EXECUTIVE REPORT

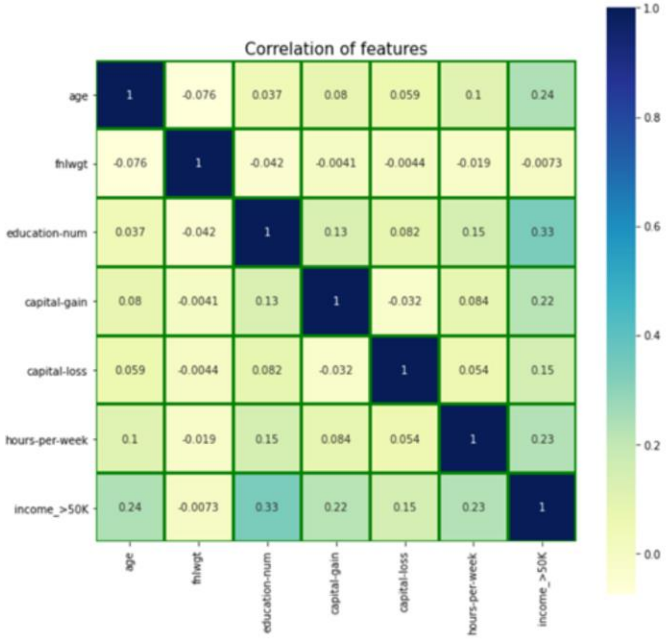
ABHINAND KANAKAN
AMEER AHMAD
PRANAV TOGGI
WOOSUNG YOON

OVERVIEW

Attributes that were identified to impact the income of an individual:

- Education
- Age
- Occupation
- Relationship

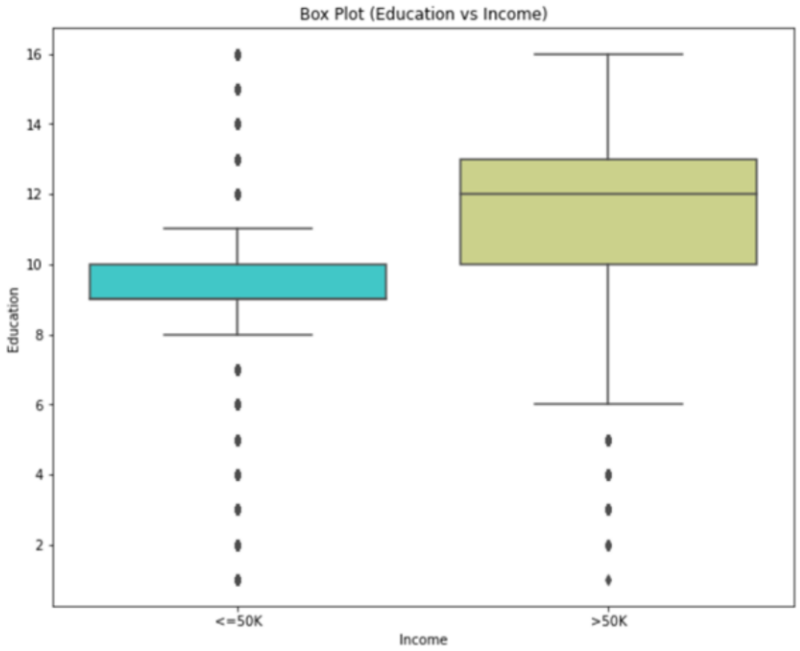
Correlation of Numerical features and Income



This visualization is a Heatmap showing the magnitude of the correlation of features by employing color as the visual variable. It gives a quick overview of how the different numerical attributes are related to income. The last row in the Heatmap indicates that features such as education, age, etc., are related to the income bracket of greater than 50k since it has positive values. This can be easily identified because of the darker shades for these cells, indicating a higher value than zero. This quick overview helps in understanding that the previously mentioned attributes might be a determinant of an individual's salary. Therefore we can proceed to explore and analyze these attributes in detail.

Slide credits: Abhinand

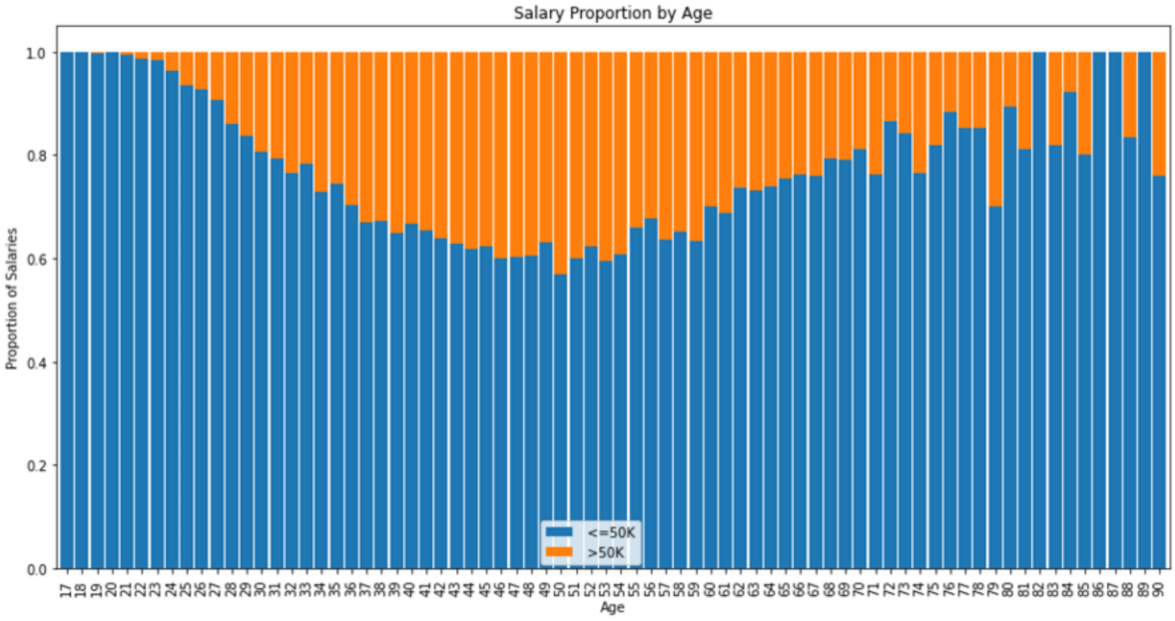
Education



Here the education number is a numerical value that directly corresponds to different education levels in individuals. Boxplots are plotted side by side for both the salary brackets of interest. By making a quick comparison, the importance of education as a predictor of income is communicated with the help of this visualization. The interquartile range for salary less than 50k lies between categories 9 and 10, while for salary greater than 50k, it lies between 10 and 13. Therefore, the median value of education for high-income individuals is outside the inter-quartile range of low-income individuals. This is a clear indication that an individual's education level plays a crucial role in deciding income.

Slide credits: Ameer

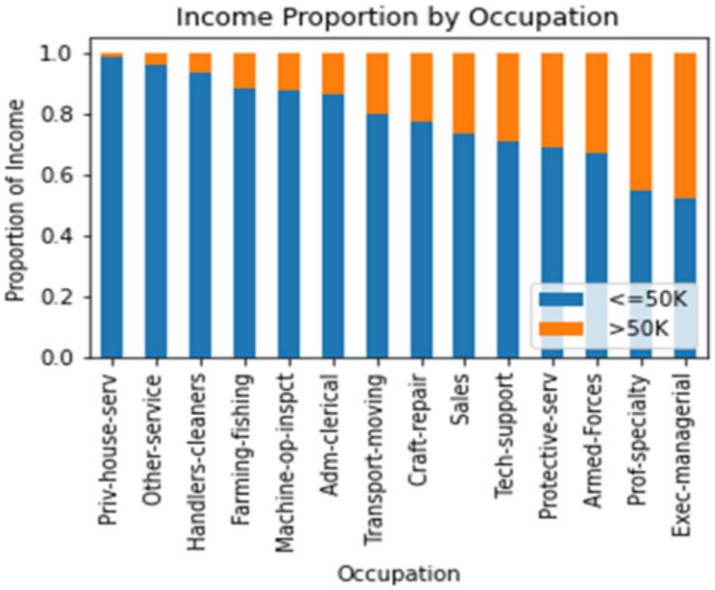
Age



The age attribute, as expected has a moderate influence on the income of the individual. From the data, age 50 gives the highest probability of income greater that $50k. And ages 21 and below showing the lowest. Apart from the apparent trend observed in the above visualization, its standard deviation of the proportion of income greater than $50k is 0.132. If there was no deviation in the proportions in the visualization, it implies that the age attribute has no influence on the income. Thus in this case, age is definitely a good candidate attribute to be considered.

Slide credits: Pranav

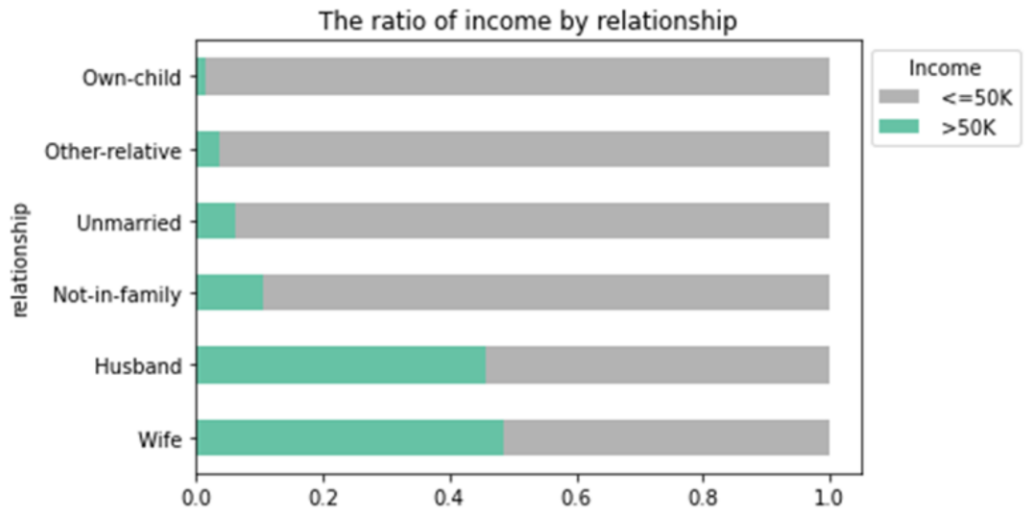
Occupation



This visualization represents the ratio of the population by occupation. There are a total of 14 occupational groups in the data. In all occupations, the probability of earnings below 50k is greater than the probability of earning above 50k. “Priv-house-serv” has the lowest probability of getting a salary greater than $50k. The probability of “Exec-managerial” and “Prof-speciality” jobs making an income greater than $50k is relatively high. Hence, occupation is an essential attribute in determining the income.

Slide credits: Woosung

Relationship



This visualization shows the percentage of income population by the relationship status. The population ratio is expressed with the help of a stacked bar chart sorted by size, making it easy to compare it with relationship categoreies in other rows. If we take a look at this visualization, we can observe that the husband and wife entries have a high-income percentage. On the other hand, if an individual was not married or had other relationship status, then an income below $50k was more likely. Therefore relationship status of person is a strong indicator of income level.

Slide credits: Woosung