Cheatography

pandas Cheat Sheet

by sankyuu via cheatography.com/144187/cs/30916/

Import Data CSV and Excel

df = pd.read_csv('filename.csv')	Read CSV into a Pandas DataFrame
df = pd.to_csv('filename.csv')	Export Pandas DataFrame to CSV
df = pd.read_excel('filename.xlsx', sheet_name="Sheet 1")	Read Excel into a Pandas DataFrame
df = pd.to_excel('filename.xlsx', sheet_name="Sheet 1")	Export Pandas DataFrame to Excel

Import Options:

header=False, Index=False, usecols=(5,6)

Can also read CSV / HTML / JSON

Initial look into the DataFrame

df.head(5)	Reads the first 5 rows
df.tail(5)	Reads the last 5 rows
df.shape()	Gives the number of columns and rows in the DataFrame
df.dtypes	Gives the datatypes for all the columns
df['ColumnName'].dtypes	Gives the datatypes for a single column

df.head(-5) can retrieve last 5 lines similar to list[-5] retrieves the 5th last.

Change Column Data Type

df['col'] = df['col'].str.rstrip('%').astype('float') / 100.0	Remove % sign, convert to float and divide by 100
df['col'] = df['col'].str[:-1].astype('float') / 100.0	Blindly removing the last char - goes to the last character
df['Column1']= df['Column1'].astype(float)	Change 'Column1' to float

Re-Order Colums

df = df[['Column3', 'Column2', 'Column1']]	Re-orders the columns to the order specified in this list

Dealing with NAN values

df= df.fillna(method='ffill')	Fills blank values using forward fill method
df= df.fillna(method='bfill')	Fills blank values using backwards fill method
df.dropna(inplace=True)	Removes rows with no values

Forward filling means fill missing values with last cell with value in the column. Backward filling means fill missing values with next cell with value.

Padding Values With Zero's

df['Column1'] = df['Column1'].astype(str).str.zfill(6)	Sets the numbe to six (6) long, which adds zeros

Obviously uses string since 0009 cannot be a valid number.

Filter Column

df2= df[(df['Column1'] >= some_number)]	Filter DataFrame by certain value and putting it in another df.
df2 = df[(df["Name"]=="Tom") & (df["Age"]==42)]	Filter DataFrame by mulitple value and putting it in another df.