Ayan Sarkar

Project idea proposal

|  |
| --- |
| NAME(s): Ayan Sarkar |
| TOPIC AREA:   * This study involves analyzing the purchasing patterns of the customers of a retail company. * One of the best examples of how we both overestimate and underestimate changes in the future is the evolution of consumer behavior throughout this century. And a lot of research has gone in this area and is one of the top areas which capitalizes the use of Analytics, to understand this ever-changing trend. * The potential to better understand this industry domain and the exciting opportunities to study the retail customer purchasing trends drove us to pick up this project. |
| BASIC RESEARCH (or ANALYSIS) QUESTIONS TO ANSWER:   * A retail company “ABC Private Limited” wants to understand the customer purchase behavior (specifically, purchase amount) against various products of different categories. They have shared purchase summary of various customers for selected high volume products from a particular month. * Now, they want to build a model to predict the purchase amount of customer against various products which will help them to create personalized offer for customers against different products. |
| WHAT TYPE OF DATA ARE YOU COLLECTING?   * This business case and its data has been taken from an open practice problem for users on an analytics website: <https://datahack.analyticsvidhya.com/contest/black-friday/> * The data was provided in .csv files * The data is divided into a training dataset (550068 rows) and a testing dataset (233600 rows) * The data set also contains customer demographics (age, gender, marital status, city\_type, stay\_in\_current\_city), product details (product\_id and product category) and Total purchase\_amount from last month.  |  |  | | --- | --- | | **Variable** | **Definition** | | **User\_ID** | User ID | | **Product\_ID** | Product ID | | **Gender** | Sex of User | | **Age** | Age in bins | | **Occupation** | Occupation (Masked) | | **City\_Category** | Category of the City (A,B,C) | | **Stay\_In\_Current\_City\_Years** | Number of years stay in current city | | **Marital\_Status** | Marital Status | | **Product\_Category\_1** | Product Category (Masked) | | **Product\_Category\_2** | Product may belong to another category also (Masked) | | **Product\_Category\_3** | Product may belong to another category also (Masked) | | **Purchase** | Purchase Amount (Target Variable) | |
| WHAT TYPE OF ANALYSIS IS PLANNED?   * K-Means clustering to identify customer segments with similar characteristics to each other, but different from the other segments. These segments can form the basis of separate, more specific marketing campaigns baes on attributes such as sex, income, socio-economic status, geographic region and more. * Application of Regression and Neural Nets to estimate the Purchase amount in our data set. |