Ecommerce sales dash board

Code:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
df = pd.read_csv("/content/Amazon Sale Report.csv.zip")

# Step 1: Data Cleaning
df['Date'] = pd.to_datetime(df['Date'], format='%m-%d-%y', errors='coerce')
df = df[df['Date'].notna() & df['Amount'].notna()]
df['Month'] = df['Date'].dt.to_period('M')

# Step 2: Monthly Sales Trend
monthly_sales = df.groupby('Month')['Amount'].sum()

# Step 3: Category-wise Sales
category_sales = df.groupby('Category')['Amount'].sum().sort_values(ascending=False)

# Step 4: State-wise Sales
state_sales = df.groupby('ship-state')['Amount'].sum().sort_values(ascending=False)

# Step 5: Order Frequency & AOV
order_metrics = df.groupby('Order ID').agg({
    'Amount': 'sum',
```

```
    'Date': 'max'
})
order_metrics['Month'] = order_metrics['Date'].dt.to_period('M')
monthly_orders = order_metrics.groupby('Month')['Amount'].agg(['count', 'mean'])

# Step 6: Cancellation Rate by Category
df['Cancelled'] = df['Status'].str.contains("Cancel", case=False)
cancel_rate = df.groupby('Category')['Cancelled'].mean().sort_values(ascending=False)

# Step 7: Visualizations
plt.figure(figsize=(16, 12))

# Monthly Sales
plt.subplot(2, 2, 1)
monthly_sales.plot(marker='o', color='teal')
plt.title("Monthly Sales Trend")
plt.xlabel("Month")
plt.ylabel("Sales (INR)")
plt.grid(True)

# Category-wise Sales
plt.subplot(2, 2, 2)
category_sales.head(10).plot(kind='barh', color='skyblue')
plt.title("Top 10 Categories by Sales")
plt.xlabel("Sales (INR)")
plt.gca().invert_yaxis()

# Order Frequency & AOV
plt.subplot(2, 2, 3)
monthly_orders['count'].plot(label='Order Count', color='navy')
monthly_orders['mean'].plot(secondary_y=True, label='Avg Order Value (AOV)', color='orange')
```

```
plt.title("Order Count & AOV Over Time")
plt.grid(True)
plt.legend(loc='upper left')

# Cancellation Rate
plt.subplot(2, 2, 4)
sns.barplot(x=cancel_rate.head(10).values, y=cancel_rate.head(10).index, palette='Reds_r')
plt.title("Top 10 Categories by Cancellation Rate")
plt.xlabel("Cancellation Rate")

plt.tight_layout()
plt.show()
```

Output: