IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mubeena Aliyarukunju

27-3-2025

Outline

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

Executive Summary

- Summary of methodologies
- Data Collection
- Data Wrangling
- EDA with Data Visualization
- EDA with SQL
- Building an Interactive Map with Folium
- Building a Dashboard with Plotly Dash
- Predictive Analysis-Classification with Machine learning
- Summary of all results
- **Exploratory Data Analysis Results**
- **Interactive Analytics demo in screenshots**
- **Predictive Analysis Results**

Introduction

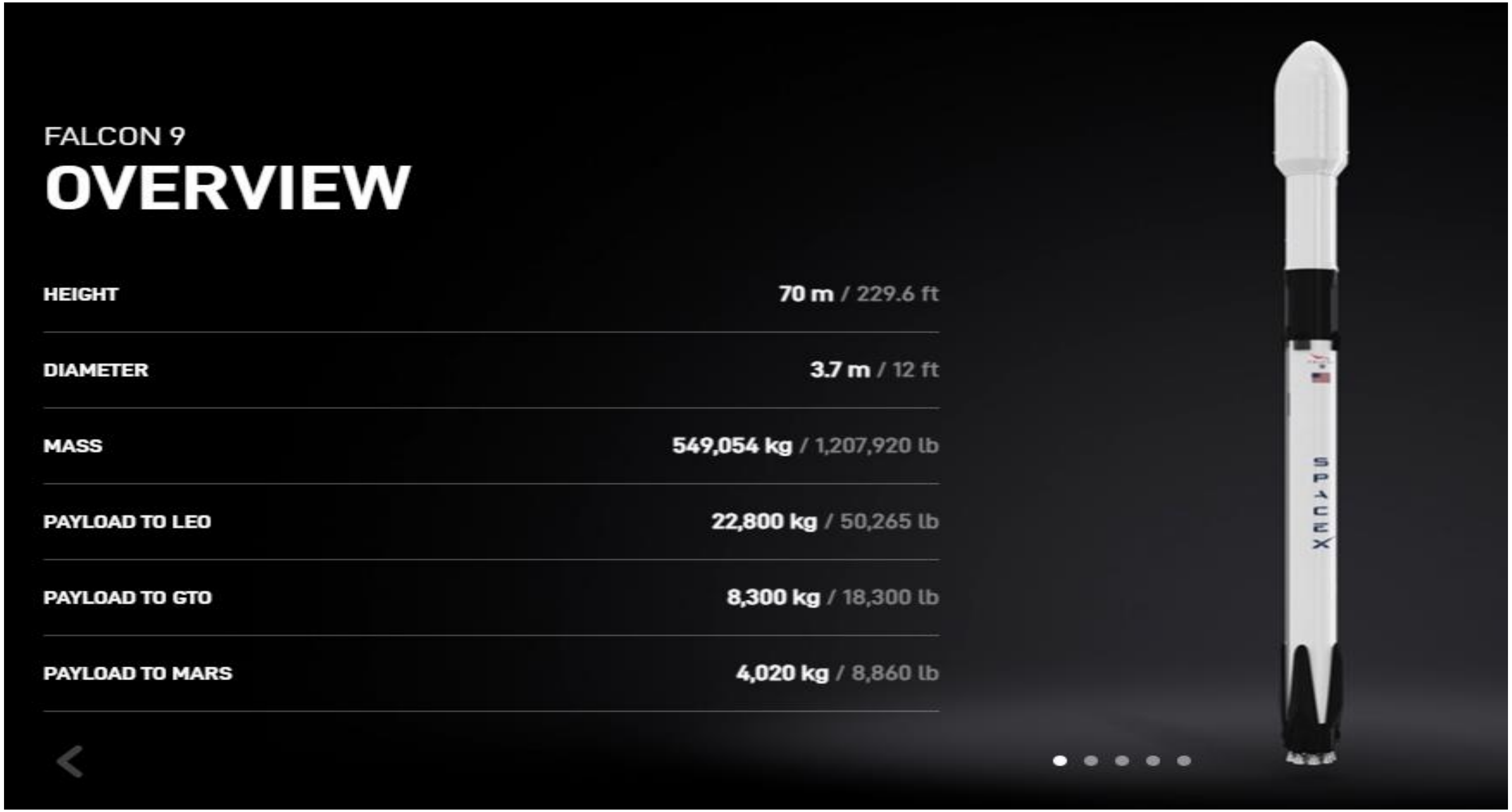
- Project background and context

In this Project, We will 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 the launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

- Problems you want to find answers

- **What factors determine if the rocket will land successfully?**
- **The interaction amongst various features that determine the success rate of a successful landing?**
- **What are the operating conditions needs to be in place to ensure a successful landing program.**

OVERVIEW OF FALCON 9


FALCON 9

OVERVIEW

HEIGHT	**70 m** / 229.6 ft
DIAMETER	**3.7 m** / 12 ft
MASS	**549,054 kg** / 1,207,920 lb
PAYLOAD TO LEO	**22,800 kg** / 50,265 lb
PAYLOAD TO GTO	**8,300 kg** / 18,300 lb
PAYLOAD TO MARS	**4,020 kg** / 8,860 lb

FALCON 9

FIRST STAGE

OVERVIEW | ENGINES | LANDING LEGS

Falcon 9's first stage incorporates nine Merlin engines and aluminum-lithium alloy tanks containing liquid oxygen and rocket-grade kerosene (RP-1) propellant.

Falcon 9 generates more than 1.7 million pounds of thrust at sea level.

FALCON 9

INTERSTAGE

The interstage is a composite structure that connects the first and second stages, and houses the pneumatic pushers that allow the first and second stage to separate during flight.

GRID FINS

Falcon 9 is equipped with four hypersonic grid fins positioned at the base of the interstage. They orient the rocket during reentry by moving the center of pressure.

FALCON 9

PAYLOAD

FAIRING | DRAGON

Made of a carbon composite material, the fairing protects satellites on their way to orbit. The fairing is jettisoned approximately 3 minutes into flight, and SpaceX continues to recover fairings for reuse on future missions.

HEIGHT	**13.1 m** / 43 ft
DIAMETER	**5.2 m** / 17.1 ft

Section 1

Methodology

Methodology

Executive Summary

- **Data collection methodology:**
  - SpaceX Rest API
  - Web Scrapping from Wikipedia page ‘List of Falcon9 and Falcon Heavy Launches’
- **Perform data wrangling**
  - Transforming categorical data using One Hot Encoding for machine learning algorithms and removing any or unnecessary information from the dataset.
- **Perform exploratory data analysis (EDA) using visualization and SQL**
- **Perform interactive visual analytics using Folium and Plotly Dash**
- **Perform predictive analysis using classification models**
  - Logistic Regression, K-Nearest Neighbors, Support Vector Machines and Decision Tree models have been developed to determine the most effective classification method.

Data Collection

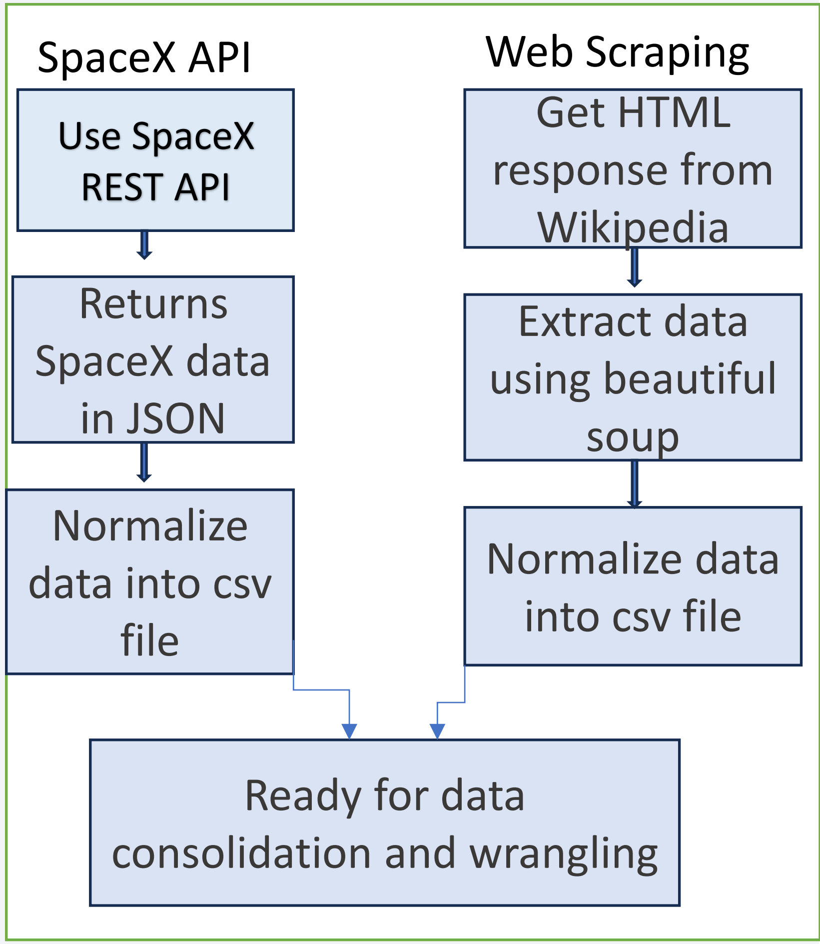
- The data sets are collected using two methods:

1. SpaceX launch data – gathered from the SpaceX REST API.

   This API will give us data about launches, including information about the rocket used , payload delivered, launch specifications, landing specifications and landing outcome.

   The SpaceX REST API endpoints, or URL starts with api.spacexdata.com/v4/.

2. Another popular data source for obtaining Falcon 9 Launch data is web scraping Wikipedia using BeautifulSoup.

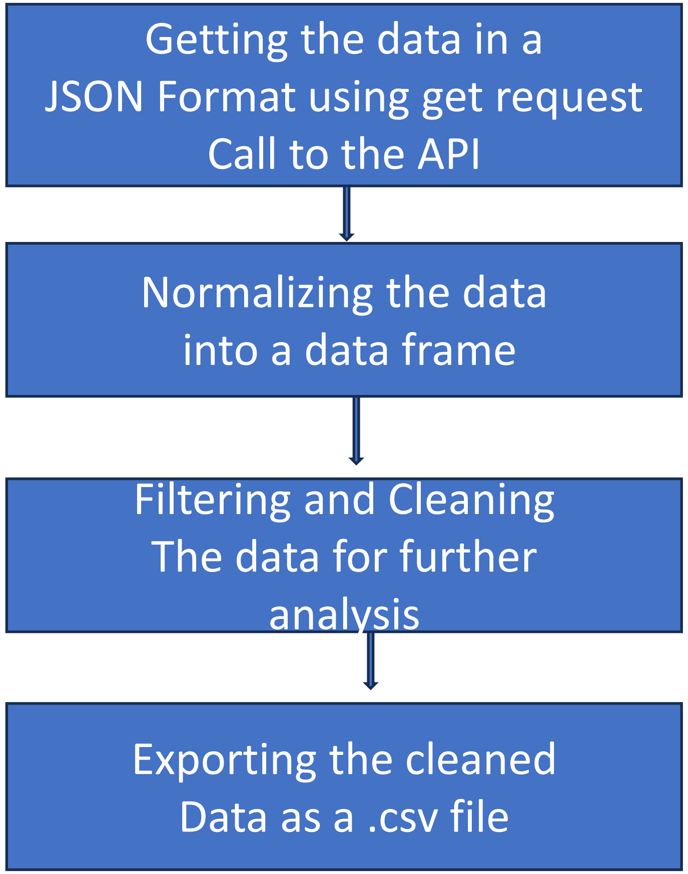


Data Collection-SpaceX API

The first dataset is collected using the SpaceX API, using get requests in Python. Then we normalize the json contents into a data frame and then using functions and pandas we extract relevant information, clean the data, and export the cleaned data.

GitHub URL of the Notebook:
https://github.com/MubeenaAli/IBM-Data-Science-Capstone-Project-SpaceX-landing-prediction.git

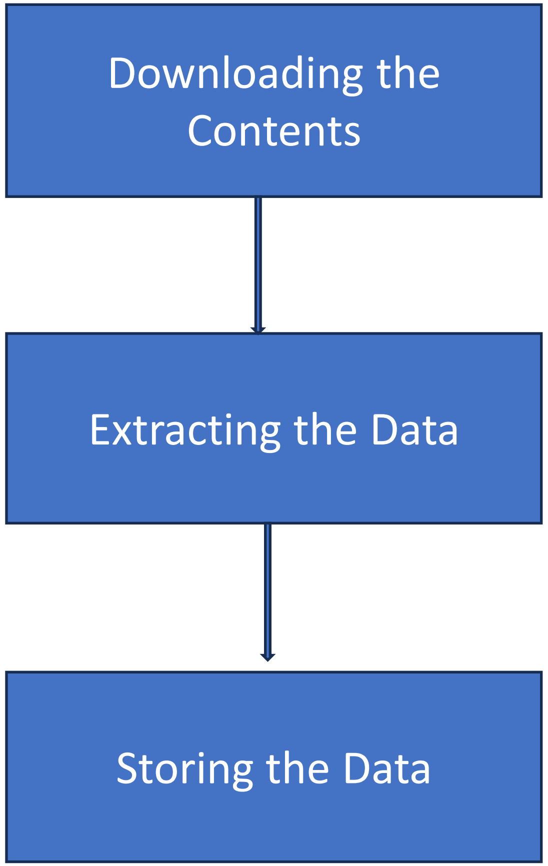
The basic flowchart of the process is shown:



Data Collection-Scraping

Web Scraping Wikipedia:

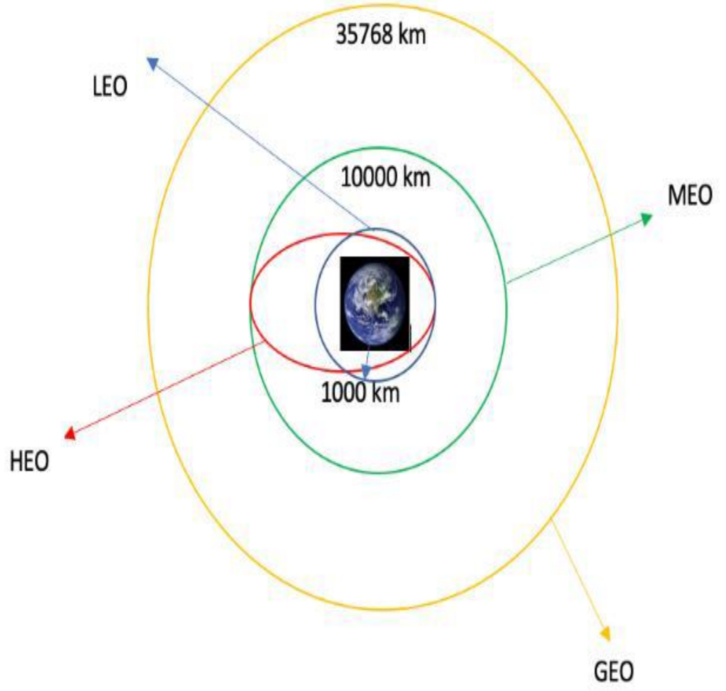
- **Identify target URL for Falcon 9 and Falcon Heavy launches**: We target the specific Wikipedia page containing the launch records.
- **Use BeautifulSoup to parse HTML content**: We use the BeautifulSoup library to parse the HTML content of the webpage.
- **Extract launch records table**: The HTML table containing the launch records is extracted.
- **Convert HTML table to Pandas DataFrame**: The extracted table is then converted into a Pandas DataFrame for further analysis.
- URL of the completed web scraping Notebook: https://github.com/MubeenaAli/IBM-Data-Science-Capstone-Project-SpaceX-landing-prediction.git



Data Wrangling

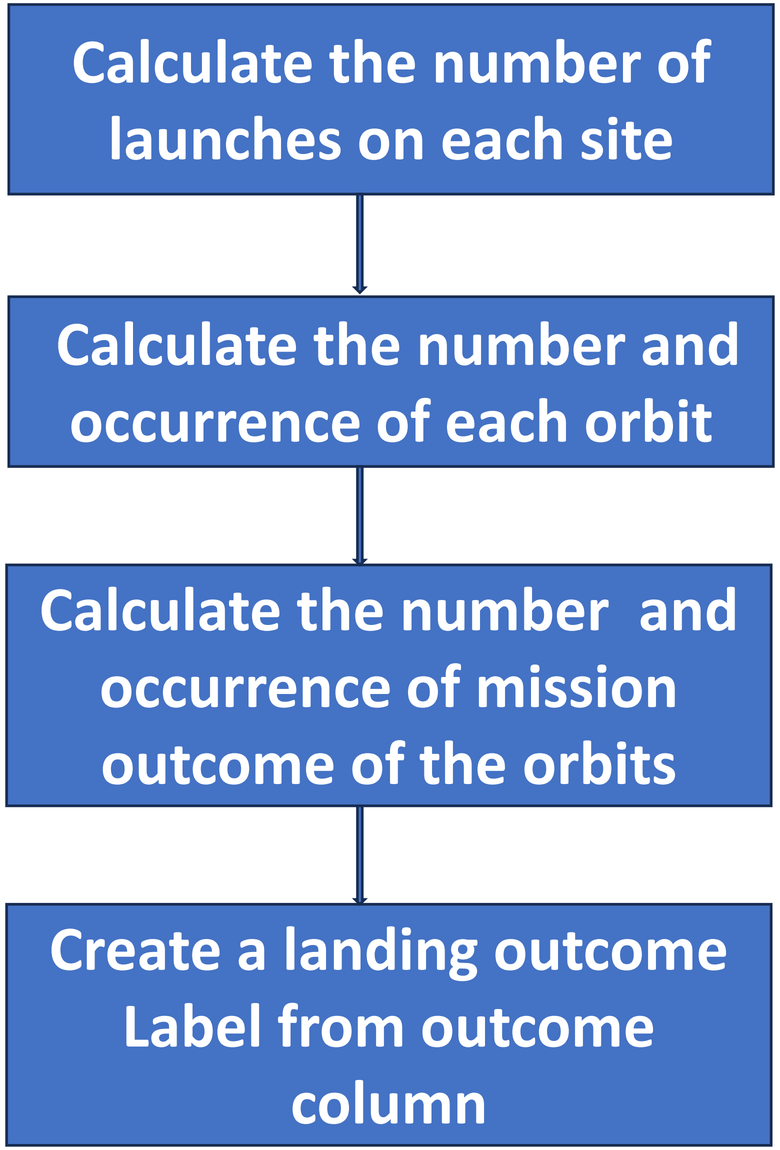
- **Calculate Number of Launches per Site**
  - Group data by launch site and count number of launches per site.
- **Calculate Number and Occurrence of Each Orbit**
  - Group data by orbit type and count number of occurrences for each orbit.
- **Calculate Mission Outcomes per Orbit**
  - Group data by orbit type and mission outcome and count number of occurrences for each mission outcome.
- **Create Landing Outcome Label**
  - Define successful and unsuccessful landing outcomes and create a new column for landing outcome labels based on existing outcome data.

