

## Description

Walmart uses both art and science to continually make progress on their core mission of better understanding and serving their customers. One way Walmart is able to improve customers' shopping experiences is by segmenting their store visits into different trip types.



Whether they're on a last-minute run for new puppy supplies or leisurely making their way through a weekly grocery list, classifying trip types enables Walmart to create the best shopping experience for every customer.

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 you 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.

## Dataset Description

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).

## Data fields

- **TripType** - a categorical id representing the type of shopping trip the customer made. This is the ground truth that you are predicting.
- **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

## Evaluation

For each visit, you must submit a TripType:

## Submission Format

You must submit a csv file with the VisitNumber and a TripType mentioning the predicted Trip Type. The file must have a header and should look like the following:

A	B	
VisitNumber	TripType	
1	0	
2	0	
3	0	
4	0	
6	0	
13	0	
14	0	
16	0	