Malbora Hajdarmataj

Milestone 5

Mat327

R link https://rstudiow.lehman.edu/s/dbb74cfc78a316af5880b/

GitHub link: https://github.com/Malbora88/ELIO/upload/main

From Milestone 4, I realized that column of season (this is a categorical variable, which can represent quantitative numbers,1 for Spring, 2 for Summer, 3 for Fall, 4 for Winter) and hour column (this represents the hour of the day, from 0 to 23. It is quantitative)

Season column.   hour column

```
mean(hour$season),        mean(hour$hr) [1] 11.54675
[1] 2.50164

median(hour$season)          median(hour$hr) 12
[1] 3

var(hour$season)            var(hour$hr)[1] 47.809
[1]1.225268

Sd(hour$season).           sd(hour$hr)[1] 6.914405
[1] 1.106918
```

In terms of the season column, we can observe that the average is around 2.50164 while the middle value (median) is 3. Based on the characteristics of the data, where seasons are represented by numbers ranging from 1 to 4, there is a tendency towards earlier seasons like Spring and Summer. This can be inferred from the fact that the average value's lower than the value indicating that a larger number of data points correspond to Spring and Summer compared to Fall and Winter.

When examining the column for hours, we see the average hour is 11.54675 and the middle value 12. These numbers are fairly like each other indicating that data is spread out evenly across hours of the day. There does not seem to be any bias towards one direction.

Standard Deviations:

When looking at the season column we can observe a deviation (SD) of 1.106918. This SD suggests that the data points for the seasons, which range from 1 to 4 are moderately spread out around the value. However, because this data represents categories numerically interpreting deviation in the way may not be as straightforward.

Regarding the hour column, we see a SD of 6.914405. Considering that hours range from 0 to 23 this standard deviation indicates a distribution of data points throughout the day. The closer the SD is to 0 the closer the data points are clustered around the value. In this case, while there is some spread in the hour column data points considering the range (0 23), it is not an extreme dispersion.

After analyzing the data, I noticed a bias towards seasons in terms of frequency in the season column with more instances falling under Spring and Summer. As for the hour column, it shows a distribution with observations spread across all hours of the day but slightly favoring midday hours. While there is variability indicated by deviation in relation to seasons values for hourly observations it suggests that activities or observations occur throughout various times of day without being heavily concentrated around any specific hour.