**ROLL NO** :21L-5653

**Hierarchy**

Concept Hierarchy is used to reduce the complexity of a dataset.

Like in the dataset after Concept Hierarchy female =1 and male=0.

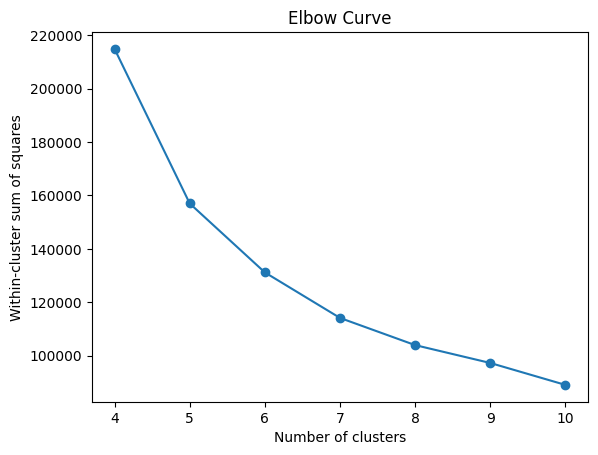
**Use of KBinsDiscretizer**

I use KBinsDiscretizer reducing the impact of outliers, simplifying model complexity.

**Use of LDA**

LDA is used to reduce the dimensionality of the dataset while retaining information relevant for predicting diabetes. It also helps to improve the accuracy and efficiency of machine learning models that use the transformed data.

**No of clusters per sum of square**

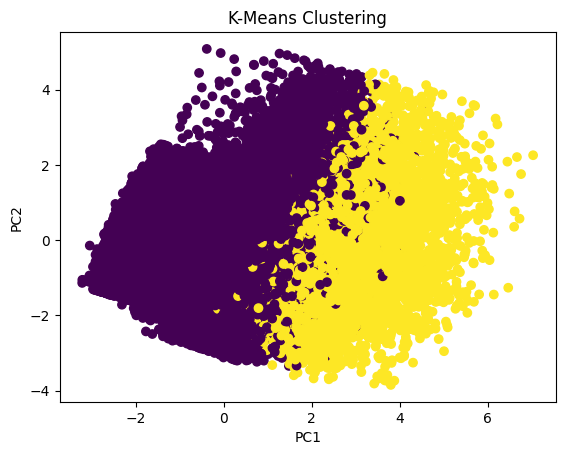
****

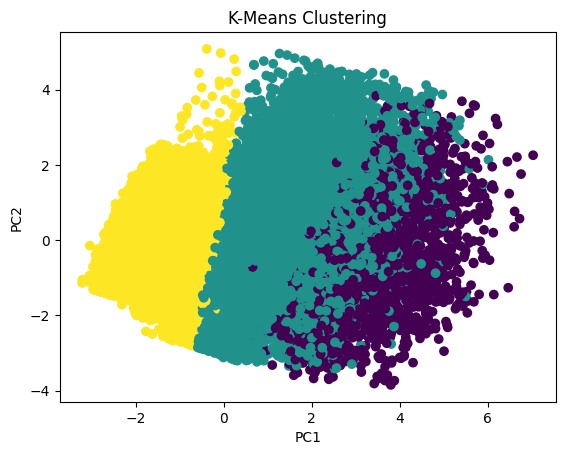
**Clusters=2**

**attribute =7**

**Arributes**

'age', 'hypertension', 'heart\_disease', 'smoking\_history', 'bmi', 'HbA1c\_level', 'blood\_glucose\_level'

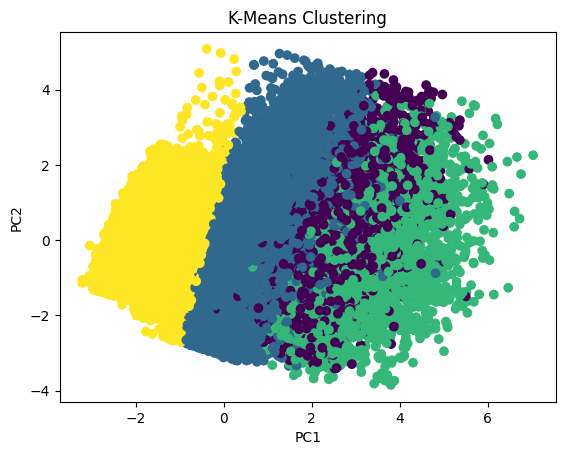




**Clusters=4**

**attribute =7**

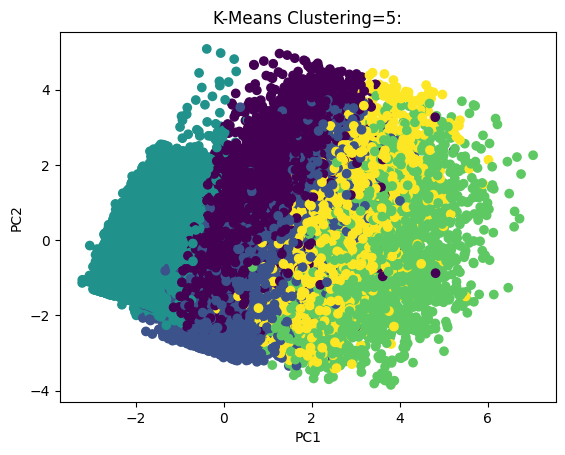
'age', 'hypertension', 'heart\_disease', 'smoking\_history', 'bmi', 'HbA1c\_level', 'blood\_glucose\_level'



**Clusters=5**

**attribute =7**

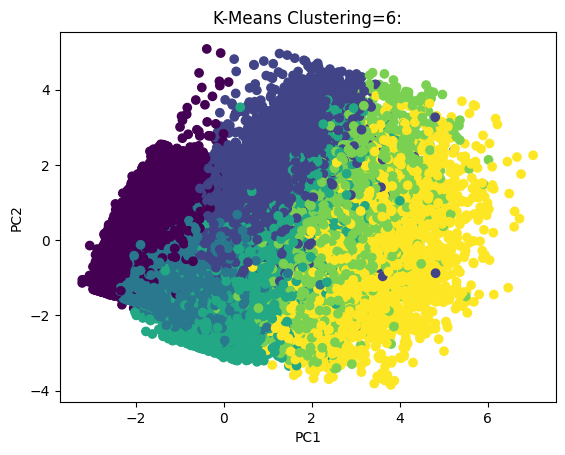
'age', 'hypertension', 'heart\_disease', 'smoking\_history', 'bmi', 'HbA1c\_level', 'blood\_glucose\_level'



**Clusters=6**

**attribute =7**

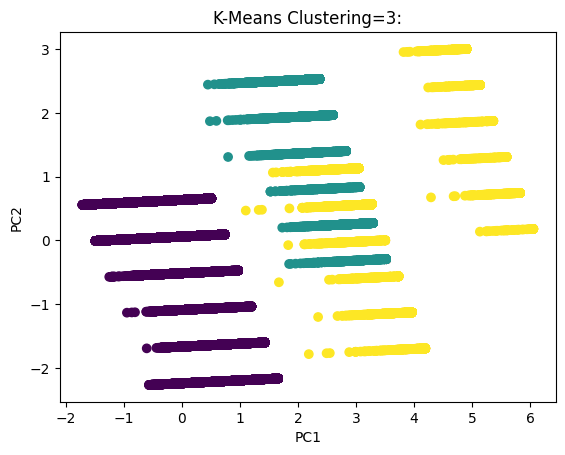
'age', 'hypertension', 'heart\_disease', 'smoking\_history', 'bmi', 'HbA1c\_level', 'blood\_glucose\_level'



**Clusters=3**

**attribute =4**

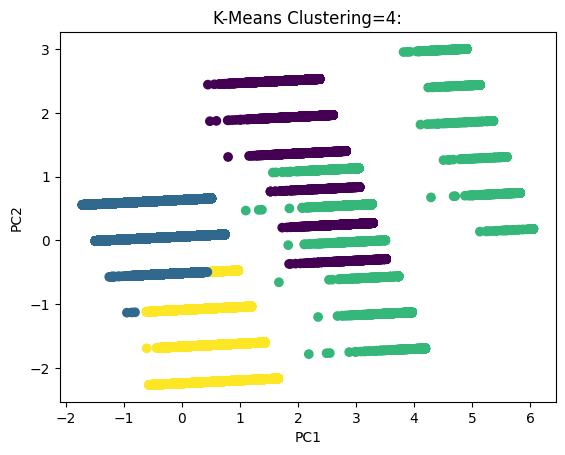
'age', 'hypertension', 'heart\_disease', 'smoking\_history'



**Clusters=4**

**attribute =4**

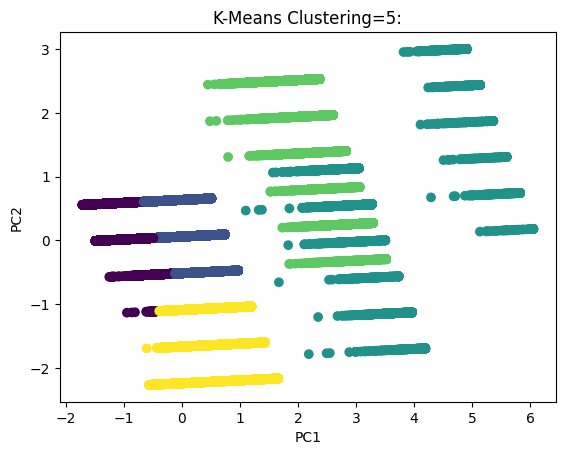
'age', 'hypertension', 'heart\_disease', 'smoking\_history'



**Clusters=5**

**attribute =4**

'age', 'hypertension', 'heart\_disease', 'smoking\_history'



**Clusters=6**

**attribute =4**

'age', 'hypertension', 'heart\_disease', 'smoking\_history'

