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

In [3]:
```
df=pd.read_csv("/home/rmdstic/Documents/TEA-14/Uber Request Data.csv
```

In [4]:
```
df
```

Out[4]:

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
0	619	Airport	1.0	Trip Completed	11/7/2016 11:51	11/7/2016 13:00
1	867	Airport	1.0	Trip Completed	11/7/2016 17:57	11/7/2016 18:47
2	1807	City	1.0	Trip Completed	12/7/2016 9:17	12/7/2016 9:58
3	2532	Airport	1.0	Trip Completed	12/7/2016 21:08	12/7/2016 22:03
4	3112	City	1.0	Trip Completed	13-07-2016 08:33:16	13-07-2016 09:25:47
5	3879	Airport	1.0	Trip Completed	13-07-2016 21:57:28	13-07-2016 22:28:59
6	4270	Airport	1.0	Trip Completed	14-07-2016 06:15:32	14-07-2016 07:13:15
7	5510	Airport	1.0	Trip Completed	15-07-2016 05:11:52	15-07-2016 06:07:52
8	6248	City	1.0	Trip Completed	15-07-2016 17:57:27	15-07-2016 18:50:51
9	267	City	2.0	Trip Completed	11/7/2016 6:46	11/7/2016 7:25
10	1467	Airport	2.0	Trip Completed	12/7/2016 5:08	12/7/2016 6:02
11	1983	City	2.0	Trip Completed	12/7/2016 12:30	12/7/2016 12:57
12	2784	Airport	2.0	Trip Completed	13-07-2016 04:49:20	13-07-2016 05:23:03
13	3075	City	2.0	Trip Completed	13-07-2016 08:02:53	13-07-2016 09:16:19
14	3379	City	2.0	Trip Completed	13-07-2016 14:23:02	13-07-2016 15:35:18
15	3482	Airport	2.0	Trip Completed	13-07-2016 17:23:18	13-07-2016 18:20:51
16	4652	City	2.0	Trip Completed	14-07-2016 12:01:02	14-07-2016 12:36:46
17	5335	Airport	2.0	Trip Completed	14-07-2016 22:24:13	14-07-2016 23:18:52
18	535	Airport	3.0	Trip Completed	11/7/2016 10:00	11/7/2016 10:31
19	960	Airport	3.0	Trip Completed	11/7/2016 18:45	11/7/2016 19:23
20	1934	Airport	3.0	Trip Completed	12/7/2016 11:17	12/7/2016 12:23
21	2083	Airport	3.0	Trip Completed	12/7/2016 15:46	12/7/2016 16:40
22	2211	Airport	3.0	Trip Completed	12/7/2016 18:00	12/7/2016 18:28
23	3096	Airport	3.0	Trip Completed	13-07-2016 08:17:29	13-07-2016 09:22:37
24	3881	Airport	3.0	Trip Completed	13-07-2016 21:54:18	13-07-2016 22:51:23
25	5254	City	3.0	Trip Completed	14-07-2016 21:23:03	14-07-2016 22:25:19

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
26	5434	City	3.0	Trip Completed	15-07-2016 02:41:38	15-07-2016 03:24:43
27	5916	City	3.0	Trip Completed	15-07-2016 10:00:43	15-07-2016 10:53:06
28	669	City	4.0	Trip Completed	11/7/2016 13:08	11/7/2016 13:49
29	1567	Airport	4.0	Trip Completed	12/7/2016 6:21	12/7/2016 7:10
...	...	...	...	...	...	...
6715	6683	City	NaN	No Cars Available	15-07-2016 22:34:01	NaN
6716	6686	Airport	NaN	No Cars Available	15-07-2016 22:36:13	NaN
6717	6688	Airport	NaN	No Cars Available	15-07-2016 22:37:37	NaN
6718	6689	Airport	NaN	No Cars Available	15-07-2016 22:43:46	NaN
6719	6693	City	NaN	No Cars Available	15-07-2016 22:49:33	NaN
6720	6696	City	NaN	No Cars Available	15-07-2016 22:49:45	NaN
6721	6697	Airport	NaN	No Cars Available	15-07-2016 22:51:18	NaN
6722	6709	Airport	NaN	No Cars Available	15-07-2016 22:56:00	NaN
6723	6706	Airport	NaN	No Cars Available	15-07-2016 22:58:15	NaN
6724	6708	City	NaN	No Cars Available	15-07-2016 23:03:23	NaN
6725	6713	City	NaN	No Cars Available	15-07-2016 23:08:14	NaN
6726	6715	City	NaN	No Cars Available	15-07-2016 23:11:41	NaN
6727	6716	City	NaN	No Cars Available	15-07-2016 23:14:36	NaN
6728	6718	Airport	NaN	No Cars Available	15-07-2016 23:14:39	NaN
6729	6720	City	NaN	No Cars Available	15-07-2016 23:16:48	NaN
6730	6722	Airport	NaN	No Cars Available	15-07-2016 23:18:21	NaN
6731	6725	Airport	NaN	No Cars Available	15-07-2016 23:21:53	NaN
6732	6728	City	NaN	No Cars Available	15-07-2016 23:26:50	NaN
6733	6730	Airport	NaN	No Cars Available	15-07-2016 23:27:55	NaN

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
6734	6732	Airport	NaN	No Cars Available	15-07-2016 23:35:50	NaN
6735	6737	Airport	NaN	No Cars Available	15-07-2016 23:39:15	NaN
6736	6744	Airport	NaN	No Cars Available	15-07-2016 23:42:51	NaN
6737	6740	City	NaN	No Cars Available	15-07-2016 23:43:54	NaN
6738	6746	City	NaN	No Cars Available	15-07-2016 23:46:03	NaN
6739	6739	City	NaN	No Cars Available	15-07-2016 23:46:20	NaN
6740	6745	City	NaN	No Cars Available	15-07-2016 23:49:03	NaN
6741	6752	Airport	NaN	No Cars Available	15-07-2016 23:50:05	NaN
6742	6751	City	NaN	No Cars Available	15-07-2016 23:52:06	NaN
6743	6754	City	NaN	No Cars Available	15-07-2016 23:54:39	NaN
6744	6753	Airport	NaN	No Cars Available	15-07-2016 23:55:03	NaN

6745 rows × 6 columns

In [5]:
```
df.head()
```

Out[5]:

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
0	619	Airport	1.0	Trip Completed	11/7/2016 11:51	11/7/2016 13:00
1	867	Airport	1.0	Trip Completed	11/7/2016 17:57	11/7/2016 18:47
2	1807	City	1.0	Trip Completed	12/7/2016 9:17	12/7/2016 9:58
3	2532	Airport	1.0	Trip Completed	12/7/2016 21:08	12/7/2016 22:03
4	3112	City	1.0	Trip Completed	13-07-2016 08:33:16	13-07-2016 09:25:47

In [6]:
```
df.head(10)
```

Out[6]:

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
0	619	Airport	1.0	Trip Completed	11/7/2016 11:51	11/7/2016 13:00
1	867	Airport	1.0	Trip Completed	11/7/2016 17:57	11/7/2016 18:47
2	1807	City	1.0	Trip Completed	12/7/2016 9:17	12/7/2016 9:58
3	2532	Airport	1.0	Trip Completed	12/7/2016 21:08	12/7/2016 22:03
4	3112	City	1.0	Trip Completed	13-07-2016 08:33:16	13-07-2016 09:25:47
5	3879	Airport	1.0	Trip Completed	13-07-2016 21:57:28	13-07-2016 22:28:59
6	4270	Airport	1.0	Trip Completed	14-07-2016 06:15:32	14-07-2016 07:13:15
7	5510	Airport	1.0	Trip Completed	15-07-2016 05:11:52	15-07-2016 06:07:52
8	6248	City	1.0	Trip Completed	15-07-2016 17:57:27	15-07-2016 18:50:51
9	267	City	2.0	Trip Completed	11/7/2016 6:46	11/7/2016 7:25

In [7]:
```
df.tail()
```

Out[7]:

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
6740	6745	City	NaN	No Cars Available	15-07-2016 23:49:03	NaN
6741	6752	Airport	NaN	No Cars Available	15-07-2016 23:50:05	NaN
6742	6751	City	NaN	No Cars Available	15-07-2016 23:52:06	NaN
6743	6754	City	NaN	No Cars Available	15-07-2016 23:54:39	NaN
6744	6753	Airport	NaN	No Cars Available	15-07-2016 23:55:03	NaN

In [8]:
```
df.tail(10)
```

Out[8]:

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
6735	6737	Airport	NaN	No Cars Available	15-07-2016 23:39:15	NaN
6736	6744	Airport	NaN	No Cars Available	15-07-2016 23:42:51	NaN
6737	6740	City	NaN	No Cars Available	15-07-2016 23:43:54	NaN
6738	6746	City	NaN	No Cars Available	15-07-2016 23:46:03	NaN
6739	6739	City	NaN	No Cars Available	15-07-2016 23:46:20	NaN
6740	6745	City	NaN	No Cars Available	15-07-2016 23:49:03	NaN
6741	6752	Airport	NaN	No Cars Available	15-07-2016 23:50:05	NaN
6742	6751	City	NaN	No Cars Available	15-07-2016 23:52:06	NaN
6743	6754	City	NaN	No Cars Available	15-07-2016 23:54:39	NaN
6744	6753	Airport	NaN	No Cars Available	15-07-2016 23:55:03	NaN

In [9]:
```
df.index
```

Out[9]:
```
RangeIndex(start=0, stop=6745, step=1)
```

In [10]:
```
df.columns
```

Out[10]:
```
Index(['Request id', 'Pickup point', 'Driver id', 'Status',
       'Request timestamp', 'Drop timestamp'],
      dtype='object')
```

In [11]:
```
df.columns.values
```

Out[11]:
```
array(['Request id', 'Pickup point', 'Driver id', 'Status',
       'Request timestamp', 'Drop timestamp'], dtype=object)
```

In [12]:
```
df.shape
```

Out[12]:
```
(6745, 6)
```

In [13]:
```
df.dtypes
```

Out[13]:
```
Request id             int64
Pickup point          object
Driver id            float64
Status                object
Request timestamp     object
Drop timestamp        object
dtype: object
```

In [14]:
```
df.describe()
```

Out[14]:

	Request id	Driver id
count	6745.000000	4095.000000
mean	3384.644922	149.501343
std	1955.099667	86.051994
min	1.000000	1.000000
25%	1691.000000	75.000000
50%	3387.000000	149.000000
75%	5080.000000	224.000000
max	6766.000000	300.000000

In [15]:
```
df.isnull()
```

Out[15]:

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
5	False	False	False	False	False	False
6	False	False	False	False	False	False
7	False	False	False	False	False	False
8	False	False	False	False	False	False
9	False	False	False	False	False	False
10	False	False	False	False	False	False
11	False	False	False	False	False	False
12	False	False	False	False	False	False
13	False	False	False	False	False	False
14	False	False	False	False	False	False
15	False	False	False	False	False	False
16	False	False	False	False	False	False
17	False	False	False	False	False	False
18	False	False	False	False	False	False
19	False	False	False	False	False	False
20	False	False	False	False	False	False
21	False	False	False	False	False	False
22	False	False	False	False	False	False
23	False	False	False	False	False	False
24	False	False	False	False	False	False
25	False	False	False	False	False	False
26	False	False	False	False	False	False
27	False	False	False	False	False	False
28	False	False	False	False	False	False
29	False	False	False	False	False	False
...	...	...	...	...	...	...
6715	False	False	True	False	False	True
6716	False	False	True	False	False	True
6717	False	False	True	False	False	True
6718	False	False	True	False	False	True

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
6719	False	False	True	False	False	True
6720	False	False	True	False	False	True
6721	False	False	True	False	False	True
6722	False	False	True	False	False	True
6723	False	False	True	False	False	True
6724	False	False	True	False	False	True
6725	False	False	True	False	False	True
6726	False	False	True	False	False	True
6727	False	False	True	False	False	True
6728	False	False	True	False	False	True
6729	False	False	True	False	False	True
6730	False	False	True	False	False	True
6731	False	False	True	False	False	True
6732	False	False	True	False	False	True
6733	False	False	True	False	False	True
6734	False	False	True	False	False	True
6735	False	False	True	False	False	True
6736	False	False	True	False	False	True
6737	False	False	True	False	False	True
6738	False	False	True	False	False	True
6739	False	False	True	False	False	True
6740	False	False	True	False	False	True
6741	False	False	True	False	False	True
6742	False	False	True	False	False	True
6743	False	False	True	False	False	True
6744	False	False	True	False	False	True

6745 rows × 6 columns

In [16]:
```
df.notnull()
```

Out[16]:

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
0	True	True	True	True	True	True
1	True	True	True	True	True	True
2	True	True	True	True	True	True
3	True	True	True	True	True	True
4	True	True	True	True	True	True
5	True	True	True	True	True	True
6	True	True	True	True	True	True
7	True	True	True	True	True	True
8	True	True	True	True	True	True
9	True	True	True	True	True	True
10	True	True	True	True	True	True
11	True	True	True	True	True	True
12	True	True	True	True	True	True
13	True	True	True	True	True	True
14	True	True	True	True	True	True
15	True	True	True	True	True	True
16	True	True	True	True	True	True
17	True	True	True	True	True	True
18	True	True	True	True	True	True
19	True	True	True	True	True	True
20	True	True	True	True	True	True
21	True	True	True	True	True	True
22	True	True	True	True	True	True
23	True	True	True	True	True	True
24	True	True	True	True	True	True
25	True	True	True	True	True	True
26	True	True	True	True	True	True
27	True	True	True	True	True	True
28	True	True	True	True	True	True
29	True	True	True	True	True	True
...	...	...	...	...	...	...
6715	True	True	False	True	True	False
6716	True	True	False	True	True	False
6717	True	True	False	True	True	False
6718	True	True	False	True	True	False

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
6719	True	True	False	True	True	False
6720	True	True	False	True	True	False
6721	True	True	False	True	True	False
6722	True	True	False	True	True	False
6723	True	True	False	True	True	False
6724	True	True	False	True	True	False
6725	True	True	False	True	True	False
6726	True	True	False	True	True	False
6727	True	True	False	True	True	False
6728	True	True	False	True	True	False
6729	True	True	False	True	True	False
6730	True	True	False	True	True	False
6731	True	True	False	True	True	False
6732	True	True	False	True	True	False
6733	True	True	False	True	True	False
6734	True	True	False	True	True	False
6735	True	True	False	True	True	False
6736	True	True	False	True	True	False
6737	True	True	False	True	True	False
6738	True	True	False	True	True	False
6739	True	True	False	True	True	False
6740	True	True	False	True	True	False
6741	True	True	False	True	True	False
6742	True	True	False	True	True	False
6743	True	True	False	True	True	False
6744	True	True	False	True	True	False

6745 rows × 6 columns

In [17]:
```
df.isna()
```

Out[17]:

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
5	False	False	False	False	False	False
6	False	False	False	False	False	False
7	False	False	False	False	False	False
8	False	False	False	False	False	False
9	False	False	False	False	False	False
10	False	False	False	False	False	False
11	False	False	False	False	False	False
12	False	False	False	False	False	False
13	False	False	False	False	False	False
14	False	False	False	False	False	False
15	False	False	False	False	False	False
16	False	False	False	False	False	False
17	False	False	False	False	False	False
18	False	False	False	False	False	False
19	False	False	False	False	False	False
20	False	False	False	False	False	False
21	False	False	False	False	False	False
22	False	False	False	False	False	False
23	False	False	False	False	False	False
24	False	False	False	False	False	False
25	False	False	False	False	False	False
26	False	False	False	False	False	False
27	False	False	False	False	False	False
28	False	False	False	False	False	False
29	False	False	False	False	False	False
...	...	...	...	...	...	...
6715	False	False	True	False	False	True
6716	False	False	True	False	False	True
6717	False	False	True	False	False	True
6718	False	False	True	False	False	True

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
6719	False	False	True	False	False	True
6720	False	False	True	False	False	True
6721	False	False	True	False	False	True
6722	False	False	True	False	False	True
6723	False	False	True	False	False	True
6724	False	False	True	False	False	True
6725	False	False	True	False	False	True
6726	False	False	True	False	False	True
6727	False	False	True	False	False	True
6728	False	False	True	False	False	True
6729	False	False	True	False	False	True
6730	False	False	True	False	False	True
6731	False	False	True	False	False	True
6732	False	False	True	False	False	True
6733	False	False	True	False	False	True
6734	False	False	True	False	False	True
6735	False	False	True	False	False	True
6736	False	False	True	False	False	True
6737	False	False	True	False	False	True
6738	False	False	True	False	False	True
6739	False	False	True	False	False	True
6740	False	False	True	False	False	True
6741	False	False	True	False	False	True
6742	False	False	True	False	False	True
6743	False	False	True	False	False	True
6744	False	False	True	False	False	True

6745 rows × 6 columns

In [18]:
```
df.notna()
```

Out[18]:

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
0	True	True	True	True	True	True
1	True	True	True	True	True	True
2	True	True	True	True	True	True
3	True	True	True	True	True	True
4	True	True	True	True	True	True
5	True	True	True	True	True	True
6	True	True	True	True	True	True
7	True	True	True	True	True	True
8	True	True	True	True	True	True
9	True	True	True	True	True	True
10	True	True	True	True	True	True
11	True	True	True	True	True	True
12	True	True	True	True	True	True
13	True	True	True	True	True	True
14	True	True	True	True	True	True
15	True	True	True	True	True	True
16	True	True	True	True	True	True
17	True	True	True	True	True	True
18	True	True	True	True	True	True
19	True	True	True	True	True	True
20	True	True	True	True	True	True
21	True	True	True	True	True	True
22	True	True	True	True	True	True
23	True	True	True	True	True	True
24	True	True	True	True	True	True
25	True	True	True	True	True	True
26	True	True	True	True	True	True
27	True	True	True	True	True	True
28	True	True	True	True	True	True
29	True	True	True	True	True	True
...	...	...	...	...	...	...
6715	True	True	False	True	True	False
6716	True	True	False	True	True	False
6717	True	True	False	True	True	False
6718	True	True	False	True	True	False

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
6719	True	True	False	True	True	False
6720	True	True	False	True	True	False
6721	True	True	False	True	True	False
6722	True	True	False	True	True	False
6723	True	True	False	True	True	False
6724	True	True	False	True	True	False
6725	True	True	False	True	True	False
6726	True	True	False	True	True	False
6727	True	True	False	True	True	False
6728	True	True	False	True	True	False
6729	True	True	False	True	True	False
6730	True	True	False	True	True	False
6731	True	True	False	True	True	False
6732	True	True	False	True	True	False
6733	True	True	False	True	True	False
6734	True	True	False	True	True	False
6735	True	True	False	True	True	False
6736	True	True	False	True	True	False
6737	True	True	False	True	True	False
6738	True	True	False	True	True	False
6739	True	True	False	True	True	False
6740	True	True	False	True	True	False
6741	True	True	False	True	True	False
6742	True	True	False	True	True	False
6743	True	True	False	True	True	False
6744	True	True	False	True	True	False

6745 rows × 6 columns

In [19]:
```
df.isnull().sum()
```

Out[19]:
```
Request id             0
Pickup point           0
Driver id           2650
Status                 0
Request timestamp      0
Drop timestamp      3914
dtype: int64
```

In [20]:
```
df.isnull().any()
```

Out[20]:
```
Request id           False
Pickup point         False
Driver id             True
Status               False
Request timestamp    False
Drop timestamp        True
dtype: bool
```

In [21]:
```
df.iloc[3]
```

Out[21]:
```
Request id                    2532
Pickup point               Airport
Driver id                        1
Status               Trip Completed
Request timestamp    12/7/2016 21:08
Drop timestamp       12/7/2016 22:03
Name: 3, dtype: object
```

In [22]:
```
df[0:3]
```

Out[22]:

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
0	619	Airport	1.0	Trip Completed	11/7/2016 11:51	11/7/2016 13:00
1	867	Airport	1.0	Trip Completed	11/7/2016 17:57	11/7/2016 18:47
2	1807	City	1.0	Trip Completed	12/7/2016 9:17	12/7/2016 9:58

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

Out[23]:

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
count	6745.000000	6745	4095.000000	6745	6745	2831
unique	NaN	2	NaN	3	5618	2598
top	NaN	City	NaN	Trip Completed	11/7/2016 8:37	12/7/2016 6:22
freq	NaN	3507	NaN	2831	6	4
mean	3384.644922	NaN	149.501343	NaN	NaN	NaN
std	1955.099667	NaN	86.051994	NaN	NaN	NaN
min	1.000000	NaN	1.000000	NaN	NaN	NaN
25%	1691.000000	NaN	75.000000	NaN	NaN	NaN
50%	3387.000000	NaN	149.000000	NaN	NaN	NaN
75%	5080.000000	NaN	224.000000	NaN	NaN	NaN
max	6766.000000	NaN	300.000000	NaN	NaN	NaN

In [24]:
```
df.isna().sum()
```

Out[24]:
```
Request id              0
Pickup point            0
Driver id            2650
Status                  0
Request timestamp       0
Drop timestamp       3914
dtype: int64
```

In [25]:
```
df.isna().sum().sum()
```

Out[25]:
```
6564
```

In [26]:
```
df["Request id"].astype("float64")
```

Out[26]:
```
0        619.0
1        867.0
2       1807.0
3       2532.0
4       3112.0
5       3879.0
6       4270.0
7       5510.0
8       6248.0
9        267.0
10      1467.0
11      1983.0
12      2784.0
13      3075.0
14      3379.0
15      3482.0
16      4652.0
17      5335.0
18       535.0
19       960.0
20      1934.0
21      2083.0
22      2211.0
23      3096.0
24      3881.0
25      5254.0
26      5434.0
27      5916.0
28       669.0
29      1567.0
         ...
6715    6683.0
6716    6686.0
6717    6688.0
6718    6689.0
6719    6693.0
6720    6696.0
6721    6697.0
6722    6709.0
6723    6706.0
6724    6708.0
6725    6713.0
6726    6715.0
6727    6716.0
6728    6718.0
6729    6720.0
6730    6722.0
6731    6725.0
6732    6728.0
6733    6730.0
6734    6732.0
6735    6737.0
6736    6744.0
6737    6740.0
6738    6746.0
6739    6739.0
6740    6745.0
```

```
6741    6752.0
6742    6751.0
6743    6754.0
6744    6753.0
Name: Request id, Length: 6745, dtype: float64
```

In [27]:
```
df.sort_values(by = "Request id")
```

Out[27]:

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
2700	1	Airport	285.0	Trip Completed	11/7/2016 0:20	11/7/2016 0:51
4098	2	Airport	NaN	No Cars Available	11/7/2016 0:23	NaN
776	3	Airport	80.0	Trip Completed	11/7/2016 0:24	11/7/2016 1:31
4101	4	City	NaN	No Cars Available	11/7/2016 0:37	NaN
2506	5	Airport	264.0	Trip Completed	11/7/2016 0:36	11/7/2016 1:35
4100	6	City	NaN	No Cars Available	11/7/2016 0:36	NaN
4099	7	Airport	NaN	No Cars Available	11/7/2016 0:30	NaN
4102	8	City	NaN	No Cars Available	11/7/2016 0:40	NaN
2245	9	Airport	235.0	Trip Completed	11/7/2016 0:45	11/7/2016 2:00
2173	10	City	228.0	Trip Completed	11/7/2016 0:54	11/7/2016 1:59
1884	11	City	198.0	Trip Completed	11/7/2016 1:00	11/7/2016 1:53
4103	12	City	NaN	No Cars Available	11/7/2016 1:08	NaN
1144	13	City	119.0	Trip Completed	11/7/2016 1:08	11/7/2016 1:58
4105	14	Airport	NaN	No Cars Available	11/7/2016 1:10	NaN
4107	15	City	NaN	No Cars Available	11/7/2016 1:15	NaN
4104	16	Airport	NaN	No Cars Available	11/7/2016 1:09	NaN
4108	17	Airport	NaN	No Cars Available	11/7/2016 1:16	NaN
3669	18	City	201.0	Cancelled	11/7/2016 1:16	NaN
4106	19	City	NaN	No Cars Available	11/7/2016 1:14	NaN
4109	20	Airport	NaN	No Cars Available	11/7/2016 1:23	NaN
1839	21	Airport	194.0	Trip Completed	11/7/2016 1:29	11/7/2016 2:05
4111	22	Airport	NaN	No Cars Available	11/7/2016 1:40	NaN
4110	23	City	NaN	No Cars Available	11/7/2016 1:38	NaN
4113	24	Airport	NaN	No Cars Available	11/7/2016 1:48	NaN
4112	25	Airport	NaN	No Cars Available	11/7/2016 1:45	NaN

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
4115	26	Airport	NaN	No Cars Available	11/7/2016 1:49	NaN
4114	27	Airport	NaN	No Cars Available	11/7/2016 1:49	NaN
4117	28	City	NaN	No Cars Available	11/7/2016 2:02	NaN
4116	29	City	NaN	No Cars Available	11/7/2016 1:55	NaN
2797	30	Airport	297.0	Trip Completed	11/7/2016 2:04	11/7/2016 2:37
...	...	...	...	...	...	...
6735	6737	Airport	NaN	No Cars Available	15-07-2016 23:39:15	NaN
2758	6738	Airport	291.0	Trip Completed	15-07-2016 23:45:13	16-07-2016 00:35:45
6739	6739	City	NaN	No Cars Available	15-07-2016 23:46:20	NaN
6737	6740	City	NaN	No Cars Available	15-07-2016 23:43:54	NaN
1419	6741	City	149.0	Trip Completed	15-07-2016 23:42:11	16-07-2016 00:09:09
2071	6742	City	218.0	Trip Completed	15-07-2016 23:38:20	16-07-2016 00:39:38
2265	6743	Airport	236.0	Trip Completed	15-07-2016 23:44:23	16-07-2016 00:30:46
6736	6744	Airport	NaN	No Cars Available	15-07-2016 23:42:51	NaN
6740	6745	City	NaN	No Cars Available	15-07-2016 23:49:03	NaN
6738	6746	City	NaN	No Cars Available	15-07-2016 23:46:03	NaN
641	6747	City	67.0	Trip Completed	15-07-2016 23:47:37	16-07-2016 00:58:57
801	6748	Airport	81.0	Trip Completed	15-07-2016 23:45:24	16-07-2016 00:51:38
623	6749	Airport	65.0	Trip Completed	15-07-2016 23:48:42	16-07-2016 00:36:53
2607	6750	Airport	274.0	Trip Completed	15-07-2016 23:51:42	16-07-2016 00:49:54
6742	6751	City	NaN	No Cars Available	15-07-2016 23:52:06	NaN
6741	6752	Airport	NaN	No Cars Available	15-07-2016 23:50:05	NaN
6744	6753	Airport	NaN	No Cars Available	15-07-2016 23:55:03	NaN
6743	6754	City	NaN	No Cars Available	15-07-2016 23:54:39	NaN

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
1095	6755	Airport	113.0	Trip Completed	15-07-2016 23:57:32	16-07-2016 01:09:24
1319	6756	City	137.0	Trip Completed	15-07-2016 23:57:25	16-07-2016 00:51:09
2401	6757	Airport	251.0	Trip Completed	15-07-2016 23:59:58	16-07-2016 00:30:17
56	6758	Airport	6.0	Trip Completed	15-07-2016 00:03:39	15-07-2016 01:07:39
3339	6759	City	125.0	Cancelled	15-07-2016 23:59:29	NaN
1294	6760	Airport	134.0	Trip Completed	15-07-2016 00:03:52	15-07-2016 00:47:44
455	6761	Airport	47.0	Trip Completed	15-07-2016 00:05:00	15-07-2016 00:47:06
2534	6762	Airport	267.0	Trip Completed	15-07-2016 00:07:29	15-07-2016 00:52:50
2137	6763	City	224.0	Trip Completed	15-07-2016 00:04:44	15-07-2016 01:06:42
2324	6764	City	243.0	Trip Completed	15-07-2016 00:06:12	15-07-2016 01:17:53
6165	6765	Airport	NaN	No Cars Available	15-07-2016 00:09:09	NaN
1042	6766	City	108.0	Trip Completed	15-07-2016 00:06:56	15-07-2016 01:10:34

6745 rows × 6 columns

In [28]:
```
df=pd.read_csv("/home/rmdstic/Documents/TEA-14/iris.csv")
```

In [29]:
```
df
```

Out[29]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
5	5.4	3.9	1.7	0.4	setosa
6	4.6	3.4	1.4	0.3	setosa
7	5.0	3.4	1.5	0.2	setosa
8	4.4	2.9	1.4	0.2	setosa
9	4.9	3.1	1.5	0.1	setosa
10	5.4	3.7	1.5	0.2	setosa
11	4.8	3.4	1.6	0.2	setosa
12	4.8	3.0	1.4	0.1	setosa
13	4.3	3.0	1.1	0.1	setosa
14	5.8	4.0	1.2	0.2	setosa
15	5.7	4.4	1.5	0.4	setosa
16	5.4	3.9	1.3	0.4	setosa
17	5.1	3.5	1.4	0.3	setosa
18	5.7	3.8	1.7	0.3	setosa
19	5.1	3.8	1.5	0.3	setosa
20	5.4	3.4	1.7	0.2	setosa
21	5.1	3.7	1.5	0.4	setosa
22	4.6	3.6	1.0	0.2	setosa
23	5.1	3.3	1.7	0.5	setosa
24	4.8	3.4	1.9	0.2	setosa
25	5.0	3.0	1.6	0.2	setosa
26	5.0	3.4	1.6	0.4	setosa
27	5.2	3.5	1.5	0.2	setosa
28	5.2	3.4	1.4	0.2	setosa
29	4.7	3.2	1.6	0.2	setosa
...	...	...	...	...	...
120	6.9	3.2	5.7	2.3	virginica
121	5.6	2.8	4.9	2.0	virginica
122	7.7	2.8	6.7	2.0	virginica
123	6.3	2.7	4.9	1.8	virginica

	sepal_length	sepal_width	petal_length	petal_width	species
124	6.7	3.3	5.7	2.1	virginica
125	7.2	3.2	6.0	1.8	virginica
126	6.2	2.8	4.8	1.8	virginica
127	6.1	3.0	4.9	1.8	virginica
128	6.4	2.8	5.6	2.1	virginica
129	7.2	3.0	5.8	1.6	virginica
130	7.4	2.8	6.1	1.9	virginica
131	7.9	3.8	6.4	2.0	virginica
132	6.4	2.8	5.6	2.2	virginica
133	6.3	2.8	5.1	1.5	virginica
134	6.1	2.6	5.6	1.4	virginica
135	7.7	3.0	6.1	2.3	virginica
136	6.3	3.4	5.6	2.4	virginica
137	6.4	3.1	5.5	1.8	virginica
138	6.0	3.0	4.8	1.8	virginica
139	6.9	3.1	5.4	2.1	virginica
140	6.7	3.1	5.6	2.4	virginica
141	6.9	3.1	5.1	2.3	virginica
142	5.8	2.7	5.1	1.9	virginica
143	6.8	3.2	5.9	2.3	virginica
144	6.7	3.3	5.7	2.5	virginica
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

150 rows × 5 columns

In [30]:
```
df.head()
```

Out[30]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

In [31]:
```
df.head(10)
```

Out[31]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
5	5.4	3.9	1.7	0.4	setosa
6	4.6	3.4	1.4	0.3	setosa
7	5.0	3.4	1.5	0.2	setosa
8	4.4	2.9	1.4	0.2	setosa
9	4.9	3.1	1.5	0.1	setosa

In [32]:
```
df.tail()
```

Out[32]:

	sepal_length	sepal_width	petal_length	petal_width	species
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

In [33]:
```
df.tail(10)
```

Out[33]:

	sepal_length	sepal_width	petal_length	petal_width	species
140	6.7	3.1	5.6	2.4	virginica
141	6.9	3.1	5.1	2.3	virginica
142	5.8	2.7	5.1	1.9	virginica
143	6.8	3.2	5.9	2.3	virginica
144	6.7	3.3	5.7	2.5	virginica
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

In [34]:
```
df.index
```

Out[34]:
```
RangeIndex(start=0, stop=150, step=1)
```

In [35]:
```
df.columns
```

Out[35]:
```
Index(['sepal_length', 'sepal_width', 'petal_length', 'petal_widt
h',
       'species'],
      dtype='object')
```

In [36]:
```
df.columns.values
```

Out[36]:
```
array(['sepal_length', 'sepal_width', 'petal_length', 'petal_widt
h',
       'species'], dtype=object)
```

In [37]:
```
from sklearn import preprocessing
```

In [38]:
```
min_max_scaler=preprocessing.MinMaxScaler()
```

In [39]:
```
x= df.iloc[:,:4]
```

In [40]:
```
x_scaled = min_max_scaler.fit_transform(x)
```

In [41]:
```
df_normalized = pd.DataFrame(x_scaled)
```

In [42]:
```
df_normalized
```

Out[42]:

	0	1	2	3
0	0.222222	0.625000	0.067797	0.041667
1	0.166667	0.416667	0.067797	0.041667
2	0.111111	0.500000	0.050847	0.041667
3	0.083333	0.458333	0.084746	0.041667
4	0.194444	0.666667	0.067797	0.041667
5	0.305556	0.791667	0.118644	0.125000
6	0.083333	0.583333	0.067797	0.083333
7	0.194444	0.583333	0.084746	0.041667
8	0.027778	0.375000	0.067797	0.041667
9	0.166667	0.458333	0.084746	0.000000
10	0.305556	0.708333	0.084746	0.041667
11	0.138889	0.583333	0.101695	0.041667
12	0.138889	0.416667	0.067797	0.000000
13	0.000000	0.416667	0.016949	0.000000
14	0.416667	0.833333	0.033898	0.041667
15	0.388889	1.000000	0.084746	0.125000
16	0.305556	0.791667	0.050847	0.125000
17	0.222222	0.625000	0.067797	0.083333
18	0.388889	0.750000	0.118644	0.083333
19	0.222222	0.750000	0.084746	0.083333
20	0.305556	0.583333	0.118644	0.041667
21	0.222222	0.708333	0.084746	0.125000
22	0.083333	0.666667	0.000000	0.041667
23	0.222222	0.541667	0.118644	0.166667
24	0.138889	0.583333	0.152542	0.041667
25	0.194444	0.416667	0.101695	0.041667
26	0.194444	0.583333	0.101695	0.125000
27	0.250000	0.625000	0.084746	0.041667
28	0.250000	0.583333	0.067797	0.041667
29	0.111111	0.500000	0.101695	0.041667
...	...	...	...	...
120	0.722222	0.500000	0.796610	0.916667
121	0.361111	0.333333	0.661017	0.791667
122	0.944444	0.333333	0.966102	0.791667
123	0.555556	0.291667	0.661017	0.708333

	0	1	2	3
124	0.666667	0.541667	0.796610	0.833333
125	0.805556	0.500000	0.847458	0.708333
126	0.527778	0.333333	0.644068	0.708333
127	0.500000	0.416667	0.661017	0.708333
128	0.583333	0.333333	0.779661	0.833333
129	0.805556	0.416667	0.813559	0.625000
130	0.861111	0.333333	0.864407	0.750000
131	1.000000	0.750000	0.915254	0.791667
132	0.583333	0.333333	0.779661	0.875000
133	0.555556	0.333333	0.694915	0.583333
134	0.500000	0.250000	0.779661	0.541667
135	0.944444	0.416667	0.864407	0.916667
136	0.555556	0.583333	0.779661	0.958333
137	0.583333	0.458333	0.762712	0.708333
138	0.472222	0.416667	0.644068	0.708333
139	0.722222	0.458333	0.745763	0.833333
140	0.666667	0.458333	0.779661	0.958333
141	0.722222	0.458333	0.694915	0.916667
142	0.416667	0.291667	0.694915	0.750000
143	0.694444	0.500000	0.830508	0.916667
144	0.666667	0.541667	0.796610	1.000000
145	0.666667	0.416667	0.711864	0.916667
146	0.555556	0.208333	0.677966	0.750000
147	0.611111	0.416667	0.711864	0.791667
148	0.527778	0.583333	0.745763	0.916667
149	0.444444	0.416667	0.694915	0.708333

150 rows × 4 columns

In [43]:
```
df
```

Out[43]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
5	5.4	3.9	1.7	0.4	setosa
6	4.6	3.4	1.4	0.3	setosa
7	5.0	3.4	1.5	0.2	setosa
8	4.4	2.9	1.4	0.2	setosa
9	4.9	3.1	1.5	0.1	setosa
10	5.4	3.7	1.5	0.2	setosa
11	4.8	3.4	1.6	0.2	setosa
12	4.8	3.0	1.4	0.1	setosa
13	4.3	3.0	1.1	0.1	setosa
14	5.8	4.0	1.2	0.2	setosa
15	5.7	4.4	1.5	0.4	setosa
16	5.4	3.9	1.3	0.4	setosa
17	5.1	3.5	1.4	0.3	setosa
18	5.7	3.8	1.7	0.3	setosa
19	5.1	3.8	1.5	0.3	setosa
20	5.4	3.4	1.7	0.2	setosa
21	5.1	3.7	1.5	0.4	setosa
22	4.6	3.6	1.0	0.2	setosa
23	5.1	3.3	1.7	0.5	setosa
24	4.8	3.4	1.9	0.2	setosa
25	5.0	3.0	1.6	0.2	setosa
26	5.0	3.4	1.6	0.4	setosa
27	5.2	3.5	1.5	0.2	setosa
28	5.2	3.4	1.4	0.2	setosa
29	4.7	3.2	1.6	0.2	setosa
...	...	...	...	...	...
120	6.9	3.2	5.7	2.3	virginica
121	5.6	2.8	4.9	2.0	virginica
122	7.7	2.8	6.7	2.0	virginica
123	6.3	2.7	4.9	1.8	virginica

	sepal_length	sepal_width	petal_length	petal_width	species
124	6.7	3.3	5.7	2.1	virginica
125	7.2	3.2	6.0	1.8	virginica
126	6.2	2.8	4.8	1.8	virginica
127	6.1	3.0	4.9	1.8	virginica
128	6.4	2.8	5.6	2.1	virginica
129	7.2	3.0	5.8	1.6	virginica
130	7.4	2.8	6.1	1.9	virginica
131	7.9	3.8	6.4	2.0	virginica
132	6.4	2.8	5.6	2.2	virginica
133	6.3	2.8	5.1	1.5	virginica
134	6.1	2.6	5.6	1.4	virginica
135	7.7	3.0	6.1	2.3	virginica
136	6.3	3.4	5.6	2.4	virginica
137	6.4	3.1	5.5	1.8	virginica
138	6.0	3.0	4.8	1.8	virginica
139	6.9	3.1	5.4	2.1	virginica
140	6.7	3.1	5.6	2.4	virginica
141	6.9	3.1	5.1	2.3	virginica
142	5.8	2.7	5.1	1.9	virginica
143	6.8	3.2	5.9	2.3	virginica
144	6.7	3.3	5.7	2.5	virginica
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

150 rows × 5 columns

In [44]:
```
df['species'].unique()
```

Out[44]:
```
array(['setosa', 'versicolor', 'virginica'], dtype=object)
```

In [45]:
```
label_encoder = preprocessing.LabelEncoder()
```

In [46]:
```
df['species'] = label_encoder.fit_transform(df['species'])
```

In [47]:
```
df['species'].unique()
```

Out[47]:
```
array([0, 1, 2])
```

In [48]:
```
from sklearn import preprocessing
```

In [49]:
```
df
```

Out[49]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0
5	5.4	3.9	1.7	0.4	0
6	4.6	3.4	1.4	0.3	0
7	5.0	3.4	1.5	0.2	0
8	4.4	2.9	1.4	0.2	0
9	4.9	3.1	1.5	0.1	0
10	5.4	3.7	1.5	0.2	0
11	4.8	3.4	1.6	0.2	0
12	4.8	3.0	1.4	0.1	0
13	4.3	3.0	1.1	0.1	0
14	5.8	4.0	1.2	0.2	0
15	5.7	4.4	1.5	0.4	0
16	5.4	3.9	1.3	0.4	0
17	5.1	3.5	1.4	0.3	0
18	5.7	3.8	1.7	0.3	0
19	5.1	3.8	1.5	0.3	0
20	5.4	3.4	1.7	0.2	0
21	5.1	3.7	1.5	0.4	0
22	4.6	3.6	1.0	0.2	0
23	5.1	3.3	1.7	0.5	0
24	4.8	3.4	1.9	0.2	0
25	5.0	3.0	1.6	0.2	0
26	5.0	3.4	1.6	0.4	0
27	5.2	3.5	1.5	0.2	0
28	5.2	3.4	1.4	0.2	0
29	4.7	3.2	1.6	0.2	0
...	...	...	...	...	...
120	6.9	3.2	5.7	2.3	2
121	5.6	2.8	4.9	2.0	2
122	7.7	2.8	6.7	2.0	2
123	6.3	2.7	4.9	1.8	2

	sepal_length	sepal_width	petal_length	petal_width	species
124	6.7	3.3	5.7	2.1	2
125	7.2	3.2	6.0	1.8	2
126	6.2	2.8	4.8	1.8	2
127	6.1	3.0	4.9	1.8	2
128	6.4	2.8	5.6	2.1	2
129	7.2	3.0	5.8	1.6	2
130	7.4	2.8	6.1	1.9	2
131	7.9	3.8	6.4	2.0	2
132	6.4	2.8	5.6	2.2	2
133	6.3	2.8	5.1	1.5	2
134	6.1	2.6	5.6	1.4	2
135	7.7	3.0	6.1	2.3	2
136	6.3	3.4	5.6	2.4	2
137	6.4	3.1	5.5	1.8	2
138	6.0	3.0	4.8	1.8	2
139	6.9	3.1	5.4	2.1	2
140	6.7	3.1	5.6	2.4	2
141	6.9	3.1	5.1	2.3	2
142	5.8	2.7	5.1	1.9	2
143	6.8	3.2	5.9	2.3	2
144	6.7	3.3	5.7	2.5	2
145	6.7	3.0	5.2	2.3	2
146	6.3	2.5	5.0	1.9	2
147	6.5	3.0	5.2	2.0	2
148	6.2	3.4	5.4	2.3	2
149	5.9	3.0	5.1	1.8	2

150 rows × 5 columns

In [51]:
```
df= pd.read_csv("/home/rmdstic/Documents/TEA-14/iris.csv")
df
```

Out[51]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
5	5.4	3.9	1.7	0.4	setosa
6	4.6	3.4	1.4	0.3	setosa
7	5.0	3.4	1.5	0.2	setosa
8	4.4	2.9	1.4	0.2	setosa
9	4.9	3.1	1.5	0.1	setosa
10	5.4	3.7	1.5	0.2	setosa
11	4.8	3.4	1.6	0.2	setosa
12	4.8	3.0	1.4	0.1	setosa
13	4.3	3.0	1.1	0.1	setosa
14	5.8	4.0	1.2	0.2	setosa
15	5.7	4.4	1.5	0.4	setosa
16	5.4	3.9	1.3	0.4	setosa
17	5.1	3.5	1.4	0.3	setosa
18	5.7	3.8	1.7	0.3	setosa
19	5.1	3.8	1.5	0.3	setosa
20	5.4	3.4	1.7	0.2	setosa
21	5.1	3.7	1.5	0.4	setosa
22	4.6	3.6	1.0	0.2	setosa
23	5.1	3.3	1.7	0.5	setosa
24	4.8	3.4	1.9	0.2	setosa
25	5.0	3.0	1.6	0.2	setosa
26	5.0	3.4	1.6	0.4	setosa
27	5.2	3.5	1.5	0.2	setosa
28	5.2	3.4	1.4	0.2	setosa
29	4.7	3.2	1.6	0.2	setosa
...	...	...	...	...	...
120	6.9	3.2	5.7	2.3	virginica
121	5.6	2.8	4.9	2.0	virginica
122	7.7	2.8	6.7	2.0	virginica

	sepal_length	sepal_width	petal_length	petal_width	species
123	6.3	2.7	4.9	1.8	virginica
124	6.7	3.3	5.7	2.1	virginica
125	7.2	3.2	6.0	1.8	virginica
126	6.2	2.8	4.8	1.8	virginica
127	6.1	3.0	4.9	1.8	virginica
128	6.4	2.8	5.6	2.1	virginica
129	7.2	3.0	5.8	1.6	virginica
130	7.4	2.8	6.1	1.9	virginica
131	7.9	3.8	6.4	2.0	virginica
132	6.4	2.8	5.6	2.2	virginica
133	6.3	2.8	5.1	1.5	virginica
134	6.1	2.6	5.6	1.4	virginica
135	7.7	3.0	6.1	2.3	virginica
136	6.3	3.4	5.6	2.4	virginica
137	6.4	3.1	5.5	1.8	virginica
138	6.0	3.0	4.8	1.8	virginica
139	6.9	3.1	5.4	2.1	virginica
140	6.7	3.1	5.6	2.4	virginica
141	6.9	3.1	5.1	2.3	virginica
142	5.8	2.7	5.1	1.9	virginica
143	6.8	3.2	5.9	2.3	virginica
144	6.7	3.3	5.7	2.5	virginica
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

150 rows × 5 columns

In [52]:
```
features_df=df.drop(columns=['species'])
features_df
```

Out[52]:

	sepal_length	sepal_width	petal_length	petal_width
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
5	5.4	3.9	1.7	0.4
6	4.6	3.4	1.4	0.3
7	5.0	3.4	1.5	0.2
8	4.4	2.9	1.4	0.2
9	4.9	3.1	1.5	0.1
10	5.4	3.7	1.5	0.2
11	4.8	3.4	1.6	0.2
12	4.8	3.0	1.4	0.1
13	4.3	3.0	1.1	0.1
14	5.8	4.0	1.2	0.2
15	5.7	4.4	1.5	0.4
16	5.4	3.9	1.3	0.4
17	5.1	3.5	1.4	0.3
18	5.7	3.8	1.7	0.3
19	5.1	3.8	1.5	0.3
20	5.4	3.4	1.7	0.2
21	5.1	3.7	1.5	0.4
22	4.6	3.6	1.0	0.2
23	5.1	3.3	1.7	0.5
24	4.8	3.4	1.9	0.2
25	5.0	3.0	1.6	0.2
26	5.0	3.4	1.6	0.4
27	5.2	3.5	1.5	0.2
28	5.2	3.4	1.4	0.2
29	4.7	3.2	1.6	0.2
...	...	...	...	...
120	6.9	3.2	5.7	2.3
121	5.6	2.8	4.9	2.0
122	7.7	2.8	6.7	2.0

	sepal_length	sepal_width	petal_length	petal_width
123	6.3	2.7	4.9	1.8
124	6.7	3.3	5.7	2.1
125	7.2	3.2	6.0	1.8
126	6.2	2.8	4.8	1.8
127	6.1	3.0	4.9	1.8
128	6.4	2.8	5.6	2.1
129	7.2	3.0	5.8	1.6
130	7.4	2.8	6.1	1.9
131	7.9	3.8	6.4	2.0
132	6.4	2.8	5.6	2.2
133	6.3	2.8	5.1	1.5
134	6.1	2.6	5.6	1.4
135	7.7	3.0	6.1	2.3
136	6.3	3.4	5.6	2.4
137	6.4	3.1	5.5	1.8
138	6.0	3.0	4.8	1.8
139	6.9	3.1	5.4	2.1
140	6.7	3.1	5.6	2.4
141	6.9	3.1	5.1	2.3
142	5.8	2.7	5.1	1.9
143	6.8	3.2	5.9	2.3
144	6.7	3.3	5.7	2.5
145	6.7	3.0	5.2	2.3
146	6.3	2.5	5.0	1.9
147	6.5	3.0	5.2	2.0
148	6.2	3.4	5.4	2.3
149	5.9	3.0	5.1	1.8

150 rows × 4 columns

In [53]:
```
enc = preprocessing.OneHotEncoder(sparse=False)
enc_df=(enc.fit_transform(df[['species']]))
x=pd.DataFrame(enc_df)
x
```

Out[53]:

	0	1	2
0	1.0	0.0	0.0
1	1.0	0.0	0.0
2	1.0	0.0	0.0
3	1.0	0.0	0.0
4	1.0	0.0	0.0
5	1.0	0.0	0.0
6	1.0	0.0	0.0
7	1.0	0.0	0.0
8	1.0	0.0	0.0
9	1.0	0.0	0.0
10	1.0	0.0	0.0
11	1.0	0.0	0.0
12	1.0	0.0	0.0
13	1.0	0.0	0.0
14	1.0	0.0	0.0
15	1.0	0.0	0.0
16	1.0	0.0	0.0
17	1.0	0.0	0.0
18	1.0	0.0	0.0
19	1.0	0.0	0.0
20	1.0	0.0	0.0
21	1.0	0.0	0.0
22	1.0	0.0	0.0
23	1.0	0.0	0.0
24	1.0	0.0	0.0
25	1.0	0.0	0.0
26	1.0	0.0	0.0
27	1.0	0.0	0.0
28	1.0	0.0	0.0
29	1.0	0.0	0.0
...	...	...	...
120	0.0	0.0	1.0
121	0.0	0.0	1.0

	0	1	2
122	0.0	0.0	1.0
123	0.0	0.0	1.0
124	0.0	0.0	1.0
125	0.0	0.0	1.0
126	0.0	0.0	1.0
127	0.0	0.0	1.0
128	0.0	0.0	1.0
129	0.0	0.0	1.0
130	0.0	0.0	1.0
131	0.0	0.0	1.0
132	0.0	0.0	1.0
133	0.0	0.0	1.0
134	0.0	0.0	1.0
135	0.0	0.0	1.0
136	0.0	0.0	1.0
137	0.0	0.0	1.0
138	0.0	0.0	1.0
139	0.0	0.0	1.0
140	0.0	0.0	1.0
141	0.0	0.0	1.0
142	0.0	0.0	1.0
143	0.0	0.0	1.0
144	0.0	0.0	1.0
145	0.0	0.0	1.0
146	0.0	0.0	1.0
147	0.0	0.0	1.0
148	0.0	0.0	1.0
149	0.0	0.0	1.0

150 rows × 3 columns

In [54]:
```
df_encode = features_df.join(x)
```

In [55]:
```
df_encode
```

Out[55]:

	sepal_length	sepal_width	petal_length	petal_width	0	1	2
0	5.1	3.5	1.4	0.2	1.0	0.0	0.0
1	4.9	3.0	1.4	0.2	1.0	0.0	0.0
2	4.7	3.2	1.3	0.2	1.0	0.0	0.0
3	4.6	3.1	1.5	0.2	1.0	0.0	0.0
4	5.0	3.6	1.4	0.2	1.0	0.0	0.0
5	5.4	3.9	1.7	0.4	1.0	0.0	0.0
6	4.6	3.4	1.4	0.3	1.0	0.0	0.0
7	5.0	3.4	1.5	0.2	1.0	0.0	0.0
8	4.4	2.9	1.4	0.2	1.0	0.0	0.0
9	4.9	3.1	1.5	0.1	1.0	0.0	0.0
10	5.4	3.7	1.5	0.2	1.0	0.0	0.0
11	4.8	3.4	1.6	0.2	1.0	0.0	0.0
12	4.8	3.0	1.4	0.1	1.0	0.0	0.0
13	4.3	3.0	1.1	0.1	1.0	0.0	0.0
14	5.8	4.0	1.2	0.2	1.0	0.0	0.0
15	5.7	4.4	1.5	0.4	1.0	0.0	0.0
16	5.4	3.9	1.3	0.4	1.0	0.0	0.0
17	5.1	3.5	1.4	0.3	1.0	0.0	0.0
18	5.7	3.8	1.7	0.3	1.0	0.0	0.0
19	5.1	3.8	1.5	0.3	1.0	0.0	0.0
20	5.4	3.4	1.7	0.2	1.0	0.0	0.0
21	5.1	3.7	1.5	0.4	1.0	0.0	0.0
22	4.6	3.6	1.0	0.2	1.0	0.0	0.0
23	5.1	3.3	1.7	0.5	1.0	0.0	0.0
24	4.8	3.4	1.9	0.2	1.0	0.0	0.0
25	5.0	3.0	1.6	0.2	1.0	0.0	0.0
26	5.0	3.4	1.6	0.4	1.0	0.0	0.0
27	5.2	3.5	1.5	0.2	1.0	0.0	0.0
28	5.2	3.4	1.4	0.2	1.0	0.0	0.0
29	4.7	3.2	1.6	0.2	1.0	0.0	0.0
...	...	...	...	...	...	...	...
120	6.9	3.2	5.7	2.3	0.0	0.0	1.0
121	5.6	2.8	4.9	2.0	0.0	0.0	1.0
122	7.7	2.8	6.7	2.0	0.0	0.0	1.0
123	6.3	2.7	4.9	1.8	0.0	0.0	1.0

	sepal_length	sepal_width	petal_length	petal_width	0	1	2
124	6.7	3.3	5.7	2.1	0.0	0.0	1.0
125	7.2	3.2	6.0	1.8	0.0	0.0	1.0
126	6.2	2.8	4.8	1.8	0.0	0.0	1.0
127	6.1	3.0	4.9	1.8	0.0	0.0	1.0
128	6.4	2.8	5.6	2.1	0.0	0.0	1.0
129	7.2	3.0	5.8	1.6	0.0	0.0	1.0
130	7.4	2.8	6.1	1.9	0.0	0.0	1.0
131	7.9	3.8	6.4	2.0	0.0	0.0	1.0
132	6.4	2.8	5.6	2.2	0.0	0.0	1.0
133	6.3	2.8	5.1	1.5	0.0	0.0	1.0
134	6.1	2.6	5.6	1.4	0.0	0.0	1.0
135	7.7	3.0	6.1	2.3	0.0	0.0	1.0
136	6.3	3.4	5.6	2.4	0.0	0.0	1.0
137	6.4	3.1	5.5	1.8	0.0	0.0	1.0
138	6.0	3.0	4.8	1.8	0.0	0.0	1.0
139	6.9	3.1	5.4	2.1	0.0	0.0	1.0
140	6.7	3.1	5.6	2.4	0.0	0.0	1.0
141	6.9	3.1	5.1	2.3	0.0	0.0	1.0
142	5.8	2.7	5.1	1.9	0.0	0.0	1.0
143	6.8	3.2	5.9	2.3	0.0	0.0	1.0
144	6.7	3.3	5.7	2.5	0.0	0.0	1.0
145	6.7	3.0	5.2	2.3	0.0	0.0	1.0
146	6.3	2.5	5.0	1.9	0.0	0.0	1.0
147	6.5	3.0	5.2	2.0	0.0	0.0	1.0
148	6.2	3.4	5.4	2.3	0.0	0.0	1.0
149	5.9	3.0	5.1	1.8	0.0	0.0	1.0

150 rows × 7 columns

In [56]:
```
df_encode
```

Out[56]:

	sepal_length	sepal_width	petal_length	petal_width	0	1	2
0	5.1	3.5	1.4	0.2	1.0	0.0	0.0
1	4.9	3.0	1.4	0.2	1.0	0.0	0.0
2	4.7	3.2	1.3	0.2	1.0	0.0	0.0
3	4.6	3.1	1.5	0.2	1.0	0.0	0.0
4	5.0	3.6	1.4	0.2	1.0	0.0	0.0
5	5.4	3.9	1.7	0.4	1.0	0.0	0.0
6	4.6	3.4	1.4	0.3	1.0	0.0	0.0
7	5.0	3.4	1.5	0.2	1.0	0.0	0.0
8	4.4	2.9	1.4	0.2	1.0	0.0	0.0
9	4.9	3.1	1.5	0.1	1.0	0.0	0.0
10	5.4	3.7	1.5	0.2	1.0	0.0	0.0
11	4.8	3.4	1.6	0.2	1.0	0.0	0.0
12	4.8	3.0	1.4	0.1	1.0	0.0	0.0
13	4.3	3.0	1.1	0.1	1.0	0.0	0.0
14	5.8	4.0	1.2	0.2	1.0	0.0	0.0
15	5.7	4.4	1.5	0.4	1.0	0.0	0.0
16	5.4	3.9	1.3	0.4	1.0	0.0	0.0
17	5.1	3.5	1.4	0.3	1.0	0.0	0.0
18	5.7	3.8	1.7	0.3	1.0	0.0	0.0
19	5.1	3.8	1.5	0.3	1.0	0.0	0.0
20	5.4	3.4	1.7	0.2	1.0	0.0	0.0
21	5.1	3.7	1.5	0.4	1.0	0.0	0.0
22	4.6	3.6	1.0	0.2	1.0	0.0	0.0
23	5.1	3.3	1.7	0.5	1.0	0.0	0.0
24	4.8	3.4	1.9	0.2	1.0	0.0	0.0
25	5.0	3.0	1.6	0.2	1.0	0.0	0.0
26	5.0	3.4	1.6	0.4	1.0	0.0	0.0
27	5.2	3.5	1.5	0.2	1.0	0.0	0.0
28	5.2	3.4	1.4	0.2	1.0	0.0	0.0
29	4.7	3.2	1.6	0.2	1.0	0.0	0.0
...	...	...	...	...	...	...	...
120	6.9	3.2	5.7	2.3	0.0	0.0	1.0
121	5.6	2.8	4.9	2.0	0.0	0.0	1.0
122	7.7	2.8	6.7	2.0	0.0	0.0	1.0
123	6.3	2.7	4.9	1.8	0.0	0.0	1.0

	sepal_length	sepal_width	petal_length	petal_width	0	1	2
124	6.7	3.3	5.7	2.1	0.0	0.0	1.0
125	7.2	3.2	6.0	1.8	0.0	0.0	1.0
126	6.2	2.8	4.8	1.8	0.0	0.0	1.0
127	6.1	3.0	4.9	1.8	0.0	0.0	1.0
128	6.4	2.8	5.6	2.1	0.0	0.0	1.0
129	7.2	3.0	5.8	1.6	0.0	0.0	1.0
130	7.4	2.8	6.1	1.9	0.0	0.0	1.0
131	7.9	3.8	6.4	2.0	0.0	0.0	1.0
132	6.4	2.8	5.6	2.2	0.0	0.0	1.0
133	6.3	2.8	5.1	1.5	0.0	0.0	1.0
134	6.1	2.6	5.6	1.4	0.0	0.0	1.0
135	7.7	3.0	6.1	2.3	0.0	0.0	1.0
136	6.3	3.4	5.6	2.4	0.0	0.0	1.0
137	6.4	3.1	5.5	1.8	0.0	0.0	1.0
138	6.0	3.0	4.8	1.8	0.0	0.0	1.0
139	6.9	3.1	5.4	2.1	0.0	0.0	1.0
140	6.7	3.1	5.6	2.4	0.0	0.0	1.0
141	6.9	3.1	5.1	2.3	0.0	0.0	1.0
142	5.8	2.7	5.1	1.9	0.0	0.0	1.0
143	6.8	3.2	5.9	2.3	0.0	0.0	1.0
144	6.7	3.3	5.7	2.5	0.0	0.0	1.0
145	6.7	3.0	5.2	2.3	0.0	0.0	1.0
146	6.3	2.5	5.0	1.9	0.0	0.0	1.0
147	6.5	3.0	5.2	2.0	0.0	0.0	1.0
148	6.2	3.4	5.4	2.3	0.0	0.0	1.0
149	5.9	3.0	5.1	1.8	0.0	0.0	1.0

150 rows × 7 columns

In [ ]: