**Week#3 Exercise 3.2 Chart Ranking Justification**

**Data set** – The data set was used for this exercise was Housing Price Prediction, which is used to predict the house price based on several different variables associated with house such as house area, lot area, number of bedrooms, bathrooms, year built etc.

**Scenario** - The scenario for analysis was to see the relationship between house sale price and number of bedrooms. In order to see the relation between two variables scatter plot is the most useful and gives better understanding on how strong the variables are related and with direction of variance if any. Since the house sale price will differ based on number of bedrooms I wanted check here the relation by binning the sale price data and see if specific range of houses related with specific number of bedrooms etc.

For this analysis I have chosen three charts for visualization of sale price vs number of bedrooms. The carts selected are – **Scatter Plot, Bar Chart, and Violin chart.**

The justification is provided based on the ranking, the order of charts are from most effecting to least effective.

1. **Scatter Plot** - The scatter plots are mainly used to observe and show the relationship between two numerical variables and see the nature of the relationship. Here in this scenario the Sale Price and number of bedrooms are both numerical variables and our goal of analysis was to see the relationship between these two variables, hence the scatter plot is most effective chart to show the relation between variables. This plot also helps understand the how much one variable has affected by another variable and the scattered data points shows how strongly or weakly they are related based on the direction of data variables.

The scatter plot helped understand whether number of bedroom increases as the sale price of house or vice versa. Based on the plotted scatter plot we do not see any strong relationship as sale prices as number of bedroom but there is trend in few houses where the house price increase as the number of bedroom increases but few house having lower sale price with more bedrooms this may be caused because of old house.

Overall scatter plot is most effective to see relationship between two variables.

1. **Bar Chart** – Bar chart is used when we want to show the distribution of data to compare it with different values across subgroups of data. The bar charts also used to see the relationship between numerical and categorical variables. Basically each category is represented by one bar and the size of the bar is represented by its numeric values. Here in this scenario we do not have categorical variable so we binned house sale price into 8 different groups based on percentile and compared with number of bedrooms.

This bar chart showed that house with less price having less number of bedrooms and as the sale prices increases (bins) the number of bedrooms also increases and based on this we can see we see some relationship between sale price and number of bedrooms.

The reason bar chart is not ranked number 1 because we have grouped the sale price and our goal was to see the trend or relationship of each sale price with number of bedrooms and also the houses with number of bedrooms are not same in each group.

1. **Violin Plot -** The violin plots are used when we want observe the distribution of numerical data and compare the distributions between multiple groups. This chart helps understand the distribution of data whether it is normal distribution or uniform or bimodal distribution. Also this plot helps show density so we can see where more data points gathered corresponding other group of values.

Here, in this scenario we plotted violin plot by grouping Sale price and compared these groups with number of bedrooms. This plot showed us that lower house prices that is less than $532K having 2 bedrooms and greater than $532K house prices having 4 bedrooms, based on this we can guess that 2 bedrooms house prices usually less that 532K and 4 bedrooms house prices are more than 532K.

The reason we ranked violin plot at number 3 is because of it does not show clear relationship between two variables, it basically helps understand the distribution of groups.

After all, all of these 3 charts are helpful and having their own benefits when used appropriately based on what we want see and observe rom the data. Here analysis was to perform the relationship between two variables hence scatter plot is the most effective and ranked number 1.