**TUGAS KELOMPOK**

**PROGRAM STUDI INDEPENDEN**

**ORBIT FUTURE ACADEMY**

**Identitas Kelompok**

Kelompok : 10

Nama Anggota : Halomoan Filipus Simarmata (Jupyter XXI)

Diana Eka Riyani (Jupyter XXI)

Nyayu Chika Marselina (Jupyter XXI)

Sukma Imelda (Cordoba)

Athiya Shinta Wulandari (Cordoba)

Coach : Ipin Sugiyarto

Program : Foundations of AI and Life Skills for Gen-Z

Hari, Tanggal : Jumat, 1 April 2022

**Tugas:** Membuat model analisa market basket dengan dataset berbeda (dari beberapa model transaksi silahkan kalian pilih) proses sama seperti coding latihan. Buat kesimpulan barang apa yang cocok disandingkan berdasarkan kekuatan korelasi (lift). Laporannya screenshoot coding & hasil serta penjelasan dari masing masing step.

**Penyelesaian:**

**Groceries Market Basket Analysis**

Market basket analysis adalah suatu metodologi untuk melakukan analisis buying habit konsumen dengan menemukan asosiasi antar beberapa item yang berbeda, yang diletakkan konsumen dalam shopping basket (keranjang belanja) yang dibeli pada suatu transaksi tertentu. Tujuan dari market basket analysis adalah untuk mengetahui produk-produk mana yang mungkin akan dibeli secara bersamaan.

Algoritma Apriori adalah suatu algoritma dasar yang diusulkan oleh Agrawal & Srikant pada tahun 1994 untuk penentuan frequent itemsets untuk aturan asosiasi boolean. Algoritma Apriori memberi kita sifat asosiatif dalam transaksi. Ini juga dikenal sebagai Aturan Asosiasi. Aturan asosiasi atau association rule adalah teknik untuk menemukan aturan asosiasi antara suatu kombinasi item. Terdapat 3 metrik untuk mengukur ketepatan aturan, yaitu:

1. Support

Support adalah indikasi seberapa sering kumpulan item muncul pada dataset. Berikut perhitungan nilai support:

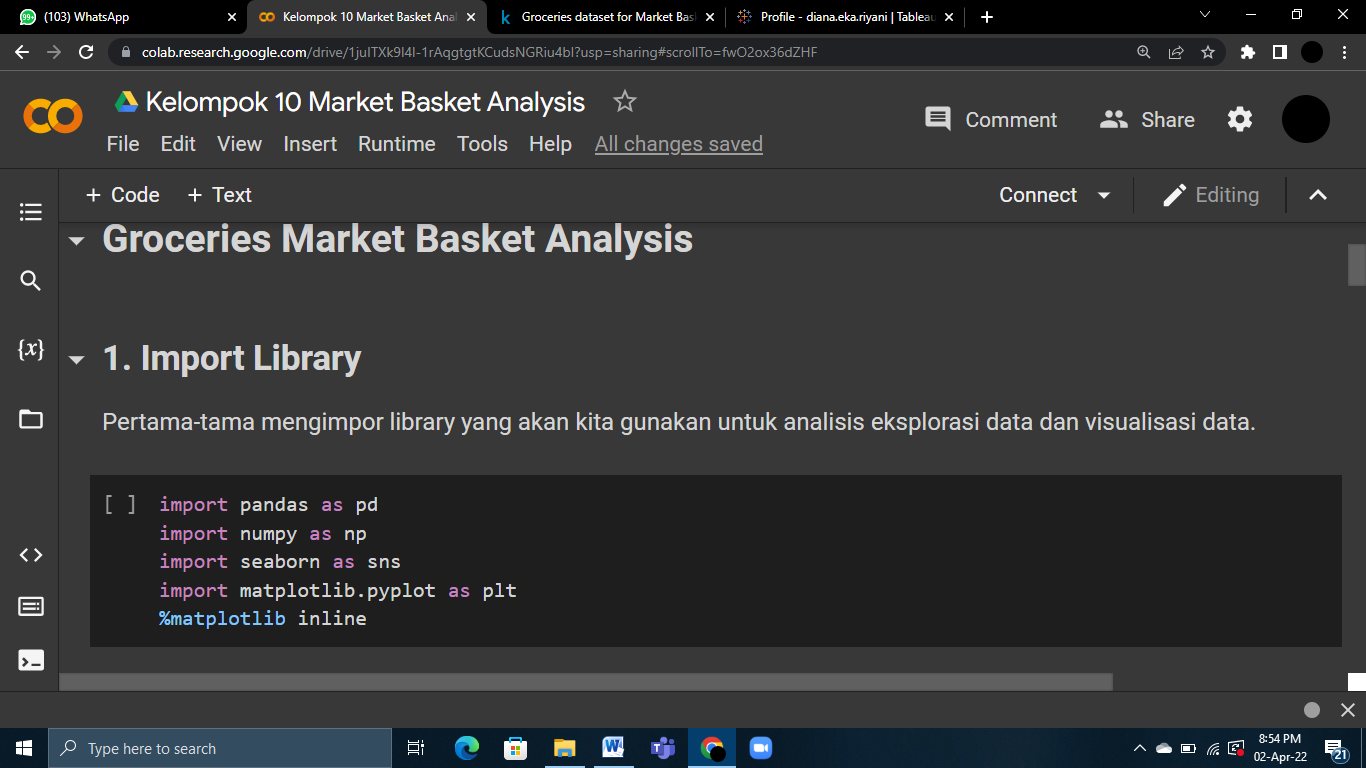
1. Confidence

Confidence adalah suatu ukuran yang menunjukkan hubungan antar dua item secara conditional (berdasarkan suatu kondisi tertentu). Berikut perhitungan nilai confidence:

1. Lift

Lift mengacu pada bagaimana peluang kedua item dibeli ketika item pertama dibeli. Berikut perhitungan nilai confidence:

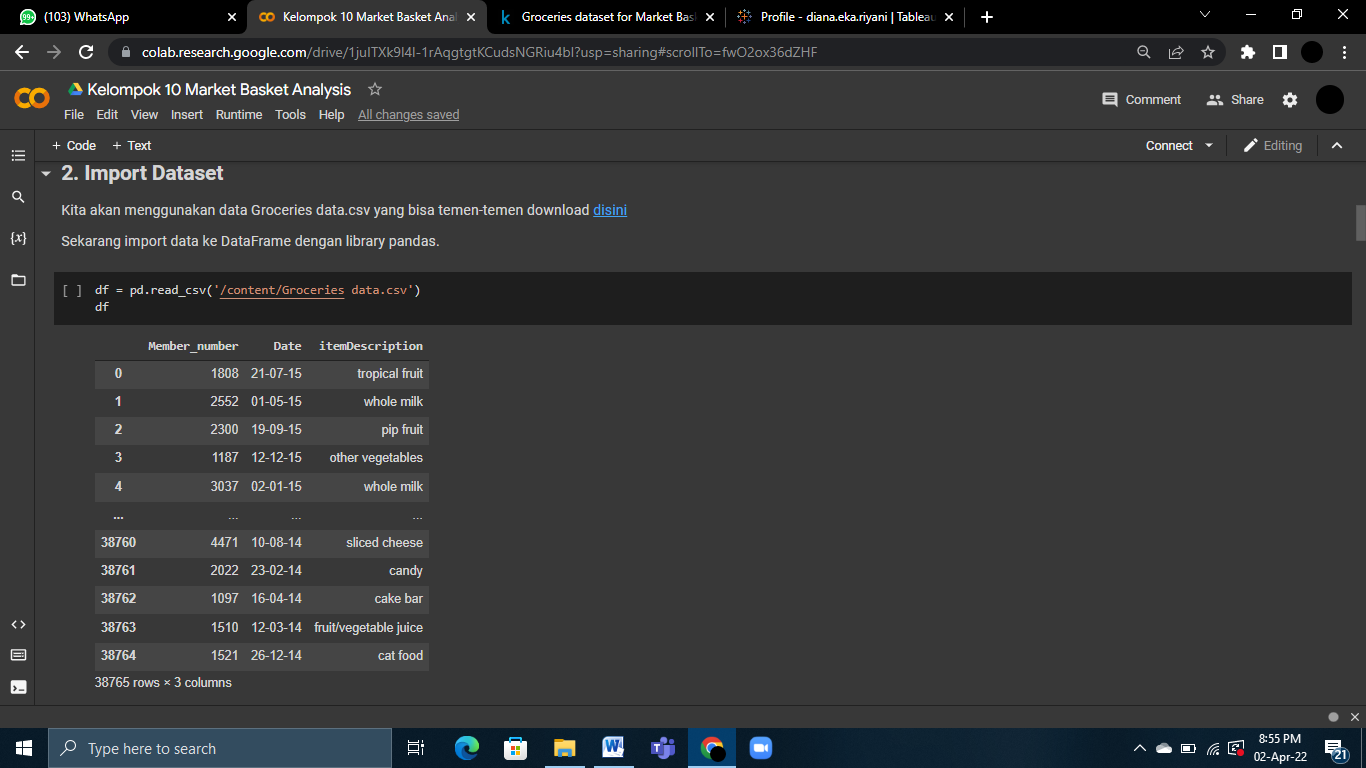
1. **Import Library**



1. **Import Dataset**

Kita akan menggunakan data Groceries data.csv yang bisa temen-temen download di <https://www.kaggle.com/datasets/rashikrahmanpritom/groceries-dataset-for-market-basket-analysismba?select=Groceries+data.csv>

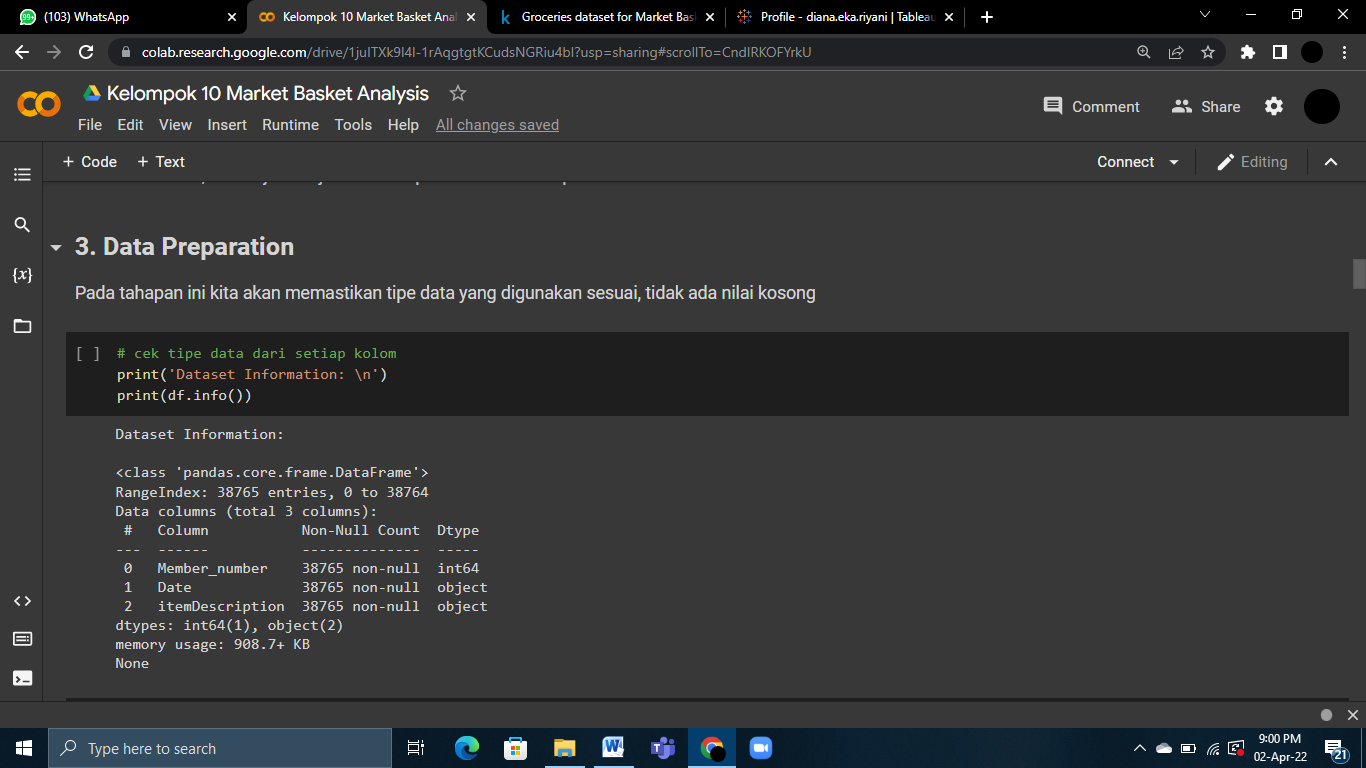
Sekarang import data ke DataFrame dengan library pandas.

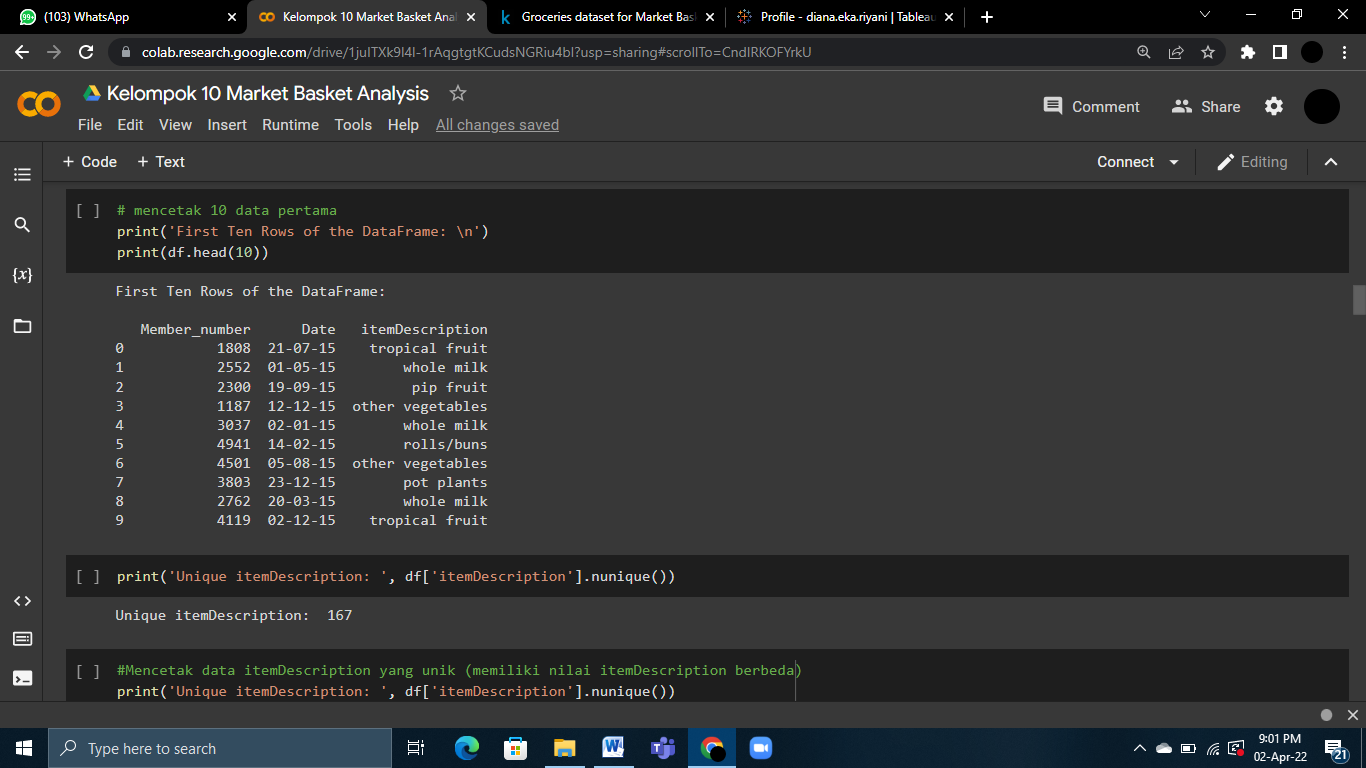


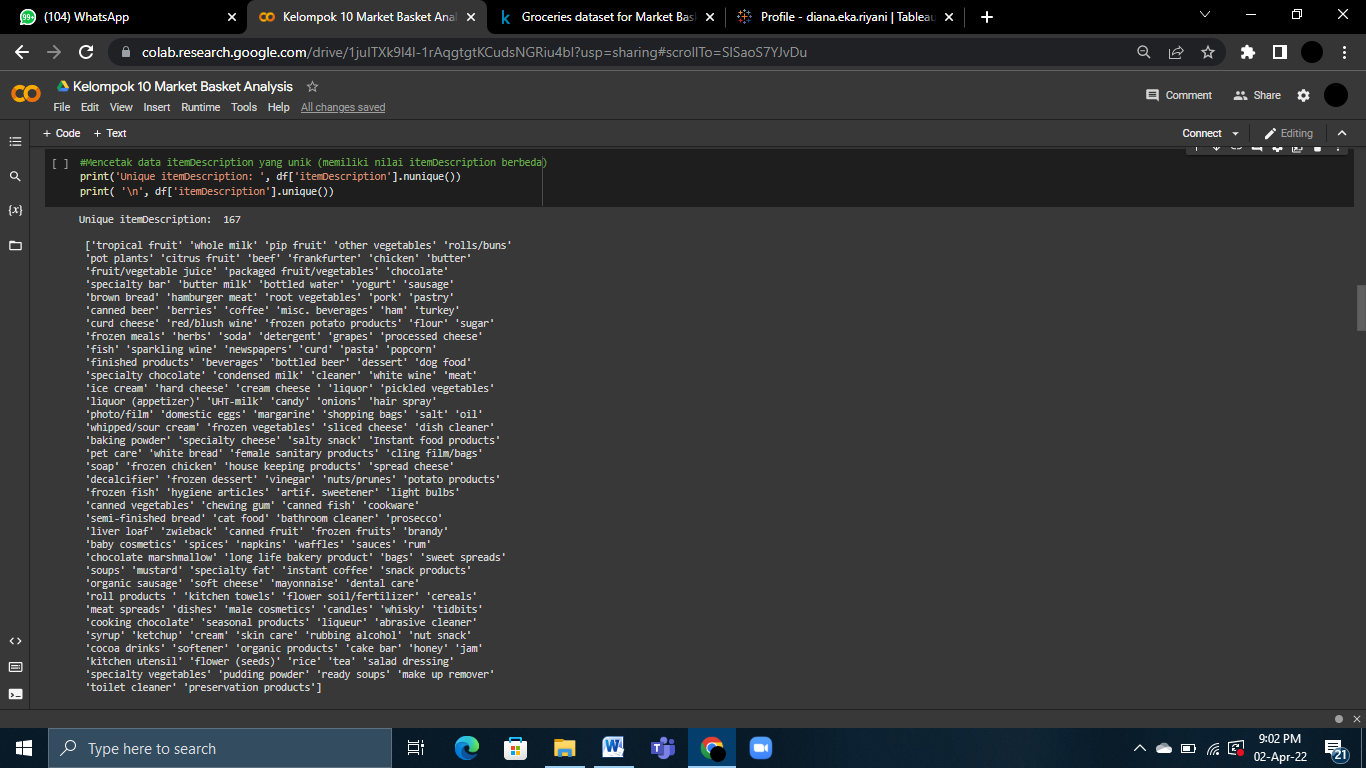
Dari tabel diatas, menunjukkan jumlah data pada dataset terdapat 38765 baris dan 3 kolom.

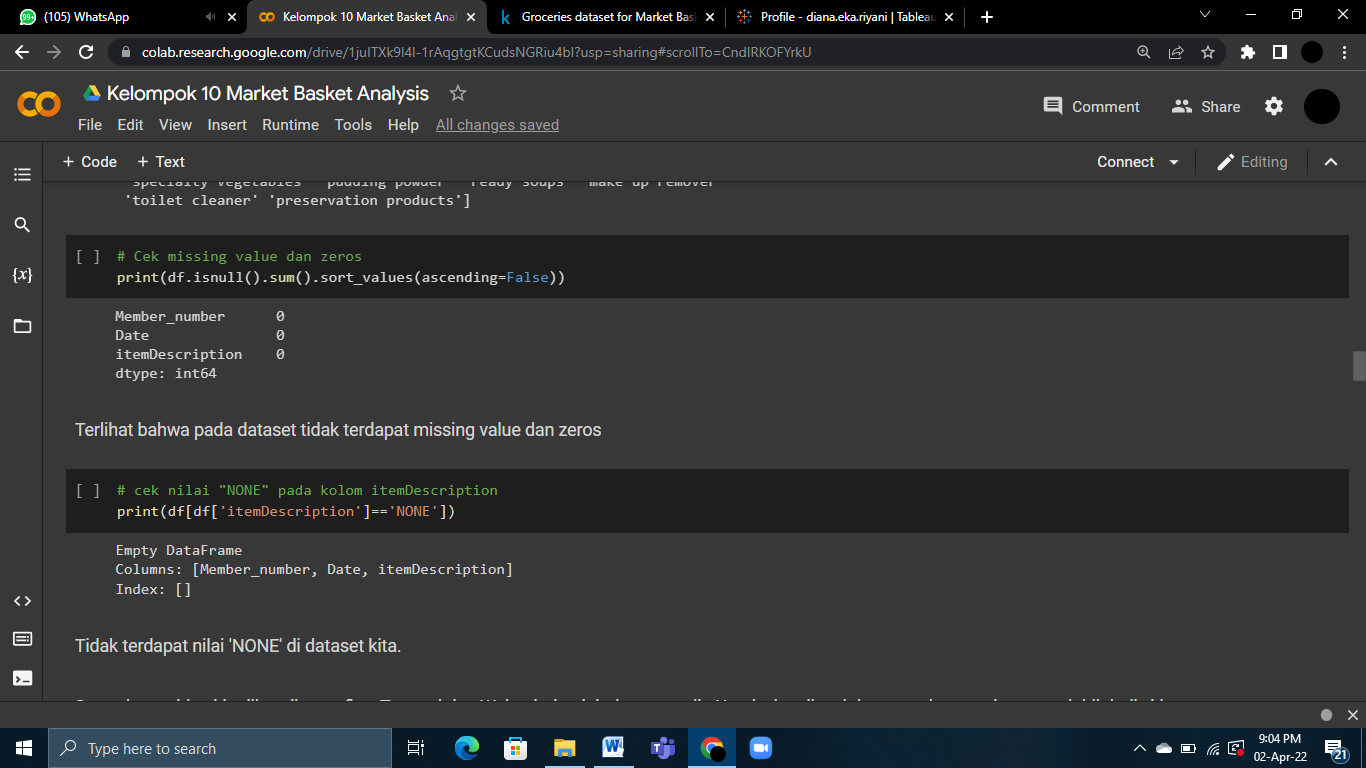
1. **Data Preparation**

Pada tahapan ini kita akan memastikan tipe data yang digunakan sesuai, tidak ada nilai kosong.

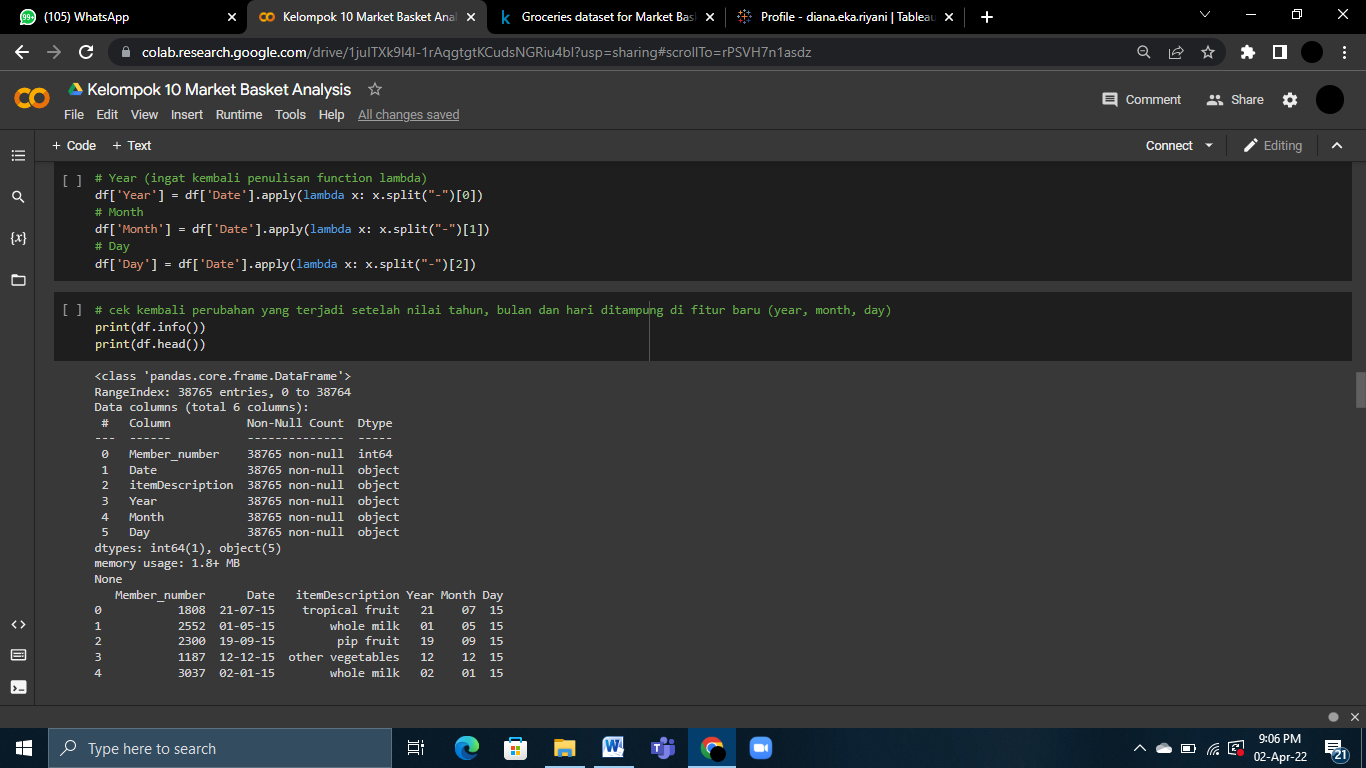






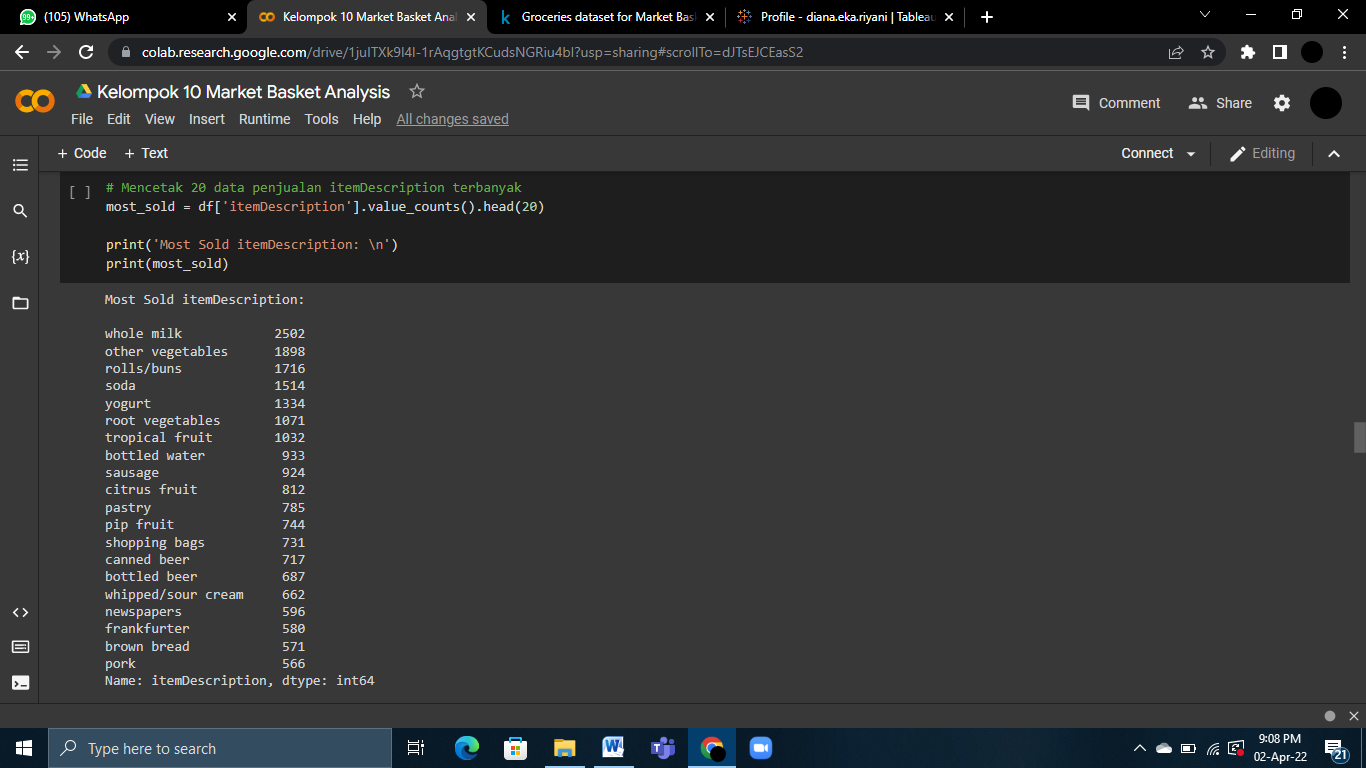


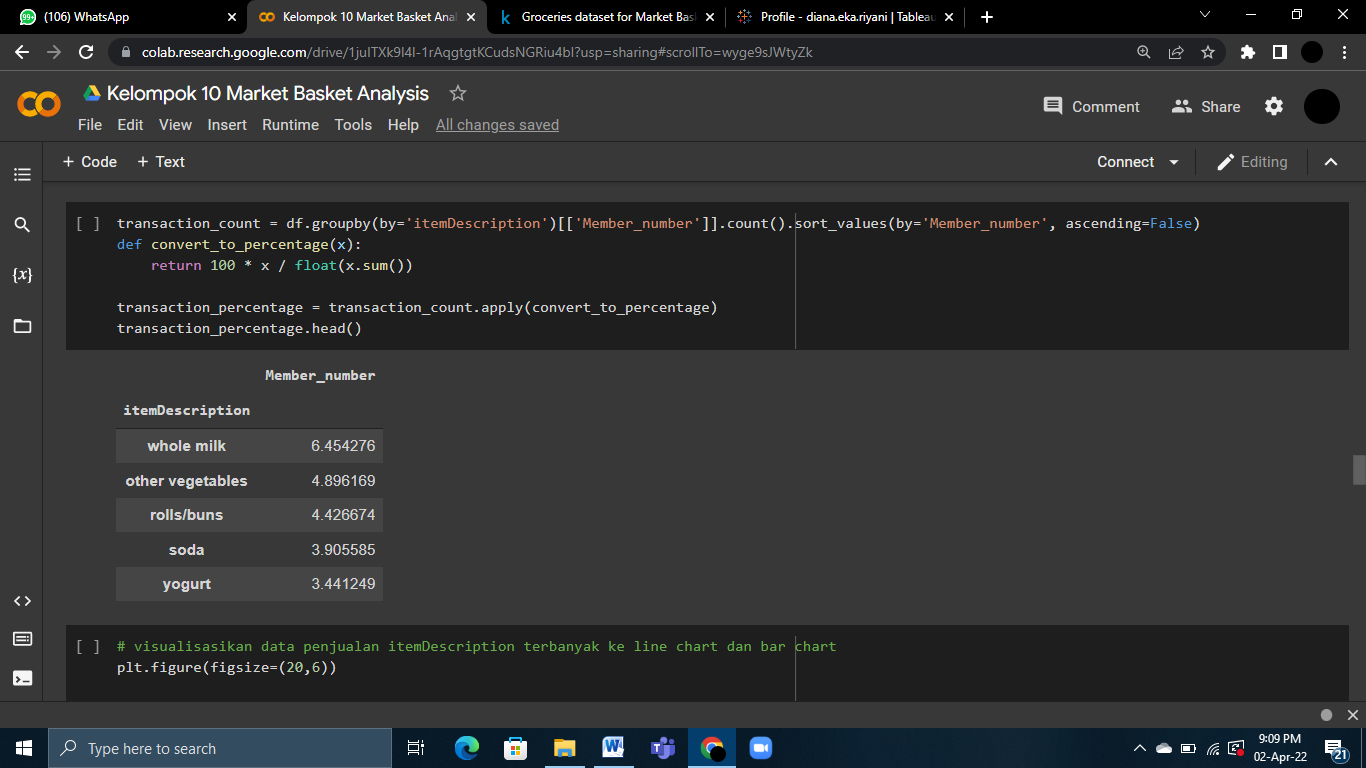
Seperti yang bisa kita lihat di atas, fitur Tanggal dan Waktu bukanlah tipe numerik. Untuk visualisasi dan pemahaman data yang lebih baik, kita bisa menambahkan beberapa fitur lagi ke DataFrame ini berdasarkan informasi dari dua fitur ini.

