

Installing packages .Loading & inspecting dataset to have BigMart Sales Dataset

```
In [1]: import numpy as np#
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [2]: df=pd.read_csv(r"D:\BigMart Saleas dataset.csv")
df
```

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Size	Outlet_Location_Type	Outlet_Type
0	FDW58	20.750	Low Fat	0.007565	Snack Foods	107.8622	OUT049	1999	Medium	Tier 1	Supermarket Type1
1	FDW14	8.300	reg	0.038428	Dairy	87.3198	OUT017	2007	NaN	Tier 2	Supermarket Type1
2	NCN55	14.600	Low Fat	0.099575	Others	241.7538	OUT010	1998	NaN	Tier 3	Grocery Store
3	FDQ58	7.315	Low Fat	0.015388	Snack Foods	155.0340	OUT017	2007	NaN	Tier 2	Supermarket Type1
4	FDY38	NaN	Regular	0.118599	Dairy	234.2300	OUT027	1985	Medium	Tier 3	Supermarket Type3
...
5676	FDB58	10.500	Regular	0.013496	Snack Foods	141.3154	OUT046	1997	Small	Tier 1	Supermarket Type1
5677	FDD47	7.600	Regular	0.142991	Starchy Foods	169.1448	OUT018	2009	Medium	Tier 3	Supermarket Type2
5678	NC017	10.000	Low Fat	0.073529	Health and Hygiene	118.7440	OUT045	2002	NaN	Tier 2	Supermarket Type1
5679	FDJ26	15.300	Regular	0.000000	Canned	214.6218	OUT017	2007	NaN	Tier 2	Supermarket Type1
5680	FDU37	9.500	Regular	0.104720	Canned	79.7960	OUT045	2002	NaN	Tier 2	Supermarket Type1

5681 rows × 11 columns

```
In [3]: df.shape
```

Out[3]: (5681, 11)

```
In [4]: df.head(10)
```

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Size	Outlet_Location_Type	Outlet_Type
0	FDW58	20.750	Low Fat	0.007565	Snack Foods	107.8622	OUT049	1999	Medium	Tier 1	Supermarket Type1
1	FDW14	8.300	reg	0.038428	Dairy	87.3198	OUT017	2007	NaN	Tier 2	Supermarket Type1
2	NCN55	14.600	Low Fat	0.099575	Others	241.7538	OUT010	1998	NaN	Tier 3	Grocery Store
3	FDQ58	7.315	Low Fat	0.015388	Snack Foods	155.0340	OUT017	2007	NaN	Tier 2	Supermarket Type1
4	FDY38	NaN	Regular	0.118599	Dairy	234.2300	OUT027	1985	Medium	Tier 3	Supermarket Type3
5	FDH56	9.800	Regular	0.063817	Fruits and Vegetables	117.1492	OUT046	1997	Small	Tier 1	Supermarket Type1
6	FDL48	19.350	Regular	0.082602	Baking Goods	50.1034	OUT018	2009	Medium	Tier 3	Supermarket Type2
7	FDC48	NaN	Low Fat	0.015782	Baking Goods	81.0592	OUT027	1985	Medium	Tier 3	Supermarket Type3
8	FDN33	6.305	Regular	0.123365	Snack Foods	95.7436	OUT045	2002	NaN	Tier 2	Supermarket Type1
9	FDA36	5.985	Low Fat	0.005698	Baking Goods	186.8924	OUT017	2007	NaN	Tier 2	Supermarket Type1

```
In [5]: df.tail()
```

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Size	Outlet_Location_Type	Outlet_Type
5676	FDB58	10.5	Regular	0.013496	Snack Foods	141.3154	OUT046	1997	Small	Tier 1	Supermarket Type1
5677	FDD47	7.6	Regular	0.142991	Starchy Foods	169.1448	OUT018	2009	Medium	Tier 3	Supermarket Type2
5678	NC017	10.0	Low Fat	0.073529	Health and Hygiene	118.7440	OUT045	2002	NaN	Tier 2	Supermarket Type1
5679	FDJ26	15.3	Regular	0.000000	Canned	214.6218	OUT017	2007	NaN	Tier 2	Supermarket Type1
5680	FDU37	9.5	Regular	0.104720	Canned	79.7960	OUT045	2002	NaN	Tier 2	Supermarket Type1

```
In [6]: df.isnull()
```

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Size	Outlet_Location_Type	Outlet_Type
0	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	True	False	False
2	False	False	False	False	False	False	False	False	True	False	False
3	False	False	False	False	False	False	False	False	True	False	False
4	False	True	False	False	False	False	False	False	False	False	False
...
5676	False	False	False	False	False	False	False	False	False	False	False
5677	False	False	False	False	False	False	False	False	False	False	False
5678	False	False	False	False	False	False	False	False	True	False	False
5679	False	False	False	False	False	False	False	False	True	False	False
5680	False	False	False	False	False	False	False	False	True	False	False

5681 rows × 11 columns

```
In [7]: df.isnull().sum()
```

```
Out[7]: Item_Identifier      0
Item_Weight      976
Item_Fat_Content      0
Item_Visibility      0
Item_Type          0
Item_MRP           0
Outlet_Identifier    0
Outlet_Establishment_Year  0
Outlet_Size      1606
Outlet_Location_Type  0
Outlet_Type          0
dtype: int64
```

```
In [8]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5681 entries, 0 to 5680
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Item_Identifier        5681 non-null  object
1   Item_Weight            4795 non-null  float64
2   Item_Fat_Content       5681 non-null  object
3   Item_Visibility        5681 non-null  float64
4   Item_Type              5681 non-null  object
5   Item_MRP               5681 non-null  float64
6   Outlet_Identifier       5681 non-null  object
7   Outlet_Establishment_Year 5681 non-null  int64
8   Outlet_Size            4075 non-null  object
9   Outlet_Location_Type   5681 non-null  object
10  Outlet_Type            5681 non-null  object
dtypes: float64(3), int64(1), object(7)
memory usage: 488.3+ KB
```

```
In [9]: df.dropna(inplace=True)
```

```
In [10]: df.describe()
```

	Item_Weight	Item_Visibility	Item_MRP	Outlet_Establishment_Year
count	3099.000000	3099.000000	3099.000000	3099.000000
mean	12.658206	0.059063	140.036189	1999.189739
std	4.616934	0.043592	61.608135	7.387906
min	4.555000	0.000000	31.990000	1987.000000
25%	8.675000	0.025824	93.310700	1997.000000
50%	12.350000	0.049154	141.315400	1999.000000
75%	16.600000	0.085354	184.993700	2004.000000
max	21.350000	0.185913	266.588400	2009.000000

```
In [11]: df.shape
```

Out[11]: (3099, 11)

```
In [12]: df1=df[['Item_Identifier','Outlet_Location_Type','Item_Weight','Outlet_Identifier','Outlet_Size','Item_Visibility','Item_MRP']]
df1
```

	Item_Identifier	Outlet_Location_Type	Item_Weight	Outlet_Identifier	Outlet_Size	Item_Visibility	Item_MRP
0	FDW58	Tier 1	20.750	OUT049	Medium	0.007565	107.8622
5	FDH56	Tier 1	9.800	OUT046	Small	0.063817	117.1492
6	FDL48	Tier 3	19.350	OUT018	Medium	0.082602	50.1034
13	FDU11	Tier 1	4.785	OUT049	Medium	0.092738	122.3098
14	DRL59	Tier 3	16.750	OUT013	High	0.021206	52.0298
...
5673	FDF46	Tier 3	7.070	OUT018	Medium	0.094053	116.0834
5674	DRL35	Tier 1	15.700	OUT046	Small	0.030704	43.2770
5675	FDW46	Tier 1	13.000	OUT049	Medium	0.070411	63.4484
5676	FDB58	Tier 1	10.500	OUT046	Small	0.013496	141.3154
5677	FDD47	Tier 3	7.600	OUT018	Medium	0.142991	169.1448

3099 rows × 7 columns

```
In [13]: data=df1[0:5]
data
```

	Item_Identifier	Outlet_Location_Type	Item_Weight	Outlet_Identifier	Outlet_Size	Item_Visibility	Item_MRP
0	FDW58	Tier 1	20.750	OUT049	Medium	0.007565	107.8622
5	FDH56	Tier 1	9.800	OUT046	Small	0.063817	117.1492
6	FDL48	Tier 3	19.350	OUT018	Medium	0.082602	50.1034
13	FDU11	Tier 1	4.785	OUT049	Medium	0.092738	122.3098
14	DRL59	Tier 3	16.750	OUT013	High	0.021206	52.0298

```
In [14]: sns.set_theme(style="whitegrid")
sns.barplot(x="Outlet_Location_Type",y="Item_Weight",data=data,hue="Outlet_Location_Type")
```

```
Out[14]: <AxesSubplot:xlabel='Outlet_Location_Type', ylabel='Item_Weight'>
```



```
In [15]: har=df.drop(["Item_Weight","Outlet_Location_Type","Item_Type","Outlet_Type"],axis=1)
har
```

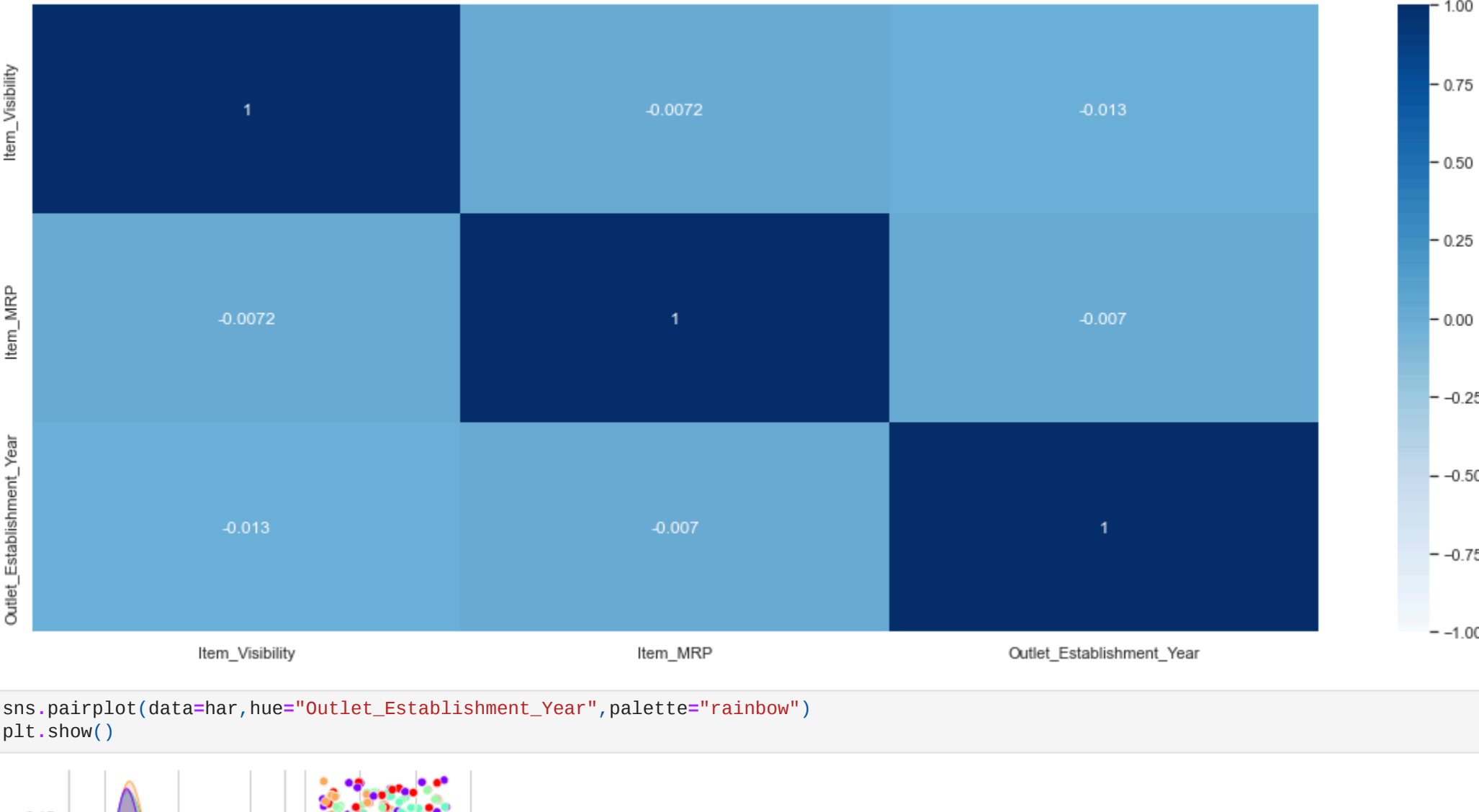
	Item_Identifier	Item_Fat_Content	Item_Visibility	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Size
0	FDW58	Low Fat	0.007565	107.8622	OUT049	1999	Medium
5	FDH56	Regular	0.063817	117.1492	OUT046	1997	Small
6	FDL48	Regular	0.082602	50.1034	OUT018	2009	Medium
13	FDU11	Low Fat	0.092738	122.3098	OUT049	1999	Medium
14	DRL59	LF	0.021206	52.0298	OUT013	1987	High
...
5673	FDF46	Low Fat	0.094053	116.0834	OUT018	2009	Medium
5674	DRL35	Low Fat	0.030704	43.2770	OUT046	1997	Small
5675	FDW46	Regular	0.070411	63.4484	OUT049	1999	Medium
5676	FDB58	Regular	0.013496	141.3154	OUT046	1997	Small
5677	FDD47	Regular	0.142991	169.1448	OUT018	2009	Medium

3099 rows × 7 columns

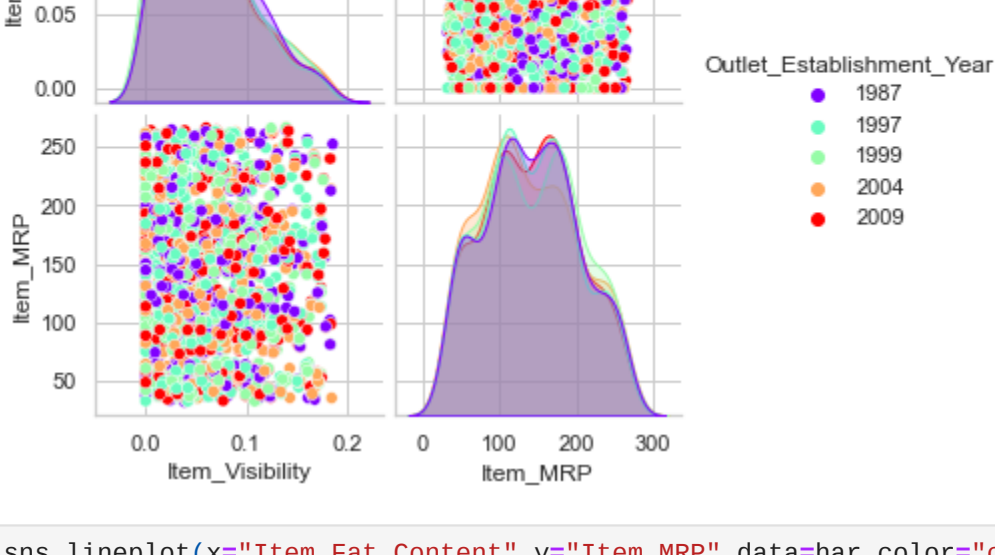
```
In [16]: har.head(6)
```

	Item_Identifier	Item_Fat_Content	Item_Visibility	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Size
0	FDW58	Low Fat	0.007565	107.8622	OUT049	1999	Medium
5	FDH56	Regular	0.063817	117.1492	OUT046	1997	Small
6	FDL48	Regular	0.082602	50.1034	OUT018	2009	Medium
13	FDU11	Low Fat	0.092738	122.3098	OUT049	1999	Medium
14	DRL59	LF	0.021206	52.0298	OUT013	1987	High
15	FDM24	Regular	0.079451	151.6366	OUT049	1999	Medium

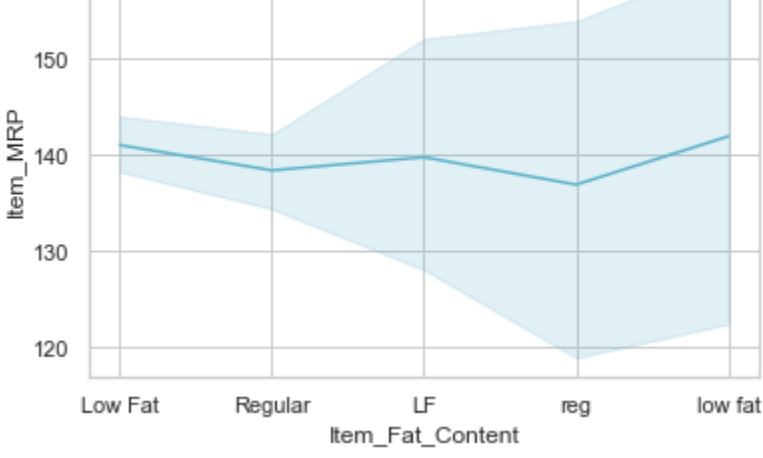
```
In [17]: plt.figure(figsize=(20,8))
sns.heatmap(data=har.corr(),annot=True,cmap="Blues",vmin=-1,vmax=1)
plt.title("Correlation",fontsize=20)
plt.show()
```



```
In [18]: sns.pairplot(data=har,hue="Outlet_Establishment_Year",palette="rainbow")
plt.show()
```



```
In [25]: sns.lineplot(x="Item_Fat_Content",y="Item_MRP",data=har,color="c")
plt.show()
```



```
In [ ]:
```