

# ABCD NFHS Project

In this file we are going to do some exploratory data analysis on the dataset that will answer the question of the second Problem statement. We will try to conduct a thorough exploratory data analysis (EDA) on the NFHS-5 dataset to investigate the impact of women's participation and partner alignment on ANC visits on overall health outcomes for mothers and their babies post-delivery. This analysis aims to uncover patterns and insights related to maternal and child health by examining factors

```
In [29]: #getting the dataset
import pandas as pd
nfhs = pd.read_csv("C:\\Users\\sujoydutta\\Desktop\\DA material\\Python\\Projects\\ABCD N
nfhs.head()
```

```
Out[29]:
```

	States/UTs	Area	Number of Households surveyed	Number of Women age 15-49 years interviewed	Number of Men age 15-54 years interviewed	Female population age 6 years and above who ever attended school (%)	Population below age 15 years (%)	Sex ratio of the total population (females per 1,000 males)	Sex ratio at birth for children born in the last five years (females per 1,000 males)	Children under age 5 years whose birth was registered with the civil authority (%)
0	India	Urban	160138	179535	26420	82.5	23.1	985.0	924	
1	India	Rural	476561	544580	75419	66.8	28.1	1037.0	931	
2	India	Total	636699	724115	101839	71.8	26.5	1020.0	929	
3	Andaman & Nicobar Islands	Urban	527	557	85	86.5	22.7	1023.0	941	
4	Andaman & Nicobar Islands	Rural	2097	1840	282	81.8	19.7	929.0	891	

5 rows × 137 columns

```
In [30]: # Convert column names to a list and print them
columns = nfhs.columns.tolist()
print(columns)
```

```
['States/UTs', 'Area', 'Number of Households surveyed', 'Number of Women age 15-49 years interviewed', 'Number of Men age 15-54 years interviewed', 'Female population age 6 years and above who ever attended school (%)', 'Population below age 15 years (%)', 'Sex ratio of the total population (females per 1,000 males)', 'Sex ratio at birth for children born in the last five years (females per 1,000 males)', 'Children under age 5 years whose birth was registered with the civil authority (%)', 'Deaths in the last 3 years registered with the civil authority (%)', 'Population living in households with electricity (%)', 'Population living in households with an improved drinking-water source1 (%)', 'Population living in households that use an improved sanitation facility2 (%)', 'Households using clean fuel for cooking3 (%)', 'Households using iodized salt (%)', 'Households with any usual member covered under a health insurance/financing scheme (%)', 'Children age 5 years who attended pre-primary school during the school year 2019-20 (%)', 'Women
```

(age 15-49) who are literate4 (%)', 'Men (age 15-49) who are literate4 (%)', 'Women (age 15-49) with 10 or more years of schooling (%)', 'Men (age 15-49) with 10 or more years of schooling (%)', 'Women (age 15-49) who have ever used the internet (%)', 'Men (age 15-49) who have ever used the internet (%)', 'Women age 20-24 years married before age 18 years (%)', 'Men age 25-29 years married before age 21 years (%)', 'Total Fertility Rate (number of children per woman)', 'Women age 15-19 years who were already mothers or pregnant at the time of the survey (%)', 'Adolescent fertility rate for women age 15-19 years5', 'Neonatal mortality rate (per 1000 live births)', 'Infant mortality rate (per 1000 live births)', 'Under-five mortality rate (per 1000 live births)', 'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Any method6 (%)', 'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Any modern method6 (%)', 'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Female sterilization (%)', 'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Male sterilization (%)', 'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - IUD/PPIUD (%)', 'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Pill (%)', 'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Condom (%)', 'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Injectables (%)', 'Total Unmet need for Family Planning (Currently Married Women Age 15-49 years)7 (%)', 'Unmet need for spacing (Currently Married Women Age 15-49 years)7 (%)', 'Health worker ever talked to female non-users about family planning (%)', 'Current users ever told about side effects of current method of family planning8 (%)', 'Mothers who had an antenatal check-up in the first trimester (for last birth in the 5 years before the survey) (%)', 'Mothers who had at least 4 antenatal care visits (for last birth in the 5 years before the survey) (%)', 'Mothers whose last birth was protected against neonatal tetanus (for last birth in the 5 years before the survey)9 (%)', 'Mothers who consumed iron folic acid for 100 days or more when they were pregnant (for last birth in the 5 years before the survey) (%)', 'Mothers who consumed iron folic acid for 180 days or more when they were pregnant (for last birth in the 5 years before the survey) (%)', 'Registered pregnancies for which the mother received a Mother and Child Protection (MCP) card (for last birth in the 5 years before the survey) (%)', 'Mothers who received postnatal care from a doctor/nurse/LHV/ANM/midwife/other health personnel within 2 days of delivery (for last birth in the 5 years before the survey) (%)', 'Average out-of-pocket expenditure per delivery in a public health facility (for last birth in the 5 years before the survey) (Rs.)', 'Children born at home who were taken to a health facility for a check-up within 24 hours of birth (for last birth in the 5 years before the survey) (%)', 'Children who received postnatal care from a doctor/nurse/LHV/ANM/midwife/other health personnel within 2 days of delivery (for last birth in the 5 years before the survey) (%)', 'Institutional births (in the 5 years before the survey) (%)', 'Institutional births in public facility (in the 5 years before the survey) (%)', 'Home births that were conducted by skilled health personnel (in the 5 years before the survey)10 (%)', 'Births attended by skilled health personnel (in the 5 years before the survey)10 (%)', 'Births delivered by caesarean section (in the 5 years before the survey) (%)', 'Births in a private health facility that were delivered by caesarean section (in the 5 years before the survey) (%)', 'Births in a public health facility that were delivered by caesarean section (in the 5 years before the survey) (%)', 'Children age 12-23 months fully vaccinated based on information from either vaccination card or mother's recall11 (%)', 'Children age 12-23 months fully vaccinated based on information from vaccination card only12 (%)', 'Children age 12-23 months who have received BCG (%)', 'Children age 12-23 months who have received 3 doses of polio vaccine13 (%)', 'Children age 12-23 months who have received 3 doses of penta or DPT vaccine (%)', 'Children age 12-23 months who have received the first dose of measles-containing vaccine (MCV) (%)', 'Children age 24-35 months who have received a second dose of measles-containing vaccine (MCV) (%)', 'Children age 12-23 months who have received 3 doses of rotavirus vaccine14 (%)', 'Children age 12-23 months who have received 3 doses of penta or hepatitis B vaccine (%)', 'Children age 9-35 months who received a vitamin A dose in the last 6 months (%)', 'Children age 12-23 months who received most of their vaccinations in a public health facility (%)', 'Children age 12-23 months who received most of their vaccinations in a private health facility (%)', 'Prevalence of diarrhoea in the 2 weeks preceding the survey (Children under age 5 years) (%)', 'Children with diarrhoea in the 2 weeks preceding the survey who received oral rehydration salts (ORS) (Children under age 5 years) (%)', 'Children with diarrhoea in the 2 weeks preceding the survey who received zinc (Children under age 5 years) (%)', 'Children with diarrhoea in the 2 weeks preceding the survey taken to a health facility or health provider (Children under age 5 years) (%)', 'Children Prevalence of symptoms of acute respiratory infection (ARI) in the 2 weeks preceding the survey (Children under age 5 years) (%)', 'Children with fever or symptoms of ARI in the 2 weeks preceding

g the survey taken to a health facility or health provider (Children under age 5 years) (%)', 'Children under age 3 years breastfed within one hour of birth15 (%)', 'Children under age 6 months exclusively breastfed16 (%)', 'Children age 6-8 months receiving solid or semi-solid food and breastmilk16 (%)', 'Breastfeeding children age 6-23 months receiving an adequate diet16, 17 (%)', 'Non-breastfeeding children age 6-23 months receiving an adequate diet16, 17 (%)', 'Total children age 6-23 months receiving an adequate diet16, 17 (%)', 'Children under 5 years who are stunted (height-for-age)18 (%)', 'Children under 5 years who are wasted (weight-for-height)18 (%)', 'Children under 5 years who are severely wasted (weight-for-height)19 (%)', 'Children under 5 years who are underweight (weight-for-age)18 (%)', 'Children under 5 years who are overweight (weight-for-height)20 (%)', 'Women (age 15-49 years) whose Body Mass Index (BMI) is below normal (BMI <18.5 kg/m2)21 (%)', 'Men (age 15-49 years) whose Body Mass Index (BMI) is below normal (BMI <18.5 kg/m2) (%)', 'Women (age 15-49 years) who are overweight or obese (BMI ≥25.0 kg/m2)21 (%)', 'Men (age 15-49 years) who are overweight or obese (BMI ≥25.0 kg/m2) (%)', 'Women (age 15-49 years) who have high risk waist-to-hip ratio (≥0.85) (%)', 'Men (age 15-49 years) who have high risk waist-to-hip ratio (≥0.90) (%)', 'Children age 6-59 months who are anaemic (<11.0 g/dl)22 (%)', 'Non-pregnant women age 15-49 years who are anaemic (<12.0 g/dl)22 (%)', 'Pregnant women age 15-49 years who are anaemic (<11.0 g/dl)22 (%)', 'All women age 15-49 years who are anaemic22 (%)', 'All women age 15-19 years who are anaemic22 (%)', 'Men age 15-49 years who are anaemic (<13.0 g/dl)22 (%)', 'Men age 15-19 years who are anaemic (<13.0 g/dl)22 (%)', 'Women age 15 years and above with high (141-160 mg/dl) Blood sugar level23 (%)', 'Women age 15 years and above with very high (>160 mg/dl) Blood sugar level23 (%)', 'Women age 15 years and above with high or very high (>140 mg/dl) Blood sugar level or taking medicine to control blood sugar level23 (%)', 'Men age 15 years and above with high (141-160 mg/dl) Blood sugar level23 (%)', 'Men age 15 years and above with very high (>160 mg/dl) Blood sugar level23 (%)', 'Men age 15 years and above with high or very high (>140 mg/dl) Blood sugar level or taking medicine to control blood sugar level23 (%)', 'Women age 15 years and above with Mildly elevated blood pressure (Systolic 140-159 mm of Hg and/or Diastolic 90-99 mm of Hg) (%)', 'Women age 15 years and above with Moderately or severely elevated blood pressure (Systolic ≥160 mm of Hg and/or Diastolic ≥100 mm of Hg) (%)', 'Women age 15 years and above with Elevated blood pressure (Systolic ≥140 mm of Hg and/or Diastolic ≥90 mm of Hg) or taking medicine to control blood pressure (%)', 'Men age 15 years and above with Mildly elevated blood pressure (Systolic 140-159 mm of Hg and/or Diastolic 90-99 mm of Hg) (%)', 'Men age 15 years and above with Moderately or severely elevated blood pressure (Systolic ≥160 mm of Hg and/or Diastolic ≥100 mm of Hg) (%)', 'Men age 15 years and above with Elevated blood pressure (Systolic ≥140 mm of Hg and/or Diastolic ≥90 mm of Hg) or taking medicine to control blood pressure (%)', 'Women (age 30-49 years) Ever undergone a screening test for cervical cancer (%)', 'Women (age 30-49 years) Ever undergone a breast examination for breast cancer (%)', 'Women (age 30-49 years) Ever undergone an oral cavity examination for oral cancer (%)', 'Men (age 30-49 years) Ever undergone an oral cavity examination for oral cancer (%)', 'Women (age 15-49 years) who have comprehensive knowledge24 of HIV/AIDS (%)', 'Men (age 15-49 years) who have comprehensive knowledge24 of HIV/AIDS (%)', 'Women (age 15-49 years) who know that consistent condom use can reduce the chance of getting HIV/AIDS (%)', 'Men (age 15-49 years) who know that consistent condom use can reduce the chance of getting HIV/AIDS (%)', 'Currently married women (age 15-49 years) who usually participate in three household decisions25 (%)', 'Women (age 15-49 years) who worked in the last 12 months and were paid in cash (%)', 'Women (age 15-49 years) owning a house and/or land (alone or jointly with others) (%)', 'Women (age 15-49 years) having a bank or savings account that they themselves use (%)', 'Women (age 15-49 years) having a mobile phone that they themselves use (%)', 'Women age 15-24 years who use hygienic methods of protection during their menstrual period26 (%)', 'Ever-married women age 18-49 years who have ever experienced spousal violence27 (%)', 'Ever-married women age 18-49 years who have experienced physical violence during any pregnancy (%)', 'Young women age 18-29 years who experienced sexual violence by age 18 (%)', 'Women age 15 years and above who use any kind of tobacco (%)', 'Men age 15 years and above who use any kind of tobacco (%)', 'Women age 15 years and above who consume alcohol (%)', 'Men age 15 years and above who consume alcohol (%)', 'Unnamed: 136']

In [33]: # Getting the relevant columns

```
selected_columns = [
    'States/UTs', 'Area', 'Number of Households surveyed', 'Number of Women age 15-49 years',
    'Number of Men age 15-54 years interviewed', 'Female population age 6 years and above',
    'Population below age 15 years (%)', 'Sex ratio of the total population (females per
```

```

'Sex ratio at birth for children born in the last five years (females per 1,000 male
'Children under age 5 years whose birth was registered with the civil authority (%)'
'Deaths in the last 3 years registered with the civil authority (%)', 'Population li
'Population living in households with an improved drinking-water source1 (%)',
'Population living in households that use an improved sanitation facility2 (%)', 'Ho
'Households using iodized salt (%)', 'Households with any usual member covered under
'Children age 5 years who attended pre-primary school during the school year 2019-20
'Women (age 15-49) who are literate4 (%)', 'Men (age 15-49) who are literate4 (%)',
'Women (age 15-49) with 10 or more years of schooling (%)', 'Men (age 15-49) with 10
'Women (age 15-49) who have ever used the internet (%)', 'Men (age 15-49) who have e
'Women age 20-24 years married before age 18 years (%)', 'Men age 25-29 years marrie
'Total Fertility Rate (number of children per woman)', 'Women age 15-19 years who we
'Adolescent fertility rate for women age 15-19 years5', 'Neonatal mortality rate (pe
'Infant mortality rate (per 1000 live births)', 'Under-five mortality rate (per 1000
'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) -
'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) -
'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) -
'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) -
'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) -
'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) -
'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) -
'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) -
'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) -
'Total Unmet need for Family Planning (Currently Married Women Age 15-49 years)7 (%)
'Unmet need for spacing (Currently Married Women Age 15-49 years)7 (%)',
'Health worker ever talked to female non-users about family planning (%)',
'Current users ever told about side effects of current method of family planning8 (%)
'Mothers who had an antenatal check-up in the first trimester (for last birth in the
'Mothers who had at least 4 antenatal care visits (for last birth in the 5 years bef
'Mothers whose last birth was protected against neonatal tetanus (for last birth in
]

```

```

In [34]: # seeing which columns are present in the DataFrame
available_columns = [col for col in selected_columns if col in nfhs.columns]
available_columns

```

```

Out[34]: ['States/UTs',
'Area',
'Number of Households surveyed',
'Number of Women age 15-49 years interviewed',
'Number of Men age 15-54 years interviewed',
'Female population age 6 years and above who ever attended school (%)',
'Population below age 15 years (%)',
'Sex ratio at birth for children born in the last five years (females per 1,000 male
s)',
'Children under age 5 years whose birth was registered with the civil authority (%)',
'Deaths in the last 3 years registered with the civil authority (%)',
'Population living in households with electricity (%)',
'Population living in households with an improved drinking-water source1 (%)',
'Population living in households that use an improved sanitation facility2 (%)',
'Households using clean fuel for cooking3 (%)',
'Households using iodized salt (%)',
'Households with any usual member covered under a health insurance/financing scheme
(%)',
'Children age 5 years who attended pre-primary school during the school year 2019-20
(%)',
'Women (age 15-49) who are literate4 (%)',
'Men (age 15-49) who are literate4 (%)',
'Women age 20-24 years married before age 18 years (%)',
'Men age 25-29 years married before age 21 years (%)',
'Total Fertility Rate (number of children per woman)',
'Women age 15-19 years who were already mothers or pregnant at the time of the survey
(%)',
'Adolescent fertility rate for women age 15-19 years5',
'Neonatal mortality rate (per 1000 live births)',
'Infant mortality rate (per 1000 live births)',

```

```
'Under-five mortality rate (per 1000 live births)',
'Health worker ever talked to female non-users about family planning (%)',
'Current users ever told about side effects of current method of family planning8 (%)',
'Mothers whose last birth was protected against neonatal tetanus (for last birth in the
5 years before the survey)9 (%)']
```

```
In [35]: # Adding additional columns if less than 50 are available
if len(available_columns) < 50:
    additional_columns = [col for col in nfhs.columns if col not in selected_columns]
    available_columns.extend(additional_columns[:50 - len(available_columns)])
```

```
In [37]: # Selecting the columns from the DataFrame
nfhsnew = nfhs[available_columns]
nfhsnew
```

Out[37]:

	States/UTs	Area	Number of Households surveyed	Number of Women age 15-49 years interviewed	Number of Men age 15-54 years interviewed	Female population age 6 years and above who ever attended school (%)	Population below age 15 years (%)	Sex ratio at birth for children born in the last five years (females per 1,000 males)	Children under age 5 years whose birth was registered with the civil authority (%)	D
0	India	Urban	160138	179535	26420	82.5	23.1	924	93.3	
1	India	Rural	476561	544580	75419	66.8	28.1	931	87.5	
2	India	Total	636699	724115	101839	71.8	26.5	929	89.1	
3	Andaman & Nicobar Islands	Urban	527	557	85	86.5	22.7	941	96.9	
4	Andaman & Nicobar Islands	Rural	2097	1840	282	81.8	19.7	891	97.8	
...	...	...	...	...	...	...	...	...	...	
105	Uttarakhand	Urban	2358	2586	316	82.4	24.7	1094	92.2	
106	Uttarakhand	Rural	9811	10694	1270	72.0	27.1	937	91.8	
107	Uttarakhand	Total	12169	13280	1586	75.2	26.3	984	91.9	
108	West Bengal	Urban	5442	6358	942	84.1	20.2	921	98.1	
109	West Bengal	Rural	12745	15050	2079	73.3	25.1	993	98.3	

110 rows × 50 columns

```
In [38]: # Replacing the '*' with NaN
import numpy as np
nfhsnew.replace('*', np.nan, inplace=True)
nfhsnew.head()
```

```
C:\Users\sujoydutta\AppData\Local\Temp\ipykernel_10300\3489800645.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
nfhsnew.replace('*', np.nan, inplace=True)
```

Out[38]:

	States/UTs	Area	Number of Households surveyed	Number of Women age 15-49 years interviewed	Number of Men age 15-54 years interviewed	Female population age 6 years and above who ever attended school (%)	Population below age 15 years (%)	Sex ratio at birth for children born in the last five years (females per 1,000 males)	Children under age 5 years whose birth was registered with the civil authority (%)	Deaths in the last 3 years registered with the civil authority (%)
0	India	Urban	160138	179535	26420	82.5	23.1	924	93.3	
1	India	Rural	476561	544580	75419	66.8	28.1	931	87.5	
2	India	Total	636699	724115	101839	71.8	26.5	929	89.1	
3	Andaman & Nicobar Islands	Urban	527	557	85	86.5	22.7	941	96.9	(%)
4	Andaman & Nicobar Islands	Rural	2097	1840	282	81.8	19.7	891	97.8	

5 rows × 10 columns

```
In [40]: #viewing the selected columns
columns = nfhsnew.columns.tolist()
print(columns)
```

```
['States/UTs', 'Area', 'Number of Households surveyed', 'Number of Women age 15-49 years interviewed', 'Number of Men age 15-54 years interviewed', 'Female population age 6 years and above who ever attended school (%)', 'Population below age 15 years (%)', 'Sex ratio at birth for children born in the last five years (females per 1,000 males)', 'Children under age 5 years whose birth was registered with the civil authority (%)', 'Deaths in the last 3 years registered with the civil authority (%)', 'Population living in households with electricity (%)', 'Population living in households with an improved drinking-water source1 (%)', 'Population living in households that use an improved sanitation facility2 (%)', 'Households using clean fuel for cooking3 (%)', 'Households using iodized salt (%)', 'Households with any usual member covered under a health insurance/financing scheme (%)', 'Children age 5 years who attended pre-primary school during the school year 2019-20 (%)', 'Women (age 15-49) who are literate4 (%)', 'Men (age 15-49) who are literate4 (%)', 'Women age 20-24 years married before age 18 years (%)', 'Men age 25-29 years married before age 21 years (%)', 'Total Fertility Rate (number of children per woman)', 'Women age 15-19 years who were already mothers or pregnant at the time of the survey (%)', 'Adolescent fertility rate for women age 15-19 years5', 'Neonatal mortality rate (per 1000 live births)', 'Infant mortality rate (per 1000 live births)', 'Under-five mortality rate (per 1000 live births)', 'Health worker ever talked to female non-users about family planning (%)', 'Current users ever told about side effects of current method of
```

family planning8 (%)', 'Mothers whose last birth was protected against neonatal tetanus (for last birth in the 5 years before the survey)9 (%)', ' Sex ratio of the total population (females per 1,000 males)', 'Women (age 15-49) with 10 or more years of schooling (%)', 'Men (age 15-49) with 10 or more years of schooling (%)', 'Women (age 15-49) who have ever used the internet (%)', 'Men (age 15-49) who have ever used the internet (%)', 'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Any method6 (%)', 'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Any modern method6 (%)', 'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Female sterilization (%)', 'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Male sterilization (%)', 'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - IUD/PPIUD (%)', 'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Pill (%)', 'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Condom (%)', 'Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Injectables (%)', 'Total Unmet need for Family Planning (Currently Married Women Age 15-49 years)7 (%)', 'Unmet need for spacing (Currently Married Women Age 15-49 years)7 (%)', 'Mothers who had an antenatal check-up in the first trimester (for last birth in the 5 years before the survey) (%)', 'Mothers who had at least 4 antenatal care visits (for last birth in the 5 years before the survey) (%)', 'Mothers who consumed iron folic acid for 100 days or more when they were pregnant (for last birth in the 5 years before the survey) (%)', 'Mothers who consumed iron folic acid for 180 days or more when they were pregnant (for last birth in the 5 years before the survey) (%)', 'Registered pregnancies for which the mother received a Mother and Child Protection (MCP) card (for last birth in the 5 years before the survey) (%)']

```
In [76]: # Replacing NaN values with the median of each column
nfhsnew = nfhsnew.apply(lambda x: x.fillna(x.mean()) if x.dtype.kind in 'biufc' else x)
nfhsnew
```

Out[76]:

	States/UTs	Area	Number of Households surveyed	Number of Women age 15-49 years interviewed	Number of Men age 15-54 years interviewed	Female population age 6 years and above who ever attended school (%)	Population below age 15 years (%)	Sex ratio at birth for children born in the last five years (females per 1,000 males)	Children under age 5 years whose birth was registered with the civil authority (%)	Deaths registered at
0	India	Urban	160138	179535	26420	82.5	23.1	924	93.3	
1	India	Rural	476561	544580	75419	66.8	28.1	931	87.5	
2	India	Total	636699	724115	101839	71.8	26.5	929	89.1	
3	Andaman & Nicobar Islands	Urban	527	557	85	86.5	22.7	941	96.9	
4	Andaman & Nicobar Islands	Rural	2097	1840	282	81.8	19.7	891	97.8	
...	...	...	...	...	...	...	...	...	...	
105	Uttarakhand	Urban	2358	2586	316	82.4	24.7	1094	92.2	
106	Uttarakhand	Rural	9811	10694	1270	72.0	27.1	937	91.8	



107	Uttarakhand	Total	12169	13280	1586	75.2	26.3	984	91.9
108	West Bengal	Urban	5442	6358	942	84.1	20.2	921	98.1
109	West Bengal	Rural	12745	15050	2079	73.3	25.1	993	98.3

110 rows × 50 columns

In [77]:

```
#checking to see null values
nfhsnew.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 110 entries, 0 to 109
```

```
Data columns (total 50 columns):
```

```
#      Column
```

```
Non-Null Count  Dtype
```

```
---  -----
```

```
-----
```

```
0      States/UTs
```

```
110 non-null    object
```

```
1      Area
```

```
110 non-null    object
```

```
2      Number of Households surveyed
```

```
110 non-null    int64
```

```
3      Number of Women age 15-49 years interviewed
```

```
110 non-null    int64
```

```
4      Number of Men age 15-54 years interviewed
```

```
110 non-null    int64
```

```
5      Female population age 6 years and above who ever attended school (%)
```

```
110 non-null    object
```

```
6      Population below age 15 years (%)
```

```
110 non-null    float6
```

```
7      Sex ratio at birth for children born in the last five years (females per 1,000 male
s)
```

```
109 non-null    object
```

```
8      Children under age 5 years whose birth was registered with the civil authority (%)
```

```
109 non-null    object
```

```
9      Deaths in the last 3 years registered with the civil authority (%)
```

```
108 non-null    object
```

```
10     Population living in households with electricity (%)
```

```
110 non-null    float6
```

```
11     Population living in households with an improved drinking-water source1 (%)
```

```
110 non-null    float6
```

```
12     Population living in households that use an improved sanitation facility2 (%)
```

```
110 non-null    float6
```

```
13     Households using clean fuel for cooking3 (%)
```

```
109 non-null    object
```

```
14     Households using iodized salt (%)
```

```
109 non-null    object
```



15	Households with any usual member covered under a health insurance/financing scheme (%)	109 non-null	object
16	Children age 5 years who attended pre-primary school during the school year 2019-20 (%)	105 non-null	object
17	Women (age 15-49) who are literate4 (%)	110 non-null	object
18	Men (age 15-49) who are literate4 (%)	109 non-null	object
19	Women age 20-24 years married before age 18 years (%)	109 non-null	object
20	Men age 25-29 years married before age 21 years (%)	96 non-null	object
21	Total Fertility Rate (number of children per woman)	109 non-null	object
22	Women age 15-19 years who were already mothers or pregnant at the time of the survey (%)	109 non-null	object
23	Adolescent fertility rate for women age 15-19 years5	110 non-null	float64
24	Neonatal mortality rate (per 1000 live births)	97 non-null	object
25	Infant mortality rate (per 1000 live births)	97 non-null	object
26	Under-five mortality rate (per 1000 live births)	110 non-null	float64
27	Health worker ever talked to female non-users about family planning (%)	109 non-null	object
28	Current users ever told about side effects of current method of family planning (%)	108 non-null	object
29	Mothers whose last birth was protected against neonatal tetanus (for last birth in the 5 years before the survey)9 (%)	109 non-null	object
30	Sex ratio of the total population (females per 1,000 males)	110 non-null	float64
31	Women (age 15-49) with 10 or more years of schooling (%)	110 non-null	object
32	Men (age 15-49) with 10 or more years of schooling (%)	109 non-null	object
33	Women (age 15-49) who have ever used the internet (%)	109 non-null	object
34	Men (age 15-49) who have ever used the internet (%)	109 non-null	object
35	Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Any method6 (%)	109 non-null	object
36	Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Any modern method6 (%)	109 non-null	object
37	Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Female sterilization (%)	109 non-null	object
38	Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) -		

```

Male sterilization (%) 109 non-null object
39 Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - IUD/PPIUD (%) 109 non-null object
40 Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Pill (%) 109 non-null object
41 Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Condom (%) 109 non-null object
42 Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Injectables (%) 109 non-null object
43 Total Unmet need for Family Planning (Currently Married Women Age 15-49 years)7 (%) 109 non-null object
ct
44 Unmet need for spacing (Currently Married Women Age 15-49 years)7 (%) 109 non-null object

45 Mothers who had an antenatal check-up in the first trimester (for last birth in the 5 years before the survey) (%) 109 non-null object

46 Mothers who had at least 4 antenatal care visits (for last birth in the 5 years before the survey) (%) 109 non-null object

47 Mothers who consumed iron folic acid for 100 days or more when they were pregnant (for last birth in the 5 years before the survey) (%) 109 non-null object
t
48 Mothers who consumed iron folic acid for 180 days or more when they were pregnant (for last birth in the 5 years before the survey) (%) 109 non-null object
t
49 Registered pregnancies for which the mother received a Mother and Child Protection (MCP) card (for last birth in the 5 years before the survey) (%) 109 non-null object

dtypes: float64(7), int64(3), object(40)
memory usage: 43.1+ KB

```

```

In [78]: # Converting columns to appropriate type
nfhsnew['States/UTs'] = nfhsnew['States/UTs'].astype(str)
nfhsnew['Area'] = nfhsnew['Area'].astype(str)

```

```

In [82]: # Function to clean and convert values
def clean_and_convert(value):
    if isinstance(value, str):
        try:
            cleaned_value = float(value.replace('(', '').replace(')', ''))
            return cleaned_value
        except ValueError:
            return None
    elif isinstance(value, int) or isinstance(value, float):
        return float(value)
    else:
        return None

```

```

In [83]: # Clean and convert specified columns
for col in float_columns:
    nfhsnew[col] = nfhsnew[col].apply(clean_and_convert)

```

```

In [84]: #viewing the data types of new columns
nfhsnew.dtypes

```

```

Out[84]: States/UTs
          object
Area
          object

```

Number of Households surveyed	float64
Number of Women age 15-49 years interviewed	float64
Number of Men age 15-54 years interviewed	float64
Female population age 6 years and above who ever attended school (%)	float64
Population below age 15 years (%)	float64
Sex ratio at birth for children born in the last five years (females per 1,000 males)	float64
Children under age 5 years whose birth was registered with the civil authority (%)	float64
Deaths in the last 3 years registered with the civil authority (%)	float64
Population living in households with electricity (%)	float64
Population living in households with an improved drinking-water source <sup>1</sup> (%)	float64
Population living in households that use an improved sanitation facility <sup>2</sup> (%)	float64
Households using clean fuel for cooking <sup>3</sup> (%)	float64
Households using iodized salt (%)	float64
Households with any usual member covered under a health insurance/financing scheme (%)	float64
Children age 5 years who attended pre-primary school during the school year 2019-20 (%)	float64
Women (age 15-49) who are literate <sup>4</sup> (%)	float64
Men (age 15-49) who are literate <sup>4</sup> (%)	float64
Women age 20-24 years married before age 18 years (%)	float64
Men age 25-29 years married before age 21 years (%)	float64
Total Fertility Rate (number of children per woman)	float64
Women age 15-19 years who were already mothers or pregnant at the time of the survey (%)	float64
Adolescent fertility rate for women age 15-19 years <sup>5</sup>	float64
Neonatal mortality rate (per 1000 live births)	float64
Infant mortality rate (per 1000 live births)	float64
Under-five mortality rate (per 1000 live births)	float64
Health worker ever talked to female non-users about family planning (%)	float64
Current users ever told about side effects of current method of family planning <sup>8</sup> (%)	float64
Mothers whose last birth was protected against neonatal tetanus (for last birth in the 5 years before the survey) <sup>9</sup> (%)	float64
Sex ratio of the total population (females per 1,000 males)	float64
Women (age 15-49) with 10 or more years of schooling (%)	float64
Men (age 15-49) with 10 or more years of schooling (%)	float64
Women (age 15-49) who have ever used the internet (%)	float64
Men (age 15-49) who have ever used the internet (%)	float64

```

Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Any
method6 (%) float64
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Any
modern method6 (%) float64
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Fema
le sterilization (%) float64
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Male
sterilization (%) float64
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - IUD/
PPIUD (%) float64
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Pill
(%) float64
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Cond
om (%) float64
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Inje
ctables (%) float64
Total Unmet need for Family Planning (Currently Married Women Age 15-49 years)7 (%)
float64
Unmet need for spacing (Currently Married Women Age 15-49 years)7 (%)
float64
Mothers who had an antenatal check-up in the first trimester (for last birth in the 5 y
ears before the survey) (%) float64
Mothers who had at least 4 antenatal care visits (for last birth in the 5 years before
the survey) (%) float64
Mothers who consumed iron folic acid for 100 days or more when they were pregnant (for l
ast birth in the 5 years before the survey) (%) float64
Mothers who consumed iron folic acid for 180 days or more when they were pregnant (for l
ast birth in the 5 years before the survey) (%) float64
Registered pregnancies for which the mother received a Mother and Child Protection (MCP)
card (for last birth in the 5 years before the survey) (%) float64
dtype: object

```

In [98]: `# Checking for null values in each column`  
`nfhsnew.info()`

```

null_counts = nfhsnew.isnull().sum()
print("\nNull value counts per column:")
print(null_counts)

```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 110 entries, 0 to 109
```

```
Data columns (total 50 columns):
```

```
# Column
```

```
Non-Null Count Dtype
```

```
--- -----
```

```
-----
```

```
0 States/UTs
```

```
110 non-null object
```

```
1 Area
```

```
110 non-null object
```

```
2 Number of Households surveyed
```

```
110 non-null float6
```

```
4
```

```
3 Number of Women age 15-49 years interviewed
```

```
110 non-null float6
```

```
4
```

```
4 Number of Men age 15-54 years interviewed
```

```
110 non-null float6
```

```
4
```

```
5 Female population age 6 years and above who ever attended school (%)
```

```
110 non-null float6
```

4	6	Population below age 15 years (%)	110 non-null	float6
4	7	Sex ratio at birth for children born in the last five years (females per 1,000 male s)	109 non-null	float6
4	8	Children under age 5 years whose birth was registered with the civil authority (%)	109 non-null	float6
4	9	Deaths in the last 3 years registered with the civil authority (%)	108 non-null	float6
4	10	Population living in households with electricity (%)	110 non-null	float6
4	11	Population living in households with an improved drinking-water source1 (%)	110 non-null	float6
4	12	Population living in households that use an improved sanitation facility2 (%)	110 non-null	float6
4	13	Households using clean fuel for cooking3 (%)	109 non-null	float6
4	14	Households using iodized salt (%)	109 non-null	float6
4	15	Households with any usual member covered under a health insurance/financing scheme (%)	109 non-null	float6
4	16	Children age 5 years who attended pre-primary school during the school year 2019-20 (%)	105 non-null	float64
4	17	Women (age 15-49) who are literate4 (%)	110 non-null	float6
4	18	Men (age 15-49) who are literate4 (%)	109 non-null	float6
4	19	Women age 20-24 years married before age 18 years (%)	109 non-null	float6
4	20	Men age 25-29 years married before age 21 years (%)	96 non-null	float6
4	21	Total Fertility Rate (number of children per woman)	109 non-null	float6
4	22	Women age 15-19 years who were already mothers or pregnant at the time of the survey (%)	109 non-null	float6
4	23	Adolescent fertility rate for women age 15-19 years5	110 non-null	float6
4	24	Neonatal mortality rate (per 1000 live births)	97 non-null	float6
4	25	Infant mortality rate (per 1000 live births)	97 non-null	float6
4	26	Under-five mortality rate (per 1000 live births)	110 non-null	float6
4	27	Health worker ever talked to female non-users about family planning (%)	109 non-null	float6

28	Current users ever told about side effects of current method of family planning (%)	108 non-null	float64
29	Mothers whose last birth was protected against neonatal tetanus (for last birth in the 5 years before the survey) (%)	109 non-null	float64
30	Sex ratio of the total population (females per 1,000 males)	110 non-null	float64
31	Women (age 15-49) with 10 or more years of schooling (%)	110 non-null	float64
32	Men (age 15-49) with 10 or more years of schooling (%)	109 non-null	float64
33	Women (age 15-49) who have ever used the internet (%)	109 non-null	float64
34	Men (age 15-49) who have ever used the internet (%)	109 non-null	float64
35	Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Any method (%)	109 non-null	float64
36	Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Any modern method (%)	109 non-null	float64
37	Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Female sterilization (%)	109 non-null	float64
38	Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Male sterilization (%)	109 non-null	float64
39	Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - IUD/PPIUD (%)	109 non-null	float64
40	Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Pill (%)	109 non-null	float64
41	Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Condom (%)	109 non-null	float64
42	Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Injectables (%)	109 non-null	float64
43	Total Unmet need for Family Planning (Currently Married Women Age 15-49 years) (%)	109 non-null	float64
44	Unmet need for spacing (Currently Married Women Age 15-49 years) (%)	109 non-null	float64
45	Mothers who had an antenatal check-up in the first trimester (for last birth in the 5 years before the survey) (%)	109 non-null	float64
46	Mothers who had at least 4 antenatal care visits (for last birth in the 5 years before the survey) (%)	109 non-null	float64
47	Mothers who consumed iron folic acid for 100 days or more when they were pregnant (for last birth in the 5 years before the survey) (%)	109 non-null	float64
48	Mothers who consumed iron folic acid for 180 days or more when they were pregnant (for last birth in the 5 years before the survey) (%)	109 non-null	float64
49	Registered pregnancies for which the mother received a Mother and Child Protection (MCP) card (for last birth in the 5 years before the survey) (%)	109 non-null	float64

dtypes: float64(48), object(2)  
memory usage: 43.1+ KB

Null value counts per column:  
States/UTs

0

Area

0

Number of Households surveyed	0
Number of Women age 15-49 years interviewed	0
Number of Men age 15-54 years interviewed	0
Female population age 6 years and above who ever attended school (%)	0
Population below age 15 years (%)	0
Sex ratio at birth for children born in the last five years (females per 1,000 males)	1
Children under age 5 years whose birth was registered with the civil authority (%)	1
Deaths in the last 3 years registered with the civil authority (%)	2
Population living in households with electricity (%)	0
Population living in households with an improved drinking-water source <sup>1</sup> (%)	0
Population living in households that use an improved sanitation facility <sup>2</sup> (%)	0
Households using clean fuel for cooking <sup>3</sup> (%)	1
Households using iodized salt (%)	1
Households with any usual member covered under a health insurance/financing scheme (%)	1
Children age 5 years who attended pre-primary school during the school year 2019-20 (%)	5
Women (age 15-49) who are literate <sup>4</sup> (%)	0
Men (age 15-49) who are literate <sup>4</sup> (%)	1
Women age 20-24 years married before age 18 years (%)	1
Men age 25-29 years married before age 21 years (%)	14
Total Fertility Rate (number of children per woman)	1
Women age 15-19 years who were already mothers or pregnant at the time of the survey (%)	1
Adolescent fertility rate for women age 15-19 years <sup>5</sup>	0
Neonatal mortality rate (per 1000 live births)	13
Infant mortality rate (per 1000 live births)	13
Under-five mortality rate (per 1000 live births)	0
Health worker ever talked to female non-users about family planning (%)	1
Current users ever told about side effects of current method of family planning <sup>8</sup> (%)	2
Mothers whose last birth was protected against neonatal tetanus (for last birth in the 5 years before the survey) <sup>9</sup> (%)	1
Sex ratio of the total population (females per 1,000 males)	0
Women (age 15-49) with 10 or more years of schooling (%)	0
Men (age 15-49) with 10 or more years of schooling (%)	1
Women (age 15-49) who have ever used the internet (%)	1
Men (age 15-49) who have ever used the internet (%)	1



```

Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Any
method6 (%) 1
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Any
modern method6 (%) 1
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Fema
le sterilization (%) 1
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Male
sterilization (%) 1
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - IUD/
PPIUD (%) 1
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Pill
(%) 1
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Cond
om (%) 1
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Inje
ctables (%) 1
Total Unmet need for Family Planning (Currently Married Women Age 15-49 years)7 (%) 1
Unmet need for spacing (Currently Married Women Age 15-49 years)7 (%) 1
Mothers who had an antenatal check-up in the first trimester (for last birth in the 5 y
ears before the survey) (%) 1
Mothers who had at least 4 antenatal care visits (for last birth in the 5 years before
the survey) (%) 1
Mothers who consumed iron folic acid for 100 days or more when they were pregnant (for l
ast birth in the 5 years before the survey) (%) 1
Mothers who consumed iron folic acid for 180 days or more when they were pregnant (for l
ast birth in the 5 years before the survey) (%) 1
Registered pregnancies for which the mother received a Mother and Child Protection (MCP)
card (for last birth in the 5 years before the survey) (%) 1
dtype: int64

```

```

In [99]: # Separating the DataFrame into numeric and non-numeric columns
numeric_columns = nfhsnew.select_dtypes(include=['number']).columns
non_numeric_columns = nfhsnew.select_dtypes(exclude=['number']).columns

```

```

In [100]: # Replace null values with mean for numeric columns

for col in numeric_columns:
    mean_value = nfhsnew[col].mean() # Calculate mean for the column
    nfhsnew[col].fillna(mean_value, inplace=True) # Replace null values with mean

```

```

In [101]: # Verifying if there are any remaining null values
null_counts_after_fill = nfhsnew.isnull().sum()
print("Null value counts after filling with mean:")
print(null_counts_after_fill)

```

```

Null value counts after filling with mean:
States/UTs
0
Area
0
Number of Households surveyed
0
Number of Women age 15-49 years interviewed
0
Number of Men age 15-54 years interviewed
0
Female population age 6 years and above who ever attended school (%)
0
Population below age 15 years (%)
0
Sex ratio at birth for children born in the last five years (females per 1,000 males)
0

```

Children under age 5 years whose birth was registered with the civil authority (%)	0
Deaths in the last 3 years registered with the civil authority (%)	0
Population living in households with electricity (%)	0
Population living in households with an improved drinking-water source1 (%)	0
Population living in households that use an improved sanitation facility2 (%)	0
Households using clean fuel for cooking3 (%)	0
Households using iodized salt (%)	0
Households with any usual member covered under a health insurance/financing scheme (%)	0
Children age 5 years who attended pre-primary school during the school year 2019-20 (%)	0
Women (age 15-49) who are literate4 (%)	0
Men (age 15-49) who are literate4 (%)	0
Women age 20-24 years married before age 18 years (%)	0
Men age 25-29 years married before age 21 years (%)	0
Total Fertility Rate (number of children per woman)	0
Women age 15-19 years who were already mothers or pregnant at the time of the survey (%)	0
Adolescent fertility rate for women age 15-19 years5	0
Neonatal mortality rate (per 1000 live births)	0
Infant mortality rate (per 1000 live births)	0
Under-five mortality rate (per 1000 live births)	0
Health worker ever talked to female non-users about family planning (%)	0
Current users ever told about side effects of current method of family planning8 (%)	0
Mothers whose last birth was protected against neonatal tetanus (for last birth in the 5 years before the survey)9 (%)	0
Sex ratio of the total population (females per 1,000 males)	0
Women (age 15-49) with 10 or more years of schooling (%)	0
Men (age 15-49) with 10 or more years of schooling (%)	0
Women (age 15-49) who have ever used the internet (%)	0
Men (age 15-49) who have ever used the internet (%)	0
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Any method6 (%)	0
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Any modern method6 (%)	0
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Female sterilization (%)	0
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Male sterilization (%)	0
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - IUD/PPIUD (%)	0
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Pill (%)	0

```

Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Condo
m (%)
0
Current Use of Family Planning Methods (Currently Married Women Age 15-49 years) - Inje
ctables (%)
0
Total Unmet need for Family Planning (Currently Married Women Age 15-49 years)7 (%)
0
Unmet need for spacing (Currently Married Women Age 15-49 years)7 (%)
0
Mothers who had an antenatal check-up in the first trimester (for last birth in the 5 y
ears before the survey) (%)
0
Mothers who had at least 4 antenatal care visits (for last birth in the 5 years before
the survey) (%)
0
Mothers who consumed iron folic acid for 100 days or more when they were pregnant (for l
ast birth in the 5 years before the survey) (%)
0
Mothers who consumed iron folic acid for 180 days or more when they were pregnant (for l
ast birth in the 5 years before the survey) (%)
0
Registered pregnancies for which the mother received a Mother and Child Protection (MCP)
card (for last birth in the 5 years before the survey) (%)
0
dtype: int64

```

```

In [105]: # Summary statistics for numeric columns
summary_stats = nfhsnew[numeric_columns].describe()
summary_stats

```

Out[105]:

	Number of Households surveyed	Number of Women age 15-49 years interviewed	Number of Men age 15- 54 years interviewed	Female population age 6 years and above who ever attended school (%)	Population below age 15 years (%)	Sex ratio at birth for children born in the last five years (females per 1,000 males)	Children under age 5 years whose birth was registered with the civil authority (%)	Deaths in the last 3 years registered with the civil authority (%)
<b>count</b>	110.000000	110.000000	110.000000	110.000000	110.000000	110.000000	110.000000	110.000000
<b>mean</b>	22987.354545	26136.836364	3675.772727	78.096364	24.677273	943.188991	93.290826	77.015741
<b>std</b>	76080.721354	86810.853826	12151.679425	9.956989	4.113630	102.241864	7.131434	18.361598
<b>min</b>	21.000000	26.000000	0.000000	52.400000	18.100000	705.000000	70.800000	33.400000
<b>25%</b>	2844.000000	2797.750000	428.500000	69.725000	22.000000	894.000000	90.950000	69.975000
<b>50%</b>	8017.500000	9047.500000	1289.000000	79.400000	23.950000	935.300000	96.150000	81.300000
<b>75%</b>	18758.500000	21701.500000	3177.750000	85.050000	26.375000	968.500000	98.100000	91.425000
<b>max</b>	636699.000000	724115.000000	101839.000000	97.600000	39.200000	1520.000000	100.000000	100.000000

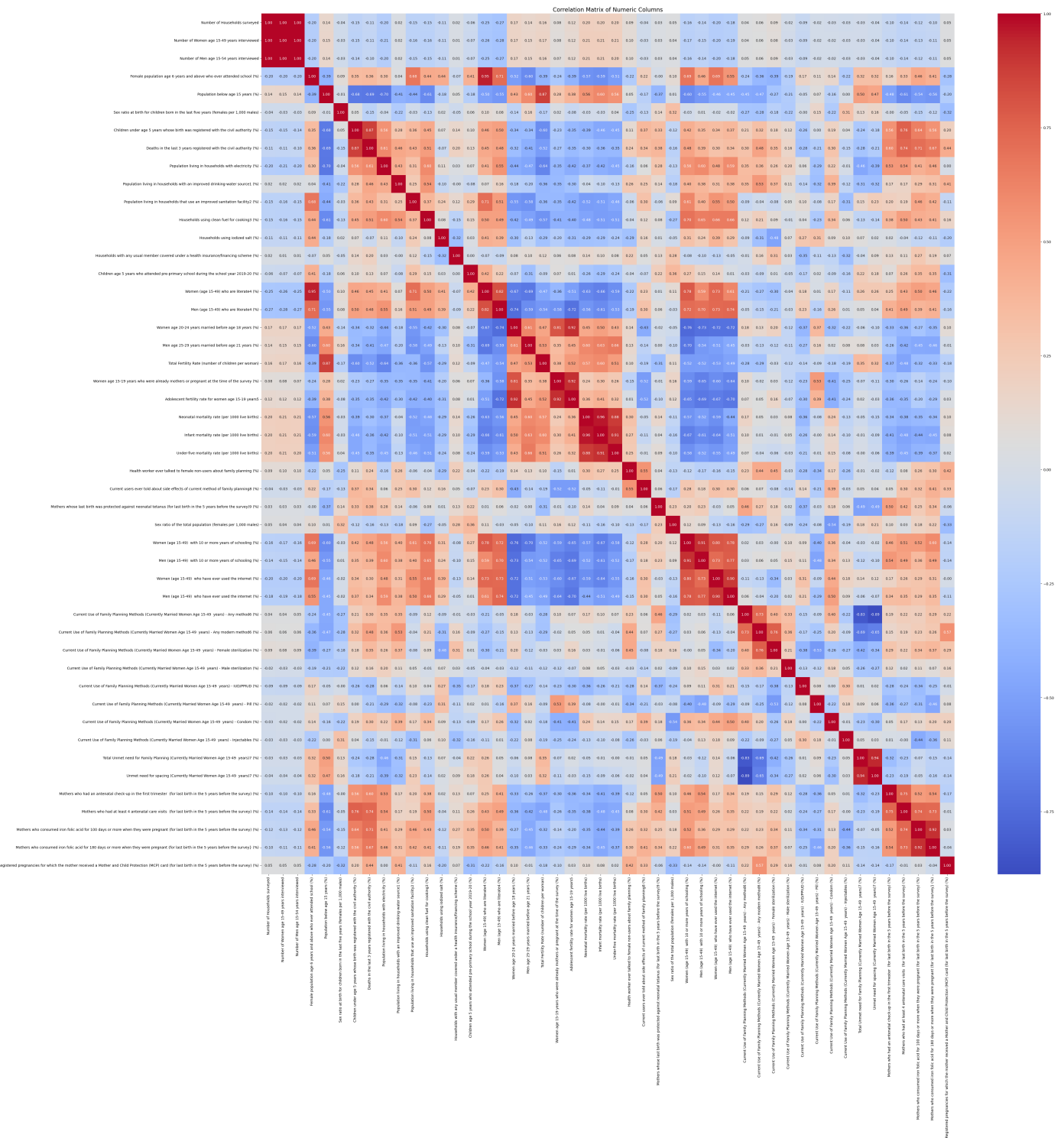
8 rows × 48 columns

```

In [93]: # Viewing the correlation matrix via heatmap
import matplotlib.pyplot as plt
import seaborn as sns
correlation_matrix = nfhsnew[numeric_columns].corr()

```

```
plt.figure(figsize=(44, 44))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Matrix of Numeric Columns', fontsize=16)
plt.show()
```



```
In [117.. # Function to convert percentages to numeric
def convert_percentages_to_numeric(series):
    try:
        return pd.to_numeric(series.str.replace('%', ''), errors='coerce')
    except AttributeError:
        # If AttributeError occurs due to non-string values, try direct coercion
        return pd.to_numeric(series, errors='coerce')
```

```
In [118.. #Let us see if Education level affects child mortality
```

```
data_filtered = nfhsnew[['Women (age 15-49) with 10 or more years of schooling (%)',
                        'Neonatal mortality rate (per 1000 live births)']].copy()
```

```

data_filtered['Women (age 15-49) with 10 or more years of schooling (%)'] = convert_per

data_filtered = data_filtered.dropna()

data_filtered['Neonatal mortality rate (per 1000 live births)'] = pd.to_numeric(data_fil

data_filtered = data_filtered.dropna()

corr_coef, p_value = stats.pearsonr(data_filtered['Women (age 15-49) with 10 or more ye
data_filtered['Neonatal mortality rate (per 1000 liv

# Print results
print(f"Pearson Correlation Coefficient (Education vs. Neonatal Mortality): {corr_coef}")
print(f"P-Value: {p_value}")

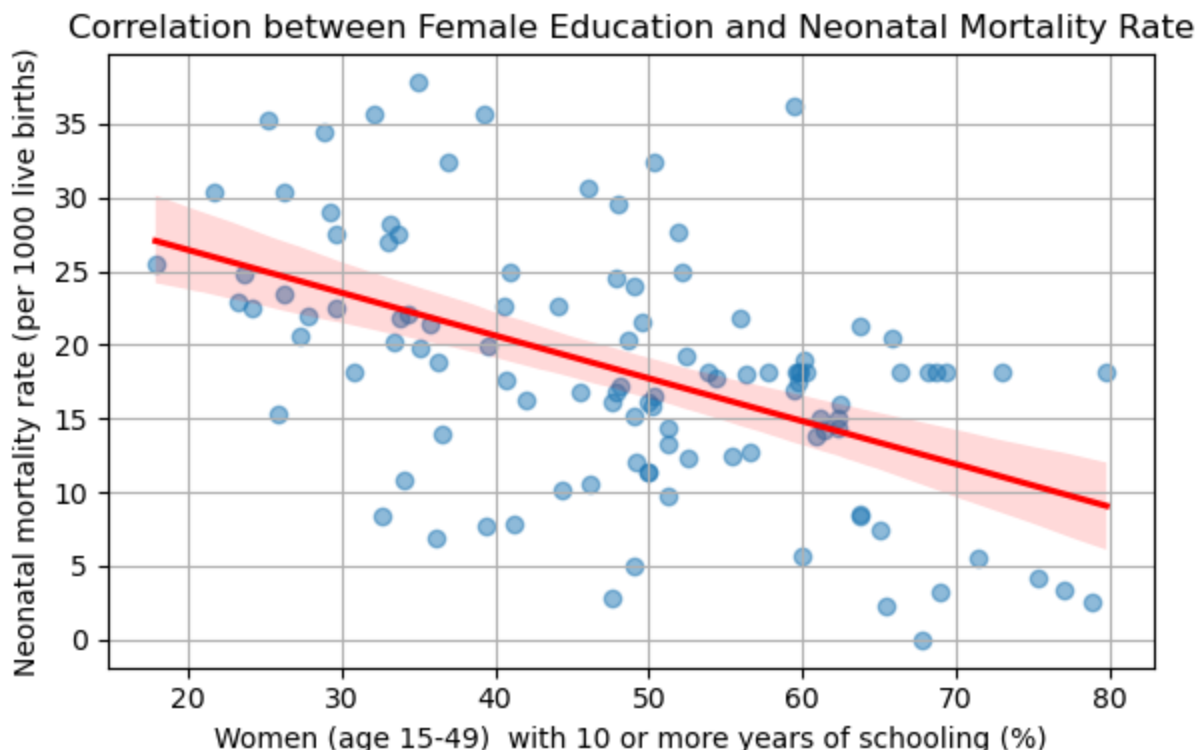
```

Pearson Correlation Coefficient (Education vs. Neonatal Mortality): -0.509484923246031  
P-Value: 1.3135618673794726e-08

```

In [119... # Plotting relation between female literacy and infant mortality
plt.figure(figsize=(6, 4))
sns.regplot(x='Women (age 15-49) with 10 or more years of schooling (%)',
            y='Neonatal mortality rate (per 1000 live births)',
            data=data_filtered,
            scatter_kws={'alpha': 0.5},
            line_kws={'color': 'red'})
plt.title('Correlation between Female Education and Neonatal Mortality Rate')
plt.xlabel('Women (age 15-49) with 10 or more years of schooling (%)')
plt.ylabel('Neonatal mortality rate (per 1000 live births)')
plt.grid(True)
plt.tight_layout()
plt.show()

```



```

In [132... #Let us see if improved Sanitation level affects child mortality
nfhsnew['Population living in households that use an improved sanitation facility2 (%)']
nfhsnew['Under-five mortality rate (per 1000 live births)'] = pd.to_numeric(nfhsnew['Und

```

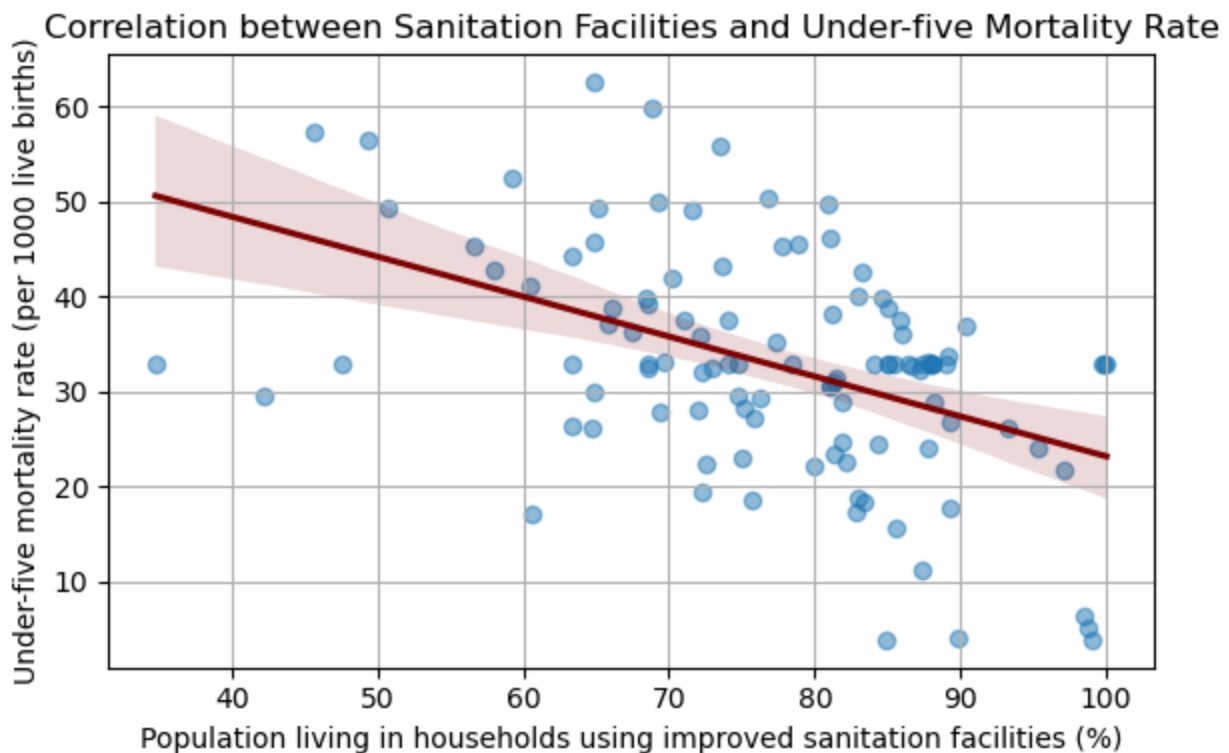
```
data_filtered = nfhsnew[['Population living in households that use an improved sanitation facility',
                        'Under-five mortality rate (per 1000 live births)']].dropna()

corr_coef, p_value = stats.pearsonr(data_filtered['Population living in households that use an improved sanitation facility'],
                                     data_filtered['Under-five mortality rate (per 1000 live births)'])

print(f"Pearson Correlation Coefficient (Sanitation Facilities vs. Under-five Mortality): {corr_coef}")
print(f"P-Value: {p_value}")
```

Pearson Correlation Coefficient (Sanitation Facilities vs. Under-five Mortality): -0.4618251575339904  
P-Value: 3.805722604758897e-07

```
In [135]: # Plotting the relation between sanitation and child death rate
plt.figure(figsize=(6, 4))
sns.regplot(x='Population living in households that use an improved sanitation facility',
            y='Under-five mortality rate (per 1000 live births)',
            data=data_filtered,
            scatter_kws={'alpha':0.5},
            line_kws={'color': 'maroon'})
plt.title('Correlation between Sanitation Facilities and Under-five Mortality Rate')
plt.xlabel('Population living in households using improved sanitation facilities (%)')
plt.ylabel('Under-five mortality rate (per 1000 live births)')
plt.grid(True)
plt.tight_layout()
plt.show()
```



```
In [104]: #seeing if all states have the same mortality rate
from scipy.stats import f_oneway
groups = [nfhsnew[nfhsnew['States/UTs'] == state]['Infant mortality rate (per 1000 live births)']

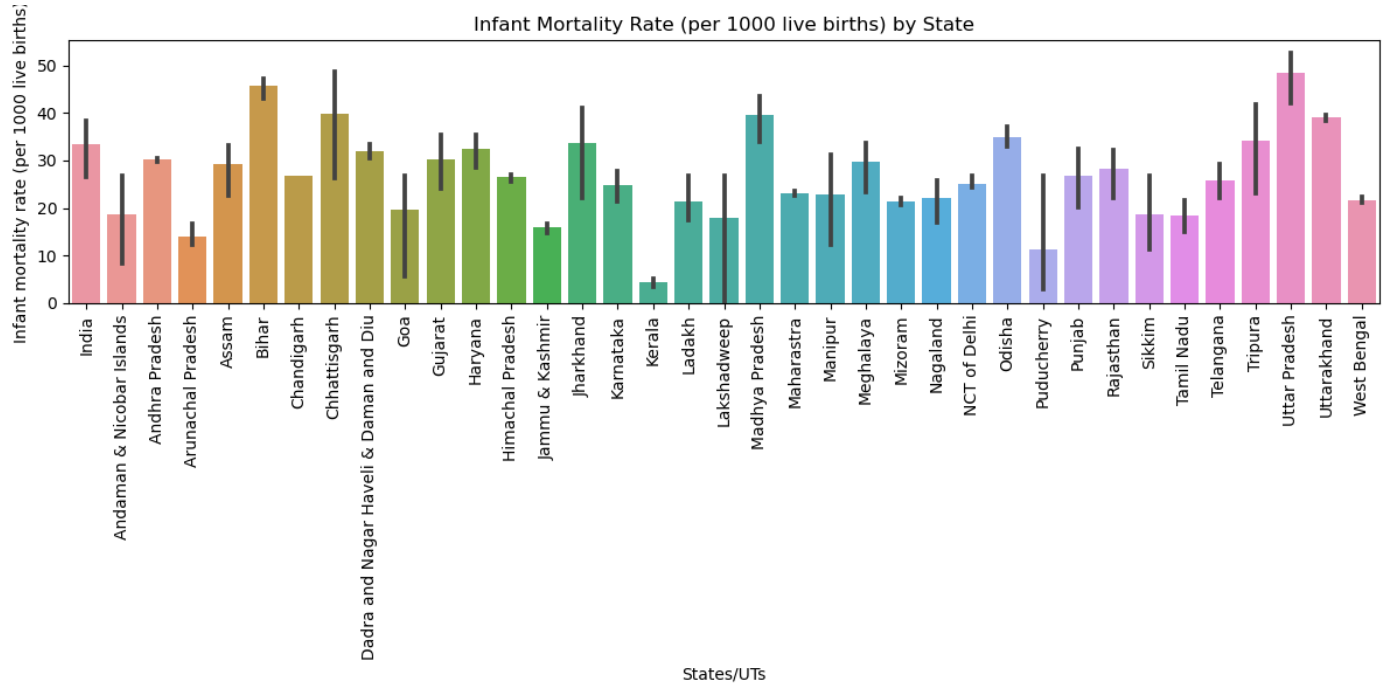
f_stat, p_value = f_oneway(*groups)

print(f"P-value for one-way ANOVA: {p_value:.4f}")

P-value for one-way ANOVA: 0.0000
```

```
In [107]: # Plotting the Infant Mortality Rate by State
plt.figure(figsize=(12, 6))
```

```
sns.barplot(x='States/UTs', y='Infant mortality rate (per 1000 live births)', data=nfhsn)
plt.xticks(rotation=90)
plt.title('Infant Mortality Rate (per 1000 live births) by State')
plt.tight_layout()
plt.show()
```



```
In [122... #TFR depends on education level
from scipy.stats import spearmanr

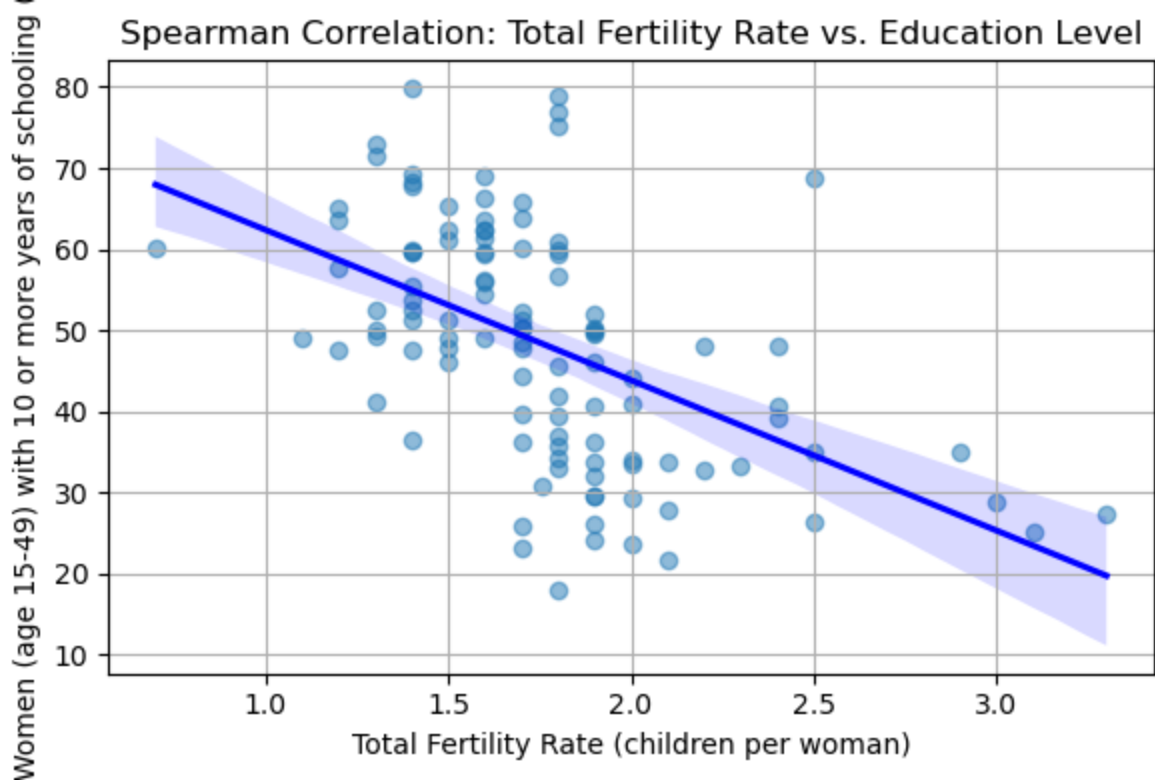
x = nfhsnew['Total Fertility Rate (number of children per woman)']
y = nfhsnew['Women (age 15-49) with 10 or more years of schooling (%)']

correlation, p_value = spearmanr(x, y)
print(f"P-value for one-way ANOVA: {p_value:.4f}")

P-value for one-way ANOVA: 0.0000
```

```
In [127... # Create scatter plot with regression line
plt.figure(figsize=(6, 4))
sns.regplot(x=x, y=y, scatter_kws={'alpha': 0.5}, line_kws={'color': 'blue'})
plt.title('Spearman Correlation: Total Fertility Rate vs. Education Level')
plt.xlabel('Total Fertility Rate (children per woman)')
plt.ylabel('Women (age 15-49) with 10 or more years of schooling (%)')
plt.grid(True)
plt.tight_layout()
plt.show()
```





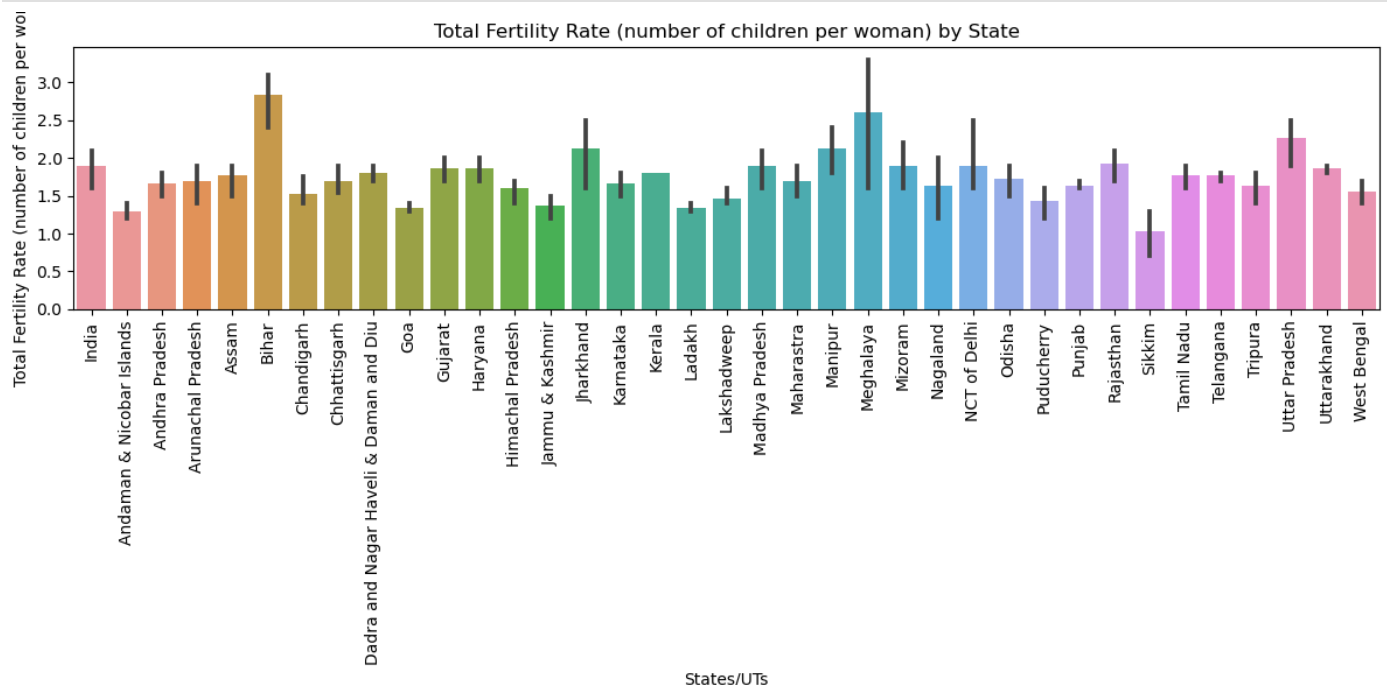
```
In [143... #TFR is different across states

groups = [nfhsnew[nfhsnew['States/UTs'] == state]['Total Fertility Rate (number of child

f_stat, p_value = f_oneway(*groups)
print(f"P-value for one-way ANOVA: {p_value:.4f}")

P-value for one-way ANOVA: 0.0000
```

```
In [144... # Plotting the TFR by State
plt.figure(figsize=(12, 6))
sns.barplot(x='States/UTs', y='Total Fertility Rate (number of children per woman)', dat
plt.xticks(rotation=90)
plt.title('Total Fertility Rate (number of children per woman) by State')
plt.tight_layout()
plt.show()
```



```
In [148... #urban rural divide in terms of sanitation
```

```

nfhsnew['Population living in households that use an improved sanitation facility2 (%)']
nfhsnew = nfhsnew.dropna(subset=['Population living in households that use an improved s

groups = nfhsnew.groupby('Area')['Population living in households that use an improved s

f_stat, p_value = stats.f_oneway(*groups)

print(f"F-Statistic: {f_stat:.4f}")
print(f"P-Value: {p_value:.4f}")

```

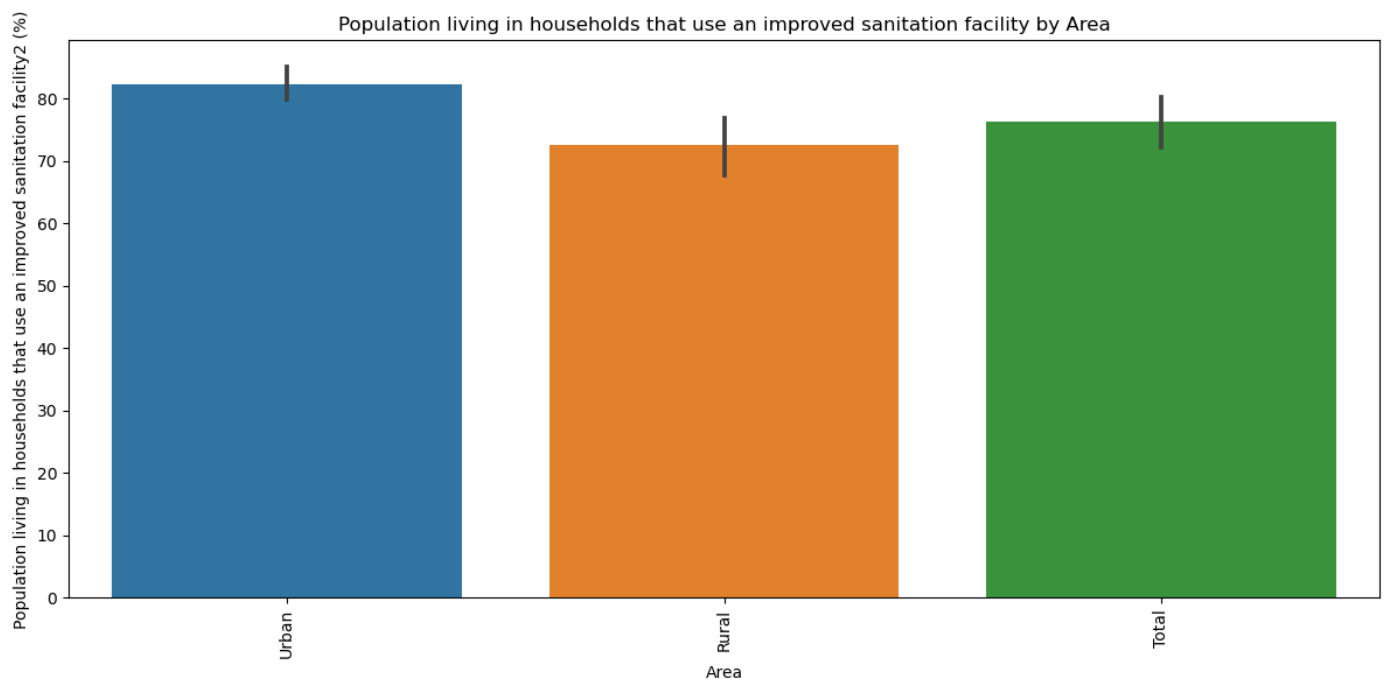
F-Statistic: 6.0237

P-Value: 0.0033

```

In [149... # Plotting the sanitation percentage by area type
plt.figure(figsize=(12, 6))
sns.barplot(x='Area', y='Population living in households that use an improved sanitation
plt.xticks(rotation=90)
plt.title('Population living in households that use an improved sanitation facility by A
plt.tight_layout()
plt.show()

```



