IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Syed Ahmed Naushad
29 August, 2023

Outline

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

Executive Summary

- Methodologies used to analyze data were:
  - Data Collection using the SpaceX API and Web Scraping
  - Exploratory Data Analysis (EDA) through SQL queries and Interactive Data Visualization
  - Machine Learning model training to predict future outcomes
- Results Obtained:
  - Important features for the successful landing of the first stage were discovered
  - Prediction of future outcomes is possible with an accuracy of around 84%

Introduction

The objective of this research was to evaluate the viability of a new aerospace manufacturer, SpaceY, to compete with SpaceX. For this, the data available on SpaceX was analyzed to get a general idea on the best practices to take in order to avoid loses.

The main focus was to determine:

- What factors result in the successful landing of the first stage of rockets to avoid additional costs
- What the best place to launch the rockets would be

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
    - Data was collected using API and WebScraping
- Perform data wrangling
    - Features were converted into integers and training labels were determined
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Multiple models were trained using different hyperparameters to determine the best parameters; The accuracy of the models was compared to find the best one

Data Collection

Describe how data sets were collected. Data was collected from two sources:

1. SpaceX API (https://api.spacexdata.com/v4/rockets/)
2. WebScraping the Falcon 9 Wikipedia Article

Data Collection – SpaceX API

- GET requests were made on the SpaceX REST API to obtain the launch data on all falcon rockets
- The data was then filtered and cleaned to suit the needs

(Jupyter Notebook for Reference)

API Request

The SpaceX REST API was called to get data in JSON format

Filtering

The data was filtered to keep only Falcon 9 launches

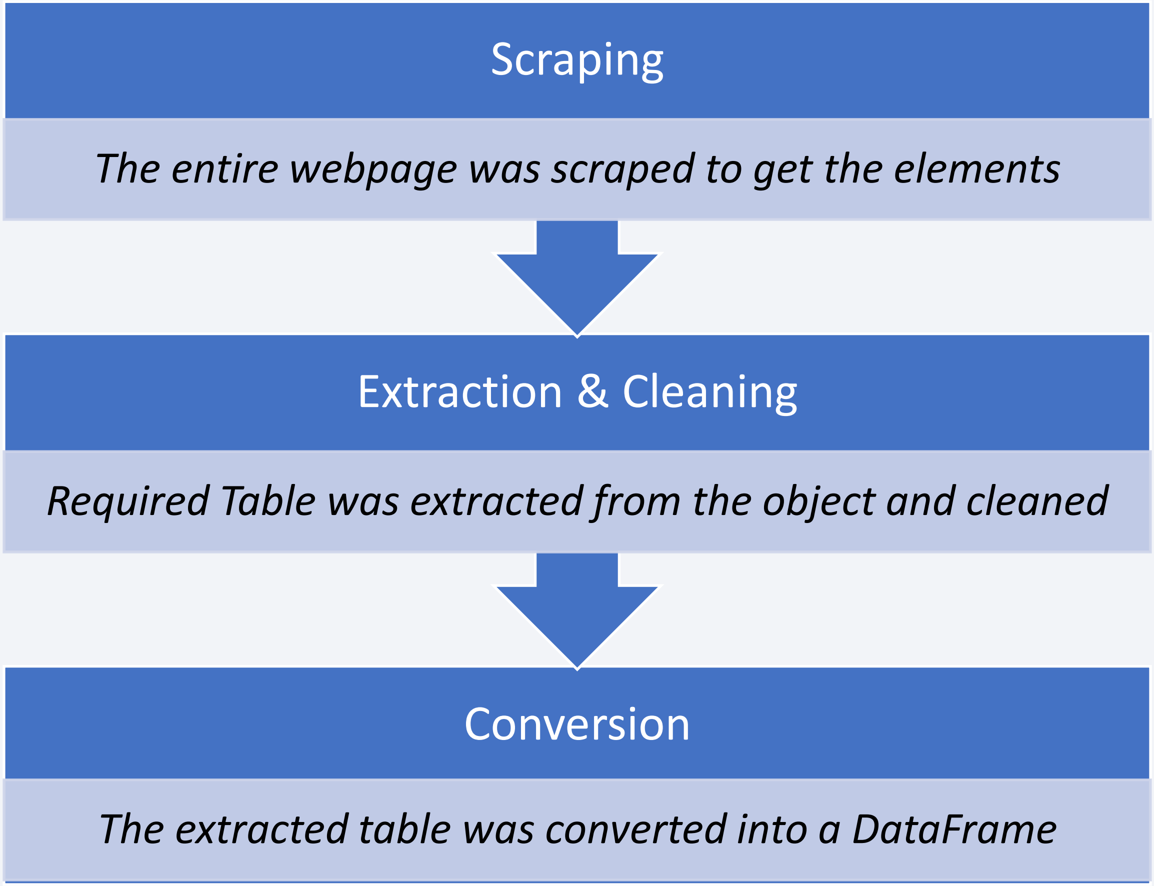
Fixing Missing Values

The missing values were filled with the overall average

Data Collection - Scraping

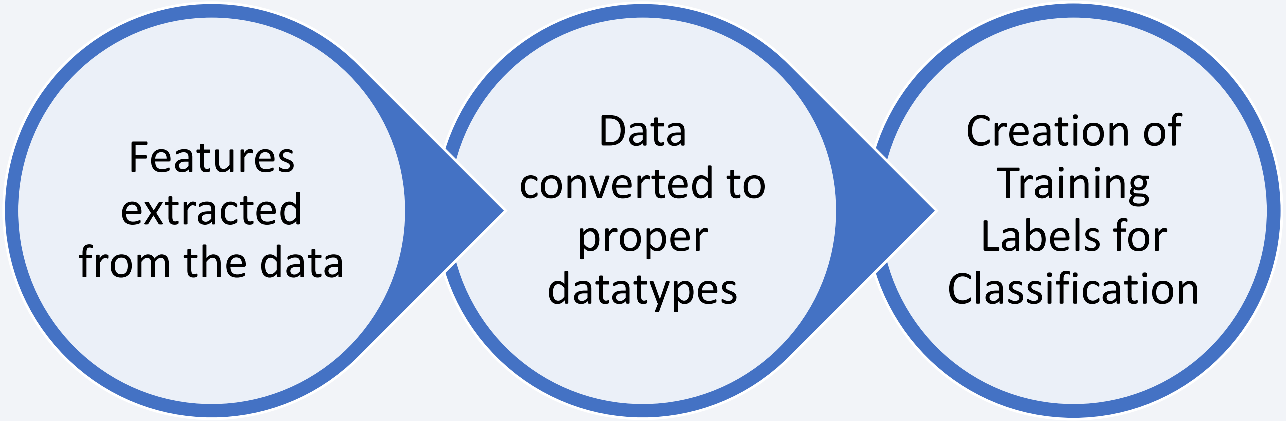
- The Wikipedia page was scraped and turned into an object
- Falcon 9 table was extracted from the object, cleaned, then converted into a DataFrame

(*[Jupyter Notebook for Reference]*)



Data Wrangling

Once all the necessary data was in hand, it was cleaned to get the necessary features. The features were standardized into the same format and training labels were determined



(*Jupyter Notebook for Reference*)

EDA with Data Visualization

Features of the data were used to plot various graphs such as:

- Scatter Plot between the *Flight Number* and *Launch Site*
- Scatter Plot between *Payload* and *Launch Site*
- Bar Plot between success rate of each orbit type
- Scatter Plot between *Flight Number* and *Orbit Type*
- Scatter Plot between *Payload* and *Orbit Type*
- Line Plot to visualize yearly trend of success rate

(*Jupyter Notebook for Reference*)

EDA with SQL

SQL queries were used to perform further EDA. The queries used were:

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

(*Jupyter Notebook for Reference*)

Build an Interactive Map with Folium

Folium was used to create an interactive map that encompassed all Falcon 9 launch sites. The map also contained elements such as:

- **Markers to** indicate points like launch sites
- **Circles** to indicate highlighted areas around specific coordinates, like NASA Johnson Space Center
- **Marker clusters to** indicate groups of events in each coordinate, like launches in a launch site
- **Lines** to indicate distances between two coordinates

(*Jupyter Notebook for Reference*)

Build a Dashboard with Plotly Dash

A Launch Records Dashboard was creating using the different Launch Sites and a range of Payload Mass.

The aim of the Dashboard was to provide key information regarding:

- The success rate of each Launch Site
- How the Payload effected the success rate

(*Source Code for Reference*)

Predictive Analysis (Classification)

Data was normalized and the class label was used to train four different classification models, which were further enhanced using different hyperparameters.

The Models were then compared to find the best accuracy

(*Jupyter Notebook for Reference*)

Standardization

- Data was prepared and standardized

Training

- Models were trained with the data and different hyperparameters

Comparison

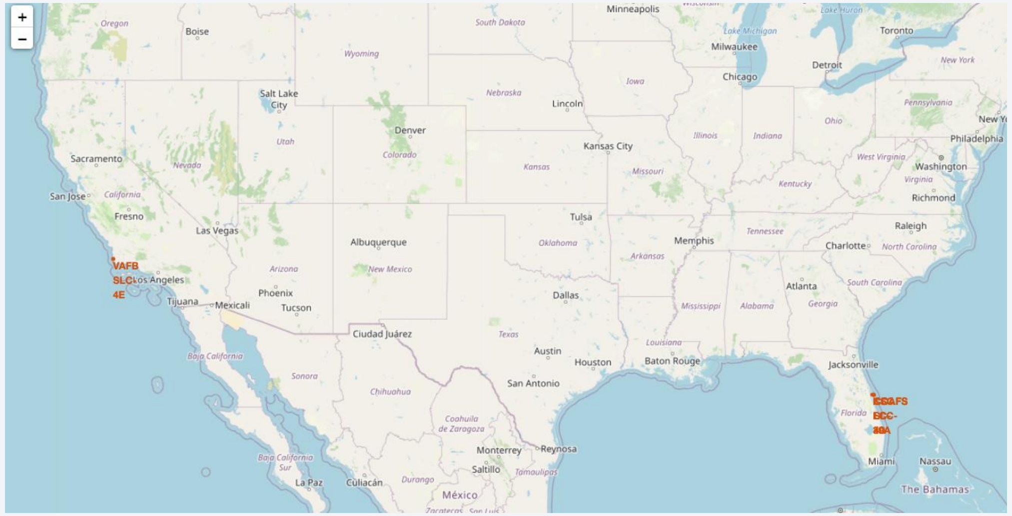
- The accuracy of the different models was compared to find the best one

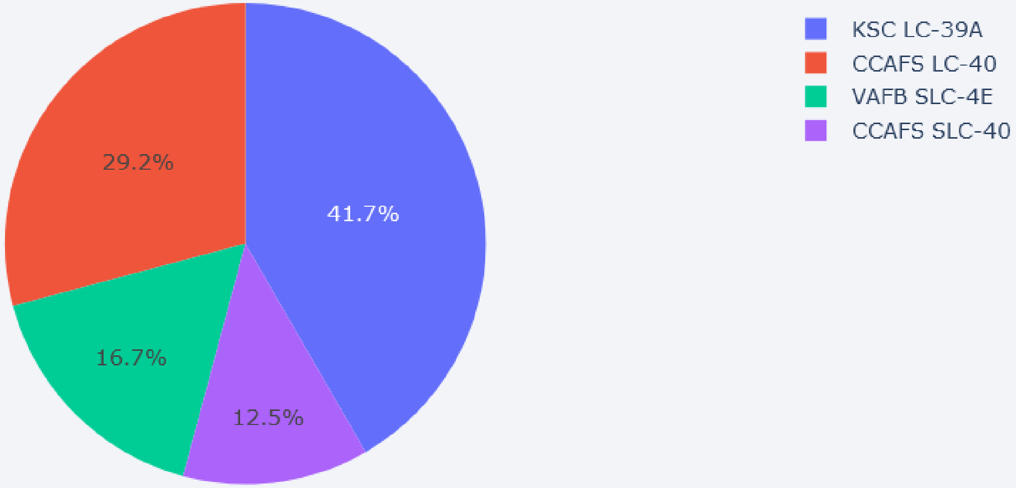
Results

- **Exploratory data analysis results**
  - The average payload mass carried by booster version F9 v1.1 is 2,928 kg
  - Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average
  - Around 98% of mission outcomes were successful
  - The number of landing outcomes became as better as years passed

Results

- **Interactive analytics result**
  - All four launch sites are located near coastlines and have good logistics
  - Most of the sites are in the east coast
  - KSC LC-39A had the most successful launches





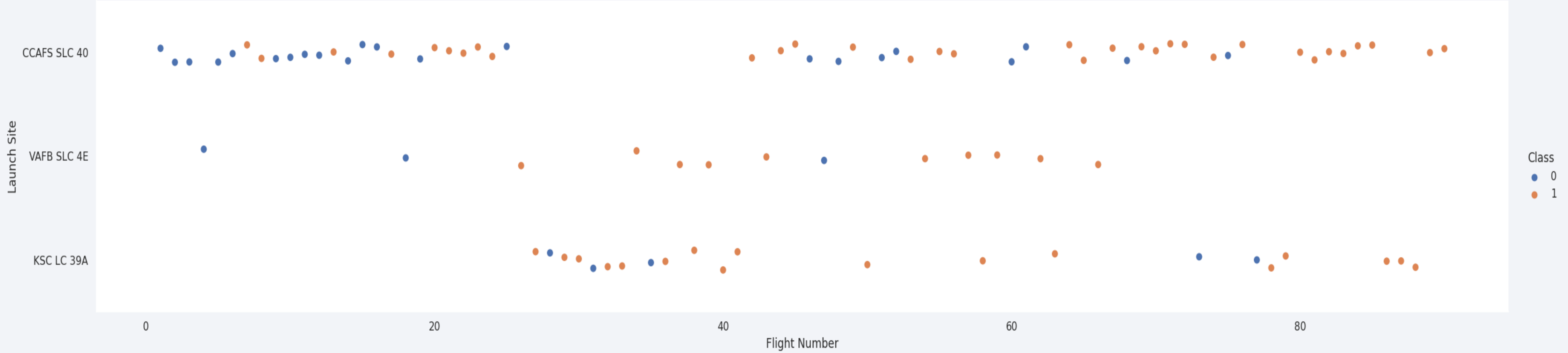
Results

- **Predictive analysis results**
  - All four models, Logistic Regression, Support Vector Machine (SVM), Decision Tree, and KNN, had a similar accuracy score.
  - It is possible to predict future stage one landing outcomes with an accuracy score of 84%

Section 2

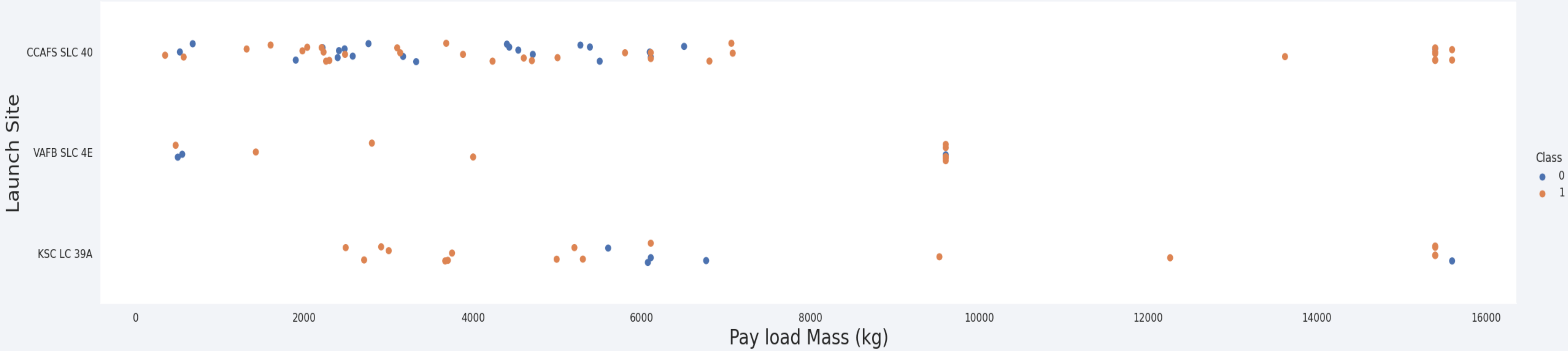
Insights drawn from EDA

Flight Number vs. Launch Site



- The best launch site seems to be CCAF5 SLC-40 with the most success rates, followed by KSC LC-39A then VAFB SLC-4E
- The overall success rate seems to be increasing with the trials

Payload vs. Launch Site

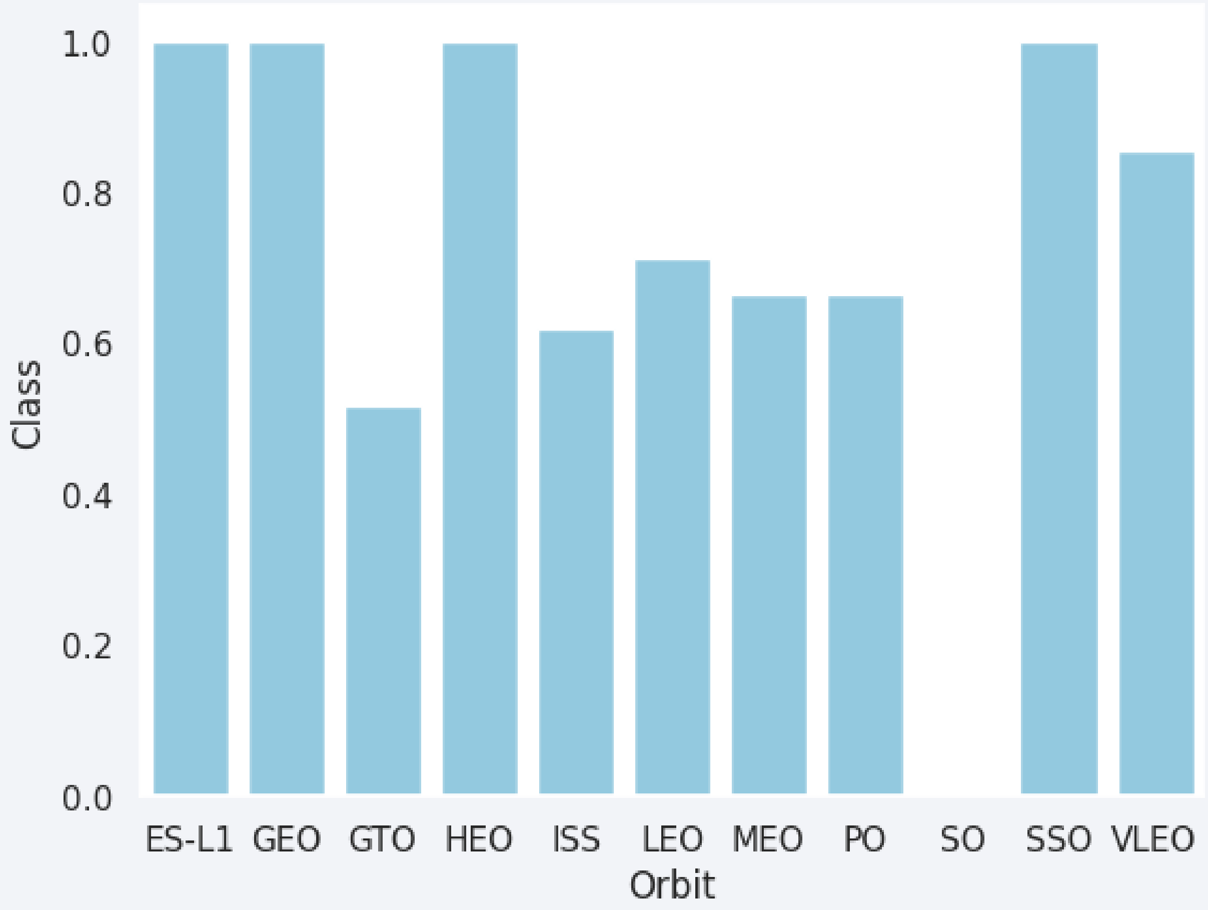


- There is an increase in success rate along with Payload Mass
- For the VAFB-SLC launch-site there are no rockets launched for Heavy Payload mass(greater than 10000)

