gTech
‹professional services/›
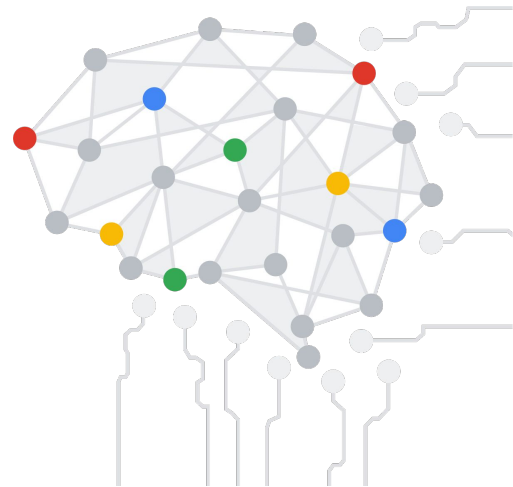
WiDS Datathon

Data Science Code Walkthrough

Manisha Arora, gTech Data Scientist

Feb 2022

What we will cover in the next 45 minutes

- Problem Statement Overview

- Understanding Data

- Exploratory Data Analysis

- Model Building

- Insights & Recommendations

gTech
‹professional services/›

What we will gain from this exercise

This workshop is ideal for you if you:

- Have some working knowledge about Python programming

- Have some understanding of machine learning through self-study or an online course

- Are looking to expand your knowledge and skills in programming and data science

What you will take away from this:

- A deeper understanding of solving end-to-end business problems

- Greater comfort with programming and machine learning skills

- Using analytical and story-telling skills to drive data-driven decision making

Problem Statement Overview

Link to Kaggle Case study:

https://www.kaggle.com/c/house-prices-advanced-regression-techniques/overview

The case study presents information about various aspects of residential homes in

Ames, Iowa. We want to understand what drives prices of residential homes in the

real-estate market and which features are most important for predicting sale prices?

gTech
‹professional services/ ›

Understanding Data

We are provided 4 data files:

- Train.csv - the training set

- Test.csv - the test set

- data_description.txt - full description of each column

- sample_submission.csv - a benchmark submission from a linear regression on year and month of sale, lot square footage, and number of bedrooms

Let's look at some codes ...

gTech
‹professional services/›