# statistic

October 13, 2022

[ ]:
```
# include the required packages
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.gridspec import GridSpec
import seaborn as sns
```

[ ]:
```
df = pd.read_csv("Mobile Device Data.csv")
df.head()
```

[ ]:

| | Model | Release Date | Release Year | \ |
|---|---|---|---|---|
| 0 | Psion Organiser II LZ 64 | 1989-01-01 | 1989.000000 | |
| 1 | Hewlett-Packard 95LX (HP Jaguar) | 1991-04-01 | 1991.250000 | |
| 2 | Psion Series 3 | 1991-06-01 | 1991.416667 | |
| 3 | Hewlett-Packard 95LX 1MB RAM (HP Jaguar) | 1992-01-01 | 1992.000000 | |
| 4 | Psion Series 3a | 1993-03-01 | 1993.166667 | |

| | Model ID | RAM Capacity (Mb) | Storage (Mb) | CPU Clock (MHz) | \ |
|---|---|---|---|---|---|
| 0 | 1 | 0.000000 | 0.000000 | 0.000000 | |
| 1 | 2 | 0.000214 | 0.000015 | 0.002230 | |
| 2 | 3 | 0.000092 | 0.000015 | 0.001895 | |
| 3 | 4 | 0.000458 | 0.000015 | 0.002230 | |
| 4 | 5 | 0.000214 | 0.000015 | 0.003386 | |

| | Display Diagonal (in) | Display Width(px) | Display Length(px) | Width (mm) | \ |
|---|---|---|---|---|---|
| 0 | 0.050360 | 0.000000 | 0.000000 | 0.369714 | |
| 1 | 0.223022 | 0.056911 | 0.047619 | 0.416098 | |
| 2 | 0.187050 | 0.056911 | 0.023810 | 0.432469 | |
| 3 | 0.223022 | 0.056911 | 0.047619 | 0.416098 | |
| 4 | 0.258993 | 0.154472 | 0.063492 | 0.432469 | |

| | Length (mm) | Depth (mm) | Volume (cubic cm) | Mass (grams) | \ |
|---|---|---|---|---|---|
| 0 | 0.118852 | 0.146889 | 0.059332 | 0.044181 | |
| 1 | 0.147541 | 0.127646 | 0.065799 | 0.053879 | |
| 2 | 0.147541 | 0.101988 | 0.056634 | 0.044181 | |
| 3 | 0.147541 | 0.127646 | 0.065799 | 0.053879 | |
| 4 | 0.147541 | 0.101988 | 0.056634 | 0.046336 | |

```
   Pixel Density (per inch) Company_ID Company Company_real
0                 0.000000      other   other        other
1                 0.030793      other   other        other
2                 0.035336      other   other        other
3                 0.030793      other   other        other
4                 0.129985      other   other        other
```

[ ]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10104 entries, 0 to 10103
Data columns (total 19 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Model                     10104 non-null  object
 1   Release Date              10104 non-null  object
 2   Release Year              10104 non-null  float64
 3   Model ID                  10104 non-null  int64
 4   RAM Capacity (Mb)         10104 non-null  float64
 5   Storage (Mb)              10104 non-null  float64
 6   CPU Clock (MHz)           10104 non-null  float64
 7   Display Diagonal (in)     10104 non-null  float64
 8   Display Width(px)         10104 non-null  float64
 9   Display Length(px)        10104 non-null  float64
 10  Width (mm)                10104 non-null  float64
 11  Length (mm)               10104 non-null  float64
 12  Depth (mm)                10104 non-null  float64
 13  Volume (cubic cm)         10104 non-null  float64
 14  Mass (grams)              10104 non-null  float64
 15  Pixel Density (per inch)  10104 non-null  float64
 16  Company_ID                10104 non-null  object
 17  Company                   10104 non-null  object
 18  Company_real              10104 non-null  object
dtypes: float64(13), int64(1), object(5)
memory usage: 1.5+ MB
```

[ ]:
```
# statistic of the data
df.describe()
```

[ ]:
```
       Release Year      Model ID  RAM Capacity (Mb)  Storage (Mb)  \
count  10104.000000  10104.000000       10104.000000  10104.000000
mean    2008.388707   1452.582838           0.143822      0.049237
std        2.994518    817.839811           0.171839      0.135451
min     1989.000000      1.000000           0.000000      0.000000
25%     2006.416667    767.000000           0.031220      0.002096
50%     2008.416667   1339.000000           0.062471      0.004194
```

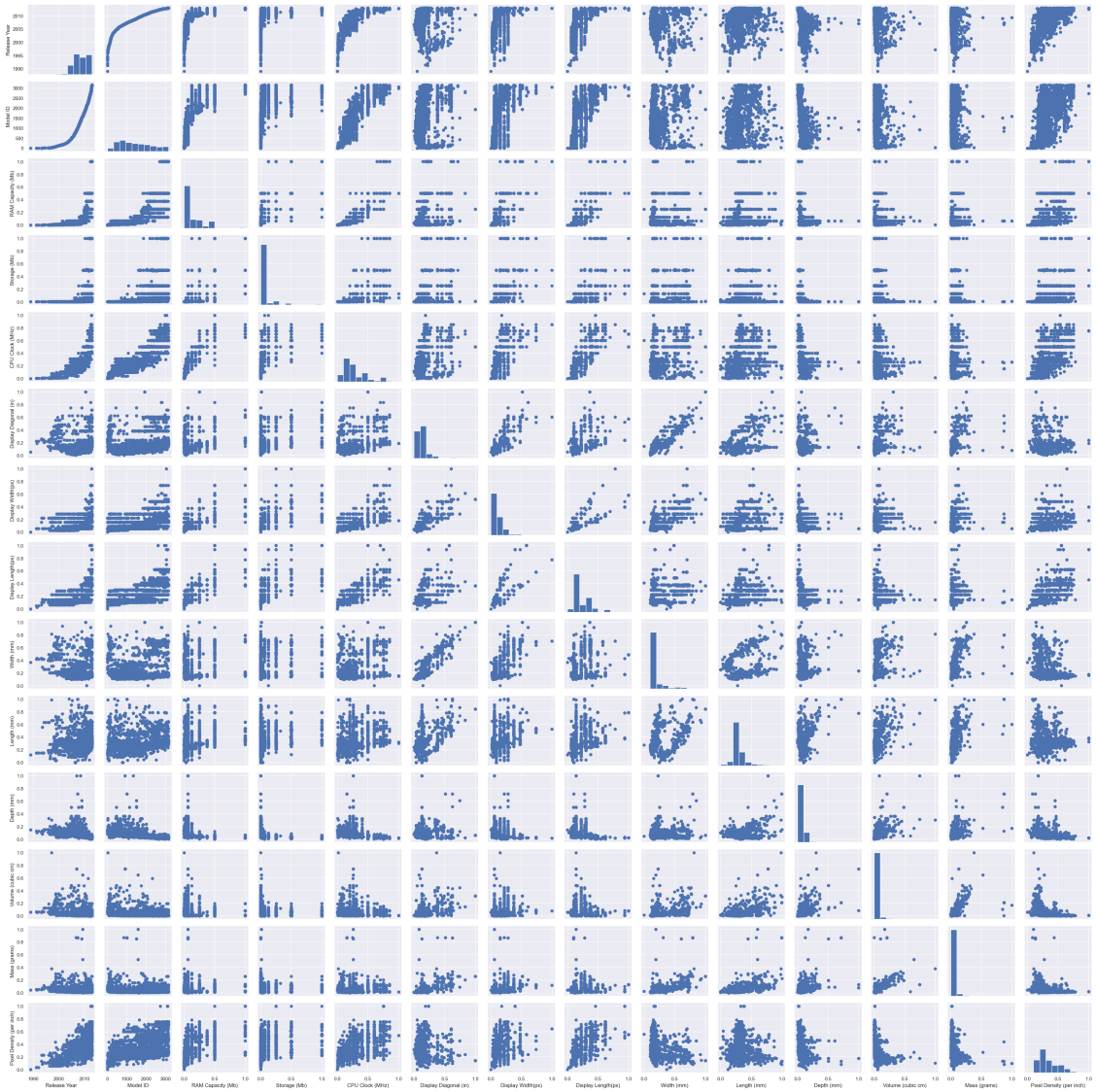
```
75%     2010.833333   2084.000000           0.249977      0.016778
max     2012.916667   3162.000000           1.000000      1.000000

       CPU Clock (MHz)  Display Diagonal (in)  Display Width(px)  \
count     10104.000000           10104.000000       10104.000000
mean          0.295382               0.137328           0.117413
std           0.193653               0.103184           0.090601
min           0.000000               0.000000           0.000000
25%           0.155615               0.079137           0.056911
50%           0.259662               0.129496           0.089431
75%           0.399726               0.165468           0.154472
max           1.000000               1.000000           1.000000

       Display Length(px)    Width (mm)   Length (mm)    Depth (mm)  \
count        10104.000000  10104.000000  10104.000000  10104.000000
mean             0.228311      0.190195      0.286308      0.069221
std              0.140831      0.113701      0.102998      0.051175
min              0.000000      0.000000      0.000000      0.000000
25%              0.142857      0.140518      0.240574      0.037845
50%              0.142857      0.151432      0.270492      0.062219
75%              0.365079      0.181173      0.311475      0.080821
max              1.000000      1.000000      1.000000      1.000000

       Volume (cubic cm)  Mass (grams)  Pixel Density (per inch)
count       10104.000000  10104.000000              10104.000000
mean            0.026265      0.027958                  0.329753
std             0.045195      0.040390                  0.147881
min             0.000000      0.000000                  0.000000
25%             0.008886      0.012931                  0.216305
50%             0.012847      0.017672                  0.276376
75%             0.025737      0.025862                  0.436648
max             1.000000      1.000000                  1.000000
```

[ ]:
```
# check for missing data
df.isna().sum()
```

[ ]:
```
Model                     0
Release Date              0
Release Year              0
Model ID                  0
RAM Capacity (Mb)         0
Storage (Mb)              0
CPU Clock (MHz)           0
Display Diagonal (in)     0
Display Width(px)         0
Display Length(px)        0
Width (mm)                0
```

```
Length (mm)                  0
Depth (mm)                   0
Volume (cubic cm)            0
Mass (grams)                 0
Pixel Density (per inch)     0
Company_ID                   0
Company                      0
Company_real                 0
dtype: int64
```

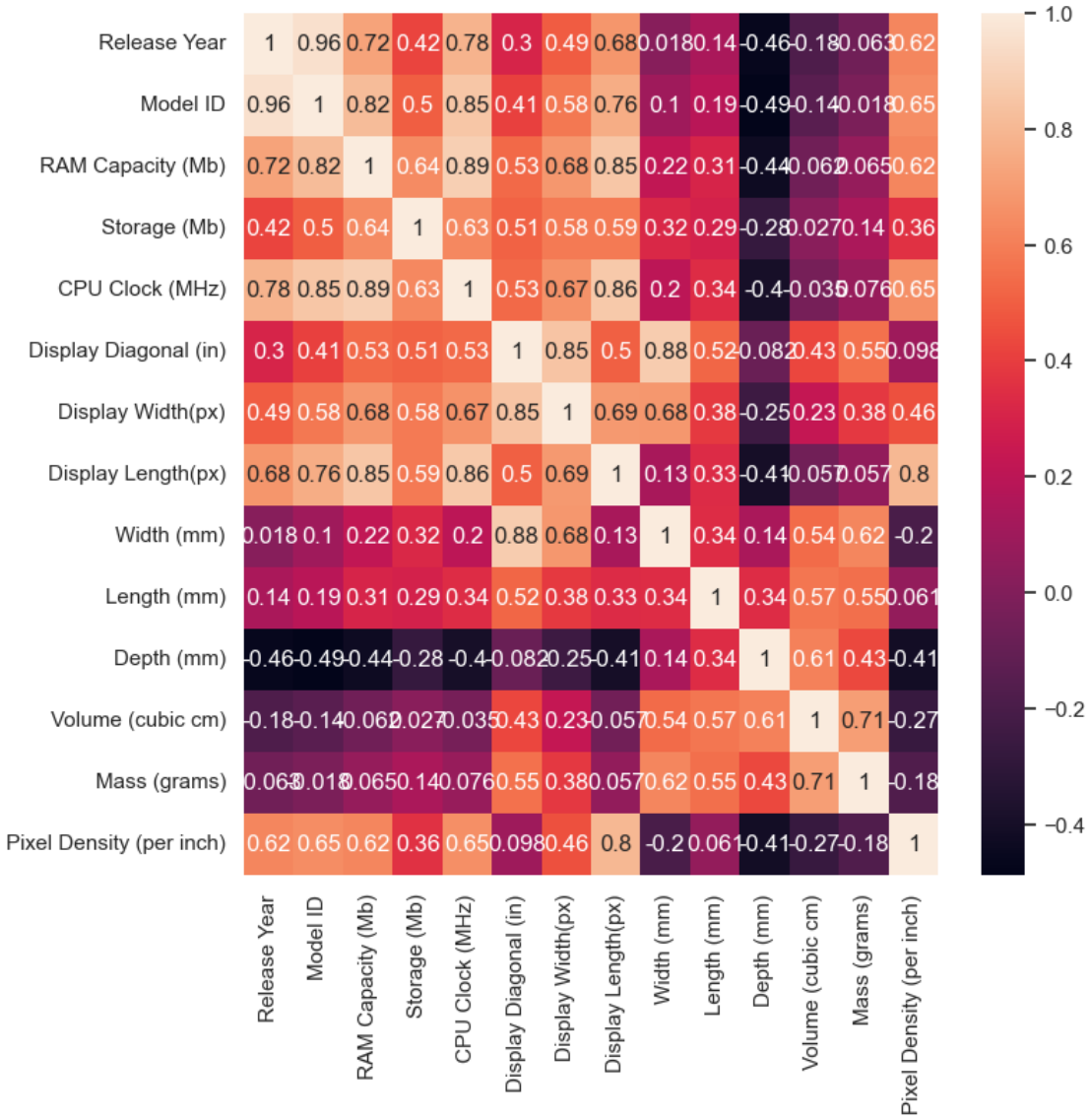
[ ]:
```
sns.set()
g = sns.PairGrid(df)
g = g.map_diag(plt.hist)
g = g.map_offdiag(plt.scatter)
```



[ ]:
```
# correlation matrix with sns heatmap
plt.figure(figsize=(8,8))
sns.heatmap(df.corr(), annot=True)
```

```
/var/folders/34/v16rh8rj48ddxlq5hfr8thb40000gn/T/ipykernel_19347/228633963.py:3:
FutureWarning: The default value of numeric_only in DataFrame.corr is
deprecated. In a future version, it will default to False. Select only valid
columns or specify the value of numeric_only to silence this warning.
  sns.heatmap(df.corr(), annot=True)
```

[ ]: <AxesSubplot: >



[ ]: