MSBA Practicum
The City of Rochester Project

Team 13

William He, Xinyu Xing, Yichan Ang,

Nam Vo, Yuqi Yang, and Suyao Xu

Table of Content

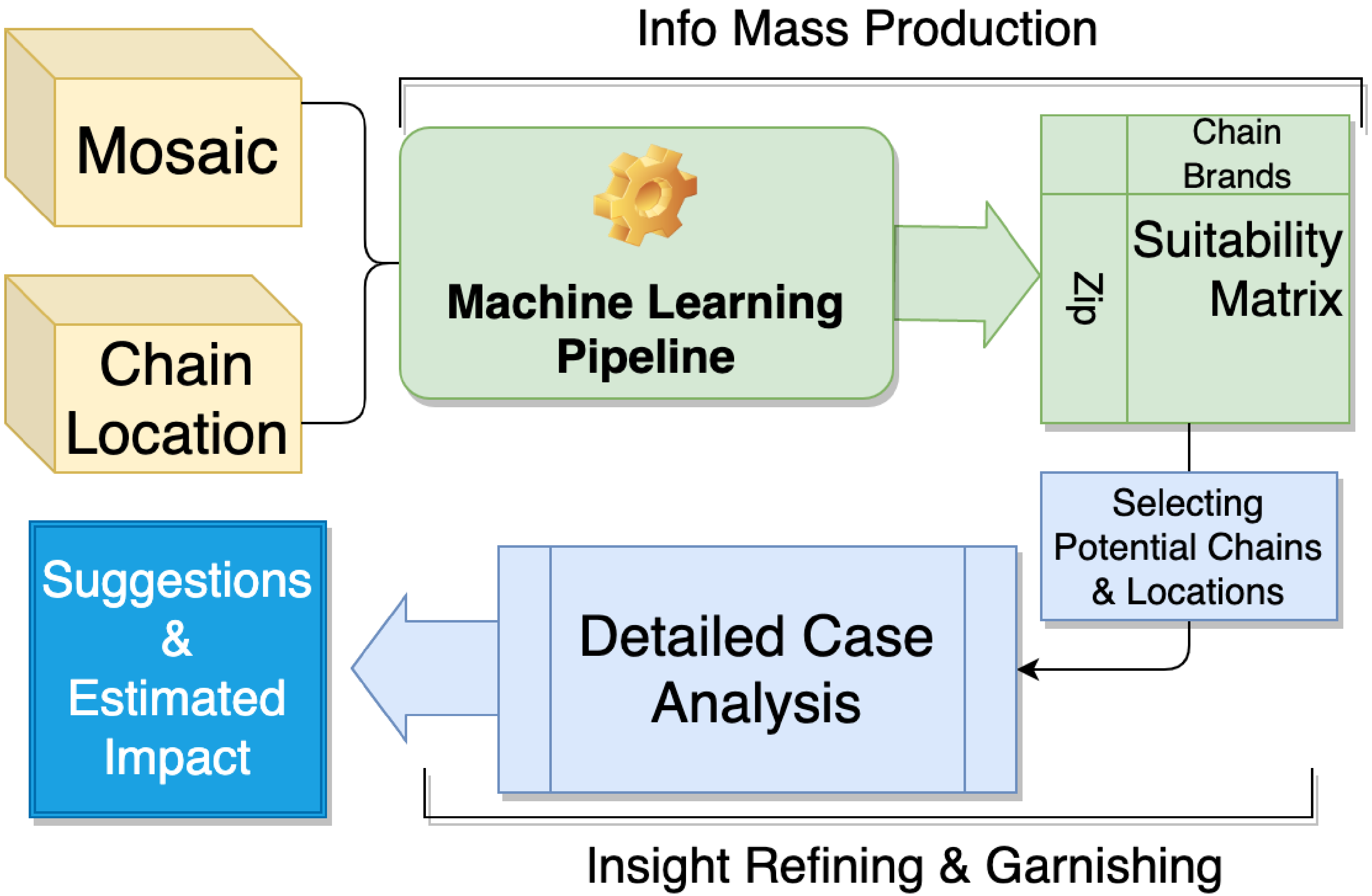
Deliverables Preview

Machine Learning Pipeline:
A pipeline process that produces machine-leaning models to predict the suitability of national restaurants or retailers in the Rochester area

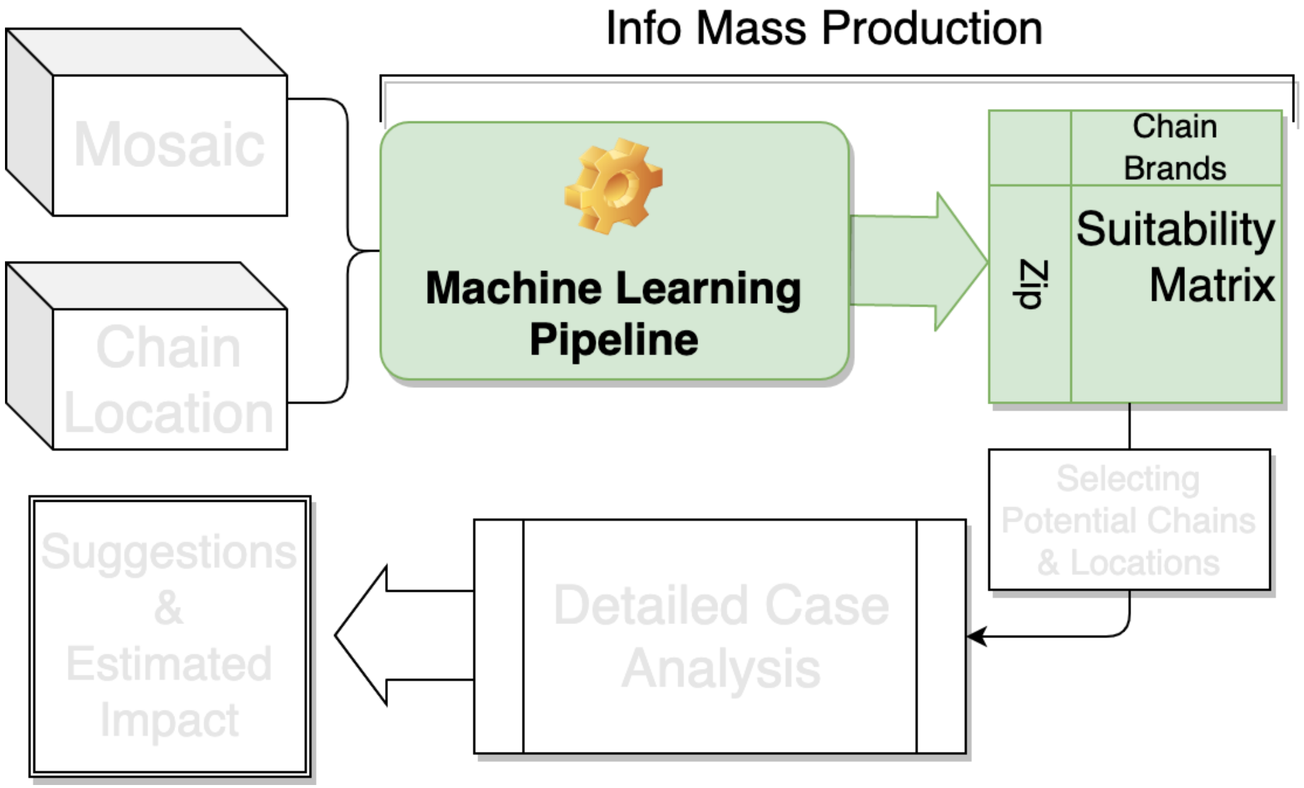
Case Analysis Findings:
From the results produced by the pipeline, four top-tier brands are matched with area suggestions based on strategy and potential revenue analyses

Final Recommendation:
Financial outlook and recommended addresses for the four brands' hypothetical stores in Rochester, which can further facilitate investment decisions

Project Overview



Phase 1 – Machine Learning Pipeline



Goal

Build a model for each brand that predicts **if each ROC zip location is suitable for that brand?**

Process

1. Data Extraction

Extract Model-Ready Datasets

2. Resampling

Resample Datasets Based On Target

3. Model Tuning

Tune Model Using Hyper-Parameters

Pipeline – Data Extraction

Model-Ready Datasets

ZipCode	IfStore	Mosaic Seg 1	Mosaic Seg 2	Mosaic Seg …	Mosaic Seg 71
10001	1	245	98	…	13
10002	0	42	23	…	6

Target

Predictors

ZipCode	IfStore	Mosaic Seg 1	Mosaic Seg 2	Mosaic Seg …	Mosaic Seg 71
10001	0	245	98	…	13
10002	0	42	23	…	6

……

10 Datasets: One each for 9 picked brands from given data plus Shake Shack, which we scraped from its website.

Binary Target: To predict classification of ifSuitable, we turned count into binary ifStore.

Pipeline – Resampling

We customized a resampling technique that prevents bias caused by **one major problem using simple down-sampling.**

No What-A-Burger

Brooklyn

Similar Demographics
Difference caused by unobserved reasons*

Bias

Has What-A-Burger

Dallas

*Note: Distribution Strategy, What-A-Burger only sell in southern states

Pipeline – Resampling

We customized a resampling technique that **random selects one nearby datapoint by zip location for each minority class point***

No What-A-Burger

Fort Worth

Eliminating Unobserved Influence

No Bias

Has What-A-Burger

Dallas

*Note: (ifStore = 1)

Pipeline – Model Tuning

Random Forest

Used Random Forest based on the nature of mass production and relatively high dimensionality

Hyperparameter Tuning

Models are tuned using a set of optimal parameters to optimize accuracy

Full Automation

Our pipeline, including tuning, is total automatized given the right data format

Tuning Example

		Actual	
Predicted		Not Suitable	Suitable
	Not Suitable	**40.56%**	**3.27%**
	Suitable	**9.44%**	**46.73%**

87% Accuracy
High False Positive Rate
Low False Negative Rate

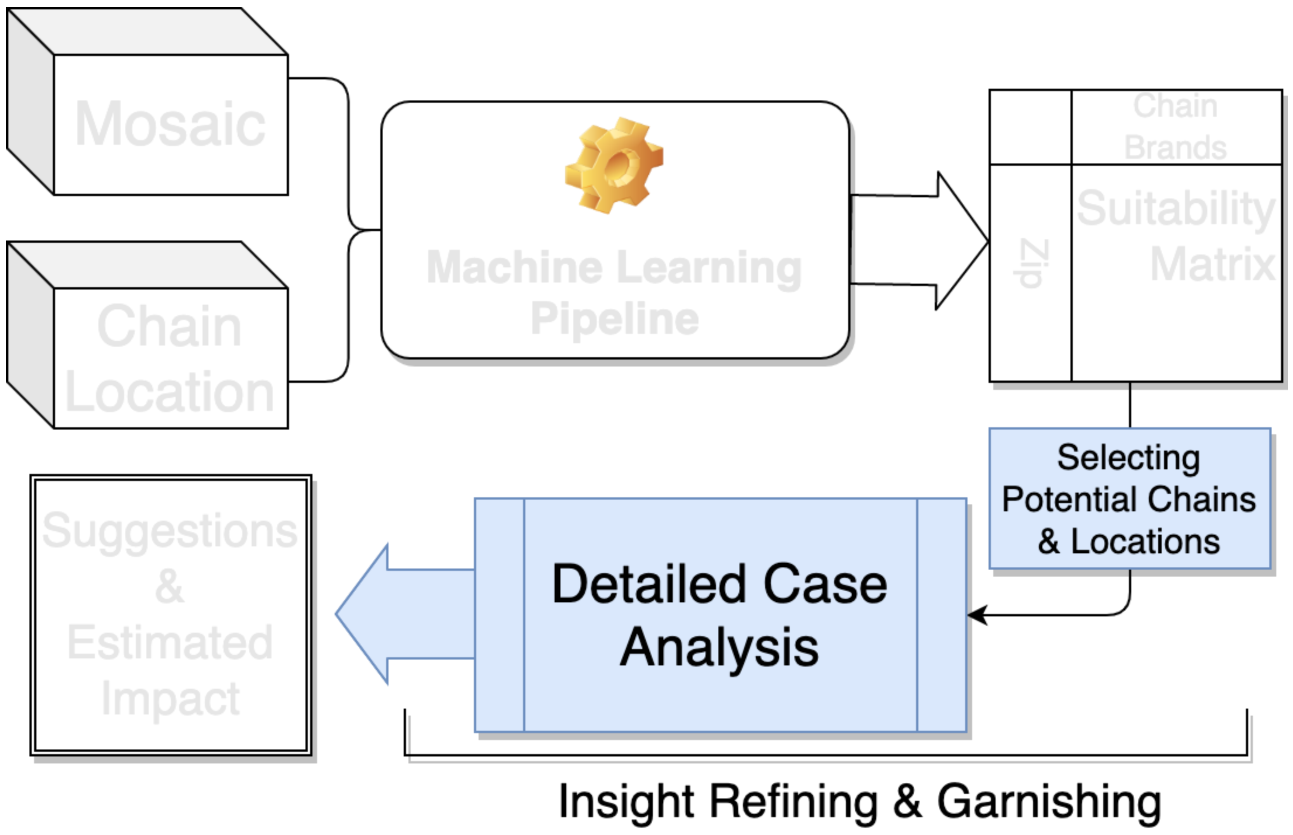
Pipeline Result – Suitability Matrix

Zip_Code	Baskin_Robbins	Bahama_Breeze	Capital_Grille	Eddie_V	Panda_Express	Seasons_52	Shake_Shack	Smashburger	WhatABurger	Yard_House
14602	5%	1%	7%	3%	15%	3%	32%	5%	1%	4%
14603	4%	0%	5%	2%	12%	0%	8%	1%	17%	1%
14604	39%	9%	72%	19%	49%	36%	66%	43%	13%	57%
14605	34%	11%	38%	6%	41%	23%	39%	29%	34%	39%
14606	54%	34%	28%	14%	36%	24%	21%	19%	61%	53%
14607	46%	63%	97%	60%	76%	69%	77%	79%	51%	86%
14608	31%	23%	83%	34%	49%	60%	59%	63%	38%	73%
14609	60%	49%	37%	28%	52%	37%	23%	33%	47%	51%
14610	71%	86%	93%	90%	65%	89%	75%	89%	43%	85%
14611	20%	14%	28%	8%	27%	21%	24%	12%	34%	39%
14612	69%	56%	26%	47%	68%	27%	12%	59%	36%	37%
14613	36%	15%	26%	11%	42%	15%	26%	13%	56%	43%
14614	12%	1%	22%	2%	15%	4%	32%	5%	4%	10%
14615	40%	23%	28%	12%	34%	20%	18%	19%	72%	45%
14616	53%	35%	20%	21%	51%	19%	16%	16%	46%	31%
14617	57%	39%	23%	36%	59%	27%	21%	31%	49%	38%
14618	74%	96%	63%	80%	68%	92%	72%	93%	33%	88%
14619	26%	9%	11%	5%	23%	11%	20%	5%	38%	26%
14620	41%	78%	79%	51%	64%	72%	62%	74%	53%	86%
14621	30%	17%	35%	14%	35%	25%	28%	23%	45%	50%
14622	28%	20%	20%	18%	26%	13%	13%	11%	20%	23%
14623	47%	46%	45%	32%	54%	45%	28%	52%	45%	74%
14624	65%	47%	17%	35%	63%	18%	4%	43%	43%	32%
14625	38%	43%	15%	49%	26%	35%	17%	34%	18%	30%
14626	73%	46%	17%	37%	66%	19%	6%	44%	40%	36%
14627	0%	3%	9%	3%	43%	2%	9%	1%	0%	2%
14692	3%	6%	6%	7%	15%	5%	21%	7%	4%	9%

Interpretation:
Model predicts **86%** probability that Bahama Breeze is suitable in Zip Location 14610

Fully Expandable
Our model pipeline works for every retailer or restaurant chain

Phase 2 – Case Analysis



Goal

Further analyze selected results from pipeline to **retrieve insights that can influence decision makers**

Process

1. Select Top Zip Codes & Chains

Interpret the suitability matrix and select potential combinations

2. Case-by-case Analysis

Conduct case-by-case analysis based on strategy and potential revenue

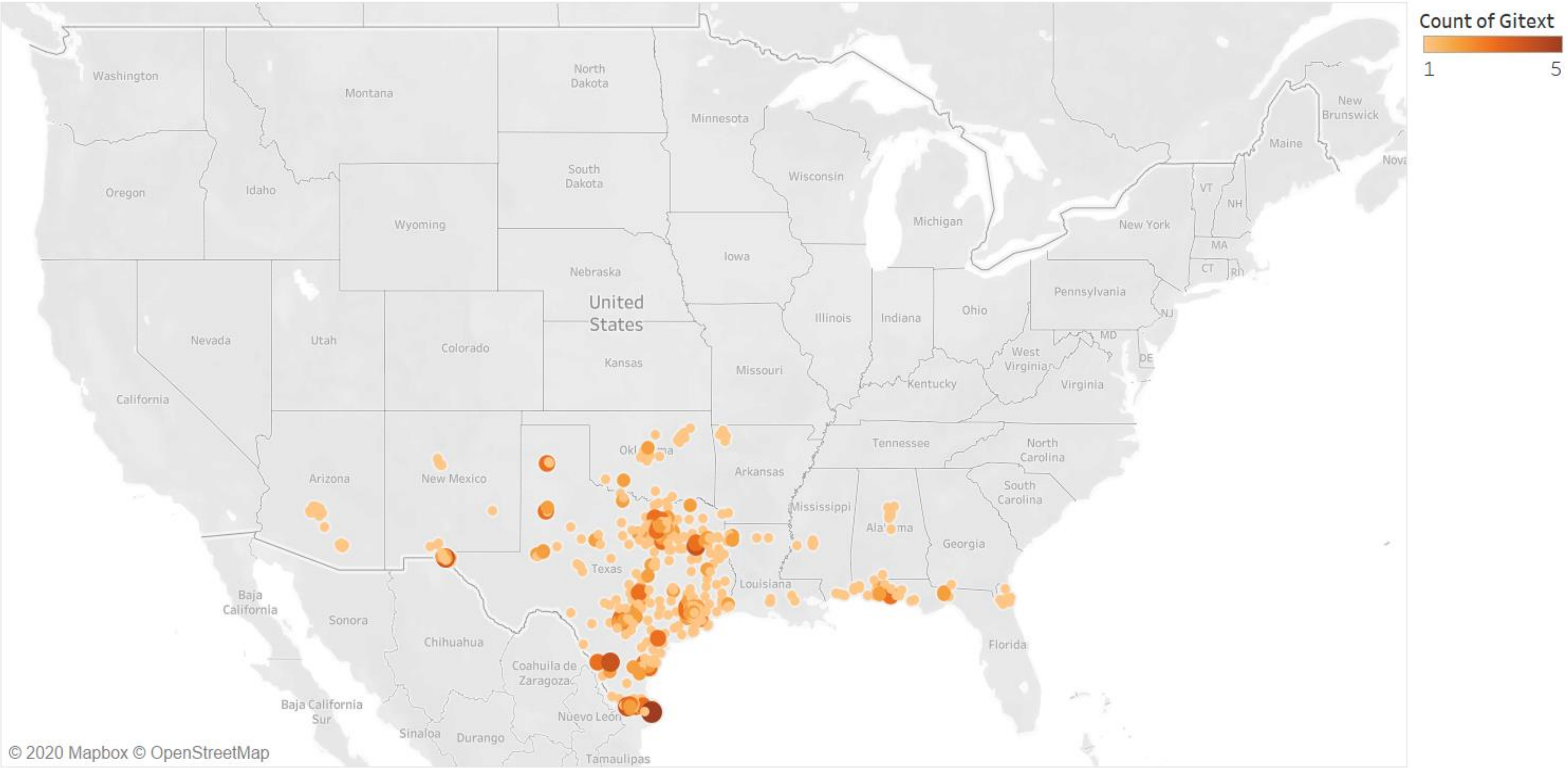
Case Analysis – Pipeline Result

According to the suitability matrix, we selected the locations which have high suitability for the restaurant chains:

Zipcode	BR baskin robbins	Bahama Breeze Island Grille	The Capital Grille	Eddie V's Prime Seafood	Panda Express Chinese Kitchen	Seasons 52 Fresh Grill	Shake Shack	Smash Burger	Whataburger	Yard House
14607			✓		✓		✓			✓
14610			✓	✓						
14615									✓	
14618	✓	✓				✓		✓		✓
14626	✓									

Case Analysis – Strategy Concern

Distribution of Whataburger in the US



What-A-burger's business strategy means it has **low probability** to settle in Rochester, although Rochester has high suitability.

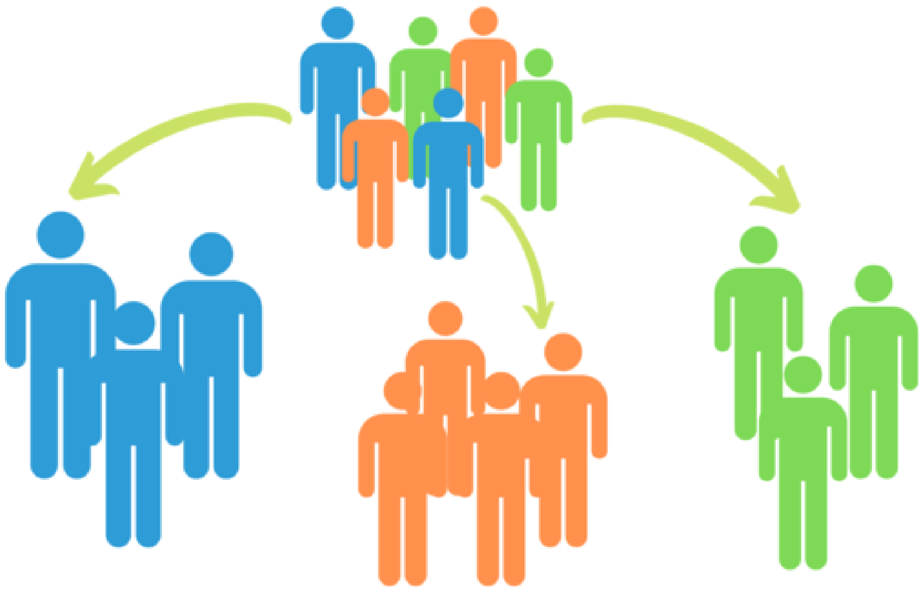
Case Analysis – Revenue Concern

Reason

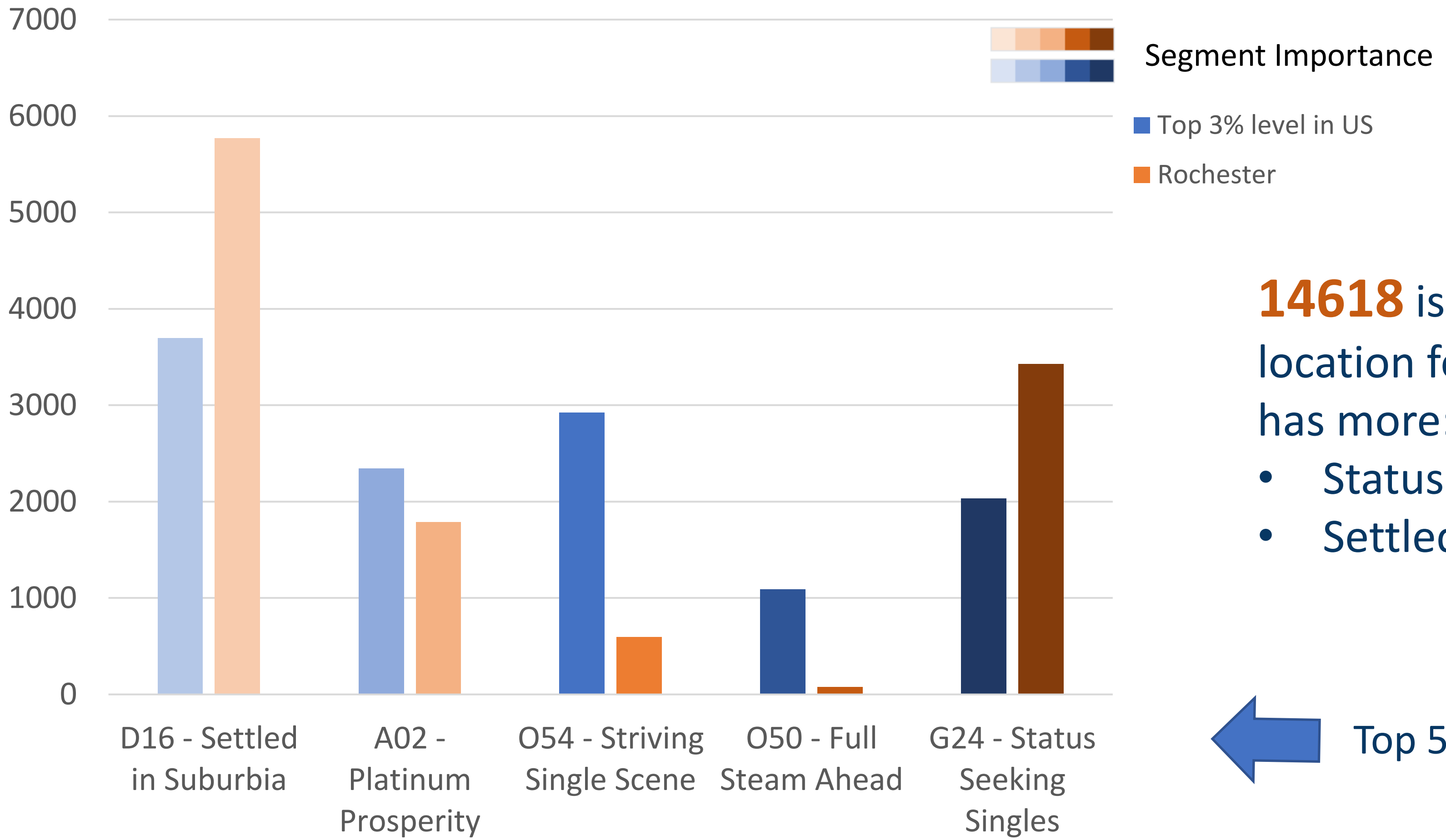
To verify the targeted location can bring satisfiable revenue, we need to make sure it has **ample potential customers**.

Method

Explore segments which have **high correlation** with the number of each restaurant brand.



Case Analysis – Revenue Concern

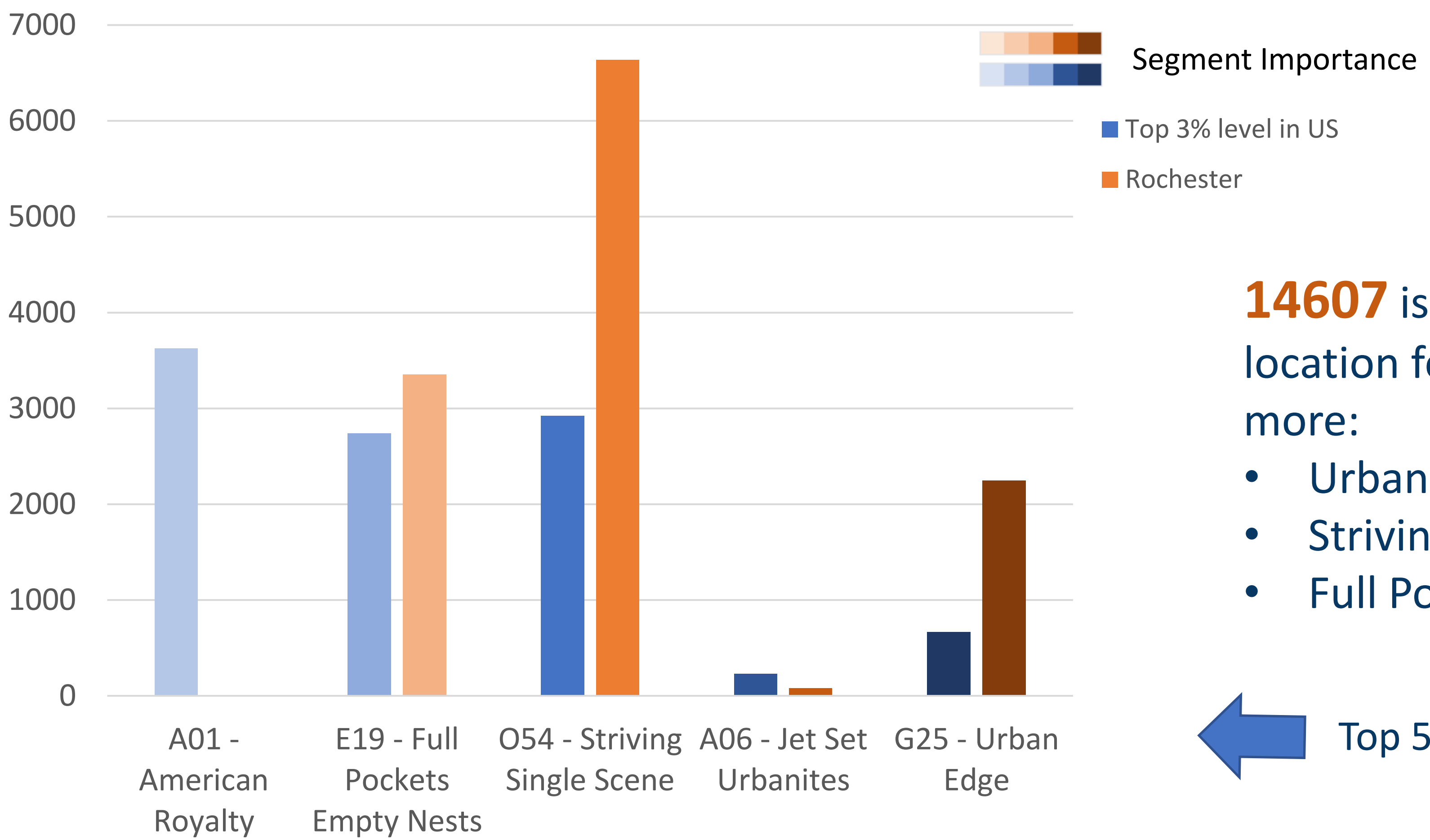


14618 is a standout profitable location for Bahama Breeze, since it has more:

- Status Seeking Singles
- Settled in Suburbia

Top 5 Customer Segments

Case Analysis – Revenue Concern

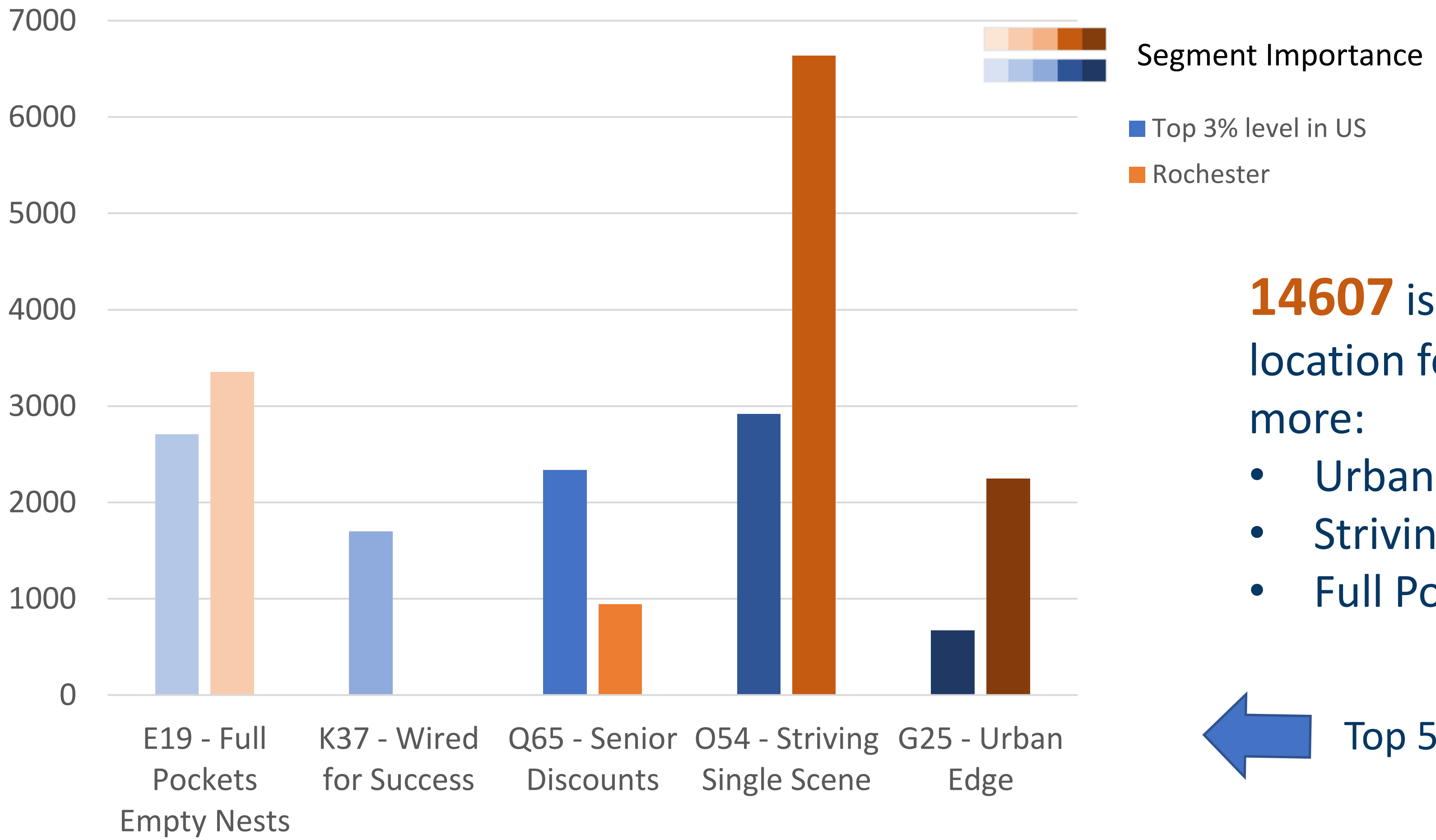


14607 is a standout profitable location for Capital Grille, since it has more:

- Urban Edge
- Striving Single Scene
- Full Pockets Empty Nests

Top 5 Customer Segments

Case Analysis – Revenue Concern



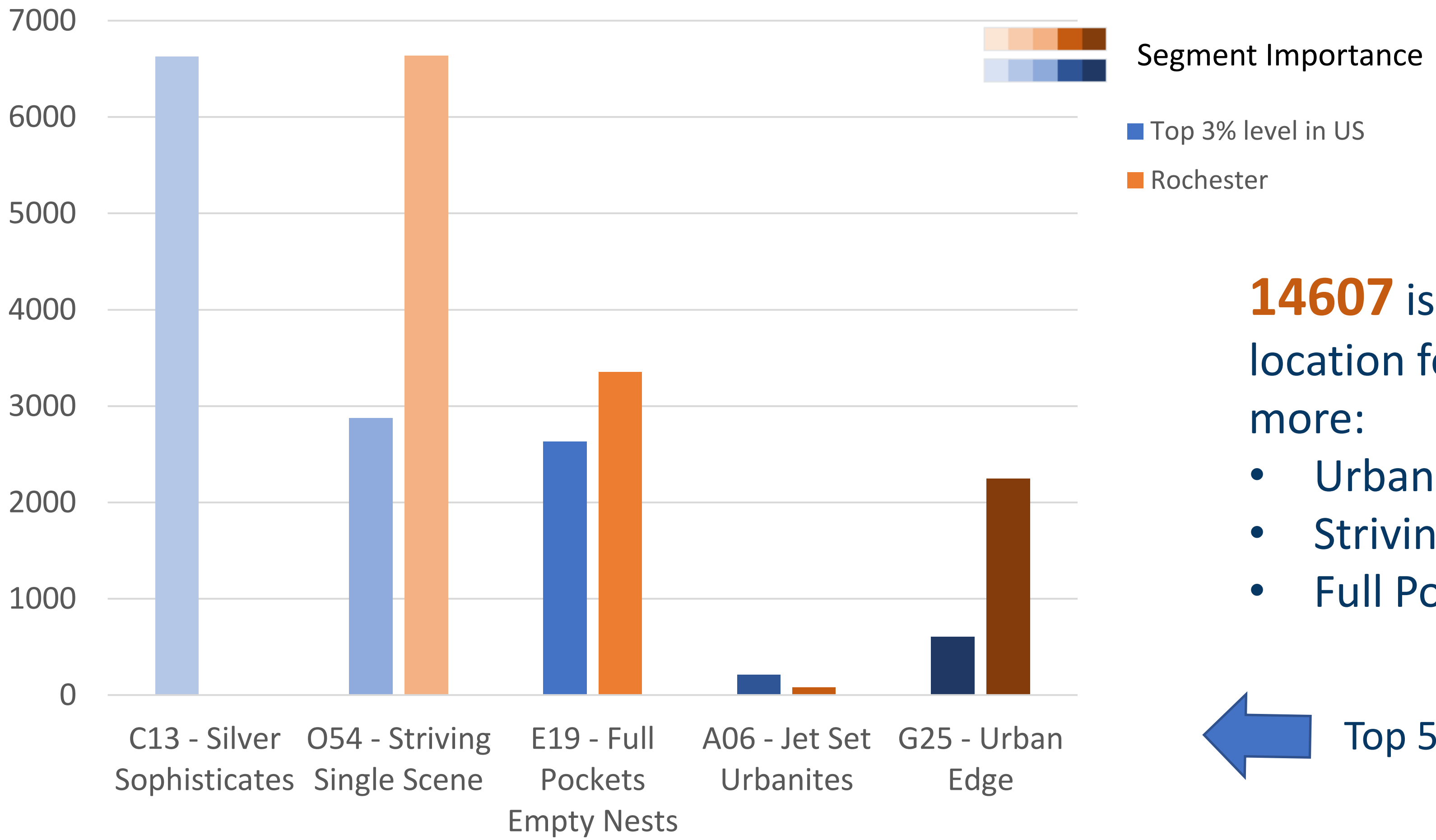
Yard House

14607 is a standout profitable location for Yard House, since it has more:

- Urban Edge
- Striving Single Scene
- Full Pockets Empty Nests

Top 5 Customer Segments

Case Analysis – Revenue Concern



14607 is a standout profitable location for Shake Shack, since it has more:

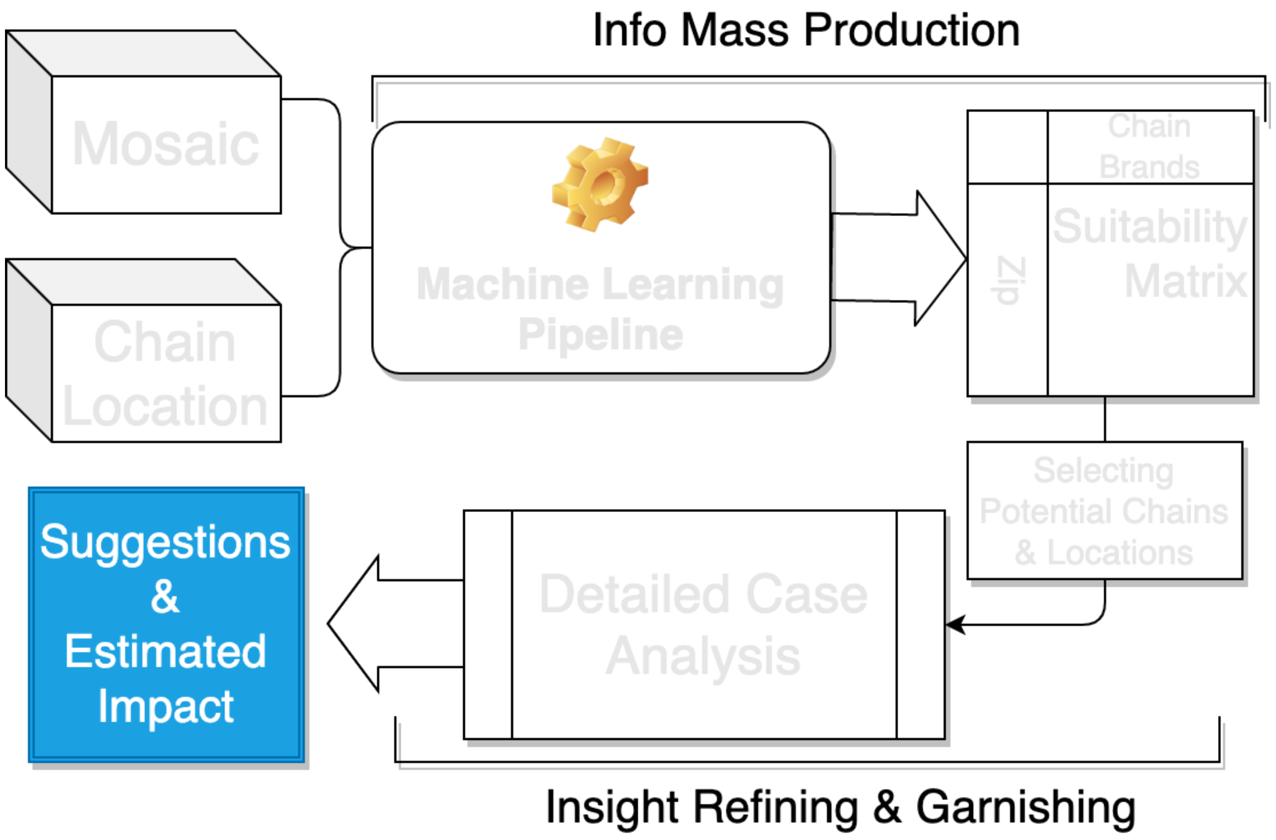
- Urban Edge
- Striving Single Scene
- Full Pockets Empty Nests

Top 5 Customer Segments

Case Analysis Result

Restaurant Chain	Suitable Location
Yard House	14607
Shake Shack	14607
The Capital Grille	14607
Bahama Breeze	14618

Final Recommendation

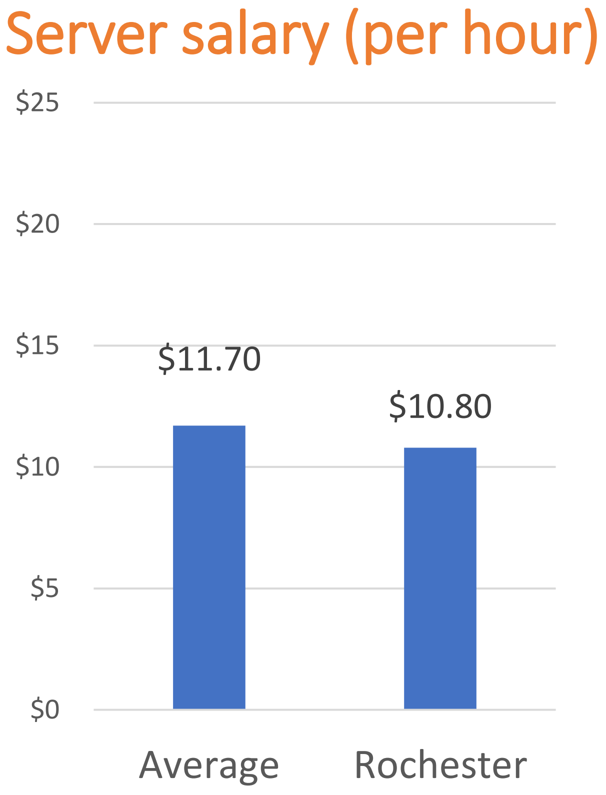


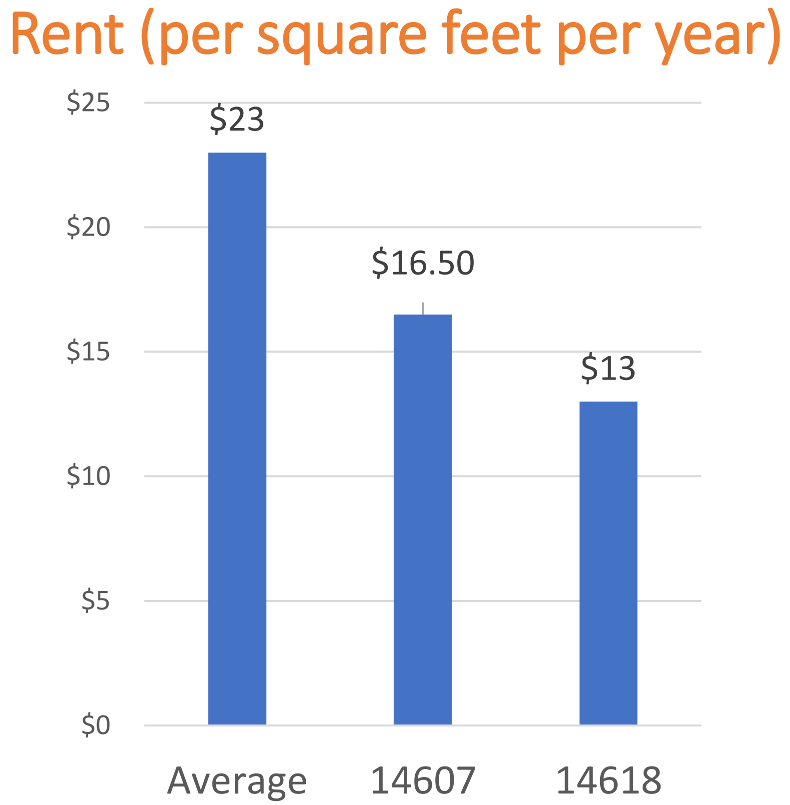
Goal

Summarize our findings and provide our own recommended **findings that the City of R can use to influence decision-makers**

Final Recommendation – Financials

- **Revenue Forecast – Achieve average**
- **Cost Forecast - Lower than average**



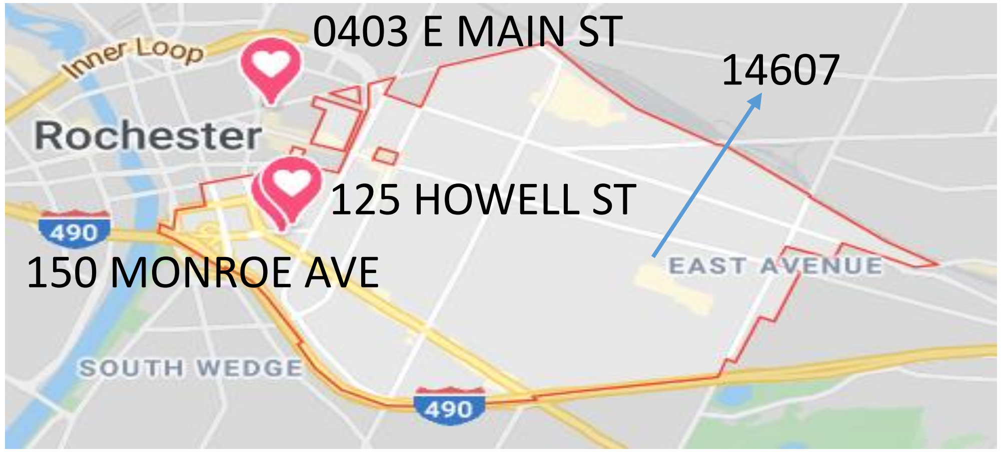


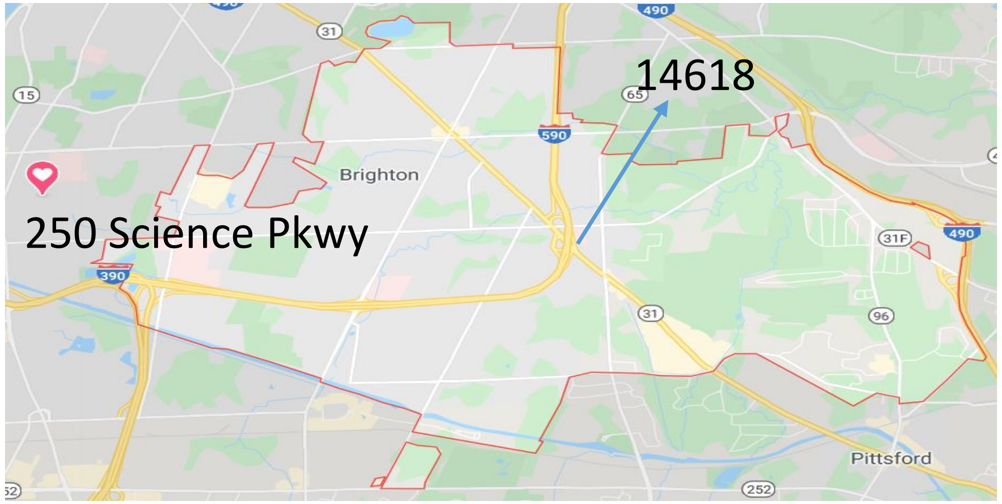
Final Recommendation – Financials

	Revenue	Cost	Annual Profit	Initial Cost	Break-Even Point	Num of Employees
Yard House	$7,343,373	$6,625,033	$718,339	$2,096,775	**2.9 yrs**	10
Capital Grille	$8,389,090	$7,568,458	$820,633	$2,395,362	**2.9 yrs**	11
Bahama Breeze	$5,869,047	$5,294,928	$574,118	$1,675,807	**2.9 yrs**	11
Shake Shack	$3,196,338	$3,059,274	$138,091	$296,680	**2.1 yrs**	20

Achieve Average **Below Average** **Community Impact for ROC**

Final Recommendation – Store Address





Possible Addresses

All recommended locations are owned by City of Rochester – Allow flexibility of rent decided by the city.

*Note: 250 Science Pkwy is technically in 14620 but is only a few blocks away from 14618.

Consolidated Recommendations

ML Pipeline

- Used to find potential locations for 10 brands
- Recommend applying this pipeline to any interested businesses

Case Analysis Findings

- Found four top-tier brands with optimal ROC locations
- The City can use our findings to persuade chains restaurants' decision makers or conduct further research

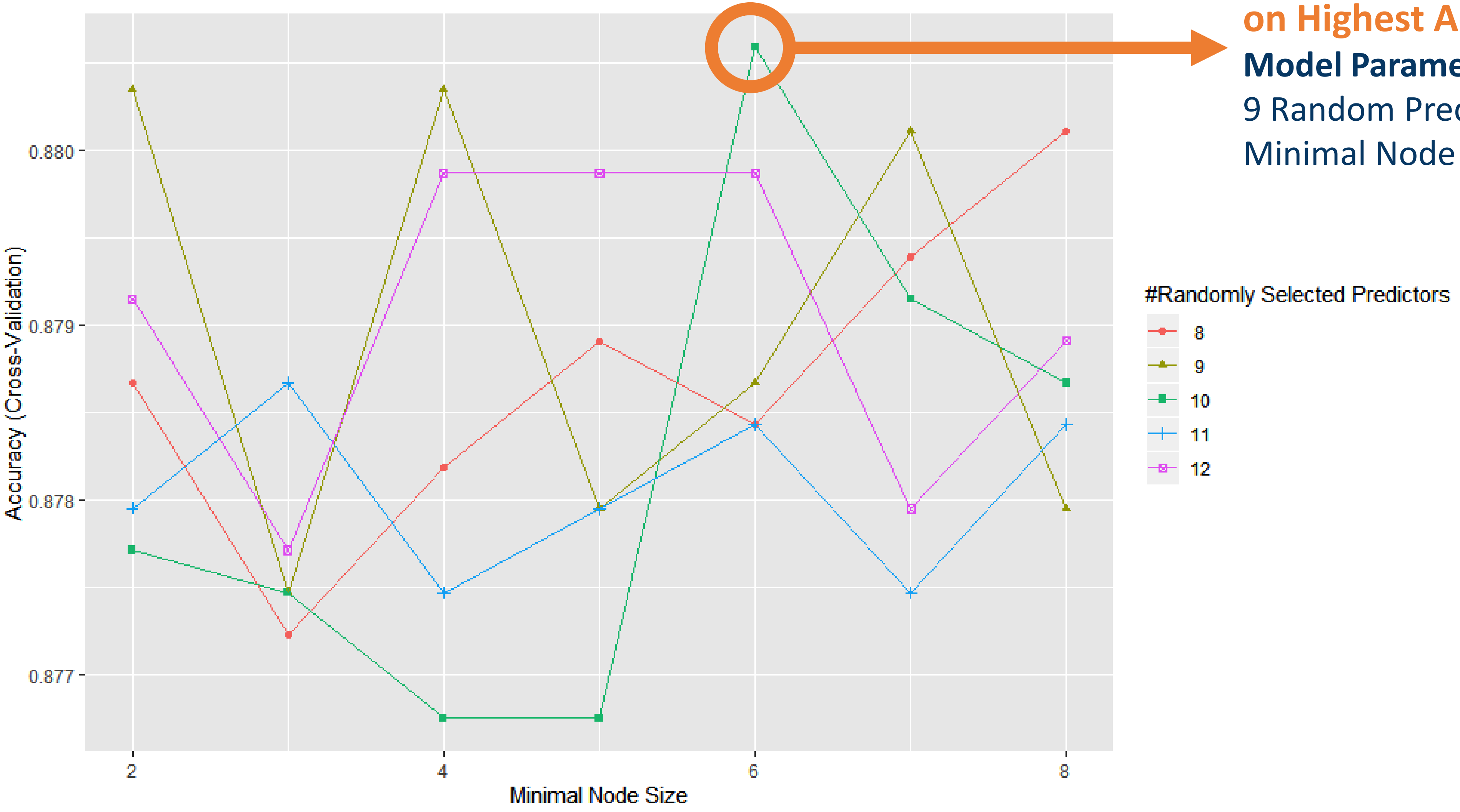
Financials & Addresses

- The most direct evidences to the decision makers
- Recommend the City to present these evidences to the chains' decision makers

Lastly…

We hope the City of Rochester can use our project results to convince the restaurant decision-makers to invest in the local communities.

Thank you!



Selected Model based on Highest Accuracy

Model Parameters:

9 Random Predictors

Minimal Node Size of 6

Current:

- Binary Classification
- Predict if a location is suitable?

Explained in the slides.

Original Design:

- Integer Regression
- Predict the degree of suitability.

Due to the problem with the distribution of restaurants count (the target), we have to use zero-inflated poisson regression to fit the model properly. However, we don’t have enough data points to complete model fitting.