QUIZ

Data Mining

2025

Quiz (1)

- You are given an Excel file named "students.xlsx" with a sheet named "Sheet1" containing the following columns: Name, Age, Grade, Score.
- Questions:
  1. Read the Excel file into a DataFrame.
  2. Display the first 3 students info.
  3. Calculate the average score of all students.
  4. Draw a histogram of the "Score" column.
  5. Draw a boxplot of the "Age" column.
  6. Count the number of students in each grade

```
# Sample Data for Excel File
data = {
    "Name": ["Alice", "Bob", "Charlie", "David", "Eva"],
    "Age": [20, 21, 19, 22, 20],
    "Grade": ["A", "B", "A", "C", "B"],
    "Score": [95, 85, 90, 70, 80]
}
df = pd.DataFrame(data)

# Save the DataFrame to an Excel file
df.to_excel("students.xlsx", index=False, sheet_name="Sheet1")
```

Quiz (2)

- You are given an Excel file named "sales.xlsx" with a sheet named "Sheet1" containing the following columns: Product, Region, Sales.
- Questions:
  1. Read the Excel file into a DataFrame.
  2. Display the DataFrame's info.
  3. Calculate the total sales for each region.
  4. Draw a histogram of the "Sales" column.
  5. Draw a boxplot of the "Sales" column.
  6. Find the product with the highest sales.

```
# Sample Data for Excel File
data = {
    "Product": ["A", "B", "C", "D", "E", "F", "G"],
    "Region": ["North", "South", "North", "East", "West", "South", "North"],
    "Sales": [200, 150, 300, 250, 400, 350, 100]
}
df = pd.DataFrame(data)

# Save the DataFrame to an Excel file
df.to_excel("sales.xlsx", index=False, sheet_name="Sheet1")
```

Quiz (3)

- You are given a Database file named "school.db" contains a table named students with the following columns: Name, Age, Grade, Score.
- Questions:
  1. List the "students" whose names start with ‘B’.
  2. Sort the students based on their scores and age in descending order.
  3. Calculate the total number of students for each grade.
  4. Draw a histogram of the "Score" column.
  5. Draw a boxplot of the "Age" column.
  6. Find the students who the minimum score.

```
# Connect to the database (or create it if it doesn't exist)
conn = sqlite3.connect("school.db")
cursor = conn.cursor()

# Create the "students" table
cursor.execute("""
CREATE TABLE IF NOT EXISTS students (
    Name TEXT,
    Age INTEGER,
    Grade TEXT,
    Score INTEGER
)
""")

# Insert sample data into the "students" table
students_data = [
    ("Alice", 20, "A", 95),
    ("Bob", 21, "B", 85),
    ("Charlie", 19, "A", 90),
    ("David", 22, "C", 70),
    ("Eva", 20, "B", 80)
]
cursor.executemany("INSERT INTO students VALUES (?, ?, ?, ?)", students_data)

# Commit the changes and close the connection
conn.commit()
conn.close()
```

Quiz (4)

- You are given an Database file named “company.db" contains a table named students with the following columns: Name, Department, Salary.
- Questions:
  1. List the "employees" who work as HR.
  2. Display the employees whose salary in range 60000 and 70000.
  3. Calculate the total salary for each department.
  4. Draw a histogram of the "Salary" column.
  5. Draw a boxplot of the "Salary" column.
  6. Find the department with the highest average salary.

```
# Connect to the database (or create it if it doesn't exist)
conn = sqlite3.connect("company.db")
cursor = conn.cursor()

# Create the "employees" table
cursor.execute("""
CREATE TABLE IF NOT EXISTS employees (
    Name TEXT,
    Department TEXT,
    Salary INTEGER
)
""")

# Insert sample data into the "employees" table
employees_data = [
    ("Alice", "HR", 50000),
    ("Bob", "Finance", 60000),
    ("Charlie", "HR", 55000),
    ("David", "IT", 70000),
    ("Eva", "Finance", 65000)
]
cursor.executemany("INSERT INTO employees VALUES (?, ?, ?)", employees_data)

# Commit the changes and close the connection
conn.commit()
conn.close()
```