

# Data Exploration

Tanetpong

06/10/2021

```
df <- read.csv("../data/df_scanner.csv")
```

Count by category

```
print(table(df$category))
```

```
##
##           Bakery
##           8152
##      Baking ingredients
##           16804
##           Bath Products
##           38113
##           Berries
##           7747
##           Biscuits
##           10897
##      Bottled water
##           13590
##           Bread
##           72148
##           Butter
##           17445
##           Canned soup
##           7326
##      Canned vegetables
##           35896
##           Cereals
##           22188
##           Cheese
##           51149
##           Chicken
##           15315
##      Chips & Salty snacks
##           34604
##           Chocolate
##           23536
##      Cleaning products
##           14235
##           Coffee
##           21342
```

##	Coffee filters
##	2197
##	Cookies
##	20648
##	Cream
##	34705
##	Diet pills
##	3059
##	Dishwasher & washing up
##	8127
##	Eggs
##	4156
##	Fish
##	38025
##	Frozen Pizza Snacks
##	8238
##	Fruit Juice
##	40491
##	Fruits
##	18599
##	Ice cream
##	25602
##	Jam
##	21395
##	LactoseFree
##	31275
##	Laundry
##	19030
##	Meat
##	22922
##	Milk
##	8806
##	Oils
##	8060
##	Pasta
##	17722
##	Pasta Sauce
##	5326
##	Pastry
##	11877
##	Raw nuts, dried fruits & granola bars
##	12630
##	Ready-made meals
##	45091
##	Rice
##	7008
##	Root vegetables
##	16993
##	Seasoning
##	29838
##	Soft drinks
##	27292
##	Sweets & Candy
##	33249

```
##           Tea
##          12397
##         Tobacco
##          12034
##       Toilet tissue
##           9734
##         Toiletries
##          22675
##        Vegetables
##          42503
##          Vinegar
##           4338
##          Yogurt
##          44027
```

Count by brand

```
View(df %>% group_by(category) %>%
      summarise(count = n_distinct(brand)))
```

```
## 'summarise()' ungrouping output (override with '.groups' argument)
```