**Fetch Rewards Coding Exercise - Data Analytics Internship**

**Part 1: Relational Data Model**

The relational data model consists of 4 tables: Users, Receipts, Receipt\_items and Brands as shown below:

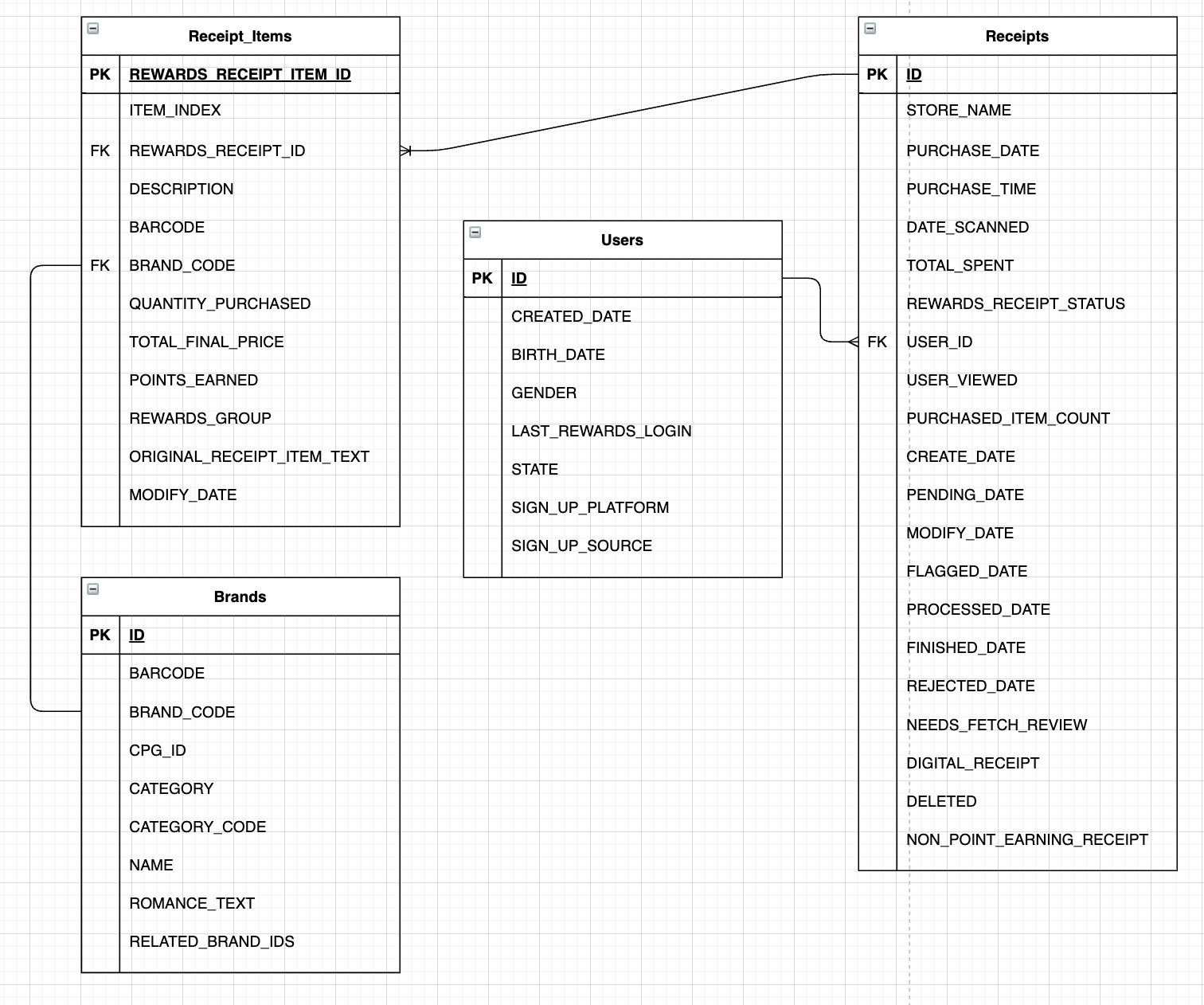
****

Diagram 1: Relational Data Model

(Software used: draw.io)

**Potential data warehouse design problem & proposed recommendation:**

Given the current scenario, the Brands table cannot be effectively joined with any of the remaining three tables. For the existing dataset, Brands.BRAND\_CODE has been used to join with Receipt\_items.BRAND\_CODE. But BRAND\_CODE has null values.

A better recommendation here would be to include BRAND\_ID in the Receipt\_items table which can then act as a joinable key with Brands.id.

**Part 2: SQL queries to answer questions from a business stakeholder**

For part 2 of this exercise, BigQuery has been used to perform analytical queries in order to answer the following questions:

1. **Which brand saw the most dollars spent in the month of June?**

**Clarifying question to be asked**: Is the stakeholder interested in knowing the most dollars spent for any specific year? Does the stakeholder want a year-wise breakup for highest dollars spent on a Brand or just a cumulative highest spending in June over the years?

GREAT VALUE saw the most dollars spent in June over the years.

Query:

SELECT r\_i.BRAND\_CODE, sum(r.TOTAL\_SPENT) as most\_spent

FROM `fetch.receipt\_items` as r\_i

join `fetch.receipts` as r

on r\_i.REWARDS\_RECEIPT\_ID = r.ID

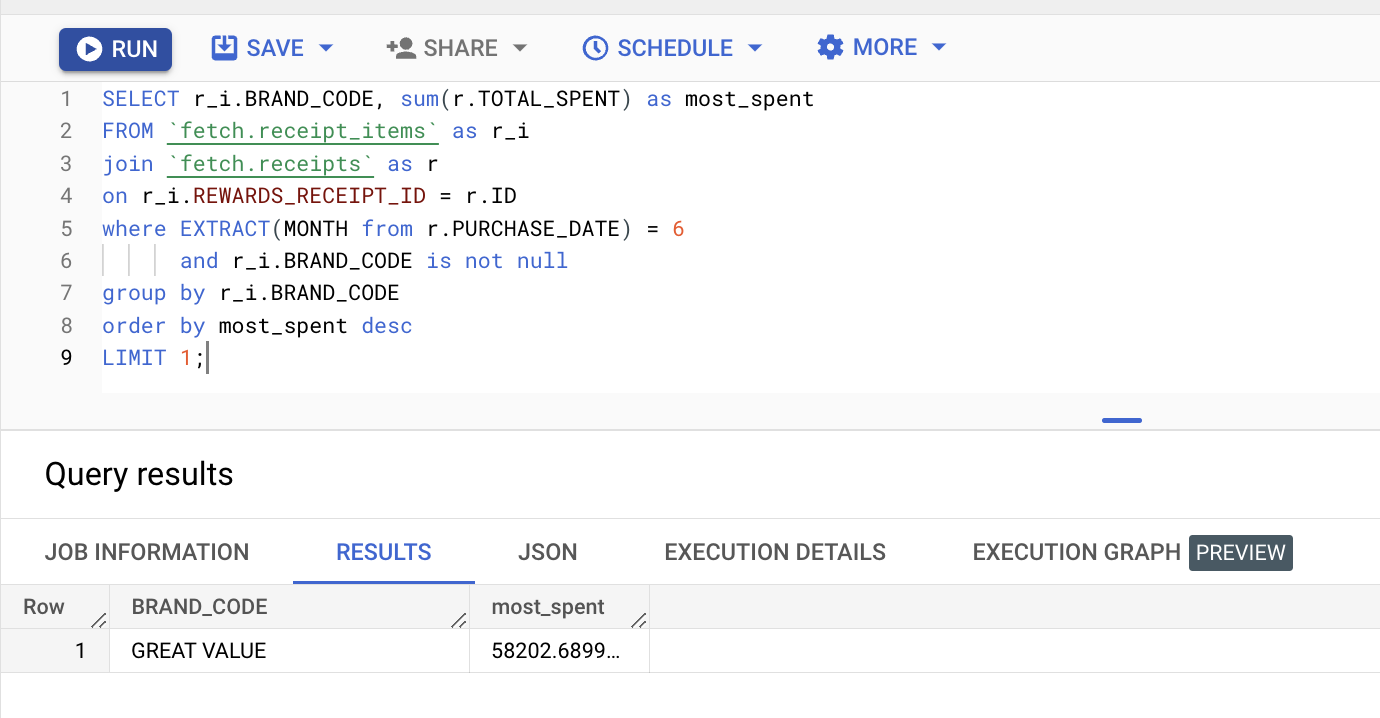
where EXTRACT(MONTH from r.PURCHASE\_DATE) = 6

and r\_i.BRAND\_CODE is not null

group by r\_i.BRAND\_CODE

order by most\_spent desc

LIMIT 1;

****

1. **Which user spent the most money in the month of August?**

**Clarifying question to be asked**: Is the stakeholder interested in knowing the highest spending user for any specific year? Does the stakeholder want a year-wise breakup for highest spending or just a cumulative highest spending in August over the years?

The user having ID as6032cb807d464912dab4dc1e has spent the most in August.

Query:

select USER\_ID, sum(TOTAL\_SPENT) as total\_spending

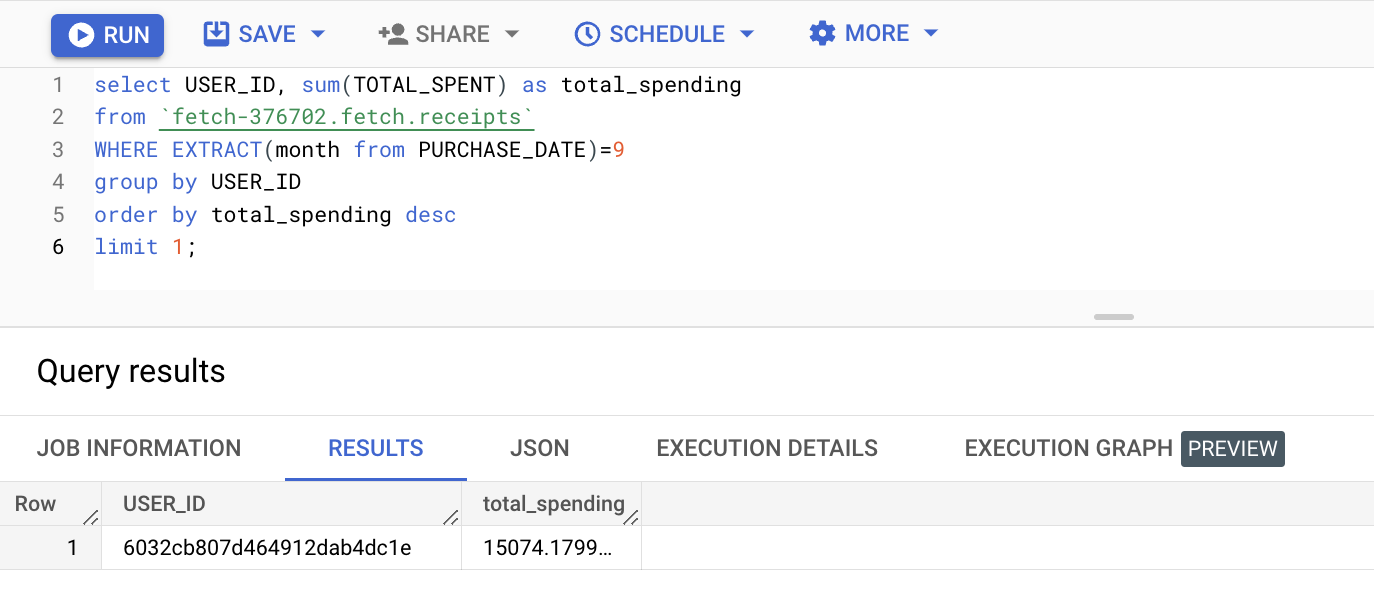
from `fetch-376702.fetch.receipts`

WHERE EXTRACT(month from PURCHASE\_DATE)=9

group by USER\_ID

order by total\_spending desc

limit 1;

****

1. **What user bought the most expensive item?**

The user having ID as617376b8a9619d488190e0b6 has bought the most expensive item.

Query:

select b.USER\_ID

from

(SELECT REWARDS\_RECEIPT\_ID,

max(TOTAL\_FINAL\_PRICE/nullif(QUANTITY\_PURCHASED,0)) as most\_expensive\_item\_price

FROM `fetch-376702.fetch.receipt\_items`

group by REWARDS\_RECEIPT\_ID, DESCRIPTION

order by most\_expensive\_item\_price desc

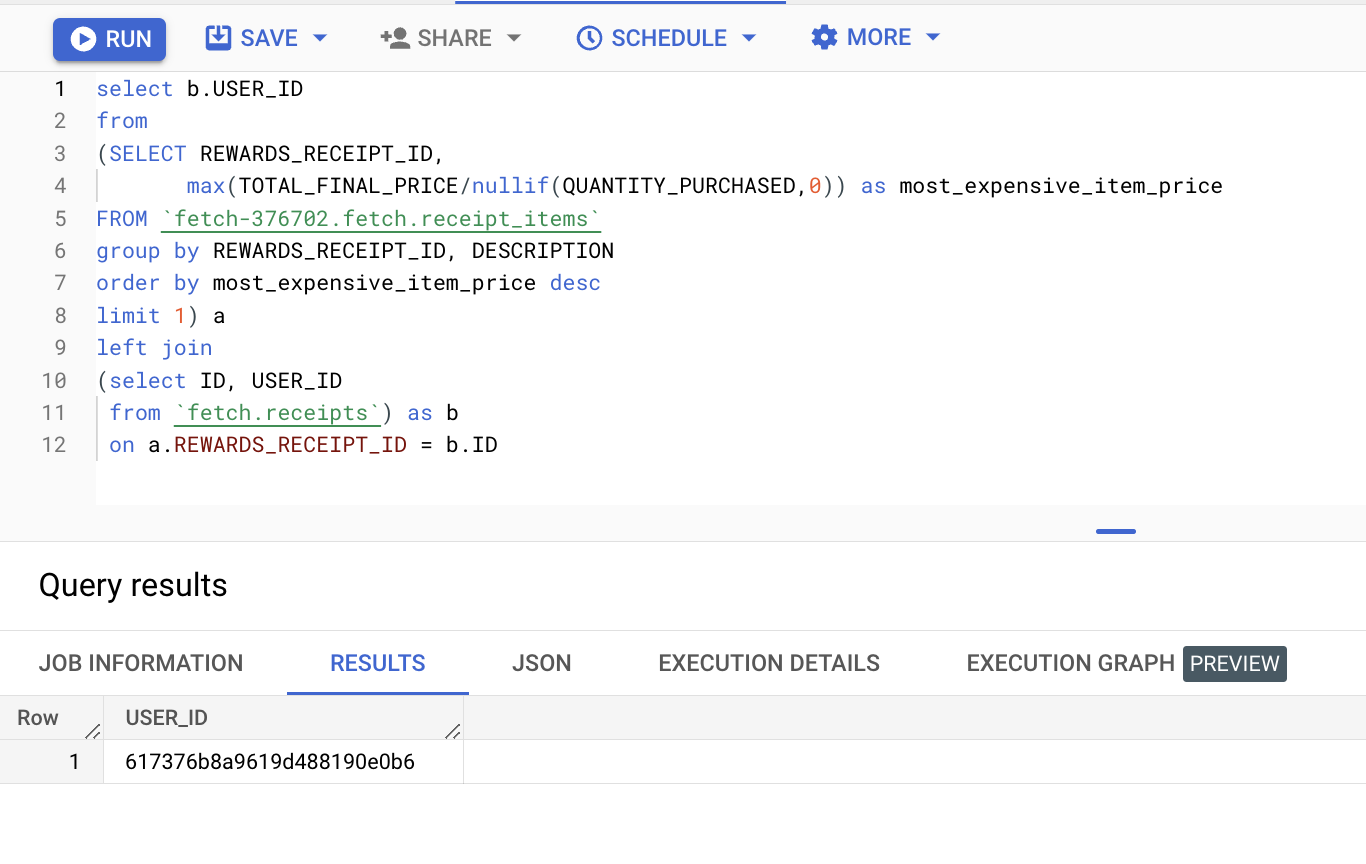
limit 1) a

left join

(select ID, USER\_ID

from `fetch.receipts`) as b

on a.REWARDS\_RECEIPT\_ID = b.ID

****

1. **What is the name of the most expensive item purchased?**

Starbucks Iced Coffee Premium Coffee Beverage Unsweetened Blonde Roast Bottle 48 Oz 1 Ct

Query:

SELECT max(TOTAL\_FINAL\_PRICE/nullif(QUANTITY\_PURCHASED,0)) as most\_expensive\_item\_price,

