Data Science and Machine Learning for Predicting SpaceX Falcon 9 Landings

Umamaheswaran V
Feb 20,2025

OUTLINE

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

EXECUTIVE SUMMARY

- Leverage data science and machine learning to predict the success of SpaceX Falcon 9 first stage landings, aiming to reduce launch costs and enhance strategic planning in the aerospace industry.
- Data Collection and Analysis : Gathered and prepared historical launch data, including flight details, launch sites, payloads, orbits, and landing outcomes. Performed exploratory data analysis to identify patterns and key factors influencing landing success.
- Machine Learning Implementation : Developed predictive models using Logistic Regression, Support Vector Machines (SVM), Decision Trees, and K-Nearest Neighbors (KNN). Applied hyperparameter tuning to optimize model performance, achieving approximately 83% accuracy across models.
- Key Findings and Impact : Identified significant features affecting landing success—such as launch site, orbit type, and payload mass. Demonstrated that predictive analytics can significantly contribute to cost optimization, mission planning, and competitive advantage in the aerospace sector.

INTRODUCTION

- The project successfully demonstrates the power of data science and machine learning in predicting SpaceX Falcon 9 first stage landing outcomes, offering valuable insights for cost optimization and strategic decision-making in future space missions.
- Data Collection and Preparation: Gathered and cleaned historical launch data, including flight numbers, dates, launch sites, payload masses, orbits, and landing outcomes. Performed feature engineering by converting categorical variables using one-hot encoding and standardized numerical features for model readiness.
- Exploratory Data Analysis (EDA):
  - Launch site, orbit type, and payload mass are critical factors influencing landing success.
- Model Development and Evaluation:
  - Trained multiple machine learning models—including Logistic Regression, Support Vector Machines (SVM), Decision Trees, and K-Nearest Neighbors (KNN)—using GridSearchCV for hyperparameter tuning. All models achieved approximately 83% accuracy on test data, with the Decision Tree model slightly outperforming others.
  - Predictive models can optimize resource allocation, reduce operational costs, and provide a competitive edge in the aerospace sector.

METHODOLOGY

- Data Collection :
  - SpaceX API
- Web Scraping :
  - BeautifulSoup
  - Request
- Exploratory Data Analysis(EDA) :
  - Pandas and Numpy
  - SQL
- Data Visualization :
  - Matplotlib and Seaborn
  - Folium
  - Dash by Plotly
- Machine Learning Prediction :
  - Scikit-Learn

METHODOLOGY

Data Collection:

- SpaceX API :
  - Using SpaceX API (https://api.spacexdata.com/v4/rockets/)
  - Ensuring the data is in the correct format.
  - The filtered and cleaned DataFrame contains information on Falcon 9 launches, including launch details, payloads, cores, and landing outcomes.
  - Missing values in the PayloadMass column have been replaced with the mean value.

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857

Discussion : How can the data analysis be applied to future launches :

- The analysis can provide significant insights for future launches in the following ways:
- Cost Prediction : By predicting the success of the Falcon 9 first stage landings, you can estimate the potential cost savings. Knowing in advance whether a first stage is likely to be reused can help in budgeting and cost planning for future missions.
- Improved Decision-Making : The data can aid in decision-making for SpaceX and other aerospace companies. For instance, understanding the factors contributing to successful landings can influence future design improvements and operational strategies.
- Enhanced Reliability : By continuously analyzing data from previous launches, SpaceX can refine its technology and procedures to improve the reliability and success rates of their missions. This can lead to more consistent successful landings, further reducing costs.
- Risk Assessment : Analyzing past data helps in identifying potential risks and their likelihood. This information can be used to develop contingency plans and mitigate risks in future launches.
- Competitive Advantage : For companies competing against SpaceX, having access to this analysis can provide a competitive edge. They can better understand the factors that make SpaceX successful and use that information to improve their own technologies and bidding strategies.
- Innovation and Development : Data-driven insights can spur innovation in aerospace technology. By understanding the successes and failures of past launches, engineers and scientists can develop new technologies and improve existing ones.
- Customized Payloads : The analysis can help in customizing payloads based on the expected performance of the rocket. This can optimize the use of the rocket's capabilities and ensure the success of the mission.
- Overall, the data analysis from past launches can play a crucial role in enhancing the efficiency, reliability, and competitiveness of future space missions.

Discussion (Contd): Key data points to be focus on successful Landing:

- To ensure the Falcon 9 first stage landing prediction is accurate and informative for future launches, focusing on the following key data points would be essential:
- Booster Version : To identify which variant of the Falcon 9 was used, as different versions may have varying success rates.
- Launch Site Details : To determine the geographical influence on landing success.
- Payload Data : Heavier payloads may affect landing success due to fuel consumption.
- Orbit Details : The type of orbit (e.g., LEO, GEO) may impact the landing trajectory and success.
- Landing Outcome : To directly measure landing success or failure.
- Landing Type : Whether it was a controlled landing on a pad or an ocean recovery.
- Number of Flights : To evaluate the reuse effectiveness of boosters.
- Gridfins Usage : Gridfins' role in stabilizing the descent.
- Reusability : Tracking the number of reuses for each core.
- Landing Pad Usage : To analyze the landing pad's role in success rates.
- Block Number : To understand improvements across different versions.
- Serial : To track the history and performance of each individual core.
- Flight Number : To track the sequence of launches.
- Launch Date : To study the effect of different seasons or times on landing success.
- Weather Conditions : Wind speed, humidity, and temperature during launch and landing.
- Sea State : For ocean landings, the sea state might impact the recovery success.
- Failures and Anomalies : Any recorded issues during flight or landing attempts.
- Recovery Attempts : Instances where recovery was attempted but not successful.
- Legs Deployment : To know if landing legs were used and their impact.
- Fuel Reserves : Amount of fuel left during the landing phase.
- By systematically analyzing these data points, you can gain a comprehensive understanding of factors influencing the success of Falcon 9 first stage landings. This, in turn, can guide future improvements and enhance predictive accuracy.

Discussion (Contd): The correlation between payload mass and landing outcomes :

To understand the correlation between payload mass and landing outcomes, we need to analyze the data to see if there is a relationship between these two variables. The steps include:

- Ensure that the dataset contains complete information on payload mass and landing outcomes.
- Handle any missing values as done in the initial data cleaning.
- Calculate the correlation coefficient between payload mass and landing outcomes
- Visualize the data using scatter plots or other suitable plots.
- Conduct statistical tests to determine if the correlation is significant.

Discussion (Contd): Cost Reduction and impact on future missions :

Cost reduction has a profound impact on the future of space missions in several ways:

- Increased Launch Frequency : Lower costs make it more feasible to launch rockets more frequently.
- Expanding Access to Space : As launch costs decrease, access to space becomes more affordable for a wider range of entities
- Advancement in Space Technology : Cost savings allow for more investment in research and development of new technologies
- Sustainable Space Exploration : Reusable rockets not only saves money but also promotes sustainability by minimizing waste and reducing the need for manufacturing new rockets for each mission
- Enabling Deep Space Missions : With savings from launch costs, more funds can be allocated to develop technologies required for deep space exploration.
- Boosting the Space Economy : A reduction in launch costs can spur economic growth in the space sector
- International Collaboration : Lower launch costs make it easier for countries to collaborate on space missions. Shared missions can pool resources and knowledge, leading to more successful and scientifically rich projects
- Improving Earth Observation and Communication : Cheaper launches mean more satellites can be placed into orbit.This can enhance weather forecasting, disaster response, and environmental monitoring

Results(Outcome):

- The data analysis provides valuable insights into SpaceX launches and the importance of data wrangling in predictive modeling.
- The data analysis from past launches can play a crucial role in enhancing the efficiency, reliability, and competitiveness of future space missions.
- The optimization of key points can guide future improvements and enhance predictive accuracy.
- Analyzing the correlation between payload mass and landing outcomes helps understand how payload weight might affect the landing success. This information can be valuable for optimizing future launches and ensuring higher success rates.
- Cost reduction in space missions enables more frequent launches, wider access, technological advancements, sustainability, ambitious deep space exploration, economic growth, international collaboration, and improved Earth services.
    - It's a game-changer that opens up a multitude of opportunities for the future of space exploration.

METHODOLOGY

Web scraping :

- BeautifulSoup and Request :
  - Use BeautifulSoup and requests to extract Falcon 9 and Falcon Heavy launch records from the Wikipedia page
- Data Extraction:
  - Extract HTML table data and convert it into a Pandas DataFrame.
- Data Parsing:
  - Create helper functions to process the extracted data and handle various table elements.

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10

Discussion :

- Request the HTML Page:
  - Use the requests library to get the HTML content of the Wikipedia page.
  - Create a BeautifulSoup object to parse the HTML content.

```
response = requests.get(static_url).text
soup = BeautifulSoup(response,"html.parser")
```

  - Use BeautifulSoup to find all tables and identify the relevant one containing launch records.
  - Extract column names from the table header.

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

Discussion (Contd) :

- Initialize a dictionary with extracted column names as keys and empty lists as values.
- Loop through each table row to extract data and append it to the corresponding lists in the dictionary.

```
column_names = []
for row in first_launch_table.find_all('th'):
    name = extract_column_from_header(row)
    if (name != None and len(name) > 0):
        column_names.append(name)
```

  - Check for missing values and handle them appropriately, such as replacing them with mean values or setting them to zero.
  - Export the cleaned DataFrame to a CSV file.

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

Result (Outcome) :

- The analysis provides a comprehensive guide to web scraping and data extraction using Python libraries such as BeautifulSoup and Pandas.
- By following the step-by-step instructions, one can effectively collect and analyze launch data from SpaceX, gaining insights into their historical launches and performance.

METHODOLOGY

Exploratory Data Analysis :

- Analyze the dataset to find patterns and insights.
- Handle missing values and identify data types.
- Determine Training Labels:
- Convert landing outcomes into binary labels (1 for successful landing, 0 for unsuccessful landing)

```
landing_outcomes=df['Outcome'].value_counts()
landing_outcomes
```

```
Outcome
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: count, dtype: int64
```

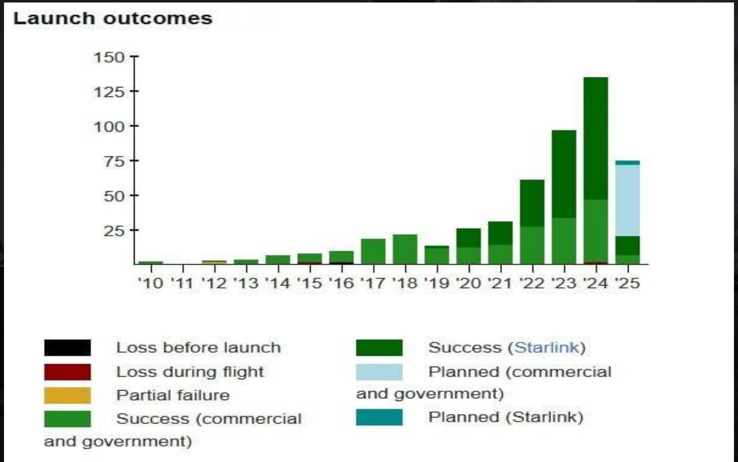
Discussion :

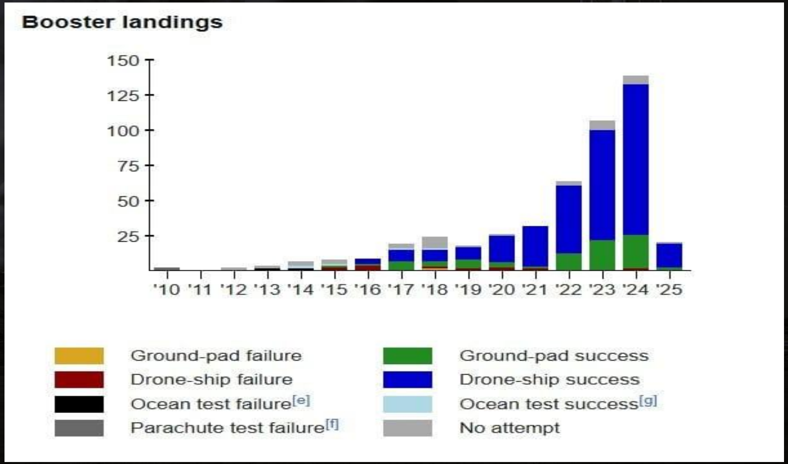
- The analysis involves cleaning and processing the data to prepare it for predictive modeling.
- Missing values are identified and handled appropriately.
- The dataset includes various launch sites and orbits, each with different characteristics and outcomes.
- By converting landing outcomes into binary labels and assign it to a new column Class, the dataset is prepared for supervised learning.

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0

Result (Outcome)

- This data analysis is a comprehensive guide to data wrangling, from loading and cleaning the dataset to preparing it for predictive modeling.
- It provides valuable insights into SpaceX Falcon 9 launches and the factors that influence landing success.





METHODOLOGY

EDA with SQL :

- The data analysis focuses on using SQL to analyze the SpaceX Falcon 9 and Falcon Heavy launch records from a dataset.
- Understand the SpaceX Dataset.
- Load the dataset into a corresponding table in a SQLite database.
- Execute SQL queries to answer specific assignment questions.

Discussion :

- Install necessary libraries such as sqlalchemy and ipython-sql.
- Establish a connection to the SQLite database using sqlite3.
- Load the SpaceX dataset from a provided URL and convert it into a Pandas DataFrame.
- Store the DataFrame into the SQL table.
- Execute various SQL queries to answer the assignment tasks.

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	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
2013-09-29	16:00:00	F9 v1.1 B1003	VAFB SLC-4E	CASSIOPE	500	Polar LEO	MDA	Success	Uncontrolled (ocean)
2013-12-03	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170	GTO	SES	Success	No attempt
2014-01-06	22:06:00	F9 v1.1	CCAFS LC-40	Thaicom 6	3325	GTO	Thaicom	Success	No attempt
2014-04-18	19:25:00	F9 v1.1	CCAFS LC-40	SpaceX CRS-3	2296	LEO (ISS)	NASA (CRS)	Success	Controlled (ocean)
2014-07-14	15:15:00	F9 v1.1	CCAFS LC-40	OG2 Mission 1 6 Orbcomm-OG2 satellites	1316	LEO	Orbcomm	Success	Controlled (ocean)
2014-08-05	8:00:00	F9 v1.1	CCAFS LC-40	AsiaSat 8	4535	GTO	AsiaSat	Success	No attempt
2014-09-07	5:00:00	F9 v1.1 B1011	CCAFS LC-40	AsiaSat 6	4428	GTO	AsiaSat	Success	No attempt

Result (Outcome) :

- Unique Launch Sites:
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40
- Five records where launch sites begin with the string 'CCA'.

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	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

Result (Outcome) :

- Total Payload Mass by NASA (CRS):
  - Total payload mass carried by boosters launched by NASA (CRS) is 45,596 kg

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

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER='NASA (CRS)'
```

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

sum(PAYLOAD_MASS__KG_)
45596

Result (Outcome) :

- Average Payload Mass by Booster Version F9 v1.1:
  - The average payload mass carried by booster version F9 v1.1 is 2928.4 kg.

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

```
%sql select avg(payload_mass__kg_) from spacextbl where Booster_Version = 'F9 v1.1'
```

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

avg(payload_mass__kg_)
2928.4

Result (Outcome) :

- First Successful Landing Outcome on Ground Pad:
  - The first successful landing outcome on the ground pad was achieved on 2015-12-22.

```
%sql select min(date) from spacextbl where Landing_Outcome = 'Success (ground pad)'
```

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

min(date)
2015-12-22

Result (Outcome) :

- Boosters with Success on Drone Ship and Payload Mass Between 4000 and 6000 kg:

```
%sql select booster_version from spacextbl where landing_outcome = 'Success (drone ship)' and payload_mass__kg_ >4000 and payload_mass__kg_<6000
```

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

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

Result (Outcome) :

- Total Number of Successful and Failure Mission Outcomes:
  - There are 100 total mission outcomes, including both successes and failures in flight.

