**Работа с Pandas**

|  |  |
| --- | --- |
| Создание объекта Series | a = [1, 3, 5, 7, 2]  b = pd.Series(a) |
| Можно задать индексы явно | b = pd.Series(a, index = [0, 1, 2, 3, 4])  b = pd.Series(a, index = ['a', 'b', 'c', 'd', 'e']) |
| Получение индексов | b.index |
| Поменять индексы | b.index = [10, 11, 12, 13, 14] |
| Задать тип данных | b = pd.Series(a, dtype = np.float64) |
| Поменять тип данных позже | b = b.astype(np.float64) |
| Создать Series из словаря (ключи становятся индексами, значения - данными) | d = {'1st':'a', '2nd':'b', '3rd':'c', '4th':'d', '5th':'e'}  b = pd.Series(d) |
| Просмотр значений массива Series | b.values |
| Доступ к данным по индексу  Доступ по списку индексов | b[0]  b[[0, 3]] |
| **Выбор данных их массива Series** |  |
| первые несколько элементов (по умолчанию - 5)  последние несколько элементов | b.head()  b.head(3)  b.tail() |
| Выборка по условию  по логическому условию "и"  по логическому условию "или" | b[b > 5]  b[(b > 5) & (b % 2 == 0)]  b[(b > 5) | (b % 3 == 0)] |
| Изменение элементов |  |
| элементу с индексом 0 присваивается значение 4  все числе в Series, которые меньше 5, приравниваются к 0  элементам под индексами 0, 1, 2 присваивается значение 1 | b[0] = 4  b[b < 5] = 0  b[[0, 1, 2]] = 1 |
| Добавление и удаление данных **Добавление данных**  **Удаление данных по индексу** | b = b.append(pd.Series({6:10, 7:11, 8:12, 9:13}))  b = b.drop([0, 1, 2]) |
| Запись и чтение файла запись Series в файл с форматом pickle  Чтение файла | b.to\_pickle('b.pkl')  b2 = pd.read\_pickle('b.pkl') |
| **Структура данных DataFrame** |  |
| Создание DataFrame | df = pd.DataFrame({'col1': ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h'],                     'col2': [1, 3, 5, 7, 9, 11, 13, 15]}) |
| **Просмотр информации о DataFrame** Просмотреть форму массива  список столбцов  список индексов  покажет кол-во строк в столбце, тип данных  статистические хар-ки | df.shape  df.columns  df.index  df.info()  df.describe() |
| **Выбор данных из массива** |  |
| просмотр нескольких первых строк DataFrame  просмотр последних 5-ти строк  вывод первых строк для отдельного столбца | df.head()  df.head(3)  df.tail()  df['col1'].head()  df.col1.head() |
| Изменение индекса | df.index = [2, 4, 6, 8, 10, 12, 14, 16] |
| **Выбор данных по индексу**  указывается индекс строки и имя столбца  если второй аргумент не указан, выбираются все столбцы  срез по строкам, верхняя граница включена | df.loc[2, 'col1']  df.loc[4]  df.loc[2:4] |
| **Выбор данных по позиции**  просмотр первой строки во всех столбцах  просмотр среза первых двух строк, верхняя граница не включена  для просмотра данных конкретного столбца указывается его индекс | df.iloc[0]  df.iloc[0:2]  df.iloc[2:6, 0] |
| **Выбор по условию**  выбираем те строки, в которых значение col1 = b  результат такой же, но выводятся не все столбца, а только col2  просмотр тех строк, в которых значение col2 больше 10, при этом выводятся только значения col1  пример логического "и"  пример логического "или"  указание отрезка, которому должен соответствовать столбец | df.loc[df['col1'] == 'b']  df.loc[df['col1'] == 'b', 'col2']  df.loc[df['col2'] > 10, 'col1']  df.loc[(df['col2'] > 10) & (df['col1'] != 'g')]  df.loc[(df['col2'] > 10) | (df['col2'] % 9 == 0)]  df.loc[df['col2'].**between**(11,13)] |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |