In [1]:

```
import numpy as np
import pandas as pd
import plotly.express as px
import plotly.graph_objects as go
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:

```
df=pd.read_csv("blinkit_data.csv")
df
```

Out[2]:

	Item Fat Content	Item Identifier	Item Type	Outlet Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Type	Item Visibility
0	Regular	FDX32	Fruits and Vegetables	2012	OUT049	Tier 1	Medium	Supermarket Type1	0.100014
1	Low Fat	NCB42	Health and Hygiene	2022	OUT018	Tier 3	Medium	Supermarket Type2	0.008596
2	Regular	FDR28	Frozen Foods	2010	OUT046	Tier 1	Small	Supermarket Type1	0.025896
3	Regular	FDL50	Canned	2000	OUT013	Tier 3	High	Supermarket Type1	0.042278
4	Low Fat	DRI25	Soft Drinks	2015	OUT045	Tier 2	Small	Supermarket Type1	0.033970
...	...	...	...	...	...	...	...	...	...
8518	low fat	NCT53	Health and Hygiene	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.000000
8519	low fat	FDN09	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.034706
8520	low fat	DRE13	Soft Drinks	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.027571
8521	reg	FDT50	Dairy	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.107715
8522	reg	FDM58	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.000000

8523 rows × 12 columns

In [3]:

```
df.head(10)
```

Out[3]:

	Item Fat Content	Item Identifier	Item Type	Outlet Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Type	Item Visibility	W
0	Regular	FDX32	Fruits and Vegetables	2012	OUT049	Tier 1	Medium	Supermarket Type1	0.100014	

	Item Fat Content	Item Identifier	Item Type	Outlet Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Type	Item Visibility	W
1	Low Fat	NCB42	Health and Hygiene	2022	OUT018	Tier 3	Medium	Supermarket Type2	0.008596	
2	Regular	FDR28	Frozen Foods	2010	OUT046	Tier 1	Small	Supermarket Type1	0.025896	
3	Regular	FDL50	Canned	2000	OUT013	Tier 3	High	Supermarket Type1	0.042278	
4	Low Fat	DRI25	Soft Drinks	2015	OUT045	Tier 2	Small	Supermarket Type1	0.033970	
5	low fat	FDS52	Frozen Foods	2020	OUT017	Tier 2	Small	Supermarket Type1	0.005505	
6	Low Fat	NCU05	Health and Hygiene	2011	OUT010	Tier 3	Small	Grocery Store	0.098312	
7	Low Fat	NCD30	Household	2015	OUT045	Tier 2	Small	Supermarket Type1	0.026904	
8	Low Fat	FDW20	Fruits and Vegetables	2000	OUT013	Tier 3	High	Supermarket Type1	0.024129	
9	Low Fat	FDX25	Canned	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.101562	

In [5]:

```
df.columns
```

Out[5]:

```
Index(['Item Fat Content', 'Item Identifier', 'Item Type',
       'Outlet Establishment Year', 'Outlet Identifier',
       'Outlet Location Type', 'Outlet Size', 'Outlet Type', 'Item Visibility',
       'Item Weight', 'Sales', 'Rating'],
      dtype='object')
```

In [4]:

```
df.dtypes
```

Out[4]:

```
Item Fat Content              object
Item Identifier               object
Item Type                     object
Outlet Establishment Year      int64
Outlet Identifier             object
Outlet Location Type          object
Outlet Size                   object
Outlet Type                   object
Item Visibility              float64
Item Weight                  float64
Sales                        float64
Rating                       float64
dtype: object
```

In [6]:

```
df.shape
```

Out[6]:

```
(8523, 12)
```

In [7]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8523 entries, 0 to 8522
Data columns (total 12 columns):
 #   Column                     Non-Null Count  Dtype  
---  ------                     --------------  -----  
 0   Item Fat Content           8523 non-null   object 
 1   Item Identifier            8523 non-null   object 
 2   Item Type                  8523 non-null   object 
 3   Outlet Establishment Year  8523 non-null   int64  
 4   Outlet Identifier          8523 non-null   object 
 5   Outlet Location Type       8523 non-null   object 
 6   Outlet Size                8523 non-null   object 
 7   Outlet Type                8523 non-null   object 
 8   Item Visibility            8523 non-null   float64
 9   Item Weight                7060 non-null   float64
 10  Sales                      8523 non-null   float64
 11  Rating                     8523 non-null   float64
dtypes: float64(4), int64(1), object(7)
memory usage: 799.2+ KB
```

In [8]:

```
df.isnull().sum()
```

Out[8]:

```
Item Fat Content                0
Item Identifier                 0
Item Type                       0
Outlet Establishment Year       0
Outlet Identifier               0
Outlet Location Type            0
Outlet Size                     0
Outlet Type                     0
Item Visibility                 0
Item Weight                  1463
Sales                           0
Rating                          0
dtype: int64
```

In [9]:

```
df[pd.isnull(df).any(axis=1)]
```

Out[9]:

	Item Fat Content	Item Identifier	Item Type	Outlet Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Type	Item Visibility
9	Low Fat	FDX25	Canned	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.101562
10	LF	FDX21	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.084555
14	Low Fat	FDH19	Meat	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.032928
18	Low Fat	FDJ56	Fruits and Vegetables	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.182515

	Item Fat Content	Item Identifier	Item Type	Outlet Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Type	Item Visibility
20	Regular	FDZ07	Fruits and Vegetables	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.000000
...	...	...	...	...	...	...	...	...	...
8518	low fat	NCT53	Health and Hygiene	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.000000
8519	low fat	FDN09	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.034706
8520	low fat	DRE13	Soft Drinks	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.027571
8521	reg	FDT50	Dairy	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.107715
8522	reg	FDM58	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.000000

1463 rows × 12 columns

In [10]:

```
df["Item Weight"]=df["Item Weight"].fillna(0)
df
```

Out[10]:

	Item Fat Content	Item Identifier	Item Type	Outlet Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Type	Item Visibility
0	Regular	FDX32	Fruits and Vegetables	2012	OUT049	Tier 1	Medium	Supermarket Type1	0.100014
1	Low Fat	NCB42	Health and Hygiene	2022	OUT018	Tier 3	Medium	Supermarket Type2	0.008596
2	Regular	FDR28	Frozen Foods	2010	OUT046	Tier 1	Small	Supermarket Type1	0.025896
3	Regular	FDL50	Canned	2000	OUT013	Tier 3	High	Supermarket Type1	0.042278
4	Low Fat	DRI25	Soft Drinks	2015	OUT045	Tier 2	Small	Supermarket Type1	0.033970
...	...	...	...	...	...	...	...	...	...
8518	low fat	NCT53	Health and Hygiene	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.000000
8519	low fat	FDN09	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.034706
8520	low fat	DRE13	Soft Drinks	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.027571
8521	reg	FDT50	Dairy	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.107715
8522	reg	FDM58	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.000000

8523 rows × 12 columns

In [11]:

```
df.isnull().sum()
```

Out[11]:

```
Item Fat Content             0
Item Identifier              0
Item Type                    0
Outlet Establishment Year    0
Outlet Identifier            0
Outlet Location Type         0
Outlet Size                  0
Outlet Type                  0
Item Visibility              0
Item Weight                  0
Sales                        0
Rating                       0
dtype: int64
```

In [12]:

```
df["Item Fat Content"].unique()
```

Out[12]:

```
array(['Regular', 'Low Fat', 'low fat', 'LF', 'reg'], dtype=object)
```

In [13]:

```
df["Item Fat Content"]=df["Item Fat Content"].replace({"low fat":"Low Fat","Replace":"Re
```

In [14]:

```
df["Item Fat Content"].unique()
```

Out[14]:

```
array(['Regular', 'Low Fat', 'LF', 'reg'], dtype=object)
```

In [15]:

```
df.describe()
```

Out[15]:

	Outlet Establishment Year	Item Visibility	Item Weight	Sales	Rating
count	8523.000000	8523.000000	8523.000000	8523.000000	8523.000000
mean	2010.831867	0.066132	10.650590	140.992782	3.965857
std	8.371760	0.051598	6.431899	62.275067	0.605651
min	1998.000000	0.000000	0.000000	31.290000	1.000000
25%	2000.000000	0.026989	6.650000	93.826500	4.000000
50%	2012.000000	0.053931	11.000000	143.012800	4.000000
75%	2017.000000	0.094585	16.000000	185.643700	4.200000
max	2022.000000	0.328391	21.350000	266.888400	5.000000

In [16]:

```
df["Sales"].sum()
```

Out[16]:

```
1201681.4808
```

In [17]:

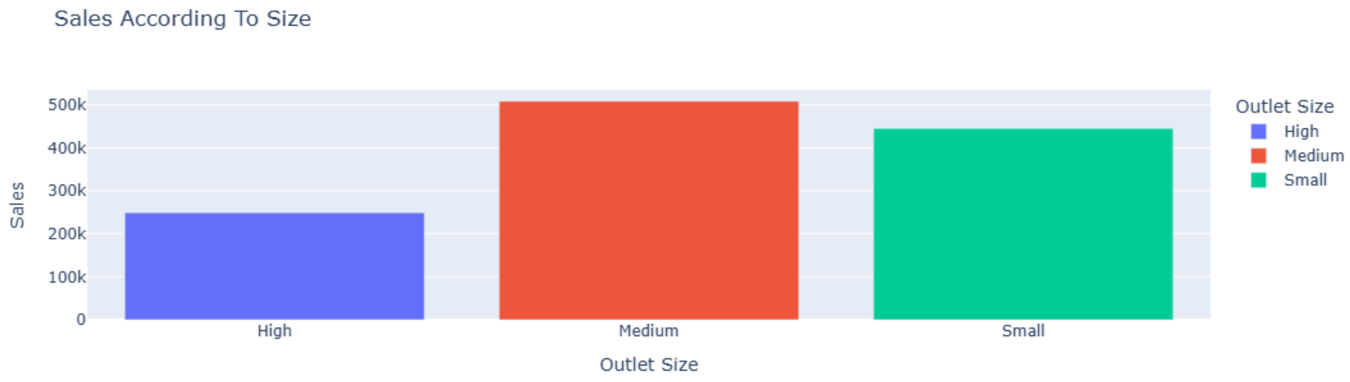
```
Display_According_To_Size_Sales=df.groupby("Outlet Size")["Sales"].sum().reset_index()
Display_According_To_Size_Sales
```

Out[17]:

	Outlet Size	Sales
0	High	248991.5840
1	Medium	507895.7284
2	Small	444794.1684

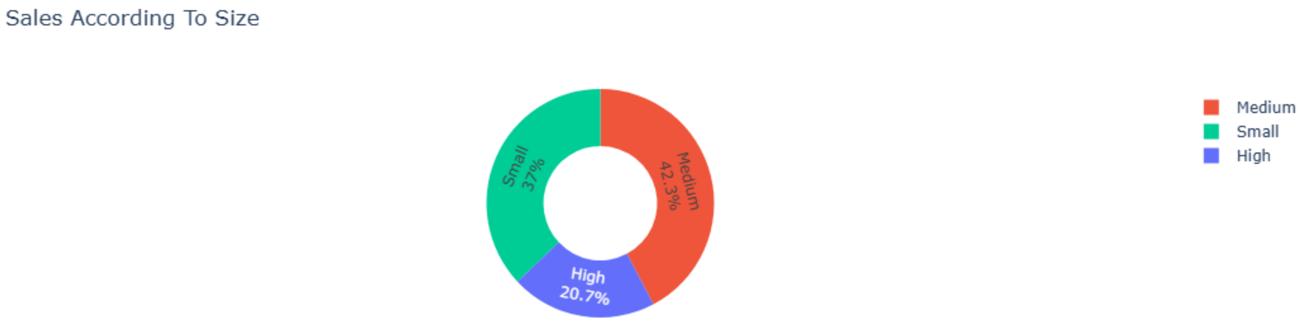
In [18]:

```
Bar_Display_According_To_Size_Sales=px.bar(Display_According_To_Size_Sales,x="Outlet Siz
Bar_Display_According_To_Size_Sales
```



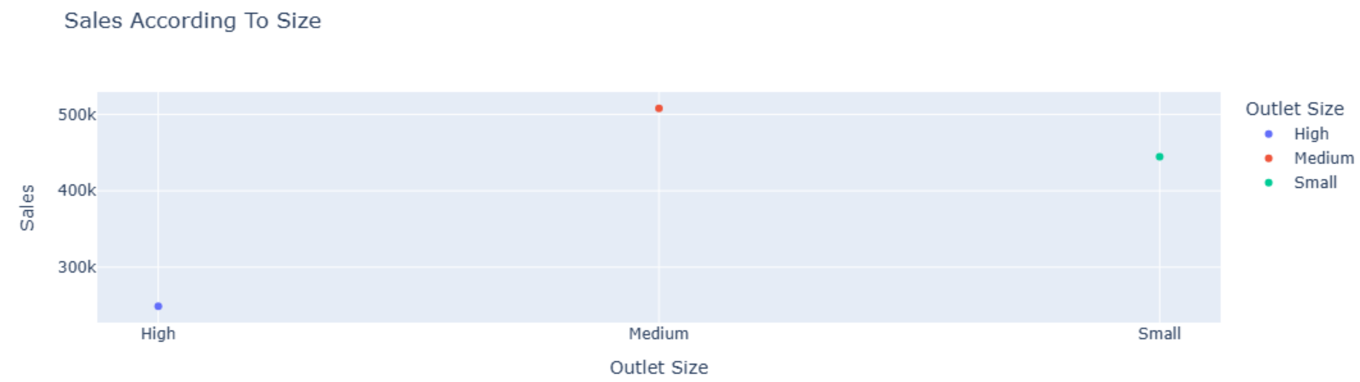
In [72]:

```
Bar_Display_According_To_Size_Sales=px.pie(Display_According_To_Size_Sales,values="Sales
Bar_Display_According_To_Size_Sales.update_traces(
    textposition="inside",
    textinfo="percent+label")
Bar_Display_According_To_Size_Sales
```



In [20]:

```
Bar_Display_According_To_Size_Sales=px.scatter(data_frame=Display_According_To_Size_Sale
Bar_Display_According_To_Size_Sales
```



In [21]:

```
df
```

Out[21]:

	Item Fat Content	Item Identifier	Item Type	Outlet Establishment Year	Outlet Identifier	Outlet Location Type	Outlet Size	Outlet Type	Item Visibility
0	Regular	FDX32	Fruits and Vegetables	2012	OUT049	Tier 1	Medium	Supermarket Type1	0.100014
1	Low Fat	NCB42	Health and Hygiene	2022	OUT018	Tier 3	Medium	Supermarket Type2	0.008596
2	Regular	FDR28	Frozen Foods	2010	OUT046	Tier 1	Small	Supermarket Type1	0.025896
3	Regular	FDL50	Canned	2000	OUT013	Tier 3	High	Supermarket Type1	0.042278
4	Low Fat	DRI25	Soft Drinks	2015	OUT045	Tier 2	Small	Supermarket Type1	0.033970
...	...	...	...	...	...	...	...	...	...
8518	Low Fat	NCT53	Health and Hygiene	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.000000
8519	Low Fat	FDN09	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.034706
8520	Low Fat	DRE13	Soft Drinks	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.027571
8521	reg	FDT50	Dairy	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.107715
8522	reg	FDM58	Snack Foods	1998	OUT027	Tier 3	Medium	Supermarket Type3	0.000000

8523 rows × 12 columns

In [63]:

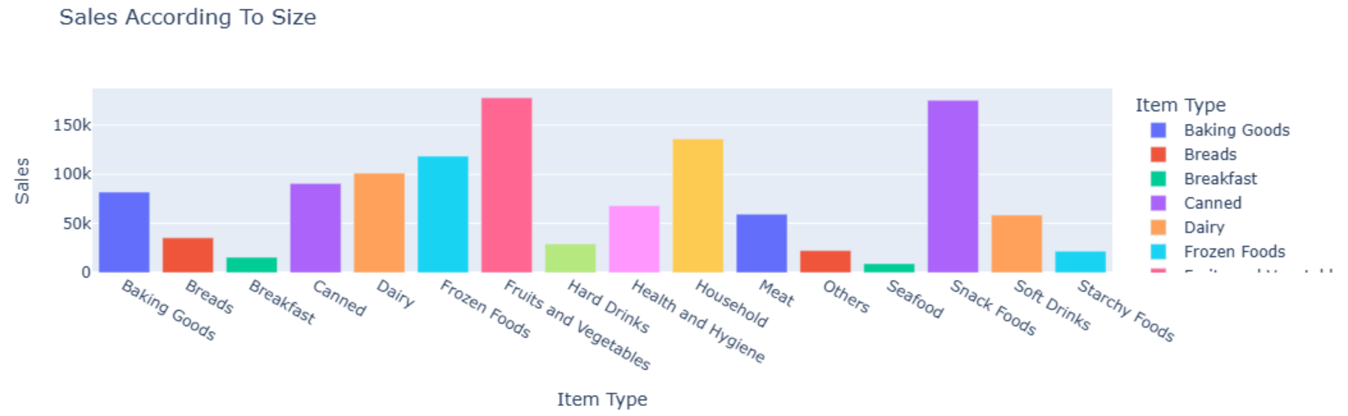
```
Display_According_To_Item_Type=df.groupby("Item Type")["Sales"].sum().reset_index()
Display_According_To_Item_Type
```

Out[63]:

	Item Type	Sales
0	Baking Goods	81894.7364
1	Breads	35379.1198
2	Breakfast	15596.6966
3	Canned	90706.7270
4	Dairy	101276.4596
5	Frozen Foods	118558.8814
6	Fruits and Vegetables	178124.0810
7	Hard Drinks	29334.6766
8	Health and Hygiene	68025.8388
9	Household	135976.5254
10	Meat	59449.8638
11	Others	22451.8916
12	Seafood	9077.8700
13	Snack Foods	175433.9204
14	Soft Drinks	58514.1650
15	Starchy Foods	21880.0274

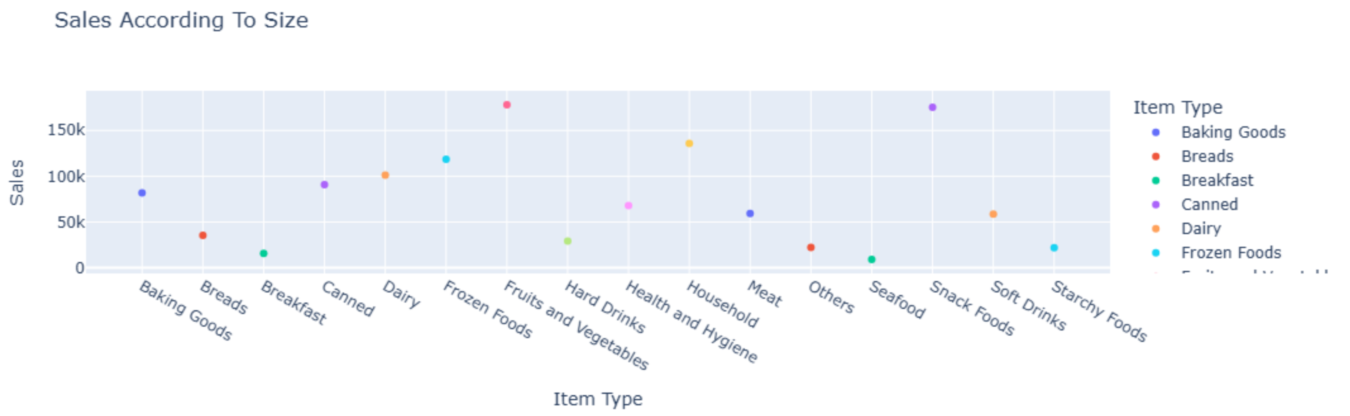
In [23]:

```
Bar_Display_According_To_Item_Type=px.bar(Display_According_To_Item_Type,x="Item Type",y
Bar_Display_According_To_Item_Type
```



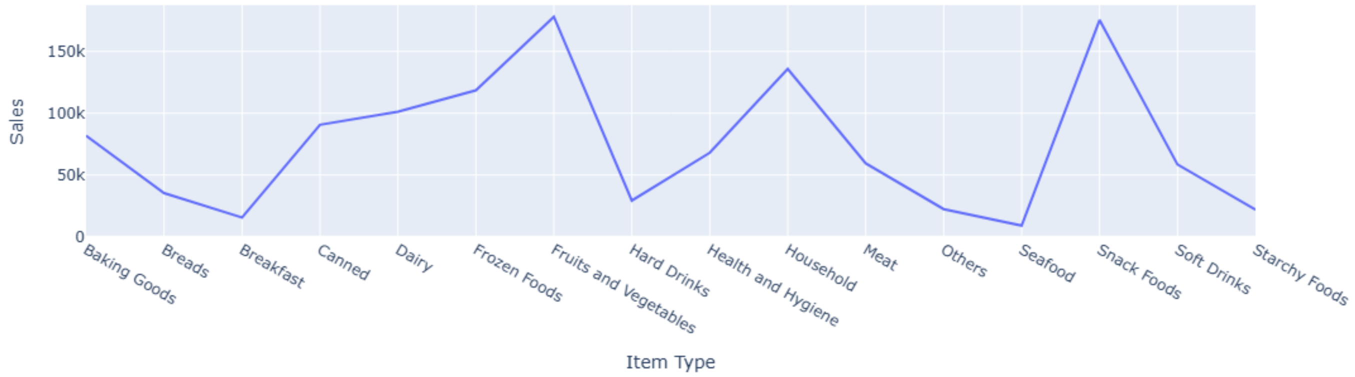
In [26]:

```
Scatter_plot_Display_According_To_Item_Type=px.scatter(data_frame=Display_According_To_I
Scatter_plot_Display_According_To_Item_Type
```



