



# Finding perfect spot for next Montessori

Using Machine learning to find best area for new Montessori school in North Texas



# Title and Content Layout with List

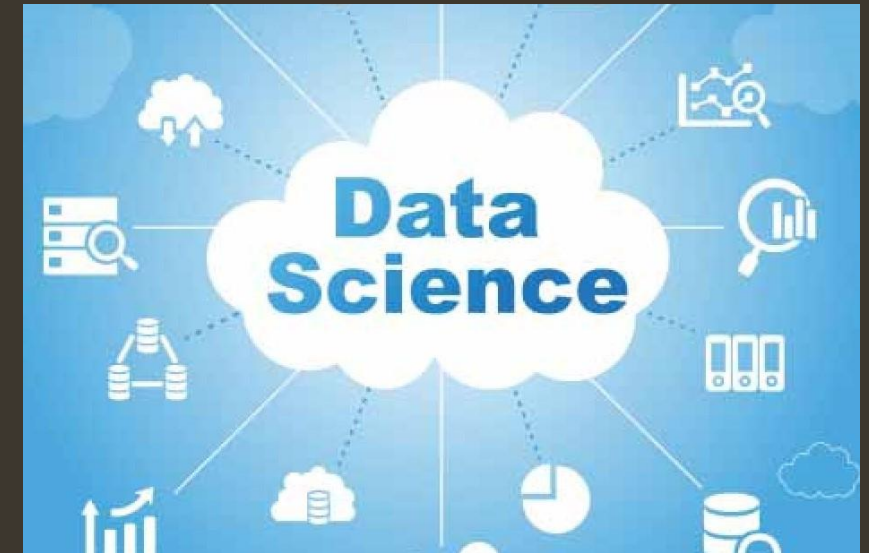
- Problem Statement
- Data requirement
- Methodology
- Results
- Conclusion

# Problem statement

A successful Montessori owner in my hometown of Frisco, TX wants to expand to new areas within the DFW metroplex. After having 2 successful schools, one in Plano TX and other one in Frisco TX, the owner now wants to expand in a similar area and wants some recommendations

Anand Lonkar to the rescue with his newly earned Machine learning skills!

In the presentation we will see how I used the knowledge learned in the **IBM Data Science Professional Certificate** program to solve this problem

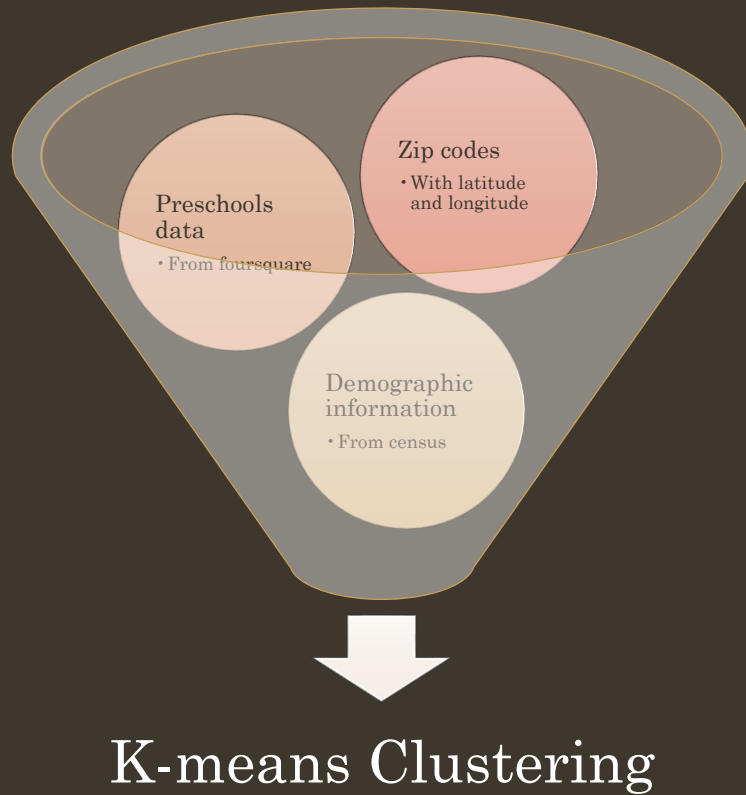


# Data requirement

- All Zip codes within 100 miles of my location - Data from <https://www.zipcodestogo.com/lookups/radius-search.php>
- Location data for all zip codes in Texas - <https://github.com/OpenDataDE/State-zip-code-GeoJSON>
- Census Data for each zip code –
- Population data – Total population and kids under 5
- Income profile for each Zip code
- Housing data for each Zip code
- Foursquare API - to find educational institutions in each of the zip codes

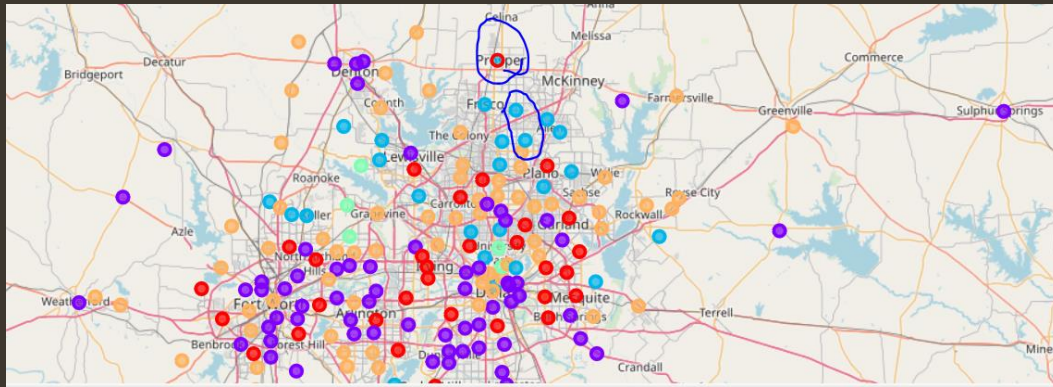


# Methodology



- Collect location data with zip codes
- Get preschools information from foursquare
- Cleanse and merge collected data
- Calculate kids to school ratio
- Perform kmeans clustering making sure that the existing preschools of the owner are in same cluster
- Select the zip code with highest kids to school ratio

# Results



- Prosper TX is the clear winner
- In the same cluster as the other 2 locations
- Has a high kids to school ratio
- Close to the other 2 locations

# Conclusion

- Given proper data set, K means is an effective way to cluster like datasets
- In this scenario, the neighborhoods with high income and higher percentage of children under age of 5 were clustered together
- Simple math was then used to find the best zip code with least competition