WINNING THE SPACE RACE WITH DATA SCIENCE

IBM Developer
SKILLS NETWORK

Joel Knapp

03/22/2024

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Insights Drawn from EDA
  - Launch Sites Proximity Analysis
  - Building a Dashboard with Plotly Dash
  - Predictive Analysis
- Conclusion
- Appendix (Github Repository)

EXECUTIVE SUMMARY

Data Collection

- API and Webscraping

Data Wrangling

Exploratory Data Analysis

- SQL Queries
- Data Visualization

Visual Analytics

- Map Analysis using Folium
- SpaceX Dashboard with Plotly Dash

Predictive Analysis

INTRODUCTION

The aim of this report is to predict if the Falcon 9 first stage will land successfully.

SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.

Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

DATA COLLECTION & WRANGLING

Data was collected through API from the SpaceX website as well as Web Scraping the Space X Wikipedia

Data was analyzed and transformed to categorize each launch as a "failure (0)" or "success (1)"

DATA COLLECTION – API

SpaceX data was collected from api.spacexdata.com/v4 and converted into a dataframe using Pandas

Request and Parse Data

- Decode content as json and convert do dataframe

Filter the Dataframe

- Filter data to isolate Falcon 9 launches

Remove Missing Values

- Use isnull() function to remove missing values

DATA COLLECTION – WEB SCRAPING

Data on Falcon 9 Launches was scraped from Wikipedia using BeautifulSoup

Request data from URL

- Use get() method to extract data from Wikipedia

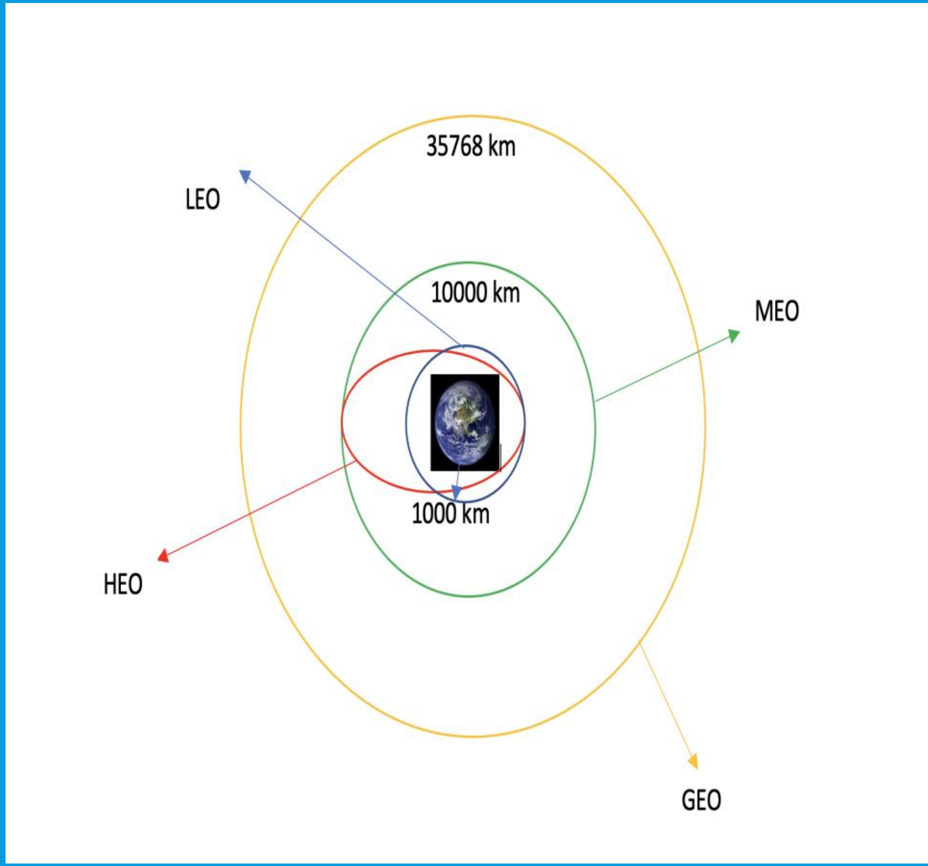
Extract Column Names

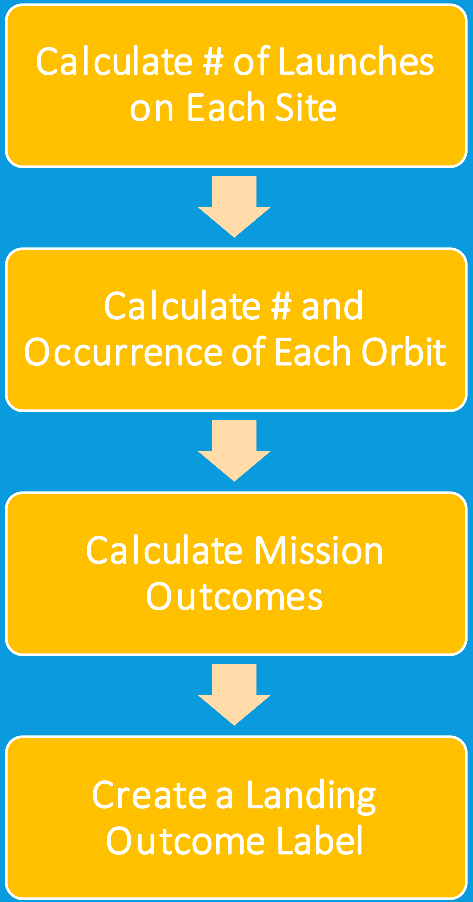
- Search tables for 'th' element and collect all column names

Create a Dataframe

- Parse the html tables into a Pandas dataframe

DATA WRANGLING





- Use method valuecounts() to determine # of launches, orbits, mission outcomes

- Use the output to create a Outcome column:

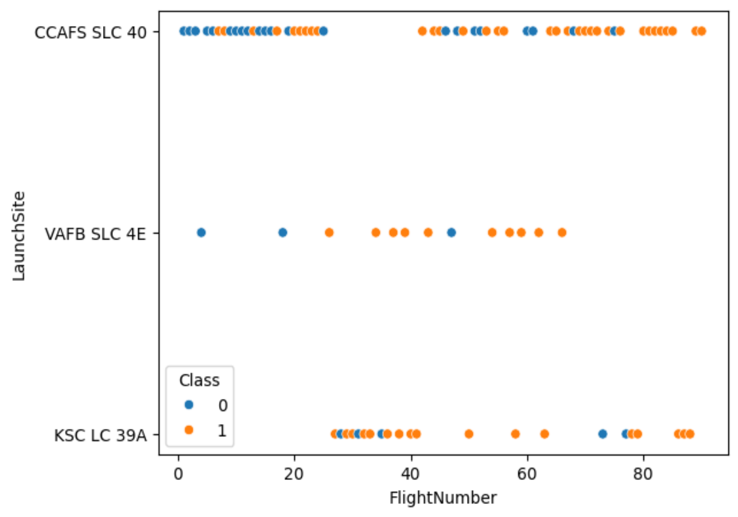
  0 = failed outcome

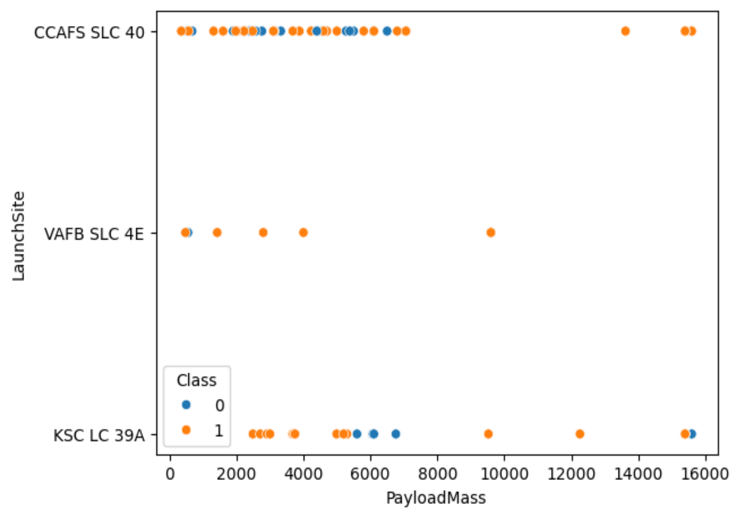
  1 = successful outcome

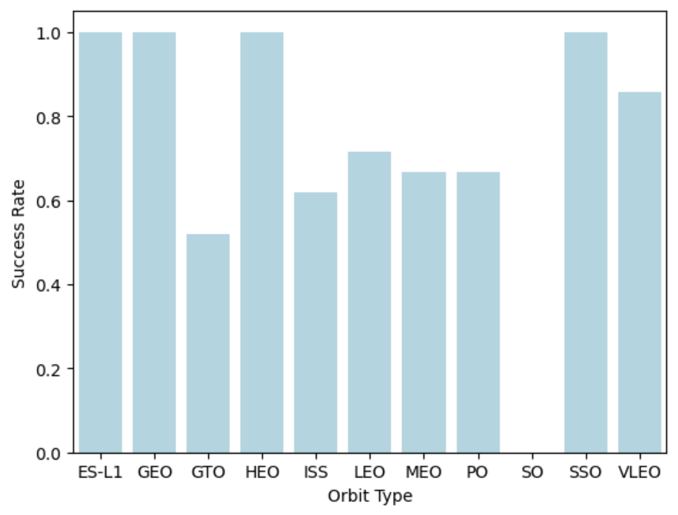
EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis was conducted utilizing Data Visualization leveraging Plotly and Seaborn. The relationships between Flight Number, Payload Mass, Orbit Type, Launch Site, and Year were analyzed in order to gain further insights into the variables affecting successful launches.

EDA WITH DATA VISUALIZATION





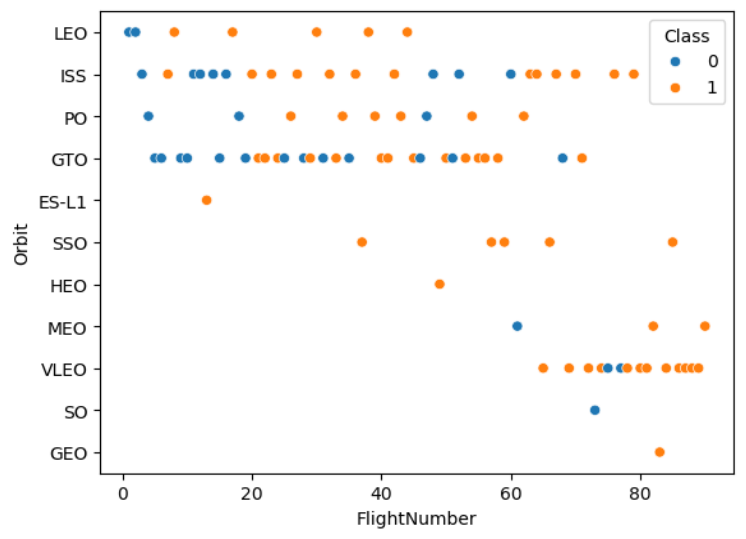


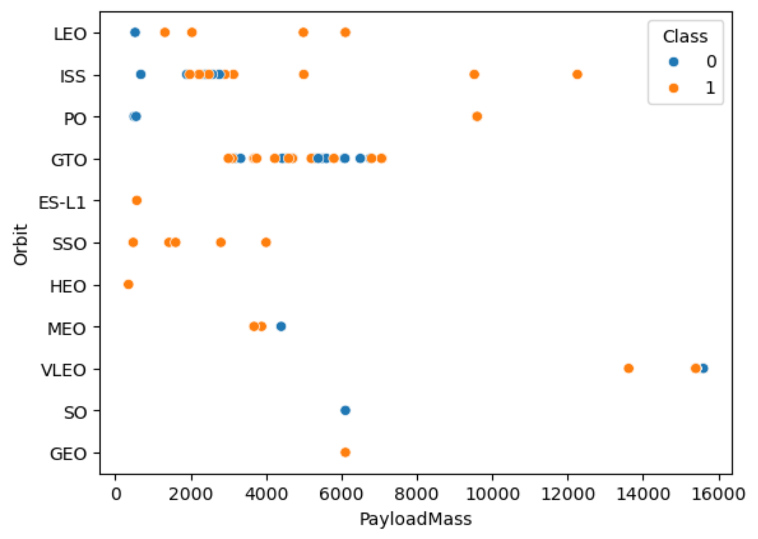
- The chances of success (orange) increases with additional launches
- CCAFS SLC 40 has seen the most launches

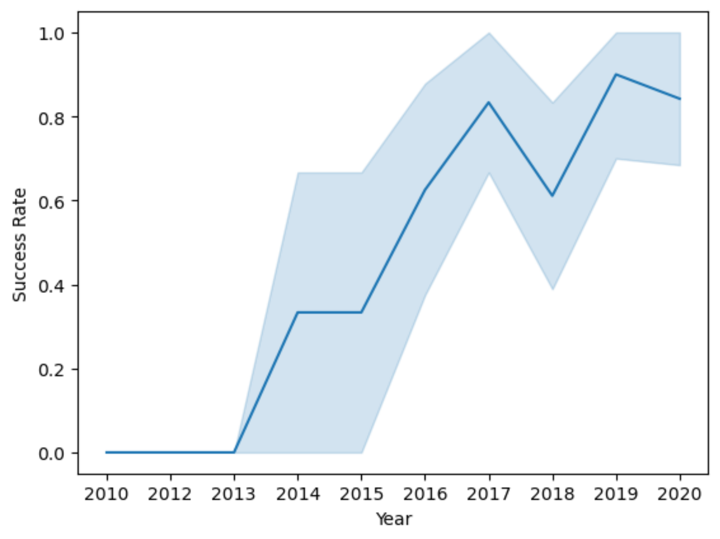
- There have been 0 heavy rockets (Payload Mass > 10,000) launched from VAFB SLC 4E launch site

- Orbit types ES-L1, GEO, HEO, and SSO exhibit the highest success rates

EDA WITH DATA VISUALIZATION







- The type of orbit has changed with successive launches
- LEO, SSO, and VLEO are the most successful orbit types

- With heavy payloads the chance of a successful landing is higher for Polar, LEO and ISS
- GTO success rate is not as easy to distinguish

- Success rates have consistently increased since 2013

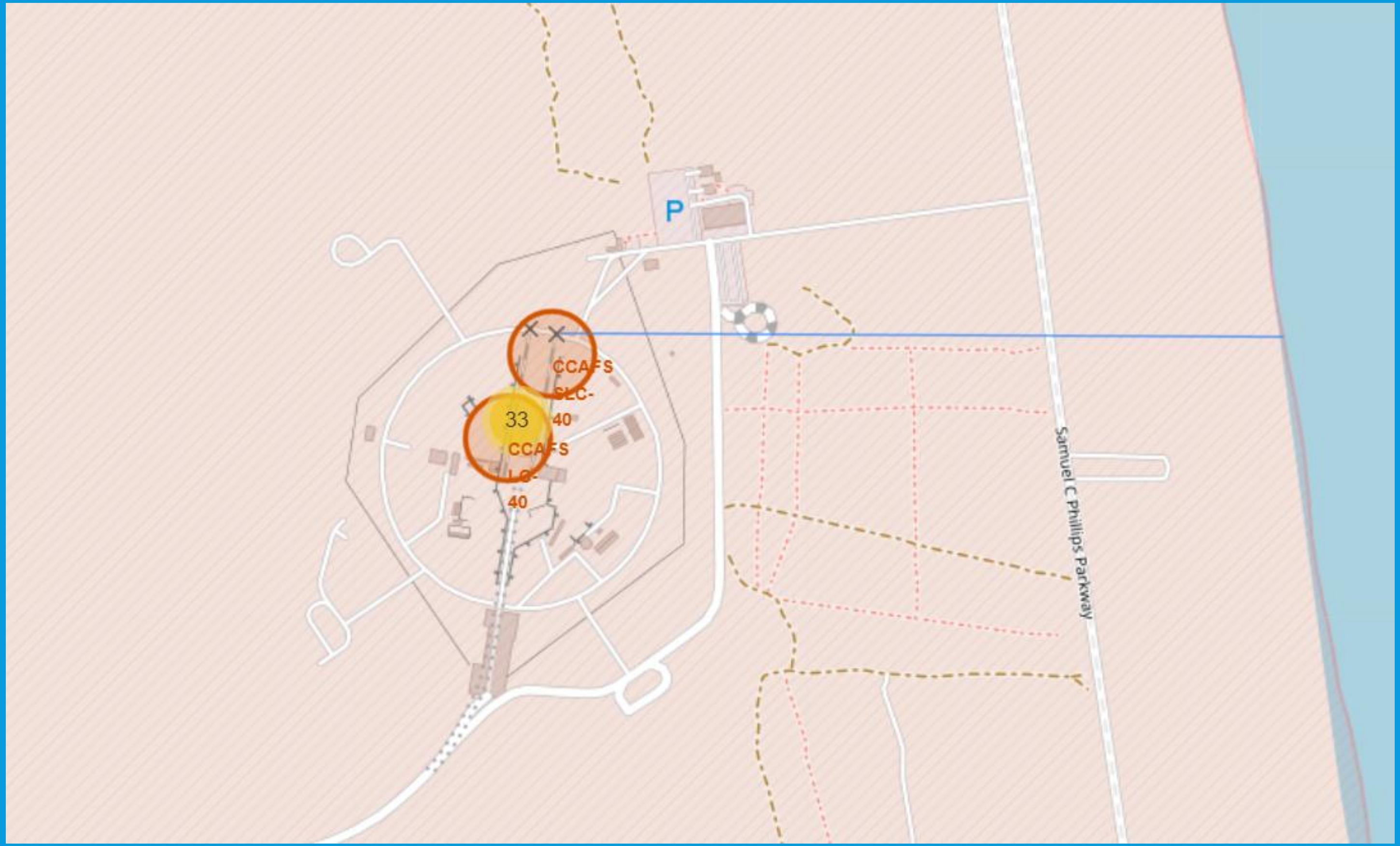
EDA WITH SQL

- Use SQL to Perform Exploratory Data Analysis:
- Select the names of the unique launch sites
- Filter records where launch sites begin with the string 'CCA'
- Calculate the total payload mass carried by boosters launched by NASA (CRS)
- Determine average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was achieved.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass

INTERACTIVE VISUAL ANALYTICS AND DASHBOARD

An interactive map was built using Folium to display the location of launch sites and the success rate of launches. A dashboard was built using Plotly Dash to examine launch records interactively.

INTERACTIVE MAP USING FOLIUM



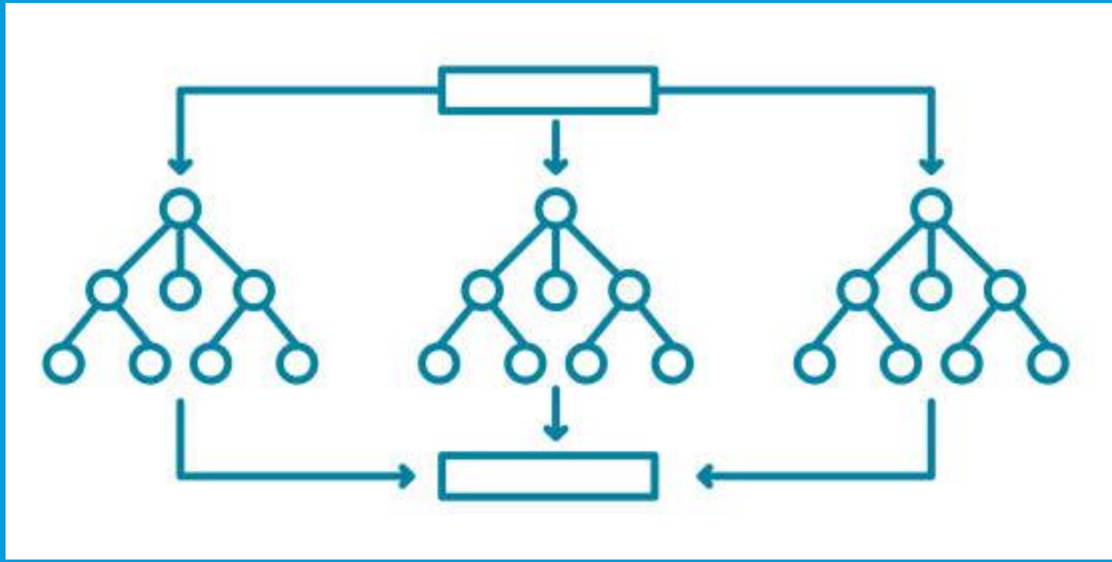
- A Folium map was created with circles for each launch site and markers indicating whether launches at that site were successful or failures
- Lines were added to the map indicating distance from coastline and nearest city

INTERACTIVE DASHBOARD USING PLOTLY DASH

- An interactive dashboard was built to examine the relationship between booster site, payload mass and failure or success of the launch

PREDICTIVE ANALYSIS (CLASSIFICATION)

- The dataset was evaluated using different machine learning models to find which model can most accurately predict the chance of a successful rocket landing
- Models examined included Support Vector Machine, K Nearest Neighbour, and Decision Tree classification models



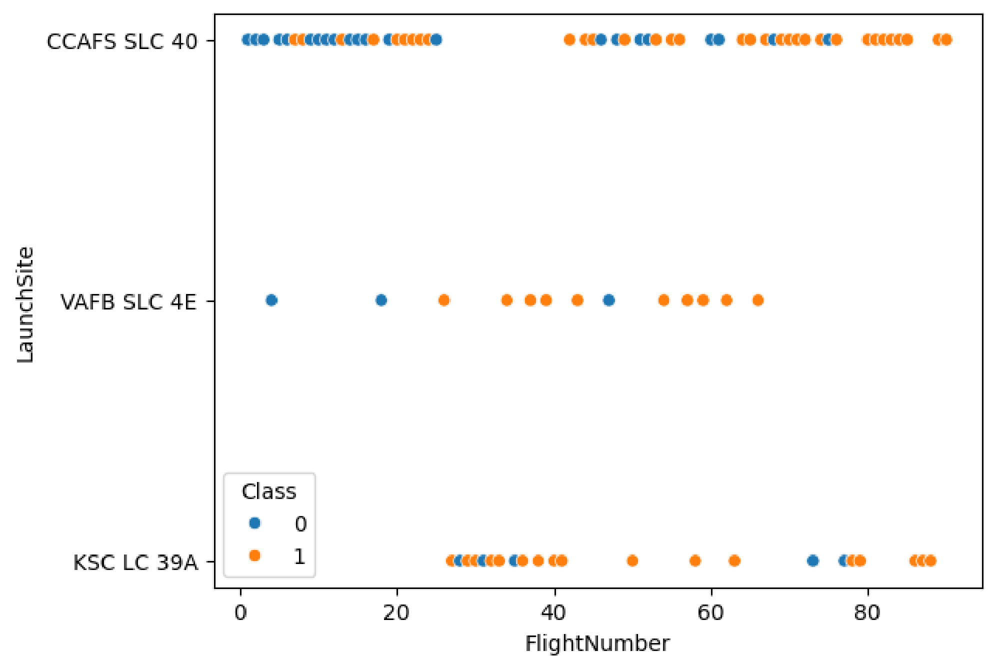
EDA DATA VISUALIZATIONS

- A review of all Data Visualizations created during the Exploratory Data Analysis phase

FLIGHT NUMBER VS. LAUNCH SITE

A scatter plot of Flight Number vs. Launch Site:

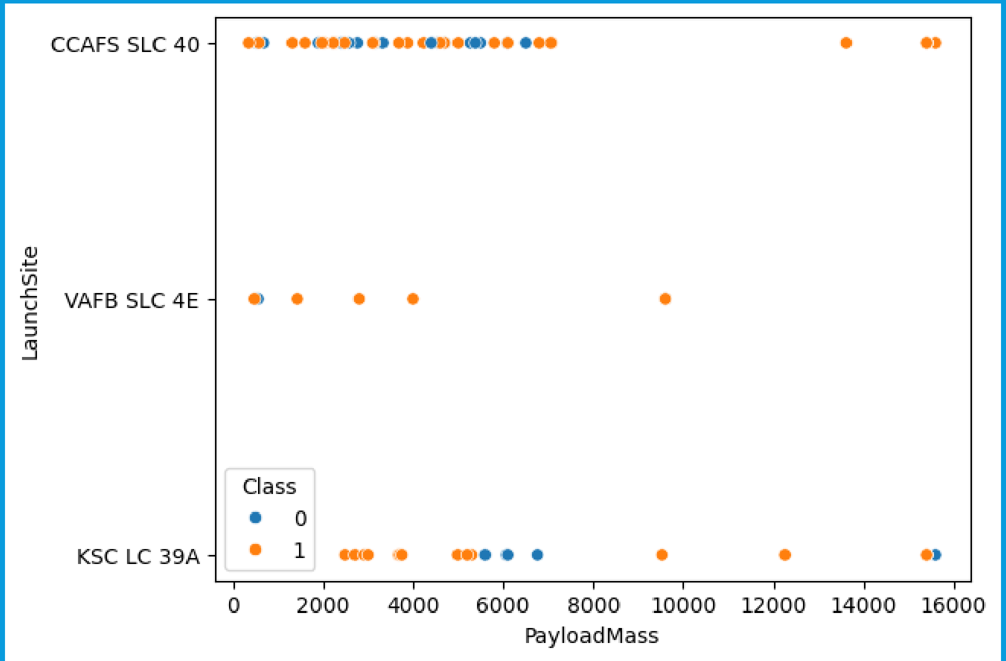
- We can see from this plot that KSC LC 39A has the highest number of successful launches.
- The launch site CCAFS SLC 40 was used the most, and VAFB SLC 4E has the fewest launches.



