EDA

**General process:**

1. Merge property csv with train csv, remove all column which does not have logerror column – the shape new dataframe: 167888 \* 60
2. Drop some useless columns (such as parcelid)
3. 21 category columns and 32 numeric columns

**Missing values and outliers:**

Numeric columns:

1. Remove columns which missing more than 90% of the values
2. Replace nan with mean values

Category columns:

1. Remove some columns (see analysis process in code)
2. Replace null value with False for binary column
3. Replace null value with mode/ or the most frequent type (according to the column definition in excel) for other category columns