IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Juan Marcos Díaz
5/11/2022

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

  Data from Space X was obtained from 2 sources: Space X API and Web Scrapping

- Perform data wrangling

  After analyzing the features, a landing outcome label was created

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
- Data was normalized, divided in training and test data sets and evaluated by four different classification models

Data Collection

Data sets were collected from:

- Space X API (https://api.spacexdata.com/v4/rockets/)
- Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)

Data Collection – SpaceX API

- SpaceX offers a public API from where data can be obtained
- https://github.com/JMarksss/IBM/blob/main/jupyter-labs-spacex-data-collection-api_JMD%20(1).ipynb

1. Request API and parse the SpaceX launch data
2. Filter data to only include Falcon 9 launches
3. Deal with Missing Values

Data Collection - Scraping

- Data from SpaceX launches can also be obtained from Wikipedia;

https://github.com/JMarksss/IBM/blob/main/jupyter-labs-webscraping_JMD%20(1).ipynb

1. Request the Falcon9 Launch Wiki page
2. Extract all column/variable names from the HTML table header
3. Create a data frame by parsing the launch HTML tables

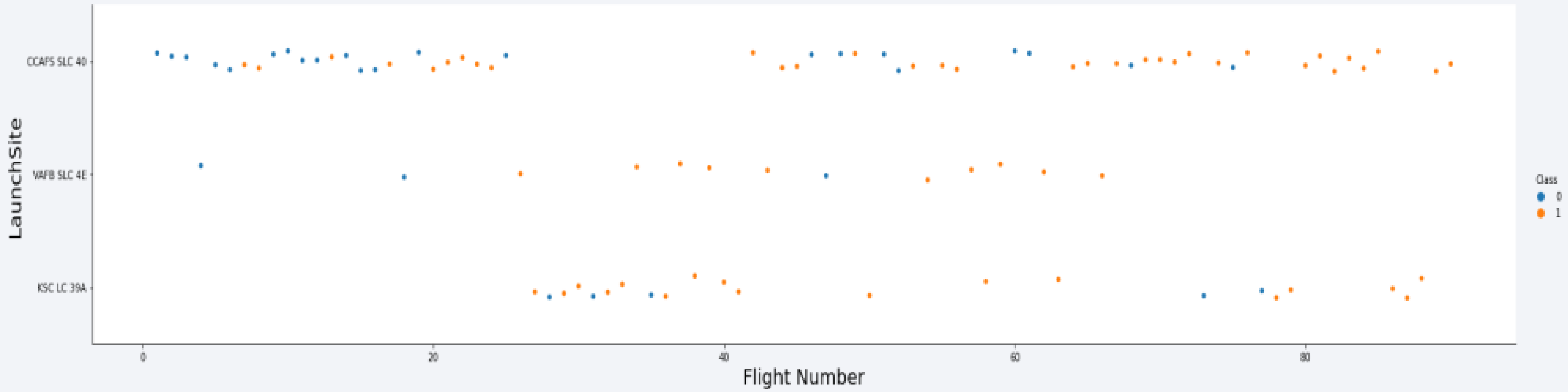
Data Wrangling

1. First some Exploratory Data Analysis was performed on the dataset.
2. Then some features as outcome per orbit type were calculated.
3. Finally, the landing outcome label was created from Outcome column.

- https://github.com/JMarksss/IBM/blob/main/labs-jupyter-spacex-Data%20wrangling_JMD%20(1).ipynb

EDA with Data Visualization

- Scatterplots and Barplots were used to visualize the relationship between pair of features as Launch Site and Flight Number



https://github.com/JMarksss/IBM/blob/main/jupyter-labs-eda-dataviz_JMD%20(1).ipynb

EDA with SQL

The SQL queries performed were:

- Names of the unique launch sites in the space mission;
- Top 5 launch sites whose name 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;
- Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg;
- Total number of successful and failure mission outcomes;
- Names of the booster versions which have carried the maximum payload mass;
- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015; and
- Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20.

https://github.com/JMarksss/IBM/blob/main/jupyter-labs-eda-sql-coursera_JMD%20(1).ipynb

Build an Interactive Map with Folium

The next objects were used with Folium Maps:

- Markers indicate points like launch sites;
- Circles indicate highlighted areas around specific coordinates;
- Marker clusters indicates groups of events in each coordinate;
- Lines are used to indicate distances between two coordinates.

https://github.com/JMarksss/IBM/blob/main/lab_jupyter_launch_site_location_JMD%20(1).ipynb

Build a Dashboard with Plotly Dash

- Percentage of launch by site and Payload range were used to visualize data and analyze the relation between payloads and launch sites, helping to identify where is best place to launch.

Predictive Analysis (Classification)

• Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbors.

1. Data preparation and standardization
2. Test of each model with combinations of hyperparameters
3. Comparison of results

https://github.com/JMarksss/IBM/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5_JMD%20(1).ipynb

Results

Exploratory data analysis results:

- Space X uses 4 different launch sites;
- The first launches were done to Space X and NASA;
- The average payload of F9 v1.1 booster is 2,928 kg;
- The first success landing outcome happened in 2015, first launch was in 2010;
- Almost 100% of mission outcomes were successful;
- Two booster versions failed at landing in drone ships in 2015;
- The number of landing outcomes increase per year.

Results

Launch sites use to be in safety places near sea, mostly at East Coast

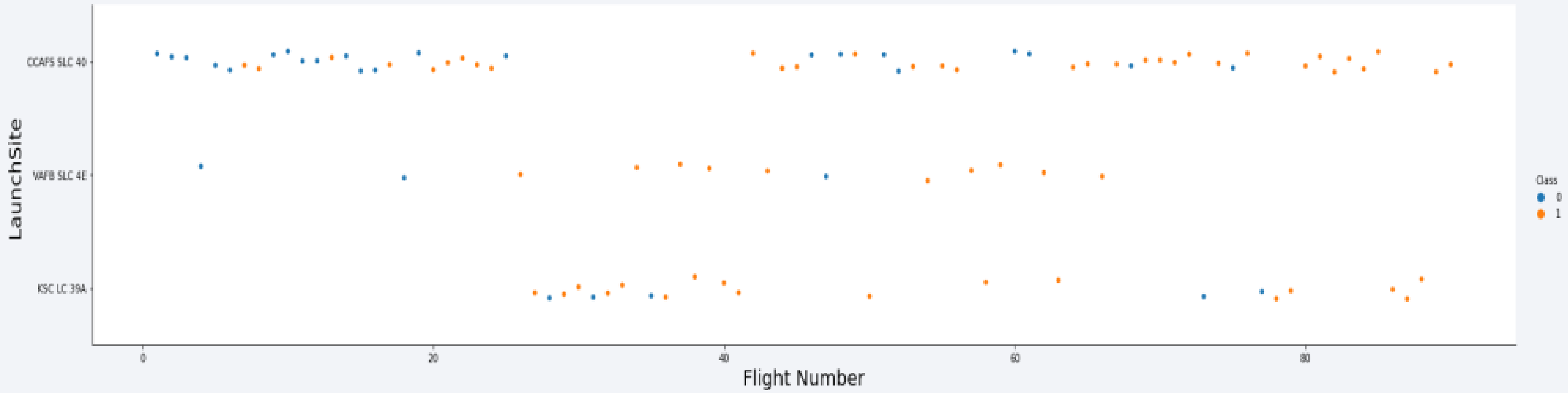
- Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings

Section 2

Insights drawn from EDA

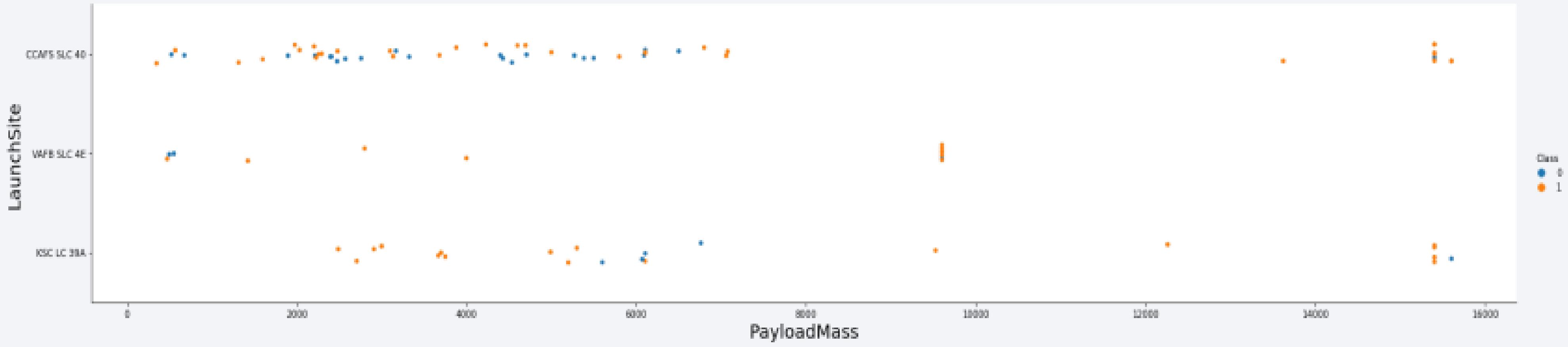
Flight Number vs. Launch Site

- The best launch site nowadays is CCAF5 SLC 40, where most of recent launches were successful;
- General success rate improved over time.



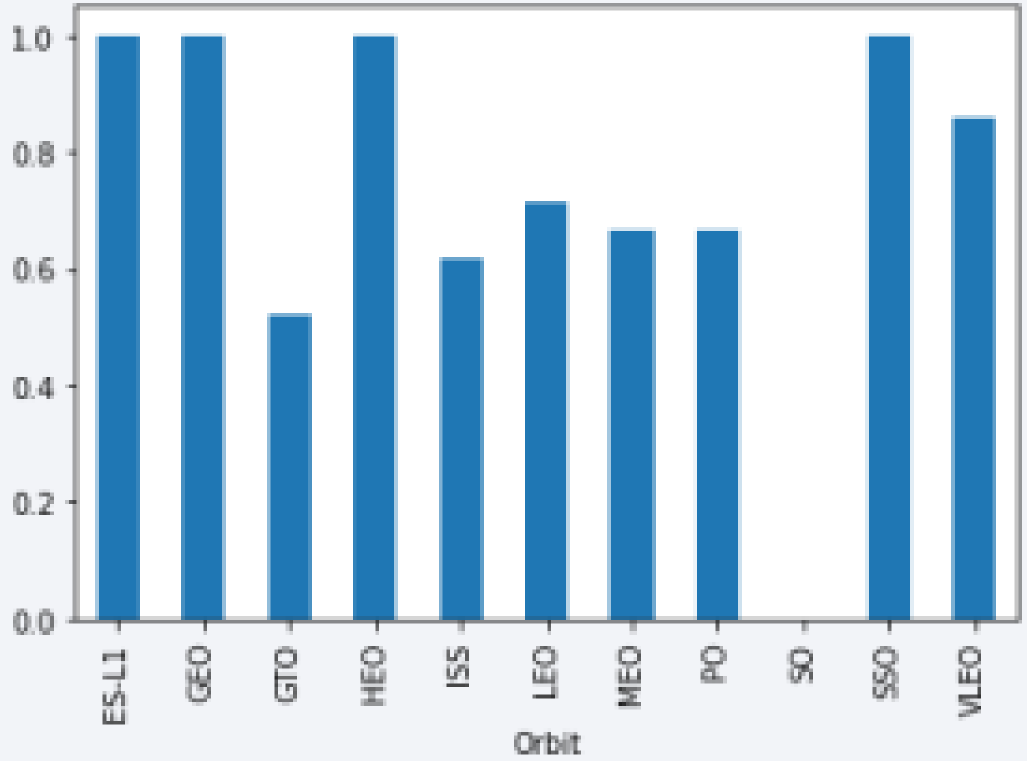
Payload vs. Launch Site

- Payloads over 9,000kg have excellent success rate
- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.



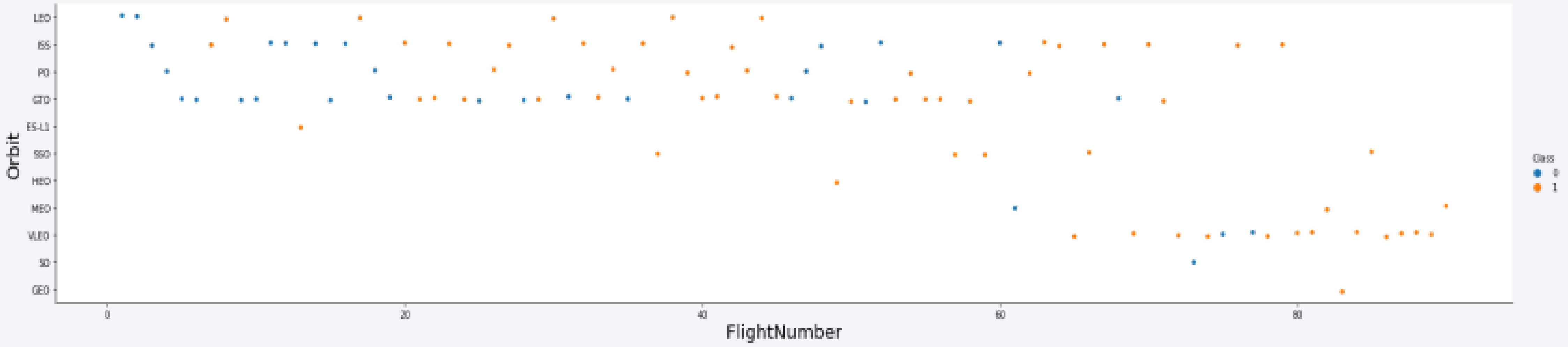
Success Rate vs. Orbit Type

- The orbits ES-L1, GEO, HEO and SSO have a perfect success rate.
- SO has the worst success rate.



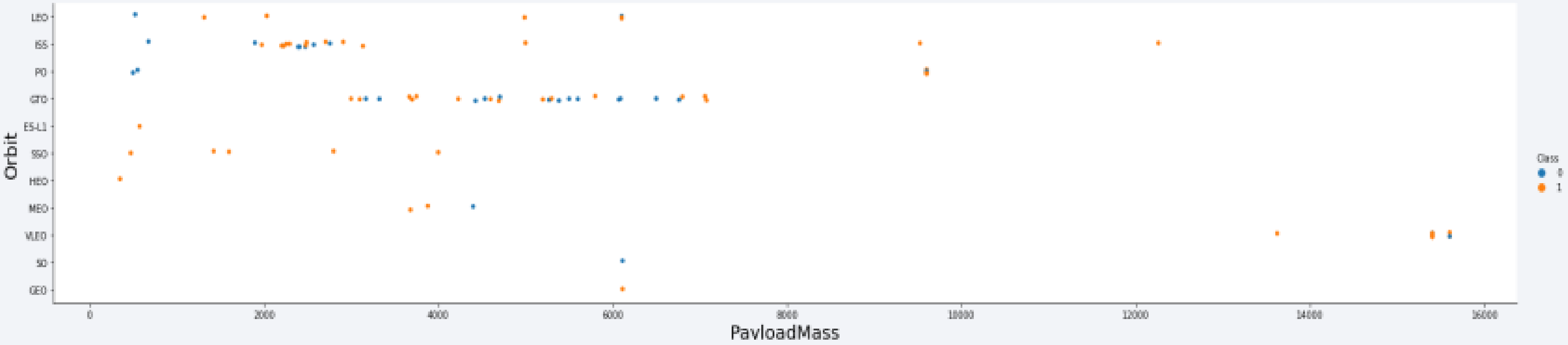
Flight Number vs. Orbit Type

- Success rate improved over time to all orbits;



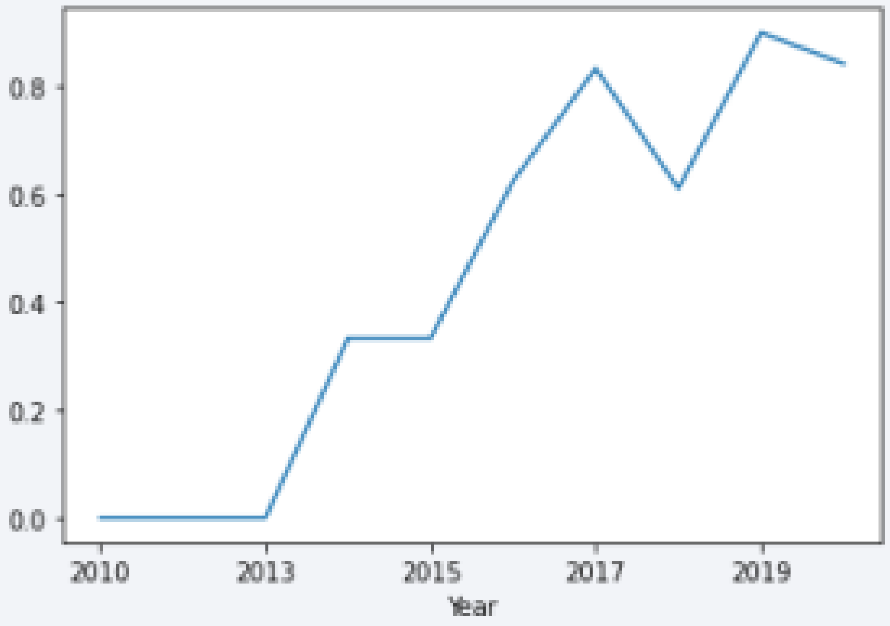
Payload vs. Orbit Type

- There is no relation between payload and success rate to orbit GTO
- There are few launches to the orbits SO and GEO.



Launch Success Yearly Trend

- The first three years there was no improvement.
- Success rate has increased since 2013.
- In 2018 it decrease a bit.



All Launch Site Names

There are four unique launch sites:

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

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	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 carried by boosters from NASA was 111.268 kg.
- Total payload calculated above, by summing all payloads whose codes contain ‘CRS’, which corresponds to NASA.

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 was 2.928 kg
- We obtained the value filtering data by the booster version and calculating the average payload mass.

First Successful Ground Landing Date

- First successful landing outcome on ground pad was 2015-12-22
- We obtained the value filtering data by successful landing outcome on ground pad and getting the minimum value for date

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:

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

Total Number of Successful and Failure Mission Outcomes

- Grouping mission outcomes and counting records for each group led us to the next summary:

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

Boosters Carried Maximum Payload

Boosters which have carried the maximum payload mass are:

- F9 B5 B1048.4
- F9 B5 B1048.5
- F9 B5 B1049.4
- F9 B5 B1049.5
- F9 B5 B1049.7
- F9 B5 B1051.3
- 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

- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

- Ranking of all landing outcomes between the date 2010-06-04 and 2017- 03-20:

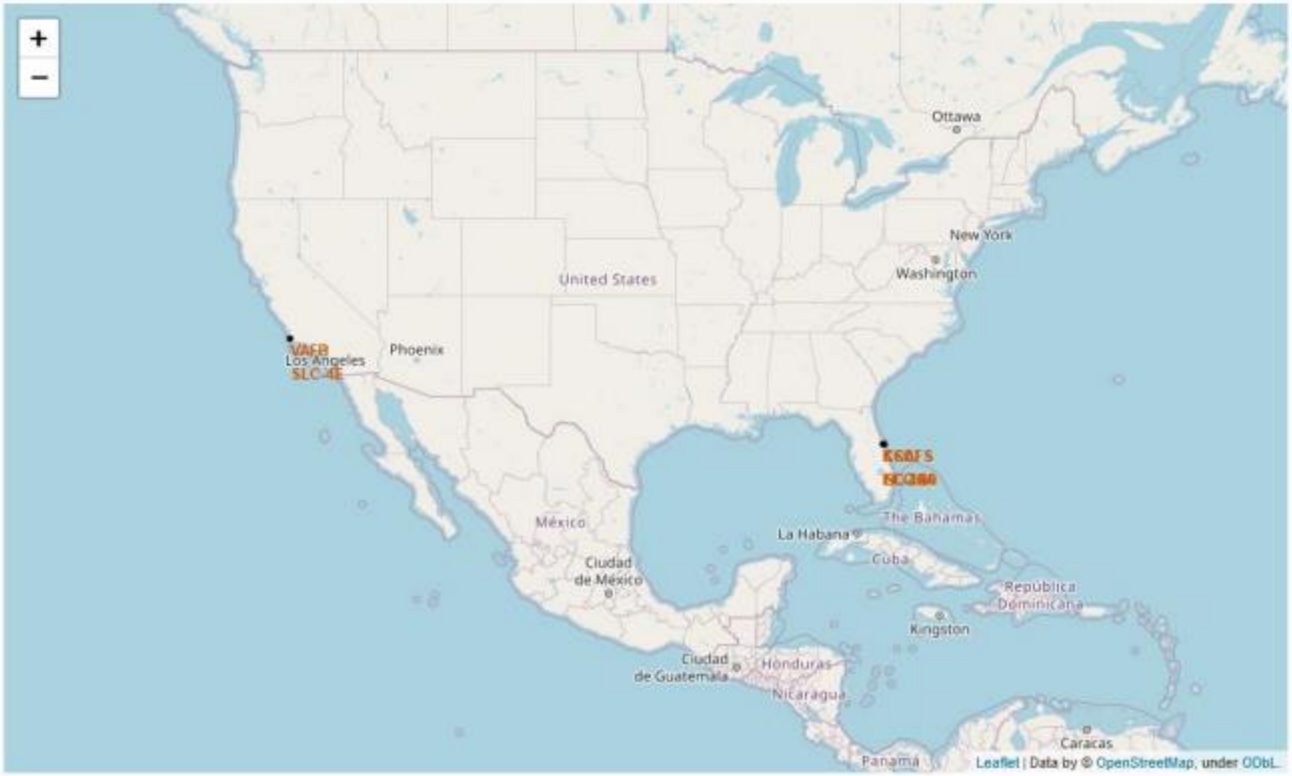
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

All launch sites

- Launch sites are near sea, probably by safety, but not too far from roads and railroads.



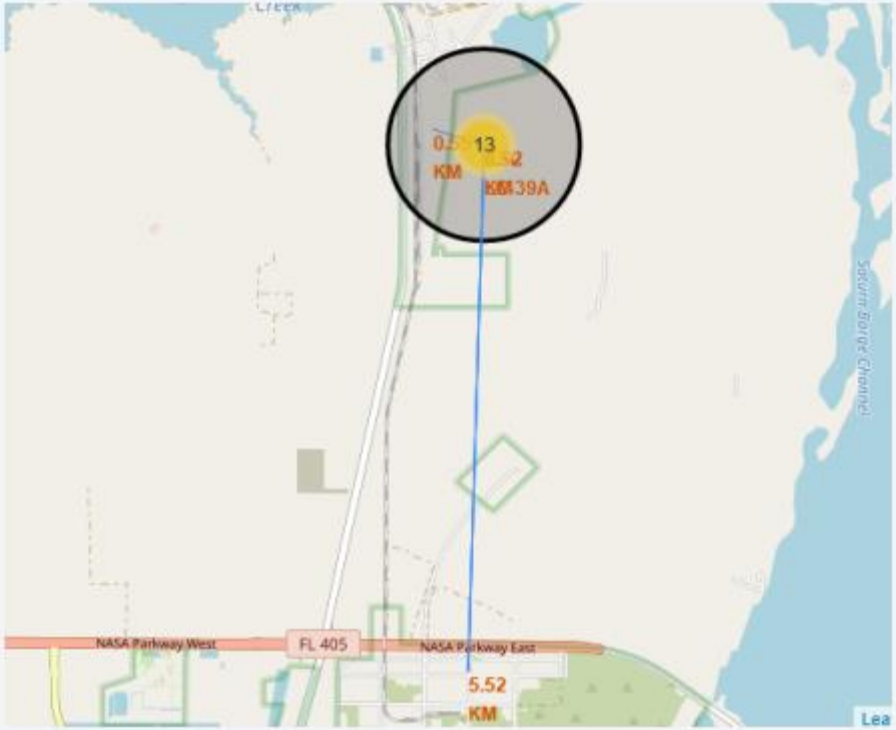
Launch Outcomes by Site

- Green markers indicate successful and red ones indicate failure.

KSC LC-39A

Logistics and Safety

- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas.

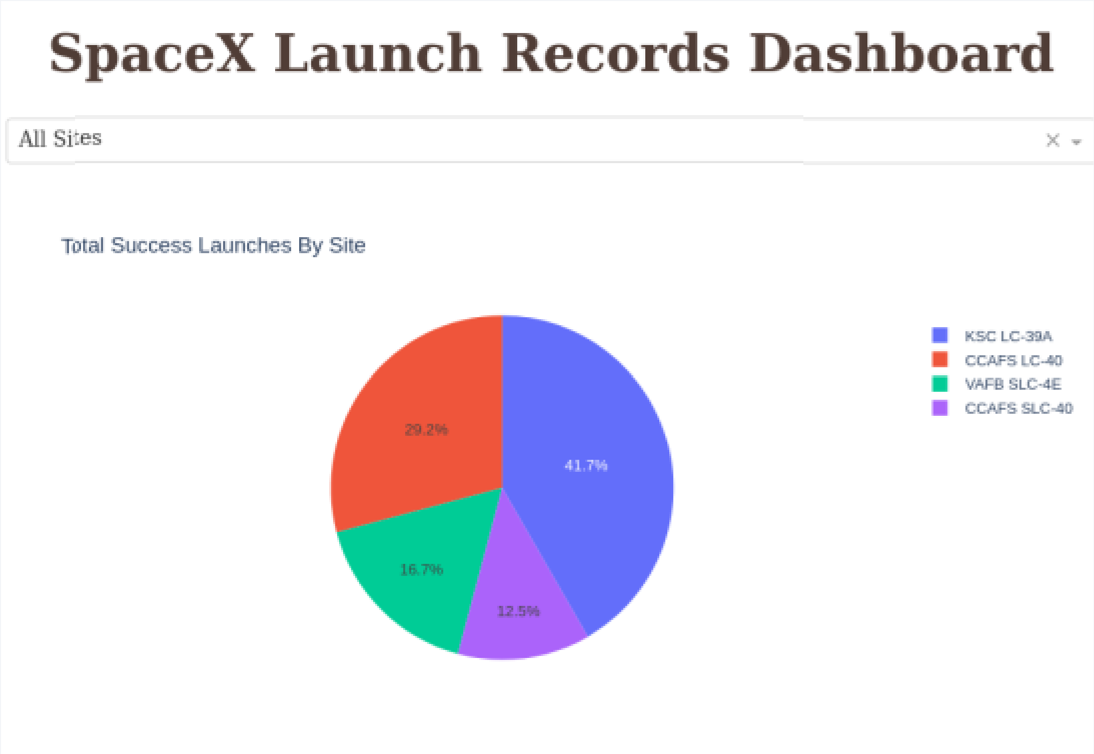


Section 4

Build a Dashboard with Plotly Dash

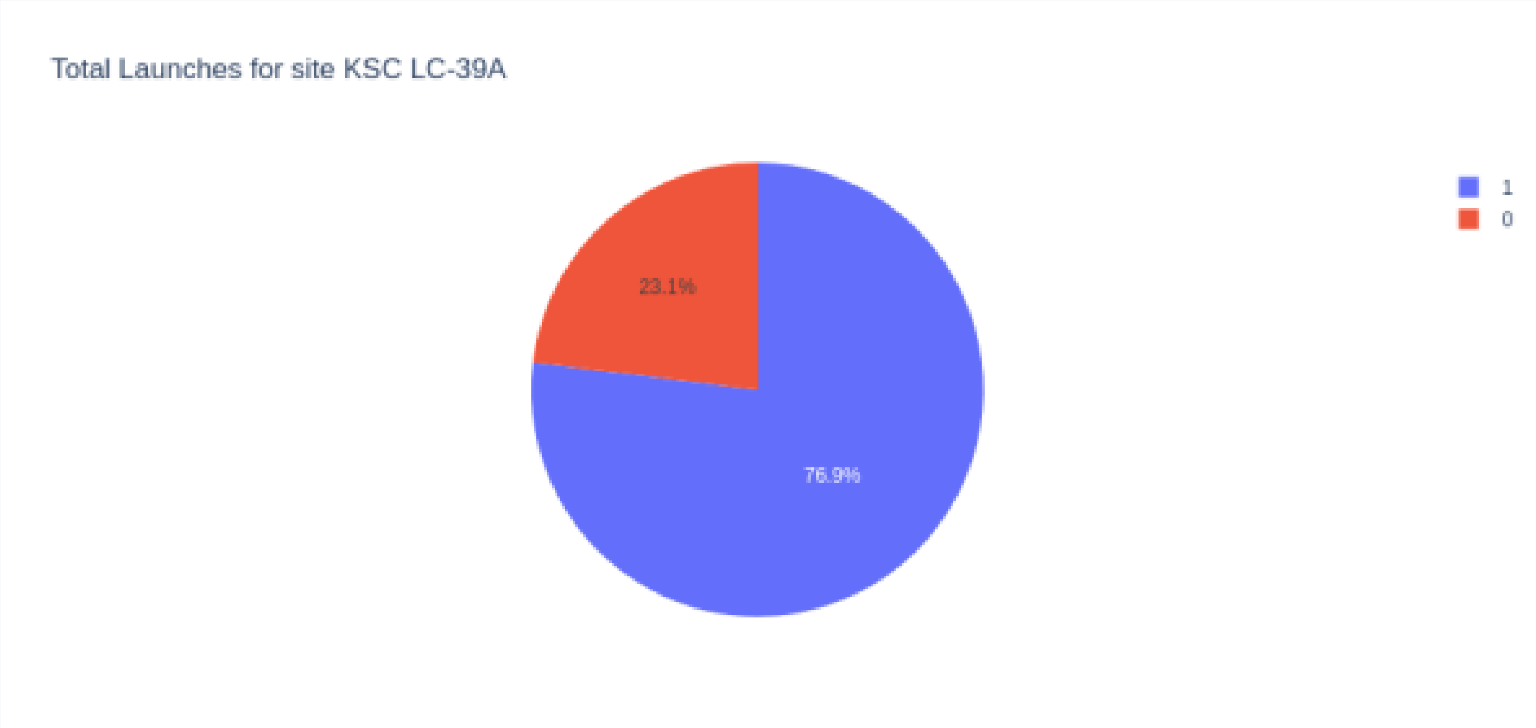
Successful Launches by Site

- The place from where launches are done is a very important factor of success.



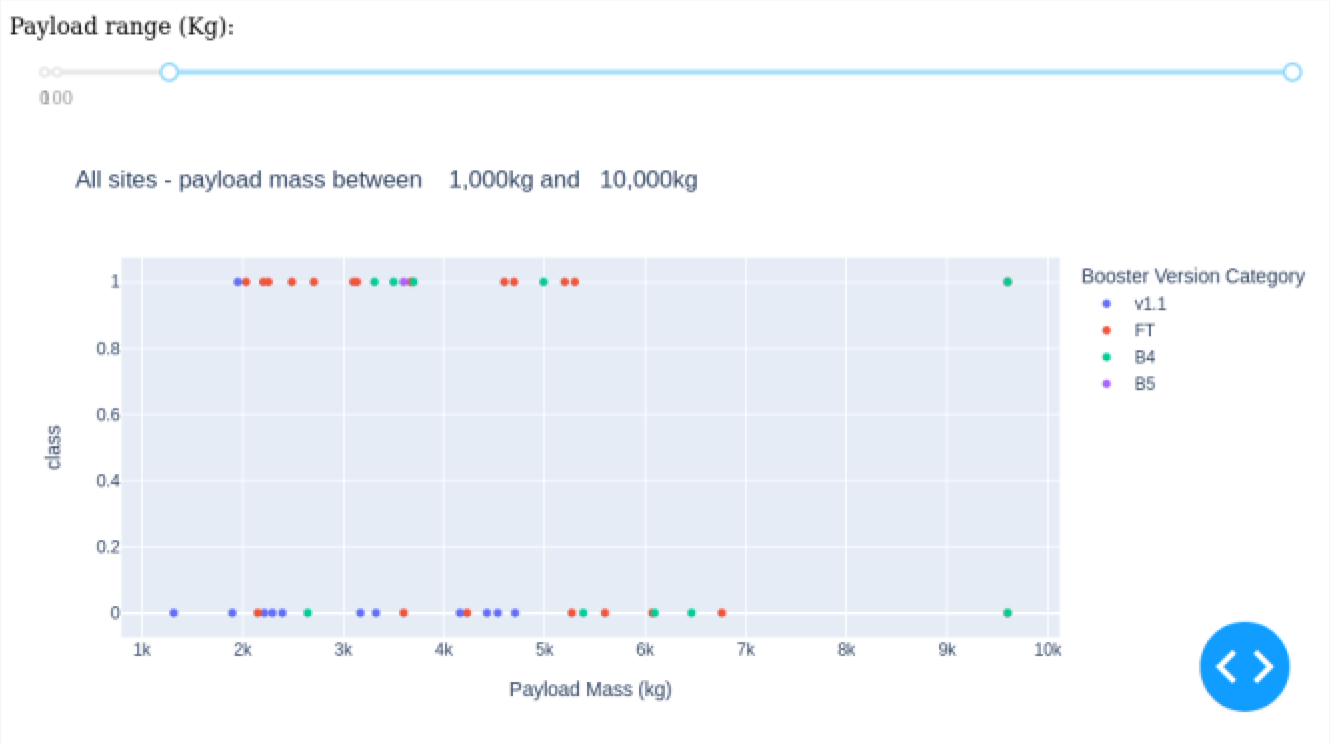
Launc Success Ratio for KSC LC-39A

- 76.9% of launches are successful in this site.



Payload vs Launch Outcome

- Payloads under 6,000kg and FT boosters are the most successful combination.

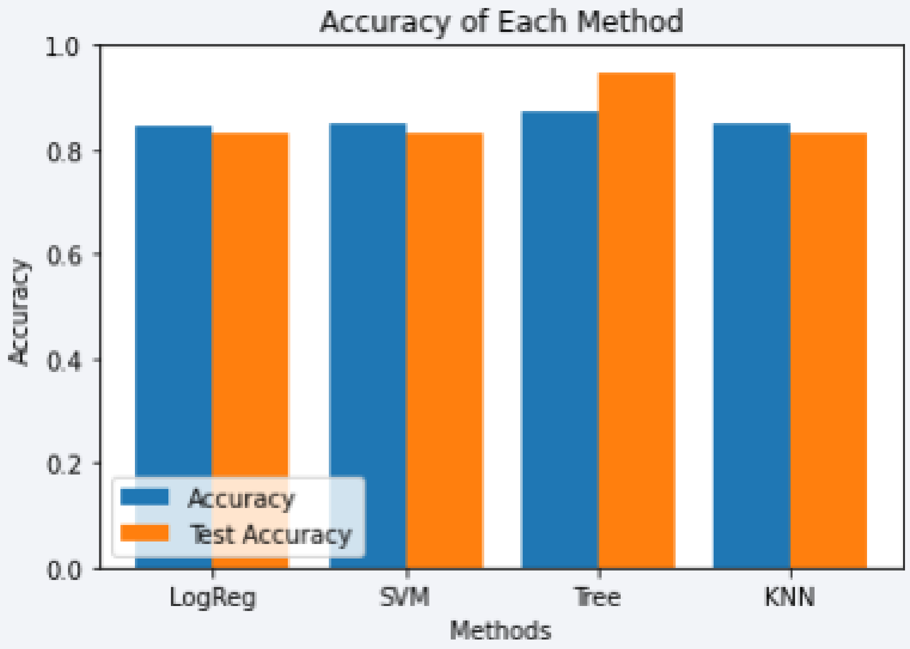


Section 5

Predictive Analysis (Classification)

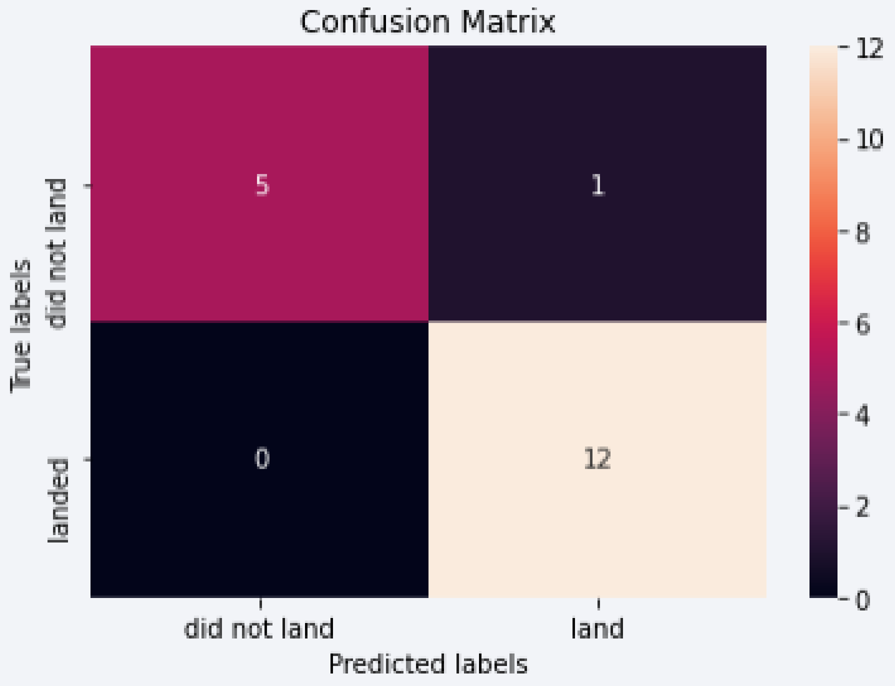
Classification Accuracy

- Four classification models were tested,
- The model with the highest classification accuracy is Decision Tree Classifier, (Accuracy over 87%).



Confusion Matrix

- Confusion matrix of Decision Tree Classifier proves its accuracy



Conclusions

- The best launch site is KSC LC-39A;
- Launches above 7,000kg are less risky;
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time;
- Decision Tree Classifier can be used to predict successful landings and increase profits.

Appendix

It's important to set a value to np.random.seed variable to replicate always the same result

Thank you!