# BA222 Project 1 – Exploring and Analyzing Interesting Data in Excel

Due: Th, Sep 28 at 12PM

Fall 2023

## Description

Data come from a variety of real world sources. In this project, you will be identifying an interesting data set from the real world and using techniques from BA222 to analyze it. This is intentionally a very open-ended project.

## Groups

Please work in groups of 4. Your partners can be from any of **my** sections. All participants should put their name on the report and on the Excel spreadsheet. All participants should upload and submit the report and spreadsheet to the Questrom Tools site; we will only assign 1 grade but that will help me keep track.

Please plan to identify your partners no later than Friday Sep 15, and submit them in this form:

`https://forms.gle/x3UsmwN6GNX5e7du7`

If you are unable to find a partner, please email the TA Nicholas Antonelli (nantonel@bu.edu) and he will help you find a partner. Sooner the better, and no later than Sep 15.

## Data Sources

Data can come from any source you choose. Points will be awarded for style and creativity in selecting the dataset. I don't want 50 projects with the BU Covid Dashboard or the stock price data.

Please make sure your dataset has at least 500 observations and fewer than 250,000 (just to make sure that it can be opened in Excel without crashing). Make sure it has at least 8 variables. You are welcome to use a subset of a larger data set.

Please do not use simulated data.

Find data you are passionate about. This is an opportunity for you to learn more about something you care about. The data can be about sports, fashion, travel, gaming, the internet, real estate, businesses, literally anything of your choosing. Convey that passion in your analysis.

# Expectations

You will conduct an analysis in an Excel spreadsheet and produce a 2-5 page written report, both of which will be submitted.

Writing a good technical report is part of the skill set this class teaches. The quality of your writing will affect your grade. Please see the writing handout on the course website.

Here are some guidelines for what this should entail. Reports are expected to be well written and detailed. Please plan to exceed these guidelines; this is just a roadmap of how a basic analysis can be conducted.

Analysis:

- What are some summary statistics of the data?

- What are the ranges of values?

- Include some visualizations. For single variables, you can consider pie charts (categorical) or histograms (numerical).

- How do the data relate to each other? Are there any interesting correlations? Are there any patterns to be found? Consider making scatterplots.

Report

- Describe where you got your data set. Are these data publicly available? Where could one reproduce this data from? If you subsetted a larger data set, what are the criteria for subsetting?

- Describe your variables. What types of data are they? Try to give the columns descriptive column names, or make sure the variables are clear in your report. Ensure units are specific; e.g. is value in dollars or euros?

- How is the data quality? Do any data not make sense? Do you think these data are representative of a broader set observations? Did you need to remove any bad data? How does that change your results?

- What are some questions that could be answered with these data? What does your analysis tell you about these questions? Tell a story with your data.