**Walmart Capstone Challenge # 1**

**Theme - Substitutes**

Develop an innovative data science solution to help Walmart gain a competitive advantage by choosing assortments that will drive customer loyalty and sales

**Walmart’s business model**

Walmart is a multinational retail corporation that operates a chain of hypermarkets, discount department stores and grocery stores across the U.S and in 28 countries in the world. Through innovation, Walmart is creating a unique experience to let customers shop anytime and anywhere online, through stores and via mobile devices.

**Walmart’s challenges and opportunities**Walmart is operating at an enormous scale: 260 million customers visit our 11,500 stores every week. In fiscal year 2016, Walmart generated revenue of $482.1 billion in revenue, making us the largest company in the world. Additionally, Walmart employs 2.3 million associates worldwide. Walmart’s mission is to provide all of our customers across the globe with the right products at the right place at the right time at the right price. By understanding how similar products are to each other we can serve our customers better by doing the following:

* Identify substitutes and complementary products within a category
* Understand which products should be placed next to each other on a shelf
* Make better decisions whether we should delete and product from a store (I might not want to delete an product if there is nothing else similar a customer could buy instead)

**Challenge**

Given Walmart product images and attributes, identify which products are substitutes. These products should be grouped into substitutable communities. Also, given additional images and attributes, determine which substitutable community the new products should belong to.

**Definitions**

An assortment is the set of products available to a customer when they visit a store.

A substitute is a similar product that a customer could by *instead* of another product. For example, if I regularly buy Yoplait strawberry single serve yogurt, the Chobani strawberry single serve as well as the Yoplait blueberry yogurts can be considered substitutes since I may buy them instead of my original Yoplait strawberry yogurt.

A community is a group of like products.

**Dataset**

There are two zipped folders – each folder has a set of pictures belonging to that dataset. The image name indicates the product identifier or UPCS. The dataset is available at the following link:

<https://www.dropbox.com/sh/c9h05pk34jqzirz/AABvjpYh0_1zsQYAKkznJhtfa?dl=0>

**Approach**

The students are allowed to use any approach that works for them, however they are encouraged to use a deep learning approach for image classification and grouping and finally taxonomy.