

1. 処理イメージ

- 1. テーブルを3つに分ける
- 2. ヘッダー加工
- 3. サブカテゴリに主要カテゴリ項目付与
- 4. 結合
- 5. 欠損埋め、ヘッダー整理

ここでは分かりやすくするため「年月 = "2023-07","2023-08"」の2ヵ月分で説明する。
※「カテゴリ」名は勝手に設定。

カテゴリ	売上高(千円)	構成比 (%)	対前年増減(-)率(%)	年月
総 額	140115227	100	12.2	2023-07
紳士服・洋品	9288658	6.6	14.5	2023-07
婦人服・洋品	20800539	14.8	18.3	2023-07
子供服・洋品	1786671	1.3	11.1	2023-07
その他衣料品	1818815	1.3	-2.9	2023-07
衣 料 品	33694683	24	15.5	2023-07
身のまわり品	24914267	17.8	21.4	2023-07
化粧品	11112711	7.9	24.4	2023-07
美術・宝飾・貴金属	15572050	11.1	2.1	2023-07
その他雑貨	4793985	3.4	14.7	2023-07
雑 貨	31478746	22.5	11	2023-07
家 具	1469317	1	2.7	2023-07
家 電	902232	0.6	24.4	2023-07
その他家庭用品	2894566	2.1	4	2023-07
家庭用品	5266115	3.8	6.6	2023-07
生 鮮 食 品	4792748	3.4	4.3	2023-07
菓 子	11481277	8.2	14.3	2023-07
惣 菜	7678389	5.5	6.6	2023-07
その他食料品	13280938	9.5	-1.5	2023-07
食 料 品	37233352	26.6	5.4	2023-07
食 堂 喫 茶	2803509	2	21.2	2023-07
サ ー ビ ス	1772802	1.3	-6.9	2023-07
そ の 他	2951753	2.1	28.2	2023-07
総 額	113562195	100	10	2023-08
紳士服・洋品	6708842	5.9	12.4	2023-08
婦人服・洋品	16047685	14.1	17.3	2023-08
子供服・洋品	1462202	1.3	17.7	2023-08
その他衣料品	1307363	1.2	-8.9	2023-08
衣 料 品	25526092	22.5	14.3	2023-08
身のまわり品	20925604	18.4	17.2	2023-08

1.1 テーブルを3つに分ける

●主要カテゴリ（PK：カテゴリ、年月）

カテゴリ	売上高(千円)	構成比（％）	対前年増減(-)率(%)	年月
衣料品	33694683	24	15.5	2023-07
身のまわり品	24914267	17.8	21.4	2023-07
雑貨	31478746	22.5	11	2023-07
家庭用品	5266115	3.8	6.6	2023-07
食料品	37233352	26.6	5.4	2023-07
食堂喫茶	2803509	2	21.2	2023-07
サービス	1772802	1.3	-6.9	2023-07
その他	2951753	2.1	28.2	2023-07
衣料品	25526092	22.5	14.3	2023-08
身のまわり品	20925604	18.4	17.2	2023-08
雑貨	29370909	25.9	17.3	2023-08
家庭用品	5074998	4.5	14	2023-08
食料品	25722458	22.7	-1.2	2023-08
食堂喫茶	2667047	2.3	25.6	2023-08
サービス	2207410	1.9	-37.3	2023-08
その他	2067677	1.8	12.4	2023-08

●主要カテゴリ総額（PK：年月）

カテゴリ	売上高(千円)	構成比（％）	対前年増減(-)率(%)	年月
総額	140115227	100	12.2	2023-07
総額	113562195	100	10	2023-08

1.2 ヘッダー加工

●主要カテゴリ

主要カテゴリ	主要カテゴリ売上高(千円)	主要カテゴリ構成比（％）	主要カテゴリ対前年増減(-)率(%)	年月
衣料品	33694683	24	15.5	2023-07
身のまわり品	24914267	17.8	21.4	2023-07
雑貨	31478746	22.5	11	2023-07
家庭用品	5266115	3.8	6.6	2023-07
食料品	37233352	26.6	5.4	2023-07
食堂喫茶	2803509	2	21.2	2023-07
サービス	1772802	1.3	-6.9	2023-07
その他	2951753	2.1	28.2	2023-07
衣料品	25526092	22.5	14.3	2023-08
身のまわり品	20925604	18.4	17.2	2023-08
雑貨	29370909	25.9	17.3	2023-08
家庭用品	5074998	4.5	14	2023-08
食料品	25722458	22.7	-1.2	2023-08
食堂喫茶	2667047	2.3	25.6	2023-08
サービス	2207410	1.9	-37.3	2023-08
その他	2067677	1.8	12.4	2023-08

●主要カテゴリ総額

主要カテゴリ総額売上高(千円)	主要カテゴリ総額対前年増減(-)率(%)	年月
140115227	12.2	2023-07
113562195	10	2023-08

●サブカテゴリ（PK：カテゴリ、年月）

カテゴリ	売上高(千円)	構成比（％）	対前年増減(-)率(%)	年月
紳士服・洋品	9288658	6.6	14.5	2023-07
婦人服・洋品	20800539	14.8	18.3	2023-07
子供服・洋品	1786671	1.3	11.1	2023-07
その他衣料品	1818815	1.3	-2.9	2023-07
化粧品	11112711	7.9	24.4	2023-07
美術・宝飾・貴金属	15572050	11.1	2.1	2023-07
その他雑貨	4793985	3.4	14.7	2023-07
家具	1469317	1	2.7	2023-07
家電	902232	0.6	24.4	2023-07
その他家庭用品	2894566	2.1	4	2023-07
生鮮食品	4792748	3.4	4.3	2023-07
菓子	11481277	8.2	14.3	2023-07
惣菜	7678389	5.5	6.6	2023-07
その他食料品	13280938	9.5	-1.5	2023-07
紳士服・洋品	6708842	5.9	12.4	2023-08
婦人服・洋品	16047685	14.1	17.3	2023-08
子供服・洋品	1462202	1.3	17.7	2023-08
その他衣料品	1307363	1.2	-8.9	2023-08
化粧品	11051240	9.7	36.1	2023-08
美術・宝飾・貴金属	13925991	12.3	4.6	2023-08
その他雑貨	4393678	3.9	21.7	2023-08
家具	1296178	1.1	13.1	2023-08
家電	1014698	0.9	43.9	2023-08
その他家庭用品	2764122	2.4	6.2	2023-08
生鮮食品	3284401	2.9	-2.4	2023-08
菓子	8217530	7.2	5.6	2023-08
惣菜	6129009	5.4	-4.8	2023-08
その他食料品	8091518	7.1	-4.2	2023-08

●サブカテゴリ

サブカテゴリ	サブカテゴリ売上高(千円)	サブカテゴリ構成比（％）	サブカテゴリ対前年増減(-)率(%)	年月
紳士服・洋品	9288658	6.6	14.5	2023-07
婦人服・洋品	20800539	14.8	18.3	2023-07
子供服・洋品	1786671	1.3	11.1	2023-07
その他衣料品	1818815	1.3	-2.9	2023-07
化粧品	11112711	7.9	24.4	2023-07
美術・宝飾・貴金属	15572050	11.1	2.1	2023-07
その他雑貨	4793985	3.4	14.7	2023-07
家具	1469317	1	2.7	2023-07
家電	902232	0.6	24.4	2023-07
その他家庭用品	2894566	2.1	4	2023-07
生鮮食品	4792748	3.4	4.3	2023-07
菓子	11481277	8.2	14.3	2023-07
惣菜	7678389	5.5	6.6	2023-07
その他食料品	13280938	9.5	-1.5	2023-07
紳士服・洋品	6708842	5.9	12.4	2023-08
婦人服・洋品	16047685	14.1	17.3	2023-08
子供服・洋品	1462202	1.3	17.7	2023-08
その他衣料品	1307363	1.2	-8.9	2023-08
化粧品	11051240	9.7	36.1	2023-08
美術・宝飾・貴金属	13925991	12.3	4.6	2023-08
その他雑貨	4393678	3.9	21.7	2023-08
家具	1296178	1.1	13.1	2023-08
家電	1014698	0.9	43.9	2023-08
その他家庭用品	2764122	2.4	6.2	2023-08
生鮮食品	3284401	2.9	-2.4	2023-08
菓子	8217530	7.2	5.6	2023-08
惣菜	6129009	5.4	-4.8	2023-08
その他食料品	8091518	7.1	-4.2	2023-08

1.3 サブカテゴリに主要カテゴリ項目作成

主要カテゴリ	サブカテゴリ	サブカテゴリ売上高(千円)	サブカテゴリ構成比(%)	サブカテゴリ対前年増減(±)率(%)	年月
衣料品	紳士服・洋品	9288658	6.6	14.5	2023-07
衣料品	婦人服・洋品	20800539	14.8	18.3	2023-07
衣料品	子供服・洋品	1786671	1.3	11.1	2023-07
衣料品	その他衣料品	1818815	1.3	-2.9	2023-07
雑貨	化粧品	11112711	7.9	24.4	2023-07
雑貨	美術・宝飾・貴金属	15572050	11.1	2.1	2023-07
雑貨	その他雑貨	4793985	3.4	14.7	2023-07
家庭用品	家具	1469317	1	2.7	2023-07
家庭用品	家電	902232	0.6	24.4	2023-07
家庭用品	その他家庭用品	2894566	2.1	4	2023-07
食料品	生鮮食品	4792748	3.4	4.3	2023-07
食料品	菓子	11481277	8.2	14.3	2023-07
食料品	惣菜	7678389	5.5	6.6	2023-07
食料品	その他食料品	13280938	9.5	-1.5	2023-07
衣料品	紳士服・洋品	6708842	5.9	12.4	2023-08
衣料品	婦人服・洋品	16047685	14.1	17.3	2023-08

1.4 結合

※Excelで作ったのでヘッダーはpandasのmerge()と違うかも。年月のddhhmmddは無視。

主要カテゴリ	サブカテゴリ	サブカテゴリ売上高(千円)	サブカテゴリ構成比(%)	サブカテゴリ対前年増減(±)率(%)	年月
衣料品	紳士服・洋品	9288658	6.6	14.5	2023/07/01 0:00:00
衣料品	婦人服・洋品	20800539	14.8	18.3	2023/07/01 0:00:00
衣料品	子供服・洋品	1786671	1.3	11.1	2023/07/01 0:00:00
衣料品	その他衣料品	1818815	1.3	-2.9	2023/07/01 0:00:00
雑貨	化粧品	11112711	7.9	24.4	2023/07/01 0:00:00
雑貨	美術・宝飾・貴金属	15572050	11.1	2.1	2023/07/01 0:00:00
雑貨	その他雑貨	4793985	3.4	14.7	2023/07/01 0:00:00

1.5 欠損埋め、ヘッダー整理

主要カテゴリ	サブカテゴリ	サブカテゴリ売上高(千円)	サブカテゴリ構成比(%)	サブカテゴリ対前年増減(±)率(%)	年月
衣料品	紳士服・洋品	9288658	6.6	14.5	2023/07/01 0:00:00
衣料品	婦人服・洋品	20800539	14.8	18.3	2023/07/01 0:00:00
衣料品	子供服・洋品	1786671	1.3	11.1	2023/07/01 0:00:00
衣料品	その他衣料品	1818815	1.3	-2.9	2023/07/01 0:00:00
雑貨	化粧品	11112711	7.9	24.4	2023/07/01 0:00:00
雑貨	美術・宝飾・貴金属	15572050	11.1	2.1	2023/07/01 0:00:00
雑貨	その他雑貨	4793985	3.4	14.7	2023/07/01 0:00:00

2.参考コード

```
# 「# 横持変換」の続きから
# 本来カラム名がすべて設定されていることが望ましい
concatenated_tables.rename(
    columns={concatenated_tables.columns[0]: "カテゴリ"}, inplace=True)
concatenated_tables.head()

category_dict = {
    "衣料品": ["紳士服・洋品", "婦人服・洋品", "子供服・洋品", "その他衣料品"],
    "身のまわり品": [""],
    "雑貨": ["化粧品", "美術・宝飾・貴金属", "その他雑貨"],
    "家庭用品": ["家具", "家電", "その他家庭用品"],
    "食料品": ["生鮮食品", "菓子", "惣菜", "その他食料品"],
    "食堂喫茶": [""],
    "サービス": [""],
    "その他": [""],
}

# 主要カテゴリ
category_keys = category_dict.keys()
df_main_category = concatenated_tables[
    concatenated_tables["カテゴリ"].isin(category_keys)
]

# サブカテゴリ
category_values = [
    item for sublist in category_dict.values() for item in sublist
]
df_sub_category = concatenated_tables[
    concatenated_tables["カテゴリ"].isin(category_values)
]

# 総額
df_total = concatenated_tables[concatenated_tables["カテゴリ"] == "総額"]

# 面倒なので無理やりカラム名変更
df_main_category.columns = [
    "主要カテゴリ", "主要カテゴリ売上高(千円)", "主要カテゴリ構成比(%)",
    "主要カテゴリ対前年増減(-)率(%)", "年月"]
df_sub_category.columns = [
    "サブカテゴリ", "サブカテゴリ売上高(千円)", "サブカテゴリ構成比(%)",
```

```

    "サブカテゴリ対前年増減(-)率( % )", "年月"]
df_total.drop(columns=["カテゴリ", "構成比(%)"], inplace=True)
df_total.columns = [
    "主要カテゴリ総額売上高(千円)", "主要カテゴリ総額対前年増減(-)率( % )",
    "年月"]

# サブカテゴリに主要カテゴリ列追加
def add_main_category_column(df, category_dict):
    category_map = {v: k for k, values in category_dict.items()
                     for v in values}
    df["主要カテゴリ"] = df["サブカテゴリ"].map(category_map)
    return df

df_sub_category = add_main_category_column(df_sub_category, category_dict)

# 結合
df_finaly = df_main_category.merge(
    df_sub_category, how="left", on=["主要カテゴリ", "年月"]) \
    .merge(df_total, how="left", on="年月")

# カラム名並び替え・欠損値補完
df_finaly = df_finaly[[
    col for col in df_finaly.columns if col != "年月"] + ["年月"]]
df_finaly.fillna("-", inplace=True)

```