

E-Commerce Project

```
In [1]: import numpy as np
import pandas as pd
import warnings
warnings.filterwarnings('ignore')

data = pd.read_csv('New data.csv',encoding = 'ISO-8859-1')
```

```
In [2]: data.head()
```

```
Out[2]:
```

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26	3.39	17850.0	United Kingdom

```
In [3]: data.shape
```

```
Out[3]: (541909, 8)
```

```
In [4]: data.Country.value_counts()
```

```
Out[4]: United Kingdom      495478
Germany            9495
France             8557
EIRE               8196
Spain              2533
Netherlands        2371
Belgium            2069
Switzerland        2002
Portugal           1519
Australia          1259
Norway             1086
Italy               803
Channel Islands    758
Finland            695
Cyprus              622
Sweden             462
Unspecified         446
Austria            401
Denmark            389
Japan               358
Poland              341
Israel              297
USA                 291
Hong Kong          288
Singapore          229
Iceland            182
Canada              151
Greece              146
Malta               127
United Arab Emirates   68
European Community    61
RSA                 58
Lebanon             45
Lithuania           35
Brazil              32
Czech Republic      30
Bahrain             19
Saudi Arabia        10
Name: Country, dtype: int64
```

```
In [5]: # total number of customer
```

```
data.describe(include = 'all')
```

Out[5]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID
count	541909	541909	540455	541909.000000	541909	541909.000000	406829.000000
unique	25900	4070	4223	NaN	23260	NaN	NaN
top	573585	85123A	WHITE HANGING HEART T- LIGHT HOLDER	NaN	10/31/2011 14:41	NaN	NaN
freq	1114	2313	2369	NaN	1114	NaN	NaN
mean	NaN	NaN	NaN	9.552250	NaN	4.611114	15287.690570
std	NaN	NaN	NaN	218.081158	NaN	96.759853	1713.600303
min	NaN	NaN	NaN	-80995.000000	NaN	-11062.060000	12346.000000
25%	NaN	NaN	NaN	1.000000	NaN	1.250000	13953.000000
50%	NaN	NaN	NaN	3.000000	NaN	2.080000	15152.000000
75%	NaN	NaN	NaN	10.000000	NaN	4.130000	16791.000000
max	NaN	NaN	NaN	80995.000000	NaN	38970.000000	18287.000000

```
In [6]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 541909 entries, 0 to 541908
Data columns (total 8 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   InvoiceNo   541909 non-null object 
 1   StockCode    541909 non-null object 
 2   Description  540455 non-null object 
 3   Quantity     541909 non-null int64  
 4   InvoiceDate  541909 non-null object 
 5   UnitPrice    541909 non-null float64
 6   CustomerID   406829 non-null float64
 7   Country      541909 non-null object 
dtypes: float64(2), int64(1), object(5)
memory usage: 33.1+ MB
```

```
In [7]: data['c_id'] = data.CustomerID.astype('object')
```

```
In [8]: data['InvoiceDate'] = pd.to_datetime(data['InvoiceDate'], format='%m/%d/%Y %H:%M')
```

In [9]: `data.head()`

Out[9]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	c_id
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	17850.0
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom	17850.0
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0

```
In [10]: data.describe(include = 'all')
```

Out[10]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID
count	541909	541909	540455	541909.000000	541909	541909.000000	406829.000000
unique	25900	4070	4223	NaN	23260	NaN	NaN
top	573585	85123A	WHITE HANGING HEART T- LIGHT HOLDER	NaN	2011-10-31 14:41:00	NaN	NaN
freq	1114	2313	2369	NaN	1114	NaN	NaN
first	NaN	NaN	NaN	NaN	2010-12-01 08:26:00	NaN	NaN
last	NaN	NaN	NaN	NaN	2011-12-09 12:50:00	NaN	NaN
mean	NaN	NaN	NaN	9.552250	NaN	4.611114	15287.690570
std	NaN	NaN	NaN	218.081158	NaN	96.759853	1713.600303
min	NaN	NaN	NaN	-80995.000000	NaN	-11062.060000	12346.000000
25%	NaN	NaN	NaN	1.000000	NaN	1.250000	13953.000000
50%	NaN	NaN	NaN	3.000000	NaN	2.080000	15152.000000
75%	NaN	NaN	NaN	10.000000	NaN	4.130000	16791.000000
max	NaN	NaN	NaN	80995.000000	NaN	38970.000000	18287.000000

```
In [11]: data.describe(include = 'all')
```

Out[11]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID
count	541909	541909	540455	541909.000000	541909	541909.000000	406829.000000
unique	25900	4070	4223	NaN	23260	NaN	NaN
top	573585	85123A	WHITE HANGING HEART T- LIGHT HOLDER	NaN	2011-10-31 14:41:00	NaN	NaN
freq	1114	2313	2369	NaN	1114	NaN	NaN
first	NaN	NaN	NaN	NaN	2010-12-01 08:26:00	NaN	NaN
last	NaN	NaN	NaN	NaN	2011-12-09 12:50:00	NaN	NaN
mean	NaN	NaN	NaN	9.552250	NaN	4.611114	15287.690570
std	NaN	NaN	NaN	218.081158	NaN	96.759853	1713.600303
min	NaN	NaN	NaN	-80995.000000	NaN	-11062.060000	12346.000000
25%	NaN	NaN	NaN	1.000000	NaN	1.250000	13953.000000
50%	NaN	NaN	NaN	3.000000	NaN	2.080000	15152.000000
75%	NaN	NaN	NaN	10.000000	NaN	4.130000	16791.000000
max	NaN	NaN	NaN	80995.000000	NaN	38970.000000	18287.000000



```
In [12]: 19052/365
```

Out[12]: 52.1972602739726

```
In [13]: data['day'] = data['InvoiceDate'].dt.day
```

```
In [14]: data.head()
```

Out[14]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	c_id
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	17850.0
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom	17850.0
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0

```
In [15]: data['invoice_time'] = data['InvoiceDate'].map(lambda x : 100 * x.year + x.month)
```

```
In [16]: data.head()
```

Out[16]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	c_id
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	17850.0
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom	17850.0
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0

◀ ▶

```
In [17]: data['invoice_time'] = data.invoice_time.astype('object')
```

```
In [18]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 541909 entries, 0 to 541908
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   InvoiceNo        541909 non-null   object 
 1   StockCode         541909 non-null   object 
 2   Description       540455 non-null   object 
 3   Quantity          541909 non-null   int64  
 4   InvoiceDate       541909 non-null   datetime64[ns]
 5   UnitPrice         541909 non-null   float64 
 6   CustomerID        406829 non-null   float64 
 7   Country            541909 non-null   object 
 8   c_id               406829 non-null   object 
 9   day                541909 non-null   int64  
 10  invoice_time      541909 non-null   object 
dtypes: datetime64[ns](1), float64(2), int64(2), object(6)
memory usage: 45.5+ MB
```

```
In [19]: data.describe(include = 'all')
```

Out[19]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID
count	541909	541909	540455	541909.000000	541909	541909.000000	406829.000000
unique	25900	4070	4223	NaN	23260	NaN	NaN
top	573585	85123A	WHITE HANGING HEART T- LIGHT HOLDER	NaN	2011-10-31 14:41:00	NaN	NaN
freq	1114	2313	2369	NaN	1114	NaN	NaN
first	NaN	NaN	NaN	NaN	2010-12-01 08:26:00	NaN	NaN
last	NaN	NaN	NaN	NaN	2011-12-09 12:50:00	NaN	NaN
mean	NaN	NaN	NaN	9.552250	NaN	4.611114	15287.690570
std	NaN	NaN	NaN	218.081158	NaN	96.759853	1713.600303
min	NaN	NaN	NaN	-80995.000000	NaN	-11062.060000	12346.000000
25%	NaN	NaN	NaN	1.000000	NaN	1.250000	13953.000000
50%	NaN	NaN	NaN	3.000000	NaN	2.080000	15152.000000
75%	NaN	NaN	NaN	10.000000	NaN	4.130000	16791.000000
max	NaN	NaN	NaN	80995.000000	NaN	38970.000000	18287.000000

In [20]: `data.head()`

Out[20]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	c_id
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	17850.0
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom	17850.0
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0



In [21]: `data = data[data.UnitPrice > 0]`

In [22]: `data.head()`

Out[22]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	c_id
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	17850.0
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom	17850.0
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0

```
In [23]: data.describe(include = 'all')
```

Out[23]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID
count	539392	539392	539392	539392.000000	539392	539392.000000	406789.000000
unique	23796	3938	4042	NaN	21802	NaN	NaN
top	573585	85123A	WHITE HANGING HEART T-LIGHT HOLDER	NaN	2011-10-31 14:41:00	NaN	NaN
freq	1114	2307	2365	NaN	1114	NaN	NaN
first	NaN	NaN	NaN	NaN	2010-12-01 08:26:00	NaN	NaN
last	NaN	NaN	NaN	NaN	2011-12-09 12:50:00	NaN	NaN
mean	NaN	NaN	NaN	9.845904	NaN	4.673648	15287.795830
std	NaN	NaN	NaN	215.412652	NaN	94.614722	1713.573064
min	NaN	NaN	NaN	-80995.000000	NaN	0.001000	12346.000000
25%	NaN	NaN	NaN	1.000000	NaN	1.250000	13954.000000
50%	NaN	NaN	NaN	3.000000	NaN	2.080000	15152.000000
75%	NaN	NaN	NaN	10.000000	NaN	4.130000	16791.000000
max	NaN	NaN	NaN	80995.000000	NaN	38970.000000	18287.000000

```
In [24]: # how many orders made by customer ?
```

```
temp1 = data.groupby(by=['c_id'], as_index = False)[['InvoiceNo']].count()
```

```
In [25]: temp1.sort_values(by = ['InvoiceNo'], ascending = False)
```

Out[25]:

	c_id	InvoiceNo
4041	17841.0	7983
1894	14911.0	5901
1299	14096.0	5128
330	12748.0	4641
1673	14606.0	2782
...
2524	15753.0	1
1317	14119.0	1
3125	16579.0	1
592	13120.0	1
2006	15070.0	1

4371 rows × 2 columns

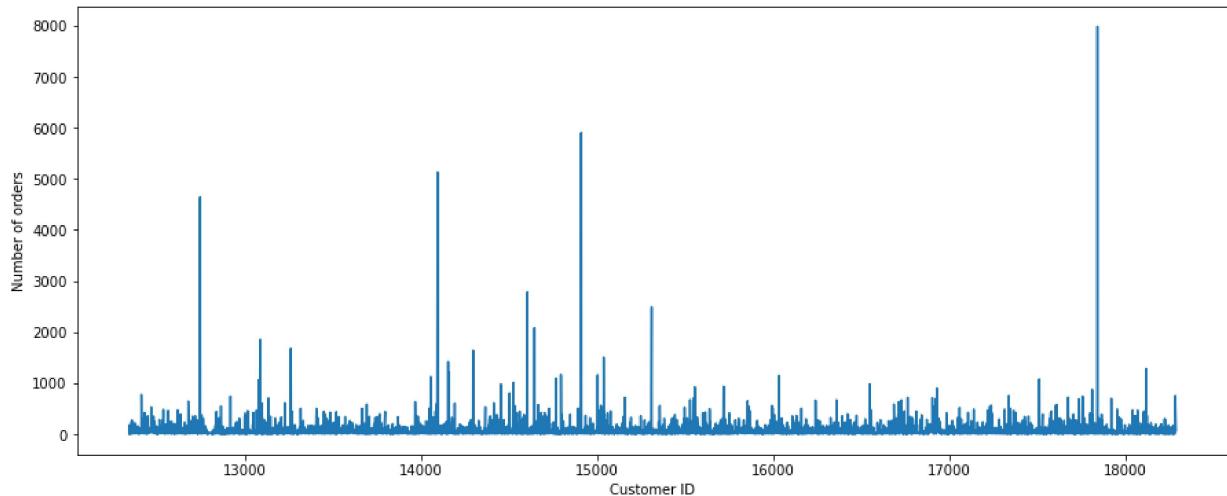
```
In [26]: temp1[temp1['InvoiceNo'] == 1]
```

Out[26]:

	c_id	InvoiceNo
125	12503.0	1
127	12505.0	1
467	12943.0	1
520	13017.0	1
577	13099.0	1
...
4240	18113.0	1
4260	18141.0	1
4286	18174.0	1
4294	18184.0	1
4333	18233.0	1

78 rows × 2 columns

```
In [27]: import matplotlib.pyplot as plt
import seaborn as sns
plt.subplots(figsize = (15,6))
plt.xlabel('Customer ID')
plt.ylabel('Number of orders')
plt.plot(temp1.c_id,temp1.InvoiceNo)
plt.show()
```



```
In [28]: # How much money spent by the customer
```

```
data['amount_spent'] = data['Quantity'] * data['UnitPrice']
```

```
In [29]: data.head()
```

Out[29]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	c_id
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	17850.0
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom	17850.0
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0

```
In [30]: temp2 = data.groupby(by = ['c_id'],as_index = False)[['amount_spent']].sum()
```

Out[30]:

	c_id	amount_spent
0	12346.0	0.00
1	12347.0	4310.00
2	12348.0	1797.24
3	12349.0	1757.55
4	12350.0	334.40
...
4366	18280.0	180.60
4367	18281.0	80.82
4368	18282.0	176.60
4369	18283.0	2094.88
4370	18287.0	1837.28

4371 rows × 2 columns

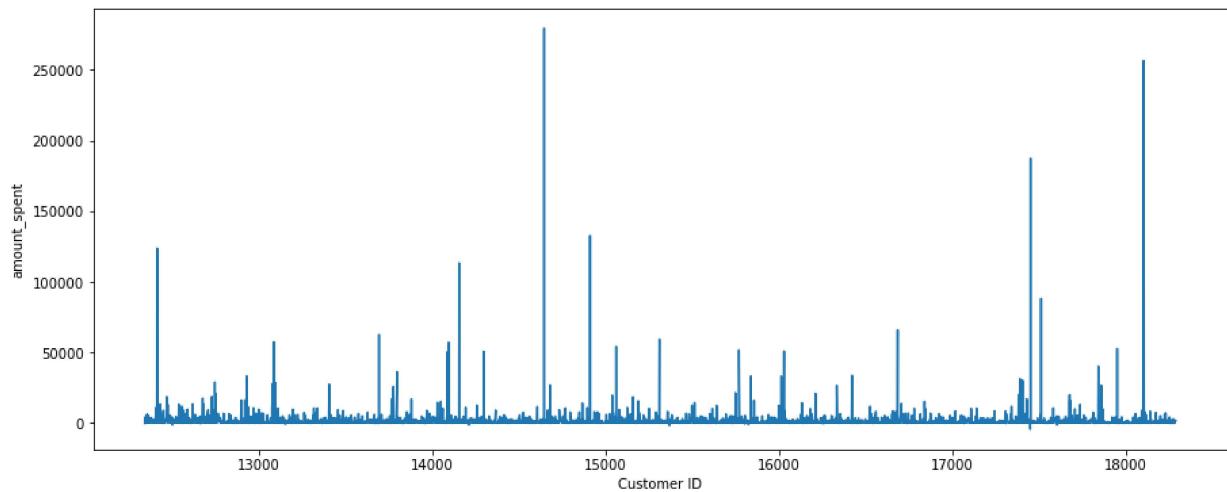
```
In [31]: temp2.sort_values(by = ['amount_spent'],ascending = False)
```

Out[31]:

	c_id	amount_spent
1702	14646.0	279489.02
4232	18102.0	256438.49
3757	17450.0	187482.17
1894	14911.0	132572.62
55	12415.0	123725.45
...
125	12503.0	-1126.00
3869	17603.0	-1165.30
1383	14213.0	-1192.20
2235	15369.0	-1592.49
3755	17448.0	-4287.63

4371 rows × 2 columns

```
In [32]: import matplotlib.pyplot as plt
import seaborn as sns
plt.subplots(figsize = (15,6))
plt.xlabel('Customer ID')
plt.ylabel('amount_spent')
plt.plot(temp2.c_id,temp2.amount_spent)
plt.show()
```



```
In [33]: data.head()
```

Out[33]:

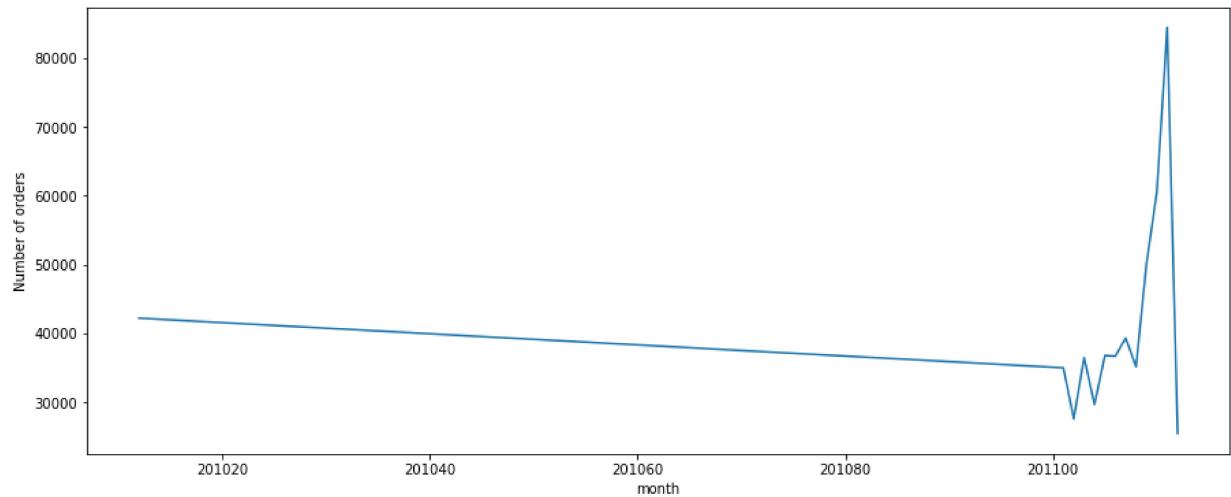
	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	c_id
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	17850.0
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom	17850.0
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0

```
In [34]: # How many order that was done per month
```

```
temp3 = data.groupby(by = ['invoice_time'],as_index = False)[['InvoiceNo']].count()
```

```
In [35]: import matplotlib.pyplot as plt
```

```
import seaborn as sns
plt.subplots(figsize = (15,6))
plt.xlabel('month')
plt.ylabel('Number of orders')
plt.plot(temp3.invoice_time,temp3.InvoiceNo)
plt.show()
```



```
In [36]: temp3
```

Out[36]:

	invoice_time	InvoiceNo
0	201012	42208
1	201101	35007
2	201102	27580
3	201103	36502
4	201104	29655
5	201105	36785
6	201106	36688
7	201107	39330
8	201108	35151
9	201109	50066
10	201110	60504
11	201111	84445
12	201112	25471

```
In [37]: temp3.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13 entries, 0 to 12
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   invoice_time    13 non-null   int64  
 1   InvoiceNo     13 non-null   int64  
dtypes: int64(2)
memory usage: 336.0 bytes
```

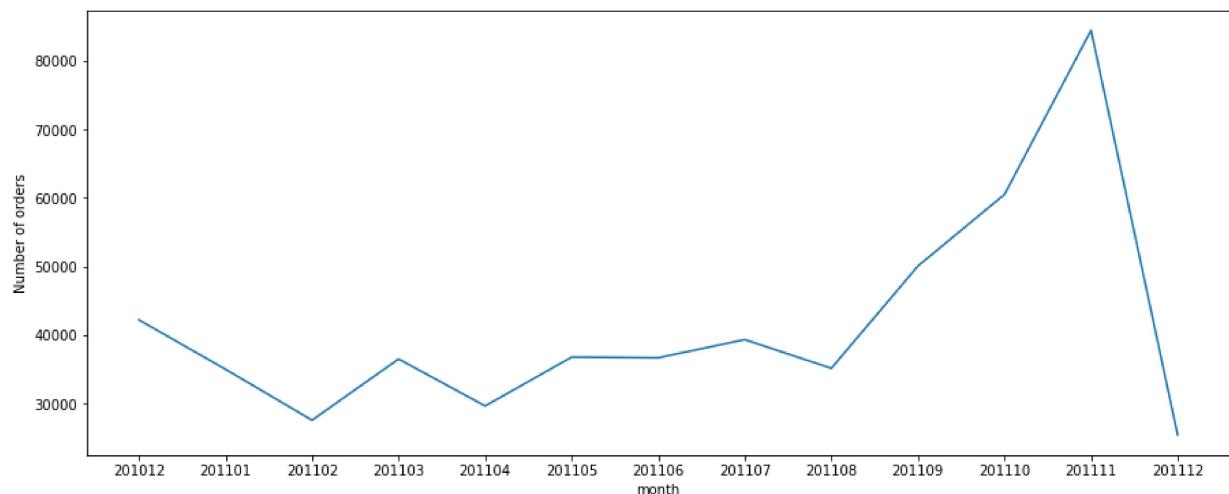
```
In [38]: temp3['month'] = temp3.invoice_time.astype('str')
```

```
In [39]: temp3
```

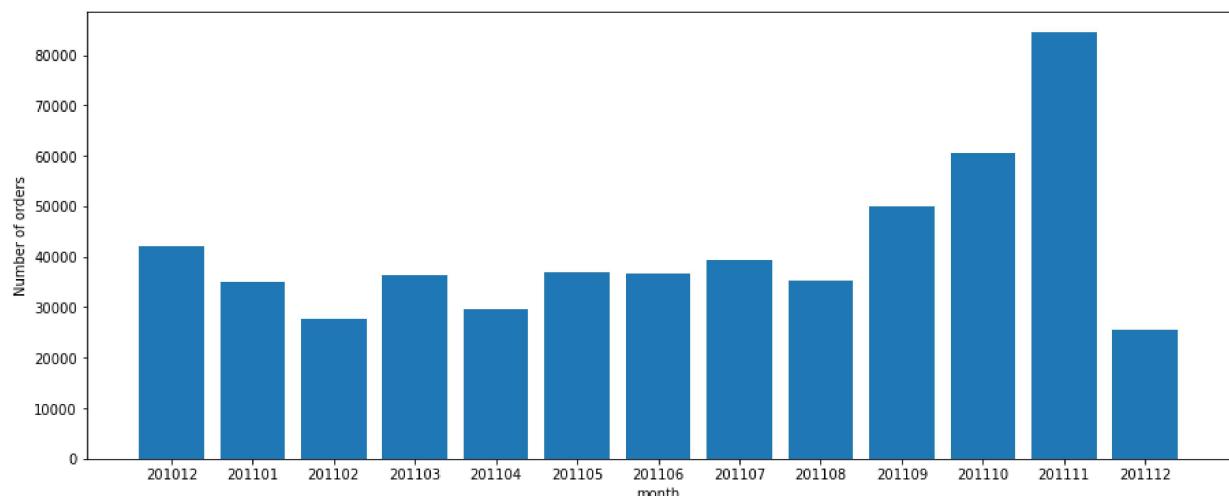
Out[39]:

	invoice_time	InvoiceNo	month
0	201012	42208	201012
1	201101	35007	201101
2	201102	27580	201102
3	201103	36502	201103
4	201104	29655	201104
5	201105	36785	201105
6	201106	36688	201106
7	201107	39330	201107
8	201108	35151	201108
9	201109	50066	201109
10	201110	60504	201110
11	201111	84445	201111
12	201112	25471	201112

```
In [40]: import matplotlib.pyplot as plt
import seaborn as sns
plt.subplots(figsize = (15,6))
plt.xlabel('month')
plt.ylabel('Number of orders')
plt.plot(temp3.month,temp3.InvoiceNo)
plt.show()
```



```
In [41]: plt.subplots(figsize = (15,6))
plt.xlabel('month')
plt.ylabel('Number of orders')
plt.bar(temp3.month,temp3.InvoiceNo)
plt.show()
```



```
In [42]: temp3 = data.groupby(by = ['invoice_time'],as_index = False)['amount_spent'].sum()  
temp3
```

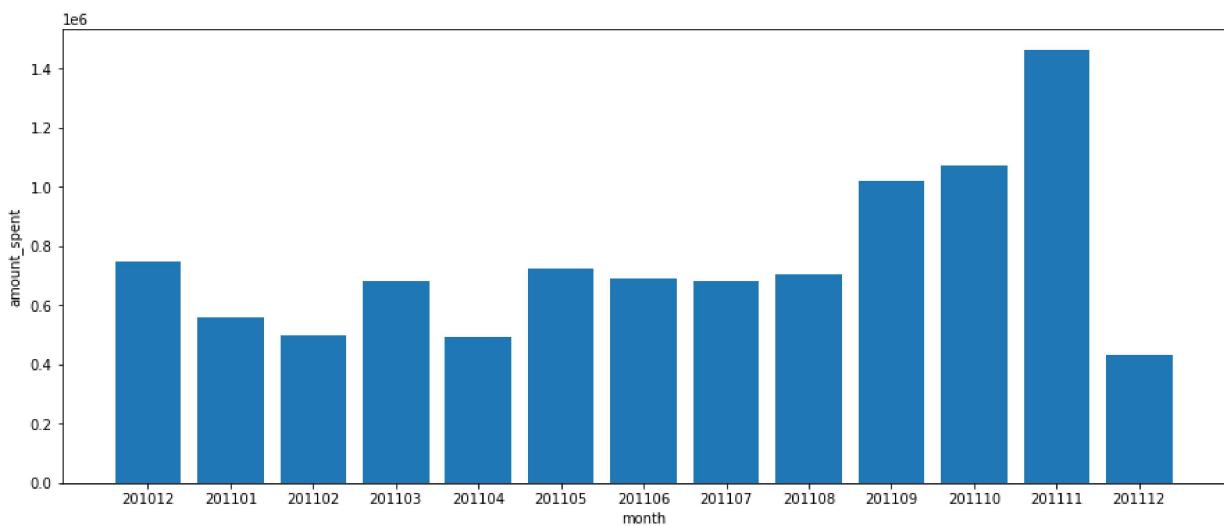
Out[42]:

	invoice_time	amount_spent
0	201012	748957.020
1	201101	560000.260
2	201102	498062.650
3	201103	683267.080
4	201104	493207.121
5	201105	723333.510
6	201106	691123.120
7	201107	681300.111
8	201108	704804.630
9	201109	1019687.622
10	201110	1070704.670
11	201111	1461756.250
12	201112	433668.010

```
In [43]: temp3['month'] = temp3.invoice_time.astype('str')  
temp3['month']
```

Out[43]: 0 201012
1 201101
2 201102
3 201103
4 201104
5 201105
6 201106
7 201107
8 201108
9 201109
10 201110
11 201111
12 201112
Name: month, dtype: object

```
In [44]: plt.subplots(figsize = (15,6))
plt.xlabel('month')
plt.ylabel('amount_spent')
plt.bar(temp3.month,temp3.amount_spent)
plt.show()
```



```
In [45]: data.head()
```

Out[45]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	c_id
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	17850.0
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom	17850.0
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0

```
In [46]: data['hour'] = data.InvoiceDate.dt.hour  
data['hour'].head()
```

```
Out[46]: 0    8  
1    8  
2    8  
3    8  
4    8  
Name: hour, dtype: int64
```

```
In [47]: data.head()
```

```
Out[47]:
```

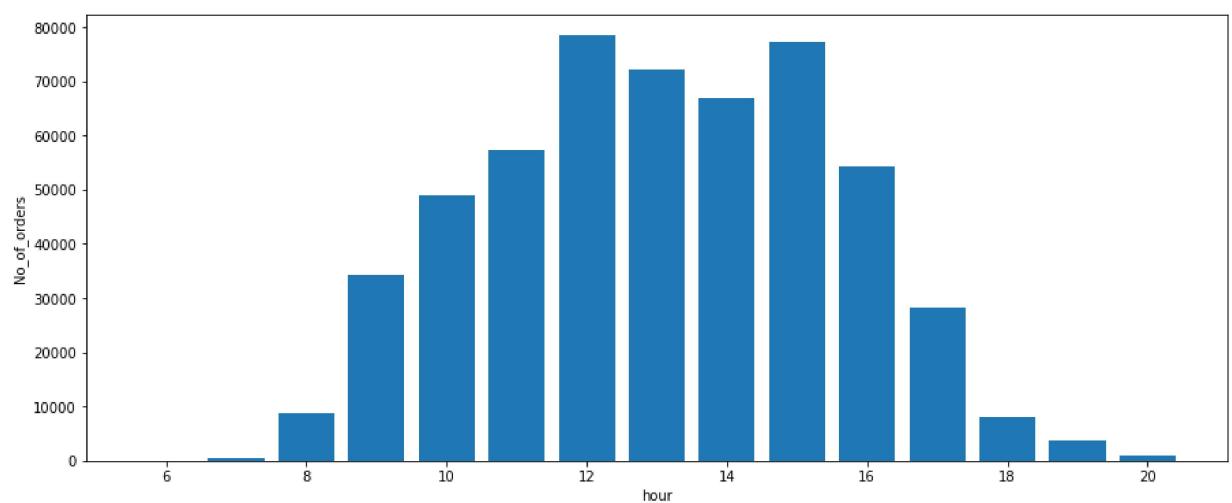
	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	c_id
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	17850.0
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom	17850.0
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0

```
In [48]: temp3 = data.groupby(by = ['hour'],as_index = False)[['InvoiceNo']].count()  
temp3
```

Out[48]:

	hour	InvoiceNo
0	6	41
1	7	383
2	8	8902
3	9	34226
4	10	48831
5	11	57332
6	12	78458
7	13	72064
8	14	66960
9	15	77139
10	16	54288
11	17	28274
12	18	7919
13	19	3704
14	20	871

```
In [49]: plt.subplots(figsize = (15,6))  
plt.xlabel('hour')  
plt.ylabel('No_of_orders')  
plt.bar(temp3.hour,temp3.InvoiceNo)  
plt.show()
```



```
In [50]: data.head()
```

Out[50]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	c_id
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	17850.0
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom	17850.0
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0

◀ ▶

```
In [51]: data_v1 = data[data['Country'] != 'United Kingdom']
data_v1.head()
```

Out[51]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	c_id
26	536370	22728	ALARM CLOCK BAKELIKE PINK	24	2010-12-01 08:45:00	3.75	12583.0	France	12583.0
27	536370	22727	ALARM CLOCK BAKELIKE RED	24	2010-12-01 08:45:00	3.75	12583.0	France	12583.0
28	536370	22726	ALARM CLOCK BAKELIKE GREEN	12	2010-12-01 08:45:00	3.75	12583.0	France	12583.0
29	536370	21724	PANDA AND BUNNIES STICKER SHEET	12	2010-12-01 08:45:00	0.85	12583.0	France	12583.0
30	536370	21883	STARS GIFT TAPE	24	2010-12-01 08:45:00	0.65	12583.0	France	12583.0

◀ ▶

```
In [52]: temp3 = data_v1.groupby(by = ['Country'],as_index = False)[['amount_spent']].sum()  
temp3
```

Out[52]:

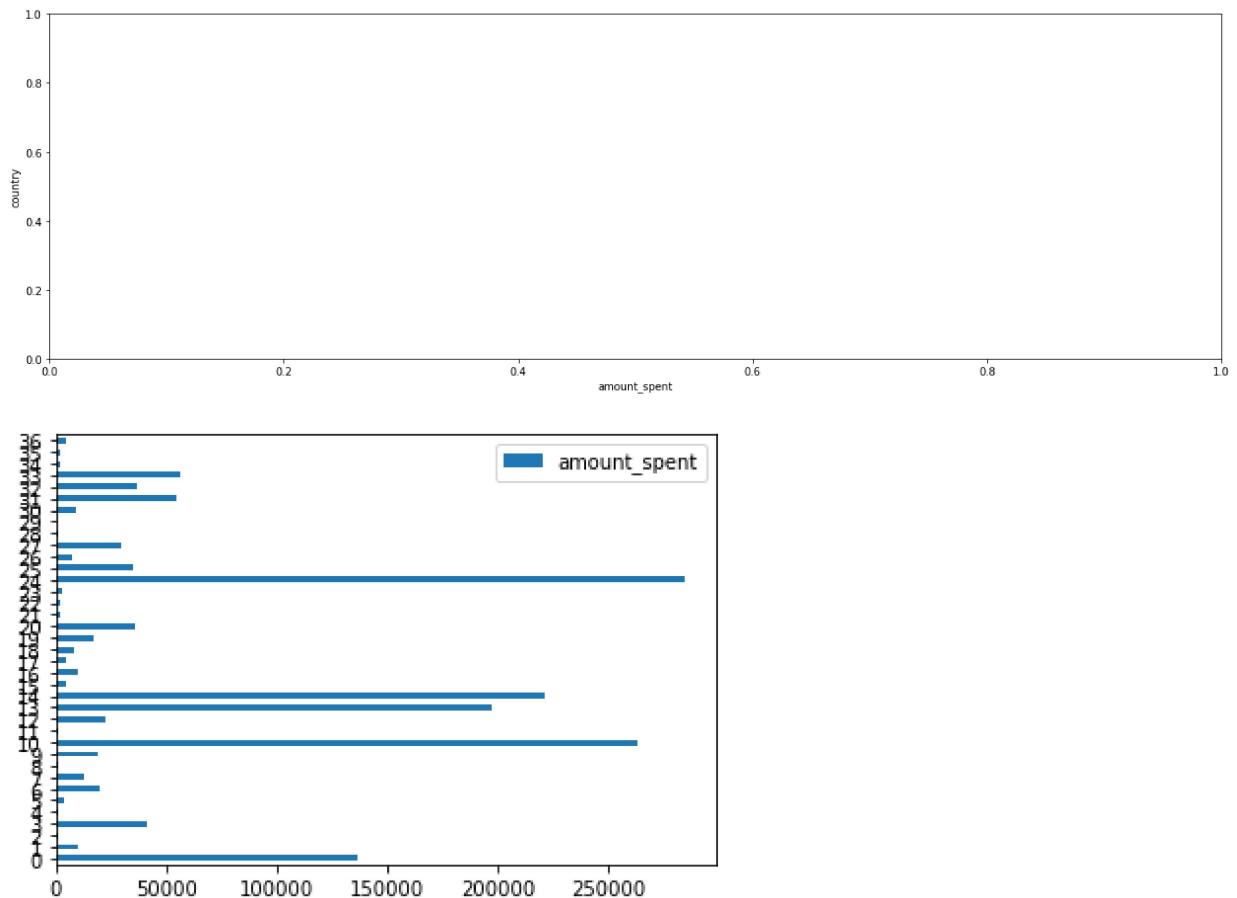
	Country	amount_spent
0	Australia	137077.27
1	Austria	10154.32
2	Bahrain	548.40
3	Belgium	40910.96
4	Brazil	1143.60
5	Canada	3666.38
6	Channel Islands	20086.29
7	Cyprus	12946.29
8	Czech Republic	707.72
9	Denmark	18768.14
10	EIRE	263276.82
11	European Community	1291.75
12	Finland	22326.74
13	France	197403.90
14	Germany	221698.21
15	Greece	4710.52
16	Hong Kong	10117.04
17	Iceland	4310.00
18	Israel	7907.82
19	Italy	16890.51
20	Japan	35340.62
21	Lebanon	1693.88
22	Lithuania	1661.06
23	Malta	2505.47
24	Netherlands	284661.54
25	Norway	35163.46
26	Poland	7213.14
27	Portugal	29367.02
28	RSA	1002.31
29	Saudi Arabia	131.17
30	Singapore	9120.39
31	Spain	54774.58
32	Sweden	36595.91
33	Switzerland	56385.35
34	USA	1730.92
35	United Arab Emirates	1902.28

```
Country  amount_spent
36        Unspecified      4749.79
```

```
In [53]: data.columns
```

```
Out[53]: Index(['InvoiceNo', 'StockCode', 'Description', 'Quantity', 'InvoiceDate',
       'UnitPrice', 'CustomerID', 'Country', 'c_id', 'day', 'invoice_time',
       'amount_spent', 'hour'],
      dtype='object')
```

```
In [54]: plt.subplots(figsize=(20, 6))
plt.xlabel('amount_spent')
plt.ylabel('country')
temp3.plot(kind='barh')
plt.show()
```



```
In [55]: temp3.sort_values(by = ['amount_spent'], ascending = False)
```

Out[55]:

	Country	amount_spent
24	Netherlands	284661.54
10	EIRE	263276.82
14	Germany	221698.21
13	France	197403.90
0	Australia	137077.27
33	Switzerland	56385.35
31	Spain	54774.58
3	Belgium	40910.96
32	Sweden	36595.91
20	Japan	35340.62
25	Norway	35163.46
27	Portugal	29367.02
12	Finland	22326.74
6	Channel Islands	20086.29
9	Denmark	18768.14
19	Italy	16890.51
7	Cyprus	12946.29
1	Austria	10154.32
16	Hong Kong	10117.04
30	Singapore	9120.39
18	Israel	7907.82
26	Poland	7213.14
36	Unspecified	4749.79
15	Greece	4710.52
17	Iceland	4310.00
5	Canada	3666.38
23	Malta	2505.47
35	United Arab Emirates	1902.28
34	USA	1730.92
21	Lebanon	1693.88
22	Lithuania	1661.06
11	European Community	1291.75
4	Brazil	1143.60
28	RSA	1002.31
8	Czech Republic	707.72
2	Bahrain	548.40

	Country	amount_spent
29	Saudi Arabia	131.17

In [56]: `data.head()`

Out[56]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	c_id
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	17850.0
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom	17850.0
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	17850.0



In [57]: `temp3 = data.groupby(by = ['Country', 'hour'], as_index = False)[['InvoiceNo']].count()`

```
In [58]: temp3
```

Out[58]:

	Country	hour	InvoiceNo
0	Australia	8	97
1	Australia	9	188
2	Australia	10	205
3	Australia	11	99
4	Australia	12	262
...
281	Unspecified	12	66
282	Unspecified	13	43
283	Unspecified	14	34
284	Unspecified	15	56
285	Unspecified	16	206

286 rows × 3 columns

```
In [59]: temp4 = temp3.groupby(by = ['Country'],as_index = False)[['InvoiceNo']].max()  
temp4.head()
```

Out[59]:

	Country	InvoiceNo
0	Australia	262
1	Austria	107
2	Bahrain	13
3	Belgium	351
4	Brazil	32

```
In [60]: temp5 = pd.merge(temp3,temp4 ,on='Country',how = 'left')
temp5
```

Out[60]:

	Country	hour	InvoiceNo_x	InvoiceNo_y
0	Australia	8	97	262
1	Australia	9	188	262
2	Australia	10	205	262
3	Australia	11	99	262
4	Australia	12	262	262
...
281	Unspecified	12	66	206
282	Unspecified	13	43	206
283	Unspecified	14	34	206
284	Unspecified	15	56	206
285	Unspecified	16	206	206

286 rows × 4 columns

```
In [61]: temp6 = temp5[temp5.InvoiceNo_x == temp5.InvoiceNo_y]
```

In [62]: temp6

Out[62]:

	Country	hour	InvoiceNo_x	InvoiceNo_y
4	Australia	12	262	262
9	Austria	10	107	107
17	Bahrain	13	13	13
25	Belgium	14	351	351
30	Brazil	10	32	32
32	Canada	10	78	78
39	Channel Islands	10	183	183
50	Cyprus	12	399	399
54	Czech Republic	8	15	15
61	Denmark	10	105	105
74	EIRE	10	1459	1459
85	European Community	10	40	40
89	Finland	9	200	200
103	France	10	1307	1307
118	Germany	13	1466	1466
126	Greece	9	36	36
135	Hong Kong	14	150	150
141	Iceland	14	60	60
143	Israel	8	171	171
153	Italy	13	200	200
161	Japan	11	121	121
168	Lebanon	10	45	45
169	Lithuania	12	34	34
171	Malta	8	57	57
181	Netherlands	11	560	560
193	Norway	13	489	489
200	Poland	10	101	101
213	Portugal	14	290	290
218	RSA	12	57	57
219	Saudi Arabia	10	9	9
222	Singapore	11	79	79
230	Spain	13	724	724
241	Sweden	13	93	93
249	Switzerland	10	555	555
260	USA	16	224	224
263	United Arab Emirates	17	37	37

	Country	hour	InvoiceNo_x	InvoiceNo_y
273	United Kingdom	15	72845	72845
285	Unspecified	16	206	206

Thankyou