# CUSTOMER SEGMENTATION USING DATA SCIENCE

**Phase** 2


## TEAM MEMBERES

## TEAM LEADER : R.RAHUL

R.R.SANJAY

V.PERMAL

R.NITHISHKUMAR

S.S.SHAKTHI


Customer segmentation is a critical aspect of any business strategy. It involves dividing a company's customer base into distinct groups with similar characteristics, behaviors, or needs. This allows businesses to tailor their marketing efforts, product offerings, and customer service to better meet the specific needs of each segment. Here's a step-by-step guide to creating a project on customer segmentation using data science:

**1. **Define Objectives and Scope**:

- Clearly state the purpose of your project. Are you aiming to increase sales, improve customer satisfaction, or something else?
- Define the scope: Which data sources will you use? What types of

segmentation will you perform?

**2. **Data Collection and Preprocessing**:

- Collect relevant data: This might include transaction history, demographic information, website behavior, customer feedback, etc.
- Clean and preprocess the data: Handle missing values, outliers, and perform data transformations (e.g., normalization, encoding categorical variables)

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

- Perform initial data exploration to understand the characteristics and distributions of the variables.
- Identify any trends, patterns, or anomalies in the data.

**4. Feature Engineering**:

- Create new features that could be useful for segmentation. For example, you might calculate metrics like recency, frequency, and monetary value (RFM) for each customer.
- Select the most relevant features for segmentation.

**5. Choose Segmentation Methods**:

- Decide on the segmentation techniques you'll use. Common methods include:
- **RFM Analysis**: Dividing customers based on recency, frequency, and monetary value of their purchases.
- **Clustering Algorithms (e.g., K-means, DBSCAN)**: Automatically grouping customers based on similarity in their features.
- **Decision Trees or Random Forests**: If you have labeled data with

clear segments, these algorithms can be used for supervised segmentation.

**6. Apply Segmentation Techniques**:

- Implement the chosen methods and segment the customer base.

**7. Evaluate Segmentation Results**:

- Use metrics like Silhouette Score, Davies-Bouldin Index (for clustering), or domain-specific metrics (e.g., customer lifetime value) to evaluate the quality of your segmentation.

**8. Profile Each Segment**:

- For each segment, create a detailed profile including demographic information, buying behavior, preferences, etc.**9. Interpretation and Actionable Insights**:

- Analyze the segments and draw actionable insights. What marketing strategies, product offerings, or communication channels are most effective for each segment?

**10. Visualizations and Reporting**:

- Create visualizations (e.g., bar charts, pie charts, heatmaps) to present your findings. Use tools like Matplotlib, Seaborn, or Tableau for visualization.
- Prepare a comprehensive report summarizing the project, methodology, findings, and recommendations.

**11. Deployment (Optional)**:

- If applicable, deploy the segmentation model in a business environment where it can be used for real-time customer segmentation.

**12. Monitoring and Iteration**:

• Continuously monitor the effectiveness of the segmentation strategy. Make adjustments as needed based on changing customer behavior or business goals.

Remember, communication is key in any data science project. Clearly articulate your findings and recommendations to stakeholders in a way that they can easily understand and act upon.

## INOVATING THE APPROACH:

**Data Collection and Preprocessing:**

- Gather relevant customer data, which might include demographics, purchase history, online behavior, and more.
- Preprocess the data by handling missing values, encoding categorical features, and normalizing or standardizing numerical features

**Feature Selection/Engineering:**

- Determine which features are most relevant to the segmentation task. Feature engineering may involve creating new variables or transforming existing ones to capture valuable insights.

**Select a Clustering Algorithm:**

- Clustering algorithms group similar customers together. Common algorithms include K-Means, Hierarchical Clustering, DBSCAN, and Gaussian Mixture Models (GMM).
- Choose an algorithm that suits your data and business objectives.

**Hyperparameter Tuning:**

- Hyperparameters are settings or configurations of the clustering algorithm that affect its performance. Hyperparameter tuning aims to find the best combination for your specific dataset.
- Consider using techniques such as grid search, random search, or Bayesian optimization to find optimal hyperparameters. Libraries like scikit-learn offer tools for hyperparameter tuning.

**Validation and Evaluation:**

- Split your data into training and testing sets to evaluate the model's performance.
- Use appropriate metrics (e.g., Silhouette Score, Davies-Bouldin Index, or within-cluster sum of squares) to assess the quality of the segmentation.

**Iterative Process:**

- Hyperparameter tuning might require several iterations. Experiment with different combinations to find the best model.
- Be cautious about overfitting, and consider using cross-validation to estimate the model's generalization performance.

**Interpretation and Visualization:**

- Once you have an optimized clustering model, visualize the customer segments to gain insights.
- Explore the characteristics of each segment to understand the differences and similarities among customers

**Implementation and Action:**

- Translate your insights into actionable strategies. This might include targeted marketing campaigns, product recommendations, or personalized customer experiences.

**Monitoring and Maintenance:**

- Customer behavior can change over time, so periodically re-run the segmentation process to adapt to evolving customer preferences and trends.

**Experiment and Innovation:**

- Continue to innovate by experimenting with different clustering algorithms and hyperparameters to improve the accuracy and relevance of customer segments.
- Explore other techniques, such as dimensionality reduction (e.g., PCA) or using advanced machine learning models like deep learning for segmentation.

## USING PAC:

**Data Collection and Preprocessing:**

- Collect and preprocess your customer data as usual. Ensure you handle missing values, encode categorical features, and normalize or standardize numerical features.

**Feature Selection/Engineering:**

- If you haven't already, determine which features are most relevant to your segmentation task and conduct feature engineering.

**PCA Dimensionality Reduction:**

- Apply PCA to reduce the dimensionality of your data. PCA works by transforming your original features into a new set of uncorrelated variables (principal components) that capture most of the variance in the data.

```
from sklearn.decomposition import PCA
```

**# Choose the number of components (or set a variance threshold)**

**pca = PCA(n_components=2)**

**transformed_data = pca.fit_transform(your_data)**

Select an appropriate number of principal components, considering the cumulative explained variance ratio. You may choose to retain a certain percentage of the total variance (e.g., 90% or 95%) or based on domain knowledge.

**Clustering Algorithm Selection:**

- After PCA, you can use the reduced dataset for customer segmentation. The reduced feature set can make clustering algorithms perform better and more efficiently.

**Hyperparameter Tuning:**

- As mentioned in the previous response, use hyperparameter tuning techniques to optimize your clustering algorithm's settings, taking into account the reduced feature space.

**Validation and Evaluation:**

- Split your data into training and testing sets and evaluate your clustering model based on appropriate metrics, as mentioned earlier.

**Interpretation and Visualization:**

- Visualize and interpret the customer segments in the reduced feature space. Keep in mind that since the features are now principal components, they may not have direct interpretability as the original features.

**Implementation and Action:**

- Translate insights from your PCA-enhanced segmentation into actionable strategies, which may include personalized marketing campaigns, product recommendations, or other tailored customer experiences.

**Monitoring and Maintenance:**

- Periodically update and re-run your customer segmentation process with PCA to adapt to changing customer behavior.

**Experiment and Innovation:**

- Continue to innovate by experimenting with different combinations of PCA and clustering algorithms, as well as exploring other advanced techniques like deep learning for segmentation.