

```
In [56]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

In [57]: train=pd.read_csv(r'E:\VAILM TRAINING\intern 1\train_u940skv (2) (1) (1) (1).csv')

In [58]: train

Out[58]:
```

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year
0	FDA15	9.300	Low Fat	0.016047	Dairy	249.8092	OUT049	
1	DRC01	5.920	Regular	0.019278	Soft Drinks	48.2692	OUT018	
2	FDM15	17.500	Low Fat	0.016760	Meat	141.6180	OUT049	
3	FDX07	19.200	Regular	0.054021	Fruits and Vegetables	182.0950	OUT010	
4	NC019	8.930	Low Fat	0.054021	Household	53.8614	OUT013	
...
8518	FD022	6.865	Low Fat	0.056783	Snack Foods	214.5218	OUT013	
8519	FD036	8.380	Regular	0.049962	Baking Goods	108.1370	OUT045	
8520	NC029	10.600	Low Fat	0.039186	Health and Hygiene	85.1224	OUT035	
8521	FD046	7.210	Regular	0.145221	Snack Foods	103.1332	OUT018	
8522	DRC01	14.800	Low Fat	0.044878	Soft Drinks	75.4670	OUT046	

8523 rows x 12 columns

```
In [59]: train=pd.read_csv(r'E:\VAILM TRAINING\intern 1\Test_u940SKV (2) (1) (1) (1).csv')

In [60]: test

Out[60]:
```

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year
0	FD058	20.750	Low Fat	0.007565	Snack Foods	107.8622	OUT049	
1	FDW14	8.300	reg	0.038428	Dairy	87.3198	OUT017	
2	NC055	14.600	Low Fat	0.096575	Others	241.7538	OUT010	
3	FD058	7.315	Low Fat	0.015388	Snack Foods	155.0340	OUT017	
4	FDY38	NaN	Regular	0.118999	Dairy	234.2300	OUT027	
...
5676	FD058	10.500	Regular	0.013496	Snack Foods	141.3154	OUT046	
5677	FD047	7.600	Regular	0.142991	Starchy Foods	169.1448	OUT018	
5678	NC017	10.000	Low Fat	0.073529	Health and Hygiene	118.7440	OUT045	
5679	FD026	15.300	Regular	0.054021	Canned	214.6218	OUT017	
5680	FDU37	9.500	Regular	0.104720	Canned	79.7960	OUT045	

5681 rows x 11 columns

```
In [61]: train.shape

Out[61]: (8523, 12)

In [62]: test.shape

Out[62]: (5681, 11)

In [63]: train.info

Out[63]:
<bound method DataFrame.info of
0      FDA15      9.300      Low Fat      0.016047      Dairy      249.8092      OUT049
1      DRC01      5.920      Regular      0.019278      Soft Drinks      48.2692      OUT018
2      FDM15     17.500      Low Fat      0.016760      Meat      141.6180      OUT049
3      FDX07     19.200      Regular      0.054021      Fruits and Vegetables      182.0950      OUT010
4      NCD19      8.930      Low Fat      0.054021      Household      53.8614      OUT013
...
8518     FD022      6.865      Low Fat      0.056783      Snack Foods      214.5218      OUT013
8519     FD036      8.380      Regular      0.049962      Baking Goods      108.1370      OUT045
8520     NC029     10.600      Low Fat      0.039186      Health and Hygiene      85.1224      OUT035
8521     FD046      7.210      Regular      0.145221      Snack Foods      103.1332      OUT018
8522     DRC01     14.800      Low Fat      0.044878      Soft Drinks      75.4670      OUT046

Item_Type      Item_MRP      Outlet_Identifier \
0      Dairy      249.8092      OUT049
1      Soft Drinks      48.2692      OUT018
2      Meat      141.6180      OUT049
3      Fruits and Vegetables      182.0950      OUT010
4      Household      53.8614      OUT013
...
8518     Snack Foods      214.5218      OUT013
8519     Baking Goods      108.1370      OUT045
8520     Health and Hygiene      85.1224      OUT035
8521     Snack Foods      103.1332      OUT018
8522     Soft Drinks      75.4670      OUT046

Outlet_Establishment_Year      Outlet_Size      Outlet_Location_Type \
0      1999      Medium      Tier 1
1      2000      Medium      Tier 3
2      1999      Medium      Tier 1
3      1999      Medium      Tier 1
4      1987      High      Tier 3
...
8518     1987      High      Tier 3
8519     2002      NaN      Tier 2
8520     2004      NaN      Tier 2
8521     2009      Medium      Tier 3
8522     1997      Small      Tier 1

Outlet_Type      Item_Outlet_Sales
0      Supermarket Type1      3725.2380
1      Supermarket Type2      443.4228
2      Supermarket Type1      2097.2708
3      Grocery Store      732.3809
4      Supermarket Type1      994.7852
...
8518     Supermarket Type1      2778.3834
8519     Supermarket Type1      549.2850
8520     Supermarket Type1      1193.1136
8521     Supermarket Type2      1845.5976
8522     Supermarket Type1      765.6700

Item_Type      Item_MRP      Outlet_Identifier \
0      Dairy      249.8092      OUT049
1      Soft Drinks      48.2692      OUT018
2      Meat      141.6180      OUT049
3      Fruits and Vegetables      182.0950      OUT010
4      Household      53.8614      OUT013
...
8518     Snack Foods      214.5218      OUT013
8519     Baking Goods      108.1370      OUT045
8520     Health and Hygiene      85.1224      OUT035
8521     Snack Foods      103.1332      OUT018
8522     Soft Drinks      75.4670      OUT046

Outlet_Type      Item_Outlet_Sales
0      Supermarket Type1      3725.2380
1      Supermarket Type2      443.4228
2      Supermarket Type1      2097.2708
3      Grocery Store      732.3809
4      Supermarket Type1      994.7852
...
8518     Supermarket Type1      2778.3834
8519     Supermarket Type1      549.2850
8520     Supermarket Type1      1193.1136
8521     Supermarket Type2      1845.5976
8522     Supermarket Type1      765.6700

In [64]: train.dtypes

Out[64]:
Item_Identifier      object
Item_Weight          float64
Item_Fat_Content     object
Item_Visibility       float64
Item_Type            object
Item_MRP             object
Outlet_Identifier     int64
Outlet_Establishment_Year      object
Outlet_Location_Type  object
Outlet_Type           object
Item_Outlet_Sales     float64
dtype: object

In [65]: type(train)

Out[65]: pandas.core.frame.DataFrame

In [66]: train.shape[1]

Out[66]: 12

In [67]: train['Item_Weight'].isnull().sum()

Out[67]: 1463

In [68]: plt.figure(figsize=(19,10))
x=train['Item_Visibility'], y=train['Item_Outlet_Sales']
plt.xlabel('Item_Visibility'), plt.ylabel('Item_Outlet_Sales')
plt.scatter(x,y,z=5,label='Sales',color='b')
plt.title('Item_Visibility vs Item_Outlet_Sales')
plt.legend()
plt.show()
```