

Elevate Labs task_1

Submission Date:

```
In [1]: import pandas as pd

# Load dataset
df = pd.read_csv("amazon.csv")
```

```
In [3]: df
```

Out[3]:

	product_id	product_name	category	d
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories Accessories&Peripherals ...	
1	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories Accessories&Peripherals ...	
2	B096MSW6CT	Sounce Fast Phone Charging Cable & Data Sync U...	Computers&Accessories Accessories&Peripherals ...	
3	B08HDJ86NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories Accessories&Peripherals ...	
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories Accessories&Peripherals ...	
...	
1460	B08L7J3T31	Noir Aqua - 5pcs PP Spun Filter + 1 Spanner ...	Home&Kitchen Kitchen&HomeAppliances WaterPurif...	
1461	B01M6453MB	Prestige Delight PRWO Electric Rice Cooker (1 ...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	
1462	B009P2LIL4	Bajaj Majesty RX10 2000 Watts Heat Convector R...	Home&Kitchen Heating,Cooling&AirQuality RoomHe...	
1463	B00J5DYCCA	Havells Ventil Air DSP 230mm Exhaust Fan (Pist...	Home&Kitchen Heating,Cooling&AirQuality Fans E...	
1464	B01486F4G6	Borosil Jumbo 1000-Watt Grill Sandwich Maker (...)	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	

1465 rows × 16 columns



```
In [7]: # Step 1: Drop duplicate rows  
df = df.drop_duplicates()
```

```
In [11]: df
```

Out[11]:

	product_id	product_name	category	d
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories Accessories&Peripherals ...	
1	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories Accessories&Peripherals ...	
2	B096MSW6CT	Sounce Fast Phone Charging Cable & Data Sync U...	Computers&Accessories Accessories&Peripherals ...	
3	B08HDJ86NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories Accessories&Peripherals ...	
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories Accessories&Peripherals ...	
...	
1460	B08L7J3T31	Noir Aqua - 5pcs PP Spun Filter + 1 Spanner ...	Home&Kitchen Kitchen&HomeAppliances WaterPurif...	
1461	B01M6453MB	Prestige Delight PRWO Electric Rice Cooker (1 ...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	
1462	B009P2LIL4	Bajaj Majesty RX10 2000 Watts Heat Convector R...	Home&Kitchen Heating,Cooling&AirQuality RoomHe...	
1463	B00J5DYCCA	Havells Ventil Air DSP 230mm Exhaust Fan (Pist...	Home&Kitchen Heating,Cooling&AirQuality Fans E...	
1464	B01486F4G6	Borosil Jumbo 1000-Watt Grill Sandwich Maker (...)	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	

1465 rows × 16 columns



```
In [21]: # Step 2: Rename column headers (lowercase, no spaces)
df.columns = df.columns.str.strip().str.lower().str.replace("_", " ")
```

```
In [23]: df.columns
```

```
Out[23]: Index(['product id', 'product name', 'category', 'discounted price',
               'actual price', 'discount percentage', 'rating', 'rating count',
               'about product', 'user id', 'user name', 'review id', 'review title',
               'review content', 'img link', 'product link'],
              dtype='object')
```

```
In [29]: # Step 3: Handle missing values
print("Missing values before:", df.isnull().sum())
df = df.dropna(subset=['rating count']) # or use df['rating count'].fillna('0',
```

```
Missing values before: product id      0
product name      0
category          0
discounted price  0
actual price      0
discount percentage  0
rating            0
rating count      0
about product     0
user id           0
user name         0
review id         0
review title      0
review content    0
img link          0
product link      0
dtype: int64
```

```
In [31]: print("Missing values after:", df.isnull().sum())
```

```
Missing values after: product id      0
product name      0
category          0
discounted price  0
actual price      0
discount percentage  0
rating            0
rating count      0
about product     0
user id           0
user name         0
review id         0
review title      0
review content    0
img link          0
product link      0
dtype: int64
```

```
In [37]: # Step 4: Convert price columns to numeric (remove ₹ and commas)
for col in ['discounted price', 'actual price']:
    df[col] = df[col].str.replace('₹', '', regex=False).str.replace(',', '', reg
```

```
In [39]: df
```

Out[39]:

	product id	product name	category	disc
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories Accessories&Peripherals ...	
1	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories Accessories&Peripherals ...	
2	B096MSW6CT	Source Fast Phone Charging Cable & Data Sync U...	Computers&Accessories Accessories&Peripherals ...	
3	B08HDJ86NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories Accessories&Peripherals ...	
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories Accessories&Peripherals ...	
...	
1460	B08L7J3T31	Noir Aqua - 5pcs PP Spun Filter + 1 Spanner ...	Home&Kitchen Kitchen&HomeAppliances WaterPurif...	
1461	B01M6453MB	Prestige Delight PRWO Electric Rice Cooker (1 ...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	
1462	B009P2LIL4	Bajaj Majesty RX10 2000 Watts Heat Convector R...	Home&Kitchen Heating,Cooling&AirQuality RoomHe...	
1463	B00J5DYCCA	Havells Ventil Air DSP 230mm Exhaust Fan (Pist...	Home&Kitchen Heating,Cooling&AirQuality Fans E...	

	product id	product name	category	disc
1464	B01486F4G6	Borosil Jumbo 1000-Watt Grill Sandwich Maker (...)	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	

1463 rows × 16 columns

```
In [47]: # Step 5: Convert discount_percentage (remove %)
df['discount percentage'] = df['discount percentage'].astype(str).str.replace('%', '')
df
```

Out[47]:

	product id	product name	category	disc
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories Accessories&Peripherals ...	
1	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories Accessories&Peripherals ...	
2	B096MSW6CT	Source Fast Phone Charging Cable & Data Sync U...	Computers&Accessories Accessories&Peripherals ...	
3	B08HDJ86NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories Accessories&Peripherals ...	
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories Accessories&Peripherals ...	
...	
1460	B08L7J3T31	Noir Aqua - 5pcs PP Spun Filter + 1 Spanner ...	Home&Kitchen Kitchen&HomeAppliances WaterPurif...	
1461	B01M6453MB	Prestige Delight PRWO Electric Rice Cooker (1 ...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	
1462	B009P2LIL4	Bajaj Majesty RX10 2000 Watts Heat Convactor R...	Home&Kitchen Heating,Cooling&AirQuality RoomHe...	
1463	B00J5DYCCA	Havells Ventil Air DSP 230mm Exhaust Fan (Pist...	Home&Kitchen Heating,Cooling&AirQuality Fans E...	

	product id	product name	category	disc
1464	B01486F4G6	Borosil Jumbo 1000-Watt Grill Sandwich Maker (...)	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	

1463 rows × 16 columns

```
In [49]: # Step 6: Convert rating and rating_count to numeric
df['rating'] = pd.to_numeric(df['rating'], errors='coerce')
df['rating count'] = df['rating count'].str.replace(',', '').astype(int)
```

```
In [51]: df
```

Out[51]:

	product id	product name	category	disc
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories Accessories&Peripherals ...	
1	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories Accessories&Peripherals ...	
2	B096MSW6CT	Source Fast Phone Charging Cable & Data Sync U...	Computers&Accessories Accessories&Peripherals ...	
3	B08HDJ86NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories Accessories&Peripherals ...	
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories Accessories&Peripherals ...	
...	
1460	B08L7J3T31	Noir Aqua - 5pcs PP Spun Filter + 1 Spanner ...	Home&Kitchen Kitchen&HomeAppliances WaterPurif...	
1461	B01M6453MB	Prestige Delight PRWO Electric Rice Cooker (1 ...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	
1462	B009P2LIL4	Bajaj Majesty RX10 2000 Watts Heat Convector R...	Home&Kitchen Heating,Cooling&AirQuality RoomHe...	
1463	B00J5DYCCA	Havells Ventil Air DSP 230mm Exhaust Fan (Pist...	Home&Kitchen Heating,Cooling&AirQuality Fans E...	

	product id	product name	category	disc
		Borosil Jumbo		
1464	B01486F4G6	1000-Watt Grill Sandwich Maker (...)	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	

1463 rows × 16 columns

```
In [53]: # Step 7: Print final data types
print("\nData types after cleaning:\n", df.dtypes)
```

Data types after cleaning:

```
product id          object
product name        object
category            object
discounted price    float64
actual price        float64
discount percentage float64
rating              float64
rating count        int32
about product       object
user id             object
user name           object
review id           object
review title        object
review content      object
img link            object
product link        object
dtype: object
```

```
In [55]: # Optional: Save cleaned data
df.to_csv("amazon_cleaned.csv", index=False)
print("\n✅ Cleaned data saved to 'amazon_cleaned.csv'")
```

✅ Cleaned data saved to 'amazon_cleaned.csv'

```
In [57]: df.isnull().sum()
```

```
Out[57]: product id      0
product name      0
category          0
discounted price  0
actual price      0
discount percentage 0
rating            1
rating count      0
about product     0
user id           0
user name         0
review id         0
review title      0
review content    0
img link          0
product link      0
dtype: int64
```

```
In [59]: df['rating'] = df['rating'].fillna(df['rating'].mean())
```

```
In [61]: df.isnull().sum()
```

```
Out[61]: product id      0
product name      0
category          0
discounted price  0
actual price      0
discount percentage 0
rating            0
rating count      0
about product     0
user id           0
user name         0
review id         0
review title      0
review content    0
img link          0
product link      0
dtype: int64
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [66]: import pandas as pd
import numpy as np

# Load the dataset (you can modify this to read your file)
df = pd.read_csv("amazon_cleaned.csv")
df
```

Out[66]:

	product id	product name	category	disc
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories Accessories&Peripherals ...	
1	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories Accessories&Peripherals ...	
2	B096MSW6CT	Source Fast Phone Charging Cable & Data Sync U...	Computers&Accessories Accessories&Peripherals ...	
3	B08HDJ86NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories Accessories&Peripherals ...	
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories Accessories&Peripherals ...	
...	
1458	B08L7J3T31	Noir Aqua - 5pcs PP Spun Filter + 1 Spanner ...	Home&Kitchen Kitchen&HomeAppliances WaterPurif...	
1459	B01M6453MB	Prestige Delight PRWO Electric Rice Cooker (1 ...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	
1460	B009P2LIL4	Bajaj Majesty RX10 2000 Watts Heat Convactor R...	Home&Kitchen Heating,Cooling&AirQuality RoomHe...	
1461	B00J5DYCCA	Havells Ventil Air DSP 230mm Exhaust Fan (Pist...	Home&Kitchen Heating,Cooling&AirQuality Fans E...	

	product id	product name	category	disc
		Borosil Jumbo		
1462	B01486F4G6	1000-Watt Grill Sandwich Maker (...)	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	

1463 rows × 16 columns

```
In [68]: # Preview columns
print(df.columns)
```

```
Index(['product id', 'product name', 'category', 'discounted price',
      'actual price', 'discount percentage', 'rating', 'rating count',
      'about product', 'user id', 'user name', 'review id', 'review title',
      'review content', 'img link', 'product link'],
      dtype='object')
```

```
In [72]: # Step 4: Split multivalue columns (like user_id, user_name, review_id, review_t
def split_multivalues(col):
    return df[col].astype(str).str.split(',')

multi_cols = ['user id', 'user name', 'review id', 'review title', 'review conte
for col in multi_cols:
    df[col] = split_multivalues(col)
```

```
In [74]: # Optional: Create a count of reviews per product
df['num reviews'] = df['user id'].apply(lambda x: len(x))
```

```
In [76]: # Step 5: Clean 'about_product' text
df['about product'] = df['about product'].astype(str).str.replace(r'\|', '. ').s
```

```
In [78]: # Step 6: Text cleaning (optional – useful for NLP tasks)
import re

def clean_text(text):
    text = re.sub(r'^a-zA-Z0-9\s', '', text) # remove special chars
    text = text.lower()
    return text

text_cols = ['about product']
for col in text_cols:
    df[col] = df[col].apply(clean_text)
```

```
In [80]: # Step 7: Handle missing values
df.fillna({'discounted price': 0, 'actual price': 0, 'rating': 0, 'rating count'
```

```
In [88]: # Optional: Save cleaned data as Excel file
df.to_excel('cleaned_products.xlsx', index=False)

print("✅ Preprocessing done! Cleaned Excel file saved as 'cleaned_products.xlsx'
✅ Preprocessing done! Cleaned Excel file saved as 'cleaned_products.xlsx'
```

In [90]: df

Out[90]:

	product id	product name	category	disc
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories Accessories&Peripherals ...	
1	B098NS6PVG	Ambrane Unbreakable 60W / 3A Fast Charging 1.5...	Computers&Accessories Accessories&Peripherals ...	
2	B096MSW6CT	Source Fast Phone Charging Cable & Data Sync U...	Computers&Accessories Accessories&Peripherals ...	
3	B08HDJ86NZ	boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...	Computers&Accessories Accessories&Peripherals ...	
4	B08CF3B7N1	Portronics Konnect L 1.2M Fast Charging 3A 8 P...	Computers&Accessories Accessories&Peripherals ...	
...	
1458	B08L7J3T31	Noir Aqua - 5pcs PP Spun Filter + 1 Spanner ...	Home&Kitchen Kitchen&HomeAppliances WaterPurif...	
1459	B01M6453MB	Prestige Delight PRWO Electric Rice Cooker (1 ...	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	
1460	B009P2LIL4	Bajaj Majesty RX10 2000 Watts Heat Convector R...	Home&Kitchen Heating,Cooling&AirQuality RoomHe...	

	product id	product name	category	disc
1461	B00J5DYCCA	Havells Ventil Air DSP 230mm Exhaust Fan (Pist...	Home&Kitchen Heating,Cooling&AirQuality Fans E...	
1462	B01486F4G6	Borosil Jumbo 1000-Watt Grill Sandwich Maker (...)	Home&Kitchen Kitchen&HomeAppliances SmallKitch...	

1463 rows × 17 columns

```
In [92]: df.columns
```

```
Out[92]: Index(['product id', 'product name', 'category', 'discounted price',  
              'actual price', 'discount percentage', 'rating', 'rating count',  
              'about product', 'user id', 'user name', 'review id', 'review title',  
              'review content', 'img link', 'product link', 'num reviews'],  
              dtype='object')
```

```
In [ ]:
```