Customer Analysis Project

Overview

This project involves analyzing customer data from the `customers.csv` dataset. The analysis focuses on understanding
customer behavior by examining metrics such as spending habits, order patterns, demographic distributions, and correlations
between key variables. The project leverages Python libraries like `pandas`, `seaborn`, and `matplotlib` for data analysis
and visualization.

Features

1. Data Processing
- **Load Data:** The dataset is read from a CSV file (`customers.csv`) and loaded into a DataFrame for analysis.
- **Data Cleanup:** Cleaned data is saved as `cleaned_data.csv` for future use.
- **Age Grouping:** Customers are categorized into age groups for demographic analysis.

2. Descriptive Statistics
- Summary statistics of the dataset are generated using `describe()`.
- Average metrics (e.g., age, spending, orders) calculated for overall and subgroup analyses.

3. Analysis by Subgroups
- **By Gender:**
  - Average spending and orders.
  - Box plots and bar charts for visualizing spending and order distributions.
- **By Job and Hobbies:**
  - Average spending and orders grouped by job titles and hobbies.
  - Bar charts for visual representation.
- **By Marital Status:**
  - Average spending and orders.
  - Bar charts for marital status-based insights.

4. Correlation Analysis
- Calculated correlation between age and spending.
- Scatter plots for age vs. spending and orders vs. spending (colored by gender).

5. Key Metrics
- **Overall Metrics:**
  - Average age, spending, and orders.
  - Total spending and total orders.
- **Recent Registrations:**
  - Analysis of orders and spending for customers registered after January 1, 2023.
- **High vs. Low Spenders:**
  - Identified and categorized customers based on spending relative to the average.

Visualizations
The following visualizations are generated:
1. **Bar Charts:**
   - Spending and orders by gender, job, and marital status.

2. **Scatter Plots:**
   - Age vs. spending.
   - Spending vs. orders (with gender hue).
3. **Box Plot:**
   - Spending by gender.

All plots are displayed using `matplotlib` and `seaborn`.

Usage

Prerequisites
- Python 3.7 or higher
- Required libraries:
  - `pandas`
  - `matplotlib`
  - `seaborn`

Steps
1. Place the `customers.csv` file in the project directory.
2. Run the script to analyze the data and generate visualizations.
3. Review the cleaned data saved as `cleaned_data.csv`.

File Descriptions
- **customers.csv:** Original dataset containing customer details.
- **cleaned_data.csv:** Cleaned version of the dataset after analysis.

Key Insights
1. Spending and orders vary significantly across demographics (e.g., gender, marital status, and job roles).
2. Age shows a weak correlation with spending, as shown in the analysis.
3. High and low spenders were identified for potential targeted marketing.

Future Improvements
- Add data validation checks for missing or invalid entries.
- Implement advanced analysis, such as predictive modeling or clustering.
- Enhance visualizations with interactive tools like Plotly or Dash.
