Task: you're given two datasets with thousands of rows of body measurements for both men and women. Both datasets have columns for training a model: gender, age, height, and weight. train.csv file also has measurements for the three variables you'll have to predict: bust_circumference, waist_circumference, and hip_circumference. Your task is to use machine learning techniques to predict these three measurements for the second dataset (test.csv). Your output should be three columns with measurements for every row of the second dataset. You can use any tools of your choice. After you're done with that, answer the following questions.

1. Briefly describe your methodology for analyzing the data.
2. Which ML algorithm did you pick for making predictions? Why? Which ML algorithm would be a poor choice for this dataset? Why?
3. Which evaluation metric did you pick? Why? What evaluation metric would you use if this was a classification problem to predict whether a person was male or female? How would your answer change if the classes were very imbalanced?
4. The body measurement prediction problem is a regression problem. In a classification problem which would worry you more: false positives or false negatives? Why?
5. How would you deal with missing data for numerical and categorical variables?
6. Explain regularization to a layperson. What is it and why would you want to use it?
7. What's the difference between L1 and L2 regularization methods?
8. What is better: 50 small decision trees or one large decision tree? Why?

Email us the following: all of the code you used to analyze the data, a csv file with the three columns of predicted body measurements, answers to the questions above.