Instacart Transaction Matching R package

Purpose: Find equivalent transaction in TSMC POS system for Instacart orders.

Requirements:

* R 3.6 or later
* R1010 package installed

Dependencies:

* R1010, Dplyr, data.table

**‘R1010’ package installation:**

The easiest way to install it is by running the following R commands:

1. download.file("https://www.1010data.com/downloads/R1010/win/R1010\_win.zip", method="libcurl", destfile="R1010\_win.zip")
2. unzip("R1010\_win.zip"
3. unzip("R1010\_**\***\_win.zip") # Change \* for the version you want to install e.g. unzip("R1010\_**4.0.0**\_win.zip")
4. install.packages('R1010\_**\***\_win/R1010.zip', type="win", repos=NULL) # Change \* for the version you want to install e.g. unzip("R1010\_**4.0.0**\_win.zip")

Once you get R1010 package up and running, it’s time to install ‘InstaMatch’ package developed by the Strategic Analytics team.

**‘InstaMatch’ package installation:**

1. devtools::install\_github("savemart-strategy/InstaMatch")
2. library(InstaMatch)

Steps:

1. Export Instacart transaction file (preferably a maximum of two days data)
2. Run R script to call the function ‘match\_orders’ passing file as an argument and 1010 credentials
3. Handle new file returned by the R script function
4. Save file as “.csv”, “.xlsx” etc.

Pseudo-code:

1. Receive Instacart data.
2. Download 1010 data with dates ranging from 4 days before the first transaction date in the Instacart file and 7 days after the last transaction in the Instacart file.
3. Find the 1010 transaction with the most matches on UPCs, UPC count, order price, for every single Instacart order.
4. For transactions with poor matching coefficient, expand the lookup to include all stores within the same city.
5. Return dataframe containing the corresponding transId from 1010 to every order on Instacart file.

Challenges:

1. Instacart transaction date may not match exactly what is recorded on the POS
2. Instacart location may not match exactly what is recorded on the POS.
3. Items may vary in the transaction from what was ordered versus what was rung on the POS.
4. UPCs are in different format.
5. Item Description, Dept, and Category may not match.