

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
Демьяненко Александр Владимирович

Проверил:
Воронкин Роман Александрович

Работа защищена с оценкой:

Ставрополь, 2021

Лабораторная работа №2

Основы Pandas

Цель работы исследовать базовые возможности системы управления базами данных Pandas

Ход работы:

1. Создал свой общедоступный репозиторий на Github, выбрав лицензию MIT и язык программирования Python.

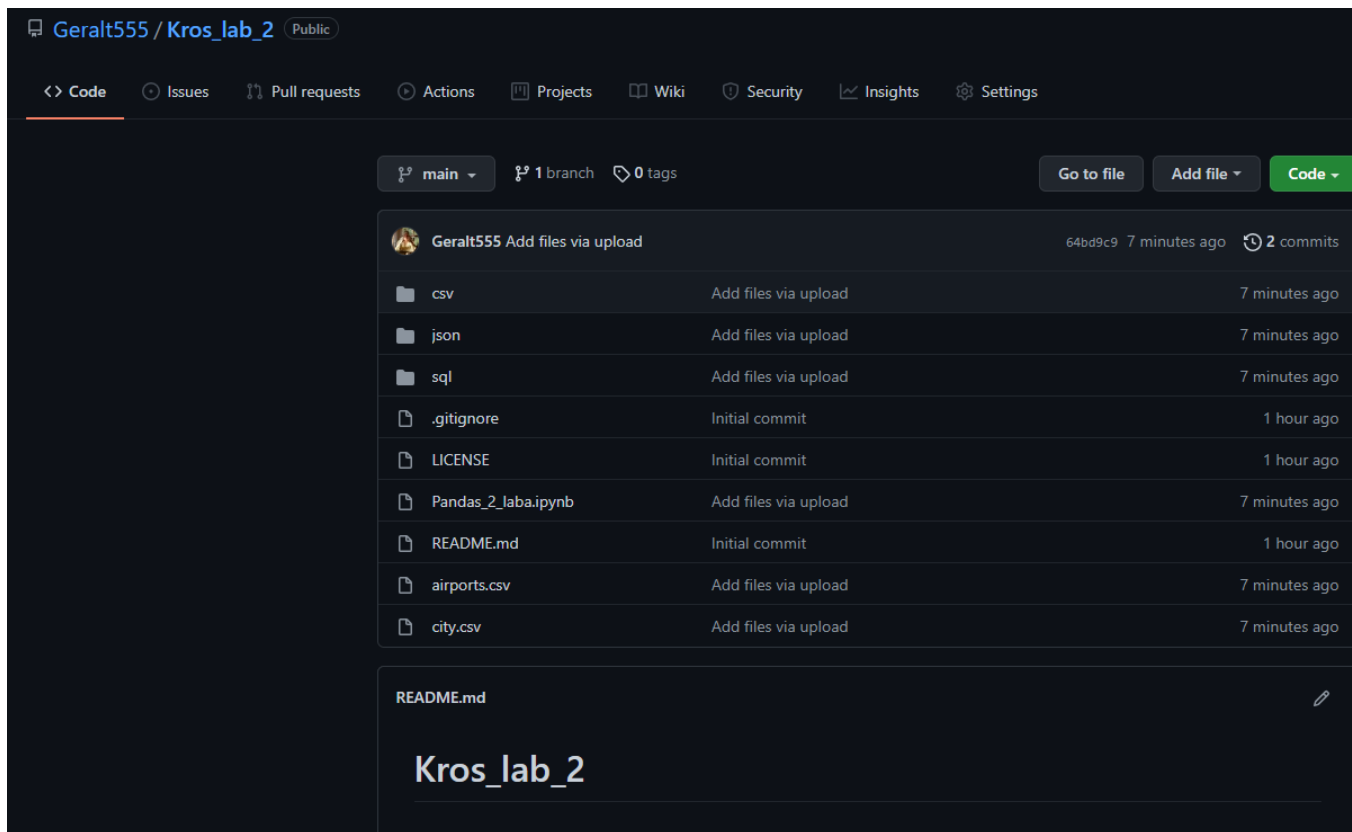


Рисунок 1. Создание нового репозитория

2. Организовал свой репозиторий в соответствии с моделью ветвления gitflow
3. Скачал файл и проверил содержимое таблицы:

```
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

4. 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
```

5. С помощью команды select вывел следующую информацию:

- 1) У каких городов в названии есть слово «Красный»?

Узнали самое длинное название города в таблице

```
In [128]: q=[]
for i in city["city"]:
    i = str(i)
    q.append(i)
print(f'Самое длинное название города в таблице имеет {len(max(list, key=len))} символов')
```

Самое длинное название города в таблице имеет 25 символов

Количество городов для каждого часового пояса в Сибирском и Приволжском федеральных округах
 Ответ: 58

```
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

- 2) Какие города появились за последние 30 лет?

- 3) Сколько городов в Приволжском и Уральском округах?

```
In [148]: q = city[['timezone', 'address']].groupby('timezone').count().sort_values('address', ascending=True)
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

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

Индивидуальное задание

```
In [6]: airports = pd.read_csv('airports.csv')
airports
```

Out[6]:

	id	ident	type	name	latitude_deg	longitude_deg	elevation_ft	continent	iso_country	iso_region	municipality	scheduled_service	gps
0	6523	00A	heliport	Total Rf Heliport	40.070801	-74.933601	11.0	NaN	US	US-PA	Bensalem	no	
1	323361	00AA	small_airport	Aero B Ranch Airport	38.704022	-101.473911	3435.0	NaN	US	US-KS	Leoti	no	
2	6524	00AK	small_airport	Lowell Field	59.947733	-151.692524	450.0	NaN	US	US-AK	Anchor Point	no	
3	6525	00AL	small_airport	Epps Airpark	34.864799	-86.770302	820.0	NaN	US	US-AL	Harvest	no	
4	6526	00AR	closed	Newport Hospital & Clinic Heliport	35.608700	-91.254898	237.0	NaN	US	US-AR	Newport	no	
...
68328	46378	ZZ-0001	heliport	Sealand Helipad	51.894444	1.482500	40.0	EU	GB	GB-ENG	Sealand	no	
68329	307326	ZZ-0002	small_airport	Glorioso Islands Airstrip	-11.584278	47.296389	11.0	AF	TF	TF-U-A	Grande Glorieuse	no	
68330	346788	ZZ-0003	small_airport	Fainting Goat Airport	32.110587	-97.356312	690.0	NaN	US	US-TX	Blum	no	
68331	342102	ZZZW	closed	Scandium City Heliport	69.355287	-138.939310	4.0	NaN	CA	CA-YT	(Old) Scandium City	no	
68332	313629	ZZZZ	small_airport	Satsuma Iōjima Airport	30.784722	130.270556	338.0	AS	JP	JP-46	Mishima	no	

68333 rows x 18 columns

7. Сортировка по имени

```
In [9]: df1 = pd.DataFrame({'id' : [1, 2], 'name': ['Harry Potter', 'Ron Weasley']})
df2 = pd.DataFrame({'id' : [3], 'name': ['Hermione Granger']})
pd.concat([df1, df2]).reset_index(drop=True)
```

```
Out[9]:
```

	id	name
0	1	Harry Potter
1	2	Ron Weasley
2	3	Hermione Granger

8. Сортировка по 2-ум параметрам

```
In [14]: airports[airports.type.isin(['heliport', 'balloonport'])]
```

```
Out[14]:
```

	id	ident	type	name	latitude_deg	longitude_deg	elevation_ft	continent	iso_country	iso_region	municipality	scheduled_service	gps_coord
0	6523	00A	heliport	Total Rf Heliport	40.070801	-74.933601	11.0	NaN	US	US-PA	Bensalem	no	00C
9	322658	00CN	heliport	Kitchen Creek Heliport	32.727374	-116.459742	3350.0	NaN	US	US-CA	Pine Valley	no	00C
12	6532	00FD	heliport	Ringhaver Heliport	28.846600	-82.345398	25.0	NaN	US	US-FL	Riverview	no	00F
15	6535	00GE	heliport	Caffrey Heliport	33.889245	-84.737930	957.0	NaN	US	US-GA	Hiram	no	00C
16	6536	00HI	heliport	Kaupulehu Heliport	19.832715	-155.980233	43.0	NaN	US	US-HI	Kailua-Kona	no	00I
...
67976	347768	ZA-0251	heliport	Eagle View Helipad	-25.390456	30.172134	NaN	AF	ZA	ZA-MP	Walkersons	no	Na
68029	339169	ZGNT	heliport	Shenzhen Nantou Heliport	22.558736	113.925612	NaN	AS	CN	CN-44	Shenzhen (Nanshan)	no	ZGN
68054	301278	ZIZ	heliport	Zamzama Heliport	26.710944	67.667250	128.0	AS	PK	PK-SD	Zamzama Gas Field	no	Na
68277	345912	ZW-0048	heliport	Chinyike Mine Heliport	-18.925900	30.274200	NaN	AF	ZW	ZW-U-A	NaN	no	Na
68328	46378	ZZ-0001	heliport	Sealand Helipad	51.894444	1.482500	40.0	EU	GB	GB-ENG	Sealand	no	Na

16598 rows x 18 columns

9. Сортировка по аэропортам в Америке

```
In [18]: airports[(airports.iso_region == 'US-CA') & (airports.type == 'large_airport')][['ident', 'name', 'municipality']]
```

```
Out[18]:
```

	ident	name	municipality
32088	KACV	California Redwood Coast-Humboldt County Inter...	Arcata/Eureka
32238	KBAB	Beale Air Force Base	Marysville
32491	KCMA	Camarillo International Airport	Camarillo
33084	KEDW	Edwards Air Force Base	Edwards
33859	KL45	Bakersfield International Airport	Bakersfield
33883	KLAX	Los Angeles International Airport	Los Angeles
34412	KOAK	Metropolitan Oakland International Airport	Oakland
34468	KONT	Ontario International Airport	Ontario
35998	KSAN	San Diego International Airport	San Diego
36038	KSFO	San Francisco International Airport	San Francisco
36059	KSJC	Norman Y. Mineta San Jose International Airport	San Jose
36081	KSMF	Sacramento International Airport	Sacramento
36387	KVIS	Visalia International Airport	Visalia

10. График

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