

Analysis of Sales Data

November 23, 2021

```
[1]: import pandas as pd
import os
import glob
```

```
[ ]: ##How to concatenate multiple csvs
```

```
[2]: path = r'/Users/anastasbaci/Desktop/pandas/Pandas-Data-Science-Tasks-master 2/
↳SalesAnalysis/Sales_Data' # use your path
all_files = glob.glob(path + "/*.csv")

li = []

for filename in all_files:
    df = pd.read_csv(filename, index_col=None, header=0)
    li.append(df)

all_months_data = pd.concat(li, axis=0, ignore_index=True)
```

```
[3]: all_months_data
```

```
[3]:      Order ID      Product Quantity Ordered Price Each \
0      295665      Macbook Pro Laptop           1      1700
1      295666      LG Washing Machine           1      600.0
2      295667      USB-C Charging Cable           1      11.95
3      295668      27in FHD Monitor            1     149.99
4      295669      USB-C Charging Cable           1      11.95
...      ...      ...      ...      ...
186845  222905      AAA Batteries (4-pack)           1        2.99
186846  222906      27in FHD Monitor            1     149.99
186847  222907      USB-C Charging Cable           1      11.95
186848  222908      USB-C Charging Cable           1      11.95
186849  222909      AAA Batteries (4-pack)           1        2.99
```

	Order Date	Purchase Address
0	12/30/19 00:01	136 Church St, New York City, NY 10001
1	12/29/19 07:03	562 2nd St, New York City, NY 10001
2	12/12/19 18:21	277 Main St, New York City, NY 10001

```

3      12/22/19 15:13      410 6th St, San Francisco, CA 94016
4      12/18/19 12:38      43 Hill St, Atlanta, GA 30301
...
186845 06/07/19 19:02      795 Pine St, Boston, MA 02215
186846 06/01/19 19:29      495 North St, New York City, NY 10001
186847 06/22/19 18:57      319 Ridge St, San Francisco, CA 94016
186848 06/26/19 18:35      916 Main St, San Francisco, CA 94016
186849 06/25/19 14:33      209 11th St, Atlanta, GA 30301

```

[186850 rows x 6 columns]

```
[4]: all_months_data.to_csv("ALL_MONTHS_DATA.csv", index=False)
```

```
[5]: all_data=pd.read_csv("ALL_MONTHS_DATA.csv")
```

```
[6]: all_data.head(10)
```

```
[6]:
```

	Order ID	Product	Quantity Ordered	Price Each	\
0	295665	Macbook Pro Laptop	1	1700	
1	295666	LG Washing Machine	1	600.0	
2	295667	USB-C Charging Cable	1	11.95	
3	295668	27in FHD Monitor	1	149.99	
4	295669	USB-C Charging Cable	1	11.95	
5	295670	AA Batteries (4-pack)	1	3.84	
6	295671	USB-C Charging Cable	1	11.95	
7	295672	USB-C Charging Cable	2	11.95	
8	295673	Bose SoundSport Headphones	1	99.99	
9	295674	AAA Batteries (4-pack)	4	2.99	

	Order Date	Purchase Address
0	12/30/19 00:01	136 Church St, New York City, NY 10001
1	12/29/19 07:03	562 2nd St, New York City, NY 10001
2	12/12/19 18:21	277 Main St, New York City, NY 10001
3	12/22/19 15:13	410 6th St, San Francisco, CA 94016
4	12/18/19 12:38	43 Hill St, Atlanta, GA 30301
5	12/31/19 22:58	200 Jefferson St, New York City, NY 10001
6	12/16/19 15:10	928 12th St, Portland, OR 97035
7	12/13/19 09:29	813 Hickory St, Dallas, TX 75001
8	12/15/19 23:26	718 Wilson St, Dallas, TX 75001
9	12/28/19 11:51	77 7th St, Dallas, TX 75001

```
[7]: all_data.shape
```

```
[7]: (186850, 6)
```

```
[8]: df_nan=all_data[all_data.isna().any(axis=1)]
```

```
[9]: df_nan
```

```
[9]:      Order ID Product Quantity Ordered Price Each Order Date \
264      NaN      NaN      NaN      NaN      NaN      NaN
648      NaN      NaN      NaN      NaN      NaN      NaN
680      NaN      NaN      NaN      NaN      NaN      NaN
1385     NaN      NaN      NaN      NaN      NaN      NaN
1495     NaN      NaN      NaN      NaN      NaN      NaN
...      ...      ...      ...      ...      ...      ...
185795    NaN      NaN      NaN      NaN      NaN      NaN
185868    NaN      NaN      NaN      NaN      NaN      NaN
185887    NaN      NaN      NaN      NaN      NaN      NaN
185960    NaN      NaN      NaN      NaN      NaN      NaN
186580    NaN      NaN      NaN      NaN      NaN      NaN
```

```
      Purchase Address
264      NaN
648      NaN
680      NaN
1385     NaN
1495     NaN
...      ...
185795    NaN
185868    NaN
185887    NaN
185960    NaN
186580    NaN
```

```
[545 rows x 6 columns]
```

```
[10]: all_data=all_data.dropna(how ='any')
```

```
[11]: all_data.head(25)
```

```
[11]:      Order ID      Product Quantity Ordered Price Each \
0      295665      Macbook Pro Laptop      1      1700
1      295666      LG Washing Machine      1      600.0
2      295667      USB-C Charging Cable      1      11.95
3      295668      27in FHD Monitor      1      149.99
4      295669      USB-C Charging Cable      1      11.95
5      295670      AA Batteries (4-pack)      1      3.84
6      295671      USB-C Charging Cable      1      11.95
7      295672      USB-C Charging Cable      2      11.95
8      295673      Bose SoundSport Headphones      1      99.99
9      295674      AAA Batteries (4-pack)      4      2.99
10     295675      USB-C Charging Cable      2      11.95
11     295676      ThinkPad Laptop      1      999.99
```

12	295677	AA Batteries (4-pack)	2	3.84
13	295678	AAA Batteries (4-pack)	2	2.99
14	295679	USB-C Charging Cable	1	11.95
15	295680	Lightning Charging Cable	1	14.95
16	295681	Google Phone	1	600
17	295681	USB-C Charging Cable	1	11.95
18	295681	Bose SoundSport Headphones	1	99.99
19	295681	Wired Headphones	1	11.99
20	295682	USB-C Charging Cable	1	11.95
21	295683	Wired Headphones	1	11.99
22	295684	AAA Batteries (4-pack)	4	2.99
23	295685	Wired Headphones	1	11.99
24	295686	USB-C Charging Cable	1	11.95

	Order Date	Purchase Address
0	12/30/19 00:01	136 Church St, New York City, NY 10001
1	12/29/19 07:03	562 2nd St, New York City, NY 10001
2	12/12/19 18:21	277 Main St, New York City, NY 10001
3	12/22/19 15:13	410 6th St, San Francisco, CA 94016
4	12/18/19 12:38	43 Hill St, Atlanta, GA 30301
5	12/31/19 22:58	200 Jefferson St, New York City, NY 10001
6	12/16/19 15:10	928 12th St, Portland, OR 97035
7	12/13/19 09:29	813 Hickory St, Dallas, TX 75001
8	12/15/19 23:26	718 Wilson St, Dallas, TX 75001
9	12/28/19 11:51	77 7th St, Dallas, TX 75001
10	12/13/19 13:52	594 1st St, San Francisco, CA 94016
11	12/28/19 17:19	410 Lincoln St, Los Angeles, CA 90001
12	12/20/19 19:19	866 Pine St, Boston, MA 02215
13	12/06/19 09:38	187 Lincoln St, Dallas, TX 75001
14	12/25/19 09:39	902 2nd St, Dallas, TX 75001
15	12/01/19 14:30	338 Main St, Austin, TX 73301
16	12/25/19 12:37	79 Elm St, Boston, MA 02215
17	12/25/19 12:37	79 Elm St, Boston, MA 02215
18	12/25/19 12:37	79 Elm St, Boston, MA 02215
19	12/25/19 12:37	79 Elm St, Boston, MA 02215
20	12/23/19 19:25	780 Elm St, Portland, OR 97035
21	12/23/19 22:46	341 Lake St, San Francisco, CA 94016
22	12/04/19 12:29	936 Church St, San Francisco, CA 94016
23	12/11/19 22:54	662 Ridge St, San Francisco, CA 94016
24	12/17/19 16:10	573 Maple St, Portland, ME 04101

```
[12]: all_data['Months']=all_data['Order Date'].str[0:2]
```

```
[13]: all_data['Months']
```

```
[13]: 0      12
      1      12
```

```

2         12
3         12
4         12
..
186845    06
186846    06
186847    06
186848    06
186849    06
Name: Months, Length: 186305, dtype: object

```

```
[14]: all_data['Price Each']=pd.to_numeric(all_data['Price Each'], errors='coerce')
```

```
[15]: all_data['Quantity Ordered']=pd.to_numeric(all_data['Quantity Ordered'],
↪errors='coerce')
```

```
[16]: all_data['Sales']=all_data['Price Each']*all_data['Quantity Ordered']
```

```
[17]: all_data['Sales'].head(25)
```

```

[17]: 0      1700.00
      1       600.00
      2       11.95
      3      149.99
      4       11.95
      5        3.84
      6       11.95
      7       23.90
      8       99.99
      9       11.96
     10       23.90
     11      999.99
     12        7.68
     13        5.98
     14       11.95
     15       14.95
     16      600.00
     17       11.95
     18       99.99
     19       11.99
     20       11.95
     21       11.99
     22       11.96
     23       11.99
     24       11.95
Name: Sales, dtype: float64

```

```
[18]: all_data.head()
```

```
[18]:   Order ID      Product  Quantity Ordered  Price Each  \
0    295665  Macbook Pro Laptop             1.0    1700.00
1    295666   LG Washing Machine             1.0     600.00
2    295667  USB-C Charging Cable             1.0     11.95
3    295668    27in FHD Monitor              1.0    149.99
4    295669  USB-C Charging Cable             1.0     11.95

      Order Date      Purchase Address  Months  Sales
0  12/30/19 00:01  136 Church St, New York City, NY 10001    12  1700.00
1  12/29/19 07:03   562 2nd St, New York City, NY 10001    12   600.00
2  12/12/19 18:21  277 Main St, New York City, NY 10001    12    11.95
3  12/22/19 15:13  410 6th St, San Francisco, CA 94016    12   149.99
4  12/18/19 12:38    43 Hill St, Atlanta, GA 30301    12    11.95
```

```
[48]: t_s=all_data.groupby(['Months'], as_index=False)['Sales'].sum()
```

```
[49]: t_s
```

```
[49]:   Months      Sales
0      01  1.822257e+06
1      02  2.202022e+06
2      03  2.807100e+06
3      04  3.390670e+06
4      05  3.152607e+06
5      06  2.577802e+06
6      07  2.647776e+06
7      08  2.244468e+06
8      09  2.097560e+06
9      10  3.736727e+06
10     11  3.199603e+06
11     12  4.613443e+06
12     Or  0.000000e+00
```

```
[50]: t_s = t_s.drop(labels=[12], axis=0)
```

```
[51]: t_s
```

```
[51]:   Months      Sales
0      01  1.822257e+06
1      02  2.202022e+06
2      03  2.807100e+06
3      04  3.390670e+06
4      05  3.152607e+06
5      06  2.577802e+06
6      07  2.647776e+06
```

7	08	2.244468e+06
8	09	2.097560e+06
9	10	3.736727e+06
10	11	3.199603e+06
11	12	4.613443e+06

```
[52]: t_s.sort_values("Sales", ascending=False)
```

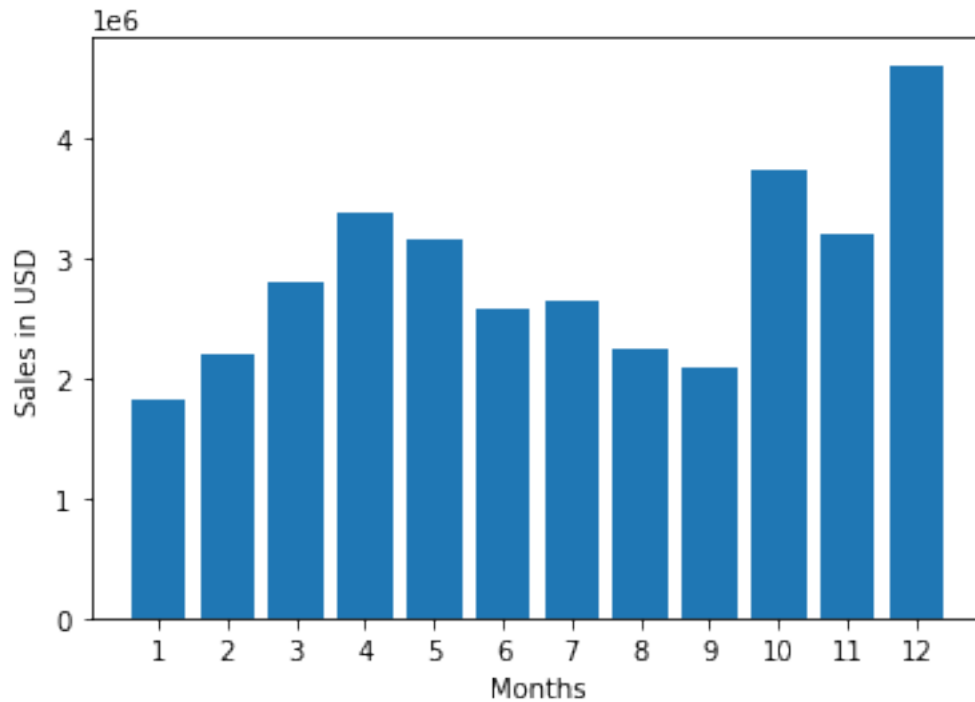
```
[52]:
```

	Months	Sales
11	12	4.613443e+06
9	10	3.736727e+06
3	04	3.390670e+06
10	11	3.199603e+06
4	05	3.152607e+06
2	03	2.807100e+06
6	07	2.647776e+06
5	06	2.577802e+06
7	08	2.244468e+06
1	02	2.202022e+06
8	09	2.097560e+06
0	01	1.822257e+06

```
[53]: import matplotlib.pyplot as plt
```

```
[54]: months=range(1,13)
```

```
[55]: plt.bar(months, t_s['Sales'])
plt.xticks(months)
plt.xlabel('Months')
plt.ylabel('Sales in USD')
plt.show()
```



```
[56]: all_data.dtypes
```

```
[56]: Order ID      object
      Product      object
      Quantity Ordered  float64
      Price Each    float64
      Order Date     object
      Purchase Address object
      Months         object
      Sales          float64
      City           object
      dtype: object
```

```
[57]: all_data['City'] = all_data['Purchase Address'].str.split(',').str.
      ↪get(1)+','+all_data['Purchase Address'].str.split(',').str.get(2)
```

```
[105]: all_data.head()
```

```
[105]:  Order ID      Product  Quantity Ordered  Price Each  \
0    295665  Macbook Pro Laptop             1.0      1700.00
1    295666   LG Washing Machine             1.0       600.00
2    295667  USB-C Charging Cable             1.0        11.95
3    295668    27in FHD Monitor             1.0      149.99
4    295669  USB-C Charging Cable             1.0        11.95
```


	Order Date	Purchase Address	Months	Sales \
0	12/30/19 00:01	136 Church St, New York City, NY 10001	12	1700.00
1	12/29/19 07:03	562 2nd St, New York City, NY 10001	12	600.00
2	12/12/19 18:21	277 Main St, New York City, NY 10001	12	11.95
3	12/22/19 15:13	410 6th St, San Francisco, CA 94016	12	149.99
4	12/18/19 12:38	43 Hill St, Atlanta, GA 30301	12	11.95

	City
0	New York City, NY 10001
1	New York City, NY 10001
2	New York City, NY 10001
3	San Francisco, CA 94016
4	Atlanta, GA 30301

```
[60]: city_sales=all_data.groupby(['City'], as_index=False)['Sales'].sum()
```

```
[62]: city_sales
```

```
[62]:
```

	City	Sales
0	Atlanta, GA 30301	2.795499e+06
1	Austin, TX 73301	1.819582e+06
2	Boston, MA 02215	3.661642e+06
3	Dallas, TX 75001	2.767975e+06
4	Los Angeles, CA 90001	5.452571e+06
5	New York City, NY 10001	4.664317e+06
6	Portland, ME 04101	4.497583e+05
7	Portland, OR 97035	1.870732e+06
8	San Francisco, CA 94016	8.262204e+06
9	Seattle, WA 98101	2.747755e+06

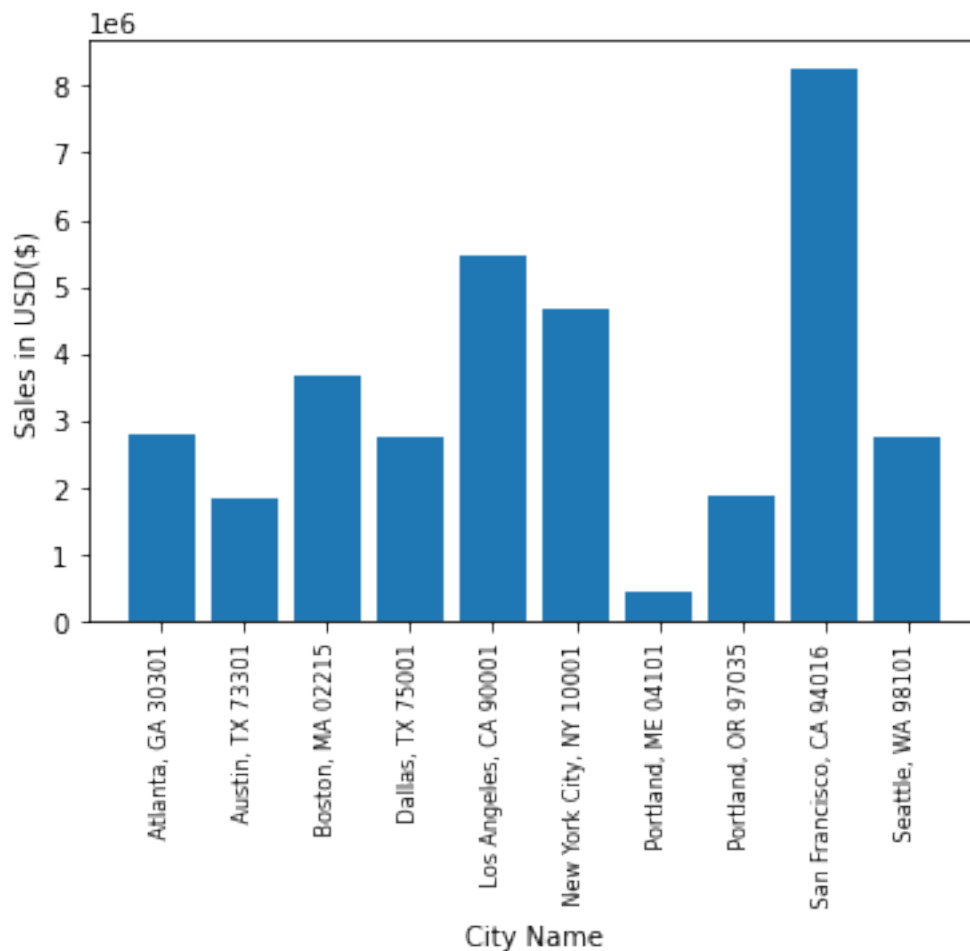
```
[63]: city_sales.sort_values("Sales", ascending=False)
```

```
[63]:
```

	City	Sales
8	San Francisco, CA 94016	8.262204e+06
4	Los Angeles, CA 90001	5.452571e+06
5	New York City, NY 10001	4.664317e+06
2	Boston, MA 02215	3.661642e+06
0	Atlanta, GA 30301	2.795499e+06
3	Dallas, TX 75001	2.767975e+06
9	Seattle, WA 98101	2.747755e+06
7	Portland, OR 97035	1.870732e+06
1	Austin, TX 73301	1.819582e+06
6	Portland, ME 04101	4.497583e+05

```
[98]: import matplotlib.pyplot as plt
cities=[city for city, df in city_sales.groupby('City')]
```

```
[101]: plt.bar(cities, city_sales['Sales'])
plt.xticks(cities, rotation='vertical', size=8)
plt.xlabel('City Name')
plt.ylabel('Sales in USD($)' )
plt.show()
```



```
[110]: all_data.head()
```

```
[110]:
```

	Order ID	Product	Quantity Ordered	Price Each	\
0	295665	Macbook Pro Laptop	1.0	1700.00	
1	295666	LG Washing Machine	1.0	600.00	
2	295667	USB-C Charging Cable	1.0	11.95	
3	295668	27in FHD Monitor	1.0	149.99	
4	295669	USB-C Charging Cable	1.0	11.95	

	Order Date	Purchase Address	Months	Sales	\
0	12/30/19 00:01	136 Church St, New York City, NY 10001	12	1700.00	

1	12/29/19 07:03	562 2nd St, New York City, NY 10001	12	600.00
2	12/12/19 18:21	277 Main St, New York City, NY 10001	12	11.95
3	12/22/19 15:13	410 6th St, San Francisco, CA 94016	12	149.99
4	12/18/19 12:38	43 Hill St, Atlanta, GA 30301	12	11.95

	City
0	New York City, NY 10001
1	New York City, NY 10001
2	New York City, NY 10001
3	San Francisco, CA 94016
4	Atlanta, GA 30301

```
[115]: all_data['Order Date'] = pd.to_datetime(all_data['Order Date'], errors='coerce')
```

```
[116]: all_data.head()
```

```
[116]:
```

	Order ID	Product	Quantity Ordered	Price Each	\
0	295665	Macbook Pro Laptop	1.0	1700.00	
1	295666	LG Washing Machine	1.0	600.00	
2	295667	USB-C Charging Cable	1.0	11.95	
3	295668	27in FHD Monitor	1.0	149.99	
4	295669	USB-C Charging Cable	1.0	11.95	

	Order Date	Purchase Address	Months	Sales	\
0	2019-12-30 00:01:00	136 Church St, New York City, NY 10001	12	1700.00	
1	2019-12-29 07:03:00	562 2nd St, New York City, NY 10001	12	600.00	
2	2019-12-12 18:21:00	277 Main St, New York City, NY 10001	12	11.95	
3	2019-12-22 15:13:00	410 6th St, San Francisco, CA 94016	12	149.99	
4	2019-12-18 12:38:00	43 Hill St, Atlanta, GA 30301	12	11.95	

	City
0	New York City, NY 10001
1	New York City, NY 10001
2	New York City, NY 10001
3	San Francisco, CA 94016
4	Atlanta, GA 30301

```
[131]: all_data['Hour'] = all_data['Order Date'].dt.hour
```

```
[129]: all_data.head()
```

```
[129]:
```

	Order ID	Product	Quantity Ordered	Price Each	\
0	295665	Macbook Pro Laptop	1.0	1700.00	
1	295666	LG Washing Machine	1.0	600.00	
2	295667	USB-C Charging Cable	1.0	11.95	
3	295668	27in FHD Monitor	1.0	149.99	
4	295669	USB-C Charging Cable	1.0	11.95	

	Order Date	Purchase Address	Months	Sales \
0	2019-12-30 00:01:00	136 Church St, New York City, NY 10001	12	1700.00
1	2019-12-29 07:03:00	562 2nd St, New York City, NY 10001	12	600.00
2	2019-12-12 18:21:00	277 Main St, New York City, NY 10001	12	11.95
3	2019-12-22 15:13:00	410 6th St, San Francisco, CA 94016	12	149.99
4	2019-12-18 12:38:00	43 Hill St, Atlanta, GA 30301	12	11.95

	City	Time	Hour
0	New York City, NY 10001	0.0	0.0
1	New York City, NY 10001	7.0	7.0
2	New York City, NY 10001	18.0	18.0
3	San Francisco, CA 94016	15.0	15.0
4	Atlanta, GA 30301	12.0	12.0

```
[132]: all_data['Minute']=all_data['Order Date'].dt.minute
```

```
[133]: all_data.head()
```

```
[133]:
```

	Order ID	Product	Quantity Ordered	Price Each \
0	295665	Macbook Pro Laptop	1.0	1700.00
1	295666	LG Washing Machine	1.0	600.00
2	295667	USB-C Charging Cable	1.0	11.95
3	295668	27in FHD Monitor	1.0	149.99
4	295669	USB-C Charging Cable	1.0	11.95

	Order Date	Purchase Address	Months	Sales \
0	2019-12-30 00:01:00	136 Church St, New York City, NY 10001	12	1700.00
1	2019-12-29 07:03:00	562 2nd St, New York City, NY 10001	12	600.00
2	2019-12-12 18:21:00	277 Main St, New York City, NY 10001	12	11.95
3	2019-12-22 15:13:00	410 6th St, San Francisco, CA 94016	12	149.99
4	2019-12-18 12:38:00	43 Hill St, Atlanta, GA 30301	12	11.95

	City	Time	Hour	Minute
0	New York City, NY 10001	0.0	0.0	1.0
1	New York City, NY 10001	7.0	7.0	3.0
2	New York City, NY 10001	18.0	18.0	21.0
3	San Francisco, CA 94016	15.0	15.0	13.0
4	Atlanta, GA 30301	12.0	12.0	38.0

```
[134]: sales_per_hour=city_sales=all_data.groupby(['Hour'], as_index=False)['Sales'].
      ↪sum()
```

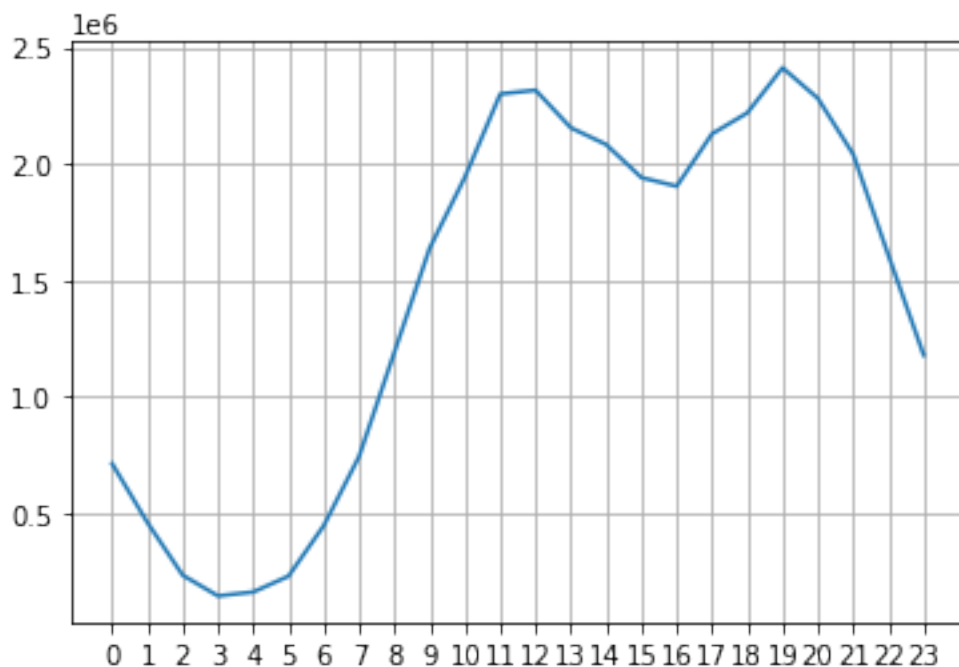
```
[135]: sales_per_hour
```

```
[135]:
```

	Hour	Sales
0	0.0	713721.27

1	1.0	460866.88
2	2.0	234851.44
3	3.0	145757.89
4	4.0	162661.01
5	5.0	230679.82
6	6.0	448113.00
7	7.0	744854.12
8	8.0	1192348.97
9	9.0	1639030.58
10	10.0	1944286.77
11	11.0	2300610.24
12	12.0	2316821.34
13	13.0	2155389.80
14	14.0	2083672.73
15	15.0	1941549.60
16	16.0	1904601.31
17	17.0	2129361.61
18	18.0	2219348.30
19	19.0	2412938.54
20	20.0	2281716.24
21	21.0	2042000.86
22	22.0	1607549.21
23	23.0	1179304.44

```
[152]: import numpy as np
import matplotlib.pyplot as plt
p=plt.plot(sales_per_hour['Hour'], sales_per_hour['Sales'])
plt.xticks(sales_per_hour['Hour'])
plt.grid()
plt.show(p)
```



```
[155]: all_data.drop('Time', inplace=True, axis=1)
```

```
[156]: all_data
```

```
[156]:
```

	Order ID	Product	Quantity Ordered	Price Each	\
0	295665	Macbook Pro Laptop	1.0	1700.00	
1	295666	LG Washing Machine	1.0	600.00	
2	295667	USB-C Charging Cable	1.0	11.95	
3	295668	27in FHD Monitor	1.0	149.99	
4	295669	USB-C Charging Cable	1.0	11.95	
...	
186845	222905	AAA Batteries (4-pack)	1.0	2.99	
186846	222906	27in FHD Monitor	1.0	149.99	
186847	222907	USB-C Charging Cable	1.0	11.95	
186848	222908	USB-C Charging Cable	1.0	11.95	
186849	222909	AAA Batteries (4-pack)	1.0	2.99	

	Order Date	Purchase Address	Months	\
0	2019-12-30 00:01:00	136 Church St, New York City, NY 10001	12	
1	2019-12-29 07:03:00	562 2nd St, New York City, NY 10001	12	
2	2019-12-12 18:21:00	277 Main St, New York City, NY 10001	12	
3	2019-12-22 15:13:00	410 6th St, San Francisco, CA 94016	12	
4	2019-12-18 12:38:00	43 Hill St, Atlanta, GA 30301	12	
...	
186845	2019-06-07 19:02:00	795 Pine St, Boston, MA 02215	06	

186846	2019-06-01	19:29:00	495 North St, New York City, NY 10001	06
186847	2019-06-22	18:57:00	319 Ridge St, San Francisco, CA 94016	06
186848	2019-06-26	18:35:00	916 Main St, San Francisco, CA 94016	06
186849	2019-06-25	14:33:00	209 11th St, Atlanta, GA 30301	06

	Sales		City	Hour	Minute
0	1700.00	New York City, NY	10001	0.0	1.0
1	600.00	New York City, NY	10001	7.0	3.0
2	11.95	New York City, NY	10001	18.0	21.0
3	149.99	San Francisco, CA	94016	15.0	13.0
4	11.95	Atlanta, GA	30301	12.0	38.0
...
186845	2.99	Boston, MA	02215	19.0	2.0
186846	149.99	New York City, NY	10001	19.0	29.0
186847	11.95	San Francisco, CA	94016	18.0	57.0
186848	11.95	San Francisco, CA	94016	18.0	35.0
186849	2.99	Atlanta, GA	30301	14.0	33.0

[186305 rows x 11 columns]