IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Dhanujie Jayapala

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- Summary of methodologies

In this exercise team used multiple methodologies to acquire and analyze the causal relationship of successful rocket landings by Space X.

- **Collect** : Data collected using SpaceX REST API and web scraping techniques acquired data from Wikipedia
- **Wrangle** : Convert raw data to create usable outcome variable
- **Explore** : Data visualization techniques to explore trends, considering factors like payload, launch site, yearly trends.
- **Analyze** : Analyzing the data with SQLlite
- **Geographic Mapping** : Geographically visualize the launch site success rates and proximity to geographical locations using folium
- **Dashboard** : An interactive dashboard website launch sites with the most success and successful payload ranges
- **Build Models** : Machine Learning to predict landing outcomes using logistic regression, support vector machine (SVM), decision tree and K-nearest neighbor (KNN)

Executive Summary

- Results can be Summarized into three buckets.
  - Exploratory Analysis
    - Success of missions has increase with time
    - We can observe a 100% success rate for ES-L1, GEO , HEO and SSO orbit types
    - There were serval drone ship landings
  - Visualization
    - Highest performing payload mass is between 2,000 and 6,000 kg
    - Space X has chosen sites near the equator. Other factors such as close to a main road and railway was also import
    - Being close to a coast makes some room for failed landings
    - Most successful site is KSC LC-39A with a success rate of 77
  - Predictive Analytics
    - Decision Tree Classifier has the best learning algorithm for the data available

Introduction

- SpaceX, officially known as Space Exploration Technologies Corp., is a revolutionary American aerospace company founded by Elon Musk in 2002.
- **Key Accomplishments:**
  - **Reusable Rockets:** SpaceX has successfully developed and implemented reusable rockets like the Falcon 9 and Falcon Heavy, significantly reducing launch costs.
  - **Human Spaceflight:** SpaceX achieved the milestone of sending astronauts to the International Space Station (ISS) with its Dragon spacecraft, marking the first time a private company has transported humans to orbit.
- **Impact on the Space Industry:**
  - **Reduced Launch Costs:** SpaceX's reusable rockets have dramatically lowered the cost of accessing space, opening up new possibilities for commercial and scientific endeavors.
  - **Increased Innovation:** The company's focus on innovation and pushing technological boundaries has spurred advancements in various areas of space technology.
- **Future Outlook:**
- SpaceX continues to drive the future of space exploration with its ambitious projects like Starship and Starlink. The company's innovative approach and commitment to reducing costs have positioned it as a major player in the global space industry, shaping the future of space travel and exploration.

Objectives of the Analysis

- Understand key considerations for the success of Space X
- Identify factors contributed to the success of the Space X program
- Learn improvement patterns from the data available on Space X
- Replicate the findings for better success for the in Strat up Space Y

Section 1

Methodology

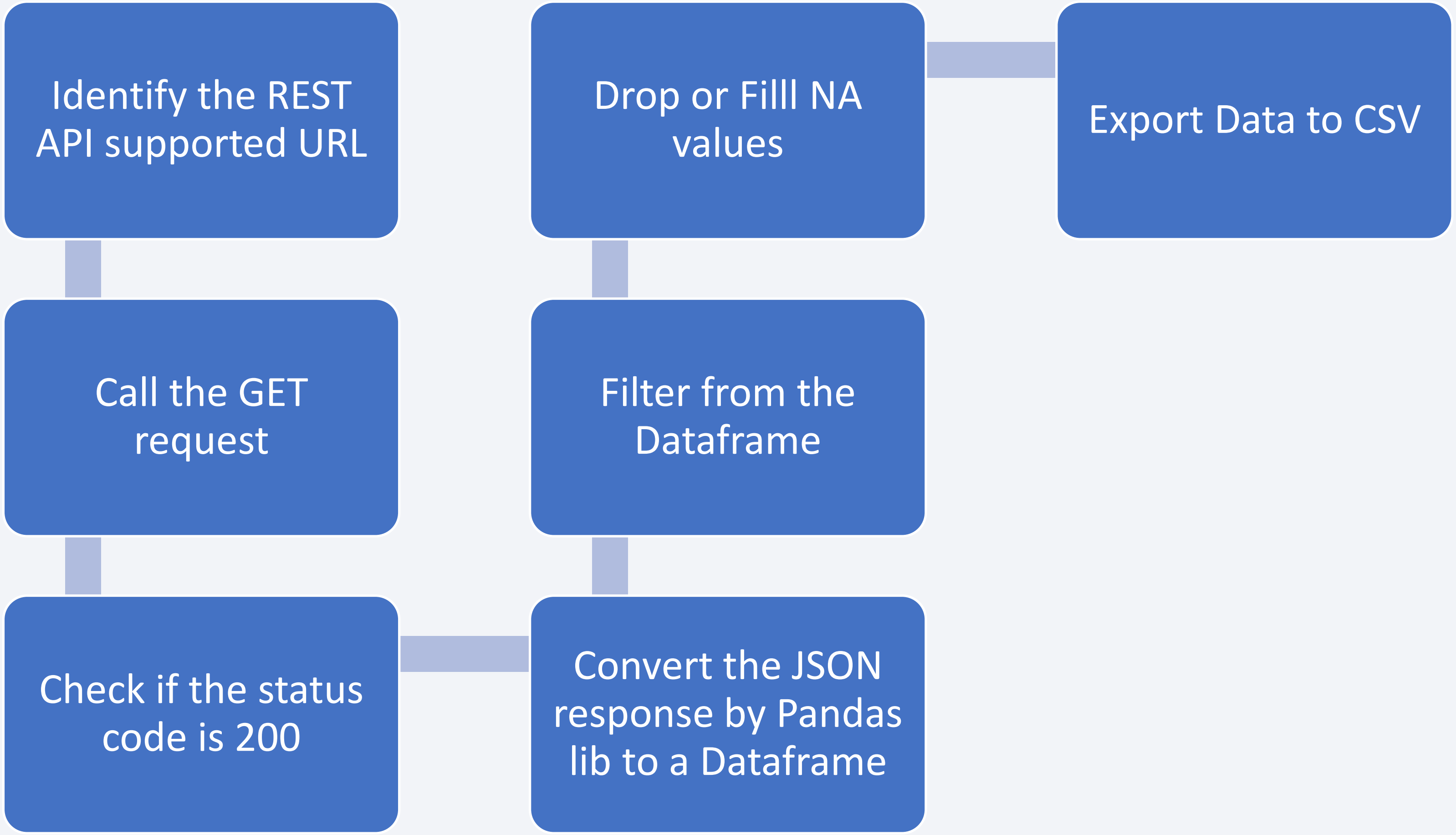
Methodology

Executive Summary

- Data collection methodology:
  - Data were collecting mainly using two methodologies, which are web-scarping through Wikipedia and using SpaceX REST API platforms.
- Perform data wrangling
  - Data preparation or wrangling was done by filtering the dataset, removing missing values and doing one hot encoding.
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Thr
- Perform interactive visual analytics using Folium and Plotly Dash

- Perform predictive analysis using classification models

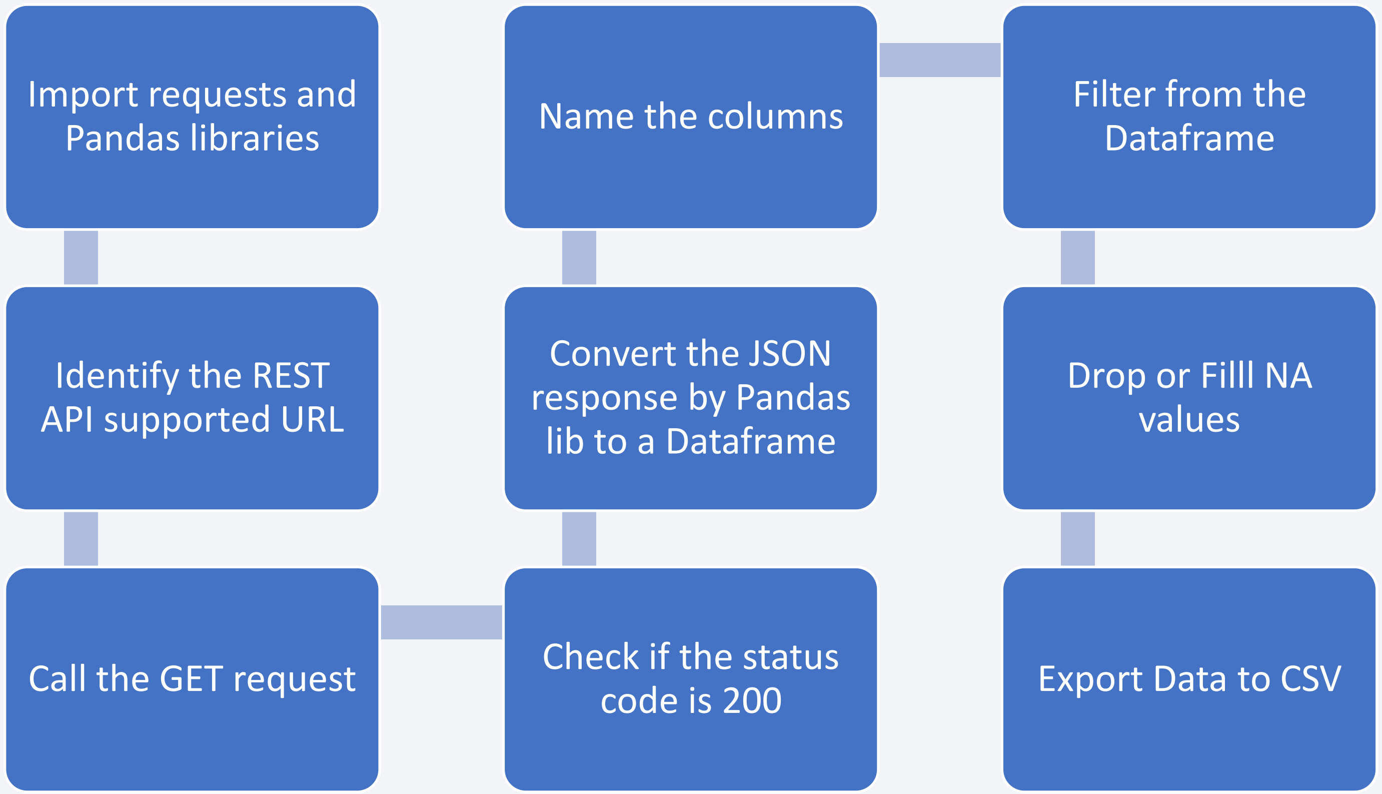
Data Collection



Data Collection – SpaceX API

- Key Code lines

```
# Requests allows us to make HTTP requests which we will use to get data from an API
import requests
# Pandas is a software library written for the Python programming language for data manipulation and analysis.
import pandas as pd
response = requests.get(spacex_url)
data = pd.json_normalize(response.json())
data.head()
```

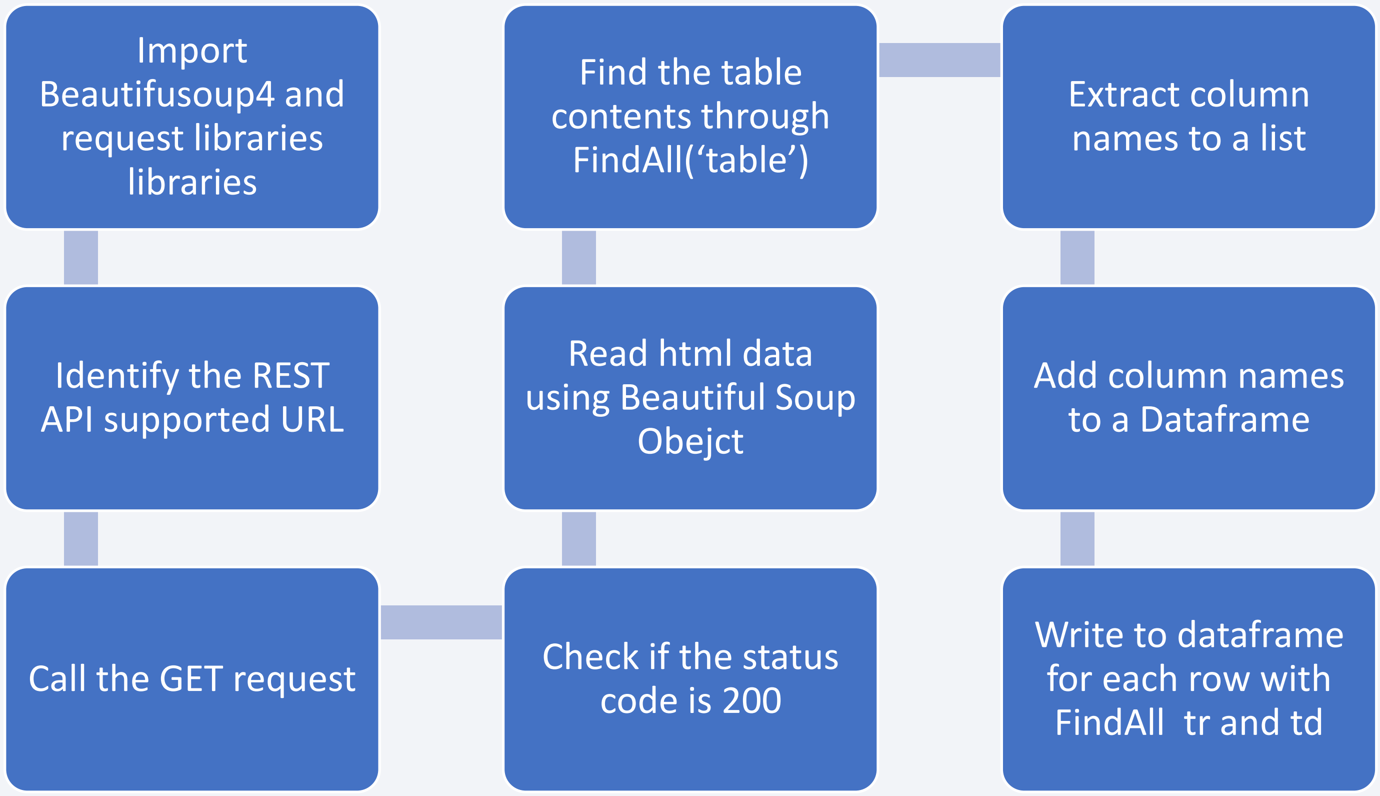


Full Code Can be found in here:

Data Collection - Scraping

- Key lines of Code

```
import requests
from bs4 import BeautifulSoup
import pandas as pd
html_data =
requests.get(static_url)
html_data.status_code
soup =
BeautifulSoup(html_data.text)
html_tables =
soup.find_all('table’)
```



Full code can be found here

Data Wrangling

- Key lines of Code

```
import pandas as pd

import numpy as np

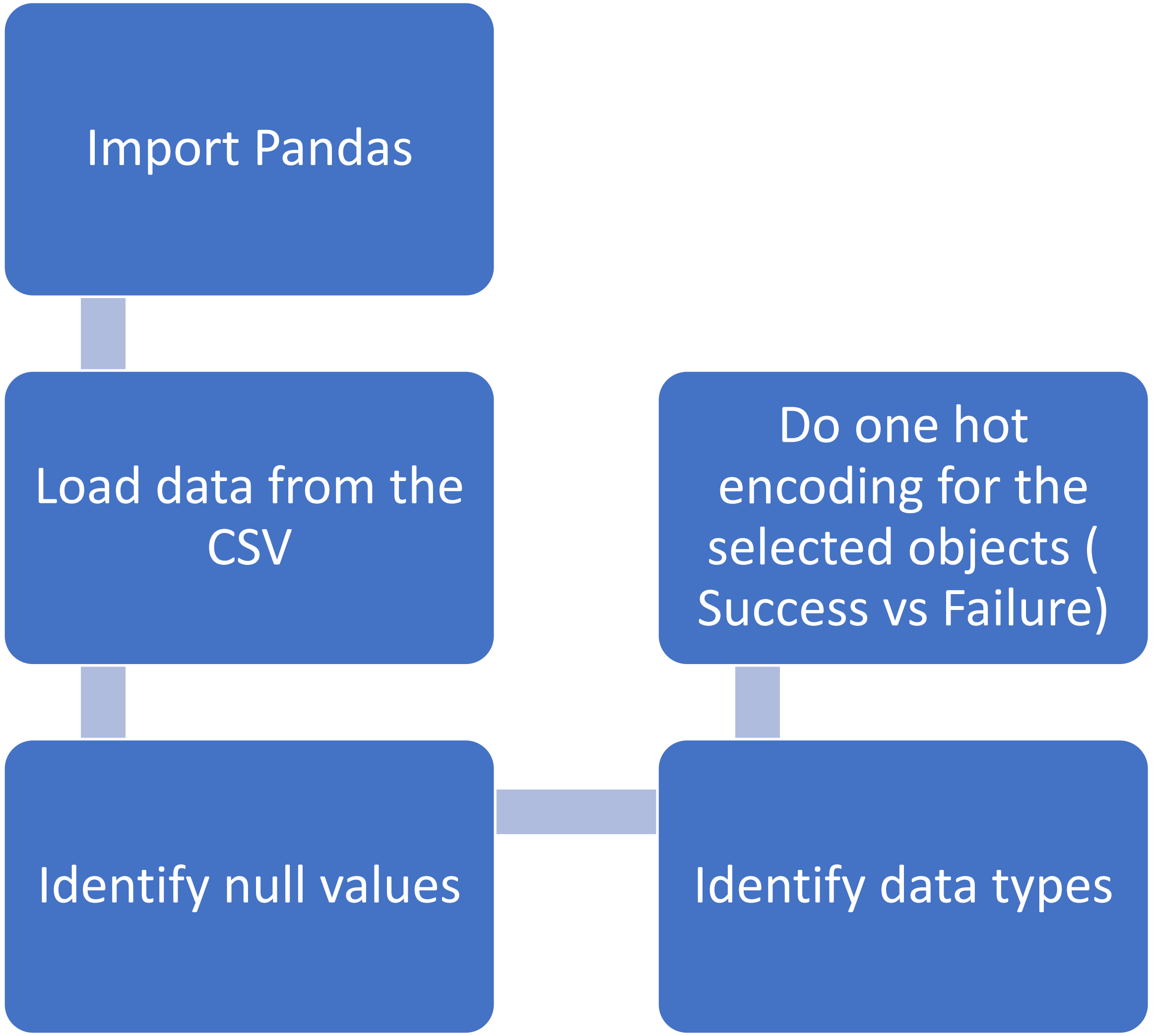
df=pd.read_csv(csv_file_path)

df.isnull().sum()/len(df)*100


good_outcomes = set(landing_outcomes.keys()).difference(bad_outcomes)

landing_class = df['Outcome'].replace(good_outcomes, 1)

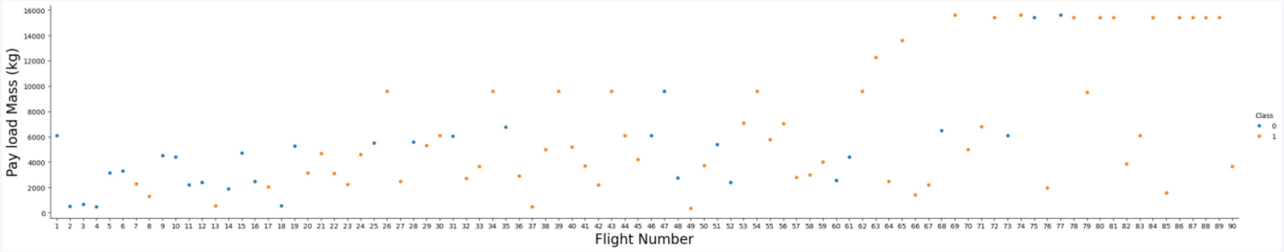
landing_class  = pd.DataFrame(landing_class).replace(bad_outcomes, 0)
```

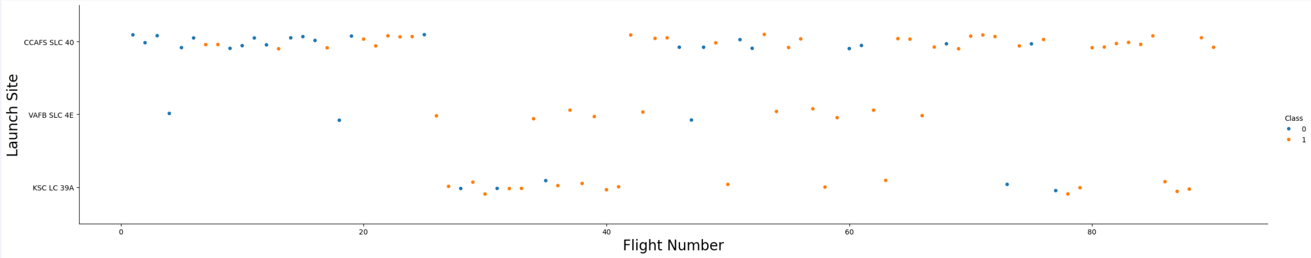


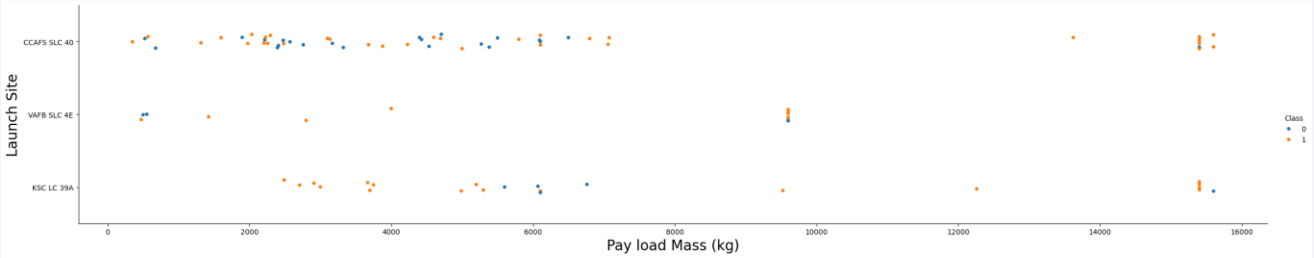
Full code can be found here

EDA with Data Visualization

- Charts plotted and why
  - Flight Number vs Payload Mass with Success in Blue : To see how payload mass increased over time and success rate increased over time
  - Flight No vs Launch Site : What are more successful launch sites and what is used nowadays
  - Payload Mass vs Launch Site : To see if Launch site supports certain Payloads





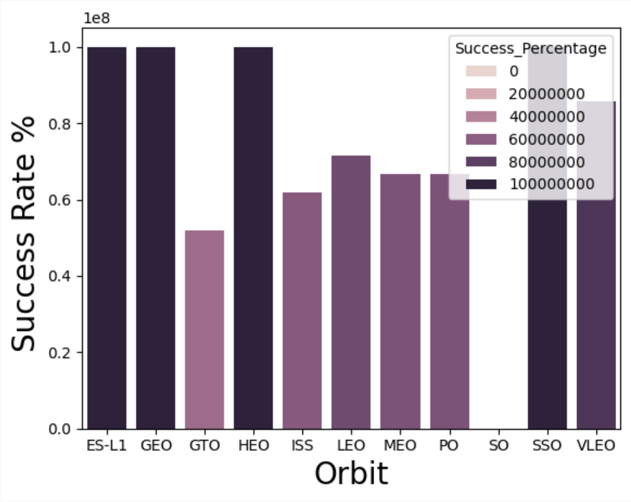


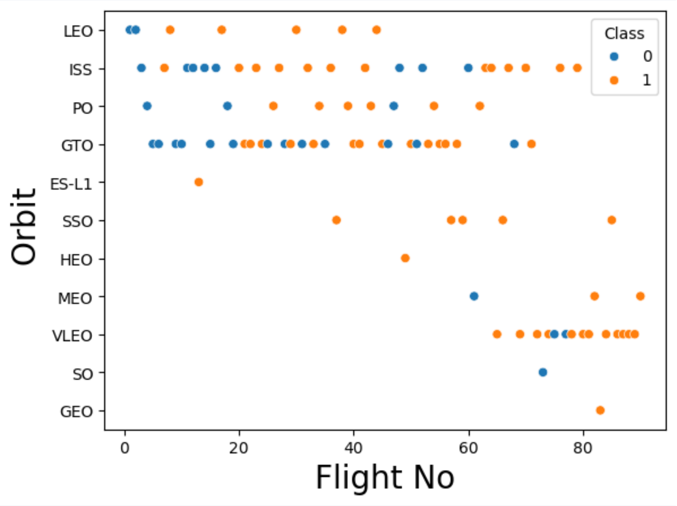
- Full workbook link :

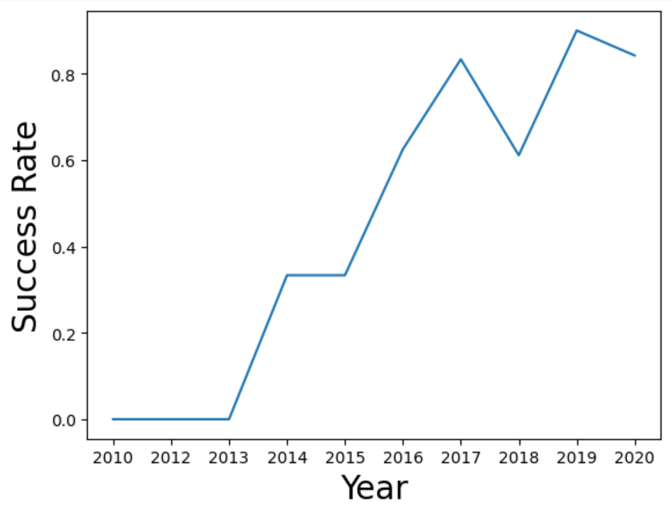
EDA with Data Visualization

- Charts plotted and why
  - Orbit Type and Success Rates : To see any relations ship to see if there are any orbits that causes mission failures
  - Payload Mass vs Orbit Type : Any specific orbits types require a certain payload mass
  - Year vs Success Rate: How identify whether the success rate has improved over the years

- Full workbook link :







EDA with SQL

- Using bullet point format, summarize the SQL queries you performed

- `%sql create table SPACEXTABLE as select * from SPACEXTBL where Date is not null`
- `%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTBL`
- `%sql SELECT * FROM SPACEXTBL WHERE "Launch_Site"  LIKE "CCA%" LIMIT 5`
- `%sql DESCRIBE SPACEXTBL`
- `%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE "Booster_Version" LIKE "F9 v1.1%"`
- `sql SELECT MIN(DATE) FROM SPACEXTBL WHERE "Mission_Outcome" IS "Success"`

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
  - `%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" BETWEEN 4000 AND 6000`
  - `%sql SELECT COUNT("Mission_Outcome") FROM SPACEXTBL WHERE "Mission_Outcome" IS "Success" OR "Success (payload status unclear)"`
  - `sql SELECT COUNT ("Mission_Outcome"),"Mission_Outcome" FROM SPACEXTBL GROUP BY "Mission_Outcome"`
  - `%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" IS ( SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL)`
  - `%sql SELECT substr("Date",6,2) as "Month" , substr("Date",0,5) as Year , "Landing_Outcome", "Booster_version", "Launch_site" FROM SPACEXTBL WHERE "Landing_Outcome" IS "Failure (drone ship)" AND  SUBSTR("Date",0,5) IS "2015"`
  - `%sql SELECT "Mission_Outcome", COUNT("Mission_Outcome") as "Events" FROM SPACEXTBL WHERE "Date" BETWEEN "2010-06-04" AND "2017-03-20" GROUP BY "Mission_Outcome"  ORDER BY "Events" DESC`
- Full code is here :

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

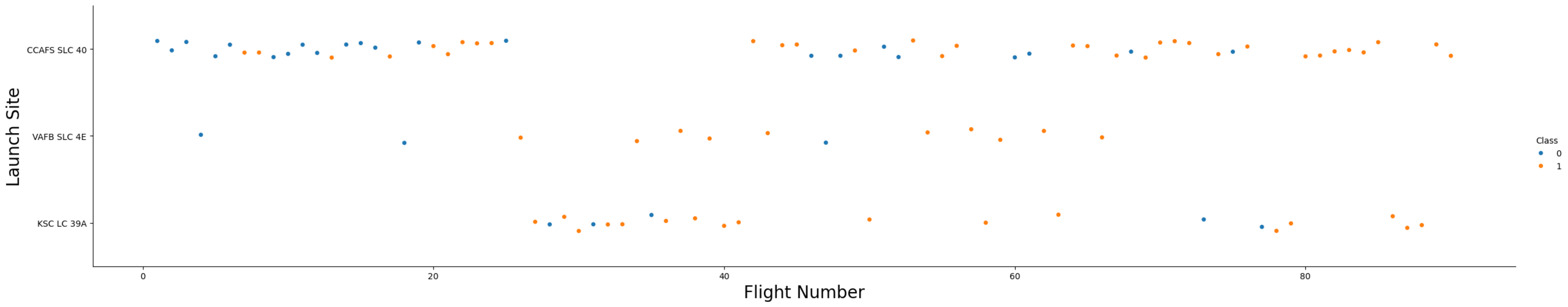
- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- Class 0 indicates failures and Class 1 indicates success.
- We can observe that there is a higher failure rate in the first few years and the failure rate as reduced over the years.
- Also, Initially the site they operated was mostly from CCAF5 5LC 40 and later moved to K5C LC 39A which has a higher success rate.

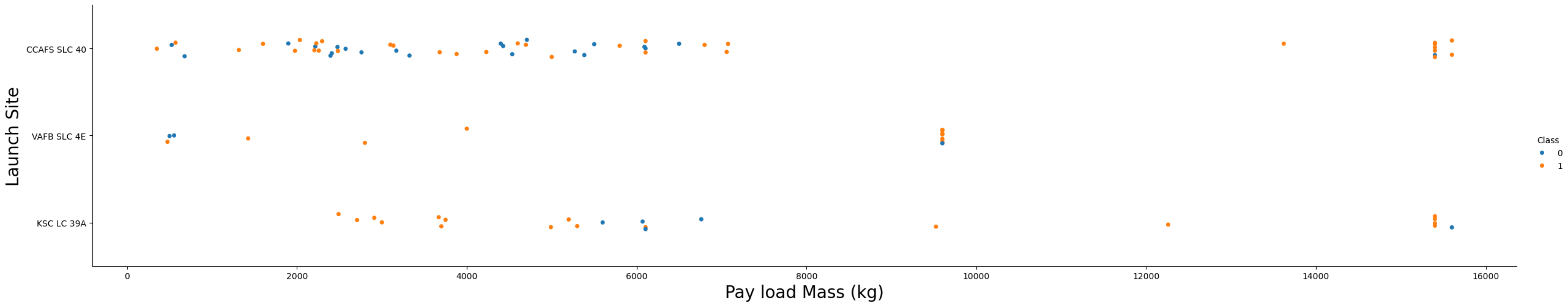


Payload vs. Launch Site

We can see the Falcon X team has done very high payloads (~ 15,000 kg) with a high success rate.

Most launches around 7000 kg were successful.

VAFB SL 4E site was not used for launches about 10,000 kg

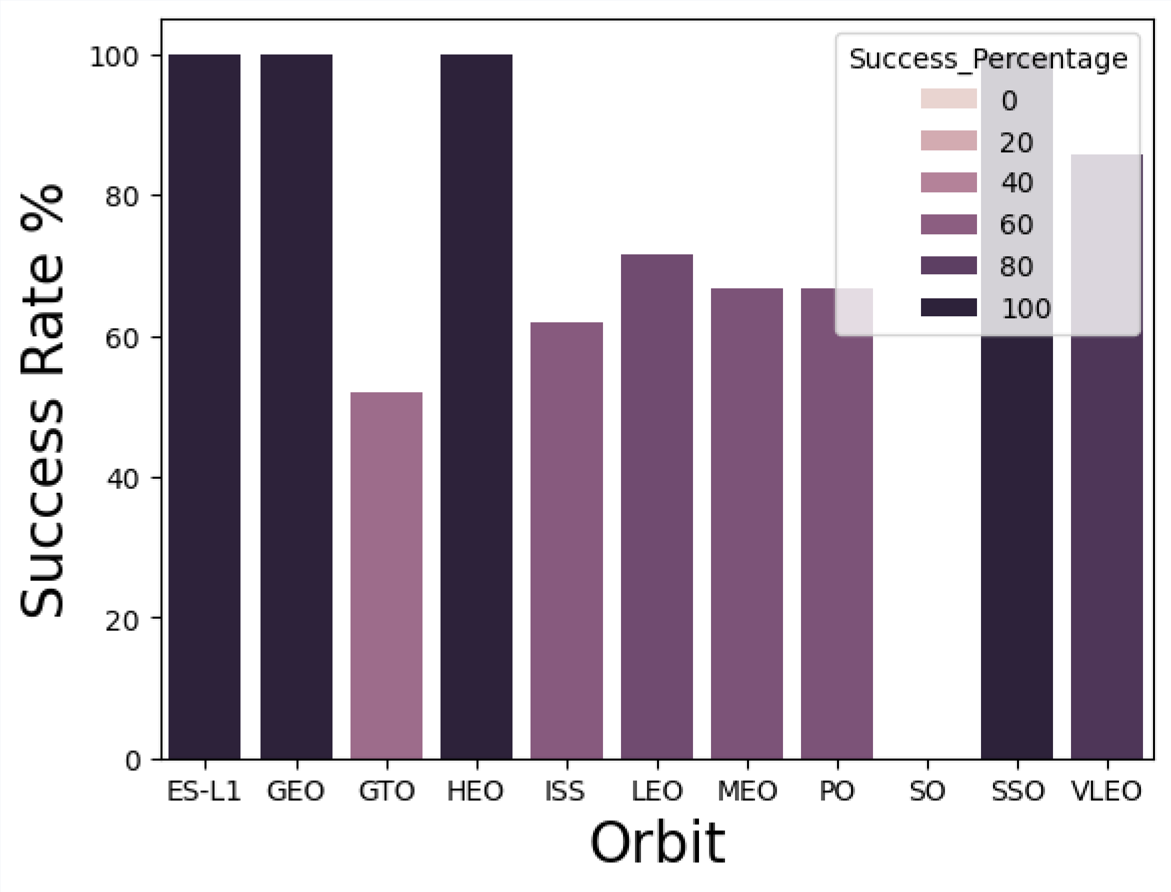


Success Rate vs. Orbit Type

- We can observe a 100% success rate for ES-L1, GEO , HEO and SSO orbit types

- SO has a 0% success rate.

- GTO, ISS , MEO, PO, LEO and VLEO has an increasing success rates starting from 50% to 90% in consecutive order.

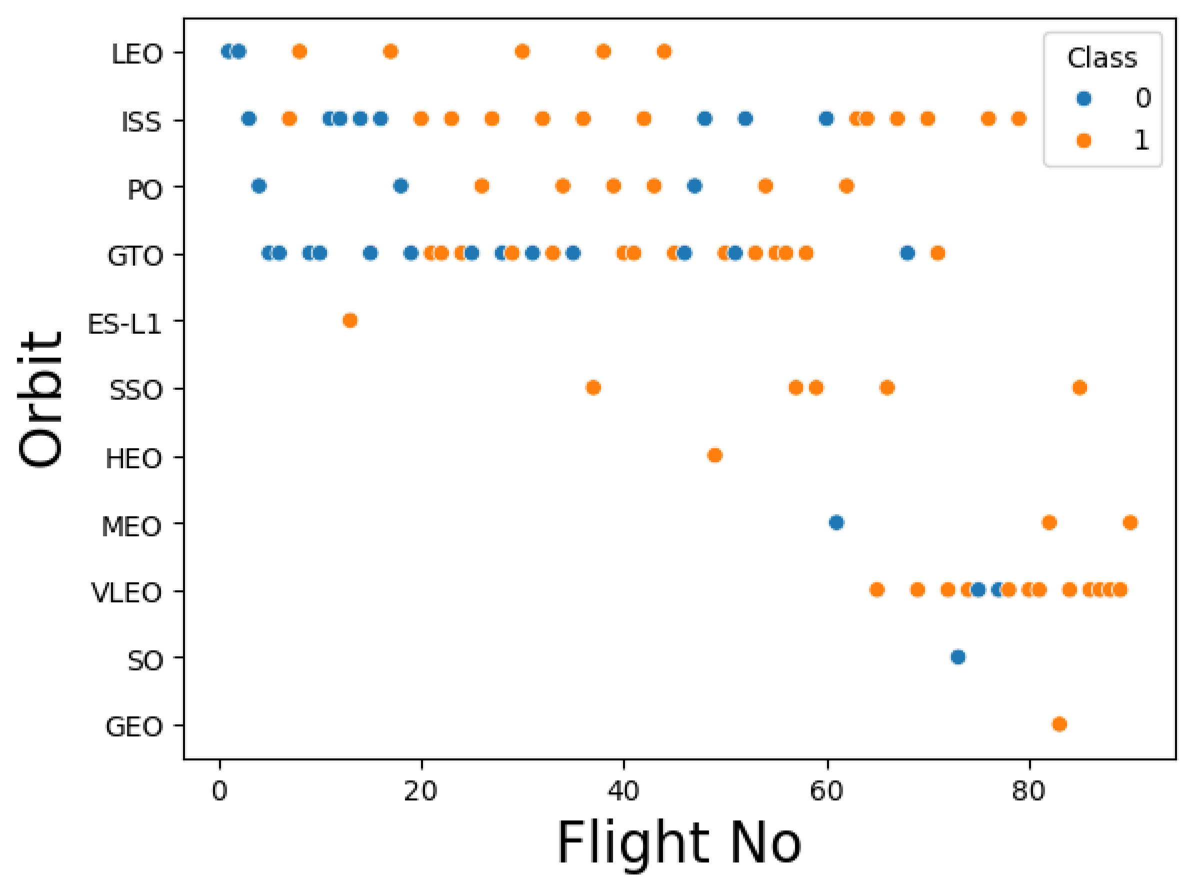


Flight Number vs. Orbit Type

Most initial launches are made from LEO, ISS, PO and GTO orbits.

For more recent flights they have preferred VLEO orbit.

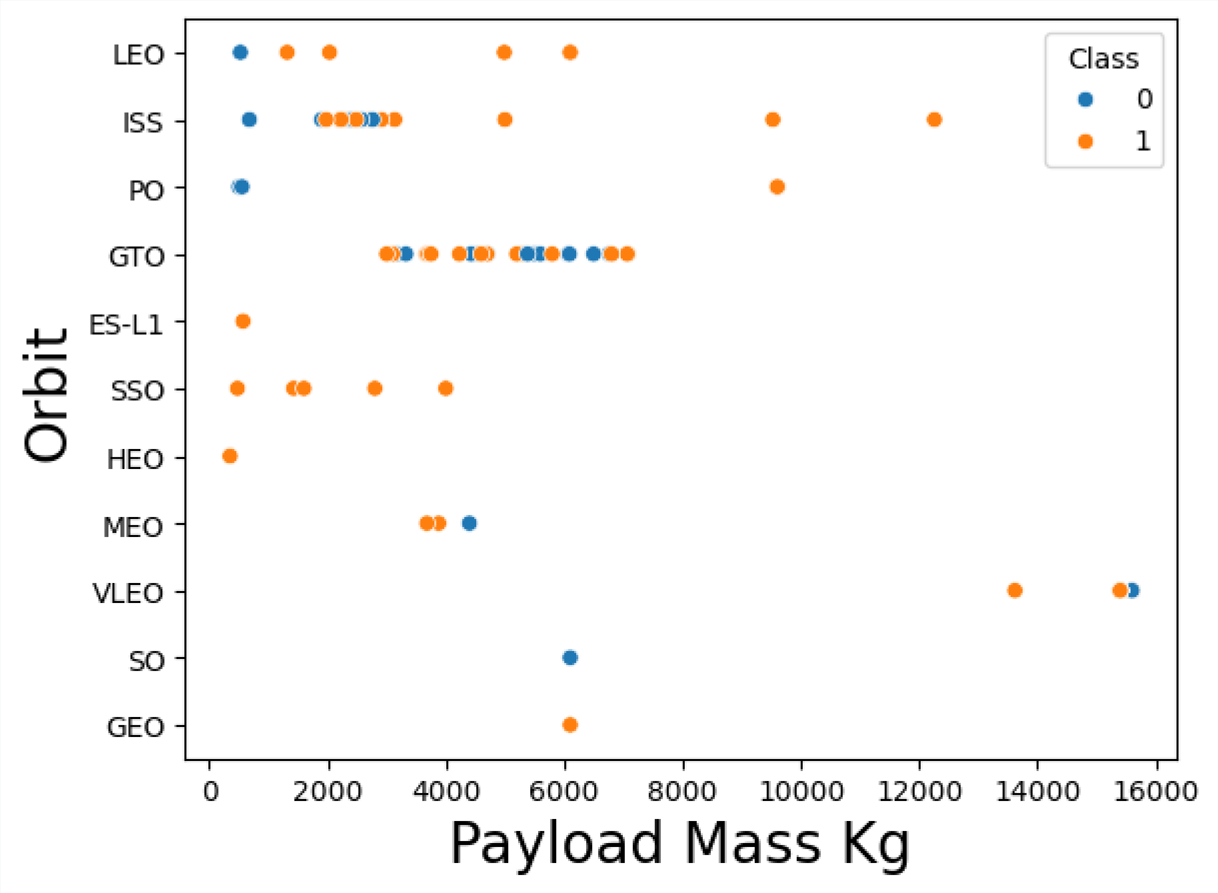
The success rate has increased as the flight no increased, meaning the recent flights with higher success rate.



Payload vs. Orbit Type

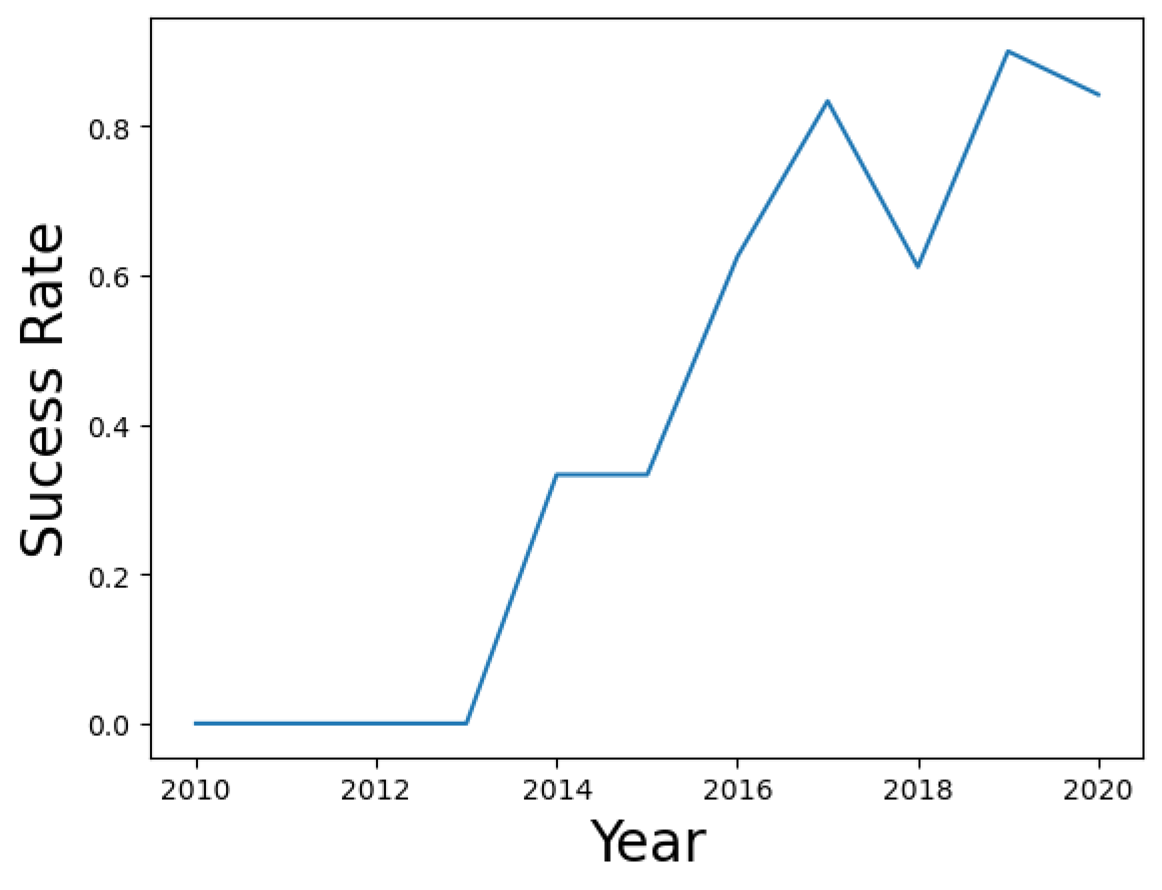
- Large payloads were sent to VLEO, ISS and PO orbits

- GTO has been attempted with multiple medium level payloads, with mixed results.



Launch Success Yearly Trend

- There were no success for the first 4 years.
- The success rate has improved over the years with an exception on year 2018 since 2013



All Launch Site Names

- Unique launch site names are
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40
- The SPACEXTBL has all the launches. Hence we need to find distinct names under “Launch Site” column to see the names of all sites.
- Query to get the above result are as below.
  - %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTBL

Launch Site Names Begin with 'CCA'

- Records starting the launch site name as CCA are to the right.
- We have selected five results from SPACEXTBL, then on Launch Site column , we have search for strings starting with CCA. % denotes trailing string with any character/s.
- The code to get the above result is as follows.
  - %sql SELECT * FROM SPACEXTBL WHERE "Launch_Site" LIKE "CCA%" LIMIT 5

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

Total Payload Mass

- Total Payload Mass for the customer NASA CRS missions are 45,596 kg

SUM(PAYLOAD_MASS__KG_)
45596

- Here we are selecting a sub table of PAYLOAD_MASS_KG , from the SPACEXTBL , when the customer is NASA CRS and summing the result.

- Code is as following
  - %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER IS 'NASA (CRS)’

Average Payload Mass by F9 v1.1

- The average payload mass by F9 v1.1 is 2928.4
- Here we do a small table of SPACEXTBL where the Booster version is F9 v1.1 and getting the average of the results.
- The SQL query goes as below.
  - `%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE "Booster_Version" IS "F9 v1.1"`

Task 4

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

[22]:
```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE "Booster_Version" IS "F9 v1.1"
```

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

[22]:

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- First successful landing outcome on ground pad was on 22/12/2015
- The query ran was
  - `%sql SELECT MIN(DATE) FROM SPACEXTBL WHERE "Landing_Outcome" IS "Success (ground pad)"`
- Finding the earliest data (minimum) of which Landing outcome is “Success (ground pad”

Task 5

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

Hint:Use min function

[25]:
```
%sql SELECT MIN(DATE) FROM SPACEXTBL WHERE "Landing_Outcome" IS "Success (ground pad)"
```

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

[25]:

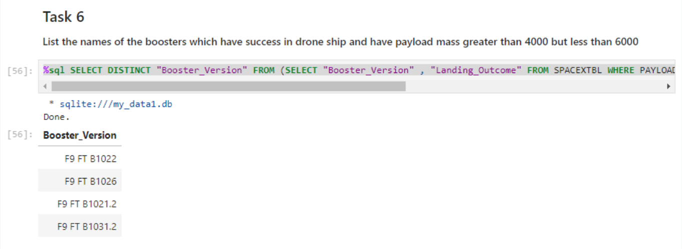
MIN(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

- SQL query is as below.
  - `%sql SELECT DISTINCT "Booster_Version" FROM (SELECT "Booster_Version" , "Landing_Outcome" FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000) WHERE "Landing_Outcome" IS "Success (drone ship)"`



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

```
[56]: %sql SELECT DISTINCT "Booster_Version" FROM (SELECT "Booster_Version" , "Landing_Outcome" FROM SPACEXTBL WHERE PAYLOAD
 * sqlite:///my_data1.db
Done.
```

[56]:

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

Total Number of Successful and Failure Mission Outcomes

- Mission Outcomes are as follows.

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

- Query result is as follows
  - `%sql SELECT COUNT ("Mission_Outcome"),"Mission_Outcome" FROM SPACEXTBL GROUP BY "Mission_Outcome"`

Boosters Carried Maximum Payload

- Following boosters carried the maximum payload of 15600 kg

  1. F9 B5 B1048.4
  2. F9 B5 B1049.4
  3. F9 B5 B1051.3
  4. F9 B5 B1056.4
  5. F9 B5 B1048.5
  6. F9 B5 B1051.4
  7. F9 B5 B1049.5
  8. F9 B5 B1060.2
  9. F9 B5 B1058.3
  10. F9 B5 B1051.6
  11. F9 B5 B1060.3
  12. F9 B5 B1049.7

SQL query is as below.

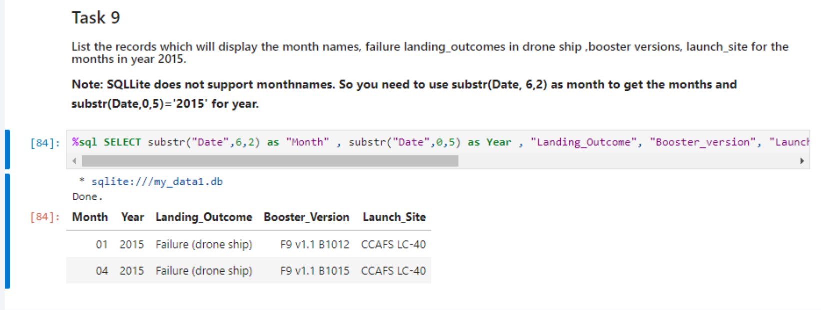
```
%sql SELECT DISTINCT "Booster_Version" FROM
SPACEXTBL WHERE "PAYLOAD_MASS__KG_" IS ( SELECT
MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL)
```

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

- SQL Query is as below
  - `%sql SELECT substr("Date",6,2) as "Month" , substr("Date",0,5) as Year , "Landing_Outcome", "Booster_version", "Launch_site" FROM SPACEXTBL WHERE "Landing_Outcome" IS "Failure (drone ship)" AND  SUBSTR("Date",0,5) IS "2015"`



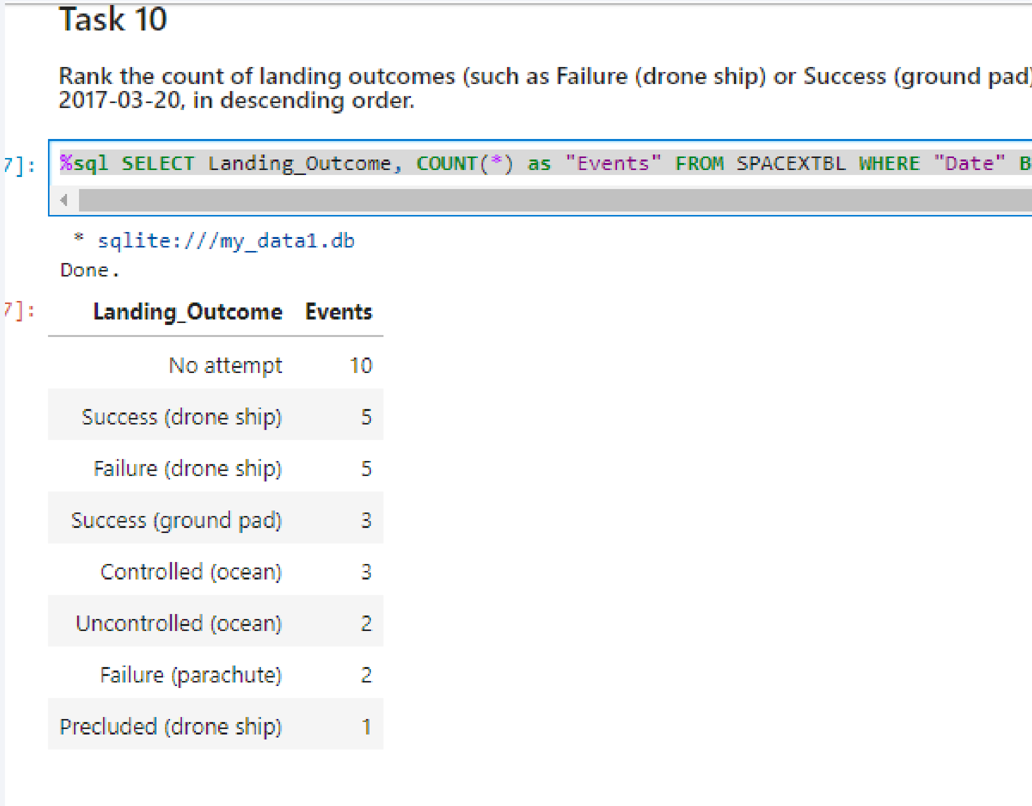
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Ranked count of landing outcomes.

- SQL query is as below.

  - ```
    %sql SELECT Landing_Outcome, COUNT(*) as "Events" FROM SPACEXTBL WHERE "Date" BETWEEN "2010-06-04" AND "2017-03-20" GROUP BY "Landing_Outcome"  ORDER BY "Events" DESC
    ```

- We are selecting the count of landing outcomes between the given dates by grouping them and ordering them in descending order.



Task 10

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad) 2017-03-20, in descending order.

```
%sql SELECT Landing_Outcome, COUNT(*) as "Events" FROM SPACEXTBL WHERE "Date" B
```

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

Landing_Outcome	Events
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
```

Section 3

Launch Sites Proximities Analysis

SpaceX Launch Sites

Key considerations selecting a space launch sites are as below, which can be observed from here and other assessments below.

1. Proximity to the Equator:

Earth's Rotational Velocity provides an initial boost to a rocket launched eastward. This boost is maximized closer to the equator, where the Earth's rotational speed is highest. And there is also Orbital Inclination which Launching closer to the equator allows for easier access to a wider range of orbital inclinations, including geostationary orbits.

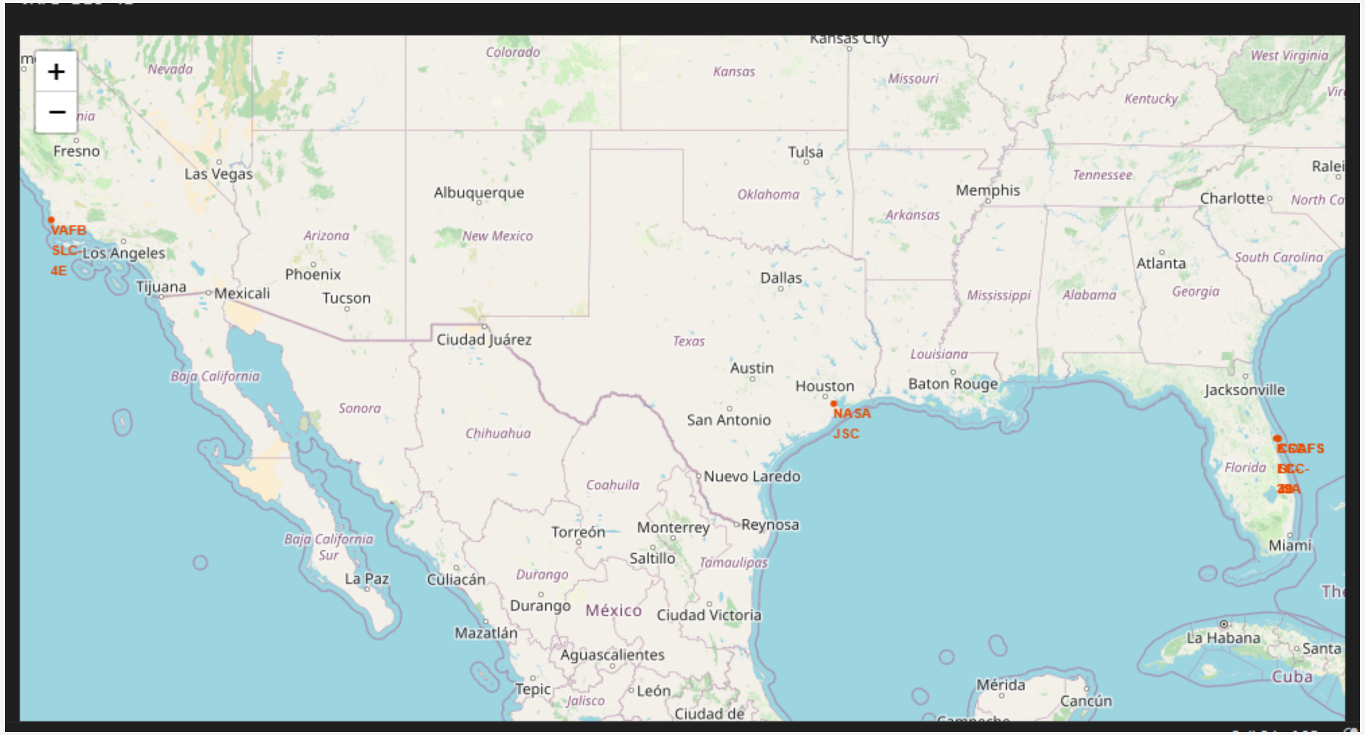
2. Downrange Safety:

Special considerations such as using unpopulated areas where Launch sites are typically located in remote areas with minimal population density to minimize the risk of casualties in case of launch failures. Further Many launch sites are situated near large bodies of water to ensure that falling debris lands in a safe, unpopulated area.

3. Infrastructure and Accessibility:

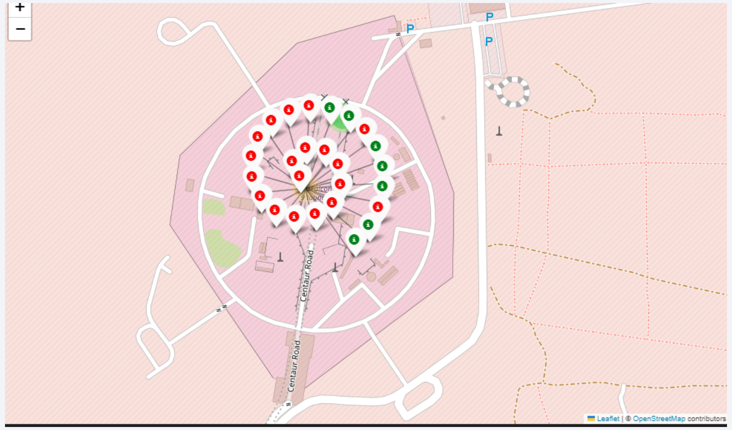
Transportation: Good transportation links are essential for transporting rocket components, fuel, and personnel to the launch site.

Support Facilities: Adequate infrastructure, including power, water, and communication systems, is necessary for the operation of the launch site.



Launch Outcomes

- Only 7/26 (27%) of the launches made from CCAFS LC - 40 were sucessfull



- Only 3/7 (43%) of the launches were successful form CCAFS SLC-40

Green for Successful Launches

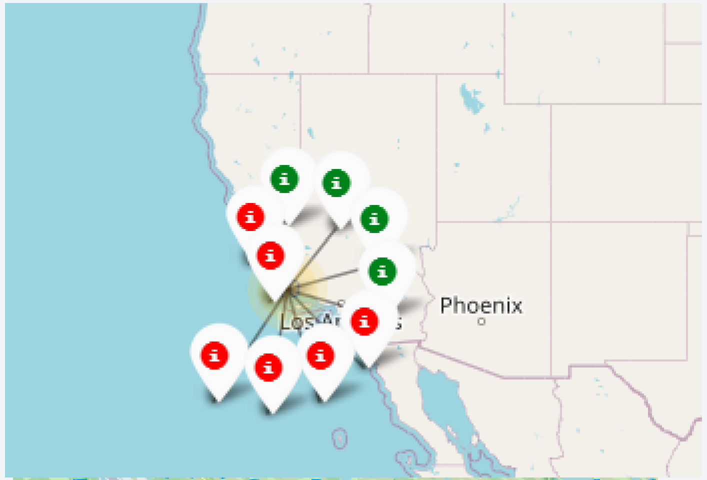
Red for Unsuccessful launches

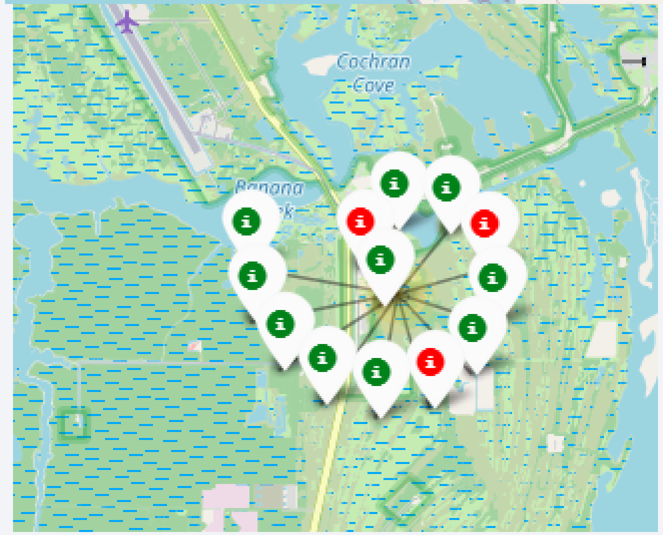
Launch Outcomes

- Only 4/10 (40%) of the launches made from VFAB SLC-4E were successful.

However

- **10/13 (77%) of the launches made from KSC-LC 39A were successful**





Green for Successful Launches

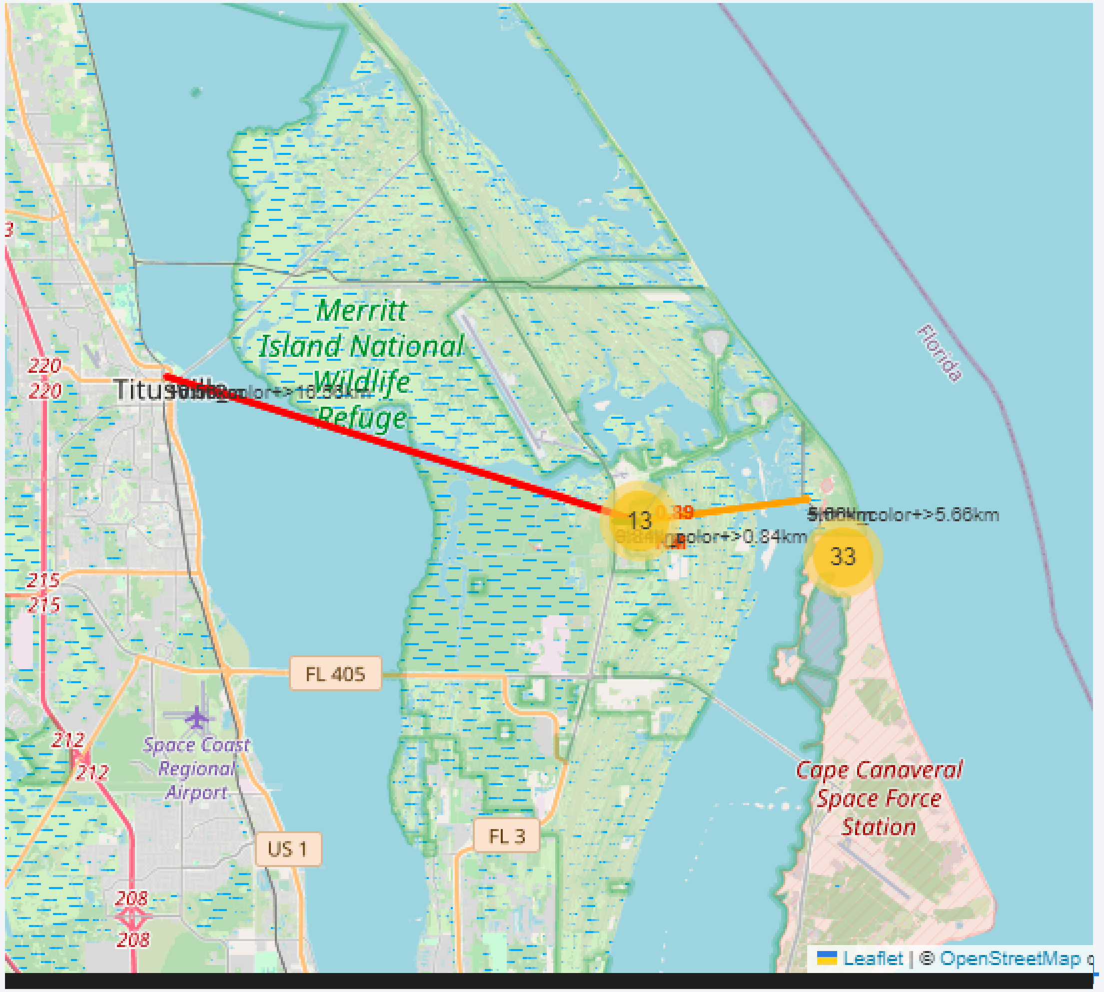
Red for Unsuccessful launches

Distance to Proximities from site VFAB SLC-4E

- VFAB SLC-4E site is 0.89 km from the sea

- And it is 16.5 km from the nearest city Titusville

- Close to main transportation locations such as 5.6 km from the railway and .84 km form the highway

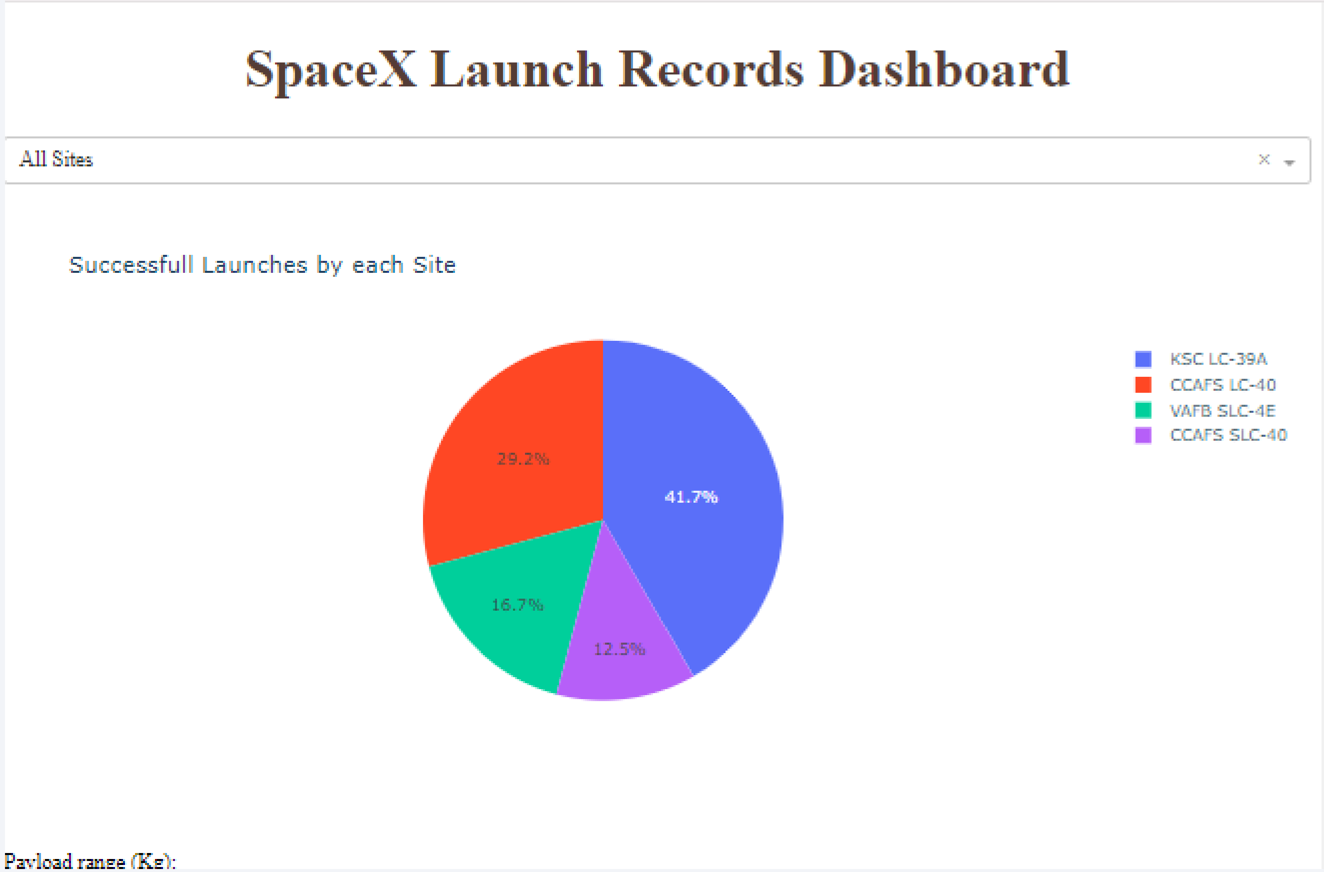


Section 4

Build a Dashboard with Plotly Dash

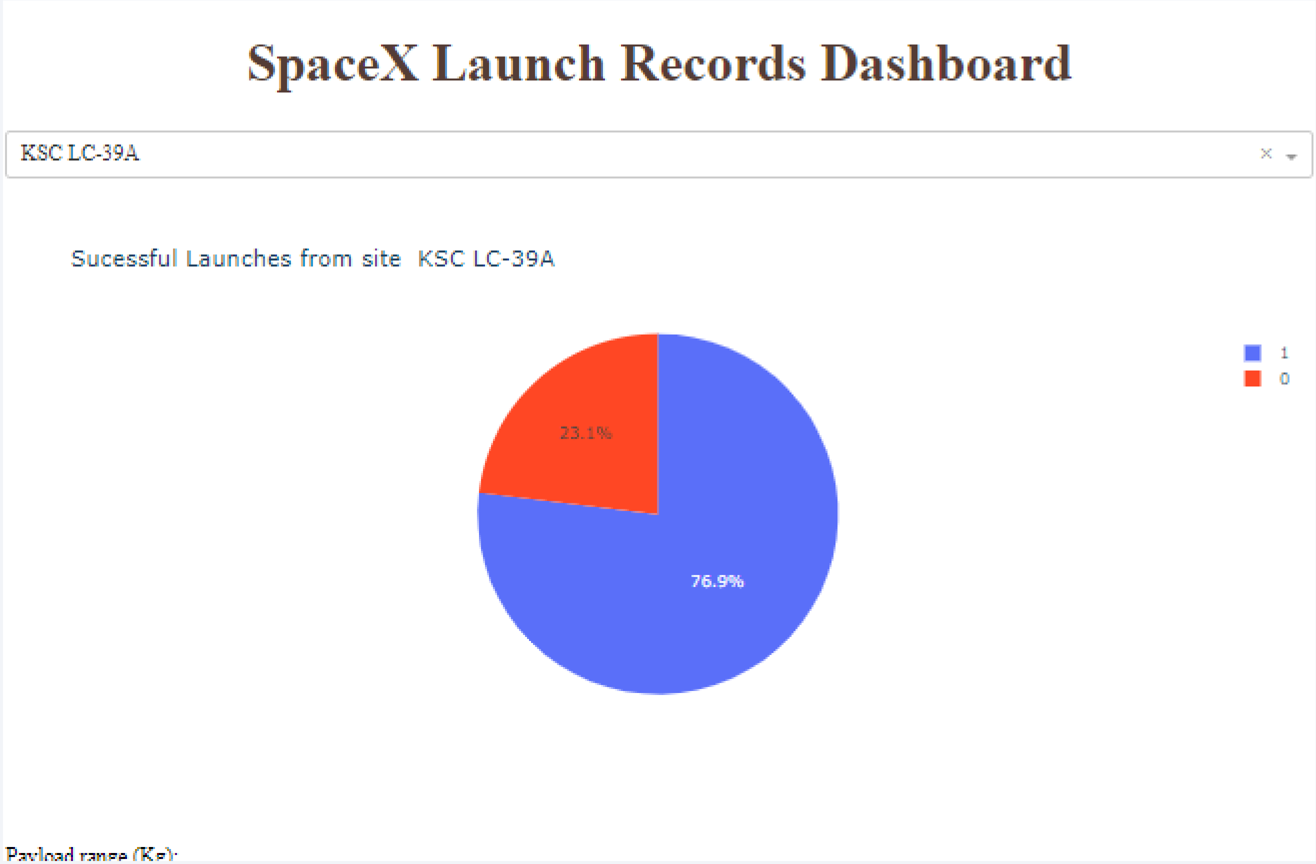
All Site Launch Success

- KSC LC-39A has the highest number of successful launches.



KSC LC-39A Site Success Rate

- There are 10(77%) Successfully launches from KSC LC-39A site

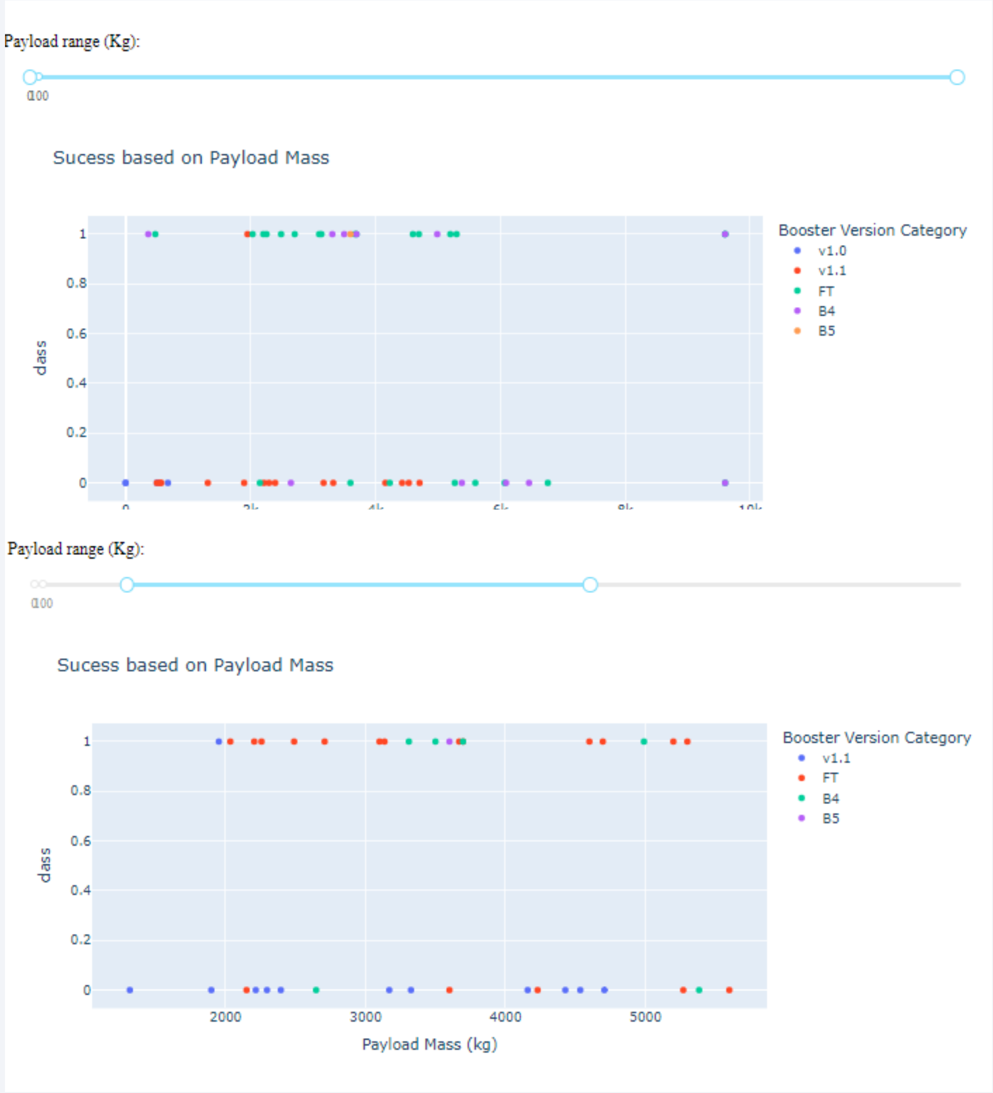


Impact on Payload Mass to Launch Success

- There seems to be limited success to launches beyond 6000 kg

- Also not much of a success for payloads lower than 2000 kg

- Hence data indicate the optimum payload range to be between 2000 and 6000

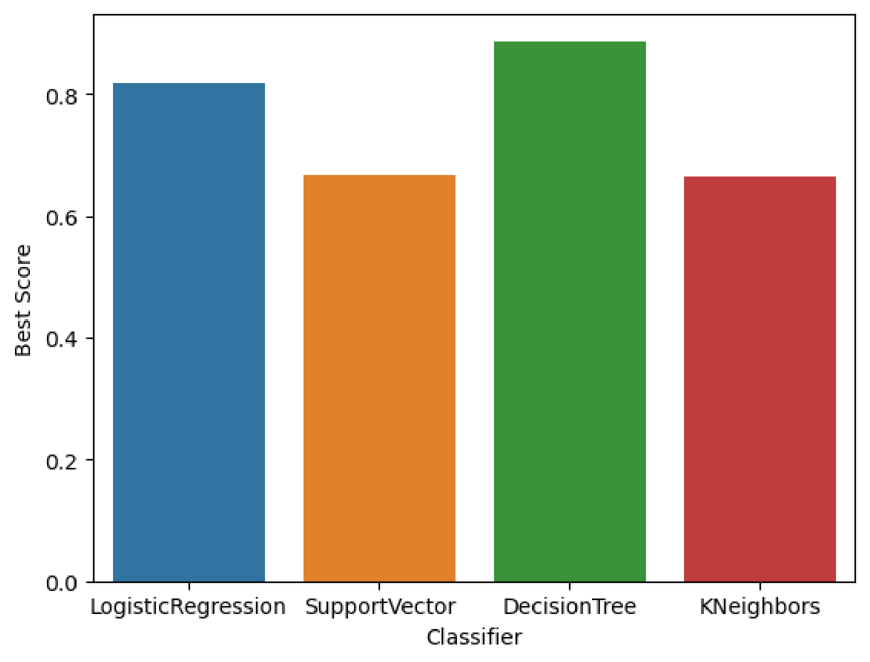


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Based on the accuracy factor, Decision Tree Classifier has the highest accuracy followed by Logistic Regression



Confusion Matrix of Decision Tree Classifier

- A confusion matrix summarizes the performance of a classification algorithm
- The fact that there are false positives (Type 1 error) is not a good indicator, which reduces the precision and F1 score
- Confusion Matrix Outputs are as below for the decision tree classifire:
  - 12 True positive
  - 2 True negative
  - **4 False positive**
  - 0 False Negative

- **Precision**= TP / (TP + FP) = 12 / 16 = .75

- **Recall**= TP / (TP + FN)12 / 12 = 1

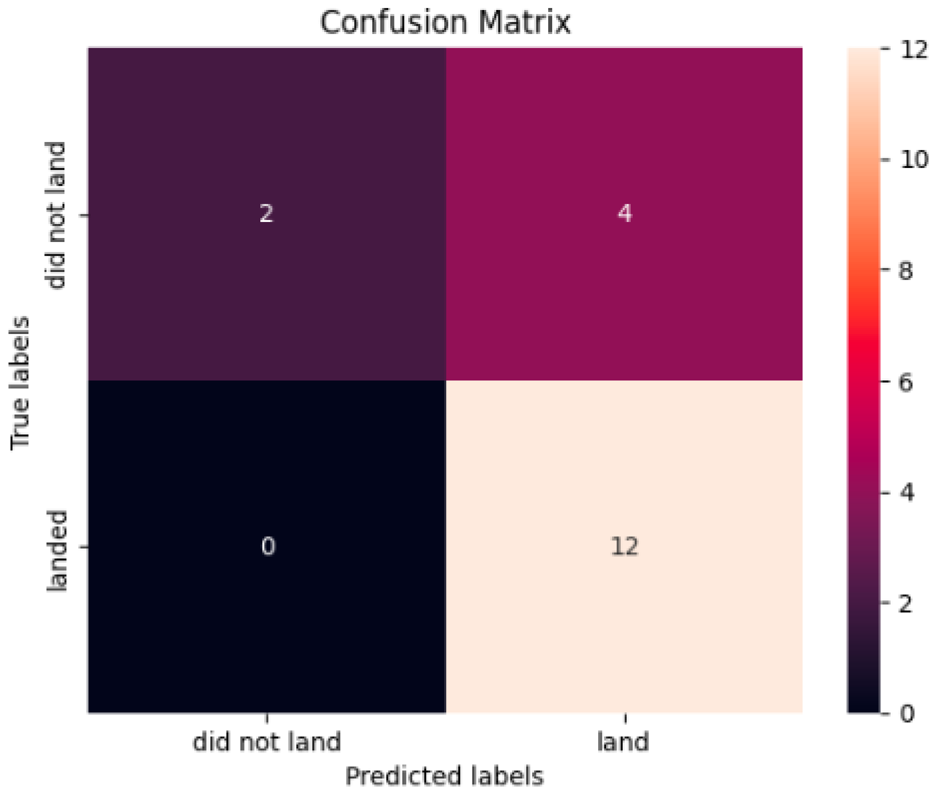
- **F1 Score**= 2 * (Precision * Recall) / (Precision + Recall)

  =2 * (.75 * 1) / (.75 + 1) = 1,5/ 1.75 = 0.857

- **Accuracy**= (TP + TN) / (TP + TN + FP + FN) = 14/18

  = 0.77

```
[25]: yhat = tree_cv.predict(X_test)
      plot_confusion_matrix(Y_test,yhat)
```



Conclusions

- Success of missions has increase with time
- We can observe a 100% success rate for ES-L1, GEO , HEO and SSO orbit types
- There were serval drone ship landings
- Highest performing payload mass is between 2,000 and 6,000 kg
- Space X has chosen sites near the equator. Other factors such as close to a main road and railway was also import
- Being close to a coast makes some room for failed landings
- Most successful site is KSC LC-39A with a success rate of 77%
- Decision Tree Classifier has the best learning algorithm for the data available

Thank you!