MSc Data Science and Analytics 2024 Scholarship Award Exam

Data Description

The data contains 80 variables that characterise the demographic and socio-economic situation of 181 galaxies over a period of at most 26 years. A composite index is given that measures their well-being. However, the demographic and socio-economic variables that influence this index is not known. We seek to determine, ***what makes the galaxies better off?***

We would like you to use the data and:

1. Tell us which variables best explain the variance of the well-being index
2. Determine the future well-being values of the galaxies

Submission Instructions and Format

We have provided you data with observed values of the well-being index of each galaxy and a validation dataset that requires the prediction of the future well-being index.

Kindly submit:

1. A report that discusses the demographic and socio-economic determinants of the galaxies' wellbeing.

Submission Format: The report should be a **pdf of a slide presentation** of not more than 5 slides

2. The **predicted future well-being index** values with the highest possible level of certainty using data in the **validation dataset.**

Submission Format: A csv file ***Saved as "firstname_lastname_DSA.csv"*** containing:

Variable	Description
ID	Unique identifier of the observations in the validation dataset
Predicted Well-Being Index	Prediction for the Well-Being Index

3. Analysis file

Submission Format: **python/R notebook** with detailed comments and organised analysis of **EDA with visualisations, well documented analytical process and test results**. Kindly follow reproducibility guides. Sample of best practice can be found [here](#) or [here](#).

Submission Details

Kindly submit your documents, compressed in one folder saved as "***firstname_lastname.zip***",

Send the zipped folder to Laura Mideva, @iLabAfrica Course Administrator.

Email address: LMideva@strathmore.edu

Timeline: **Wednesday 27th March 2024** at **1.00**

EAT

Submission Evaluation Criteria

The submission will be evaluation against the following criteria:

1. The reports will be assessed on the quality of the presentation of the findings: narrative and visualisations
2. The future predictions of the well-being index will be evaluated using RMSE metric
3. . The reproducibility of your codebase