

# Take Home Test Instructions

## Overview of the exercise

The goal of this exercise is to suggest to Instacart customers items that they might want to purchase based on past orders. The point is not to create an award-winning algorithm, but to place yourself in the shoes of a data science consultant. Many positions at Capgemini require you to become familiar with a technique quickly, even if your prior knowledge is not in that subject area. Your job is to get smart on the subject, create a working prototype and draft a presentation that can be given to a client using only 15 hours of work time.

## Instacart dataset

First, get your data! You can download the dataset [here](#).

The data dictionary, explaining all the features in your dataset is available [here](#).

## Instructions

Your client has asked you to build a recommendation engine based on this data. The goal is to recommend to clients products that they are likely to purchase. In other words, your task here is to predict as accurately as possible what clients will purchase!

The “order\_products\_\_prior.csv” file contains information on orders that clients have previously placed.

The “order\_products\_\_train\_cap.csv” (train set) and “order\_products\_\_test\_cap.csv” (test set) contain new orders. **The goal of the exercise is to predict the items in each order of the test set, given the order\_id field and the data on the order contained in the other files in the folder.**

**You will get back to your client with the following:**

- A presentation containing an exploration of the data, a description of your approach, the results you obtained, and any interesting insights you are able to generate
- Your code. Python or R is preferred for the coding, you may feel free to use other relevant tools for data visualization
- Be prepared to present your results and answer questions

## Guiding questions

To help you find your way through this exercise, we provide a list of questions that might be interesting to answer. These questions are not mandatory, you don't have to answer all of them - they are merely here to help you get started and find your way through this exercise!

### 1) Data Exploration

- What initial insights can you get from a first exploration of the dataset?

- Do some variables seem to have more importance than other? What transformations might be needed?

## 2) Prediction

- Describe your overall approach: how did you formulate the problem to make recommendations?
- What model did you choose? Why?
- How did you assess performance of your model? Which metrics seemed particularly important to you?
- Please present your results and model performance
- How would you suggest delivering the recommendations to your client?

## 3) Insights and Next Steps

- What insights can you extract from your analysis?
- What are some of the things you would suggest as next steps for your client?

## Important Note

The recruiting team is aware that this was a Kaggle competition in 2017. As data scientists, using publicly accessible examples and code (with proper citing) is part of our practice. However, we ask that you refrain from using the kernels and notebook solutions that have been posted in the Instacart Kaggle competition. The interview team, and data scientists who will review your work, are familiar with these solutions. That being said, you are free to use any other resources that you see fit. **Please be prepared to discuss your code and how methodology during the interview phase.**

\*\*\*Day of the week begins on Sunday in this dataset.