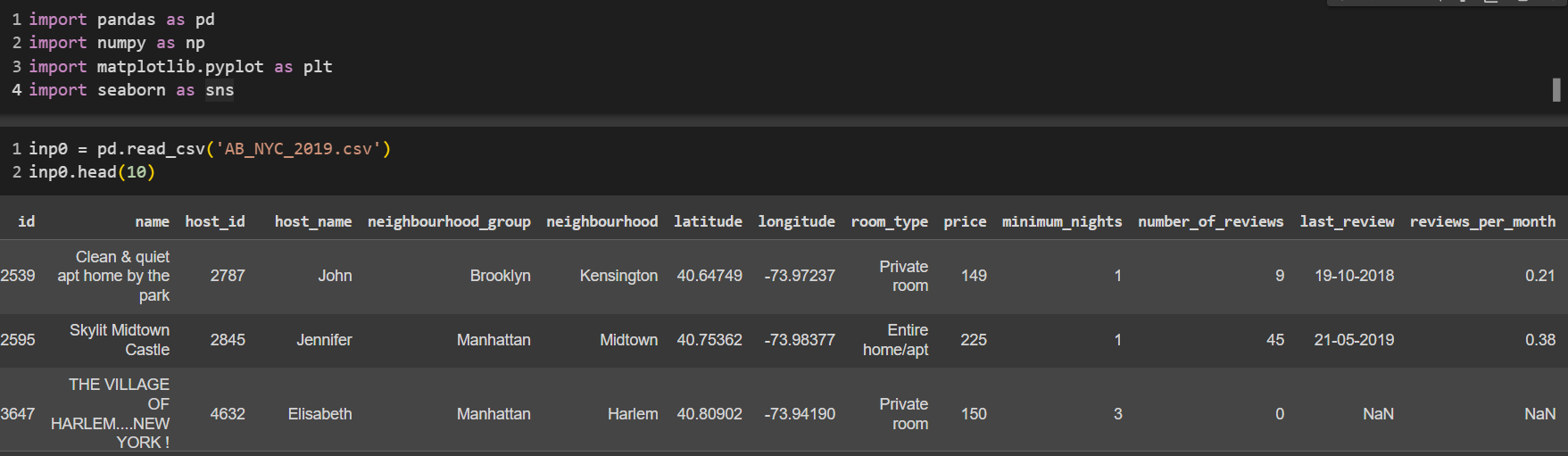
Storytelling Case Study: Airbnb, NYC

* Prashant Tiwari, Shreya Sunil Phade and Lipika Kandari

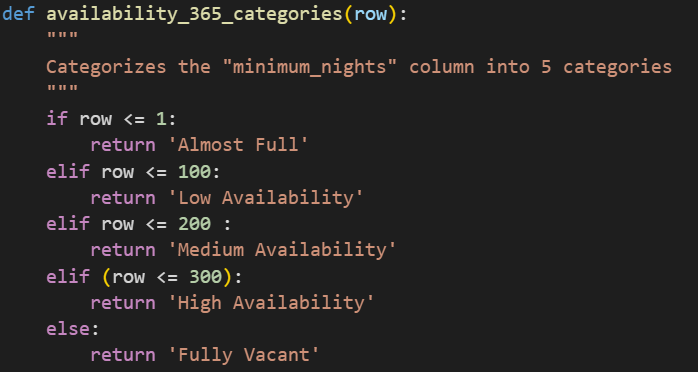
1. **Import libraries and read the data.**

