**Senior BI Analyst assignment.**

Considering that Village enterprise has a preference for STATA for data cleaning and analysis (Feel free to use R or related software although we would prefer STATA) and Power BI for Business Intelligence, you have been provided with dataset1 which contains information on **household weekly consumption/expenditure** and savings. You are required to attempt the following questions.

Question 1:

1. How many observations and variables are in the dataset?
2. What is the unit of observation?
3. Data contains some duplicates in terms of household ID (hh\_id). How many duplicates are there in this data set?
4. List country hh\_id and respondent given name 1 (given\_name\_household\_member\_1) for all duplicate observations
5. Drop excess observations as follows,

Where hh\_id is 174179, drop the observation that has the respondent as Scovia

Where hh\_id is 178903, drop the observation that has the respondent as David

Where hh\_id is 187589, drop the observation that has the respondent as Immaculate

1. What is the proportion of households from Uganda?

Question 2:

1. Rename the following variables

|  |  |
| --- | --- |
| **Old variable** | **New Variable** |
| how\_much\_does\_a\_unit\_of\_sweet\_po | ucost\_sweetpotatoes |
| quantity\_other\_food | quantity\_otherfood |
| ucode\_other\_food | ucode\_otherfood |
| consumption\_avocado | consumption\_avocados |
| consumption\_otherfruitveg\_descri | specify\_otherfruitveg\_descri |

Save the clean dataset with the name “Dataset1 clean”

1. This data is wide by **hh\_id**. Reshape the data to a long format for all **consumption\_** **quantity\_ ucode\_** **ucost\_**. Save the new variable created as “**Product**”
2. How many observations and variables now exist?

Question 3:

1. Replace both quantity and **ucost** to zero, where there was not consumption (consumption=No)
2. Replace both quantity and **ucost** to missing where the quantity captured is either equal to **99**, **98** or **999**
3. For all values of **ucost** less than 100 shillings in **Uganda**, replace them to **missing**
4. For all values of **ucost** less than 5 shillings in **Kenya**, replace them to **missing**
5. By **product** and **country**, calculate the median **ucost** and **quantity**
6. Replace all missing values with median ucost and median quantity

Question 4:

Analyze the dataset and develop a Power BI dashboard that provides insights into the following areas:

1. Household Demographics:

* Analyse and present the distribution of households by the number of household members.
* Show the proportion of households from Uganda compared to other countries.

1. Product Consumption and Costs:

* Identify the top 5 most commonly consumed products based on quantity. - ***Done***
* Calculate and display the average cost per unit for each product category.

1. Comparative Analysis:

* Compare the average consumption quantities and costs of products between different subcounties.
* Highlight any significant differences or trends.

1. Geographical Insights:

* Map the households based on their subcounty and show key metrics such as average household size and average product consumption.

1. Any other analysis and visualization you deem necessary.

**Deliverables**

1. Cleaned and transformed datasets
2. Analysis report (can be in PowerPoint or Word format) summarizing your findings from the dataset.
3. STATA or R code describing your workings with clear comments
4. Comprehensive Power BI dashboard based on the insights outlined above. Ensure the dashboard is interactive and user-friendly. Share the .pbix file of your Power BI dashboard.

**Good luck!**