IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Lakshi VS
04-07-2024

Outline

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

Executive Summary

- **Summary of methodologies**

-Data Collection through API

-Data Collection with Web Scraping

-Data Wrangling

-Exploratory Data Analysis with SQL

-Exploratory Data Analysis with Data Visualization

-Interactive Visual Analytics with Folium

-Machine Learning Prediction

- **Summary of all results**

-Exploratory Data Analysis result

-Interactive analytics in screenshots

-Predictive Analytics result

Introduction

- Project background and context

  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. The goal of the project is to create a machine learning pipeline to predict if the first stage will land successfully.

- Problems you want to find answers

  -What factors determine if the rocket will land successfully?

  -The interaction amongst various features that determine the success rate of a successful landing.

  -What operating conditions needs to be in place to ensure a successful landing program.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using SpaceX API and web scraping from Wikipedia.
- Perform data wrangling
  - One hot encoding was applied to categorical 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
  - How to build, tune, evaluate classification models

Data Collection

- The data was collected using various methods:

  -Data collection was done using get request to the SpaceX API

  -Next, we decoded the response content as a Json using json() function call and turn it into a pandas dataframe using json_normalize

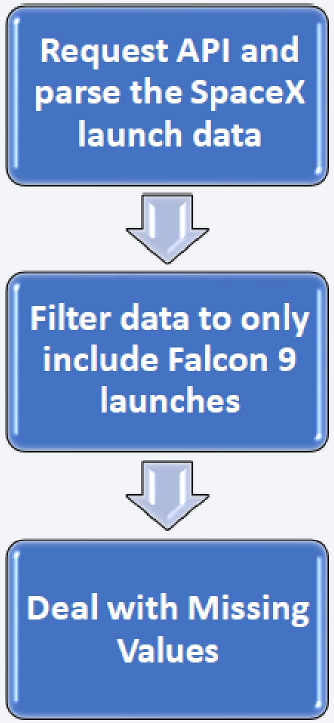
  -We then cleaned the data, checked for missing values and fill in missing values where necessary

  -In addition, we performed web scraping from Wikipedia for Falcon 9 launch records with BeautifulSoup

  -The objective was to extract the launch records as HTML table, parse the table and convert it to a pandas dataframe for future analysis

Data Collection – SpaceX API

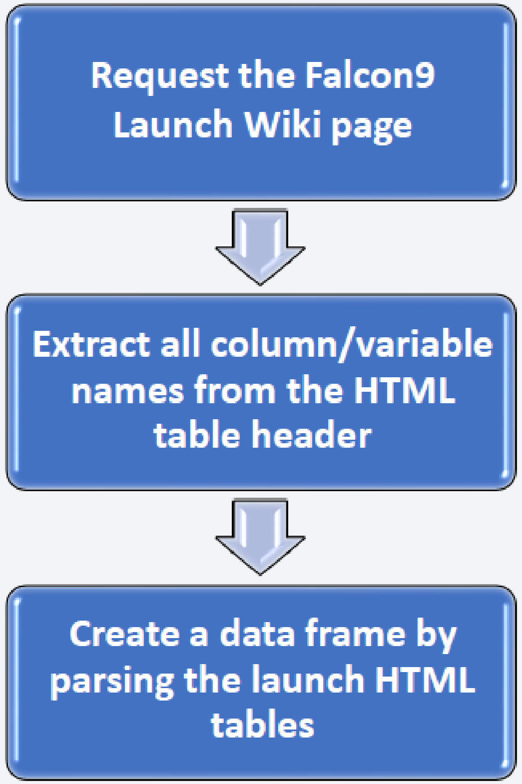
- We used the get request to the SpaceX API to collect data, clean the requested data and did some basic data wrangling and formatting.
- GitHub URL:
  https://github.com/LakshiVS/IBM-Data-Science-Professional-Certificate/blob/main/03.Capstone%20Project/Data%20Collection%20API.ipynb



Data Collection - Scraping

- We applied web scrapping to webscrap Falcon 9 launch records with BeautifulSoup
- We parsed the table and converted it into a pandas data frame
- GitHub URL:

  https://github.com/LakshiVS/IBM-Data-Science-Professional-Certificate/blob/main/03.Capstone%20Project/Data%20Collection%20with%20Web%20Scraping.ipynb



Data Wrangling

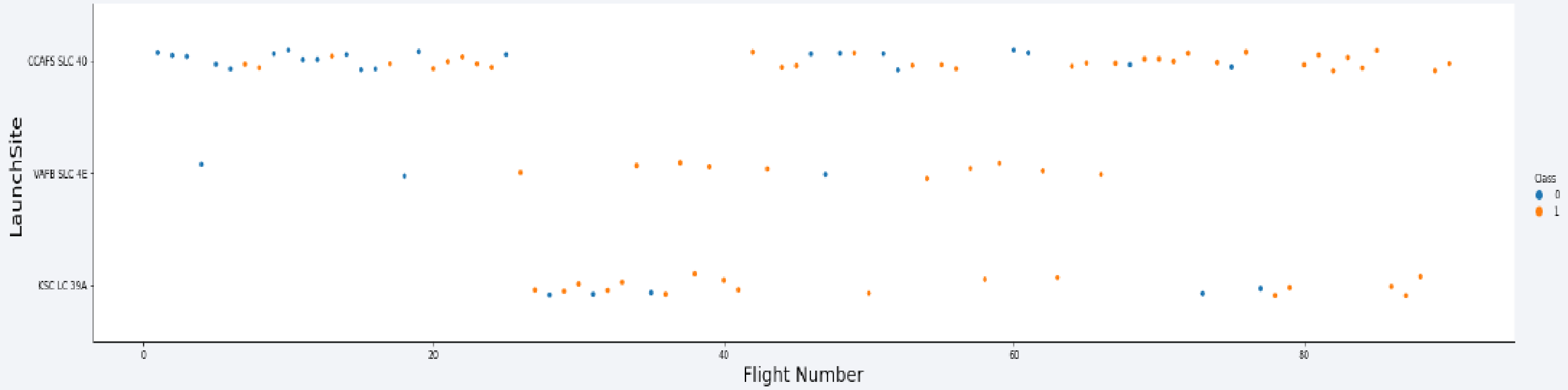
- We performed exploratory data analysis and determined the training labels.
- We calculated the number of launches at each site, and the number and occurrence of each orbits
- We created landing outcome label from outcome column and exported the results to csv.
- GitHub URL :

  https://github.com/LakshiVS/IBM-Data-Science-Professional-Certificate/blob/main/03.Capstone%20Project/Data%20Wrangling.ipynb

EDA with Data Visualization

- To explore data, scatterplots and barplots were used to visualize the relationship between pair offeatures:
- Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, Orbit and Flight Number, Payload andOrbit
- GitHub URL:

https://github.com/LakshiVS/IBM-Data-Science-Professional-Certificate/blob/main/03.Capstone%20Project/EDA%20with%20Data%20Visualization.ipynb



EDA with SQL

- SQL Queries Performed:

  1. Names of the unique launch sites in the space mission;
  2. Top 5 launch sites whose name begin with the string 'X';
  3. Total payload mass carried by boosters launched by NASA;
  4. Average payload mass carried by booster version F9 v1.1;
  5. Date when the first successful landing outcome in ground pad was achieved;
  6. Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg;
  7. Total number of successful and failure mission outcomes;
  8. Names of the booster versions which have carried the maximum payload mass;
  9. Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015;
  10. 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.

  Github link: https://github.com/LakshiVS/IBM-Data-Science-Professional-Certificate/blob/main/03.Capstone%20Project/EDA.ipynb

Build an Interactive Map with Folium

- •Markers, circles, lines and marker clusters were used with Folium Maps
- Markers indicate points like launch sites
- Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center
- Marker clusters indicate groups of events in each coordinate, like launches in a launch
- Lines are used to indicate distances between two coordinates.

- GitHub URL:

  https://github.com/LakshiVS/IBM-Data-Science-Professional-Certificate/blob/main/03.Capstone%20Project/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

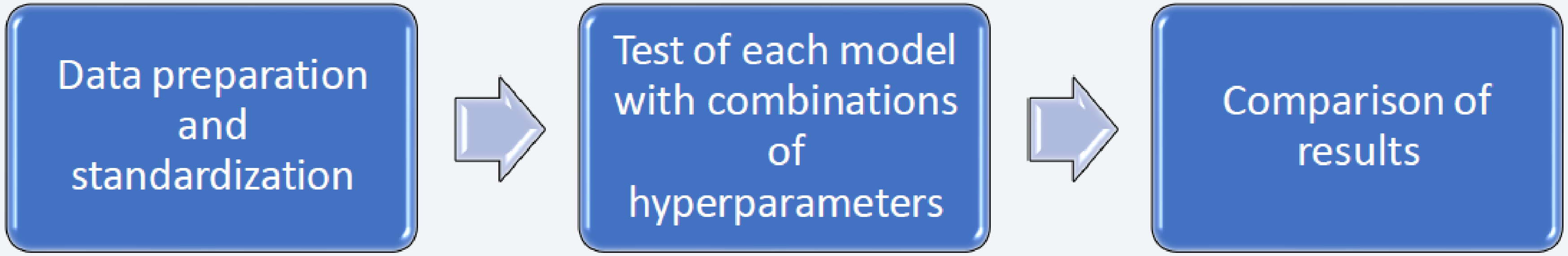
- The following graphs and plots were used to visualize data
  - Percentage of launches by site
  - Payload range
- This combination allowed us to quickly analyze the relation between payloads and launch sites, helping to identify the best place to launch according to payloads.

Github url: https://github.com/LakshiVS/IBM-Data-Science-Professional-Certificate/blob/main/03.Capstone%20Project/spacex_dash_app.py

Predictive Analysis (Classification)

- Four classification models were compared: logistic regression, support vector machine, decision tree and k nearestneighbors.
- GitHub URL:

  https://github.com/LakshiVS/IBM-Data-Science-Professional-Certificate/blob/main/03.Capstone%20Project/Machine%20Learning%20Prediction.ipynb



Results

Exploratory data analysis results:

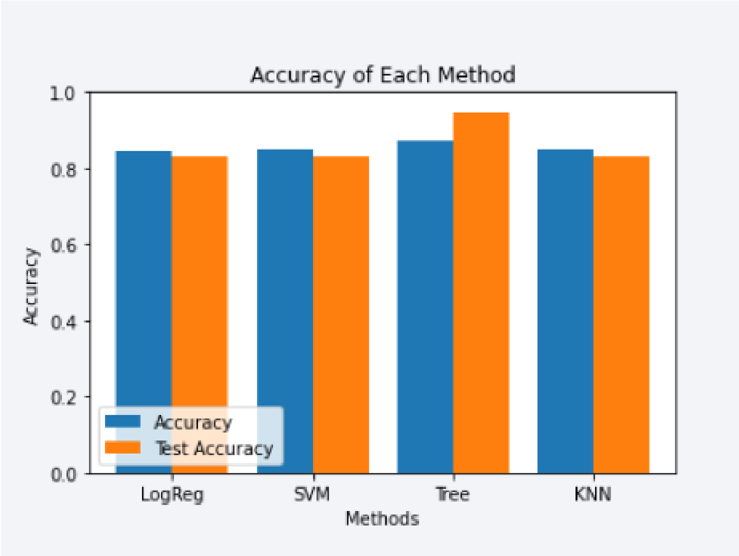
- Space X uses 4 different launches;
- The first launches were done to Space Xitself and NASA;
- The average payload of F9 v1.1 booster is 2,928kg;
- The first successful landing outcome happened in 2015 five year after the first launch;
- Many Falcon 9 booster versions were successful at landing in drone ships having payloads above the average;
- Almost 100% of mission outcomes were successful;
- Twoboosterversionsfailedatlandingindroneshipsin2015:F9v1.1B1012andF9v1.1B1015;
- The number of landing outcomes became as better as the years passed.

Results

- Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around.
- Most launches happens at east cost launch sites.

Results

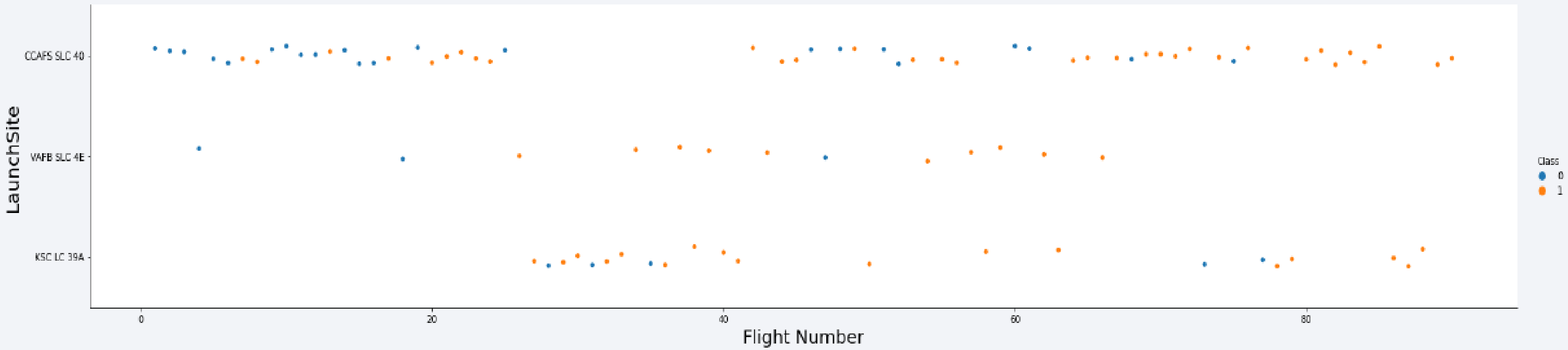
- Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%.



Section 2

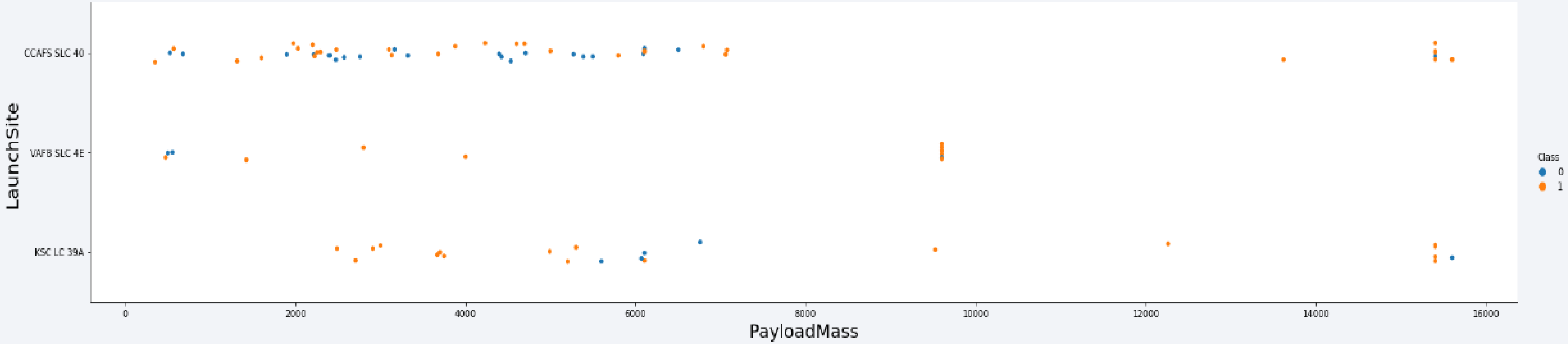
Insights drawn from EDA

Flight Number vs. Launch Site



- According to the plot above, it's possible to verify that the best launch site nowadays is CCAF5 SLC40, where most of recent launches were successful;
- In second place VAFB SLC4E and third place KSCLC 39A;
- It's also possible to see that the general success rate improved over time.

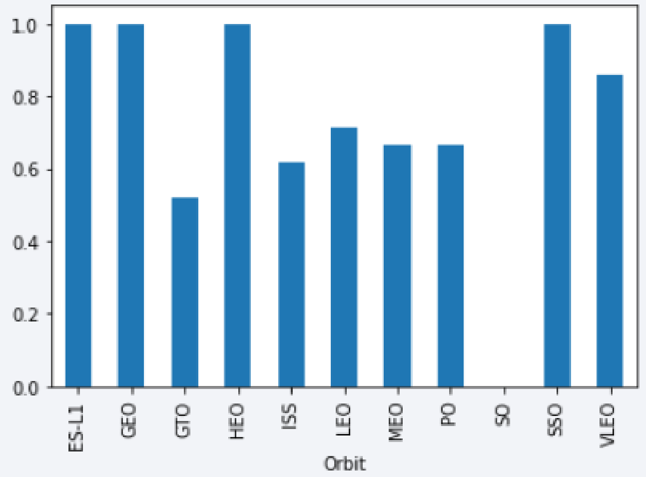
Payload vs. Launch Site



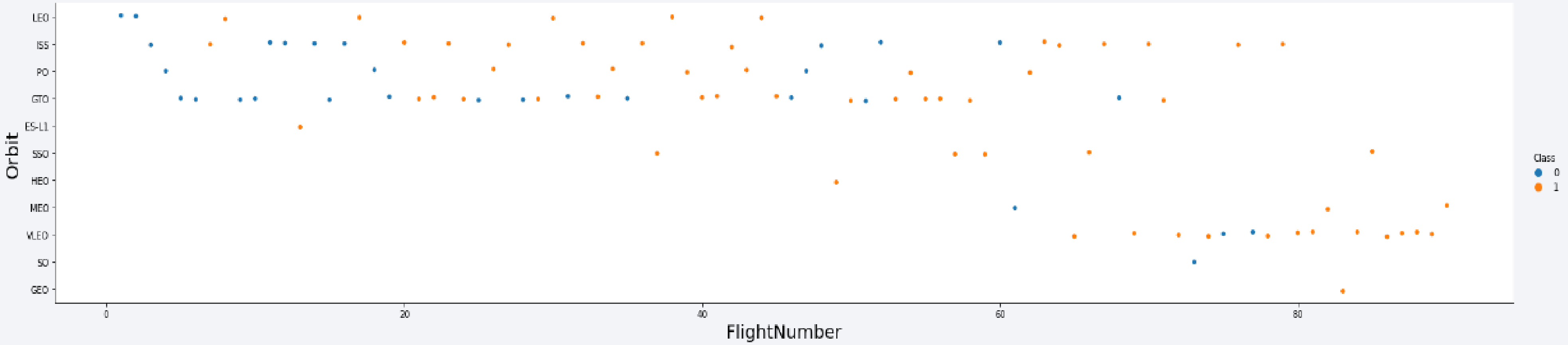
- Payloads over 9,000kg (about the weight of a school bus) have excellent success rate;
- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type

- The biggest success rates happens to orbits:
  - ES-L1;
  - GEO;
  - HEO; and
  - SSO.
- Followed by:
  - VLEO (above 80%); and
  - LFO (above 70%).

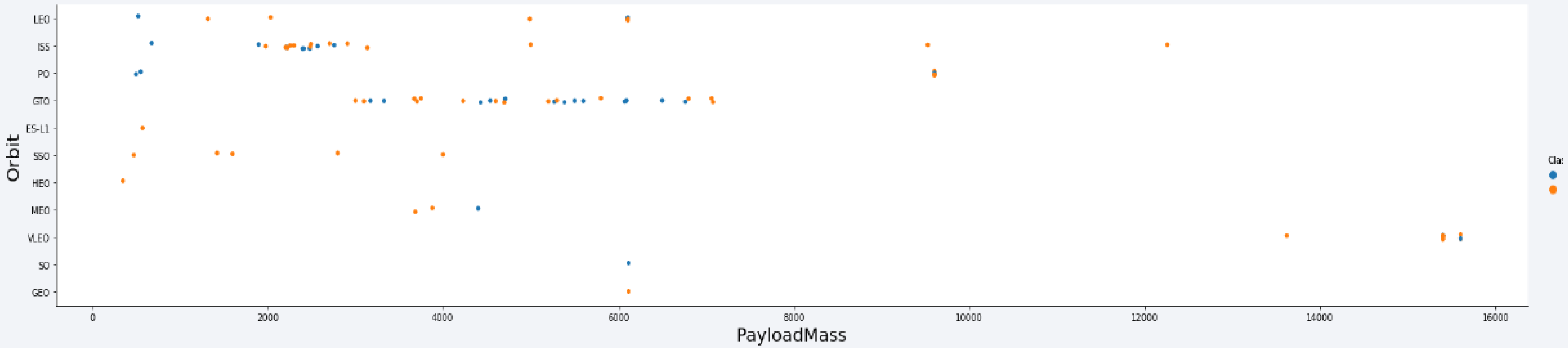


Flight Number vs. Orbit Type



- Apparently, success rate improved over time to all orbits;
- VLEO orbit seems a new business opportunity, due to recent increase of its frequency.

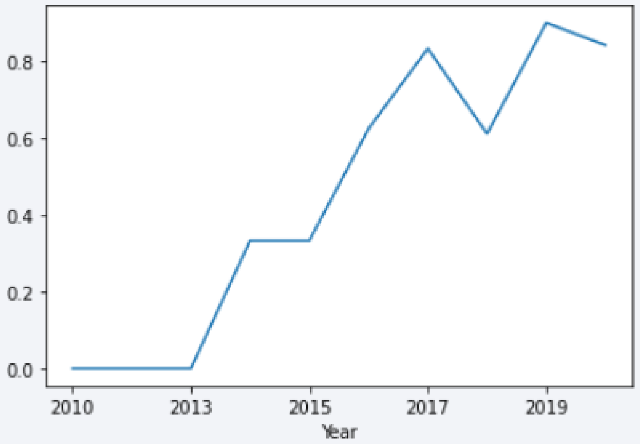
Payload vs. Orbit Type



- Apparently, there is no relation between payload and success rate to orbit GTO;
- ISS orbit has the widest range of payload and a good rate of success;
- There are few launches to the orbits SO and GEO.

Launch Success Yearly Trend

- Success rate started increasing in 2013 and kept until 2020;
- It seems that the first three years were a period of adjusts and improvement of technology.



All Launch Site Names

- According to data, there are four launch sites:

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

- They are obtained by selecting unique occurrences of “launch_site” values from the dataset.

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

- Here we can see five samples of Cape Canaveral launches.

Total Payload Mass

- Total payload carried by boosters from NASA:

Total Payload (kg)
111.268

- Total payload calculated above, by summing all payloads whose codes contain 'CRS', which corresponds to NASA.

Average Payload Mass by F9 v1.1

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

Avg Payload (kg)
2.928

- Filtering data by the booster version above and calculating the average payload mass we obtained the value of 2,928 kg.

First Successful Ground Landing Date

- First successful landing outcome on ground pad:

Min Date
2015-12-22

- 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, that happened on 12/22/2015.

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

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

- Selecting distinct booster versions according to the filters above, these 4 are the result.

Total Number of Successful and Failure Mission Outcomes

- Number of successful and failure mission outcomes:

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

- Grouping mission outcomes and counting records for each group led us to the summary above.

Boosters Carried Maximum Payload

- Boosters which have carried the maximum payload mass

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

- These are the boosters which have carried the maximum payload mass registered in the dataset.

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

- The list above has the only two occurrences.

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:

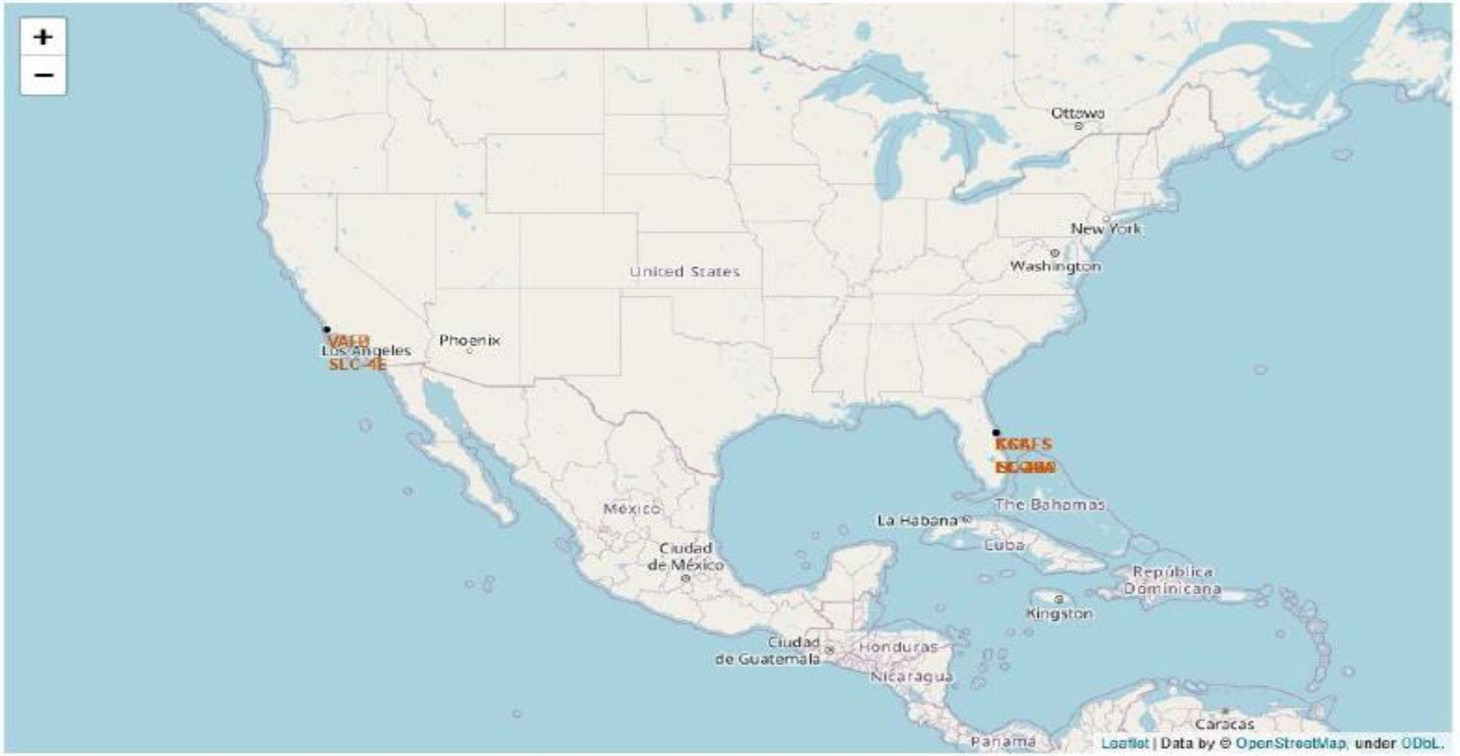
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- This view of data alerts us that “No attempt” must be taken in account.

Section 3

Launch Sites Proximities Analysis

All launch sites



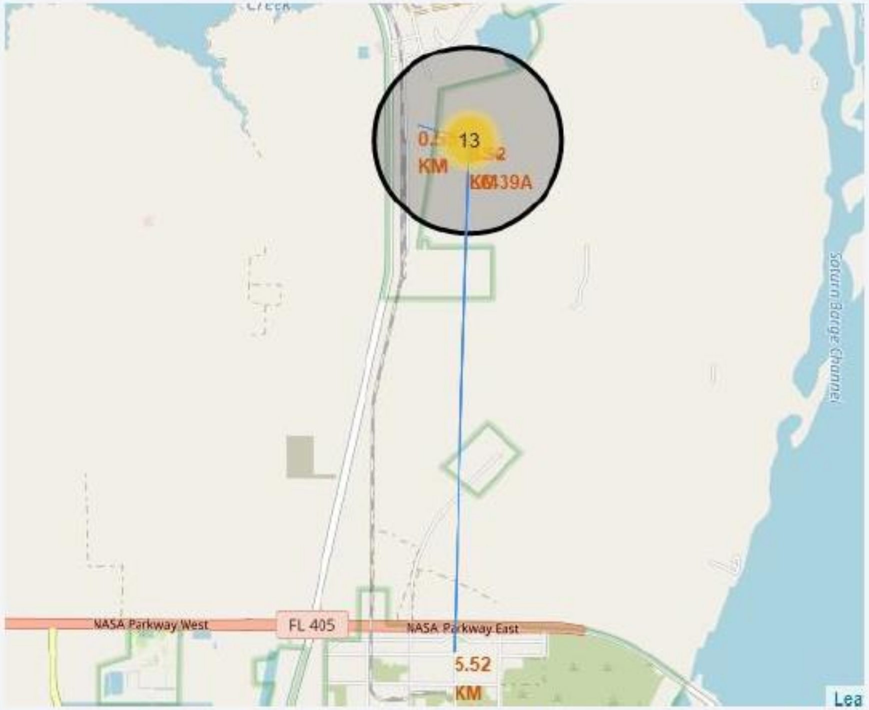
- Launch sites are near sea, probably by safety, but not too far from roads and railroads.

Launch Outcomes by Site

- Example of KSCLC-39A launch site launch outcomes

- Green markers indicate successful and red ones indicate failure.

Logistics and Safety

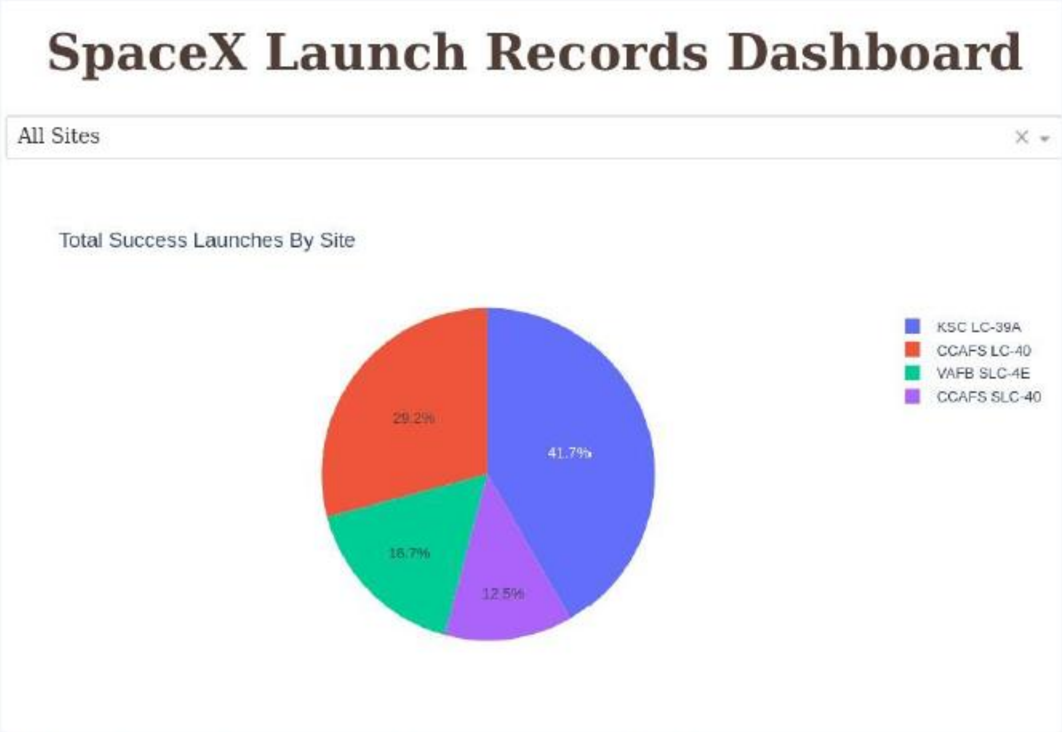


- Launch site KSCLC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas.

Section 4

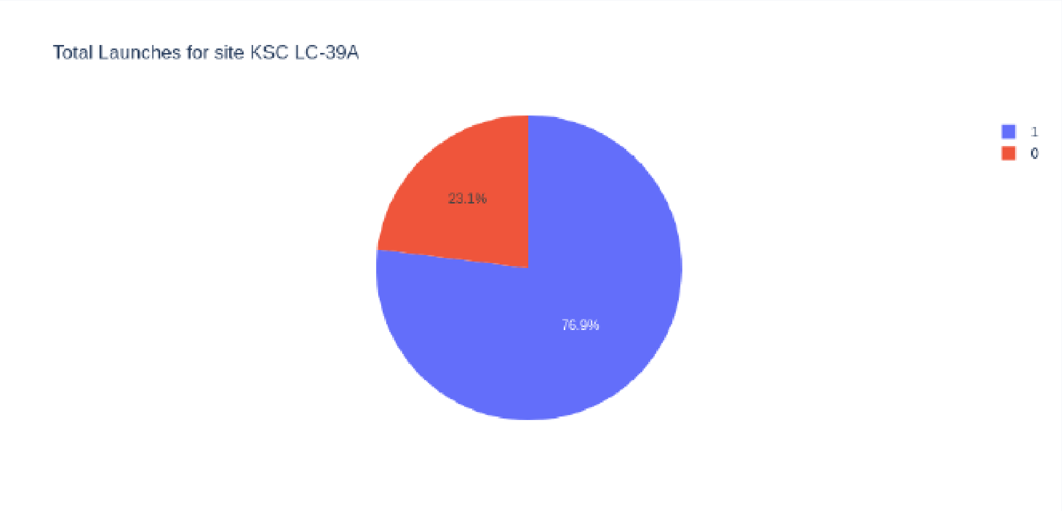
Build a Dashboard with Plotly Dash

Successful Launches by Site



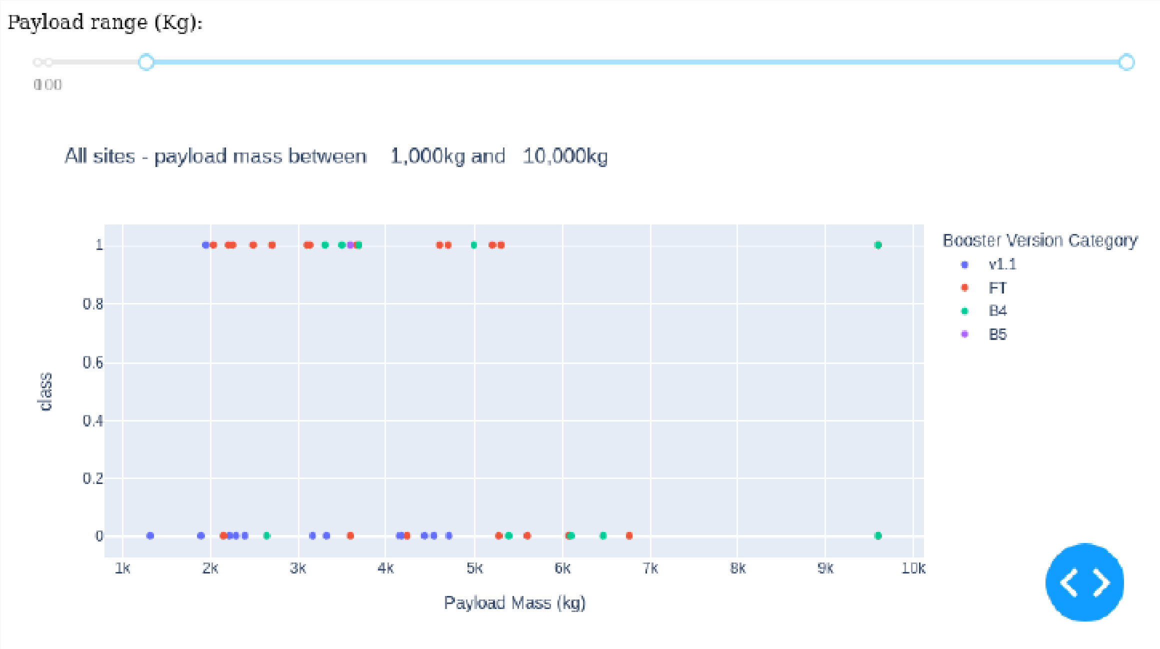
- The place from where launches are done seems to be a very important factor of success of missions.

Launch Success Ratio for KSCLC-39A



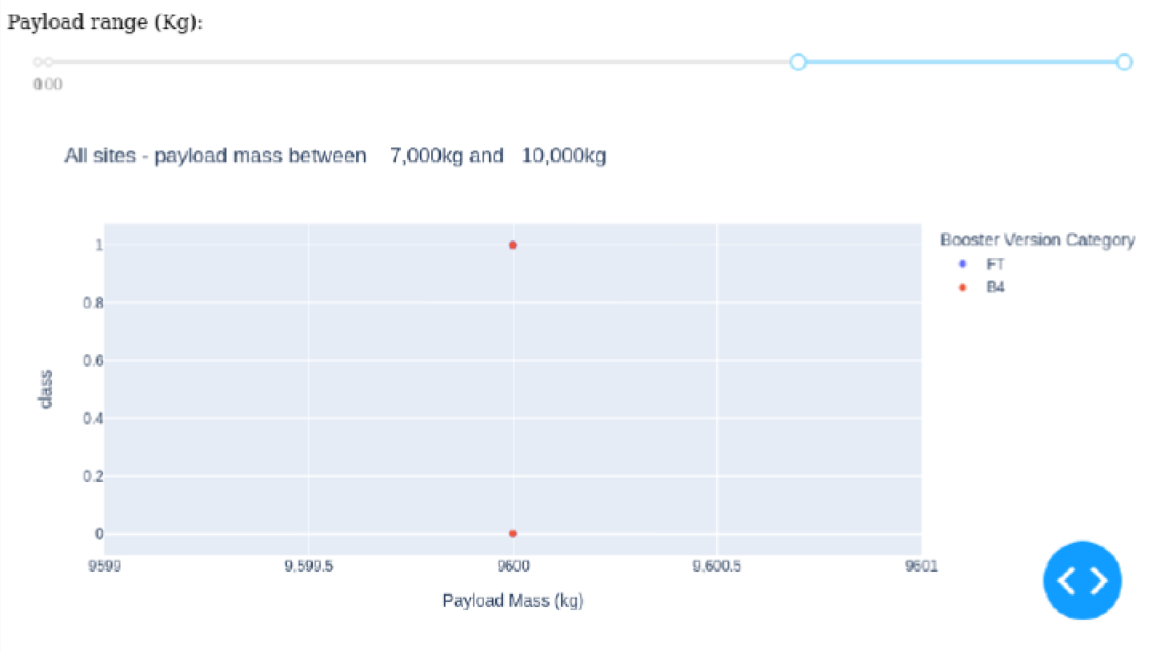
- 76.9% of launches are successful in this site.

Payload vs. Launch Outcome



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

Payload vs. Launch Outcome



- There's not enough data to estimate risk of launches over 7,000kg

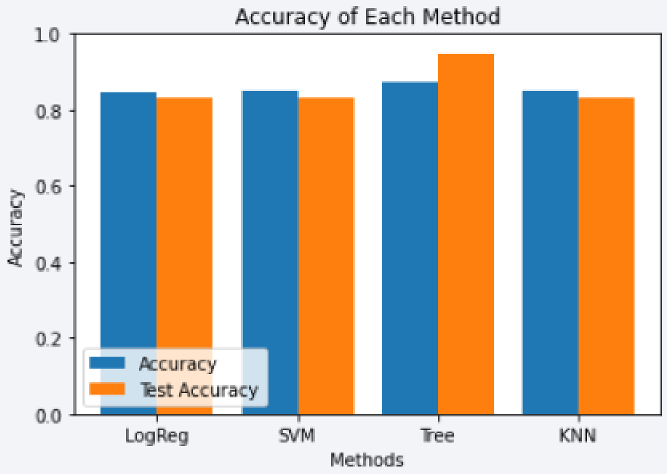
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Four classification models were tested, and their accuracies are plotted beside;

- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87%.



Confusion Matrix

- Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.

Conclusions

- Different data sources were analyzed, refining conclusions along the process
- The best launch site is KSCLC-39A;
- Launches above 7,000kg are less risky;
- Although most mission outcomes are successful, successful landing outcomes seem to improve over time, according to the evolution of processes and rockets;
- Decision Tree Classifier can be used to predict successful landings increase profits.

Thank you!