Objective

The primary objective of this project was to analyze quarterly relationships between economic indicators - Consumer Price Index (CPI) and Spending Momentum Index (SMI) - and e-commerce spending. By leveraging unsupervised learning methods like K-Means and DBSCAN, we sought to uncover meaningful clusters, identify outliers, and provide actionable insights for businesses and stakeholders.

Summary of Key Findings

1. K-Means Clustering:

   - Optimal number of clusters: 5

   - Silhouette Score: 0.63, indicating good cluster separation.

   - Captured distinct groupings of quarters based on CPI, SMI, and spending but forced all data points into clusters, including outliers.

2. DBSCAN Clustering:

   - Optimal Parameters: eps=1.5, min_samples=5

   - Silhouette Score: 0.63, slightly outperforming K-Means.

   - Strengths:

     - Identified meaningful clusters while preserving outliers as a separate category.

     - Highlighted anomalies in quarters (Cluster -1) with significantly low CPI, SMI, and e-commerce spending.

Cluster Insights:

- Cluster 0 (Majority Behavior): Represents typical quarters with moderate CPI, SMI, and spending.

- Cluster -1 (Outliers): Captures quarters characterized by low CPI, SMI, and e-commerce spending,

such as Q2 and Q3 2020 during the COVID-19 pandemic.

Recommendations

1. Analyze Outliers:

   - Investigate further economic events contributing to anomalies in Q2 and Q3 2020.

   - Consider external factors like policy changes, stimulus efforts, and consumer confidence.

2. Data Enhancement:

   - Include additional features like employment rates, wage growth, and consumer confidence index.

3. Modeling Improvements:

   - Apply time-series models to explore lagged relationships between CPI, SMI, and spending.

   - Experiment with advanced clustering techniques, such as Gaussian Mixture Models (GMM) and Hierarchical Clustering.

4. Business Actions:

   - Use trends in CPI and SMI to predict and adapt to future spending behavior.

   - Tailor marketing strategies during low-SMI periods to focus on essential goods or promotions.

Conclusion

This project successfully demonstrated the value of unsupervised learning techniques in analyzing economic indicators and consumer behavior. The combination of K-Means and DBSCAN clustering provided meaningful insights into quarterly trends and outlier periods. These findings offer actionable recommendations for businesses and policymakers to adapt to changing economic conditions.

Spending Analysis Project Report

By incorporating additional data and modeling techniques in future analyses, this approach can be further refined to drive deeper insights and more robust strategies for stakeholders.