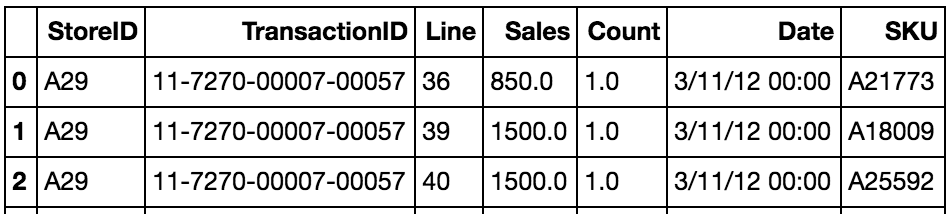
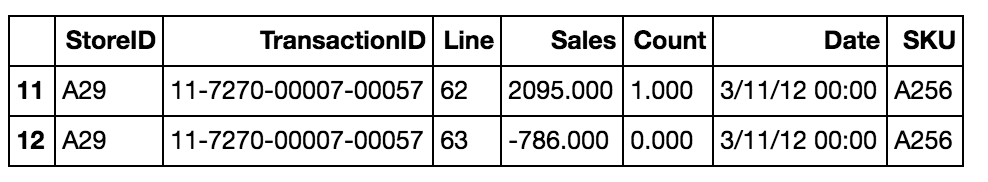
**All Scenarios**

**If I discount this pencil, will that drive up the demand per transaction. I end up buying a lot of other stuffs.**

**There are total 1M rows and 7 features in the dataset**

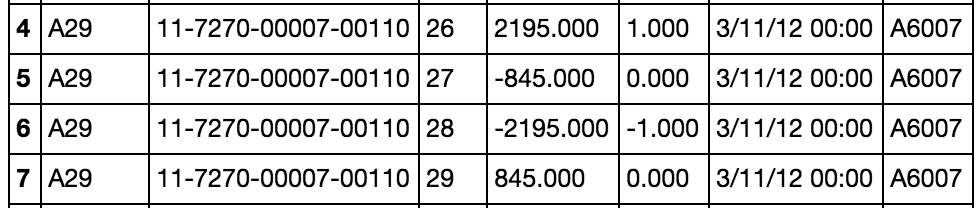
****

**scenario 1**: when count=0, sales is negative, then 'it is a discount'

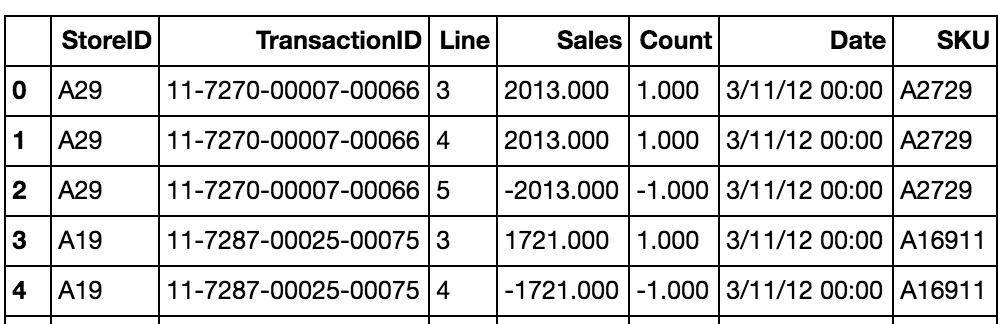


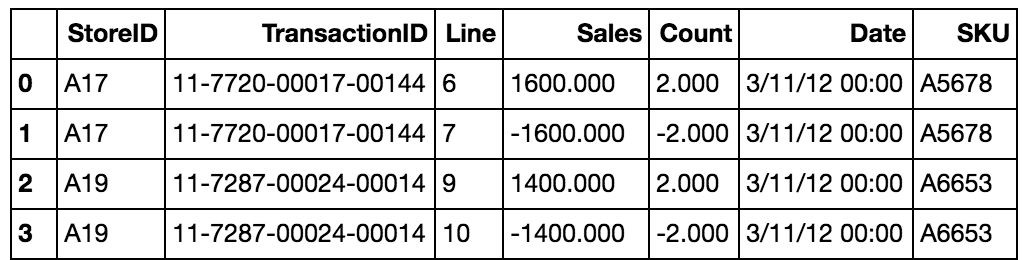
**scenario 2**: when count=0, sales is positive, then 'it is a return'

The positive sales amount is basically the credit back to the store for the discount extended to the customer earlier. (everthing will cancle it out) (is that credit back to orphan or regular return)



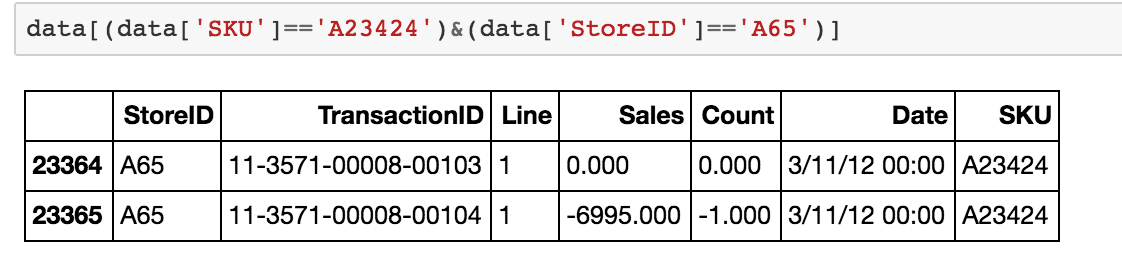
**scenario 3**: when count=-1 or -2, basically a negative number, then 'it could be a promotion strategy like buy 1 get 1 free, or a return', in this analysis, I treated it as a return.

****

****

**scenario 4:** no **corresponding** positive sales when count<0 for SKU in a specfic store, I found count is negative meaning is a return, but no corresponding sales associated with it. **Meaning that it returned sth bought earlier from other stores or that is not on the record.** Or when item is returned, I can’t find a prior mathcing sales record for it.

**Number of SKUs that exhibit such behavior across all the stores, there might be duplicate SKUs in this number**: 1020

****

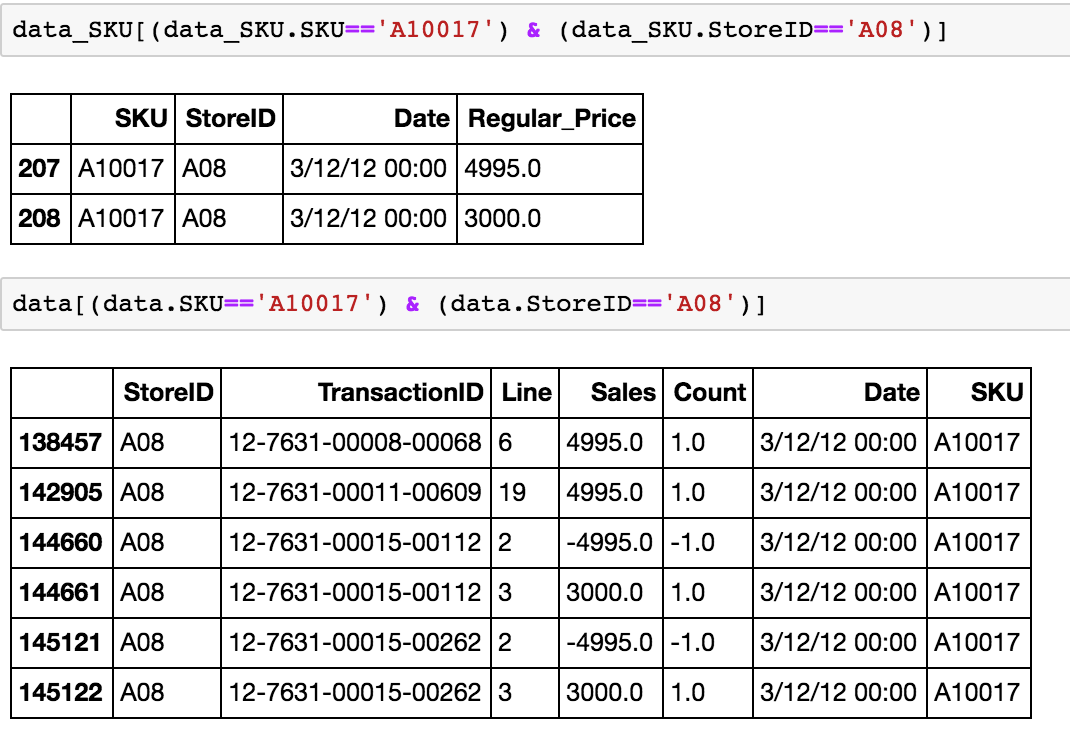
**Next scenarios will be based on SKU level per store, per day**

**(Total count of row numbers of SKU on the per day, per store level that exhibit such behaviors: 200K)**

**scenario 5: 2 different regular prices for the same SkU within the same store on the same date**

Number of cases on SKU level per store, per day: 1270

a. Promotion can start anytime

****

**Explanation:** a store's promotion can start anytime during a day. It could happen that the transaction captures sale of an item before and after a promotion starts.

17.9.16 Do some imputation. (add rows, figure out probabaly data entry error, to correct the records through the imputation to fill in the missing records)

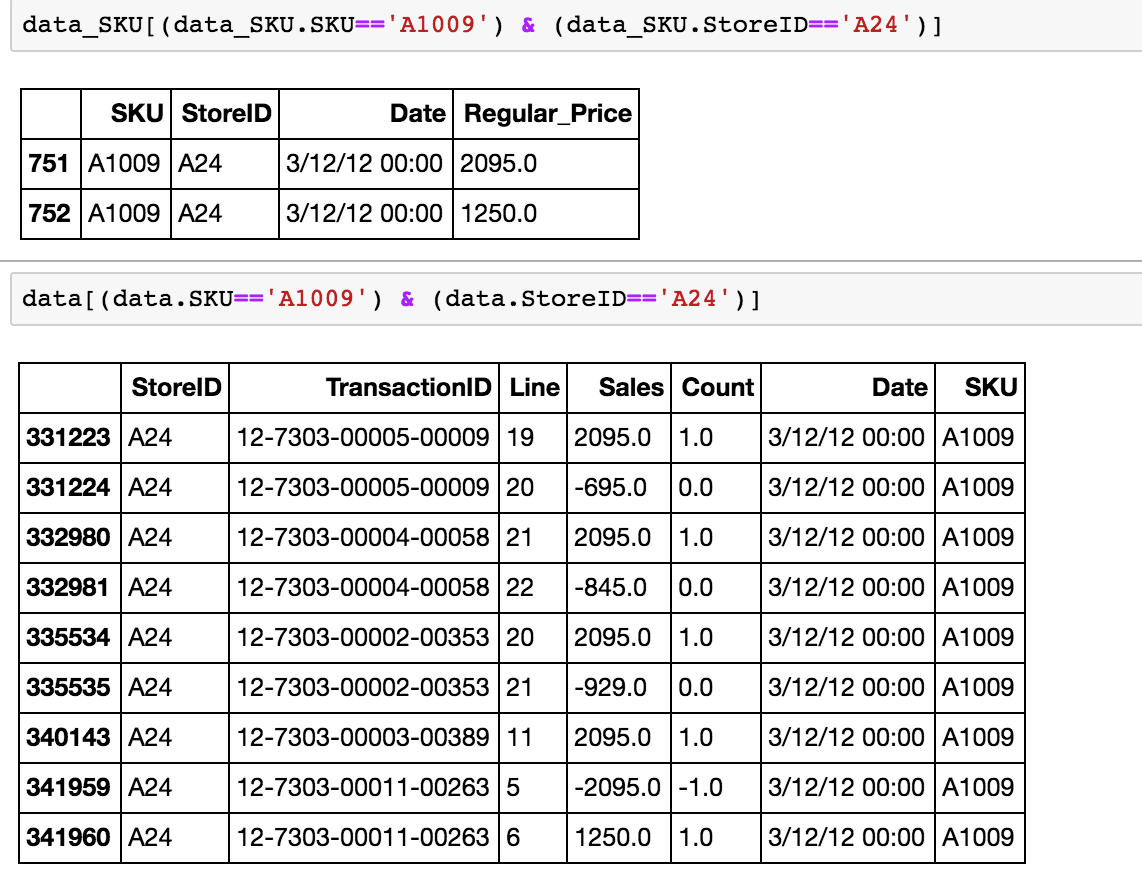
leave that to me

different regular prices(actually discounted), just remove orphan returns, we

**I don’t want you spend a time on it (we can just filter out)**

b. Different discount value and regular price equals to the price after discount

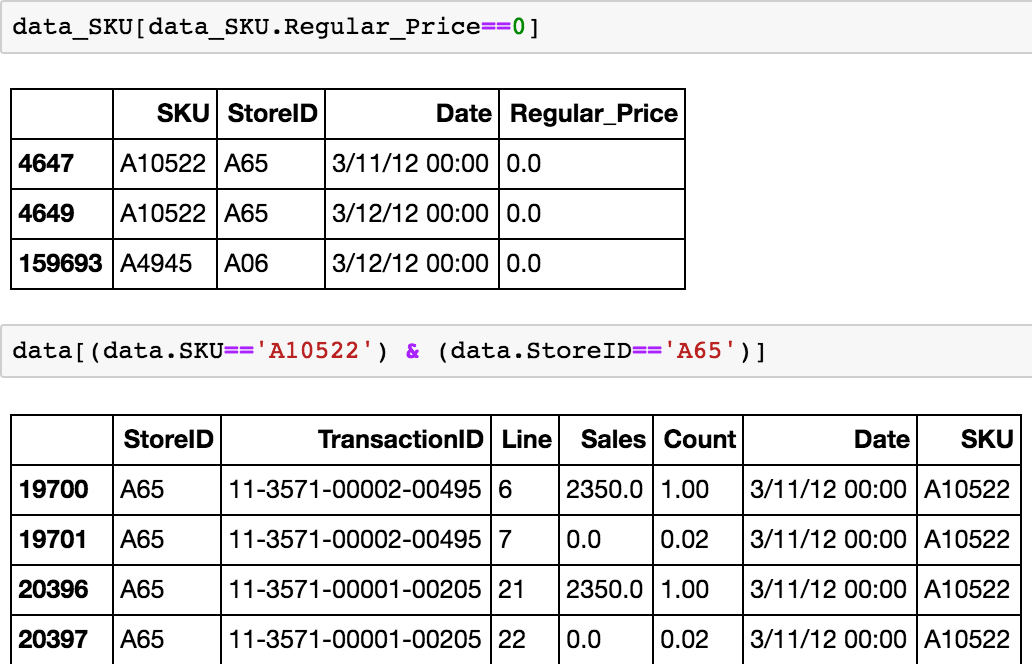
leave it (different membership, it won’t skew the analysis)

****

c. Sales=0 and Count is a small number

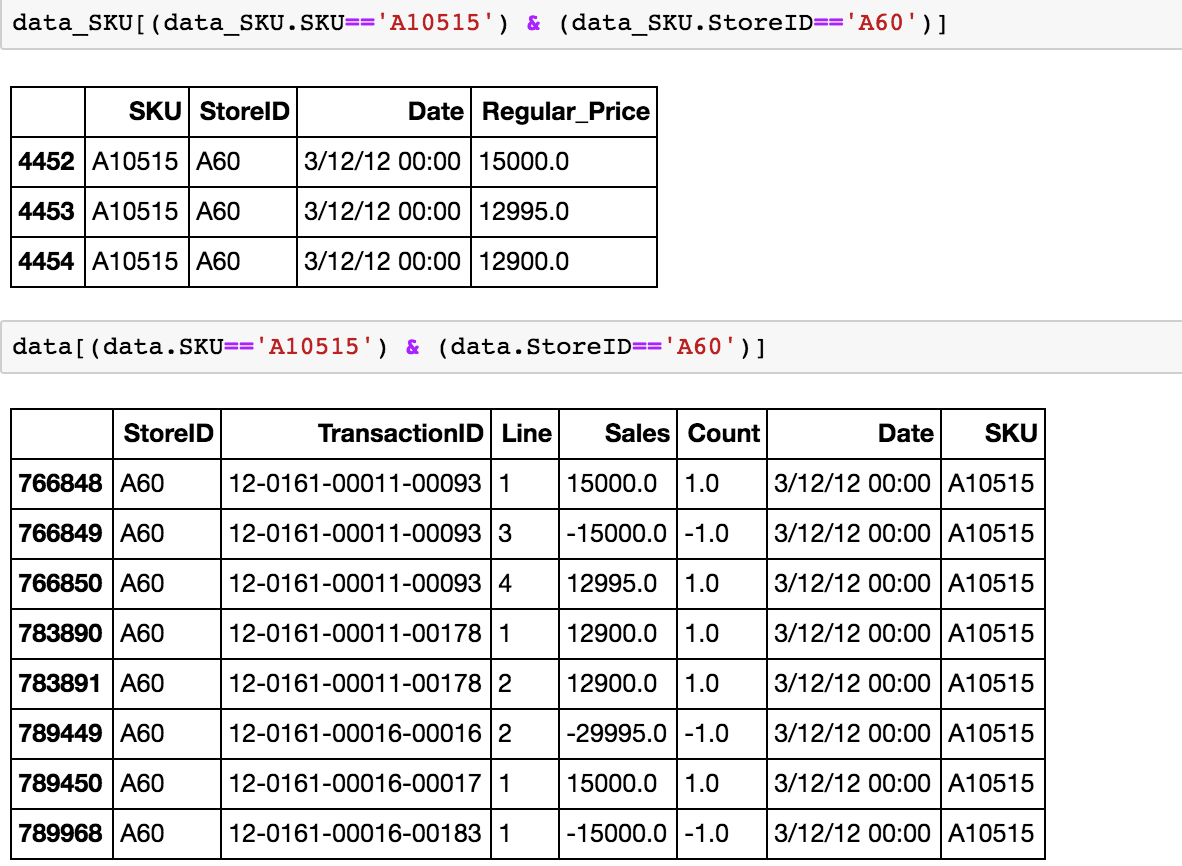
There are 3 cases

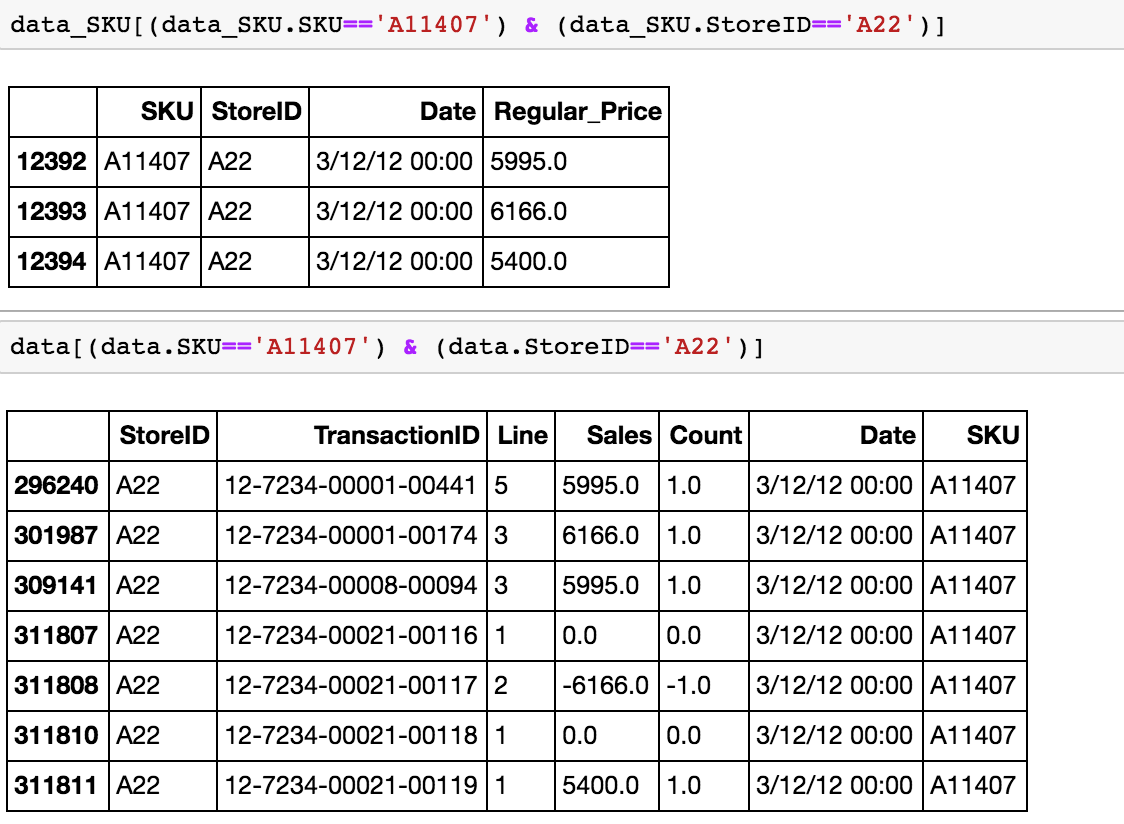
Number of cases on SKU level per store, per day: 3

****

**scenario 6:** 3 distinct regular price on SKU level per store, per day

Number of cases on SKU level per store, per day: 30



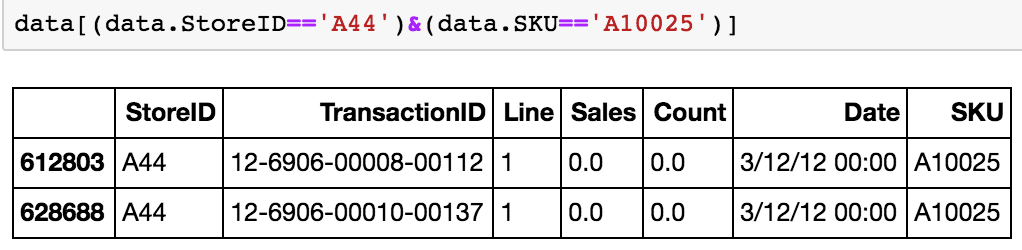
****

**Let’s assume the same store we can’t charge at diff price on same SkU within same date, we decide to filter out those scenarios.**

**选一些record做imputation，其他做filter out**

**scenario 7:** There are only obersvations when Sales=0 & Count=0 on SKU level per store, per day

Number of cases on SKU level per store, per day: 477

****

**Gift card for 10$, we are not making money on it.**

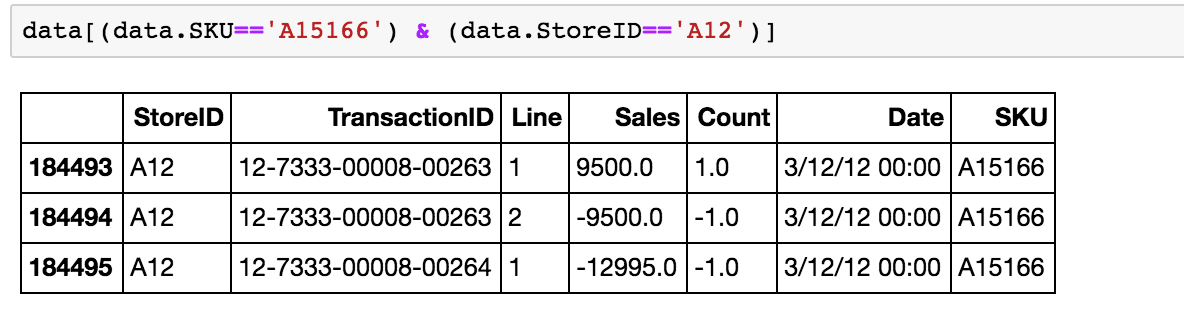
**In cases, when there is 0 count and 0 sales, We can just leave it here.**

**I will filter it out.**

**scenario 8:** Unit Price after sales is bigger than regular price on SKU level per store, per day

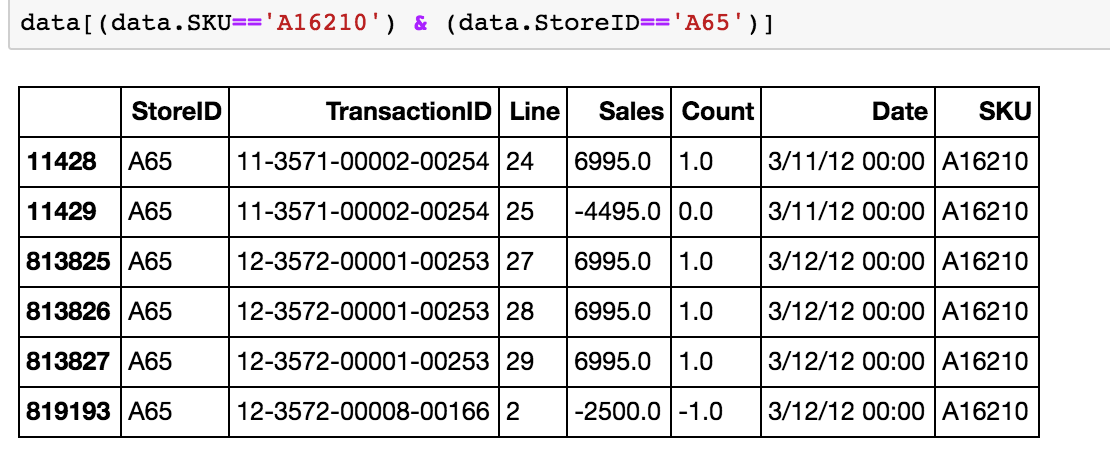
Number of cases on SKU level per store, per day: 22

a.

****

**Explanation:** It only have returned history, but no purchased history. It could be the reason that the item was bought outside timeperiod of dataset, so it was not recorded in this dataset.

b.

****

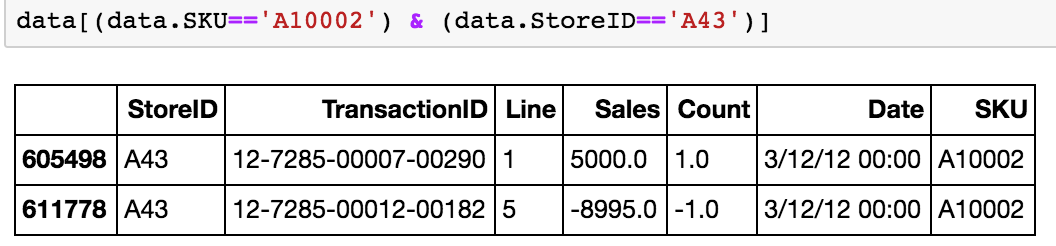
**Explanation:** Returned price is lower than the regular price. (Maybe date is wrong?)

**scenario 9:** Discount rate is missing value

Number of cases on SKU level per store, per day: 1041

a. Total Sales<0, Total Count=0

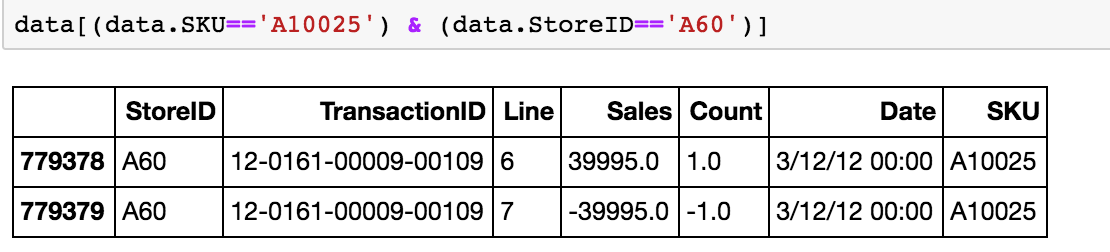
Number of cases on SKU level per store, per day: 51



