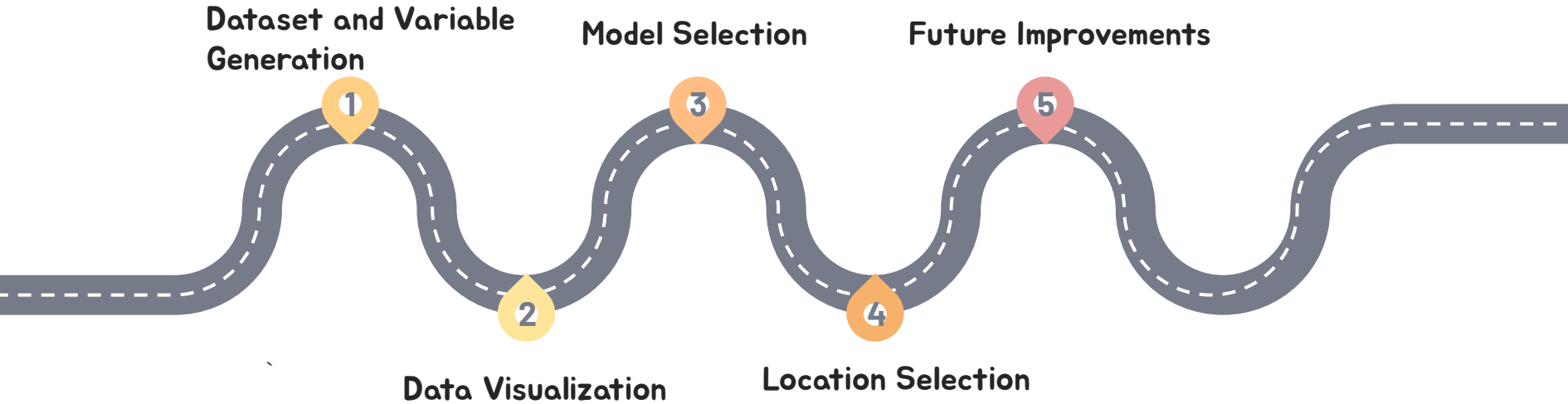
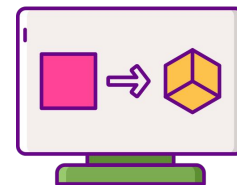
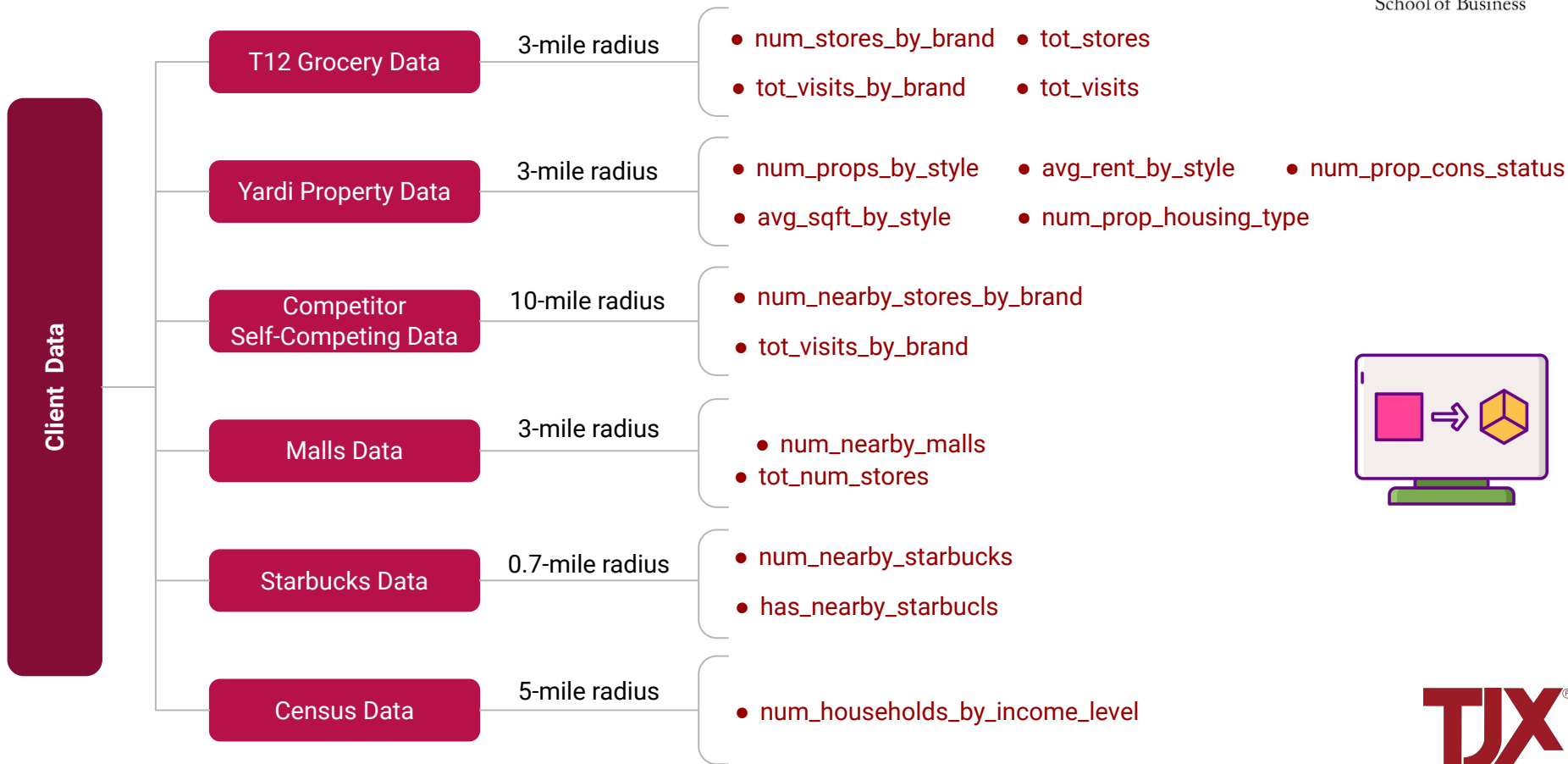


Roadmap



Datasets and Variable Generation



Objectives

Finding locations for future HomeGoods, Marshalls, and T.J. Maxx stores which can generate **optimal annual visits** in 3 states:

Alabama

5

T.J. maxx

3

Marshalls

3

HomeGoods

Florida

10

T.J. maxx

10

Marshalls

10

HomeGoods

Tennessee

5

T.J. maxx

3

Marshalls

3

HomeGoods

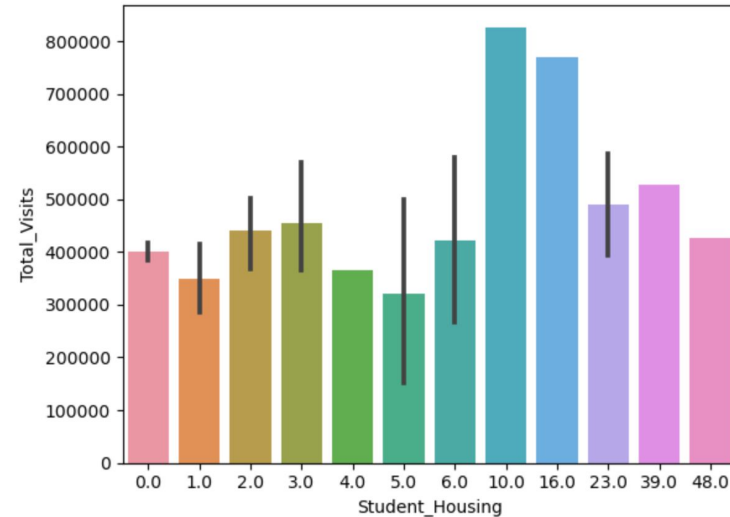
52 new stores



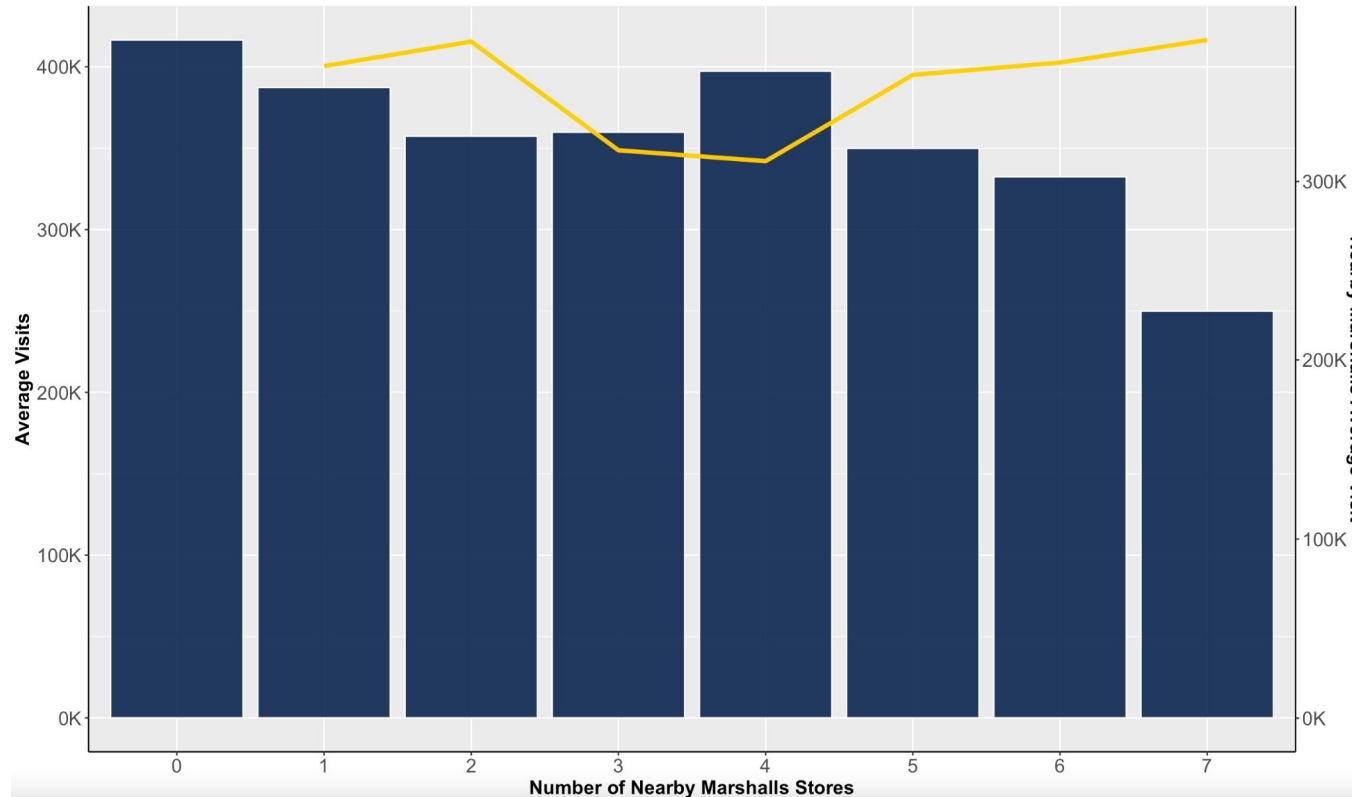
Visualizations

	Marshalls	HomeGoods	TJMax
	Client_Visits	Client_Visits	Client_Visits
num_stores_Aldi	-0.14	0.32	0.05
num_stores_Publix	-0.12	0.21	0.15
num_stores_The_Fresh_Market	-0.07	0.22	0.20
num_stores_Total_Wine_&_More	0.03	0.12	0.05
num_stores_Trader_Joe's	-0.07	0.06	-0.01
num_stores_Walmart_Neighborhood_Market	-0.17	0.13	0.10
num_stores_Whole_Foods_Market	-0.01	0.12	-0.13
num_stores_Winn-Dixie	-0.20	0.00	-0.03
num_stores_ABC_Fine_Wine_&_Spirits	-0.03	0.01	-0.10
num_stores_Sprouts_Farmers_Market	-0.01	0.05	-0.04
Total_Grocery_Visit	-0.13	0.17	0.15
num_Discretionary	0.03	-0.06	0.00
num_High-Mid-Range	-0.09	0.11	-0.08
num_Low-Mid-Range	-0.14	-0.15	0.07
num_Workforce	-0.04	-0.11	0.02
avg_sqft_Discretionary	0.09	0.65	0.45
avg_sqft_High_Mid-Range	0.30	0.17	0.16
avg_sqft_Low-Mid-Range	0.10	-0.21	0.31
avg_sqft_Workforce	-0.09	-0.15	-0.08
avg_rent_Discretionary	-0.13	0.39	0.01
avg_rent_High_Mid-Range	0.10	0.02	-0.09
avg_rent_Low-Mid-Range	0.01	-0.02	0.07
avg_rent_Workforce	0.02	-0.16	-0.07
Student_Housing	0.11	-0.06	0.32
Affordable_Housing	-0.09	-0.10	-0.07
Military_Housing		0.14	0.04
Senior_Housing	-0.09	-0.02	-0.07
Adaptive_Reuse	0.03	-0.06	-0.06
Single_Family_Rental	-0.08	0.24	-0.05
Fractured_Condo	-0.03	-0.17	0.00

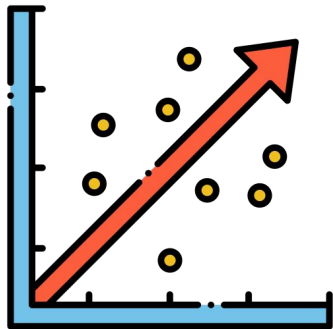
Number of Student Housing Nearby



Marshalls Store Loses Its Customers To Its Nearby Stores

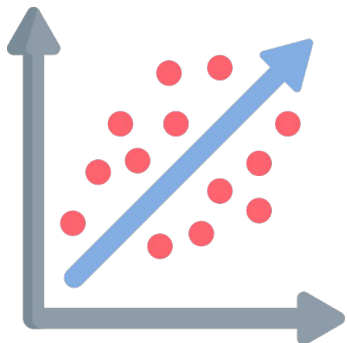


Modeling: Regression Models Don't Work



OLS Regression !

- Training R-square: 58% - 64%
- Predicting Accuracy: 50% - 60%



Lasso Regression !

- Training R-square: <10%
- Predicting Accuracy: 68% - 79%

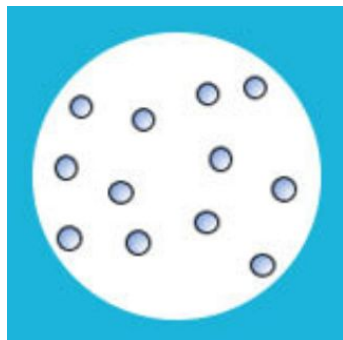
OLS Overfitting

Due to having too many variables, but not enough observations

Lasso Low R-square

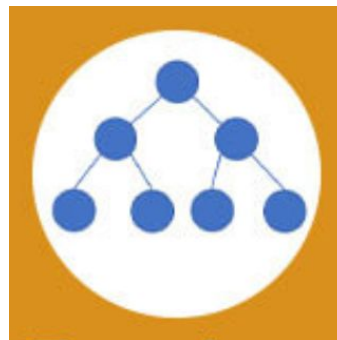
Lasso R-square of testing split is negative, plus Lasso R-square of training split is very low → Unreliable

Modeling: Tree-based Models



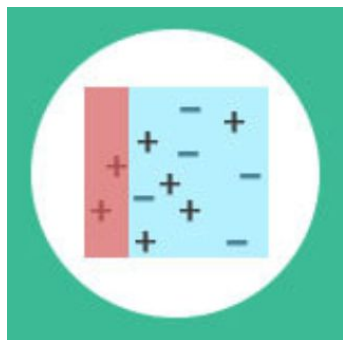
Gradient Boosting !

- Training Accuracy: ~98%
- Predicting Accuracy: 58% - 72%



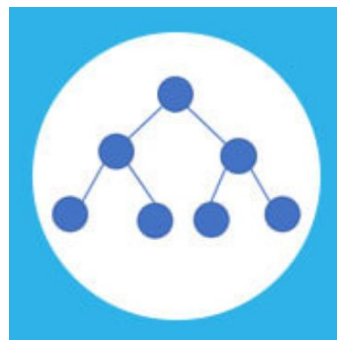
Random Forest ✓

- Training Accuracy: ~81%
- Predicting Accuracy: 68% - 80%



AdaBoost !

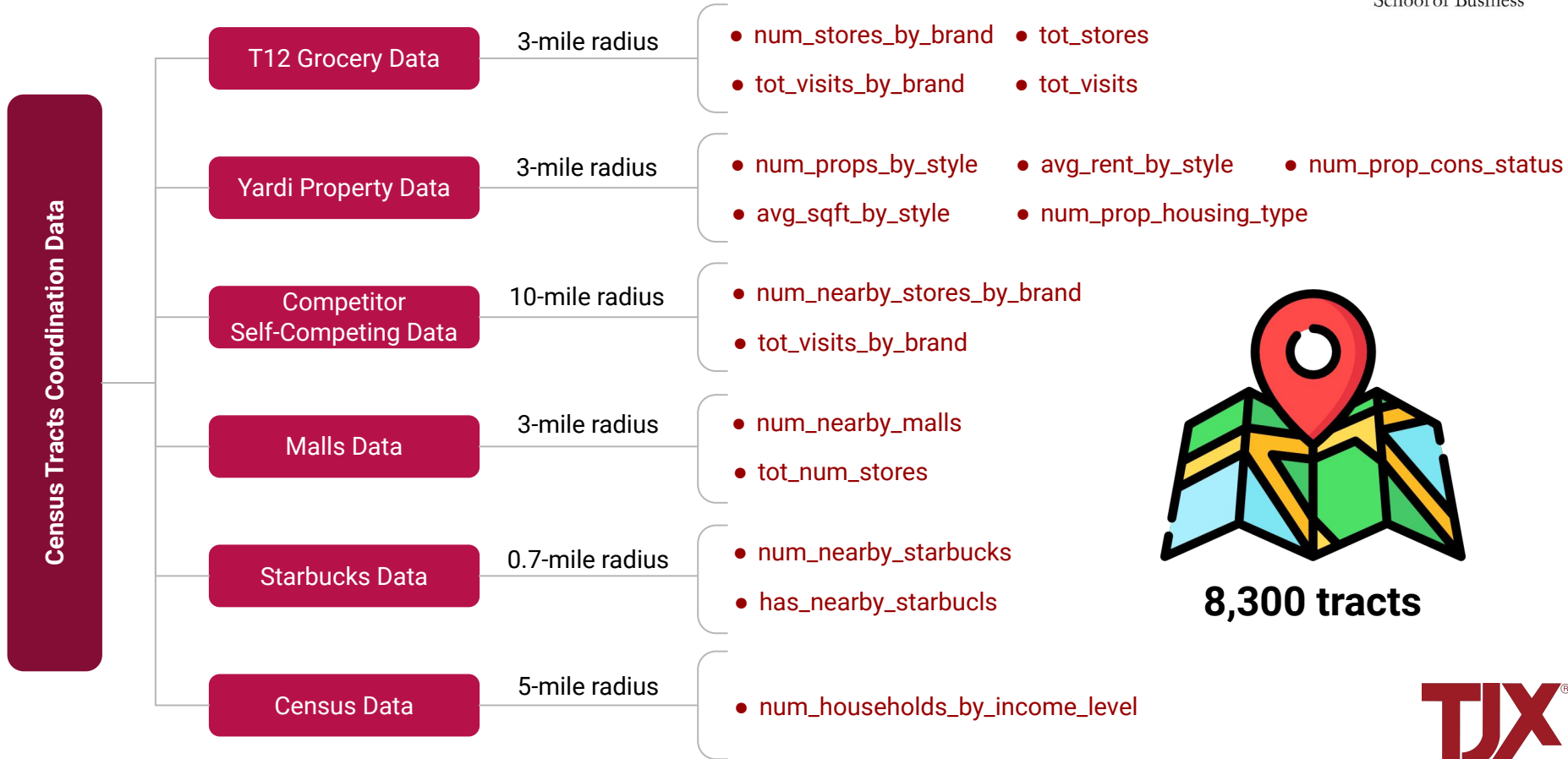
- Training Accuracy: 90% - 96%
- Predicting Accuracy: 65% - 76%



XGBoost !

- Training Accuracy: ~100%
- Predicting Accuracy: 63% - 74%

Prepare Predicting Dataset at Tract Level



Site Selection: Strategies



Populated, metro areas, near campuses, malls, and highways



Avoid cities of high similar store density and open to new cities to expand the store network

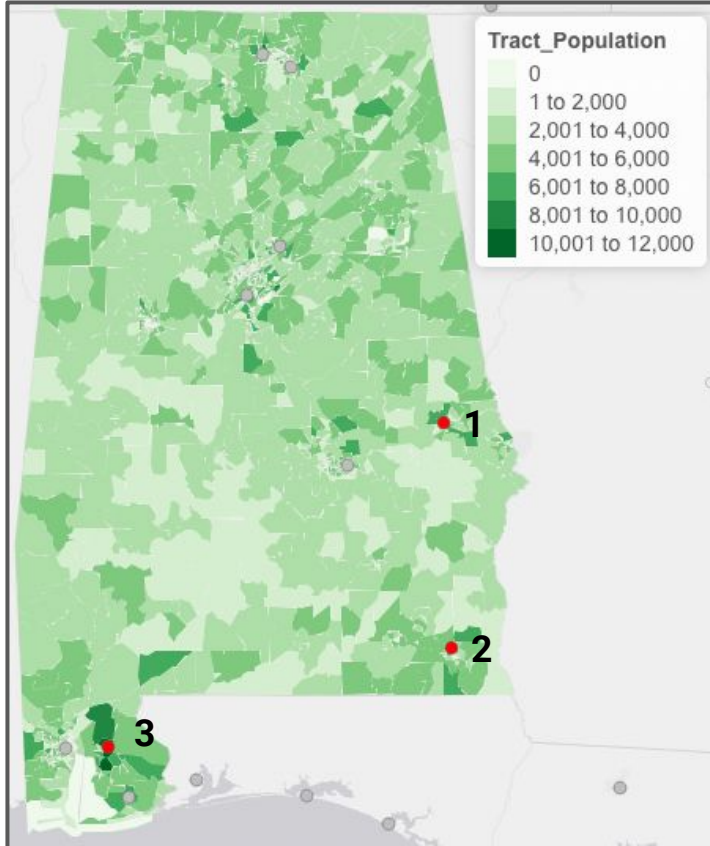


Same brand selected sites should be 10 miles away, in different counties



Try to optimize selections with highest predicted number of visits

Site Selection: Alabama - Marshalls

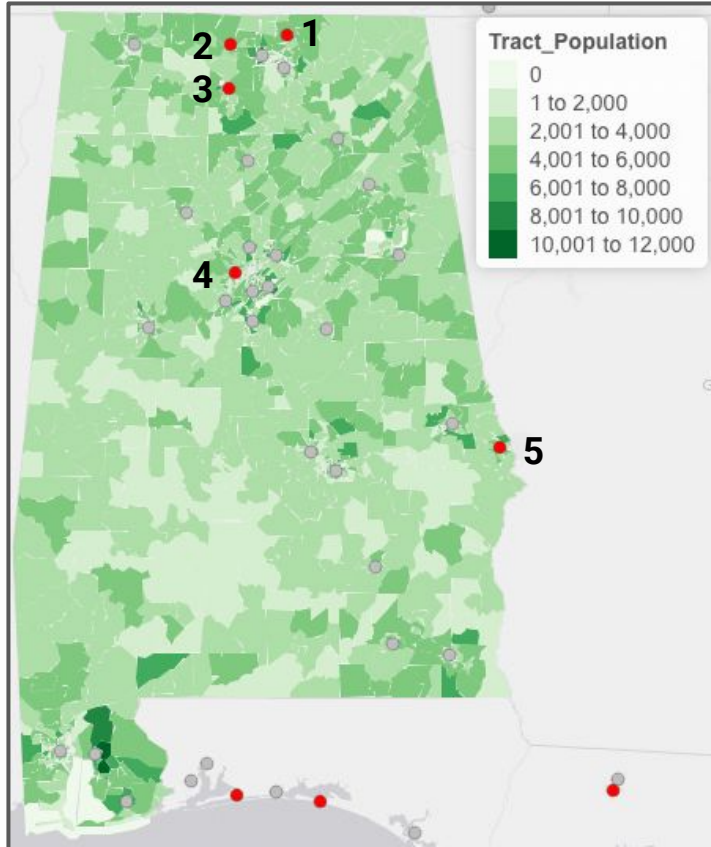


- | | |
|--|---------|
| 1. Auburn (32.6130368,-85.4808959) | 459,695 |
| - 1 T.J.Maxx nearby | |
| 2. Dothan (31.2755021,-85.4348261) | 431,699 |
| - 1 T.J.Maxx and 1 HomeGoods nearby | |
| 3. Spanish Fort (30.6835581,-87.8297072) | 544,369 |
| - 1 T.J.Maxx nearby | |

430,030

Current Average
Marshalls visits

Site Selection: Alabama - T.J. Maxx

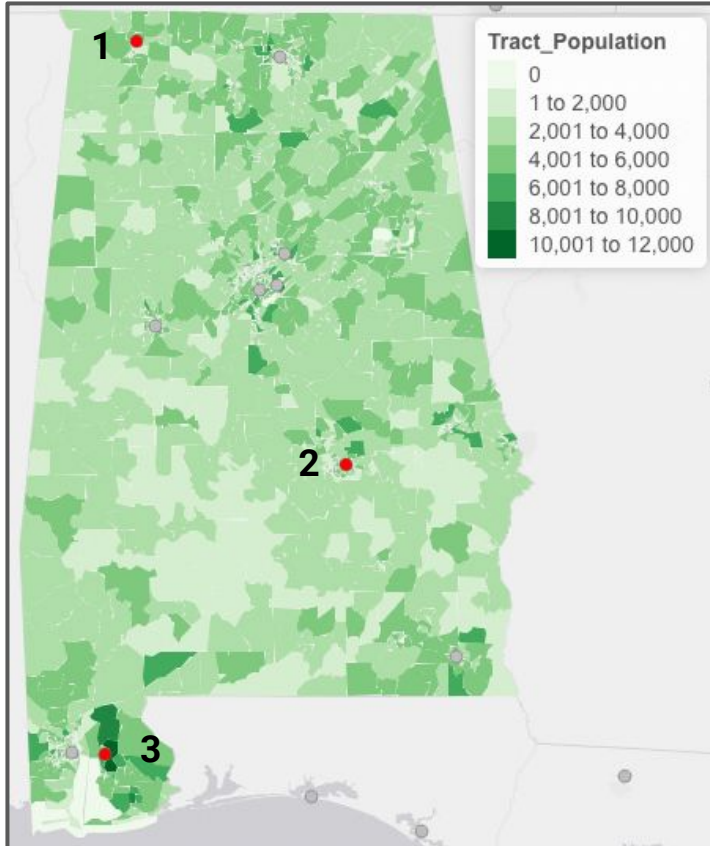


- | | |
|---|---------|
| 1. Meridianville (34.8774654,-86.5629052) | 476,513 |
| 2. Athens (34.8205875,-86.9517709) | 472,847 |
| 3. Decatur (34.5712507,-86.9706149) | 627,805 |
| 4. Birmingham (34.5712507,-86.9706149) | 531,529 |
| 5. Phenix City (34.5712507,-86.9706149) | 525,261 |

512,641

Current Average
T.J. Maxx visits

Site Selection: Alabama - HomeGoods



1. Florence (34.8344088,-87.662508)

338,598

- 1 T.J.Maxx nearby

2. Montgomery (32.3716245,-86.2001367)

348,124

- 1 T.J.Maxx and 1 Marshalls nearby

3. Daphne (30.6477601,-87.8838007)

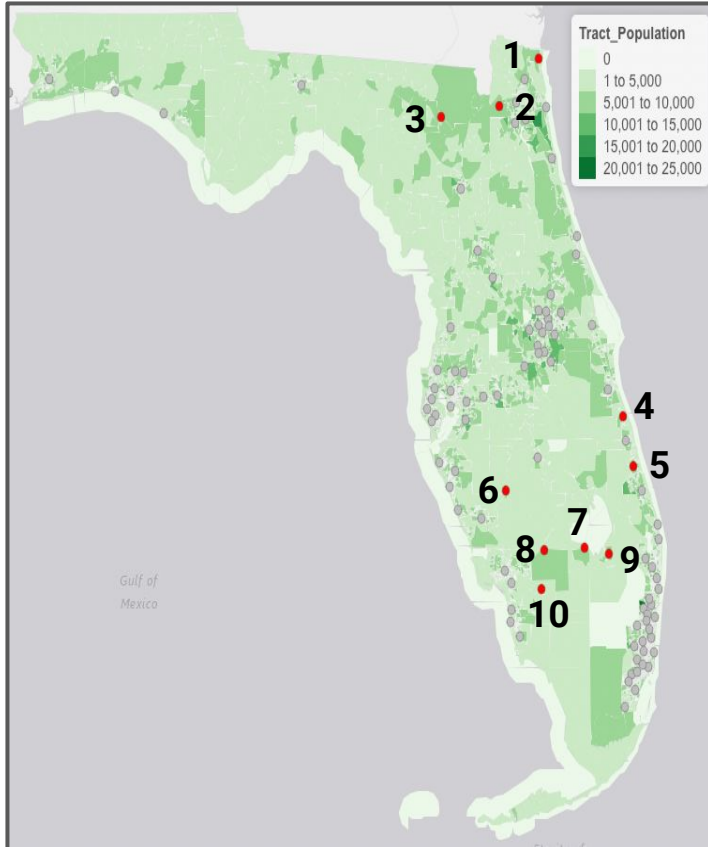
362,571

- 1 T.J.Maxx nearby

335,477

Current Average
HomeGoods visits

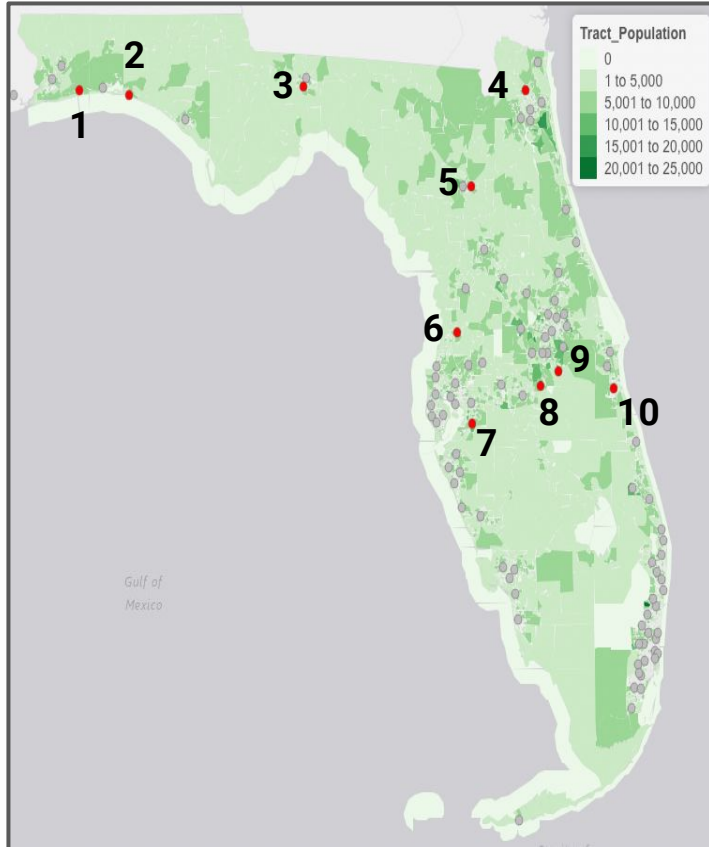
Site Selection: Florida - Marshalls



- | | |
|--|---------|
| 1. Fernandina Beach (34.8774654,-86.5629052) | 494,715 |
| - 1 T.J.Maxx and 1 HomeGoods nearby | |
| 2. Jacksonville (30.2737719,-81.9400913) | 537,915 |
| - 1 HomeGoods nearby | |
| 3. Lake City (30.188127,-82.6226571) | 405,083 |
| 4. Sebastian (27.825037,-80.4855166) | 531,842 |
| 5. Fort Pierce (27.4671297,-80.3674548) | 533,489 |
| - 1 HomeGoods nearby | |
| 6. Arcadia (27.2340401,-81.861203) | 524,197 |
| 7. Clewiston (26.7789195,-80.9336042) | 533,954 |
| 8. Port LaBelle (26.7527222,-81.4118758) | 507,572 |
| 9. Belle Glade (26.7302019,-80.6509221) | 538,419 |
| 10. Immokalee (26.4371756,-81.4408339) | 510,789 |

384,745 Current Average Marshalls visits

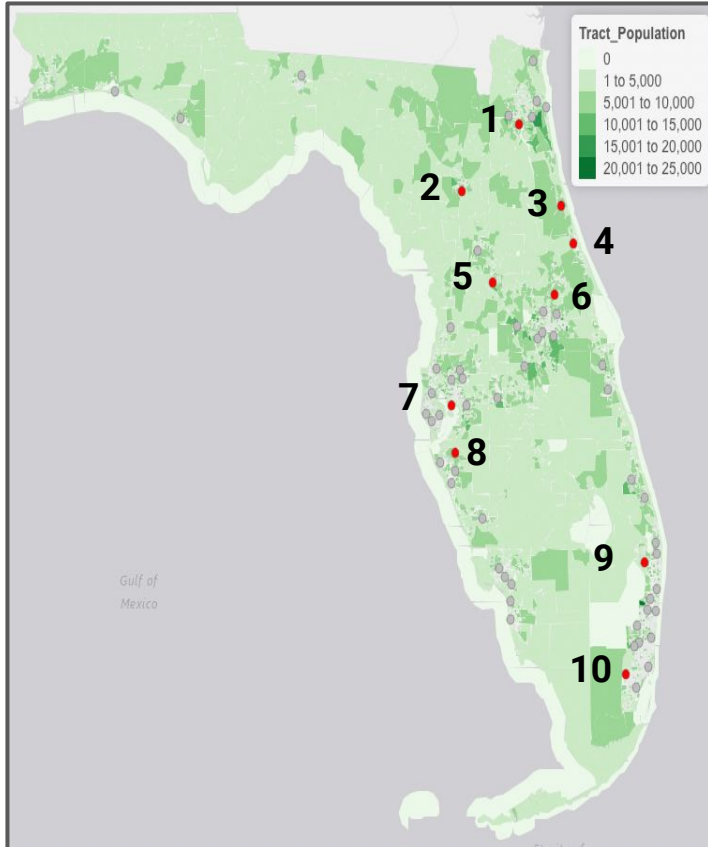
Site Selection: Florida - T.J. Maxx



- | | |
|---|---------|
| 1. Navarre (30.406463,-86.9104493) | 503,112 |
| 2. Miramar Beach (30.370037,-86.329751)
- 1 HomeGoods and 1 Marshalls nearby | 480,251 |
| 3. Tallahassee (30.4384196,-84.2910109)
- TJX's brands are nearby | 586,867 |
| 4. Jacksonville (30.4096454,-81.6861334)
- 2 Marshalls nearby | 455,783 |
| 5. Gainesville (29.6568455,-82.3169268)
- 1 T.J. Maxx and 1 Marshalls nearby | 516,161 |
| 6. Brooksville (28.5101164,-82.4899396)
- 1 HomeGoods and 1 Marshalls nearby | 493,723 |
| 7. Wimauma (27.7845132,-82.3053481)
- 1 HomeGoods and 1 Marshalls nearby | 505,945 |
| 8. Kissimmee (28.082933,-81.5027699) | 484,050 |
| 9. St Cloud (28.1933498,-81.2971483)
- 1 Marshalls nearby | 473,791 |
| 10. West Melbourne (28.0642764,-80.6459155)
- 1 Marshalls nearby and 1 HomeGoods nearby | 468,691 |

425,768 Current Average T.J. Maxx visits

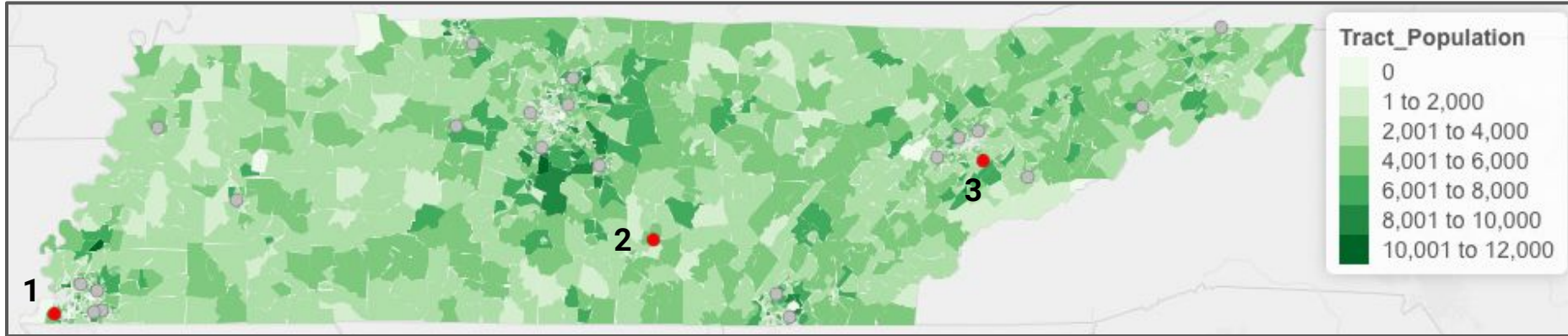
Site Selection: Florida - HomeGoods



- | | |
|--|---------|
| 1. Lakeside (30.1329919,-81.7104913) | 364,926 |
| - 1 HomeGoods, 2 Marshalls, and 2 T.J. Maxx nearby | |
| 2. Gainesville (29.612285,-82.3770724) | 480,251 |
| - 1 T.J. Maxx and 1 Marshalls nearby | |
| 3. Palm Coast (29.4933524,-81.2074328) | 350,989 |
| - 1 T.J. Maxx nearby | |
| 4. Daytona Beach (29.2030861,-81.0643388) | 332,260 |
| - 2 Marshalls and 1 T.J. Maxx nearby | |
| 5. The Villages (28.891737,-82.0102088) | 358,872 |
| - 1 T.J. Maxx and 1 Marshalls nearby | |
| 6. Sanford (28.796322,-81.2839486) | 352,155 |
| - 1 T.J. Maxx and 1 Marshalls nearby | |
| 7. Tampa (27.9168559,-82.5003139) | 336,444 |
| - 2 T.J. Maxx and 2 Marshalls nearby | |
| 8. Bradenton (27.5370315,-82.4522426) | 363,540 |
| - 1 T.J. Maxx nearby | |
| 9. Wellington (26.6588261,-80.2376705) | 382,735 |
| - 2 T.J. Maxx and 2 Marshalls nearby | |
| 10. West End (25.7494715,-80.4567622) | 358,597 |
| - 4 T.J. Maxx and 3 Marshalls nearby | |

341,634 Current Average Marshalls visits

Site Selection: Tennessee - Marshalls

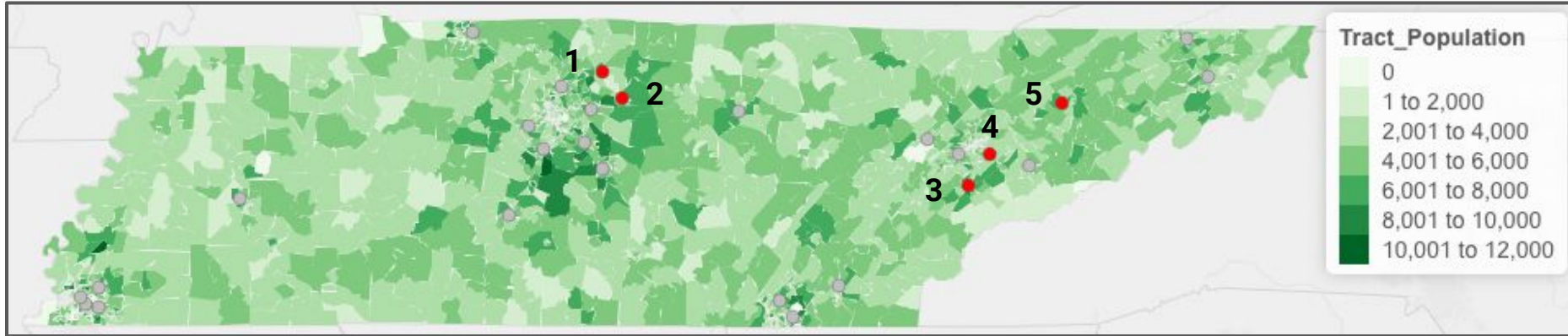


- | | |
|--|---------|
| 1. Memphis (35.0452222,-90.074135) | 467,473 |
| 2. Manchester (35.4561774,-86.0698831) | 531,009 |
| 3. Knoxville (35.8806407,-83.8611812) | 527,138 |

383,480

Current Average
Marshalls visits

Site Selection: Tennessee - T.J. Maxx

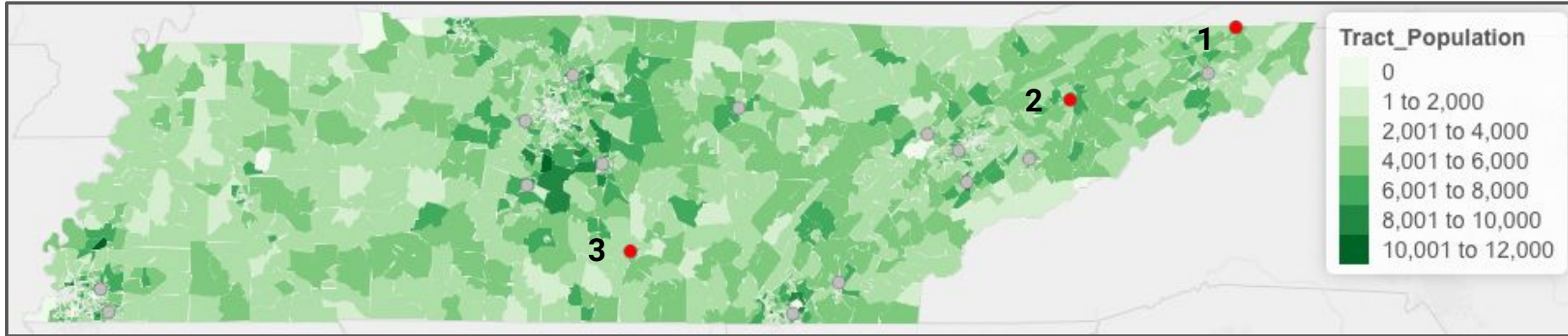


1. Gallatin (36.3704939,-86.4302447)	524,278
2. Lebanon (36.2259424,-86.2993944)	433,867
3. Maryville (35.7539049,-83.9833371) - 1 HomeGoods nearby	529,280
4. Knoxville (35.9216627,-83.834614) - 1 Marshalls nearby	518,699
5. Morristown (36.2019729,-83.356935)	527,214

519,313

Current Average
T.J. Maxx visits

Site Selection: Tennessee - HomeGoods



- | | |
|---|---------|
| 1. Tullahoma (35.3827475,-86.2464592) | 305,637 |
| 2. Morristown (36.203256,-83.2983276) | 337,157 |
| 3. Bristol (36.5918617,-82.1934684)
- 1 Marshalls nearby | 314,965 |

328,208

Current Average
HomeGoods visits

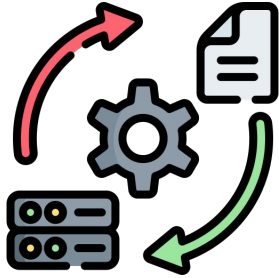
Our Sites Selection Is Expected to ...

generate

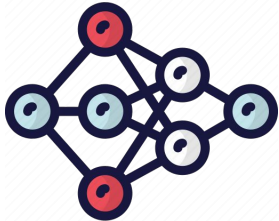
~24,000,000

MORE annual store visits

What MORE Can Be DONE



Extract better variables that can run in regression models for better interpretation



Design an algorithm to help optimize site selection rather than manually selection



Tune existing models and experiment with new machine learning models like LightGBM, CNN to potentially increase higher prediction accuracy

THANK YOU!

New Marshalls Tract Locations

Name	State	City	County_Code	GEOID	Lat	Long	Predicted_Visits
New Marshalls	TN	Memphis	157	47157022310	35.0452222	-90.074135	467473
New Marshalls	TN	Manchester	31	47031970502	35.4561774	-86.0698831	531009
New Marshalls	TN	Knoxville	93	47093000000	35.8806407	-83.8611812	527138
New Marshalls	FL	Fort Pierce	111	12111380902	27.4255514	-80.3654763	533489
New Marshalls	FL	Jacksonville	31	12031000000	30.2737719	-81.9400913	537915
New Marshalls	FL	Immokalee	21	12021011305	26.4371756	-81.4408339	510789
New Marshalls	FL	Belle Glade	99	12099008101	26.7302019	-80.6509221	538419
New Marshalls	FL	Clewiston	51	12051000102	26.7789195	-80.9336042	533954
New Marshalls	FL	Fernandina Beach	89	12089050204	30.6427763	-81.4691355	494715
New Marshalls	FL	Arcadia	27	12027010200	27.2340401	-81.861203	524197
New Marshalls	FL	Port LaBelle	51	12051000403	26.7527222	-81.4118758	507572
New Marshalls	FL	Sebastian	61	12061050808	27.825037	-80.4855166	531842
New Marshalls	FL	Lake City	23	12023110400	30.188127	-82.6226571	405083
New Marshalls	AL	Spanish Fort	3	1003010708	30.6835581	-87.8297072	544369
New Marshalls	AL	Auburn	81	1081040903	32.6130368	-85.4808959	459695
New Marshalls	AL	Dothan	69	1069040302	31.2755021	-85.4348261	431699

New T.J. Maxx Tract Locations

Name	State	City	County_Code	GEOID	Lat	Long	Predicted_Visits
New T.J. Maxx	TN	Morristown	63	47063100900	36.2019729	-83.356935	527214
New T.J. Maxx	TN	Lebanon	189	47189030500	36.2259424	-86.2993944	433867
New T.J. Maxx	TN	Knoxville	93	47093005501	35.9216627	-83.834614	518699
New T.J. Maxx	TN	Maryville	9	47009010500	35.7539049	-83.9833371	529280
New T.J. Maxx	TN	Gallatin	165	47165020904	36.3704939	-86.4302447	524278
New T.J. Maxx	FL	Navarre	113	12113010821	30.406463	-86.9104493	503112
New T.J. Maxx	FL	Wimauma	57	12057013916	27.7845132	-82.3053481	505945
New T.J. Maxx	FL	Brooksville	53	12053040909	28.5101164	-82.4899396	493723
New T.J. Maxx	FL	Kissimmee	105	12105014127	28.082933	-81.5027699	484050
New T.J. Maxx	FL	St Cloud	97	12097043207	28.1933498	-81.2971483	473791
New T.J. Maxx	FL	Tallahassee	73	12073000502	30.4384196	-84.2910109	586867
New T.J. Maxx	FL	Gainesville	1	12001000500	29.6568455	-82.3169268	516161
New T.J. Maxx	FL	West Melbourne	9	12009065022	28.0642764	-80.6459155	468691
New T.J. Maxx	FL	Miramar Beach	131	12131950609	30.370037	-86.329751	480251
New T.J. Maxx	FL	Jacksonville	31	12031011000	30.4096454	-81.6861334	455783
New T.J. Maxx	AL	Decatur	103	1103000300	34.5712507	-86.9706149	627805
New T.J. Maxx	AL	Birmingham	73	1073003500	33.5056238	-86.924273	531529
New T.J. Maxx	AL	Athens	83	1083020700	34.8205875	-86.9517709	472847
New T.J. Maxx	AL	Phenix City	81	1081042008	32.4858346	-85.0763311	525261
New T.J. Maxx	AL	Meridianville	89	1089010703	34.8774654	-86.5629052	476513

New HomeGoods Tract Locations

Name	State	City	County_Code	GEOID	Lat	Long	Predicted_Visits
New HomeGoods	TN	Morristown	63	47063100300	36.203256	-83.2983276	337157
New HomeGoods	TN	Bristol	163	47163042703	36.5918617	-82.1934684	314965
New HomeGoods	TN	Tullahoma	31	47031971001	35.3827475	-86.2464592	305637
New HomeGoods	FL	Gainesville	1	12001001519	29.612285	-82.3770724	388626
New HomeGoods	FL	Bradenton	81	12081001909	27.5370315	-82.4522426	363540
New HomeGoods	FL	The Villages	119	12119911204	28.891737	-82.0102088	358872
New HomeGoods	FL	Sanford	117	12117020401	28.796322	-81.2839486	352155
New HomeGoods	FL	Palm Coast	35	12035060213	29.4933524	-81.2074328	350989
New HomeGoods	FL	Tampa	57	12057006200	27.9168559	-82.5003139	336444
New HomeGoods	FL	Wellington	99	12099007750	26.6588261	-80.2376705	382735
New HomeGoods	FL	West End	86	12086015103	25.7494715	-80.4567622	358597
New HomeGoods	FL	Daytona Beach	127	12127082303	29.2030861	-81.0643388	332260
New HomeGoods	FL	Lakeside	19	12019030705	30.1329919	-81.7104913	364926
New HomeGoods	AL	Florence	77	1077010902	34.8344088	-87.6625081	338598
New HomeGoods	AL	Daphne	3	1003010704	30.6477601	-87.8838007	362571
New HomeGoods	AL	Montgomery	101	1101005403	32.3716245	-86.2001367	348124

Data Cleaning

1. **Collected and Combined** over **90+** related features
2. **Cleaned the Numerical Data** by replacing **missing values** and **outliers** with **zero**
3. **Handled the Categorical Data** with **one-hot encoding** technique

