Table 1: input_dataset for Question 1 & 2 [1]

BerkeleyHeights	NJ	39	5320	1	11980	3.1	1.37	0
Marpletownship	PA	45	47616	1	23123	2.82	0.8	0
Tigardcity	OR	?	?	1	29344	2.43	0.74	3
Gloversvillecity	NY	35	29443	1	16656	2.4	1.7	0
Bemidjicity	MN	7	5068	1	11245	2.76	0.53	0
Springfieldcity	MO	?	?	1	140494	2.45	2.51	7
Norwoodtown	MA	21	50250	1	28700	2.6	1.6	0
Andersoncity	IN	?	?	1	59459	2.45	14.2	8
Fargocity	ND	17	25700	1	74111	2.46	0.35	0
Wacocity	TX	?	?	1	103590	2.62	23.14	29

1. **Develop a Simple Linear Regression model in Python to predict murders (last column) while the factor is Population. 05**
2. **Develop a Multiple Linear Regression model in Python to predict murders while the factors are State, Population and household size. 05**

3. **Write a Python program to solve the following issue: 05**
   The IRS offers taxpayers the choice of allowing the IRS to compute the amount of their tax refund. During the busy filing season, the number of returns received at the Springfield Service Center which request this service follows a Poisson distribution with a mean of three per day. What is the probability that on a particular day:
   a. There are no requests?
   b. Exactly three requests appear?

4. ***Explain the following Python script:*** **05**

```
x = input_data[:,7] # independent variable
y = input_data[:,12] # dependent variable
plt.scatter(x,y) # Scatter plot for Attendance vs Final exam mark
plt.show()
dataframe1.boxplot()
input_data.boxplot(column='GRE Score');
```

[1] *Dataset description:*

-- communityname: Community name - not predictive - for information only (string)
-- state: US state (by 2 letter postal abbreviation)(nominal)
-- countyCode: numeric code for county - not predictive, and many missing values (numeric)
-- communityCode: numeric code for community - not predictive and many missing values (numeric)
-- fold: fold number for non-random 10 fold cross validation, potentially useful for debugging, paired tests - not predictive (numeric - integer)
-- population: population for community: (numeric - expected to be integer)
-- householdsize: mean people per household (numeric - decimal)
-- racepctblack: percentage of population that is african american (numeric - decimal)

-- murders: number of murders in 1995 (numeric - expected to be integer) potential GOAL attribute (to be predicted)