SpaceX Success Evolution

Lawrence A
March 21, 2024

CONTENT

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

EXECUTIVE SUMMARY

Methodologies Summary

- Data collection
- Data Cleaning
- Data Visualization
- SQL Exploratory Data Analysis
- Building an interactive map with Folium
- Plotly Dash Dashboard Building
- Machine Learning Methods

EXECUTIVE SUMMARY

Results Summary

- Exploratory Data Analysis
- Interactive analytics
- Predictive analysis

INTRODUCTION

Founded by Elon Musk in 2002, SpaceX is a leading private aerospace company known for its innovative approach to space flight. SpaceX has revolutionized the space industry with its Falcon rocket series of rockets and pioneering work in rocket booster reusability. Their notable achievements include, but not limited to, successful cargo and crew missions to the ISS. SpaceX aiming to enable crewed missions to Mars and other planets.

INTRODUCTION

Questions to be answered:

- What variables affect the success of booster landing?
- How has the success rate changed over the years?
- How best can we binarily classify them?

METHODOLOGY

Data acquisition methodology:

- Web Scraping from Wikipedia and API requests from SpaceX

Data Cleaning:

- Filtering data to remove or replace missing values

SQL and visualization exploratory data analysis (EDA)

Folium and Plotly Dash interactive visual analytics

Classification models predictive analysis

classification models Building, tuning and evaluating

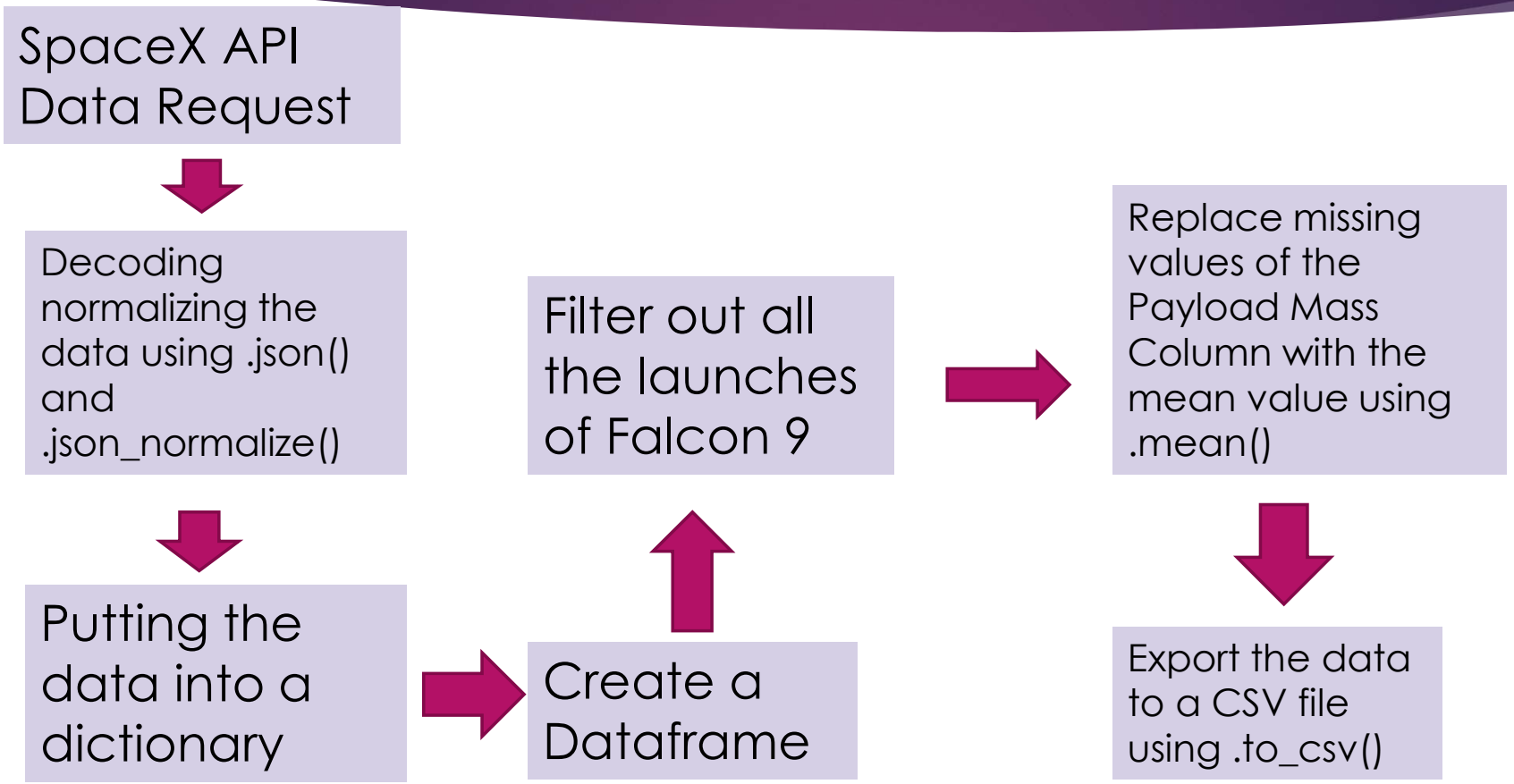
DATA COLLECTION

API requests from SpaceX REST API and Web Scraping were used to collect data from a table in SpaceX's Wikipedia entry.

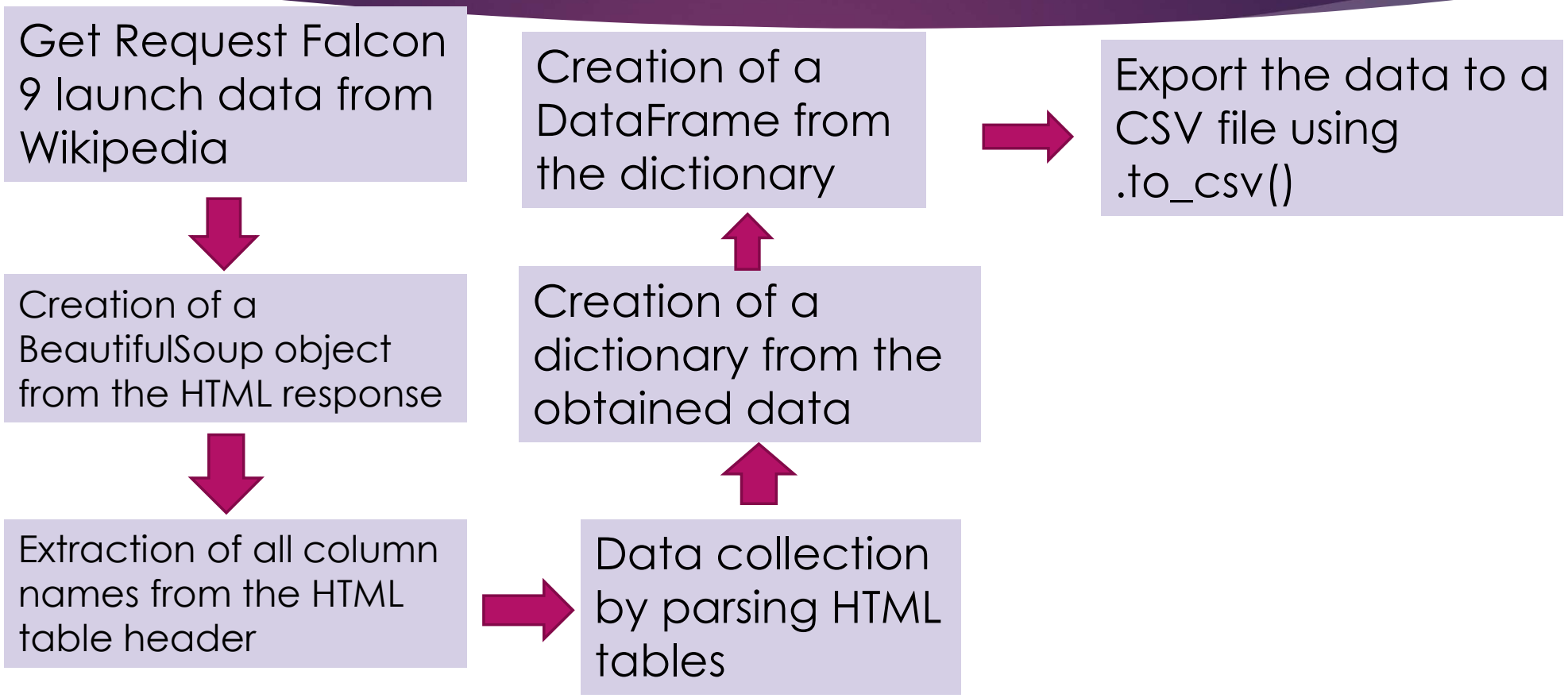
Columns obtained by using SpaceX REST API are: FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude

Columns obtained by using Wikipedia Web Scraping are: Flight Nunber, Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time

SpaceX API Data Collection Sequence



Web Scraping Data Collection Sequence



Data Wrangling

The dataset shows different cases for the booster landings.

True Ocean = mission was successful, booster landed in the ocean.

False Ocean = booster unsuccessful landing in the ocean.

True RTLS = mission was successfully, booster landed on a ground pad

False RTLS = mission was unsuccessfully, booster didn't land on ground pad.

True ASDS = mission was successful, booster landed on a drone ship

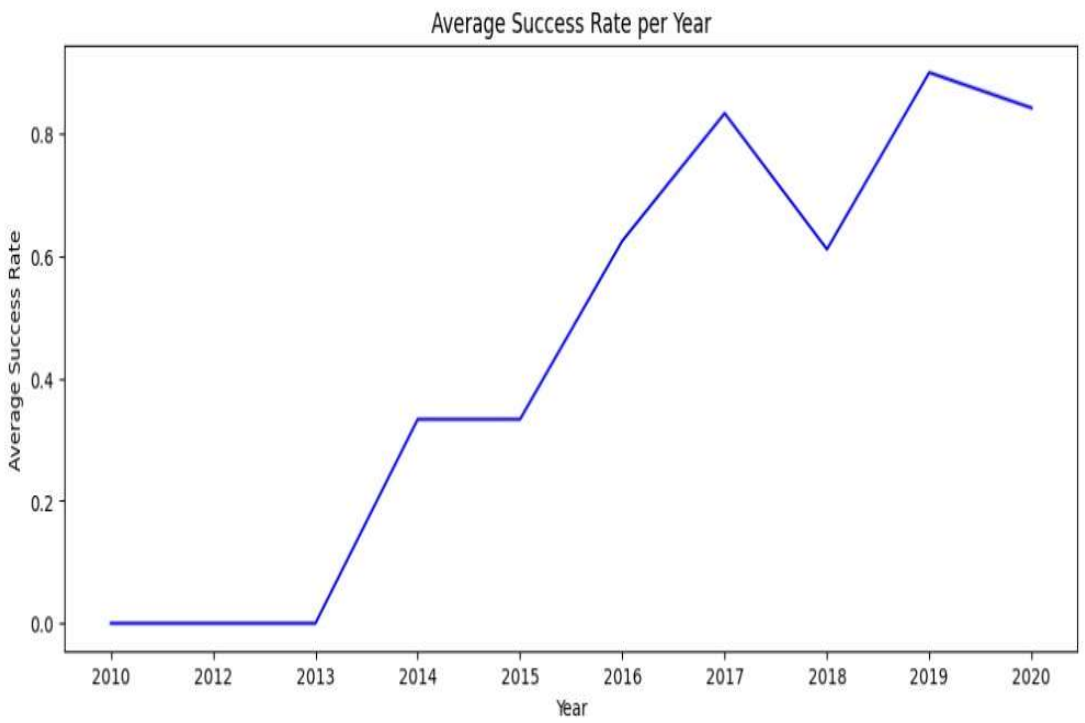
False ASDS = mission was unsuccessful, booster didn't land on a drone ship.
During the processing of the data, the landing outcomes where changed into boolean(1 for successful landing and 0 for unsuccessful).

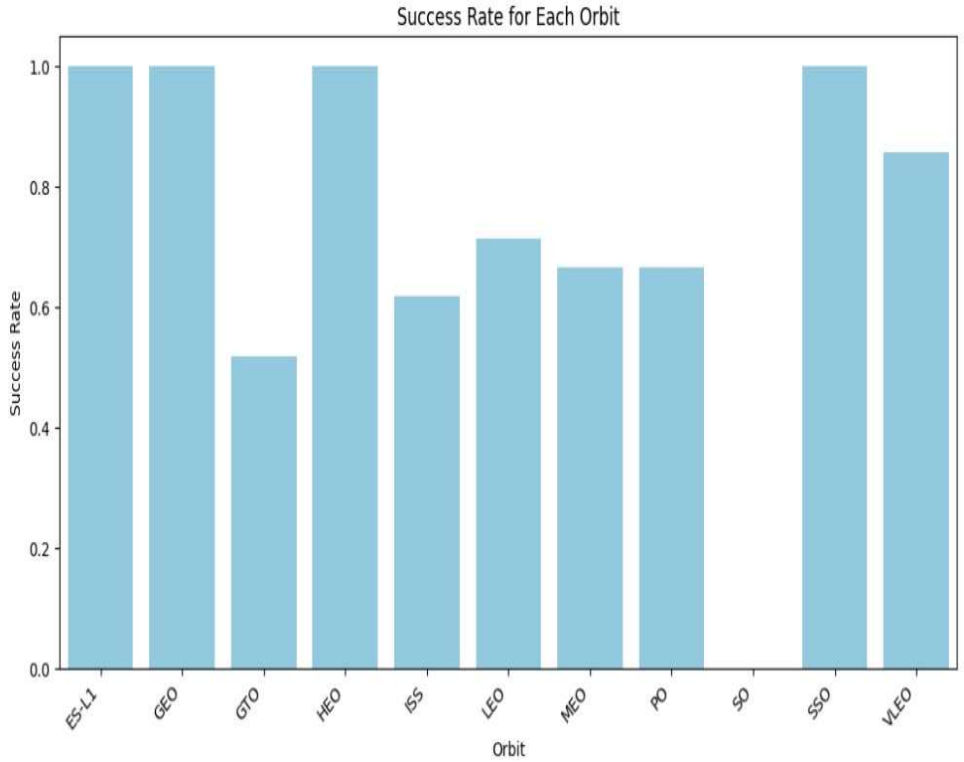
Data Visualization

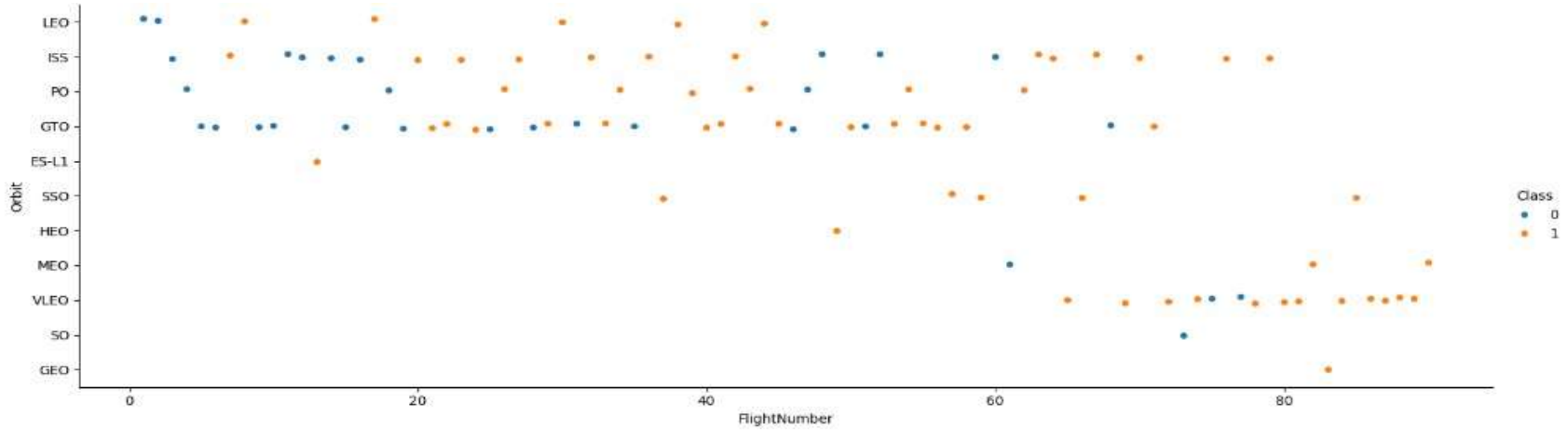
The following charts were plotted:

Flight Number vs. Launch Site, Flight Number vs. Payload Mass, Payload Mass vs. Launch Site, Orbit Type vs. Success Rate, Flight Number vs. Orbit Type, Payload Mass vs Orbit Type and Success Rate Yearly Trend

The scatter plots were used to examined The relationships between the variables. If any relationship exists, then those variables could be used to create machine learning models. Bar charts were used to show comparisons among discrete categories. Line charts were used to show trends in data over time


Average Success Rate per Year
Average Success Rate
0.0
0.2
0.4
0.6
0.8
2010
2012
2013
2014
2015
2016
2017
2018
2019
2020
Year


Success Rate for Each Orbit
Success Rate
0.0
0.2
0.4
0.6
0.8
1.0
ES-L1
GEO
GTO
HEO
ISS
LEO
MEO
PO
SO
SSO
VLEO
Orbit


Orbit
LEO
ISS
PO
GTO
ES-L1
SSO
HEO
MEO
VLEO
SO
GEO
0
20
40
60
80
FlightNumber
Class
0
1

SQL EDA

The following SQL queries were performed:

- Display of unique launch sites names
- Display of the first 5 rows where launch site name begins with 'CCA'
- Display of the total sum of payload carried by NASA (CRS) boosters
- Display of the average payload carried by F9 v1.1 booster
- Display the date when the first successful landing on ground pad was achieved
- Display of boosters which had success on drone ship landing and had payload mass greater than 4000kg but less than 6000kg
- Display the sum of successful and failure mission
- Listing the booster versions which carried the maximum payload mass
- Listing all failed landing on drone ship, their booster versions and launch site in year 2015

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

In [20]:
```
%%sql SELECT "Booster_Version", "PAYLOAD_MASS__KG_"  FROM SPACEXTABLE
WHERE "Landing_Outcome" LIKE "Success (drone ship)"
AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000;
```

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

Out[20]:

Booster_Version	PAYLOAD_MASS_KG_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

Task 7

List the total number of successful and failure mission outcomes

In [21]:
```
%%sql SELECT "Mission_Outcome", count("Mission_Outcome") FROM SPACEXTABLE
GROUP BY ("Mission_Outcome" LIKE "Success%");
```

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

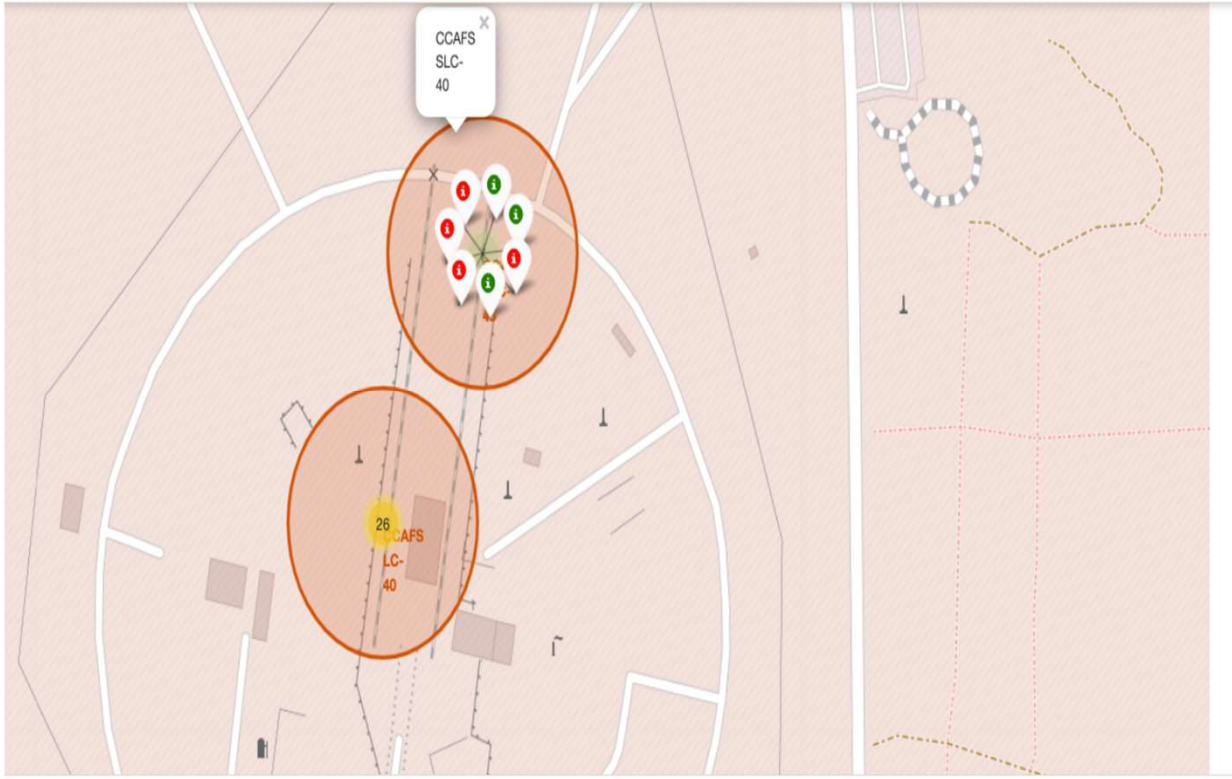
Out[21]:

Mission_Outcome	count("Mission_Outcome")
Failure (in flight)	1
Success	100

Folium Interactive Map

Folium maps displayed Launch Sites, successful and unsuccessful booster landings, and proximity examples to key locations: Railway, Highway, Coast, and City.

The maps provide insight into the reasons behind the selection of launch site. They also visualize successful landings in relation to their geographic location, aiding in location analysis



Plotly Dash Dashboard

- The dashboard features charts: a pie chart and a scatter plot.
- The pie chart allows selection to display the distribution of successful landings across all launch sites or individual launch site success rates.
- The scatter plot allows selection of either all sites or individual sites, with a slider for payload mass ranging from 0 to 10000 kg.
- The pie chart visualizes launch site success rates. The scatter plot helps in observing success rates across launch sites, payload mass, and booster version categories.

Predictive Analysis

Sequence:

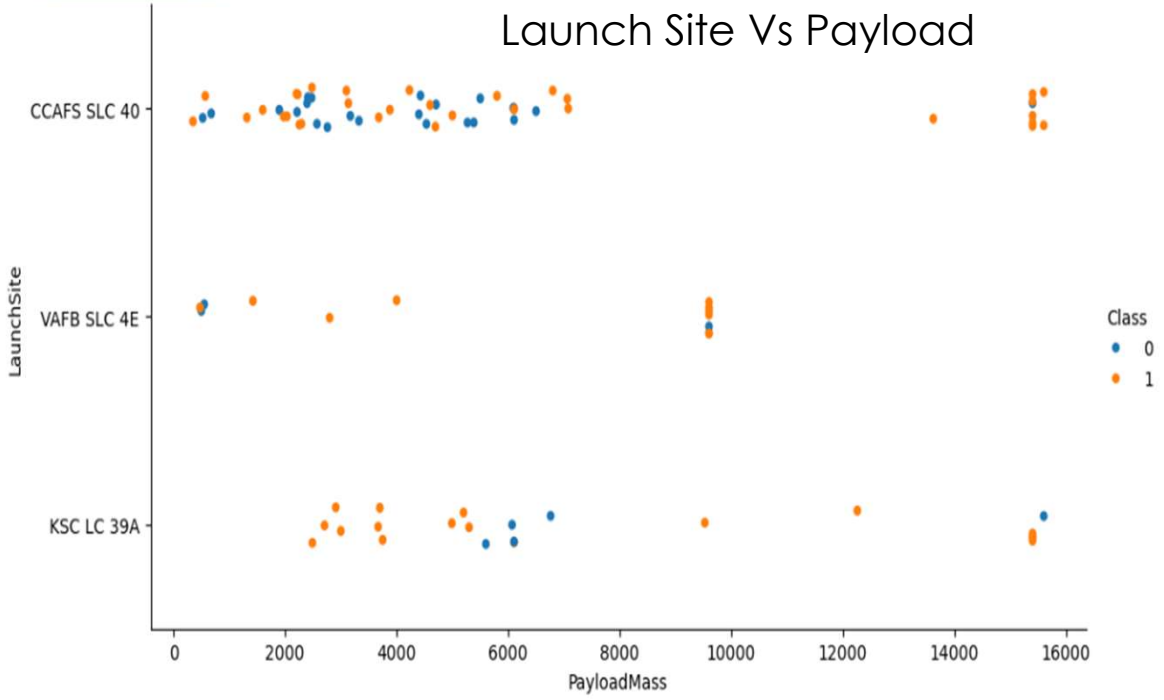
- Creation of a NumPy array from the column "Class"
- Standardizing of data with StandardScaler and fitting and transforming the data
- Splitting the data into train sets and test sets
- Creation of a GridSearchCV object
- Application of GridSearchCV on LogReg, SVM, Decision Tree, and KNN models
- Calculation of test data accuracy using the method .score() for all models Examining the confusion matrix for all models
- Locating the best performing method by examining the Jaccard_score and F1 score metrics

Results

The Plotly dashboard shows the outcomes of SQL and visualization Exploratory Data Analysis (EDA) through, an Interactive Folium Map, and the outcome of the predictive model which has an accuracy of 83%

EDA with Visualization

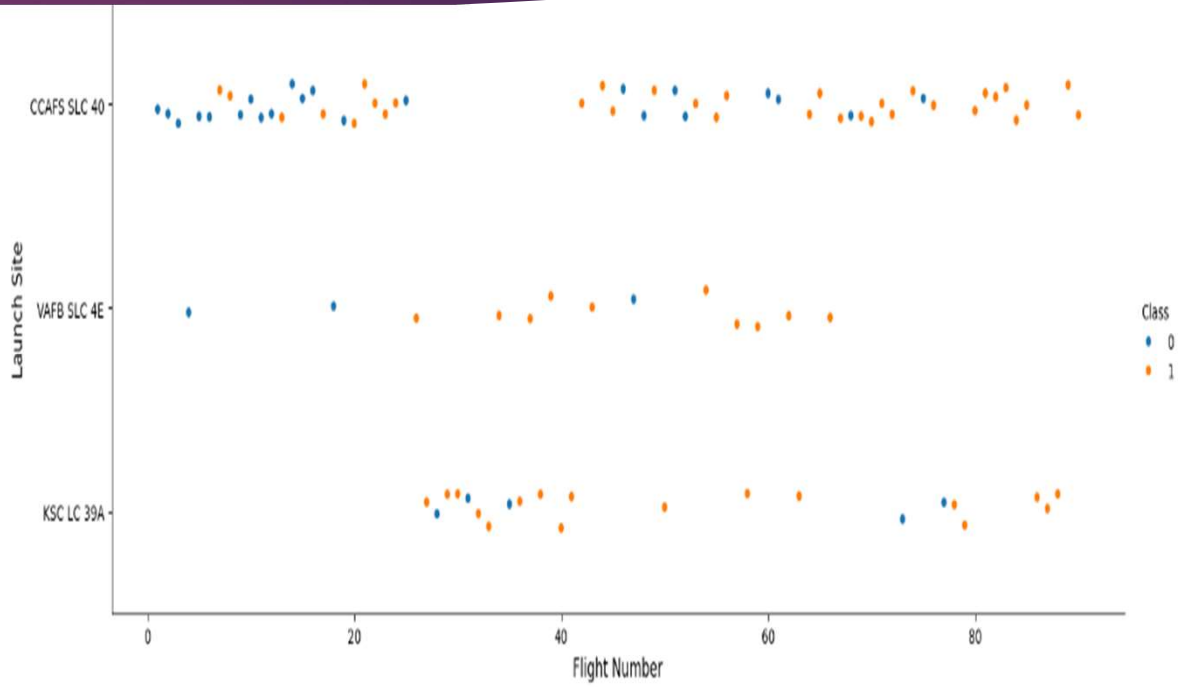
Observation shows that most of payload range from 0-6000 kg.
CCAFS SLC 40 had a high success rate when payload exceeds 6000kg



Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

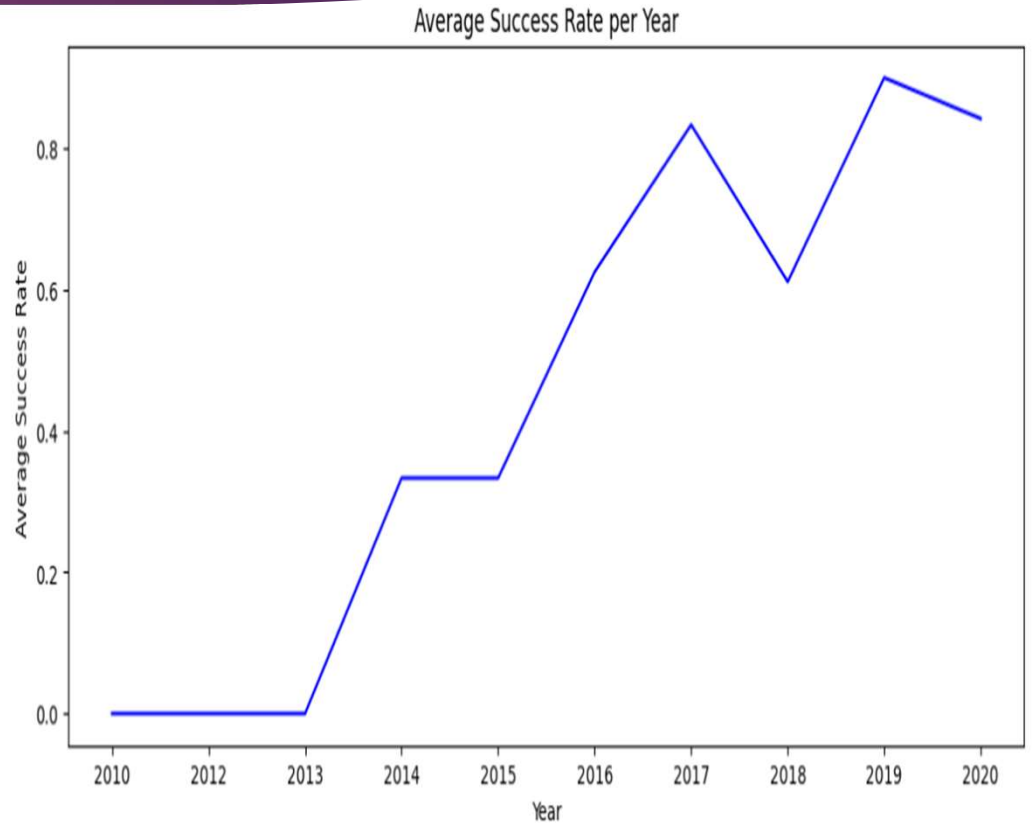
EDA with Visualization

The graph indicates a increase in success with increasing time After flight 20, there is a significant rise in success rates. Also, CCAFS might be the primary launch site as it had higher volume of launches compared to other launch sites.



EDA with Visualization

Success rate has increased significantly over time. It slightly dropped in 2018.Currently, success rate is around 84%.



EDA with SQL

This query retrieved all the unique launch site names from the SQL database

There are 4 unique launch sites.

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

```
In [9]: %sql SELECT distinct(Launch_Site) FROM  SPACEXTABLE;
```

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

Out[9]:

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

EDA with SQL

This query retrieved the first 5 rows from the database where the Launch Site name begins with the string "CCA"

In [18]:
```
%sql select * from SPACEXTABLE where "Launch_Site" like'CCA%' limit 5 ;
```

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

Out[18]:

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

EDA with SQL

- This SQL query Average Payload Mass by F9 v1.1
- It is 2535kg

Display average payload mass carried by booster version F9 v1.1

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTABLE where "Booster_Version" like 'F9 v1.1%';
```

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

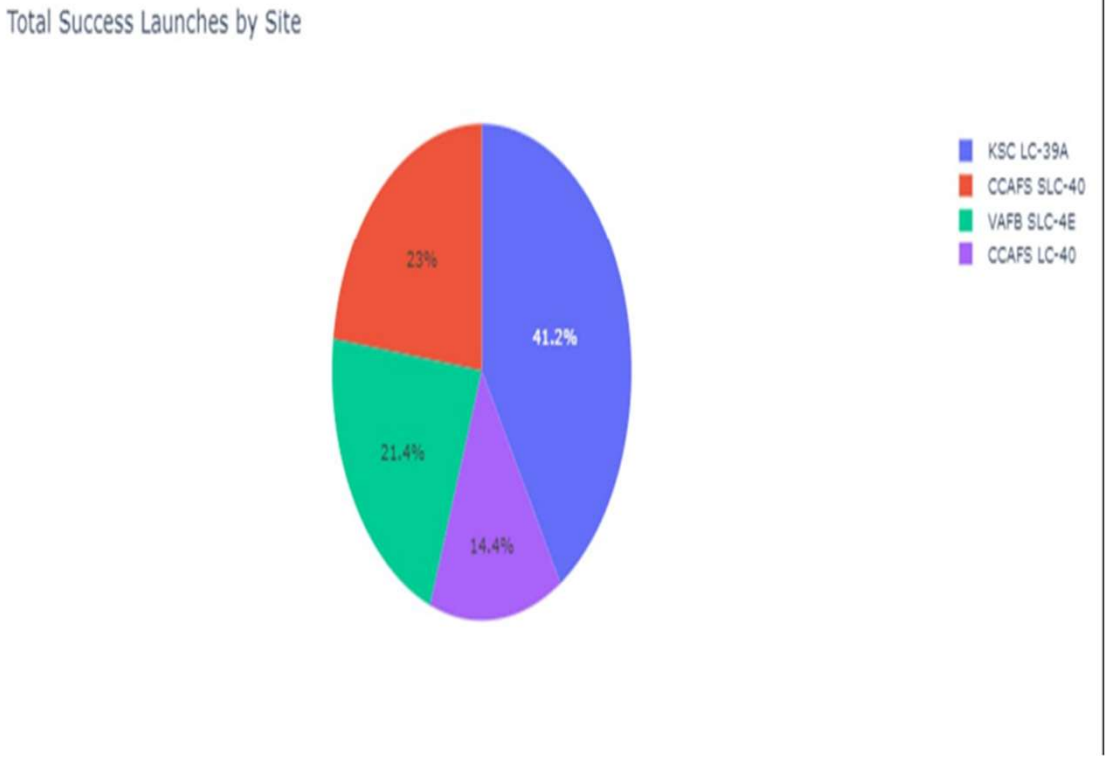
avg(PAYLOAD_MASS__KG_)
2534.6666666666665

Launch Analysis

- Almost all the launch sites are located closer to the equator. This is because earth moves faster at the equator.
- The launch sites are closer to the coast and are away from big cities in order to minimise exposing people to harm in case a rocket fails or explodes

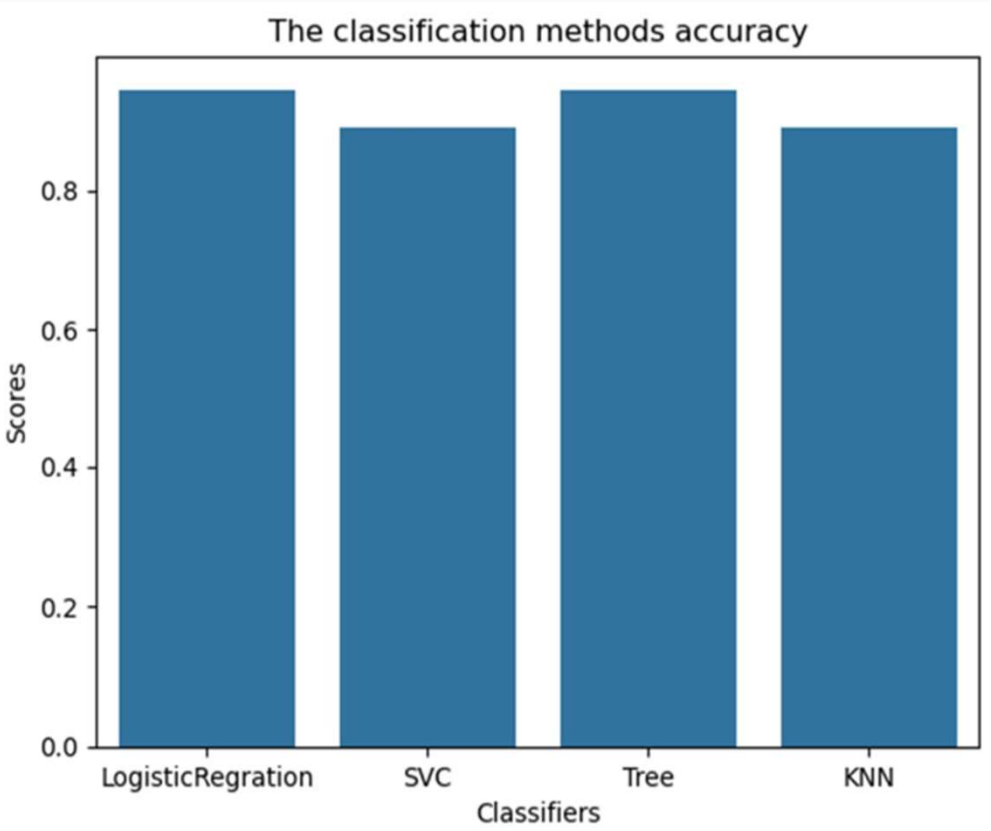
Build a Dashboard with Plotly Dash

CCAFS SLC-40, and KSC had the same number of successful landings. VAFB had the smallest proportion of successful landings



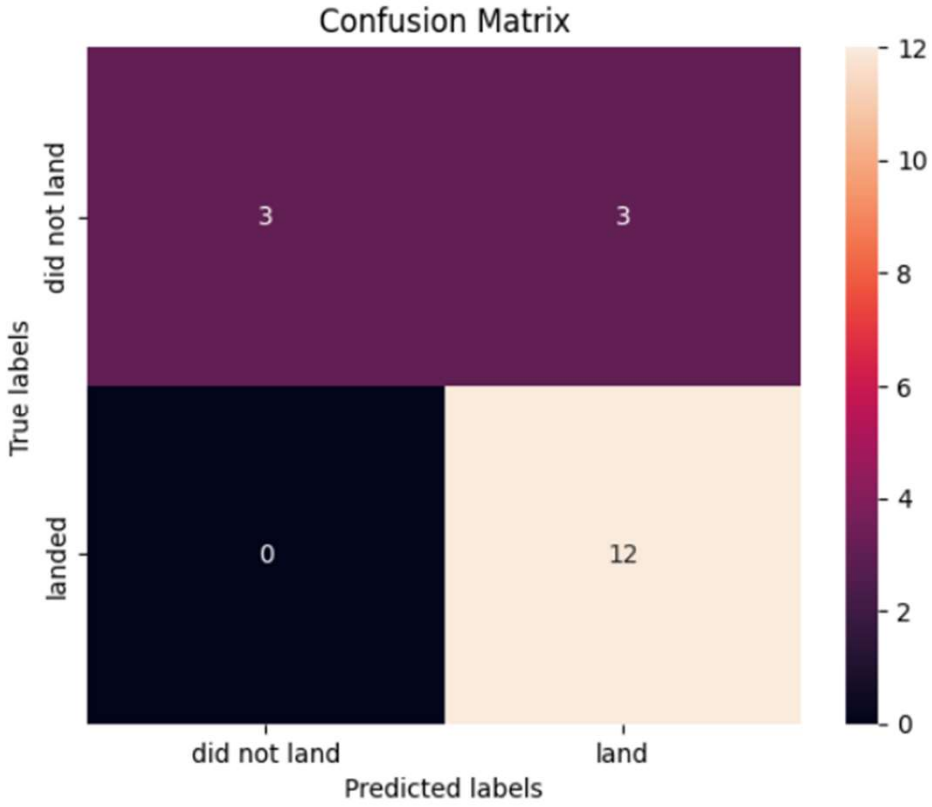
Predictive Analysis (Classification)

- All the predictive models achieved over 80% accuracy .
- However, logistic regression method gives the best results.
- Therefore, it was used for the predictive analysis.



Confusion Matrix Analysis

- True Positive: The models correctly predicted 12 successful landings.
- True Negative (TN): The models correctly predicted 3 unsuccessful landings.
- False Positive (FP): The models incorrectly predicted 3 successful landings



Conclusion

- Launches that had payload mass between 0-6000kg showed higher success than launches payload mass greater than 6000kg.
- All launch sites are closer to the equator and a coast line
- Launch success rate increased over time.
- KSC LC-39A had the highest success rate of all the launches sites

APPENDIX

Thanks to all instructors of this course:

https://www.coursera.org/professional-certificates/ibm-data-science?#instructors

GitHub repository:https://github.com/CODE-RULES/IBM-Capstone