# Trends in Contraception Awareness And Child Health: Analysis of DHFS data

Shaurya Garg[#1], Akshat Johar[*2], Saksham Bansal[#3]

BITS Pilani, Hyderabad Campus

*Abstract*— **This project explores the problem of awareness of contraception among the Indian population differentiating based on child health and a number of children in different demographic areas and across different wealth indices.**
**Family planning and the use of contraceptives have been found to have significant effects on how mothers and children behave in households. There is a stark difference in the number of children born in each household in rural v/s urban areas and even in different geographical regions of India as divided into zones by us, also the awareness of contraception and the longevity of its use is correlated with all of the above findings.**

*Keywords*— **DHFS, contraceptives, birth weight**

*Colab Link*—
Data Extraction and visualisation
*Data Analysis*

## I. INTRODUCTION

The majority of Indians are aware of various forms of birth control. And more than 98% of men and women who are currently married and are between the ages of 15 and 49 are familiar with at least one form of contraception. Both men (52%) and women (52%) who are currently married are aware of emergency contraception.

The adolescent population now knows more about contraception. Nonetheless, despite these statistics, 20% of men think a woman who uses contraception may become promiscuous, and 35% of men think contraception is a woman's issue.

Birth control tablets lower the risk of pregnancy. Moreover, they lessen the chance that a fertilised egg may implant outside the uterus (ectopic pregnancy). Most frequently, an ectopic pregnancy develops in one of the tubes that transfer eggs from the ovaries to the uterus (fallopian tubes). There is a marginally increased risk of an ectopic pregnancy if you become pregnant while taking a progestin-only birth control pill, sometimes known as the mini pill.

In this survey, the main focus has been on general health parameters, the wealth index, rural versus urban households and its effects on wanted versus unwanted pregnancy. This, in turn, is checked through the child's health whether which of the above reasons has more correlation.

In addition, patterns were discovered that might be used to forecast children's health since mothers' health and wise health-care choices made during pregnancy were proven to have a significant impact. And it was shown that awareness of contraception affected institutional births and a number of other aspects of women's and children's health.

## II. PROBLEM DEFINITION

The present study aims to examine the influence of household contraception awareness and its diverse determinants. The impact of different factors on children's health is also analysed.

## III. DATASET DESCRIPTION

The DHS Program, i.e. Demographic and Health Survey is responsible for providing great tech support to a lot of the countries all around the globe since the dawn of 1984.The surveys which come under the DHS gather data on different things such as child and maternal health, reproductive health , various diseases such as malaria , HIV and AIDS, data related to individuals and where the live and also data related to nutrition and personal hygiene of all men , women and children. The main aim of the DHS Program is to gather and analyse this data to draw important insights which can lead to policy making by the government of that specific country/locality for the betterment of the population.

Political Leaders, government policy makers, professionals in the healthcare department, media representatives, basically anyone who can have an affect over the current situation of the society is provided with the datasets captured and surveyed by the DHS. These datasets are open for all and anyone can access these without any physical cost.

The DHS Program has the responsibility of collecting and distributing the data surveyed on the topics such as population health, education, diseases, hygiene, etc, to the respective nations so that the policymakers of that particular nation can take the action and formulate programs timely. The program is supported globally by some of the top non-profit organizations such as WHO, UNICEF, etc.

The technical teams creating and supporting the surveys work closely together, and the DHS and Multiple Indicator Cluster Surveys are very comparable.

## A. Content of the dataset
- 1500 Discrete valued attributes (including but not limited to):
  - Child Health Indicators
  - Women Health Indicators
  - Contraception Methods
  - Literacy rates
  - Wealth indices
- A total of 4 files with each having three timelines of study each which have over y attributes each.

### IV. DATASET DESCRIPTION

This section describes the preprocessing techniques used on the dataset and the reasons for selecting those techniques.

This Dataset contained information on every child ever born to interviewed women.

### A. Data Filtering

Filtering or attribute subset selection is the process of extracting a smaller portion of the main dataset. This portion basically consists of the attributes/columns which highly contribute to the problem statement that one is addressing to. With the help of data filtering , unnecessary columns are removed from the dataset which helps in reducing the dimensionality of the dataset. This eases out the processing phase by a great extent.

Based on the domain knowledge and our problem statement we finalized on attributes some of which are listed below
- Highest Education level
- Wealth Index Combined

- Total Children Ever Born
- Heard of Family Planning Awareness
- Size of child at birth
- Does smoking
- Drinks Alcohol

### B. Data Filtering

Cleaning basically refers to making the data consistent in order to gain the correct insights from the analysis phase. It includes removing the missing values, replacing the missing values with a common value, and eliminating duplicity in the dataset.

As our dataset was already very large and contained over 12.75 lakh rows we calculated the frequency of different values inside each attribute as our data was mostly discrete. This gave us the leverage to fill the missing values with NaN and then delete all the rows which had at least one NaN value or a value which didn't contribute to the survey. This reduced our data from 1274250 rows to 14554 rows which is a reduction of 98.87% of rows.

For the states, we calculated the number of entries for each state and removed less than 100 entries so that we get a more holistic view of the region and remove the outliers. This analysis brings the number of rows to 14218.

### C. Feature Creation

The process of creating new features from a given dataset using data mining techniques is known as feature creation. Understanding data and finding patterns that can enhance their performance is helpful to machine learning algorithms. Here we have used data aggregation as a method to create new features.

The underlying issue and features existing in the raw data heavily influence the development of new features. Faster training and more precise predictions are produced by better features.

Based on the column of states, we divided the data into six zones namely Southern, Western, Eastern, North Eastern, Central, and North Zone.

This would help us analyze the results better and correlate them region-wise as due to this data aggregation, the results will be more uniform when seen over large data over a large timeframe.

This results in 14218 × 18 columns.

## D. Normalization

Normalization is one of the vital pre-processing techniques in store. It basically comes in helpful when the attributes belong to different scales and in order to gain the correct insights from the analysis phase, all these attributes are brought down to a single scale.

With the help of normalization, each attribute, now being brought down to a consistent scale, can be given an equal amount of importance in the analysis part, and no attribute dominates the other, which may lead to wrong prediction. In Machine Learning, with the help of normalization, the prediction accuracy increases to a great extent as the model is trained correctly.

We will start by discussing min-max normalization. It is the original unstructured data's linear transformation. The data is scaled from 0 to 1. The formula used is

$$v' = \frac{v - \min F}{\max F - \min F}(new\_max_F - new\_min_F) + new\_min_F$$

We normalized the attribute of child size and wealth index. Both were brought to the range of 0 to 1.

## E. Sampling

When we are faced with datasets consisting of a lot of entries or rows, it is quite difficult to gain insights out of it. With the help of Sampling, a smaller portion of the dataset is made available which maintains the data integrity as well as eases down the analysis phase.

Now we have used a sampling rate of 25 samples from each state to reduce the number of rows for ease of analysis.

This brings number of rows down to 650 across 18 attributes.

### V. DATA VISUALIZATION

Relationships between gender, literacy, home amenities, women's and children's health, etc. were found using all the major visualization approaches (histogram, boxplot, pie-chart, scatter plot, etc.). These visualizations gave us some fascinating directions to investigate. It became clear how varied the interactions of elements are in urban and rural households as well as across the different zones of the country.

