**Topic: Network Analytics**

**1) Use the flights\_hault and connecting routes datasets to find betweenness, closeness centrality, page rank, directed and undirected graphs.**

**Dataset: Flights\_hault**

**column names:**

**ID","Name","City","Country","IATA\_FAA","ICAO","Latitude","Longitude","Altitude","Time","DST","Tz database time"**

**Dataset: Connecting routes**

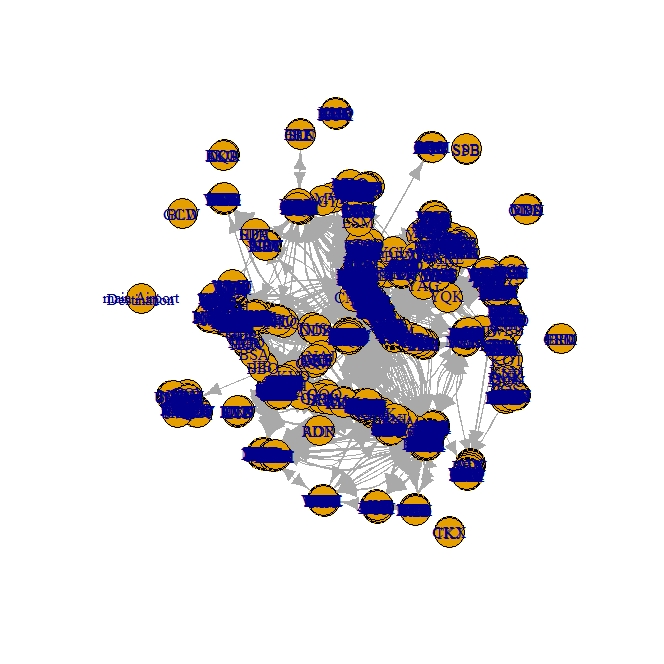
**column names:**

**"flights", " ID", "main Airport","main Airport ID", "Destination ", "Destination ID", "haults", "machinary"**

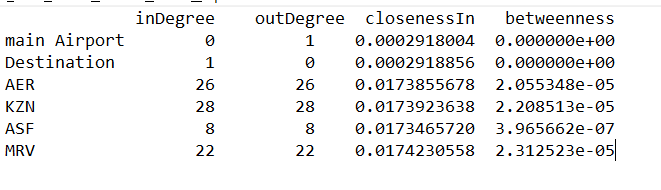
**Ans:**

* Loading the dataset of flights\_haul
* As there are no proper column names, the new columns are given to the dataset
* Loading the dataset of flights connecting routes
* In this dataset as well, there are no proper column names, the new columns are given to the dataset

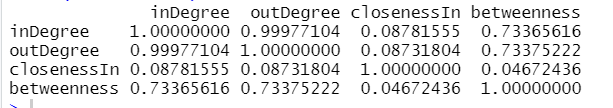
**Graphical Visualization:**

****

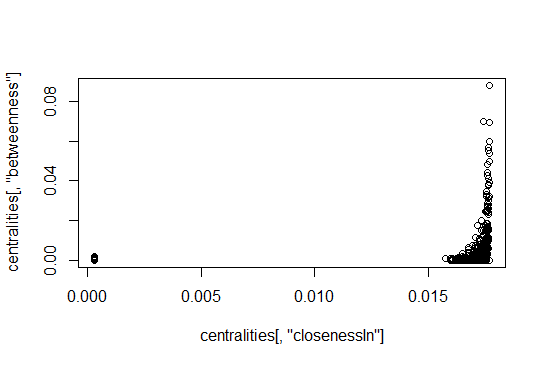
* Using the function **vcount** we can see how many airports there is the network, here it shows that **3427** airports are available in this network.
* With the help of **ecount** function, we can come to know that there are 67664 connections in the network.
* With the help of **degree** function, we can clearly say that Atlanta airport has most flight incoming with the count of 911 flights and the given airport id is 3584.
* With the help of **degree** function, we can clearly say that Atlanta airport has most flight out going with the count of 915 flights and the given airport id is 3584.
* With the help of closeness function, we can find out which of the airport is close to most of the airport, here FRA is the closet airport with closeness value of 0.017 with the airport code 338.
* Using betweenness function, we find that LAX as the most common routes and hence LAX is the international hub for all the points in center
* Taking the degree, closeness and betweenness centralities below are the main airports



