**Overview**

Financial Inclusion remains one of the main obstacles to economic and human development in Africa. For example, across Kenya, Rwanda, Tanzania, and Uganda only 9.1 million adults (or 13.9% of the adult population) have access to or use a commercial bank account.

Traditionally, access to bank accounts has been regarded as an indicator of financial inclusion. Despite the proliferation of mobile money in Africa and the growth of innovative fintech solutions, banks still play a pivotal role in facilitating access to financial services. Access to bank accounts enables households to save and facilitate payments while also helping businesses build up their credit-worthiness and improve their access to other financial services. Therefore, access to bank accounts is an essential contributor to long-term economic growth.

The research problem is to figure out how we can predict which individuals are most likely to have or use a bank account. Your solution will help provide an indication of the state of financial inclusion in Kenya, Rwanda, Tanzania, and Uganda, while providing insights into some of the key demographic factors that might drive individuals’ financial outcomes.

In order to work on the above problem, you need to do the following:

* + 1. Define the question, the metric for success, the context, experimental design taken and the appropriateness of the available data to answer the given question
    2. Find and deal with outliers, anomalies, and missing data within the dataset.
    3. Perform **univariate**, **bivariate** and **multivariate** analysis recording your observations.
    4. Implement the solution by performing the respective analysis i.e. factor analysis, principal component analysis, and discriminant analysis.
    5. Challenge your solution by providing insights on how you can make improvements.

NB: Remember to go through the rubric [[Link]](https://moringaschool.instructure.com/courses/93/assignments/432) to get an understanding of how you will be graded.

**Dataset**

The main dataset contains demographic information and what financial services are used by individuals across East Africa. This data was extracted from various Finscope surveys ranging from 2016 to 2018, and more information about these surveys can be found here:

* + - FinAccess Kenya 2018. [[Link]](https://fsdkenya.org/publication/finaccess2019/)
    - [(Links to an external site.)](https://fsdkenya.org/publication/finaccess2019/)
    - Finscope Rwanda 2016. [[Link]](http://www.statistics.gov.rw/publication/finscope-rwanda-2016)
    - [(Links to an external site.)](http://www.statistics.gov.rw/publication/finscope-rwanda-2016)
    - Finscope Tanzania 2017. [[Link]](http://www.fsdt.or.tz/finscope/)
    - [(Links to an external site.)](http://www.fsdt.or.tz/finscope/)
    - Finscope Uganda 2018. [[Link]](http://fsduganda.or.ug/finscope-2018-survey-report/)
    - [(Links to an external site.)](http://fsduganda.or.ug/finscope-2018-survey-report/)

Dataset Files:

* + 1. Variable Definitions: <http://bit.ly/VariableDefinitions>
    2. [(Links to an external site.)](http://bit.ly/VariableDefinitions)
    4. Dataset: <http://bit.ly/FinancialDataset>
    5. [(Links to an external site.)](http://bit.ly/FinancialDataset)

