**NEDB2101 – Database Administration**

**Semester 1, 2024-2025**

**Course Coordinator: Dr. Murtaza Ali Khan**

**Lab 8 - Data Visualization using SQLite, Python and Google Colab**

**Objectives**

* Load a database in Python Using sqlite
* Apply SQL queries to the database
* Read the table of database in panda’s dataframe
* Use functions of Python libraries such as panda, matplotlib, and seaborn for statistical analysis and visualization of the records in the table

**Down the Database File**

Download the database file from the following link and store in your local hard disk

[**https://drive.google.com/file/d/1nNNWb-2l09P8hwWZcUci3hWUzgCUt9Ve/view?usp=sharing**](https://drive.google.com/file/d/1nNNWb-2l09P8hwWZcUci3hWUzgCUt9Ve/view?usp=sharing)

**Upload the Database File in Google Colab**

1. Create a new notebook in the Colab
2. On the left side menu click on the **Files** option
3. Click on **Upload to session storage** icon
4. Choose and TRADE\_DB which you downloaded. It will be uploaded to the session storage

|  |  |  |
| --- | --- | --- |
| **1 (Files)** | **2 (Upload to session storage)** | **3 (TRADE\_DB)** |
|  |  |  |

**TRADE\_DB**

1. It is the database
2. The TRADE\_DB is an existing database in the current working directory
3. The TRADE\_DB has a table called ITEM
4. The TRADE table has following five fields (columns):
   * ItemCode
   * PurchaseDate
   * ItemName
   * ItemPrice
   * Category

**Python libraries for data handling and visualization**

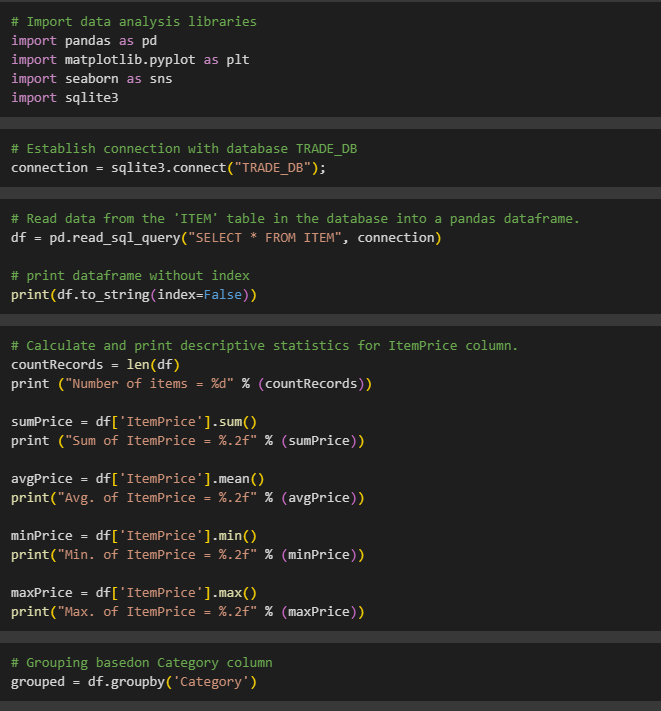
Python has following libraries for data handling and visualization:

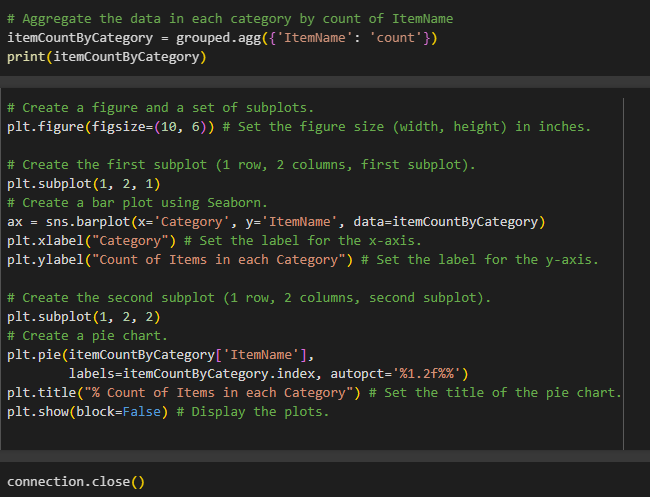
* Pandas (import pandas as pd): Used for data manipulation and analysis, particularly with DataFrames.
* Matplotlib (import matplotlib.pyplot as plt): A library for creating static and interactive visualizations.
* Seaborn (import seaborn as sns): Built on Matplotlib, it simplifies the creation of attractive statistical graphics.
* SQLite3 (import sqlite3): Allows interaction with SQLite databases for executing SQL commands and managing data.

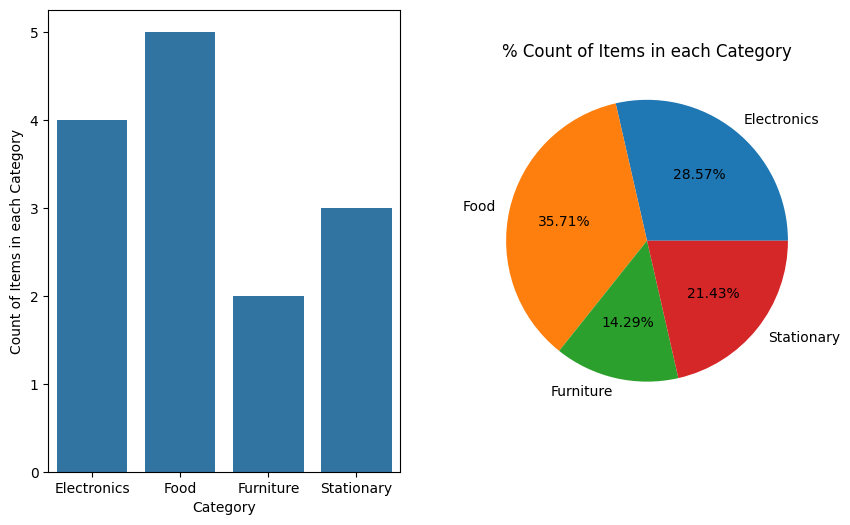
**Lab Task**

1. Open the new notebook in Colab and write the following code.
2. Save your code with name **Sqlite\_Python\_DataVisualization**
3. Run the code

We will discuss this code in the lab







**Loading, processing, and visualization of data in Python**

1. Import necessary libraries: pandas, matplotlib.pyplot, seaborn, and sqlite3 are imported for data analysis, visualization, and database interaction.
2. Connect to the database: A connection is established to the SQLite database file named "TRADE\_DB".
3. Read data: Data from the 'ITEM' table in the database is read into a pandas DataFrame called df.
4. Descriptive statistics: The code calculates and prints descriptive statistics (count, sum, average, minimum, maximum) for the 'ItemPrice' column of the DataFrame.
5. Grouping by category: The DataFrame is grouped by the 'Category' column.
6. Aggregation: The code counts the number of items in each category based on the 'ItemName' column and stores the results in a new DataFrame called itemCountByCategory.
7. A bar plot is created to visualize the count of items in each category.
8. A pie chart is created to show the percentage distribution of items across categories.
9. Close connection: The connection to the database is closed.

**Summary:**

In essence, the code reads data from a database, performs basic data analysis, and then visualizes the distribution of items across different categories using a bar plot and a pie chart.