**PHARMA DATA ANALYSIS USING SQL**

Data analysis is a process of finding answers to questions or solving problems using data.

A pharmaceutical dataset of ‘Poland’ and ‘Germany’ Where we will be looking at these 2 cities collectively.

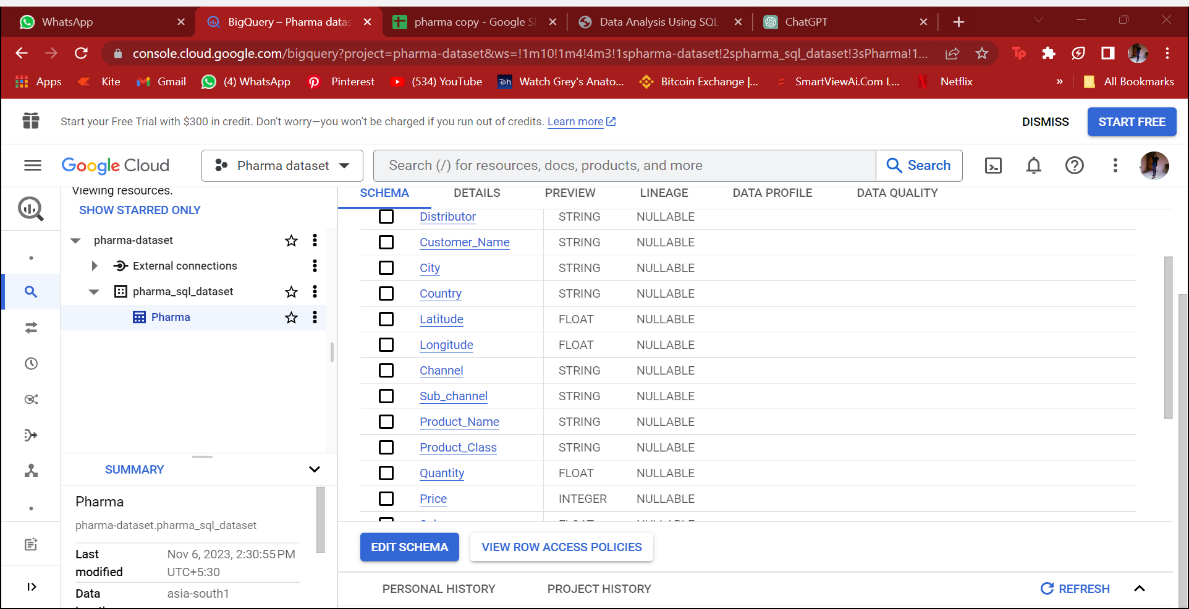
**Problem Statement:**

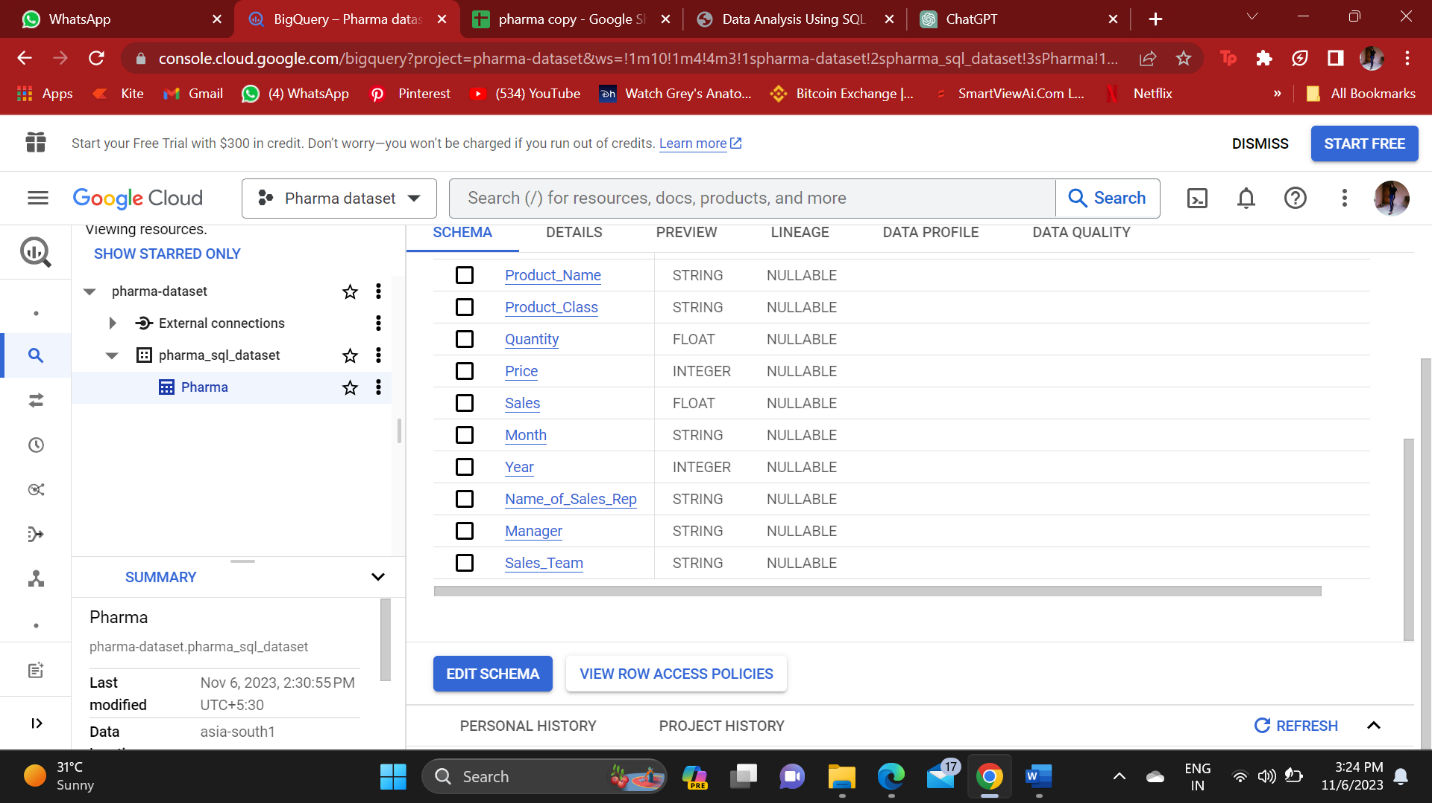
*"The pharmaceutical industry in Germany plays a pivotal role in the country's healthcare system, contributing significantly to its economy and public health. In this dynamic and highly regulated environment, it is crucial to analyze pharmaceutical datasets comprehensively. This study focuses on conducting an in-depth analysis of pharmaceutical data in Germany. The analysis aims to explore various facets of the industry, including drug utilization patterns, market trends, regulatory compliance, and the impact of healthcare policies on pharmaceutical companies, healthcare providers, and patients.*

Starting with the basic collection of data.

**1. Retrieve all columns for all records in the dataset.**

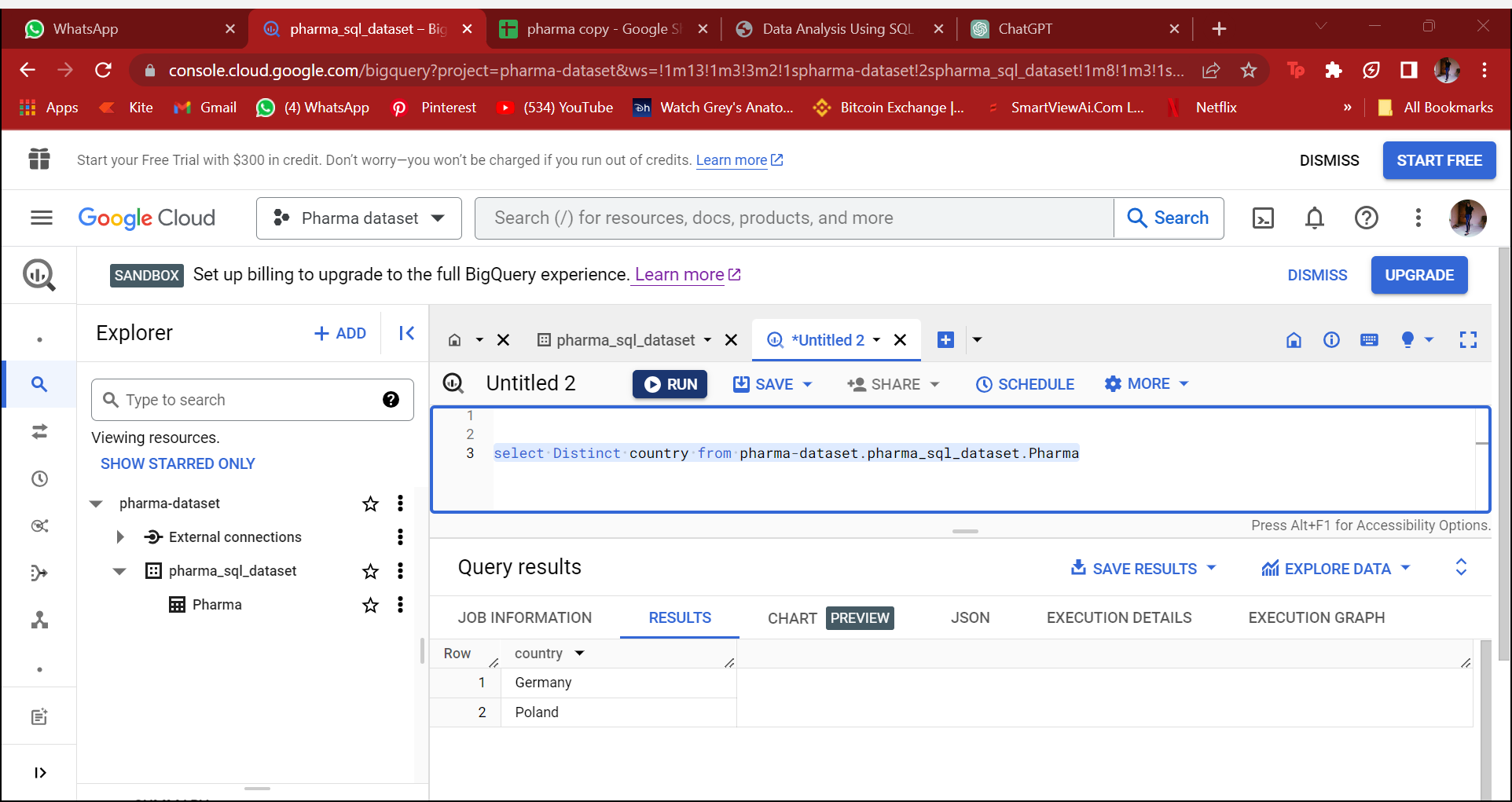
- **Select \* from** pharma-dataset.pharma\_sql\_dataset.Pharma





**2. How many unique countries are represented in the dataset?**

- select Distinct country from pharma-dataset.pharma\_sql\_dataset.Pharma



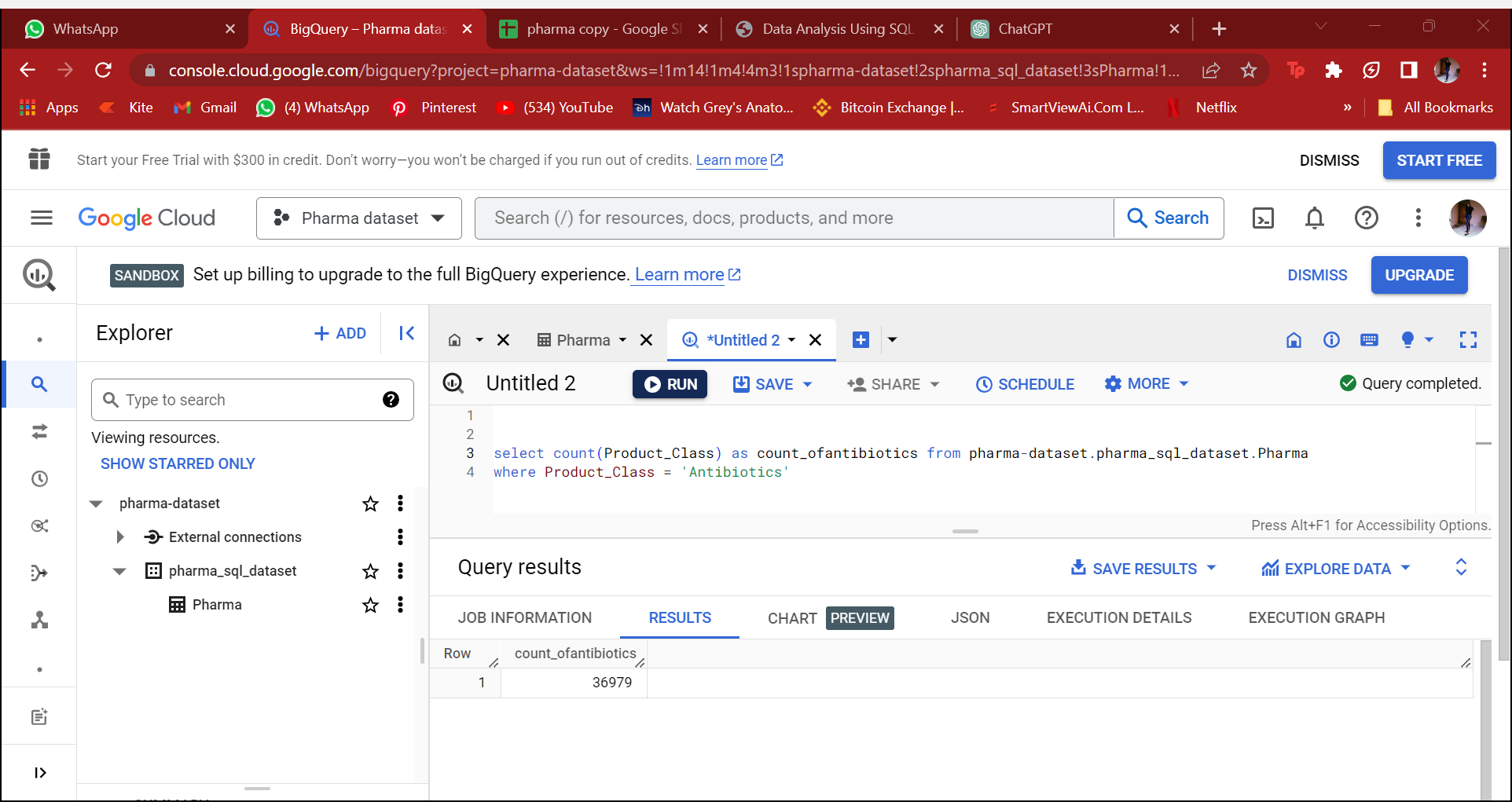
Analysis: There are 2 countries present in this dataset, but we cannot compare any of the data with Poland, as the records are only of year 2018.

**3. Find the total quantity sold for the ' Antibiotics' product class.**

- select count(Product\_Class)as count\_ofantibiotics

from pharma-dataset.Pharma.pharma\_table

where Product\_Class = 'Antibiotics'

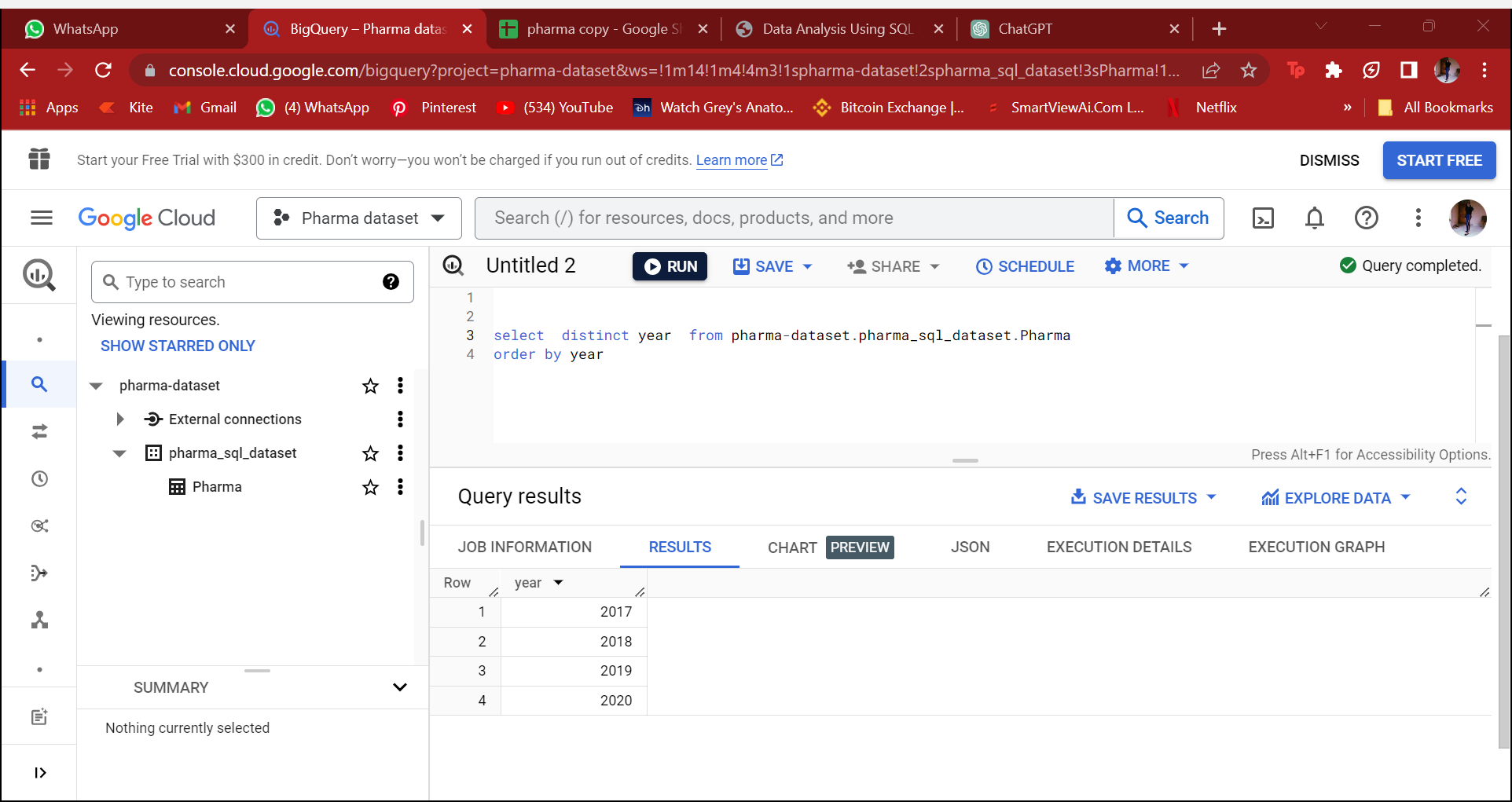


Analysis: 36,979 is the quantity that is sold specifically for Antibiotics.

**4. List all the distinct year and months present in the dataset**

* select year from pharma-dataset.pharma\_sql\_dataset.Pharma

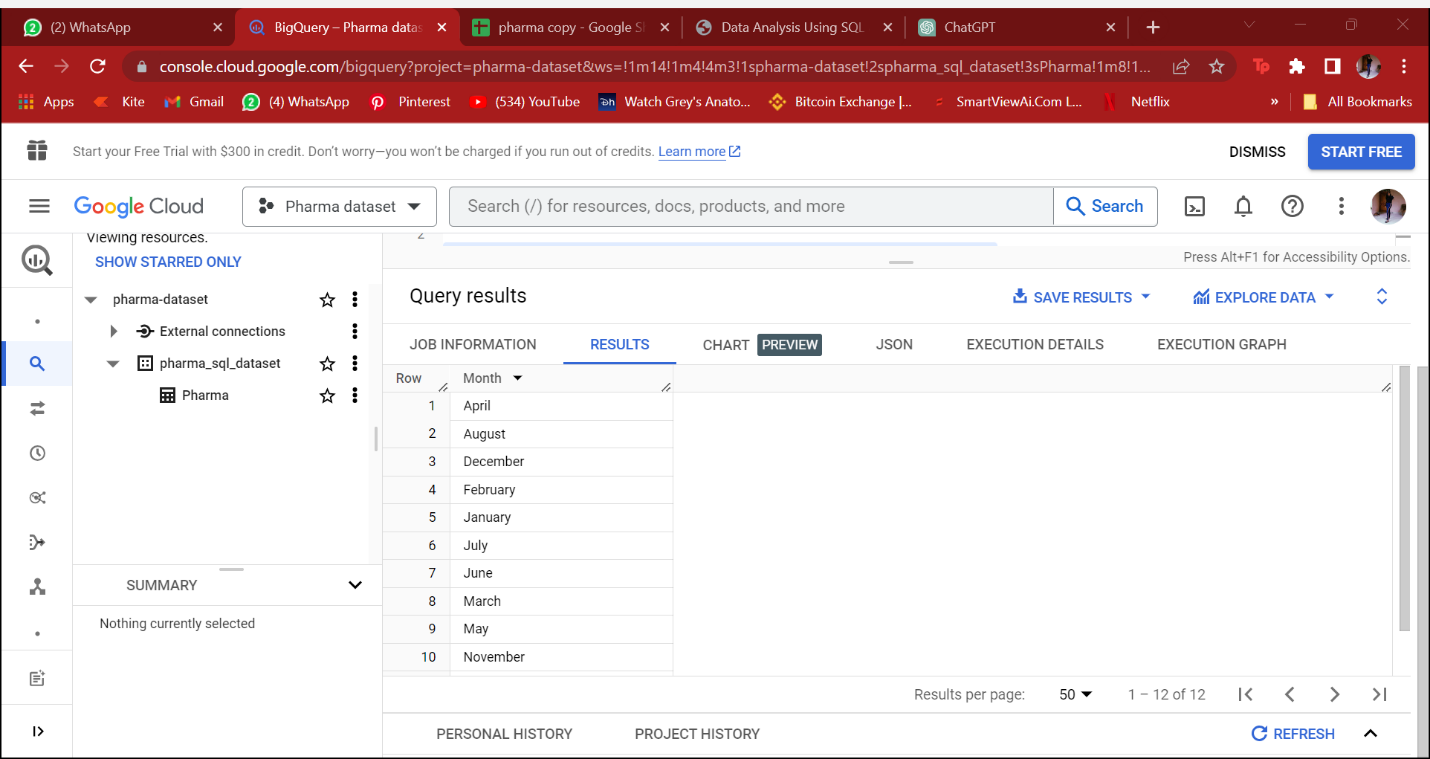
order by year asc



Analysis: This dataset has records of 4 years and those are : 2017, 2018,2019,2020

-select  distinct Month from pharma-dataset.Pharma.pharma\_table

order by month asc

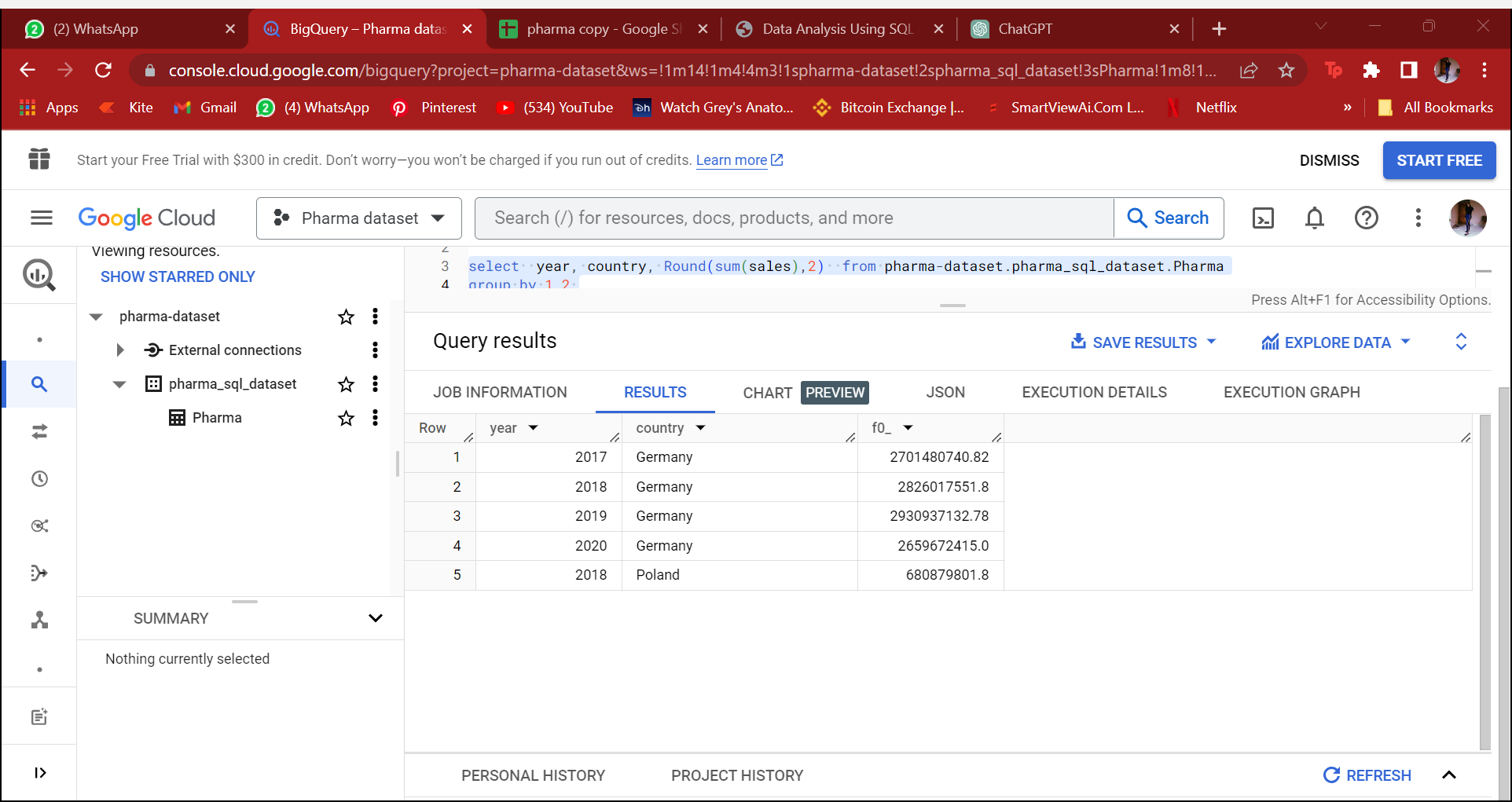


Some numerical data

**5. Calculate the total sales for each year.**

- select  year, country, Round(sum(sales),2)  from pharma-dataset.Pharma.pharma\_table

group by 1,2



Analysis:

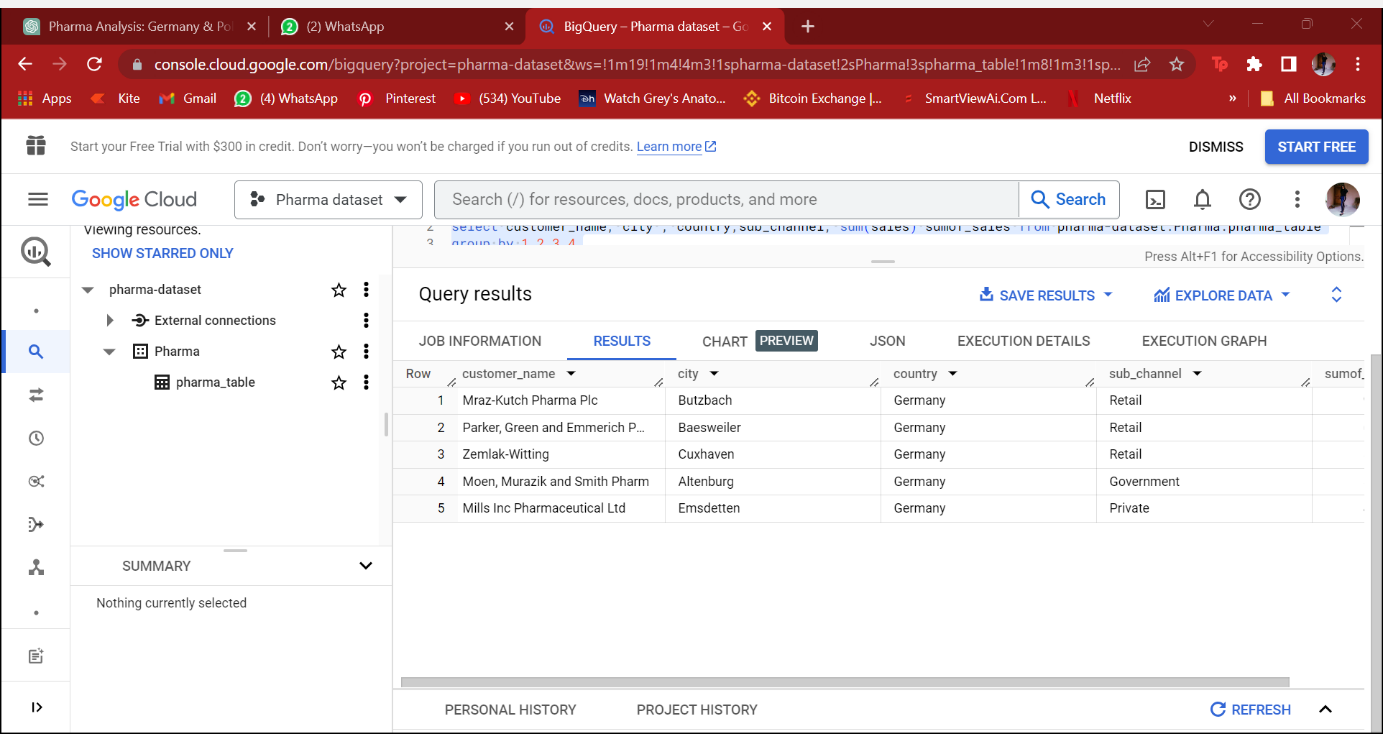
**6. Find the top 5 customer with the highest sales value.**

- select customer\_name, city , country,sub\_channel, sum(sales) sumof\_sales from pharma-dataset.Pharma.pharma\_table

group by 1,2,3,4

order by sumof\_sales desc

limit 5



Analysis: Above are the top 5 customers who have the highest sales value and each of them belongs to Germany, And if we look at the sub-channel top 3 sellers belongs to Retail market.

7. **Which regions areas in Germany have the lowest and highest market penetration , and what strategies can be implemented to expand market reach?**

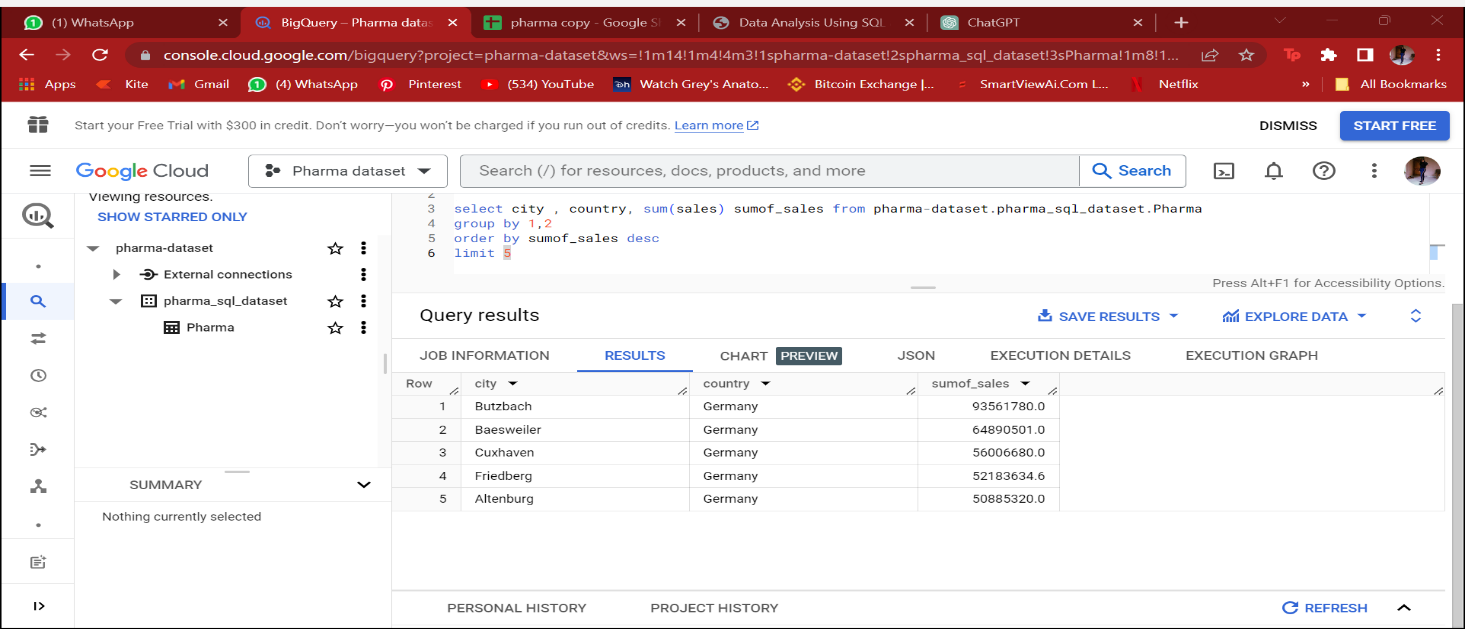
**HIGHEST:**

- select city , country, sum(sales) sumof\_sales from pharma-dataset.pharma\_sql\_dataset.Pharma

group by 1,2

order by sumof\_sales desc

limit 5

- 

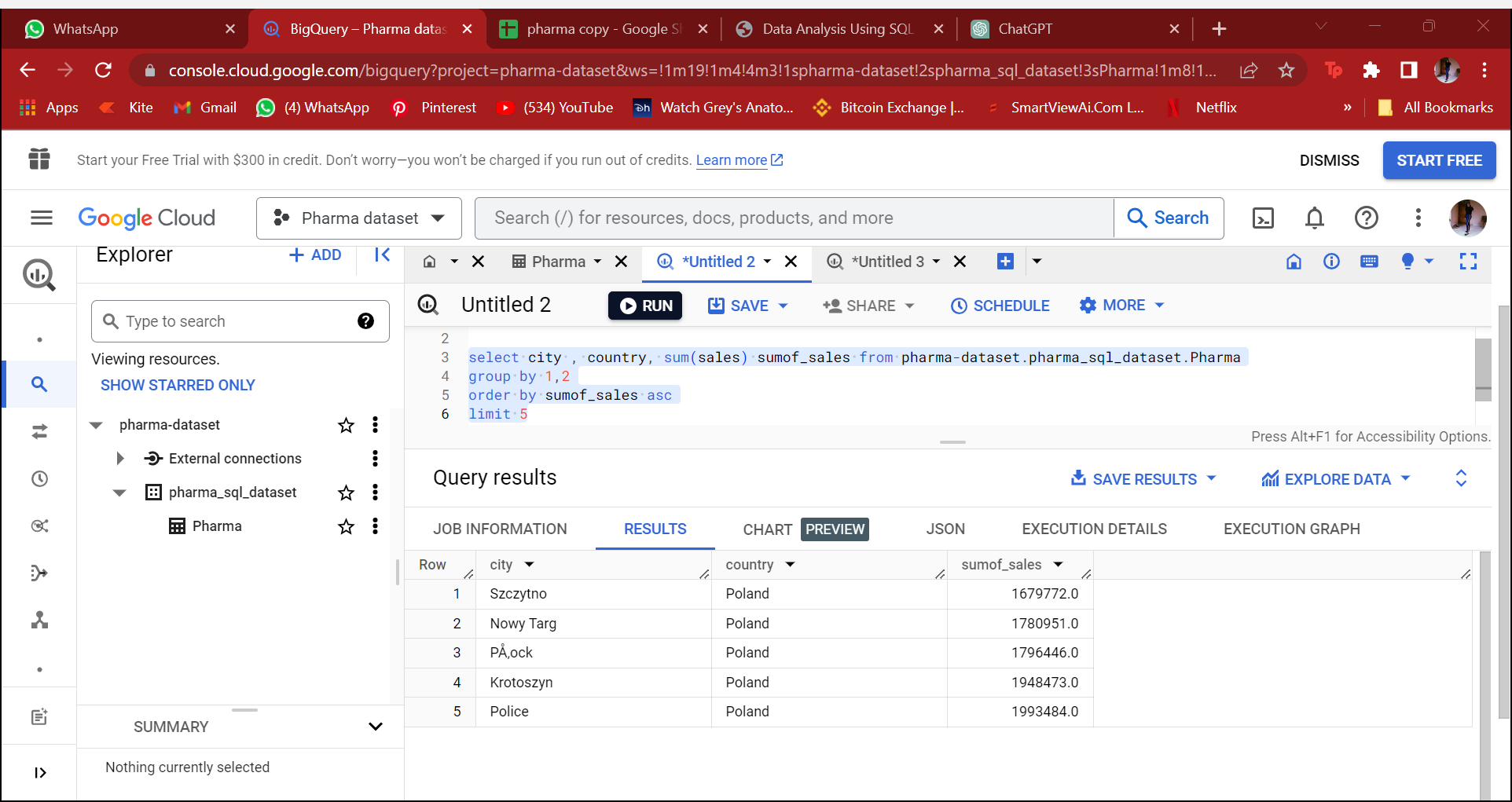
**LOWEST:**

select city , country, sum(sales) sumof\_sales from pharma-dataset.pharma\_sql\_dataset.Pharma

group by 1,2

order by sumof\_sales asc

limit 5



Analysis: Most of the cities from Poland has the lowest market penetration as the data is also specifically for the year 2018. May be further in few years it might take a lead on distribution and sales of pharma .

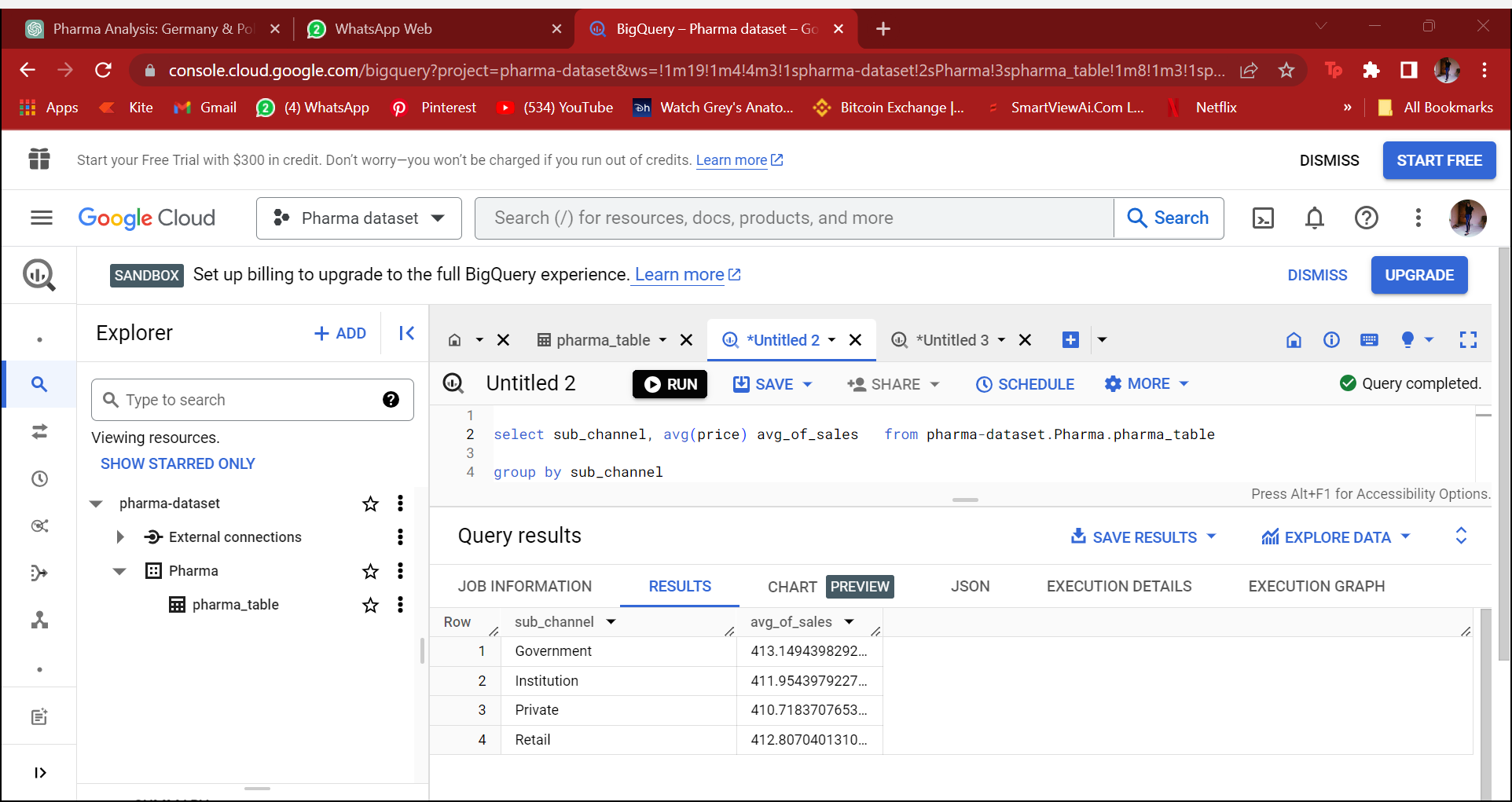
We will have to cater more to hospitals and private and in the retail market .

**8. Calculate the average price of products in each sub-channel**

- select sub\_channel, avg(price) avg\_of\_sales   from pharma-dataset.pharma\_sql\_dataset.Pharma

group by sub\_channel

order by sub\_channel



Analysis: The average of sales has been almost similar in all sub-channels, Government being the top most but not with a very big difference.

**Market Overview and Trends:**

9. **Retrieve top 5 products in terms of sales along with the respective sales trends over the years.**

with top5 as

 (select Product\_name,  sum(sales) sumof\_sales  from pharma-dataset.Pharma.pharma\_table

   group by product\_name

   order by sumof\_sales desc

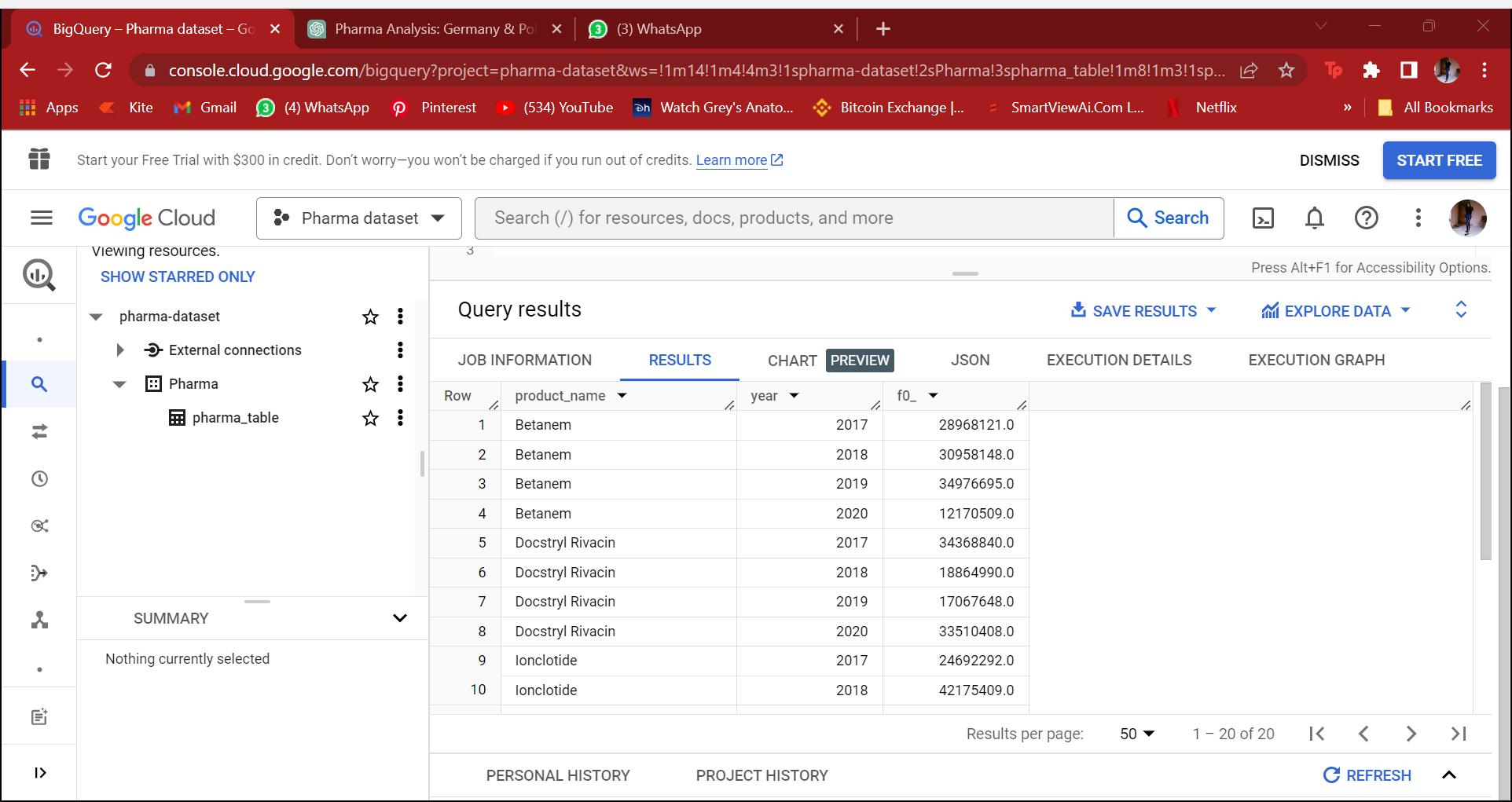
     limit 5)

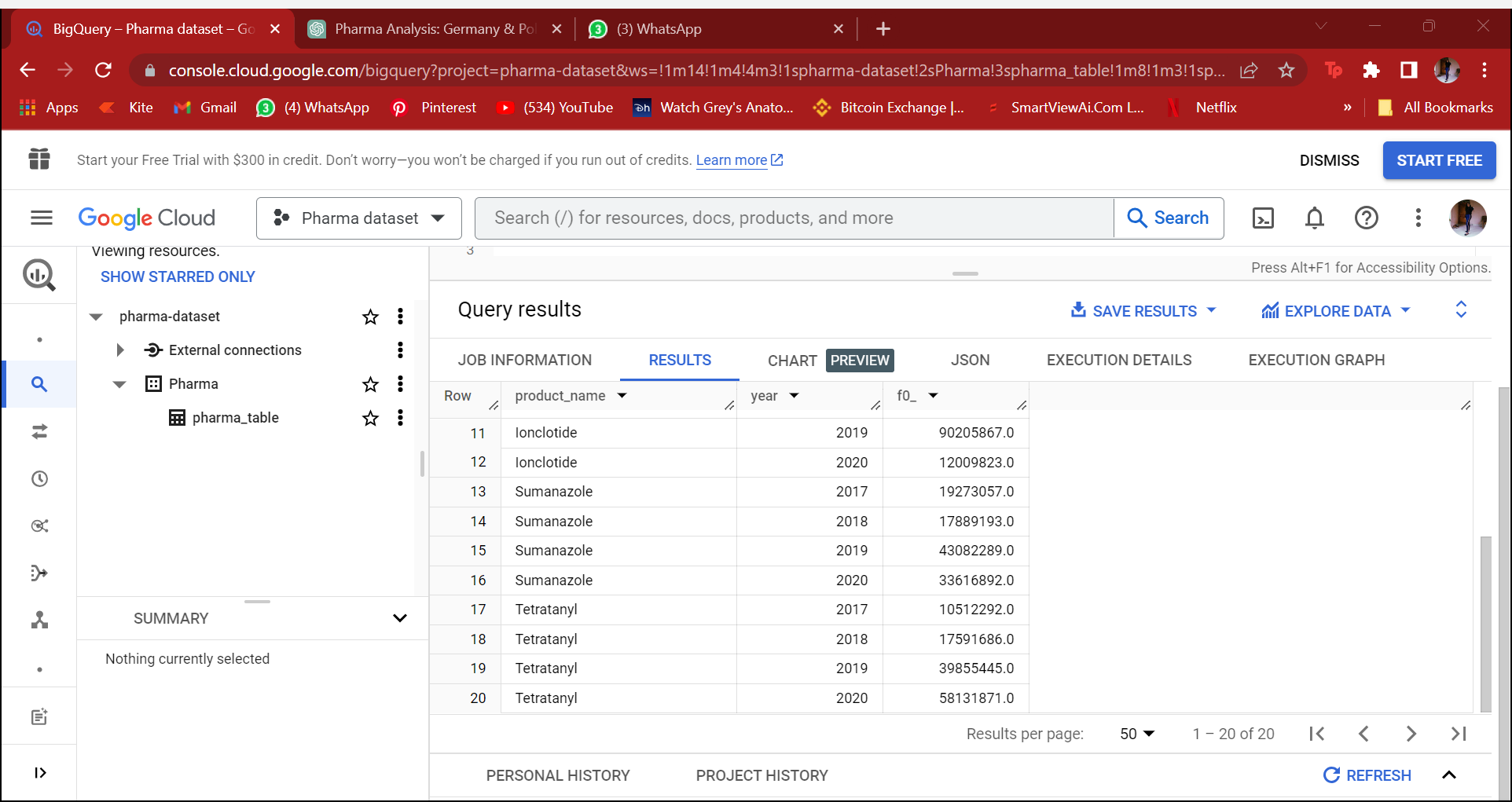
select ph.product\_name , year, sum(sales) from pharma-dataset.Pharma.pharma\_table ph

 join top5 t5 on ph.product\_name = t5.product\_name

 group by 1,2

order by 1,2





Analysis: As per my findings Betanem and Lonclotide has been the most increasing in every year apart from 2020, 2020 sales has been down comparatively previous three years.

Docstryl has been down in sales in 2018 and 2019 but in 2020 the sales increased massively.

Sumanazole- The first two years were down in sales but last 2 years(2019,2020) has been a boost in sales.

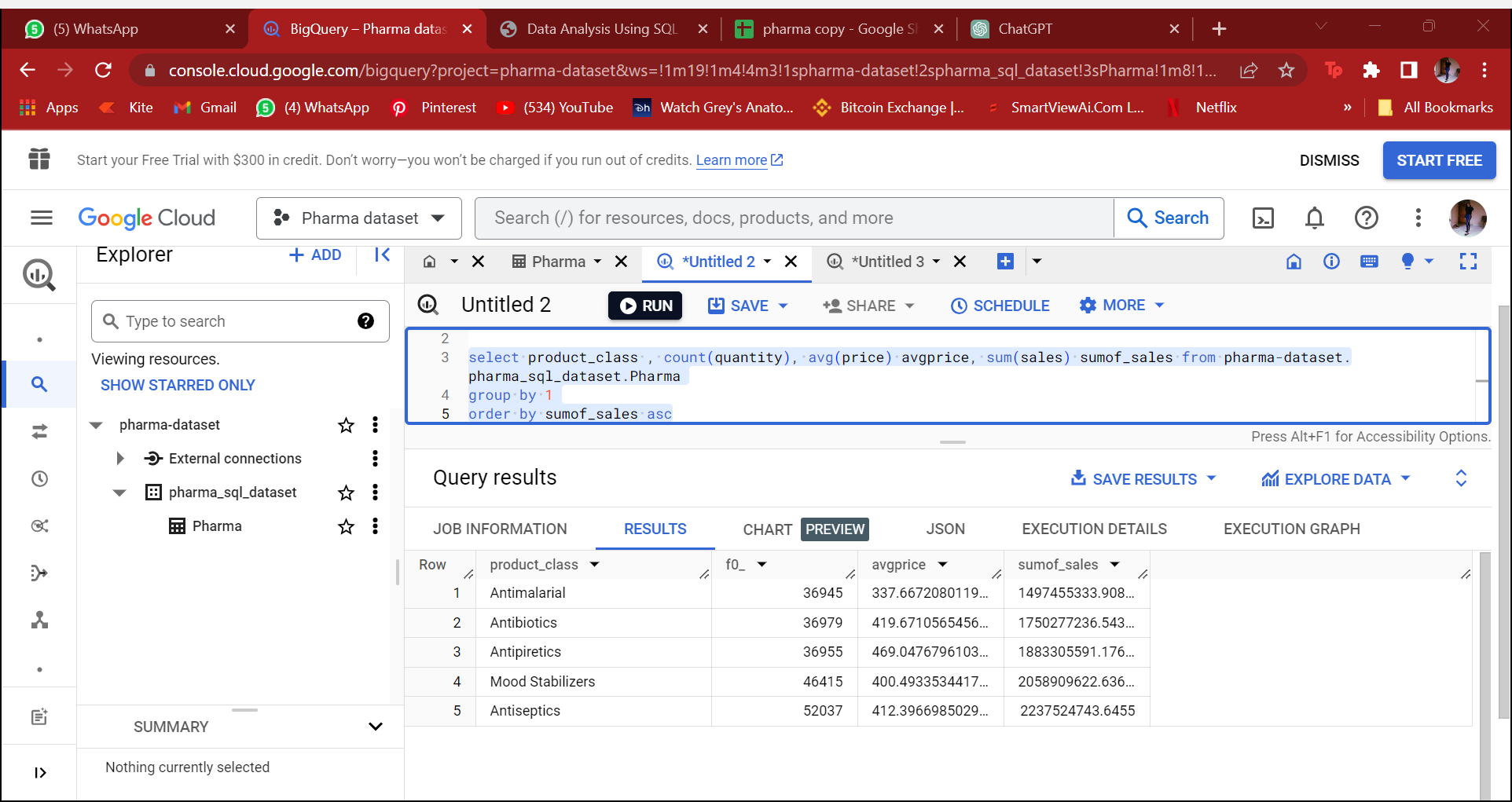
ONLY Trtratanyl has been constantly growing each and every year. There is no decline in the growth and sales of this particular product.

**10. Create a summary report that includes the total sales, average price, and total quantity sold for each product class**

- select product\_class , count(quantity), avg(price) avgprice, sum(sales) sumof\_sales from pharma-dataset.pharma\_sql\_dataset.Pharma

group by 1

order by sumof\_sales asc



Analysis: Antiseptics has been the most sold and highest sales product class among the others.

Antimalarial and Antibiotics has the almost similar quantity sold but has a big average price difference.

Additionally Antibiotics sales is more than the Antimalarial.

11. **what is the overall market share of different distributor in Germany?**

SELECT distinct Distributor,

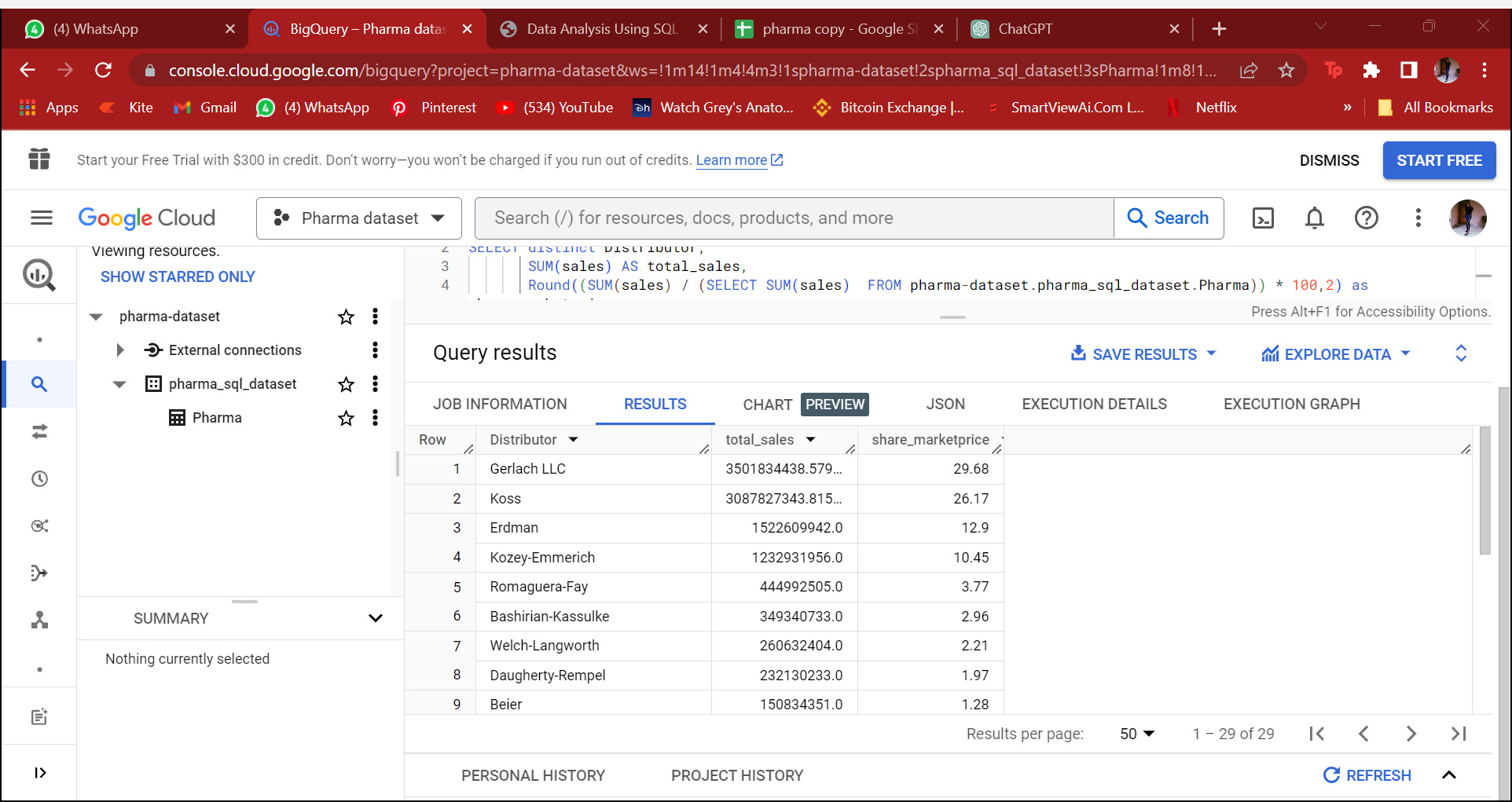
   SUM(sales) AS total\_sales,

       Round((SUM(sales) / (SELECT SUM(sales)  FROM pharma- dataset.pharma\_sql\_dataset.Pharma)) \* 100,2) as share\_marketprice

     from pharma-dataset.pharma\_sql\_dataset.Pharma

     group by Distributor

order by share\_marketprice desc



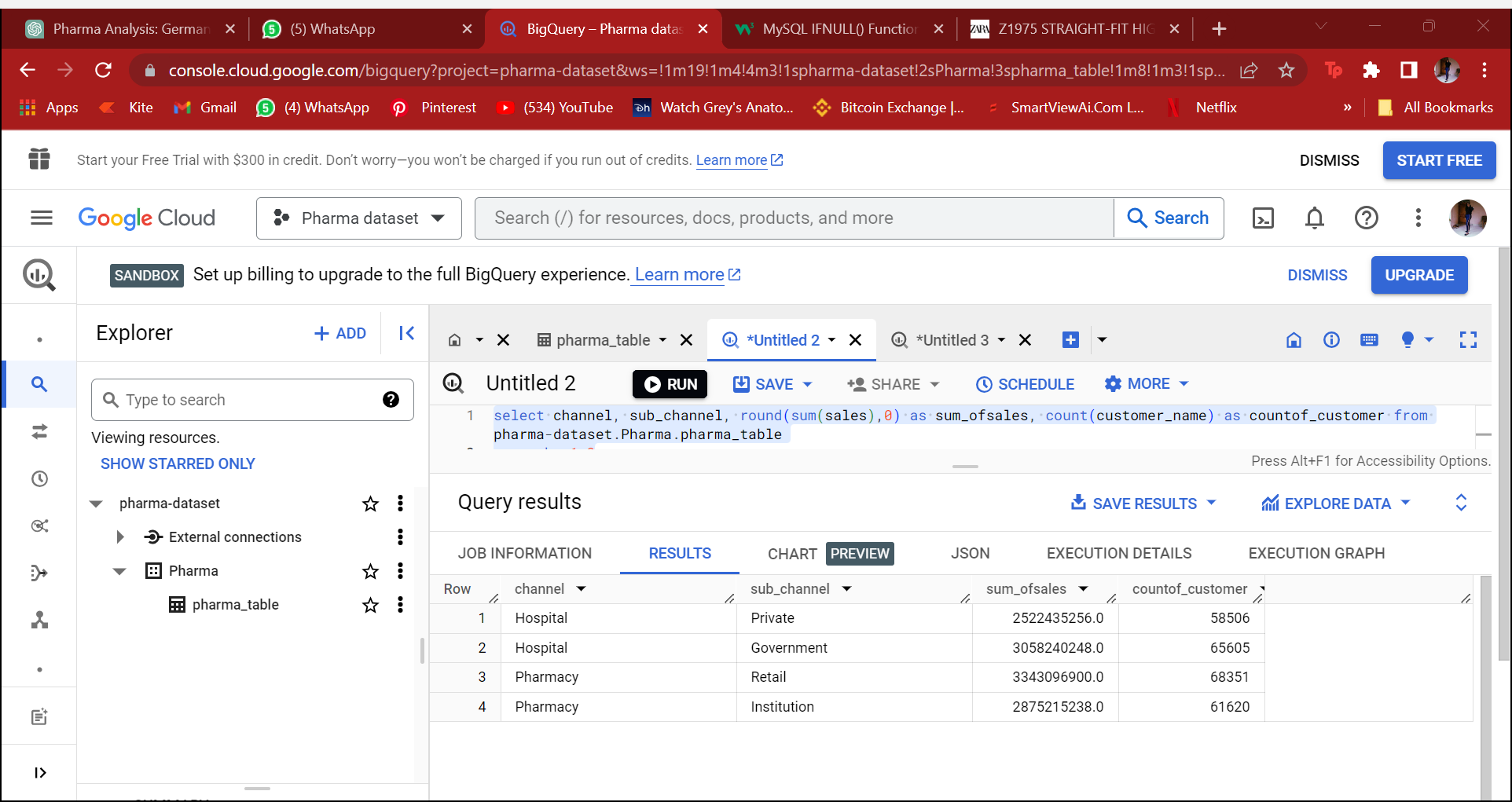
Analysis: As I can see that only 4 top distributor has contributed the most and almost 75% of sales are made by them. And the other distributor has contributes the rest 25% .

**Channel and Sub-channel Analysis**

**12. Which sales channels and sub-channels are the most effective in terms of sales and customer reach?**

select channel, sub\_channel, round(sum(sales),0) as sum\_ofsales, count(customer\_name) as countof\_customer from pharma-dataset.Pharma.pharma\_table

group by 1,2



Analysis: As per my findings Pharmacy channel(retail) has the most number of sales and customer reach and secondly hospital channel(government) has also the most customer reach

**Sales Team and Representative Performance:**

**How do different sales teams compare in terms of total sales and growth rates?**

select \*,

(previousyear - currentyear)/previousyear\*100 as change\_in\_growth

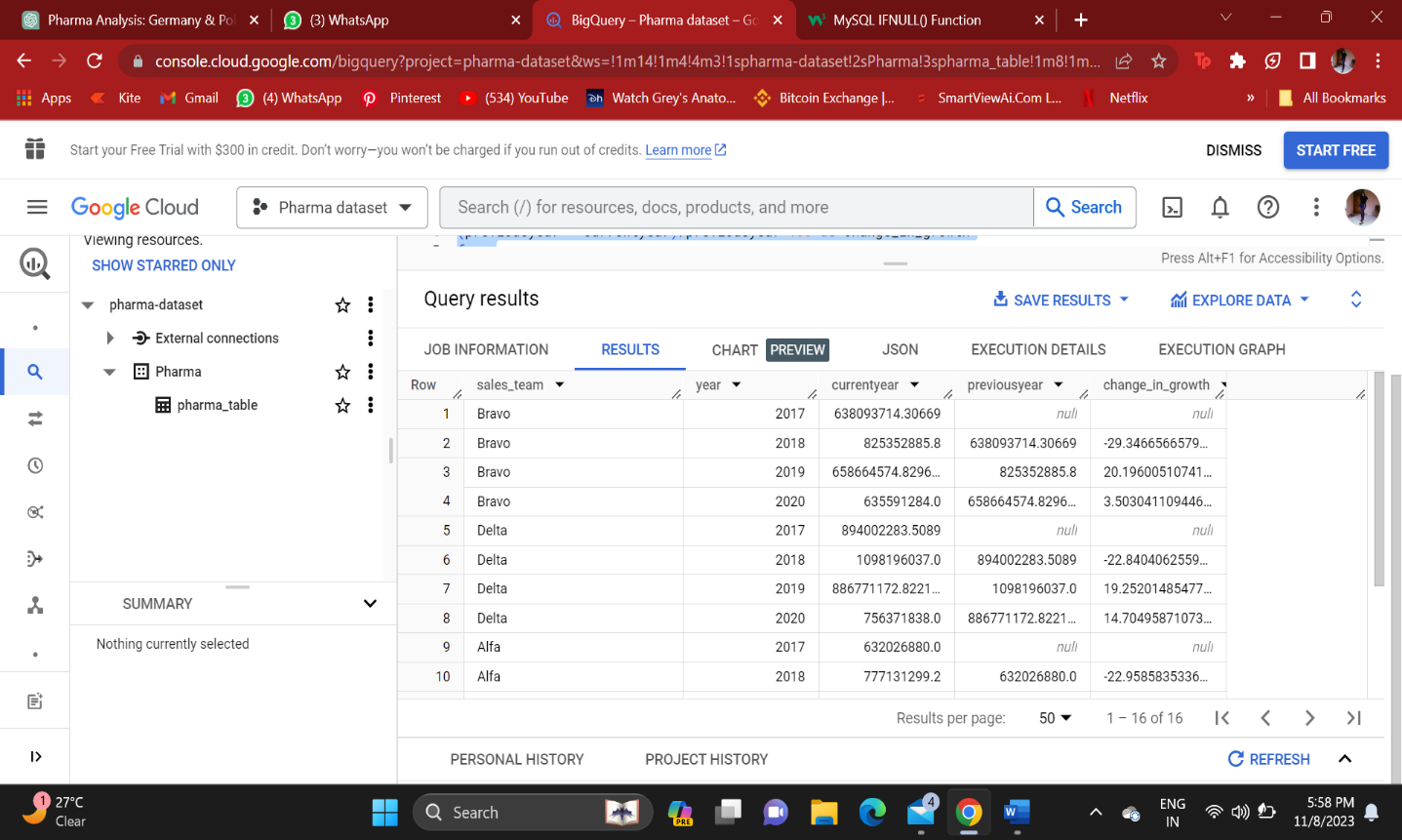
from

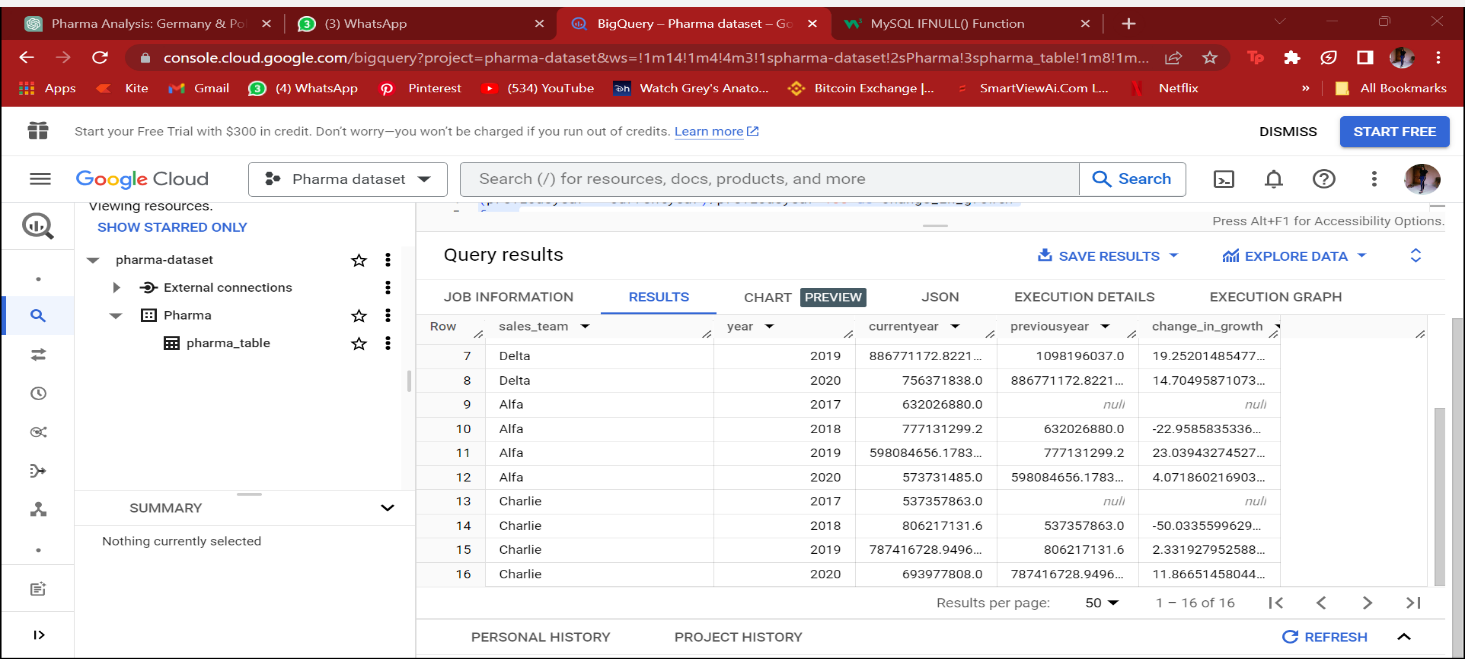
(select sales\_team,year, sum(ifnull(sales,0)) as currentyear,

lag(sum(ifnull(sales,0))) over (partition by sales\_team order by year) as previousyear,

 from pharma-dataset.Pharma.pharma\_table

group by 1,2)





Analysis: Delta and Charlie sales team has a good and massive growth, even while

Decline there is less difference in the amount of sales.

**14. Are there specific products or channels where certain sales teams or representatives excel?**

select sales\_team , product\_class, total\_quantity from

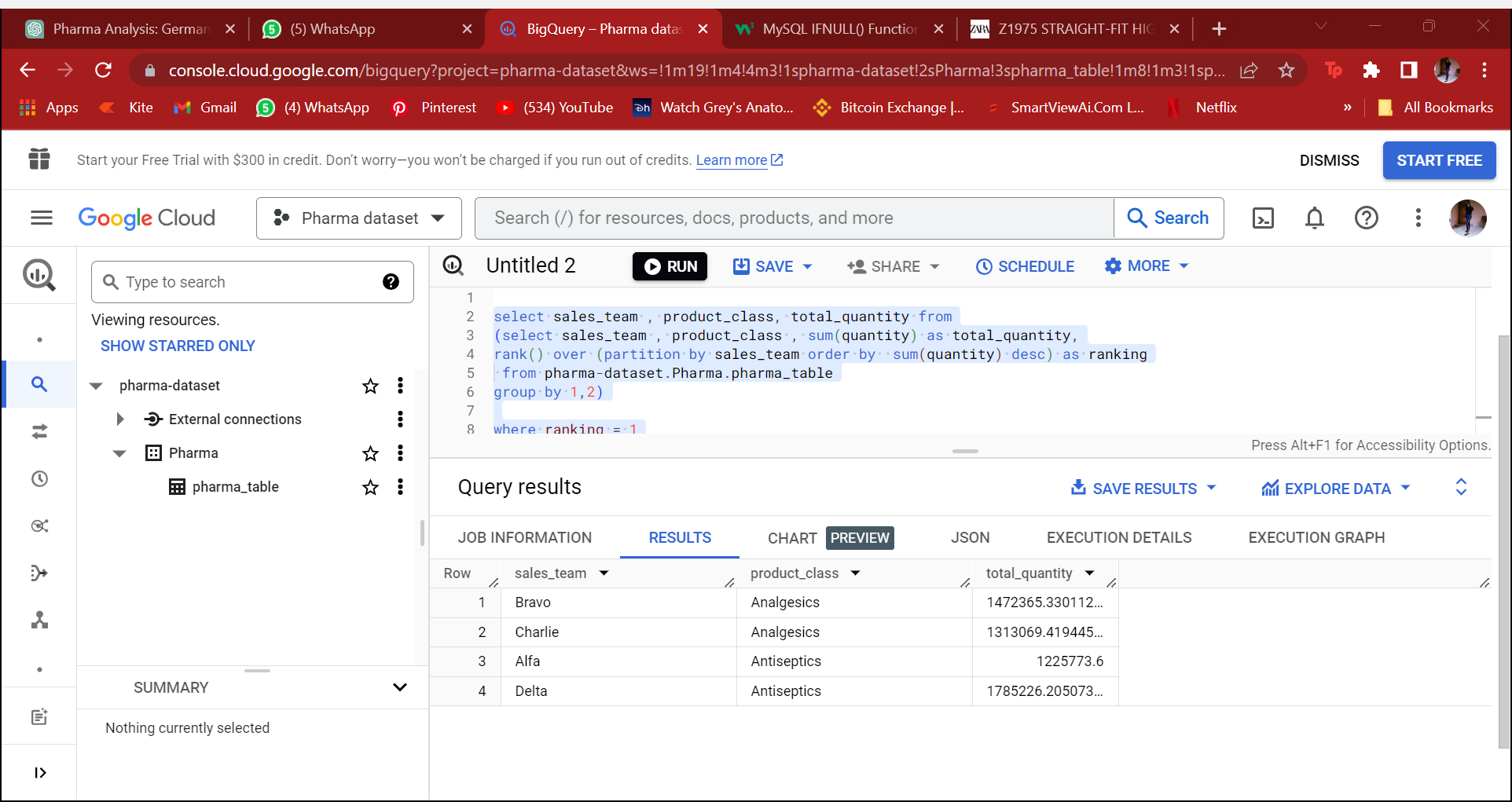
(select sales\_team , product\_class , sum(quantity) as total\_quantity,

rank() over (partition by sales\_team order by  sum(quantity) desc) as ranking

 from pharma-dataset.Pharma.pharma\_table

group by 1,2)

where ranking = 1



Analysis: there are 2 product class where sales team excel as the demand of Antiseptics is high and the sales too.