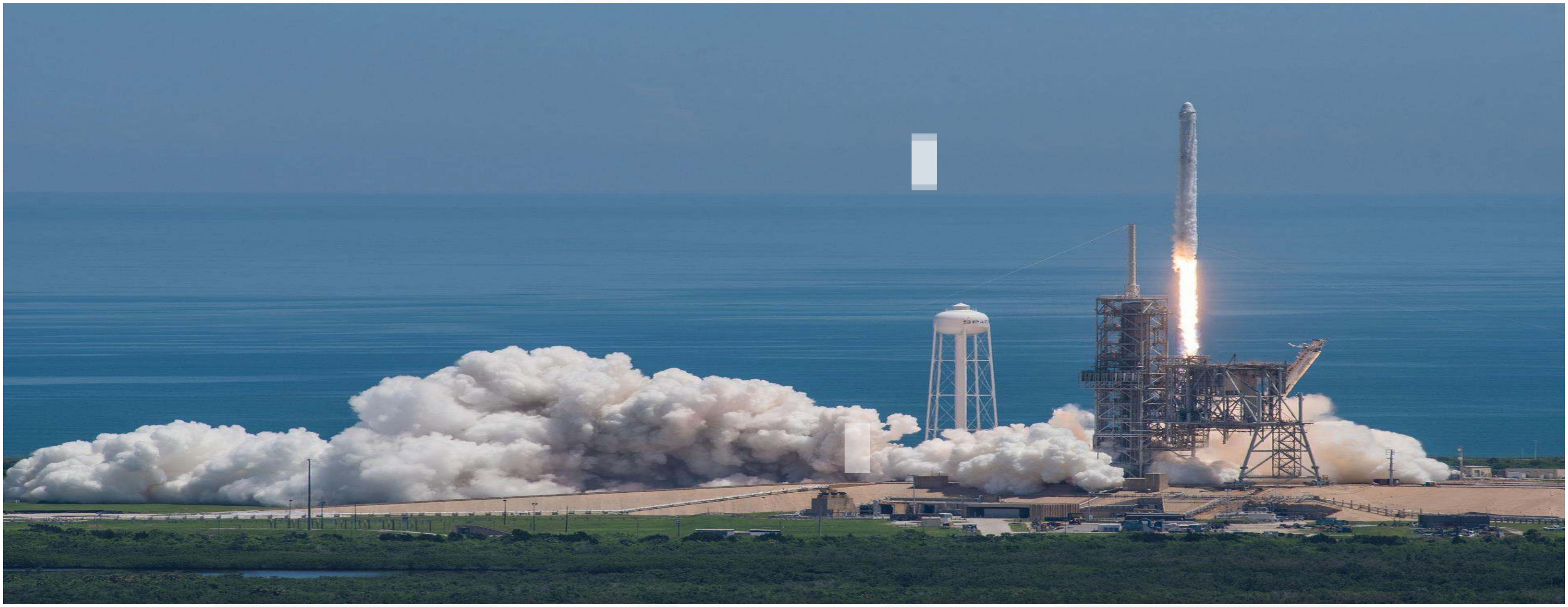
Applied Data Science Capstone
SpaceX Falcon9



Angel Martinez
08-20-2023

IBM Developer

SKILLS NETWORK

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

IBM Developer

SKILLS NETWORK

EXECUTIVE SUMMARY

- This capstone project revolves around the prediction of the successful landing of SpaceX's Falcon 9 first stage through the utilization of various machine learning classification algorithms. The project's core phases encompass data collection, meticulous preprocessing, and effective formatting. These are followed by an insightful exploratory data analysis phase, augmented by interactive data visualization, culminating in the application of machine learning algorithms for predictive modeling.
- Our comprehensive analysis reveals intriguing correlations between distinct features of rocket launches and their ultimate outcomes, whether categorized as successes or failures. The visual representations highlight these correlations, shedding light on key aspects influencing the success of Falcon 9 first stage landings.
- Upon rigorous evaluation, the decision tree algorithm emerges as a promising candidate for accurately predicting the success of Falcon 9 first stage landings. Its robust performance throughout the analysis showcases its potential to yield precise predictions in this context.

INTRODUCTION

In this capstone project, our primary objective is to predict the successful landing of the Falcon 9 first stage, a critical aspect of SpaceX's launch operations. SpaceX has positioned itself as a leader in cost-efficient space travel, evident in the stark contrast between their launch costs and those of their competitors. While traditional providers charge over 165 million dollars per launch, SpaceX's innovative reuse of the first stage allows them to advertise a cost of 62 million dollars, revolutionizing the economics of space travel. Consequently, our predictive analysis not only holds the potential to determine the outcome of a launch but also to estimate its associated costs, which can be invaluable information for competitors vying for rocket launch contracts.

It's important to note that the landscape of unsuccessful landings has evolved. SpaceX now occasionally executes planned unsuccessful landings, choosing controlled ocean landings for certain missions. This nuanced approach highlights the complexity of the predictive challenge at hand.

Amidst these dynamics, we confront a pivotal question: given a comprehensive set of parameters encompassing payload mass, orbit type, launch site, and more, can we accurately anticipate the success of the Falcon 9's first stage landing?

This project is rooted in a broader context – the rapid integration of private space travel into mainstream culture. Recent milestones in private space exploration have contributed to the democratization of space, but the substantial costs associated with launches persist as a formidable entry barrier for new contenders in the space race. SpaceX's pioneering approach to reusing the first stage has set a precedent that grants them a substantial competitive edge. With launch costs significantly lower than the industry average, SpaceX's innovation is reshaping the landscape and challenging traditional norms.

By delving into the intricate relationships between various parameters and landing outcomes, we aim to extract valuable insights. Our analysis seeks to decipher correlations between launch sites and success rates, further enriching our understanding of the factors that underpin successful landings. Through this multifaceted exploration, we strive to contribute to the broader discourse on space travel economics and technology.

In essence, this capstone project amalgamates rigorous data analysis with the evolving dynamics of the space industry. By dissecting the factors that contribute to the success of Falcon 9 first stage landings, we aim to provide actionable insights that can shape decisions and strategies within the competitive landscape of modern space exploration.

METHODOLOGY

The methodology employed in this project integrates various stages to achieve its objectives effectively:

- Data Collection and Preprocessing: The foundation of this study relies on meticulous data collection, which is achieved through multiple approaches:
    - The SpaceX API serves as a primary source for acquiring relevant data.
    - Complementary data is extracted through web scraping techniques, focusing on Falcon 9 and Falcon Heavy launch records obtained from Wikipedia. Subsequent data wrangling ensures data consistency and quality.
    - By transforming mission outcomes into training labels (0 for unsuccessful and 1 for successful), a robust training dataset is formed to support the supervised machine learning models.
- Exploratory Data Analysis (EDA): Pandas, NumPy and SQL are utilized to perform an in-depth exploratory data analysis. This phase unveils patterns, trends, and potential insights within the data, further setting the stage for subsequent analysis.
- Data Visualization: Effective data visualization techniques are harnessed to make complex information comprehensible:
    - Matplotlib and Seaborn contribute to static visualizations, enhancing the understanding of the data's characteristics.
    - Geographic insights are unlocked using Folium, offering an interactive perspective.
    - Dashboard with Plotly Dash
- Machine Learning Prediction: The predictive aspect of this project involves applying diverse classification algorithms to achieve accurate forecasts:
    - Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN) serve as the core classification models.
    - The classification process involves several steps: introducing a 'class' column, standardizing and transforming data, segmenting data into training and test sets, and rigorously evaluating each model's performance using the test data.

