

1. Instacart Market Basket Analysis: Predict customer reorder pattern, build recommendation system.

Data source: Kaggle

Instacart is a grocery ordering and delivery app, which allows you to select products through their app, and then personal shoppers review your order and do in the in-store shopping and delivery for you. In other words, Instacart delivers groceries from your favorite stores to your door. The company is expanding its platform to cover 90 millions US household in 2018. With millions of transactions in real time, Instacart's problem is a representative of a problem I would like to work on as a data scientist: predict customer behaviors with large amount of data. This project will focus on:

Which products a user would buy again, try for the first time, or add to their cart next during a session?

2. Zillow Prize: Zillow's Home Value Prediction (Zestimate)

Data source: Kaggle

For this project, I will build a home valuation algorithm from the ground up,

"Zestimates" are estimated home values based on 7.5 million statistical and machine learning models that analyze hundreds of data points on each property. And, by continually improving the median margin of error (from 14% at the onset to 5% today), Zillow has since become established as one of the largest, most trusted marketplaces for real estate information in the U.S. and a leading example of impactful machine learning.

3. Customer churns prediction

4. Fraud Detection