

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
ФГАОУ ВО «СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт математики и информационных технологий имени
профессора Н.И.Червякова

Кафедра инфокоммуникаций.

Дисциплина: Технологии программирования

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №2

Основы Pandas

Выполнил: студент 3 курса
09.03.01 «Информатика и вычислительная
техника» группы ИВТ-б-о-19-1
Винников Никита Александрович

Проверил:
Воронкин Роман Александрович

Работа защищена с оценкой:

Лабораторная работа №2

Основы Pandas

Цель работы исследовать базовые возможности системы управления базами данных Pandas

Ход работы:

```
In [110]: import pandas as pd
import numpy as np
import json
import matplotlib.pyplot as plt
%matplotlib inline
```

```
In [111]: city = pd.read_csv('city.csv')
```

Столбцы и количество строк в таблице city

1. С помощью команды `select` вывел информацию о том, сколько городов в каждом из федеральных округов:

```
In [124]: city.columns
```

```
Out[124]: Index(['address', 'postal_code', 'country', 'federal_district', 'region_type',
               'region', 'area_type', 'area', 'city_type', 'city', 'settlement_type',
               'settlement', 'kladr_id', 'fias_id', 'fias_level', 'capital_marker',
               'okato', 'oktmo', 'tax_office', 'timezone', 'geo_lat', 'geo_lon',
               'population', 'foundation_year'],
              dtype='object')
```

```
In [123]: len(city)
```

```
Out[123]: 1117
```

Узнали самое длинное название города в таблице

```
Ввод [137]: 1 q=[]
2 for i in city["city"]:
3     i = str(i)
4     q.append(i)
5     print(f'Самое длинное название города в таблице имеет {len(max(list, key=len))} символов')
```

Самое длинное название города в таблице имеет 12 символов

```
Ввод [138]: 1 time = city[city['federal_district'].isin(['Приволжский', 'Сибирский'])]
2 time[['timezone', 'address']].groupby('timezone').count().rename({'address': 'Количество городов'}, axis=1)
```

Out[138]:

Количество городов	
timezone	
UTC+3	101
UTC+4	41
UTC+5	58
UTC+6	6
UTC+7	86
UTC+8	22

2. С помощью команды select вывел следующую информацию: Сколько городов в Приволжском и Уральском округах?

```
In [130]: time = city[city['federal_district'].isin(['Приволжский', 'Сибирский'])]
time[['timezone', 'address']].groupby('timezone').count().rename({'address': 'Количество городов'}, axis=1)
```

Out[130]:

Количество городов	
timezone	
UTC+3	101
UTC+4	41
UTC+5	58
UTC+6	6
UTC+7	86
UTC+8	22

3. Сформировал подзапрос: “Сколько городов было основано в каждом веке?»:

```
In [148]: q = city[['timezone', 'address']].groupby('timezone').count().sort_values('address', ascending=False)
q.rename({'address': 'Количество городов'}, axis=1)
```

Out[148]:

Количество городов	
timezone	
UTC+3	660
UTC+5	173
UTC+7	86
UTC+4	66
UTC+9	31
UTC+8	28
UTC+10	22
UTC+2	22
UTC+11	17

Индивидуальное задание:

```
In [1]: import pandas as pd
import numpy as np
import json
import matplotlib.pyplot as plt
%matplotlib inline
```

```
In [2]: spotify = pd.read_csv('spotify.csv')
spotify
```

```
Out[2]:
```

	Unnamed: 0	artist	album	track_name	track_id	energy	danceability	key	loudness	acousticness	speechiness	instrumentalness
0	0	The Weeknd	After Hours	Blinding Lights	0VjjW4GIUZAMYd2vXMi3b	0.730	0.514	1	-5.934	0.00146	0.0598	
1	1	Tones And I	Dance Monkey	Dance Monkey	1rgnBhdG2JDFTbYKYRZAku	0.593	0.825	6	-6.401	0.68800	0.0988	
2	2	Roddy Ricch	Please Excuse Me For Being Antisocial	The Box	0nbXyq5TXYPcO7pr3N8S4I	0.586	0.896	10	-6.687	0.10400	0.0559	
3	3	SAINT JHN	Roses (Imanbek Remix)	Roses - Imanbek Remix	2Wo6QQD1KMDWeFkkjLqwx5	0.721	0.785	8	-5.457	0.01490	0.0506	
4	4	Dua Lipa	Future Nostalgia	Don't Start Now	3PflrDoz19wz7qK7tYeu62	0.793	0.793	11	-4.521	0.01230	0.0830	
5	5	DaBaby	BLAME IT ON A REVOLUTION	ROCKSTAR (feat. Roddy Ricch)	7ytR5pFWmSizHJleQkqoq4	0.690	0.746	11	-7.956	0.24700	0.1640	

Все поля таблицы данных

```
In [3]: spotify.columns
```

```
Out[3]: Index(['Unnamed: 0', 'artist', 'album', 'track_name', 'track_id', 'energy',
'danceability', 'key', 'loudness', 'acousticness', 'speechiness',
'instrumentalness', 'liveness', 'valence', 'tempo', 'duration_ms',
'genre'],
dtype='object')
```

```
In [ ]: Общее количество полей
```

```
In [4]: len(spotify)
```

```
Out[4]: 50
```

```
In [ ]: Вывел с помощью сортировки первых 10 артистов и их данные
```

```
In [6]: spotify.sort_values(by=['artist'], ascending=True).head(10)
```

```
Out[6]:
```

	Unnamed: 0	artist	album	track_name	track_id	energy	danceability	key	loudness	acousticness	speechiness	instrumentalness
23	23	24kGoldn	Mood (feat. iann dior)	Mood (feat. iann dior)	3tjFYV6RSftuktYI3ZtYcq	0.722	0.700	7	-3.558	0.2210	0.0369	0.00001
46	46	Ariana Grande	Stuck with U	Stuck with U (with Justin Bieber)	4HBZA5f1ZLE435QTztThqH	0.450	0.597	8	-6.658	0.2230	0.0418	0.00001
11	11	Arizona Zervas	ROXANNE	ROXANNE	696DnlkuDOXcMAnKITgXXK	0.601	0.621	6	-5.616	0.0522	0.1480	0.00001
33	33	BENEE	Hey u x	Supalonely (feat. Gus Dapperton)	3GZoWLVbmxkBys6g0DLFLf	0.631	0.862	7	-4.746	0.2910	0.0515	0.00021
32	32	BTS	Dynamite (DayTime Version)	Dynamite	0t1kP63rueHleOhQKYSXFY	0.765	0.746	6	-4.410	0.0112	0.0993	0.00001
43	43	Bad Bunny	YHLQMDLG	Safaera	2DEZmgHKAvm41k4J3R2E9Y	0.829	0.607	2	-4.074	0.0103	0.3790	0.00001
26	26	Billie Eilish	WHEN WE ALL FALL ASLEEP, WHERE DO WE GO?	bad guy	2Fxmhks0bxGSBdJ92vM42m	0.425	0.701	7	-10.965	0.3280	0.3750	0.13001
24	24	Billie Eilish	everything i wanted	everything i wanted	3ZCTVFBt2Brf31RLEnCkWWJ	0.225	0.704	6	-14.454	0.9020	0.0994	0.65701
44	44	Billie Eilish	lovely (with Khalid)	lovely (with Khalid)	0u2P5u6lvoDfwTYjAADbn4	0.296	0.351	4	-10.109	0.9340	0.0333	0.00001
41	41	Black Eyed Peas	Translation	RITMO (Bad Boys For Life)	4NCsrTzgVfsDo8nWyp8PPc	0.704	0.723	10	-7.088	0.0259	0.0571	0.00101

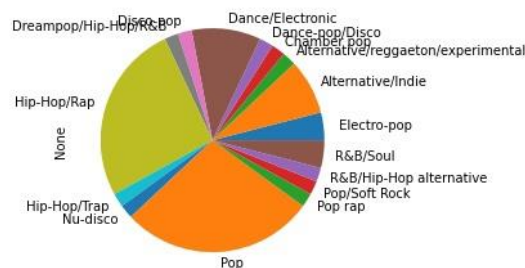
Отобразил последние 10 полей данных

41	41	Black Eyed Peas	Translation	RITMO (Bad Boys For Life)	4NCsrTzgVfsDo8nWyP8PPc	0.704	0.723	10	-7.088	0.02590	0.0571
42	42	THE SCOTTS	THE SCOTTS	THE SCOTTS	39Yp9wwQisRIDOvrVg7mbk	0.537	0.716	0	-7.648	0.23300	0.0514
43	43	Bad Bunny	YHLQMDLG	Safaera	2DEZmgHKAvm41k4J3R2E9Y	0.829	0.607	2	-4.074	0.01030	0.3790
44	44	Billie Eilish	lovely (with Khalid)	lovely (with Khalid)	0u2P5u6lvoDfwTYjAADbn4	0.296	0.351	4	-10.109	0.93400	0.0333
45	45	Juice WRLD	Goodbye & Good Riddance	Lucid Dreams	285pBltuF7vW8TeWk8hdRR	0.566	0.511	6	-7.230	0.34900	0.2000
46	46	Ariana Grande	Stuck with U	Stuck with U (with Justin Bieber)	4HBZA5fIZLE435QTztThqH	0.450	0.597	8	-6.658	0.22300	0.0418
47	47	JP Saxe	If the World Was Ending (feat. Julia Michaels)	If the World Was Ending - feat. Julia Michaels	2kJwzbxV2ppxnQoYw4GLBZ	0.473	0.464	1	-10.086	0.86600	0.1290
48	48	Dua Lipa	Future Nostalgia	Physical	3AzjcOeAmA57TIO9zF1ZW	0.844	0.647	0	-3.756	0.01370	0.0457
49	49	Travis Scott	ASTROWORLD	SICKO MODE	2xLMifQCjDGFmkHkpNLD9h	0.730	0.834	8	-3.714	0.00513	0.2220

In []: Вывел количество жанров музыки и разделил их с помощью круговой диаграммы

In [9]: `spotify.groupby('genre').size().plot(kind = 'pie')`

Out[9]: <AxesSubplot:ylabel='None'>



Контрольные вопросы:

1. Какие существуют средства для импорта данных в SQLite?

```
.import --csv city.csv city
```

```
.mode csv
```

```
.import city.csv city
```

2. В чем недостатки локальных и централизованных СКВ?

Локальные СКВ

Легко запутаться в файлах и в каком конкретно ты находишься, можно изменить не тот файл или сохранить в не нужной директории

Централизованные

Централизованный сервер является уязвимым местом всей системы. Если сервер выключается на час, то в течение часа разработчики не могут взаимодействовать, и никто не может сохранить новые версии. Если же повреждается диск с центральной базой данных и нет резервной копии, вы теряете абсолютно всё - всю историю проекта, разве что за исключением

нескольких рабочих версий, сохранившихся на рабочих машинах пользователей. Локальные системы управления версиями подвержены той же проблеме: если вся история проекта хранится в одном месте, вы рискуете потерять всё.

3. Каково назначение команды `.schema` ? показывает список и структуру всех таблиц в базе

4. Как выполняется группировка и сортировка данных в запросах SQLite?

Группировка производится, когда в оператора `group` используется функция `by`

5. Каково назначение "табличных выражений" в SQLite?

Наглядная демонстрация содержимого в таблице

6. Как осуществляется экспорт данных из SQLite в форматы CSV и JSON?

```
sqlite> .mode csv sqlite> .once samara.csv sqlite> select  
kladr_id, city from city where region = 'Самарская'; sqlite> .exit  
.mode json select kladr_id, city from city where region =  
'Самарская' limit 3;
```

7. Какие еще форматы для экспорта данных Вам известны?

```
.mode insert cities
```

```
.mode markdown
```

```
.mode html
```