

In [40]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import os
```

In [41]:

```
os.chdir('C:\\Users\\USER\\Desktop\\MobileDataSets')
df=pd.read_csv('MobileTrain.csv')
```

In [42]:

df

Out[42]:

ck_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt	n_cores	...	px_height	px_wid
2.2	0	1	0	7	0.6	188	2	...	20	7
0.5	1	0	1	53	0.7	136	3	...	905	19
0.5	1	2	1	41	0.9	145	5	...	1263	17
2.5	0	0	0	10	0.8	131	6	...	1216	17
1.2	0	13	1	44	0.6	141	2	...	1208	12
...
0.5	1	0	1	2	0.8	106	6	...	1222	18
2.6	1	0	0	39	0.2	187	4	...	915	19
0.9	1	1	1	36	0.7	108	8	...	868	16
0.9	0	4	1	46	0.1	145	5	...	336	6
2.0	1	5	1	45	0.9	168	6	...	483	7

In [43]:

```
df.head()
```

Out[43]:

	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt
0	842	0	2.2	0	1	0	7	0.6	188
1	1021	1	0.5	1	0	1	53	0.7	136
2	563	1	0.5	1	2	1	41	0.9	145
3	615	1	2.5	0	0	0	10	0.8	131
4	1821	1	1.2	0	13	1	44	0.6	141

5 rows × 21 columns

In [44]:

```
df.tail()
```

Out[44]:

	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt
1995	794	1	0.5	1	0	1	2	0.8	106
1996	1965	1	2.6	1	0	0	39	0.2	187
1997	1911	0	0.9	1	1	1	36	0.7	108
1998	1512	0	0.9	0	4	1	46	0.1	145
1999	510	1	2.0	1	5	1	45	0.9	168

5 rows × 21 columns

In [45]:

```
df.shape
```

Out[45]:

```
(2000, 21)
```

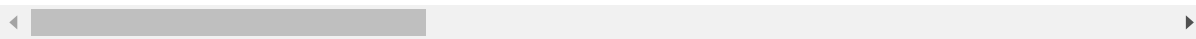
In [46]:

```
df.describe()
```

Out[46]:

	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_men
count	2000.000000	2000.0000	2000.000000	2000.000000	2000.000000	2000.000000	2000.000000
mean	1238.518500	0.4950	1.522250	0.509500	4.309500	0.521500	32.040000
std	439.418206	0.5001	0.816004	0.500035	4.341444	0.499662	18.140000
min	501.000000	0.0000	0.500000	0.000000	0.000000	0.000000	2.000000
25%	851.750000	0.0000	0.700000	0.000000	1.000000	0.000000	16.000000
50%	1226.000000	0.0000	1.500000	1.000000	3.000000	1.000000	32.000000
75%	1615.250000	1.0000	2.200000	1.000000	7.000000	1.000000	48.000000
max	1998.000000	1.0000	3.000000	1.000000	19.000000	1.000000	64.000000

8 rows × 21 columns



In [47]:

```
df.columns
```

Out[47]:

```
Index(['battery_power', 'blue', 'clock_speed', 'dual_sim', 'fc', 'four_g',
       'int_memory', 'm_dep', 'mobile_wt', 'n_cores', 'pc', 'px_height',
       'px_width', 'ram', 'sc_h', 'sc_w', 'talk_time', 'three_g',
       'touch_screen', 'wifi', 'price_range'],
      dtype='object')
```

In [48]:

```
df.nunique()
```

Out[48]:

```
battery_power    1094
blue              2
clock_speed      26
dual_sim         2
fc               20
four_g           2
int_memory       63
m_dep            10
mobile_wt        121
n_cores          8
pc               21
px_height        1137
px_width         1109
ram              1562
sc_h             15
sc_w             19
talk_time        19
three_g          2
touch_screen     2
wifi             2
price_range      4
dtype: int64
```

In [49]:

```
df['n_cores'].unique()
```

Out[49]:

```
array([2, 3, 5, 6, 1, 8, 4, 7], dtype=int64)
```

```
#CLEANING THE DATA
```

In [50]:

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

Out[50]:

```
battery_power    0
blue              0
clock_speed      0
dual_sim         0
fc               0
four_g           0
int_memory       0
m_dep            0
mobile_wt        0
n_cores          0
pc               0
px_height        0
px_width         0
ram              0
sc_h             0
sc_w             0
talk_time        0
three_g          0
touch_screen     0
wifi             0
price_range      0
dtype: int64
```

In [51]:

```
m=df.drop(['blue','clock_speed'],axis=1)
```

m

In [52]:

```
m.head()
```

Out[52]:

	battery_power	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt	n_cores	pc	px_height
0	842	0	1	0	7	0.6	188	2	2	2
1	1021	1	0	1	53	0.7	136	3	6	90
2	563	1	2	1	41	0.9	145	5	6	126
3	615	0	0	0	10	0.8	131	6	9	121
4	1821	0	13	1	44	0.6	141	2	14	120

In [53]:

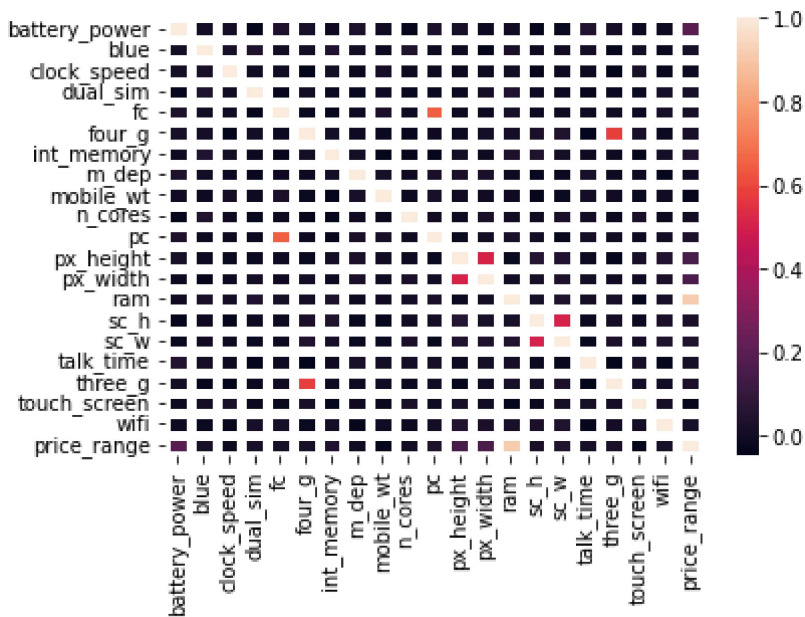
```
corelation=df.corr()
```

In [54]:

```
sns.heatmap(correlation,xticklabels=correlation.columns,yticklabels=correlation.columns,linewidth=0.5)
```

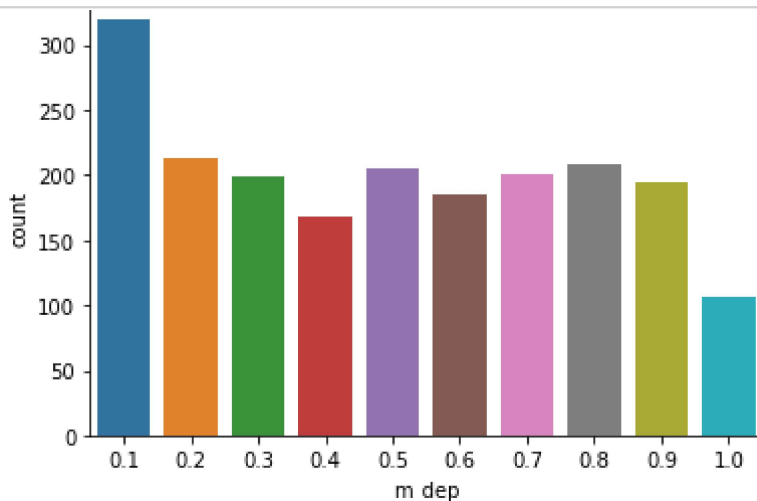
Out[54]:

<AxesSubplot:>



In [56]:

```
for i in df:
    if(df[i].nunique()<50):
        sns.countplot(x=df[i])
        plt.show()
```



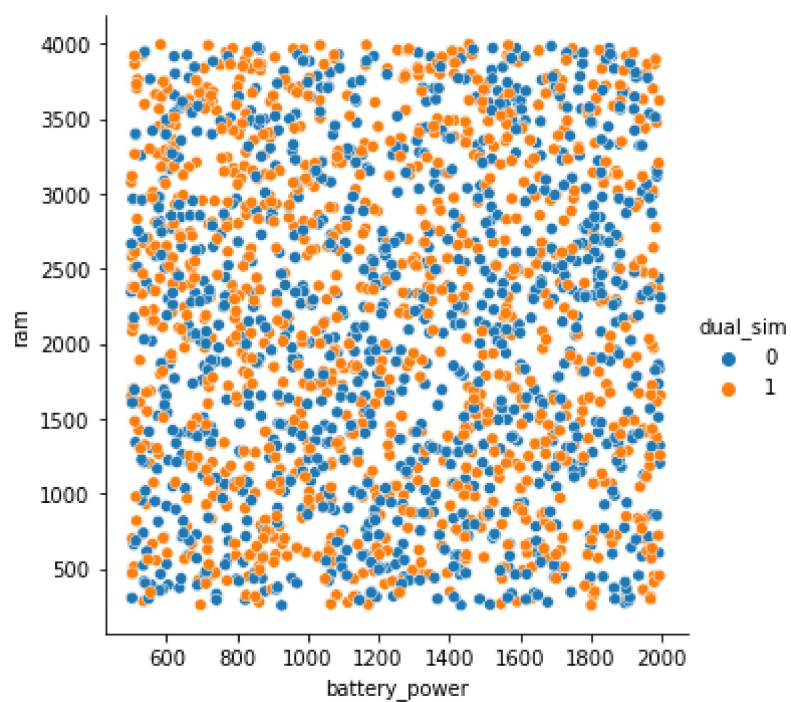
In []:

In [24]:

```
sns.relplot(x='battery_power',y='ram',hue='dual_sim',data=m)
```

Out[24]:

<seaborn.axisgrid.FacetGrid at 0x17ad2351460>



In [67]:

```

import warnings
warnings.filterwarnings('ignore')
plt.figure(figsize=(50,20))
plt.subplot(331)
sns.distplot(df['battery_power'])
plt.subplot(332)
sns.distplot(df['blue'])
plt.subplot(333)
sns.distplot(df['clock_speed'])
plt.subplot(334)
sns.distplot(df['dual_sim'])
plt.subplot(335)
sns.distplot(df['fc'])
plt.subplot(336)
sns.distplot(df['four_g'])
plt.subplot(337)
sns.distplot(df['int_memory'])
plt.subplot(338)
sns.distplot(df['m_dep'])
plt.subplot(339)
sns.distplot(df['mobile_wt'])
plt.subplot(341)
sns.distplot(df['pc'])
plt.subplot(342)
sns.distplot(df['px_height'])
plt.subplot(343)
sns.distplot(df['px_width'])
plt.subplot(344)
sns.distplot(df['ram'])
plt.subplot(345)
sns.distplot(df['sc_h'])
plt.subplot(346)
sns.distplot(df['sc_w'])
plt.subplot(347)
sns.distplot(df['talk_time'])
plt.subplot(348)
sns.distplot(df['three_g'])
plt.subplot(349)
sns.distplot(df['touch_screen'])
plt.subplot(351)
sns.distplot(df['price_range'])
plt.show()

```

