Deloitte Virtual Internship – Task 1: Transaction Data Analysis

This task involved examining a large transaction dataset to identify structure, issues, and key trends. Python was used to load and analyze the data using the pandas library.

This is my code:

```
# Show total number of rows
print(f" 🔸 Total number of transactions: {df.shape[0]}")

# Show column names
print("\n 🔹 Columns in the file:")
print(df.columns.tolist())

#checking for missing or null values
print(df.isnull().sum())

#total sales per store or top selling products
print(df.groupby("STORE_NBR")["TOT_SALES"].sum().sort_values(ascending=False))

#fixing the data format
df['DATE'] = pd.to_datetime(df['DATE'], unit='D', origin='1899-12-30')
print(df['DATE'].head())
print(df.describe())
```

📊 Terminal Output (Results of the Code):
Total number of transactions: 264836

Columns in the file:
['DATE', 'STORE_NBR', 'LYLTY_CARD_NBR', 'TXN_ID', 'PROD_NBR', 'PROD_NAME', 'PROD_QTY', 'TOT_SALES']

```
DATE            0
STORE_NBR       0
...
```

This analysis verified data integrity, confirmed no missing values, and helped identify high-performing stores by sales volume. The cleaned dataset is ready for further exploration and segmentation.