**Project collaboration with Dr. Angshu**

Project**: Implementation of Data Science techniques on retail data**

Dataset: <https://archive.ics.uci.edu/ml/datasets/online+retail>

This is an excellent open source dataset available on UCI’s repository (also owns most popular datasets like Iris, Cars etc). It has 500k records therefore is appropriate for retail domain.

**Tasks:**

1. Frequent itemset mining
2. Guest segmentation based on purchase patterns

**Algorithms used:**

1. Apriori and FP Growth for frequent itemset

This shows how frequent item sets can be mined quickly to give combo offers in retail. It uses one of the fastest frequent items set mining approach FPGrowth.

Furthermore, these frequent itemset can be used for deriving association rules for some of the common pricing and promotion strategies like:

* Loss leader pricing – one item sold at loss while this loss is gained from another item’s higher price

1. Clustering based on - Guest purchase patterns