

# Instacart Market Basket Analysis

Data Science Challenge – Round 2

May, 2022

Proposed By:  
**Noran Hany**  
**Toka Khaled**

Under the Supervision of:  
**Dr. Doaa Mahmoud**



The background features a light gray color with several thin, dark gray circles of varying sizes. A large, bold, light gray number '1' is positioned on the left side, partially overlapping the word 'Introduction'.

# 1 Introduction

This project exploits the power of data analysis and machine learning to take business to the next level!

# 2 Project Description

Data can be divided into 3 parts:

- Prior data : Order history of every user . This data contains nearly 3–100 past orders per user
- Train data : Current order data of every user . This data contains only 1 order per user
- Test data : Future order data of every user . This data will not contain any product information (We will predict this)

# 2 Data Description

## Data exist across 6 csv files:

orders.csv — consists of order details placed by any user — shape: (3421083, 7)

- Order\_id : Unique for every order
- User\_id : Unique for every user
- Eval\_set : ( prior / train / test)
- Order\_number : ith order placed by user
- Order\_dow : Day of week
- Order\_hour\_of\_day : Time of day in hr
- Days\_since\_prior\_order : difference in days between 2 orders

products.csv — details of a product, shape: (49688, 4)

- product\_id : product ID of item
- product\_name : name of product
- aisle\_id : aisle id of the product
- department\_id : department id of the product

Aisles.csv — details of aisles, shape: (134,2)

- aisle\_id : aisle ID
- aisle\_name : name of aisle

department.csv — details of department, shape: (21,2)

- department\_id : department ID
- department\_name : name of department

# 2 Data Description

`order_products__prior.csv` — consists of all product details for any prior order, shape: (32434489, 4)

- `order_id` : Unique order id for every order
- `product_id` : product ID of item
- `add_to_cart_order` : denotes the sequence in which products were added to cart.
- `reordered` : product is reordered ? (1/0)

`order_products__train.csv` — consists of all product details for a train order, shape: (1384617, 4)

- `order_id` : Unique order id for every order
- `product_id` : product ID of item
- `add_to_cart_order` : denotes the sequence in which products were added to cart.
- `reordered` : product is reordered ? (1/0)

# Models to be developed

## Model 1: Predicting Future Order.

Predict the products which are most likely to be ordered in a user's future order based on the user's purchasing history and preferences of products.

## Model 2: Associations between Products.

Applying an association rules algorithm to find rules and associations between products independent of any user. This is a non-user-centric algorithm.

The association rules will be later used in business to:

- Predict the next likely product the customer would purchase during the ordering process.
- Design positions of products in the software application's home page, so that products most likely to be ordered together are placed next to each other.
- Make offers by bundling a low frequently buying product with a high frequent one having high association with each other.