**Week 12 Individual Project**

### **1.BUSINESS UNDERSTANDING.**

#### **Problem Statement**

A Kenyan entrepreneur has created an online cryptography course and would want to advertise it on her blog. She currently targets audiences originating from various countries. In the past, she ran ads to advertise a related course on the same blog and collected data in the process.

She would now like to employ my services as a Data Science Consultant to help her identify which individuals are most likely to click on her ads.

## **Main Objective**

To identify the individuals who are most likely to click on the ads.

## **Specific Objectives**

1. To determine how Age affects ads clicks
2. To determine how daily internet usage affects ads click
3. To determine the correlation between various factors
4. To see how gender affects the rate in which people click on ads

#### **Business Success Criteria**

We will try to determine the factors that individuals who click on ads have that greatly affect their decision. This will be visualised by various scatter plots and graphs that we will plot

**Tools used in this primary project**

We used several tools in our project for various purposes:

* R studios
* Github
* Google slides

### **2. DATA UNDERSTANDING.**

#### **Data Understanding Overview.**

Our dataset was obtained from the Moringa course outline

Our dataset contains the following columns:

* Age :The age of the individuals
* Area income: Income of the area
* Daily internet usage : Usage of the internet
* Male: Gender of the people
* City:City of the person
* Ad topic line: Topic of ad
* country: Country of the person
* Timestamp: date and time
* Clicked on ad: Whether a person clicked on an ad

### **3**. **DATA PREPARATION.**

The dataset has 1,001 rows and 10 columns. I started by finding the unique value counts in columns of interest, just to get a feel of our data. Most of the data was object.

Next I found out that there were no missing values or duplicates within our dataset.

There were no outliers within our dataset therefore we decided to not drop any rows or columns. This is because we wanted to work with a complete dataset and the outliers would be good for comparison.

#### **Findings from EDA**

* There is a positive correlation between daily time spent and daily internet usage
* There is no correlation between age and area income
* There is no correlation between Age and duration of time spent daily on the internet.
* People from all ages end up clicking on the ads. We can also see that people who click on ads are also from the age of 50-60.
* Gender isn’t a factor that affects whether people click on ads or not.
* People who click on ads are the ones who spend most time daily on the internet.

**Conclusions**

From the graphs above we can conclude that majority of the people who will click on the ads are the ones who spend more time on the internet, are from all ages including the age from 50-60 and use more internet usage since it has a positive correlation with daily internet time.

### **5. RECOMMENDATIONS**

The company should target people from all ages mostly from the age of 50-60 maybe by advertising something close to what that age group may like. The company should focus less on gender and more on people who spend more daily internet time, maybe by giving free internet usage on clicking the ads.