Data science homework 4 - Mayank Sharma - ms14662

```
import pandas as pd

# Sample Data Generation
sales_data = {'Date': ['2023-03-18']*3 + ['2023-01-15']*2,
              'Order_id': [101, 102, 103, 104, 105],
              'Item_id': [201, 202, 203, 201, 202],
              'Customer_id': [1, 2, 3, 1, 2],
              'Quantity': [2, 1, 4, 1, 2],
              'Revenue': [200, 150, 400, 100, 300]}

items_data = {'Item_id': [201, 202, 203],
              'Item_name': ['Item_A', 'Item_B', 'Item_C'],
              'price': [100, 150, 100],
              'department': ['Dept1', 'Dept2', 'Dept3']}

customers_data = {'customer_id': [1, 2, 3],
                  'first_name': ['John', 'Jane', 'John'],
                  'last_name': ['Doe', 'Smith', 'Doe'],
                  'Address': ['Addr1', 'Addr2', 'Addr3']}

sales = pd.DataFrame(sales_data)
items = pd.DataFrame(items_data)
customers = pd.DataFrame(customers_data)

# Q1: Total orders completed on 18th March 2023
q1 = sales[sales['Date'] == '2023-03-18'].shape[0]

# Q2: Total orders by 'John Doe' on 18th March 2023
merged_q2 = pd.merge(sales, customers, left_on='Customer_id', right_on='customer_id')
q2 = merged_q2[(merged_q2['Date'] == '2023-03-18') &
               (merged_q2['first_name'] == 'John') &
               (merged_q2['last_name'] == 'Doe')].shape[0]

# Q3: Total customers and average spend in January 2023
jan_data = sales[sales['Date'].str.startswith('2023-01')]
q3_total_customers = jan_data['Customer_id'].nunique()
q3_avg_spend = jan_data['Revenue'].mean()

# Q4: Departments with revenue less than $600 in 2022
merged_q4 = pd.merge(sales, items, on='Item_id')
q4 = merged_q4.groupby('department')['Revenue'].sum()
q4_result = q4[q4 < 600]
```

```
# Q5: Most and least revenue generated by an order
q5_max = sales.groupby('Order_id')['Revenue'].sum().max()
q5_min = sales.groupby('Order_id')['Revenue'].sum().min()

# Q6: Orders in the most lucrative order
max_order_id = sales.groupby('Order_id')['Revenue'].sum().idxmax()
q6 = sales[sales['Order_id'] == max_order_id]
```

To display the results:

```
print("Q1:", q1)
print("Q2:", q2)
print("Q3 - Total Customers:", q3_total_customers, ", Avg Spend:", q3_avg_spend)
print("Q4:", q4_result)
print("Q5 - Most Revenue:", q5_max, ", Least Revenue:", q5_min)
print("Q6:", q6)
```