Bellevue University

Exercise 6.2

DSC500: Introduction to Data Science

Felipe Rodriguez

October 9, 2022

For this scenario, a consumer is asking that their energy usage be analyzed to determine what times of day they consume energy the most. The consumer has access to their energy consumption per hour and will send the data to begin the analysis. The data provided should give total cost per two hours of usage. This is what the data will begin to look when documented:

Table

Description automatically generated

So far, the data gathered can be filtered by hour and compared, but once there are more variables the data will be more difficult to analyze. With this initial data, a chart can be made and then added to it once more data is received. This chart can begin to take shape and look like this:

Chart, scatter chart

Description automatically generated

As data gets added, it can be added to the chart. K-Means clustering will be important in this analysis. With K-Means Clustering, we can see the correlation of different times and the cost associated. This will aid the consumer to determine when energy is being used the most and, by clustering into 3 or 5 groups, they will be able to figure out ways to minimize their usage. Once there are more days and data, this same analysis can be done to check days of the week, which days energy is consumed the most, or even which months are consumed the most. With clustering, days, hours, or months with similar costs can be grouped together. The benefit of this clustering is that there is “one parameter to tune and you can easily see the direct impact of adjusting the value of parameter K” (Bento 2018).

Reference:

Bento, C. (2018, December 3). *K-means in real life: Clustering workout sessions*. Medium. Retrieved October 9, 2022, from <https://towardsdatascience.com/k-means-in-real-life-> clustering-workout-sessions-119946f9e8dd