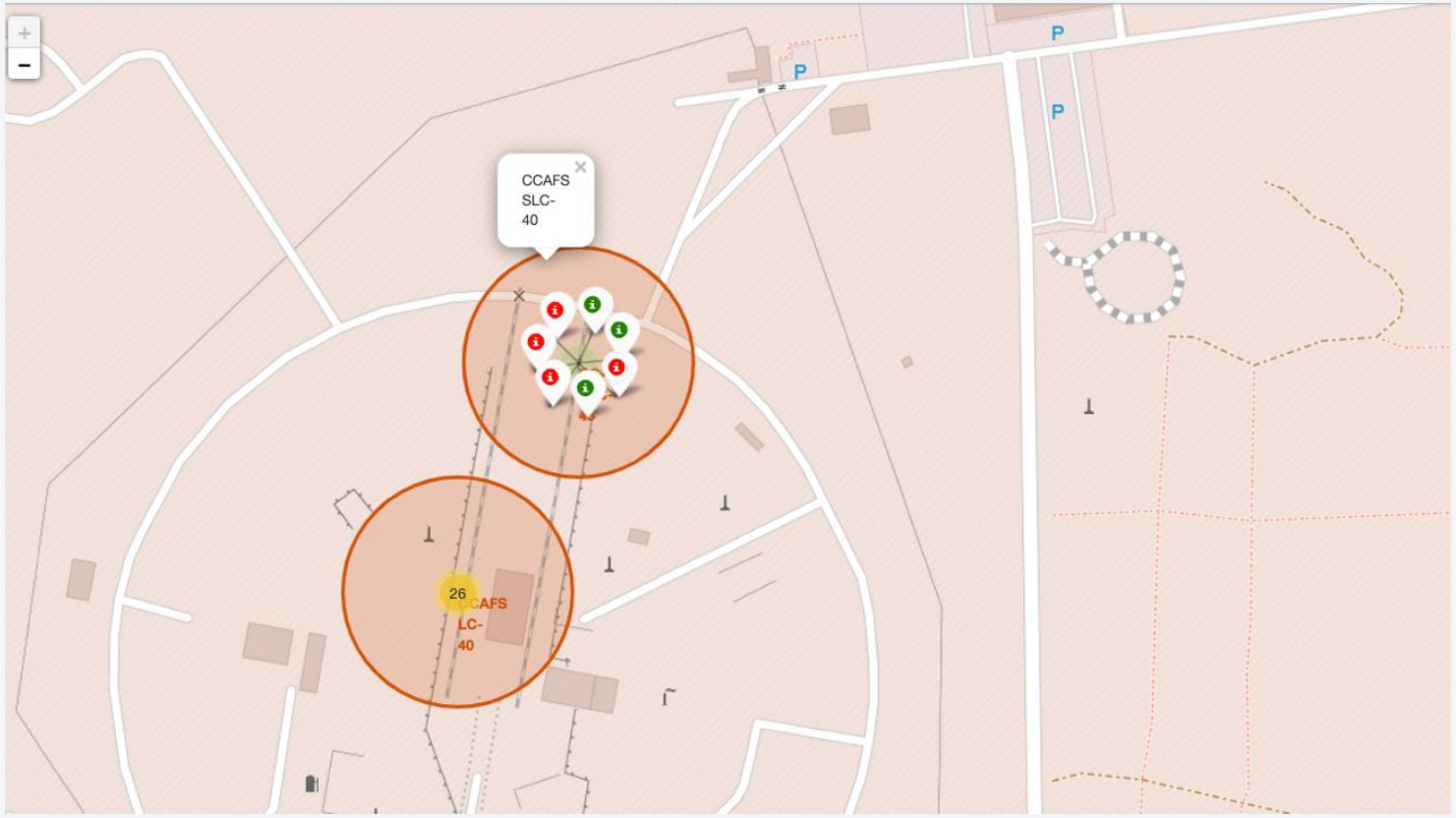
All launch sites

- Launch sites are near sea



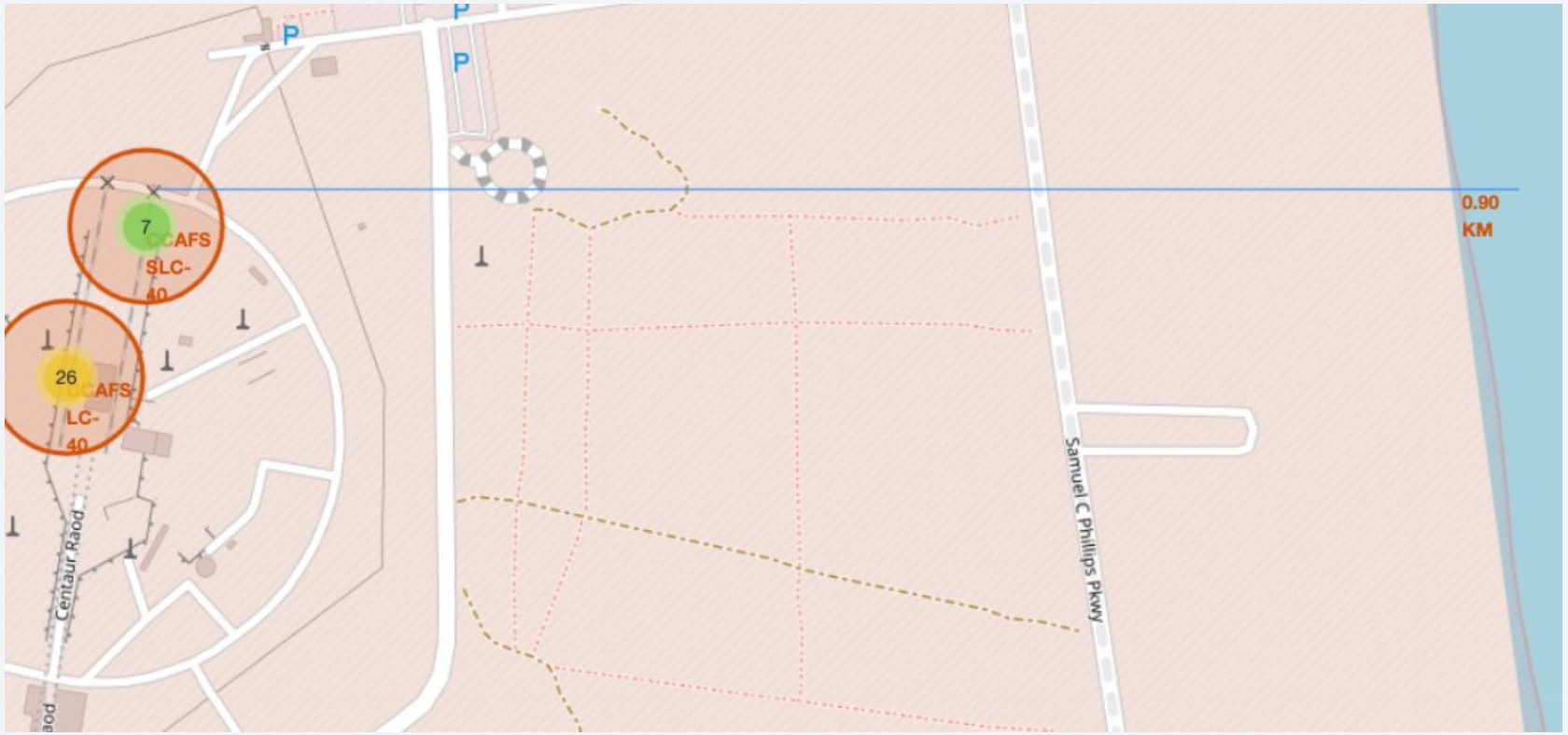
Launch outcomes by site

- Green marks the successful outcomes, and red otherwise.



The proximities of launch sites

- The CCAFSSLC-40 is 0.9km far away from the coastline, also being close to railroad and highway

