DATA SCIENCE CAPSTONE PROJECT

IBM Developer
SKILLS NETWORK

Horacio A. Rodriguez Carracedo

July 10th, 2024

OUTLINE

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

IBM Developer SKILLS NETWORK

EXECUTIVE SUMMARY

Summary of Methodologies

- Data Collection.
- Data Wrangling.
- Exploratory Data Analysis with Data Visualization and SQL.
- Building an Interactive Map with Folium.
- Building an Interactive Map with Plotly Dash.
- Predictive Analysis for each Classification Model.

Summary of All Results

- Exploratory Data Analysis Results.
- Interactive Analytics Demo in Screenshots.
- Predictive Analysis Results.

IBM Developer
SKILLS NETWORK

INTRODUCTION

Project Background and Context

Utilizing advanced engineering, SpaceX has drastically reduced the cost of space travel. Key to these savings is the reusability of the Falcon 9's first stage, a breakthrough that disrupts traditional single-use launch models.
The ability to predict and ensure the successful landing of the Falcon 9's first stage is vital. Success in this area directly correlates to cost savings and operational sustainability, which are critical in maintaining SpaceX's competitive pricing of $62 million per launch, compared to over $165 million charged by competitors.

Research Objectives

- ✓ *Predictive Modeling*: To determine if the first stage will land successfully.
- ✓ *Key Factors*: Analyze how payload mass, launch site, flight number, and orbit affect landing success.
- ✓ *Algorithm Evaluation*: Identify the best algorithm for classification of landing outcomes.

IBM Developer
SKILLS NETWORK

METHODOLOGY

IBM Developer
SKILLS NETWORK

METHODOLOGY

A.- Data Collection

Sources

SpaceX API: For up-to-date launch and vehicle data.

Wikipedia (Web Scraping): For historical and contextual data.

Data Preparation

DataFrame Creation: Structuring raw data for easier handling.

Data Cleaning: Filtering and dealing with missing values.

One Hot Encoding: Transforming categorical data for analysis.

B.- Data Analysis

Exploratory Data Analysis (EDA)

Tools: Visualizations and SQL for uncovering trends and relationships.

Interactive Visual Analytics

Using Folium & Plotly-Dash: For geospatial and interactive dashboard creation.

C.- Predictive Analysis

Model Development

Classification Models: Building and tuning to predict outcomes accurately.

Model Evaluation: Ensuring optimal performance through rigorous testing.

IBM Developer
SKILLS NETWORK

DATA COLLECTION

Data Sources & Methods

SpaceX API

- *Collection*: Utilize GET requests to fetch launch data.
- *Processing*: Decode JSON responses, normalize and load into a pandas DataFrame.
- *Data Cleaning*: Identify and fill missing values to ensure data integrity.

Wikipedia Web Scraping

- *Collection*: Scrape Falcon 9 launch records using *BeautifulSoup*.
- *Processing*: Extract HTML tables, parse and convert data into a structured pandas DataFrame.
- Export the data to a CSV file.

[Github Link: Data Collection API]

```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json'
```

We should see that the request was successfull with the 200 status response code

```
response.status_code
```

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False
4	6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	Landing
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb234
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb234
92	89	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e534
93	90	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234

IBM Developer SKILLS NETWORK

WEB SCRAPPING

- Getting HTML response
- Creating BeautifulSoup Object
- Finding Tables
- Obtaining Column Names
- Creating a dictionary and appending data
- Dictionary to dataframe and df to csv file.

```python
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
response = requests.get(static_url)

soup = BeautifulSoup(response.text, 'html.parser')

html_tables = soup.find_all('table')
first_launch_table = html_tables[2]

th_elements = first_launch_table.find_all('th')
for th in th_elements:
    name = extract_column_from_header(th)
    if name is not None and len(name) > 0:
        column_names.append(name)

launch_dict= dict.fromkeys(column_names)

df= pd.DataFrame({ key:pd.Series(value) for key, value in launch_dict.items() })
df.to_csv('spacex_web_scraped.csv', index=False)
```

Data Extraction

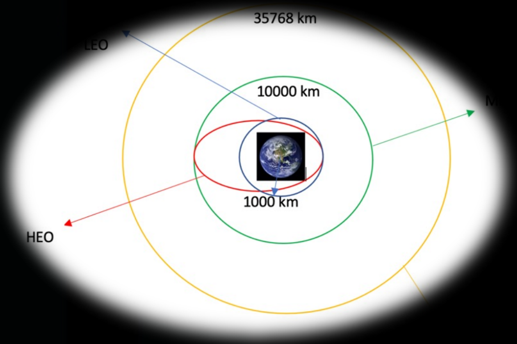
```
1
4 June 2010
18:45
F9 v1.0B0003.1
CCAFS
Dragon Spacecraft Qualification Unit
Dragon Spacecraft Qualification Unit
LEO
SpaceX
Success

Failure
2
8 December 2010
15:43
F9 v1.0B0004.1
CCAFS
Dragon
Dragon
LEO
NASA
Success
Failure
3
22 May 2012
...
CCSFS
Transporter-1
Transporter-1
```

[Github Link: Data Collection with Web Scraping](#)

DATA WRANGLING

Objective

- ✓ Clean and unify messy and complex datasets for analysis
- ✓ Convert outcomes into binary training labels for predictive modeling.



Steps Involved

- ❖ *Load Data*: Import data using API calls and web scraping and load them into Pandas DataFrame.
- ❖ *Data Cleaning and Transformation*: Identify and fill missing values, remove irrelevant data and convert textual data into structured formats.
- ❖ *Feature Engineering*:
    1. Boolean Encoding: Convert outcome into binary labels.
    2. Aggregation: Calculate statistics such as launch counts per site and occurrences per orbit.
- ❖ *Export Processed Data*: Export the structured data to a CSV file.

```
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
Name: Outcome, dtype: int64
```

Github Link: Data Wrangling

IBM Developer
SKILLS NETWORK

EDA AND VISUAL ANALYTICS

Objective

Perform comprehensive Exploratory Data Analysis (EDA) to uncover trends, patterns, and relationships within the SpaceX launch dataset.

Data Visualization Techniques

➢ **Scatter Plots**

Analyze relationships between variables:

i) Payload and Flight Number.
ii) Flight Number and Launch Site.
iii) Payload Mass and Launch Site
iv) Flight Number and Orbit Type
v) Payload Mass and Orbit Type

➢ **Bar Charts**

Illustrate the success rate across different orbit types to pinpoint the most reliable orbits. Compare mission outcomes to assess performance trends over time.

➢ **Line Graphs**

Track changes in launch success rates over the years to detect stability and improvement in launch operations.

Highlight dependency and potential correlations relevant for model predictions.

[Github Link: Data Visualization]

IBM Developer SKILLS NETWORK

EDA WITH SQL

Objective

Extract and analyze key metrics from the SpaceX dataset using SQL queries to provide actionable insights.

SQL Queries Performed

- ✓ Identify all launch site names.
- ✓ Determine the launch site names that begin with 'CCA'
- ✓ Obtain the total payload mass carried.
- ✓ Obtain average payload mass carried by booster version F9 v1.1
- ✓ Obtain the date when the first successful landing outcome in ground pad was achieved.
- ✓ List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
- ✓ List the total number of successful and failure mission outcomes.
- ✓ List the names of the booster versions which have carried the maximum payload mass.
- ✓ Launch site for the months in year 2015.
- ✓ Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

[Github Link: EDA with SQL]

IBM Developer SKILLS NETWORK

BUILDING AN INTERACTIVE MAP WITH FOLIUM

Objective

Enhance decision-making with a dynamic map showing launch success rates and site proximity to key infrastructures using Folium.

Map Features

Markers and Colors
Added markers with circle popup labels and text labels using latitude and longitude coordinates. We used the latitude and longitude coordinates for each launch site and added a circle marker with a label.
Distance Lines
Show distances from launch sites to infrastructures like railways and highways to assess risks and logistics.

Data Handling
Classify outcomes as 0 (failure) or 1 (success) for clarity.

Cluster Analysis
Simplify visualization with marker clusters that highlight success rate trends.

[Github Link: Map with Folium]()

IBM Developer
SKILLS NETWORK

INTERACTIVE DASHBOARD WITH PLOTLY DASH

Objective

Create a dynamic dashboard to analyze launch data using Plotly Dash, providing interactive visualizations to facilitate data-driven decision-making.

Dashboard Features

Launch Sites Dropdown

Allows users to select specific launch sites for targeted data analysis.

Pie Chart

Displays the total count of successful launches versus failed launches for all sites or a selected site, enhancing comparative insights

Scatter Chart

Shows the relationship between payload mass and launch success rate, helping to identify trends across different booster versions.

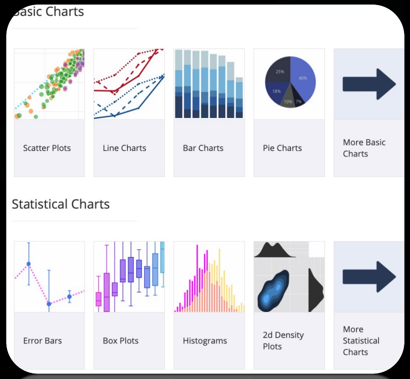
[Github Link: Plotly Dashboard]

Payload Mass Slider

Enables filtering of launch data based on payload mass ranges, allowing users to explore specific payload scenarios.

Interactivity and Data Sources

Interactive elements like dropdowns, sliders, and live charts provide a hands-on approach to exploring launch data.
Data sourced from comprehensive databases via APIs and web scraping, ensuring up-to-date and accurate information.



IBM Developer SKILLS NETWORK

PREDICTIVE ANALYSIS (CLASSIFICATION)

Data Preparation

Transformation: Convert data into NumPy arrays.

Standardization: Utilize StandardScaler to normalize data.

Setup and Training

Data Splitting: Divide into training and testing sets.

Model Selection: Use SVM, Decision Tree, KNN, and Logistic Regression.

Hyperparameter Optimization: Apply GridSearchCV with cross-validation to fine-tune models.

Model Evaluation

Performance Metrics: Evaluate using Jaccard Score, F1 Score, and confusion matrix.

Best Model Selection: Determine the most efficient model based on performance metrics.

Visualization and Results

Confusion Matrix: Graphical representation of the model's effectiveness.

Optimal Model Identification: Highlight the best model in terms of accuracy and performance.

Github Link: Predictive Analysis

IBM Developer
SKILLS NETWORK

RESULTS

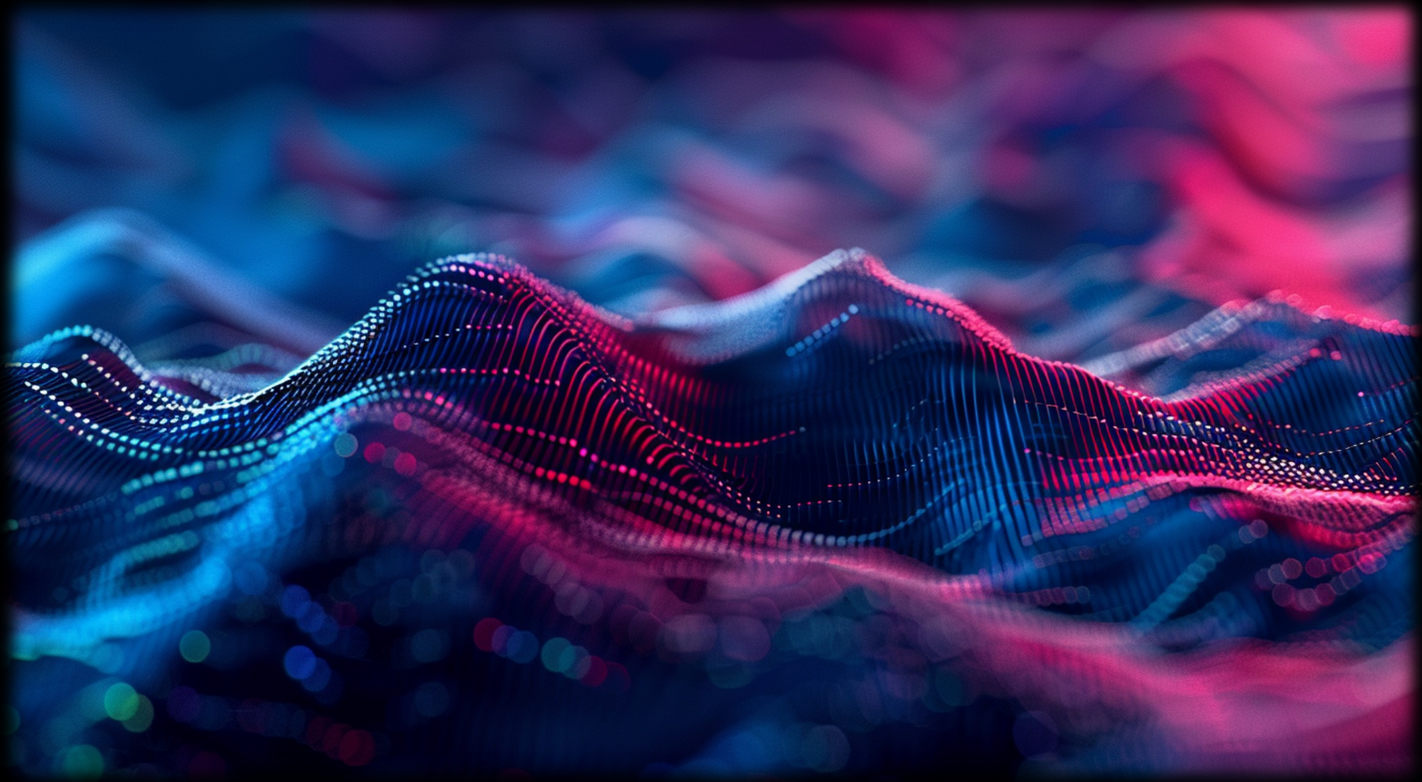
Exploratory Data Analysis Results

Interactive Analytics Demo in Screenshots

Predictive Analysis Results

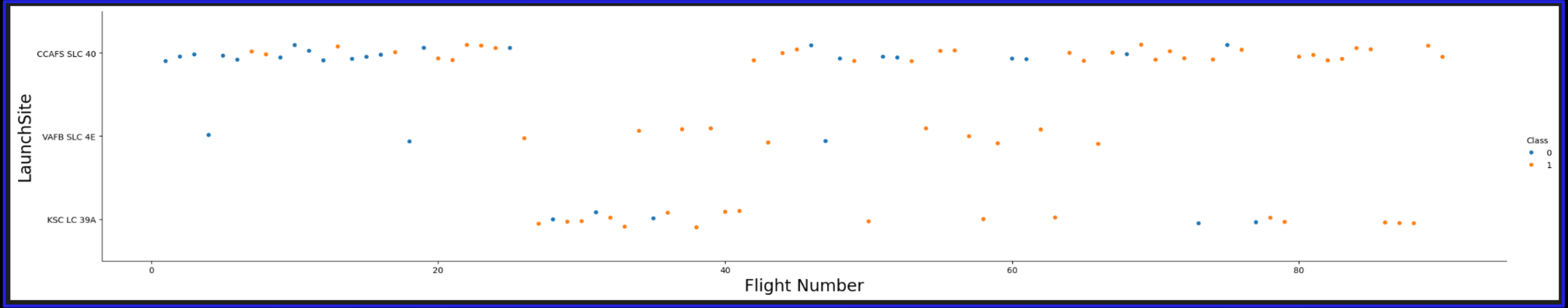
IBM Developer SKILLS NETWORK

EDA WITH VISUALIZATION



IBM Developer
SKILLS NETWORK

FLIGHT NUMBER VS LAUNCH SITE



General Trend

- *CCAFS SLC 40*: Hosts the most launches; success rates vary, not strictly improving with more launches.
- *VAFB SLC 4E & KSC LC 39*: Fewer flights but generally higher success rates, suggesting efficiency isn't solely based on volume.
- Early flights show lower success; recent flights across all sites have improved.

Key Insights

Experience contributes to success.

IBM Developer SKILLS NETWORK

PAYLOAD MASS VS LAUNCH SITE

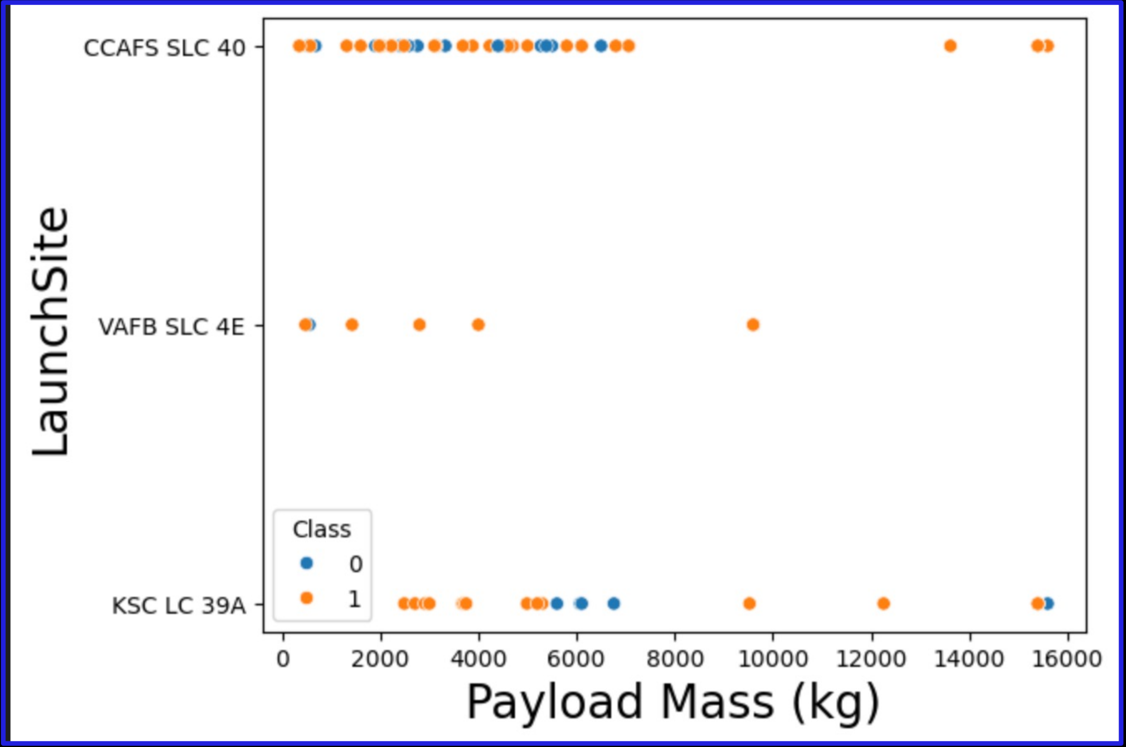
General Trend

- Higher payload masses (>7000 kg) generally correlate with increased launch success.
- *KSC LC 39A*: Exceptional success across payload sizes, perfect success rate for payloads under 5500 kg.
- *VAFB SLC 4E:* Demonstrates positive trend towards success with larger payloads.

Key Insights

Success dependency on payload mass varies by site.

KSC LC 39A's consistent success across payload scales suggests superior operational efficiency.



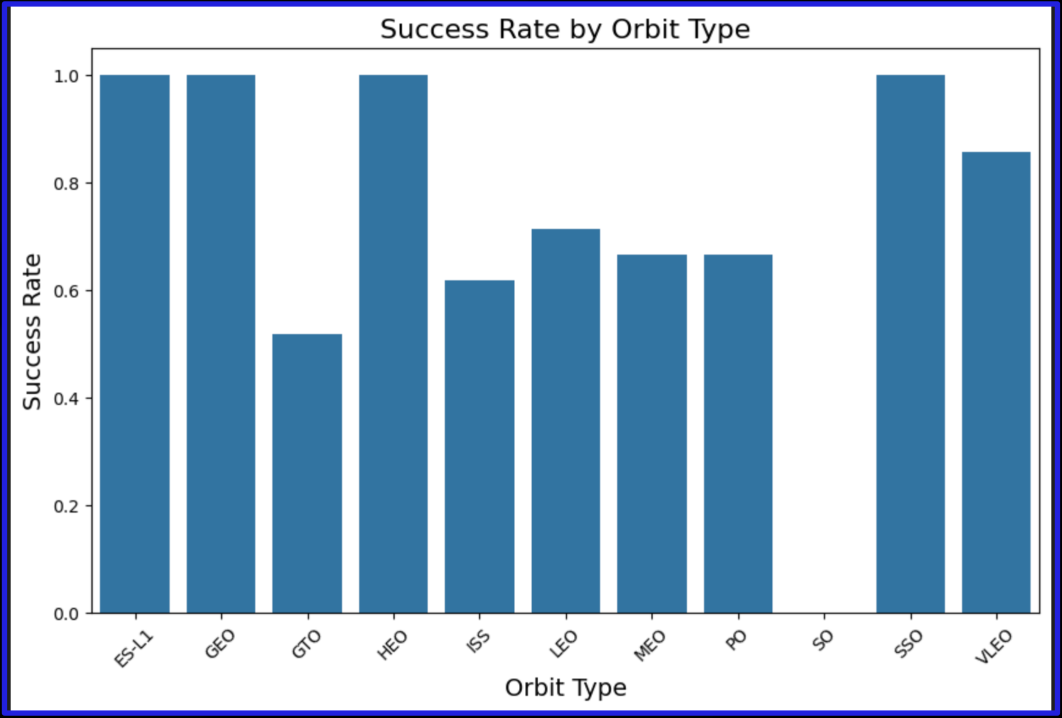
SUCCESS RATE VS ORBIT TYPE

General Trend

- Orbits ES-L1, GEO, HEO, SSO, VLEO: Near or at 100% success rate, indicating high reliability.
- Orbits GTO, ISS, LEO, MEO, PO: Success rates between 50% and 85%, reflecting variable outcomes.
- Orbit SO: 0% success, indicating significant challenges.

Key Insights

Tailoring mission strategies to orbit-specific success trends can optimize overall mission success.



IBM Developer SKILLS NETWORK

FLIGHT NUMBER VS ORBIT TYPE

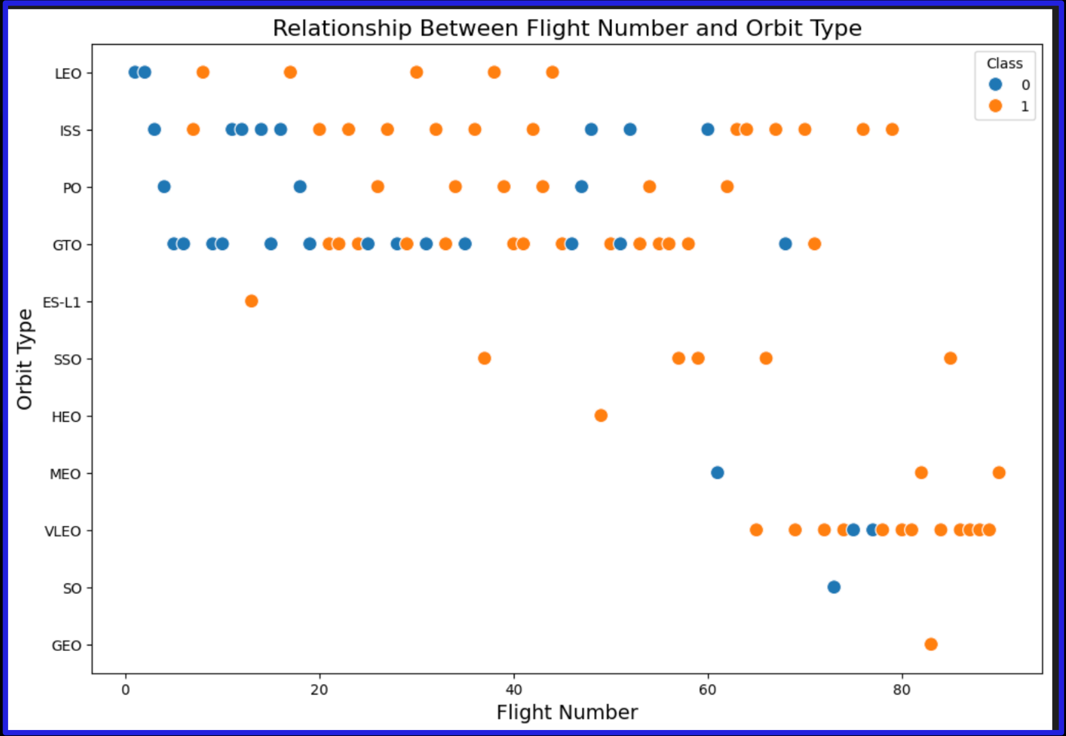
General Trend

- Success rates in LEO orbits show a positive correlation with the number of flights, suggesting improved outcomes as experience accumulates.

- No clear relationship between the number of flights and success rates in GTO orbits

Key Insights

The absence of a pattern in GTO might reflect technical or environmental challenges that affect mission outcomes irrespective of flight experience.

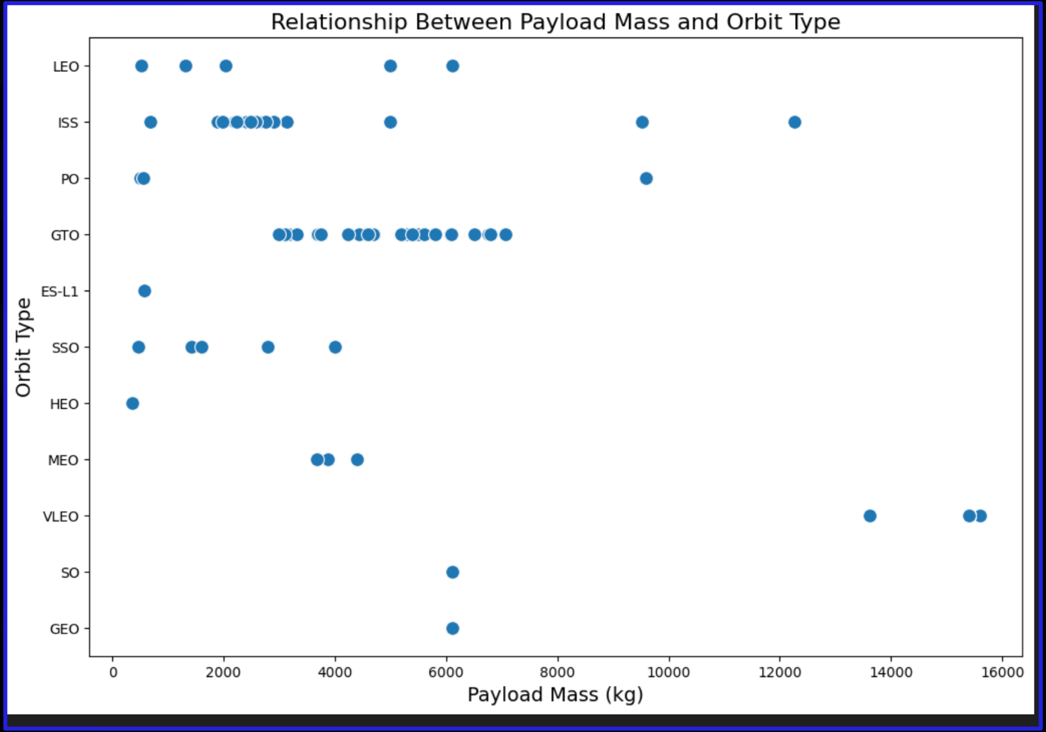


IBM Developer SKILLS NETWORK

PAYLOAD MASS VS ORBIT TYPE

General Trend

- *LEO & Polar Orbits*: Higher payload masses enhance success rates.
- *MEO, GTO, & VLEO Orbits*: Heavier payloads decrease success rates.



Key Insights

Tailor payload strategies to orbit type to maximize mission success.

Necessary to evaluate and optimize payload capacities for specific orbits.

IBM Developer SKILLS NETWORK

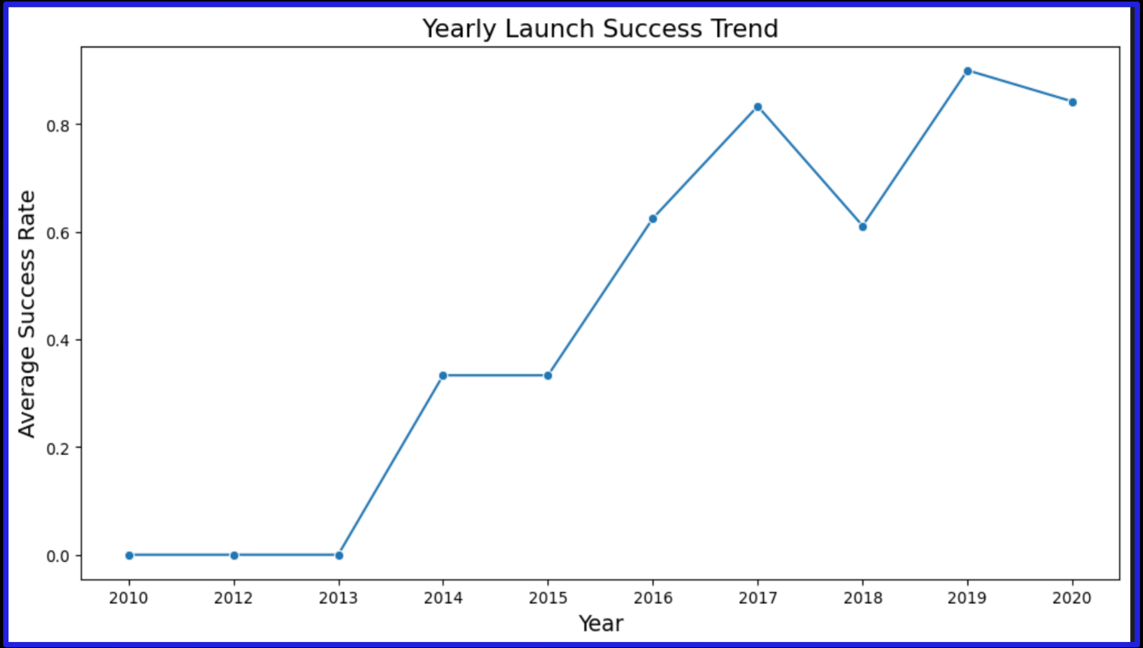
LAUNCH SUCCESS YEARLY TREND

General Trend

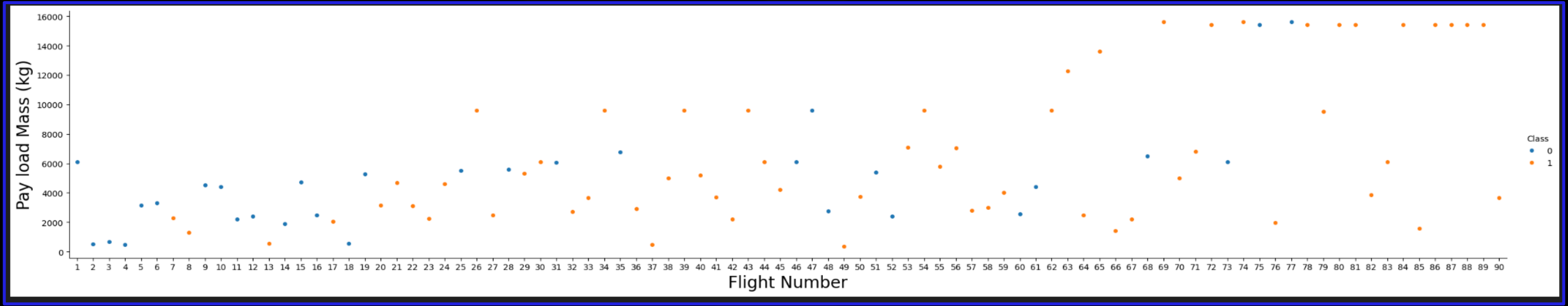
- Success rates have shown a consistent upward trend from 2013 to 2020.
- There is a slight dip in success rates after 2019, suggesting challenges or adjustments in launch operations.

Key Insights

The dip in 2019-2020 calls for a closer look into potential factors affecting launch success to ensure resilience and adaptability in future missions.



PAYLOAD VS. FLIGHT NUMBER



General Trend

- Payload mass generally increases with flight numbers, indicating a trend towards larger payloads in later flights.

- Early flights show a mix of successes and failures with lower payload masses.

Key Insights

The trend indicates a strategic shift towards more ambitious missions with larger payloads as flight experience grows.

Later flights, which carry larger payloads, tend to have higher success rates.

IBM Developer SKILLS NETWORK

EDA WITH SQL

ALL LAUNCH SITE NAMES

```
%sql create table SPACEXTABLE as select * from SPACEXTBL where Date is not
null
```

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

Showing the names of the distinct launch sites involved in the space mission.

IBM Developer
SKILLS NETWORK

LAUNCH SITE NAMES BEGIN WITH 'CCA'

```
query = """
SELECT * FROM SPACEXTBL
WHERE Launch_Site LIKE 'CCA%'
LIMIT 5;
"""
cur.execute(query)
```

```
('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')
```

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

TOTAL PAYLOAD MASS

```
query = """
SELECT SUM(Payload_Mass__kg_) AS Total_Payload_Mass
FROM SPACEXTBL
WHERE Customer LIKE '%NASA (CRS)%';
"""
cur.execute(query)
```

```
Total Payload Mass for NASA (CRS): 48213 kg
```

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

IBM Developer
SKILLS NETWORK

AVERAGE PAYLOAD MASS BY F9 V1.1

```
query = """
SELECT AVG(Payload_Mass__kg_) AS Average_Payload_Mass
FROM SPACEXTBL
WHERE Booster_Version = 'F9 v1.1';
"""
cur.execute(query)
```

```
Average Payload Mass for Booster Version F9 v1.1: 2928.4 kg
```

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

FIRST SUCCESSFUL GROUND LANDING DATE

```
query = """
SELECT MIN(Date) AS First_Successful_Landing
FROM SPACEXTBL
WHERE Landing_Outcome = 'Success (ground pad)';
"""
cur.execute(query)
```

```
Date of the first successful landing on a ground pad: 2015-12-22
```

List the date when the first successful landing outcome in ground pad was achieved.

IBM Developer
SKILLS NETWORK

SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD BETWEEN 4000 AND 6000

```
query = """
SELECT Booster_Version
FROM SPACEXTBL
WHERE Landing_Outcome = 'Success (drone ship)'
        AND Payload_Mass__kg_ > 4000
        AND Payload_Mass__kg_ < 6000;
"""
cur.execute(query)
```

```
Boosters with success in drone ship and payload mass between 4000 and 6000 kg:
F9 FT B1022
F9 FT B1026
F9 FT  B1021.2
F9 FT  B1031.2
```

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

IBM Developer
SKILLS NETWORK

TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

```python
query = """
SELECT
    CASE
        WHEN Mission_Outcome LIKE '%Success%' THEN 'Success'
        WHEN Mission_Outcome LIKE '%Failure%' THEN 'Failure'
        ELSE 'Other'
    END AS Outcome_Type,
    COUNT(*) AS Outcome_Count
FROM SPACEXTBL
GROUP BY Outcome_Type;
"""
cur.execute(query)
```

```
Mission outcomes counts:
Failure: 1
Success: 100
```

List the total number of successful and failure mission outcomes.

BOOSTERS CARRIED MAXIMUM PAYLOAD

```
query = """
SELECT Booster_Version
FROM SPACEXTBL
WHERE Payload_Mass__kg_ = (SELECT MAX(Payload_Mass__kg_) FROM SPACEXTBL);
"""
cur.execute(query)
```

```
Booster versions that carried the maximum payload mass:
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
```

Listing the names of the booster versions which have carried the maximum payload mass.

IBM Developer SKILLS NETWORK

2015 LAUNCH RECORDS

```
query = """
SELECT substr(Date, 6, 2) AS Month, Booster_Version,
Launch_Site, Landing_Outcome
FROM SPACEXTBL
WHERE substr(Date, 0, 5) = '2015'
  AND Landing_Outcome LIKE '%Failure%'
  AND Landing_Outcome LIKE '%drone ship%'
ORDER BY Date;
"""
cur.execute(query)
```

```
Records from 2015 with failure landing outcomes in drone ship:
Month: January, Booster Version: F9 v1.1 B1012, Launch Site:
CCAFS LC-40, Landing Outcome: Failure (drone ship)
Month: April, Booster Version: F9 v1.1 B1015, Launch Site:
CCAFS LC-40, Landing Outcome: Failure (drone ship)
```

Enumerating the failed drone ship landings, along with their booster versions and launch site names for the months in 2015.

IBM Developer
SKILLS NETWORK

RANK SUCCESS COUNT BETWEEN 2010-06-04 AND 2017-03-20

```
query = """
SELECT Landing_Outcome, COUNT(*) AS Outcome_Count
FROM SPACEXTBL
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY Landing_Outcome
ORDER BY Outcome_Count DESC;
"""
cur.execute(query)
```

```
Ranking of landing outcomes between 2010-06-04 and 2017-03-20:
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
```

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.

IBM Developer
SKILLS NETWORK

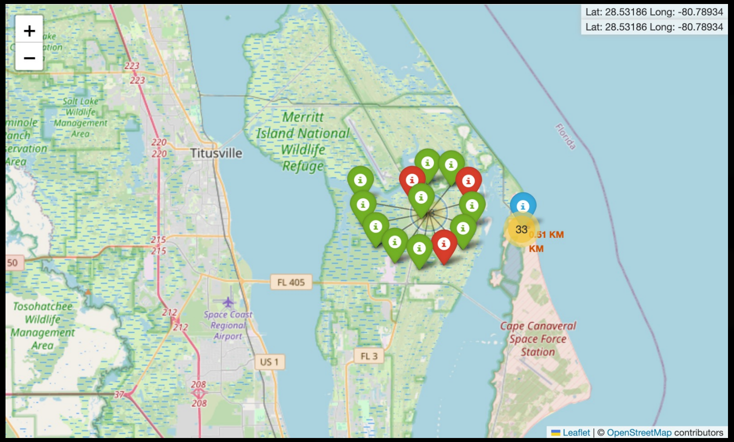
INTERACTIVE MAP WITH FOLIUM

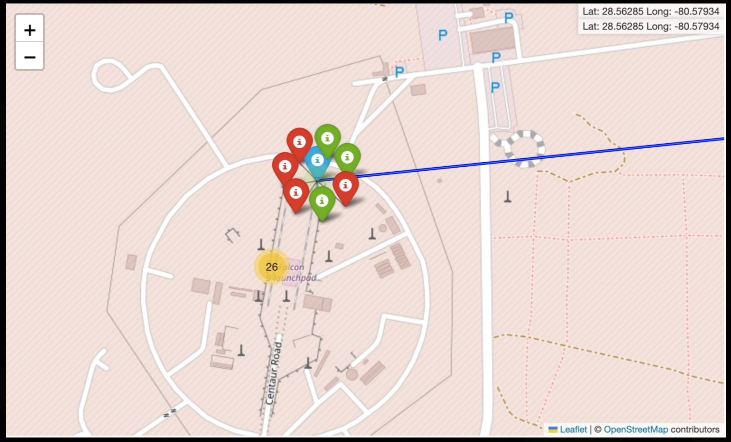
ALL LAUNCH SITES ON FOLIUM MAP

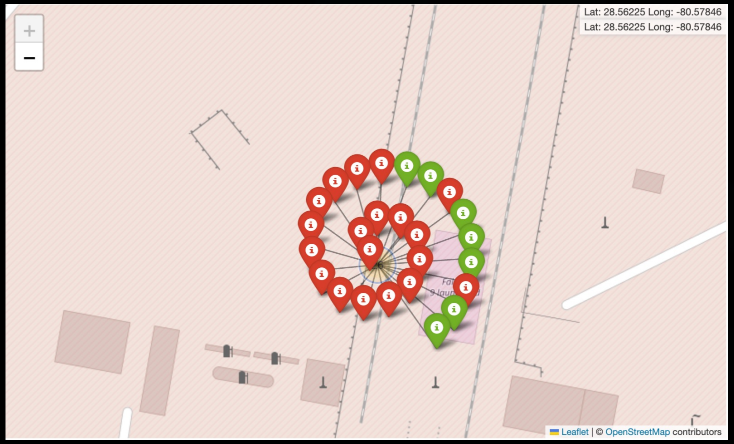
Launch sites are close to the coast to minimize the risk of debris falling or exploding near populated areas.

IBM Developer
SKILLS NETWORK

SUCCESS/FAILED LAUNCHES FOR EACH SITE ON THE MAP







Using the color-coded markers, we can identify which launch sites have relatively high success rates.

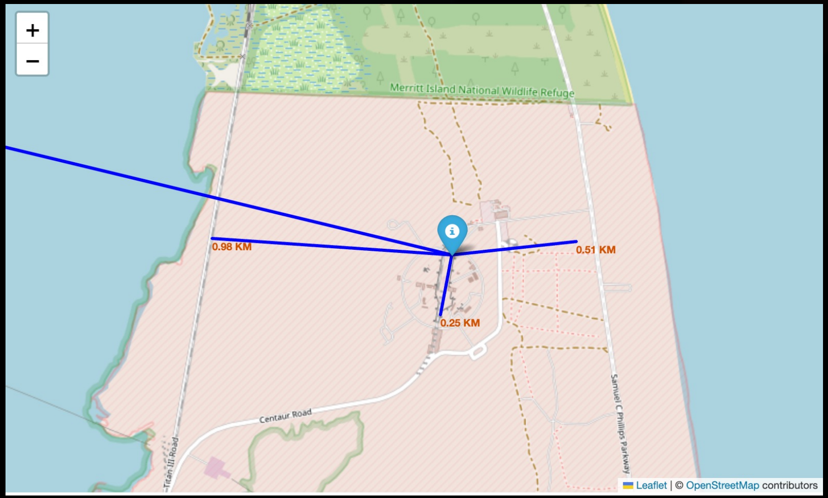
Green Marker = Successful Launch

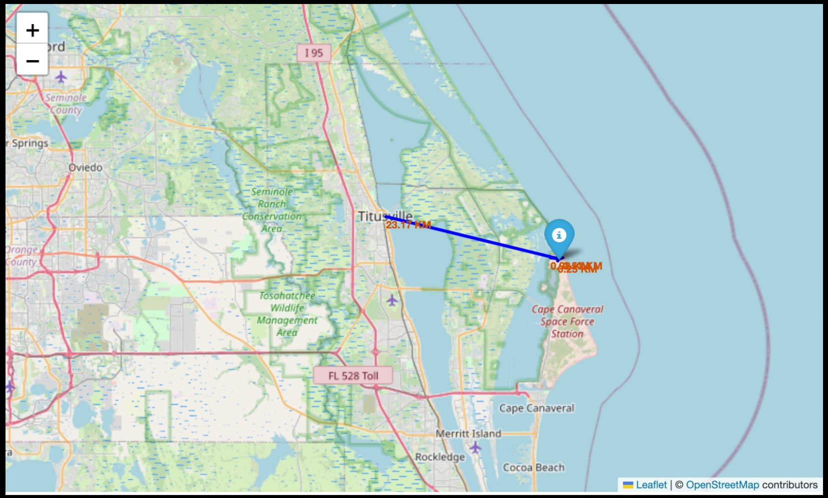
Red Marker = Failed Launch

Launch Site **KSC LC-39A** shows a very high success rate.

IBM Developer SKILLS NETWORK

PROXIMITY TO RELEVANT LOCATIONS





From the analysis of the launch site CCAFS SLC-40, we can determine that:

It is located at a distance of:

0.98 km from the closer railway.
0.51 km from the closer highway.
0.25 km from the closer street.

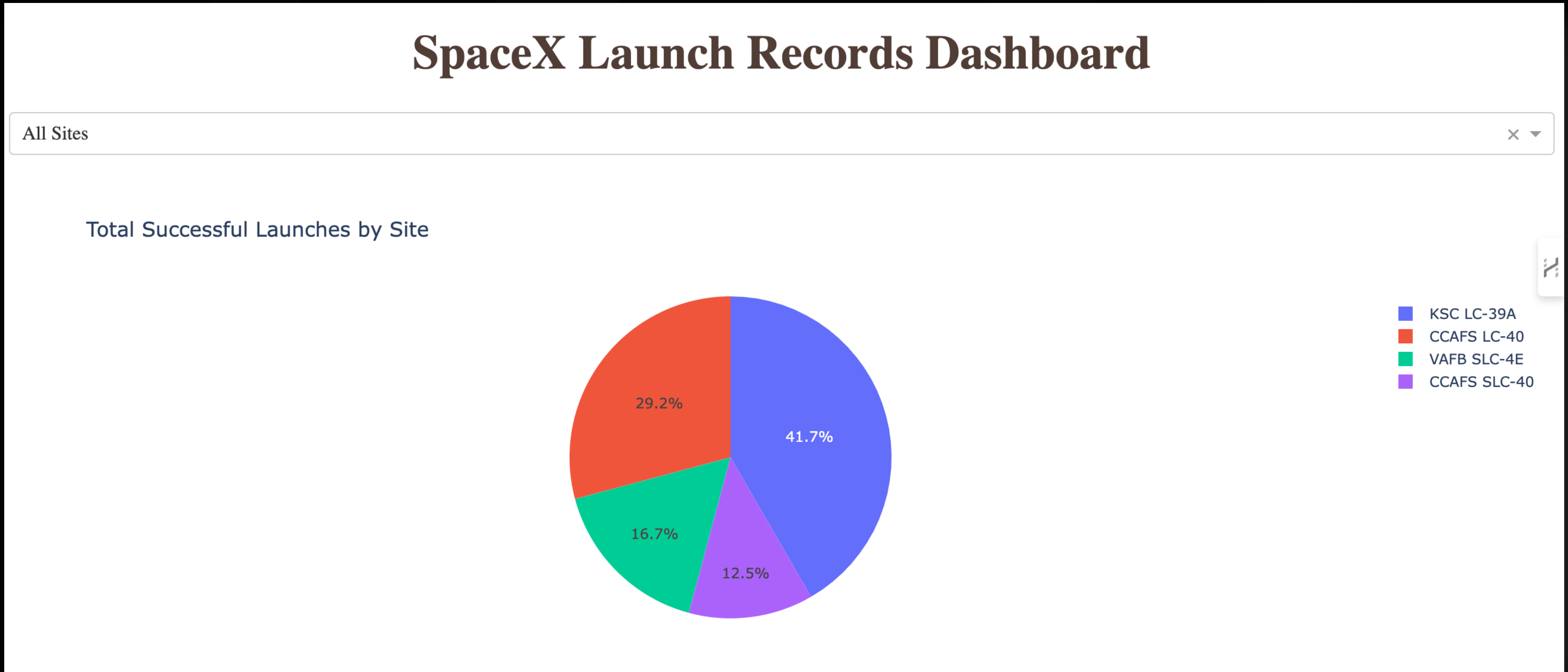
The closest city is Titusville.

IBM Developer SKILLS NETWORK

BUILD A DASHBOARD WITH PLOTLY DASH

IBM Developer
SKILLS NETWORK

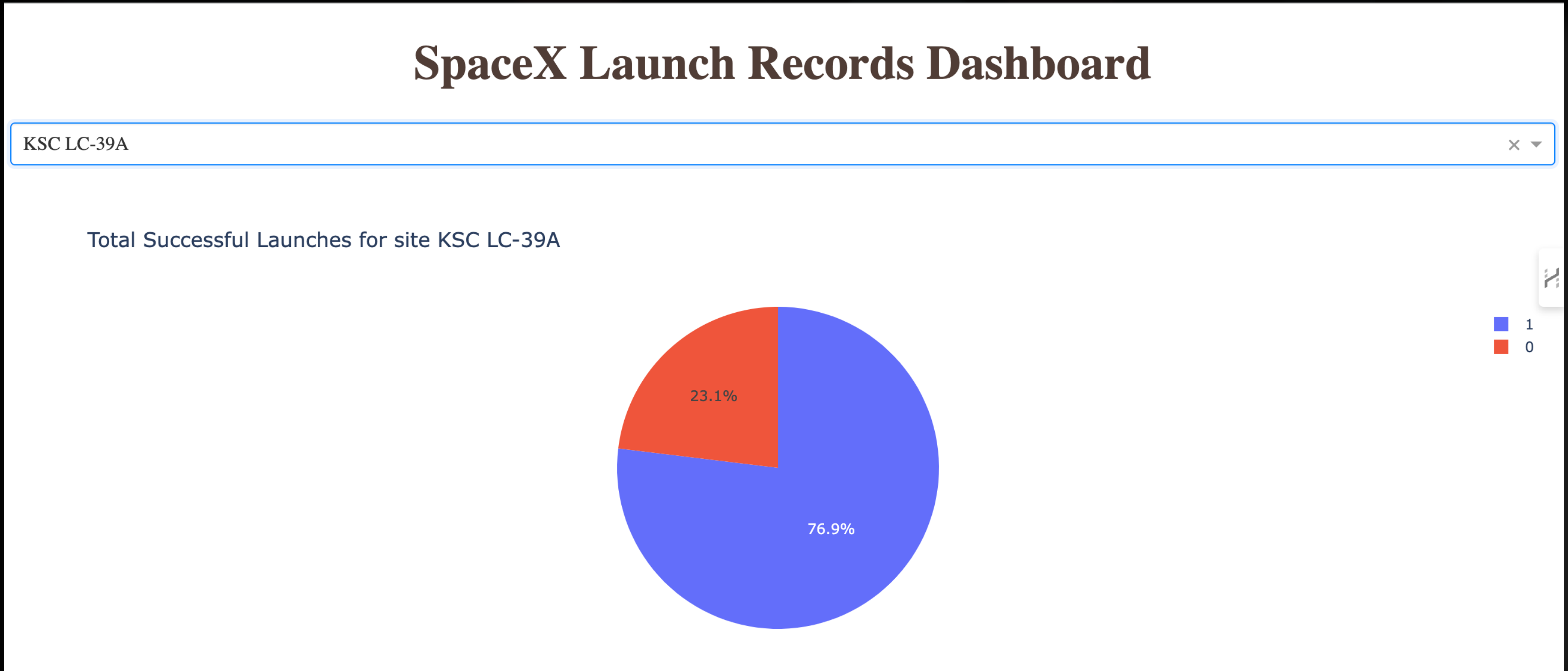
TOTAL SUCCESSFUL LAUNCHES FOR ALL SITES



The chart shows that **KSC LC-39A** has the most successful launches.

IBM Developer SKILLS NETWORK

LAUNCH SITE WITH HIGHEST SUCCESS RATIO



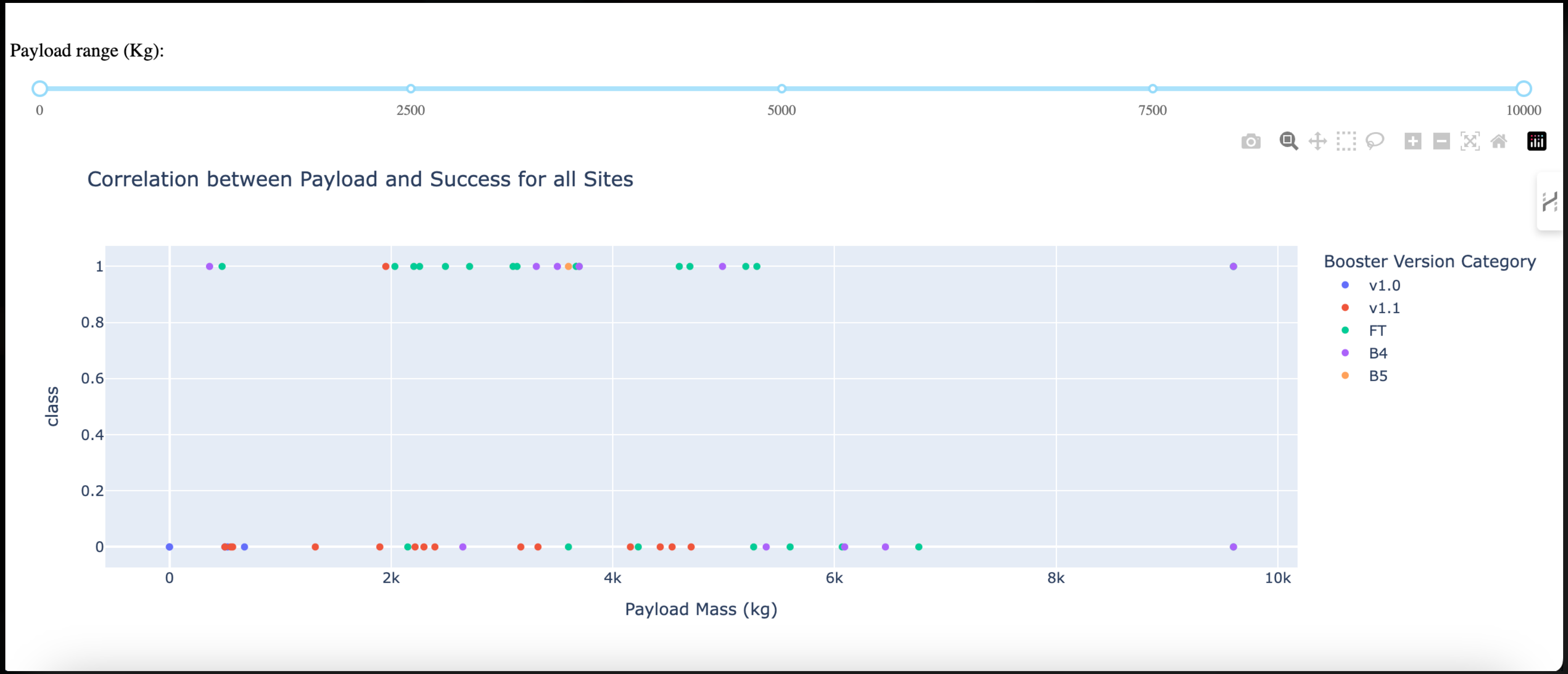
KSC LC-39A achieved a 76.9% success rate and a 23.1% failure rate.

IBM Developer
SKILLS NETWORK

SCATTER PLOT
PAYLOAD MASS VS LAUNCH SUCCESS

Launch Success and Payload Mass

There is a strong correlation between launch success and payload mass.
Most launches with payloads under 5500 kg tend to be successful.



Booster Version Category

There does not appear to be a clear relationship between different booster versions (v1.0, v1.1, FT, B4, B5) and launch success.

IBM Developer SKILLS NETWORK

PREDICTIVE ANALYSIS (CLASSIFICATION)

IBM Developer
SKILLS NETWORK

CLASSIFICATION ACCURACY

The **DECISION TREE CLASSIFIER** is the optimal machine learning algorithm for this task.

```python
print("tuned hpyerparameters :(best parameters) ",tree_cv.
best_params_)
print("accuracy :",tree_cv.best_score_)
```

Python

```
tuned hpyerparameters :(best parameters)  {'criterion': 'entropy',
'max_depth': 6, 'max_features': 'sqrt', 'min_samples_leaf': 2,
'min_samples_split': 5, 'splitter': 'random'}
accuracy : 0.8892857142857142
```

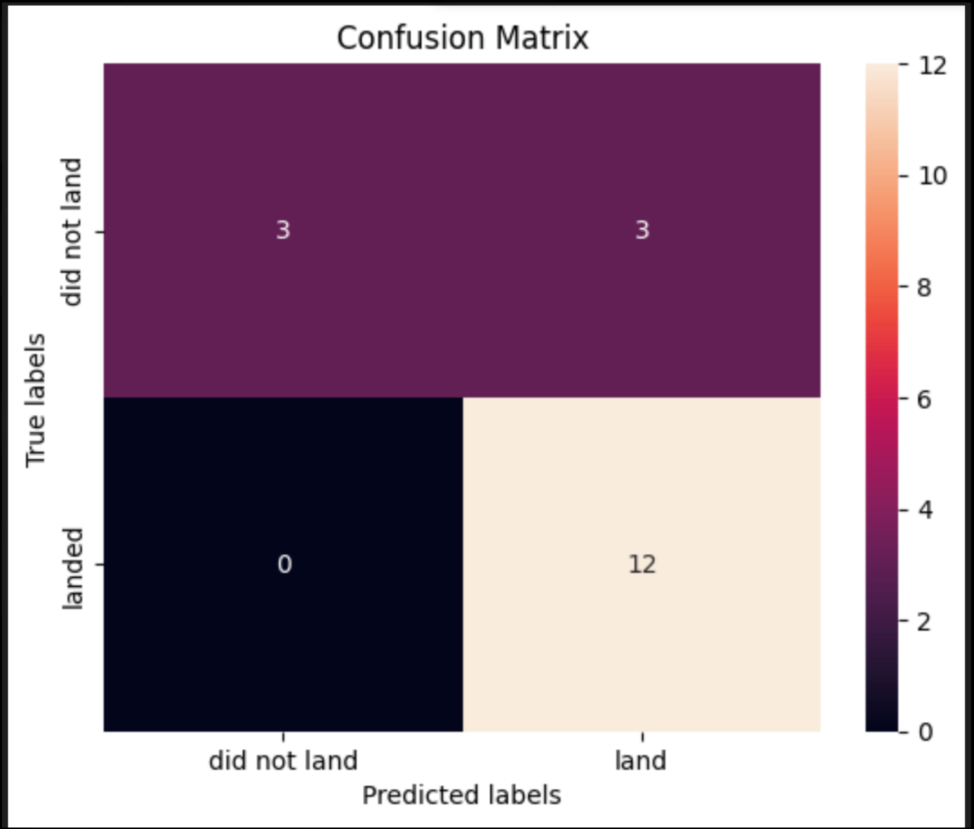
IBM Developer
SKILLS NETWORK

CONFUSION MATRIX

The confusion matrix for the decision tree classifier indicates that the classifier can distinguish between the different classes.

However, the main issue is the false positives, meaning unsuccessful landings are marked as successful landings by the classifier.

Decision Tree



CONCLUSIONS

- ✓ **The DECISION TREE CLASSIFIER** is the best machine learning algorithm for this task.
- ✓ A higher number of flights at a launch site correlates with a higher success rate at that site.
- ✓ Orbits ES-L1, GEO, HEO, SSO, VLEO demonstrated the highest success rates.
- ✓ Launches with smaller payload masses achieved better results compared to those with larger payload masses.
- ✓ KSC LC-39A recorded the highest number of successful launches among all sites.
- ✓ The success rate of launches increases over the years.

THANK YOU!

IBM Developer
SKILLS NETWORK