Task: Replace the unidentified values with mean values

Step 1: Import the essential libraries and read the uploaded csv file

- copy the path from the uploaded file

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.preprocessing import *

ds=pd.read_csv("/content/sample_data/A_2018.02.02_11.21.55.csv")
ds
```

	Timestamp	Longitude	Latitude	Speed	Operatorname	CellID	NetworkMode	RSRP	RSRQ	SNR	CQI	RSSI	DL_bitrate	UL_bi
0	2018.02.02_11.21.55	-8.656289	52.132759	44	A	2	LTE	-99	-11	4	6	-86	12906	
1	2018.02.02_11.21.56	-8.656298	52.132711	42	A	2	LTE	-99	-11	4	6	-86	12906	
2	2018.02.02_11.21.56	-8.656298	52.132711	42	A	2	LTE	-103	-13	4	6	-86	14248	
3	2018.02.02_11.21.57	-8.656298	52.132711	42	A	2	LTE	-103	-13	4	6	-86	20612	
4	2018.02.02_11.21.58	-8.656298	52.132711	42	A	2	LTE	-103	-13	4	6	-84	22312	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
608	2018.02.02_11.34.14	-8.656298	52.132711	42	A	6	HSPA+	-81	-2	-	-	-	515	
609	2018.02.02_11.34.15	-8.656298	52.132711	42	A	6	HSPA+	-81	-2	-	-	-	380	
610	2018.02.02_11.34.16	-8.656298	52.132711	42	A	6	HSPA+	-83	-2	-	-	-	449	
611	2018.02.02_11.34.18	-8.656298	52.132711	42	A	6	HSPA+	-83	-2	-	-	-	567	
612	2018.02.02_11.34.18	-8.656298	52.132711	42	A	6	HSPA+	-83	-2	-	-	-	567	

613 rows × 20 columns

Step 2: Describe the data set

- By this we'll be able to study the dataset

```
ds.describe()
```

	Longitude	Latitude	Speed	CellID	RSRP	DL_bitrate	UL_bitrate
count	6.130000e+02	613.000000	613.000000	613.000000	613.000000	613.000000	613.000000
mean	-8.656298e+00	52.132711	42.003263	3805.942904	-85.939641	5413.768352	114.344209
std	3.635066e-07	0.000002	0.080779	13759.425286	17.190340	6446.983454	127.216662
min	-8.656298e+00	52.132711	42.000000	0.000000	-200.000000	0.000000	0.000000
25%	-8.656298e+00	52.132711	42.000000	2.000000	-102.000000	1152.000000	24.000000
50%	-8.656298e+00	52.132711	42.000000	6.000000	-81.000000	2930.000000	68.000000
75%	-8.656298e+00	52.132711	42.000000	6.000000	-73.000000	7023.000000	159.000000
max	-8.656289e+00	52.132759	44.000000	57094.000000	-51.000000	47543.000000	822.000000

Step 2: Replace the '-' with NaN values

```
symbol_to_replace="-"
ds.replace(symbol_to_replace, np.nan, inplace= True)
ds
ds.to_csv("/content/cleanedds.csv", index=False)
```

Step 3: Now replace the NaN value of acquired column with the mean value of the entire column which is to be replaced

```
ds_1=pd.read_csv("/content/cleanedds.csv")
```

```
columns_to_replace = ['RSSI', 'NRxRSRP', 'NRxRSRQ']
ds_1[columns_to_replace] = ds_1[columns_to_replace].fillna(ds_1.mean())
ds_1
ds_1.to_csv("/content/final_ds.csv", index=False)
```

```
<ipython-input-24-3b8168a45961>:2: FutureWarning: The default value of numeric_only in DataFrame.mean is deprecated. In a future ve
  ds_1[columns_to_replace] = ds_1[columns_to_replace].fillna(ds_1.mean())
```

```

```