## 

## Information Technology and data analytics

MASY1-GC 1240-104

Spring Semester 2024

**Individual Project**

The objective

You have just been hired as a consultant for a new and emerging home improvement store called “Tool Time.” The year is 2012 and Tool Time has 15 stores and 1500 employees. Tool Time’s flagship store is located in Essex county Massachusetts and they have a corporate headquarters in downtown Essex. They have also recently expanded their store footprint. Tool Time now has stores in Rhode Island and New Hampshire. Each of its 15 stores are doing very well in revenue and profit. They have hired you to consult them on potentially expanding and becoming a franchise in the larger New England Region.

Tool Time specializes in home improvement and DIY (Do-it-Yourself) supplies. Its store inventory and customers are similar to that of Home Depot and Lowes. The store footprint of Tool Time is not that of Home Depot and Lowes but they believe they have a competitive advantage in the small communities of the New England area.

You have been provided with a vast amount of data on Lowes and Home Depot including store locations by county, Census 2000 and 2010 data, highway counts per county and property tax data by county. A data scientist working for Tool Time has already provided data visualizations, built machine learning models on projected store sites and done an immense amount of research on its competitors (Home Depot and Lowes).

**Deliverables and Specifications for your report**

*- Less than 10 pages total or 2,750 words double spaced (including charts, graphs, data visualizations, etc)*

- Purpose of the Strategy and CRM solution (1 page)

* The Strategy (3 pages)
* The CRM Solution (3 pages)
* Choose an analytical solution for CRM to analyze, prescribe or predict for management (1/2 page - 1 page)
* New technology innovations, data sources or data pipelines to consider for the business and the CRM solution (1/2 page - 1 page)

1. Strategic

Your objective is to provide a strategic plan to identify how Home Depot and Lowes have selected their store locations in the New England Region. Once you have reverse engineered their strategy, you must develop one for Tool Time and its potential areas to expand its stores and become a franchise. You must select the next 3 store locations in the North East for Tool Time. You are to select the next 3 Tool Time stores where there is NO presence of Home Depot or Lowes in the county. Tool Time believes that they can differentiate themselves in small communities with superior customer service and they can do so without a presence from Home Depot or Lowes in the county.

1. CRM

To effectively compete against Home Depot and Lowes, Tool Time must intelligently manage its customer relationships to ensure its processes, people and strategy are aligned and fully consistent with its goals of expanding in the New England region. While expanding they must also be fully consistent with its goals of increasing revenues and net profits while enhancing the customer experience. Develop a plan for Tool Time’s CRM implementation, application to its customers, integration with its IT department and success criteria or KPI’s for its use in the corporate headquarters of Tool Time.

CRM Process considerations

1. Identify New Prospects
2. Track and Monitor Sales Opportunities
3. Organize and Engage Current Customers
4. Establish a Relationship with Current Customers
5. Manage Customer Life Cycle

* Components for consideration of your report:
* Phased implementation of the Strategy and the CRM solution
* Strategic Planning
* Performance Standards or KPIs of the CRM solution
* Custom Reporting and Analytic Solutions
* Data migration, data pipelines and data architecture
* Maintenance and Support
* Data Privacy and Responsible AI considerations
* Social Media Integration
* Change Management
* A Data warehouse and data analytics
* Security
* A customer loyalty program
* Cloud or on-prem considerations
* Risk Assessment

**Competitor Background**

***Home Improvement and DIY store history***

Home Depot and Lowes have similar store sizes, inventory and slight variations in their customer base. Typically, the stores are 120,000 square feet with a 20,000-square foot joining garden center. Inventory between the stores is quite similar with minor varieties here and there. Through your research and insight from your new colleagues at Tool Time, they have determined that customer demographics, property taxes and a store location with a close proximity to highways are the most important in making strategic decisions on store locations.

***Home Depot’s About Us*** (https://corporate.homedepot.com/about)

“Our founders’ vision of one-stop shopping for the do-it-yourselfer came to fruition when they opened the first two Home Depot stores on June 22, 1979, in Atlanta, Georgia. The first stores, at around 60,000 square feet each, were cavernous warehouses that dwarfed the competition and stocked 25,000 products, much more than the average hardware store at that time.

Today, The Home Depot is the world’s largest home improvement retailer with nearly 400,000 orange-blooded associates and more than 2,200 stores in the U.S., Canada and Mexico. The typical store today averages 105,000 square feet of indoor retail space, interconnected with an e-commerce business that offers more than one million products for the DIY customer, professional contractors, and the industry’s largest installation business for the Do-It-For-Me customer.”

***Lowe’s About Us*** (http://careers.lowes.com/media/39421/loweshistory.pdf)

Founded in 1946, Lowe’s has grown from a small hardware store to the 2nd largest home improvement retailer worldwide, the 8th largest retailer in the U.S. and 19th largest in the world.

Anticipating DIY needs while still accommodating contractors, Lowe’s began to enlarge its stores and expanded its merchandise offerings. The modern Lowe’s began in 1994, when its new store expansion consisted only of stores greater than 85,000 square feet of selling space. Expansion continues in North America as Lowe's plans to open 10 stores in 2012, with 103,000-square-foot stores in large markets and 94,000-square-foot stores in small to mid-size markets.

***Customer Demographics***

The customers for Home Depot and Lowes are split into the following categories:

1. *Contractors* involved in the construction and repairs of homes. This group comes to the store to buy home construction supplies, raw materials and tools
2. *DIY consumers* are homeowners or renters that live in a dwelling and come to the stores to buy materials and request consulting on their projects from the store staff
3. *Finished product* consumers –are those that come to the store to purchase goods that have been completed by the manufacturer but may need installation assistance like refrigerators, light fixtures, washer and dryers, etc.

**The Data**

***Files***

Home\_Depot\_Lowes\_Data.csv

RDC\_MarketHotness\_Monthly.csv

state\_region.csv

highways.csv

HDLData.csv

Property\_Tax.csv

*Data Considerations*

The New England Region encompasses the following states – Maine, New Hampshire, Connecticut, Vermont, Rhode Island and Massachusetts

Unfortunately, you do not have raw data on the above 3 customer demographic categories. However, the next best data points which relate to these customers are free and available to you – Census Data, Real Estate Data, Property Tax data and US & State Highway data.

The census data provides a broad but telling story of demographics for each county in the United States. Many businesses use this data for market research. The types of customers that Home Depot and Lowes are targeting can be gleaned from this census and store location data. A data dictionary is referenced in the appendix section below.

The property tax data is a dataset from smartasset.com and it lists property taxes by each county. The DIY stores have a large footprint with estimates of 103,000-square-feet in each store’s size. A larger property tax would imply a larger cost of operations per store given the county it is located in. Tool Times store size is similar in square footage to Home Depot and Lowes.

*Subjective Insight from Tool Time corporate*

The Vice President of Marketing at Tool Time has mentioned that she noticed Lowes and Home Depot stores are usually located near major highways. The data scientist has collected a count of US Highway, State Highways and Tool Roads per county in the North East Region. The VP is not certain if her theory is to related to the success with Home Depot and Lowes supply chain management or if vehicle traffic off of highway ramps is high and therefore leading to more customers stopping at their stores.

Tool Time executive management also asked the data scientist to research real estate and population data. The data scientist has performed some correlation research and theorized his findings. It is up to you, the consultant, to use or not use this research in your strategy and CRM planning.

Tool Time prides itself on its relationship with its small community locations. Their in-store employees are highly knowledgeable in DIY, tools and getting the job done. Tool Time has received many customer service awards. In addition to the success of their in-store employees, the IT and marketing teams collect survey data and conduct email marketing campaigns to continue to improve their customer loyalty. These considerations should be included in the CRM solution.

**Data Research Notes from the Tool Time’s Data Scientist:**

1. Data Transformation
   1. Performed an inner join on the HDLo and the Region data
   2. Isolated data frame to the New England Region
   3. Created 3 new columns
      1. Lowes present in county
      2. Home Depot present in county
      3. Home Depot and Lowes not present in county
   4. Performed an left join merging HDLo data set and the highway data
      1. Merged by both county name and state abbreviation
   5. Performed a left join HDLo and property tax data data setMerge by both county name and state abbreviation
2. Performed Exploratory Data Analysis on the stores
   1. Corrections to variables
      1. Watch out for missing values. Its ok to remove them or use median values.
      2. Use the Boolean values of present or missing Home Depot or Lowes. There should be a correlation against Lowes and one correlation against Home Depot for each of the independent variables
      3. There is collinearity among the census variables. Subtract the year \_2010 values from the year \_2000 to devise a unique variable free from collinearity. Take the absolute value of the subtracted values.
         1. Remove columns with \_2000 and \_2010 once after you have created the normalized field
3. Exploratory Data Analysis
   1. Total store counts of Home Depot and Lowes in the New England Region
   2. Created one dummy variable for Home Depot and one dummy variable for Lowes that identifies if the store is located in a county
4. *Data Visualizations* 
   1. Used a United States map with FIPS locations to plot the count of store locations of both Lowes and Home Depot by county

\*NOTE: The FIPS numbers are in the “county” column and should be used in conjunction with the map packages

R map packages: maps, maptools, blscrapeR, tmap, ggcorrplot

Python packages: importlib, plotly, plotly.figure\_factory, basemap, mpl\_toolkits.basemap

An example of populations by county from plot.ly can be seen here - <https://plot.ly/python/county-choropleth/#single-state>

Created the following visualizations to assist with data understanding:

North East Maps of the following:

1. North East Lowes store presence by count
2. North East Home Depot Presence by county
3. Census 2010 Population in North East
4. Highway county by County
5. Median Annual Property Tax Payment by county

Through a scatter plot visualization prove or disprove the following theories:

Note - Use the seaborn package with scatterplot

1. Determined if their is a relationship of income and population with store\_present by county

x="income", y="population",

hue="county", size="Store\_present"

2. Determined if their is a relationship of highway count and population with store\_present by county

x="highway\_count", y="population",

hue="county", size="Store\_present"

3. Determined if their is a relationship of Median Annual Property Tax Payment and population with store\_present by county

x="highway\_count", y="population",

hue="county", size="Store\_present"

1. Through a pandas\_profiling report and a spearmen correlation, I found the strength of the correlation for the variables to identify the following
   1. Found the variables that are most import to Home Depot
   2. Found the variables that are most import to both Home Depot and Lowes
   3. Found how are the home improvement store chains similar in their decision making
   4. Found how they are different
2. Found the top 3 counties that can be predicted as potential candidates for new locations for both Lowes and Home Depot?
   1. Used a logistic regression, RandomForestClassifier and a Decision Tree algorithms
   2. Leveraged a confusion matrix and precision, recall and F1 for evaluations

* Data Dictionary for Census Data
  + Areaname – city / town name
  + County – zipcode
  + State
  + Lcount – count of Lowes stores in the town/city
  + Hdcount – count of Home Depot in the town/city
  + Pop\_2000 – population in 2000
  + Pop\_2010 – population in 2010
  + Income\_2000 – avg income in 2000
  + Income\_2010 – avg income in 2010
  + pct\_U18\_2000 – percent under 18 in 2000
  + pct\_U18\_2010 – percent under 18 in 2010
  + Pct\_college\_2000 – percent in college per town in 2000
  + Pct\_college\_2010 – percent in college per town in 2010
  + Ownhome\_2000 – percent owned home in 2000
  + Ownhome\_2010 – percent owned home in 2010
  + Density\_2000 – percent density per town in 2000
  + Density\_2010 – percent density per town in 2010
  + Pct\_white\_2000 – percent of Caucasian in town in 2000
  + Pct\_white\_2010 – percent of Caucasian in town in 2010
  + Pct\_black\_2000 – percent African American in 2000
  + Pct\_black\_2010 – percent African American in 2010

Data Dictionary for realtor.com data

|  |  |
| --- | --- |
| **Realtor** |  |
| Median Listing Price | The median listing price within the specified geography during the specified month. |
| Avg Listing Price | The average listing price within the specified geography during the specified month. |
| Median Listing Square Feet | The median home size in square feet within the specified geography during the specified month. |
| Median List Price Per Sqft | The median listing price per square foot within the specified geography during the specified month. |
| Active Listing Count | The count of active listings within the specified geography during the specified month. The active listing count tracks the number of for sale properties on the market, excluding pending listings where a pending status is available. This is a snapsot measure of how many active listings can be expected on any given day of the specified month. |
| Median Days on Market | The median number of days property listings spend on the market within the specified geography during the specified month. Time spent on the market is defined as the time between the initial listing of a property and either its closing date or the date it is taken off the market. |
| Days on Market Under 30 Share | The share of properties spending fewer than 30 days on the market. |
| New Listing Count | The count of new listings added to the market within the specified geography during the month. |
| Price Increase Count | The count of listings which have had their price increased within the specified geography during the month. |
| Price Decrease Count | The count of listings which have had their price reduced within the specified geography during the month. |
| Pending Listing Count | The count of pending listings within the specified geography during the specified month, if a pending definition is available for that geography. This is a snapsot measure of how many pending listings can be expected on any given day of the specified month. |
| Total Listing Count | The total of both active listings and pending listings within the specified geography during the specified month. This is a snapsot measure of how many total listings can be expected on any given day of the specified month. |
| Pending Ratio | The ratio of the pending listing count to the active listing count within the specified geography during the specified month. |
| Views | The count of views to all Realtor.com listings in the geographic area. |
| Views per Property | The total count of views to active Realtor.com listings divided by the active listing count for the geographic area. |
| Market Hotness |  |
| Nielsen HH Rank | The specified zip code, county, or metro area’s rank by household count compared to other zip codes, counties and metro areas. A rank value of 1 is the highest by household count. |
| Hotness Rank Within County | In the case of a zip code, this metric represents the zip code’s hotness rank, by hotness score, compared to all other zip codes within its county. A rank value of 1 is considered the hottest (highest hotness score). |
| Hotness Rank Within CBSA | In the case of a zip code or county, this metric represents the zip code or county’s hotness rank, by hotness score, compared to all other zip codes or counties within its metro area. A rank value of 1 is considered the hottest (highest hotness score). |
| Hotness Rank | The specified zip code, county, or metro area’s hotness rank, by hotness score, compared to all other zip codes, counties and metro areas. A rank value of 1 is considered the hottest (highest hotness score). |
| Hotness Rank (Prev) | The specified zip code, county, or metro area’s hotness rank in the previous month. |
| Hotness Rank (Change) | The change in hotness rank from the previous month. A positive value indicates that the geography’s hotness has decreased (moved down in ranking), and a negative value indicates that its hotness has increased (moved up in ranking). |
| Hotness Score | The hotness score is an equally-weighted composite metric of a geography’s supply score and demand score. |
| Supply Score | The supply score is an index representing a zip code, county or metro’s median days on market ranking compared to other zip codes, counties or metros. |
| Demand Score | The supply score is an index representing a zip code, county or metro’s listing page views per property ranking compared to other zip codes, counties or metros. |
| Median DOM | The median number of days property listings spend on the market within the specified geography during the specified month. Time spent on the market is defined as the time between the initial listing of a property and either its closing date or the date it is taken off the market. |
| Median DOM (vs US) | The median days on market in the specified geography divided by the median days on market for the US overall during the same month. |
| Views Per Property  (vs US) | The count of views a typical property receives in the specified geography divided by the count of views a typical property receives in the US overall during the same month. |
| Median Listing Price | The median listing price within the specified geography during the specified month. |
| Median Listing Price  (vs US) | The median listing price within the specified geography divided by the median listing price for the US overall during the same month. |