Data Engineer Assessment – Urbanisation Prospects

Problem

You are a Data Engineer at a small Global Investment Management company who have enlisted your team to assist them with assessing global population & urbanization growth rates together with mean income statistics to determine the viability of expanding their product offering reach to additional countries.

Requirement

1. Create a data processing and analytics solution which consumes the data sources listed below, stores, and prepares the data for consumption by internal systems and users and provides visualizations for use by the company's sales and marketing teams*.
2. Be creative in terms of the insights that could be gleaned from the data sources.
3. You will be required to showcase (present and demonstrate) the end-to-end solution and will be required to talk through and explain each step & component of your solution.
4. Where possible, code should preferably be published and made available to us via GitHub.

*Note: We understand you have limited time available to complete this assessment, so components of your solution which could be time-consuming to implement can be reduced in scope or excluded. However, please be prepared to discuss these.

Technologies & Resources

Technologies

You are welcome to use any data, analytics, and reporting platforms, however, the following technologies are aligned to those in use at Ninety One. Each is available as a free trial.

- Databricks
- Microsoft Fabric
- PowerBI

Resources

Please use the following public sample datasets as your source data

- Urban and rural population 1950-2050 - UN World Urbanization Prospects 2018
- World Inequality Database (WID) - Pretax income.csv
- Countries Continents
- OWID country to WHO regions

Evaluation Criteria

- The requirements in this assessment are deliberately vague as rather than assessing the most technically correct solution, we are interested in your creativity, thought processes, and approach to deriving a technical solution.
- This assessment should be seen as an opportunity to showcase your knowledge and highlight those aspects of data engineering which you believe to be important.
- The following areas will be of interest:
    - Ability to create a robust, maintainable, testable solution, with thought given to reusability, performance, and redundancy.
    - Ability to understand requirements, understand the source data & relationships between the data, and translate those into creative, meaningful outcomes.