# Changelog

# Appendix

Data Journal

slides /// Presentation

# Data Cleaning Procedure

1. Check for duplicates

Each ID should be a unique value. After checking for duplicates 26 were found and removed

Graphical user interface, text, application, email

Description automatically generated

1. In the married column the data is labeled “M” for “married” and “S” for “single” To make sure that is clear I will use the find and replace function to write out the complete word.

Graphical user interface, text, application

Description automatically generated

1. I’ll repeat this process for the “S” replacing it with “Single”
2. I’ll also repeat this process for the Gender column changing “M” to “Male” and “F” to “Female.”
3. For the income column, the currency values are showing two decimal points. For the sake of readability and that all the values are whole numbers, I will change the format to not show the decimal points.

Table

Description automatically generated

Table

Description automatically generated

1. We have ages ranging 25 to 89. For Visualization purposes that we be used later, I’ll create 3 categories to sort the Age ranges into. To do this I will use a nested if statement.

   
This formula groups ages older than 54 into the “Old” category. Ages younger than 54 but older than 32 will be “Middle Age.” Ages that are younger than 32 will be considered “Adolescent.”

Table

Description automatically generated

Now, The data is cleaned and properly formatted. We’ll move to using pivot tables to analyze the dataset and see what kind of insights we can find with the data.

# ANALYZE

1. I create a chart showing average income for both male and female and whether they did or did not purchase a bike. This chart is to see if income is a factor in people deciding to purchase a bike. If so, should we cater to them? These are questions I Plan to answer soon.

Chart, bar chart

Description automatically generated

2. Next, I want to create a line chart showing the distance of customer’s commute and weather they did or did not purchase a bike. This will help us decide if the distance the customer has to commute is a factor in their decision to purchase a bike

Chart, line chart

Description automatically generated

3. I’ll utilize another line graph to show which age groups are purchasing/ Looking at bikes.

Chart, line chart

Description automatically generated

Lastly, with the adding of slicer menus on the left side, the charts will now change based on the metrics selected.

Chart, line chart

Description automatically generated