

# Second Project Guidelines

ECO225, Professor Khazra

Please start working on your project ASAP. Thinking about the best visualization format, understanding the structure of a map, coding, and fine-tuning your maps and visualizations will take awhile. Please plan to finish, one or two days before the deadline, to account for unprecedented situations and get help.

Your second project should include your first project. You will have two main sections for your project (and some subsections that you title yourself). The first main title should be "Project One," and the second main title should be "Project Two". You can add titles by adding "#" sign before your title in your markdown cell; check this link for more details on titling: <https://www.datacamp.com/community/tutorials/markdown-in-jupyter-notebook>

## Formatting

You will lose **15 points** if your submitted project does not meet any of the following requirements:

1. Submitted projects must be in PDF. Any other format, including a .ipynb file, is not accepted.
2. You should use markdown to write in your notebooks (Comments are just for readability of your codes). Check out lab notebooks to see how to have a title, subtitle, bold text, etc. You will lose points by not having a clear notebook.
3. The projects should have clear "section titles." Besides the subsections you choose to have, you should have two main sections; Project One and Project Two.
4. In case you missed this in your first project: choose a title for your paper. What is the question that you would like to answer using this dataset? The question can be the title of your paper.
5. You should write your project in Jupyter Notebook (Python) and submit it in pdf format. If you have problems converting your notebook to pdf, first download it as HTML, then print/save the HTML version in pdf.
6. This is an individual project. However, you are encouraged to check the projects on the Kaggle.com website that use similar data. We have provided some "useful

links" on the data list on Quercus. You can use these sources, but the coding and explanations **must be yours**. Do not copy and paste the same chunk of code in your project.

Please note that if you include a graph or a table, you should **explain what you learn from it**. Do not add an output (graph/table/number) without any explanation. All projects should have an introduction and a conclusion. Suppose you want to send your project to a company or school that you are applying to. The final product should be a clean and comprehensive report.

- Do not include unnecessary chunks of code or outputs and errors. You will lose mark for including these.
- Any graph, summary, or output should have an explanation following. Why do you include it in your report, and what do you understand from it? Graphs and output with no explanations or no economic intuition will not be awarded any marks.

**IMPORTANT:** Show your code and results, and provide economic explanations for your results for all parts. We will evaluate your work's quality and accuracy; simply answering all questions does not guarantee a full mark. These are the minimums to pass, try to go beyond the instructions.

## Second Project Details

We will build upon the first project towards a full academic paper on a hot topic using real-world data and cool Python techniques, just excellent! Here's what we are going to do for this project.

### Part One

1. **(10 points)** Incorporate the comments that you have received in your first report. This is the first part of your second project, which will be graded again. We will check the first section of your second project (your updated first project) and grade it again. If you have not received any comments, try to improve on your first report by doing a better job on the introduction or literature review or fine-tuning your visualizations, or adding more meaningful analysis. You will always revisit your past projects until the final project (your masterpiece).
2. **(10 points)** Fine-tune all graphs in your first and second projects. (e.g. proper title, labels, colour, scale, etc.). It is recommended to avoid pie plots unless necessary.

## Part Two

From the visualization lecture we learned that there are 4 steps in creating an informative visualization:

1. Identify the message.
2. Describe your visualization.
3. Create a draft of the visualization (and verify your data!).
4. Fine tune the visualization details.

1. **(10 points)** Think about the main message/question of your paper. You still don't know the precise answer to that question, but you can use visualization to give us some intuition about the possible answer(s). Please write down your message (title it THE MESSAGE!) and go to step two.
2. Use pen and paper to draw what you have in mind.
  - Can you plot your Y (dependent variable) by subgroups? Examples: by gender, type (whatever type makes sense for your project), population (create a variable high/low pop), developing or developed countries, characteristics of the firm, income (create a column high/low), share of minorities, etc.
  - Notice that you can create histograms, scatter plots, line plots, etc.

What is on the Y axes? What is on the X-axis? What colors are you choosing? What line style are you going to choose if it is a line plot? Think about all the details and draw it on a paper.

3. **(15 points)** Fine-tune all graphs in your first and second projects.

## Part Three

For this part, you will complete and fine-tune your maps from the maps lecture exercise. This is now part of the pdf that you will turn in and should be the last section of your paper (before conclusion). The guidelines below are very similar to what you saw for the exercise. **(45 points)**

- Create at least two maps. One for your outcome variable (Y) and one for your most important independent variable (X). Clearly explain your maps. (e.g. What can you learn from the maps? Do you see the same relation between X and Y as before in your maps? How are they related to your previous findings? Do you see any outlier location?)
  - If your dataset does not have a proper X variable for mapping, you can map your outcome variable in for two different periods of times.

- **(10 bonus points)** Alternatively, you can merge your data with another relevant and interesting dataset and the same geographic level. If your dataset already have the proper Xs you can still earn this bonus point by complimenting your data with other datasets. Have a clear title and explain the new dataset if you use this.
- You should know the geographic level that you are working with. Is it Zip code? Is it county? Is it state? Is it a country? Check the column that contains this information. Is it nice and clean? What is the format? Are there any missing data?
- Go to the "useful links" column on the spreadsheet where you chose your preferred dataset. I have added the spreadsheet to the Module for your convenience as well. Check the column titled "useful links." Check the sample code and see if there is code for a map and if you would like to modify the code for your use.
- Familiarize yourself with the U.S. Census Shapefiles. Look at the Wisconsin voting example in the maps notebook. I have provided the Census Shapefile links in the spreadsheet in the 4th and 10th row. You can use these links as in the Wisconsin example and create your maps.
- You should create a new data frame with the unique location names. All of your datasets are in panel format, which means you have more than one row for each location. Therefore you need to first create a dataset with the unique values for each geographic level (whatever value you want to plot) and the unique names or IDs for the geographic levels (zip codes, cities, etc.), put it in a geo-spatial data frame, and then map it. For example, if you have created an average measure of SAT grades for each zip code. You want a data frame with these columns: (zipcode, average SAT for each zip). You can then merge this with your zip code shapefile (like we merged county with poll results in the Wisconsin example) and plot. Another strategy is to plot all the points (if you have long/lats for all your observations) and have a heat-map style visualization.
- There will be more complications. For instance, the name of the city or county, or state may have extra strings, may be in lower case or upper case, and therefore your merge with the shapefile may fail. Always check the values in both your dataset (containing the original location) and your shapefile. Leave 3-4 days for errors and for figuring out the sample code.
- Make sure to take advantage of the existing notebooks on Kaggle. Feel free to read, understand, and use their code. Look for a map and try to understand the code first and then apply it to your notebook. This is not cheating; it's how you learn to code. Simply copy-pasting someone's code is cheating.
- You do NOT need to use the same libraries as we used in the lecture necessarily. What I showed you was a popular library, but there are many more. Get creative and explore. You can also look at other suggested notebooks for the other Excel spreadsheet datasets to get inspired.
- Polishing your map should be easy at this point because you have a draft from your week's exercise!

- **(5 bonus points)** for nicer maps that have a special feature like hovers or maps that are very well fine-tuned! If you create a map with bokeh library, it may not show the feature on your pdf. In that case, add a link of your HTML version to your projects (and make it somehow visible so that we don't miss it). As always, you cannot appeal to the bonus points.

**(10 points)** Don't forget to include your updated introduction and conclusion. Write about your new exciting findings. You should re-write your intro and conclusion and include the comments you received. If there are no meaningful changes to your previous intro/concl, the 10 points will not be awarded. You need one introduction and one conclusion for the entire paper (not two intros and two conclusions). You will keep revising these parts in the next project and in your final project.

**Upload your Jupyter Notebook (your code and explanation) in pdf format on Crowdmark for marking.**

Enjoy being a creator!