

678 Midterm Project Proposal

Qianhui Rong

11/11/2018

Introduction

This project will focus on customer segmentation, which is an important topic in Business-to-Customer marketing field. Customer segmentation is a concept of dividing customers into subsets, each of which has one or some shared characteristics that cause them to have the same demands. Therefore, if the company can successfully and efficiently segment its customers, its advertising methods will have a higher probability of success and its profitability will then be improved.

Objectives

The main data I'll use in this project is "E-Commerce Data - Actual transactions from UK retailer" from <https://www.kaggle.com/carrie1/ecommerce-data> with eight columns: InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID and Country for each customer who has completed at least a purchase.

Two objectives are expected:

- Clusters of customers
- Model(s) to predict each cluster's purchase amount next year

Implementation

Data Cleaning

Up till now, I've found that some data (like unit price and quantity) have negative value or zero. I'll simply delete them or do some other transformations(tbd).

Exploratory Data Analysis

Some interesting questions for EDA:

- Highest number of orders and money spent on purchases
- How many orders (per hour/month/day)?
- Transactional patterns for each country (highest amount, most frequent)
- Transactional patterns for unit price
- Which kinds of products are the most popular?

Modeling

To be determined after EDA.