



Spain Electricity Shortfall

Pinnacle Data Solutions

Table of Contents



01

Our Company



02

Overview



03

Exploratory Data Analysis



04

Data Preprocessing and
Feature Engineering



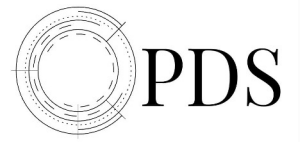
05

Model Development,
Evaluation and Deployment



06

Conclusion



01

Our Company

About us



Pinnacle Data Solutions is a new leading technology company founded in 2022.



We are a proudly **African company**.



We have experience across **multiple industries**.



Efficiency



Collaboration



Innovation

Meet the Team



Thepe Mashala
Technical Project
Manager



Sibusiso Mashabela
Senior Data
Scientist/Engineer



Maryam Ojulari
Senior Business Architect



Ereshia Gabier
Lead Data
Scientist/Engineer



Sibusiso Sibiya
AWS Cloud Specialist

02

Overview

Introduction

Overview

Fossil fuels power economies by currently supplying **>80%** of world's **energy**.

Address the impacts of **climate change** on health and to reduce pollutants.

Renewable energy is the key to address **sustainable** living.



Problem Statement and Objective

Overview



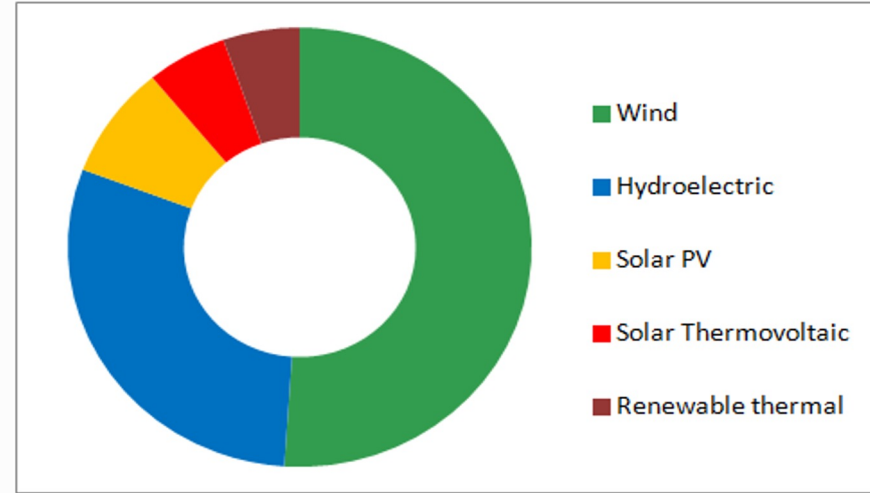
Advancement of any developmental **economy**.



Spain's **2030 emission reduction** target is 23%.

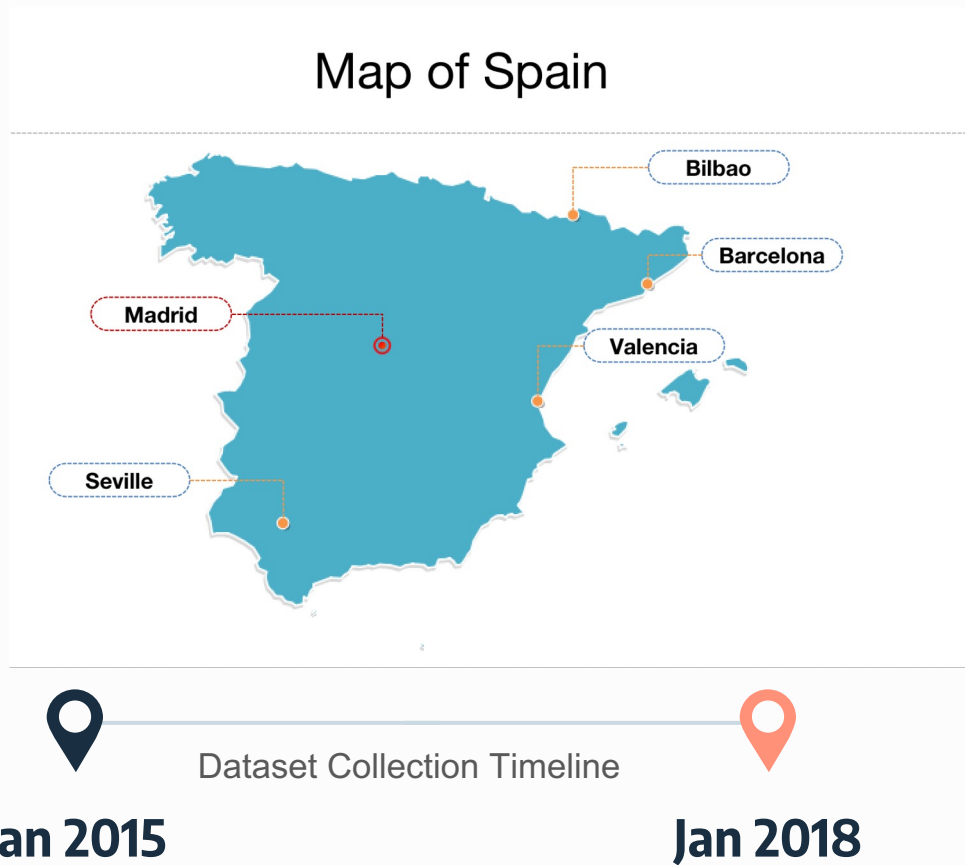


The **aim** is to provide a solution that would guide in decision-making on infrastructure development for renewable energy sources.



By afloresm - SOLUCAR PS10, CC BY 2.0,
<https://commons.wikimedia.org/w/index.php?curid=2821733>

Dataset



Data source: **Spain Electricity Shortfall Challenge 2022** on Kaggle

03

Exploratory Data Analysis

Training Dataset

- 8763 rows (observations)
- 49 columns (features).



Categorical Features -
3 features (Object)

Numerical Features -
46 features (int64 and float64)



**Features with missing
Values -**
1 feature

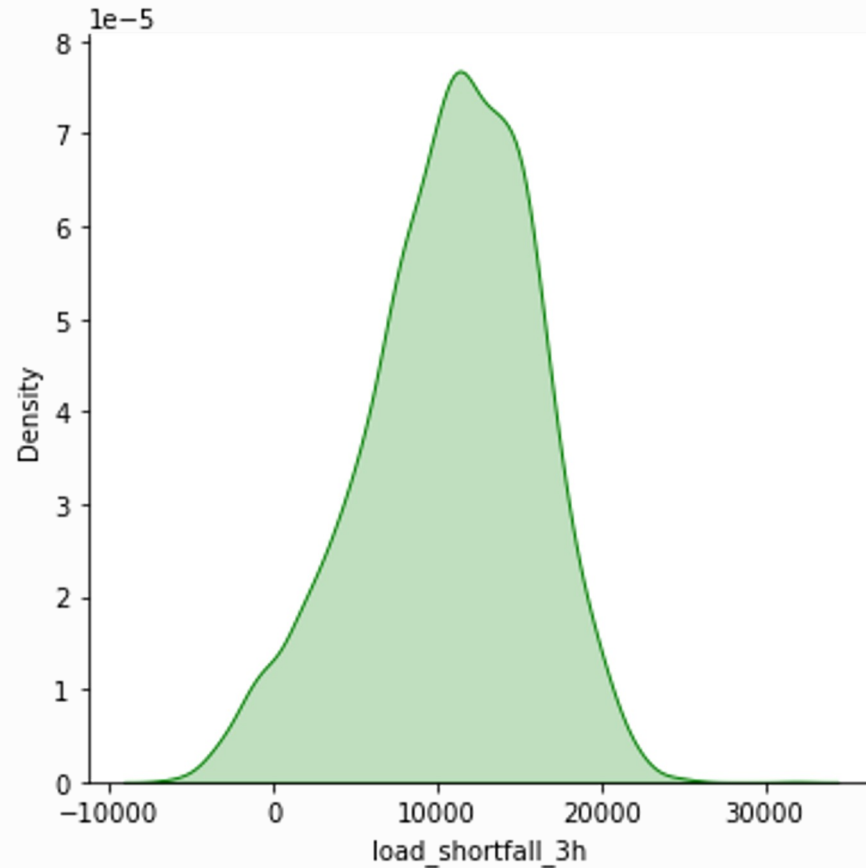
Test Dataset

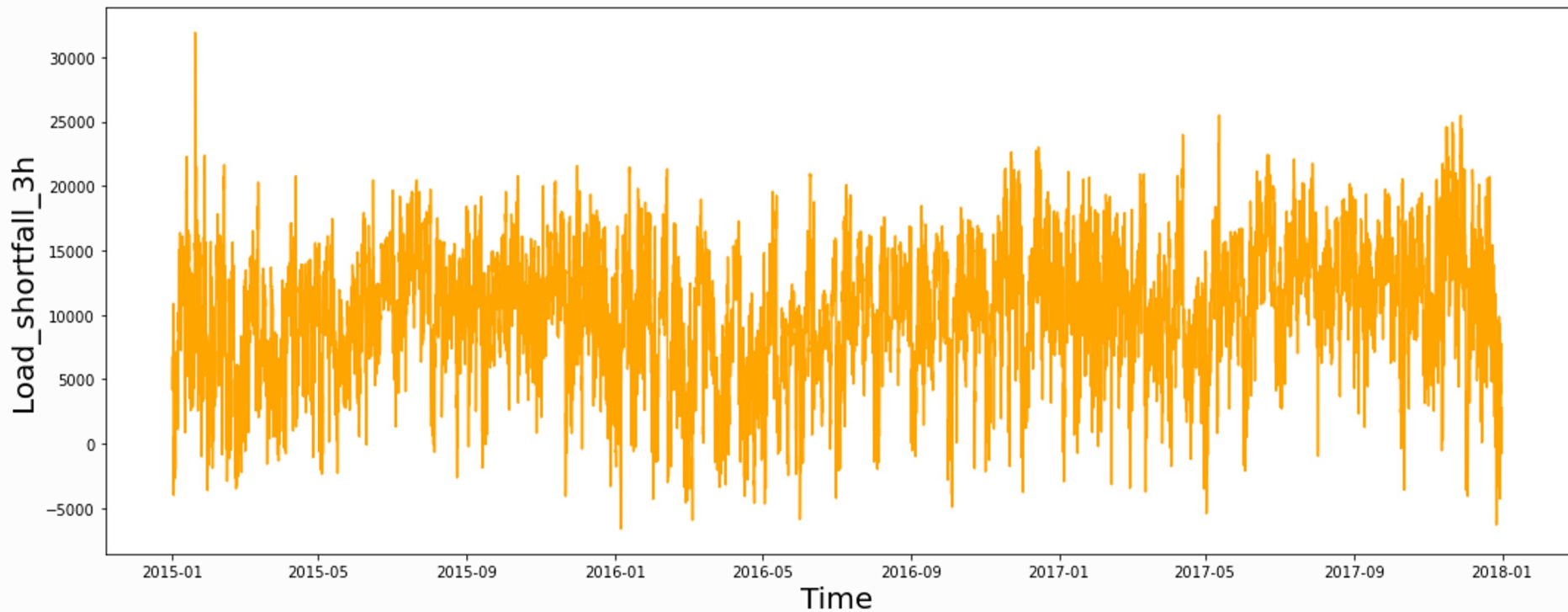
- 2920 rows (observations)
- 48 columns (features)



Categorical Features -
3 features (Object)

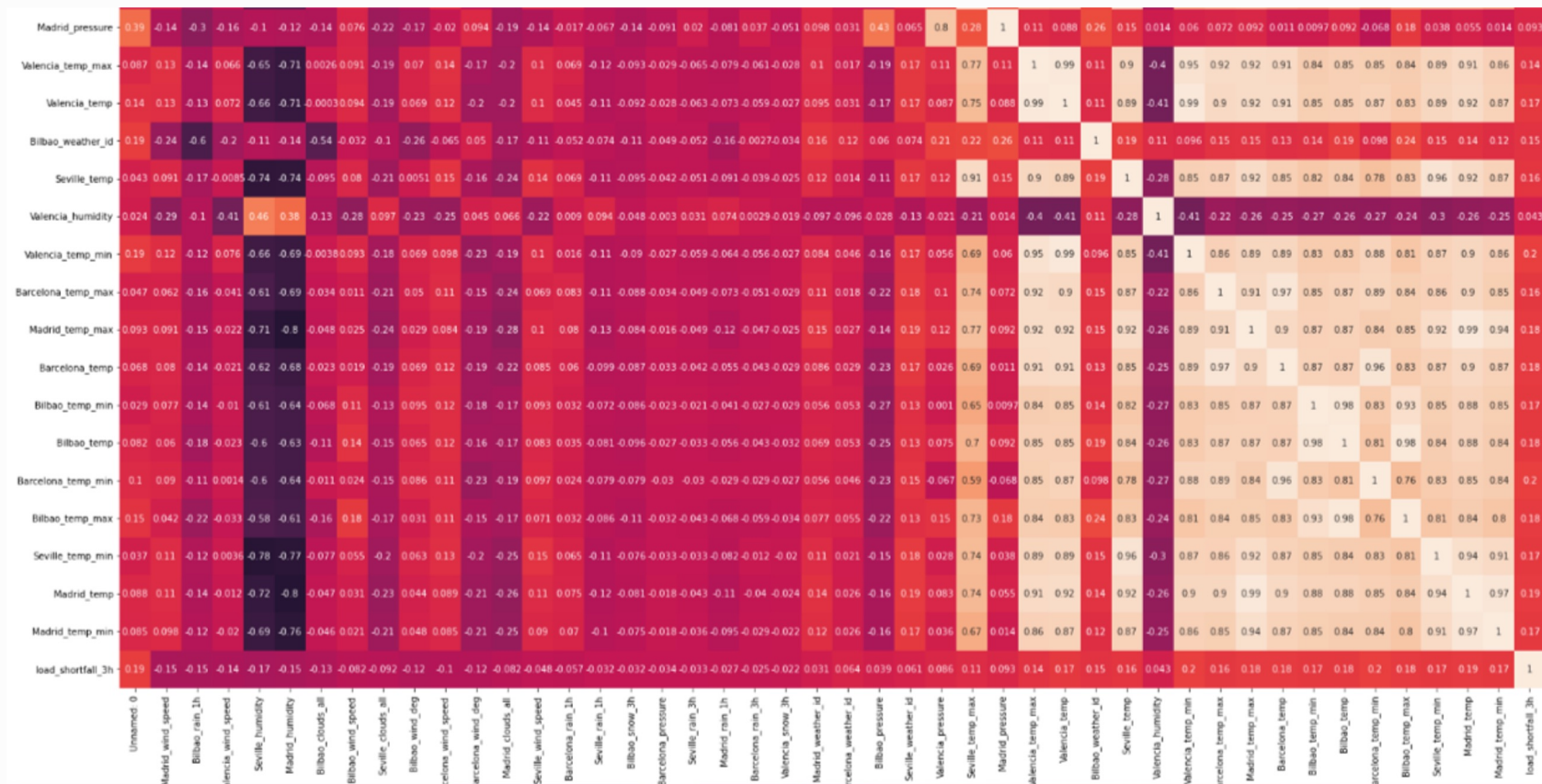
Numerical Features -
45 features (int64 and float64)





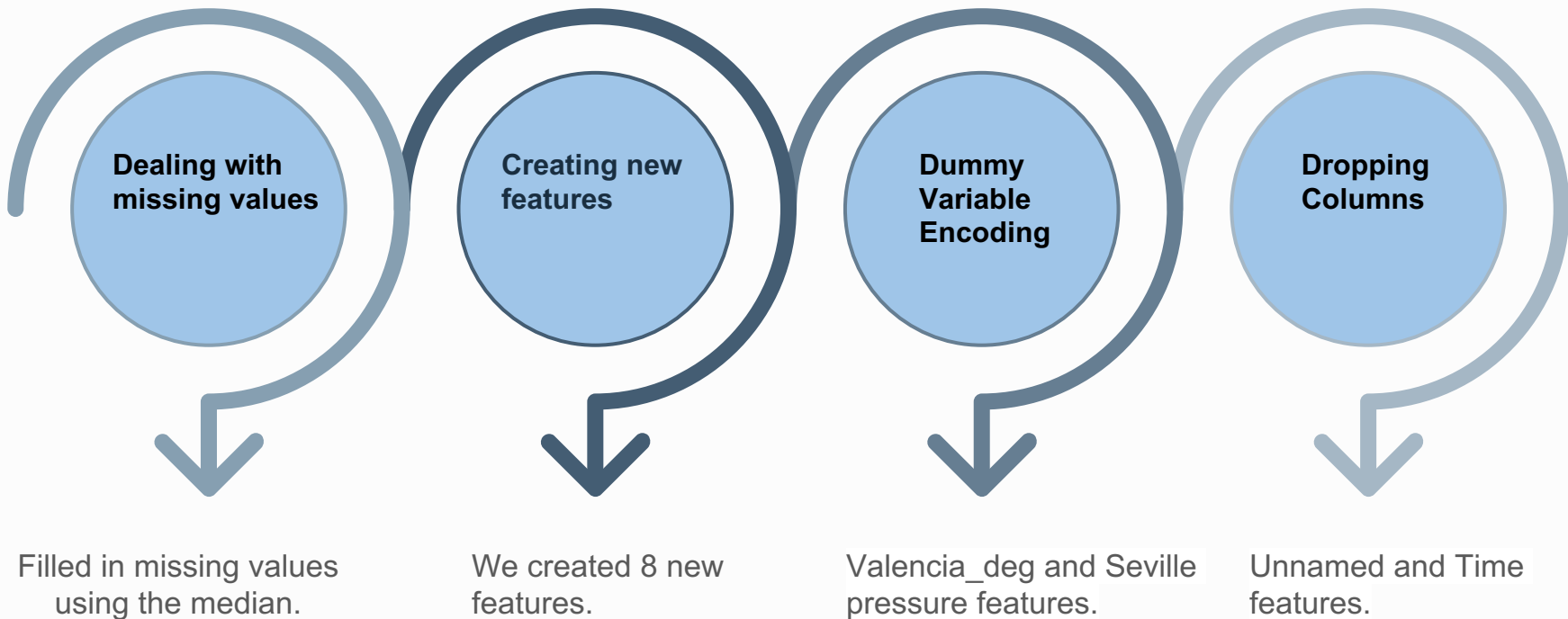
Features relationship Heatmap

Exploratory Data Analysis



04

Data Preprocessing and Feature Engineering



05

Model Development and Evaluation

1

Linear Regression

Assumes a linear relationship between the predictor variables and predicted variables.

2

Ridge Regression

This is a shrinkage method that reduces large coefficients.

3

Random Forest

An ensemble method of decision trees for each feature in our dataset.

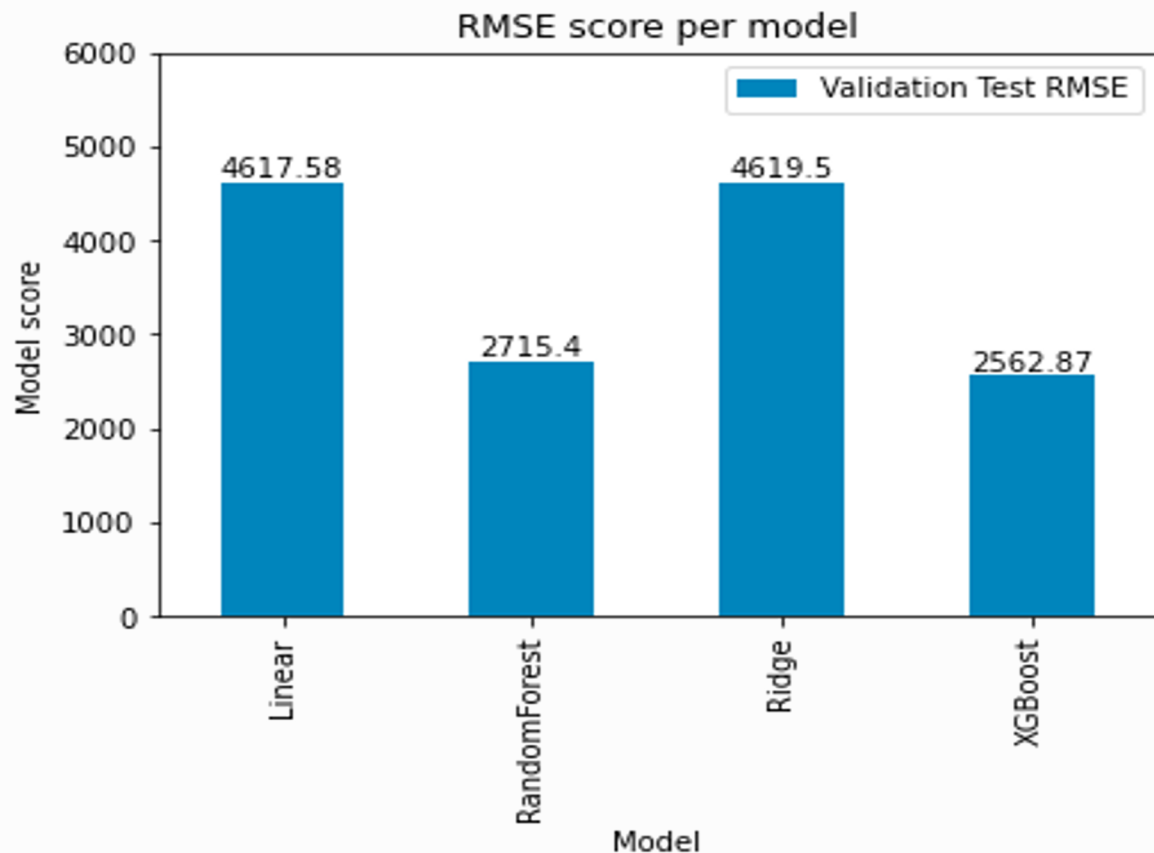
4

XGBRegressor

This ensemble model uses gradient boosting.

Root Mean Square Error (RMSE)

Model Evaluation



Deployed model via API

Model Deployment

1



2



3



```
random-forest_model.pkl 100% 108MB 688.1KB/s 02:40

Humani@ComputerOne MINGW64 ~/Documents/GitHub/EC2-Test (main)
$ scp api.py df_test.csv df_train.csv model.py request.py explore-student@18.203.115.119:EC2-Test
explore-student@18.203.115.119's password:
api.py 100% 2223 9.2KB/s 00:00
df_test.csv 100% 1207KB 315.0KB/s 00:03
df_train.csv 100% 3638KB 262.4KB/s 00:13
model.py 100% 4552 20.5KB/s 00:00
request.py 100% 1833 7.3KB/s 00:00

Humani@ComputerOne MINGW64 ~/Documents/GitHub/EC2-Test (main)
$ python request.py
Sending POST request to web server API at: http://18.203.115.119:5000/api_v0.1

Querying API with the following data:
[8764, '2018-01-01 03:00:00', 4.66666666667, 'level 8', 0.0, 5.33333333333, 89.0, 78.0, 0.0, 3.66666666667, 0.0, 143.3333
33333, 4.66666666667, 266.6666666667, 0.0, 0.6666666667, 0.0, 'sp25', 0.0, 0, 1020.3333333333, 0.0, 0.0, 0.0, 0, 800.0,
00.3333333333, 1026.666666667, 800.0, nan, 282.4833333333, 1030.3333333333, 284.15, 284.15, 721.0, 281.6733333333, 53.
666666667, 284.15, 284.816666667, 280.4833333333, 284.19, 277.816666667, 281.01, 283.4833333333, 284.15, 281.15, 279.
933333333, 278.15]

Received POST response:
*****
API prediction result: 11042.770000000004
The response took: 0.547264 seconds
*****

Humani@ComputerOne MINGW64 ~/Documents/GitHub/EC2-Test (main)
$
```

06

Conclusion

Conclusion

The Government of Spain should expand its energy mix by extending its use of renewable resources such as wind, solar and hydro power over and above the use of fossil fuels.

The deployed API would enable informed decisions based on the weather data from the cities.

Invest to improve or add renewable energy infrastructure.





Thank You

