IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

"That rocket is launching my hope and dream"
-Rakioz-

24/11/2021

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

We try to create a model that best predicts the landing results of rockets from SpaceX missions using their past data.

We found that the decision tree classification has the best accuracy among all the methods we tried.

Introduction

If I am being honest, this entire project was created solely for me to pass the IBM Data Science course. I don't want to role-play my way out of this.

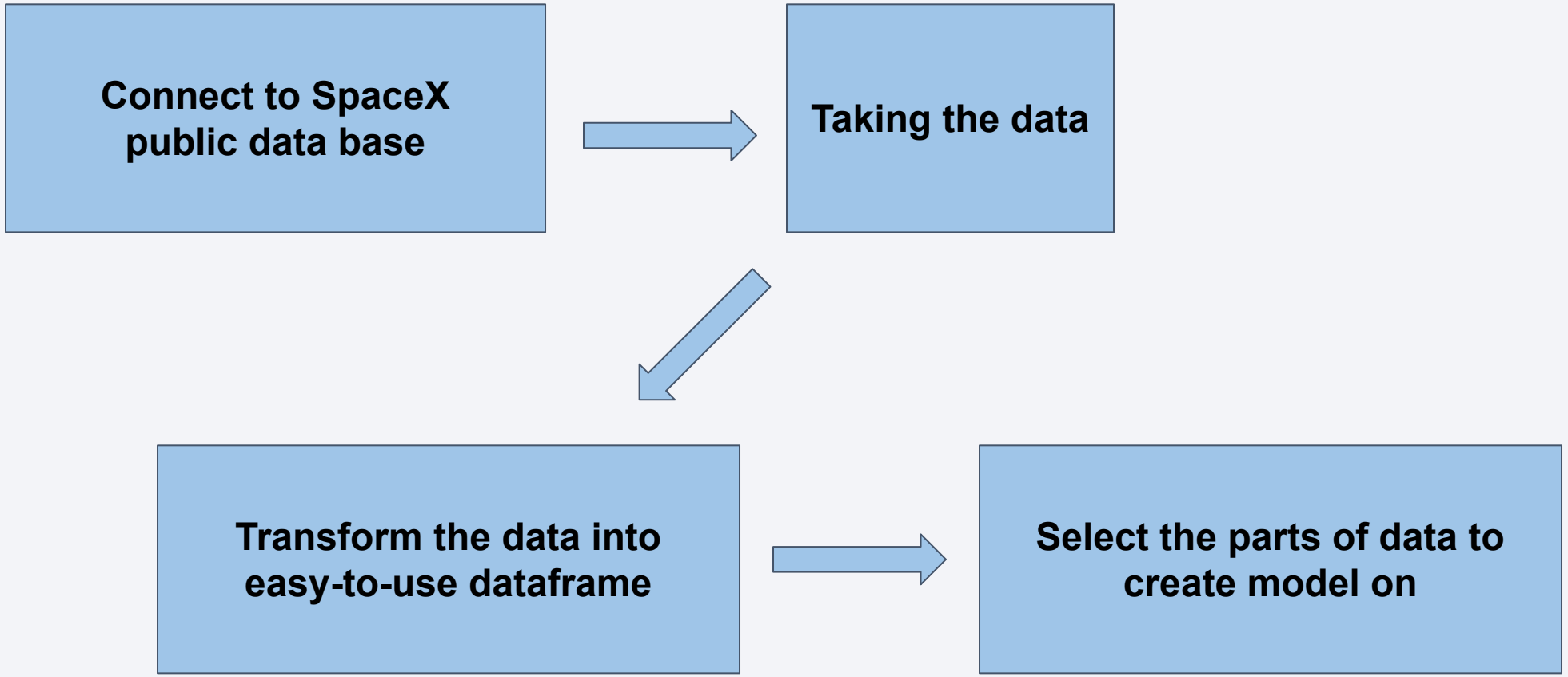
Lacking as I am, I made sure that everything works as the notebook instructed. All the Python codes and their results, including Watson Studio links, can be found the Github repository here.

This presentation will follow the structure provided by the course.

Section 1

Methodology

Data Collection



Data Collection – SpaceX API

We ended up with a dataframe on Falcon 9 launches.

We keep only the free variables we are interested in building the model on such as Orbit, Launch Site, Payload Mass, Coordinate.

Our dependent variable is the outcome.

The result can be found here.

Data Collection - Scraping

We change the parts of our dataframe that are categorical variables into numerical ones so that we can use machine learning to create a model later.

Most importantly, we create the column of numbers 0 and 1 indicating the landing outcome instead of ‘Success’ and ‘Failure’.

The result can be found here.

EDA with Data Visualization

We plot graphs to display relationships between the variables such as Payload, Flight number, Orbit type.

The most important ones are those concerning the success rate of the missions. In the last part, we can clearly see the success rate going up sharply after 2013.

The result can be found here.

EDA with SQL

We use SQL queries to see the total count and mean of different variables.

We pay special attention to the booster versions with successful landing outcome.

The result can be found here.

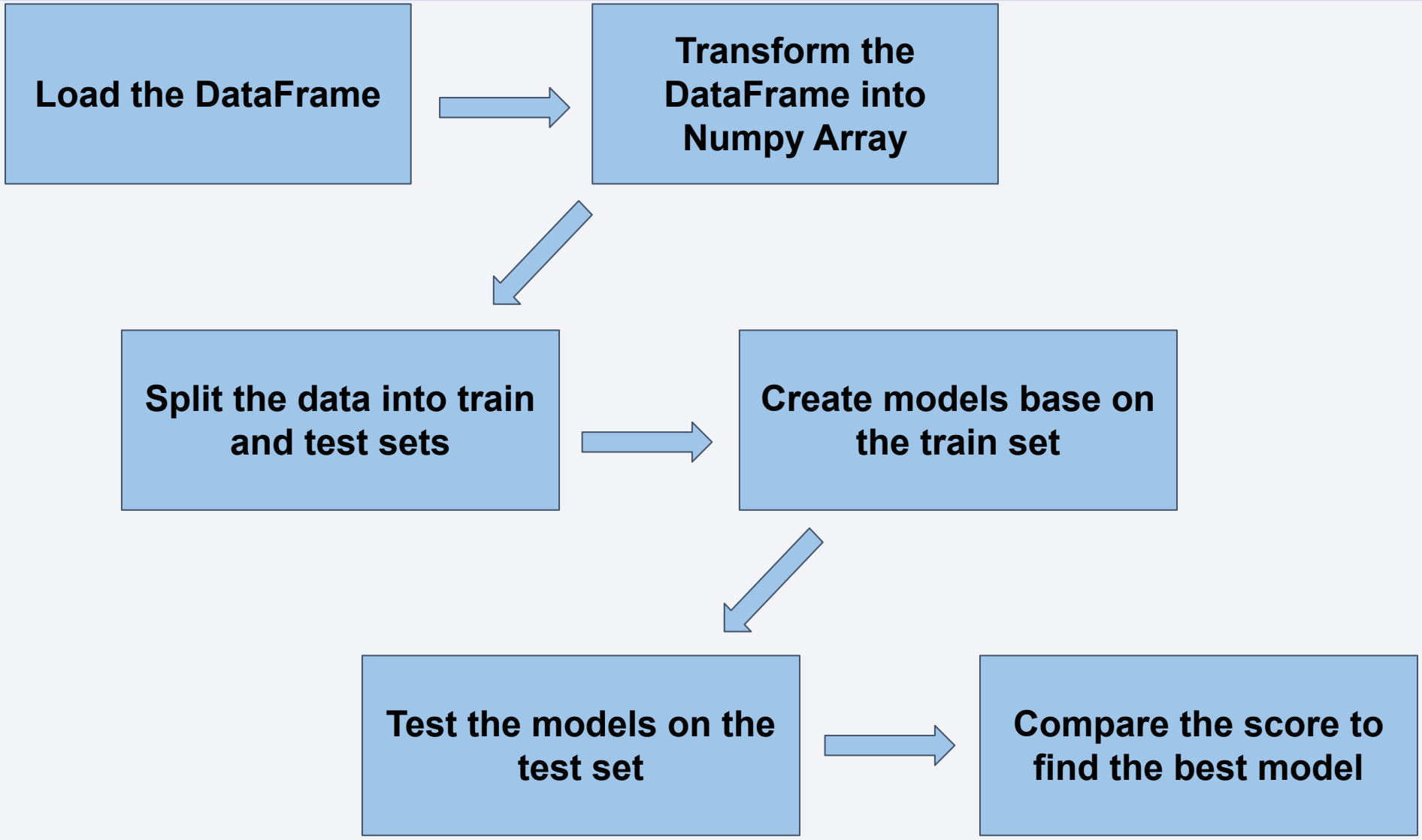
Build an Interactive Map with Folium

We create Folium markers to indicate the location of launch sites on the map of USA.

We also add information about successful attempts at each launch site and create a sample of distance measuring marker.

The result can be found here.

Predictive Analysis (Classification)



Predictive Analysis (Classification)

By comparing the accuracy score of each model, we conclude that the decision tree model has the best score.

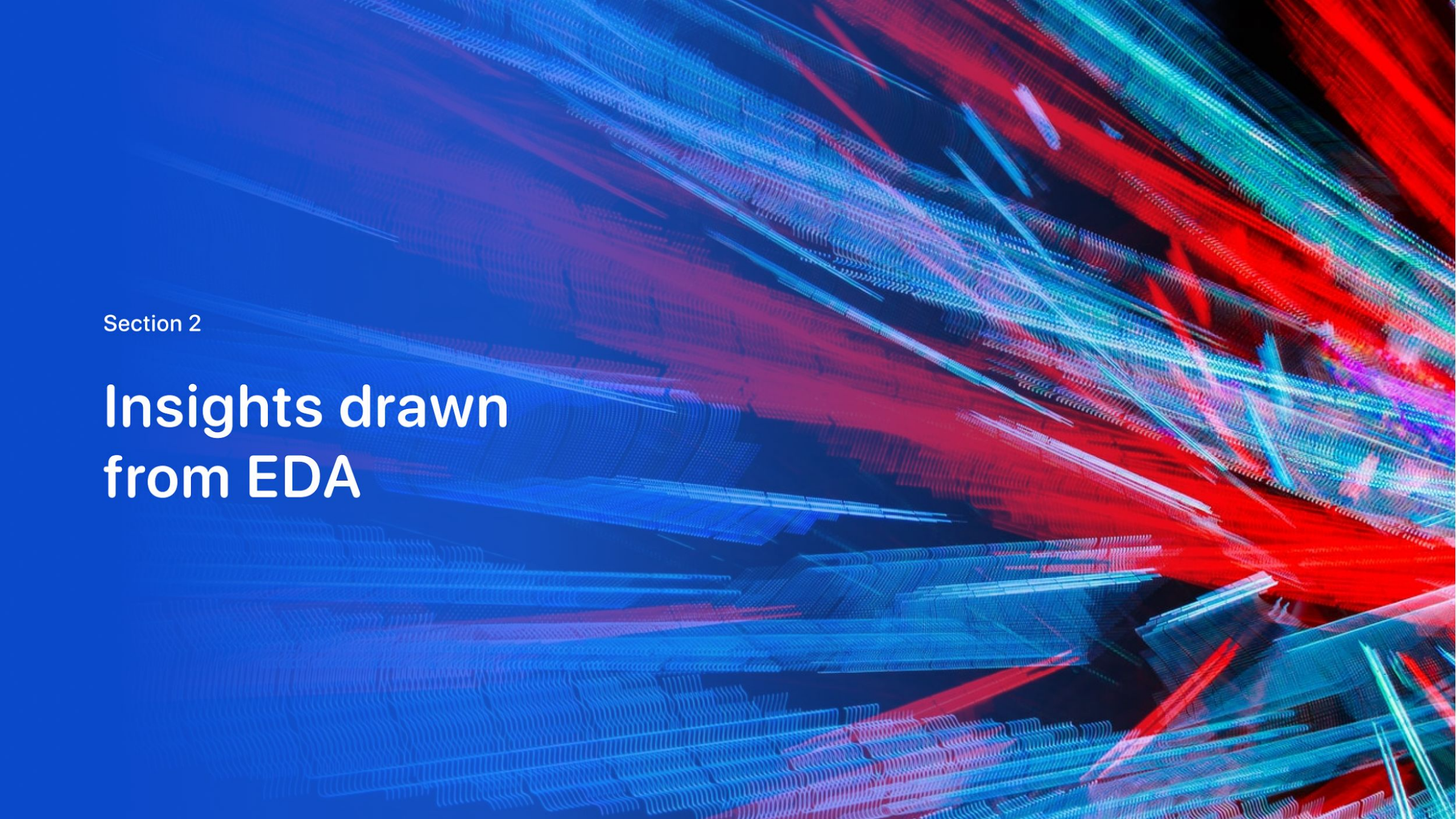
The result can be found here.

Results

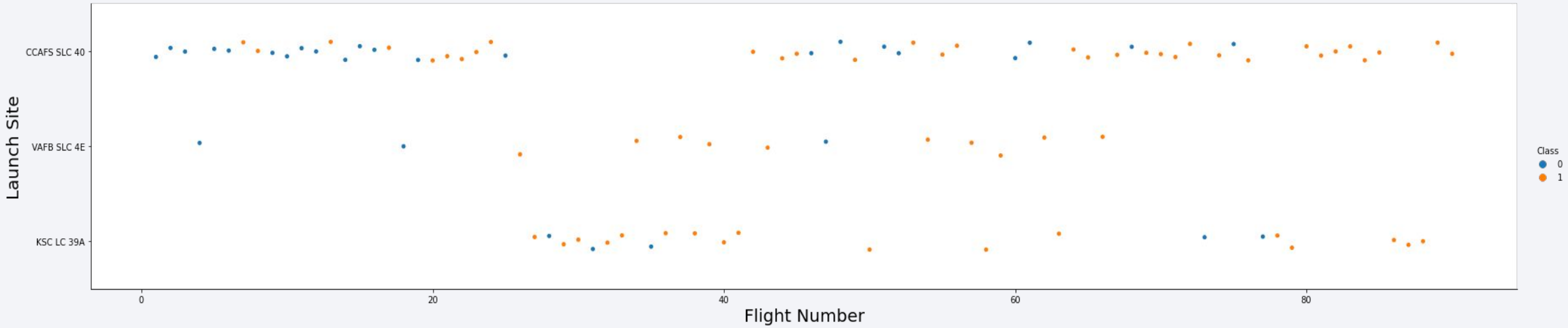
Exploratory data analysis results

Interactive analytics demo in screenshots

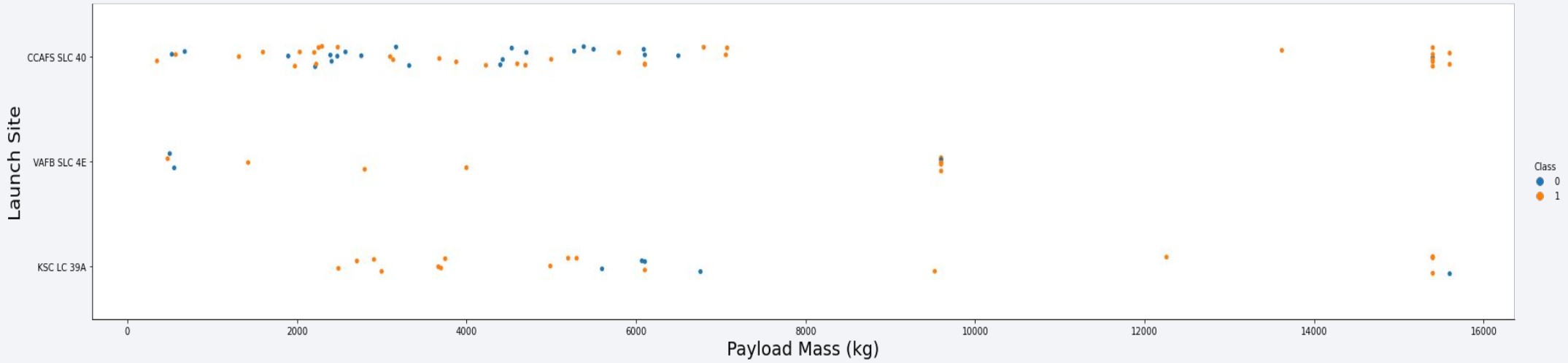
Predictive analysis results


Section 2
Insights drawn from EDA

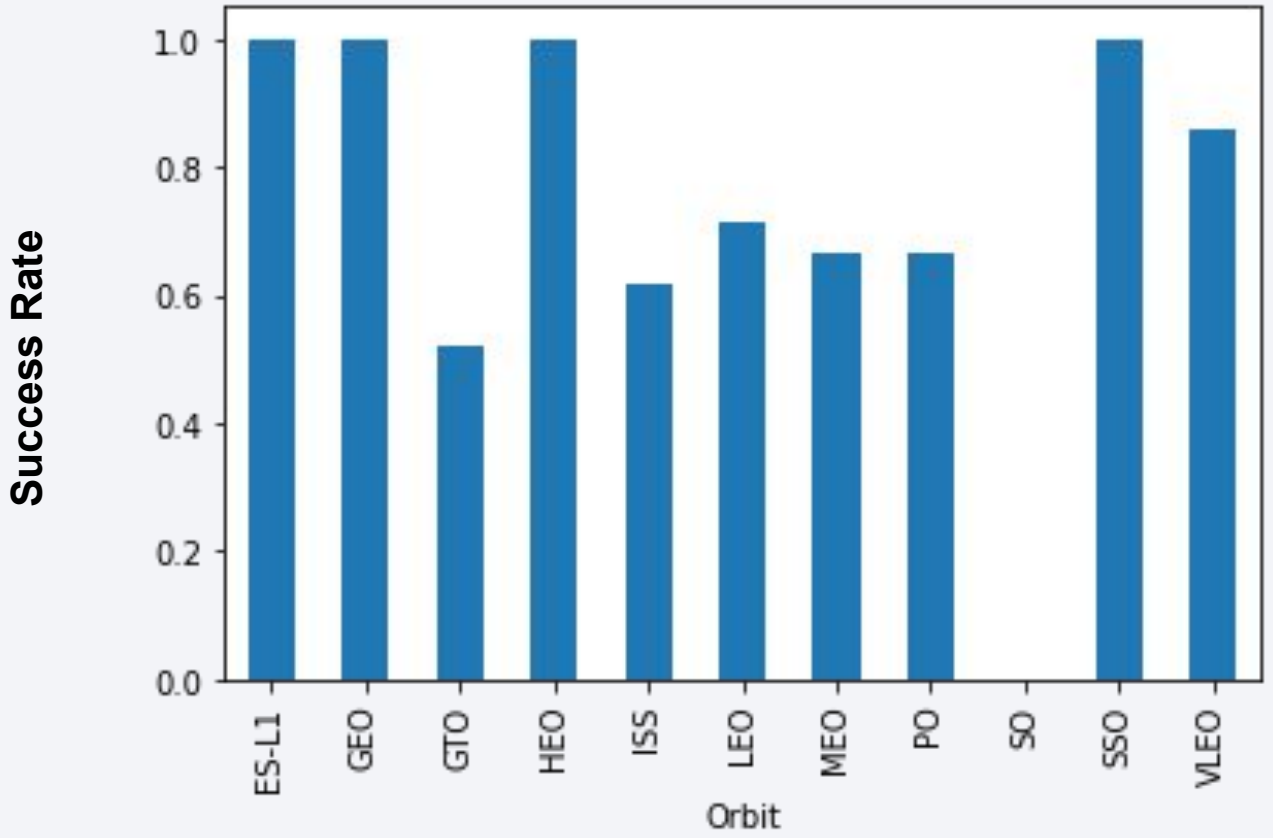
Flight Number vs. Launch Site



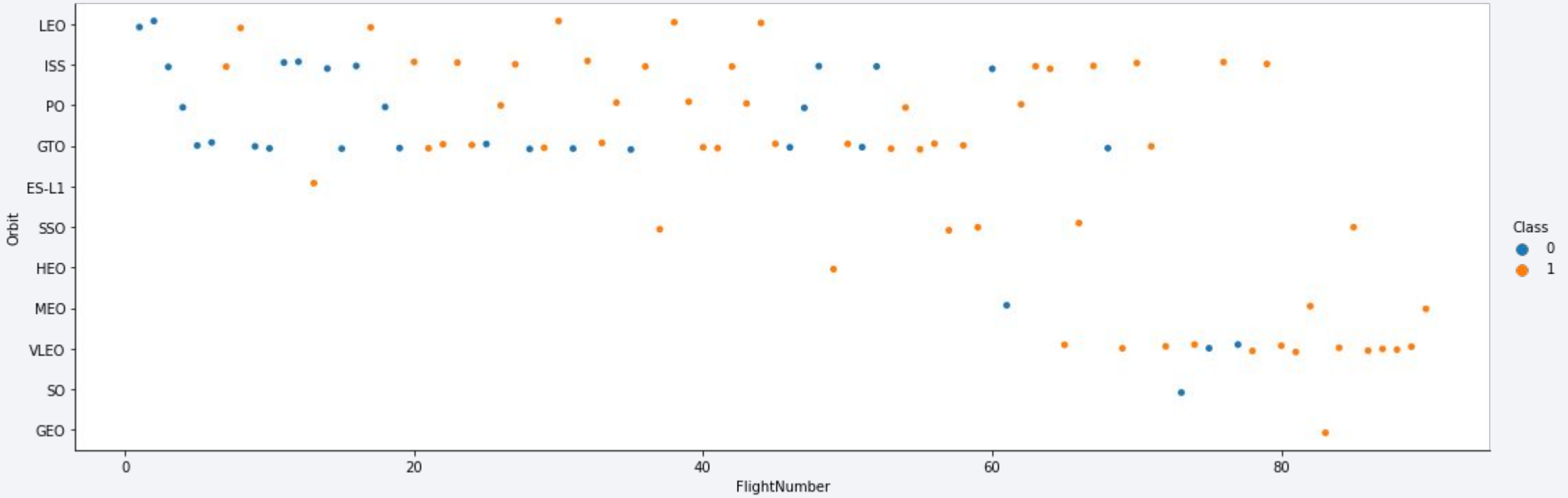
Payload vs. Launch Site



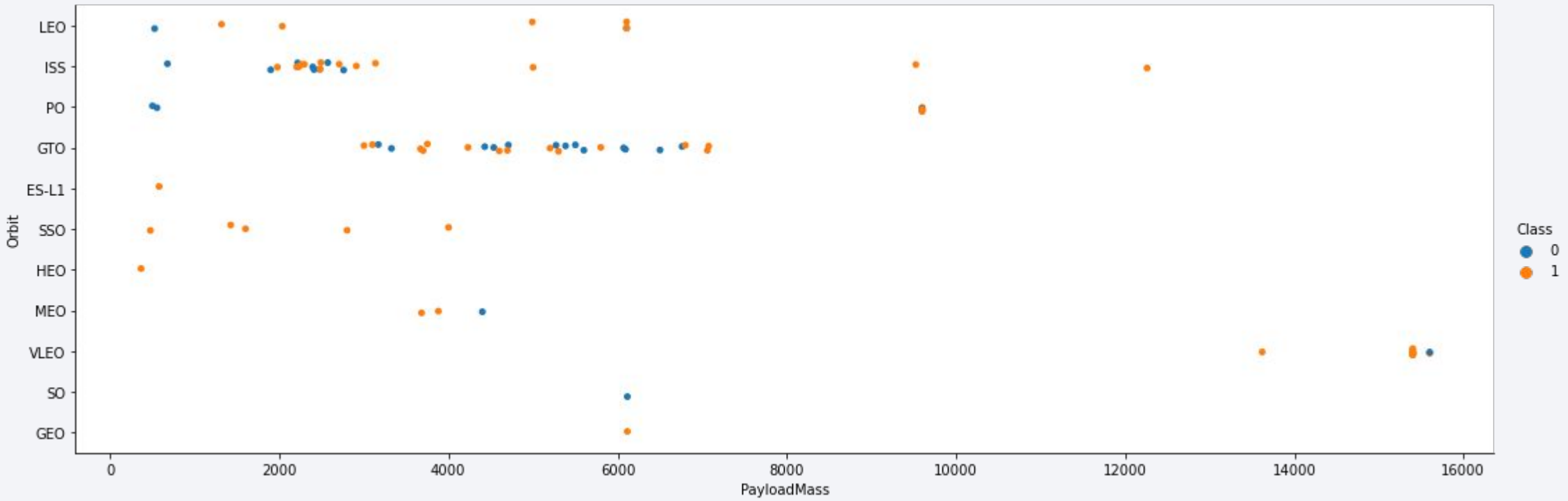
Success Rate vs. Orbit Type



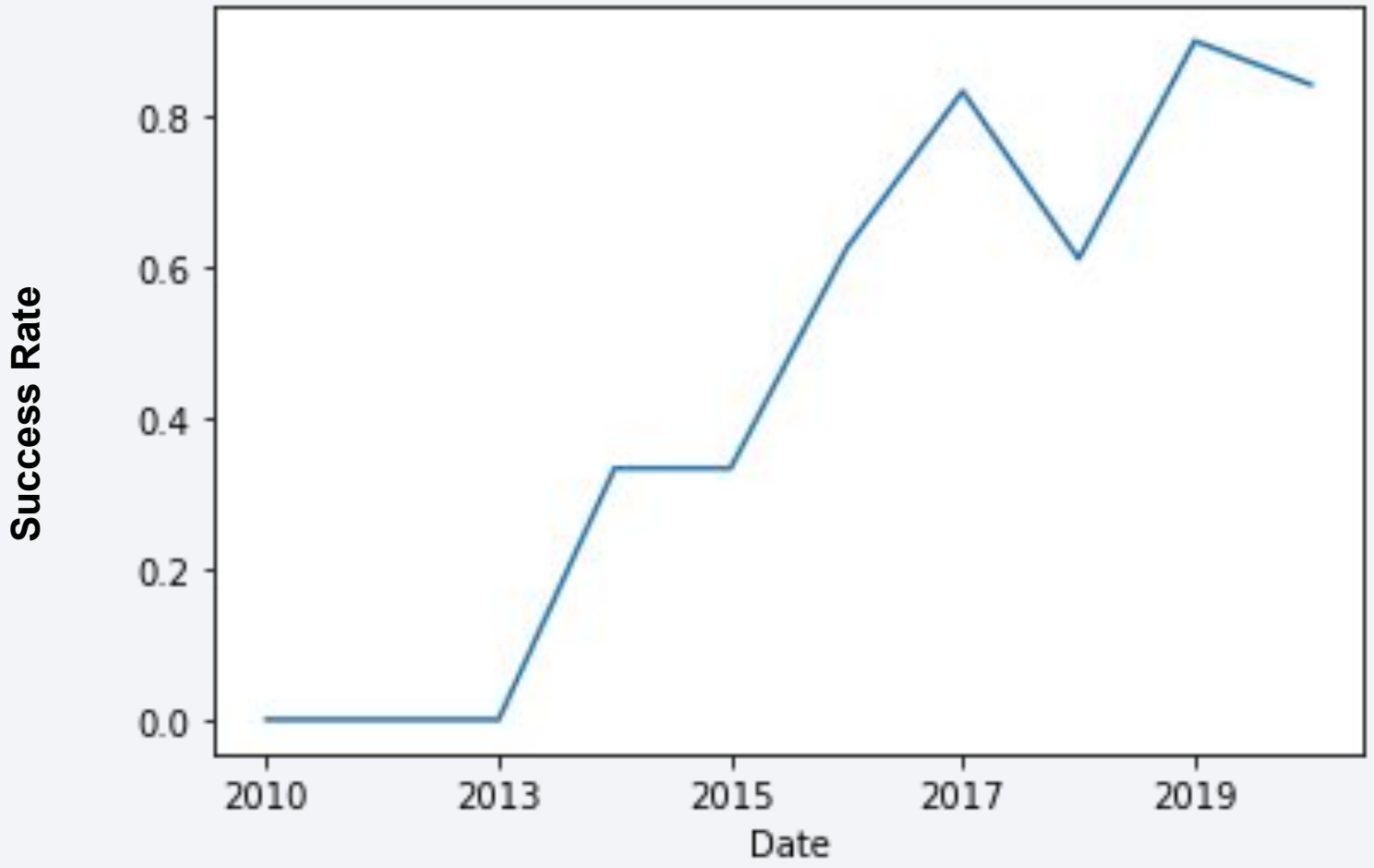
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

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

Launch Site Names 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 attempt
2012-10-08	0:35:00	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 attempt

Total Payload Mass

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql SELECT SUM(payload_mass__kg_) FROM SPACEX where CUSTOMER like 'NASA (CRS)'
```

```
 * ibm_db_sa://xhz91834:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u
Done.
```

1
45596

Total Payload Mass

Average Payload Mass by F9 v1.1

Display average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG(payload_mass__kg_) FROM SPACEX where BOOSTER_VERSION like 'F9 v1.1'
```

```
 * ibm_db_sa://xhz91834:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.
Done.
```

1

2928

Avg Payload Mass

First Successful Ground Landing Date

DATE	time_utc_	booster_version	launch_site	payload	payload_mass_kg_
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034

Successful Drone Ship Landing with Payload between 4000 and 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

List the total number of successful and failure mission outcomes

In [132…

```
%sql SELECT COUNT(case when mission_outcome like '%Success%' t
```

```
 * ibm_db_sa://xhz91834:***@98538591-7217-4024-b027-8baa776ffa
Done.
```

Out[132…

success	failure
100	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

2015 Launch Records

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

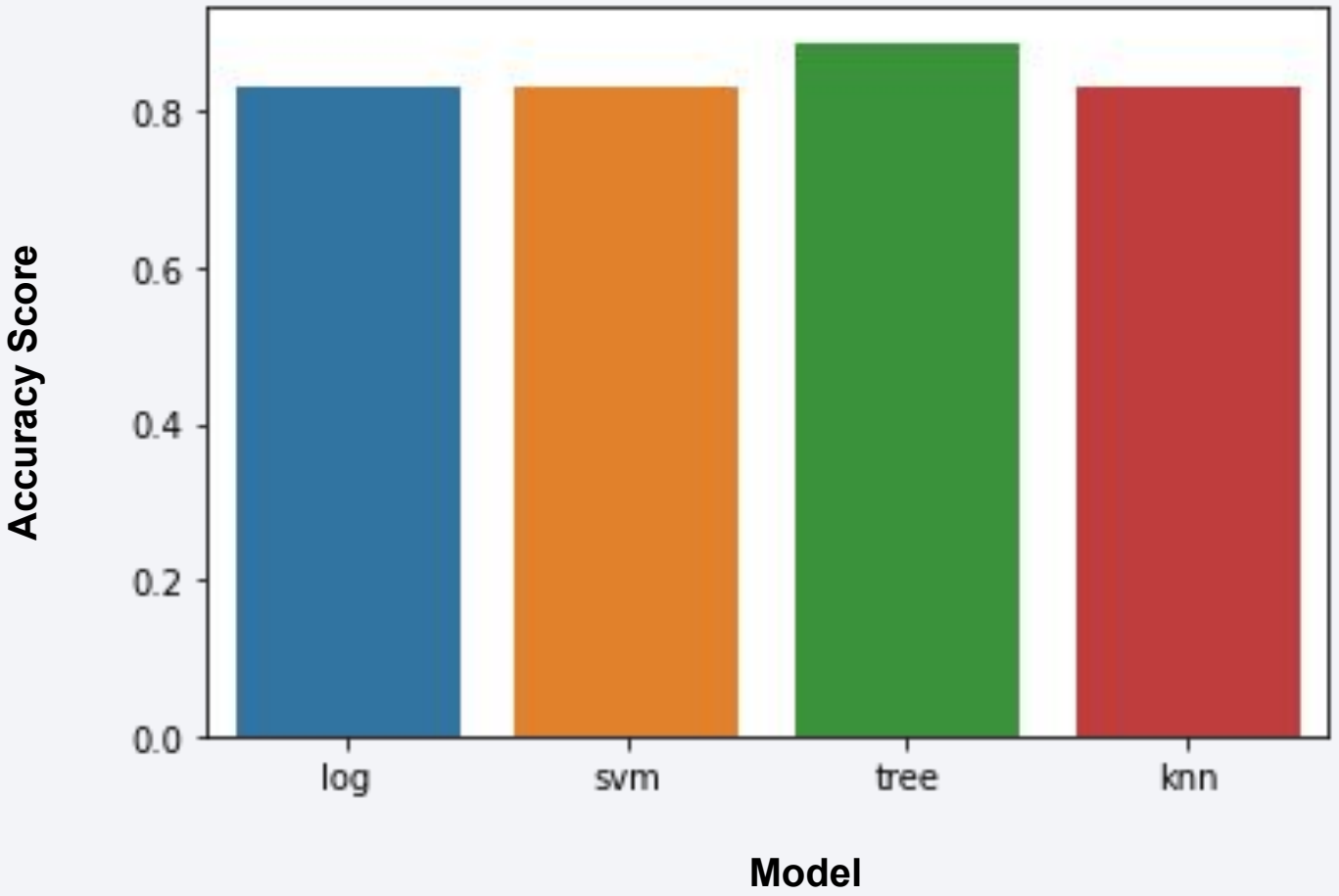
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

