Insights obtained by analyzing the data

The data consists of details about users, including their purchase receipts and information about the brands associated with the items they bought. This information can be analyzed to uncover patterns in sales, identify frequent buyers, and determine which stores are most popular. I have found some key insights from this data that can be communicated to non-technical stakeholders.

1. Quality of data:

By observing the data, we can find that many data values are missing or are NULL. Different data cleaning techniques can be used to improve the quality of the data and also to increase the correctness of analysis.
Few examples of missing data are: Null brand codes, quantity purchased and reward points in the receipt_items csv. These values should either be made 0 or appropriate values should be filled as NULL values cause several issues like data integrity issues, query and performance issues.

2. Determining the top 10 stores with the highest number of purchases.

Query used:

```
SELECT COUNT(*) as no_of_purchases, STORE_NAME FROM fetch.receipts r GROUP BY STORE_NAME ORDER BY no_of_purchases DESC;
```

Results:

	no_of_purchases	STORE_NAME
1	6,943	WALMART
2	2,787	AMAZON
3	1,836	
4	1,805	TARGET
5	1,780	DUNKIN DONUTS
6	1,632	MCDONALD'S
7	1,403	DOLLAR TREE STORES
8	1,393	WALGREENS
9	1,225	PUBLIX
10	1,102	COSTCO

Analysis: Maximum number of purchases was made at Walmart among the 7119 stores found in the data and least number of purchases were made at several stores. This information can be used to determine what qualities a store possesses that attracts more clients and steps can be taken to make appropriate changes in stores that are not performing well.
Some of the things that can influence this is the store location, customer experience, variety of products offered etc.
We can also observe that the third entry in the above list doesn't have a store name. Data cleaning will help us avoid such results and will increase data correctness.

3. Determining the top 10 stores with the highest total sales.

Query used:

SELECT SUM(TOTAL_SPENT) **as** total_spenttt, STORE_NAME **FROM fetch**.receipts r **GROUP BY** STORE_NAME **ORDER BY** total_spenttt **DESC**;

Results:

	total_spenttt	STORE_NAME
1	386,335.79	WALMART
2	158,594.05	BURLINGTON
3	144,646.99	THE HOME DEPOT
4	131,248.32	COSTCO
5	125,721.85	380 LENOX MEAT CORPORATION
6	105,926.14	AMAZON
7	95,306.62	SAM'S CLUB
8	84,588.84	TARGET
9	68,176.92	CHASE
10	65,731.87	LOWE'S HOME IMPROVEMENT

Analysis: It is interesting to see that total sales and max purchases queries give us two different list of stores. Through this we can know that even if the number of purchases were less in a few stores, the total sales amount is high. This is solely determined by the type of goods sold at that store and the price point of these goods. Hence, during data analysis we need to compare stores of similar category to get more meaningful results.

4. Determining the top 10 customers who spent the most:

Query used:

SELECT SUM(TOTAL_SPENT) **as** total_spenttt, USER_ID **FROM fetch**.receipts r **GROUP BY** USER_ID **ORDER BY** total_spenttt **DESC**;

Results:

	total_spenttt	USER_ID
1	218,211.45	5ffb49a847903912705e9a64
2	213,197.99	617376b8a9619d488190e0b6
3	192,838.56	609ab37f7a2e8f2f95ae968f
4	134,434.34	6032cb807d464912dab4dc1e
5	83,072.44	607cfe881c7f7e6d7249b73a
6	82,538.82	60047e8a2d7db612a69d2e18
7	73,577.4	60b7b2011d501f6c02387b62
8	73,315.84	5fef29605b73fc128b245f36
9	70,354.37	601769bf3dedd212c85f049b
10	70,187.77	60aeac67642e1616f4256972

Analysis: Through such statistics we can determine which customers are actively buying. Targeted sales can be done, or incentives can be provided to loyal customers. We can also dive deeper and analyze the kind of products that these customers are purchasing to understand the current market better.