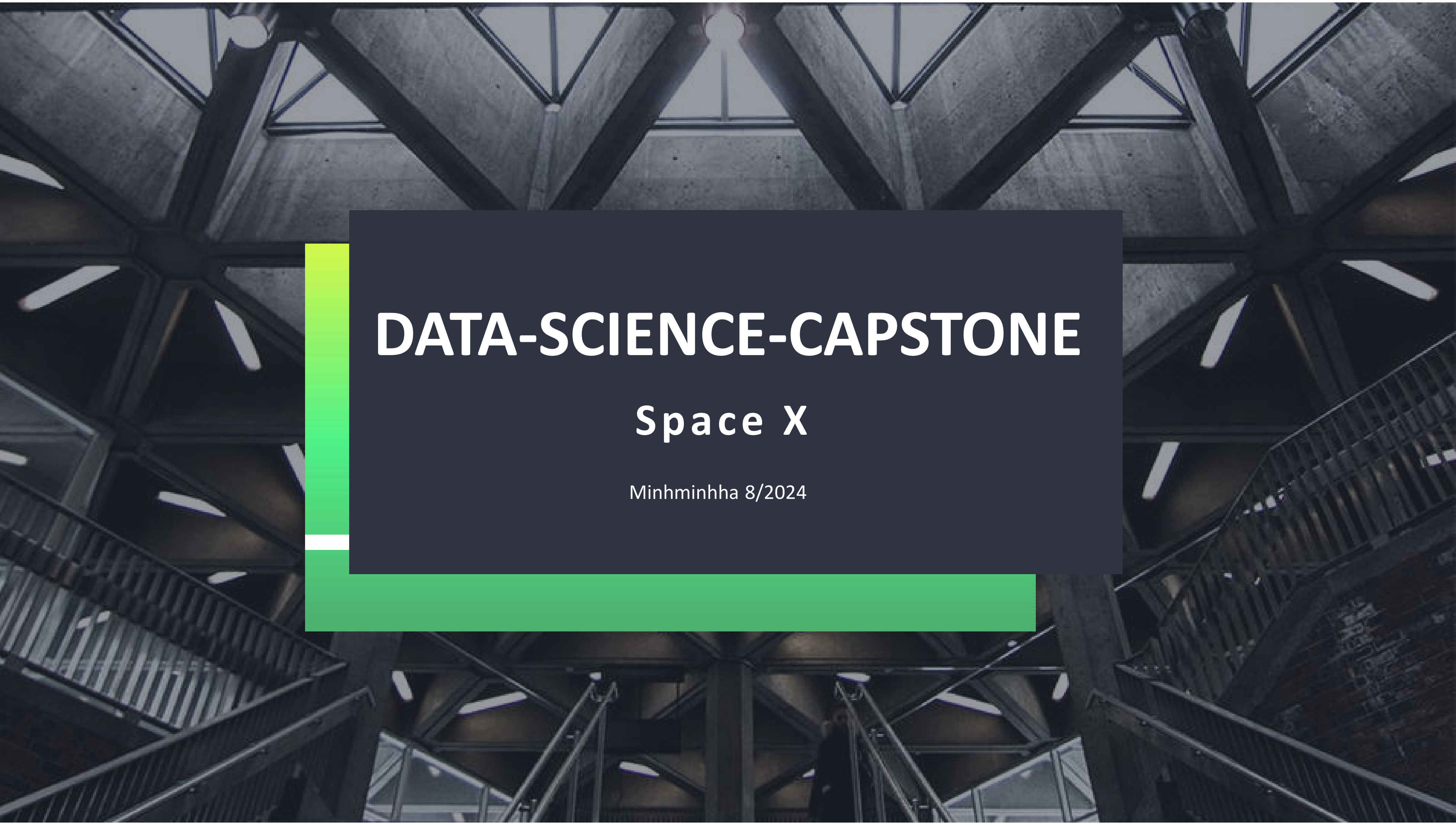


DATA-SCIENCE-CAPSTONE

Space X

Minhminhha 8/2024

CONTENT

1. Executive Summary
2. Introduction
3. Methodology
4. Results
5. Conclusions

1. EXECUTIVE SUMMARY

Summary of methodologies

Our approach encompassed a dual methodology involving API integration and web scraping techniques for data collection. Following the acquisition phase, we employed a suite of Python data manipulation methods to meticulously process and cleanse the dataset. Subsequently, SQL queries were employed to extract pertinent information from the refined dataset. Early insights were garnered through systematic data visualization and trend analysis. Concluding our analytical framework, we implemented supervised machine learning models to formulate predictions regarding the success of landing events. We applied supervised machine learning models to make predictions about the success of the landing event. this domain.

Summary of all results

Through meticulous data analysis, we identified discernible patterns and correlations among variables directly influencing the success of landing events. Leveraging these insights, we developed and trained a predictive model that demonstrated a notable capability to accurately forecast the probability of a successful landing event. Notably, the model achieved a commendable accuracy rate of 83%, underscoring its effectiveness in providing reliable prognostications within this domain.

2.INTRODUCTION

• SpaceX's commitment to reusable rockets has significantly mitigated space travel costs by strategically focusing on the retrieval of the first rocket phase. The recovery of this initial phase is paramount in preserving and reusing expensive components, contributing directly to cost reduction. An in-depth analysis of the success rate of these retrieval events serves as a valuable metric for evaluating efficiency and cost-effectiveness in SpaceX's pioneering approach. This particular project is geared towards predicting the success of the first phase retrieval event, thereby offering predictive insights aimed at enhancing decision-making within the space industry.• Our objective is to forecast the success of first-phase rocket retrieval, with the overarching aim of optimizing resource allocation. By achieving this predictive capability, we seek to enhance mission success rates and contribute to substantial cost savings.

3.METHODOLOGY

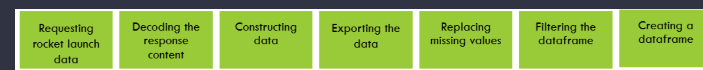
Executive Summary

- Data collection
- Data wrangling
- EDA using visualization and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis using classification models

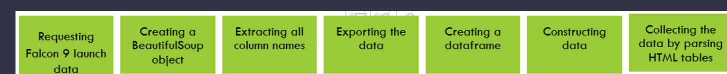
Data collection

- Data collection process involved a combination of API requests from SpaceX REST API and Web Scraping data from a table in SpaceX's Wikipedia entry.We had to use both of these data collection methods in order to get complete information about the launches for a more detailed analysis.
- Data Columns are obtained by using SpaceX REST API: FlightNumber, Date, Booster Version, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude

Requests from SpaceX API



Requests from Web scraping



DATA WRANGLING

In the data set, there are several different cases where the booster did not land successfully. Sometimes a landing was attempted but failed due to an accident; for example, True Ocean means the mission outcome was successfully landed to a specific region of the ocean while False Ocean means the mission outcome was unsuccessfully landed to a specific region of the ocean. True RTLS means the mission outcome was successfully landed to a ground pad False RTLS means the mission outcome was unsuccessfully landed to a ground pad.True ASDS means the mission outcome was successfully landed on a drone ship False ASDS means the mission outcome was unsuccessfully landed on a drone ship.We mainly convert those outcomes into Training Labels with "1" means the booster successfully landed, "O" means it was unsuccessful.

- Perform exploratory Data Analysis and determine Training Labels
- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurrence of mission outcome per orbit type
- Create a landing outcome label from Outcome column
- Exporting the data to CSV

EDA using visualization

We explored the data by visualizing the relationship between flight number and launch Site, payload and launch site, success rate of each orbit type, flight number and orbit type, the launch succes

EDA with SQL

- Displaying the names of the launch sites. Displaying 5 records where launch sites begin with the string ‘CCA’.
- Displaying the total payload mass carried by booster launched by NASA (CRS). Displaying the average payload mass carried by booster version F9 v1.1.
- Listing the date when the first successful landing outcome in ground pad was achieved.
- Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000. Listing the total number of successful and failure mission outcomes.
- Listing the names of the booster_versions which have carried the maximum payload mass.

INTERACTIVE VISUAL ANALYTICS USING FOLIUM AND PLOTLY DASH

INTERACTIVE MAP WITH FOLIUM

Markers of all Launch Sites:

- Added Marker with Circle, Popup Label and Text Label of NASA Johnson Space Center using its latitude and longitude coordinates as a start location.
- Added Markers with Circle, Popup Label and Text Label of all Launch Sites using their latitude and longitude coordinates to show their geographical locations and proximity to Equator and coasts.

Coloured Markers of the launch outcomes for each Launch Site:

- Added coloured Markers of success (Green) and failed (Red) launches using Marker Cluster to identify which launch sites have relatively high success rates.

Distances between a Launch Site to its proximities:

- Added coloured Lines to show distances between the Launch Site KSC LC-39A (as an example) and its proximities like Railway, Highway, Coastline and Closest City.

