

# Chicago Area Bakery

# Overview

- Client wishes to open a bakery within the county of DuPage, Illinois
- Requests review of popular establishments in each city/town along
- Would like a few solid recommendations along with the reasoning behind them

# Data Acquisition

- Cities and towns in DuPage County Illinois. Obtained from Wikipedia entry for Dupage County via webscraping.
- Venues by location – obtained by utilizing Foursquare API and functions introduced in Applied Data Science Capstone clustering lab.
- Household income of locations in DuPage County obtained from 2019 American Community Survey provided by the United States Census Bureau (<https://data.census.gov/>)
- Sales and payroll figures for the 'Accommodation and Food Service' sector obtained from 2017 Economic Census (<https://www.census.gov/eos/www/naics/>)

# Data Cleaning

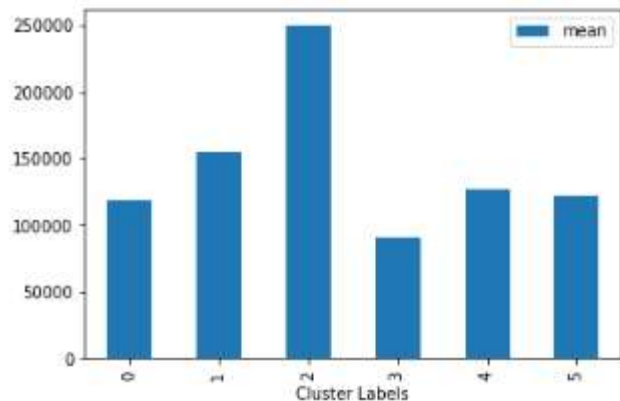
- The vast majority of cleaning was related to the different formats the sources used for the locations. All sources needed to have the same formatting so that the sources could be merged
- Household income data contained a handful of text entries which need to be removed to have the income figures be of a numeric data type
- One Hot Encoding was used to convert categorical data to numerical data as required by K-Means clustering

# K-Means Clustering

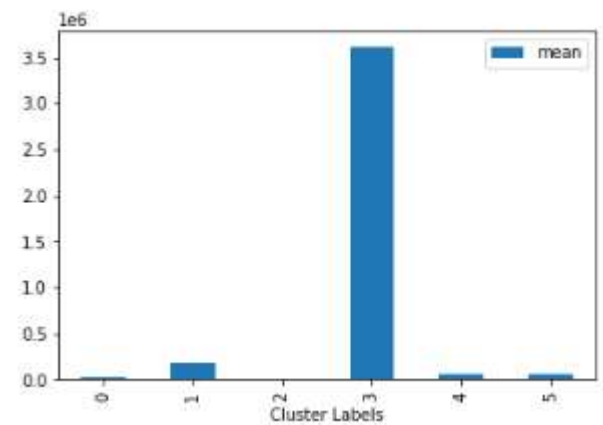
- There are 36 locations prior to clustering.
- Requesting identification of 6 clusters
- Dataset normalized with `StandardScaler()` function.

# Results

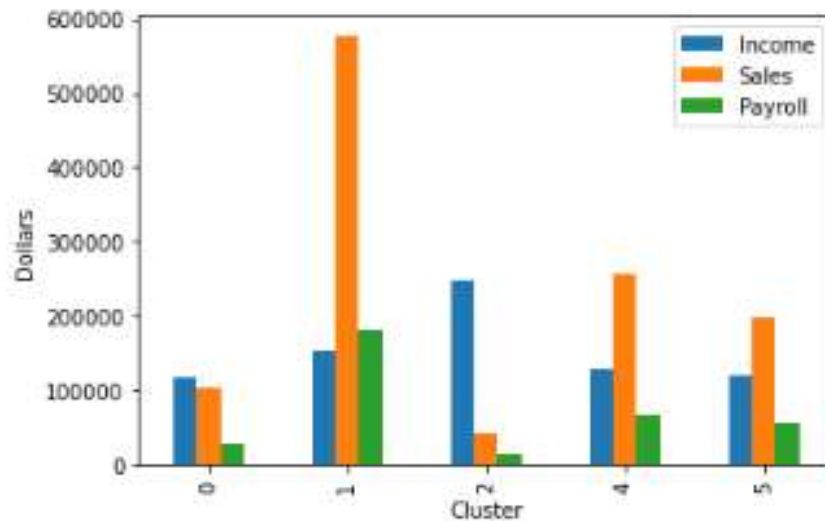
- Mean Income between 5 of the 6 clusters relatively similar



- Bar graphs for sales and payroll for each cluster resembled the figure to the right. This cluster contains only the city of Chicago.



- Removing the cluster containing Chicago, the comparisons between the remaining clusters for sales and payroll are more easily seen.



- From this figure, the annual sales of cluster 1 far outreach those of the other cluster. Although the annual payroll is also larger, it is not to the same degree as the others.
- Based on this, the recommendation to the client would be to open their bakery in Naperville, IL which is the only location within Cluster 1.