Local Art Coupon Effectiveness Analysis

Based on T-test method

Xudong ZHANG

USC ID: 2406221580

Key words: t-test, coupon, age, visit frequency, join date, customer segmentation

Contents

Part I: Business Understanding

Background

Local Art Gallery, as an artwork dealer, makes revenue from commission (Paintings 8%, Jewelry 8%, Mosaics 12%, Sculpture 12%).

And two co-founders are arguing about whether the 10% coupon promotion campaign will promote revenue(commission).

In this case, t-test is suggested to rest this debate and serve for the future promotion plans.

Task

Use t-test to determine whether this 10%-off coupon will improve revenues (commission) Should LocalArt offer the 10%-off coupon to all, some, or none of its customers? Which ones? Should it adopt some other type of promotion? Why?

Develop Other business recommendations based on insights from the data. How can these data be used to improve LocalArt's profits?

What other data might you suggest, in an ideal world tracking at LocalArt to help business? Why?

Part II: Preparation

Data Preview

The first data set consists of characteristic and single item spending information on the 5000 customers in the study. And the second contains the customer ID of those 1017 customers that are randomly sampled from these 5000 to test with a 10% coupon. These two data set could be joined together with identical customer ID.

Exhibit 1 Two initial data sets joined by "CustID"

```
> str(data)
'data.frame':   5000 obs. of  10 variables:
 $ X         : int  1 2 3 4 5 6 7 8 9 10 ...
 $ CustID    : num  86753091 86753092 86753093 86753094 867530
 $ JoinDate  : Factor w/ 412 levels "2014-03-18","2014-04-28",
 $ DOB       : Factor w/ 3790 levels "1964-08-01","1964-11-15"
 $ Gender    : Factor w/ 2 levels "Female","Male": 1 1 1 1 2 2
 $ Visits    : int  61 69 84 32 19 21 58 73 40 29 ...
 $ Paintings: num  550.2 96.1 10.1 106 38.6 ...
 $ Jewelry   : num  192 211 292 344.6 17.6 ...
 $ Mosaics   : num  70.6 285.1 265.5 216.1 446.3 ...
 $ Sculpture: num  38.3 15.6 23.5 318 355.6 ...
```

```
> str(coupon)
'data.frame':   1017 obs. of  2 variables:
 $ X     : int  1 2 3 4 5 6 7 8 9 10 ...
 $ CustID: num  8.68e+07 8.68e+07 8.68e+08 8.68e+08 8.68e
```

Join

CustID	JoinDate	DOB	Gender	Visits	Paintings	Jewelry	Mosaics	Sculpture	Coupon
86753091	2015-01-28	1993-06-12	Female	61	550.2223644	192.045731	70.6183075	38.3413245	no
86753092	2015-01-16	1972-04-25	Female	69	96.1207129	210.991971	285.1215228	15.6448939	no
86753093	2015-03-14	1971-11-15	Female	84	10.0682782	292.032648	265.5397291	23.4522460	no

I plotted the density distribution curves for all four kinds of spending, which are far from normally distributed. I cannot tell many things from this plot, even though the Welch's t-test helps to revise the statistical drawbacks of this data set

and proves the significant coupon effects with its super small p-value. To dig deeper, I decide to segment all the 5000 customers into smaller subgroups and try to find their corresponding coupon effects

