Project 2: Exploring Customer Socio-Demographics

1. Uncovering the relationship between high-income and socio-economic factors of customers will allow a company to constantly refresh its 'prediction' of a customer's income without the necessity of invasive interaction (such as calls, questionnaires, etc.). Eliminating such invasive interactions will reduce communication cost. Finally, the knowledge of income status can aid a company devise a more targeted marketing and product alignment.

2. The entire purpose of this question is to get the students to think about 'Causality' and 'Correlation' and to achieve a self-discovery that although tempting to infer causality (i.e. 'these drivers are causing X') it may not be the case. The purpose of statistical models is to identify 'correlation', the co-occurrence of data. These may indeed be a result of causality, however, the difference must be made clear.

3. Students are to highlight any interesting insights (in their opinion) using distribution graphs, charts, bubble plots, etc. The objective is to explore the data.

4. See chapter 3, from page 44 on Supervised Learning.

5. See chapter 3, from page 44 on Supervised Learning.

6. Answers to this question will heavily depend on the selected model and any transformation that had been applied to the data. However, two key areas to focus on are the interpretability of the selected variables and their respective weights (especially is using a nonlinear technique, or if the selected variables are transformed variables). Secondly, focus should be on how the business might leverage on the insights from these variables.

7. Open ended question to get students to explore what is possible beyond the specific guided question. For instance, are there any other questions that can be answered using this data? Can other types of techniques be used and what would be their potential output/usage?

- You would want student to provide a detailed step-by-step representation from a thought-process-implementation representing what has been done, challenges faced and solution/workarounds.

- Students should analyze and combine the insights from the previous project questions to identify whether anything can be said about a) the products or b) the purchase pattern (i.e., customers). Students should make use of the data analysis, the model development as well as identified drivers to reach their conclusions

- The primary question that needs to be addressed is "How does the business make use of these insights to a) sell better b) understand customers better – and what can be done with these insights?"