



Retail Real Estate Analytics: Determining M-Express Locations

Industry Practicum
MS BAIM



Mitchell E. Daniels, Jr.
School of Business

TEAM



Veda Samhitha Alluri

Email:
alluriv@purdue.edu

LinkedIn:
www.linkedin.com/in/veda-samhitha-alluri



Shubhankar Bhajekar

Email:
sbhajeka@purdue.edu

LinkedIn:
www.linkedin.com/in/shubhankar-bhajekar



Harshraj Jadeja

Email:
hjadeja@purdue.edu

LinkedIn:
<https://www.linkedin.com/in/n/Harshraj-jadeja/>



Husam Saleh

Email:
hasaleh@purdue.edu

LinkedIn:
www.linkedin.com/in/husam-saleh-421hs



Srinija Srimamilla

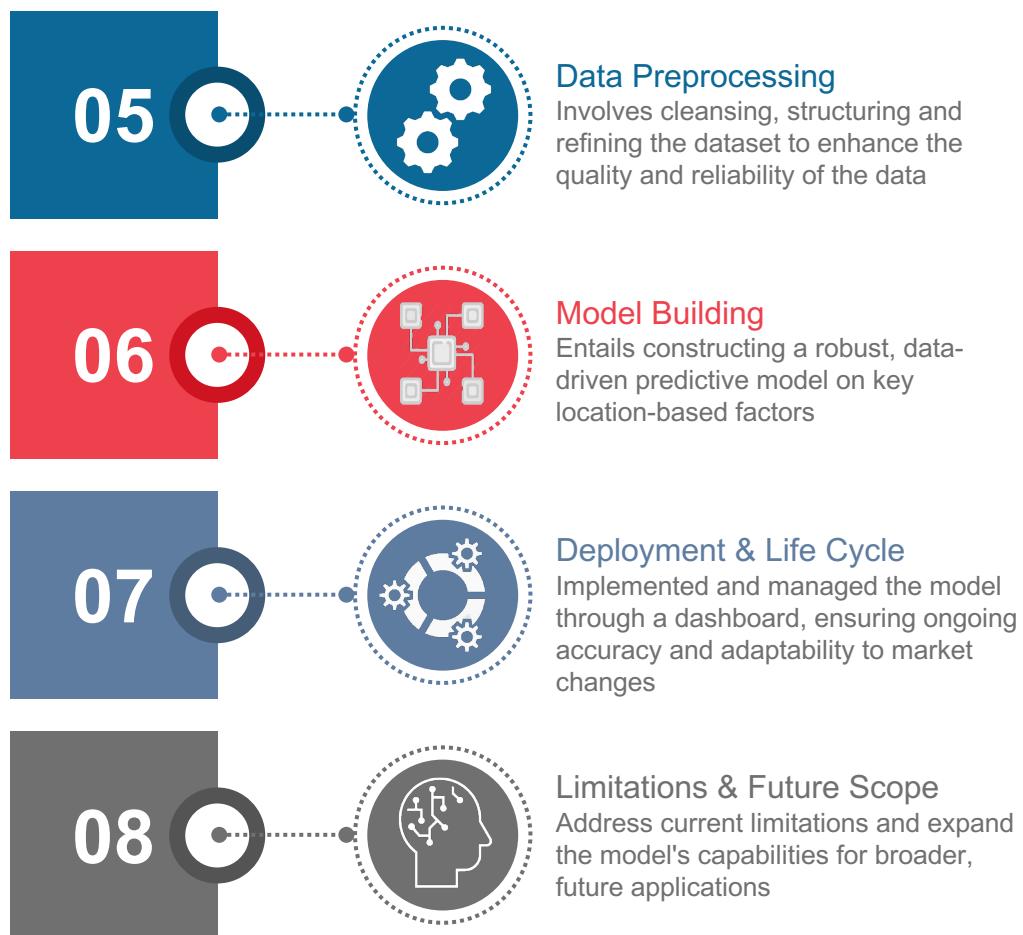
Email:
ssrimami@purdue.edu

LinkedIn:
www.linkedin.com/in/srinija-a-s

AGENDA

- 01**  **Project Scope**
To develop a scoring algorithm to identify optimal locations for Meijer's retail footprint expansion
- 02**  **Methodology**
Encompasses data collection, preprocessing and EDA, culminating in the development of a predictive scoring model for strategic site selection
- 03**  **Data**
Crafted by web-scraping with API's, focusing on geospatial and customer-centric attributes for Meijer and its competitors
- 04**  **Exploratory Data Analysis**
Conducted to identify trends, patterns, and correlations in data for feature selection and optimization

AGENDA



01

Project Scope

Defining project outline



Business Problem

To identify optimal locations for new mExpress stores across Michigan that maximize high foot traffic and revenue, with a vision for nationwide scalability.



Project Outcome

Help Meijer prioritize locations by their score, ensuring resources are allocated to the most promising sites.



Analytics Problem

To transform raw location-based data into actionable insights by analyzing Meijer and its competitor's Store level data (Geospatial + Customer Centric).



Key Deliverables

A scoring model that evaluates the potential success of any given location in terms of customer visits pursuing it as a location for Meijer's next mExpress.

