Rwanda Revenue Authority

Data Analytics Challenge

[2]

Facilitator Name

Kudakwashe Dandajena
University Name
African Leadership University

Group Members Details

Francis Walker
Wenseslaus Raphael
Jean Damascene Uwingeneye
Rachel Christelle Sarh Ayateke

0.1 Challenge Overview

The Rwanda Revenue Authority(RRA) is an institution that is mandated to collect tax revenue on behalf of the Republic of Rwanda. As a result of the prevailing COVID- 19 pandemic, its tax collection figures for the 2020-2021 fiscal year deteriorated by over 10% to a specific value of US1,3 Billion Dollars.

They are now want to improve their fiscal tax revenue for the year 2021-2022. In line with its digital agenda, the institution is planning to establish an effective revenue collection platform to influence its decision-making processes.

They gave out this data analytics challenge to groups that specialize in data analytics hoping to gain provide better insights on data that is already available.

0.2 Problem Overview

One of the goals RRA would like to accomplish with this challenge is to digitize their data on collected revenue. This is due to the deterioration of the tax figures by over 10% , the value of US 1.3 Billion Dollars. To achieve their goal, RRA needs to see what data can be used for and how it can help them draw better conclusions every year. For this challenge, we will need to find a dataset from an African country to demonstrate what the RRA can do with tax revenue data. Once the dataset has been found and cleaned, it will be used to demonstrate visualizations that could be potentially used and accomplished with data from Rwanda as well.

0.3 Team Details - Data Analysts

- Francis Walker

- Rachel Christelle Sarah

- Jean Demascene

- Wenseslaus Raphael

0.4 Environment Needed

To perform the analysis effectively we used the following resources

- Jupyter Notebook

- Anaconda packages

- Python (Latest Version)

- Different libraries example Matplotlib for visualization, NumPy, pandas, and IPython for display.

0.5 Dataset At-A-Glance

To give the RRA an idea of what could be done with the data that is collected we have to use two different datasets. The datasets will be cleaned and used to help solve our problem. Datasets are:

- South Africa Tax Revenue Dataset (1990-2018)

- Kenya County Funds Allocation information

The South African dataset provides total tax revenue and non-tax revenue about every year between 1990 to 2018. This will allow us to understand the different changes in tax revenue over the years. This dataset will be used to create different visualizations to draw insights and prepare a short recommendation to RRA.

Kenya County Funds dataset will be used to try different pandas, matplotlib, and other library functions to manipulate the data. This will allow us to demonstrate our data manipulation skills.

0.6 Solution

As described in the bid description the main objective of the bid is to simulate the scenario of the proposed solution or proof of concept (POC). So for full disclosure, the data and analysis in this solution section, do not reflect the ground covered by RRA or its affiliates. The solution designed here is the blueprint of what can be done to solve the problem mentioned above to help RRA implement its mandate effectively using data.

6.1 Solution breakdown

The solution is broken down into three main parts:

- **Data Ingestion from Different Sources:** Datasets used, were from different sources. The data collected by RRA come in different shapes and formats, so to depict the same scenario, we used different data sources in this POC.

  To implement the data ingestion for RRA, we will use data pipelines into all data sources in RRA existing systems to feed the datasets with the latest data. RRA collects a lot of data from an e-taxation system called Electronic Billing Machine(**EBM**). Using the data collected every day, the data pipelines will update the datasets, therefore creating real-time data analysis.

- **Data integration:** To gain the full picture from ingested data, there is a need for integration. The datasets need to be analyzed and brought together to build upon each other.

  To implement the data integration with RRA, as depicted in the Jupyter Notebook attached, the datasets will have a common virtual ground, where analytics tools such as matplotlib, Pandas, etc.. will be used to draw the collective analysis on the data.
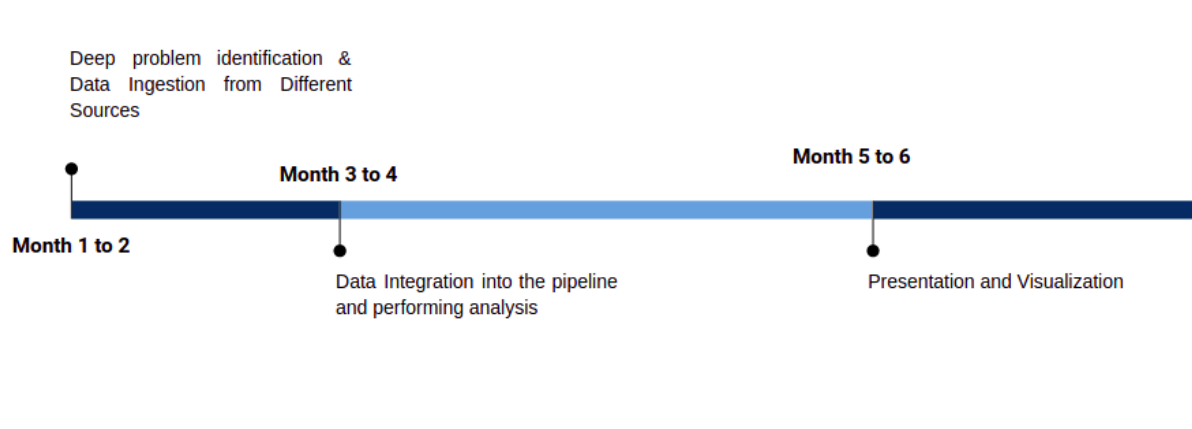
- **Presentation and visualization:** To finally make a sense of what the datasets are about, the analysis included data break down and visualizations to help to draw the meaning of the data in the presentation.

  Implementation of the visualization with RRA will be similar to the POC in the Jupyter Notebook provided. After creating the collective ground made of data pipelines from RRA systems, and after integrating then the last stage will be data presentation through mostly visualizations. Visualizations will be updated in real-time as more data from data pipelines are added to the existing datasets.

  At this stage, the main purpose of influencing the decision made by RRA should be fulfilled, to make informed **decisions based on data**. Effective timely data analysis and visualizations will provide the tools that are needed by the executives to make better decisions.

6. 2 Implementation timeline

To effectively implement the solution here is a timeline breakdown of how we will go about solving the problem. The whole implementation process will take half a quarter (6 weeks).

Deep problem identification & Data Ingestion from Different Sources

Month 5 to 6

Month 3 to 4

Month 1 to 2

Data Integration into the pipeline and performing analysis

Presentation and Visualization

1) Budget

- The budget for this project is an estimate of how much it will cost to implement the analysis solution. The charges are hourly ($ 10 per hour) and require about 40hours a week. This brings a total of $ **9,600.** In case of any additional expenses/work hours, we'll revisit the budget and adjust accordingly.

2) Risk

With this solution, taxation being implemented there is a high risk that consumers' spending might go down. This is because there will be an added cost to products which generally increases the price of products and services in the market.

This can have an impact in other areas like raw material production which will cause scarcity in production hence price fluctuations in the market. It will also affect the number of entrepreneurs due to the fear of starting a business. All this will hinder a business from growing and hence fewer tax is collected.

References

1 "Details of Public Revenues - South Africa", Stats.oecd.org, 2021. [Online]. Available: https://stats.oecd.org/Index.aspx?DataSetCode=REVZAF. [Accessed: 07- Mar- 2021]

1) [2]"How Data, Analysis, and Reports Can Improve Customer Service | CommBox (BumpYard)", *CommBox (BumpYard)*, 2021. [Online]. Available: https://www.commbox.io/how-data-analysis-and-reports-can-improve-customer-service/. [Accessed: 07- Mar- 2021]

2) Dandajena, K., 2021. *AI Project 2- Data Analytics Assignment.pdf.* [online] Google Docs. Available at: <https://drive.google.com/file/d/1tcd-xKSKxwZoCzpfqWWLXOA1QXMj4met/view> [Accessed 7 March 2021].

3) Dataoneorg.github.io. 2021. [online] Available at: <https://dataoneorg.github.io/Education/lessons/09_analysis/09_a: [Accessed 7 March 2021].