

РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное
образовательное учреждение высшего образования
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Кафедра инфокоммуникаций
«Основы работы с DuckDB»

Отчет по лабораторной работе
по дисциплине «Программирование на Python»

Выполнил студент группы ИВТ-б-о-21-1

Шайдеров Дмитрий Викторович.

«20» мая 2023г.

Подпись студента _____

Работа защищена « » _____ 20__ г.

Проверил Воронкин Р.А. _____
(подпись)

Ставрополь 2023

Цель работы: исследовать базовые возможности системы управления базами данных DuckDB.

Порядок выполнения работы:

1. Решить задачи:

Задание 1. Выполните команды. Что вернула команда .schema ?

```
D create table customer(name string);
D select * from customer;

name
varchar

0 rows

D .schema customer
CREATE TABLE customer("name" VARCHAR);;
```

Рисунок 1 - Результат выполнения задания 1

Команда .schema customer вывела запрос CREATE, соответствующий таблице customer.

Задание 2. С помощью команды .help найдите команду, которая отвечает за вывод времени выполнения запроса.

```
D .timer on
D select count(*) from read_csv_auto('./city.csv');

count_star()
int64

1117

Run Time (s): real 0.078 user 0.046875 sys 0.000000
```

Рисунок 5 - Результат выполнения задания 2

Задание 3. Загрузить файл city.csv. Выполните запрос. Какое число он вернул?

```
D select max(length(city)) from read_csv_auto('./city.csv');

max(length(city))
int64

25
```

Рисунок 6 - Результат выполнения задания 3

Задание 4. загрузите файл city.csv, но без использования опции --csv.

```
D create table city as select * from read_csv_auto('./city.csv');
D
```

Рисунок 7 - Результат выполнения задания 4

Задание 5. Напишите запрос, который посчитает количество городов для каждого часового пояса в Сибирском и Приволжском федеральных округах. Выведите столбцы `timezone` и `city_count`, отсортируйте по значению часового пояса.

```
D select timezone, count(city) as city_count from read_csv_auto('./city.csv')
> where federal_district in ('Приволжский', 'Сибирский')
> group by timezone
> order by timezone ASC;
```

timezone varchar	city_count int64
UTC+3	101
UTC+4	41
UTC+5	58
UTC+6	6
UTC+7	86
UTC+8	22

Рисунок 8 - Результат выполнения задания 5

Задание 6. Напишите запрос, который посчитает количество городов в каждом часовом поясе. Отсортируйте по количеству городов по убыванию.

```
D create table city as select * from read_csv_auto('./city.csv');
D select
> timezone, count(city) as city_count from city
> group by timezone
> order by city_count DESC;
```

timezone varchar	city_count int64
UTC+3	656
UTC+5	173
UTC+7	86
UTC+4	66
UTC+9	31
UTC+8	28
UTC+10	22
UTC+2	22
UTC+11	17
UTC+12	6
UTC+6	6
11 rows	2 columns

Рисунок 9 - Результат выполнения задания 6

2. Выполните индивидуальное задание. Каждый запрос к базе данных сохраните в файл с расширением sql.

Загрузите в SQLite выбранный Вами датасет в формате CSV. Сформируйте более пяти запросов к таблицам БД. Выгрузите результат выполнения запросов в форматы CSV и JSON.

Запрос 1. Выбрать 10 книг с самым большим рейтингом. Отсортировать в порядке убывания рейтинга.

```
D create table books as select * from read_csv_auto('./books.csv');
D copy (select title, authors, average_rating from books
> where length(title) < 50 order by average_rating
> DESC
> limit 10) to 'request_1.csv' (header, delimiter ',');
D copy (select title, authors, average_rating from books
> where length(title) < 50 order by average_rating
> DESC
> limit 10) to 'request_1.json';
D select title, authors, average_rating from books
> where length(title) < 50 order by average_rating desc limit 10;
```

title varchar	authors varchar	average_rating double
Willem de Kooning: Late Paintings	Julie Sylvester/David Sylvester	5.0
Middlesex Borough (Images of America: New :	Middlesex Borough Heritage Committee	5.0
Winchester Shotguns	Dennis Adler/R.L. Wilson	5.0
Taxation of Mineral Rents	Ross Garnaut	5.0
The New Big Book of America	Todd Davis/Marc Frey	5.0
Fanning the Flame: Bible Cross and Mission	Chris Green/Chris Wright/Paul Douglas Gardner	5.0
Tyrannosaurus Wrecks (Stanley #1)	Laura Driscoll/Alisa Klayman-Grodsky/Eric Weiner	5.0
The Irish Anatomist: A Study of Flann O'Br:	Keith Donohue	5.0
Bill Gates: Computer Legend (Famous Lives)	Sara Barton-Wood	5.0
Existential Meditation	Simon Cleveland	4.91
10 rows		3 columns

Рисунок 10 - Результат выполнения запроса 1 индивидуального задания

Запрос 2. Выбрать 5 книг на английском языке с наименьшим числом оценок. Отсортировать в порядке возрастания числа оценок.

```
D copy(
> select title, ratings_count from books
> order by ratings_count desc limit 5) to 'request_2.csv' (header, delimiter ',')
D copy(
> select title, ratings_count from books
> order by ratings_count desc limit 5) to 'request_2.json';
D select title, ratings_count from books
> order by ratings_count desc limit 5;
```

title varchar	ratings_count int64
Twilight (Twilight #1)	4597666
The Hobbit or There and Back Again	2530894
The Catcher in the Rye	2457092
Angels & Demons (Robert Langdon #1)	2418736
Harry Potter and the Prisoner of Azkaban (Harry Potter #3)	2339585

Рисунок 11 - Результат выполнения запроса 2 индивидуального задания

Запрос 3. Посчитать количество книг на каждом языке.

```
D copy(> select language_code as language, count(title) as book_count from books
> group by language) to 'request_3.csv' (header, delimiter ',');
D copy(> select language_code as language, count(title) as book_count from books
> group by language) to 'request_3.json';
D select language_code as language, count(title) as book_count from books
> group by language;
```

language varchar	book_count int64
eng	8908
en-US	1408
fre	144
spa	218
en-GB	214
mul	19
grc	11
enm	3
en-CA	7
ger	99
jpn	46
ara	1
nl	1
zho	14
lat	3
por	10
srp	1
ita	5
rus	2
msa	1
glg	1
wel	1
swe	2
nor	1
tur	1
gla	1
ale	1
27 rows	2 columns

Рисунок 12 - Результат выполнения запроса 3 индивидуального задания

Запрос 4. Выбрать 10 авторов с наивысшим общим рейтингом за все написанные ими книги. Отсортировать в порядке убывания рейтинга.

```

D copy(
> select authors, (sum(average_rating) / count(title)) as rating from books where length(authors) < 50
> group by authors order by rating desc limit 10) to 'request_4.csv' (header, delimiter ',');
D copy(
> select authors, (sum(average_rating) / count(title)) as rating from books where length(authors) < 50
> group by authors order by rating desc limit 10) to 'request_4.json';
D select authors, (sum(average_rating) / count(title)) as rating from books where length(authors) < 50
> group by authors order by rating desc limit 10;

```

authors varchar	rating double
John Diamond	5.0
Elena N. Mahlow	5.0
Ian Martin/Katie Elliott	5.0
Dennis Adler/R.L. Wilson	5.0
R. McL. Wilson	5.0
Nicholas Evans/Rhonda Evans	5.0
Aristophanes/F.W. Hall/W.M. Geldart	5.0
Julie Sylvester/David Sylvester	5.0
Middlesex Borough Heritage Committee	5.0
Chris Green/Chris Wright/Paul Douglas Gardner	5.0

10 rows 2 columns

Рисунок 13 - Результат выполнения запроса 4 индивидуального задания

Запрос 5. Выбрать любые 15 книг с числом письменных рецензий от 1000 до 2000.

```

D copy(
> select title, text_reviews_count from books
> where length(title) < 50 and text_reviews_count between 1000 and 2000 limit 15) to 'request_5.csv' (header, delimiter ',');
D copy(
> select title, text_reviews_count from books
> where length(title) < 50 and text_reviews_count between 1000 and 2000 limit 15) to 'request_5.json';
D select title, text_reviews_count from books
> where length(title) < 50 and text_reviews_count between 1000 and 2000 limit 15;

```

title varchar	text_reviews_count int64
Anna Karenina	1851
I am Charlotte Simmons	1688
The Portrait of a Lady	1951
Slouching Towards Bethlehem	1825
Play It As It Lays	1706
The New York Trilogy	1873
The Brooklyn Follies	1157
Killing Yourself to Live: 85% of a True Story	1109
We the Living	1086
Digging to America	1841
Quicksilver (The Baroque Cycle #1)	1735
The Design of Everyday Things	1379
Pompeii	1726
The Egypt Game	1130
Gates of Fire	1629

15 rows 2 columns

Рисунок 14 - Результат выполнения запроса 5 индивидуального задания

Запрос 6. Выбрать любые 5 книг на французском или испанском языке с рейтингом выше 4 и числом страниц больше 500.

```

D copy(
> select title, language_code as language, num_pages from books
> where length(title) < 50 and language_code in ('fre', 'spa') and num_pages > 500
> limit 15) to 'request_6.csv' (header, delimiter ',');
D copy(
> select title, language_code as language, num_pages from books
> where length(title) < 50 and language_code in ('fre', 'spa') and num_pages > 500
> limit 15) to 'request_6.json';
D select title, language_code as language, num_pages from books
> where length(title) < 50 and language_code in ('fre', 'spa') and num_pages > 500
> limit 15;

```

title varchar	language varchar	num_pages int64
Timbuktu / Leviathan / Moon Palace	fre	1075
Da Vinci Code (Robert Langdon #2)	fre	744
Endymion	fre	572
V.	fre	640
Angeles y demonios (Robert Langdon #1)	spa	508
El Coleccionista De Huesos (Lincoln Rhyme #1)	spa	640
Hyperion (Los Cantos de Hyperion #1)	spa	618
El codigo Da Vinci (Robert Langdon #2)	spa	557
Chaman (Familia Cole #2)	spa	768
L'Echiquier du mal	fre	999
Dreamcatcher	fre	688
Los Versos Sata?nicos	spa	679
El nombre de la rosa	spa	782
Insomnia	spa	890
El Clan Del Oso Cavernario	spa	548
15 rows		3 columns

Рисунок 15 - Результат выполнения запроса 6 индивидуального задания

Вывод: были исследованы базовые возможности системы управления базами данных DuckDB.