* Below are the correlation of the centralities

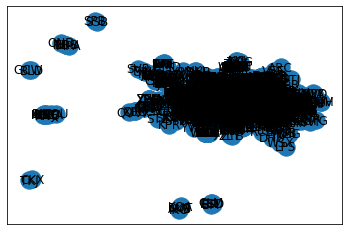


**Pair Plot for closeness and betweenness:**

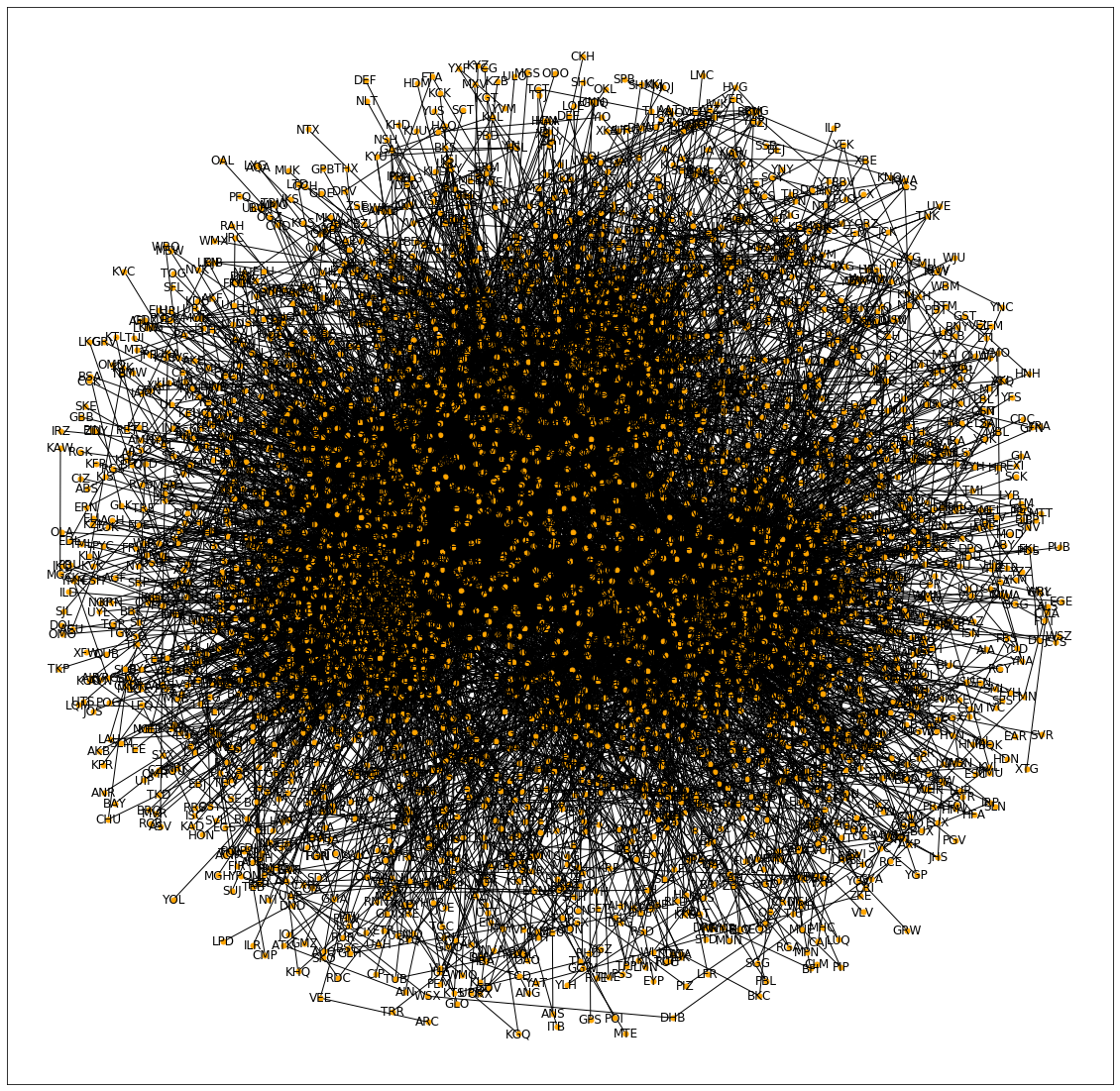
****

* Using eigen centrality function, we say that Atlanta is the most important airport with eigen value 0.27
* Using page rank function, we say that Atlanta is ranked as the best airport in google page rank system

**Eigen-Vector Centrality:**

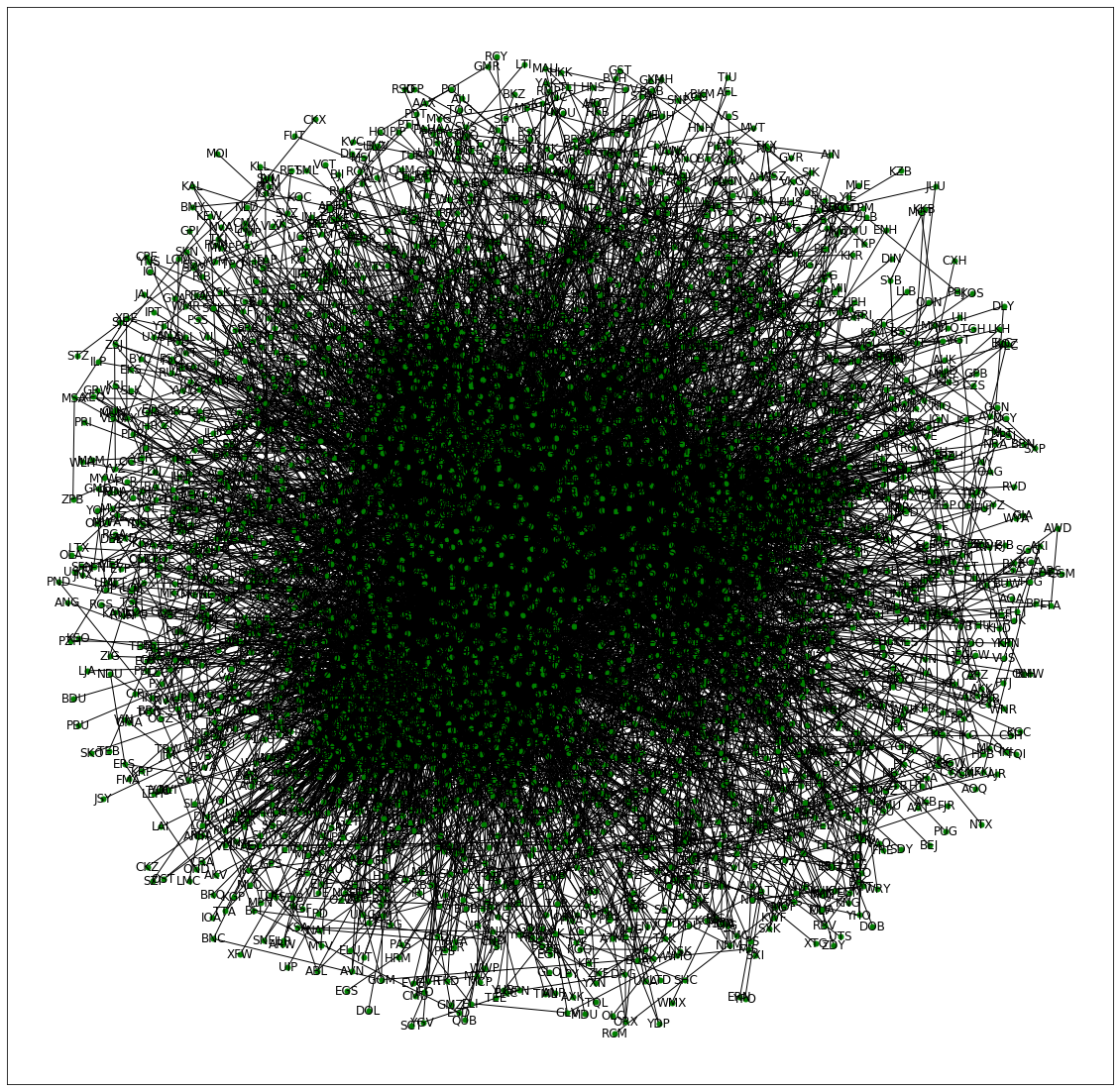
****

**Betweenness Centrality:**

****

**Closeness Centrality:**

* Number of nodes: 3425
* Number of edges: 19257
* Average degree: 11.2450



**2) Business Problem: Three data sets of social media networks are attached.**

**- plot circular graph**

**- Star Graph for the three data sets**

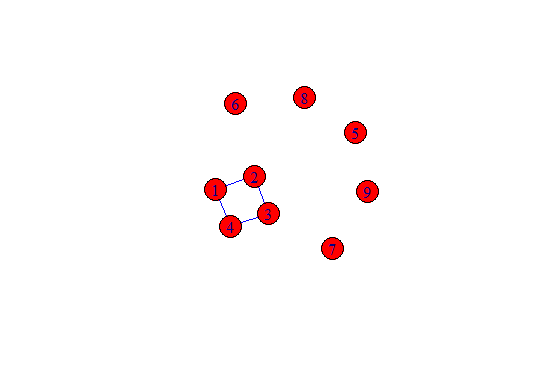
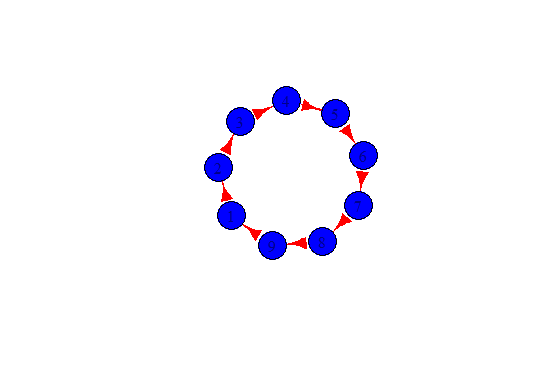
**- Check for its nodes and directed and undirected graphs**

**Draw your insights on the data.**

**Ans:**

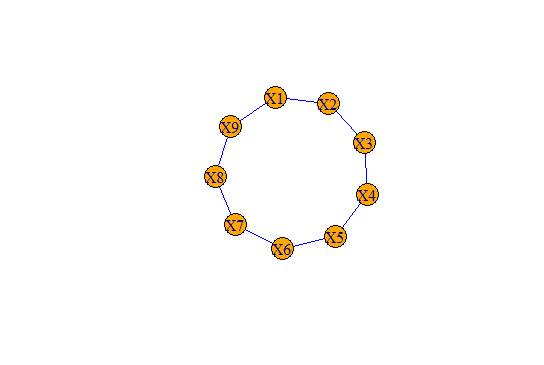
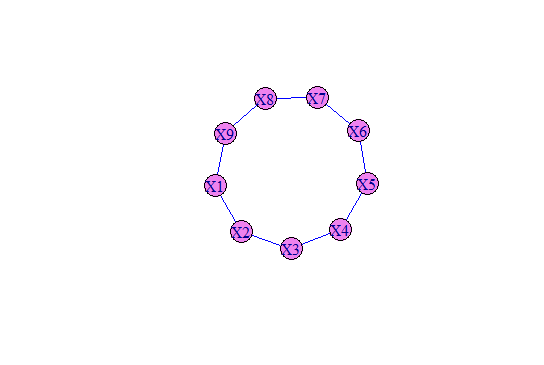
**Graphical Representation of Facebook**

**Edge Graphs:**

****

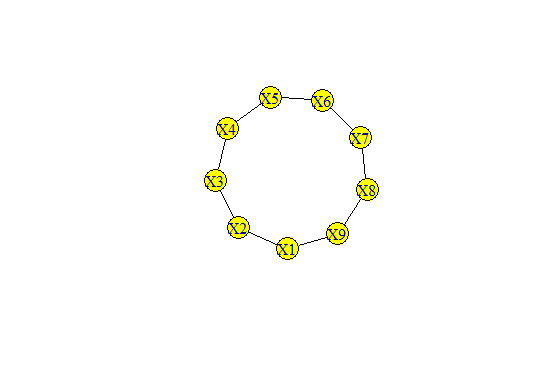
* In the above both the graphs, the set of vertices “V” = {1,2,3,4,5,6,7,8,9} and the set of edges “E” = { 12, 23, 34, 41} in diagram 2 which are undirected but whereas, in diagram1 we can see they are all directed to each other.

**Star Graph:**

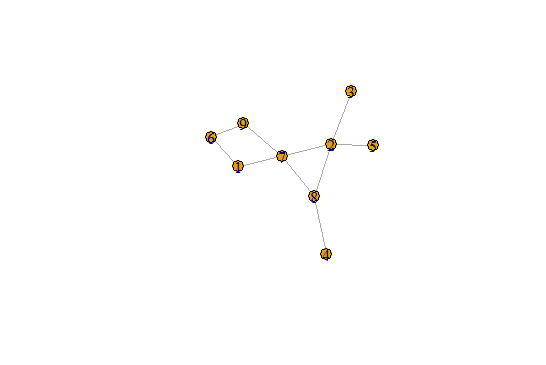
****

**Circular Graph:**

* In the above graph, we have 9 nodes which are undirected graph i.e., the nodes are connected but they do not have any direction.

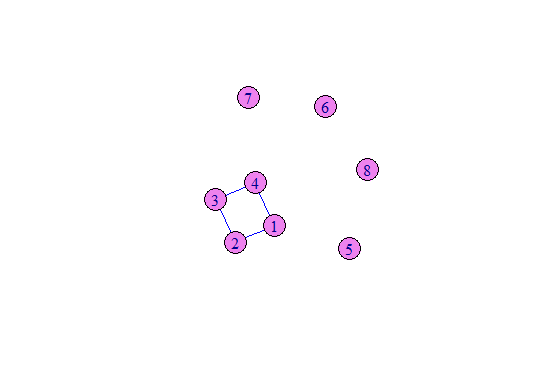
****

**GNP Graph:**

****

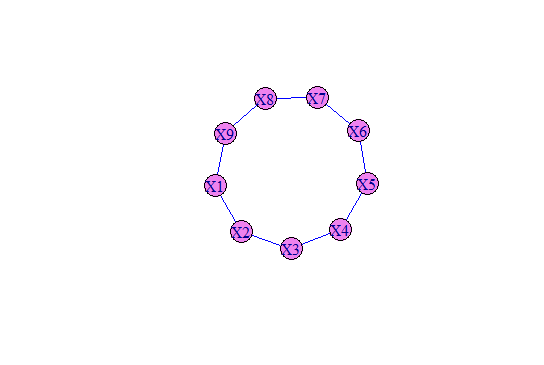
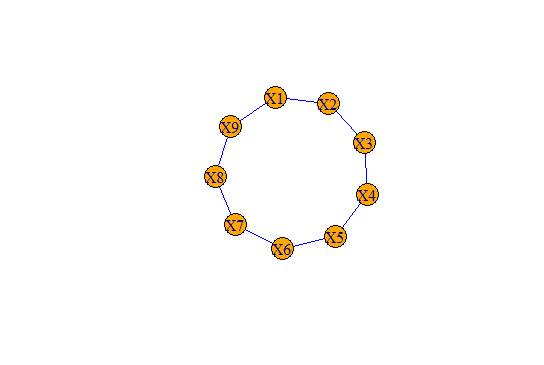
**Graphical Representation of Instagram:**

**Edge Graphs:**

****

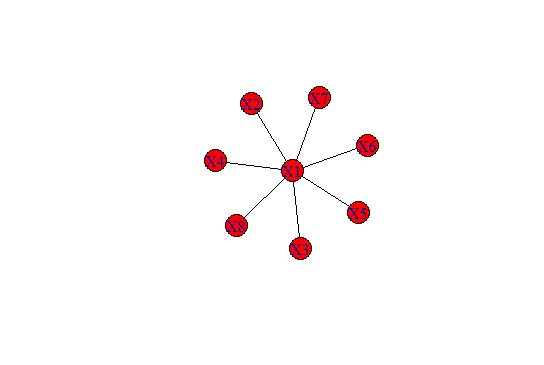
* In the above both the graphs, the set of vertices “V” = {1,2,3,4,5,6,7,8,9} and the set of edges “E” = { 12, 23, 34, 41} in diagram 2 which are undirected but whereas, in diagram1 we can see they are all directed to each other.

**Star Graph:**



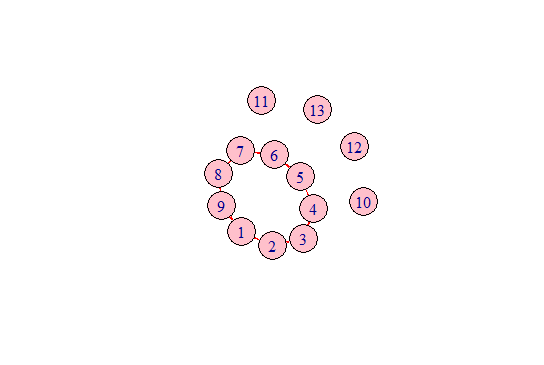
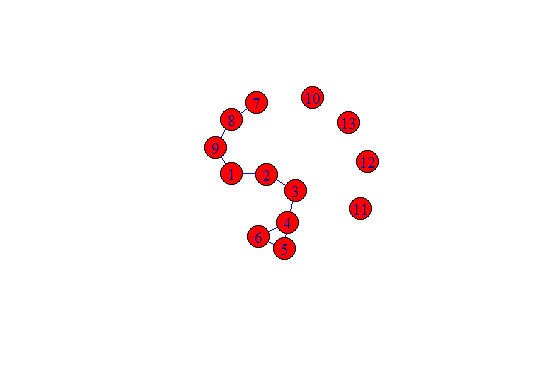
**Circular Graph:**

* In the above graph, we have 7 nodes which are undirected graph i.e., the nodes are connected but they do not have any direction.

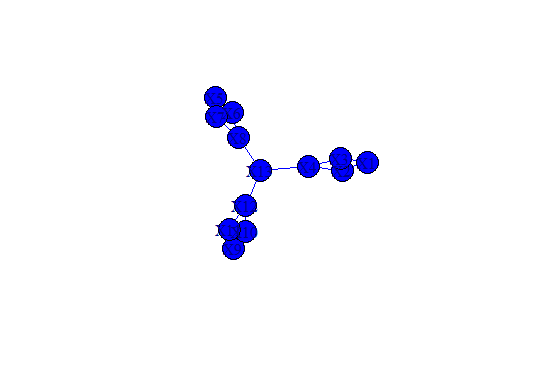
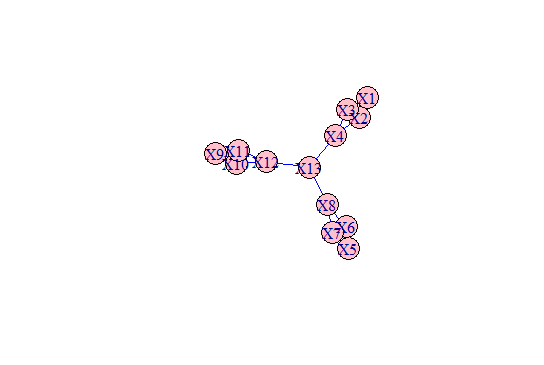


**Graphical Representation of Linkedin:**

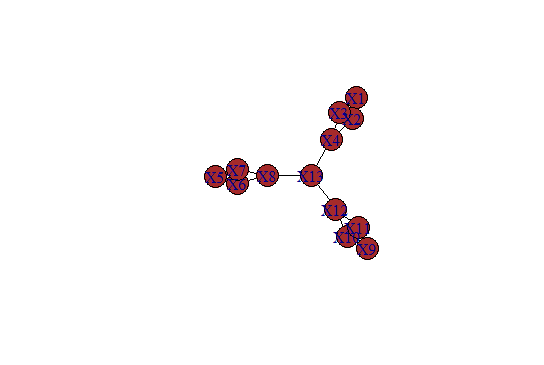
**Edge Graphs:**



**Star Graphs:**

**Circular Graphs:**

****

* In the above graph, Node 13 is connected to node 12 , node 4 & node 8 and a group of 4 nodes are connected ( 1, 2, 3, 4) , (5, 6, 7, 8) , (9, 10, 11, 12) and these group are not having directions respectively. However, following each other is seen in few patterns like ( X2 & X3 followed by X1 & X2 ) ,

( X6 & X7 followed by X5 & X8) , ( X10 & X11 followed by X9 & X12 ).

**Instructions:**

**1. Business Problem**

* 1. **Objective**
  2. **Constraints (if any)**

**Using R and Python codes perform:**

**2. Data Pre-processing**

**2.1 Data cleaning, Feature Engineering etc.**

**3. Exploratory Data Analysis (EDA):**

**3.1. Summary**

**3.2. Univariate analysis**

**3.3. Bivariate analysis**

**4. Model Building**

**4.1 Build the Network model on the given data sets.**

**4.2 Perform the Network Analytics.**

**4.3 Model(s) Improvement (Try with different no. of clusters)**

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

**6. problem statement information and data dictionary is given in next page.**

**Note:**

**The assignment should be submitted in the following format:**

* **R code**
* **Python code**
* **Code Modularization should be maintained**
* **Documentation of the modules (elaborating on steps mentioned above).**