GitHub URL: https://github.com/MubeenaAli/IBM-Data-Science-Capstone-Project-SpaceX-landing-prediction.git



Different types of orbits

Data wrangling Flowchart



EDA with Data Visualization

1. Visualization of relationship between Flight number and launch site (bar & scatter)
2. Visualization of relationship between payload and launch site (bar & scatter)
3. Visualization of relationship between success rate and each orbit (bar & scatter).So that side-by-side comparison could be done.
4. Visualization of relationship between flight number and orbit type (bar & scatter)
5. Visualization of relationship between payload and orbit type (bar & scatter)
6. Visualization of yearly launch success trend

GitHub URL: https://github.com/MubeenaAli/IBM-Data-Science-Capstone-Project-SpaceX-landing-prediction.git

EDA with SQL

1. Display the names of the unique launch sites in the space mission
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 successful 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 successful landing outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

GitHub **URL:https://github.com/MubeenaAli/IBM-Data-Science-Capstone-Project-SpaceX-landing-prediction.git**

Build an Interactive Map with Folium

- Generated initial site map with coordinates and zoom
- Created and added folium circle and folium marker for each launch site to be able to distinctly identify them on the map
- Created a new column in data frame caller “marker color” to store color based on class, with green being success and red being failure
- Marked the success/failed launches for each site on the map using a marker cluster to enhance map readability and appearance
- Calculated the distance between launch site and the nearest coastline, highway and city respectively
- Created and added folium marker to map and drew a line from launch site to each of the above to visualize and analyze proximities
- GitHub [URL:https://github.com/MubeenaAli/IBM-Data-Science-Capstone-Project-SpaceX-landing-prediction.git](https://github.com/MubeenaAli/IBM-Data-Science-Capstone-Project-SpaceX-landing-prediction.git)

Build a Dashboard with Plotly Dash

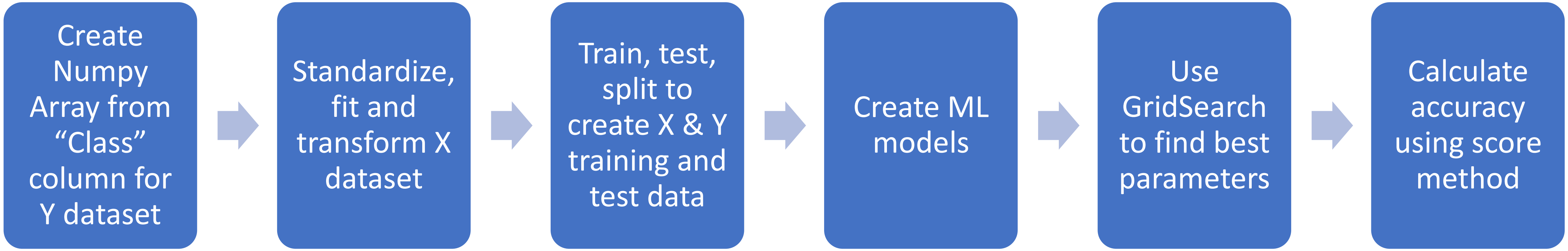
- Built a dashboard with dropdown options for each launch site (or all)
- Displayed a pie chart showing success and failure percentage of each site to show the breakdown of launches for each site as well as a summary chart of percentage launch success contributed by each site when "All" is selected
- Displayed a slider control scatter plot of success and booster version category with slider controlling payload mass to show the booster version category success rate based on payload mass

GitHub URL: https://github.com/MubeenaAli/IBM-Data-Science-Capstone-Project-SpaceX-landing-prediction.git

Predictive Analysis (Classification)

- Created a Numpy array from the “Class” column of the data
- Standardized, fit and transformed X data
- Used the function train_test_split to create X and Y training and test datasets
- Created logistic regression, support vector machine, decision tree and k nearest neighbor models
- Used GridSearch to find the best parameters for each model as well as used the score method to calculate accuracy of each
- GitHub URL: https://github.com/MubeenaAli/IBM-Data-Science-Capstone-Project-SpaceX-landing-prediction.git

Predictive Analysis Flowchart



Results

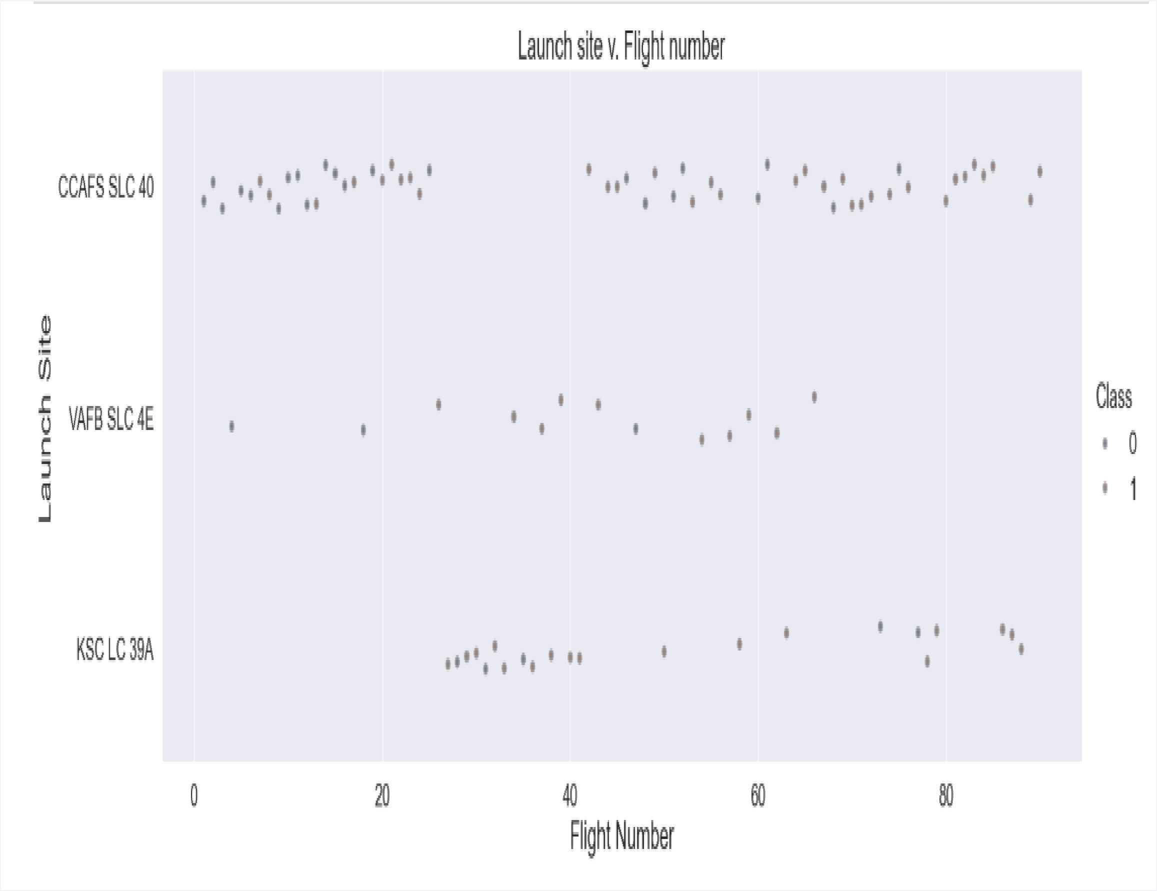
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

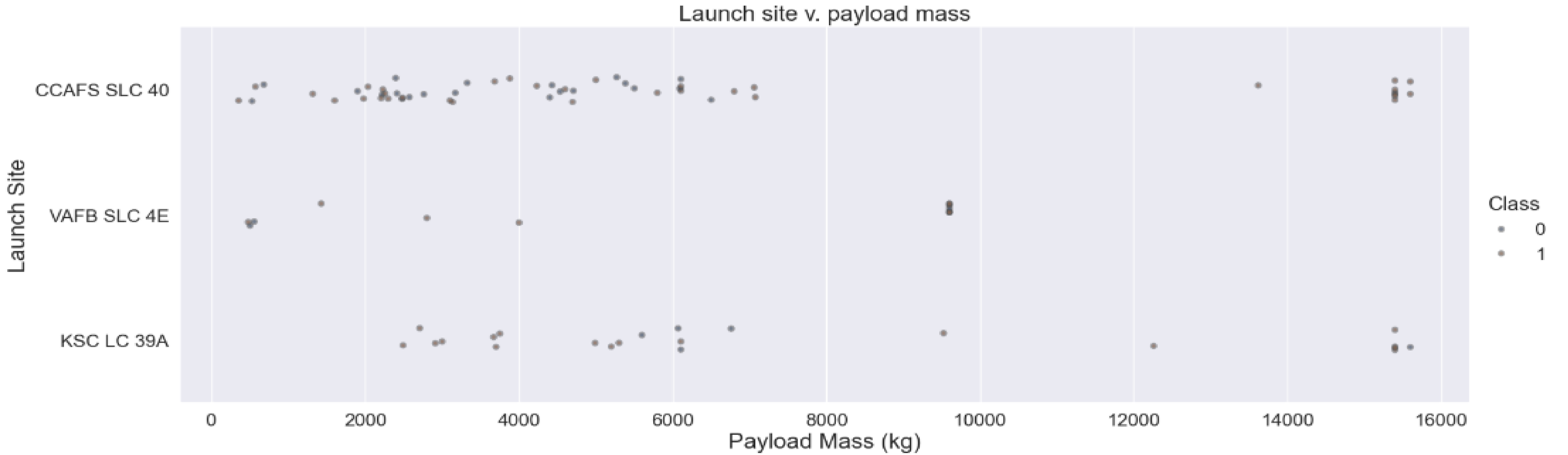
Insights drawn from EDA

Flight Number vs. Launch Site

- Most flights seem to be from CCAFSSLC 40, followed by KSC LC 39A.
- Fewer launches are from VAFB SLC 4E.
- More recent flights seem to have higher success rate.
- As flight numbers increase, a higher proportion of launches appear to have successful landings, indicating technological improvements over time.



Payload vs. Launch Site



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).

- CCAFS SLC 40 and KSC LC 39A have a wide range of payloads, including heavy payloads up to 16,000 kg.
- VAFB SLC 4E seems to have only lighter payloads(below 10,000 kg) suggesting it may not handle heavy launches.
- Heavier payloads (above 10,000 kg) appear to be launched mostly from CCAFS SLC 40

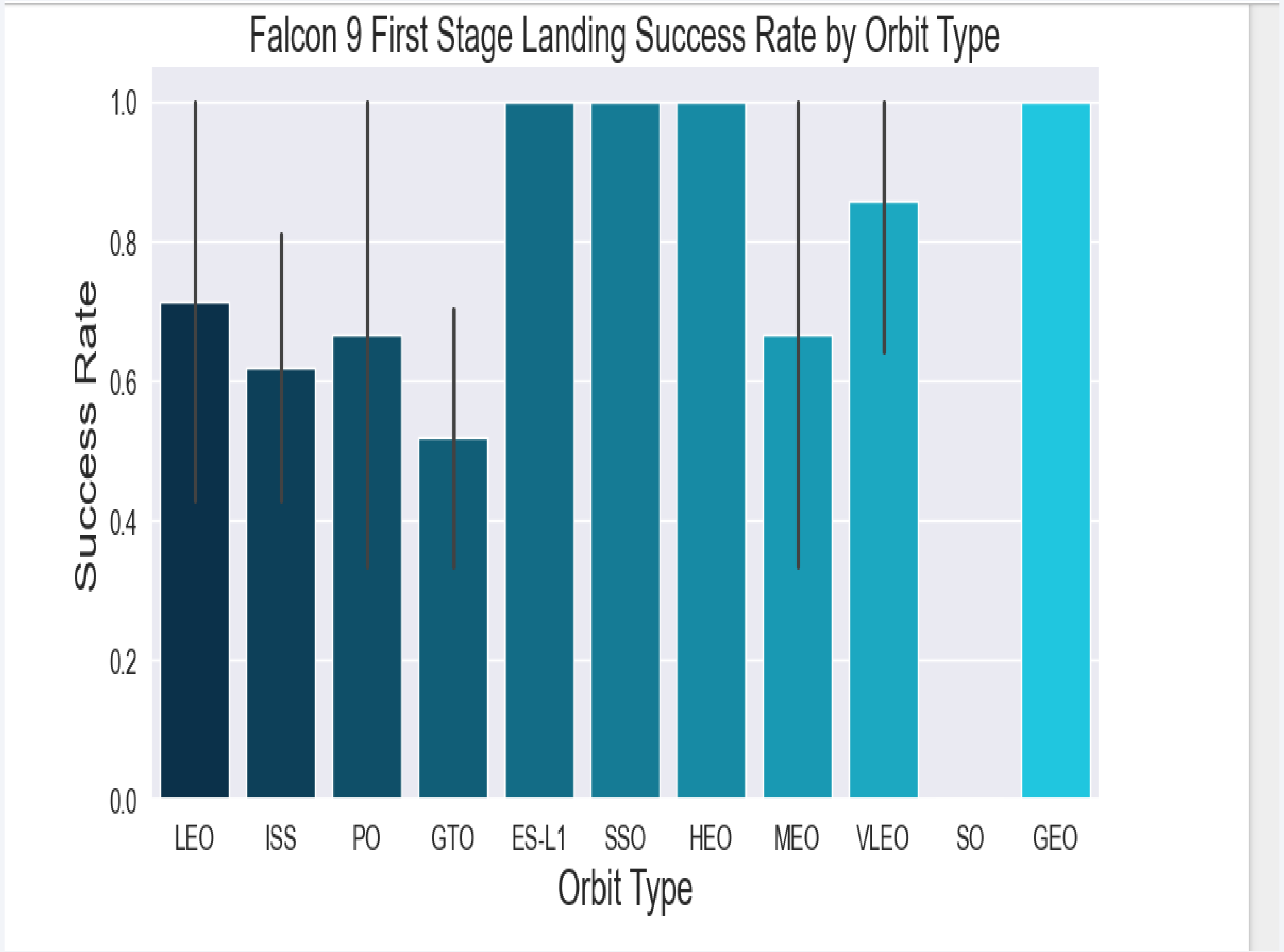
  and KSC LC 39A, indicating that these sites are preferred for high-mass

  missions.
- The VAFB SLC 4E launch site does not handle heavy payloads(>10000 kg), while

  CCAFS SLC 40 and KSC LC 39A support both light and heavy payloads.

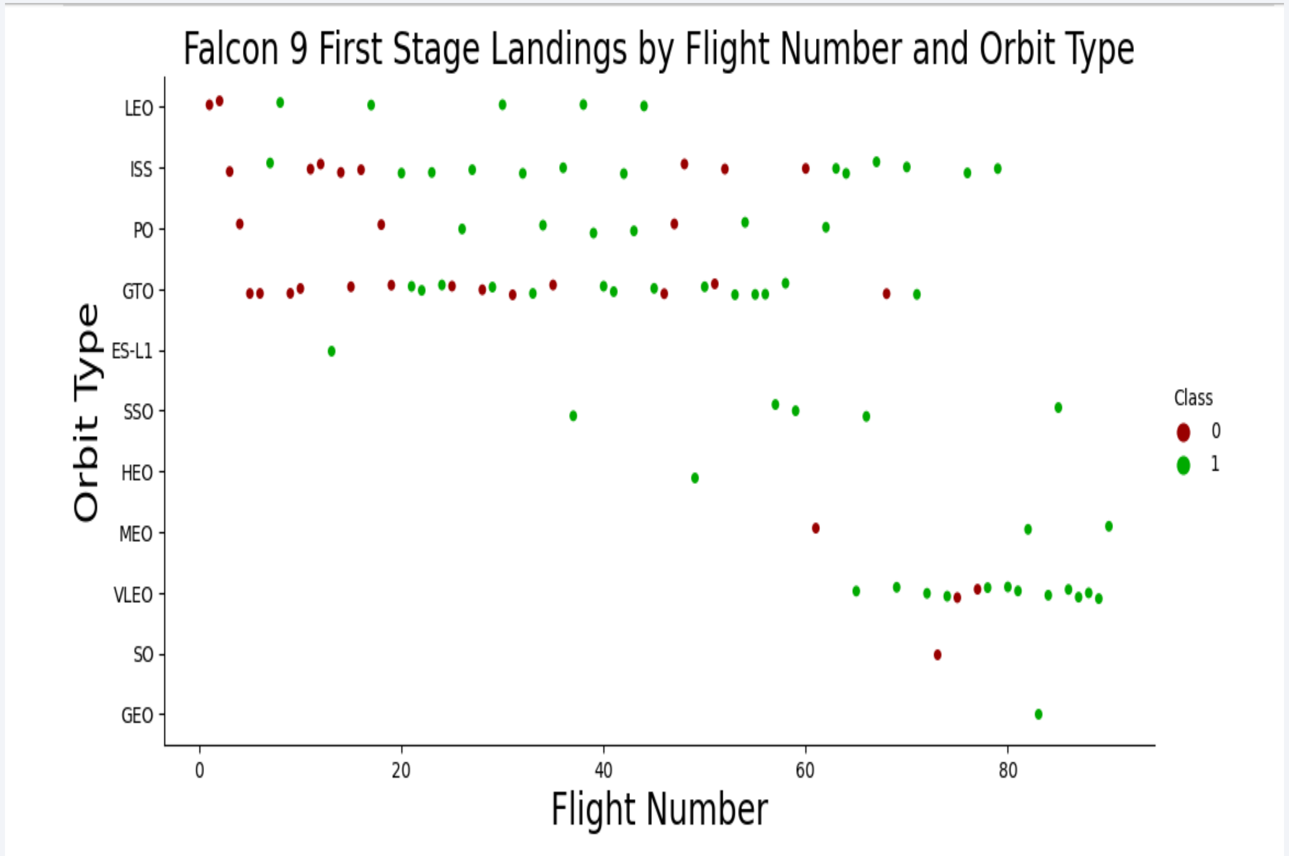
Success Rate vs. Orbit Type

- GEO, ES-L1,SSO,HEO: Show the highest landing success rates, close to 100%.
- LEO and PO: Have relatively lower success rates, indicating more frequent failures or variability.
- MEO: Has a lower success rate with higher variability(as indicated by larger error bars).
- Higher-energy orbits (like GEO) require more fuel for reentry, affecting landing success.



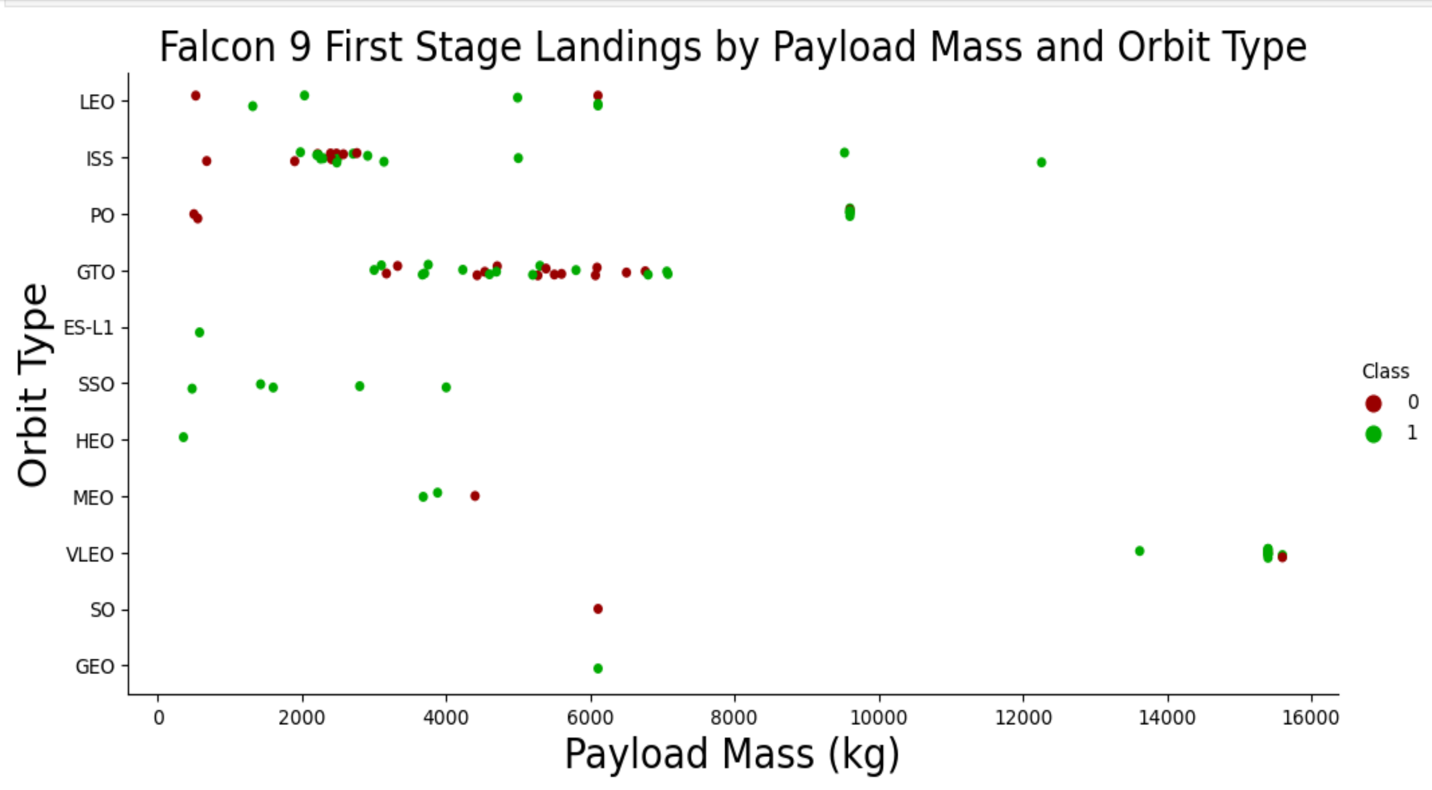
Flight Number vs. Orbit Type

- Based on the previous slide, ES-L1, GEO, HEO, and SSO have the best success rate for the orbit destination.
- However, for ES-L1, HEO, and GEO, there is only one launch for each orbit destination. Thus, no clear relationship can be said about these orbits. The same can be said for GTO orbit, too, due to a series of failures and successes across all flight numbers.
- The success of LEO orbit does appear to be related to the number of flights due to having a higher success on the later number of a flights.

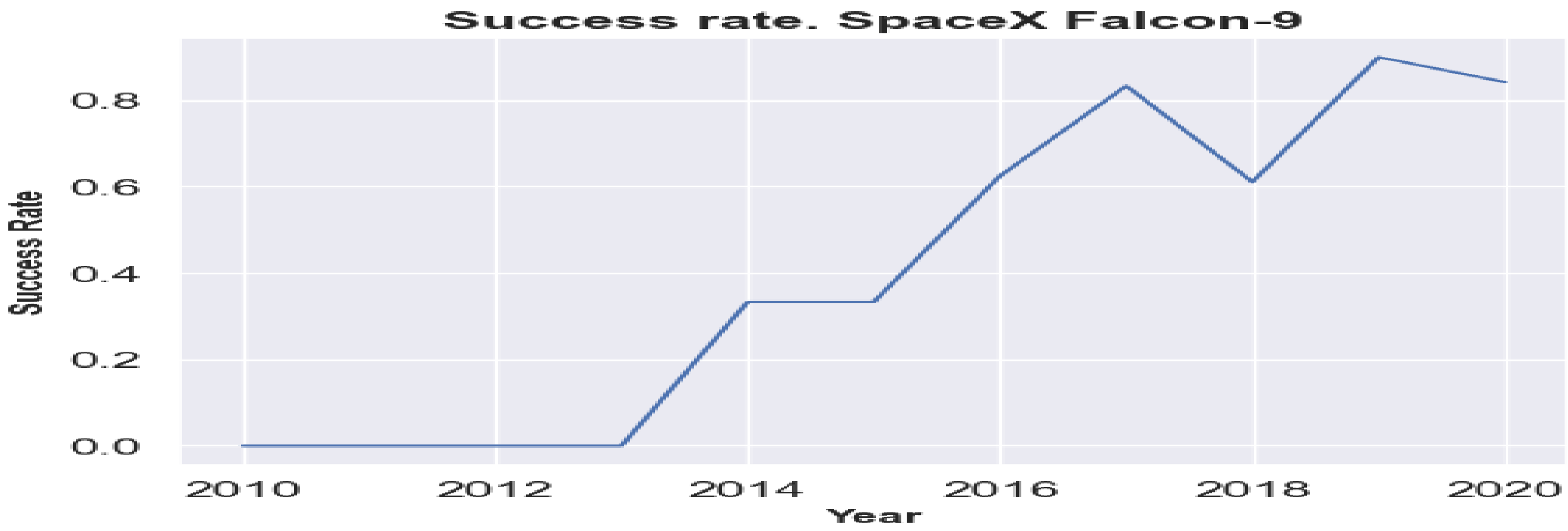


Payload vs. Orbit Type

- With heavy payloads the successful landing or positive landing rate are more for PO, LEO and ISS orbits.
- However, for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.



Launch Success Yearly Trend



You can observe that the success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.

- **The success rate since 2013 kept increasing till 2020**
- **2010-2013 period had no success rate**

All Launch Site Names

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

```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE

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

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

The DISTINCT statement used in this query ensures that the query only returns unique launch site names from the Launch_ Site column.

Launch Site Names Begin with 'CCA'

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

```
%sql SELECT * FROM SPACEXTABLE 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	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

This query used the LIKE operator and limit keyword to display only 5 records, of site names beginning with CCA.

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

This query used the SUM function to total the amount of payload mass from the column Payload_Mass__Kg_ launched by the customer ‘NASA (CRS)’.

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 = 'F9 v1.1';
```

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

AVG(PAYLOAD_MASS__kg_)
2928.4

This query used the AVG function to retrieve the average payload mass from column Payload_Mass__kg_ that is carried by the booster 'F9 v1.1'.

First Successful Ground Landing Date

```
%sql SELECT  MIN("Date") AS "First_Successful_Landing_On_Ground_Pad" FROM  SPACEXTBL WHERE "Landing_Outcome" = 'Success (ground pad)';
```

 * sqlite:///my_data1.db
Done.

First_Successful_Landing_On_Ground_Pad
2015-12-22

The MIN function on the launch date to retrieve the first date of which the condition of the "Landing_Outcome" is 'success (ground pad)'.

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  "Booster_Version" AS "Booster_Name" FROM  SPACEXTBL WHERE  "Landing_Outcome" = 'Success (drone ship)'  AND "Payload_Mass__kg_" > 4000  AND "
```

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

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

This query was used to retrieve the booster version that successfully landed on a drone ship and carried a payload mass between 4000kg and 6000kg.

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%sql SELECT "Mission_Outcome", COUNT(*) AS "Total" FROM SPACEXTBL GROUP BY "Mission_Outcome";
```

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

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

In this query used count() aggregate function and 'GROUP BY' statement to find total number of successful and failed outcomes.

Boosters Carried Maximum Payload

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

```
%sql SELECT "Booster_Version",Payload, "PAYLOAD_MASS__KG_" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL);
```

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

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

This query was used to retrieve booster versions and payload mass with the condition that the booster carried the maximum payload mass. This query also used sub-query as its condition to find the maximum payload mass.

2015 Launch Records

```
%sql SELECT CASE substr(Date, 6, 2) WHEN '01' THEN 'January' WHEN '02' THEN 'February' WHEN '03' THEN 'March' WHEN '04' THEN 'April' WHEN '05' THEN 'May'
```

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

Month_Name	Landing_Outcome	Booster_Version	Launch_Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

This query retrieved the month, landing outcome, booster version, and launch site for launch in the year 2015 and has an outcome of 'Failure (drone ship)'

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(*) AS "Count" FROM  SPACEXTBL WHERE  "Date" BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY  "Landing_Outcome" ORDER
```

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

Landing_Outcome	Count
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

This query was used to retrieve the number of launches that resulted in each landing outcome between 2010-06-04 and 2017-03-20. The GROUP BY statement is to group the count of each landing outcome and then ORDER BY was used to rank the landing outcome columns based on the number of each total launch in descending order.

Section 3

Launch Sites Proximities Analysis

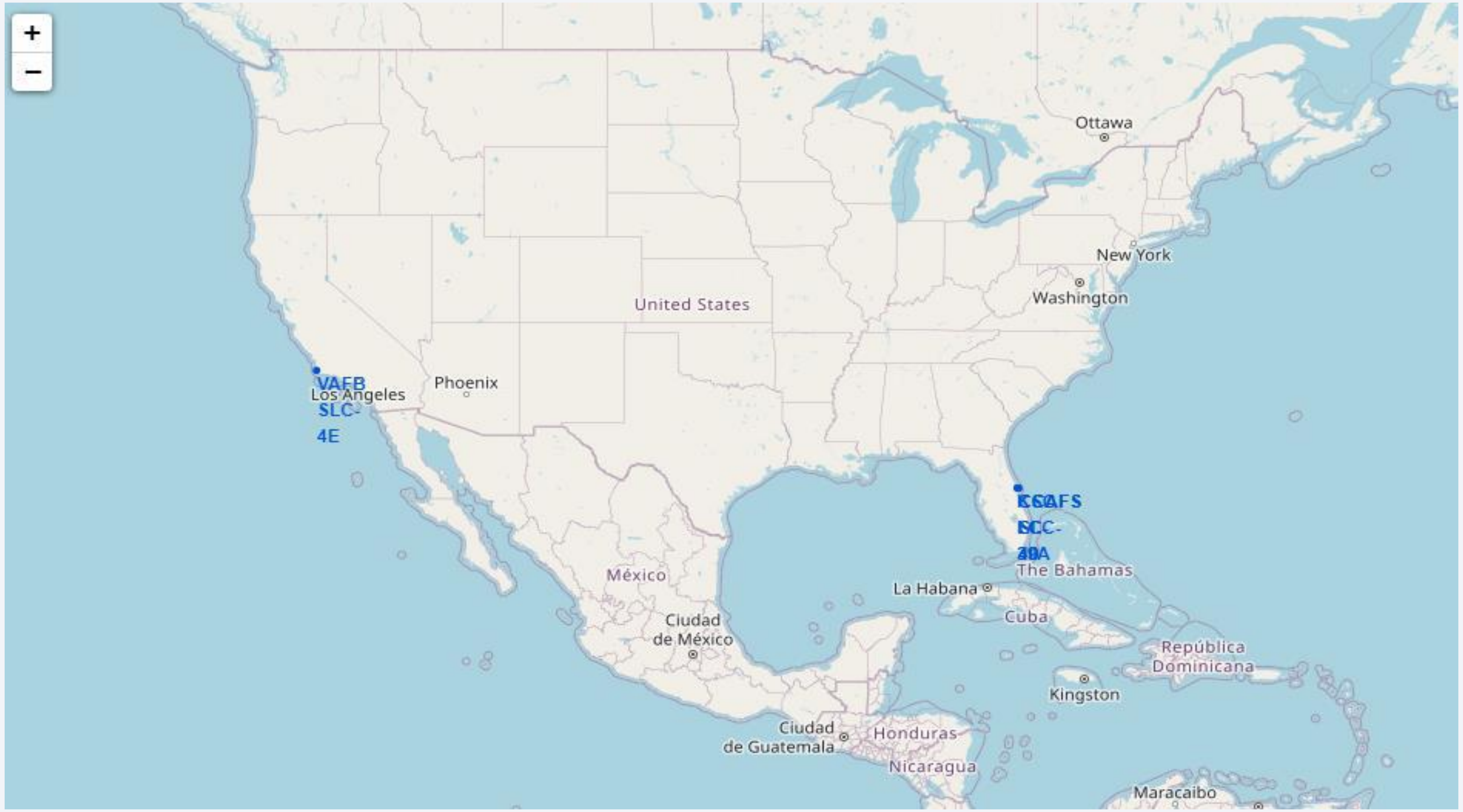
Map showing All Launch Sites

The map shows all the sites of Falcon 9 launch.

• We can see that the sites are located very close to the coastline as the failure rates of rockets are high (about 5-10%) and civilian areas must be avoided.

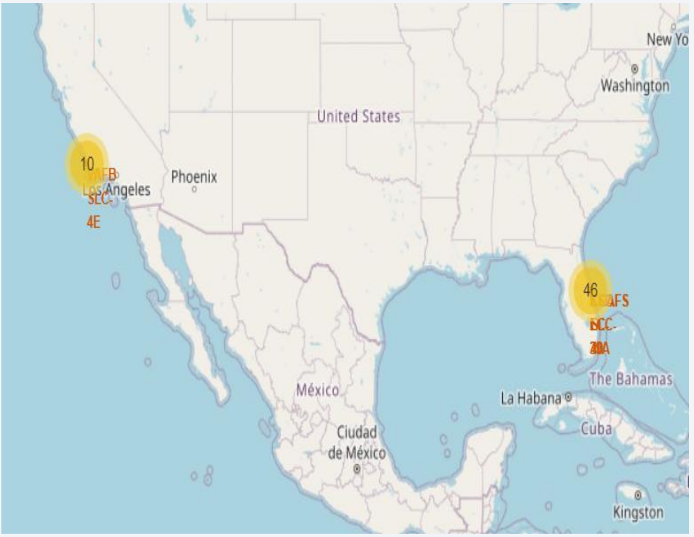
• 3 launch sites are located on the east coast of the US and 1 on the west coast

• all the sites are located close to the equator as it takes less fuel to launch a rocket from the equator

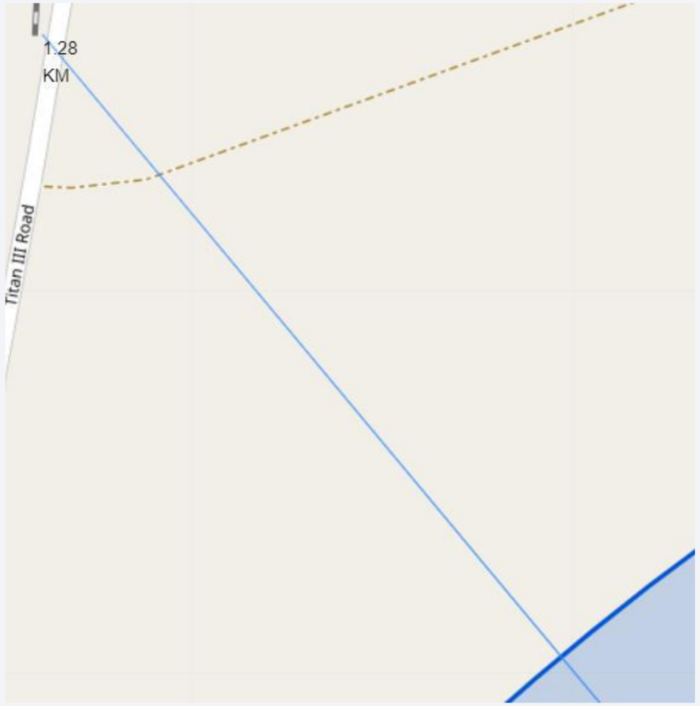


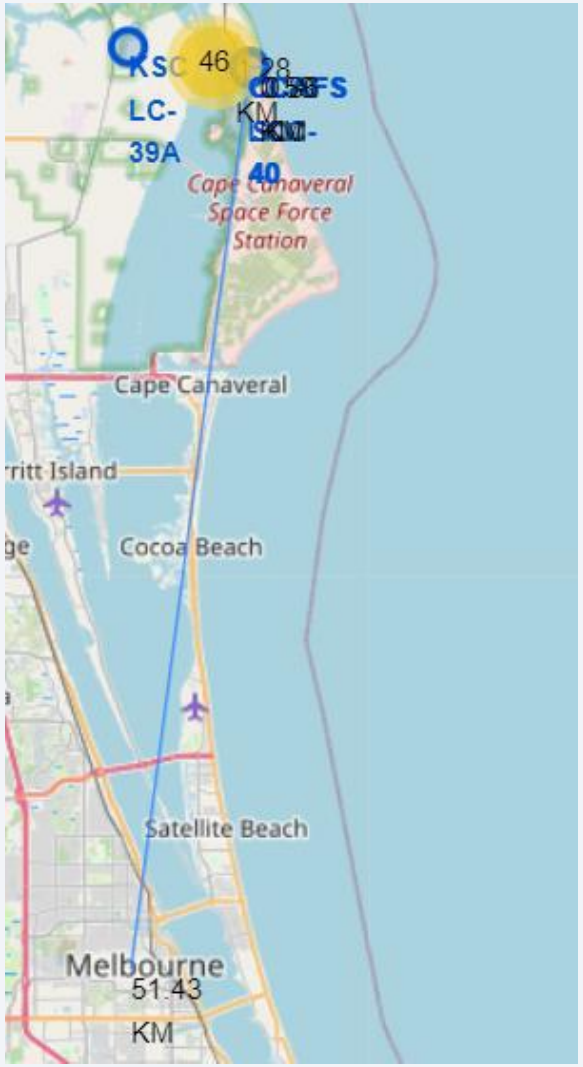
Success/failed launches marked for each site on the map

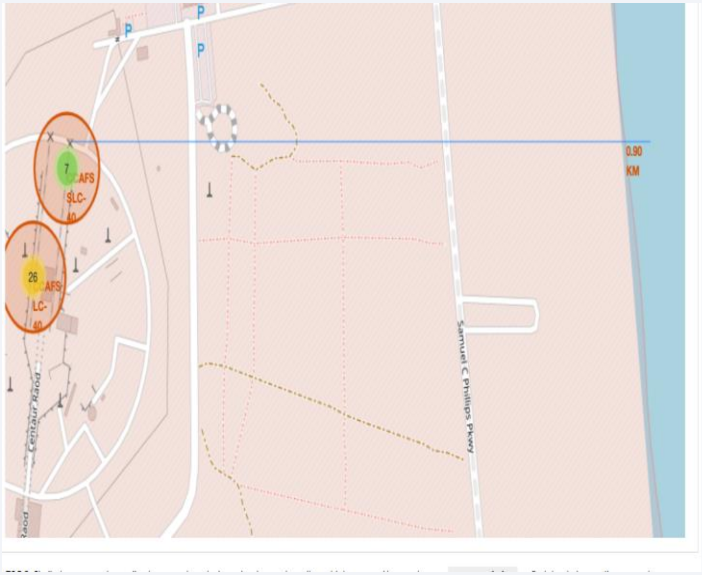
- From the colourable markers in marker clusters, it can be easily identified which launch sites have relatively high success rates.
- The green markers show successful launches and red markers show unsuccessful ones.
- Most launches happen on the east coast



The distances between a launch site to its proximities





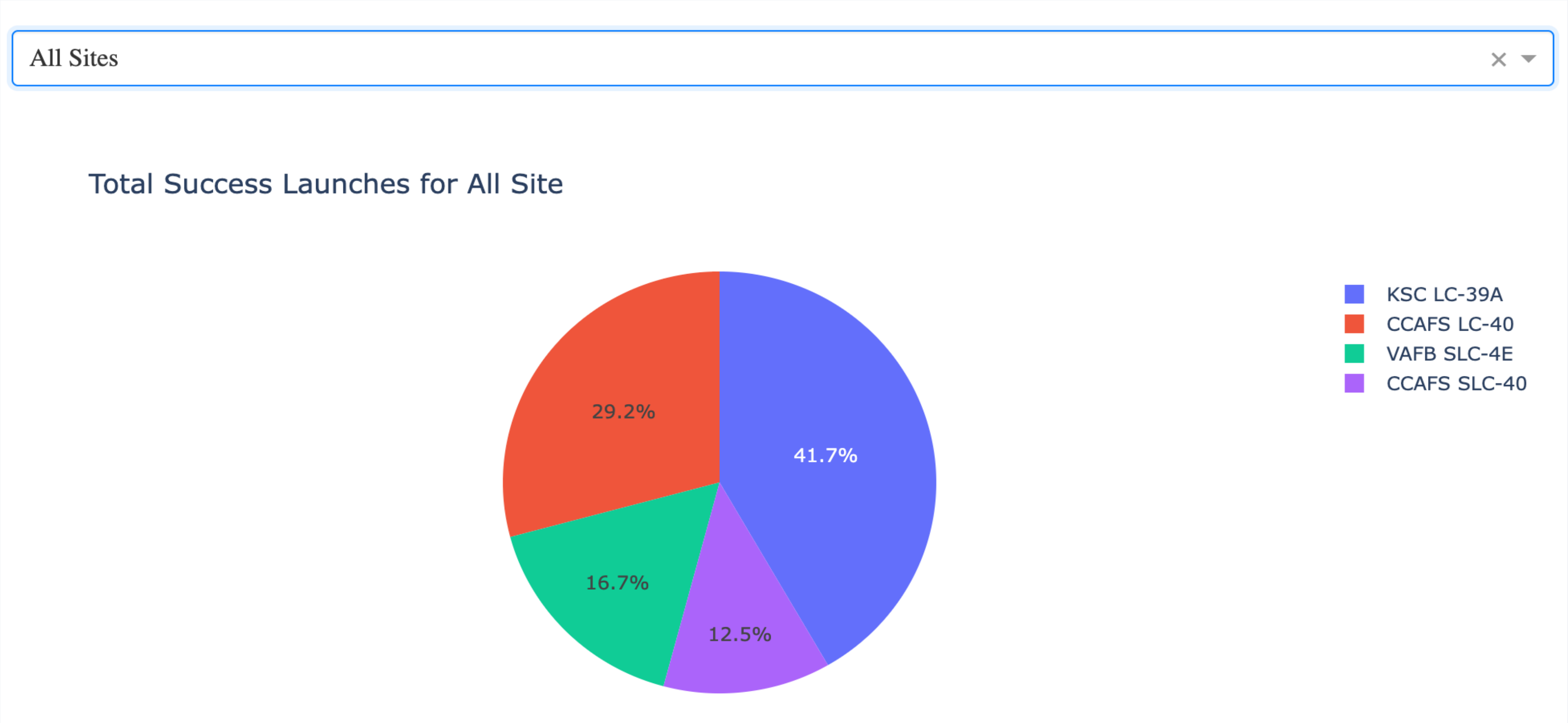


- launch sites are close proximity to railways, highways, coastline
- Launch sites keep certain distance away from cities

Section 4

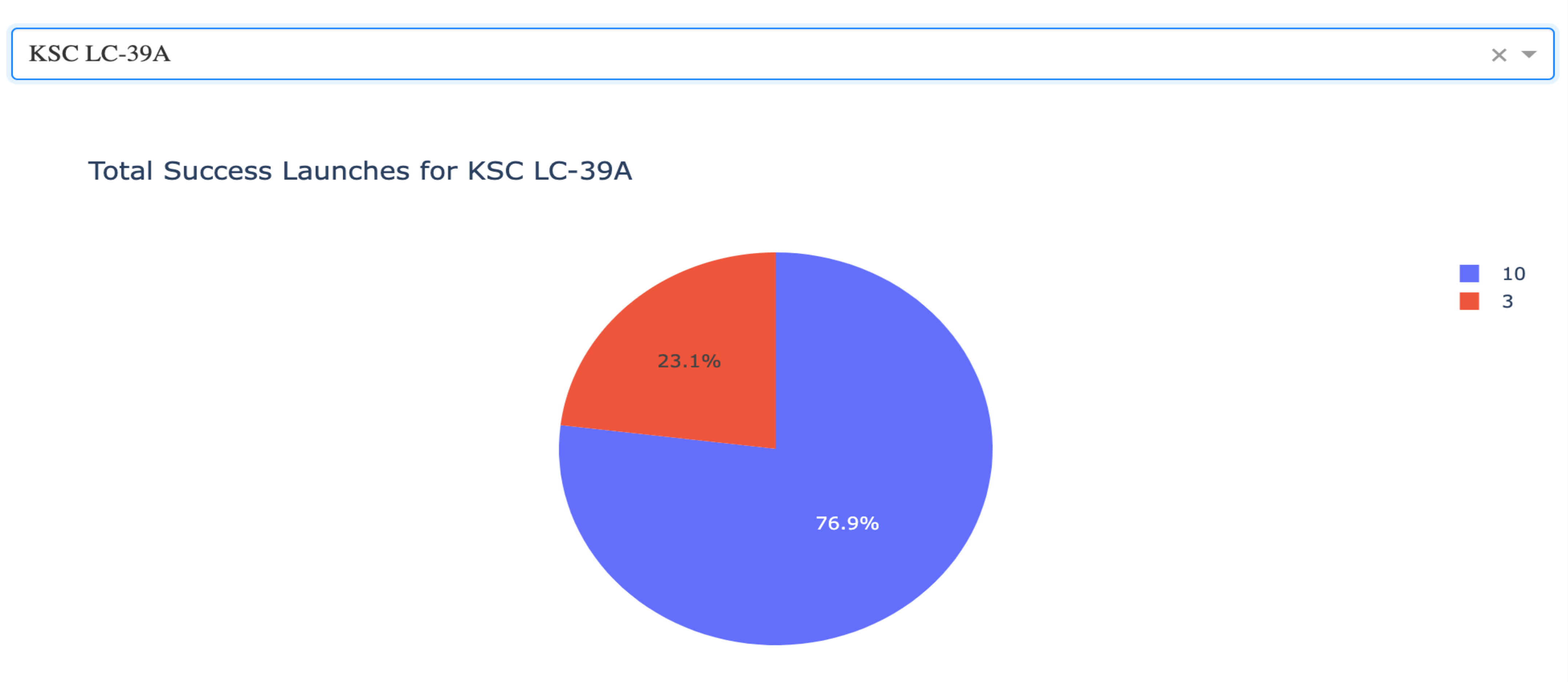
Build a Dashboard with Plotly Dash

Success Percentages for all Launch Sites



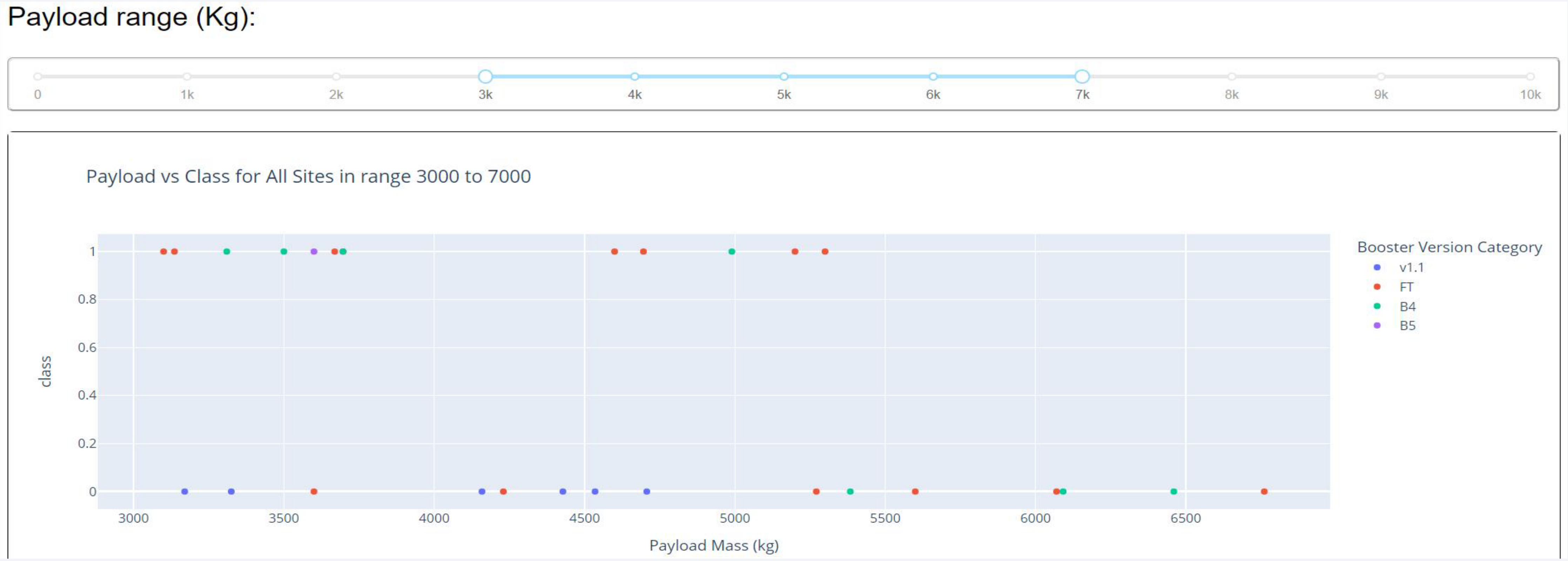
From this pie chart, we can see that Kennedy Space Center Launch Complex 39A (KSC LC-39A) had the highest share of a successful landing.

Percentages for Launch Site with Most Success Ratio



- From this pie chart, Kennedy Space Center Launch Complex 39A (KSC LC-39A) managed to achieve almost 77% of success (class 1) and 23% of landing (class 0) for Falcon 9 first-stage landing
- User can select data for all launch sites using the dropdown menu above the pie chart.

Payload vs Launch Outcome For All Sites



Payload Range 3000 to 7000 kg is selected:

- **B4 booster has high success rate**
- **FT booster has average success rate while having a high failure rate for high payload**

Payload range (Kg):

Payload Range 6000 to 10000 kg is selected:

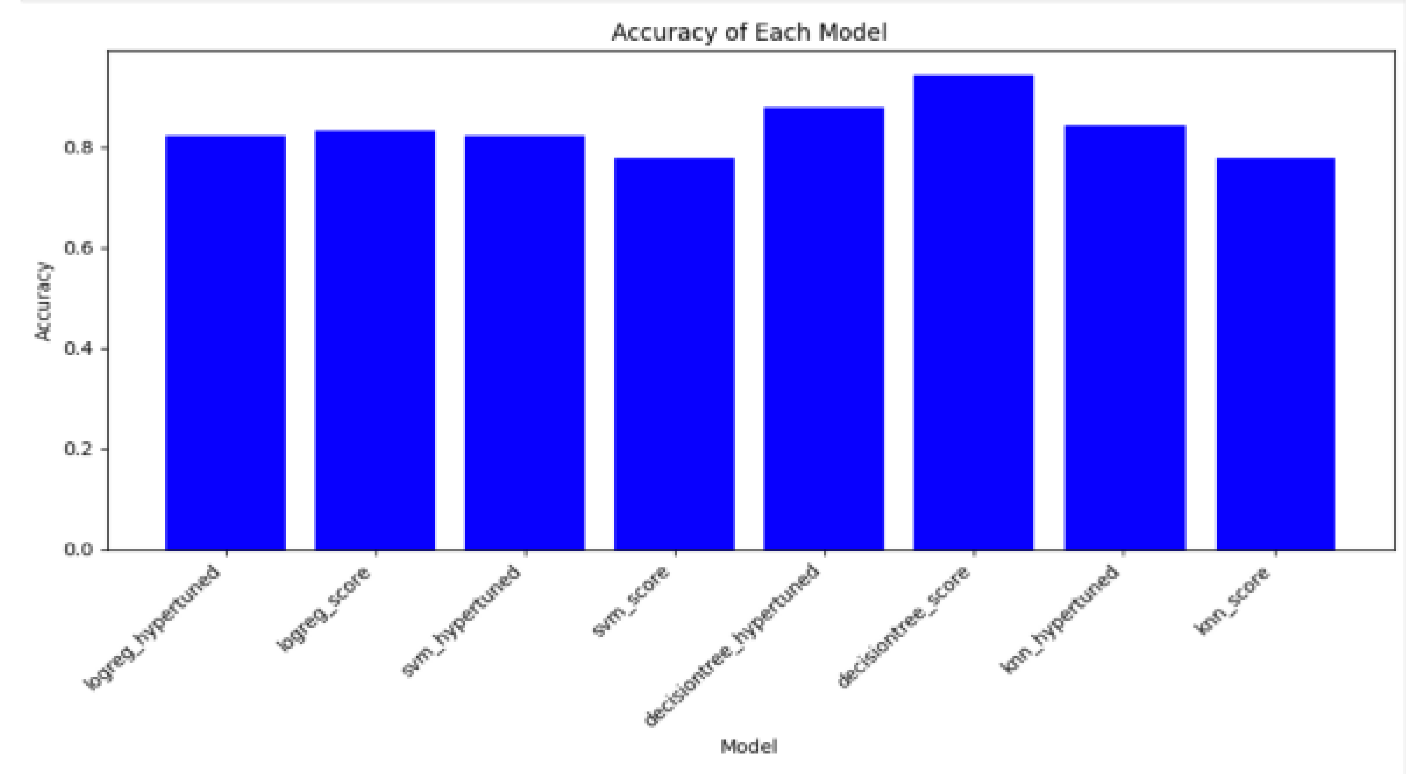
- **There is a high failure rate for this payload range**

Section 5

Predictive Analysis (Classification)

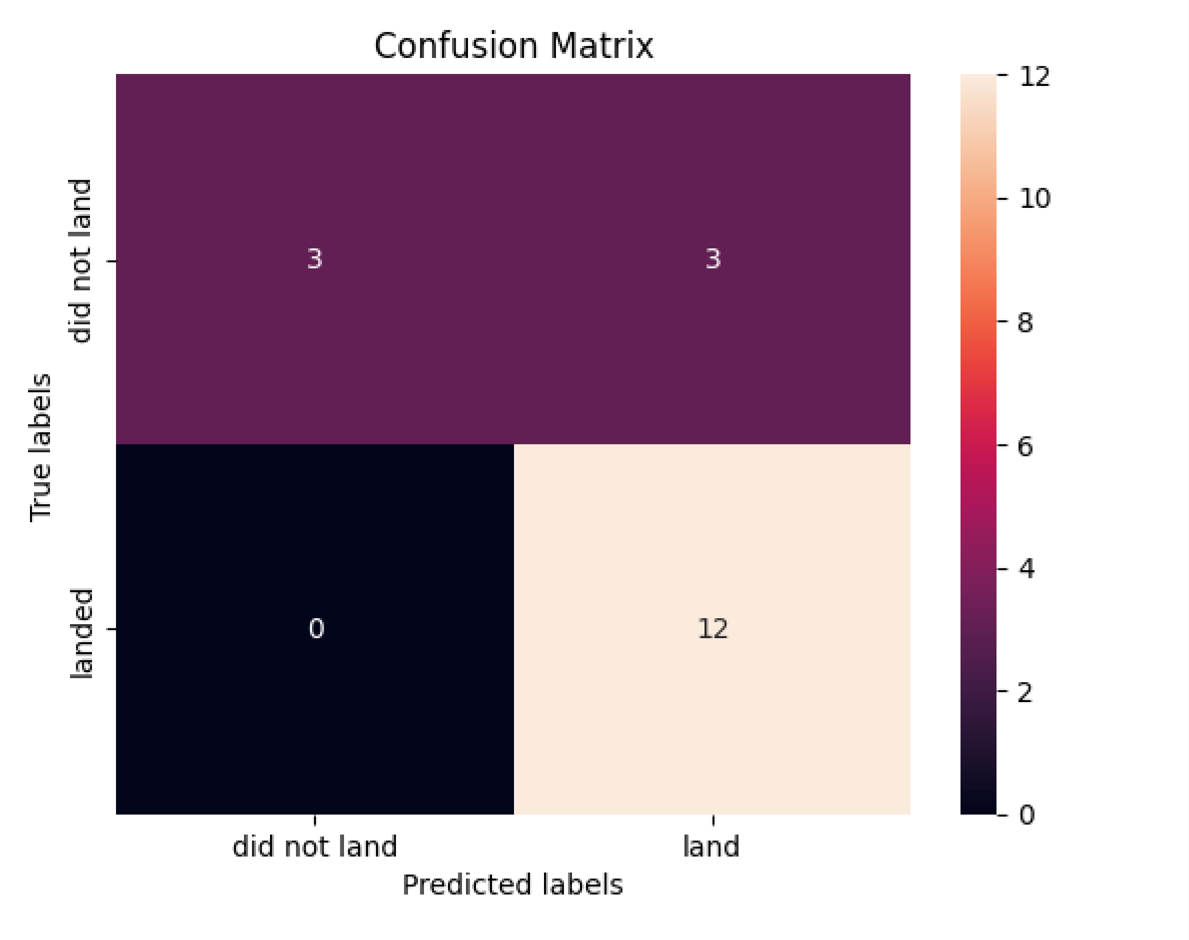
Classification Accuracy

- Visualized the accuracy of all classification models used in the project in a bar chart, as in the figure beside:
- We can see that Decision Tree Classifier has the highest accuracy.



Confusion Matrix of Decision Tree Model

- The confusion matrix above shows 12 true positives (correctly predicted), 3 true negatives (correctly predicted)
- 3 false positives (incorrectly predicted) and 0 false negatives (incorrectly predicted).



Conclusions

- Kennedy Space Center Launch Complex 39A (KSC LC-39A) had the highest share of successful landing
- For the orbit destination, ES-L1, GEO, HEO, and SSO have the best success rate.
- Low weighted payloads perform better than the heavier payloads
- Failure rate of new launches are low.
- Decision Tree Classifier is the best model for the problem and can be used to predict the success or failure of upcoming launches

Appendix

1. To make the requested JSON results more consistent, we will use the following static response object for this project:

```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json'
```

We should see that the request was successfull with the 200 status response code

```
response.status_code
```

```
200
```

2. Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia

https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

3.SQL query

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

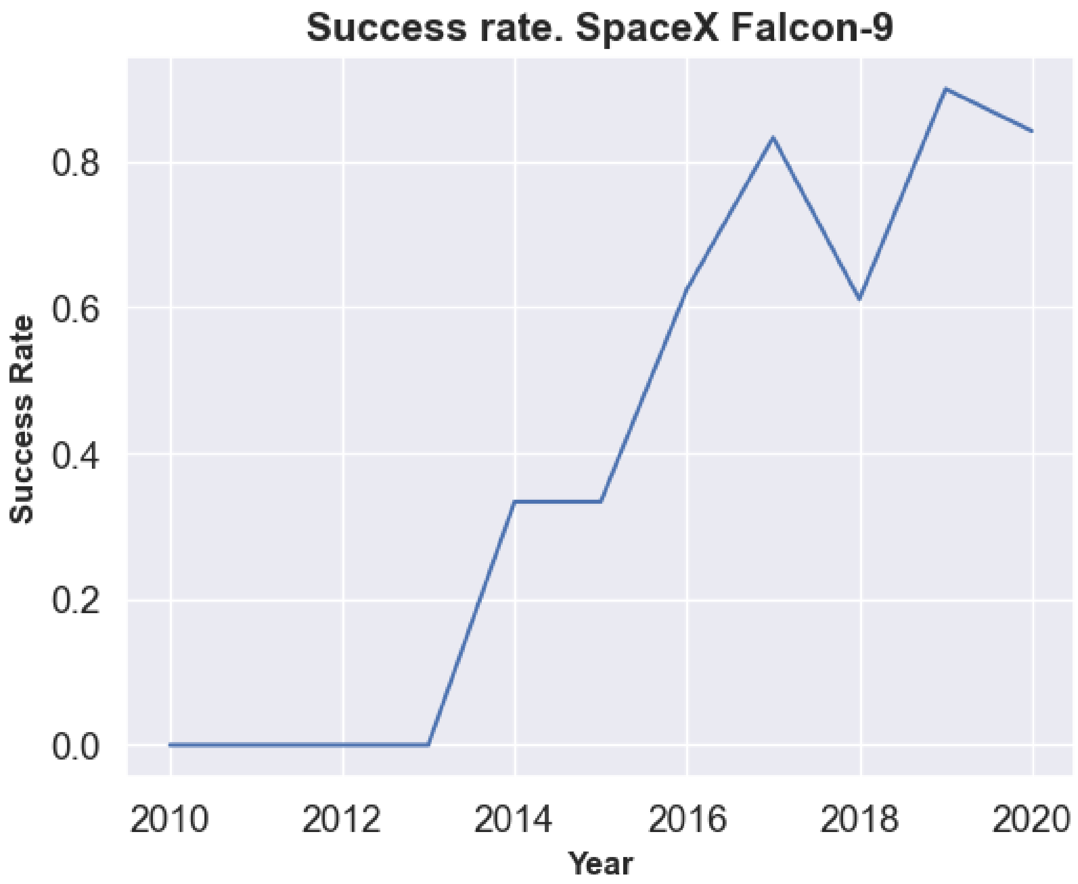
Hint:Use min function

```
%sql SELECT  MIN("Date") AS "First_Successful_Landing_On_Ground_Pad" FROM  SPACEXTBL WHERE "Landing_Outcome" = 'Success (ground pad)';
```

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

First_Successful_Landing_On_Ground_Pad
2015-12-22

4.



You can observe that the success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.

Thank you!