IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

ELDO MARTADJAJA

November 11, 2023

Outline

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

Executive Summary

Methodology Overview:

- Data Retrieval via API
- Web Scraping for Data Collection
- Data Preprocessing and Wrangling
- Exploratory Data Analysis using SQL
- Exploratory Data Analysis with Data Visualization
- Interactive Data Visualization with Folium
- Machine Learning Predictive Modeling
- Comprehensive Results Summary

A Recap of All Findings:

- Insights from Exploratory Data Analysis
- Highlights from Interactive Analytics (including Screenshots)
- Outcomes from Predictive Analytics
- Conclusions
- Scientific Suggestions

Introduction

Project background and context

- SpaceX tries to distinguish itself by offering reusable Falcon 9 rocket launches at a cost of 62 million dollars, a substantial contrast to the 165 million dollars charged by other providers, with a record loss of just 1 payload. A key source of these cost savings is SpaceX's capability to reuse the first rocket stage. Predicting the successful landing of this first stage can have a significant impact on launch cost estimation, providing valuable insights for potential competitors in the rocket launcher industry.
- In this project I will aim to construct a machine learning pipeline forecasting the successful landing of the reusable first stage rocketry.

Questions to answer by this project

- What factors that impact the successful landing of reusable first stage rockets?
- Is the achievement of a successful landing very much dependent on intricate various characteristics at play?
- What operational conditions are crucial for ensuring a successful landing sequence?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

How data sets were collected:

- Setting Objectives: Establish precise objectives for data utilization.
- Identifying Sources: Determine the locations where essential data can be retrieved, such as databases or APIs like SpaceX.
- Data Acquisition: Retrieve data from the identified sources, which may entail techniques like web scraping or making API requests.
- Data Refinement: Refine the collected data using libraries like Pandas and Numpy to eliminate duplicates, manage missing values, or transform data formats.
- Data Preservation: Archive the refined data in a database or file system, potentially employing SQL for relational databases.
- Data Examination: Examine the stored and refined data to extract valuable insights.

Data Collection – SpaceX API

- Performed a GET request to the SpaceX API for data retrieval, followed by data cleansing and basic data manipulation and formatting.

- https://github.com/eldoma/IBM-SpaceX

Task 1: Request and parse the SpaceX launch data using the GET request

To make the requested JSON results more consistent, we will use the following static response object for this project:

```
[11]: static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNet
```

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

```
[12]: response.status_code
[12]: 200
```

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_norm`

```
[ ]: # Use json_normalize meethod to convert the json result into a dataframe
     # Eldo Writes
     import requests
     import json
     import pandas as pd

     # request the SpaceX launch data
     res = requests.get(static_json_url)
     print(res.content)
```

Using the dataframe `data` print the first 5 rows

```
[25]: static_json_df = res.json()
      # apply json_normalize
      data = pd.json_normalize(static_json_df)
      # Get the head of the dataframe
      data.head(5)
```

[25]:

	static_fire_date_utc	static_fire_date_unix	tbd	net	window	rocket	success	details	crew
0	2006-03-17T00:00:00.000Z	1.142554e+09	False	False	0.0	5e9d0d95eda69955f709d1eb	False	Engine failure at 33 seconds and loss of vehicle	[]

Data Collection - Scraping

- Web scraping Falcon 9 launch records with BeautifulSoup

- https://github.com/eldoma/IBM-SpaceX

```python
            # Payload
            payload = row[3].a.string if row[3].a else "N/A"
            launch_dict['Payload'] = payload

            # Payload Mass
            payload_mass = get_mass(row[4])
            launch_dict['Payload mass'] = payload_mass

            # Orbit
            orbit = row[5].a.string if row[5].a else "N/A"
            launch_dict['Orbit'] = orbit

            # Customer
            customer = row[6].a.string if row[6].a else "N/A"
            launch_dict['Customer'] = customer

            # Launch outcome
            launch_outcome = list(row[7].strings)[0] if row[7].strings else "N/A"
            launch_dict['Launch outcome'] = launch_outcome

            # Booster landing
            booster_landing = landing_status(row[8])
            launch_dict['Booster landing'] = booster_landing

            # Append the launch_dict to the launch_list
            launch_list.append(launch_dict)

# Print the first few launch dictionaries as a sample
for launch in launch_list[:5]:
    print(launch)
```

```
{'Flight No.': '78', 'Date': '7 January 2020', 'Time': '02:19:21', 'Version Booster': 'F9 B5', 'Launch Site': 'CCSFS', 'Payl
Customer': 'SpaceX', 'Launch outcome': 'Success\n', 'Booster landing': 'Success'}
{'Flight No.': '79', 'Date': '19 January 2020', 'Time': '15:30', 'Version Booster': 'F9 B5', 'Launch Site': 'KSC', 'Payload'
', 'Orbit': 'Sub-orbital', 'Customer': 'NASA', 'Launch outcome': 'Success\n', 'Booster landing': 'No attempt\n'}
{'Flight No.': '80', 'Date': '29 January 2020', 'Time': '14:07', 'Version Booster': 'F9 B5', 'Launch Site': 'CCSFS', 'Payloa
stomer': 'SpaceX', 'Launch outcome': 'Success\n', 'Booster landing': 'Success'}
{'Flight No.': '81', 'Date': '17 February 2020', 'Time': '15:05', 'Version Booster': 'F9 B5', 'Launch Site': 'CCSFS', 'Paylo
ustomer': 'SpaceX', 'Launch outcome': 'Success\n', 'Booster landing': 'Failure'}
{'Flight No.': '82', 'Date': '7 March 2020', 'Time': '04:50', 'Version Booster': 'F9 B5', 'Launch Site': 'CCSFS', 'Payload':
ustomer': 'NASA', 'Launch outcome': 'Success\n', 'Booster landing': 'Success'}
```

After you have fill in the parsed launch record values into `launch_dict`, you can create a dataframe from it.

[21]:
```python
df= pd.DataFrame({ key:pd.Series(value) for key, value in launch_dict.items() })
```

We can now export it to a CSV for the next section, but to make the answers consistent and in case you have difficulties finishing this lab.

Following labs will be using a provided dataset to make each lab independent.

`df.to_csv('spacex_web_scraped.csv', index=False)`

Data Cleaned & Exported to CSV for Analysis

2020 [edit]

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[490] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[491]

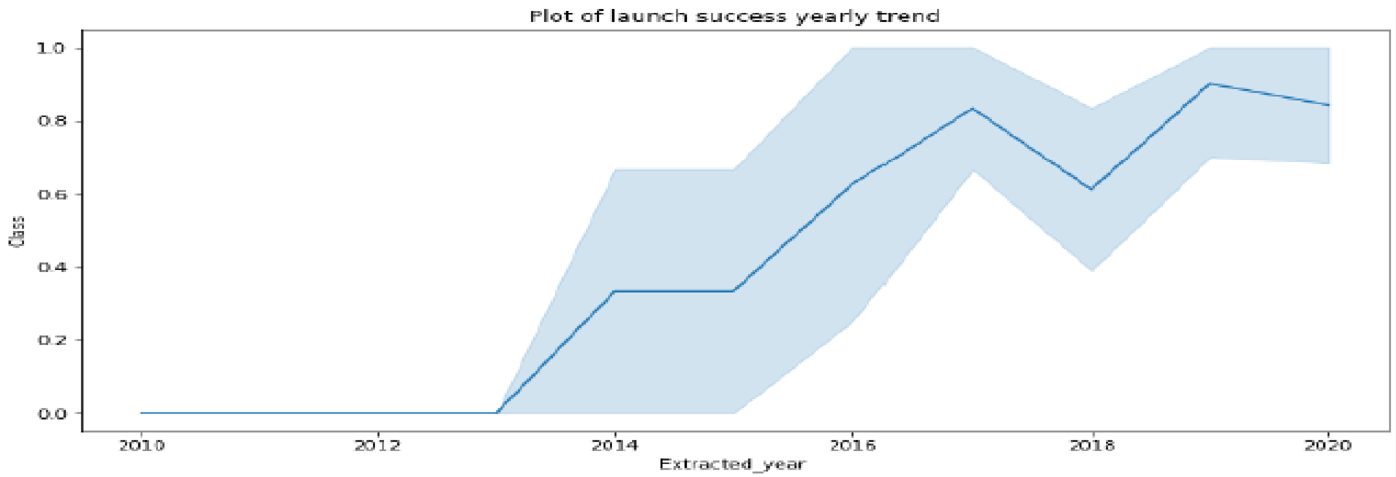
[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 ♺ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 ♺ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule,[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[501]	F9 B5 ♺ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[502]								
81	17 February 2020, 15:05[503]	F9 B5 ♺ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 ♺ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 ♺)	1,977 kg (4,359 lb)[507]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries Bartolomeo, an ESA platform for hosting external payloads onto ISS.[508] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 ♺ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the payload still reached the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 ♺ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

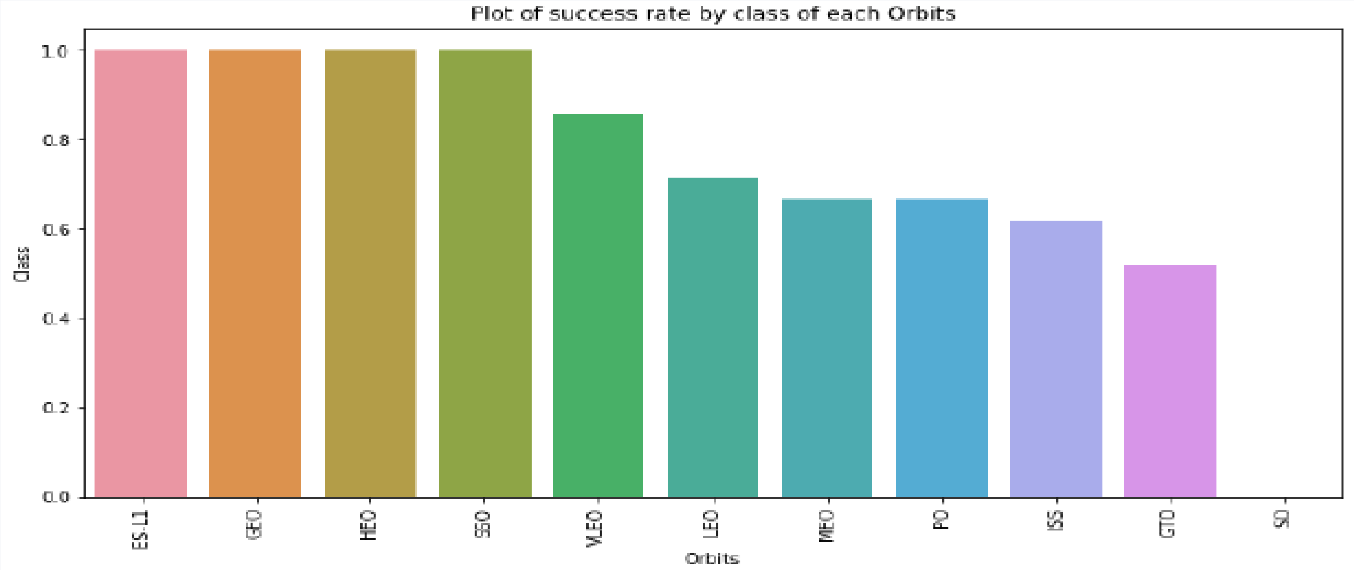
Data Wrangling

- Performed an exploratory analysis of the data and determined the training labels. Calculated the number of launches at each site and analyzed the frequency and count of each orbit. Generated the landing outcome label from the outcome column and stored the findings in a CSV file.
- https://github.com/eldoma/IBM-SpaceX

EDA with Data Visualization

- Visualizing data by exploring the correlation between flight number and launch site, payload and launch site, the success rate of each orbit type, flight number, and orbit type, as well as the annual trend in launch success.
- https://github.com/eldoma/IBM-SpaceX



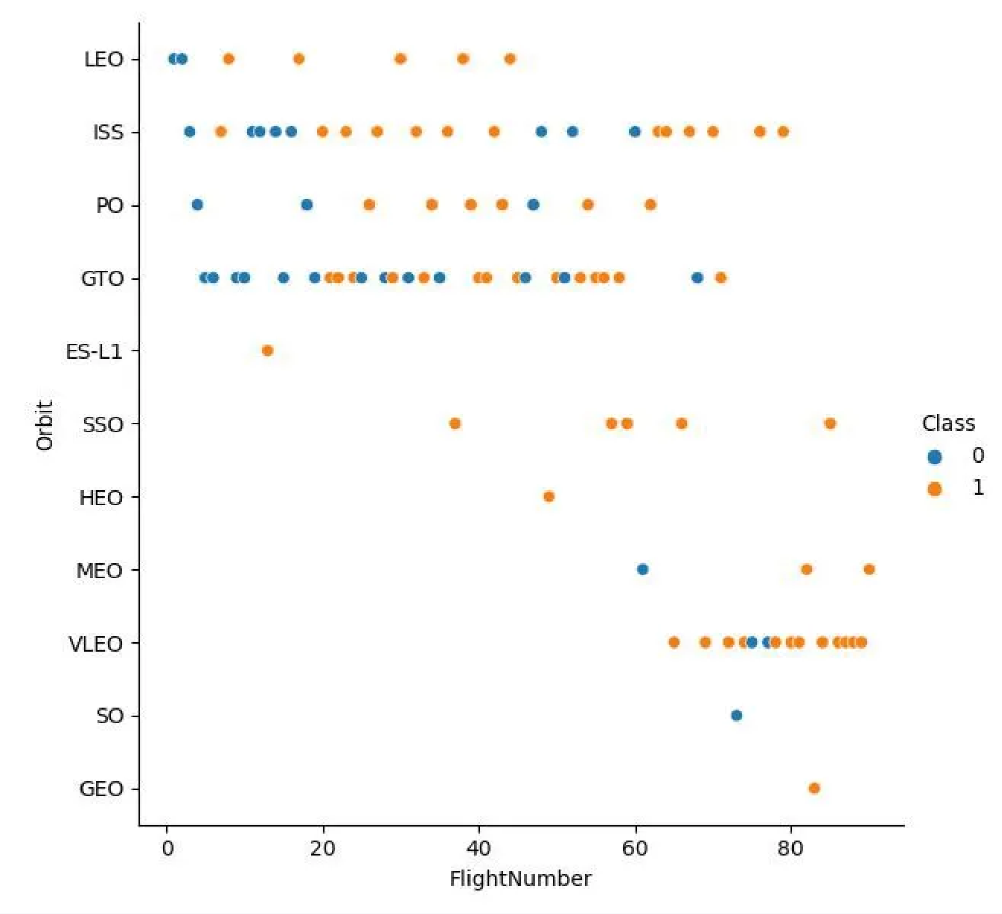


EDA with Data Visualization

- We utilize Python's Matplotlib and Seaborn libraries to create visual representations of the dataset, allowing us to explore the relationships present within it.

- Visualizations pertains to the success rate in each orbit:
  Class 1 corresponds to "Success.“
  Class 0 corresponds to "Failure.“

  https://github.com/eldoma/IBM-SpaceX



EDA with SQL

- Identifying the unique launch sites in space missions.
- Calculating the total payload mass of NASA's (CRS) launched boosters.
- Determining the average payload mass of booster version F9 v1.1.
- Counting the total number of successful and failed mission outcomes.
- Identifying the failed landing outcomes on drone ships and collecting data on their booster versions and launch site names.

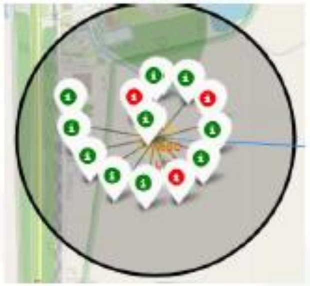
https://github.com/eldoma/IBM-SpaceX

Interactive Map with Folium: Launch Success / Failure Locations

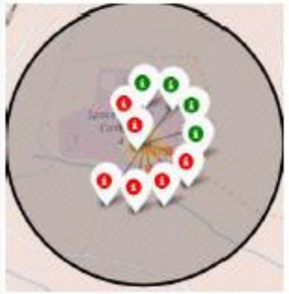
RED = Failure GREEN = Success

CCAFS LC-40

CCAFS SLC-40



KSC LC-39A

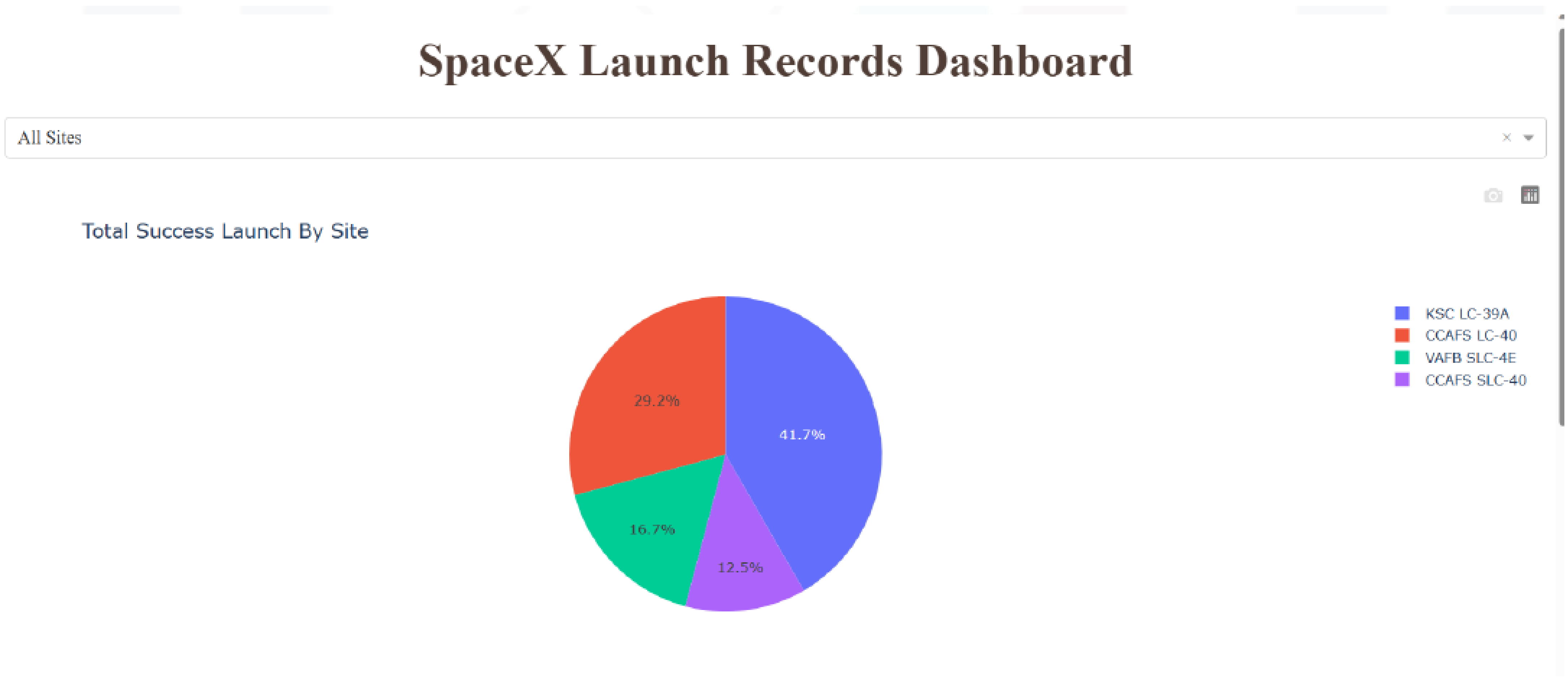


VAFB SLC-4E

Dashboard with Plotly Dash

- We loaded the data using NumPy and Pandas, performed data transformation, and then divided our data into training and testing sets. Additionally, we constructed various machine learning models and fine-tuned their hyperparameters using GridSearchCV.
- Finally, we created scatter plots to visualize the relationship between the outcome and payload mass (in kilograms) for different booster versions.
- https://github.com/eldoma/IBM-SpaceX

Launch Success Probabilities



Predictive Analysis (Classification)

- Create NumPy array from the Class column
- Standardize the data with StandardScaler.
- Fit and transform the data.
- Split the data using train_test_split
- Create a GridSearchCV object with cv=10 for parameter optimization
- Apply GridSearchCV on different algorithms: logistic regression
- (LogisticRegression()), support vector machine (SVC()), decision tree
- (DecisionTreeClassifier()), K Nearest Neighbor (KNeighborsClassifier())

Results

Exploratory Data Analysis:

- Over time, there is a significant noticeable improvement in launch successes.
- Among landing sites, KSC LC 39A stands out with the highest success rate.
- Orbits ES-L1, GEO, HEO, and SSO have a perfect 100% success rate.

Visual Analytics:

- Most launch sites located near the equator, closest to desired orbits, situated along the coast.
- Launch sites are strategically positioned, ensuring that a failed launch would not cause collateral damage to cities, highways, or railways, yet they remain accessible for logistical support.

Predictive Analytics:

- The Decision Tree model emerges as the most effective predictive model for this dataset.

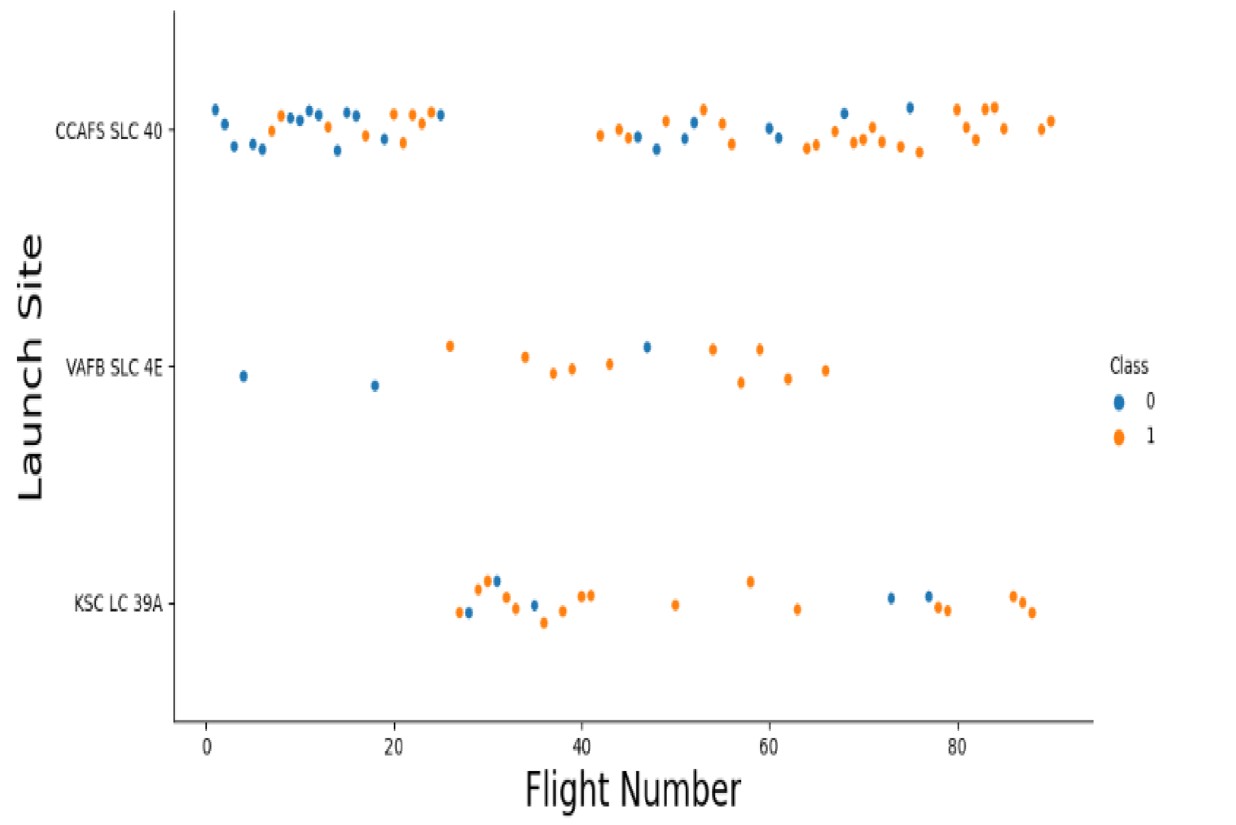
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

Exploratory Data Analysis insights:

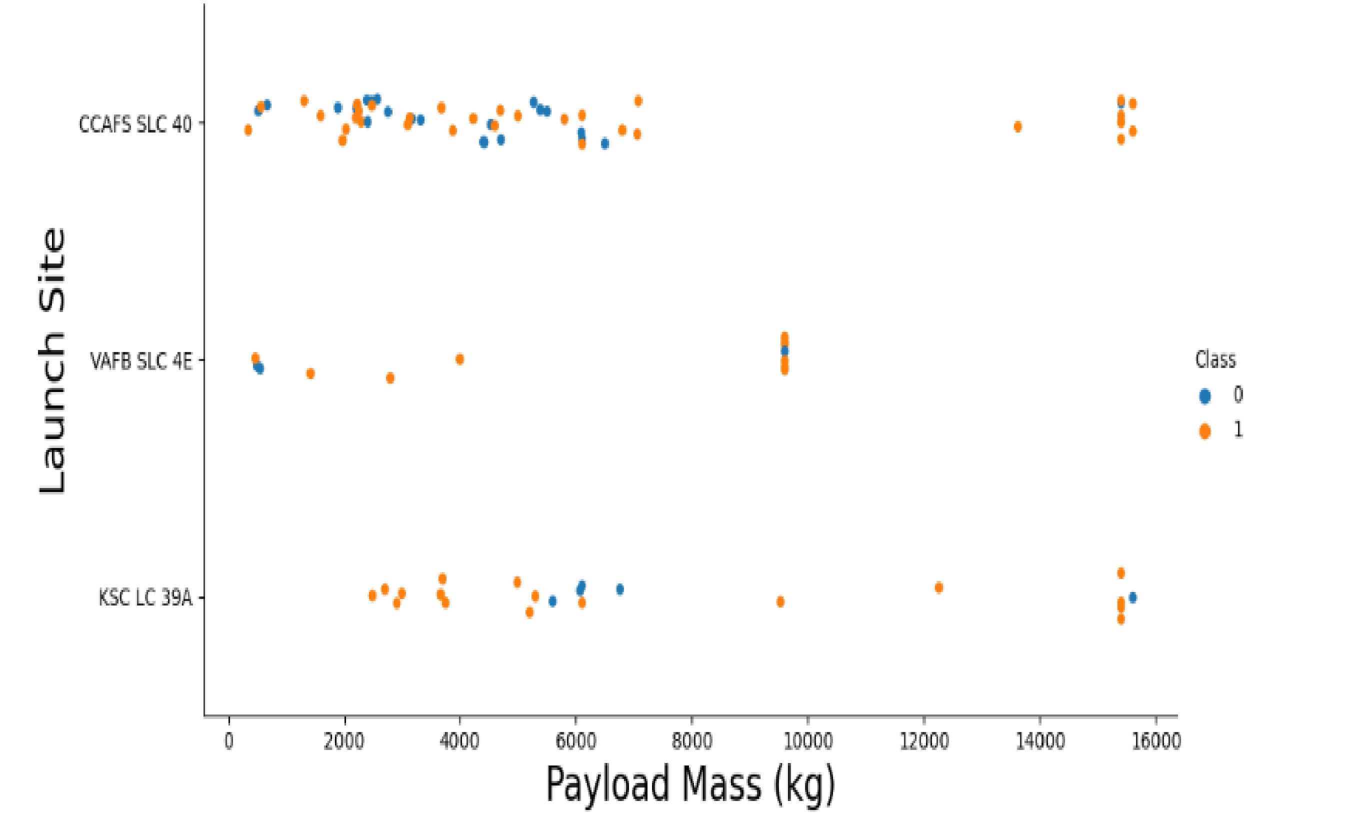
- Success rates for earlier flights were lower (indicated by blue), while success rates for later flights were higher (indicated by orange).
- Approximately half of the launches originated from the CCAFS SLC 40 launch site.
- Both VAFB SLC 4E and KSC LC 39A have notably higher success rates.
- It can be inferred that newer launches tend to have higher success rate.



Payload vs. Launch Site

Exploratory Data Analysis insights:

- Generally, there is a positive correlation between payload mass (in kg) and success rate.
- Most launches with a payload exceeding 7,000 kg resulted in success.
- KSC LC 39A achieved a 100% success rate for launches with payloads less than 5,500 kg.
- Notably, VAFB SLC 4E has not launched payloads greater than approximately 10,000 kg.

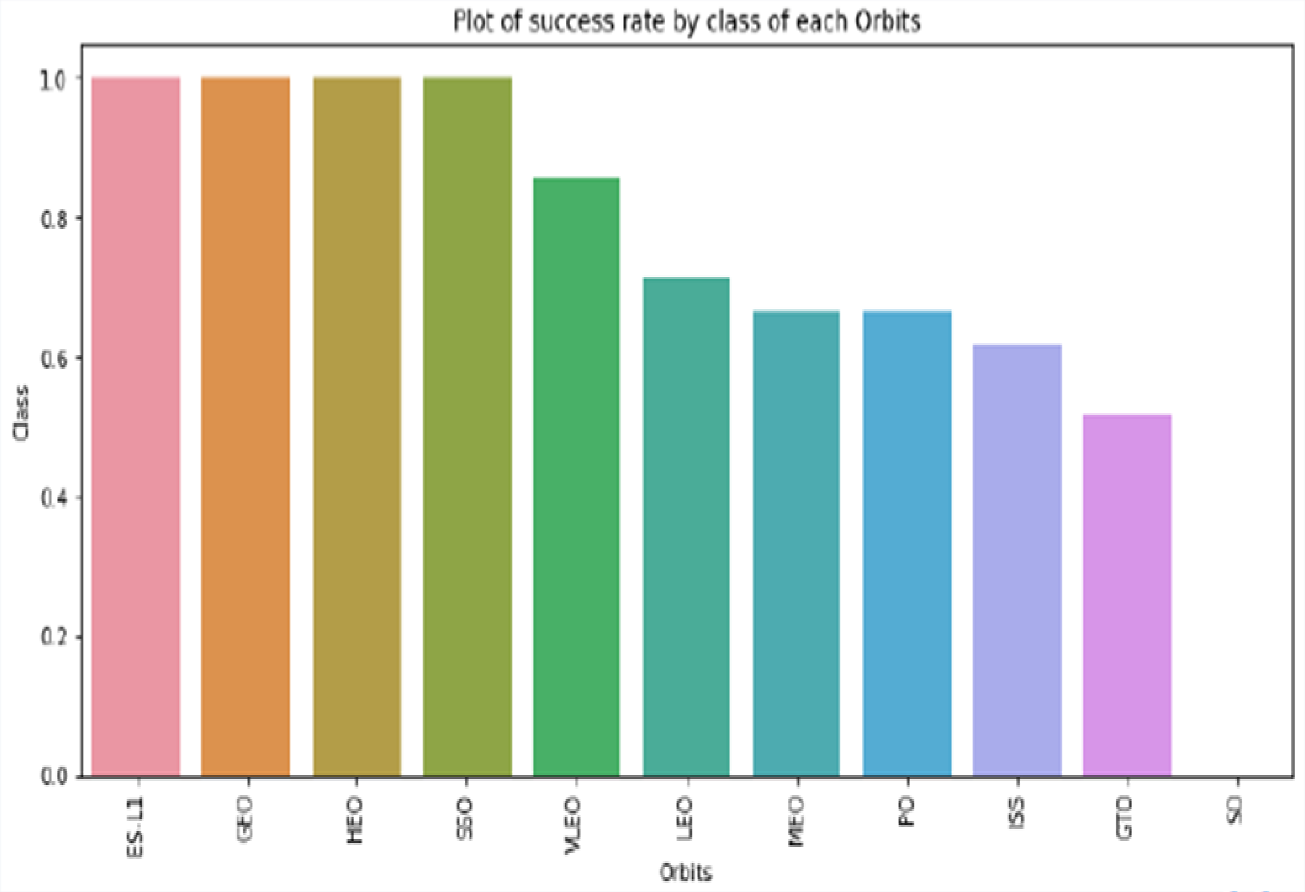


Success Rate vs. Orbit Type

There are several types of Earth orbit, and each offers certain advantages and capabilities: LEO, MEO, GEO, GSO, Polar, SSO and HEO.

Orbit Type Success insights

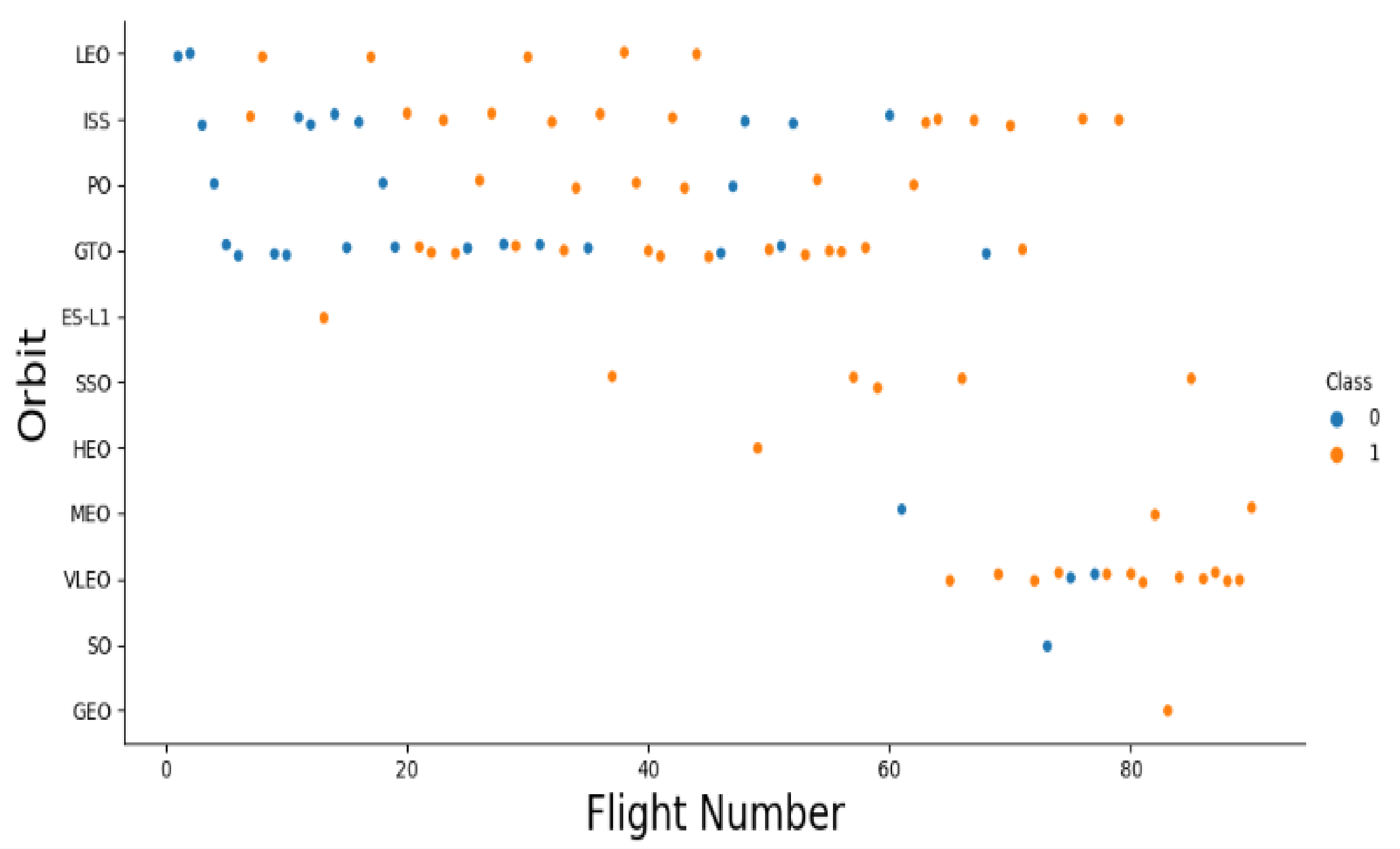
- 100% Success Rate :ES L1, GEO, HEO and SSO
- 50%- 80% Success Rate : GTO, ISS, LEO, MEO, PO
- 0% Success Rate : SO



Flight Number vs. Orbit Type

Exploratory Data Analysis insights

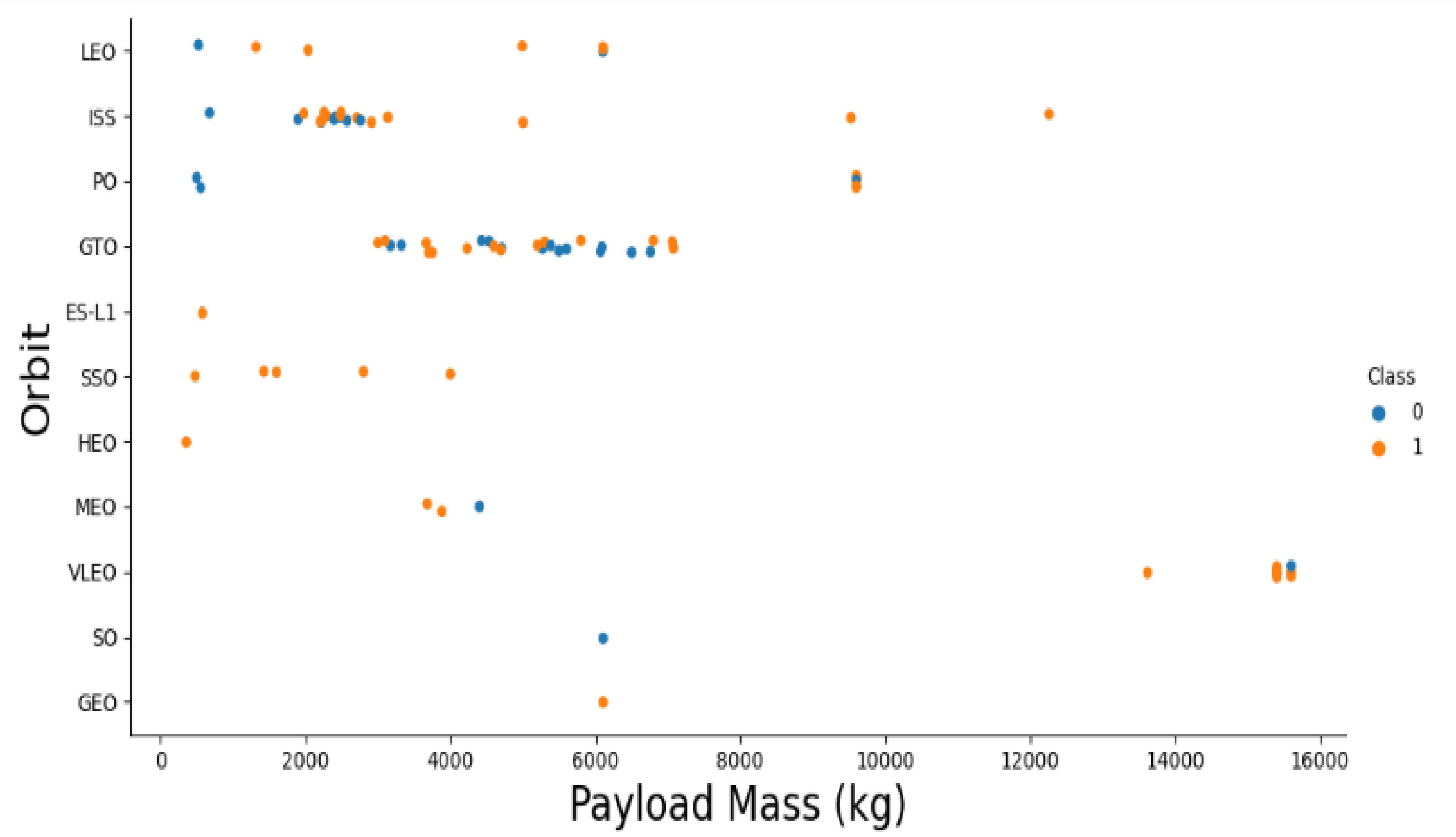
- The success rate generally rises with the number of flights for each orbit.
- This pattern is especially evident for the Low Earth Orbit (LEO).
- However, the Geostationary Transfer Orbit (GTO) does not exhibit the same trend.