# Rubric

**Some rubric (2)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Criteria** | **Ratings** | **Pts** |  |
| This criterion is linked to a learning outcome  Univariate Analysis | |  |  |  |  | | --- | --- | --- | --- | | **4.0 Pts**  **Excellent**  The student exhaustively performs univariate analysis by calculating and interpreting measures of central tendency (mode, median, mean) for a set of data. They also select an appropriate measure of dispersion and correctly calculate and interpret the statistic i.e They show evidence of computing and explaining the range, the interquartile range, the standard deviation, variance, skewness, and kurtosis in a given dataset. They also create and interpret a histogram, a bar chart, a box plot, and a frequency table to analyse variables where necessary. They go a step further to provide recommendations for effectively performing univariate analysis while working on the current problem. | **3.0 Pts**  **Meets Expectations**  The student adequately performs univariate analysis i.e calculating and interpreting measures of central tendency (mode, median, mean) for a set of data. They selected an appropriate measure of dispersion and correctly calculated and interpreted the statistic i.e They show evidence of computing and explaining the range, the interquartile range, the standard deviation, variance, skewness, and kurtosis in a given dataset. They employ appropriate visualization tools. | **2.0 Pts**  **Approaches Expectations**  The student inadequately performs univariate analysis i.e calculating and interpreting measures of central tendency (mode, median, mean) for a set of data. Fails to show evidence of computing and explaining the range, the interquartile range, the standard deviation, variance, skewness, and kurtosis in a given dataset. | **0.0 Pts**  **Needs Improvement**  The student does not show evidence of performing univariate analysis. | | 4.0 pts |  |
| This criterion is linked to a learning outcome  Bivariate Analysis | |  |  |  |  | | --- | --- | --- | --- | | **4.0 Pts**  **Excellent**  The student exhaustively performs bivariate analysis by creating visualizations i.e. scatter plots, line charts to determine the relationship between variables, by describing the relationship between two quantitative variables, and computing and interpreting the Pearson correlation coefficient. They go a step further to provide recommendations for effectively performing bivariate analysis while working on the current problem. | **3.0 Pts**  **Meets Expectations**  The student adequately performs bivariate analysis by creating visualizations i.e. scatter plots, line charts to determine the relationship between variables, by describing the relationship between two quantitative variables, and computing and interpreting the Pearson correlation coefficient. | **2.0 Pts**  **Approaches Expectations**  The student inadequately performs bivariate analysis by creating visualizations i.e. scatter plots, line charts to determine the relationship between variables, by describing the relationship between two quantitative variables, and computing and interpreting the Pearson correlation coefficient. | **0.0 Pts**  **Needs Improvements**  The student does not show evidence of performing bivariate analysis. | | 4.0 pts |  |
| This criterion is linked to a learning outcome  Multivariate Analysis | |  |  |  |  | | --- | --- | --- | --- | | **4.0 Pts**  **Excellent**  The student exhaustively performs multivariate analysis by finding relationships between variables using multiple regression analysis and applying reduction techniques to their analysis. They go a step further to provide recommendations for effectively performing univariate analysis while working on the current problem. They go a step further to provide recommendations for effectively performing multivariate analysis while working on the current problem. | **3.0 Pts**  **Meets Expectations**  The student adequately performs multivariate analysis by finding relationships between variables using multiple regression analysis and applying reduction techniques to their analysis. | **2.0 Pts**  **Approaches Expectations**  The student inadequately performs multivariate analysis by finding relationships between variables using multiple regression analysis and applying reduction techniques to their analysis. | **0.0 Pts**  **Needs Improvement**  The student does not show evidence of performing multivariate analysis. | | 4.0 pts |  |
| This criterion is linked to a learning outcome  Exploratory Data Analysis and Documentation | |  |  |  |  | | --- | --- | --- | --- | | **4.0 Pts**  **Excellent**  The student exhaustively applies the exploratory data analysis approach though defining the question, the metric for success, the context, experimental design taken and the appropriateness of the available data to answer the given question. The students find and deal with outliers, anomalies, and missing data within the dataset. More on, challenge their solution by providing insights on how they can make improvements. | **3.0 Pts**  **Meets Expectations**  The student adequately applies the exploratory data analysis approach though defining the question, the metric for success, the context, experimental design taken and the appropriateness of the available data to answer the given question. The students find and deal with outliers, anomalies, and missing data within the dataset. | **2.0 Pts**  **Approaches Expectations**  The student inadequately applies the exploratory data analysis approach though defining the question, the metric for success, the context, experimental design taken and the appropriateness of the available data to answer the given question. The students find and deal with outliers, anomalies, and missing data within the dataset. | **0.0 Pts**  **Needs Improvement**  The student does not show evidence of applying the exploratory data analysis approach | | 4.0 pts |  |
| Total points: 16.0 | | | |