

CPG Data Description

in progress

Owen Eagen, Yifan Xu

2018/11/25

Contents

1	Replicating Previous Data Cleaning Process	1
2	Exploring Files in the “externals” Directory	1
2.1	dff.zip	1
2.2	IRI.zip	1
2.3	OMNI.zip	2

This file documents observations made during the data cleaning process for the CPG project.

1 Replicating Previous Data Cleaning Process

This section deals with the process and results of replicating the cleaning code provided by Professor Joo in the RA2/data_cleaning directory.

2 Exploring Files in the “externals” Directory

This section stores our findings when exploring the files in the “externals” directory. It mainly includes our interpretation of the variable names and the comparison we make between the private and public data.

2.1 dff.zip

DFF, Dominick’s Finer Foods, was a leading supermarket chain in Chicago.

2.2 IRI.zip

2.2.1 A bit about IRI

IRI, Information Resources Inc, is a market research company founded in Chicago in 1979, and was acquired by [Symphony Technology Group](#) in 2003.¹ IRI developed Apollo system, which provides desktop-based solutions that cover category management process, including assortment management and on-shelf planogram (IRI is mentioned on page 304 in Dreze’s paper).

2.2.2 Files in IRI.zip

There are in total 28 files in the folder. Names of all files have 6 characters, starting with “acv.” The last 3 characters are identifiers for the type of good documented. A map of the identifiers to the product names is stored in `externals/POG/doc/category_desc.txt`.

¹Source: [Wikipedia](#)

2.2.3 Variable Names

```
[1] "UPC"      "WEEK"      "DACVFD"    "DACVF"     "DACVD"     "DACVP"     "DINCREM"   "DACVFA"
[9] "DACVFB"    "DKEYCAT"   "MINCREM"   "DCINCREM"   "MCINCREM"   "DBVOL"     "DBP"       "DBF"
[17] "DBFD"      "DVNP"      "DBVNP"     "DBD"        "BD"         "DAVGFD"    "DAVGF"     "DAVGD"
[25] "DAVGP"     "DWTAVG"
```

We have not yet figured out the exact meanings of the variables yet. But we grouped them according to observed patterns and made some guesses based on the patterns.

2.2.3.1 Grouping of variable names

- DACVF, DACVP, DACVD
 - ACVFB, DACVFD, DACVFA
- DINCREM, MINCREAM, DCINCREM, MCINCREM
- DBVOL, DBP, DBF, DBD, DBFD
- DVNP, DBVNP
- BD
- DAVGFD, DAVGF, DAVGD, DAVGP, DWTAVG

2.2.3.2 Guesses

- D-: Daily?
- -AC-: Accumulated? Average cost?
- -ACV-:
- V: Volume
- -AVG-: Average
- WTAVG: Weighted average
- -P: Price
- -BP: Bundle price?
- -D:
- -F:
- -INCREM: Increment
- -C-: Cumulative? Cost?
- M-: Monthly

2.3 OMNI.zip

[Omni Resources](#) is a technology consulting firm found in 1984.