

## Final Project Format and Instructions

### Introduction to Data Analytics – ITE 5201 - Humber College

#### **Section 1: Introduction**

For the final project, each Group will choose a small dataset, define a problem statement, perform statistical analysis, build a model, and report results and observations. Here are the steps:

1. Find a dataset that you are interested in to work on. Here is list of suggested resources:

a. <https://www.kaggle.com/>

b. [https://www.cs.ucr.edu/~eamonn/time\\_series\\_data\\_2018/](https://www.cs.ucr.edu/~eamonn/time_series_data_2018/)

2. Choose a subset of the features and samples of the dataset that you think is appropriate and define a problem statement. The problem statement must be short and in a single line. For example: "Prediction/Analysis of house market in down-town Toronto using immigration data".

3. Discuss the dataset and your observations. For example, perform visualization, univariate, bivariate, and multivariate analysis on the selected dataset.

4. Build an appropriate model to address the proposed problem statement. For example, a regression model or classification model. You need to discuss why you have selected a model.

5. Train and evaluate the model on the training and test datasets, respectively, and present the results and observations. Use visualization methods and supporting paragraphs.

Each Group will deliver the followings:

- A final report, according to Section 2.
- A final presentation, according to Section 3

#### **Section 2. Final Report Structure**

A template for the report is provided. Follow the following instructions in writing your project report:

1. First (title) page: Project title, Student names and ID
2. Submission format: .pdf
3. Maximum number of pages (excluding the title page): 2 pages

#### **Section 3. Final Presentation Structure**

1. Each Group will have 15 minutes to present (6 minutes each member of the group + 3 minutes questions and answers). (This allocated time is anticipated and subject to change)

2. Each Group presentation must have a cover slide + six slides:

a. Slide 1 & 2:

i. Problem Statement.

- ii. Dataset description (e.g. #samples, features, etc.).
- iii. Statistical plots of the dataset (e.g. bivariate analysis).

b. Slide 3&4:

- i. Analytical/Predictive model diagram (relationship between inputs, model, and outputs).
- ii. Discuss why and how this model is designed to address the problem statement.

c. Slide 5&6:

- i. Visualization of analytical and inference results.
- ii. Discuss what you can infer from the visualizations.

3. A presentation template is provided in the attachment.

#### **Section 4. Evaluation and Timelines**

The final mark will be calculated based on the quality of the report, presentation, and response to the questions during the presentation. A 20% penalty will be applied if the presentation or report templates and instructions are not followed. No late submission is allowed.

Presentation days: April 12<sup>th</sup> and April 19<sup>th</sup>, Lecture time.

Report submission deadline: April 19<sup>th</sup>, at 11:59pm

If there are any questions, please send me an email.