IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

<Name>
<Date>

# Outline

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

# Executive Summary

- Summary of using methodologies
  - Data Collection using with a SpaceX API and web scraping
  - Exploratory Data Analysis with data wrangling and search database
  - data visualization and create interactive map
  - Machine Learning Prediction
- Summary of all results
  - Possible to collected valuable data from public sources
  - Exploratory Data Analysis is possible to identify which features are the be  to better predict success of launchings
  - Use machine Learning predictions showed the success or failure of the launch can be predicted

# Introduction

- Project background and context
  - In the space industry, where governments have traditionally played a key role, there is now more room for private companies to enter the field
- Problems you want to find answers
  - Analyze the results of SpaceX, the company that has already taken the lead, to determine whether or not they will be able to enter the field

Section 1

# Methodology

# Methodology

## Executive Summary

- Data collection methodology:
  Get data from the following sources:
  - Get SpaceX launch data from API
  - Get web scraping data from wikipedia
- Perform data wrangling

  Data obtained from API or the web is formatted using the following processes:
  - Data validation
  - Feature selection
  - Missing value handling
  - Replace categorical data with numeric values

# Methodology

## Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
  - Aggregate data by axes such as “Launch Site" and "Landing Outcome" to understand the overall trend of the data
- Perform interactive visual analytics using Folium and Plotly Dash
  - Marking launch sites on maps using Folium
  - Visualizing quantitative information using Dash

# Methodology

## Executive Summary

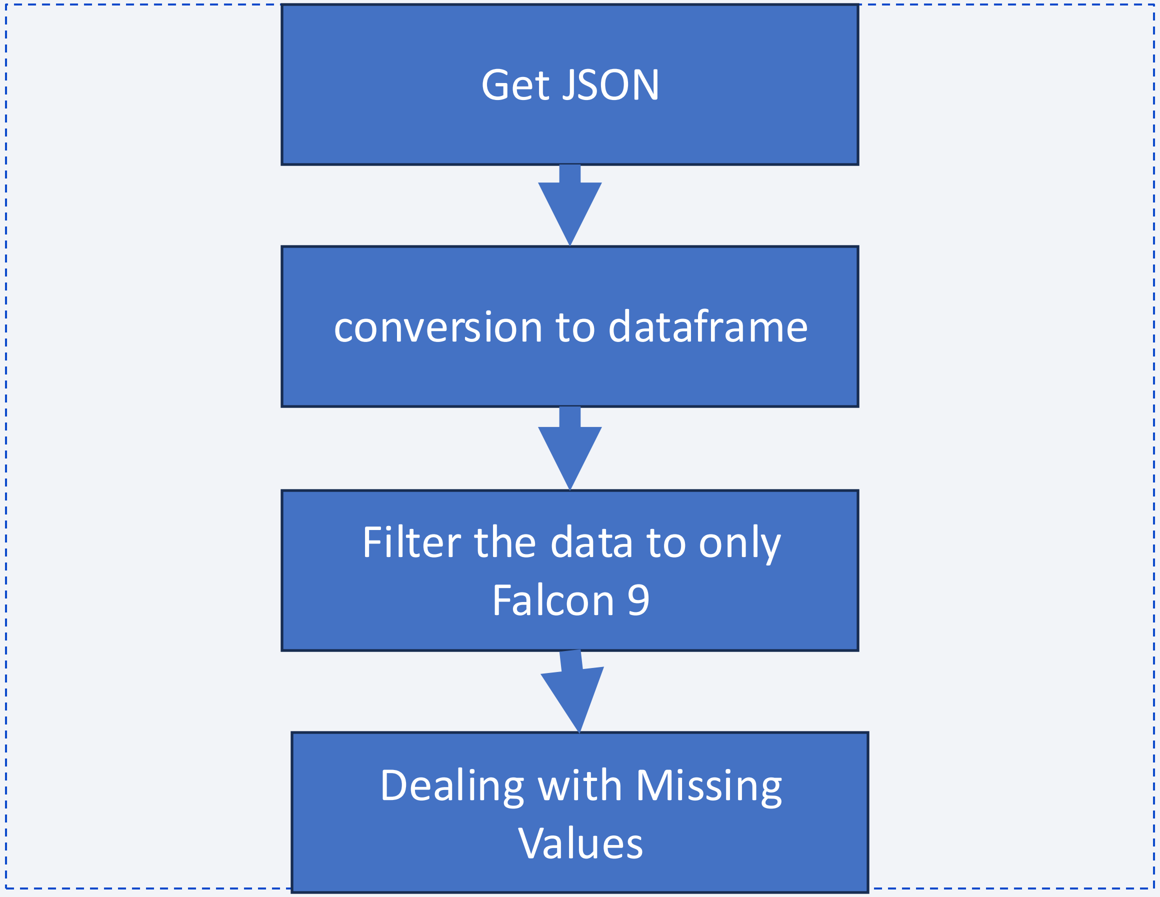
- Perform predictive analysis using classification models
    - How to build, tune, evaluate classification models
        - Cast the target variable to a numeric type
        - Normalize features
    - Construct a predictive model using the following techniques:
        - Logistic Regulation
        - Decision Tree
        - SVM
        - KNN

# Data Collection

- Data source
  - SpaceX API(JSON)
    - https://api.spacexdata.com/v4/launches/past
  - Wikipedia(Scraping WEB)
    - https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches

# Data Collection – SpaceX API

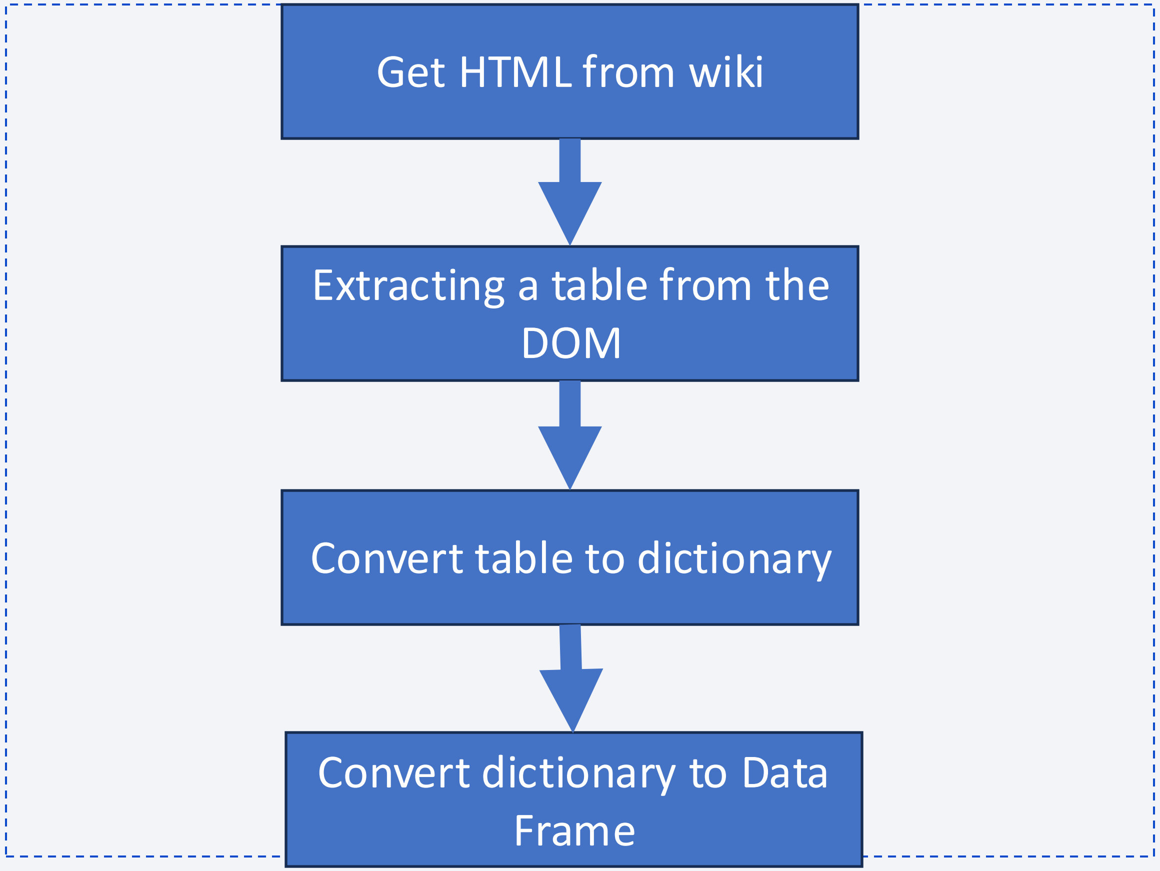
- Request the API provided by SpaceX to obtain JSON format data
- Parse the JSON format data and convert it to a Data Frame
- Complete missing values and extract Falcon 9 data




SpaceX API calls notebook (https://github.com/nekonabesan/IBM_DataScience/blob/main/Capstone/jupyter-labs-spacex-data-collection-api.ipynb)

# Data Collection - Scraping

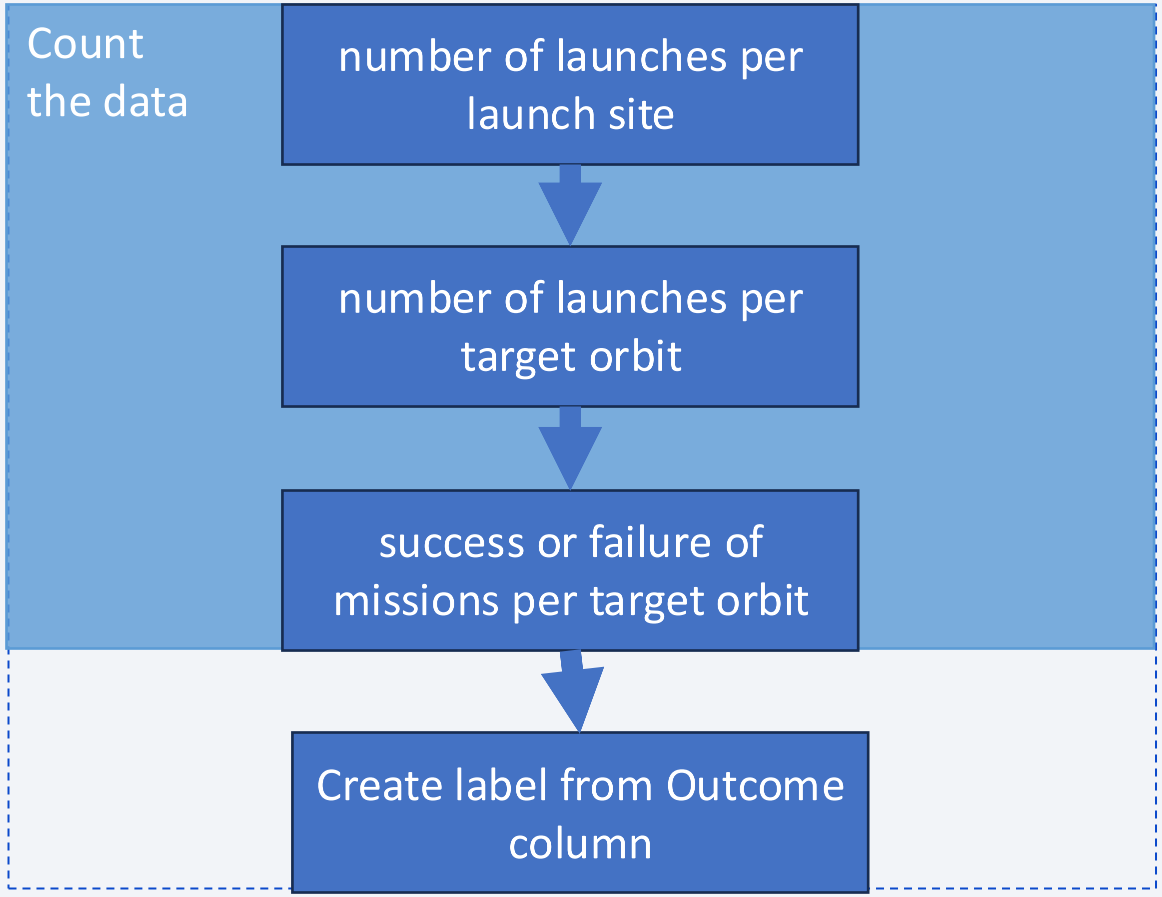
- Request the "Falcon9 Launch" page on Wikipedia and get the HTML
- Extract all column and variable names from the HTML table header
- convert to data frame by parsing the launch HTML tables




Web scraping notebook (https://github.com/nekonabesan/IBM_DataScience/blob/main/Capstone/jupyter-labs-webscraping.ipynb)

# Data Wrangling

- Counting the number of launches per launch site
- Counting the number of launches per target orbit
- Counting the success or failure of missions per target orbit
- Create a landing outcome label from Outcome column

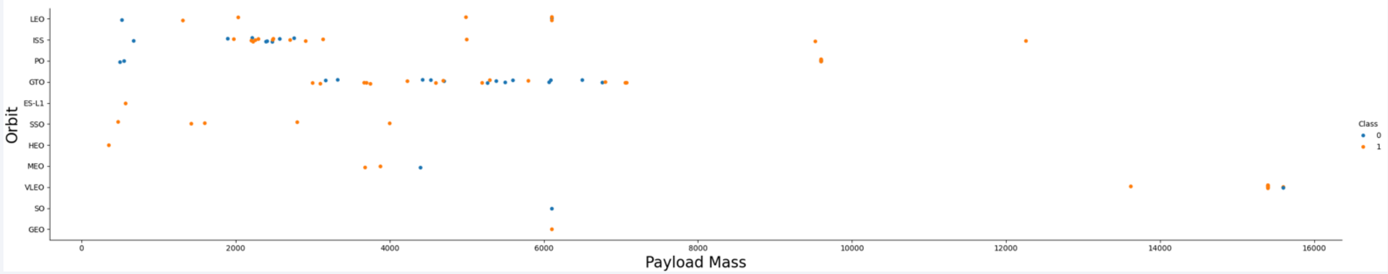


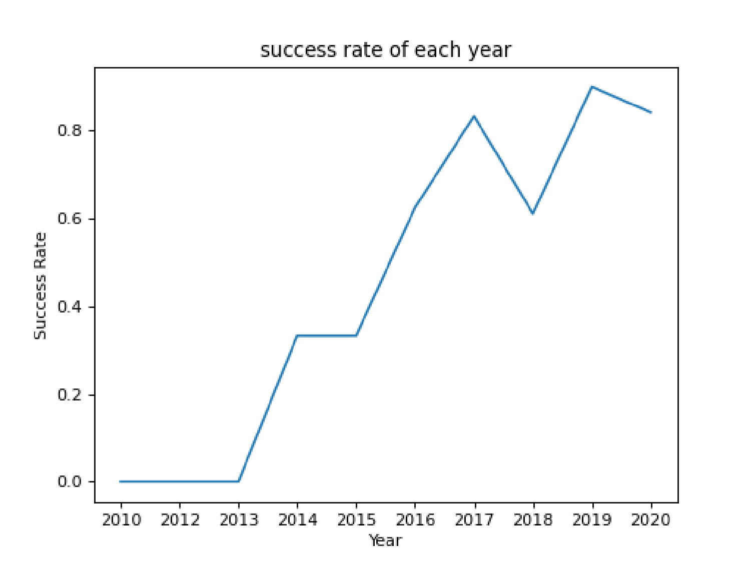

Data wrangling notebook (https://github.com/nekonabesan/IBM_DataScience/blob/main/Capstone/labs-jupyter-spacex-Data%20wrangling.ipynb)

# EDA with Data Visualization

- Visualize and compare relationships between features such as "Launch Site", "Payload Mass", and "Orbit Type”

- Finally, visualize the annual launch success rate



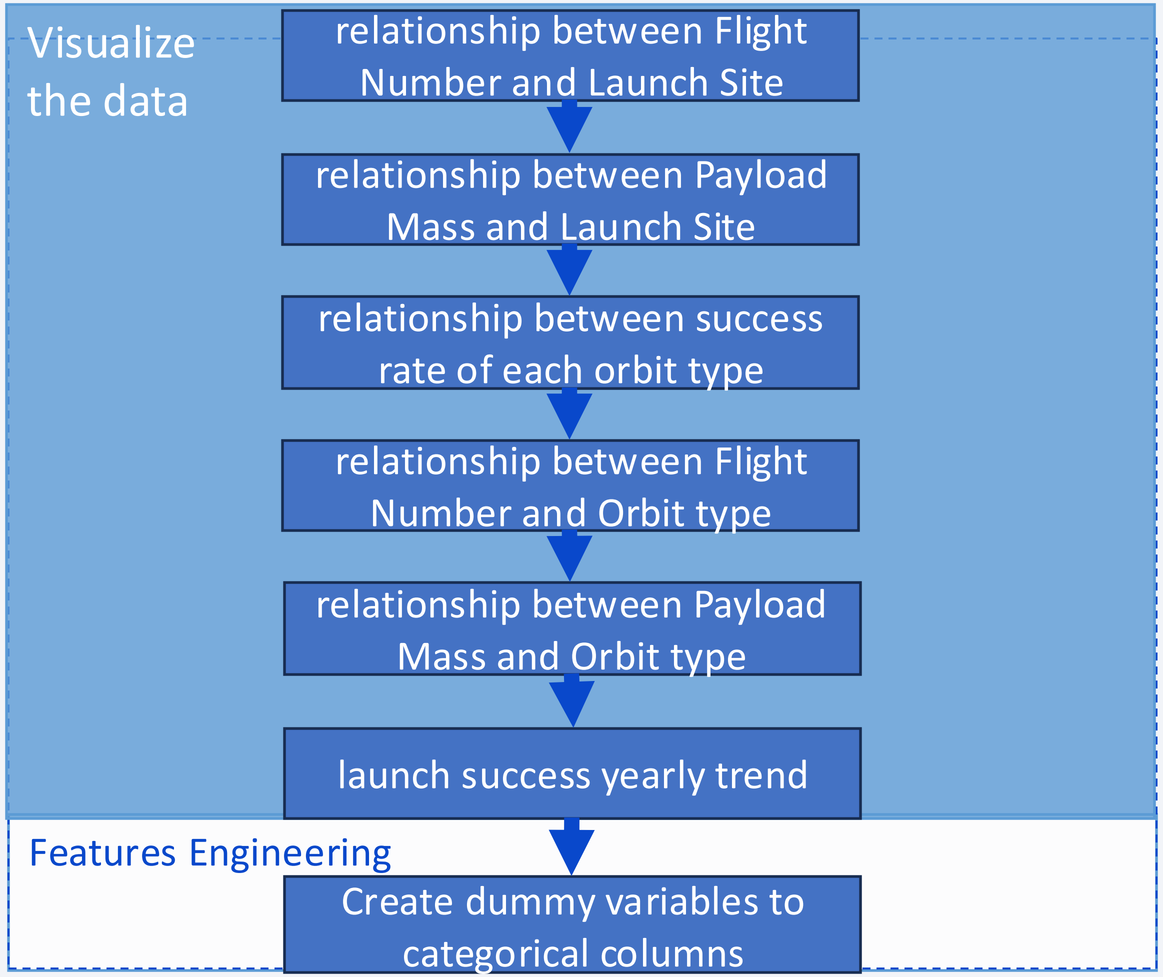






- EDA with data visualization notebook (https://x.gd/FdBSn)

# EDA with SQL

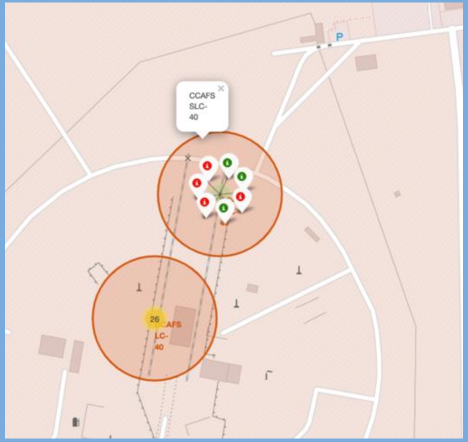
- List and visualize some numbers from launch data
- Create List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year of 2015
- Caluculate Rank the count of landing outcomes such as Failure or Success between the date 2010-06-04 and 2017-03-20
- EDA with data visualization notebook (https://x.gd/UzWQt)



# Build an Interactive Map with Folium

- Mark all launch sites on a map with use folium circle
- Mark the success or failed launches for each site on the map
- Calculate the distances between a launch site to its proximities
- Visualize the distance between the launch site and major roads or urban areas on a map and analyze the location

Mark information
On map

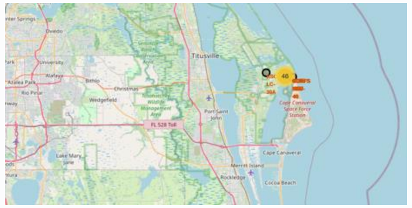


Mark all launch sites on a map

Mark the success or failed launches

Calculate the distances

Analize location

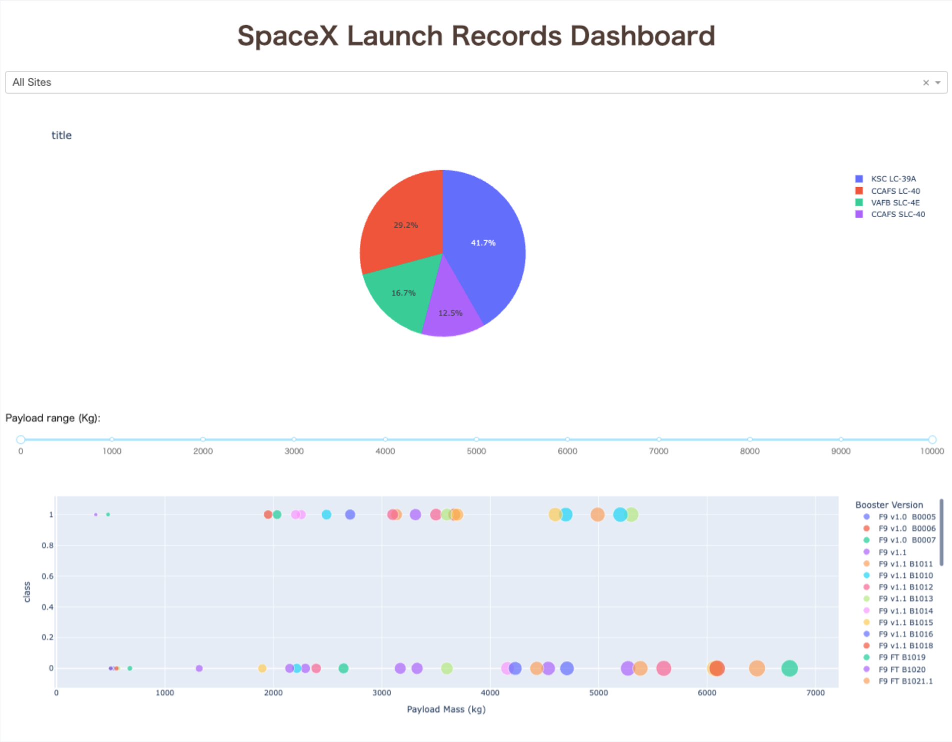


Analyze the location

create interactive map notebook
(https://github.com/nekonabesan/IBM_DataScience/blob/main/Capstone/lab_jupyter_launch_site_location.ipynbß)

# Build a Dashboard with Plotly Dash

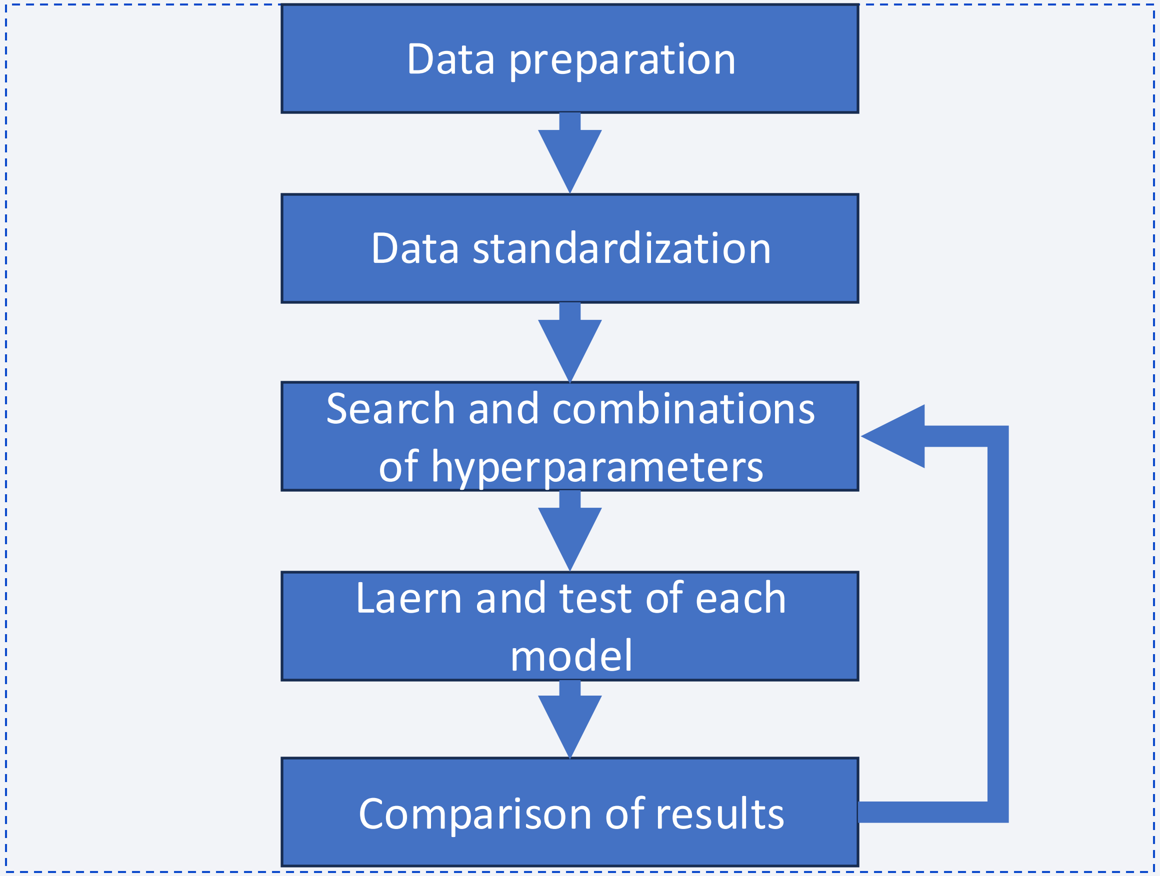
- Implementation a dropdown list to enable Launch Site selection, In addition to that Implement a callback function associated with the list selection state
- Implementation a pie chart to show the total successful launches count for all sites, If a specific launch site was selected, show the Success vs. Failed counts for the site
- Implementation a slider to select payload range slider, and Implementation callback function with Linked to slider action




create Dash app script (https://github.com/nekonabesan/IBM_DataScience/blob/main/Capstone/spacex-dash-app.py)

# Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- Prepare and standardize the data, in addition to that the data set split training and test data.
- Training a model and alculate the accuracy on the test data using the method score:



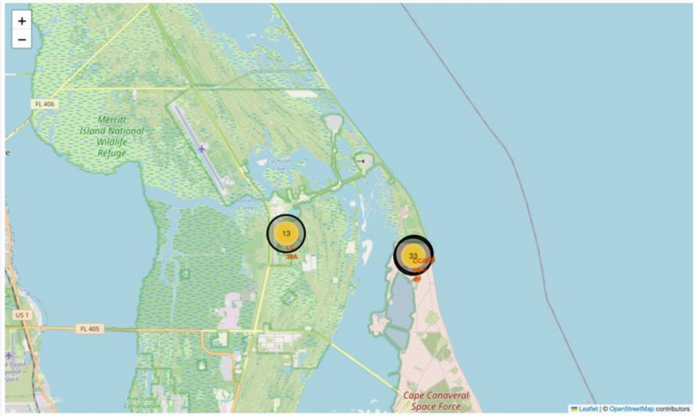

create Dash app script
(https://github.com/nekonabesan/IBM_DataScience/blob/main/Capstone/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb)

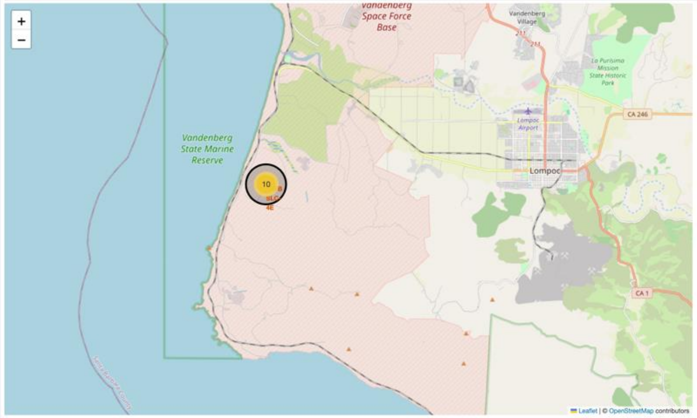
# Results

- Summary of exploratory data analysis results
  - SpaceX conducted 90 launches from four launch sites between 2010 and 2020, with a success rate of just under 67%.
  - The orbit types with the highest success rates are SSO, HEO, GEO, and ES-L1.
  - From 2013 to 2020, launch success rates have consistently increased.

# Results

- Interactive analytics of launch site location
  - The launches were carried out from the following facilities:
    - CCAFS LC-40 (Florida)
    - CCAFS SLC-40(Florida)
    - KSC LC 39A(Florida)
    - VAFB SLC-4E(Santa Barbara County CA)
  - The launch site is located next to the coast and in the southern United States, relatively close to the equator

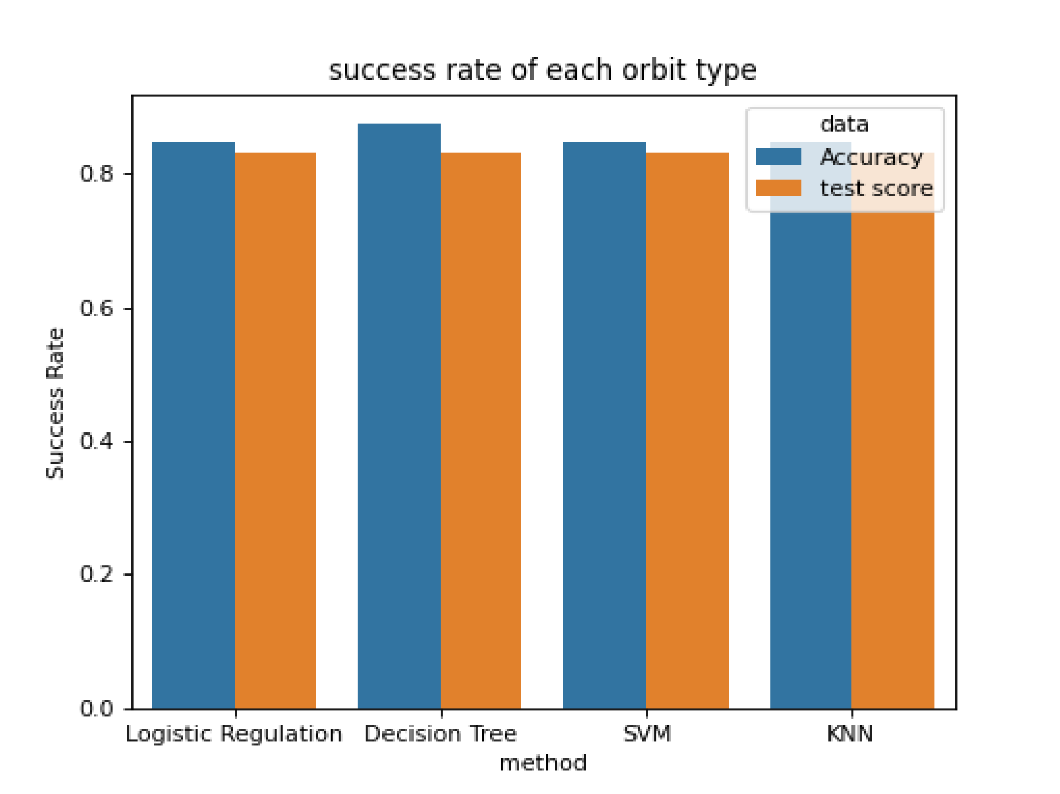


# Results

- Predictive analysis results
  - Construct a predictive model using four methods
    - Logistic Regulation
    - Decision Tree
    - SVM
    - KNN
  - The amount of data was small, so there was no significant difference in the results.

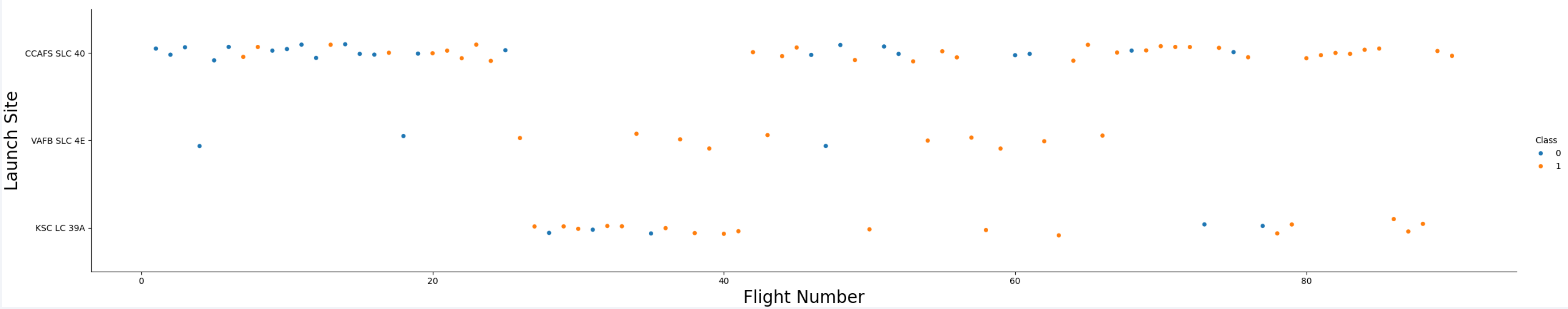
| | method | Accuracy | test score |
|---|---|---|---|
| 0 | Logistic Regulation | 0.846429 | 0.833333 |
| 1 | Decision Tree | 0.875000 | 0.833333 |
| 2 | SVM | 0.848214 | 0.833333 |
| 3 | KNN | 0.848214 | 0.833333 |



Section 2

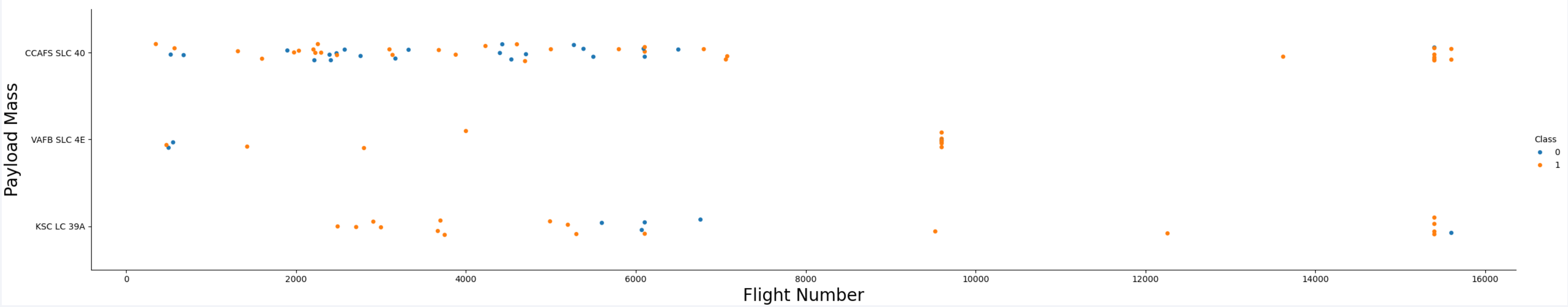
# Insights drawn from EDA

# Flight Number vs. Launch Site




- The launch site with the most attempts is CCAF5 SLC 40.
- The launch site with the highest success rate is CCAF5 SLC 40
- Success rate is improving over time

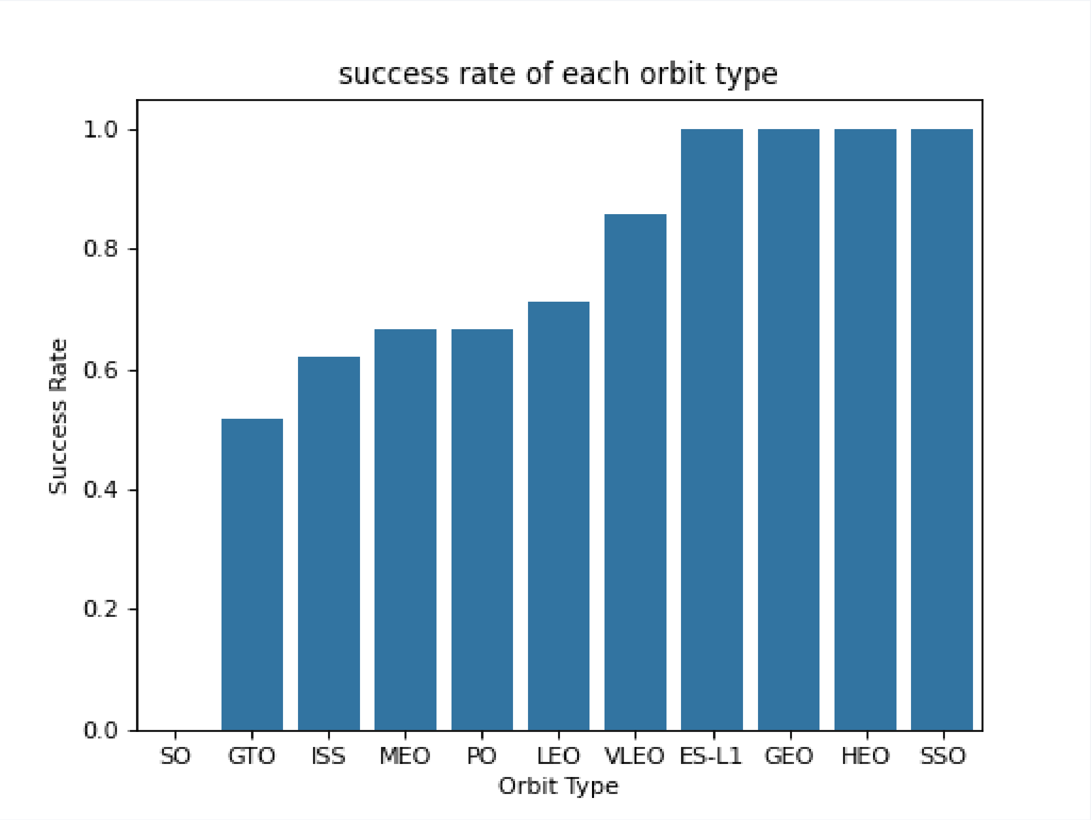
# Payload vs. Launch Site



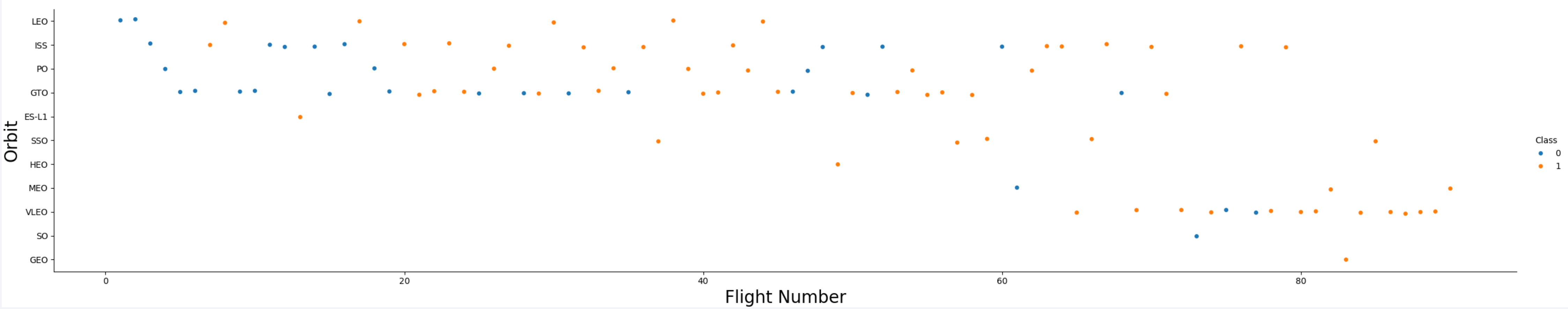

- The positive efficiency improves as the payload increases
- The above is thought to be due to the payload being increased as the booster matures

# Success Rate vs. Orbit Type

- The most successful orbit types are:
  - SSO
  - HEO
  - GEO
  - ES-L1

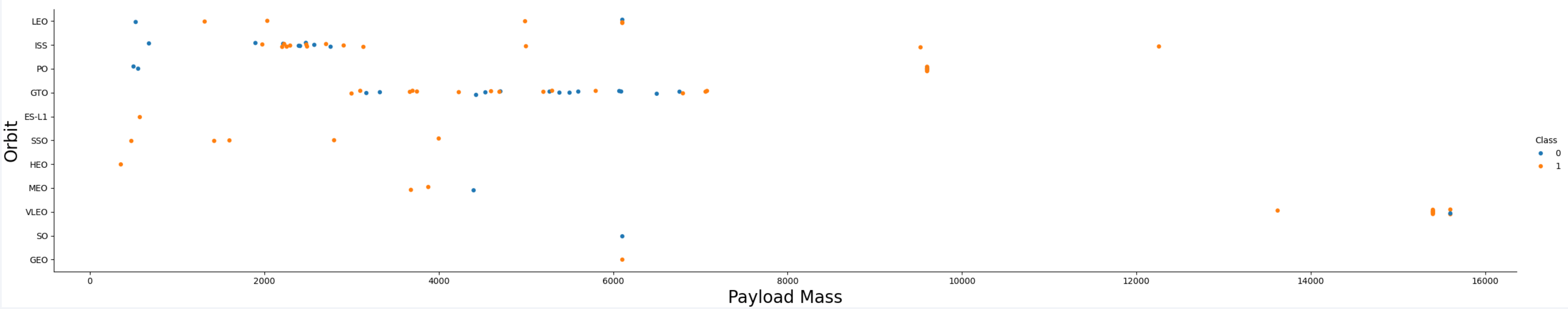


# Flight Number vs. Orbit Type




- Regardless of orbit type, orbit insertion success rates have improved over time

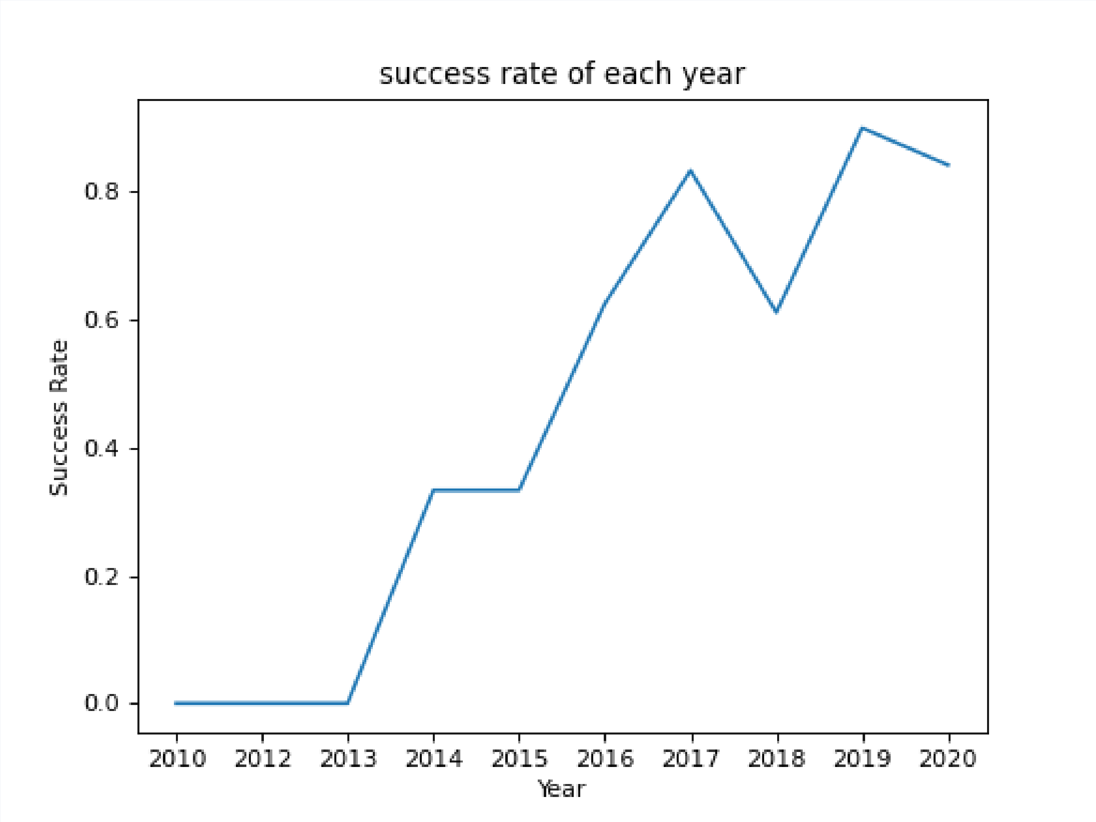
# Payload vs. Orbit Type




- Looking at the GTO and ISS orbits, which have been tried many times, there is no clear difference in success rate between orbit types and payload axes

# Launch Success Yearly Trend

- The overall success rate of launches is improving year by year.



# All Launch Site Names

- Names of the unique launch sites

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

- query result

SELECT `Launch_Site` FROM SPACEXTBL GROUP BY `Launch_Site`

# Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

| 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 |

- Extracted to launch data of Cape Canaveral Space Launch Complex Sites

```
SELECT * FROM SPACEXTBL WHERE `Launch_Site` LIKE 'CCA%' LIMIT 5
```

# Total Payload Mass

- total payload mass carried by boosters launched by NASA (CRS)

| total payload mass[kg] |
|---:|
| 48213 |

- Extract from `Booster_Version` column with WHERE clause

```
SELECT SUM(`PAYLOAD_MASS__KG_`) AS `total payload mass[kg]`
FROM SPACEXTBL WHERE `Customer` LIKE '%NASA (CRS)%'
```

# Average Payload Mass by F9 v1.1

Average payload mass carried by booster version F9 v1.1

| payload mass[kg] |
| ---: |
| 2928.4 |

- query result with a short explanation here

```
SELECT AVG(`PAYLOAD_MASS__KG_`) AS `payload mass[kg]` FROM SPACEXTBL
WHERE `Booster_Version` = 'F9 v1.1'
```

# First Successful Ground Landing Date

- First successful landing outcome on ground pad

| first succesful date |
| ---: |
| 2015-12-22 |

- `Date` Sort in ascending order and filter by LIMIT clause

```
SELECT `Date` AS `first succesful date` FROM SPACEXTBL
WHERE `Landing_Outcome` = 'Success (ground pad)' ORDER BY `Date` ASC LIMIT 1
```

# Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of 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 |

- Present your query result with a short explanation here

```
SELECT `Booster_Version` AS `Booster Version` FROM SPACEXTBL
WHERE `Landing_Outcome` = 'Success (drone ship)’
AND (`PAYLOAD_MASS__KG_` >= 4000 AND `PAYLOAD_MASS__KG_` <= 6000)
```

# Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

| Outcome | counts |
|---|---|
| Failure | 1 |
| Success | 100 |

- GROUP BY with `Outcome` and use CASE statement to separate cases

```
SELECT
CASE
   WHEN `Mission_Outcome` LIKE '%Success%' THEN 'Success'
   ELSE 'Failure'
END AS `Outcome`
,COUNT(`Mission_Outcome`) AS `counts`
FROM SPACEXTBL GROUP BY `Outcome`
```

# Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Specify the list obtained by the subquery as the IN condition

```
SELECT `Booster_Version`,`PAYLOAD_MASS__KG_` AS
`MAX_PAYLOAD` FROM SPACEXTBL AS TBL1
WHERE `MAX_PAYLOAD` IN (SELECT
MAX(`PAYLOAD_MASS__KG_`) AS `MAX_PAYLOAD` FROM
SPACEXTBL AS TBL2)
GROUP BY `Booster_Version`
ORDER BY `Booster_Version`
```

| Booster_Version | MAX_PAYLOAD |
|---|---|
| F9 B5 B1048.4 | 15600 |
| F9 B5 B1048.5 | 15600 |
| F9 B5 B1049.4 | 15600 |
| F9 B5 B1049.5 | 15600 |
| F9 B5 B1049.7 | 15600 |
| F9 B5 B1051.3 | 15600 |
| F9 B5 B1051.4 | 15600 |
| F9 B5 B1051.6 | 15600 |
| F9 B5 B1056.4 | 15600 |
| F9 B5 B1058.3 | 15600 |
| F9 B5 B1060.2 | 15600 |
| F9 B5 B1060.3 | 15600 |

# 2015 Launch Records

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

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

- SELECT SPACEXTBL and edit `Date` column

```
SELECT
substr(Date, 6,2) AS `month`
,`Landing_Outcome`
,`Booster_Version`
,`Launch_Site`
FROM SPACEXTBL
WHERE substr(Date,0,5)='2015' AND `Landing_Outcome` = 'Failure (drone ship)'
```

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

- Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20
- SELECT SPACEXTBL and sort `count` as DESC

```
SELECT
COUNT(*) AS `count`
,`Landing_Outcome`
FROM SPACEXTBL
WHERE `Date` >= '2010-06-04' AND '2017-03-20' >= `Date`
GROUP BY `Landing_Outcome`
ORDER BY `count` DESC
```

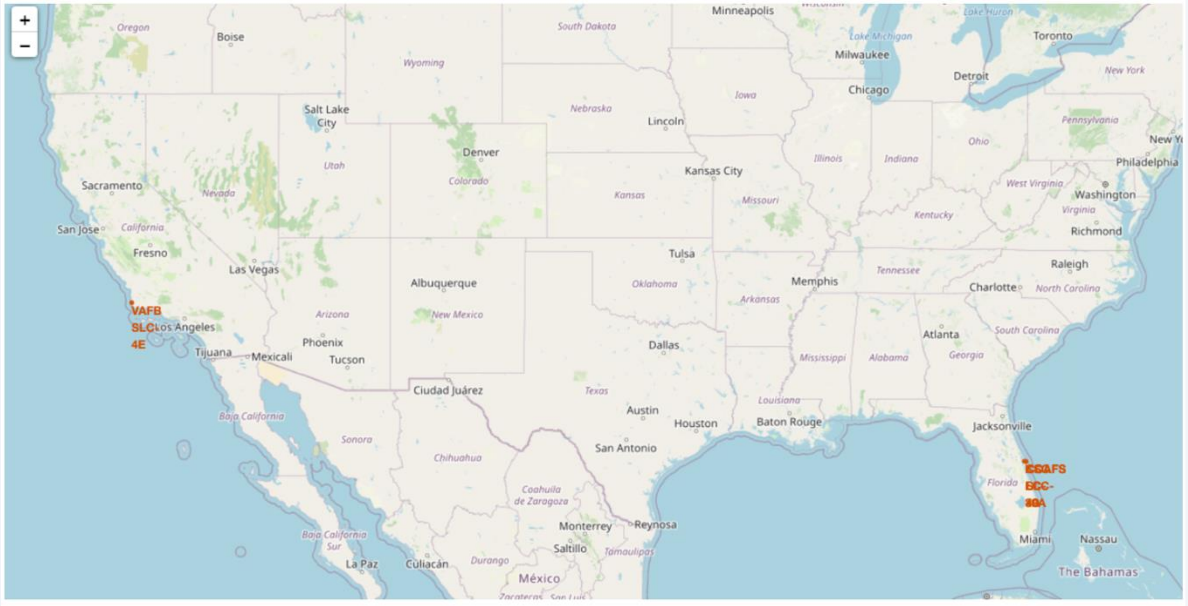
| count | Landing_Outcome |
|---:|---:|
| 10 | No attempt |
| 5 | Success (drone ship) |
| 5 | Failure (drone ship) |
| 3 | Success (ground pad) |
| 3 | Controlled (ocean) |
| 2 | Uncontrolled (ocean) |
| 2 | Failure (parachute) |
| 1 | Precluded (drone ship) |

Section 3

# Launch Sites Proximities Analysis

# Drawing launch sites on folium map

- Folium map on Launch site location
- Four launch sites are marked on the map.
- The launch site is located in the most favorable location in the United States for orbital launch.

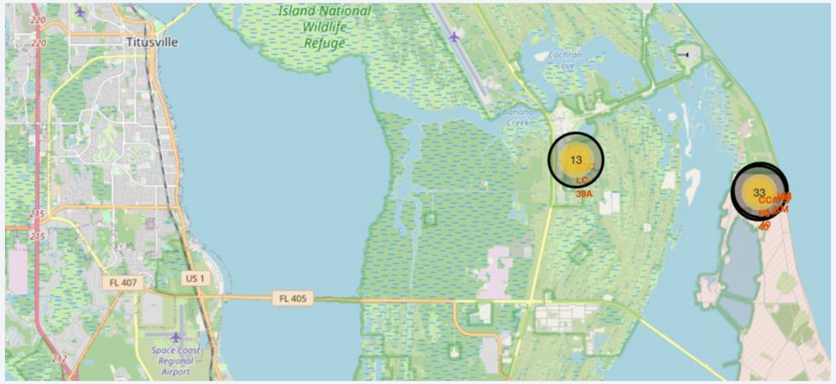


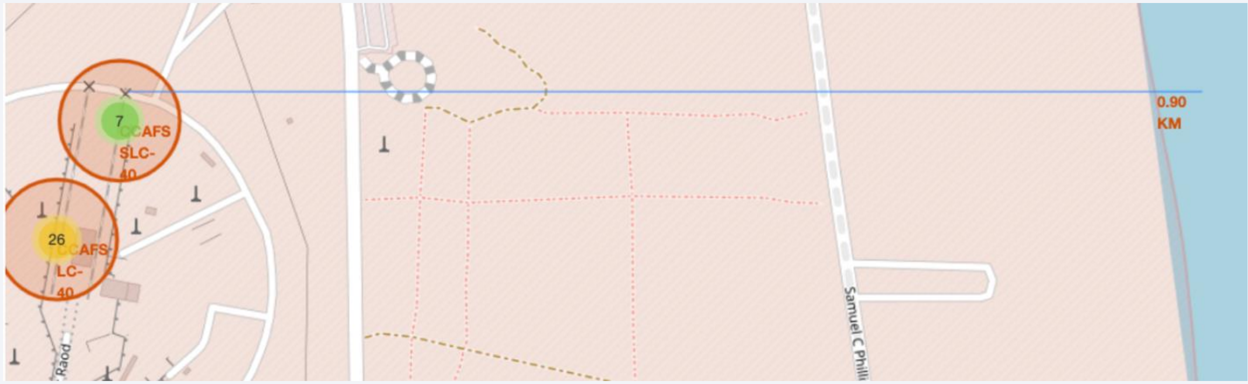
# Launch Outcomes on Sites

- success or failure of each launch outcome on draw folium map


# Location and environment of launch sites

- Visualize the relationship between launch sites and important infrastructure such as roads and railways, as well as the surrounding environment



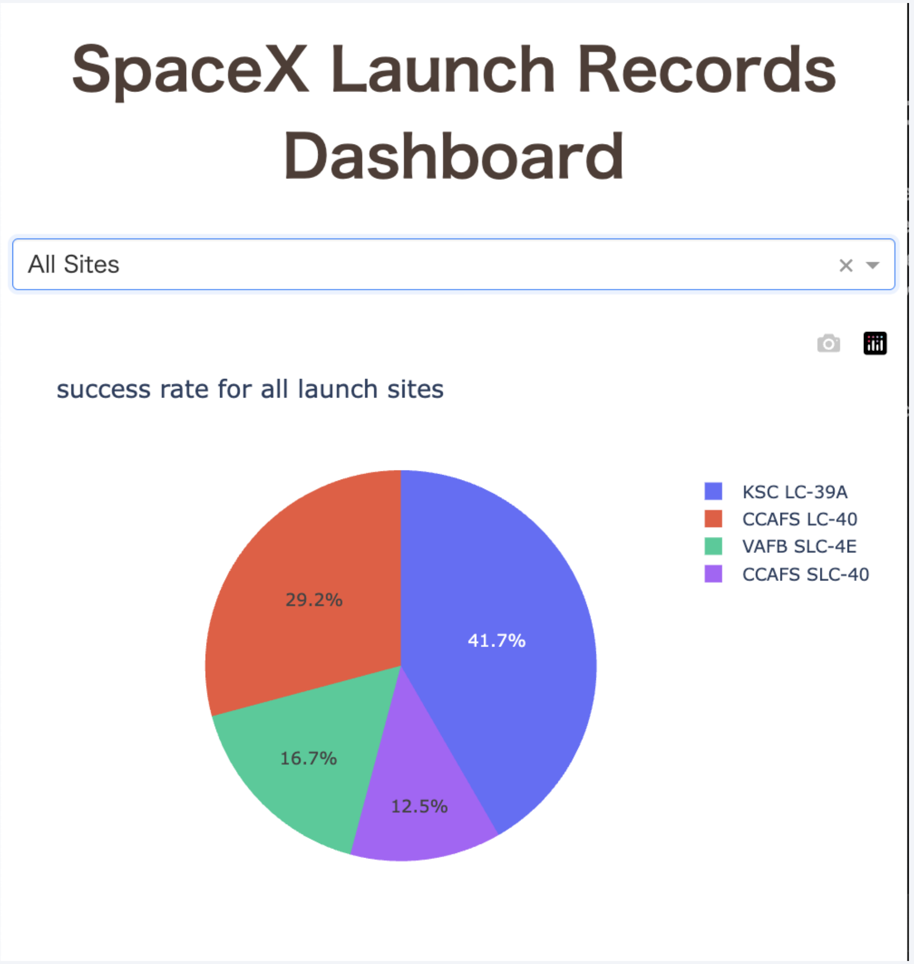



Section 4

# Build a Dashboard with Plotly Dash

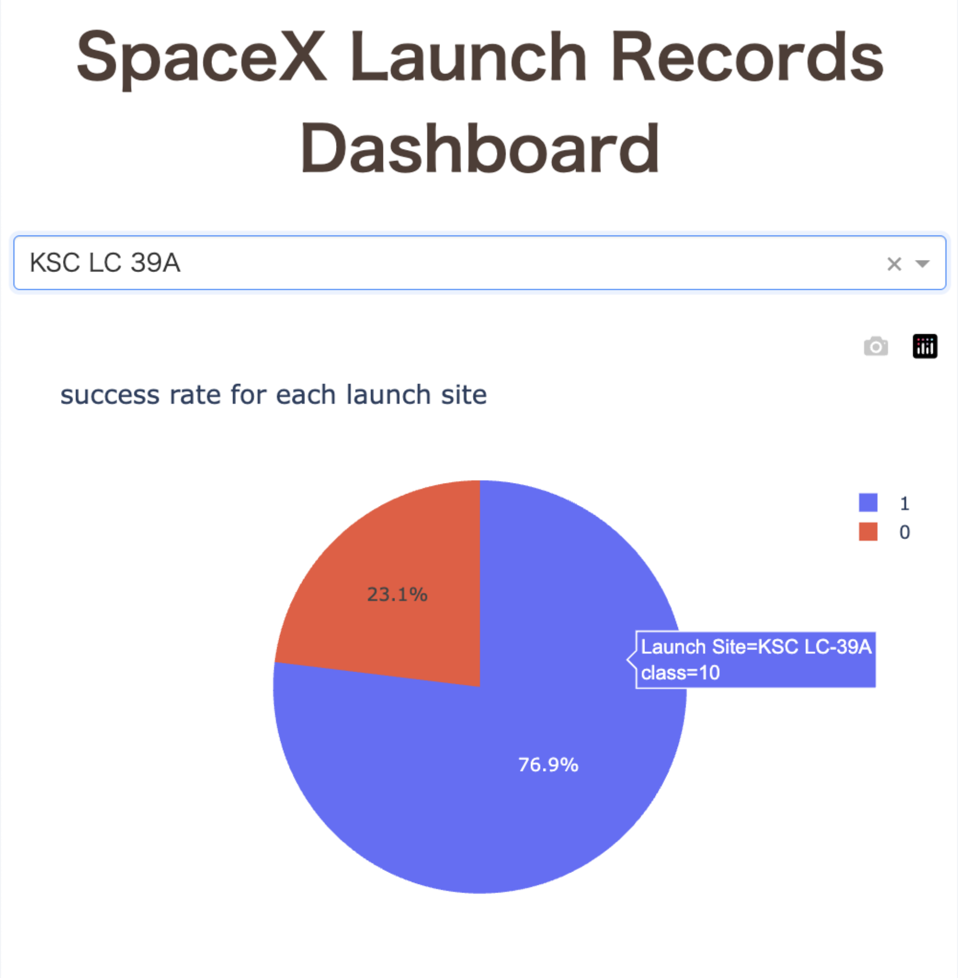
# Success rate for all launch site

- Visualize launch success rate for each launch site using the Plotly Dash application



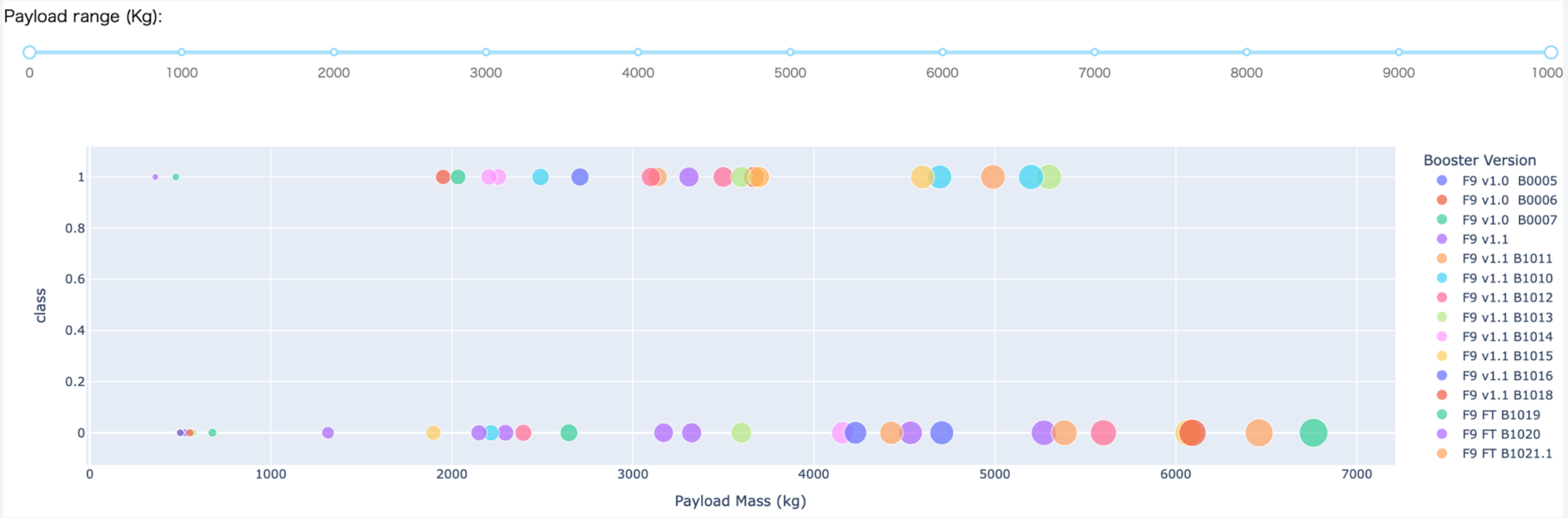
# The most successful launch site

- The launch site with the highest success rate was Launch Complex 39 (LC-39) at the John F. Kennedy Space Center in Merritt Island, Florida



# Payload and launch results

- The success rate of launch decreases when the payload exceeds 5000 kg
- If the payload exceeds 6000kg, the launch will fail
- The FT version has the highest success rate among boosters with more than 10 launches

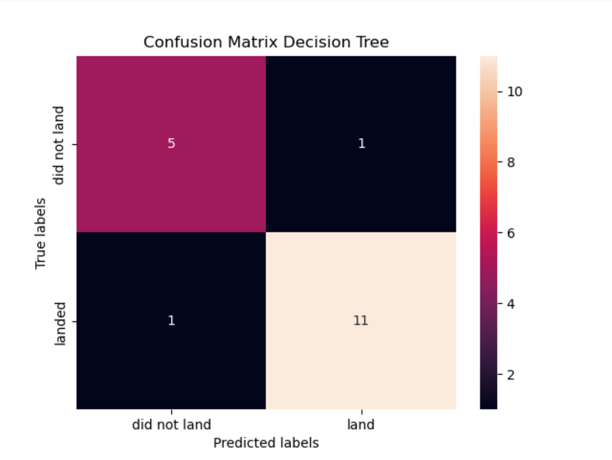


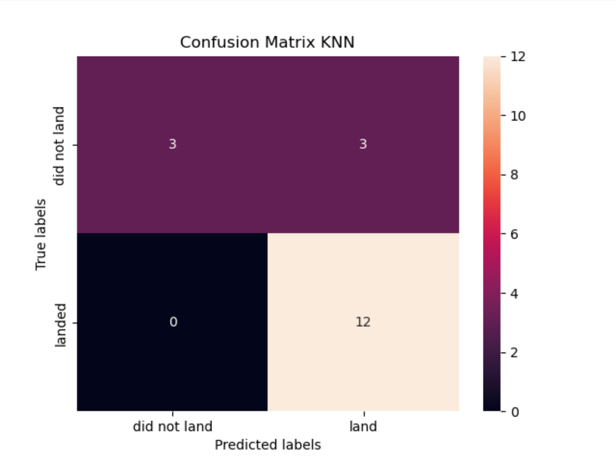
Section 5

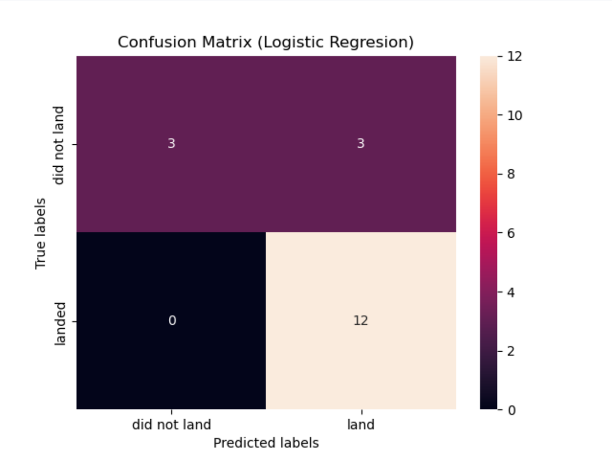
# Predictive Analysis (Classification)

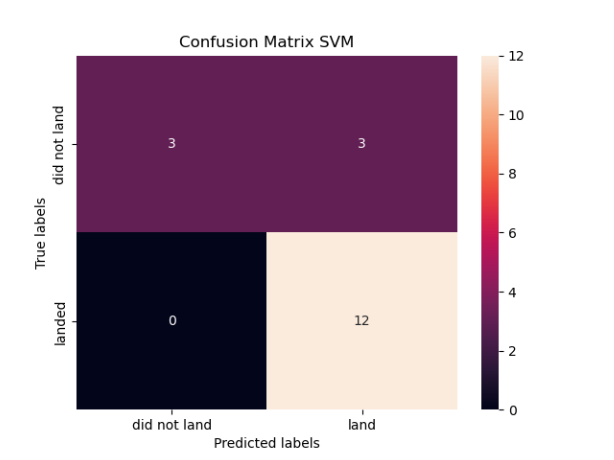
# Classification Accuracy

- The most accurate model in the training dataset was Decision Tree
- On the other hand, there was no difference in accuracy when looking at the test scores.



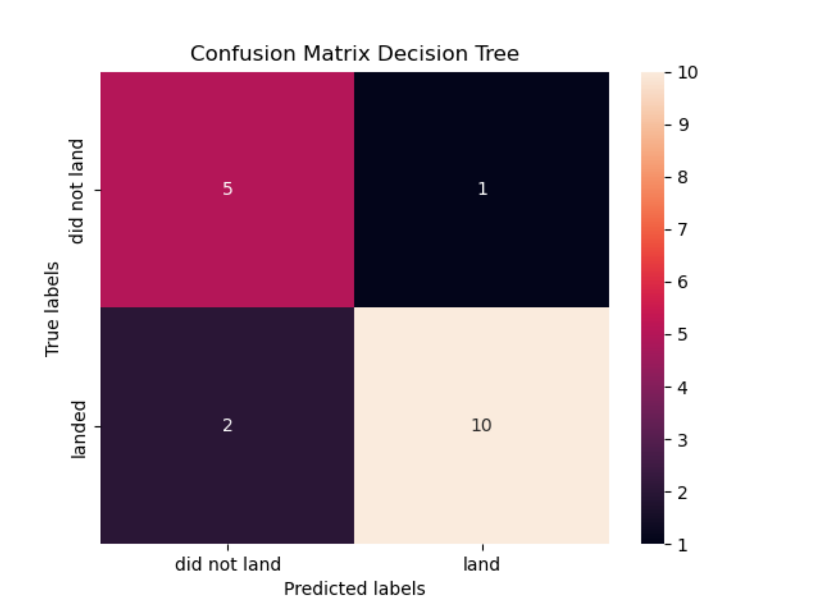



# Confusion Matrix

- Confusion matrix of Decision Tree model
- The classification accuracy details are shown in the table below




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 1.00 | 0.50 | 0.67 | 6 |
| 1 | 0.80 | 1.00 | 0.89 | 12 |
| accuracy | | | 0.83 | 18 |
| macro avg | 0.90 | 0.75 | 0.78 | 18 |
| weighted avg | 0.87 | 0.83 | 0.81 | 18 |

| | method | Accuracy | test score |
|---|---|---|---|
| 0 | Logistic Regulation | 0.846429 | 0.833333 |
| 1 | Decision Tree | 0.900000 | 0.888889 |
| 2 | SVM | 0.848214 | 0.833333 |
| 3 | KNN | 0.848214 | 0.833333 |

# Conclusions

- The most successful orbit types are SSO and HEO and GEO and ES-L1
- The launch site with the highest success rate was KSC LC-39
- Comparing the payload and launch results axes, the success rate of launch decreases when the payload exceeds 5000 kg
- The overall success rate of launches is improving year by year.

Thank you!