```
In [1]: # Do imports
        import pandas as pd
        import os
```

```
In [3]: df = pd.read_csv(os.path.join("Data", "bsc5.csv"))
```

```
In [4]: print(df.shape)
```

```
(9110, 53)
```

```
In [5]: print(df.head(5))
```

```
   HR      Name          DM     HD       SAO      FK5 IRflag r_IRflag  \
0   1       NaN  BD+44 4550    3.0   36042.0      NaN    NaN      NaN
1   2       NaN  BD−01 4525    6.0  128569.0      NaN    NaN      NaN
2   3  33  Psc   BD−06 6357   28.0  128572.0   1002.0      I      NaN
3   4  86  Peg   BD+12 5063   87.0   91701.0   2004.0    NaN      NaN
4   5       NaN  BD+57 2865  123.0   21085.0      NaN    NaN      NaN

   Multiple   ADS  ... RadVel n_RadVel  l_RotVel  RotVel  u_RotVel Dmag   Sep  \
0       NaN  46.0  ...  −18.0      NaN       NaN   195.0       NaN  4.2  21.6
1       NaN   NaN  ...   14.0        V       NaN     NaN       NaN  NaN   NaN
2       NaN   NaN  ...   −6.0     SB10         <    17.0       NaN  2.5   0.0
3       NaN   NaN  ...   −2.0       V?       NaN     NaN       NaN  NaN   NaN
4       NaN  61.0  ...  −12.0        V       NaN     NaN       NaN  0.8   1.4

   MultID  MultCnt  NoteFlag
0      AC      3.0       NaN
1     NaN      NaN       NaN
2     NaN      3.0         *
3     NaN      NaN       NaN
4     NaN      NaN         *

[5 rows x 53 columns]
```

```
In [6]: print(df.tail(5))
```

```
        HR Name          DM       HD       SAO      FK5 IRflag r_IRflag  \
9105  9106  NaN  CP−73 2346  225233.0  255629.0      NaN    NaN      NaN
9106  9107  NaN  BD+33 4828  225239.0   53622.0   2002.0    NaN      NaN
9107  9108  NaN  CP−72 2800  225253.0  255631.0   1001.0    NaN      NaN
9108  9109  NaN  BD+25 5068  225276.0   73731.0      NaN      I      NaN
9109  9110  NaN  BD+60 2667  225289.0   10962.0      NaN    NaN      NaN

      Multiple   ADS  ... RadVel n_RadVel  l_RotVel  RotVel  u_RotVel Dmag  \
9105       NaN   NaN  ...    8.0      NaN       NaN     NaN       NaN  NaN
9106       NaN   NaN  ...    4.0        V       NaN     NaN       NaN  NaN
9107       NaN   NaN  ...   −3.0        V       NaN     NaN       NaN  NaN
9108       NaN  42.0  ...   −5.0      NaN       NaN     NaN       NaN  4.5
9109       NaN   NaN  ...   14.0        V       NaN    50.0       NaN  NaN

       Sep  MultID  MultCnt  NoteFlag
9105   NaN     NaN      NaN       NaN
9106   NaN     NaN      NaN       NaN
9107   NaN     NaN      NaN       NaN
9108  17.5     NaN      NaN       NaN
9109   NaN     NaN      NaN         *

[5 rows x 53 columns]
```

```
In [7]: print(df.describe())
```

```
                   HR              HD             SAO            FK5            ADS  \
count      9110.00000     9096.000000     9071.000000    3077.000000    1435.000000
mean       4555.50000   107334.581794   134193.509536    1954.447514    8278.804878
std        2629.97481    67739.321279    75516.511021    1209.016612    4700.699670
min           1.00000        3.000000      106.000000       1.000000       1.000000
25%        2278.25000    44330.500000    71659.000000     820.000000    4154.500000
50%        4555.50000   103449.500000   131443.000000    2104.000000    8406.000000
75%        6832.75000   167648.750000   200666.500000    3046.000000   11822.500000
max        9110.00000   225289.000000   258996.000000    3997.000000   17175.000000

              RAh1900      RAm1900      RAs1900      DEd1900      DEm1900  ...  \
count     9096.000000  9096.000000  9096.000000  9096.000000  9096.000000  ...
mean        11.555739    29.665347    29.990699    33.864886    29.355761  ...
std          6.787416    17.189771    17.301530    21.200809    17.454815  ...
min          0.000000     0.000000     0.000000     0.000000     0.000000  ...
25%          6.000000    15.000000    14.800000    16.000000    14.000000  ...
50%         11.000000    30.000000    30.300000    33.000000    29.000000  ...
75%         18.000000    45.000000    44.500000    50.000000    45.000000  ...
max         23.000000    59.000000    59.900000    89.000000    59.000000  ...

                  U–B           R–I         pmRA         pmDE      Parallax  \
count     7206.000000  2551.000000  9091.000000  9096.000000  3288.000000
mean         0.374167     0.368491     0.001825    -0.019220     0.024515
std          0.804599     0.458768     0.163263     0.138356     0.032333
min         -1.110000    -0.530000    -3.646000    -5.813000    -0.032000
25%         -0.080000     0.020000    -0.024000    -0.031000     0.009000
50%          0.100000     0.310000    -0.002000    -0.008000     0.018000
75%          0.880000     0.570000     0.020000     0.007000     0.032000
max          7.400000     3.710000     4.136000     3.208000     0.751000

                RadVel        RotVel         Dmag          Sep       MultCnt
count      9088.000000  3932.000000  2888.000000  2951.000000  1148.000000
mean          0.600464    87.750763     3.536946    38.440291     3.710801
std          23.637570    83.451620     2.789877    70.584071     1.367829
min        -140.000000     0.000000     0.000000     0.000000     3.000000
25%         -14.000000    19.000000     1.000000     1.300000     3.000000
50%           0.000000    60.000000     3.200000    10.800000     3.000000
75%          15.000000   132.000000     5.700000    41.950000     4.000000
max         289.000000   455.000000    13.800000  1480.000000    17.000000

[8 rows x 31 columns]
```

In [8]: `print(df.info())`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9110 entries, 0 to 9109
Data columns (total 53 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   HR          9110 non-null   int64
 1   Name        3157 non-null   object
 2   DM          9096 non-null   object
 3   HD          9096 non-null   float64
 4   SAO         9071 non-null   float64
 5   FK5         3077 non-null   float64
 6   IRflag      1743 non-null   object
 7   r_IRflag    84 non-null     object
 8   Multiple    1577 non-null   object
 9   ADS         1435 non-null   float64
 10  ADScomp     403 non-null    object
 11  VarID       2182 non-null   object
 12  RAh1900     9096 non-null   float64
 13  RAm1900     9096 non-null   float64
 14  RAs1900     9096 non-null   float64
 15  DE          9096 non-null   object
 16  DEd1900     9096 non-null   float64
 17  DEm1900     9096 non-null   float64
 18  DEs1900     9096 non-null   float64
 19  RAh         9096 non-null   float64
 20  RAm         9096 non-null   float64
 21  RAs         9096 non-null   float64
 22  DE.1        9096 non-null   object
 23  DEd         9096 non-null   float64
 24  DEm         9096 non-null   float64
 25  DEs         9096 non-null   float64
 26  GLON        9096 non-null   float64
 27  GLAT        9096 non-null   float64
 28  Vmag        9096 non-null   float64
 29  n_Vmag      306 non-null    object
 30  u_Vmag      1 non-null      object
 31  B-V         8786 non-null   float64
 32  u_B-V       30 non-null     object
 33  U-B         7206 non-null   float64
 34  u_U-B       46 non-null     object
 35  R-I         2551 non-null   float64
 36  n_R-I       303 non-null    object
 37  SpType      9096 non-null   object
 38  n_SpType    432 non-null    object
 39  pmRA        9091 non-null   float64
 40  pmDE        9096 non-null   float64
 41  n_Parallax  393 non-null    object
 42  Parallax    3288 non-null   float64
 43  RadVel      9088 non-null   float64
 44  n_RadVel    4962 non-null   object
 45  l_RotVel    725 non-null    object
 46  RotVel      3932 non-null   float64
 47  u_RotVel    285 non-null    object
 48  Dmag        2888 non-null   float64
 49  Sep         2951 non-null   float64
 50  MultID      1116 non-null   object
 51  MultCnt     1148 non-null   float64
 52  NoteFlag    4477 non-null   object
dtypes: float64(30), int64(1), object(22)
memory usage: 3.7+ MB
None
```

In [ ]: