In [1]:
```
### Querying MySQL database tractortek ###
```

In [2]:
```
from sqlalchemy import create_engine
import pandas as pd
```

In [3]:
```
engine = create_engine('mysql+pymysql://root:pw123@localhost/tractortek')

df=pd.read_sql('prod_info', engine)
print(df.head())
```

```
    prod_id                  prod_name prod_manufc   esp_id
0  PROD_001              Gator XUV 590M  John Deere  ESP_001
1  PROD_002                     CUV82         Cat  ESP_002
2  PROD_003   1025R Sub-Compact Tractor  John Deere  ESP_003
3  PROD_004  CT1021 Sub-Compact Tractor      Bobcat  ESP_004
4  PROD_005                   UV34 Gas       Bobcat  ESP_005
```

In [4]:
```
df=pd.read_sql('prod_sales', engine)
print(df.head(20))
```

```
    sales_id   prod_id  emp_id  year week  quantity
0          1  PROD_001  EMP244  2021   W0        23
1          2  PROD_001  EMP244  2021   W1        27
2          3  PROD_001  EMP244  2021   W2        37
3          4  PROD_001  EMP244  2021   W3        47
4          5  PROD_001  EMP244  2021   W4        42
5          6  PROD_001  EMP244  2021   W5        55
6          7  PROD_001  EMP244  2021   W6        59
7          8  PROD_001  EMP244  2021   W7        66
8          9  PROD_001  EMP244  2021   W8        91
9         10  PROD_001  EMP244  2021   W9        70
10        11  PROD_001  EMP244  2021  W10        62
11        12  PROD_001  EMP244  2021  W11        30
12        13  PROD_001  EMP244  2021  W12        47
13        14  PROD_001  EMP244  2021  W13        49
14        15  PROD_001  EMP244  2021  W14        44
15        16  PROD_001  EMP244  2021  W15        74
16        17  PROD_001  EMP244  2021  W16        87
17        18  PROD_001  EMP244  2021  W17       113
```

```
18        19  PROD_001  EMP244  2021  W18        94
19        20  PROD_001  EMP244  2021  W19        87
```

In [5]:

```
df=pd.read_sql('warranty_sales', engine)
print(df.head(10))
```

```
   sales_id   esp_id  emp_id  year week  quantity
0         1  ESP_001  EMP234  2020   W0         3
1         2  ESP_001  EMP234  2020   W1         3
2         3  ESP_001  EMP234  2020   W2         3
3         4  ESP_001  EMP234  2020   W3         6
4         5  ESP_001  EMP234  2020   W4        12
5         6  ESP_001  EMP234  2020   W5         7
6         7  ESP_001  EMP234  2020   W6         6
7         8  ESP_001  EMP234  2020   W7         7
8         9  ESP_001  EMP234  2020   W8         5
9        10  ESP_001  EMP234  2020   W9         3
C:\Users\mdbro\anaconda3\lib\site-packages\sqlalchemy\dialects\mysql\reflection.py:62: SAWarning: Unknown schema content:
'  KEY `esp_id` (`esp_id`) /*!80000 INVISIBLE */,'
  util.warn("Unknown schema content: %r" % line)
```

In [6]:

```
df=pd.read_sql('warranty_prices', engine)
print(df.head(8))
```

```
   id   esp_id  price_2020  price_2021
0   1  ESP_006         372         450
1   2  ESP_005         843         989
2   3  ESP_004         843         989
3   4  ESP_007        1027        1100
4   5  ESP_002         843         989
5   6  ESP_001         843         989
6   7  ESP_008         843         989
7   8  ESP_003         978        1010
```

In [7]:

```
df=pd.read_sql('employees', engine)
print(df.head())
```

```
   emp_id              name paygrade region
0  EMP234    Bachmann, Jane      C13     NW
1  EMP244       Evans, Gina      C12     NW
2  EMP256     Lawson, Harry      C11     NW
3  EMP267  Clement, Beverly      C14     SW
4  EMP290      Allen, Maude      C12     SW
```

