PART 2- EXTENSION PLAN

Motivation/problem statement: **2**

Research questions and/or hypotheses: **2**

Data to be used: **3**

Unknowns and dependencies: **4**

Methodology: **5**

Timeline to completion: **5**

References: **6**

Motivation/problem statement

When covid hit US hard in March 2020, the economic downturn was quite evident that led to closure of many businesses when strict quarantine protocols and mask mandates were in place. This impact was also reported in US census Bureau's economic surveys[1]. The economic impact could be seen across many sectors like retail, wholesale and manufacturing. Many people lost jobs during this time and others scrambled to make their needs meet. But after the situation normalized a bit, many companies resorted to work from home policy to keep their wheels running. I want to assess this economic impact by backtracking the unemployment rate in Hamilton county during the covid times. Although the initial shock of covid had put many people out of jobs but later when things started to get normal and new work policies were brought to place, unemployment rate would also have improved. I want to weigh the impact of covid on the unemployment rate in Hamilton county and understand if we can make our jobs resilient to such pandemics or any disaster by implementing innovative work policies.

Secondly, I want to assess the impact of covid on the housing market. An average person would dream of owning a house in their lifetime and during covid, many people took this as an opportunity to make their dreams come true by leveraging the plummeting mortgage rate. As an individual, I would expect people having financial trouble to invest in properties at a time when many people were losing jobs. According to a report by Vox[2] , between sep 2019 and sep 2020, homeowners accumulated a collective $1 trillion in additional home equity. The S&P CoreLogic Case-Shiller National Home Price NSA Index, which tracks the housing prices, reported that the housing prices had risen by 9.5% in November 2020. While this came as good news for many homeowners, many American dreamers were completely shut out of the housing market.

The primary reason for the market boom could be the demand and supply problem. While the demand increased due to low mortgage rates, the supply couldn't match with the demand and hence the high housing prices. But still if one does the math correctly, the decrease in mortgage rate compensate for the increase in housing prices.Through this analysis, I want to track the market sentiments in Hamilton county during the covid.

Research questions and/or hypotheses

Research question:

- How was the unemployment rate affected during covid 19 and were all economic sectors affected equally?
- How was the housing market sentiment of people during covid19 when many people were facing financial burden and what segment of people benefitted from this boom?

Hypothesis:

- While the overall unemployment rate increased , I expect Hamilton county to show similar trends in unemployment rate.
- With the boom in the housing industry in the USA, the market hotness score will also increase in Hamilton county, Ohio.

Data to be used

I plan to use the data from St.Louis FRED database as my primary source. FRED stands for Federal Reserve Economic Database which consists of hundreds of thousands of economic data time series and it's free to use for non-commercial purposes. The Common analysis was around the trend of active confirmed cases of covid and as the cases were increasing , more jobs were impacted and this data would aid in answering my research question.The use of FRED data requires proper citation and some data which is owned by a third party, requires permission of the owners.

License terms and conditions- *https://fred.stlouisfed.org/legal/#full-fred-terms*

- **Unemployment time series of Hamilton county**- This time series data is available from the year 1990 to 2022 and contains the monthly unemployment rate. This data will be used to answer my first research question.

  Link to data - https://fred.stlouisfed.org/series/OHHAMI1URN

- **Market Hotness Index data of Hamilton county** - This time series data is available from the year 2018 to 2022 and contains the Market Hotness Index which is a score. This data will be used to answer my second research question. The MHI is defined by *realtor.com* as a score that reflects how the locals experience the fast moving supply and rising demand. It breaks down the demand and supply dynamics to rank the counties, zip codes as compared to the rest of the country. *'Realtor.com examines listing views by market as an indicator of demand and median days on market as an indicator of supply. '* The MHI may not necessarily indicate the rising and dropping of housing prices but could be used as a good proxy for it.

  Link to data - https://fred.stlouisfed.org/series/HOSCCOUNTY39061

My second data source is datausa.io. It uses the US government data which is cleaned by their team and is made free to access for public use. With the rise in covid cases, the housing market boomed owing to the decrease in mortgage rates and hence this data will help me to answer my second research question. The use of data is meant for informational and educational purposes and the content in the website is copyrighted by GNU Affero General Public License v3.0 (GPLv3) license.

License terms and conditions- https://datausa.io/about/usage/

- **Employment by different sectors** - I am planning to use this data to answer my first research question. It will help to check if all the sectors were equally affected or if some sectors were more impacted than others in the covid leaving some people in a more disadvantageous position than others. This is a monthly time series data indicating the number of employees in each super sector. This data is available for the cincinnati which is the major metro around Hamilton and I will use it as a proxy for Hamilton.
- **House ownership data**- This is a yearly time series data of house ownership percentage by different counties in the Ohio state. I will use this data to check how the house ownership percentage has changed during, before and after covid. I will try to find if a correlation exists between the Market Hotness Index and House ownership. This may or may not directly indicate towards the housing market sentiments of people during the pandemic.

  Link to data - https://datausa.io/profile/geo/cincinnati-oh/#housing

Unknowns and dependencies

For my first research question, the data for employment by different sectors is unavailable for hamilton county. So, the data for Cincinnati is used instead. For my second research question, I wanted to check if there is any increase in housing prices in Hamilton county due to change in demand supply dynamics during covid. But I couldn't find any direct data for housing prices. The closest data I could find was the Housing Price Index data for Cincinnati county and Market Hotness Index for Hamilton county. So, I have decided to use the Market Hotness Index data for Hamilton county. But if this data doesn't help me to address my hypotheses , I may use the Housing Price Index data for Cinicinnati as a proxy for Hamilton.

The data I have gathered is on a monthly level but the data used during the collective analysis is on a daily level. To compare the trends fairly, I will aggregate the daily cases to the monthly level and compare. While averaging out will smoothen the curve and I may lose important data points, I can assume that the covid pandemic in general impacted the unemployment and housing market as an individual event and a small period wouldn't be responsible for triggering unemployment rate and housing market conditions.

Studying the housing market conditions is tricky since we need to track the housing prices , availability of listings in the market for supply and maybe the data of people applying for home loans could be a good indicator of demand. The home loan data though is not usually publicly

available. Even if I just make assumptions based on the housing prices, it could lead to wrong inferences. I cannot say that housing prices increased because the demand has increased. What if the supply has decreased while the demand remained the same.

In any case, I don't have any of the above discussed data. I will be using the market hotness index data which takes into account the rising supply and demand to derive a score. But the only problem is that the indicator is relative. For example if the score of New York is higher than Houston, it means the market in New York is hotter than Houston. The caveat in using this is that if the market in New York becomes hotter in one year for some reason, that will affect the score of Houston. My assumption here is that during covid, covid was the only triggering event affecting the market. So any rise or decrease in the market hotness index can be attributed to covid. Hence, we need to carefully choose the period to track the market hotness index.

Methodology

After cleaning the data and ensuring the continuity of all the time series data, I will first use the Pearson correlation to check the correlation between unemployment rate and covid confirmed cases, and the market hotness index and covid confirmed cases. It will give me a preliminary sense if my hypothesis is right or wrong. To validate the results further, I will perform regression analysis keeping just one predictor at a time.

I will check the relationship between the dependent (unemployment rate) and the independent variable (covid confirmed cases) through the beta coefficient. I will also test the significance of the beta coefficients using the f-test and its associated p-value. I understand that any inference drawn from here cannot imply causation which is okay because I am just trying to find any potential correlation between the dependent and independent variables.

Null Hypothesis : The beta coefficient is not significantly different from zero.

Alternative Hypothesis : The beta coefficient is significantly different from zero.

I will repeat the same but with the other dependent variable i.e market hotness index. But any inference drawn from regression analysis would require certain conditions to be met and I am planning to make necessary assumptions if all the conditions are not met. Regression analysis is chosen since I am trying to model the relationship between one or more variables. It aligns with the objective of my hypothesis which is to understand if the covid induced recession drove the unemployment rate and the change in housing market sentiments.

For the purpose of presentation, I will present my results in the form of charts like the scatterplots to display the relationships between variables, normality and heteroscedasticity charts for showing that the assumptions of the regression analysis holds true. I am planning to show a univariate analysis of a few variables like employment by different sectors. I will also plot the different time series data that I am using for my analysis.

Timeline to completion

Data collection and cleaning- Collecting data from the sources described above and cleaning it for further analysis. - *November 16th 2022*

Exploratory data analysis- Plot time series and perform univariate analysis , correlation analysis - *November 18th 2022*

Regression Analysis- Regression assumptions check, regression model building and beta coefficients statistical tests. - *November 22nd 2022*

Presentation- Create visualizations and prepare presentations using insights from the previous analysis. - *November 24th 2022*

Documentation- Explain the process of answering the research questions and clearly summarizing the results. - *December 2nd 2022*

References

[1]https://www2.census.gov/library/publications/2022/economics/coronavirus-pandemics-economic-impact.pdf.

[2]https://www.vox.com/22264268/covid-19-housing-insecurity-housing-prices-mortgage-rates-pandemic-zoning-supply-demand.