Module 8 Live Coding Assignment: Experimental Design, Casual Research, and Targeting Analysis

Throughout this assignment, please remember to import all the necessary libraries at the beginning of every exercise.

Question 1

Learning about the data you are working with is essential when performing exploratory analysis. Import the data set heart.csv and assign to a dataframe "df". Use the modules "info", "describe", and "shape" to retrieve information about the database. Get the cholesterol (column "chol") distribution across the different values of "cp". Save your results in a dataframe named "df1". Print to screen the first 5 rows of "df1".

Question 2

Using the same dataframe as in the previous question, get the composition of different resting heartbeats (i.e.,"restecg") in each "cp" group. Save your results in a dataframe called "df2". Make sure your resulting database has "cp" and "resteg" as indexes and the column "sex". Print to screen the first 5 rows of "df2".

Question 3

Using the original dataframe from Question 1, find the slope distributions across all different "cp" groups. Name the the final dataframe "df3". Make sure your resulting database as "slope" as indexes, and MultiIndex column for "cp" and "sex" . Print to screen the first 5 rows of "df3".

Question 4

Using the dataframe "df", create a two-sided T-test for the null hypothesis such that two "cp" values 1 and 2 have identical average (expected) values. Do not assume equal population variance. Save your results in two variables called "t1" and "t2". **Hint:** Use the module "stats.ttest_ind" for this Welch's t-test and the "sex" variable as values. Next, repeat the same exercise as above, but assume equal variance across the population. Assign

your expected values to "te1" and "te2", respectively. Print all the results you have found to screen.

Question 5

Using the dataframe "df", create a boolean variable for whether or not a person has a threshold of beats per minute ("trestbps") lower than 125. Name this variable "low_trest" and plot your results in a histogram.

Question 6

Similarly to what you have done in the previous question, create another boolean variable "high_tach" that has value 1 if the person has a "thalach" value greater than 160. Show your results in a histogram.