DASHBOARD WITH PLOTLY DASH

Launch Sites Dropdown List:

- Added a dropdown list to enable Launch Site selection.

Pie Chart showing Success Launches (All Sites/Certain Site):

- Added a pie chart to show the total successful launches count for all sites and the Success vs. Failed counts for the site, if a specific Launch Site was selected.

Slider of Payload Mass Range:Added a slider to select Payload range.Scatter Chart of Payload Mass vs. Success Rate for the different Booster Versions:

- Added a scatter chart to show the correlation between Payload and Launch Success.

PREDICTIVE ANALYSIS USING CLASSIFICATION MODELS

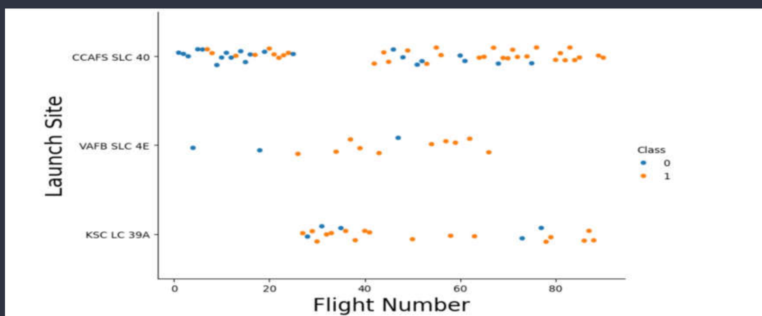
- Creating a NumPy array from the column "Class" in data
- Finding the method performs best by examining the Jaccard_score and F1_score metrics
- Standardizing the data with StandardScaler, then fitting and transforming it
- Examining the confusion matrix for all models
- Splitting the data into training and testing sets with train_test_split function
- Calculating the accuracy on the test data using the method.score() for all models
- Creating a GridSearchCV object with cv = 10 to find the best parameters
- Applying GridSearchCV on LogReg, SVM, Decision Tree, and KNN models

4.RESULTS

- Exploratory data analysis
- Interactive analytics demo in screenshots
- Predictive analysis

EDA USING VISUALIZATION

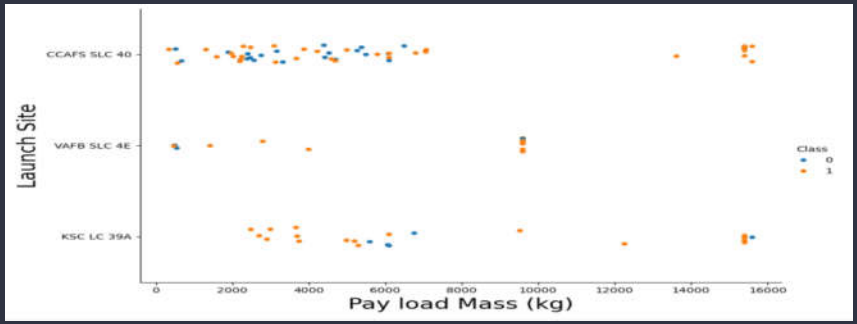
FLIGHT NUMBER VS. LAUNCH SITE



From the plot, we found that the larger the flight amount at a launch site, the greater the success rate at a launch site

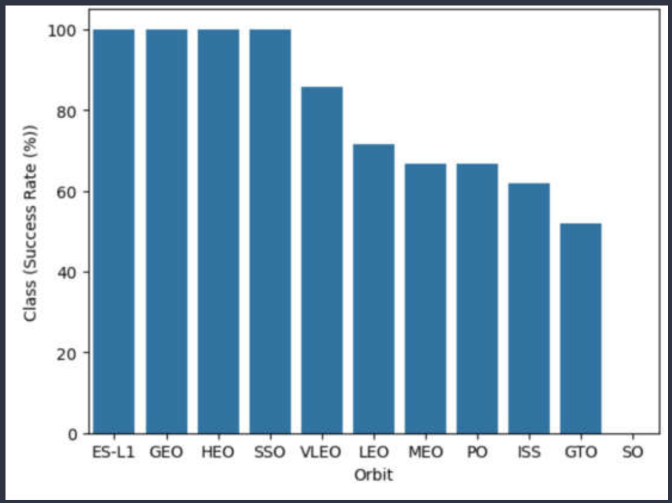
PAYLOAD VS. LAUNCH SITE

The greater the payload mass for launch site CCAFS SLC 40 the higher the success rate for the rocket.



EDA USING VISUALIZATION

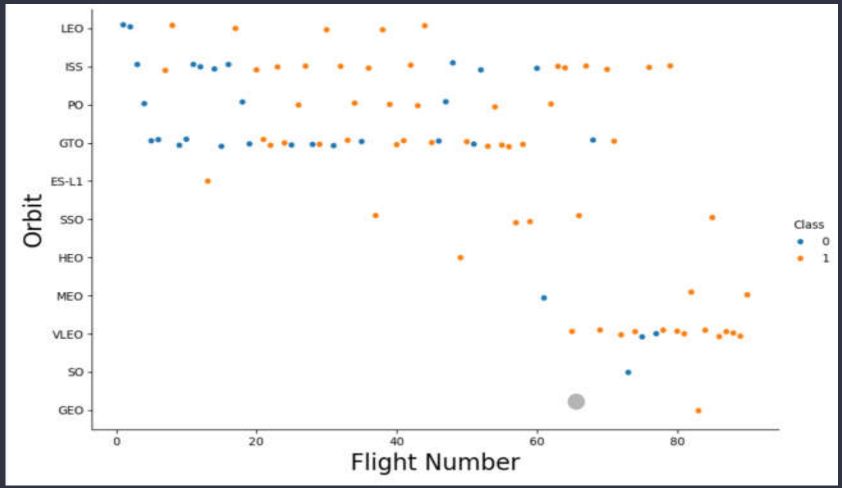
SUCCESS RATE VS. ORBIT TYPE



From the plot, we can see that ES-L1, GEO, HEO, SSO, VLEO had the most success rate

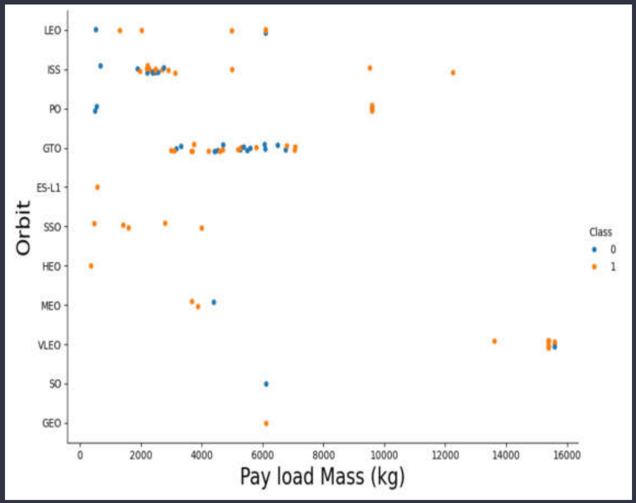
FLIGHT NUMBER VS. ORBIT TYPE

The plot below shows the Flight Number vs. Orbit type. We observe that in the LEO orbit, success is related to the number of flights whereas in the GTO orbit, there is no relationship between flight number and the orbit.



EDA USING VISUALIZATION

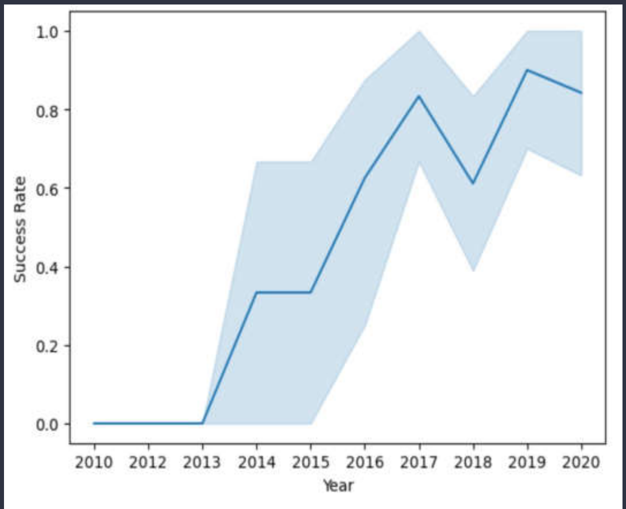
PAYLOAD VS. ORBIT TYPE



We can observe that with heavy payloads, the successful landing are more for PO,LEO and ISS orbits

LAUNCH SUCCESS YEARLY TREND

From the plot, we can observe that success rate since 2013 kept on increasing till 2020



EDA WITH SQL

ALL LAUNCH SITE NAMES

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

```
In [31]: %sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEXTBL;
 * sqlite:///my_data1.db
Done.
```

Out[31]:

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

Used the key word DISTINCT to show only unique launch sites from the SpaceX data

LAUNCH SITE NAMES BEGIN WITH 'CCA'

Used the query above to display 5 records where launch sites begin with CCA

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

```
%sql SELECT * FROM 'SPACEXTBL' 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	Landin _Outcom
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failu (parachut
08-12-2010	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	Failu (parachut
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attem
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attem
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attem

EDA WITH SQL

TOTAL PAYLOAD MASS

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

```
In [17]: %sql SELECT SUM(PAYLOAD_MASS__KG_) as "Total Payload Mass(Kgs)", Customer FROM 'SPACEXTBL' WHERE Customer = 'NASA
 * sqlite:///my_data1.db
Done.
```

Out[17]:

Total Payload Mass(Kgs)	Customer
45596	NASA (CRS)

Calculated the total payload carried by boosters from NASA as 45596 using the query below

AVERAGE PAYLOAD MASS BY F9 V1.1

Calculated the average payload mass carried by booster version F9 v1.1 B1003 as 2534.66666666666

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

```
In [19]: %sql SELECT AVG(PAYLOAD_MASS__KG_) as "Payload Mass Kgs", Customer, Booster_Version FROM 'SPACEXTBL' WHERE Booste
 * sqlite:///my_data1.db
Done.
```

Out[19]:

Payload Mass Kgs	Customer	Booster_Version
2534.6666666666665	MDA	F9 v1.1 B1003

EDA WITH SQL

FIRST SUCCESSFUL GROUND LANDING DATE

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

Hint:Use min function

In [21]:
```
%sql SELECT MIN(DATE) FROM 'SPACEXTBL' WHERE "Landing _Outcome" = "Success (ground pad)";
```

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

Out[21]:

MIN(DATE)
01-05-2017

We observed that the dates of the first successful landing outcome on ground pad was 1st May 2017

SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD BETWEEN 4000 AND 6000

We used the WHERE clause to filter for boosters which have success fully landed on drone-ship and applied the and condition to determine successful landing with payload mass greater than 4000 but less than 6000.

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

In [26]:
```
# %sql SELECT * FROM 'SPACEXTBL'
```

In [27]:
```
%sql SELECT DISTINCT Booster_Version, Payload FROM SPACEXTBL WHERE "Landing _Outcome" = "Success (drone ship)" AN
```

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

Out[27]:

Booster_Version	Payload
F9 FT B1022	JCSAT-14
F9 FT B1026	JCSAT-16
F9 FT B1021.2	SES-10
F9 FT B1031.2	SES-11 / EchoStar 105

EDA WITH SQL

TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

List the total number of successful and failure mission outcomes

```
In [28]: %sql SELECT "Mission_Outcome", COUNT("Mission_Outcome") as Total FROM SPACEXTBL GROUP BY "Mission_Outcome";

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

Out[28]:

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

Total number of Mission outcome was a success or a failure

BOOSTERS CARRIED MAXIMUM PAYLOAD

We determined the booster that have carried the maximum payload using a subquery in the WHERE clause and the MAX() function

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

```
In [30]: %sql SELECT "Booster_Version",Payload, "PAYLOAD_MASS__KG_" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX

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

Out[30]:

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 B5 B1048.4	Starlink 1 v1.0, SpaceX CRS-19	15600
F9 B5 B1049.4	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600
F9 B5 B1051.3	Starlink 3 v1.0, Starlink 4 v1.0	15600
F9 B5 B1056.4	Starlink 4 v1.0, SpaceX CRS-20	15600
F9 B5 B1048.5	Starlink 5 v1.0, Starlink 6 v1.0	15600
F9 B5 B1051.4	Starlink 6 v1.0, Crew Dragon Demo-2	15600
F9 B5 B1049.5	Starlink 7 v1.0, Starlink 8 v1.0	15600
F9 B5 B1060.2	Starlink 11 v1.0, Starlink 12 v1.0	15600
F9 B5 B1058.3	Starlink 12 v1.0, Starlink 13 v1.0	15600
F9 B5 B1051.6	Starlink 13 v1.0, Starlink 14 v1.0	15600
F9 B5 B1060.3	Starlink 14 v1.0, GPS III-04	15600
F9 B5 B1049.7	Starlink 15 v1.0, SpaceX CRS-21	15600

EDA WITH SQL

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.

