Uncleaning the dataset

Calculations

48895 rows excluding header were present in the document

Used only 26 % of entire dataset i.e. 12713 removed and applied variations into it. 1% extra to involve margin of error.

Recipe:

**Completeness :**

To make data in complete selected filtering the neighbourhood group Brooklyn

Replaced all the some of the column’s like latitude, no of reviews, reviews per month values for it as blank.

**Consistency:**

To make data inconsistent selected rows by filtering the neighbourhood group Queens.

Unique column id is replaced with duplicate values in all the rows

Replaced the date with MM-DD-YYYY instead of DD-MM-YYYY.

**Accuracy**:

To make data inaccurate selected rows by filtering the neighbourhood group Manhattan.

Replaced Lat, long with wrong co-ordinates

**Validity:**

To make data invalid selected rows by filtering the neighbourhood group Manhattan.

Replaced price with negative value

To make data invalid selected rows by filtering the neighbourhood group Bronx.

Replaced Minimum Nights with negative values

Replaced last review date with string value

**Uniqueness:**

To make data not unique selected rows by filtering the neighbourhood group Queens.

Unique column id is replaced with duplicate values in all the rows

**Conformity:**

To remove conformity selected rows by filtering the neighbourhood group Queens.

Replaced the date with MM-DD-YYYY instead of DD-MM-YYYY.