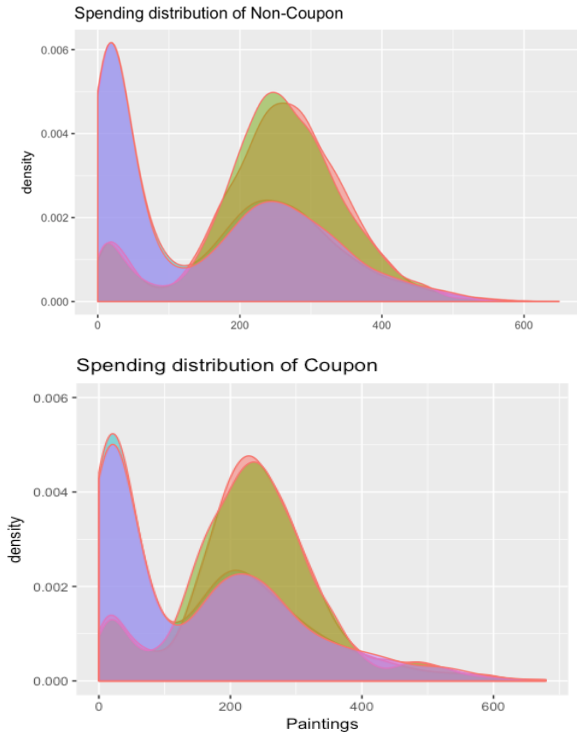
Exhibit 2 Initial t-test and visualization on single item spending (Non-Conpon vs Coupon)

```
arts = c("Jewelry","Paintings","Mosaics","Sculpture")

for (art in arts) {
  t_test_trial = t.test(coupon0[,art], coupon1[,art])
  print(art)
  print(c("p-value",t_test_trial$p.value))
  print(c (t_test_trial$estimate[1],t_test_trial$estimate[2]))

}
```

```
## [1] "Jewelry"
## [1] "p-value"                "5.66599031945904e-05"
## mean of x mean of y
##  250.7350  235.1547
## [1] "Paintings"
## [1] "p-value"           "0.195609016336686"
## mean of x mean of y
##  150.4586  144.1750
## [1] "Mosaics"
## [1] "p-value"                "1.16104680266866e-05"
## mean of x mean of y
##  248.0689  231.5345
## [1] "Sculpture"
## [1] "p-value"           "0.646429245219605"
## mean of x mean of y
##  152.1421  149.8159
```

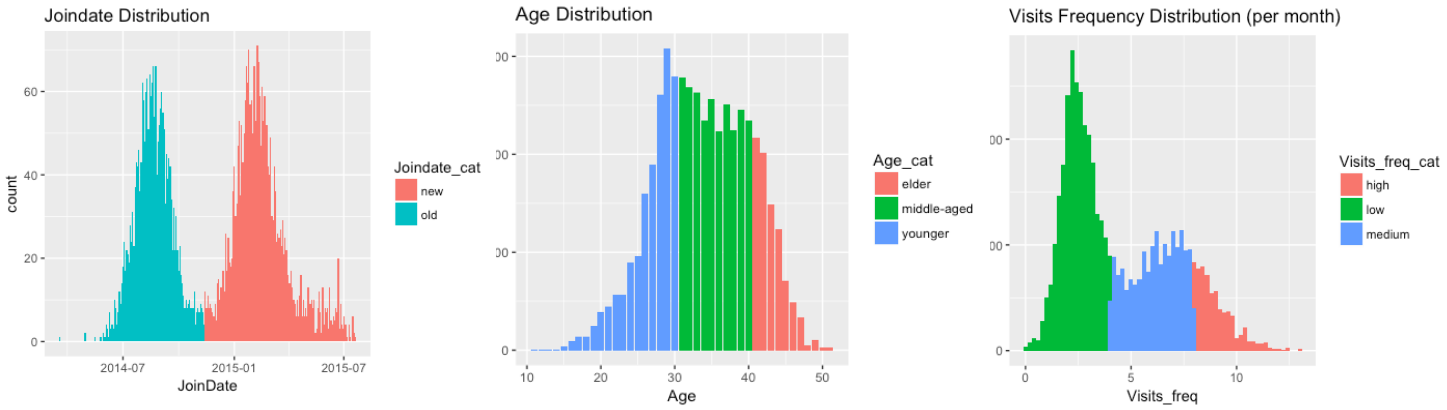


How should I to segment the customer group?

Customer segmentation

I start by getting to know our data. We start by getting to know our data. I explore the data little by plotting distribution. Trying to make more economical sense, I segment the customer into smaller group by defining different Join date levels, age level and visit frequency levels.

Exhibit 3 Three main metrics' (joindate, age, visits frequency) distribution and classification



And I also get a new 72-row categorized data set, with each row stands for a customer subgroup.

Exhibit 4 Enriched data set with new categorized information

	Mosaics	Sculpture	Coupon	Joindate_cat	Age	Age_cat	Visits_freq	Visits_freq_cat	Total_spending
31	70.6183075	38.3413245	no	new	22	younger	5.4302671	medium	851.2277
71	285.1215228	15.6448939	no	new	43	elder	5.9312321	medium	607.8791
48	265.5397291	23.4522460	no	new	44	elder	8.6301370	high	591.0929
44	216.0719710	317.9841788	no	old	30	younger	1.9161677	low	984.6340
93	446.3177941	355.5880940	no	new	34	middle-aged	2.7804878	low	858.1108
03	460.3095654	328.5865293	no	new	35	middle-aged	3.7724551	low	833.3576
36	299.3788191	210.2508320	yes	old	33	middle-aged	3.2044199	low	860.5653
45	261.0116162	4.7191375	no	new	38	middle-aged	6.0000000	medium	581.4142
13	277.7877753	176.9410317	yes	old	32	middle-aged	2.4048096	low	981.6664
23	218.6143695	196.0915753	yes	old	33	middle-aged	1.5992647	low	953.7892
20	267.2510518	1.7476496	no	new	42	elder	4.2485549	medium	576.4431

Modeling: T-tests

One of the most common tests in statistics is the t-test, used to determine whether the means of two groups are equal to each other. The assumption for the test is that both groups are sampled from normal distributions with equal variances.

The null hypothesis is that the two means are equal, and the alternative is that they are not. It is known that under the null hypothesis, we can calculate a t-statistic that will follow a t-distribution with n1 + n2 - 2 degrees of freedom. There is also a widely used modification of the t-test, known as Welch's t-test that adjusts the number of degrees of freedom when the variances are thought not to be equal to each other. Before we can explore the test much further, we need to find an easy way to calculate the t-statistic.

Welch's t-test is designed for unequal variances, but the assumption of normality is maintained. And it could be directly called by t.test() function in R.

Part III: T-test for 72 scenarios

Based on the above 18 segmentations and the 4 types of spending, I got 72 (4*3*3*2 = 72) scenarios of customer spending and conducted 72 pair of t-test results.

Exhibit 5 Four-level loops to run 72 t-tests

```
coupon1 = filter(cooldata, Coupon == "yes")
coupon0 = filter(cooldata, Coupon == "no")
arts = c("Jewelry", "Paintings", "Mosaics", "Sculpture")
t_data = NULL


## 4 levels of for loops

for (art in arts) {
   for (age in levels(cooldata$Age_cat)) {
      for (joindate in levels(cooldata$Joindate_cat)) {
         for (visit_freq in levels(cooldata$Visits_freq_cat)) {

            t = t.test(coupon0[coupon0$Age_cat == age,]
                       [coupon0$Joindate_cat == joindate,]
                    [coupon0$Visits_freq_cat == visit_freq,]$art,

                    coupon1[coupon1$Age_cat == age,]
                    [coupon1$Joindate_cat == joindate,]
                    [coupon1$Visits_freq_cat == visit_freq,]$art)
           t_data = data.frame(rbind(t_data,
                          list(art, age, joindate, visit_freq,
                    t$p.value, t$estimate[1],t$estimate[2])))

            colnames(t_data) = c("Art_type", "Age_cat", "Joindate_cat", "Visit_freq_cat",
"P_value","Est_mean_NonCoupon","Est_mean_Coupon")
         }
      }
   }
}
```

Then I get a new dataset like this.

Exhibit 6 T-test results list

Item (4) Customer segmenting (3*3*2) To filter(<.01) T-test x & y (difference)

X	Art_type	Age_cat	Joindate_cat	Visit_freq_cat	P_value	Est_mean_NonCoupon	Est_mean_Coupon
1	Jewelry	elder	new	high	4.272747e-01	272.8844	259.5878
2	Jewelry	elder	new	low	5.209378e-03	281.2410	251.2789
3	Jewelry	elder	new	medium	1.743256e-06	282.0920	220.8128
4	Jewelry	elder	old	high	1.581697e-02	264.6331	213.9504
5	Jewelry	elder	old	low	1.271298e-01	270.7151	253.6285
6	Jewelry	elder	old	medium	1.372562e-01	276.5434	254.9521

The data contains not only the characteristics (like age, joindate, etc.) of this customer segment, but also the t-test results (like p-value and the mean of t-test). Based on these data, we implemented the following analysis to check the coupon effects on the spending of these customers (significant or insignificant, positive or negative).

Set the significance level to be 1%, we can get 49 remaining t-test that represent significant coupon effects on customer spending.

Before we dive into the final revenue, we can analyze at a single item spending level. In this way, we can test customer purchasing power, coupon sensitivity, and customer aesthetics.

Part IV: T-test Analysis (spending level)

One dimension: Age category

Here I use bar charts to describe the single item spending btw different age groups.

It's clear to see that there are [missing values] for the [Painting and Sculpture] spending of elder group. We can infer that coupon has no significant increasing or decreasing effects on elder customers' spending on these two categories.

There will be more missing values in the following visualizations, and they may bring us similar findings.

Btw, for the values that are not missing, coupon exerts negative effects on elder people. We can simply conclude that do not send coupons to elder customers in this case.

Exhibit 7 Bar charts of single item spending between different Age group

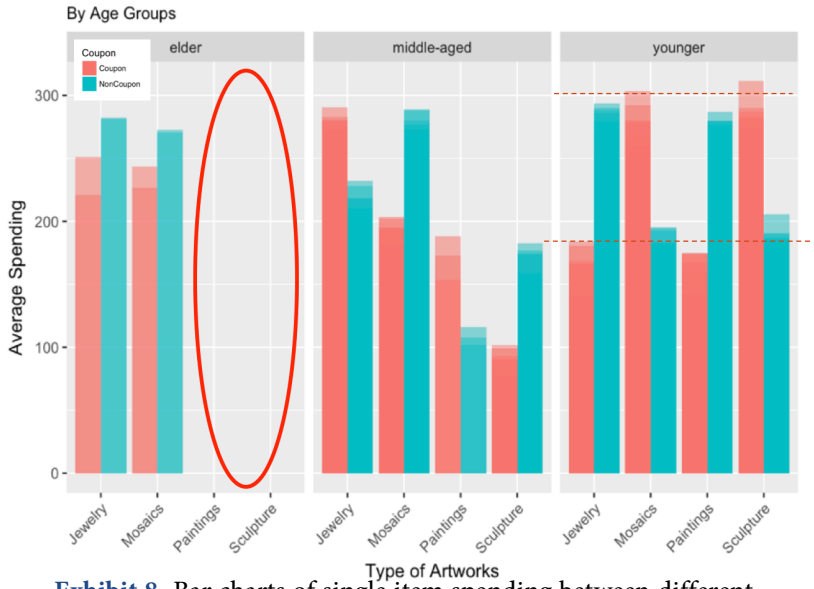
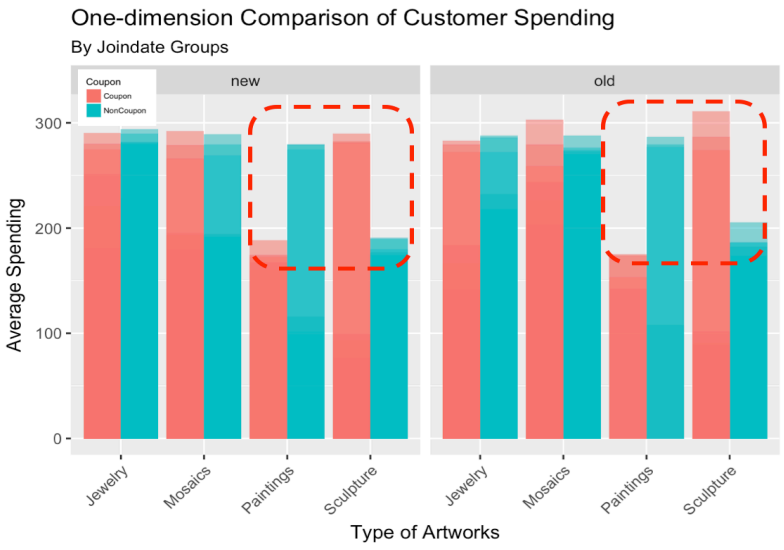


Exhibit 8 Bar charts of single item spending between different Join date group



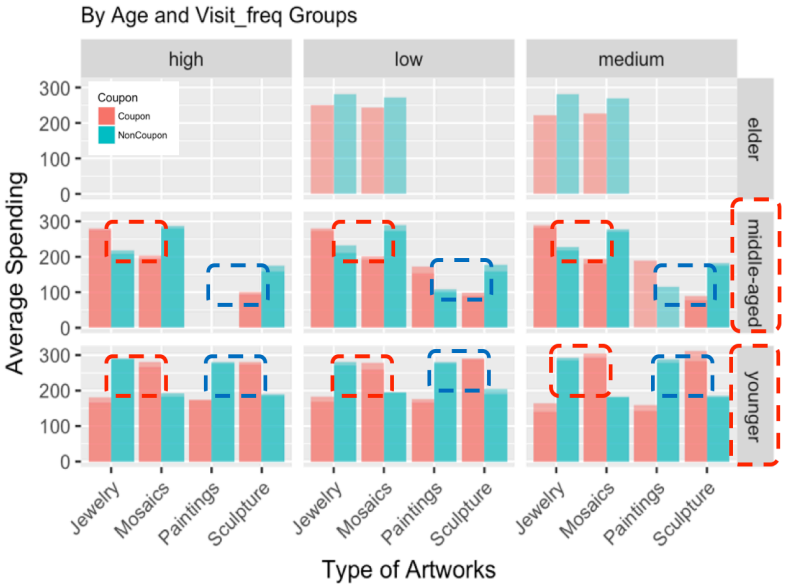
One dimension: Joindate category

Change the dimension to joindate category, no much difference between new and old users. But we can see very clear reversed coupon effects on paintings and sculpture

Two dimensions: Age and Visit frequency

Then I do two-dimensional to visualize more precisely. Taking Age category and Visit Frequency category as an example, we can find some magically reversed effects on young and middle-aged. This could be simply checked by the relative height of green (non-coupon) bars and red (coupon) bars. So, it's really necessary to spend more effort to design a smart coupon plan.

Exhibit 9 Bar charts of single item spending between different Age and Visit frequency group



Then go deeper to see the effects on the commission level. This may help to find guidelines for future coupon promotions, as well as to maximize revenue.

Part V: T-test Analysis (commission level)

The most responsive and contributing purchase

With which function, we can easily iterate the t-test results and find the biggest change in commission. In this way, we can find the most sensitive and responsive group and its corresponding single type of artwork.

Exhibit 10 The most responsive and contributing purchase

```
# find the most sensitive segement that comes with biggest commission increase
com_data[which.max(com_data$commission_change),]
```

```
##      X  Art_type Age_cat Joindate_cat Visit_freq_cat      P_value
## 49 72 Sculpture younger          old         medium 4.509981e-09
##    Est_mean_NonCoupon Est_mean_Coupon spending_change commission_change
## 49           185.8996        311.3574        125.4578          15.05494
```

Most contributing item

Most contributing group characteristic

Commission increased by 10% coupon per person

The results are shown as above. That's to say, among all the 18 segmented subgroups, # younger(below 25) + old(joined later than Nov 2014) + medium user (monthly 4-8 visits) customers # are most sensitive to buy more sculpture with coupon promotion.

Positive change on average commission (Three dimensions)

To see the heat map of the commission change in different subgroups. In this graph, I use all the data that brings positive spending change. It's clear to see sculpture and mosaics benefit most from this coupon. With coupon stimulating, almost every sub-groups buy more, except for middle-aged group. It's also worth mentioning that the behavior changes dramatically as yon grow older. As we can see from the rightmost part, the pattern here between younger and middle-aged are totally different.

Exhibit 11 Heat map of positive commission change between four artworks and 18 customer segments

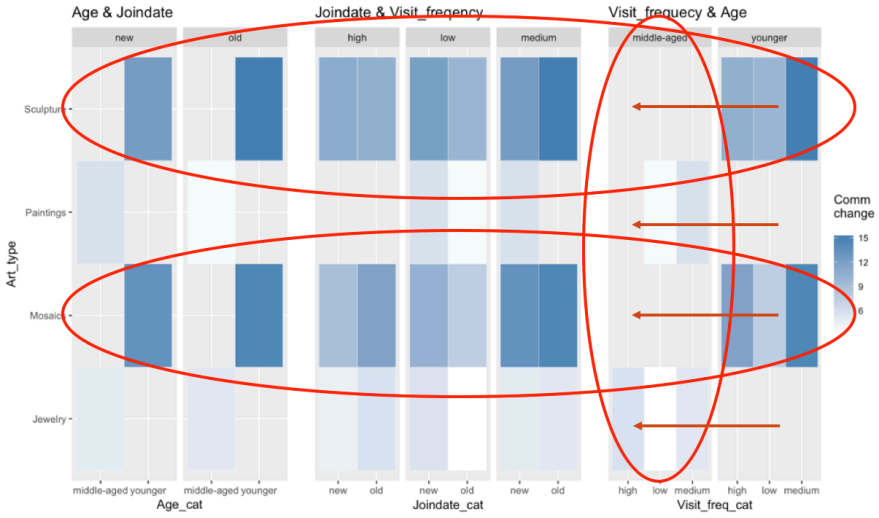
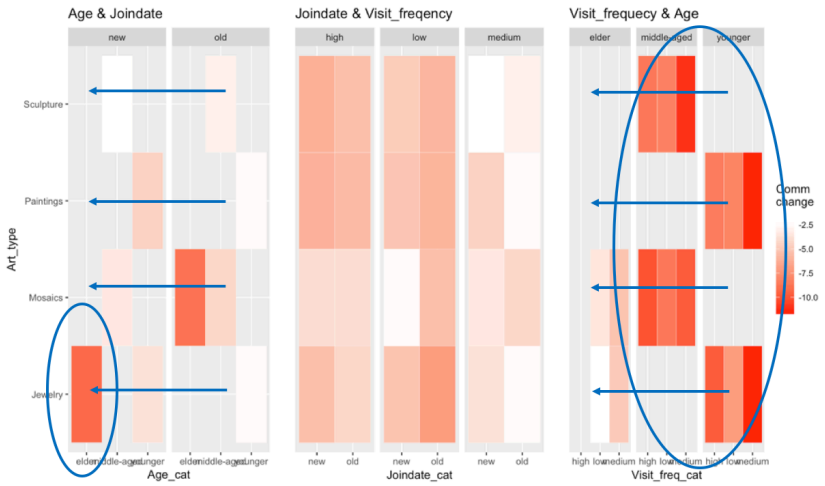


Exhibit 12 Heat map of negative commission change between four artworks and 18 customer segments



Negative change on average commission

To see from another side, I plot all the negative commission change. We can use similar reasoning and prove that the purchasing and sensitivity pattern are totally different between the two subgroups we mentioned above. It's worth mentioning that the behavior changes significantly from youngers not only to middle-aged, but also to elder customers.

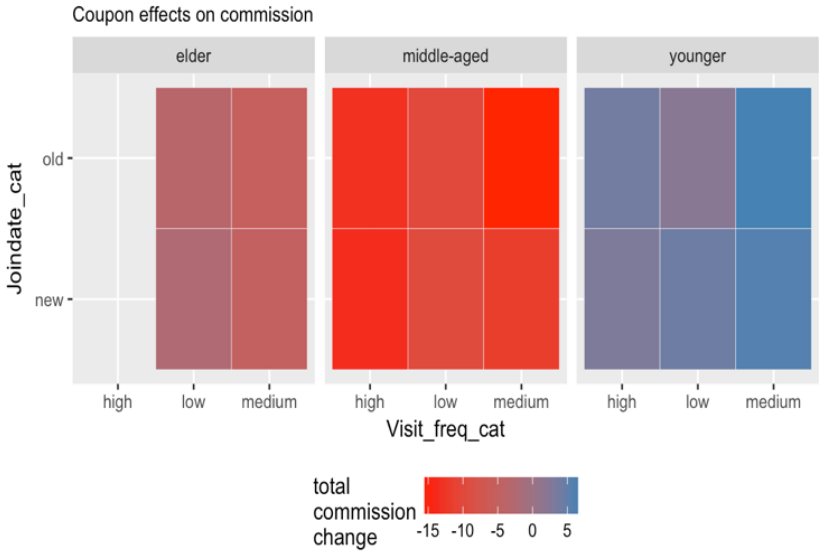
Aggregated commissions change (three dimensions)

To see from the level of aggregated commission, I plotted this heat map. It's clear that younger customers (the blue tiles) are the only subgroup that are positively boosted by this coupon, elders are kind of neutral and slightly negative, and

middle-age customers are more easily to take advantage of coupon promotion. Which means, this segment may fail our coupon promotion plan.

With regard to different join date level, there is no much difference between the upper row and lower row. The same thing happens on the visit frequency level. Anyways, these two factors are significant to affect the natural spending, but there is no much internal difference.

Exhibit 13 Heat map of aggregated commission change between 18 customer segments



Part VI: Suggestions

1. Spot outstanding subgroup characteristics

   The company should follow the nature of different age group during promotion. All in all, Local Art should target the young, avoid the middle, and comfort the old.

2. Conditional coupon tips

   If the company want to design conditional coupon. Here are some tips that they should bear in mind.

   a. Stimulate on sculpture and mosaics purchase

   b. Reversed effects between young and middle-aged

   c. Aesthetic changes between different age groups

3. Cultivate more young customers

   Because younger generations have strong purchasing power and they are easily motivated by 10% coupon in this case, Local Art should never give up cultivate this user group and spend more time designing how to make this segmentation marketing.

Part VII: Summary

Statistical techniques

In this case, I mainly explore the data with A/B test (t-test) method. I am not very sure if the way I segment the data is statistically correct. And I even plot all the spending curves grouping them by the same sub-category, and I find only few of them are occasionally normally distributed. I cannot find a better way to solve the absence of this key assumption, but I still conducted the T-test as suggested by the case.

Why not do clustering? Actually, I tried to use unsupervised clustering to create subgroups. But what prevented me from digging this is that I cannot accurately describe the demographical characteristics of the best-four clusters. There might be many overlaps (like age, visit frequency) between them and bias will follow up if we try to find and target a certain cluster.

Statistical insights

Most findings are talked in the Part VI. One thing could be done in the following analysis would be to segment the 5000 customers from another dimension. I would suggest to dig deeper into the below total spending (aggregated spending of four types of artworks per person) density distribution.

It's easy to find that the natural density area (red, no coupon) has three peaks and for some reason, the middle one shifts rightward and the other two are shifting leftward. There might be significant coupon effects on different total spending ranges. If given more time, it would be great to explore this.

Exhibit 13 Density distribution of aggregated spending per person and the guess of their moving trend