IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

YASH MAHALE
18-August-2023

Outline

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

Executive Summary

Summary of methodologies

- Data collection
- Web scrapping
- Data wrangling
- Exploratory data analysis with SQL
- Exploratory data analysis with data visualization
- Building a dashboard with plotly dash
- Machine learning algorithm

Summary of all results

- Exploratory data analysis results
- Interactive analytics image screenshots
- Predictive analysis results

Introduction

Project background and context

SpaceX stands as the premier achiever in the era of commercial space exploration, significantly reducing the financial barriers to space travel. On its website, the company promotes Falcon 9 rocket launches, which come at a price of 62 million dollars. In contrast, other providers demand a substantially higher cost of around 165 million dollars for each launch. A major factor behind these remarkable savings lies in SpaceX's ability to recycle and reuse the first stage of their rockets. By accurately predicting the successful landing of the first stage, we can determine the overall cost of a launch. Drawing insights from publicly available information and employing advanced machine learning models, our endeavor is to forecast whether SpaceX will effectively recycle the first stage, thus unveiling the cost-efficiency of a launch.

Problems you want to find answers

- How do variables such as payload mass, launch site, number of flights, and orbits affect the success of the first stage landing?
- Does the rate of successful landings increase over the years?
- What is the best algorithm that can be used for binary classification in this case?

Section 1

Methodology

Methodology

Data collection methodology:

- Using spaceX rest API
- Using web scrapping from Wikipedia

Perform data wrangling

- Filtering the data
- Dealing with missing values
- Using One Hot Encoding to prepare the data to a binary classification

Perform exploratory data analysis (EDA) using visualization and SQL

Perform interactive visual analytics using Folium and Plotly Dash

Perform predictive analysis using classification models

- Building, tuning and evaluation of classification models to ensure the best results

Data Collection

Data collection process involved a combination of API requests from SpaceX REST API and Web Scraping data from a table in SpaceX's Wikipedia entry.

We had to use both of these data collection methods in order to get complete information about the launches for a more detailed analysis.

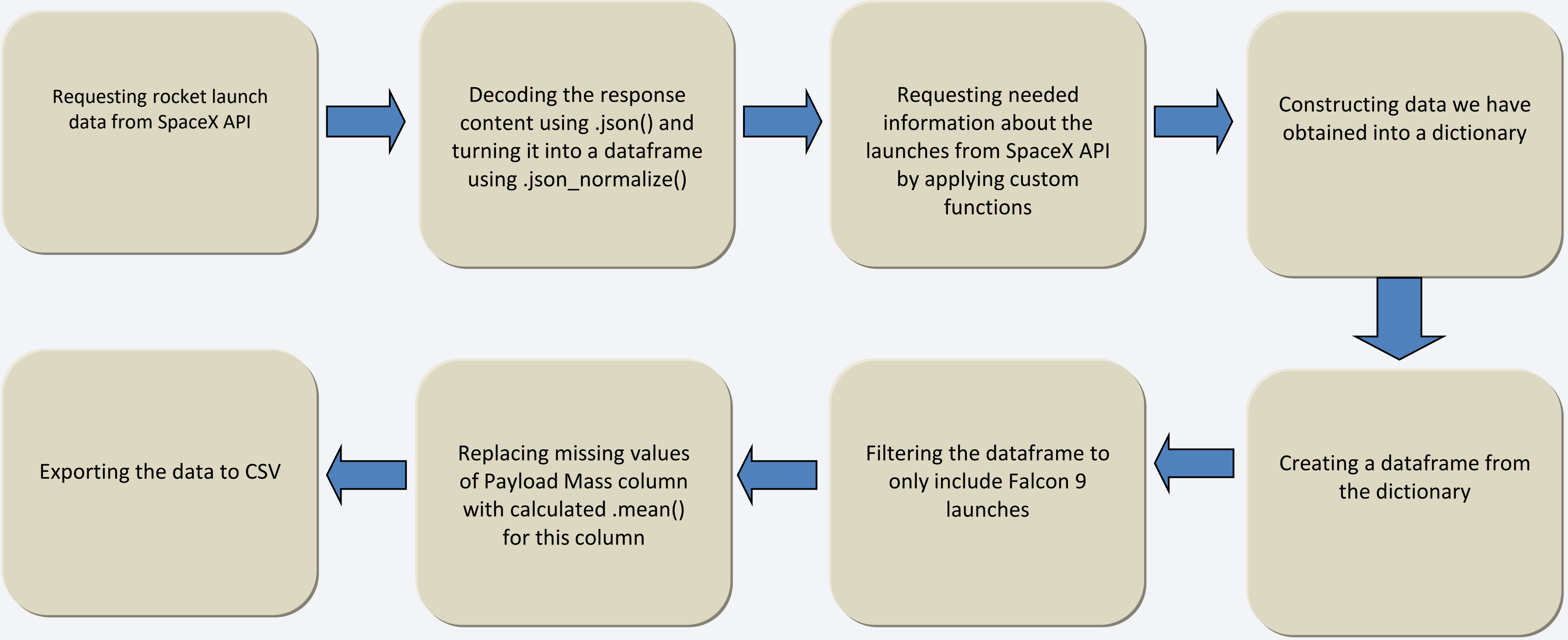
Data Columns are obtained by using SpaceX REST API:

FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude

Data Columns are obtained by using Wikipedia Web Scraping:

Flight No., Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time

Data Collection – SpaceX API



Requesting rocket launch data from SpaceX API

Decoding the response content using .json() and turning it into a dataframe using .json_normalize()

Requesting needed information about the launches from SpaceX API by applying custom functions

Constructing data we have obtained into a dictionary

Creating a dataframe from the dictionary

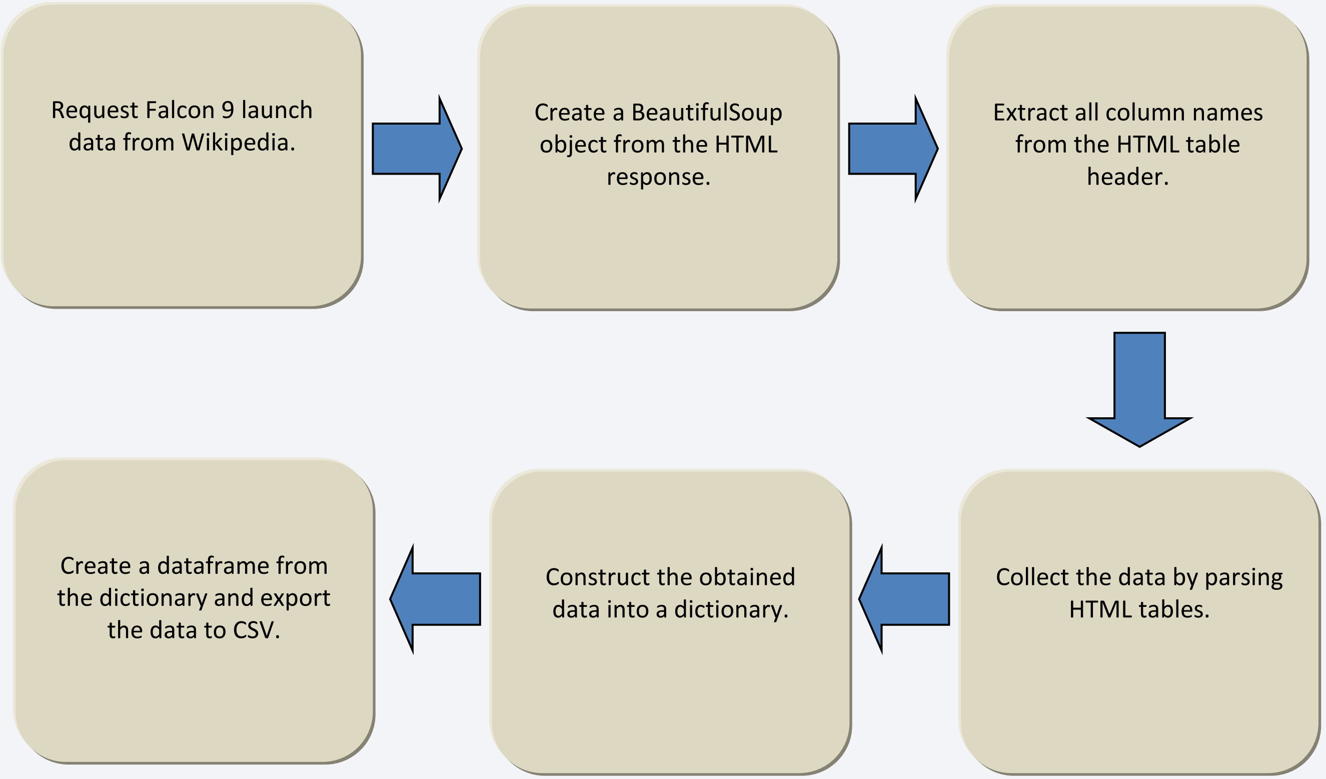
Filtering the dataframe to only include Falcon 9 launches

Replacing missing values of Payload Mass column with calculated .mean() for this column

Exporting the data to CSV

https://github.com/YASH-MAHALE-07/YASH_IBM_CAPSTONE_PROJECT/blob/main/Data%20collection%20api.ipynb

Data Collection - Scraping



Request Falcon 9 launch data from Wikipedia.

Create a BeautifulSoup object from the HTML response.

Extract all column names from the HTML table header.

Collect the data by parsing HTML tables.

Construct the obtained data into a dictionary.

Create a dataframe from the dictionary and export the data to CSV.

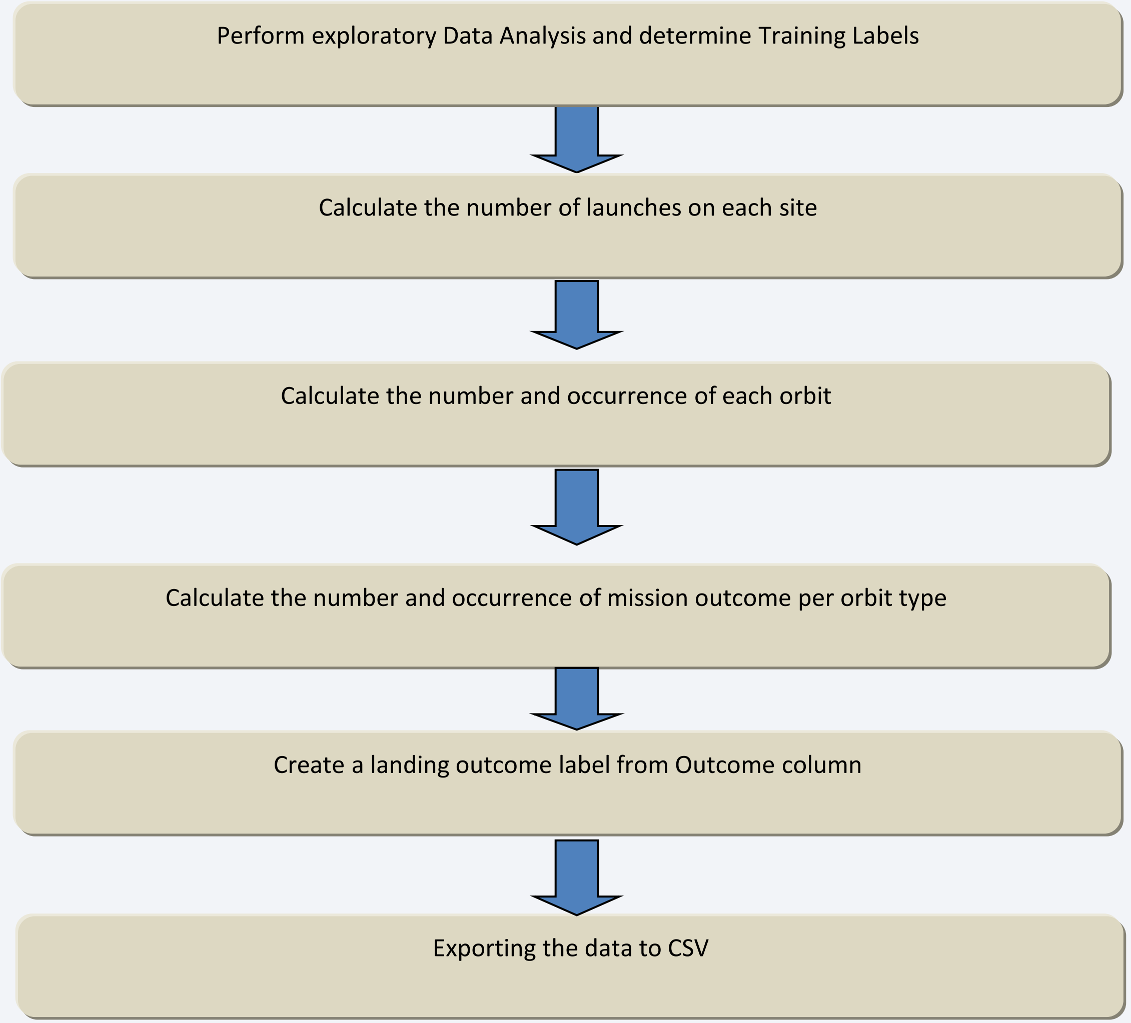
https://github.com/YASH-MAHALE-07/YASH_IBM_CAPSTONE_PROJECT/blob/main/Webscraping.ipynb

Data Wrangling

In the dataset, various scenarios exist where the booster's landing was not successful. At times, a landing attempt was made but failed due to an accident. For instance, "True Ocean" indicates that the mission outcome involved a successful landing in a specific ocean region, while "False Ocean" signifies an unsuccessful landing in a specific ocean region.

Similarly, "True RTLS" indicates a successful ground pad landing, while "False RTLS" indicates an unsuccessful ground pad landing. "True ASDS" denotes a successful landing on a drone ship, whereas "False ASDS" signifies an unsuccessful landing on a drone ship.

To streamline these outcomes, we converted them into training labels, wherein "1" signifies a successful booster landing, and "0" indicates an unsuccessful landing.



Perform exploratory Data Analysis and determine Training Labels

Calculate the number of launches on each site

Calculate the number and occurrence of each orbit

Calculate the number and occurrence of mission outcome per orbit type

Create a landing outcome label from Outcome column

Exporting the data to CSV

https://github.com/YASH-MAHALE-07/YASH_IBM_CAPSTONE_PROJECT/blob/main/Data%20wrangling.ipynb

EDA with Data Visualization

Various charts were generated for analysis:

Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch Site, Orbit Type vs. Success Rate, Flight Number vs. Orbit Type, Payload Mass vs. Orbit Type, Yearly Trend of Success Rate

Scatter plots were utilized to illustrate the potential relationships between variables. If a correlation is observed, these relationships could be leveraged in a machine learning model.

Bar charts were employed to make comparisons across discrete categories. The aim was to showcase the connection between the specific categories being compared and a measurable value.

Line charts were used to visualize trends in data over time, particularly for time series analysis.

https://github.com/YASH-MAHALE-07/YASH_IBM_CAPSTONE_PROJECT/blob/main/EDA%20Data%20visualization.ipynb

EDA with SQL

Performed SQL queries:

- Displayed the names of the unique launch sites in the space mission
- Displayed 5 records where launch sites start with the string ‘CCA'
- Displayed the total payload mass carried by boosters launched by NASA (CRS)
- Displayed the average payload mass carried by booster version F9 v1.1
- Listed the date when the first successful landing outcome on a ground pad was achieved
- Listed the names of boosters that achieved success on a drone ship and had payload mass greater than 4000 but less than 6000
- Listed the total number of successful and failure mission outcomes
- Listed the names of booster versions that carried the maximum payload mass
- Listed the failed landing outcomes on a drone ship, their booster versions, and launch site names for the months in year 2015
- Ranked the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the dates 2010-06-04 and 2017-03-20 in descending order

https://github.com/YASH-MAHALE-07/YASH_IBM_CAPSTONE_PROJECT/blob/main/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

Markers of all Launch Sites:

- Added Markers with Circles, Popup Labels, and Text Labels of NASA Johnson Space Center using its latitude and longitude coordinates as a starting location.
- Added Markers with Circles, Popup Labels, and Text Labels of all Launch Sites using their latitude and longitude coordinates to display their geographical locations and proximity to the Equator and coasts.

Colored Markers of the launch outcomes for each Launch Site:

- Added colored Markers representing success (Green) and failed (Red) launches using a Marker Cluster to identify which launch sites have relatively high success rates.

Distances between a Launch Site and its proximities:

- Added colored Lines to illustrate the distances between the Launch Site KSC LC-39A (as an example) and its proximities such as Railway, Highway, Coastline, and the Closest City.

https://github.com/YASH-MAHALE-07/YASH_IBM_CAPSTONE_PROJECT/blob/main/Data%20Analysis%20on%20map.ipynb

Build a Dashboard with Plotly Dash

Launch Sites Dropdown List:

- Added a dropdown list to allow the user to select a Launch Site.

Pie Chart displaying Success Launches (All Sites/Specific Site):

- Included a pie chart to visualize the total count of successful launches for all sites and the breakdown of Success vs. Failed counts for a specific Launch Site if chosen.

Payload Mass Range Slider:

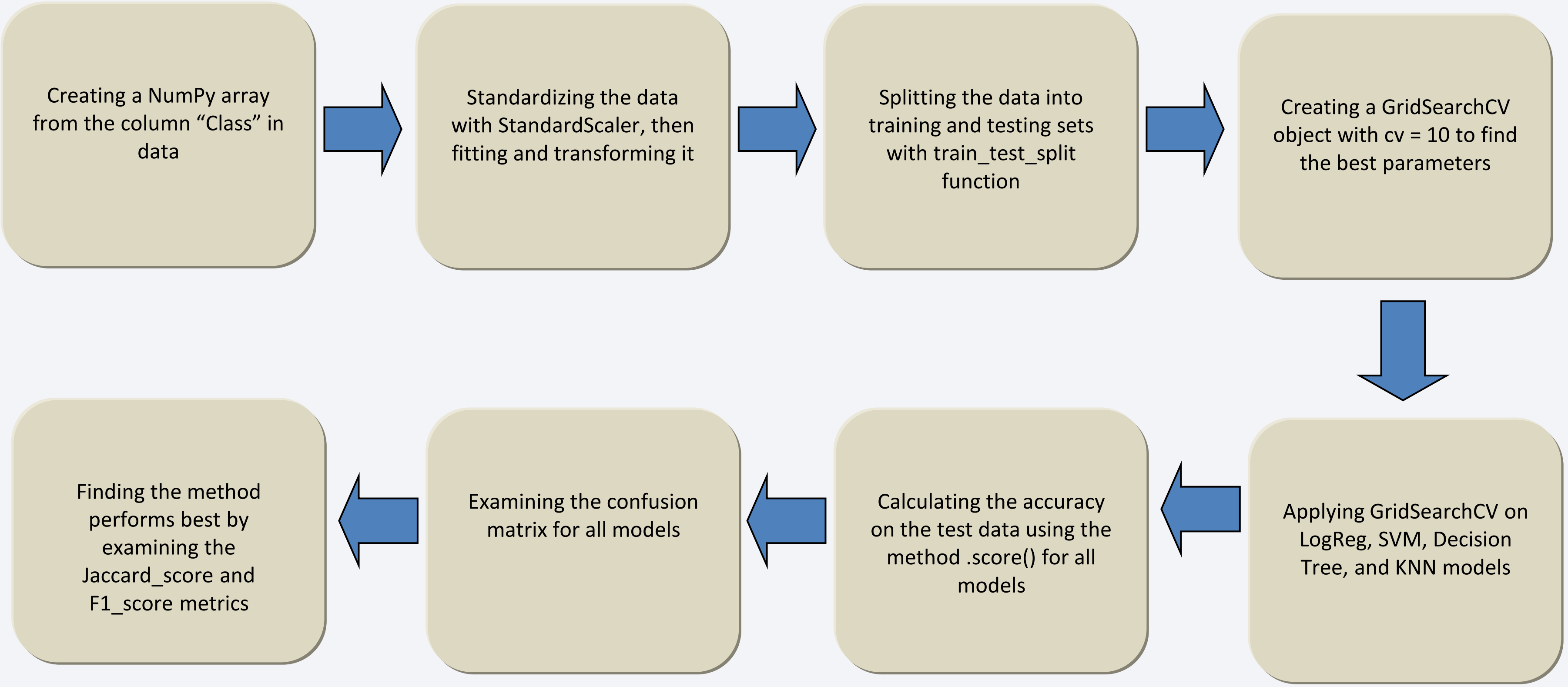
- Incorporated a slider to enable the user to select a range of Payload masses.

Scatter Chart illustrating Payload Mass vs. Success Rate for various Booster Versions:

- Created a scatter chart to depict the relationship between Payload Mass and Launch Success, differentiated by different Booster Versions.

https://github.com/YASH-MAHALE-07/YASH_IBM_CAPSTONE_PROJECT/blob/main/Spacex%20dash%20app.ipynb

Predictive Analysis (Classification)



Creating a NumPy array from the column “Class” in data

Standardizing the data with StandardScaler, then fitting and transforming it

Splitting the data into training and testing sets with train_test_split function

Creating a GridSearchCV object with cv = 10 to find the best parameters

Applying GridSearchCV on LogReg, SVM, Decision Tree, and KNN models

Calculating the accuracy on the test data using the method .score() for all models

Examining the confusion matrix for all models

Finding the method performs best by examining the Jaccard_score and F1_score metrics

https://github.com/YASH-MAHALE-07/YASH_IBM_CAPSTONE_PROJECT/blob/main/Machine%20Learning%20Prediction%20SpaceX.ipynb

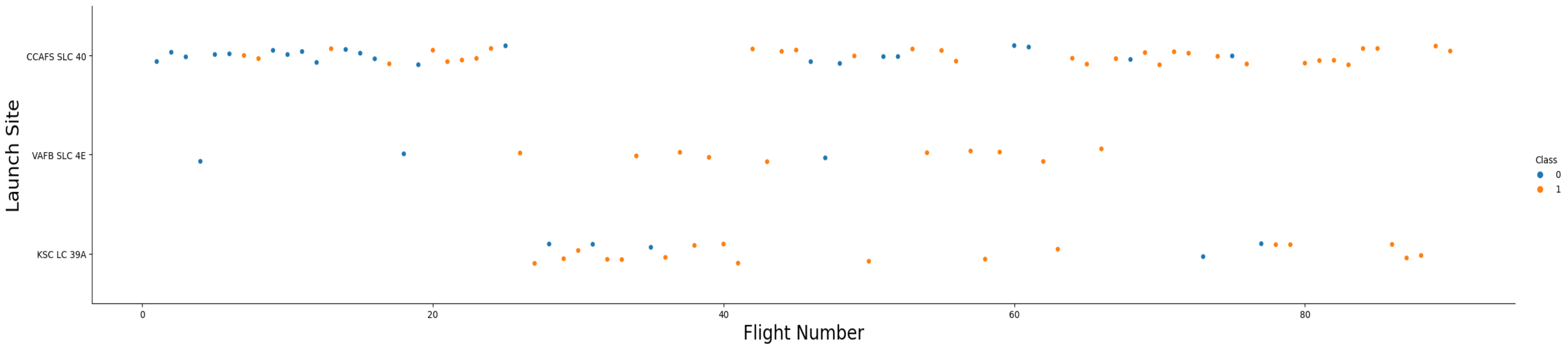
Results

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

Section 2

Insights drawn from EDA

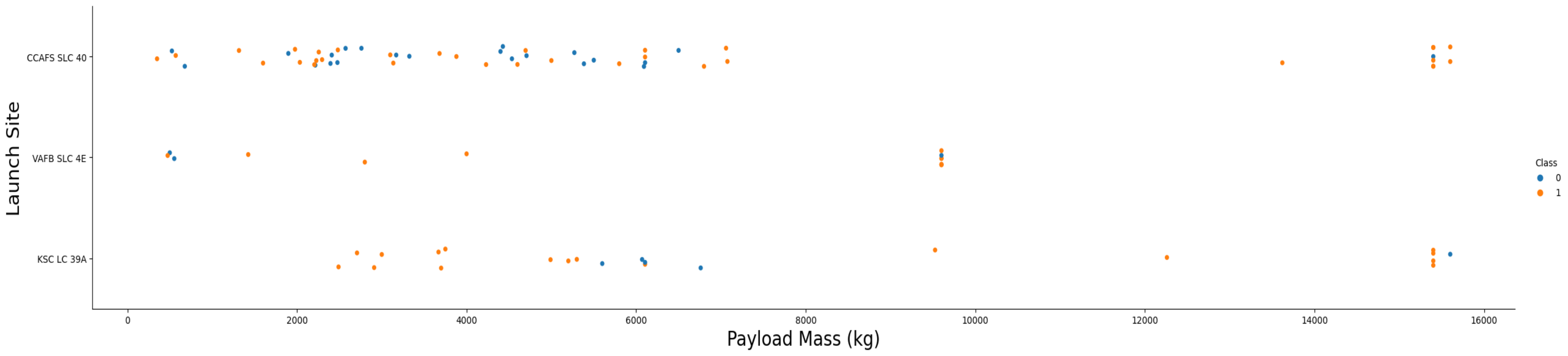
Flight Number vs. Launch Site



Explanation:

- Initial flights were unsuccessful, whereas recent flights have been successful.
- CCAFS SLC 40 launch site accounts for approximately half of all launches.
- VAFB SLC 4E and KSC LC 39A display higher success rates.
- It can be inferred that the success rate tends to increase with newer launches.

Payload vs. Launch Site



Explanation:

- There is a positive correlation between payload mass and success rate for each launch site.
- Launches with payload mass exceeding 7000 kg tend to have a higher success rate.
- KSC LC 39A maintains a 100% success rate for payload mass under 5500 kg as well.

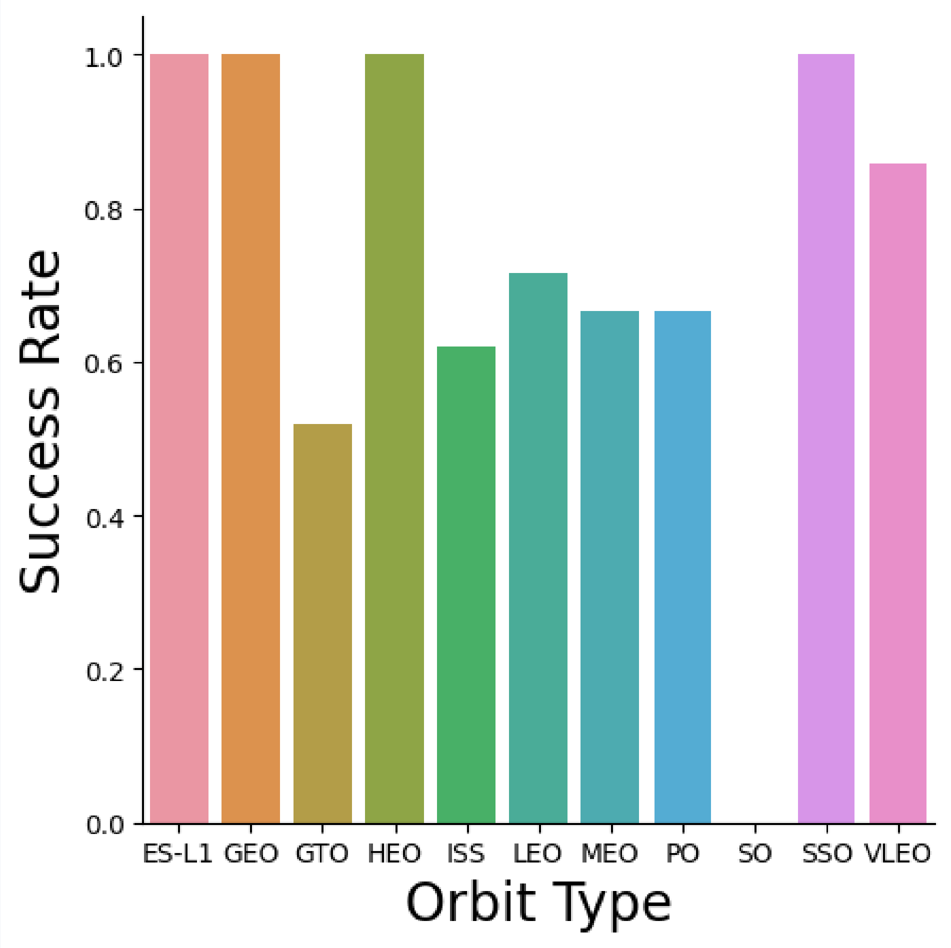
Success Rate vs. Orbit Type

Explanation:

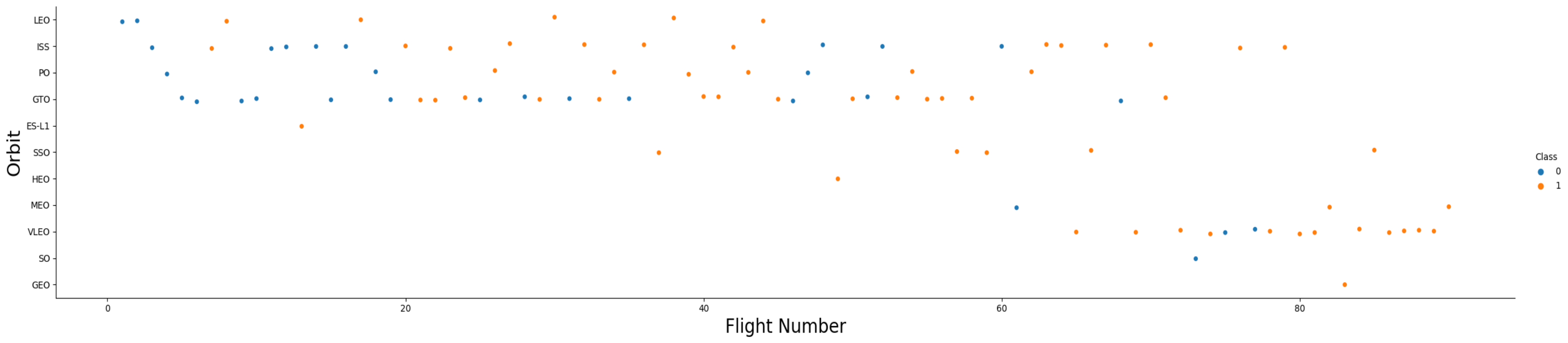
- Orbits with 100% success rate: - ES-L1, GEO, HEO, SSO

- Orbits with 0% success rate: - SO

- Orbits with success rate between 50% and 85%: - GTO, ISS, LEO, MEO, PO



Flight Number vs. Orbit Type

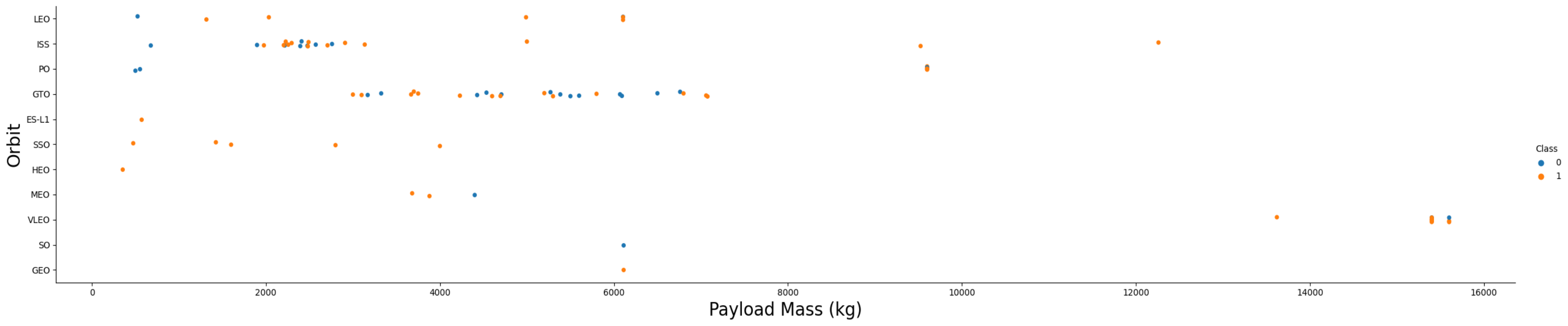


Explanation:

- In the Low Earth Orbit (LEO), the success rate appears to be positively related to the number of flights. As the number of flights increases, the success rate tends to increase as well.

- However, in the Geostationary Transfer Orbit (GTO), there seems to be no clear relationship between flight number and success rate. The success rate remains relatively consistent regardless of the number of flights.

Payload vs. Orbit Type



Explanation:

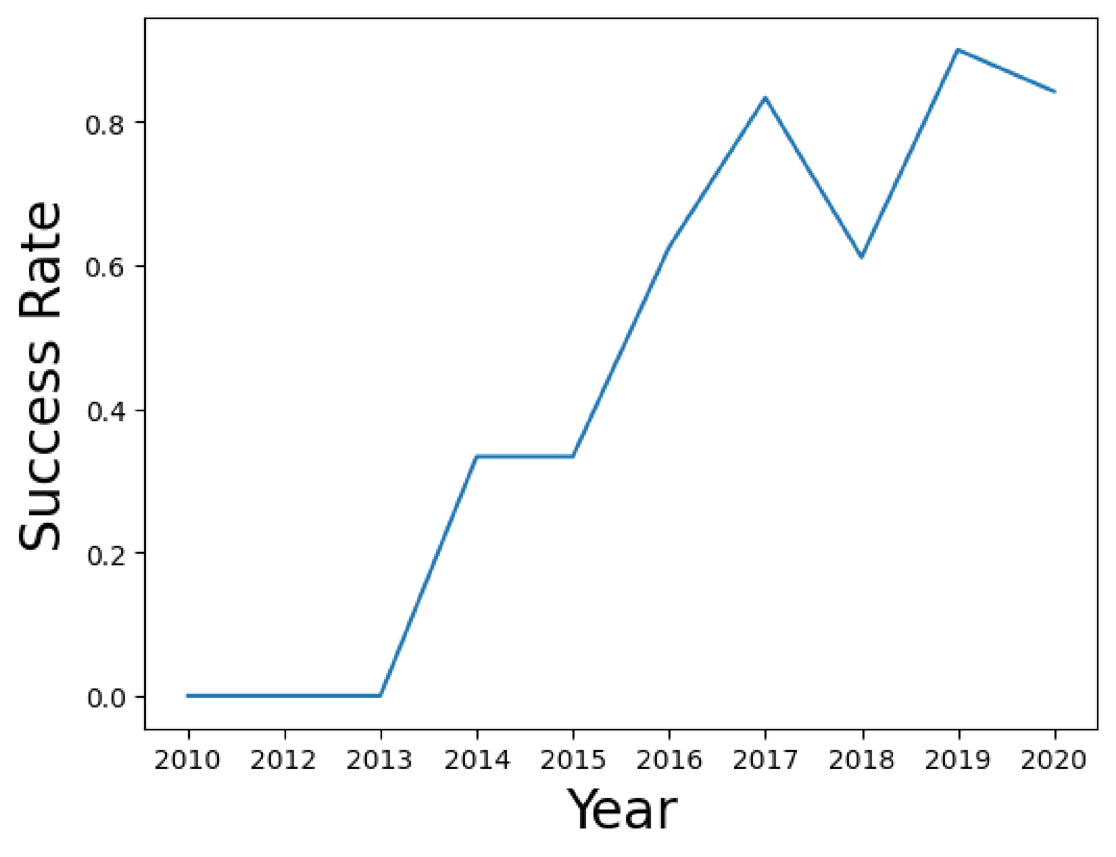
- Heavier payloads seem to have a negative influence on success rates in Geostationary Transfer Orbit (GTO) orbits. As payload mass increases, the success rate tends to decrease in GTO orbits.

- On the other hand, heavier payloads appear to have a positive influence on success rates in Geostationary Transfer Orbit (GTO) and Polar Low Earth Orbit (LEO) orbits, such as the International Space Station (ISS) orbit. As payload mass increases, the success rate tends to increase in these orbits.

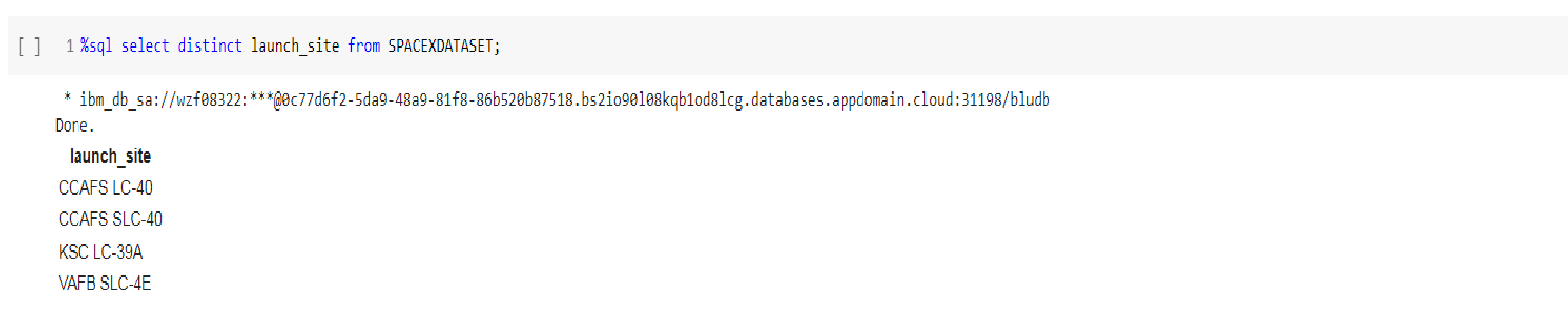
Launch Success Yearly Trend

Explanation:

- The success rate has shown a consistent increase from 2013 to 2020.
- However, in 2021, there was a slight decrease in the success rate compared to the previous year. This could be due to various factors affecting individual launch outcomes.



All Launch Site Names



```
[ ]  %sql select distinct launch_site from SPACEXDATASET;
```

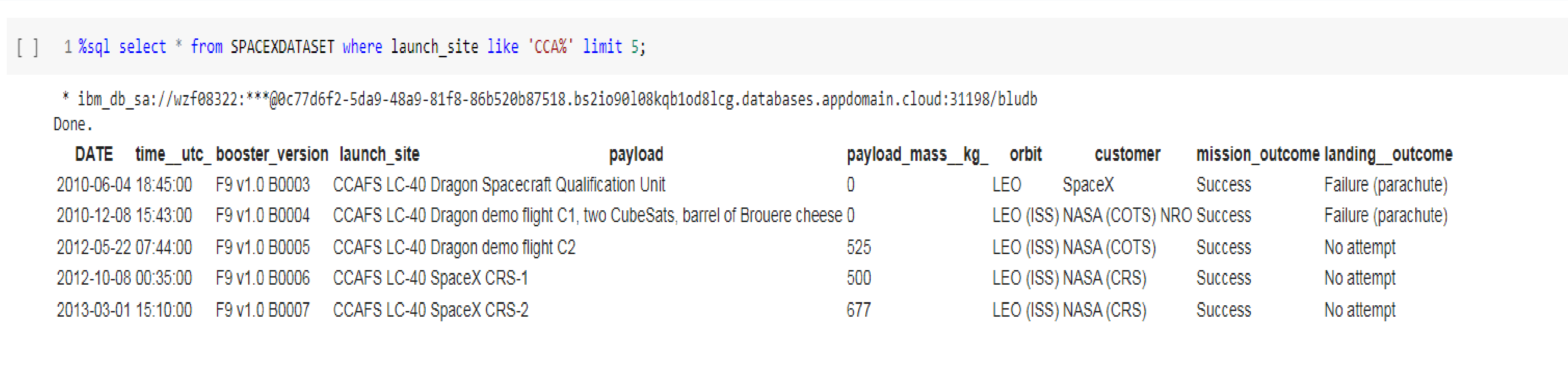
```
 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

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

Explanation:

- Displaying the names of the unique launch sites in the space mission allows us to identify the different locations from which SpaceX launches its missions.
- This information is crucial for understanding the distribution of launch sites and their significance in the overall mission operations.

Launch Site Names Begin with 'CCA'



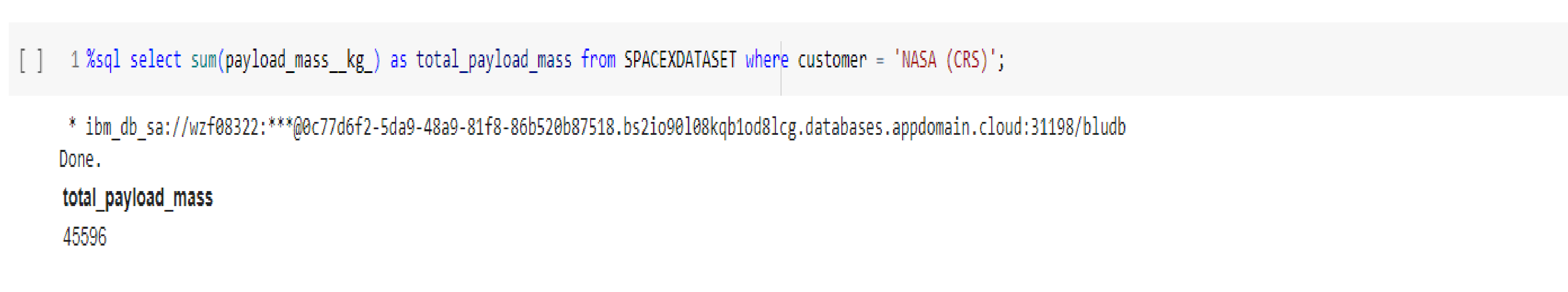
```
[ ]  %sql select * from SPACEXDATASET where launch_site like 'CCA%' limit 5;
```

```
 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
   DATE     time__utc_ booster_version  launch_site                        payload                                payload_mass__kg_   orbit      customer      mission_outcome landing__outcome
2010-06-04 18:45:00   F9 v1.0 B0003   CCAFS LC-40 Dragon Spacecraft Qualification Unit                           0                  LEO        SpaceX         Success         Failure (parachute)
2010-12-08 15:43:00   F9 v1.0 B0004   CCAFS LC-40 Dragon demo flight C1, two CubeSats, barrel of Brouere cheese 0                  LEO (ISS) NASA (COTS) NRO Success         Failure (parachute)
2012-05-22 07:44:00   F9 v1.0 B0005   CCAFS LC-40 Dragon demo flight C2                                          525                LEO (ISS) NASA (COTS)     Success         No attempt
2012-10-08 00:35:00   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
```

Explanation:

- Displaying 5 records where launch sites begin with the string 'CCA' helps us quickly identify and analyze launches associated with a specific launch site or location.
- This information is useful for investigating trends, success rates, and patterns related to launches from those particular sites.

Total Payload Mass



```
[ ]  %sql select sum(payload_mass__kg_) as total_payload_mass from SPACEXDATASET where customer = 'NASA (CRS)';
```

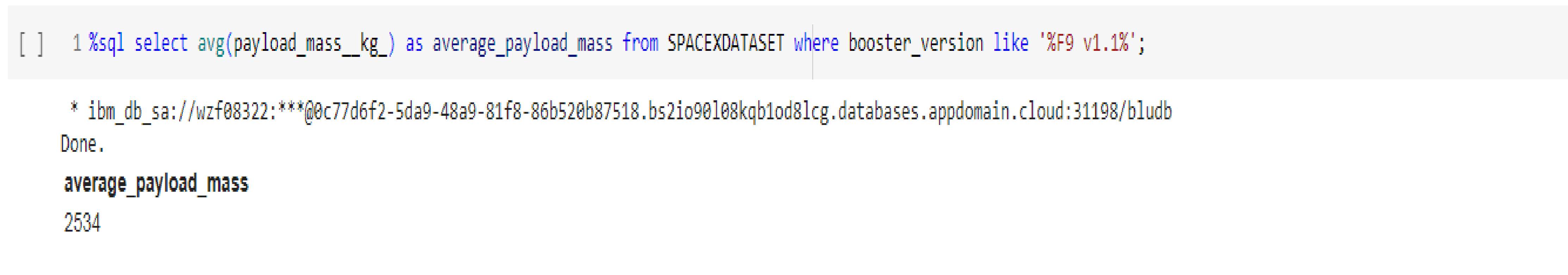
```
 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
total_payload_mass
45596
```

Explanation:

- Displaying the total payload mass carried by boosters launched by NASA (CRS) gives us insights into the magnitude of cargo transported in collaboration with NASA for various missions.

- This information helps us understand the scope and scale of SpaceX's involvement in supporting NASA's initiatives and space missions.

Average Payload Mass by F9 v1.1



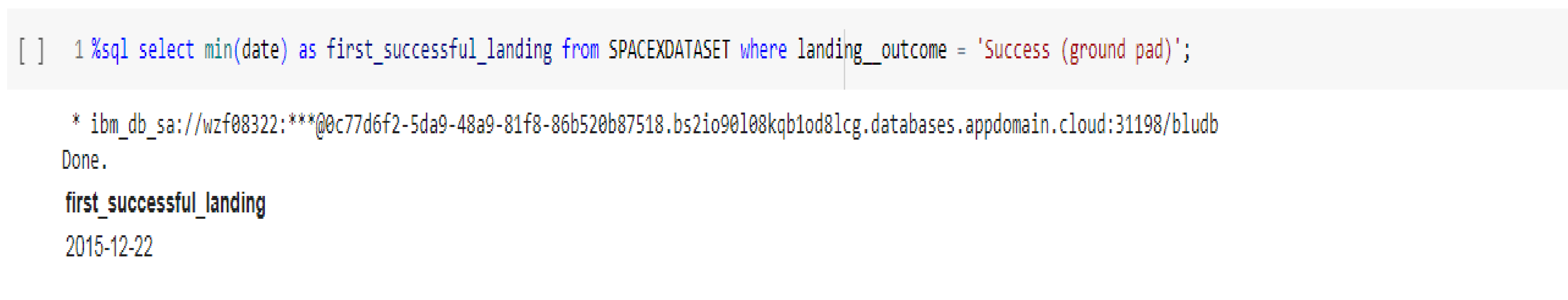
```
[ ]  %sql select avg(payload_mass__kg_) as average_payload_mass from SPACEXDATASET where booster_version like '%F9 v1.1%';
```

```
 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
average_payload_mass
2534
```

Explanation:

- Displaying the average payload mass carried by booster version F9 v1.1 helps us understand the typical payload capacity of this specific booster model.
- This information is valuable for assessing the capabilities of this particular booster version in terms of payload delivery and its relevance for different types of missions.

First Successful Ground Landing Date



```
[ ]  %sql select min(date) as first_successful_landing from SPACEXDATASET where landing_outcome = 'Success (ground pad)';
```

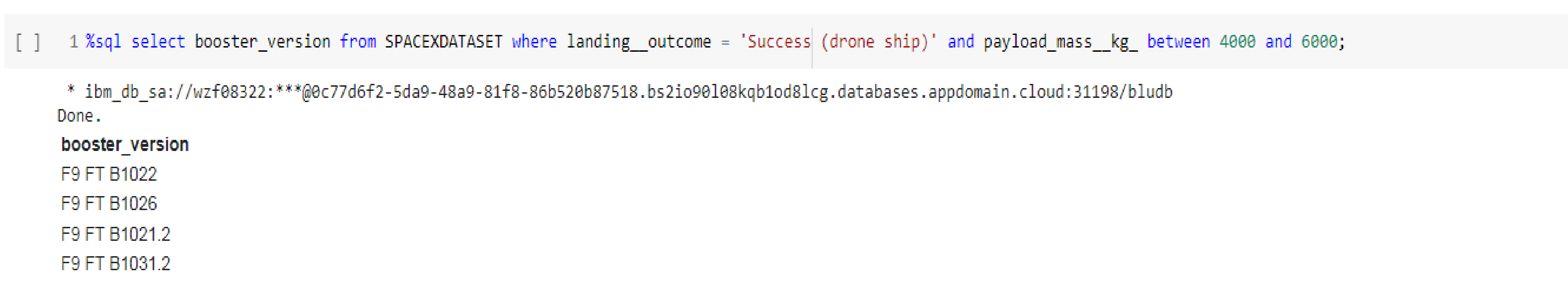
```
 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

first_successful_landing
2015-12-22

Explanation:

- Listing the date when the first successful landing outcome on a ground pad was achieved provides insights into SpaceX's progress in achieving reusability and precision landing capabilities.
- This milestone marked a significant advancement in space technology and paved the way for more cost-effective and sustainable space travel.

Successful Drone Ship Landing with Payload between 4000 and 6000



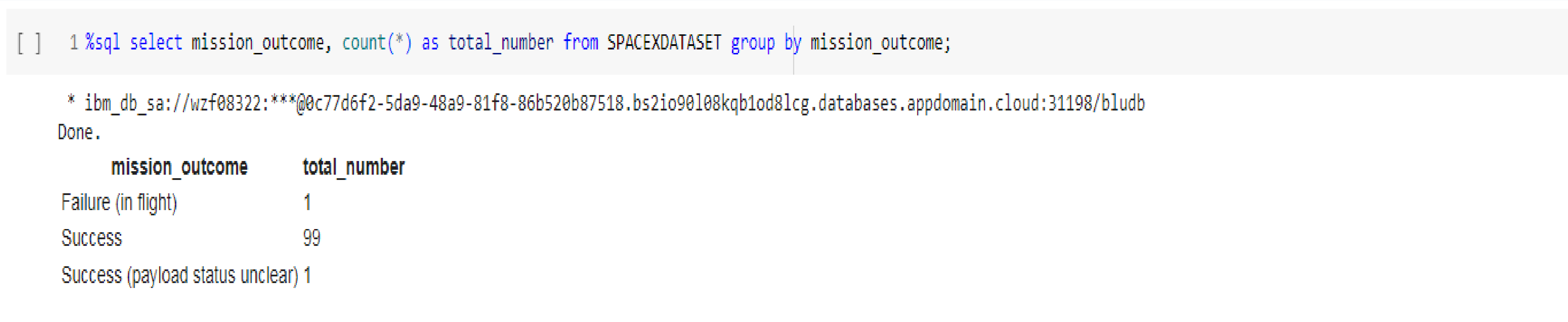
```
[ ]  %sql select booster_version from SPACEXDATASET where landing__outcome = 'Success (drone ship)' and payload_mass__kg_ between 4000 and 6000;

 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
booster_version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2
```

Explanation:

- Listing the names of the boosters which have successfully landed on a drone ship and carried a payload mass greater than 4000 kg but less than 6000 kg provides insights into the capabilities of SpaceX's reusability strategy for medium-sized payloads.

- This information highlights SpaceX's ability to recover and reuse boosters for a specific payload range, contributing to cost savings and operational efficiency.

Total Number of Successful and Failure Mission Outcomes



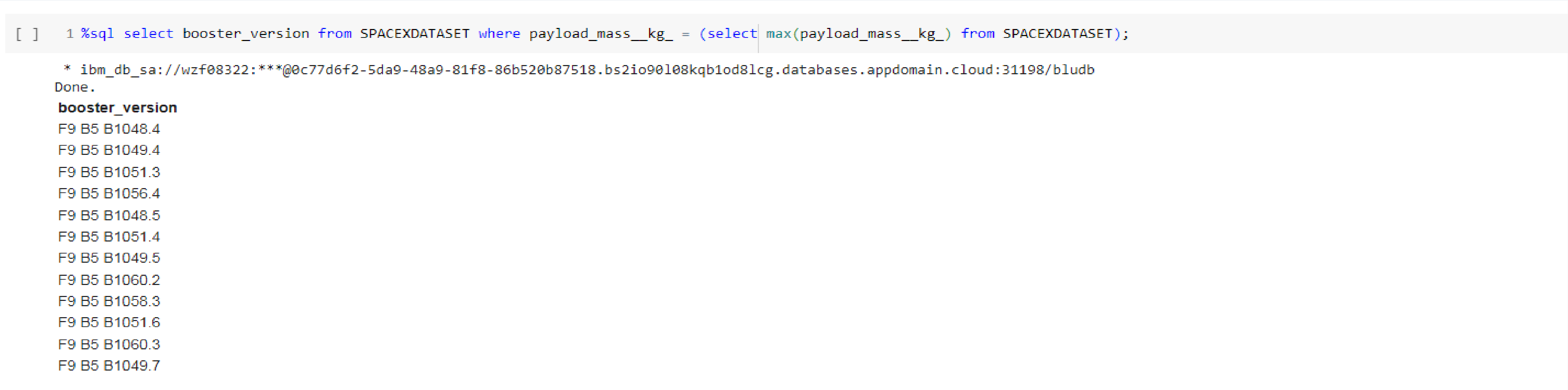
```
[ ]  %sql select mission_outcome, count(*) as total_number from SPACEXDATASET group by mission_outcome;
```

```
 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
         mission_outcome        total_number
Failure (in flight)             1
Success                         99
Success (payload status unclear) 1
```

Explanation:

- Listing the total number of successful and failed mission outcomes provides an overview of the overall success rate of SpaceX launches.
- This data allows us to assess the company's performance in terms of mission success, which is crucial for maintaining customer trust and achieving business objectives.

Boosters Carried Maximum Payload



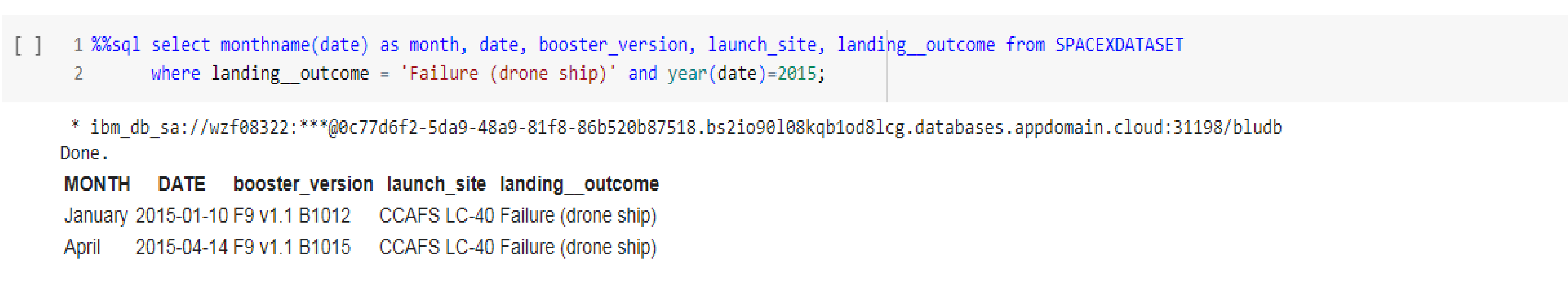
```
[ ]  %sql select booster_version from SPACEXDATASET where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXDATASET);
```

```
 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
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
```

Explanation:

- Listing the names of the booster versions that have carried the maximum payload mass provides insight into the capabilities of different booster versions.
- This information helps us identify which booster versions are capable of handling heavier payloads, which is important for planning and optimizing future missions.

2015 Launch Records



```
%%sql select monthname(date) as month, date, booster_version, launch_site, landing__outcome from SPACEXDATASET
      where landing__outcome = 'Failure (drone ship)' and year(date)=2015;
```

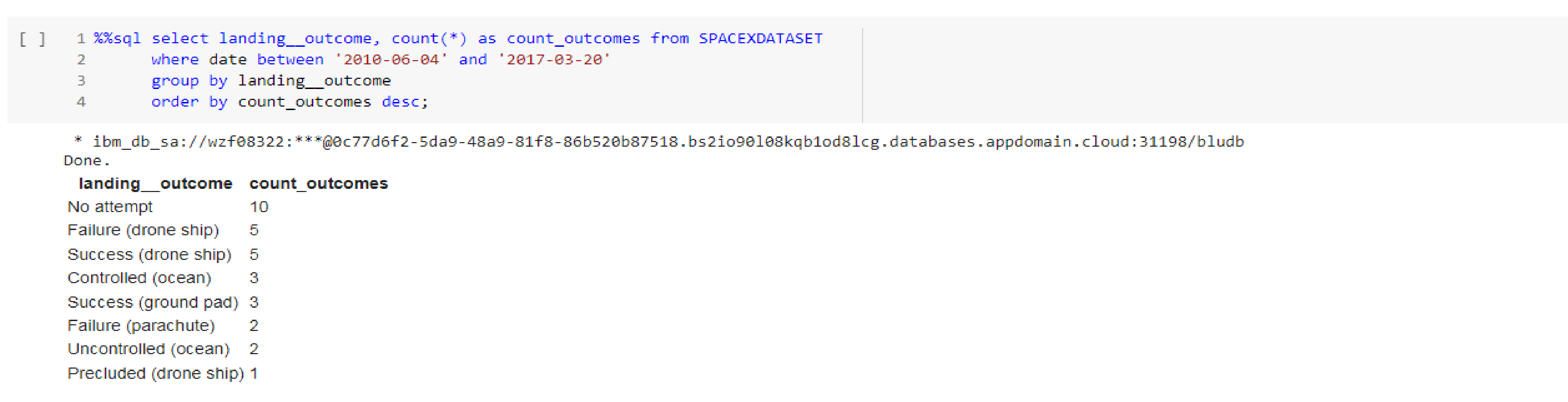
```
 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

MONTH	DATE	booster_version	launch_site	landing__outcome
January	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

Explanation:

- Listing the failed landing outcomes in drone ship, along with their booster versions and launch site names for the months in the year 2015, allows us to identify specific time periods and locations where landing failures were more common.
- This information can help us analyze potential patterns or factors contributing to landing failures during that time frame.

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



```
%%sql select landing__outcome, count(*) as count_outcomes from SPACEXDATASET
          where date between '2010-06-04' and '2017-03-20'
          group by landing__outcome
          order by count_outcomes desc;
```

```
 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

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

Explanation:

- Ranking the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the dates 2010-06-04 and 2017-03-20 in descending order helps us understand the distribution of successful and unsuccessful landings during that specific time frame.

- This analysis can provide insights into the trend of landing outcomes over the years, potentially highlighting periods of improvement or challenges in SpaceX's landing endeavors.

Section 3

Launch Sites Proximities Analysis

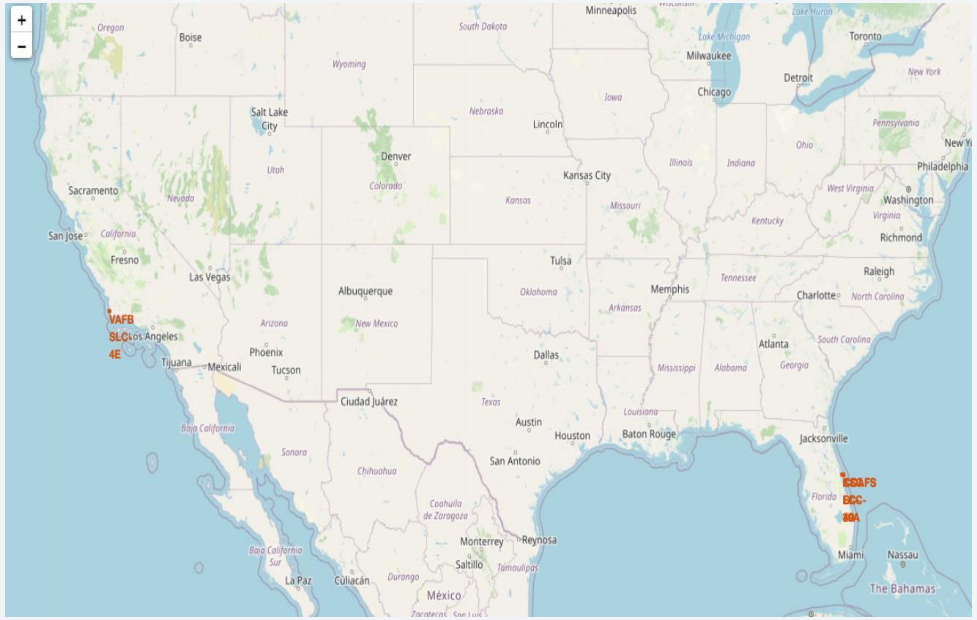
All launch sites’ location markers on a global map

Explanation:

- Most of the Launch sites are strategically located in proximity to the Equator line. The Earth's rotation is fastest at the equator, with a speed of 1670 km/hour. By launching from the equator, a spacecraft inherits this speed, aiding it in achieving and maintaining orbit.

- All the launch sites are situated near coastlines. Launching rockets over the ocean helps mitigate the risk of falling debris or explosions occurring in populated areas, enhancing safety during space missions.

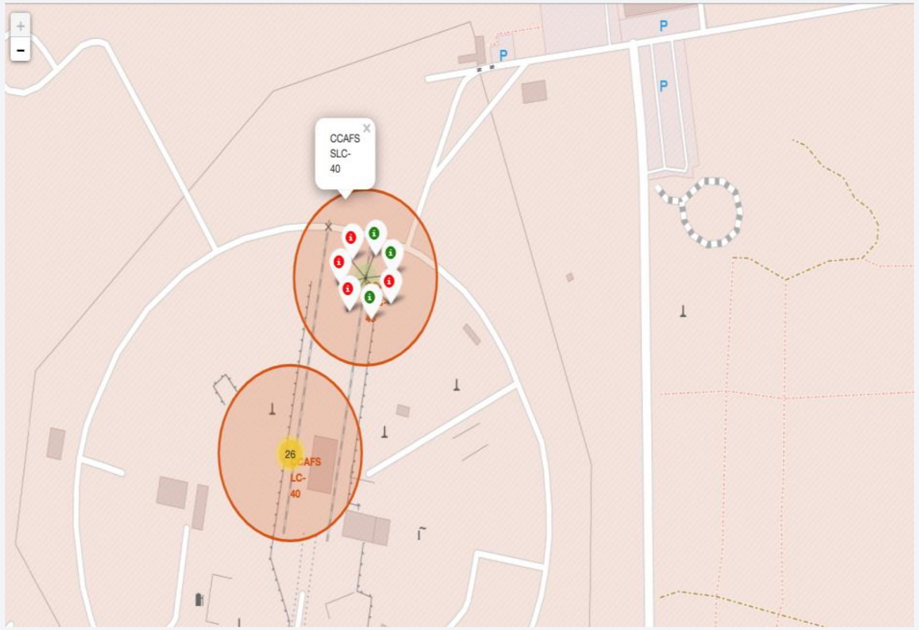
- Additionally, the selection of launch sites close to the equator and coastlines could be influenced by regulatory considerations, trajectory optimization, and the availability of suitable facilities for assembling and launching rockets.



Colour-labeled launch records on the map

Explanation:

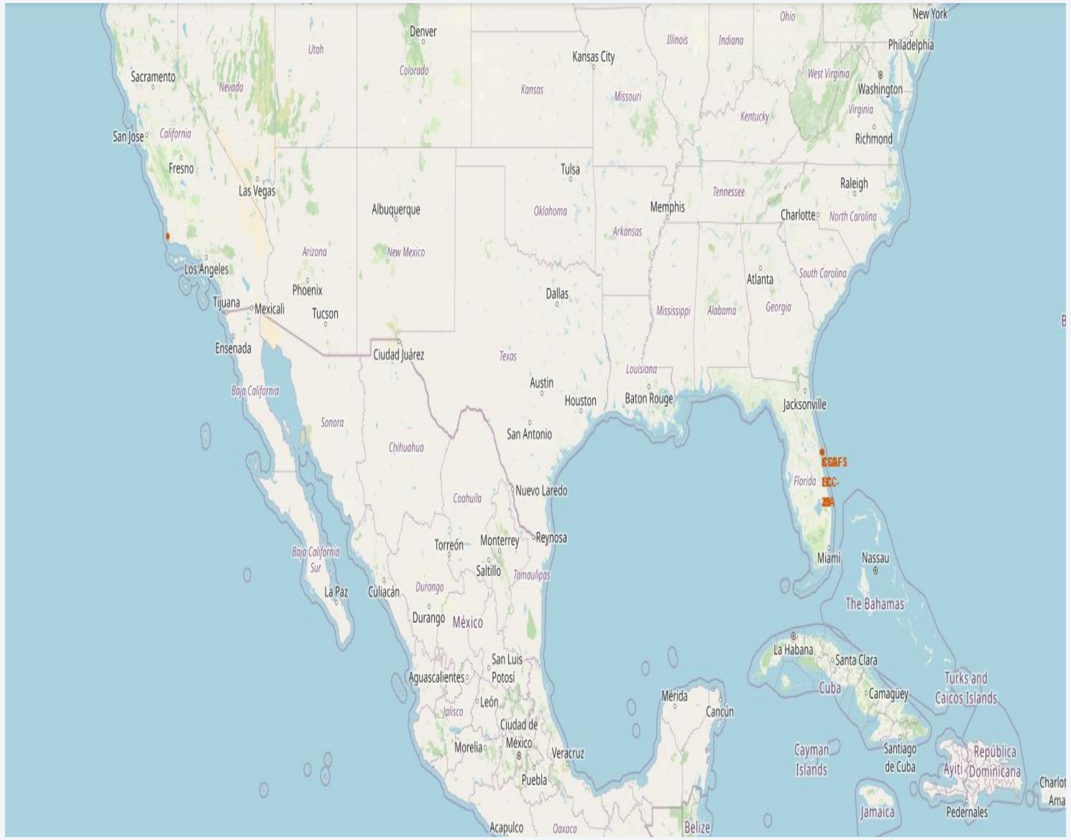
- By observing the color-coded markers on the map, it becomes evident which launch sites have achieved notable success rates.
  - ➢ Green Marker = Successful Launch
  - ➢ Red Marker = Failed Launch
- One can deduce that Launch Site KSC LC-39A exhibits an exceptionally high Success Rate based on the markers associated with it.
- The varying success rates among different launch sites may result from factors such as operational procedures, technical capabilities, and historical mission performance.

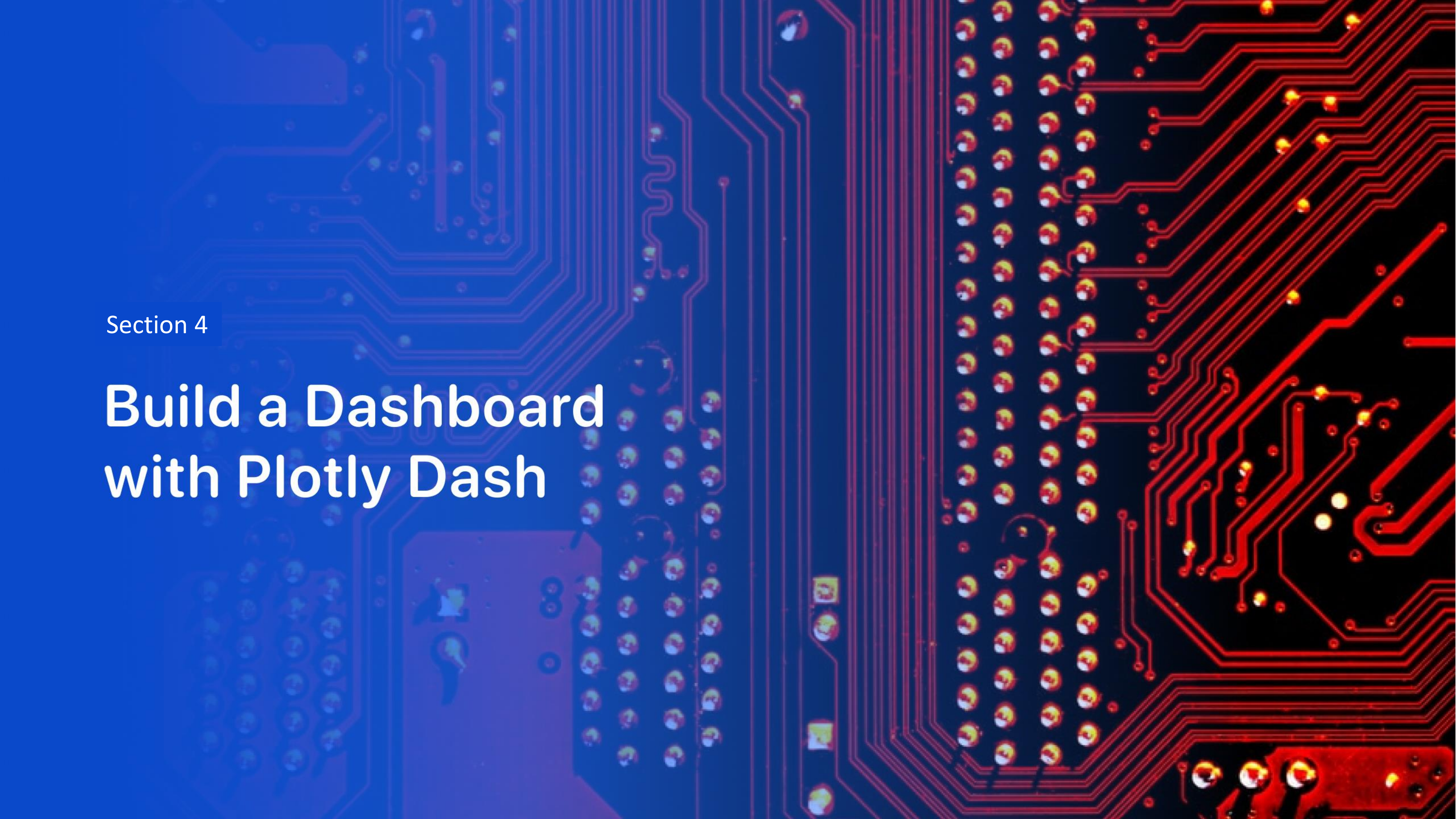


Distance from the launch site KSC LC-39A to its proximities

Explanation:

- A visual analysis of launch site KSC LC-39A provides clear insights:
- It is in close proximity to a railway (approximately 15.23 km).
- It is in close proximity to a highway (approximately 20.28 km).
- It is in close proximity to the coastline (approximately 14.99 km).
- Additionally, launch site KSC LC-39A is relatively near its closest city, Titusville (approximately 16.32 km).
- It's worth noting that a failed rocket, given its high velocity, can cover distances of 15-20 km in mere seconds. This presents potential dangers to populated areas, underscoring the significance of selecting launch sites strategically to minimize risks to human settlements.

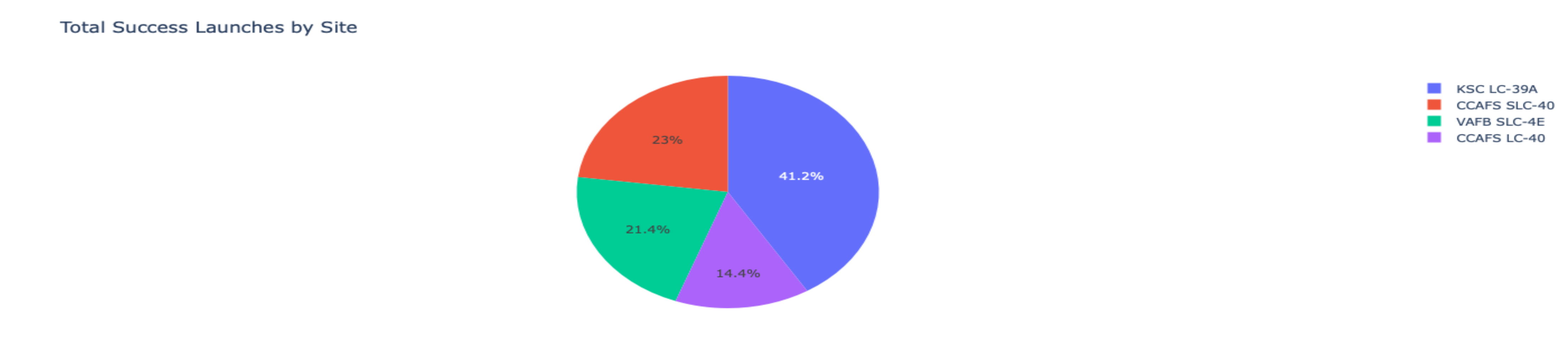




Section 4

Build a Dashboard with Plotly Dash

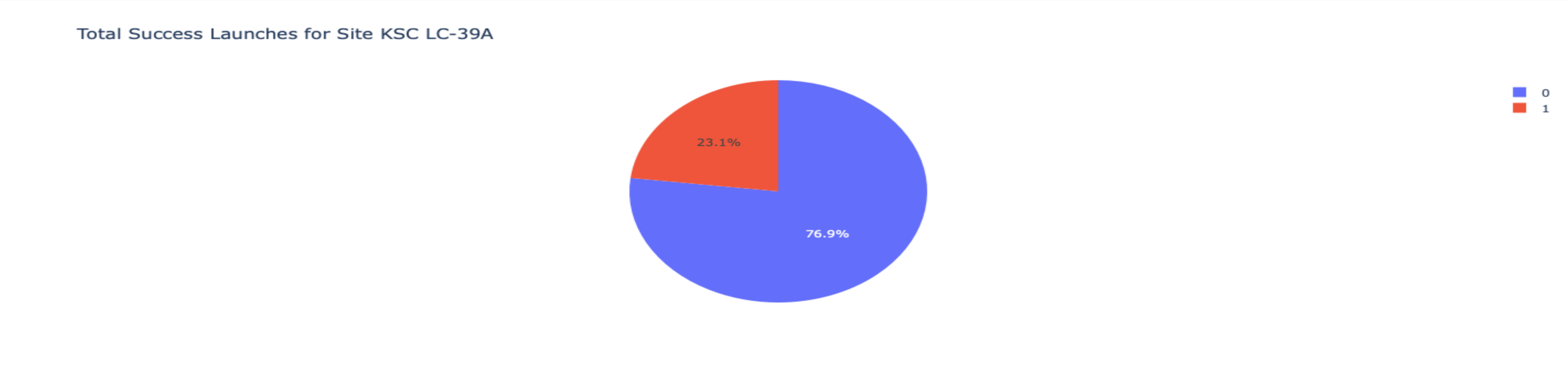
The success percentage by each sites.



Explanation:

- The chart unequivocally illustrates that among all sites, KSC LC-39A boasts the highest success rate.
- Furthermore, the data underscores the pivotal role of booster reusability, which contributes to SpaceX's success in reducing launch costs.
- The graphical representation provides a clear and unambiguous comparison of success rates across all launch sites, with KSC LC-39A emerging as the leader.
- The success rate of KSC LC-39A and the success of reusability highlight SpaceX's commitment to continuous innovation and technological advancement.

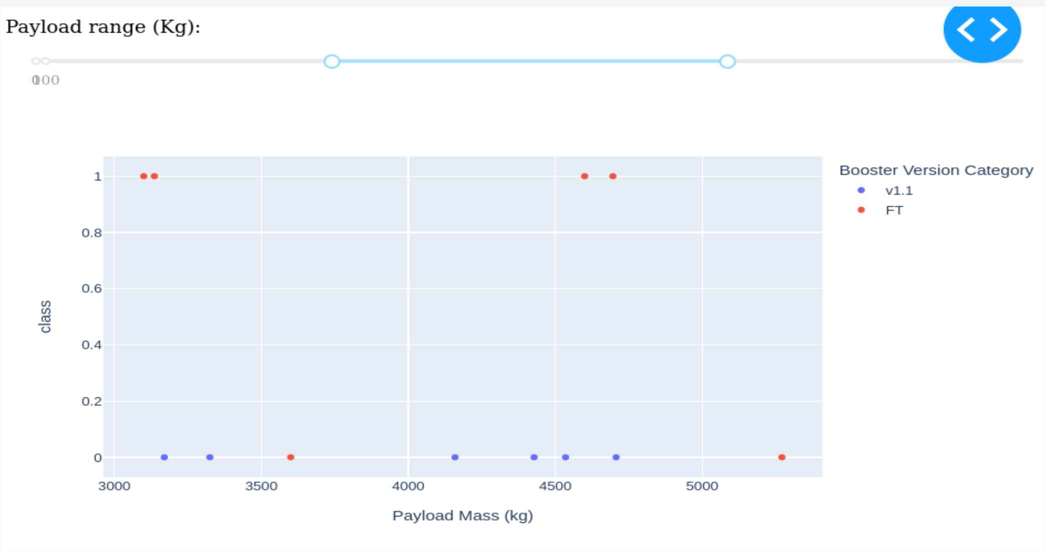
The highest launch-success ratio: KSC LC-39A



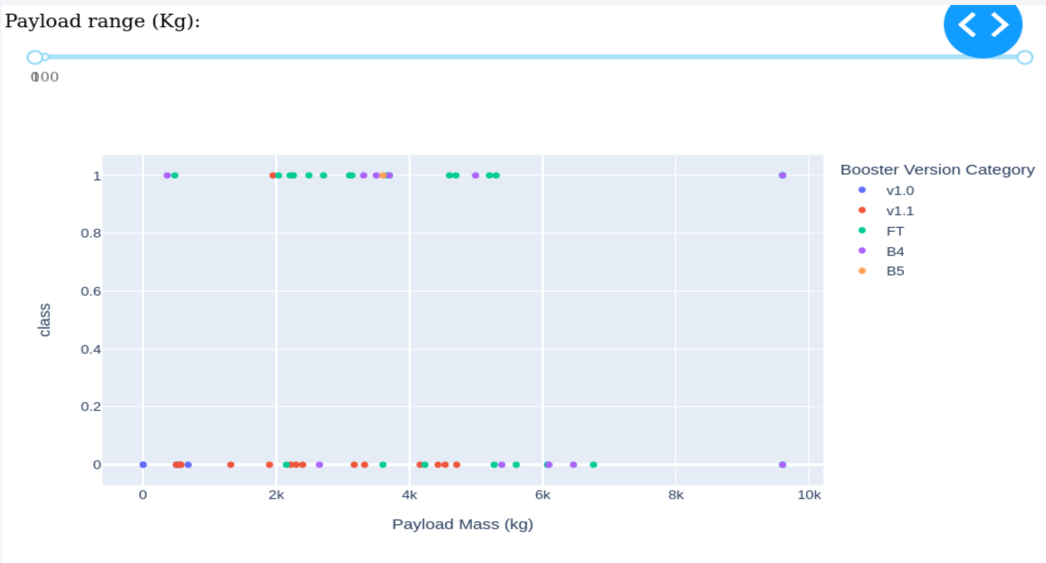
Explanation:

- The data analysis clearly indicates that KSC LC-39A has achieved a remarkable launch success rate of 76.9%, which is the highest among the various launch sites considered.
- This achievement is supported by a track record of 10 successful landings out of 13 attempted launches, with only 3 instances of failure. This high success rate reflects the proficiency and reliability of SpaceX's operations at this site.
- Additionally, VAFB SLC-4E and CCAFS LC-40 have also demonstrated notable success rates, further emphasizing their importance in SpaceX's achievements. While their exact success rates may not be as high as KSC LC-39A's, they still play a significant role in the company's overall launch accomplishments.

Payload vs Launch Outcome Scatter Plot



Payload range (Kg):



Payload range (Kg):

Explanation:

- The data visualizations clearly depict a trend where payloads falling within the weight range of 2000 to 5500 kg demonstrate the highest success rate in SpaceX's launch history.

- This range showcases a consistent success rate, which suggests that payloads of this size enjoy optimal conditions for successful launches and missions.

- Interestingly, payloads exceeding 7000 kg also exhibit a favorable success rate. This observation highlights SpaceX's ability to handle and launch larger payloads, reflecting their capacity to adapt their launch vehicles and processes to accommodate different payload masses.

Section 5

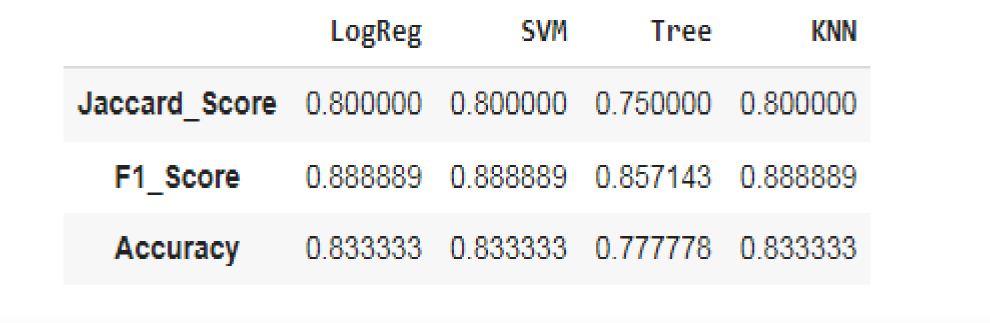
Predictive Analysis (Classification)

Classification Accuracy

Explanation:

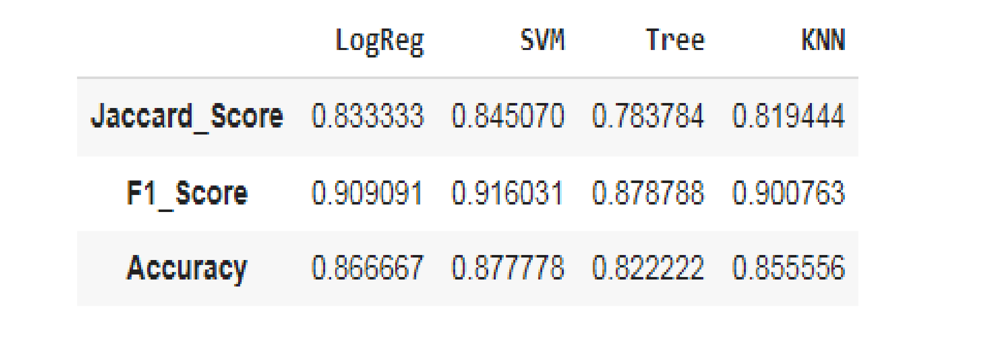
- The outcomes of the Test Set scores do not allow us to definitively determine the most effective method.
- The uniform Test Set scores could potentially stem from the limited size of the test sample (18 samples). Thus, we broadened our analysis to encompass the entire Dataset.
- The comprehensive evaluation across the entire Dataset reaffirms the SVM Model's superiority. This model not only yields superior scores but also attains the highest level of accuracy.

Scores and Accuracy of the Test Set



	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.750000	0.800000
F1_Score	0.888889	0.888889	0.857143	0.888889
Accuracy	0.833333	0.833333	0.777778	0.833333

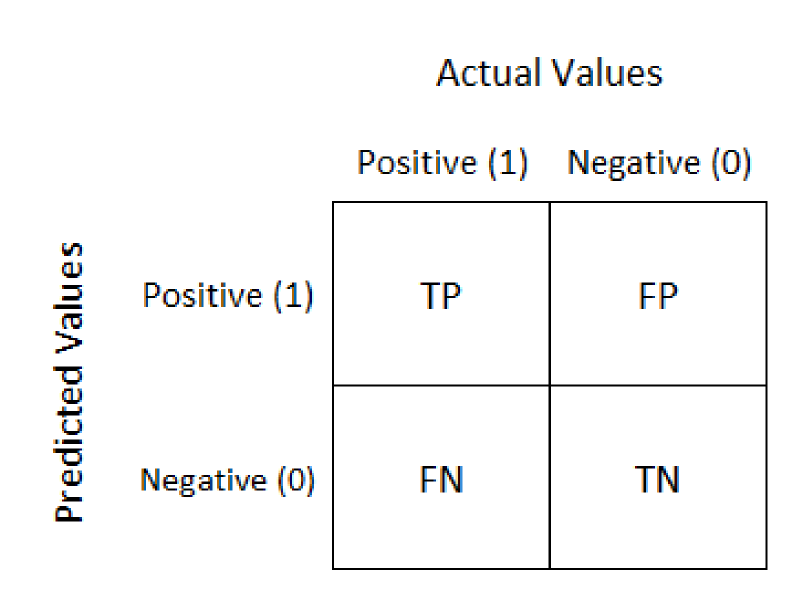
Scores and Accuracy of the Entire Data Set



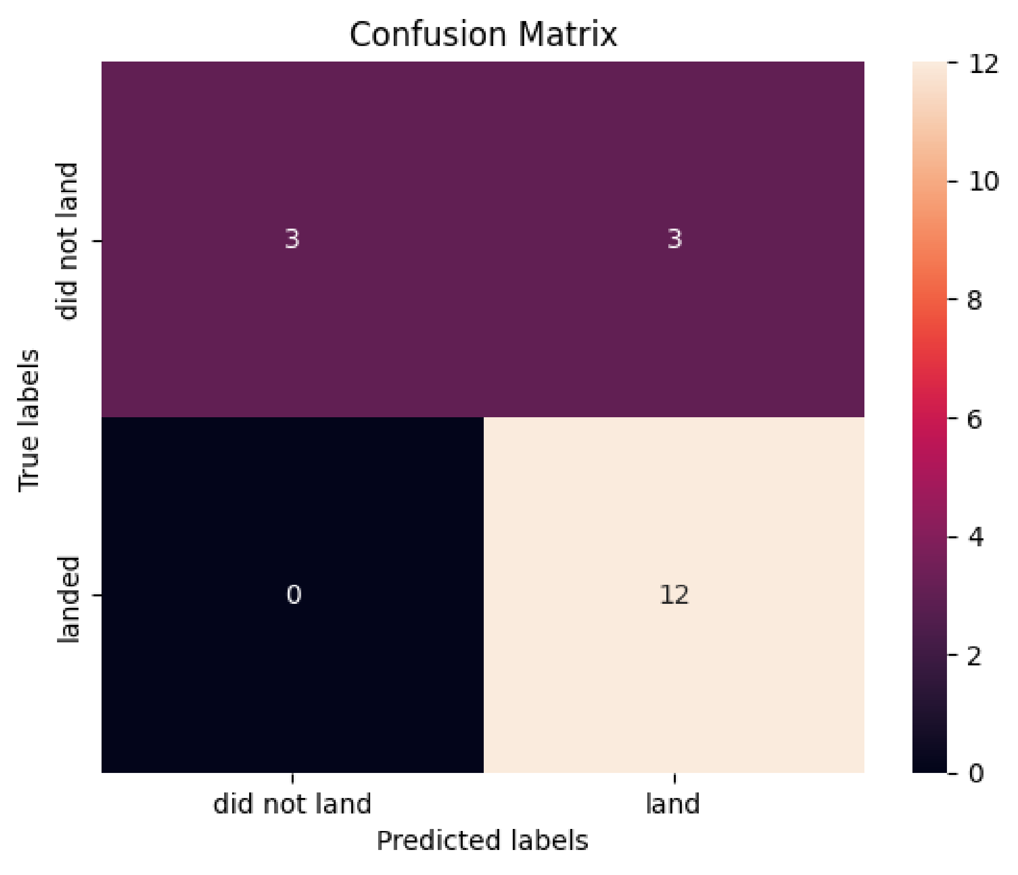
	LogReg	SVM	Tree	KNN
Jaccard_Score	0.833333	0.845070	0.783784	0.819444
F1_Score	0.909091	0.916031	0.878788	0.900763
Accuracy	0.866667	0.877778	0.822222	0.855556

Confusion Matrix

The confusion matrix reveals the performance of the SVM model. It correctly predicted 12 successful landings (True Positives) and 3 unsuccessful landings (True Negatives). However, it misclassified 3 successful landings as unsuccessful (False Negatives), and also 3 unsuccessful landings as successful (False Positives).



		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN



Conclusions

- SVM Model excels as the most suitable algorithm for this dataset.
- Lower payload launches exhibit superior outcomes compared to heavier payloads.
- Launch sites are primarily near the Equator, ensuring efficient launches. All sites are very close to coastlines, mitigating risks.
- Launch success rates consistently climb over the years.
- KSC LC-39A stands out with the highest success rate among launch sites.
- Orbits ES-L1, GEO, HEO, and SSO achieve a flawless 100% success rate.

Appendix

Python Code Snippets:

- Code for requesting rocket launch data from SpaceX API and converting it into a DataFrame.
- Code for performing SQL queries to extract specific information from the dataset.
- Code for creating various charts using libraries like Plotly and Matplotlib.

SQL Queries:

- SQL queries used to retrieve information such as unique launch sites, payload statistics, successful and failed mission outcomes, etc.

Charts and Visualizations:

- Scatter plots depicting relationships between Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch Site, etc.
- Bar charts illustrating comparisons among different categories, such as launch outcomes.
- Line charts showcasing trends in data over time, including the success rate over the years.

Jupyter Notebook Outputs:

- Outputs from Jupyter Notebook detailing data exploration, analysis, visualization, and machine learning model development.

Data Sets:

- Original data collected from SpaceX API and Wikipedia through web scraping.
- Processed and cleaned data used for analysis and visualization.

Thank you!