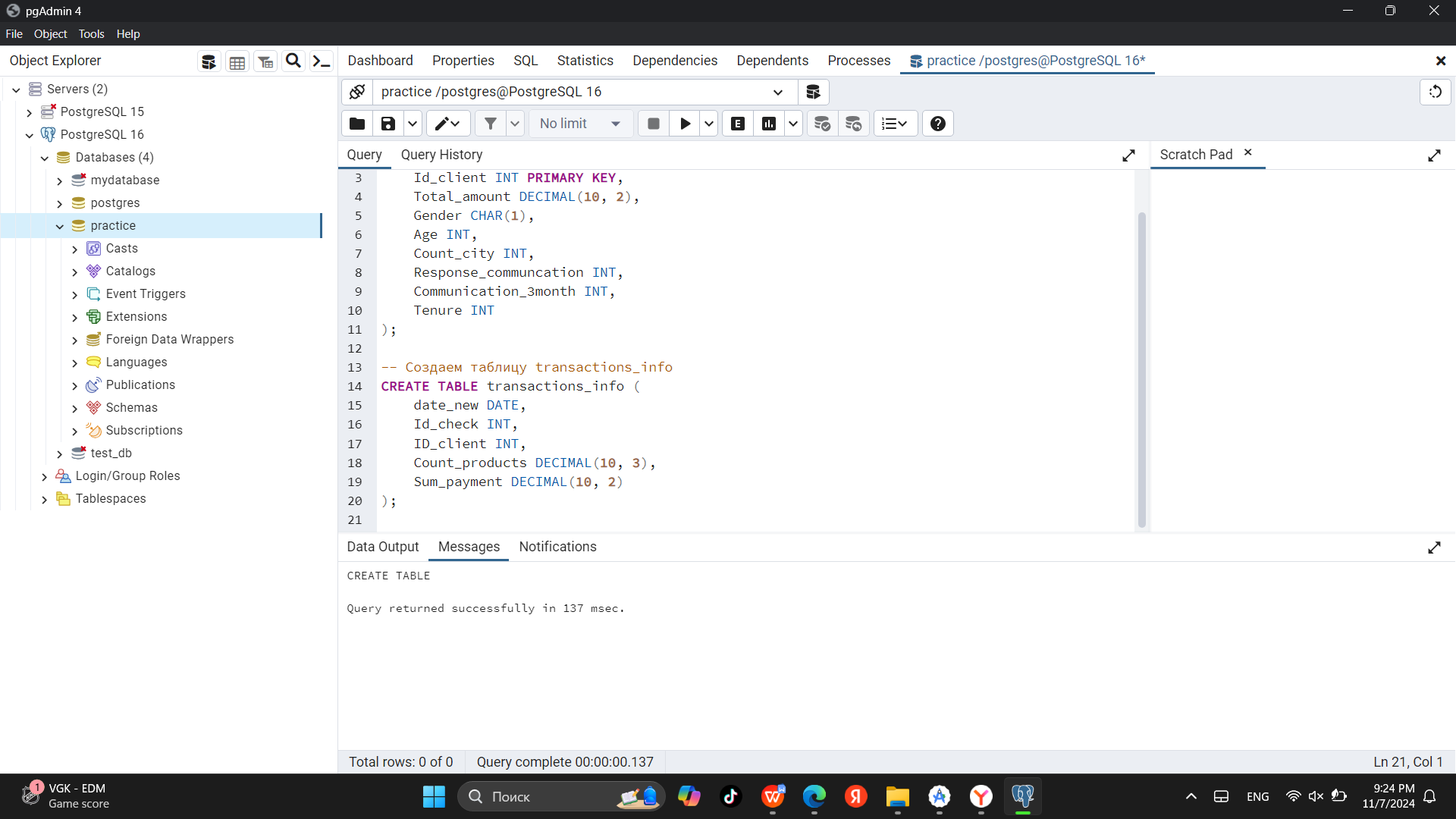
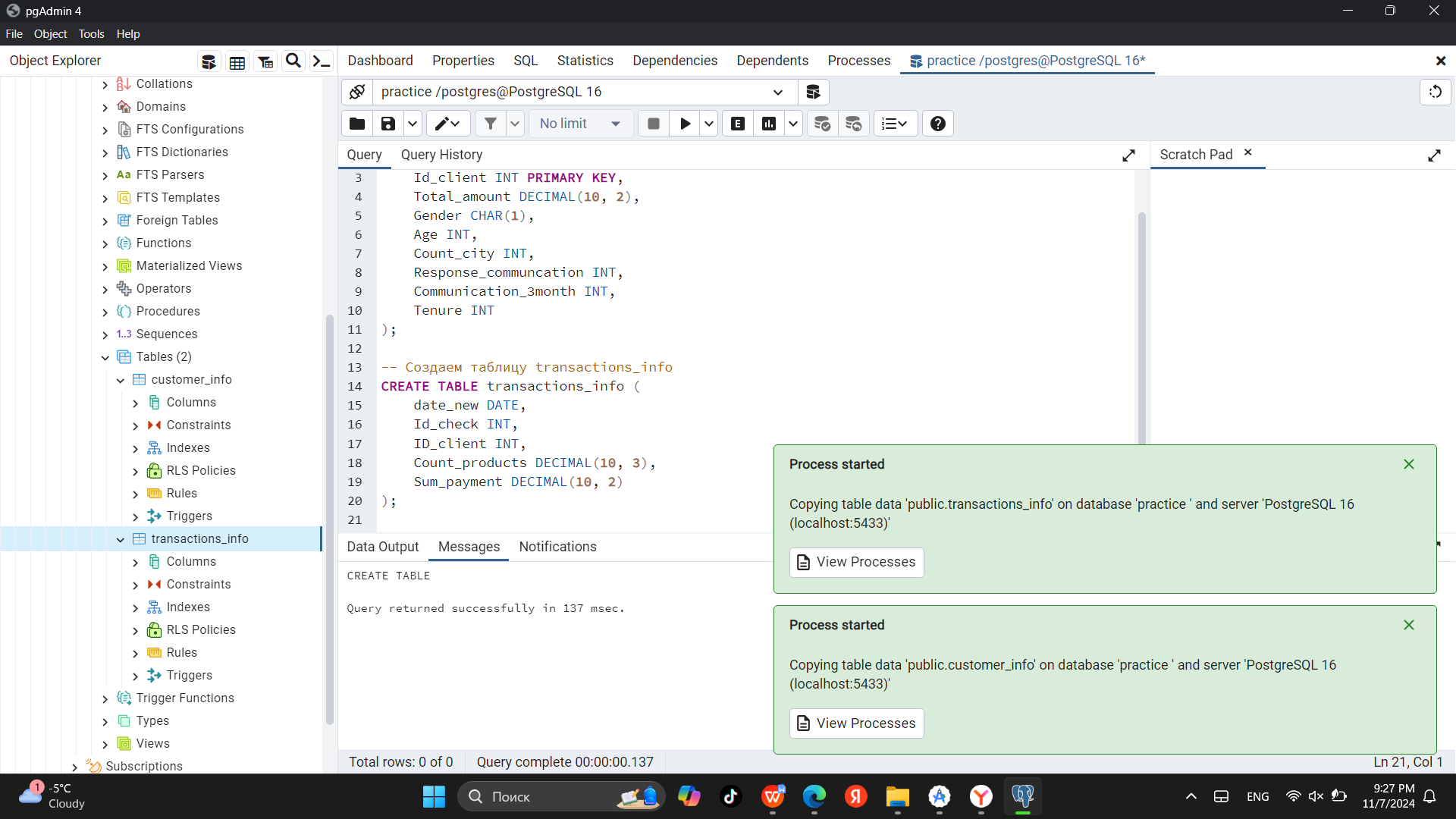
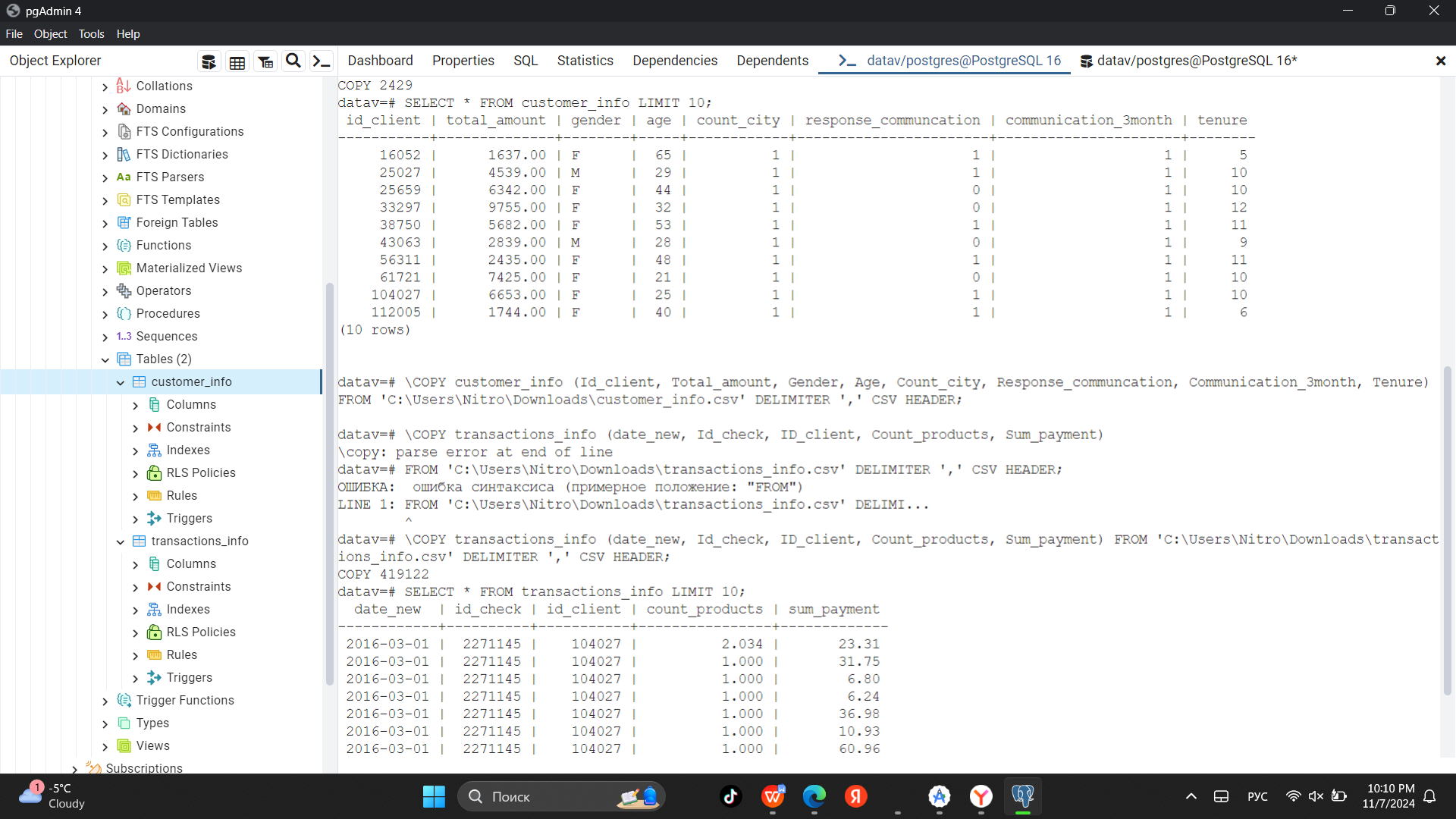
SQL



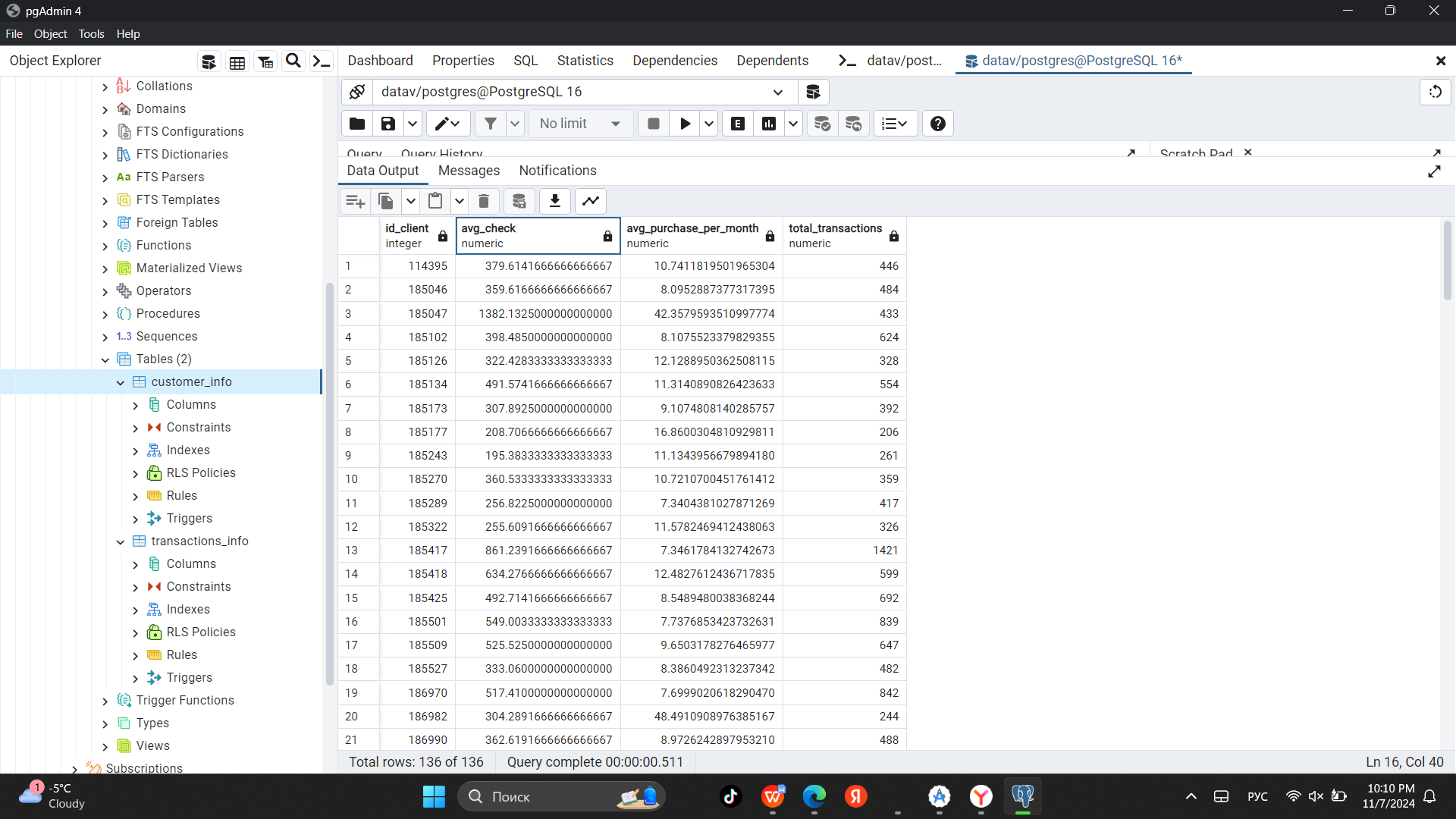


SELECT \* FROM transactions\_info LIMIT 10;

SELECT \* FROM customer\_info LIMIT 10;



1. список клиентов с непрерывной историей за год, то есть каждый месяц на регулярной основе без пропусков за указанный годовой период, средний чек за период с 01.06.2015 по 01.06.2016, средняя сумма покупок за месяц, количество всех операций по клиенту за период;



WITH MonthlyPurchases AS (

SELECT

ID\_client,

TO\_CHAR(date\_new, 'YYYY-MM') AS month, -- Преобразуем дату в год-месяц

SUM(Sum\_payment) AS total\_spent,

COUNT(\*) AS num\_transactions

FROM transactions\_info

WHERE date\_new BETWEEN '2015-06-01' AND '2016-06-01' -- Фильтруем по датам

GROUP BY ID\_client, TO\_CHAR(date\_new, 'YYYY-MM') -- Группируем по клиентам и месяцам

),

ClientsWithFullHistory AS (

SELECT

ID\_client

FROM MonthlyPurchases

GROUP BY ID\_client

HAVING COUNT(DISTINCT month) = 12

),

AggregatedData AS (

SELECT

c.ID\_client,

AVG(mp.total\_spent) AS avg\_check, -- Средний чек

AVG(mp.total\_spent / mp.num\_transactions) AS avg\_purchase\_per\_month, -- Средняя сумма покупок за месяц

SUM(mp.num\_transactions) AS total\_transactions -- Общее количество операций

FROM MonthlyPurchases mp

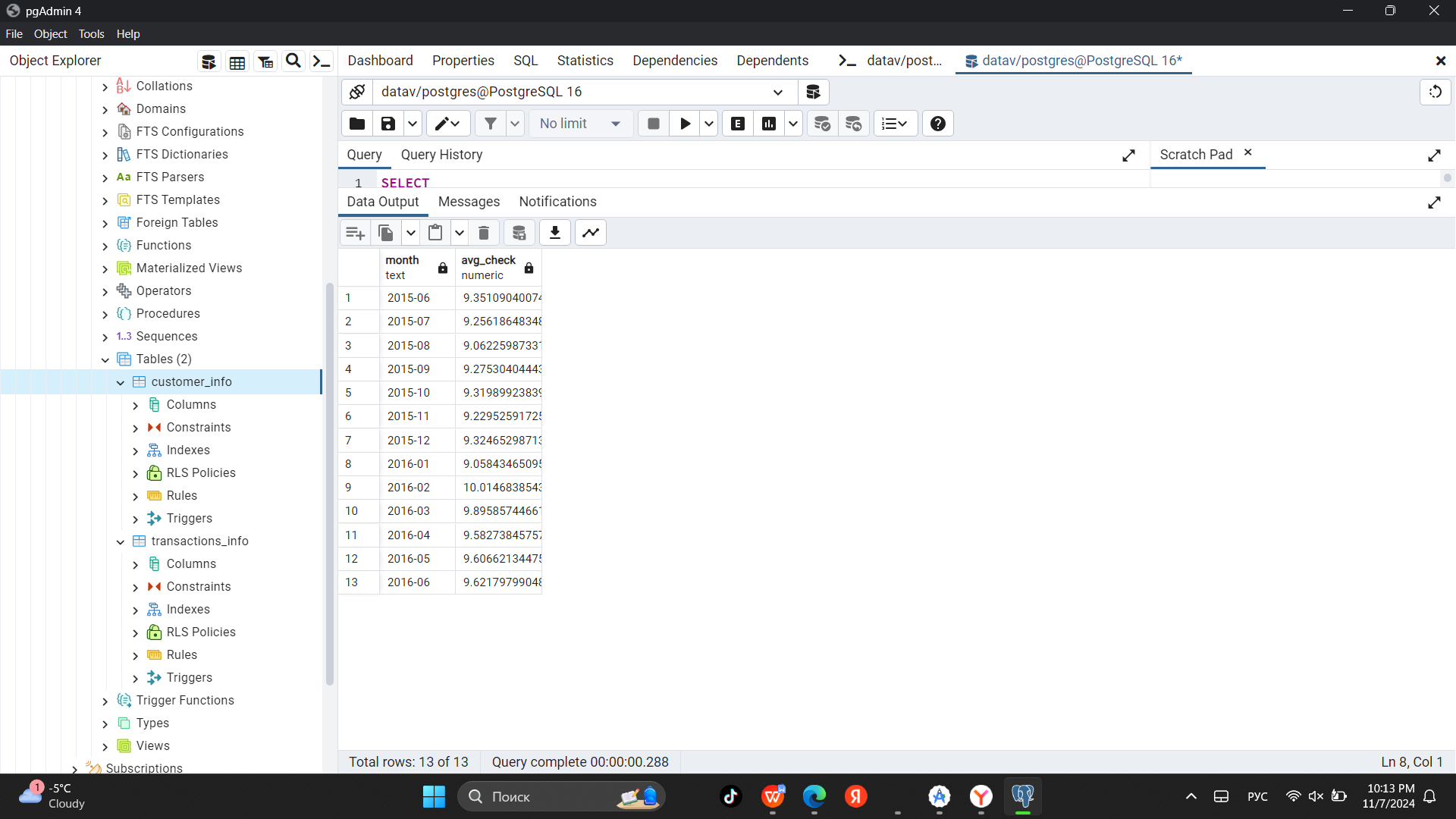
JOIN ClientsWithFullHistory c ON mp.ID\_client = c.ID\_client

GROUP BY c.ID\_client

)

SELECT \* FROM AggregatedData;

1. информацию в разрезе месяцев:
2. средняя сумма чека в месяц;



SELECT

TO\_CHAR(date\_new, 'YYYY-MM') AS month, -- Преобразуем дату в год-месяц

AVG(Sum\_payment) AS avg\_check -- Средняя сумма чека за месяц

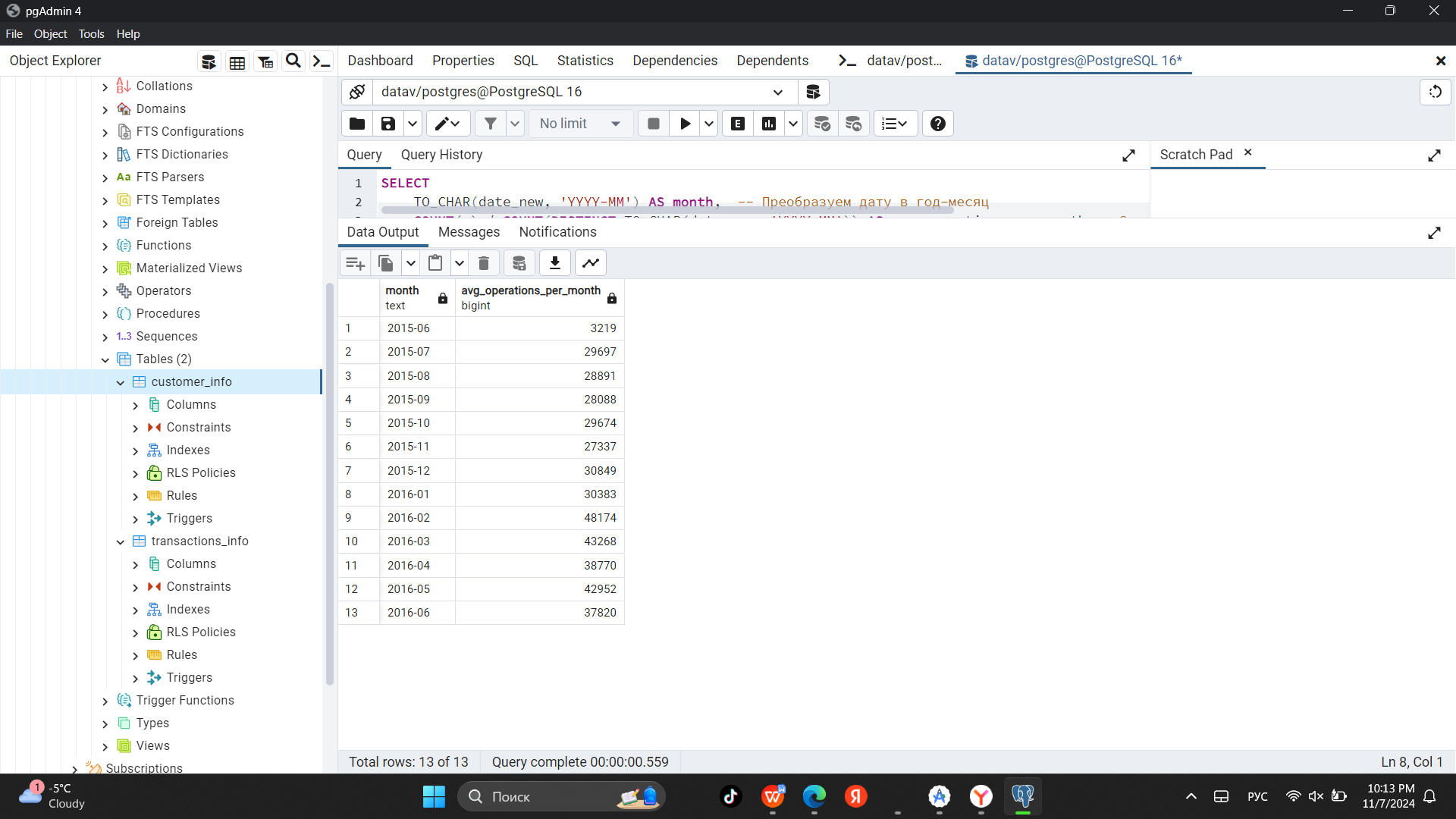
FROM transactions\_info

WHERE date\_new BETWEEN '2015-06-01' AND '2016-06-01' -- Фильтруем по датам

GROUP BY TO\_CHAR(date\_new, 'YYYY-MM') -- Группируем по месяцам

ORDER BY month; -- Сортировка по месяцам

1. среднее количество операций в месяц;



SELECT

TO\_CHAR(date\_new, 'YYYY-MM') AS month, -- Преобразуем дату в год-месяц

COUNT(\*) / COUNT(DISTINCT TO\_CHAR(date\_new, 'YYYY-MM')) AS avg\_operations\_per\_month -- Среднее количество операций в месяц

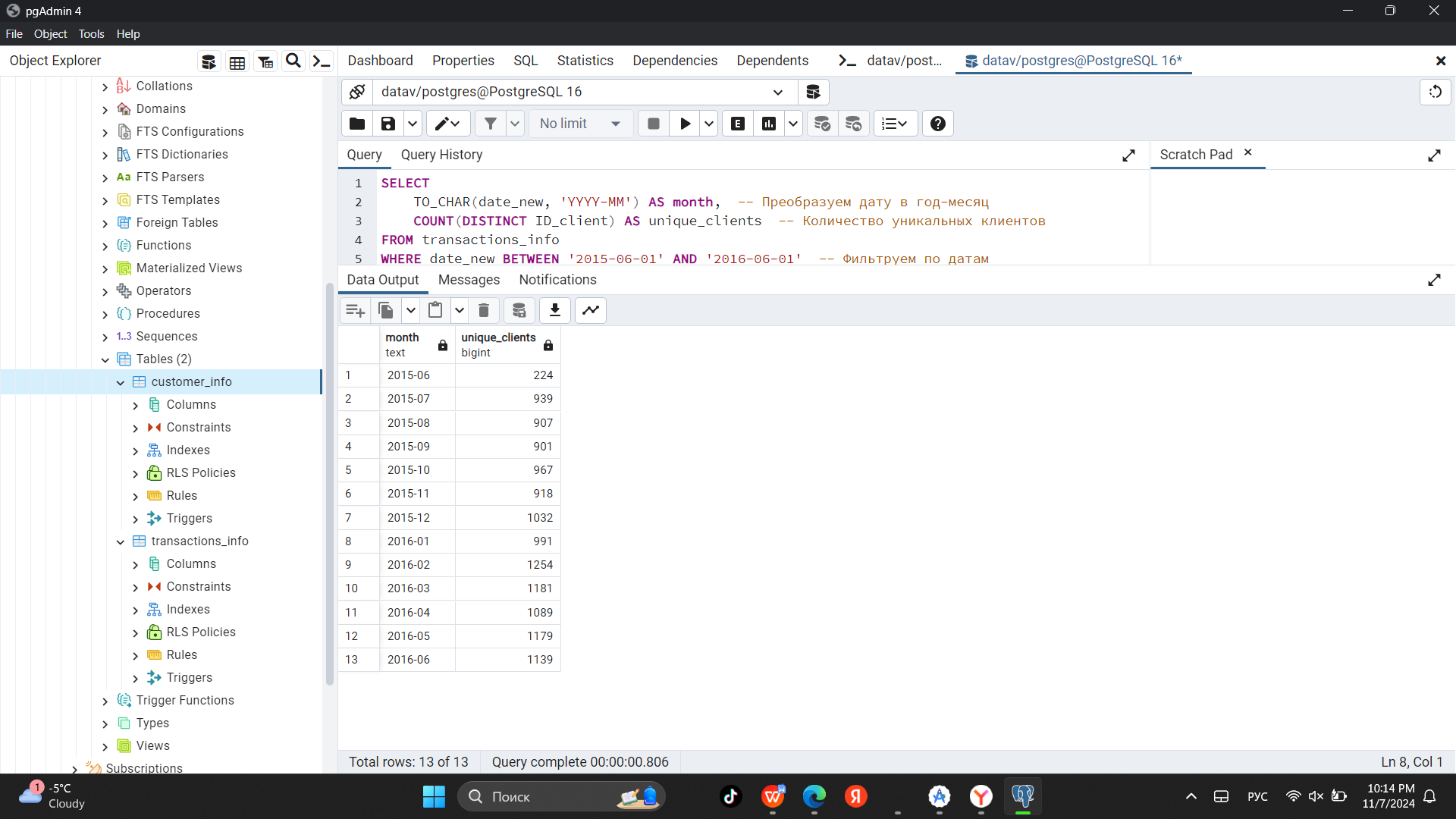
FROM transactions\_info

WHERE date\_new BETWEEN '2015-06-01' AND '2016-06-01' -- Фильтруем по датам

GROUP BY TO\_CHAR(date\_new, 'YYYY-MM') -- Группируем по месяцам

ORDER BY month; -- Сортировка по месяцам

1. среднее количество клиентов, которые совершали операции;



SELECT

TO\_CHAR(date\_new, 'YYYY-MM') AS month, -- Преобразуем дату в год-месяц

COUNT(DISTINCT ID\_client) AS unique\_clients -- Количество уникальных клиентов

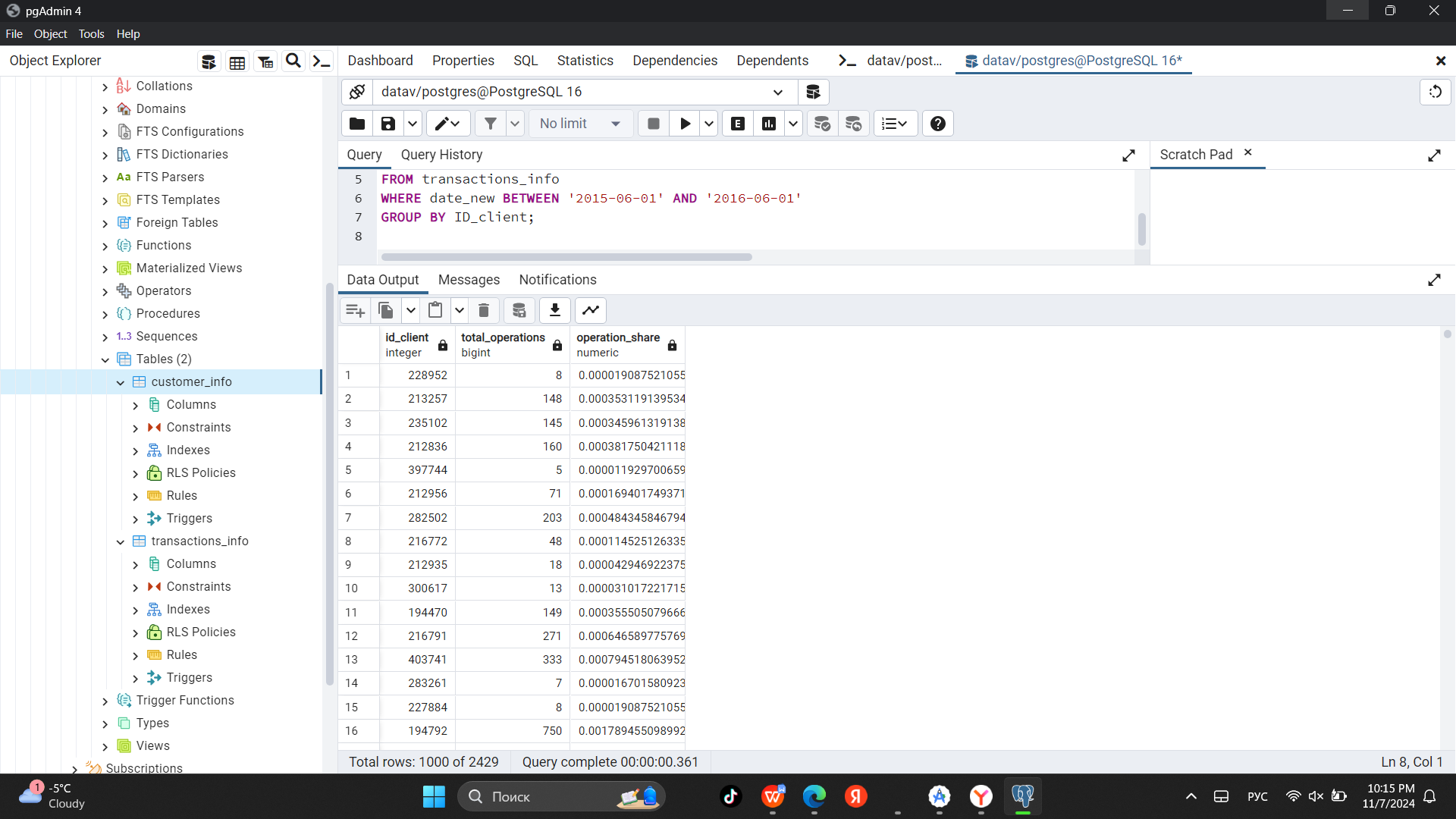
FROM transactions\_info

WHERE date\_new BETWEEN '2015-06-01' AND '2016-06-01' -- Фильтруем по датам

GROUP BY TO\_CHAR(date\_new, 'YYYY-MM') -- Группируем по месяцам

ORDER BY month;

1. долю от общего количества операций за год и долю в месяц от общей суммы операций;



SELECT

ID\_client,

COUNT(\*) AS total\_operations, -- Количество операций по клиенту

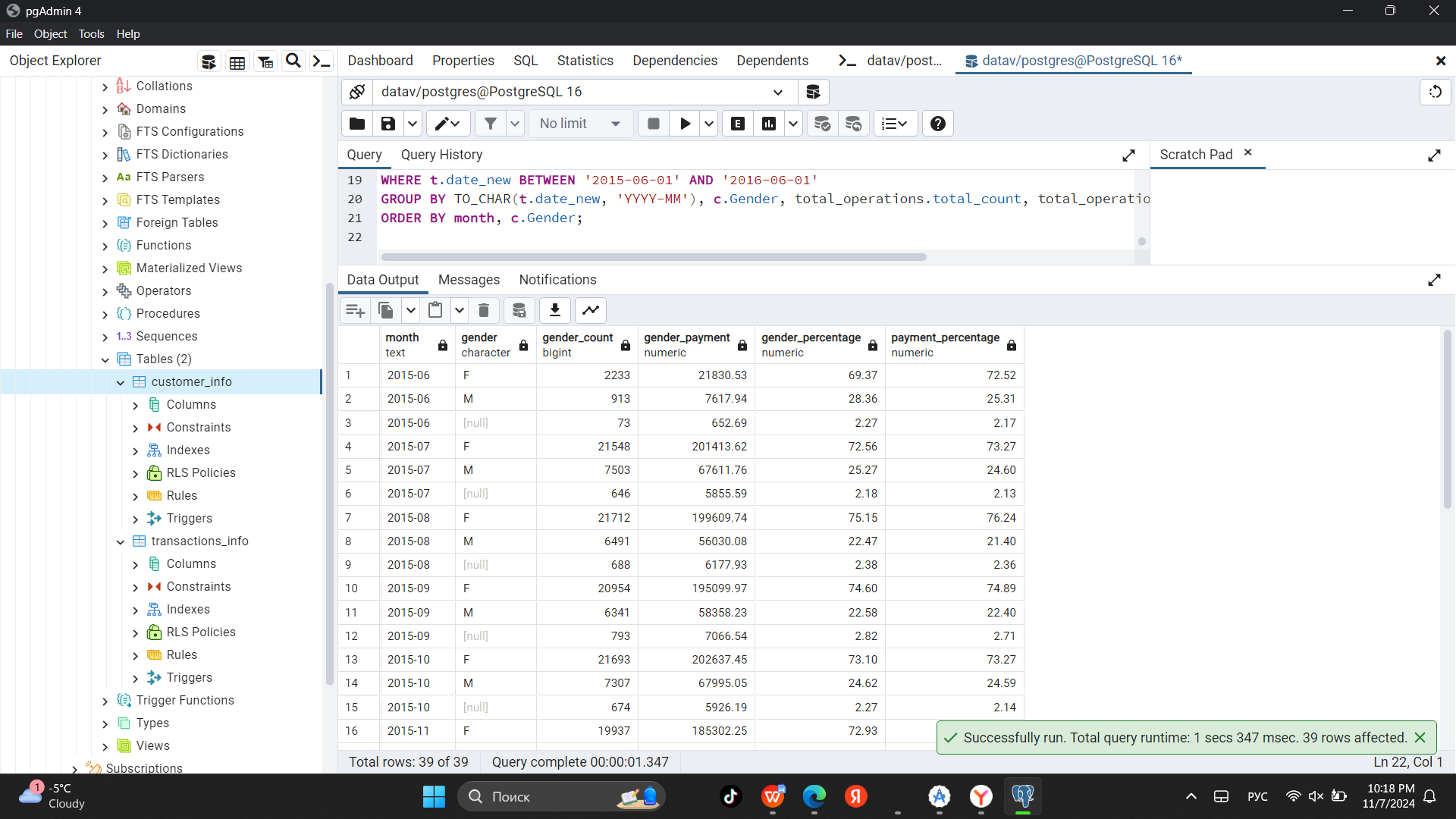
COUNT(\*) \* 1.0 / (SELECT COUNT(\*) FROM transactions\_info WHERE date\_new BETWEEN '2015-06-01' AND '2016-06-01') AS operation\_share -- Доля операций этого клиента от общего количества

FROM transactions\_info

WHERE date\_new BETWEEN '2015-06-01' AND '2016-06-01'

GROUP BY ID\_client;

1. вывести % соотношение M/F/NA в каждом месяце с их долей затрат;



SELECT

TO\_CHAR(t.date\_new, 'YYYY-MM') AS month, -- Преобразуем дату в год-месяц

c.Gender,

COUNT(\*) AS gender\_count, -- Количество клиентов по полу

SUM(t.Sum\_payment) AS gender\_payment, -- Сумма операций по полу

ROUND(COUNT(\*) \* 100.0 / total\_operations.total\_count, 2) AS gender\_percentage, -- Процентное соотношение по клиентам

ROUND(SUM(t.Sum\_payment) \* 100.0 / total\_operations.total\_payment, 2) AS payment\_percentage -- Процентная доля по затратам

FROM transactions\_info t

JOIN customer\_info c ON t.ID\_client = c.Id\_client -- Объединяем по ID клиента

JOIN (

SELECT

TO\_CHAR(date\_new, 'YYYY-MM') AS month,

COUNT(\*) AS total\_count, -- Общее количество операций в месяц

SUM(Sum\_payment) AS total\_payment -- Общая сумма операций в месяц

FROM transactions\_info

WHERE date\_new BETWEEN '2015-06-01' AND '2016-06-01'

GROUP BY TO\_CHAR(date\_new, 'YYYY-MM')

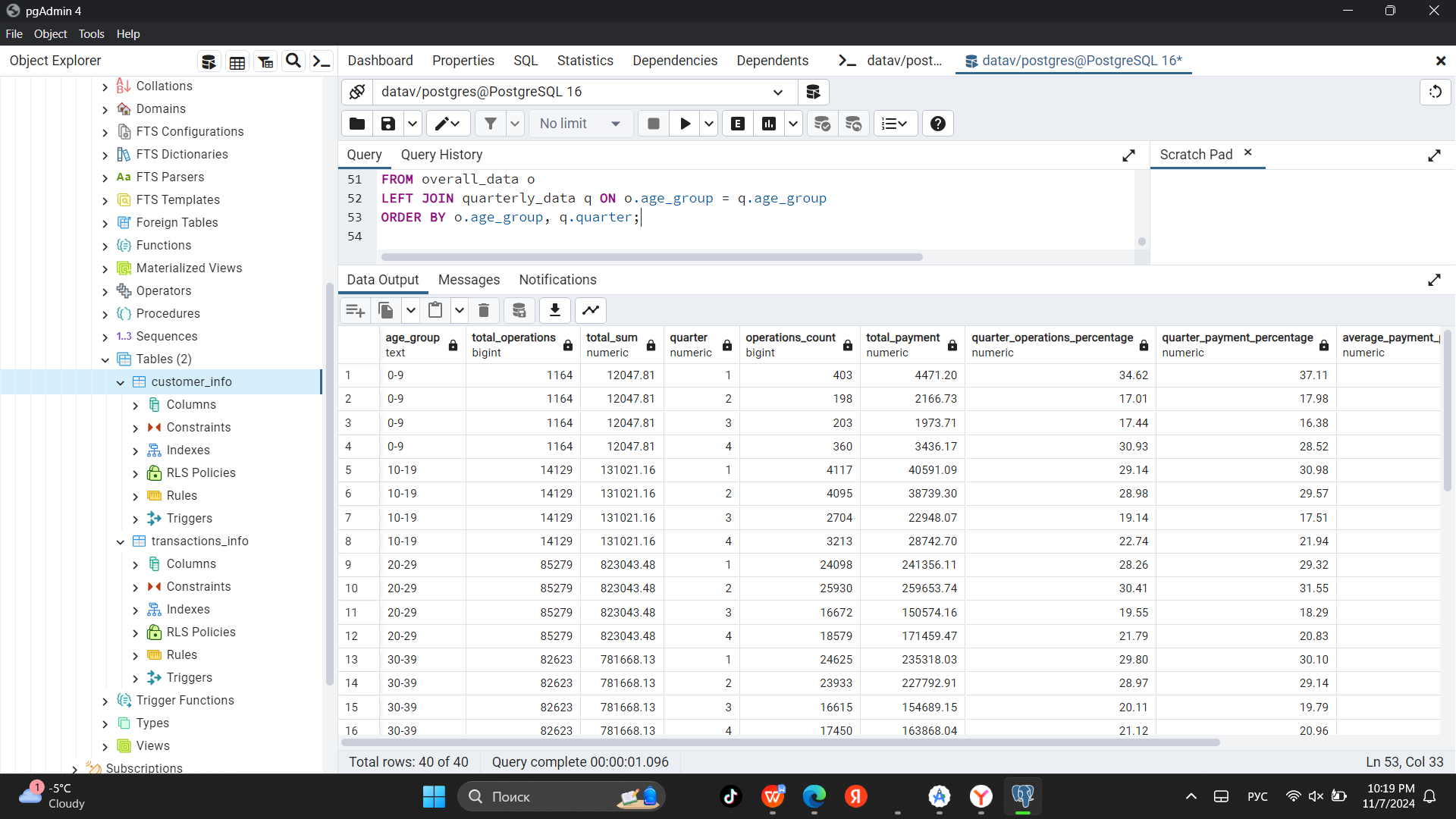
) total\_operations ON TO\_CHAR(t.date\_new, 'YYYY-MM') = total\_operations.month

WHERE t.date\_new BETWEEN '2015-06-01' AND '2016-06-01'

GROUP BY TO\_CHAR(t.date\_new, 'YYYY-MM'), c.Gender, total\_operations.total\_count, total\_operations.total\_payment

ORDER BY month, c.Gender;

1. возрастные группы клиентов с шагом 10 лет и отдельно клиентов, у которых нет данной информации, с параметрами сумма и количество операций за весь период, и поквартально - средние показатели и %.





WITH age\_groups AS (

SELECT

CASE

WHEN c.Age IS NULL THEN 'Unknown'

WHEN c.Age BETWEEN 0 AND 9 THEN '0-9'

WHEN c.Age BETWEEN 10 AND 19 THEN '10-19'

WHEN c.Age BETWEEN 20 AND 29 THEN '20-29'

WHEN c.Age BETWEEN 30 AND 39 THEN '30-39'

WHEN c.Age BETWEEN 40 AND 49 THEN '40-49'

WHEN c.Age BETWEEN 50 AND 59 THEN '50-59'

WHEN c.Age BETWEEN 60 AND 69 THEN '60-69'

WHEN c.Age BETWEEN 70 AND 79 THEN '70-79'

WHEN c.Age BETWEEN 80 AND 89 THEN '80-89'

ELSE '90+'

END AS age\_group, -- Группируем по возрастным категориям

c.Id\_client,

t.Sum\_payment,

t.date\_new

FROM customer\_info c

JOIN transactions\_info t ON c.Id\_client = t.ID\_client

WHERE t.date\_new BETWEEN '2015-06-01' AND '2016-06-01'

),

quarterly\_data AS (

SELECT

age\_group,

EXTRACT(QUARTER FROM date\_new) AS quarter, -- Определяем квартал из даты

COUNT(\*) AS operations\_count, -- Количество операций в квартале

SUM(Sum\_payment) AS total\_payment -- Общая сумма операций в квартале

FROM age\_groups

GROUP BY age\_group, EXTRACT(QUARTER FROM date\_new)

),

overall\_data AS (

SELECT

age\_group,

COUNT(\*) AS total\_operations, -- Общее количество операций по возрастной группе

SUM(Sum\_payment) AS total\_sum -- Общая сумма операций по возрастной группе

FROM age\_groups

GROUP BY age\_group

)

SELECT

o.age\_group,

o.total\_operations,

o.total\_sum,

q.quarter,

q.operations\_count,

q.total\_payment,

ROUND(q.operations\_count \* 100.0 / o.total\_operations, 2) AS quarter\_operations\_percentage, -- Процент операций по кварталу

ROUND(q.total\_payment \* 100.0 / o.total\_sum, 2) AS quarter\_payment\_percentage, -- Процент суммы по кварталу

ROUND(q.total\_payment / NULLIF(q.operations\_count, 0), 2) AS average\_payment\_per\_operation, -- Средний чек в квартале

ROUND(q.total\_payment / NULLIF(o.total\_operations, 0), 2) AS average\_payment\_per\_client -- Средний чек на клиента

FROM overall\_data o

LEFT JOIN quarterly\_data q ON o.age\_group = q.age\_group

ORDER BY o.age\_group, q.quarter;

**Блок Power BI**

* Задание 1

Нужно на основе приложенных таблиц сделать отчет в Power BI и отправить файл *.pbix*. Ссылка на таблицы:

<https://drive.google.com/drive/u/1/folders/1NKr6UriAvwIedyKspHhNdyHyAmfmYEJY>

Таблицы:

- warehouses

- products

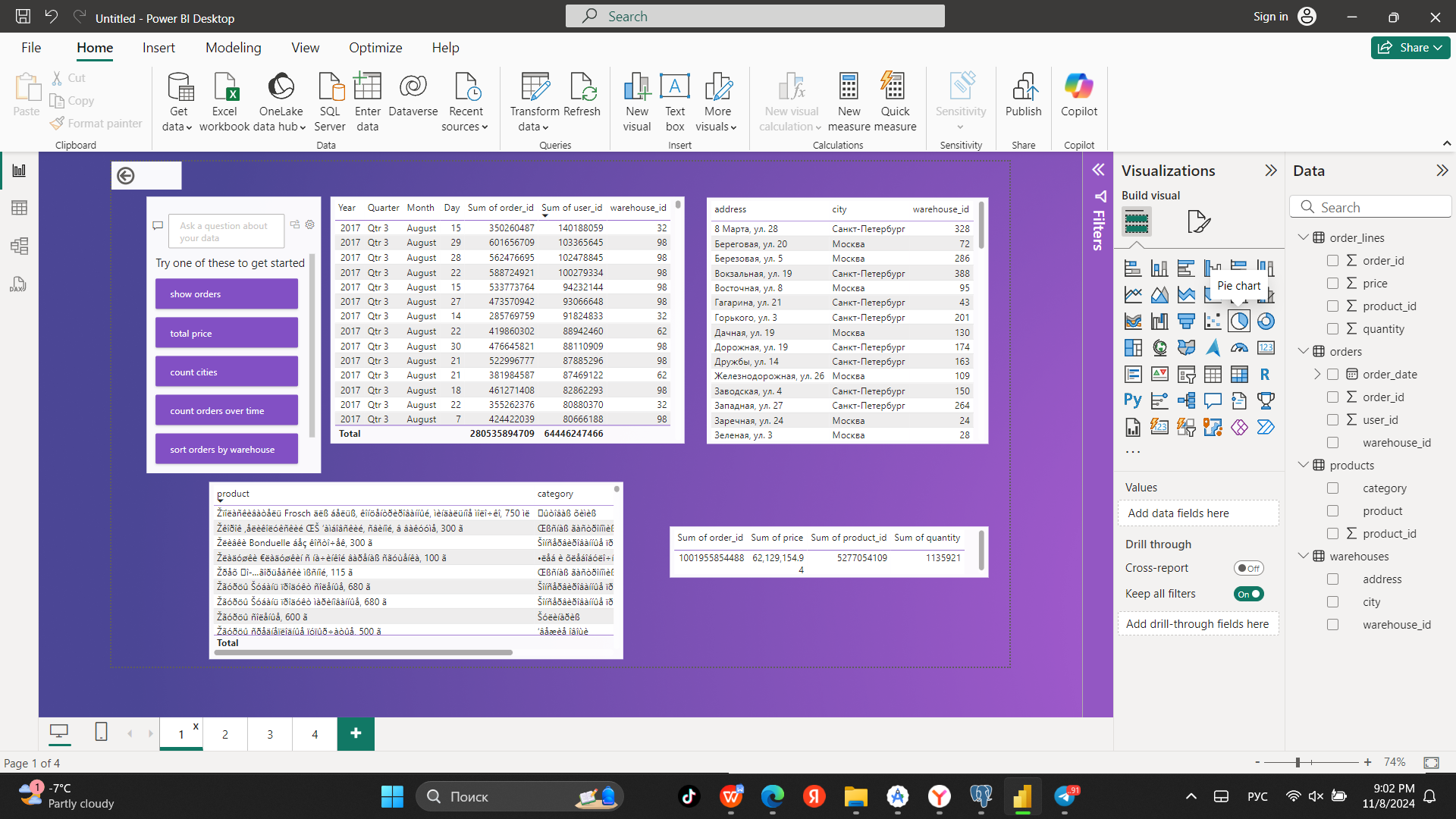
- orders

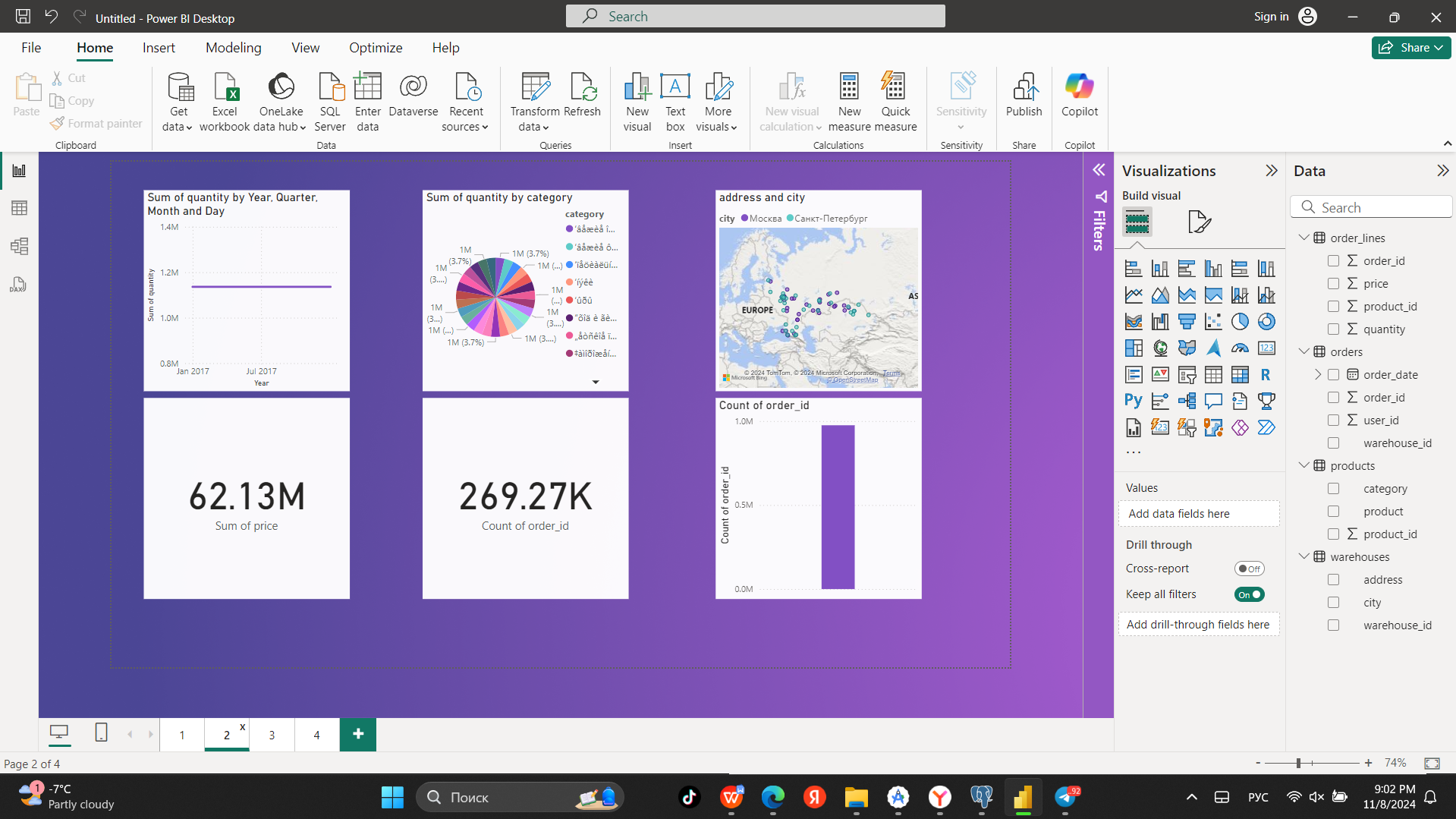
- order\_lines

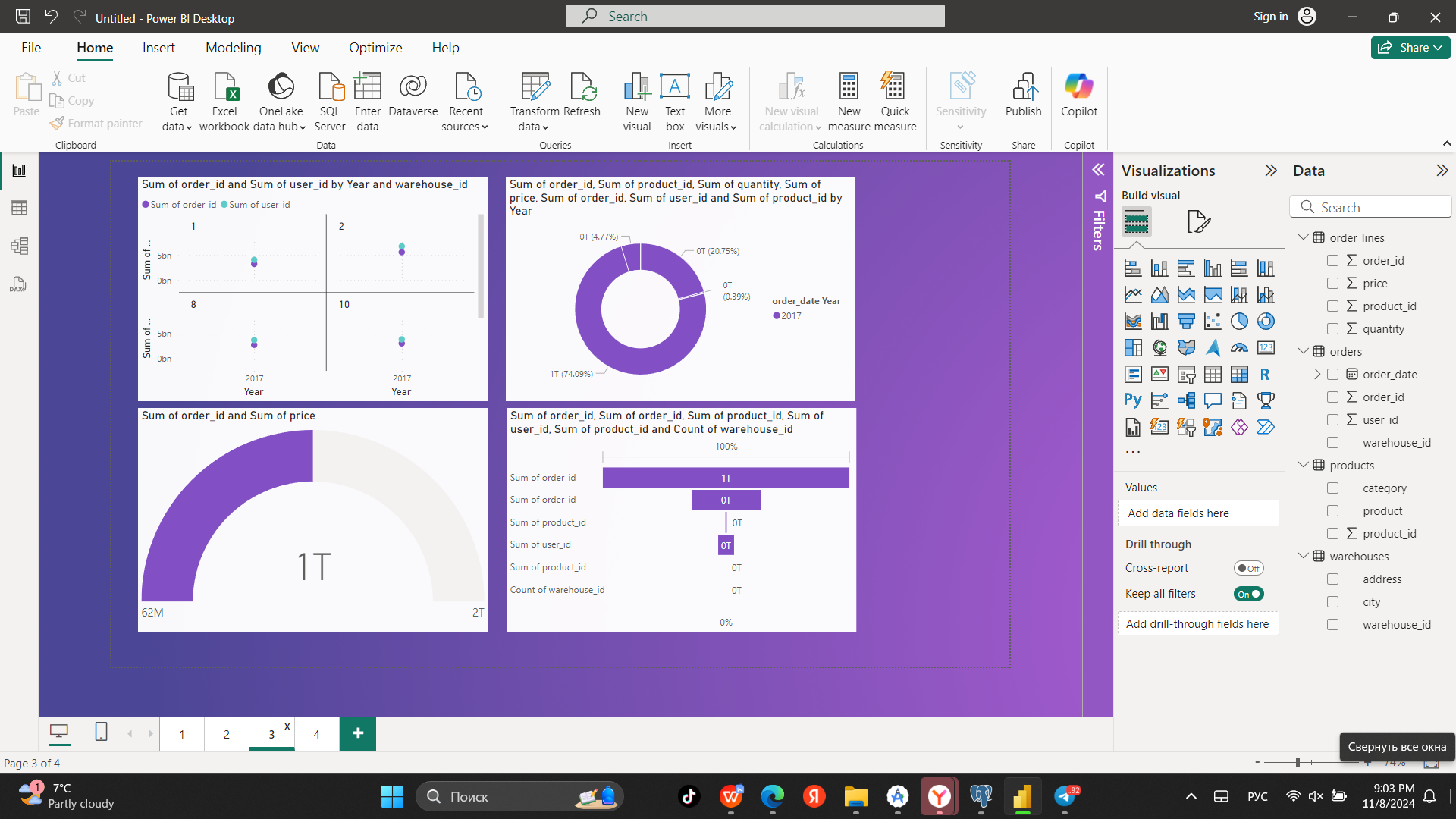
\*В таблицах рандомные данные, но стоит воспринимать их как фактические.

Всю обработку нужно делать внутри Power BI - исходные файлы оставьте как есть.

**Требования к отчету - представьте, что это все данные, которые вообще есть и вы делаете первый в компании отчет. Ваша задача выжать из этих данных максимум пользы. Какие показатели посчитать, какие визуализации выбрать и как организовать отчет - ваш выбор.**

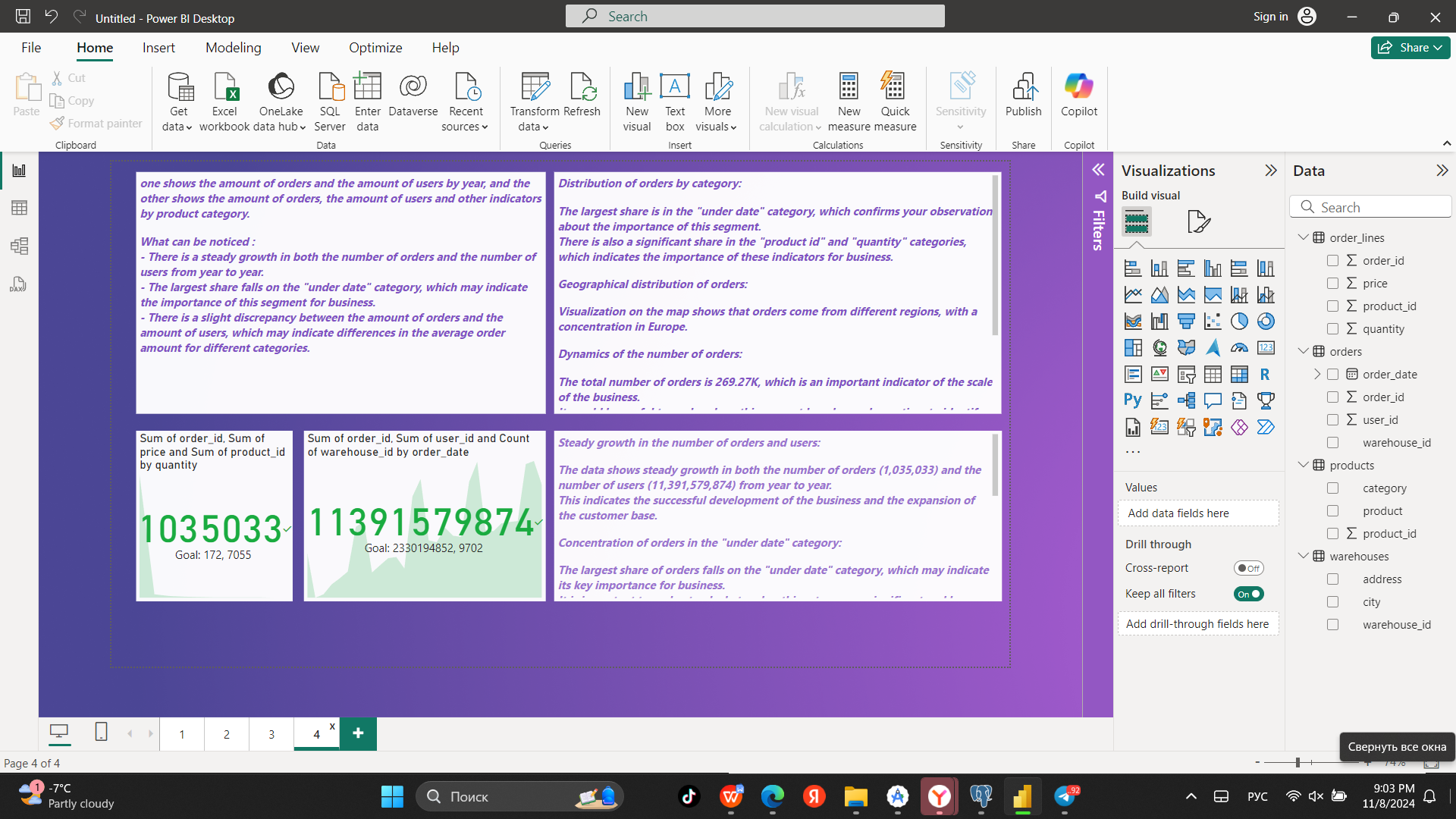






**Бонус:**

**Если вы найдете что-то интересное (аномалии, тренды, зависимости, etc.) - оформите свои мысли в виде комментария на отдельном листе отчета.**



One shows the amount of orders and the amount of users by year, and the other shows the amount of orders, the amount of users and other indicators by product category.

What can be noticed :

* - There is a steady growth in both the number of orders and the number of users from year to year.
* - The largest share falls on the "under date" category, which may indicate the importance of this segment for business.
* - There is a slight discrepancy between the amount of orders and the amount of users, which may indicate differences in the average order amount for different categories.

Distribution of orders by category:

* The largest share is in the "under date" category, which confirms your observation about the importance of this segment.
* There is also a significant share in the "product id" and "quantity" categories, which indicates the importance of these indicators for business.

Geographical distribution of orders:

* Visualization on the map shows that orders come from different regions, with a concentration in Europe.

Dynamics of the number of orders:

* The total number of orders is 269.27K, which is an important indicator of the scale of the business.
* It would be useful to analyze how this amount has changed over time to identify trends and seasonality.

Thus, the main directions for further analysis may include:

* Detailed study of the sales structure by category
* Analysis of geographical distribution and key markets
* Investigation of the dynamics of the number of orders over time

Steady growth in the number of orders and users:

* The data shows steady growth in both the number of orders (1,035,033) and the number of users (11,391,579,874) from year to year.
* This indicates the successful development of the business and the expansion of the customer base.

Concentration of orders in the "under date" category:

* The largest share of orders falls on the "under date" category, which may indicate its key importance for business.
* It is important to understand what makes this category so significant and how this insight can be used for further development.

There is a slight discrepancy between the number of orders and the number of users:

* The visualization shows a slight discrepancy between the total number of orders and the total number of users.
* This may indicate differences in the average order amount for different product categories or customer segments.
* Further analysis of the average checks by category could shed light on this pattern.

Thus, the key areas for further research may be:

* Detailed analysis of the sales structure by category, especially the "under date" category
* Studying the dynamics of average checks and their distribution by category
* Identification of key drivers of growth in the number of orders and users
* Задание 2

Напишите SQL запрос по базе данных из задания 1, который выведет список тех пользователей, которые купили за период 1-15 августа 2 любых корма для животных, кроме "Корм Kitekat для кошек, с кроликом в соусе, 85 г". Приложите его в текстовом документе.

SELECT DISTINCT o.user\_id

FROM orders o

JOIN order\_lines ol ON o.order\_id = ol.order\_id

JOIN products p ON ol.product\_id = p.product\_id

WHERE o.order\_date BETWEEN '2017-08-01' AND '2017-08-15'

AND p.product\_name != 'Корм Kitekat для кошек, с кроликом в соусе, 85 г'

GROUP BY o.user\_id

HAVING COUNT(DISTINCT p.product\_id) >= 2;

user\_id

52410

10155

17185

60455

91612

18517

7598

950934

* Задание 3

Напишите SQL запрос, который выведет список топ 5 самых часто встречающихся товаров в заказах пользователей в СПб за период 15-30 августа. Приложите его в том же текстовом документе, где вы написали запрос из предыдущего пункта.

SELECT p.product, COUNT(ol.product\_id) AS product\_count

FROM orders o

JOIN order\_lines ol ON o.order\_id = ol.order\_id

JOIN products p ON ol.product\_id = p.product\_id

JOIN warehouses w ON o.warehouse\_id = w.warehouse\_id

WHERE o.order\_date BETWEEN '2017-08-15' AND '2017-08-30'

AND w.city = 'Санкт-Петербург'

GROUP BY p.product

ORDER BY product\_count DESC

LIMIT 5;

product product\_count

Корм для собак "Chappi" 150

Корм для кошек "Whiskas" 120

Игрушка для кошек "Moulin Rouge" 95

Корм для собак "Pedigree" 80

Корм для кошек "Кошачий рацион" 70