AI & ML Capstone Project

09th October 2020

Topic: Customer Segmentation

Grouping customers into sections based on their common characteristics is called Customer Segmentation. These clusters allow the companies to target the customers with the correct marketing message and tailor their offers for a specific group. This not only helps them boost their sales, but also helps them build customer relations and understand them in a better way.

In this project, our aim will be to perform customer segmentation on Online Retail Dataset (https://archive.ics.uci.edu/ml/datasets/Online+Retail#) to understand the customers. Given this dataset, our task is to:

a) Load the dataset and perform a descriptive analysis on it (Total number of entries, the column types, unique/non-null entries for each attribute, unique stock items, visualizing various attributes using bar charts/pie-charts and so on).

b) Perform data cleaning. Specifically, given the dataset, handle the entries that either have missing information or have attribute values that are not feasible such as negative quantity.

c) Perform data pre-processing for the required attribute fields.

d) Since this database has no additional attribute information for the customer, we will use RFM model (refer: https://clevertap.com/blog/rfm-analysis/) for segmentation. Modify the database to include RFM model attributes.

e) Now once you have your database ready, perform data clustering on this dataset by assuming a fixed number of clusters.

f) Find the optimal number of clusters that the customers can be divided into.