Outline
Data collection, wrangling, and formatting

SpaceX API:
Utilize the SpaceX API at https://api.spacexdata.com/v4/rockets/.
Retrieve data about various rocket launches carried out by SpaceX.
Filter the data to include only Falcon 9 launches, as relevant to the project.

Now let's start requesting rocket launch data from SpaceX API with the following URL:

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

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

The data from these requests will be stored in lists and will be used to create a new dataframe.

```
#Global variables
BoosterVersion = []
PayloadMass = []
Orbit = []
LaunchSite = []
Outcome = []
Flights = []
GridFins = []
Reused = []
Legs = []
LandingPad = []
Block = []
ReusedCount = []
Serial = []
Longitude = []
Latitude = []
```

1. API Request/Response

2. After data has been normalized into data frame, data is then stored and used to create a new data frame as Global variables

Outline

Data collection, wrangling, and formatting

SpaceX API

Invoke auxiliary functions to extract pertinent information from columns containing identifiers (e.g., "rocket" column denoting identification numbers):

- Utilize **`getBoosterVersion(data)`** function.
- Utilize **`getLaunchSite(data)`** function.
- Utilize **`getPayloadData(data)`** function.
- Utilize **`getCoreData(data)`** function.

Create a dataset using the obtained data and consolidate columns into a dictionary structure.

```
launch_dict = {'FlightNumber': list(data['flight_number']),
'Date': list(data['date']),
'BoosterVersion':BoosterVersion,
'PayloadMass':PayloadMass,
'Orbit':Orbit,
'LaunchSite':LaunchSite,
'Outcome':Outcome,
'Flights':Flights,
'GridFins':GridFins,
'Reused':Reused,
'Legs':Legs,
'LandingPad':LandingPad,
'Block':Block,
'ReusedCount':ReusedCount,
'Serial':Serial,
'Longitude': Longitude,
'Latitude': Latitude}
```

Transform the dictionary into a data frame, apply a filter to select Falcon 9 launches, and subsequently convert it to a CSV format.

```
# Create a data from launch_dict
df_launch = pd.DataFrame(launch_dict)
```

```
# Hint data['BoosterVersion']!='Falcon 1'
data_falcon9 = df_launch[df_launch['BoosterVersion']!= 'Falcon 1']
```

```
data_falcon9.to_csv('dataset_part\_1.csv', index=False)
```

Outline

Data collection, wrangling, and formatting

Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia

IBM Developer

SKILLS NETWORK

Web scraping Falcon 9 launch records with BeautifulSoup:

- Extract a Falcon 9 launch records HTML table from Wikipedia

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)

Web scraping Falcon 9 launch records with BeautifulSoup:

- Extract a Falcon 9 launch records HTML table from Wiki

1. Execute an HTTP GET request to retrieve an HTML page.
2. Generate a Beautiful Soup object.
3. Extract column names from the HTML table header.
4. Construct a dictionary using the extracted column names as keys.
5. Utilize helper functions to populate the dictionary with launch records.
6. Transform the dictionary into a DataFrame.

1

```
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
```

2

```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(response.text)
```

3

```
html_tables = soup.findAll("table")
```

4 / 5

```
launch_dict= dict.fromkeys(column_names)

# Remove an irrelvant column
del launch_dict['Date and time ( )']

# Let's initial the launch_dict with each value to be an empty list
launch_dict['Flight No.'] = []
launch_dict['Launch site'] = []
launch_dict['Payload'] = []
launch_dict['Payload mass'] = []
launch_dict['Orbit'] = []
launch_dict['Customer'] = []
launch_dict['Launch outcome'] = []
# Added some new columns
launch_dict['Version Booster']=[]
launch_dict['Booster landing']=[]
launch_dict['Date']=[]
launch_dict['Time']=[]
```

6

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

Data Wrangling

Conducted Exploratory Data Analysis (EDA) to uncover patterns within the dataset and establish labels for training supervised models. The dataset included diverse mission outcomes, which were transformed into Training Labels: 1 indicated a successful booster landing, while 0 represented an unsuccessful landing. To define these labels, the following landing scenarios were taken into account:

"True Ocean": Mission outcome successfully landed in a designated area of the ocean.

"False Ocean": Mission outcome unsuccessfully landed in a designated area of the ocean.

"RTLS": Mission outcome successfully landed on a ground pad (Return to Launch Site).

"False RTLS": Mission outcome unsuccessfully landed on a ground pad.

"True ASDS": Mission outcome successfully landed on a drone ship (Autonomous Spaceport Drone Ship).

"False ASDS": Mission outcome unsuccessfully landed on a drone ship.

Data Wrangling-con't

1. Import the dataset into a data frame.
2. Uncover data patterns. Number of launches, orbit for each site and number and occurrence for each site.
3. Generate labels for landing outcomes.

1.

```
URL = 'https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_
resp = await fetch(URL)
dataset_part_1_csv = io.BytesIO((await resp.arrayBuffer()).to_py())
```

Load Space X dataset, from last section.

```
df=pd.read_csv(dataset_part_1_csv)
df.head(10)
```

2.1

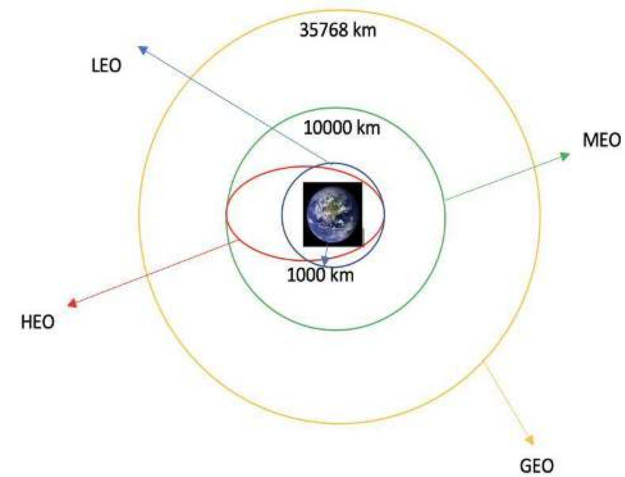
```
df['LaunchSite'].value_counts()
```

```
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
Name: LaunchSite, dtype: int64
```

2.2

```
df['Orbit'].value_counts()
```

```
GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
ES-L1     1
HEO       1
SO        1
GEO       1
Name: Orbit, dtype: int64
```



Each launch aims to an dedicated orbit, and here are some common orbit types:

- **LEO**: Low Earth orbit (LEO)is an Earth-centred orbit with an altitude of 2,000 km (1,200 mi) or less (approximately one-third of the radius of Earth),[1] or with at least 11.25 periods per day (an orbital period of 128 minutes or less) and an eccentricity less than 0.25.[2] Most of the manmade objects in outer space are in LEO [1].
- **VLEO**: Very Low Earth Orbits (VLEO) can be defined as the orbits with a mean altitude below 450 km. Operating in these orbits can provide a number of benefits to Earth observation spacecraft as the spacecraft operates closer to the observation[2].
- **GTO** A geosynchronous orbit is a high Earth orbit that allows satellites to match Earth's rotation. Located at 22,236 miles (35,786 kilometers) above Earth's equator, this position is a valuable spot for monitoring weather, communications and surveillance. Because the satellite orbits at the same speed that the Earth is turning, the satellite seems to stay in place over a single longitude, though it may drift north to south," NASA wrote on its Earth Observatory website [3].
- **SSO (or SO)**: It is a Sun-synchronous orbit also called a heliosynchronous orbit is a nearly polar orbit around a planet, in which the satellite passes over any given point of the planet's surface at the same local mean solar time [4].
- **ES-L1** :At the Lagrange points the gravitational forces of the two large bodies cancel out in such a way that a small object placed in orbit there is in equilibrium relative to the center of mass of the large bodies. L1 is one such point between the sun and the earth [5] .
- **HEO** A highly elliptical orbit, is an elliptic orbit with high eccentricity, usually referring to one around Earth [6].
- **ISS** A modular space station (habitable artificial satellite) in low Earth orbit. It is a multinational collaborative project between five participating space agencies: NASA (United States), Roscosmos (Russia), JAXA (Japan), ESA (Europe), and CSA (Canada) [7]
- **MEO** Geocentric orbits ranging in altitude from 2,000 km (1,200 mi) to just below geosynchronous orbit at 35,786 kilometers (22,236 mi). Also known as an intermediate circular orbit. These are "most commonly at 20,200 kilometers (12,600 mi), or 20,650 kilometers (12,830 mi), with an orbital period of 12 hours [8]
- **HEO** Geocentric orbits above the altitude of geosynchronous orbit (35,786 km or 22,236 mi) [9]
- **GEO** It is a circular geosynchronous orbit 35,786 kilometres (22,236 miles) above Earth's equator and following the direction of Earth's rotation [10]
- **PO** It is one type of satellites in which a satellite passes above or nearly above both poles of the body being orbited(usually a planet such as the Earth

Data Wrangling-con't

3.

```
# landing_outcomes = values on Outcome column
landing_outcomes = df['Outcome'].value_counts()
landing_outcomes
```

```
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
```

EDA with Data Visualization

- Exploratory Data Analysis (EDA): Pandas, NumPy, and SQL are utilized to perform an in-depth exploratory data analysis. This phase unveils patterns, trends, and potential insights within the data, further setting the stage for subsequent analysis.
- Data Visualization: Effective data visualization techniques are harnessed to make complex information comprehensible:
  - Matplotlib and Seaborn contribute to static visualizations, enhancing the understanding of the data's characteristics.
  - Geographic insights are unlocked using Folium, offering an interactive perspective.
  - EDA with Data Visualization - Link

EDA with SQL

To gain a deeper understanding of the SpaceX dataset, the following SQL queries/operations were executed on an IBM DB2 cloud instance:

1. Retrieve distinct launch site names from the space missions.
2. Display 5 records where launch sites start with the 'CCA' string.
3. Calculate the total payload mass carried by boosters launched by NASA (CRS).
4. Calculate the average payload mass carried by booster version F9 v1.1.
5. Identify the date of the first successful landing outcome on a ground pad.
6. List booster names that achieved success on a drone ship with payload mass greater than 4000 and less than 6000.
7. Count and display the total number of successful and failed mission outcomes.
8. List the names of booster versions that carried the maximum payload mass using a subquery.
9. List the details of failed landing outcomes on a drone ship, including booster versions and launch site names, specifically in the year 2015.
10. Rank the count of landing outcomes (e.g., Failure (drone ship) or Success (ground pad)) between the dates 2010-06-04 and 2017-03-20, in descending order.

EDA with SQL-Results

1. Retrieve distinct launch site names from the space missions.

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

2. Display 5 records where launch sites start with the 'CCA' string.

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

EDA with SQL-Results con't

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

   45596

4. Calculate the average payload mass carried by booster version F9 v1.1.

   2928

5. Identify the date of the first successful landing outcome on a ground pad.

   2015-12-22

6. List booster names that achieved success on a drone ship with payload mass greater than 4000 and less than 6000.

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

7. Count and display the total number of successful and failed mission outcomes.

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

EDA with SQL-Results con't

8. List the names of booster versions that carried the maximum payload mass using a subquery.

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

9. List the details of failed landing outcomes on a drone ship, including booster versions and launch site names, specifically in the year 2015.

DATE	booster_version	launch_site
2015-01-10	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	F9 v1.1 B1015	CCAFS LC-40

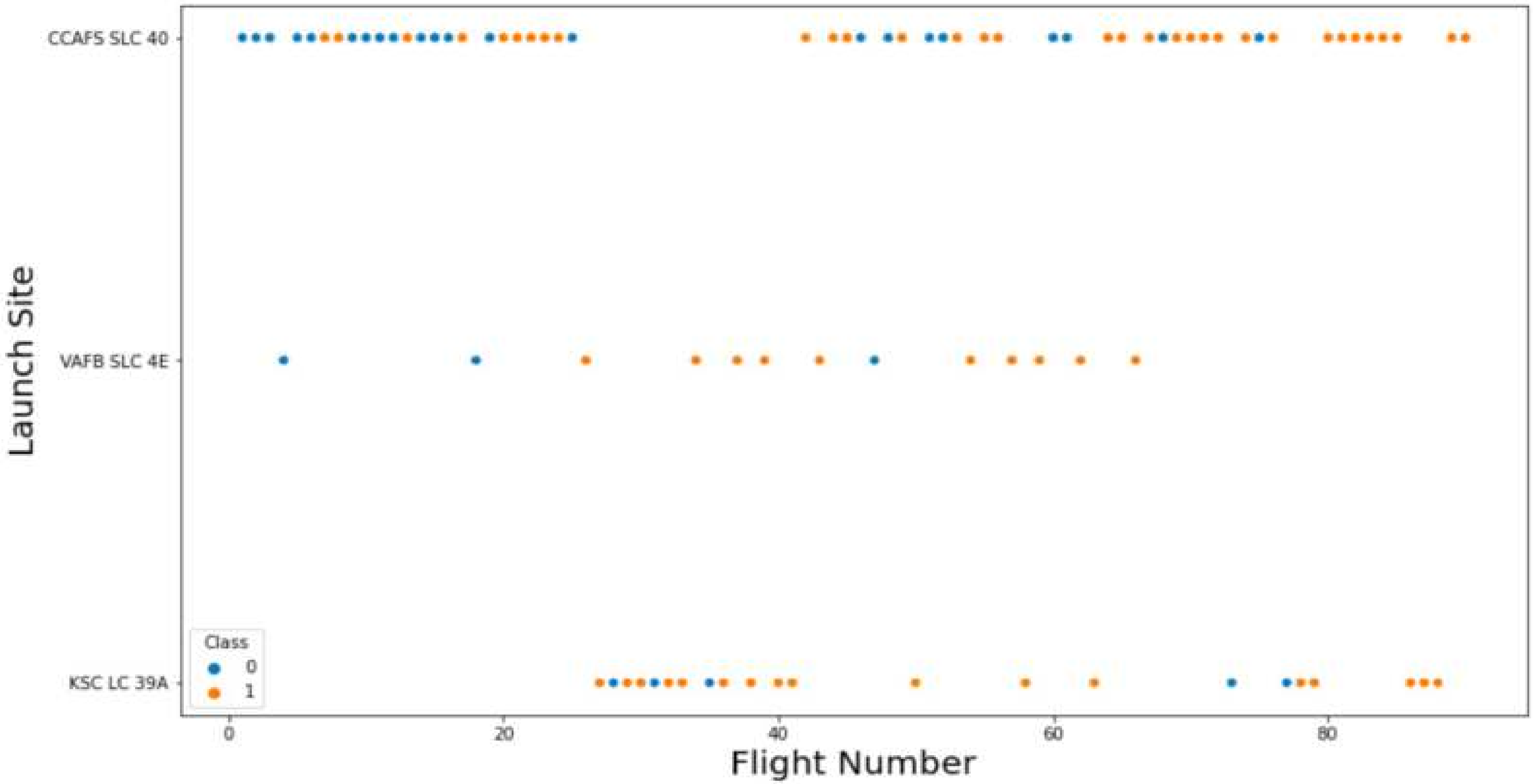
EDA with SQL-Results con't

10. Rank the count of landing outcomes (e.g., Failure (drone ship) or Success (ground pad)) between the dates 2010-06-04 and 2017-03-20, in descending order.

landing__outcome	landing_count
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

EDA with Visualization with Matplotlib and Seaborn

The relationship between flight number and launch site

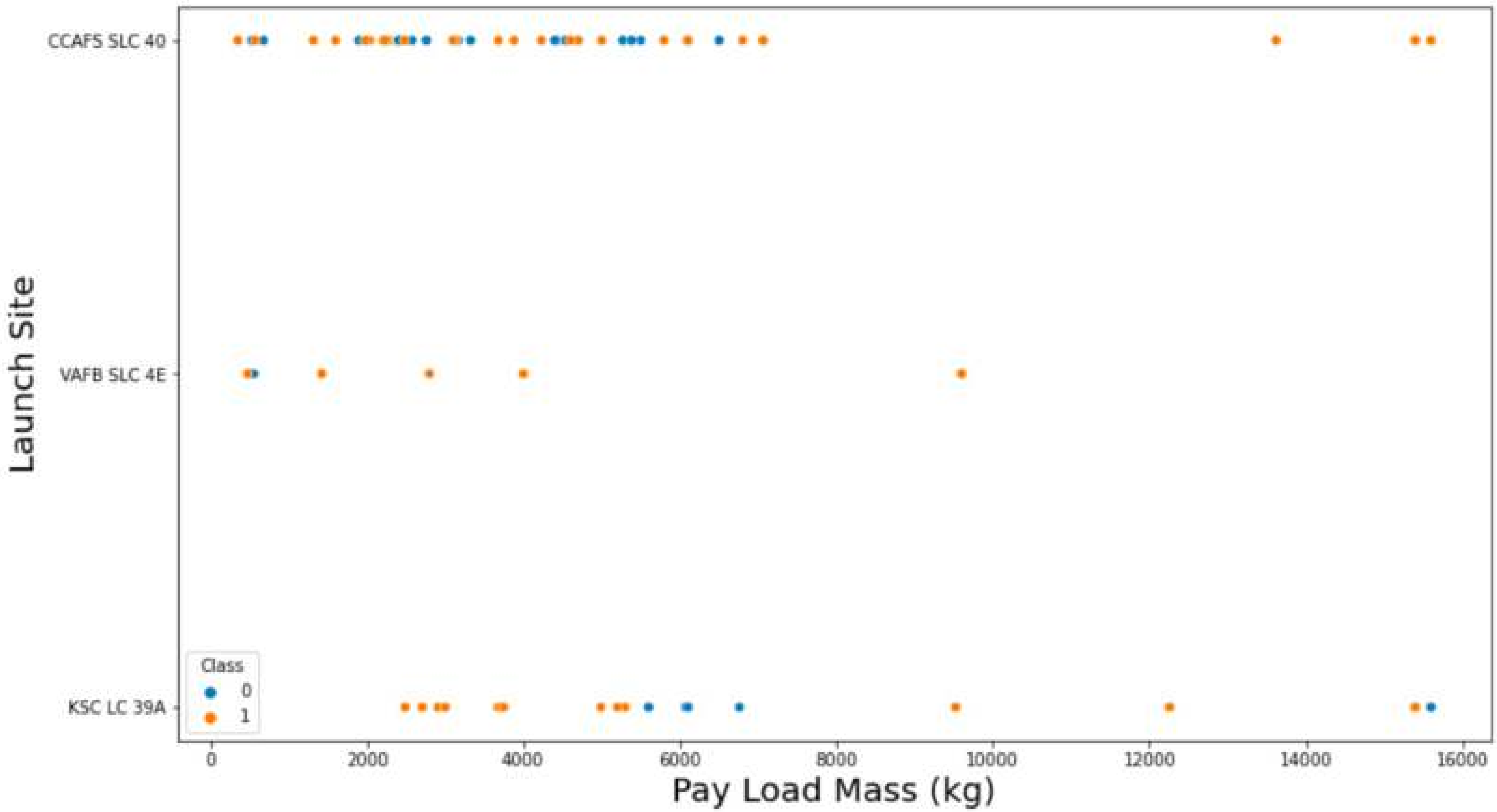


IBM Developer

SKILLS NETWORK

EDA with Visualization with Matplotlib and Seaborn con't

The relationship between payload mass and launch site

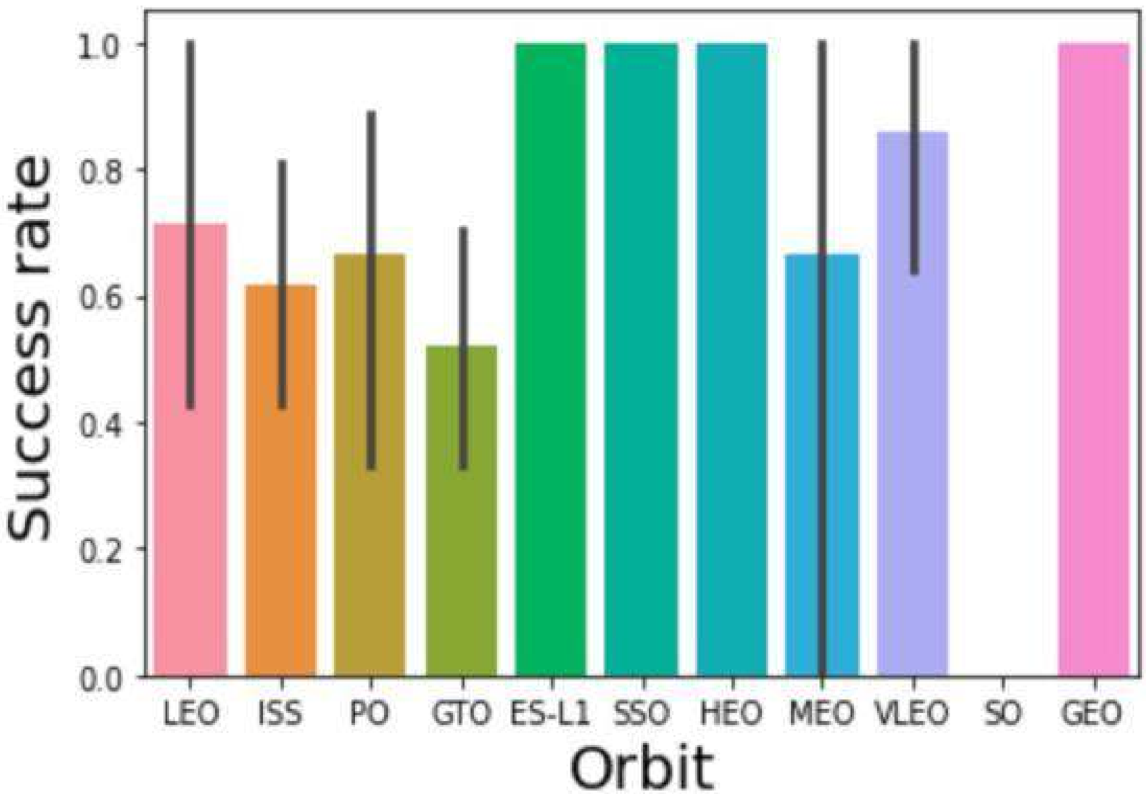


IBM Developer

SKILLS NETWORK

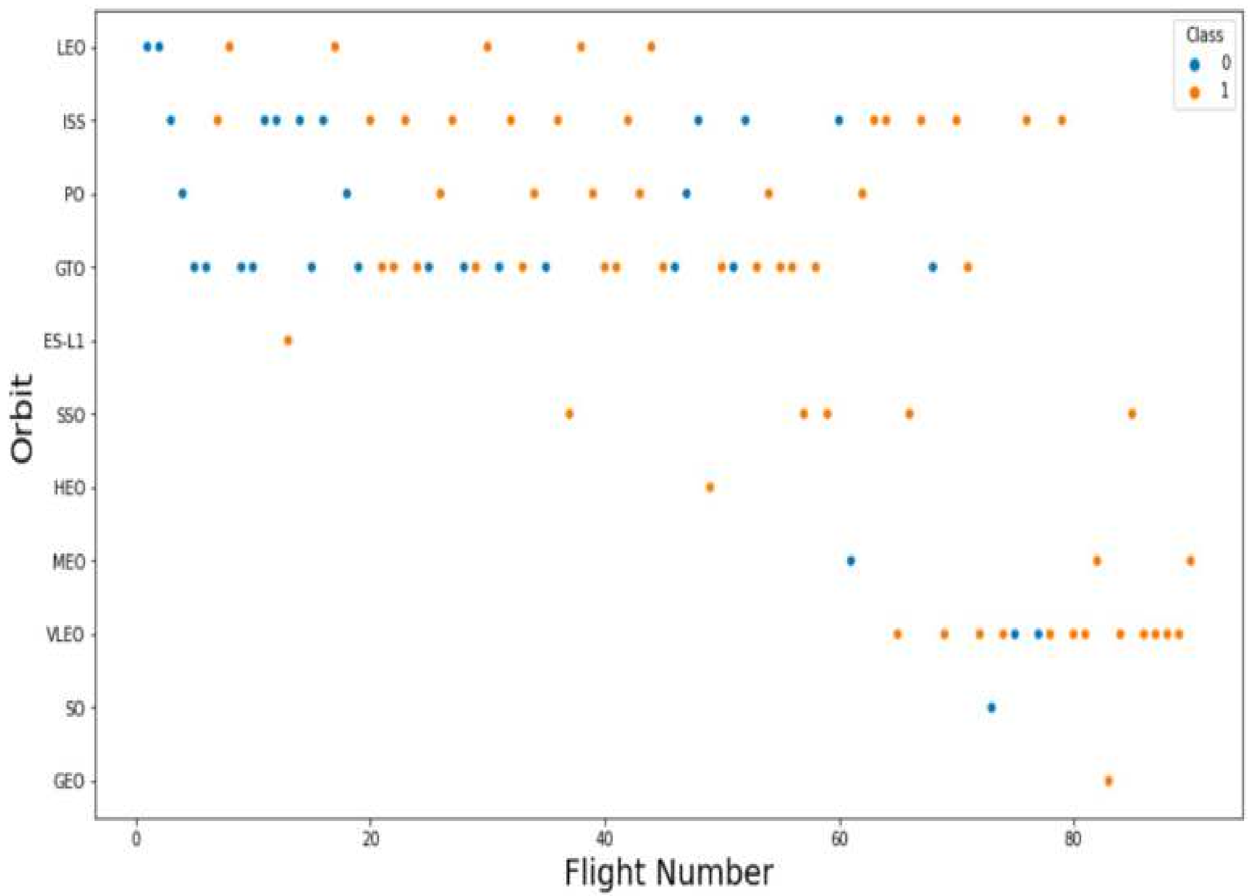
EDA with Visualization with Matplotlib and Seaborn con't

The relationship between success rate and orbit type



EDA with Visualization with Matplotlib and Seaborn con't

The relationship between flight number and orbit type

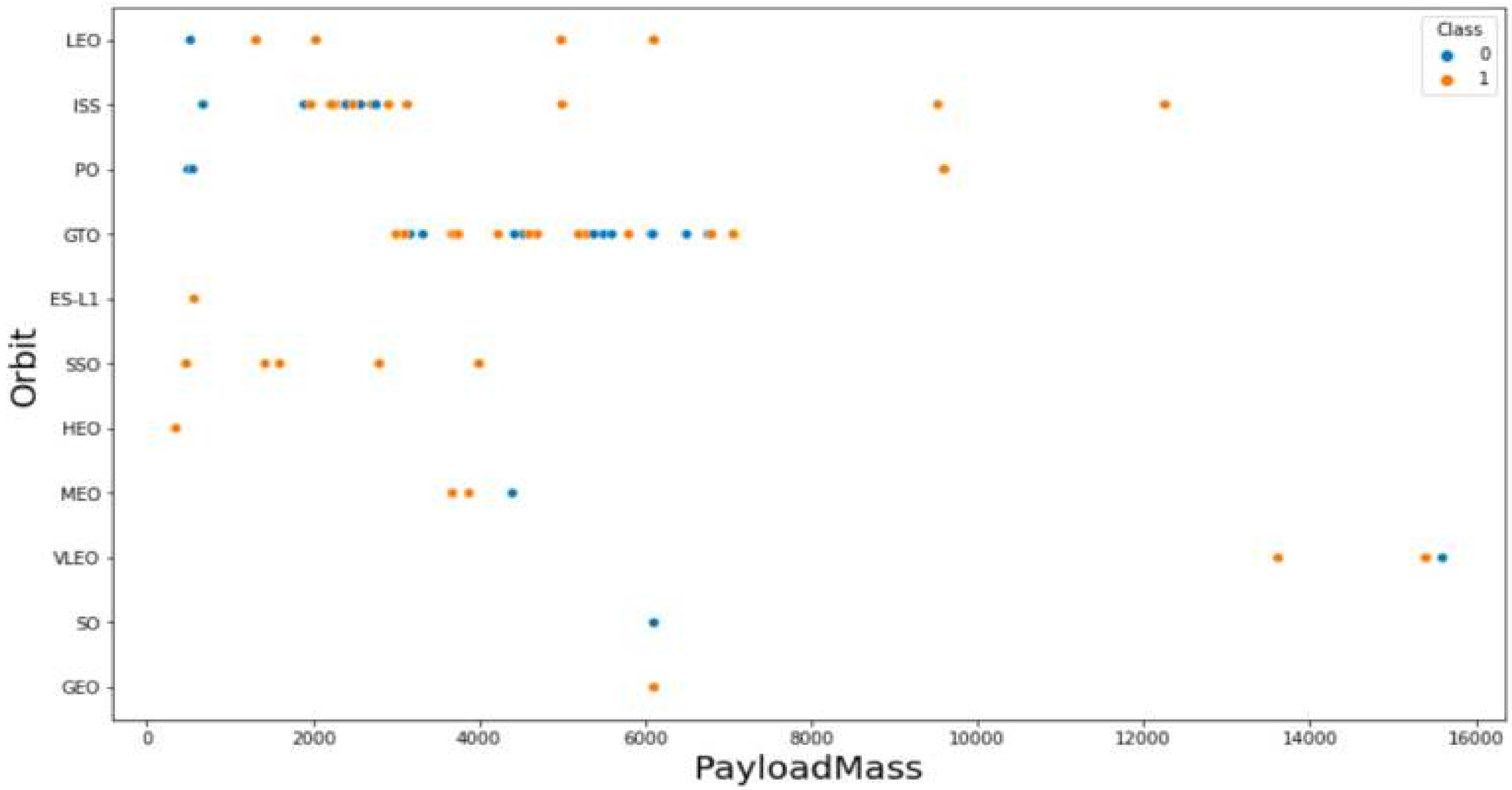


IBM Developer

SKILLS NETWORK

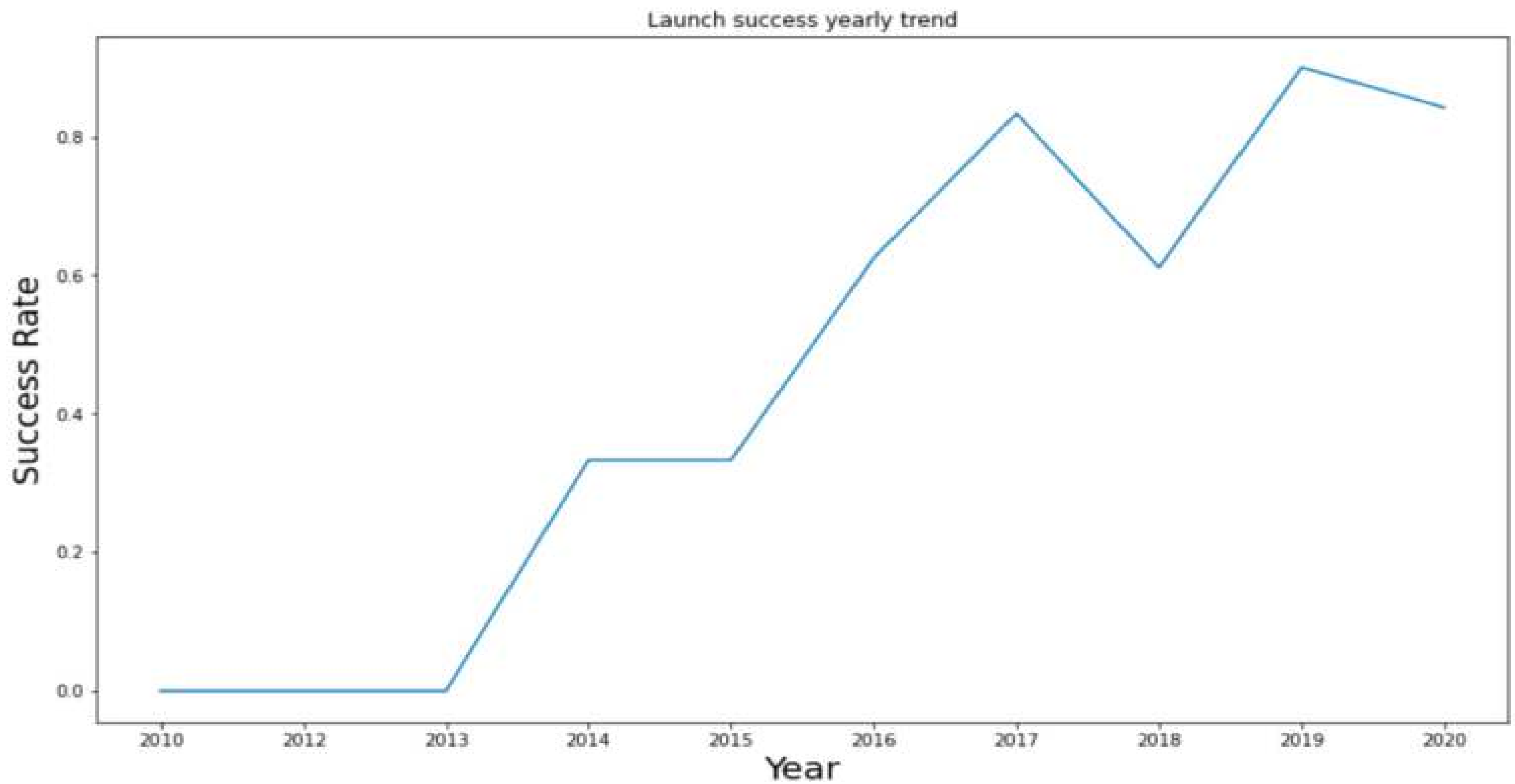
EDA with Visualization with Matplotlib and Seaborn con't

The relationship between payload mass and orbit type



EDA with Visualization with Matplotlib and Seaborn con't

The launch success yearly trend



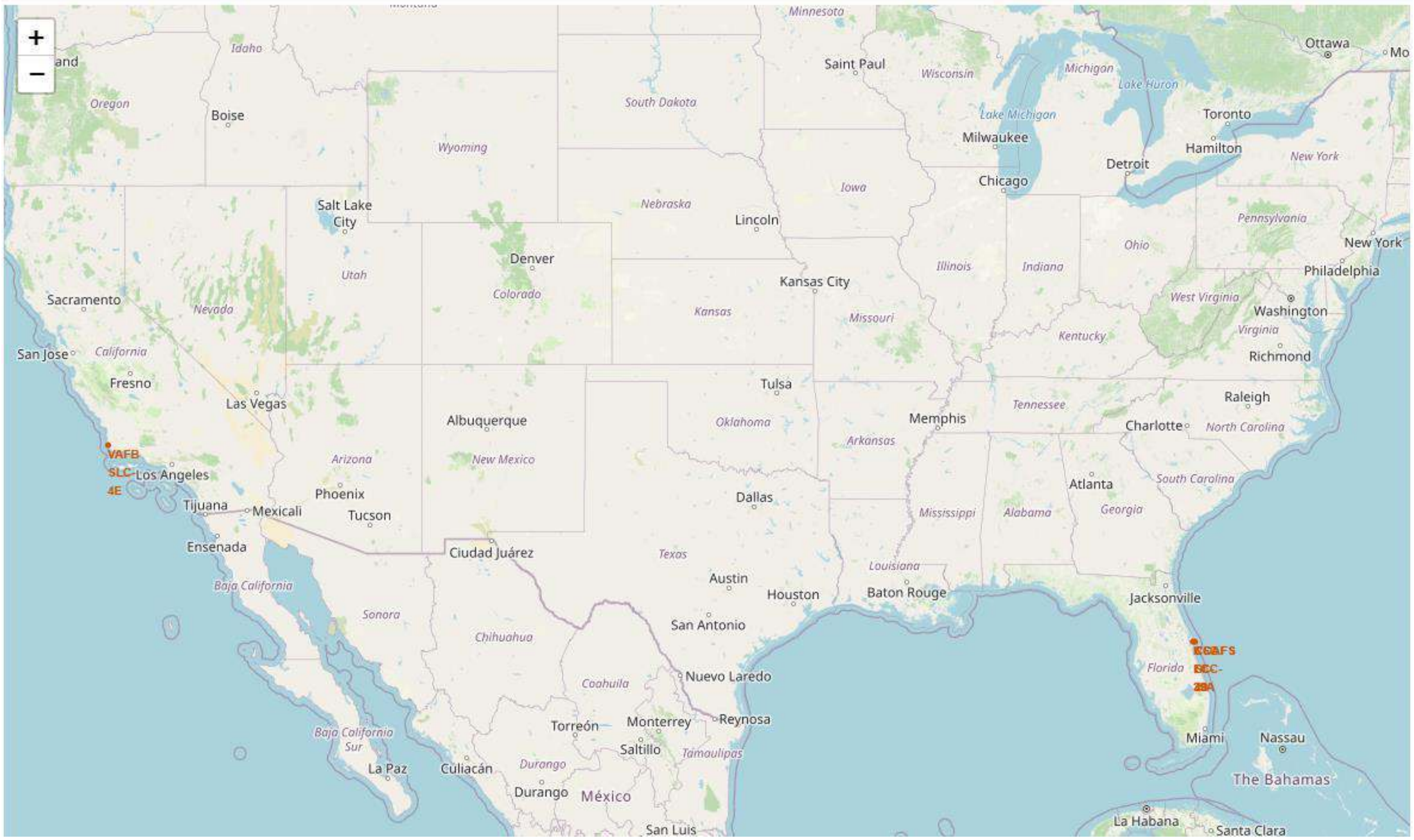
IBM Developer

SKILLS NETWORK

Geographic insights are unlocked using Folium

Interactive Map with Folium, offering an interactive perspective

Each launch site on the site map



IBM Developer

SKILLS NETWORK

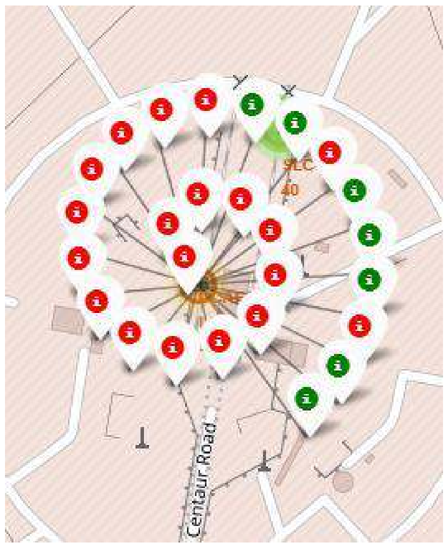
Geographic insights are unlocked using Folium con't

Success/Failed launches for each site on the map

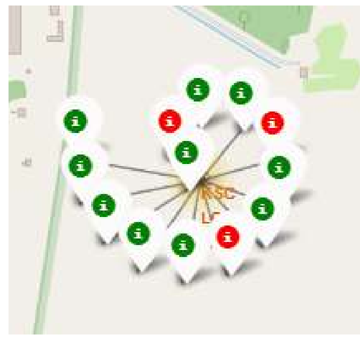
When we focus on a specific launch site, we observe a combination of green and red markers. Each green marker signifies a successful launch, whereas each red marker indicates a failed launch.

Site CCAF LC-40

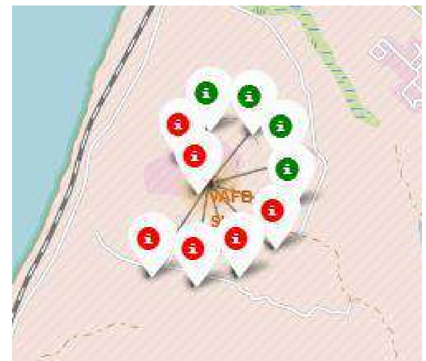
Site CCAF SLC-40



Site KSC LC-39A



Site VAFB SLC-4E

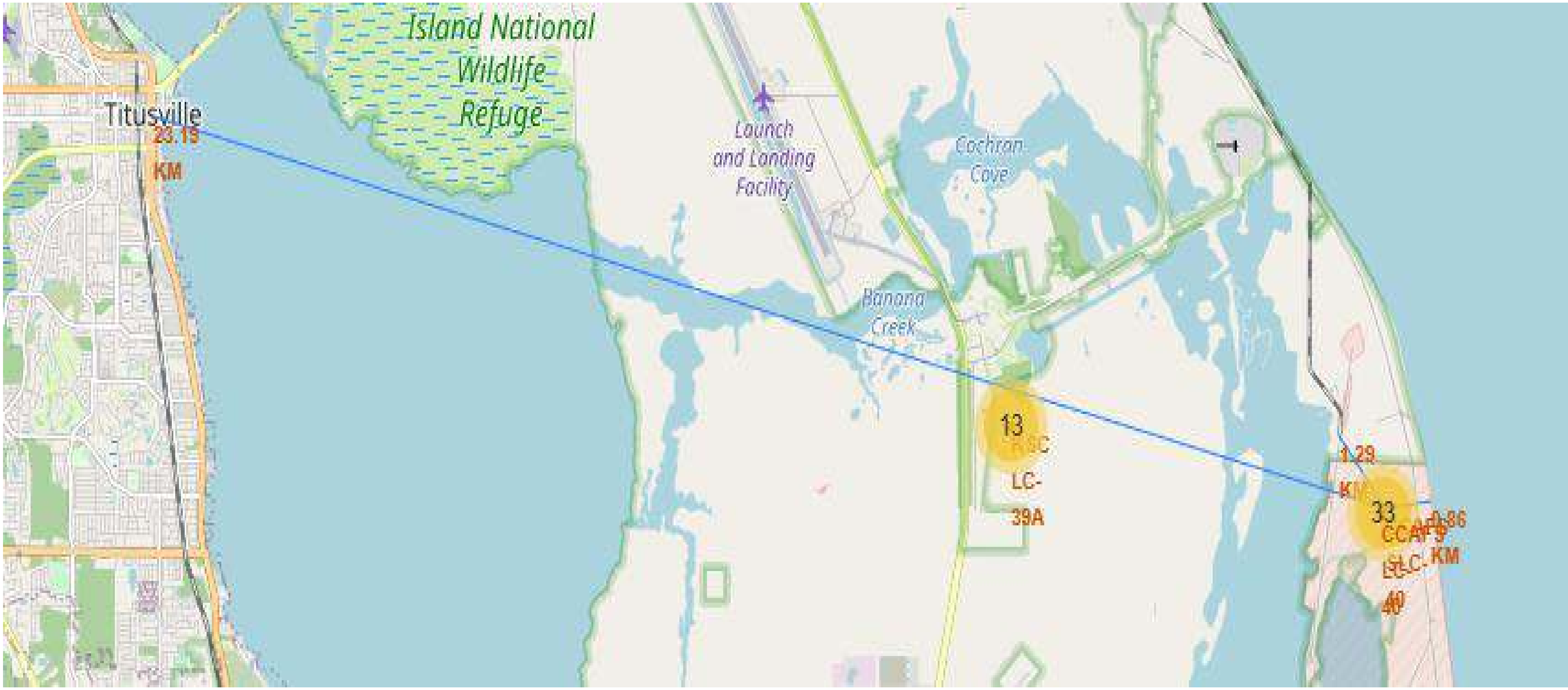


Geographic insights are unlocked using Folium con't

The measurements of distances between a launch site and its nearby locations, such as the nearest city, railway, or highway, are considered. The image provided depicts the distance between the CCAFS SLC-40 launch site and the closest coastline.

