IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Yogita Gohiya
23-10-2025

Outline

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

Executive Summary

Summary of Methodologies

- Collected and cleaned SpaceX launch data
- Performed EDA to find key factors affecting landing success
- Built and tested ML models (Logistic Regression, Decision Tree, SVM, Random Forest)
- Evaluated models using accuracy

Summary of Results

- EDA Results
- Interactive Analytics
- Predictive Analytics

Introduction

Background:

SpaceX revolutionized space launches by reusing rocket stages, drastically reducing costs — about $62M per launch compared to $165M for competitors. Its success in recovering the Falcon 9 first stage since 2010 has drawn global attention.

Business Problem:

Launch cost depends on whether the first stage lands successfully. Predicting this outcome helps estimate costs and understand SpaceX’s pricing edge. Competing companies can use this insight to optimize bids and pricing strategies.

Section 1

Methodology

Methodology

Executive Summary

Data Collection & Wrangling

- Collected launch data from **SpaceX API** and **Wikipedia**
- Cleaned and transformed data (handled missing values, encoded categories)

Exploratory Data Analysis (EDA)

- Used **visualizations** and **SQL queries** to find trends and key factors
- Explored payload, orbit type, launch site, and success rate relationship

Interactive Visual Analytics

- Built **Folium maps** to show launch locations and landing outcomes
- Created **Plotly Dash dashboard** for interactive exploration

Predictive Analysis

- Applied **classification models** (Logistic Regression, SVM, KNN, Decision Tree)
- **Tuned models** using Grid Search and **evaluated** with accuracy
- Identified **best model** for predicting landing success

Data Collection

Data Sources:

- SpaceX API – official launch data (dates, payload, rocket type, launch outcome)
- Wikipedia – supplementary data (orbit type, landing location, booster version)

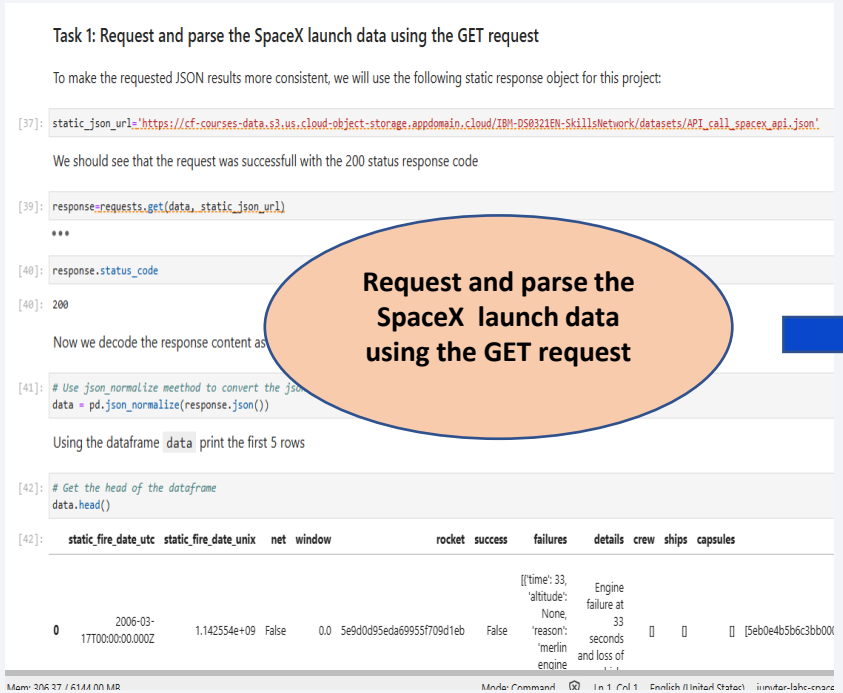
Tools Used:

- Python libraries: requests, pandas, BeautifulSoup
- API calls for JSON data retrieval
- Web scraping for additional launch info

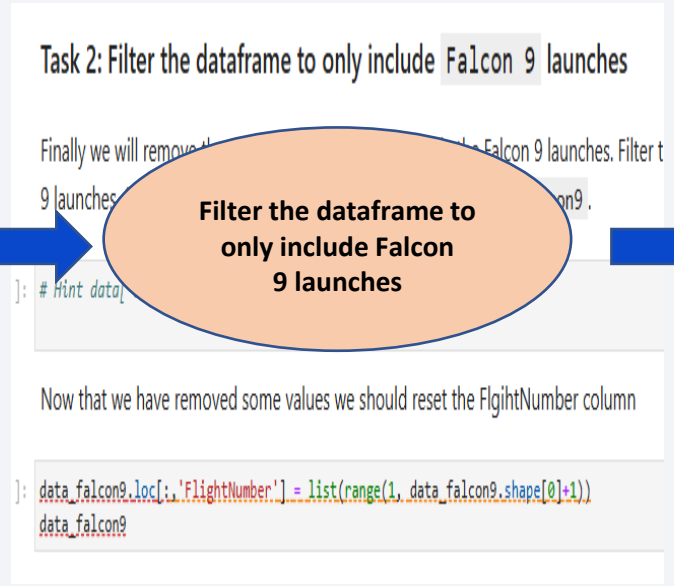
Process Summary:

- **Access SpaceX REST API** → Retrieve raw launch data in JSON format
- **Web Scraping (Wikipedia)** → Collect missing attributes (orbit, landing type)
- **Combine Datasets** → Merge API and scraped data
- **Data Cleaning** → Handle missing values & standardize formats
- **Store Dataset** → Save as CSV for further analysis

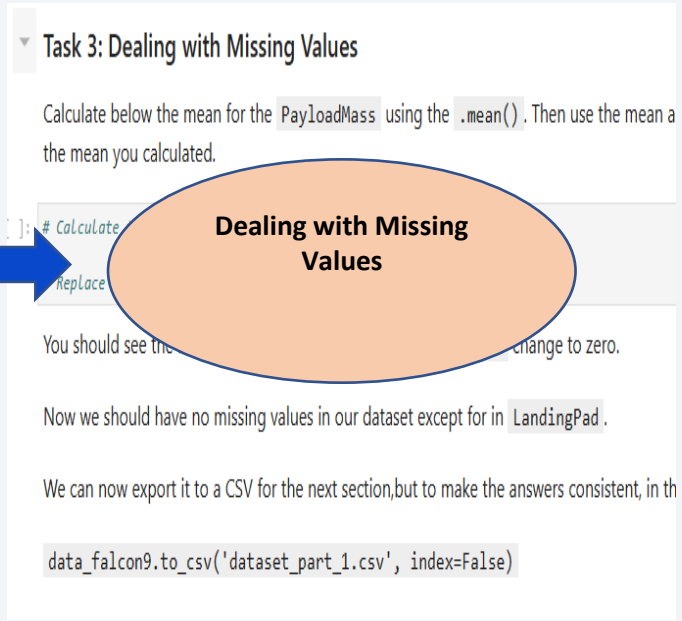
Data Collection – SpaceX API



Request and parse the SpaceX launch data using the GET request



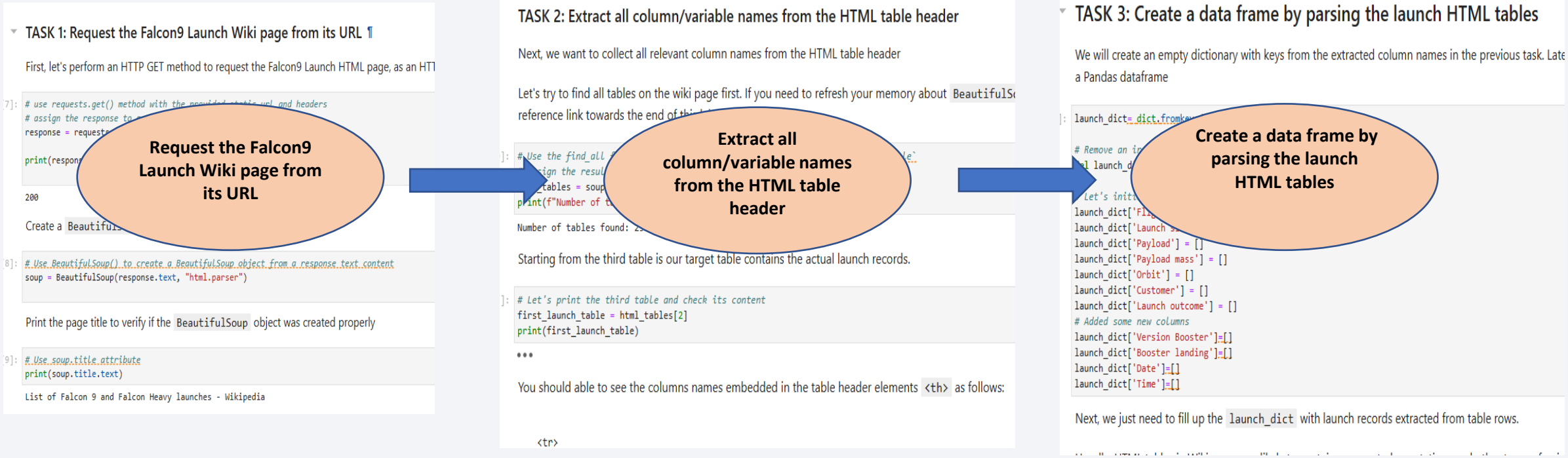
Filter the dataframe to only include Falcon 9 launches



Dealing with Missing Values

GitHub URL : https://github.com/YogitaGohiya/Space-X-Falcon-9-First-Stage-Landing-Prediction/blob/main/spacex-data-collection-api.ipynb

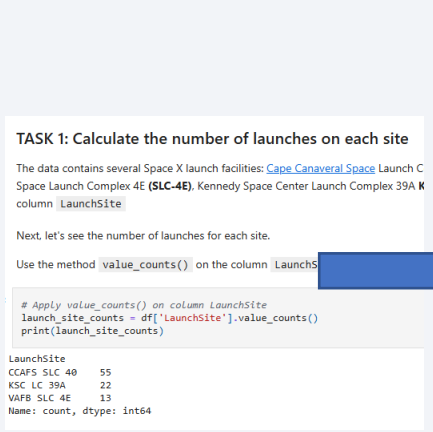
Data Collection - Scraping

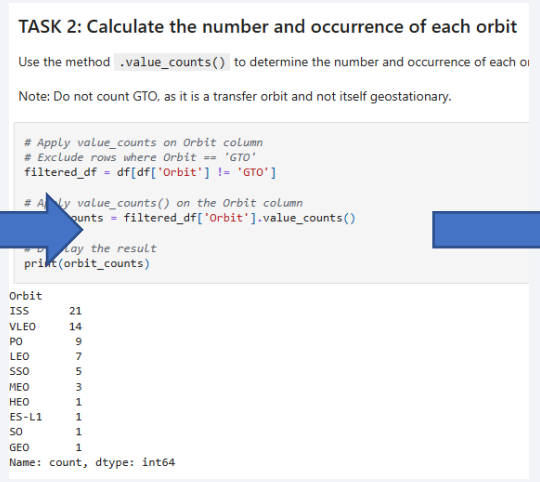


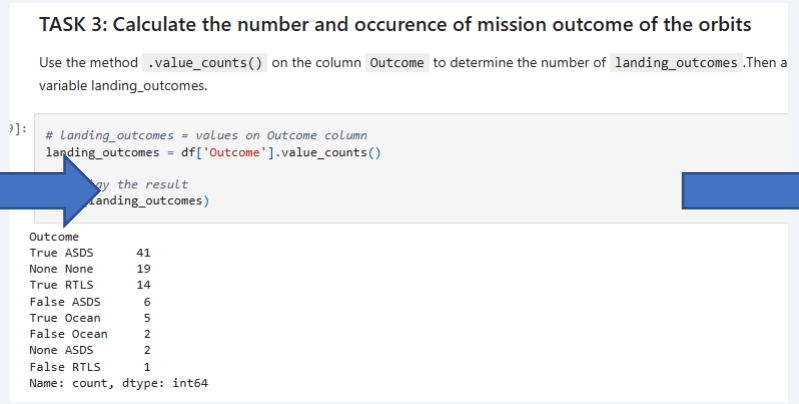
GitHub URL : https://github.com/YogitaGohiya/Space-X-Falcon-9-First-Stage-Landing-Prediction/blob/main/webscraping.ipynb

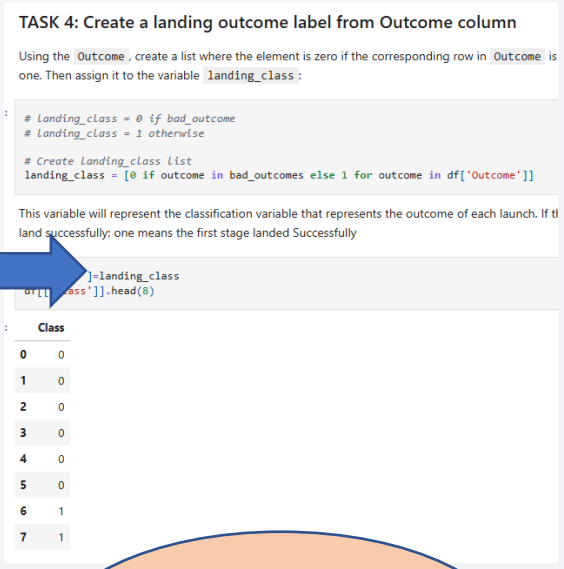
Data Wrangling

Process of **cleaning and transforming raw data** into a usable format





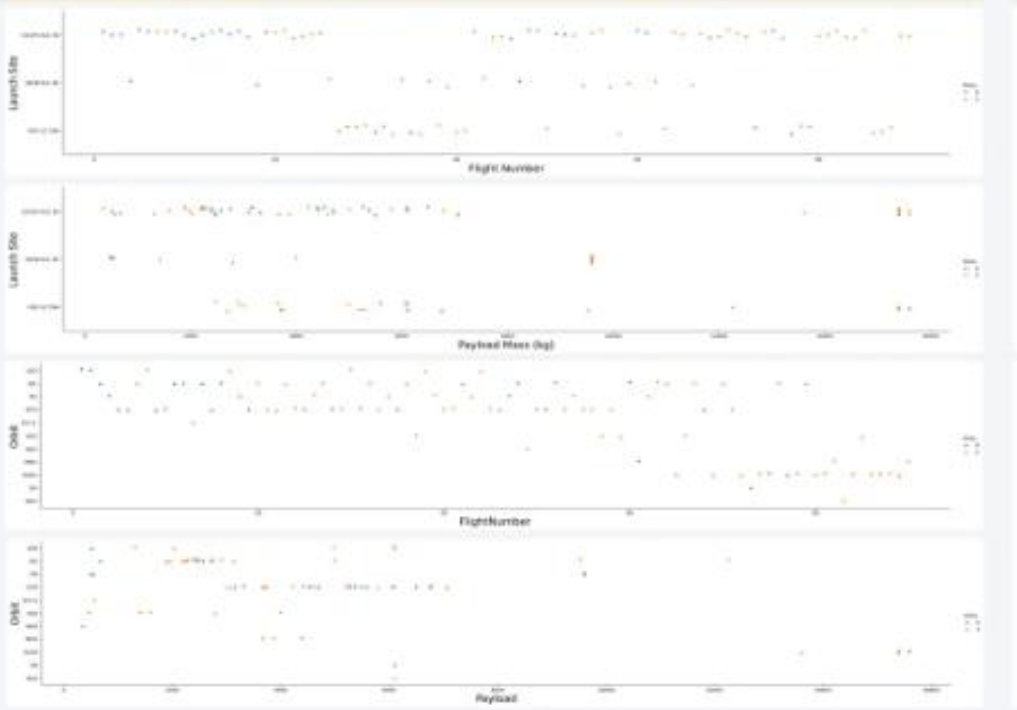




Calculate the number of launches on each site → **Calculate the number and occurrence of each orbit** → **Calculate the number and occurence of mission outcome of the orbits** → **Create a landing outcome label from Outcome column**

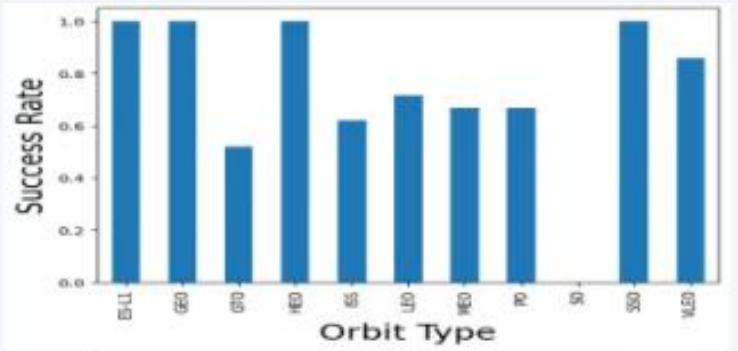
GitHub URL : https://github.com/YogitaGohiya/Space-X-Falcon-9-First-Stage-Landing-Prediction/blob/main/spacex-Data%20wrangling.ipynb

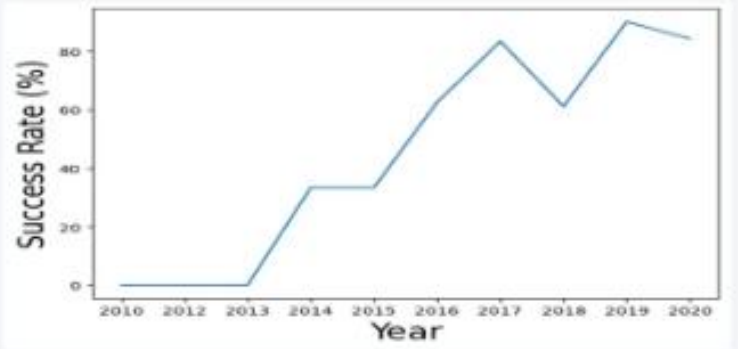
EDA with Data Visualization



Using scatter point chart visualize relationship between

1. Flight Number and Launch Site
2. Payload Mass and Launch Site
3. FlightNumber and Orbit type
4. Payload Mass and Orbit type





1. Bar chart for success rate of each orbit type
2. Line chart for launch success yearly trend

GitHub URL https://github.com/YogitaGohiya/Space-X-Falcon-9-First-Stage-Landing-Prediction/blob/main/EDA%20with%20Visualization%20Lab.ipynb

EDA with SQL

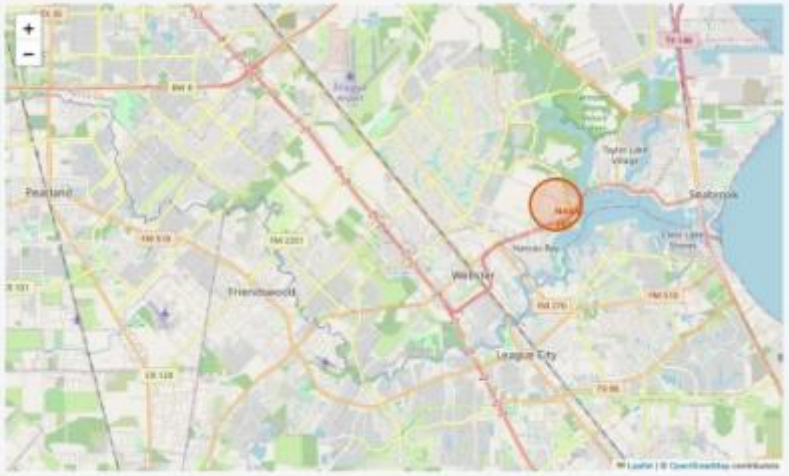
- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first succesful landing outcome in ground pad was acheived.
- 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 all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.
- 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.
- 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.

GitHub URL: https://github.com/YogitaGohiya/Space-X-Falcon-9-First-Stage-Landing-Prediction/blob/main/eda-sql.ipynb

Build an Interactive Map with Folium

Purpose

- To **visually identify** where launches occur across different sites
- To **analyze success patterns** by location and visualize reusability performance
- To make data **interactive and intuitive** for understanding geographical impact on landing success reference and peer-review purpose

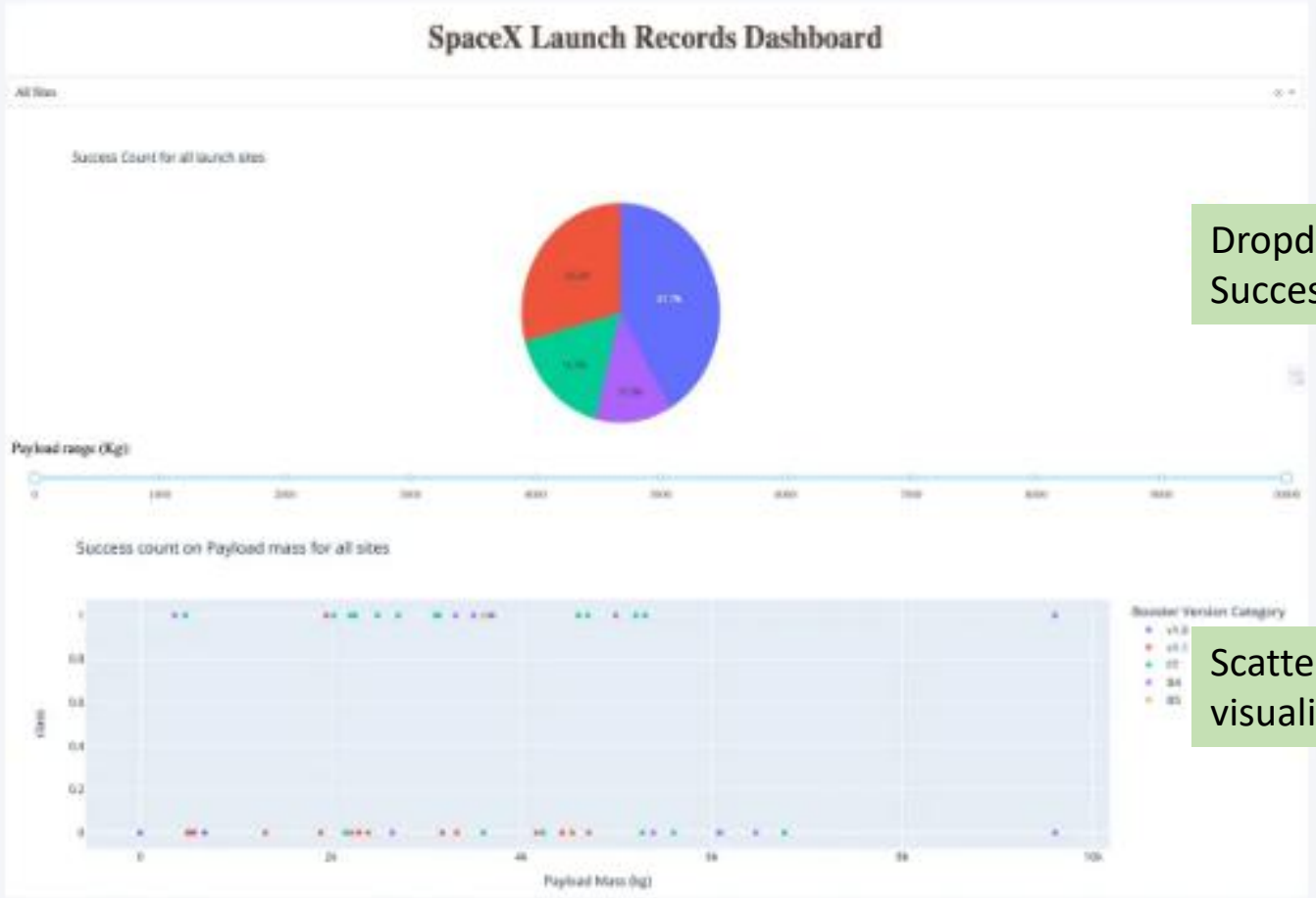


Circle marker at NASA Johnson Space Center's coordinate

Distance marker to show distances between a launch site to its proximities

GitHub URL: https://github.com/YogitaGohiya/Space-X-Falcon-9-First-Stage-Landing-Prediction/blob/main/eda-sql.ipynb

Build a Dashboard with Plotly Dash



Dropdown option of a pie chart created to visualize Successful launches of each site

Scatter point with a Range Slider to Select Payload to visualize Successful launches

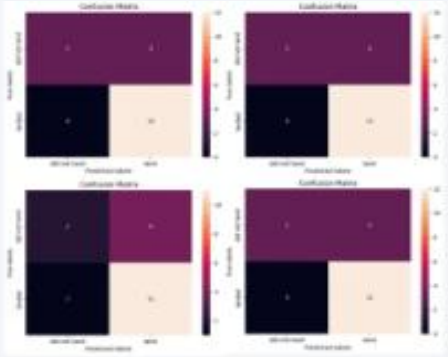
GitHub URL: https://github.com/YogitaGohiya/Space-X-Falcon-9-First-Stage-Landing-Prediction/blob/main/dashboard%20spacex

Predictive Analysis (Classification)

To predict if the Falcon 9 first stage will land successfully we used Logistic Regression, Decision Tree, K-Nearest Neighbors (KNN), Support Vector Machine (SVM)

Steps:

- Split Data – Train/test split for unbiased evaluation
- Build Models – Train multiple classification models
- Evaluate Models – Compare accuracy(highest 83.33%)
- Tune Hyperparameters – Use GridSearchCV for optimal settings
- Select Best Model – Choose the highest-performing classifier



GitHub URL: https://github.com/YogitaGohiya/Space-X-Falcon-9-First-Stage-Landing-Prediction/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5%20(1).ipynb

Results

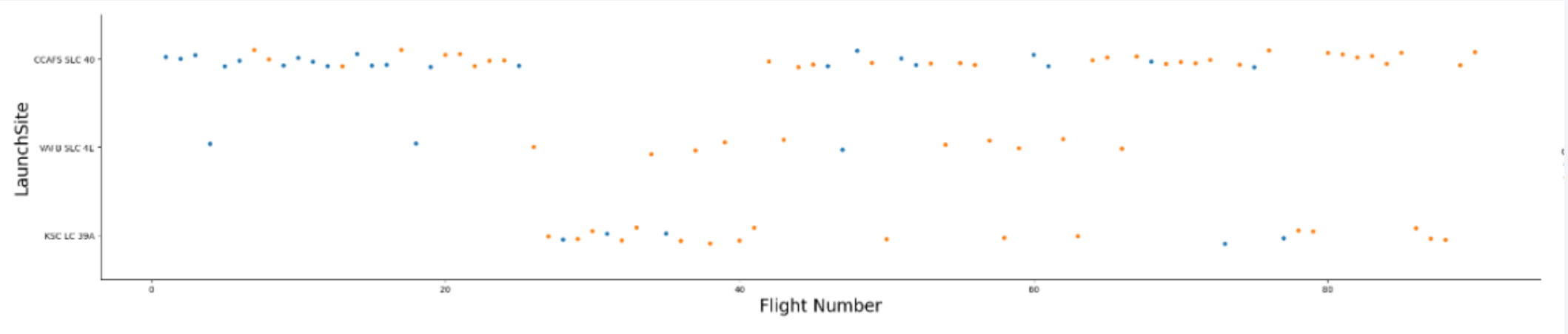
- **Top Models**: LR, SVM, and KNN gave the best prediction performance
- **Lighter Payloads**: Higher success rates than heavier ones
- **Best Site & Orbits**: KSC LC-39A and GEO/HEO/SSO/ES L1 orbits show the most successful launches

KSC LC-39A has the most successful launches overall

Section 2

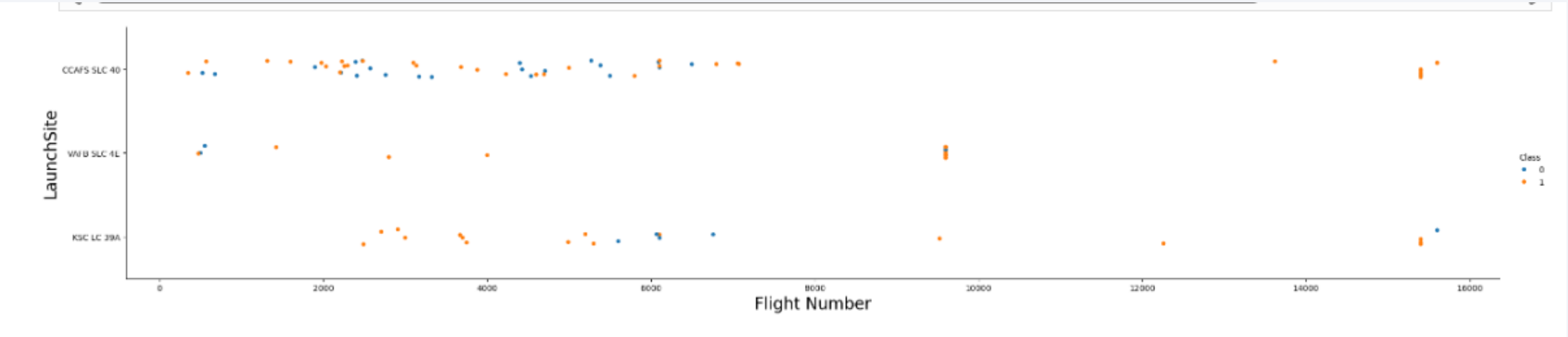
Insights drawn from EDA

Flight Number vs. Launch Site



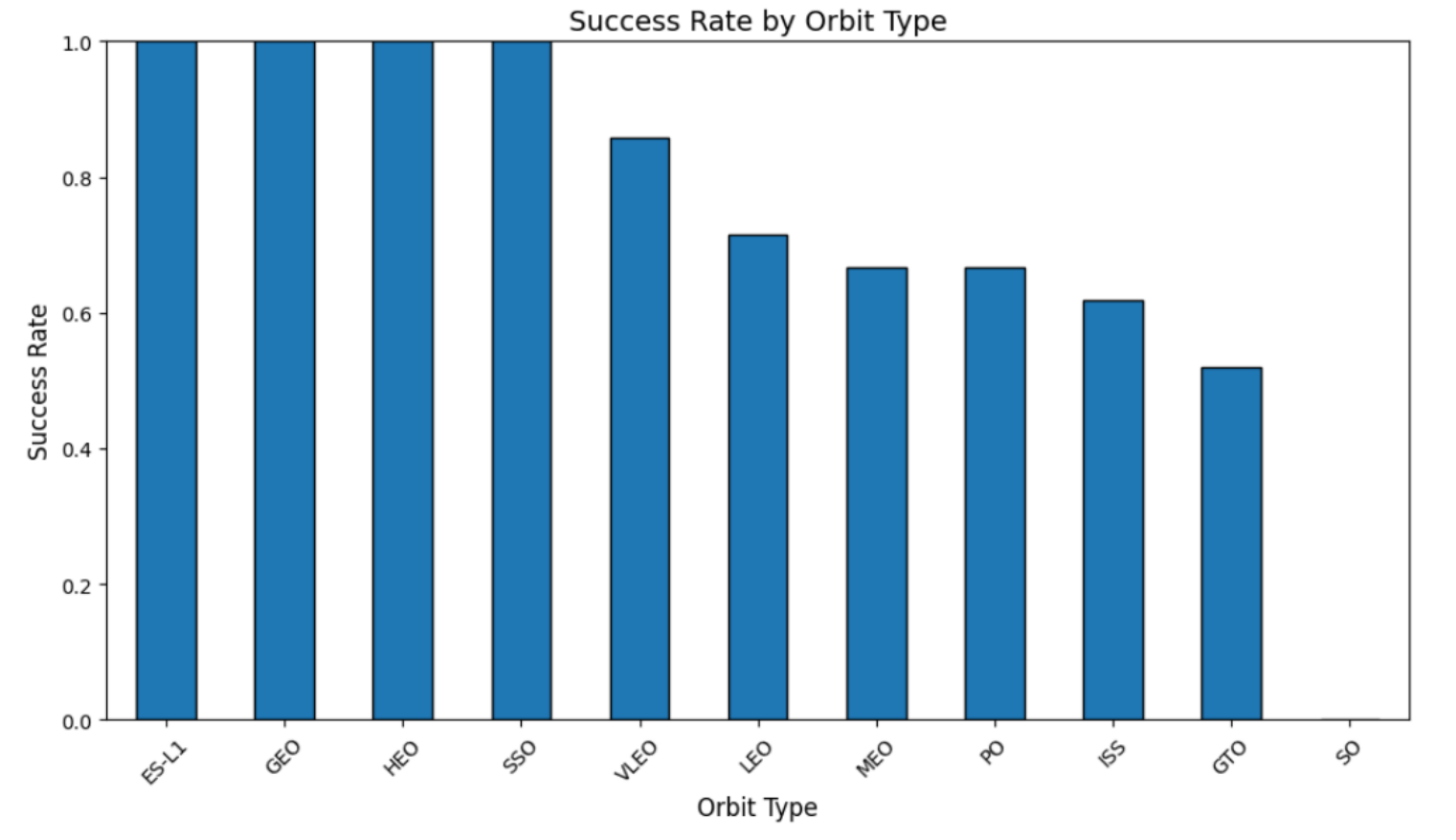
CCAFS SLC-40 recorded the **highest number of launches** among all launch sites.

Payload vs. Launch Site



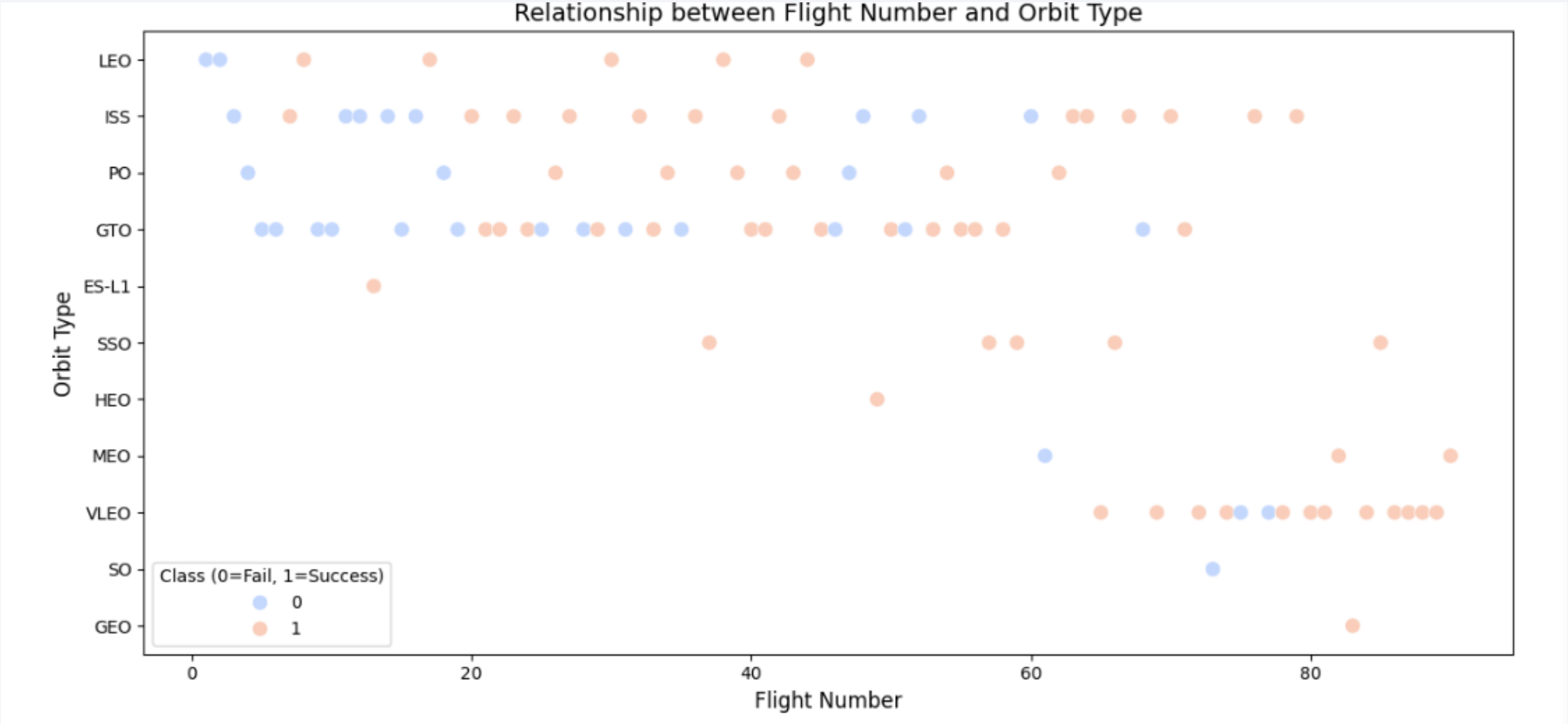
Payloads with lower mass are have more launches compared to those with higher mass across all three launch sites

Success Rate vs. Orbit Type



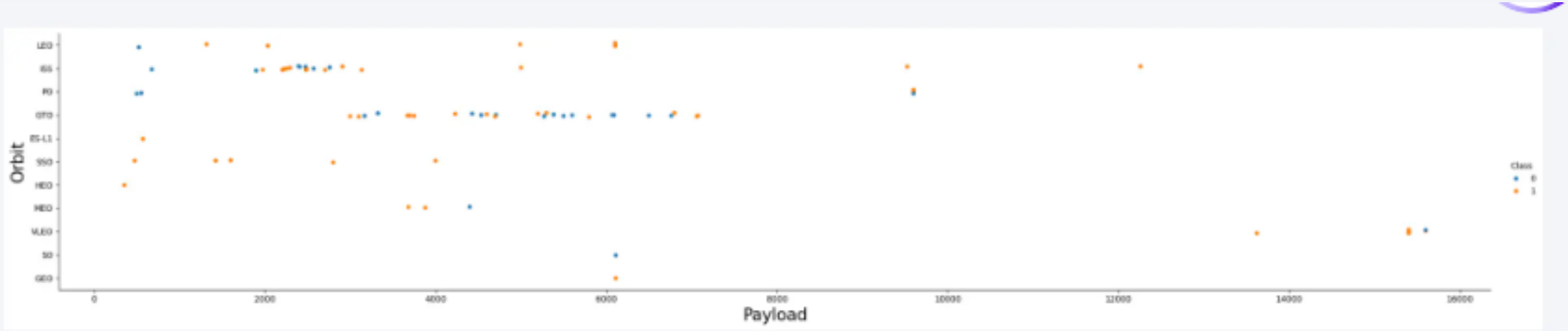
Orbit types ES-L 1,GEO,HEO,SSO have the highest success rate among all

Flight Number vs. Orbit Type



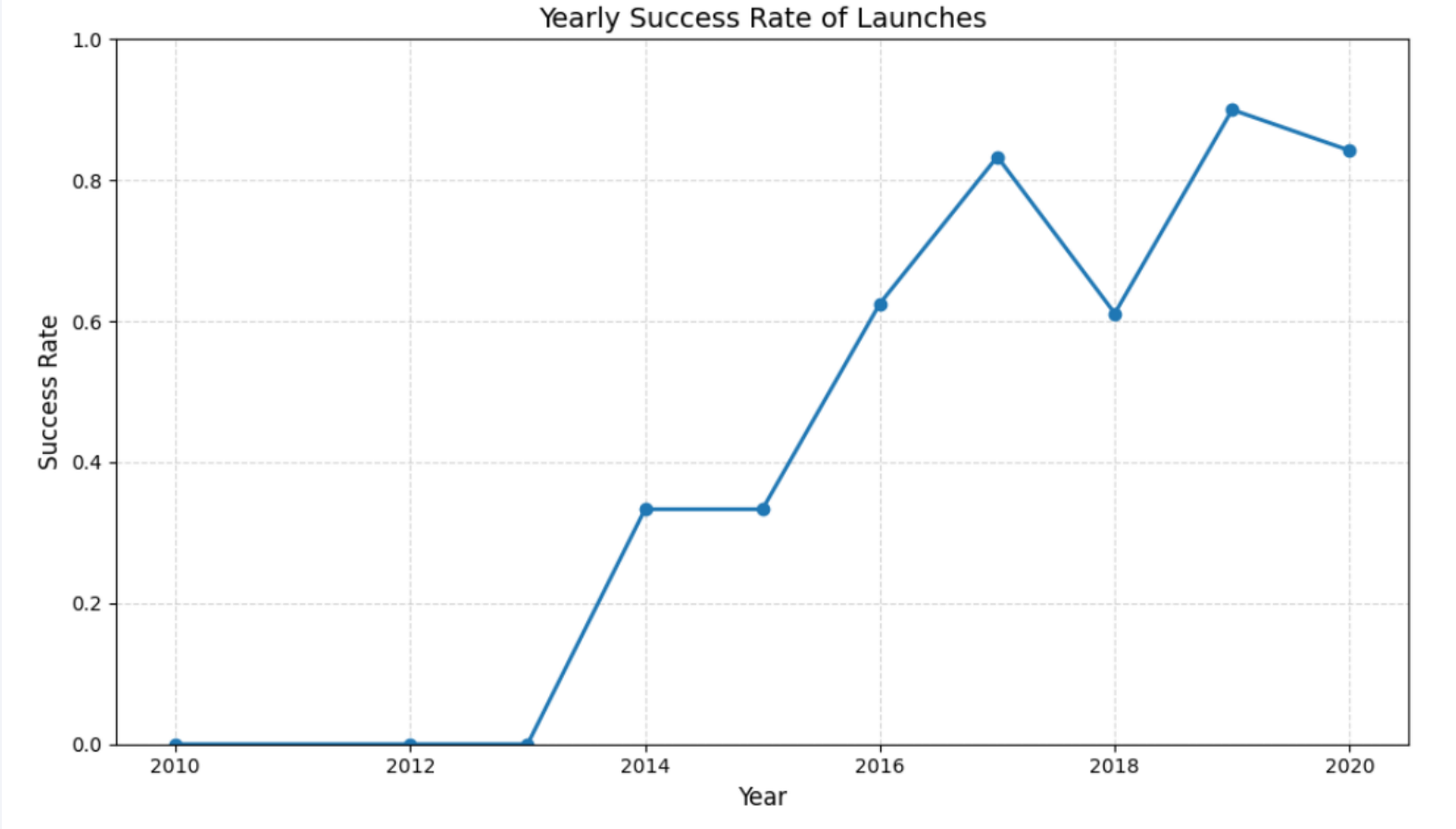
LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success

Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS. However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend



Observed that the success rate since 2013 kept increasing till 2020

All Launch Site Names

Task 1

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

```
n [34]: %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
 * sqlite:///my_data1.db
Done.
```

ut[34]:

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

Performed an SQL query to obtain all launch site names

Launch Site Names Begin with 'CCA'

Task 2

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

In [44]:
```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 16
```

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

Out[44]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__K
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Brouere cheese	
2012-	7:44:00	F9 v1.0 B0005	CCAFS LC-	Dragon demo flight	5

Performed an SQL query to obtain 5 launch site names that begin with 'CCA'

Total Payload Mass

Task 3

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

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") AS Total_Payload_Mass FROM SPACEXTABLE WHERE "Customer" = 'NASA (CRS)';
```

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

Total_Payload_Mass
45596

Performed an SQL query to display the total payload mass carried by boosters launched by NASA (CRS)

Average Payload Mass by F9 v1.1

Task 4

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

```
']:  %sql SELECT AVG(PAYLOAD_MASS__KG_) AS average_payload FROM SPACEXTABLE WHERE Booster_Version = 'F9 v1.1';
```

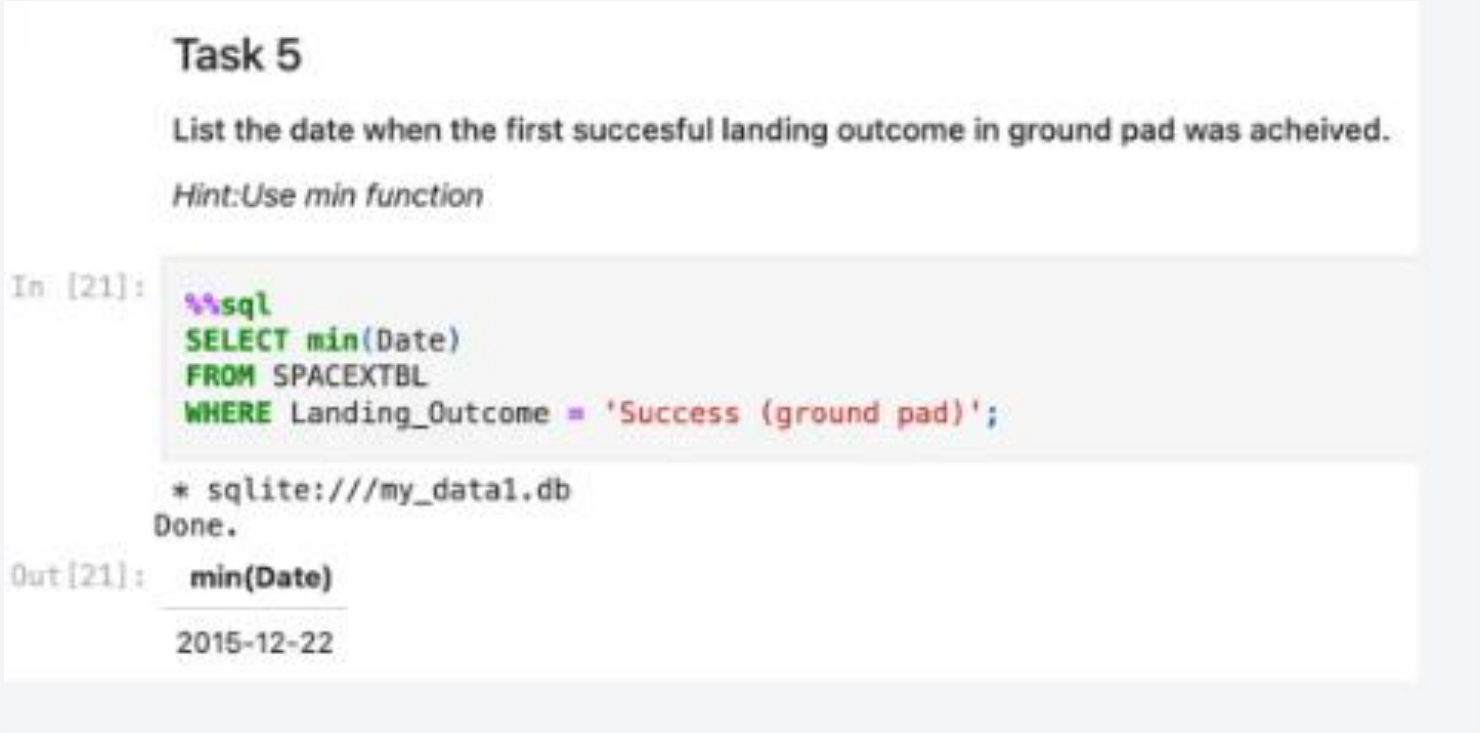
```
 * sqlite:///my_data1.db
Done.
```

']:

average_payload
2928.4

Performed an SQL query to display average payload mass carried by booster version F9 v1.1

First Successful Ground Landing Date



Task 5

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

Hint:Use min function

In [21]:

```
%%sql
SELECT min(Date)
FROM SPACEXTBL
WHERE Landing_Outcome = 'Success (ground pad)';
```

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

Out[21]:

min(Date)
2015-12-22

Performed an SQL query to list the date when the first succesful landing outcome in ground pad was acheived

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
]: %sql SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (drone ship)'AND PAYLOAD_MASS__KG_ > 4000 AND
```

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

]:

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

Performed an SQL query to list the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

[53]:
```
%sql SELECT Mission_Outcome, COUNT(*) AS total FROM SPACEXTABLE GROUP BY Mission_Outcome;
```

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

[53]:

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

Performed an SQL query to List the total number of successful and failure mission outcomes

Boosters Carried Maximum Payload

Task 8

List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.

```
55]: %sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE
```

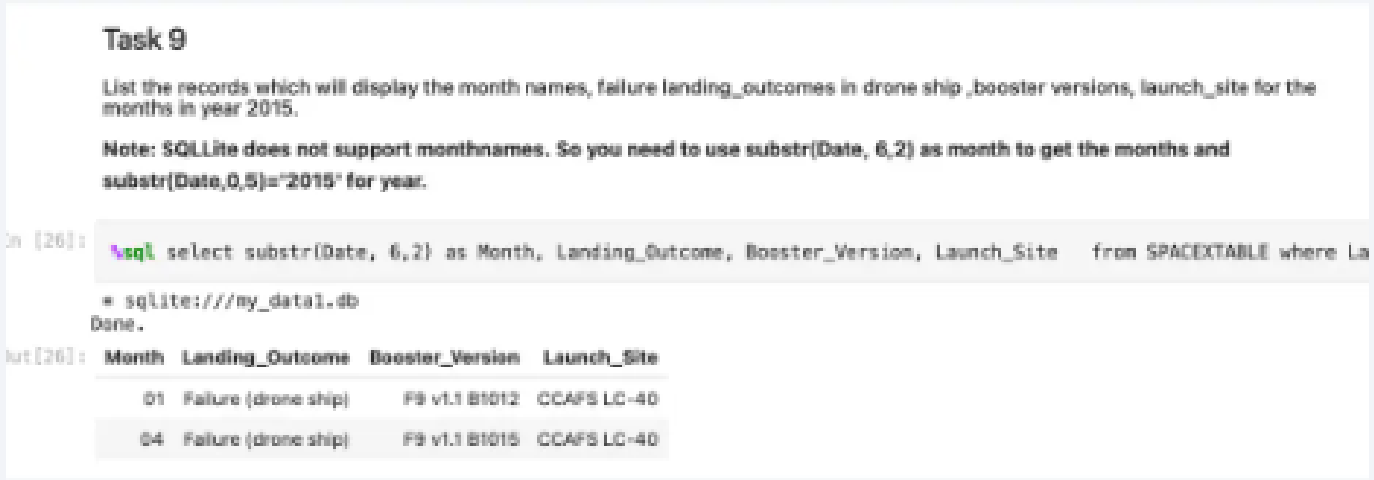
```
 * sqlite:///my_data1.db
Done.
```

55]:

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

Performed an SQL query to List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.

2015 Launch Records



Task 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.

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 [26]:
```
%sql select substr(Date, 6,2) as Month, Landing_Outcome, Booster_Version, Launch_Site   from SPACEXTABLE where La
```

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

Out[26]:

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

Performed an SQL query to 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.

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

Task 10

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 [27]:
```
%sql select Landing_Outcome, count(*) as 'Count' from SPACEXTABLE where Date between '2010-06-04' and '2017-03-20
```

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

Out[27]:

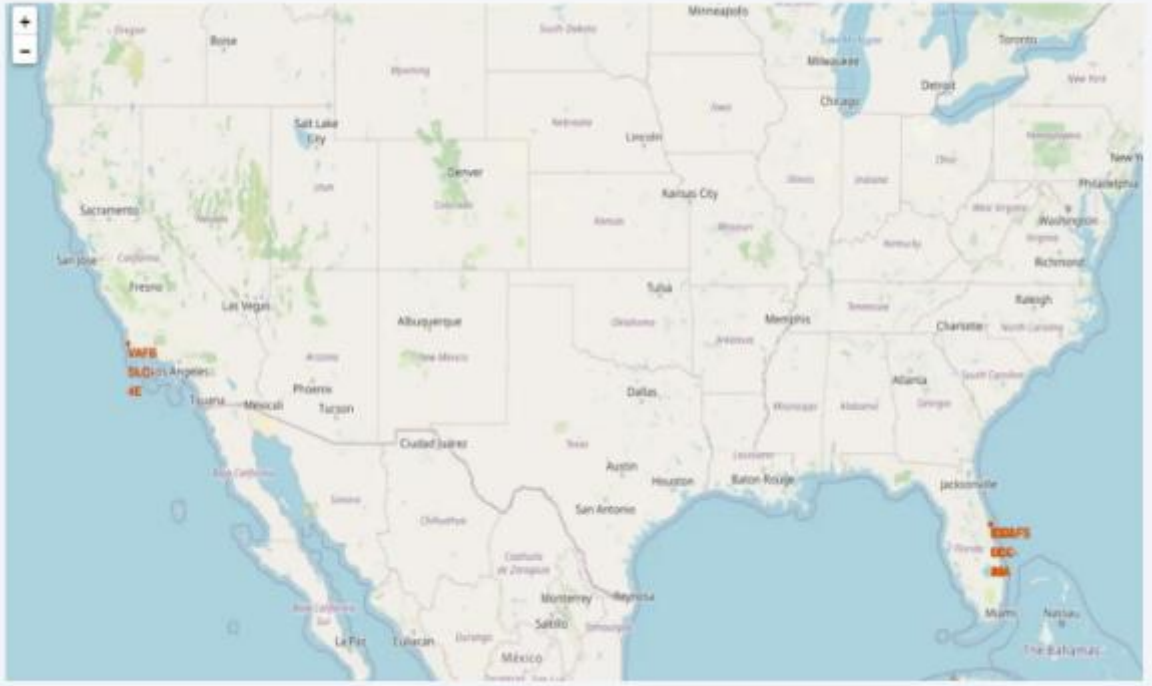
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Performed an SQL query to 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


Section 3
Launch Sites
Proximities Analysis

All launch sites on a map

The launch sites are labelled by a marker with their names on the map

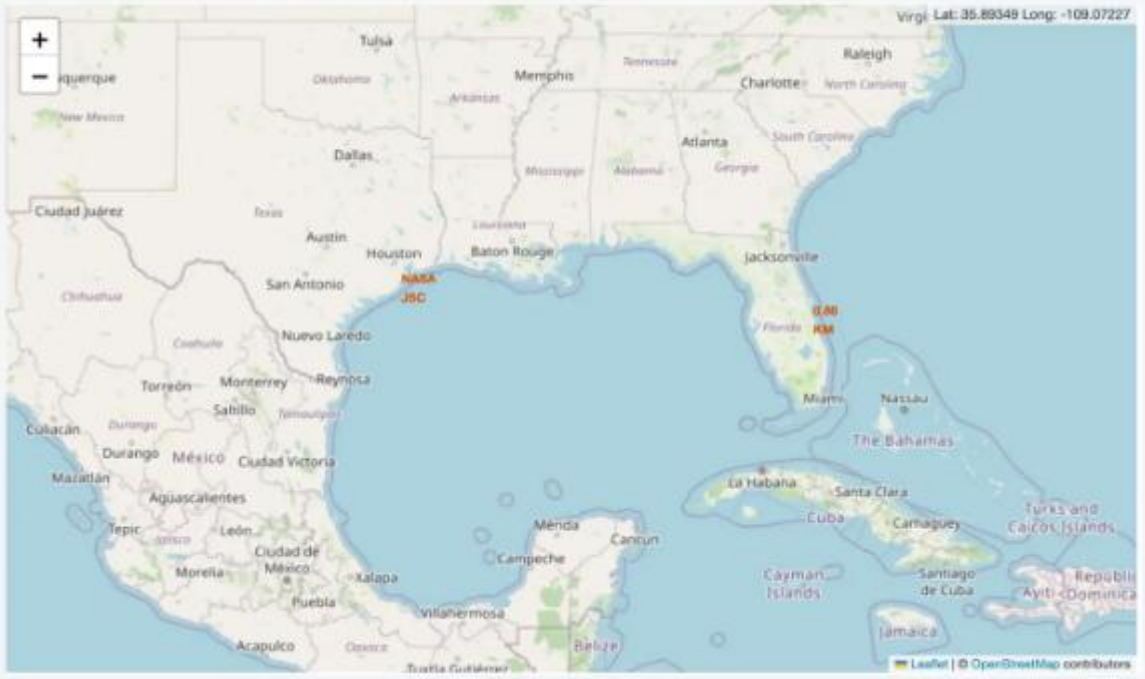


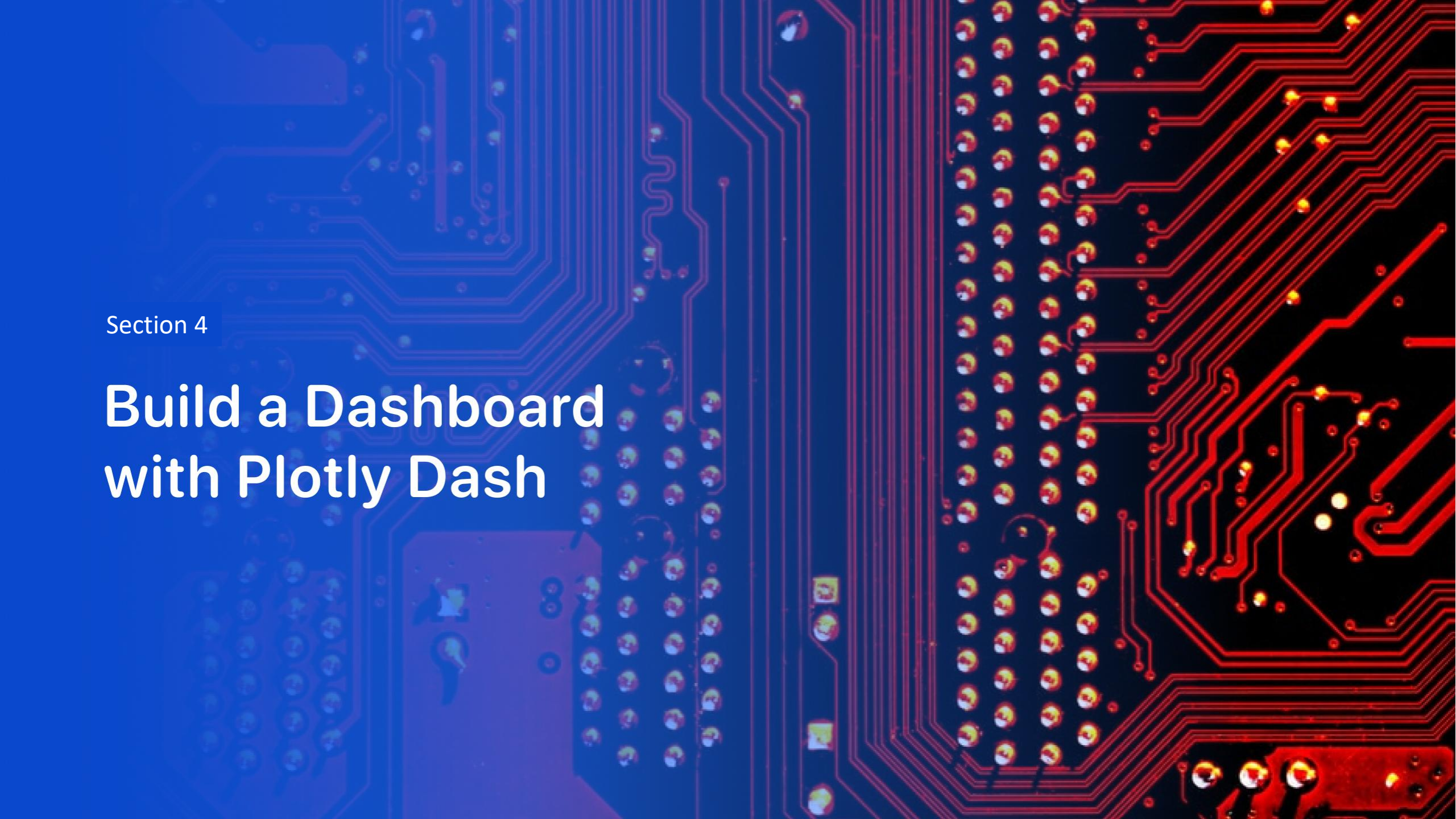
All success and failed launches for each sites on a map

The launch records are grouped in clusters on the map, then labelled by green markers for successful launches and re markers for failed ones

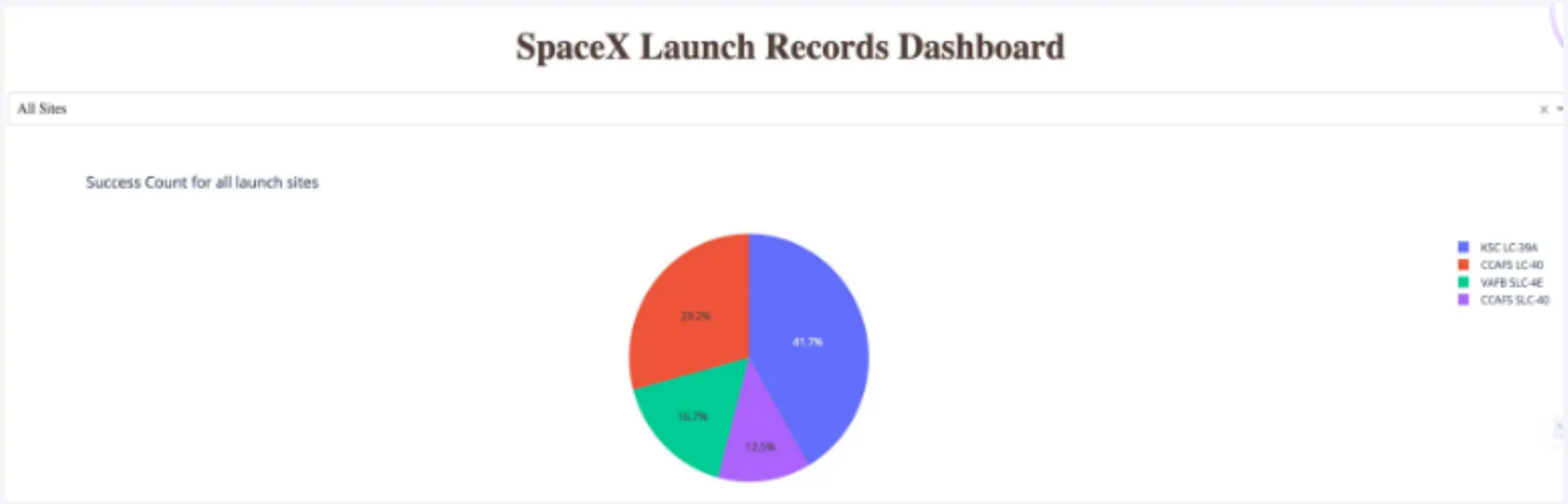
Distance between a launch site to is proximities

The closest coastline from NASA JSC is marked as a point using Mouseposition and the distance between the coastline point and the launch site is aproxx. 0.86 KM



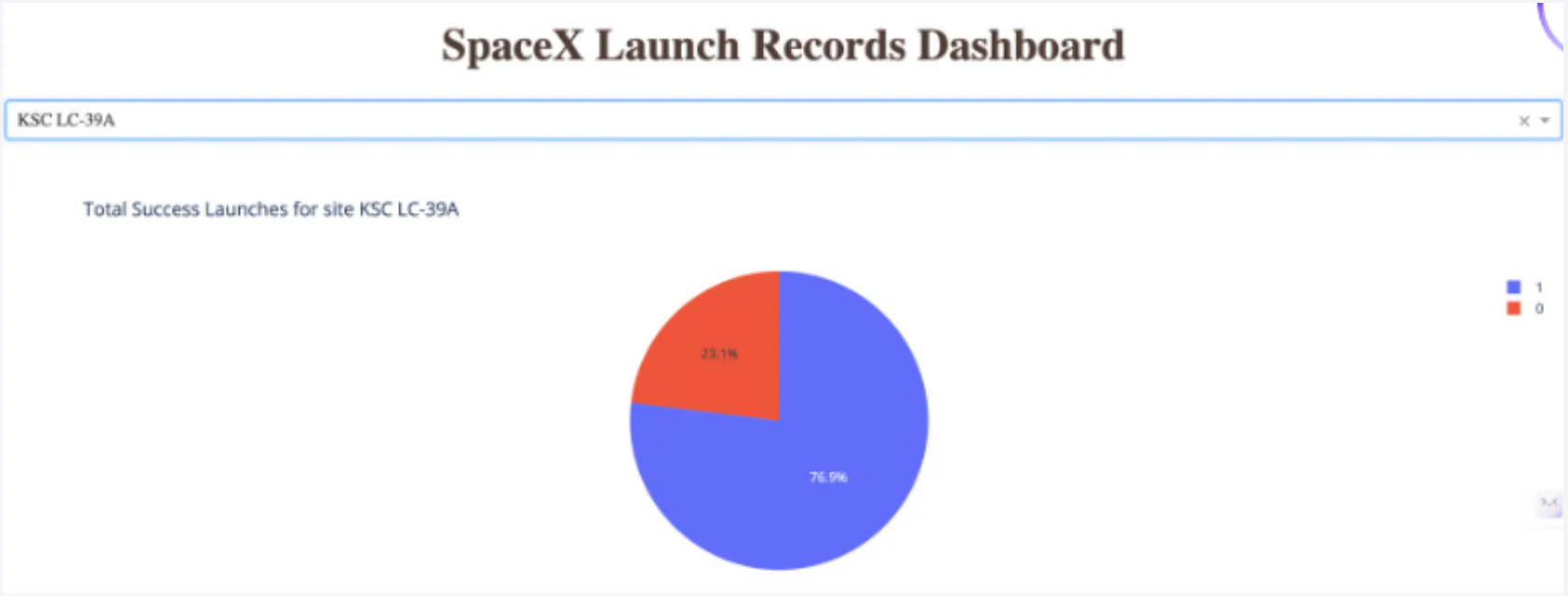

Section 4
Build a Dashboard with Plotly Dash

Total successful launches for all sites



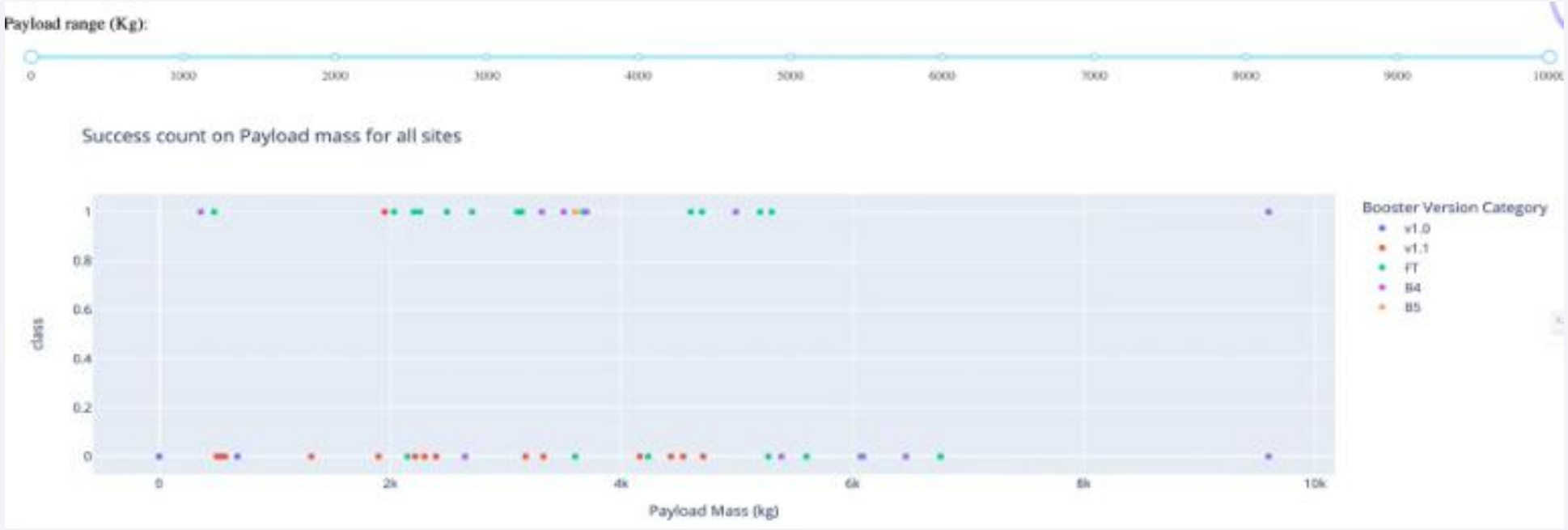
KSC LC-39A has the highest amount of successful launches with 41.7% from the entire record, whereas CCAFS SLC-40 has the lowest amount of successful launches with only 12.50%

Success ratio of the launch site with the highest successful launches



KSC LC-39A which is the launch site with highest amount of success, has a 76.90% success rate for the launches from its site and 23.10% failure rate

Payload VS Launch outcome

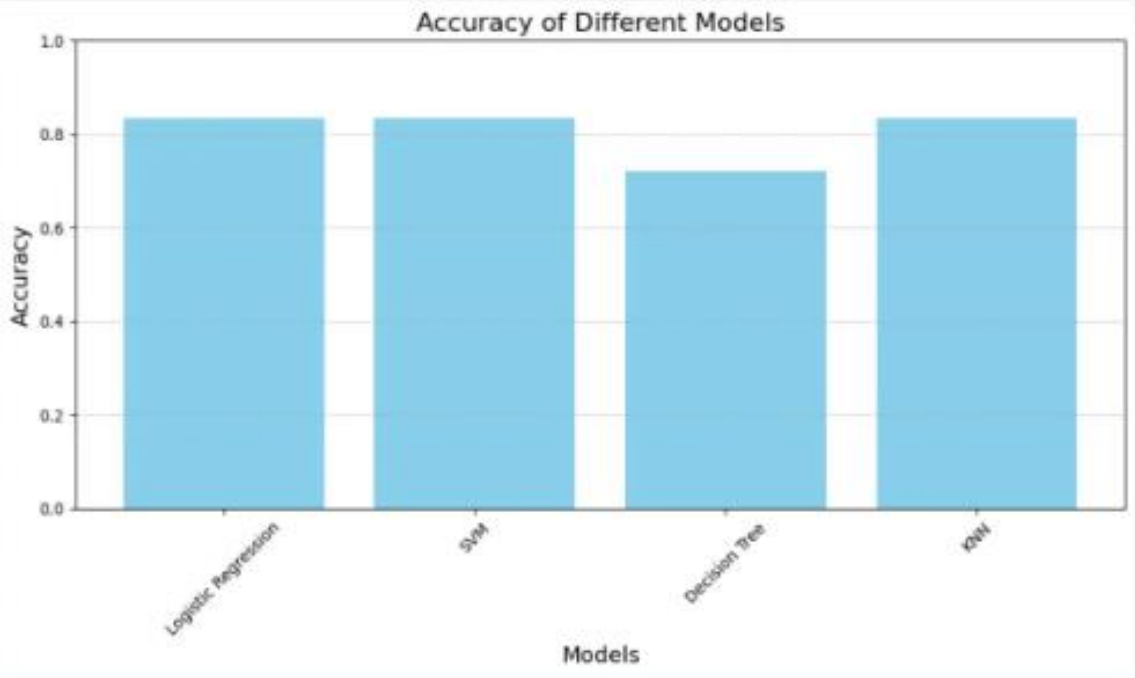


The payload range that has the highest success launches is between 2000 to 4000 kg , which can be seen by the most number of plots in that range, followed by the payload range of 4000 to 6000 kg with the second most number of plots

Section 5

Predictive Analysis (Classification)

Classification Accuracy



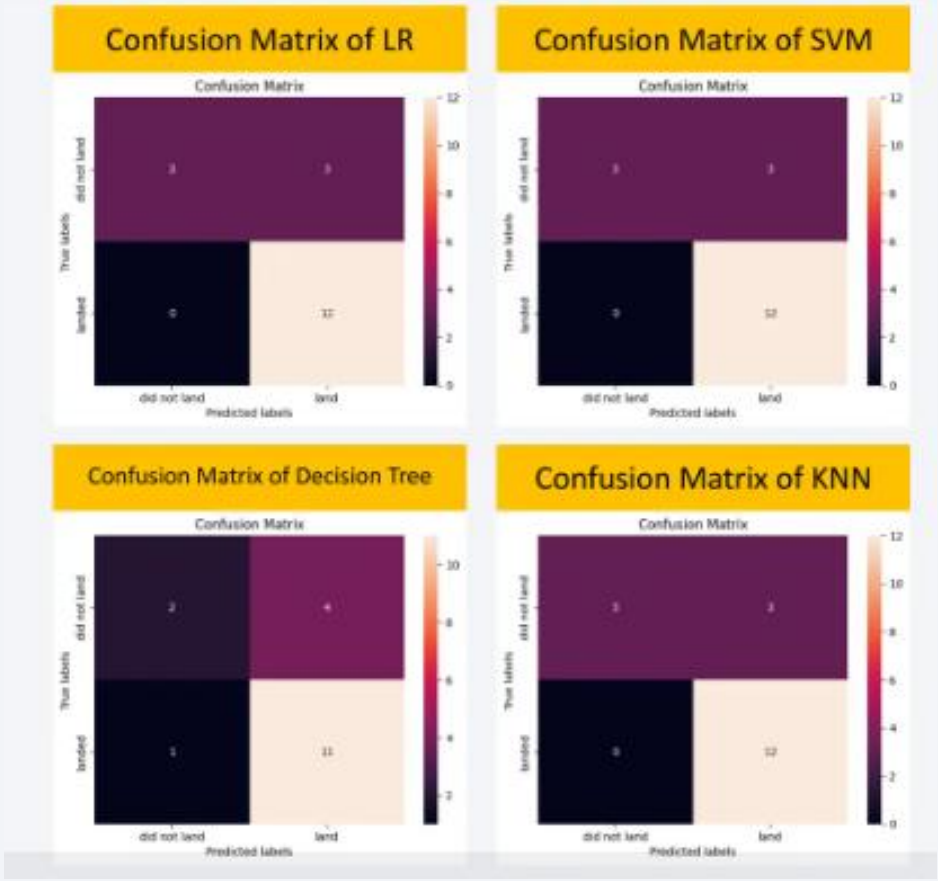
Logistic Regression (LR), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) are the top-performing models for predicting landing outcomes with thw accuracy of 83.33%

Confusion Matrix

LR,SVM,KNN models are good as their confusion matrix shows that they predict all 12 successful landing correctly, with zero error.

However, the decision tree model only predicted 11 successful landing with on wrong prediction

Logistic Regression (LR), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) are the top-performing models for predicting landing outcomes with same accuracy 83.33%



Conclusions

- **Logistic Regression (LR), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN)** are the top-performing models for predicting landing outcomes.
- Lighter payloads show **a higher success** rate compared to heavier ones.
- GEO, HEO, SSO, and ES L1 orbit types demonstrate the **highest rates** of successful launches.
- Launch Complex 39A (**KSC LC-39A**) at the Kennedy Space Center records the most successful launches overall compared to other sites.

Appendix

Link:
https://github.com/YogitaGohiya/Space-X-Falcon-9-First-Stage-Landing-Prediction/tree/main

Thank you!