

<https://archive-beta.ics.uci.edu/dataset/352/online+retail>

InvoiceNo: Номер счета. 6-значный целочисленный номер, однозначно присвоенный каждой транзакции. Если этот код начинается с буквы «с», это означает отмену.

StockCode: Код продукта (элемента). 5-значный целочисленный номер, однозначно присвоенный каждому отдельному продукту.

Description: Название продукта (элемента).

Quantity: Количество каждого продукта (элемента) за транзакцию.

InvoiceDate: Дата и время счета. День и время создания каждой транзакции.

UnitPrice: Цена товара за единицу в фунтах стерлингов.

ID клиента: номер клиента. 5-значный целочисленный номер, однозначно присвоенный каждому покупателю.

Country: Название страны, в которой проживает каждый клиент.

Импортирование pandas и os

```
1 import pandas as pd
2 import os
3 pd.set_option('display.max_columns', 5) # set max_columns (to print all columns)
4 pd.set_option('display.width', None) # set width for print (to remove line breaks)
Executed in 2ms, 17 Apr at 01:24:25
```

Функция и её вызов

```
1 def user_function(file_name:str):
2
3     # os.walk(".") возвращает три аргумента - имя корневого каталога(текущего), список имен вложенных папок, список файлов текущего каталога
4     for root, dirs, files in os.walk("."):
5         # выводим список всех файлов
6         print("list of files: ")
7         for filename in files:
8             print(filename)
9     data = pd.read_excel(file_name)
10    print("data: ")
11    print(data)
12    print("Count of data: ",data.shape[0])
13    print("info about data: ")
14    print(data.info())
15
16    exist_datetime=False
17    for i in data:
18        if data[i].dtype == "datetime64[ns]":
19            print("Datetime column is exists")
20            exist_datetime = True
21            break
22    if not exist_datetime:
```

```
22     if not exist_datetime:
23         # 1.01.2011
24         data.loc[:, "datetime"] = pd.date_range(start='1/1/2011', periods=data.shape[0], freq='1H')
25         data.to_excel('new_data_with_datetime.xlsx', index=False)
26         print("Column with datetime created with name \"new_data_with_datetime.xlsx\"")
27         file = str(input("Enter file name: "))
28         data=pd.read_excel(file)
29         print(data)
30 # ../../dataset/online_retail.xlsx
31 # ../../dataset/online_retail_without_datetime.xlsx
32 file_name = str(input("Enter file name: "))
33 user_function(file_name)
34
```

Вывод:

list of files:

new_data_with_datetime.xlsx

z_2.4.ipynb

data:

	InvoiceNo	StockCode	...	CustomerID	Country
0	536365	85123A	...	17850.0	United Kingdom
1	536365	71053	...	17850.0	United Kingdom
2	536365	84406B	...	17850.0	United Kingdom
3	536365	84029G	...	17850.0	United Kingdom
4	536365	84029E	...	17850.0	United Kingdom
...
19996	537999	22034	...	NaN	United Kingdom
19997	538000	22942	...	12748.0	United Kingdom
19998	538000	79321	...	12748.0	United Kingdom
19999	538001	35958	...	NaN	United Kingdom
20000	538002	22690	...	14932.0	Channel Islands

[20001 rows x 7 columns]

Count of data: 20001

info about data:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 20001 entries, 0 to 20000

Data columns (total 7 columns):

#	Column	Non-Null Count	Dtype
0	InvoiceNo	20001 non-null	object
1	StockCode	20001 non-null	object
2	Description	19930 non-null	object
3	Quantity	20001 non-null	int64
4	UnitPrice	20001 non-null	float64
5	CustomerID	13348 non-null	float64
6	Country	20001 non-null	object

dtypes: float64(2), int64(1), object(4)

memory usage: 1.1+ MB

None

Column with datetime created with name "new_data_with_datetime.xlsx"

	InvoiceNo	StockCode	...	CustomerID	Country
0	536365	85123A	...	17850.0	United Kingdom
1	536365	71053	...	17850.0	United Kingdom
2	536365	84406B	...	17850.0	United Kingdom
3	536365	84029G	...	17850.0	United Kingdom
4	536365	84029E	...	17850.0	United Kingdom
...
19996	537999	22034	...	NaN	United Kingdom
19997	538000	22942	...	12748.0	United Kingdom
19998	538000	79321	...	12748.0	United Kingdom
19999	538001	35958	...	NaN	United Kingdom
20000	538002	22690	...	14932.0	Channel Islands

[20001 rows x 7 columns]