PAYLOAD VS. LAUNCH SITE

A scatter plot of Payload vs. Launch Site:

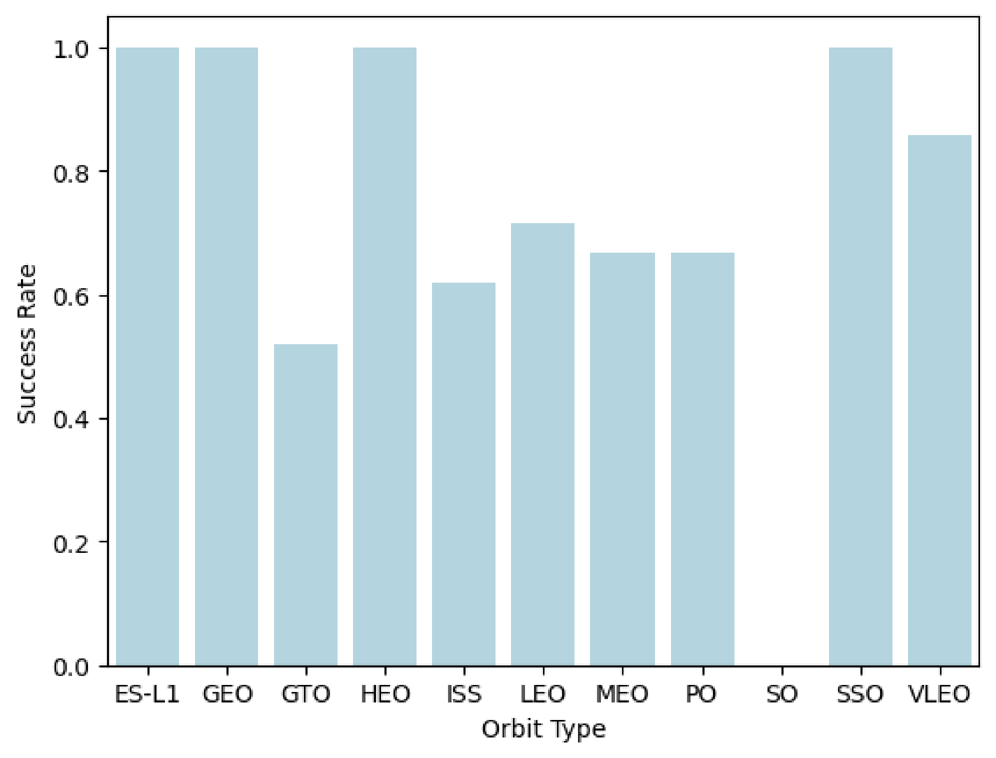
- There is a sweet spot for payload mass. Too heavy or too light and the landing will be unsuccessful.
- Heavy rockets have not been launched from VAFB SLC 4E.



SUCCESS RATE VS. ORBIT TYPE

A bar chart for the success rate of each orbit type:

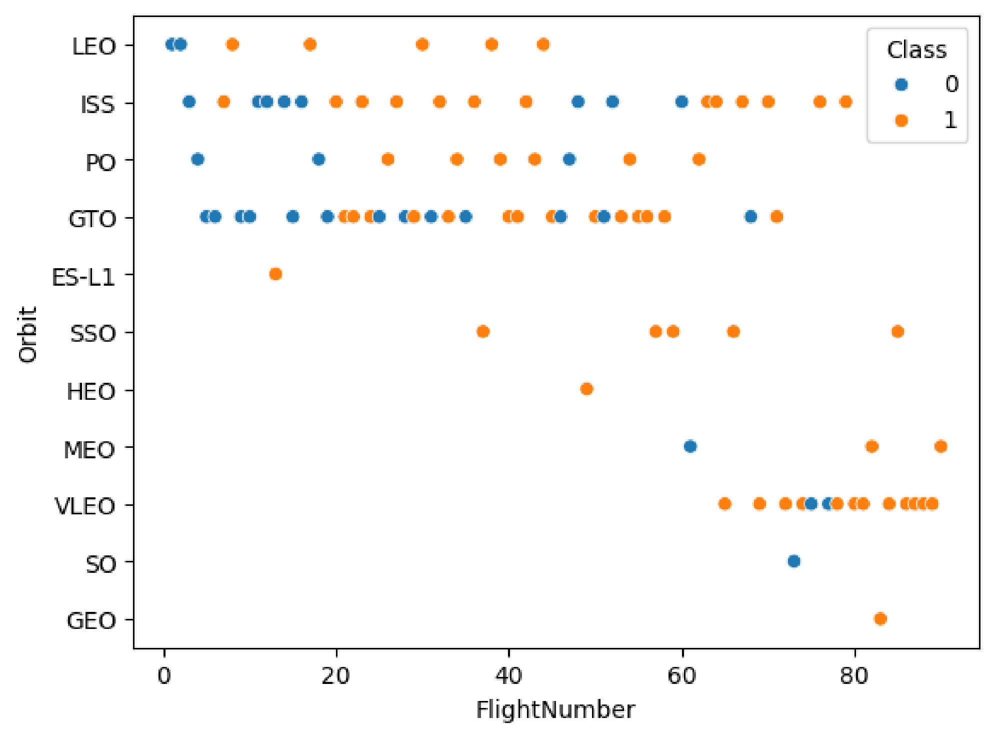
- ES-L1, GEO, HEO, and SSO orbit types have 100% success rate
- GTO Orbit type has the lowest success rate (approximately 50%)



FLIGHT NUMBER VS. ORBIT TYPE

A scatter plot of Flight number vs. Orbit type:

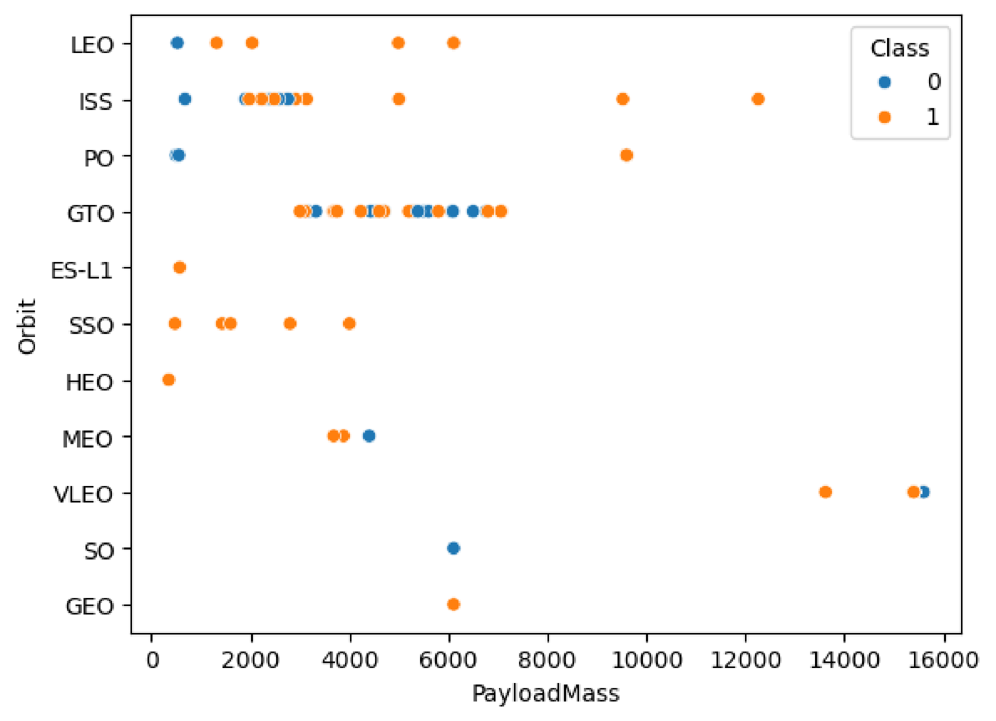
- More recent orbit types are more successful than earlier models
- GEO, HEO, SSO, and ES-L1 have 100% success rates
- Orbit type has changed over time



PAYLOAD VS. ORBIT TYPE

Scatter plot examining Payload compared to Orbit Type:

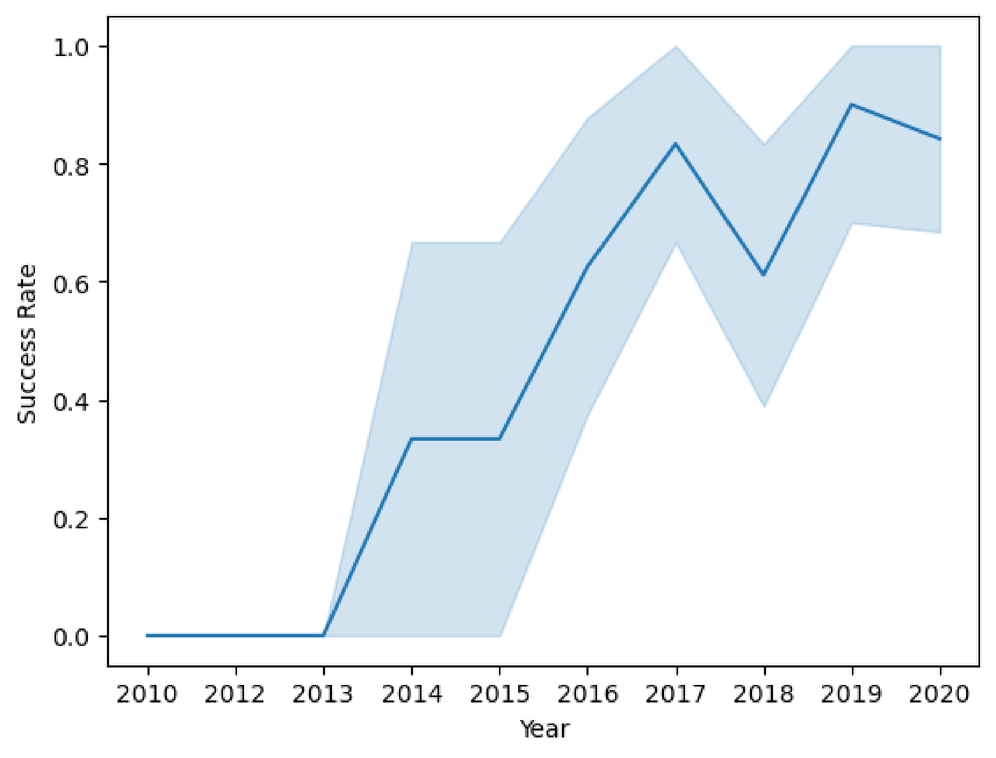
- With heavy payloads the chance of a successful landing is higher for Polar, LEO and ISS
- GTO success rate is not as easy to distinguish



LAUNCH SUCCESS YEARLY TREND

Line chart of yearly average success rate:

- Since 2013, the success rate of launches has consistently increased
- In 2018 there was a slight decline in success rate, but this improved in 2019
- There was more variability in success rate from 2013-2015



EDA USING SQL

A review of all SQL queries built during the Exploratory Data Analysis phase.

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

NAMES OF LAUNCH SITES

Selects all Launch Site names from the dataset

In [10]:
```
%sql SELECT * FROM SPACEXTABLE WHERE "LAUNCH_SITE" LIKE "CCA%" LIMIT 5
```

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

Out[10]:

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

LAUNCH SITE NAMES BEGIN WITH 'CCA'

Selects all Launch Sites from SpaceX data that begin with CCA

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

In [11]:
```
%sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "CUSTOMER" = "NASA (CRS)"
```

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

Out[11]:

SUM("PAYLOAD_MASS__KG_")
45596

TOTAL PAYLOAD MASS

Selects the sum total of Payload Mass for boosters launched by NASA

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

```
In [12]: %sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "BOOSTER_VERSION" LIKE "%F9 V1.1"

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

Out[12]:

AVG("PAYLOAD_MASS__KG_")
2928.4

AVERAGE PAYLOAD MASS BY F9 V1.1

Displays the average Payload Mass for booster version F9 v1.1

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

Hint:Use min function

In [13]:

```
%sql SELECT MIN("DATE") FROM SPACEXTABLE WHERE "MISSION_OUTCOME" LIKE "%SUCCESS"
```

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

Out[13]:

MIN("DATE")
2010-06-04

FIRST SUCCESSFUL GROUND LANDING DATE

Displays the date of the first successful landing on ground pad

In [14]:
```
%sql SELECT "BOOSTER_VERSION" FROM SPACEXTABLE WHERE "LANDING_OUTCOME" LIKE "%SUCCESS%" \
AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000
```

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

Out[14]:

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 B4 B1043.1
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD BETWEEN 4000 AND 6000

Displays the booster versions of successful landing outcomes with Payload between 4000kg and 6000kg

In [15]:

```
%sql SELECT (SELECT COUNT("MISSION_OUTCOME") FROM SPACEXTBL WHERE "MISSION_OUTCOME" LIKE '%Success%') AS SUCCESS, \
(SELECT COUNT("MISSION_OUTCOME") FROM SPACEXTBL WHERE "MISSION_OUTCOME" LIKE '%Failure%') AS FAILURE
```

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

Out[15]:

SUCCESS	FAILURE
100	1

TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

Displays the total number of "Successful" launch missions compared to "Failed" missions

In [16]:

```
%sql SELECT DISTINCT "BOOSTER_VERSION" FROM SPACEXTBL \
WHERE "PAYLOAD_MASS__KG_" = (SELECT max("PAYLOAD_MASS__KG_") FROM SPACEXTBL)
```

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

Out[16]:

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

BOOSTERS CARRIED MAXIMUM PAYLOAD

Displays the booster versions of launches that carried the highest Payload Mass

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, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

In [29]:

```
%sql SELECT substr("DATE", 6,2) AS MONTH, "BOOSTER_VERSION", "LAUNCH_SITE" FROM SPACEXTBL\
WHERE "LANDING_OUTCOME" LIKE '%Failure%' and substr("DATE",0,5) = '2015'
```

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

Out[29]:

MONTH	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

2015 LAUNCH RECORDS

Displays failed landing outcomes and booster versions and launch sites for drone ships in 2015

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 [26]:

```
%sql SELECT "LANDING_OUTCOME", COUNT("LANDING_OUTCOME") FROM SPACEXTBL\
WHERE "DATE" >= '2010-06-04' and "DATE" <= '2017-03-20' and "LANDING_OUTCOME" LIKE '%%'\
GROUP BY "LANDING_OUTCOME" \
ORDER BY COUNT("LANDING _OUTCOME") DESC ;
```

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

Out[26]:

Landing_Outcome	COUNT("LANDING_OUTCOME")
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

RANK LANDING OUTCOMES BETWEEN 2010 AND 2017

Counts landing outcomes 2010-06-04 to 2017-03-20 and ranks the outcomes in a descending order based on count of outcome

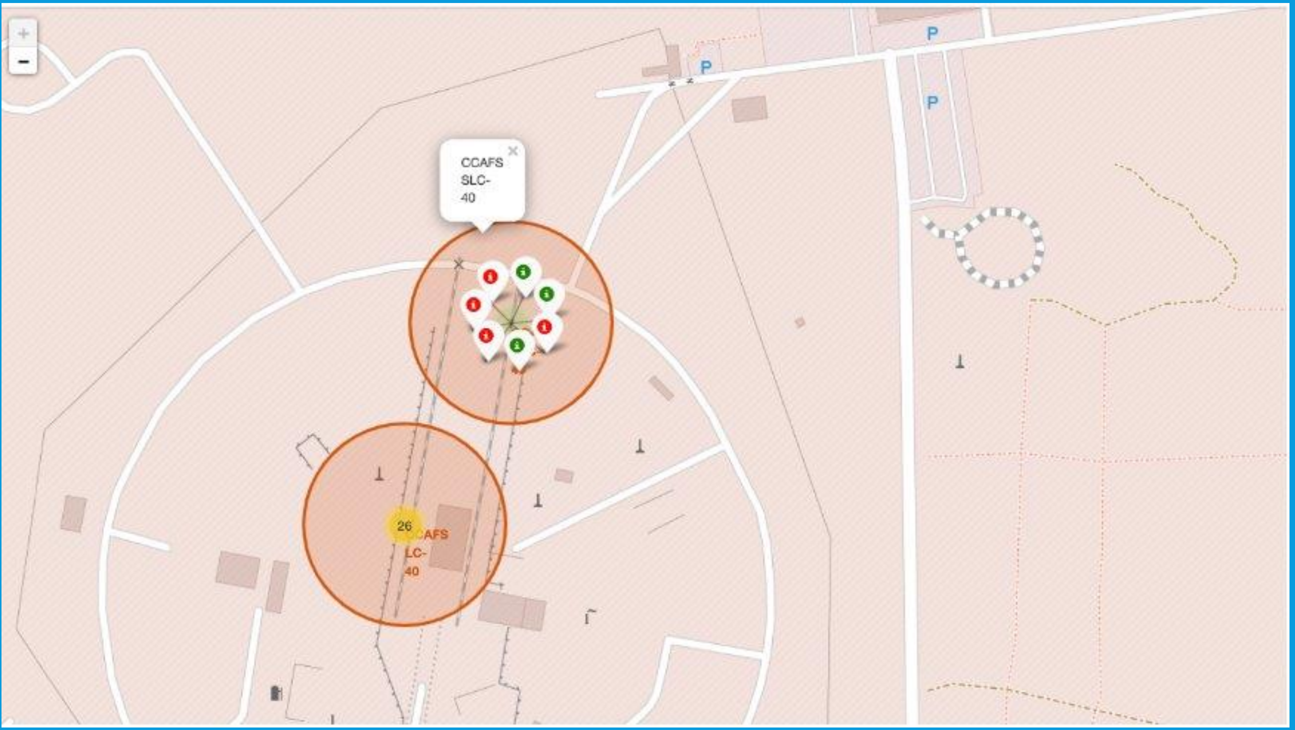
SQL QUERIES COMPILED

- **%sql** SELECT DISTINCT "LAUNCH_SITE" FROM SPACEXTABLE
- **%sql** SELECT * FROM SPACEXTABLE WHERE "LAUNCH_SITE" LIKE "CCA%" LIMIT 5
- **%sql** SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "CUSTOMER" = "NASA (CRS)"
- **%sql** SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "BOOSTER_VERSION" LIKE "%F9 V1.1"
- **%sql** SELECT MIN("DATE") FROM SPACEXTABLE WHERE "MISSION_OUTCOME" LIKE "%SUCCESS"
- **%sql** SELECT "BOOSTER_VERSION" FROM SPACEXTABLE WHERE "LANDING_OUTCOME" LIKE "%SUCCESS%" \ AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000
- **%sql** SELECT (SELECT COUNT("MISSION_OUTCOME") FROM SPACEXTBL WHERE "MISSION_OUTCOME" LIKE '%Success%') AS SUCCESS, (SELECT COUNT("MISSION_OUTCOME") FROM SPACEXTBL WHERE "MISSION_OUTCOME" LIKE '%Failure%') AS FAILURE
- **%sql** SELECT DISTINCT "BOOSTER_VERSION" FROM SPACEXTBL \ WHERE "PAYLOAD_MASS__KG_" = (SELECT max("PAYLOAD_MASS__KG_") FROM SPACEXTBL)

LAUNCH SITES PROXIMITY ANALYSIS

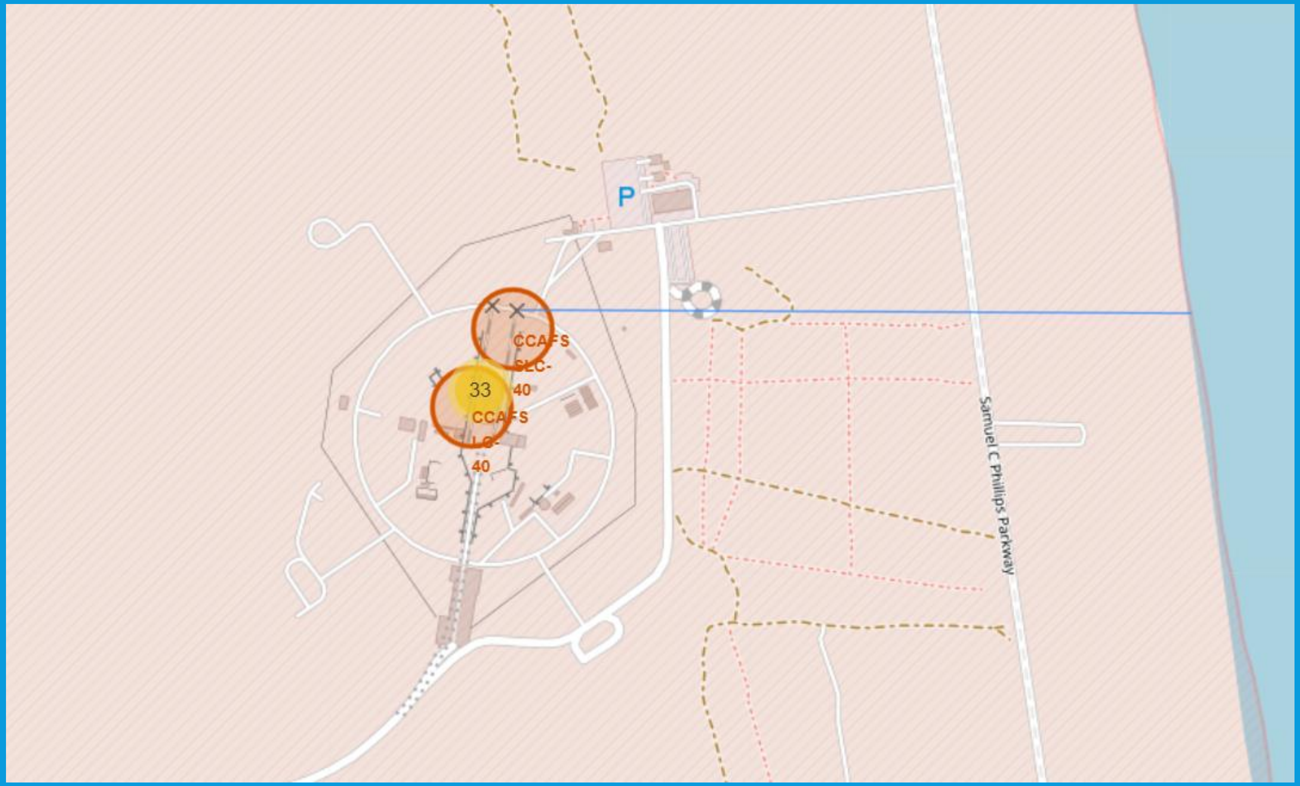
A review of interactive Folium maps displaying launch site and information such as success/failure of launches and proximity to geographic features such as coastline and nearest city.

SUCCESS AND FAILURE OF LAUNCHES



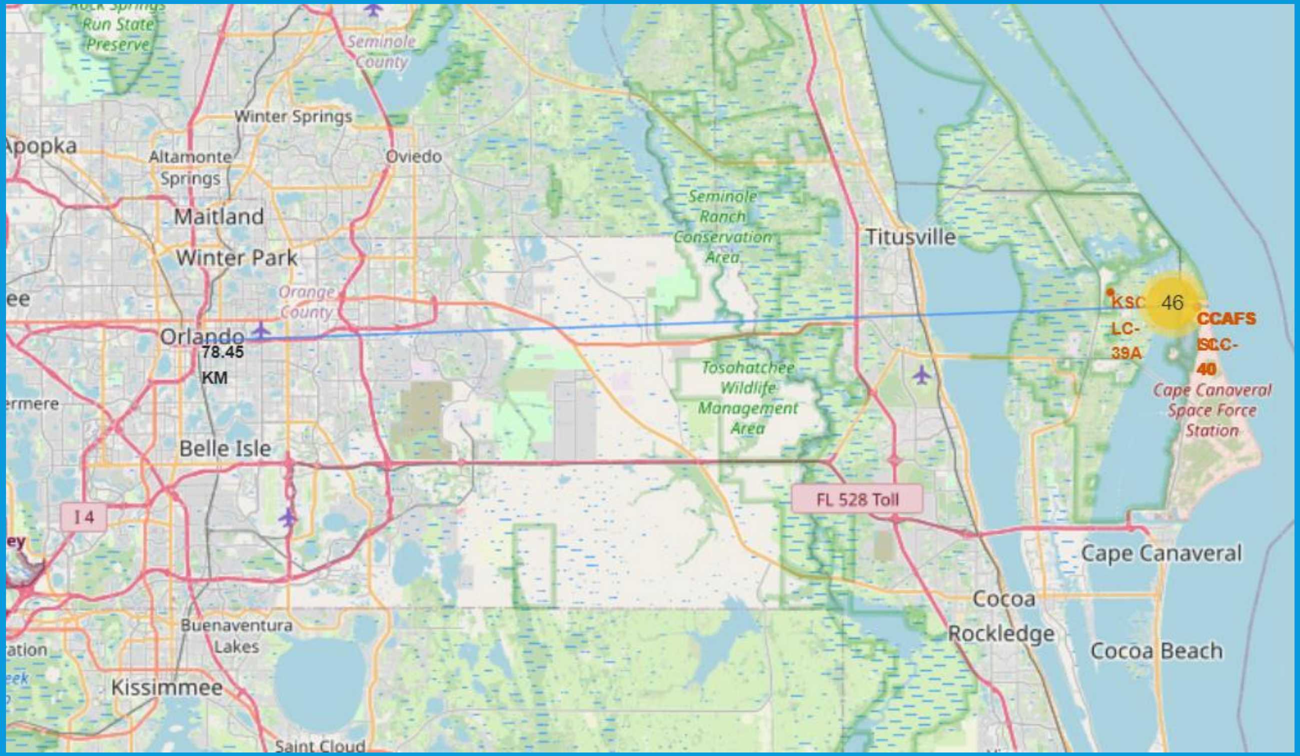
- This map displays launches from launch site CCAFS SLC-40
- Green markers indicate success
- Red markers indicate failure
- We can see this launch site has equal successes and failures

DISTANCE OF LAUNCH SITE FROM COASTLINE



- This map displays the distance from the coastline of launch sites CCAFS SLC 40 and CCAFS LC 40
- Both of these launch sites are quite close to the coastline

DISTANCE OF LAUNCH SITE FROM NEAREST CITY

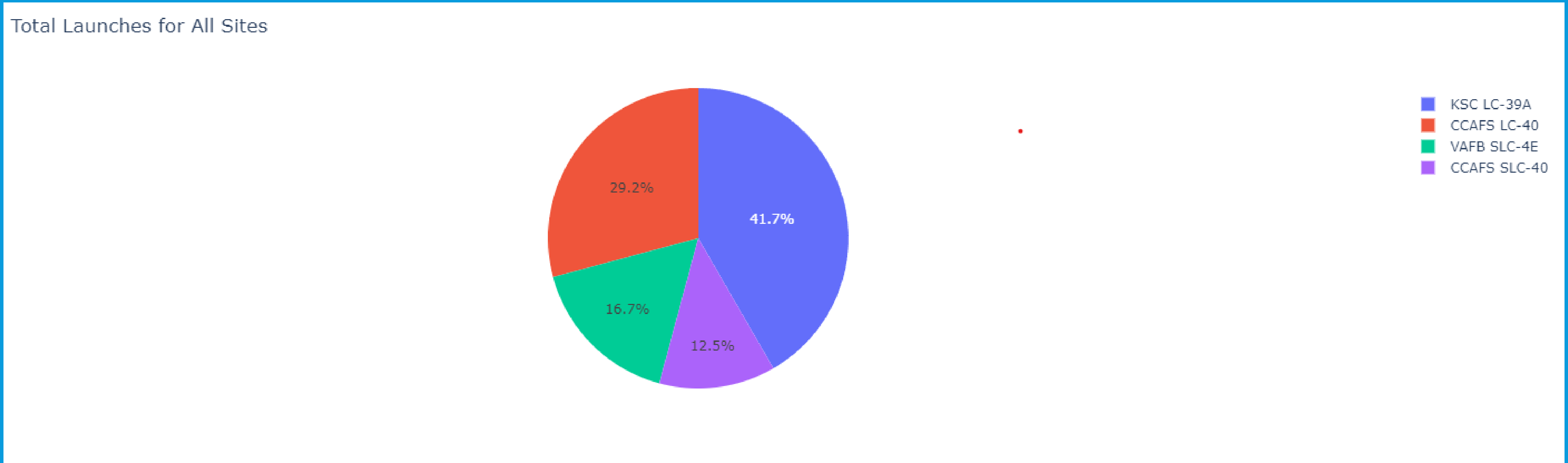


- This map shows the distance from launch sites on the East Coast to the nearest city (Orlando)
- We can see that the launch sites are roughly 80km from the city centre

LAUNCH ANALYSIS DASHBOARD

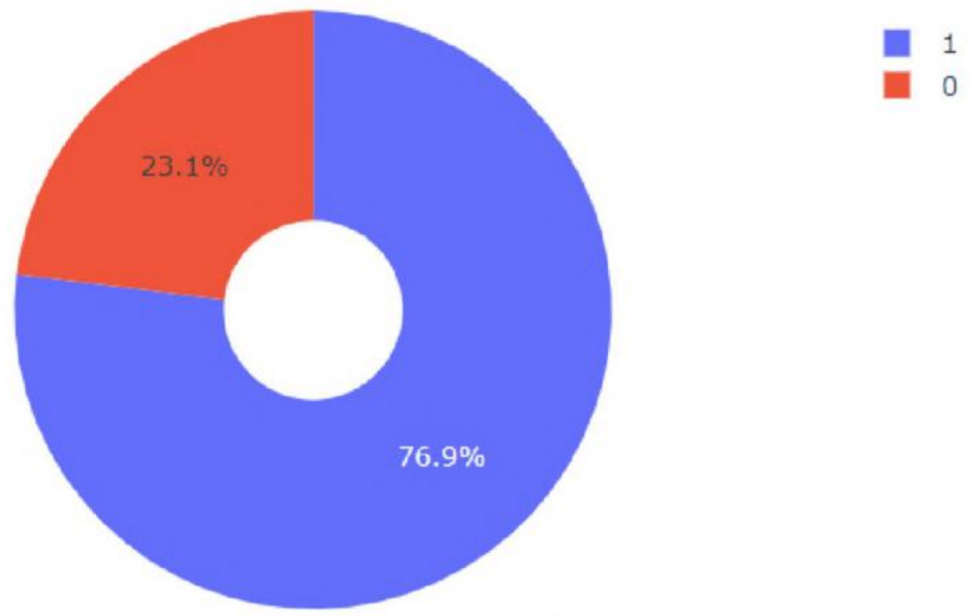
Review of interactive dashboard built using Plotly Dash which can be manipulated to explore relationship between Launch Site or Payload Mass and success rate.

DASHBOARD SCREENSHOT – TOTAL LAUNCHES



- From this Dashboard screenshot we can see the breakdown of number of launches by Launch Site. The KSC LC-39A Launch Site has almost half of the total launches.

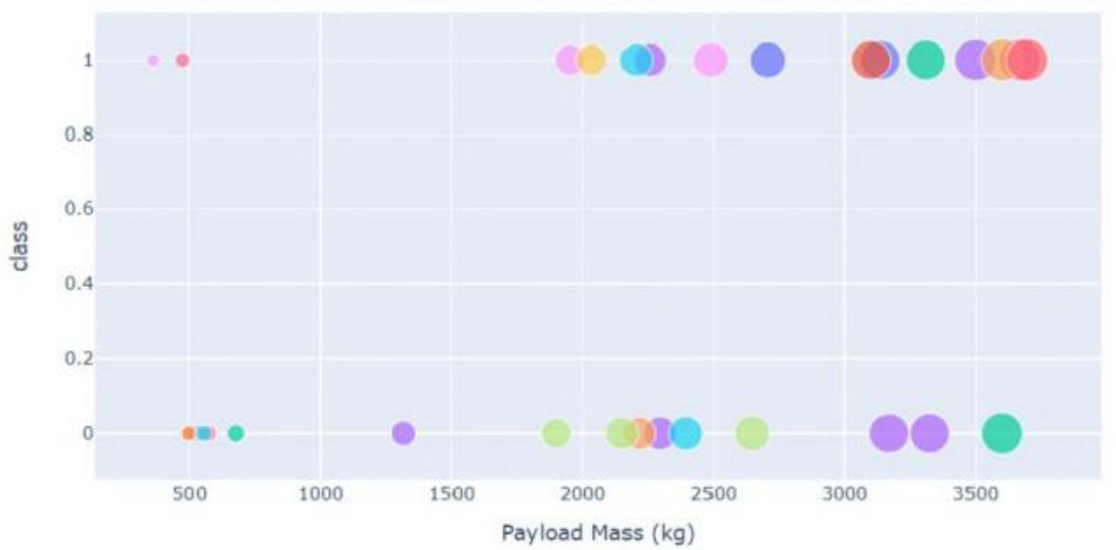
DASHBOARD SCREENSHOT – KSC LC-39A LAUNCHES

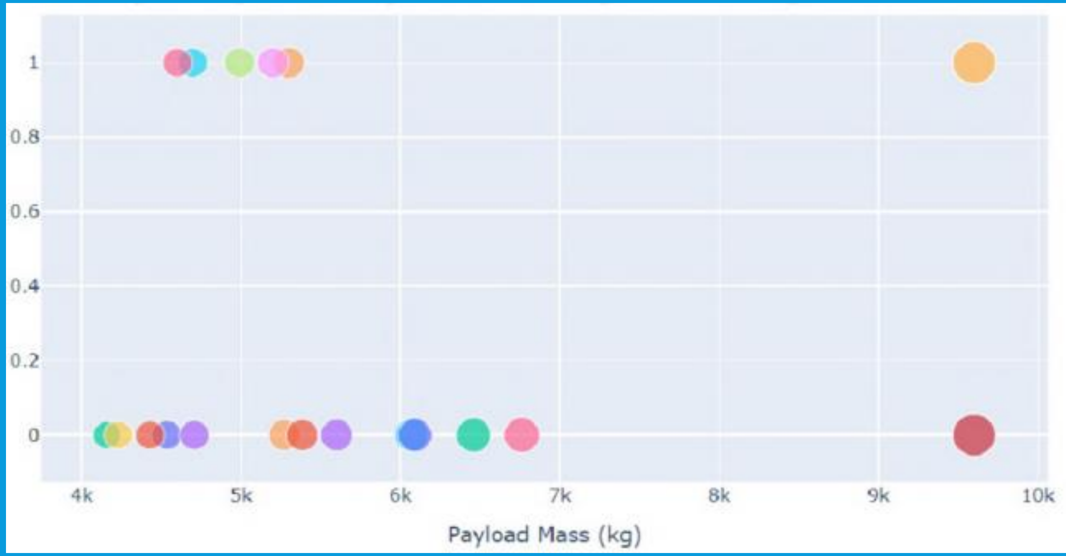


KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

- This chart shows the results of launches from KSC LC-39A Launch Site
- We can see that this site had a 76.9% success rate, the highest among all launch sites

DASHBOARD SCREENSHOT – KSC LC-39A LAUNCHES





When the Payload Mass is too low, the launches are prone to failure

When the Payload Mass is too high, the launches are also prone to failure

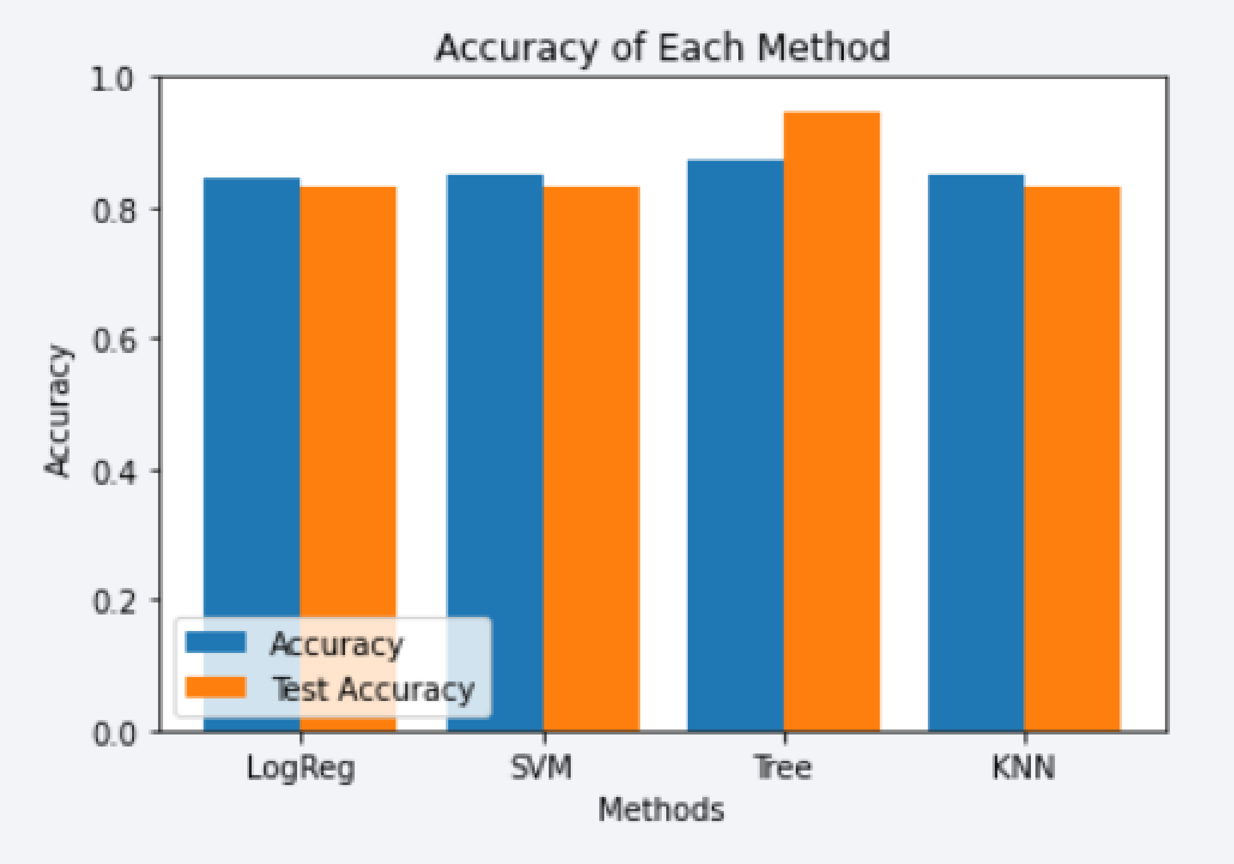
A Payload Mass between 2000kg and 5000kg appears to result in the highest chance of success

43

PREDICTIVE ANALYSIS (CLASSIFICATION)

A review of the machine learning models explored for predicting the success or failure of future launches.

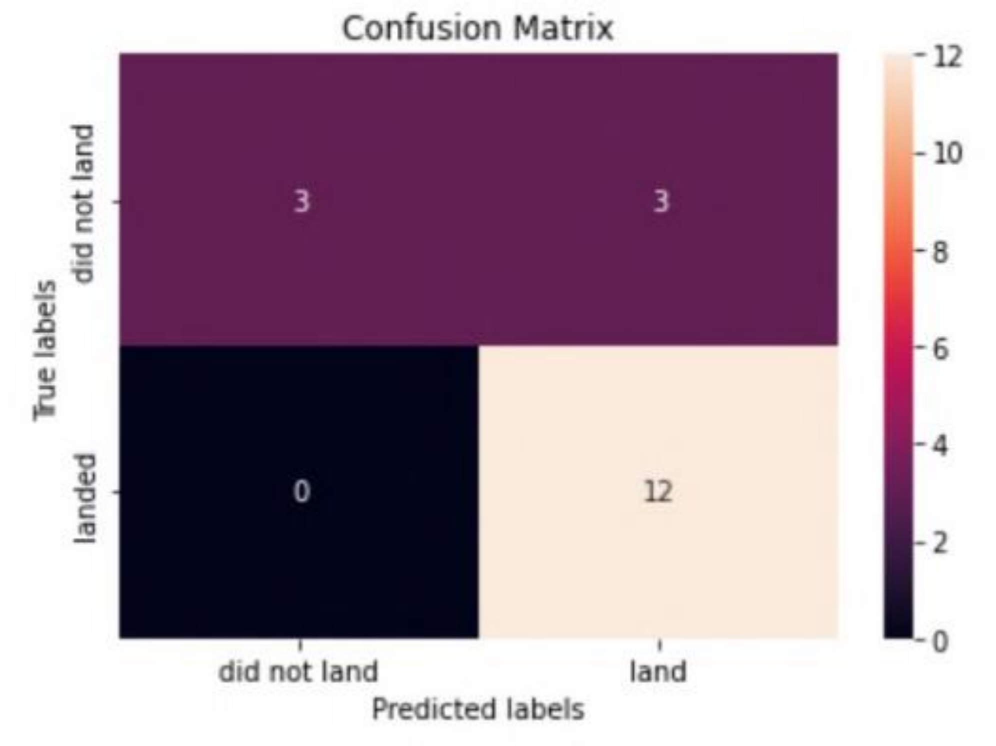
CLASSIFICATION ACCURACY



- The bar chart to the left clearly shows that a Decision Tree Model results in the highest accuracy in regards to predicting the successful or failure of future launches

CONFUSION MATRIX

- The Confusion Matrix to the right indicates that while a Decision Tree Model is 100% accurate when predicting landed rockets
- The model is slightly less accurate when predicting that rockets will not land successfully



CONCLUSIONS

- Chances of a successful launch have increased over time
- KSC LC 39A is the most successful launch site with 76.9% success rate
- Medium payloads (2000-5000kg) succeed more often than small or large payloads
- ES-L1, GEO, HEO, and SSO orbit types have 100% success rate
- The most accurate model for predicting the success of future launches is a Decision Tree Model
- The Decision Tree model is extremely accurate at predicting whether a rocket will successfully land

APPENDIX

All relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, and data sets created during this project can be found in the following Github Repository:

https://github.com/Jknapp01/IBM-Data-Science-Capstone-SpaceX

A special thanks goes out to everybody who put effort into building the courses for the IBM Data Science Professional Certificate.

I've learned so much through these 10 courses and I'm extremely grateful to all of you.

Thankyou from the bottom of my heart for creating Data Science content that
was interesting and engaging.

THANKYOU!