```
In [68]: %sql SELECT substr(Date,7,4), substr(Date, 4, 2),"Booster_Version", "Launch_Site", Payload, "PAYLOAD_MASS__KG_",

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

Out[68]:

substr(Date,7,4)	substr(Date, 4, 2)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Mission_Outcome	Landing _Outcome
2015	01	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	Success	Failure (drone ship)
2015	04	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	Success	Failure (drone ship)

We used a combinations of the WHERE clause, LIKE, AND, and Between conditions to filter for failed landing outcomes in drone ship, their booster versions, and launch site names for year 2015.

BOOSTERS CARRIED MAXIMUM PAYLOAD

We selected Landing outcome sand the COUNT of landing outcomes from the data and-used the WHERE clause to filter for landing outcomes between 2010-06-04 to 2010-03-20.

• We applied the group by clause to group the landing outcomes and the order by clause to order the grouped landing outcomes in descending order.

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.

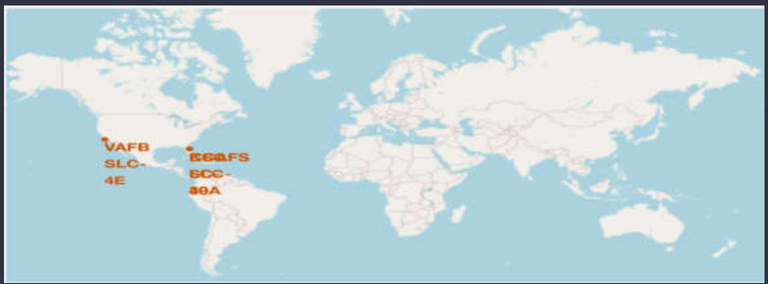
```
In [45]: %sql SELECT LANDING_OUTCOME, COUNT(*) AS COUNT_LAUNCHES FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-

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

Out[45]:

Landing_Outcome	COUNT_LAUNCHES
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

LAUCH SITES PROXIMITIES ANALYSIS

ALL LAUNCH SITES GLOBAL MAP MARKER



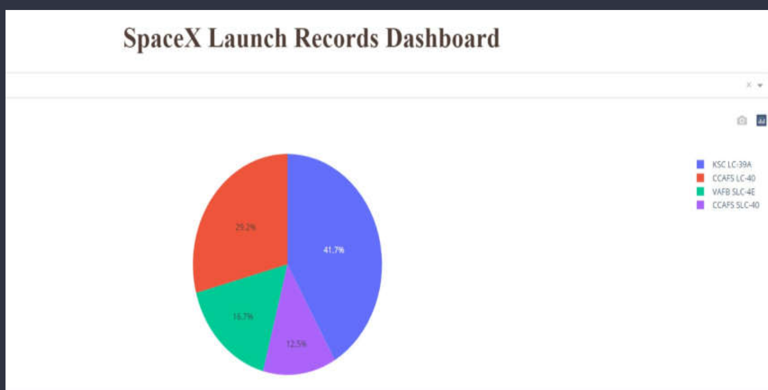
We can see that the Space launch sites are in the United States of America coasts. Florida and California

MARKERS SHOWING LAUNCH SITES WITH COLOR LABELS

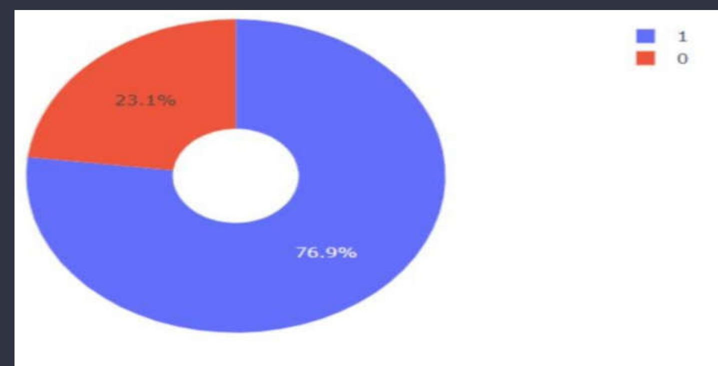
Green marker showing Successful Launches.
Red marker showing Unsuccessful Launches

BUILD DASHBOARD WITH PLOTY DASH

TOTAL SUCCESS LAUNCHES BY ALL SITES

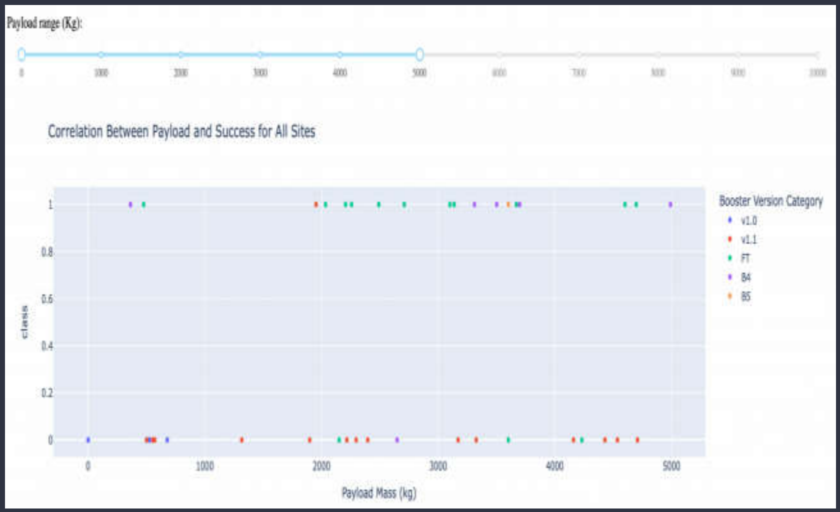


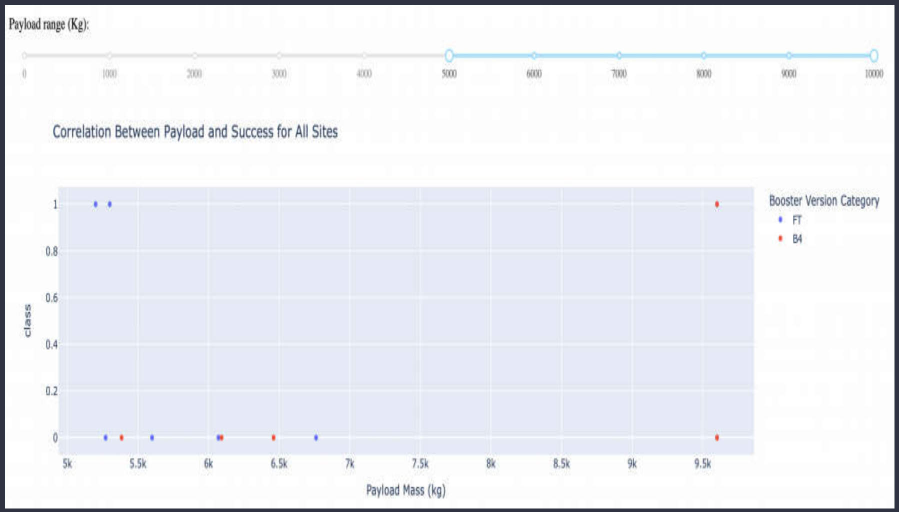
TOTAL SUCCESS LAUNCHES FOR KSC LC-39A



BUILD DASHBOARD WITH PLOTY DASH

PAYLOAD MAS VS LAUNCH OUTCOME FOR ALL SITES





