

## ▼ Wholesale Dataset

ไฟล์ "Wholesale customers data.csv" (download จาก <https://archive.ics.uci.edu/ml/machine-learning-databases/wholesale/customers/wholesale.csv>)  
การขายสินค้าประเภทต่างๆ ในแต่ละภูมิภาค ผ่าน 2 ช่องทาง คือ Horeca (Hotel/Restaurant/Cafe) หรือ Retail โดยแถวแรกเป็นชื่อคอลัมน์ และ แถวที่ 2 เป็นต้นไปเก็บข้อมูล

ประเภทของสินค้า คือ fresh products, milk products, grocery products, frozen products, detergents  
คอลัมน์ CHANNEL เป็นช่องทางการขาย ที่มีค่าเป็น Horeca (Hotel/Restaurant/Cafe) หรือ Retail (1,2)  
Oporto หรือ Other (1,2,3) คอลัมน์ FRESH, MILK, GROCERY, FROZEN, DETERGENTS\_PAPER, DELICATES  
ในภูมิภาคนั้น ผ่านช่องทางนั้นในหนึ่งปี

- 1) เขียนโปรแกรมเพื่อแปลงข้อมูลช่องทางการขายและภูมิภาค จากตัวเลขให้เป็นชื่อของแต่ละช่องทางใหม่ที่มีชื่อว่า Rename\_Wholesale.csv โดยจะใช้ไฟล์นี้เป็นหลักในข้อถัดไป
- 2) เขียนโปรแกรมเพื่อหายอดขายรวมจากทุกช่องทาง ในทุกภูมิภาค สำหรับสินค้าแต่ละประเภทแล้ววาง
- 3) เขียนโปรแกรมเพื่อหายอดขายรวมของสินค้าทุกประเภท ในทุกช่องทาง สำหรับแต่ละภูมิภาคแล้ววาง
- 4) เขียนโปรแกรมเพื่อหายอดขายรวมของสินค้าสำหรับทำความสะอาดในแต่ละช่องทางการขาย โดยแต่ละภูมิภาค
- 5) เขียนโปรแกรมเพื่อหาจำนวน record ที่ยอดขายของอาหารสดรวมกับอาหารแช่แข็งผ่านช่องทาง Ho
- 6) เขียนโปรแกรมเพื่อหาภูมิภาคที่มีการบริโภคนมมากที่สุด และในภูมิภาคนั้นมีการบริโภคนมมากกว่า

```
1 from google.colab import drive
2 drive.mount('/content/drive')
```

➡ Go to this URL in a browser: [https://accounts.google.com/o/oauth2/auth?client\\_id=...](https://accounts.google.com/o/oauth2/auth?client_id=...)  
Enter your authorization code:  
.....  
Mounted at /content/drive

- 1) เขียนโปรแกรมเพื่อแปลงข้อมูลช่องทางการขายและภูมิภาค จากตัวเลข
- ▼ ทางการขายและภูมิภาคนั้น ๆ จากนั้นบันทึกเป็นไฟล์ใหม่ที่มีชื่อว่า Rena  
ไฟล์นี้เป็นหลักในข้อถัดไป

```
1 #Solution
2 import pandas as pd
3 import numpy as np
4 import matplotlib.pyplot as plt
```

```

1 df = pd.read_csv("Wholesale customers data.csv")
2 df['Channel'].replace(1, 'Horeca', inplace=True)
3 df['Channel'].replace(2, 'Retail' , inplace=True)
4 df['Region'].replace(1, 'Lisnon' , inplace=True)
5 df['Region'].replace(2, 'Oporto' , inplace=True)
6 df['Region'].replace(3, 'Other' , inplace=True)
7 df.to_csv('Rename_Wholesale.csv')
8 df = pd.read_csv('Rename_Wholesale.csv')

```

## 2) เขียนโปรแกรมเพื่อหายอดขายรวมจากทุกช่องทาง ในทุกภูมิภาค สำหรับ pie chart

```

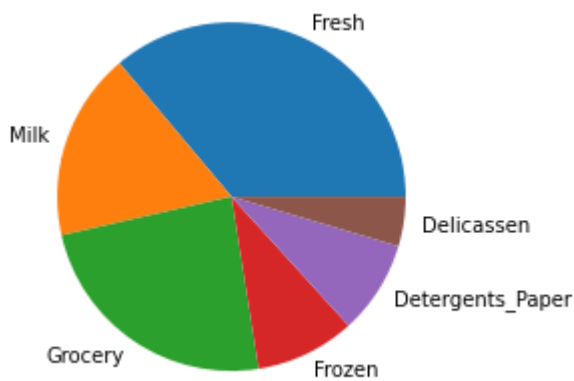
1 #Solution
2 data2 = [['Fresh', df['Fresh'].sum()], ['Milk', df['Milk'].sum()], ['Grocery', df[
3 df2 = pd.DataFrame(data2, columns = ['Goods', 'Total_sale'])
4 plt.pie(df2['Total_sale'], labels=df2['Goods'])

```

```

[<matplotlib.patches.Wedge at 0x7f83cad8b438>,
 <matplotlib.patches.Wedge at 0x7f83cad8b978>,
 <matplotlib.patches.Wedge at 0x7f83cad8be80>,
 <matplotlib.patches.Wedge at 0x7f83cada23c8>,
 <matplotlib.patches.Wedge at 0x7f83cada28d0>,
 <matplotlib.patches.Wedge at 0x7f83cada2dd8>],
 [Text(0.4646943537423714, 0.9970251539454559, 'Fresh'),
 Text(-1.0426809169323177, 0.3504518589837719, 'Milk'),
 Text(-0.6167281602164499, -0.9108492610723428, 'Grocery'),
 Text(0.47556761347886634, -0.9918847942226032, 'Frozen'),
 Text(0.931507510707914, -0.5850587641380527, 'Detergents_Paper'),
 Text(1.0885865665919394, -0.1580483692910923, 'Delicassen')]]

```



## 3) เขียนโปรแกรมเพื่อหายอดขายรวมของสินค้าทุกประเภท ในทุกช่องทาง สำหรับ pie chart

```

1 #Solution
2 data3 = []
3 data3 = df.loc[df['Region'] == 'Other'].sum()
4 Other_total_sale = data3[3]+data3[4]+data3[5]+data3[6]+data3[7]+data3[8]

```

```
5 Other_total_sale
```

```
↳ 10677599
```

```
1 data3 = []
2 data3 = df.loc[df['Region'] == 'Oporto'].sum()
3 data3
4 Oporto_total_sale = data3[3]+data3[4]+data3[5]+data3[6]+data3[7]+data3[8]
5 Oporto_total_sale
```

```
↳ 1555088
```

```
1 data3 = []
2 data3 = df.loc[df['Region'] == 'Lisbon'].sum()
3 data3
4 Lisbon_total_sale = data3[3]+data3[4]+data3[5]+data3[6]+data3[7]+data3[8]
5 Lisbon_total_sale
```

```
↳ 2386813
```

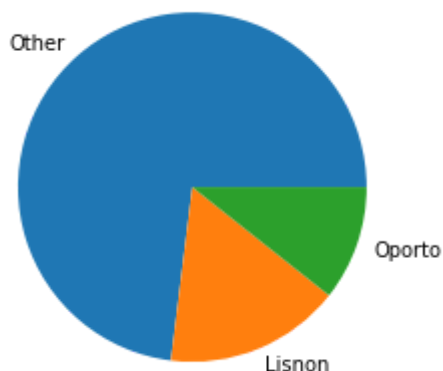
```
1 data3 = []
2 data3 = [['Other',Other_total_sale], ['Lisbon',Lisbon_total_sale], ['Oporto', 0
3 df3 = pd.DataFrame(data3, columns = ['Goods', 'Total_sale'])
4 df3
```

```
↳
```

	Goods	Total_sale
0	Other	10677599
1	Lisbon	2386813
2	Oporto	1555088

```
1 plt.pie(df3['Total_sale'],labels=df3['Goods'])
```

```
↳ ([<matplotlib.patches.Wedge at 0x7f83ca036c88>,
<matplotlib.patches.Wedge at 0x7f83ca044208>,
<matplotlib.patches.Wedge at 0x7f83ca044710>],
[Text(-0.7283936086362807, 0.824283173974707, 'Other'),
Text(0.41774442250412963, -1.0175900930466508, 'Lisbon'),
Text(1.0391497641077394, -0.36078770455051506, 'Oporto')])
```



#### 4) เขียนโปรแกรมเพื่อหายอดขายรวมของสินค้าสำหรับทำความสะอาด แบ่งเป็นของแต่ละภูมิภาค พร้อมพล็อตกราฟของแต่ละภูมิภาค

```
1 #Solution
2 data4 = []
3 data4 = df.loc[df['Region'] == 'Oporto'].sum()
4 Oporto_Detergents_Paper = data4[7]
5 Oporto_Detergents_Paper
```

173311

```
1 data4 = []
2 data4 = df.loc[df['Region'] == 'Lisbon'].sum()
3 Lisbon_Detergents_Paper = data4[7]
4 Lisbon_Detergents_Paper
```

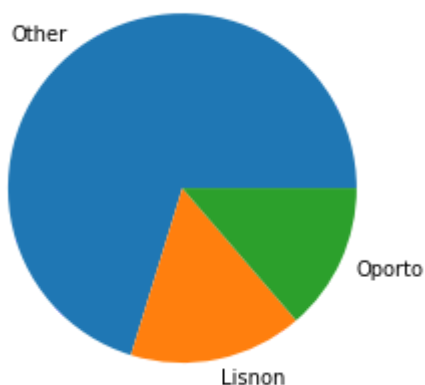
204136

```
1 data4 = []
2 data4 = df.loc[df['Region'] == 'Other'].sum()
3 Other_Detergents_Paper = data4[7]
4 Other_Detergents_Paper
```

890410

```
1 data4 = []
2 data4 = [['Other', Other_Detergents_Paper], ['Lisbon', Lisbon_Detergents_Paper],
3 df4 = pd.DataFrame(data4, columns = ['Goods', 'Total_sale'])
4 plt.pie(df4['Total_sale'], labels=df4['Goods'])
```

```
[<matplotlib.patches.Wedge at 0x7f83ca00b7b8>,
<matplotlib.patches.Wedge at 0x7f83ca00bcf8>,
<matplotlib.patches.Wedge at 0x7f83ca018240>],
[Text(-0.6529640341712075, 0.8852332856817249, 'Other'),
Text(0.22509365532528425, -1.0767231985669772, 'Lisbon'),
Text(1.0001175451539583, -0.4580009780286722, 'Oporto')]]
```



## 5) เขียนโปรแกรมเพื่อหาจำนวนปีที่ยอดขายของอาหารสดรวมกับอาหาร Horeca(Hotel/Restaurant/Cafe) ไม่เกิน 1,000 ชิ้น

```
1 #Solution
2 df5 = pd.read_csv('Rename_Wholesale.csv')
3 df5[(df5['Channel']=='Horeca') & (df5['Fresh']+df5['Frozen']<1000)].count()
```

```
↳ Unnamed: 0      10
   Channel         10
   Region         10
   Fresh          10
   Milk           10
   Grocery        10
   Frozen         10
   Detergents_Paper 10
   Delicassen     10
   dtype: int64
```

## 6) เขียนโปรแกรมเพื่อหาภูมิภาคที่มีการบริโภคนมมากที่สุด และในภูมิภาคอาหารสดทั้งหมดกี่ปี

```
1 #Solution
2 data6 = []
3 data6 = df.loc[df['Region'] == 'Oporto'].sum()
4 Oporto_Milk = data6[4]
5 Oporto_Milk
6
```

```
↳ 239144
```

```
1 data6 = []
2 data6 = df.loc[df['Region'] == 'Lisbon'].sum()
3 Lisbon_Milk = data6[4]
4 Lisbon_Milk
```

```
↳ 422454
```

```
1 data6 = []
2 data6 = df.loc[df['Region'] == 'Other'].sum()
3 Other_Milk = data6[4]
4 Other_Milk
```

```
↳ 1888759
```

```
1 if Oporto_Milk>Lisbon_Milk:
2     if Oporto_Milk>Other_Milk:
3         print('The most is Oporto')
4         else: print('The most is Other')
```

```
4     else : print('The most is Other')
5 if Lisnon_Milk>Other_Milk:
6     print('The most is Lisnon')
7 else : print('The most is Other')
```

➤ The most is Other

```
1 df6 = pd.read_csv('Rename_Wholesale.csv')
2 df6[(df6['Region']=='Other')&(df6['Milk'] > df6['Fresh'])].count()
```

➤

Unnamed: 0	107
Channel	107
Region	107
Fresh	107
Milk	107
Grocery	107
Frozen	107
Detergents_Paper	107
Delicassen	107
dtype: int64	