Final Project

Instacart

- **Assumptions :**

I prepared a dummy dataset for date table. Apart from that I have used all the real dataset for order table, user table, product table, aisles table, department table. For each user, dataset provides between 4 and 100 orders, with the sequence of products purchased in each order. Dataset also, provides the week and hour of day the order was placed, and a relative measure of time between orders.

- **Design Process :**

The size of the dataset is 205.77MB

The dataset contains a sample of over 3 million grocery orders from more than 200,000 Instacart users.

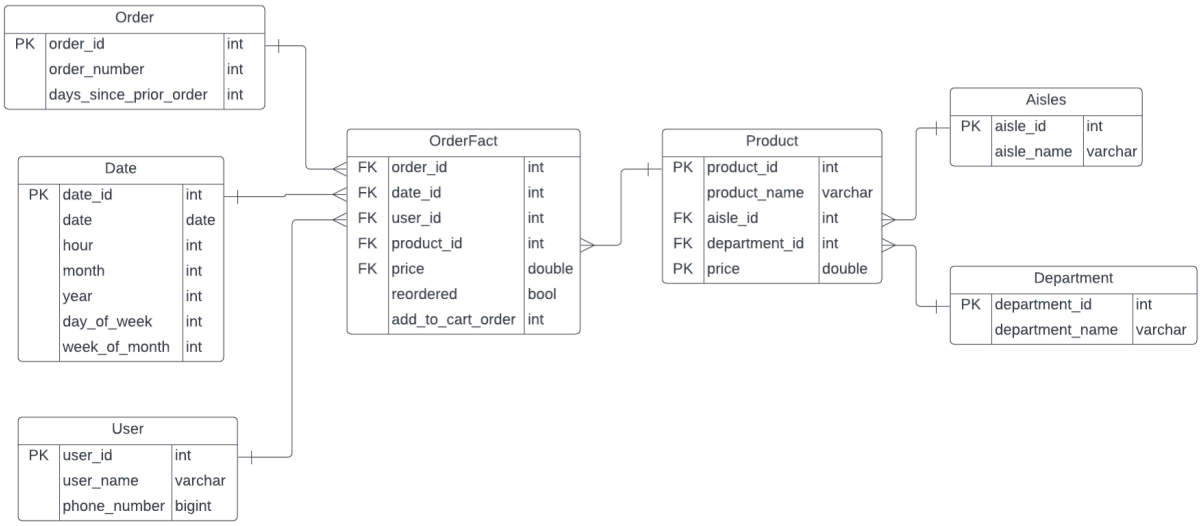
In all I have designed 6 dimension table, including a date dimension :

- Order
- User
- Product
- Aisles
- Department
- Date

There are two hierarchies of product table namely aisle and department tables.
I have also designed a fact table called OrderFact.
There are around 250,000 records in the fact table.

Column phone_number in User table is used as SCD (type 2) as it usually changes relatively slow.

Star Schema Diagram :



- **ETL Process :**

ETL is performed by using Python and Jupyter Notebook.
First, we import the Pandas library that will be used for performing ETL.
All the fields are loaded from excel sheet to a pandas dataframe.
We check the shape of the fact table.

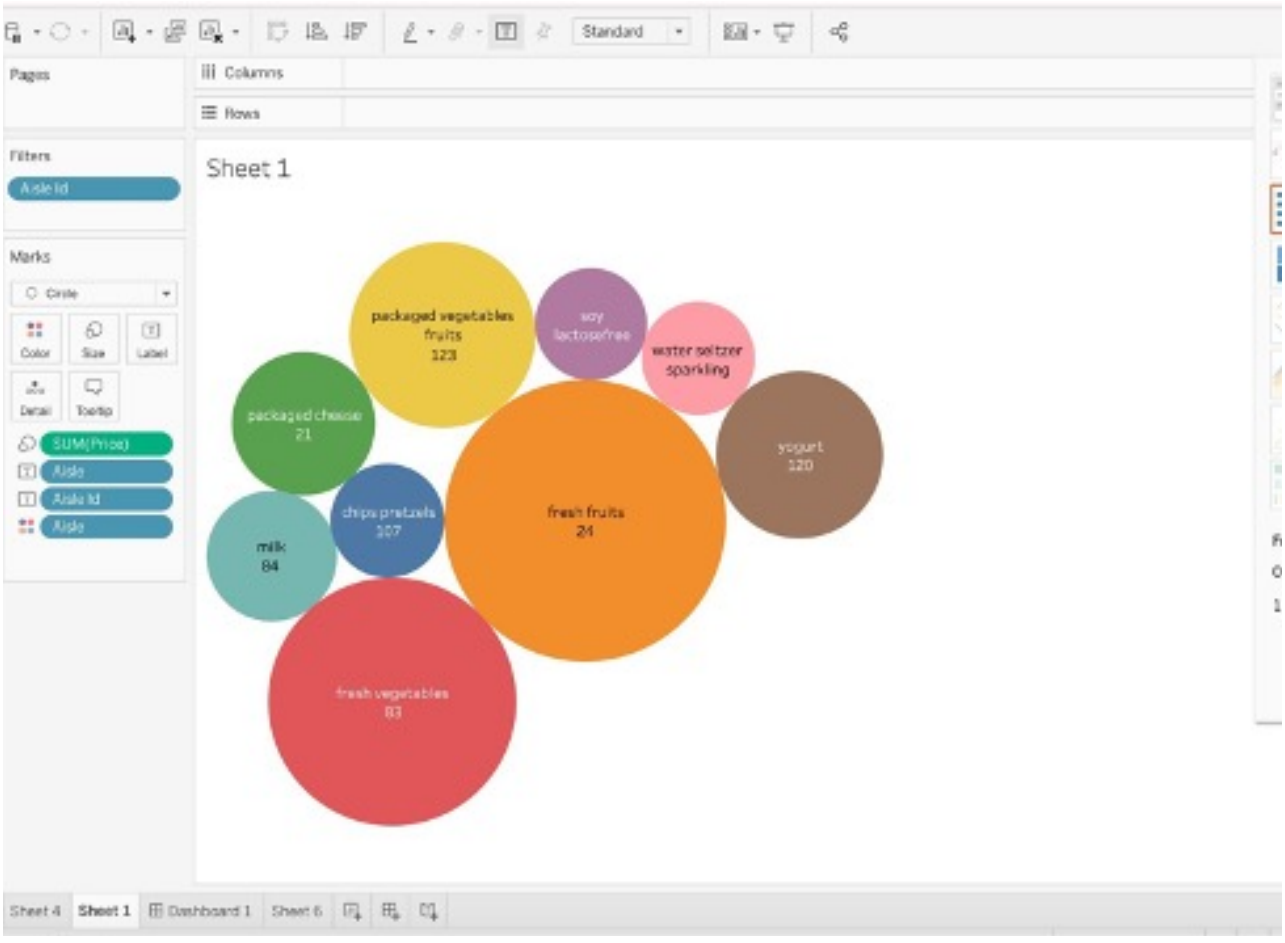
During this process I first loaded all the dataset and imported using pandas. Then I checked the size of the dataset. Dropped all the duplicates from product table. Removed all the products that didn't have product id. Joined the date and fact table to see when a particular

transaction took place and categorized it by month and year. Later dropped duplicated user table. Removed all the rows from the column price in the product table where the value is negative or null as these are anomalies. Price for any product cannot be negative.
Lastly, checked the shape for product table.

I have attached the python script file for your reference.

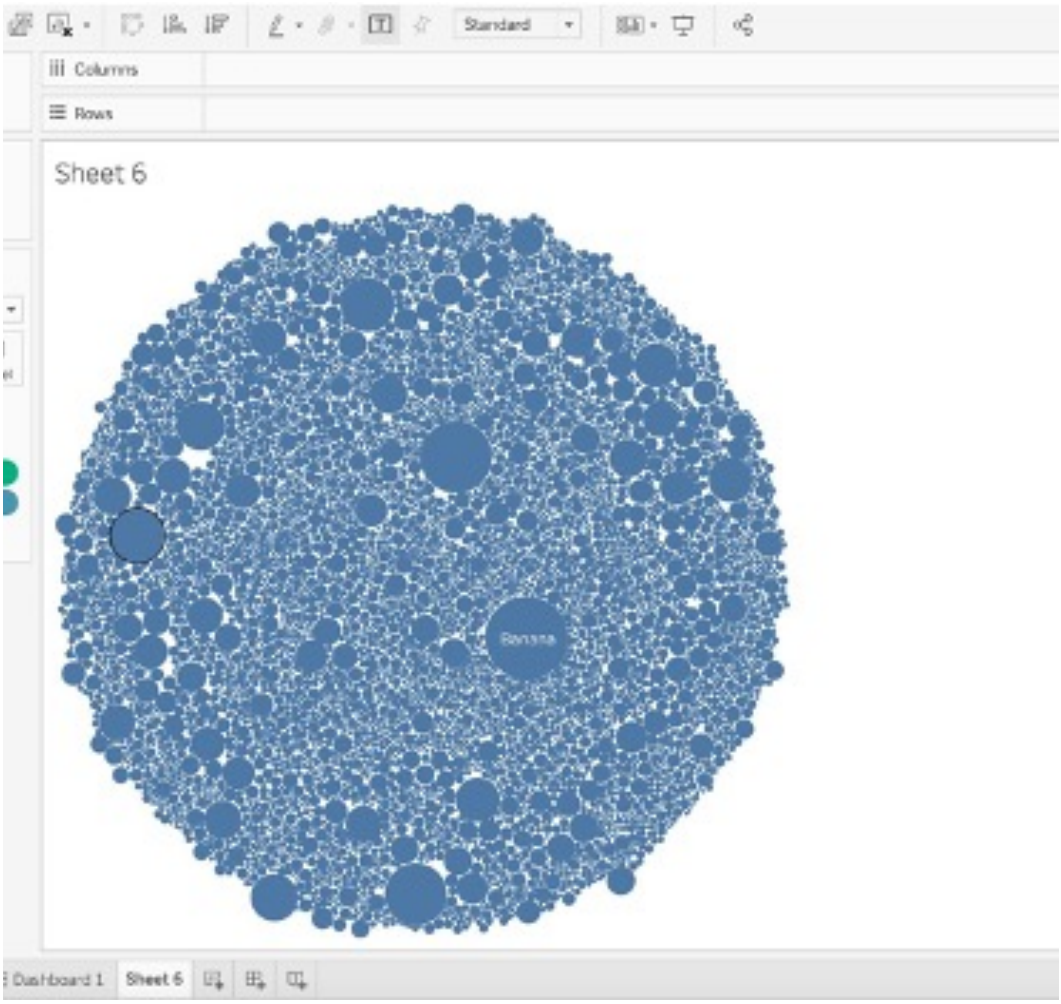
- **Visualization :**

1. Which aisle generated the most sales?



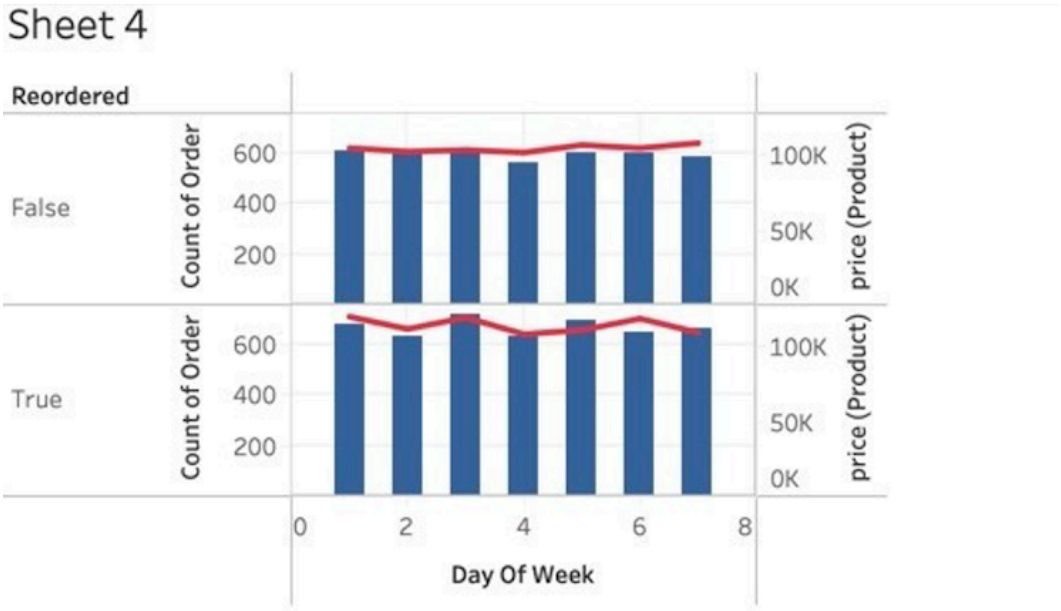
Here you can clearly see that the aisle which generated the maximum amount of sales is packed vegetable fruits which is around 123 is the highest sales of all.

2. What are the most reordered products?

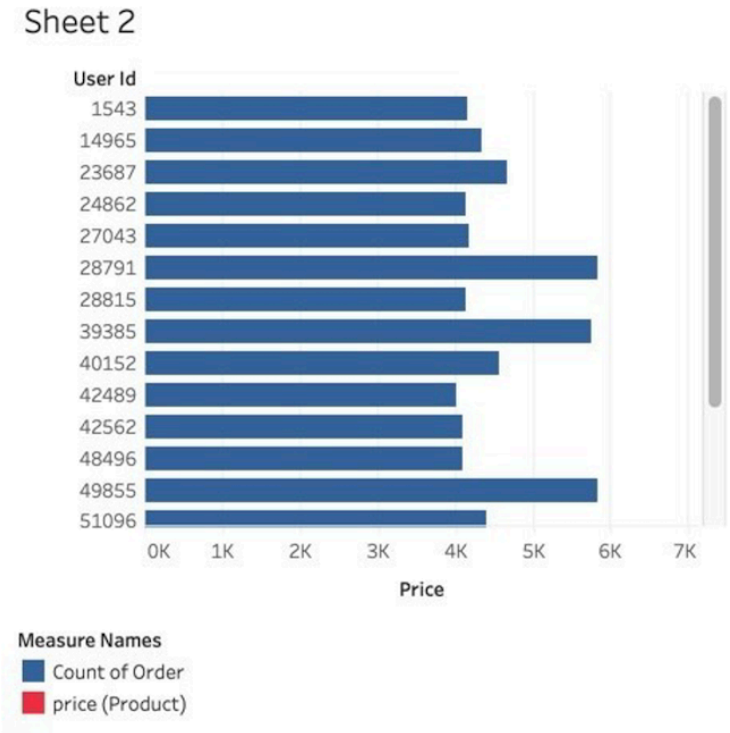


Here you can see the most reordered products are the bananas.

3. How many users made the same order using Instacart app in the same day of the week?

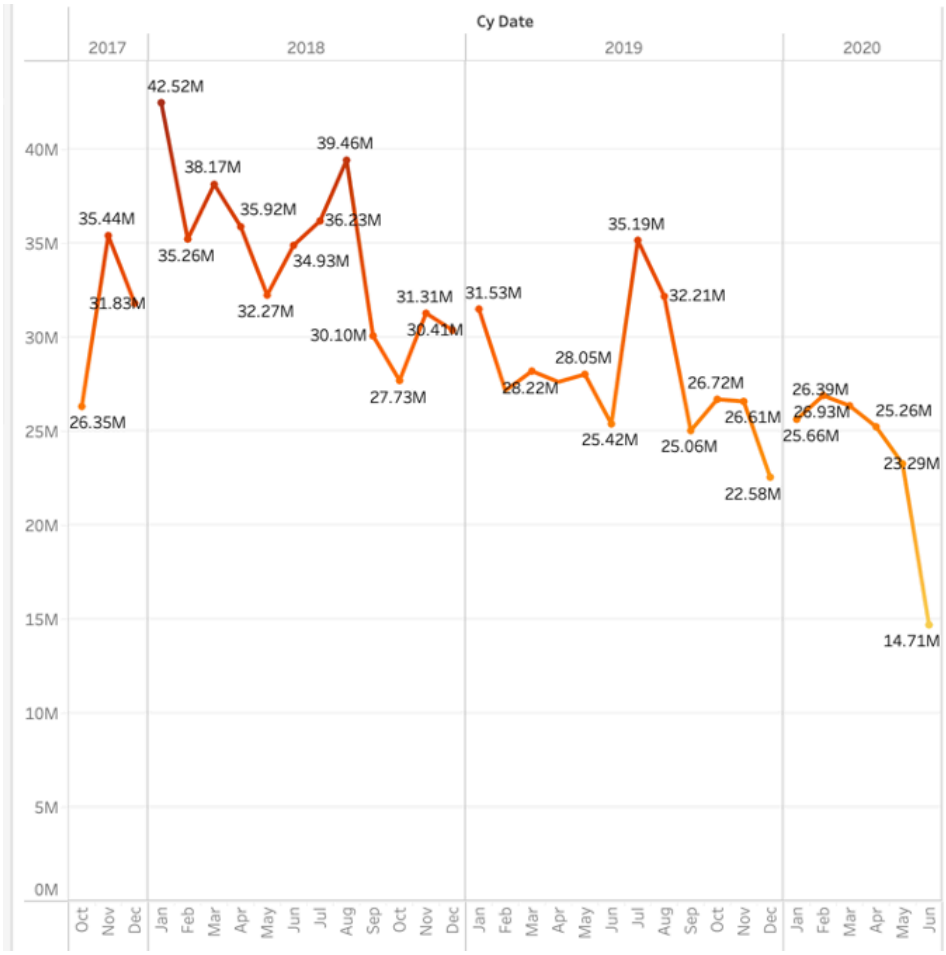


4. What is the average spending for a monthly user?



Here, it show what's the average spending of a monthly buyer which is around 4k.

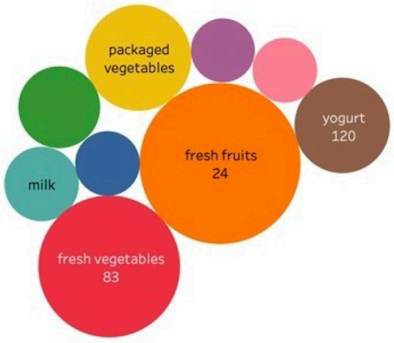
5. What is the total profit made by the end of each year?



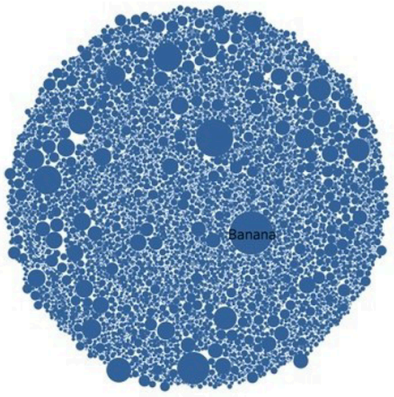
From the above visualization we can see that maximum profit was made in the beginning of the year 2018 which was around 42.52 Million

Dashboard :

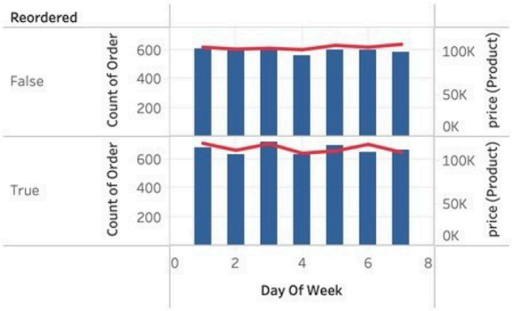
Sheet 1



Sheet 6



Sheet 4



Sheet 2