National Chain Retailer Attraction Plan

- Business Analytics Project (City of Rochester)

Xiaodan Ding
Xiaodan.Ding@simon.rochester.edu

Pin Li
Pin.Li@simon.rochester.edu

Jiawen Liang
Jiawen.Liang@simon.rochester.edu

Ruiling Shen
Ruiling.Shen@simon.rochester.edu

Chenxi Tao
Chenxi.Tao@simon.rochester.edu

April 26th, 2020

I. Introduction

The project **objective** is to attract a national retailer and identify an optimal location for it in downtown Rochester with consideration of economic status, demographic attributes and business opportunities. The **ultimate goal** is to improve residents' quality of life and increase the city's employment opportunities, sales tax revenue, building occupancy, and retail offerings.

II. Key Findings and Recommendation

Throughout the analysis process, we got the following **key findings**:

1. The government has provided varieties of business assistance plans/programs to boost economic development in Rochester.
2. Compared to peer cities, Rochester has the least number of Walgreens and HomeGoods stores.
3. Rochester has a similar demographic composition to Cincinnati, which outperforms all comparable cities in terms of number of retailers and retail sales, so Rochester can learn from the success of Cincinnati.
4. Walgreens, CVS and Grocery are three brands that have much higher probabilities of establishing a new business in the urban core than other retailers do, combined with peer city analysis, we choose Walgreens as our target.
5. The probability of Walgreens opening a store in zip code 14614 reaches up to 98.83%.

Our **recommendation** is that Walgreens can start a new business at Midtown Parcel On Main (275 E Main St; 40 Cortland St). This development-ready site is located at midtown with good visibility, complementary stores, a nearby parking lot and medium size rent cost.

III. Description of Brand-Selection Model

Our **methodology** follows the steps below:

Step 1: verify that downtown Rochester is an ideal business destination.

Step 2: understand economic diversity and narrow down retailer candidates by conducting comparative analysis.

Step 3: justify the choice of Walgreens through Multi-Label Deep Neural Network and LightGBM Classifier modeling.

Step 4: select an optimal site for the retailer with consideration of multiple business needs.

In terms of **data** overview, we used *business location data* which shows retailer names and their specific locations in each zip code area. In addition, we explored *Experian Mosaic USA* dataset which presents consumer classification and corresponding demographic characteristics in each zip code area. We also collected more granular demographic data from the *US Census Bureau* to supplement the existing data. Furthermore, background resources included *2019 Commercial Corridor Study*, *Federal Reserve Bank of Chicago's Peer Identification Tool* and *Opportunity Zones Investment Prospectus*.

The **modeling** methodology has two phases. The first one is to select potential retailers using *Multi-Label Deep Neural Network*. Inputs are 246 columns of demographic attributes and mosaic segments of 9855 zip codes, and outputs are boolean variables indicating if there is a store located in the area. Having trained the model, we used it to predict the probabilities of 20 retailers opening a new store in downtown Rochester. The result shows that CVS, Walgreens and Grocery Markets have much higher probability. By combining the results of comparative analysis and the first-phase model, we chose Walgreens as the target retailer. The second phase is to justify our choice using *LightGBM Classifier*. Inputs are the same as the first model, and output is a binary label specifically for Walgreens. Aftering training the model, we used it to predict the probabilities of Walgreens opening a new store in downtown Rochester. The result shows that zip code 14614 has the highest probability, laying a foundation for site selection.

The **techniques** involved in this project include exploratory data analysis, web scraping, data augmentation, descriptive analysis, comparative analysis, and machine learning.