PYTHON FOR DATA ANALYSIS FUNDAMENTALS

1. Importing a CSV file to Google Colab via code

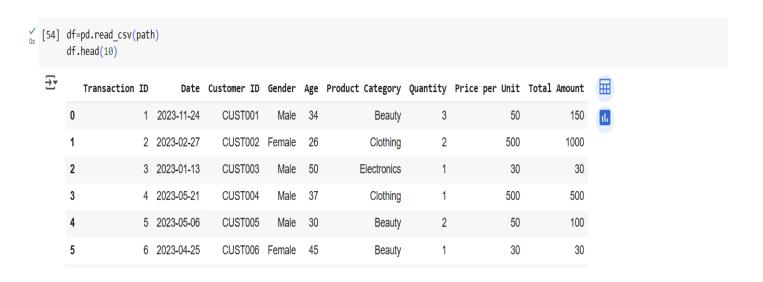


2. Importing Pandas and NumPy for data analysis and creating a path for the dataset.

```
import pandas as pd import numpy as np

y [4] path="/content/retail_sales_dataset.csv"
```

3. Checking if data is loaded successfully while limiting it to 10 rows.



4. Investigating the properties of the dataset given.

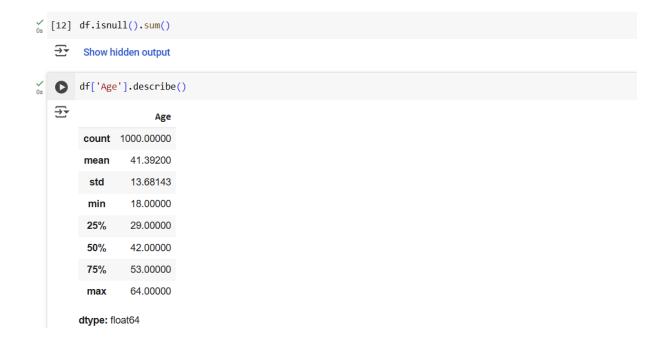
```
[8] df.shape
  → (1000, 9)

↓ ◆ ⇔ ■ ‡ 见 □ :
  df.info()
  RangeIndex: 1000 entries, 0 to 999
      Data columns (total 9 columns):
          Column
                          Non-Null Count Dtype
          Transaction ID 1000 non-null
                                        int64
                          1000 non-null
          Date
                                        object
          Customer ID
                         1000 non-null
                                        object
          Gender
                          1000 non-null
          Age
                          1000 non-null
                                        int64
          Product Category 1000 non-null
                                        object
          Quantity
Price per Unit
                          1000 non-null
                                        int64
                         1000 non-null
          Total Amount
                          1000 non-null
                                        int64
      dtypes: int64(5), object(4)
      memory usage: 70.4+ KB
```

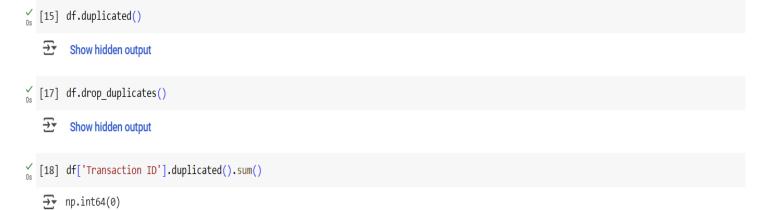
5. Investigating the properties of the dataset given.



6. Investigating the properties of the dataset given, focusing on nulls, and describe statement.



7. Checking for any duplicates in the dataset.



8. Syntax for age groups

```
[24] def categorize_age(age):
       if age < 12:
         return 'Child'
       elif age < 18:
         return 'Teen'
       elif age < 65:
         return 'Adult'
         return 'Senior'
     df['AgeGroup'] = df['Age'].apply(categorize_age)
     display(df)
<del>_</del>
           Transaction ID
                                 Date Customer ID Gender
                                                            Age Product Category Quantity Price per Unit Total Amount AgeGroup
       0
                        1 2023-11-24
                                          CUST001
                                                                                                                        150
                                                                                                                                 Adult
                                                                            Beauty
       1
                        2 2023-02-27
                                          CUST002 Female
                                                                           Clothing
                                                                                                          500
                                                                                                                       1000
                                                                                                                                 Adult
       2
                           2023-01-13
                                          CUST003
                                                             50
                                                                         Electronics
                                                                                                           30
                                                                                                                         30
                                                                                                                                 Adult
                                                                           Clothing
       3
                        4 2023-05-21
                                          CUST004
                                                                                                          500
                                                      Male
                                                             37
                                                                                                                        500
                                                                                                                                 Adult
```

9. Extracting the day of the week to analyse the best performing day of the week.

```
import pandas as pd
√<sub>0s</sub> [46]
            df['Date'] = pd.to_datetime(df['Date'])
            df['day of the week'] = df['Date'].dt.day_name()
            display(df)
   <del>_</del>₹
                                                                                            Price
               Transaction
                                        Customer
                                                                     Product
                                                                                                      Total
                                                                                                                          Day of the
                                                                                                                                                       day of the
                                                                                                                                       day_of_month
                                Date
                                                   Gender Age
                                                                              Quantity
                                                                                                              AgeGroup
                                                                                              per
                         ID
                                              ID
                                                                    Category
                                                                                                     Amount
                                                                                                                                Week
                                                                                                                                                              week
                                                                                             Unit
                               2023-
           0
                                        CUST001
                                                     Male
                                                             34
                                                                      Beauty
                                                                                      3
                                                                                               50
                                                                                                        150
                                                                                                                  Adult
                                                                                                                               Friday
                                                                                                                                                  24
                                                                                                                                                             Friday
                               11-24
                               2023-
           1
                           2
                                        CUST002 Female
                                                             26
                                                                     Clothing
                                                                                      2
                                                                                              500
                                                                                                       1000
                                                                                                                  Adult
                                                                                                                              Monday
                                                                                                                                                  27
                                                                                                                                                           Monday
                               02-27
                               2023-
                                                                                                                                                            Friday
           2
                           3
                                        CUST003
                                                             50
                                                                   Electronics
                                                                                       1
                                                                                               30
                                                                                                         30
                                                                                                                  Adult
                                                                                                                               Friday
                                                                                                                                                  13
                                                     Male
                               01-13
                               2023-
           3
                                        CUST004
                                                     Male
                                                             37
                                                                     Clothing
                                                                                       1
                                                                                              500
                                                                                                        500
                                                                                                                  Adult
                                                                                                                              Sunday
                                                                                                                                                  21
                                                                                                                                                           Sunday
                               05-21
                               2023-
                                        CUST005
                                                     Male
                                                             30
                                                                      Beauty
                                                                                               50
                                                                                                         100
                                                                                                                  Adult
                                                                                                                             Saturday
                                                                                                                                                          Saturday
                               05-06
```

10. Deleting a duplicated column from the data frame.

```
[47] import pandas as pd
          df.drop(columns=['Day of the Week'], inplace=True)
          print(df)
```

Show hidden output

11. Extracting the day of the month from the date.

```
import pandas as pd

df['Date'] = pd.to_datetime(df['Date'])

df['day_of_month'] = df['Date'].dt.day

display(df)
```

[†]		Transaction ID	Date	Customer ID	Gender	Age	Product Category	Quantity	Price per Unit	Total Amount	AgeGroup	day_of_month	day of the week
	0	1	2023-11- 24	CUST001	Male	34	Beauty	3	50	150	Adult	24	Friday
	1	2	2023- 02-27	CUST002	Female	26	Clothing	2	500	1000	Adult	27	Monday
	2	3	2023- 01-13	CUST003	Male	50	Electronics	1	30	30	Adult	13	Friday
	3	4	2023- 05-21	CUST004	Male	37	Clothing	1	500	500	Adult	21	Sunday

12. Extracting the month of the year from the date to analyse the best-performing month in the dataset.

```
df['Date'] = pd.to_datetime(df['Date'])
df['month'] = df['Date'].dt.month
display(df)
```

5	Transaction II	Date	Customer	Gender	Age	Product Category	Quantity	Price per Unit	Total Amount	AgeGroup	day_of_month	day of the week	month
0	,	1 2023 11-24		Male	34	Beauty	3	50	150	Adult	24	Friday	11
1	2	2023 02-27		Female	26	Clothing	2	500	1000	Adult	27	Monday	2
2	;	3 2023 01-13		Male	50	Electronics	1	30	30	Adult	13	Friday	1
3	4	2023 05-21		Male	37	Clothing	1	500	500	Adult	21	Sunday	5
4	ţ	2023 05-06		Male	30	Beauty	2	50	100	Adult	6	Saturday	5
99	5 996	2023		Male	62	Clothing	1	50	50	Adult	16	Tuesday	5

13. Extracting the month and year from the dataset.

```
os D
         import pandas as pd
        df['Date'] = pd.to_datetime(df['Date'])
df['year_month'] = df['Date'].dt.strftime('%Y-%m')
        display(df)
   Price
                                                                                                                                    day of
the week
                                                                  Product
               Transaction
                                      Customer Gender Age
                                                                                                Total
                             Date
                                                                           Quantity
                                                                                        per
Unit
                                                                                                        AgeGroup day_of_month
                                                                                                                                               month year_month
                         ID
                                                                 Category
                                                                                               Amount
                             2023-
11-24
          0
                                      CUST001
                                                   Male
                                                                   Beauty
                                                                                          50
                                                                                                   150
                                                                                                            Adult
                                                                                                                              24
                                                                                                                                        Friday
                                                                                                                                                   11
                                                                                                                                                           2023-11
                             2023-
02-27
          1
                                      CUST002 Female
                                                           26
                                                                  Clothing
                                                                                   2
                                                                                          500
                                                                                                  1000
                                                                                                            Adult
                                                                                                                              27
                                                                                                                                      Monday
                                                                                                                                                    2
                                                                                                                                                           2023-02
                             2023-
01-13
                          3
                                      CUST003
                                                                                                                                        Friday
                                                                                                                                                           2023-01
                                                   Male
                                                           50 Electronics
                                                                                          30
                                                                                                    30
                                                                                                            Adult
                                                                                                                              13
                          4 2023-
05-21
                                                                                                                                                           2023-05
                                      CUST004
                                                   Male
                                                                  Clothing
                                                                                          500
                                                                                                  500
                                                                                                            Adult
                                                                                                                              21
                                                                                                                                       Sunday
                                                                                                                                                    5
                                      CUST005
                                                                                   2
                                                                                                   100
                                                                                                                                      Saturday
                                                                                                                                                           2023-05
                                      CUST996
                                                           62
                                                                                                    50
                                                                                                            Adult
                                                                                                                                                           2023-05
                                                   Male
                                                                  Clothing
                                                                                                                                      Tuesday
```

14. Syntax for spending buckets where spending is grouped.

```
def categorize_total_amount(total_amount):
    if total_amount <=99:
        return '0-99'
    elif total_amount <=199:
        return '100-199'
    elif total_amount <=299:
        return '200-299'
    elif total_amount <=499:
        return '300-499'
    else:
        return '500-2000'

df['Spending_Bucket'] = df['Total Amount'].apply(categorize_total_amount)
    display(df)</pre>
```

15. The entire table after analysis has been performed.

₹	Transa	action ID	Date	Customer ID	Gender	Age	Product Category	Quantity	Price per Unit	Total Amount	AgeGroup	day_of_month	day of the week	month	year_month	Spending_Bucket
	0	1	2023-11- 24	CUST001	Male	34	Beauty	3	50	150	Adult	24	Friday	11	2023-11	100-199
	1	2	2023-02- 27	CUST002	Female	26	Clothing	2	500	1000	Adult	27	Monday	2	2023-02	500-2000
	2	3	2023-01- 13	CUST003	Male	50	Electronics	1	30	30	Adult	13	Friday	1	2023-01	0-99
	3	4	2023-05- 21	CUST004	Male	37	Clothing	1	500	500	Adult	21	Sunday	5	2023-05	500-2000
	4	5	2023-05- 06	CUST005	Male	30	Beauty	2	50	100	Adult	6	Saturday	5	2023-05	100-199
9	995	996	2023-05- 16	CUST996	Male	62	Clothing	1	50	50	Adult	16	Tuesday	5	2023-05	0-99
9	996	997	2023-11- 17	CUST997	Male	52	Beauty	3	30	90	Adult	17	Friday	11	2023-11	0-99
5	997	998	2023-10- 29	CUST998	Female	23	Beauty	4	25	100	Adult	29	Sunday	10	2023-10	100-199
9	998	999	2023-12- 05	CUST999	Female	36	Electronics	3	50	150	Adult	5	Tuesday	12	2023-12	100-199
9	999	1000	2023-04- 12	CUST1000	Male	47	Electronics	4	30	120	Adult	12	Wednesday	4	2023-04	100-199