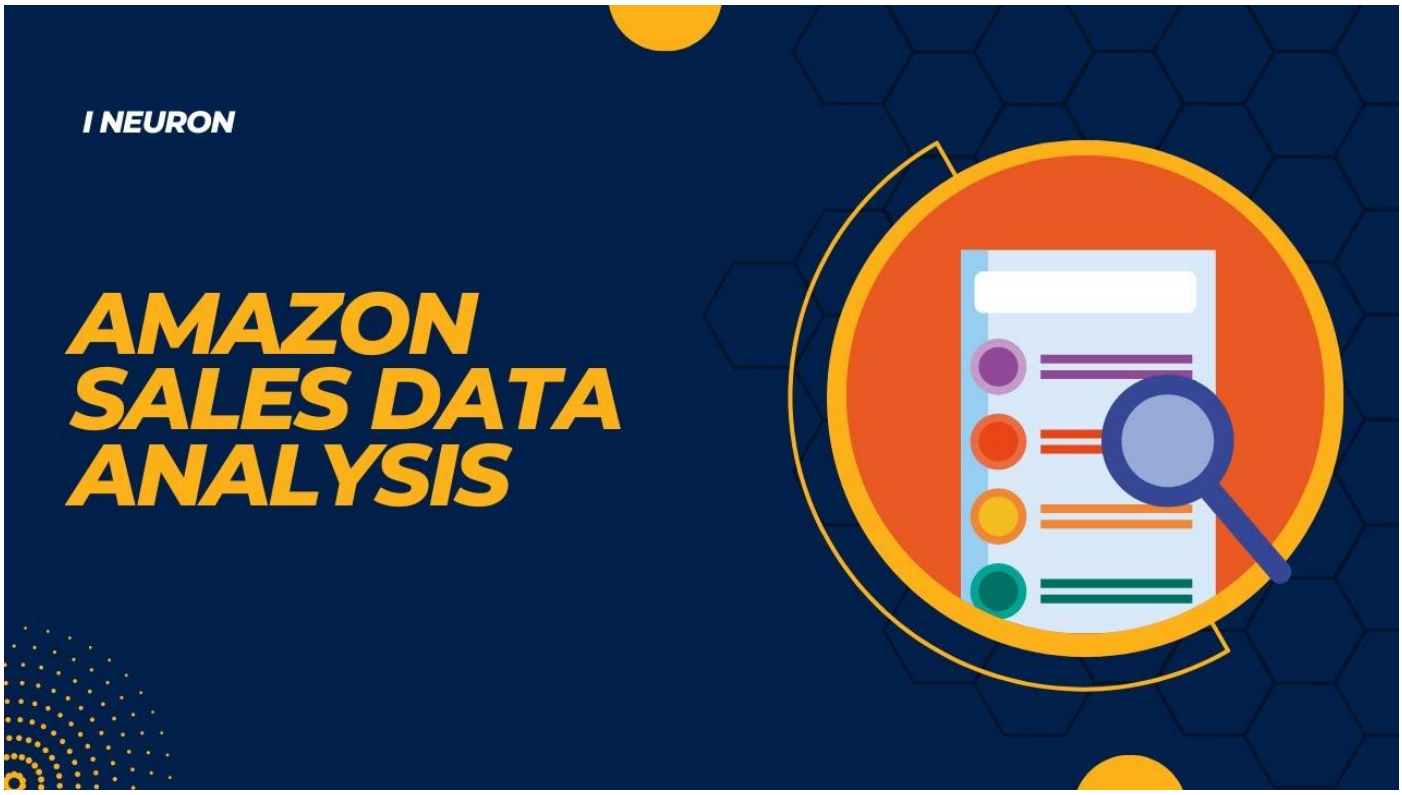
DATA ANALYSIS REPORT OF AMAZON DATA USING PYTHON BY: UDIT

Goal: Increase Revenue and Improve Product Category Performance



The primary goal of this analysis is to increase revenue and enhance the performance of product categories. The focus will be on diversification, reducing dependence on any single category, and increasing revenue for underperforming product categories.

Key Performance Indicators (KPI):

- **Diversification KPI**: Calculate the percentage of revenue generated by the top 3 product categories. The goal is to increase the percentage of revenue for the lowest-performing product category within the top 3 by 5% by the end of Q3 2022.
- **Revenue**: Monitoring the total revenue trends and addressing the decline in Q2 2022.
- **Category Performance**: Analyzing the performance of key product categories such as Set, Kurta, and Western Dress to balance the portfolio.

Preliminary Analysis

Based on an initial analysis of Amazon India's Q2 2022 data, we can draw the following insights:

- **Revenue Trends**:
  The total revenue decreased by 18.77% over the quarter, with May revenue dropping by 9.06% from April and June revenue declining by 10.68% from May. This decline is a major concern and needs to be addressed.
- **Product Category Performance**:
  - The product category "Set" dominates revenue, contributing 49.88% of total revenue.
  - Kurta generates 27.09% of total revenue.
  - Western Dress accounts for 14.28% of total revenue.
  - These categories represent opportunities for both revenue generation and diversification.
- **High-Value Products**:
  The top 5 product categories by average price in dollars are:
  - Set ($9.43)
  - Saree ($9.14)
  - Western Dress ($8.75)
  - Ethnic Dress ($8.26)
  - Top ($6.09)
    These high-value products represent the potential for increased profitability.
- **Order Cancellations and Returns**:
  - The total number of canceled and returned orders was 49,178, representing 17.53% of all orders.
  - 14.22% of orders were canceled, and 1.64% were returned.
  - A focus on reducing cancellations can improve profitability.
- **Customer Segmentation**:
  - Business customers' average order amount is $8.21.
  - Regular customers' average order amount is $7.37.
- **Regional Preferences**:
  The most popular product categories vary by state, suggesting that marketing efforts should be tailored regionally to capitalize on local preferences.

Data Cleaning Operations

To ensure accurate analysis, the dataset underwent the following data cleaning steps:

- **Date Range**: April 1, 2022, to June 29, 2022, was selected for the analysis, excluding data from March due to insufficient entries.
- **Dropped Columns**:
  - `Unnamed: 22` (irrelevant data)
  - `fulfilled-by` (only one value: Amazon Courier "easy-ship")
  - `ship-country` (all data pertains to India)
  - `currency` (all data in INR)
  - `Sales Channel` (assumed to be Amazon)

- drop columns

```
columns_to_drop=['Unnamed: 0','fulfilled-by','ship-country','currency','Sales Channel ']
amazon_data.drop(columns=columns_to_drop,axis=1,inplace=True)
```

- **Handling Duplicates**:
  - Duplicates in the `Order` and `ASIN` columns were removed.
  - drop duplicates

```
[ ] duplited_col=amazon_data.apply(lambda col:col.duplicated().any())
    duplited_col
```

Show hidden output

```
[ ] amazon_data=amazon_data.drop_duplicates(subset=['ASIN'])

[ ] amazon_data=amazon_data.drop_duplicates(subset=['Order ID'])
```

- **Filling Missing Values**:
  - `Courier Status`: Filled missing values with 'Unknown.'

```
amazon_data['Courier Status']=amazon_data['Courier Status'].fillna('Unknown')
```

  - `promotion-ids`: Filled missing values with 'No Promotion.'

```
amazon_data['promotion-ids']=amazon_data['promotion-ids'].fillna('No Promotion')
```

  - `Amount`: Filled missing amounts with 0, as 97% of the orders with missing
  - amounts were canceled.

```
amazon_data['Amount']=amazon_data['Amount'].fillna(0)
```

- **Column Renaming:**
  - `B2B`: Renamed to `customer_type` with values changed to 'business' and 'regular.'
  - renaming columns

```
amazon_data=amazon_data.rename(columns={'B2B':'customer-type'})
```

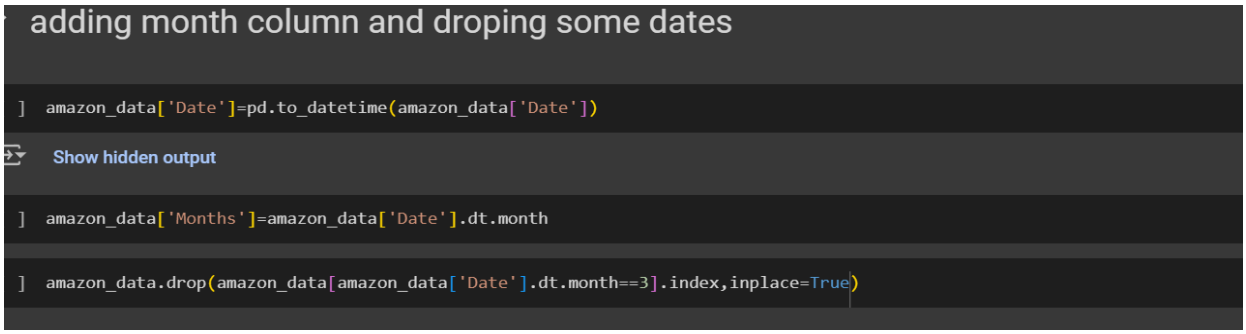
  - `Amount`: Renamed to `order_amount_($)` and converted from INR to USD using the exchange rate 1 INR = 0.0121 USD.

```
amazon_data=amazon_data.rename(columns={'Amount':'order_amount_($)'})
```

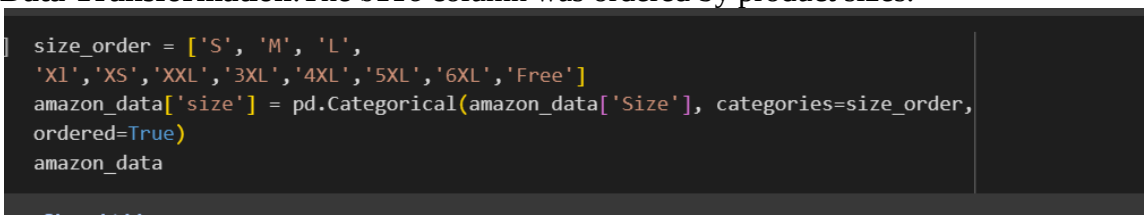
  - conversion INR to USD

```
exchange=0.0120988
amazon_data['order_amount_($)']=amazon_data['order_amount_($)'].apply(lambda x:x*exchange)
```

- **Column Creation**:
  - Created a `month` column to enable grouping and analysis by month.



- **Data Transformation**:The `size` column was ordered by product sizes.



  -

---

final insights and visualizations

Top 3 Categories by Sales

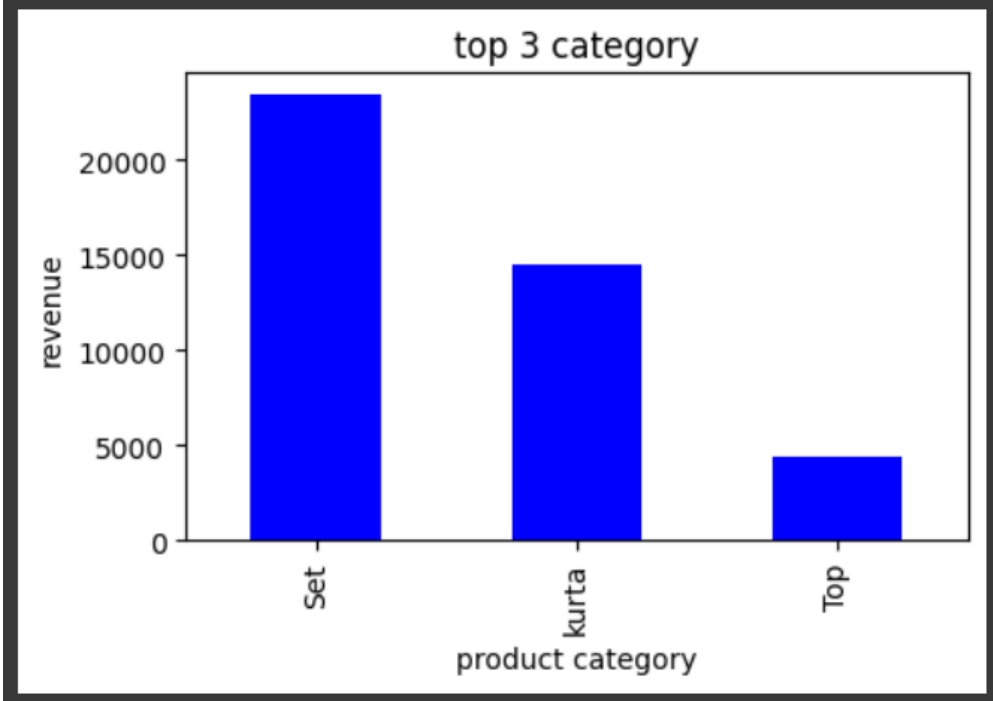
```
#top 3 category_revenue products
category_revenue=amazon_data.groupby('Category')['order_amount_($)'].sum()
top_3=category_revenue.sort_values(ascending=False).head(3)
print(top_3)
```

```
Category
Set      23363.865401
kurta    14430.321395
Top       4409.685327
Name: order_amount_($), dtype: float64
```

```
#percentage_top3 products
total_revenue=category_revenue.sum()
top_3_per=(top_3/total_revenue)*100
print(top_3_per)
```

```
Category
Set      47.829310
kurta    29.541016
Top       9.027282
Name: order_amount_($), dtype: float64
```

```python
#plotting_revenue of top 3 products
plt.figure(figsize=(5,3))
top_3.plot(kind='bar',color='blue')
plt.title('top 3 category')
plt.xlabel('product category')
plt.ylabel('revenue')
plt.show()
```



- **Set**: Total sales of $23363.86, contributing 47.82% to the total revenue.
- **Kurta**: Total sales of $14430.3221, contributing 14.30%.
- **Top**: Contributes 9.02% (category name and amount missing from the dataset).

Interpretation:
These top 3 categories dominate sales, with Set contributing nearly half. Special attention should be given to their performance, especially kurta, to diversify revenue.

Bottom 3 Categories by Sale

```
bottom_3=category_revenue.sort_values().head(3)
print(bottom_3)
```

```
Category
Dupatta      7.380268
Saree      617.323969
Blouse     707.310851
Name: order_amount_($), dtype: float64
```

- **Dupatta**: Sales of $7.380268.
- **Saree**: Sales of $617.32.
- **Blouse**: Sales of $707.31.

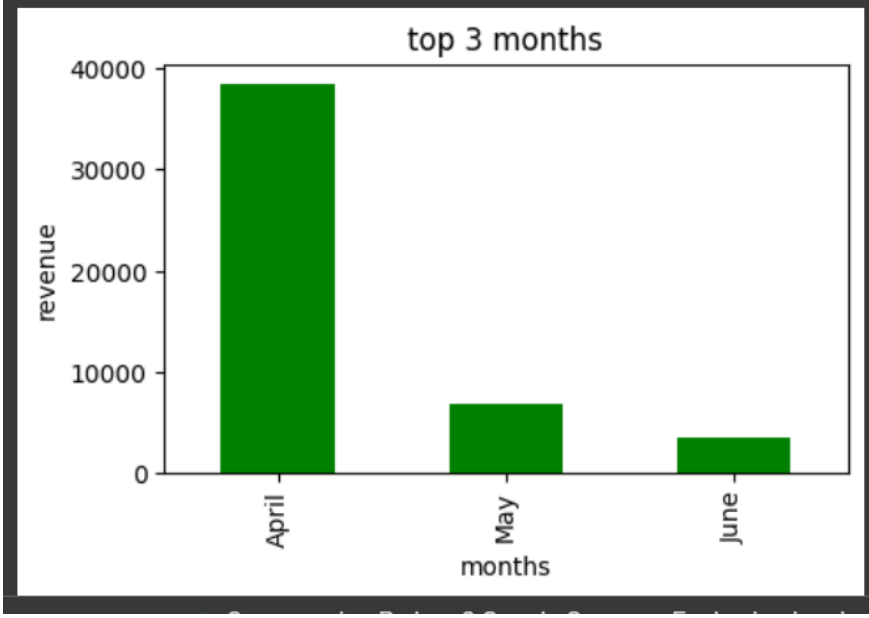
Interpretation:
These categories are underperforming. Targeted marketing and promotions can help boost sales.

Month-wise Revenue

```
monthwise_revenue=amazon_data.groupby('Months')['order_amount_($)'].sum()
top_months=monthwise_revenue.sort_values(ascending=False).head(3)
top_months.index = pd.to_datetime(top_months.index, format='%m').strftime('%B')
print(top_months)
```

```
Months
April    38486.943878
May       6823.806077
June      3537.676900
Name: order_amount_($), dtype: float64
```

```
#plotting_revenue of top 3 months
plt.figure(figsize=(5,3))
top_months.plot(kind='bar',color='green')
plt.title('top 3 months')
plt.xlabel('months')
plt.ylabel('revenue')
plt.show()
```



- **April**: Revenue = $38486.94.
- **May**: Revenue = $6823.80.
- **June**: Revenue = $3557.67.

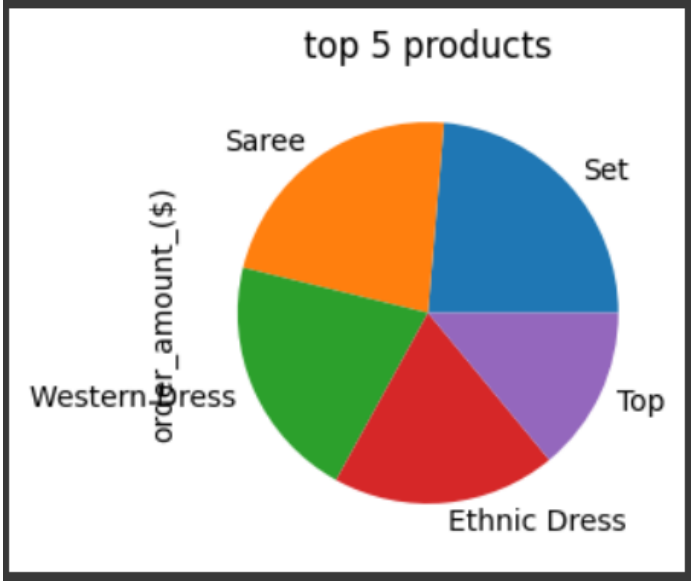
Interpretation:
Month-over-month decline in sales requires interventions like promotions or product launches.

Top 5 Product Categories by Average Price

```python
#top 3 category_revenue products acc to average price
average_category_revenue=amazon_data.groupby('Category')['order_amount_($)'].mean()
top_5_average=average_category_revenue.sort_values(ascending=False).head(5)
print(top_5_average)
```

```
Category
Set              9.674478
Saree            9.213791
Western Dress    8.505658
Ethnic Dress     7.775856
Top              5.726864
Name: order_amount_($), dtype: float64
```

```python
#plotting_revenue_average of top 5 products
plt.figure(figsize=(5,3))
top_5_average.plot(kind='pie')
plt.title('top 5 products')
plt.show()
```



- **Set**: $9.67.
- **Saree**: $9.21.
- **Western Dress**: $8.05.
- **Ethnic Dress**: $7.77.
- **Top**: $5.72.

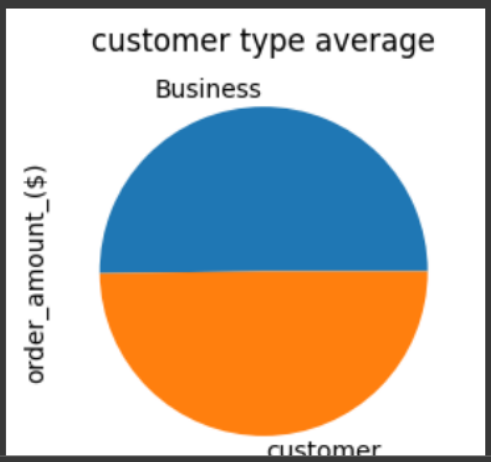
Interpretation:
These are high-value categories. Focusing on these premium products could increase overall revenue.

Average Order Amount by Customer Type

```
#average_revenue acc to customer type
average_ctype_revenue=amazon_data.groupby('customer-type')['order_amount_($)'].mean()
customer_type_average=average_ctype_revenue.sort_values(ascending=False)
print(customer_type_average)
```

```
customer-type
Business    7.102907
customer    7.049373
Name: order_amount_($), dtype: float64
```

```
#plotting_revenue_average of top 5 products
plt.figure(figsize=(5,3))
customer_type_average.plot(kind='pie')
plt.title('customer type average')
plt.show()
```



- **Business customers**: $7.10.
- **Regular customers**: $7.04.

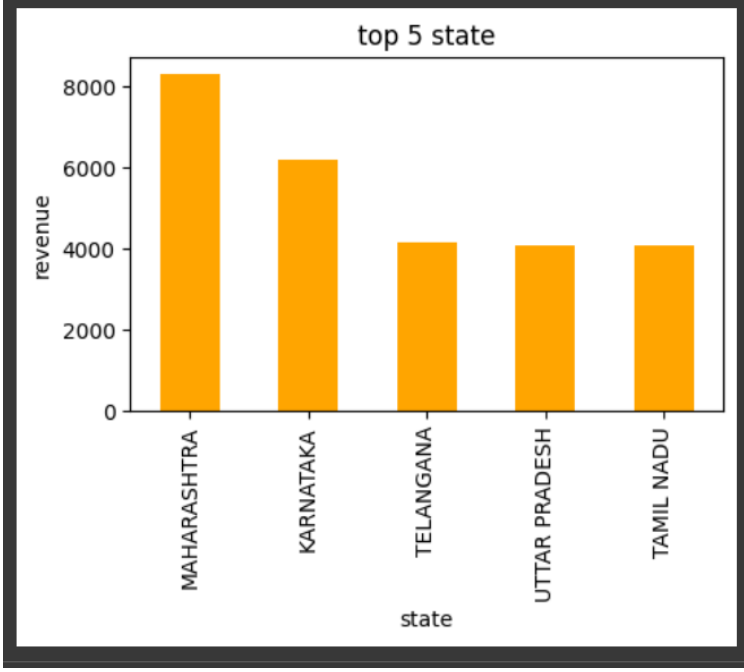
Interpretation:
Business customers tend to spend more. Offering deals or programs targeting business customers could increase purchase volume.

Top 5 States by Sales

```
2]  #top 5 state_revenue acc to average price
    total_state_revenue=amazon_data.groupby('ship-state')['order_amount_($)'].sum()
    top_5_state=total_state_revenue.sort_values(ascending=False).head(5)
    print(top_5_state)
```

```
ship-state
MAHARASHTRA      8320.383839
KARNATAKA        6193.388182
TELANGANA        4140.931416
UTTAR PRADESH    4084.165904
TAMIL NADU       4069.489696
Name: order_amount_($), dtype: float64
```

```
#plotting_revenue of top 5 state
plt.figure(figsize=(5,3))
top_5_state.plot(kind='bar',color='orange')
plt.title('top 5 state')
plt.xlabel('state')
plt.ylabel('revenue')
plt.show()
```



- **Maharashtra**: $8320.38.
- **Karnataka**: $6193.38.
- **Telangana**: $4140.93.
- **Uttar Pradesh**: $4084.16.
- **Tamil Nadu**: $4069.48.

Interpretation:
These states lead in sales, making them key regions for marketing and promotional campaigns.

Promotion-wise Sales

- **With promotion**: 68.2% of sales.
- **Without promotion**: 31.8% of sales.

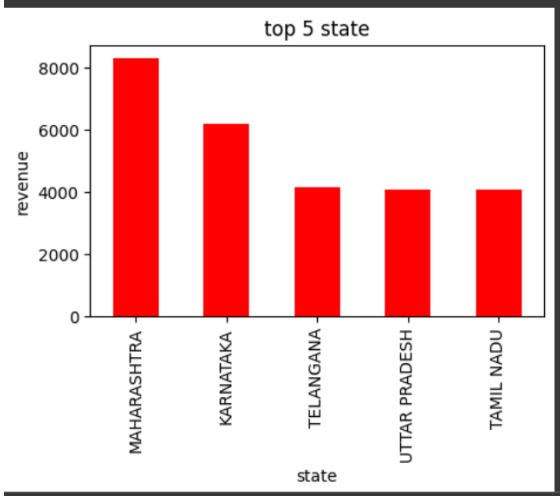
Interpretation:
Promotions drive over two-thirds of sales. Increasing the frequency or targeting them can significantly boost revenue.

Western Dresses Sales by State

```
#top 5 state_revenue acc to western dress
total_state_wes=amazon_data.query("Category=='Western Dress'").groupby('ship-state')['order_amount_($)'].sum()
top_5_state_wes=total_state_wes.sort_values(ascending=False).head(5)
print(top_5_state_wes)
```

```
ship-state
MAHARASHTRA      620.215582
KARNATAKA        533.847451
TELANGANA        475.147582
TAMIL NADU       303.923066
UTTAR PRADESH    238.751549
Name: order_amount_($), dtype: float64
```

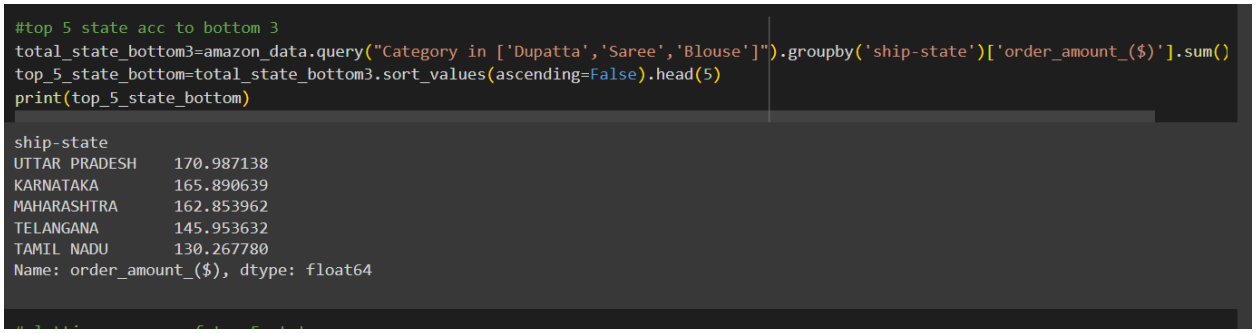
```
#plotting_revenue of top 5 state
plt.figure(figsize=(5,3))
top_5_state.plot(kind='bar',color='red')
plt.title('top 5 state')
plt.xlabel('state')
plt.ylabel('revenue')
plt.show()
```



- **Maharashtra**: $620.21.
- **Karnataka**: $533.84.
- **Telangana**: $475.14.
- **Tamil Nadu**: $303.92.
- **Kerala**: $238.75.

Interpretation:
Western dresses perform well in these states. More targeted marketing in these regions would boost sales further.

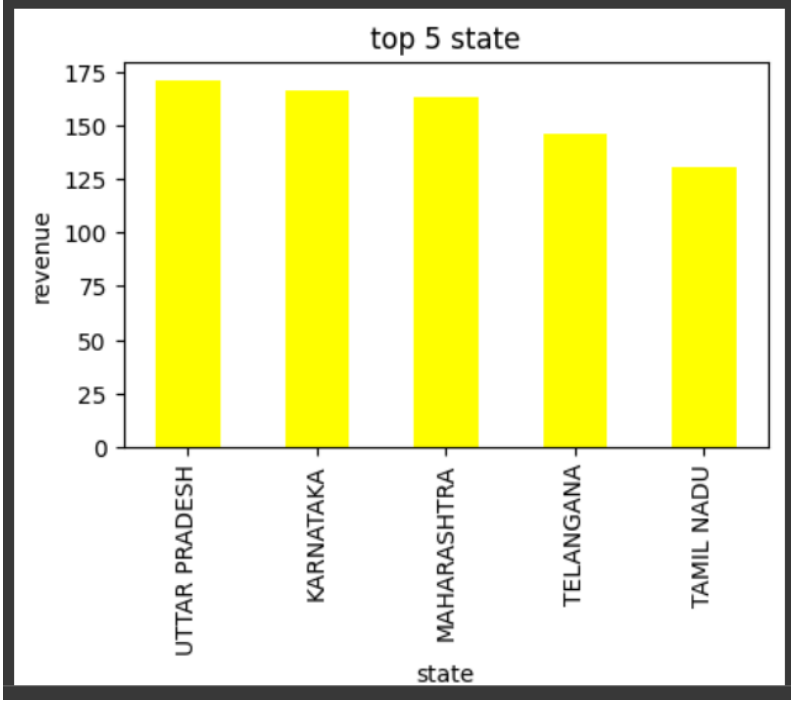
Bottom Three Categories' Performance



```
#top 5 state acc to bottom 3
total_state_bottom3=amazon_data.query("Category in ['Dupatta','Saree','Blouse']").groupby('ship-state')['order_amount_($)'].sum()
top_5_state_bottom=total_state_bottom3.sort_values(ascending=False).head(5)
print(top_5_state_bottom)
```

```
ship-state
UTTAR PRADESH    170.987138
KARNATAKA        165.890639
MAHARASHTRA      162.853962
TELANGANA        145.953632
TAMIL NADU       130.267780
Name: order_amount_($), dtype: float64
```

```
#plotting_revenue of top 5 state
plt.figure(figsize=(5,3))
top_5_state_bottom.plot(kind='bar',color='yellow')
plt.title('top 5 state')
plt.xlabel('state')
plt.ylabel('revenue')
plt.show()
```



- **Dupatta, Saree, and Blouse** perform well in uttar pradesh, Karnataka, Maharashtra ,T elangana and Tamil Nadu.

Interpretation:
Even though these are low-performing categories, they show potential in certain states. Promotions in these regions could improve their sales.

Recommendations

1. **Promote Underperforming Categories**:
   Focus on promotional efforts for the Western Dress category, aiming to increase its revenue share by at least 5%. Target top-performing states with region-specific marketing campaigns.

2. **Reduce Cancellations**:
   Analyze reasons for cancellations and returns. Implement strategies to streamline order fulfillment and reduce canceled orders.
3. **Geographically Targeted Campaigns**:
   Invest in marketing campaigns that align with regional preferences, leveraging state-wise popular product categories to maximize sales.
4. **Monitor Strategy Impact**:
   Track the performance of promotions and revenue growth in real time. Measure the effectiveness of efforts to boost Western Dress sales and reduce cancellations.