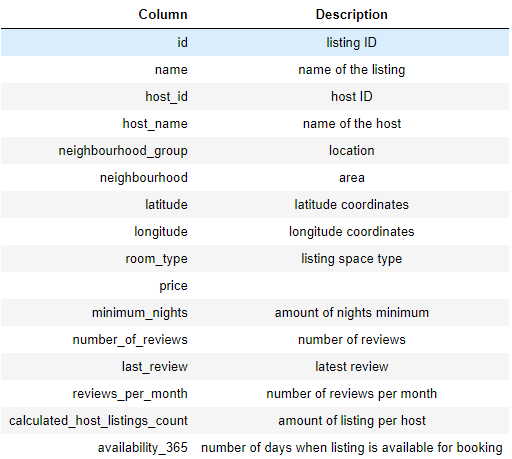
**Methodology**

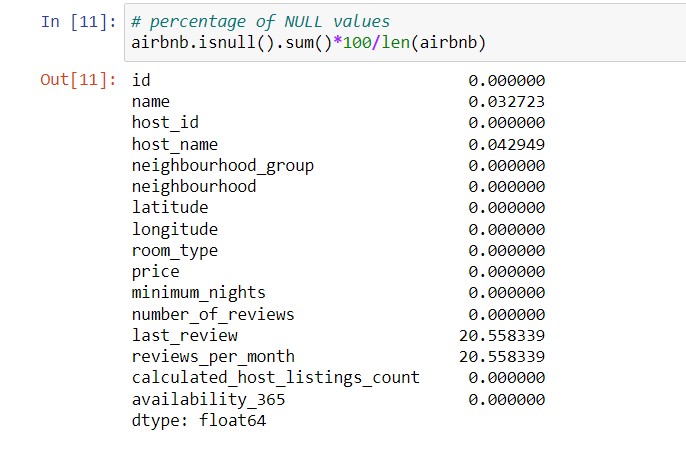
# **Data**

* PYTHON (JUPYTER) used for initial Data understanding and processing.
* Here is a snapshot of data dictionary:



# **Data Pre-processing**

1. NULL values treatment



* Deleting all rows where either “name” or “host\_name” is missing

C:\Users\Kunal Sadana\Desktop\UpGrad\COURSES\5. Advanced ML and Story Telling\10. Airbnb CaseStudy\images\PYTHON\Screenshot_2.jpg

* Filling NULL values with 0 for “reviews\_per\_month” since the same rows which had NULL values for this column had 0 for “number\_of\_reviews”.

C:\Users\Kunal Sadana\Desktop\UpGrad\COURSES\5. Advanced ML and Story Telling\10. Airbnb CaseStudy\images\PYTHON\Screenshot_3.jpg

1. Conversion of datatype

* Converting “last\_review” to datetime

C:\Users\Kunal Sadana\Desktop\UpGrad\COURSES\5. Advanced ML and Story Telling\10. Airbnb CaseStudy\images\PYTHON\Screenshot_4.jpg

1. Saving the file

* Saving the pre-processed file to be used in TABLEAU for visualisation

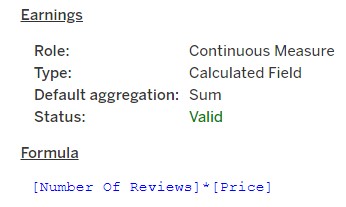
C:\Users\Kunal Sadana\Desktop\UpGrad\COURSES\5. Advanced ML and Story Telling\10. Airbnb CaseStudy\images\PYTHON\Screenshot_5.jpg

# **Data Visualisation**

TABLEAU used for the Data Visualisation purposes

* Creation of new features/variables to help in better analysis via visualisation:

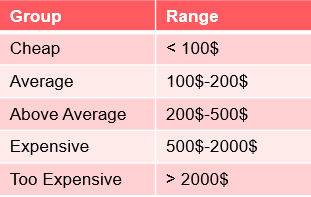
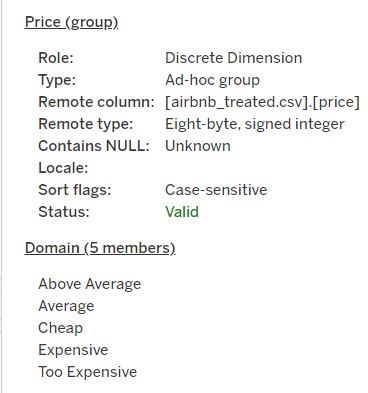
1. Earnings: “Earnings” is a calculated feature derived by given formula.



1. Availability 365 (groups): “Availability 365 (groups)” is a calculated feature where Availability has been binned into following categories:



1. Price (groups): “Price (groups)” is a calculated feature where Price has been binned into following categories:



* “Preference” of customers is based on “SUM (number\_of\_reviews)”.
* Various types of plots used:
  + Bar graphs
  + Stacked Bar charts
  + Scatter plots
  + Bubble charts
  + Maps
  + Box plots
  + Pie Charts