Rules

• Internet and Books are ALLOWED

• Name your file as following: StudentID_ChineseName/EnglishName_quiz#
Ex: 123456789_安瓦_quiz1 or 123456789_Anvar_quiz1

• **Extension of your file or your file type should be .py.**
NO .ipynb, .7z, .zip and so on! PLEASE

• If MOSS (Measure of Software Similarity) detects that any two files have more than 50% similarity, both students will get 0 for this quiz

• If you submit the code within an hour you will get the full score. Otherwise, you will have 24 hours to finish it and you will get 80% of your final score.

Rules are simple. If you don't follow them you get 0 automatically.

FOLLOW AS TEMPLATE

[ ]

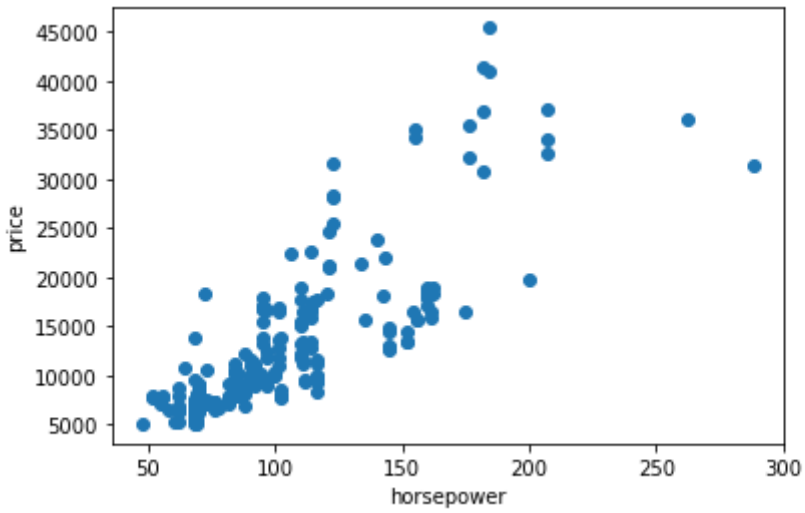
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
'Add more library modules here'


def problem1():
  #Your code here
  #You do not need to pass any arguments to the function. Just do all the printin
g and plotting in the functions
  '.....'

def problem2():

def problem3():


def problem4():
```

```
problem1()
problem2()
problem3()
problem4()
```

Problem 1

- Read the csv file with pandas module
- **Print** out the **columns** labels of the data with **pandas function**
- **Print** out the **first 10 rows for all the columns** with **pandas function**
- Plot out the **scatter plot** with horsepower as the x-axis and price as the y-axis

In this case, **horsepower** is our data predictor and **price** is our label



Problem 2

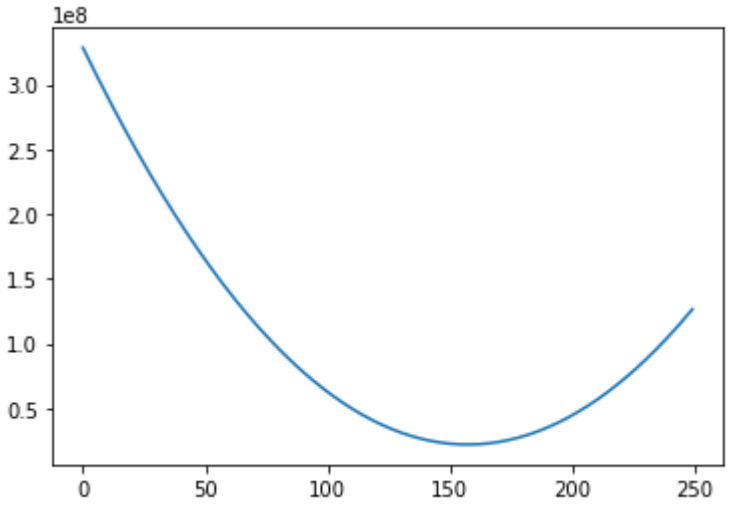
- Remember that the formula of linear regression function model is $\hat{y} = \beta_0 + \beta_1 * x$
- The formula for Mean Squared Error is $MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2$

Follow the procedures!

- Create variables β_0 and β_1
- Assign value to β_0 to -3000 and β_1 in a range of [0,250]
- Following the previous problem (**horsepower** as data, **price** as label), fit the value of the coefficients with the data to get prediction.
- Plot the MSE for all β_1

Hints:

- Create empty lists
- Set a range of values for β_1
- Calculate the prediction
- Calculate MSE for varying β_1

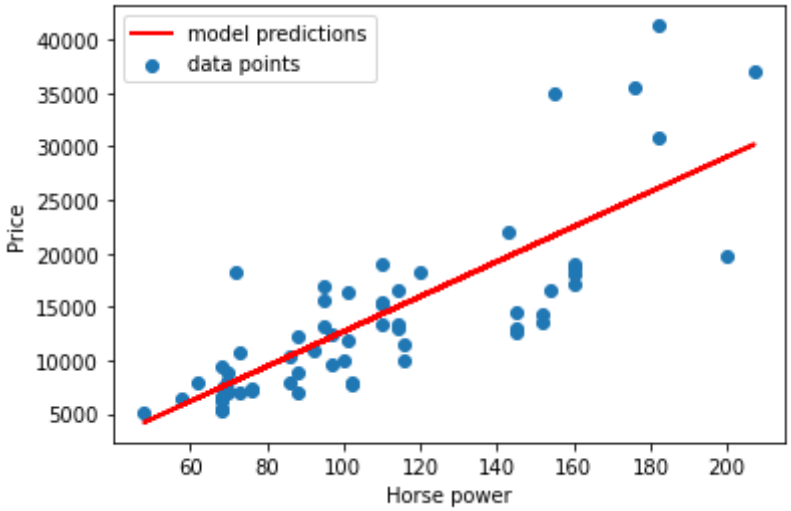


Problem 3

Use sklearn with our data from previous problems. **STRICTLY** use sklearn.

1. Use **train_test_split** function (from sklearn) to split the dataset into training and testing sets (70% training, 30% test).
   Use **horsespower** as your data and **price** as label.
2. Use the **LinearRegression** function (from sklearn) to make a model.
3. Fit (from sklearn) the model on the training set
4. Predict (from sklearn) on the testing set using the fit model.

5. Estimate the fit of the model using mean_squared_error function (from sklearn). **Print out the MSE value!**
6. Plot the dataset along with the predictions to visualize the fit. There will be 2 plots in one figure (one scatter plot of your test data and its label, another one is plot of your model prediction)



Problem 4

Problem 4 is similar to Problem 3. The only different thing is you need to set multiple features as your data to predict

---

1. Use **train_test_split** function (from sklearn) to split the dataset into training and testing sets (70% training, 30% test). Use **all the four numeric data from the csv (horsepower, peakrpm, citympg, highwaympg)** as your data and **price** as label.

   This is the only difference from problem 3.
2. Use the **LinearRegression** function (from sklearn) to make a model.
3. Fit (from sklearn) the model on the training set
4. Predict (from sklearn) on the testing set using the fit model.
5. Estimate the fit of the model using mean_squared_error function (from sklearn). **Print out the MSE value!**

!!NO PLOTTING IN THIS PROBLEM!!!

Your code would be very similar to problem 3.