02

Methodology

Analytical process framework



Meijer and Competitor Stores

Meijer Supercenters, M-Express store, and Competitor Stores



Select Store Attributes

Categorize attributes, such as Traffic, Nearby places, and County Demographics



Identify Locations

Identify existing and potential store locations for which data needs to be sourced



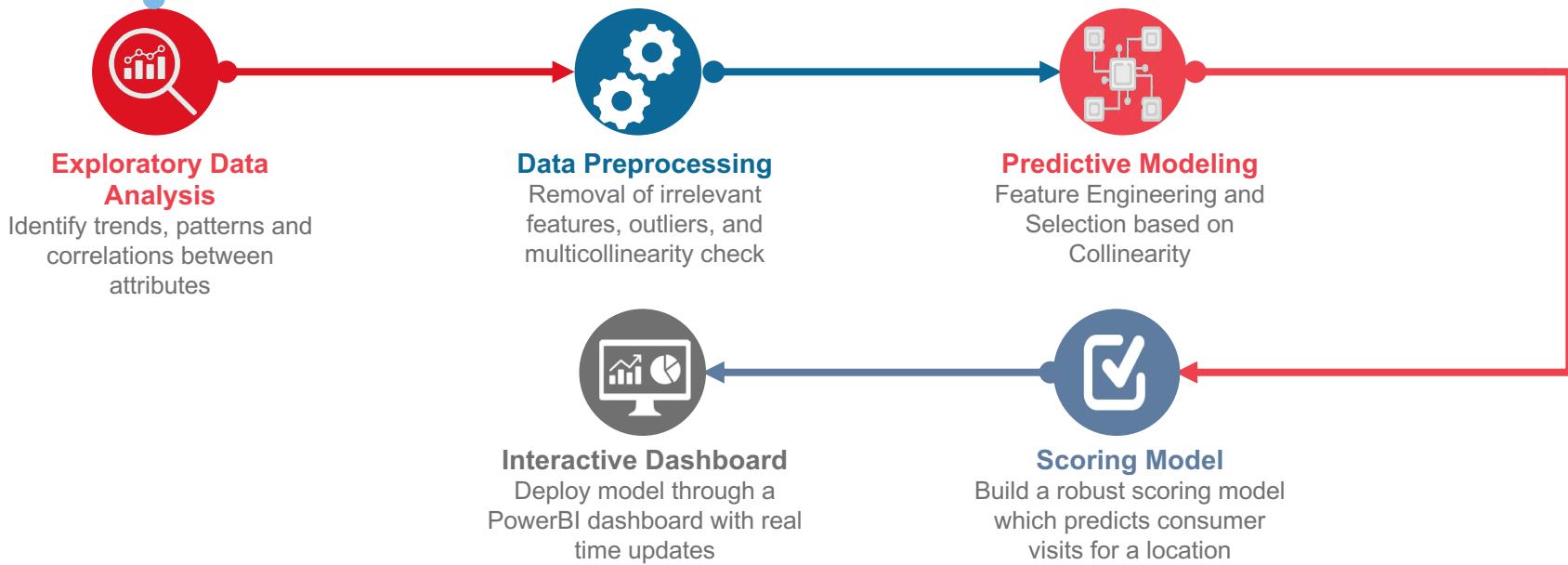
Data Frame Creation

Create separate data frames for Meijer and its Competitors as well as a master data frame



Web Scraping

Collect data using Python Libraries and APIs, such as Google Maps and Overpass API



03



Data

Collection of geospatial and customer-centric attributes

Geo Spatial Data



Traffic Volume
AADT



Distance to Highway
Overpass API,
GoogleMapsAPI



Duration to Highway
Overpass
API, GoogleMapsAPI



Nearby Places
OpenStreetMap, Overpass
API



Customer Sentiment
Yelp, Google

Customer Data



County Demographics
Geocodio



Traffic Volume AADT



Identify States



Data from Department
of Transportation



Identify Store
Latitude & Longitude



Annual Average
Daily Traffic

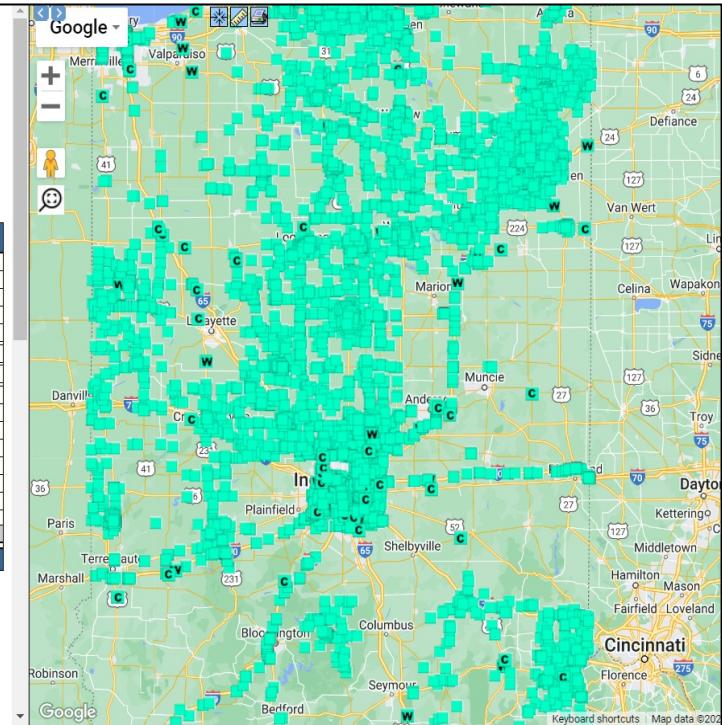
Indiana Department of Transportation **MS2**
Transportation Data Management System

[Home](#) [Login](#) [+ Locate](#) [+ Locate All](#) [Email This](#)
[Auto-Locate OFF](#)

Disclaimer: The data is provided pursuant to the Indiana Open Records Act. It represents accurate reproductions of the records on file with the Indiana Department of Transportation; however, [more](#)

List View All DRs

Record	◀	▶	1	▶	◀	of 8324	Goto Record	Go
Location ID: 100190							MPO ID:	210_031_1702
Type: SPOT							On HPMS:	Yes
On NHS: No							LRS Loc Pt.:	1.761588
LRS ID: 2000000310000001							SF Group:	U1_SWG
AF Group: U2_A							Route Type:	US Route
GF Group: U1_SWG							Route:	31
Class Dist Grp:							Active:	Yes
Seas Clss Grp:							Category:	
WIM Group:							QC Group:	JUR2SHORT
Fnc'l Class:	Principal Arterial (Freeway/Exprswy)						Milepost:	
Located On:								
Loc On Alias:	US 31							
More Detail ▶								
STATION DATA								
Directions:	2-WAY	NEG	POS	?				
	1	2	3	1	2	3		
AADT ?								
Year	AADT	DHV-30	K %	D %	PA	BC	Src	
2023	9,749				9,388 (96%)	360 (4%)		
2022	9,391 ³						Grown from 2021	
https://www.ms2soft.com								





Distance to Highway

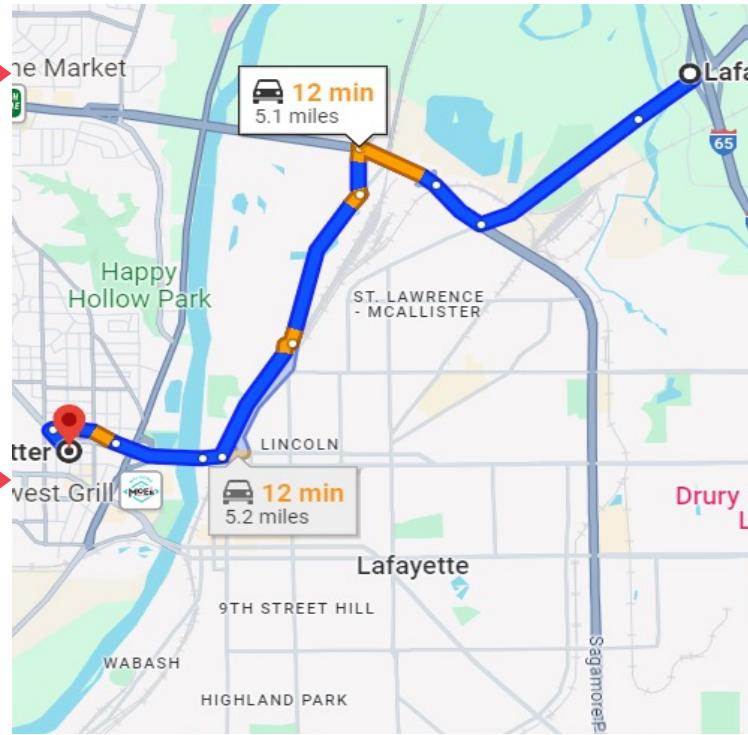
Overpass API,
GoogleMapsAPI



Coordinates of the
Potential Store Location



Driving Distance between
Highway Exit & Store



Duration to Highway¹⁰

Overpass
API, GoogleMapsAPI



Coordinates of the nearest
entry/exit ramp of a highway



Google Maps API

Duration from Highway
Exit to reach the Store





Nearby Places

OpenStreetMap, Overpass
API



County Demographics

Geocodio

Grocery Stores

Fuel Stations

1

Residential

Commercial

Educational

3

Hospitals

Apparel Stores

5

Hotels

Tourist

Airports



Population

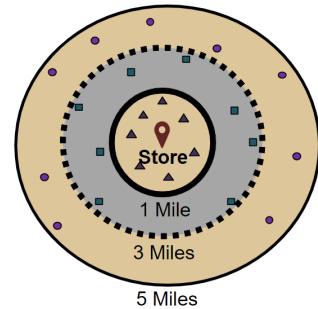
No. of Households

Income

Household Value

Age

Gender



Convenience store landscape

Michigan



Meijer Stores

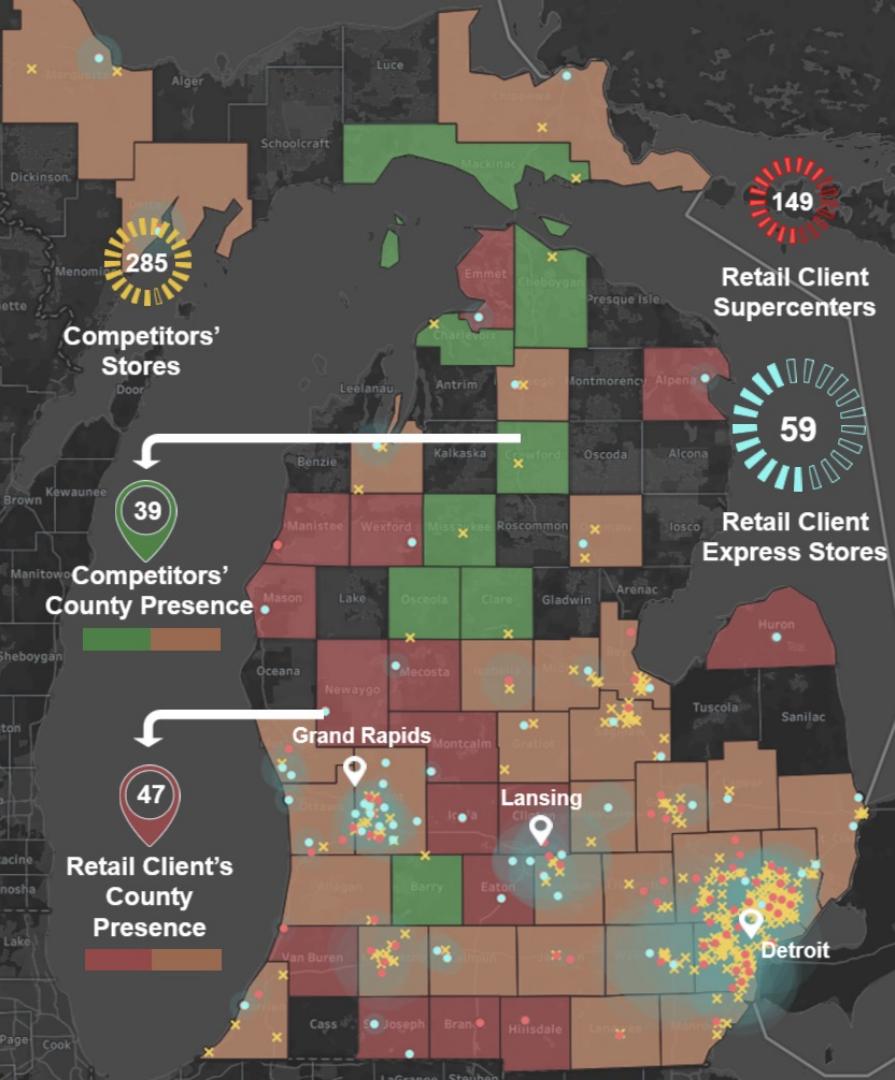
Identifying Meijer's presence in each of the counties in Michigan. Visualizing the extent and density of Meijer Supercenters as well as Express Stores.

meijer



Competitor Stores

Identifying Competitors' presence in each of the counties in Michigan. Visualizing the extent and density of their Stores.

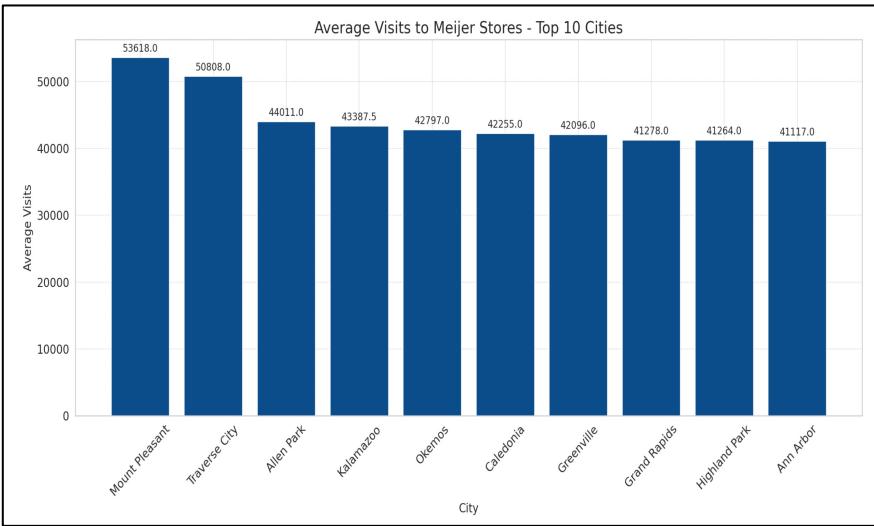


04

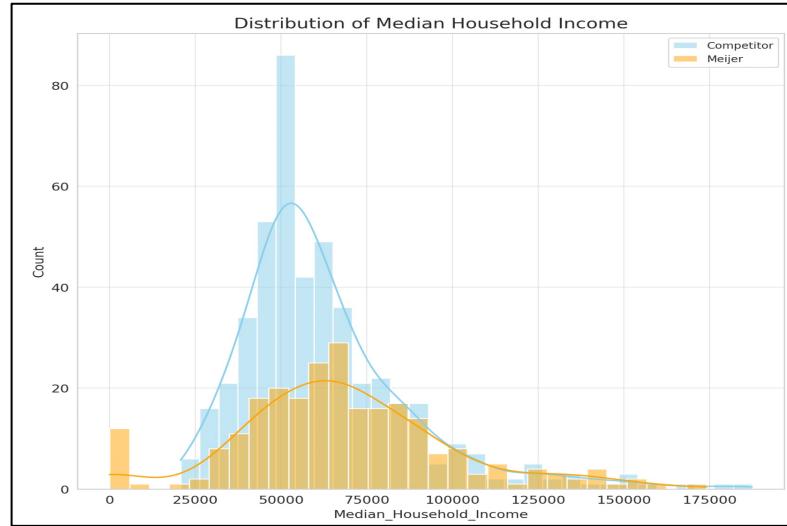


Exploratory Data Analysis

To identify trends, patterns, and correlations in data for feature selection and optimization



Mount pleasant has the highest average daily store visits followed by Traverse City



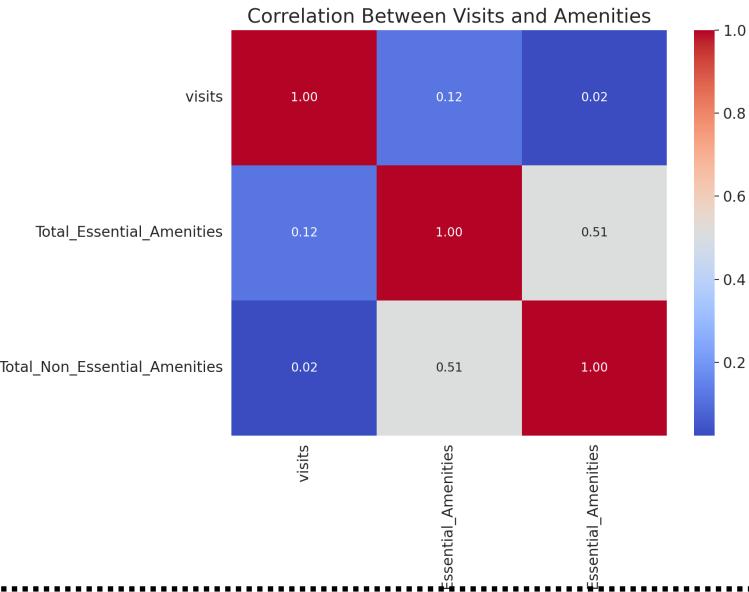
Both Meijer and its Competitors target the mid income level consumers. Although Meijer's target consumers come from a marginally higher household income bucket

04

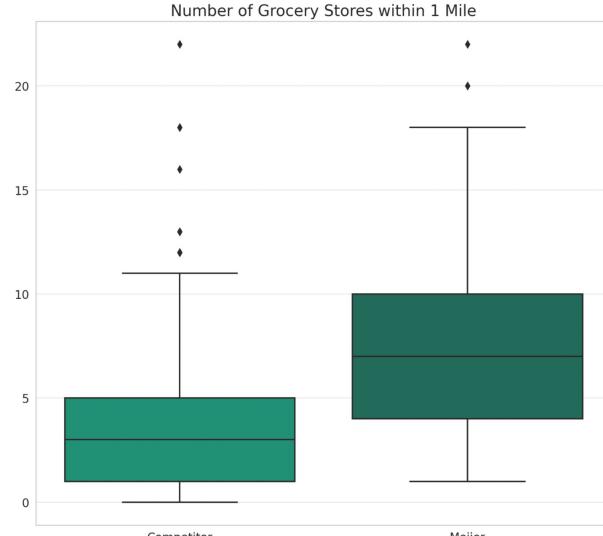


Exploratory Data Analysis

To identify trends, patterns, and correlations in data for feature selection and optimization



Presence of Essential amenities (Education, Hospitals & Fuel) is more likely to garner visits than non-essential visits



Meijer locations tend to have a slightly wider distribution, suggesting some locations are in more competitive areas than its competitors. (Reason for the higher visits – Anchoring effect)

05



Data Preprocessing

Enhances the quality and reliability of the data, paving the way for more accurate and meaningful insights

Why?



Enhance Data Quality

Removing irrelevant and highly correlated columns improves the overall quality and reliability of the dataset



Reduce Overfitting Risk

By eliminating multicollinearity and correlated features, we lead to more generalized and robust models



Ensure Data Integrity

Removing nulls from columns helps in maintaining the integrity of the dataset, ensuring that the analysis is not skewed

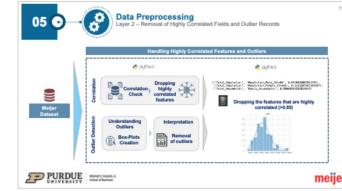
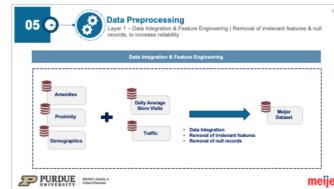


Removal of Outliers

Removing outliers can enhance model accuracy by reducing noise and ensuring the training process focuses on the more representative patterns of the data

Layered Approach

Data Preprocessing Framework

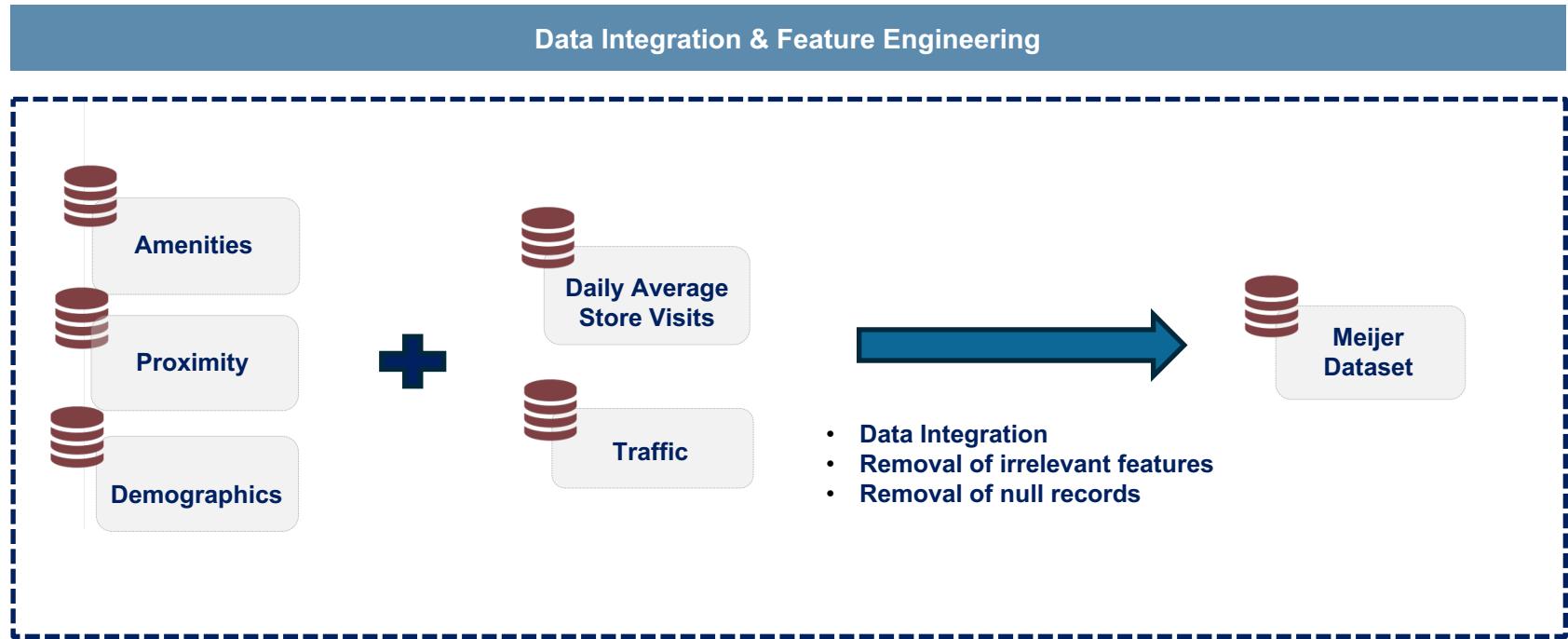


05



Data Preprocessing

Layer 1 – Data Integration & Feature Engineering | Removal of irrelevant features & null records, to increase reliability

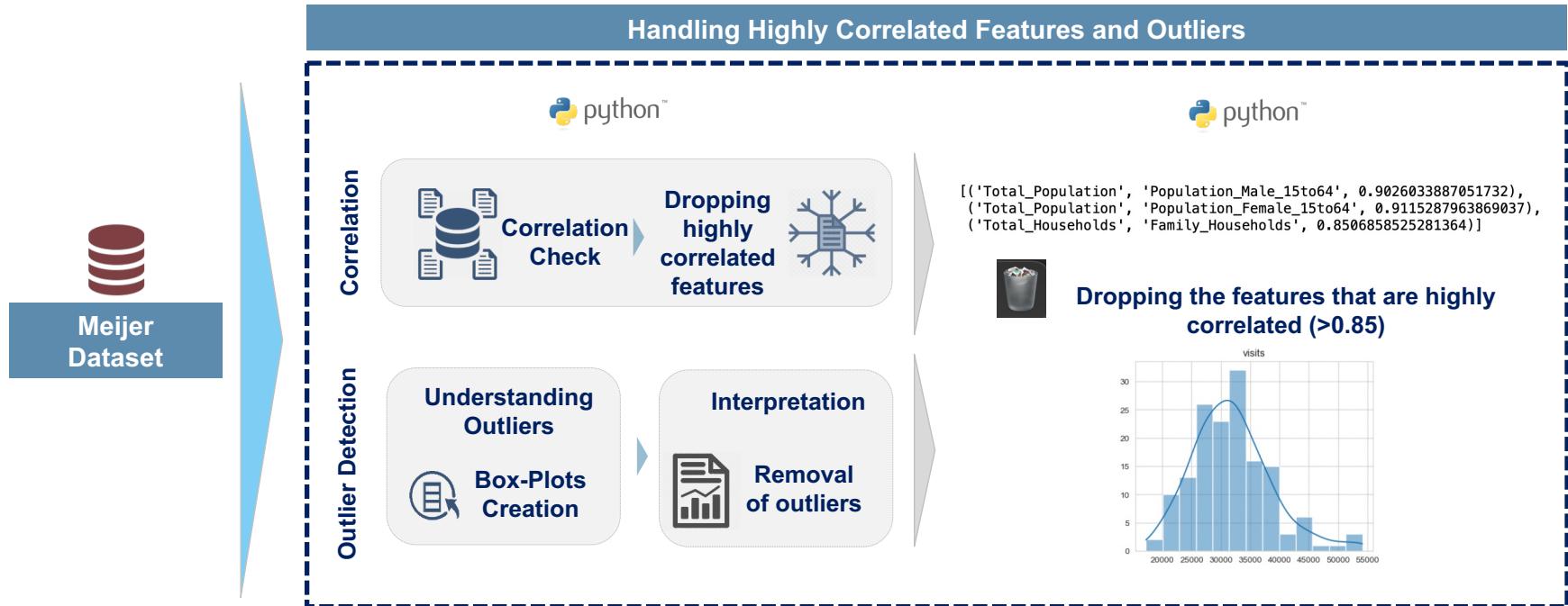


05



Data Preprocessing

Layer 2 – Removal of Highly Correlated Fields and Outlier Records



05



Data Preprocessing

Layer 3 – Multicollinearity Check | Removal of multicollinear features to increase the model accuracy and to mitigate the issue with correlated predictor variables

Multicollinearity Check

	Variable	VIF		Variable	VIF
13	Family_Households	inf		Population_Female_15to64	13.038979
14	Non_Family_Households	inf		Population_Male_15to64	12.257272
25	Population_Female_65+	inf		Grocery_Stores_1mile	11.514007
24	Population_Female_15to64	inf	0	Fuel_Stations_Car_Wash_1mile	8.374988
23	Population_Female_Under15	inf	5	Population_Female_65+	7.624846
22	Population_Male_65+	inf	20	Population_Male_65+	7.509603
21	Population_Male_15to64	inf	17	Food_Outlets_1mile	7.407307
20	Population_Male_Under15	inf	1	Population_Female_Under15	6.674956
19	Household_Income_200k+	inf	18	Household_Income_100to200k	5.424931
18	Household_Income_100to200k	inf		Household_Income_50to100k	5.383863
17	Household_Income_50to100k	inf		Population_Male_Under15	4.558936
11	Total_Population	inf	13	Household_Income_Under50k	3.919316
12	Total_Households	inf	12	Airports_5miles	3.356540
16	Household_Income_Under50k	inf	15	Hospitals_3miles	3.029923
0	Grocery_Stores_1mile	11.862665	11	Hotels_5miles	2.296802
15	Median_Household_Income	9.900051	10	Household_Income_200k+	2.179150
5	Fuel_Stations_Car_Wash_1mile	8.667405	6	Tourist_Destinations_5miles	2.003704
1	Food_Outlets_1mile	7.592041	9	Apparels_3miles	1.961856
10	Airports_5miles	3.490082	14	Commercial_Housing_1mile	1.700336
6	Hospitals_3miles	3.035254	8	Educational_Institutions_1mile	1.550978
9	Hotels_5miles	2.379762	7	driving distance to nearest motorway (in miles)	1.360390
8	Tourist_Destinations_5miles	2.003834	4	Residential_Housing_1mile	1.292327
7	Apparels_3miles	1.992710	21		
4	Commercial_Housing_1mile	1.730377			
2	Educational_Institutions_1mile	1.565611			
26	driving distance to nearest motorway (in miles)	1.439911			
3	Residential_Housing_1mile	1.346156	3		

06



Model Building

Stage 1 | Predictive Model Building

Data Preparation

One-Hot Encoding

Train – Test Split

Scaling

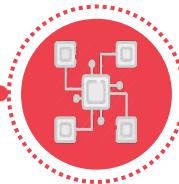
Recursive Feature
Elimination

Hyperparameter Tuning using GridSearchCV

```
param_grid = {  
    'n_estimators': [100, 150, 200],  
    'learning_rate': [0.01, 0.05, 0.1, 0.15],  
    'max_depth': [3, 7, 10],  
    'min_child_weight': [0, 5, 10],  
    'gamma': [0, 0.2, 0.4],  
    'subsample': [0, 0.6, 0.75, 1.0],  
    'lambda': [0, 0.75, 1.5, 2],  
    'alpha': [0, 0.5, 1.0]  
}
```

```
Fitting 3 folds for each of 15552 candidates, totalling 46656 fits  
Best Parameters Found: {'alpha': 0.5, 'gamma': 0, 'lambda': 1.5, 'learning_rate': 0.05, 'max_depth': 10, 'min_chil  
d_weight': 0, 'n_estimators': 100, 'subsample': 0.75}
```

06



Model Building

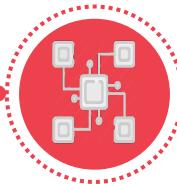
Stage 1 | Performance Metrics



Results for Meijer Stores

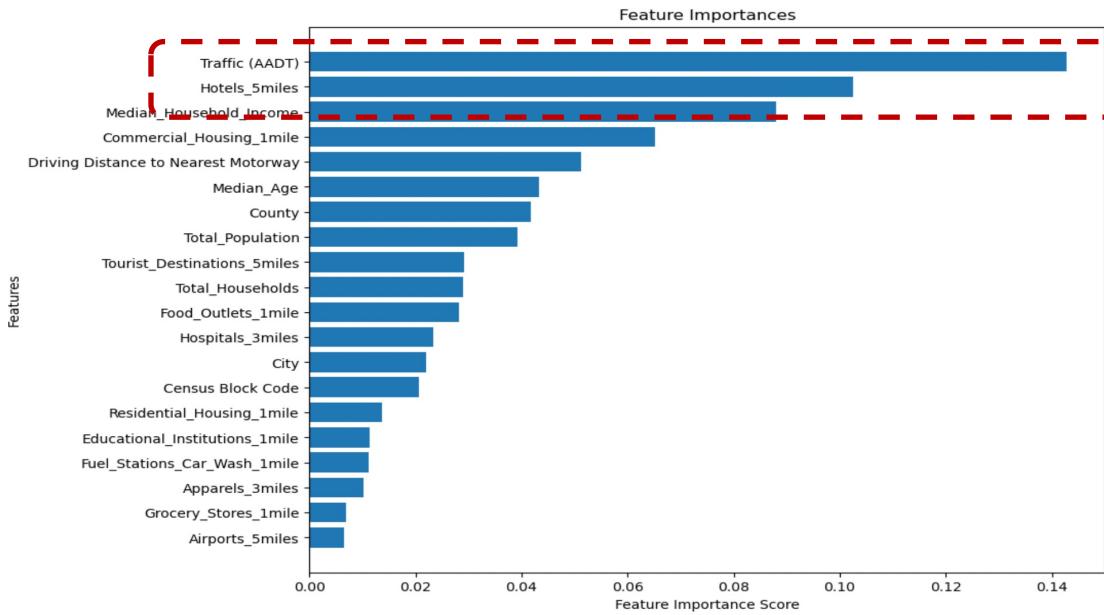
XG Boost Regressor (With the best parameters)	Method	RMSE	MAE	MAPE
	Train-Test (70-30)	6335.45	4779.50	15.80%
	6-Fold Cross Validation	6503.98	5062.84	16.49%

06

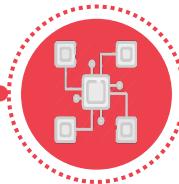


Model Building

Stage 1 | Predictive Model Building



06



Model Building

Stage 2 | Heuristic Scoring Model



Traffic



Proximity



Essential Amenities



Non-Essential Amenities



Demographics

25%



25%



20%



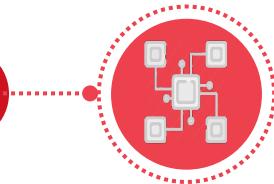
10%



20%

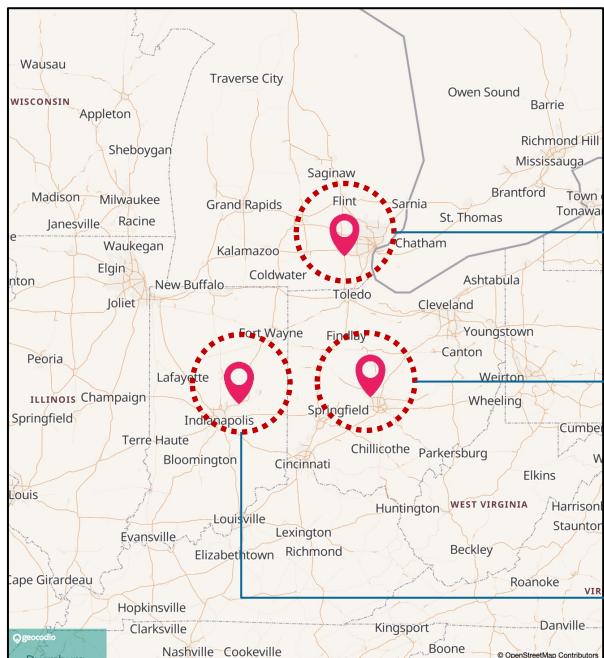
- Normalized traffic volume and distances to highways to a 0-5 scale, highlighting the importance of easy access.
- Essential and non-essential amenities are evaluated separately, with scores adjusted to reflect their relative importance.
- Population dynamics within a specific age and gender demographic are scored for potential market size.
- The final location suitability score is a weighted sum of all components, balanced according to predetermined weights.

06



Model Building

Stage 3 | Predicting Shortlisted Site Visits



Michigan
Predicted Visits: 32929
Heuristic Score: 3.752
Traffic (AADT): 9308

Ohio
Predicted Visits: 32816
Heuristic Score: 3.487
Traffic (AADT): 2390

Indiana
Predicted Visits: 30717
Heuristic Score: 3.104
Traffic (AADT): 2077

07



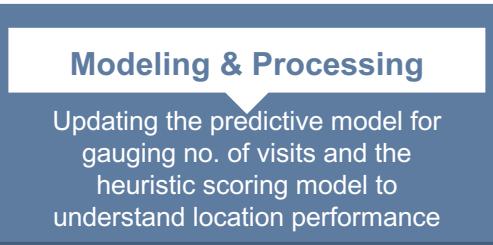
Deployment and Life Cycle

Microsoft Azure

 aws²⁴

Data Integration & Storage

Building data frameworks in spreadsheets or SQL for Meijer and Competitor stores



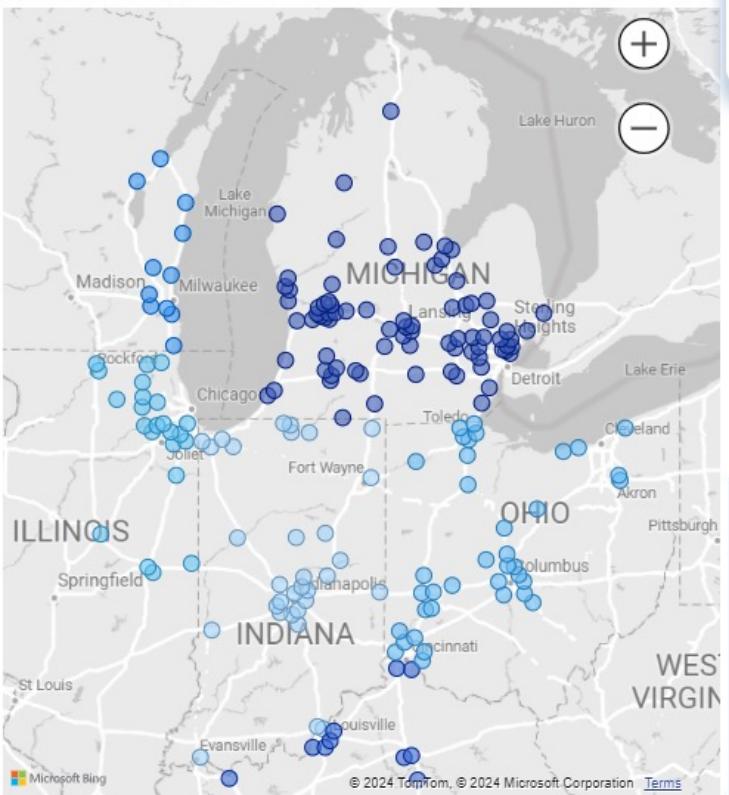
Advanced Real Estate Analytics for Retail

[Reset All Filters](#)

State	County	City	Store ID	Latitude	Longitude
All	All	All	All	All	All

Store Locations

State ● IL ● IN ● KY ● MI ● OH ● WI



30.44K

Store Visits (Predicted)

2.4

Location Score (Calculated)

16.07K

Annual Avg Daily Traffic

196

No. of Stores

7.94

Duration to Highway (Mins)

41.77

Median Age

3.96

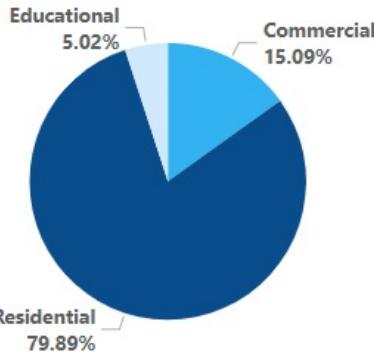
Distance to Highway (Miles)

73.90K

Median Household Income

Within 1 Mile Radius of the Store

● Commercial ● Residential ● Educational



Food Outlets

1601

3172

Hospitals

1406

1406

Fuel Stations

979

924

Airports

861

597

Tourist D...

720

597

08



Limitations and Future Scope

Address current limitations and expand the model's capabilities for broader, future applications

01

Web-Scraping Geospatial Data

- Traffic AADT data could also be obtained for the remaining states in the US, although not all states offer a direct download option.
- It is crucial to verify the accuracy of data obtained from free APIs, such as Overpass, to ensure reliability.



02

Customer Data

The inclusion of features related to customer sentiments, such as Google and Yelp reviews, was considered but not implemented due to the exhaustion of free API trial search calls, preventing the scraping of reviews for all Meijer stores.



03

Model Performance & Accuracy

Enhancing the model's accuracy is achievable by expanding the dataset to include more no. of stores from competitors, effectively increasing the size of the training dataset and subsequently of the R-Squared metric to explain variability.



Thank You