Project revision

Instructions:

1. Use the dataNew.xls to analyze the data. Download the file from MS Team.
2. Plot the following bar charts (No. of calories versus year in ascending order) for the 3 different 10 years range:
   1. 1900 to 1910
   2. 1911 to 1920
   3. 1921 to 1930
3. Calulate the total and mean value of the calories for the 3 years ranges in step 2.
4. The sample result is shown belo fyr.
5. Create a unit testcase for the total and the mean values. You are to show 1 fail and 1 pass. Screen shot your result.

|  |
| --- |
| Chart, bar chart  Description automatically generated  Total calories = 3637.2  Mean value = 330.65 |
| Chart, bar chart  Description automatically generated  Total calories = 2782.2  Mean value = 278.22 |
| Chart, bar chart  Description automatically generated  Total calories = 2939.2  Mean value = 293.92 |

1. Create a project file
2. Install all packages…numpy, pandas, matplotlib
3. Copy your data file into project(xxx.xls or xxx.csv)

Start coding:

1. Read the file (pd.read\_csv/pd.read\_excel/.xlsx)
2. Print the dataframe (print(df)) to see rows n columns
3. Print columns heads(e.g. model, mpg, cyl……) àcars.columns
4. E.g. cars.head()à first 5 rows, cars.head(3)àfirst 3 rows
5. Split columns (step 23 in lab10). Print df after split
6. Check datatype of each column, ensure the years are numeric (convert if not)
7. Sort the data, calculate mean, sum……..
8. Filter the years and plot the bar chart
9. Create unittest case