Success Rate vs. Orbit Type

- Launches in the orbit ES-L1, GEO, HEO, and SSO have a 100% success rate

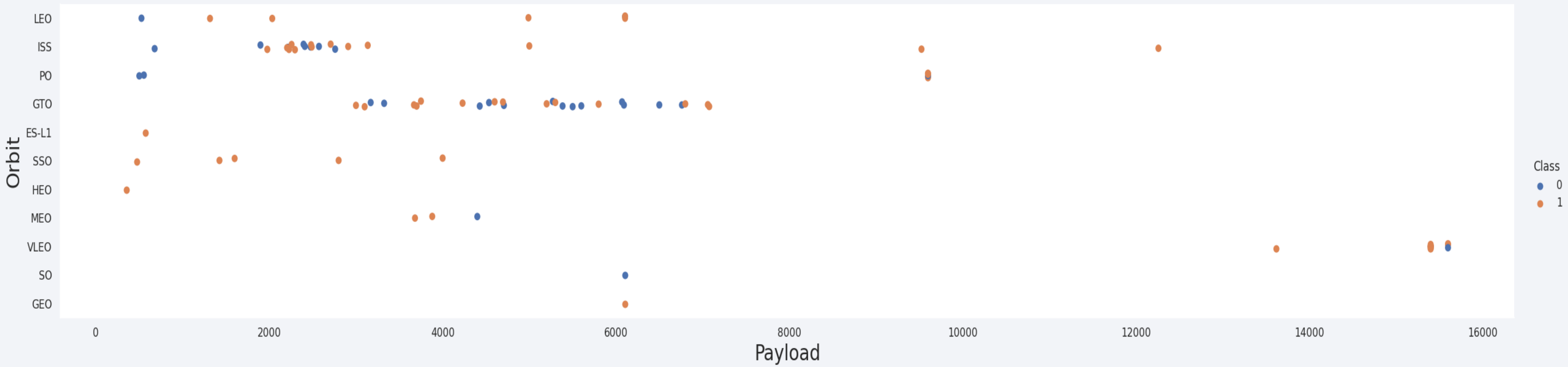
- VLEO and LEO also show 80% and 70% success rate respectively



Flight Number vs. Orbit Type

- 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

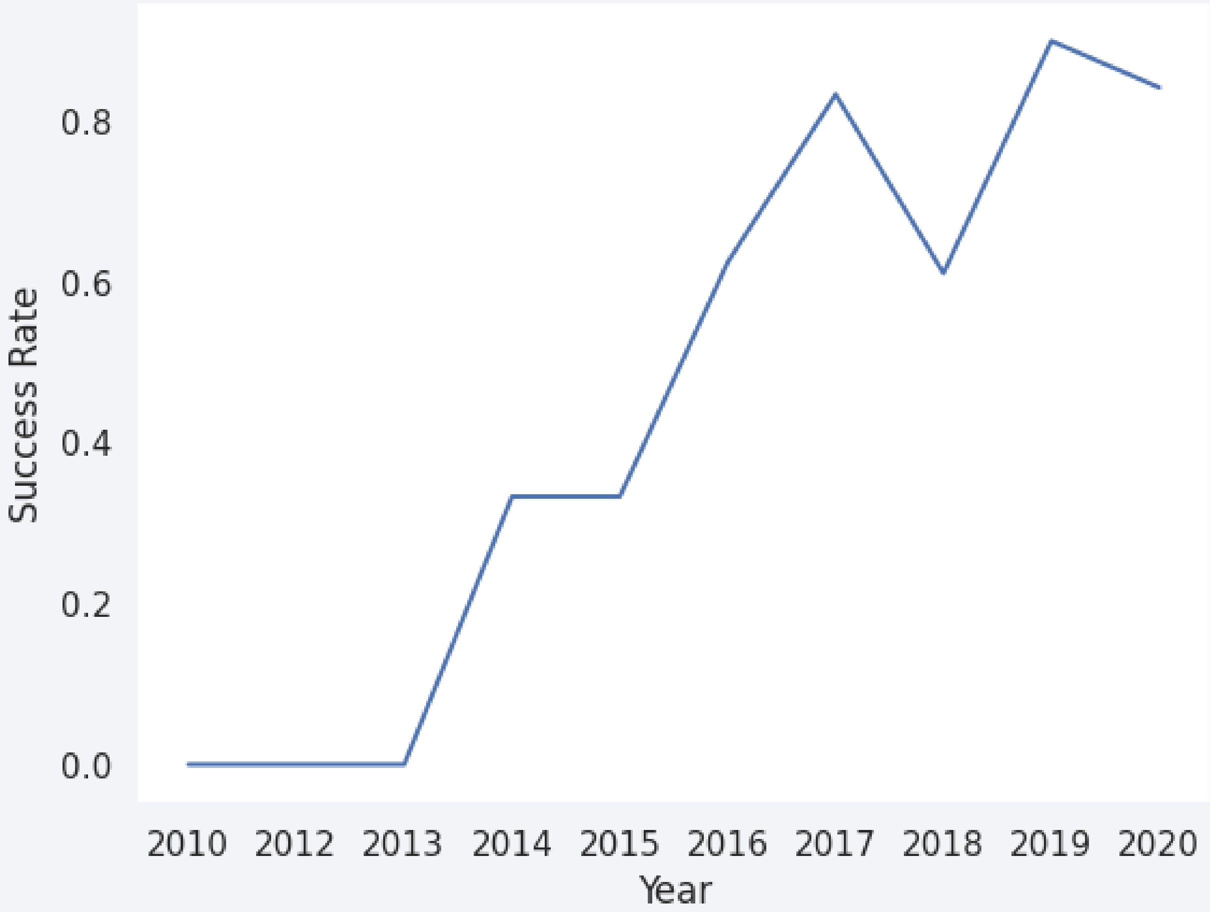
Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS
- However, for GTO, we cannot distinguish this well as both positive landing rate and negative landing (unsuccessful mission) are both there here

Launch Success Yearly Trend

- This line plot shows that the success rate since 2013 kept increasing till 2020



All Launch Site Names

- There are four Launch Sites:

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

- These were obtained from the 'Launch_Site' column of the data

Launch Site Names Begin with 'CCA'

- These are first five Launch Sites that 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	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 attemp

Total Payload Mass

- The Total Payload Mass carried by the boosters by Nasa is **111268 Kg**
- This was calculated by taking the sum of the boosters starting with ‘CRS’, that is, the boosters launched by NASA

Average Payload Mass by F9 v1.1

- The Average Payload Mass carried by booster version F9 v1.1 is **2534.667 Kg**
- This was calculated by taking the sum of the Payload mass carried by the booster version F9 v1.1 and dividing it by the number of F9 v1.1 boosters launched

First Successful Ground Landing Date

- The first successful landing outcome on ground pad was on **22nd December, 2015**
- By filtering data by successful landing outcome on ground pad and getting the minimum value for date it's possible to identify the first occurrence

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are:

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 are as follows:

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

Boosters Carried Maximum Payload

- The twelve boosters that have carried the maximum payload are:

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

Booster Version (…)
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 the year 2015 are:

Month	Landing Outcome	Booster Version	Launch Site
10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- This result was obtained using ta subquery

