

DSC1111

## Formative Assessment 2

Zion John Yousef T. Ramilo

### Reflection

Within the cleaning process of the datasets, the only place where the handling of different date ranges was performed was when we had to let the data entries for reviews be unique to the listing ID and the date. Where the process was, we had to group by listing ID and date to ensure the uniqueness of the data entry, then join the intersection between the original dataset and the grouped data.

The data introduced that there were entries that would be removed or trimmed due to the nature of the process, since there has been an imbalance or different usage of the listing ID and date. This imbalance was handled by using the intersection join, where, although there would be some trimmed information, it would be accurate and representative of the goal of the analysis.

Consistent time filtering is important in multi-source data sets because you have to determine whether or not the data entries would be consistent and would not be able to introduce any kind of inconsistencies, as once measured, the data would be inaccurate and misrepresented.