**Project 1, Task 2: Report to Blackwell Electronics**

Customer Age as an Indicator of Purchasing Habits

Anders Dowd - Spring 2022 Data Analytics Certification Program

1. *Overview*

Are there differences in the age of Blackwell customers between regions? If so, can the age of a customer in a region be predicted 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?

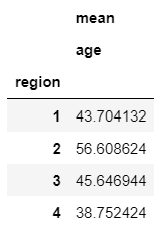
1. *Data*

Using the same sales data from our previous analysis, we built predictive models through machine learning algorithms to assess the relationship between a number of demographic variables.

1. *Analysis*

The first question is easy enough to answer. Taking an average of all customer ages by region, we can see that there is indeed a significant variance in average age by region. Table 1 shows a breakdown, with Regions 1, 2, 3, and 4 corresponding to North, South, East, and West, respectively. The South region has the oldest average customer age at 56.6 years, while the West has the youngest at 38.8 years.

**Table 1:** Average Customer Age By Region

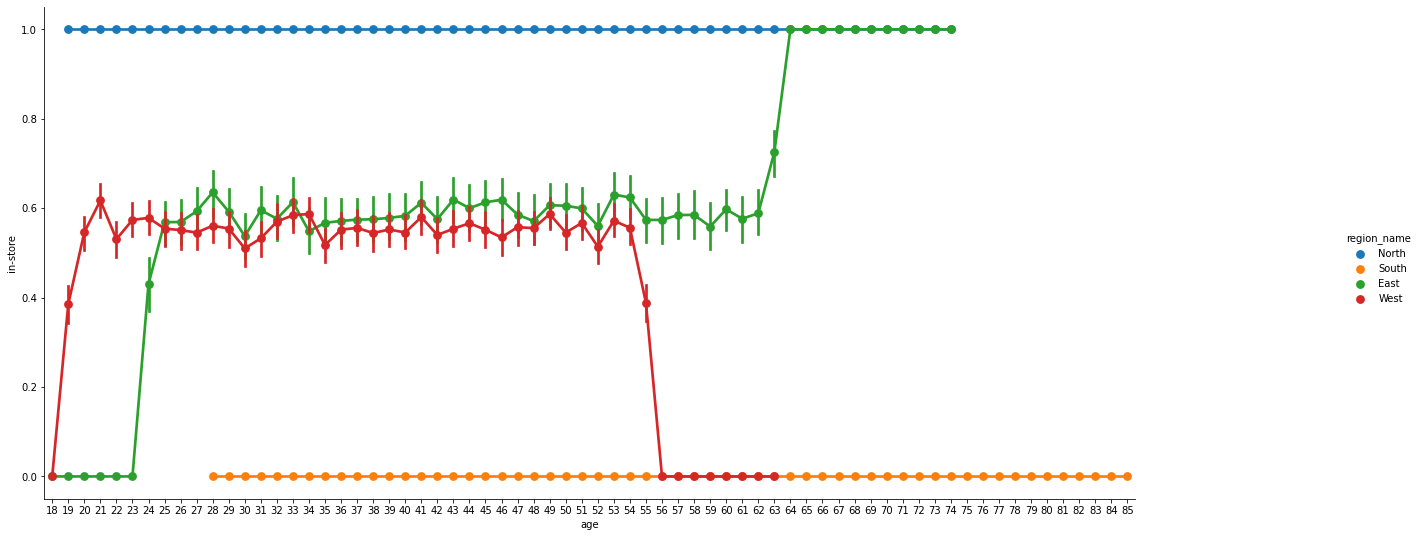
****

But can independent demographic variables be used to predict the age of an individual customer? Through machine modeling experiments it becomes evident that while the data points we have at our disposal are insufficient to accurately predict a precise age, we can predict with reasonable success an age bracket if we divide the data into two groups. Since Martin’s hypothesis is that ‘older’ customers shop in store and spend more money than ‘younger’ customers, let’s draw a line between these two descriptors. We could debate all day at what age a person officially becomes ‘old’, but it’s fair to say the world really starts keeping score at 30, so let’s make that our cutoff (not to mention the fact that to an average 18 year old, 30 sounds ancient). When we use our remaining demographic data points (in-store vs online sales, items purchased, amount spent, and region of sale) as our predictors of customer age, we can construct a model that predicts with an 82% success rate if a customer is over or under the age of 30.

If we want to set our cutoff for youth a bit higher, we can use another model to predict whether the customer is over or under 40. With this model, we can be 64% confident in our prediction. But can this information be used to predict pertinent purchasing habits that could help marketing efforts?

Let’s take a look at the breakdown of online vs. in-store sales by region and how it relates to customer age. Table 2 illustrates this relationship, with 0 on the ‘in-store’ axis representing online and 1 representing in-store. There are a few interesting bits of information we can glean from this table. Firstly, 100% of sales in the North region occurred in-store, while 100% of sales in the South region occurred online. This is of particular note since the South is the oldest region on average; this is an immediate indicator that there may be an issue with Martin’s hypothesis. However, the East region displays behavior that supports Martin’s inclination that older customers are more likely to shop in-store and younger customers are more likely to shop online. We can see that 100% of purchases by customers aged 18-23 in the East were in fact online, while 100% of purchases by those 64 and older were in-store. The age range in between rests near the middle, but leans toward in-store. Finally, in the West we see a unique tendency of a more-or-less balanced distribution of online vs in-store sales until we get beyond the age of 55, where we see 100% of sales online. But what can we actually predict based on this data?

**Table 2:** Customer Age Versus Sales Location By Region



When we use our ‘in-store’ feature on its own to try to elicit customer age, we can predict with 83% confidence whether that customer is over 30 or not, and with 61% whether they are over or under 40. When we try to go the other way, we’re essentially no better than a coin flip. Age alone is not a good indicator of online or in-store purchasing habits. However, if both age and region are known, we can achieve an accuracy of 77% in our predictions. Furthermore, if the amount spent is also added to the equation, we can achieve an impressive 86% success rate of predicting whether the purchase was made online or in-store. Knowing the number of items purchased had negligible impact on prediction success.

1. *Conclusion*

While a customer age range can be predicted using the other established variables, there is not a strong predictive formula to nail down a precise age of Blackwell customers based on the existing data. However, we can establish a strong predictive model to determine a tendency toward online or in-store purchasing habits based on age, region, and amount spent, which could provide valuable insight to inform a future marketing strategy. At the end of the day, region of purchase is the strongest indicator of both age and online vs. in-store shopping and should therefore be the primary consideration in future marketing strategies.