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

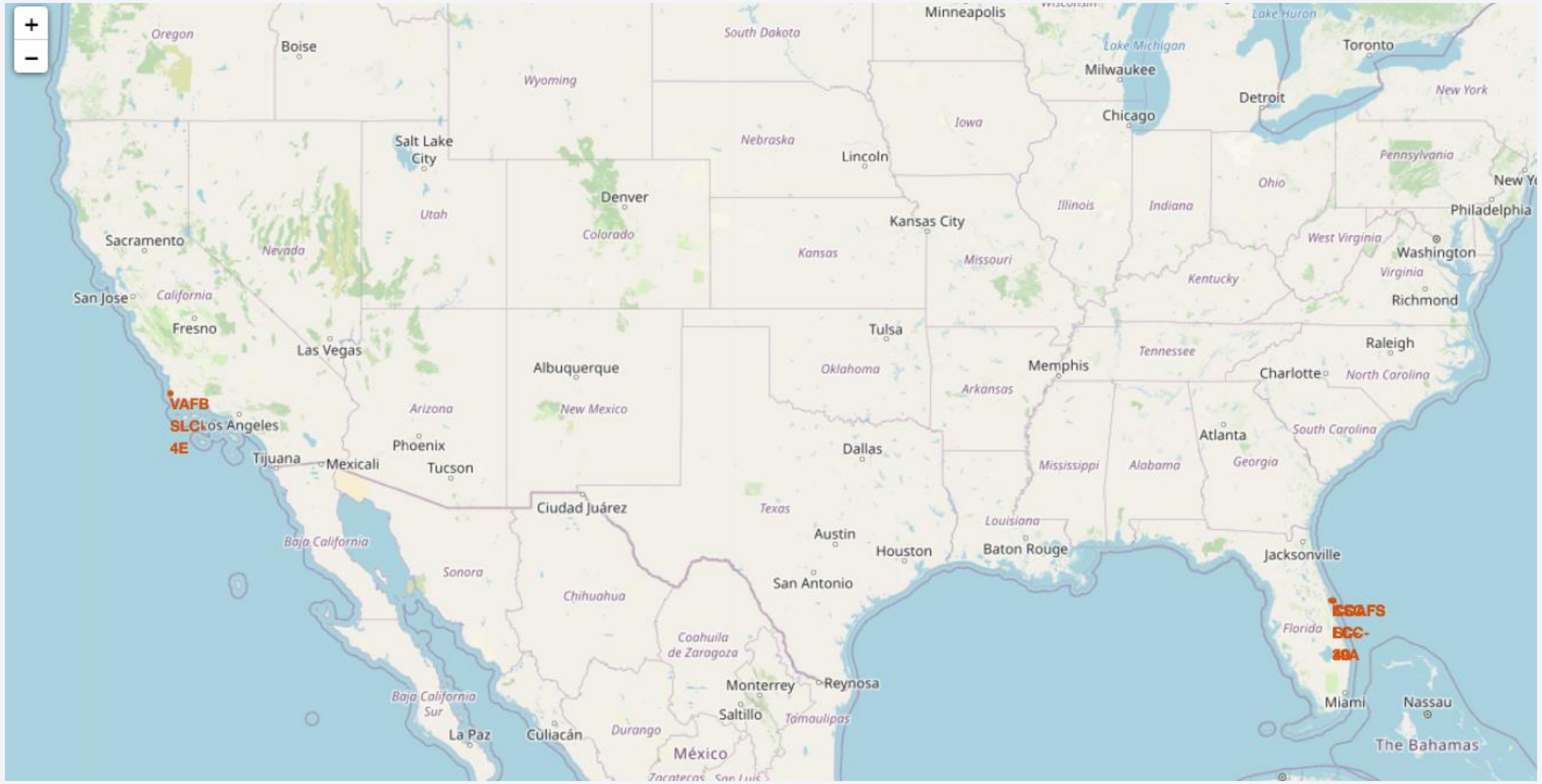
- Ranking all the landing outcomes between 2010-06-04 and 2017-03-20 gives the following result:

