Mojahid Osman, DATS 6401, Fall 2019

Final Technical Report

Natural Disasters effect in Housing Pricing in USA

**Abstract**

Natural Disasters is the one of the most threaten to the people live in US, the natural disaster include but not limited to hurricane, tornado, storm, high water, wind-driven water, tidal wave, tsunami, earthquake, volcanic eruption, landslide, mudslide, and snowstorm. In many cases the president of united states can declare an emergency when the president determines a federal assistance may needed providing emergency services, such as the protection of lives, property, public health, and safety. The impact of these disaster has a direct effect on the resident in different way, some these effects are economics implications or sometimes psychological or moral effects.

**Objectives**

The objective of this project is to study the behavior of these natural disasters and their impact on the US economy, and we will be focusing on the effects on Housing pricing. It’s very clear that some types of these natural disaster like hurricane, tsunami and earthquake has a major impact on the properties like houses and building, and the areas that declare as a potential disasters area has more risk than other and properties in these areas has least attract people and developers. In this project I am going to study this hypothesis by using a natural disaster data for all disaster declared by the president since 1953 created by (Federal Emergency Management Agency – FEMA). I am going to use data for only 10 years from 2007-2017 in this project and join the disaster data with data of Housing Price Index created by (Zillow – one of the major real state databases in US) also for the period of 10 years from 2007-2017.

**Functional Requirements**

The main requirement for this project is to give the users the ability to navigate and see different views of the data by slicing and dicing the data using filters and controls. The users should also be able to see the historical movement of the house pricing and the price trend over the time, we are also building a viz that help the users to see the coloration between the house prices and disaster declaration.

To achieve this requirement, first we need to work with underlying data and formatted in the way that gives this flexibility. First I am creating a details level which accommodate the low level of the data that includes the house price per counties for specific month and join this data with disaster data, and aggregated level which has the over view about the data and give the users the high level view of the house performance and the prices change on State and region level.

**Non-Functional Requirements**

In this section we are defining the look and feel of the web interface and the visualization of the project, included but not limit to colors, images, icons and all other type of design we are going to use to make it easy for the users to navigate and use the website. So I am going to use bootstrap CCS to implement all look and feel.

**System Architecture and Description**

I downloaded the FEMA data form “kaggle.com” in CSV format, the file has data since 1953, but in this project, I am going to use only the data from 2008-2018. The Housing pricing dataset which has all the information we need about house pricing per month for each county in the USA since 1996, the data was also in CSV format form Zillow website.

First, I loaded both file to MySQL DB to be easy for me to explore the data and do some data profiling and data dictionary for both files and, but after sometimes on that I found its easy to use Tableau Prep – Light ETL Tool comes with Tableau Desktop – I found it very useful in term of joining the two files and do data cleansing and formatting. The good things about the data prep you can save this data formatting in a MAP and use it later whenever you have data update in the future.

|  |
| --- |
| Map image |

**Development Platforms**

In this project I am building a web interface using HTML and CSS/JS and create different charts and visualization to represent the analysis, starting with high-level map of the US to see how the disaster data distributed among US states, and give the ability to drill down to see more information on the county level. I am still didn't decides how will implement hat but most probably I will be using Tableau as it has more flexibility with mapping and drill-down functionality. The underlying data will mange by Tableau Data perp, so whenever we got new data, we can just use the current logic we have in Tableau prep to generate the output file.

**Proposed Visualizations**

In the first page will have a high overview about the data project and summary of House pricing and disaster declaration per region and per years. We also have some charts to show different segment of data using Pie chart for Region pricing and Bar Char for the count of disaster.

For this high-level view, I am using D3.js to on the home page and have to aggregations numbers for. For the details information and all breaks down, I am using Tableau, because its more powerful interim of creating geographical chart (maps) filters and controls and give the user ability to get the answers for many questions using the same view as, Top 10 House Price State, Most effected State/County by disaster and too many other analyses.

**Experimental Analyses and Conclusions**

The data show that the house price in some state specially in East and West coast like (CL, FL,NY and WA) is really high compare to other states although these states are more hit by disaster. The

***Experimental Analyses and Conclusions***

*The data first revealed that there were significantly more counties that voted Republican than voted Democrat (over five-fold). The demographic factors that differed the most between Democratic and Republican counties were the percent of the population with bachelor’s degree and percent of the population that is white. The scatter plots for both demographic factors further revealed that differences in these factors become increasingly distinct as the margin of the vote increases. Particularly, as the vote margin of the Republican winning counties increase, so does the percentage of the white population. Age and poverty rates were other demographic factors with slighter Republican versus Democratic differences, with the average age of Democratic counties being lower and poverty rates being slightly higher. There was seemingly little difference between the ideologically different counites in terms of percent of the population with high school diplomas, median income, and health care insurance coverage.*

*While correlation does not prove causation, there are clear differences in racial makeup and college educated populations as it pertains to political ideologies, which may influence how people in this country vote. Interestingly, while health care is perhaps one of the most hotly debated political topics, Democratic and Republican counties share similar health care coverage rates.*