

Name : Manmath Maroti Kornule

Roll_no: CS7-59

PRN : 202401110045

Subject : Essentials of Data Science

Assignment:

Make Up Session Assignment for Theory

1. Take one real dataset. Find any 10 grains for the given dataset. Find solutions using pandas.

Data set link(Kaggle): <https://www.kaggle.com/datasets/atharvasoundankar/fashion-retail-sales>

Data set : fashion retail sales

	A	B	C	D	E	F	G	H	I
1	Customer Reference ID	Item Purchased	Purchase Amount (USD)	Date Purchase	Review Rating	Payment Method			
2	4018	Handbag	4619	05-02-2023		Credit Card			
3	4115	Tunic	2456	11-07-2023	2	Credit Card			
4	4019	Tank Top	2102	23-03-2023	4.1	Cash			
5	4097	Leggings	3126	15-03-2023	3.2	Cash			
6	3997	Wallet	3003	27-11-2022	4.7	Cash			
7	4080	Onesie	2914	11-12-2022	4.5	Credit Card			
8	4055	Jacket	2571	08-07-2023	1.3	Cash			
9	3973	Trousers	2419	10-11-2022	4.6	Cash			
10	4044	Jeans	4771	19-05-2023	4.1	Cash			
11	4010	Loafers	4233	11-06-2023		Credit Card			
12	4108	Slippers	2356	19-03-2023	4.8	Credit Card			
13	4067	Bowtie	4418	21-11-2022	3.4	Cash			
14	4068	Pajamas	3728	09-12-2022		Credit Card			
15	4102	Trench Coat	2130	29-01-2023	4.8	Cash			
16	4044	Handbag	2122	01-08-2023	1.2	Credit Card			
17	4096	Poncho	2383	10-04-2023		Credit Card			
18	4017	Gloves	2895	17-07-2023	3.6	Credit Card			
19	4001	Trench Coat	2952	13-06-2023	2.2	Cash			
20	3976	Slippers	4069	18-06-2023	4.9	Credit Card			
21	4103	Romper	4465	21-11-2022		Credit Card			
22	4081	T-shirt	2108	25-03-2023		Cash			
23	3986	Gloves	4298	05-04-2023	4.2	Cash			
24	4108	Jeans	2814	11-02-2023	4.4	Credit Card			
25	4109	Flip-Flops	4932	20-12-2022	4	Credit Card			
26	4075	Tunic	4661	12-04-2023	1.2	Credit Card			
27	4040	Shorts	4872	25-08-2023	3.8	Cash			
28	4099	Blazer	4232	25-10-2022	2.7	Credit Card			

Reading the dataset and displaying first 5 rows for verification.

The screenshot shows a Jupyter Notebook interface. The code cell on the left reads the 'Fashion_Retail_Sales.csv' file and displays the first 5 rows. The console on the right shows the output of the code, which is a table of 5 rows and 6 columns.

```
1 # -*- coding: utf-8 -*-
2 """
3 Created on Wed May 7 19:14:26 2025
4
5 @author: manmath
6 """
7
8 import pandas as pd
9
10 df = pd.read_csv(r"C:\Users\manmath\Downloads\Fashion_Retail_Sales.csv")
11 print(df.head())
12
```

	Customer Reference ID	Item Purchased	Review Rating	Payment Method
0	4018	Handbag	NaN	Credit Card
1	4115	Tunic	2.0	Credit Card
2	4010	Tank Top	4.1	Cash
3	4007	Leggings	3.2	Cash
4	3997	Wallet	4.7	Cash

GRAINS:

1. Average Purchase Amount Per Year?

The screenshot shows a Jupyter Notebook interface. The code cell on the left calculates the yearly average purchase amount. The console on the right shows the output of the code, which is a table with 2 rows and 2 columns.

```
1 # -*- coding: utf-8 -*-
2 """
3 Created on Wed May 7 19:14:26 2025
4
5 @author: manmath
6 """
7
8 import pandas as pd
9
10 df = pd.read_csv(r"C:\Users\manmath\Downloads\Fashion_Retail_Sales.csv")
11 df['Date Purchase'] = pd.to_datetime(df['Date Purchase'], dayfirst=True)
12 df['Year'] = df['Date Purchase'].dt.year
13 yearly_avg = df.groupby('Year')['Purchase Amount (USD)'].mean()
14 print("Yearly Average Purchase Amount:\n", yearly_avg)
15
```

Year	Yearly Average Purchase Amount:
2022	162.994220
2023	154.596696

2. Display the details of Most Expensive Item Purchased?

The screenshot shows a Jupyter Notebook interface. The code cell on the left finds the most expensive item. The console on the right shows the output of the code, which is a table with 1 row and 6 columns.

```
1 # -*- coding: utf-8 -*-
2 """
3 Created on Wed May 7 19:14:26 2025
4
5 @author: manmath
6 """
7
8 import pandas as pd
9
10 df = pd.read_csv(r"C:\Users\manmath\Downloads\Fashion_Retail_Sales.csv")
11 max_item = df[df['Purchase Amount (USD)'] == df['Purchase Amount (USD)'].max()]
12 print("Most Expensive Purchase:\n", max_item)
13
```

	Customer Reference ID	Item Purchased	Review Rating	Payment Method
23	4109	Flip-Flops	4.0	Credit Card

3. How many purchases were made with each payment method?

The screenshot shows a Jupyter Notebook interface. The code cell on the left calculates the payment method counts. The console on the right shows the output of the code, which is a table with 2 rows and 2 columns.

```
1 # -*- coding: utf-8 -*-
2 """
3 Created on Wed May 7 19:14:26 2025
4
5 @author: manmath
6 """
7
8 import pandas as pd
9
10 df = pd.read_csv(r"C:\Users\manmath\Downloads\Fashion_Retail_Sales.csv")
11 payment_counts = df['Payment Method'].value_counts()
12 print("Payment Method Counts:\n", payment_counts)
13
```

Payment Method	Payment Method Counts:
Credit Card	1770
Cash	1630

4. What is the average purchase amount for each item?

```
C:\Users\manmath\untitled0.py
untitled0.py X
1  # -*- coding: utf-8 -*-
2  """
3  Created on Wed May  7 19:14:26 2025
4
5  @author: manmath
6  """
7
8  import pandas as pd
9
10 df = pd.read_csv(r"C:\Users\manmath\Downloads\Fashion_Retail_Sales.csv")
11
12 avg_purchase_per_item = df.groupby('Item Purchased')['Purchase Amount (USD)'].mean()
13 print("Average Purchase per Item:\n", avg_purchase_per_item)
14
15
16
```

```
Console 1/A X
In [6]: runfile('C:/Users/manmath/untitled0.py', wdir='C:/Users/manmath')
Average Purchase per Item:
Item Purchased
Backpack      112.145455
Belt           151.942029
Blazer        184.907407
Blouse        183.719298
Boots         228.673077
Bowtie        197.840909
Camisole      98.967742
Cardigan      151.967213
Coat          106.807692
Dress         93.340909
Flannel Shirt 102.377358
Flip-Flops    275.829268
Gloves        237.115385
Handbag       214.711864
Hat           103.267857
Hoodie        154.500000
Jacket        157.980000
Jeans         272.250000
Jumpsuit      114.217391
Kimono        149.672131
Leggings      165.040816
Loafers       183.416667
Onesie        161.847458
Overalls      103.222222
Pajamas       209.803279
Pants         130.915493
Polo Shirt    105.953488
Poncho        190.366667
Raincoat      123.218182
Romper        197.104167
Sandals       118.531915
```

5. Count how many purchases happened in each year.

```
C:\Users\manmath\untitled0.py
untitled0.py X
1  # -*- coding: utf-8 -*-
2  """
3  Created on Wed May  7 19:14:26 2025
4
5  @author: manmath
6  """
7
8  import pandas as pd
9
10 df = pd.read_csv(r"C:\Users\manmath\Downloads\Fashion_Retail_Sales.csv")
11
12 df['Date Purchase'] = pd.to_datetime(df['Date Purchase'], dayfirst=True)
13 df['Year'] = df['Date Purchase'].dt.year
14 yearly_purchases = df['Year'].value_counts().sort_index()
15 print("Purchases Per Year:\n", yearly_purchases)
16
17
```

```
Console 1/A X
In [7]: runfile('C:/Users/manmath/untitled0.py', wdir='C:/Users/manmath')
Purchases Per Year:
Year
2022      853
2023     2547
Name: count, dtype: int64

In [8]:
```

6. Show all purchases with rating less than 3.

```
File Edit Search Source Run Debug Consoles Projects Tools View Help
C:\Users\manmath\untitled0.py
untitled0.py X
1  # -*- coding: utf-8 -*-
2  """
3  Created on Wed May  7 19:14:26 2025
4
5  @author: manmath
6  """
7
8  import pandas as pd
9
10 df = pd.read_csv(r"C:\Users\manmath\Downloads\Fashion_Retail_Sales.csv")
11
12 low_rated = df[df['Review Rating'] < 3]
13 print("Low-Rated Purchases:\n", low_rated)
14
15
```

```
Console 1/A X
In [8]: runfile('C:/Users/manmath/untitled0.py', wdir='C:/Users/manmath')
Low-Rated Purchases:
Customer Reference ID Item Purchased ... Review Rating Payment Method
1      4115      Tunic ...      2.0      Credit Card
6      4055      Jacket ...      1.3      Cash
14     4044      Handbag ...      1.2      Credit Card
17     4001      Trench Coat ...      2.2      Cash
24     4075      Tunic ...      1.2      Credit Card
...      ...      ...      ...      ...
3395     4118      Shorts ...      2.7      Cash
3396     4056      Slippers ...      2.5      Cash
3397     3991      Onesie ...      2.4      Credit Card
3398     4042      Sandals ...      1.8      Cash
3399     4001      Backpack ...      1.4      Cash

[1488 rows x 6 columns]

In [9]:
```

7. What is the average spending per customers who spends above overall average?

```
C:\Users\manmath\untitled0.py
untitled0.py
1  # -*- coding: utf-8 -*-
2  """
3  Created on Wed May 7 19:14:26 2025
4
5  @author: manmath
6  """
7
8  import pandas as pd
9
10 df = pd.read_csv(r"C:\Users\manmath\Downloads\Fashion_Retail_Sales.csv")
11
12
13 overall_avg = df['Purchase Amount (USD)'].mean()
14 avg_spending_per_customer = df.groupby('Customer Reference ID')['Purchase Amount (USD)'].mean()
15 above_avg_customers = avg_spending_per_customer[avg_spending_per_customer > overall_avg]
16 above_avg_customers = above_avg_customers.round(2)
17 print("Customers Who Spend Above the Overall Average:\n", above_avg_customers)
18
19
```

```
Console 1/A X
In [13]: runfile('C:/Users/manmath/untitled0.py', wdir='C:/Users/manmath')
Customers Who Spend Above the Overall Average:
Customer Reference ID
3972    274.33
3973    203.68
3974    209.16
3976    342.28
3980    218.05
3981    392.00
3984    451.93
3986    304.85
3997    277.79
4001    235.33
4002    287.27
4010    342.79
4012    312.00
4017    227.82
4018    387.00
4019    257.00
4035    429.25
4040    536.50
4044    485.83
4054    332.10
4055    275.20
4060    296.06
4067    326.40
4068    253.35
4070    236.38
4075    296.64
4080    312.21
4081    188.21
4083    251.47
4096    329.30
4097    490.75
4099    375.12
4102    215.00
4103    318.75
4108    361.26
4109    745.00
4110    212.71
4115    218.50
4122    201.86
Name: Purchase Amount (USD), dtype: float64
```

8. Display all purchases made on a specific date.

```
File Edit Search Source Run Debug Consoles Projects Tools View Help
C:\Users\manmath\untitled0.py
untitled0.py
1  # -*- coding: utf-8 -*-
2  """
3  Created on Wed May 7 19:14:26 2025
4
5  @author: manmath
6  """
7
8  import pandas as pd
9
10 df = pd.read_csv(r"C:\Users\manmath\Downloads\Fashion_Retail_Sales.csv")
11
12
13 specific_date = '2022-12-11' # Replace with your chosen date
14 df['Date Purchase'] = pd.to_datetime(df['Date Purchase'], dayfirst=True)
15 purchases_on_date = df[df['Date Purchase'] == specific_date]
16 print("Purchases on Date:\n", purchases_on_date)
17
18
19
```

```
Console 1/A X
In [14]: runfile('C:/Users/manmath/untitled0.py', wdir='C:/Users/manmath')
Purchases on Date:
Customer Reference ID Item Purchased ... Review Rating Payment Method
5      4080      Onesie ...      4.5      Credit Card
127    4072      Onesie ...      1.2      Cash
555    4046      Poncho ...      2.8      Cash
828    4090      Umbrella ...      3.8      Cash
910    4079      Sun Hat ...      1.6      Credit Card
1101   4098      Poncho ...      3.0      Cash
1927   4066      Sunglasses ...      NaN      Credit Card
1958   3976      Leggings ...      2.2      Credit Card
2051   4008      Hoodie ...      1.5      Cash
2296   4117      Pajamas ...      2.5      Credit Card
2401   4012      Tank Top ...      1.4      Cash
2501   3979      Flannel Shirt ...      1.2      Credit Card

[12 rows x 6 columns]

In [15]:
```

9. Show all purchases above \$500.

```
C:\Users\manmath\untitled0.py
untitled0.py
1  # -*- coding: utf-8 -*-
2  """
3  Created on Wed May 7 19:14:26 2025
4
5  @author: manmath
6  """
7
8  import pandas as pd
9
10 df = pd.read_csv(r"C:\Users\manmath\Downloads\Fashion_Retail_Sales.csv")
11
12
13 high_purchases = df[df['Purchase Amount (USD)'] > 500]
14 print("Purchases Above $500:\n", high_purchases)
15
16
17
```

```
Console 1/A X
In [16]: runfile('C:/Users/manmath/untitled0.py', wdir='C:/Users/manmath')
Purchases Above $500:
Customer Reference ID Item Purchased ... Review Rating Payment Method
0      4018      Handbag ...      NaN      Credit Card
1      4115      Tunic ...      2.0      Credit Card
2      4019      Tank Top ...      4.1      Cash
3      4097      Leggings ...      3.2      Cash
4      3997      Wallet ...      4.7      Cash
5      4080      Onesie ...      4.5      Credit Card
6      4055      Jacket ...      1.3      Cash
7      3973      Trousers ...      4.6      Cash
8      4044      Jeans ...      4.1      Cash
9      4010      Loafers ...      NaN      Credit Card
10     4108      Slippers ...      4.8      Credit Card
11     4067      Bowtie ...      3.4      Cash
12     4068      Pajamas ...      NaN      Credit Card
13     4102      Trench Coat ...      4.8      Cash
14     4044      Handbag ...      1.2      Credit Card
15     4096      Poncho ...      NaN      Credit Card
16     4017      Gloves ...      3.6      Credit Card
17     4001      Trench Coat ...      2.2      Cash
18     3976      Slippers ...      4.9      Credit Card
19     4103      Romper ...      NaN      Credit Card
20     4081      T-shirt ...      NaN      Cash
```


10. How many high-value purchases (above average amount) are made per item?

```
File Edit Search Source Run Debug Consoles Projects Tools View Help
C:\Users\manmath\untitled0.py
untitled0.py* X
1 #-*- coding: utf-8 -*-
2 """
3 Created on Wed May 7 19:14:26 2025
4
5 @author: manmath
6 """
7
8 import pandas as pd
9
10 df = pd.read_csv(r"C:\Users\manmath\Downloads\Fashion_Retail_Sales.csv")
11
12 avg_amount = df['Purchase Amount (USD)'].mean()
13 high_value = df[df['Purchase Amount (USD)'] > avg_amount]
14 high_value_count = high_value['Item Purchased'].value_counts()
15 print("High-Value Purchases per Item:\n", high_value_count)
16
17
18
19
```

```
In [17]: runfile('C:/Users/manmath/untitled0.py', wdir='C:/Users/manmath')
High-Value Purchases per Item:
Item Purchased
Blouse      20
Scarf       20
Shorts      19
Handbag     18
Poncho      17
Skirt       17
T-shirt     17
Overalls    16
Onesie      16
Jeans       16
Loafers     16
Swimsuit    16
Raincoat    16
Boots       15
Slippers    15
Pajamas     15
Hoodie      15
Cardigan    14
Tunic       14
Sweater     14
Kimono      14
Leggings    14
Flannel Shirt 14
Trench Coat 14
Sandals     14
Jacket      14
Coat        14
Umbrella    13
Vest        13
Polo Shirt  13
Camisole    13
Sneakers    13
Wallet      13
Hat         13
Flip-Flops  13
Tank Top    13
Jumpsuit    12
Romper      12
Backpack    12
Pants       12
Belt        11
Gloves      11
Sunglasses  11
```