Overview of What I've Done

1. **Imported Libraries**
   - Brought in `pandas` for data manipulation.
2. **Loaded the Dataset**
   - Read the Kaggle "Medical Appointments" CSV (May 2016) into a DataFrame.
   - Displayed initial `head()`, `info()` to understand shape and dtypes.
3. **Exploratory Data Analysis (EDA)**
   - Checked for duplicate rows.
   - Generated summary statistics (`describe()`) for numerical columns.
4. **Data Cleaning & Preprocessing**
   - **Invalid Ages**: Removed entries where `age` is negative or zero.
   - **Column Names**: Lowercased and replaced spaces/special characters with underscores.
   - **Renaming**: Standardized certain column names (e.g., `no_show` → `no_show`, if you renamed).
   - **Handicap Encoding**: Converted the `handicap` column to binary (0 = no handicap, 1 = any handicap).
   - **Date Parsing**: Cast appointment and scheduled dates to `datetime` objects for time-series readiness.
   - **No-Show Encoding**: Mapped "Yes"/"No" to 1/0 to facilitate modeling later.
   - **Gender Standardization**: Stripped whitespace, uppercased, and replaced `M`/`F` with `Male`/`Female`.
   - **Neighborhood Formatting**: Title-cased neighborhood names so they're human-readable and consistent.

5. **Final Checks & Export**

    - Ran `info()`, `describe()`, and `head()` again to verify cleaning.
    - Saved the cleaned DataFrame to `cleaned_medical_appointments.csv`.