

ABCD Behaviour Data Hackathon

Team : The Misfit (A lone wolf)

Introduction

I am thrilled to be a part of the 1st Edition of the Behavior Data Hackathon, and would personally thank Atlas of Behavior Change in Development (ABCD) which is in collaboration with the Centre for Social and Behavior Change (CSBC). Also I would love to personally thank my mentor Rounak Banik for the guidance and Vrinda Agarwal for the chance to work on this project. I admit I am not a statistics or a sociology wizard but I would try my best to understand the vast dataset that I have been presented and gather insights and develop models out of it that would address the problem I have been presented with.

Problem Statement

Context: 59% of women had at least four Ante Natal Care (ANC) visits during their last pregnancy. However, only 52% of women are allowed to go to the health facility on their own. Further, 21% of men didn't attend ANC visits because they did not feel it was necessary, nor did they allow their partner to go.

Question: Using data from the NFHS-5 dataset, identify the effect that women participation and partner alignment on ANC visits has on overall health outcomes of the mother and the child post-delivery.

Data Analysis and Tools Deployed

Data Cleaning and Preparation

The dataset we received was too huge to process, in fact leading to frequent computer and program crashes. Later when we were able to run it on excel we were met with numerous variables that required careful consideration. A lot of our initial efforts was dedicated to thoroughly selecting the columns most important to our analysis. Out of an original set of 137 columns, we identified and retained 50 that were highly relevant to our objectives.

The initial stages of data formatting and cleaning were conducted using Excel. This helped us address some immediate issues and prepare the data for more intensive processing. However, the bulk of our data preparation was carried out in Python. The powerful libraries like pandas, NumPy etc were used for more sophisticated cleaning and transformation tasks.

One of the initial challenges we encountered was dealing with symbols such as * and %, which were prevalent in the dataset and posed significant barriers to effective analysis. These symbols were initially replaced with blank values, which were then converted to `np.nan` to standardize the handling of missing data. Subsequently, these `np.nan` values were imputed with mean values to ensure robust statistical analysis. After thorough consideration, we determined that outlier removal was unnecessary for our specific analysis and chose to omit this step. This decision was based on the nature of our data and the objectives of our study.

Exploratory Data Analysis (EDA)

In the initial exploration of the dataset that is without any statistical tests, following insights were gathered from the dataset. Most of these indicators show that things in India are changing for the better as a lot of work has been done by the government to address several key issues like child mortality, female foeticide etc. However a lot of things like high teenage fertility rate, a lot of regions not having electricity, sanitation etc shows that India is still a third world country. Here are the following insights:

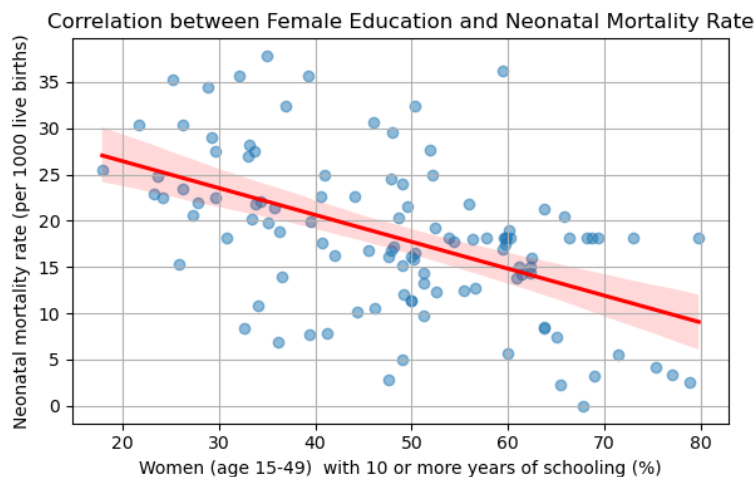
- **Surveyed Households:** A whopping amount 6.36 lakh households were interviewed from both rural and urban parts of various states and UTs of India thus giving an extremely huge sample size.
- **Gender wise Interviews:** A total of 7.25 lakh women and 1.1 lakh men between the age of 15 to 54 were interviewed. However, it shows that the men are just 15% of the sample size which shows the gender disparity in the survey. Also, it shows men are mostly insensitive or indifferent to issues suffered by women.
- **Population Below Age 15:** About a quarter of the population is below 15 years, with a range from 18.1% to 39.2% in the regions which shows that there is a significant young population in the country.
- **Amenities:** 98% of households have electricity, 95% of households have access to improved drinking water and almost 80% of households use improved sanitation. However these numbers are only for half of the regions which shows the significant urban rural divide in this country.
- **Adolescent Fertility Rate:** The mean adolescent fertility rate is 32.33 births per 1,000 women aged 15-19 years which is alarming as the legal age for marriage is 21. Being a mother at a young age also increases the risk of child mortality.
- **Sex Ratio:** The average sex ratio is 1,008 females per 1,000 males which is a good sign as India had a poor sex ratio. Beti bachao campaign by the government has worked wonders.
- **High Antenatal Check-up Rates:** On average, 93.29% of mothers had an antenatal check-up in the first trimester which shows significant awareness among women and the efforts of government bearing fruit.
- **Variable Adherence to Multiple Visits:** About 77.02% of mothers had at least four antenatal care visits, showing considerable awareness.
- **Nearly Universal Iron Folic Acid Use:** Around 98% of mothers consumed iron folic acid for 100 days or more during pregnancy which shows that they are aware of anaemia, low birth weight etc.
- **Widespread MCP Card Use:** Mothers received a Mother and Child Protection (MCP) card for nearly 95% of registered pregnancies, highlighting widespread usage.
- **Moderate Unmet Family Planning Needs:** The average total unmet need for family planning among married women aged 15-49 years was 10.07%, indicating some gaps in contraceptive coverage.
- **Early Marriage Concerns:** A substantial proportion of women (52.60%) married before age 18 years, highlighting ongoing challenges in addressing early marriage practices.
- **Variable Use of Modern Contraceptive Methods:** While female sterilization is prevalent with an average of 2 women out of 3 getting sterilized, a bit more than half of the couple using condoms and around one third using injectables. It demonstrates that there is a significant awareness of population control methods among couples.
- **High Literacy Among Women:** The mean percentage of women aged 15-49 who are literate is approximately 93%, indicating widespread literacy among women in this age group. This suggests a positive trend towards education and empowerment among women. The Beti Padhao campaign seems to have been working!

Insights and Context

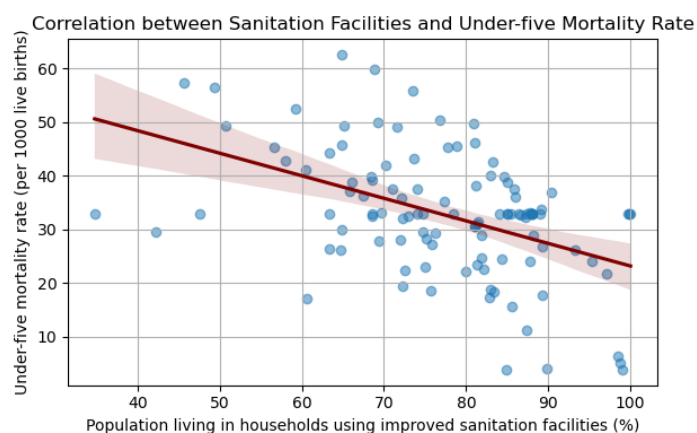
Key Findings with Interpretation

A very high negative correlation between Education vs. Neonatal Mortality was observed and it had an insignificant p-value which means the correlation was significant.

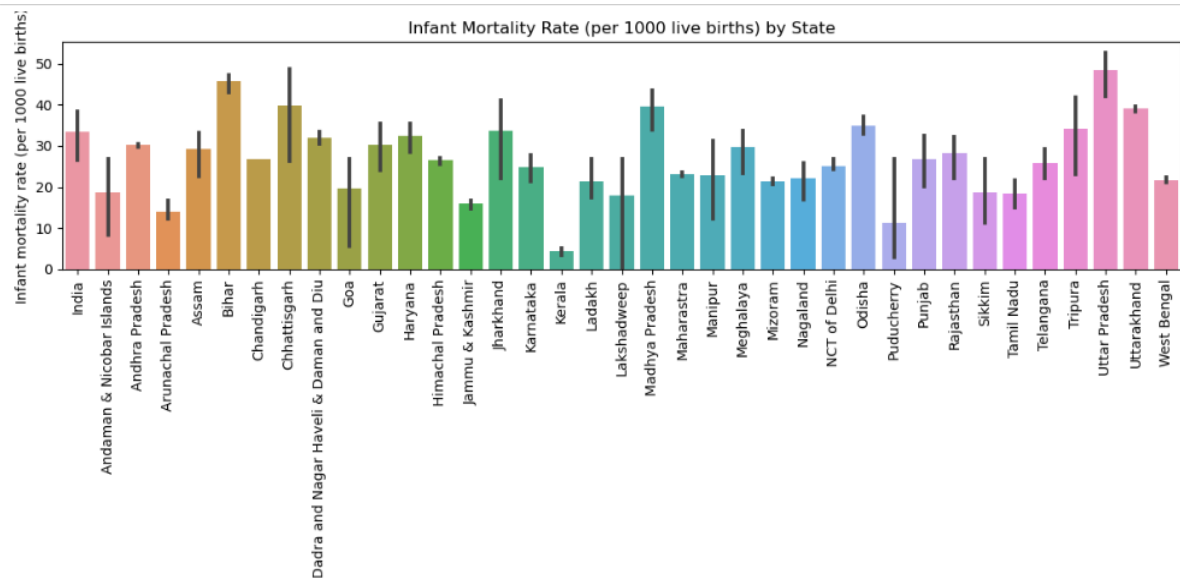
We can interpret this as the power of education as it was observed here that the states/area where more percentage of women are educated, the child mortality rate was lower.



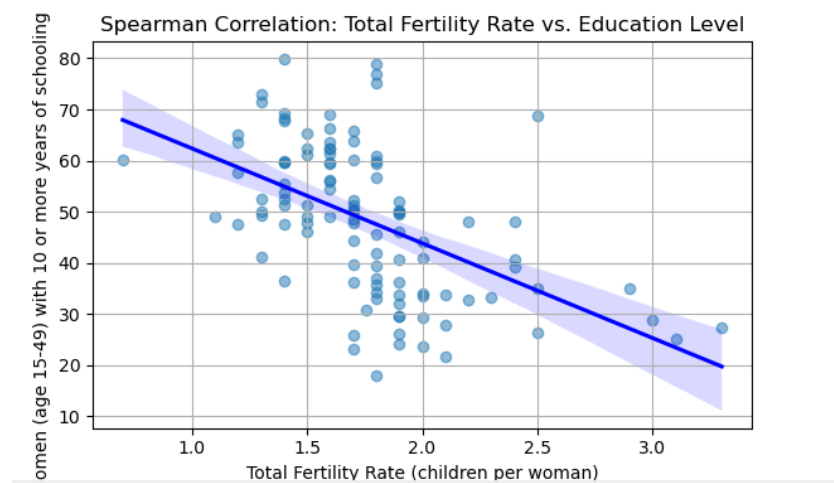
A very high negative correlation has been seen with the percentage of households who have improved sanitary conditions and infant mortality rate. Also the correlation component was very negative and had a low p-value indicating significance of the test. Hence we can interpret that children are more likely to die in their early childhood due to diseases spread by poor sanitary conditions.



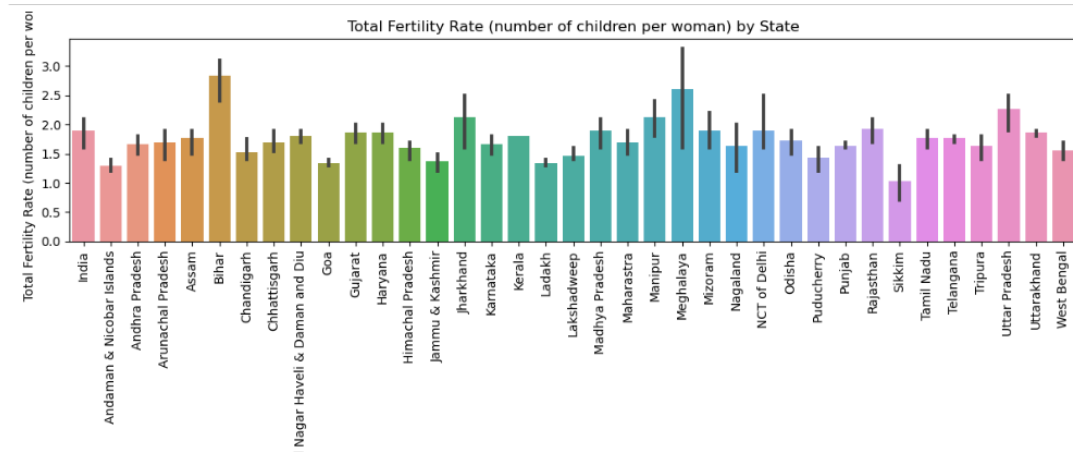
It has been observed with a low p value of course that mean infant mortality rates across states are varied with the states in the BIMARU belt that is Bihar, Madhya Pradesh, Uttar Pradesh and Rajasthan (Also Chhattisgarh and Jharkhand which is extension of BIMARU belt) displaying high infant mortality rates than the average for India.



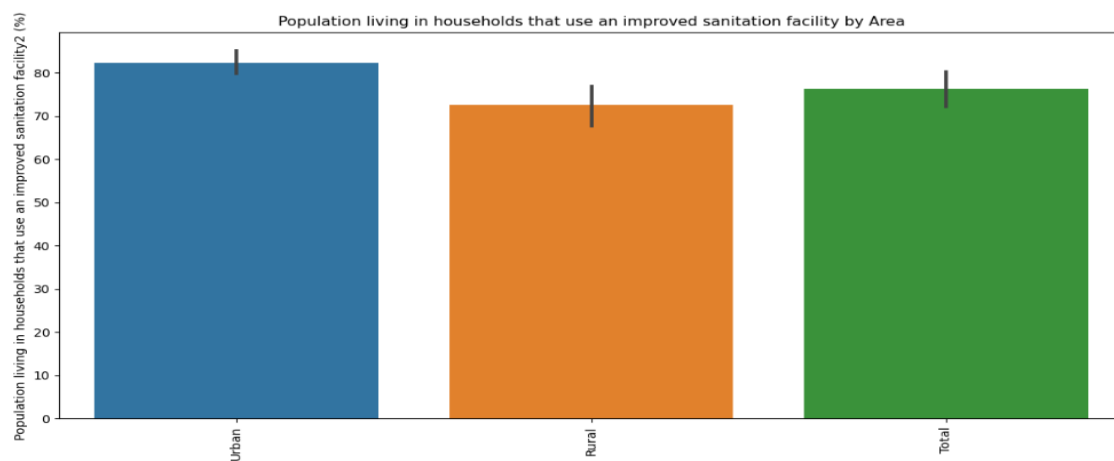
It has been observed with a low p value of course that women with low TFR have high literacy mostly and of course the test has a low value of p to conclude it is valid. So, a very high negative correlation has been observed with the percentage of women who have 10+ years of schooling and the TFR (That is kids per woman). That means highly educated women will prefer having less babies.



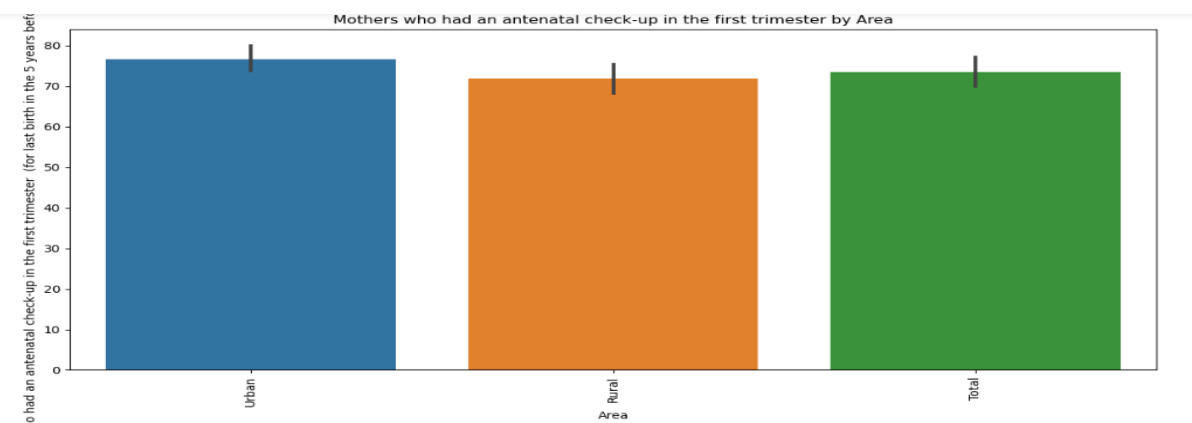
It has been observed with a low p value of course that Fertility rates across states are varied with the states in the BIMARU belt that is Bihar, Madhya Pradesh, Uttar Pradesh and Rajasthan (Also Chhattisgarh and Jharkhand which is extension of BIMARU belt) displaying high total fertility rates than the average for India. In addition to this list are Meghalaya which is the state with the highest birth rate in the north east while Sikkim displays the lowest.



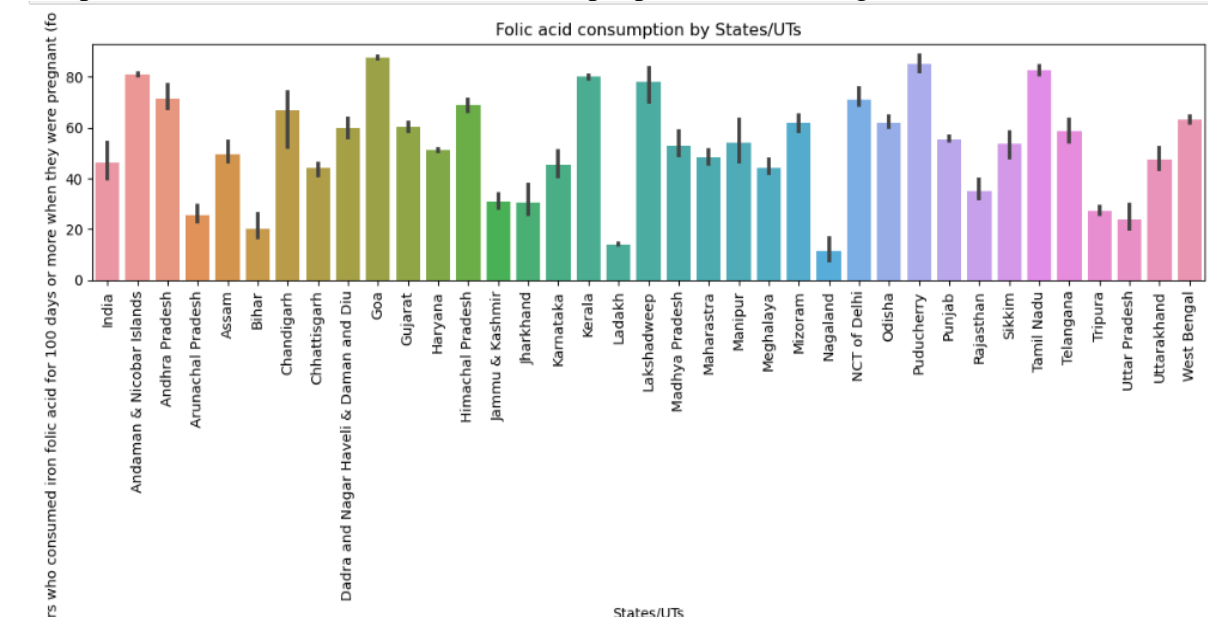
Our hypothesis of the urban rural divide in India was also confirmed as we came to know that this is not a fiction but a harsh reality of this country where resources are not divided equitably. With a very low p-score of less than 5%, we came to know that for instance that people in urban areas are more likely to live with better sanitation than their rural counterparts which is lower than the average.



Now this brings us to other point of discussion that whether the distribution of medical infrastructure is unfair as well. We did a hypothesis test which gave us a low value of p and we could not be happier. It made us conclude that women in the rural and urban area of almost all the states have the same percentage of visits for ANC. It shows that government has done significant work to improve accessibility and awareness.



Now we have talked at length about the urban rural divide, let us come to the divide in the varied regions of India namely the states and the union territories. Here as the p value is really low, we can conclude that the test is significant and we can say that there is disparity among the states in the regard of pregnant women consuming folic acid. The BIMARU belt that we spoke of earlier didn't disappoint (pun intended) as they are again in the bottom of this list however MP performed well but is far behind states/UTs like West Bengal, Lakshadweep, Andaman and Nicobar, Tamil Nadu so and so which top the list in the terms of folic acid consumption. We can attribute this to these areas being coastal and hence it has been proven that fish-based diet which these people consume is a great source of folic acid.



Data Analysis methodologies used

Data Cleaning and Preprocessing: Ensured data integrity by handling missing values and converting relevant variables into numeric formats for statistical analysis.

Exploratory Data Analysis (EDA): Examined data distributions, correlations, and trends to identify key variables influencing maternal and neonatal health outcomes.

Hypothesis Testing: Conducted ANOVA and Chi-square tests to explore relationships between categorical variables and health indicators.

Machine Learning Modeling: Employed Gradient Boosting Regression to predict neonatal mortality rates based on socio-demographic factors and maternal health behaviors.

Model Evaluation: Assessed model performance using metrics like R-squared to gauge predictive accuracy and interpretability.

Behavioral Insights

- Men's participation in the survey was notably low at just 15%, revealing a concerning gender disparity in engagement with women's health issues.
- Adolescent fertility rates, averaging 32.33 births per 1,000 women aged 15-19, highlight persistent challenges in addressing early pregnancies and their associated health risks.
- Disparities between urban and rural areas, particularly in access to basic amenities like sanitation, underscore ongoing developmental gaps across different parts of India.
- Higher levels of female education correlate strongly with lower total fertility rates, indicating the pivotal role of education in shaping reproductive health choices.
- Regional differences in maternal health indicators suggest uneven distribution of healthcare resources and infrastructure, affecting health outcomes across states.
- Nearly universal adoption of iron folic acid during pregnancy indicates increasing awareness about maternal nutrition and health among expectant mothers.
- Varied usage rates of modern contraceptives emphasize the need for targeted initiatives to promote effective family planning practices.
- Positive impacts of government initiatives like Beti Bachao Beti Padhao are reflected in improved sex ratios and higher literacy rates among women.
- Early marriage practices contribute significantly to higher fertility rates, highlighting the urgency of addressing cultural norms that impact maternal health.
- Differences in folic acid consumption between coastal and inland regions underscore the influence of regional dietary habits on maternal health outcomes.

Conclusion

Based on the regression analysis conducted, the model achieved an impressive R-squared score of 0.90, indicating that 90% of the variability in neonatal mortality rates can be explained by the selected independent variables. Among these, important factors influencing neonatal health included early marriage practices among men, total fertility rates, and maternal consumption of iron folic acid for at least 100 days.

These findings display the critical impact of socio-demographic factors and maternal health behaviours on neonatal outcomes. We would recommend delaying early marriages and improving maternal health through proper nutrition and prenatal care emerge as pivotal strategies to potentially reduce neonatal mortality rates.

The analysis highlights the importance of targeted health interventions aimed at promoting maternal health education, improving access to family planning methods, and fostering awareness among men about the repercussions of early marriages on neonatal health.

