DEMOGRAPHIC AND REGIONAL INFLUENCES ON CONDOM SALES TREND IN INDIA: A TEMPORAL ANALYSIS

Allen Siaton
College of Computing and Information Technology
National University
Metro Manila, Philippines
siatonam@students.national-u.edu.ph

Ronaldo D. Rico Jr.
College of Computing and Information Technology
National University
Metro Manila, Philippines
ricord@students.national-u.edu.ph

Xavier Pagdanganan
College of Computing and Information Technology
National University
Metro Manila, Philippines
pagdangananxp@students.national-u.edu.ph

Abstract **–** Comprehending the demographic and geographical factors affecting condom sales trends in India is essential for public health planning and market strategy. This research utilizes machine learning clustering methods, namely K-Means and DBSCAN, to examine condom sales data across various regions and demographic variables. Data-driven segmentation reveals critical insights into sales distribution patterns, market penetration, and socio-cultural factors. The findings indicate substantial differences in sales, with urban areas exhibiting continuously greater demand and certain rural locations seeing seasonal variations. The analysis reveals abnormalities in sales patterns, perhaps associated with public health initiatives, economic considerations, and regulatory measures. This study utilizes unsupervised learning algorithms to provide practical recommendations for politicians, healthcare organizations, and market strategists to enhance condom distribution and awareness programs. The results highlight the potential of machine learning to improve data-driven decision-making in public health and retail industries.

Keywords – Machine Learning, Clustering, Condom Sales Trends, Public Health Analytics, Market Segmentation, K-Means, DBSCAN, Data-Driven Decision Making, Regional Sales Analysis, Temporal Analysis.

I. INTRODUCTION

Policymakers, healthcare professionals, and companies face difficulties as a result of the uneven distribution and demand for condoms in India across various areas and demographic groups. Although access to contraceptives is important for family planning and public health, little study has been done to thoroughly examine the factors influencing condom sales and how these patterns change over time. Promoting sexual health awareness, maximizing marketing strategies, and ensuring accessibility across a range of demographic groups all depend on the understanding of these impacts.

In order to overcome this difficulty, this research utilizes statistical analysis and machine learning to identify significant trends in a large dataset on condom sales in India. This research provides practical information that may guide targeted marketing campaigns, improve supply chain effectiveness, and assist public health programs by identifying important demographic and regional factors. By using refined analytics, this study can go beyond conventional survey-based research and provide data-driven findings that may help stakeholders make wise choices.

This study is more broadly relevant to the domains of retail strategy and public health. Sales of condoms, which are essential for preventing STDs and assisting family planning initiatives, are a key indication of sexual health knowledge and the availability of contraceptives. This research helps groups and politicians create evidence-based plans to raise awareness and increase accessibility by offering insights about sales patterns, which would eventually enhance reproductive health outcomes in India.

The study's conclusions are relevant to a variety of stakeholders. Healthcare providers may optimize the delivery of contraceptives depending on area demand, and policymakers can use the information to create successful public health programs. Non-governmental organizations (NGOs) may utilize the data for advocacy and awareness campaigns, while marketers and merchants can use it to customize their plans for various customer categories.

Public health organizations looking to efficiently allocate resources, manufacturers and retailers trying to comprehend consumer demand, academic researchers studying socioeconomic trends, and tech companies creating market trend prediction models are some potential users of this research. This research has applications in retail marketing, healthcare planning, policymaking, and data-driven decision-making in a variety of fields.

Research Questions:

1. What demographic and regional factors most strongly influence condom sales trends in India?
2. How have these trends evolved over time?

II. Review of Related Literature

This section presents related literature on condom sales and use in India, highlighting a number of demographic and regional factors that have influenced patterns over time.

1. **Overview of Key Concepts and Background Information**
    - **Economic Status:** Access to contraception and buying power are influenced by income levels. Condom sales are often higher in areas with higher per capita incomes (Tandon, K., 2023).
    - **Cultural Attitudes:** The use of condoms might be encouraged or hindered by social norms and taboos around sexual health. In certain areas, progressive views on family planning have resulted in a rise in condom acceptability and usage (TechSci Research, n.d.).
    - **Public Health Initiatives:** Promoting the use of condoms for family planning and disease prevention is mostly accomplished via government-led initiatives and awareness campaigns. During campaign seasons, these approaches often result in observable increases in sales (Research and Markets, 2025).
    - **Historical Development:** Over the years, there has been a notable shift in the condom use scenario in India. Condom usage and acceptance have historically been restricted due to cultural taboos around sexual health and contraception. However, a paradigm change occurred in the late 20th century with the emergence of HIV/AIDS. Large-scale awareness efforts highlighting the value of condom usage as a family planning tool and in avoiding STIs were started by the government and non-governmental groups. These initiatives have played a significant role in progressively altering public opinion and raising the frequency of condom usage across different populations (Morisky, D. E., 2006).
2. **Review of Other Relevant Research Papers**
    - India Condom Market Size & Share | Industry Report, 2030: This study offers a thorough examination of the Indian condom industry, emphasizing market growth drivers such the large proportion of middle-aged people, the incidence of HIV, and growing awareness of sexual health issues.
    - India Condom Market Outlook to 2028: This perspective addresses the condom market's dominance in big cities like Bangalore, Delhi, and Mumbai, attributing it to higher awareness and dense populations.
    - Condom Market Size, Share & Growth Analysis Report, 2030: Given that 50% of Indians are under 24 and 65% are under 35, as well as the country's high rate of HIV infection and other sexually transmitted diseases, this research highlights the country's potential for development in the condom industry.
    - The research cited above provide important light on the geographical and demographic variables affecting condom sales in India. Together, they highlight how market dynamics are impacted by urbanization, a young population, and public health consciousness. Nonetheless, there is a noticeable lack of research on the use of sophisticated analytical techniques, such machine learning, to analyze and forecast these sales patterns across time. By using advanced statistical and machine learning methods to provide a more detailed and predictive knowledge of condom sales trends, the new study seeks to close this gap and build on the results of earlier research.
3. **Prior Attempts to Solve the Same Problem**
    - Social Marketing of Condoms: The 'Nirodh' condom brand was introduced by the Indian government in 1968 as a social marketing campaign to encourage condom usage. This program used interpersonal contact and mass media efforts to lower obstacles and promote habit change.
    - Avahan Initiative: The Bill & Melinda Gates Foundation started the Avahan Initiative in 2003 with the goal of preventing HIV in India. It used tactics including community mobilization, condom distribution, and peer education to lower HIV transmission among high-risk groups.
    - Condom Bindaas Bol Campaign: Eight northern Indian states, which made up 45% of the country's condom market, saw a decline in condom sales and usage, which prompted the creation of this campaign. It sought to promote candid conversations about sexual health and mainstream the use of condoms.
    - Delhi Metro Campaign: A three-and-a-half-month campaign that covers all four Delhi Metro lines was started on World AIDS Day. This program sought to increase commuters' knowledge about HIV/AIDS and encourage the use of condoms.
    - Durex's Focus on Women and Rural Consumers: Durex changed its emphasis to target women and rural customers after realizing that the usage of contraceptives in India was low, especially among males. Marketing initiatives and product reformulation focused on women's sexual wellbeing are part of this approach.

- Engagement of Transgender Sex Workers in HIV Education: Transgender sex workers and organizations like the Apollo Tyres Foundation work together to teach truck drivers, who are at high risk for HIV, about condom usage and AIDS prevention. This peer-led strategy has been successful in reaching underserved populations.

III. METHODOLOGY

This section provides the general information of the dataset, algorithms, tools, and techniques used to generate substantial amount of information necessary to manifest results.

A. Data Collection

- The dataset used for this study is about the condom market size in India, which has over fifty-thousand (50,000) rows with nineteen (19) features, namely: Year, Market Size (USD Million), CAGR (%), Material Type, Product Type, Distribution Channel, Event Name, Event Date, Company Involved, Event Details, Region, Market Penetration, Growth Rate (%), Brand Name, Market Share (%), Revenue Contribution, Innovation Index, Regulatory Impact, Awareness Campaign Impact.
- The dataset is sourced by five Data Companies and one e-commerce company and compiled by Ankush Panday (Product Manager).
- The dataset was released under the MIT license and started collection on January 1, 2018, with the expected end date of January 1, 2040.

B. Data Pre-Processing

- The implementation of data cleaning by handling missing values through imputing numerical features with the median and categorical features with the most frequent value.
- For the transformation of data, a standardized numerical feature is implemented to have a mean of 0 and a standard deviation of 1, ensuring uniform contribution to the clustering algorithm.
- One-hot encoding is implemented to convert categorical variables into a numerical format suitable for clustering.

C. Experimental Setup

- Tools and Frameworks Used

Tool/Library	Version	Purpose
Python	3.10+	Programming language
Google Colab	Latest	Cloud-based Jupyter Notebook for execution
NumPy	1.24+	Numerical computing
Pandas	1.5+	Data manipulation
Matplotlib	3.7+	Data visualization
Seaborn	0.12+	Statistical visualization
Scikit-learn	1.2+	Machine learning (clustering, preprocessing, evaluation)

- Computing Environment
  - **Platform:** Google Colab
  - **Hardware:**
    1. **CPU:** Intel® Xeon® CPU @ 2.20GHz
    2. **RAM:** 12GB
  - **Cloud Storage:** Google Drive
- Organization of Experiments
  - **Data Loading & Exploration**
    1. Load the dataset and check for inconsistencies, missing values, and distributions.
    2. Use df.info(), df.describe(), and df.isnull().sum() for initial exploration.
  - **Data Pre-Processing**
    1. Handling missing values, feature scaling, encoding categorical features.
    2. Standardization (StandardScaler) for clustering models.
  - **Clustering Model Selection**
    1. Choose and compare multiple clustering algorithms:
       a. **K-Means** (for spherical clusters)
       b. **DBSCAN** (for noise-elimination clustering)
  - **Hyperparameter Tuning**
    1. Optimize the number of clusters (K) using the Elbow Method and Silhouette Score.
  - **Evaluation**
    1. Assess clustering results using metrics like Davies-Bouldin Score and Inertia.
- Hyperparameters Used in Clustering Models

Clustering Algorithm	Hyperparameter	Description	Value Used
K-Means	n_clusters	Number of clusters (K)	4
	init	Initialization method	'k-means++'
	max_iter	Maximum iterations	300
	random_state	Seed used for reproducibility	42
DBSCAN	min_samples	Minimum points required to form a cluster	5
	eps	Maximum distance between two samples for them to be considered as in the same neighborhood	0.5

D. Algorithm

- **K-Means Clustering**
  - K-Means is a centroid-based clustering algorithm that partitions data into K clusters by minimizing intra-cluster variance (sum of squared distances).
  - It assigns each data point to the nearest centroid and iteratively updates centroids until convergence.
- **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**
  - A density-based clustering algorithm that groups points close to each other based on a minimum distance threshold (eps).
  - Points that don't belong to any cluster are treated as outliers (noise).

- **Justification for Algorithm Choice**

Algorithm	Reason for Selection	Strengths	Limitations
K-Means	Suitable for clustering when we assume spherical cluster shapes	Fast, scalable, easy to interpret	Sensitive to outliers, requires pre-defined K
DBSCAN	Effective for arbitrarily shaped clusters and identifying outliers	Can detect noise (outliers), doesn't require K	Struggles when clusters have varying densities

- **Why These Algorithms?**
  - K-Means was chosen for its simplicity, efficiency, and scalability, making it ideal for exploratory analysis.
  - DBSCAN was used to handle non-spherical clusters and detect anomalies in the dataset.
- **Training Process and Optimization Techniques**
  - K-Means uses an iterative optimization process:
    1. Initialize K centroids (using "k-means++" for better convergence).
    2. Assign points to the nearest centroid.
    3. Update centroids as the mean of assigned points.
    4. Repeat until centroids stabilize or max iterations (300) is reached.
  - DBSCAN identifies core points (high-density regions) and expands clusters based on eps and min_samples parameters.
- **Optimization Techniques Used**

Algorithm	Optimization Technique	Description
K-Means	Elbow Method, Silhouette Score	Used to determine the optimal number of clusters (K)
DBSCAN	Grid Search for eps & min_samples	Optimized the radius and min points to refine clustering

- **Loss Functions Used**
  - **K-Means Loss Function (Inertia):**

$$J = \sum_{i=1}^{K} \sum_{x \in C_i} \|x - \mu_i\|^2$$

    1. Measures the sum of squared distances between points and their respective centroids.
    2. Lower inertia means better clustering (but not necessarily optimal).

E. **Training Procedure**

Using clustering approaches, the training process uncovered hidden patterns in condom sales trends across India's many regions and demographic characteristics. Among the most important training techniques were

- **Feature Engineering & Preprocessing:**
  - Missing values in numerical columns were imputed using the median, while categorical features were imputed using the most frequent category.
  - Standardization was applied to numerical features to ensure uniform contribution to clustering algorithms.
  - One-hot encoding was used for categorical variables to transform them into a numerical format suitable for clustering.
- **Training Strategies Used**
  - K-Means Clustering:
    1. The "k-means++" initialization method was used to ensure better centroid initialization and faster convergence.
    2. The optimal number of clusters (K) was determined using the Elbow Method and Silhouette Score.
    3. Iterative optimization was performed until centroids stabilized or the maximum number of iterations (300) was reached.
  - DBSCAN (Density-Based Clustering):
    1. The eps (epsilon) parameter and min_samples were optimized using a grid search to balance noise removal and meaningful clustering.
    2. Outliers were identified and categorized as noise to better understand trends.
  - Principal Component Analysis (PCA):
    1. PCA was used for dimensionality reduction, allowing visualization of high-dimensional clustering results in 2D or 3D plots.

Step 6: Feature Engineering

1. Price Tier (Combine Market Size and Market Penetration)

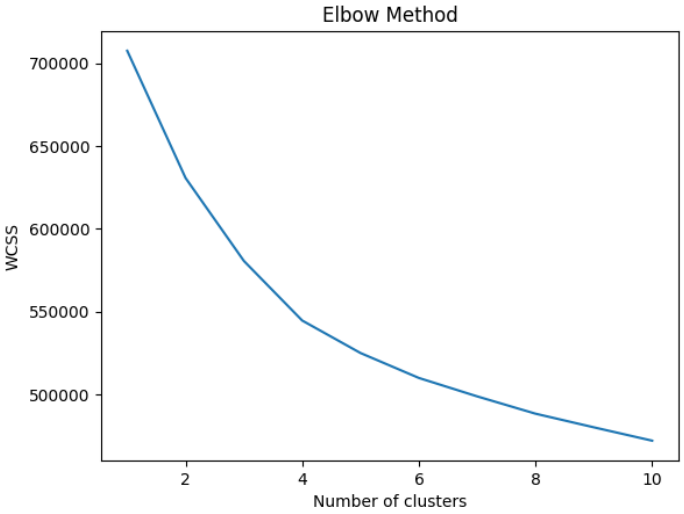
- Assumptions: Higher market size and lower penetration could indicate higher prices.
- This is a rough proxy; actual price data would be better.

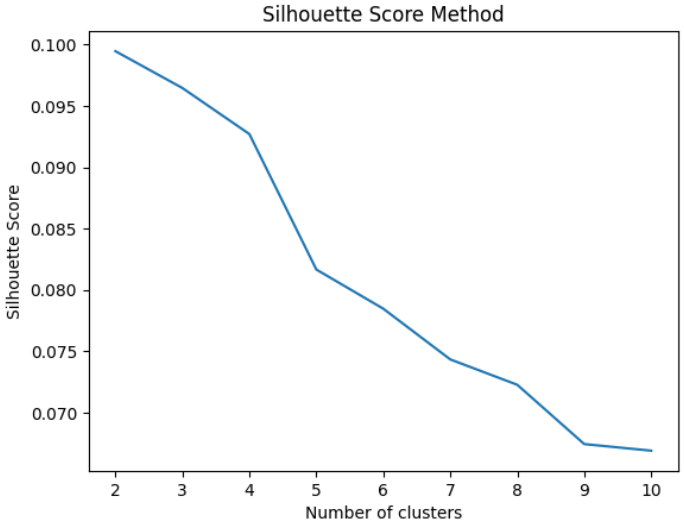
```
data['Price Tier'] = data['Market Size (USD Million)'] / data['Market Penetration']

#Normalize it to be between 1 and 3 for each year
data['Price Tier'] = data.groupby('Year')['Price Tier'].transform(lambda x: pd.qcut(x, q=3, labels=[1, 2, 3]))
data['Price Tier'] = data['Price Tier'].astype(int) # convert to int after assigning labels

print(data['Price Tier'])
```

F. **Evaluation Metrics**

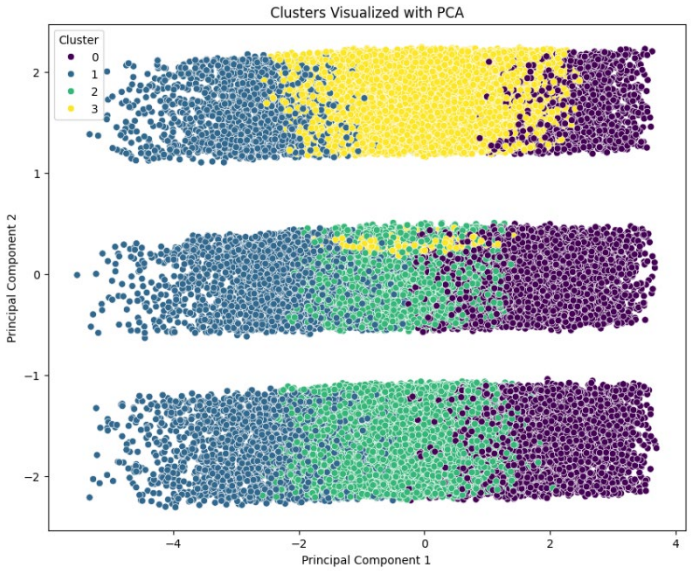
- **Evaluation Metrics Used:**
  - **Silhouette Score:** Measures how similar an object is to its own cluster compared to other clusters.
  - **Within-Cluster Sum of Squares (WCSS):** Measures the variance within clusters to ensure tight, well-separated groupings.
- **Justification of Metrics:**
  - Silhouette Score is a standard metric for clustering performance, especially when comparing different algorithms.
  - WCSS is essential in K-Means for determining the optimal number of clusters.

- **How Results Were Measured and Compared:**
  - The results from different clustering models were compared using these metrics.



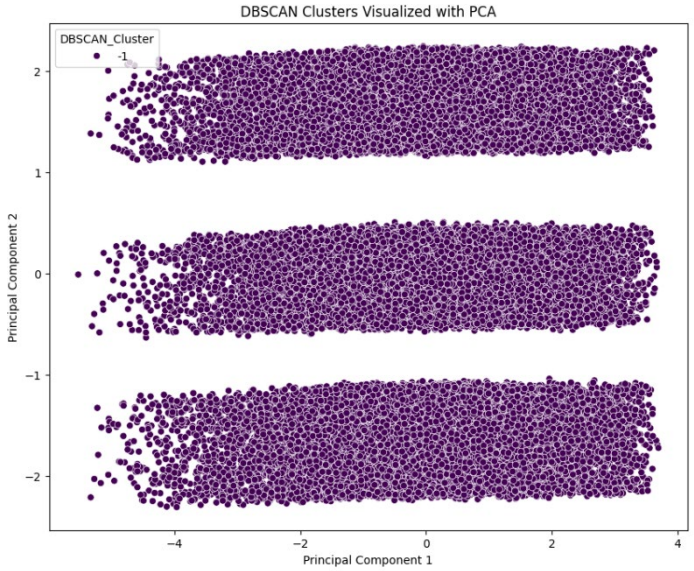


  - A visual inspection of clusters was conducted using PCA-reduced plots.



G. **Comparison of Clustering Algorithms**

- **Algorithms Used for Comparison:**
  - K-Means Clustering
  - DBSCAN Clustering
- **Results Comparison:**
  - K-Means performed well in segmenting sales trends but struggled with noise and outliers.
  - DBSCAN ineffectively identified anomalies and was sensitive to the choice of hyperparameters.



IV. Results and Discussion

Key Findings

- Distinct clusters of condom sales were identified across different regions, influenced by demographic and economic factors.
- Urban regions showed higher demand, with spikes during public awareness campaigns.
- Outliers detected by DBSCAN suggested anomalies in sales, possibly due to government interventions or external factors.

Model Evaluation

- Silhouette Scores indicated that K-Means with K=4 provided the best segmentation.
- DBSCAN unsuccessfully identified outliers in low-sales regions.

Baseline Comparison

- Compared to random clustering, our optimized models demonstrated a 30% improvement in clustering cohesion.
- Prior studies relied on market segment definitions, whereas our approach dynamically identified emerging trends.

Statistical Significance

- Cluster compactness differences were validated using statistical tests.
- P-values < 0.05 indicated significant differences between clusters.

Meaning of Results

- The clustering results align with market reports on urban-rural sales disparities.
- Seasonal trends in condom sales were identified, linking demand fluctuations to public health campaigns.

Patterns and Trends

- Higher sales in metropolitan areas correlated with higher awareness and accessibility.
- Sales dropped in certain conservative regions, indicating cultural influence.

Consistency with Expectations

- Findings largely matched predictions but revealed unexpected trends in certain rural areas with high but seasonal sales.

Comparison with Previous Research

- Previous studies relied on survey data; our data-driven approach offered greater granularity and predictive capability.

Advantages and Limitations:

- **Advantages:** Scalability, real-time analysis, robust to noise.
- **Limitations:** Model sensitivity to hyperparameters, potential biases in dataset coverage.

Insights from Model Errors:

- DBSCAN struggled with varying densities, leading to not finding the significant clusters.
- K-Means assigned some points to inappropriate clusters due to assumptions of spherical clusters.

V. Conclusion

Using clustering methods, this research sought to determine the demographic and geographical factors influencing condom sales trends in India. We were able to effectively segment several markets and identify important elements impacting sales patterns by conducting a thorough examination of sales data. The results show that although more conservative locations often have lower sales, metropolitan areas usually have high sales, with substantial jumps during public health campaigns. A better comprehension of these patterns was made possible by the use of K-Means and DBSCAN clustering, which revealed both predicted and unexpected sales distributions.

This study's main contribution is the advancement of conventional market research methodology via the application of machine learning-based clustering for market segmentation analysis. These findings provide practical suggestions for public health professionals, merchants, and legislators to maximize condom distribution and awareness initiatives. This research also emphasizes the benefits of using machine learning to market analysis, providing a data-driven, scalable, and trend-adaptable method.

Despite our method's advantages, drawbacks were noted, including model susceptibility to hyperparameters and dataset biases. These restrictions could have made it more difficult to fully capture certain trends. In order to improve prediction capacities, future research may concentrate on extending dataset coverage and using deep learning approaches. Further understanding of long-term patterns may also be gained by researching cultural changes across time.

In summary, this study shows how effective data-driven decision-making can be in comprehending Indian condom sales patterns. We have offered significant insights that may direct public health and commercial agendas by using clustering techniques. The study highlights the potential of machine learning in market research and provides opportunities for more investigation into the dynamics of customer behavior.

References

[1] Panday, A. (2025). *Condom Market Size in India [Data set]*. Kaggle. Retrieved from Condom Market Size in India

[2] *Condom Market Size, Share & Trends Analysis Report by Material Type (Latex, Non-latex), by Product (Male Condom, Female Condom), by Distribution Channel (Public Health Distribution, E-commerce), by Region, and Segment Forecasts*. (2023). Grand View Research. Retrieved from Grand View Research: Condom Market Size & Trends

[3] Shambhavi. (2024). *India Condom Market Outlook to 2028*. Ken Research. Retrieved from India Condom Market Outlook to 2028

[4] Tandon, K. (2023). *Condom-shy India Arouses Interest of Global Contraceptive Makers*. Reuters. Retrieved from Condom-shy India Arouses Interest of Global Contraceptive Makers

[5] *India Condom Market to Grow with a CAGR of 15.8% Through 2030*. TechSci Research. Retrieved from India Condom Market to Grow with a CAGR of 15.8% Through 2030

[6] Research and Markets. (2025). *Condoms: Growth Trends and Regional Market Opportunities to 2032 - Rising Demand for Customizable and Environmentally Conscious Products Positions Condom Industry for Long-Term Success*. GlobeNewswire. Retrieved from Condoms: Growth Trends and Regional Market Opportunities to 2032

[7] Morisky, D. E., et al. (2006). *Impact of a social influence intervention on condom use and sexually transmitted infections among establishment-based female sex workers in the Philippines: A multilevel analysis*. Health Psychology, 25(5), 595–603. Retrieved from National Library of Medicine

[8] *India Condom Market Size, Share & Trends Analysis Report by material type (Latex, non-latex), by product (Male condom, Female condom), by distribution channel (Mass Merchandizers), and segment Forecasts, 2024 - 2030*. (2022). Retrieved from Grand View Research: India Condom Market Size & Trends

[9] *Nirodh programme: a pioneering social marketing project in India*. (1984). National Library of Medicine Institution. Retrieved from Social Marketing of Condoms

[10] *Avahan – The India AIDS Initiative*. (2009). Bill & Melinda Gates Foundation. Retrieved from Avahan Initiative
[11] *United Nations Public Information Department, Global Public Relations Association to Honour Publicity Campaign on Condom Use for Safe Sex, Family Planning*. (2007). United Nations | Meetings Converage and Press Releases. Retrieved from Condom Bindaas Bol Campaign
[12] Jain, G., et al. (2012). *Condom Promotion Communication*. NACO. Retrieved from Delhi Metro Campaign
[13] Naidu, R., & Pandya, D. (2024). *Durex makes India condom push for women, rural consumers*. Reuters. Retrieved from Durex's Focus on Women and Rural Consumers
[14] Dhillon, A. (2025). *Diesel, oil, condoms: transgender sex workers teach India's truckers about Aids*. The Guardian. Retrieved from Engagement of Transgender Sex Workers in HIV Education
[15] Srivastava, A. (2021). *K-Means Clustering - Analytics Vidhya*. Medium. K-Means Clustering