

EXPLORE || DIGITAL SKILLS

Python for Data Science
Predict

Contents

[Problem Context/Domain](#)

[Project Requirements](#)

[Function Descriptions](#)

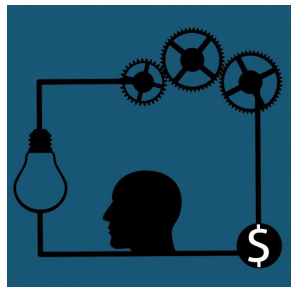
[Predict FAQs](#)



Your Project | Building Functions to Calculate Metrics using Eskom Data

Problem Statement: You need to build python functions which calculate/analyse data from Eskom

Your Context



- You are a **Data Scientist at Eskom**
- **Eskom** requires certain **metrics** to be calculated for their analytics team

Your role



- **Build 7 functions** using **python**
- These **7 functions** will need to process both **numeric & text data**

Stuff you need to Know



- List manipulation
- Dictionaries
- Basic Statistics and Aggregations
- Function definitions
- PEP8 coding style

Your Job



- **Write 7 functions** which outputs **metrics**
- **Submit functions** at the end of the sprint

Your Project | Building Functions to Calculate Metrics using Eskom Data



For this project your main task will be to build a **module** that looks at the following::

1

Numerical Metrics
(Statistics & Time Series
analysis)

2

Twitter Data Processing
(Text/Language)

What are you building?

Functions which take in a list or a pandas dataframe and returns either a **dictionary** or a **list**

Functions which take in text data as either a list or as a dataframe and returns a **list** or **pandas dataframe**

What do you need to do?

Write and submit 3 functions of increasing difficulty:

- **2 Functions** on **calculated metrics**
- **1 Function** on **time data**

Write and submit 4 functions of increasing difficulty

Your Project | Building Functions to Calculate Metrics using Eskom Data

For this project you will need to do the following:

Write 7 Functions:

Project | Function 1

Build a function that summarises data into a dictionary:

1

Metric
Dictionary

Task Instructions

Write a function that **calculates metrics** from Eskom data and outputs the metrics as a **dictionary**. These metrics should include:

- Mean
- Median
- Maximum
- Minimum
- Standard Deviation
- Variance

Rules:

No extra packages may be imported

Project | Function 2

Build a function which calculates the 5 number summary:

2

5 Number
Summary

Task Instructions

Write a function that **returns a dictionary of the 5 number summary**:

Rules:

No extra packages can be imported

Project | Function 3

Build a function which takes a date and converts it to a certain format:

3

Date Parser

Task Instructions

Write a function which takes in a date in one format and converts it into another format:

- The function must take in a **date** and output a **string** with the correct format

Rules:

No extra functions may be used.

Project | Function 4

Build a function that returns a portion of grouped data:

4

Hashtag &
Municipality
Extractor

Task Instructions

Write a function that returns **a dataframe** :

- **With an added column of extracted hashtags from each tweet**
- **With an added column of the municipality mentioned in each tweet**

Rules:

Only pandas may be used

Project | Function 5

Build a function that returns a portion of grouped data:

5

Number of
Tweets per
day

Task Instructions

Write a function that returns **The number of tweets per day :**

Rules:

Only pandas may be used

Project | Function 6

Build a function that returns a portion of grouped data:

6

Word
Splitter

Situation

Write a function that **splits a sentence into a list of individual words**:

- **Takes in a dataframe and outputs a dataframe with a new column of a list of individual words**

Rules:

Only pandas may be used

Project | Function 7

Build a function that returns a portion of grouped data:

7

Stop Word
Remover

Task Instructions

Write a function that returns **a dataframe with::**

- **An added column of the tweet without stopwords**
- A dictionary of stop words will be provided

Rules:

Only pandas may be used

Predict-related FAQs

This page will be updated periodically with common predict-related questions which may arise during the Sprint. Consider consulting this space before asking your course facilitator a question.

