

E-COMMERCE CUSTOMER SEGMENTATION

CONTENTS:

CONTENTS:

- *Importing Essential Libraries.*
- *Loading the Dataset.*
- *Understanding the Data.*
- *Handling Duplicate & Missing Values.*
- *Exploratory Data Analysis (EDA).*
- *New Dataset & New Column.*
- *Data Pre-Processing for Clustering.*
- *Optimal K Value.*
- *K-Means Clustering.*
- *Analysing Clusters.*
- *Conclusion.*

Importing Essential Libraries:

- *numpy & pandas for data manipulation.*
- *matplotlib.pyplot & seaborn for visualizing the data.*
- *sklearn libraries for clustering (KMeans) & evaluation (silhouette_score).*
- *MinMaxScaler for normalizing data.*
- *yellowbrick.cluster.KElbowVisualizer to determine the optimal number of clusters.*

Loading the Dataset:

~~>Loading the Dataset:~~

- *We are reading a CSV file named cust_data.csv into a Pandas DataFrame.*

Understanding the Data: Quasi-reading the Data

- *info(): Shows DataTypes & Non-Null values.*
- *shape: Prints the number of Rows & Columns.*
- *describe(): Provides descriptive statistics.*

Handling Duplicate and Missing Values:

- *Check for duplicate values.*
- *Check for missing values.*
- *Filling the missing values in the Gender column using Mode (most frequently occurred value).*
- *Checking the missing values again.*

Exploratory Data Analysis (EDA):

- *Generated count of Male & Female customers.*
- *Plotted the count of Male & Female customers.*
- *Showed the comparison of Total Orders & Gender-wise distribution of orders.*
- *Generated a heatmap to visualize correlations between numerical columns.*
- *Plotted histograms for all numerical columns.*

New Dataset and New Column:

Новый Датасет и Новый Колонн:

- *Created a new dataset named “df1”.*
- *Created a new column named “Total_Search” by summing values from the third column onward.*
- *Displayed the new dataset in descending order using “Total_Search” column.*
- *Displayed the top 10 customers based on total searches using Barplot.*

Data Pre-processing for Clustering:

Data Pre-processing for Clustering:

- *Extracted the numerical data (excluding first two columns).*
- *Applied “MinMaxScaler” to scale features between 0 & 1.*

Optimal K Value: Obtained via K-Means:

- Computed the inertia (sum of squared distances) for K values from 1 to 15 using Elbow Method.
- Plotted the Elbow Method graph.
- Used KElbow Visualizer to determine the optimal K.
- Performed Silhouette analysis to validate the number of clusters.

K-Means Clustering:

K-Μέσων Κλασσεύσιμων:

- *Trained the K-Means model with $K = 3$ & assigned cluster labels to customers.*
- *Saved clustered data to “Cluster_data.csv”*
- *Plotted the number of customers in each cluster.*

Analysing Clusters:

- *Read clustered data & calculated total searches.*
- *Analysed gender distribution & total searches for Cluster 0.*
- *Analysed gender distribution & total searches for Cluster 1.*
- *Analysed gender distribution & total searches for Cluster 2.*
- *Created a final dataframe grouping all Clusters.*
- *Plotted Countplot to show total customers on each Cluster.*
- *Plotted Barplot to show number of times customers searched the products & their past orders.*

Conclusion:

CONCLUSION

- *Cluster 0: Moderate searches & regular customers with high previous orders.*
- *Cluster 1: Low searches & needs engagement strategies with moderate previous orders.*
- *Cluster 2: High search activity & potential high-value customers with very few previous orders.*

THANK YOU