**Data Cleaning and Processing:**

1. **Step 1: Importing Libraries**
   * Imported essential libraries such as numpy, pandas, seaborn, matplotlib, and scipy for data manipulation, visualization, and statistical analysis.
   * Utilized MinMaxScaler and StandardScaler from sklearn for feature scaling.
2. **Step 2: Loading the Dataset**
   * Loaded the "NFL Play by Play 2009-2016" dataset using pd.read\_csv() to begin data cleaning and inspection.
3. **Step 3: Inspecting the Data**
   * Initial inspection included checking:
     + Dataset shape: df.shape
     + First few rows: df.head()
     + Descriptive statistics: df.describe()
     + Data types and missing values: df.info(), df.isnull().sum()
4. **Step 4: Handling Missing Values**
   * **Column: down**
     + Filled missing values using the mean for numerical consistency.
   * **Column: time**
     + Dropped rows with missing values since the missing data was crucial for analysis.
   * **Column: PlayTimeDiff**
     + Similar approach as time, dropped rows with missing values for accurate time analysis.
   * **Column: SideofField**
     + Replaced missing values with the mode (most frequent value).
   * **Column: yacEPA & Home\_WP\_pre**
     + Filled missing values with their respective column means.
   * **Automatic Filling**
     + Filled missing values in all numeric columns with their mean.
     + Filled missing values in non-numeric (object) columns using the mode.
5. **Step 5: Handling Outliers**
   * Visualized outliers using boxplots to identify extreme data points.
   * Used the Interquartile Range (IQR) method to detect and list the indices of outliers.
6. **Step 6: Feature Engineering**
   * Created new features such as:
     + TotalYards by adding Yards.Gained and AirYards.
     + GameTimeLeft by subtracting PlayTimeDiff from TimeSecs to track time left in the game.
7. **Step 7: Normalization and Scaling**
   * Applied MinMaxScaler to scale newly created features like TotalYards and GameTimeLeft for modeling purposes.
8. **Step 8: Statistical Visualization**
   * Visualized the distribution of key variables (e.g., TotalYards) using histograms and density plots to understand data spread.
   * Created a correlation matrix to examine relationships between variables, identifying multicollinearity and key patterns.
9. **Step 9: Final Data Quality Check**
   * After all cleaning steps, performed a final check for missing values and inconsistencies using df.isnull().sum() to ensure data integrity.