1. Project Description : <https://www.kaggle.com/c/sberbank-russian-housing-market#description>

Housing costs demand a significant investment from both consumers and developers. And when it comes to planning a budget—whether personal or corporate—the last thing anyone needs is uncertainty about one of their biggest expenses. [Sberbank](https://www.kaggle.com/sberbank), Russia’s oldest and largest bank, helps their customers by making predictions about realty prices so renters, developers, and lenders are more confident when they sign a lease or purchase a building.

Although the housing market is relatively stable in Russia, the country’s volatile economy makes forecasting prices as a function of apartment characteristics a unique challenge. Complex interactions between housing features such as number of bedrooms and location are enough to make pricing predictions complicated. Adding an unstable economy to the mix means Sberbank and their customers need more than simple regression models in their arsenal.

In this competition, Sberbank is challenging Kagglers to develop algorithms which use a broad spectrum of features to predict realty prices. Competitors will rely on a rich dataset that includes housing data and macroeconomic patterns. An accurate forecasting model will allow Sberbank to provide more certainty to their customers in an uncertain economy.

Submissions are evaluated on the [RMSLE](https://www.kaggle.com/wiki/RootMeanSquaredLogarithmicError) between their predicted prices and the actual data. The target variable, called **price\_doc** in the training set, is the sale price of each property.

1. Plan from Week 5 to Week 10 on completing the project,

Week 5: Data exploration & Visualization.

Week 6: Data Munging.

Week 7: Feature Engineering.

Week 8: Create model using linear regression.

Week 9: Train, test model and calculate RMLSE.

Week 10: Improve model and document the process.

1. Team Name.

Data Science Rocks: Capstone Project Group 23

1. Dataset

The aim of this competition is to predict the sale price of each property. The target variable is called **price\_doc** in train.csv.

The training data is from August 2011 to June 2015, and the test set is from July 2015 to May 2016. The dataset also includes information about overall conditions in Russia's economy and finance sector, so you can focus on generating accurate price forecasts for individual properties, without needing to second-guess what the business cycle will do.

## Data Files

* **train.csv, test.csv:** information about individual transactions. The rows are indexed by the "id" field, which refers to individual transactions (particular properties might appear more than once, in separate transactions). These files also include supplementary information about the local area of each property.
* **macro.csv:** data on Russia's macro economy and financial sector (could be joined to the train and test sets on the "timestamp" column)
* **sample\_submission.csv:** an example submission file in the correct format
* **data\_dictionary.txt:** explanations of the fields available in the other data files