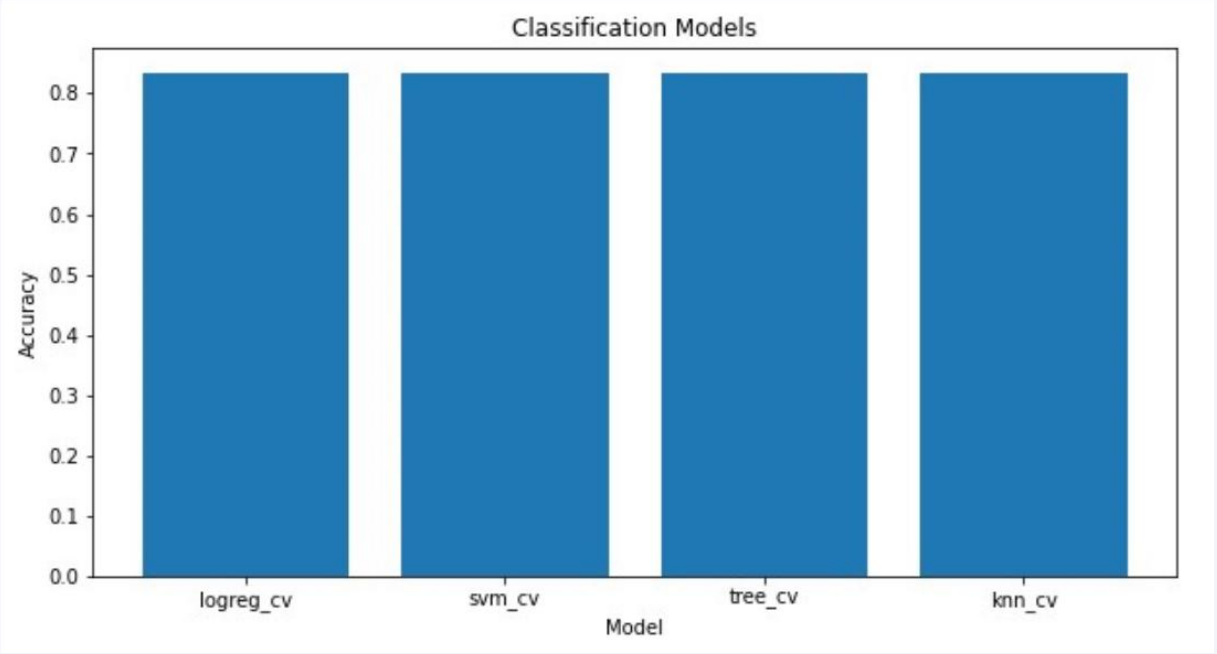


Section 4

Predictive Analysis (Classification)

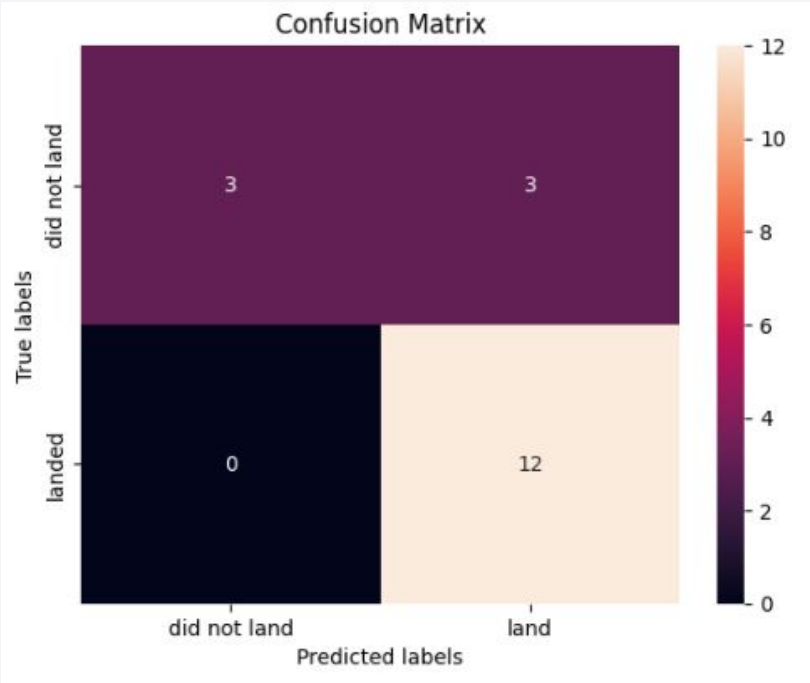
Classification Accuracy

- Four models have same accuracy



Confusion Matrix

- Four model has the same accuracy, which means their confusion matrix was exactly the same.



Conclusions

- All the machine learning model presents the same accuracy.
- KSC LC-39A is the launch site where has most successes generally, while CCAFS SLC 40 may be the best in recent years.
- The greater payload mass, the higher the likelihood of a successful landing outcomes

Appendix

- GitHub repository: https://github.com/christine1117/SpaceX.git

Thank you!