Data Academy project feedback

Learner: Thomas Langlois

Marker: Robert Jirschik

Module: Association Rules

Indicative grade: Merit

Overall comments:

Overall, this is an almost perfect project (just narrowly missed distinction) that could have benefitted from an explanation of the metrics: support, confidence and lift; a justification of the thresholds and their meaning (e.g. connection between support and sample size) and some visualisations, if only to make the report more engaging.
A further, possibly more meaningful analysis when choosing different sizes for the itemsets (1-2 antecedents; 3+ consequents) may provide more interesting insights.

Data pipeline

The table below contains specific feedback on each section of the data science pipeline.

Stages	Feedback
Creating a data question *Collect, manipulate, and collate data from a range of sources to solve specific problems*	**Strengths** + Clearly identifies an appropriate data question and states testable hypotheses and initial expectations + Sources relevant data and explains its application to the organisation **Improvements**
Exploration *Use statistical tools and visualisations to explore & summarise data*	**Strengths** + Examines and discusses core features of the data + Uses a range of visualisation methods and interprets the results with detail and clarity: Love the heatmap and the use of the bokeh world map to display KYC forl different countries + Uses a range of well-chosen statistical methods to analyse data + Explores in detail the relevance of identified relationships and uses insights to guide the project **Improvements**
Preparation *Deal effectively with data issues and inconsistencies*	**Strengths** + Employs a range of appropriate data cleaning/preparation methods + Explains the preparation process and justifies decisions based on specific context + Discusses the analytical implications of chosen techniques and decisions + Shows awareness of potential issues with techniques chosen **Improvements**

Analysis *Use a range of analytical techniques to model and understand the data*	**Strengths** + Selected an analytical technique to model the data + Successfully employed an analytical technique + Additional analysis done based on country of residence to gain further insights **Improvements** - Justify and explain the parameters chosen (e..g support threshold) - Why was a 2->1 rule chosen? I would imagine based off one antecedent being able to confidently predict 2 or more consequents would have been more interesting?
Interpretation *Draw clear conclusions and insights from data*	**Strengths** + Draws detailed and well-considered conclusions from the model + Explores the limitations of the model and its applicability beyond the data + Fully considers the data question and the hypotheses **Improvements** - Could have explained what actionable insights for SG could be gained from the analysis - Suggest a direction of further research
Presentation *Clearly communicate ideas, insights and processes*	**Strengths** + The project is well-structured & presented consistently + The project has a clear narrative from beginning to end and is comprehensible by a reader without specialist knowledge + The text is well articulated and engaging + The pandas_profiling is a great touch! **Improvements** - Could have visualised more; either during the data exploration phase or when presenting the results of the Association Rules analysis