

Assignment feedback: Jurgita Cepure

What you did well:

Analysis:

Jurgita, well done for delivering excellent Python and R scripts. You demonstrated a solid command of linear regression and decision tree models in your Jupyter Notebook. On top of this, you performed a thorough exploratory analysis of all the variables and their relationship with loyalty points which was great to see. Your use of both single and multiple linear regression clearly showed the key variables which influence customer loyalty points. In addition, I was pleased to see that you performed both TextBlob and VADER sentiment analyses on the dataset and you gave a balanced commentary to their strengths and weaknesses. Your use of Markdown commentary was excellent throughout, and you structured your work in an accessible manner. You demonstrated strong data wrangling skills in your natural language processing and the word clouds produced informed stakeholders of the underlying customer sentiments across product lines. Your R file was well structured and your insights were sound and clearly presented. Great work!

Report:

Your report is well structured and your analytical approach gives a helpful overview of the analytical process. Your exploratory analysis gives targeted insight into high-value customers which is oriented to the business scenario. It was clear that you engaged with the dataset through your extensive analysis which demonstrated your analytical rigour. The visualisations you produced linked your commentary to the analysis and you highlighted important findings through the use of emphasis which was useful. Your patterns and insights were sound and the recommendations were relevant to business objectives.

Presentation

This was a good presentation. It was clearly laid out, with an introduction, contextual exposition and recommendations. I thought your slides were professionally curated and you interpreted your visualisations nicely in relation to the business scenario. I liked that you started with the high-level explanations of the driver of loyalty points, before delving deeper into your analysis. Your choice of visualisations was great and they supported your narrative well. In addition, I was pleased that you elaborated on your clusters and suggested examples of tailored marketing to each of them. Nice job!

Areas for growth:

Analysis:

This was an excellent notebook, so what follows is minor commentary. You tested for multicollinearity in your multiple linear regression which was good, however you should have included tests for autocollinearity and heteroskedasticity for full credit. I was pleased to see you used functions with docstrings in your sentiment analysis. However, you could have extended this functionalisation across a wider proportion of your codebase for repetitive actions such as plotting and modelling. That being said, it is clear you left no stone unturned in your exploratory analysis and possess the mindset of a real analyst. I have no doubt that you will go far!

Report:

To make your analysis more reproducible, consider dedicating more space to the data cleaning decisions made in your analytical approach. I disagreed with your decision to prune the tree to depth 10, as a visual check of the tree shows splitting of sub-branches with very few samples. This increases the tree sensitivity to slight changes in the data, so when testing on new data there is a greater risk that the model will overreact to anomalies. However, you do note that optimal tree depth should be validated on unseen data which is correct. In addition, you correctly identify 5 prominent clusters using the silhouette and elbow method but decide to use 8 clusters due to the secondary peak on the silhouette. I take no issue with your justification here, however, I expected you to use this information to generate target customer profiles. This should have featured more prominently in your report, given its potential application to marketing. Your overview of patterns and predictions was very brief in relation to your extensive analysis which left the reader unclear about your most salient findings.

Presentation:

Your recommendations were logical from a data analyst's perspective, however think about your audience. The objective of the presentation is to sell your analysis to the stakeholders. As such, you should emphasise recommendations that have a clear (and immediate) value-add to the business. For example, which customer segment would you suggest targeting to increase ROI for the business? Think about how you can quantify the value of each customer segment are to the business in terms that stakeholders can understand (e.g. KPIs). The more specific your recommendations, the more likely they will be adopted by the business.

Jurgita, overall, this was an excellent submission. Great job!