Payload vs. Orbit Type

Exploratory Data Analysis insights

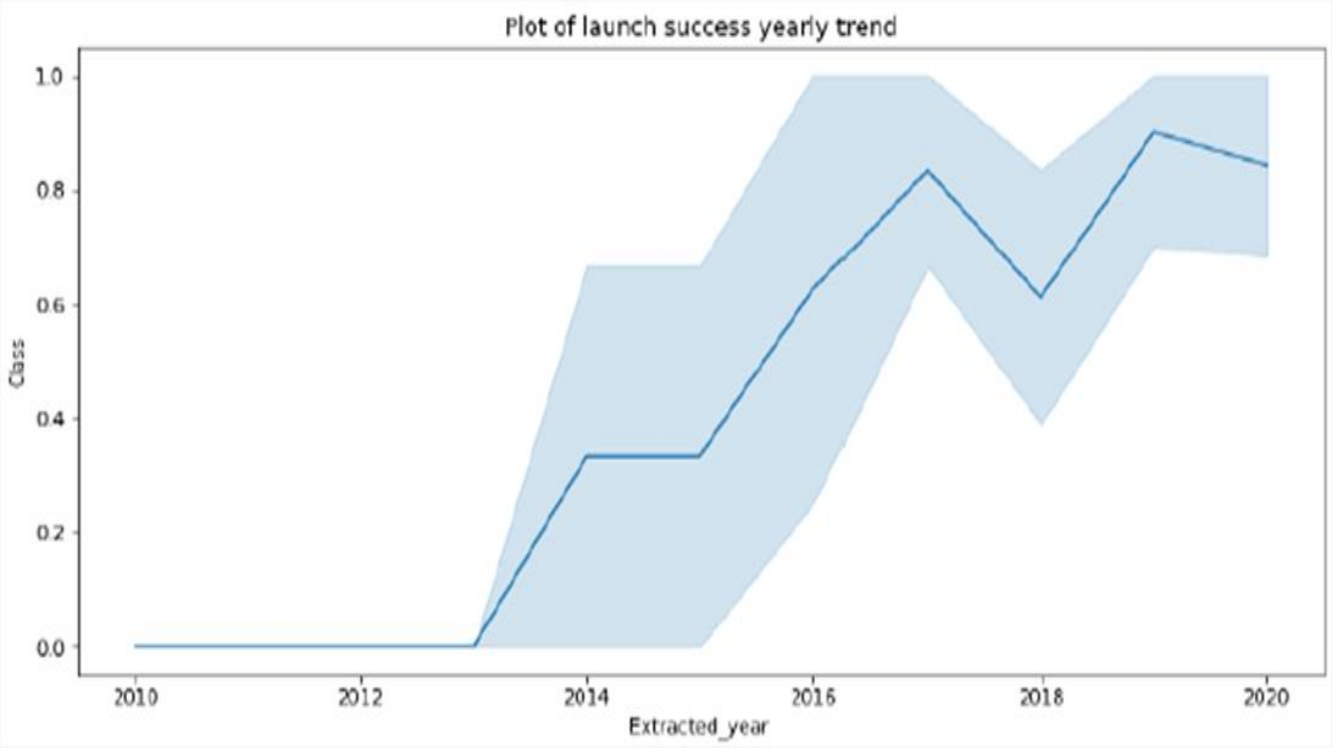
- For heavy payloads, the Low Earth Orbit (LEO), International Space Station (ISS), and Polar Orbit (PO) are more suitable, the successful landing are better for PO, LEO and ISS orbits.
- In contrast, the Geostationary Transfer Orbit (GTO) shows varying success rates with heavier payloads.



Launch Success Yearly Trend

Exploratory Data Analysis insights

- During the years 2013 to 2017 and again from 2018 to 2019, there was an improvement in the success rate.
- However, there was a fall in the success rate from 2017 to 2018 and from 2019 to 2020.
- In summary, the success rate has shown an overall improvement since 2013.



All Launch Site Names

Query using DISTINCT to display Launch Site names

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

Records with Launch Site Starting with CCA

Displaying 5 records below

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

In [10]:
```
task_1 = '''
        SELECT DISTINCT LaunchSite
        FROM SpaceX
'''
create_pandas_df(task_1, database=conn)
```

Out[10]:

	launchsite
0	KSC LC-39A
1	CCAFS LC-40
2	CCAFS SLC-40
3	VAFB SLC-4E

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

In [11]:
```
task_2 = '''
        SELECT *
        FROM SpaceX
        WHERE LaunchSite LIKE 'CCA%'
        LIMIT 5
        '''
create_pandas_df(task_2, database=conn)
```

Out[11]:

	date	time	boosterversion	launchsite	payload	payloadmasskg	orbit	customer	missionoutcome	landingoutcome
0	2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	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
3	2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Launch Site Names Begin with 'CCA'

- Query using ‘Unique’ to display 5 records where launch sites names begin with ‘CCA’

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

In [11]:

```
task_2 = '''
        SELECT *
        FROM SpaceX
        WHERE LaunchSite LIKE 'CCA%'
        LIMIT 5
        '''
create_pandas_df(task_2, database=conn)
```

Out[11]:

	date	time	boosterversion	launchsite	payload	payloadmasskg	orbit	customer	missionoutcome	landingoutcome
0	2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-08-12	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	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
3	2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	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

- Query displaying 45,596 kg total Payload mass carried by boosters launched by NASA

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

In [12]:
```
task_3 = '''
        SELECT SUM(PayloadMassKG) AS Total_PayloadMass
        FROM SpaceX
        WHERE Customer LIKE 'NASA (CRS)'
        '''
create_pandas_df(task_3, database=conn)
```

Out[12]:

	total_payloadmass
0	45596

Average Payload Mass by F9 v1.1

- Query displaying 2,928 kg (average payload mass) carried by Falcon booster version F9 v1.1

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

In [13]:

```
task_4 = '''
        SELECT AVG(PayloadMassKG) AS Avg_PayloadMass
        FROM SpaceX
        WHERE BoosterVersion = 'F9 v1.1'
        '''
create_pandas_df(task_4, database=conn)
```

Out[13]:

	avg_payloadmass
0	2928.4

First Successful Ground Landing Date

- Query to display dates of the first successful landing outcome on ground pad, which shows December 22nd, 2015.
- The maiden flight of the Falcon 9 Full Thrust version, on the evening of December 21, 2015, a successful landing which the first stage rocket was fully recovered the following day.

In [14]:
```
task_5 = '''
        SELECT MIN(Date) AS FirstSuccessfull_landing_date
        FROM SpaceX
        WHERE LandingOutcome LIKE 'Success (ground pad)'
        '''
create_pandas_df(task_5, database=conn)
```

Out[14]:

	firstsuccessfull_landing_date
0	2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Query reveals names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are:
  – F9 FT B1022, F9 FT B1026, F9 FT B1021.2, F9 FT B1031.2
- Payloads carried: JSCAT-14, JSCAT-16, SES-10, SES-11 / EchoStar 105

```
%sql SELECT PAYLOAD \
FROM SPACEXTBL \
WHERE LANDING__OUTCOME = 'Success (drone ship)' \
AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```

```
 * ibm_db_sa://yyy33800:***@1bbf73c5-d84a-4bb0-85b9
   sqlite:///my_data1.db
Done.
```

payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

```
In [15]: task_6 = '''
                 SELECT BoosterVersion
                 FROM SpaceX
                 WHERE LandingOutcome = 'Success (drone ship)'
                     AND PayloadMassKG > 4000
                     AND PayloadMassKG < 6000
                 '''
         create_pandas_df(task_6, database=conn)
```

Out[15]:

	boosterversion
0	F9 FT B1022
1	F9 FT B1026
2	F9 FT B1021.2
3	F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Query using wildcard like '%' to filter for WHERE MissionOutcome
- So far, one rocket and its payload were destroyed on the launch pad during the fueling process before a static fire test was set to occur:
  SpaceX Falcon 9 rocket catastrophic explosion destroys the rocket & Amos-6 Israeli satellite payload at launch pad 40 at Cape Canaveral Air Force Station, FL, on Sept. 1, 2016. A static hot fire test was planned ahead of scheduled launch on Sept. 3, 2016. Credit: USLaunchReport

List the total number of successful and failure mission outcomes

In [16]:

```
task_7a = '''
        SELECT COUNT(MissionOutcome) AS SuccessOutcome
        FROM SpaceX
        WHERE MissionOutcome LIKE 'Success%'
        '''

task_7b = '''
        SELECT COUNT(MissionOutcome) AS FailureOutcome
        FROM SpaceX
        WHERE MissionOutcome LIKE 'Failure%'
        '''
print('The total number of successful mission outcome is:')
display(create_pandas_df(task_7a, database=conn))
print()
print('The total number of failed mission outcome is:')
create_pandas_df(task_7b, database=conn)
```

```
The total number of successful mission outcome is:
```

	successoutcome
0	100

```
The total number of failed mission outcome is:
```

Out[16]:

	failureoutcome
0	1

Boosters Carried Maximum Payload

A total of twelve (12) F9 boosters which have 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

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

In [17]:

```
task_8 = '''
        SELECT BoosterVersion, PayloadMassKG
        FROM SpaceX
        WHERE PayloadMassKG = (
                                SELECT MAX(PayloadMassKG)
                                FROM SpaceX
                                )
        ORDER BY BoosterVersion
        '''
create_pandas_df(task_8, database=conn)
```

Out[17]:

	boosterversion	payloadmasskg
0	F9 B5 B1048.4	15600
1	F9 B5 B1048.5	15600
2	F9 B5 B1049.4	15600
3	F9 B5 B1049.5	15600
4	F9 B5 B1049.7	15600
5	F9 B5 B1051.3	15600
6	F9 B5 B1051.4	15600
7	F9 B5 B1051.6	15600
8	F9 B5 B1056.4	15600
9	F9 B5 B1058.3	15600
10	F9 B5 B1060.2	15600
11	F9 B5 B1060.3	15600

2015 Launch Records

- failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015 are: F9 v1.1 81012 & 81015, at launch site CCAFS LC-40, Cape Canaveral Space Launch Complex 40.
- Falcon lands on droneship, but the lockout collet doesn't latch on one the four legs, causing it to tip over post landing. Root cause may have been ice buildup due to condensation from heavy fog at liftoff. Credit: SPACENEWS

List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

In [18]:

```
task_9 = '''
        SELECT BoosterVersion, LaunchSite, LandingOutcome
        FROM SpaceX
        WHERE LandingOutcome LIKE 'Failure (drone ship)'
            AND Date BETWEEN '2015-01-01' AND '2015-12-31'
        '''
create_pandas_df(task_9, database=conn)
```

Out[18]:

	boosterversion	launchsite	landingoutcome
0	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
1	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

- Rank 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
- There is no definitive conclusion to derive whether drone ship or ground landings are better off because it depends on the specific mission's needs and goals.
- Overall it depends on SpaceX's ability to choose between these two landing methods which provides flexibility that enables them to optimize the rocket recovery process for each mission.

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad))

In [19]:

```
task_10 = '''
        SELECT LandingOutcome, COUNT(LandingOutcome)
        FROM SpaceX
        WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
        GROUP BY LandingOutcome
        ORDER BY COUNT(LandingOutcome) DESC
        '''
create_pandas_df(task_10, database=conn)
```

Out[19]:

	landingoutcome	count
0	No attempt	10
1	Success (drone ship)	6
2	Failure (drone ship)	5
3	Success (ground pad)	5
4	Controlled (ocean)	3
5	Uncontrolled (ocean)	2
6	Precluded (drone ship)	1
7	Failure (parachute)	1

Section 3

Launch Sites Proximities Analysis

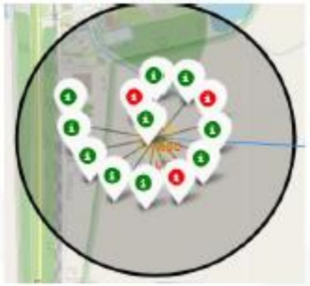
Folium Map : SpaceX Launch Sites on the World

- Proximity to the Equator: Launch sites located closer to the Earth's equator enjoy advantages when launching rockets into equatorial orbits. These sites benefit from the Earth's natural rotational speed, providing a built-in velocity boost that reduces the need for additional fuel and boosters. This proximity results in cost savings for reusable rocket launches.

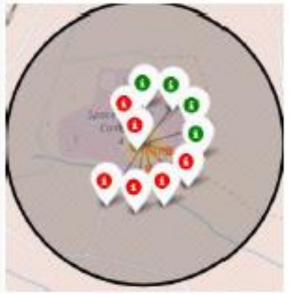
RED = Failure GREEN = Success

CCAFS LC-40

CCAFS SLC-40

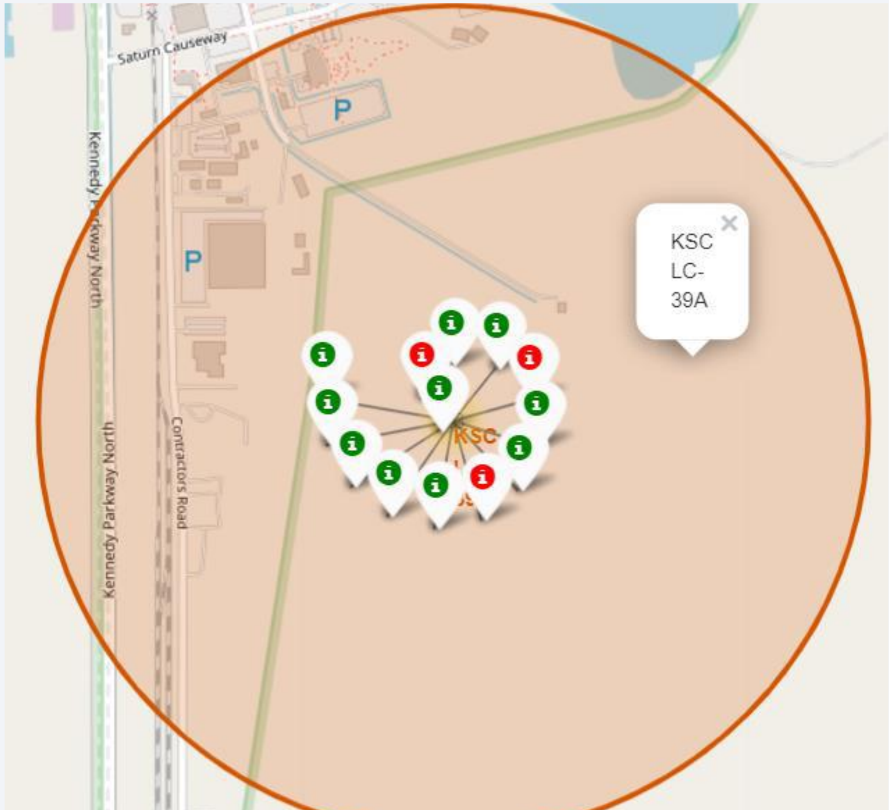


KSC LC-39A

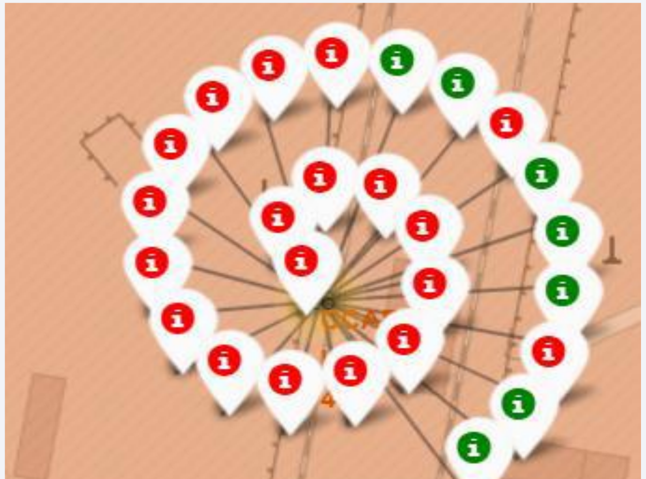


VAFB SLC-4E

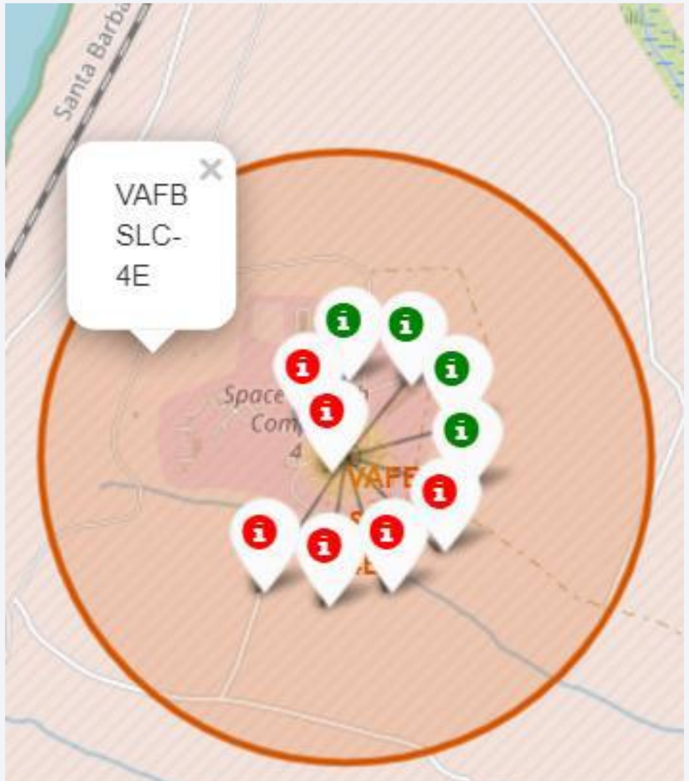
Folium Map: Success/Failed Launches for Each Launch Site



KSC LC-39A
Kennedy Space Center Launch Complex 39A
Florida

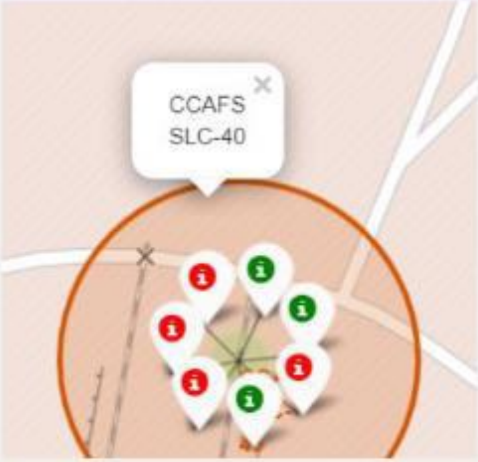


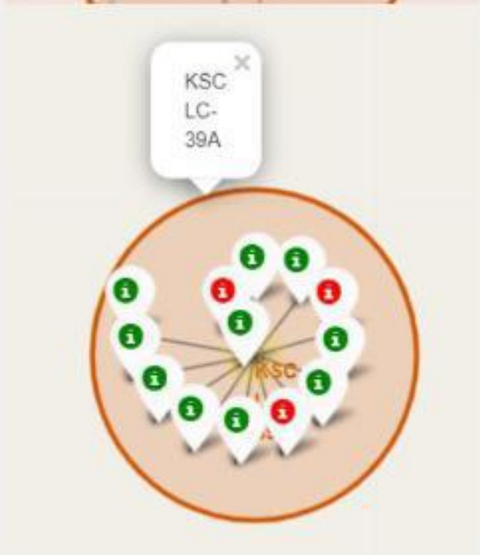
CCAFS SLC-40
Cape Canaveral Space Launch Complex 40
Florida

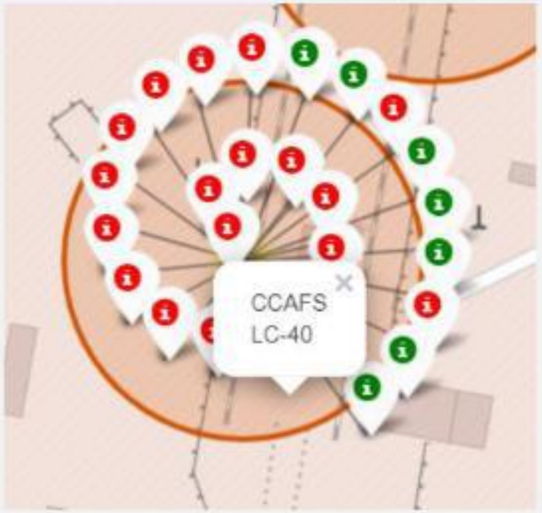


Space Launch Complex 4,
Vandenberg Space Force
Base, California

Folium Map: Success/Failed Launches Markers for Each Launch Site





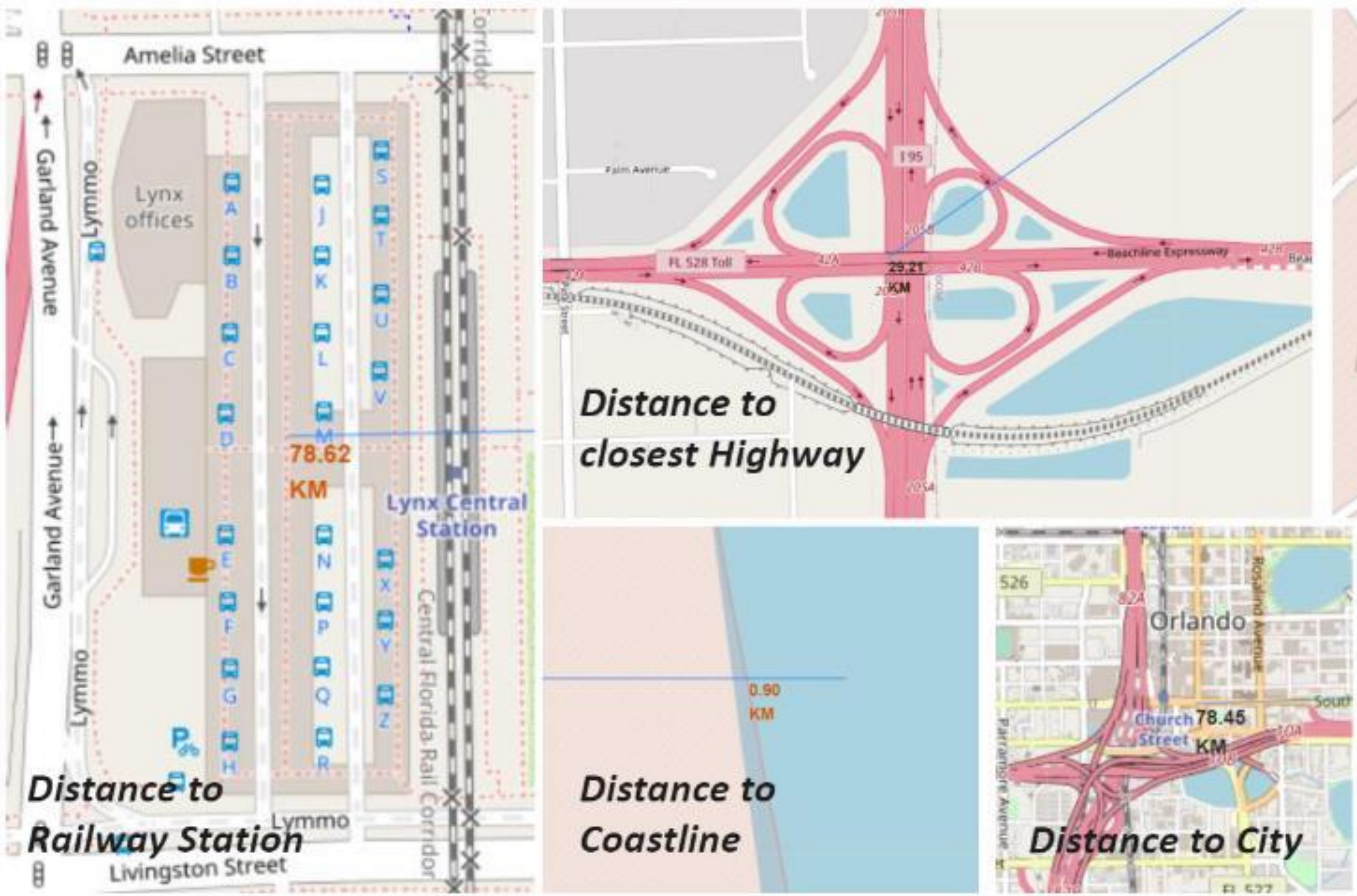


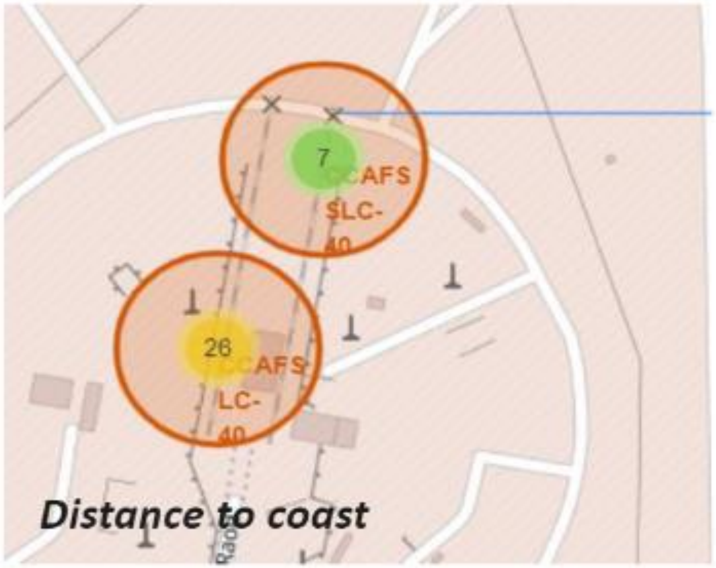
Florida Launch Sites

Green Marker shows successful Launches and Red Marker shows Failures

California Launch Site

Folium Map: Distances





- Are launch sites in close proximity to railways? No
- Are launch sites in close proximity to highways? No
- Are launch sites in close proximity to coastline? Yes
- Do launch sites keep certain distance away from cities? Yes

Section 4

Build a Dashboard with Plotly Dash

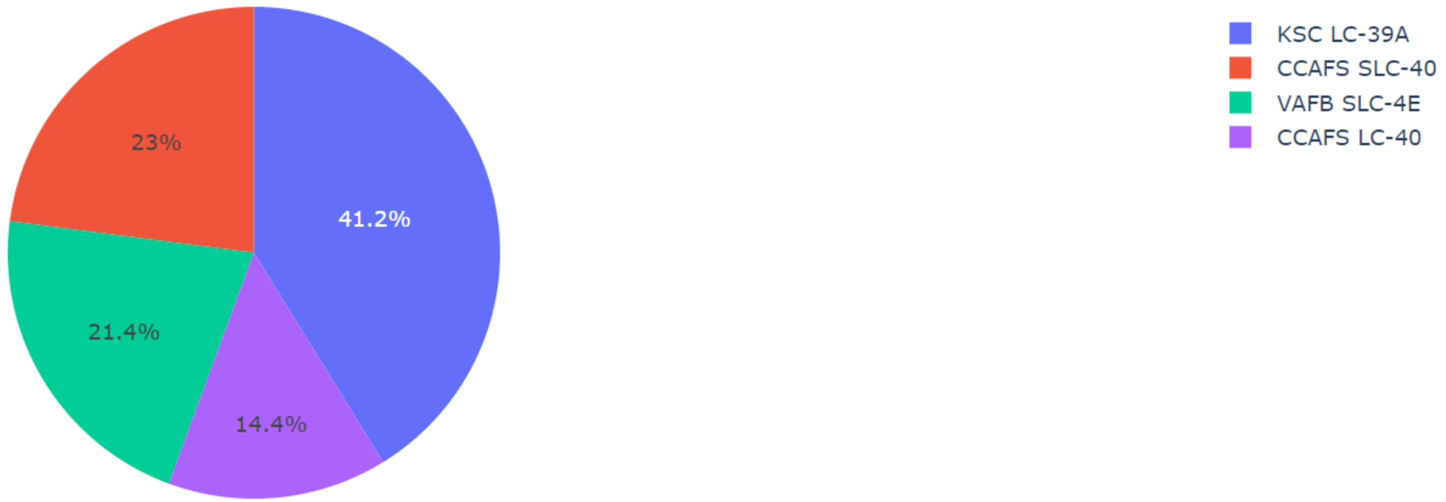
Launch Record Success by Site

SpaceX Launch Records Dashboard

All Sites

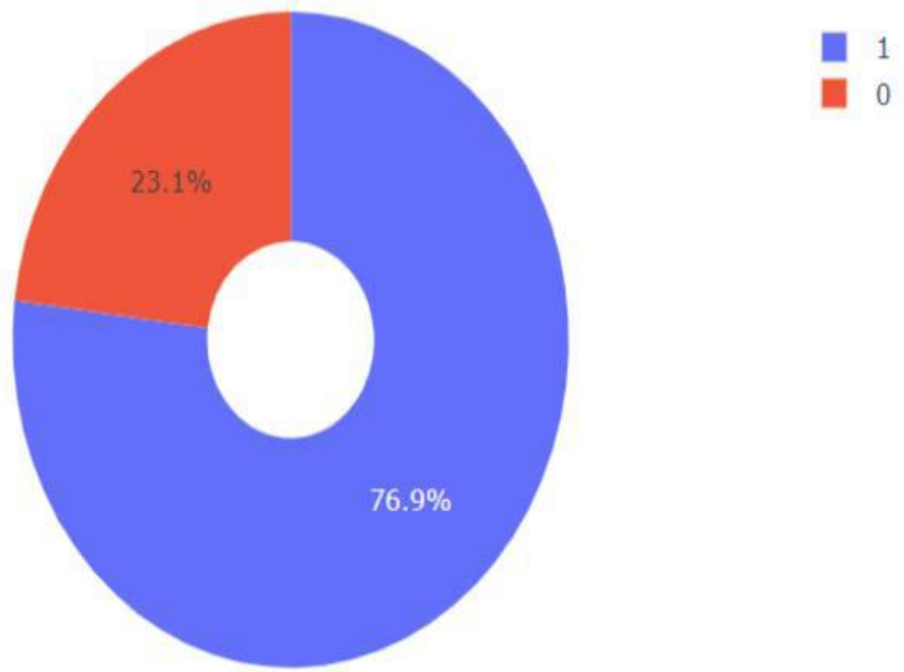
Total Success Launches by Site

- KSC LC 39A by far was the most successful launch site record among launch sites by 41.2%



Launch Success of site # KSC LC-39A

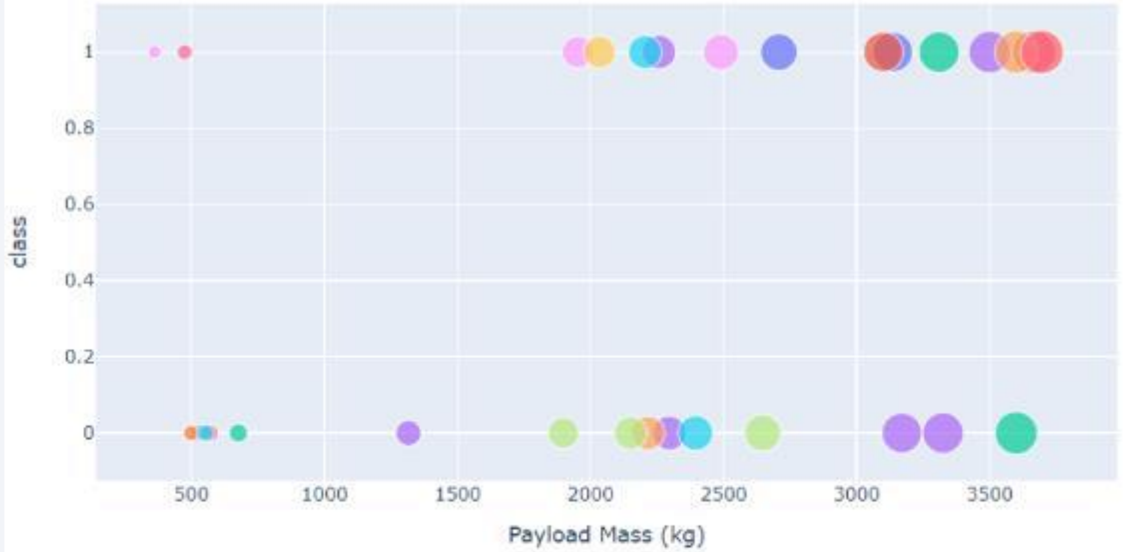
- KSC LC- 39A has the highest success rate compared to its failure rate (76.9%)
- 10 successful launches, 3 failed launches



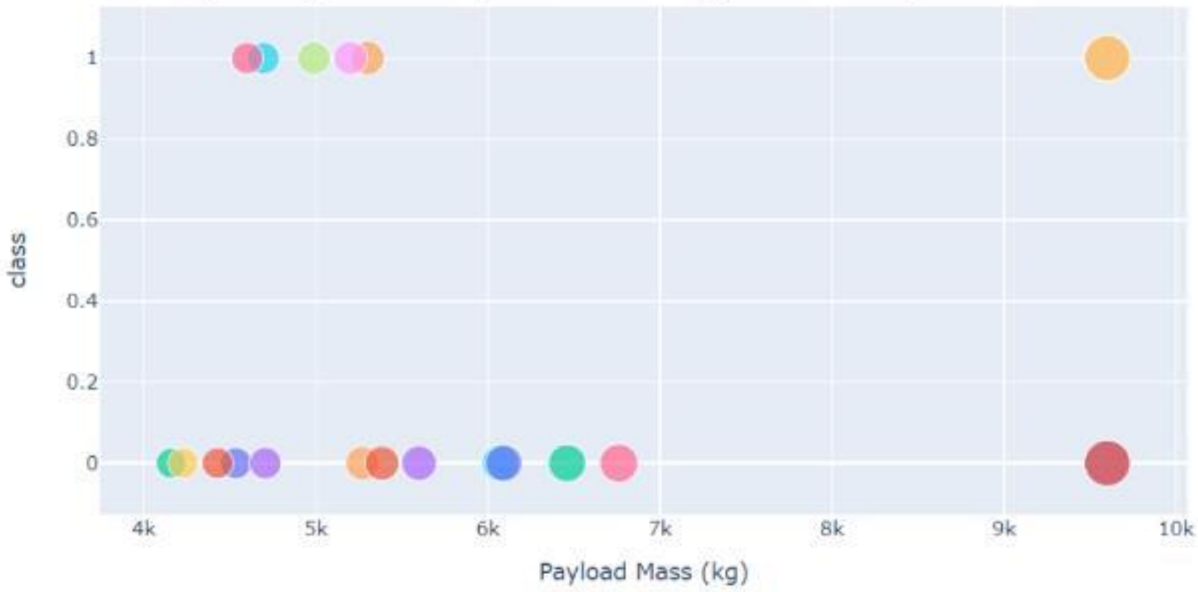
KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Payload vs Launch Outcome Success Rates

Low Weighted Payload 0kg – 4000kg



Heavy Weighted Payload 4000kg – 10000kg



We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

Section 5

Predictive Analysis (Classification)

Classification Accuracy

```
models = {'KNeighbors':knn_cv.best_score_,
              'DecisionTree':tree_cv.best_score_,
              'LogisticRegression':logreg_cv.best_score_,
              'SupportVector': svm_cv.best_score_}

bestalgorithm = max(models, key=models.get)
print('Best model is', bestalgorithm,'with a score of', models[bestalgorithm])
if bestalgorithm == 'DecisionTree':
    print('Best params is :', tree_cv.best_params_)
if bestalgorithm == 'KNeighbors':
    print('Best params is :', knn_cv.best_params_)
if bestalgorithm == 'LogisticRegression':
    print('Best params is :', logreg_cv.best_params_)
if bestalgorithm == 'SupportVector':
    print('Best params is :', svm_cv.best_params_)
```

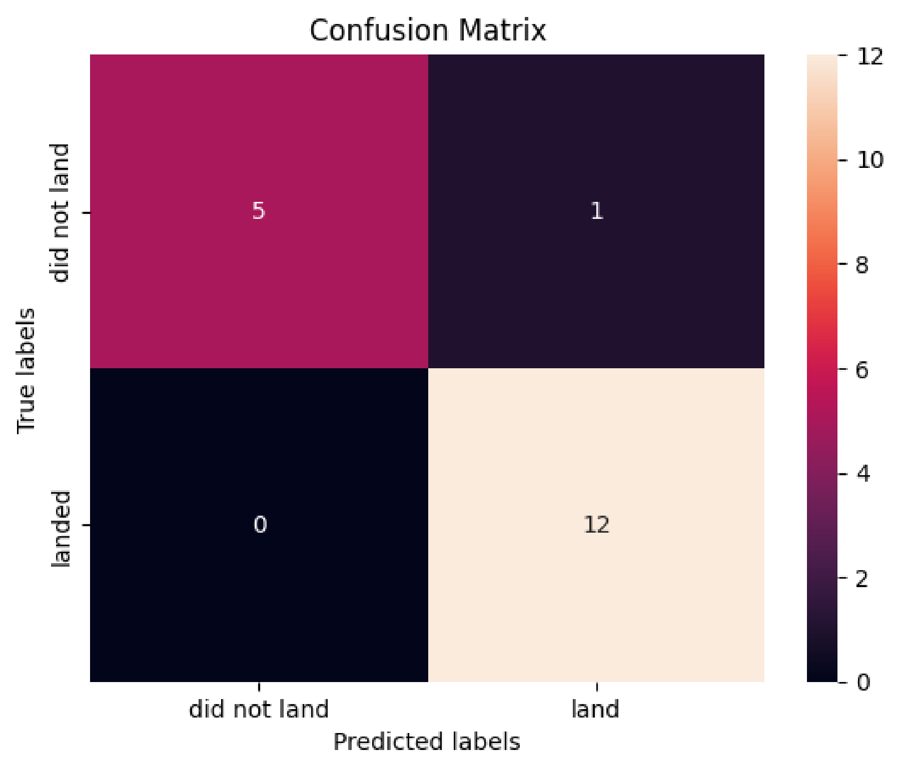
```
Best model is DecisionTree with a score of 0.8732142857142856
Best params is : {'criterion': 'gini', 'max_depth': 6, 'max_features': 'auto', 'min_samples_leaf': 2, 'min_samples_split': 5, 'splitter': 'random'}
```

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

- All models exhibited similar performance with equivalent scores and accuracy.
- Notably, the decision tree classifier outperformed the others, achieving the highest classification accuracy of 0.875.

Confusion Matrix

- As observed in the Classification Accuracy, the Decision Tree classifier model yields the most favorable outcomes for our dataset.
- However, the confusion matrix for the Decision Tree Classifier reveals a noteworthy concern: the classifier generates a considerable number of false positives.
- In other words, it frequently predicts a successful landing when, in reality, the rocket did not land successfully.
- It's worth noting that all the confusion matrices produced identical results.
- The presence of false positives (Type 1 errors) is a significant issue, with unsuccessful landings being erroneously classified as successful by the classifier.



Conclusions

- A greater number of flights at a specific launch site shows a positive correlation with a higher success rate at that particular site.
- Among the orbits, ES-L1, GEO, HEO, SSO, and VLEO, we observe the highest success rates.
- When dealing with heavier payloads, we see an increase in the success rate, particularly for LEO, ISS, and Polar orbits.
- The data indicates that the launch success rate has experienced an overall increase from 2013 to 2020, with minor fluctuations throughout this timeframe.
- It's noteworthy that Launch Site KDC LC 39 A boasts the most impressive launch success rate.
- When it comes to predicting the successful landing of SpaceX Falcon 9 rocket's first stage for reuse, the Decision Tree Classifier emerges as the most effective machine learning algorithm/model.

Scientific Suggestions: EQUATORIAL LAUNCH SITES

- Most launch sites should ideally located near the equator to leverage the Earth's rotational speed, which provides an additional natural boost for launching rockets into equatorial orbits.
- This rotational speed of the Earth contributes to the rocket's initial velocity and helps save the cost of using extra fuel and boosters.
- When launching from sites near the equator, rockets receive the inherent speed from the Earth's rotation in the prograde direction, which is beneficial for achieving the desired orbit with less energy expenditure.

This, in turn, reduces the cost and complexity of rocket launches, making equatorial launch sites advantageous for space agencies and companies.

Scientific Suggestions: COASTAL LAUNCH SITES

Launch sites are typically located close to the coast for several practical reasons:

- Safety: Launching near the coast provides a buffer zone in case of launch failures. If a rocket veers off course or encounters problems shortly after liftoff, launching over water reduces the risk to populated areas. It also minimizes the potential impact on critical infrastructure.
- Over-Ocean Trajectory: Rockets launched from coastal sites have a trajectory that takes them over the open ocean, reducing the risk of overflying populated areas or other sensitive locations. This over-ocean trajectory is particularly important during early stages of a launch.
- Range Safety: Coastal launch sites often have well-defined safety corridors that extend over the ocean. These safety corridors are designed to keep the rocket and its payload away from ships, aircraft, and other hazards. Range safety officers closely monitor the rocket's flight path.
- Access to Transportation: Being near the coast provides convenient access to transportation by sea. This facilitates the delivery of rocket components and payload to the launch site and the transportation of personnel.
- Water-Based Recovery: Coastal sites are ideal for missions involving rocket stages that are intended to be recovered by splashing down in the ocean, as it provides a nearby recovery area.

Overall, coastal launch sites offer practical advantages in terms of safety, range, and logistical support, making them a common choice for space launch facilities.

Appendix

Data is collected from SpaceX's official API and its endpoints.

- https://api.spacexdata.com
- https://api.spacexdata.com/v4/rockets/
- https://api.spacexdata.com/v4/launchpads/
- https://api.spacexdata.com/v4/payloads/
- https://api.spacexdata.com/v4/cores/
- https://api.spacexdata.com/v4/launches/past

Thank you!