



Data Science Assignment 01

Executive Summary

Group 09 (DS_2019_007)

MSc in Computer Science 2019

Department of Computer Science and Engineering

University of Moratuwa

Executive Summary

After analysing the PSID data which is based on families and individuals in United States, we have found out below details.

Specially during the hypothesis testing, where we considered earnings and education level in which we found out, when education rate is becoming higher, the earnings are also increased.

Also we have observed following key factors during the analysis:

- Hours and earnings have a strong positive correlation inferring higher hours spent allow higher earnings
- Education and earnings have a significant positive correlation inferring higher education levels allow higher earnings
- Education and kids have significant negative correlation inferring residents with relatively higher education levels tend to have less children
- Earnings and kids have a significant negative correlation inferring families with relatively high income tend to raise less children

We used T-test as the method of hypothesis testing.