Pandas Data Analysis and Business Applications

Task 7: Analysis Report

The initial data for this project was provided as dictionaries, which were converted into Pandas DataFrames. Once the data was structured, various cleaning and transformation steps were performed to ensure consistency and usability.

Missing values in the provided data were identified and the missing values in the "Price" column were replaced with the mean price of the products, and missing "Stock" values were filled with 0.

```
#Fill missing "Price" values with the mean price and "Stock" values with
0.
df_nan.fillna({'Price': df_nan['Price'].mean()}, inplace = True)
df_nan.fillna({'Stock': 0}, inplace=True)
```

‘Total Value’ and ‘Discounted Price’ were created and added to the product data frame where total value was calculated as the stock quantity multiplied by the price, and discounted price after reducing the price by 10%. The column names were then standardized for better usability.

```
# For df_products, add a new column "Total Value" that is the product of
"Price" and "Stock"
df_products['Total Value'] = df_products['Price']*df_products['Stock']

# Create another new column in df_products named "Discounted Price" that
applies a 10% discount to "Price".
df_products['Discounted Price'] = df_products['Price']*0.9

df_products.rename(columns={'Product': 'Item', 'Price' : 'Unit Price',
'Stock': 'Inventory'}, inplace=True)
```

Along with this, the data was filtered to display all products priced over $300 as well as those that contain the letter ‘a’ in their names using numeric and string-based filtering.

```
pricey_items = df_products[df_products['Unit Price']>300]
items_with_a      =      df_products[df_products['Item'].str.contains('a',
case=False)]
```

Sales data was grouped by region to find the total and average sales which helps to identify the highest and lowest sales regions: South as the highest, and East being the lowest sales region. Average sales varied for every region showing differences in demand. This information would help in the formulation of regional sales strategies as well as identifying low-selling areas and areas of improvement.

```
#Group df_sales by "Region" and calculate the total sales for each region.
```

```
grouped_region = df_sales.groupby('Region').sum('Sales')
print(f"Total Sales by Region:\n {grouped_region} \n")

#Group df_sales by "Region" and compute both the sum and average of sales.
aggregated = df_sales.groupby('Region').agg({'Sales': ['sum', 'mean']})
print(f"\n Total Sales and Mean by Region:\n {aggregated}")
```

The customer and order data was merged using an inner join on Customer ID to combine relevant details about orders placed with the business. This helps to analyze customer buying behavior.

```
#Merge df_customers and df_orders on "CustomerID" using an inner join
merged = pd.merge(df_customers, df_orders, on='CustomerID', how='inner')
print(f"Customers and Orders Data: \n{merged}")
```

Finally, pivot tables for regional sales summary provided a structured view of sales distribution. This helps to create a clearer view of regional sales trends helping businesses efficiently allocate their resources.

```
#Create a pivot table from df_sales summarizing the total sales by region
pivot = pd.pivot_table(df_sales, values='Sales', index='Region',
aggfunc='sum')
print(f"Total Sales by Region:\n {pivot}\n")
```

The techniques applied in this analysis are widely used in business settings, helping companies optimize operations and make data-driven decisions. Performing aggregate calculations helps businesses track stock along with allowing them to implement sales strategies effectively. Grouping and aggregation provide valuable insights into which regions or locations generate the most revenue. Pivot tables allow for quick analysis of different metrics, such as revenue per region or product category. Merging data enables businesses to track customer behavior, personalize offers, and improve customer satisfaction.

The step-by-step cleaning, transformation, and analysis of this dataset demonstrated essential data manipulation techniques used in business analytics. Grouping, merging, and pivot tables provided key insights into regional sales performance and customer purchase behavior. These techniques are powerful tools that companies use to enhance decision-making, optimize sales, and improve overall business efficiency.