

■ Dataset Overview

Command	Description
<code>df.shape</code>	Check dataset size (rows x cols)
<code>df.head()</code>	Preview first rows
<code>df.tail()</code>	Preview last rows
<code>df.sample(5, random_state=42)</code>	Random sample of rows
<code>df.info()</code>	Column types & non-null counts
<code>df.describe()</code>	Summary statistics
<code>df.columns</code>	List column names
<code>df.nunique()</code>	Unique values per column

■ Missing Values

Command	Description
<code>df.isnull().sum()</code>	Count missing values
<code>df.dropna(inplace=True)</code>	Drop rows with NaN values
<code>df.fillna(value, inplace=True)</code>	Fill with constant value
<code>df['col'].fillna(df['col'].mean(), inplace=True)</code>	Fill numeric NaNs with mean
<code>df['col'].fillna(df['col'].mode()[0], inplace=True)</code>	Fill categorical NaNs with mode
<code>df.interpolate()</code>	Fill missing values with interpolation

■ Duplicates

Command	Description
<code>df.duplicated().sum()</code>	Check duplicates
<code>df.drop_duplicates(inplace=True)</code>	Remove duplicates

■ Data Types & Conversion

Command	Description
<code>df.dtypes</code>	Check datatypes
<code>pd.to_numeric(df['col'], errors='coerce').astype('Int64')</code>	Safe convert to integer
<code>pd.to_datetime(df['date'])</code>	Convert to datetime
<code>df['col'].astype('category')</code>	Convert to category type

■ String Cleaning

Command	Description
<code>df['col'].str.strip()</code>	Remove spaces
<code>df['col'].str.lower()</code>	Convert text to lowercase
<code>df['col'].str.replace('[^a-zA-Z0-9]', '', regex=True)</code>	Remove special characters (keep spaces)

<code>df['col'].str.split()</code>	Split string into list
------------------------------------	------------------------

■ **Outliers & Numeric Cleaning**

Command	Description
<code>df.describe()</code>	Detect outliers with stats
<code>df['col'].quantile([0.01,0.99])</code>	Cutoffs for outliers
<code>df[df['col'] < df['col'].quantile(0.95)]</code>	Trim top 5% extreme values
<code>df['col'].clip(lower=0, upper=df['col'].quantile(0.95))</code>	Clip values

■ **Column & Index Management**

Command	Description
<code>df.rename(columns={'old':'new'})</code>	Rename columns
<code>df.drop(['col1','col2'], axis=1)</code>	Drop unnecessary columns
<code>df.reset_index(drop=True)</code>	Reset index
<code>df['new'] = df['col1'] + df['col2']</code>	Create new column

■ **Final Checks**

Command	Description
<code>df.info()</code>	Verify dataset structure
<code>df.isnull().sum()</code>	Confirm no missing values
<code>df.duplicated().sum()</code>	Check no duplicates
<code>df.describe(include='all')</code>	Review distributions