



## EDA Internship Task Week 6

### Birth Rate Analysis

Let's take a look at the freely available data on births in the United States, provided by the Centers for Disease Control (CDC). Find Dataset "births.csv" and implement EDA techniques.

Q.1: Add a new column "Decade" by calculating. For example 1969 will be 1960, 1988 will 1980 etc.

**Q.2: Show the descriptive statistics of the data.**

Q.3: Check if your data contains any missing values

**Q.4: What is the trend of male & female births every decade?**

Q.5: To remove outliers from dataset following technique to include only those values that fall within 5 standard deviations from the mean. This is a common statistical technique used to focus on the central tendency of the data while excluding extreme values.

Use this technique to remove outliers.

**Q.6: Plot births by weekday for several decades. Write down your observation.**

**Q.7: Group the data by month and day separately**

**Q.8: Focusing on the month and day only, you have a time series reflecting the average number of births by date of the year. From this, plot the data.**

