IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Teji Singh
1st Sept 2023

Outline

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

Executive Summary

- Summary of methodologies (on SpaceX Data)
    a. Data Collection using SpaceX API
    b. Data Collection with Web Scraping
    c. Data Wrangling
    d. Exploratory Data Analysis using SQL
    e. EDA DataViz Using Python Pandas and Matplotlib
    f. SpaceX Launch Sites Analysis with Folium and Creating Interactive Visual Analytics Dashboards using Dash Modules
    g. SpaceX Machine Learning Landing Prediction

- Summary of all results
    a. EDA Results
    b. Interactive Visual Analytics Dashboards
    c. Predictive Analytics

Introduction

- **Project background and context**

  SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

- **Problems you want to find answers**

  In this capstone, we will predict if the Falcon 9 first stage will land successfully.

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

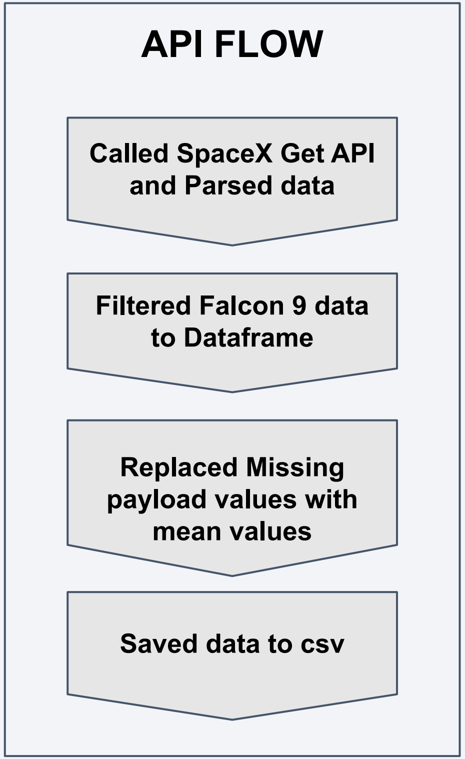
Data Collection

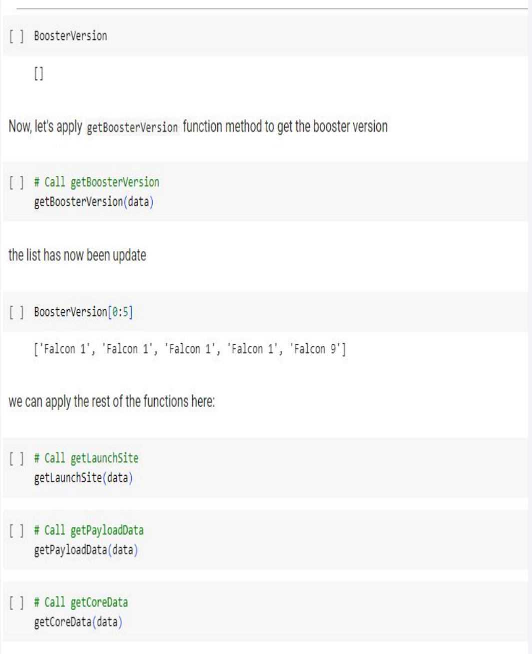
- **SpaceX API Data Retrieval**:
    a. Utilized SpaceX's RESTful API, employing helper functions for data extraction.
    b. Made GET requests to the API, decoded JSON results, and converted them into a Pandas dataframe.
- **Web Scraping for Falcon 9 Launch Records**:
    a. Conducted web scraping on a Wikipedia page titled "List of Falcon 9 and Falcon Heavy launches."
    b. Used BeautifulSoup and request libraries to extract, parse, and transform HTML table records into a Pandas dataframe.
- **Data Compilation and Analysis**:
    a. Combined data from SpaceX API and web scraping to create a comprehensive dataset.
    b. This dataset was ready for analysis, manipulation, and visualization as required for the project's objectives.

Data Collection – SpaceX API

- The data was acquired through SpaceX's RESTful API by initiating a GET request to their server. Subsequently, the SpaceX launch data was retrieved, parsed, and decoded from the response content, resulting in a JSON format. This JSON data was then converted into a Pandas dataframe for further processing and analysis.

- *Link*: [https://github.com/TjRathore/IBM-Data Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb](https://github.com/TjRathore/IBM-Data%20Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb)





Data Collection - Scraping

- Here, I Conducted web scraping to acquire historical Falcon 9 launch records from a Wikipedia page using BeautifulSoup and request libraries. Then, Extracted the Falcon 9 launch records from an HTML table within the Wikipedia page and created a dataframe by parsing the launch HTML.

- *Link*: https://github.com/TjRathore/IBM-DataScience-Capstone/blob/main/jupyter-labs-webscraping.ipynb

Web Scraping Flow

Extract a Falcon 9 launch records HTML table from Wikipedia

Parse the table and convert it into a Pandas dataframe

Replaced Missing payload values with mean values

Saved data to csv

```
# Let's print the third table and check its content
first_launch_table = html_tables[2]
print(first_launch_table)

<table class="wikitable plainrowheaders collapsible" style="width: 100%;">
<tbody><tr>
<th scope="col">Flight No.
</th>
<th scope="col">Date andtime ()
</th>
<th scope="col">
</th>
<th scope="col">Launch site
</th>
<th scope="col">Payload
</th>
<th scope="col">Payload mass
</th>
<th scope="col">Orbit
</th>
<th scope="col">Customer
</th>
<th scope="col">Launchoutcome
</th>
<th scope="col">
</th></tr>
<tr>
<th rowspan="2" scope="row" style="text-align:center;">1
</th>
<td>4 June 2010,<br/>18:45
</td>
<td><a href="/wiki/Falcon_9_v1.0" title="Falcon 9 v1.0">F9 v1.0</a><sup class="reference" id="cite_ref-MuskMay2012_13-0"><a href="#cite_note-MuskMay2012-13">
</td>
<td><a href="/wiki/Cape_Canaveral_Space_Force_Station" title="Cape Canaveral Space Force Station">CCAFS</a>,<br/><a href="/wiki/Cape_Canaveral_Space_Launch_C
</td>
<td><a href="/wiki/Dragon_Spacecraft_Qualification_Unit" title="Dragon Spacecraft Qualification Unit">Dragon Spacecraft Qualification Unit</a>
</td>
<td>
</td>
<td><a href="/wiki/Low_Earth_orbit" title="Low Earth orbit">LEO</a>
</td>
```

Data Wrangling

- Data was initially filtered within the Pandas DataFrame to retain only Falcon 9 launches based on the BoosterVersion column. Additionally, missing values in the LandingPad and PayloadMass columns were addressed.

- Exploratory Data Analysis (EDA) was conducted to identify data patterns and determine a suitable label for training supervised models. For PayloadMass, missing values were replaced with the mean column value during this process.

Link: https://github.com/TjRathore/IBM-DataScience-Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_1_L3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

Calculated*:*

- ***Number of launches on each site***
- ***number and occurrence of each orbit***
- ***number and occurrence of missions outcome per orbit type***

```
[ ]  df['Class']=landing_class
     df[['Class']].head(8)
```

	Class
0	0
1	0
2	0
3	0
4	0
5	0
6	1

Created Feature: 'Class'
Class = 0: booster landing failure
Class = 1: booster landing success

EDA with Data Visualization

1. Performed data Analysis and Feature Engineering using Pandas and Matplotlib .i.e.

    a. Exploratory Data Analysis

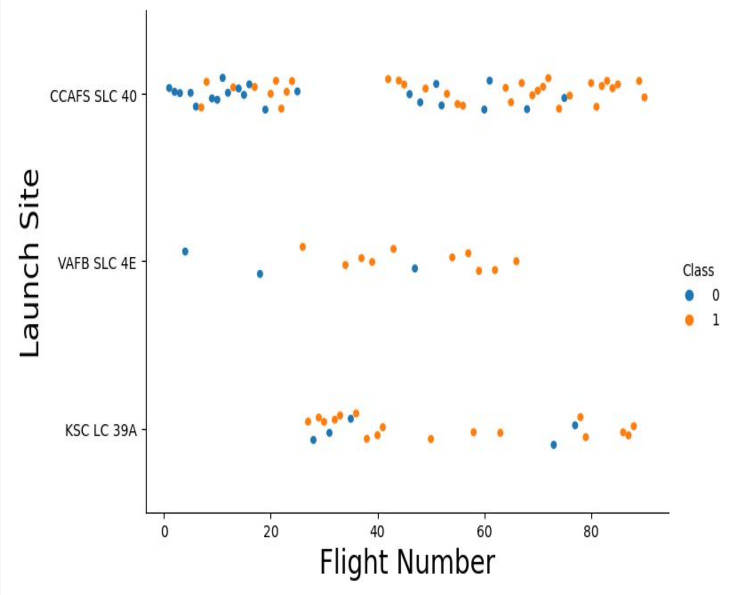
    b. Preparing Data Feature Engineering

2. Used scatter plots to Visualize the relationship between Flight Number and Launch Site, Payload and Launch Site, FlightNumber and Orbit type, Payload and Orbit type.

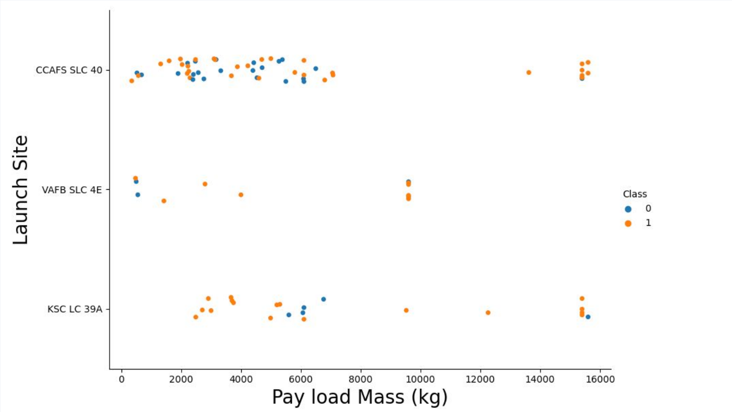
3. Used Bar chart to Visualize the relationship between success rate of each orbit type

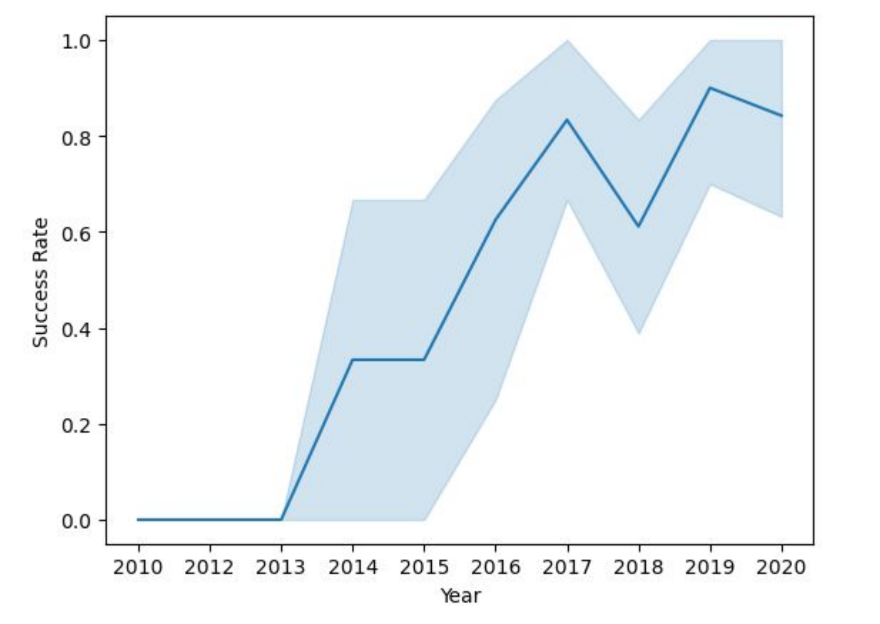
4. Line plot to Visualize the launch success yearly trend.

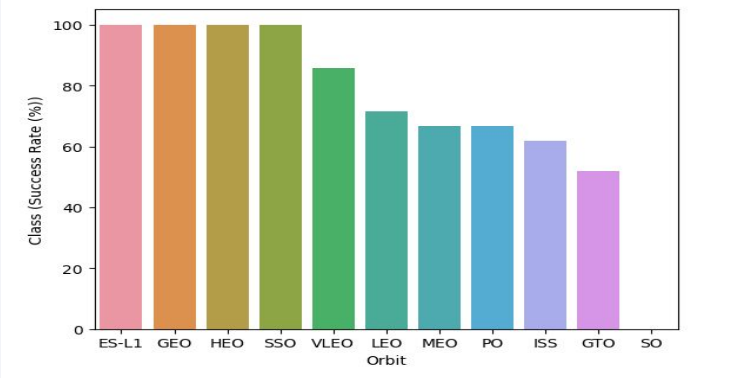
Link:
https://github.com/TjRathore/IBM-DataScience-Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb



EDA with Data Visualization *(Contd…)*







EDA with SQL

- The following SQL queries were performed for EDA:

```
[ ]  %sql SELECT DISTINCT LAUNCH_SITE as "Launch Sites" FROM SPACEXTBL;

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

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

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
[ ]  %sql SELECT MIN(DATE) FROM SPACEXTBL WHERE Landing_Outcome = "Success (ground pad)";
```

EDA with SQL

Task 2

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

```
[ ] %sql SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

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

```
[ ] %sql SELECT SUM(PAYLOAD_MASS__KG_) as "Total Payload Mass(Kgs)", Customer FROM SPACEXTBL WHERE Customer = 'NASA (CRS)';
```

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

```
[ ] %sql SELECT DISTINCT Booster_Version, Payload FROM SPACEXTBL WHERE Landing_Outcome = "Success (drone ship)" AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

EDA with SQL

List the total number of successful and failure mission outcomes

```
[ ]  %sql SELECT Mission_Outcome, COUNT(Mission_Outcome) as Total FROM SPACEXTBL GROUP BY Mission_Outcome;
```

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

```
[ ]  %sql SELECT Booster_Version,Payload, PAYLOAD_MASS__KG_ FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.

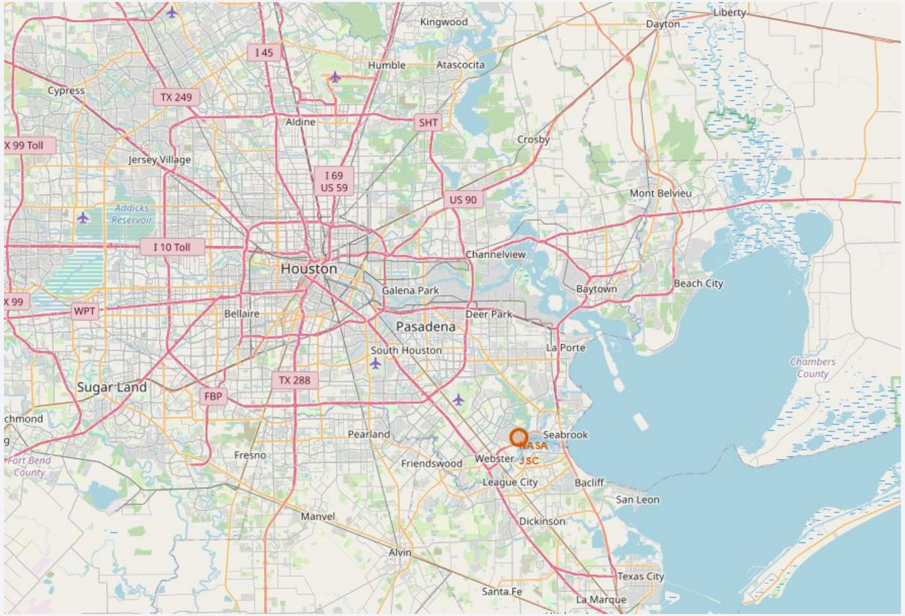
```
[ ]  %sql SELECT substr(Date, 4, 1) as 'Month', Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTBL WHERE substr(Date, 1, 4)='2015' AND Landing_Outcome = 'Failure (drone ship)';
```

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

```
[ ]  %sql SELECT Landing_Outcome, count(Landing_Outcome) as Landing_Count FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' group by Landing_Outcome ORDER BY count(Landing_Outcome) DESC;
```

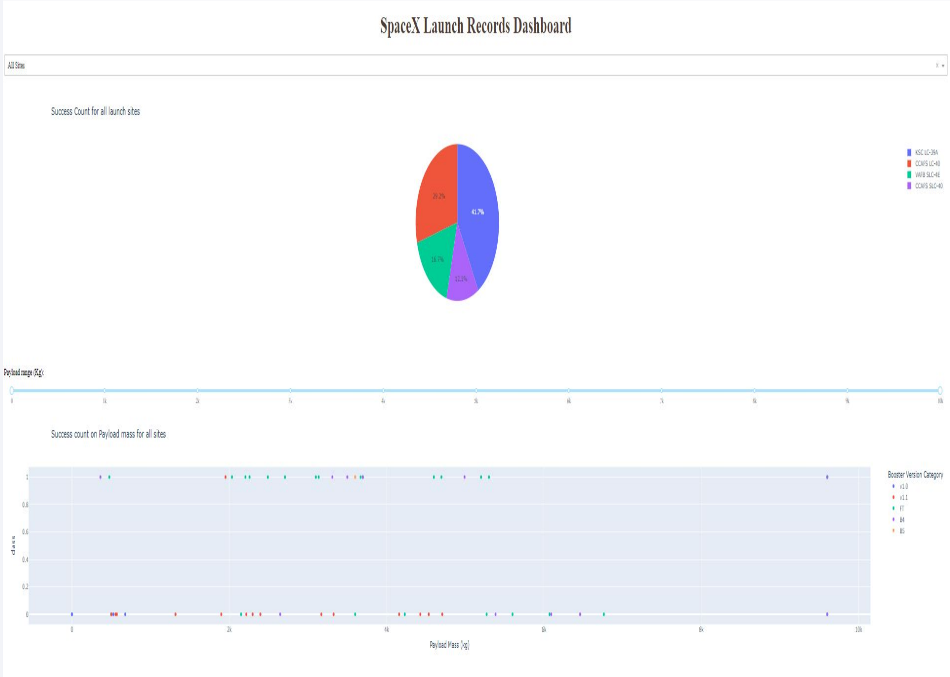
Build an Interactive Map with Folium

- Created folium map to marked all the launch sites, and created map objects such as markers, circles, lines to mark the success or failure of launches for each launch site.
- Created a launch set outcomes (failure=0 or success=1).
- *Link*: https://github.com/TjRathore/IBM-DataScience-Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite.ipynb



Build a Dashboard with Plotly Dash

- Built an interactive dashboard application with Plotly dash by:
    a. Adding a Launch Site Drop-down Input Component
    b. Adding a callback function to render success-pie-chart based on selected site dropdown
    c. Adding a Range Slider to Select Payload
    d. Added a callback function to render the success-payload-scatter-chart scatter plot.
- *Link*: https://github.com/TjRathore/IBM-DataScience-Capstone/blob/main/spacex_dash_app.py
- Screenshots can be found here.



Predictive Analysis (Classification)

Here is a summary of the process involved in building, evaluating, improving, and finding the best-performing classification model:

- **Data Loading and Exploration**:

    a. Initially, the data was loaded into a Pandas DataFrame for analysis.
    b. Exploratory Data Analysis (EDA) was conducted to gain insights into the dataset.

- **Training Labels Determination**:

    a. To define the training labels, a NumPy array was created from the 'Class' column using the to_numpy() method. This array was assigned to the variable 'Y' as the outcome variable.

- **Feature Engineering**:

    a. The feature dataset (denoted as 'x') was standardized to ensure uniformity in the data. This standardization was performed using the preprocessing.StandardScaler() function from the Scikit-Learn library.

- **Data Splitting**:

    a. The dataset was split into training and testing sets using the train_test_split function from Scikit-Learn's model_selection module.
    b. The 'test_size' parameter was set to 0.2, meaning that 20% of the data was reserved for testing.
    c. A 'random_state' value of 2 was specified for reproducibility.

Contd…

To determine the optimal machine learning model among SVM, Classification Trees, k-nearest neighbors, and Logistic Regression, the following approach was employed:

- **Model Initialization and Hyperparameter Tuning:**
    a. Initially, individual objects were created for each of the machine learning algorithms.
    b. Subsequently, a GridSearchCV object was established for each model, encompassing a predefined set of hyperparameters for tuning.
- **Hyperparameter Optimization:**
    a. Each model was subjected to hyperparameter optimization through the GridSearchCV object.
    b. The optimization process utilized a cross-validation (cv=10) approach on the training dataset to identify the best hyperparameters.
- **Parameter Assessment and Validation:**
    a. Following the hyperparameter tuning process, the results from GridSearchCV were examined for each model.
    b. The best parameters were extracted using the 'best_params_' attribute, and the corresponding accuracy on the validation data was obtained via the 'best_score_' attribute.
- **Model Evaluation on Test Data:**
    a. The performance of each model was assessed using the 'score' method, which calculated the accuracy on the test dataset.
    b. Additionally, confusion matrices were generated for each model by comparing the actual test outcomes with the predicted results.

Contd…

- The table on the right shows the test data accuracy score for each of the methods comparing them to show which performed best using the test data between Logistic Regression, SVM, Decision Trees and KNN:

Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.833333
KNN	0.833333

Link:
https://github.com/TjRathore/IBM-DataScience-Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

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

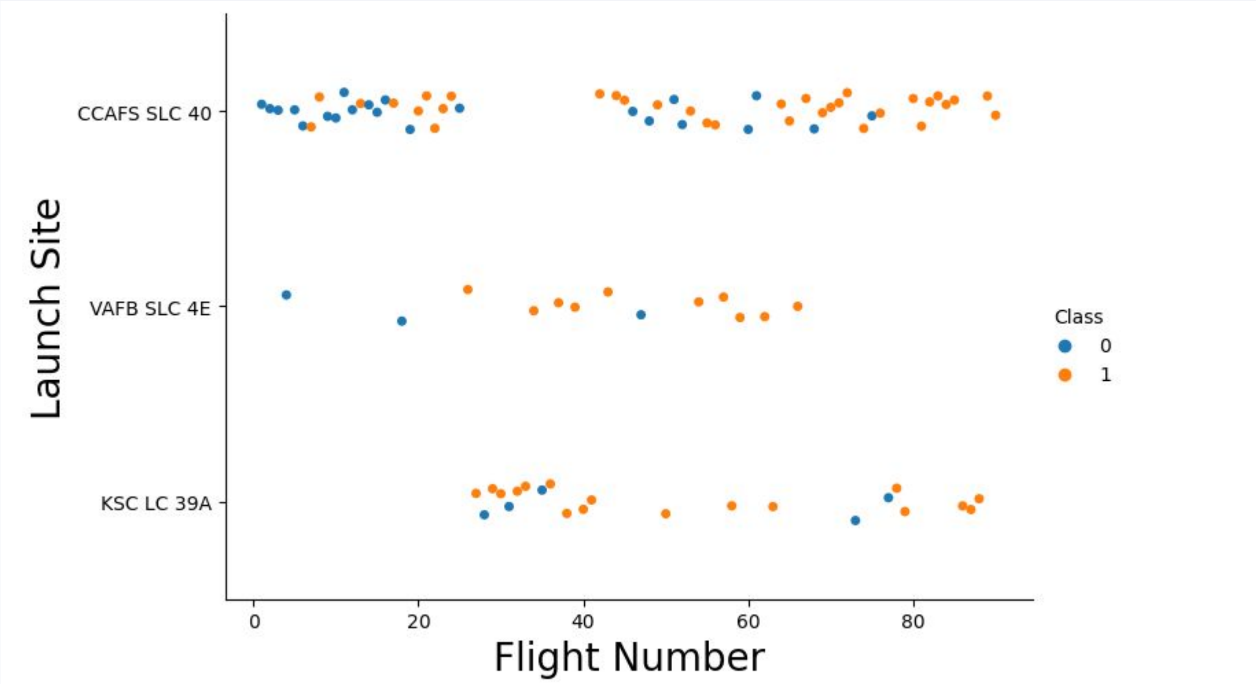
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

A scatter plot of Flight Number vs. Launch Site

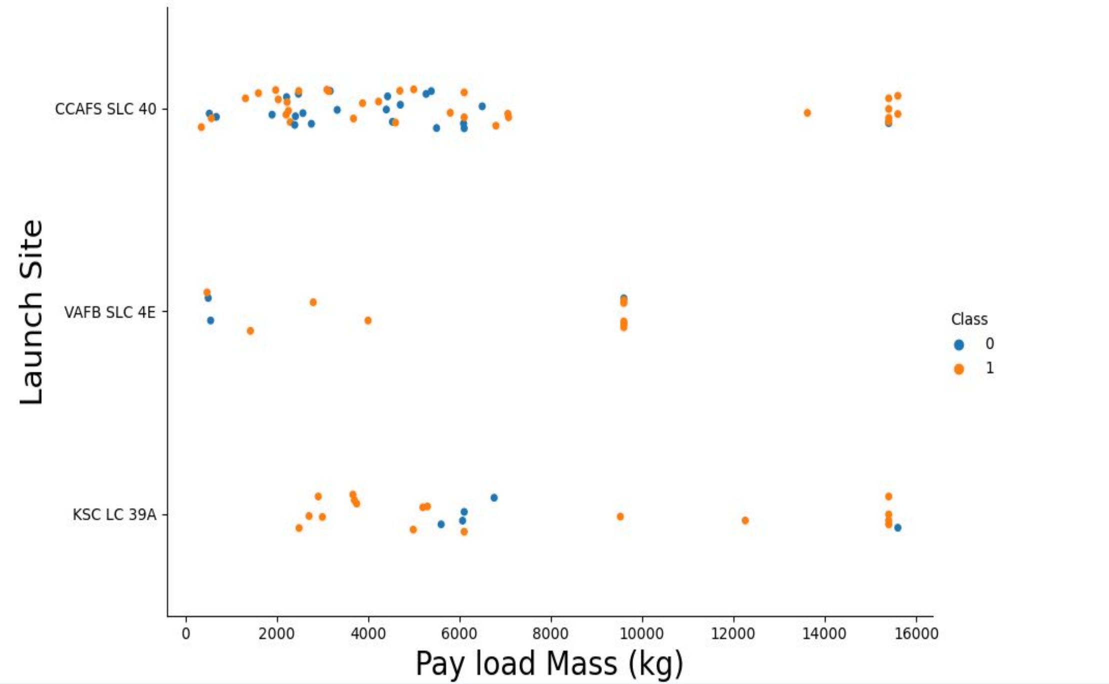
- Cape Canaveral CCAFS-SLC 40 is the most used launch site.
- CCAFS-SLC 40 has most no. of failures, particularly in the early stage of Falcon9 project.



Payload vs. Launch Site

Scatter plot of Payload vs. Launch Site

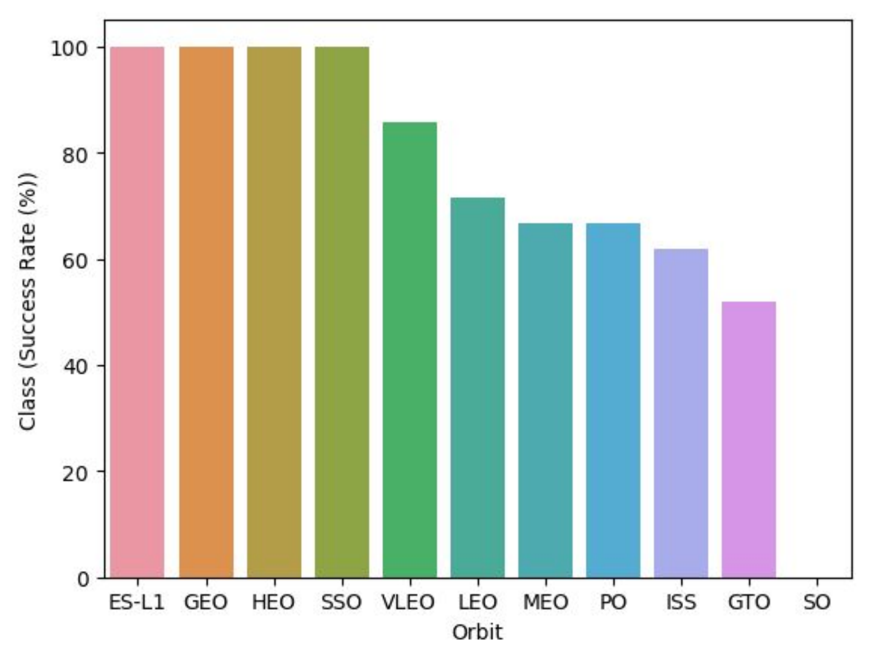
- Given Falcon9 specifications, heavy payloads > 10000 kg are sent to low/medium orbits LEO/MEO only.
- It is observed that the % of failures is lower for heavy payload which indicates that low orbits are less risky to the success of the mission (recovery of booster).



Success Rate vs. Orbit Type

Bar chart for the success rate of each orbit type

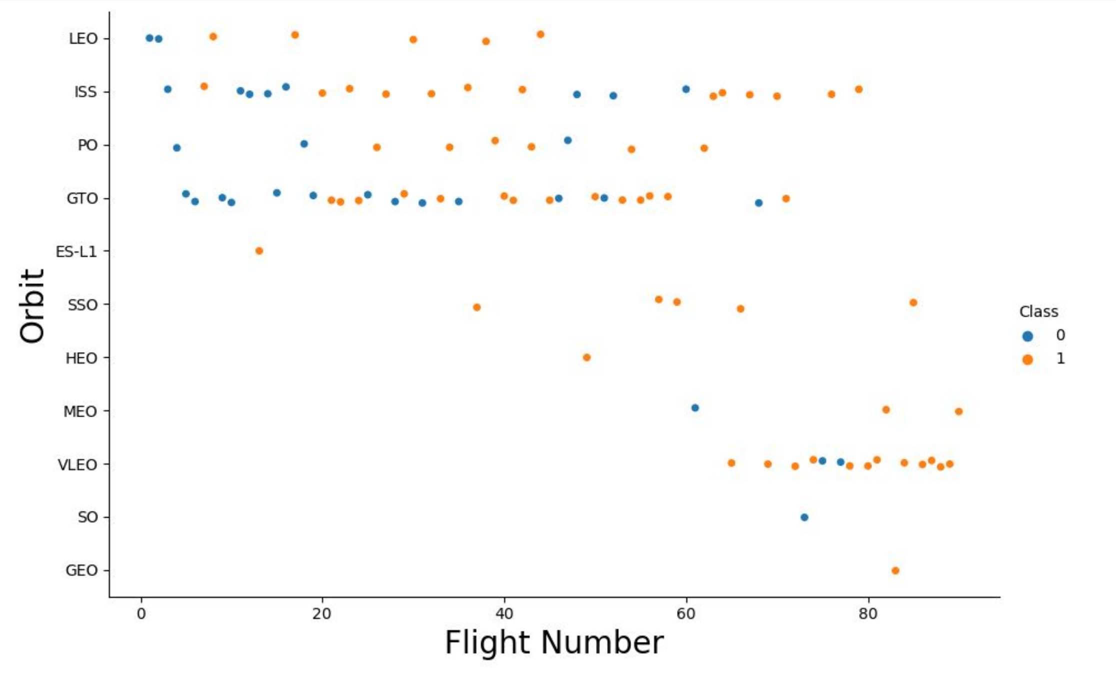
- GTO sees the lowest success rate while SSO, HEO, GEO & ES-L1 have highest success rates.
- Orbit SO has 0% success rate.
- Based on analysis, it can be said that success rate may depend on payload mass and orbit.



Flight Number vs. Orbit Type

- It can be said that PO, SSO, ISS, VLEO are low orbits.
- GTO is a transfer orbit to GEO.
- It looks like GTO are higher risk missions, low orbits are lower risk.

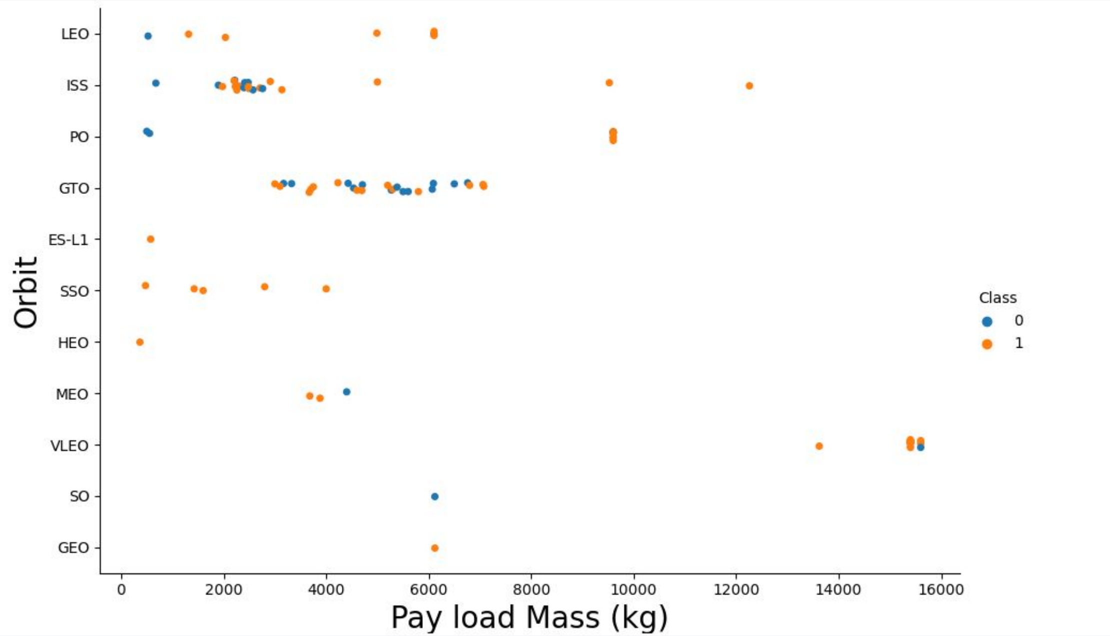
Scatter point of Flight number vs. Orbit type



Payload vs. Orbit Type

- The highest success rate is associated with low Earth orbits (excluding ISS) and payloads with low mass.
- The chart reveals that 5 out of 8 failures occurred during the early stages of the Falcon 9 project when the rocket's reliability was relatively low.
- Across a payload mass range of 2000 to 7500 kg, the success rate for GTO missions appears to be relatively evenly distributed. However, GTO missions pose a consistent risk to mission success, irrespective of payload mass.
- Although Falcon 9's reliability has improved over time, recent instances of failed booster recoveries following GTO launches suggest that challenges persist in this orbit category.

Scatter point of payload vs. orbit type



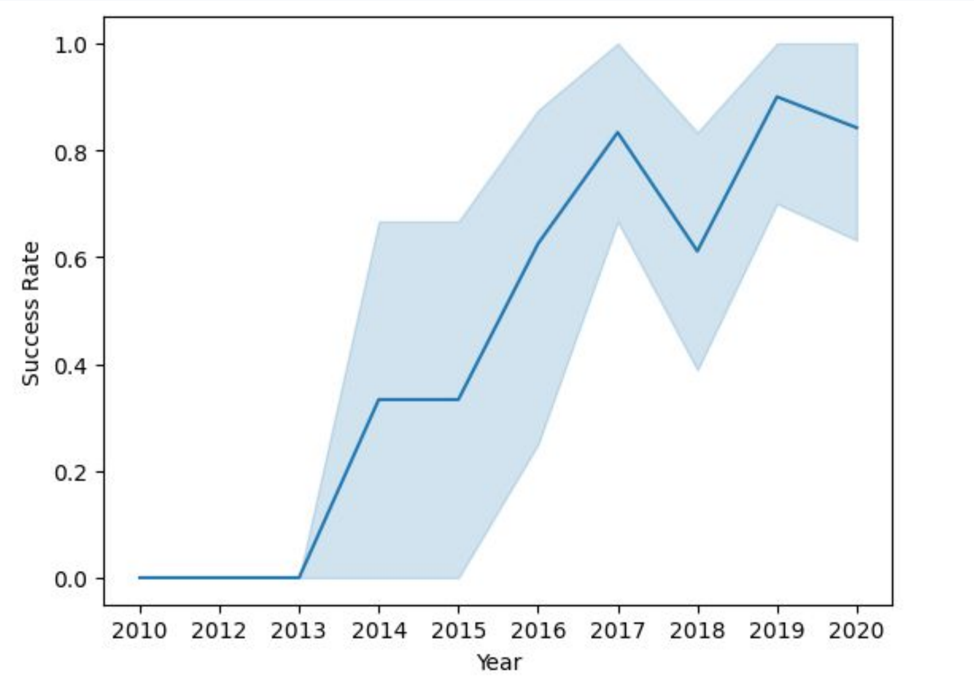
Launch Success Yearly Trend

- Falcon 9's reliability has seen significant improvement over time, with a success rate tied to factors such as payload mass, orbit type, and other variables under investigation.

- Ariane 5 has maintained an impressive close-to-100% success rate across 82 flights since 2003, while Falcon 9, on average, achieves a 66% success rate in booster recovery.

- Despite this difference, Falcon 9's current success rate remains adequate to sustain the financial viability of SpaceX.

Line chart of yearly average success rate



All Launch Site Names

Find the names of the unique launch sites

```
%sql SELECT DISTINCT LAUNCH_SITE as "Launch Sites" FROM SPACEXTBL;
```

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

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

Observation: It can be inferred that there are four distinct launch sites. These launch sites likely represent the locations from which SpaceX conducts its rocket launches.

Launch Site Names Begin with 'CCA'

Find 5 records where launch sites begin with `CCA`

```
[ ] %sql SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%' LIMIT 5;

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Observation: some of the missions had successful mission outcomes, while others had landing failures (specifically "Failure (parachute)") or no landing attempt.

Total Payload Mass

Calculate the total payload carried by boosters from NASA

```
[ ] %sql SELECT SUM(PAYLOAD_MASS__KG_) as "Total Payload Mass(Kgs)", Customer FROM SPACEXTBL WHERE Customer = 'NASA (CRS)';

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

Total Payload Mass(Kgs)	Customer
45596	NASA (CRS)

Observation: The query calculates the total payload mass (in kilograms) for missions where the customer is 'NASA (CRS)'.

Average Payload Mass by F9 v1.1

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

```
[ ]  %sql SELECT AVG(PAYLOAD_MASS__KG_) as "Avg Payload Mass", Customer, Booster_Version FROM SPACEXTBL WHERE Booster_Version LIKE 'F9 v1.1%';

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

Avg Payload Mass	Customer	Booster_Version
2534.6666666666665	MDA	F9 v1.1 B1003

Observation: This information provides insights into the average payload mass for missions using the specified Falcon 9 booster version and the associated customer.

First Successful Ground Landing Date

Find the dates of the first successful landing outcome on ground pad

```
%sql SELECT MIN(DATE) FROM SPACEXTBL WHERE Landing_Outcome = "Success (ground pad)";
```

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

MIN(DATE)
2015-12-22

Observation: This information provides insight into the timeline of successful ground pad landings achieved by SpaceX for its missions.

Successful Drone Ship Landing with Payload between 4000 and 6000

List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
[ ] %sql SELECT DISTINCT Booster_Version, Payload 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	Payload
F9 FT B1022	JCSAT-14
F9 FT B1026	JCSAT-16
F9 FT B1021.2	SES-10
F9 FT B1031.2	SES-11 / EchoStar 105

Observation: These missions achieved successful landings on drone ships and had payload masses within the specified range.

Total Number of Successful and Failure Mission Outcomes

Calculate the total number of successful and failure mission outcomes

```
[ ]  %sql SELECT Mission_Outcome, COUNT(Mission_Outcome) as Total FROM SPACEXTBL GROUP BY Mission_Outcome;
```

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

Mission_Outcome	Total
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Observation: It can be inferred that with given set of data, max. Missions succeeded.

Boosters Carried Maximum Payload

List the names of the booster which have carried the maximum payload mass

```
[ ]  %sql SELECT Booster_Version,Payload, PAYLOAD_MASS__KG_ FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);

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

Booster_Version	Payload	PAYLOAD_MASS__KG_
F9 B5 B1048.4	Starlink 1 v1.0, SpaceX CRS-19	15600
F9 B5 B1049.4	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600
F9 B5 B1051.3	Starlink 3 v1.0, Starlink 4 v1.0	15600
F9 B5 B1056.4	Starlink 4 v1.0, SpaceX CRS-20	15600
F9 B5 B1048.5	Starlink 5 v1.0, Starlink 6 v1.0	15600
F9 B5 B1051.4	Starlink 6 v1.0, Crew Dragon Demo-2	15600
F9 B5 B1049.5	Starlink 7 v1.0, Starlink 8 v1.0	15600
F9 B5 B1060.2	Starlink 11 v1.0, Starlink 12 v1.0	15600
F9 B5 B1058.3	Starlink 12 v1.0, Starlink 13 v1.0	15600
F9 B5 B1051.6	Starlink 13 v1.0, Starlink 14 v1.0	15600
F9 B5 B1060.3	Starlink 14 v1.0, GPS III-04	15600
F9 B5 B1049.7	Starlink 15 v1.0, SpaceX CRS-21	15600

Observation: Multiple booster versions were used for these missions (e.g., "F9 B5 B1048.4," "F9 B5 B1049.4"), highlighting SpaceX's reuse of boosters for multiple high-payload missions.

2015 Launch Records

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

```
[ ] %sql SELECT substr(Date, 4, 1) as 'Month', Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTBL WHERE substr(Date, 1, 4)='2015' AND Landing_Outcome = 'Failure (drone ship)';

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

Month	Landing_Outcome	Booster_Version	Launch_Site
5	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
5	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Observation: These records represent SpaceX missions in May 2015 that experienced failed drone ship landings at Cape Canaveral Air Force Station Launch Complex 40.

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

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

```
%sql SELECT Landing_Outcome, count(Landing_Outcome) as Landing_Count FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' group by Landing_Outcome ORDER BY count(Landing_Outcome) DESC;
```

