Final Project

2022-08-20

# MIS503 – Final Project Part 4

## Hagen, Tabitha

### Zillow Home Value Index Analysis

### R Setup

knitr::opts\_chunk$set(echo = TRUE)  
library(tidyverse)  
library(readr)  
SingleFamilyResidenceSales <- read\_csv("SingleFamilyResidenceSales.csv") # copy original file

### Relocation Home Value Comparison

#### You have been given a new opportunity to relocate for a new position within your company. They have given you the option of 4 different areas in the country (Chicago, Denver, Houston or New York).

##### a1. Based on your analysis, which city’s housing is most affordable?

Houston appears to be the most affordable in the Woodlands and Sugar Land areas. These are the outskirts of Houston and are not within the “Loop” so there is less density and more land to build retail establishments.

##### a2. Least affordable?

New York is the least affordable in the Newark and Jersey City areas which are actually in New Jersey. This area has a high level of public provided resources such as transportation, rent assistance, and food assistance. However, recently, those in that area have been moving more and more down south.

##### b1. Which cities saw the largest change in prices over the past 5 years?

In the last 5 years (as opposed to the past 5 years which are not on the graph), had Denver in the Aurora and Lakewood areas sharply increase in Home Value. I suspect that in the next few years, that will change as there is a serious multi-year drought coming out of that area and leading all the way to California.

##### b2. Which city has remained more consistent (i.e., no huge swings up or down in home values)?

Houston has remained the most consistent in home values. Even in the year 2005, with all the Hurricane Katrina refugees from New Orleans, their home values remained consistent.

##### c1. During the market downturn in 2012, which cities were most impacted?

In 2012, both Chicago and New York were affected negatively but Denver and Houston barely noticed.

##### c2.Which cities have recovered?

All the cities had recovered by 2018. I am sure that all these same areas are seeing some turmultous times now with the COVID lockdowns and housing issues seen after the lockdowns.

# create a new dataset that will be used for all commands after the pipe sign  
NationalHomeSales <- SingleFamilyResidenceSales %>% select (RegionName,   
 State,   
 Metro,   
 ends\_with("-05"))   
  
# Modify the variables names by taking off the suffix "-05" and removing NA's  
NationalHomeSales <- NationalHomeSales %>% # piping reduces redundant commands   
 filter(!is.na(RegionName)) %>% # don't use the na observations  
 filter(!is.na(State)) %>% # don't use the na observations  
 filter(!is.na(Metro)) %>% # don't use the na observations  
 rename\_at(vars(matches("-05")), ~str\_remove(., "-05")) %>%  
 # remove suffix from Variables  
 pivot\_longer(c('1996':"2018"),names\_to='YR',values\_to='ZHVI')%>%   
 # rewrite columns to rows  
 filter(!is.na(YR)) %>% # don't use the na observations  
 filter(!is.na(ZHVI)) %>% # don't use the na observations  
 # filter out homes in Illinois  
 filter(State=="IL" | State=="TX" | State=="NY" | State=="CO")   
 # filter out Chicago, Denver, Houston, or NY  
   
NationalHomeSales<- subset(NationalHomeSales, RegionName %in% c('Chicago', 'Denver', 'Houston', 'New York'))

# scatter plot of the info  
ggplot(NationalHomeSales,mapping=aes(YR,ZHVI)) + # x axis = YR, y axis = ZHVI  
   
 geom\_point() +   
 facet\_wrap(~Metro) +  
 labs(title = "City Home Value Comparison", x = "Year",   
 y="Home Values (in $)")+ # Graph Labels  
 # changes x-axis labels to display vertical values  
 theme(axis.text.x = element\_text(angle = 90, vjust=0.5, size=5))+   
 scale\_y\_continuous(labels = scales::comma) # labels get comma values

