■図2-6：データ補正前の集計結果(金額)

```
In [7]: res = uriage_data.pivot_table(index="purchase_month", columns="item_name", values="item_price", aggfunc="sum", fill_value=0)
        res
```

Out[7]:

item_name	商品W	商品n	商品E	商品M	商品P	商品S	商品W	商品X	商品O	商品Q	...	商品k	商品l	商品o	商品p	商品r	商品s	商品t	商品v	商品x	商品y
purchase_month																					
201901	0	1400	0	0	0	0	0	0	0	0	...	1100	1200	1500	0	0	0	0	0	0	0
201902	0	0	0	0	0	0	0	2400	0	0	...	0	0	0	0	0	1900	2000	2200	0	0
201903	0	0	500	1300	1600	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
201904	2300	0	0	0	0	0	0	0	0	1700	...	0	0	0	0	0	1900	0	0	0	0
201905	0	0	0	0	0	1900	0	0	0	0	...	0	1200	0	0	0	0	0	0	0	2500
201906	0	0	0	0	0	0	2300	0	0	0	...	0	0	0	1600	0	0	0	0	2400	0
201907	0	0	0	0	0	0	0	0	0	0	...	0	0	1500	0	1800	0	0	0	0	0

7 rows × 99 columns

すでに日付の年月処理は行われていますので、pivot_tableにて集計処理を行うだけです。

こちらも同じ様に正しい集計結果になっていない事が確認できました。

このように、データの揺れが残ったまま集計・分析を行っても、**全く意味のない結果**となってしまいますので、いかにデータ加工が分析の前処理として重要かが分かるかと思います。

それでは、以降ではデータの揺れを補正するデータ加工に挑戦していきましょう。

ノック14：商品名の揺れを補正しよう

まずは商品名の揺れを補正していきましょう。

今回のケースは比較的簡単なデータの揺れで、「スペースの有無」「半角・全角」の揺れを補正するだけで解決できそうです。

まずは現状の確認から実施します。補正後の結果が正しい結果かどうかを判定するために、現状の把握はとても重要です。

```
print(len(pd.unique(uriage_data.item_name)))
```


■図2-7：商品名のユニーク数確認

```
In [8]: print(len(pd.unique(uriage_data["item_name"])))
        99
```

売上履歴のitem_nameの重複を除外したユニークなデータ件数をpd.uniqueで確認する事ができます。

前のノックでも確認したように、本来A～Zの26商品が99個に増えてしまっています。

さっそく、データの揺れを解消していきましょう。

```
uriage_data["item_name"] = uriage_data["item_name"].str.upper()
uriage_data["item_name"] = uriage_data["item_name"].str.replace("　", "")
uriage_data["item_name"] = uriage_data["item_name"].str.replace(" ", "")
uriage_data.sort_values(by=["item_name"], ascending=True)
```

■図2-8：処理結果

```
In [8]: uriage_data["item_name"] = uriage_data["item_name"].str.upper()
        uriage_data["item_name"] = uriage_data["item_name"].str.replace("　", "")
        uriage_data["item_name"] = uriage_data["item_name"].str.replace(" ", "")
        uriage_data.sort_values(by=["item_name"], ascending=True)
```

Out[8]:

	purchase_date	item_name	item_price	customer_name	purchase_month
0	2019-06-13 18:02:34	商品A	100.0	深井栗々美	201906
1748	2019-05-19 20:22:22	商品A	100.0	松川綾女	201905
223	2019-06-25 08:13:20	商品A	100.0	板橋隆	201906
1742	2019-06-13 16:03:17	商品A	100.0	小平陽子	201906
1738	2019-02-10 00:28:43	商品A	100.0	松田浩正	201902
1721	2019-02-24 19:18:05	商品A	100.0	楠哲平	201902
1708	2019-03-27 17:10:06	商品A	100.0	西脇礼子	201903
1707	2019-03-25 21:42:02	商品A	100.0	逸見広司	201903
234	2019-03-23 09:32:03	商品A	100.0	赤木だん吉	201903
1684	2019-02-17 20:25:57	商品A	100.0	手塚雅之	201902
497	2019-02-11 03:54:57	商品A	100.0	堀江佑	201902
1761	2019-05-04 14:44:55	商品A	100.0	八木雅彦	201905
1658	2019-03-23 20:08:49	商品A	100.0	片瀬貴利	201903
1544	2019-01-21 11:49:01	商品A	100.0	藤広之	201901
1513	2019-04-04 18:58:05	商品A	100.0	河村由樹	201904
1500	2019-02-28 17:25:59	商品A	100.0	赤木愛梨	201902
1498	2019-01-19 12:15:21	商品A	100.0	今西	201901