

# Introduction to Data Science

## Course Project – Assignment

### 1 Data Collection

Describe the source of your data or provide details of how you collected the data.

### 2 Data Format Description

Describe the format of your dataset in 1-2 paragraphs. E.g.: What files are included? What is the file format / structure of the data? What are the most relevant attributes within each file and what do they mean?

### 3 Descriptive Statistics

Analyze some of the basic characteristics of the data values. For example: What is the range of values? What is the mean, standard deviation, etc.? Can you plot the distribution of values for numerical variables? What is the distribution of categories for categorical variables? You can optionally visualize some of these.

*Note:* This is an open-ended task and the kinds of statistics you can compute will depend on a lot on your dataset. For example, for textual data, you could report top keywords, average length, etc.

### 4 Data Analysis, Visualization, and Insights

Are there any interesting insights that you can derive from the data? This may include the questions you originally planned to address in your course project proposal, as well as any other insights you have found. Examples include interesting correlations, interesting

differences in a variable value between categories of another variable (e.g., the average salary of people computed separately for people from each country), etc.

Include at least one visualization based on altair/vega/vega-lite (include the source code). Other visualizations can be created using any suitable library or tool.

*Note:* This is a very open-ended task and should make up the majority of your submission (typically, at least 2 pages of text/visualizations). Your final report (submitted towards the end of the semester) will build on and extend whatever you write here, so you will still have further time to continue working towards answering the questions you originally planned to address.

## 5 Future Plans

Briefly describe (in 1 short paragraph) what you plan to do next.