DSC520 Week8/9 Exercise 8.3

Venkat Jagadeesh Jampani

February 12th 2022

Project: Impact of Covid 19 on Home Prices/IT Jobs/Migration in USA.

Introduction

Covid 19 Coronavirus disease 2019 (COVID-19) is a contagious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The first known case was identified in Wuhan, China, in December 2019. The disease has since spread worldwide, leading to an ongoing pandemic.

Symptoms of COVID-19 are variable, but often include fever,cough, headache,fatigue, breathing difficulties, loss of smell, and loss of taste. Symptoms may begin one to fourteen days after exposure to the virus. At least a third of people who are infected do not develop noticeable symptoms. Of those people who develop symptoms noticeable enough to be classed as patients, most (81%) develop mild to moderate symptoms (up to mild pneumonia), while 14% develop severe symptoms (dyspnea, hypoxia, or more than 50% lung involvement on imaging), and 5% suffer critical symptoms (respiratory failure, shock, or multiorgan dysfunction). Older people are at a higher risk of developing severe symptoms. Some people continue to experience a range of effects (long COVID) for months after recovery, and damage to organs has been observed. Multi-year studies are underway to further investigate the long-term effects of the disease.

Due to COvid-19, many people working in the Information technology, started to work remotely (WFH - work from home). That directly effected the demand for home prices. I would like to address the consequences of the COVid 19 on information technology jobs and the demand for the migration, work from home and the affect on the increase and decrease of Housing prices.

Research Questions

1. How are the pre covid housing prices in the different states in USA?
2. In which states the housing prices are impacted by Covid,worf from home, remote work.
3. What kind of homes are most preferred ?
4. To which states information technology jobs and people are migrating?
5. What are the average prices of different house models pre covid ?
6. What are the average prices of different house models post covid ?
7. What are the other factors impacting the housing prices in USA?

Approach

Analyze data to discover correlations and patterns and create a model to predict how COVID-19 affected the Job marked, demand for Work for home, and better office setup at homes, demand for large homes increasing in rentals, migration f people, increase and decrease in house proces etc.

1. Find the key predictors for the regression model.
2. Calculate R2, Adjusted R2, p-value, betas etc
3. Check confidence intervals if the model is likely to be representative of the true population.
4. Perform variance, residuals, leverage, cooks distance, covariance ratio analysis on the model(s).
5. Check if model is unbiased and use it for predictions.

How this approach addresses (fully or partially) the problem

This approach is based on the learnings in the course. Focus is to get enough data points to be able to address the problem. The predictors for the model and then the various calculations involved in the approach will give enough data points to analyze and draw conclusions on the correlations, model performance and if the model is unbiased to be able to use effectively for the predictions.

Data (Minimun of 3 datasets - but no requirement on number of fields/rows)

1. COid-19 cases, deaths across USA, both historical and latest data. Data Source: https://www.nytimes.com/interactive/2021/us/covid-cases.html

2. Hosing sales and prices in USA Data Source:
   1. https://www.realtor.com/research/january-2022-data/

   2. https://www.redfin.com/

   3. https://www.zillow.com/

3. IT Jobs and People migration across USA Data Source:
   1. https://www.dice.com/

   2. https://www.linkedin.com/

   3. https://www.monster.com/

Required Packages

The below packages might be required for data transformations, regression diagnostics, different statistical tests, plotting, visualizations and evaluations.

1. dplyr
2. purrr

3. QuantPsyc
4. lm.test
5. car
6. ggplot2
7. qqplotr
8. rcmdr

Plots and Table needs

1. Histogram
2. Scatter Plot
3. QQ Plot
4. Density Plot

Questions for future steps

1. How to ensure the data quality for the analysis ?
2. Can a different model be used for predictions ?
3. How to ensure the model that is selected is best fit for anlaysis?