The underlying dataset you receive is home sales in California. I want you to calculate some summary statistics and answer a few questions listed below.

Data dictionary:
- FA_TRANSACTION_ID: this is the unique ID for a given home sale
- PROPERTY_ID: this is the unique ID for a given property
- BUYER_BORROWER1_NAME: this is the name of the buyer of the home
- BUYER_BORROWWER1_CORP_IND: this is whether or not the buyer was a corporation
- SELLER1_NAME: this is the seller's name
- SALE_AMT: the amount the home was purchased for
- RECORDING_DATE: the date the property sold
- FULL_STREET_ADDRESS, CITY, ZIP_CODE: all parts of the address
- MSA: geographic region the property belongs to
- SUM_BUILDING_SQ_FT: the size of the home in square feet
- FIRST_MTG_LENDER_NAME: the name of the lender
- YEAR BUILT: the year the home was built

Tasks:
1. Upload the data to your database of choice (must be SQL based)
2. Answer the following questions in SQL & share your SQL:
   a. What percentage of properties were purchased by corporations? How did that trend over time? What about by MSA?
   b. Find the 2nd largest home buyer in every MSA and the number of homes they purchased
   c. Plot the average sale amount by MSA (you can use your choice of software to make the plot)
   d. Who were the top 5 largest lenders for corporate purchases?
   e. Is there anything else interesting in the dataset that you noticed?

RESULT :

A. What percentage of properties were purchased by corporations? How did that trend over time? What about by MSA?

Answer :

Part 1 : What percentage of properties were purchased by corporations?

File : *Question A Part 1.sql*

```
/* Question A Part 1 : What percentage of properties were purchased by corporations?*/
/*
Total Amount of Data : 336344
*/
SELECT
(COUNT(COALESCE(sfrpurchases_data.BUYER_BORROWER1_CORP_IND))/336344)*10
0
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.BUYER_BORROWER1_CORP_IND IN ('Y');
# Result : 20.13%
```

Part 3 : What about by MSA?

File : *Question A Part 3.sql*

```
/* Question A Part 3 : What about by MSA?*/
/*
Total Amount of Data : 336344
*/
SELECT (COUNT(COALESCE(sfrpurchases_data.MSA))/336344)*100
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA NOT IN ('');
# Result : 98.43%
```

B. Find the 2nd largest home buyer in every MSA and the number of homes they purchased

Answer :

File : *Question B.sql*

```
/* Question B : Find the 2nd largest home buyer in every MSA and the number of homes
they purchased*/
/*
MSA : Bakersfield, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Bakersfield, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 5887000
-- -- -- -- -- -- --
/*
MSA : Chico, CA
```

```
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Chico, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 4043536
-- -- -- -- -- -- --
/*
MSA : Clearlake, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Clearlake, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 6481818
-- -- -- -- -- -- --
/*
MSA : Crescent City, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Crescent City, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 1200000
-- -- -- -- -- -- --
/*
MSA : El Centro, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('El Centro, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 4750000
-- -- -- -- -- -- --
/*
MSA : Eureka-Arcata, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Eureka-Arcata, CA')
```

```
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 1600000
-- -- -- -- -- -- --
/*
MSA : Fresno, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Fresno, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 36008500
-- -- -- -- -- -- --
/*
MSA : Hanford-Corcoran, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Hanford-Corcoran, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 1467500
-- -- -- -- -- -- --
/*
MSA : Kennewick-Richland, WA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Kennewick-Richland, WA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : -- (There is only one buyer from Kennewick-Richland, WA, so there is no the 2nd
largest home buyer from Kennewick-Richland, WA)
-- -- -- -- -- -- --
/*
MSA : Los Angeles-Long Beach-Anaheim, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Los Angeles-Long Beach-Anaheim, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 1130090909
```

```
-- -- -- -- -- -- --
/*
MSA : Madera, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Madera, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 14657000
-- -- -- -- -- -- --
/*
MSA : Modesto, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Modesto, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 24898272
-- -- -- -- -- -- --
/*
MSA : Merced, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Merced, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 31402000
-- -- -- -- -- -- --
/*
MSA : Napa, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Napa, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 11900000
-- -- -- -- -- -- --
/*
MSA : Oxnard-Thousand Oaks-Ventura, CA
*/
```

```
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Oxnard-Thousand Oaks-Ventura, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 24700000
-- -- -- -- -- -- --
/*
MSA : Red Bluff, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Red Bluff, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 525000
-- -- -- -- -- -- --
/*
MSA : Redding, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Redding, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 1936326
-- -- -- -- -- -- --
/*
MSA : Sacramento-Roseville-Folsom, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Sacramento-Roseville-Folsom, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 650000000
-- -- -- -- -- -- --
/*
MSA : Riverside-San Bernardino-Ontario, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Riverside-San Bernardino-Ontario, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
```

```sql
LIMIT 2;
# Result : 454545454
-- -- -- -- -- -- --
/*
MSA : Salinas, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Salinas, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 36284000
-- -- -- -- -- -- --
/*
MSA : San Diego-Chula Vista-Carlsbad, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('San Diego-Chula Vista-Carlsbad, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 160025454
-- -- -- -- -- -- --
/*
MSA : San Francisco-Oakland-Berkeley, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('San Francisco-Oakland-Berkeley, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 690000000
-- -- -- -- -- -- --
/*
MSA : San Luis Obispo-Paso Robles, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA LIKE ('San Luis Obispo-Paso Robles, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 73000000
-- -- -- -- -- -- --
/*
```

```
MSA : San Jose-Sunnyvale-Santa Clara, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('San Jose-Sunnyvale-Santa Clara, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 28500000
-- -- -- -- -- -- --
/*
MSA : Santa Rosa-Petaluma, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Santa Rosa-Petaluma, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 74500000
-- -- -- -- -- -- --
/*
MSA : Santa Maria-Santa Barbara, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Santa Maria-Santa Barbara, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 41697000
-- -- -- -- -- -- --
/*
MSA : Santa Cruz-Watsonville, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Santa Cruz-Watsonville, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 10000000
-- -- -- -- -- -- --
/*
MSA : Truckee-Grass Valley, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
```

```sql
WHERE sfrpurchases_data.MSA IN ('Truckee-Grass Valley, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 5500000
-- -- -- -- -- -- --
/*
MSA : Sonora, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Sonora, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 2975000
-- -- -- -- -- -- --
/*
MSA : Stockton, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Stockton, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 14237500
-- -- -- -- -- -- --
/*
MSA : Susanville, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Susanville, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 2900000
-- -- -- -- -- -- --
/*
MSA : Ukiah, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Ukiah, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 3450000
```

```
-- -- -- -- -- -- --
/*
MSA : Vallejo, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Vallejo, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 10842727
-- -- -- -- -- -- --
/*
MSA : Visalia, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Visalia, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 5900000
-- -- -- -- -- -- --
/*
MSA : Yuba City, CA
*/
SELECT COALESCE(sfrpurchases_data.MSA), sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE sfrpurchases_data.MSA IN ('Yuba City, CA')
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 2;
# Result : 3000000
```
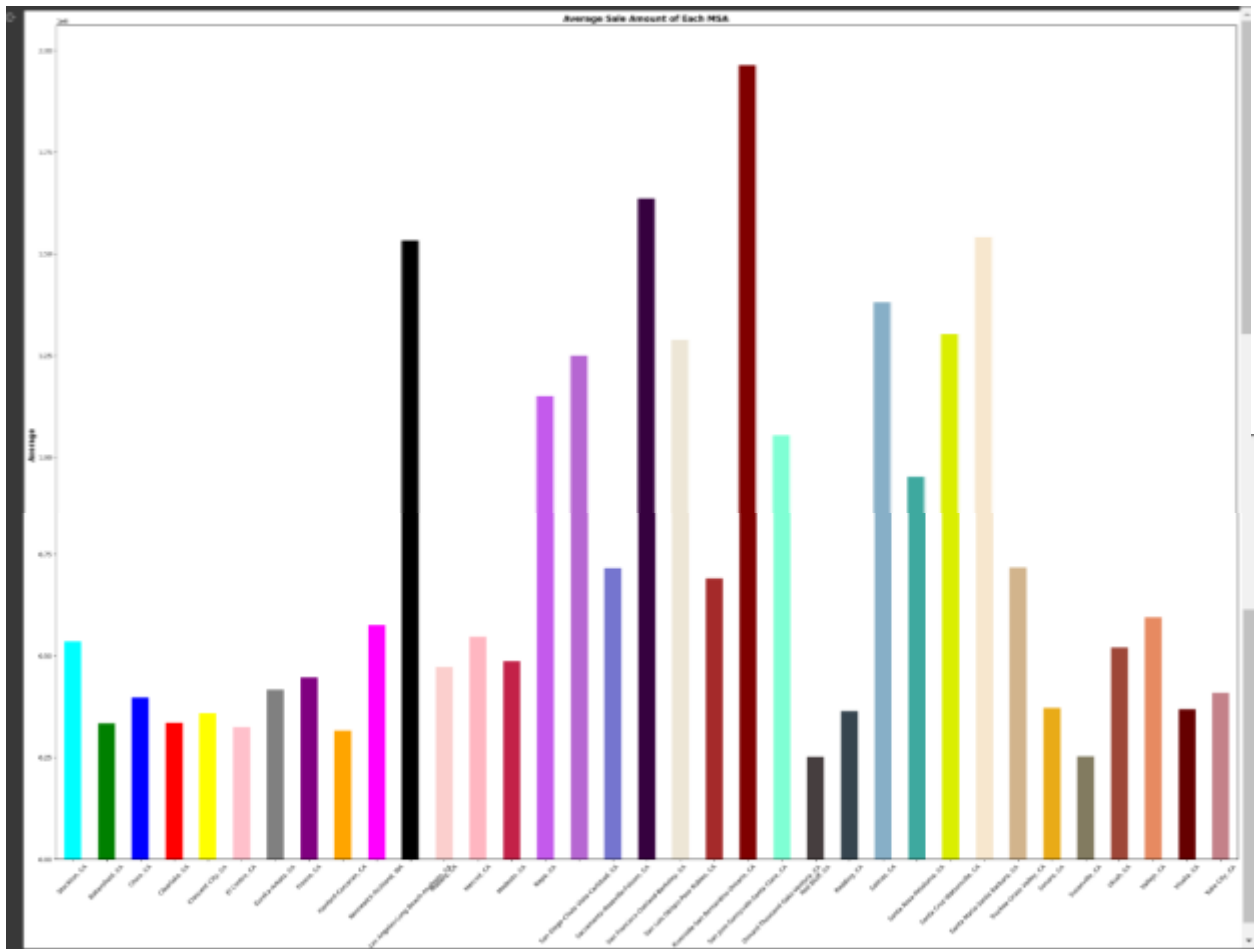
C. Plot the average sale amount by MSA (you can use your choice of software to make the plot)
   Answer :

Average Sale Amount of Each MSA

Source Code

Tools : *Google Colaboratory*

File : *Question C.ipynb*

Cell 1 (Import Libraries)

```
# Import Libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

Cell 2 (Import Dataset)

```
# Import Dataset SRF .csv (Top Football League Scores Dataset)
df = pd.read_csv('SFR Purchases.csv')
```

Cell 3 (Missing / Empty Value Check)

```
MSA_Empty = df['MSA'].isnull().sum()


MSA_Total = len(df) - MSA_Empty
```

```
print('MSA Empty Values : ', MSA_Empty)
print('Total Amount of MSA in the Dataset : ', MSA_Total)
```

Cell 4 (Dropping the Missing Values from MSA column)
```
df['MSA'].dropna()

df[df['MSA'] == ''].shape
```

Cell 5 (Average of each MSA)
```
avg_stockton = (np.sum(df['SALE_AMT'][df['MSA'] == 'Stockton, CA']))/(l
en(df[df['MSA'] == 'Stockton, CA']))
print('Average Stockton : \n', avg_stockton)

avg_bakersfield = (np.sum(df['SALE_AMT'][df['MSA'] == 'Bakersfield, CA'
]))/(len(df[df['MSA'] == 'Bakersfield, CA']))
print('Average Bakersfield : \n', avg_bakersfield)

avg_chico = (np.sum(df['SALE_AMT'][df['MSA'] == 'Chico, CA']))/(len(df[
df['MSA'] == 'Chico, CA']))
print('Average Chico : \n', avg_chico)

avg_clearlake = (np.sum(df['SALE_AMT'][df['MSA'] == 'Clearlake, CA']))/
(len(df[df['MSA'] == 'Clearlake, CA']))
print('Average Clearlake : \n', avg_clearlake)

avg_crescent = (np.sum(df['SALE_AMT'][df['MSA'] == 'Crescent City, CA']
))/(len(df[df['MSA'] == 'Crescent City, CA']))
print('Average Crescent City : \n', avg_crescent)

avg_centro = (np.sum(df['SALE_AMT'][df['MSA'] == 'El Centro, CA']))/(le
n(df[df['MSA'] == 'El Centro, CA']))
print('Average El Centro : \n', avg_centro)

avg_eureka = (np.sum(df['SALE_AMT'][df['MSA'] == 'Eureka-
Arcata, CA']))/(len(df[df['MSA'] == 'Eureka-Arcata, CA']))
print('Average Eureka-Arcata : \n', avg_eureka)

avg_fresno = (np.sum(df['SALE_AMT'][df['MSA'] == 'Fresno, CA']))/(len(d
f[df['MSA'] == 'Fresno, CA']))
print('Average Fresno : \n', avg_fresno)

avg_hanford = (np.sum(df['SALE_AMT'][df['MSA'] == 'Hanford-
Corcoran, CA']))/(len(df[df['MSA'] == 'Hanford-Corcoran, CA']))
print('Average Hanford-Corcoran : \n', avg_hanford)
```

```python
avg_kennewick = (np.sum(df['SALE_AMT'][df['MSA'] == 'Kennewick-
Richland, WA']))/(len(df[df['MSA'] == 'Kennewick-Richland, WA']))
print('Average Kennewick-Richland : \n', avg_kennewick)

avg_losangelesanaheim = (np.sum(df['SALE_AMT'][df['MSA'] == 'Los Angele
s-Long Beach-Anaheim, CA']))/(len(df[df['MSA'] == 'Los Angeles-
Long Beach-Anaheim, CA']))
print('Average Los Angeles-Long Beach-
Anaheim : \n', avg_losangelesanaheim)

avg_madera = (np.sum(df['SALE_AMT'][df['MSA'] == 'Madera, CA']))/(len(d
f[df['MSA'] == 'Madera, CA']))
print('Average Madera : \n', avg_madera)

avg_merced = (np.sum(df['SALE_AMT'][df['MSA'] == 'Merced, CA']))/(len(d
f[df['MSA'] == 'Merced, CA']))
print('Average Merced : \n', avg_merced)

avg_modesto = (np.sum(df['SALE_AMT'][df['MSA'] == 'Modesto, CA']))/(len
(df[df['MSA'] == 'Modesto, CA']))
print('Average Modesto : \n', avg_modesto)

avg_napa = (np.sum(df['SALE_AMT'][df['MSA'] == 'Napa, CA']))/(len(df[df
['MSA'] == 'Napa, CA']))
print('Average Napa : \n', avg_napa)

avg_carlsbad = (np.sum(df['SALE_AMT'][df['MSA'] == 'San Diego-
Chula Vista-Carlsbad, CA']))/(len(df[df['MSA'] == 'San Diego-
Chula Vista-Carlsbad, CA']))
print('Average San Diego-Chula Vista-Carlsbad : \n', avg_carlsbad)

avg_sacramento = (np.sum(df['SALE_AMT'][df['MSA'] == 'Sacramento-
Roseville-Folsom, CA']))/(len(df[df['MSA'] == 'Sacramento-Roseville-
Folsom, CA']))
print('Average Sacramento-Roseville-Folsom : \n', avg_sacramento)

avg_berkeley = (np.sum(df['SALE_AMT'][df['MSA'] == 'San Francisco-
Oakland-Berkeley, CA']))/(len(df[df['MSA'] == 'San Francisco-Oakland-
Berkeley, CA']))
print('Average San Francisco-Oakland-Berkeley : \n', avg_berkeley)

avg_sanluis = (np.sum(df['SALE_AMT'][df['MSA'] == 'San Luis Obispo-
Paso Robles, CA']))/(len(df[df['MSA'] == 'San Luis Obispo-
Paso Robles, CA']))
```

```python
print('Average San Luis Obispo-Paso Robles : \n', avg_sanluis)

avg_riverside = (np.sum(df['SALE_AMT'][df['MSA'] == 'Riverside-
San Bernardino-Ontario, CA']))/(len(df[df['MSA'] == 'Riverside-
San Bernardino-Ontario, CA']))
print('Average Riverside-San Bernardino-Ontario : \n', avg_riverside)

avg_sanjose = (np.sum(df['SALE_AMT'][df['MSA'] == 'San Jose-Sunnyvale-
Santa Clara, CA']))/(len(df[df['MSA'] == 'San Jose-Sunnyvale-
Santa Clara, CA']))
print('Average San Jose-Sunnyvale-Santa Clara : \n', avg_sanjose)

avg_oxnard = (np.sum(df['SALE_AMT'][df['MSA'] == 'Oxnard-Thousand Oaks-
Ventura, CA']))/(len(df[df['MSA'] == 'Oxnard-Thousand Oaks-
Ventura, CA']))
print('Average Oxnard-Thousand Oaks-Ventura : \n', avg_oxnard)

avg_redbluff = (np.sum(df['SALE_AMT'][df['MSA'] == 'Red Bluff, CA']))/(
len(df[df['MSA'] == 'Red Bluff, CA']))
print('Average Red Bluff : \n', avg_redbluff)

avg_redding = (np.sum(df['SALE_AMT'][df['MSA'] == 'Redding, CA']))/(len
(df[df['MSA'] == 'Redding, CA']))
print('Average Redding : \n', avg_redding)

avg_salinas = (np.sum(df['SALE_AMT'][df['MSA'] == 'Salinas, CA']))/(len
(df[df['MSA'] == 'Salinas, CA']))
print('Average Salinas : \n', avg_salinas)

avg_santarosa = (np.sum(df['SALE_AMT'][df['MSA'] == 'Santa Rosa-
Petaluma, CA']))/(len(df[df['MSA'] == 'Santa Rosa-Petaluma, CA']))
print('Average Santa Rosa-Petaluma : \n', avg_santarosa)

avg_santacruz = (np.sum(df['SALE_AMT'][df['MSA'] == 'Santa Cruz-
Watsonville, CA']))/(len(df[df['MSA'] == 'Santa Cruz-
Watsonville, CA']))
print('Average Santa Cruz-Watsonville : \n', avg_santacruz)

avg_santamaria = (np.sum(df['SALE_AMT'][df['MSA'] == 'Santa Maria-
Santa Barbara, CA']))/(len(df[df['MSA'] == 'Santa Maria-
Santa Barbara, CA']))
print('Average Santa Maria-Santa Barbara : \n', avg_santamaria)

avg_truckee = (np.sum(df['SALE_AMT'][df['MSA'] == 'Truckee-
Grass Valley, CA']))/(len(df[df['MSA'] == 'Truckee-Grass Valley, CA']))
```

```
print('Average Truckee-Grass Valley : \n', avg_truckee)

avg_sonora = (np.sum(df['SALE_AMT'][df['MSA'] == 'Sonora, CA']))/(len(d
f[df['MSA'] == 'Sonora, CA']))
print('Average Sonora : \n', avg_sonora)

avg_susanville = (np.sum(df['SALE_AMT'][df['MSA'] == 'Susanville, CA'])
)/(len(df[df['MSA'] == 'Susanville, CA']))
print('Average Susanville : \n', avg_susanville)

avg_ukiah = (np.sum(df['SALE_AMT'][df['MSA'] == 'Ukiah, CA']))/(len(df[
df['MSA'] == 'Ukiah, CA']))
print('Average Ukiah : \n', avg_ukiah)

avg_vallejo = (np.sum(df['SALE_AMT'][df['MSA'] == 'Vallejo, CA']))/(len
(df[df['MSA'] == 'Vallejo, CA']))
print('Average Vallejo : \n', avg_vallejo)

avg_visalia = (np.sum(df['SALE_AMT'][df['MSA'] == 'Visalia, CA']))/(len
(df[df['MSA'] == 'Visalia, CA']))
print('Average Visalia : \n', avg_visalia)

avg_yuba = (np.sum(df['SALE_AMT'][df['MSA'] == 'Yuba City, CA']))/(len(
df[df['MSA'] == 'Yuba City, CA']))
print('Average Yuba City : \n', avg_yuba)
```

Cell 6 (Creating Final Dataset to do Data Visualization / Creating the Plot)

```
data = np.array([avg_stockton.astype(float), avg_bakersfield.astype(flo
at), avg_chico.astype(float), avg_clearlake.astype(float), avg_crescent
.astype(float), avg_centro.astype(float), avg_eureka.astype(float), avg
_fresno.astype(float), avg_hanford.astype(float), avg_kennewick.astype(
float), avg_losangelesanaheim.astype(float), avg_madera.astype(float),
avg_merced.astype(float), avg_modesto.astype(float), avg_napa.astype(fl
oat), avg_carlsbad.astype(float), avg_sacramento.astype(float), avg_ber
keley.astype(float), avg_sanluis.astype(float), avg_riverside.astype(fl
oat), avg_sanjose.astype(float), avg_oxnard.astype(float), avg_redbluff
.astype(float), avg_redding.astype(float), avg_salinas.astype(float), a
vg_santarosa.astype(float), avg_santacruz.astype(float), avg_santamaria
.astype(float), avg_truckee.astype(float), avg_sonora.astype(float), av
g_susanville.astype(float), avg_ukiah.astype(float), avg_vallejo.astype
(float), avg_visalia.astype(float), avg_yuba.astype(float)])

# Index column
index = ['Stockton, CA', 'Bakersfield, CA', 'Chico, CA', 'Clearlake, CA
', 'Crescent City, CA', 'El Centro, CA', 'Eureka-
```

```
Arkata, CA', 'Fresno, CA', 'Hanford-Corcoran, CA', 'Kennewick-
Richland, WA', 'Los Angeles-Long Beach-
Anaheim, CA', 'Madera, CA', 'Merced, CA', 'Modesto, CA', 'Napa, CA', 'S
an Diego-Chula Vista-Carlsbad, CA', 'Sacramento-Roseville-
Folsom, CA', 'San Francisco-Oakland-Berkeley, CA', 'San Luis Obispo-
Paso Robles, CA', 'Riverside-San Bernardino-Ontario, CA', 'San Jose-
Sunnyvale-Santa Clara, CA', 'Oxnard-Thousand Oaks-
Ventura, CA', 'Red Bluff, CA', 'Redding, CA', 'Salinas, CA', 'Santa Ros
a-Petaluma, CA', 'Santa Cruz-Watsonville, CA', 'Santa Maria-
Santa Barbara, CA', 'Truckee-
Grass Valley, CA', 'Sonora, CA', 'Susanville, CA', 'Ukiah, CA', 'Vallej
o, CA', 'Visalia, CA', 'Yuba City, CA']

# Column headers
columns = ['Avg']

# Create DataFrame
df_final = pd.DataFrame(data, index=index, columns=columns)
df_final.head()
```

Cell 7 (Data Visualization)
```
plotbar = df_final['Avg']
colors = ['cyan', 'green', 'blue', 'red', 'yellow', 'pink', 'grey', 'pu
rple', 'orange', 'magenta', 'black', '#FBCFCD', '#FFB6C1', '#C32148', '
#C45AEC', '#B666D2', '#7575CF', '#36013F', '#EDE6D6', 'brown', '#800000
', '#7FFFD4', '#463E3F', '#36454F', '#87AFC7', '#3EA99F', '#DAEE01', '#
F7E7CE', '#D2B48C', '#E9AB17', '#827B60', '#9E4638', '#E78A61', '#66000
0', '#C48189']
plotbar.plot(kind='bar', figsize=(35,25), color=colors)
plt.xticks(rotation=45)
plt.xlabel('MSA', fontsize = 12, fontweight = "semibold")
plt.ylabel('Average', fontsize = 12, fontweight = "semibold")
plt.title('Average Sale Amount of Each MSA', fontsize = 14, fontweight
= "semibold")
plt.show()
```

D. Who were the top 5 largest lenders for corporate purchases?
    Answer :
    1. MORGAN STANLEY PRIVATE BANK NA (SALE_AMT : **1619311818**)
    2. RIVERBEND FUNDING LLC (SALE_AMT : **640000000**)
    3. CITY NATIONAL BANK (SALE_AMT : **431803790**)
    4. EASY STREET CAPITAL CA LLC (SALE_AMT : **305202918**)
    5. UIF CORPORATION (SALE_AMT : **200530000**)
    File : *Question D.sql*

```sql
SELECT COALESCE(sfrpurchases_data.FIRST_MTG_LENDER_NAME),
sfrpurchases_data.SALE_AMT
FROM sfrpurchases.sfrpurchases_data
WHERE (sfrpurchases_data.BUYER_BORROWER1_CORP_IND IN ('Y')) AND
(sfrpurchases_data.FIRST_MTG_LENDER_NAME NOT IN (''))
ORDER BY sfrpurchases_data.SALE_AMT DESC
LIMIT 5;
/*
Result :
-> TOP 5 LARGEST LENDERS FOR CORPORATE PURCHASES
  1. MORGAN STANLEY PRIVATE BANK NA (SALE_AMT : 1619311818)
  2. RIVERBEND FUNDING LLC (SALE_AMT : 640000000)
  3. CITY NATIONAL BANK (SALE_AMT : 431803790)
  4. EASY STREET CAPITAL CA LLC (SALE_AMT : 305202918)
  5. UIF CORPORATION (SALE_AMT : 200530000)
*/
```