|  |  |
| --- | --- |
| ­­ | **2017** |
|  | Wal-Mart Stores, Inc.  Omker Mahalanobish Somedip Karmakar |

|  |
| --- |
| **[DEMOGRAPHIC IDENTIFICATION OF A CASH CUSTOMER FROM HER SHOPPING PATTERN]** |
| 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 of our cash customers, since they are not identifiable. Our algorithm would help identify the demographics of a cash customer, depending on her purchase pattern. In this way, we wouldn’t miss out on the information of our customers, in the process enriching our database, which would in turn help us to make better decisions to serve our customers. |

Contents

[DEMOGRAPHIC IDENTIFICATION OF A CASH CUSTOMER FROM HER SHOPPING PATTERN 2](#_Toc486596828)

[Overview 2](#_Toc486596829)

[Existing System Constraint 2](#_Toc486596830)

[Solution 2](#_Toc486596831)

[Overall Solution 2](#_Toc486596832)

[Analysis Explanation Diagram 4](#_Toc486596833)

[System Implementation 4](#_Toc486596834)

[Use Cases 5](#_Toc486596835)

## DEMOGRAPHIC IDENTIFICATION OF A CASH CUSTOMER FROM HER SHOPPING PATTERN

## Overview

* The idea is to know and infer more about our customer base.
* We lose a lot of information on our customer base, owing to the mode of transaction of a major chunk of our customers being cash.
* Cash customers are very difficult to identify, for there not being any key identifier to trace them successfully.
* Here we propose a statistical machine learning algorithm to identify a cash customer from her purchase pattern, thereby enriching our database.
* Knowing our customer better would in turn help us make better customer facing decisions.
* The algorithm builds up on the assumption that different demographic groups have varied habits which gets reflected in their shopping patterns.

## ­­Existing System Constraint

* In the existing system, we do not have any information about our cash customers.
* Even if we try to acquire customer information from external sources, like Nielsen, there is always a cost factor attached to it.

## Solution­­

### Overall Solution

* Creation of Training Data:
  + The Training data is created at a store level, using the demographic information of our identifiable customer base (i.e. the card customers), which is readily available to us.
  + With this data, is joined the visit-level purchase metrics in order to obtain a fully demographically annotated set of identified customers, along with their purchase pattern.
  + The Demographic Information of interest could be features like :
    - Gender
    - Marital Status
    - Ethnicity
    - Age-Group
    - Family Size
    - Educational Background
    - Income Level
    - House Property Ownership
  + The Purchase Pattern Metrics could include :
    - Trip type (eg: Grocery, Meat, Medicine, Services, GM, Miscellaneous)
    - Trip time (eg : Weekdays, Weeknights, Weekend Evenings, Weekend Mornings)
    - Basket Size (Total Size of the Basket, in terms of Units)
    - Basket Value (Total Value of the Basket, in terms of Dollars)
    - Departments actively visited (No. of departments purchased from)
    - Promotional Items Purchased (Percentage of promotional items purchased)
    - Average Sales Amount
    - Variation in Sales Amount
    - Share of wallet from different price buckets
    - Share of wallet from different brands and finelines
* Analysis of Training Data :
  + Once the Training Data is created, it has to be analyzed to obtain insights and decision rules from the same.
  + The traceable transactions also have the demographic information. We divide the traceable data into a training set and a validation set.
  + We first cluster the traceable transactions based on transaction metrics 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 based on the transaction metrics to classify each transaction into a distinct demographic customer segment.
  + Any customer group thus formed would consist of customers with homogeneous shopping pattern, whereas any two customer group would have customers who are starkly different from each other, in terms of their purchase pattern.
  + Each customer group thus formed is extensively analyzed to create demographic profiles using their Demographic Information.
  + In this way, we would have a demographic profile attached to each of our customer group.
  + This decision tree model is validated based on the validation set to fine-tune the model parameters and increase the model accuracy.
* Creation of Test Data :
  + The Test Data would consist of only the Purchase Pattern Metrics of the un-identified customers, at a state level.
  + This data is readily available to us, among our scan tables.
* Attaching Demographic Information to un-identified cash customers:
  + We run our Test Data through the Decision-Tree that we had built using the Training Data, to obtain similar Customer Group.
  + Since, the customers in a given Customer Group have similar purchase patterns, according to our assumption, they would also have similar demographic profiles.
  + Thus, we would attach to each customer group, their respective Demographic Profiles, as obtained using the Training Data.
* Introduction of Machine Learning:
  + Now, each un-identified customer in our database has a demographic profile.
  + Once we have built the Decision Tree on the Training Data, we can at real time, attach Demographic Profiles to each cash customer.
  + We would use this to flash the profile of the customer in our Point of Sale machines, which could be either up-voted or down-voted by the teller, depending on how much the visually identifiable demographics of the customer matches our profile predictions.
  + The algorithm will use the learning thus gained and train itself for better predictions with time.

## Analysis Explanation Diagram



## System Implementation with Feedback Loop



## Use Cases

* This algorithm could be easily deployed in the Wal-Mart stores.
* The system will help identifying customers who usually prefer cash to be their mode of purchase, without probing into their privacy.
* In the end, this would lead to enrichment of the Wal-Mart data, helping the Analysts to come up with better insights, which would help the leaders take better decisions to serve our customers.