

```
In [1]: pip install ucimlrepo
```

```
Requirement already satisfied: ucimlrepo in c:\users\dell\anaconda3\lib\site-packages (0.0.7)
Requirement already satisfied: pandas>=1.0.0 in c:\users\dell\anaconda3\lib\site-packages (from ucimlrepo) (1.5.3)
Requirement already satisfied: certifi>=2020.12.5 in c:\users\dell\anaconda3\lib\site-packages (from ucimlrepo) (2022.12.7)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\dell\anaconda3\lib\site-packages (from pandas>=1.0.0->ucimlrepo) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\dell\anaconda3\lib\site-packages (from pandas>=1.0.0->ucimlrepo) (2022.7)
Requirement already satisfied: numpy>=1.21.0 in c:\users\dell\anaconda3\lib\site-packages (from pandas>=1.0.0->ucimlrepo) (1.23.5)
Requirement already satisfied: six>=1.5 in c:\users\dell\anaconda3\lib\site-packages (from python-dateutil>=2.8.1->pandas>=1.0.0->ucimlrepo) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

```
In [2]: from ucimlrepo import fetch_ucirepo
```

```
# fetch dataset
wholesale_customers = fetch_ucirepo(id=292)

# data (as pandas dataframes)
X = wholesale_customers.data.features
y = wholesale_customers.data.targets

# metadata
print(wholesale_customers.metadata)

# variable information
print(wholesale_customers.variables)
```

```
{'uci_id': 292, 'name': 'Wholesale customers', 'repository_url': 'https://archive.ics.uci.edu/dataset/292/wholesale+customers', 'data_url': 'https://archive.ics.uci.edu/static/public/292/data.csv', 'abstract': 'The data set refers to clients of a wholesale distributor. It includes the annual spending in monetary units (m.u.) on diverse product categories', 'area': 'Business', 'tasks': ['Classification', 'Clustering'], 'characteristics': ['Multivariate'], 'num_instances': 440, 'num_features': 7, 'feature_types': ['Integer'], 'demographics': [], 'target_col': ['Region'], 'index_col': None, 'has_missing_values': 'no', 'missing_values_symbol': None, 'year_of_dataset_creation': 2013, 'last_updated': 'Mon Feb 05 2024', 'dataset_doi': '10.24432/C5030X', 'creators': ['Margarida Cardoso'], 'intro_paper': None, 'additional_info': {'summary': None, 'purpose': None, 'funded_by': None, 'instances_represent': None, 'recommended_data_splits': None, 'sensitive_data': None, 'preprocessing_description': None, 'variable_info': '1)\tFRESH: annual spending (m.u.) on fresh products (Continuous);\r\n2)\tMILK: annual spending (m.u.) on milk products (Continuous);\r\n3)\tGROCERY: annual spending (m.u.) on grocery products (Continuous);\r\n4)\tFROZEN: annual spending (m.u.) on frozen products (Continuous)\r\n5)\tDETERGENTS_PAPER: annual spending (m.u.) on detergents and paper products (Continuous) \r\n6)\tDELICATESSEN: annual spending (m.u.) on a nd delicatessen products (Continuous); \r\n7)\tCHANNEL: customersâ€™ Channel - Horeca (Hotel/Restaurant/CafÃ©) or Retail channel (Nominal)\r\n8)\tREGION: customersâ€™ Region â€“ Lisbon, Oporto or Other (Nominal)\r\n\r\nDescriptive Statistics:\r\n\r\n\t(Minimum, Maximum, Mean, Std. Deviation)\r\n\r\nFRESH (\t3, 112151, 12000.30, 12647.329)\r\n\r\nMILK \t(55, 73498, 5796.27, 7380.377)\r\n\r\nGROCERY\t(3, 92780, 7951.28, 9503.163)\r\n\r\nFROZEN \t(25, 60869, 3071.93, 4854.673)\r\n\r\nDETERGENTS_PAPER (3, 40827, 2881.49, 4767.854)\r\n\r\nDELICATESSEN (3, 47943, 1524.87, 2820.106)\r\n\r\n\r\nREGION\tFrequency\r\n\r\nLisbon\t77\r\n\r\nOporto\t47\r\n\r\nOther Region\t316\r\n\r\nTotal\t440\r\n\r\n\r\nCHANNEL\tFrequency\r\n\r\nHoreca\t298\r\n\r\nRetail\t142\r\n\r\nTotal\t440\r\n\r\n', 'citation': None}}
```

	name	role	type	demographic	description	units	\
0	Channel	Feature	Categorical	None	None	None	
1	Region	Target	Categorical	None	None	None	
2	Fresh	Feature	Integer	None	None	None	
3	Milk	Feature	Integer	None	None	None	
4	Grocery	Feature	Integer	None	None	None	
5	Frozen	Feature	Integer	None	None	None	
6	Detergents_Paper	Feature	Integer	None	None	None	
7	Delicassen	Feature	Integer	None	None	None	

	missing_values
0	no
1	no
2	no
3	no
4	no
5	no
6	no
7	no

```
In [3]: print(X)
```

	Channel	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicassen
0	2	12669	9656	7561	214	2674	1338
1	2	7057	9810	9568	1762	3293	1776
2	2	6353	8808	7684	2405	3516	7844
3	1	13265	1196	4221	6404	507	1788
4	2	22615	5410	7198	3915	1777	5185
..
435	1	29703	12051	16027	13135	182	2204
436	1	39228	1431	764	4510	93	2346
437	2	14531	15488	30243	437	14841	1867
438	1	10290	1981	2232	1038	168	2125
439	1	2787	1698	2510	65	477	52

[440 rows x 7 columns]

```
In [4]: print(y)
```

	Region
0	3
1	3
2	3
3	3
4	3
..	...
435	3
436	3
437	3
438	3
439	3

[440 rows x 1 columns]

```
In [5]: X.head()
```

	Channel	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicassen
0	2	12669	9656	7561	214	2674	1338
1	2	7057	9810	9568	1762	3293	1776
2	2	6353	8808	7684	2405	3516	7844
3	1	13265	1196	4221	6404	507	1788
4	2	22615	5410	7198	3915	1777	5185

```
In [6]: X.isnull().sum()
```

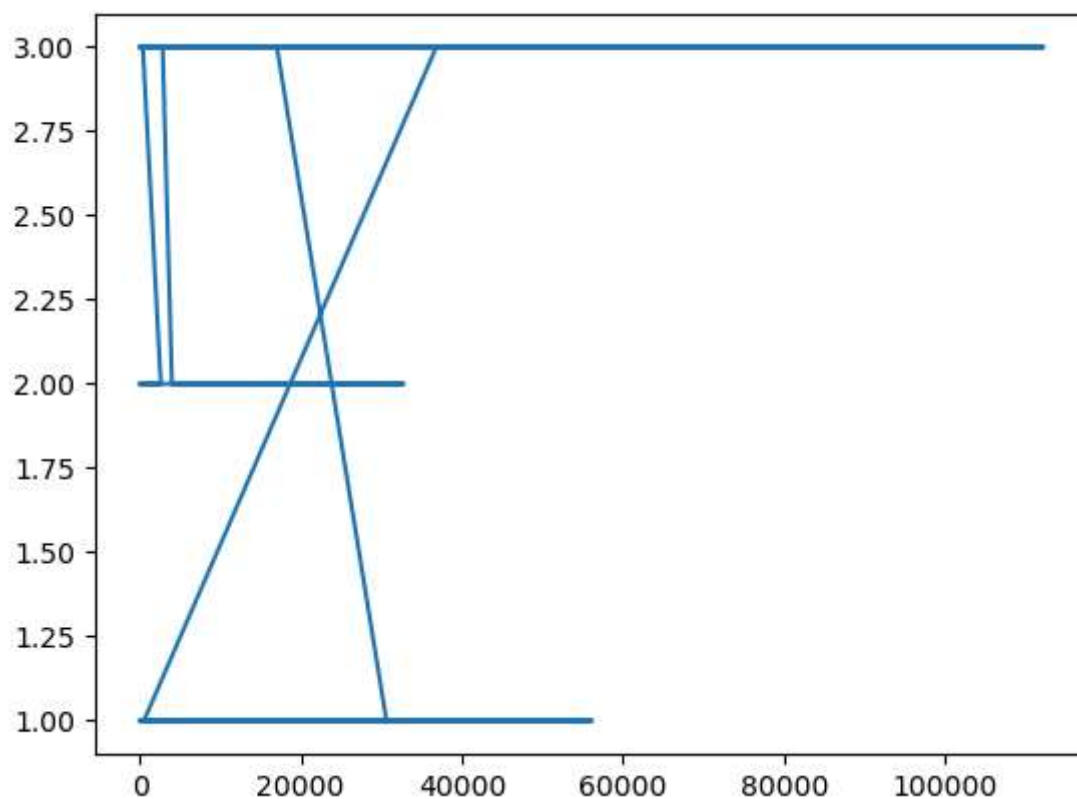
Channel	0
Fresh	0
Milk	0
Grocery	0
Frozen	0
Detergents_Paper	0
Delicassen	0
dtype:	int64

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

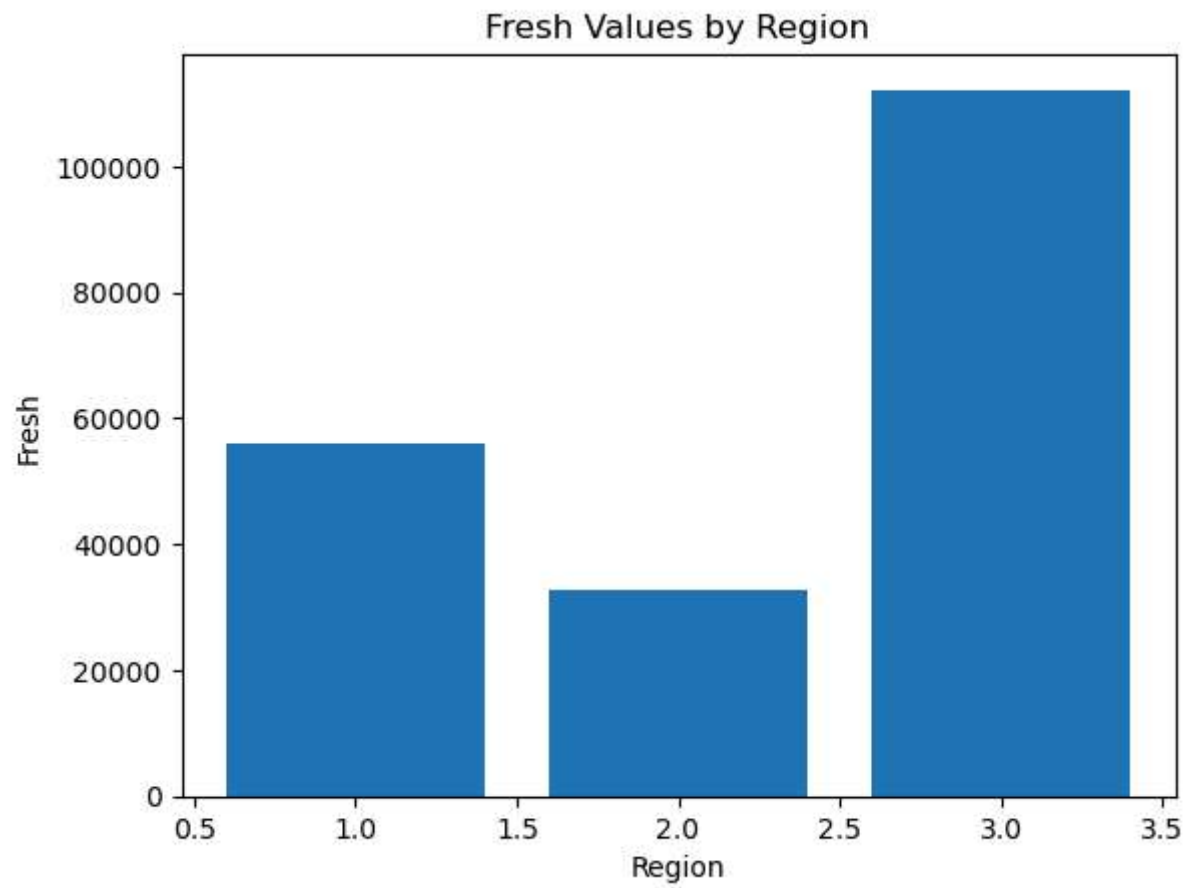
Region	0
dtype:	int64

```
In [10]: import matplotlib.pyplot as plt  
plt.plot(X["Fresh"],y["Region"])
```

```
Out[10]: [matplotlib.lines.Line2D at 0x25a3e530>]
```



```
In [12]: import matplotlib.pyplot as plt  
plt.bar(y["Region"], X["Fresh"])  
plt.xlabel('Region')  
plt.ylabel('Fresh')  
plt.title('Fresh Values by Region')  
plt.show()
```



In []: