IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Dikshant Buwa
26 January

Outline

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

Executive Summary

- **Summary of methodologies:**
- Data Collection using web scraping and SpaceX API.
- Exploratory Data Analysis (EDA), including data wrangling, data visualization and interactive. Visual analytics.
- Machine Learning Prediction.

- **Summary of all results:**

- It was possible to collected valuable data from public sources.
- EDA allowed to identify which features are the best to predict success of launchings.
- Machine Learning Prediction showed the best model to predict which characteristics are important to drive this opportunity by the best way, using all collected data.

Introduction

- Project background and context:
- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars and other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.

- Problems you want to find answers:
- We need to determine the Falcon 9 rocket will land on first stage or not.
- If we could solve above problem then we can determine the cost of a launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data Collection using SpaceX API and using json normalize method
- Perform data wrangling
  - Collected data was enriched by creating a landing outcome label based on outcomedata after summarizing and analysing features
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Import CSV data using IBM DB2 resource and extract the needful data for analysis.
- Perform interactive visual analytics using Folium and Plotly Dash
  - Display the data using Folium for Maps and Plotlyfor Dashboard.
- Perform predictive analysis using classification models
  - Data that was collected until this step were normalized, divided in training and test data sets and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters.

Data Collection

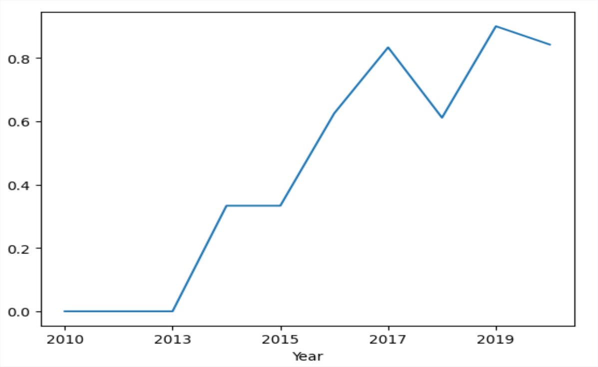
- Requesting rocket launch data from SpaceX API with the following URL
- https://api.spacexdata.com/v4/launches/past
- we then decode the response content as a Json using .json() and turn it into a Pandas dataframe using .json_normalize()

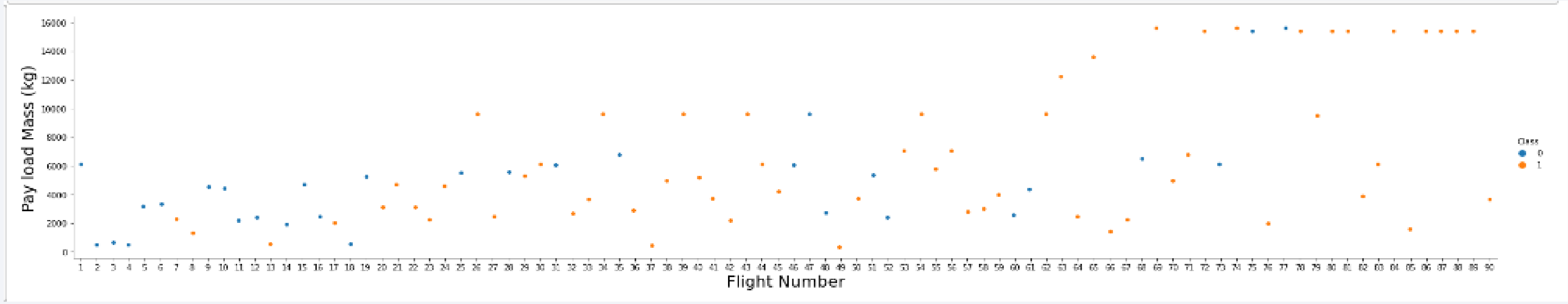
Data Wrangling

- Initially some Exploratory Data Analysis (EDA) was performed on the dataset.
- Then the summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated.
- Finally, the landing outcome label was created from Outcome column.

EDA with Data Visualization

- We see that different launch sites have different success rate's. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39AandVAFB SLC 4E has a success rate of 77%.
- The success rate since 2013 kept increasing till 2020.





EDA with SQL

- The following SQL queries were performed:
  - 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 F9v1.1.
  - 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 6000kg.
  - Total number of successful and failure mission outcomes.
  - Names of the booster versions which have carried the maximum payload mass.
  - Failedlandingoutcomesindroneship,theirboosterversions,andlaunchsitenamesforin year 2015.
  - 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.

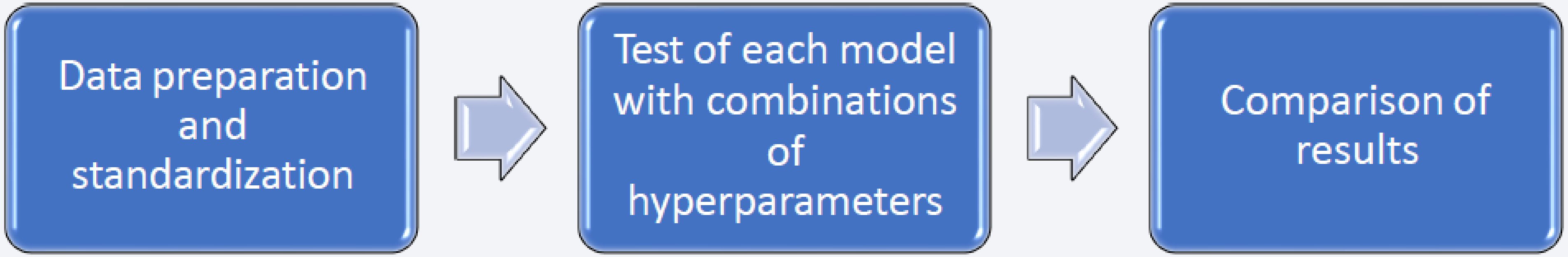
Build an Interactive Map with Folium

- Marker’s, 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 indicates groups of events in each coordinate, like launches in a launch site.
- Lines are used to indicate distances between two coordinates.

Build a Dashboard with Plotly Dash

- We created a Plotly Dash application to perform interactive visual analytics on SpaceX launch data in real-time.
- We used pie charts, rangeslider and scatter plots to visualize data.
- Pie charts for the percentage of successful launches by site, in order to determine the best launch site.
- Rangeslider allows to select a payload mass in a range.
- Scatter plots to study the relation between payloads and launch sites, in order to better understand the best launch sites according to payloads.

Predictive Analysis (Classification)



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

Results

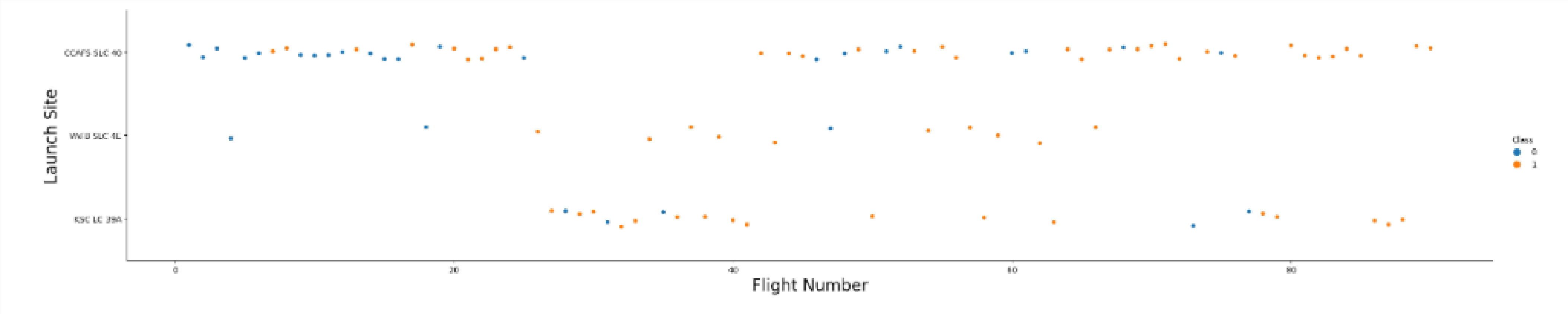
- Space X uses 4 different launch sites.
- The first launches were done to SpaceX itself and NASA.
- The average payload of F9 v1.1 booster is 2,928kg.
- The first success landing outcome happened in 2015 fiver year after the first launch.
- Many Falcon 9 booster versions were successful at landing in 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:F9v1.1B1012andF9v1.1B1015.
- The number of landing outcomes became as better as years passed.

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- Most of first launches were performed in CCAFS SLC 40.
- The success rate improves over time for every site.

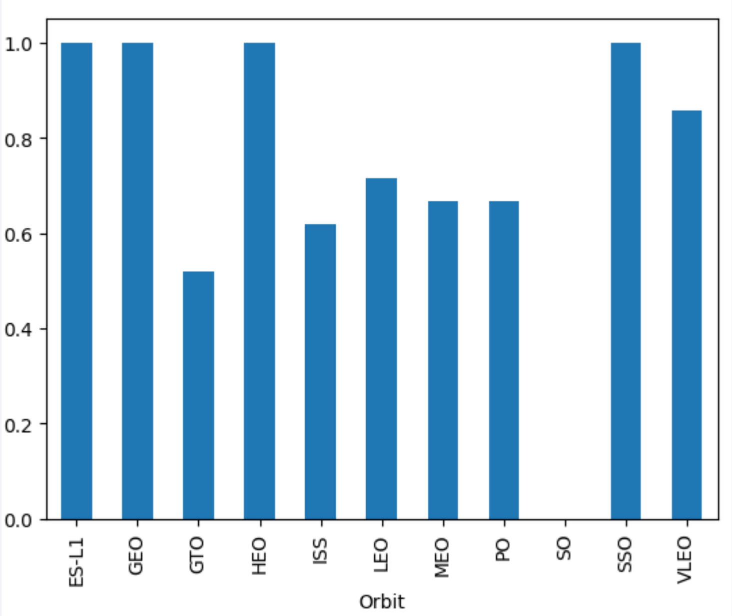


Payload vs. Launch Site

- 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 KSCLC 39A launch sites.

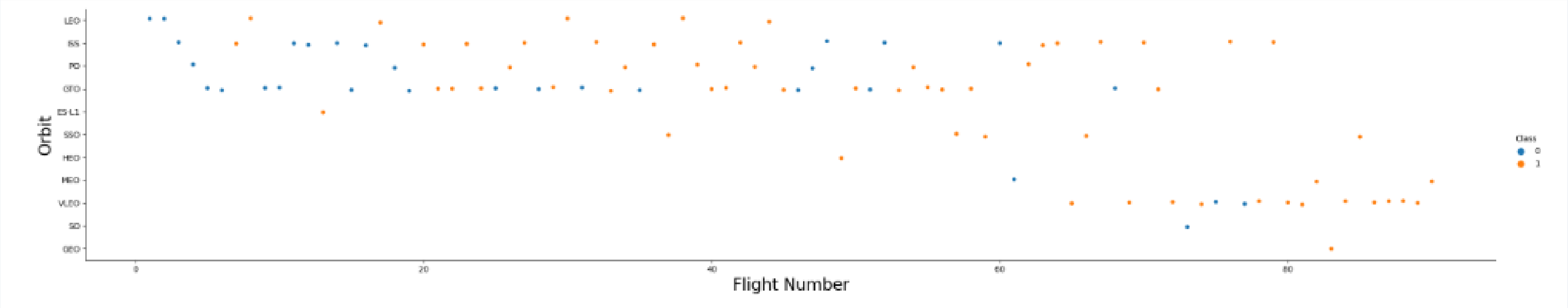
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO have the highest success rate
- SO and GTO have the worst success rate



Flight Number vs. Orbit Type

- LEO: the success related to the number of flights.
- GTO: No relation between flight number and success.
- VLEO: Most of successful launches in the last period.

