Customer Demographics Report

Prepared by Thomas Higginbotham

# Overview

At Danielle Sherman’s direction, I have used data mining methods to explore the customer transaction data collected from recent online and in-store sales to infer any insights and conclusions about customer purchasing behavior, specifically:

* Are there differences in the age of customers between regions? If so, can we predict the age of a customer in a region based on other demographic data?
* Is there any correlation between age of a customer and if the transaction was made online or in the store? Or do other factors correlate to an online or in-store transaction?

# There are differences in the age of customers between regions

Chart, bar chart

Description automatically generatedFor this analysis, it was determined that the average age of customers was about 45 years, and for simplicity, four age brackets were used: 18-29, 30-44, 45-59, and 60-85. When these splits are grouped by region, it is apparent that there are differences in the age makeup of customers from region to region. The North region has more customers in the 30-44 age bracket, followed by the 45-59, 18-29, and 60-85 brackets. The South region does not have many customers in the 18-29 bracket, roughly equal numbers of customers in the 30-44 and 45-59 brackets, and the highest share of 60-85 customers for all regions. The East region has roughly equal numbers 30-44 and 45-59 customers, followed by 60-85 and finally 18-29 customers. The West region has the highest share of 18-29 customers of all regions, the highest share of 30-44 customers for all regions, the highest share of 45-59 customers for all regions, and relatively few 60-85 customers.

# It is not possible to predict a customer’s age based on either the region or other demographic information.

Table

Description automatically generatedUsing various methods, I attempted to predict a customer’s age based on the demographic information provided. This chart shows that the available methods could only predict any of the age brackets less than half the time. If we expanded the number of brackets to use decades, the accuracy would be even less. So in this case, my conclusion is that it’s not possible to predict the customer’s age based on the demographic information provided.

# As age increases, the amount purchased decreases

Chart, histogram

Description automatically generatedAt the transaction level, we see a very strong relationship between age and the amount purchased, demonstrating that as the customer age increases, the amount purchased in a transaction decreases. In this graph, we have a scatter plot showing every transaction plotted with the customer’s age and the amount purchased as a blue dot. Then, we see a black line showing the linear regression. It is noticeable that as customers reach their mid-60’s, the amount purchased declines sharply. There is another noticeable decline in purchases as customers reach their mid-70’s.

# Conclusions

Based on the analysis conducted, the following conclusions were reached

* There is a difference in the age of customers between the different regions.
* It is not possible to predict the age of customers based on demographic information.
* As the age of customers increases, the amount purchased is less.

# 