Day_12_Task

```
import pandas as pd
import numpy as np
# Step 1: Create Dataset
data = {
    "Customer ID": [101,102,103,104,105,101],
    "Order ID": [1001,1002,1003,1004,1005,1001],
    "Gender": ["m","f","m","f","m","m"],
    "Age": [25, -5, 130, 40, 35, 25],
    "City": ["Hyderabad","Chennai",None,"Mumbai","Chennai",None],
    "Product Name": ["Laptop","Shirt","Mobile","Shoes","Tablet","Laptop"],
    "Product Category": ["Electronics","Clothing
","ELECTRONICS","Footwear","electronics","Electronics"],
    "Quantity": [1,2,1,3,2,1],
    "Price": [50000,1000,20000,3000,15000,50000],
    "Order Date": ["2025-01-10","2025-02-15","2025-03-20","2025-04-25","2025-05-05","2025-01-
10"]
}
df = pd.DataFrame(data)
# Step 2: Remove Duplicate Orders
df = df.drop_duplicates(subset="Order ID")
# Step 3: Fix Gender Column
df["Gender"] = df["Gender"].replace({"m":"male","f":"female"})
# Step 4: Handle Age Issues
median_age = df[(df["Age"] > 0) & (df["Age"] < 100)]["Age"].median()
df["Age"] = df["Age"].apply(lambda x: median_age if x < 0 or x > 100 else x)
# Step 5: Fill Missing City with Mode
mode_city = df["City"].mode()[0]
df["City"] = df["City"].fillna(mode_city)
# Step 6: Standardize Product Category
df["Product Category"] = df["Product Category"].str.strip().str.lower()
# Step 7: Convert Order Date to Datetime
```

```
df["Order Date"] = pd.to_datetime(df["Order Date"])
# Step 8: Extract Month & Day Names
df["Month Name"] = df["Order Date"].dt.month_name()
df["Day Name"] = df["Order Date"].dt.day_name()
# Step 9: Rename Columns to Lowercase
df.columns = df.columns.str.lower()
# Step 10: Calculate Revenue
df["revenue"] = df["quantity"] * df["price"]
total_revenue = df["revenue"].sum()
top_cities = df.groupby("city")["revenue"].sum().sort_values(ascending=False)
category_sales = df.groupby("product category")["revenue"].sum()

# Final Outputs
print("Final Cleaned Data:\n")
print(df)
print("\nTotal Revenue:")
print(total_revenue)
print("\nTop Cities by Sales:")
print(top_cities)
print("\nCategory Wise Sales:")
print(category_sales)
```

```
TERMINAL

PS C:\Users\mouni\Downloads> python ecommerce_project.py
>>
Final Cleaned Data:

   customer id  order id  gender  ...  month name  day name revenue
0          101      1001    male  ...     January    Friday   50000
1          102      1002  female  ...    February  Saturday    2000
2          103      1003    male  ...       March  Thursday   20000
3          104      1004  female  ...       April    Friday    9000
4          105      1005    male  ...         May    Monday   30000

[5 rows x 13 columns]

Total Revenue:
111000

Top Cities by Sales:
city
Chennai      52000
Hyderabad    50000
Mumbai        9000
Name: revenue, dtype: int64

Category Wise Sales:
product category
clothing         2000
electronics    100000
footwear         9000
Name: revenue, dtype: int64
PS C:\Users\mouni\Downloads>
```