Geographic insights are unlocked using Folium con't

The measurements of distances between a launch site and its nearby locations, such as the nearest city, railway, or highway, are considered. The image provided depicts the distance between the CCAFS SLC-40 launch site and the closest city & railway.



IBM Developer

SKILLS NETWORK

Dashboard with Plotly Dash

Launch Success Counts For All Sites

Payload vs Launch Outcome Scatter Plots for All Sites

The majority of successful launches fall within the payload range of approximately 2000 to 5500. Among booster version categories, 'FT' stands out as having the highest number of successful launches. Notably, the sole booster version that achieved success with a payload exceeding 6000 is 'B4'.

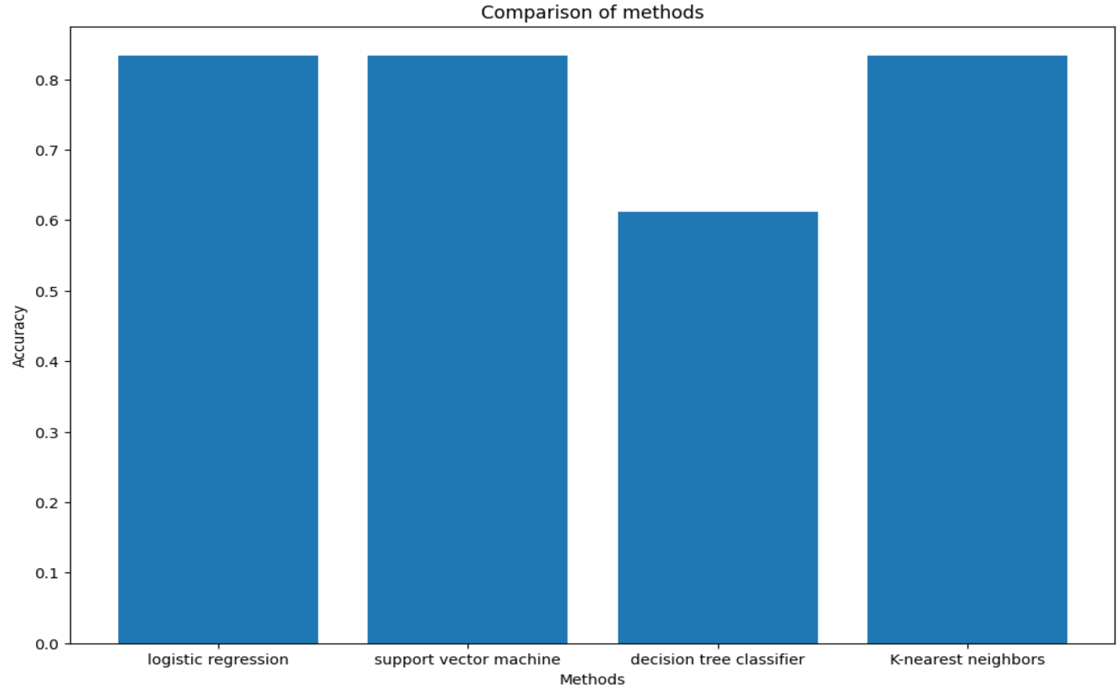
IBM Developer

SKILLS NETWORK

Predictive Analysis

Machine Learning Prediction: The predictive aspect of this project involves applying diverse classification algorithms to achieve accurate forecasts: Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN) serve as the core classification models.

	Algo Type	Accuracy Score	Test Data Accuracy Score
2	Decision Tree	0.875000	0.833333
3	KNN	0.848214	0.833333
1	SVM	0.848214	0.833333
0	Logistic Regression	0.846429	0.833333



IBM Developer

SKILLS NETWORK

Predictive Analysis-Results

Comparing the outcomes of all four models side by side, it's evident that they exhibit identical accuracy scores and confusion matrices when tested against the designated test set. This parity in performance metrics necessitates the utilization of their GridSearchCV best scores as the basis for ranking. By aligning with their GridSearchCV best scores, the models are hierarchically ordered as follows, with the first representing the most optimal and the last indicating the least preferable:

1. Decision Tree
   Score: 0.8892857142857142
2. K-Nearest Neighbors
   Score: 0.8482142857142858
3. Support Vector Machine
   Score: 0.8482142857142856
4. Logistic Regression
   Score: 0.8464285714285713

Conclusion & Insights

As the frequency of flights increases, there is a heightened likelihood of a successful landing for the first stage.
Although success rates demonstrate an inclination to rise with the augmentation of payload mass, a distinct correlation between payload mass and success rates remains elusive.
Over the duration from 2013 to 2020, the launch success rate underwent a substantial escalation of approximately 80%.
Among the launch sites, 'KSC LC-39A' stands out with the highest launch success rate, while 'CCAFS SLC 40' exhibits the lowest.
When evaluating orbits, it becomes evident that ES-L1, GEO, HEO, and SSO orbits boast the highest launch success rates, whereas the GTO orbit records the lowest.
Strategic positioning characterizes the launch sites, ensuring their distance from urban areas while prioritizing proximity to coastlines, railroads, and highways.
The zenith of our analysis points toward the Decision Tree machine learning classification model as the most proficient performer, boasting an accuracy of around 87.5%. This model's competence remains consistent when subjected to the test dataset, maintaining an accuracy score of about 83% across all models. It is plausible that a larger dataset might be requisite to refine and potentially enhance the models, thereby discovering a more optimal fit for predictive purposes.

Discussion

Within this endeavor, our focal objective lies in forecasting the outcome of the initial stage's landing for a given Falcon 9 launch. This projection bears the pivotal role of elucidating the associated launch cost, which holds significance in a realm of cost-efficient space exploration.

Diving into the intricacies of Falcon 9 launches, we recognize that each feature embedded within the launch—be it payload mass, orbit type, or other variables—wields the potential to exert distinctive influences on the ultimate mission outcome.

Our approach employs the power of diverse machine learning algorithms, acting as conduits for deciphering patterns entrenched within historical Falcon 9 launch data. By gleaning insights from this trove of past endeavors, we meticulously craft predictive models poised to forecast the trajectory of a Falcon 9 launch's outcome.

In the pursuit of identifying the most adept model, we evaluate the performance of several machine learning algorithms. It is noteworthy that amidst this comprehensive assessment, the predictive model engendered by the decision tree algorithm emerges as the frontrunner. Its prowess in capturing the underlying intricacies within the data resonates as it outperforms the array of four distinct algorithms harnessed in this endeavor.