Olarn Pornpitaksuk

11/13/2016

**Data Science Intensive – Project Proposals**

1. Yelp - What does it take to be a Yelp Elite?

Becoming a yelp elite has prestige and tangible benefits especially in urban cities such as New York. However, it can sometimes be somewhat of a mystery as to what it takes to become a yelp elite as they have a black-box application process. For aspiring Yelpers, what kind of commitment does it take to become a Yelp Elite Member, and what does it to take to maintain your elite status?

The deliverable will be a list requirements that are likely to allow any yelper to gain elite status in NYC.

The data for this analysis will be scraped from the Yelp website. We will analyze two sets of data:

1. Yelp users who were recently appointed "Elite Status" in 2016
   1. We will want to understand what actions on the app lead to gaining elite status
      1. The reviews
         1. Volume of reviews
            1. Range of cuisine types
            2. Geographic range of restaurants
            3. # of "New" restaurant review OR reviews for restaurants with limited amount of likes
         2. Rating Given
         3. Length of Text
         4. Number of likes
            1. Cool
            2. Useful
            3. Funny
         5. Photos uploaded
2. Yelp users who lost their elite status
   1. Understand what their user activity was during the previous years and then see how their activity had changed that resulted in the loss of "Elite" status
      1. Years Prior
         1. Number of reviews written
         2. Rating Given
         3. Length of Review
         4. Number of Likes received
         5. Photos uploaded
      2. Year of Elite Status Lost
         1. Number of reviews written
         2. Rating Given
         3. Length of Review
         4. Number of Likes received
         5. Photos uploaded

1. Yelp - Competitive Landscape analysis of All Thai Restaurants in NYC

My uncle owns a Thai Restaurant in Bayside Queens, I believe it would helpful for him to understand the feedback he's received from his Yelp reviews over the years, and while also understanding how his competitors are doing. Within only my Uncle's restaurant, I'd like to identify key value drivers that he brings to his customers, in addition to areas of improvement. When analyzing other restaurants, I'd like to understand what they've been doing well, not so well, and which dishes have been their most popular. Competitors within a 5 mile radius will be included in the study.

All data will be scraped from the Yelp profiles of each of the Thai restaurants being studied.

1. Uncle's restaurant - Understanding what customers do and don't like
   1. The reviews
      1. Ratings Given
      2. Text Review
         1. Dishes Mentioned
         2. Sentiments
            1. Natural Language processing
      3. Weighting reviews based on user history
         1. Yelp Elite Status
         2. # of reviews provided
         3. # of good reviews vs. bad reviews

1. Identifying the most popular dishes at each restaurant
   1. The reviews
      1. Ratings Given
      2. Text Review
         1. Dishes Mentioned
         2. Sentiments
            1. Natural Language processing
      3. Weighting reviews based on user history
         1. Yelp Elite Status
         2. # of reviews provided
         3. # of good reviews vs. bad review

1. Medicare Health Advantage Participants - What were the leading causes of dissatisfaction amoung patients in the earliest co-horts in each data set, and what recommendations could we make about improve the experience for these patients (given they have the longest amount of time on the beneficiary plans)?

ISPCR from the University of Michigan has collected data on the physical and mental well-being of randomly selected Medicare Advantage Plan participants from different insurance companies over two years. Two waves of surveys were submitted to these patients to observer how the mental and physical well being of patients have changed over each wave. For example, Cohort 1 between 1998-2000 has two data sets – the baseline and the follow up which represent each wave of research. There has been a total of 3 cohorts captured in the study, spanning from 1998 to 2002. The data can be found in the following link: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/23380?classification=ICPSR.IX.&q=&sortBy=7#method>

I would like to study the leading causes of dissatisfaction among the youngest Medicare participants (ages 62.5 – 70) in each cohort. This will allow the Medicare agencies to understand what improvements to existing Medicare Plan guidelines could be provided to managed care organizations creating these plans.

The answers to the following survey questions will be evaluated:

Machine generated alternative text:
SF-36S 
survey questions are comprised Of the following: 
General health question (Q I ) 
Health transition question (Q2) 
Vigorous activities question (Q3a) 
Moderate activities question (Q3b) 
L ifting or carrying groceries question (Q3c) 
Climbing several flights of stairs question (Q3d) 
Climbing one flight of stairs question (Q3e) 
Bending, kneeling, or stooping question (Q30 
Walking more than a mile question (Q3g) 
Walking several blocks question (Q3h) 
Walking one block question (Q3i) 
Bathing or dressing question (Q3j) 
Physical health limiting time spent on activities question (Q4a) 
Physical health limiting amount accomplished question (Q4b) 
Physical health limiting the kind of activities question (Q4c) 
Physical health causing difficulty performing activities question (Q4d) 
Emotional problems limiting time spent on activities question (Q5a) 
Emotional problems limiting amount accomplished question (Q5b) 
Emotional problems limiting carefulness (Q5c) 
Extent health interfering with social activities question (Q6) 
Bodily pain question (Q7) 
pain interfering with work question (08) 
Full of pep question (Q9a) 
Nervous question (Q9b) 
Down in the dumps question (Q9c) 
Calm and peaceful question (Q9d) 
Lots of energy question (Q9e) 
Downhearted and blue question (Q9t) 
Feeling worn out question (Q9g) 
Happy question (Q9h) 
Feeling tired question (Q9i) 
Amount of time health interfering with social activities question (QIO) 
Sick easier question (QI la) 
As healthy question (QI 1b) 
Future health question (QI I c) 
Excellent health question (QI Id) 

Machine generated alternative text:
General 
with the the HOS items nut provide 
to risk may or a plan's SLEh 

Machine generated alternative text:
Living: 
(Q 12b) 
120 
geumg 
(0120 
Chef' 
• Chest hen (Q 3b) 
o' bream n (Q 
of n 'ham (Q 
of bream flight (Q 
or legs 'he (Q 
espec.auy at night 
Decreased hod 
'hat heal (Q 
paralysis or (Q 
'0 read (QI 
'0 hear IQ 
(022) 
(020 
(QJ,O 
Of With 