Precipitation Prediction based on NOAA JFK STATION

University of Toronto – Data Analytics

Project 4 Machine learning

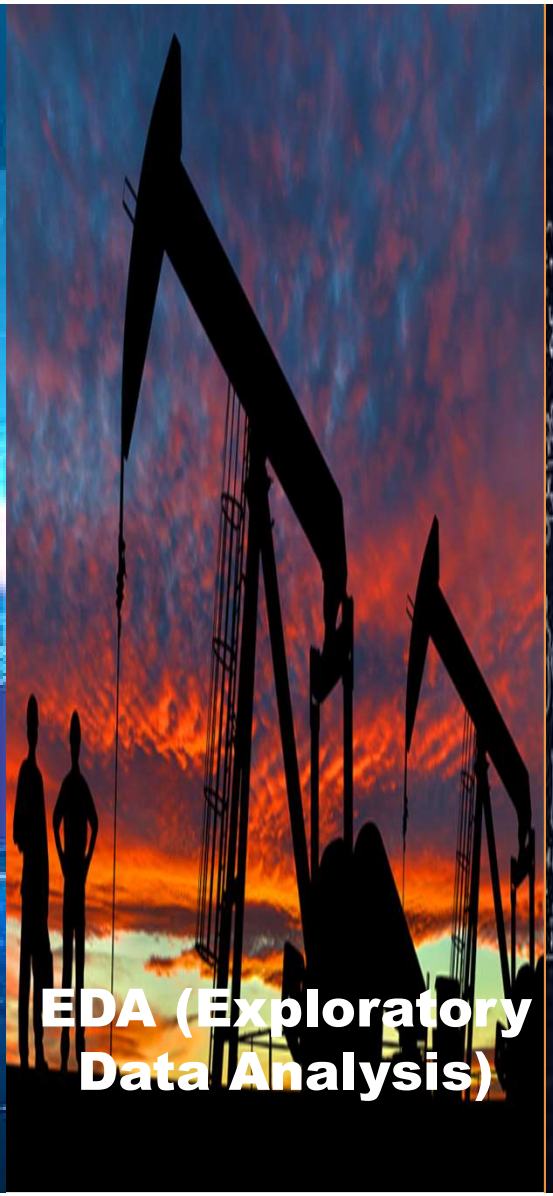
Group Member Names:
Daniel Marquez
Peiran (Diana) Cao
Lailah Libay

Data-Alchemists

About the project

- Predictive precipitation modelling of weather patterns using NOAA station data (JFK)
- Select monthly data from 2010-2022
- Variables select wind, precipitation, temperature, pressure, fog, thunderstorms
- Supervise Machine learning Model: Regression, linear model, means square
- Visualization: Tableau


ETL （Extract，
Transform，
Load)

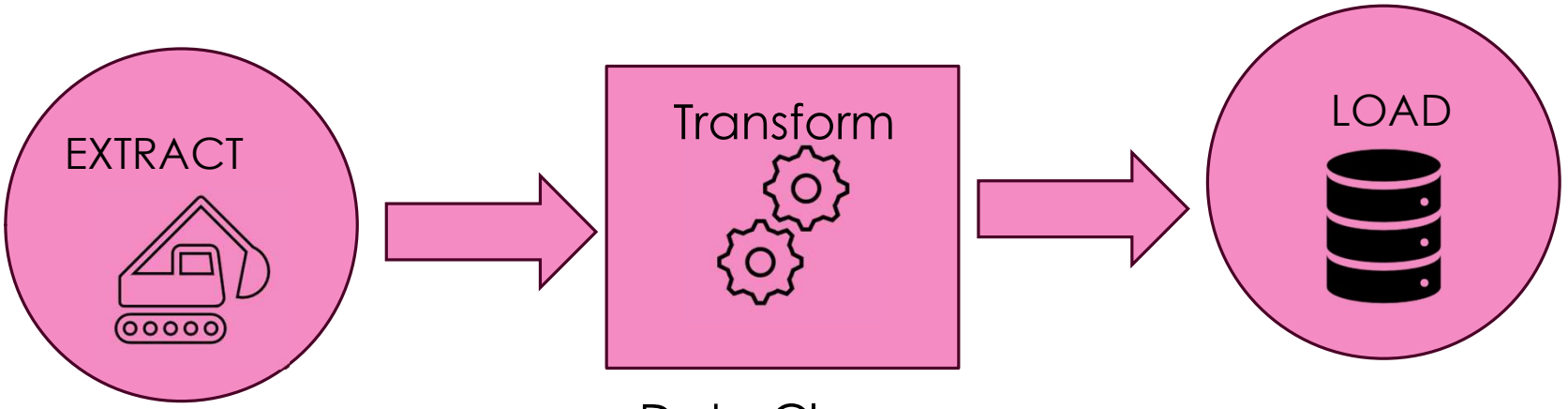

EDA (Exploratory
Data Analysis)


Machine
Learning


Visualizations

ETL

Main Menu



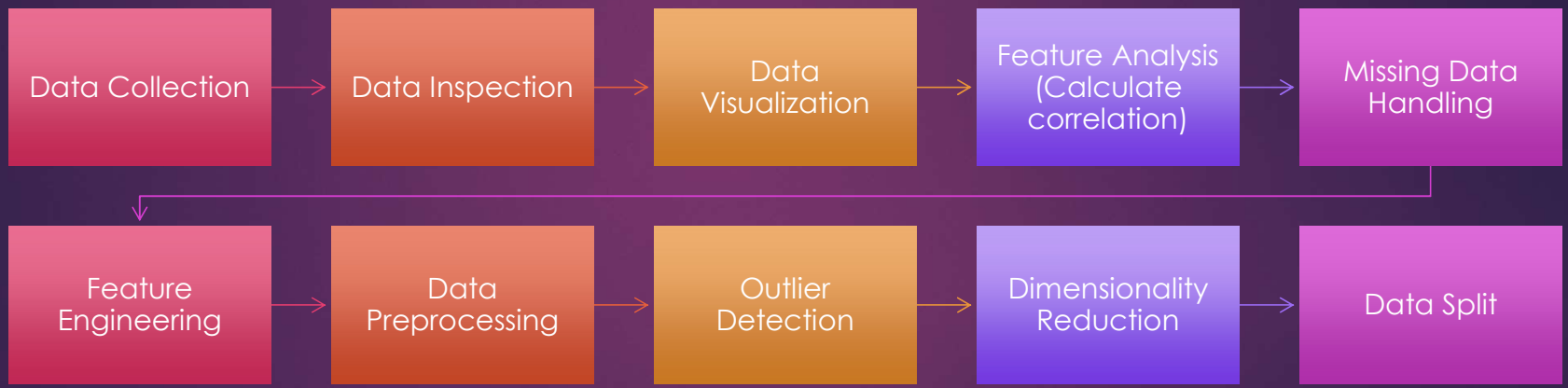
NOAA.GOV API

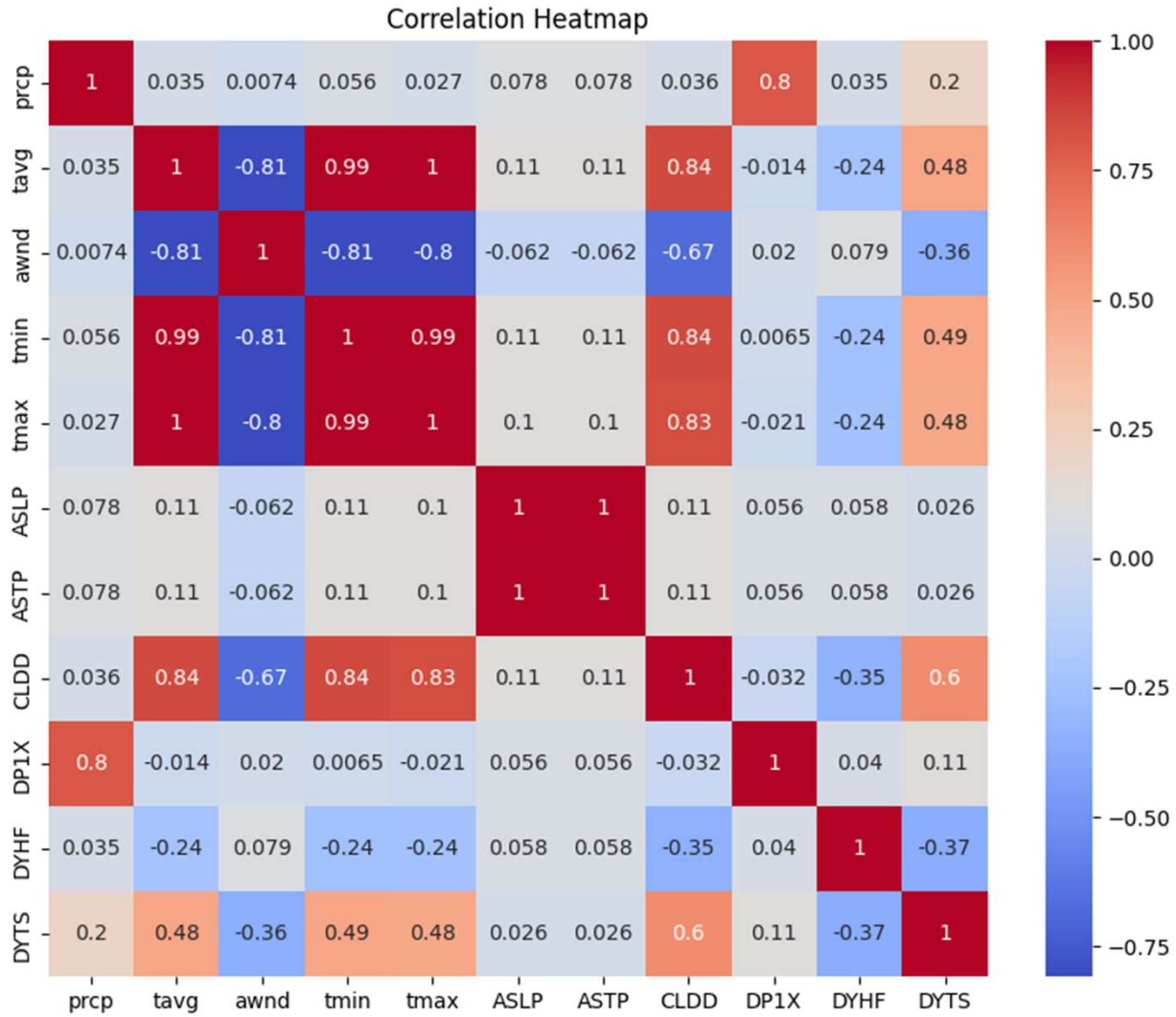
Data Clean
Filter
Summarize

PostgreSQL

Main Menu

EDA (exploratory Data analysis)





Main Menu

Feature Analysis (Calculate correlation)

Scikit-learn's train_test_split function

Main Menu

Machine Learning: Process

Import	Use	Utilize	Scale	Assign	Apply	Print	Identify
Import Dependencies for machine learning	Use extracted data obtained in ETL	Utilize dependent and independent variables selected in EDA	Scale Data	Assign Train & Test data using Scikit-learn's train-test split function	Apply Supervised Learning Linear Regression models: •Linear_model •Ridge Regression •Lasso Regression •Elastic Net Regression	Print R2 and mean-squared error scores	Identify the optimal Linear Regression model to use

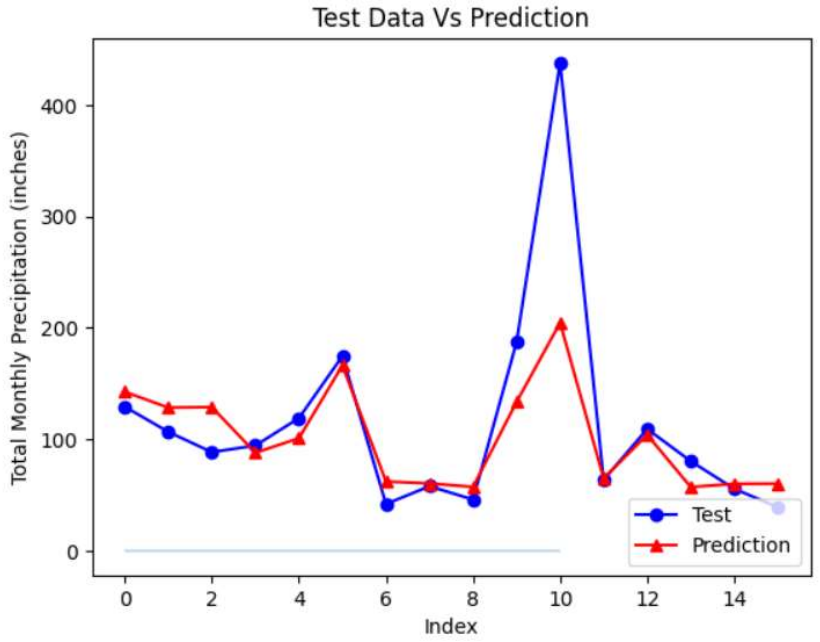
Machine Learning: Test VS Predicted Data

Main Menu

```
#Merge y_test & y_predict DataFrames

test_predict = pd.merge(y_test3, y_predict3, left_index=True, right_index=True)
test_predict.head()
```

	y_test	Prediction
0	129.3	142.677156
1	106.7	128.468837
2	88.5	128.932716
3	94.4	87.986374
4	119.0	100.952022



Machine Learning: MSE & R2 Scores

```
# Test  mean squared error
mse=mean_squared_error(y_test, y_predict)
print(f" MSE - Mean Squared error for linear_model is: {mse}")
```

```
 MSE - Mean Squared error for linear_model is: 3851.2390384534388
```

```
from sklearn.metrics import r2_score
score = r2_score(y_test, y_predict)
print("The accuracy of our model is {}%".format(round(score, 2)*100))
```

```
The accuracy of our model is 56.00000000000001%
```

```
Average Temperature (F): 13
Average windpeed (miles/hr): 4.9
Minimum Temperature (F): 8.1
```

User Input Data

Main Menu

```
Average Temperature (F): 13
Average windpeed (miles/hr): 4.9
Minimum Temperature (F): 8.1
Maximum Temperature (F): 17.9
Average Sea Level Pressure(mb) e.g 1014: 1012.9
Average Station Level Pressure(mb) e.g 1014: 1012
# of days more than 65(F) / 18.3(C) : 4
# of days with heavy precipitation : 1
# of days with heavy fog : 2
# of days with thunderstorm : 0
Predicted Total Precipitation for the month (Inches):[[86.96269029]]
```

Predicted result

Machine Learning (Applying Prediction Model)

NOAA JFK (Airport) 2010-2022 Weather

Overview | Precipitation | Average Temperature VS Precipitation

Measure Names
Avg. Precipitat

Average Temperature (°F) VS Precipitation (inch)

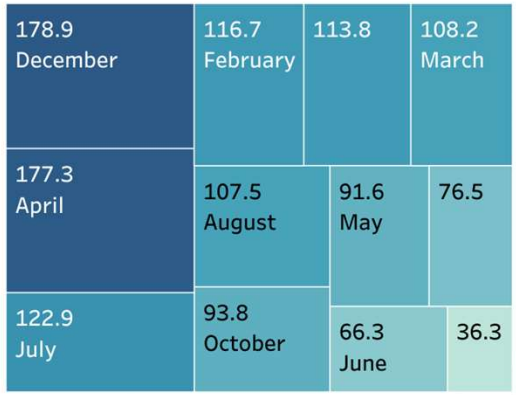
Average Temperature (°F)/Monthly

Select Year
2014

Avg. Precipitation
36.3 178.9

Avg. Temperatur..
-1.84 24.15

Average Precipitation (inch)/Monthly



Main Menu

Visualizations

Click Graphic to open Tableau

Main Menu

QUESTION TIME

Main
Menu

Thank you