

Cleaning the Data

Duplicate Rows:

Removed duplicate rows, based on ID (as there must only be one row for each customer, whom is uniquely identified, via the "ID" feature).

Similar values in "Marital Status" & "Gender":

"M" shows up in both "Marital Status" and "Gender", which could be somewhat confusing. So to help clarify, the characters were substituted for words, as follows:

For "Marital Status":

- M = Married
- S = Single

For "Gender":

- M = Male
- F = Female

Sorting issue with "Commute distance":

Third, replaced "10+ Miles", with "Beyond 10 Miles", to fix sorting issue. (The feature is a string, according to excel, and is thus ordered by the first character, then the next, etc. Now, the first char of "10+ Miles" is a 1, and the first char of "2-5 Miles" is a 2, thus "10+ Miles" would go before "2-5 Miles", which is out of order).

Age as a category:

"age" is currently quantitative (individual years), and as this project centres around making a dashboard, it could be quite awkward working with age as individual years (in-case you wanted a histogram, or pie chart, etc, which requires far fewer unique values, hence age brackets would be much more appropriate).

So, a new feature, "Age Bracket" was created, using a VBA function and SELECT CASE statements.

Now that the data has been cleaned, we can begin to slice out bits of it, via Pivot Tables, and start building some dashboard elements.

Creating Dashboard Elements (Pivot Tables)

So, first, I figured that commute distance, and age, would be the biggest impact on whether someone would be a customer. (For example, those who commute 20 miles or more, are very unlikely to be customers, same with those aged 80 years or more). Thus, I looked at the distribution for both of these features, individually:

Commute Distance:

In the line chart of commute distance, we can see that most of our customers tend to have a shorter commute distance, and as a commute gets longer, we get less and less customers. Thus, we would likely want to increasingly lean away from advertising to customers, that have longer and longer commute distances.

So, what we could do as an advertising idea, is determine which locations have a heavy concentration of jobs, then look at all residential areas within a short commute distance, and advertise to the people there, that we offer an affordable, environmentally friendly, healthy, alternative means of transportation.

Age:

With the line chart on age, we can see that most of our customers tend to be older, around 31-45 years, with a decent amount at 46-60 years. So advertising to these age brackets would be wisest, as they're most likely to be customers (unless the market is considered completely tapped, in which case a risky expansion manoeuvre could be tapping into the under 18's market, as we have NO customers there).

Income Brackets, & Region:

Next, I looked at what income brackets our customers were in, for each region. The clustered column shows most of our customers in North America and the Pacific, have an income around \$60,000. Interestingly, most of our European customers have a much lower income, of around \$25,000.

Thus, it may be worth marketing new, smaller, more affordable accessories/products, to the European market.

Number of Children, by Region:

Finally, in our clustered bar chart, we can see that we have a rather large number of customers in North America, that have between 2 to 4 children. Thus, to help determine how viable it would be to expand into the under 18's market, we could survey these customers, to see how interested they would be, in a new line of bike products, dedicated to under 18's, such as their own children.