In [14]:
```
import numpy as np
import pandas as pd
```

Reading dataset using pandas

In [15]:
```
data=pd.read_csv('DSDataLastThreeMonths.csv')
```

In [16]:
```
data.head(10)
```

Out[16]:

	CASTNO	HM_WT	AIM_S	HM_S	HM_C	HM_SI	HM_TI	HM_MN	CAC2	MG	HM_TEMP	CAC2_INJ_TIME	MG_INJ_TIME	DS_S
0	V99790	171	0.005	0.040	4.572	0.518	0.069	0.040	347	49	1387.0	10	4	0.009
1	V99791	161	0.005	0.035	4.580	0.559	0.070	0.041	361	45	1386.0	10	4	0.007
2	V99789	170	0.005	0.042	4.590	0.468	0.044	0.036	355	52	1368.0	9	4	0.013
3	V99787	159	0.005	0.034	4.530	0.737	0.077	0.040	313	44	1401.0	9	4	0.010
4	V99782	162	0.005	0.042	4.560	0.443	0.056	0.037	360	55	1346.0	10	5	0.013
5	V99781	165	0.005	0.035	4.520	0.480	0.056	0.037	323	49	1407.0	9	4	0.010
6	V99779	167	0.005	0.040	4.560	0.630	0.069	0.037	349	51	1395.0	10	5	0.009
7	V99778	163	0.005	0.040	4.560	0.547	0.057	0.036	346	78	1373.0	9	5	0.012
8	V99777	164	0.005	0.050	4.500	0.397	0.049	0.037	373	72	1386.0	11	7	0.004
9	V99776	162	0.005	0.039	4.520	0.622	0.069	0.040	344	56	1357.0	9	5	0.007

In [17]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6697 entries, 0 to 6696
Data columns (total 14 columns):
CASTNO          6697 non-null object
```

```
HM_WT             6697 non-null int64
AIM_S             6697 non-null float64
HM_S              6697 non-null float64
HM_C              6697 non-null float64
HM_SI             6697 non-null float64
HM_TI             6697 non-null float64
HM_MN             6697 non-null float64
CAC2              6697 non-null int64
MG                6697 non-null int64
HM_TEMP           6094 non-null float64
CAC2_INJ_TIME     6697 non-null int64
MG_INJ_TIME       6697 non-null int64
DS_S              6697 non-null float64
dtypes: float64(8), int64(5), object(1)
memory usage: 732.6+ KB
```

In [24]:
```
# Above we can see all columns have 6697 entries except HM_TEMP
# HM_TEMP has only 6097 entries.
# To Fill missing values I have taken mean of HM_TEMP values and fitted the resulting mean into missing positions.
```

In [18]:
```
data_fit=data.fillna(data["HM_TEMP"].mean())
```

In [19]:
```
data_fit.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6697 entries, 0 to 6696
Data columns (total 14 columns):
CASTNO            6697 non-null object
HM_WT             6697 non-null int64
AIM_S             6697 non-null float64
HM_S              6697 non-null float64
HM_C              6697 non-null float64
HM_SI             6697 non-null float64
HM_TI             6697 non-null float64
HM_MN             6697 non-null float64
```

```
CAC2              6697 non-null int64
MG                6697 non-null int64
HM_TEMP           6697 non-null float64
CAC2_INJ_TIME     6697 non-null int64
MG_INJ_TIME       6697 non-null int64
DS_S              6697 non-null float64
dtypes: float64(8), int64(5), object(1)
memory usage: 732.6+ KB
```

In [21]:
```
data_fit.head(10)
```

Out[21]:

	CASTNO	HM_WT	AIM_S	HM_S	HM_C	HM_SI	HM_TI	HM_MN	CAC2	MG	HM_TEMP	CAC2_INJ_TIME	MG_INJ_TIME	DS_S
0	V99790	171	0.005	0.040	4.572	0.518	0.069	0.040	347	49	1387.0	10	4	0.009
1	V99791	161	0.005	0.035	4.580	0.559	0.070	0.041	361	45	1386.0	10	4	0.007
2	V99789	170	0.005	0.042	4.590	0.468	0.044	0.036	355	52	1368.0	9	4	0.013
3	V99787	159	0.005	0.034	4.530	0.737	0.077	0.040	313	44	1401.0	9	4	0.010
4	V99782	162	0.005	0.042	4.560	0.443	0.056	0.037	360	55	1346.0	10	5	0.013
5	V99781	165	0.005	0.035	4.520	0.480	0.056	0.037	323	49	1407.0	9	4	0.010
6	V99779	167	0.005	0.040	4.560	0.630	0.069	0.037	349	51	1395.0	10	5	0.009
7	V99778	163	0.005	0.040	4.560	0.547	0.057	0.036	346	78	1373.0	9	5	0.012
8	V99777	164	0.005	0.050	4.500	0.397	0.049	0.037	373	72	1386.0	11	7	0.004
9	V99776	162	0.005	0.039	4.520	0.622	0.069	0.040	344	56	1357.0	9	5	0.007

In [25]:
```
type(data_fit)
```

Out[25]:
```
pandas.core.frame.DataFrame
```

In [29]:
```
data_fit.to_csv('dataset.csv',index=False)
```

In [ ]:
```

```