```
[1]: import pandas as pd
     import numpy as np
     import matplotlib as mpl
     %matplotlib inline
     %matplotlib inline
```

```
[2]: data = pd.read_csv('Train.csv', sep=',')
     data.head()
```

```
[2]:   Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin  BMI  DiabetesPedigreeFunction  Age  Outcome
       0           1       85             66             29        0   26                     0.351   31        0
       1           8      183             64              0        0   23                     0.672   32        1
       2           1       89             66             23       94   28                     0.167   21        0
       3           0      137             40             35      168   43                     2.288   33        1
       4           5      116             74              0        0   25                     0.201   30        0
```

```
[3]: data.info()
```

```
[3]: <class 'DataFrame'>
     RangeIndex: 2000 entries, 0 to 1999
     Data columns (total 9 cols):
     Pregnancies               2000 non-null int64
     Glucose                   2000 non-null int64
     BloodPressure             2000 non-null int64
     SkinThickness             2000 non-null int64
     Insulin                   2000 non-null int64
     BMI                       2000 non-null int64
     DiabetesPedigreeFunction  2000 non-null int64
     Age                       2000 non-null int64
     Outcome                   2000 non-null int64
     dtypes: int64(9)
     memory usage: 140.6KB
```

```
[4]: data.shape
```

```
[4]: (2000, 9)
```

```
[5]: data.isnull().sum()
```

```
[5]: Pregnancies               0
     Glucose                   0
     BloodPressure             0
     SkinThickness             0
     Insulin                   0
     BMI                       0
     DiabetesPedigreeFunction  0
     Age                       0
     Outcome                   0
     dtype: int64
```

```
[6]: data.describe()
```

```
[6]:        Pregnancies      Glucose  BloodPressure  SkinThickness      Insulin          BMI  DiabetesPedigreeFunction          Age      Outcome
     count  2000.000000  2000.000000    2000.000000    2000.000000  2000.000000  2000.000000               2000.000  2000.000000  2000.000000
     mean      3.703000   120.923000      69.025000      20.535000    79.799000    31.992000                  0.471    33.240500     0.685000
     std       3.306855    31.978468      19.351561      15.941829   115.244002     7.884160                  0.331    11.690764     0.464500
     min       0.000000     0.000000       0.000000       0.000000     0.000000     0.000000                  0.078    21.000000     0.000000
     25%       1.000000   100.000000      62.000000       0.000000     0.000000    27.375000                  0.244    24.000000     0.000000
     50%       3.000000   117.000000      72.000000      23.000000    30.500000    32.000000                  0.376    29.500000     1.000000
     75%       5.000000   141.000000      80.000000      32.000000   127.250000    36.600000                  0.626    41.000000     1.000000
     max      14.000000   199.000000     122.000000      99.000000   846.000000    67.100000                  2.420    81.000000     1.000000
```

```
[7]: import PIL ? ImageTkAgg ? ImageTkAgg? ?FigureCanvasTkAgg ?FigureCanvasTkAgg?
     %matplotlib inline
     %matplotlib inline ?matplotlib.inline ?matplotlib.inline
```

```
[8]: data['Pregnancies'].hist(bins=50)
     data['Glucose'].hist(bins=50)
     data['BloodPressure'].hist(bins=50)
     data['SkinThickness'].hist(bins=50)
     data['Insulin'].hist(bins=50)
     data['BMI'].hist(bins=50)
     data['DiabetesPedigreeFunction'].hist(bins=50)
     data['Age'].hist(bins=50)
     data['Outcome'].hist(bins=50)
     data.hist(bins=50)
     data.hist(bins=50, figsize=(20,15))
     data.hist(bins=50, figsize=(20,15))
     data.hist(bins=50, figsize=(20,15))
```

Output(s)

```
0.0%
0.0%
0.0%
```