

CSE 578 - Data Visualization

Executive Report Submission

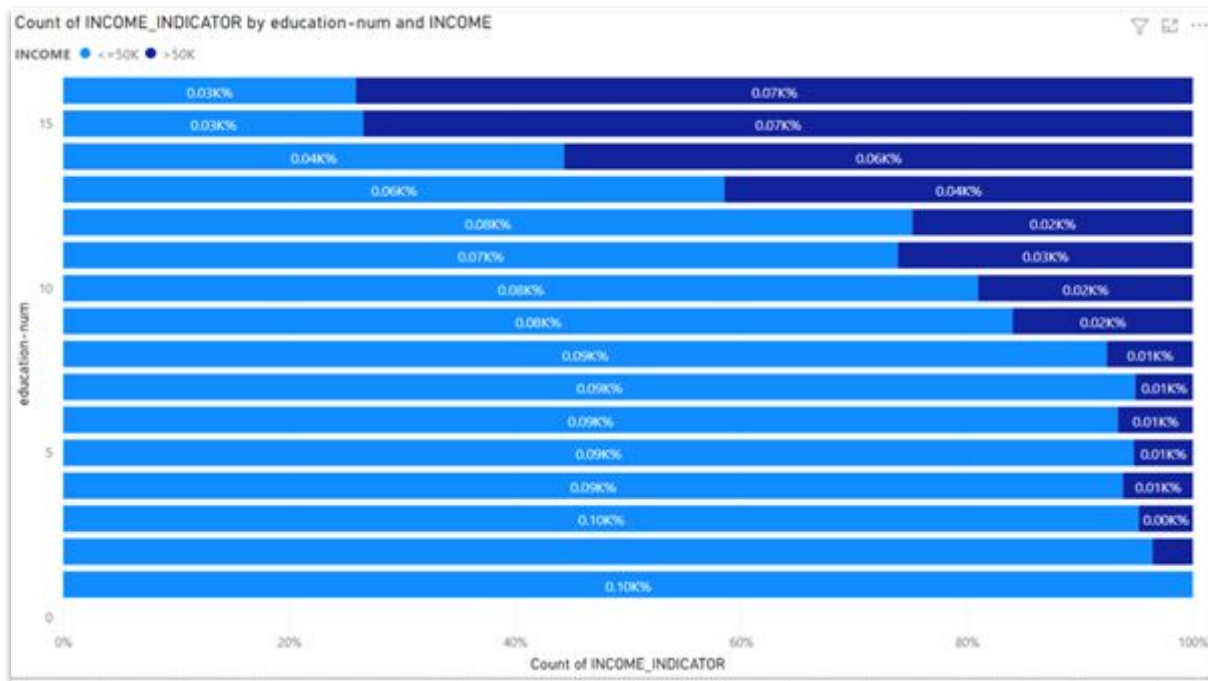


Project Statement

UVW college provides a dataset to be analyzed to check which attributes affect the salary ($>50k$, or $\leq 50k$). We as XYZ's analysis team, analyzed the data and based on our results we provide this report.

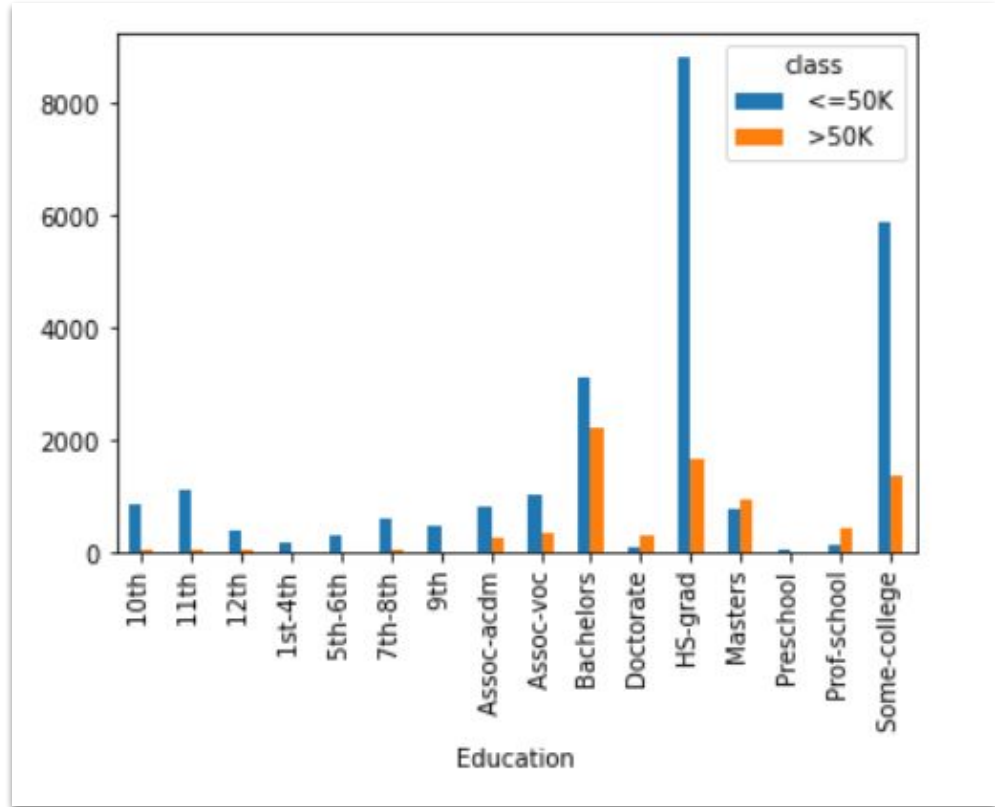
1- Education-num: We want to figure out if the education-num of an individual affects their income label (>50k, or <=50k)

The stacked bar chart shows that education number is an important attribute in determining the income of an individual. It shows that as education number increases there is an increase in those that are earning above 50k and a significant decrease in those that are earning below 50k. If an individual has a lower education number there is a high chance that they will earn less than 50k but for those with high education num they will have an increased chance of earning above 50k.

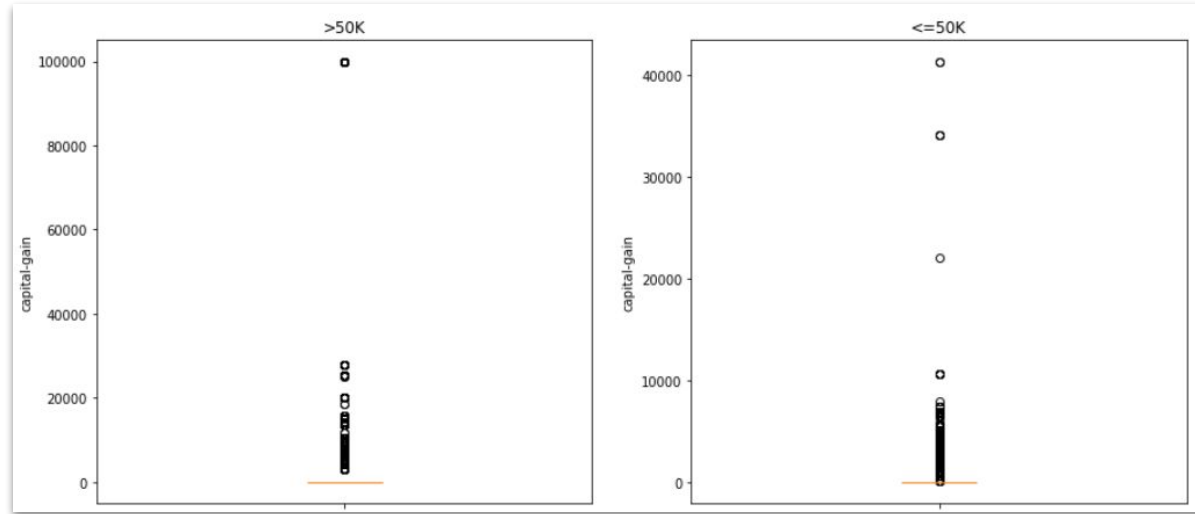


2- Education: Prof-School, Masters and Doctorate are more likely to gain salary >50k

From the unstacked bar chart in the right, which shows the classes for individuals (>50k, < 50k) in terms of education attribute. We can see that the options (Prof-School, Masters and Doctorate) are more likely to gain salary above 50k with probability of more than 60%, especially for the Prof-School individuals, which reaches around 80%. However, all the other options are more likely to gain a salary of <50k.

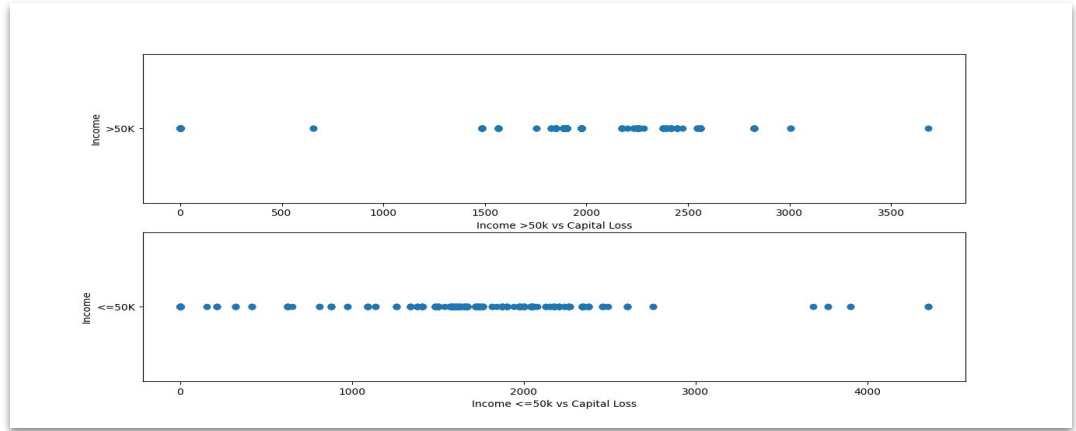
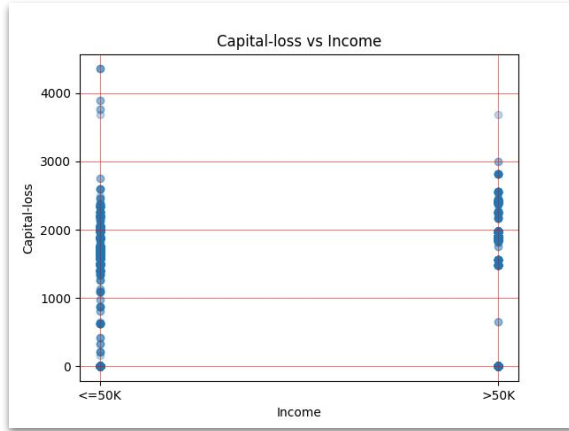


3- Capital Gain: >10k capital gains indicate very high probability of >50k salary



The boxplots show the distribution of capital gains for the group with salary >50k and the group with salary <=50k. With the median of both groups being 0, this feature is not a useful salary group indicator for more than half of the individuals with no or low capital gains. However, as the majority in the salary <=50k group have capital gains less than 10k (more than 99.9%), it can be concluded that if an individual has capital gains more than 10k, he or she is almost certainly in the salary >50k group.

4- We want to figure out if the capital-loss of an individual affects their income label (>50k, or <=50k)



The chart provides us with an insights of capital loss against income. We can see capital range where they have both label. However, there are another range goes to one of them. For example, range from 2800:3800 goes to >50k class. For 1:1400 goes to <=50k. So, if an individual's capital loss is above 3000 he is absolutely getting >50k.

Conclusion

Summary:

Based on our analysis, these are the top four classifying attributes that affect the income label.

Visualization Roles:

Goabaone: education-num, age and marital status

Bingxi: capital-gain, relationship and workclass

Islam: capital-loss, occupation, country-native and sex.

Mohamad: education, race, hours-per-week and workclass

Ahmed: hours-per-week, race and fnlwgt