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

Section 3

Launch Sites Proximities Analysis

Falcon 9 Launch Sites

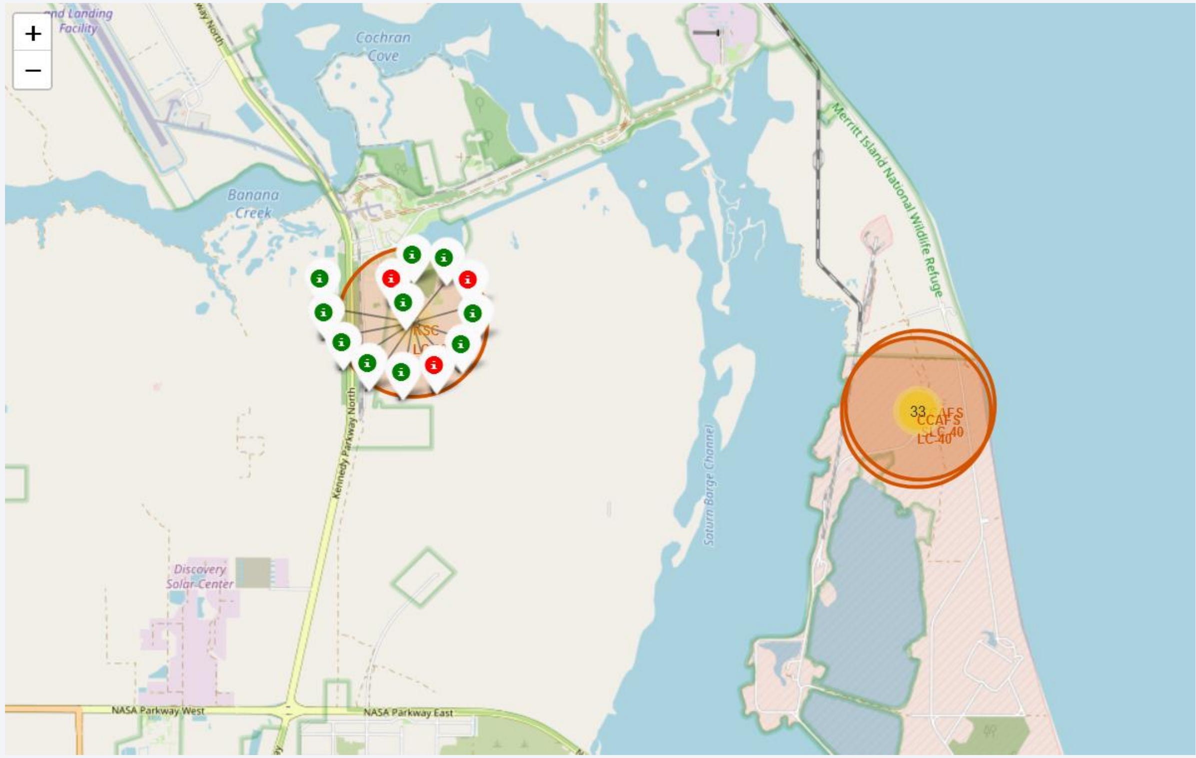


- All four launch are near the coastline, ensuring safety during launches

Launch Outcomes

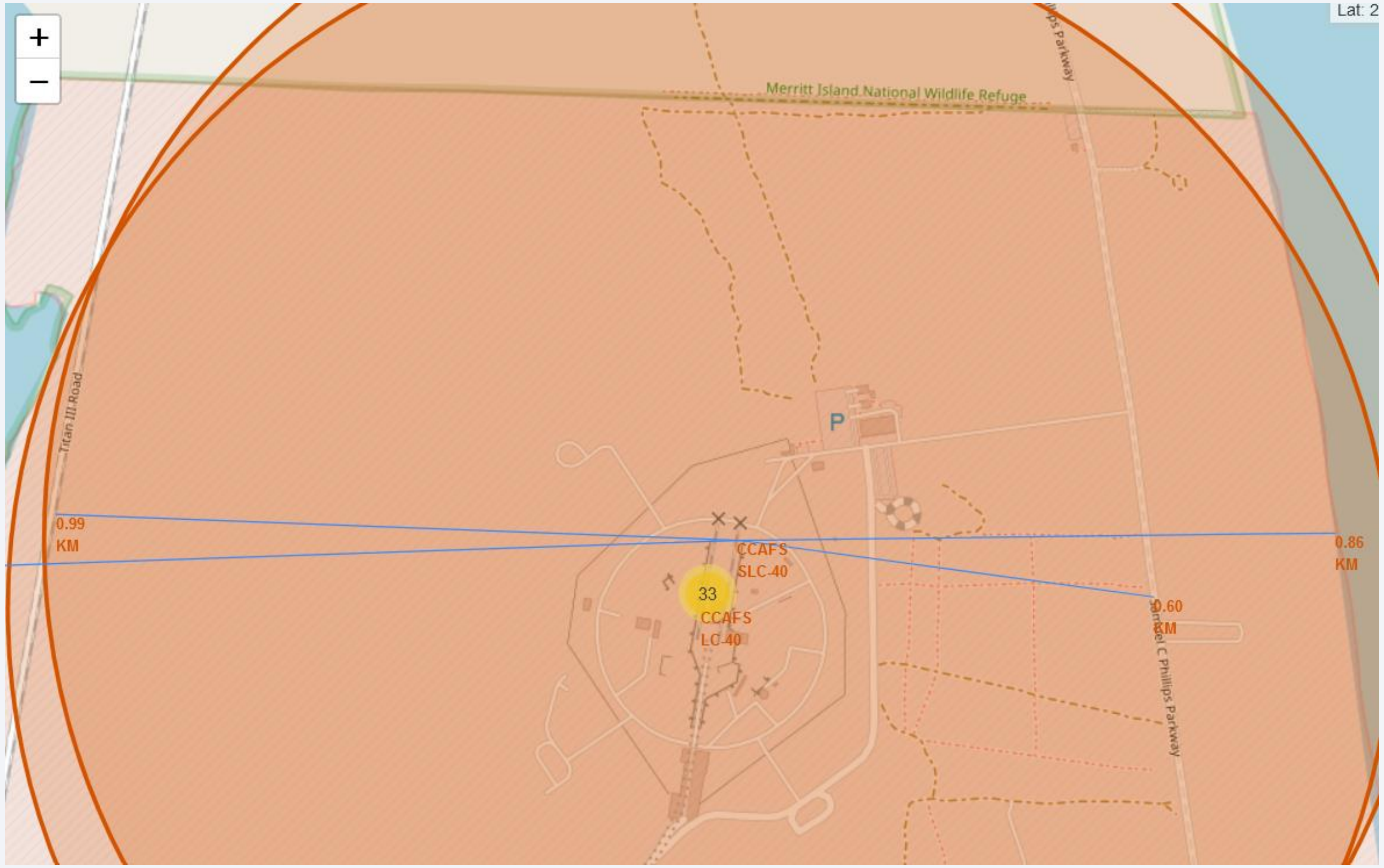
- The image to the right displays the launch outcomes of the site KSC LC-39A

- The successful launches are marked with green while the failed launches are marked with red



Logistics and Safety

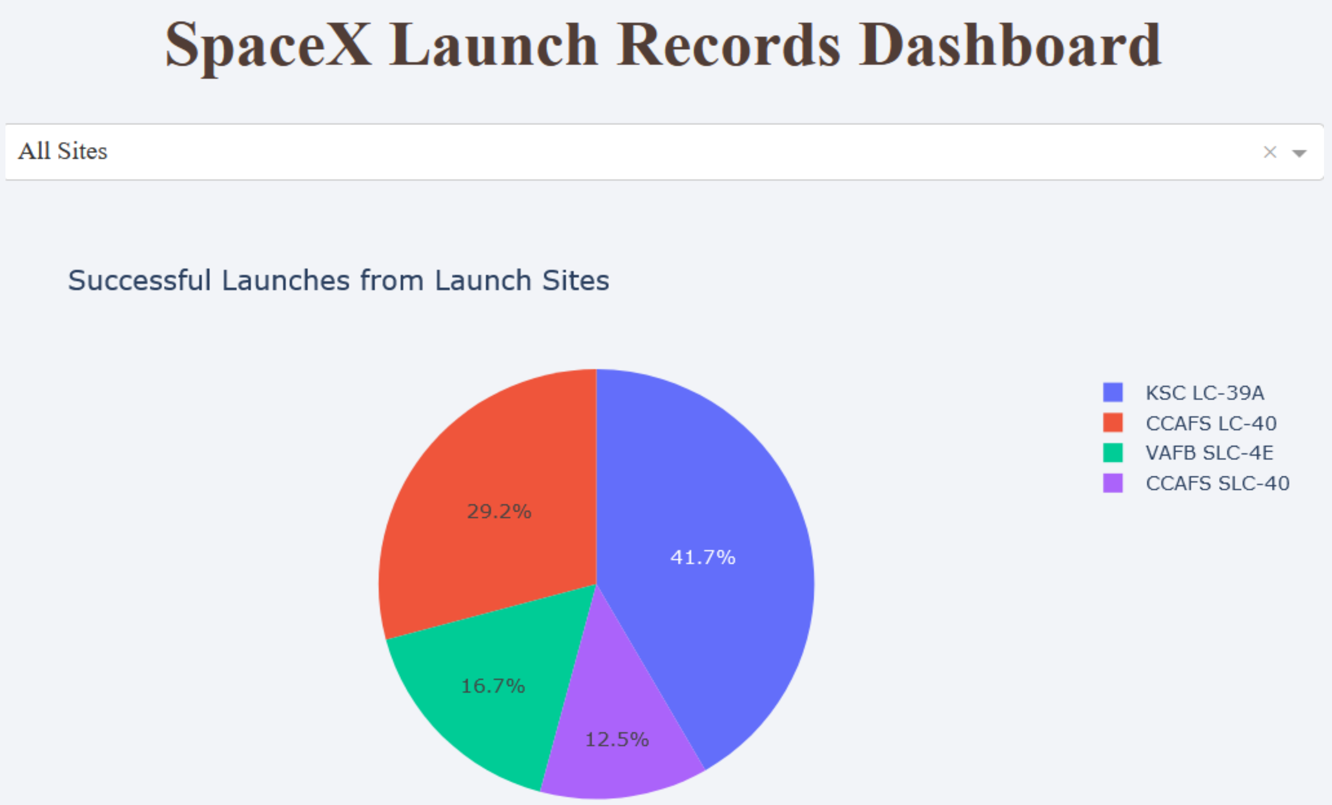
- The image to the right shows how important logistics are to the launch site
- We can see that the coastline, highway, and railway lines are all within 1 km away from the launch site CCAFS SLC-40



Section 4

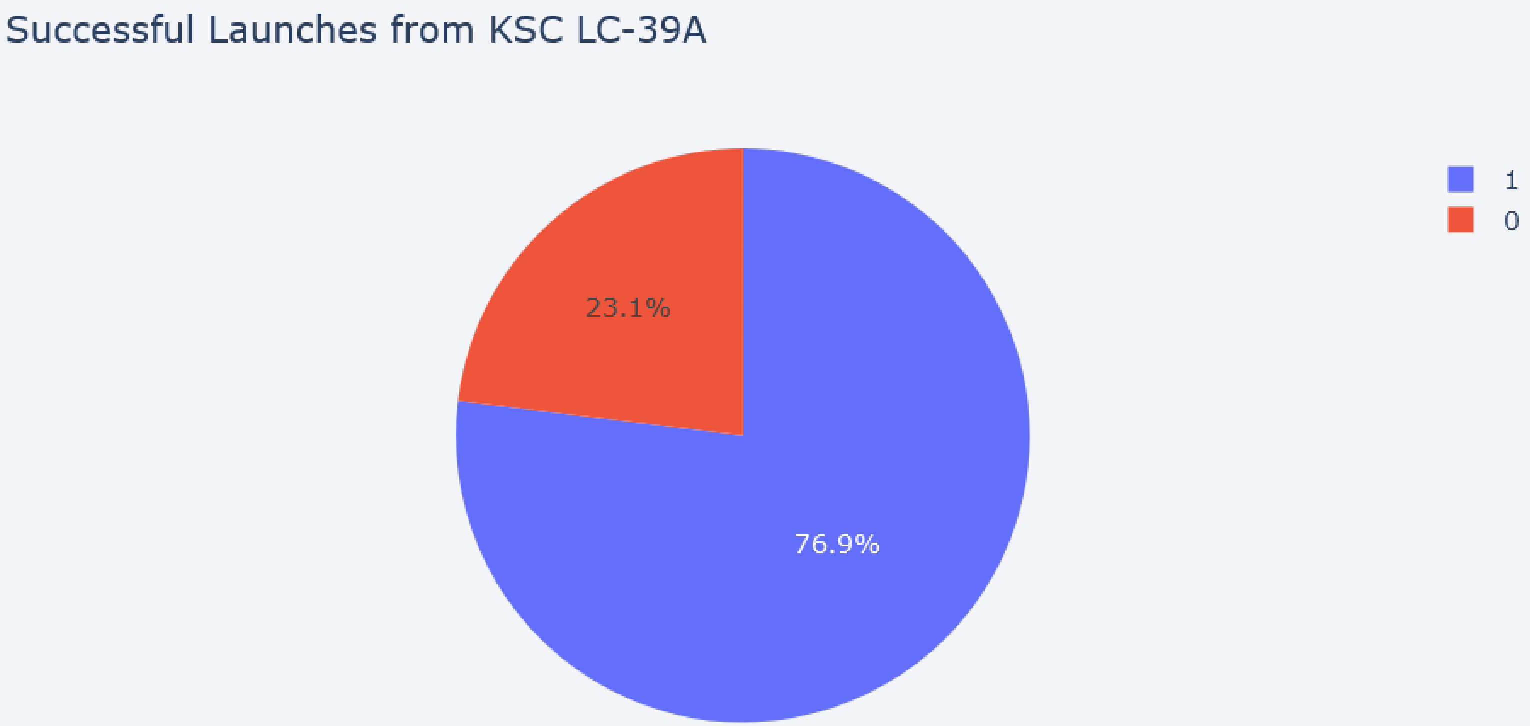
Build a Dashboard with Plotly Dash

Successful Launches of All Sites



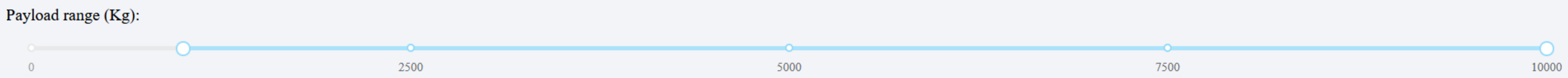
- The success rate of the launches of all sites is as followed

