

ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ НИЖЕГОРОДСКИЙ
ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. Р. Е. АЛЕКСЕЕВА

Кафедра «Прикладная математика»

Лабораторная работа №1

по дисциплине «Базы данных»

Тема: «Python, Pandas»

Студент

(Подпись) Валькова.Н.П.
(Фамилия, И., О.)

18-ПМ
(Группа)
(Дата сдачи)

(Подпись) **Проверил**
Моисеев А.Е
(Фамилия, И., О.)

Отчет защищен «__» _____2021_г.
с оценкой _____

Нижний Новгород, 2021

Оглавление

1. Введение.....	3
2. Постановка задачи.....	4
3. Решение.....	5

1	Вып.	Валькова.Н.П.			Лабораторная работа по предмету «Базы данных»-НГТУ-(18-ПМ)	Лист
2	Пров.	Моисеев А.Е				№
№		Ф.И.О.	Подп.	Дата		2

Введение

Pandas— программная библиотека на языке Python для обработки и анализа данных. Работа pandas с данными строится поверх библиотеки NumPy, являющейся инструментом более низкого уровня. Предоставляет специальные структуры данных и операции для манипулирования числовыми таблицами и временными рядами.

Основные возможности библиотеки:

- Объект DataFrame для манипулирования индексированными массивами двумерных данных
- Инструменты для обмена данными между структурами в памяти и файлами различных форматов
- Встроенные средства совмещения данных и способы обработки отсутствующей информации
- Переформатирование наборов данных, в том числе создание сводных таблиц
- Срез данных по значениям индекса, расширенные возможности индексирования, выборка из больших наборов данных
- Вставка и удаление столбцов данных
- Возможности группировки позволяют выполнять трёхэтапные операции типа «разделение, изменение, объединение».
- Слияние и объединение наборов данных
- Иерархическое индексирование позволяет работать с данными высокой размерности в структурах меньшей размерности
- Работа с временными рядами: формирование временных периодов и изменение интервалов и так далее

1	Вып.	Валькова.Н.П.			Лабораторная работа по предмету «Базы данных»-НГТУ-(18-ПМ)	Лист
2	Пров.	Моисеев А.Е				№
№		Ф.И.О.	Подп.	Дата		3

Постановка задачи:

Разобрать датасет с сайта ilostst.olo.org по различным параметрам.

Для разбора был выбран датасет с средней номинальной почасовая заработной платой работников в разбивке по полу и роду занятий .

1	Вып.	Валькова.Н.П.			Лабораторная работа по предмету «Базы данных»-НГТУ-(18-ПМ)	Лист
2	Пров.	Моисеев А.Е				№
№		Ф.И.О.	Подп.	Дата		4

Решение

- Подключаем numpy и pandas

```
import numpy as np
import pandas
```

- Считываем данные из csv файла и оставляем нужные нам столбцы.

```
data =
pandas.read_csv('EAR_4MTH_SEX_OCU_CUR_NB_A.csv')
data =
data[['ref_area','sex','classif1','classif2','time','obs_value']]
```

```
data_ref_area = pandas.read_csv('ref_area_en.csv', index_col
= 'ref_area')
```

- Переименовываем столбец в таблице со странами.

```
data_ref_area = data_ref_area.rename(columns={'
ref_area.label':'ref_area_label'})
data_ref_area = data_ref_area[['ref_area_label']]
```

- Добавляем столбец чтобы лучше понимать к какой стране относятся данные.

```
data = data.join(data_ref_area, on = ['ref_area'])
```

1	Вып.	Валькова.Н.П.			Лабораторная работа по предмету «Базы данных»-НГТУ-(18-ПМ)	Лист
2	Пров.	Моисеев А.Е				№
№		Ф.И.О.	Подп.	Дата		5

- Выведем данные по России за период от 2010 до 2015 года.

```
print(data[(data.ref_area == 'RUS') & ((data.time >= 2010) &
(data.time <= 2015))])
```

```
/home/nadezhda/Документы/DB/l1/venv/bin/python /home/nadezhda/Документы/DB/l1/main.py
ref_area  sex      classif1  ...  time  obs_value  ref_area_label
60194     RUS  SEX_T    OCU_ISC088_TOTAL  ...  2011    22717.00  Russian Federation
60195     RUS  SEX_T    OCU_ISC088_TOTAL  ...  2011     1217.74  Russian Federation
60196     RUS  SEX_T    OCU_ISC088_TOTAL  ...  2011      773.15  Russian Federation
60197     RUS  SEX_T      OCU_ISC088_1  ...  2011    41581.00  Russian Federation
60198     RUS  SEX_T      OCU_ISC088_1  ...  2011     2228.94  Russian Federation
...      ...      ...      ...      ...      ...      ...
60450     RUS  SEX_F      OCU_ISC008_8  ...  2015      996.72  Russian Federation
60451     RUS  SEX_F      OCU_ISC008_8  ...  2015      402.76  Russian Federation
60452     RUS  SEX_F      OCU_ISC008_9  ...  2015    14660.00  Russian Federation
60453     RUS  SEX_F      OCU_ISC008_9  ...  2015      595.36  Russian Federation
60454     RUS  SEX_F      OCU_ISC008_9  ...  2015      240.57  Russian Federation
```

- Максимальная средняя почасовая зарплата в долларах по миру(любой год):

```
print(data[(data.classif2 == 'CUR_TYPE_USD') & (data.sex
== 'SEX_T') & (data.classif1 ==
'OCU_ISCO08_TOTAL')].obs_value.max())
```

Вывод:
132957.31

- Средняя средняя почасовая зарплата в долларах по миру(любой год):

```
print(data[(data.classif2 == 'CUR_TYPE_USD') & (data.sex
== 'SEX_T') & (data.classif1 ==
'OCU_ISCO08_TOTAL')].obs_value.mean())
```

Вывод:
1565.5936007130124

1	Вып.	Валькова.Н.П.			Лабораторная работа по предмету «Базы данных»-НГТУ-(18-ПМ)	Лист
2	Пров.	Моисеев А.Е				№
№		Ф.И.О.	Подп.	Дата		6

- Медианная средняя почасовая зарплата в долларах по миру(любой год):

```
print(data[(data.classif2 == 'CUR_TYPE_USD') & (data.sex == 'SEX_T') & (data.classif1 == 'OCU_ISCO08_TOTAL')].obs_value.median())
```

Вывод:
620.72

- Топ 20 стран с максимальной почасовой зарплатой за все годы:

```
/home/nadezhda/Документы/DB/Labs1/venv/bin/python /home/nadezhda/Документы/DB/Labs1/ma
```

	ref_area	sex	classif1	...	time	obs_value	ref_area_label
2633	AUS	SEX_T	OCU_ISCO08_TOTAL	...	2016	132957.31	Australia
12240	CHE	SEX_T	OCU_ISCO08_TOTAL	...	2014	7969.21	Switzerland
49582	NOR	SEX_T	OCU_ISCO08_TOTAL	...	2013	7234.04	Norway
49420	NOR	SEX_T	OCU_ISCO08_TOTAL	...	2011	7065.62	Norway
49501	NOR	SEX_T	OCU_ISCO08_TOTAL	...	2012	7047.70	Norway
49663	NOR	SEX_T	OCU_ISCO08_TOTAL	...	2014	6950.54	Norway
33795	ISL	SEX_T	OCU_ISCO08_TOTAL	...	2018	6758.99	Iceland
33714	ISL	SEX_T	OCU_ISCO08_TOTAL	...	2017	6664.20	Iceland
49339	NOR	SEX_T	OCU_ISCO08_TOTAL	...	2010	6303.60	Norway
33875	ISL	SEX_T	OCU_ISCO08_TOTAL	...	2019	6190.52	Iceland
12330	CHE	SEX_T	OCU_ISCO08_TOTAL	...	2015	5966.10	Switzerland
49258	NOR	SEX_T	OCU_ISCO08_TOTAL	...	2009	5820.30	Norway
49108	NOR	SEX_T	OCU_ISCO08_TOTAL	...	2018	5608.36	Norway
39884	LUX	SEX_T	OCU_ISCO08_TOTAL	...	2014	5587.67	Luxembourg
48937	NOR	SEX_T	OCU_ISCO08_TOTAL	...	2015	5580.24	Norway
17933	DNK	SEX_T	OCU_ISCO08_TOTAL	...	2014	5570.46	Denmark
33633	ISL	SEX_T	OCU_ISCO08_TOTAL	...	2016	5537.55	Iceland
49197	NOR	SEX_T	OCU_ISCO08_TOTAL	...	2019	5373.86	Norway
49018	NOR	SEX_T	OCU_ISCO08_TOTAL	...	2017	5356.84	Norway
33387	IRL	SEX_T	OCU_ISCO08_TOTAL	...	2014	5019.08	Ireland

1	Вып.	Валькова.Н.П.			Лабораторная работа по предмету «Базы данных»-НГТУ-(18-ПМ)	Лист
2	Пров.	Моисеев А.Е				№
№		Ф.И.О.	Подп.	Дата		7


```
print(data[(data.classif2 == 'CUR_TYPE_USD') &
(data.ref_area == 'RUS') & (data.sex == 'SEX_T') &
(data.classif1_label != 'Total')].sort_values(by = 'obs_value',
ascending = False)[:20])
```

```
/home/nadezhda/Документы/DB/Labs1/venv/bin/python /home/nadezhda/Документы/D
obs_value ... classif1_label
60280 1633.34 ... 1. Legislators, senior officials and managers
60199 1415.17 ... 1. Legislators, senior officials and managers
60548 1303.88 ... 1. Managers
60460 1277.60 ... 1. Managers
60118 1055.63 ... 1. Managers
60283 1036.68 ... 2. Professionals
60370 996.41 ... 1. Managers
60301 878.41 ... 8. Plant and machine operators and assemblers
60298 877.69 ... 7. Craft and related trades workers
60202 850.48 ... 2. Professionals
60286 811.88 ... 3. Technicians and associate professionals
60217 799.02 ... 8. Plant and machine operators and assemblers
60214 786.32 ... 7. Craft and related trades workers
60550 754.55 ... 2. Professionals
60463 743.21 ... 2. Professionals
60552 692.98 ... 3. Technicians and associate professionals
60562 677.92 ... 8. Plant and machine operators, and assemblers
60560 663.35 ... 7. Craft and related trades workers
60205 645.29 ... 3. Technicians and associate professionals
60481 641.50 ... 8. Plant and machine operators, and assemblers

[20 rows x 5 columns]
```

1	Вып.	Валькова.Н.П.			Лабораторная работа по предмету «Базы данных»-НГТУ-(18-ПМ)	Лист
2	Пров.	Моисеев А.Е				№
№		Ф.И.О.	Подп.	Дата		9

- Топ 20 профессий (на которых работают женщины) с самой высокой почасовой оплатой(в долларах) за 2019 год

```
print(data[(data.classif2 == 'CUR_TYPE_USD') & (data.time == 2019) & (data.sex == 'SEX_F')& (data.classif1_label != 'Total')].sort_values(by = 'obs_value', ascending = False)[:20])
```

```

/home/nadezhda/Документы/DB/Labs1/venv/bin/python /home/nadezhda/Документы/DB/Labs1/main.py
obs_value ref_area ... time classif1_label
59877 10052.47 QAT ... 2019 1. Legislators, senior officials and managers
33913 8963.62 ISL ... 2019 1. Managers
61535 7380.37 SGP ... 2019 1. Managers
59879 7112.09 QAT ... 2019 2. Professionals
37995 6885.44 KOR ... 2019 1. Managers
49239 6879.55 NOR ... 2019 1. Managers
24677 6262.34 FIN ... 2019 1. Managers
59881 6151.37 QAT ... 2019 3. Technicians and associate professionals
33915 6149.74 ISL ... 2019 2. Professionals
72029 5953.18 USA ... 2019 1. Managers
49241 5672.73 NOR ... 2019 2. Professionals
33917 5586.97 ISL ... 2019 3. Technicians and associate professionals
3746 5286.16 AUT ... 2019 1. Managers
49243 5243.18 NOR ... 2019 3. Technicians and associate professionals
72032 5119.95 USA ... 2019 2. Professionals
61537 5003.08 SGP ... 2019 2. Professionals
33919 4632.70 ISL ... 2019 4. Clerical support workers
49245 4463.64 NOR ... 2019 4. Clerical support workers
26193 4443.20 GBR ... 2019 1. Managers
24695 4433.12 FIN ... 2019 0. Armed forces occupations

[20 rows x 6 columns]

```

1	Вып.	Валькова.Н.П.			Лабораторная работа по предмету «Базы данных»-НГТУ-(18-ПМ)	Лист
2	Пров.	Моисеев А.Е				№
№		Ф.И.О.	Подп.	Дата		10