EXPLORE | DIGITAL SKILLS

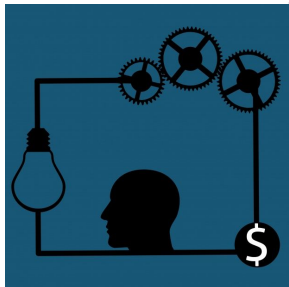Python for Data Science
Predict

Contents

EXPLORE | DIGITAL SKILLS

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

EXPLORE | DIGITAL SKILLS

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::

	What are you building?	**What do you need to do?**
1 Numerical Metrics (Statistics & Time Series analysis)	**Functions** which take in a list or a pandas dataframe and returns either a **dictionary** or a **list**	**Write and submit 3 functions** of increasing difficulty: ➢ **2 Functions** on **calculated metrics** ➢ **1 Function** on **time data**
2 Twitter Data Processing (Text/Language)	**Functions** which take in text data as either a list or as a dataframe and returns a **list or pandas dataframe**	**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:

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

1

Metric
Dictionary

EXPLORE DIGITAL SKILLS

Project | Function 2

Build a function which calculates the 5 number summary:

Task Instructions
Write a function that **returns a dictionary of the 5 number summary**: **Rules:** No extra packages can be imported

2

5 Number
Summary

Project | Function 3

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

Task Instructions

3

Date Parser

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.

EXPLORE || DIGITAL SKILLS

Project | Function 4

Build a function that returns a portion of grouped data:

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

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

Rules:
Only pandas may be used

EXPLORE ║ DIGITAL SKILLS

Project | Function 6

Build a function that returns a portion of grouped data:

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

6

Word Splitter

Project | Function 7

Build a function that returns a portion of grouped data:

Task Instructions

Stop Word
Remover

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

EXPLORE ‖ DIGITAL SKILLS

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.