Data Analysis (Assignment #1)

Story:

An E-Commerce portal has collected data regarding the customer activity on the website, app and their purchase portal. The portal management wanted to study the relationship between customers' length of Memberships and yearly money amount spent by the customer buying stuff on their portal (Last two columns in the datasheet). Unfortunately, some information regarding some customers are missing e.g. Length of Membership.

A. You have three approaches that needed to be tackled:
   1) Fill the missing data using 'zeros'
   2) Fill the missing data using the mean of their column
   3) Fill the missing data using median of their column

   After trying each of them you need to study the effect of your new data proposition on the correlation between "Length of Membership" and "Yearly Amount Spent

B. Using Principal Components Analysis (PCA), Reduce the number of independent variables to only two(use any library you want for pca)

Data:

"assignment 1.csv" includes data collected by the company to be used in any further analysis

Deliverables:

1) Python code that performs each approach handling missing data and report the correlation coefficient after..
2) correlation after each approach
3) Your opinion regarding each approach and which one you recommend to the company.

Due Date: 6th of December midnight.
Delay Policy: Each day of delay cost you 2 grades out of 10 and after three days of delay you lose full assignment grade.
Submissions will be via moodle.
Submit your assignments individually.
If plagiarism is detected or suspected, tough actions will be applied.