<pre>files = [file for file in os.listdir(path) all_months_data = pd.DataFrame() for file in files: current_data = pd.read_csv(path+"/"+fil all_months_data = pd.concat([all_months] all_months_data.to_csv("all_data_copy.csv",</pre>	one csv file	
<pre>df = pd.read_csv('Sales_Data/all_data_copy. df.head()</pre>	s_data, current_data]) , index=False) .csv')	
2 176559 Bose SoundSport Headphones 3 176560 Google Phone 4 176560 Wired Headphones df.shape (186850, 6)	dered Price Each Order Date Purchase Address 2 11.95 04/19/19 08:46 917 1st St, Dallas, TX 75001 NaN NaN NaN NaN 99.99 04/07/19 22:30 682 Chestnut St, Boston, MA 02215 1 600 04/12/19 14:38 669 Spruce St, Los Angeles, CA 90001 1 11.99 04/12/19 14:38 669 Spruce St, Los Angeles, CA 90001	
df.info() <class 'pandas.core.frame.dataframe'=""> RangeIndex: 186850 entries, 0 to 186849 Data columns (total 6 columns): # Column Non-Null Count Dty</class>	ject ject ject ject ject	
dtypes: object(6) memory usage: 8.6+ MB df.dtypes Order ID object Product object Quantity Ordered object Price Each object Order Date object Purchase Address object dtype: object		
<pre>df.isna().sum() Order ID</pre>		
Order ID	tity Ordered Price Each Order Date Purchase Address 1 3.84 03/17/19 11:00 459 1st St, San Francisco, CA 94016	
161874 266124 Macbook Pro Laptop 20329 238537 34in Ultrawide Monitor 113149 169542 20in Monitor 165626 269705 Lightning Charging Cable 123201 196664 Bose SoundSport Headphones 130283 203390 Bose SoundSport Headphones 186453 258983 AAA Batteries (4-pack) 130390 203488 USB-C Charging Cable 184847 257445 AA Batteries (4-pack)	1 1700 10/25/19 16:38 136 Hickory St, Austin, TX 73301 1 379.99 08/29/19 19:45 833 5th St, Boston, MA 02215 1 109.99 03/05/19 15:39 964 Hill St, Boston, MA 02215 1 14.95 10/16/19 10:26 543 2nd St, San Francisco, CA 94016 1 99.99 05/31/19 20:19 427 Madison St, Dallas, TX 75001 1 99.99 05/07/19 20:33 913 Lake St, Los Angeles, CA 90001 1 2.99 09/01/19 09:14 522 West St, Los Angeles, CA 90001 1 11.95 05/07/19 08:29 953 Meadow St, Seattle, WA 98101 1 3.84 09/28/19 21:22 451 River St, Atlanta, GA 30301	
<pre>Data Cleaning # converting the Order ID column to int df[-df['Order ID'].str.isdigit()] to_drop = df[-df['Order ID'].str.isdigit()] df = df.drop(to_drop) df[-df['Order ID'].str.isdigit()] Order ID Product Quantity Ordered Price Each Order</pre>		
<pre>df['Order ID'] = df['Order ID'].astype(int) df['Order ID'].dtype dtype('int32') # converting the Quantity Order column to i df['Quantity Ordered'] = df['Quantity Order # Price Each column to float df['Price Each'] = df['Price Each'].astype(</pre>	<pre>int red'].astype(int) (float)</pre>	
<pre>df['Order Date'] = pd.to_datetime(df['Order df['Order Date'].dtype dtype('<m8[ns]') 1,="" 12,="" 2,="" 3,="" 4,="" 5,="" 7,="" 8,="" 9,="" array([="" date'].dt.mor="" df.dtypes="" df['order="" id<="" month']="df['Order" month'].unique()="" order="" td=""><td>nth</td><td></td></m8[ns]')></pre>	nth	
Product object Quantity Ordered int32 Price Each float64 Order Date datetime64[ns] Purchase Address object Order Month int32 dtype: object df['Total Price'] = df['Price Each']*df['Qudf.head() Order ID Product Quantity Order 176558 USB-C Charging Cable	dered Price Each Order Date Purchase Address Order Month Total Price 2 11.95 2019-04-19 08:46:00 917 1st St, Dallas, TX 75001 4 23.90	
2 176559 Bose SoundSport Headphones 3 176560 Google Phone 4 176560 Wired Headphones 5 176561 Wired Headphones EDA Univariate analysis df['Total Price'].describe()	1 99.99 2019-04-07 22:30:00 682 Chestnut St, Boston, MA 02215 4 99.99 1 600.00 2019-04-12 14:38:00 669 Spruce St, Los Angeles, CA 90001 4 600.00 1 11.99 2019-04-12 14:38:00 669 Spruce St, Los Angeles, CA 90001 4 11.99 1 11.99 2019-04-30 09:27:00 333 8th St, Los Angeles, CA 90001 4 11.99	
	on\Python310\lib\site-packages\seaborn_oldcore.py:1498: FutureWarning:	
is_categorical_dtype is deprecated and will <axes: xlabel="Total Price"></axes:>	l be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead	
0 500 1000 1500 2000	2500 3000 3500	
fig= px.box(df['Total Price']) fig.show() 3500 3000	•	
2500 2000 1500 1000 500		
<pre>df = df[df['Total Price'] < 1000] df['Total Price'].describe() count 181210.000000</pre>	Total Price variable	
count 181210.000000 mean 145.865217 std 228.161760 min 2.990000 25% 11.950000 75% 150.000000 max 999.990000 Name: Total Price, dtype: float64 fig= px.box(df['Total Price']) fig.show()		
1000		
400 200 0	Total Price variable	
is_categorical_dtype is deprecated and will C:\Users\top10\AppData\Local\Programs\Pytho	<pre>Price':'sum'})</pre>	
<pre></pre>		
60000 - 40000 - 20000 - 0 250 500 750 10	000 1250 1500 1750 2000 I Price	
<pre>fig= px.box(Orders['Total Price']) fig.show()</pre> 2000 1500		
1000 500		
conclusion: most customers make orders with total pri month_total = df.groupby('Order Month')['To months = range(1, 13) fig= px.bar(month_total,x = months, y= mont fig.show()	otal Price'].sum()	
3.5M 3M 2.5M		
2M 1.5M 1M 0.5M 0		
# the last three months of the year have the # December is the best month for sales df.head() Order ID Product Quantity Order ID USB-C Charging Cable 176558 USB-C Charging Cable 176559 Bose SoundSport Headphones	dered Price Each Order Date Purchase Address Order Month Total Price 2 11.95 2019-04-19 08:46:00 917 1st St, Dallas, TX 75001 4 23.90	
3 176560 Google Phone 4 176560 Wired Headphones 5 176561 Wired Headphones Bivariate analysis sns.barplot(data= df, x= 'Order Month', y= plt.ylabel('Total Price mean')	1 600.00 2019-04-12 14:38:00 669 Spruce St, Los Angeles, CA 90001 4 600.00 1 11.99 2019-04-12 14:38:00 669 Spruce St, Los Angeles, CA 90001 4 11.99 1 11.99 2019-04-30 09:27:00 333 8th St, Los Angeles, CA 90001 4 11.99	
C:\Users\top10\AppData\Local\Programs\Pytho is_categorical_dtype is deprecated and will C:\Users\top10\AppData\Local\Programs\Pytho	l be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead on\Python310\lib\site-packages\seaborn_oldcore.py:1498: FutureWarning: l be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead on\Python310\lib\site-packages\seaborn_oldcore.py:1498: FutureWarning: l be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead	
140 - 120 - 80 - 100 -		
# the mean of the Total Price is about 150	nth	
# Adding city column df['City'] = df['Purchase Address'].apply(1	Lambda x: x. split(',')[1]) tity Ordered Price Each Order Date Purchase Address Order Month Total Price City 2 11.95 2019-04-19 08:46:00 917 1st St, Dallas, TX 75001 4 23.90 Dallas 1 99.99 2019-04-07 22:30:00 682 Chestnut St, Boston, MA 02215 4 99.99 Boston 1 600.00 2019-04-12 14:38:00 669 Spruce St, Los Angeles, CA 90001 4 600.00 Los Angeles 1 11.99 2019-04-30 09:27:00 333 8th St, Los Angeles, CA 90001 4 11.99 Los Angeles	
186845 259353 AAA Batteries (4-pack) 186846 259354 iPhone 186847 259355 iPhone 186848 259356 34in Ultrawide Monitor 186849 259357 USB-C Charging Cable 181210 rows × 9 columns	3 2.99 2019-09-17 20:56:00 840 Highland St, Los Angeles, CA 90001 9 8.97 Los Angeles 1 700.00 2019-09-01 16:00:00 216 Dogwood St, San Francisco, CA 94016 9 700.00 San Francisco 1 700.00 2019-09-23 07:39:00 220 12th St, San Francisco, CA 94016 9 700.00 San Francisco 1 379.99 2019-09-19 17:30:00 511 Forest St, San Francisco, CA 94016 9 379.99 San Francisco 1 11.95 2019-09-30 00:18:00 250 Meadow St, San Francisco, CA 94016 9 11.95 San Francisco	
sales_per_city= df.groupby('City').agg({'Qu fig = px.bar(sales_per_city, labels={'x':'C fig.show() Quantity ordered per City 50k	City', 'y':'Quantity ordered'}, title= 'Quantity ordered per City', color= sales_per_city.index) City Atlanta Austin Boston Dallas Los Angeles Los Angeles	
30k 20k 10k	Los Angeles New York City Portland San Francisco Seattle	
# San Freancisco is the best City for Sales df['Hour'] = df['Order Date'].dt.hour df.head() Order ID Product Quantity Ord 176558 USB-C Charging Cable		
2 176559 Bose SoundSport Headphones 3 176560 Google Phone 4 176560 Wired Headphones 5 176561 Wired Headphones # relationship between day hour and Quantit df['Hour'] = df['Hour'].astype('category') fig = px.bar(df['Hour'].value_counts()) fig.show()	1 99.99 2019-04-07 22:30:00 682 Chestnut St, Boston, MA 02215 4 99.99 Boston 22 1 600.00 2019-04-12 14:38:00 669 Spruce St, Los Angeles, CA 90001 4 600.00 Los Angeles 14 1 11.99 2019-04-12 14:38:00 669 Spruce St, Los Angeles, CA 90001 4 11.99 Los Angeles 14 1 11.99 2019-04-30 09:27:00 333 8th St, Los Angeles, CA 90001 4 11.99 Los Angeles 9	
12k 10k 8k	variable count	
6k 4k 2k 0	5 10 15 20 Hour	
# Orders decrease after 12 AM till 7 AM as		

Importing needed libraries