**Demographic Identification of Cash Customer from Shopping Pattern**

**Omker Mahalanobish**

Senior Statistical Analyst

GDAP-Bangalore(Assortment)

[omker.mahalanobish@walmart.com](mailto:omker.mahalanobish@walmart.com)

**Somedip Karmakar**

Senior Statistical Analyst

GDAP-Bangalore(Assortment)

Somedip.karmakar@walmart.com

**Business problem:** Walmart services millions of customers in a store every day, and a major part of our customer base indulge in cash transactions, instead of going for credit cards. Although we could buy information from external companies regarding our customers’ demographic, we lose out on a lot of information pertaining to our cash customers, since they are not identifiable. This ‘incompleteness’ of data hinders us from identifying our true customer base and in turn portrays a rather biased image of our customers, posing a major roadblock in analyzing the POS data to come up with better customer facing decisions.

**Methodology:** In order to obtain optimum assortment decisions, knowledge of the true customer base is a bare minimum necessity.

We could fairly assume that different demographic groups have varied habits which gets reflected in their shopping patterns. Hence, assigning demographic features to our cash customers would enrich our database and in turn would let us gain deeper insights into our shoppers, without violating their privacy. Our algorithm would help identify the demographics of a cash customer, depending on her purchase pattern.

The model is trained on the POS scan data for identifiable customer base (i.e. our card customers), to obtain groups of customers who shop similarly.

The groups of customers thus obtained are profiled demographically to attach demographic features.

The basket of an un-identified shopper is then passed through the model to identify which customer group the purchase entails to, and the respective demographic features are assigned to the same.

Machine Learning techniques are used to validate and improve the models.

**Business Impact:** The output of the algorithm could be used to enrich our database, helping the analysts come up with better customer insights driven assortment strategy, in turn helping the leaders take better customer facing decisions.

***Keyword:*** *Machine Learning , Assortment Optimization , Market Basket Analysis*