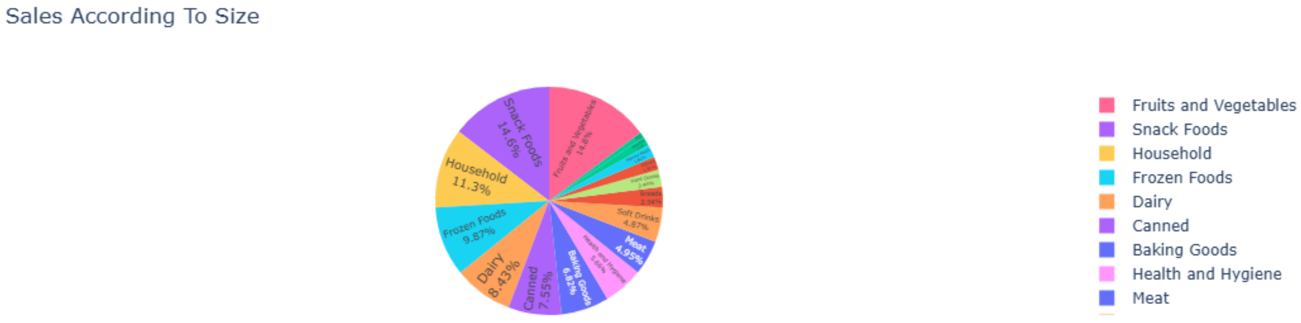
In [64]:

```
Line_plot_Display_According_To_Item_Type=px.line(Display_According_To_Item_Type,x="Item
Line_plot_Display_According_To_Item_Type
```



In [71]:

```
Scatter_plot_Display_According_To_Item_Type=px.pie(Display_According_To_Item_Type,values
Scatter_plot_Display_According_To_Item_Type.update_traces(
    textposition="inside",
    textinfo="percent+label")
Scatter_plot_Display_According_To_Item_Type
```



In [28]:

```
df.columns
```

Out[28]:

```
Index(['Item Fat Content', 'Item Identifier', 'Item Type',
       'Outlet Establishment Year', 'Outlet Identifier',
       'Outlet Location Type', 'Outlet Size', 'Outlet Type', 'Item Visibility',
       'Item Weight', 'Sales', 'Rating'],
      dtype='object')
```

In [53]:

```
Yearly_sale=df.groupby("Outlet Establishment Year")["Sales"].sum().reset_index()
Yearly_sale
```

Out[53]:

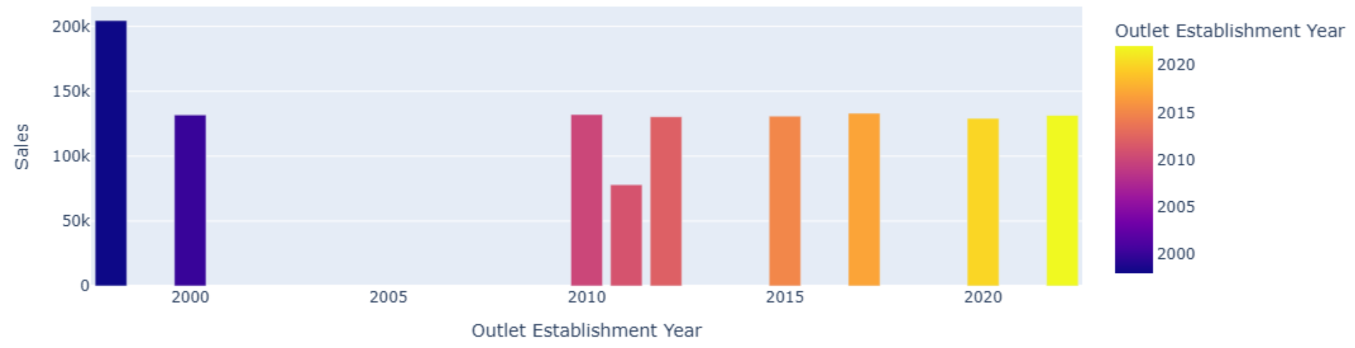
	Outlet Establishment Year	Sales
0	1998	204522.2570
1	2000	131809.0156
2	2010	132113.3698
3	2011	78131.5646
4	2012	130476.8598
5	2015	130942.7782
6	2017	133103.9070
7	2020	129103.9564
8	2022	131477.7724

In [56]:

```
Yearly_sales=px.bar(Yearly_sale,x="Outlet Establishment Year",y="Sales",color="Outlet Es
```

In [57]:

```
Yearly_sales
```

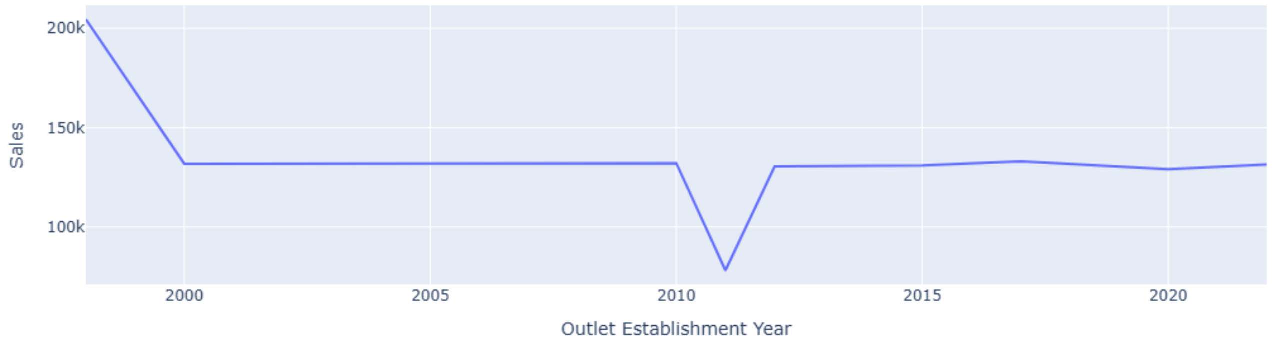


In [58]:

```
Outlet_Establishment_Year=px.line(Yearly_sale,x="Outlet Establishment Year",y="Sales")
```

In [59]:

```
Outlet_Establishment_Year
```



In [62]:

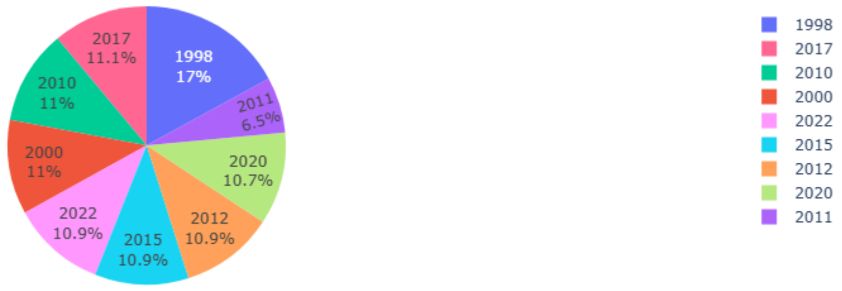
```
df["Outlet Establishment Year"].head(100)
```

Out[62]:

```
0     2012
1     2022
2     2010
3     2000
4     2015
      ... 
95    1998
96    1998
97    1998
98    1998
99    1998
Name: Outlet Establishment Year, Length: 100, dtype: int64
```

In [70]:

```
Outlet_Establishment_Year=px.pie(Yearly_sale,values="Sales",names="Outlet Establishment
Outlet_Establishment_Year.update_traces(
    textposition="inside",
    textinfo="percent+label")
Outlet_Establishment_Year
```



In [ ]: