Detailed Report

# **1. Sampling and bias**

As a team, we analyzed the dataset for sampling. The data set has some uncertain data like having 98 children and the education level like 99. In the preprocessing step, we removed the uncertain data. And the given data has the age group 30 to 50. In the population, kids to adults included. Here the data set given has the age range of 30 to 50. So we make our hypothesis within that age range only.

# **2. Hypothesis Test**

Earning increase with the education level of people who are in the age group of 30 to 50.

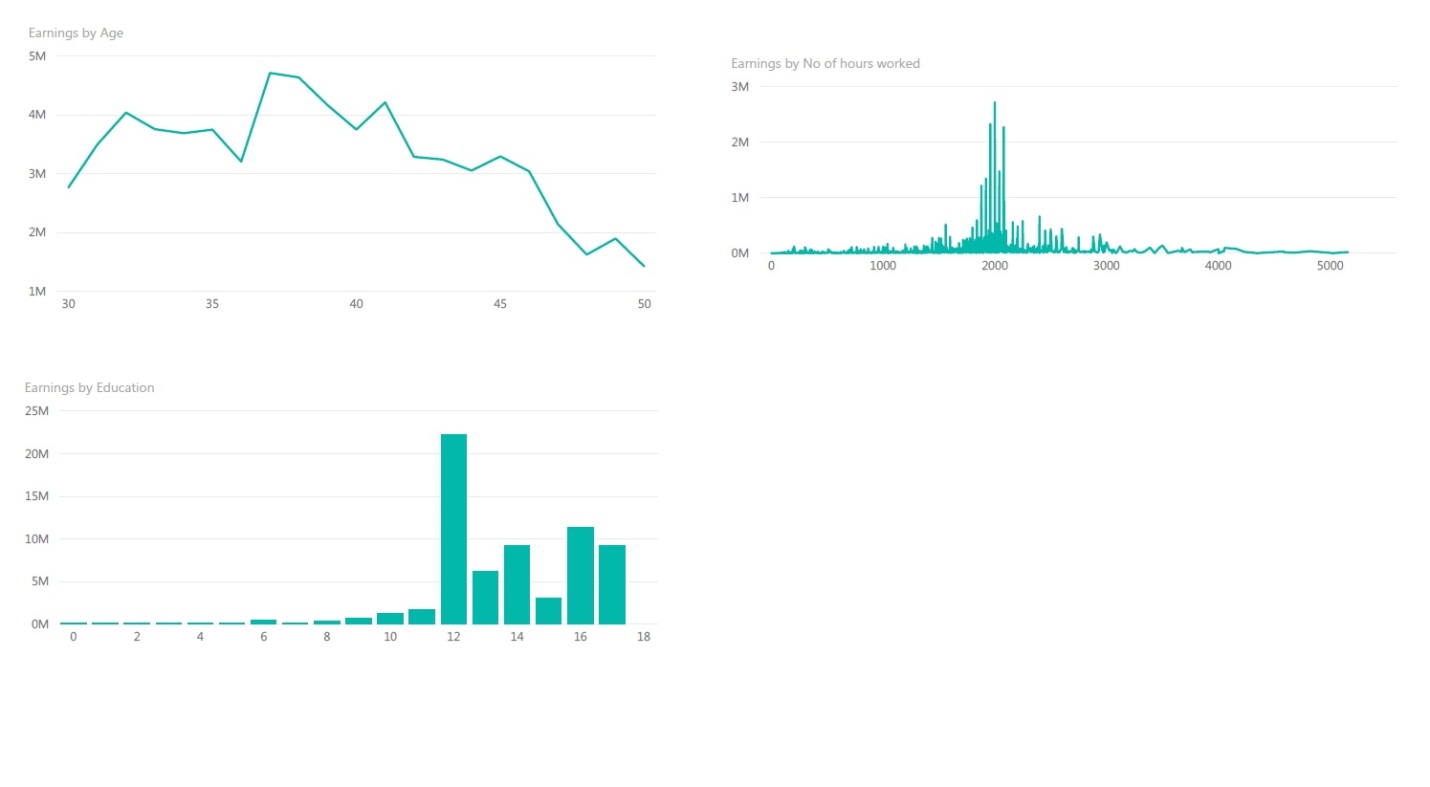
# **3. Analyzing and visualizing**

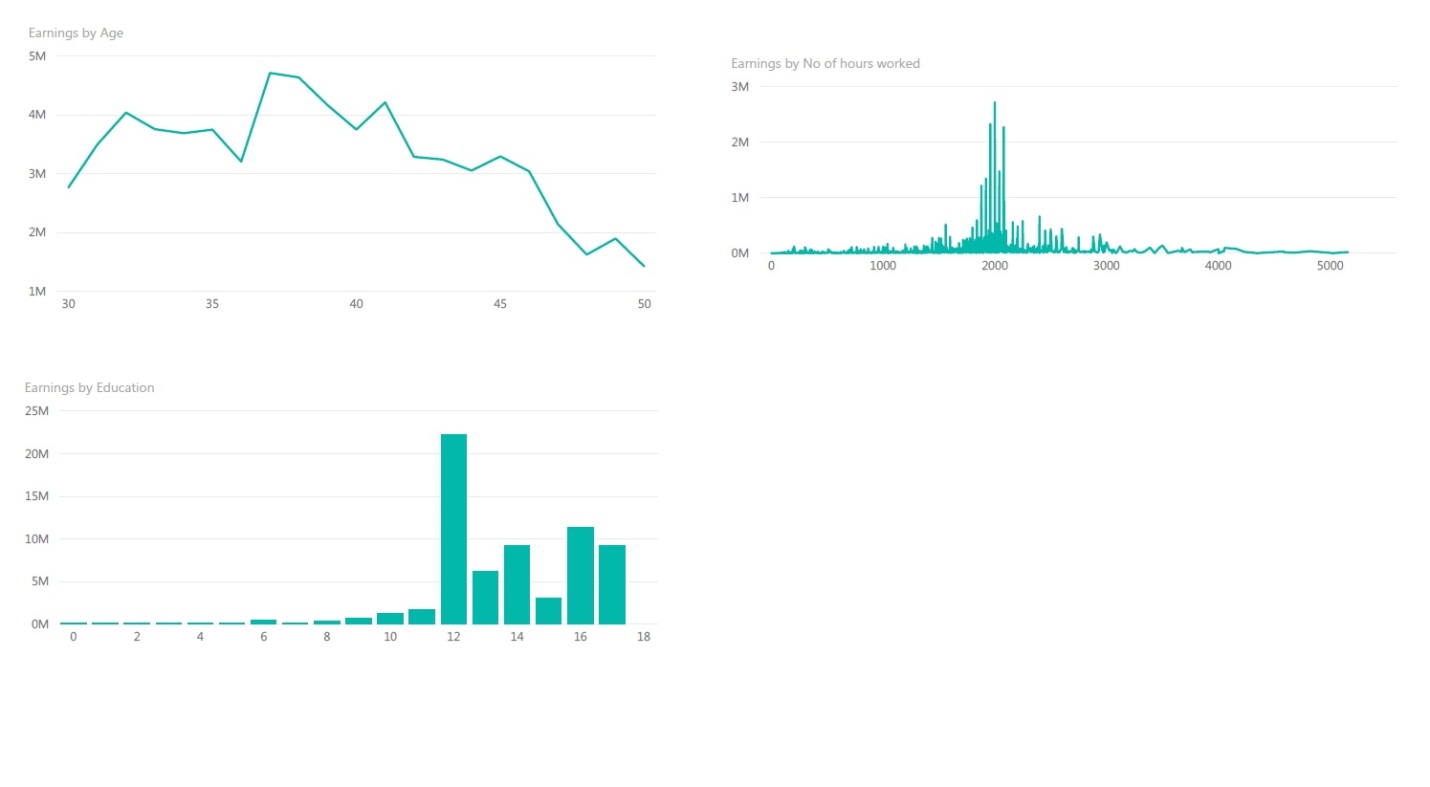
1. Source Code

The link to the source code is given below.

<https://github.com/UpekkaA/datascience-assignment1/blob/master/source-code/visualisation.py>

1. Results





# **4. Other useful analysis.**