**Assignment 2**

**DS-5003/2003 Statistical and Mathematical tools for Data Science (Fall 2021)**

City Controller Office has released payroll information for all Los Angeles City employees on a quarterly basis since 2013. Data includes department titles, job titles, projected annual salaries (with breakdowns of quarterly pay), bonuses, and benefits information of employees. Government is interested in knowing answers to the questions mentioned below for future betterment of different departments. You are required to answer the following questions by using **hypothesis testing**. None of the questions can be answered as **Yes** or **No**. You are required to setup and run the experiments to reach a conclusion. Ensure that your conclusions are sound by keeping in mind the Dos and Don'ts for each concept, e.g. sample size, randomization and so on. All the intermediate steps, graphs and experimental results must be included while answering.

You are required to use the attached City Payroll Dataset. There are many empty fields, you should devise a mechanism to handle them.

**Questions:**

**Question 1.** Do the employees working as Police Officer-II have a better chance of getting Temporary Bonus Pay?

**Question 2.** Employees who get Permanent Bonus Pay are most likely to be from Public Works-Sanitation Department?

**Question 3.** Do the employees working in Water and Power (DWP) Department have a better chance of being employed overtime?

**Question 4.** In 2014, employees of Recreation and Parks Department were complaining that they have been denied the Longevity Bonus Pay. Confirm their complaint from the given dataset.

**Question 5.** Senior Clerk Typist from Harbor (Port of LA) Department has been telling Senior Clerk Typist of Water and Power (DWP) Department that they have more Average Health Cost than them. Accept or Reject this claim according to the dataset given.

**Guidelines:**

Many concepts in the asked questions are subjective and are left for you to define, i.e. “highest pay raise” and others. Note that in some cases, you may need to look at multiple columns to get the required information.

**Submission:**

You are required to submit a report containing response to each question with supporting graphs, tables or stats. A single R code file containing all the questions must be submitted separately with proper partitioning of questions. Furthermore, you have to mention your definition of the subjective terms used while answering.