# Bike Sales Customer Insights

This project is based on Alex the [Analyst's YouTube video](https://youtu.be/opJgMj1IUrc) which he gives permission for the data and analysis to be reproduced, reused and modified.

## ****TLDR****

I cleaned the data, created two dashboards to give the client greater customer insight, then created a presentation to share this information.

[excel-project-dataset-bike-sales-finalDownload](https://michaelcollinsdataanalyst.files.wordpress.com/2022/04/excel-project-dataset-bike-sales-2.xlsx)

**Aim of the analysis:**   
Create a dashboard that stakeholders can use to better understand their customers. In particular, who bought bikes and who didn't after visiting one of their stores.   
Guide stakeholders to understand who they are converting into customers rather than just people browsing.

**Question:**In what ways do customers who bought or didn't buy products differ: age, income, distance to store etc.

**Procedure:**   
I will follow this workflow, outlined in the Google Data Analytics Course:  
1. Ask  
2. Prepare  
3. Process  
4. Analyse  
5. Share  
6. Act

## ****ASK****

1. What data does the company have on their customers/prospective customers? What information did they ask of each individual?
2. Does the company have any particular areas to focus on e.g. household income, distance from the store?
3. Where does the data come from? Is it internal or external? How were customers asked these questions?

**Assumptions:**   
- The data is internally collected, customers' responses are reliable to their individual experience, this sample group is typical of 'the average' customer and is representative of their diverse customer base.   
- Insights found will stay internal and be shared with the sales department, advertising as appropriate.

**My deliverable:**  
Clear and relevant charts to allow comparisons between different types of customers, relating to the data gathered.

[excel-project-dataset-bike-sales-rawDownload](https://michaelcollinsdataanalyst.files.wordpress.com/2022/04/excel-project-dataset-bike-sales-raw.xlsx)

Above is the raw data supplied by the client.

## ****PREPARE****

* The raw data supplied is under the company's license. Their team gathered it and sought appropriate permissions from everyone answering questions.
* None of the data allows for anyone to be personally identified (no addresses, emails, names).
* This creates some limitations: follow-up questions, if a customer visits or answers questions multiple times.
* The data is mostly fairly clean, with limited issues to resolve.
* One significant limitation of the data is there is no column to indicate how much the customer spent. As these are bike shops, the purchase of a bike, compared with a small accessory is a big difference.

## ****PROCESS****

As this data has been supplied for a case study with a focus on creating a dashboard (rather than as a cleaning exercise), unsurprisingly, there are few errors to deal with.   
1. First I familiarised myself with the data, looking at what demographic information was collected from each customer.   
2. Then I checked for duplicates (26 rows were removed), leaving behind 1000 rows.   
3. For clarity, I changed the Marital Status and Gender columns to full words: Married/ Single and Female/ Male.   
4. I checked the data types for each column, e.g. Numbers for a numeric field.   
5. Looking at filters for rows E-L: I checked for typos, unexpected entries. e.g. Children and cars should be whole numbers, probably under 10.  
6. For Age and Income there are lots of unique values, making it difficult to group customers. Using nested IF statements, I created Age Brackets (insert age boundaries). Then I used an IFS for Income Brackets (insert income brackets) to aggregate the data.

-- Nested IF statement to create age brackets =IF(M2>50,"3. Older", IF(M2>=35,"2. Middle age", IF(M2<35,"1. Younger", "Invalid"))) -- IFS statement to create income brackets =IFS(D2<=30000,"1 Lower Income",D2<=60000,"2 Lower Middle Income",D290000,"4 Higher Income")

## ****ANALYSE****

**Creating Pivot Tables:**   
1. Average Income by Gender  
The first thing I noticed is those who made purchases, whether male or female, was they are better off. This could have an impact relating to which income brackets to target for advertising.

Chart, bar chart

Description automatically generated

2. Comparing Customer Age brackets

Chart, line chart

Description automatically generated

I played around a little with the age boundaries: the minimum age was 25, with an average age at a little over 44. I chose 35 as a cut-off for being younger (perhaps my age had a little influence here) and 50 being the start of being older. This made sense to me given the data and that it is for bicycles, which might skew to a younger customer than many businesses. I would want to talk with the stakeholders more about these boundaries though before finalising a dashboard.   
The immediate observation is that only the middle age bracket is more likely to make a purchase than not. This could influence the sales teams to make sure younger and older customers have all the information they need to make a purchase.

3. Customer Commute

Chart, line chart

Description automatically generated

This chart is most useful when combined with further location information. Firstly, where an individual store is: is it in a city centre or more on the outskirts. Secondly, where does the customer live and work? This second piece of information would be a useful addition to further customer surveys.

Chart, line chart

Description automatically generated

Where this dashboard starts to really deliver interesting insights is when Slicers are added. In the gif above, you can see the data adjust as we look at just Married respondents, then Single people. This is easier to see in the Excel file:

[excel-project-dataset-bike-sales-dashboard1Download](https://michaelcollinsdataanalyst.files.wordpress.com/2022/04/excel-project-dataset-bike-sales-dashboard1.xlsx)

Similar changes can be observed when adjusting for region or level of education.

## ****SHARE****

As the dashboards will be used directly by the client, I will hide sheets they do not need to access to minimise the possibility of breaking any of the tables. This will leave the two dashboards and the workspace (so they can update data).

#### ****Presentation****

[customer-insights-for-bike-store-chainDownload](https://michaelcollinsdataanalyst.files.wordpress.com/2022/05/customer-insights-for-bike-store-chain.pptx)

Text

Description automatically generated with medium confidenceGraphical user interface

Description automatically generatedGraphical user interface, text, application

Description automatically generatedGraphical user interface, text, application

Description automatically generatedGraphical user interface, text, application

Description automatically generatedChart, line chart

Description automatically generatedChart

Description automatically generatedChart

Description automatically generatedChart, line chart

Description automatically generatedChart

Description automatically generated with medium confidenceChart, bar chart

Description automatically generatedChart

Description automatically generatedChart, bar chart

Description automatically generatedChart, bar chart

Description automatically generatedChart, bar chart

Description automatically generatedChart, bar chart

Description automatically generatedChart

Description automatically generatedGraphical user interface, text, application

Description automatically generatedGraphical user interface, application

Description automatically generatedGraphical user interface

Description automatically generated with low confidence

## ACT

The main takeaway from my presentation is the regionality of the data. I would invite further discussions about what the data should guide to do looking at each region specifically.

Europe: Clerical and manual workers are both most likely to make a purchase or browse. These groups should be kept in mind both for publicity and for the sales teams.

North America: There is a huge discrepancy between skilled manual workers and non-skilled. Why don't manual workers come to the stores, let alone make a purchase? More research should be done, as this is potentially an area for growth.

Pacific: They have the most evenly spread occupations. Is there something they are doing that other regions are not in attracting customers from across different professions?

#### ****Suggestions:****

This data should be paired with data about the locations of stores and population demographics they serve before any bold decisions should be made.