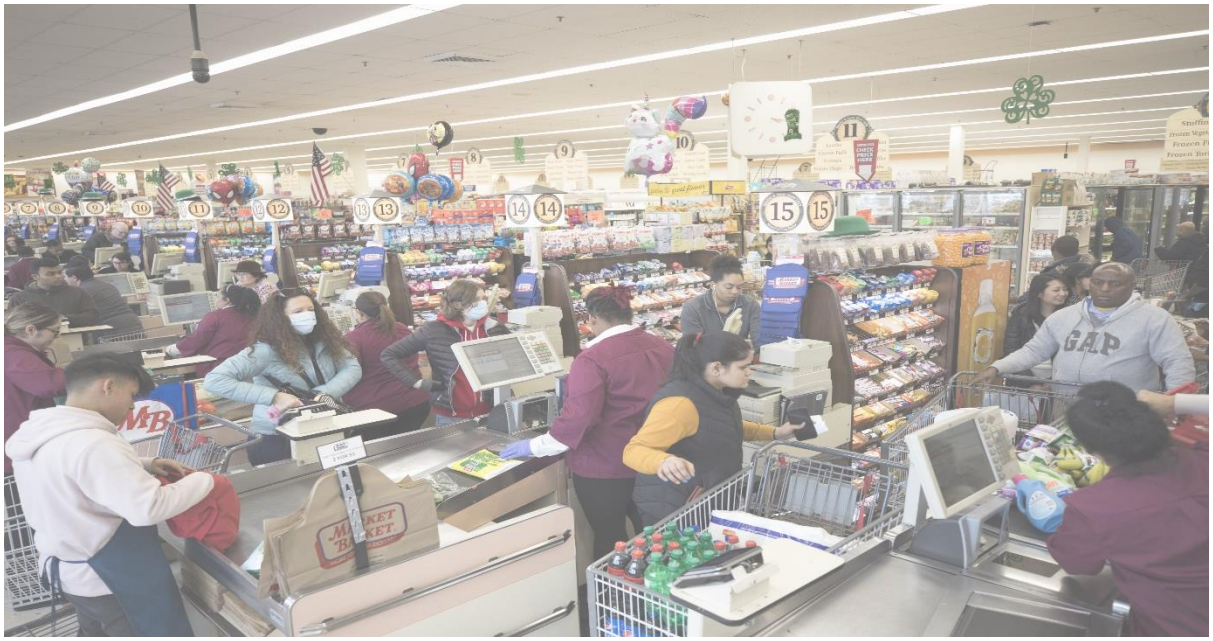
Case Study -:

Gartner receives transaction-level scanning data (**POS**) data from its partner stores regularly. Stores sharing POS Data include bigger format store types such as supermarkets, hypermarkets as well as smaller traditional trade grocery stores (Kirana Stores), medical stores, etc using a POS machine.

While in a bigger format store, all items for all transactions are scanned using a POS machine, smaller and more localized shops do not have a 100% compliance rate in terms of scanning and inputting information into the POS machine for its all transactions.

A transaction involving a single packet of chips or a single piece of candy may not be scanned and recorded to spare customers the inconvenience or during rush hours when the store is crowded with customers.



The picture is during the COVID -19

Thus, the data received from such stores are often incomplete and lacks complete information of all transaction completed within a day.

Additionally, apart from an incomplete transaction date in a day. It is observed that certain stores do not share data for all active days. Stores share data ranging from 2 to 28 days in a month. While it is possible to impute/extrapolate data for 2 days of a month using 28 days of actual historical data, the vice versa is not recommended.

Today a blanket call is taken to include or not to include the data for a store given its compliance rate and data quality. Inactivity for a couple of hours currently disqualifies the store for the whole day. These limits are effective, it leads to a high wastage of available data.

Gartner expects you to bring out some insights form the data which can be crucial for the business to take relevant decisions. The insights can be both operational and strategic also.

About Datasets

In this hackathon, you are provided with the dataset that contains store-level data by brands and categories for selected stores.

Dataset Description:

Dataset_Ideal_data – The file contains brand-level data for 10 stores for the last 3 months. This can be referred to as the ideal data.

Dataset_Working_data – This contains data for selected stores that are missing and/or incomplete.

Dataset_mapping_file – This file is provided to help understand the column names in the dataset.