

Welcome to Jupyter!

```
In [2]: import pandas as pd
import csv
```

```
In [5]: pd.__version__
```

```
Out[5]: '1.1.5'
```

```
In [ ]:
```

```
In [6]: !python --version
```

```
Python 3.6.13
```

```
In [4]: df = pd.read_csv ('Sales.csv')
print(df.head())
```

	Region	Country	Item Type \
0	Australia and Oceania	Tuvalu	Baby Food
1	Central America and the Caribbean	Grenada	Cereal
2	Europe	Russia	Office Supplies
3	Sub-Saharan Africa	Sao Tome and Principe	Fruits
4	Sub-Saharan Africa	Rwanda	Office Supplies

	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold \
0	Offline	H	5/28/2010	669165933	6/27/2010	9925
1	Online	C	8/22/2012	963881480	9/15/2012	2804
2	Offline	L	5/2/2014	341417157	5/8/2014	1779
3	Online	C	6/20/2014	514321792	7/5/2014	8102
4	Offline	L	2/1/2013	115456712	2/6/2013	5062

	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit
0	255.28	159.42	2533654.00	1582243.50	951410.50
1	205.70	117.11	576782.80	328376.44	248406.36
2	651.21	524.96	1158502.59	933903.84	224598.75
3	9.33	6.92	75591.66	56065.84	19525.82
4	651.21	524.96	3296425.02	2657347.52	639077.50

```
In [17]: df.shape
```

```
Out[17]: (100, 14)
```

```
In [24]: df.columns
```

```
Out[24]: Index(['Region', 'Country', 'Item Type', 'Sales Channel', 'Order Priority',
               'Order Date', 'Order ID', 'Ship Date', 'Units Sold', 'Unit Price',
               'Unit Cost', 'Total Revenue', 'Total Cost', 'Total Profit'],
              dtype='object')
```

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

Out[29]:

	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	Unit Price	U
0	Australia and Oceania	Tuvalu	Baby Food	Offline	H	5/28/2010	669165933	6/27/2010	9925	255.28	159.
1	Central America and the Caribbean	Grenada	Cereal	Online	C	8/22/2012	963881480	9/15/2012	2804	205.70	117.
2	Europe	Russia	Office Supplies	Offline	L	5/2/2014	341417157	5/8/2014	1779	651.21	524.
3	Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	C	6/20/2014	514321792	7/5/2014	8102	9.33	6.
4	Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L	2/1/2013	115456712	2/6/2013	5062	651.21	524.



In [32]:

df.tail()

Out[32]:

	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	U
95	Sub-Saharan Africa	Mali	Clothes	Online	M	7/26/2011	512878119	9/3/2011	888	109
96	Asia	Malaysia	Fruits	Offline	L	11/11/2011	810711038	12/28/2011	6267	9
97	Sub-Saharan Africa	Sierra Leone	Vegetables	Offline	C	6/1/2016	728815257	6/29/2016	1485	154
98	North America	Mexico	Personal Care	Offline	M	7/30/2015	559427106	8/8/2015	5767	81
99	Sub-Saharan Africa	Mozambique	Household	Offline	L	2/10/2012	665095412	2/15/2012	5367	668



In [36]:

len(df)

Out[36]: 100

In [38]:

df.sample(5)

Out[38]:

	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	Unit Price
88	Middle East and North Africa	Kuwait	Fruits	Online	M	4/30/2012	513417565	5/18/2012	522	9.33
7	Sub-Saharan Africa	Burkina Faso	Vegetables	Online	H	7/17/2012	871543967	7/27/2012	8082	154.06
12	Asia	Bangladesh	Clothes	Online	L	1/13/2017	187310731	3/1/2017	8263	109.28
69	Sub-Saharan Africa	Madagascar	Clothes	Offline	L	4/25/2015	610425555	5/28/2015	7342	109.28
34	Sub-Saharan Africa	Djibouti	Snacks	Online	M	2/25/2017	756274640	2/25/2017	7327	152.58



```
In [39]: df.Region
```

```
Out[39]: 0          Australia and Oceania
1  Central America and the Caribbean
2          Europe
3  Sub-Saharan Africa
4  Sub-Saharan Africa
...
95  Sub-Saharan Africa
96          Asia
97  Sub-Saharan Africa
98  North America
99  Sub-Saharan Africa
Name: Region, Length: 100, dtype: object
```

```
In [43]: df['Country']
```

```
Out[43]: 0          Tuvalu
1          Grenada
2          Russia
3  Sao Tome and Principe
4          Rwanda
...
95          Mali
96  Malaysia
97  Sierra Leone
98          Mexico
99  Mozambique
Name: Country, Length: 100, dtype: object
```

```
In [44]: df["Region","Country"]
```

KeyError

Traceback (most recent call last)

/srv/conda/envs/notebook/lib/python3.6/site-packages/pandas/core/indexes/base.py in get_

```

loc(self, key, method, tolerance)
    2897         try:
-> 2898             return self._engine.get_loc(casted_key)
    2899         except KeyError as err:

pandas/_libs/index.pyx in pandas._libs.index.IndexEngine.get_loc()

pandas/_libs/index.pyx in pandas._libs.index.IndexEngine.get_loc()

pandas/_libs/hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.get_
item()

pandas/_libs/hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.get_
item()

KeyError: ('Region', 'Country')

```

The above exception was the direct cause of the following exception:

```

KeyError                                Traceback (most recent call last)
<ipython-input-44-ce65fbcad0b8> in <module>
----> 1 df["Region","Country"]

/srv/conda/envs/notebook/lib/python3.6/site-packages/pandas/core/frame.py in __getitem__
(self, key)
    2904         if self.columns.nlevels > 1:
    2905             return self._getitem_multilevel(key)
-> 2906         indexer = self.columns.get_loc(key)
    2907         if is_integer(indexer):
    2908             indexer = [indexer]

/srv/conda/envs/notebook/lib/python3.6/site-packages/pandas/core/indexes/base.py in get_
loc(self, key, method, tolerance)
    2898         return self._engine.get_loc(casted_key)
    2899         except KeyError as err:
-> 2900             raise KeyError(key) from err
    2901
    2902         if tolerance is not None:

KeyError: ('Region', 'Country')

```

In [42]:

```
df[["Region","Country"]]
```

Out[42]:

	Region	Country
0	Australia and Oceania	Tuvalu
1	Central America and the Caribbean	Grenada
2	Europe	Russia
3	Sub-Saharan Africa	Sao Tome and Principe
4	Sub-Saharan Africa	Rwanda
...
95	Sub-Saharan Africa	Mali
96	Asia	Malaysia
97	Sub-Saharan Africa	Sierra Leone
98	North America	Mexico

	Region	Country
99	Sub-Saharan Africa	Mozambique

100 rows × 2 columns

```
In [ ]:
```

```
In [47]: # Descriptive numeritive values
df.describe()
```

Out[47]:

	Order ID	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	Total Profit
count	1.000000e+02	100.000000	100.000000	100.000000	1.000000e+02	1.000000e+02	1.000000e+02
mean	5.550204e+08	5128.710000	276.761300	191.048000	1.373488e+06	9.318057e+05	4.416820e+05
std	2.606153e+08	2794.484562	235.592241	188.208181	1.460029e+06	1.083938e+06	4.385379e+05
min	1.146066e+08	124.000000	9.330000	6.920000	4.870260e+03	3.612240e+03	1.258020e+03
25%	3.389225e+08	2836.250000	81.730000	35.840000	2.687212e+05	1.688680e+05	1.214436e+05
50%	5.577086e+08	5382.500000	179.880000	107.275000	7.523144e+05	3.635664e+05	2.907680e+05
75%	7.907551e+08	7369.000000	437.200000	263.330000	2.212045e+06	1.613870e+06	6.358288e+05
max	9.940222e+08	9925.000000	668.270000	524.960000	5.997055e+06	4.509794e+06	1.719922e+06

```
In [48]: df.describe(include="object")
```

Out[48]:

	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Ship Date
count	100	100	100	100	100	100	100
unique	7	76	12	2	4	100	99
top	Sub-Saharan Africa	The Gambia	Clothes	Online	H	2/10/2012	11/17/2010
freq	36	4	13	50	30	1	2

```
In [49]: df.isnull().values.any()
```

Out[49]: False

```
In [ ]:
```