

```
In [1]: import pandas as pd  
  
df = pd.read_csv("Expense_data.csv.csv")  
df.head()
```

```
Out[1]:
```

	Date	Item	Amount	Category	Time	day
0	1/3/2023	chai	7	alone	7:00	Wednesday
1	1/3/2023	chai	20	friend	10:00	Wednesday
2	1/3/2023	juice	15	friend	13:00	Wednesday
3	1/3/2023	rikshow	12	alone	14:00	Wednesday
4	1/3/2023	coffee	12	alone	15:00	Wednesday

```
In [2]: df.shape
```

```
Out[2]: (145, 6)
```

```
In [3]: df.columns
```

```
Out[3]: Index(['Date', 'Item', 'Amount', 'Category', 'Time', 'day'], dtype='object')
```

```
In [4]: df.isnull().sum()
```

```
Out[4]: Date      0  
Item       0  
Amount     0  
Category   1  
Time       0  
day        0  
dtype: int64
```

```
In [5]: df.head(10)
```

Out[5]:

	Date	Item	Amount	Category	Time	day
0	1/3/2023	chai	7	alone	7:00	Wednesday
1	1/3/2023	chai	20	friend	10:00	Wednesday
2	1/3/2023	juice	15	friend	13:00	Wednesday
3	1/3/2023	rikshow	12	alone	14:00	Wednesday
4	1/3/2023	coffee	12	alone	15:00	Wednesday
5	1/3/2023	chai with snaks	25	alone	17:00	Wednesday
6	1/3/2023	coldrink	10	alone	21:30	Wednesday
7	1/3/2023	others	10	alone	12:00	Wednesday
8	2/3/2023	chai	7	alone	7:00	Thursday
9	2/3/2023	juice	15	friend	13:00	Thursday

In [9]: `df['Category'] = df['Category'].fillna("Other")`In [10]: `df.isnull().sum()`

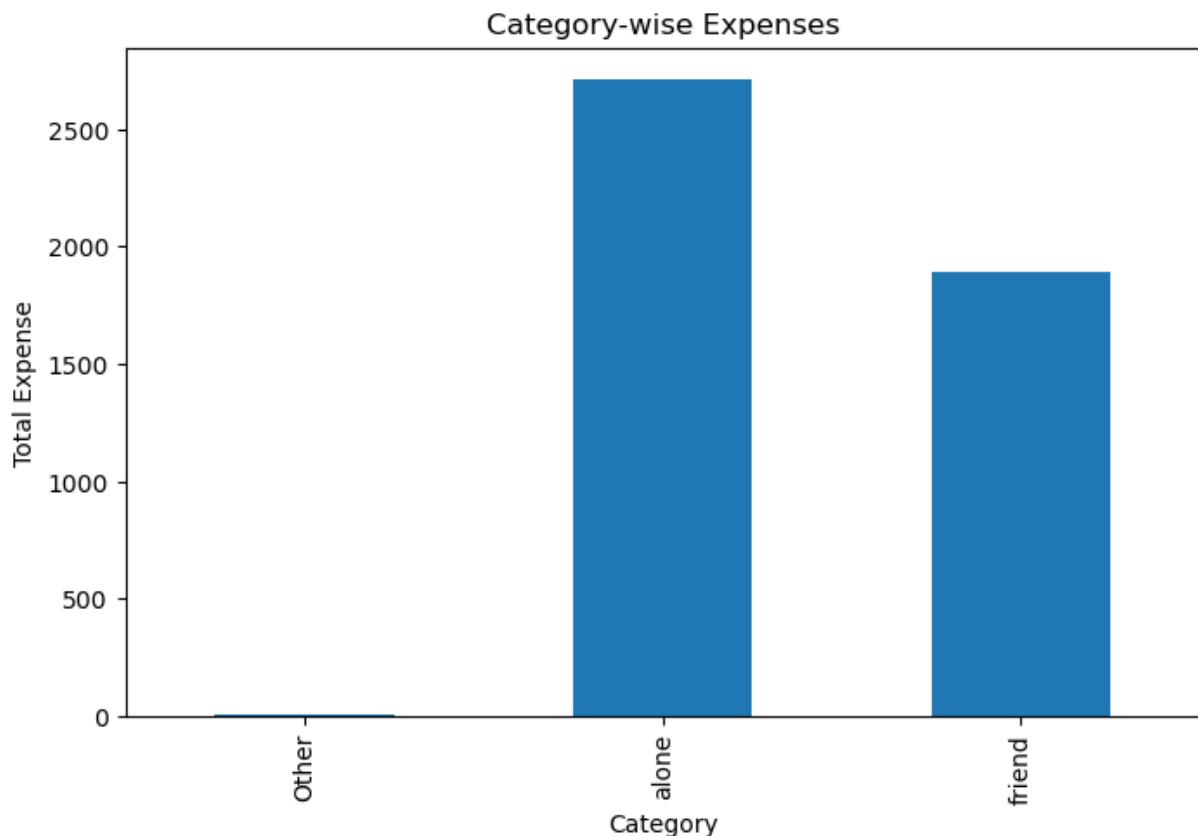
Out[10]:

