

Finding And Analyzing data.

Aadi Sriram Kumar

Department of Business, Concordia University Wisconsin

DBA 9110 : Business Intelligence & Predictive Analytics

Dr. James Eden

Assignment Due Date : 07-07-2024

Finding and Analyzing Data.

This report will include a detailed analysis, data mining techniques, and their applications. The final report will detailed description of the data set.

Introduction:

This report aims to analyze the "Household Consumption of Various Goods and Services in India" data set to explore consumer behavior patterns across different demographics. The data set is obtained from the National Sample Survey (NSS) 66th round, conducted between July 2009 and June 2010. The analysis will employ various data mining techniques to uncover insights and trends that can be useful for business and Policymaker.

Fig 1: Pattern of pulse consumption, all-India

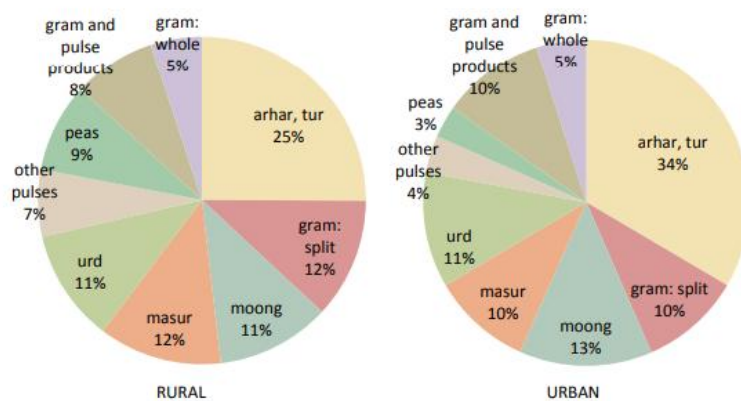


Fig 6.1: Percentages of households possessing washing machines

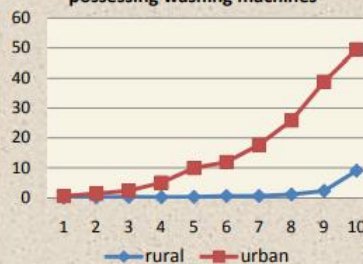


Fig 6.2: Percentages of households possessing refrigerators

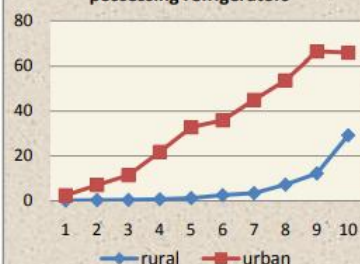


Fig 6.3: Percentages of households possessing bicycles

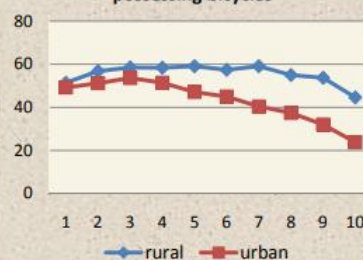
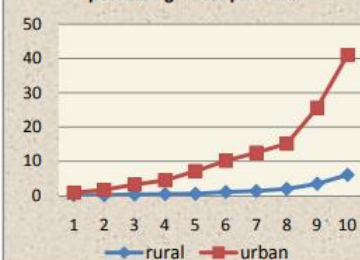


Fig 6.4: Percentages of households possessing water purifiers



Q1: Provide a short description of the data set in terms of data types and attributes.

Origin and Structure:

The data set is sourced from the National Sample Survey Office (NSSO) and includes data on household consumption across India. It comprises various attributes detailing household demographics and expenditure across multiple categories.

Data Types and Attributes:

Household Demographics
Household size
Income levels
Geographic location (rural/urban)

Consumption Categories:

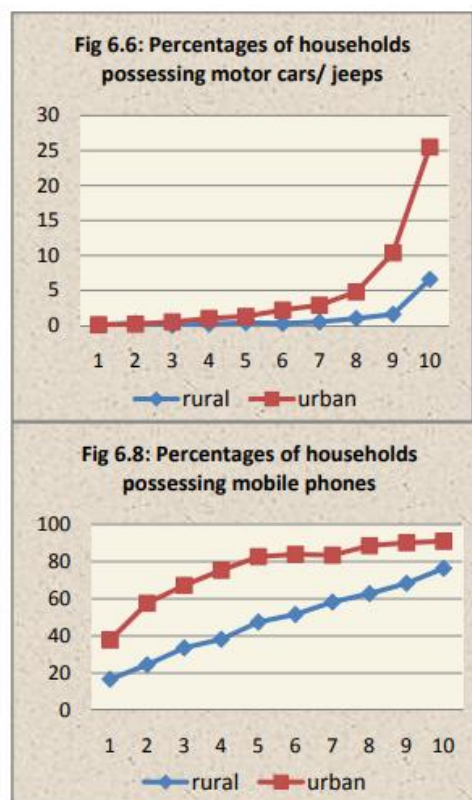
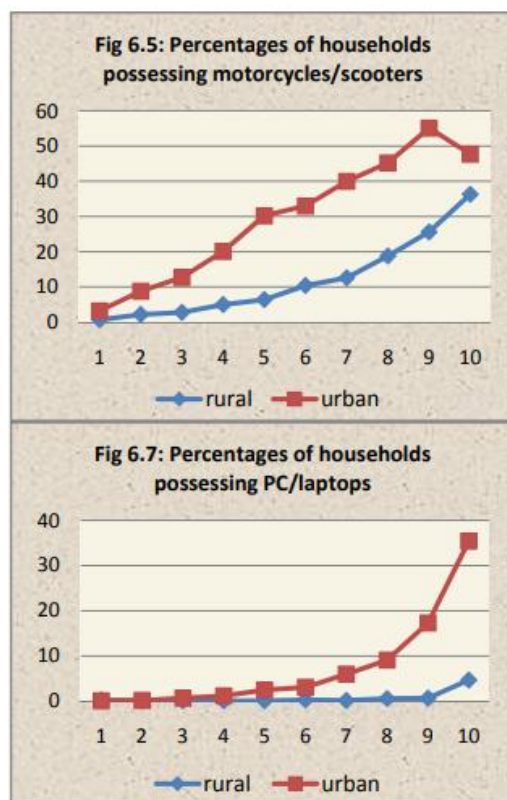
Food and beverages
Housing
Education
Health care
Transportation
Other goods and services

Quantitative Data:

Amounts spent in each category
Frequency of purchases

Categorical Data:

Types of goods and services consumed



Q2: Explain why you selected this dataset over others.

Answer:

1. **Relevance:** The dataset provides a detailed overview of consumer behavior in India, crucial for understanding economic trends and market demands.
2. **Richness:** It covers a broad range of consumption categories, making it suitable for various analyses, including predictive modeling and segmentation.
3. **Accessibility:** The data is publicly available and comprehensive, facilitating detailed analysis without additional costs.

Q3: Explain what you might be able to do with the data set using data mining techniques, models, why you wanted to do it.

Answer:

Predictive Analysis

Objective: Forecast future consumption trends based on historical data.

Techniques: Time series analysis, regression models.

Application: Businesses can use these forecasts for inventory management and supply chain optimization.

Customer Segmentation:

Objective: Segment households into distinct groups based on consumption patterns.

Techniques: Clustering algorithms (K-means, hierarchical clustering).

Application: Marketers can develop targeted strategies to appeal to different consumer segments, enhancing marketing efficiency and customer satisfaction.

Recommendation Systems:

Objective: Suggest products or services to households based on past consumption.

Techniques: Collaborative filtering, association rule mining.

Application: Retailers can improve customer experience by offering personalized recommendations, increasing sales and loyalty.

Insights and Trends:

From the visualizations and initial analysis, we can derive several insights:

Spending Patterns: Households spend the most on food and housing, with significant variations across income levels and geographic locations.

Education and Healthcare: There is a notable difference in expenditure on education and healthcare between urban and rural households.

Predictive Modeling: Time series analysis can be applied to forecast future spending trends, helping businesses plan better.

Conclusion

This report analyzed the "Household Consumption of Various Goods and Services in India" dataset using data mining techniques to uncover consumer behavior patterns. The findings indicate significant insights that can help businesses and policymakers make informed decisions. The dataset's richness and relevance make it a valuable resource for various analytical applications.

References

National Sample Survey Office. (2012). Household Consumption of Various Goods and Services in India. Government of India, Ministry of Statistics and Programme Implementation.

Larose, D. T., & Larose, C. D. (2015). Data Mining and Predictive Analytics (2nd ed.). Wiley.

Sahay, B. S. (2012). Data Analytics and Decision Support Systems. Springer.