Jolie Ng

EDA LAB

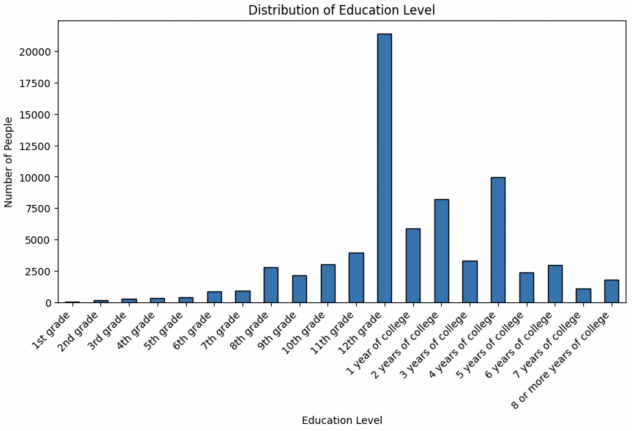
1. Download a small (5-15) set of variables of interest. Done in notebook.
2. Write a short description of the data you chose, and why. (1 page)

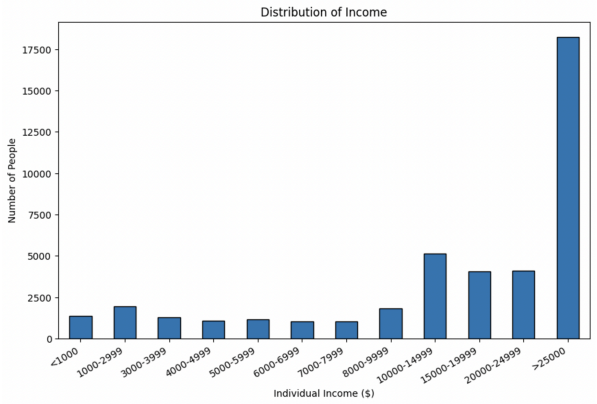
I chose to analyze data on **income (both family income and individual income), age, race, religion, political stance, and education level**, because there are many stereotypes surrounding these variables that make us biased. For example, we might unintentionally make assumptions about someone's income based on their race, religion, or education level. It's a common assumption that people of certain races, particularly Black or Hispanic individuals, earn less money than White individuals, even though income disparities are obviously influenced by many factors beyond just race. I'm interested in seeing just how true this stereotype actually is and if other factors also play a role in determining income.

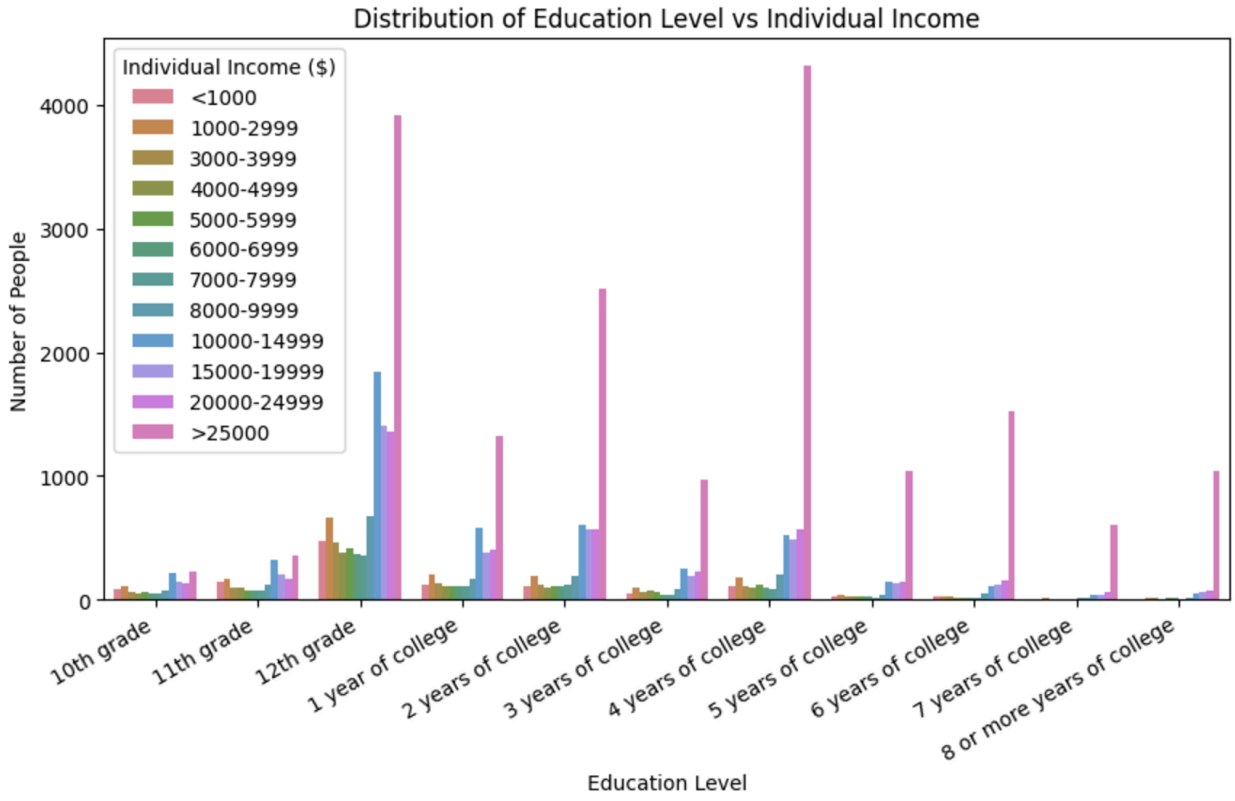
The reason I focused mainly on education level and income in my data extraction and analysis is because this is something I hear a lot of talk about in my life. People around me always say things like, "You need to get a degree to get a good job," or "Without a degree, you won't make good money." There's a lot of pressure to pursue higher education because people believe it directly leads to higher-paying jobs. I wanted to see if this is true, or if there are other factors at play. People are often judged based on their background, but, for example, a really smart person might not have had the opportunity to get a higher education degree, but they can definitely fulfil the requirements or knowledge needed for a high-paying job. Someone without a degree can also start a business and make much more money than those with corporate jobs.

3. Load the data using Pandas. Clean them up for EDA. Do this in a notebook with comments or markdown chunks explaining your choices. Done in notebook.

4. Produce some numeric summaries and visualizations. (1-3 pages). Done in notebook, but here are pictures of the visualizations I made for reference:







5. Describe your findings in 1-2 pages.

From the visualizations I created, we can see that in general, there are significantly more people in this dataset with a high income (more than $25000). This is another issue we have to consider, since the people who have time to fill in a survey are probably the ones who don't have two back-to-back minimum wage jobs and no time on their hands. From looking at the Education Level vs. Individual Income graph, we can see that the data does agree with the fact that a more advanced education level is correlated with higher income (proportionately, the percentage of people who have a high income compared to lower incomes increases by education level). However, after four years of college, that proportion seems to stay stagnant i.e. more years of college after obtaining a bachelor's degree does not seem to affect income significantly, if at all.

I think this makes a lot of sense, and aligns with my perspective that in college and beyond, you have to be self-motivated to learn. A lot of material has to be self-taught or explored by yourself since there aren't strict academic curriculums the way there is in elementary, middle, and high school. People with just a bachelor's degree can learn the information needed for a high-paying job without getting a master's or even higher degree, and might be able to show more technical skills during the job interview than someone who has a higher degree. I think this shows that most jobs think it is important for applicants to at least have a bachelor's degree to confirm their academic foundation. Beyond that, it is up to each person to decide how much they want to learn and what skills they can provide to the jobs they want.