

Midterm #2 → March 27
Due → March 15

Data Analysis Assignment ✕

Overview

we need graph and word doc.

This assignment asks you to utilize a data source, perform some analysis, present that data visually, and draw a conclusion and recommendation for action based on that analysis.

Data Source

Data can be sourced from anywhere that you can find something pertinent and suitable. The guidelines are:

- There needs to be a substantial amount of data. What this means will vary depending on the type of data you are examining – in general it will mean that several thousand rows of data are in the source.
- You need to be able to draw some recommendation for action from the data. E.g. visualizing the distribution of different colors of jellybeans isn't useful as there's no action to take with the results.

You can use data from almost any source, several are posted on Moodle under the assignment's module.

You will receive an extra mark for connecting your analysis tool to an external data source, e.g. storing your data in a database and connecting to that database with the analysis tool.

Visualization

The visualization that you are to produce is open ended and up to you. The key point is that it supports or illustrates your analysis and conclusion.

Visualizing data is not something that has one 'correct' answer. You are using this data to communicate your analysis and support your conclusion. In grading, more points will be given for visualizations that are:

- More clear.
- More illustrative.
- More visually engaging.

The visualization can be created in any number of different formats, different formats can also be combined if they support your point:

- An Excel graph.
- An Excel pivot table.
- A Tableau visualization.
- A report developed in Access' reporting tool.

Some factors that may help you develop an impactful visualization are:

- Choosing a suitable type of graph or chart for the data.
- Utilizing visual effects such as color coding, varying size, or customizing layout or axis settings.
- Creating interactive elements in your visualization.

There will be no difference in grading based on the tool used, only on the quality of the final product.

Analysis and Recommendation

In the analysis section you should interpret the data and explain what it means, while your recommendation should be a conclusion drawn on the data. An example is the Tableau that I demonstrated in class, based on the hotel data from the previous assignment – certain countries tend to rate hotels higher and have more total reviewers, so it may make sense for a hotel chain to target their advertising in those countries.

Deliverables

Your deliverables for this project will be:

- Visualization element or elements.
- Written analysis and conclusion in Word. (Probably half a page to a page of writing).

If your analysis produces a static graphic(s), please include that in the Word document. If your analysis produces something interactive, please either include a link to where I can find it (e.g. Tableau public) or the appropriate file (e.g. Excel pivot table).

Recommended, but not Mandatory, Process

- Find a set of data that appears like it may lead to some relevant possible recommendations. E.g. sales data that shows sales broken down by product and region may lead you to suspect that certain products might be more successful in different areas.
- Load the data into your tool.
- Provide a simple level of analysis, such as grouping and summing to see if it looks like there are any trends obvious in the data. E.g. if you have sales data as mentioned above, you might group by product and region (or make a pivot table doing that) to see if there are any trends.
 - Doing this should be quick, 15 - 30 minutes or so! You're just checking to see if it looks like there's anything there to analyze.
- Assuming you see a trend, think about what you can draw from that trend. Create an initial draft of your analysis and conclusion.
- Refine your visualization to focus on the trend and illustrate your intended point better.
 - This can include things such as adding coloring, adjusting the scale on charts, creating additional visualizations that support your points, adding conditional formatting in Excel, filtering out non-relevant information, etc...
- Revise your analysis into a final version that is in sync with the visualization.

Grading

The grading on this project will be split into two main categories:

- Does the project demonstrated the stated goals (10 marks):
 - Data is imported or connected to correctly.
 - Bonus for connecting to external data source (e.g. importing data into a DB, then connecting with Excel/Tableau/Etc...)
 - Data is presented visually.
 - A sensible conclusion/recommendation is drawn from the data presented.
- Quality of the presentation (5 marks):
 - Presenting data is more art than science. A higher quality visualization will be scored higher here. Some things that may bring quality are interactivity, multiple angles of visualization, clarity/understandability, etc...

When completing the project, focus on the first part initially. 2/3 of the marks is obviously enough to pass, so if you do a reasonable attempt, you'll get at least an OK score. The second part will give more marks for the better/more elaborate/more ambitious you're able to make your presentation.

We'll talk about this in class, so if you don't have a super clear idea on what this is asking, we can clear that up next week.