Self-Assessment

Over the course of this project, I played several roles. First off, I assisted my team in analyzing our data by creating pivot tables and graphs in the raw data file. This helped us understand the type of data we were working with, and potential questions we could answer. Second, I helped my team clean the data and create CSV files using Python and Pandas. Third, I helped build the Database in SQL, in addition to creating usable CSV files. Fourth, I assisted my team with creating visualizations in Tableau. Lastly, I helped build the final presentation, which included screenshots of code, written analysis, and more.

My greatest personal challenge was finding the best way to merge CSVs in SQL. Since the raw data was very messy, I had to think hard about the best way to merge the necessary data into the right tables. This required lots of trial and error but eventually, I was able to create a schema that correctly linked the CSV files in our SQL Database.

While I did not directly build the Machine Learning Model, I offered helpful feedback to my team. This required using Stack Overflow, prior examples, and other resources to help create our model. I also offered suggestions on how to build our presentations for each segment and I offered peer reviews as necessary.

Project Summary

For this project, we decided to analyze Household Income and Expenditures. Since personal finance is an important topic for everyone, we thought it would be interesting to try and predict if the income and expenditures would increase or decrease in the coming years. This would allow us to plan effectively for the future and make smart financial decisions. Because income and expenditure balances are continuous variables, we decided to use Linear Regression. In the end, our results were very close to the prior year’s results. However, we noticed that both income and expenditures are projected to decrease in the next year. While there was still room for improvement, we were mostly satisfied with our results.

Team Summary

In addition to regularly scheduled class time, our team would meet on Zoom throughout the week. While we did not assign specific roles, everyone worked together on each segment of the project, and this proved to be successful. It was also great to see each person’s strengths, as this helped us become more efficient in completing the project. A few challenges we encountered were picking the right questions to ask, creating our CSVs, and choosing the correct Machine Learning Model. We resolved these challenges by communicating with each other and playing off each other’s strengths. If we could do one thing differently, I would spend more time planning out the questions to ask and try grouping the expenditures into categories. This would have required a Classification Model which could have been even more insightful and provided better results.