Group 5 : Project Proposal

Research Question: Does looks of a person have any significant impact on wage?

Data Source :
We will be using the dataset 'beauty' available in the Wooldrige package.

Link: http://hec-cloud34.segi.ulg.ac.be/~moodle/mod/folder/view.php?id=8086&lang=en

Data Description:
The dataset contains 1260 records with 17 independent variables. It includes various variables related to individuals' employment and demographic characteristics. The "wage" variable represents the hourly wage of an individual, while "lwage" is the natural log of the hourly wage, "belavg" is a binary variable indicating whether an individual's perceived looks are rated as 2 or lower. Conversely, "abvavg" takes the value of 1 if looks are rated as 4 or higher, "exper" represents the years of workforce experience, and "looks" is an ordinal variable from 1 to 5, indicating appearance. Binary variables include "union" for union membership, "goodhlth" for good health, "black" for Black ethnicity, "female" for gender, "married" for marital status, "south" for residence in the South, "bigcity" for residence in a big city, and "smllcity" for residence in a small city. "service" is a binary variable for employment in the service industry. "expersq" is the squared value of workforce experience, and "educ" represents the years of schooling completed by the individual.

Motivation:
The motivation behind choosing this topic was that all the team members are nearing graduation, and we will be looking for jobs as our next big step in our professional careers. Our interest is to see that apart from education, what are the other factors that have a significant impact on wage, specifically looks.

Project idea:
To apply the knowledge gained in this class, we'll use multiple linear regression with the OLS estimation technique to identify significant parameters in predicting employee wages. We aim to run regression models with wage and log(wage) as dependent variables, incorporating selected independent variables and their interaction effects. Following the regressions, we will conduct hypothesis testing to study the join significance between the variables and compute R-squared, Adjusted R-squared and Mean Squared Error (MSE) to pinpoint and interpret the coefficients of the significant variables.

Team Members:

1. Sai Madhav Reddy Nomula - SXN210168
2. Shreyas Nellore - SXN210092
3. Dhatri Chunchu - DXC220013
4. Almitha Veigas - AXV220022
5. Jeevithesh Reddy Narravula Reddy - JXN210032
6. Rithwik Reddy Koripelly - RXK210087
7. Pidaparthi Murthy - PSM220000
8. Tej Sai Pranav Reddy Kagitala - TXK220023