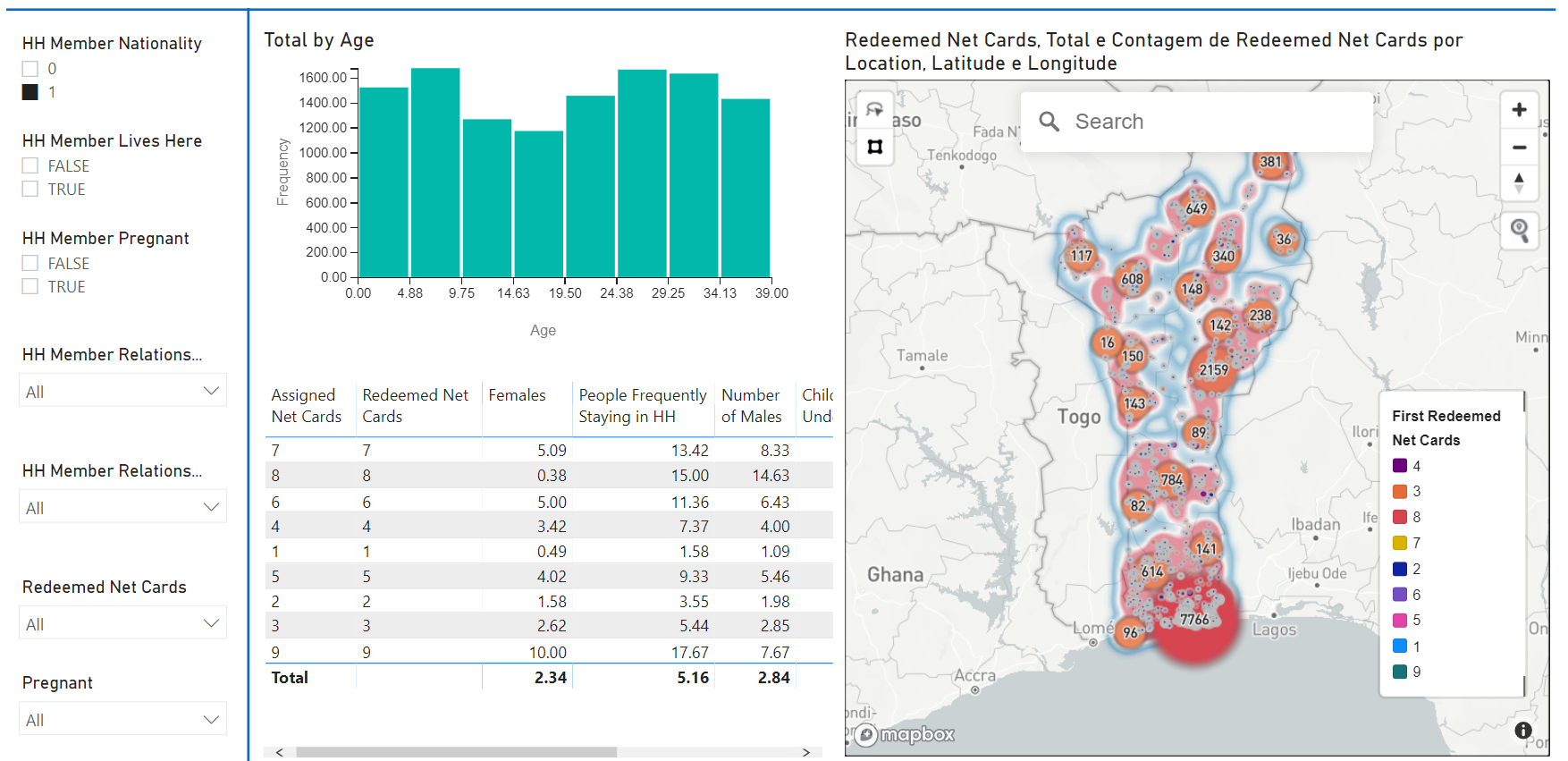
Implementation of the Who Model we missed



**Redeemed Net Cards, Redeemed Net Cards Total and Count by Location, Latitude and Longitude.**

**Definição do Problema:** In 2020, CRS provided technical and management support to and worked closely with Benin’s National Malaria Control Program (PNLP) and National Coordination Bureau (INC) to distribute 7.6 million ITNs to 13.6 million people in 3 million households. We achieved an ITN redemption rate of 94%.

To understand what most impacted the bed net redemption rate we ran the Who Did We Miss? model. Using the insights gained, we were able to make the following recommendations to CRS programming and MEAL staff to consider for future campaigns:

•National boarders, highways, and parks had higher miss rates.  Adopt a diversified strategy that changes based on the characteristics of each particular location. E.g., distribute more nets in locations on borders and close to main highways.

•Families that receive 2-4 bed nets represented 80% of the program.  Reduce dropout in these groups to optimize the program and gain scale.

•Families that receive only 1 bed net had the highest number of non-redemptions.  Develop a strategy to improve redemption rate by those who only receive 1 bed net.

**Database: The**  data used comes from the Redhose system in CSV format. The information from questionnaires was grouped into text files for manipulation.

**Data Analysis:** The data will be analyzed in powerbi for the generation of graphs and information to characteristicize the strategic problems that should be solved.

**Data preprocessing:** Identification of missing data and correction by column median, when the number of missing values is more than 30% of the total column records. If the total missing values are less than 5% of the column, it is suggested that the data be deleted. Correction of outliers by the median of the column.

**Model Construction:**  The model will be built on unsupervised learning using the cluster method to identify groups and their characteristics. To predict the number of mosquito nets, the supervised learning model of random forest will be used.

**Model Deploy:**  The model can be used in powerbi data visualization tools or Flask web application tools.

**Metric You**  want: Accuracy greater than 80%.

**Necessary tools and libraries:** Python, Sklearn, Pycaret, Jupyter notebook, powerbi, numpy, pandas.

**Repository:** https://ml.azure.com/?tid=b80c308c-d08d-4b07-915c-11a92d9cc6bd&wsid=/subscriptions/c8edbbc9-5923-493c-a5da-dc1e04d1d0d5/resourcegroups/CRS\_mira\_goldcopyDev\_ml/workspaces/ml\_benin

Creation of the Model

1. Installation of non-native libraries:

Graphical user interface, Text, Application, Word

Description generated automatically

1. Loading libraries:

Text

Description generated automatically

1. Access to the cloud environment of data repositories:

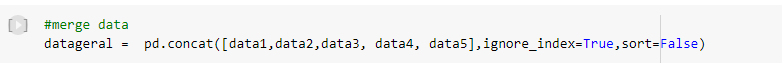
Text

Description automatically generated with low confidence

1. Data import and dataframe conversion:



1. Union of text files into a single dataframe:



1. Dataframe inspection:

Graphical user interface, Text

Description generated automatically

Graphical user interface, Text, Application

Description generated automatically

An image containing Table

Description generated automatically

1. Identification of missing values through chart:

Graph, Bar Chart

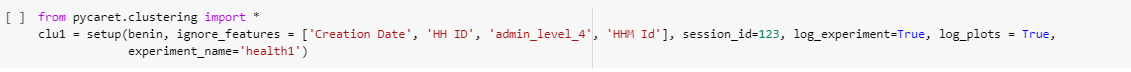
Description generated automatically

1. Correction of missing values through median and imputation per column:

Text

Description generated automatically

1. Import of libraries for Kmeans generation with Pycarert:



1. Variables/Columns used in cluster model construction:

Text

Description automatically generated with low confidence

1. Cluster models that best fit the data model offered by pycaret:

Graphical user interface, Table

Description automatically generated with average trust

1. Definition of the number of clusters and result of accuracy, with measurement of accuracy the silhouette score: 81%.

Graphical user interface, Text, Application, Email

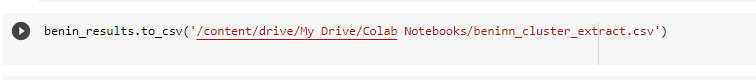
Description generated automatically

1. Dataframe view with cluster IDENTIFICATION:

Computer screen with black text on white background

Description generated automatically

1. Export of dataset



1. Accuracy of clusters:

Graph, Bar Chart

Description generated automatically