


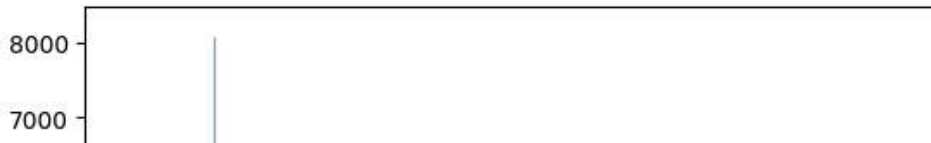
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Load your data into a DataFrame
df = pd.read_csv('/content/amazon_sales_data_cleaned.csv') # Replace 'your_data.csv' with your actual file

# Example: Distribution of a numerical column
sns.histplot(df['Amount'], kde=True)
plt.show()

# Example: Relationship between two variables
sns.scatterplot(x='ship-postal-code', y='Amount', data=df)
plt.show()
```

 <ipython-input-3-194c442c6675>:6: DtypeWarning: Columns (17) have mixed types. Specify dtype option on
 df = pd.read_csv('/content/amazon_sales_data_cleaned.csv') # Replace 'your_data.csv' with your actua



```
# Example: Creating a new feature based on existing ones
df['Date'] = df['Amount'] / df['ship-postal-code']
```


```
df['Date']
```

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
```

```
# Split data into training and testing sets
X = df[['ship-postal-code', 'Date']] # Example features
y = df['Amount'] # Target variable
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
# Train a linear regression model
model = LinearRegression()
model.fit(X_train, y_train)
```

```
# Predict and evaluate
y_pred = model.predict(X_test)
print("Mean Squared Error:", mean_squared_error(y_test, y_pred))
```

 Mean Squared Error: 39903.4982636495

