

Cheatsheet 06: Pandas

Installation

pip

```
pip install pandas
```

conda

```
conda install pandas
```

Importing pandas

```
import pandas as pd
```

Creating a DataFrame

```
# From a dictionary
data = {'A': [1, 2, 3], 'B': [4, 5, 6], 'C': [7, 8, 9]}
df = pd.DataFrame(data)

# From a list of dictionaries
data = [
    {'A': 1, 'B': 4, 'C': 7},
    {'A': 2, 'B': 5, 'C': 8},
    {'A': 3, 'B': 6, 'C': 9}
]
df = pd.DataFrame(data)

# From a CSV file
df = pd.read_csv('filename.csv')
```

Viewing Data

```
# First n rows (default 5)
df.head(n)

# Last n rows (default 5)
df.tail(n)

# Basic information about the DataFrame
df.info()

# Summary statistics
df.describe()
```

Selection

```
# Selecting a single column
df['A']

# Selecting multiple columns
df[['A', 'B']]

# Selecting rows by index
df.loc[0] # row with index 0
df.iloc[0] # first row

# Selecting rows by condition
df[df['A'] > 1]

# Selecting specific rows and columns
df.loc[[0, 2], ['A', 'B']]
```

Sorting

```
# Sort by a single column
df.sort_values(by='A', ascending=True)

# Sort by multiple columns
```

```
df.sort_values(by=['A', 'B'], ascending=[True, False])
```

Modifying Data

```
# Adding a new column
df['D'] = [10, 11, 12]

# Renaming columns
df.rename(columns={'A': 'Col_A', 'B': 'Col_B'}, inplace=True)

# Dropping a column
df.drop('C', axis=1, inplace=True)

# Dropping a row by index
df.drop(0, axis=0, inplace=True)
```

Handling Missing Data

```
# Checking for missing data
df.isnull()

# Dropping missing data
df.dropna(axis=0) # drop rows
df.dropna(axis=1) # drop columns

# Filling missing data
df.fillna(value=0)
```

Grouping

```
# Group by a single column
grouped = df.groupby('A')

# Group by multiple columns
grouped = df.groupby(['A', 'B'])

# Aggregate functions
```

```
grouped.sum()
grouped.mean()
grouped.count()
```

Merging DataFrames

```
# Concatenating DataFrames
df_concat = pd.concat([df1, df2], axis=0) # Vertical concatenation
df_concat = pd.concat([df1, df2], axis=1) # Horizontal concatenation

# Merging DataFrames on a common key
df_merged = pd.merge(df1, df2, on='key')

# Merging with different key names
df_merged = pd.merge(df1, df2, left_on='key1', right_on='key2')
```

Exporting DataFrames

```
# To CSV
df.to_csv('filename.csv', index=False)

# To Excel
df.to_excel('filename.xlsx', index=False)
```