Exploratory Data Analysis on a Dataset

**Objective**:

The main goal of this assignment is to conduct a thorough exploratory analysis of the "cardiographic.csv" dataset to uncover insights, identify patterns, and understand the dataset's underlying structure. You will use statistical summaries, visualizations, and data manipulation techniques to explore the dataset comprehensively.

Dataset:

1. **LB** - Likely stands for "Baseline Fetal Heart Rate (FHR)" which represents the average fetal heart rate over a period.
2. **AC** - Could represent "Accelerations" in the FHR. Accelerations are usually a sign of fetal well-being.
3. **FM** - May indicate "Fetal Movements" detected by the monitor.
4. **UC** - Likely denotes "Uterine Contractions", which can impact the FHR pattern.
5. **DL** - Could stand for "Decelerations Late" with respect to uterine contractions, which can be a sign of fetal distress.
6. **DS** - May represent "Decelerations Short" or decelerations of brief duration.
7. **DP** - Could indicate "Decelerations Prolonged", or long-lasting decelerations.
8. **ASTV** - Might refer to "Percentage of Time with Abnormal Short Term Variability" in the FHR.
9. **MSTV** - Likely stands for "Mean Value of Short Term Variability" in the FHR.
10. **ALTV** - Could represent "Percentage of Time with Abnormal Long Term Variability" in the FHR.
11. **MLTV** - Might indicate "Mean Value of Long Term Variability" in the FHR.

Tools and Libraries:

* Python or R programming language
* Data manipulation libraries
* Data visualization libraries (Matplotlib and Seaborn in Python)
* Jupyter Notebook for documenting your analysis

Tasks:

1. **Data Cleaning and Preparation:**
   * Load the dataset into a DataFrame or equivalent data structure.
   * Handle missing values appropriately (e.g., imputation, deletion).
   * Identify and correct any inconsistencies in data types (e.g., numerical values stored as strings).
   * Detect and treat outliers if necessary.
2. **Statistical Summary:**
   * Provide a statistical summary for each variable in the dataset, including measures of central tendency (mean, median) and dispersion (standard deviation, interquartile range).
   * Highlight any interesting findings from this summary.
3. **Data Visualization:**
   * Create histograms or boxplots to visualize the distributions of various numerical variables.
   * Use bar charts or pie charts to display the frequency of categories for categorical variables.
   * Generate scatter plots or correlation heatmaps to explore relationships between pairs of variables.
   * Employ advanced visualization techniques like pair plots, or violin plots for deeper insights.
4. **Pattern Recognition and Insights:**
   * Identify any correlations between variables and discuss their potential implications.
   * Look for trends or patterns over time if temporal data is available.
5. **Conclusion:**
   * Summarize the key insights and patterns discovered through your exploratory analysis.
   * Discuss how these findings could impact decision-making or further analyses.

Deliverables:

* A detailed Jupyter Notebook file containing the code, visualizations, and explanations for each step of your analysis.
* A brief report summarizing your findings, insights, and any recommendations for further analysis or actions based on the data.

During my time at Shivakumar & Co., I developed a strong interest in data analysis. Observing a lack of dedicated data analysts within the company, I proposed to my Managing Director the creation of such a role. While the data analyst position was not directly added to Shivakumar & Co., the company recognised the value of this role and created it within our affiliated firm, Om Associates. As a result, I transitioned to the data analyst role at Om Associates to pursue my passion for data analysis and to contribute more effectively to our joint firm's success.

After getting married in 2020, my spouse and I decided to embark on a new journey together, which included pursuing my Master's degree in the United States. This decision necessitated my departure from [Company Name] in 2020 to facilitate our relocation to the USA. Subsequently, I enrolled in a Master's program in Big Data Analytics at SDSU, which has been an enriching experience, allowing me to deepen my expertise in this field. Now, having successfully completed my studies, I'm excited to re-enter the workforce with enhanced skills and knowledge.

Out[150]: