In [130]:
```python
import pandas as pd
import numpy as np
```

In [148]:
```python
filePath = 'G:/PRCTICE/Python/'
inputFileName = 'SalesOrderHeader.xlsx'
outputFileName = 'SalesOrderHeader_Analysis_Result.xlsx'
inputFile = filePath + inputFileName
outputFile = filePath + outputFileName
```

In [149]:
```python
data = pd.read_excel(inputFile)
```

Chage index name and serial:

In [150]:
```python
data.index.name = "Index"
data.index = data.index + 1
```

Substring:

In [151]:
```python
SalesOrderCode  = data['SalesOrderNumber'].astype(str).str[:3]
```

Insert new column in a specific position:

In [152]:
```python
data.insert(loc=7, column='SalesOrderCode', value=SalesOrderCode)
```

In [153]:
```python
data.head(3)
```

Out[153]:

	SalesOrderID	RevisionNumber	OrderDate	DueDate	ShipDate	Status	OnlineOrderFlag	SalesOrderCode	SalesOrderNumber	PurchaseOrderNumbe
Index										
1	43659	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43659	PO52214578
2	43660	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43660	PO1885012750
3	43661	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43661	PO1847318962

3 rows × 27 columns

Replace:

In [ ]:
```python
data['DueDate']=data['DueDate'].replace('1970-01-01', np.nan)
```

Apply IN, Between and & with where clause:

In [155]:
```python
condition1 = data['SalesOrderCode'].isin(['SO4','SO5'])
condition2 = data['OrderDate'].between('2011-05-31','2013-05-31')
data[condition1 & condition2].head(2)
```

Out[155]:

	SalesOrderID	RevisionNumber	OrderDate	DueDate	ShipDate	Status	OnlineOrderFlag	SalesOrderCode	SalesOrderNumber	PurchaseOrderNumbe
Index										
1	43659	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43659	PO52214578
2	43660	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43660	PO1885012750

2 rows × 27 columns

In [156]:
```python
data[condition1 & condition2].tail(2)
```

Out[156]:

	SalesOrderID	RevisionNumber	OrderDate	DueDate	ShipDate	Status	OnlineOrderFlag	SalesOrderCode	SalesOrderNumber	PurchaseOrderNumbe
Index										
7543	51201	8	2013-05-31	2013-06-12	2013-06-07	5	1	SO5	SO51201	NaN
7544	51202	8	2013-05-31	2013-06-12	2013-06-07	5	1	SO5	SO51202	NaN

2 rows × 27 columns

In [211]:
```python
data.where(condition1 & condition2).head(2)
```

Out[211]:

	SL#	SalesOrderID	RevisionNumber	OrderDate	DueDate	ShipDate	Status	OnlineOrderFlag	SalesOrderCode	SalesOrderNumber	...	CreditCardID
0	NaN	NaN	NaN	NaT	NaT	NaT	NaN	NaN	NaN	NaN	...	NaN
1	2.0	43660.0	8.0	2011-05-31	2011-06-12	2011-06-07	5.0	0.0	SO4	SO43660	...	5618.0

2 rows × 28 columns

In [212]:
```python
data.query("RevisionNumber in [8,7]").head(3)
```

Out[212]:

	SL#	SalesOrderID	RevisionNumber	OrderDate	DueDate	ShipDate	Status	OnlineOrderFlag	SalesOrderCode	SalesOrderNumber	...	CreditCardID
0	1	43659	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43659	...	16281.0
1	2	43660	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43660	...	5618.0
2	3	43661	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43661	...	1346.0

3 rows × 28 columns

group by with and without meging index column:

In [157]:
```python
data_groupedby = data.groupby(['SalesOrderCode'], as_index=False).agg('count')
```

In [158]:
```python
data_groupedby.head(3)
```

Out[158]:

	SalesOrderCode	SalesOrderID	RevisionNumber	OrderDate	DueDate	ShipDate	Status	OnlineOrderFlag	SalesOrderNumber	PurchaseOrderNumber	..
0	SO4	6341	6341	6341	6341	6341	6341	6341	6341	1758	..
1	SO5	10000	10000	10000	10000	10000	10000	10000	10000	1147	..
2	SO6	10000	10000	10000	10000	10000	10000	10000	10000	720	..

3 rows × 27 columns

In [159]:
```python
data_groupedby = data.groupby(['SalesOrderCode']).agg('count')
```

In [160]:
```python
data_groupedby.head()
```

Out[160]:

	SalesOrderID	RevisionNumber	OrderDate	DueDate	ShipDate	Status	OnlineOrderFlag	SalesOrderNumber	PurchaseOrderNumber	Acco
SalesOrderCode										
SO4	6341	6341	6341	6341	6341	6341	6341	6341	1758	
SO5	10000	10000	10000	10000	10000	10000	10000	10000	1147	
SO6	10000	10000	10000	10000	10000	10000	10000	10000	720	
SO7	5124	5124	5124	5124	5124	5124	5124	5124	181	

4 rows × 26 columns

Export every column with distinct value to excel:

In [175]:
```python
writer = pd.ExcelWriter(outputFile, engine='xlsxwriter')
```

In [176]:
```python
data_groupedby.to_excel(writer, 'Tally')
```

In [177]:
```python
total = len(data.index)
```

In [ ]:
```python
data.insert(loc=0, column='SL#', value=data.index)
```

In [178]:
```python
for column in data:
    if(data[column].name == 'SalesOrderCode' or  data[column].name == 'SL#'):
        continue
    else:
        col_names =  [data[column].name]
        my_df  = pd.DataFrame(columns = col_names)
        my_df[data[column].name] = data[column].unique().tolist()
        if((my_df[column].nunique()/total) <= 0.02):
            grouped_data = data[['SalesOrderCode',data[column].name,'SL#']]
            grouped_data = grouped_data.groupby(['SalesOrderCode',data[column].name], as_index=False).agg('count')
            #Rename index column here.
            grouped_data = grouped_data.rename(columns={'SL#': 'Count'})
            if(len(grouped_data) > 0):
                grouped_data.to_excel(writer, sheet_name = data[column].name)
```

In [179]:
```python
writer.save()
writer.close()
```

Set, Reset index:

In [184]:
```python
data.set_index("SL#",inplace= True)
```

In [185]:
```python
data.head()
```

Out[185]:

	SalesOrderID	RevisionNumber	OrderDate	DueDate	ShipDate	Status	OnlineOrderFlag	SalesOrderCode	SalesOrderNumber	PurchaseOrderNumber
SL#										
1	43659	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43659	PO522145787
2	43660	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43660	PO18850127500
3	43661	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43661	PO18473189620
4	43662	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43662	PO18444174044
5	43663	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43663	PO18009186470

5 rows × 27 columns

In [187]:
```python
data.reset_index(drop=False, inplace=True)
```

In [194]:
```python
data.head(2)
```

Out[194]:

	SL#	SalesOrderID	RevisionNumber	OrderDate	DueDate	ShipDate	Status	OnlineOrderFlag	SalesOrderCode	SalesOrderNumber	...	CreditCardID
0	1	43659	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43659	...	16281.0
1	2	43660	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43660	...	5618.0

2 rows × 28 columns

Show data by loc & iloc:

In [198]:
```python
data.iloc[0:1]
```

Out[198]:

	SL#	SalesOrderID	RevisionNumber	OrderDate	DueDate	ShipDate	Status	OnlineOrderFlag	SalesOrderCode	SalesOrderNumber	...	CreditCardID
0	1	43659	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43659	...	16281.0

1 rows × 28 columns

In [207]:
```python
data.iloc[0,6:8]
```

Out[207]:
```
Status             5
OnlineOrderFlag    0
Name: 0, dtype: object
```

Rename column:

In [210]:
```python
data.rename(columns= {'ShipDate':'ShipingDate'},inplace= False).head(3)
```

Out[210]:

	SL#	SalesOrderID	RevisionNumber	OrderDate	DueDate	ShipingDate	Status	OnlineOrderFlag	SalesOrderCode	SalesOrderNumber	...	CreditCardID
0	1	43659	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43659	...	16281.0
1	2	43660	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43660	...	5618.0
2	3	43661	8	2011-05-31	2011-06-12	2011-06-07	5	0	SO4	SO43661	...	1346.0

3 rows × 28 columns

In [247]:
```python
import pandas as pd
import numpy as np
```

In [252]:
```
%%HTML
<style type="text/css">
    table.dataframe td, table.dataframe th {
        border-style: inset;
    }
</style>
```

In [234]:
```python
filePath = 'G:/PRCTICE/Python/'
inputFileName = 'BusinessEntity.xlsx'
outputFileName = 'SalesOrderHeader_Analysis_Result.xlsx'
inputFile = filePath + inputFileName
outputFile = filePath + outputFileName
```

In [239]:
```python
data = pd.read_excel(inputFile)
```

In [240]:
```python
data.head()
```

Out[240]:

	BusinessEntityID	Title	FirstName	MiddleName	LastName	Suffix	PhoneNumber	PhoneNumberType	EmailAddress	EmailPromotion	AddressT
0	1699	Mr.	David	R.	Robinett	NaN	238-555-0100	Home	david22@adventure-works.com	1	Ho
1	1700	Ms.	Rebecca	A.	Robinson	NaN	648-555-0100	Cell	rebecca3@adventure-works.com	0	Ho
2	1701	Ms.	Dorothy	B.	Robinson	NaN	423-555-0100	Cell	dorothy3@adventure-works.com	2	Ho
3	1702	Ms.	Carol Ann	F.	Rockne	NaN	439-555-0100	Cell	carolann0@adventure-works.com	0	Ho
4	1703	Mr.	Scott	M.	Rodgers	NaN	989-555-0100	Cell	scott10@adventure-works.com	0	Ho

In [246]:
```python
data.set_index(keys=['CountryRegionName','StateProvinceName','City']).
```

Out[246]:

			BusinessEntityID	Title	FirstName	MiddleName	LastName	Suffix	PhoneNumber	PhoneNumber
CountryRegionName	StateProvinceName	City								
Germany	Nordrhein-Westfalen	Solingen	1699	Mr.	David	R.	Robinett	NaN	238-555-0100	H
Australia	Victoria	Seaford	1700	Ms.	Rebecca	A.	Robinson	NaN	648-555-0100	
		Geelong	1701	Ms.	Dorothy	B.	Robinson	NaN	423-555-0100	
United Kingdom	England	Lancaster	1702	Ms.	Carol Ann	F.	Rockne	NaN	439-555-0100	
Australia	Queensland	East Brisbane	1703	Mr.	Scott	M.	Rodgers	NaN	989-555-0100	
United Kingdom	England	Esher-Molesey	1704	Mr.	Jim	NaN	Rodman	NaN	899-555-0100	H
United States	California	Concord	1705	Mr.	Eric	NaN	Rothenberg	NaN	326-555-0100	H
Australia	New South Wales	St. Leonards	1706	Mr.	Michael	L.	Rothkugel	NaN	358-555-0100	
	Queensland	East Brisbane	1707	Mr.	Pablo	NaN	Rovira Diez	NaN	786-555-0100	
	Victoria	Seaford	1708	Ms.	Linda	R.	Rousey	NaN	369-555-0100	H
United States	California	Beverly Hills	1709	Mr.	Luke	J.	Roy	NaN	583-555-0100	H
Australia	Queensland	Rockhampton	1710	Ms.	Lisa	K.	Roy	NaN	953-555-0100	
France	Seine Saint Denis	Saint-Denis	1711	Mr.	Michael	NaN	Ruggiero	NaN	227-555-0100	
Australia	Victoria	Seaford	1712	Ms.	Pearlie	J.	Rusek	NaN	633-555-0100	H
France	Seine Saint Denis	Pantin	1713	Ms.	Andrea	NaN	Rusko	NaN	587-555-0100	
Germany	Hessen	Hamburg	1714	Mr.	Andy	NaN	Ruth	NaN	856-555-0100	
Australia	Victoria	Warrnambool	1715	Ms.	Justine	J.	Ryan	NaN	498-555-0100	H
Germany	Hessen	Salzgitter	1716	Ms.	Deanna	N.	Sabella	NaN	529-555-0100	H
France	Seine (Paris)	Paris	1717	Mr.	Lane	NaN	Sacksteder	NaN	114-555-0100	H
	Pas de Calais	Boulogne-sur-Mer	1718	Mr.	Peter	NaN	Saddow	NaN	132-555-0100	H
Australia	Queensland	East Brisbane	1719	Ms.	Sharon	NaN	Salavaria	NaN	109-555-0100	
	New South Wales	Coffs Harbour	1720	Mr.	Irving	W	Schmidt	NaN	111-555-0100	
Germany	Saarland	Stuttgart	1721	Mr.	Raymond	NaN	Sam	NaN	313-555-0100	
United Kingdom	England	London	1722	NaN	Mandar	NaN	Samant	NaN	129-555-0100	H
Canada	British Columbia	Cliffside	1723	Mr.	John	J	Sandstone	NaN	114-555-0100	
Australia	New South Wales	Lavender Bay	1724	Ms.	Madalena	J.	Sanchez	NaN	749-555-0100	
United States	California	Colma	1725	Mr.	Thomas	T.	Sanchez	Jr.	127-555-0100	
France	Essonne	Les Ulis	1726	Mr.	Ken	NaN	Sánchez	NaN	967-555-0100	H
Germany	Nordrhein-Westfalen	Solingen	1727	Mr.	Mikael	NaN	Sandberg	NaN	854-555-0100	
Australia	New South Wales	Milsons Point	1728	Ms.	Mary	R.	Sandidge	NaN	113-555-0100	
...	...	...	...	...	...	...	...	...	...	
Germany	Nordrhein-Westfalen	Paderborn	20749	NaN	Crystal	NaN	Liu	NaN	1 (11) 500 555-0194	H
United States	Oregon	Lake Oswego	20750	NaN	Isabella	M	Stewart	NaN	126-555-0181	H
Australia	New South Wales	Newcastle	20751	NaN	Crystal	A	Yang	NaN	1 (11) 500 555-0112	H
United States	California	Colma	20752	NaN	Isabella	D	Sanchez	NaN	917-555-0121	H
		El Cajon	20753	NaN	Isabella	M	Morris	NaN	693-555-0146	
Australia	Queensland	Gold Coast	20754	NaN	Crystal	L	Huang	NaN	1 (11) 500 555-0127	
United Kingdom	England	Kirkby	20755	NaN	Crystal	W	Wu	NaN	1 (11) 500 555-0151	H
France	Seine Saint Denis	Drancy	20756	NaN	Crystal	NaN	Lin	NaN	1 (11) 500 555-0113	
Australia	Queensland	East Brisbane	20757	NaN	Carol	L	Zhou	NaN	1 (11) 500 555-0141	
United States	Washington	Seattle	20758	NaN	Isabella	NaN	Rogers	NaN	199-555-0140	
		Renton	20759	NaN	Isabella	NaN	Reed	NaN	754-555-0118	
		Renton	20759	NaN	Isabella	NaN	Reed	NaN	754-555-0118	
		Lynnwood	20760	NaN	Isabella	D	Cook	NaN	248-555-0151	H
Canada	British Columbia	Haney	20761	NaN	Isabella	L	Morgan	NaN	118-555-0127	H
Australia	South Australia	Findon	20762	NaN	Crystal	NaN	Zhao	NaN	1 (11) 500 555-0161	H
	Queensland	Rockhampton	20763	NaN	Isabella	NaN	Bell	NaN	1 (11) 500 555-0157	H
	Victoria	Geelong	20764	NaN	Crystal	A	Lu	NaN	1 (11) 500 555-0199	
	New South Wales	Silverwater	20765	NaN	Francis	D	Ruiz	NaN	1 (11) 500 555-0190	
		North Ryde	20766	NaN	Crystal	E	Xu	NaN	1 (11) 500 555-0127	
Germany	Brandenburg	Eilenburg	20767	NaN	Crystal	E	Sun	NaN	1 (11) 500 555-0119	
France	Moselle	Metz	20768	NaN	Isabella	NaN	Murphy	NaN	1 (11) 500 555-0117	H
Germany	Hamburg	Hamburg	20769	NaN	Crystal	C	Zhu	NaN	1 (11) 500 555-0180	H
	Hessen	Duesseldorf	20770	NaN	Crystal	J	Gao	NaN	1 (11) 500 555-0136	
United States	California	Beverly Hills	20771	NaN	Isabella	NaN	Bailey	NaN	808-555-0174	
United Kingdom	England	London	20772	NaN	Crystal	J	Liang	NaN	1 (11) 500 555-0120	H
		W. York	20773	NaN	Crystal	NaN	Guo	NaN	1 (11) 500 555-0171	
United States	California	Torrance	20774	NaN	Isabella	F	Richardson	NaN	910-555-0166	
	Washington	Bremerton	20775	NaN	Crystal	S	He	NaN	813-555-0148	H
France	Yveline	Versailles	20776	NaN	Crystal	NaN	Zheng	NaN	1 (11) 500 555-0171	H
Australia	New South Wales	Darlinghurst	20777	NaN	Crystal	NaN	Hu	NaN	1 (11) 500 555-0126	H

18508 rows × 15 columns

In [269]:
```python
data_count = data.count()
```

In [270]:
```python
data_count.head()
```

Out[270]:
```
BusinessEntityID    18508
Title                 101
FirstName           18508
MiddleName          10666
LastName            18508
dtype: int64
```

In [271]:
```python
col_names =  ["Column","Value"]
my_df  = pd.DataFrame([data_count],columns = col_names)
```

In [272]:
```python
my_df
```

Out[272]:

	Column	Value
0	NaN	NaN

In [ ]: