Data Wrangling I - Command Cheat Sheet (LIP DDS TCC)

L -- Load

```
import pandas as pd               # Import pandas library
import numpy as np               # Import numpy library
df = pd.read_csv('student-por.csv')   # Load CSV into DataFrame
```

I -- Inspect

```
df.head()                         # Display first 5 rows
df.columns                        # Display column names
df.dtypes                         # Display data types of each column
```

P -- Preprocessing

```
df.isnull().sum()               # Count missing values in each column
df.describe()                   # Summary statistics for numerical columns
df.describe(include='all')     # Summary statistics for all columns
df.info()                       # Structure, non-null counts, and types
```

D -- Dimensions

```
df.shape                        # Output number of rows and columns
df.size                         # Output total number of elements
df.ndim                         # Output number of dimensions (should be 2)
```

D -- Drop Missing

```
df = df.dropna()                # Drop rows with missing values (new DataFrame)
df.dropna(inplace=True)         # Drop rows with missing values (in-place)
```

S -- Set Data Types

```
df['age'] = df['age'].astype(int)                 # Convert 'age' to integer
df['Pclass'] = df['Pclass'].astype('category')      # Convert 'Pclass' to category type
```

T -- Transform Categorical

```
df['sex'].replace({'F': 0, 'M': 1}, inplace=True)   # Convert 'sex' to numeric values
df['school'].replace({'GP': 0, 'MS': 1}, inplace=True) # Convert 'school' to numeric
values
```

C -- Check Again

```
df.dtypes                          # Check updated data types
df['sex'].value_counts()           # Frequency count of 'sex' column values
```

C -- Clean Duplicates

```
df.duplicated().sum()              # Count duplicate rows
df.drop_duplicates(inplace=True) # Remove duplicate rows in-place
```