**Explanation:** Returned price is higher than the regular price and total count equalts to 0.

b. Total Sales=0, Total Count=0

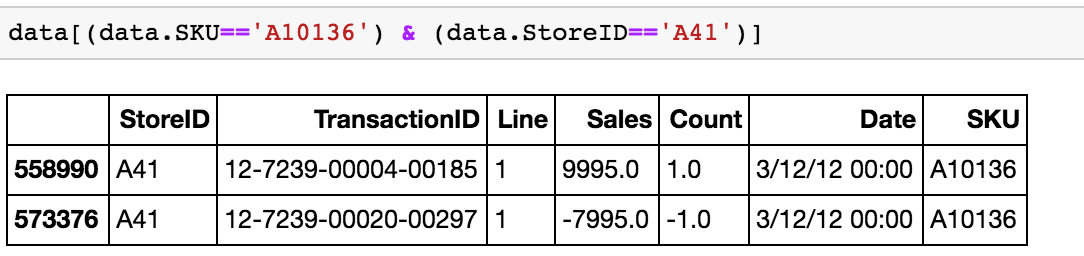
Number of cases on SKU level per store, per day: 957



**Explanation:** Sales cancel each other out.

c. Total Sales>0, Total Count=0

Number of cases on SKU level per store, per day: 33

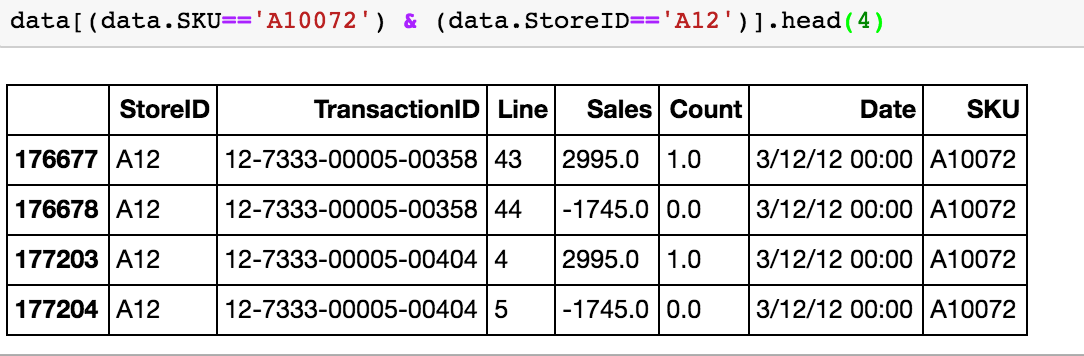


**Explanation:** Returned price is lower than the regular price and total count equals to 0.

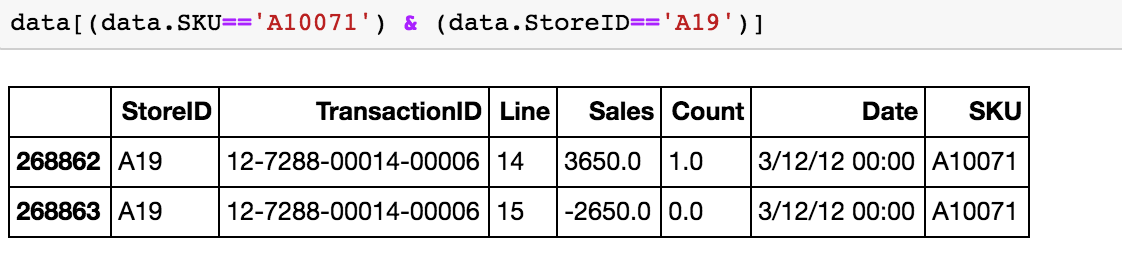
**scenario 10:** Discount rate is higher than 30%

Number of cases on SKU level per store, per day: 11831

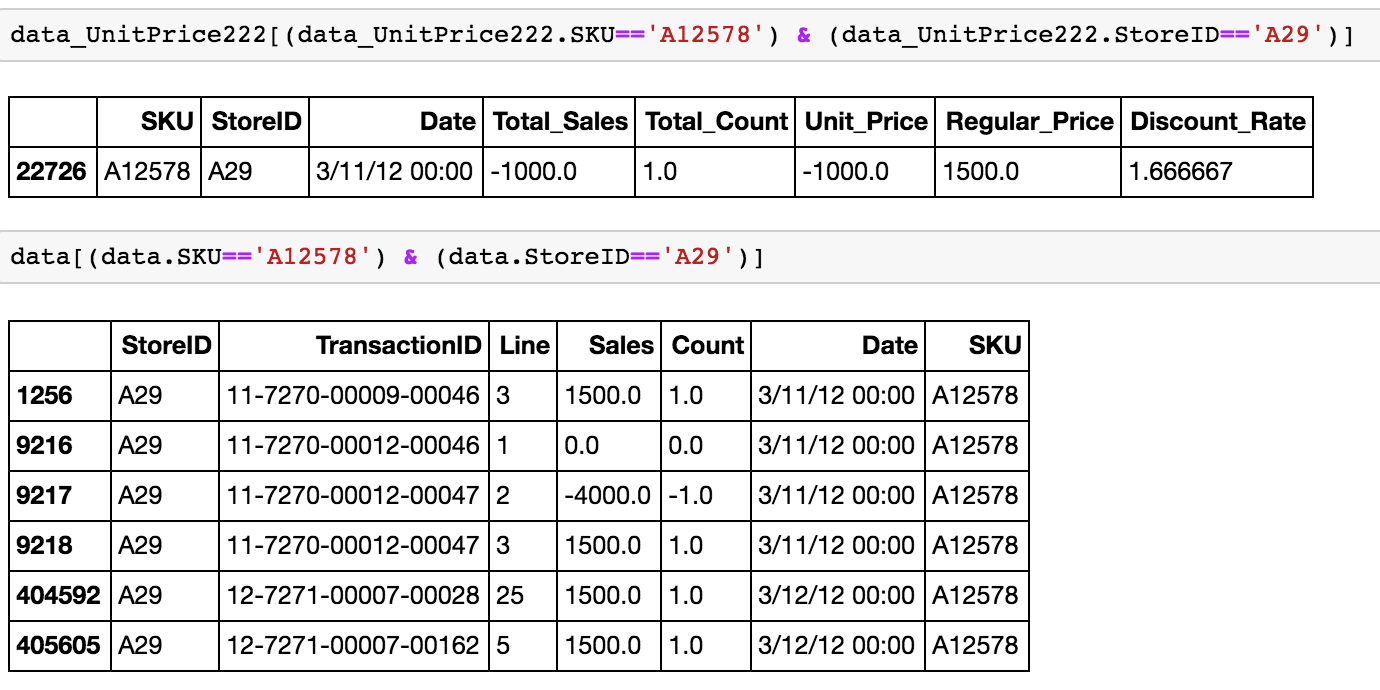
a. Discount Rate is 58%

****

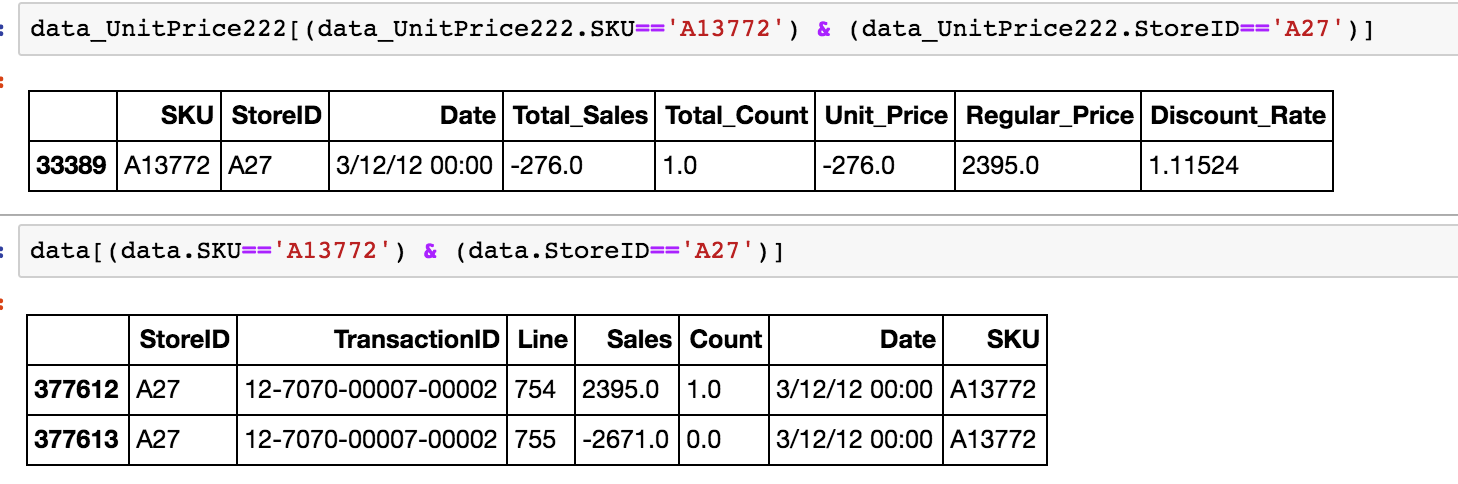
b. Discount Rate is 72%



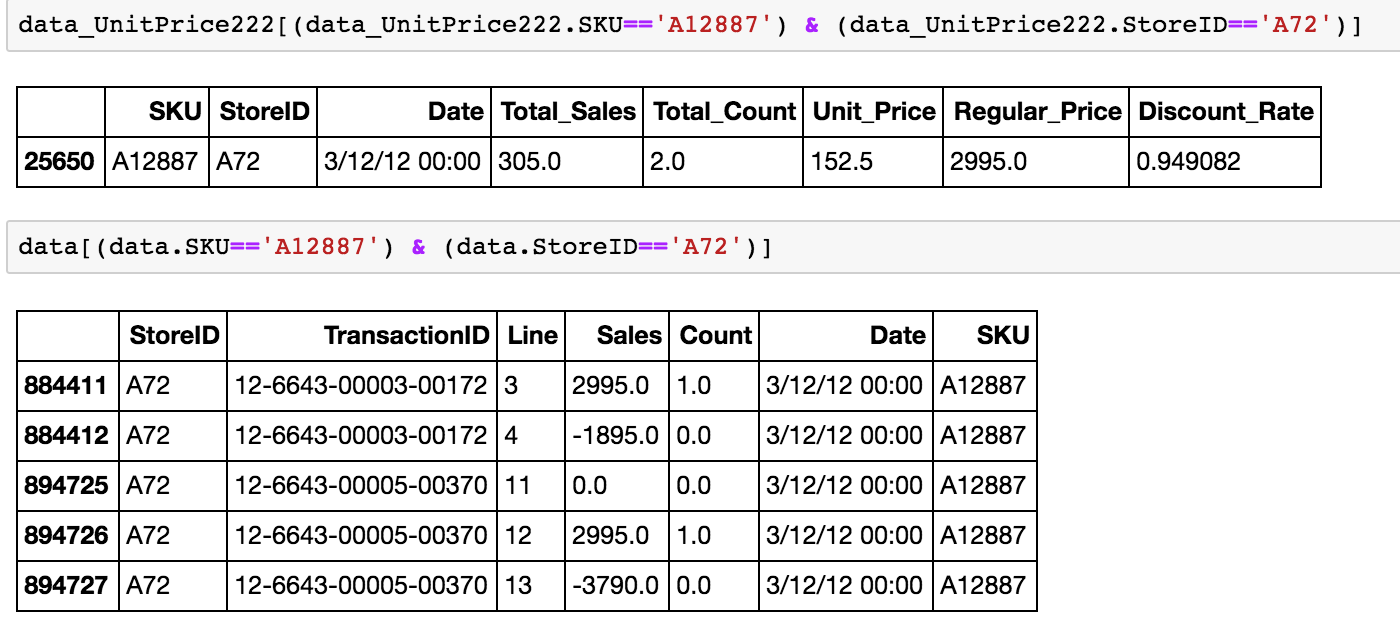
c. Returned price is higher than the regular price



d. Discount is larger than the regular price

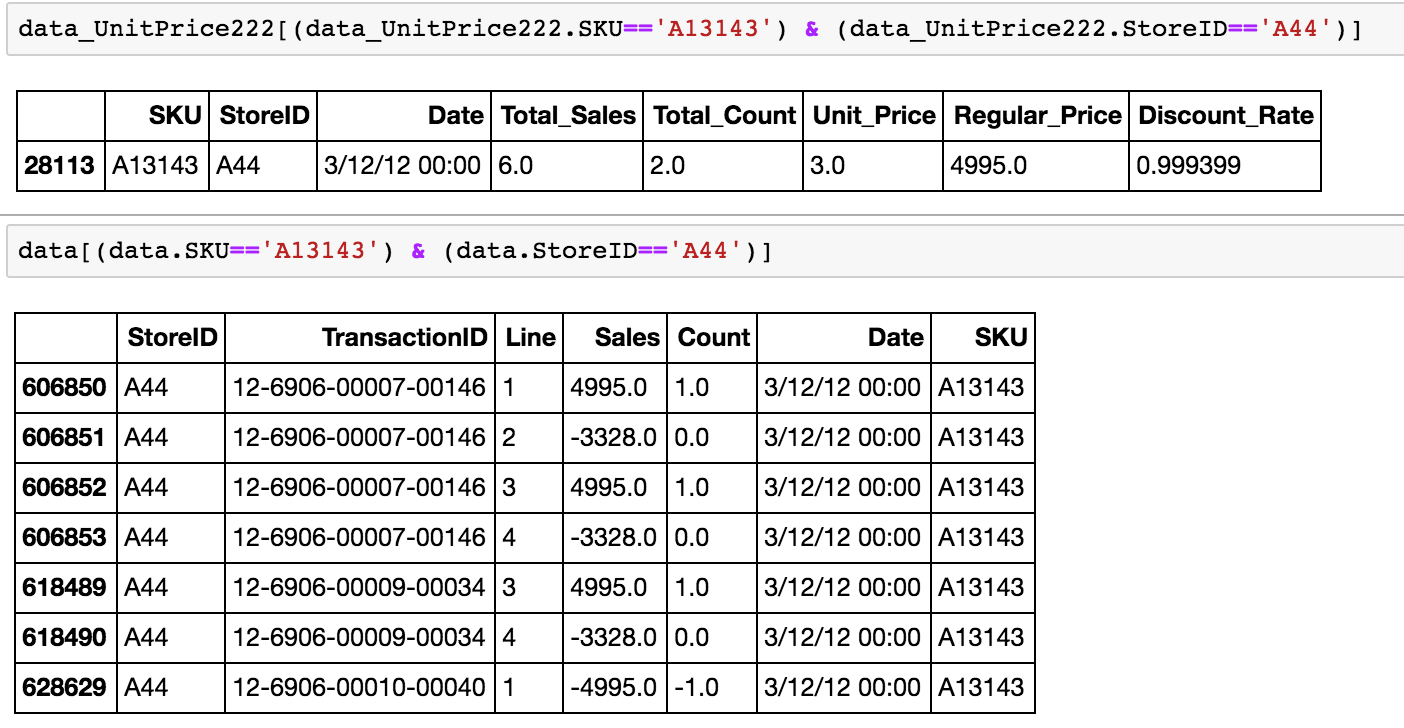


e. Different discount value and discount is even larger than regular price



**Anomaly(not normal): we list out 10 diferent , later on, if these cases are true anormaly or not. We do imputation to fill or we fillter out**

f. No credit back to the store when returning the discounted items



Clean data

Whole project is like an EDA. Let’s call it an EDA. Let’s wrap up.

Go back to your data, make sure everything is Okay.

Metrics:

On a per SKU level, look at proportion of the discounted item on per basket level. Discount pencil 10 to 5, drive up the demand. (decrease in sales will that drive up increase in demand)

SKU level: elastictiy of demand, across different stores, select top 50 features.

Average: Per store basis, per day basis

All sales with discount on a

Association Analysis

What are the features: access all sort of data. Cost Marketing campaign data (coupon) to the value, the product, revenue is provided to me. / Retail : / Profit margin / Inventory cost: low inventory level for those unpopluar data (use this space to store item that once discount will drive up our profit) / Markdown: discount is a money loss, how much I am making / Opportunity Cost

Seasonality: in summer, what are the products tend to buy the most /Geographic Data: do they have different trend

Promotion Strategy

additional percentage of customers

你就是我这里下一个要帮忙找到工作的学生