

Assignment 2

Yuqi Zheng

AD 571 Business Analytics Foundations

Boston University

2019 Fall

Table of Contents

Executive Summary 3

Analysis of Data Process 3

Reference..... 5

Appendix 5

Executive Summary

In this report, the following question would be answered. I) how to import data into R-Studio; II) how to determine the average price of Sunnyside per year; III) how to filter data by removing fields with 0; IV) Comparing the result with other neighborhoods.

After analyzing, the average price of 1 square foot of residential real estate in SUNNYSIDE was lower than WHITESTONE at most of the years, except 2009 and 2013; also, was lower than ANNADALE at the most of times before 2011. ANNADALE was more stable than WHITESTONE and SUNNYSIDE on the whole. Although ANNADALE was stable, but its average price per square foot never reached to 300 from 2003 to 2016. WHITESTONE dropped intensely from 2007 to 2009 and increased profoundly from 2013 to 2014, but the average price of per square foot of WHITESTONE was the highest among three neighborhoods which is around 470. SUNNYSIDE changed sharply between 2005 and 2009, which means sales of SUNNYSIDE were fluctuating. But SUNNYSIDE's highest average price of per square foot among those years was around 300, which is lower than WHITESTONE and higher than ANNADALE. Overall, SUNNYSIDE, ANNADALE and WHITESTONE were increasing between 2003 and 2016. But WHITESTONE was more profitable than SUNNYSIDE and ANNADALE, especially in 2006 and 2016. Thus, I would recommend WHITESTONE instead of SUNNYSIDE and ANNADALE based on my analysis.

Analysis of Data Process

After installing packages on R, we need to connect SQL server by "met-sql19.bu.edu". Once we connect the server, we could find the data set on the connections. And we need to use "dbReadTable" to read all data tables from the server. We use "left_join" function to match the Neighborhood ID from the NYCHistorical data table and also need to match BuildingNameID.

And then we use "group_by" function to aggregate data by year. To filter SUNNYSIDE in residential status, use the filter function by Neighborhood name and Status. Since we need to determine the average price of 1 square foot of residential real estate in SUNNYSIDE for each year, we use "summarise" function to define the TotalSales and Total FT. TotalSales equals to the sum up of SalePrice and TotalFT equals to the amount up of GrossSqFt. Then use "mutate" function to create a column which is named "Average" by TotalSales divide by TotalFT.

For question three, we need to filter data by removing fields with 0 for price or square feet. To process this step, we need to define a new table (to keep clear from before process) and run the same process as above. The only difference is using filter function and use " $\neq 0$ " to filter out all the 0 numbers in GrossSqFt, SalePrice, and LandSqFt. After we remove the 0 numbers, we need to run the "summarise" and "mutate" function again. Then, for my neighborhood, SUNNYSIDE, the yearly average of TotalSales divide by TotalFT is created.

To compare other neighborhoods, I pick two neighborhoods, WHITESTONE and ANNADALE, randomly. And I do the same process as above. Then I get a table of the yearly average of WHITESTONE and ANNADALE too.

Finally, I use "ggplot" function to plot SUNNYSIDE, WHITESTONE, ANNADALE's yearly average and get a graph, see Appendix 1. The red line represents WHITESTONE, the brown line represents ANNADALE and the blue line represents SUNNYSIDE.

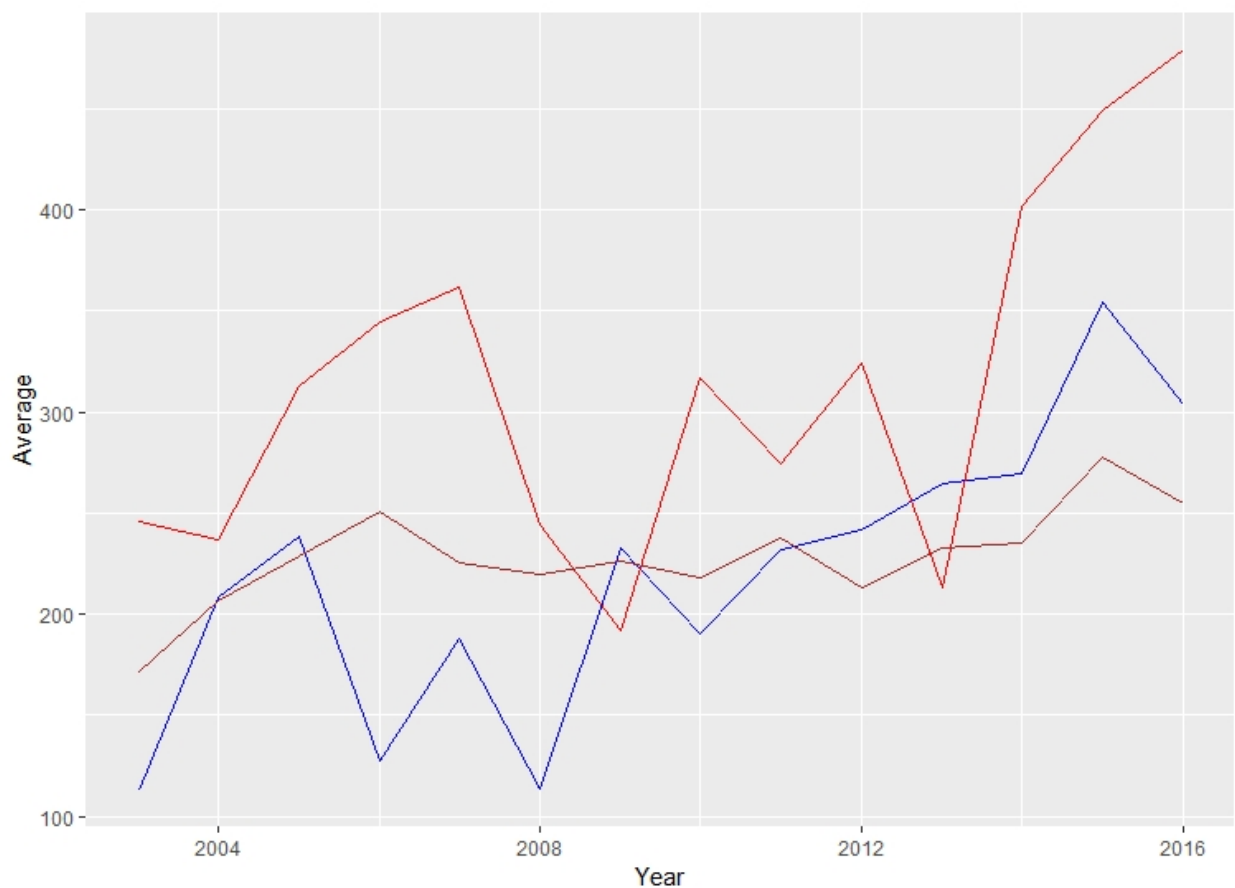
Reference

ggplot2. (0AD). Retrieved from

<https://www.rdocumentation.org/packages/ggplot2/versions/3.2.1/topics/ggplot>.

(0AD). Retrieved from <https://study.com/academy/lesson/mutate-function-in-r-programming.html>.

Appendix



Appendix 1