**Summary**

**Team and Contribution**

**Anojan Kaneshathas** :

**A.M.H.D Karunathilaka** : Correlations, Summary Report, Hypothesis 1, Insights

**S.A.C Fernando** : Basic visualization, Cleaning data, insights on sampling and bias, Hypothesis 2

**Git Repo Link**

<https://github.com/HashanAbeysinghe/data-science-assignment-1.git>

**Hypothesis / Questions**

1. We explored whether there is a relationship between education and earnings.

H0 : There is no correlation (c =0)

Ha : There is a correlation (c > 0)

2. We explored whether married people earn more than others.

H0: Mean of married people and others is equal or less (*μM <= μO*)

Ha: Mean of married people is greater than others (*μM > μO*)

3. We explored whether there is a relationship between work hours and earnings .

H0 : There is no correlation (c =0)

Ha : There is a correlation (c > 0)

**Assumptions**

1. Data set is taken in a completely random manner.
2. Data set reflects population.
3. Number of kind above 15 are outliers and removed when cleaning.
4. Education level above 30 removed as outliers.
5. Education level below 10 removed.

**References**

1. <https://pandas.pydata.org/pandas-docs>
2. <https://towardsdatascience.com/ways-to-detect-and-remove-the-outliers-404d16608dba>
3. <https://medium.com/we-are-orb/introduction-to-data-visualization-with-pandas-21709985ff67>
4. <https://machinelearningmastery.com/statistical-hypothesis-tests-in-python-cheat-sheet/>
5. <https://www.texasgateway.org/resource/124-testing-significance-correlation-coefficient-optional>