**1. Unique value count**

One of the first things which can be useful during data exploration is to see how many unique values are there in categorical columns. This gives an idea of what is the data about.

# 2. Frequency Count

Frequency count is finding how frequent individual values occur in column.

# 3. Variance

When it comes to analysing numeric values, some basic information such as minimum, maximum and variance are very useful. Variance gives a good indication how the values are spread.

# 4. Pareto Analysis

Pareto analysis is a creative way of focusing on what is important. Pareto 80–20 rule can be effectively used in data exploration.

# 5. Histogram

Histogram are one of the data scientists favourite data exploration techniques. It gives information on the range of values in which most of the values fall. It also gives information on whether there is any skew in data.

# 6. Correlation Heat-map between all numeric columns

The term correlation refers to a mutual relationship or association between two things. In almost any business or for personal reasons, it is useful to express something in terms of its relationship with others. Finding correlation is very useful in data exploration, as it gives an idea on how the columns are related to each other

And one of the best ways to see correlation between numeric columns is using a heat-map.

# 7. Pearson Correlation and Trend between two numeric columns

Once you have visualised correlation heat-map , the next step is to see the correlation trend between two specific numeric columns.

# 8. Cramer-V correlation between all Categorical columns

Cramer-V is a very useful data exploration technique to find the correlation between categorical variables. And the result of Cramer-V can also be visualised using heat-map.

# 9. Correlation between two specific categorical columns

Once you have checked correlation between categorical columns using Cramer-V correlation matrix, you can further explore correlation between any two categorical columns. This can be done using a bubble plot between the two columns with size of the bubble indicating the number of occurrences

# 10. Cluster size Analysis

We live in a world with immense amount of data. It is very easy to get bogged down by data overload. In order to survive in this ever-increasing data world, we need to look things from a high-level perspective.

Grouping things together allows us to have that high-level perspective. Groups of data allows us to first look at the groups rather than individual data point. What would you prefer — looking at millions of data records or looking at few groups of data? The answer is obviously later as we humans prefer understanding in a top-down way

Data science can help us this amazing feat of creating few groups out of lots of data. In data science terminology, the process of grouping is also called clustering or segmentation. And making segments is an excellent data exploration technique as it gives an very good overview of data

As a first step in segmentation, it is useful to make an analysis of cluster size. The cluster size analysis shows on how can data can be split into different groups.

# 12. Outlier overview

Finding something unusual in data is called Outlier detection (also known as anomaly detection). These outliers represent something unusual, rare , anomaly or something exceptional. Outliers does not necessarily mean something negative. Outlier analysis helps tremendously to enhance the quality of exploratory data analysis

Outlier values in numeric columns can be obtained by various techniques such as standard deviation analysis, or algorithms such as Isolation forest. An outlier overview analysis gives overview of outliers in all numeric columns.

# 13. Outlier analysis for individual numeric column

Once you have checked which columns have very high or very low values, you can analyse individual columns.

# 14. Outlier analysis for multiple columns

One of the important step of exploratory data analysis is finding outlier based on multiple column (at row level). This can be obtained using various algorithms such as Isolation forest

A scatterplot is shown and outliers are marked in different colour (with label 1). The axis of the scatterplot is based on columns due to which the row is an outlier.

# 15. Specialised Visualisation

Till now most of the visualization you have seen are classic ones such as Bar chart, scatter plot etc... However during the data exploration it is very valuable to add some specialized visualization such as Radar Chart, Neural Network visualisation or Sankey charts

It helps a lot to understand the data much better. Radar chart can help in comparison. While Neural network visualisation can help understand what combination of columns could be important features or also to understand hidden or latent features. Sankey charts can be very useful in making path analysis