Hello. I’m Gene Barnes and this is my MIS 670 Capstone Project, “Finding the best Kansas colleges and dorms”.

This project began with searching web sites for reviews of Kansas colleges, and more specifically, their reviews of residential areas (dorms, apartments) of those colleges. We used the website “Niche” as they have compiled a lot of data for more than just the best of the best, but also show on their dashboard other areas of interest such as, can a student be accepted based on their SAT scores?; how successful is the college in terms of enrollment, graduation, and job placement?

As I have been an employee of K-State’s department of Housing and Dining services for quite a while, I wanted to use this project to help with our business and data understanding of how to ensure we have the best residential areas available to our residential students. Our business intelligence suggests that we always try to improve on maintaining those residential areas. The data provided will allow the department to not only help itself in efficiently managing living spaces on campus, but will also further the role of the university by providing a safe, clean environment for our residential students to be successful in their academic pursuits.

During this project, I also wanted to introduce some analysis methods that we either did not cover or briefly covered in class. The first method used is part of the web scraping or data collection, called Selenium. This package allows you to select a web browser (Chrome in this project), send it to a URL, scrape the data, then, and this is the most interesting part, automate the selection of the “next” button at the end of page to advance to the next page for scraping. There is a loop that tests if there is data available (while true), then will end if there are no longer any pages to scrape. This allowed me to obtain all the reviews for both K-State and KU.

After data cleaning, processing and saving our data into a data frame, we use descriptive analysis methods of word frequency, word cloud, pivot tables and descriptive statistics to analyze all the text in the reviews. For KSU, there were 3,047 reviews and for KU there were 2, 967 reviews. The most used word for both schools was “student”, however, “ruin” (KSU) and “bat” (KU) were also listed as the lowest words in the reviews. I was also able to produce a couple of wordclouds for these as well.

Using content analysis, we determined what each review’s sentiment or feeling was with KSU having the most positive reviews (2,480) vs. KU (2,324). KU had more negative reviews (408) against KSU (347). Both schools were almost equal in number of neutral reviews (KSU: 220, KU: 235).

The second new analysis method I found to use was clustering groups using the k-value method. Since it was difficult to produce any type of other network analysis (metrics, networks, etc.), it was determined that clustering may prove a better method. This was found inconclusive as the clusters produced were too close in sentiment, however, all three clusters seem to be positive in nature. One type of plot for clustering is called a dendrogram and because our dataset was so large, it was unable to easily break down the graph so it would be readable.

In conclusion, what started out as a simple project of finding reviews for the best colleges and dorms in Kansas turned out to be a “Sunflower Showdown” between KSU and KU, though that was not the intent. We found that KSU won this showdown in all analysis – most reviews, most positive sentiment, and better statistics. The managerial implications from this project are that Housing and Dining services needs to ensure they maintain a full compliment of support staff to keep the residential areas safe, secure, clean, and the residential students happy. Doing so will assist the university as a whole by maintaining higher levels of enrollment and greater revenue to continue providing the best education and student life possible during the college years.

Thank you for viewing this presentation.