COMP 4304 – Course Project

MEMORIAL UNIVERSITY

Project Proposal Due: **11:59pm, Feb. 27, 2023**
Project Checkpoint Due: **11:59pm, Mar. 20, 2023**
Final Project Due: **11:59pm, Apr. 5, 2023**

Introduction

You will act as a freelance data visualization specialist for your course project. You will have a client ask for your help to answer their key business needs or questions using a data set that they have provided. You will need to create visualizations from this data set to satisfy their needs.

Client

Your client is the Transportation Safety Board of Canada. They are a governmental agency responsible for advancing the safety of airway, railway, and marine transportation in Canada.

They have provided you a data set (`flights.csv`) containing information about all commercial and private aviation occurrences (accidents) occurring within Canada over the past 25 years. Occurrences are, for example, emergency landings, an aircraft leaving a runway, or two aircraft colliding. Information is included about the aircraft involved, the time of the occurrence, the location, the extent of any damage or injured personnel, and more.

Their broad goal is to answer the question, "Where should efforts be focused to improve the safety of Canadian air space?"

Project Overview

Using the provided data set, you will need to explore the data to find the interesting stories contained within that address the broad goal. This is a *broad* goal. The client's question can be answered in multiple possible contexts. For example, you might find interesting differences by geographic region or type of occurrence (among many other possibilities).

You do not need to solve the problem of Canadian air space safety. Instead, highlight for the client some areas where there is potential or need for increased safety, or, conversely, areas where further safety efforts may be unnecessary.

You must identify two sub-goals within the data that you plan to convey to your client. These, along with the broad goal, provide the underlying motivations for the visualizations that you will create. Each should be a more specific interpretation of the broad goal. The sub-goals could be in the form of a statement or question.

You will deliver 4 visualizations to your client. These should relate to the main goal or the specific sub-goals that you have identified. Of the 4 visualizations, there must be:

- 1 explanatory visualization,
- 1 exploratory visualization,
- 1 interactive visualization, and
- 1 visualization that contains derived data types.

Each visualization may contain more than one of the above elements, but all 4 of the above elements must be represented at least once in your set of visualizations. For example, you could create 2 explanatory visualizations using derived data, 1 explanatory visualization with interactive elements, and 1 exploratory visualization.

Furthermore, of your 4 visualizations, there needs to be at least 3 different standard types of visualizations (e.g., bar plot, scatter plot, etc). That is, you will need to create a diversity of visualization types. For example, do not include 3 histograms in your set of visualizations, although it would be acceptable to have 2 histograms, a heat map, and a box plot.

Groups

You are allowed to work in pairs of 2 for the course project. This is not a requirement, and you are allowed to work on the project individually if you so choose.

Please email me with who is your project partner before you submit the Project Proposal. Also include you and your partner's name in every submission. Only one submission is required per group, but make sure you identify who is your partner.

Project Proposal (10%)

Due: 11:59pm, Feb. 27, 2023

Submit a proposal to your client outlining the two sub-goals you have identified and at least 2 initial visualizations you have made as part of your exploration of the data. These initial visualizations do not need to be finished products. They only need to be functional, in terms of labelling, appropriateness of plot type for the data, and correctness of any data manipulation required to generate the visualization. Fine-tuning their visual appeal can come in later stages of the project.

Describe the purpose of each initial visualization and how each is connected to the main goal or your sub-goals. Describe your planned next steps for your project. These next steps should describe two out of three of the following areas: any deeper exploration of the data that you expect to accomplish, any refinements to the current visualizations that you intend to make, and/or additional visualizations that you envision creating for the final product.

Your proposal should be submitted as a `.pdf` with standard 8.5×11 size (US Letter). This is your initial proposal to your client, and should look professional. Include your name(s). Text should be 12 pt font size. A maximum of 2 pages is allowed.

Proposals will be graded on their:

- Length of proposal (2 pages or less). (**1%**)
- Formatting and overall visual appeal of the proposal. (**1%**)
- Identified sub-goals. (**2%**)
- Functionality of the two initial visualizations. (**2%**)
- Described purpose of each visualization. (**2%**)
- Planned next steps for the project. (**2%**)

Project Checkpoint (10%)

Due: 11:59pm, Mar. 20, 2023

Submit in-progress versions of your 4 visualizations for review by your client. At least 1 visualization should be complete in terms of its final composition and finishing touches. The remaining 3 visualizations may be initial drafts, similar to the draft visualizations submitted for the proposal.

Label each of the 4 visualizations as either explanatory or exploratory. If it is explanatory, write the interpretation the visualization is intended to convey. What is the takeaway message the viewer should understand from this visualization? If it is exploratory, write two inquiries that you anticipate your viewer might explore.

Submit your project checkpoint as a `.pdf` with standard 8.5×11 size (US Letter). The pdf need only contain your 4 visualizations along with accompanying explanations. Include your name(s). Text should be 12 pt font size.

Your project checkpoint will be graded according to:

- Exploratory and explanatory descriptions to each visualization. (**3%**)
- Diversity of visualization types. At least 3 standard visualization types are expected. (**2%**)
- Functionality of the 3 draft visualizations. (**3%**)
- Final product visualization
    - Elements of style, such as colour, annotations, and spatial layout. (**1%**)
    - Choices on angle, frame and focus of the visualization. (**1%**)

Final Project Submission (30%)

Due: 11:59pm, Apr. 5, 2023

Your final project submission consists of 3 parts:

- Your 4 final visualizations. (**12%**)
- The code used to create those visualizations from the original dataset. (**8%**)
- A pre-recorded video presentation to your client walking through your visualizations. (**10%**)

Include a text file describing the files in your project submission.

Your visualizations will be graded according to the following:

- Submission has 4 visualizations (and only 4). (**1%**)
- Representation of explanatory and exploratory visualizations. (**1%**)
- Diversity of visualization types. At least three visualization types are expected. (**1%**)
- Appropriateness of visualization type for the chosen data and alignment with stated goal or sub-goal. (**1%**)
- Functionality, responsiveness, and usefulness of interactive components. (**2%**)
- Elements of style, such as choices of colour, annotation, line or point style, and other stylistic components. (**4%**)
- Clarity of each visualization (decisions on angle, frame and focus). (**2%**)

Your code will be graded for its correctness according to the following:

- Code to derive new data types and any pre-processing of data. At least one of your visualizations must include a derived data type. (**3%**)
- Code implementation of each visualization. (**4%**)
- Your notebook(s) work as expected when Restart Kernel and Run All is performed. (**1%**)

For your video presentation, you will walk your client through the visualizations that you have made. Introduce the problem in terms of the stated broad goal and any sub-goals you have identified. Describe each visualization to your client. Provide context for each visualization in terms of how it relates to the goal and sub-goals, and why you have chosen this design. Your presentation is your opportunity to tell the stories contained within your visualizations to your client.

Do not walk your client through your code in your presentation. The focus of your presentation is on the stories you are telling through your visualizations.

Your video presentation will be graded as follows:

- Length of video (5–10 minutes). Videos over 10 minutes will receive no marks on this aspect. Videos under 5 minutes will receive half marks. (**1%**)
- Explanation of each visualization. How should your client understand your visualization? How can your client use any interactive elements? What is the purpose of the visualization? How does it help the client achieve what they need? What are the key takeaways that your client should read from each visualization? (**6%**)
- Overall cohesiveness of the story that you tell through your presentation. (**3%**)

Submission

Submit your project through Brightpsace.

Your 4 visualizations can be submitted in whichever file format makes sense. They could be `.pdf` or `.png` files, or submitted inside a Jupyter notebook (`.ipynb`; this may make the most sense for interactive elements).

The code to handle any data manipulation and to create your visualizations should be in Python, either inside of a Jupyter Notebook or as Python source files.

Your pre-recorded video presentation can be in an uploaded video file or a YouTube link.

Only one member of your group needs to submit your project files, but make sure you clearly state who is your project partner.

Late submissions will be subject to a 20% penalty for each day past the deadline (proposal, checkpoint and final submissions).

Attribution

Submissions should include an attribution section indicating any sources of material, ideas or contribution of others to the submission.

Submissions must represent your independent work.

You are encouraged to use any resources to help with your solution, but your solution must represent independent work. If your submitted work includes unacknowledged collaboration, code materials, ideas or other elements that are not your original work, it may be considered plagiarism or some other form of cheating under MUN general regulations 6.12.4.2 (4.12.4.2 for graduate students) and academic penalties will be applied accordingly.

Avoid academic penalties by properly attributing any contribution to your submission by others, including internet sources and classmates. This will also help distinguish what elements of the submission are original. You may not receive full credit if your original elements are insufficient, but you can avoid penalties for plagiarism or copying if you acknowledge your sources.

Github

I encourage you to store and version your work on GitHub. It is good practice to do so as everyone uses git in the real world.

However, **it is a requirement that git repositories containing assignment material be private.** University regulations (undergraduate 6.12.4.2 and graduate 4.12.4.2) consider it cheating if you allow your work to be copied. There will be zero tolerance for this.