

BA222 Problem Set 1

Fall 2023

Due: Th, Sep 14 2023 at 12PM Eastern Time

No Late Problem Sets
(12PM is NOON not midnight)

Setup

Make sure you have Excel installed on your computer.

Download the NYC Property Sales data from Blackboard. Open the data in Excel.

Create a new tab to conduct your analysis and store your answers.

The glossary of terms for this dataset can be found **at this link**

Or if that link doesn't work:

https://www1.nyc.gov/assets/finance/downloads/pdf/07pdf/glossary_rsf071607.pdf

Writeup Instructions

Answer each of the questions below. Please type your answers into a document, using complete sentences. Submit that document as a PDF, and include a screenshot of your excel analysis as a page in the PDF. Please do not submit your Excel file.

Questions

1. Describe the data. What does each observation correspond to? How many variables are there? How many observations are there? What different types of variables do we observe? You do not need to be exhaustive, but write a few complete sentences.
2. What is the mean of the Year Built column?

Look at the Year Built column. What are the minimum and maximum values? Do these seem reasonable?

Sort your data by year built, and delete all 0 values. What is the mean of the remaining data? (Bonus: As an advanced move, there are many ways of looking at the mean of nonzero values without sorting or modifying the data.)

3. Using the COUNTIF function, count how many observations in the Zip Code column are from zip code "10011".

Look at a map of New York City on google maps. Where is zip code 10011?

4. What is the sum of the recorded sale prices? Format this value in a cell with commas and no decimal places. Why are some sales prices 0? Read the data glossary to find out.
5. Consider the total units column. What are the median number of units in a sale? What is the standard deviation of the number of units? What do you think this tells us about the distribution of the number of units? Write a sentence telling us about the distribution.
6. Now, plot a histogram of the number of the Total Units column. Give the histogram useful labels (X axis title, Y axis title, graph title)
7. How many commercial units were sold? How many total units were sold? What fraction of the total units sold were commercial units?