Name: Chimge Tsendnyam

Project: Merging the Data and Storing in a Database/Visualizing Data

Three databases that I loaded into an SQLite database are ‘Glassdoor wage data’ from Kaggle, ‘Median Earnings data’ from US BLS and ‘Currency Conversion data’ from Currency Layer API. Loading them into one database and saving them as tables was pretty straightforward because I had saved the clean dataframes as CVSs beforehand. However, I found merge & join in SQLite to be very complex. Especially, all 3 of my datasets did not have any common immediate relationship. I tried to create commonality between the first 2 tables by adding department column in table#2. However, when I joined them using LEFT joins, it got merged very wrong, I had duplicate rows. I tried to join them by CROSS or FULL join, but I was getting errors. Because of that I had no way to join the datasets correctly. I think the biggest lesson learned is to make sure pick datasets that have commonality especially one to one and one to many relationships. I don’t think my datasets had any commonality and the fact that I was not able to join them as CROSS or FULL join left with me no choice. However, the goal of my picking wage related data was to see if there is wage gap in gender which I was able to do some visualization to clearly see that. Some of the biggest insight I found from my visualizations are :

* Age comes with experience amongst everything and as you get older the general trend is your base pay goes up
* One surprising founding was - bonus pay was higher for younger people
* Median base pay for male was $100k versus median base pay for female was around $90k
* There was not much difference in bonus pay for males and females
* Marketing associate is the title that has the highest frequency for females versus software engineer is the title that has the highest frequency for males
* When it comes to an education level, males hold the larger number of master’s degrees compared to females; however, there was not much difference in frequency for other education levels
* More males usually have higher performance rates compared to females