IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Tatiana Gebrayel
7-7-22

Outline

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

Executive Summary

- **Summary of methodologies**
  - Data collection
  - Data wrangling
  - EDA with data visualization
  - EDA with SQL
  - Building an interactive map with folium
  - Building a dashboard with Plotty Dash
  - Predictive analysis (Classification)
- **Summary of methodologies**
  - Exploratory data analysis results
  - Interactive analytics demo in screenshots
  - Predictive analysis results

Introduction

- Project background and context

We predicted if the Falcon 9 first stage will land successfully.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 candetermine 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

- What influences if the rocket will land successfully?

- The effect each relationship with certain rocket variables willimpact in determining the success rate of a successful landing

- What conditions does Spacex have to achieve to get the best results and ensure the best rocket success landing rate.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX Rest API
  - (web Scrapping) from wikipedia
- Perform data wrangling
  - Transforming data for Machine Learning
- 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

- data sets were collected using he API call from several websites

  https://api.spacexdata.com/v4

1. Collecting data with API call
2. Converting to data frame using JSON
3. Updating columns and rows
4. Filtering the data to keep Falcon 9 launches only
5. Convert data to csv file.

Data Collection – SpaceX API

1. Collecting data with API call

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

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

2. Converting to data frame using JSON

```
# Use json_normalize meethod to convert the json result into a dataframe
data = pd.json_normalize(response.json())
```

3. Cleaning Data

```
# Call getLaunchSite
getLaunchSite(data)
```

```
# Call getPayloadData
getPayloadData(data)
```

```
# Call getCoreData
getCoreData(data)
```

4. Assign list to dictionary

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

5. Filter dataframe and Convert data to csv file.

```
data_falcon9.loc[:,'FlightNumber'] = list(range(1, data_falcon9.shape[0]+1))
data_falcon9
```

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

https://github.com/tatianagebrayel/Data-science-capstone/blob/master/collecting%20data.ipynb

Data Collection - Scraping

- web scraping from Wipedia
- https://github.com/tatianagebrayel/Data-science-capstone/blob/master/web%20scraping.ipynb

1. Getting Response from HTML

```
data = requests.get(static_url).text
```

2. Creating BeautifulSoup Object

```
soup = BeautifulSoup(data, 'html5lib')
```

3. Finding Tables

```
html_tables = soup.find_all('table')
```

4. Getting Column names

```
column_names = []
```

5. Creating dictionary

```
launch_dict= dict.fromkeys(column_names
del launch_dict['Date and time ( )']
launch_dict['Flight No.'] = []
launch_dict['Launch site'] = []
launch_dict['Payload'] = []
launch_dict['Payload mass'] = []
launch_dict['Orbit'] = []
launch_dict['Customer'] = []
launch_dict['Launch outcome'] = []
launch_dict['Version Booster']=[]
launch_dict['Booster landing']=[]
launch_dict['Date']=[]
launch_dict['Time']=[]
```

6. Appending Data to Keys

```
extracted_row = 0
#Extract each table
for table_number,table in enumerate(soup.find_all('table',"wikitable plainrowheaders collapsible")):
   # get table row
```

7. Converting dictionary to dataframe

```
df=pd.DataFrame(launch_dict)
```

8. Converting Dataframe to csv

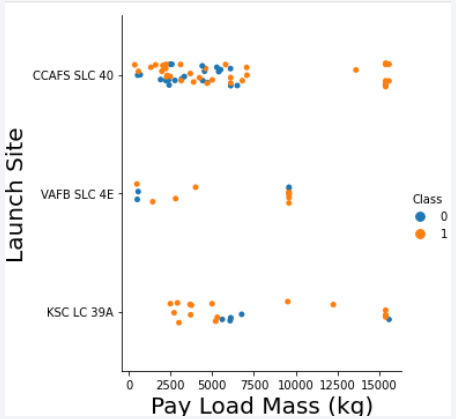
```
df.to_csv('spacex_web_scraped.csv', index=False)
```

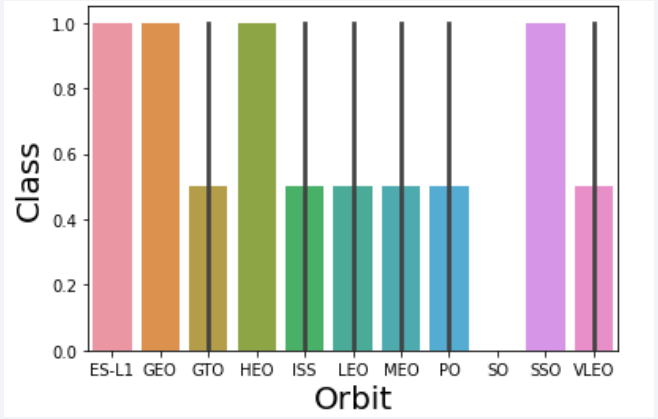
Data Wrangling

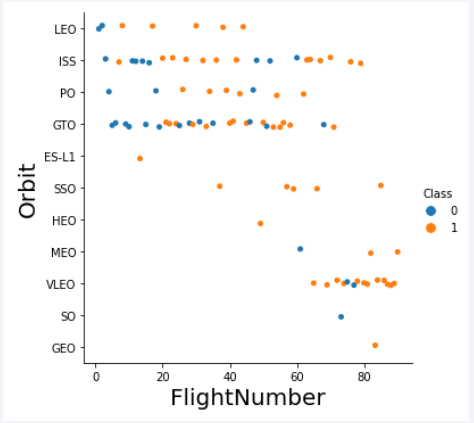
- Check null values
- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurrence of mission outcome per orbit type
- Create a landing outcome label from Outcome column
- Handle null values

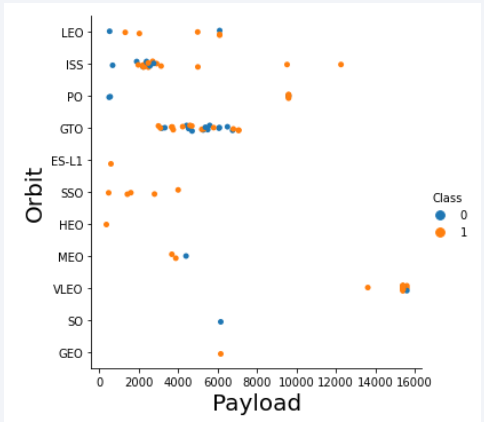
https://github.com/tatianagebrayel/Data-science-capstone/blob/master/data%20wrangling.ipynb

EDA with Data Visualization









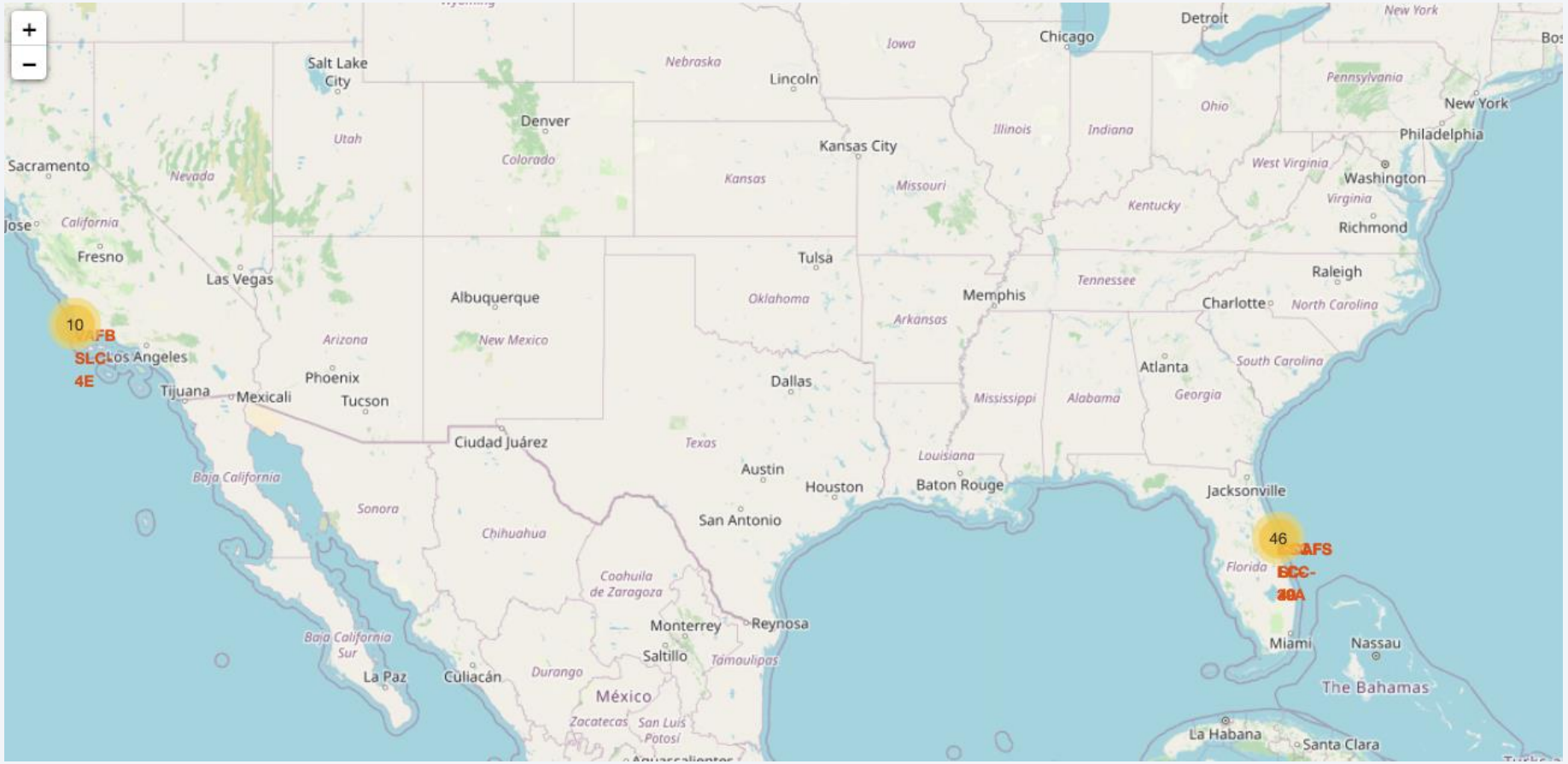
- https://github.com/tatianagebrayel/Data-science-capstone/blob/master/EDA%20with%20data%20visualization.ipynb

EDA with SQL

- Displaying names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with ‘KSC’
- Displaying the total and average payload mass carried by boosters
- Listing the date where the successful landing outcome in drone ship was achieved
- Listing names of the boosters which payload mass is between 4000 and 6000
- Listing the total number of successful and failure mission outcomes…

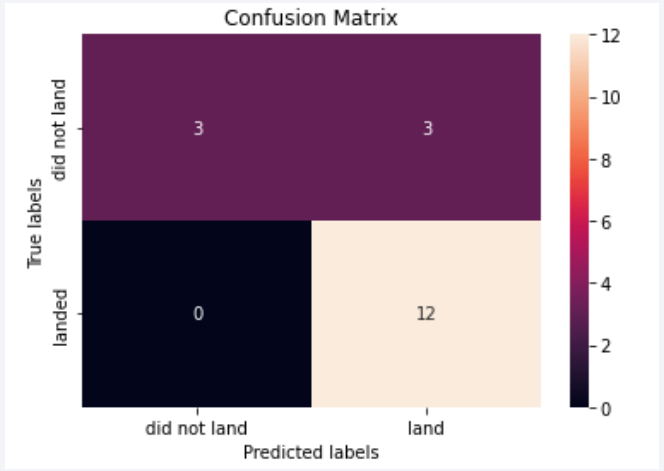
- https://github.com/tatianagebrayel/Data-science-capstone/blob/master/EDA%20with%20SQL.ipynb

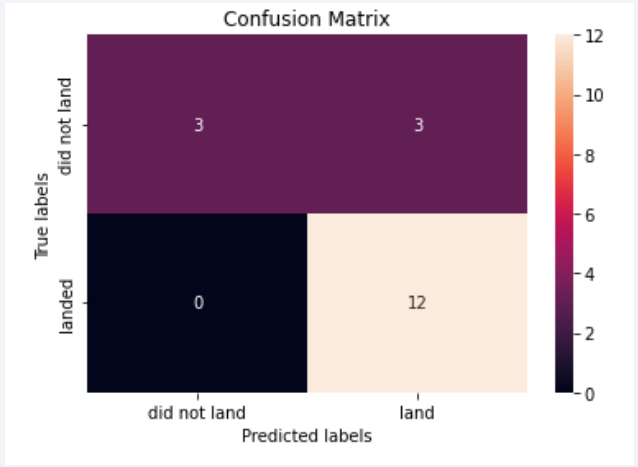
Build an Interactive Map with Folium

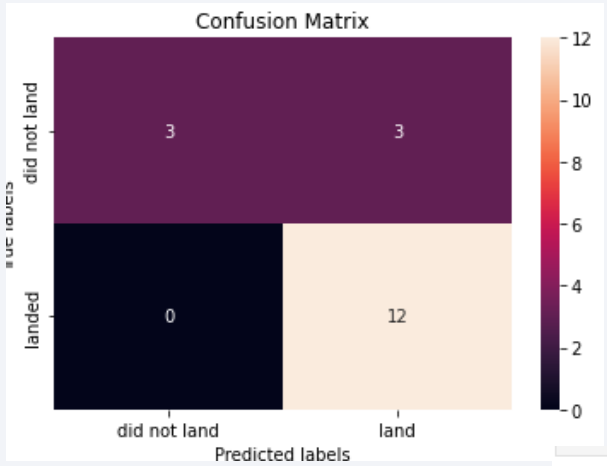


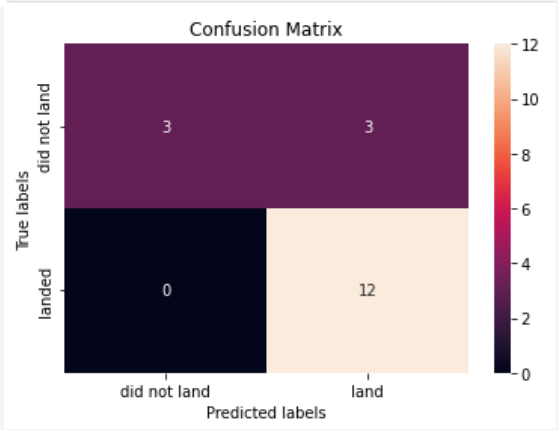
- https://github.com/tatianagebrayel/Data-science-capstone/blob/master/folium%20lab.ipynb

Predictive Analysis (Classification)









- Highest accuracy : 83.3%
- SVM performs the best in terms of Area Under the Curve at 0.958

- https://github.com/tatianagebrayel/Data-science-capstone/blob/master/machine%20learning%20prediction.ipynb

Results

- The SVM ,KNN and Logistic Regression models are the best in terms of prediction accuracy for this dataset
- Low weighted payloads perform better than the heavier payloads
- The success rates for SpaceX launches is directly proportional time in years
- KSC LC 39A had the most successful launches from all sites
- Orbit GEO,HEO,SSO, ES L1 have the best success rate

Thank you!