

# Pandas Analysis of Music, Anime, and Figurine Sales

December 29, 2024

## 1 Dataset

The following dataset contains information about music, anime, and figurine sales.

Table 1: Music, Anime, and Figurine Sales Data

Order ID	Product Name	Category	Quantity Sold	Price per Unit
1	Demon Slayer: Kimetsu no Yaiba - Tanjiro Figure	Figurines	5	12.00
2	Attack on Titan Season 1 DVD Box Set	Anime	3	15.00
3	My Hero Academia: All Might Pop Figure	Figurines	2	10.00
4	Spirited Away Soundtrack Vinyl	Music	1	18.00
5	One Piece: Luffy Figure	Figurines	4	10.00
6	Naruto Shippuden: Kakashi Figure	Figurines	10	8.00
7	Jujutsu Kaisen: Yuuta Okkotsu Figure	Figurines	2	12.00
8	Your Name Blu-ray	Anime	3	10.00
9	Cowboy Bebop: Complete Series DVD	Anime	1	20.00
10	Studio Ghibli: The Art of Spirited Away	Book	4	10.00
11	Demon Slayer: Nezuko Figure	Figurines	6	10.00
12	Attack on Titan Season 2 DVD Box Set	Anime	5	10.00
13	One Punch Man: Saitama Figure	Figurines	2	10.00
14	Spirited Away: Chihiro Figure	Figurines	3	10.00
15	Cowboy Bebop Soundtrack	Music	1	18.00
16	My Neighbor Totoro Plush	Figurines	8	10.00
17	Howl's Moving Castle DVD	Anime	3	10.00
18	Naruto: Shippuden DVD Box Set	Anime	4	10.00
19	Dragon Ball Z: Goku Figure	Figurines	2	10.00
20	Fullmetal Alchemist: Brotherhood DVD Box Set	Anime	5	10.00

## 2 Questions

Here are some analytical questions based on the dataset:

1. What are the top three products in terms of total sales revenue? Can you provide a breakdown of their sales figures?
2. Can you identify which city has the highest number of orders? What does this tell us about our customer base in that location?
3. How do the average ratings differ across product categories (e.g., Figurines, Anime, Music)? Are there any categories that consistently receive higher ratings?
4. Is there any correlation between delivery time and customer ratings? How does delivery time affect customer satisfaction?
5. What is the overall return rate for the dataset? Are there specific products or categories that have a higher return rate?
6. How do sales figures compare between online and retail channels? Which channel performs better, and what might be the reasons for this?
7. Can you segment customers based on their purchase behavior? For example, can you identify high-value customers versus occasional buyers?
8. Based on the historical sales data, can you build a predictive model to forecast next month's sales? What features would you include in your model?
9. If we were to run a promotional campaign on a specific product category, how would you assess its potential impact on sales? What metrics would you track?
10. Are there any observable seasonal trends in the sales data? For example, do certain products sell better during specific months or seasons?

### 3 Explanations and Python Pandas Code

Here are the explanations and Python Pandas code for each question:

#### 3.1 1. Top Three Products by Total Sales

To find the top three products in terms of total sales revenue, we can use the following Pandas code:

```
# Top three products by total sales
top_products = df.groupby('Product_Name')['Total_Sales']
                .sum().nlargest(3)
print(top_products)
```

### 3.2 2. City with Highest Number of Orders

To identify which city has the highest number of orders, we can use:

```
# City with the highest number of orders
top_city = df['City'].value_counts().idxmax()
print("City with the highest number of orders:",
      top_city)
```

### 3.3 3. Average Ratings by Product Category

To compare average ratings across product categories, we can use:

```
# Average ratings by category
average_ratings = df.groupby('Category')['Rating'].
    mean()
print(average_ratings)
```

### 3.4 4. Correlation Between Delivery Time and Customer Ratings

To analyze the correlation between delivery time and ratings, we can use:

```
# Correlation between delivery time and ratings
correlation = df['Delivery Time (Days)'].corr(df['
    Rating'])
print("Correlation between delivery time and ratings:",
      , correlation)
```

### 3.5 5. Overall Return Rate

To calculate the overall return rate, we can use:

```
# Overall return rate
overall_return_rate = df['Return Rate (%)'].mean()
print("Overall return rate:", overall_return_rate)
```

### 3.6 6. Sales Channel Performance

To compare sales figures between online and retail channels, we can use:

```
# Sales figures by channel
sales_channel_performance = df.groupby('Sales Channel'
    )['Total Sales'].sum()
print(sales_channel_performance)
```

### 3.7 7. Customer Segmentation

To segment customers based on their purchase behavior, we can analyze total sales per customer:

```
# Total sales per customer
customer_segmentation = df.groupby('Customer_Name')['Total_Sales'].sum().sort_values(ascending=False)
print(customer_segmentation)
```

### 3.8 8. Predictive Model for Next Month's Sales

For predictive modeling, you would typically export the relevant data and use a machine learning library (like scikit-learn in Python) to build your model. However, you can prepare your dataset with:

```
# Preparing data for predictive modeling
monthly_sales = df.resample('M', on='Order_Date')['Total_Sales'].sum()
print(monthly_sales)
```

### 3.9 9. Assessing Impact of Promotional Campaigns

To assess the potential impact of a promotional campaign, you might analyze sales before and after the campaign:

```
# Sales before and after a promotional campaign
sales_before_after = df[df['Order_Date'] < '2024-01-15']
                        .groupby('Order_Date')['Total_Sales'].sum()
print(sales_before_after)
```

### 3.10 10. Seasonal Trends

To identify seasonal trends, you can group sales by month:

```
# Monthly sales trends
monthly_trends = df.groupby(df['Order_Date'].dt.month)
                  ['Total_Sales'].sum()
print(monthly_trends)
```