Supermarket Sales

DEPI Data Analysis First Project Using Python

By: Ahmed Essam Abdallah

❖ Introduction:

This project aims to analyze and visualize supermarket sales data to gain insights into customer behavior, sales trends, and product performance. The project involves data cleaning, transformation, and visualization using Python (Colab Notebook) libraries such as Pandas, NumPy and Matplotlib.

	Invoice ID	Branch	Yangon	Naypyitaw	Mandalay	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	Rating
0	750-67-8428	A	1	0	0	Normal	Male	Health and beauty	74.69	7	26.1415	NaN	1/5/2019	13:08	Ewallet	9.1
1	226-31-3081	C	0	1	0	Normal	Male	Electronic accessories	15.28	5	3.8200	80.2200	3/8/2019	10:29	Cash	9.6
2	631-41-3108	A	1	0	0	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	3/3/2019	13:23	Credit card	7.4
3	123-19-1176	A	1	0	0	Normal	Male	Health and beauty	58.22	8	NaN	489.0480	1/27/2019	8 - 30 PM	Ewallet	8.4
4	373-73-7910	A	1	0	0	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	2/8/2019	10:37	Ewallet	5.3
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1001	861-77-0145	C	0	1	0	Member	Male	Electronic accessories	81.97	10	40.9850	860.6850	3/3/2019	14:30	Cash	9.2
1002	479-26-8945	B	0	0	1	Member	Female	Sports and travel	16.49	2	1.6490	34.6290	2/5/2019	11:32	Ewallet	4.6
1003	210-67-5886	C	0	1	0	Member	Female	Health and beauty	98.21	3	14.7315	309.3615	2/5/2019	10:41	Credit card	7.8
1004	227-78-1148	B	0	0	1	Normal	Female	Fashion accessories	72.84	7	25.4940	535.3740	2/15/2019	12:44	Cash	8.4
1005	645-44-1170	A	1	0	0	Member	Male	Home and lifestyle	58.07	9	26.1315	548.7615	1/19/2019	20:07	Ewallet	4.3

1006 rows × 16 columns

By: Ahmed Essam Abdallah

❖ Project Index:

1. Gather:
    a. File in hand 'Project Data - Supermarket Sales.csv'.

2. Assess:
    a. Quality Issues.
    b. Tidiness Issues.

3. Clean:
    a. Duplicated Date.
    b. Data Redundancy.
    c. Inaccurate Data.
    d. Invalid Data.
    e. Missing Data.
    f. Inconsistent Data.
    g. Tidiness Issues.

4. Storing.

5. Data Visualization:
    a. Sales Insights.
    b. Customer Behavior Insights.

By: Ahmed Essam Abdallah

❖ Gather:

- File in hand 'Project Data - Supermarket Sales.csv'. You can find it in the repository files.

❖ Assess:

1. Quality Issues:

a. Duplicate Data:
- I found 6 identical records appear more than once.

b. Data Redundancy:
- Branch column was repeated column, that adds no value for the data (The Branch location column was enough).

c. Missing Data:
- Nulls represented as dashes (-) in Customer type column.
(I tried to figure out the missing data, but I couldn't so I left it as I found it)
- Tax 5% column was missing value.
- Total column was missing value.

d. Inaccurate Data:
- Input mistake in Customer type column (Memberr instead of Member).
- Input mistake in Rating column (97 instead of 9.7).

e. Invalid Data:
- Unit price column sometimes has USD and was an object not a float.
- Quantity column has negative values.

By: Ahmed Essam Abdallah

f. Inconsistent Data:
- Date column has different formats.
- Time column has different formats.

2. Tidiness Issues:
a. Columns headers as values not variables (Columns: Yangon, Naypyitaw, Mandalay).

❖ Clean:

1. Fixing Quality Issues:
a. Duplicated Date:
- Remove the duplicate records.

b. Data Redundancy:
- Drop Branch column.

c. Inaccurate Data:
- Correct input mistake in Customer type column (Change Meberr into Member).
- Correct input mistake in Rating column (Change 97 into 9.7).

d. Invalid Data:
- Remove USD from Unit price column and change its type to float.
- Change Quantity column negative values.

e. Missing Data:
- Recalculate the Tax 5% column.

$$(Tax\ 5\% = Unit\ price * Quantity * 0.05)$$

- Recalculate the Total column.

$$(Total = Unit\ price * Quantity + Tax\ 5\%)$$

By: Ahmed Essam Abdallah

f. Inconsistent Data:
   - Change Date column format.
   - Change Time column format.

2. Fixing Tidiness Issues:
   a. Conver the Yangon, Naypyitaw and Mandalay columns to Branch location.

   b. Drop the Yangon, Naypyitaw and Mandalay columns.

❖ Storing:

- After gathering, Assessing and Cleaning data. I stored the data.
You can find it in the repository files 'Cleaned_Supermarket_Sales'.

❖ Data Visualization:

1. Sales Insights:
   - Create visualizations to analyze sales trends, total sales and product line distribution.

2. Customer Behavior Insights:
   - Create visualizations to examine customer purchasing patterns and their preferences.

Thank You

By: Ahmed Essam Abdallah