# DAVA – Truth or Myth?

1. The **first** belief I am investigating is: The higher the number of reviews, the better the review ratings.

(10 marks)

Conclusion: Based on data, this conventional belief is true.

Justification:

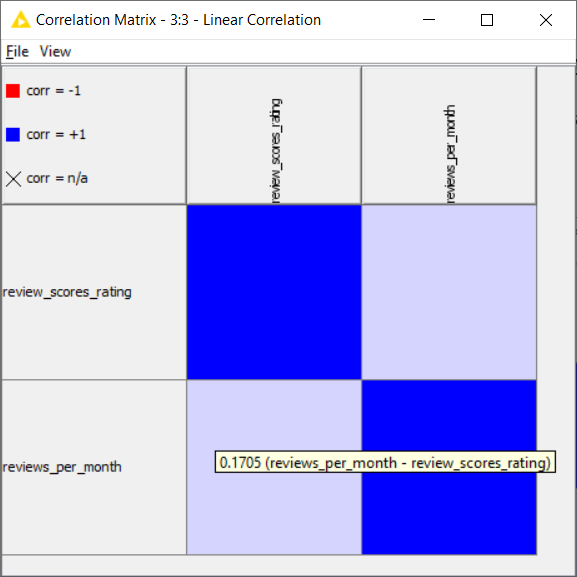
From the **linear correlation chart** is shows a weak positive correlation of 0.1705 between

review\_score\_rating and reviews\_per\_month.

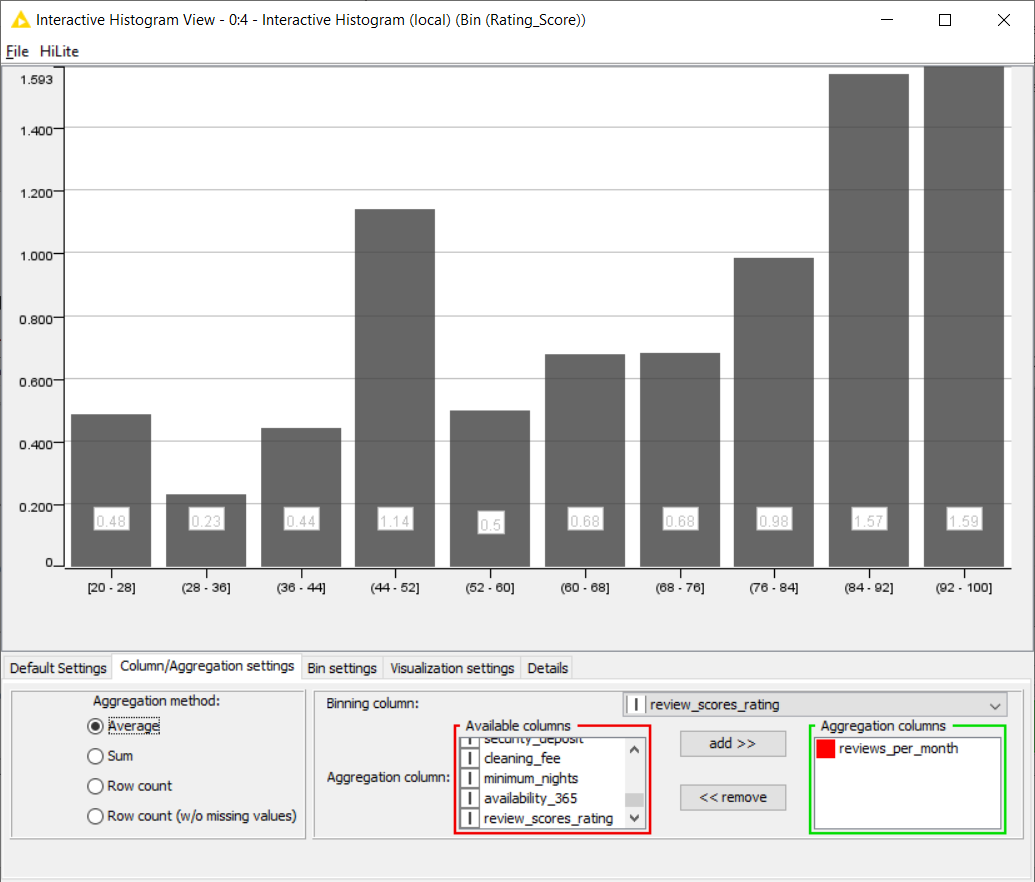
In the **histogram chart**, it generally shows there is a higher average reviews\_per\_month for higher review\_score\_ratings. However, it was observed that **bin 4 (rating score 44-52)** had a higher average reviews\_per\_month than some of the bins with higher rating score **(bin 5, 6, 7 and 8).**

The correlation is **positively correlated** though it is **somewhat weak** (See chart below)

**Correlation Matrix :** review\_score\_rating vs reviews\_per\_month



**Histogram Chart :** review\_score\_rating vs reviews\_per\_month

****

1. The **second** belief I am investigating is: **The central region has higher average listing price.**

(10 marks)

Conclusion: Based on data, this conventional belief is true.

Justification:

From the **Histogram chart**, it shows the CENTRAL region has the highest average price of $162.76.

From the **Box-Plot chart**, it shows the CENTRAL region has some extreme outliers. This will push up the mean listing price.

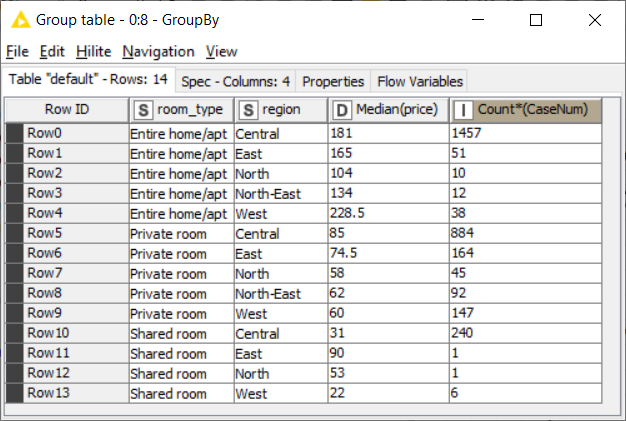
From the **GroupBy table**, beside showing the median price of each room type in each region, it also indicates the number of transactions recorded. The number of cases for entire home/appt in CENTRAL region is **1457** which has many more cases compare to other regions. This will push up the average listing price, if room type is not considered, since the price of an entire home/appt is normally higher than a room. Thus, it is important to take room-type into consideration when looking at the average listing price.

Some interesting observations were made from GroupBy table; the WEST region for entire home/appt ($228.5) has a higher average price than the CENTRAL region ($181). Also noted that CENTRAL region for shared room is not the highest, instead the EAST and NORTH region has higher price, but it is not meaningful comparison because they have only one transaction recorded (see Count(CaseNum) column).

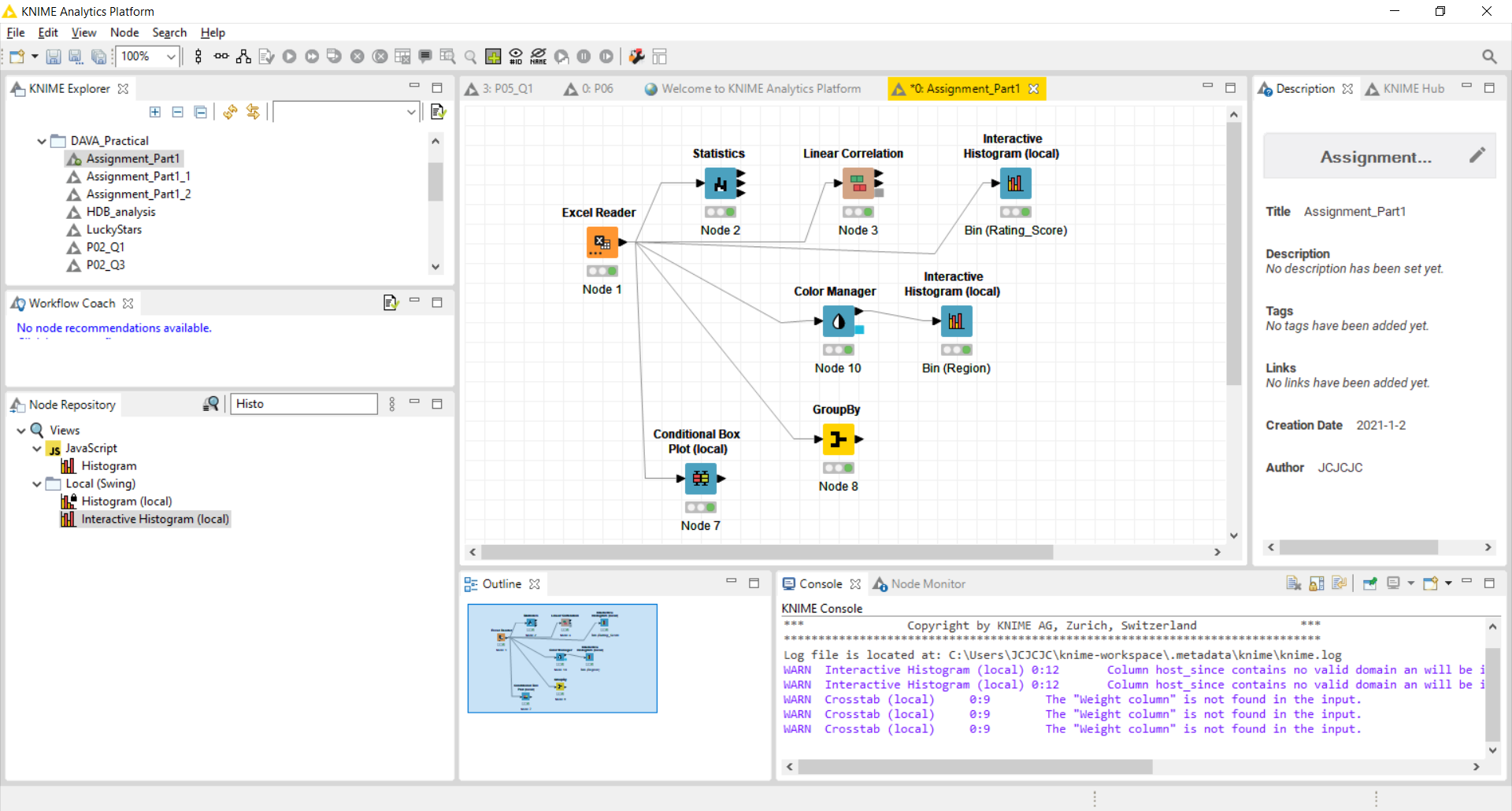
On general the CENTRAL region can command higher price, except for entire home/appt room type in the WEST region.

|  |  |  |  |
| --- | --- | --- | --- |
| **Histogram Chart :** Region vs Average Price | | **Box-plot :** Indicating quite a few extreme outliers in the central region for the transacted price | |
|  | |

**GroupBy Table :** Group by room type and region showing the median price and the number of cases transacted

****

# Screenshot of my overall KNIME workflow is shown below.



\*\*\*\*\* END OF PART 1 \*\*\*\*\*