We discovered some interesting relationships between variation of birth weight within different zones.
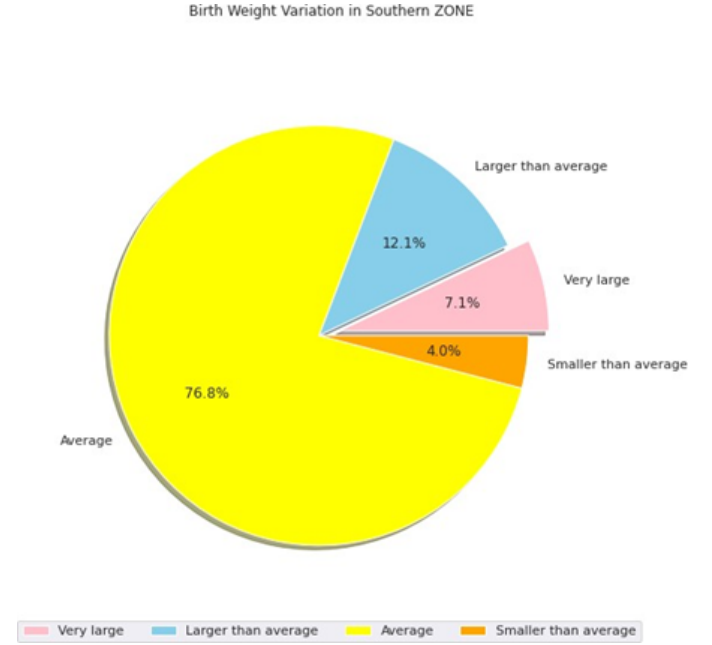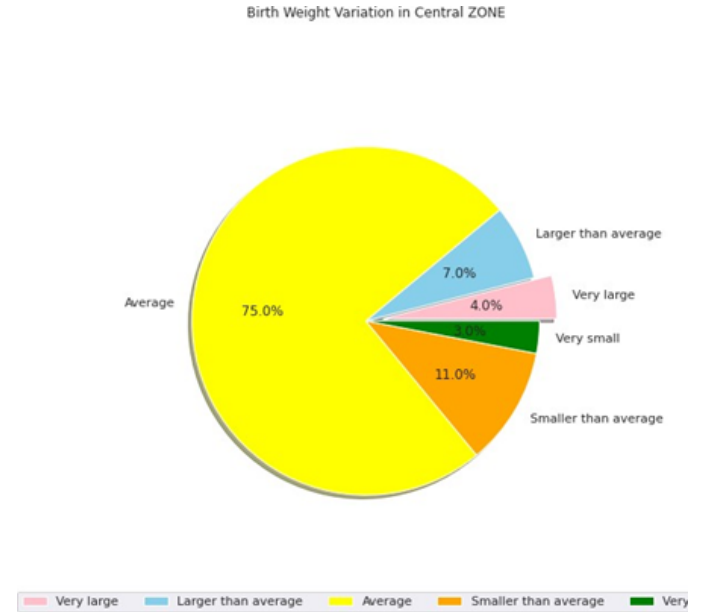


*Fig.1*



*Fig.2*

In Fig. 1 and Fig. 2 we can see, that average sized child is more frequent but the central zone has more frequency of smaller than average children whereas the southern zone has more children who

are larger than average size indicating the presence of better health of a newborn child in the southern region as compared to the central region.

As can be seen from the figures below that in urban India[Fig.3] 75% of families have at most 2 children. Here the key is (pink - 1 child, blue - 2 children, yellow - 3 children, orange - 4 children, green - 5 children, red - 7 children).
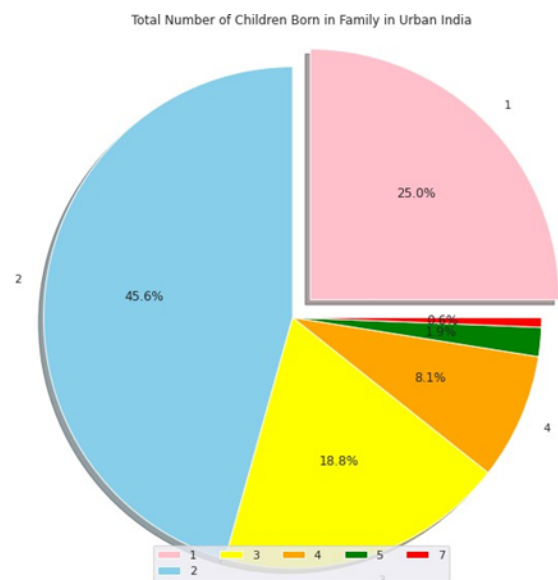


*Fig.3*

Whereas in rural India[Fig.4], there is a higher preference for having more children. As can be inferred that 25% of households have 5 children. With this comparison, we can infer that there is a lack of family planning awareness when it comes to the rural sector.
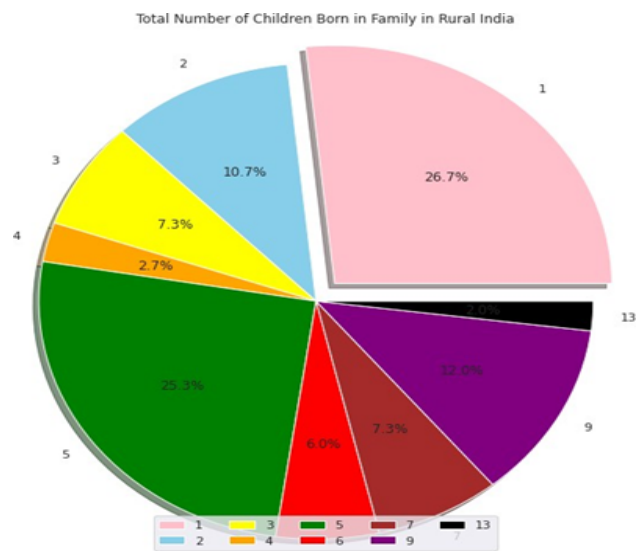


*Fig.4*

Fig.5 depicts that there is a general tendency of 2 children per household but whereas the northeastern zone has over 45% of households having at least 3 children, the north zone has 76% of households with at most 2 children.
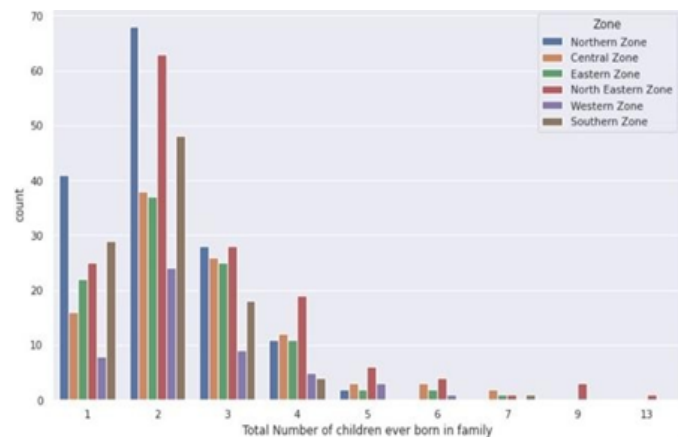


*Fig.5*

Fig.6 and Fig,7 showcase the choice of contraceptive use across the northern as well as the northeastern zone.
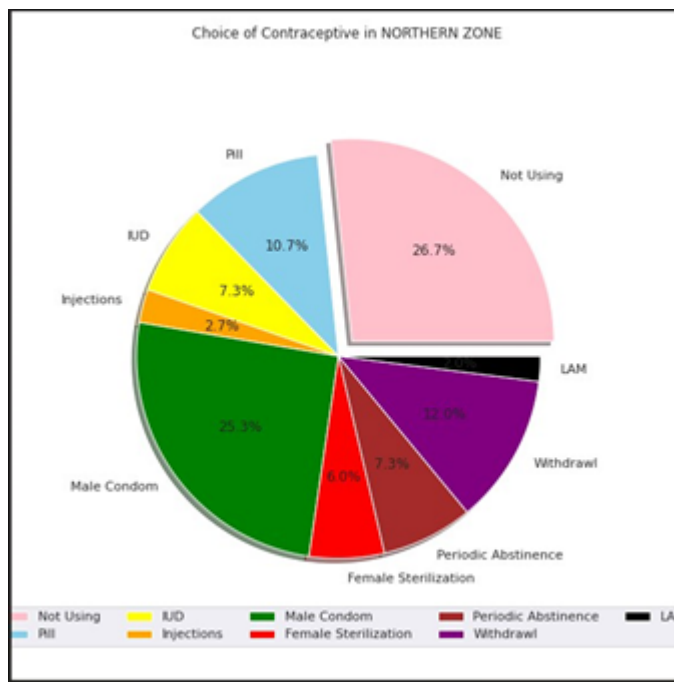
*Fig.6*

In the Northern Region about 26.7% are not using any contraceptive methods whereas in the North Eastern region about 38% of the households are not using any contraceptive methods. This analysis can be concurred with Fig.5 results which depict more number of total children born in a north eastern family as compared to a family belonging to the northern region.
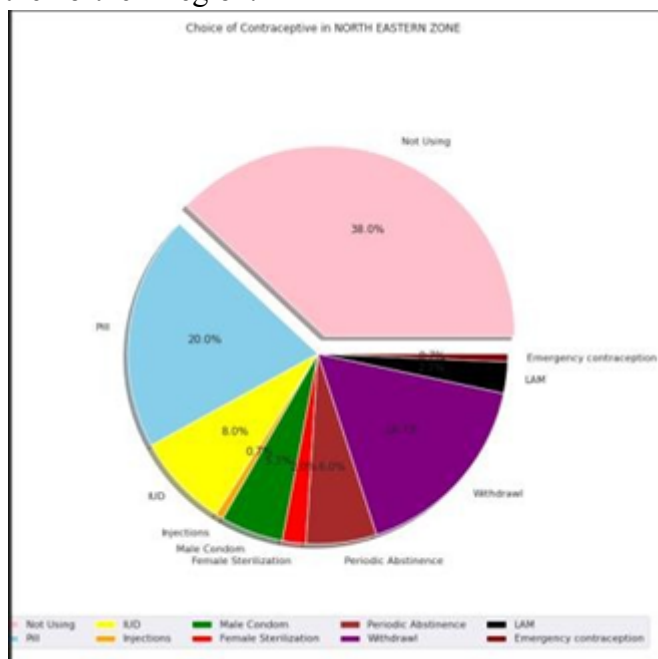


*Fig.7*

VI.       LITERATURE SURVEY

[1] Increased contraceptive usage in 3rd world countries has reduced the frequency of maternal mortality by 40% over the last 2 decades, just by decreasing unwanted births among the population. By taking preventive measures about high-risk pregnancies, particularly in women with high parities, and those that would have resulted in unsafe abortions, increasing contraceptive usage has decreased the mortality ratio of mothers by nearly 26 in every 100 live births in just a span of 10 years. Also, if the unfulfilled demand of various contraceptives is met then a reduction of more than 30% is still possible. With non-contraceptive benefits of individual techniques included, modern contraceptives give more benefits to women's health than drawbacks. Contraception can contribute higher child survival rates by allowing a longer gap between 2 pregnancies for women. There is very high risk if early death when difference between children was less than 2 years ,as high as 60%. Low birthweight is also a result of premature pregnancy.

[2] For the study, 590 women in total were recruited. 50.2% of the women used contraception overall, with traditional methods being utilised by 19.5% and contemporary methods by 30.7%. The more effective methods of contraception were being used and preferred by more women than in the past. Education level and family planning discussions during prenatal care were significant predictors of current contraceptive usage. Effective family planning and encouraging marital talk about contraception has contributed to an increase in the use of contraception after birth.

[3] The effectiveness and consistency of adult contraception were examined using logistic regression to examine the relationships between teenage attitudes towards contraception, confidence in one's knowledge of contraceptives and knowledge of condoms and reproduction. More accurate condom knowledge increased the likelihood of using more effective means of

contraception as opposed to less effective or no methods of contraception in adulthood. Conclusions were that knowledge of condoms and reproduction one acquires in adolescence are predictive of the use of contraception in maturity. Sex education proved healthier in the long run.

[4] Just 42% of women who are currently married and aware of contraception actually use family planning. Various estimates of fertility, infant and child mortality etc. at the national and state levels are provided by the survey, which is carried under Indian Ministry of Health and Family Welfare. The survey was conducted by IIPS throughout India and consulting businesses.

[5] The intentional termination of pregnancy via the use of many methods, including contraceptive methods, sexual practices, chemicals, drugs, and surgical procedures is very important to know. Effective contraception gives a flexibility to have children when women want to and allows them to interact physically without fear of becoming pregnant unintentionally. The objective is to give contraception in the most private and comfortable setting possible at the lowest expense. Sterilization, especially of females, predominates in all current practices. Yet, as teenagers still have years of their reproductive life ahead of them, they require reversible and non-invasive therapies to postpone or spread out pregnancies.

## VII. METHODS OF DATA ANALYSIS (CONTRACEPTION AWARENESS AND CHILD HEALTH)

In the initial stages, a few interesting relationships were discovered (shown in the scatter plots below) like that of household sanitation facilities, literacy/education rates, found to be correlated with institutional births, and a number of other aspects of women's and children's health.
Regression Analysis, Clustering Analysis, and Association Rule Mining were employed to determine the relationship between contraception awareness and child health with different factors.

*A. Regression Analysis*

On the basis of correlation analysis and literature review, variables that were pertinent to contraception awareness and child health were chosen. First, a complete regression model using all 12 variables was run, and the p-value for each characteristic was examined to identify risk factors for child health and anaemia. Then, a new regression model was created using only variables with p-values lower than 0.05. The remaining factors were eliminated. This was carried out until all final variables were determined to be significant (p-value 0.05). The regression model was created, and p-values were discovered using the Python stat models package.

*B. Clustering Analysis*

To explore how they affected contraception, attributes (dimensions) were individually picked, and *K-Means Clustering* algorithm was implemented and run along the picked dimension(s) to cluster the data points. The Z-Score normalised data was fed to the clustering algorithm.
Firstly we calculated the Value of K using the Elbow method, which gives an appropriate range to choose the number of clusters from.

*C. Association Rule Mining*

*Apriori algorithm* was implemented to mine association rules about how *contraception awareness and child health* predicts other factors related to the health of women and children. The thresholds for support and confidence were x and t, respectively.

## VIII. RESULTS (CONTRACEPTION AWARENESS)

*A. Regression*

A multivariate regression analysis was performed to examine ten attributes, and the p-values for each attribute were analysed. In the regression analysis, four attributes were excluded from the model due to their p-values exceeding the predetermined threshold of 0.05. A novel regression model has been developed for the remaining attributes. The study found that all six attributes were significant in the new model with a p-value of less than 0.05. Therefore, they were not removed. In this study, we

analysed the regression model coefficients to investigate their impact on Contraception Awareness.

A positive correlation was observed between the following attributes and increased Total Children Born

- Contraceptive Method Used

Negative Correlation was found with

- Wealth Index
- Highest Education Level
- Family Planning Awareness

The positive correlation of children born with contraception in the context of current dataset translates to less usage of traditional contraceptive measures like pills and condoms as these have been set on lower values. Also, lower education levels, wealth and family planning awareness lead to the birth of more children.

A similar analysis was done keeping the correlation with respect to wealth, and then 3 columns were dropped in the first iteration, and the rest 7 had a p-value less than 0.5.

Attributes having a positive correlation with wealth were:

- Highest Education Level
- Heard about Family Planning
- Age of respondent at first birth
- Husband/partners education level

Negative correlation

- Type of residence (rural)
- Total children born

From the above analysis it can be inferred that more wealth leads to higher education levels, more age at first birth and higher awareness about family planning whereas lesser wealth leads to more children in the household and a rural household.

It is suspected that the results obtained may be biased due to a high degree of multicollinearity in the independent variables. This is a common issue in multivariate regression. In this study, we further

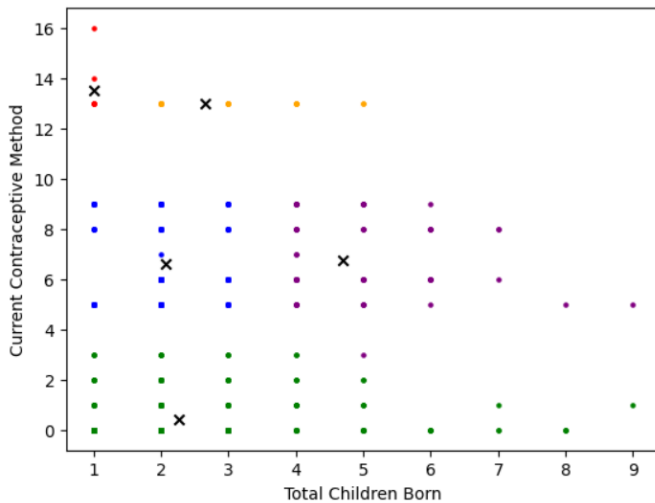investigated the impact of these variables by employing clustering analysis and association rule mining techniques.

*B. Clustering*

Clustering on the basis of "Total Children Born" and "Current Contraceptive Method" gave us 5 clusters with centroids C1 = [1.0, 13.5], C2 = [2.0646, 6.6007], C3 = [2.2631, 0.4270], C4 = [2.6571, 13.0], C5 = [4.6930, 6.7425].

Cluster 1 has 0.59% has data where 1 child is born and emergency contraception or female condoms being used, cluster 2 has 40% of data with upto 3 children born with sterilization or abstinence being used, cluster 3 has 49.2% of data where 0-9 children are born and the contraceptive used is either none or medications eg. pills, cluster 4 has 2.59% has data where 2-5 children are born and the contraceptive used is Lactational amenorrhea, and cluster 5 has 7.48% of data where 4-9 children are born and the method used is sterilization, male condom or withdrawl.
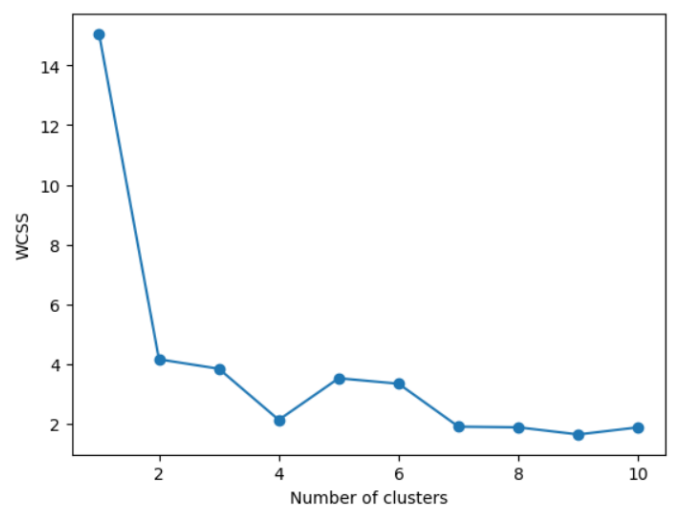


The plot above shows the elbow method that is used to find the value of K for k means algorithm . The optimal value is 5 clusters

*A.   Regression*

Doing the analysis with respect to drinking alcohol firstly 11 attributes were taken. After the first iteration as per a p-value of less than 0.05, 5 attributes made it to the next iteration. The F-statistic value 8.625 was observed and it's probability (4.68e-08) The positive correlation was found between

- Smokes
- Age of respondent at first birth

Negative correlation was found as

- highest education level
- wealth index
- Anaemia level

We can infer that the people who don't drink alcohol also don't smoke and have a higher age at first birth which is considered a safer practice. Whereas, more alcohol consumption is prevalent in less wealthy households , less educated households and also leads to more anemia in children.

Doing analysis with respect to Heard about family planning we took p value as 0.12. This resulted in a final group of 6 attributes with a F statistic value 27.15, Probability of the F statistic 1.13e-30. The positive attributes were

- Place of residence
- Highest Education level
- wealth index combine

Negative correlation

- Total Children Born
- Size of child at birth
- Ever had a terminated pregnancy

From the above analysis we can conclude that family planning awareness is given more importance as the education level increases, and wealth index increases. On the other hand as family planning awareness decreases it leads to more
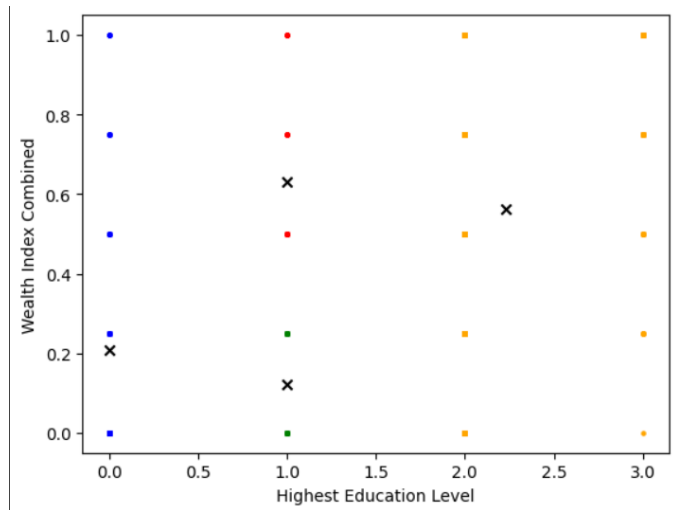
*C. Association*

The association rules were obtained in the format of the format: (antecedent)—>(consequent) (confidence). The following rules were obtained:

1) ('Reason Of Last Discontinuation (wanted to become pregnant)')', 'Highest Education Level (Secondary') -> ('Husbands Education Level (secondary)') Confidence is: 0.745152354570637
2) ('Total Children Born (1)) -> ('Heard about family planning (Yes)') Confidence is: 0.5946843853820598
3) ('Total Children Born (3)',) -> ('Heard about family planning (No)',) Confidence is: 0.4913494809688581
4) ('Current Contraceptive Method (Female Sterilization)',) -> ('Place Of Residence (Rural)',) Confidence is: 0.8666666666666666
5) ('Wealth Index Combined (poorest)',) -> ('Heard about family planning (No)',) Confidence is: 0.7034482758620689
6) ('Husbands Education Level (Higher)') -> ('Heard about family planning (yes)',) Confidence is: 0.6666666666666667

children in the household and more chances of terminated pregnancy.

### B. Clustering

Clustering on the basis of "Highest Education Level" and "Wealth Index combined" gave us 4 clusters with centroids C1 (medium education level) = [1.0, 0.6298], C2 (low education level) = [0.0, 0.2071], C3 (medium education level) = [1.0, 0.1206], C4 (high education level) = [2.2318, 0.5623].

Cluster 1 has 3.58% of people with low wealth index and medium education, cluster 2 has 16% of people with low education level but variable wealth index, cluster 3 has 8.59% of people with medium education level and medium to high wealth index and cluster 4 has 71.55% of people with high education level and low to high variable wealth index.

The optimal value found from elbow method is 4 clusters.
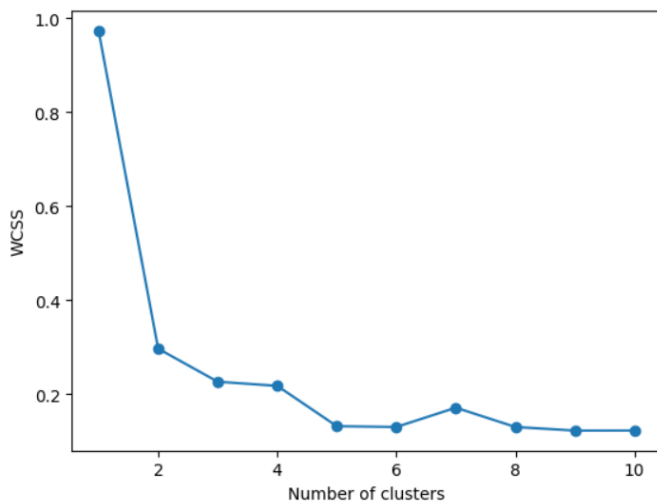




### C. Association

1) ('Ever had a terminated pregnancy (No)', 'Size of Child at Birth (average)') -> ('Smokes (no)')Confidence is: 0.9965675057208238

2) ('Drink Alcohol (no)',) -> ('Smokes (no)',)Confidence is: 0.9977151561309977

3) ('Heard about family planning (yes),) -> ('Drink Alcohol (no)', 'Smokes (no)')Confidence is: 0.9823609226594301

4) ('Highest Education Level (secondary),) -> ('Drink Alcohol (no)', 'Smokes (no)')Confidence is: 0.9797843665768194

5) ('Anemia Level (not anaemic)',) -> ('Drink Alcohol (no)', 'Smokes (no)')Confidence is: 0.972027972027972

### X. CONCLUSION

According to the aforementioned analysis, it can be concluded that there is a correlation between increased education levels of both partners and greater awareness of contraception. The increased awareness regarding family planning can be attributed to the government's efforts in implementing family planning awareness programmes. The most commonly utilised contraceptive methods are condoms and pills. The awareness of contraception has been found to have a positive impact on the birth weight of children, contributing to their overall health. The incidence of

childhood anaemia is significantly decreased when parents abstain from smoking and drinking. The relationship between birth weight and terminated pregnancies is direct. A decrease in the number of terminations results in a healthier birth weight for the child.

Akshat Johar - provided the code for Apriori and did visualization plots

Shaurya Garg - code for K means and report writing

Saksham Bansal - code for regression analysis and report writing

REFERENCES

[1] Cleland, J., Conde-Agudelo, A., Peterson, H., Ross, J., & Tsui, A. (2012). Contraception and health. The Lancet, 380(9837), 149-156

[2] Wuni, C., Turpin, C. A., & Dassah, E. T. (2018). Determinants of contraceptive use and future contraceptive intentions of women attending child welfare clinics in urban Ghana. BMC public health, 18(1), 1-8.

[3] Guzzo, K. B., & Hayford, S. R. (2018). Adolescent reproductive and contraceptive knowledge and attitudes and adult contraceptive behavior. Maternal and child health journal, 22, 32-40.

[4] Ramesh, B. M., Gulati, S. C., & Retherford, R. D. (1996). Contraceptive use in India, 1992-93.

[5] Jain, R., & Muralidhar, S. (2011). Contraceptive methods: needs, options and utilization. The journal of obstetrics and gynecology of India, 61, 626-634.