MA4142 End-term 2023 Time: 2:05pm-3:55pm

Please find the dataset on 20 cities. A real estate investor requires your advice on which city to invest in. Determine the best city to live in. You may choose 'Housing' variable from the data as your dependent variable. Give justification for each and every step, and interpretation for each result. Include your conclusions and recommendations. Include your graphs and codes and explanation everything in one document. [40marks]

a. Explore the data by examining basic plots using these variables and their descriptive statistics and any other EDA that you have learnt (data pre-processing).
b. Fit an appropriate regression model to the data. You may use transformation wherever necessary.
c. Explain the output.
d. Interpret the model coefficients.
e. Can you make causal statements? Why or why not?
f. Are all the variables required in the model? Find the better model? Explain and justify why.
g. Check for outliers.
h. Diagnose the model.
i. Make necessary changes wherever required.
j. Give your interpretation of the results, conclusion and business insights.

City	Income	Commute	Literacy	Job Growth	Physicians	Rape Rate	Restaurants	Housing	Median Age	Household Income
A	26,000	49.2	5.15	10.8	1987	51.3	5582	109,400	35.3	68,000
B	29,300	45.3	5.97	9.5	517	50.8	9988	97,000	43.2	70,400
C	24,800	39.8	9.41	8.2	592	77.7	20511	114,700	29.5	60,500
D	27,900	46.8	4.61	7.6	3310	51.2	8946	99,100	40.5	65,900
E	37,500	39.9	5.64	12.2	975	40.1	4000	122,200	47.1	84,700
F	31,900	49.5	4.80	7.7	2238	38.0	8970	145,300	39.3	75,800
G	25,300	44.4	6.84	5.4	611	38.8	9570	99,500	38.6	62,600
H	22,000	44.8	2.79	6.2	272	65.7	19101	76,400	41.6	54,800
I	29,400	44.9	4.48	7.8	381	48.7	12099	112,500	41.8	72,900
J	42,400	44.7	5.16	8.0	1812	45.4	10953	143,500	41.2	100,000
K	40,500	40.0	6.41	10.9	294	69.6	2655	173,600	41.7	102,000
L	24,700	38.7	1.66	9.0	196	19.0	15796	129,200	33.4	65,300
M	24,400	41.1	5.60	8.7	404	77.2	16001	126,500	30.6	62,200
N	22,400	42.8	2.16	8.3	534	57.9	16712	102,700	34.5	59,200
O	22,200	37.8	2.72	8.4	166	50.9	11856	110,300	35.4	57,100
P	27,500	48.4	4.03	8.1	1553	83.6	12348	107,400	34.3	72,000
Q	23,100	44.5	2.07	4.7	502	42.7	65804	116,000	38.5	59,400
R	25,000	41.4	3.61	13.9	172	17.8	36151	120,000	52.7	57,300
S	25,800	53.5	5.03	5.3	4143	57.4	14310	132,800	36.2	71,900
T	22,600	45.0	5.29	6.5	526	52.2	8878	86,500	41.5	54,000