

DATA SCIENCE

TRAINING PROGRAM

Lab 2: Descriptive Analysis and Basic Visualisation

Nico Rotstein, Kodit.io
2019-10-16

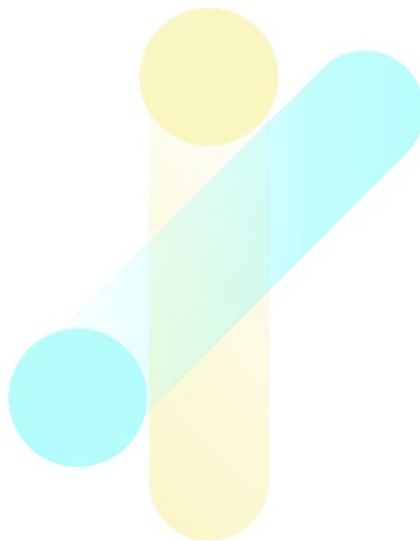
How to Lab

- Lab projects: dataset + challenge (no tests!)
- Groups of 1-3 people
- Get staff help on Mondays after lecture, and Wednesdays
- Questions to peers on the Community Platform [www.theshortcutcommunity.org]



Dataset 1

- Real estate transactions
<https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data>
- Register to Kaggle and download
- Features:
 - SalePrice: sale price in dollars (target variable to predict)
 - LotArea: lot size in square feet
 - Utilities: type of utilities available
 - BldgType: type of dwelling
 - YearBuilt: original construction date
 - ...



Lab 2

- Apply techniques learnt on [Monday](#)
- The dataset, inside out
 - use descriptive analysis metrics
 - are all features useful?
 - how "balanced" is the target variable?
- Show me the data
 - pick and use visualisation techniques
 - seeing the data is helping you how?
 - feeling like going back to descriptive analysis?

