Topic: Hierarchical Clustering

1.) Perform clustering (Both hierarchical and K means clustering) for the airlines data to obtain optimum number of clusters.

Draw the inferences from the clusters obtained.

	ID.	Balance	Qual_miles	cc1_miles	cc2_miles	cc3_miles	Bonus_miles	Bonus_trans	Flight_miles_12mo	Flight_trans_12	Days_since_enroll	Award.
1	1	28143	0	1	1	1	174	1	0	0	7000	0
2	2	19244	0	1	1	1	215	2	0	0	6968	0
3	3	41354	0	1	1	1	4123	4	0	0	7034	0
4	4	14776	0	1	1	1	500	1	0	0	6952	0
5	5	97752	0	4	1	1	43300	26	2077	4	6935	1
6	6	16420	0	1	1	1	0	0	0	0	6942	0
7	7	84914	0	3	1	1	27482	25	0	0	6994	0
8	8	20856	0	1	1	1	5250	4	250	1	6938	1
9	9	443003	0	3	2	1	1753	43	3850	12	6948	1
10	10	104860	0	3	1	1	28426	28	1150	3	6931	1
11	11	40091	0	2	1	1	7278	10	0	0	6959	0
12	12	96522	0	5	1	1	61105	19	0	0	6924	1
13	13	43382	0	2	1	1	11150	20	0	0	6924	0
14	14	43097	0	1	1	1	3258	6	0	0	6918	0
15	15	17648	0	1	1	1	0	0	0	0	6912	0

2.) Perform Clustering for the crime data and identify the number of clusters formed and draw inferences.

	X	Murder	Assault	UrbanPop	Rape
1	Alabama	13.2	236	58	21.2
2	Alaska	10.0	263	48	44.5
3	Arizona	8.1	294	80	31.0
4	Arkansas	8.8	190	50	19.5
5	California	9.0	276	91	40.6
6	Colorado	7.9	204	78	38.7
7	Connecticut	3.3	110	77	11.1
8	Delaware	5.9	238	72	15.8
9	Florida	15.4	335	80	31.9
10	Georgia	17.4	211	60	25.8
11	Hawaii	5.3	46	83	20.2
12	Idaho	2.6	120	54	14.2
13	Illinois	10.4	249	83	24.0

Hints:

1. Business Problem

    1.1. Objective
    1.2. Constraints (if any)

2. Data Pre-processing

    2.1 Data cleaning, Feature Engineering, EDA etc.

3. Model Building

    3.1 Partition the dataset
    3.2 Model(s) - Reasons to choose any algorithm
    3.3 Model(s) Improvement steps
    3.4 Model Evaluation
    3.5 Python and R codes

4. Deployment

    4.1 Deploy solutions using R shiny and Python Flask.

5. Result Share the benefits/impact of the solution - how or in what way the business (client)

    gets benefit from the solution provided.

Note:

1. For each assignment the solution should be submitted in the format

2. Research and Perform all possible steps for improving the model(s) accuracy

    Ex: Feature Engineering, Hyper Parameter tuning, etc.

3. All the codes (executable programs) are running without errors

4. Documentation of the module should be submitted along with R & Python codes, elaborating on every step mentioned here