**Consumer Behavior – Walmart Recruiting: Trip Type Classification**

Use market basket analysis to classify shopping trips

Currently, Walmart's trip types are created from a combination of existing customer insights ("art") and purchase history data ("science"). In their third recruiting competition, Walmart is challenging Kagglers to focus on the (data) science and classify customer trips using only a transactional dataset of the items they've purchased. Improving the science behind trip type classification will help Walmart refine their segmentation process.

For this competition, you are tasked with categorizing shopping trip types based on the items that customers purchased. To give a few hypothetical examples of trip types: a customer may make a small daily dinner trip, a weekly large grocery trip, a trip to buy gifts for an upcoming holiday, or a seasonal trip to buy clothes.

Walmart has categorized the trips contained in this data into 38 distinct types using a proprietary method applied to an extended set of data. You are challenged to recreate this categorization/clustering with a more limited set of features. This could provide new and more robust ways to categorize trips.

The training set (**train.csv**) contains a large number of customer visits with the TripType included. You must predict the TripType for each customer visit in the test set (**test.csv**). Each visit may only have one TripType. You will not be provided with more information than what is given in the data (e.g. what the TripTypes represent or more product information).

**The test set file is encrypted**. You must [complete this brief survey](https://form.jotform.com/52928247647972) to receive the password.

Data fields

* **TripType** - a categorical id representing the type of shopping trip the customer made. This is the ground truth that you are predicting. TripType\_999 is an "other" category.
* **VisitNumber** - an id corresponding to a single trip by a single customer
* **Weekday** - the weekday of the trip
* **Upc** - the UPC number of the product purchased
* **ScanCount** - the number of the given item that was purchased. A negative value indicates a product return.
* **DepartmentDescription** - a high-level description of the item's department
* **FinelineNumber** - a more refined category for each of the products, created by Walmart