WINNING SPACE RACE WITH DATA SCIENCE

30/08/2024

OUTLINE OF THE PRESENTATION

- **Executive**
- **Summary**
- **Introduction**
- **Methodology**
- **Results**
- **Conclusion**

EXECUTIVE SUMMARY

- Data Collection
- Data Wrangling
- Exploratory Data Analysis (EDA)
- Interactive Visual Analytics
- Predictive Modeling

INTRODUCTION

The goal of this project is to predict whether the first stage of a Falcon 9 rocket will successfully land, which significantly influences the cost-effectiveness of the launch. Understanding the factors that lead to a successful landing can help other companies compete with SpaceX by bidding more effectively for launches.

METHODOLOGY

- **Data Collection:** Data was obtained from the SpaceX API and through web scraping. This included information on various aspects of each launch, such as the rocket version, payload, launch site, and more.

In [5]:

```
# Takes the dataset and uses the cores column to call the API and append the data to the list
def getCoreData(data):
    for core in data['cores']:
            if core['core'] != None:
                response = requests.get("https://api.spacexdata.com/v4/cores/"+core['core']).
                Block.append(response['block'])
                ReusedCount.append(response['reuse_count'])
                Serial.append(response['serial'])
            else:
                Block.append(None)
                ReusedCount.append(None)
                Serial.append(None)
            Outcome.append(str(core['landing_success'])+' '+str(core['landing_type']))
            Flights.append(core['flight'])
            GridFins.append(core['gridfins'])
            Reused.append(core['reused'])
            Legs.append(core['legs'])
            LandingPad.append(core['landpad'])
```

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

In [6]:

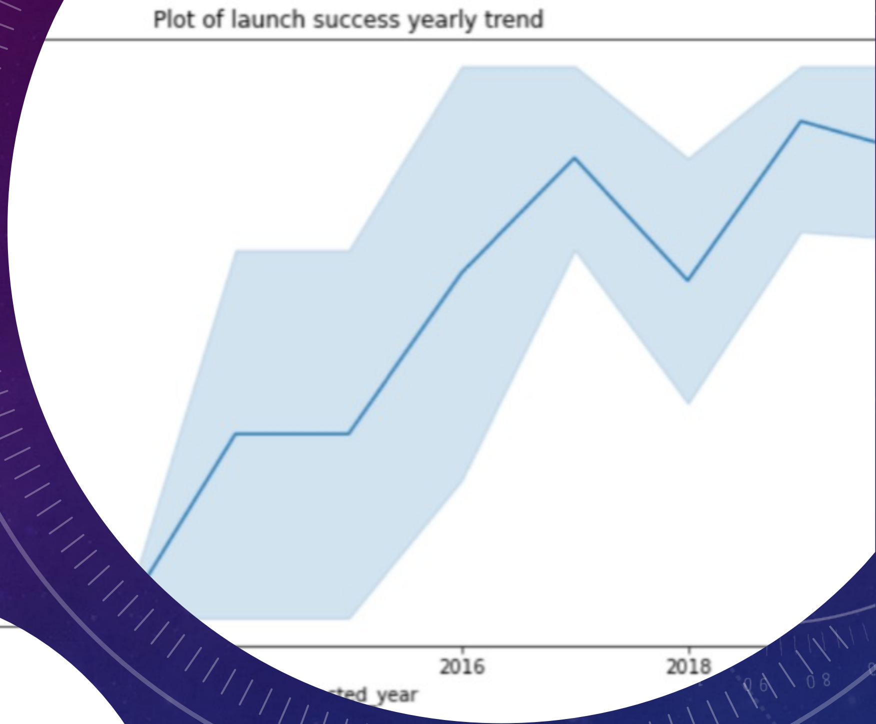
```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

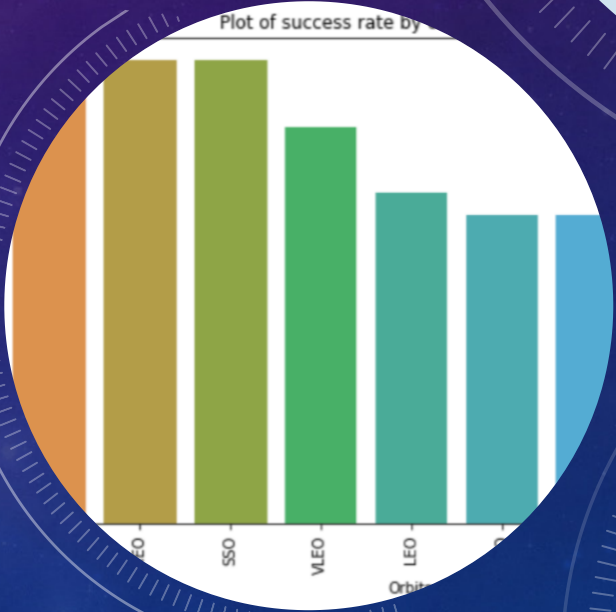
METHODOLOGY

- **Data Wrangling:** The data was cleaned to remove inconsistencies and missing values. Categorical variables were transformed using one-hot encoding, and the data was prepared for analysis.

EDA WITH VISUALIZATION

- **Exploratory Data Analysis (EDA):** The data was analyzed using SQL and visualizations to identify key patterns and relationships. For example, the relationship between flight numbers and launch success was explored, as well as the impact of payload mass on landing outcomes.





EDA WITH SQL

Variable Relationships: Analyzing relationships between variables like Flight Number, Launch Site, Payload, and Orbit to uncover trends and patterns.

SQL Queries: Examples include showing unique launch sites, payload mass by booster version, and mission outcomes.

Details: These pages describe how exploratory data analysis was conducted, including both visualizations and SQL queries, to gain insights into the data.

We applied EDA with SQL to get insight from the data. We wrote queries to find out for instance:

- The names of unique launch sites in the space mission.
- The total payload mass carried by boosters launched by NASA (CRS)
- The average payload mass carried by booster version F9 v1.1
- The total number of successful and failure mission outcomes
- The failed landing outcomes in drone ship, their booster version and launch site names.

EDA WITH SQL RESULTS

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

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

	total_payloadmass
0	45596

Task 4

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

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

	avg_payloadmass
0	2928.4

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

```
task_1 = '''
        SELECT DISTINCT LaunchSite
        FROM SpaceX
'''
create_pandas_df(task_1, database=conn)
```

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

Task 2

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

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

	date	time	boosterversion	launchsite	payload	payloadmasskg	orbit	customer	missio
0	2010-04-	18:45:00	F9 v1.0 B0003	CCAFS	Dragon Spacecraft	0	LEO	SpaceX	

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

	successoutcome
0	100

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

	failureoutcome
0	1

Task 8

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

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

	boosterversion	payloadmasskg
0	F9 B5 B1048.4	15600
1	F9 B5 B1048.5	15600
2	F9 B5 B1049.4	15600
3	F9 B5 B1049.5	15600
4	F9 B5 B1049.7	15600
5	F9 B5 B1051.3	15600

Task 5

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

Hint:Use min function

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

	firstsuccessfull_landing_date
0	2015-12-22

Task 6

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

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

	boosterversion
0	F9 FT B1022
1	F9 FT B1026

INTERACTIVE MAP WITH FOLIUM

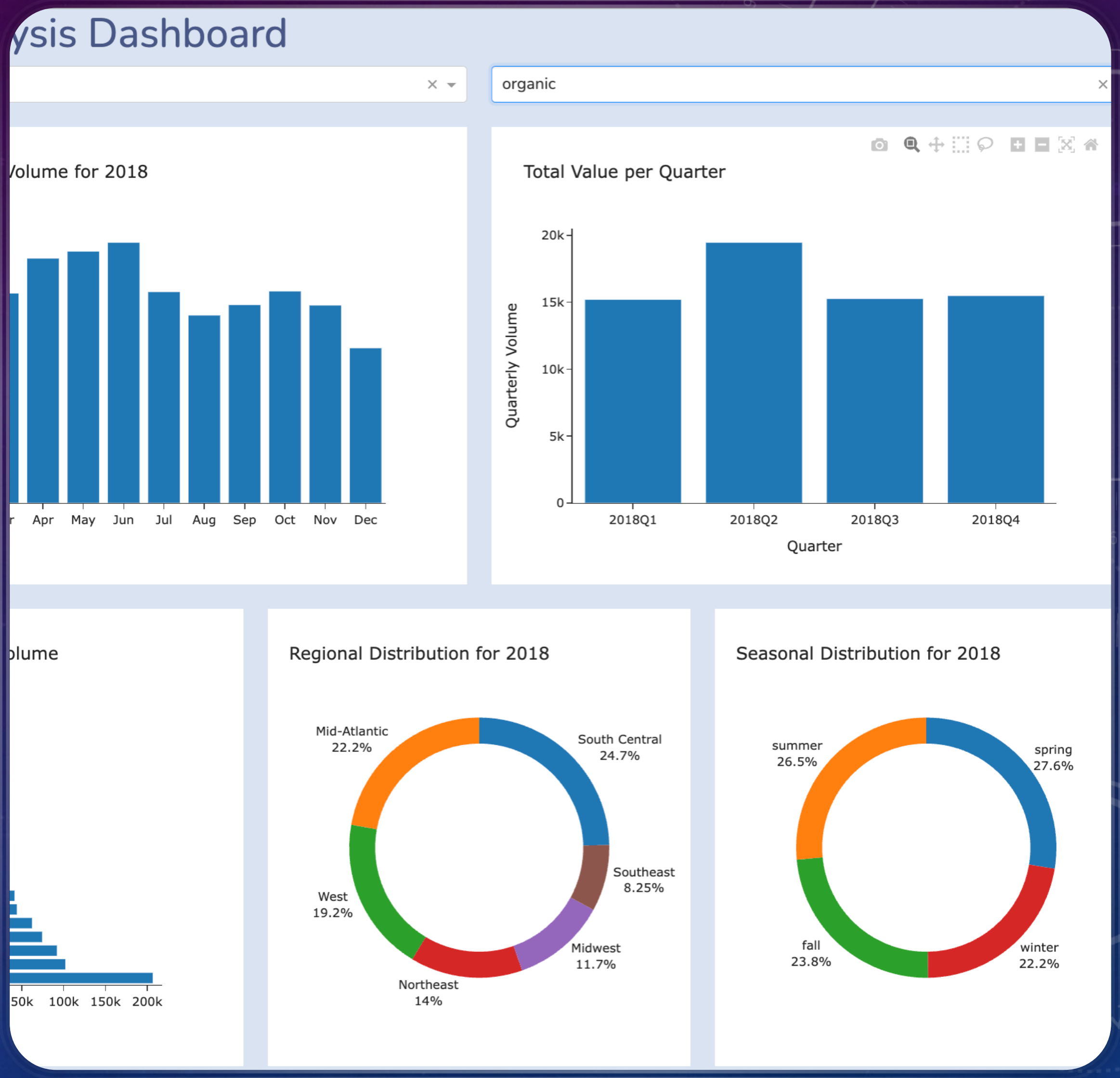
- We marked all launch sites, and added map objects such as markers, circles, lines to mark the success or failure of launches for each site on the folium map.
- We assigned the feature launch outcomes (failure or success) to class 0 and 1.i.e., 0 for failure, and 1 for success.
- Using the color-labeled marker clusters, we identified which launch sites have relatively high success rate.
- We calculated the distances between a launch site to its proximities. We answered some question for instance:
  - Are launch sites near railways, highways and coastlines.
  - Do launch sites keep certain distance away from cities.

INTERACTIVE DASHBOARD WITH PLOTLY DASH

Dashboard Features:

- **Dropdown Selector:** Allows users to filter the data by Launch Site, affecting visualizations like pie charts and scatter plots.
- **Range Selector:** A Payload Mass range selector filters data points on the scatter plot.

Details: This page explains the interactive features of the dashboard created with Plotly Dash, enabling dynamic data exploration.

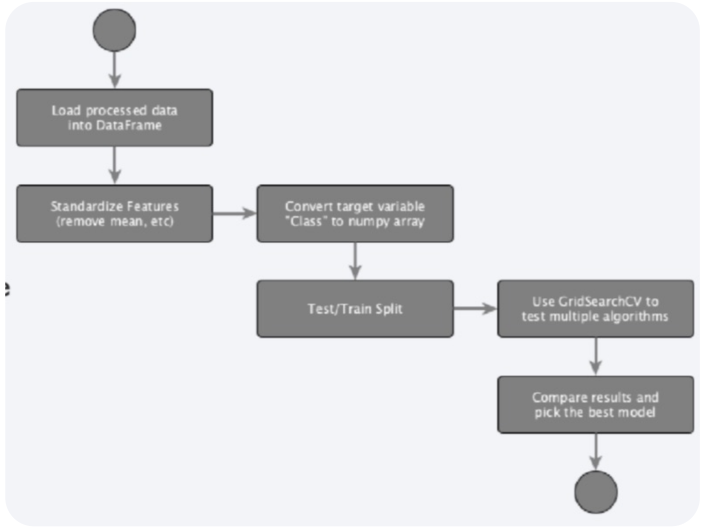


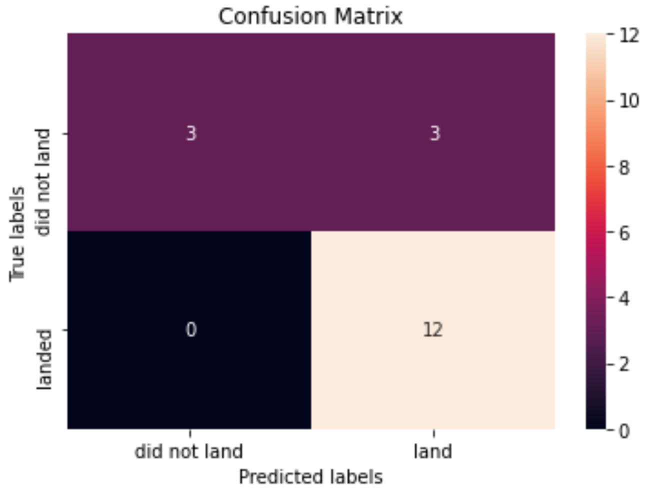
PREDICTIVE MODELING

Machine Learning Techniques:

- Algorithms Tested: Logistic Regression, SVC, Decision Tree Classifier, K Neighbors Classifier.
- Hyperparameter Tuning: Used GridSearchCV to optimize model performance.

Details: This page provides details on the machine learning models tested and the methodology used to evaluate their performance.





CONCLUSION

- The "Winning the Space Race with Data Science" project effectively used data science methodologies to predict the success of SpaceX Falcon 9 rocket landings. By analyzing various factors like launch sites, payloads, and orbits, the project identified key elements that influence successful landings. The Decision Tree Classifier emerged as the most accurate predictive model, demonstrating the practical application of machine learning in aerospace operations. This project underscores the importance of data-driven decision-making in achieving higher success rates and optimizing costs in space exploration.