Factors associated with the Asian American *Quality of Life* in Austin TX

Contents

Purpose

The Report on the Asian American Quality of Life is the outcome of a survey conducted from August to December 2015. Self-identified Asians aged 18 and older living in the Austin area were eligible to participate. The survey aimed to include representative samples of the five largest Asian groups in Austin (Asian Indian, Chinese, Vietnamese, Korean, and Filipino); however, other Asian groups were also included. The survey was done by a research team in the School of Social Work at the University of Texas at Austin for the City of Austin.

As a broad racial category, Asian Americans are the fastest-growing minority group in the United States (U.S. Census Bureau, 2012). **An analysis of the factors that are associated with Asian American's opinions on their "Quality of Life"** in this survey could be leveraged by the community to identify if there are areas that can help to improve this quality in Austin.

Data Set & Variables

The dataset contains 2601 observations, each representing an individual response in the survey, and 231 variables. This makes it a fairly complex dataset from the perspective of having a large number of predictor variables. Before any data mining techniques are applied, the data set will need to be cleaned to address missing or incorrect data.

As shown in the referenced survey sample, answers where requested for 7 broad topics:

1. **Demographic information** such as Age, Marital status, employment and income
2. **Immigration/Acculturation** such as fluency in English and familiarity with American culture and customs
3. **Health information** such as overall health, present conditions and insurance

4. **Emotional well-being** such as a rating for overall Quality of Life on a scale of 1 to 10 (this is our response variable) and opinions on mental health
5. **Special interest health topics** such as Alzheimer's and Advanced Directives awareness
6. **Social and Community resources** such as family or attendance of religious activities
7. **Life in the city of Austin** such as opinion on Austin life and public services used

The data collected from questions in the first 6 sections will be used for data mining as section 7 is just inquiring about Austin's services and usage.

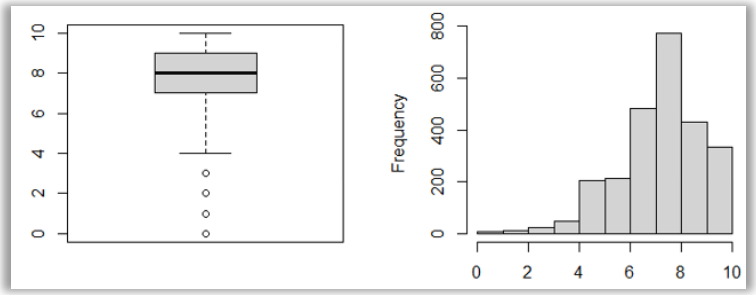
Data mining techniques

Two data mining techniques were selected based on their ability to find associations within the variables and predict a response variable when a large number of predictors are involved.

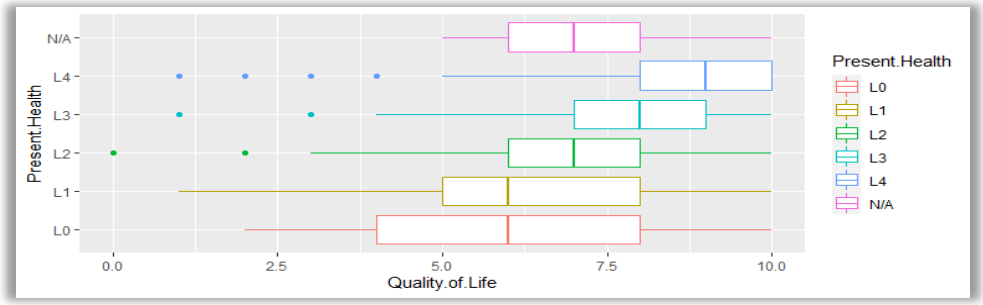
1. **Association Rules**: This unsupervised learning approach will help discover patterns of co-occurrence in the large data set by identifying entities that frequently appear together in a group. Patterns found will be summarized by association rules, which will provide those conditional probabilities associated with specific selection of *Quality of Life* in the survey data set.
2. **Decision Tree – Random Forest**: It makes sense to use this supervised learning method for 2 reasons - there could be strong correlations between predictors (i.e. questions in the survey) and secondly a subset of predictors could be much more informative about the response variable than other predictors.

Analysis

Analysis of the data collected found that survey response were recorded differently for different topics. For example, for one column a Yes/No response may have been recorded as "Y"/"N" but for another column it was recorded as "0"/"Yes". Such instances were transformed into a consistent format across columns while addressing missing values wherever possible. The response variable *Quality of Life* (also referred to in this document as "QoL") that we attempt to predict ranges from 0 (Very Poor) to 10 (Excellent). It was found that a larger number of responses selected a QoL on the higher end of the scale as shown in below charts.



Prior to applying the specific data mining techniques, relationships between some of the variables were also explored. One such relationship between QoL and present Health was charted as shown below and a pattern can be detected.



Next the 2 data mining techniques were applied.

For the **Association Rules**, rules that appear in 10% of them were mined, with 20% confidence. The resulting 142,000 rules were filtered to have at least one of the QoL values on the right hand side of the equation, resulting in 242 observations and 7 variables.

For the **Random Forest** method, the QoL was grouped under 4 categories - 1-6: Low (20% of data), 7: Fair (19.2% of data), 8: Good (30.6% of data), 9-10: Excellent (30.2% of data). When predicted, the out of the box error rate was estimated at 52%.

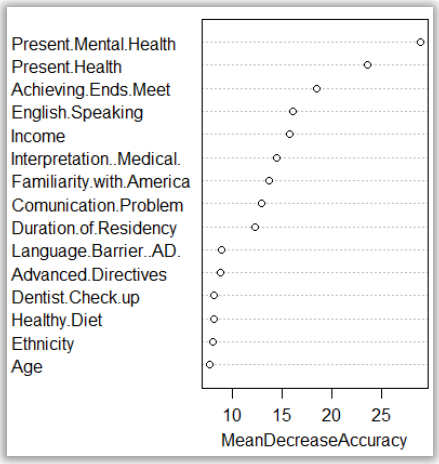
Findings from Association Rules

- Association rules were not identified for the lower QoL which likely was because the median of QoL was 8 in the data set
- Healthy Diet was included 67% of the QoL:9 rules, 24% of the QOL:8 rules and none for the lower QOL rules
- QoL: 9 - ALL of the rules were made up Health related features - Healthy Diet, Health/Dental Insurance, Physical Checkups, Dental Checkup, Primary Care, and Regular Exercise. One of the rules was Healthy Diet and having a Spouse.
- QoL: 8 - Health Insurance, Healthy Diet and Physical Checkups were the most common among these rules as well. But also at the top of the list of association rules here were Counseling (Have you ever received psychological counseling or treatment?), having English as a Primary Language and having a Spouse. For rules that included income, it was always $70,000 and above.
- QoL: 7 - There were only 5 association rules which had Health Insurance as a common factor with the other QoL. But most telling were the simple rules with psychological/Depression Counseling preference (If you use counseling, would you prefer a counselor of your own ethnic group?), Doctor Preference (If you could choose, would you prefer to be treated by a doctor of your own ethnic group?) and having Family as source of health information (Where do you get health-related information?). These feature that

were not in the higher QoL would seem to indicate a need for assistance in a manner where a patient's ethnic or cultural background may be understood.

Findings from Random Forest

- Some of the finding here aligned with the association rules. These were importance of Present Mental Health, Mental Health and fluency in English and related Communication Problem (Have you had an experience that you could not understand what the doctor/nurse said?)
- Financial features also were prominent here with yearly household Income and Achieving.End.Meet (Thinking of your household's total monthly income, would you say that your household is able to make ends meet?) being 2 of the top 5 in importance.
- The top 15 features in order of relative importance are shown in the below chart.



Conclusion

Following are the conclusions drawn from the analysis and data mining of this specific data set of the Asian American community surveyed in Austin in 2015 and recommendations to enhance the quality of life if this community.

1. Present Health was the primary predictor of a higher Quality of Life. This includes a Healthy Diet, Health insurance, regular physical/dental checkups. The organizations and government serving this community will do well to pay close attention to increasing access to health care and enabling the people to make healthier choices in their life.
2. Fluency in English contributed either directly or indirectly a higher Quality of Life. Organization serving this community could make English speaking resources available to the adult community preferably through easily accessible methods.
3. For those that may still not have access to learning or the time or desire, the community should be enabled to include health care professionals and counselors from their own ethnic groups. This could be through access to education or scholarships within the community so that graduating students could return to serve their communities.

References

- Austin government's final report on Asian American Quality of Life
- CSV file download link
- Survey Questionnaire (blank)