Price prediction for Airbnb in Tokyo

Data Science Related
- Hacking Skills
  - Python
- Math & Statics Knowledge
  - Chose attributes based on correlation between ‘price’ and other attributes
- Substantive Expertise
  - Chose famous areas for tourists

Hands-On
- Python
- Library
  - tensorflow (keras), pandas, matplotlib

Data Acquisition
- Inside Airbnb - http://insideairbnb.com/get-the-data.html
- 106 columns -> 35 columns (The absolutes value of correlation with ‘price’ > 0.05)

Learning
- Feature selection (Process 1, 5)
- Remove outliers (Process 3)
- One-hot encoding (Process 4)
- Normalization (Process 4)
- Neural Network (Process 6)
- Preventing overfitting (Process 6)

Feedback
- Model: Neural Network
- The reason why I chose specific area is that each area have similar features. It helps to get high accurate.

Process
1. Drop apparently useless attributes
   - Id, listing_url, host_name, etc…
2. Extract data based on areas
   - Shibuya, Shinjuku, Toshima, Taito, Minato
3. Drop outliers, missing data.
4. Compute correlation between ‘price’ and other attributes
5. Extract attributes which have relatively high correlation
   - accommodates: 0.579
   - guests_included: 0.491
   - cleaning_fee: 0.484 etc…
6. Build and train neural network model
   1. Hidden layer
      - Dimensionality of the output space: 256
      - Activation function: Exponential linear unit
   2. Optimizer
      - RMSprop
7. Predict ‘price’ with the model