Python & Pandas Cheat Sheet (Workshop 1 Recap)

1. Core Python Data Types

int, float, str, list, dict.

Check type: type(x)

Formatted print: print(f"{name} is {age}")

2. Creating & Inspecting DataFrames

Create DataFrame: df = pd.DataFrame(dict_students)

View top rows: df.head()

View shape: df.shape

Summary statistics: df.describe()

Set index: df = df.set_index('name')

3. Selecting Data

Select column: df['math']

Select multiple columns: df[['math','econ']]

Select row by label: df.loc['Beth']

4. Filtering with .query()

General form: df.query('condition')

Example: df.query('math < 60')

Example: df.query('econ >= 60 and math >= 60')

Example: df.query('(math + econ)/2 >= 60')

Filter by category: df.query("gender == 'Male'")

5. Creating New Variables

Mean mark: df['mean_mark'] = (df['math'] + df['econ']) / 2

Create indicator: df['is_female'] = (df['gender']=='Female').astype(int)

6. Summary Statistics

Column means: df[['math','econ']].mean()

Correlation: df[['math','econ']].corr()

7. Grouping with .groupby()

Group means: df.groupby('gender').mean()

Specific variable: df.groupby('gender')['econ'].mean()

Multiple stats: df.groupby('gender')['econ'].agg(['mean','min','max'])

8. Cleaning Data

Remove duplicates: df.drop_duplicates()

Remove missing values: df.dropna()

9. Simple Plots

Bar plot: df[['math','econ']].plot(kind='bar')

Scatter plot: df.plot(kind='scatter', x='math', y='econ')