Improve ur Business with my Analysis

Capstone Project
By : Mo Amer.

step1: Python Operations:
1. Extraction — Python → CSV:
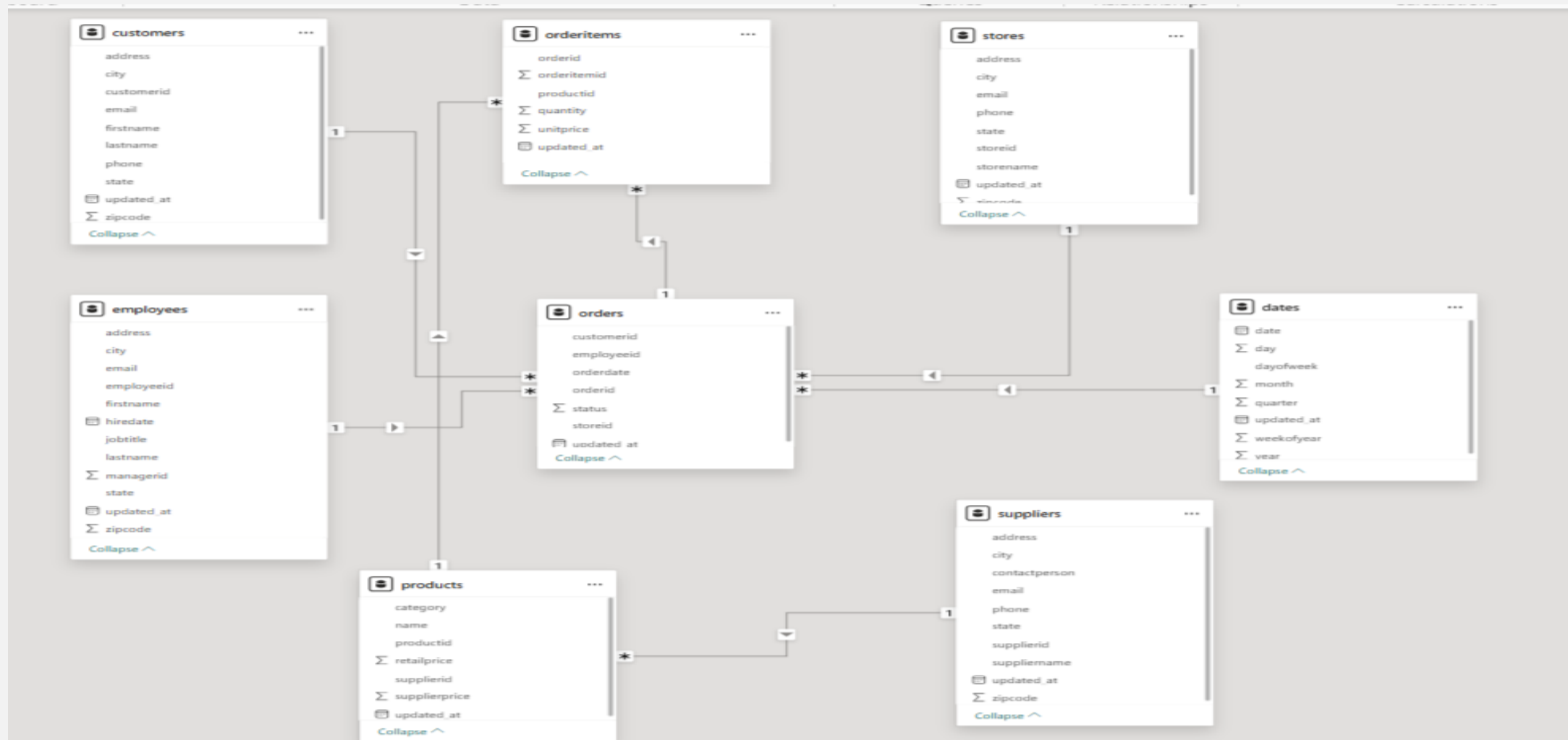
```python
import os
import pandas as pd
import psycopg2

DATABASE_URL = 'postgresql://neondb_owner:npg_nf0Bpoz8SKtF@ep-solitary-recipe-ah9qonat-pooler.c-3.us-east-1.aws.neon.tech/neondb?sslmode=r

conn = psycopg2.connect(DATABASE_URL)
cur = conn.cursor()

os.makedirs("Data",exist_ok=True)

cur.execute("""
    SELECT table_name
    FROM information_schema.tables
    WHERE table_type='BASE TABLE' and table_schema ='oms_core';
""")

tables = [t[0] for t in cur.fetchall()]
print("Tables in data:",tables)

dfs = {}
for table in tables:
    query = f"SELECT * FROM oms_core.{table};"
    df = pd.read_sql(query,conn)
    dfs[table] = df
    csv_path = f"Data/{table}.csv"
    df.to_csv(csv_path,index = False)
    print(f"saved{csv_path},columns num:{len(df)}")

print(dfs)
```

2.Data Quality Validation:

```python
30    print(dfs)
31
32    if "customers" in dfs:
33        df = dfs["customers"]
34        null_ids = df[df["customerid"].isnull()]
35        print("❌ Customerid Null rows: "if len(null_ids)>0 else "✅ Customerid check passed")
36        if len(null_ids)>0:
37            print(null_ids)
38
39    if "orders" in dfs:
40        df = dfs["orders"]
41        null_dates = df[df["orderdate"].isnull()]
42        print("❌ orderdate NULL rows:" if len(null_dates) > 0 else "✅ orderdate check passed")
43        if len(null_dates) > 0:
44            print(null_dates)
45
46    if "orderitems" in dfs:
47        df = dfs["orderitems"]
48        bad_qty = df[df["quantity"] <= 0]
49        print("❌ quantity <= 0 rows:" if len(bad_qty) > 0 else "✅ quantity check passed")
50        if len(bad_qty) > 0:
51            print(bad_qty)
52
53    if "products" in dfs:
54        df = dfs["products"]
55        bad_price = df[df["retailprice"] < df["supplierprice"]]
56        print("❌ retailprice < supplierprice rows:" if len(bad_price) > 0 else "✅ price check passed")
57        if len(bad_price) > 0:
58            print(bad_price)
59
60    if "customers" in dfs:
61        df = dfs["customers"]
62        invalid_email = df[~df["email"].str.contains(r"@.*\.", na=True)]
63        print("❌ invalid email rows:" if len(invalid_email) > 0 else "✅ email check passed")
64        if len(invalid_email) > 0:
65            print(invalid_email)
66
67    if "employees" in dfs:
68        df = dfs["employees"]
69        df['hiredate'] = pd.to_datetime(df['hiredate'])
70        bad_hire = df[df["hiredate"] > pd.Timestamp.today()]
71        print("❌ hiredate in future rows:" if len(bad_hire) > 0 else "✅ hiredate check passed")
72        if len(bad_hire) > 0:
73            print(bad_hire)
```

Step 2: Power bi Operations:
1. Data Model:

2. Create Business KPIs (DAX):

1. Total Orders
2. Average Order Value
3. Total Revenue
4. Cost
5. Profit
6. Gross Margin %
7. Repeat Customers

3.visualizations

➢ Showing the important (KPIs)

➢ Matrices

       1. To clarifies top employees

       2. To clarifies top products

       3. To clarifies top customers

➢ Line Chart

       To clarifies Revenue by month

➢ Bar chart

       To clarifies Revenue by category

➢ Donut chart

       Revenue by store name

➢ Slicers

       To explain impact of any Employee, Customer, Product, and Date on your Business

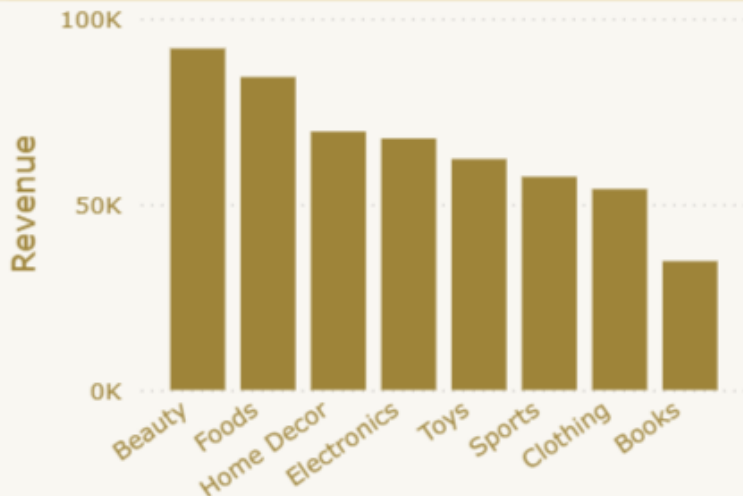**Total Orders**
300
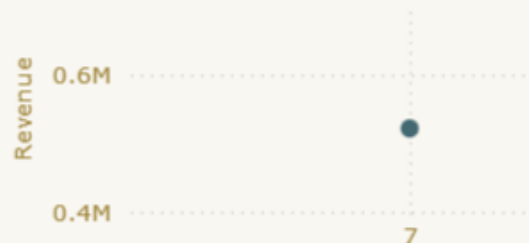
Revenue
522K

Average Order Value
$1.74K

Gross Margin
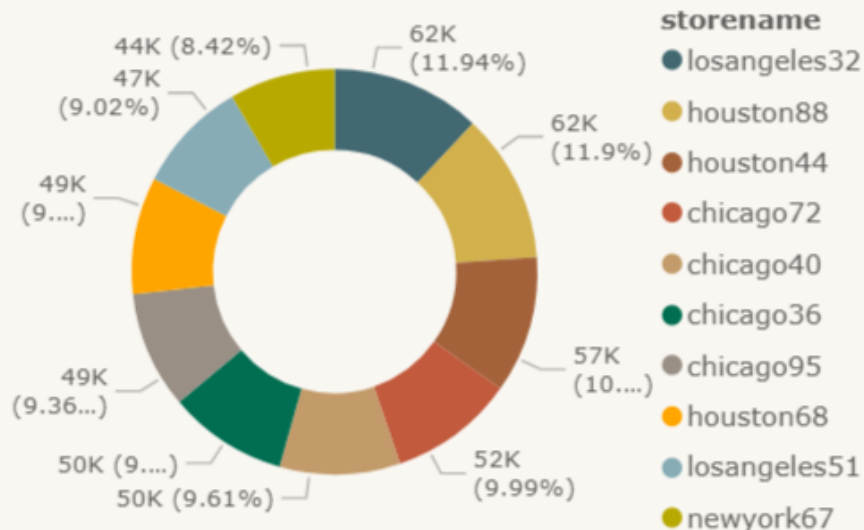5.02%

Repeat Customers
75

Revenue by category

(bar chart: categories – Beauty, Foods, Home Decor, Electronics, Toys, Sports, Clothing, Books; Revenue axis 0K–100K)

Revenue by month

(scatter/line chart: Revenue axis 0.4M–0.6M)

Top Employees

employeeid	Orders	Revenue	Profit
505616	6	15393	775
508575	6	15557	760
507196	5	14586	756
501617	7	11338	653
503955	5	12418	637
503576	5	9773	612
Total	**300**	**521833**	**26179**

Top Customers

customerid	Orders	Revenue	Profit
26677	6	15542	817
26218	6	14141	760
20677	8	11807	711
23728	7	14670	710
21348	8	13732	650
22531	9	12964	605
20012	6	11337	560
Total	**300**	**521833**	**26179**

Revenue by storename

(donut chart)
- 62K (11.94%)
- 62K (11.9%)
- 57K (10....)
- 52K (9.99%)
- 50K (9.61%)
- 50K (9....)
- 49K (9.36...)
- 49K (9....)
- 47K (9.02%)
- 44K (8.42%)

storename
- losangeles32
- houston88
- houston44
- chicago72
- chicago40
- chicago36
- chicago95
- houston68
- losangeles51
- newyork67

Top Products by Profit

product name	Profit
Doll	1505
Perfume	1487
Laptop	1341
Basketball	1328
Lipstick	1224
Sushi	1219
Face Cream	1191
Eyeshadow Palette	888
Vase	877
Burger	871
Salad	821
Tablet	816
Tennis Racket	786
Dress	781
Total	**26179**

Slicers: employeeid (All), customerid (All), product name (All), date (All)

Click on this folder to open the documents of project

Compressed
(zipped) Folder

Thank u eng. Abdelhamid for ur effort with me