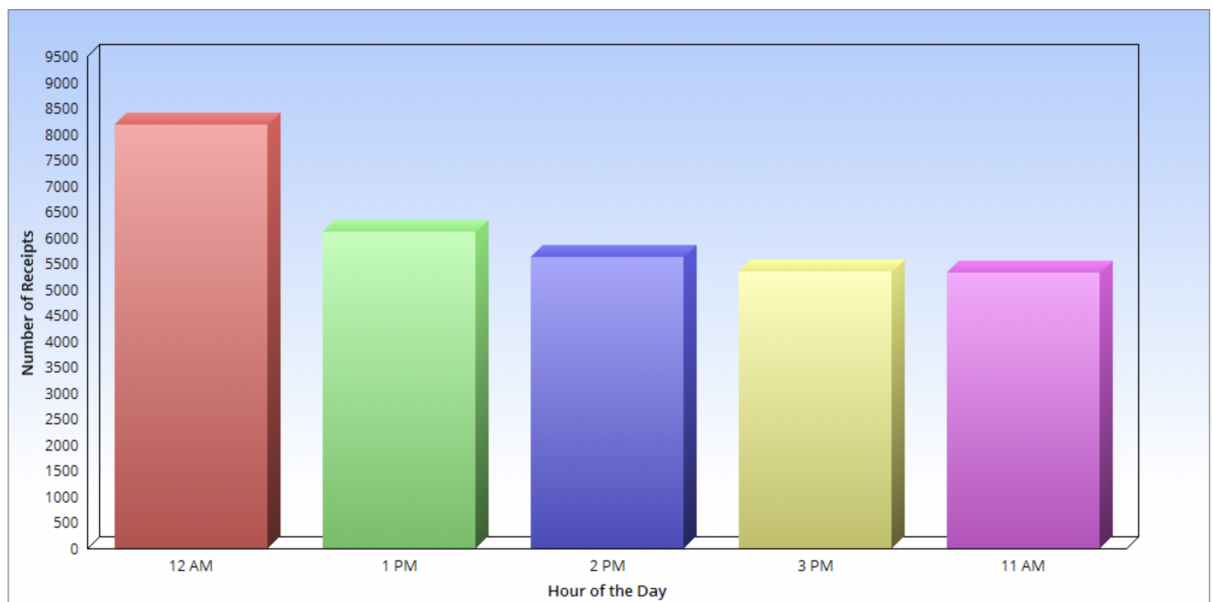


Buyer trends and Observations:

- **Companies:** Top 5 companies continue to consistently hold leading position w.r.t. customer spending
- **Sign Up:** 31.09% users signed up using Apple. To attract other non-Apple users we can reward extra points to them on signing up with Google, Facebook or Email.
- **Brands:** Cheerios holds most of the related brand ids in the shopping list. This can help us understand which brand gives us the most data as compared to others.
- **Buyers:** Fetch Rewards users mostly belong to the age group of 40 to 55 years.
- **Stores:** Top 2 stores which have maximum share in scanned receipts are:
 - Walmart: 12.89%
 - Amazon: 5.17%

The average age of customers who shop at Walmart is 46.5 years and of customers who shop at Amazon is 44.34 years.

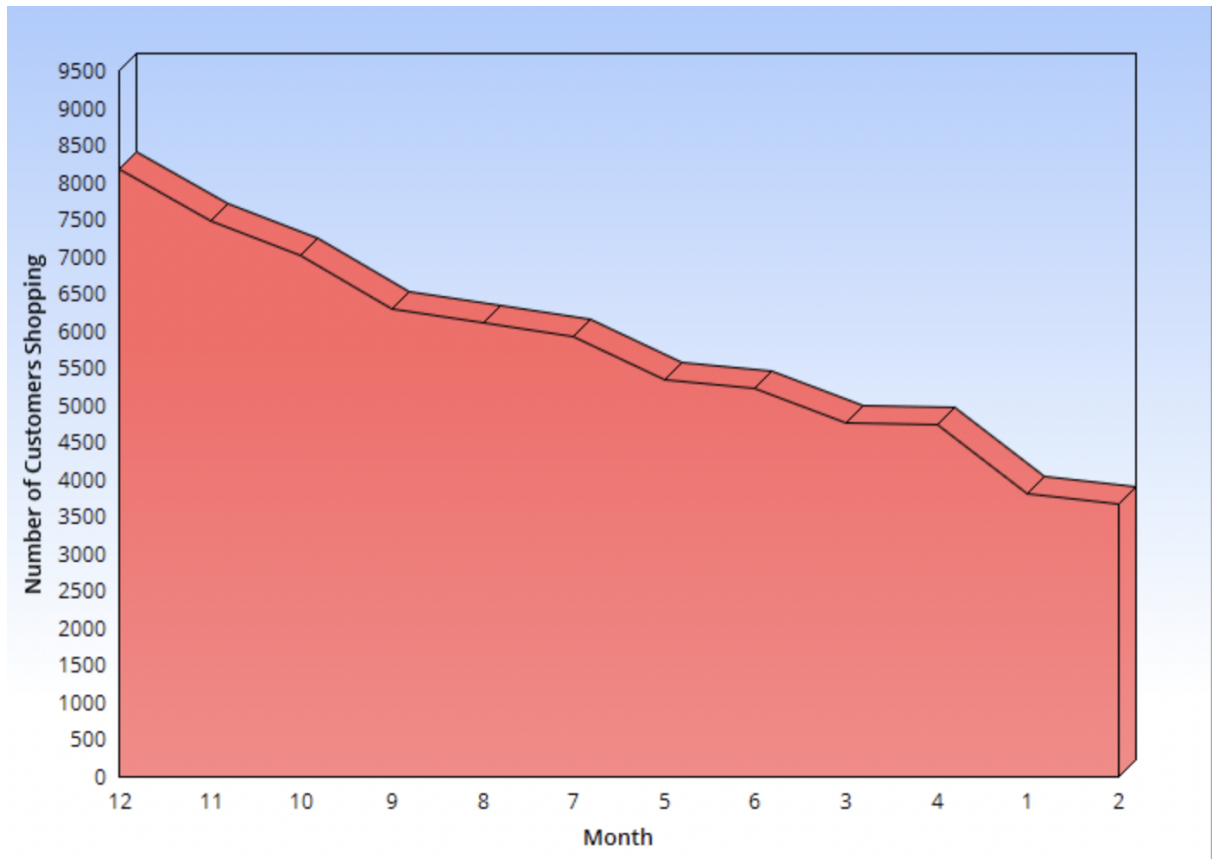
- **Shopping Time:** The peak shopping time is either at midnight, which accounts for 12.4% of the shopping, or in the afternoon between 11am to 3pm which accounts for 31.14%.



- Ratio between digital receipts and non-digital receipts is 4700:65901 which means customers generally prefer going to the regular brick and mortar. Marketing and variety of products can be increased in store. We

can also ask customers to fill a survey asking why they prefer this over online shopping and work on that further.

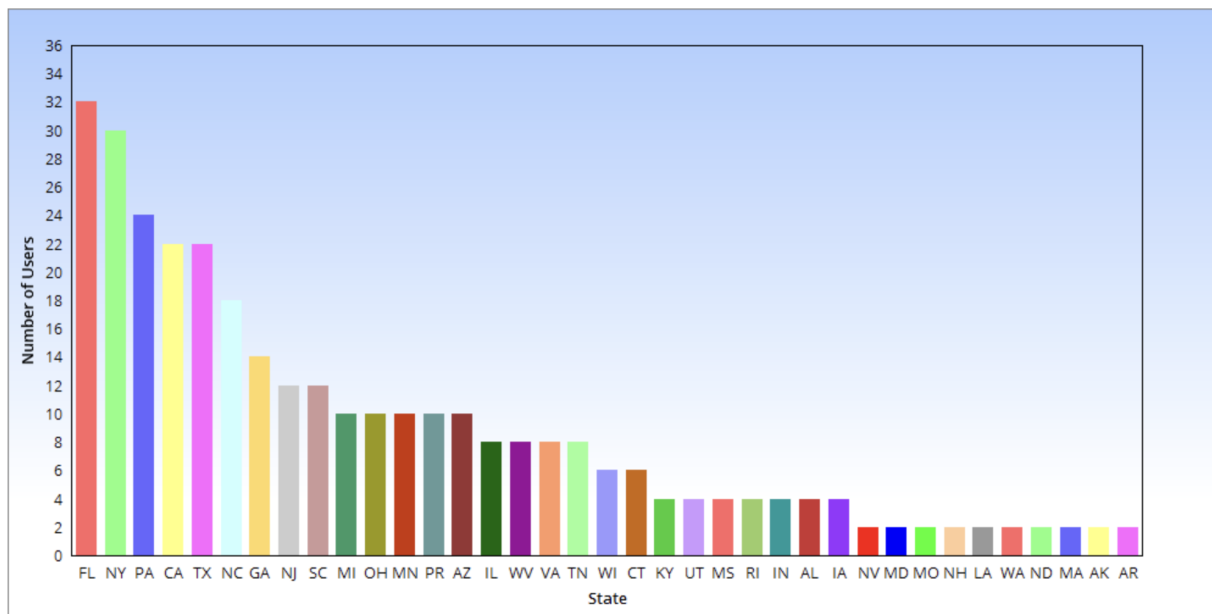
- Customers shop the most during the end of year than the start of the year.



- Most shopped categories are the following:

ABC CATEGORY ▼	123 Cnt ▼
Snacks	54
Beverages	44
Grocery	29
Frozen	29

- **Cities:** Evenly spread with 3 states contributing > 26 % of scanned receipts. Number of users state wise:



Quality of Data:

- Field in the brands table called 'related_brands' has multiple : separated values making it a task to query it.
- There are times when the receipt is too old to be scanned. In such cases we can add an entry depicting the same hence helping the stakeholders understand the reason behind failed scans.
- There are a lot of entries which are either Blank or NULL. These values are also present in receipt and receipt_items tables, which are the basis of our market analysis leading to a false depiction of the buyer trends. Since the dataset itself has such values for key fields the insights we gather from the data can be misleading. (Like in the case of purchase_time where the NULL or blank values got converted to 12AM when the time was converted to the 24 hour format for analysis).
- A check can be added while scanning receipts for these fields and only those receipts can be added in the user's account which have all necessary field data clearly visible.