**K - Means Clustering**

**Instructions:**

Please share your answers filled in-line in the word document. Submit code separately wherever applicable.

Please ensure you update all the details:

**Name: \_\_\_\_\_\_\_\_\_\_\_\_\_ Batch ID:** \_\_\_\_\_\_\_\_\_\_\_

**Topic: K Means Clustering**

**Problem Statements:**

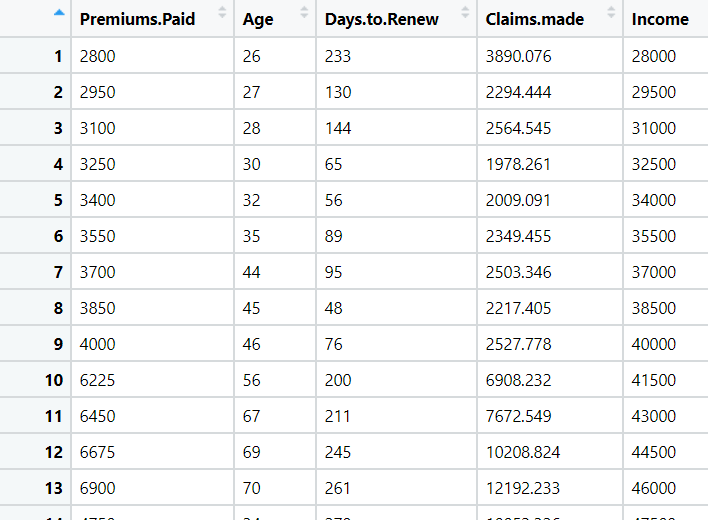
1. Perform K means clustering on the airlines dataset to obtain optimum number of clusters. Draw the inferences from the clusters obtained. Refer to EastWestAirlines.xlsx dataset.



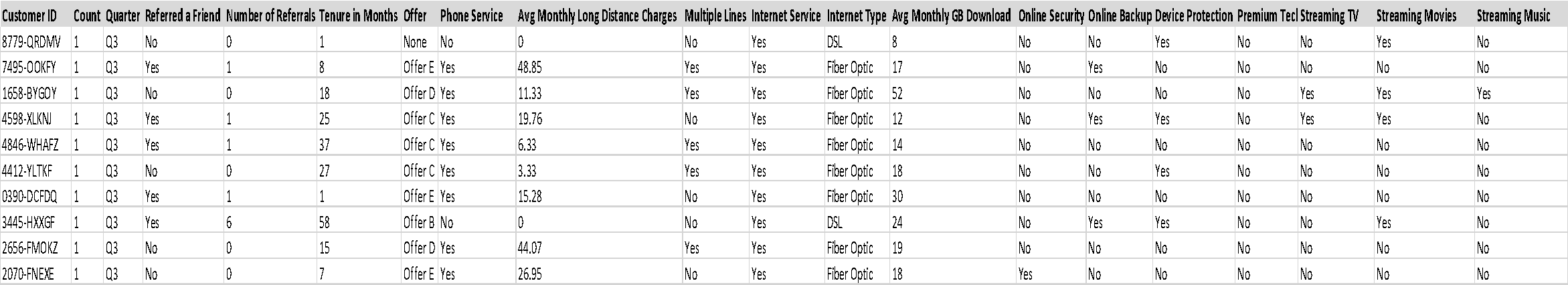
1. Perform clustering for the crime data and identify the number of clusters formed and draw inferences. Refer to crime\_data.csv dataset.



1. Analyze the information given in the following ‘Insurance Policy dataset’ to create clusters of persons falling in the same type. Refer to Insurance Dataset.csv

****

1. Perform clustering analysis on the telecom dataset. The data is a mixture of both categorical and numerical data. It consists of the number of customers who churn. Derive insights and get possible information on factors that may affect the churn decision. Refer to Telco\_customer\_churn.xlsx dataset.



1. Perform clustering on mixed data. Convert the categorical variables to numeric by using dummies or label encoding and perform normalization techniques. The dataset has the details of customers related to their auto insurance. Refer to Autoinsurance.csv dataset.



**Hints:**

**1. Business Problem**

* 1. **What is the business objective?**
  2. **Are there any constraints?**

**2. Work on each feature of the dataset to create a data dictionary as displayed in the below image:**



**3. Data Pre-processing**

**3.1 Data Cleaning, Feature Engineering, etc.**

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

**4.1. Summary.**

**4.2. Univariate analysis.**

**4.3. Bivariate analysis.**

**5. Model Building**

**5.1 Build the model on the scaled data (try multiple options).**

**5.2 Perform K- means clustering and obtain optimum number of clusters using scree plot.**

**5.3 Validate the clusters (try with different number of clusters) – label the clusters and derive insights (compare the results from multiple approaches).**

**6. Write about the benefits/impact of the solution - in what way does the business (client) benefit from the solution provided?**