Venture Creed: Customer Segmentation Solution

Presented by Bhaskar Kumar

Identify customer groups based on purchasing behavior using advanced clustering techniques.

Data Collection

Sources

- Transactions and demographics
- Online behavior tracking
- Surveys and social media

Scale

Minimum 100,000 customer records collected

Example: Average transaction value by segment

Data Processing & Implementation

1

Data Cleaning

Handle missing values & outliers to improve quality.

2

Feature Engineering

Create variables like recency, frequency, monetary value.

3

Model Selection

Use K-means & hierarchical clustering; pick best model.

4

Evaluation

Assess cluster stability and interpret results.

5

Implementation

Integrate customer segments into marketing and CRM tools.

Target 80% accuracy in data and 90% explainability in models.

Venture Creed Solution: Transforming Marketing

Traditional Segmentation

Falls short, lacks personalization and precision.

Our Advantage

Provides data-driven insights for targeted marketing.

Real Impact

Boosts customer engagement and increases ROI effectively.

Results

Personalized campaigns increase conversion rates by 35%.

Data Collection

Sources

- Transactions and demographics
- Online behavior tracking
- Surveys and social media

Scale

Minimum 100,000 customer records collected

Example: Average transaction value by segment

Data Processing & Implementation

1

Data Cleaning

Handle missing values & outliers to improve quality.

2

Feature Engineering

Create variables like recency, frequency, monetary value.

3

Model Selection

Use K-means & hierarchical clustering; pick best model.

4

Evaluation

Assess cluster stability and interpret results.

5

Implementation

Integrate customer segments into marketing and CRM tools.

Target 80% accuracy in data and 90% explainability in models.

Clustering and Segmentation for Venture Creed

This presentation covers the development of clustering logic, feature selection, visualization, outlier detection, data quality assessment, and final validation for Venture Creed’s customer segmentation. We begin with data preparation and modeling, followed by insights into feature relevance and visual validation of clusters. The process ensures robust, business-relevant segments that support targeted marketing strategies. Finally, we address data quality issues and conclude with quality checks and export of actionable outputs.

Clustering Logic & Model Development

Data Cleaning

Started with 3,000+ customer records, dropping non-behavioral fields like Customer ID and address details.

Handling Missing Values

Fields with >30% missing got a ‘Missing’ category; others were imputed by mode grouped by Revenue Bucket.

Encoding & Scaling

Ordinal fields encoded with domain order; features scaled using StandardScaler to avoid bias in KMeans.

Cluster Selection

Tested K=2 to 10 using Elbow and Silhouette methods; optimal clusters = 4 based on highest silhouette score.

Feature Selection Strategy

Key Features Retained

Selected 11 features tied to purchasing behavior, market potential, and loyalty, such as Churn Segment and Market Share.

Dropped high-cardinality, low-value fields like address and postal info to improve model effectiveness.

Missing Data & Encoding

Imputed categorical values using 'Unknown' and numeric features using global medians. No revenue-based grouping was used in the current implementation.

Ordinal Encoder converted categorical data to numeric format, assuming implicit order. binary fields mapped to 0 and 1 for consistency.

PCA-Based 2D Visualization

Dimensionality Reduction

Used PCA to reduce features to two components capturing 68% of variance for clear visualization.

Cluster Visualization

Plotted customers in 2D space, color-coded by clusters: Premium Power, Strategic Seasonal, Recovery Priority, Emerging High-Potential.

Validation & Insights

Clear cluster separation validates segmentation and aids strategic marketing and retention discussions.

Outlier Detection

Visual and statistical methods were used to monitor potential outliers in purchasing behavior and customer segments. Techniques such as distribution plots, Z-score analysis, and PCA visualizations helped identify records with unusual patterns for further review. These records were not removed but can be flagged for downstream exception handling or business inspection.

Data Quality Issues Identified

Missing Values

High missingness in fields like EA_Segment and Seasonality_Segment required categorical imputation using most frequent values or ‘Unknown’.

Categorical Inconsistencies

Capitalization and trailing spaces caused inconsistent entries such as 'High ', ' high', and 'HIGH'.

Undefined Segments

Some categorical segments like EA_Segment lacked definitions in the data dictionary.

Ambiguous Buckets

Bucketed features like Revenue and Volume had overlapping or unclear values needing harmonization.

Final Validation & Quality Checks

Feature Verification

Confirmed that all numeric clustering features were standardized using Standard Scaler to eliminate scale bias.

Categorical variables were encoded using Ordinal Encoder, assuming implicit order where applicable.

Cluster count was validated statistically and visually using both the Silhouette Score and Elbow Method, confirming the choice of 8 clusters for optimal separation.

Data Integrity & Exports

Ensured cluster assignments were consistent with no duplicates or mismatches.

Exported customer_clusters.xlsx, cluster_statistics.xlsx, and cluster_summary.xlsx for business use.

Summary and Next Steps

We developed a robust clustering model with eight meaningful customer segments, built on strong feature engineering and data cleaning practices.
PCA-based visualization showed clear cluster separation, while statistical metrics (e.g., silhouette score) confirmed segment quality.
Data quality issues—such as ambiguous buckets and inconsistent categorical labels—were identified for future refinement.

Next Steps Include:

- Integrating clusters into CRM systems for personalized outreach
- Designing targeted marketing strategies per segment
- Continuously monitoring data quality and updating clusters quarterly to reflect customer behavior evolution

.