In [11]: `category_expense = df.groupby("Category")["Amount"].sum()  
category_expense`

Out[11]:

In [12]: `import matplotlib.pyplot as plt`

```
plt.figure(figsize=(8,5))
category_expense.plot(kind="bar")
plt.title("Category-wise Expenses")
plt.xlabel("Category")
plt.ylabel("Total Expense")
plt.show()
```



```
In [14]: df['Date'] = pd.to_datetime(df['Date'], errors='coerce')
```

```
In [15]: df.columns
```

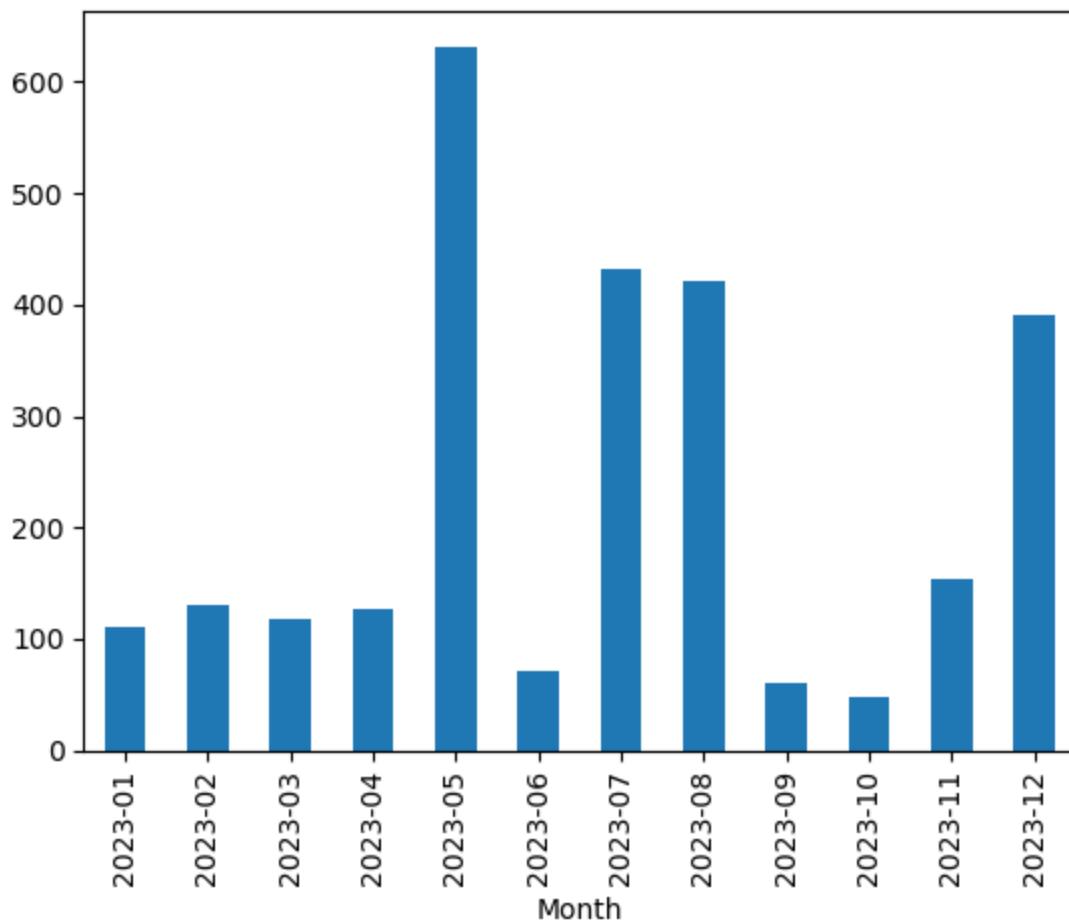
```
Out[15]: Index(['Date', 'Item', 'Amount', 'Category', 'Time', 'day'], dtype='object')
```

```
In [16]: df['Month'] = df['Date'].dt.to_period('M')
monthly_expense = df.groupby('Month')['Amount'].sum()
monthly_expense
```

```
Out[16]: Month
2023-01    111
2023-02    131
2023-03    117
2023-04    126
2023-05    632
2023-06     71
2023-07    432
2023-08    422
2023-09     60
2023-10     47
2023-11    154
2023-12    390
Freq: M, Name: Amount, dtype: int64
```

```
In [17]: monthly_expense.plot(kind='bar')
```

```
Out[17]: <Axes: xlabel='Month'>
```



```
In [18]: category_insights = df.groupby('Category')['Amount'].sum().sort_values(ascending=False)
category_insights
```

```
Out[18]: Category
alone    2710
friend   1889
Other     10
Name: Amount, dtype: int64
```

```
In [19]: highest = df.groupby("Category")["Amount"].sum().idxmax()
lowest = df.groupby("Category")["Amount"].sum().idxmin()

highest, lowest
```

```
Out[19]: ('alone', 'Other')
```

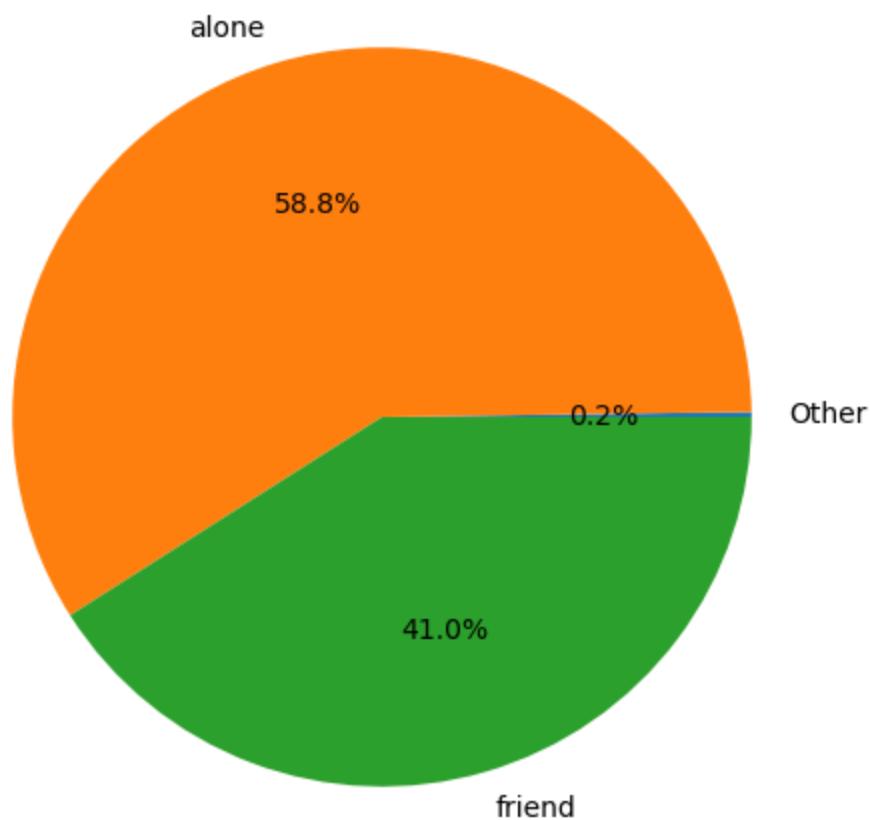
```
In [20]: summary = {
    "Total Transactions": len(df),
    "Total Expense": df["Amount"].sum(),
    "Highest Expense Category": df.groupby("Category")["Amount"].sum().idxmax(),
    "Lowest Expense Category": df.groupby("Category")["Amount"].sum().idxmin()
}

summary
```

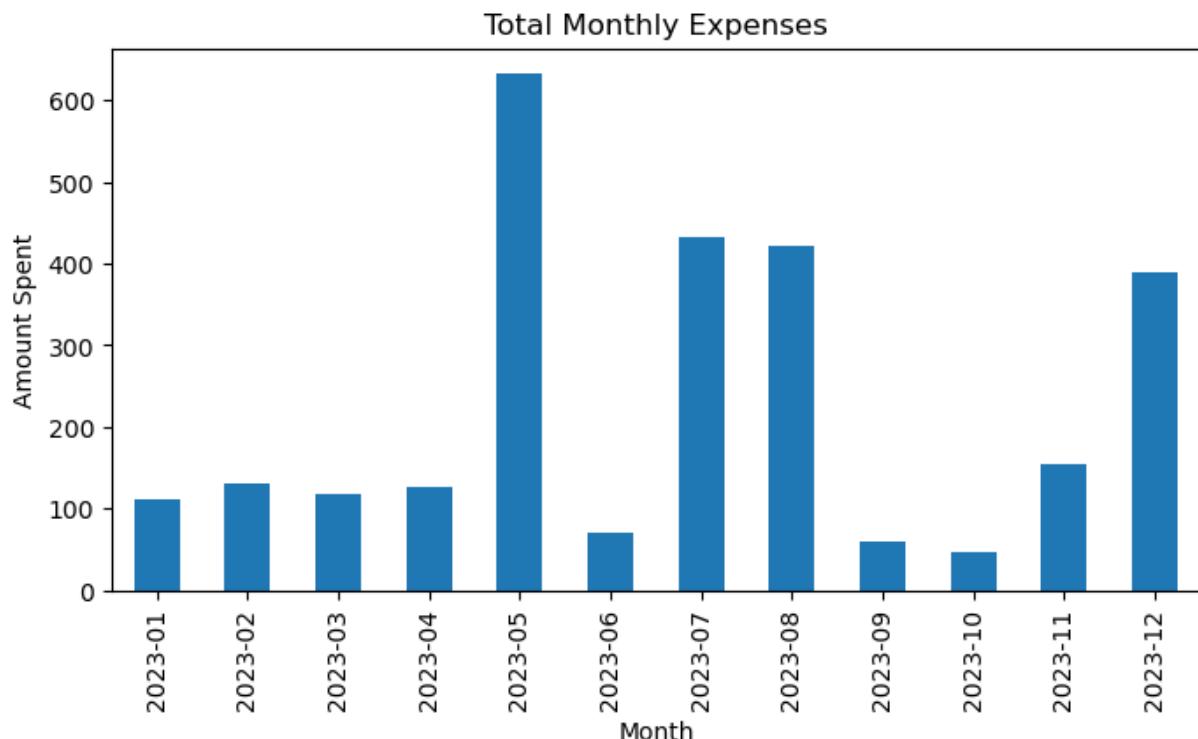
```
Out[20]: {'Total Transactions': 145,  
          'Total Expense': np.int64(4609),  
          'Highest Expense Category': 'alone',  
          'Lowest Expense Category': 'Other'}
```

```
In [21]: import matplotlib.pyplot as plt  
  
plt.figure(figsize=(6,6))  
df.groupby('Category')['Amount'].sum().plot(kind='pie', autopct='%1.1f%%')  
plt.title("Expense Distribution by Category")  
plt.ylabel("") # extra Label remove  
plt.show()
```

Expense Distribution by Category



```
In [22]: monthly_expense = df.groupby('Month')['Amount'].sum()  
  
plt.figure(figsize=(8,4))  
monthly_expense.plot(kind='bar')  
plt.title("Total Monthly Expenses")  
plt.xlabel("Month")  
plt.ylabel("Amount Spent")  
plt.show()
```



In [ ]: