Vivek A

Click here to view the github

Analysing SpaceX Falcon 9 Landings

IBM DATA SCIENCE CAPSTONE PROJECT

Outline

Executive Summary

Summary of methodologies

-Data collection

-Data wrangling

-EDA with data visualization -EDA with SQL

-Building an interactive map with Folium

-Building a Dashboard with Plotly Dash

-Predictive analysis (Classification)

Summary of all results

-Exploratory data analysis results

-Interactive analytics demo in screenshots

-Predictive analysis results

Introduction

Project Background and Context

Objective: Predict if the Falcon 9 first stage will land successfully.

- SpaceX advertises Falcon 9 rocket launches for 62 million dollars, while other providers cost upwards of 165 million dollars.
- Much of the cost savings are due to SpaceX's ability to reuse the first stage of the rocket.
- Predicting the landing success of the first stage can help determine the cost of a launch.
- This information can be valuable for alternate companies looking to bid against SpaceX for rocket launches.

Introduction

Problems to Address

Objective: Predict if the Falcon 9 first stage will land successfully.

- Prediction Goal: Determine the likelihood of a rocket successfully landing.
- Influencing Factors: Identify the factors that influence the successful landing of a rocket.
- Optimal Conditions: Determine the optimal conditions to achieve a successful landing.

Methodology

- Data collection methodology
- Perform data wrangling
- 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 Collection

Data collection process involved a combination of API requests from Space X public API and web scraping data from a table in Space X's Wikipedia entry.

Space X API DataColumns:

FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude

Wikipedia Webscrape Data Columns:

Flight No., Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time

Data Source

Data Features

Data collection - Web Scrapping

Github url for data collection using web scrapping

01 Request Wikipedia html

02 BeautifulSoup html5lib Parser

03 Find launch info html table

04 Create dictionary

05 Iterate through table cells to extract data to dictionary

06 Cast dictionary to DataFrame

Data Wrangling

Github url for data collection using web scrapping

Training label

- The outcome field details two components: ‘mission outcome’ and ‘landing location’
- We want to create a training label ‘Class’ to indicate successful landing = 1; unsuccessful landing = 0
- Value mapping:
  - Outcomes ‘True ASDS’, ‘True RTLS’, & ‘True Ocean’ – set Class to -> 1
  - Outcomes ‘None None’, ‘False ASDS’, ‘None ASDS’, ‘False Ocean’, ‘False RTLS’ – set Class to -> 0

EDA With Sql

Github url for data collection using web scrapping

Goal

- To better understand the data, the dataset is loaded into IBM DB2 Database and queried using SQL magic in Python.

Queries

1. Unique Launch Sites: Display the names of unique launch sites in the space mission.
2. Launch Sites Starting with 'CAA': Display 5 records where launch sites begin with the string 'CAA'.
3. Total Payload Mass by NASA (CRS): Display the total payload mass carried by boosters launched by NASA (CRS).
4. Average Payload Mass by Booster Version F9 v1.1: Display average payload mass carried by booster version F9 v1.1.
5. Successful Drone Ship Landings: List the dates where successful landing outcomes on drone ships were achieved.
6. Boosters with Ground Pad Success and Specific Payload Mass: List the names of boosters with success in ground pad landings and payload mass between 4000 and 6000.
7. Successful and Failed Mission Outcomes: List the total number of successful and failed mission outcomes.
8. Boosters with Maximum Payload Mass: List the names of booster versions that carried the maximum payload mass.
9. 2015 Launch Data: List records showing the month names, successful ground pad landings, booster versions, and launch sites for the year 2015.
10. Ranking Successful Landings (2010-2017): Rank the count of successful landing outcomes between 2010-06-04 and 2017-03-20 in descending order.

EDA With Data Visualization

Github url for data collection using web scrapping

Goal

- Conduct exploratory data analysis (EDA) on variables such as ‘Flight Number,’ ‘Payload Mass,’ ‘Launch Site,’ ‘Orbit,’ ‘Class,’ and ‘Year’ to investigate relationships between these variables.

Charts Plotted

- Scatter Charts:
  - Flight Number vs. Payload Mass
  - Flight Number vs. Launch Site
  - Payload Mass vs. Launch Site
  - Orbit vs. Flight Number
  - Payload Mass vs. Orbit Type
  - Orbit vs. Payload Mass
- **Bar Charts:**
  - Mean Payload Mass vs. Orbit Type
- **Line Charts:**
  - Success Rate vs. Year

Build an Interactive Map with Folium

Github url for data collection using web scrapping

- Mapping Launch Data: Used Folium to create interactive maps of launch sites.
- Labelled Circle Markers: Added circle markers at each launch site using latitude and longitude coordinates.
- Marker Colors: Green for successful landings, red for unsuccessful outcomes, implemented with MarkerCluster().
- Distance Calculation: Calculated distances to key locations such as the nearest railway, highway, coast, and city.
- Visualizing Distances: Marked lines on the map to visualize these distances.

Predictive Analysis

Github url for data collection using web scrapping

01 Split label column ‘Class’ from dataset

02 Use StandardScaler() to fit and transform data

03 Split the data into training and test sets

04 For each model, use test data set to determine accuracy

05 For each model, use GridSearchCV to tune optimal hyper parameters

06 Use train data to fit models: logistic regression, SVM, decision tree, KNN

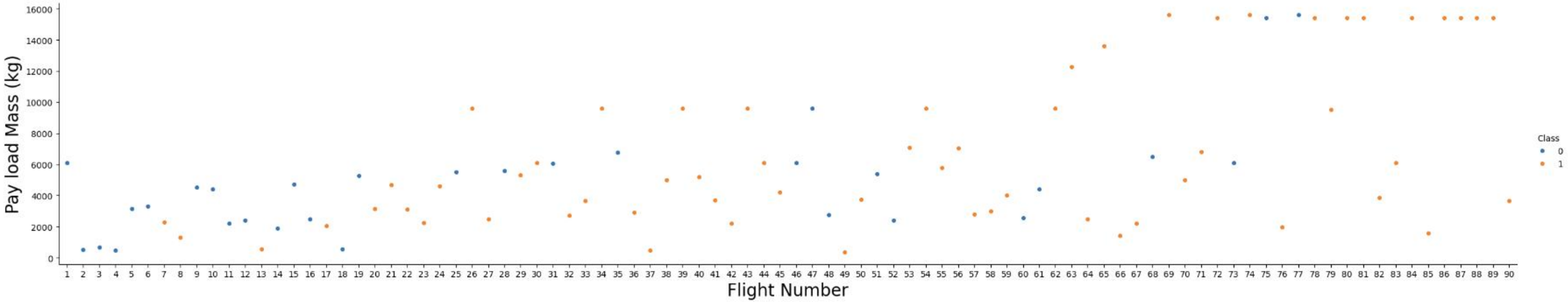
07 For each model, plot a confusion matrix

08 Plot a bar chart to compare accuracy scores of models

Insights from EDA

Payload mass vs flight number

Blue dot indicates unsuccessful and orange indicate successfull landing

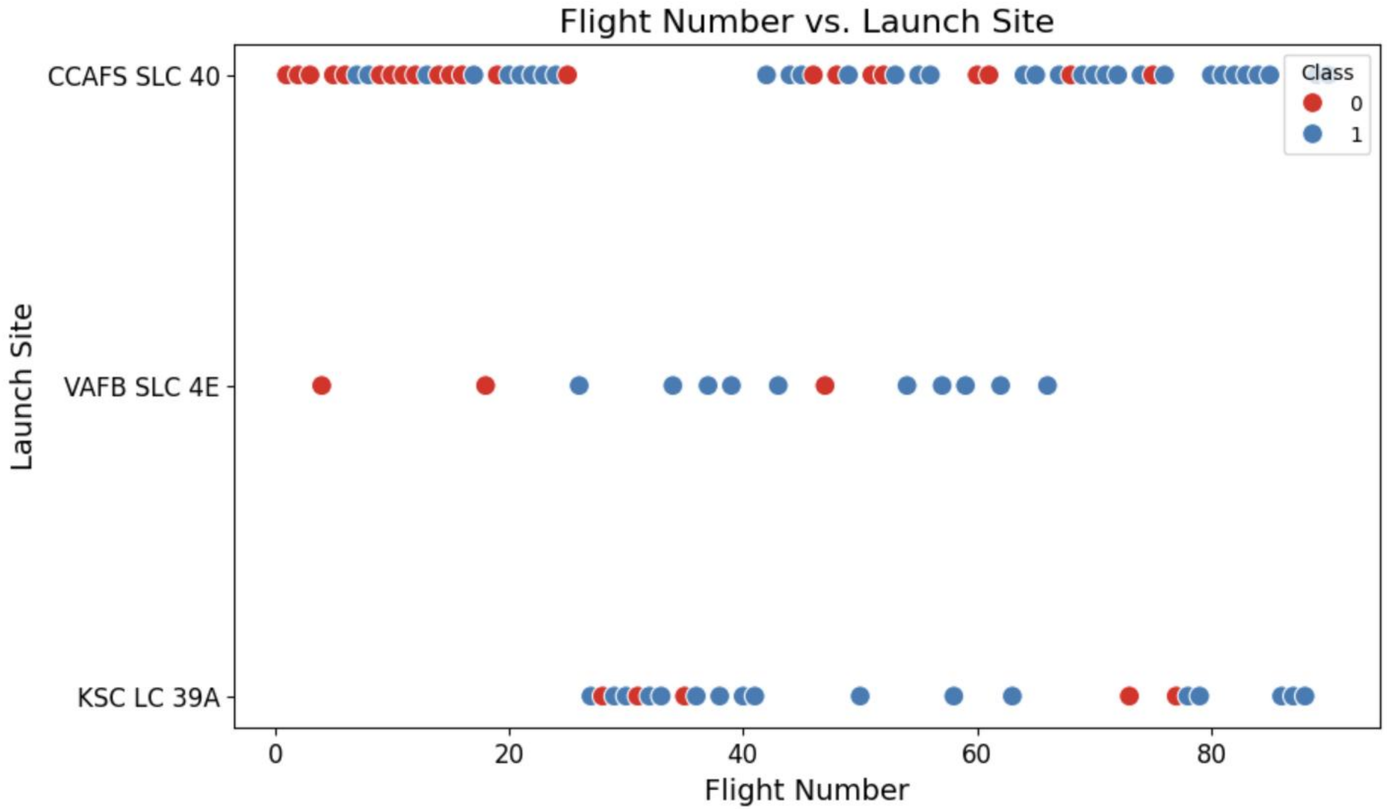


- We see that as the flight number increases, the first stage is more likely to land successfully.

Flight Number vs. Launch Site

Blue dot indicates successful and red indicates unsuccessfull landing

- Unsuccessful launches were more frequent in the early flight numbers, success rate has improved for more recent flights.
- VAFB SLC 4E is least used lauch site

