

Module 1

ASSIGNMENT 1: CLUSTERING

Five theatres that stage non-musical plays in Chicago joined forces to develop a marketing plan. The group of theatres is called North Shore Live. In addition to the segment of current theatre attendees, the group wants to attract people who do not currently attend theatre productions. In developing a marketing plan, the group conducted a marketing research study in which they recruited a random sample of approximately 2,700 adults living in the Chicago area to complete the survey. Data was recorded in the theater.csv file. The Marketing Director from the North Shore Live group would like to find different segments of people with respect to their attitudes toward theatre so they could invest in productions that are most likely to attract and enjoyed by the different segments.

What is your team's (assume you are consultants) recommendation to the Director of Marketing? The Director of Marketing is requesting a *.docx or *.pdf document showing your analysis. She would like to share the document with her team of statistical analysts so your methods and syntax/codes should be clear. Follow formal report protocols for the HW document (summary-no more than 4 pages.)

Data: theater.csv

Variables: The survey asked the respondents their attitudes and beliefs about theater using semantic differential scales. There are five scale questions and four demographic variables in this dataset.

1. *Attitude*: Respondents were asked to rate the experience of attending live theater using a seven-point scale on 10 questions. The 10 questions were very highly correlated and thus, were averaged to form the first variable called "attitude".

The 10 questions are 1=not fun and 7=fun, 1=not exciting and 7=exciting, 1=boring and 7=stimulating, bad and 7=good, 1=uncomfortable and 7=comfortable, 1=cannot appreciate and 7=can appreciate, 1=irritating and 7=relaxing, 1=disliked and 7=liked and 1=not educational and 7=educational.

2. *Planning*: 1=spur of the moment and 7=requires planning

3. *Parents*: 1 = my parents disliked plays and 7=my parents liked plays

4. *Goodval*: 1=too expensive and 7=good value for money

5. *Getto*: 1= hard to get to and 7=easy to get to

6. *Age*=Binned into categories where higher number indicates older people

7. *Educ*=Binned into categories where higher number indicates more educated

8. *Income*: Binned into categories where higher number indicates higher income

9. *Cnty*: Represents location of the respondents (county)