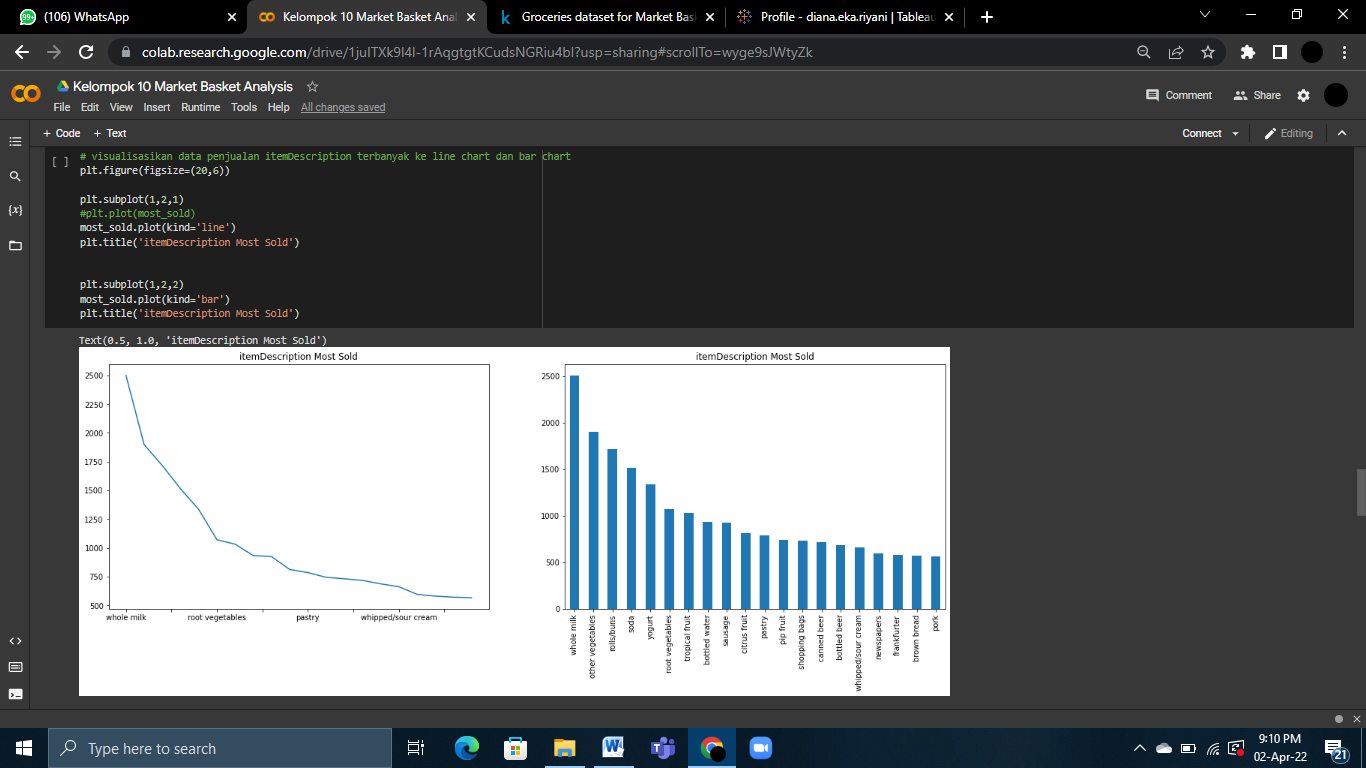


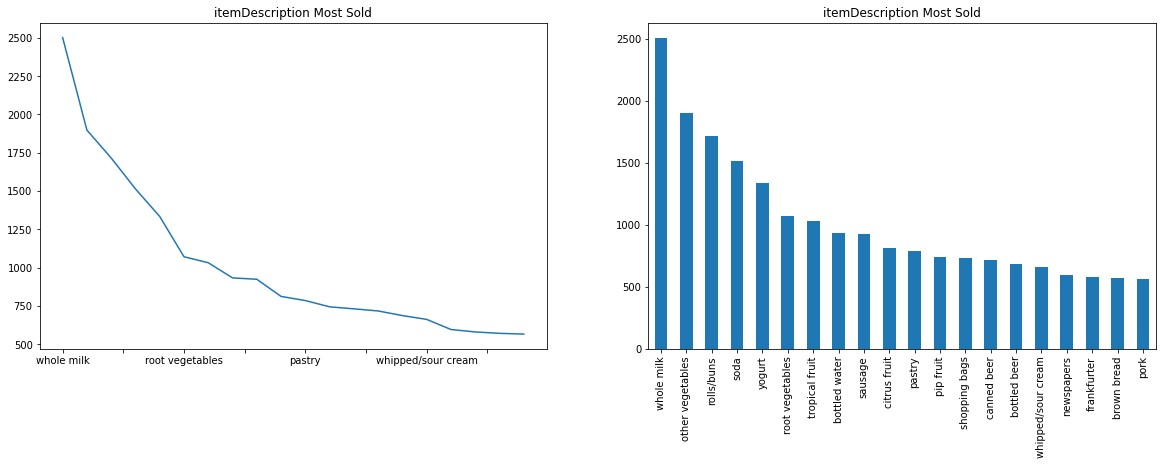
1. **Visualisasi dan Memahami Data**

Kita tahu bahwa dataset ini direkam dari 01/01/2014 hingga 30/12/2015. Sebelum kita masuk dalam pemodelan, kita harus mengeksplorasi dan memvisualisasikan penjualan dalam periode waktu ini. Bahan makanan apa yang paling banyak dibeli pelanggan? Bulan mana yang lebih sukses? Mari kita jawab ini secara visual.

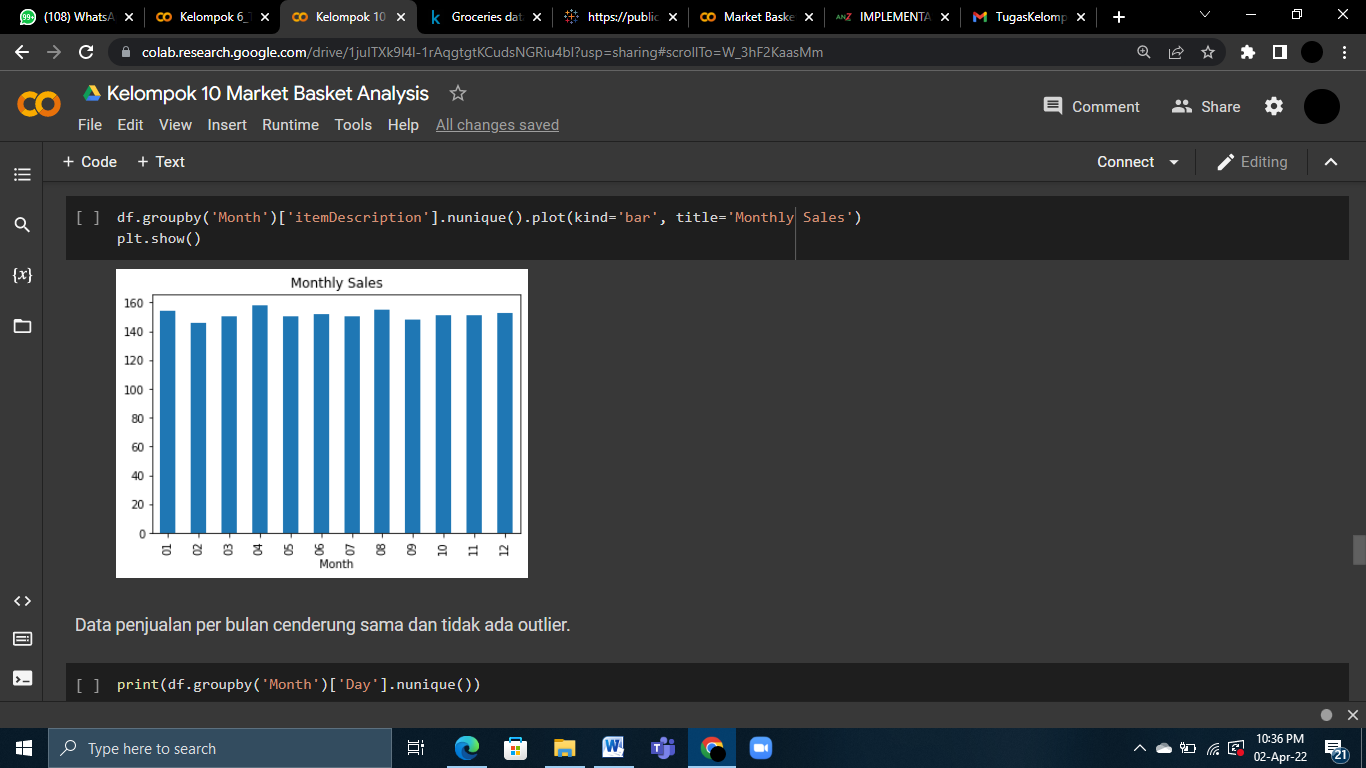


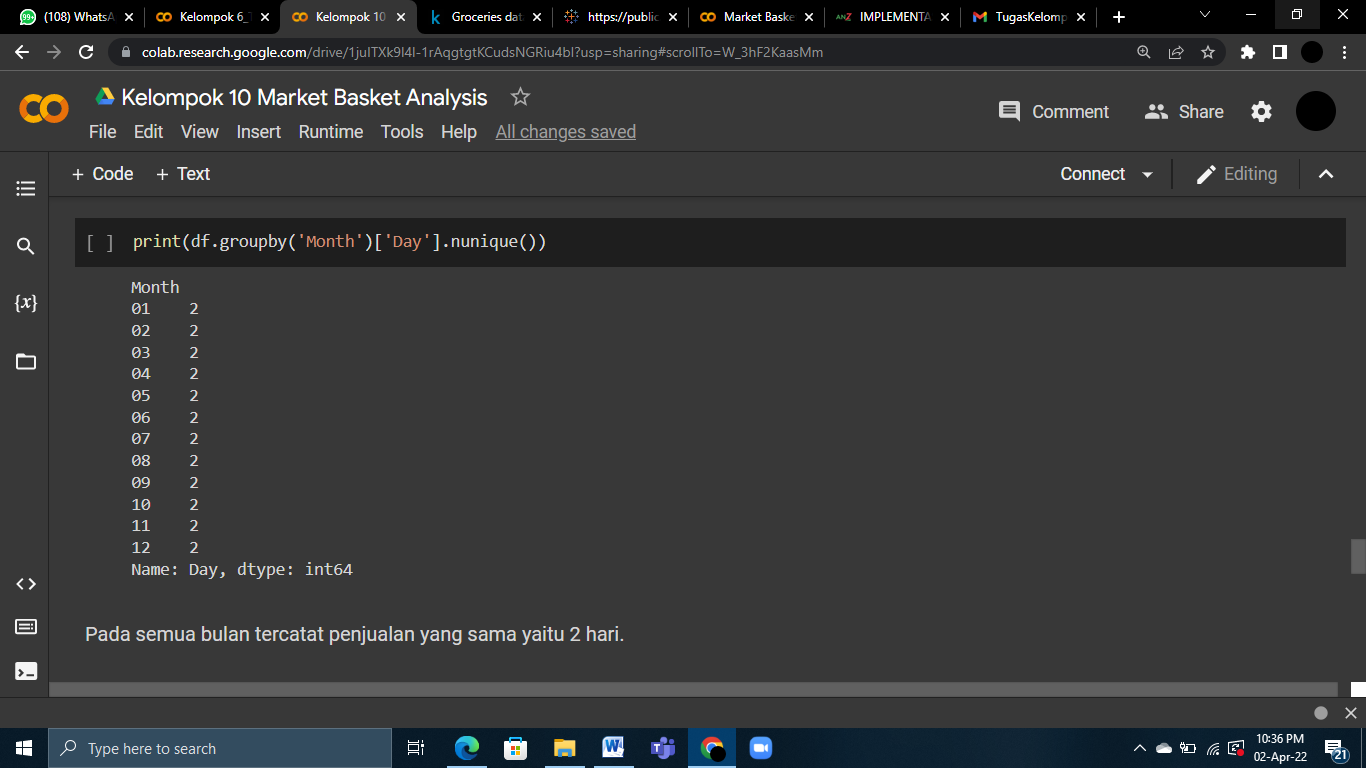




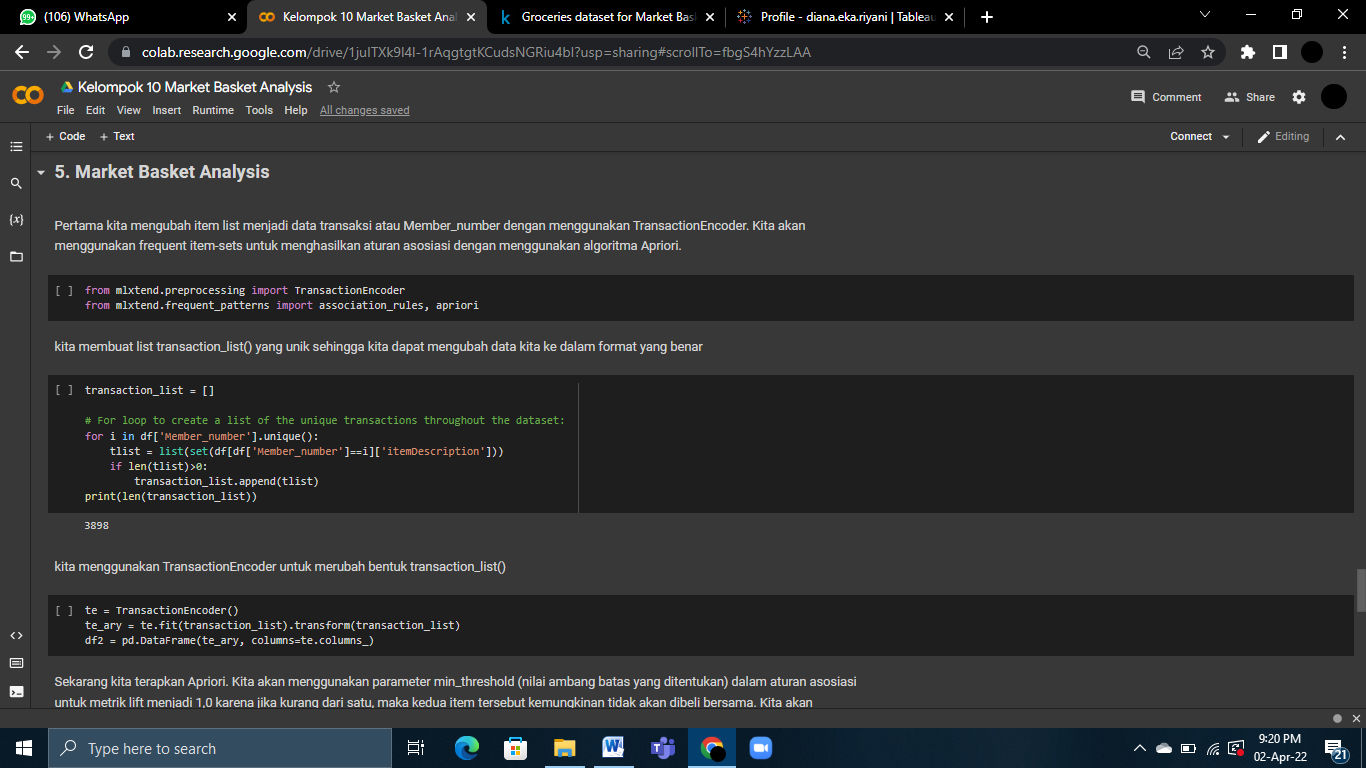


Berdasarkan hasil visualisasi, whole milk adalah bahan makanan yang paling banyak terjual, diikuti oleh other vegetables, rolls/buns, soda, dan yogurt. Ini masuk akal untuk toko groceries. Sekarang setelah kita mengetahui item mana yang paling populer, mari kita lihat bulan mana yang menghasilkan penjualan paling banyak.





1. **Market Basket Analysis**



Sekarang kita terapkan Apriori. Kita akan menggunakan parameter min\_threshold (nilai ambang batas yang ditentukan) dalam aturan asosiasi untuk metrik lift menjadi 1,0 karena jika kurang dari satu, maka kedua item tersebut kemungkinan tidak akan dibeli bersama. Kita akan mengurutkan nilai berdasarkan keyakinan untuk melihat kemungkinan suatu bahan makanan dibeli jika pendahulunya dibeli.



**Kesimpulan**

Dapat dilihat dari analisis di atas, di mana semakin tinggi nilai peningkatan, semakin kuat korelasi antar item. Data dengan jelas menunjukkan bahwa whole milk adalah item paling populer. Mari kita lihat korelasi item yang lebih menarik (format: ):

Bisnis selalu mencari cara untuk mengoptimalkan pengaturan mereka dan meningkatkan penjualan mereka. Karena kita sekarang mengetahui korelasi antara itemDescription dan kepentingan bersama pelanggan, bisnis dapat membuat keputusan berdasarkan temuan ini. Misalnya, toko groceries mungkin tertarik untuk mengadakan promosi itemDescription gratis, mengingat kemungkinan besar item lain dijual sebagai hasilnya (misalkan jika mereka mengadakan promo buy 1 get 1 khusus special event, itu mungkin tidak hanya menarik pelanggan baru yang sering datang, tetapi juga ada peluang yang sangat bagus bahwa pelanggan masih akan menghabiskan uang untuk whole milk).