**FINAL PROJECT**

College of Professional Studies, Northeastern University

ALY 6980 : DATA ANALYTICS CAPSTONE

Dr. Eric Liu

10th December 2020

Piyush Nath Gupta

|  |
| --- |
| **INTRODUCTION** |

This project seeks to analyze the home furnishings and décor industry with At home Group as a case study, the business, strategy, challenges, key players, trends, and opportunities. The home furnishing industry is segmented by Product (Home Furniture, Home Textile, Flooring, Wall Décor, Tableware and Cookware, Lighting and Lamps, Sanitaryware, Home Accessories, Other Products) and by Distribution Channel (Supermarkets and Hypermarkets, Home Decor Stores, Gift Shops, Direct to Consumer, Online, Others) with home furniture having the biggest market share (Anon, 2017).

With a growing number of individuals attaining homes, particularly in developed countries such as the United States, the home decor market is growing tremendously (Anon, 2020 ). The global Home Furnishings market is valued at 808700 million USD in 2020 is expected to reach 1149970 million USD by the end of 2026, growing at a CAGR of 5.1% during 2021-2026 (Anon, 2020). One factor responsible for the rapid increase in the home décor market is the expansion of the real estate industry (Anon, 2017). A large number of people are accessing online retail outlets to decorate their homes. Other factors that affect demand growth are the increase in household incomes and the enhancement of people's lifestyles (Anon, 2019). Some key players in the industry are IKEA, Walmart, Bed Bath and Beyond, Macy’s, Wayfair, At Home group, Future Group, Haworth, Ashley Furniture, Carrefour, J.C. Penny, Create and Barrel, Fred Meyer, Herman Miller, Home Depot, Badcock and American Furniture Warehouse etc. The home furnishings market size segment by applications can be divided into:

* E-Commerce Sales
* In-store Sales

**ATHOME GROUP**

At Home is the leading home decor and furniture store in the United States. The business (formerly known as Garden Ridge) was founded in 1979 in Garden Ridge, Texas, suburb of San Antonio, and currently operates 219 stores in 37 states throughout the United States (At Home Group, 2019). The business focuses on offering the widest selection of items for every space, style and budget, with more than 70% of its items unbranded, privately labeled or expressly built for the Athome Group. Some of the goods are imported from foreign manufacturers in countries such as China, Vietnam , India, Turkey and Hong Kong (At Home Group, 2019).

The business presently has two distribution centres, one in Plano, Texas, with a capacity to serve 220 stores, and the other in Carlisle, Pennsylvania, which became completely functional in the fiscal year ended January 25, 2020 ('Fiscal Year 2020') with a capacity to serve 220 stores, along with more than 350 stores (At Home Group, 2019).

The company’s goals are:

* Provide the biggest collection of Home Décor at a great value
* Create an enjoyable self-help store experience for customers
* Expand the company’s brand to generate demand
* Run a profitable business
* Be a great place to work and grow

Its growth strategies include:

* Increasing its store base
* Drive competitive store sales
* Build the at home brand and create awareness
* Utilize digital solutions to improve shopping experience

While the company claims that they do not have a major competitor, some items sold by the company compete with the following categories shown in the table below:

|  |  |  |
| --- | --- | --- |
| S/N | PRODUCTS | COMPETITORS |
| 1 | Specialty Home Décor / Organization and Furniture retailers | Bed Bath & Beyond, The Container Store, Ethan Allen, Havertys, Home Goods, Pier 1 Imports and Williams-Sonoma |
| 2 | Mass / Club retailers | Costco, Target and Wal-Mart Stores |
| 3 | Arts / Craft / Hobby retailers | Hobby Lobby, Jo-Ann Stores and Michaels Stores |
| 4 | Discount Home Décor retailers | Big Lots, Burlington and Tuesday Morning |
| 5 | Home Improvement and repair retailers | Home Depot and Lowe’s |
| 6 | Online home décor retailers | Wayfair, Amazon |

***Table 1*** *At home group products and competitors*

Over the past six fiscal years , the company has also invested in IT processes and technology, including Merchandising, Planning and Allocation (JDA), POS (IBM / Toshiba), Logistics Center (PKMS) and Financing and Accounting (SAP), to ensure its processes are reliable and efficient (At Home Group, 2020). In addition, the organization has also invested in data centers and increased IT team members to provide the necessary resources and project management capability required for expansion (At Home Group, 2019).

**EFFECT OF PANDEMIC ON THE INDUSTRY**

The COVID-19 pandemic poses itself as a major challenge which has a huge effect on individuals, companies and the global economy. As a result, companies and organisations in all industries and sectors have faced economic and organizational complexities, ranging from payroll to auditing practices, as well as the way organizations perform day-to-day activities and have had to adapt to meet this situation (Anon, 2020).

In the home furnishing industry, a major trend was a disruption in global supply chains. Several well-known companies have furniture manufacturing factories in China, and most good quality leathers come from Italy, adding extra pressures on the supply chain (Vidojevic, 2020). Many manufacturers were also required to shut down the first two months of the pandemic, which in turn prolonged average waiting times. Some retailers failed to keep up as job losses rose, many consumers canceled, limited or diverted their expenditure, and the retail industry experienced two consecutive months of low sales and consumer demand declines (Anon, 2020).

Another major trend during the pandemic was the rise in consumer spending on furniture and home furnishings during the three months of the lockdown, especially in May (Anon, 2020). This was very interesting, as several retail outlets were shut down in those months as a result of the lockdown. This was because one of the first things people did when they were unexpectedly expected to work from home in the early days of the lockdown was to buy a new desk, an office chair, or improved lighting and with the inflow of support checks from the state, some new and improved home furniture.

E-commerce is a blessing, particularly during the outbreak, when most home decor and furniture stores closed and the need for furniture soared, shoppers had to turn to online shopping. Companies that invested in improving their online presence coped more easily. Direct-to-consumer (DTC) brands were among the first to close their shops because they already had a robust online presence (Anon, 2020). For instance, Reuters reported that Ikea would sell products on the Chinese e-commerce platform Tmall, the first time the company has sold through a third party in its 77-year history (Anon, 2020).

Furniture companies with retail outlets developed additional sanitation protocols and took extra cleaning measures throughout the day to ensure safety. Stores also revised their operating hours closing either at an earlier time or opening at later times. Additionally, employees were coached on extra preventative measures and had to submit daily checklists to executive management ensuring they comply with our newly developed sanitation standards (Anon, 2020).

**ATHOME GROUP DURING THE PANDEMIC**

Unlike other home furnishings outlets, At Home furniture sales grew through the pandemic, with buyers flocking to their websites and shops for new outdoor furniture and office accessories. The company’s sales surpassed while retailers in its category like pier1 imports went out of business entirely (Thomas, 2020). Its quarterly same-store revenues are forecast to rise by 42% and its second-quarter net income is estimated to be at least $82 million, with revenue of about $515 million. At Home stock have rose by more than 123% this year, with a market value of $778.6 million (Thomas, 2020). Some key questions to be answered with regards to this are:

* How was the company able to boost sales during the pandemic?
* What strategies did the company implement to boost sales?

**BUSINESS CHALLENGES**

While increasing customer demand, strong growth in housing start-ups and disposable income levels are projected to fuel sales of home furnishing items, the rise in the percentage of online retailers posed a challenge to substantial sales growth for retail stores during the lock-down period (Anon, 2020). Another challenge the company faced was high external competition from discount department stores and online retailers (Anon, 2020). In addition, the production and shipment of overseas furniture was interrupted due to a pandemic, leaving many products back ordered or out of stock (Anon, 2020).

**BUSINESS STRATEGY**

As per capita disposable income and consumer spending are predicted to decline by 2020, demand for manufactured products is likely to decline, and consumers are likely to move away from expensive items, negatively affecting revenues (Anon, 2020). This caused the company to take cost-cutting measures. The company uses a low-price policy that offers the customers the best possible value without the need for frequent discounts or offers (At Home Group, 2019). Currently, 80% of its net sales are made at full retail, with the rest attributed to planned price increases used to clear a slow-moving product or a post-seasonal product (At Home Group, 2019). In addition, the company uses an overall price point of less than $15 to make it a pioneer in the home decor industry. It uses in-store merchandising, visual management, sourcing and pricing methods and makes a self-service model to eliminate in-store staffing requirements, while adhering to social distance regulations providing customers with superior value, maintaining strong which consistent quality across product ranges (At Home Group, 2019).

During the pandemic, net profits crossed the $1 billion plateau for the first time in the company's history, rising 23 per cent to $1,166 billion, marking five consecutive years and 19 straight quarters at least 20 per cent of revenue growth. Comparable store sales improved 2.7% and the excellent results encouraged the company to reinvest in its operation, make progress against our strategic goals and improve revenue growth (Thomas, 2020).

|  |
| --- |
| **DATA & METHODS** |

For this project, the data collected is based on demographic and store revenue information for At home stores nationwide. All data collected is based on demographic information using store location and zip code information. The dataset contains 220 rows and 84 columns. The information in this dataset are grouped based on Store location Information, Rendered servicesPopulation, Population Type and Population growth, Income, Housing type, Competition and Store Revenue. The data on “store information” was gotten from the At home website <https://www.athome.com/store-find/> and using the zip code information on all stores the following sources were used to derive more additional demographic information on the data set based on zip code. The data dictionary is shown in the table below:

|  |  |  |
| --- | --- | --- |
| S/N | VARIABLES | DESCRIPTION |
| 1 | Store Name | Indicates the name of the store |
| 2 | Address | Indicates the address of the store |
| 3 | States | Indicates the state where the store is located |
| 4 | Region | Indicates the region where the store is located |
| 5 | Country | Indicates the country where the store is located |
| 6 | City | Indicates the city where the store is located |
| 7 | Zip Code | Indicates the zip code where the store is located |
| 8 | Lat | Indicates the latitude of the store location |
| 9 | Lng | Indicates the latitude of the store location |
| 10 | Service Type | Indicates the delivery service provided by each store |
| 11 | Instore | Indicates the “instore” delivery service noted by (1 if available or 0 if unavailable) |
| 12 | Pickup Curbside | Indicates the “pickup curbside” delivery service noted by (1 if available or 0 if unavailable) |
| 13 | Local Delivery | Indicates the “local delivery” delivery service noted by (1 if available or 0 if unavailable) |
| 14 | Total Population 2017 | Indicates the total population of the store location in 2017 |
| 15 | Total Population 2018 | Indicates the population of the store location in 2018 |
| 16 | Total Population 2019 | Indicates the population of the store location in 2019 |
| 17 | Zipcode Population 2013 | Indicates the zip code population of the store location in 2013 |
| 18 | Zipcode Population 2014 | Indicates the zip code population of the store location in 2014 |
| 19 | Zipcode Population 2015 | Indicates the zip code population of the store location in 2015 |
| 20 | Current Zipcode Population | Indicates the current population of the store location |
| 21 | Population density | Indicates the population density of the store location |
| 22 | Pop growth 2010 to 2020 | Indicates the population growth in percentage of the store location from 2010 to 2020 |
| 23 | Predicted growth from 2020 to 2025 | Indicates the predicted population growth in percentage of the store location from 2020 to 2025 |
| 24 | Total pop growth 2017 | Indicates the total population growth rate in 2017 |
| 25 | Total pop growth 2018 | Indicates the total population growth rate in 2018 |
| 26 | Total pop growth 2019 | Indicates the total population growth rate in 2019 |
| 27 | House Affordability Index | Indicates housing affordability |
| 28 | Male | Indicates the current male population in the store location |
| 29 | Female | Indicates the current female population in the store location |
| 30 | Median Age | Indicates the median age of the store location |
| 31 | Male median age | Indicates the male median age of the store location |
| 32 | Female median age | Indicates the female median age of the store location |
| 33 | Median household income | Indicates the median household income of the store location |
| 34 | Average home value | Indicates the average home value of the store location |
| 35 | Median home value | Indicates the median home value of the store location |
| 36 | White | Indicates the population of white people in the store location |
| 37 | Blacks African American | Indicates the population of blacks and African American people in the store location |
| 38 | American Indian | Indicates the population of American Indian people in the store location |
| 39 | Asian | Indicates the population of Asian people in the store location |
| 40 | Native Hawaiian | Indicates the population of native Hawaiian people in the store location |
| 41 | Other race | Indicates the population of other race in the store location |
| 42 | Two or more | Indicates the population of mixed race in the store location |
| 43 | Competition | Indicates the number of competition around the store location |
| 44 | HOH Owner 15 - 24 | Indicates the population of homeowners between age 15 to 24 in the store location |
| 45 | HOH Owner 25 - 34 | Indicates the population of homeowners between age 25 to 34 in the store location |
| 46 | HOH Owner 35 - 44 | Indicates the population of homeowners between age 35 to 44 in the store location |
| 47 | HOH Owner 45 - 54 | Indicates the population of homeowners between age 45 to 54 in the store location |
| 48 | HOH Owner 55 - 64 | Indicates the population of homeowners between age 55 to 64 in the store location |
| 49 | HOH Owner 65 - 74 | Indicates the population of homeowners between age 65 to 74 in the store location |
| 50 | HOH Owner 75 - 84 | Indicates the population of homeowners between age 75 to 84 in the store location |
| 51 | HOH Owner 85 Plus | Indicates the population of homeowners between aged 85 and above in the store location |
| 52 | HOH Owner Total | Indicates the total number of home owners in the store location |
| 53 | HOH Renter 15 - 24 | Indicates the population of home renters between age 15 to 24 in the store location |
| 54 | HOH Renter 25 - 34 | Indicates the population of home renters between age 25 to 34 in the store location |
| 55 | HOH Renter 35 - 44 | Indicates the population of home renters between age 35 to 44 in the store location |
| 56 | HOH Renter 45 - 54 | Indicates the population of home renters between age 45 to 54 in the store location |
| 57 | HOH Renter 55 - 64 | Indicates the population of home renters between age 55 to 64 in the store location |
| 58 | HOH Renter 65 - 74 | Indicates the population of home renters between age 65 to 74 in the store location |
| 59 | HOH Renter 75 - 84 | Indicates the population of home renters between age 75 to 84 in the store location |
| 60 | HOH Renter 85 Plus | Indicates the population of home renters between aged 85 and above in the store location |
| 61 | HOH Rental Total | Indicates the total number of home renters in the store location |
| 62 | Family Households | Indicates the number of family households in the store location |
| 63 | Singles Guardian | Indicates the number of singles guardian in the store location |
| 64 | Singles | Indicates the number of singles in the store location |
| 65 | Singles with Roommates | Indicates the number of singles with roommates in the store location |
| 66 | Houses built 2010 or later | Indicates the number of modern houses in the store location |
| 67 | Studio Apartment | Indicates the number of studio apartments in the store location |
| 68 | 1 Bedroom | Indicates the number of 1-bedroom houses in the store location |
| 69 | 2 Bedroom | Indicates the number of 2-bedroom houses in the store location |
| 70 | 3+ Bedroom | Indicates the number of 3 + bedroom houses in the store location |
| 71 | Housing Units | Indicates the number of housing units in the store location |
| 72 | Occupied Housing Units | Indicates the number of occupied housing units in the store location |
| 73 | Wealth Index |  |
| 74 | H income <$25k | Indicates the number of household income less than $25000 in the store location |
| 75 | H income $25k | Indicates the number of household income equal to $25000 in the store location |
| 76 | H income $45k | Indicates the number of household income equal to $45000 in the store location |
| 77 | H income $60k | Indicates the number of household income equal to $60000 in the store location |
| 78 | H income $100k | Indicates the number of household income equal to $100000 in the store location |
| 79 | H income $150k | Indicates the number of household income equal to $150000 in the store location |
| 80 | H income >$200k | Indicates the number of household income greater than $200000 in the store location |
| 81 | Sales 2017 | Indicates the total sales in 2017 for each store location |
| 82 | Sales 2018 | Indicates the total sales in 2018 for each store location |
| 83 | Sales 2019 | Indicates the total sales in 2019 for each store location |
| 84 | Sales 2020 | Indicates the total sales in 2020 for each store location |

***Table 2*** *Showing the data dictionary for the collected dataset*

|  |
| --- |
| **RESEARCH QUESTIONS** |

This project seeks to analyze At home Group using demographic information and store revenue information to determine key players and opportunities for the company and to provide a meaningful insight to be considered for choosing future store locations. To begin this analysis and utilizing the data provided, some essential business questions with relation to the business can be examined, these questions are:

* What areas show the most amount of percentage change in revenue growth in the last three years?
* What stores will generate the most revenue in sales in the future?
* What areas generate the most revenue?
* What are the key players in profitability?
* Is there a relationship between profitability and type of store service rendered?
* How does population affect the sales?

|  |
| --- |
| **ANALYTICAL PROCEDURES** |

***What areas show the most amount of percentage change in revenue growth in the last three years?***

In this question, we want to find out whether the sales revenue of At Home stores in the past three years has special characteristics geographically. We intend to conduct a region analysis using the data. First of all, we created a new measure variable, which is “continuous growth rate”. To calculate this variable, we used the formula **[(Sales 2019 – Sales 2017)/ Sales 2017**]. Since the Sales for 2020 does not include the data of the whole year, we did not include it in the analysis. Thus, we will analyze the geographical characteristics of store sales from 2017 to 2019.

Chart

Description automatically generated

**Figure 1** Percentage continuous revenue growth in the past three years for the different states

The chart above shows the growth rate of each state's combined sales over three years. The top five states for sales growth are Michigan (MI), Texas (TX), Ohio (OH), Arizona (AZ), and Arkansas (AR). We can see sales growth of over 250% in each of the first five states for three years, and even 480% in Michigan. On the other hand, states like Minnesota (MN), Nebraska (NE), Wisconsin (WI), Iowa (IA), Virginia (VA) show a drop in sales with VA having the highest of over -50%.

Graphical user interface, application

Description automatically generated

**Figure 2** Median revenue growth rate for over three years for each state

If we look at the median sales growth rate, the results are quite different. In the graph of States’ medium sales revenue growth, only Arkansas is still among the top five states, and New Jersey goes from seventh to the first. From the comparison of the two charts, we can conclude that the sum sales revenue growth rate in some states, although seemingly high, are contributed by some extremely high sales of a few successful stores. In particular, Michigan and Texas, mid-level stores of these two states saw sales growth of little more than zero in the past three years.

Chart, bar chart

Description automatically generated

**Figure 3** Sum revenue growth rate for over three years for each region

Next, we decided to expand the state to region to analyze. The United States is divided into 6 regions. Stores in the Midwest had the highest growth rates in the past three years. However, the high growth rate came from stores in almost only three states, Michigan (MI), Missouri (MO) and Ohio (OH). Stores in the southwest region had the second highest three-year sales growth rate among the six regions. Like the Midwest region, sales growth in the south-west was almost entirely contributed by two states, which are Texas (TX) and Arizona (AZ).

Map

Description automatically generated

**Figure 4** Sales Revenue Growth Distribution

If we put metrics in the map, it will be more perspicuous. The map shows that 2019 sales were concentrated in the south and east of the United States. Texas topped the list, well ahead of other states. In terms of sales growth, several stores with high growth rates are marked yellow. In conclusion, there are eight stores that had experienced significant growth in sales over the past three years. These stores are shown in table 2 below.

|  |  |  |  |
| --- | --- | --- | --- |
| STATE | STORE NAME | REGION | ZIP CODE |
| MI | Michigan Troy | Midwest | 48083 |
| OH | Ohio Columbus | Midwest | 43240 |
| NJ | New Jersey Cherry Hill | Mid-Atlantic | 08002 |
| MO | Missouri Springfield | Midwest | 65807 |
| AR | Arkansas Rogers | South | 72758 |
| TX | Texas Richmond | Southwest | 77406 |
| TX | Texas Pearland | Southwest | 77584 |
| AZ | Arizona Gilbert | Southwest | 85295 |

***Table 3*** *Stores with significant sales growth in the last three years*

***Which stores will generate the most revenue on sales in the future?***

To answer this question, we utilized R studio. Some variable names were not standardized, so we used some functions perform data cleaning. We found one N/A record because the area covered by the zip code is too small that no relevant population data could be found. Next, we join the demographic data with sales revenue data in advance. For this question, however, we believe that it would be best to combine sales revenue data as one component to make a decision tree model. Hence, we drop some four-year sales data.

Text

Description automatically generated

Figure Code used to combine sales revenue as one component for decision tree model

Before building the decision tree model, we join the sales revenue data with demographic data based on zip codes. We choose sales revenue as the target variable, and we split 80% data as train data.

Text

Description automatically generated

**Figure 6** Code for building decision tree model

Text, letter

Description automatically generated

Text, letter

Description automatically generated

**Figure 7** Estimate value of each node

After building the decision tree model, we find that in the terminal nodes, the estimated value of Pop\_Growth\_per\_year\_from\_2010\_to\_2020>=7.01 is the largest. According to this condition, we obtained these stores after sifting through all the data:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| STORE NAME | ADDRESS | STATES | REGION | COUNTY | ZIPCODE |
| Texas Alliance | Michigan Troy | TX | Southwest | Fort Worth | 76177 |
| Texas Conroe | Ohio Columbus | TX | Southwest | Conroe | 77384 |

***Table 4*** *Stores predicted to have the best performance in the future*

The stores showed in table 3 above are predicted to have the best performance in the future.

Following that, we determined the variables that were of most importance to this analysis.

Chart, bar chart

Description automatically generated

**Figure 8** Chart showing feature importance

By rearranging the predictive variables, the difference between the response variable and predictive variables was calculated. The greater the difference, the more important the variable. As the graph shows, the top 4 important variables that affect the response variable are **Total\_Population\_2019**, **Wealth\_index,** **Predicted\_Growth\_per\_year\_from\_2020\_to\_2025,** and **Total\_Population\_2017**.

Graphical user interface

Description automatically generated

**Figure 9** Partial dependence plot for four important variables

We select those four important variables to draw a partial dependence plot. The partial dependence plot shows the marginal effect of selected variables to the response variable. As the plot shows, Total\_Population\_2019, which is the most important variable, a little change in this variable will result in nearly $1,750,000 change in sales revenue. Change from 2.5% to 2.7% in Predicted growth rate per year from 2020 to 2025 will result in nearly $1,600,000 change in sales revenue.

***What areas generate the most revenue and what are the key players in profitability?***

From our given dataset, we considered sales for 2020 as our target variable. To determine the distribution of this variable, we plotted a distribution plot shown below.

Chart

Description automatically generated

**Figure 10** Distribution plot to show variation in value of sales 2020

The plot above shows that the variables are normally distributed, hence, no firther scaling was required.

Next, we divided our dataset into X and Y where X consists of all the predictor variables, and Y contains our target variable i.e. Sales\_2020:

Text, letter

Description automatically generated

**Figure 11** Variables for testing and training data

Here, we divided our training and testing data into ratio 80:20 and imported linear regression model to fit our training model:

Graphical user interface, text, application, email

Description automatically generated

**Figure 12** Code for linear regression

Results from regression is shown in the figure 13 below with the following values for all predictor variables.

Table

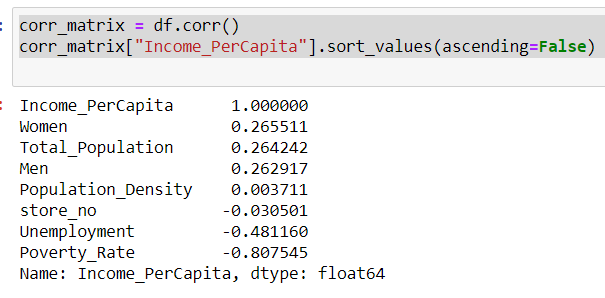
Description automatically generated

Table

Description automatically generated

**Figure 13** Results for linear regression model

From the above values, we were able to figure out the important demographic attributes which impact our Sales the most. For instance, a lower t value of Region and Population Density affirms their strong correlation with the target sales value. To affirm the connection of 'Income\_PerCapita' with different properties, following relationship framework was shaped proposing 'Total\_Population' to be the most noteworthy contributing component.



**Figure 14** relationship framework for variables

Additionally, we plotted the population density using a join plot against our sales\_2020 data as follows:

Chart, scatter chart

Description automatically generated

**Figure 15** Join plot for comparing population density and sales 2020

The above plot shows negative correlation among populaiton density and Sales\_2020 as lower the value of population density leads higher sales, then areas with higher population density. Similary, creating a jointplot using region distribution as follows:

A picture containing graphical user interface

Description automatically generated

**Figure 16** Joint plot for comparing region wise distribution for sales 2020

The above chart shows that the most number of sales are obtained from South and Southwest regions.

Chart, bar chart

Description automatically generated

**Figure 17** County wise distribution of total population by states

***Is there a relationship between profitability and type of store service rendered?***

To find the relationship between profitability and type of store services, we check the density of the three type of store service firstly. The “1” means the store has this service, the “0” means the store do not has this service. Obviously, there are 214 stores has in-store service, there are 5 stores do not have in-store service. Meanwhile, there are 218 stores has the pickup-curbside service, and only one store does not have the pickup-curbside service. However, only 155 stores have the local-delivery service, 64 stores does not has local-delivery service.

Graphical user interface, text, application, email

Description automatically generated

**Figure 18** Screen shot for density of in-store service, pickup-curbside service, and local-delivery service.

Next, we create the distribution plot to check the density of sales in each year. Also, we calculate the average sales for each year and create a table to shows the result. It is clear that the average sales decrease year by year ranging from 9264444.32 to 8062629.54.

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

Figure 19. Average Sales Distribution in 2017, 2018, 2019, and 2020

|  |  |
| --- | --- |
| years | Average sales |
| 2017 | 9264444.32 |
| 2018 | 9,331,472.44 |
| 2019 | 9157967.56 |
| 2020 | 8062629.54 |

Table 5. *the average sales of each year.*

On the other hand, we check the sales for the local delivery services under different population density, average household income, and vacant housing unite. We can see that the population density and vacant housing unite has similar distribution in figures. The average household income may have great effect to the sales in local delivery service.

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

***Figure 20.*** *Sales For The Local Delivery Services Under Different Population Density, Average Household Income, And Vacant Housing Unite*

For modelling, we create a correlation heatmap to check the relationship between 2020 years sales and store services. the correlation heatmap shows three types of services all has positive correlation to the sales, especially, local delivery service has highest correlation to the sales in 2020. Thus, this analysis result could prove the delivery service is really necessary for residents during the covid-19 period of time.

A picture containing bar chart

Description automatically generated

***Figure 21.*** *Decision Tree Model*

This decision tree has 4 layer, and five leaves. The first layer is about the local delivery service, which divides the data into two parts according to whether the store has local delivery. When local delivery service is not provided, sales are 6,754,800, while when local delivery service is provided, its sales Are 8,370,738. So, the local delivery service will bring nearly 200,000 sales to the store.

The second layer is the in-store service. For these stores do not have local delivery service and in-store service, sales are 5,707,126, while when in-store service is provided, its sales is 6,777,576.65. Besides, for these stores has local delivery service, it divides the data into two parts according to whether the store has in-store service. When in-store is not provided, we can see that its sales are 6,786,403.75, and when in-store is provided, sales are 8,422,683.623. Therefore, in-store service is still significant when a store has local delivery services and in-store service, sales is the highest, with 8,422,683.623.

The last layer is the pickup curbside service. According to whether there is a pickup curbside service, the data is divided into two parts: when there is a pickup curbside service, the sales is 6,800,702.0; when there is no pickup curbside service, the sales is 5,732,929.0. Therefore, the pickup curbside service is essential because of the comparison for stores that do not provide in-store service. The sales of these stores that provide instore service are about 1,000,000.0 higher.

Therefore, even in a particular period, having three services is still very attractive. However, it is worth noting that stores with local delivery service will have more sales. Therefore, we can say that all services are necessary for sales. At the same time, for future business, we suggest that delivery service and pickup curbside services should be added to all of the store, which will help increase sales in particular periods.

Diagram

Description automatically generated

***Figure 22.*** *Decision Tree Model*

***How does population affect the sales per location?***

To determine the question above, we obtained the top stores which performed well in terms of sales. To do this, we chose top ten stores with maximum sales when compared to other stores from 2017 to 2020.

Figures this, that , that and that below shows the top ten stores with high sales and its population growth for 2017, 2018, 2019 and 2020 respectively.

Chart, bar chart

Description automatically generated

**Figure 23** Population growth and sales for top 10 stores for 2017

Chart, bar chart

Description automatically generated

**Figure 24** Population growth and sales for top 10 stores for 2018

Chart, bar chart

Description automatically generated

**Figure 25** Population growth and sales for top 10 stores for 2019

Chart, bar chart

Description automatically generated

**Figure 26** Population growth and sales for top 10 stores for 2020

Based on the sales data we can see that there are six stores which are performing well and in the top ten for four years. These stores are:

* Texas Conroe
* Texas Webster
* North Carolina Charlotte
* Oklahoma Oklahoma City
* Oklahoma Tulsa
* Virginia Chesapeake

In order to observe whether population affects the sales of these stores, we decided to compare the sales with the population growth for the above six stores.

**TEXAS CONROE**

First, we observed the “Texas Conroe” store sales data with population growth.Here, we can see as the population growth increases the sales also increased. However, in 2020 the sales decreased which may be due to Covid effect.

Graphical user interface, application, Word

Description automatically generated

**Figure 27** Population growth and sales for Texas conroe store

**TEXAS WEBSTER**

Graphical user interface

Description automatically generated

**Figure 28** Population growth and store sales for Texas Webster store

Here, we can clearly see that for the year 2018 the population growth decreased significantly compared to 2017, however, sales for that year increased significantly, being the year with the highest sales record for this store. In this scenario and for the four years, population does not affect the sales.

**NORTH CAROLINA CHARLOTTE**

Graphical user interface

Description automatically generated

**Figure 29** Population growth and sales for North Carolina Charlotte store

For Charlotte stores, the sales went up when the population growth decreased in 2018. However, in 2019 there was an increase in population growth but sales decreased.

**OKLAHOMA OKLAHOMA CITY**

Graphical user interface, diagram

Description automatically generated

**Figure 30** Population growth and store sales for Oklahoma Oklahoma city store

For this store, the sales went up while population increased for 2018 and the sales went up even though population growth came down in 2019 so we can see that the population does not affect the sales.

**OKLAHOMA TULSA**

Graphical user interface, line chart

Description automatically generated

**Figure 31** Population growth and sales for Oklahoma Tulsa store

For this store, as population increased in 2018, sales slightly decreased. However, in 2019, there was a significant increase in population, but the sales slightly reduced.

**VIRGINIA CHESAPEAKE**

Graphical user interface

Description automatically generated

**Figure 32** Population growth and sales for Virginia Chesapeake store

For this store, the sales increased as population also reduced in 2018. On the other hand, in 2019 as population increased, sales decreased slightly.

**INTEGRATING ALL STORES DATA IN TO SINGLE GRAPH**

Graphical user interface, diagram

Description automatically generated

**Figure 33** Integrated population and sales growth for all top 10 stores

**SUPERVISED LEARNING MODEL – LINEAR REGRESSION**

This purpose of this model is to enable the company to use demographic information of a new location to provide the predictive sales as output which can help in deciding the profitability of proposed store locations.

Here, we would apply a predictive model to provided some important demographic information about a hypothetical location and will obtain output of predicted sales, , which will be inspired from the history and other attributes from our data. This would be helpful in important decision making for future store locations, as an estimate of future sales at new considered location would be effective for better capital/resource management.

Initially, we have removed the Nan and missing values from our dataset and coverted some categorical attributes into numbers, for example:

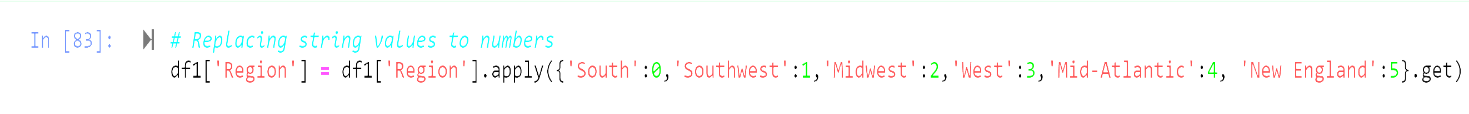


Figure 34 Code for removing missing values

Performing OLS (Ordinary Least Squares) Regression on our dataset as it is used to predict values of a continuous response variable using one or more explanatory variables and can also identify the strength of the relationships between these variables (these two goals of regression are often referred to as prediction and explanation).

Graphical user interface

Description automatically generated

Graphical user interface, text, application

Description automatically generated

**Figure 35** Output for Ordinary Least Squares (OLS) Regression

Here we observe that our prediction model has Durbin Watson test value close to 2.  A value of 2.0 means that there is no autocorrelation detected in the sample. In addition, More than 1 value for Jarque-Bera test suggests that the errors present in our dataset are not normally distributed. Moreover, due to presence of 82 attributes here, we observe that our R squared value is close to 1. Hence, we should be building our predictive model by creating a new sub-dataframe, which would only consists of important demographic information, which we consider impacts the target variable (‘Sales\_2020’) the most. Hence, we consider Region, Population density, Median Age, Median household income and median home value, to be the demographic information we need to provide to obtain predicted sales as output. Following is the correlation matrix for the same:

Graphical user interface, application, Teams

Description automatically generated

**Figure 36** Correlation matrix for predictor/input variables

From the above matrix, we observe that Population density, median household income and Median home value are negatively correlated to each other. Moving forward, we have created X1 (predictor) and Y1 (target) datasets as shown below:

Graphical user interface, text, application

Description automatically generated

Figure 37

Here, we have applied regression on our fitted model after dividing the dataset into train and test samples. Now, we try to predict the sales for our first 5 entries:

A picture containing text

Description automatically generated

Figure 38

Plotting our predicted values from above against out test dataset, we observe that most of the values fit exactly as in training dataset:

Chart, scatter chart

Description automatically generated

Figure 39

Finding MAE and RMSE values for Train and Test datasets:

Text

Description automatically generated

Figure 40

Hence, lower values of MAE and RMSE for our training models gives us the idea that the prediction model is pretty good. However, test model does not testify the same.

Therefore, we can conclude that for our predictive model, the predictions are quite accurate. But for the values which are not in the range (outliers), the accuracy in result outcome of predicted Sales may be affected which will be taken care of with the increasing entries in our datasets over time.

Final Model:

Graphical user interface, text, application, email

Description automatically generated

Figure 41

Here, we can enter the values as prompted regarding the new location demographics and obtain a predicted ‘Sales\_2020’ value for the considered location. Here, the sales value of 8425110 USD obtained should be compared with the total capital investment by senior leadership of AT Home group to decide whether to move forward with the location for opening a new store will be profitable or not.

Here we will include our interpretations for the questions posed in the data selection assignment. We observe that our response variable ‘Sales 2020’ representing total revenue for an individual store depends directly upon our predictor variables.

For our predictive model, we conclude that the predictions are quite accurate. But for the values which are not in the range (outliers), the accuracy in result outcome of predicted insurance cost may be affected which will be taken care of with the increasing entries in our datasets over time.

|  |
| --- |
| **CONCLUSION & RECOMMENDATIONS** |

There can be no doubt that the home furnishing industry had some interesting trends during the pandemic. From having a high demand in sales, retail outlets being forced to shut down and supply chain networks being disrupted. Although most customers still want and have missed the physical act of purchasing and going to retail outlets, the previous few months have demonstrated the need and strength of e-commerce and online ordering, especially in the home furnishings industry. For several consumer items, online purchasing may potentially change buying patterns within the first three months of the pandemic. On a positive note, though, particularly for the furniture industry, customers seem to have missed the physical shopping experience a great deal. With the provisions provided by At home group, customers still have the ability to walk in to get furniture or order online (Anon, 2020). With its great pricing strategy, the company was able to out-do most of its competitors and reaping huge monetary benefits from the pandemic than it has ever in its years of operation.

Using demographic statistics such as population, population rise, density, age, shop position, income and sales data from our analysis, our findings indicates some very valuable observations for the At Home audience to find areas for profitability. Utilizing the gathered and merged data, we faced several difficulties, some of which were lacking up-to-date sales and demographic data for some stores, however ultimately we were able to create visualizations and model our dataset to produce valuable information for the business.

Following our analysis, we discovered that some of our results were in accordance with our assumptions to the data question, while others were not, such as population growth not being a factor for increase in store sales. In addition, we saw that certain areas with a larger number of stores have not generated large profits over the years, so expansion could be essential in other productive areas. We can also conclude that areas with all service types may be more profitable that areas with some service types hence the company can look expanding its service types to more profitable areas based on our analysis.

Furthermore, some key players in profitability for At home group leading to 2020 value are Region, Population Density, Median age, Median household income and Median House value which were very helpful variables used in our analysis. From the comparison of some of our visualizations, we can conclude that the sum sales revenue growth rate in some states, although seemingly high, are contributed by some extremely high sales of a few successful stores. In particular, Michigan and Texas, mid-level stores of these two states saw sales growth of little more than zero in the past three years. Lastly, the pandemic had a huge effect on store revenue performance as all store locations saw a decrease in sales for 2020. However, stores with local delivery option was seen to do better than stores without local delivery option.

Based on our analysis, some specific recommendations to the company are as follows:

* The company can open new stores in the future in Texas, Ohio, Arkansas, Arizona and Missouri as these states have the potential for higher sales
* At home group can consider providing local delivery to more stores as stores with this option have much more profitability. Also, as the uncertainty of the pandemic increases, the company stands to make more profit by providing home delivery option to its customers
* Most of the store location was based in the south and southwest region, the company can look at spreading out to more profitable states in other regions like New England and mid-Atlantic regions
* Areas with the lower population density, median age, higher household income and home value should be prioritized over other as seen from our analysis in opening store locations

|  |
| --- |
| **REFERENCES:** |

* Anonymous. (2020). *Coronavirus Business & Economy Impact News | Business Insider*. Www.Businessinsider.Com. https://www.businessinsider.com/coronavirus-business-impact
* Thomas, L. (2020, July 31). *At Home CEO says retailer could grow to more than 600 stores as sales soar during pandemic*. CNBC. https://www.cnbc.com/2020/07/31/at-home-ceo-we-could-grow-to-more-than-600-stores-as-sales-soar.html
* *Anonymous Mordor intelligence*. (2017). Mordorintelligence.Com. <https://www.mordorintelligence.com/industry-reports/us-home-decor-market>
* *Global Home Furnishings Market Size in 2020 Industry Demand, Market Share, Trend, Industry News, Business Growth, Top Key Players, Showing Impressive Growth by 2026*. (n.d.). MarketWatch. Retrieved September 30, 2020, from <https://www.marketwatch.com/press-release/global-home-furnishings-market-size-in-2020-industry-demand-market-share-trend-industry-news-business-growth-top-key-players-showing-impressive-growth-by-2026-2020-09-24>
* *United States Home Decor Market Analysis & Forecast Report 2019-2024: Analyzed by Product & End-user - ResearchAndMarkets.com*. (2019, May 23). AP NEWS. https://apnews.com/press-release/pr-businesswire/56fd2a73fd6143cf9098e13e5dc66d07
* At Home Group. (2019). 2019 10-K form. [www.athomegroup.com](http://www.athomegroup.com)
* Vidojevic, B. (n.d.). *The Furniture Industry in the Time of Coronavirus: First Outlook*. Blog.Cylindo.Com. Retrieved September 30, 2020, from <https://blog.cylindo.com/furniture-industry-in-the-time-of-coronavirus-first-outlook>
* IKEA to sell through third party for first time on Tmall in China. (2020, March 10). *Reuters*. <https://www.reuters.com/article/us-ikea-tmall-china/ikea-to-sell-through-third-party-for-first-time-on-tmall-in-china-idUSKBN20X0AH>
* IBISWorld. (2020, June 1). *Coronavirus Update: Industry Fast Facts | IBISWorld Industry Insider*. Www.Ibisworld.Com. https://www.ibisworld.com/industry-insider/coronavirus-insights/coronavirus-update-industry-fast-facts/
* Anonymous(2020). Hfbusiness.Com. <http://hfbusiness.com/CurrentIssue/CurrentIssueModule/articleid/20384/statistically-speaking-three-months-of-the-pandemic-the-impact-on-the-furniture-and-home-furnishings-industry>
* *How the coronavirus altered DTC’s relationship with brick and mortar*. (n.d.). Retail Dive. Retrieved September 30, 2020, from <https://www.retaildive.com/news/how-the-coronavirus-altered-dtcs-relationship-with-brick-and-mortar/582772/>
* Chapter 3: Ordinary Least-Squares Regression. (n.d.). Retrieved December 10, 2020, from <https://methods.sagepub.com/book/the-multivariate-social-scientist/d49.xml>
* Kenton, W. (2020, October 21). Understanding the Durbin Watson Statistic. Retrieved December 10, 2020, from https://www.investopedia.com/terms/d/durbin-watson-statistic.asp

|  |  |  |
| --- | --- | --- |
| SOURCE | INFORMATION | LINK |
| At home Website | Store information i.e store name, zip code, address and service type | <https://www.athome.com/store-find/> |
| Yelp | Information on surrounding store competition per zip code | <https://www.yelp.com/> |
| United States Zip codes | Information on population, population type, trends, growth, income and housing. | <https://www.unitedstateszipcodes.org/35226/> |