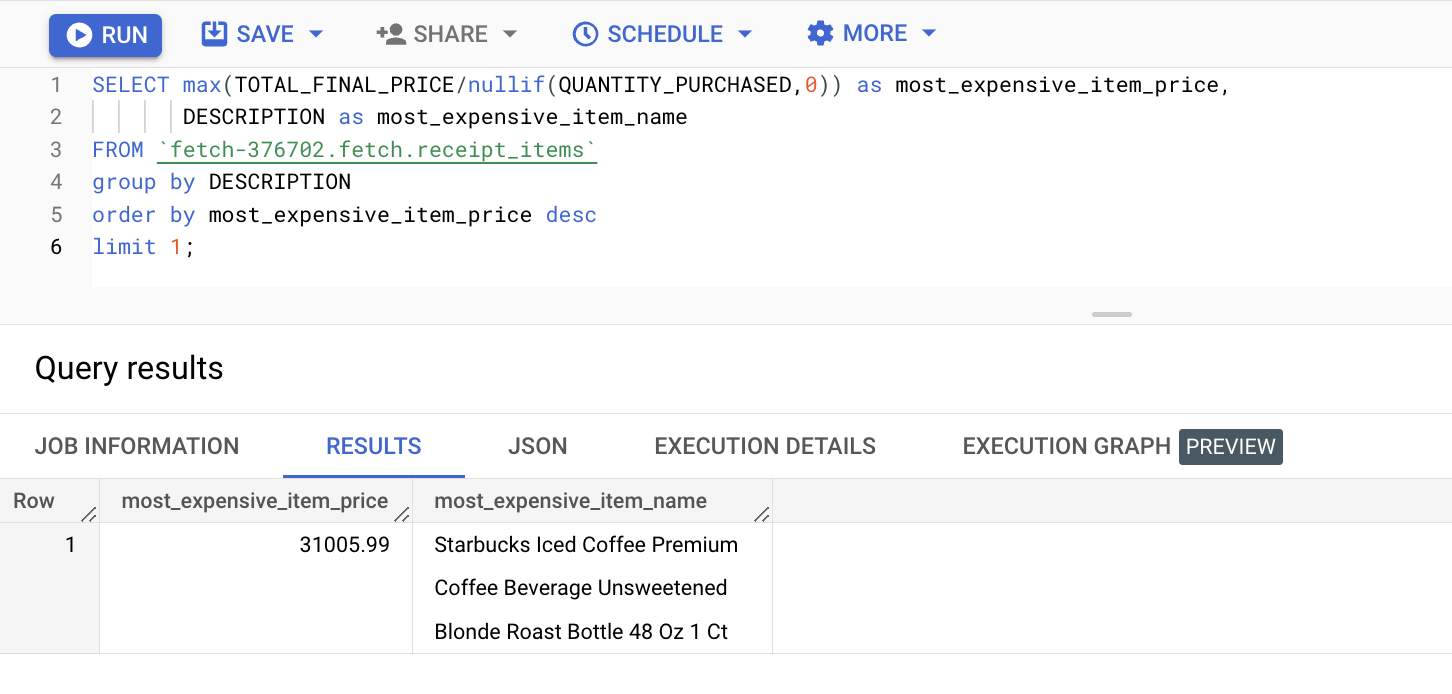
DESCRIPTION as most\_expensive\_item\_name

FROM `fetch-376702.fetch.receipt\_items`

group by DESCRIPTION

order by most\_expensive\_item\_price desc

limit 1;



1. **How many users scanned in each month?**

Total unique users scanned each month:

select EXTRACT(year from DATE\_SCANNED) AS year,

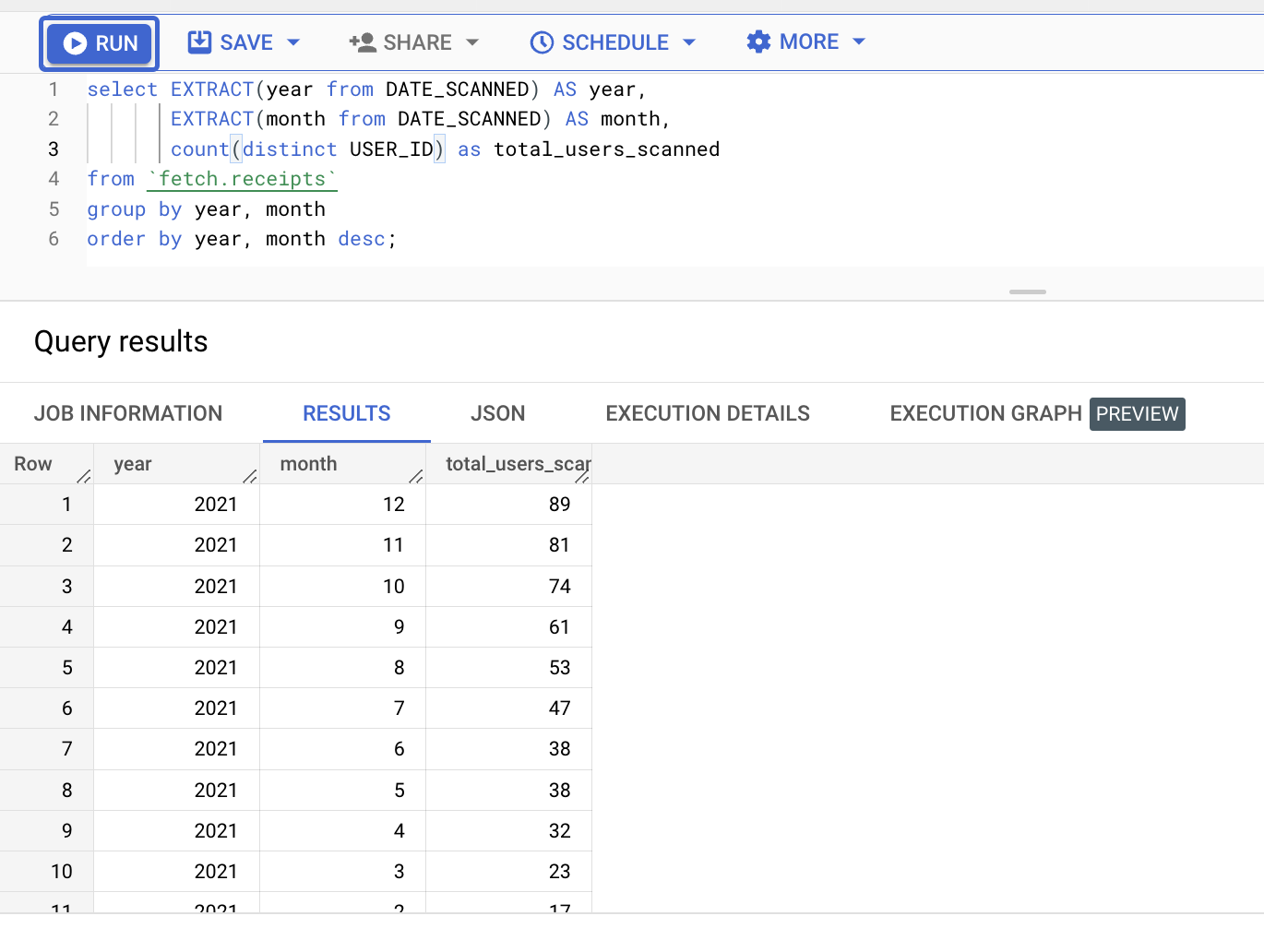
EXTRACT(month from DATE\_SCANNED) AS month,

count(distinct USER\_ID) as total\_scans

from `fetch.receipts`

group by year, month

order by year, month desc;

****

Total scans per month:

Query:

select EXTRACT(year from DATE\_SCANNED) AS year,

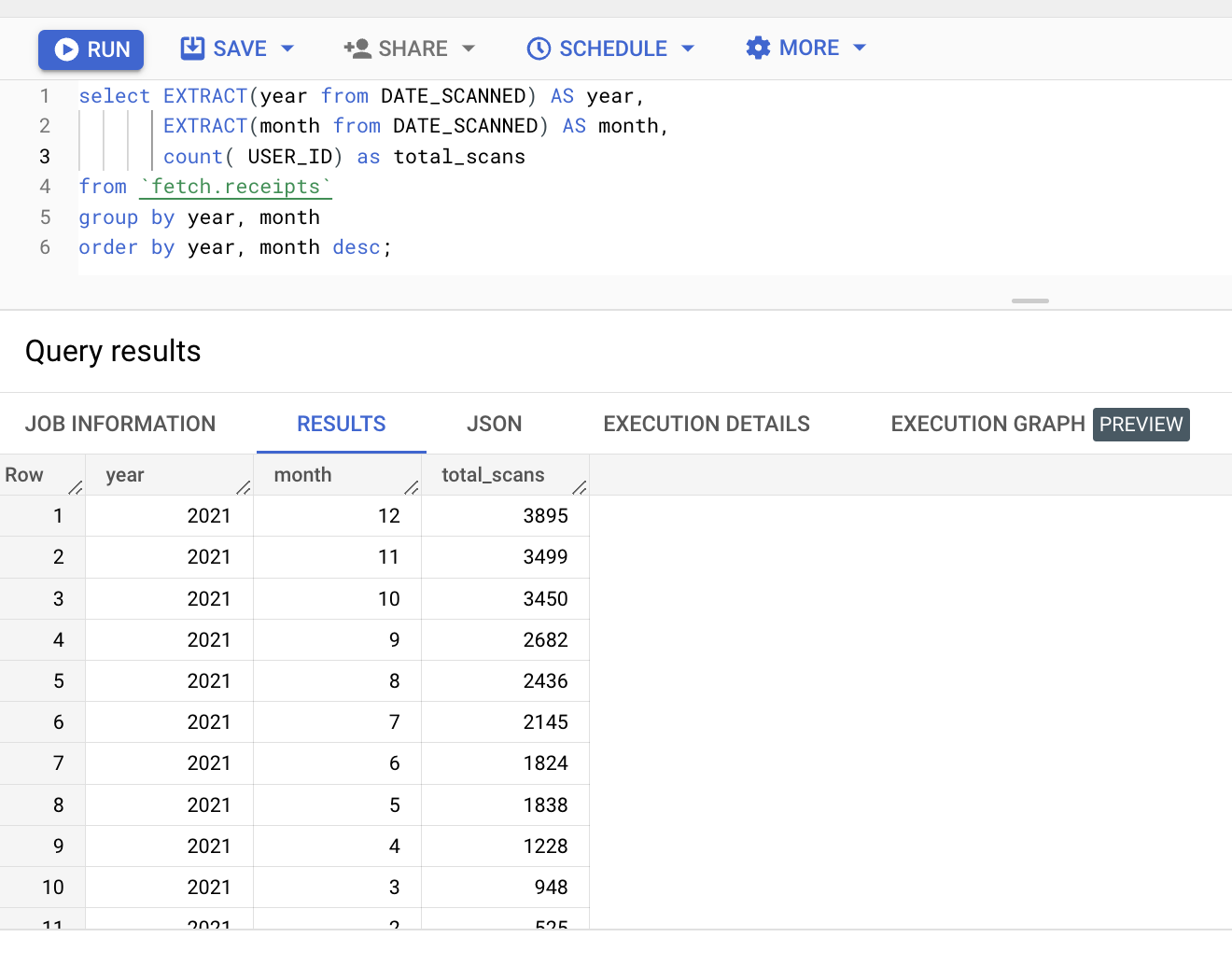
EXTRACT(month from DATE\_SCANNED) AS month,

count( USER\_ID) as total\_scans

from `fetch.receipts`

group by year, month

order by year, month desc;



**Part 3: Data Insights & Recommendations**

**Database design Recommendation:**

The Brands table cannot be effectively joined with any of the remaining three tables. For the existing dataset, Brands.BRAND\_CODE has been used to join with Receipt\_items.BRAND\_CODE. But BRAND\_CODE has null values.

**Recommendation:** include BRAND\_ID in the Receipt\_items table which can then act as a joinable key with Brands.ID.

**Analysing the receipts data:**

Tool used for analysis: Tableau

Looking at the number of receipts generated year-wise it is observed that majority of the receipts are generated from 2021 to 2023. Very few entries are from 1970 which can potentially be an error in storing the date and should be checked upon. Similar is the case for the years 2018, 2019, 2020 as seen in Chart 1.

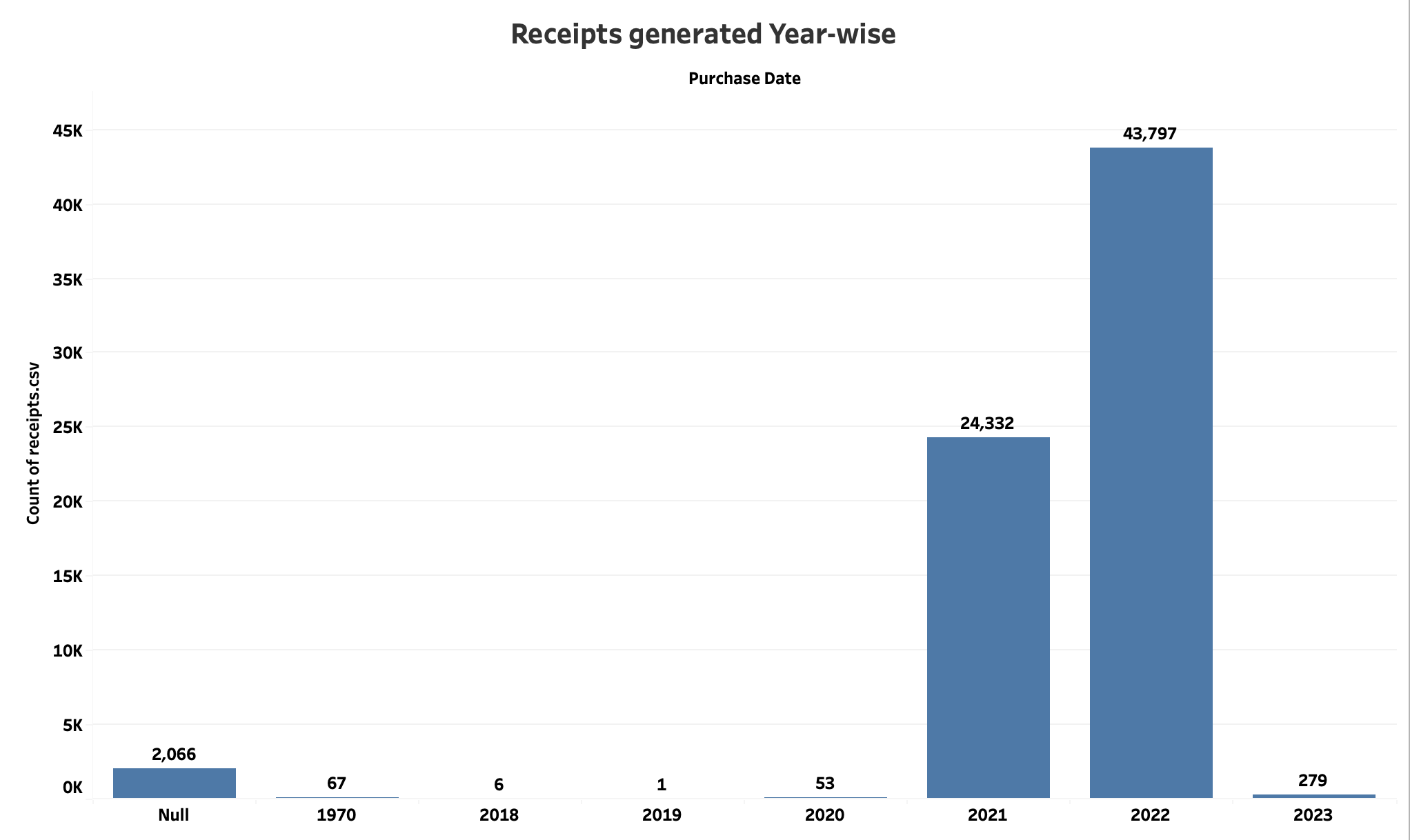


Chart 1: Receipts generated Year-wise

**Charting some monthly and yearly trends :**

*Including receipts data only from 2021 to 2023*

Peak spending month was July for 2021 while for 2022 it was August. The second most highest spending occurs in the Month of December.

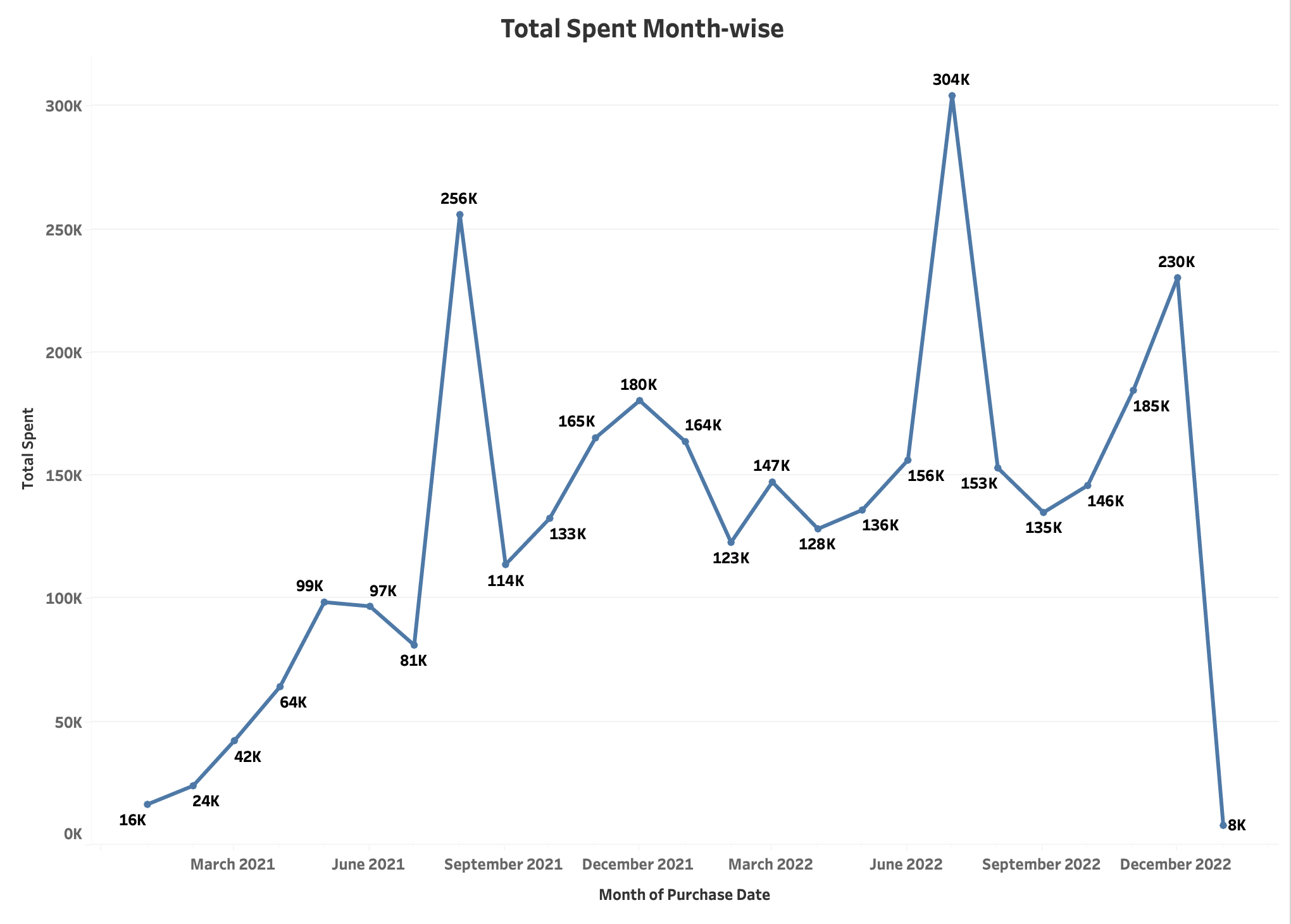


Chart 2: Total spent Month-wise

**Highest spending users:**

Identifying the highest spending users can help to not only understand who spends the most, but also device strategies such as offering discounts for highest spending customers. Providing some incentives/points for the customers who do not spend the most, but can potentially engage more in terms of purchasing products if introduced to such benefits.

Various targeted marketing strategies can be implemented by identifying different customer segments.

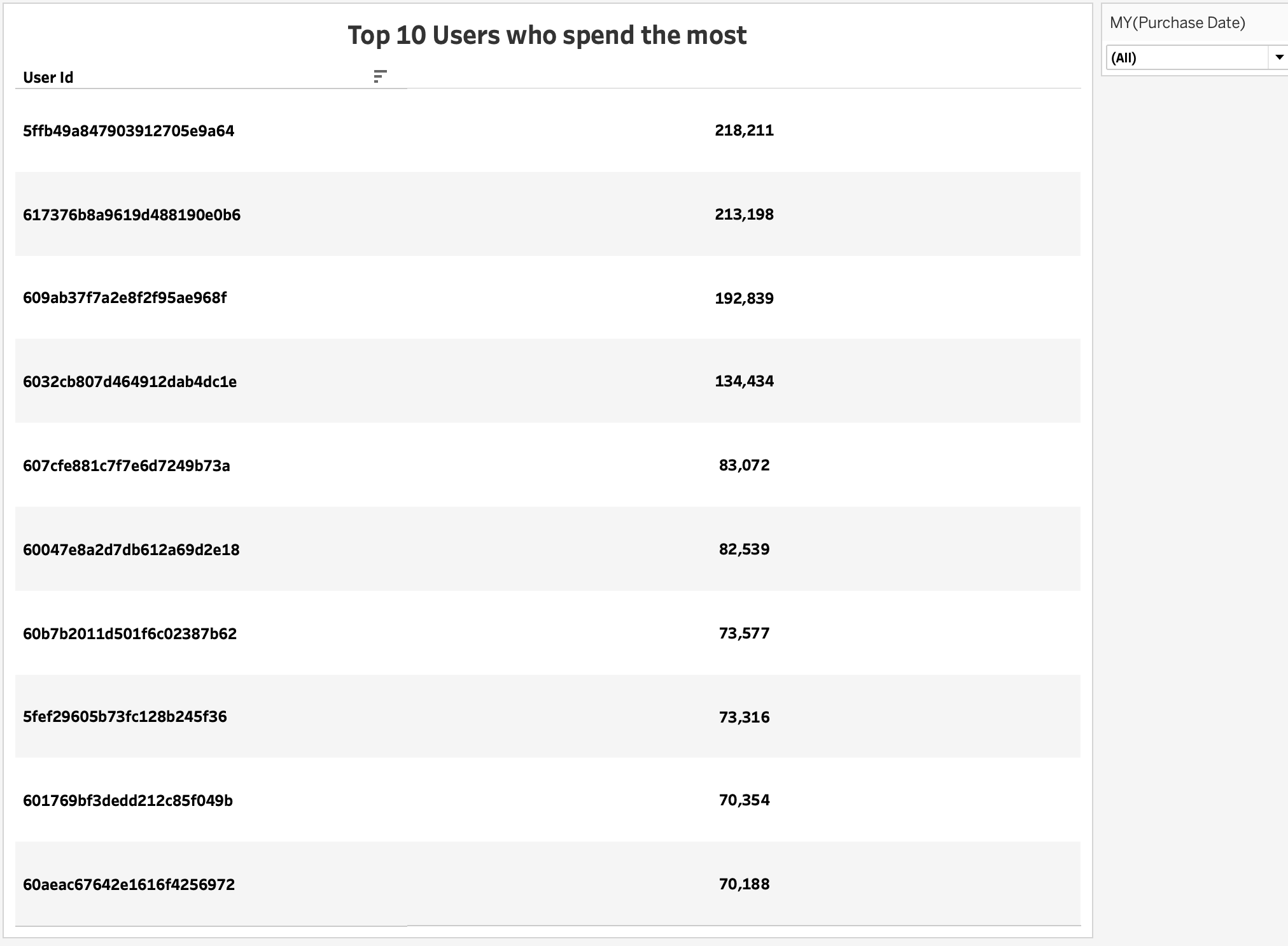


Chart 3: Top 10 users who spend the most

**Top stores based on most spending:**

Use of variable parameter is done here. This implies the user can select Top X stores to be displayed. If the user selects X as 10 then the top 10 stores will be displayed as can be seen in Chart 4.

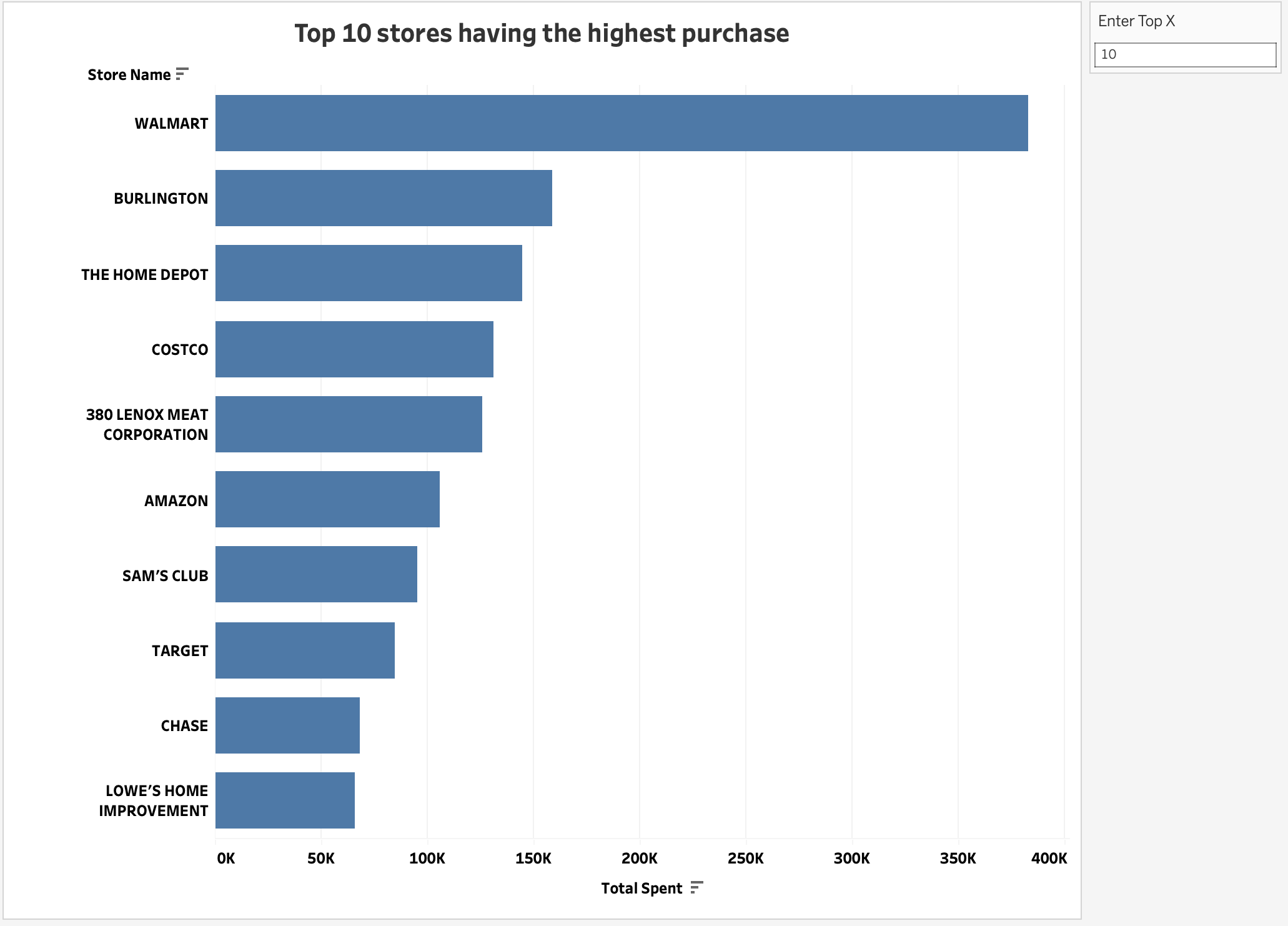


Chart 4: Top X stores based on spending

A User Activity Dashboard summarizing all the insights: [Published on Tableau Public](https://public.tableau.com/app/profile/neha.bhujbal/viz/Fetch_Data_Assignment/Dashboard1?publish=yes)