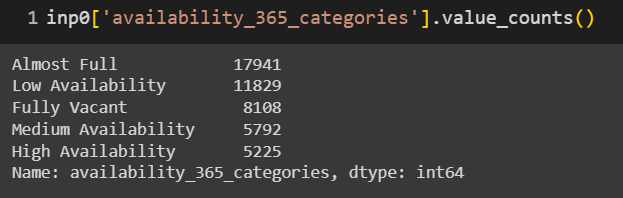


1. **Creating features.**

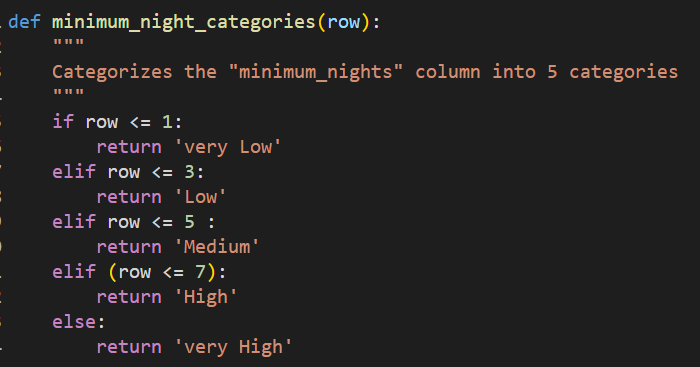
We can better grasp correlations and connections between items and explain our results by categorizing.

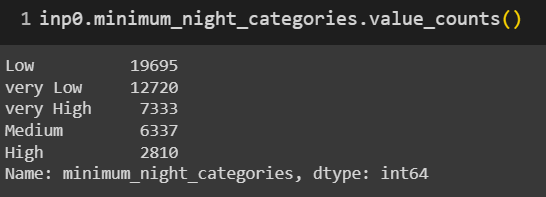
* 1. Categorizing the "availability\_365" column into 5 categories.



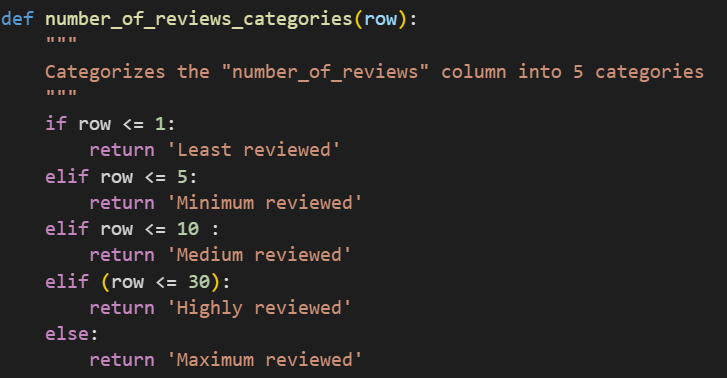


* 1. Categorizing the "minimum\_nights" column into 5 categories.

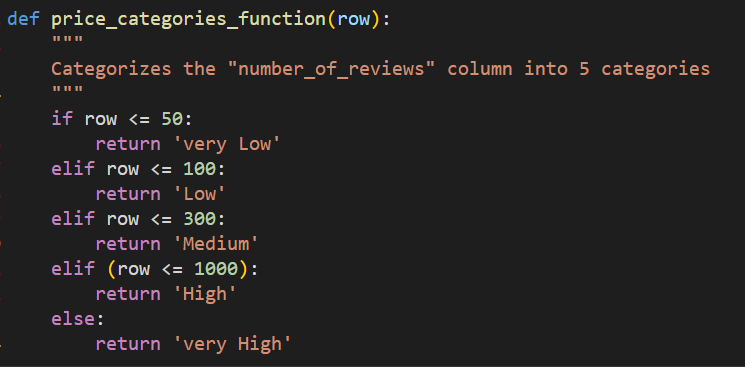




* 1. Categorizing the "number\_of\_reviews" column into 5 categories.

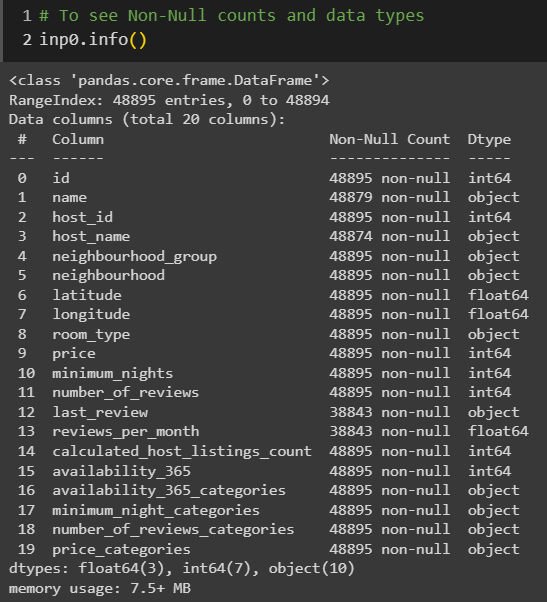


* 1. Categorizing the "price" column into 5 categories.



1. **Fixing columns.**

Check to see the non-null values and the data types corresponding to them.



Fixing reviews\_per\_month is necessary. It is a Dtype object. A suitable Dtype for this column is datetime64.



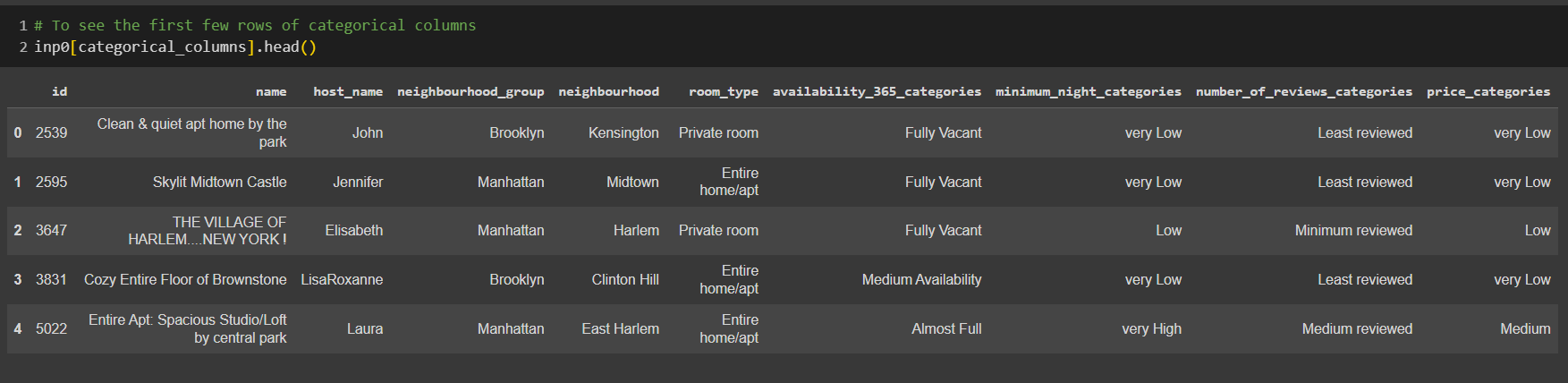
There are no longer any Dtypes that need to be repaired, and the data is free of errors like displaced columns that prevent it from aligning properly. Additionally, the columns required for the subsequent analysis are derived.

1. **Data types.**

Get the different types of data based on their data types.

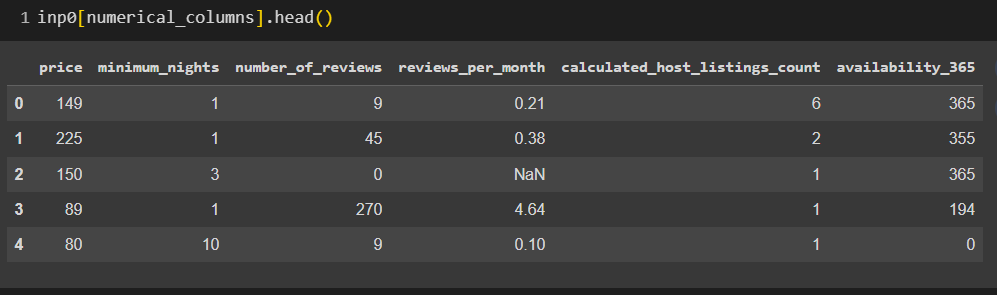
* 1. Categorical.



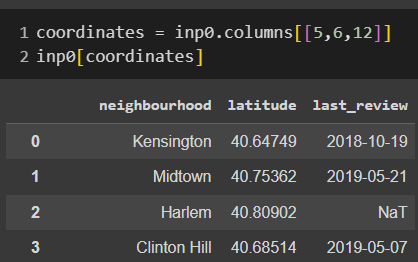


* 1. Numerical.

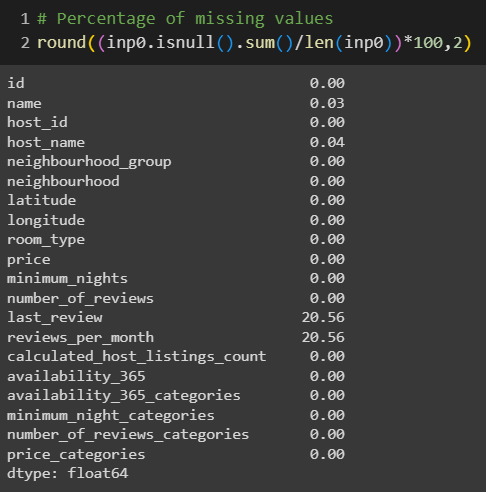




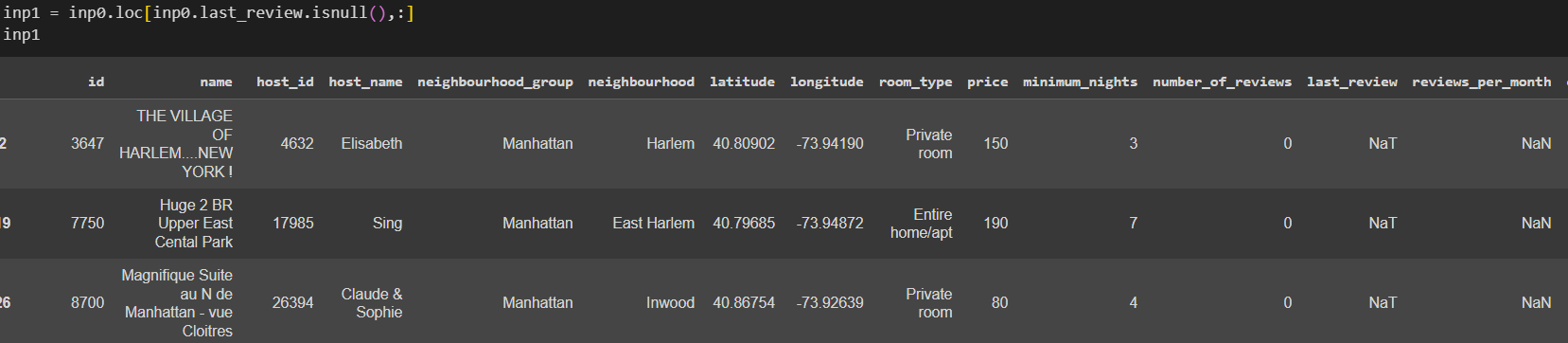
* 1. Coordinates and dates.



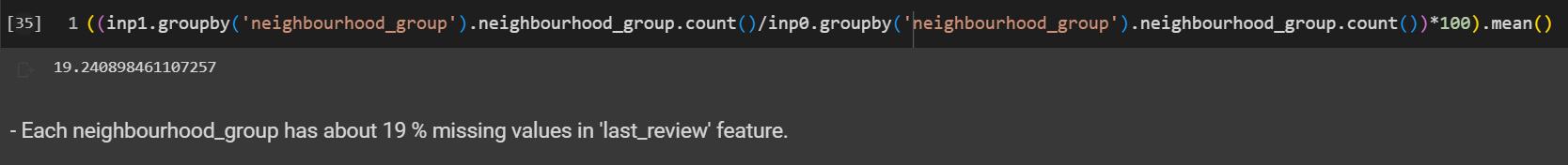
1. **Missing values.**

****

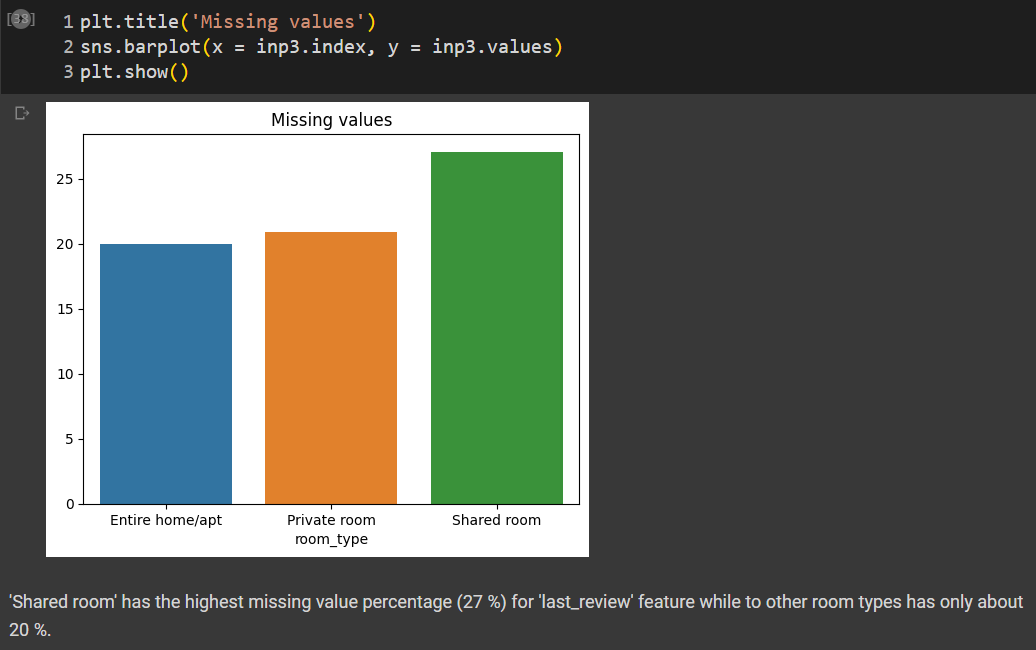
* About 20.56% of the values in two columns (last\_review and reviews\_per\_month) are missing.
* 0.3% and 0.4% of the values for name and host\_name are missing.
* Checking for Missing Completely at Random (MCAR) values is essential.If a value is absent but not at random (MNAR), the cause is independent of either of those two conditions. The absence of the value has a specific cause.
* We are simply evaluating the dataset and not creating a model, therefore there is no discarding or imputation of columns.The majority of the features are crucial for our analysis as well.
  1. Missing values analysis.



* 1. Missing values Analysis ('neighbourhood\_group' feature).

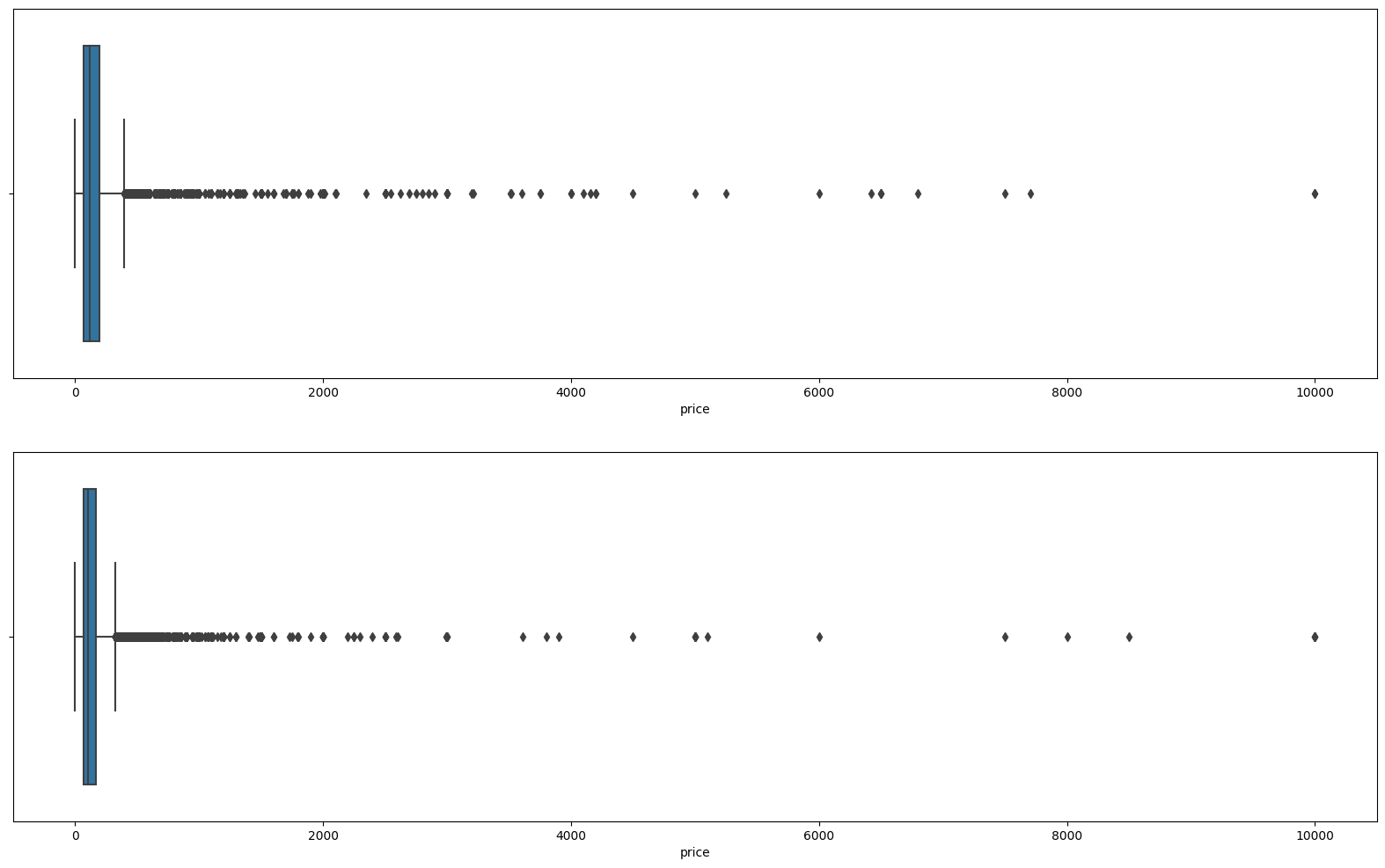


* 1. Missing values Analysis ('room\_type' feature).



The plot provided below reveals the following:

* When the 'last\_review' function is absent, the price is higher.
* Reviews are less likely to be posted for shared rooms - reviews are less likely to be posted when the costs are high.
* The analysis above appears to indicate that the missing values in this case are not MCARs (missing entirely at random).



1. **Data Visualization**

After performing above steps, we used the Excel file to create various visualizations through Tableau. There were no codes involved in Tableau only filtration was done.