

Pandas pocket card

Twenty commands that cover most facility data work.

Task	Command
Load a CSV export	<code>df = pd.read_csv("file.csv")</code> <code>sep=";"</code> or <code>skiprows=2</code> for awkward files
Look at it	<code>df.head()</code> <code>df.shape</code> <code>df.columns</code> <code>df.dtypes</code>
One column, several columns	<code>df["anc4"]</code> <code>df[["facility", "month", "anc4"]]</code>
Rows by condition	<code>df[df["facility"] == "Gondar CSH"]</code> <code>df[(a) & (b)]</code>
Real spellings	<code>df["facility"].unique()</code> <code>df["facility"].value_counts()</code>
Sort	<code>df.sort_values("anc4", ascending=False)</code>
Missing values	<code>df.isna().sum()</code> <code>df[df["anc4"].isna()]</code>
Make numbers numbers	<code>pd.to_numeric(df["anc4"], errors="coerce")</code>
Dates	<code>pd.to_datetime(df["date"], format="%d/%m/%Y")</code>
Duplicates on a key	<code>df.duplicated(subset=["facility", "period"]).sum()</code>
Describe a column	<code>df["deliveries"].describe()</code>
Group and summarise	<code>df.groupby("facility")["deliveries"].agg(["count", "sum", "mean"]).reset_index()</code>
New column for every row	<code>df["anc4_cov"] = df["anc4"] / df["expected_pregnancies"] * 100</code>
Save	<code>df.to_csv("output/clean.csv", index=False)</code> <code>df.to_excel("output/clean.xlsx", index=False)</code>

Based on the ERAS Africa Basics of Data Analysis using Python course, 2026, the WHO Data Quality Review toolkit, the Ethiopian HMIS indicator reference and the pandas and matplotlib documentation. Free to copy and adapt with attribution.