```
%sql select count(Mission_Outcome) from spacextbl where Mission_Outcome = 'Success' or Mission_Outcome = 'Failure (in flight)'
```

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

count(Mission_Outcome)
100

Result (Outcome) :

- Boosters with Maximum Payload Mass:

```
%sql select Booster_Version from spacextbl where Payload_Mass__kg_=(select max(payload_mass__kg_) from spacextbl)
```

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

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

Result (Outcome) :

- Records of Failure Landing Outcomes in Drone Ship for 2015:

```
%sql select * from spacextbl where (Landing_Outcome = 'Failure' or Landing_Outcome = 'Failure (parachute)' or Landing_Outcome = 'Failure (drone ship)')
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2015-01-10	9:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)

- Landing outcomes ranked by count in descending order between the specified dates.

```
%sql SELECT Landing_Outcome, COUNT(*) AS QTY FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY QTY DESC
```

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

Landing_Outcome	QTY
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

METHODOLOGY

Data Visualization :

- Analyze and visualize key relationships between variables in the dataset.
- Identify patterns and trends that could impact the success of Falcon 9 first stage landings.
- Prepare and transform data for future predictive modeling.

	FlightNumber	PayloadMass	Flights	GridFins	Reused	Legs	Block	ReusedCount	Orbit_ES-L1	Orbit_GEO	...	Serial_B1048	Serial_B1049	Serial_B1050	Serial_B1051	Serial_B1054	Serial_B1056	Serial_B1058	Serial_B1059	Serial_B1060	Serial_B1062
0	1.0	6104.959412	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	2.0	525.000000	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	3.0	677.000000	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	4.0	500.000000	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	5.0	3170.000000	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

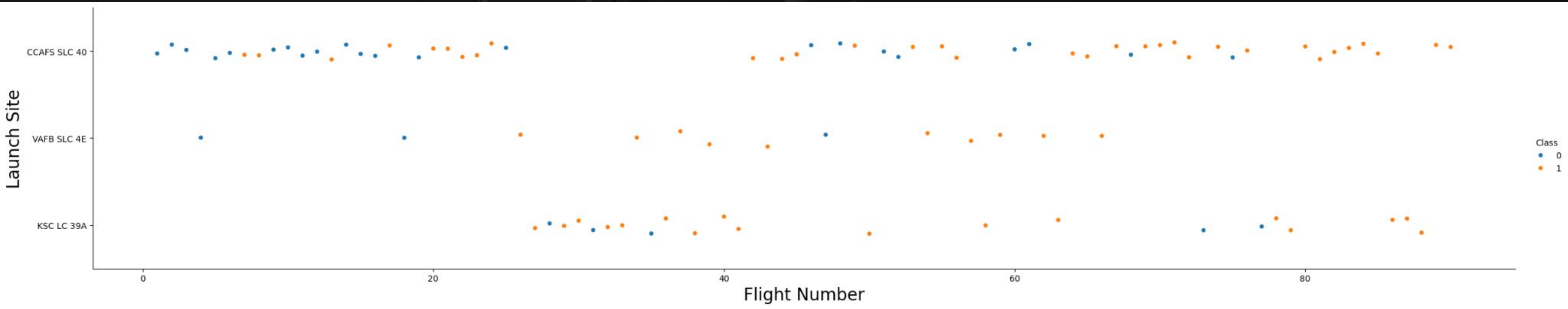
5 rows × 80 columns

Discussion :

- Import Libraries and Define Auxiliary Functions
- Loading the Dataset The dataset is loaded from a specified URL into a Pandas DataFrame
- A scatter plot is created to visualize the relationship between flight number, payload mass, and launch outcome shows that as the flight number increases, the likelihood of a successful landing increases
- A scatter plot is created to visualize the relationship between flight number and launch site, with the launch outcome (class) as the hue helps identify patterns in the launch site data
- A scatter plot is created to observe the relationship between payload mass and launch site helps determine if certain launch sites are associated with specific payload masses
- A bar chart is created to visualize the success rate for each orbit type helps identify which orbits have the highest success rates
- A scatter plot is created to observe the relationship between flight number and orbit type helps determine if the number of flights impacts the success rate for different orbits
- A scatter plot is created to observe the relationship between payload mass and orbit type helps determine if payload mass impacts the success rate for different orbits
- A line chart is created to visualize the yearly trend of the average success rate helps identify how the success rate has changed over time
- Create Dummy Variables for Categorical Columns
- One-hot encoding is applied to categorical columns such as Orbit, LaunchSite, LandingPad, and Serial. This transformation converts categorical variables into numerical format, making them suitable for predictive modeling
- Cast All Numeric Columns to Float64
- The features_one_hot DataFrame, which now contains only numerical values, is cast to the float64 data type. This ensures consistency in data types for future modeling.

Result (Outcome)

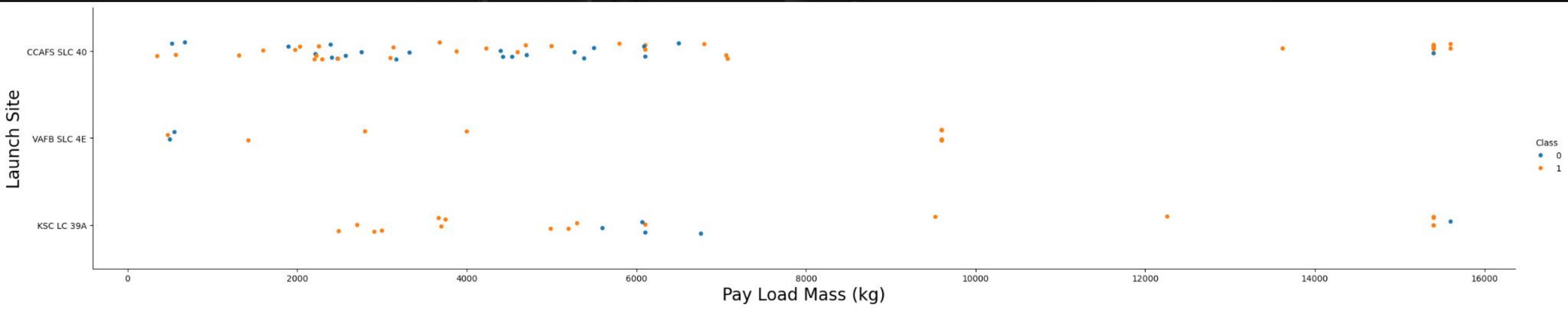
- The relationship between flight number and launch site



By observation,CCAFS SLC 40 appears to be the most of early 1st stage failure took place

Result (Outcome)

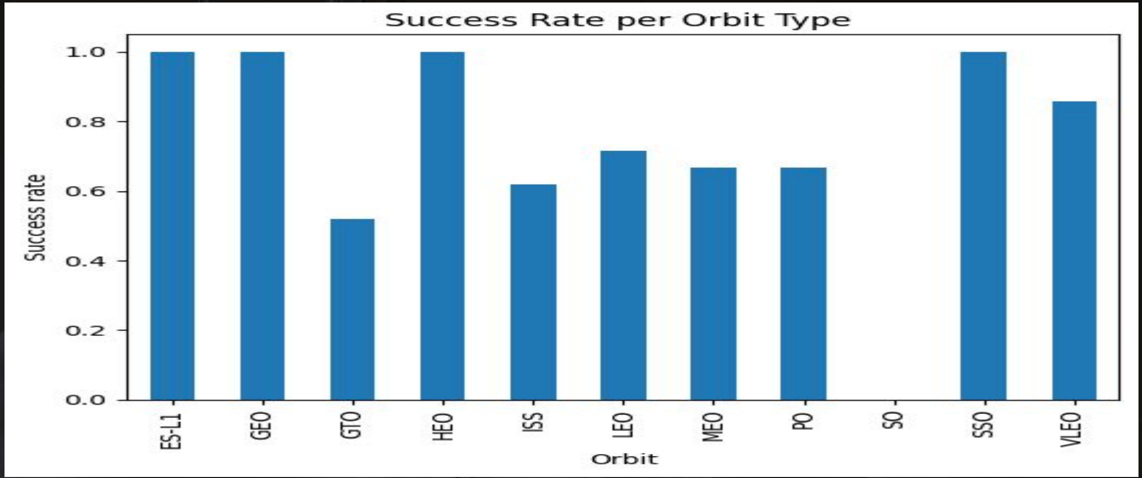
- ○ The relationship between payload and launch site



CCAFS SLC 40 and KSC LC 39 A are in favour of heavy payloads

Result (Outcome)

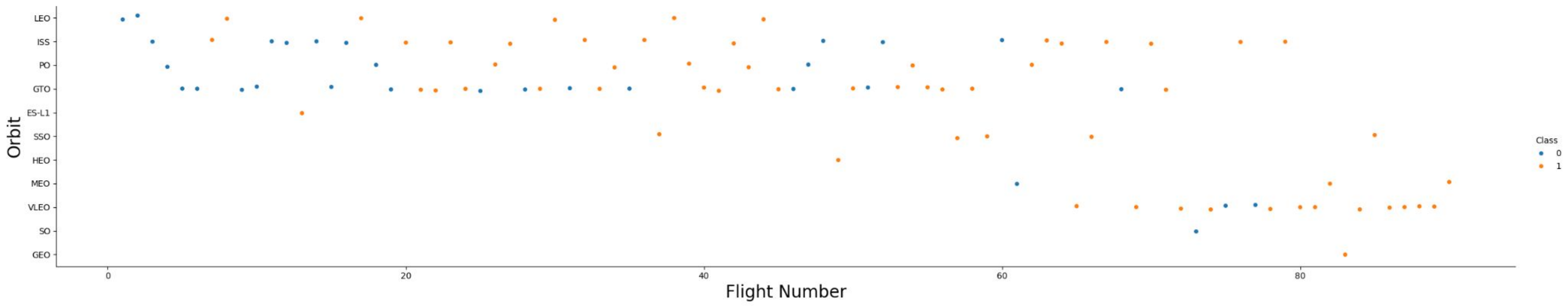
- The relationship between success rate and orbit type



Orbits other than SO have successful 1st stage landing

Result (Outcome)

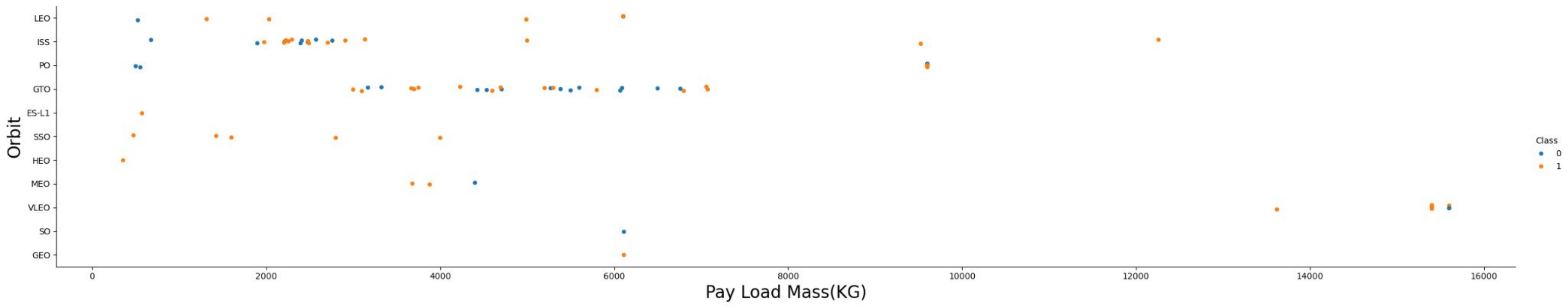
- The relationship between flight number and orbit type



Orbit between LEO and GTO are showing results proportional with the flight number

Result (Outcome)

- The relationship between payload and orbit type

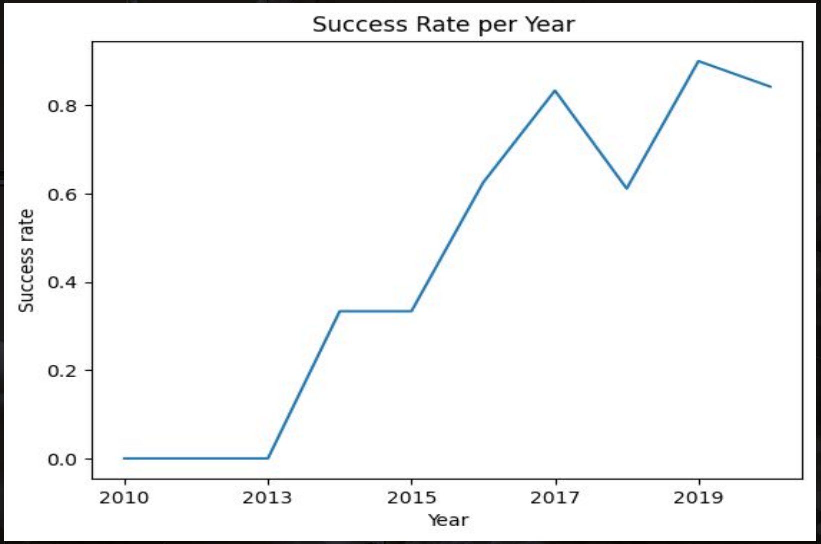


With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Result (Outcome)

- Launch Success Yearly Trend



The success rate since 2013 kept increasing till 2020

METHODOLOGY

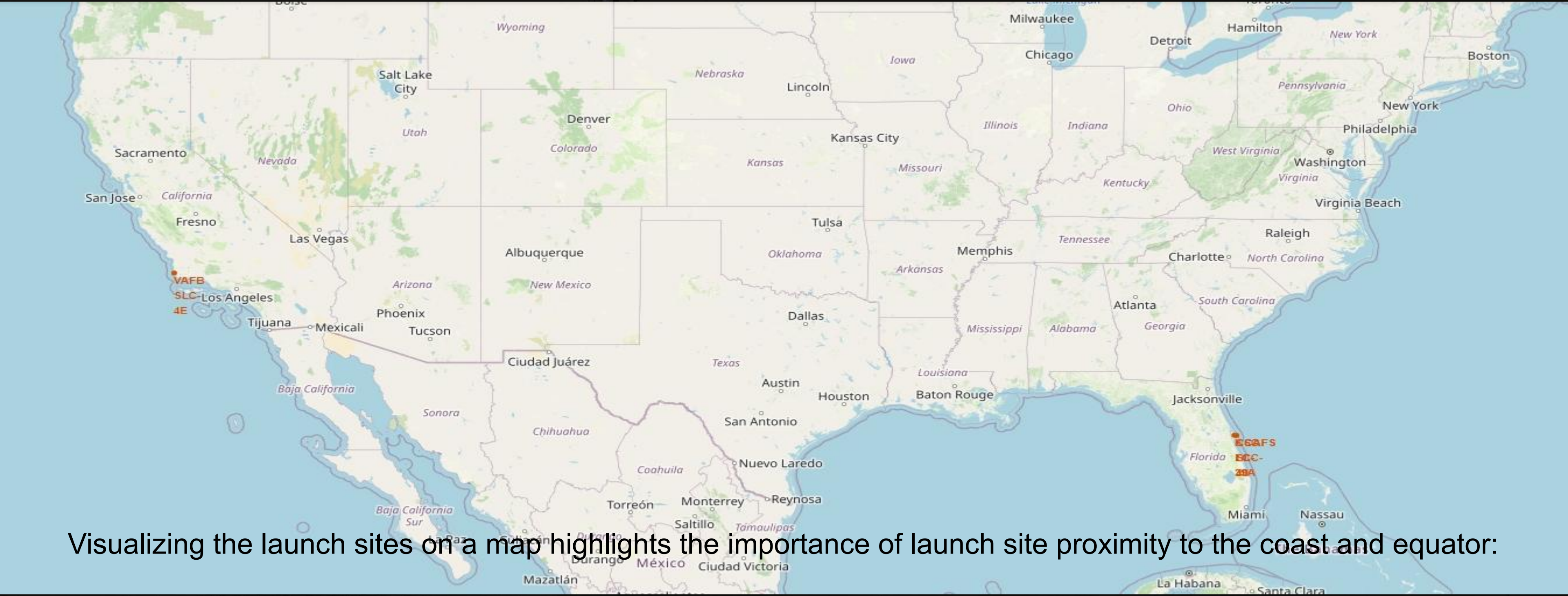
Data Visualization using Folium :

- Analyze and visualize key relationships between variables in the dataset.
- Identify patterns and trends that could impact the success of Falcon 9 first stage landings.
- Prepare and transform data for future predictive modeling.

Discussion :

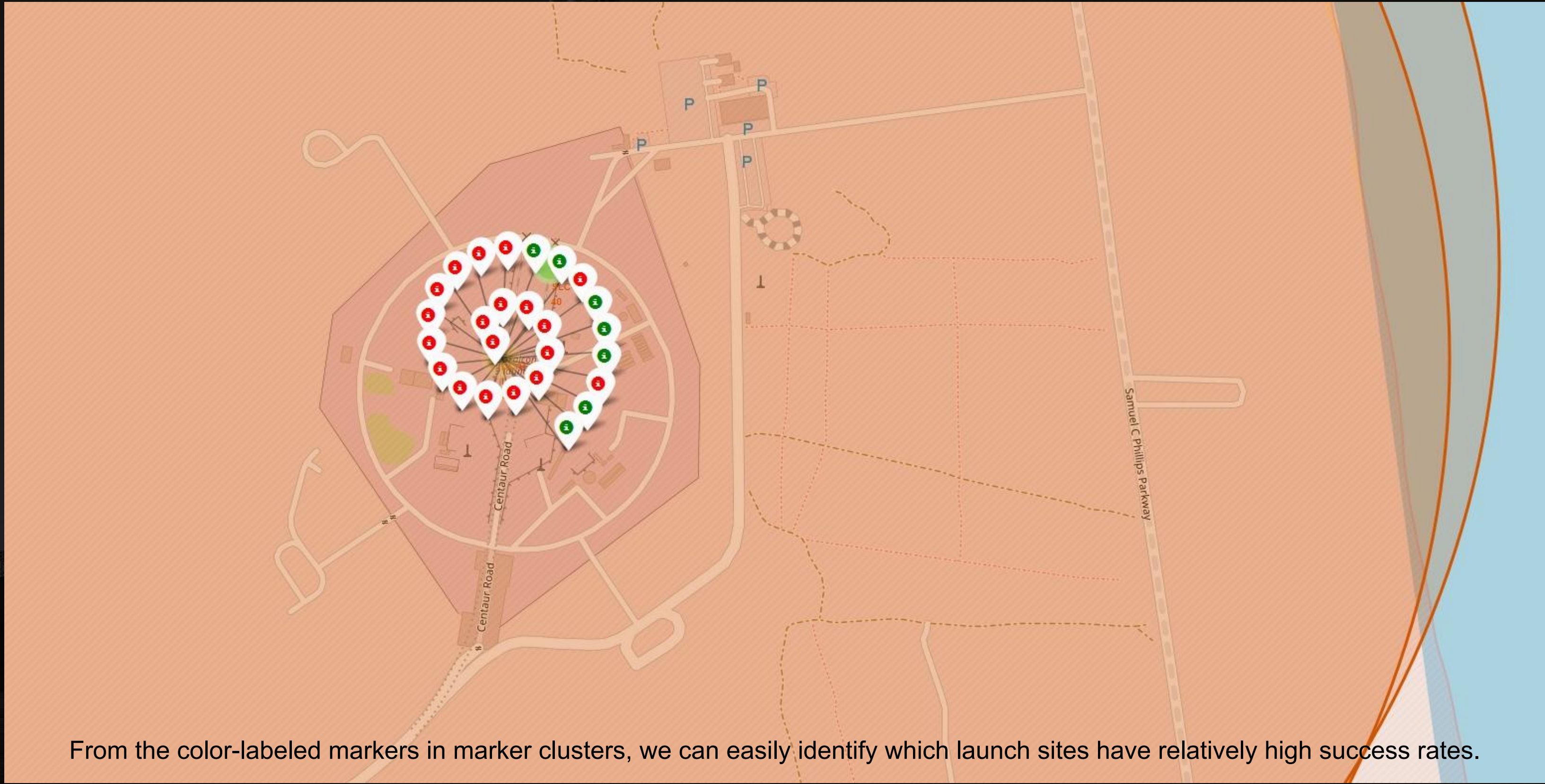
- Loading the Dataset The dataset is loaded from a specified URL into a Pandas DataFrame
- A scatter plot is created to visualize the relationship between flight number, payload mass, and launch outcome shows that as the flight number increases, the likelihood of a successful landing increases
- A scatter plot is created to visualize the relationship between flight number and launch site, with the launch outcome (class) as the hue helps identify patterns in the launch site data
- A scatter plot is created to observe the relationship between payload mass and launch site helps determine if certain launch sites are associated with specific payload masses
- A bar chart is created to visualize the success rate for each orbit type helps identify which orbits have the highest success rates
- A scatter plot is created to observe the relationship between flight number and orbit type helps determine if the number of flights impacts the success rate for different orbits
- A scatter plot is created to observe the relationship between payload mass and orbit type helps determine if payload mass impacts the success rate for different orbits
- A line chart is created to visualize the yearly trend of the average success rate helps identify how the success rate has changed over time
- Create Dummy Variables for Categorical Columns and Cast All Numeric Columns to Float64

Result (Outcome) :



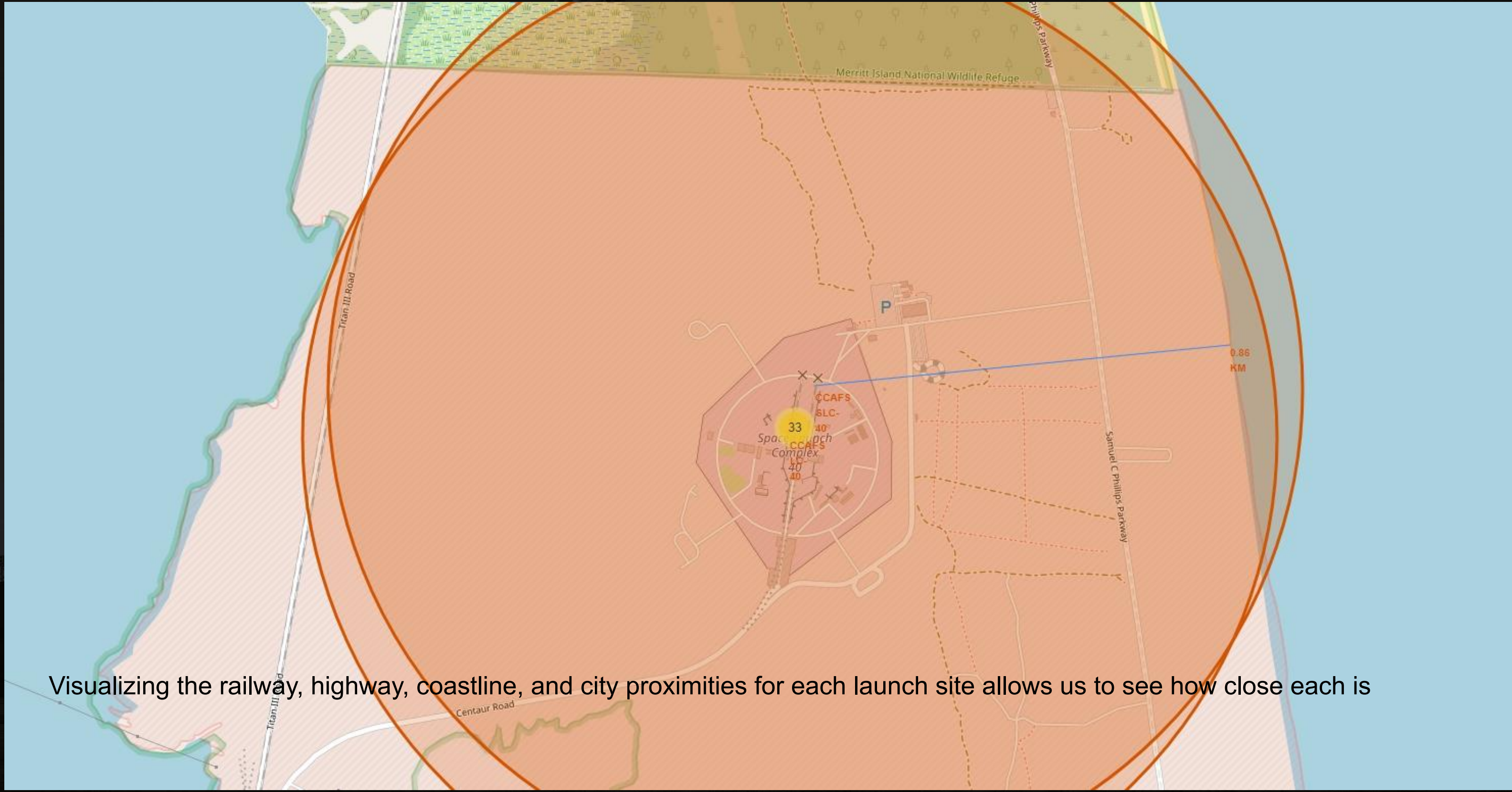
Visualizing the launch sites on a map highlights the importance of launch site proximity to the coast and equator:

Result :



From the color-labeled markers in marker clusters, we can easily identify which launch sites have relatively high success rates.

Result :



Visualizing the railway, highway, coastline, and city proximities for each launch site allows us to see how close each is

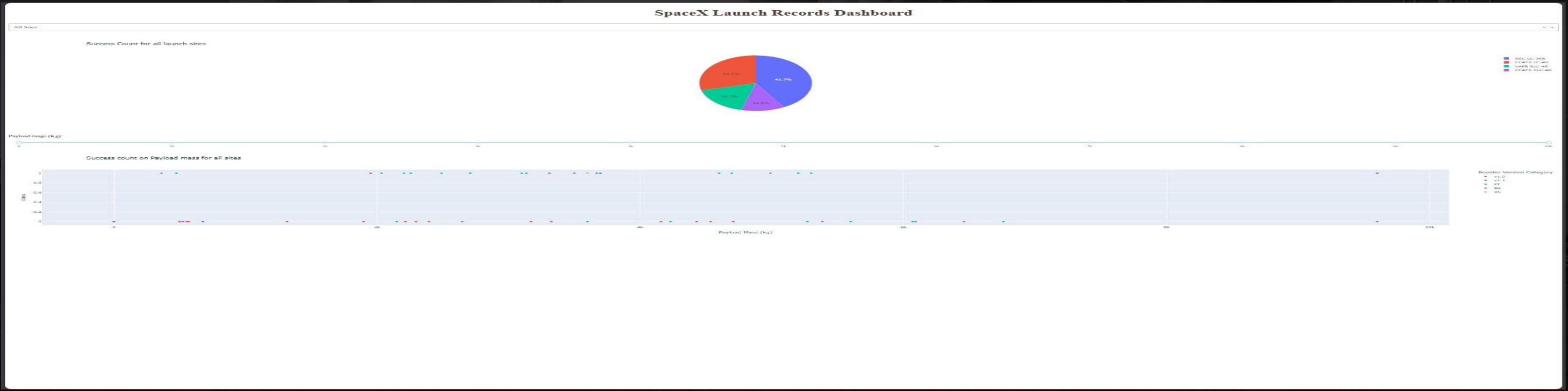
Result :


Cochran Cove
Futch Cove
Broadaxe Creek
Ross Creek
Complex 39A
Kennedy Parkway North
Banana Creek
False Cape
Conrad Creek
East Conrad Creek
Tea Creek
Saturn Barge Channel
Roberts
6.61 KM
13
LC 39A
33
CCAFS
LC 40

METHODOLOGY

Dashboard using plotly :

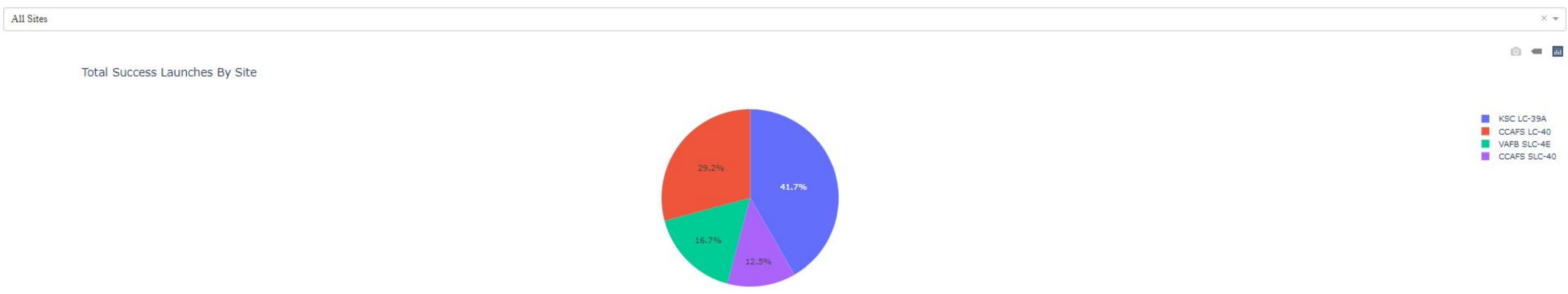
- Added a Launch Site Drop-down Input Component
- Added a callback function to render success-pie-chart based on selected site dropdown
- Added a Range Slider to Select Payload
- Added a callback function to render the success-payload-scatter-chart scatter plot



Discussion :

- Install python packages required to run the application
- Add a Launch Site Drop-down Input Component
- Add a callback function to render success-pie-chart based on selected site dropdown
- Observed if variable payload is correlated to mission outcome it is easy to select different payload range and identify some visual patterns.
- Plot a scatter plot with the x axis to be the payload and the y axis to be the launch outcome to observe how payload may be correlated with mission outcomes for selected site(s) and color-label the Booster version on each scatter point so that we may observe mission outcomes with different boosters.

Result :

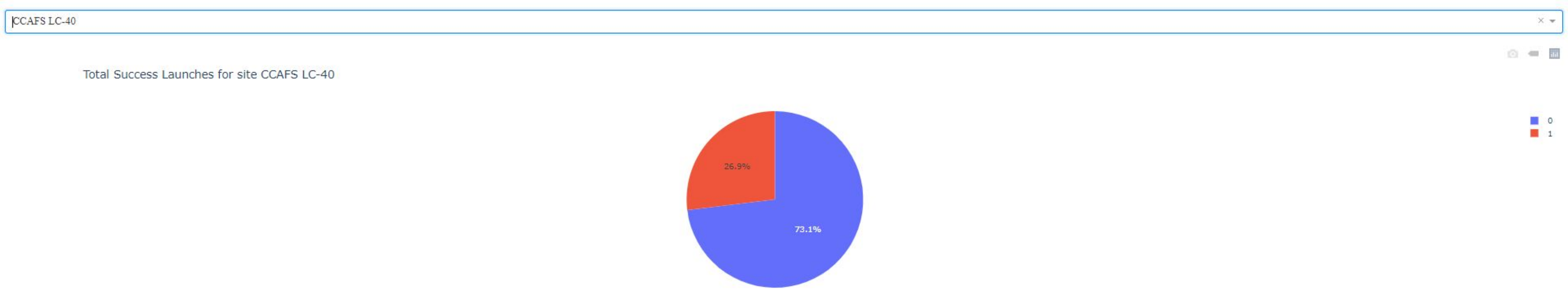


Pie chart showing booster landing success rate Drop down menu to choose between all sites and individual launch sites

Color coded by launch site

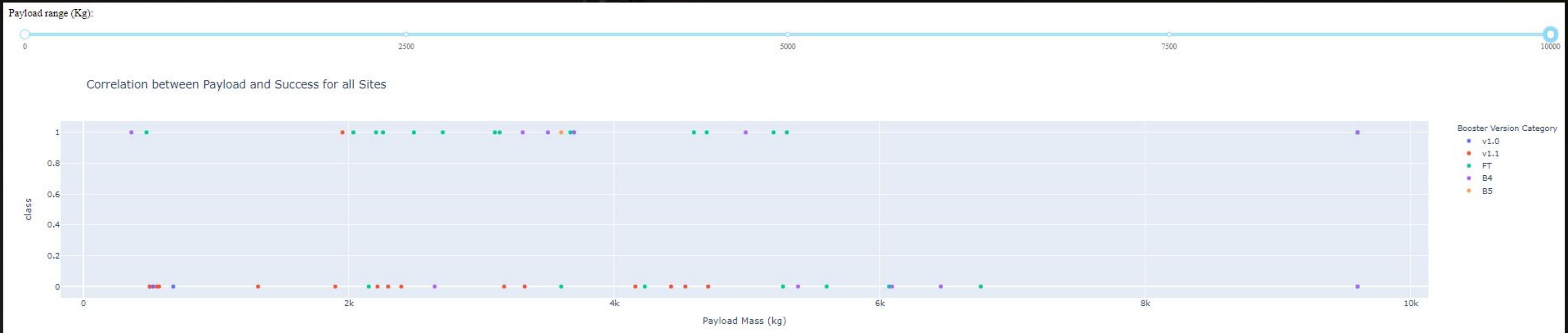
- KSC LC-39A
- CCAFS LC-40
- VAFB SLC-4E
- CCAFS SLC-40

Result :



- The picture below shows a pie chart when launch site CCAFS LC-40 is chosen.
- 0 represents failed launches while 1 represents successful launches. We can see that 73.1% of launches done at CCAFS LC-40 are failed launches

Result :



- The picture below shows a scatterplot when the payload mass range is set to be from 2000 kg to 8000 kg.
- Class 0 represents failed launches while class 1 represents successful launches

METHODOLOGY

plotly

Machine Learning Prediction :

- Perform Exploratory Data Analysis and Determine Training Labels
- Standardize the Data
- Split Data into Training and Testing Sets
- Find the Best Hyperparameters for Various Models
- Evaluate Model Performance on Test Data

	FlightNumber	PayloadMass	Flights	Block	ReusedCount	Orbit_ES-L1	Orbit_GEO	Orbit_GTO	Orbit_HEO	Orbit_ISS	...	Serial_B1058	Serial_B1059	Serial_B1060	Seria
0	1.0	6104.959412	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	
1	2.0	525.000000	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	
2	3.0	677.000000	1.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	...	0.0	0.0	0.0	
3	4.0	500.000000	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	
4	5.0	3170.000000	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
85	86.0	15400.000000	2.0	5.0	2.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0	
86	87.0	15400.000000	3.0	5.0	2.0	0.0	0.0	0.0	0.0	0.0	...	1.0	0.0	0.0	
87	88.0	15400.000000	6.0	5.0	5.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	
88	89.0	15400.000000	3.0	5.0	2.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0	
89	90.0	3681.000000	1.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	

90 rows × 83 columns

Discussion :

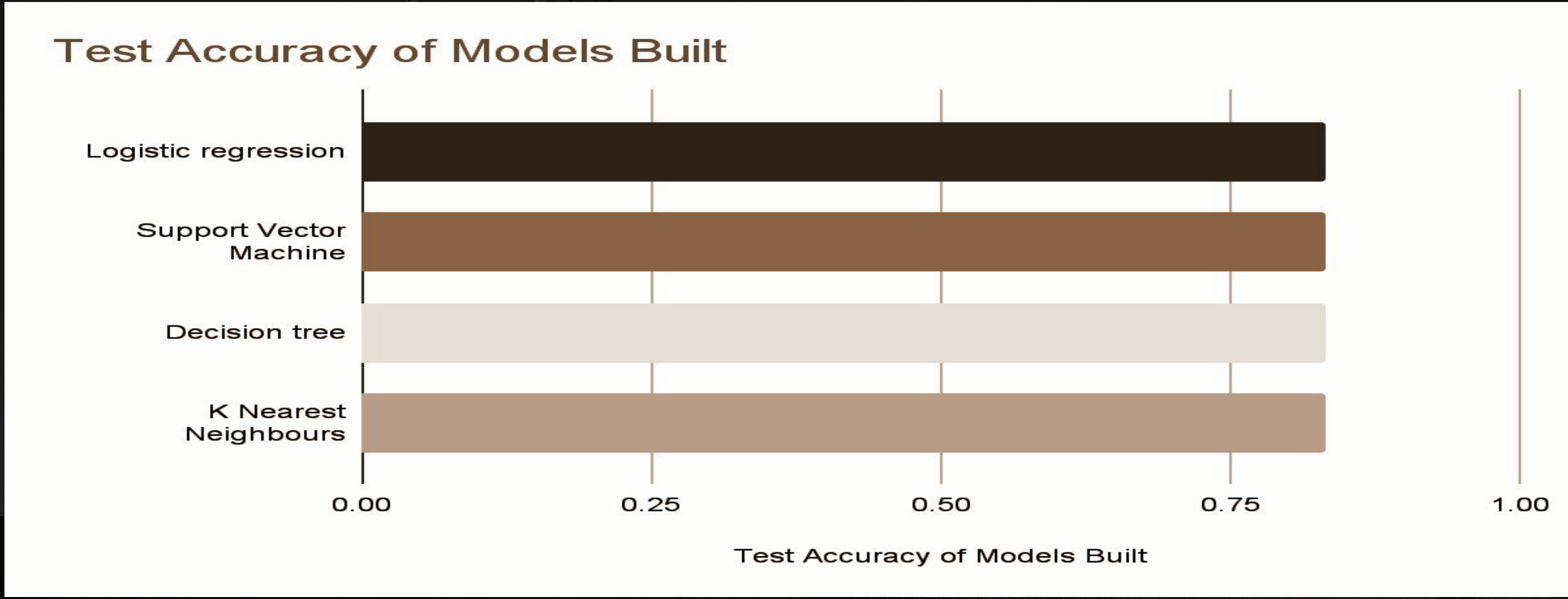
- Import Libraries and Define Auxiliary Functions
- The data is loaded from provided URLs and converted into Pandas DataFrames
- A NumPy array is created from the Class column in the data DataFrame
- The data in X is standardized using StandardScaler
- Split the Data into Training and Testing Sets
- The model created and tuned using GridSearchCV with an accuracy of 83.333%:
  - Logistic regression
  - Support Vector Machine
  - Decision Tree
  - K-Nearest Neighbours

Result :

- All four models (Logistic Regression, SVM, Decision Tree, KNN) have shown similar accuracy on the test data (~83.333%).
- Hyperparameter tuning through GridSearchCV has helped optimize the models for better performance.
- The confusion matrix for each model indicates that while they are good at predicting successful landings, there are some false positives (predicting a landing when there was none).
- The models have been trained and evaluated, showing promising accuracy in predicting the success of Falcon 9 first stage landings.
- Further improvements can be made by experimenting with different models, feature engineering, and data augmentation

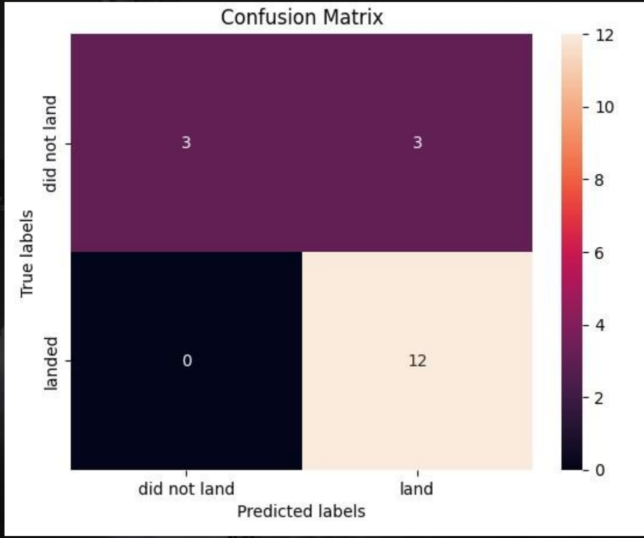
Result :

Classification Accuracy :



Result :

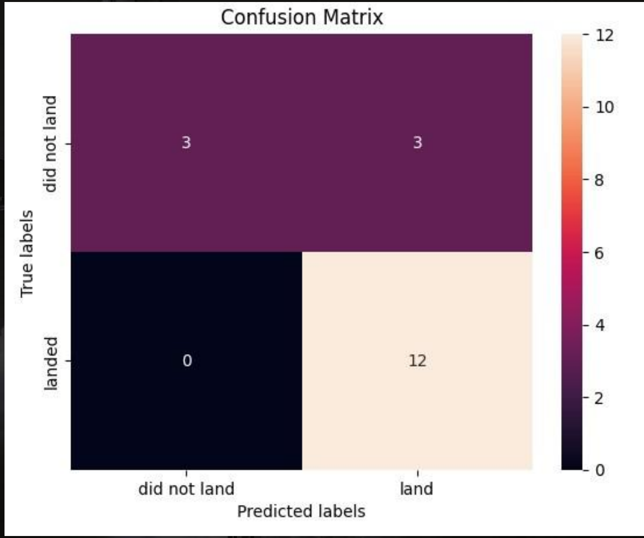
Confusion Matrix for Linear Regression :



GridSearchCV best score : 0.8464285714285713
Accuracy score on test :0.8333333333333334

Result :

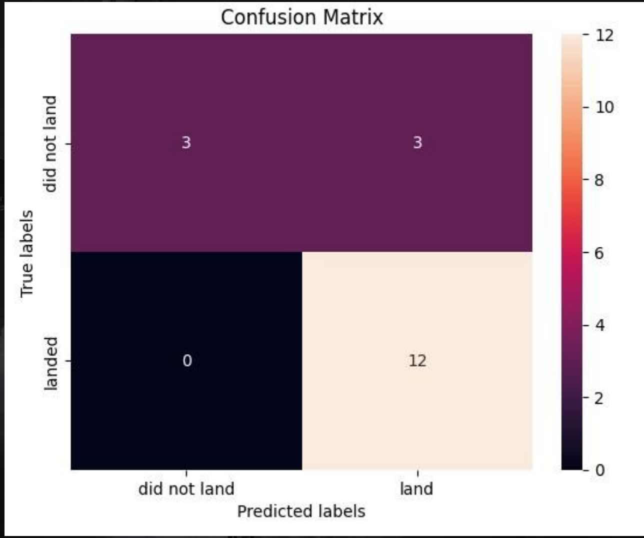
Confusion Matrix for Support Vector Machine:



GridSearchCV best score : 0.8482142857142856
Accuracy score on test :0.8333333333333334

Result :

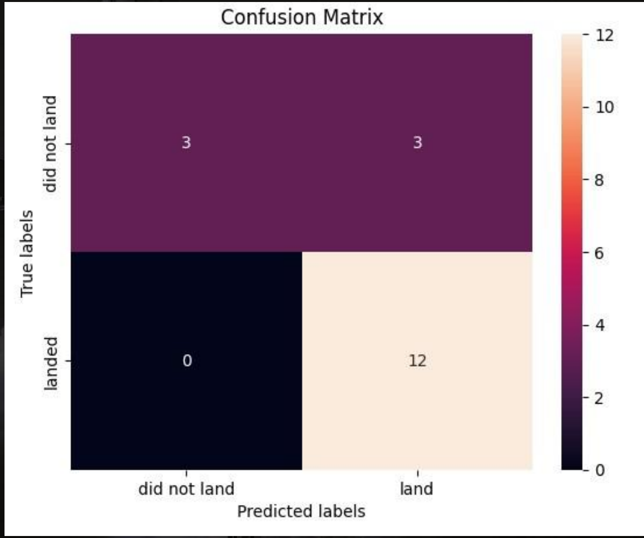
Confusion Matrix for Tree:



GridSearchCV best score : 0.8607142857142858
Accuracy score on test :0.8333333333333334

Result :

Confusion Matrix for K-Nearest Neighbours:



GridSearchCV best score : 0.8482142857142858
Accuracy score on test :0.8333333333333334

Suggestion : Satellite Launch Risk Assessment Tool

Considering the advancements SpaceX has achieved with their Falcon 9 landings, it could be valuable to develop a Satellite Launch Risk Assessment Tool. This tool would help SpaceX and other aerospace companies better understand and mitigate risks associated with satellite launches. Here are some key points:

- Risk Identification : Use historical launch data to identify common risks, such as payload issues, weather conditions, and equipment malfunctions.
- Predictive Modeling : Implement machine learning models to predict potential risks for upcoming launches based on factors like launch site, payload mass, and weather forecasts.
- Visualization Dashboard : Create an interactive dashboard that displays risk levels for each scheduled launch. This dashboard could include visualizations of risk factors and their impact on mission success.
- Recommendations : Provide actionable recommendations for mitigating identified risks. For instance, suggest alternative launch windows or additional checks for high-risk payloads.
- Continuous Improvement : Continuously update the tool with data from new launches to improve its predictive accuracy and reliability.
- This tool could be a valuable asset for improving mission success rates, reducing costs, and enhancing safety in the aerospace industry. It would also demonstrate the practical application of data science and machine learning beyond predicting Falcon 9 landings.

CONCLUSION

- Utilize data science and machine learning to predict the success of SpaceX Falcon 9 first stage landings, aiming to reduce costs and enhance strategic planning in the aerospace industry
- Data Collection and Preparation : Gathered and cleaned historical launch data, including flight numbers, dates, launch sites, payload masses, orbits, and landing outcomes. Performed feature engineering by converting categorical variables using one-hot encoding and standardized numerical features for model readiness.
- Exploratory Data Analysis (EDA) :
    - Conducted analysis to identify key patterns:
        - Higher flight numbers and specific payload masses correlate with increased landing success.
        - Certain launch sites, like KSC LC 39A, show higher success rates.
        - Missions to Low Earth Orbit (LEO) have higher landing success rates.
        - An upward trend in success rates over the years indicates technological advancements.
- Model Development and Evaluation : Trained multiple machine learning models—including Logistic Regression, Support Vector Machines (SVM), Decision Trees, and K-Nearest Neighbors (KNN)—using GridSearchCV for hyperparameter tuning. All models achieved approximately 83% accuracy on test data, with the Decision Tree model slightly outperforming others.
- Significant Features : Launch site, orbit type, and payload mass are critical factors influencing landing success.
- Predictive Consistency : High accuracy across models indicates robust predictive capability.
- Industry Implications : Predictive models can optimize resource allocation, reduce operational costs, and provide a competitive edge in the aerospace sector.
- The project successfully demonstrates the power of data science and machine learning in predicting SpaceX Falcon 9 first stage landing outcomes, offering valuable insights for cost optimization and strategic decision-making in future space missions.

Appendix

- Reference Links :
  - Data Collection
  - Data Preparation
  - Data Wrangling
  - EDA with visualization
  - EDA with SQL
  - EDA with Folium
  - Dashboard using Plotly
  - Predictive Analysis by Machine Learning