EastWestAirlines.xlsx

problem statement 1.

Perform clustering for the airlines data to obtain optimum number of clusters. Draw the inferences from the clusters obtained. Refer to EastWestAirlines.xlsx dataset

Inference:-

The data which we used in our analysis comprises of 12 variables including continuous and categorical variables while there were almost 5,000  observers or respondents from whom the data has been collected and then the final analysis was performed.

**1. Business Problem**

**What is the business objective?**

**Minimize:-**

**Maximizes :-**

**Are there any constraints?:-**

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

|  |  |  |  |
| --- | --- | --- | --- |
| Name of feature | Description | Type | Relevance |
| ID# | here the id is unique | Ordinal data | Irrelevant |
| Balance | Number of miles eligible for award travel | Discrete data | relevant |
| Qual\_miles | Number of miles counted as qualifying for Topflight status | Discrete data | Relevant |
| cc1\_miles | Has member earned miles with airline freq. flyer credit card in the past 12 months | Discrete data | Relevant |
| cc2\_miles | Has member earned miles with Rewards credit card in the past 12 months (1=Yes/0=No)? | Discrete data | Relevant |
| cc3\_miles | Has member earned miles with Small Business credit card in the past 12 months (1=Yes/0=No)? | Discrete data | Relevant |
| Bonus\_miles | Number of miles earned from non-flight bonus transactions in the past 12 months | Discrete data | Relevant |
| Bonus\_trans | Number of non-flight bonus transactions in the past 12 months | Discrete data | Relevant |
| Flight\_miles\_12mo | Number of flight miles in the past 12 months | Discrete data | Relevant |
| Flight\_trans\_12 | Number of flight transactions in the past 12 months | Discrete data | Relevant |
| Days\_since\_enroll | Number of days since Enroll\_date | Discrete data | Relevant |
| Award? | Dummy variable for Last\_award (1=not null, 0=null) | Ordinal | Irrelevant |

**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 the hierarchical clustering and visualize the clusters using dendrogram.**

**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?**