CUSTOMER SEGMENTATION USING CLUSTERING

** Project Overview**

I developed a machine learning model to segment customers based on their behavior and demographics using the Mall Customer Segmentation dataset. This project involved various stages including data cleaning, feature engineering, exploratory data analysis, model selection, evaluation, and deployment using Python and key data science libraries. The goal was to build an accurate and efficient clustering model to provide insights into the different customer segments for targeted marketing strategies.

****Problem Statement****

The objective of this project was to segment mall customers into distinct groups based on their demographic and behavioral attributes. By identifying different customer segments, businesses can tailor their marketing efforts to better meet the needs of each group.

****Tools & Technologies****

- **Python**: The primary programming language used.
- **Pandas & NumPy**: For data manipulation and analysis.
- **Scikit-Learn**: For building and evaluating clustering models.
- **Matplotlib & Seaborn**: For data visualization and exploratory data analysis.
- **Flask**: For deploying the model as an interactive web application.

****Data Preparation****

- **Data Cleaning**: Checked for missing values and handled any inconsistencies in the dataset.
- **Feature Selection**: Selected relevant features such as age, annual income, and spending score for clustering analysis.
- **Feature Scaling**: Standardized the features to ensure they contribute equally to the clustering process.

****Exploratory Data Analysis (EDA)****

- Visualized distributions and relationships between features using histograms, scatter plots, and box plots.

- Used heatmaps to identify correlations between features and understand their impact on clustering.

Model Development

- **Algorithm Selection**: Evaluated multiple clustering algorithms including:
  - K-Means - Hierarchical Clustering - DBSCAN
- **Optimal Number of Clusters**: Used the Elbow method and Silhouette scores to determine the optimal number of clusters.
- **Model Training**: Trained the selected model (K-Means) with the optimal number of clusters.
- **Evaluation Metrics**: Assessed model performance using metrics such as inertia and silhouette score to ensure robust clustering.

Results

- **Optimal Clustering**: Identified 5 distinct customer segments based on the Elbow method and Silhouette score.
- **Key Insights**:
  - **Cluster 1**: High income and high spending score – targeted for premium marketing.
  - **Cluster 2**: Low income and low spending score – potential for budget marketing.
  - **Cluster 3**: Moderate income and moderate spending score – general marketing.
  - **Cluster 4**: High income and low spending score – potential for upselling.
  - **Cluster 5**: Low income and high spending score – price-sensitive marketing.

Key Learnings

- **Data Preprocessing**: The importance of scaling and normalizing data to improve clustering performance.
- **Model Evaluation**: Understanding the significance of metrics like inertia and silhouette score in evaluating clustering models.
- **Deployment**: Gaining practical experience in deploying a clustering model and creating an interactive user interface for real-world applications.

GitHub Repository

Explore the complete project and code here: https://github.com/sravanthi224/Quanta-5.git