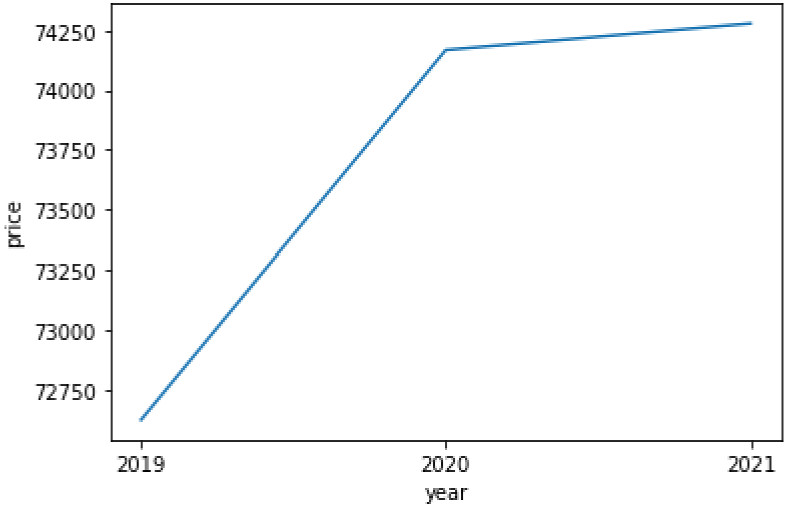
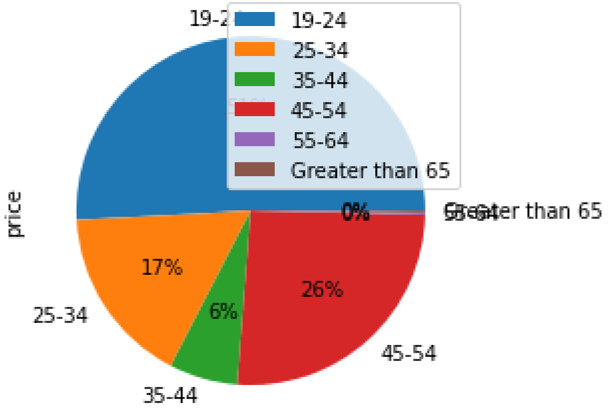
In [63]:

```
#importing  the required libraries (All 5 libraries are used)
import pandas as pd
import matplotlib.pyplot as plt
import plotly.graph_objects as go
import seaborn as sns
import datetime
datetime.datetime.strptime

#getting excel data in dataframe
df = pd.read_csv('dataset_2019_2022.csv')

#Change datetime format
df.transaction_date = pd.to_datetime(df.transaction_date)

 #removing/filtering year 2022 as it only have data till may 2022 ; We will be using this dataframe df throught the ex
ercise.
df = df[df.transaction_date < pd.Timestamp(2022,1,1)]


#remove duplicate customers - drop Keep is a parameter of method drop duplicate. Keep first record and drop everything
after/before
df.drop_duplicates(keep= 'first', inplace=True)

# creating line graph- Figure 0 : Total revenue in last 3 years
df["year"] = df.transaction_date.dt.year
df['year'] = df['year'].astype(str).str.split('.').str[0] # getting location 0 until character '.' in df column "year"

NewDf_sum_per_year = df.groupby('year')['price'].sum().reset_index()
sns.lineplot(x = "year", y = "price", data=NewDf_sum_per_year)

plt.show()
```



In [66]:

```
#creating pie chart - Figure 1 : Age Group percentage in SuperFoodsMax (below)
df.groupby(['age_band']).sum().plot(kind='pie', y='price',autopct='%1.0f%%',)
plt.show()
```



In [3]:

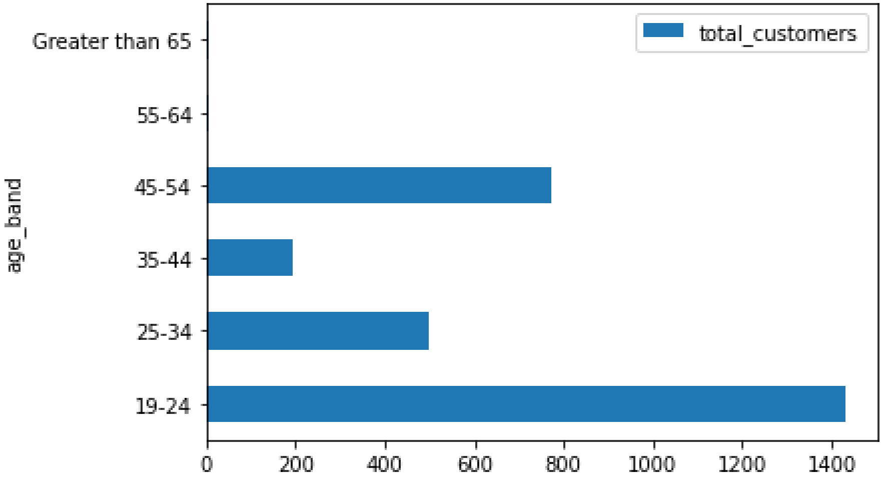
```
# creating table - Figure 2 : Age Band total revenue in SuperFoodsMax (below)
df.groupby('age_band').agg(Total_Revenue=('price',sum))
```

Out[3]:

	Total_Revenue
age_band	
19-24	112322.24
25-34	36713.56
35-44	14046.41
45-54	57213.87
55-64	396.85
Greater than 65	379.46

In [4]:

```
# creating horizontal bar graph- Figure 3 : Age Band total customers as per Age-Band (below)
df.groupby('age_band').agg(total_customers=('customer_id',pd.Series.nunique)).plot(kind='barh')
plt.show()
```



In [5]:

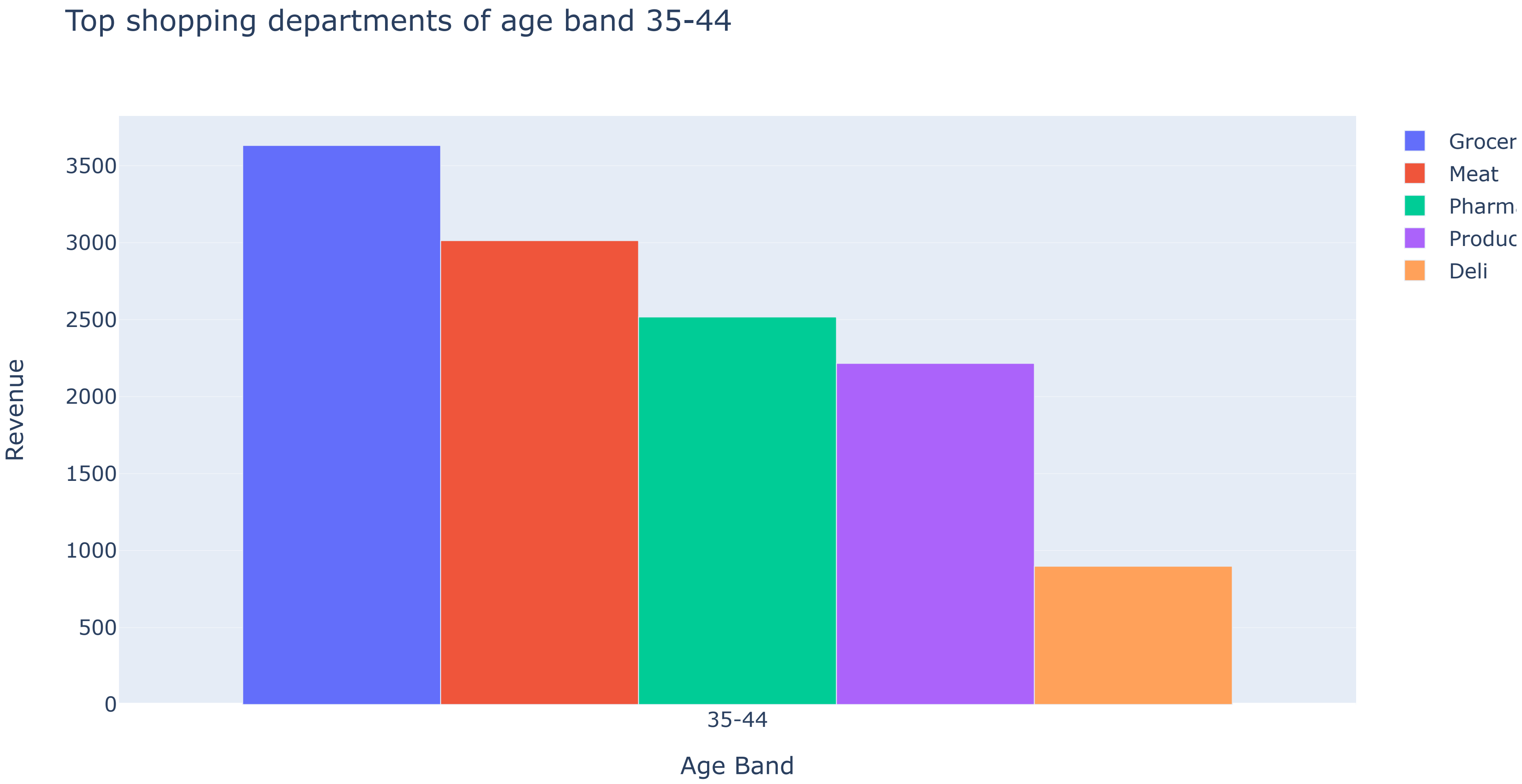
```
# creating Table- Figure 4 : Top 5 revenue generating departments for AgeBand 35-44 (below)
#Set up top five department types in descending order of revenue by age band type for 35-44
tmp = df.groupby(['age_band','department',]).agg(total_revenue=('price', sum)).reset_index()
tmp = tmp[tmp.age_band == '35-44']
top_5 = pd.concat(
    [tmp[tmp.age_band == hh] \
         .sort_values('total_revenue', ascending=False) \
     .head(5) for hh in tmp.age_band.unique()]).reset_index(drop=True)
top_5.head()
```

Out[5]:

	age_band	department	total_revenue
0	35-44	Grocery	3631.70
1	35-44	Meat	3013.54
2	35-44	Pharmaceutical	2517.39
3	35-44	Produce	2215.98
4	35-44	Deli	898.21

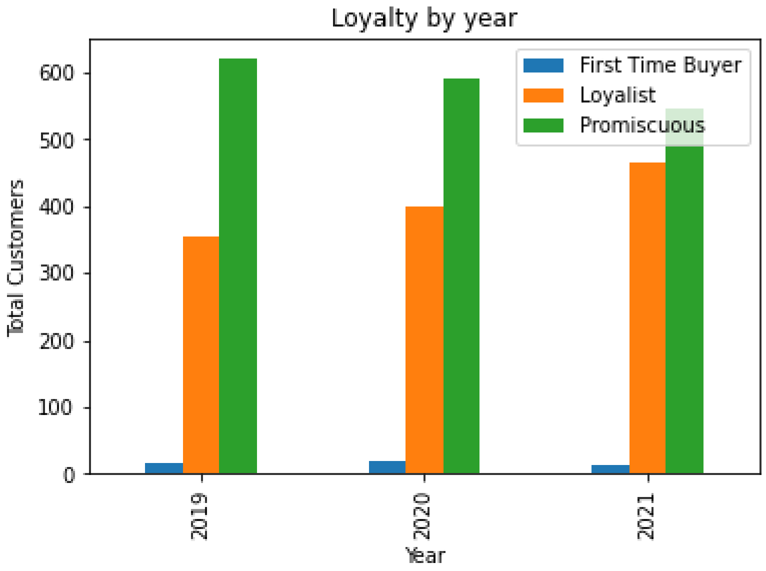
In [6]:

```
# creating Bar chart- Figure 5 : Top 5 revenue generating departments for AgeBand 35-44 (below)
data = [] #creating data array
for d in top_5.department.unique():
    tmp1 = top_5[top_5.department==d].groupby(['age_band']).agg(revenue=('total_revenue', sum)).reset_index()
    data.append(go.Bar(x=tmp1.age_band, y=tmp1.revenue, name = d))

go.Figure(
    data = data,
    layout = go.Layout(
        title ='Top shopping departments of age band 35-44',
        yaxis=dict(
            title='Revenue'
        ),
          xaxis=dict(
            title='Age Band'
        )
    )
).show(renderer = 'iframe')
```


Top shopping departments of age band 35-44
Revenue
3500
3000
2500
2000
1500
1000
500
0
35-44
Age Band
Grocer
Meat
Pharm
Produc
Deli

In [59]:

```
#Figure 6 : Bar graph of Loyalty segments in last 3 years
grouped_data = df.groupby(['year', 'loyalty']).agg(total_customers=('customer_id', pd.Series.nunique))

# Unstack the 'loyalty' level of the index to create separate bars for each loyalty category
unstacked_data = grouped_data.unstack()

# Change legend names
unstacked_data.columns = ['First Time Buyer', 'Loyalist', 'Promiscuous']

# Plot the unstacked bar plot
unstacked_data.plot(kind='bar', stacked=False, legend = True)
plt.title("Loyalty by year")
plt.xlabel("Year")
plt.ylabel("Total Customers")
plt.show()
```



```
In [67]: #printing unstacked_data dataframe created above
         display(unstacked_data)
```

	First Time Buyer	Loyalist	Promiscuous
year			
2019	17	356	620
2020	18	399	591
2021	12	464	545