Individual Contributions and Learnings

Ajay Kannan

Reflections:

The dataset provides us with the household income details which was obtained from the U.S Census. We were tasked with providing user stories based on the individual's salary feature that would help the companies make business decisions. Through the learnings with the course work, though challenging, I was capable of representing the visualisations in a marketable fashion.

I cleaned the data before working on it and I disregarded some data which was missing. Then, we divided ourselves in to various sets of the features in the project. I took the Marital status, Sex, Occupation, education number and Salary.

I started playing the variables, not based on salary at first. I created common plots like bar chart, pie chart, scatter plots etc. This helped me understand the distributions of the data within the sub-categories. Then I began to interleave thr sub-categories. I found **Mosaic plot** the most useful because the mosaic plot gives a larger area if has a unique distribution.

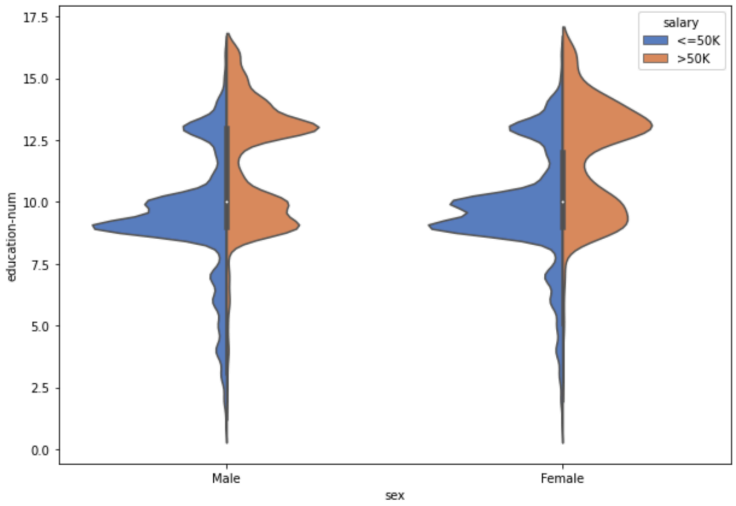
Through the use of **Exploratory Data Analysis** (EDA) I found out many odd distributions in the data, which later I presented it to my peers. I was able to view some stark inferences with Occupation, Sex and Education num based on the salary income, which is elaborated on the Analysis of my findings.

Other students had some interesting visualisation. It was a valuable experience and I learnt a lot from the peers as well. I am making to new representations which are different, so that I can learn better insights. Team work is very important and I was able to do this because of the insights made by teammates.

Analysis:

I sorted the marital-status into 4 types which are Not-Married, Married, Separated and Widowed. And the segregated them based on salary. All showed the same kind of distribution when graphed with educational number.

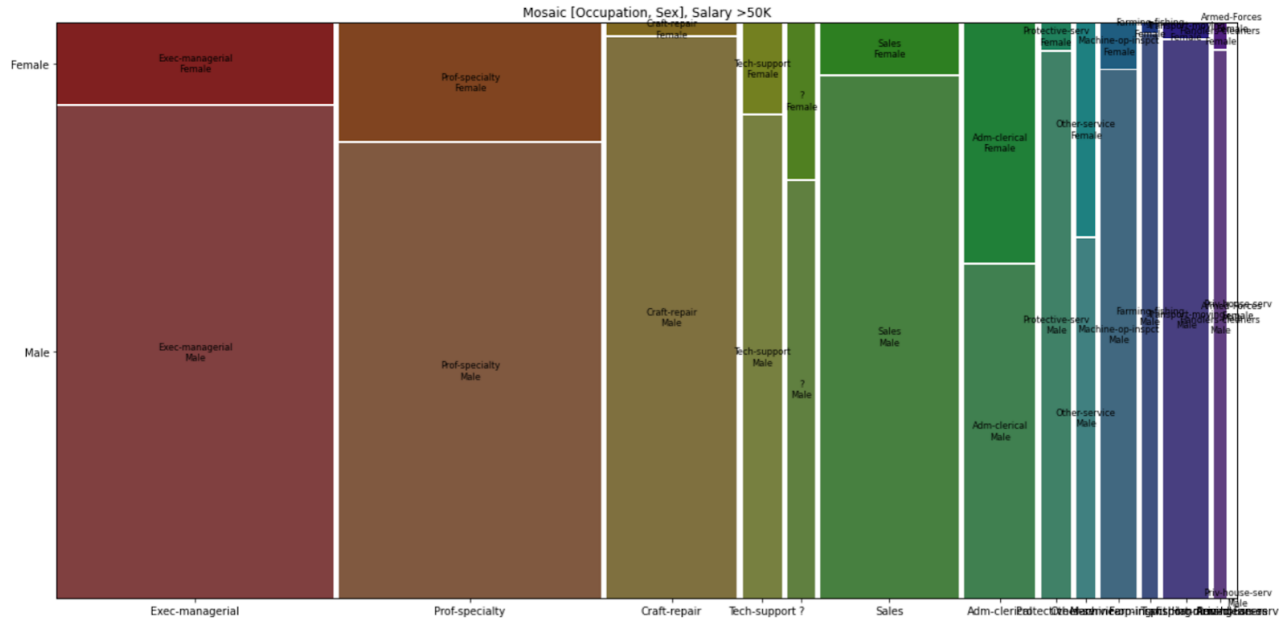
Then, I checked the Sex vs education status along with salary. I inferred that irrespective of marital status, there is a peak at edu-num 9 [8,10] and 13[12,14] which is **High school** and **Bachelors**. One interesting thing is education level corresponds to the salary. People with higher education tend to earn more.



Education-num represent the highest educational qualification obtained by an individual. There seems to be a positive correlation between salary and education-num. As people obtain higher valued education they seem to get paid more.

Similarly I began clubbing Occupation and salary. These are mosaic plots segregated through Occupation and Sex based on Salary. The inference is that for people earning below 50K, the salary is distributed fairly among various occupations for both genders. But it can't be said the same for high paying jobs (<50K). The data is skewed for people earning more than 50K towards males for most of the occupations.

It can be inferred the number of males is demographically more than females in the given dataset. Analysing them based on salary, we find that the data is skewed on male dominance. It becomes more apparent when seeing >50k class.



Similarly the educational vs salary was also insightful as male dominance can also be seen predominantly here. We also used bar charts, pie charts and scatter plots/matrix as a tool to show the distributions of salary between sex, education-num and marital status. I also checked the native country, only to as that 92% of the given sample is native to the United States. So, the data is skewed inherently to native country.

Some key learning which I learnt were,

1. Though some features are so apparently correlated, we can check all the features and it might come as a surprise. A peer made correlation matrix and we used that to view possible inferences. That means we cannot assume pre-fixed notions.
2. Interleaving the data, we can see some key insights as occupation vs sex based on salary. Two or three may give the data/visuals that might be used for key business ideas.
3. **Data visualisation is a key to successful business development as it elaborates the insights and it visually evokes the ideas, which I learnt first hand in this project.**

<u>Thank You!!</u>