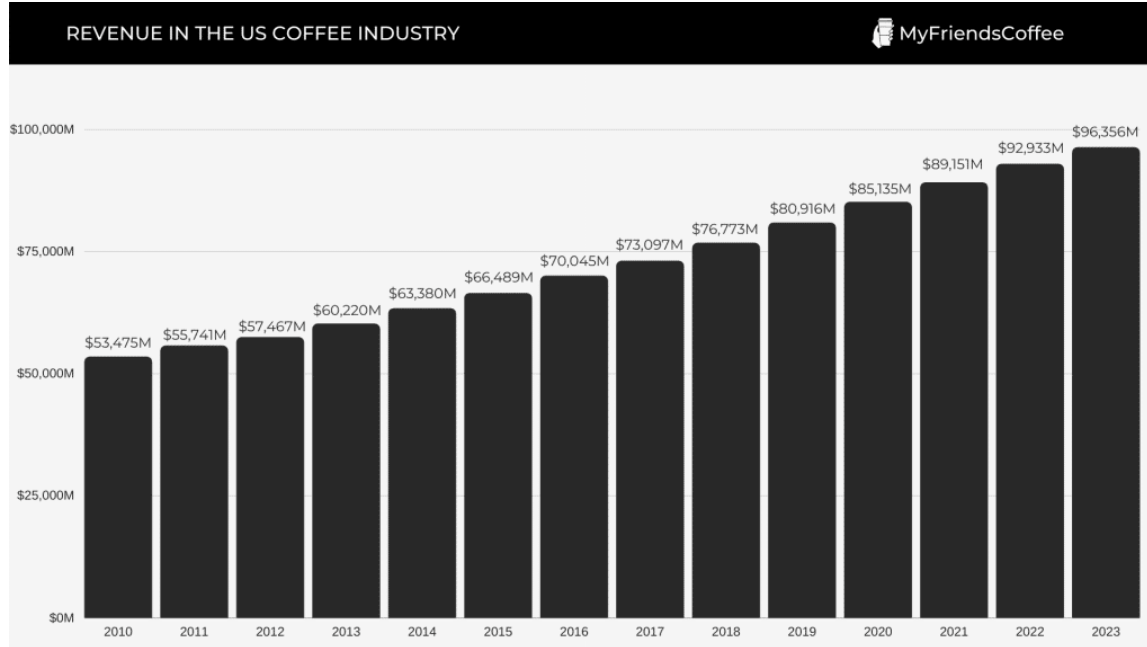


# Starting a coffee shop in New York City



**IBM Data Science Capstone Project**  
By Yu-Hung Chen

# Introduction



Fast-growing  
market with  
total revenue of  
**96 trillion dollars**

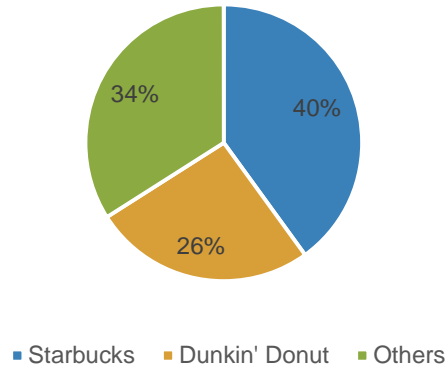
But...  
is there still  
market left?



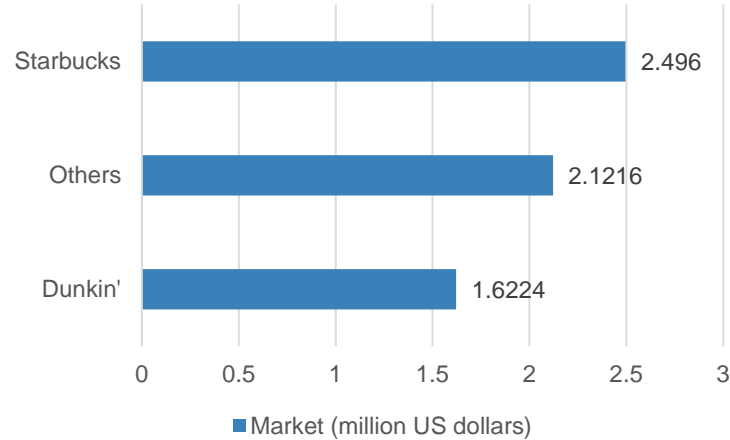
# Current Situation in NYC



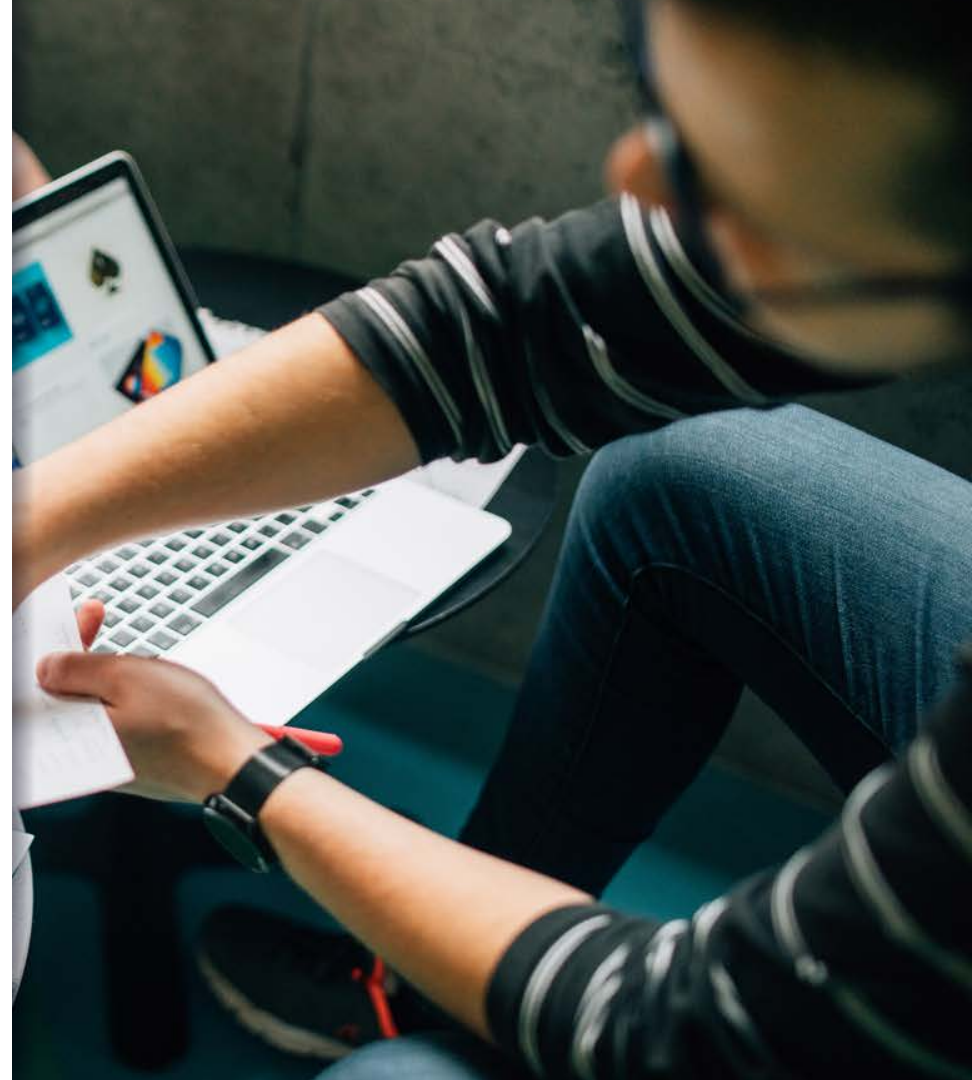
Popularity (as market share)



Transform popularity to revenue in NYC



So...  
where is  
the most  
ideal place  
to start?





# Methodology: Background



## Unequal Distribution of the Population and Coffee Shop

### Population Distribution



### Coffee Shop Distribution



# Methodology



## Target Area

Using the demographic data of each community in NYC to cluster them into groups.  
Filter for the community whose number of coffee shop is below the average of the cluster group.

## Rental Cost Estimation

Using the median rent of the community as the cost of starting the coffee shop (rent is the majority of the cost)

## Market Size Estimation

Using the number of population and population density of the neighborhood as the market potential

## Profitability Estimation

Rank the market size and cost of each community, and get the rank difference as the profitability index. The higher the index, the more profitable.

Profitable Areas Get!



# Methodology: Where Machine Learning Takes Part



## Target Area

Using the demographic data of each community in NYC to cluster them into groups.  
Filter for the community whose number of coffee shop is below the average of the cluster group.

### Machine Learning Applied in Clustering:

1. Using the elbow law to find out the best number of cluster
2. Using the following demographic data to cluster the neighborhoods:
  - Car-free commute
  - Households with childre under 18
  - Mean travel time to work
  - Median household income
  - Population
  - Population aged 25+without highschool diploma
  - Population density
  - Poverty rate
  - Serious crime rate
  - Unemployment rate
  - Units authorized by new residential building permits
  - Units issued new certificates of occupancy



# Data Used



## 1. NYC Neighborhood Data Profiles (<https://furmancenter.org/neighborhoods>)

The main data source is the data provided by the government. By selecting the district can the relevant demographic data be shown.

## 2. Foursquare API

This API will be used to collect the information of the venues located in each area.

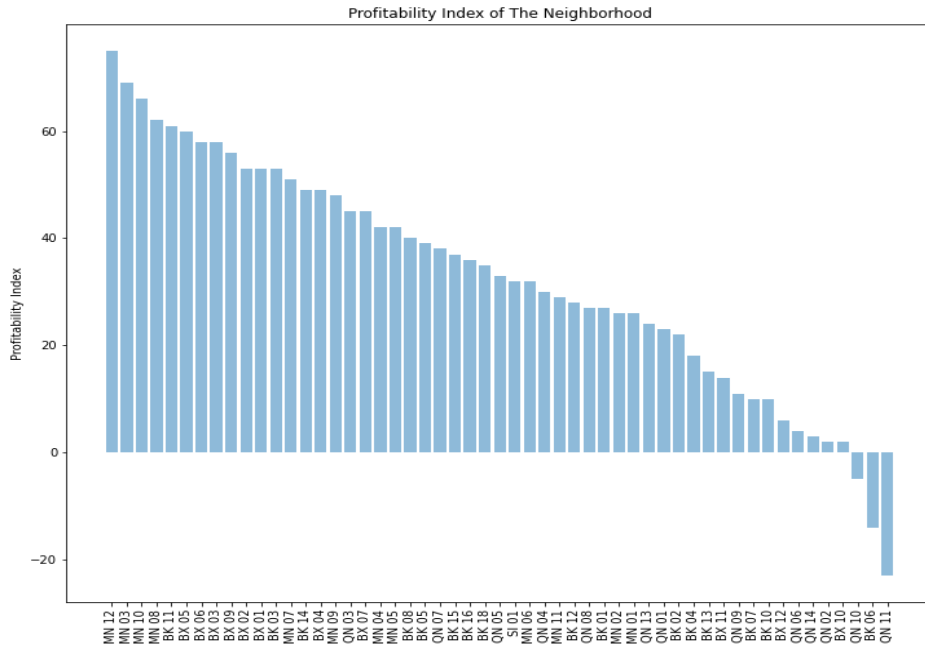


# Results

# Result Plots

Column Chart of the Profitability Index by Community

Heat Map of the Profitability Index by Community



# Go! It's Time!

By starting the coffee shop in the most profitable unsaturated areas can the chance of sucessully running the business be higher. Of course there are still lots of things to be consider additionally.

