

Name: Sajib Biswas

Roll: 180028

Date: 10<sup>th</sup> November, 2021

## Project-05

### Q-01: How to use this knowledge?

Answer: In this final section, we will investigate ways that you can make use of the clustered data. First, we will consider how the different groups of customers, the *customer segments*, may be affected differently by a specific delivery scheme. Then, we will consider how giving a label to each customer (which *segment* that customer belongs to) can provide for additional features about the customer data.

Companies will often run A/B tests when making small changes to their products or services to determine whether making that change will affect its customers positively or negatively. The wholesale distributor is considering changing its delivery service from currently 5 days a week to 3 days a week. However, the distributor will only make this change in delivery service for customers that react positively.

### Q-02: How can the wholesale distributor use the customer segments to determine which customers, if any, would react positively to the change in delivery service?

Answer: Making the change to the delivery service means that products will be delivered fewer times in a week.

The wholesale distributor can identify the clusters to conduct the A/B test on, but the test should be done on one cluster at a time because the two clusters represent different types of customers, so their delivery needs might be different, and their reaction to change will, thus, be different. In other words, the control and experiment groups should be from the same cluster, at a time.

Additional structure is derived from originally unlabeled data when using clustering techniques. Since each customer has a *customer segment* it best identifies with (depending on the clustering algorithm applied), we can consider '*customer segment*' as an engineered feature for the data. Assume

the wholesale distributor recently acquired ten new customers and each provided estimates for anticipated annual spending of each product category. Knowing these estimates, the wholesale distributor wants to classify each new customer to a *customer segment* to determine the most appropriate delivery service.

**Q-03: How can the wholesale distributor label the new customers using only their estimated product spending and the customer segment data?**

Answer: To label the new customers, the distributor will first need to build and train a supervised learner on the data that we labeled through clustering. The data to fit will be the estimated spends, and the target variable will be the customer segment i.e. 0 or 1 (i.e. grocery store or restaurant). They can then use the classifier to predict segments for new incoming data.