IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rishika Vishvakarma
02-08-2023

Outline

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

Executive Summary

- The methodology consist of web scraping the data, cleaning it, wrangling it, make an Explaratory Data Analisis calculating the success rate of each launch site, each orbit, each landing site among others, visualizing relationships among variables, and finally make a predictive analysis splitting the data and creating a model to determine the success rate with given variables.
- The Main results consist of a created model that returns a correct outcome with an accuracy of 83%.

Introduction

- Falcon 9 is a rocket from the SpaceX, This company has achieved a lot, the fist private company to throw a rocket into orbit, recover the ship after the travel, it has been a lot of failed attemps of landing, recently the company has started achieving the landing and making history.
- We want to predict if the rocket will land successfully or not, depending on many factors.

Section 1

Methodology

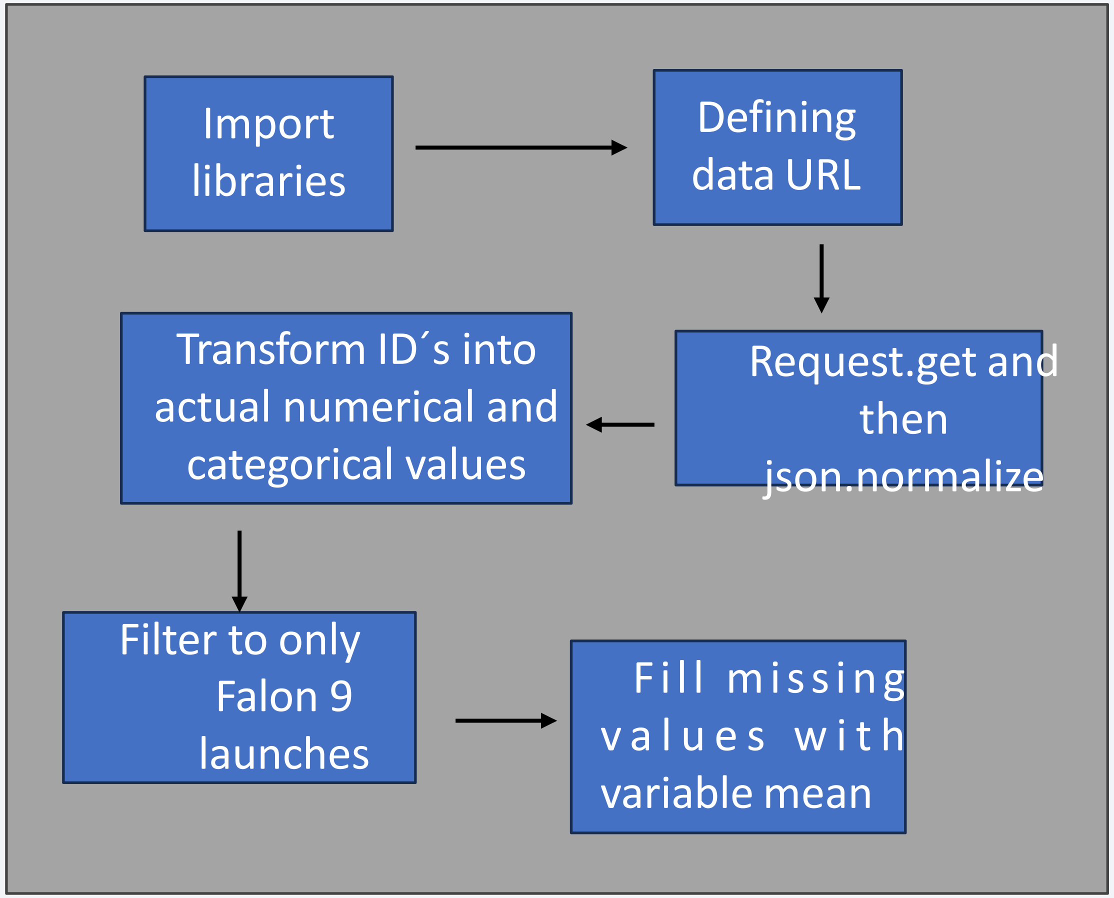
Methodology

Executive Summary

- Data collection methodology:
  - We have Scraped the data from from the official SpaceX website, filtered by Falcon 9 launches
- Perform data wrangling
  - We replaced the missing values with the variable mean, created a column with a binary variable depending of the success.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Create a column for the class, standardize the data, split the data and find best hyuperparameters.

Data Collection and Scraping - SpaceX API

- Data collection API (LINK)



Data Wrangling

- Data Wrangling.ipynb

Calculate number of launches of each site → Calculate the number and ocurrence of each orbit → Calculate the number and ocurrence of mission outcomes of the orbits → Create a landing outcome laberel from outcome column

EDA with Data Visualization

- The chart that we plotted were:

1. Payload mass vs Flight number
2. Flight number vs launch site
3. Payload Mass vs Launch Site
4. Success rate per orbit
5. Flightnumber and orbit type (sort by class)
6. Payload and orbit type (sort by class)
7. Finally the launch success yearly trend

In every one of each graph we printed the success dots with one color and the failed with other, this way we can identify some patterns in the variables talked before.

EDA PYTHON.ipynb

EDA with SQL

- With SQL Queries, we performed:

[EDA SQL.ipynb]

1. Display unique launch sites
2. Display 5 records where launch sites begin with the string 'CCA'
3. Display the total payload mass carried by boosters launched by NASA (CRS)
4. Display average payload mass carried by booster version F9 v1.1
5. List the date when the first succesful landing outcome in ground pad was achieved
6. List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
7. List the total number of successful and failure mission outcomes
8. List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
9. List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
10. 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.

Build an Interactive Map with Folium

- What we did with folium was:
- Marked all launch sites on the map
- Marked the success/failed launches for each site on the map
- Clustered the launchsite
- Calculate the distances between a launch site to its proximities
- Drew a `PolyLine` between a launch site to the selected coastline point

We added those markers in the map to identify and analyse the success rate of each launch site on the map

[Interactive visual analytics and dashboard.ipynb]

Dashboard with Plotly Dash

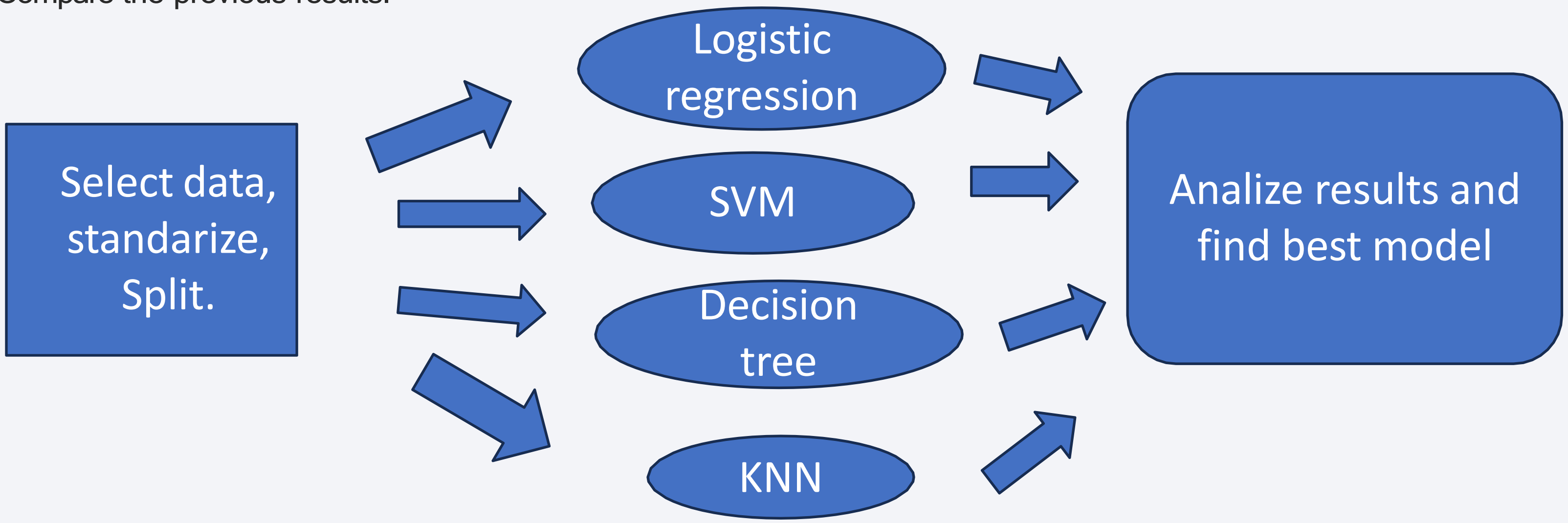
- Dropdown List with Launch Sites
- Allow user to select all launch sites or a certain launch site
- Pie Chart Showing Successful Launches
- Allow user to see successful and unsuccessful launches as a percent of the tota
- Slider of Payload Mass Range
- Allow user to select payload mass range
- Scatter Chart Showing Payload Mass vs. Success Rate by Booster Version
- Allow user to See the correlation between Payload and Launch Success

Predictive Analysis (Classification)

- Fist we select de data to use, standardize it, split the data into train/test.

1. create a logistic regression, create a gridsearchCV and find the best parameters, we calculate the accuracy, analize the confusion matrix.
2. Create a support vector machine object, then create a gridsearch object, fit the object and find best parameters, calculate de accuracy.
3. Create a decision tree classifier, then a gridsearchCV, fit the object and find best parameters, calculate accuracy.
4. Create k nearest neightbor object then a gridsearchCV, fit the object and find best parameters, calculate accuracy
5. Compare the previous results.

Predictive analysis.ipynb



Results

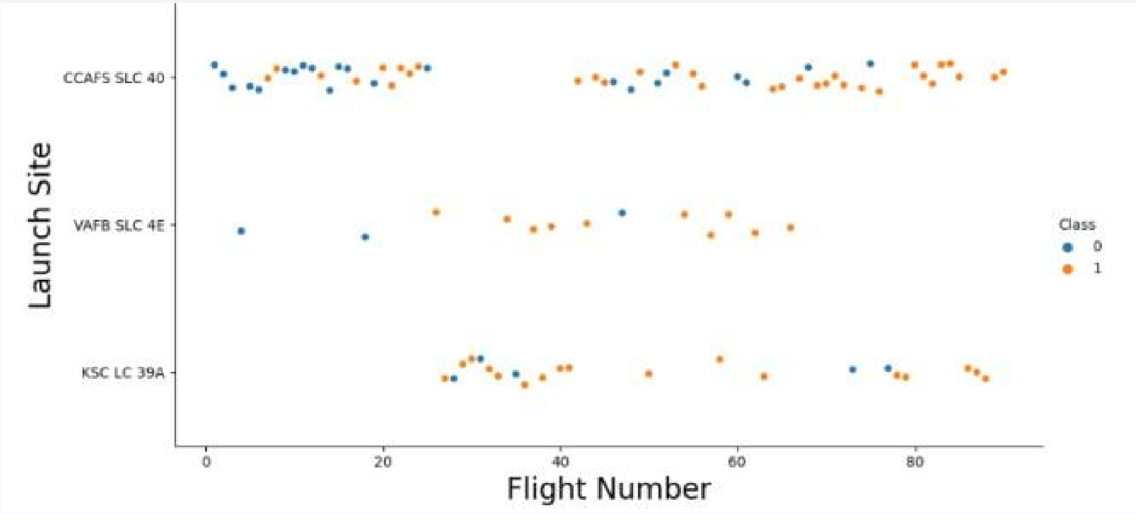
- Exploratory Data Analysis
- • Launch success has improved over time
- • KSC LC-39A has the highest success rate among landing sites
- • Orbits ES-L1, GEO, HEO and SSO have a 100% success rate
- Visual Analytics
- • Most launch sites are near the equator, and all are close to the coast
- • Launch sites are far enough away from anything a failed launch can damage
- (city, highway, railway), while Still close enough to bring people and material
- ta support launch activities
- Predictive Analytics
- • Decision Tree model is the best predictive model for the dataset

Section 2

Insights drawn from EDA

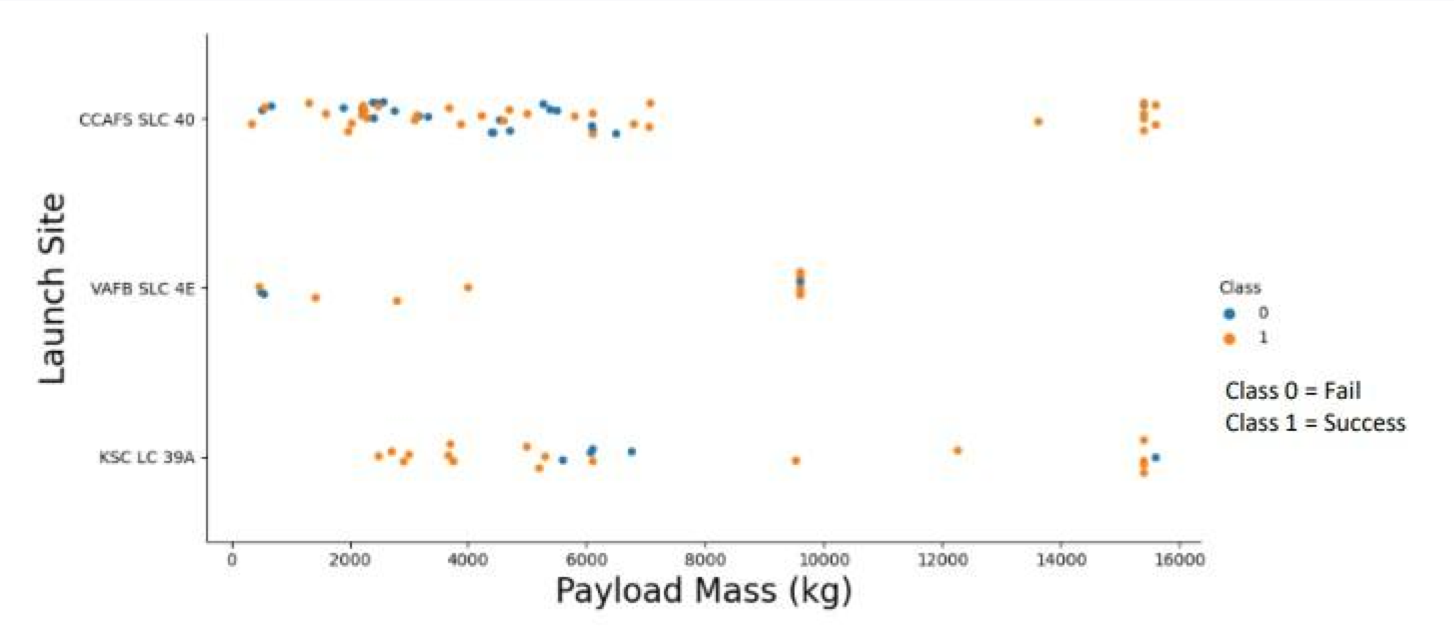
Flight Number vs. Launch Site

- Exploratory Data Analysis
- Earlier flights had a lower success rate (blue ones)
- Later flights had a higher success rate (orange ones)
- VAFB SLC 4E and KSC LC 39A have higher success rates



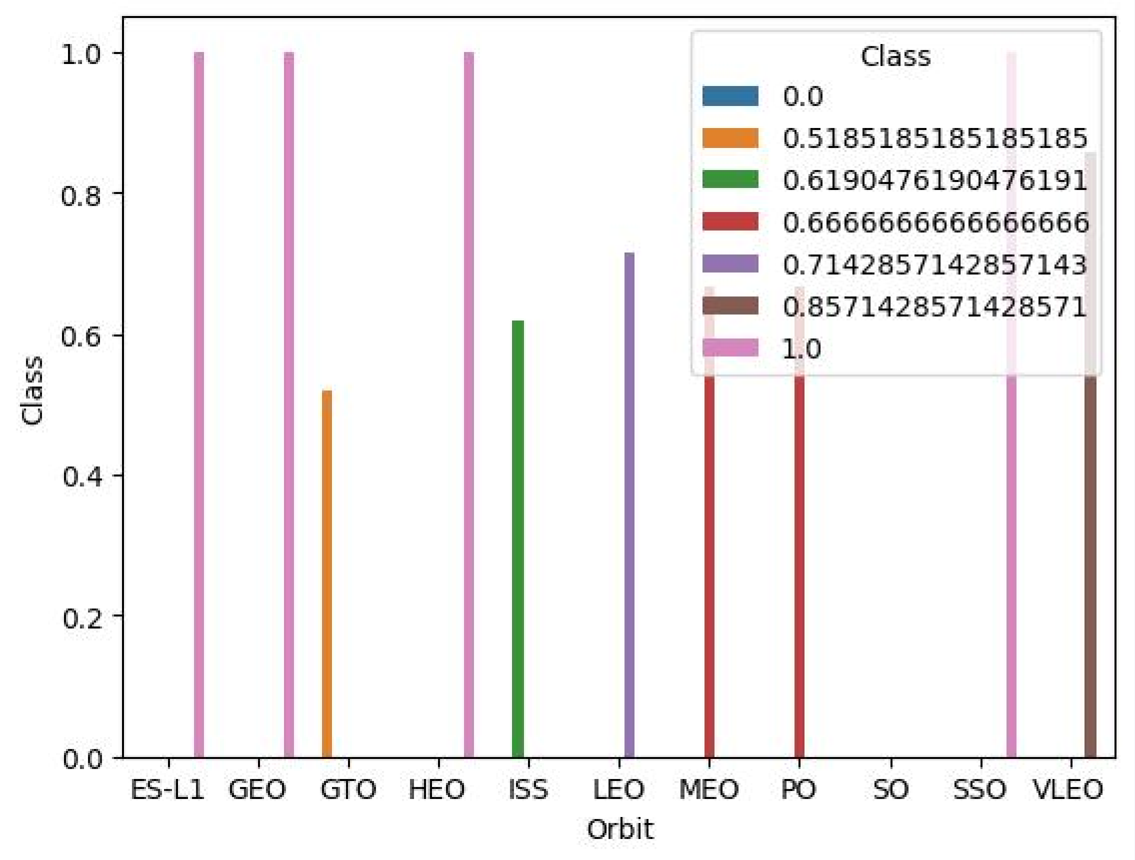
Payload vs. Launch Site

- The higher the payload mass (kg), the higher the success rate
- Most launces with a payload greaterthan 7,000 kg were successful
- VAFB SKC 4E has not launched anything greaterthan ~10,000 kg



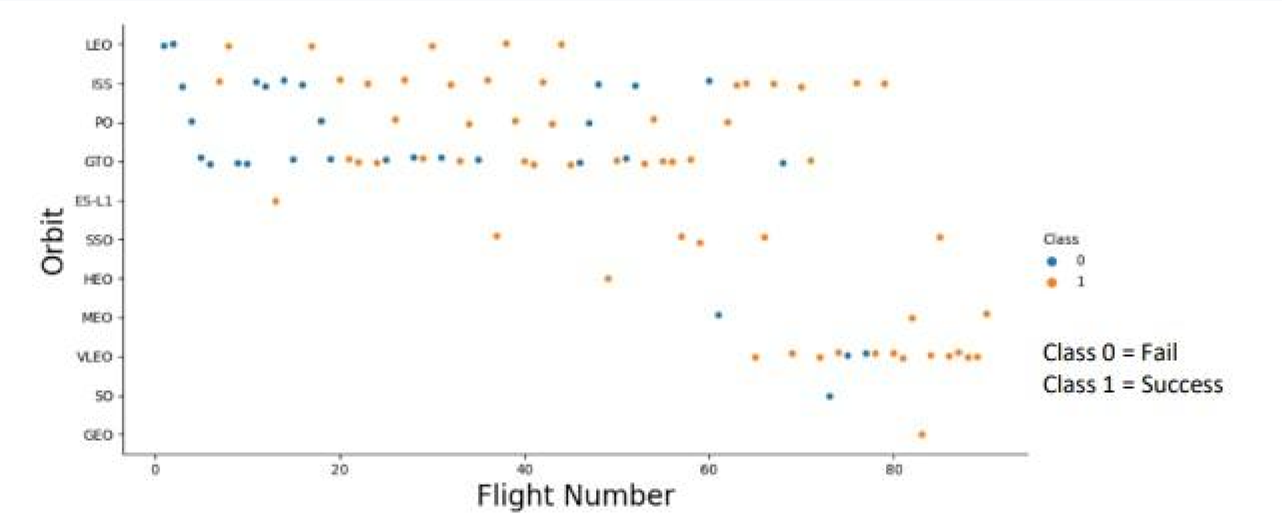
Success Rate vs. Orbit Type

- 100% Success Rate: ES-LI, GEO, HEO and SSO
- Good Success Rate: GTO, ISS, LEO, MEO, PO
- 0% Success Rate: SO



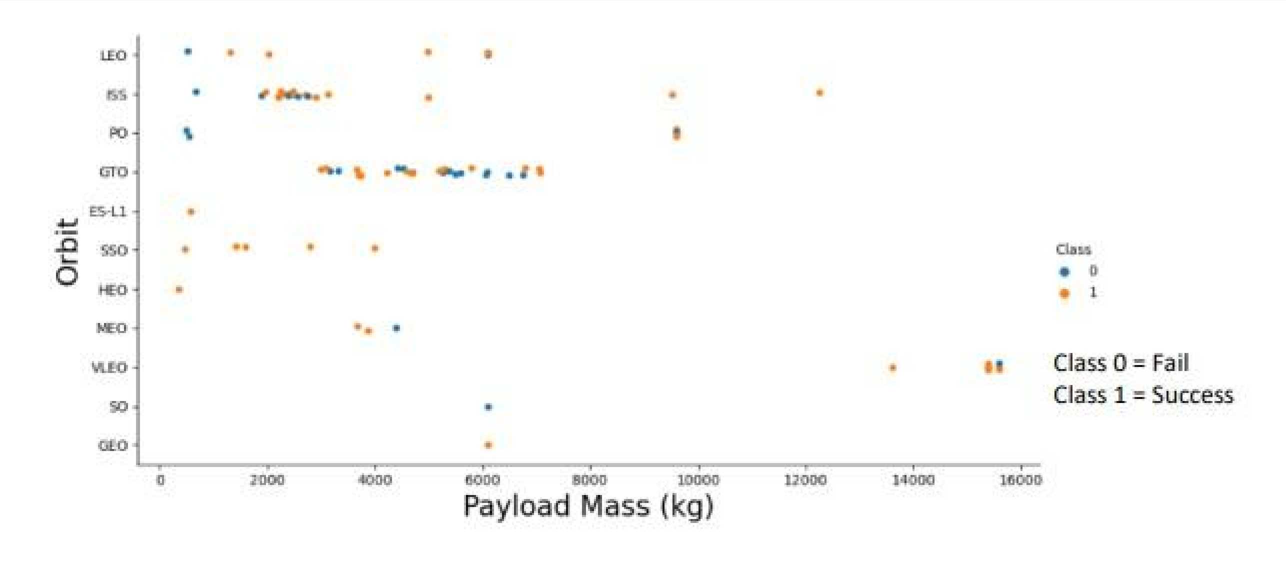
Flight Number vs. Orbit Type

- The success rate typically increases with the number of flights for each orbit
- This relationship is highly apparent for the LEO orbit



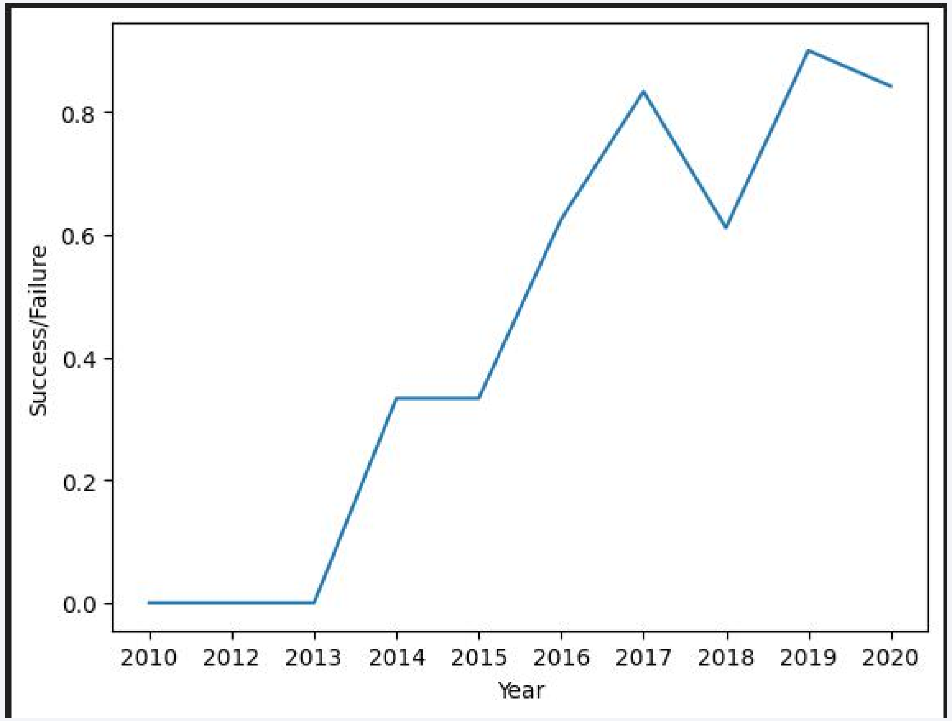
Payload vs. Orbit Type

- Heavy payloads are better with LEO, ISS and PO orbits
- The GTO orbit has mixed success with heavier payloads



Launch Success Yearly Trend

- The success rate improved from 2013 and 2018
- The success rate decreased in 2017 and 2019
- Overall, the success rate has improved since 2013



All Launch Site Names

Display the names of the unique launch sites in the space mission

```
%sql select distinct(launch_site) from SPACEXTBL;
```

```
 * 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'

Display 5 records where launch sites begin with the string 'CCA'

+ Code + Markdown

```
%sql select * from SPACEXTBL where launch_site like 'CCA%' limit 5;
```

Python

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

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	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

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

```
%sql select sum(payload_mass__kg_) from SPACEXTBL where customer = 'NASA (CRS)';
```

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

sum(payload_mass__kg_)
45596

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 SPACEXTBL where booster_version like 'F9 v1.1%';
```

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

avg(payload_mass__kg_)
2534.6666666666665

First Successful Ground Landing Date

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
1  %sql select min(DATE) from SPACEXTBL where "Landing_Outcome" = 'Success (ground pad)';
```

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

min(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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 distinct(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

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%sql select  mission_outcome,count(*) as Counter from SPACEXTBL group by mission_outcome;
```

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

Mission_Outcome	Counter
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%sql select distinct(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

2015 Launch Records

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.

```
%sql select  landing_outcome,booster_version,launch_site,date from SPACEXTBL where Date like '2015%' and landing_outcome = 'Failure (drone ship)';

```

* sqlite:///my_data1.db
Done.

Landing_Outcome	Booster_Version	Launch_Site	Date
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015-10-01
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	2015-04-14

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.

```
%sql select landing_outcome,count(landing_outcome) as count from SPACEXTBL where DATE BETWEEN '2010-06-04' and '2017-03-20' GROUP BY landing_outcome ORDER BY count(landing_outcome
```

Python

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

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

Section 3

Launch Sites Proximities Analysis

Launch Sites

We can see the two launching sites of spaceX, one in each coast

Launch outcomes

Greens are success
Red are failed launches

Distance to proximities

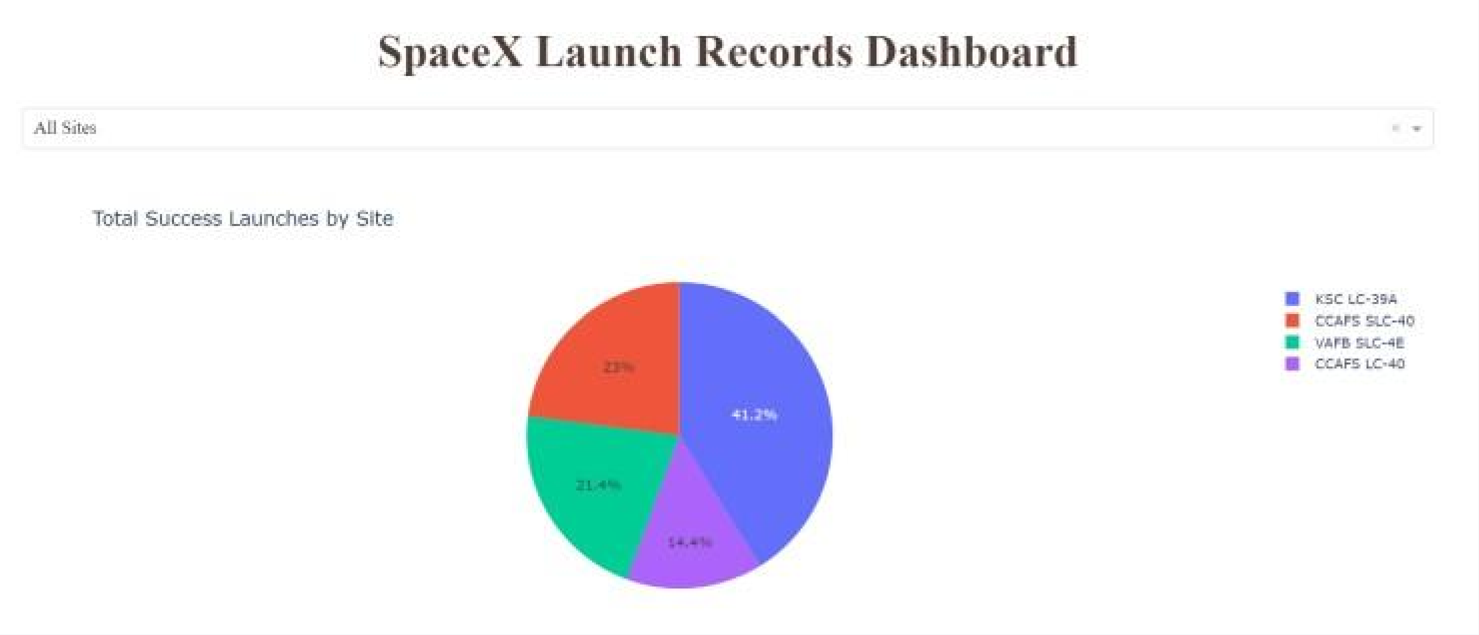
It needs to be a safe zone, so that nothing falls on people or private property, in an exclusión zone so nobady gets in.

Section 4

Build a Dashboard with Plotly Dash

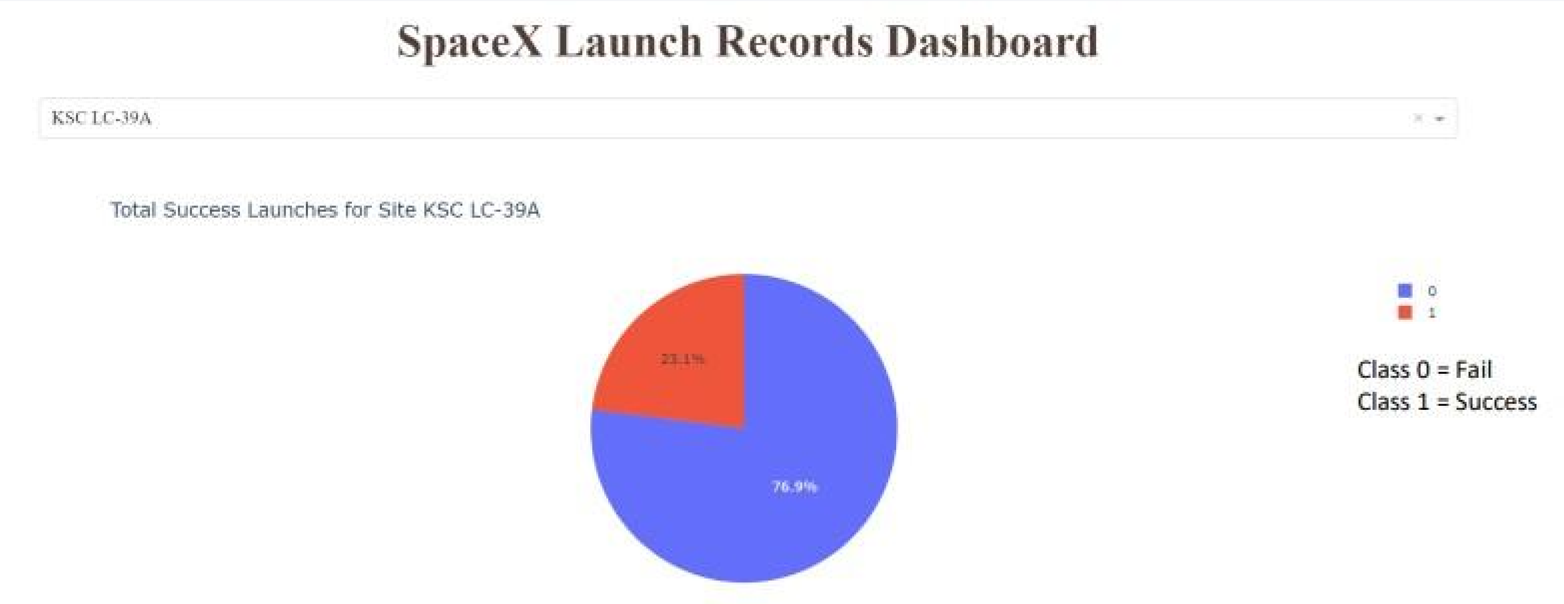
Launch success by site

- KSC LC-39A has the most successul launches



Launch Success (KSC LC-29A)

KSC LC-39A has the highest success rate amongst launch sites (76.9%)



Payload Mass and Success

- Payloads between 2,000 kg and 5,000 kg have the highest success rate

Section 5

Predictive Analysis (Classification)

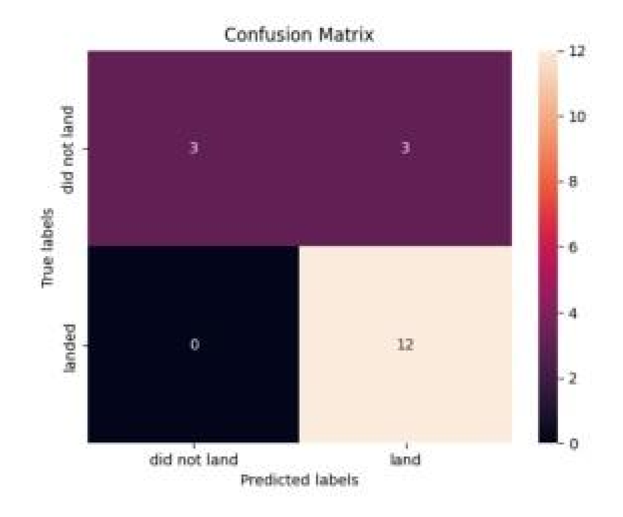
Classification Accuracy

- All the models performed at about the same level and had the same scores and accuracy. This is likely due to the small dataset. The Decision Tree model slightly outperformed the rest when looking at .best_score_

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

Confusion Matrix

- A confusion matrix summarizes the performance of a classification algorithm
- All the confusion matrices were identical



Conclusions

- Coast: All the launch sites are close to the coast.
- Orbits: ES-LI, GEO, HEO, and SSO have a 100% success rate.
- Payload Mass: Across all launch sites, the higher the payload mass (kg), the higher the success rate.
- Model Performance: The models performed similarly on the test set, with the decision tree model slightly outperforming others.
- Equator: Most of the launch sites are near the equator for an additional natural boost - due to the rotational speed of Earth - which helps save the cost of putting in extra fuel and boosters.
- Launch Success: Increases over time.
- KSC LC-39A has the highest success rate among launch sites and achieves a 100% success rate for launches less than 5,500 kg (payload).

Thank you!