This document is a brief write-up for 3 proposed capstone projects.

Project 1:

Data: from kaggle competition “Santander Product Recommendation”, Santander is a bank which offers financial products and services.

The whole dataset is 1.5 years and has monthly records of multiple columns that store customer’s information, such as customer code, status of employment, customer’s country residence, age, gender of customer, gross income of the household, etc. There are also 24 product columns storing product items a customer has, such as "credit card", "savings account", “mortgage”, “short-term deposits”, “medium-term deposits”, “long-term deposits”, etc.

Link of data source: <https://www.kaggle.com/c/santander-product-recommendation>

Goal: To predict what products Santander’ customers will purchase in the next month based on past data.

Project 2:

Data: from kaggle competition “TalkingData Mobile User Demographics”, TalkingData is China’s largest third-party mobile data platform.

Fields in the dataset include gender, age, event, apps’ labels, label categories, device ids, brand and models.

Link of data source: <https://www.kaggle.com/c/talkingdata-mobile-user-demographics>

Goal: To predicting users’ demographic characteristics (gender and age) based on their app downloads, geolocation, and mobile device properties.

Project 3:

Data: from kaggle competition “Two Sigma Connect: Rental Listing Inquiries”, Two Sigma is a company running business in technology and investment management.

Dataset is consisted of 16 columns, namely bathroom number, bedroom number, building id, created, description, display address, features, latitude, listing id, longitudes, manager’s id, photo links, price, street address, interest level(target field).

Link of data source: https://www.kaggle.com/c/two-sigma-connect-rental-listing-inquiries

Goal: To predict how much interest each rental post will receive based on apartment info provided in the data.