Happiness Index: a neural network model of American cities' happiness

By: Samira Azhar, Varsha Kalangari, Lauren Kroll, and Colleen Palmer

Abstract:

This study uses the World Happiness Index and WalletHub's rankings of US cities' happiness levels as a benchmark and comparison of demographic, statistical, and economic data. From this, a Happiness Index will be created to decide which major American metropolitan city is happiest, along with the creation of a numerical ranking. This data will be used to analyze the most important factors that contribute to the overall happiness rating. The information gathered will then be used to rank major cities- Chicago, Houston, Los Angeles, and New York- from different regions of the United States based on the happiness of their inhabitants.

Introduction:

Happiness is a very important aspect of every human's life and well-being. The state of being happy is necessary to live, but it can be derived from different values within a person's life. Whether it be wealth, family, or power, the emotion is very personal. Different external factors, such as the economic environment and social atmosphere also influence happiness. The many variables make the term difficult to define, but it is so influential to everyday life. Through this data analysis project, the goal is to find a deeper understanding of happiness in a broader sense of the word. It is one of the most subjective aspects of life, yet it is the driving force and motivation behind everyday actions. Happiness is determinant upon personal preferences; however, those preferences can be heavily swayed by the environment around oneself.

The implications of the research are abundant and far-reaching as our research will provide a qualitative association for something as subjective as happiness. By analyzing which major metropolitan city is the happiest, this study can build off our conclusions to find actionable ways to make the quality of life better in other metropolitan cities across the U.S. The happiness index rankings can be used to identify what specific factors or variables make a city "happy," and in turn, individuals can look toward improving these specific factors in their own areas. Placing qualitative measures on happiness is beneficial in that it provides concrete ways to improve overall well-being. This, in turn, has ripple effects toward creating a more prosperous and productive society. Taking the unconventional choice to measure happiness provides individuals with a unique perspective and starting point for improving quality of life as happy cities lead to more opportunities; both in terms of personal well-being and overall growth for the community.

Cities can identify the factors that primarily affect overall happiness of their population in order to discern where they may be lacking in resources. For example, a city may implement new social programs or laws in order to mitigate problems in their city that are discovered to adversely affect happiness. This research can serve as a guideline for improving the lives of those living in a specific geographic region through concrete legislation.

Data Set:

The World Happiness Index has analyzed numerous countries around the world, using an assortment of factors and variables to determine the happy from the unhappy. On a smaller scale, WalletHub's "Happiest Cities in America" article demonstrated how geography, even within a country, can play a major role in the well-being and happiness of its inhabitants. From these articles, the major economic, demographic, and environmental factors were used to create a new happiness index. It was modeled off The World Happiness Index and used WalletHub's study as a control group. From this, external factors for WalletHub's top 100 US cities were compiled. Exhibit 1 depicts the first 25 cities on our list. From this, you can see the inputs used: median income, poverty rate, median age, etc. This was the data set used in SPSS to model a U.S. Major City Happiness Index.

A noticeable feature in the study, however, was the prominence of small towns and cities within the happiest. Living in the major metropolitan cities in the United States can bring about stress, fear, and feelings of unworthiness due to comparison. From the SPSS models, a case will be built to determine the happiest of America's four largest cities: Los Angeles, Houston, Chicago, and New York. Although all four cities are major metropolitan areas, each is composed of a unique culture that may impact the overall happiness of its inhabitants. By studying these areas, regional differences in overall happiness can be determined by analyzing the social, environmental, and economic climate they reside in.

Analysis:

In SPSS Modeler, there were four different data analysis approaches: association analysis, neural networks, cluster analysis and a CHAID decision tree. The neural network modeler was the most effective model in analyzing our set of data. Compared to the other three models tested, the neural network had the highest accuracy. The output determined by this model was most relevant to our desired conclusion.

Cluster Analysis: The dataset was run through the cluster analysis model in order to determine if it yielded relevant results. This model was considered because of its ability to identify clusters with common characteristics within our collected data. The cluster analysis could have identified groups of higher ranked cities, and which factors were similar within that cluster. On the other hand, it could also identify factors common among the lower ranked cities. For example, if there was a common factor of low poverty rate in the higher ranker cities, the cluster analysis may have grouped these cities together, and an important input could easily be identified. However, when the data was run through the cluster analysis model, it only identified two clusters. This was not effective in determining common characteristics within the group because it was too broad. A narrower scope was needed in order to see why certain cities ranked higher than others.

Association Analysis: Association Analysis identifies items that occur together. This model did not work for the data because it did not have simply categorical data. For this model to work, it would need data that is yes/no or fail/pass.

CHAID Decision Tree: The CHAID analysis creates a decision tree to determine how variables best merge together to predict the dependent variable. In our case, our dependent variable was the city happiness

rankings and we were looking to see what variable or combinations of variables best come together to define a happy city. We were looking to identify the top predictors of happiness. For this specific model it was found that median income has the most significant influence on happiness, followed by the poverty rate, as seen in exhibit two. In addition, the CHAID model produced multiple nodes that give us more ways of looking at how to section the data. While the CHAID model did provide interesting insight, it was not the most effective choice based on our dataset.

Neural Network Model: The first phase of analysis centered on WalletHub's study regarding the "Happiest Cities in America." In order to analyze the rankings given by the study, various socio-economic indicators for the 100 cities listed in the rankings. When run through SPSS modeler, the neural network model was the best fit as it clearly laid out the different independent variables that influenced the relative happiness in all of the cities. Based on the predictor importance output as seen in exhibit three, the model found that median income and commute time are the two highest influencers of happiness. More specifically, cities with higher median incomes and shorter commute times are more likely to be ranked higher in terms of overall happiness. The predictor ranking output was very significant as the two most important variables are both qualitative quantities, thus bringing more structure to the qualitative measurement of happiness. Leaders in many of these areas can now take actionable measures, such as, increasing the minimum wage or making the city more commuter-friendly in order to increase overall happiness.

In order to further test the validity of the model, a plot node was run against the actual happiness rankings compared to the predicted happiness rankings. The results from the plot, as seen in exhibit 3, show a nearly straight line, which indicates that the various socio-economic factors inputted in the model to predict happiness were consistent with how the rankings established by the WalletHub study. The validity of our model, as indicated by the plot, allowed us to expand our analysis to cities outside of the WalletHub study.

Conclusion:

Through the analysis of these factors, it was determined which major United States cities hold the happiest citizens. Despite the fact that only two of the four chosen cities were included in the original study, the results from the data analysis allowed an application to the remaining two cities. The data analysis proved that the ranking of the four most populated metropolitan areas in the United States, from happiest to least happy, is as follows: (1) Los Angeles, (2) New York City, (3) Chicago, (4) Houston. Not only does this information provide insight into the cities themselves, but also what input factors Americans prioritize for their happiness levels.

This Happiness Index, narrowed and focused on the United States, allows for a better understanding of the nation as a whole. The model can be used to inform and educate both locals and tourists alike. It provides insight into the less recognizable aspects of each major city, while also explaining the values and desires of the inhabitants.

Meaning: A definitive ranking of the happiest cities in the United States was critical to this study. Data from 2018 was used in the model, because it provided a complete set of metrics and figures relevant to the desired result. The WalletHub study ranks 182 cities in the U.S. in terms of their

happiness. In the research case, analyzed the top 100 happiest cities in the U.S. With these observed cities, data was compiled on socio-economic, physical, and environmental factors to the corresponding cities. Some examples of factors include median income, median age, divorce rates, and poverty rates. The goal was to identify which factors were most critical to the happiness level in each city. Since a definitive ranking of the happiest US cities in America was used, the factors that were most commonly present in the higher-ranking cities, such as Plano, Texas, the happiest city in America, could be observed.

Application: There are a myriad of applications to this study of making sense of the Happiness Index for major cities in the United States. The results, which make sense of the happiness rankings provided by WalletHub, provides a quantitative measurement to the idea of happiness. This study has identified concrete and tangible ways to improve overall well-being, such as tying healthcare spending and unemployment rates to overall happiness. It is necessary for a city to prioritize the happiness and overall well-being of its citizens, as these factors directly contribute to a more productive society. Cities, using this data, can work to improve certain elements, such as the poverty and unemployment rate in order to increase happiness. This study has allowed something that was thought to be unquantifiable (happiness), to be broken up into quantifiable areas in order to improve the overall community.

Benefits: The Happiness Index for major metropolitan areas in the United States allows for a better understanding of the country's inhabitants. While there are certain aspects of life that are expected to inherently make people happier, such as a well-paying career or a big home, this study exposes the weight these, and other factors, may carry on one's happiness. The Happiness Index is applied to the four largest cities in the United States, whose locations are scattered into the overarching demographic regions of the country (East, Midwest, South, West). This Index allows us to determine which region holds the happiest city, while understanding the factors behind the "why" of each place. The major benefits from this research stem from more perfect information. Whether it be finding a place to move, to visit, or to leave, this Happiness Index shares the morals, values, and desires of each major metropolitan U.S. city.

Limitations: The Happiness Index calculated through the data mining techniques considered many variables. However, there are infinite factors that can affect Happiness Levels. The data set considered used economic, demographic, and educational inputs. There are many external factors not considered, and this could implement implicit bias to the data found. The Happiness Index calculated had a high accuracy rate, so the direct impacts of including more inputs may not be extreme, but they are still an important consideration. The Happiness Index also used data from American cities listed on WalletHub's list. This list limited the scope of the study to the 100 cities studied. This study may have followed WalletHub's data sampling and analytical procedures, thus implementing their biases into this data as well. Through due diligence and analysis, WalletHub's data appears to be credible and to maintain the integrity of the data.
The data sets used to analyze each individual city's happiness also may have some limits. The data used was from 2014-2019. While this data was most recent to study, it is also a time of limited political

and economic turmoil. External events, like an economic downturn or time of war, impact cities differently based on their individual characteristics. This data cannot be used to generalize during events similar to those. This study can only represent a range of years after the study with a similar economic, demographic, and educational climate that the study took place.

Implications/Recommendations:

Within the analysis done, more analytical methods could have been conducted to analyze the inputs on a micro-level. For example, an A Priori analysis could have been beneficial as a second neural network. This would help determine the most accurate model to use for this study. In addition, the data could have been used to compare the results from the CHAID neural network. This could help discover sensitivities within the inputs. A cluster analysis was done to determine common factors within the happiest cities, but more analysis could have been done here. This could also be done with cities that are not considered to be happy. This would give more insight on the factors that cities can utilize to improve the wellbeing of citizens.

Like any other study, a larger sample size facilitates generalizations and extrapolation of data. If more cities were studied, it would help with the generalization of all US cities. This could be done by sampling different cities using different techniques: dividing cities into different regions or basing samples off of population distribution.

Creating an index with more inputs would allow for more factors to be analyzed for a higher accuracy. Beyond this, data from years previous would give this Happiness Index more predictive power. This study included data from 2018 only. Future studies into this topic may utilize the existing data and add more years and cities into the dataset. This will reinforce the results of the model. As mentioned in the limitations section, this study was using data from a stable environment, economically, politically, and environmentally. It would be beneficial to compare data from periods of recession or disaster. This insight would help predict the future happiness of these cities, as these events will occur in uncertain times.

The model indicates that median income is the most influential factor on happiness. Commute time is the second most important factor in determining happiness. A lower commute time and a higher median income correlate to a higher level of happiness within the observed urban areas. Therefore, legislative entities in U.S. cities should focus on these measures in order to improve the overall happiness in their respective areas. To improve commute time, city governments should implement policies to create easier access to mass transit for individuals. In addition, improving the condition and quality of these mass transit options will encourage widespread use. This can decrease the commute time of individuals, and thus improve their happiness. Median income highly impacted happiness, so further information can be gathered on discretionary income. This can help economists and politicians decide proper legislation catered to their city's needs through increasing wages or subsidizing necessary expenses in order to increase median income without increasing the cost of living. Overall, many of these factors can be used to micromanage a city through legislation.

References:

https://www.tuck.com/best-worst-us-cities-sleep/#Huge-Table

https://datausa.io/profile/geo/plano-tx/?compare=irvine-ca

https://www.census.gov/library/visualizations/interactive/marriage-divorce-rates-by-state.html

https://wallethub.com/edu/happiest-places-to-live/32619/

https://ggweather.com/camelot.htm

Exhibits:

Exhibit 1

City	Median Income	Poverty Rate	Median Age (in y	Median Property	Commute Time (in	Sleep Rate	Health Care	Spe
Plano, TX	$ 93,012	7.03%	38.3	$ 341,100	24.4	64.66	$	6,9
Irvine, CA	$ 95,573	13.20%	34.4	$ 753,400	23.3	77.03	$	7,5
Madison, WI	$ 70,463	11.70%	36.4	$ 258,700	20.5	85.78	$	8,7
Fremont, CA	$ 122,191	4.85%	37.7	$ 785,700	31.7	70.83	$	7,5
Huntington Beach, CA	$ 88,079	8.94%	41.9	$ 668,700	27	76.25	$	7,5
Fargo, ND	$ 50,561	13.90%	30.3	$ 190,200	14.6	78.32	$	9,8
Grand Prarie, TX	$ 62,589	12.60%	32.9	$ 139,600	27.8	47.93	$	6,9
San Jose, CA	$ 113,036	10.00%	37	$ 968,500	29.7	67.17	$	7,5
Scottsdale, AZ	$ 80,306	8.91%	46.9	$ 433,500	19.8	74.15	$	6,4
San Francisco, CA	$ 112,376	11.70%	38.3	$ 1,200,000	30.9	69.55	$	7,5
Bismarck, ND	$ 61,477	9.52%	37.1	$ 223,600	17.3		$	9,8
Overland Park, KS	$ 78,217	5.79%	37.9	$ 248,100	19.2	85.14	$	7,5
Santa Rosa, CA	$ 67,144	11.80%	38.1	$ 655,200	23.5	67.19	$	7,5
Austin, TX	$ 71,543	15.40%	33.6	$ 365,500	22	66.87	$	6,9
Sioux Falls, SD	$ 56,714	11.10%	34.3	$ 168,600	16.4	80.08	$	8,9
Pearl City, HI	$ 91,122	4.61%	40.9	$ 626,300	28.6		$	7,2
Glendale, CA	$ 58,657	15.60%	41.2	$ 675,300	26.3	57.55	$	7,5
San Diego, CA	$ 79,646	14.50%	35.4	$ 654,700	22.5	67.54	$	7,5
St. Paul, MN	$ 59,033	20.40%	32.6	$ 216,100	22.5	76.72	$	8,8
Charleston, SC	$ 61,367	14.60%	34.4	$ 286,200	20.7	67.54	$	7,3
Gilbert, AZ	$ 87,566	5.87%	33.6	$ 286,400	25.6	94.9	$	6,4
Anaheim, CA	$ 76,154	16.00%	34.4	$ 606,600	27.8	56.22	$	7,5
Raleigh, NC	$ 65,695	14.00%	34.1	$ 268,900	22.5		$	7,2
Cape Coral, FL	$ 53,653	12.70%	45.9	$ 196,800	28.4	57.22	$	8,0
Cedar Rapids, IA	$ 56,828	11.40%	36	$ 138,500	17.2	78.38	$	8,2

Exhibit 2

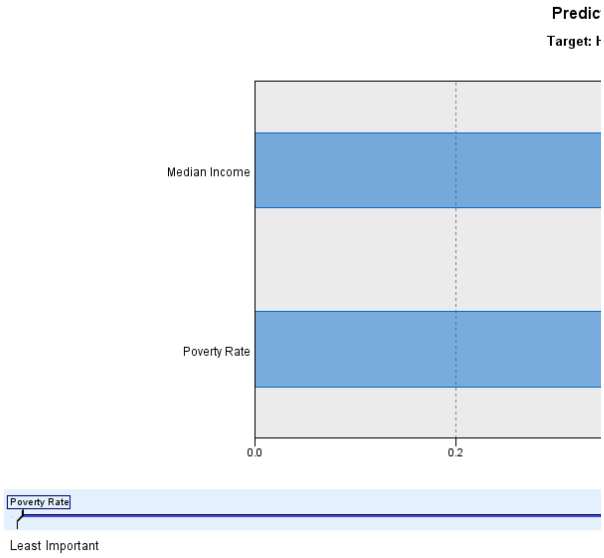

Exhibit 3

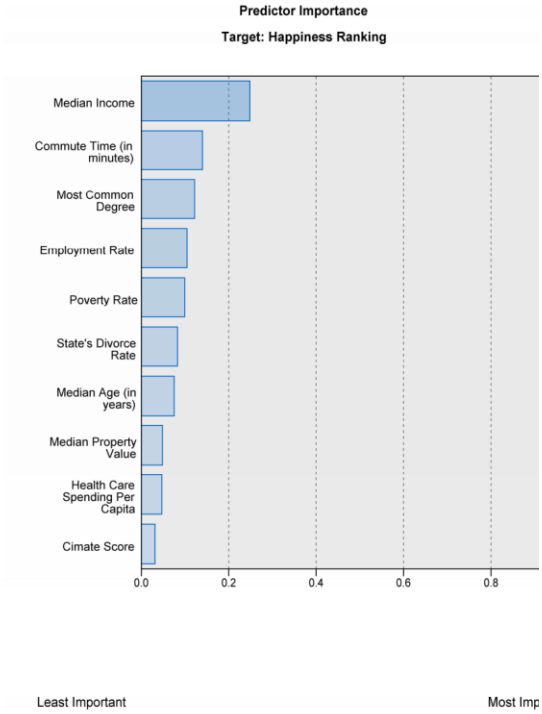

Exhibit 4