**Santander Predictive Model**

**Brief**

# Can you pair products with people?

Ready to make a downpayment on your first house? Or looking to leverage the equity in the home you have? To support needs for a range of financial decisions, [Santander Bank](https://www.santanderbank.com/us/personal) offers a lending hand to their customers through personalized product recommendations.



Under their current system, a small number of Santander’s customers receive many recommendations while many others rarely see any resulting in an uneven customer experience. In their second competition, Santander is challenging Kagglers to predict which products their existing customers will use in the next month based on their past behavior and that of similar customers.

With a more effective recommendation system in place, Santander can better meet the individual needs of all customers and ensure their satisfaction no matter where they are in life.

**Data**

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

So, the data is available in CSV format. I must profile and cleanse this data also likely you will need to make additional ratios as columns derived from those in the dataset.

We need to find significant columns in predicting what products the various buckets of customer types will likely buy in the final month June 2016.

Preparation work which will help you are below:

**Python scikit-learn regression models:**

<http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html>

<https://www.youtube.com/watch?v=dSYJVbj4Eew>

**Statistics and Regressions, -p value and correlations**

<https://www.coursera.org/learn/basic-statistics/home/week/2>

I can will also send you over some nice templates for data profiling with Pandas and plot.ly.