**Project Goal**

The goal of this project is to import at least 3 datasets that interest you, define a business and analytics problem, work with the data to clean and tidy, and perform data analysis; all while using R Markdown to produce an HTML report that is fully reproducible.

### Project Data

The final data set should have more than 15-20 data points of interest (variables/attributes) and at least one thousand rows/records, but more is welcome and likely needed.

**Possible Data Sources**

### <https://hub.mph.in.gov/>

### <https://catalog.data.gov/dataset>

### <https://github.com/awesomedata/awesome-public-datasets>

### <https://datasetsearch.research.google.com/>

### Project Report

You will write an [R Markdown](http://wfu-r.github.io/r_markdown) HTML report that provides the sections in the project outline below. You will need to import, assess, clean & tidy the data, and then come up with your own research questions that you would like to answer from the data by performing exploratory or explanatory data analysis. Some thoughts to help you:

* [Make a storyboard](https://www.vox.com/culture/2018/11/20/18099001/animated-films-storyboards-script-story-ralph-breaks-the-internet). Your project should be a logical, cohesive story–not simply a bunch of graphs created for the sake of making them.
* Simple descriptive statistics can (and usually) yield more of an immediate impact than a complicated model if presented correctly, but predictive and prescriptive models should be used
* Here is an example project but yours should be more detailed in Domains I-II: <https://rpubs.com/yash91sharma/dw_project_ys>
* You can import Tableau sheets into R Markdown if you would prefer to use Tableau than R

### Project Outline

**Proposal**– Provide background and context of the topic and business/organization of the applied analytics project, data source, and project plan for *Applied Analytics Project.*

**Introduction- Domain I & II** (150 points) – (I) Understand a business problem and (II) reformulate the problem into an analytics problem with a potential analytics solution. In short, this is an introduction to your project. *This should be like a white paper you would submit to key stakeholders and will require A LOT research on the business/organization/or analytics problem you are trying to address.*

* + - Provide an introduction that explains the problem statement you are addressing. Why should we be interested in this?
    - Provide an explanation of how you plan to address this problem statement (the data used and the methodology employed)
    - Discuss your current proposed approach/analytic technique you think will address (fully or partially) this problem.
    - Explain who are your stakeholders and how your analysis will help the stakeholders of your analysis.

**Data Preparation- Domain III** (75 points) – Work effectively with data to help identify potential relationships that will lead to refinement of the business and analytics problem. In short, this section includes data preparation

* 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.).
  + You should use at least 3 separate datasets
* Document data cleaning steps (tell me why you are doing the data cleaning activities that you perform).
* Provide summary information about the variables of concern in your cleaned data set.

**Analysis- Domain IV & V** (150 points) – (IV) Identify, select, and (V) build approaches for solving the business problem. In short, this section includes the evaluation of R packages, selection of R packages, and exploratory analysis. You can also use Tableau or any other statistical package.

* All R packages described/included at the beginning of the script, so the reader knows which are required to replicate the analysis.
* Explanation is provided regarding the purpose of each package
* Uncover new information in the data that is not self-evident (i.e. do not just plot the data as it is like you do above; rather, slice and dice the data in different ways, create new variables)
* Provide findings in the form of plots and tables. Show me you can display findings in different ways.
* Graph(s) are carefully tuned for desired purpose. One graph illustrates one primary point and is appropriately formatted (plot and axis titles, legend if necessary, scales are appropriate, appropriate, etc.).
* Table(s) carefully constructed to make it easy to perform important comparisons. Careful styling highlights important features. Size of table is appropriate.
* Insights obtained from the analysis are thoroughly, yet succinctly, explained. Easy to see and understand the interesting findings that you uncovered.

**Report Summary- Domain VI** (200 points) - deploy the selected model to help solve the business problem

* Summarize the problem statement you addressed.
* Summarize how you addressed this problem statement (the data used and the methodology/packages employed).
* Summarize the interesting insights that your analysis provided.
* Summarize the implications to the consumer of your analysis.
* Discuss the limitations of your analysis and how you, or someone else, could improve or build on it.
* Proper coding style is followed and code is well commented (<http://uc-r.github.io/basics#style>)
* Rmd fully executes without any errors and HTML produced matches the HTML report submitted by student.
* Upon submission you will upload the final HTML report to RPubs and provide me with the URL. You will also submit the .Rmd file that produced the HTML report, your data, and any other files your .Rmd file leverages (images, .bib file, etc.).

Here is a great source to use throughout the project: <http://uc-r.github.io/introduction>

**Academic Reflection Paper (150).** The reflection paper should provide an overview of the theories, methods, and tools used during the Applied Analytics Project. I would recommend using the Certified Analytics Professional Guide, CRISP-DM material, any articles or textbooks used within the MSDA program to back up the decisions you made during the project.

* The paper should be roughly 10 pages
* Should address the below rubric points

**Analytics Project-- Zoom Presentation** *(1 @ 100 each).*The report summary will be recorded via Zoom and should be an overview of the Report Summary and Academic Reflection Paper. In other words, it should be a synthesis of the entire capstone experience.

* + The presentation should be roughly 30 minutes.

***Academic Reflection Paper Grading Guidelines***.

**Content/Development—70%**

|  |  |
| --- | --- |
| *Subject Matter—40%*   * Key elements of assignment covered (summarize the findings and then discuss theories, methods, and tools used during the Applied Analytics Project) * Content is comprehensive/accurate * Displays an understanding of relevant theory * Major points supported by specific details/examples * Research is adequate/timely * Writer has included appropriate number of sources (8-10) |  |
| *Higher-Order Thinking—30%*   * Writer compares/contrasts/integrates theory/subject matter with process used during Applied Analytics Project * At an appropriate level, the writer analyzes and synthesizes theory/practice to develop new ideas and ways of conceptualizing and performing |  |

**Organization—10%**

|  |  |
| --- | --- |
| * The introduction provides a sufficient background on the topic and previews major points * Central theme/purpose is immediately clear * Structure is clear, logical, and easy to follow * Subsequent sections develop/support the central theme * Conclusion/recommendations follow logically from the body of the paper |  |

**Style/Mechanics—20%**

|  |  |
| --- | --- |
| *Format--10%*   * Citations/reference page follows guidelines * Properly cites ideas/info from other sources using APA * Paper is laid out effectively--uses, heading and other reader-friendly tools * Paper is neat/shows attention to detail   *Grammar/Punctuation/Spelling--10%*   * Rules of grammar, usage, punctuation are followed * Spelling is correct   *Readability/Style--10%*   * Sentences are complete, clear, and concise * Sentences are well-constructed with consistently strong, varied structure * Transitions between sentences/paragraphs/sections help maintain the flow of thought * Words used are precise and unambiguous * The tone is appropriate to the audience, content, and assignment |  |

Comments / Grade

|  |
| --- |
|  |