**DSC 520 Final Project Template**

This handout is meant to provide you with some structure for your final project. Note that there are two parts to this handout. First is the description of what you will be doing in each section and their general due dates. Second is the template itself with the various headings.

You are free to adjust this as needed. However, given what you’ll be doing, it is advantageous to you to just follow the template as it will help keep you focused. Should you have any questions, please ask!

**Section 1 – Week 8 – Getting Started**

* Provide an introduction that explains the problem statement you are addressing. Why would someone be interested in this?
* Draft 5-10 Research questions that focus on the problem statement.
* Provide a concise explanation of how you plan to address this problem statement.
* Discuss how your proposed approach will address (fully or partially) this problem.
* Do some digging on a dataset that you can use to address the issue.
  + Original source where the data was obtained is cited and, if possible, hyperlinked.
  + Source data is thoroughly explained (i.e. what was the original purpose of the data, when was it collected, how many variables did the original have, explain any peculiarities of the source data such as how missing values are recorded, or how data was imputed, etc.).
* Identify the packages that are needed for your project.
* What types of plots and tables will help you to illustrate the ﬁndings to your research questions?
* What do you not know how to do right now that you need to learn to answer your research questions?

**Section 2 – Week 9 – Cleaning Your Data and Exploratory Data Analysis**

* Data importing and cleaning steps are explained in the text and in the DataCamp exercises (tell me why you are doing the data cleaning activities that you perform) and follow a logical process.
* With a clean dataset, show what the final data set looks like. However, do not print off a data frame with 200+ rows; show me the data in the most condensed form possible.
* What do you not know how to do right now that you need to learn to import and cleanup your dataset?
* Discuss how you plan to uncover new information in the data that is not self-evident.
* What are different ways you could look at this data to answer the questions you want to answer?
* Do you plan to slice and dice the data in different ways, create new variables, or join separate data frames to create new summary information? Explain.
* How could you summarize your data to answer key questions?
* What types of plots and tables will help you to illustrate the findings to your questions? Ensure that all graph plots have axis titles, legend if necessary, scales are appropriate, appropriate geoms used, etc.).
* What do you not know how to do right now that you need to learn to answer your questions?
* Do you plan on incorporating any machine learning techniques to answer your research questions? Explain.

Suggestion from the course professor: Some additional questions you may want to consider asking yourself as you work through this section of the project:

1. What features could you ﬁlter on?
2. How could arranging your data in different ways help?
3. Can you reduce your data by selecting only certain variables?
4. Could creating new variables add new insights?
5. Could summary statistics at different categorical levels tell you more?
6. How can you incorporate the pipe (%>%) operator to make your code more efﬁcient?

**Section 2 – Week 10**

* Overall, write a coherent narrative that tells a story with the data as you complete this section.
* Summarize the problem statement you addressed.
* Summarize how you addressed this problem statement (the data used and the methodology employed).
* Summarize the interesting insights that your analysis provided.
* Summarize the implications to the consumer (target audience) of your analysis.
* Discuss the limitations of your analysis and how you, or someone else, could improve or build on it.
* In addition, submit your completed Project using R Markdown or provide a link to where it can also be downloaded from and/or viewed.

Name:   
Date:   
Title:   
  
**Section 1**

* Introduction
* Research questions
* Approach
* How your approach addresses (fully or partially) the problem.
* Data
* Required Packages
* Plots and Table Needs
* Questions for future steps.

**Section 2**

* How to import and clean my data
* What does the final data set look like?
* Questions for future steps.
* What information is not self-evident?
* What are different ways you could look at this data?
* How do you plan to slice and dice the data?
* How could you summarize your data to answer key questions?
* What types of plots and tables will help you to illustrate the findings to your questions?
* Do you plan on incorporating any machine learning techniques to answer your research questions? Explain.
* Questions for future steps.

**Section 3**

* A story / narrative that emerged from your data. Follow this structure.
  + Introduction.
  + The problem statement you addressed.
  + How you addressed this problem statement
  + Analysis.
  + Implications.
  + Limitations.
  + Concluding Remarks