**ASSIGNMENT 3**

In this assignment, you will use a very large dataset that contains customer orders for a grocery store.  Since the data is voluminous, your job is to explore the data using PCA and Clustering to determine if you can find some interesting patterns in the data.  Be sure to visualize a variety of principal components in your analysis and compare the results to pick the best solution for your clusters.

**Description of the Data**

The dataset contains order information from a grocery chain for customers in a single transaction file.

* **order\_id -** order number
* **product\_id -** product number
* **cart\_order** - the sequence in which they have been added to the cart in that order
* **reordered** - indicator whether the customer has a previous order that contains the product
* **product\_name** - Product name
* **aisle\_id** - the grocery store aisle ID where the product can be found
* **aisle** - the grocery store aisle descriptor where the product can be found
* **department\_id** - internal grocery store department ID
* **cust\_id** - customer identifier
* **source** - Prior means that this data came from the history of purchases (you can ignore this)
* **order\_number** - Internal order ID
* **order\_dow** - the day of the week in which the order was placed
* **order\_hour\_of\_day** - the hour of the day in which the order was placed
* **days\_since\_prior\_order** - the number of days since the last order