

TECH ALPHARETTA

2972 Webb Bridge Rd, Alpharetta, GA 30009

**Market Basket Analysis**

Market Basket Analysis is an invaluable exercise for retail and e-commerce stores.

In this challenge, you'll practice exploratory analysis, data wrangling, clustering, and general problem solving.

**Background**

**Market Basket Analysis** is the study of items that are likely to be purchased together.

The deliverable is to create groups of items that are frequently purchased together so that stores can better organize their layouts.

By doing so, stores can make complementary items easier to find, thus improving sales.



A **basket** is simply defined as all the items that were purchased together in a single order. Note: Some stores might define a "single order" as any items bought within X number of days.

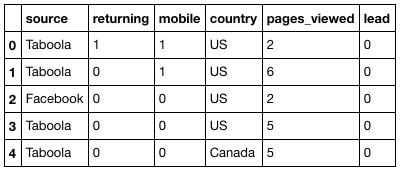
In this challenge, you will help a boutique clothing retailer with Market Basket Analysis.

You will use their customer's purchase history to create clusters of items that should be placed in the same section.

**Data**

The first table is called **customer\_baskets.csv**.

It contains the purchase history for 23,000 customers across 36,000 purchases.

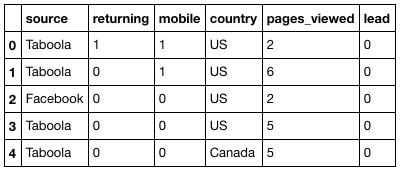


*Data Dictionary:*

* **customer\_id**- Unique customer ID.
* **basket** - Comma-separated list of item categories bought in the basket.

The second table is **clothing\_categories.csv**.

It contains a description of each clothing category offered in the store.



*Data Dictionary:*

* **category\_id**- Unique ID of the item category.
* **name** - Name of the category.
* **description** - Description of the category.

**Objectives**

Here's your challenge:

* First, explore the dataset. Answer the following questions.
  + Which item category was the most popular?
  + Which customer has bought the greatest number of items?
  + What is the distribution of basket sizes?
  + Tip: You may need to wrangle the dataset beforehand.
* Next, for each item category, find the ID of the customer who purchased that category the most. The retailer would like to interview them.
* Finally, create clusters of item categories that have the highest likelihood of being purchased together.
  + Tip: Each item should belong to only one cluster.