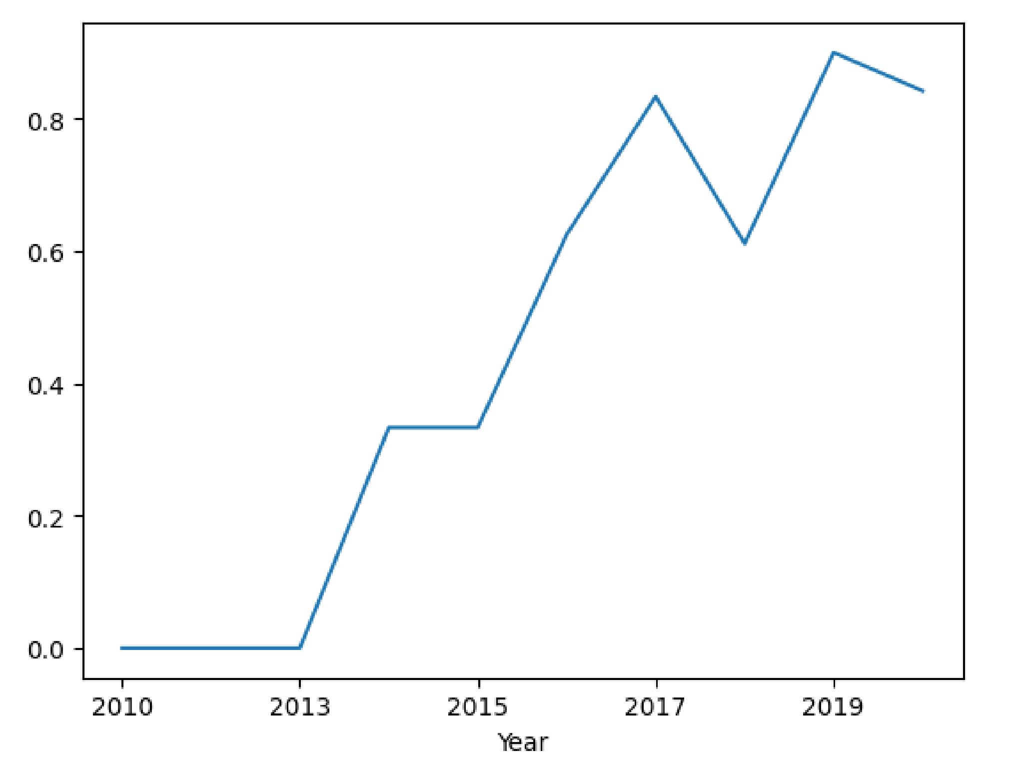


Payload vs. Orbit Type

- 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
- Success rate Zero between 2010 and 2013



All Launch Site Names

- According to data, there are four launchsites:

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.27

Launch Site Names Begin with 'CCA'

- First 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
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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
2012-08-10	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Total payload mass (in kg) carried by boosters launched by NASA (CRS).

total_payload_mass_nasa_crs
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1

avg_payload_mass_f9v11
2928

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad.

date_first_groundpad_success
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 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

- The total number of successful mission outcomes are 99 and failure mission outcomes are only 1.

mission_outcome	total
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass.

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

- List the 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

- 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.

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

Section 3

Launch Sites Proximities Analysis

<Folium Map Screenshot 1>

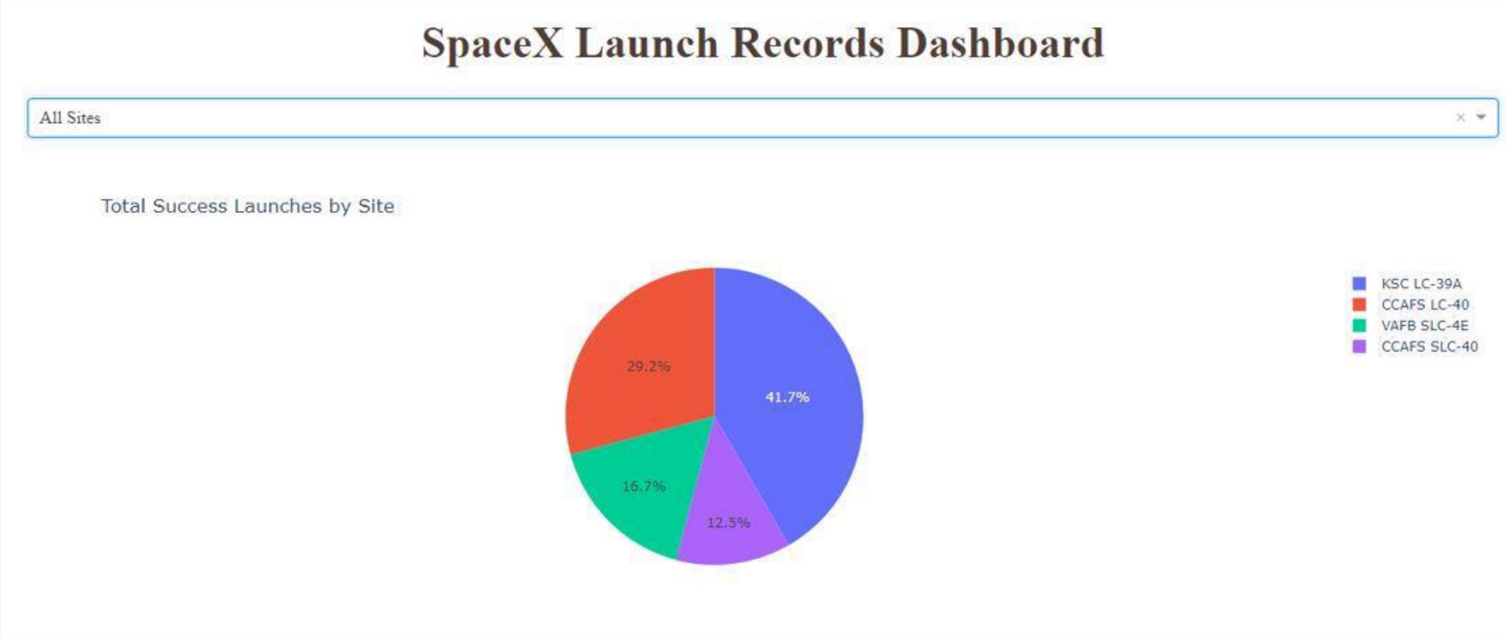
- It Includes all launch sites' location markers -

Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- This pie-chart shows us the success rate for launches by site.

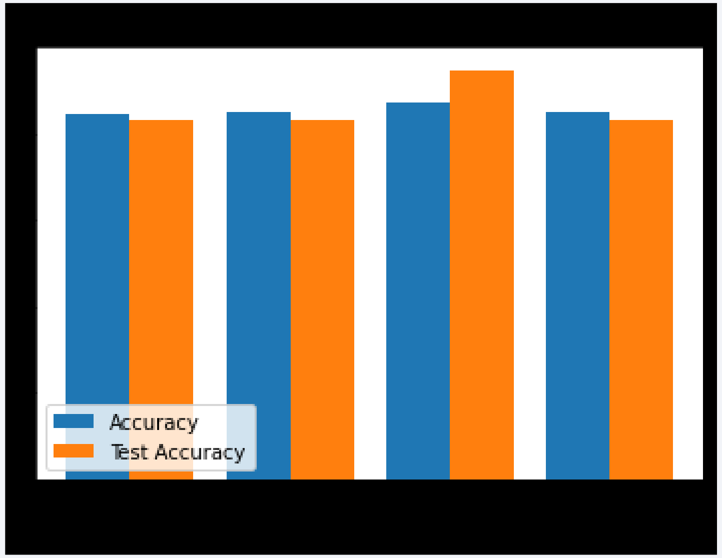


Section 5

Predictive Analysis (Classification)

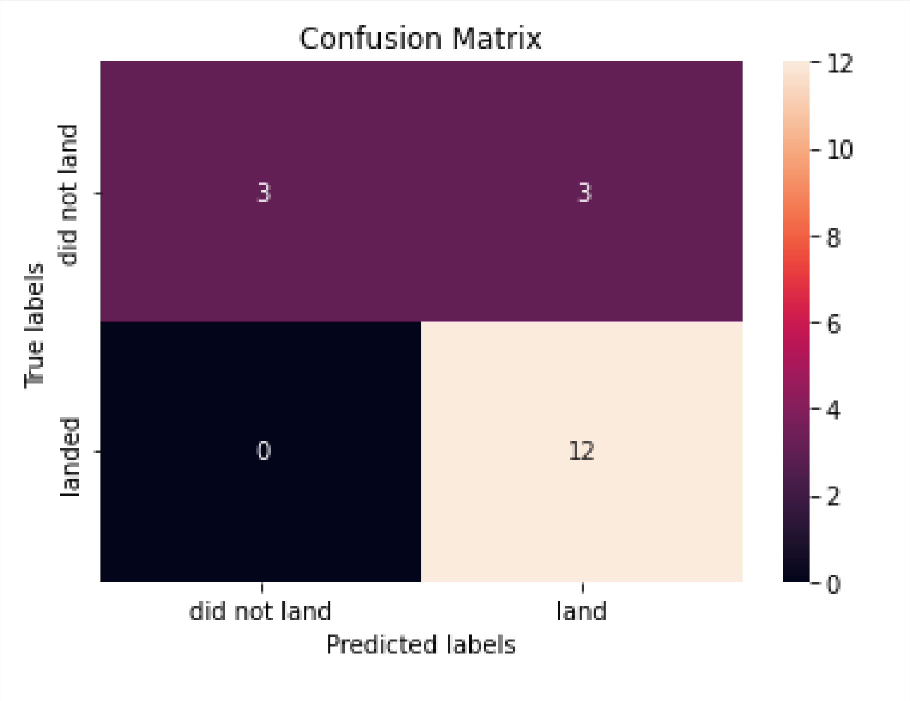
Classification Accuracy

- 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

- The Decision Tree model performed the best.



Conclusions

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

Thank you!