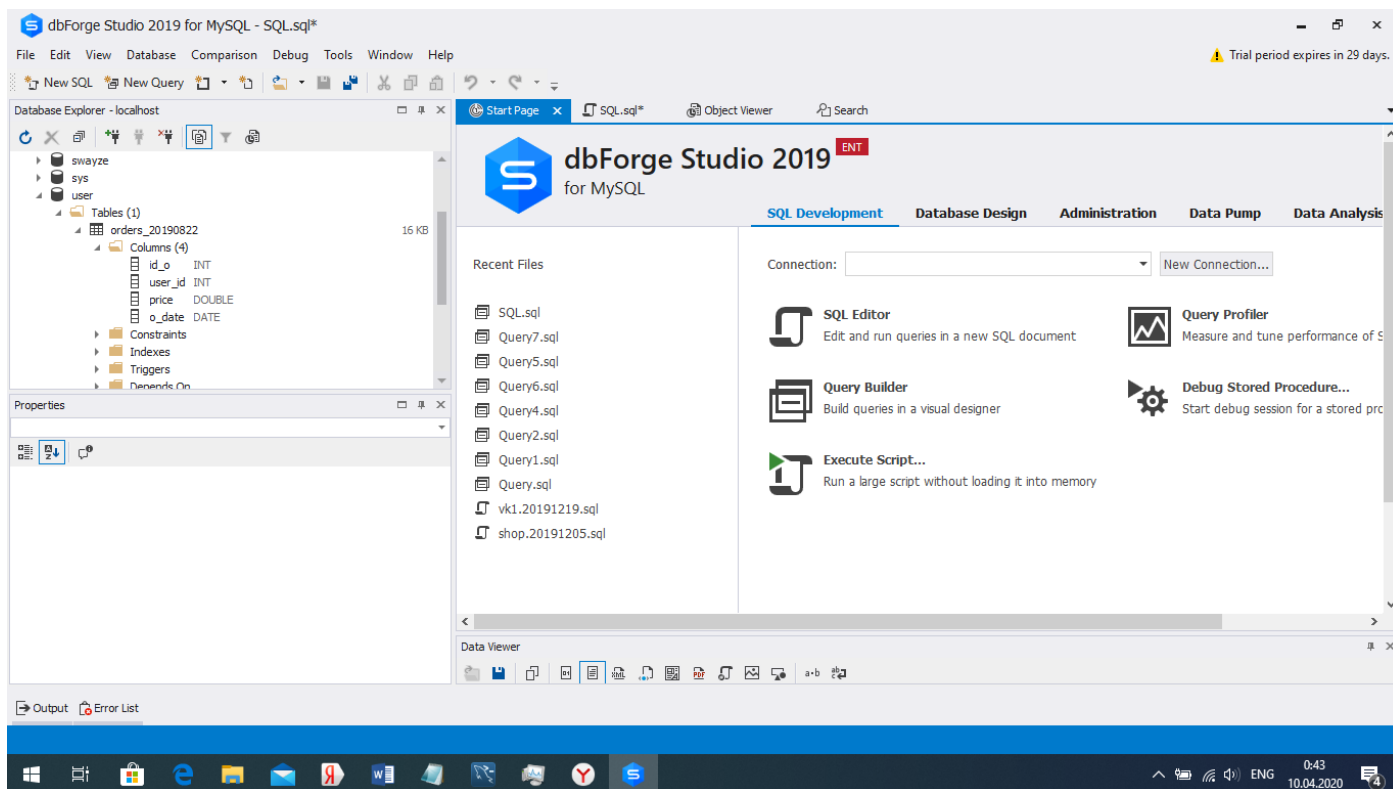


Базы данных для аналитиков. Основы ETL

Домашнее задание 1. Аналитика в бизнес-задачах

1. Залить в свою БД данные по продажам



2. Проанализировать, какой период данных выгружен

SELECT

MIN(orders_20190822.o_date) AS StartDate,

MAX(orders_20190822.o_date) AS `End Date`

FROM orders_20190822

dbForge Studio 2019 for MySQL - Query.sql

File Edit View Database Comparison Data Debug Tools Window Help

Database Explorer - localhost

- swayze
- sys
- user
 - Tables (1)
 - orders_20... 16 KB
 - Columns (4)
 - id_o INT
 - user_id INT
 - price DOI
 - o_date DAT
 - Constraints
 - Indexes
 - Triggers

Properties

Query.sql

StartDate	End Date
DATE	DATE
01.01.2016	31.12.2017

Record 1 of 1

Query executed successfully. 0:00:03.153 localhost (8.0) root user

Data Viewer

Output Error List

Ready

Windows taskbar: 0:51 10.04.2020

dbForge Studio 2019 for MySQL - Query.sql

File Edit View Database Comparison Data Debug Tools Window Help

Database Explorer - localhost

- swayze
- sys
- user
 - Tables (1)
 - orders_20... 16 KB
 - Columns (4)
 - id_o INT
 - user_id INT
 - price DOI
 - o_date DAT
 - Constraints
 - Indexes
 - Triggers

Properties

Query.sql

StartDate	End Date
DATE	DATE
01.01.2016	31.12.2017

Record 1 of 1

Query executed successfully. 0:00:03.153 localhost (8.0) root user

Data Viewer

Output Error List

Ready

Windows taskbar: 0:51 10.04.2020

3. Посчитать кол-во строк, кол-во заказов и кол-во уникальных пользователей, кот совершали заказы.

SELECT

COUNT(*) AS Total_Orders,

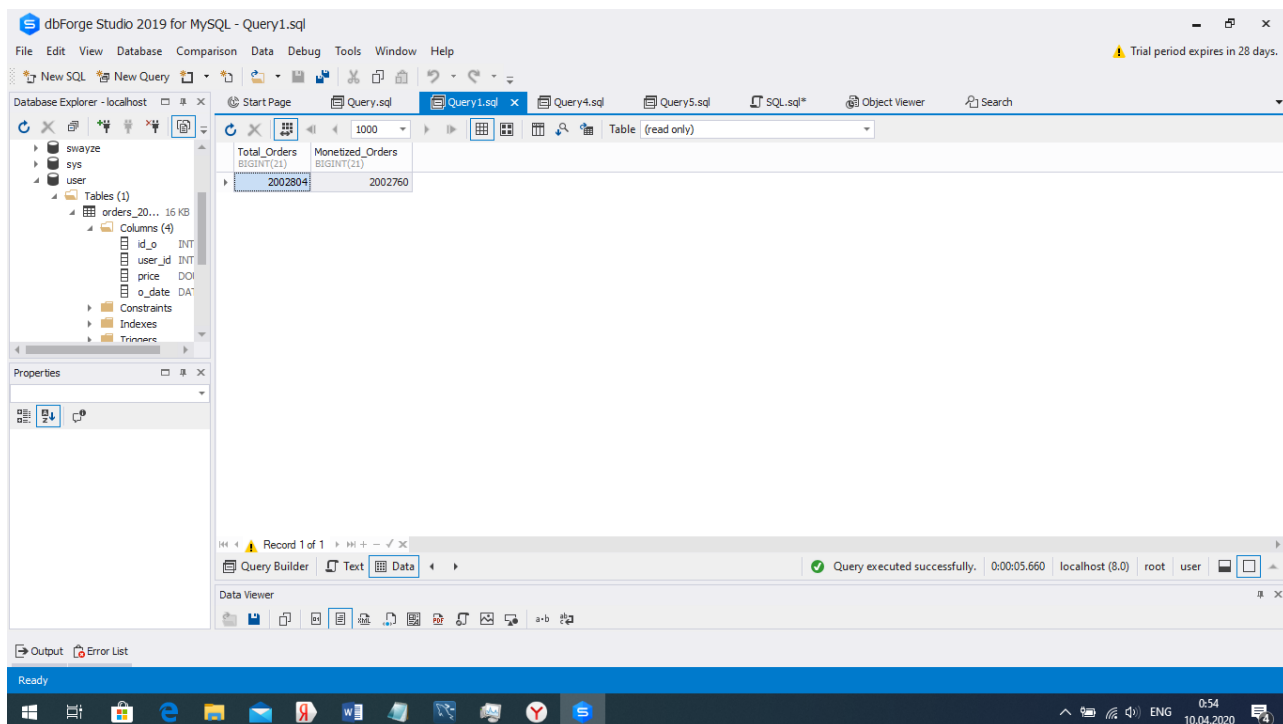
(SELECT COUNT(DISTINCT(id_o)) AS Number_of_orders

FROM orders_20190822

WHERE price > 0 AND price IS NOT NULL

) AS Monetized_Orders

FROM orders_20190822



dbForge Studio 2019 for MySQL - Query1.sql

File Edit View Database Comparison Data Debug Tools Window Help

Start Page Query1.sql Query4.sql Query5.sql SQL.sql* Object Viewer Search

Database Explorer - localhost

- swayze
- sys
- user
 - Tables (1)
 - orders_20... 16 KB
 - Columns (4)
 - id_o INT
 - user_id INT
 - price DOI
 - o_date DAT
 - Constraints
 - Indexes
 - Triggers

Properties

Table (read only)

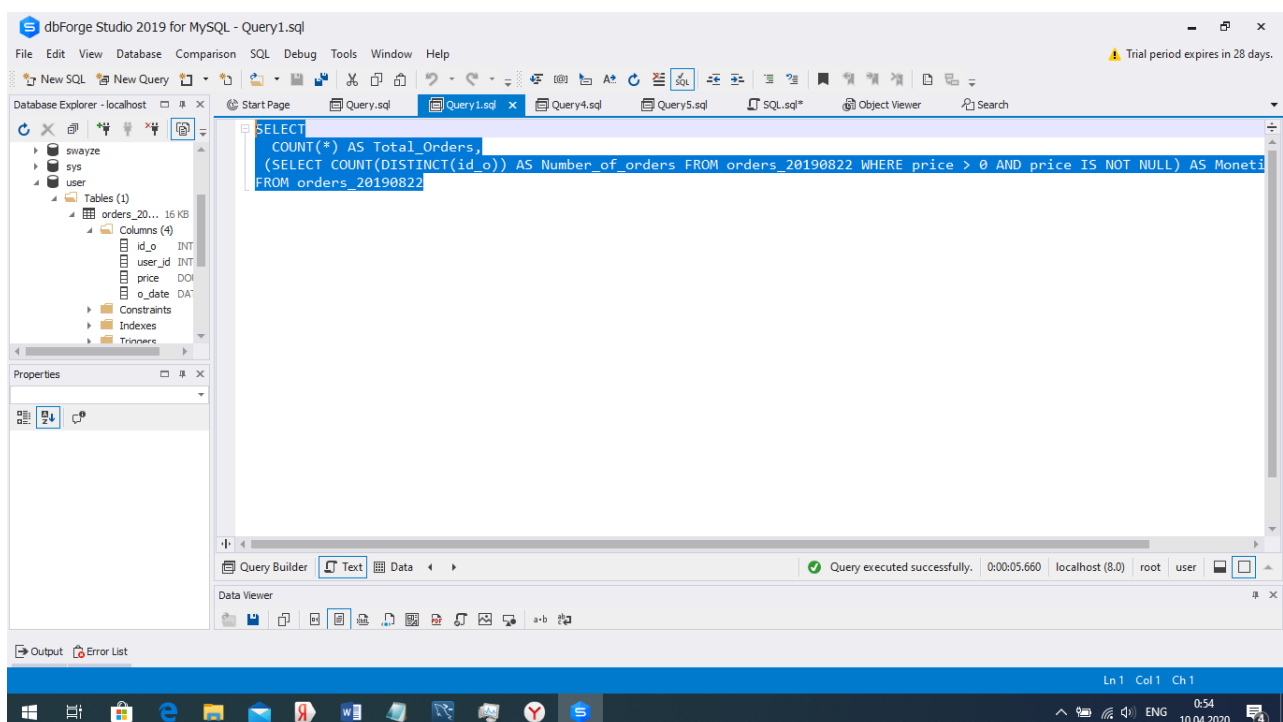
Total_Orders	Monetized_Orders
2002804	2002760

Record 1 of 1

Query executed successfully. 0:00:05.660 localhost (8.0) root user

Output Error List

Ready



dbForge Studio 2019 for MySQL - Query1.sql

File Edit View Database Comparison SQL Debug Tools Window Help

New SQL New Query

Start Page Query1.sql Query4.sql Query5.sql SQL.sql* Object Viewer Search

Database Explorer - localhost

- swayze
- sys
- user
 - Tables (1)
 - orders_20... 16 KB
 - Columns (4)
 - id_o INT
 - user_id INT
 - price DOI
 - o_date DAT
 - Constraints
 - Indexes
 - Triggers

Properties

Query Builder Text Data

```
SELECT
COUNT(*) AS Total_Orders,
(SELECT COUNT(DISTINCT(id_o)) AS Number_of_orders FROM orders_20190822 WHERE price > 0 AND price IS NOT NULL) AS Monetized_Orders
FROM orders_20190822
```

Query executed successfully. 0:00:05.660 localhost (8.0) root user

Output Error List

Ln1 Col1 Ch1

Ready

```
SELECT COUNT(DISTINCT(user_id)) AS unique_users  
FROM orders_20190822
```

The screenshot shows the dbForge Studio 2019 for MySQL interface. The Database Explorer on the left shows the 'user' table. The main window displays the results of a query in the 'Data' tab. The query is: `SELECT COUNT(DISTINCT(user_id)) AS unique_users FROM orders_20190822`. The result is a single row with the value 1015119. The status bar at the bottom indicates 'Query executed successfully. 0:00:04.441 localhost (8.0) root user'.

unique_users
1015119

The screenshot shows the dbForge Studio 2019 for MySQL interface. The Database Explorer on the left shows the 'user' table. The main window displays the SQL query editor with the query: `SELECT COUNT(DISTINCT(user_id)) AS unique_users FROM orders_20190822`. The status bar at the bottom indicates 'Query executed successfully. 0:00:04.441 localhost (8.0) root user'.

4. По годам посчитать средний чек, среднее кол-во заказов на пользователя, сделать вывод , как изменялись эти показатели Год от года.

Вопрос, я пытался через LEAD and LAG прописать PERCENTAGE CHANGE но вместо цифр я получал NULL. Подскажите пожалуйста как прописать PERCENTAGE CHANGE через WINDOW FUNCTION чтобы указать Прирост на 14.4%

НЕ РАБОЧИЙ ВАРИАНТ

USE user

```
SELECT the_table.avg_sales AS Average_Sales,
       the_table.Year,
       100*((the_table.next - the_table.prev)/the_table.prev) AS PCT_CHANGE
FROM
(
    SELECT
        ROUND(AVG(price)) AS avg_sales,
        LEAD(AVG(price),1) OVER (ORDER BY year(o_date) ) AS next,
        LAG(AVG(price),1) OVER (ORDER BY year(o_date) ) AS prev,
        year(o_date) AS Year
    FROM orders_20190822
    GROUP BY Year
) AS the_table;
```

РАБОЧИЙ ВАРИАНТ

```
SELECT
    ROUND(AVG(curr.price)) AS avg_sales,
    year(curr.o_date) AS Year
FROM orders_20190822 curr
GROUP BY Year;
```

dbForge Studio 2019 for MySQL - SQL.sql*

File Edit View Database Comparison Data SQL Debug Tools Window Help

Database Explorer - localhost

Tables (1)

- orders_20... 16 KB
 - Columns (4)
 - id_o INT
 - user_id INT
 - price DOI
 - o_date DAT

Properties

Query1 Query2

avg_sales Year

avg_sales DOUBLE(17, 0)	Year INT(5)
2096	2016
2398	2017

Record 2 of 2

Query executed successfully. 0:00:07.013 localhost (8.0) root user

Output Error List

Windows taskbar: 1:01 10.04.2020

dbForge Studio 2019 for MySQL - SQL.sql*

File Edit View Database Comparison Data SQL Debug Tools Window Help

Database Explorer - localhost

Tables (1)

- orders_20... 16 KB
 - Columns (4)
 - id_o INT
 - user_id INT
 - price DOI
 - o_date DAT

Properties

Query1 Query2

Query2

```
USE user

SELECT the_table.avg_sales AS Average_Sales,
       the_table.Year,
       100*((the_table.next - the_table.prev)/the_table.prev) AS PCT_CHANGE
FROM(
    SELECT
        ROUND(AVG(price)) AS avg_sales,
        LEAD(AVG(price),1) OVER (ORDER BY year(o_date) ) AS next,
        LAG(AVG(price),1) OVER (ORDER BY year(o_date) ) AS prev,
        year(o_date) AS Year
    FROM orders_20190822
    GROUP BY Year
) AS the_table;

SELECT
    ROUND(AVG(curr.price)) AS avg_sales,
    year(curr.o_date) AS Year
FROM orders_20190822 curr
GROUP BY Year;
```

Query executed successfully. 0:00:07.013 localhost (8.0) root user

Output Error List

Windows taskbar: 1:02 10.04.2020

```

SELECT ROUND(orders_per_client,2) AS Per_Client_Orders, Year # HOW TO ADD A PERCENT CHANGE HERE

FROM (

SELECT

COUNT(id_o)/COUNT(DISTINCT(user_id)) AS orders_per_client,

year(o_date) AS Year

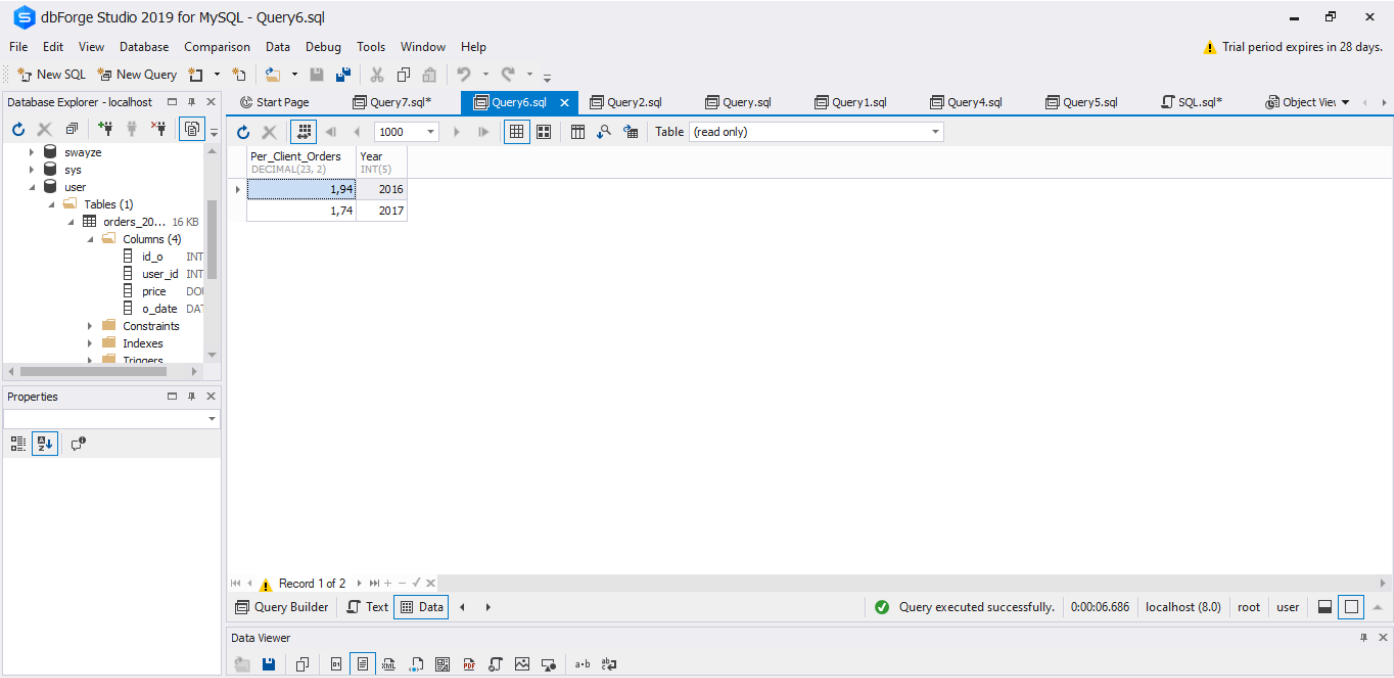
FROM orders_20190822

GROUP BY Year

)

AS the_table

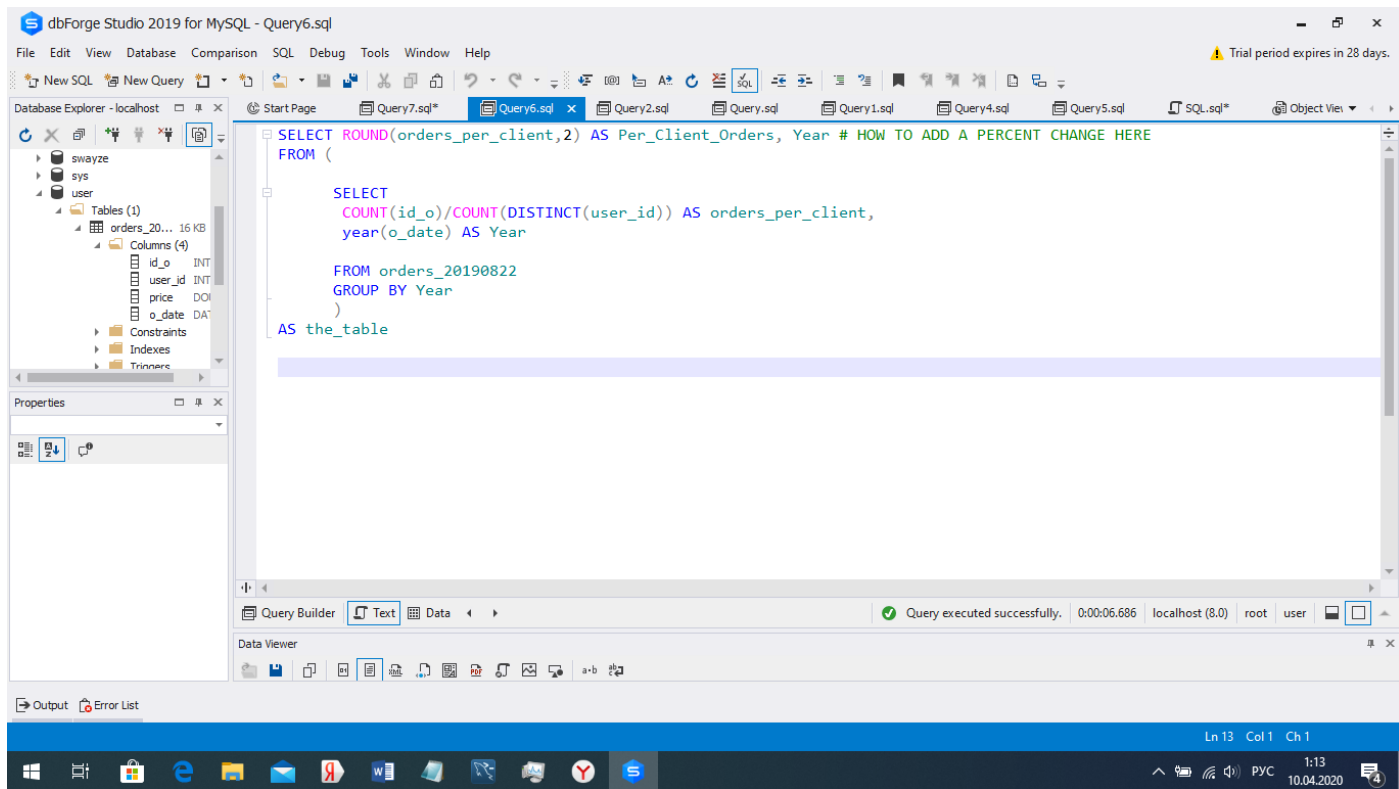
```



The screenshot shows the dbForge Studio 2019 for MySQL interface. The 'Query6.sql' window displays the results of a query in a table view. The table has two columns: 'Per_Client_Orders' (DECIMAL(23, 2)) and 'Year' (INT(5)). The results show two rows of data for the years 2016 and 2017.

Per_Client_Orders	Year
1,94	2016
1,74	2017

The status bar at the bottom indicates that the query was executed successfully on 10.04.2020 at 1:13.



Снижение на 10%

5. Найти кол-во пользователей, которые покупали в одном году и перестали покупать в следующем.

Два рабочих варианта

```

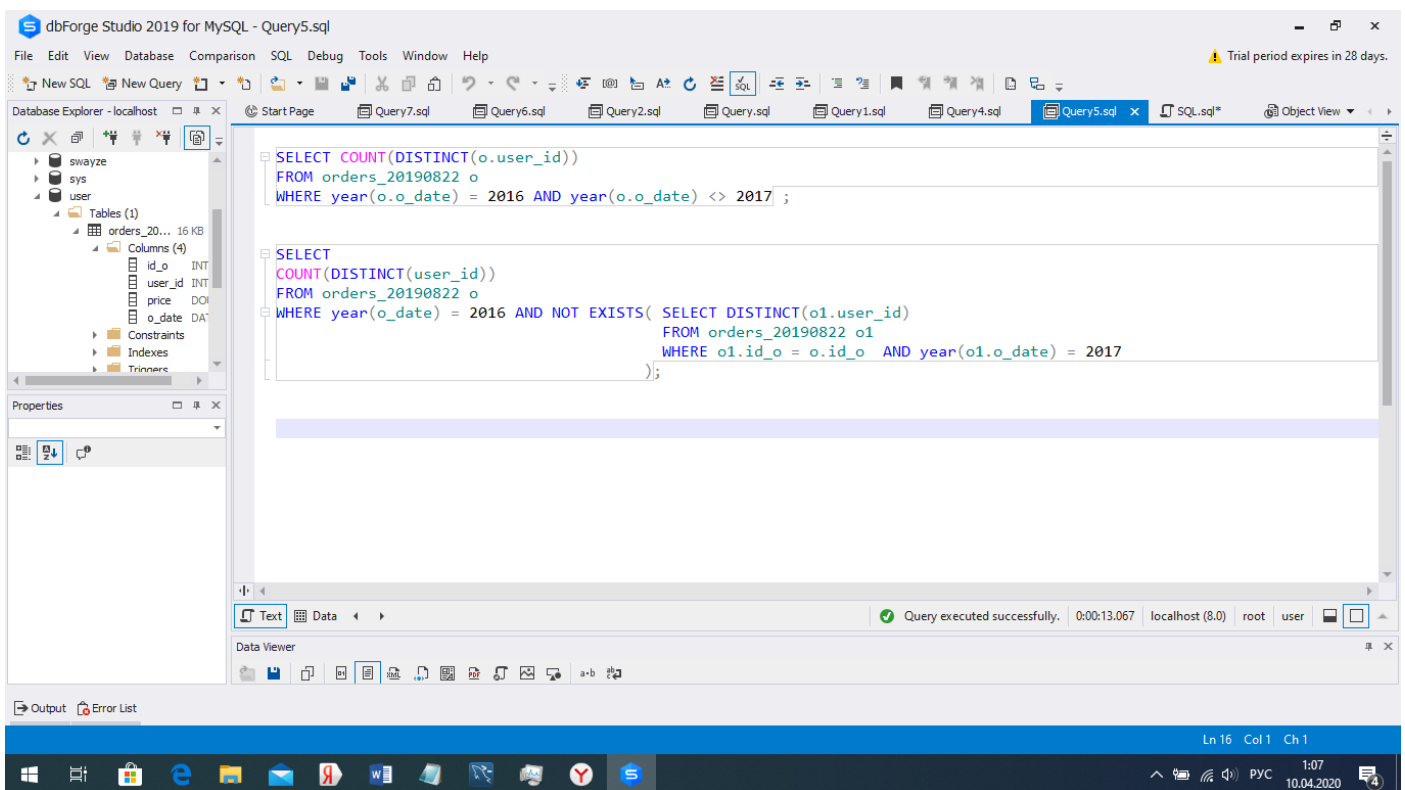
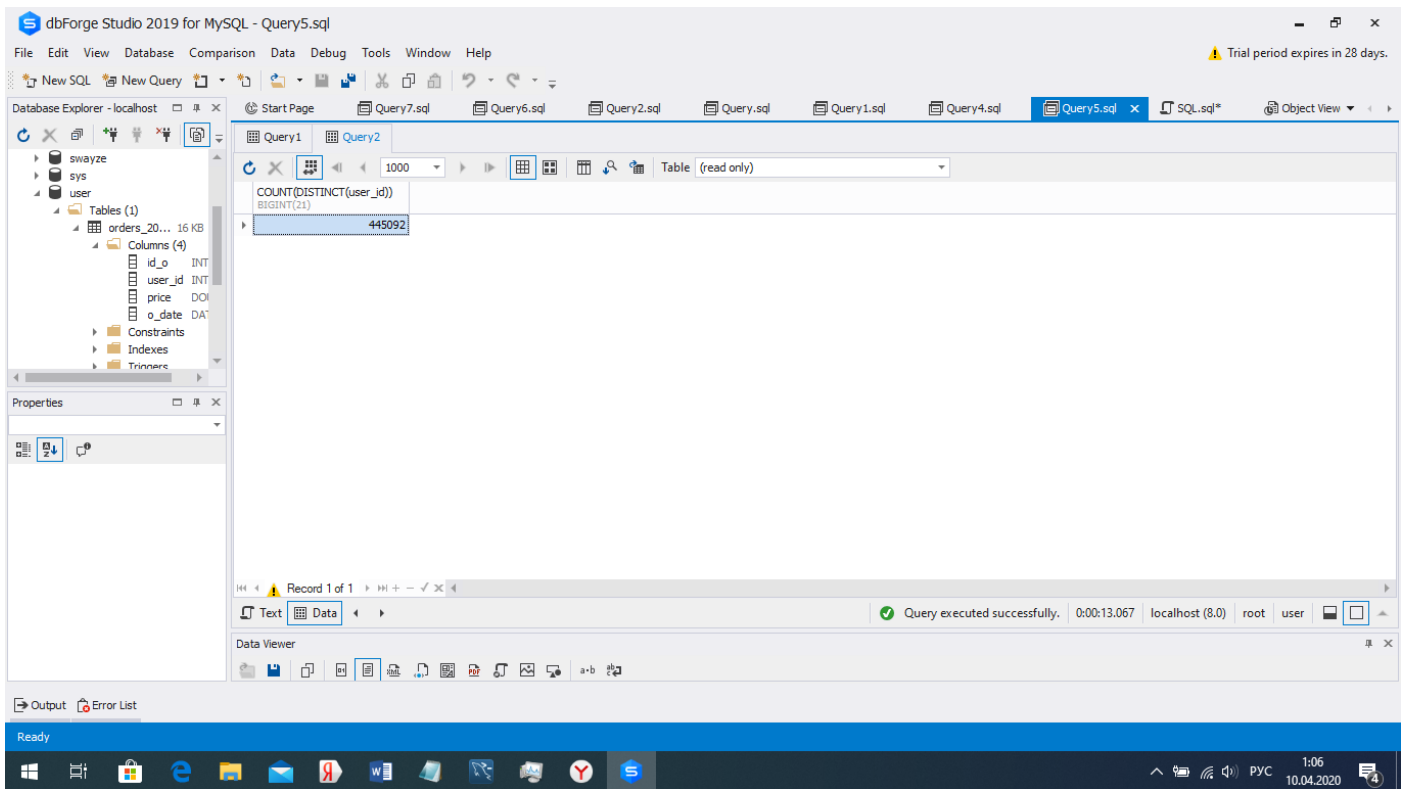
SELECT COUNT(DISTINCT(o.user_id))
FROM orders_20190822 o
WHERE year(o.o_date) = 2016 AND year(o.o_date) <> 2017 ;

```

```

SELECT
COUNT(DISTINCT(user_id))
FROM orders_20190822 o
WHERE year(o.o_date) = 2016 AND NOT EXISTS( SELECT DISTINCT(o1.user_id)
FROM orders_20190822 o1
WHERE o1.id_o = o.id_o AND year(o1.o_date) = 2017
);

```

6. Найти ID самого активного по кол-ву покупок пользователя.

```

SELECT DISTINCT(o.user_id), COUNT(*) AS ORDERS
FROM orders_20190822 o
GROUP BY o.user_id
ORDER BY ORDERS DESC LIMIT 1

```

dbForge Studio 2019 for MySQL - Query7.sql*

File Edit View Database Comparison Data Debug Tools Window Help

New SQL New Query

Database Explorer - localhost

Start Page Query7.sql* Query6.sql Query2.sql Query.sql Query1.sql Query4.sql Query5.sql SQL.sql* Object View

swayze sys user Tables (1) orders_20... 16 KB Columns (4) id_o INT user_id INT price DOI o_date DAT Constraints Indexes Triggers

Properties

Record 1 of 1

Query Builder Text Data

Query executed successfully. 0:00:05.268 localhost (8.0) root user

Data Viewer

Output Error List

Ready

Windows taskbar with icons for File Explorer, Mail, Word, etc.

System tray with date 10.04.2020 and time 1:08