```

In [156... # ANC rates differ by the type of area
checkout_rates = [nfhsnew[nfhsnew['Area'] == area]['Mothers who had an antenatal check-up

f_stat, p_value = stats.f_oneway(*checkout_rates)

print(f"ANOVA Test for Antenatal Check-up Rates by Area")
print(f"F-Statistic: {f_stat:.4f}")
print(f"P-Value: {p_value:.4f}")

```

ANOVA Test for Antenatal Check-up Rates by Area

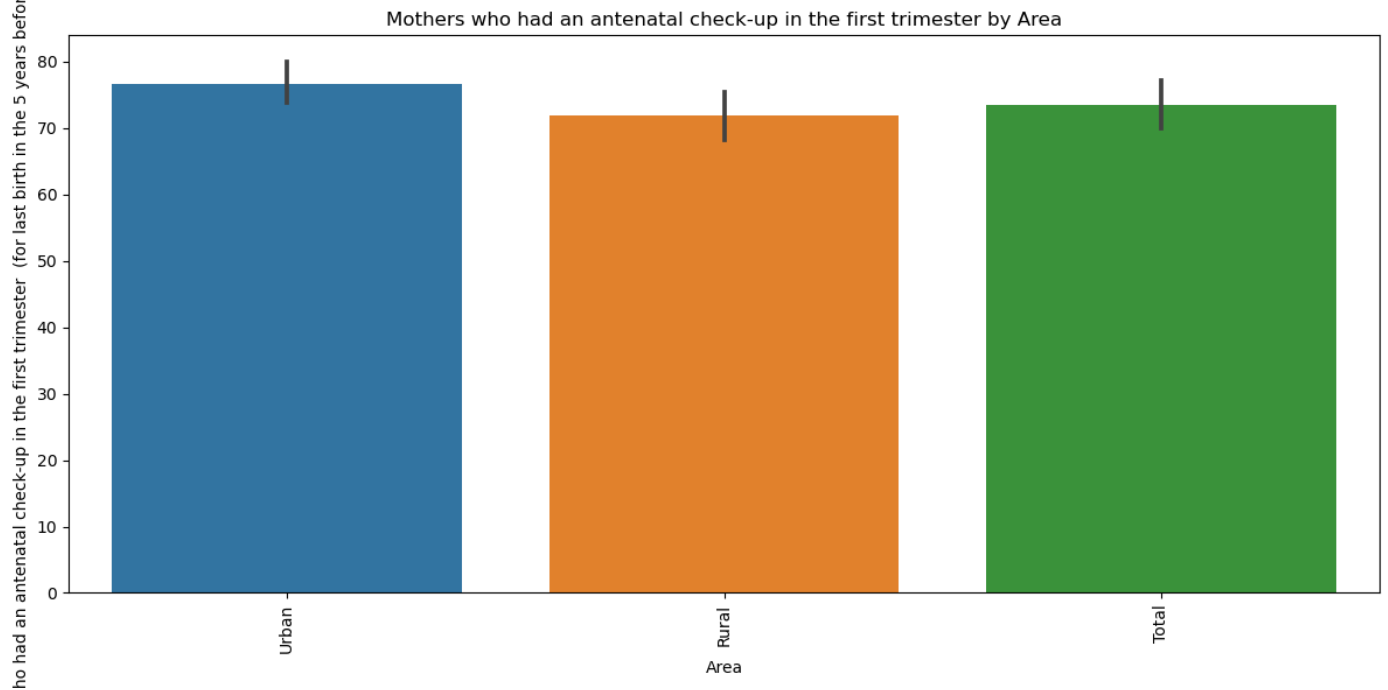
F-Statistic: 1.8658

P-Value: 0.1598

```

In [157... # Plotting the ANC rates differ by the type of area
plt.figure(figsize=(12, 6))
sns.barplot(x='Area', y='Mothers who had an antenatal check-up in the first trimester (
plt.xticks(rotation=90)
plt.title('Mothers who had an antenatal check-up in the first trimester by Area')
plt.tight_layout()
plt.show()

```



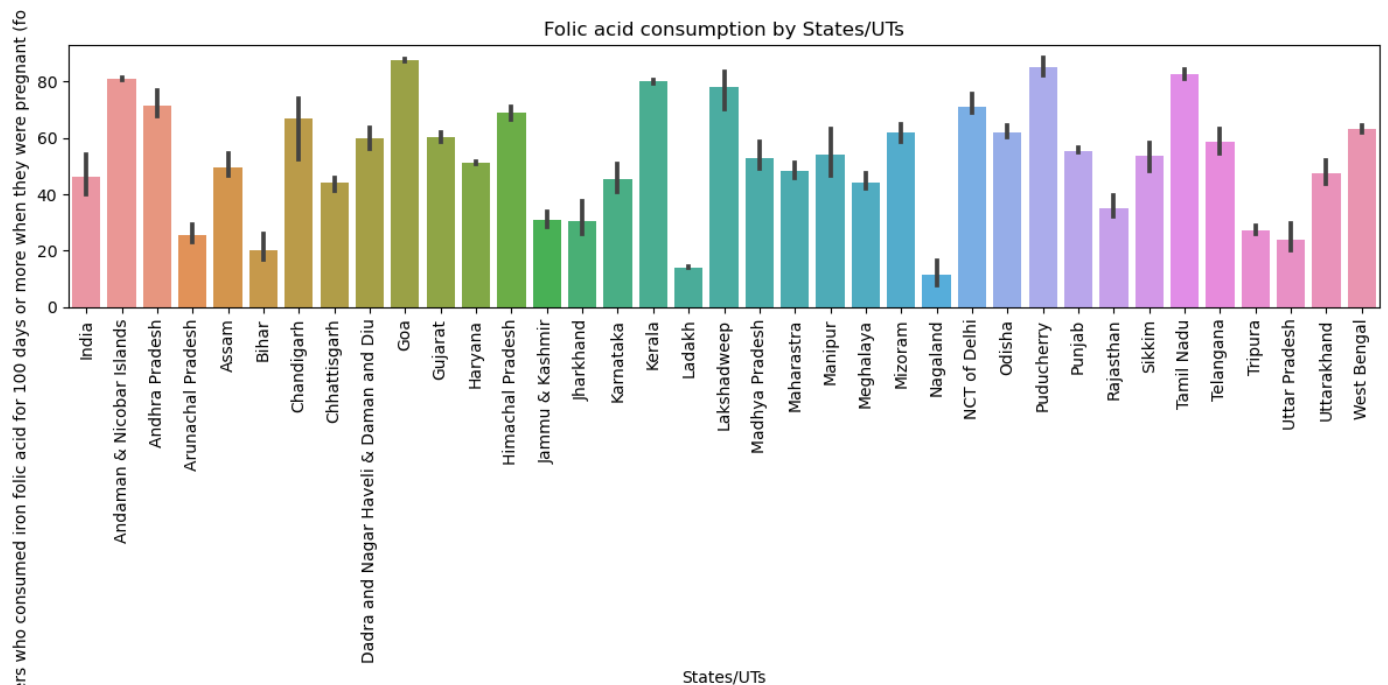
```
In [159... # Performing ANOVA test by states to see if folic acid consumption varies or not
iron_folic_acid = [nfhsnew[nfhsnew['States/UTs'] == StatesUTs]['Mothers who consumed iron
folic acid for 100 days or more']

f_stat, p_value = stats.f_oneway(*iron_folic_acid)

print(f"ANOVA Test for Iron Folic Acid Consumption for 100 Days or More by States/UTs")
print(f"F-Statistic: {f_stat:.4f}")
print(f"P-Value: {p_value:.4f}")
```

ANOVA Test for Iron Folic Acid Consumption for 100 Days or More by States/UTs  
 F-Statistic: 63.7476  
 P-Value: 0.0000

```
In [161... # Plotting the folic acid consumption rates by the type of state
plt.figure(figsize=(12, 6))
sns.barplot(x='States/UTs', y='Mothers who consumed iron folic acid for 100 days or more')
plt.xticks(rotation=90)
plt.title('Folic acid consumption by States/UTs')
plt.tight_layout()
plt.show()
```



```

In [162... # Selecting relevant variables including partner alignment related features
X = nfhsnew[['Total Fertility Rate (number of children per woman)',
            'Women (age 15-49) with 10 or more years of schooling (%)',
            'Health worker ever talked to female non-users about family planning (%)',
            'Mothers who had at least 4 antenatal care visits (for last birth in the 5
            'Mothers who consumed iron folic acid for 100 days or more when they were p
            'Men age 25-29 years married before age 21 years (%)',
            'Current Use of Family Planning Methods (Currently Married Women Age 15-49
            'Current Use of Family Planning Methods (Currently Married Women Age 15-49
            'Current Use of Family Planning Methods (Currently Married Women Age 15-49

y = nfhsnew['Neonatal mortality rate (per 1000 live births)']

In [163... # Ensuring all selected variables are numeric
X = X.apply(pd.to_numeric, errors='coerce')

In [164... # Handling missing values
X.fillna(X.mean(), inplace=True)
y.fillna(y.mean(), inplace=True)

In [166... # Splitting the data into training and testing sets
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

In [173... # Initialize the Random Forest Regressor
from sklearn.ensemble import GradientBoostingRegressor
gb_regressor = GradientBoostingRegressor(random_state=42)

In [174... # Fitting the model
gb_regressor.fit(X_train, y_train)

Out[174]: ▾ GradientBoostingRegressor
GradientBoostingRegressor(random_state=42)

In [175... # Predicting on the test set
y_pred = gb_regressor.predict(X_test)

In [178... # Model evaluation
from sklearn.metrics import r2_score

# Print the adjusted R-squared score

print(f"R^2 Score: {abs_r2:.2f}")

R^2 Score: 0.90

In [179... # Printing feature importances
feature_importances = pd.DataFrame({'Feature': X.columns, 'Importance': gb_regressor.fea
print("\nFeature Importances:")
print(feature_importances.sort_values(by='Importance', ascending=False))

Feature Importances:

Feature Importance
5 Men age 25-29 years married before age 21 year... 0.396571
0 Total Fertility Rate (number of children per w... 0.201906
4 Mothers who consumed iron folic acid for 100 d... 0.099990
2 Health worker ever talked to female non-users ... 0.091767
3 Mothers who had at least 4 antenatal care visi... 0.054852
6 Current Use of Family Planning Methods (Curren... 0.048849
7 Current Use of Family Planning Methods (Curren... 0.037732

```

1	Women (age 15-49) with 10 or more years of sc...	0.036135
8	Current Use of Family Planning Methods (Curren...	0.032198