DATA 311 - Fall 2020

Assignment 5 - Due Weds, Nov 11 by midnight JEONG MO YANG

For this assignment,

- Get all of our sales rep rules/data loaded into the database
- Also load the October 2020 sales data
- Produce a few summaries!

In [6]:
```
import sqlite3
import Store
import pandas as pd
```

In [7]:
```
conn = sqlite3.connect('Store.db')
curs = conn.cursor()
curs.execute("PRAGMA foreign_keys=ON;")
```

Out[7]: <sqlite3.Cursor at 0x7ffb562550a0>

In [8]:
```
Store.Rebuild()
```

Out[8]: 1

In [9]:
```
records = []
for x in range(1, 13):
    if x < 10:
        records.append("20190" + str(x))
    else:
        records.append("2019" + str(x))
for x in range(1, 11):
    if x <10:
        records.append("20200" + str(x))
    else:
        records.append("2020"+ str(x))
```

In [10]:
```
for y in records:
    filename = './data/Sales_' + y + '.csv'
    data = pd.read_csv(filename, dtype={'zip':str})
    Store.LoadData(data)
```

In [11]:
```
Store.LoadMoreData()
```

Out[11]: 1

In [12]:
```
#Helpful reminder - you can list out all of the tables and column names with this code
x = pd.read_sql("""SELECT name
                FROM sqlite_master
                WHERE type = 'table'
                AND name LIKE 't%';""",conn)
for table in x.values:
    sql = "PRAGMA table_info(" + table[0] + ");"
    print(table)
    print(pd.read_sql(sql,conn))
    print('Wn')
```

```
['tProd']
   cid        name     type  notnull dflt_value  pk
```

```
0    0     prod_id  INTEGER        0       None   1
1    1   prod_name     TEXT        1       None   0
2    2  unit_price  NUMERIC        1       None   0


['tState']
   cid   name  type  notnull dflt_value  pk
0    0     st  TEXT        0       None   1
1    1  state  TEXT        1       None   0
2    2    div  TEXT        1       None   0


['tZip']
   cid  name  type  notnull dflt_value  pk
0    0   zip  TEXT        0       None   1
1    1  city  TEXT        1       None   0
2    2    st  TEXT        1       None   0


['tCust']
   cid        name     type  notnull dflt_value  pk
0    0     cust_id  INTEGER        0       None   1
1    1  first_name     TEXT        1       None   0
2    2   last_name     TEXT        1       None   0
3    3     address     TEXT        1       None   0
4    4         zip     TEXT        0       None   0


['tOrder']
   cid      name     type  notnull dflt_value  pk
0    0  order_id  INTEGER        0       None   1
1    1   cust_id  INTEGER        0       None   0
2    2       day  INTEGER        1       None   0
3    3     month  INTEGER        1       None   0
4    4      year  INTEGER        1       None   0


['tOrderDetail']
   cid      name     type  notnull dflt_value  pk
0    0  order_id  INTEGER        0       None   1
1    1   prod_id  INTEGER        0       None   2
2    2       qty  INTEGER        1       None   0


['tDivToReg']
   cid name  type  notnull dflt_value  pk
0    0  div  TEXT        0       None   1
1    1  reg  TEXT        1       None   0


['tRep']
   cid      name     type  notnull dflt_value  pk
0    0    rep_id  INTEGER        0       None   1
1    1  rep_name     TEXT        1       None   0


['tRepByDiv']
   cid    name     type  notnull dflt_value  pk
0    0     div     TEXT        0       None   1
1    1  rep_id  INTEGER        0       None   0


['tRepByReg']
   cid    name     type  notnull dflt_value  pk
0    0     reg     TEXT        0       None   1
1    1  rep_id  INTEGER        0       None   0


['tRepByState']
```

```
   cid    name     type  notnull dflt_value  pk
0    0      st     TEXT        0       None   1
1    1  rep_id  INTEGER        0       None   0


['tRepByZip']
   cid    name     type  notnull dflt_value  pk
0    0     zip     TEXT        0       None   1
1    1  rep_id  INTEGER        0       None   0


['tRepByCust']
   cid     name     type  notnull dflt_value  pk
0    0  cust_id  INTEGER        0       None   1
1    1   rep_id  INTEGER        0       None   0
```

---

1) What are our total sales for all data in the database, grouped by sales rep?

Return two columns:

- Sales rep name
- Total sales

In [13]:
```
pd.read_sql("""SELECT rep_name AS Sales_rep_Name, sum(unit_price*qty) AS Total_Sales
                FROM tCust
                JOIN tOrder USING (cust_id)
                JOIN tOrderDetail USING (order_id)
                JOIN tProd USING (prod_id)
                JOIN tRepByCust USING (cust_id)
                JOIN tRep USING (rep_id)
                GROUP BY rep_name;""",conn)
```

Out[13]:

	Sales_rep_Name	Total_Sales
0	Alice	1502423.25
1	Bob	1936065.27
2	Cathy	5194235.92
3	Diane	2895373.04
4	Edgar	214196.54
5	Frank	15740.38

In [ ]:
```

```

---

2) How many customers are assigned to each sales rep?

Return two columns:

- Sales rep name
- Number of customers assigned

In [31]:
```
pd.read_sql("""SELECT rep_name AS Sales_rep_Name, count(DISTINCT cust_id) as NumCustAs
                FROM tCust
                JOIN tRepByCust USING (cust_id)
```

```
                    JOIN tRep USING (rep_id)
                    GROUP BY rep_name""",conn)
```

Out[31]:

	Sales_rep_Name	NumCustAssg
0	Alice	41
1	Bob	55
2	Cathy	135
3	Diane	72
4	Edgar	6
5	Frank	1

---

3) What are our total sales for all data in the database, grouped by region?

Return two columns:

- Region
- Total sales

In [15]:
```
pd.read_sql("""SELECT reg AS Region,sum(unit_price*qty) AS Total_Sales
                    FROM tCust
                    JOIN tOrder USING (cust_id)
                    JOIN tOrderDetail USING (order_id)
                    JOIN tProd USING (prod_id)
                    JOIN tZip USING (zip)
                    JOIN tState USING (st)
                    JOIN tDivToreg USING (div)
                    GROUP BY reg;""",conn)
```

Out[15]:

	Region	Total_Sales
0	Midwest	2553259.73
1	Northeast	1972839.06
2	PR	191336.42
3	South	4129485.77
4	West	2911113.42

In [ ]:
```
conn.close()
```

In [ ]:

In [ ]:

In [ ]: