| **Step (English)** | **SAS Code** | **R Code** | **Python Code (pandas)** | **Explanation** |
| --- | --- | --- | --- | --- |
| 1. Load the dataset | data mydata; set mylib.dataset; run; | mydata <- read.csv("dataset.csv") | import pandas as pd<br>mydata = pd.read\_csv("dataset.csv") | Loads the dataset into your workspace/memory for analysis. |
| 2. View first few rows | proc print data=mydata (obs=5); run; | head(mydata) | mydata.head() | Shows the first few rows so you can see what the data looks like. |
| 3. See variable names (column names) | proc contents data=mydata; run; | names(mydata) | mydata.columns | Lists all column (variable) names in your dataset. |
| 4. Get dataset dimensions (rows, columns) | proc contents data=mydata; run; | dim(mydata) | mydata.shape | Tells you how many rows (observations) and columns (variables) the dataset has. |
| 5. Summary statistics for all variables | proc means data=mydata; run; | summary(mydata) | mydata.describe() | Gives you quick stats: min, max, mean, etc. for each column (numeric). |
| 6. Check for missing values | proc means data=mydata nmiss; run; | colSums(is.na(mydata)) | mydata.isnull().sum() | Counts missing values in each column. |
| 7. Check data types (numeric, character, etc.) | proc contents data=mydata; run; | str(mydata) | mydata.dtypes | Shows what type (number, string, etc.) each variable is. |
| 8. View unique values for a variable | proc freq data=mydata; tables var1; run; | unique(mydata$var1) | mydata['var1'].unique() | Lists all unique values in a column (useful for categorical data). |