Hi Professor Provost,

We would like to work on a next purchase prediction project from the dataset provided by Instacart.

https://www.instacart.com/datasets/grocery-shopping-2017

The business use case would be for Instacart to increase the average basket size by recommending the next item the model predicts the customer will purchase.

We have 3.4 million orders to train and test.

Our target variable is the reordered variable and we can use the following variable as predictors:

* add\_to\_cart\_order: order in which each product was added to cart
* product\_name: name of the product
* aisle: the name of the aisle
* department: the name of the department
* order\_number: the order sequence number for this user
* order\_dow: the day of the week the order was placed on
* order\_hour\_of\_day: the hour of the day the order was placed on
* days\_since\_prior\_order: days since the last order

This data was also used for a Kaggle competition.

https://www.kaggle.com/c/instacart-market-basket-analysis

For this project we will have to repurpose the data since we won’t have the test set that Kaggle set aside for the competition.

by dropping the 75K test set from the orders set since we won’t have access to those true values and setting a portion of the order data for the new test set.