CSCE 5320 SCIENTIFIC DATA VISUALIZATION
PROJECT PROPOSAL

Description:

Team Members:

Bibek Lamsal
Mo Omar
Jordan Wolfe

Title:

Analysis of Data Science Salaries based on Socioeconomic Factors of a Country

Introduction to Domain:

We will be focusing on socioeconomic.

Goals and Objectives:

Motivation:

This project will help us to understand if there is any correlation between the economic and social factors of a country in relation to the salary of their data science professionals. This will help companies and professionals to assess the market value of data science professionals to ensure that they are paid a competitive salary amount when factoring in socioeconomic factors for their country. This takes into socioeconomic domain because there are certain social factors that can determine the market value of data science professionals, which is related to finance and the overall economy is a factor into the market value as well.

Significance:

There is a pressing need to capitalize on assessing the true market value and potential salaries of data science professionals to provide transparency of these figures. Transparency of these figures allows for both companies to provide competitive salaries to their data science professionals to ensure high retention rates of their employees and for data science professionals to know what their market salary is based on their location, other locations if they thought of moving, and to guarantee that they are not paid far below the median salary for their position in their country. This project will leverage Tableau, D3.js, and Python to create advanced scientific visualizations to convey our findings. Improvements of socioeconomic factors like education and wages based on the output of the visualization may lead us to finding improvements in the professional level of data science occupation of that country.

Objectives:

Our main objective would be undergoing correlation analysis between the different datasets to assess the overall objective of finding the significance of socioeconomic factors having an effect on the market salary value of data science professionals. To start, we would need to filter down noisy data that is far outside of most of the data points to prevent the noise from effecting our results. From there we would have to compare different parameters using different visualizations to get an idea of things such as the overall distribution and trends that preside inside of the dataset. After we are able to pick up on a few patterns, cycles, or trends we would then take that trend and focus our visuals on highlighting that trend while also forecasting future values if possible. After all is done, we will then evaluate all the visualizations and assess whether or not there is a strong enough correlation between socioeconomic factors and the salaries of data science professionals to a set confidence level.

Features:
We will use the Data Science Salaries dataset and the Global Country Information Datasets for our analysis. We will utilize Tableau, D3.js, seaborn, and matplotlib for generating visualizations. For seaborn and matplotlib we will utilize these libraries to perform time series analysis and for D3.js and Tableau we will focus on other visualization types such as bar graphs, geo graphs, and area charts to create dashboards based on the datasets.

According to the National Center for Education Statistics, “The countries with the two lowest percentages of education expenditures for all education institutions [are Luxembourg and Ireland]” for 2019. And also the “The countries with the two highest percentages of education expenditures for all education institutions [are Norway and Chile]” for 2019. For our research, we would like to find the countries with the highest and lowest percentages of education expenditures for all education institutions for 2023.

References:

Elgiriyewithana, Nidula. “Global Country Information Dataset 2023.” *Kaggle*, 8 July 2023, www.kaggle.com/datasets/nelgiriyewithana/countries-of-the-world-2023.

Islam, Sazidul. “Data Science Salaries 2024.” *Kaggle*, 20 Jan. 2024, www.kaggle.com/datasets/sazidthe1/data-science-salaries.

Te_Jota, tejota. “Government Expenditure on Education.” *Kaggle*, 17 Nov. 2023, www.kaggle.com/datasets/tejota/government-expenditure-on-education-world-bank.

National Center for Education Statistics. “Education Expenditures by Country.” *Condition of Education. U.S. Department of Education, Institute of Education Sciences.*, Aug. 2023, nces.ed.gov/programs/coe/indicator/cmd.