The objective of this invention is to better understand the customer purchasing behavior by identifying the demographics of the customers, especially those using cash transactions.

Customer demographics are very useful in determining the suitable assortment of products which should be carried in the stores. Currently there is no method of identifying the cash customers.

In developed markets like US and UK, we lose about 40% information because of the intractability of the cash transactions, and this proportion increases significantly in Mexico, Canada, SA and China.

Even if the customer is making card purchases, demographic information is not always available or has to be bought from external vendors at a high cost, with no measure of accuracy.

The POS data is divided into traceable and cash transactions. The traceable transactions also have the tagged demographic information. We divide the traceable data into a training set and a validation set.

We first group the traceable transactions based on transaction metrics, like … Trip type, Trip time, basket metrics, variation in purchase pattern, share of wallet from different product attributes to get the groups of customers who have very similar purchase pattern. The groups are profiled on the demographics to get unique distinguishable demographic features for each customer segment.

Next the transactions in the training set are used to create a decision tree using a cross-validation technique, based on the transaction metrics to classify each transaction into a distinct demographic customer segment.

Next the cash transactions are fed into the decision tree model to get distinct customer segments.

**Implementation :**

While a transaction is made in the POS, the model runs depending upon the basket metrics and yields out the predicted demographic profile of the customer. This predicted profile is flashed on the POS screen, which is upvoted or downvoted by the associate managing the POS machine. This feedback is taken into account and the model parameters are fine-tuned based on the same.

Upon an up-vote, the predicted profile is directly integrated into the Walmart data, for further use in different analytical models, with a high score of confidence. A down-voted prediction is accompanied by a low score of confidence, when written into the Walmart database.

Use-cases :

* Customer Decision Tree
* Loyalty & Substitution
* Tested on UK and Canada mkts
* RFM Analysis