

Assignment 3: Understanding Customer Tipping Behavior

Dataset description

The dataset contains information about restaurant bills and tips. Each row represents **one dining group**.

- Key variables:
 - `total_bill` : total cost of the meal
 - `tip` : tip amount
 - `sex` : gender of the person paying
 - `smoker` : whether the group included smokers
 - `day` : day of the week
 - `time` : lunch or dinner
 - `size` : group size
-

Let us make it more interesting:

You are a **data analyst working for a restaurant chain**.

Management is curious about **customer tipping behavior** and wants data-driven insights:

- Do people tip more on weekends?
- Does gender or smoking status affect tips?
- How does the total bill influence tipping?

To answer these questions, you will analyze the famous **Tips dataset** from Seaborn and gradually clean, explore, and visualize the data until you reach a meaningful conclusion.

Let's make this more interesting:

Imagine you are a data analyst working for a restaurant chain. Management is curious about customer tipping behavior and has asked you to provide data driven insights. Specifically, they want to know:

- Do customers tip more on weekends?
- Does gender or smoking status influence tipping?
- How does the total bill affect the size of the tip?

Your task is to clean, explore, and visualize the data step by step, and then present meaningful conclusions based on your findings.

Q1: Loading and Understanding the Dataset (10 marks)

1. Load the `tips` dataset from seaborn.
2. Display the first 5 rows.
3. Print:
 - Dataset shape
 - Column names

```
In [5]: #import lib
```

```
In [2]: #Q1.1: Load dataset
```

```
In [3]: #Q1.2: Display first rows
```

```
In [4]: #Q1.3: Dataset shape and columns
```

Q2: Basic Filtering – Who Tips More? (25 marks)

1. Filter only **weekend customers** (Saturday & Sunday). You may create two datasets, `weekend` and `weekdays`
2. Among them, find:
 - Average tip
 - Average total bill
3. Compare this with weekday customers.

```
In [ ]:
```

```
In [7]: #Q2.1: Weekend
```

```
In [8]: # weekday
```

```
In [9]: #Q2.2: Average tip, Average total bill
```

Answer Q2.3: Compare this with weekday customers.

--

Q3: Sorting and Identifying High-Value Customers (15 marks)

1. Sort customers by `total_bill` (descending).
2. Display the **top 10 highest bills**.
3. What do you notice about their group sizes?

```
In [17]: # Q3.1: Sort customers by total_bill (descending).
```

```
In [19]: # Q3.2: Display the top 10 highest bills.
```

```
In [ ]:
```

Answer Q3.3: What do you notice about their group sizes?

--

Q4: Data Quality Check & Cleaning (10 marks)

1. Check if the dataset has missing values.
2. **Why missing values are dangerous** in machine learning.

```
In [22]: #Q4.1: missing values
```

Answer Q4.2: Why missing values are dangerous in machine learning.

--

Q5: Visualization – What Influences Tips? (25 marks)

1. Plot:
 - Tip vs Total Bill (scatter plot)
 - Average tip by day (bar plot)
 - Tip distribution (histogram)
2. Interpret each plot in 1–2 sentences.

```
In [26]: #Q5.1: Tip vs Total Bill (scatter plot)
```

```
In [28]: # Q5.1: Average tip by day (bar plot)
```

```
In [30]: # Q5.1: Tip distribution (histogram)
```

Answer Q5.2: Interpret each plot in 1–2 sentences.

--

Q6: Final Insight & Conclusion (25 marks)

In 6–8 sentences, summarize:

- Who tips more?
- When tips are highest?
- How this insight could help restaurant management.

Answer Q6:

Customers --

Restaurant managers --

In []: