House Pricing

Intermediate Presentation

Annela Pajumets, Kadi-Liis Kuum,
Kadi Kilgi, Laura Anna Tammesoo

Who we are

- Four master students from Actuarial and Financial Engineering

Kadi

Annela

Kadi-Liis

Laura

Idea of the project

- Kaggle competition

- Model the sale prices

- 79 feature variables to use

  - Training - 1460 examples of different houses

  - Test - 1459

  - USA, Iowa, Ames

What we have managed to do

- Primary data analysis

- 32 numerical, 47 categorical

- Sale Price - numerical
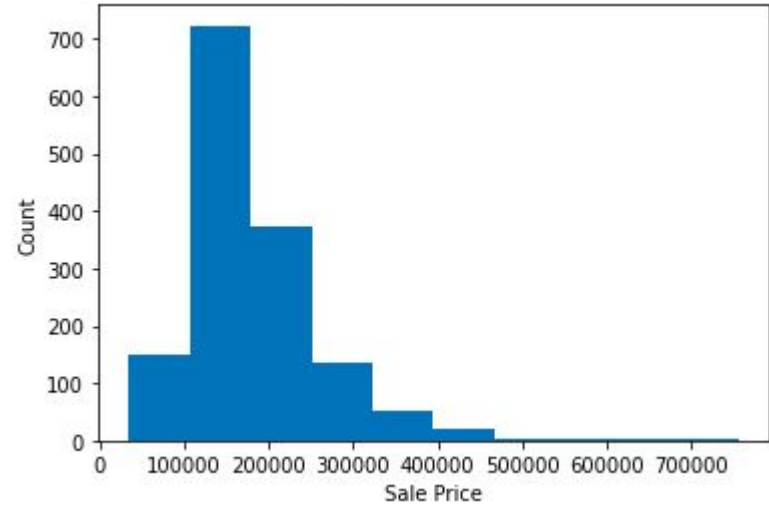
  - Min - 35 000$

  - Max - 760 000$

  - Mean - 180 000$

Figure 1. Distribution of sales prices

What is up next for us

- Regression
- Tree methods
- Ensemble
- PCA

What difficulties we have had

- Find the time
- Data is overwhelming
- Missing values
- More categorical values than numerical

Thank you for listening!

Reference

Picture taken from:

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