# Data description

You are allowed to use **only** the datasets that are provided here and any features extracted from the contextual layers data accessed from the Africa GeoPortal provided below.

**Main data set**

The main dataset contains demographic information as well as information about the kinds of financial services that are used by approximately 10,000 individuals across Tanzania. The main data set has been split into two .csv files, one containing the independent features and the other the target variable, called data.csv and data\_target.csv, respectively. A description of all the features contained in these datasets may be found in the file titled *Variable\_Codebook*

Your goal concerning the prediction task is to accurately classify each individual into four mutually exclusive categories:

1. **No\_financial\_services**: Individuals who do not use mobile money, do not save, do not have credit, and do not have insurance
2. **Other\_only**: Individuals who do not use mobile money, but do use at least one of the other financial services (savings, credit, insurance)
3. **Mm\_only**: Individuals who use mobile money only
4. **Mm\_plus**: Individuals who use mobile money and also use at least one of the other financial services (savings, credit, insurance)

**Financial\_Access\_Map**

This dataset is the geospatial mapping of all cash outlets in Tanzania in 2012. Cash outlets in this case included commercial banks, community banks, ATMs, microfinance institutions, mobile money agents, bus stations and post offices. This data was collected by FSDT. More about this dataset [here](http://www.fsdt.or.tz//wp-content/uploads/2015/12/GIS-Map-FinAccess-1.pdf)

**Prediction data set**

To test the performance of your models, you are requested to submit a file, in .csv format containing your predictions for the observations contained in the file titled pred\_data\_set.csv to ryan@capetownai.com.