IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Edwin Berbey
09-19-2023

Presentation Contents

P3	Executive Summary
P4	Introduction
P6	Methodology
P16	EDA With Visualization
P17	EDA With SQL
P19	Interactive Map with Folium
P45	Dashboard With Plotly
P50	Predictive Analysis
P52	Conclusion

Executive Summary

Methodology Summary

- Data Collection: We used various techniques to gather data:
    - APIs: We extracted data about SpaceX launches using their API.
    - Web Scraping: We implemented web scraping techniques to obtain additional data that was not available through the API.
- Exploratory Data Analysis:
    - SQL: We performed SQL queries to explore and analyze the launch dataset.
    - Data Visualization: We used tools like matplotlib and seaborn to visualize and understand trends and patterns in our data.
- Launch Site Location Analysis: Using Folium, we conducted an interactive analysis of launch site locations and how they may influence the launch success rate.
- Machine Learning Prediction: We implemented machine learning models to predict whether SpaceX's Falcon 9 first stage will successfully land or not.
- Interactive Visualization: We developed an interactive application using Plotly Dash to visualize the data and results in a more intuitive and understandable manner.

Executive Summary

- Summary of Results
  - We have identified that choosing an optimal location for building a launch site is crucial and can significantly influence the launch success rate.
  - Through visual analysis, we were able to identify specific trends and patterns related to different factors, such as payload mass, orbit type, among others.
  - Our machine learning model provides a valuable tool for predicting the success of the first stage landing, which can be crucial in determining the cost and feasibility of a launch.

Introduction

- Project background and context
  - SpaceX has revolutionized the space industry by reusing the first stage of its Falcon 9 rockets, significantly reducing launch costs. Unlike other providers that charge over 165 million dollars per launch, SpaceX offers launches for 62 million dollars thanks to this reusability.
- Problems you want to find answers
  - Landing Prediction: Is it possible to predict the success of the Falcon 9's first-stage landing?
  - Launch Site Location: Does the location affect the launch success rate?
  - Success Factors: What factors, besides location, influence the success of the launch?


Methodology

Methodology

- Executive Summary
- Data Collection Methodology
    - APIs: We used SpaceX's API to obtain detailed data on launches.
    - Web Scraping:We implemented web scraping techniques to supplement the information obtained from the API.
    - Data Processing (Data Wrangling):
    - I transformed the data to remove missing values and outliers.
    - converted and normalized variables to ensure consistency in the dataset.
- Exploratory Data Analysis (EDA):
    - Visualization: We used tools like matplotlib and seaborn to visualize trends and patterns.
    - SQL: We performed specific queries to discover relationships and correlations among different variables.
- Interactive Visual Analysis:
    - Folium: We analyzed launch site locations and their impact on success rates.
    - Plotly Dash: We developed interactive visualizations to facilitate data understanding and exploration.
- Predictive Analysis:
    - Classification Models: We built and fine-tuned models to predict the success of the first-stage landing.
    - Construction: We used algorithms such as logistic regression and decision trees.
    - Tuning: We optimized model parameters to improve accuracy.
    - Evaluation: We assessed model performance using metrics such as accuracy, recall, and F1-score.

Data Collection

Data collection was essential for our analysis, and it was carried out using multiple techniques to obtain a comprehensive and robust dataset.

Description of the Data Collection Process:

SpaceX API:

- We accessed SpaceX's official API to obtain detailed information about launches.
- This API provided us with data such as launch date, launch site location, landing success, and other relevant details.

Web Scraping:

- To complement the data obtained from the API, we turned to web scraping.
- We used specialized Python libraries to extract information from websites related to rocket launches and space events.
- This technique allowed us to gather additional data that was not available through the API.

Data Collection

Below is a simplified flowchart of the process:

```
Start
|
v
Access SpaceX API
|
v
Complete Data?
| No
+-----> Web Scraping
|
v
Data Consolidation
|
v
End
```

Data Collection – SpaceX API

- Data Collection – SpaceX API

To access detailed and up-to-date information regarding SpaceX launches, we utilized SpaceX's REST API. This API offers a wide array of endpoints that furnish data on rockets, capsules, launches, and more.

Process of SpaceX API Calls

Launches Endpoint:

We initiated calls to this endpoint to acquire data concerning all SpaceX launches.

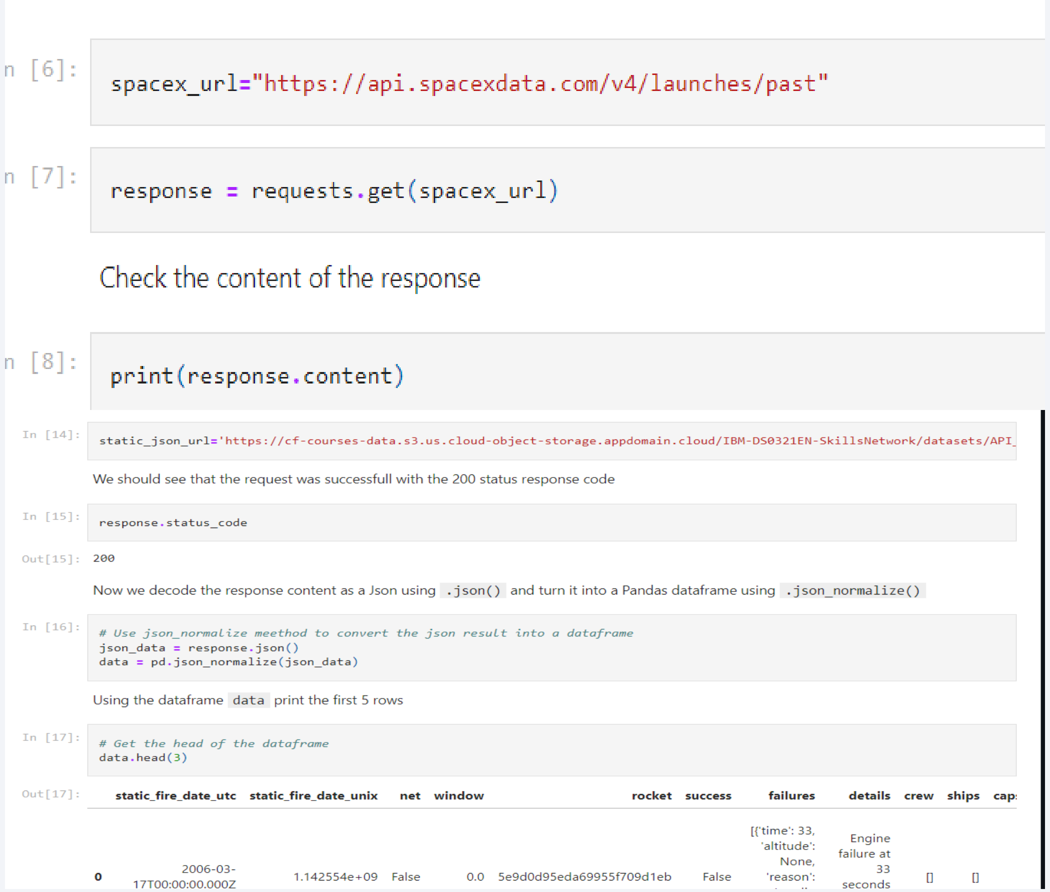
Each call yielded details such as the mission, launch date, landing success, payload, and other pertinent information.

Response Processing:

API responses were received in JSON format.

We parsed and transformed these responses to seamlessly integrate them into our primary dataset.

- External Reference: For a detailed review of the code and API call responses, you can refer to the Jupyter notebook containing all SpaceX API calls at the following GitHub link: https://github.com/TIMMY11-90/IBM-Data-Science-Capstone_SpaceX/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

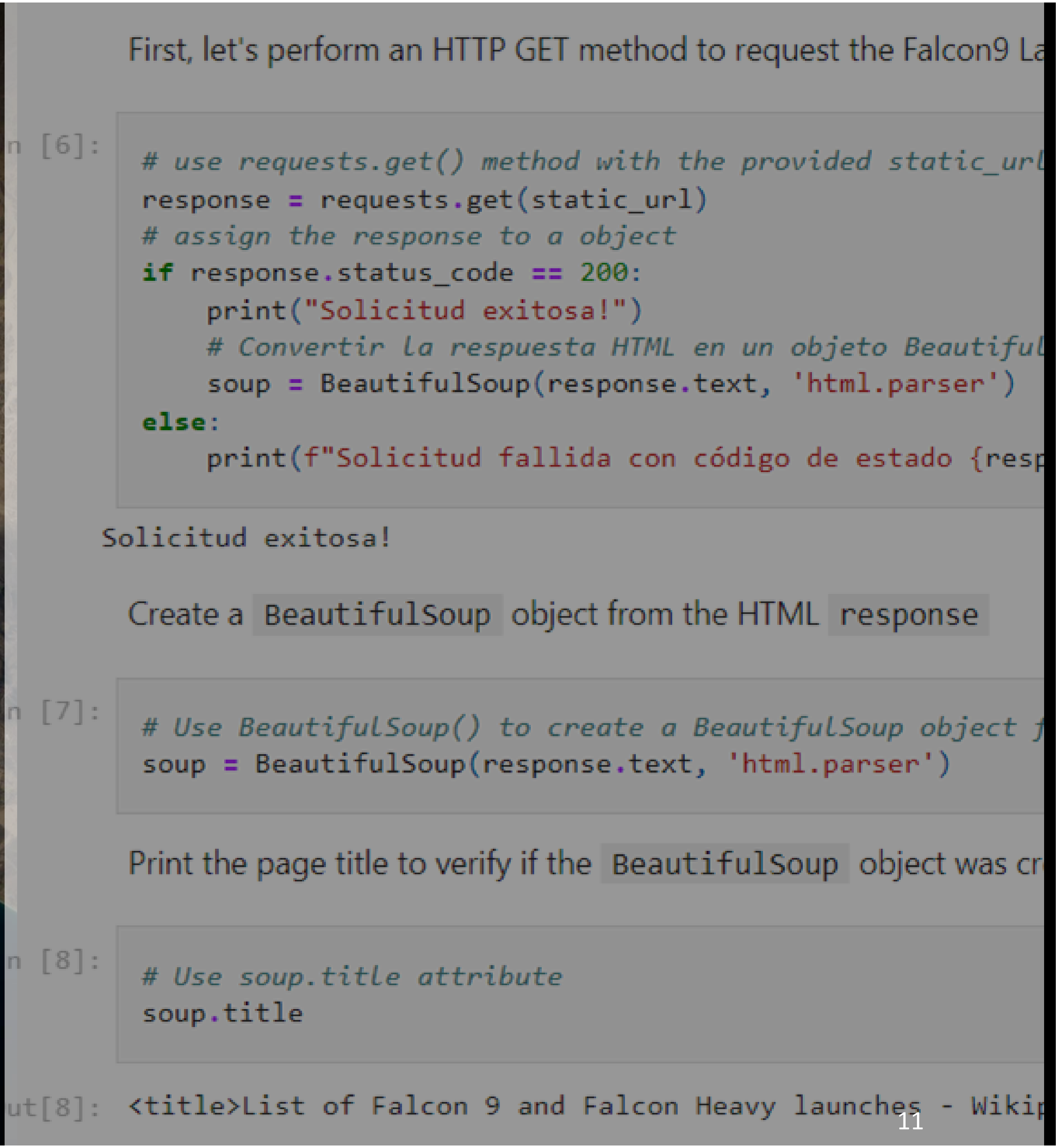


Data Collection - Web Scraping

Web scraping is a powerful technique that allowed us to gather additional data not available through the SpaceX API. We used specialized Python libraries to extract information from websites related to rocket launches and space events.

Web Scraping Process

- Source Identification:
- We selected specific websites that provide detailed information about SpaceX launches.
- Page Request:
- We made HTTP requests to fetch the webpage's content.



First, let's perform an HTTP GET method to request the Falcon9 La

In [6]:

```python
# use requests.get() method with the provided static_url
response = requests.get(static_url)
# assign the response to a object
if response.status_code == 200:
    print("Solicitud exitosa!")
    # Convertir la respuesta HTML en un objeto Beautiful
    soup = BeautifulSoup(response.text, 'html.parser')
else:
    print(f"Solicitud fallida con código de estado {resp
```

```
Solicitud exitosa!
```

Create a `BeautifulSoup` object from the HTML `response`

In [7]:

```python
# Use BeautifulSoup() to create a BeautifulSoup object f
soup = BeautifulSoup(response.text, 'html.parser')
```

Print the page title to verify if the `BeautifulSoup` object was cr

In [8]:

```python
# Use soup.title attribute
soup.title
```

Out[8]:

```
<title>List of Falcon 9 and Falcon Heavy launches - Wiki
```

Data Collection – Web Scraping

Data Extraction: Specific selectors were used to extract relevant data from the HTML structure of the page.

I transformed the extracted data into a structured format for further analysis.

Storage:

The extracted data was stored in our main dataset for in-depth analysis.

Below is a simplified flowchart of the web scraping process:

```
Start
  |
  v
Source Identification
  |
  v
HTTP Page Request
  |
  v
Successful Response?
  |         No
  +-----> Error Logging
  |
  v
Data Extraction with Specific Selectors
  |
  v
Data Transformation and Storage
  |
  v
End
```

Data Wrangling

Data processing, also known as "data wrangling," is a crucial stage in any data analysis project. In this phase, we prepare and transform the data to facilitate subsequent analysis.

Data Cleaning Phase:

- Identification and handling of missing values.
- Removal of outliers or atypical values.
- Data Transformation Phase:
- Conversion of variables into suitable formats for analysis (e.g., dates).
- Normalization and standardization of variables.

- Data Integration Phase:
  - Combining data from different sources (API and web scraping) to create a unified dataset.
  - Ensuring coherence and consistency in the combined dataset.

Key Transformation: Launch Outcome Classification

```
def classify_outcome(outcome):
    return 0 if outcome in bad_outcomes else 1

df['Class'] = df['Outcome'].apply(classify_outcome)
```

Data Wrangling

- Storage:
  - The processed dataset is saved in a structured format for further analysis and modeling.

External Reference: For a detailed review of the data processing process and the code used, you can refer to the related Jupyter notebooks at the following GitHub link: https://github.com/TIMMY11-90/IBM-Data-Science-Capstone_SpaceX/blob/main/3-labs-jupyter-spacex-Data%20wrangling.ipynb

Below is a simplified flowchart of the data processing process:

```
Start
  |
  v
Identification of Missing Values and Outliers
  |
  v
Data Transformation and Variable Conversion
  |
  v
Integration of Data from Different Sources
  |
  v
Storage of Processed Dataset
  |
  v
End
```

EDA with Data Visualization

- Data: SpaceX rocket launches.
- Key Findings:
  - As the flight number increases, the probability of success also increases.
  - CCAFS LC-40 has a success rate of 60%; KSC LC-39A and VAFB SLC 4E have a success rate of 77%.
  - Clear patterns exist in the relationship between payload mass, orbit type, and landing success.

EDA with Data Visualization

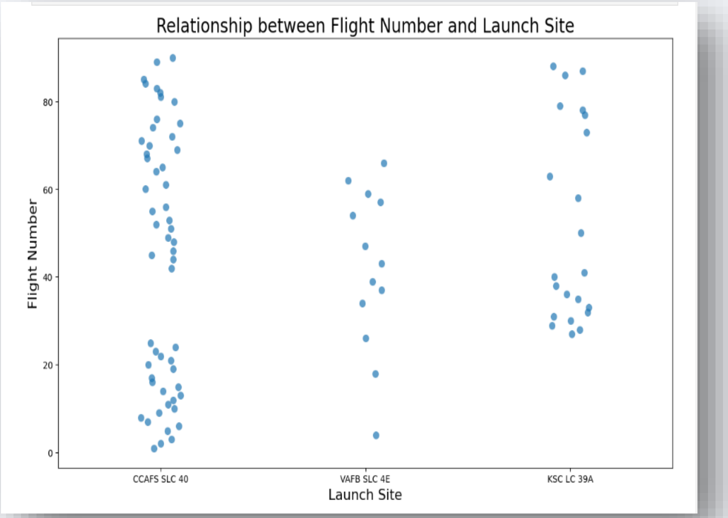
Data Visualization and Feature Engineering

- FlightNumber vs. PayloadMass
- Success Rate by Site and Orbit Type
- Annual Success Trend

Feature Engineering:

- Selection of specific attributes for prediction.
- One-Hot encoding for categories and conversion to float64.

GitHub URL: https://github.com/TIMMY11-90/IBM-Data-Science-Capstone_SpaceX/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

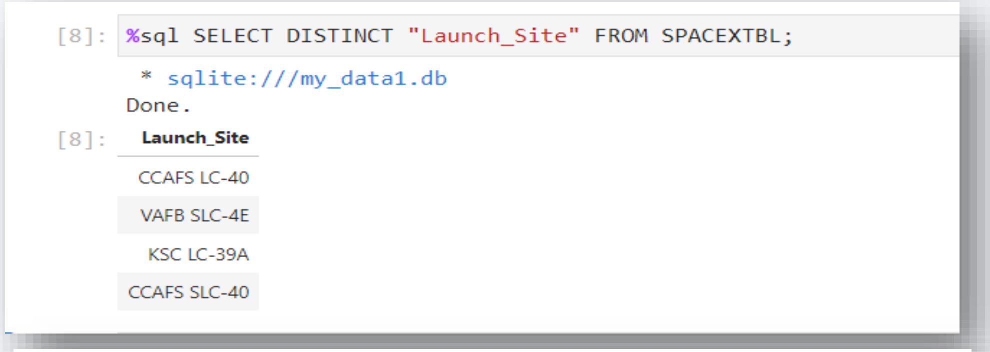


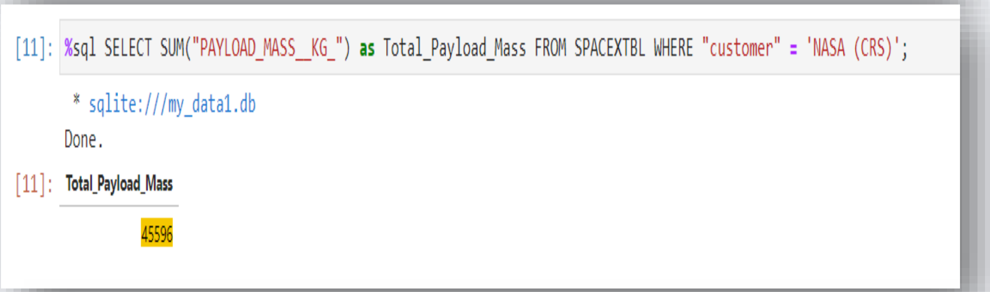
EDA with SQL

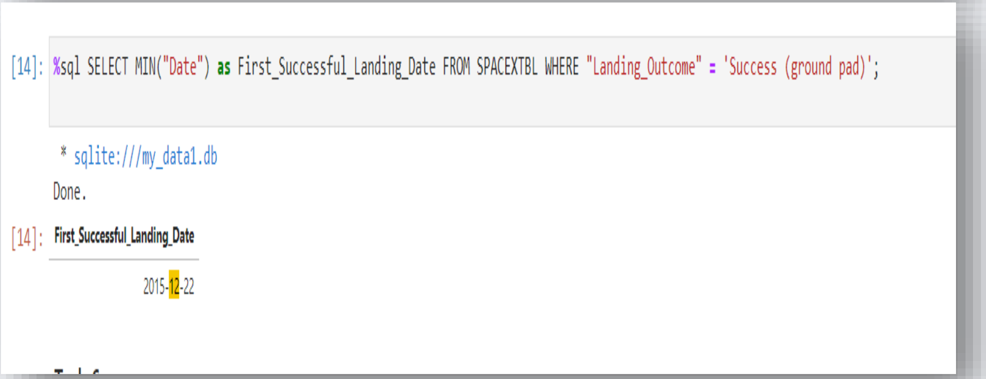
- Introduction:
  - Exploration of SpaceX dataset.
  - Loading the dataset into an SQLite database.
  - Executing SQL queries to answer lab questions.
  - Basic Dataset Information:
  - Contains details of each payload during a SpaceX mission.
  - Key Columns: Date, Launch Site, Payload Mass, Customer, Mission Outcome, and Landing Outcome.
- GitHub URL: https://github.com/TIMMY11-90/IBM-Data-Science-Capstone_SpaceX/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

EDA with SQL

- Examples of Queries Performed
- Unique launch sites in space missions.
  - Result: CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, CCAFS SLC-40.
- Total payload mass carried by rockets launched by NASA (CRS).
  - Result: 45596 kg.
- Date of the first successful ground platform landing.
  - Result: 2015-12-22.
- Rocket versions that have carried the maximum payload mass.
- Examples: F9 B5 B1048.4, F9 B5 B1049.4, F9 B5 B1051.3.
- Range of landing outcomes between '2010-06-04' and '2017-03-20'.
  - Examples: No attempt (10), Successful ground platform (5), Successful drone ship (5).

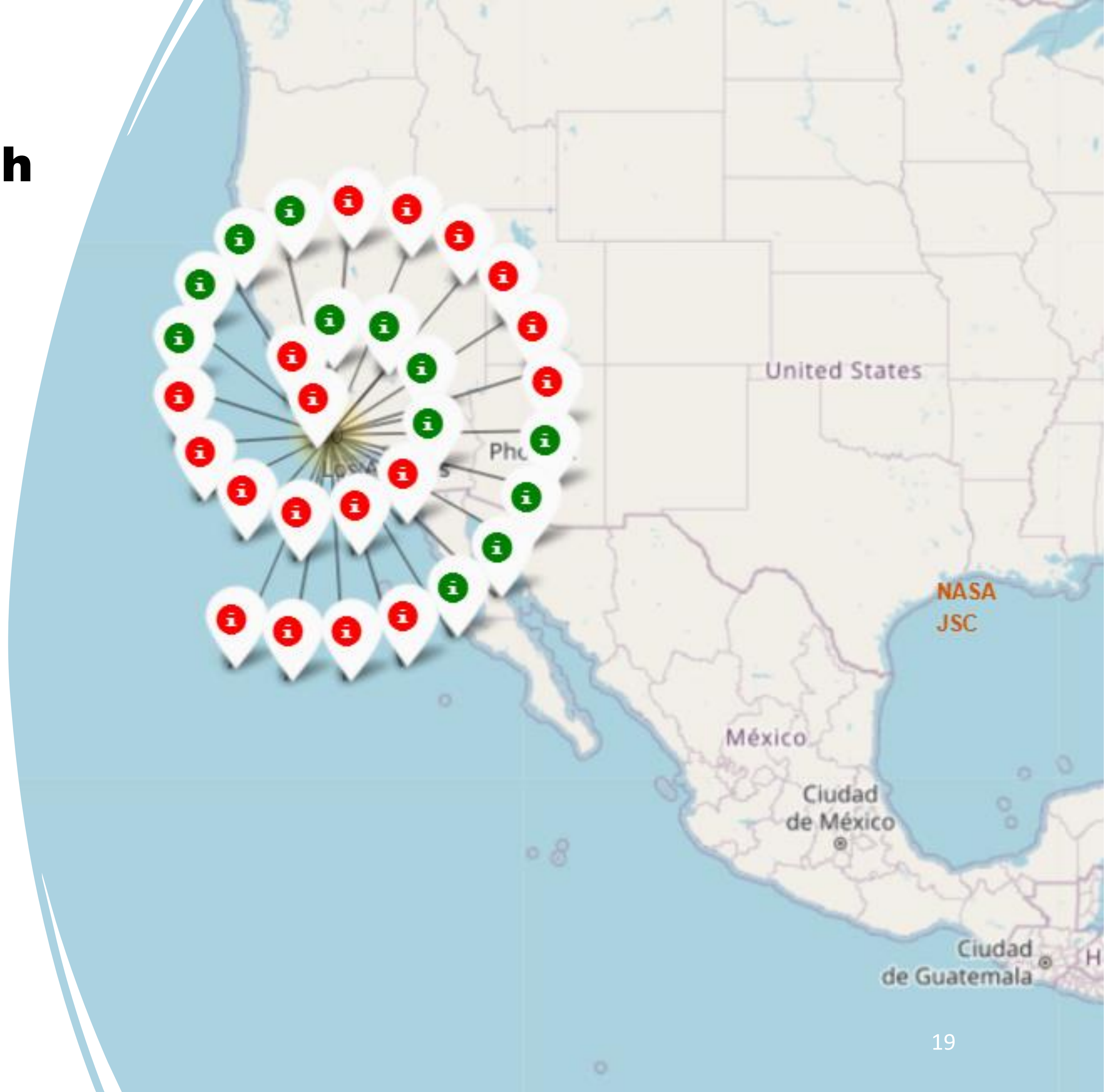






Build an Interactive Map with Folium

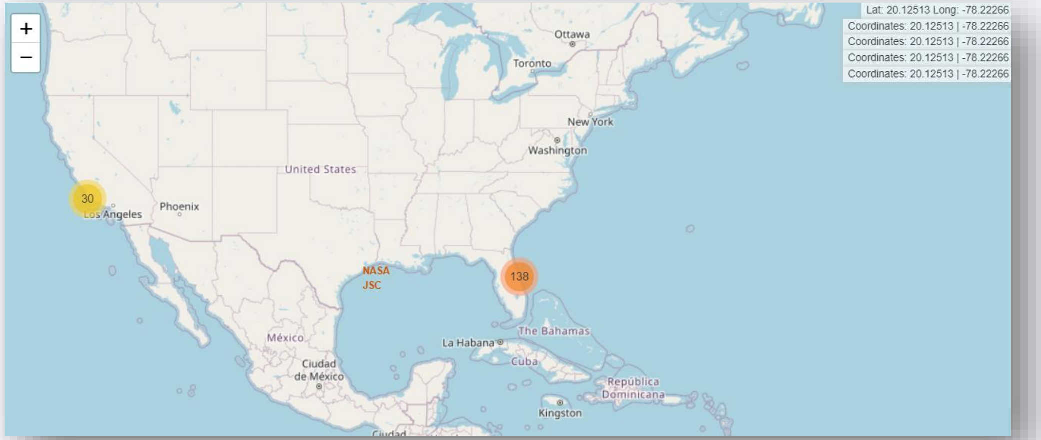
- Objectives:
  - Visualize launch site locations and analyze proximities to discover geographical patterns.
  - Utilize Folium for interactive visual analysis.
- Tasks Performed:
  - Task 1: Marking all launch sites on a map.
  - Task 2: Representing successful and unsuccessful launches for each site on the map.
  - Task 3: Calculating distances between a launch site and its proximities.



Build an Interactive Map with Folium

- Key Findings:
  - Proximity to the equator and the coast can influence launch success rates.
  - Marker clusters allow for easy visualization of success rates for different launch sites.
  - Folium's MousePosition tool facilitates coordinate identification for further analysis.

GitHub URL: https://github.com/TIMMY11-90/IBM-Data-Science-Capstone_SpaceX/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb



Build a Dashboard with Plotly Dash

Objectives: Visualize launch site locations and analyze proximities to discover geographical patterns.

Utilize Folium for interactive visual analysis.

Tasks Performed:

Task 1: Marking all launch sites on a map.

Task 2: Representing successful and unsuccessful launches for each site on the map.

Task 3: Calculating distances between a launch site and its proximities.

Key Findings:Key Findings:

Proximity to the equator and the coast can influence launch success rates.

Marker clusters allow for easy visualization of success rates for different launch sites.

Folium's MousePosition tool facilitates coordinate identification for further analysis.

Predictive Analysis (Classification)

- Objective: Predict whether the first stage of SpaceX's Falcon 9 will successfully land based on historical data.
- 1. Model Development Process:

Data Preprocessing:

- Data loading from URL.
- Creation of feature matrix X and label vector Y.
- Data standardization with preprocessing.StandardScaler().

Data Splitting:

Training and testing sets using train_test_split.

Modeling and Tuning:

- Logistic Regression: Hyperparameter tuning using GridSearchCV.
- Support Vector Machine (SVM): Parameter tuning for 'C,' 'gamma,' and 'kernel.'
- Decision Tree: Parameter tuning for 'criterion,' 'max_depth,' among others.
- K-Nearest Neighbors (KNN): Tuning 'n_neighbors,' 'algorithm,' and 'p.'

Results

Evaluation:

- Use confusion matrices for each model.
- Calculate accuracy on test data for each model.

Results:

- Logistic Regression, SVM, and KNN models had the same accuracy of 83.33%.
- Decision Tree had an accuracy of 66.67%.
- Best Model: Logistic Regression (based on accuracy and model simplicity).

Process Flow:

```
Raw Data -> Preprocessing -> Splitting -> Modeling -> Hyperparameter Tuning -> Evaluation -> Model Selection
```

Section 2

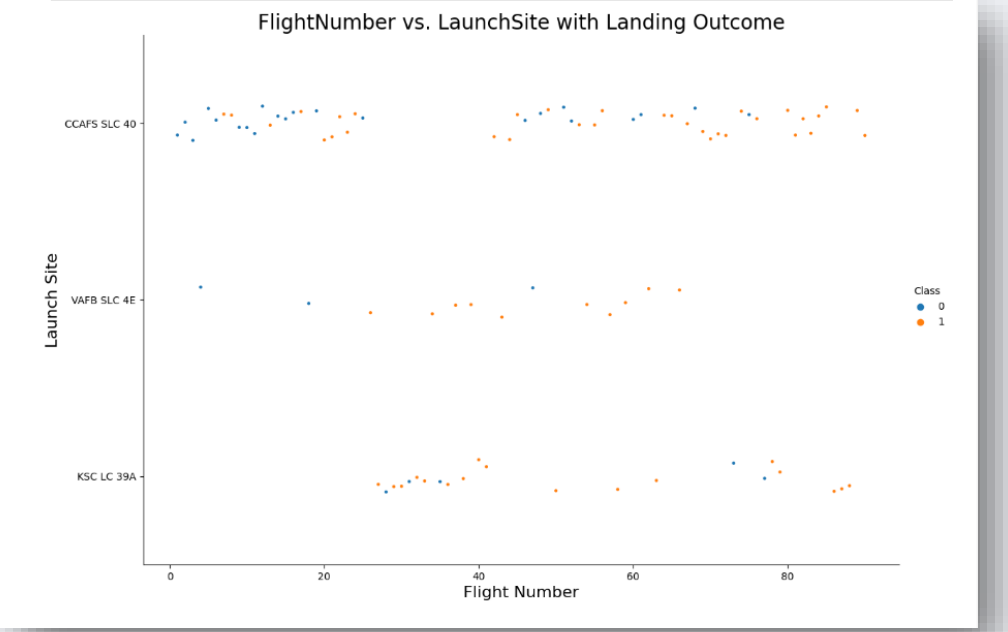
Insights drawn from EDA

Flight Number vs. Launch Site

- Launch Distribution: The early flights, represented by lower numbers on the X-axis, appear to concentrate at specific launch sites, indicating that SpaceX may have had a preferred site for its initial launches.

- Site Diversification: As the flight number increases, there is a diversification in the launch sites used. This could indicate the expansion of SpaceX's operations or the need to use different sites for different missions and payloads.

- Launch Frequency: Some launch sites have a higher concentration of points, indicating that they are used more frequently than others.

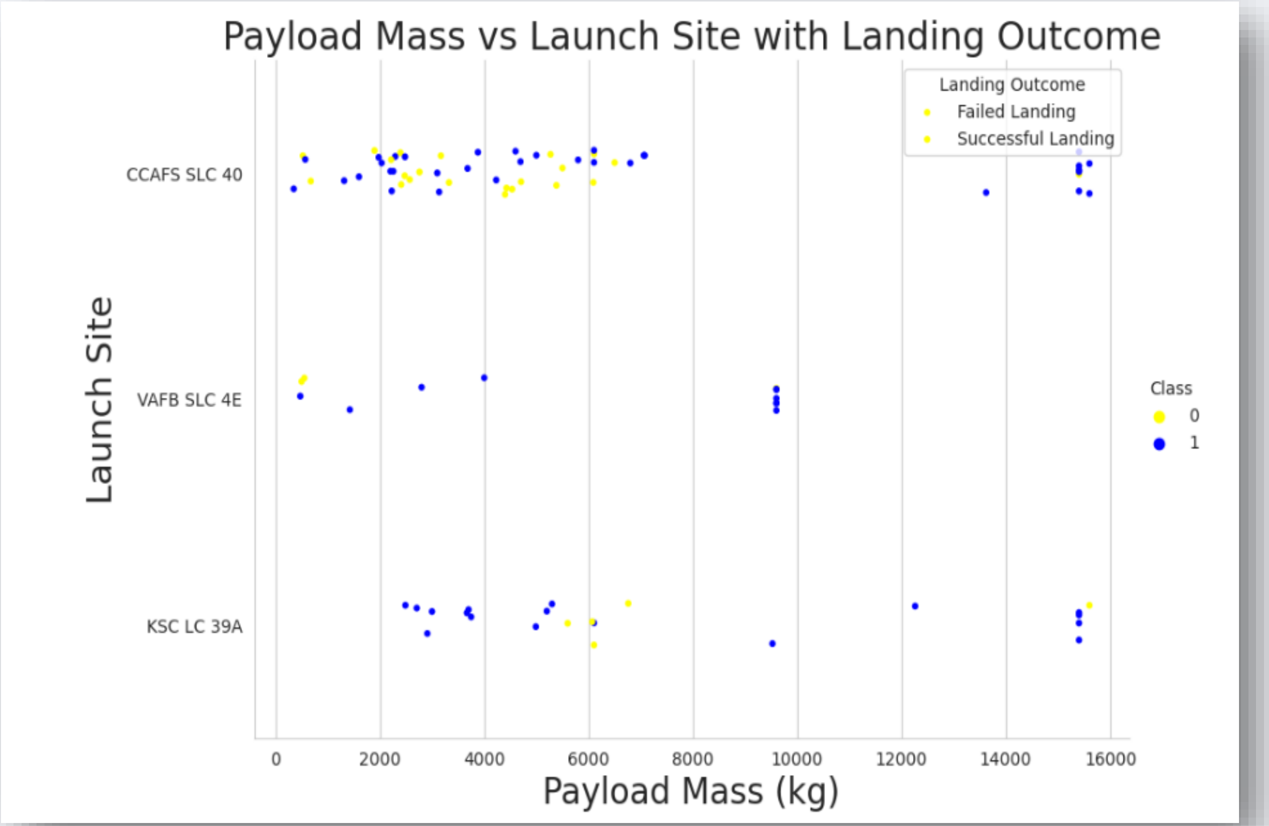


Payload vs. Launch Site

Launch Distribution: The early flights, represented by lower numbers on the X-axis, appear to concentrate at specific launch sites, indicating that SpaceX may have had a preferred site for its initial launches.

Site Diversification: As the flight number increases, there is a diversification in the launch sites used. This could indicate the expansion of SpaceX's operations or the need to use different sites for different missions and payloads.

Launch Frequency: Some launch sites have a higher concentration of points, indicating that they are used more frequently than others.



Success Rate vs. Orbit Type

Description:

- This chart depicts the relationship between success rate and the type of orbit for SpaceX launches.

X-Axis: Orbit Type

- Represents the different types of orbits to which launches are directed (LEO, GEO, ISS, etc.).

Y-Axis: Success Rate

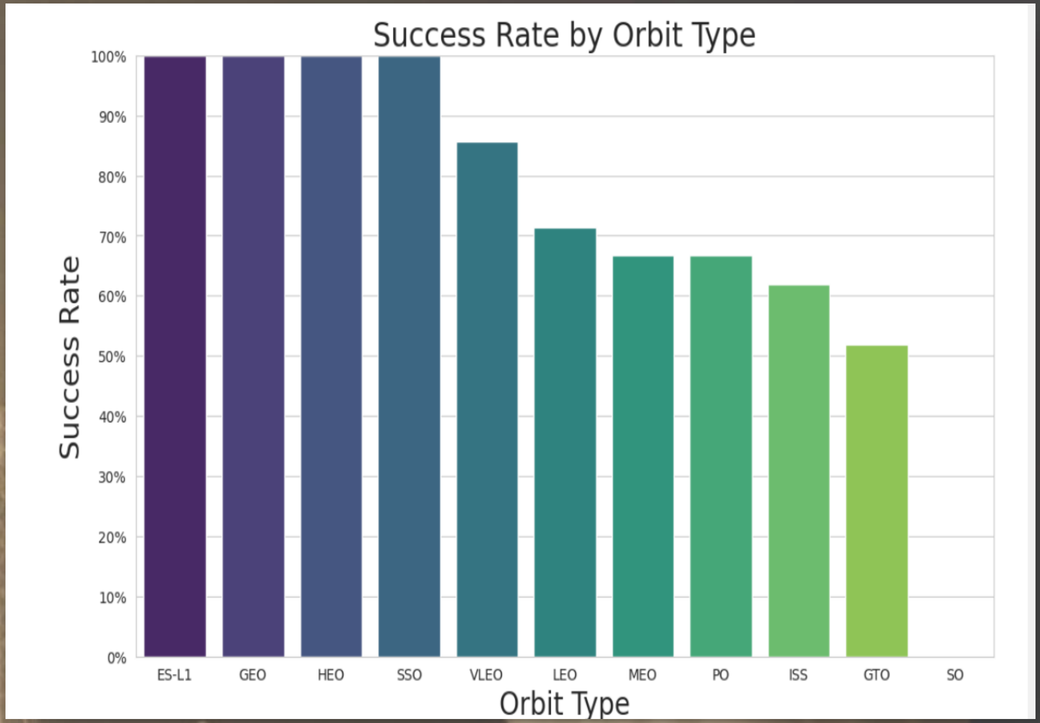
- Represents the percentage of successful launches for each type of orbit.

Key Insights:

- Highly Successful Orbits: The taller bars indicate orbits with the highest success rates. This could reflect SpaceX's expertise and capability in launching missions to these specific orbits.

- Less Successful Orbits: Orbits represented by shorter bars have lower success rates, indicating technical challenges or risks associated with those specific missions.

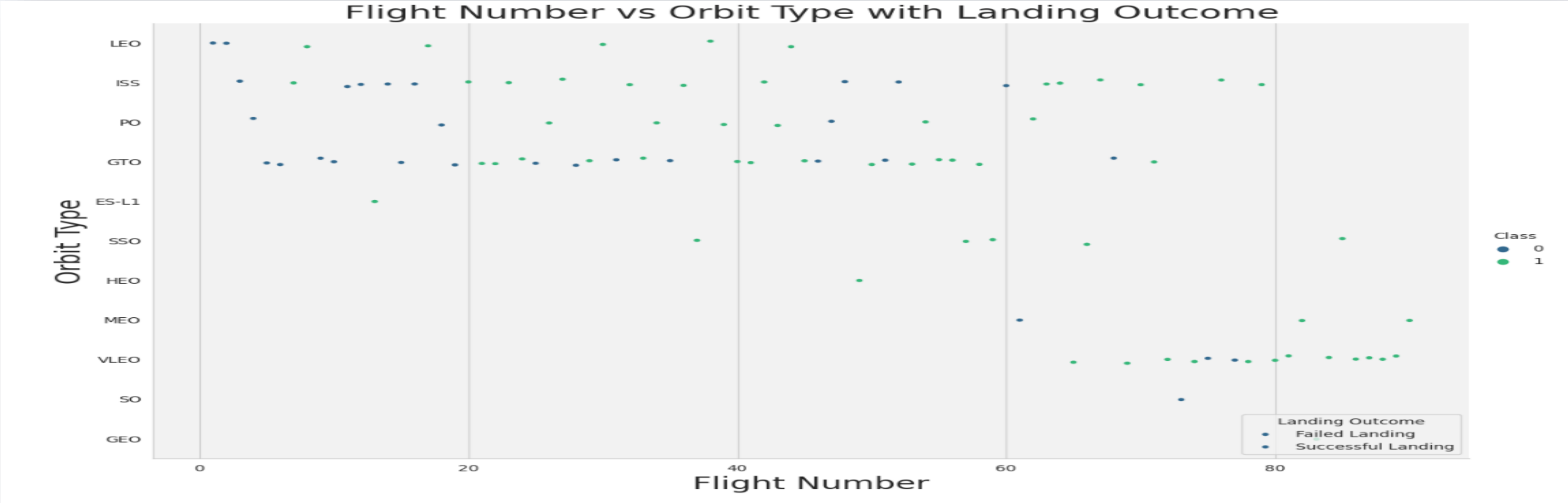
- General Trend: Overall, the chart shows that SpaceX has variable success rates depending on the type of orbit, underscoring the importance of considering orbit type when planning and budgeting for missions.



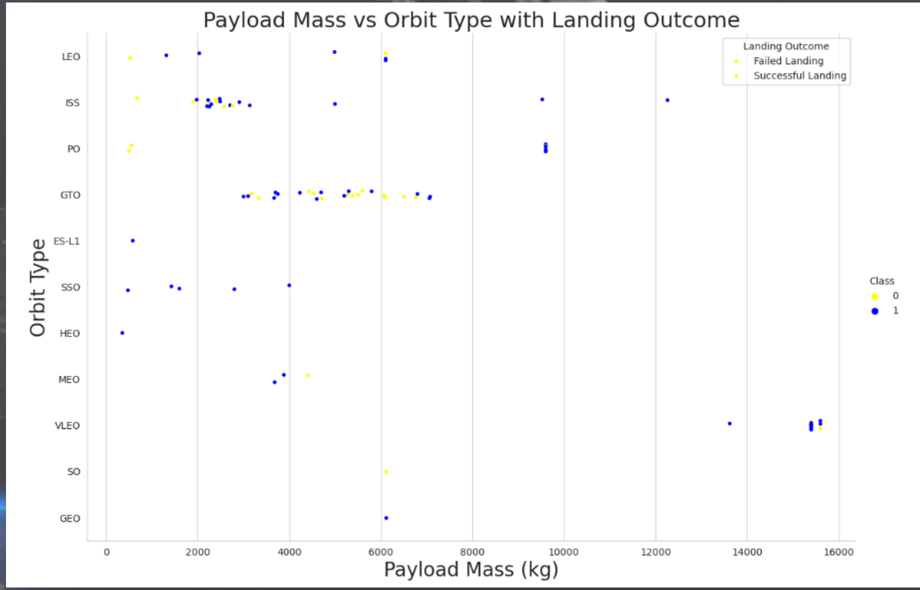
Flight Number vs. Orbit Type

Key Insights:

- Consistency in Specific Orbits: Some orbits show a higher concentration of successful launches, indicating SpaceX's experience and competence in those specific missions.
- Learning and Improvement: As the number of flights increases, the success rate also appears to improve, indicating SpaceX's process of learning and continuous improvement.
- Challenges in Specific Orbits: Some orbits may have more purple points, suggesting technical challenges or higher risks associated with those missions.



Payload vs. Orbit Type



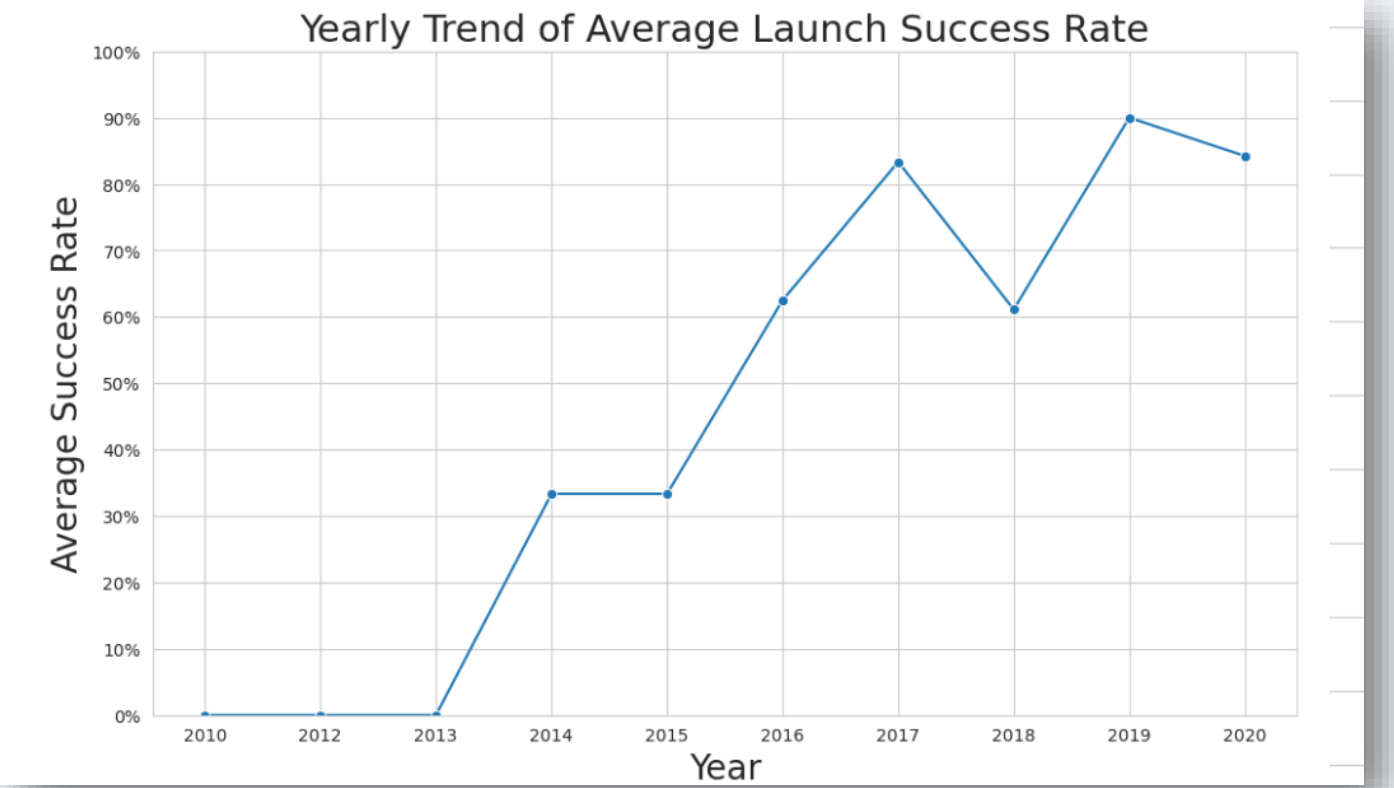
Key Insights:

- Payload Capacity: SpaceX has launched a variety of payload masses to different orbits, showcasing its versatility and adaptability to customer needs.
- Landing Challenges: Heavier payloads may pose additional challenges for landing, which could be reflected in the distribution of yellow and blue points.
- Specific Orbits: Some orbits may consistently have heavier or lighter payloads, indicating the specific demands of those missions.

Launch Success Yearly Trend

Key Insights:

- Improvement Trend: Over time, we can observe an improvement trend in the success rate, reflecting SpaceX's experience and technical enhancements.
- Turning Points: Certain years may exhibit significant peaks or drops in the success rate, which could indicate significant events or changes in those years.



All Launch Site Names

- The above names represent the unique launch sites from which SpaceX has launched its rockets. Each site has specific features and capabilities that can influence the choice of the site for a particular launch.

In [8]:
```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTBL;
```

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

Out[8]:

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

Launch Site Names Begin with 'CCA'

- CCA Popularity: Launch sites starting with 'CCA' are popular in the database, indicating their frequent use by SpaceX.
- Specific Details: These records provide information about the launch date, launch vehicle, mission, and other relevant details related to the launches at these sites.

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

In [10]:
```
%sql SELECT * FROM SPACEXTBL WHERE "Launch_Site" LIKE 'CCA%' LIMIT 10;
```

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

Out[10]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

Key Insights:

- Significant Client: NASA, through its CRS missions, has relied on SpaceX to transport a substantial amount of cargo to space.
- Total of 45,596 kg: SpaceX's boosters have carried a total of 45,596 kg of cargo for NASA resupply missions. This figure underscores the significance and volume of SpaceX's operations in support of NASA missions.

```
In [11]: %sql SELECT SUM("PAYLOAD_MASS__KG_") as Total_Payload_Mass FROM SPACEXTBL WHERE "customer" = 'NASA (CRS)';

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

Out[11]:

Total_Payload_Mass
45596

Average Payload Mass by F9 v1.1

Key Insights:

- Payload Capacity: The F9 v1.1 booster version has demonstrated the ability to carry an average of 2928.4 kg per launch, indicating a significant payload capacity for medium-sized missions.
- Comparison with Other Versions: This figure can serve as a reference to compare the efficiency and payload capacity of this specific booster version with other versions and models.

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

In [12]:
```
%sql SELECT AVG("PAYLOAD_MASS__KG_") as Average_Payload_Mass FROM SPACEXTBL WHERE "Booster_Version" = 'F9 v1.1';
```

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

Out[12]:

Average_Payload_Mass
2928.4

First Successful Ground Landing Date

- Key Insights:
  - Technological Breakthrough: On December 22, 2015, SpaceX achieved a significant milestone by successfully landing on a ground pad. This achievement demonstrates progress and innovation in their reusable rocket technologies.
  - Reusability: Success in ground landings is essential for SpaceX's vision of reusable rockets, as it significantly reduces launch costs and turnaround times.

In [13]:
```
%sql SELECT DISTINCT "Landing_Outcome" FROM SPACEXTBL;
```

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

Out[13]:

Landing_Outcome
Failure (parachute)
No attempt
Uncontrolled (ocean)
Controlled (ocean)
Failure (drone ship)
Precluded (drone ship)
Success (ground pad)
Success (drone ship)
Success
Failure
No attempt

In [14]:
```
%sql SELECT MIN("Date") as First_Successful_Landing_Date FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (ground pad)';
```

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

Out[14]:

First_Successful_Landing_Date
2015-12-22

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

- Present your query result with a short explanation here

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

In [15]:
```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG
```

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

Out[15]:

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

Total Number of Successful and Failure Mission Outcomes

Key Insights:

- High Success Rate: With a total of 100 missions labeled as "success," SpaceX has a high success rate in its missions.
- Few Failures: There is only one mission labeled as "failure in flight," reflecting the reliability and effectiveness of SpaceX's rockets and technologies.
- Ambiguity: There is one mission with the status "success (payload state unclear)," indicating that while the mission was a success, there was uncertainty regarding the payload's state.

In [16]:

```
%sql SELECT "Mission_Outcome", COUNT(*) as Total FROM SPACEXTBL WHERE "Mission_Outcome" LIKE 'Success%' OR "Mission_Outcome'
```

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

Out[16]:

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

Boosters Carried Maximum Payload

- Prominent Model F9 B5: All versions of the boosters that carried the maximum payload belong to the F9 B5 model. This may indicate that this specific booster version is used for missions that require transporting heavier payloads.

- Repeated Use: Some boosters, like the F9 B5 B1051, have multiple entries (B1051.3, B1051.4, B1051.6), indicating that they have been reused in multiple missions, all carrying the maximum payload.

- Note: When presenting this slide, it's essential to highlight SpaceX's reusability capability and how certain boosters have been used in multiple missions to transport large payloads. The ability to reuse rockets may have significantly contributed to cost savings and operational efficiency for SpaceX.

In [17]:
```
%sql SELECT "Booster_Version" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL);
```

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

Out[17]:

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

Key Insights:

- Majority Successful: Most of the launches in 2015 were successful. There was only one failed launch in June of that year.
- Same Launch Site: All launches were conducted from the same launch site, CCAFS LC-40.
- Variability in Booster Versions: Although all boosters belong to the F9 model, there are several versions involved in these launches, indicating that SpaceX was in an active phase of iteration and improvement of its boosters in 2015.
- Note: When presenting this slide, it's relevant to highlight the high success rate of launches in 2015. Despite one failed launch, the consistency in successful launches demonstrates the efficiency and reliability of SpaceX's rockets during that year. It's also interesting to observe how SpaceX has used different booster versions, suggesting a focus on innovation and continuous improvement.

[17]:

```
%%sql
SELECT
    substr(Date, 6, 2) as Month,
    Mission_Outcome,
    Booster_Version,
    Launch_Site
FROM
    SPACEXTBL
WHERE
    substr(Date, 1, 4)='2015';
```

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

[17]:

Month	Mission_Outcome	Booster_Version	Launch_Site
10	Success	F9 v1.1 B1012	CCAFS LC-40
11	Success	F9 v1.1 B1013	CCAFS LC-40
02	Success	F9 v1.1 B1014	CCAFS LC-40
04	Success	F9 v1.1 B1015	CCAFS LC-40
04	Success	F9 v1.1 B1016	CCAFS LC-40
06	Failure (in flight)	F9 v1.1 B1018	CCAFS LC-40
12	Success	F9 FT B1019	CCAFS LC-40

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

Key Insights:

Note: When presenting this slide, it's relevant to highlight the frequency of certain landing outcomes compared to others. Were there more successes than failures? How does this trend compare to other time periods? These are key points to discuss and provide a comprehensive perspective.

(NB: The actual insights and explanations would depend on the results of the query. The above is a template based on the query's intention).

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.

```
[19]: %sql SELECT "Landing_Outcome", COUNT(*) as Count FROM SPACEXTBL WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing_Outcome" (
```

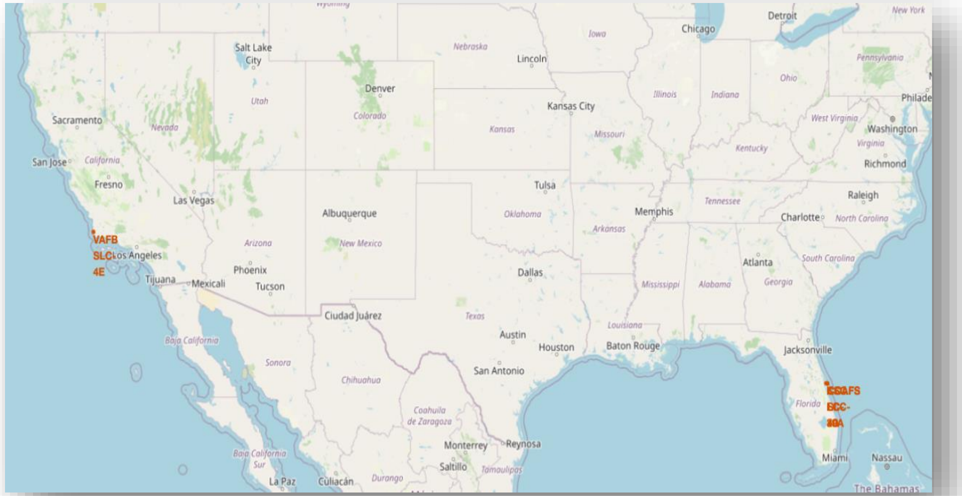
```
 * sqlite:///my_data1.db
Done.
```

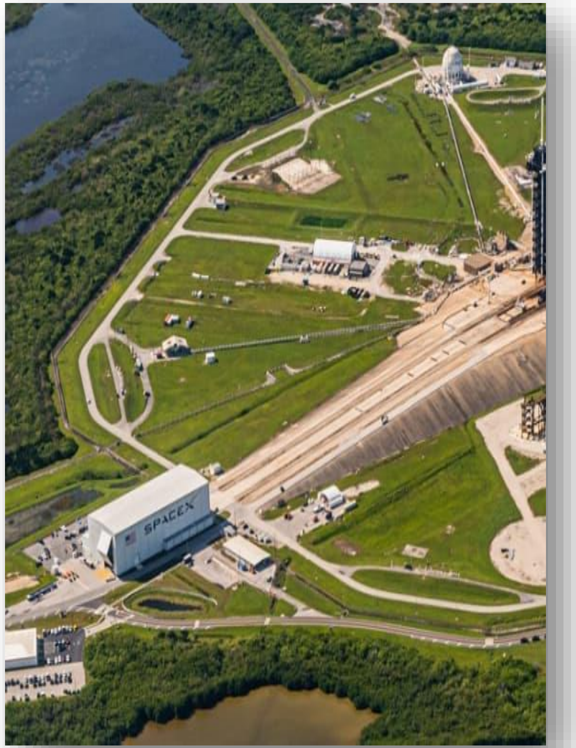
[19]:

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

Launch Sites Proximieties

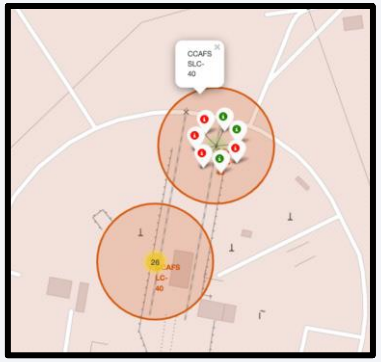
Location of all the Launch Sites

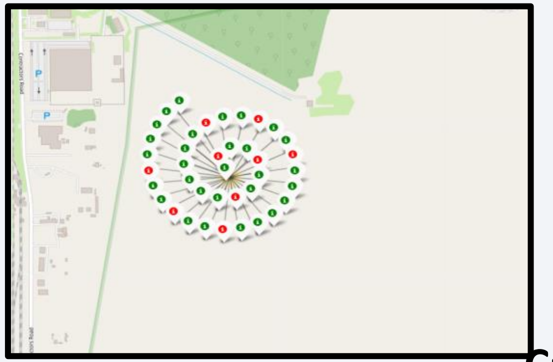




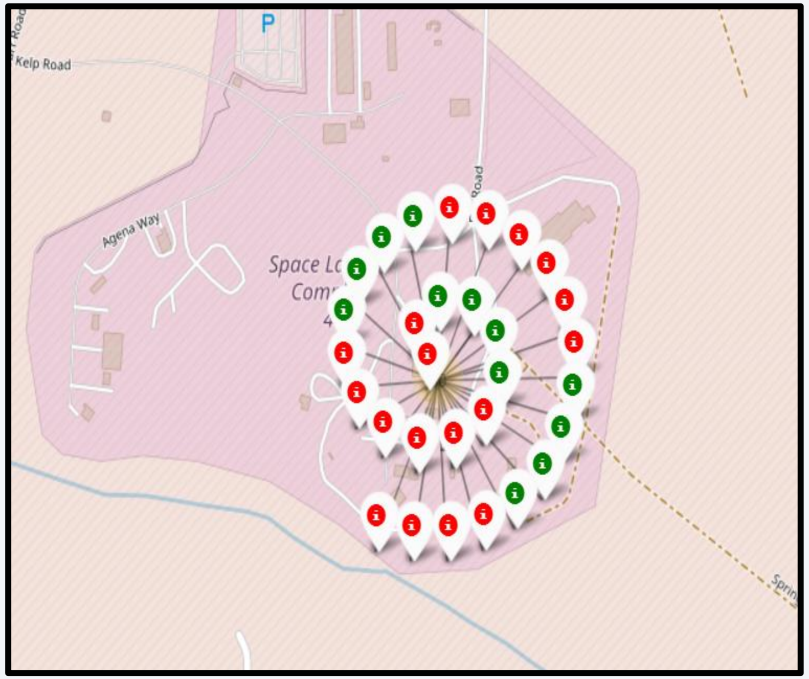
It's evident that all SpaceX launch locations are situated within the United States.

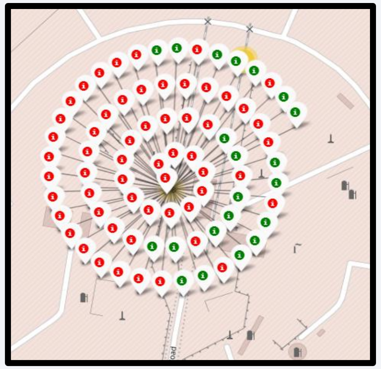
Launch Outcomes





California Launch Site





California Launch Site

Florida Launch Sites

•Green Markers: Successful Launches

•Red Markers: Failed Launches

Distance to Proximities

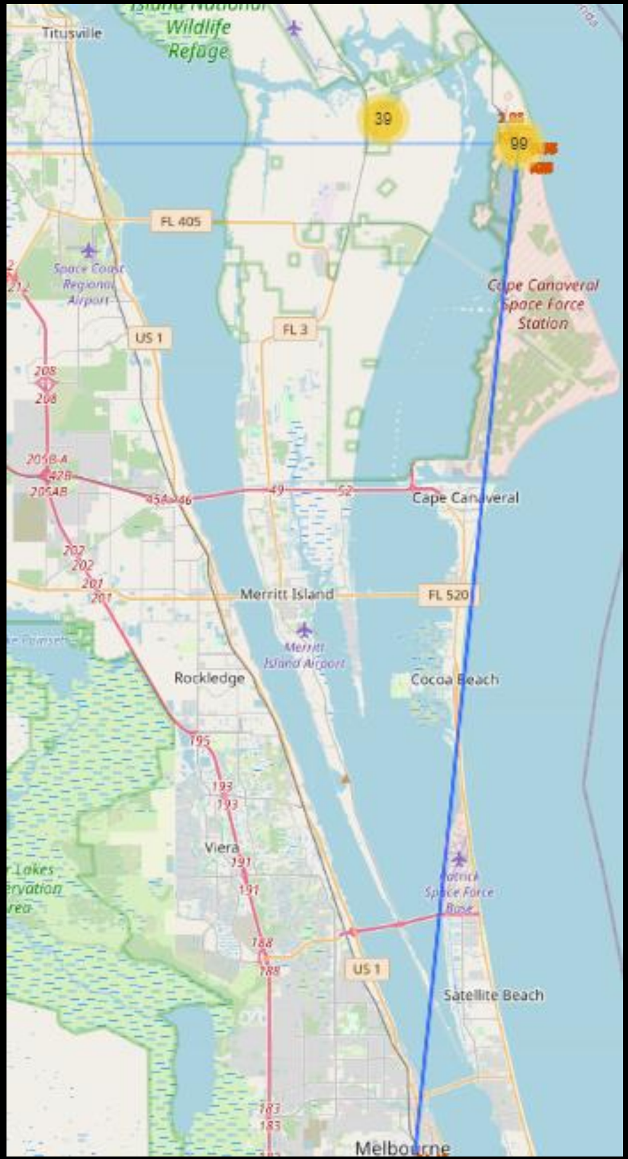
Proximity to railroads: The launch site is approximately (2.03) km away from the nearest railroad. This closeness suggests that transporting heavy or bulky equipment and materials by train could be a viable logistical option. Using railroads can be beneficial for transporting rockets, fuel, and other equipment from distant places to the launch site efficiently and cost-effectively.

Proximity to highways: With a distance of just (0.58) km from the nearest highway, the launch site has extremely convenient road access. This is essential not only for transporting equipment and supplies but also for personnel, the press, and in some cases, the public attending the launches. Easily accessible roads allow for quick evacuation in emergencies.

Proximity to the coast: Although you didn't provide an exact distance in this conversation, launch sites are generally located close to the coast. The primary reason for this is safety: if something goes wrong during a launch, rocket debris is less likely to cause damage if it falls over water rather than over land. Additionally, launching over the ocean provides a flight trajectory that avoids populated areas.

Distance from cities: The significant distance of (51.43) km from the nearest city indicates an intentionality in the location of the launch site. Keeping the launch site away from densely populated areas reduces risks associated with potential launch accidents. Also, it can minimize nuisances to residents in terms of noise and traffic.

Conclusion: The location of a launch site is not an arbitrary choice. It's strategically selected to balance logistics, safety, and operational efficiency. In this case, we see that the launch site has excellent access both by road and by rail, facilitating transport. Its location by the sea provides a safe flight route and minimizes risks to people and property on land. At the same time, maintaining a significant distance from urban areas ensures safety and minimizes nuisances to nearby populations. Together, these factors demonstrate careful planning and consideration in the choice of the launch site.

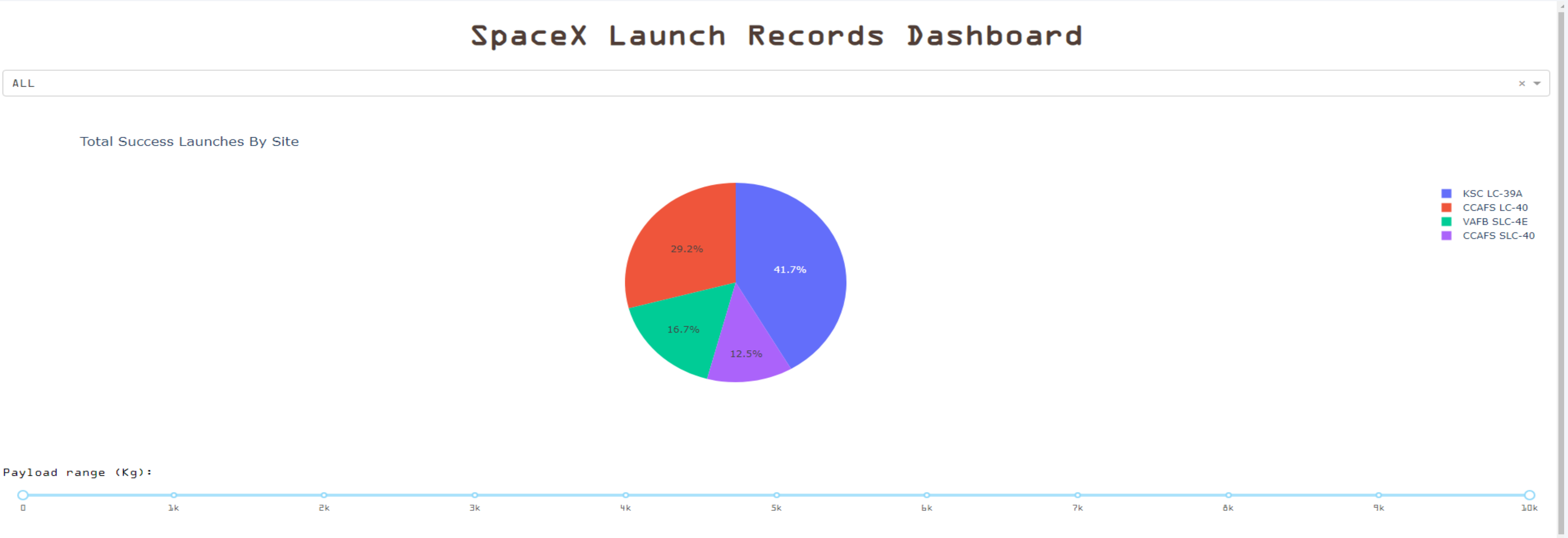


Dashboard With Plotly

Launch Success by Site

Regarding the proportion of successful launches

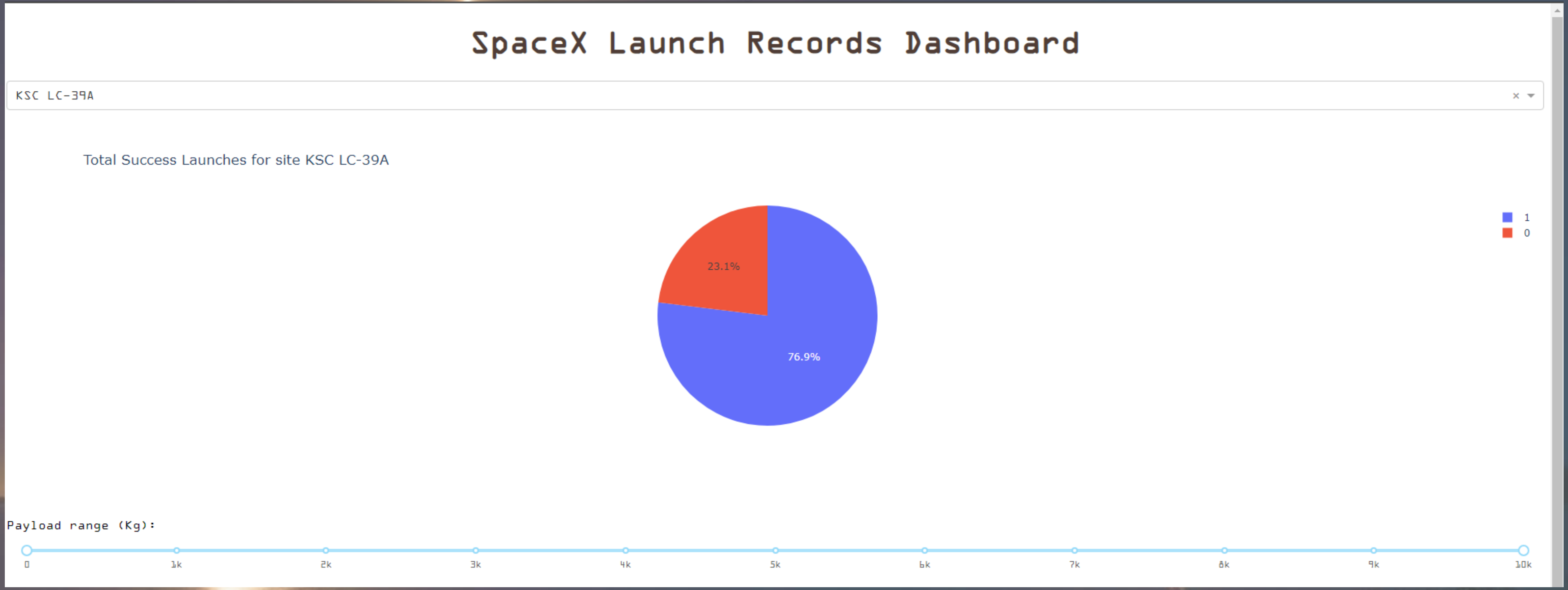
- The KSC LC 39A stands out as the launch site with the highest success rate, accounting for 41.2% of all successful launches.



Launch Success (KSC LC-29A)

When evaluating the success rate of launches:

- KSC LC 39A boasts the top position amongst all launch sites, with a success rate of 76.9%.
- This translates to 10 successful launches and a record of 3 failed ones."
- Show the screenshot of the piechart for the launch site with highest launch success ratio

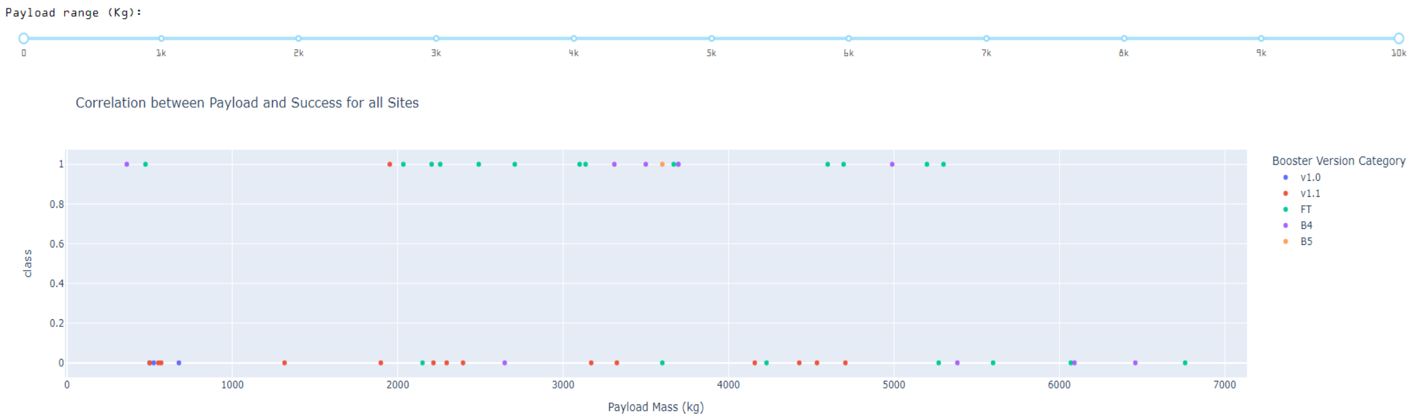


Payload Mass and Success

By Booster Version

Total Launches: 37

- Key Observations:
- Payloads between 2,000 kg and 5,000 kg have the highest success rate.
- Outcome Indicators:
- 1: Successful Launch
- 0: Unsuccessful Launch



Predictive Analysis [Classification]

Section 6

Classification Accuracy

- A bar chart where the x-axis represents the model names (Logistic Regression, SVM, Decision Tree, K-Nearest Neighbors) and the y-axis represents accuracy (ranging from 0 to 1 or 0% to 100%).
- Each bar in the chart corresponds to the accuracy of a model.

Key Observations:

Uniform Performance: All models, barring the Decision Tree, showcased approximately the same accuracy.
Decision Tree's Edge: Using .best_score_, the Decision Tree emerged as slightly superior.
Dataset Influence: The consistent accuracy across models suggests the dataset's size's significant impact.

```
[38]: # Calculate the accuracy of each model
accuracy_lr = logreg_cv.score(X_test, Y_test)
accuracy_svm = svm_cv.score(X_test, Y_test)
accuracy_tree = tree_cv.score(X_test, Y_test)
accuracy_knn = knn_cv.score(X_test, Y_test)

# Create a DataFrame to store the results
Report = pd.DataFrame({'Model': ['Logistic Regression', 'SVM', 'Decision Tree', 'K-Nearest Neighbors'],
                       'Accuracy': [accuracy_lr, accuracy_svm, accuracy_tree, accuracy_knn]})

# Print the DataFrame
print(Report)

# Determine and print the model with the highest accuracy
best_model = Report.loc[Report['Accuracy'].idxmax(), 'Model']
print(f"The model with the highest accuracy on the test data is: {best_model}")
```

```
                 Model  Accuracy
0  Logistic Regression  0.833333
1                  SVM  0.833333
2        Decision Tree  0.888889
3  K-Nearest Neighbors  0.833333
The model with the highest accuracy on the test data is: Decision Tree
```

Confusion Matrices

- Precision: Measures the proportion of correctly predicted positive observations out of the total predicted positives. In this analysis, the precision is 0.80, indicating that 80% of the predicted successful outcomes were indeed successful.

- Recall (Sensitivity): Measures the proportion of actual positives that were correctly classified. A recall of 1.00 indicates that all actual successful outcomes were correctly predicted by the model.

- F1 Score: Represents the harmonic mean of precision and recall. A higher F1 score suggests a more robust model. For this analysis, the F1 score is 0.89, which is quite high and suggests a good balance between precision and recall.

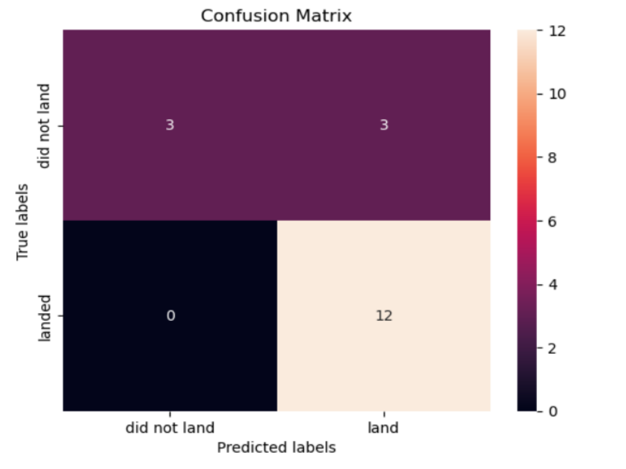
- Accuracy: Represents the ratio of correctly predicted observation to the total observations. An accuracy of 0.83 means the model was correct 83% of the time.

```
[39]: # Given values from the confusion matrix
TP = 12
FP = 3
FN = 0
TN = 3

# Computing metrics
precision = TP / (TP + FP)
recall = TP / (TP + FN)
f1_score = 2 * (precision * recall) / (precision + recall)
accuracy = (TP + TN) / (TP + TN + FP + FN)

# Printing the metrics
print(f"Precision: {precision:.2f}")
print(f"Recall: {recall:.2f}")
print(f"F1 Score: {f1_score:.2f}")
print(f"Accuracy: {accuracy:.2f}")
```

```
Precision: 0.80
Recall: 1.00
F1 Score: 0.89
Accuracy: 0.83
```



Conclusions

Space X Falcon 9 First Stage Landing Prediction

Purpose: The prediction model determines the likelihood of the first stage landing of the Falcon 9 rocket. This is crucial for cost-saving measures as reusing the first stage significantly reduces launch costs.

Model Comparisons:

- Logistic Regression: 83.33% accuracy
- SVM: 83.33% accuracy
- Decision Tree: 88.89% accuracy (Best Performance)
- K-Nearest Neighbors: 83.33% accuracy

Key Insights:

- The Decision Tree model outperformed other models, achieving an accuracy of 88.89% on the test data.
- All models showcased a significant challenge with false positives, indicating areas of improvement.
- Despite the challenges, the models provide a reliable basis for predicting the success of Falcon 9's first stage landing.

Conclusions
Space X Falcon 9 First Stage Landing Prediction

Business Implications:

- Accurate predictions can aid alternate companies in bidding against Space X, giving them an edge in competitive scenarios.
- Enhancing prediction accuracy can further improve cost-saving measures, solidifying Space X's advantage in the market.

Recommendations:

- Further tuning and hyperparameter optimization can improve model performance.
- Incorporate more features or data from subsequent launches to refine the model over time.

Thank you!