PREDICTIVE ANALYSIS

CLASSIFICATION ACCURACY

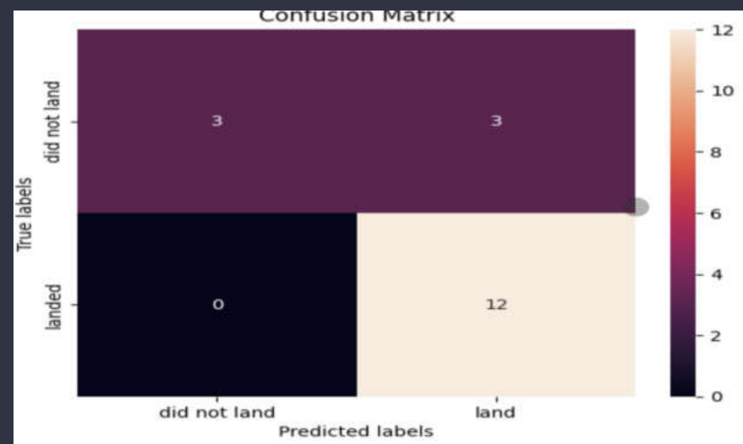
Scores and Accuracy of test set

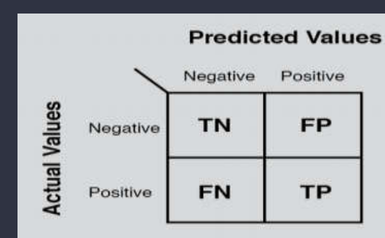
	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

Scores and Accuracy of the entire data set

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.833333	0.845070	0.882353	0.819444
F1_Score	0.909091	0.916031	0.937500	0.900763
Accuracy	0.866667	0.877778	0.911111	0.855556

CONFUSION MATRIX





5.CONCLUSIONS

The larger the flight amount at a launch site, the greater the success rate at a launch site.

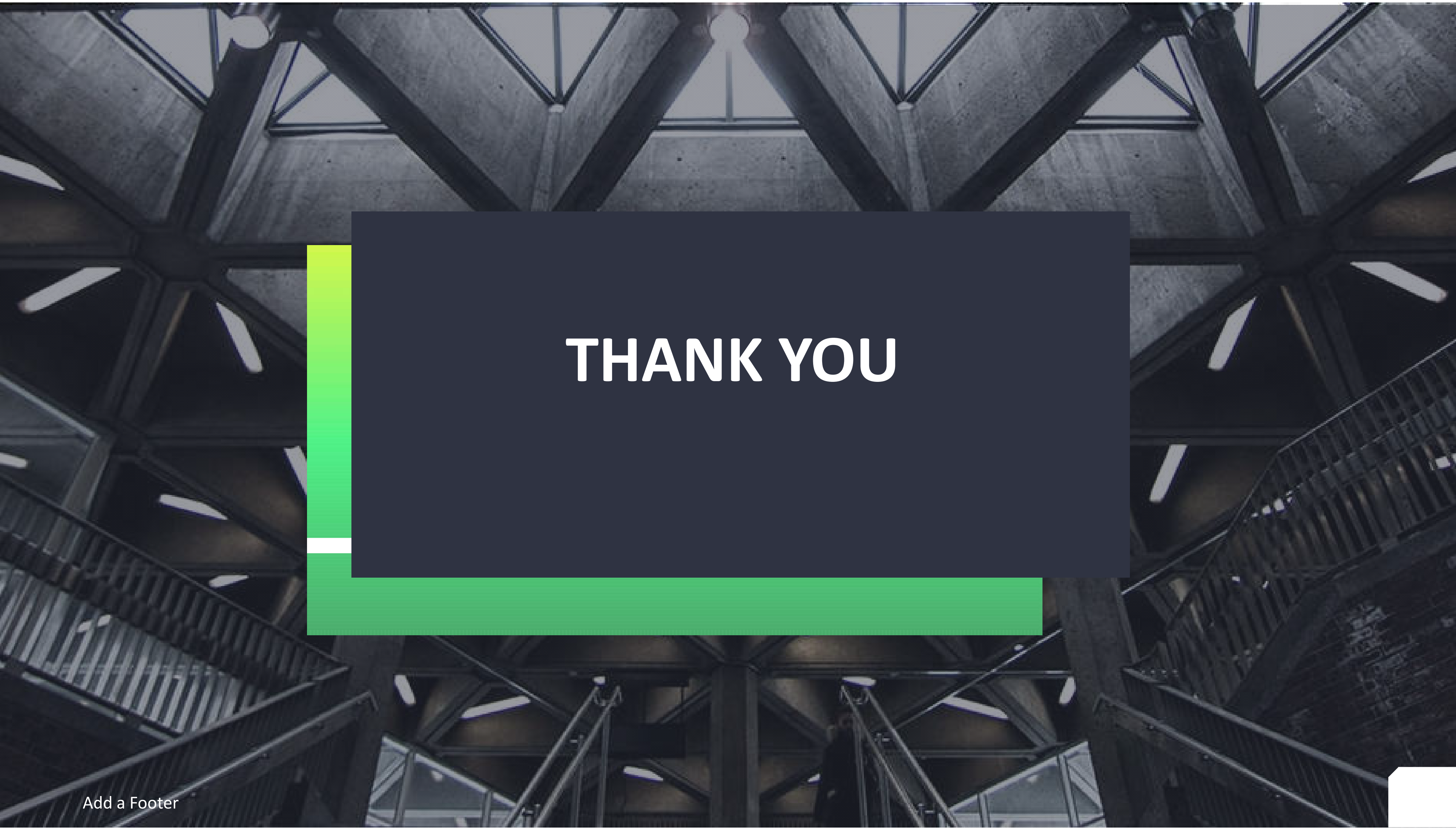
Launch success rate started to increase in 2013 till 2020. Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.

KSC LC-39A had the most successful launches of any sites.

The Decision tree classifier is the best machine learning algorithm for this task.

HEREFlight Number vs. Launch Site

Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Ut gravida eros erat. Proin a tellus sed risus lobortis sagitti



THANK YOU