FUNDAMENTALS OF DATA SCIENCE

LAB EXPERIMENTS DAY 2

DATE: 07/01/2026

DHARANI M

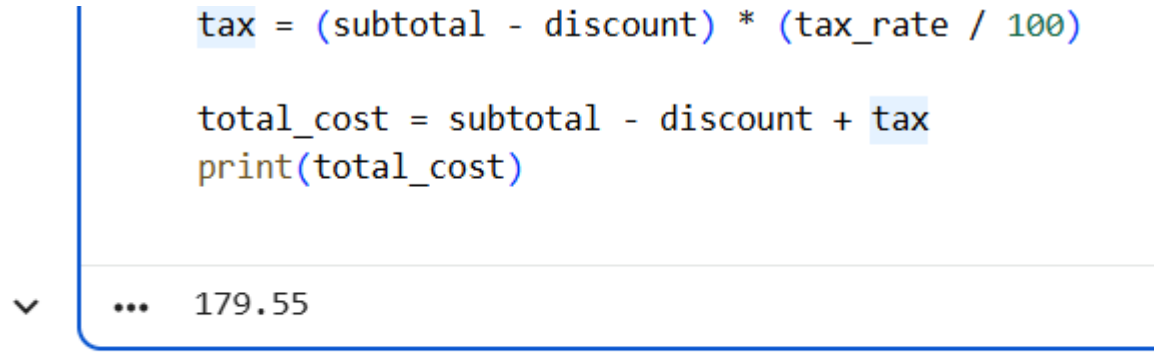
192324008

1. Question: Use arithmetic operations to calculate the total cost of a customer's purchase, including discounts and taxes, given the item prices, quantities, discount rate, and tax rate?

CODE:

```
prices = [50, 30, 20]

quantities = [2, 1, 3]

discount_rate = 10

tax_rate = 5

subtotal = sum(p * q for p, q in zip(prices, quantities))

discount = subtotal * (discount_rate / 100)

tax = (subtotal - discount) * (tax_rate / 100)


total_cost = subtotal - discount + tax

print(total_cost)
```

OUTPUT:



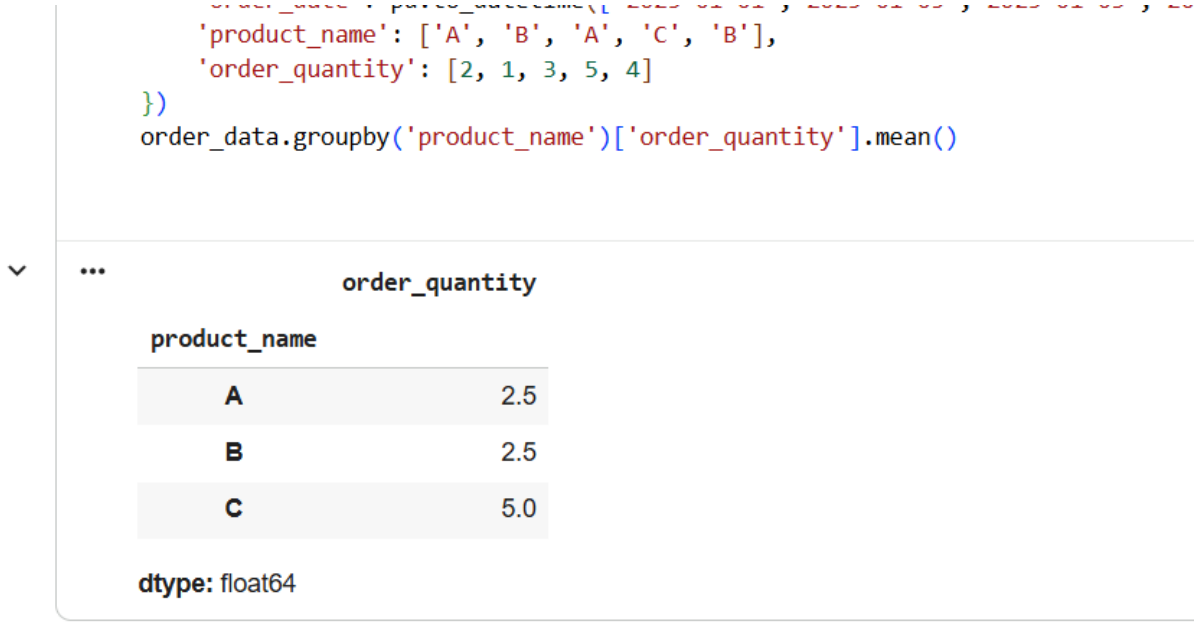
2. Question: Using Pandas DataFrame operations, how would you find the following information

from the order_data DataFrame:

CODE:

```
import pandas as pd


order_data = pd.DataFrame({
   'customer_id': [1, 1, 2, 2, 3],
   'order_date': pd.to_datetime(['2025-01-01','2025-01-05','2025-01-03','2025-01-10','2025-01-07']),
   'product_name': ['A', 'B', 'A', 'C', 'B'],
   'order_quantity': [2, 1, 3, 5, 4]
})
order_data.groupby('product_name')['order_quantity'].mean()
```

OUTPUT:



```
    'product_name': ['A', 'B', 'A', 'C', 'B'],
    'order_quantity': [2, 1, 3, 5, 4]
})
order_data.groupby('product_name')['order_quantity'].mean()
```

product_name	order_quantity
A	2.5
B	2.5
C	5.0

dtype: float64

3. Question: How would you find the top 5 products that have been sold the most in the past month?

CODE:

```
import pandas as pd


order_data = pd.DataFrame({
   'product_name': ['A', 'B', 'A', 'C', 'B', 'A'],
   'order_quantity': [2, 4, 3, 5, 1, 6]
})


order_data.groupby('product_name')['order_quantity'] \.sum() \.sort_values(ascending=False) \.head(5)
```

OUTPUT:

product_name	order_quantity
A	11
B	5
C	5

dtype: int64

4. Question: Using Pandas DataFrame operations, how would you find the following information from the property_data DataFrame:

CODE:

```
import pandas as pd


property_data = pd.DataFrame({
   'property_id': [101, 102, 103, 104],
   'location': ['CityA', 'CityB', 'CityA', 'CityC'],
   'bedrooms': [3, 5, 6, 4],
   'area_sqft': [1200, 2000, 2500, 1800],
   'listing_price': [5000000, 8000000, 9000000, 7000000]
})


avg_price = property_data.groupby('location')['listing_price'].mean()
count_4plus = len(property_data[property_data['bedrooms'] > 4])
largest_area = property_data.loc[property_data['area_sqft'].idxmax()]


avg_price, count_4plus, largest_area
```

OUTPUT:

```
(location
 CityA    7000000.0
 CityB    8000000.0
 CityC    7000000.0
 Name: listing_price, dtype: float64,
 2,
 property_id          103
 location           CityA
 bedrooms               6
 area_sqft           2500
 listing_price    9000000
 Name: 2, dtype: object)
```

5. Question: Using Pandas DataFrame operations, how would you find the following information from the property_data DataFrame:

CODE:

```
import matplotlib.pyplot as plt


months = ['Jan', 'Feb', 'Mar', 'Apr']

sales = [200, 300, 250, 400]


plt.plot(months, sales)

plt.xlabel("Month")

plt.ylabel("Sales")

plt.title("Monthly Sales Line Plot")

plt.show()


plt.bar(months, sales)

plt.xlabel("Month")

plt.ylabel("Sales")

plt.title("Monthly Sales Bar Plot")

plt.show()
```

OUTPUT:

```
plt.bar(months, sales)
plt.xlabel("Month")
plt.ylabel("Sales")
plt.title("Monthly Sales Bar Plot")
plt.show()
```

...