In [1]:
```
import os, pandas as pd, numpy as np
```

In [3]:
```
os.chdir('E:\\Data Science Course\\raw_data')
```

In [4]:
```
nortel=pd.read_csv('nortel.csv')
```

In [5]:
```
nortel.shape
```

Out[5]:
```
(3163, 18)
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3163 entries, 0 to 3162
Data columns (total 18 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   Acc_no       3163 non-null   object 
 1   REG_No       527 non-null    object 
 2   Reg_date     2925 non-null   object 
 3   Comp_date    2874 non-null   object 
 4   Month_year   3163 non-null   object 
 5   RP           3163 non-null   object 
 6   RS           3163 non-null   object 
 7   Unit         3163 non-null   object 
 8   IR           3163 non-null   object 
 9   Performance  3163 non-null   object 
 10  OverallPer   3163 non-null   int64  
 11  t_respond    3163 non-null   int64  
 12  Understand   3163 non-null   int64  
 13  t_resolve    3163 non-null   int64  
 14  s_equipment  0 non-null      float64
 15  s_Del        486 non-null    float64
 16  speak        3057 non-null   float64
 17  adv_status   3017 non-null   float64
dtypes: float64(4), int64(4), object(10)
memory usage: 444.9+ KB
```

In [7]:
```
nortel.head()
```

Out[7]:

	Acc_no	REG_No	Reg_date	Comp_date	Month_year	RP	RS	Unit	IR	Perf
0	CST-H514-H3H	NaN	5/14/2004	5/25/2005	2005-05	4 - L	4-C_problem	GPK	Solved	
1	CST-H7H6-HH28	CZK3B39B	6/7/2004	8/6/2005	2005-06	3 - M	3-Minor	GPK	Solved	
2	CST-H9H1-HH86	NaN	1/9/2004	3/5/2005	2005-05	3 - M	4-C_problem	RPLCDS	Solved	
3	CST-1H29-H233	NaN	10/29/2004	5/27/2005	2005-05	2 - H	2-Major	RPLCDS	Solved	
4	CST-11H2-H312	CZK26998	2/11/2004	9/5/2005	2005-05	3 - M	3-Minor	GPK	Solved	

In [8]:
```
nortel['OverallPer'].mean()
```

Out[8]:
```
77.73316471704078
```

In [10]:
```
nortel['Unit'].value_counts()
```

Out[10]:
```
RPLCDS    2156
GPK        993
GPA         11
Not          3
Name: Unit, dtype: int64
```

In [11]:
```
nortel.groupby('Unit')['OverallPer'].mean()
```

Out[11]:
```
Unit
GPA       70.909091
GPK       82.124874
Not       56.666667
RPLCDS    75.774583
Name: OverallPer, dtype: float64
```

In [13]:
```
nortel.groupby(['Unit','IR'])['OverallPer'].mean()
```

Out[13]:
```
Unit    IR
GPA     Solved    70.909091
GPK     No        47.916667
        Solved    82.972136
Not     Solved    56.666667
RPLCDS  No        37.719298
        Solved    77.899119
Name: OverallPer, dtype: float64
```

In [15]:
```
nortel.groupby(['Unit'])['OverallPer'].agg(['min','max','mean','std'])
```

Out[15]:

Unit	min	max	mean	std
GPA	40	90	70.909091	17.002674
GPK	0	100	82.124874	24.740926
Not	0	100	56.666667	51.316014
RPLCDS	0	100	75.774583	29.115391

In [17]:
```
nortel.groupby(['Performance'])['OverallPer'].mean()
```

Out[17]:
```
Performance
Green    98.006198
None     75.872928
Red      35.506494
Name: OverallPer, dtype: float64
```

In [18]:
```
nortel.groupby(['Unit','Performance'])['OverallPer'].mean()
```

Out[18]:
```
Unit    Performance
GPA     Green          90.000000
        None           75.000000
        Red            45.000000
GPK     Green          97.712766
        None           76.679104
        Red            45.802469
Not     Green         100.000000
        None           35.000000
RPLCDS  Green          98.203390
        None           75.601266
        Red            32.682119
Name: OverallPer, dtype: float64
```

In [22]:
```
nortel2=nortel.groupby(['Unit','Performance'])['OverallPer'].mean().reset_index()
```

In [23]:
```
nortel2
```

Out[23]:

	Unit	Performance	OverallPer
0	GPA	Green	90.000000
1	GPA	None	75.000000
2	GPA	Red	45.000000
3	GPK	Green	97.712766
4	GPK	None	76.679104
5	GPK	Red	45.802469
6	Not	Green	100.000000
7	Not	None	35.000000
8	RPLCDS	Green	98.203390
9	RPLCDS	None	75.601266
10	RPLCDS	Red	32.682119

In [24]:
```
nortel2.to_excel('nortel_stat.xlsx')
```

In [44]:
```
nortel3=nortel[(nortel['Unit']=='GPK') & (nortel['Performance']=='Green')]
```

In [45]:
```
nortel3.groupby('Unit')['Performance'].value_counts()
```

Out[45]:
```
Unit  Performance
GPK   Green          376
Name: Performance, dtype: int64
```

In [47]:
```
nortel3.head(5)
```

Out[47]:

	Acc_no	REG_No	Reg_date	Comp_date	Month_year	RP	RS	Unit	IR	Perform
10	AKM-H125-HH62	CZK296B7	1/25/2005	5/24/2005	2005-05	2 - H	3-Minor	GPK	Solved	G
17	AKM-H2H9-H318	CZK3BB17	9/2/2005	1/5/2005	2005-05	3 - M	3-Minor	GPK	Solved	G
23	AKM-H224-HH39	CZK3145B	2/24/2005	5/7/2005	2005-07	2 - H	3-Minor	GPK	Solved	G
28	AKM-H3H1-H3H1	CZK3B723	1/3/2005	5/18/2005	2005-05	3 - M	3-Minor	GPK	Solved	G
41	AKM-H316-H15H	NaN	3/16/2005	6/29/2005	2005-06	4 - L	4-C_problem	GPK	Solved	G

In [48]:
```
nortel_stat=nortel.groupby(['Unit','Performance'])['OverallPer','t_respond','Unde
```

```
<ipython-input-48-764d312d9cd5>:1: FutureWarning: Indexing with multiple keys
(implicitly converted to a tuple of keys) will be deprecated, use a list instea
d.
  nortel_stat=nortel.groupby(['Unit','Performance'])['OverallPer','t_respon
d','Understand'].describe()
```

In [50]:
```
nortel_stat.reset_index()
```

Out[50]:

	Unit	Performance	OverallPer								...	t_
			count	mean	std	min	25%	50%	75%	max	...	
0	GPA	Green	1.0	90.000000	NaN	90.0	90.0	90.0	90.0	90.0	...	
1	GPA	None	8.0	75.000000	11.952286	50.0	70.0	80.0	80.0	90.0	...	9
2	GPA	Red	2.0	45.000000	7.071068	40.0	42.5	45.0	47.5	50.0	...	6
3	GPK	Green	376.0	97.712766	4.205704	90.0	100.0	100.0	100.0	100.0	...	1(
4	GPK	None	536.0	76.679104	23.596642	0.0	70.0	80.0	90.0	100.0	...	1(
5	GPK	Red	81.0	45.802469	33.274021	0.0	20.0	40.0	80.0	100.0	...	9
6	Not	Green	1.0	100.000000	NaN	100.0	100.0	100.0	100.0	100.0	...	1(
7	Not	None	2.0	35.000000	49.497475	0.0	17.5	35.0	52.5	70.0	...	1(
8	RPLCDS	Green	590.0	98.203390	3.842306	90.0	100.0	100.0	100.0	100.0	...	1(
9	RPLCDS	None	1264.0	75.601266	24.364334	0.0	70.0	80.0	90.0	100.0	...	1(
10	RPLCDS	Red	302.0	32.682119	26.459239	0.0	10.0	30.0	40.0	100.0	...	8

11 rows × 26 columns

In [51]:
```
##Overallper ,, std, mean, min
# t_understand,, mean, max
```

In [52]:
```
norel_stat2=nortel.groupby('Unit').agg({'OverallPer':['std','mean','min'],'t_resp
```

In [53]:
```
norel_stat2
```

Out[53]:

	OverallPer			t_respond	
	std	mean	min	mean	max
Unit					
GPA	17.002674	70.909091	40	67.272727	100
GPK	24.740926	82.124874	0	86.203424	100
Not	51.316014	56.666667	0	100.000000	100
RPLCDS	29.115391	75.774583	0	80.997217	100

In [56]:
```
nortel.rename(columns={'Acc_no':'Account_number'},inplace=True)
```

In [57]:
```
nortel
```

Out[57]:

ı_date	Comp_date	Month_year	RP	RS	Unit	IR	Performance	OverallPer	t_respon
4/2004	5/25/2005	2005-05	4 - L	4-C_problem	GPK	Solved	None	60	6
7/2004	8/6/2005	2005-06	3 - M	3-Minor	GPK	Solved	None	80	9
9/2004	3/5/2005	2005-05	3 - M	4-C_problem	RPLCDS	Solved	Red	40	10
9/2004	5/27/2005	2005-05	2 - H	2-Major	RPLCDS	Solved	None	60	9
1/2004	9/5/2005	2005-05	3 - M	3-Minor	GPK	Solved	Red	40	8
...	...	...	...	...	...	...	...	...	.
NaN	NaN	2006-03	2 - H	4-C_problem	RPLCDS	Solved	Green	100	10
NaN	NaN	2006-03	3 - M	4-C_problem	GPK	Solved	Green	100	10
NaN	NaN	2006-03	4 - L	4-C_problem	RPLCDS	Solved	None	90	8
NaN	NaN	2006-03	4 - L	4-C_problem	RPLCDS	Solved	None	80	10
NaN	NaN	2006-03	3 - M	3-Minor	RPLCDS	Solved	None	90	9

In [69]:
```
nortel['total_score']=nortel[['OverallPer','t_respond','t_resolve']].sum(axis=1)
nortel['avg_score']=nortel[['OverallPer','t_respond','t_resolve']].mean(axis=1)
nortel['Std_score']=nortel[['OverallPer','t_respond','t_resolve']].std(axis=1)
nortel['CV']=nortel['Std_score']/nortel['avg_score']*100
```

In [70]:
```
nortel.head(5)
```

Out[70]:

th_year	RP	RS	Unit	IR	Performance	...	Understand	t_resolve	s_equipment	s_D
2005-05	4 - L	4-C_problem	GPK	Solved	None	...	60	60	NaN	Na
2005-06	3 - M	3-Minor	GPK	Solved	None	...	90	90	NaN	80
2005-05	3 - M	4-C_problem	RPLCDS	Solved	Red	...	100	50	NaN	Na
2005-05	2 - H	2-Major	RPLCDS	Solved	None	...	90	80	NaN	Na
2005-05	3 - M	3-Minor	GPK	Solved	Red	...	80	10	NaN	100

In [ ]: