Online Shopper Clustering

Questions:

1. When you prepare the data for training the models, how did you deal with the categorical attributes, for the clustering models we examine here only takes numerical features?

Answer: Please find below table for more information,

Categorical attribute/s	Encoding
Weekend, Revenue	Converted the bool columns from TRUE > 1 and FALSE > 0
Month	Mean encoding
VisitorType	one hot encoding using pd.get_dummies
OperatingSystems	one hot encoding using pd.get_dummies
Browser	one hot encoding using pd.get_dummies
Region	one hot encoding using pd.get_dummies
TrafficType	one hot encoding using pd.get_dummies

Also, removed outliers for "ProductRelated_Duration" by Top coding/capping at 30,000.

2. Which model is better considering their RI scores?

Answer: Considering RI scores, AgglomerativeClustering gives a better model.

3. 3. Which model is better considering their DBI scores?

Answer: Considering DBI scores, AgglomerativeClustering gives a better model.

Results:

KMeans DBI score : 0.5638426101283036

KMeans RI score : 0.5749850822183398

AgglomerativeClustering DBI score : 0.4315864920507808

AgglomerativeClustering RI score : 0.7328076143278328