1.

The technique which I chose to perform on the dataset is CCA which is known as Canonical Correlation Analysis. This method was suitable for my dataset as I chose to find the association betting odds on various platform and the goals at various points in a match. From the CCA it is clear that only one variate needs to be taken for learning about the association. This can be known from the Bartlett’s Chi Squared test.

Text

Description automatically generated

We can also conclude that about 48% of betting is explained in goals. This can be seen from the output of Canonical Correlations.

Text

Description automatically generated

From the loadings, we can infer that Bet365 Away Odds, Bet&Win Away Odds, Interwetten Away Odds, etc. are majority of the variables that are important in correlation of betting odds with the goals. There is a pattern observed here that says that away odds contribute more towards the correlation of goals scored. In similar way, Full time Home Goals and Half time Home Goals explain more about the betting patterns. This can be clearly seen in the below images of Loading outputs.

Table

Description automatically generated

From this findings we can say that the away odds of various betting platforms and the home goals at full time and halftime are significant in explaining the relationship between goals and betting.

2a. The data seems to be very suitable as the authors are trying to study association between burnout symptoms and work satisfaction. Further, the data was collected in the from of an online survey which further supports the suitability of the collected data for Canonical Correlation Analysis.

2b.

Prior to applying Canonical Correlation Analysis, the researchers performed Confirmatory Factor Analysis before finalizing the subsets and moving on with the Canonical Correlation Analysis. The two groups of variables that are being correlated are Work Satisfaction and Burnout. The authors certainly use Confirmatory Factor Analysis with enough support from the Cronbach’s α coefficient test.

2c.

For comparing independent groups, the authors used the Welch test and to compare the categorical variables the authors used chi squared test. The authors evaluate the stability of the components and the quality of the correlation by comparing the loadings of the variates.

2d.

The authors concentrate on two correlates in their analysis. The correlates consist of exhaust, personal, and depersonal variables in Burnout subset and occupation, colleagues, development, payment, supervisor, and job variables in the Work Satisfaction subset. The authors do explain the correlates in terms of the original variables through helio plots, tables, and words.

2e.

Canonical Correlation allows them to draw correlation between two set of variables namely Burnout and Work Satisfaction. It is found there is a moderate correlation between the two set of variables. Increasing the work satisfaction of the nurses has the potential to increase the quality of treatment that the nurses give to the patients. This correlation can further improve the quality of healthcare provisions in the field of gastroenterology.

3 1a.

Text

Description automatically generated

From the Bartlett’s Chi-Squared Test, we can conclude that for variates 1 to 4, we accept the null hypothesis. This is because the Pr values are less than 0.05 for variates 1 to 4. In remaining cases we reject the null hypothesis. The test statistics, degree of freedom, and the p-value can be seen in the above snippet in a tabular form.

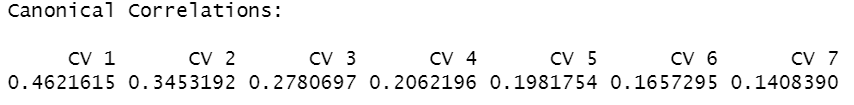
3 1b.

For finding the number of significant variates, we need to look at the Bartlett’s Chi-Squared Test statistics.

Here we can that until CV4, the Pr value is less than 0.05. CV5 onwards the Pr value crosses 0.05 mark. Hence, we can conclude that only 4 significant variates exist.

3 1c.

The first two canonical correlations also known as Cancor can be seen in the first section of the Canonical Correlation Analysis summary.



From the figure, the first two values corresponding to CV 1 and CV 2 denote the first two canonical correlation (Cancor) which are 0.4621615 and 0.3453192 respectively.

3 1d.

From the above same figure, we can conclude that about 46% of overlapping variance in variate 1 that is in CV 1 is explained by the variables in music and spending. Following similar method, 34% of the overlapping is explained by variate 2 that is CV 2.

Further conclusions can be drawn by looking at the below figure. In variate 1 for music subset that is in CV 1, the drivers for music mainly focus on Pop and Latino music. While for variate 2 of music subset that is in CV 2, more focus is on the Folk and Country music.

Table

Description automatically generated

Similarly, for variates talking about the spending subset, the variate 1 that is CV 1 mainly focuses on Shopping Centers and Branded Clothing. On the other hand, variate 2 that is CV 2 mainly focuses on Finances. This can be clearly seen in the below snippet of the output.

A screenshot of a computer

Description automatically generated with medium confidence

3 2a.

The formula for the first canonical variate will be

CV1 = 0.21(Music) + (-0.05)(Slow.songs.or.fast.songs) + 0.48(Dance) + (-0.1)(Folk) + 0.1(Country) + ……………… + 0.34(Latino) + 0.12(Techno.Trance) + (-0.13)(Opera) + 0.12(Finances) + 0.85(Shopping.centres) + ………… + 0.09(Spending.on.gadgets) + 0.001(Spending.on.healthy.eating)

3 2b.

The correlations between the first canonical variate for music and spending is given by the pictures below.

A screenshot of a computer

Description automatically generated with low confidence

Text

Description automatically generated

3 2c.

From the above figures we can make a conclusion that Pop and Latino genre of the music influence the Music variate the most. Similarly in case of Spending variate, Shopping Centres and Spending on Looks contribute to majority of the Spending Variate.