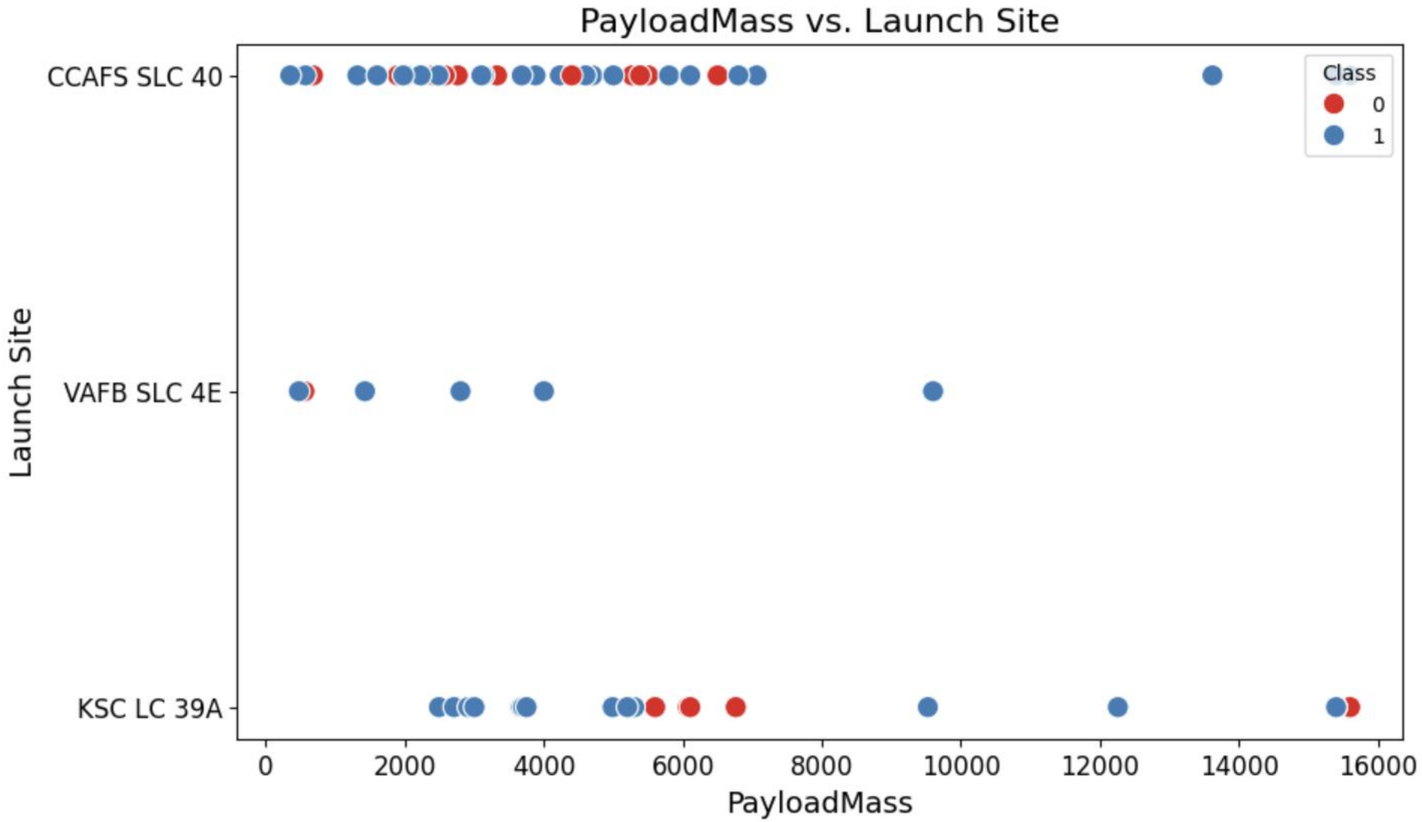


Payload vs. Launch Site

Blue dot indicates successful and red indicates unsuccessfull landing

- In VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000)

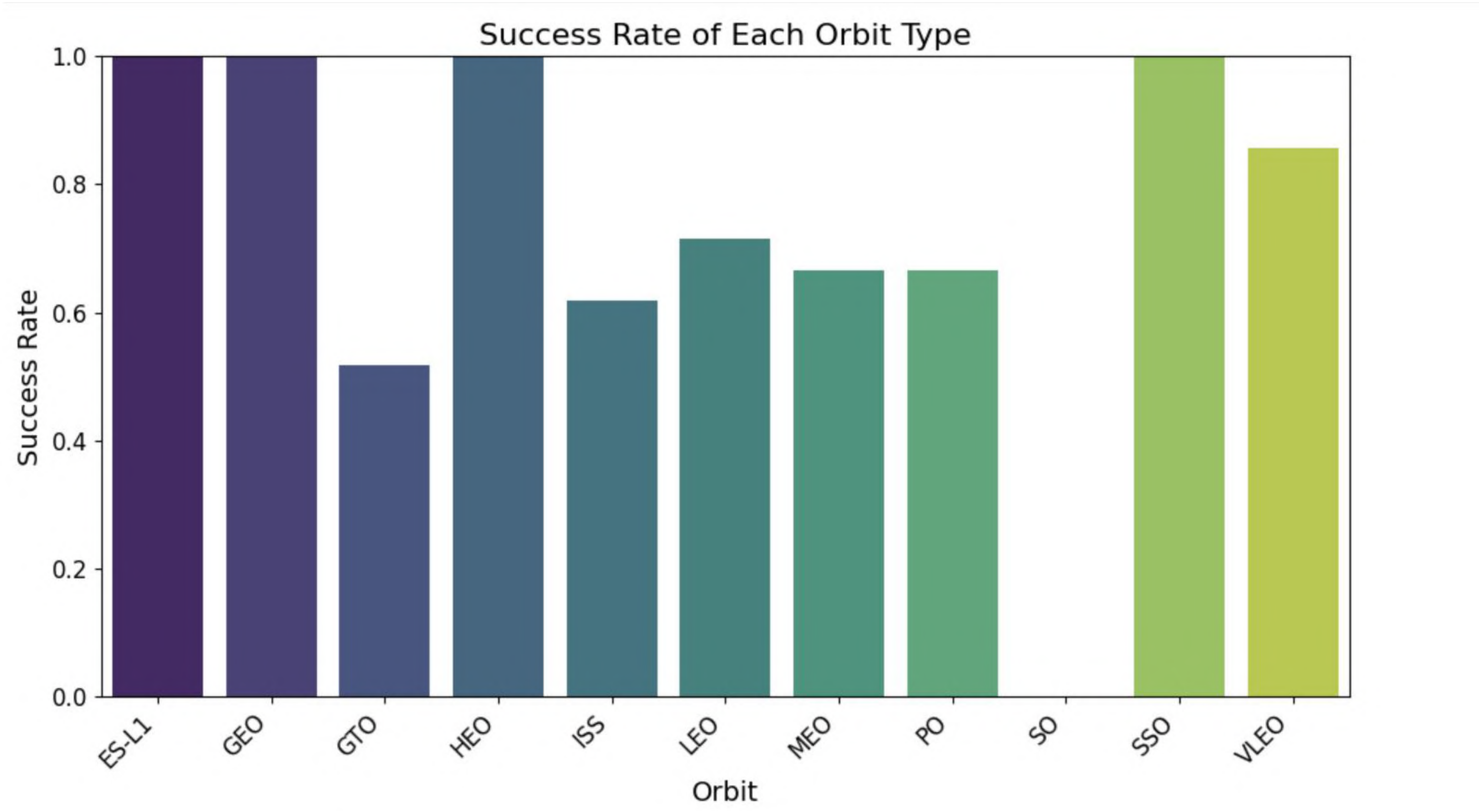
- Unsuccessful launches are more frequent in flights with mid-lower pay load mass.



Success rate vs Orbit type

Different color codes used for different orbits

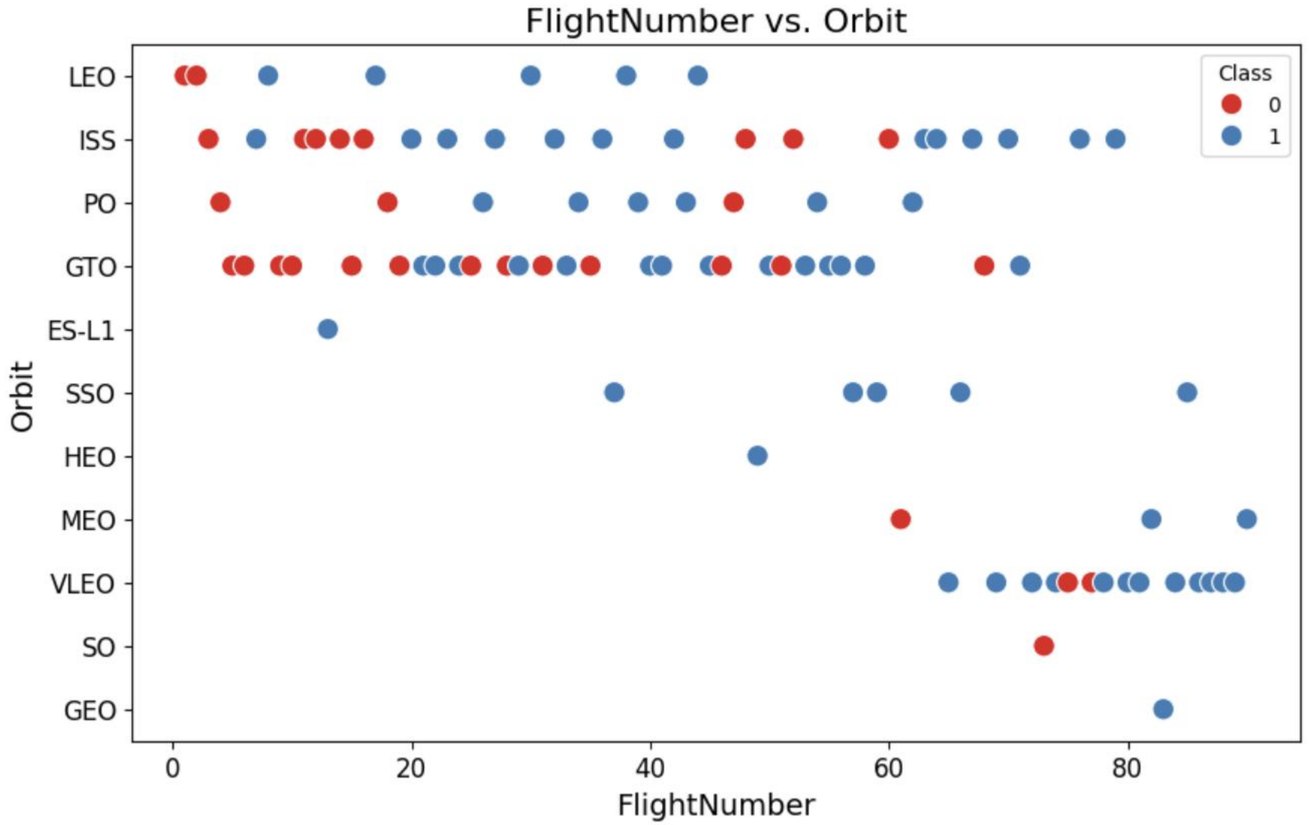
- ES-L1, GEO, HEO, SSO orbits have 100% successful launch rate.
- SO orbits have 0% successful launch rate



Flight Number vs. Orbit Type

Blue dot indicates successful and red indicates unsuccessfull landing

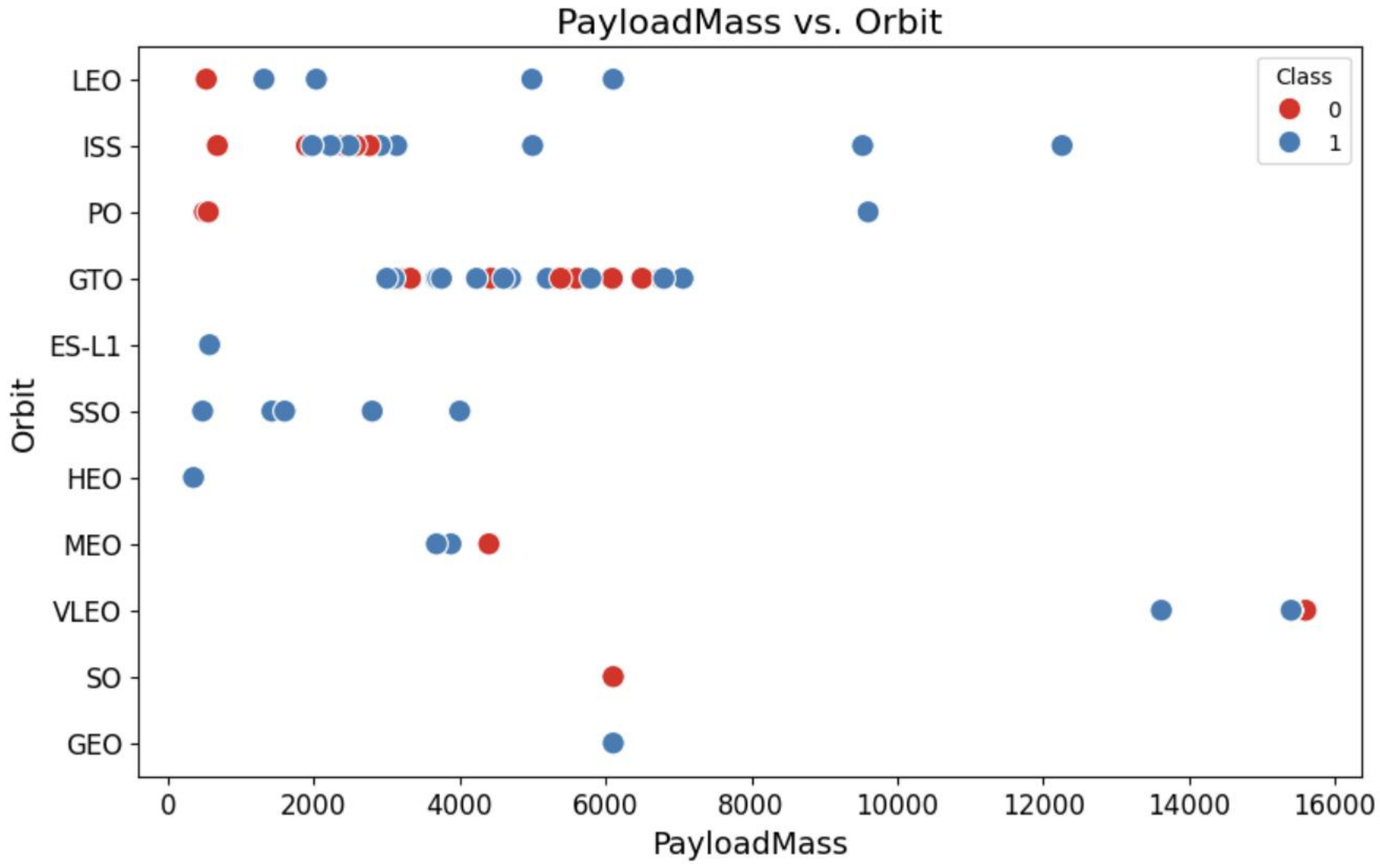
- SpaceX started with LEO orbits which saw moderate success LEO and returned to VLEO in recent launches
- SpaceX appears to perform better in lower orbits or Sun-synchronous orbits



Payload vs. Orbit Type

Blue dot indicates successful and red indicates unsuccessfull landing

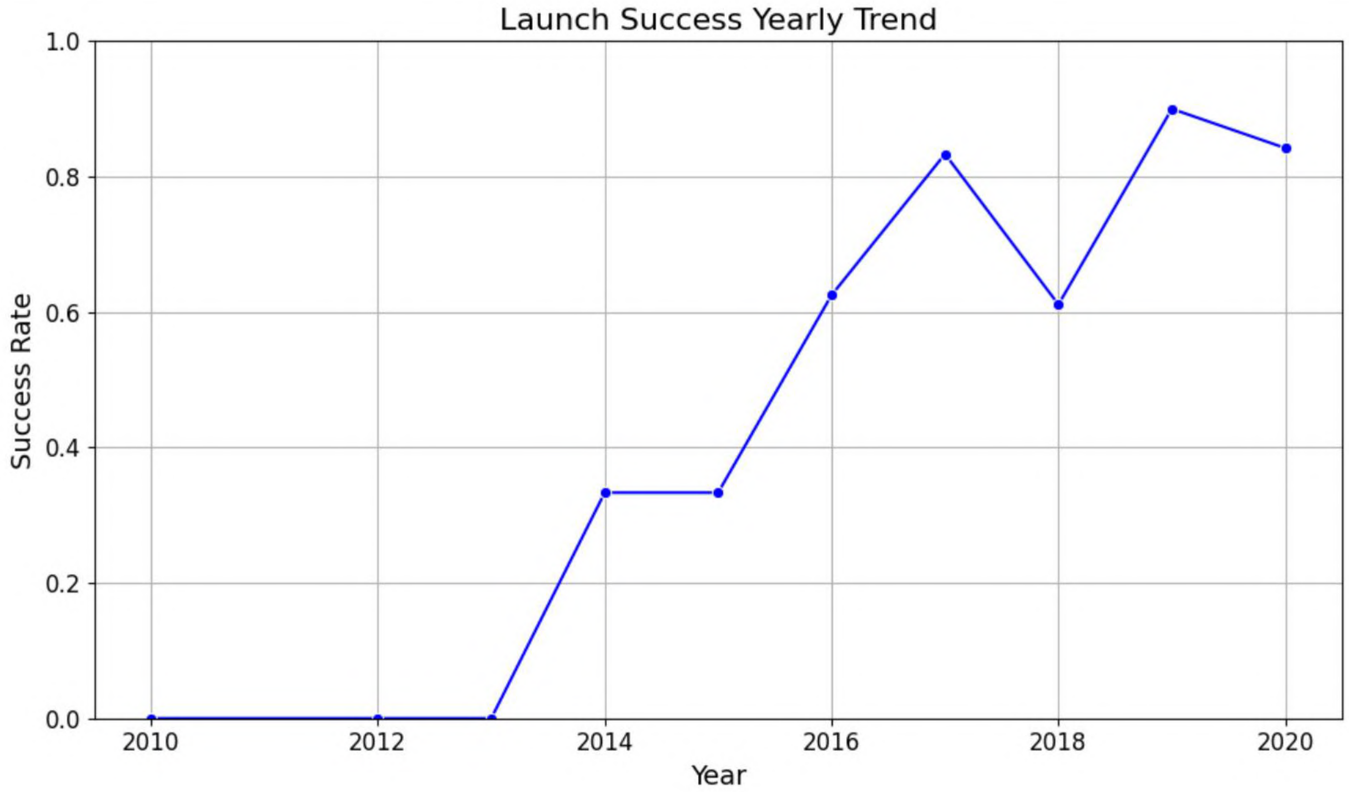
- GEO and SO orbits only have payloads that are more than 6000
- LEO and SSO seem to have relatively low payload mass
- The other most successful orbit VLEO only has payload mass values in the higher end of the range



Launch Success Yearly Trend

Blue dot indicates successful and red indicates unsuccessfull landing

- Success generally increases over time since 2013 with a slight dip in 2018
- Success in recent years at around 80%



EDA from SQL

All Launch Site Names

The sql query and answers is shown in the picture

- CCAFS SLC-40 and CCAFSSLC-40 likely all represent the same launch site with data entry errors.
- Likely only 3 unique launch_site values: CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E

```
# Task 1: Display the names of the unique launch sites in the space mission
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
```

```
 * sqlite:///my_data1.db
Done.
```

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

Launch Site Names Begin with 'CCA'

The sql query and answers is shown in the picture

```
# Task: Display 5 records where launch sites begin with the string 'CCA'
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

```
 * sqlite:///my_data1.db
Done.
```

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

- First five entries in database with Launch Site name beginning with CCA.

Total Payload Mass

The sql query and answers is shown in the picture

```
# Task 3: Display the total payload mass carried by boosters launched by NASA (CRS)
%sql SELECT SUM(PAYLOAD_MASS__KG_) as total_payload FROM SPACEXTBL WHERE Customer = 'NASA (CRS)'
```

```
 * sqlite:///my_data1.db
Done.
```

total_payload
45596

- This query sums the total payload mass in kg where NASA was the customer.
- CRS stands for Commercial Resupply Services which indicates that these payloads were sent to the International Space Station (ISS).

Average Payload Mass by F9 v1.1

The sql query and answers is shown in the picture

```
# Task 4: Display average payload mass carried by booster version F9 v1.1
%sql select avg(payload_mass__kg_) as Average from SPACEXTABLE where booster_version like 'F9 v1.1%'
```

```
 * sqlite:///my_data1.db
Done.
```

Average
2534.6666666666665

- This query calculates the average payload mass or launches which used booster version F9 v1.1

First Successful Ground Landing Date

The sql query and answers is shown in the picture

```
# Task 5: List the date when the first successful landing outcome on a ground pad was achieved
%sql SELECT MIN(Date) AS first_successful_landing_date FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (ground pad)';
```

```
 * sqlite:///my_data1.db
Done.
```

first_successful_landing_date
2015-12-22

- First ground pad landing wasn't until the end of 2015.
- Successful landings in general appear starting 2014.

Successful Drone Ship Landing with Payload between 4000 and 6000

The sql query and answers is shown in the picture

```
# Task: List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
%sql SELECT "Booster_Version" FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000;
```

```
 * sqlite:///my_data1.db
Done.
```

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

- This query returns the four booster versions that had successful drone ship landings and a payload mass between 4000 and 6000 noninclusively.

Total Number of Successful and Failure Mission Outcomes

The sql query and answers is shown in the picture

```
# Task 7: List the total number of successful and failed mission outcomes
# Task 7: List the total number of successful and failed mission outcomes
%sql SELECT SUM(CASE WHEN "Mission_Outcome" LIKE 'Success%' THEN 1 ELSE 0 END) AS successful_missions, SUM(CASE WHEN "Mission_Outcome" LIKE 'Failure%' THEN 1 ELSE 0 E
```

```
 * sqlite:///my_data1.db
Done.
```

successful_missions	failed_missions
100	1

- SpaceX appears to achieve its mission outcome nearly 99% of the time.
- This means that most of the landing failures are intended.

Boosters Carried Maximum Payload

The sql query and answers is shown in the picture

- These booster versions are very similar and all are of the F9 B5 B10xx.x variety.

- • This likely indicates payload mass correlates with the booster version that is used.

```
# Task 8: List the names of the booster_versions which have carried the maximum payload mass
%sql SELECT "Booster_Version" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL);
```

```
 * sqlite:///my_data1.db
Done.
```

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

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

The sql query and answers is shown in the picture

```
# Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20 in descending order
# Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20 in descending order
%sql SELECT "Landing_Outcome", COUNT(*) AS outcome_count FROM SPACEXTBL WHERE substr(Date, 1, 10) BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing_Outcome" ORD
```

```
 * sqlite:///my_data1.db
Done.
```

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

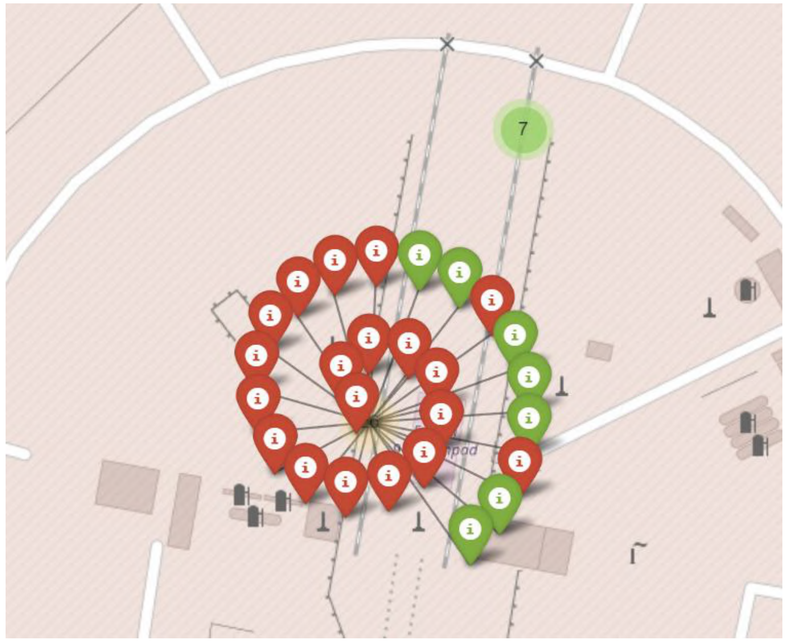
- Total 8 success, 5 in drone ship and 3 in ground pad

Interactive Map with Folium

Folium Map: Launch site locations

- We can see that all launch sites are located in North America and that all launch sites are located near to coastlines, specifically the coasts of Florida and California.

Folium Map: Color-labeled launch outcomes

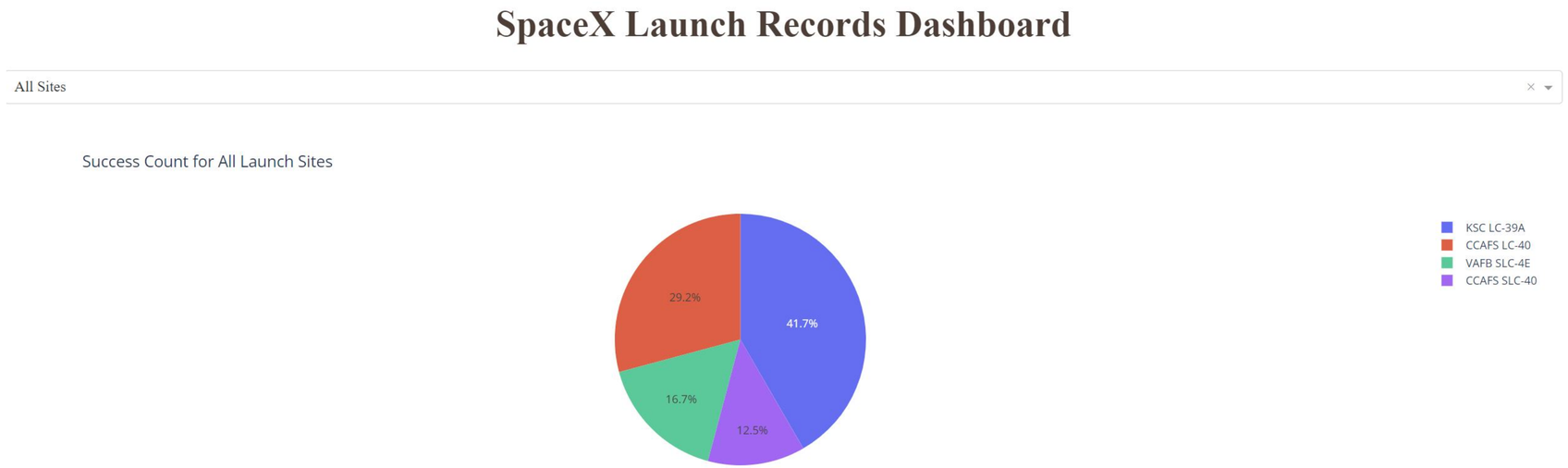


- The right screenshot tells us that 26 launch records are clustered at this launch location (CCAFS LC-40). We can drill down by clicking on the cluster, expanding the image like shown in the right screenshot. This tells us that there were 7 successful landings (green) and 19 unsuccessful landings (red).

Build a dashboard with plotydash

Plotly Dash: Successful Stage 1 Landings By Launch Site

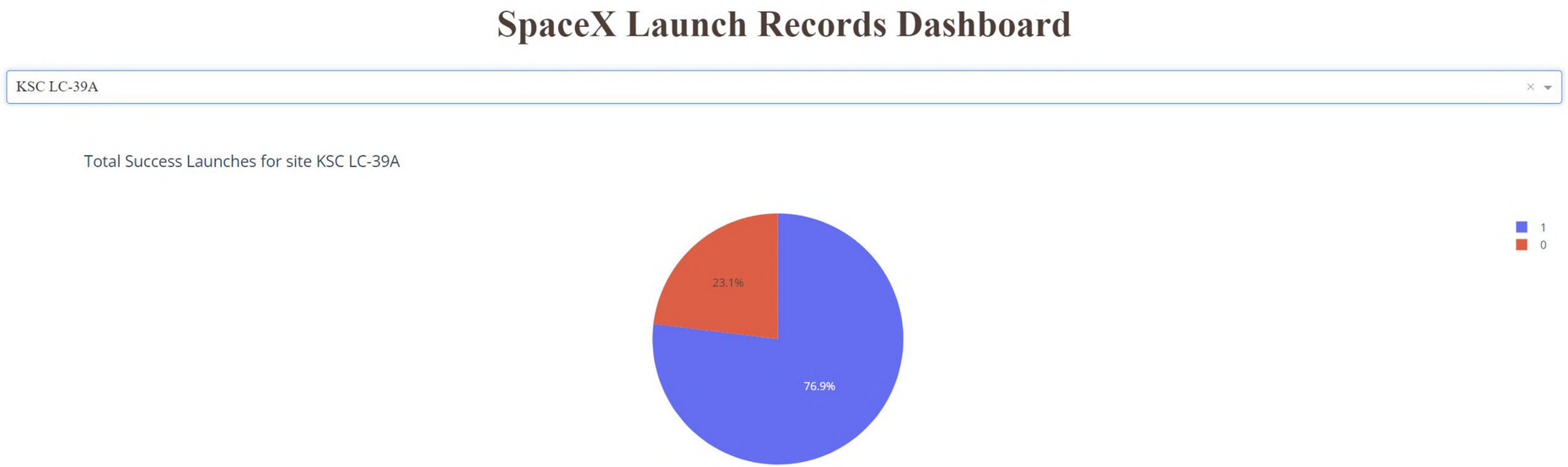
You are viewing live dashboard app.



- We can see that most successful landings were launches from KSC LC-39A. The least successful landings were launches from CCAFS SLC-40.

Plotly Dash: Successful Stage 1 Landings for KSC LC-39A

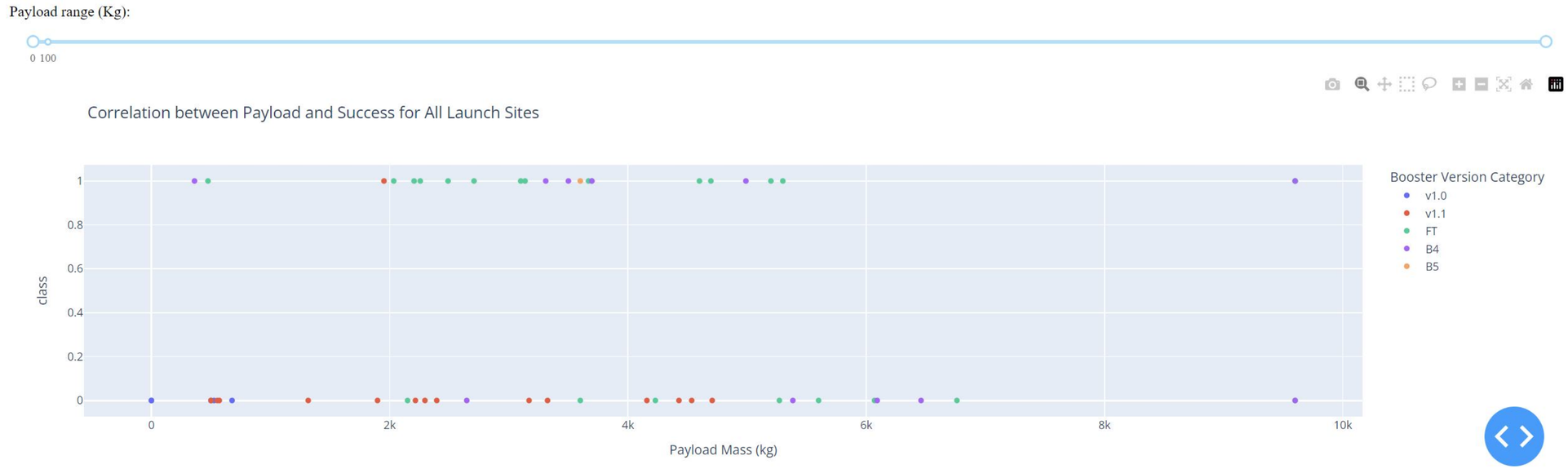
You are viewing live dashboard app.



- Upon examining the most successful launch site, KSC LC-39A, we observe the distribution of successes versus failures. Despite contributing to many successful launches, this site actually has a relatively low success rate.

Plotly Dash: Payload Mass vs. Success vs. Booster Version Category

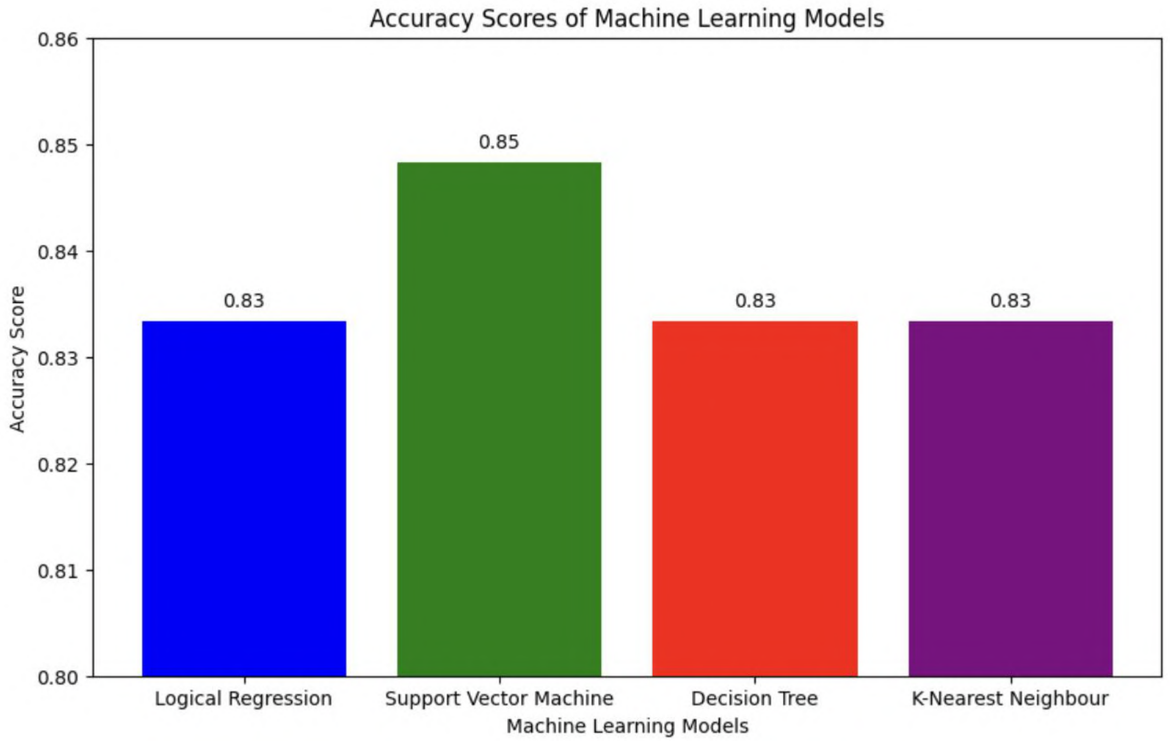
You are viewing live dashboard app.



- We can see that booster version category FT has many successes and few failures. In contrast, v1.1 has many failures and few successes.

Predictive Analysis (Classification)

Classification accuracy

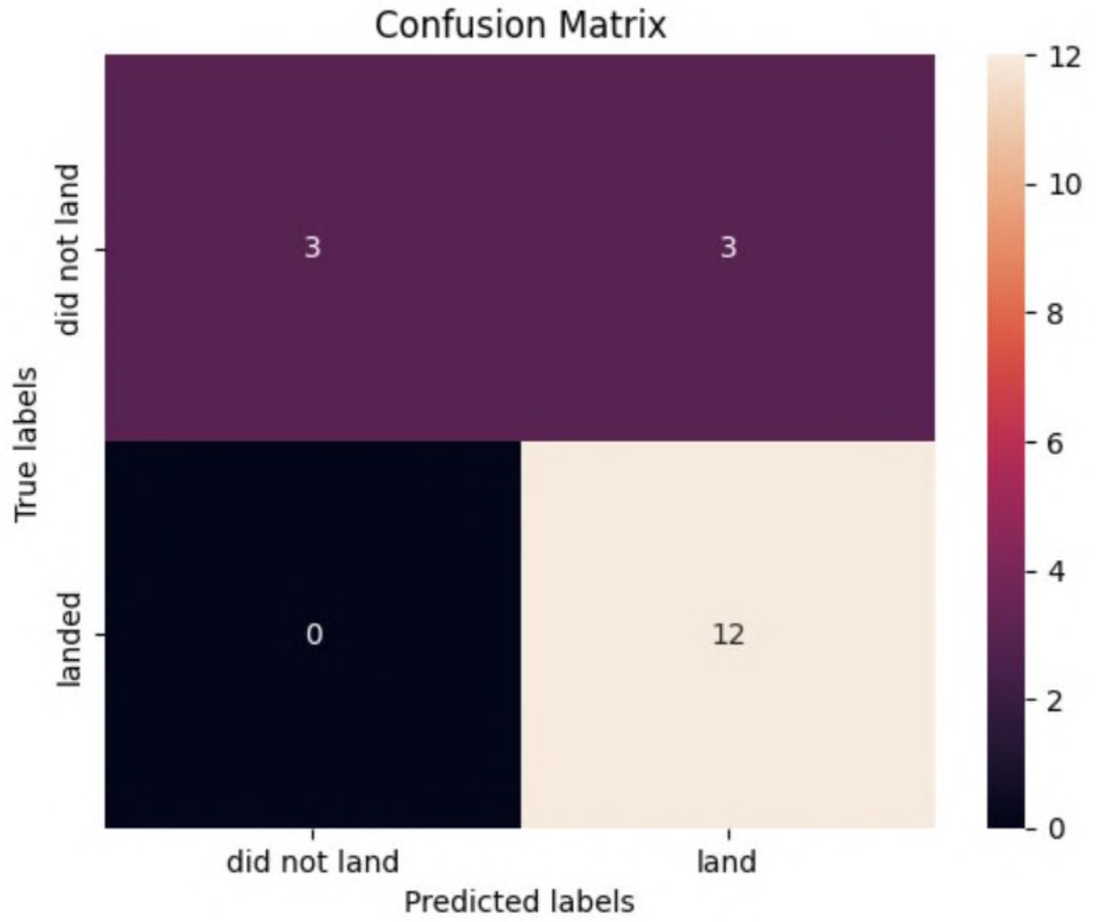


- All models had virtually the same accuracy on the test set at 83.33% accuracy. except svm that shown a slight change in accuracy

Confusion Matrix

- The models predicted 3 unsuccessful landings when the true label was unsuccessful landing.

- The models predicted 3 successful landings when the true label was unsuccessful landings (false positives). Our models over predict successful landings.



Conclusion

- Thesuccess rates for SpaceXlaunches is directly proportional time in years they will eventually perfect the launches

- Launch sites are typically located close to coastlines.

- ES-L1, GEO, HEO, SSO orbits have the best success rate.

- We created a machine learning model with an accuracy of 83%. Allon Mask of SpaceY can use this model to predict, with relatively high accuracy, whether a Stage 1 landing will be successful before the launch. This prediction can help determine whether or not to proceed with the launch.

- ◦Created a dashboard for visualization

- SpaceX appears to achieve its mission outcome nearly 99% of the time. This means that most of the landing failures are intended.

Appendix

01 Github URL

02 Special thanks to IBM instructors for delivering the IBM Data Science Professional Certificate courses!

Thanks

Thank you everyone who been with me with this journey.
Check the github repository for more info about the project