Project File

Objective

The notebook performs an **Exploratory Data Analysis (EDA)** on Amazon product reviews, focusing on understanding review distribution, detecting missing data, and analyzing sentiment.

Key Steps & Workflow

1. **Library Setup**
    - Installs and imports essential libraries for:
        - Data processing: `pandas`, `numpy`
        - Text and sentiment analysis: `nltk`, `TextBlob`, `SentimentIntensityAnalyzer`
        - Visualization: `matplotlib`, `seaborn`, `plotly`, `wordcloud`
    - Configures warnings, plot settings, and inline notebook mode.
2. **Data Loading**
    - Reads the dataset `amazon.csv`.
    - Sorts the dataset by the column `wilson_lower_bound` in descending order.
    - Removes unnecessary `Unnamed: 0` column.
3. **Data Inspection**
    - **`check_dataframe(df)`**:
        - Prints shape, data types, missing values, duplicate counts, and numeric quantiles.
    - **`check_class(df)`**:
        - Lists each column with the number of unique values.

4. **Visualization Functions**
    - `categorical_variable_summary(df, column_name)`:
        - Creates a Plotly subplot with:
            - **Count plot** for frequency
            - **Pie chart** for percentage distribution
    - Example: Analysis of the `overall` rating column.
5. **Text Analysis**
    - Extracts and inspects `reviewText` for further sentiment analysis.
    - (Later cells likely include sentiment scoring, word clouds, and keyword extraction.)

Main Outcomes

- **Dataset structure** and missing data profile are clearly identified.
- **Categorical review distribution** visualized interactively.
- **Base sentiment analysis** prepared using NLTK & TextBlob.
- **Review importance ranking** by `wilson_lower_bound`.

Executable Highlights

To **run this project end-to-end**, you:

1. Install dependencies (`wordcloud`, `nltk`, `plotly`, etc.).
2. Load `amazon.csv` in the same folder.

Execute:

python

CopyEdit

```
check_dataframe(df)
check_class(df)
categorical_variable_summary(df, 'overall')
```

3.
4. Extend sentiment analysis on `reviewText`.