```
import pandas as pd


from google.colab import files
uploaded = files.upload()
```

```
Choose Files fifa21_raw_data.csv
• fifa21_raw_data.csv(text/csv) - 8675837 bytes, last modified: 10/22/2020 - 100% done
Saving fifa21_raw_data.csv to fifa21_raw_data.csv
```

```
df = pd.read_csv('fifa21_raw_data.csv')
```

```
/tmp/ipython-input-830073151.py:1: DtypeWarning: Columns (76) have mixed types. Specify dtype option on import or set lo
  df = pd.read_csv('fifa21_raw_data.csv')
```

```
print("Shape of dataset:", df.shape)
print("\nColumn names:\n", df.columns.tolist())
print("\nData types:\n", df.dtypes)
print("\nMissing values:\n", df.isnull().sum())
print("\nBasic Statistics:\n", df.describe(include='all'))
```

```
50%                                                   NaN         NaN
75%                                                   NaN         NaN
max                                                   NaN         NaN

       Positions          Name           Age          ↓OVA           POT  \
count      18979         18979  18979.000000  18979.000000  18979.000000
unique       640         17919           NaN           NaN           NaN
top           CB  J. Rodríguez           NaN           NaN           NaN
freq        2441            13           NaN           NaN           NaN
mean         NaN           NaN     25.194583     65.718636     71.136098
std          NaN           NaN      4.710753      6.968999      6.114176
min          NaN           NaN     16.000000     47.000000     47.000000
25%          NaN           NaN     21.000000     61.000000     67.000000
50%          NaN           NaN     25.000000     66.000000     71.000000
75%          NaN           NaN     29.000000     70.000000     75.000000
max          NaN           NaN     53.000000     93.000000     95.000000

           Team & Contract  ...      A/W     D/W     IR           PAC  \
count                18979  ...    18979   18979  18979  18979.000000
unique                9023  ...        3       3      5           NaN
top     \n India\nFree\n\n  ...   Medium  Medium    1 ★           NaN
freq                    29  ...    12700   13956  17628           NaN
mean                   NaN  ...      NaN     NaN    NaN     67.454239
std                    NaN  ...      NaN     NaN    NaN     10.678058
min                    NaN  ...      NaN     NaN    NaN     25.000000
25%                    NaN  ...      NaN     NaN    NaN     61.000000
50%                    NaN  ...      NaN     NaN    NaN     68.000000
75%                    NaN  ...      NaN     NaN    NaN     75.000000
max                    NaN  ...      NaN     NaN    NaN     96.000000

                SHO           PAS           DRI           DEF           PHY  \
count   18979.000000  18979.000000  18979.000000  18979.000000  18979.000000
unique           NaN           NaN           NaN           NaN           NaN
top              NaN           NaN           NaN           NaN           NaN
freq             NaN           NaN           NaN           NaN           NaN
mean       53.457716     57.681069     62.875494     49.865904     64.368618
std        13.827229     10.081914      9.927875     16.442730      9.601665
min        16.000000     25.000000     25.000000     12.000000     28.000000
25%        44.000000     51.000000     57.000000     35.000000     58.000000
50%        56.000000     58.000000     64.000000     53.000000     65.000000
75%        64.000000     64.000000     69.000000     63.000000     71.000000
max        93.000000     93.000000     95.000000     91.000000     91.000000

         Hits
count   18979
unique    408
top       \n1
freq     3216
mean      NaN
std       NaN
min       NaN
25%       NaN
50%       NaN
75%       NaN
max       NaN

[11 rows x 77 columns]
```

```
for col in df.columns:
    if df[col].dtype in ['int64', 'float64']:
        # fill numeric with mean
        df[col] = df[col].fillna(df[col].mean())
```

```
            # OR use median:
            # df[col] = df[col].fillna(df[col].median())
        else:
            # fill categorical with mode
            df[col] = df[col].fillna(df[col].mode()[0])


print("\nMissing values after cleaning:\n", df.isnull().sum().sum())
print("Dataset is cleaned ✅")
```

```
Missing values after cleaning:
 0
Dataset is cleaned ✅
```

```
df.to_csv("cleaned_dataset.csv", index=False)
print("\nCleaned dataset saved as 'cleaned_dataset.csv'")
```

```
Cleaned dataset saved as 'cleaned_dataset.csv'
```

Start coding or generate with AI.