

Field:Retail corporation

Review literature

This is a multi national company that is trying to establish a branch in a certain place with campaign and trying to capture the audience interest in products according to their behaviour

Data Source: Kaggle

Objective : To know the product interest and Customer behaviours

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
import datetime as dt
import matplotlib.pyplot as plt # To visualize
import plotly.express as px
```

```
In [2]: df = pd.read_csv('C:\Users\vandre\Desktop\data\customeranalysis.csv')
df
```

	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	DT_Customer	Recency	MntWines	...	NumWebVisitsMonth	AcceptedCmp3	AcceptedCmp4	AcceptedCmp5	AcceptedCmp1	AcceptedCmp2	Co
0	5524	1957	Graduation	Single	58138.0	0	0	04-09-2012	58	635	88	635	...	7	0	0	0	0
1	2174	1954	Graduation	Single	46344.0	1	1	08-03-2014	38	11	...	11	...	5	0	0	0	0
2	4141	1965	Graduation	Together	71613.0	0	0	21-08-2013	26	426	...	4	...	4	0	0	0	0
3	6182	1984	Graduation	Together	26646.0	1	0	10-02-2014	26	11	...	6	...	6	0	0	0	0
4	5324	1981	PhD	Married	58293.0	1	0	19-01-2014	94	173	...	5	...	5	0	0	0	0
...
2232	10870	1967	Graduation	Married	61223.0	0	1	13-06-2013	46	709	...	5	...	5	0	0	0	0
2233	4001	1946	PhD	Together	64014.0	2	1	10-06-2014	56	406	...	7	...	7	0	0	0	1
2234	7270	1981	Graduation	Divorced	56981.0	0	0	25-01-2014	91	908	...	6	...	6	0	1	0	0
2235	8235	1956	Master	Together	69245.0	0	1	24-01-2014	8	428	...	3	...	3	0	0	0	0
2236	9405	1954	PhD	Married	52869.0	1	1	15-10-2012	40	84	...	7	...	7	0	0	0	0

2237 rows x 29 columns

1.Checking for null values

```
In [3]: df.isnull().sum()
```

	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	DT_Customer	Recency	MntWines	...	NumWebVisitsMonth	AcceptedCmp3	AcceptedCmp4	AcceptedCmp5	AcceptedCmp1	AcceptedCmp2	Co
0	5524	1957	Graduation	Single	58138.0	0	0	04-09-2012	58	635	88	635	...	7	0	0	0	0
1	2174	1954	Graduation	Single	46344.0	1	1	08-03-2014	38	11	...	11	...	5	0	0	0	0
2	4141	1965	Graduation	Together	71613.0	0	0	21-08-2013	26	426	49	127	111	21	1	1	8	10
3	6182	1984	Graduation	Together	26646.0	1	0	10-02-2014	26	11	4	20	10	3	2	2	2	4
4	5324	1981	PhD	Married	58293.0	1	0	19-01-2014	94	173	43	118	46	27	5	5	5	6
...
2232	10870	1967	Graduation	Married	61223.0	0	1	13-06-2013	46	709	...	5	...	5	0	0	0	0
2233	4001	1946	PhD	Together	64014.0	2	1	10-06-2014	56	406	...	7	...	7	0	0	0	1
2234	7270	1981	Graduation	Divorced	56981.0	0	0	25-01-2014	91	908	...	6	...	6	0	1	0	0
2235	8235	1956	Master	Together	69245.0	0	1	24-01-2014	8	428	...	3	...	3	0	0	0	0
2236	9405	1954	PhD	Married	52869.0	1	1	15-10-2012	40	84	...	7	...	7	0	0	0	0

2237 rows x 29 columns

2.Dropping null values

```
In [4]: df.dropna(inplace=True)
```

Dropping unwanted columns

```
In [5]: df.drop(['AcceptedCmp1','AcceptedCmp2','AcceptedCmp3','AcceptedCmp4','AcceptedCmp5','Z_Revenue','Z_CostContact','Complain','NumCatalogPurchases','Marital_Status','MntGoldProds'],axis=1,inplace=True)
```

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

Checking the response rate on average

```
In [7]: df['response_rate'] = df['Response'] / df['NumWebVisitsMonth']
avg = round(df['response_rate'].mean(), 2)
fig = px.histogram(df, x='response_rate', histnorm='percent', title='Ratio of Responses per Visits')
fig.show()
```



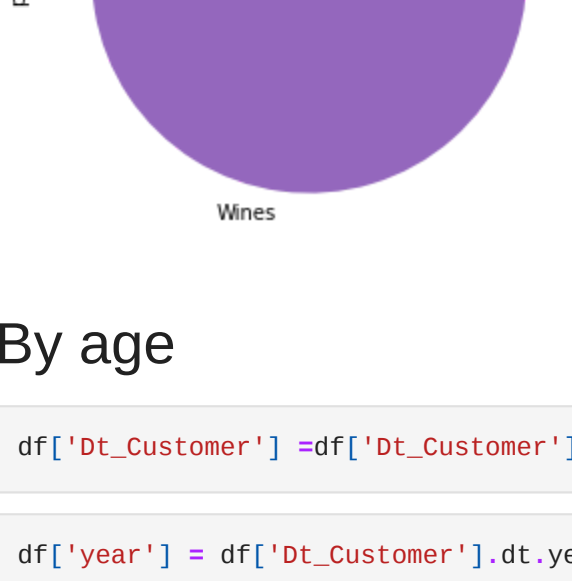
```
In [8]: print('The average rate:', avg*100, '%')
The average rate: 4.0 %
```

Know your customer

By Product

```
In [9]: M = df['MntMeatProducts'].sum()
F1 = df['MntFishProducts'].sum()
S = df['MntSweetProducts'].sum()
F = df['MntFruits'].sum()
W = df['MntWines'].sum()
data2 = pd.DataFrame({'points':[M,F1,S,F,W]})
```

```
In [10]: labels = ['Meat','Fish','sweets','Fruits','Wines']
plot = data2.plot.pie(y='points', labels=labels, figsize=(10, 5))
```



By age

```
In [11]: df['Dt_Customer'] = df['Dt_Customer'].astype('datetime64[ns]')
```

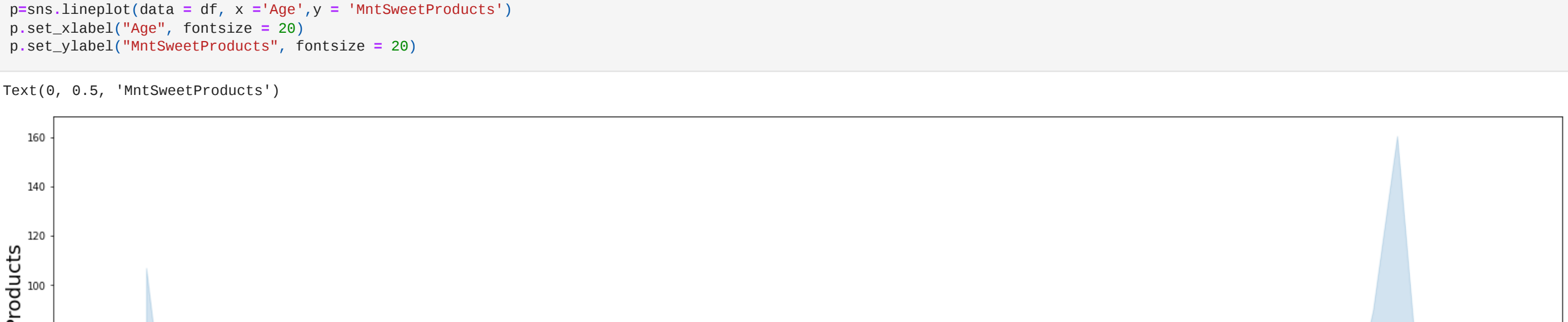
```
In [12]: df['year'] = df['Dt_Customer'].dt.year
```

```
In [13]: df['Age'] = df['year'] - df['Year_Birth']
```

```
In [14]: df
```

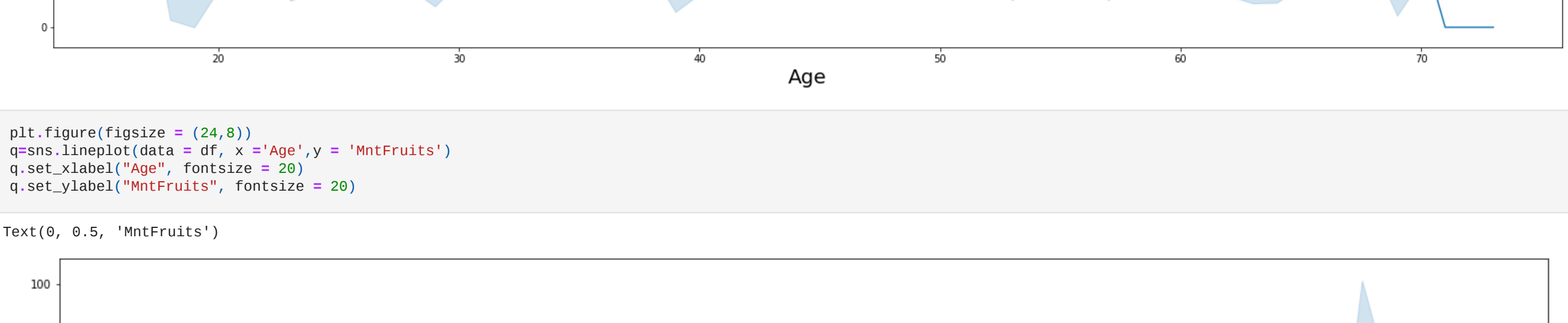
```
In [15]: plt.figure(figsize = (27,8))
sns.lineplot(data = df, x = 'Age', y = 'MntSweetProducts')
p.set_xlabel("Age", fontsize = 20)
p.set_ylabel("MntSweetProducts", fontsize = 20)
```

```
Out[15]: Text(0, 0.5, 'MntSweetProducts')
```



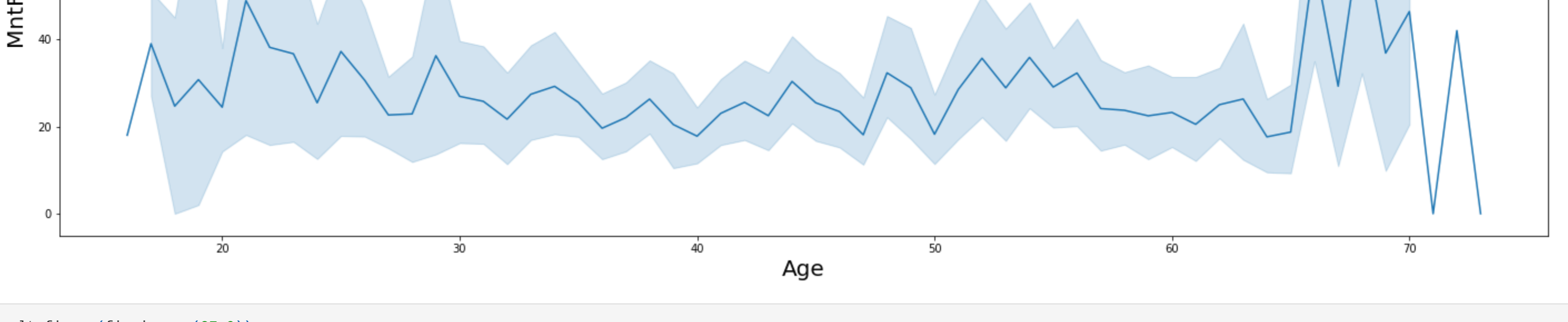
```
In [16]: plt.figure(figsize = (24,8))
sns.lineplot(data = df, x = 'Age', y = 'MntFruits')
q.set_xlabel("Age", fontsize = 20)
q.set_ylabel("MntFruits", fontsize = 20)
```

```
Out[16]: Text(0, 0.5, 'MntFruits')
```



```
In [17]: plt.figure(figsize = (27,9))
r = sns.lineplot(data = df, x = 'Age', y = 'MntMeatProducts')
r.set_xlabel("Age", fontsize = 20)
r.set_ylabel("MntMeatProducts", fontsize = 20)
```

```
Out[17]: Text(0, 0.5, 'MntMeatProducts')
```



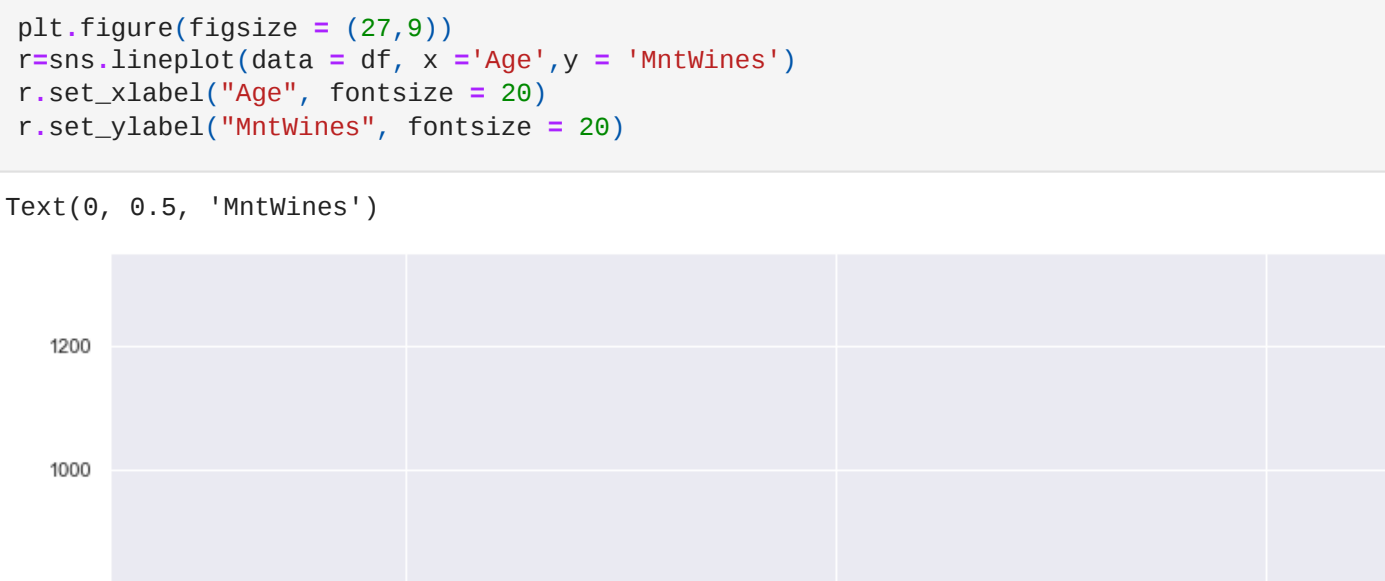
```
In [74]: plt.figure(figsize = (27,9))
r = sns.lineplot(data = df, x = 'Age', y = 'MntWines')
r.set_xlabel("Age", fontsize = 20)
r.set_ylabel("MntWines", fontsize = 20)
```

```
Out[74]: Text(0, 0.5, 'MntWines')
```



```
In [79]: sns.barplot(data = df, x = 'Kidhome', y = 'MntSweetProducts')
```

```
Out[79]: <AxesSubplot:xlabel='Kidhome', ylabel='MntSweetProducts'>
```



Relationship between products

```
In [73]: sns.scatterplot(data=df, x='MntMeatProducts', y='MntFishProducts')
```

```
Out[73]: <AxesSubplot:xlabel='MntMeatProducts', ylabel='MntFishProducts'>
```



```
In [80]: sea.scatterplot(data=df, x='MntMeatProducts', y='MntSweetProducts')
```

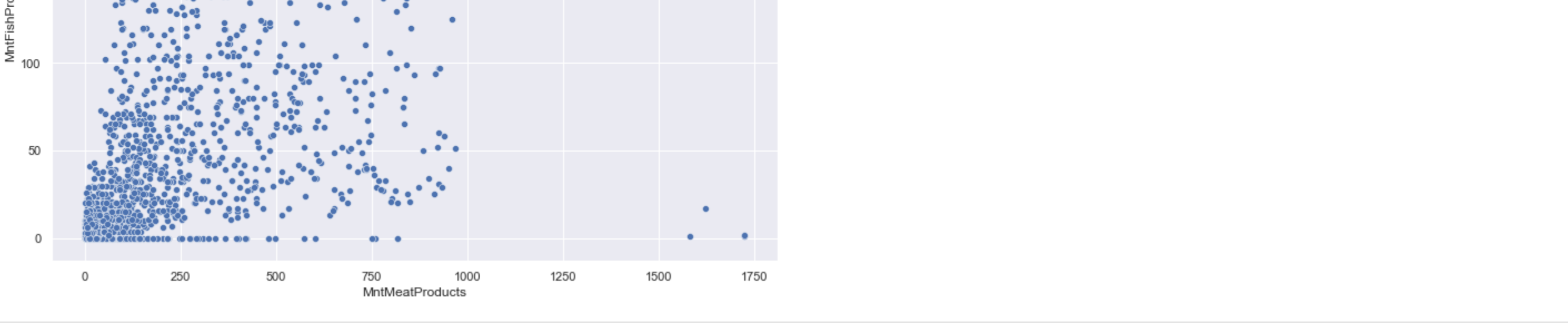
```
Out[80]: <AxesSubplot:xlabel='MntMeatProducts', ylabel='MntSweetProducts'>
```



Other factors relationships towards the products

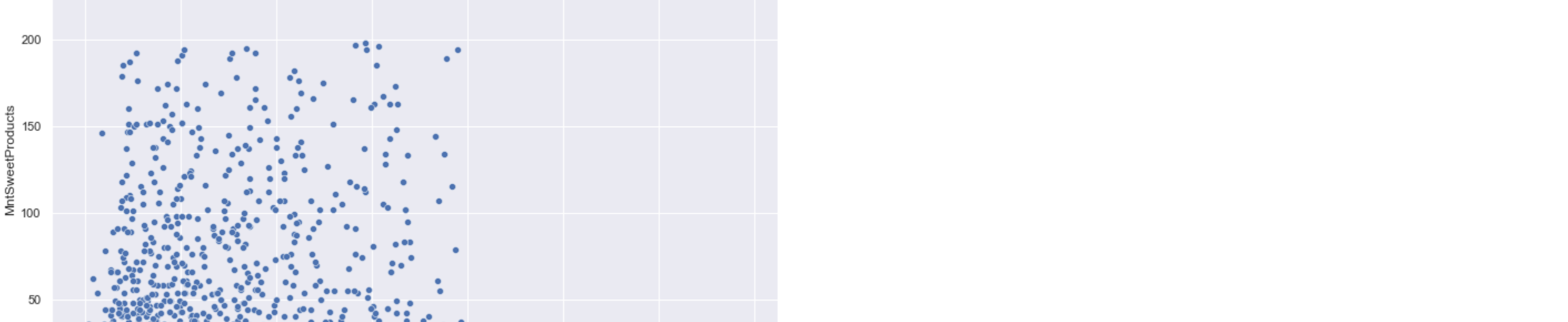
```
In [86]: plt.figure(figsize = (27,9))
s = sea.jointplot(data=df, x='Education', y='NumWebPurchases')
s.set_xlabel("Education", fontsize = 20)
s.set_ylabel("NumWebPurchases", fontsize = 20)
```

```
Out[86]: Text(0, 0.5, 'NumWebPurchases')
```



```
In [87]: plt.figure(figsize = (27,9))
t = sea.jointplot(data=df, x='Education', y='NumStorePurchases')
t.set_xlabel("Education", fontsize = 20)
t.set_ylabel("NumStorePurchases", fontsize = 20)
```

```
Out[87]: Text(0, 0.5, 'NumStorePurchases')
```



```
In [ ]:
```