GE Aviation Data Analysis

In [58]:

```
import pandas as pd
```

In [59]:

```
df1 = pd.read_csv("av_engine_data_aic_psql.csv")
```

In [60]:

```
df1
```

Out[60]:

dataset	esn	unit	flight_cycle	datetime	operator	depart_icao	destination_icao	hpc_eff_mod	hpc_flow_mod	tra	t2	t24	t30	t50	p2	p15	p30	nf

In [61]:

```
df2 = pd.read_csv("av_engine_data_axm_psql.csv")
```

In [62]:

```
df2
```

Out[62]:

	dataset	esn	unit	flight_cycle	datetime	operator	depart_icao	destination_icao	hpc_eff_mod	hpc_flow_mod	tra	t2	t24	t30
0	test_FD001	999126	26	1	2017-12-31 18:33:07	AXM	VTBD	VTUV	-0.0027	0.0006	100	518.67	642.00	1582.88
1	test_FD001	999126	26	2	2017-12-31 19:59:49	AXM	VTUV	VTBD	-0.0029	0.0002	100	518.67	642.35	1589.01
2	test_FD001	999126	26	3	2017-12-31 21:44:38	AXM	VTBD	VMMC	0.0008	0.0001	100	518.67	642.69	1590.16
3	test_FD001	999126	26	4	2018-01-01 00:50:02	AXM	VMMC	VTBD	-0.0026	0.0005	100	518.67	641.76	1583.37
4	test_FD001	999126	26	5	2018-01-01 04:25:32	AXM	VTBD	VTSP	0.0020	0.0005	100	518.67	642.54	1591.71
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
21183	train_FD001	999096	96	264	2018-02-13 22:32:30	AXM	WMKK	VDSV	-0.0034	0.0005	100	518.67	643.45	1603.59
21184	train_FD001	999096	96	265	2018-02-14 00:48:00	AXM	VDSV	WMKK	0.0015	0.0004	100	518.67	644.20	1603.88
21185	train_FD001	999096	96	266	2018-02-14 02:06:00	AXM	WMKK	WBKL	-0.0028	-0.0002	100	518.67	643.62	1599.47
21186	train_FD001	999096	96	267	2018-02-14 03:23:32	AXM	WMKK	WBKL	0.0001	0.0005	100	518.67	644.13	1595.57
21187	train_FD001	999096	96	268	2018-02-14 17:04:38	AXM	WMKK	VDSR	-0.0031	-0.0005	100	518.67	643.74	1599.20

21188 rows × 32 columns

In [63]:

```
df3 = pd.read_csv("av_engine_data_fron_psql.csv")
```

In [64]:

```
df3
```

Out[64]:

	dataset	esn	unit	flight_cycle	datetime	operator	depart_icao	destination_icao	hpc_eff_mod	hpc_flow_mod	tra	t2	t24	t30
0	train_FD001	999050	50	1	2018-01-06 06:01:09	FRON	KMCO	KMSY	-0.0029	-0.0002	100	518.67	642.66	1591.79
1	train_FD001	999050	50	2	2018-01-06 07:41:00	FRON	KMSY	KSAT	-0.0002	-0.0005	100	518.67	642.28	1587.84
2	train_FD001	999050	50	3	2018-01-06 08:41:18	FRON	KMSY	KSAT	-0.0010	-0.0005	100	518.67	642.21	1586.89
3	train_FD001	999050	50	4	2018-01-06 10:14:00	FRON	KSAT	KSAN	-0.0061	-0.0002	100	518.67	643.19	1587.36
4	train_FD001	999050	50	5	2018-01-06 11:12:52	FRON	KSAT	KSAN	-0.0002	0.0001	100	518.67	642.47	1584.96
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7285	train_FD001	999086	86	271	2018-02-08 08:24:59	FRON	KMIA	KLAS	-0.0017	0.0000	100	518.67	643.82	1599.22
7286	train_FD001	999086	86	272	2018-02-08 13:16:00	FRON	KLAS	KIND	0.0002	-0.0004	100	518.67	643.50	1600.49
7287	train_FD001	999086	86	273	2018-02-08 14:16:02	FRON	KLAS	KIND	0.0017	0.0002	100	518.67	643.41	1596.95
7288	train_FD001	999086	86	274	2018-02-08 18:21:45	FRON	KIND	KMCO	0.0003	0.0003	100	518.67	643.03	1602.04
7289	train_FD001	999086	86	275	2018-02-09 09:05:00	FRON	KMCO	KBUF	-0.0002	-0.0005	100	518.67	643.15	1600.16

7290 rows × 32 columns

In [65]:

```
df4 = pd.read_csv("av_engine_data_pgt_psql.csv")
```

In [66]:

```
df4
```

Out[66]:

	dataset	esn	unit	flight_cycle	datetime	operator	depart_icao	destination_icao	hpc_eff_mod	hpc_flow_mod	tra	t2	t24	t30
0	train_FD001	999056	56	1	2018-01-01 06:33:13	PGT	LTBJ	LTCR	0.0012	-0.0004	100	518.67	642.75	1586.44
1	train_FD001	999056	56	2	2018-01-01 09:40:21	PGT	LTCR	LTBJ	0.0012	-0.0004	100	518.67	642.47	1584.96
2	train_FD001	999056	56	3	2018-01-01 12:23:01	PGT	LTBJ	LTFJ	0.0026	0.0005	100	518.67	642.52	1587.64
3	train_FD001	999056	56	4	2018-01-01 14:11:10	PGT	LTFJ	LTCG	0.0034	-0.0002	100	518.67	642.51	1587.80
4	train_FD001	999056	56	5	2018-01-01 21:10:50	PGT	LTCG	LTFJ	0.0024	-0.0001	100	518.67	643.08	1593.15
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
6295	train_FD001	999084	84	257	2018-06-06 08:22:54	PGT	LFSB	LTFJ	-0.0027	0.0000	100	518.67	643.34	1599.89
6296	train_FD001	999084	84	258	2018-06-06 12:01:39	PGT	LTFJ	LTAJ	0.0026	0.0001	100	518.67	643.87	1598.81
6297	train_FD001	999084	84	259	2018-06-06 14:01:52	PGT	LTAJ	LTFJ	-0.0013	0.0000	100	518.67	643.20	1605.59
6298	train_FD001	999084	84	260	2018-06-06 18:41:30	PGT	OKBK	LTFJ	-0.0023	0.0001	100	518.67	643.68	1606.08
6299	train_FD001	999095	95	266	2018-02-08 01:30:23	PGT	LTFJ	EDDS	0.0019	0.0003	100	518.67	643.43	1609.16

6300 rows × 32 columns

Concatenate all data into one dataframe

In [67]:

```
df = pd.concat([df1,df2,df3,df4],axis=0)
```

In [68]:

```
df
```

Out[68]:

	dataset	esn	unit	flight_cycle	datetime	operator	depart_icao	destination_icao	hpc_eff_mod	hpc_flow_mod	tra	t2	t24	t30
0	test_FD001	999126	26	1	2017-12-31 18:33:07	AXM	VTBD	VTUV	-0.0027	0.0006	100	518.67	642.00	1582.88
1	test_FD001	999126	26	2	2017-12-31 19:59:49	AXM	VTUV	VTBD	-0.0029	0.0002	100	518.67	642.35	1589.01
2	test_FD001	999126	26	3	2017-12-31 21:44:38	AXM	VTBD	VMMC	0.0008	0.0001	100	518.67	642.69	1590.16
3	test_FD001	999126	26	4	2018-01-01 00:50:02	AXM	VMMC	VTBD	-0.0026	0.0005	100	518.67	641.76	1583.37
4	test_FD001	999126	26	5	2018-01-01 04:25:32	AXM	VTBD	VTSP	0.0020	0.0005	100	518.67	642.54	1591.71
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
6295	train_FD001	999084	84	257	2018-06-06 08:22:54	PGT	LFSB	LTFJ	-0.0027	0.0000	100	518.67	643.34	1599.89
6296	train_FD001	999084	84	258	2018-06-06 12:01:39	PGT	LTFJ	LTAJ	0.0026	0.0001	100	518.67	643.87	1598.81
6297	train_FD001	999084	84	259	2018-06-06 14:01:52	PGT	LTAJ	LTFJ	-0.0013	0.0000	100	518.67	643.20	1605.59
6298	train_FD001	999084	84	260	2018-06-06 18:41:30	PGT	OKBK	LTFJ	-0.0023	0.0001	100	518.67	643.68	1606.08
6299	train_FD001	999095	95	266	2018-02-08 01:30:23	PGT	LTFJ	EDDS	0.0019	0.0003	100	518.67	643.43	1609.16

34778 rows × 32 columns

Resetting index

In [69]:

```
df.reset_index(drop=True, inplace=True)
```

In [70]:

```
df.head(10)
```

Out[70]:

	dataset	esn	unit	flight_cycle	datetime	operator	depart_icao	destination_icao	hpc_eff_mod	hpc_flow_mod	tra	t2	t24	t30	
0	test_FD001	999126	26	1	2017-12-31 18:33:07	AXM	VTBD	VTUV	-0.0027	0.0006	100	518.67	642.00	1582.88	139
1	test_FD001	999126	26	2	2017-12-31 19:59:49	AXM	VTUV	VTBD	-0.0029	0.0002	100	518.67	642.35	1589.01	139
2	test_FD001	999126	26	3	2017-12-31 21:44:38	AXM	VTBD	VMMC	0.0008	0.0001	100	518.67	642.69	1590.16	139
3	test_FD001	999126	26	4	2018-01-01 00:50:02	AXM	VMMC	VTBD	-0.0026	0.0005	100	518.67	641.76	1583.37	138
4	test_FD001	999126	26	5	2018-01-01 04:25:32	AXM	VTBD	VTSP	0.0020	0.0005	100	518.67	642.54	1591.71	140
5	test_FD001	999126	26	6	2018-01-01 06:25:56	AXM	VTSP	VTBD	-0.0032	-0.0001	100	518.67	641.98	1592.68	139
6	test_FD001	999126	26	7	2018-01-01 08:37:05	AXM	VTBD	VTCC	0.0009	0.0003	100	518.67	642.06	1583.81	139
7	test_FD001	999126	26	8	2018-01-02 08:52:57	AXM	VTBD	VOTR	0.0035	0.0004	100	518.67	642.06	1583.40	140
8	test_FD001	999126	26	9	2018-01-05 02:37:00	AXM	VTBD	WSSS	-0.0012	-0.0001	100	518.67	641.68	1587.24	140
9	test_FD001	999126	26	10	2018-01-05 03:55:29	AXM	VTBD	WSSS	-0.0031	0.0007	100	518.67	642.24	1584.60	140

Saving data into CSV file

In [71]:

```
df.to_csv("avenginedata.csv",index=False)
```

In [72]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 34778 entries, 0 to 34777
Data columns (total 32 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   dataset           34778 non-null  object 
 1   esn               34778 non-null  object 
 2   unit              34778 non-null  object 
 3   flight_cycle      34778 non-null  object 
 4   datetime          34778 non-null  object 
 5   operator          34778 non-null  object 
 6   depart_icao       33791 non-null  object 
 7   destination_icao  33367 non-null  object 
 8   hpc_eff_mod       34778 non-null  float64
 9   hpc_flow_mod      34778 non-null  float64
 10  tra               34778 non-null  object 
 11  t2                34778 non-null  float64
 12  t24               34778 non-null  float64
 13  t30               34778 non-null  float64
 14  t50               34778 non-null  float64
 15  p2                34778 non-null  float64
 16  p15               34778 non-null  float64
 17  p30               34778 non-null  float64
 18  nf                34778 non-null  float64
 19  nc                34778 non-null  float64
 20  epr               34778 non-null  float64
 21  ps30              34778 non-null  float64
 22  phi               34778 non-null  float64
 23  nrf               34778 non-null  float64
 24  nrc               34778 non-null  float64
 25  bpr               34778 non-null  float64
 26  farb              34778 non-null  float64
 27  htbleed           34778 non-null  object 
 28  nf_dmd            34778 non-null  object 
 29  pcnfr_dmd         34778 non-null  object 
 30  w31               34778 non-null  float64
 31  w32               34778 non-null  float64
dtypes: float64(20), object(12)
memory usage: 8.5+ MB
```

In [73]:

```
# scalling the tempararute column into standard format
df['t24'] = df['t24'] + 459.67
```

In [74]:

```
df.loc[:,['t24']]
```

Out[74]:

	t24
0	1101.67
1	1102.02
2	1102.36
3	1101.43
4	1102.21
...	...
34773	1103.01
34774	1103.54
34775	1102.87
34776	1103.35
34777	1103.10

34778 rows × 1 columns

In [75]:

```
df
```

Out[75]:

	dataset	esn	unit	flight_cycle	datetime	operator	depart_icao	destination_icao	hpc_eff_mod	hpc_flow_mod	tra	t2	t24	t3
0	test_FD001	999126	26	1	2017-12-31 18:33:07	AXM	VTBD	VTUV	-0.0027	0.0006	100	518.67	1101.67	1582.8
1	test_FD001	999126	26	2	2017-12-31 19:59:49	AXM	VTUV	VTBD	-0.0029	0.0002	100	518.67	1102.02	1589.0
2	test_FD001	999126	26	3	2017-12-31 21:44:38	AXM	VTBD	VMMC	0.0008	0.0001	100	518.67	1102.36	1590.1
3	test_FD001	999126	26	4	2018-01-01 00:50:02	AXM	VMMC	VTBD	-0.0026	0.0005	100	518.67	1101.43	1583.3
4	test_FD001	999126	26	5	2018-01-01 04:25:32	AXM	VTBD	VTSP	0.0020	0.0005	100	518.67	1102.21	1591.7
...	...	...	...	...	...	...	...	...	...	...	...	...	...	.
34773	train_FD001	999084	84	257	2018-06-06 08:22:54	PGT	LFSB	LTFJ	-0.0027	0.0000	100	518.67	1103.01	1599.8
34774	train_FD001	999084	84	258	2018-06-06 12:01:39	PGT	LTFJ	LTAJ	0.0026	0.0001	100	518.67	1103.54	1598.8
34775	train_FD001	999084	84	259	2018-06-06 14:01:52	PGT	LTAJ	LTFJ	-0.0013	0.0000	100	518.67	1102.87	1605.5
34776	train_FD001	999084	84	260	2018-06-06 18:41:30	PGT	OKBK	LTFJ	-0.0023	0.0001	100	518.67	1103.35	1606.0
34777	train_FD001	999095	95	266	2018-02-08 01:30:23	PGT	LTFJ	EDDS	0.0019	0.0003	100	518.67	1103.10	1609.1

34778 rows × 32 columns

In [76]:

```
df['t24'].value_counts()
```

Out[76]:

```
1102.16    354
1102.21    337
1102.17    336
1102.12    336
1102.34    332
          ...
1104.11      1
1100.93      1
1101.00      1
1103.88      1
1104.01      1
Name: t24, Length: 306, dtype: int64
```

In [77]:

```
df.to_csv("avenginedata.csv",index=False)
```

Inner join

Use a code recipe or visual recipe to INNER JOIN the manufacturing tables to create a table that has the KPIs (Key Performance Indicators) of each part and the engine serial number (or ESN) they associate with.

In [78]:

```
import numpy as np
import pandas as pd


#sets the default autosave frequency in seconds
%autosave 60

import warnings
warnings.filterwarnings('ignore')

pd.set_option('display.max_columns',None)
#pd.set_option('display.max_rows',None)
pd.set_option('display.width', 1000)
np.random.seed(0)
np.set_printoptions(suppress=True)
```

Autosaving every 60 seconds

In [79]:

```
df = pd.read_csv("avenginedata.csv")
```

In [80]:

```
df.head(10)
```

Out[80]:

	dataset	esn	unit	flight_cycle	datetime	operator	depart_icao	destination_icao	hpc_eff_mod	hpc_flow_mod	tra	t2	t24	t30	
0	test_FD001	999126	26	1	2017-12-31 18:33:07	AXM	VTBD	VTUV	-0.0027	0.0006	100	518.67	1101.67	1582.88	13
1	test_FD001	999126	26	2	2017-12-31 19:59:49	AXM	VTUV	VTBD	-0.0029	0.0002	100	518.67	1102.02	1589.01	13
2	test_FD001	999126	26	3	2017-12-31 21:44:38	AXM	VTBD	VMMC	0.0008	0.0001	100	518.67	1102.36	1590.16	13
3	test_FD001	999126	26	4	2018-01-01 00:50:02	AXM	VMMC	VTBD	-0.0026	0.0005	100	518.67	1101.43	1583.37	13
4	test_FD001	999126	26	5	2018-01-01 04:25:32	AXM	VTBD	VTSP	0.0020	0.0005	100	518.67	1102.21	1591.71	14
5	test_FD001	999126	26	6	2018-01-01 06:25:56	AXM	VTSP	VTBD	-0.0032	-0.0001	100	518.67	1101.65	1592.68	13
6	test_FD001	999126	26	7	2018-01-01 08:37:05	AXM	VTBD	VTCC	0.0009	0.0003	100	518.67	1101.73	1583.81	13
7	test_FD001	999126	26	8	2018-01-02 08:52:57	AXM	VTBD	VOTR	0.0035	0.0004	100	518.67	1101.73	1583.40	14
8	test_FD001	999126	26	9	2018-01-05 02:37:00	AXM	VTBD	WSSS	-0.0012	-0.0001	100	518.67	1101.35	1587.24	14
9	test_FD001	999126	26	10	2018-01-05 03:55:29	AXM	VTBD	WSSS	-0.0031	0.0007	100	518.67	1101.91	1584.60	14

In [81]:

```
df1 = df.sample(frac=0.5,random_state=0)
```

In [82]:

```
1 df1
```

Out[82]:

	dataset	esn	unit	flight_cycle	datetime	operator	depart_icao	destination_icao	hpc_eff_mod	hpc_flow_mod	tra	t2	t24	t3
10457	train_FD001	999097	97	194	2018-03-04 01:03:00	AXM	VRMM	VTBD	-0.0008	-0.0002	100	518.67	1103.19	1605.4
24716	train_FD001	999015	15	116	2018-01-16 11:14:19	FRON	KDEN	KDSM	0.0042	-0.0004	100	518.67	1102.48	1590.5
21745	test_FD001	999159	59	15	2018-01-04 17:23:30	FRON	KIND	KMCO	-0.0019	0.0002	100	518.67	1101.59	1581.4
23353	test_FD001	999121	21	76	2018-01-10 18:39:21	FRON	KMSP	KDEN	-0.0007	0.0002	100	518.67	1102.14	1587.8
12172	train_FD001	999054	54	40	2018-01-10 02:50:42	AXM	WIII	WMKK	0.0010	0.0001	100	518.67	1101.58	1586.9
...	...	...	...	...	...	...	...	...	...	...	...	...	...	.
31345	test_FD001	999153	53	147	2018-01-25 11:30:02	PGT	LTFJ	LTAC	-0.0028	0.0000	100	518.67	1102.85	1590.4
11080	train_FD001	999054	54	1	2018-01-01 06:27:22	AXM	VTBD	WMKK	-0.0036	-0.0001	100	518.67	1101.27	1580.9
15982	test_FD001	999131	31	132	2018-04-03 06:26:10	AXM	NaN	NaN	0.0001	0.0003	100	518.67	1102.56	1590.6
3635	test_FD001	999110	10	85	2018-01-11 15:46:00	AXM	WMKK	VMMC	0.0035	-0.0003	100	518.67	1101.73	1583.9
17246	test_FD001	999110	10	170	2018-01-24 07:27:44	AXM	VMMC	WMKK	0.0035	0.0004	100	518.67	1102.19	1586.0

17389 rows × 32 columns

Build a supporting KPI table

In [83]:

```
1 suply_chain = pd.read_csv("av_manufacturing_supply_chain_psql.csv")
```

In [84]:

```
1 suply_chain
```

Out[84]:

	sn	pn	op	part_desc	kc	msmts	max	min
0	7837606115	54321P01	op116	shroud	1	31.983503	33.061659	21.160852
1	5039651920	54321P01	op116	shroud	1	34.456691	33.061659	21.160852
2	7837606115	54321P01	op220	shroud	2	27.895096	30.303501	17.044897
3	5039651920	54321P01	op220	shroud	2	32.920628	30.303501	17.044897
4	9856636092	44321P02	op420	blade	1	12.640872	16.346054	10.600079
...	...	...	...	...	...	...	...	...
63995	6299766913	54321P01	op116	shroud	1	20.554360	33.061659	21.160852
63996	4512061920	54321P01	op116	shroud	1	22.756896	33.061659	21.160852
63997	6299766913	54321P01	op220	shroud	2	29.583411	30.303501	17.044897
63998	4512061920	54321P01	op220	shroud	2	30.475523	30.303501	17.044897
63999	4567412522	44321P02	op016	blade	2	19.191280	27.987527	11.183152

64000 rows × 8 columns

In [85]:

```
1 suply_chain1 = suply_chain.sample(frac=0.05, random_state=0)
```

In [86]:

```
1 suply_chain1
```

Out[86]:

	sn	pn	op	part_desc	kc	msmts	max	min
11277	5897563181	54321P01	op116	shroud	1	24.172607	33.061659	21.160852
55819	3880534351	65421P11	op630	disk	120	141.158446	271.153922	99.827763
43223	4370057280	54321P01	op220	shroud	2	25.754601	30.303501	17.044897
1351	360209222	54321P01	op116	shroud	1	26.147983	33.061659	21.160852
9247	2915142739	65421P11	op630	disk	87	123.437705	271.153922	99.827763
...	...	...	...	...	...	...	...	...
41757	1140450245	54321P01	op220	shroud	2	25.778052	30.303501	17.044897
56906	9495107047	44321P02	op420	blade	1	15.922864	16.346054	10.600079
51743	944648874	65421P11	op630	disk	114	171.295726	271.153922	99.827763
21167	6936049070	54321P01	op116	shroud	1	22.841329	33.061659	21.160852
48668	3335633849	44321P02	op016	blade	2	30.345865	27.987527	11.183152

3200 rows × 8 columns

In [87]:

```
1 av_bom = pd.read_csv("av_bom_manufacturing_psql.csv")
```

In [88]:

```
1 av_bom
```

Out[88]:

	esn	pn	sn	desc	vstream
0	999010	54321P01	822106416	shroud	cmc
1	999010	54321P01	664475698	shroud	cmc
2	999010	54321P01	2430976214	shroud	cmc
3	999010	54321P01	1277358392	shroud	cmc
4	999010	54321P01	8668054501	shroud	cmc
...	...	...	...	...	...
20195	999093	44321P02	1003439575	blade	machined_airfoils
20196	999093	44321P02	3829220140	blade	machined_airfoils
20197	999093	44321P02	4571829989	blade	machined_airfoils
20198	999093	44321P02	8136478509	blade	machined_airfoils
20199	999093	65421P11	4508428560	disk	rotating_parts

20200 rows × 5 columns

In [89]:

```
1 av_bom1 = av_bom.sample(frac=0.1,random_state=0)
```

Merging table with inner join

In [90]:

```
1 kpi_table = pd.merge(left=suply_chain1, right=av_bom1, how='inner',on='pn')
```

In [91]:

```
kpi_table
```

Out[91]:

	sn_x	pn	op	part_desc	kc	msmts	max	min	esn	sn_y	desc	vstream
0	5897563181	54321P01	op116	shroud	1	24.172607	33.061659	21.160852	999046	6804819408	shroud	cmc
1	5897563181	54321P01	op116	shroud	1	24.172607	33.061659	21.160852	999063	8776103838	shroud	cmc
2	5897563181	54321P01	op116	shroud	1	24.172607	33.061659	21.160852	999095	5414855032	shroud	cmc
3	5897563181	54321P01	op116	shroud	1	24.172607	33.061659	21.160852	999021	122236869	shroud	cmc
4	5897563181	54321P01	op116	shroud	1	24.172607	33.061659	21.160852	999011	8282194799	shroud	cmc
...	...	...	...	...	...	...	...	...	...	...	...	...
2104525	3335633849	44321P02	op016	blade	2	30.345865	27.987527	11.183152	999046	6279092079	blade	machined_airfoils
2104526	3335633849	44321P02	op016	blade	2	30.345865	27.987527	11.183152	999118	7238358029	blade	machined_airfoils
2104527	3335633849	44321P02	op016	blade	2	30.345865	27.987527	11.183152	999042	3450305789	blade	machined_airfoils
2104528	3335633849	44321P02	op016	blade	2	30.345865	27.987527	11.183152	999083	934140491	blade	machined_airfoils
2104529	3335633849	44321P02	op016	blade	2	30.345865	27.987527	11.183152	999178	4761692240	blade	machined_airfoils

2104530 rows × 12 columns

In [92]:

```
kpi_table.duplicated().sum()
```

Out[92]:

```
0
```

In [93]:

```
kpi_table1 = kpi_table.sample(frac=0.05, random_state=0)
```

In [94]:

```
kpi_table1
```

Out[94]:

	sn_x	pn	op	part_desc	kc	msmts	max	min	esn	sn_y	desc	vstream
697164	3628661032	54321P01	op116	shroud	1	30.978226	33.061659	21.160852	999146	3905519313	shroud	cmc
1319265	1823811079	54321P01	op220	shroud	2	23.890846	30.303501	17.044897	999159	7981949910	shroud	cmc
388623	4591973097	54321P01	op220	shroud	2	20.538837	30.303501	17.044897	999099	826350563	shroud	cmc
1023015	5235783172	54321P01	op220	shroud	2	17.307180	30.303501	17.044897	999093	896556326	shroud	cmc
47153	8868037964	54321P01	op116	shroud	1	18.394660	33.061659	21.160852	999013	761257699	shroud	cmc
...	...	...	...	...	...	...	...	...	...	...	...	...
1130642	9661359233	54321P01	op116	shroud	1	31.282314	33.061659	21.160852	999106	7816334277	shroud	cmc
1092772	7640548765	54321P01	op116	shroud	1	27.368041	33.061659	21.160852	999159	4544512214	shroud	cmc
80923	5177880439	54321P01	op220	shroud	2	29.283795	30.303501	17.044897	999064	37027425	shroud	cmc
1951398	9788484674	44321P02	op016	blade	2	24.368014	27.987527	11.183152	999095	8355436195	blade	machined_airfoils
1592697	5077460631	44321P02	op420	blade	1	13.545880	16.346054	10.600079	999193	9829537549	blade	machined_airfoils

105226 rows × 12 columns

Join the new KPI table to the consolidated flights table

In [95]:

```
df1.head(1)
```

Out[95]:

	dataset	esn	unit	flight_cycle	datetime	operator	depart_icao	destination_icao	hpc_eff_mod	hpc_flow_mod	tra	t2	t24	t3
10457	train_FD001	999097	97	194	2018-03-04 01:03:00	AXM	VRMM	VTBD	-0.0008	-0.0002	100	518.67	1103.19	1605.4

In [96]:

```
kpi_table1.head(1)
```

Out[96]:

	sn_x	pn	op	part_desc	kc	msmts	max	min	esn	sn_y	desc	vstream
697164	3628661032	54321P01	op116	shroud	1	30.978226	33.061659	21.160852	999146	3905519313	shroud	cmc

In [97]:

```
# joining above two table
df2 = pd.merge(left=df1, right=kpi_table1, how='inner',on='esn')
```

In [98]:

```
df2
```

Out[98]:

	dataset	esn	unit	flight_cycle	datetime	operator	depart_icao	destination_icao	hpc_eff_mod	hpc_flow_mod	tra	t2	t24
0	train_FD001	999097	97	194	2018-03-04 01:03:00	AXM	VRMM	VTBD	-0.0008	-0.0002	100	518.67	1103.19	160
1	train_FD001	999097	97	194	2018-03-04 01:03:00	AXM	VRMM	VTBD	-0.0008	-0.0002	100	518.67	1103.19	160
2	train_FD001	999097	97	194	2018-03-04 01:03:00	AXM	VRMM	VTBD	-0.0008	-0.0002	100	518.67	1103.19	160
3	train_FD001	999097	97	194	2018-03-04 01:03:00	AXM	VRMM	VTBD	-0.0008	-0.0002	100	518.67	1103.19	160
4	train_FD001	999097	97	194	2018-03-04 01:03:00	AXM	VRMM	VTBD	-0.0008	-0.0002	100	518.67	1103.19	160
...	...	...	...	...	...	...	...	...	...	...	...	...	...
9094606	test_FD001	999101	1	11	2018-01-02 23:28:31	PGT	LTFJ	LTFH	0.0007	-0.0004	100	518.67	1101.71	158
9094607	test_FD001	999101	1	11	2018-01-02 23:28:31	PGT	LTFJ	LTFH	0.0007	-0.0004	100	518.67	1101.71	158
9094608	test_FD001	999101	1	11	2018-01-02 23:28:31	PGT	LTFJ	LTFH	0.0007	-0.0004	100	518.67	1101.71	158
9094609	test_FD001	999101	1	11	2018-01-02 23:28:31	PGT	LTFJ	LTFH	0.0007	-0.0004	100	518.67	1101.71	158
9094610	test_FD001	999101	1	11	2018-01-02 23:28:31	PGT	LTFJ	LTFH	0.0007	-0.0004	100	518.67	1101.71	158

9094611 rows × 43 columns

Join final table with RUL table to get remaining useful life (RUL) for each engine

In [99]:

```
rul = pd.read_csv("av_esn_rul_psql.csv")
```

In [43]:

```
rul
```

Out[43]:

	esn	rul
0	999175	123
1	999197	95
2	999123	141
3	999122	122
4	999126	162
...	...	...
95	999124	20
96	999151	134
97	999186	80
98	999168	8
99	999111	150

100 rows × 2 columns

In [44]:

```
df1.head(1)
```

Out[44]:

	dataset	esn	unit	flight_cycle	datetime	operator	depart_icao	destination_icao	hpc_eff_mod	hpc_flow_mod	tra	t2	t24	t3
10457	train_FD001	999097	97	194	2018-03-04 01:03:00	AXM	VRMM	VTBD	-0.0008	-0.0002	100	518.67	1103.19	1605.4

In [45]:

```
final_df = pd.merge(left=df1, right=rul, how='inner',on='esn')
```

In [46]:

```
final_df
```

Out[46]:

	dataset	esn	unit	flight_cycle	datetime	operator	depart_icao	destination_icao	hpc_eff_mod	hpc_flow_mod	tra	t2	t24	t30
0	test_FD001	999159	59	15	2018-01-04 17:23:30	FRON	KIND	KMCO	-0.0019	0.0002	100	518.67	1101.59	1581.43
1	test_FD001	999159	59	1	2017-12-31 20:29:57	FRON	KMCO	KCVG	-0.0008	-0.0004	100	518.67	1101.92	1589.29
2	test_FD001	999159	59	55	2018-01-09 12:13:13	FRON	KLAX	KATL	0.0008	-0.0001	100	518.67	1101.80	1577.44
3	test_FD001	999159	59	6	2018-01-02 06:49:26	FRON	KMKE	KMIA	-0.0009	-0.0002	100	518.67	1101.40	1579.85
4	test_FD001	999159	59	66	2018-01-10 09:13:00	FRON	KSAT	KONT	0.0003	0.0001	100	518.67	1101.92	1585.21
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
7858	test_FD001	999101	1	29	2018-01-05 07:48:43	PGT	LTAU	LTFJ	0.0014	0.0001	100	518.67	1101.62	1587.15
7859	test_FD001	999101	1	24	2018-01-04 12:21:10	PGT	LTCA	LTFJ	-0.0006	-0.0001	100	518.67	1101.99	1594.29
7860	test_FD001	999101	1	4	2018-01-01 07:27:35	PGT	EDDF	LTFJ	0.0042	0.0000	100	518.67	1102.11	1584.12
7861	test_FD001	999101	1	31	2018-01-05 11:49:42	PGT	LTAU	LTFJ	-0.0006	0.0004	100	518.67	1102.25	1581.22
7862	test_FD001	999101	1	11	2018-01-02 23:28:31	PGT	LTFJ	LTFH	0.0007	-0.0004	100	518.67	1101.71	1581.03

7863 rows × 33 columns

In [47]:

```
final_df.to_csv("final.csv",index=False)
```

In [48]:

```
df.isnull().sum()
```

Out[48]:

```
dataset                0
esn                    0
unit                   0
flight_cycle           0
datetime               0
operator               0
depart_icao          987
destination_icao    1411
hpc_eff_mod            0
hpc_flow_mod           0
tra                    0
t2                     0
t24                    0
t30                    0
t50                    0
p2                     0
p15                    0
p30                    0
nf                     0
nc                     0
epr                    0
ps30                   0
phi                    0
nrf                    0
nrc                    0
bpr                    0
farb                   0
htbleed                0
nf_dmd                 0
pcnfr_dmd              0
w31                    0
w32                    0
dtype: int64
```

In [49]:

```
df.nunique()
```

Out[49]:

```
dataset                 2
esn                   160
unit                   95
flight_cycle          303
datetime            16350
operator                3
depart_icao           248
destination_icao      245
hpc_eff_mod           162
hpc_flow_mod           14
tra                     1
t2                      1
t24                   306
t30                  3012
t50                  4019
p2                      1
p15                     2
p30                   508
nf                     51
nc                   6308
epr                     1
ps30                  158
phi                   421
nrf                    54
nrc                  5956
bpr                  1914
farb                    1
htbleed                12
nf_dmd                  1
pcnfr_dmd               1
w31                   123
w32                  4740
dtype: int64
```

Exploratory Data Analysis and Data Visualization

In [50]:

```
1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 34778 entries, 0 to 34777
Data columns (total 32 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   dataset           34778 non-null  object 
 1   esn               34778 non-null  int64  
 2   unit              34778 non-null  int64  
 3   flight_cycle      34778 non-null  int64  
 4   datetime          34778 non-null  object 
 5   operator          34778 non-null  object 
 6   depart_icao       33791 non-null  object 
 7   destination_icao  33367 non-null  object 
 8   hpc_eff_mod       34778 non-null  float64
 9   hpc_flow_mod      34778 non-null  float64
 10  tra               34778 non-null  int64  
 11  t2                34778 non-null  float64
 12  t24               34778 non-null  float64
 13  t30               34778 non-null  float64
 14  t50               34778 non-null  float64
 15  p2                34778 non-null  float64
 16  p15               34778 non-null  float64
 17  p30               34778 non-null  float64
 18  nf                34778 non-null  float64
 19  nc                34778 non-null  float64
 20  epr               34778 non-null  float64
 21  ps30              34778 non-null  float64
 22  phi               34778 non-null  float64
 23  nrf               34778 non-null  float64
 24  nrc               34778 non-null  float64
 25  bpr               34778 non-null  float64
 26  farb              34778 non-null  float64
 27  htbleed           34778 non-null  int64  
 28  nf_dmd            34778 non-null  int64  
 29  pcnfr_dmd         34778 non-null  int64  
 30  w31               34778 non-null  float64
 31  w32               34778 non-null  float64
dtypes: float64(20), int64(7), object(5)
memory usage: 8.5+ MB
```

In [51]:

```
df.describe().T
```

Out[51]:

	count	mean	std	min	25%	50%	75%	max
esn	34778.0	999098.774225	5.770268e+01	999002.0000	999052.000000	999096.0000	999149.0000	999200.0000
unit	34778.0	53.599114	2.950036e+01	1.0000	28.000000	54.0000	80.0000	100.0000
flight_cycle	34778.0	86.693973	5.924258e+01	1.0000	38.000000	78.0000	125.0000	303.0000
hpc_eff_mod	34778.0	-0.000026	2.195340e-03	-0.0087	-0.001500	0.0000	0.0014	0.0087
hpc_flow_mod	34778.0	0.000002	2.928849e-04	-0.0006	-0.000200	0.0000	0.0003	0.0007
tra	34778.0	100.000000	0.000000e+00	100.0000	100.000000	100.0000	100.0000	100.0000
t2	34778.0	518.670000	6.116440e-11	518.6700	518.670000	518.6700	518.6700	518.6700
t24	34778.0	1102.291373	4.810640e-01	1100.8000	1101.950000	1102.2500	1102.5900	1104.1100
t30	34778.0	1589.862846	5.900716e+00	1570.1200	1585.750000	1589.5200	1593.5200	1616.9100
t50	34778.0	1406.505339	6.677583e+00	1384.3900	1401.490000	1406.9100	1411.4450	1425.7450
p2	34778.0	14.620000	3.908041e-12	14.6200	14.620000	14.6200	14.6200	14.6200
p15	34778.0	21.609769	1.500980e-03	21.6000	21.610000	21.6100	21.6100	21.6100
p30	34778.0	553.475444	8.446940e-01	550.3500	552.950000	553.5600	554.0900	555.8100
nf	34778.0	2388.089528	6.867953e-02	2387.8900	2388.040000	2388.0800	2388.1300	2388.5600
nc	34778.0	9063.098529	1.909157e+01	9023.8500	9052.420000	9059.8100	9067.8500	9244.5900
epr	34778.0	1.300000	4.287743e-13	1.3000	1.300000	1.3000	1.3000	1.3000
ps30	34778.0	47.506024	2.540322e-01	46.8400	47.320000	47.4800	47.6500	48.5300
phi	34778.0	521.503129	7.048724e-01	518.8300	521.070000	521.5700	522.0200	523.7600
nrf	34778.0	2388.089404	6.873433e-02	2387.8800	2388.040000	2388.0800	2388.1300	2388.5600
nrc	34778.0	8142.236322	1.650522e+01	8099.9400	8133.050000	8139.9800	8147.1600	8293.7200
bpr	34778.0	8.437699	3.594247e-02	8.3279	8.412125	8.4345	8.4596	8.5836
farb	34778.0	0.030000	2.032778e-14	0.0300	0.030000	0.0300	0.0300	0.0300
htbleed	34778.0	393.031945	1.478561e+00	389.0000	392.000000	393.0000	394.0000	400.0000
nf_dmd	34778.0	2388.000000	0.000000e+00	2388.0000	2388.000000	2388.0000	2388.0000	2388.0000
pcnfr_dmd	34778.0	100.000000	0.000000e+00	100.0000	100.000000	100.0000	100.0000	100.0000
w31	34778.0	38.836900	1.727510e-01	38.1400	38.730000	38.8500	38.9600	39.4300
w32	34778.0	23.302293	1.031078e-01	22.8942	23.238900	23.3105	23.3756	23.6229

Skewness

In [52]:

```
#skewness check
df.skew()
```

Out[52]:

```
esn              0.070061
unit            -0.136080
flight_cycle     0.683975
hpc_eff_mod     -0.007578
hpc_flow_mod     0.006306
tra              0.000000
t2               0.000000
t24              0.369767
t30              0.333719
t50             -0.167031
p2               0.000000
p15             -6.355405
p30             -0.477177
nf               0.500236
nc               2.768784
epr              0.000000
ps30             0.569353
phi             -0.511116
nrf              0.513532
nrc              2.519770
bpr              0.442942
farb             0.000000
htbleed          0.385959
nf_dmd           0.000000
pcnfr_dmd        0.000000
w31             -0.409259
w32             -0.422198
dtype: float64
```

Covariance

In [53]:

```
df.cov()
```

Out[53]:

	esn	unit	flight_cycle	hpc_eff_mod	hpc_flow_mod	tra	t2	t24	t30	t50	
esn	3.329599e+03	8.615394e+02	-2.650114e+02	-7.789060e-04	1.256023e-05	0.0	0.0	-5.428390e+00	-6.520881e+01	-7.613817e+01	-5.0282
unit	8.615394e+02	8.702712e+02	5.800668e+01	-1.244464e-03	-1.006595e-05	0.0	0.0	1.133348e+00	1.211771e+01	2.196283e+01	1.754164
flight_cycle	-2.650114e+02	5.800668e+01	3.509683e+03	-6.613827e-04	-4.816860e-05	0.0	0.0	1.284017e+01	1.550586e+02	1.909896e+02	-1.5460
hpc_eff_mod	-7.789060e-04	-1.244464e-03	-6.613827e-04	4.819516e-06	-1.220166e-09	0.0	0.0	8.308540e-06	-5.429758e-05	1.069250e-04	-1.5594
hpc_flow_mod	1.256023e-05	-1.006595e-05	-4.816860e-05	-1.220166e-09	8.578156e-08	0.0	0.0	-1.331272e-06	-6.243348e-06	-1.525232e-05	2.126570
tra	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.0	0.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000
t2	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.0	0.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000
t24	-5.428390e+00	1.133348e+00	1.284017e+01	8.308540e-06	-1.331272e-06	0.0	0.0	2.314225e-01	1.624588e+00	2.136656e+00	4.182632
t30	-6.520881e+01	1.211771e+01	1.550586e+02	-5.429758e-05	-6.243348e-06	0.0	0.0	1.624588e+00	3.481845e+01	2.429011e+01	2.962010
t50	-7.613817e+01	2.196283e+01	1.909896e+02	1.069250e-04	-1.525232e-05	0.0	0.0	2.136656e+00	2.429011e+01	4.459012e+01	3.887406
p2	-5.028208e-26	1.754164e-28	-1.546043e-28	-1.559484e-35	2.126570e-35	0.0	0.0	4.182632e-28	2.962010e-28	3.887406e-28	1.262214
p15	-1.984378e-03	1.975009e-03	7.573847e-03	-9.696988e-10	-2.216283e-09	0.0	0.0	9.312229e-05	9.975048e-04	1.621091e-03	1.776492
p30	9.518804e+00	-2.780490e+00	-2.425988e+01	-1.541655e-05	9.786530e-07	0.0	0.0	-2.753754e-01	-3.191613e+00	-4.213192e+00	-1.5423
nf	-5.798894e-01	2.560116e-01	1.514814e+00	5.648940e-07	-1.277067e-07	0.0	0.0	2.133364e-02	2.374410e-01	3.290473e-01	-4.7695
nc	-2.413267e+02	-4.634963e+01	4.919717e+02	-2.517502e-04	7.466862e-06	0.0	0.0	2.270118e+00	3.223021e+01	2.886355e+01	7.763663
epr	-3.142630e-27	1.096353e-29	-9.662769e-30	-9.746777e-37	1.329106e-36	0.0	0.0	2.614145e-29	1.851256e-29	2.429629e-29	7.888836
ps30	-3.167883e+00	8.140771e-01	7.876069e+00	8.004031e-06	-8.706183e-07	0.0	0.0	8.749265e-02	1.000589e+00	1.324870e+00	5.946319
phi	8.263874e+00	-2.372454e+00	-2.067258e+01	-1.256272e-06	2.269823e-06	0.0	0.0	-2.385865e-01	-2.720149e+00	-3.626785e+00	4.580524
nrf	-5.448831e-01	2.786071e-01	1.501968e+00	1.388108e-06	-1.652673e-07	0.0	0.0	2.134964e-02	2.362979e-01	3.274677e-01	-1.5659
nrc	-1.852274e+02	-5.026389e+01	3.590624e+02	-2.935142e-04	1.555090e-05	0.0	0.0	1.132351e+00	1.851961e+01	1.206240e+01	1.917688
bpr	-4.200113e-01	9.803274e-02	1.023113e+00	9.020311e-07	-4.936258e-08	0.0	0.0	1.129396e-02	1.289559e-01	1.700020e-01	1.863306
farb	-1.473108e-28	9.000777e-32	-5.981161e-31	-3.322765e-38	3.876559e-38	0.0	0.0	1.225358e-30	8.670506e-31	1.145058e-30	3.697892
htbleed	-1.651491e+01	3.521016e+00	4.086615e+01	1.138679e-05	1.075398e-07	0.0	0.0	4.254408e-01	4.934317e+00	6.403706e+00	-6.2575
nf_dmd	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.0	0.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000
pcnfr_dmd	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.0	0.0	0.000000e+00	0.000000e+00	0.000000e+00	0.000000
w31	1.950448e+00	-4.545839e-01	-4.880340e+00	-2.864759e-06	1.285292e-07	0.0	0.0	-5.284525e-02	-6.054956e-01	-8.006554e-01	-8.5391
w32	1.198976e+00	-2.607224e-01	-2.926587e+00	-1.720644e-06	3.179034e-07	0.0	0.0	-3.181004e-02	-3.641741e-01	-4.759632e-01	-8.9369

In [54]:

```
df.describe(include = 'all').T
```

Out[54]:

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
dataset	34778	2	train_FD001	19067	NaN	NaN	NaN	NaN	NaN	NaN	NaN
esn	34778.0	NaN	NaN	NaN	999098.774225	57.702679	999002.0	999052.0	999096.0	999149.0	999200.0
unit	34778.0	NaN	NaN	NaN	53.599114	29.50036	1.0	28.0	54.0	80.0	100.0
flight_cycle	34778.0	NaN	NaN	NaN	86.693973	59.24258	1.0	38.0	78.0	125.0	303.0
datetime	34778	16350	2018-01-16 17:18:00	12	NaN	NaN	NaN	NaN	NaN	NaN	NaN
operator	34778	3	AXM	21188	NaN	NaN	NaN	NaN	NaN	NaN	NaN
depart_icao	33791	248	WMKK	8408	NaN	NaN	NaN	NaN	NaN	NaN	NaN
destination_icao	33367	245	WMKK	6868	NaN	NaN	NaN	NaN	NaN	NaN	NaN
hpc_eff_mod	34778.0	NaN	NaN	NaN	-0.000026	0.002195	-0.0087	-0.0015	0.0	0.0014	0.0087
hpc_flow_mod	34778.0	NaN	NaN	NaN	0.000002	0.000293	-0.0006	-0.0002	0.0	0.0003	0.0007
tra	34778.0	NaN	NaN	NaN	100.0	0.0	100.0	100.0	100.0	100.0	100.0
t2	34778.0	NaN	NaN	NaN	518.67	0.0	518.67	518.67	518.67	518.67	518.67
t24	34778.0	NaN	NaN	NaN	1102.291373	0.481064	1100.8	1101.95	1102.25	1102.59	1104.11
t30	34778.0	NaN	NaN	NaN	1589.862846	5.900716	1570.12	1585.75	1589.52	1593.52	1616.91
t50	34778.0	NaN	NaN	NaN	1406.505339	6.677583	1384.39	1401.49	1406.91	1411.445	1425.745
p2	34778.0	NaN	NaN	NaN	14.62	0.0	14.62	14.62	14.62	14.62	14.62
p15	34778.0	NaN	NaN	NaN	21.609769	0.001501	21.6	21.61	21.61	21.61	21.61
p30	34778.0	NaN	NaN	NaN	553.475444	0.844694	550.35	552.95	553.56	554.09	555.81
nf	34778.0	NaN	NaN	NaN	2388.089528	0.06868	2387.89	2388.04	2388.08	2388.13	2388.56
nc	34778.0	NaN	NaN	NaN	9063.098529	19.091571	9023.85	9052.42	9059.81	9067.85	9244.59
epr	34778.0	NaN	NaN	NaN	1.3	0.0	1.3	1.3	1.3	1.3	1.3
ps30	34778.0	NaN	NaN	NaN	47.506024	0.254032	46.84	47.32	47.48	47.65	48.53
phi	34778.0	NaN	NaN	NaN	521.503129	0.704872	518.83	521.07	521.57	522.02	523.76
nrf	34778.0	NaN	NaN	NaN	2388.089404	0.068734	2387.88	2388.04	2388.08	2388.13	2388.56
nrc	34778.0	NaN	NaN	NaN	8142.236322	16.505216	8099.94	8133.05	8139.98	8147.16	8293.72
bpr	34778.0	NaN	NaN	NaN	8.437699	0.035942	8.3279	8.412125	8.4345	8.4596	8.5836
farb	34778.0	NaN	NaN	NaN	0.03	0.0	0.03	0.03	0.03	0.03	0.03
htbleed	34778.0	NaN	NaN	NaN	393.031945	1.478561	389.0	392.0	393.0	394.0	400.0
nf_dmd	34778.0	NaN	NaN	NaN	2388.0	0.0	2388.0	2388.0	2388.0	2388.0	2388.0
pcnfr_dmd	34778.0	NaN	NaN	NaN	100.0	0.0	100.0	100.0	100.0	100.0	100.0
w31	34778.0	NaN	NaN	NaN	38.8369	0.172751	38.14	38.73	38.85	38.96	39.43
w32	34778.0	NaN	NaN	NaN	23.302293	0.103108	22.8942	23.2389	23.3105	23.3756	23.6229

In [55]:

```
df.columns
```

Out[55]:

```
Index(['dataset', 'esn', 'unit', 'flight_cycle', 'datetime', 'operator', 'depart_icao', 'destination_icao', 'hpc_eff_mod',
'hpc_flow_mod', 'tra', 't2', 't24', 't30', 't50', 'p2', 'p15', 'p30', 'nf', 'nc', 'epr', 'ps30', 'phi', 'nrf', 'nrc', 'bp
r', 'farb', 'htbleed', 'nf_dmd', 'pcnfr_dmd', 'w31', 'w32'], dtype='object')
```

Correlation

In [56]:

```
df.corr()
```

Out[56]:

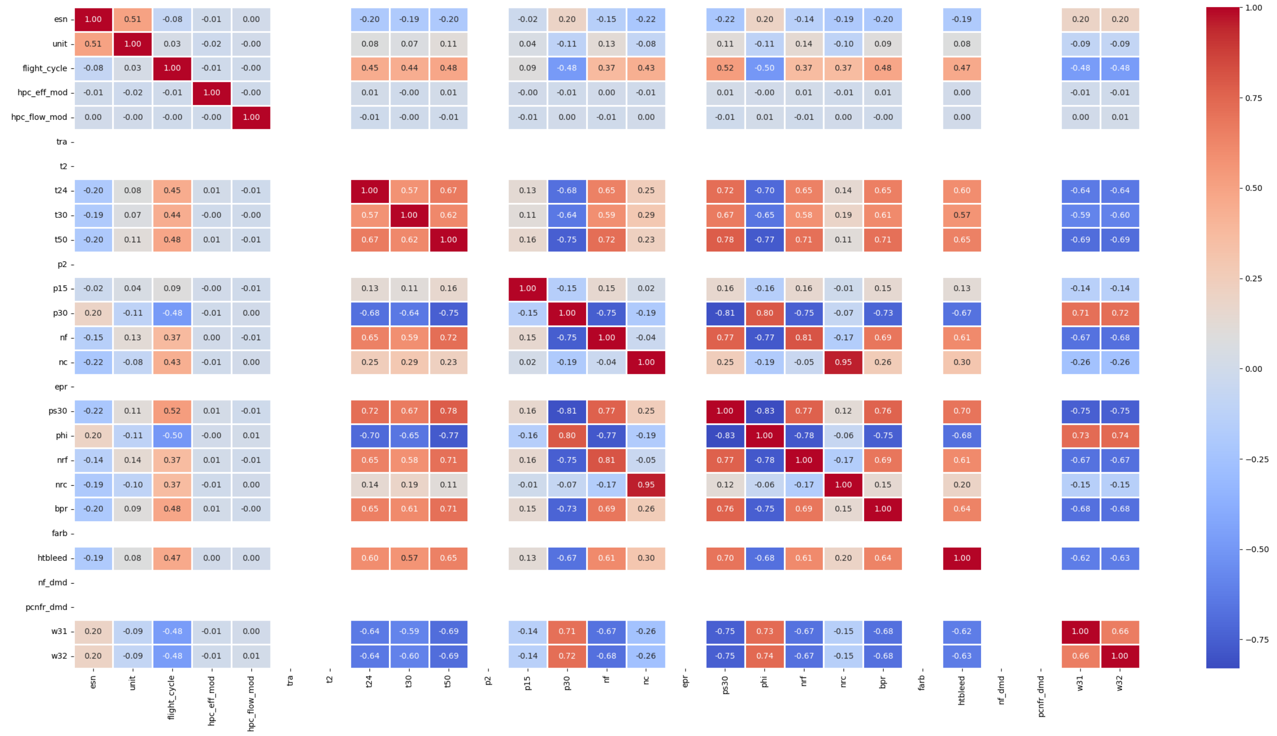
	esn	unit	flight_cycle	hpc_eff_mod	hpc_flow_mod	tra	t2	t24	t30	t50	p2	p15	p30	
esn	1.000000	0.506118	-0.077524	-0.006149	0.000743	NaN	NaN	-0.195556	-0.191516	-0.197600	NaN	-0.022911	0.195293	-0.1
unit	0.506118	1.000000	0.033191	-0.019216	-0.001165	NaN	NaN	0.079861	0.069613	0.111491	NaN	0.044603	-0.111582	0.1
flight_cycle	-0.077524	0.033191	1.000000	-0.005085	-0.002776	NaN	NaN	0.450541	0.443565	0.482788	NaN	0.085174	-0.484792	0.3
hpc_eff_mod	-0.006149	-0.019216	-0.005085	1.000000	-0.001898	NaN	NaN	0.007867	-0.004192	0.007294	NaN	-0.000294	-0.008314	0.0
hpc_flow_mod	0.000743	-0.001165	-0.002776	-0.001898	1.000000	NaN	NaN	-0.009449	-0.003613	-0.007799	NaN	-0.005041	0.003956	-0.0
tra	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
t2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
t24	-0.195556	0.079861	0.450541	0.007867	-0.009449	NaN	NaN	1.000000	0.572316	0.665139	NaN	0.128966	-0.677677	0.6
t30	-0.191516	0.069613	0.443565	-0.004192	-0.003613	NaN	NaN	0.572316	1.000000	0.616461	NaN	0.112625	-0.640333	0.5
t50	-0.197600	0.111491	0.482788	0.007294	-0.007799	NaN	NaN	0.665139	0.616461	1.000000	NaN	0.161738	-0.746952	0.7
p2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
p15	-0.022911	0.044603	0.085174	-0.000294	-0.005041	NaN	NaN	0.128966	0.112625	0.161738	NaN	1.000000	-0.153024	0.1
p30	0.195293	-0.111582	-0.484792	-0.008314	0.003956	NaN	NaN	-0.677677	-0.640333	-0.746952	NaN	-0.153024	1.000000	-0.7
nf	-0.146326	0.126359	0.372304	0.003747	-0.006349	NaN	NaN	0.645706	0.585900	0.717483	NaN	0.150408	-0.753262	1.0
nc	-0.219062	-0.082296	0.434975	-0.006007	0.001335	NaN	NaN	0.247175	0.286099	0.226406	NaN	0.016518	-0.193214	-0.0
epr	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
ps30	-0.216115	0.108630	0.523344	0.014352	-0.011702	NaN	NaN	0.715946	0.667517	0.781025	NaN	0.159529	-0.808395	0.7
phi	0.203178	-0.114093	-0.495051	-0.000812	0.010995	NaN	NaN	-0.703611	-0.654000	-0.770534	NaN	-0.155699	0.797268	-0.7
nrf	-0.137383	0.137401	0.368853	0.009199	-0.008209	NaN	NaN	0.645675	0.582615	0.713470	NaN	0.155946	-0.750544	0.8
nrc	-0.194486	-0.103230	0.367210	-0.008100	0.003217	NaN	NaN	0.142612	0.190154	0.109444	NaN	-0.009655	-0.071348	-0.1
bpr	-0.202515	0.092456	0.480487	0.011432	-0.004689	NaN	NaN	0.653184	0.608035	0.708316	NaN	0.146051	-0.727624	0.6
farb	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
htbleed	-0.193571	0.080724	0.466542	0.003508	0.000248	NaN	NaN	0.598132	0.565566	0.648594	NaN	0.125890	-0.668312	0.6
nf_dmd	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
pcnfr_dmd	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
w31	0.195667	-0.089200	-0.476865	-0.007554	0.002540	NaN	NaN	-0.635891	-0.593999	-0.694074	NaN	-0.137595	0.711355	-0.6
w32	0.201522	-0.085716	-0.479111	-0.007601	0.010527	NaN	NaN	-0.641312	-0.598567	-0.691293	NaN	-0.141173	0.715259	-0.6

In [101]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
import datetime
```

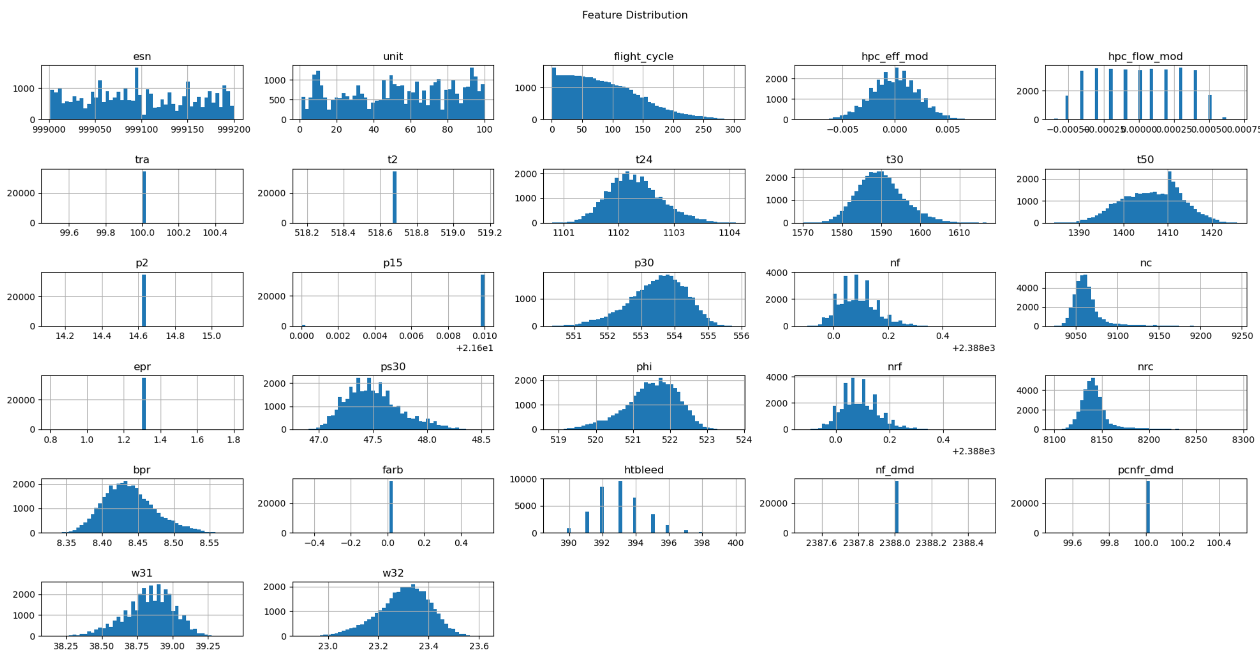
In [102]:

```
plt.figure(figsize=(30,15))
sns.heatmap(df.corr(),cmap="coolwarm",annot=True,fmt='.2f',linewidths=2)
plt.show()
```



In [103]:

```
df.hist(bins=50, figsize=(20,10))
plt.suptitle('Feature Distribution', x=0.5, y=1.02, ha='center', fontsize='large')
plt.tight_layout()
plt.show()
```



In [104]:

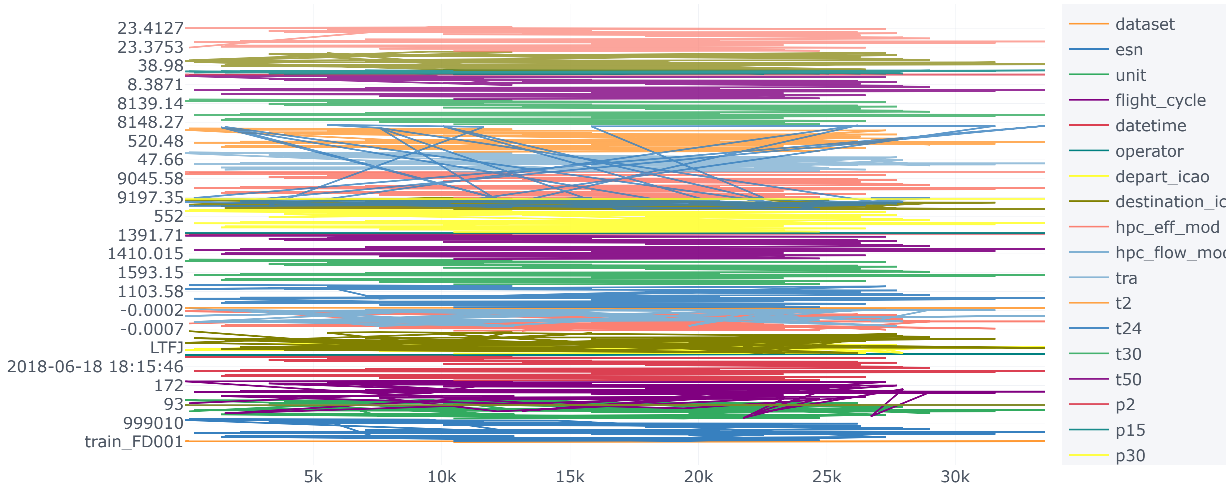
```
import cufflinks as cf
```

In [105]:

```
cf.go_offline()
```

In [106]:

```
1 df.sample(50).iplot()
```

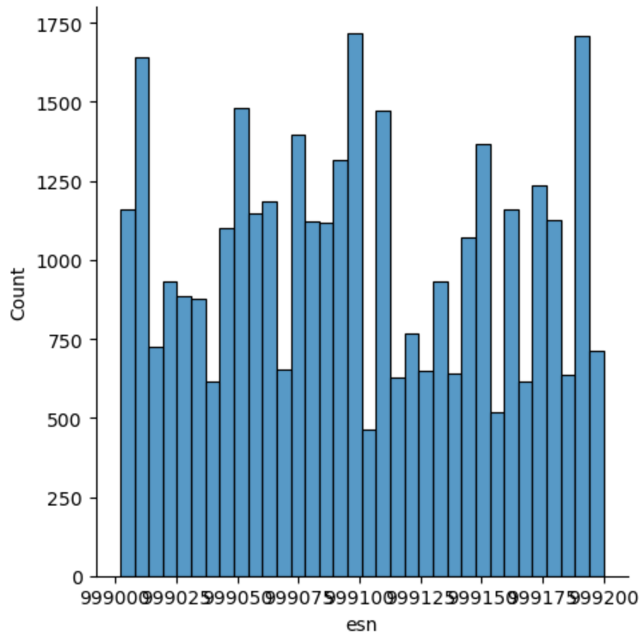


In [107]:

```
1 sns.displot(df['esn'])
```

Out[107]:

<seaborn.axisgrid.FacetGrid at 0x1be9b685730>

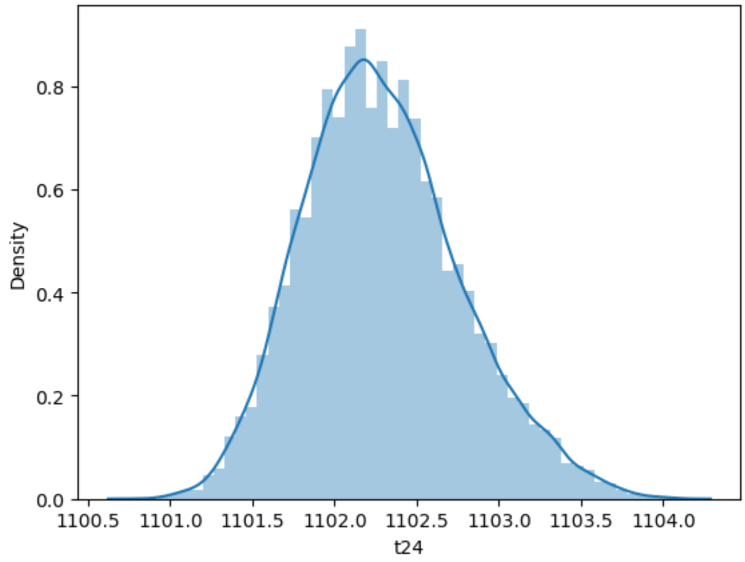


In [108]:

```
sns.distplot(df['t24'])
```

Out[108]:

```
<AxesSubplot:xlabel='t24', ylabel='Density'>
```

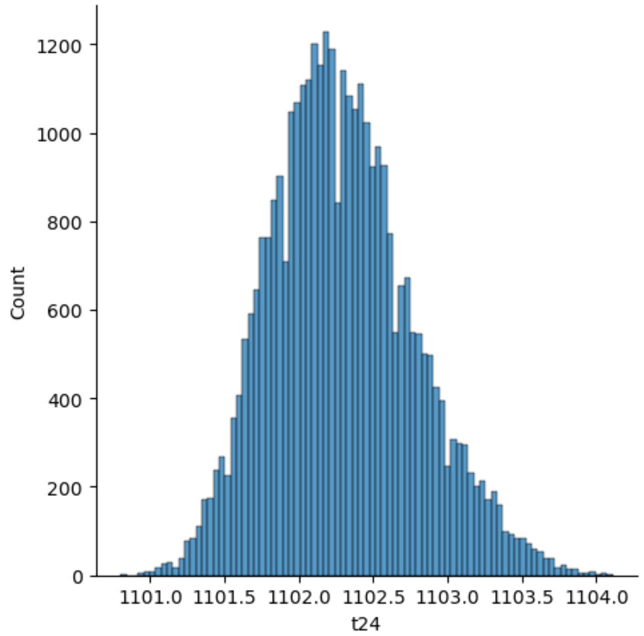


In [109]:

```
sns.displot(df['t24'])
```

Out[109]:

```
<seaborn.axisgrid.FacetGrid at 0x1be971d18b0>
```



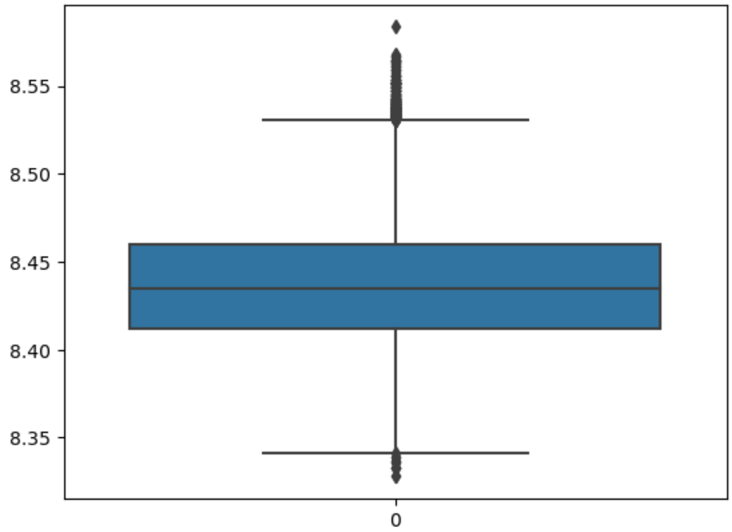
Outliers

In [110]:

```
1 sns.boxplot(data=df['bpr'])
```

Out[110]:

<AxesSubplot:>



In [111]:

```
1 df.head(1)
```

Out[111]:

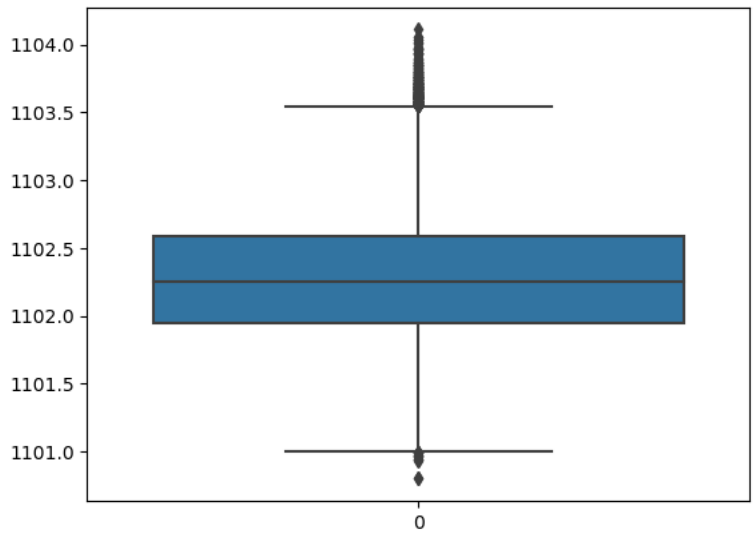
	dataset	esn	unit	flight_cycle	datetime	operator	depart_icao	destination_icao	hpc_eff_mod	hpc_flow_mod	tra	t2	t24	t30	
0	test_FD001	999126	26	1	2017-12-31 18:33:07	AXM	VTBD	VTUV	-0.0027	0.0006	100	518.67	1101.67	1582.88	13

In [112]:

```
1 sns.boxplot(data=df['t24'])
```

Out[112]:

<AxesSubplot:>

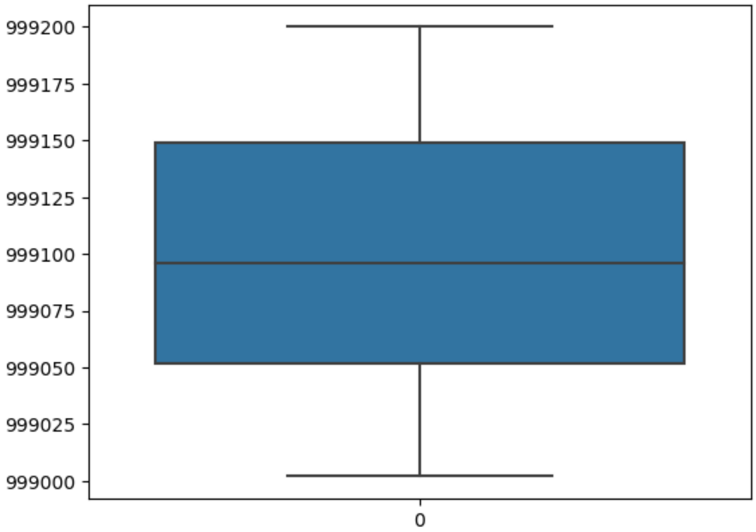


In [113]:

```
sns.boxplot(data=df['esn'])
```

Out[113]:

```
<AxesSubplot:>
```

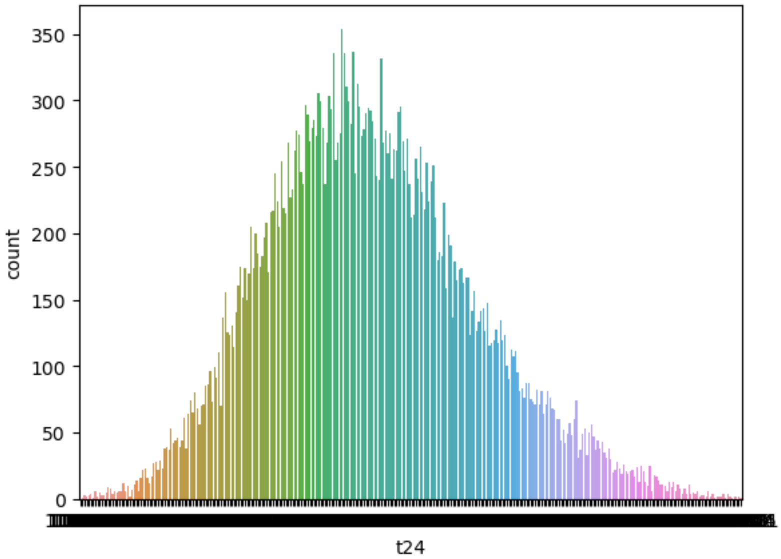


In [114]:

```
sns.countplot(df['t24'])
```

Out[114]:

```
<AxesSubplot:xlabel='t24', ylabel='count'>
```

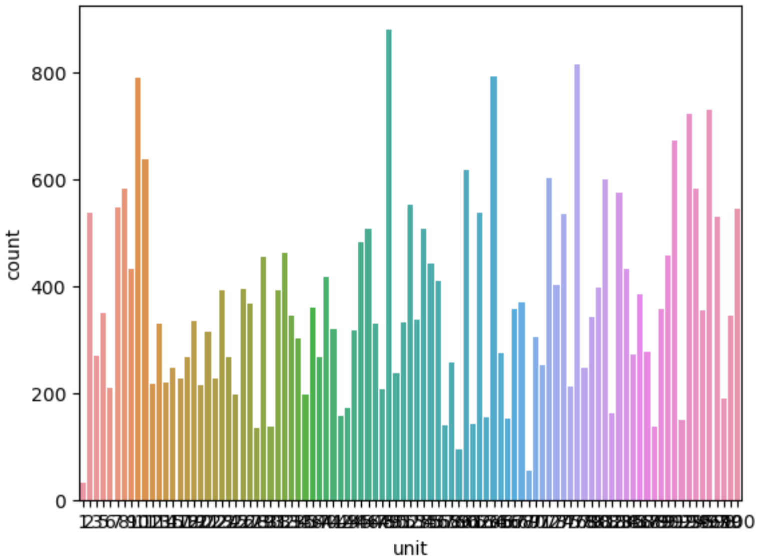


In [115]:

```
sns.countplot(df['unit'])
```

Out[115]:

<AxesSubplot:xlabel='unit', ylabel='count'>



In [ ]:

```

```

In [ ]:

```

```