# CAPSTONE — IBM DATA SCIENCE

**Where to open a mid-range Restaurant in the suburbs of North Dallas**

**Llorenç Aguila**
llorenc.Aguila@gmail.com

# Key facts about Dallas Fort Worth

- The Dallas–Fort Worth metroplex, officially designated the Dallas–Fort Worth–Arlington, TX Metropolitan Statistical Area by the U.S. Office of Management and Budget,[3] is a metropolitan area in the U.S. state of Texas encompassing 11 counties. It is the economic and cultural hub of North Texas. Residents of the area also refer to it as DFW, or the Metroplex.

- The Dallas–Fort Worth metroplex's population is 7,539,711 according to the U.S. Census Bureau's 2018 population estimates,[4] making it the most populous metropolitan area in both Texas and the South, the fourth-largest in the U.S., and the tenth-largest in the Americas. In 2016, Dallas–Fort Worth ascended to the number one spot in the nation in year-over-year population growth.[

# WHERE TO OPEN A MID-RANGE RESTAURANT IN THE SUBURBS OF NORTH DALLAS

## Context

1. Location is important but distance is more important

2. Suburbs are car dependent

3. Restaurant and retail venues are clustered in commercial areas across the zone

## Criteria for ideal location

1. Proximity to other successful midrange restaurants

2. Proximity to other businesses (groceries, shops, etc) or low range restaurants are regarded

3. Population density

4. Population affluency

# DATA SOURCES : DEMOGRAPHICS

## Census.org

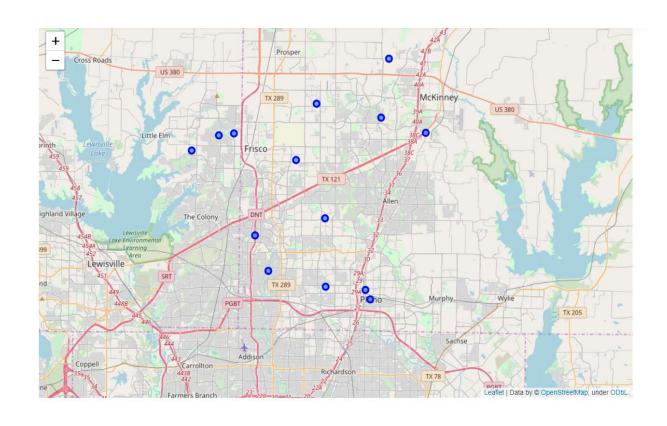| Fact | McKinney | Frisco | Plano |
|---|---|---|---|
| Population estimates, July 1, 2018, | 191,645 | 188,170 | 288,061 |
| Population estimates base, April 1, | 131,160 | 117,170 | 259,857 |
| Population, percent change - April | 46.10% | 60.60% | 10.90% |
| Population, Census, April 1, 2010 | 131,117 | 116,989 | 259,841 |
| Persons under 5 years, percent | 7.70% | 6.60% | 5.50% |
| Persons under 18 years, percent | 29.60% | 31.60% | 23.40% |
| Persons 65 years and over, percen | 9.40% | 7.50% | 12.10% |
| Female persons, percent | 51.20% | 50.80% | 50.90% |
| White alone, percent | 77.30% | 69.40% | 67.00% |
| Black or African American alone, p | 11.50% | 7.40% | 8.30% |
| American Indian and Alaska Native | 0.40% | 0.40% | 0.50% |
| Asian alone, percent | 6.20% | 16.80% | 19.40% |
| Native Hawaiian and Other Pacific | 0.00% | 0.10% | 0.10% |

## Zip-code.com bestplaces.net

| ZIP | City | Population | Median-Income | Median-Home-Price |
|---|---|---|---|---|
| 75023 | Plano | 49563 | 73602 | $285,700 |
| 75024 | Plano | 42405 | 89119 | $413,300 |
| 75025 | Plano | 53559 | 106301 | $368,900 |
| 75074 | Plano | 48977 | 55152 | $246,600 |
| 75075 | Plano | 35933 | 72576 | $303,300 |
| 75093 | Plano | 48021 | 99378 | $467,900 |
| 75033 | Frisco | 39301 | N/A | N/A |

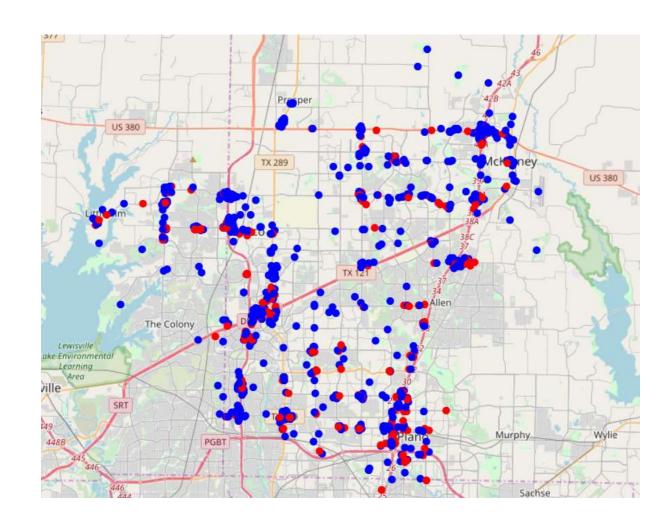Cities in scope are : Frisco, McKinney and Plano

# DATA SOURCES : GEOGRAPHICAL INFORMATION

Representation of zip codes
Belonging to Frisco, McKinney &
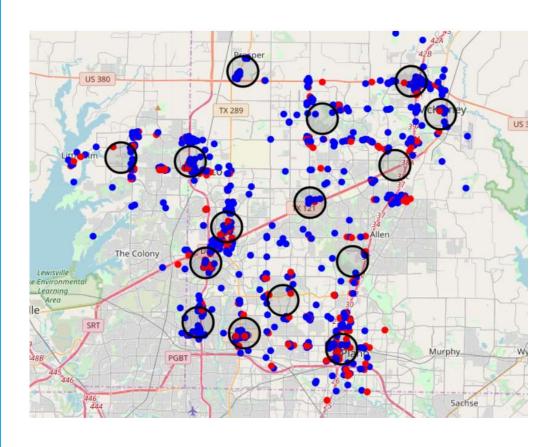Plano (Northern Dallas)
Source: Geopy (Google)

# DATA SOURCES : VENUE INFORMATION

Representation of Venues (blue) &
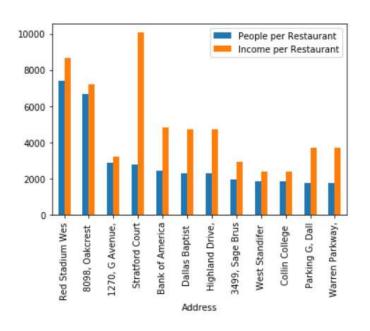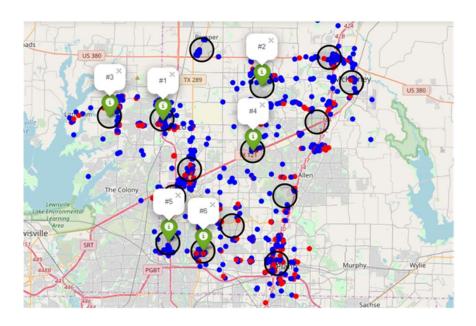Restaurants (red) within the zip codes
Source: Foursquare

# METHODOLOGY & ANALYSIS

Use of k-means to create venue
clusters by geographical proximity

# METHODOLOGY AND ANALYSIS : PICK THE OPTIMAL LOCATIONS



Top picks based on people per restaurant and income per restaurant data

# CONCLUSION AND FUTURE DIRECTION

## Conclusions

1. Existing demographic data is only available at zip level.

2. Foursquare information combined with k-means analysis provided the necessary insights

3. Final decision accuracy limited to existing commercial information

## Future Direction

1. Household level granular information will greatly improve accuracy

2. Availability of current and past company commercial data could be used to create a predictive model to forecast the success rate of a new restaurant