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

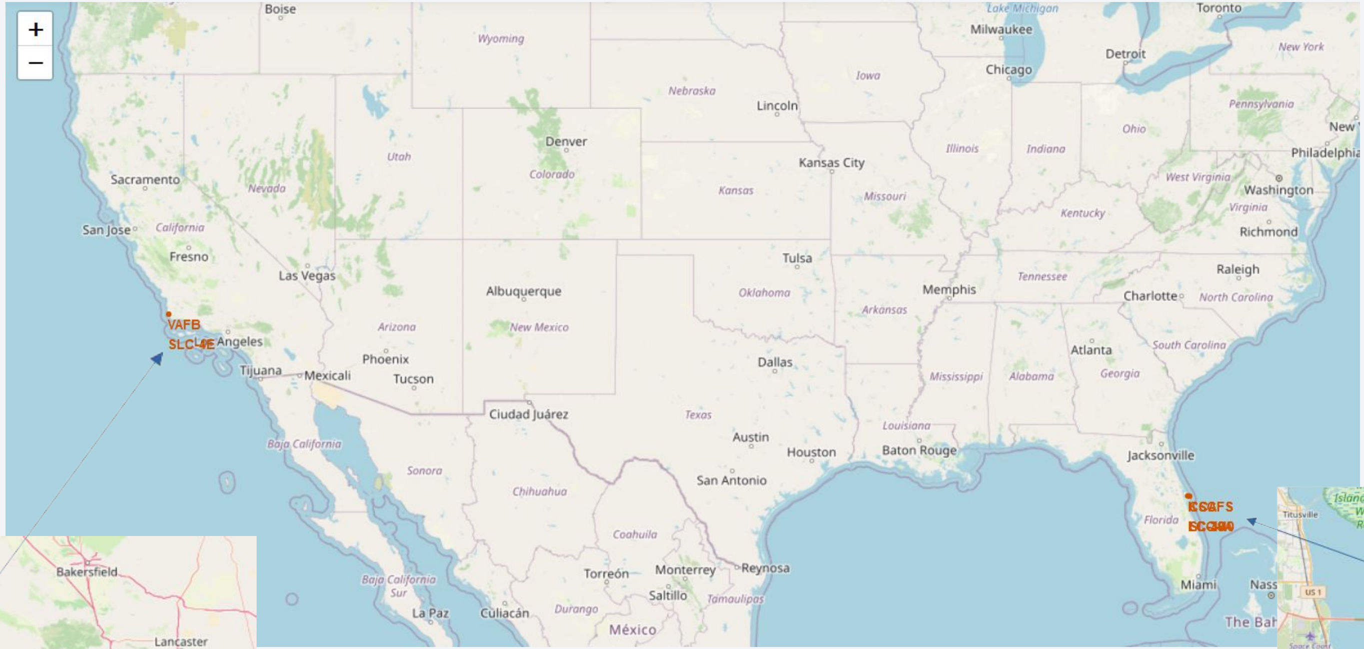
Landing_Outcome	Landing_Count
No attempt	10
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

Observation: The most frequent landing outcome during this period was "No attempt," which occurred 10 times. This suggests that a significant portion of missions during this time did not involve attempts at booster recovery or controlled landings.

Section
3

Launch Sites Proximities Analysis

SpaceX: All Launch Sites (Map)



- VAFB SLC-4E: Vandenberg Space Launch Complex 4 (CA)
- KSC-LC29A: Kennedy Space Center - Merritt Island (FL)
- CCAFS-LC40: Cape Canaveral Launch Complex 40 (FL)
- CCAF-SLC40: Cape Canaveral Space Launch Complex 40(FL)

SpaceX Launch Outcomes for each site

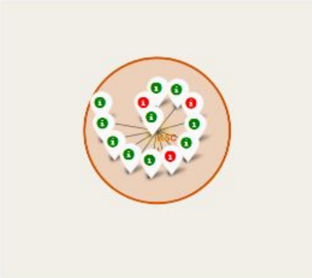
Vandenberg Space Launch Complex 4 (CA)
VAFB SLC-4E

Cape Canaveral (FL)
CCAFS-LC40

Launch Site	class	
CCAFS LC-40	0	19
	1	7
CCAFS SLC-40	0	4
	1	3
KSC LC-39A	0	3
	1	10
VAFB SLC-4E	0	6
	1	4

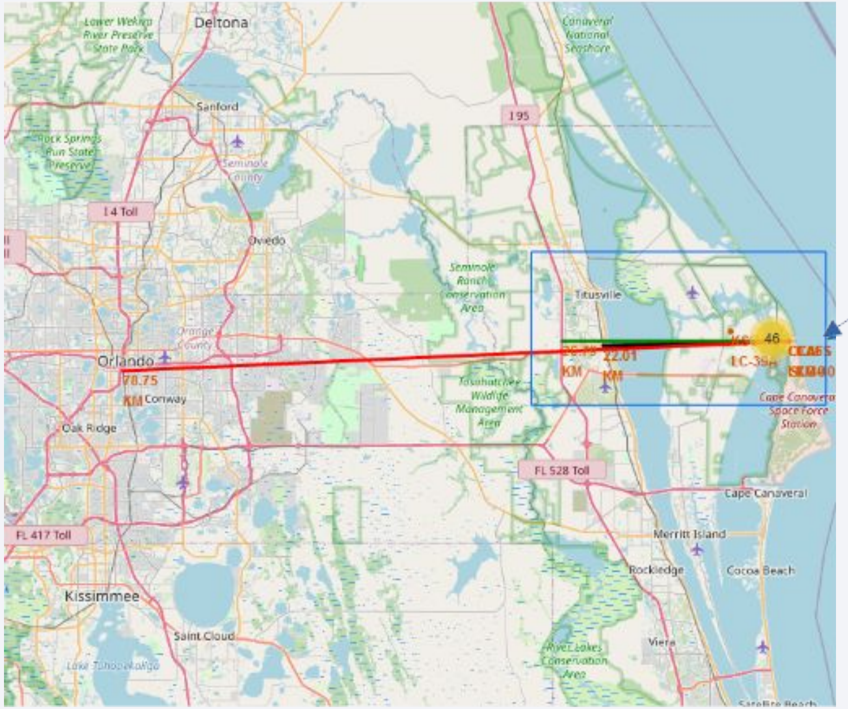
Table: Synthesis of launches outcomes
Class 0= failure
Class 1= success

Cape Canaveral (FL)
CCAFS-SLC40



Kennedy Space Center (FL)
KSC LC 39A

Launch Site & its proximities



Distance from CCAFS_SLC40 to:

- Closest coast: ~900 m
- Florida East Coast Railway: 22.0 km
- Highway I 95: 26.8 km
- Orlando: 78.75 km

Section
4

Build a Dashboard with Plotly Dash

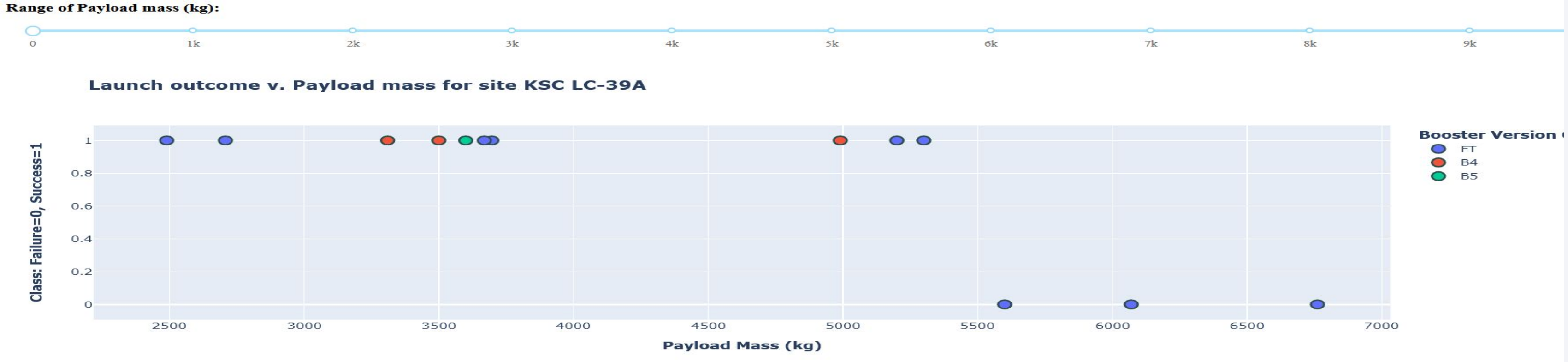
SpaceX Falcon9: Launch Record Dashboard

- The interactive dashboard facilitates the visualization and analysis of successful Falcon launches, consolidating scattered data into a cohesive view.
- Through this tool, it becomes straightforward to identify the launch site with the highest success rate among all sites.
- Worth noting, the Kennedy Space Center, situated in Florida, is a prominent contributor to the overall success of Falcon launches.

Falcon9: Highest launch success ratio

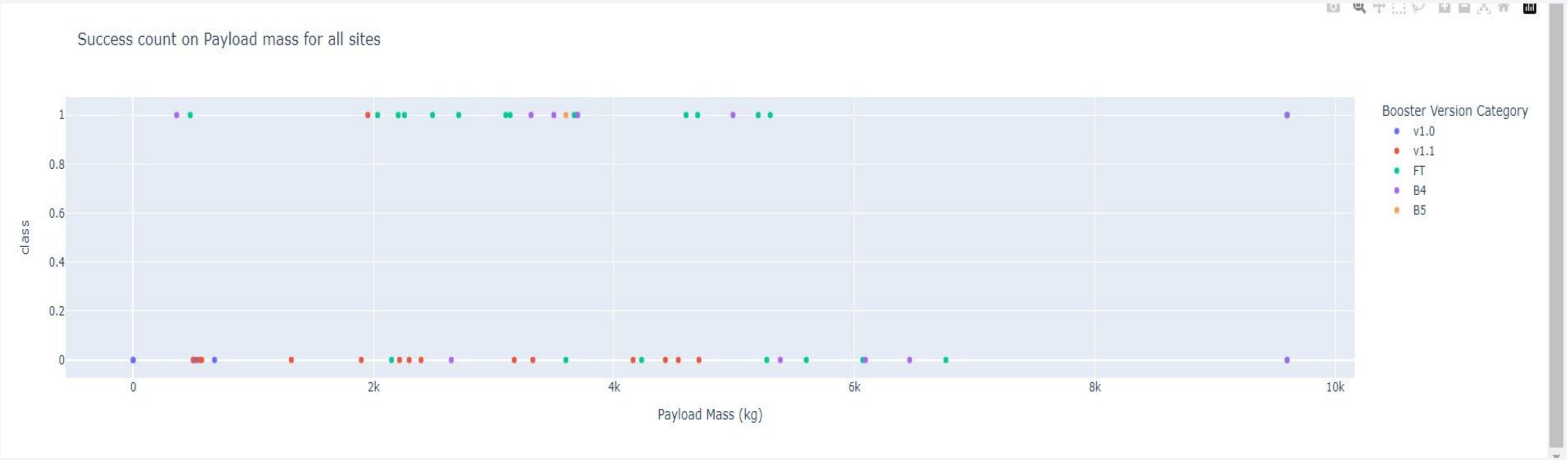
Launch Site: KSC LC-39A
Location: Kennedy Space Center, Florida
Total Flights: 13
Successful Missions: 10

Notably, missions carrying heavy payloads are considered high risk. Mission success does not appear to be significantly influenced by the booster versions when dealing with low-mass payloads (payload mass < 5,500 kilograms).



Launch outcome vs. Payload mass (all sites)

- Early versions (V1.0 and v1.1) had lower reliability.
- Landing legs, introduced in v1.1, didn't lead to intact landings and were discontinued in 2015.
- "FT" (Full Thrust) is the next-gen with the highest success rate for payloads < 6 tons, often using reused launchers. Heavy payloads pose higher risk.



Section
5

Predictive Analysis (Classification)

Classification Accuracy

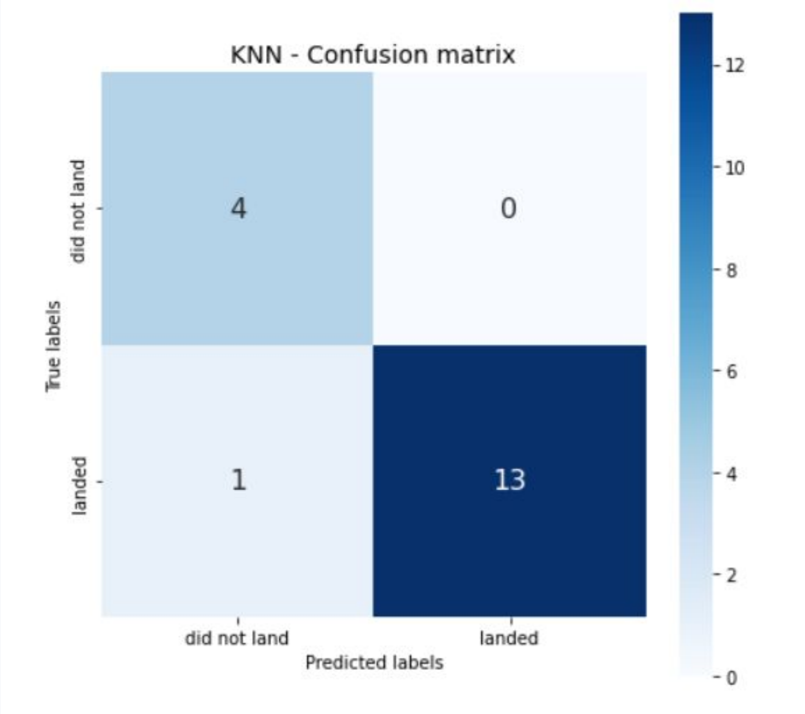
- All the Models performed with same accuracy on given test data.
- Accuracy of models: 83.3%

Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decision Tree	0.833333
KNN	0.833333

Confusion Matrix

- k-nearest neighbors algorithm (k-NN) is the best “predictor”
- The model perfectly predicts mission failure
- 1 false negative for successful booster landing (recovery)

With train test split → random_state as 0.3



Conclusions

Below are the key observations and conclusion:

- **Differential Success Rates by Launch Site:**

    a. Launch sites exhibit varying success rates.
    b. CCAFS LC-40 has a 60% success rate, while KSC LC-39A and VAFB SLC 4E boast higher rates at 77%.

- **Positive Correlation between Flight Number and Success:**

    a. Across all three launch sites, an upward trend in success rate is evident as the flight number increases.
    b. For example, VAFB SLC 4E achieves a 100% success rate after the 50th flight, while KSC LC 39A and CCAFS SLC 40 both achieve 100% success rates after the 80th flight.

- **Payload vs. Launch Site Analysis:**

    a. The scatter point chart reveals that VAFB-SLC has not launched rockets with heavy payloads exceeding 10,000 kg.

Conclusions *(contd…)*

- **Orbit-Related Success Rates:**

  a. Orbits such as ES-L1, GEO, HEO, and SSO consistently achieve 100% success rates.
  b. In contrast, the SO orbit has the lowest success rate at approximately 50%, with 0% success observed.
  c. For LEO orbits, success appears to be correlated with the number of flights, while GTO orbits do not exhibit a clear relationship with flight number.

- **Landing Success with Heavy Payloads:**

  a. Heavy payloads show higher positive landing rates for Polar, LEO, and ISS orbits.
  b. In GTO orbits, distinguishing between positive and negative landing rates is challenging due to the presence of both successful and unsuccessful missions.

- **Increasing Success Rate Over Time:**

  a. The data suggests that the success rate has steadily increased since 2013, reaching higher levels by 2020.

Appendix

- Entire Project Repository Link: https://github.com/TjRathore/IBM-DataScience-Capstone/tree/main
- Interactive Dashboard Application Screenshots: https://github.com/TjRathore/IBM-DataScience-Capstone/tree/main/screenshots

Thank you!