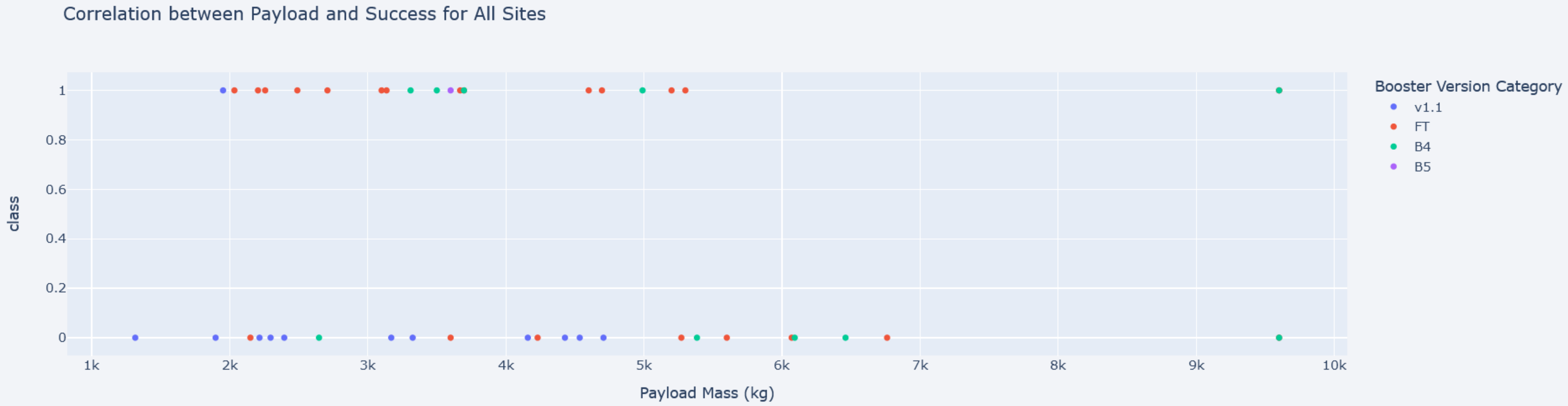
Success Rate of KSC LC-39A



- This is the Launch Site with the highest success rate of all four launch sites

Payload vs. Launch Outcome





- Booster FT under 6,000 Kg show the most successful outcome

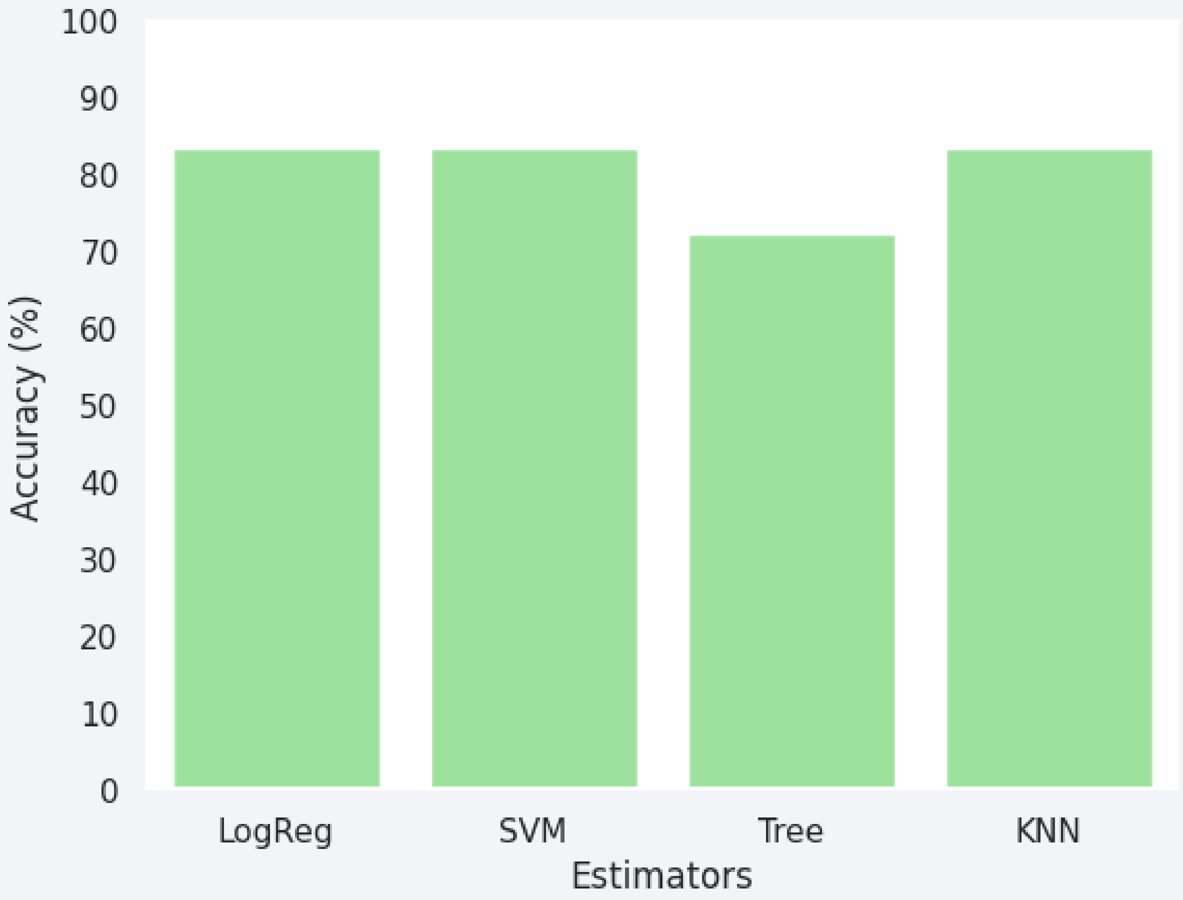
Section 5

Predictive Analysis (Classification)

Classification Accuracy

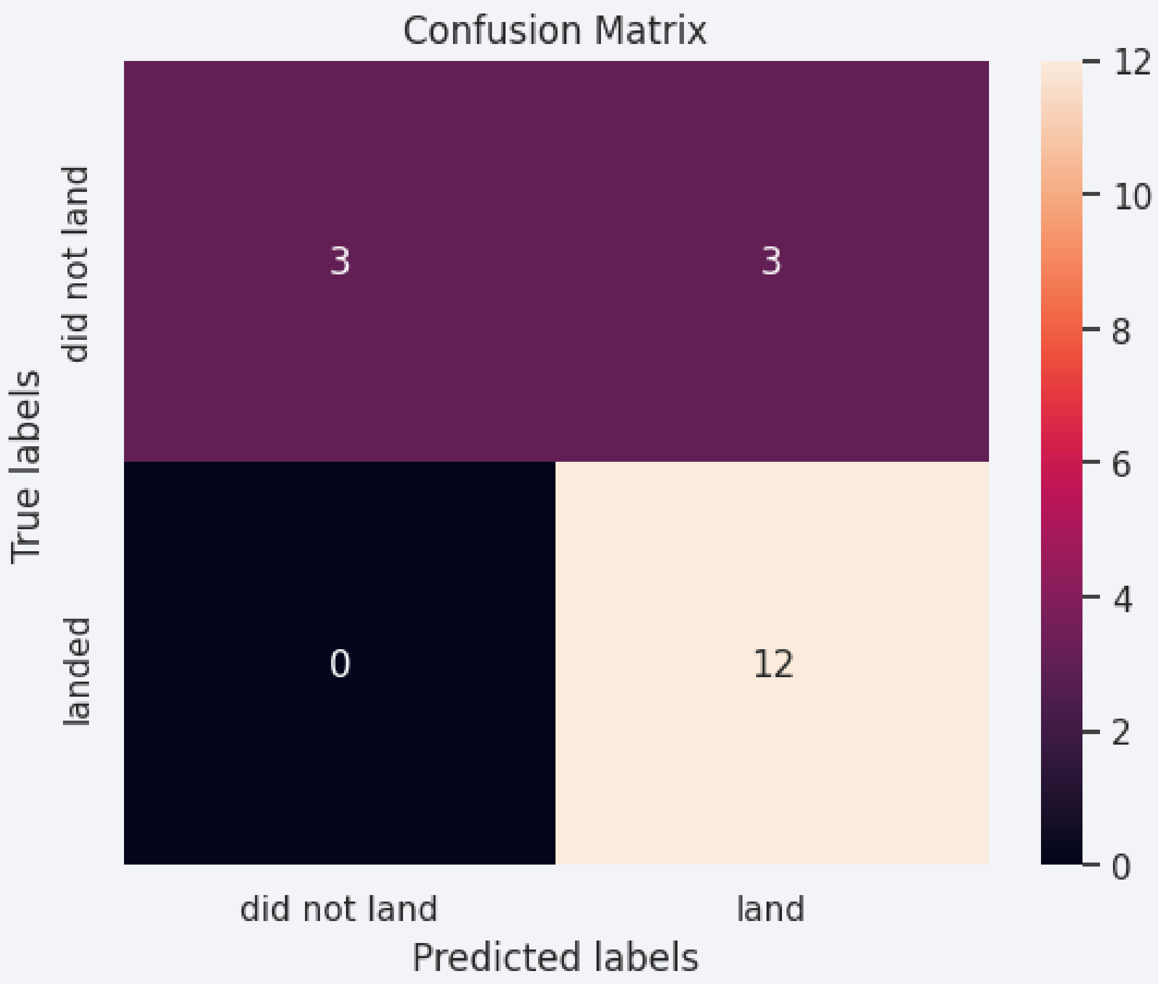
- Four models (estimators) were created to predict the outcome of an unknown launch

- Logistic Regression, Support Vector Machine, K-Nearest Neighbors have the same accuracy.



Confusion Matrix

- Since three models had the same accuracy, here is the confusion matrix of the Support Vector Machine (SVM) model



Conclusions

- Data from different sources was collected, cleaned, and analyzed to produce various insights
- Launches with payload mass above 7,000 Kg
- Successful launches seem to improve over time, seemingly due to the technological advancements
- Launch sites need to have good logistics, and are situated near the coastline
- The best launch site is KSC LC-39A
- Logistic Regression, SVM, KNN, all have the same accuracy in predicting future outcome

Appendix

- A list of hyperparameters should be prepared before training a model so that the best hyperparameters are selected

Thank you!