In [8]:
```
df=pd.read_sql('prod_prices',engine)
print(df.head(25))
```

```
    id   prod_id quarter  year  price
0    1  PROD_006      Q1  2020   4575
1    2  PROD_006      Q2  2020   4575
2    3  PROD_006      Q3  2020   4575
3    4  PROD_006      Q4  2020   4669
4    5  PROD_006      Q1  2021   4949
5    6  PROD_006      Q2  2021   4949
6    7  PROD_006      Q3  2021   4949
7    8  PROD_006      Q4  2021   5100
8    9  PROD_005      Q1  2020  13995
9   10  PROD_005      Q2  2020  13995
10  11  PROD_005      Q3  2020  13995
11  12  PROD_005      Q4  2020  14550
12  13  PROD_005      Q1  2021  16525
13  14  PROD_005      Q2  2021  16525
14  15  PROD_005      Q3  2021  16525
15  16  PROD_005      Q4  2021  17010
16  17  PROD_004      Q1  2020  11385
17  18  PROD_004      Q2  2020  11385
18  19  PROD_004      Q3  2020  11385
19  20  PROD_004      Q4  2020  11999
20  21  PROD_004      Q1  2021  12220
21  22  PROD_004      Q2  2021  12220
22  23  PROD_004      Q3  2021  12220
23  24  PROD_004      Q4  2021  12999
24  25  PROD_007      Q1  2020  22987
```

In [9]:
```
data = pd.read_sql('prod_sales', engine)
print(data.head())
```

```
   sales_id   prod_id  emp_id  year week  quantity
0         1  PROD_001  EMP244  2021   W0        23
1         2  PROD_001  EMP244  2021   W1        27
2         3  PROD_001  EMP244  2021   W2        37
3         4  PROD_001  EMP244  2021   W3        47
4         5  PROD_001  EMP244  2021   W4        42
```

In [29]:
```
data = pd.read_sql_query('SELECT emp_id, quantity FROM prod_sales', engine)
data.sort_values(by='emp_id').head(35)
```

Out[29]:

	emp_id	quantity
4162	EMP 234	15
4161	EMP 234	15
4160	EMP 234	15
1738	EMP234	87
1737	EMP234	63
1736	EMP234	58
1735	EMP234	77
1734	EMP234	98
1733	EMP234	57
1732	EMP234	72
1731	EMP234	73
1730	EMP234	37
1729	EMP234	44
1739	EMP234	59
1728	EMP234	43
1726	EMP234	71
1725	EMP234	48
1724	EMP234	56
1723	EMP234	78
1722	EMP234	61
1721	EMP234	53
1720	EMP234	38
1719	EMP234	48
1718	EMP234	29
1717	EMP234	33

	emp_id	quantity
1727	EMP234	43
1716	EMP234	27
1740	EMP234	49
1742	EMP234	31
1764	EMP234	22
1763	EMP234	27
1762	EMP234	16
1761	EMP234	23
1760	EMP234	19
1759	EMP234	17

In [30]:
```
byEmp = data.groupby('emp_id')
```

In [31]:
```
## mean(avg) number of products sold grouped by employee id ##
byEmp.mean()
```

Out[31]:

emp_id	quantity
EMP 234	15.000000
EMP234	31.229567
EMP244	41.234375
EMP256	20.453125
EMP267	42.419471
EMP290	13.959135

In [32]:
```
byEmp.sum()
```

Out[32]:

emp_id	quantity
EMP 234	45
EMP234	25983
EMP244	34307
EMP256	17017
EMP267	35293
EMP290	11614

In [24]:

```
warrdata = pd.read_sql_query('SELECT emp_id, quantity FROM warranty_sales',engine)
print(warrdata.head(10))
```

```
   emp_id  quantity
0  EMP234         3
1  EMP234         3
2  EMP234         3
3  EMP234         6
4  EMP234        12
5  EMP234         7
6  EMP234         6
7  EMP234         7
8  EMP234         5
9  EMP234         3
```

In [25]:

```
warbyEmp = warrdata.groupby('emp_id')
```

In [26]:

```
## mean(avg) number of warranties sold grouped by employee ID ##
warbyEmp.mean()
```

Out[26]:

emp_id	quantity
EMP234	3.558894
EMP244	8.167067

emp_id	quantity
EMP256	1.841346
EMP267	6.913462
EMP290	0.521635

In [27]:
```
## sum of warranties sold per employee ##
warbyEmp.sum()
```

Out[27]:

emp_id	quantity
EMP234	2961
EMP244	6795
EMP256	1532
EMP267	5752
EMP290	434

In [35]:
```
df = pd.read_sql_query('SELECT emp_id, esp_id, quantity FROM warranty_sales',engine)
print(df.head(25))
```

```
    emp_id   esp_id  quantity
0   EMP234  ESP_001         3
1   EMP234  ESP_001         3
2   EMP234  ESP_001         3
3   EMP234  ESP_001         6
4   EMP234  ESP_001        12
5   EMP234  ESP_001         7
6   EMP234  ESP_001         6
7   EMP234  ESP_001         7
8   EMP234  ESP_001         5
9   EMP234  ESP_001         3
10  EMP234  ESP_001         1
11  EMP234  ESP_001         1
12  EMP234  ESP_001         0
```

```
13  EMP234  ESP_001         1
14  EMP234  ESP_001         2
15  EMP234  ESP_001         4
16  EMP234  ESP_001         8
17  EMP234  ESP_001         6
18  EMP234  ESP_001         7
19  EMP234  ESP_001         5
20  EMP234  ESP_001         7
21  EMP234  ESP_001         9
22  EMP234  ESP_001        11
23  EMP234  ESP_001        11
24  EMP234  ESP_001         8
```

In [36]:
```
## grouping by EMP ID, summing warranty sales quantities and only returning EMP ID that = emp234 ##
df.groupby('emp_id').sum().loc['EMP234']
```

Out[36]:
```
quantity    2961
Name: EMP234, dtype: int64
```

In [37]:
```
df.groupby('emp_id').sum().loc['EMP244']
```

Out[37]:
```
quantity    6795
Name: EMP244, dtype: int64
```

In [40]:
```
## max number of warranty sales per Extended Service Plan ##
df = pd.read_sql_query('SELECT esp_id, quantity FROM warranty_sales', engine)
df.groupby('esp_id').max()
```

Out[40]:

esp_id	quantity
ESP_001	30
ESP_002	31
ESP_003	24
ESP_004	29
ESP_005	33
ESP_006	46
ESP_007	28

esp_id	quantity
ESP_008	31

In [41]:
```
df = pd.read_sql_query('SELECT prod_id, quantity FROM prod_sales', engine)
df.groupby('prod_id').sum()
```

Out[41]:

prod_id	quantity
PROD_001	20071
PROD_002	17446
PROD_003	17037
PROD_004	16315
PROD_005	14466
PROD_006	13630
PROD_007	13319
PROD_008	11975

In [42]:
```
## AVG number of products sold per Product ID ##
df.groupby('prod_id').mean()
```

Out[42]:

prod_id	quantity
PROD_001	38.598077
PROD_002	33.357553
PROD_003	32.763462
PROD_004	31.375000
PROD_005	27.819231

prod_id	quantity
PROD_006	26.211538
PROD_007	25.613462
PROD_008	23.028846

In [43]:
```
# max number of products sold per Product ID ##
df.groupby('prod_id').max()
```

Out[43]:

prod_id	quantity
PROD_001	124
PROD_002	102
PROD_003	106
PROD_004	103
PROD_005	97
PROD_006	103
PROD_007	98
PROD_008	87

In [44]:
```
## least amount of products sold per Product ID ##

df.groupby('prod_id').min()
```

Out[44]:

prod_id	quantity
PROD_001	5
PROD_002	3

	quantity
prod_id	
PROD_003	2
PROD_004	3
PROD_005	3
PROD_006	1
PROD_007	1
PROD_008	2

In [45]:
```
df.groupby('prod_id').describe()
```

Out[45]:

	quantity							
	count	mean	std	min	25%	50%	75%	max
prod_id								
PROD_001	520.0	38.598077	21.521732	5.0	23.00	33.0	51.0	124.0
PROD_002	523.0	33.357553	20.638846	3.0	17.00	28.0	46.5	102.0
PROD_003	520.0	32.763462	20.481483	2.0	16.00	28.0	45.0	106.0
PROD_004	520.0	31.375000	19.463412	3.0	16.00	27.0	44.0	103.0
PROD_005	520.0	27.819231	18.948516	3.0	14.00	23.0	38.0	97.0
PROD_006	520.0	26.211538	18.515875	1.0	12.00	22.0	35.0	103.0
PROD_007	520.0	25.613462	18.794052	1.0	11.75	21.0	34.0	98.0
PROD_008	520.0	23.028846	16.520937	2.0	10.00	18.0	31.0	87.0

Concatenating tables

In [46]:
```
df1 = pd.read_sql('prod_sales', engine)
```

In [51]:

```
df2 = pd.read_sql('prod_prices', engine)
```

In [49]:
```
df3 = pd.read_sql('warranty_sales', engine)
```

In [50]:
```
df4 = pd.read_sql('warranty_prices', engine)
```

In [53]:
```
pd.concat([df2,df4])
```

Out[53]:

	id	prod_id	quarter	year	price	esp_id	price_2020	price_2021
0	1	PROD_006	Q1	2020.0	4575.0	NaN	NaN	NaN
1	2	PROD_006	Q2	2020.0	4575.0	NaN	NaN	NaN
2	3	PROD_006	Q3	2020.0	4575.0	NaN	NaN	NaN
3	4	PROD_006	Q4	2020.0	4669.0	NaN	NaN	NaN
4	5	PROD_006	Q1	2021.0	4949.0	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...
3	4	NaN	NaN	NaN	NaN	ESP_007	1027.0	1100.0
4	5	NaN	NaN	NaN	NaN	ESP_002	843.0	989.0
5	6	NaN	NaN	NaN	NaN	ESP_001	843.0	989.0
6	7	NaN	NaN	NaN	NaN	ESP_008	843.0	989.0
7	8	NaN	NaN	NaN	NaN	ESP_003	978.0	1010.0

72 rows × 8 columns

In [54]:
```
## merging product sales table and warranty sales table on emp_id column ##

pd.merge(df1, df3, on=['emp_id'])
```

Out[54]:

	sales_id_x	prod_id	emp_id	year_x	week_x	quantity_x	sales_id_y	esp_id	year_y	week_y	quantity_y
0	1	PROD_001	EMP244	2021	W0	23	105	ESP_001	2020	W0	6

	sales_id_x	prod_id	emp_id	year_x	week_x	quantity_x	sales_id_y	esp_id	year_y	week_y	quantity_y
1	1	PROD_001	EMP244	2021	W0	23	106	ESP_001	2020	W1	7
2	1	PROD_001	EMP244	2021	W0	23	107	ESP_001	2020	W2	9
3	1	PROD_001	EMP244	2021	W0	23	108	ESP_001	2020	W3	9
4	1	PROD_001	EMP244	2021	W0	23	109	ESP_001	2020	W4	13
...	...	...	...	...	...	...	...	...	...	...	...
3461115	4160	PROD_008	EMP290	2020	W51	3	4156	ESP_008	2021	W47	0
3461116	4160	PROD_008	EMP290	2020	W51	3	4157	ESP_008	2021	W48	0
3461117	4160	PROD_008	EMP290	2020	W51	3	4158	ESP_008	2021	W49	0
3461118	4160	PROD_008	EMP290	2020	W51	3	4159	ESP_008	2021	W50	0
3461119	4160	PROD_008	EMP290	2020	W51	3	4160	ESP_008	2021	W51	0

3461120 rows × 11 columns