IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection through API
  - Data Collection with Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analytics with Folium
  - Machine Learning Prediction
- Summary of all results
  - Exploratory Data Analysis result
  - Interactive analytics in screenshots
  - Predictive Analytics result from Machine Learning Lab

Introduction

SpaceX is a revolutionary company who has disrupt the space industry by offering a rocket launches specifically Falcon 9 as low as 62 million dollars; while other providers cost upward of 165 million dollar each. Most of this saving thanks to SpaceX astounding idea to reuse the first stage of the launch by re-land the rocket to be used on the next mission. Repeating this process will make the price down even further. As a data scientist of a startup rivaling SpaceX, the goal of this project is to create the machine learning pipeline to predict the landing outcome of the first stage in the future. This project is crucial in identifying the right price to bid against SpaceX for a rocket launch.

The problems included:

- Identifying all factors that influence the landing outcome.
- The relationship between each variables and how it is affecting the outcome.
- The best condition needed to increase the probability of successful landing.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data collected using SpaceX REST API and web scrapping from Wikipedia
- Perform data wrangling
  - Data was processed using one-hot encoding for categorical features
- 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

Method 1: Using SpaceX REST API

Target Endpoint: Using API endpoint which refers to a specific URL within the SpaceX REST API that provides access to desired data. In this case, we target api.spacexdata.com/v4/launches/past to retrieve information about past launches.

GET Request with requests library: We use the requests library in Python to send a GET request to the API endpoint. This initiates a request to retrieve data.

JSON Response: The response from the API will be in JSON format, a structured data format. This response can be viewed using the .json() method.

Converting JSON to DataFrame: Python's pandas library provides the json_normalize function. This function converts the nested JSON data (list of JSON objects representing launches) into a flat table format, which is more suitable for analysis.

Method 2: Web Scraping with BeautifulSoup

Target Source: Web scraping from “Falcon 9 Launch Wiki page." This method involves extracting data from HTML tables on a webpage that contains launch records.

BeautifulSoup for Parsing: The BeautifulSoup package helps navigate and parse the HTML structure of the webpage.

Extracting Table Data: We use BeautifulSoup methods to find specific tables and extract data from table cells.

Converting to DataFrame: Similar to API data, the extracted table data can be converted into a pandas DataFrame for further analysis.

Data Collection – SpaceX API

Flowchart for data collection using SpaceX REST calls.

Method 1 (API):

```
+-----------+         +---------+         +-----------------+
|  Start    |--------->| Target  |--------->|  GET Request    |
|           |         | Endpoint|         | (requests library)|
+-----------+         +---------+         +-----------------+
                   |
                   V
+-----------+         +---------+         +-----------------+
|  Response |--------->| JSON    |--------->| Convert to      |
| (JSON)    |         | Data    |         | DataFrame       |
+-----------+         +---------+         +-----------------+
                   |
                   V
+-----------+         +---------+
|   Clean   |--------->|  Data   |
|   Data    |         |  Ready  |
+-----------+         +---------+
```

Data Collection - Scraping

Flowchart data using webscraping

Method 2 (Web Scraping):

```
+-----------+         +---------+         +-----------------+
|  Start    |--------->| Target  |--------->| Parse HTML      |
|           |         | Webpage |         | (BeautifulSoup)  |
+-----------+         +---------+         +-----------------+
                   |
                   V
+-----------+         +---------+         +-----------------+
|  Extract  |--------->| Table   |--------->| Convert to      |
|  Table    |         | Data    |         | DataFrame       |
+-----------+         +---------+         +-----------------+
                   |
                   V
+-----------+         +---------+
|   Clean   |--------->|  Data   |
|   Data    |         |  Ready  |
+-----------+         +---------+
```

Data Wrangling

Handling Categorical DataLaunch Site:
Convert categorical values like "CCAFS SLC 40" into numerical representations
Orbit: Similarly, convert categorical orbit into numerical values.

Handling Boolean Data
Outcome: Convert boolean values into binary values (0 and 1) where 0 represents a failed landing and 1 represents a successful landing.

Creating a Classification Variable
Y: Create a new binary variable Y to represent the classification target. Assign 0 to failed landings and 1 to successful landings.

```
-------------------+         +-------------------+         +--------------
 Raw Data         |--------->| Handle Categorical |--------->| Handle Bool
(e.g., CSV)      |         | Data              |         | Data            |
-------------------+         +-------------------+         +--------------
                  |                                              |
                  V                                              V
-------------------+         +-------------------+         +--------------
 Create Y         |--------->| Handle Missing    |--------->| Feature Engin
 Variable        |         | Values            |         | (if necessary)
-------------------+         +-------------------+         +--------------
                  |                                              |
                  V                                              V
-------------------+         +-------------------+
 Normalize/Scale |--------->|  Clean Data       |
 Features        |         |  Ready for        |
-------------------+         |  Analysis         |
-------------------+         +-------------------+
```

Handling Missing Values
Payload Mass: Imput the mean or median value, using a regression model to predict missing values based on other features. Removing rows with missing values if the number of missing values is small if there are missing values in the PayloadMass column.

Feature Engineering
Create new features: Based on the dataset, created new features that could be informative for predicting landing outcomes.

Normalization and Scaling
Normalize or scale features: normalizing or scaling them to prevent features with larger magnitudes from dominating the model, when they have different scales.

EDA with Data Visualization

Scatter plot is used to:

- Visualize and explain the pattern of the relationship between Payload and Flight Number.
- Visualize and explain the pattern of the relationship between Flight Number and Launch Site
- Visualize and explain the pattern of the relationship between Payload and Launch Site.
- Visualize and explain the pattern of the relationship between Flight Number and orbit Type
- Visualize and explain the pattern of the relationship between payload and orbit Type
- Visualize and explain the pattern of the relationship between the success rate of each orbit type.

Once we get a hint of the relationships using scatter plot. We will then use further visualization tools such as bar graph and line plots graph for further analysis.

Bar graphs is one of the easiest way to interpret the relationship between the attributes. In this case, we will use the bar graph to determine which orbits have the highest probability of success.

We then use the line graph to show a trends or pattern of the attribute over time which in this case, is used for see the launch success yearly trend.

We then use Feature Engineering to be used in success prediction in the future module by created the dummy variables to categorical columns.

EDA with SQL

Display:

- Names of unique launch sites
- 5 records where launch site begins with 'CCA'
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1.

List:

- Date of first successful landing on ground pad
- Names of boosters which had success landing on drone ship and have payload mass greater than 4,000 but less than 6,000
- Total number of successful and failed missions
- Names of booster versions which have carried the max payload
- Failed landing outcomes on drone ship, their booster version and launch site for the months in the year 2015
- Count of landing outcomes between 2010-06-04 and 2017-03-20 (desc)

2023

Build an Interactive Map with Folium

Map with Folium

Markers Indicating Launch Sites

- Added **blue circle** at **NASA Johnson Space Center's coordinate** with a **popup label** showing its name using its latitude and longitude coordinates
- Added **red circles** at **all launch sites coordinates** with a **popup label** showing its name using its name using its latitude and longitude coordinates

Colored Markers of Launch Outcomes

- Added **colored markers** of **successful** (**green**) and **unsuccessful** (**red**) **launches** at each launch site to show which launch sites have high success rates

Distances Between a Launch Site to Proximities

- Added **colored lines** to **show distance between** launch site **CCAFS SLC-40 and** its proximity to the **nearest coastline, railway, highway, and city**

Build a Dashboard with Plotly Dash

Dashboard with Plotly Dash

Dropdown List with Launch Sites

- Allow user to select all launch sites or a certain launch site

Pie Chart Showing Successful Launches

- Allow user to see successful and unsuccessful launches as a percent of the total

Slider of Payload Mass Range

- Allow user to select payload mass range

Scatter Chart Showing Payload Mass vs. Success Rate by Booster Version

- Allow user to see the correlation between Payload and Launch Success

Predictive Analysis (Classification)

The dataset was split into training and testing Sets.

- Logistic Regression, SVM (Support Vector Machine), Decision Tree, and KNN (k-Nearest Neighbors) machine learning models were trained on the training data set.

- Hyper-parameters were evaluated using GridSearchCV() and the best was selected using '.best_params_'.

- Using the best hyper-parameters, each of the four models were scored on accuracy by using the testing data set.

Results

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

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

Scatter plot of Flight Number vs. Launch Site

Payload vs. Launch Site

Most launches with payload mass under 10,000 kg are from any launch site, but heavier ones happens mainly at CCAFS SLC 40 and KSC LC 39A

Success Rate vs. Orbit Type

Bar chart for the success rate of each orbit type

Flight Number vs. Orbit Type

Scatter point of Flight number vs. Orbit type. in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type

Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.

Launch Success Yearly Trend

The line chart of yearly average success rate shows that the success rate since 2013 kept increasing till 2020.

All Launch Site Names

Query: sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL ORDER BY 1;

Query results: * ibm_db_sa://fvp19040:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb Done.

Names of the unique launch sites:

CCAFS LC-40

CCAFS SLC-40

KSC LC-39A

VAFB SLC-4E

Launch Site Names Begin with 'CCA'

Query: sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;

Query results: * ibm_db_sa://fvp19040:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb Done.

5 records where launch sites begin with `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	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

Total Payload Mass

Query:

```
sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD FROM SPACEXTBL
WHERE PAYLOAD LIKE '%CRS%';
```

Query results:

```
* ibm_db_sa://fvp19040:***@54a2f15b-5c0f-46df-8954-
7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb Done.
```

Total payload carried by: boosters from NASA

111268

Average Payload Mass by F9 v1.1

Query:

```
sql SELECT AVG(PAYLOAD_MASS__KG_) AS AVG_PAYLOAD FROM SPACEXTBL
WHERE BOOSTER_VERSION = 'F9 v1.1';
```

Query results:

```
* ibm_db_sa://fvp19040:***@54a2f15b-5c0f-46df-8954-
7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb Done.
```

Average payload payload mass carried by booster version F9 v1.1: 2928

First Successful Ground Landing Date

Query:

```
sql SELECT MIN(DATE) AS FIRST_SUCCESS_GP FROM SPACEXTBL WHERE
LANDING__OUTCOME = 'Success (ground pad)';
```

Query results:

```
* ibm_db_sa://fvp19040:***@54a2f15b-5c0f-46df-8954-
7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb Done.
```

dates of the first successful landing outcome on ground pad: 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Query:

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

Query results:

```
* ibm_db_sa://fvp19040:***@54a2f15b-5c0f-46df-8954-
7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb Done.
```

Successful Drone Ship Landing with Payload between 4000 and 6000:

F9 FT B1021.2

F9 FT B1031.2

F9 FT B1022

F9 FT B1026

Total Number of Successful and Failure Mission Outcomes

Query:

```
sql SELECT MISSION_OUTCOME, COUNT(*) AS QTY FROM SPACEXTBL GROUP BY
MISSION_OUTCOME ORDER BY MISSION_OUTCOME;
```

Query results:

```
* ibm_db_sa://fvp19040:***@54a2f15b-5c0f-46df-8954-
7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb Done.
```

total number of successful and failure mission outcomes:

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

Boosters Carried Maximum Payload

Query:

```
sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ =
(SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL) ORDER BY BOOSTER_VERSION;
```

Query results:

```
* ibm_db_sa://fvp19040:***@54a2f15b-5c0f-46df-8954-
7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb Done.
```

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

F9 B5 B1048.4

F9 B5 B1048.5

F9 B5 B1049.4

F9 B5 B1049.5

F9 B5 B1049.7

F9 B5 B1051.3

F9 B5 B1051.4

F9 B5 B1051.6

F9 B5 B1056.4

F9 B5 B1058.3

F9 B5 B1060.2

F9 B5 B1060.

2015 Launch Records

Query:

```
sql SELECT BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL WHERE
LANDING__OUTCOME = 'Failure (drone ship)' AND DATE_PART('YEAR', DATE) = 2015;
```

Query results:

```
* ibm_db_sa://fvp19040:***@54a2f15b-5c0f-46df-8954-
7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb Done.
```

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

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

Query:

```
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;
```

Query results:

```
* ibm_db_sa://fvp19040:***@54a2f15b-5c0f-46df-8954-
7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb Done.
```

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:

landing__outcome	qty
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

Section 3

Launch Sites Proximities Analysis

Falcon 9 Launch Site Locations

VAFB SLC-4E (California, USA)
- Vandenberg Air Force Base Space Launch Complex 4E

KSC LC-39A (Florida, USA)
- Kennedy Space Center Launch Complex 39A

CCAFS LC-40 (Florida, USA)
- Cape Canaveral Air Force Station Launch Complex 40

CCAFS SLC-40 (Florida, USA)
- Cape Canaveral Air Force Station Space Launch Complex 40

Map Markers of Success/Failed Landings

VAFB SLC-4E

KSC LC-39A

CCAFS LC-40

CCAFS SLC-40

The markers display the mission outcomes (Success/Failure) for Falcon 9 first stage landings. They are grouped on the map to be associated with the geographical coordinates for the launch site.

Distance from Launch Site to Proximities

- The CCAFS LC-40 and CCAFS SLC-40 launch sites have coordinates that are close to being, but are not exactly, right on top of each other.
- The perimeter road around CCAFS LC-40 is 0.19 km away from the launch site coordinates.
- The coastline is 0.92 km away from CCAFS LC-40.
- The rail line is 1.33 km away from CCAFS LC-40.

Section 4

Build a Dashboard with Plotly Dash

Launch Success Count for All Sites

The dropdown menu allows the selection of one or all launch sites.

- With all launch sites selected, the pie chart displays the distribution of successful Falcon 9 first stage landing outcomes between the different launch sites.
- The greatest share of successful Falcon 9 first stage landing outcomes (at 41.7% of the total) occurred at KSC LC-39A.

<Dashboard Screenshot 2>

- Falcon 9 first stage failed landings are indicated by the ‘0’ Class (blue wedge in the pie chart) and successful landings by the ‘1’ Class (red wedge in the pie chart).

- CCAFS SLC-40 was the launch site that had the highest Falcon 9 first stage landing success rate (42.9%).

<Dashboard Screenshot 3>

These screenshots are of the Payload vs. Launch Outcome scatter plots for all sites, with different payload selected in the range slider.

- The payload range from about 2,000 kg to 5,000 kg has the largest success rate.

- The ‘FT’ booster version category has the largest success rate.

CCAFS LC-40

CCAFS SLC-40

KSC LC-39A

VAFB SLC-4E

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- All models performed equally well except for the Decision Tree model which performed poorly relative to the other models.

Confusion Matrix

- Shown here is the confusion matrix for the Logistic Regression model.
- Confusion matrices can be read as:
- Prediction Breakdown:
- 12 True Positives and 3 True Negatives
- 3 False Positives and 0 False Negatives

Conclusions

- SpaceX does not have a perfect track record of Falcon 9 first stage landing outcomes.

- SpaceX’s Falcon 9 first stage landing outcomes have been trending towards greater success as more launches are made.

- The machine learning models can be used to predict future SpaceX Falcon 9 first stage landing outcomes.

Appendix

SpaceX API (JSON): https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json

- Wikipedia (Webpage): https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

- SpaceX (CSV): https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321ENSkillsNetwork/labs/module_2/data/Spacex.csv?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetworkChannel-SkillsNetworkCoursesIBMDS0321ENSkillsNetwork26802033-2022-01-01

- Launch Geo (CSV): https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_geo.csv

- Launch Dash (CSV): https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_dash.csv

Thank you!