

Instacart dataset

Dataset link: <https://www.kaggle.com/competitions/instacart-market-basket-analysis/data>

Task:

Perform EDA for this dataset. Prepare a well commented python notebook and document all your answers to the following questions -

1. What is the shape of input data tables? Are there any null values? Summary stats for each table.
2. How is the frequency of orders varying across weekdays, hour of day?
3. What is the avg gap between two consecutive orders for a particular customer? Are there any patterns?
4. What is the avg basket size? Which products are majorly reordered?
5. Identify top product categories/departments, categories with higher reorder rate.
6. Which products are added to the cart first? Explore relationship between the order in which products are added to the cart and the reorder frequency.
7. How is the variation in the basket size of customers across the history of prior orders?
8. Do customers with higher no. of orders have higher proportion of reordered products?

You can also explore some other questions which you are curious about.