IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Krishna Chaitanya BVS
30-11-2025

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- **The objective of this report is to predict whether the first stage of a SpaceX Falcon 9 rocket launch will successfully land.**

- SpaceX launches cost approximately **$62 million**, while other providers may charge **$165 million or more** for similar missions. A major reason for SpaceX's lower cost is its ability to **reuse the Falcon 9 first-stage booster**.

- By accurately predicting the likelihood of a successful landing, we can **estimate the overall launch cost**, which can provide valuable insights for companies competing with SpaceX.

Section 1

Methodology

Methodology

1. **Data Collection Methodology -** Collected SpaceX launch data through **API extraction** and **web scraping** using the BeautifulSoup HTML parser.
2. **Perform Data Wrangling (Describe how data was processed) -** Processed, cleaned, and transformed datasets using **Pandas**, including loading, shaping, and preparing data for analysis.
3. **Perform Exploratory Data Analysis (EDA) using Visualization and SQL**
   - Used **SQL** for sorting and filtering launch records.
   - Conducted visual EDA using **Plotly** and **Seaborn** to uncover trends and patterns.
4. **Perform Interactive Visual Analytics using Folium and Plotly Dash**
   - Built **interactive launch site maps** with Folium.
   - Developed a **SpaceX analytics dashboard** using Plotly Dash for dynamic, user-driven insights.
5. **Perform Predictive Analysis Using Classification Models -** Applied supervised machine learning with **scikit-learn**, including model building, hyperparameter tuning, and evaluation to predict Falcon 9 first-stage landing success.

Data Collection

1. Combined all sources into a **single unified DataFrame** for analysis
2. Retrieved launch data using the **SpaceX REST API**
3. Web-scraped additional details from the **SpaceX page**

Data Collection – SpaceX API

1. Request/Parse Data
   - Request data from API
   - Decode content as .JSON format
   - Convert to data frame
2. Filter Dataframe
   - Isolate Falcon 9 launches within the data frame
3. Remove Missing Values
   - Use isnull() to replace missing values with calculated .mean()
   - Export to CSV
4. GitHub URL - SpaceX-Analysis-Project/1.jupyter-labs-spacex-data-collection-api.ipynb at main · Cbandaru/SpaceX-Analysis-Project

Data Collection - Scraping

1. Pull data from URL
   - Use get() method to extract data
   - Create Beautiful Soup object from html
2. Extract Column Names
   - Extract all of the column names from the html table header (‘th’ element)
3. Create a Dataframe
   - Parse the html tables into a Pandas DataFrame
   - Export to CSV
4. GitHub URL - SpaceX-Analysis-Project/2.jupyter-labs-webscraping.ipynb at main · Cbandaru/SpaceX-Analysis-Project

Data Wrangling

1. Data Analysis Tasks
   - Cleaned dataset using Pandas and Numpy
   - Calculated the number of launches on each site
   - Calculated the number and occurrence of each orbit
   - Calculated the number and occurence of mission outcome of the orbits
   - Created a landing outcome label from Outcome column
2. Convert complicated outcomes into Boolean results
   - 1 = booster successfully landed
   - 0 = booster failed to land
3. GitHub URL - SpaceX-Analysis-Project/3.labs-jupyter-spacex-Data wrangling.ipynb at main · Cbandaru/SpaceX-Analysis-Project

EDA with SQL

1. Explored SpaceX mission data by identifying launch sites, filtering specific records, and summarizing payload statistics (total, average, and maximum).
2. Analyzed mission performance by tracking successful/failed landings, finding first success dates, and checking conditions like payload ranges and landing types.
3. Performed time-based and ranking analyses to understand trends (e.g., outcomes by years/months and landing outcome rankings over a date range).
4. GitHub URL - https://github.com/Cbandaru/SpaceX-Analysis-Project/blob/main/4.jupyter-labs-eda-sql-coursera_sqllite.ipynb

EDA with Data Visualization

1. Analyzed SpaceX launch data by visualizing key relationships such as flight number, launch site, payload mass, orbit type, and yearly success trends to understand mission patterns.

2. Evaluated performance metrics like success rates across different orbit categories to identify which orbits and launch conditions lead to better mission outcomes.

3. Prepared and optimized the dataset through feature engineering—creating dummy variables and converting numeric columns—ensuring the data is ready for model building and deeper analysis.

4. GitHub URL - [SpaceX-Analysis-Project/5.edadataviz.ipynb at main · Cbandaru/SpaceX-Analysis-Project]()

Build an Interactive Map with Folium

1. Added markers and circle markers to clearly show the locations of all SpaceX launch sites and indicate individual launch outcomes (success or failure), making the map visually informative at a glance.

2. Used lines and distance calculations to display proximity between launch sites and nearby facilities, helping understand geographic relationships and logistical factors.

3. Overall goal: create an interactive map that visually represents launch locations, mission results, and spatial distances, making the data easier to interpret and analyze.

4. GitHub URL -[SpaceX-Analysis-Project/6.lab_jupyter_launch_site_location.ipynb at main · Cbandaru/SpaceX-Analysis-Project]

Build a Dashboard with Plotly Dash

1. Added interactive visualizations such as a pie chart (showing launch success distribution) and a scatter plot (showing payload vs. success), enabling users to explore mission outcomes visually.
2. Included dynamic controls like a launch-site dropdown and a payload-range slider, allowing users to filter and update the graphs instantly based on their selections.
3. Built a responsive dashboard where charts automatically adjust according to user inputs, making the analysis intuitive and fully interactive.
4. GitHub URL
   - SpaceX-Analysis-Project/7.spacex-dash-app.py at main · Cbandaru/SpaceX-Analysis-Project
   - SpaceX-Analysis-Project/7.1.dash screenshot.png at main · Cbandaru/SpaceX-Analysis-Project

Predictive Analysis (Classification)

1. Prepared the data properly by creating the target label, standardizing features, and splitting the dataset into training and testing sets to ensure fair model evaluation.
2. Trained and tuned multiple models (SVM, Decision Tree, Logistic Regression) by testing different hyperparameters to improve accuracy and overall performance.
3. Compared model results on test data to identify which algorithm performed best, selecting the most accurate and reliable classification model.
4. GitHub URL - SpaceX-Analysis-Project/8.SpaceX_Machine Learning Prediction_Part_5.ipynb at main · Cbandaru/SpaceX-Analysis-Project

Results

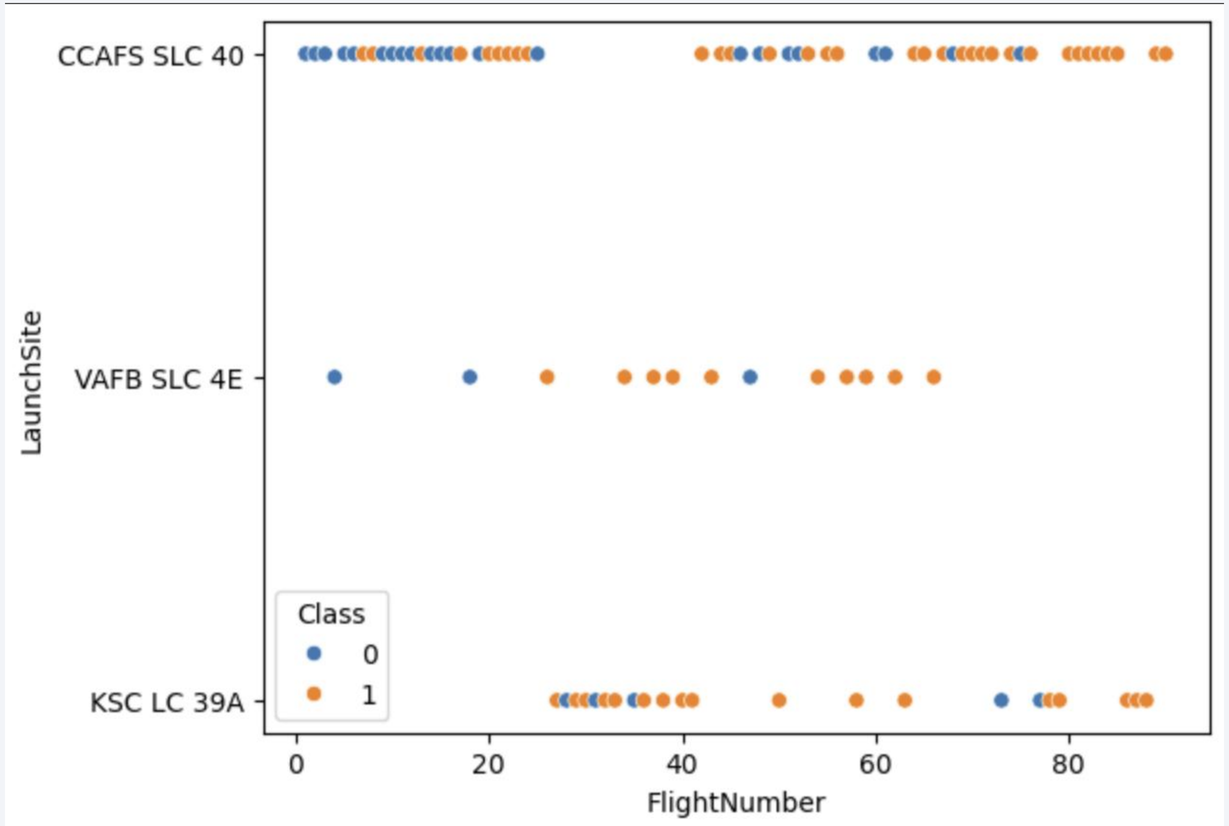
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

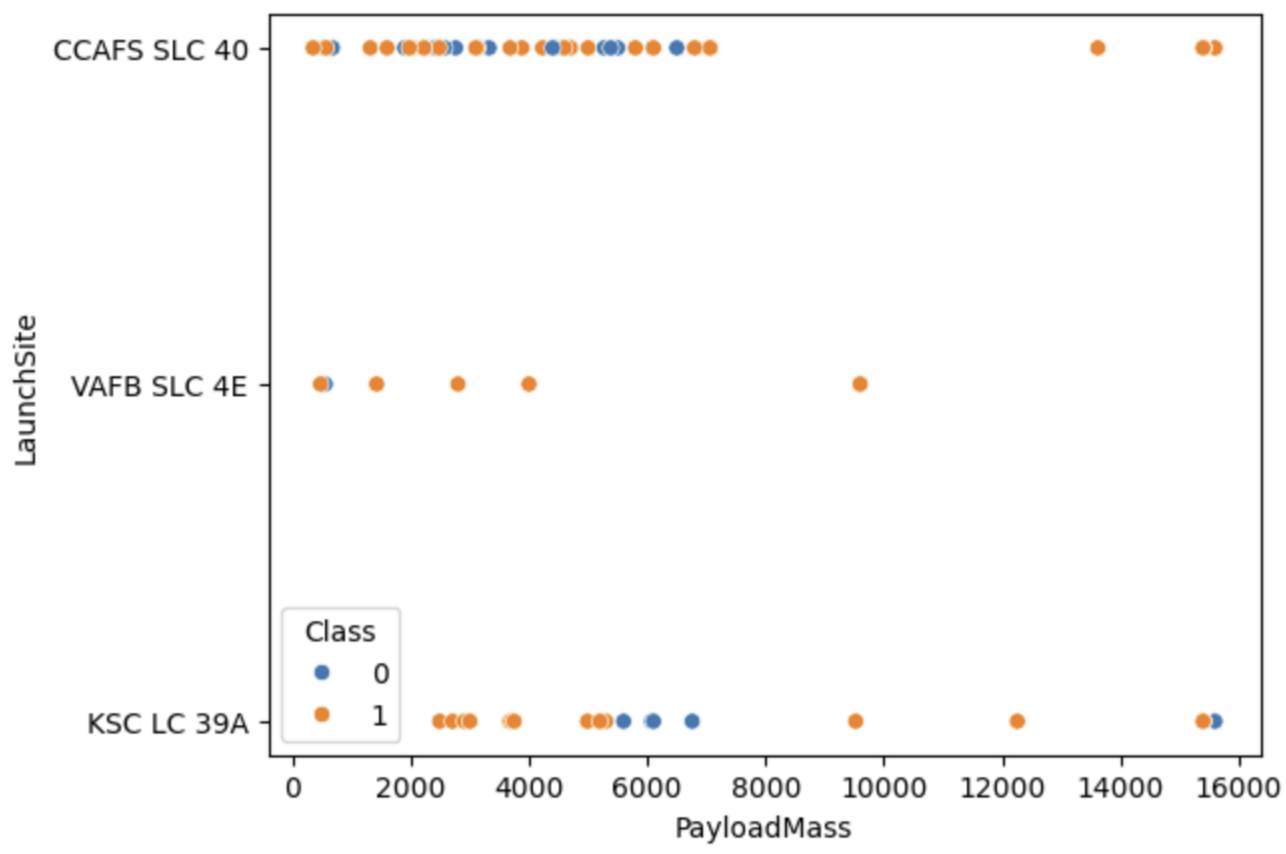
Flight Number vs. Launch Site

- Shown below is a scatter plot of Flight Number vs. Launch Site



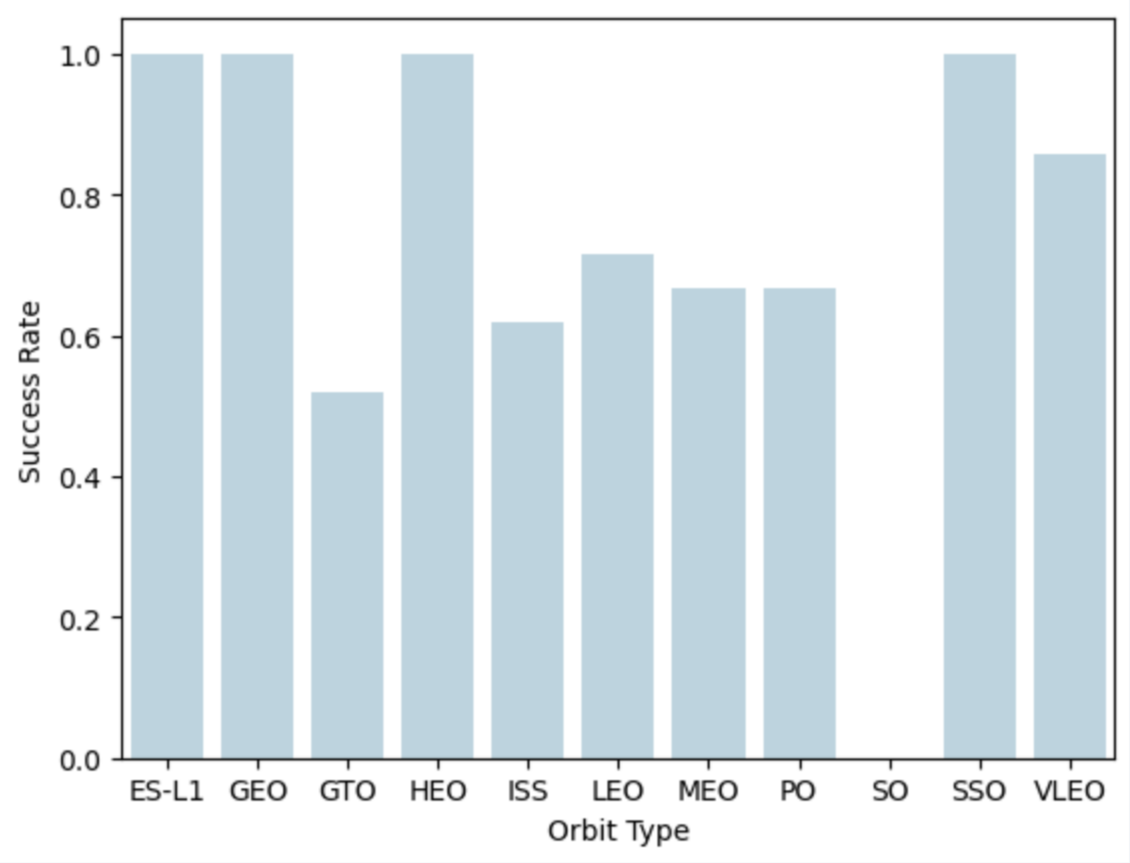
Payload vs. Launch Site

- Shown below is a scatter plot of Payload vs. Launch Site



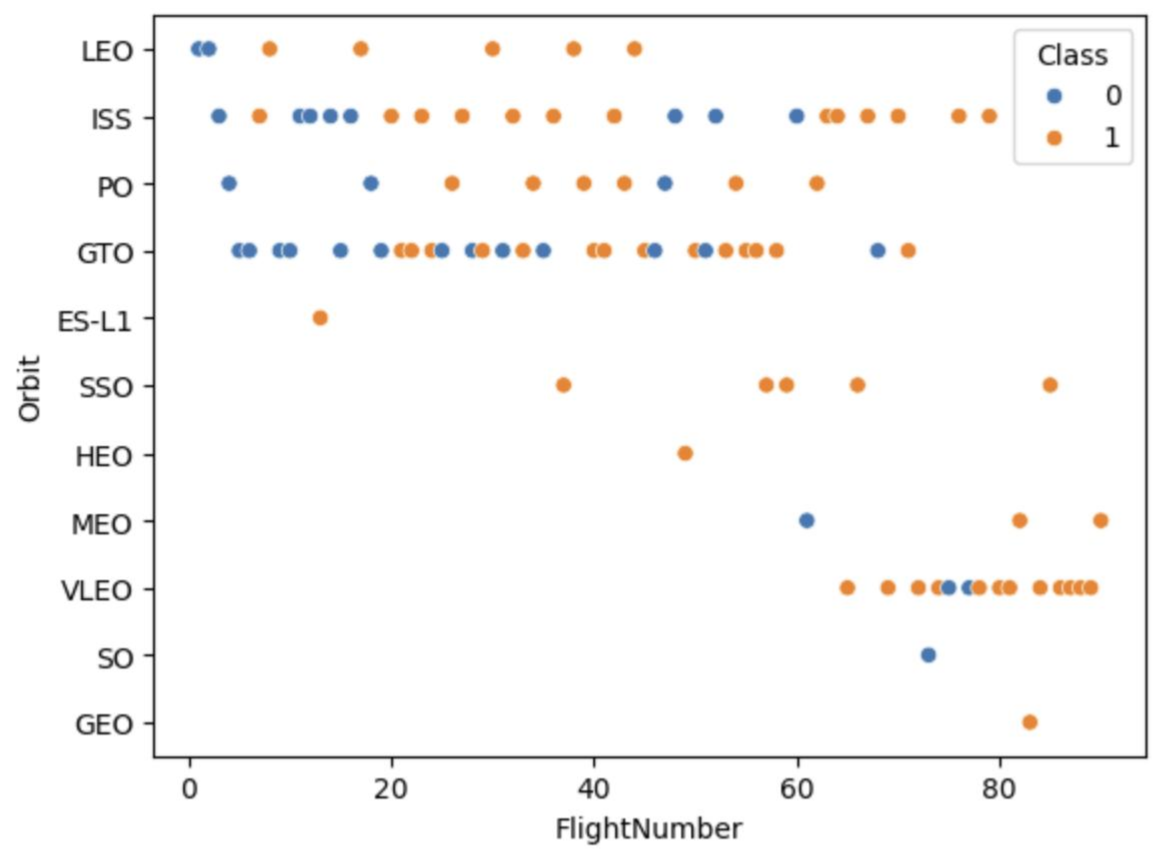
Success Rate vs. Orbit Type

- Shown below is a bar chart for the success rate of each orbit type



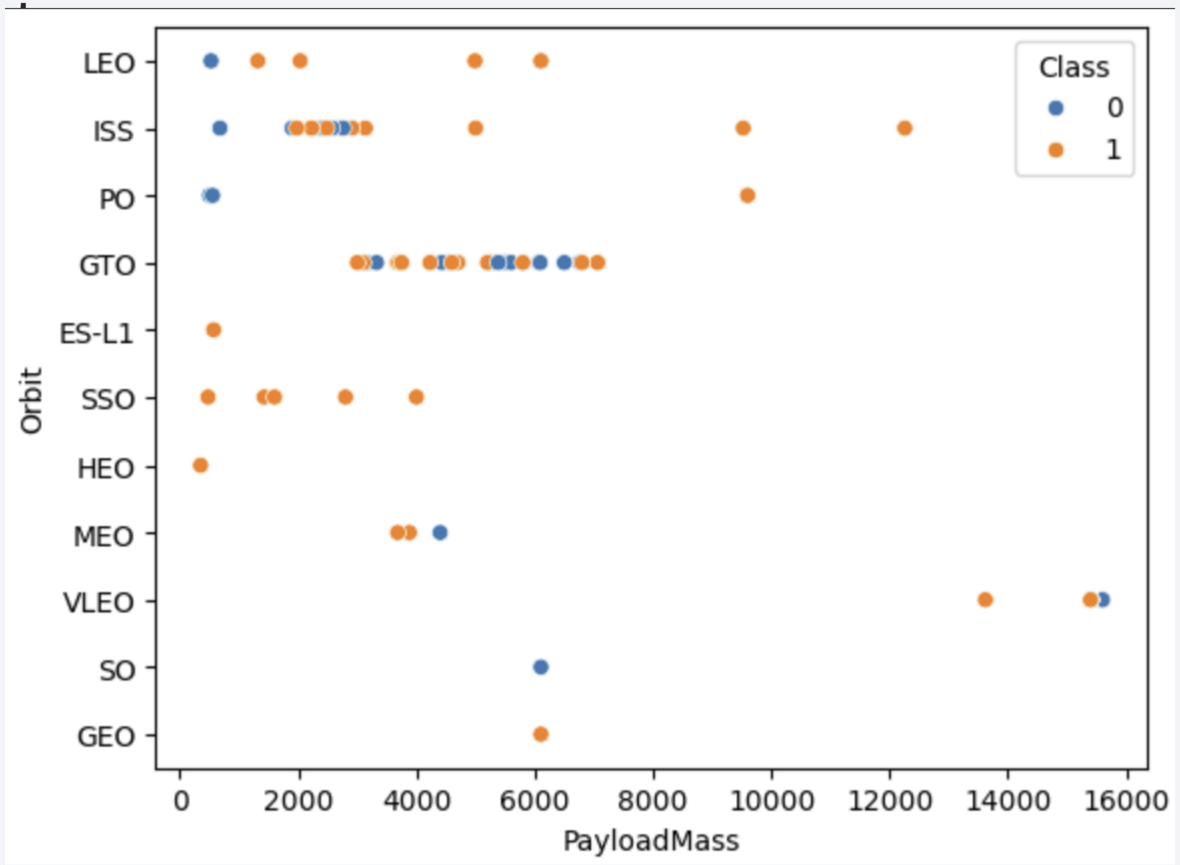
Flight Number vs. Orbit Type

- Shown below is a scatter point of Flight number vs. Orbit type



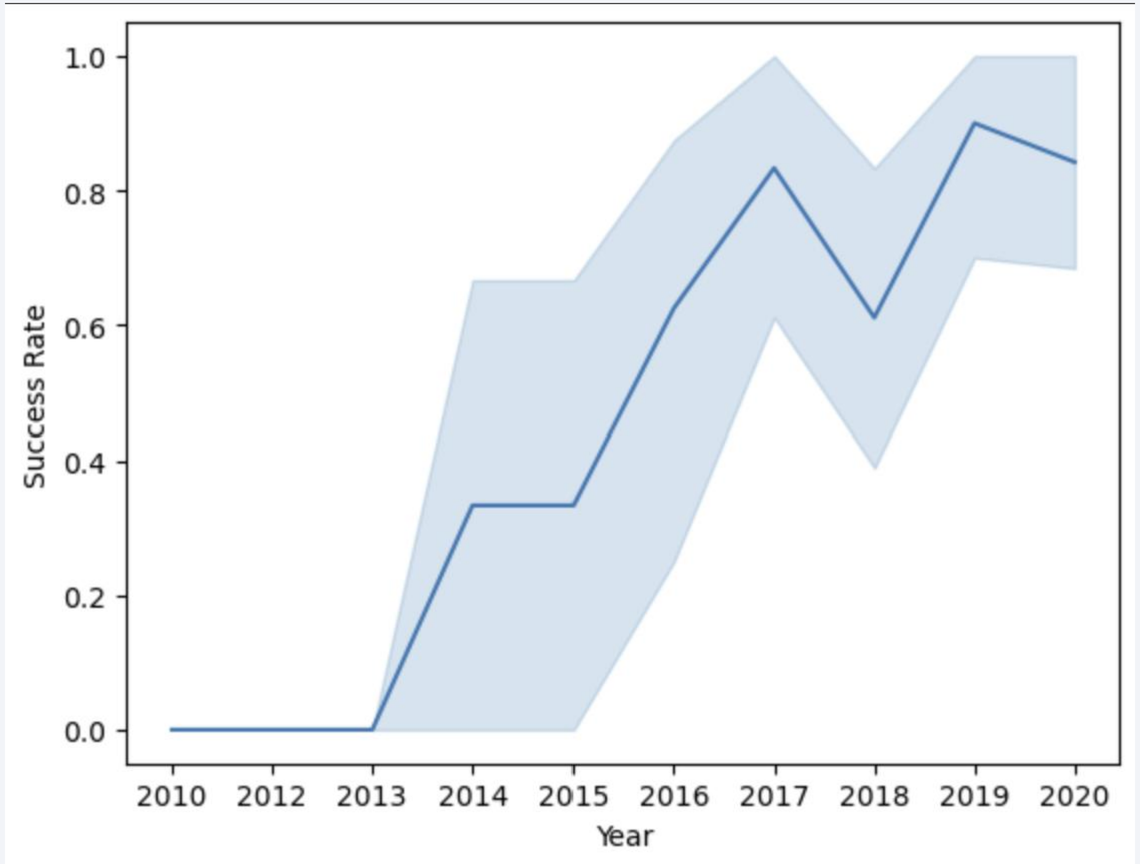
Payload vs. Orbit Type

- Shown below is a scatter point of payload vs. orbit



Launch Success Yearly Trend

- Shown below is a line chart of yearly average success rate



All Launch Site Names

- Shown below are the names of the unique launch sites

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

```
In [11]: %sql select distinct Launch_Site from SPACEXTABLE

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

Out[11]:

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

Launch Site Names Begin with 'CCA'

- Shown below are the 5 records where launch sites begin with `CCA`

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

[13]:
```
%sql select * from SPACEXTBL where Launch_Site like'CCA%' limit 5
```

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

[13]:

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

- Shown below is the calculation of the total payload carried by boosters from NASA

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

```
[13]: %sql select sum(PAYLOAD_MASS__KG_) as 'Total Payload Mass' from SPACEXTBL where Customer='NASA (CRS)'

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

[13]:

Total Payload Mass
45596

Average Payload Mass by F9 v1.1

- Shown below is the calculation of the average payload mass carried by booster version F9 v1.1

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

```
[23]: %sql select avg(PAYLOAD_MASS__KG_) as 'Average Payload Mass' from SPACEXTBL where Booster_Version like 'F9 v1.1%'
```

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

[23]:

Average Payload Mass
2534.6666666666665

First Successful Ground Landing Date

- Shown below are the dates of the first successful landing outcome on ground pad

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

Hint:Use min function

[15]:

```
%sql select  date from SPACEXTABLE where Landing_Outcome= 'Success (ground pad)'
```

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

[15]:

Date
2015-12-22
2016-07-18
2017-02-19
2017-05-01
2017-06-03
2017-08-14
2017-09-07
2017-12-15
2018-01-08

Successful Drone Ship Landing with Payload between 4000 and 6000

- Shown below are the list of names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
[17]: %sql select  Booster_Version from SPACEXTABLE where Landing_Outcome= 'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000

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

[17]:

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

Total Number of Successful and Failure Mission Outcomes

- Shown below are the calculation the total number of successful and failure mission outcomes

List the total number of successful and failure mission outcomes

[18]:

```
%sql select  Mission_Outcome, count(*) as Totals from SPACEXTABLE group by Mission_Outcome
```

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

[18]:

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

Boosters Carried Maximum Payload

- Shown below is list of names of the booster which have carried the maximum payload mass

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

[24]:
```
%sql select distinct Booster_Version from SPACEXTABLE where  PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTABLE)
```

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

[24]:

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

2015 Launch Records

- Shown below are the list of the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

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.

[28]:
```
%sql SELECT  substr(Date, 6, 2) AS Month,Landing_Outcome, Booster_Version, Launch_Site \
    FROM SPACEXTABLE WHERE substr(Date, 1, 4) = '2015'  AND Landing_Outcome ='Failure (drone ship)'
```

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

[28]:

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

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

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.

[30]:
```
%sql SELECT  Landing_Outcome, COUNT(*) AS OutcomeCount FROM SPACEXTABLE \
    WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' \
GROUP BY Landing_Outcome ORDER BY OutcomeCount DESC;
```

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

[30]:

Landing_Outcome	OutcomeCount
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

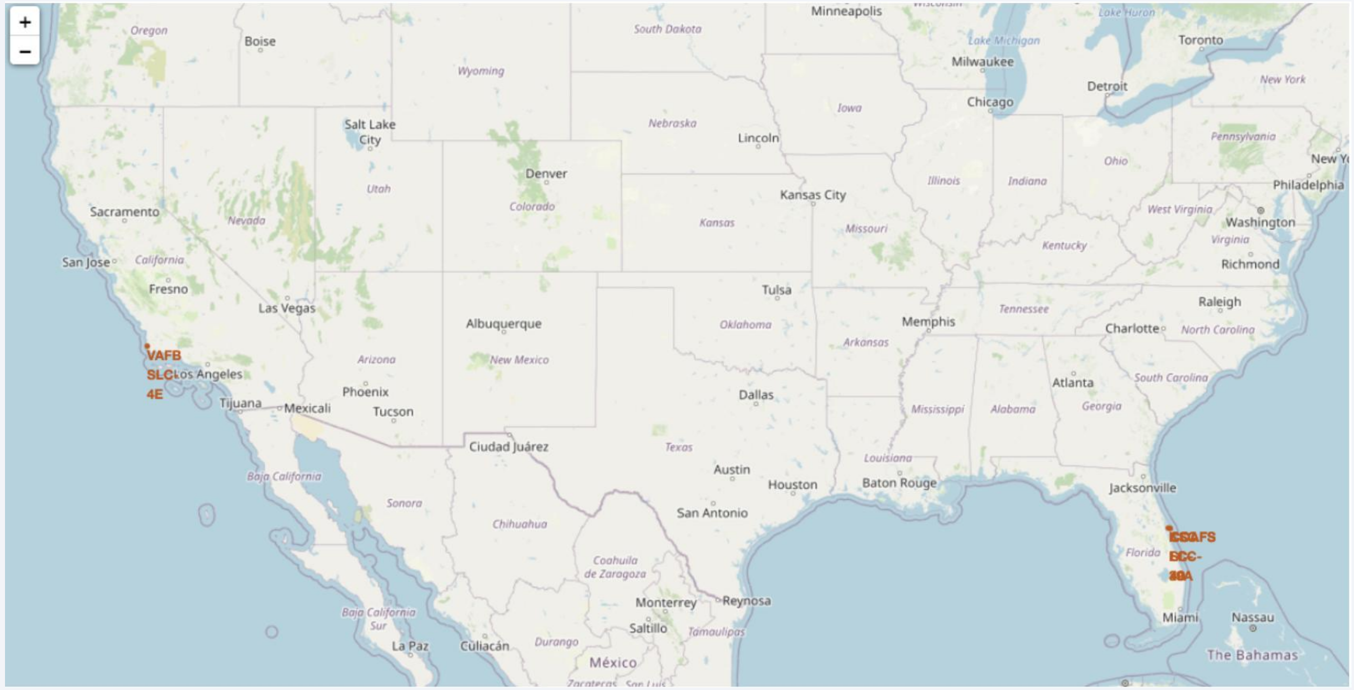
Section 3

Launch Sites Proximities Analysis

Mark all launch sites on a map

This map displays the location of SpaceX launch sites

- Note that all launch sites are in proximity to the equator line
  - The land at the equator is moving 1670 km per hour, so launching from the equator makes the spacecraft move faster and stay in orbit once it is launched
- Note that all launch sites are very close to the coast
  - Rockets launched over the ocean reduce the risk of debris dropping or exploding near people



Mark the success/failed launches on the map

This map displays the success rates of launches at the KSC LC-39A launch site

- Green = Successful Launch
- Red = Failed Launch

We can clearly verify from this graphic that the KSC LC-39A launch site has a very high success rate

Calculate the distances between a launch site to its proximities

- This map displays the distance from KSC LC-39A launch site to some important geographic locations:
- Failed rockets can cover distances such as 15-20km in a matter of seconds, which could be dangerous in populated areas.

Section 4

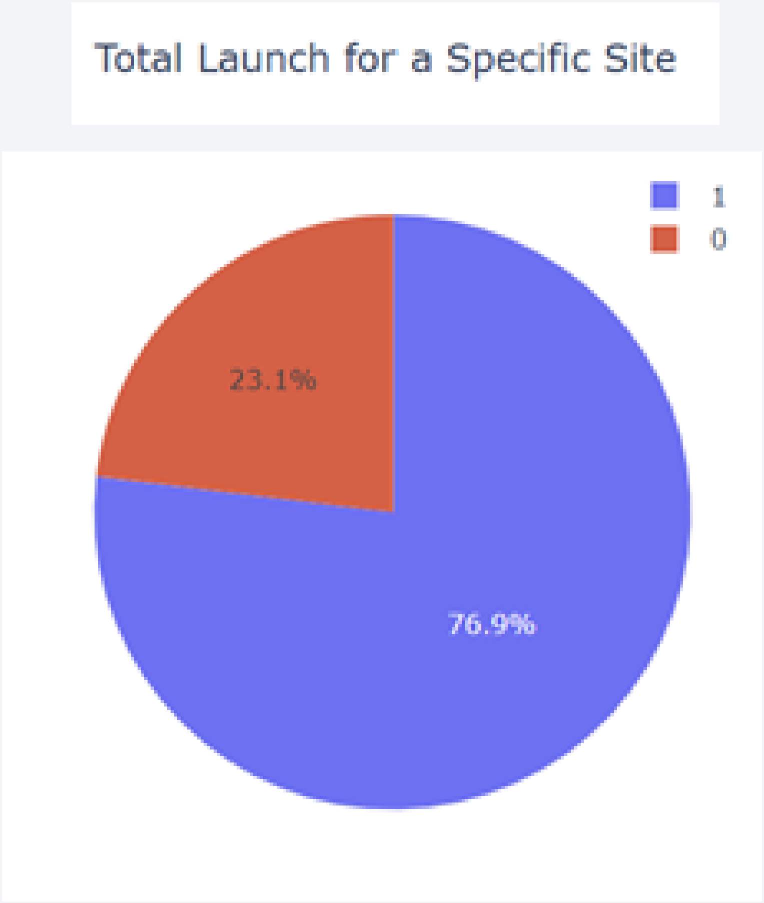
Build a Dashboard with Plotly Dash

Launch Analysis Dashboard

- This screenshot from the Launch Analysis Dashboard displays the total number of launches organized by launch site.
- We can see that the KSC LC-39A launch site accounts for 41.7% of all launches
- We could assume that the high success rate at KSC LC-39A might be correlated with the high number of launches

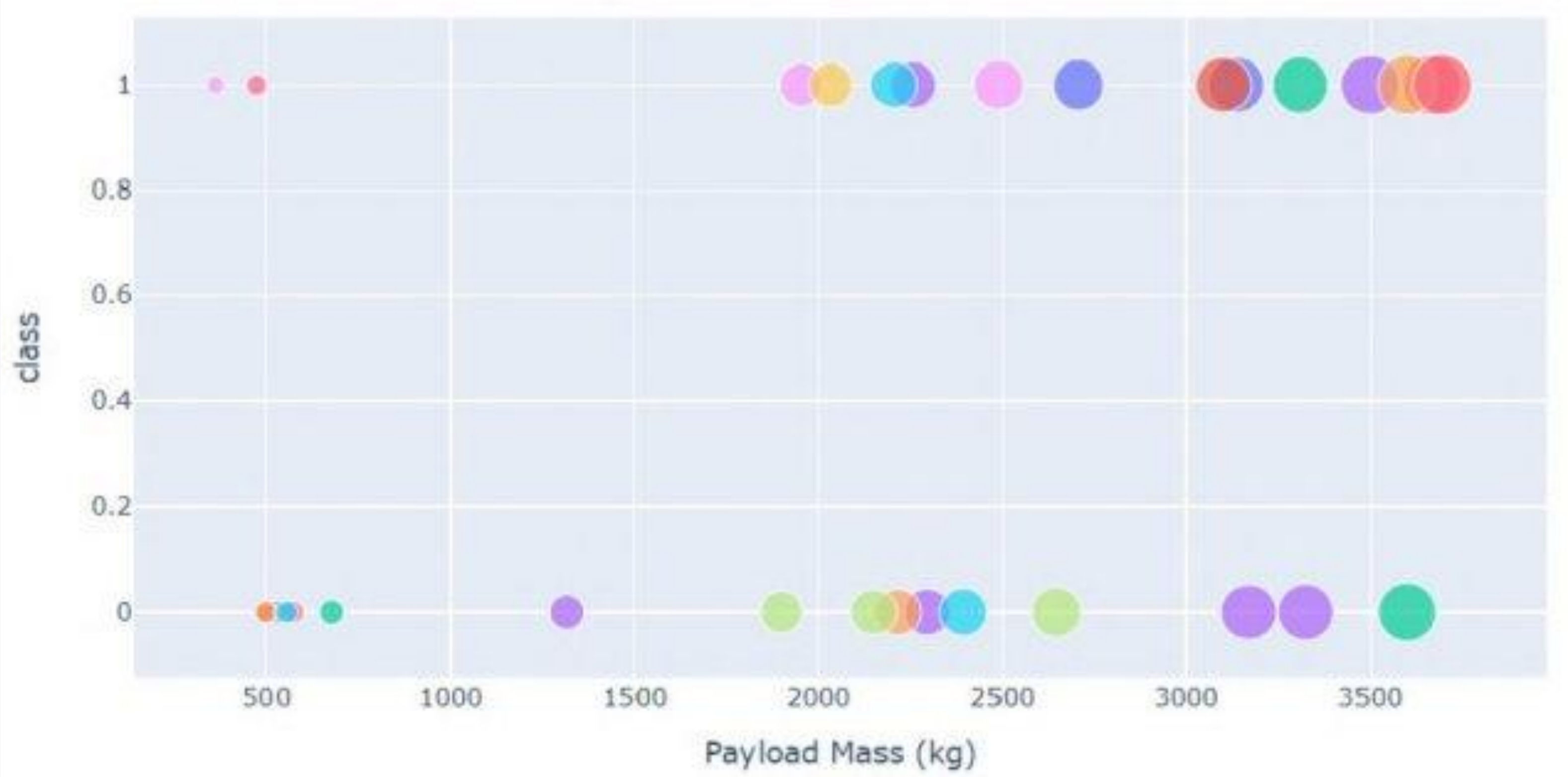
Launch Analysis Dashboard

- This screenshot from the dashboard visualizes the success rate of launches from KSC LC-39A.
- We can see that this site had a 76.9% success rate, the highest among all launch sites



Launch Analysis Dashboard

- This chart visualizes success rate by orbit type and payload mass.
  - 1=Success
  - 0=Failure
- When payload mass is too low, the launches are prone to failure.



Section 5

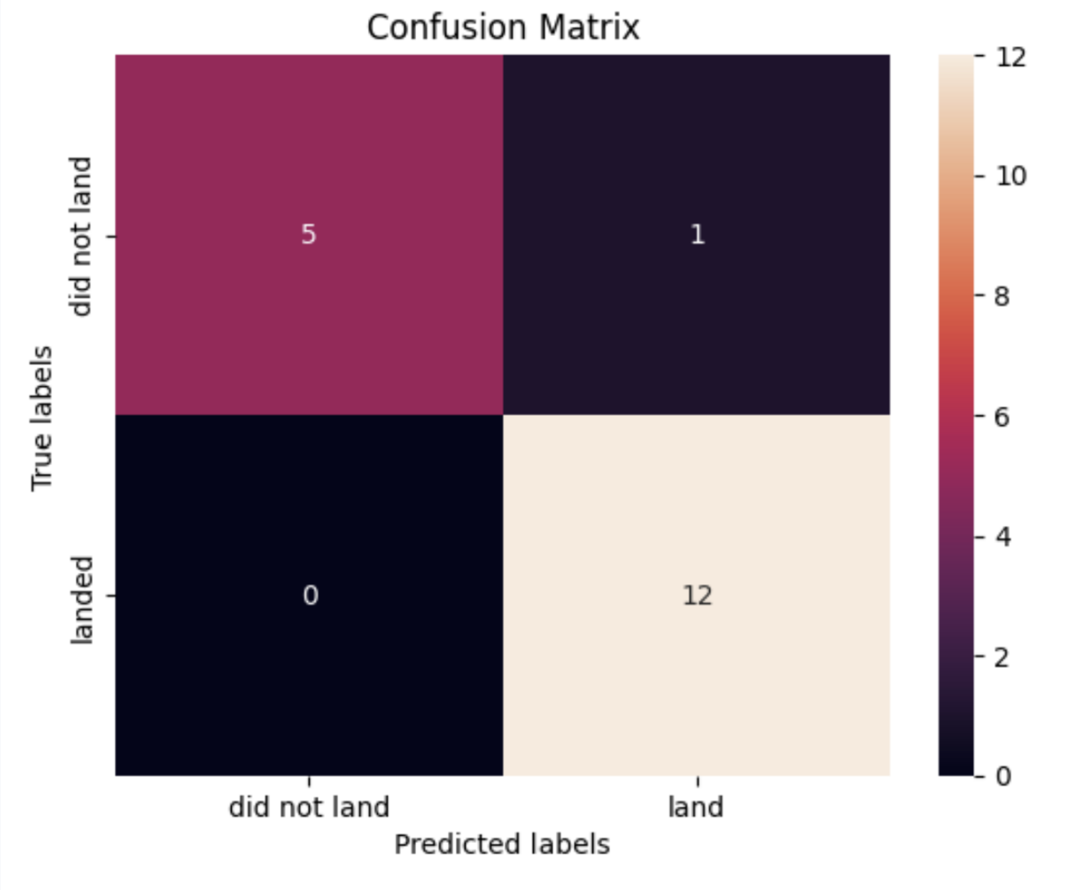
Predictive Analysis (Classification)

Classification Accuracy

- Tested four different supervised machine learning classification algorithms on the dataset.
- The bar chart to the left clearly shows that a Decision Tree Model resulted in the highest accuracy in predicting the success or failure of future SpaceX Falcon 9 rocket launches.

Confusion Matrix

- The confusion matrix displayed on the left indicates that a decision tree model was 83.33% accurate in predicting whether rockets will land successfully compared to the testing data set.



Conclusions

- Chances of success have increased over time
- KSC LC 39A is the most successful launch site with 76.9% success rate
- Medium payloads (2000-5000kg) succeed most often compared to small or large payloads
- The most accurate model for predicting the success of launches is a Decision Tree Model
- The Decision Tree model is very accurate (83.33%) at predicting whether a rocket will successfully land

Appendix

- Included all the relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project are present in the GitHub link - Cbandaru/SpaceX-Analysis-Project

Thank you!