4/26/2021

Nadir Hussain

023-18-0025

Data Science

Lab 06

Data Science/ Python for data Analysis.

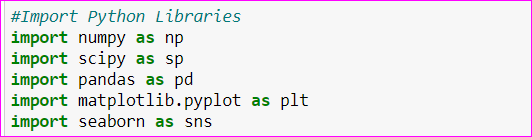
Libraries used

1. Numpy
2. Scipy
3. Pandas
4. Matplotlib
5. Seaborn
6. Scikit learn
7. Statsmodels.formula.api

Why to use

1. Numpy: Numerical python library, contains multi dimensional array and matrix data structures. Used to perform mathematical operations on arrays and matrices like arithmetic, algebraic, statistical and trigonometric.
2. Scipy: Used for mathematical , engineering and technical problems. Provides commands to visualize the data.
3. Pandas: Most widely used for machine learning, data science tasks. Built on top of Numpy. It works similar to Excel. Provides tabular data visualization. Can be used to create 1d homogeneous arrays called Series, 2d labeled heterogeneous arrays called DataFrame.
4. Matplotlib: Used to work with large data sets by visualization.
5. Seaborn: Built on matplotlib, provides attractive visualization and graphical interaction

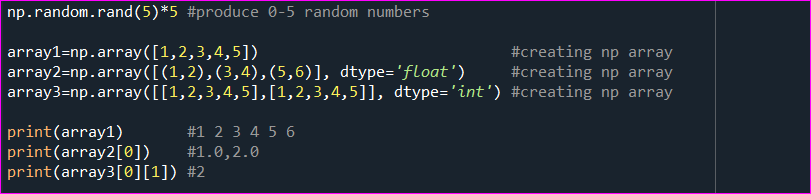
Import all of above libraris



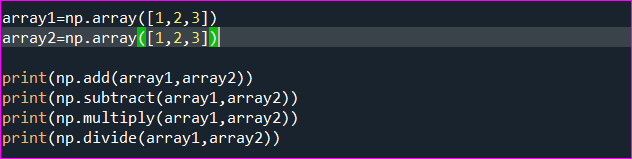
Working with Pandas—Series

Numpy functions

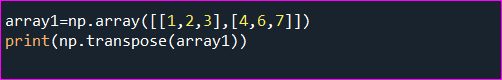
Creating arrays



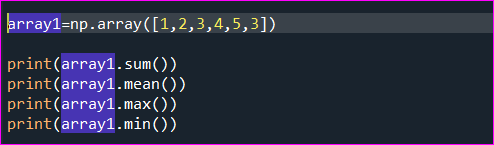
We can add, subtract, multiply and divide arrays of same dimension



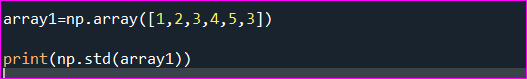
Finding transpose of array



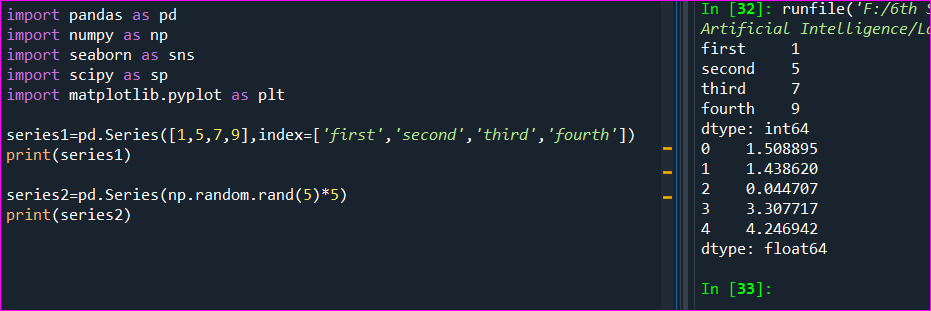
Basic statistics on array



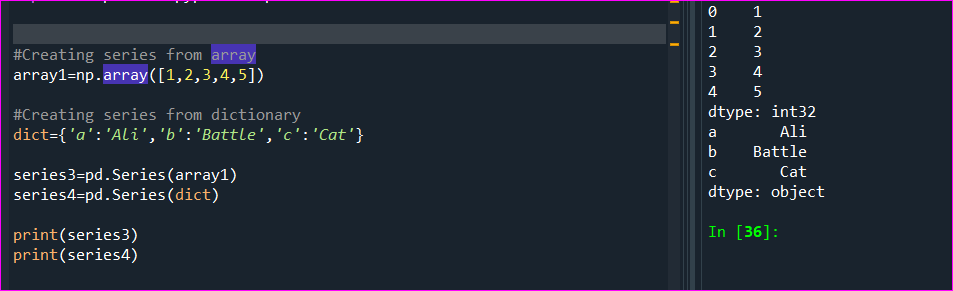
Standard deviation of data



Pandas Functions



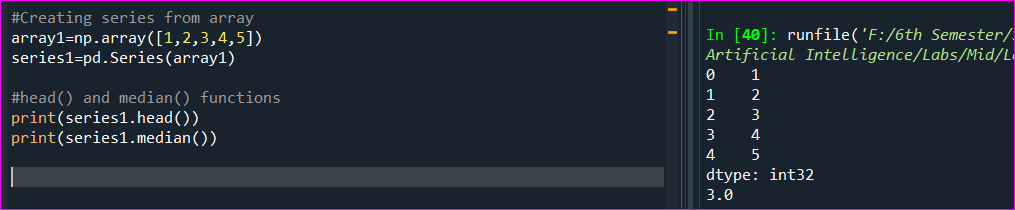
Creating Series from array and dictionary



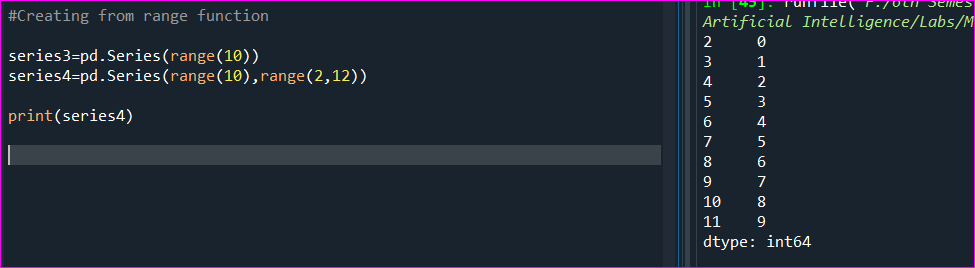
Slicing Series elements



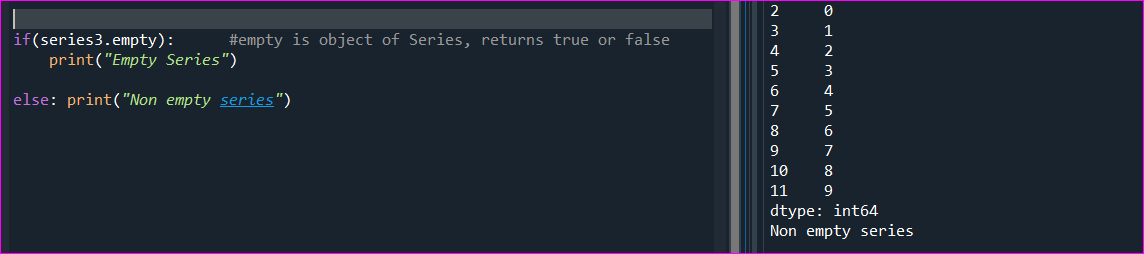
Head and median functions of pandas series



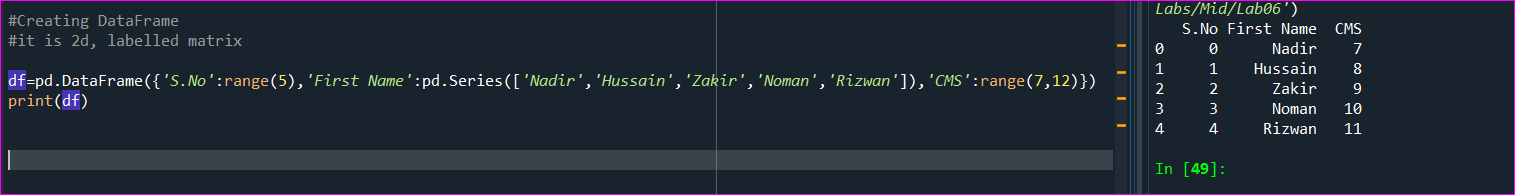
Creating series from range function



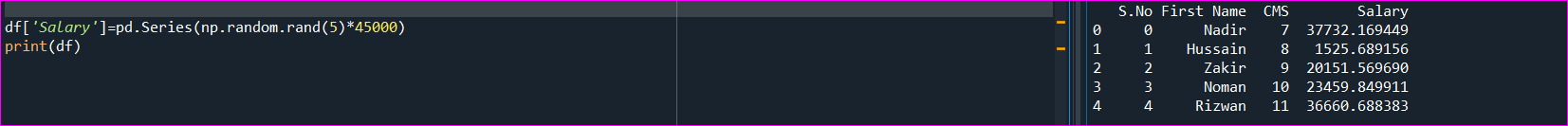
empty property or object of Series



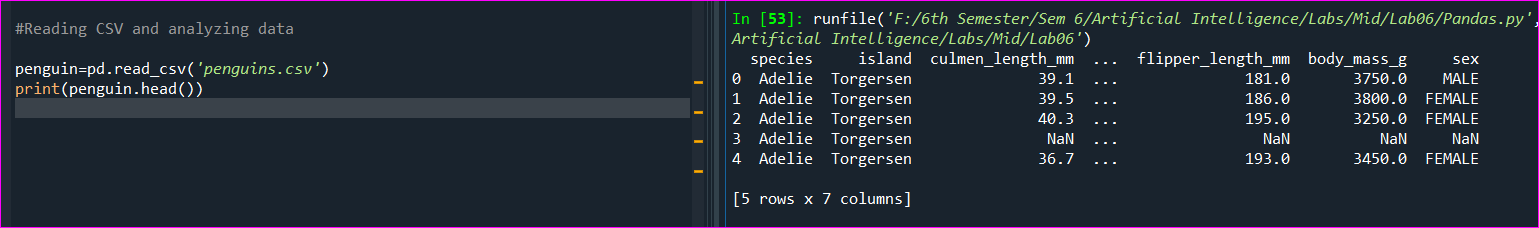
Pandas Data Frame



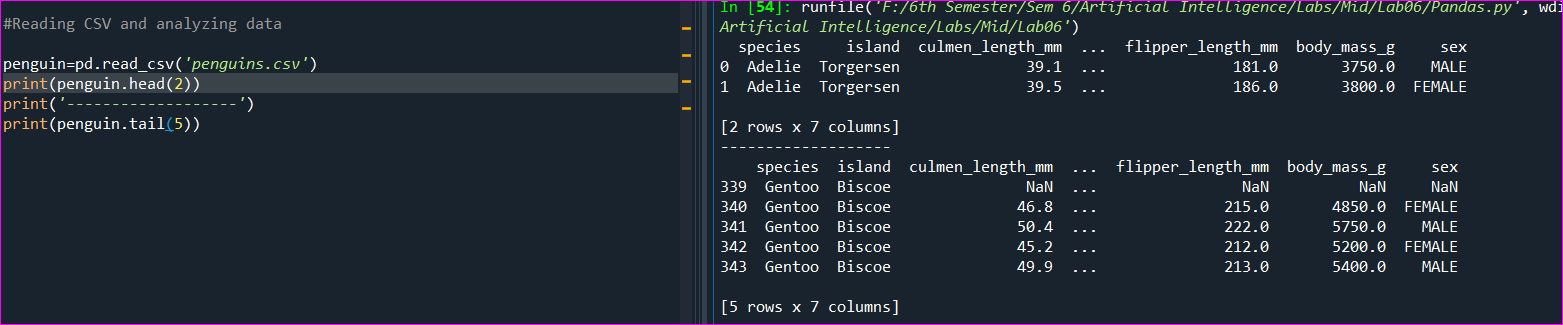
Adding new column to dataframe



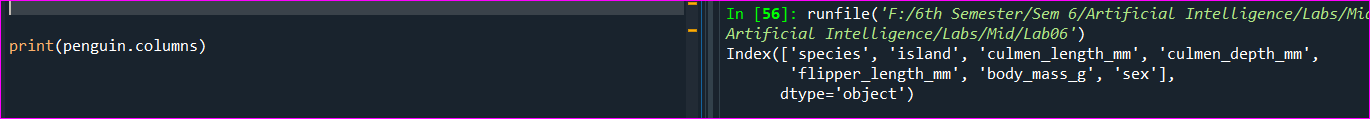
Reading and analyzing CSV dataset in pandas



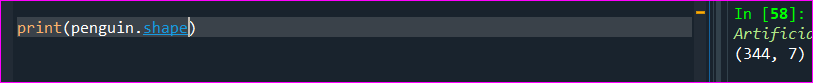
Reading from head and tail of dataset



Get names of columns of dataset

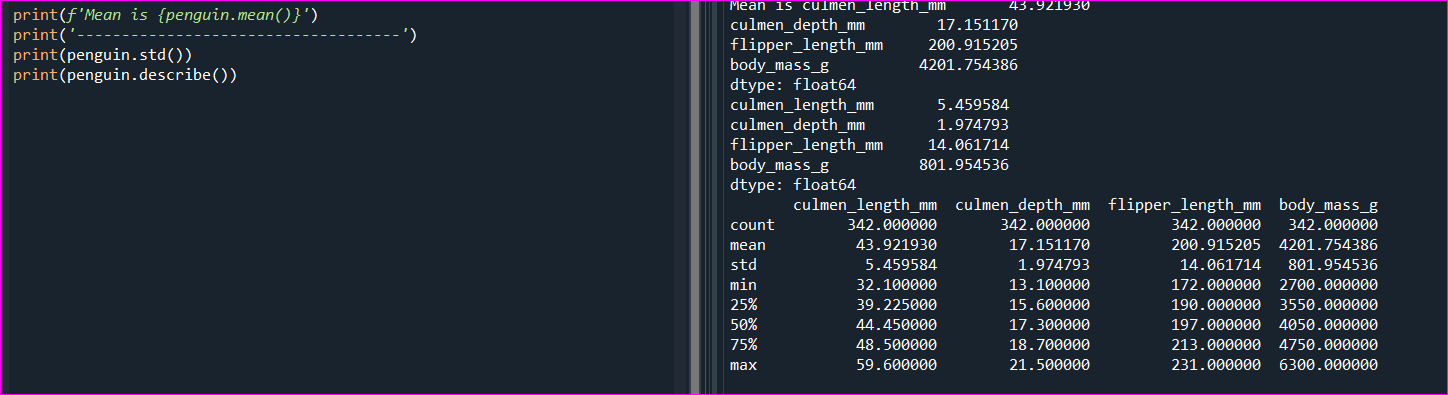


Get number of records in dataset

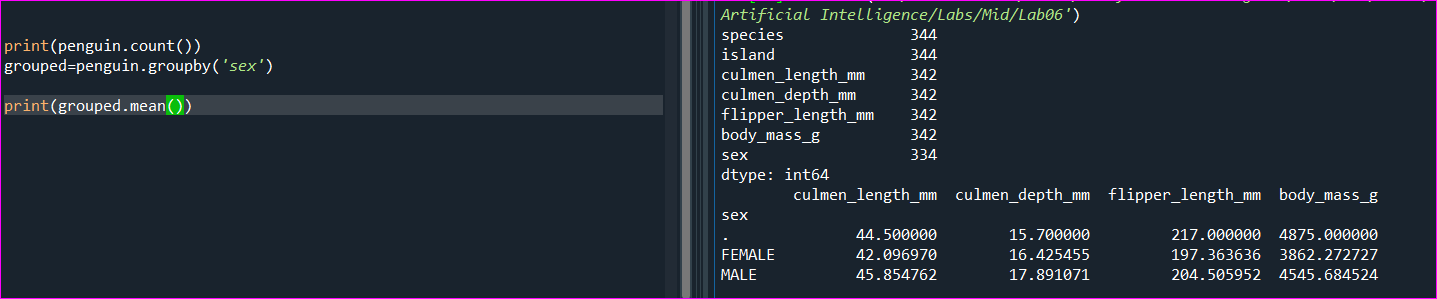


344 rows with 7 columns

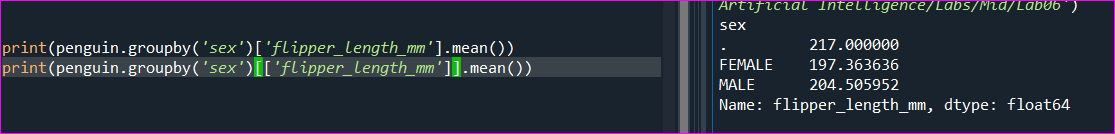
Get mean, deviation or all



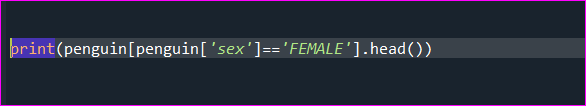
Count records and group by column



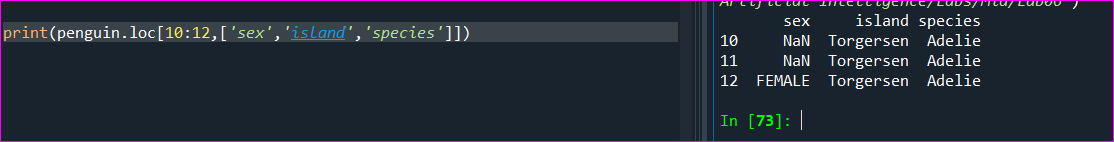
Group by and calculate mean of particular column



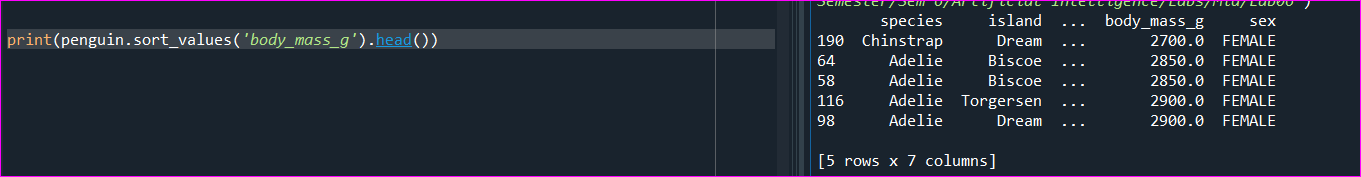
If sex is female, select 5 penguins

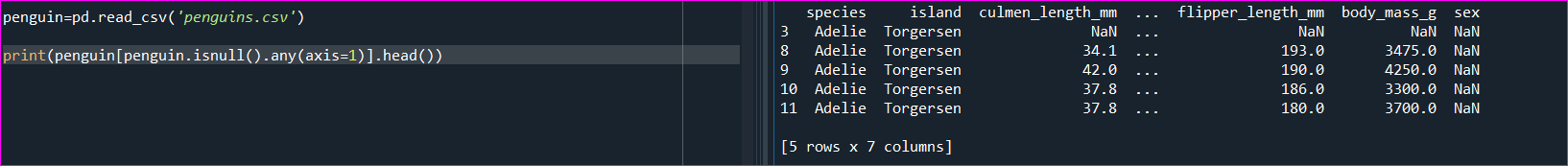


SELECT 10 to 12 records with 3 columns

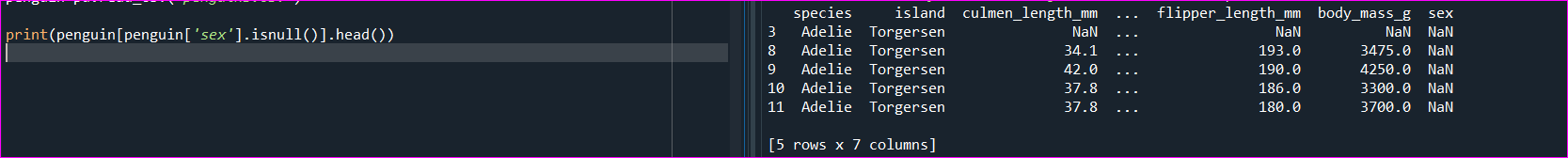


Sort penguins with body\_mass\_g



Select rows that have atleast one missing value 

Select all rows where ‘sex’ column value is missing



Remove all rows with any missing value

