Project Report

My final project is focused on the housing market in the United States and how the walkability factor may have a potential impact on housing prices. The housing market is a seemingly endless topic of conversation among all Americans and as someone who is looking to potentially buy a house in the not too distant future this is of personal interest for me. I have three main questions I would like to address with this project. The first is, what city and what state had the highest housing price index since the start of the data (1975), and where has had the lowest housing price index. The second will be answering which city has highest walkability and which city has the lowest. The final question centers around whether or not walkability has an impact on the housing price index. In order to answer these questions I used two databases from data.gov. The first of which was the FHFA House Price Indexes (HPIs), and the second was the Walkability Index from the U.S. Environmental Protection Agency. I chose both of them as they were conducted by government agencies which made me feel confident in the validity of the data, and they both had a wide enough scope of the entire country to make me feel like there was enough information to work with. The housing price index measures the changes in single family home values for all 50 states, 400 cities and extends back to the mid 70's. Again this made me feel confident that this data had the depth needed to get some real information from. The walkability index focuses on the likelihood of an individual to travel by walking due to city environmental characteristics. I felt these two data sets complimented each other extremely well and I had always assumed the higher walkability of a city the higher housing costs would be(thinking of NYC in particular) and this felt like a good opportunity to test that hypothesis. The first step to this project in terms of python was to install the proper packages, and then was to create the connection to my digital ocean SQL database with mysqlconnector. To start analyzing I decided to use pandas (obviously) to manage the data and holoview/ hvplot to create the interactive interface that will allow a user to make any graphs they want. This was achieved by using the hvexplorer option and allows full user customization. I went for this method because it felt the most efficient to me, as the HPI data source is extremely large and instead of having many many graphs I thought it would be best to allow the user to create just about any graph they would want with the hvexplorer option. After that I ran bar graphs to highlight the max and min HPI value by state, and by city. To do this I used matplotlib.pyplot and ploty and created a query that would return full state names, and then city names based on using a wildcard for state abbreviation. It had to be grouped by max and min in order to make the graph work, and go-figure was used to create the figure of the graphs that answered the first question posed. The next step was to ingest the walkability data. After trying and failing to use Arcgis to generate an api call to the walkability index data, I had to simply ingest the csv file into my jupyter notebook. After referencing the csv file and creating the data frame, by reading the csv, I had to group he walkability data by CSA_Name and filter to the max, to get the walkability index. I then created a figure and gave it hover features and then called it to show. I used the plotly and pandas packages to accomplish this. I recorded the high and low ends and then went ot create another hvexplorer plot for the to allow the user for full interactive control of walkability figures. I personally think in it's current state I will need to incorporate more data sets to even attempt to try to predict housing prices. I think this is expected as there are entire industries dedicated to this practice. I think it is best to view my dashboard as an up close zoom view of two factors and

it would require zooming out or adding different data sets to attempt to get more information or the whole picture.