**1 Page summary**

**Team & their responsibilities**

Heshan – Investigating about the biasness of the data set. Perform Statistical Tests for Analysis 1 -4.

Shanaka – Perform Statistical Tests for Analysis 5 -6

Nicum – Perform correlation tests for Work vs Earnings, Perform correlation tests for Educaion Level

vs. Earnings

Navoda – Perform correlation tests for Age vs Earnings

**Git Repo Link -** <https://github.com/shanakaChathu/DataScienceAssignment1>

**Hypothesis / Questions**

* People with age over 40 years earnings are different than the people with age less than 40 years earnings.
* People with age over 40 years working hours are different than the people with age less than 40 years working hours.
* Earnings of people with high educational levels are different than earnings of the people with low educational levels
* Working hours of the people with high educational levels are different than the working hours of the people with low educational levels
* There is a difference between earnings of married and other people
* There is a difference between average earnings of people who are having 2 or few kids with respect to the people who are having more than 2 kids.
* Earnings of a person depends with the age of the person
* Earnings of a person depends on how many hours they work
* Earnings of a person depends on the education level of the person

**Assumptions**

* NA, 98, 99 are codes to identify missing values. Because of that we excluded records with NA, 98, 99 from our analysis.
* Earnings and Hours contains zeros. We assumed these people might not possible to be the Unemployed. These zeros are there because of those individuals didn’t respond to that question in the survey.