**Final Project Proposal**

As supply chain management and retail tech is growing at faster pace, maintaining log and predicting customer’s next buy is much a highly valuable use case for a Data science student. Leveraging the skills and coding techniques we learnt in DATA643 our group wants to develop a recommender system that predicts which product the user would buy again.

* **Project Name**: Instacart Market Basket Analysis
* **Data set**: The dataset for project is a relational set of files describing customers' orders over time. The dataset is anonymized and contains a sample of over 3 million grocery orders from more than 200,000 Instacart users.
* **Data Source:** <https://www.kaggle.com/c/instacart-market-basket-analysis/data>
* **Model:** Recommender System(Visualizations included)
* **End Goal:** Predicting next best purchase for every order\_id (key-id for customer)-Real Time model and application
* **Though process: (Roadmap Planned)**
* Different Type of recommender Algorithm.
* Using Advanced Mathematical techniques.
* Using Spark or Hadoop for distributed computing.
* Project is officially declared as open source and open challenge by Instacart.com.
* **More info and references:**
* <https://tech.instacart.com/3-million-instacart-orders-open-sourced-d40d29ead6f2>
* Team Members: 1) Rose Koh

2) Harpreet Shoker