landing_outcome	2
Controlled (ocean)	3
Failure (drone ship)	5
No attempt	9
Precluded (drone ship)	1
Success (drone ship)	5
Success (ground pad)	3
Uncontrolled (ocean)	2

Section 6

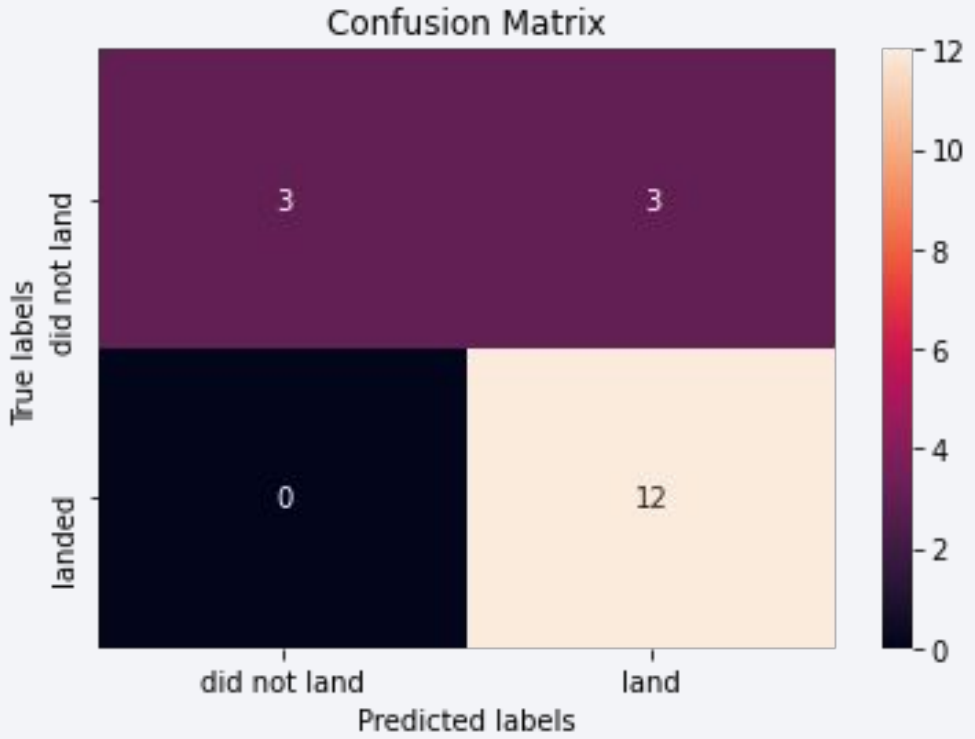
Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix

Decision Tree Classification



Conclusions

We found that the decision tree classification has the most accuracy out of all the model we create.

However, we want to make an observation that the number of data points available for the modeling and analysis is quite small and the insight we obtain from such condition is likely to be flawed.

We believe it is very important for the developer to investigate further into the real world causes behind the past failures in order to prevent the future ones.

Thank you!