香　港　大　學

THE UNIVERSITY OF HONG KONG

STAT 3613 Marketing Analytics

Group Project Report

PICO VR Headset Market Analysis

Course Co-ordinator: Dr. CW Kwan

Group Member and Contribution:
Wang Ziyu 3035777547 (40%)
Wu Honghao 3035719040 (30%)
Jokinen Ilari Tapio 3035998701 (30%)

1. Introduction:

What's the greatest invention in the 21st century? Some people think that practical artificial intelligence is a great invention. However, the millennials wouldn't think so. The top three important inventions are smartphones, Wi-Fi, and air-conditioners. The reason behind this is that most youngsters prefer playing with their phones while staying in their air-conditioned houses. Playing phones generally means playing video games through their phones. Therefore, we see great potential in the gaming industry. But we need to try to confirm if the market potential is as great as we expected. Among all forms of gaming experiences, we believe playing games in virtual reality is one of the most stimulating, which is why we chose to analyze the market of VR gaming equipment. There are some known issues with the existing VR techniques such as motion sickness for some people, high costs, heaviness, and the relative scarcity of available games compared to conventional gaming platforms. The above issues show us that improvements can be made to VR techniques. Interested in the market potential, we propose to design and launch a questionnaire to collect data according to the variables of our study. Then we will proceed with three analyses, which are factor analysis, cluster analysis, and regression analysis, on the data collected. We believe factor analysis can be used to reduce the dimensions for further analysis and to observe correlations between customer preferences for certain features of VR gaming equipment. By utilizing clustering analysis, we could understand whether different identifiable categories or archetypes of customers and customer needs exist, which would improve our further marketing efforts. Through logistic regression, we will gain a more quantitative and accurate understanding of the market potential of our proposed product. At the end of the study, we will discuss the limitations of our study.

2. Method

Our primary target group is young adults who are interested in VR gaming, and we believe the subset of people with previous experiences with VR gaming equipment to be of special importance. Young adults are targeted because they simultaneously represent the future direction for trends in the gaming industry in terms of customer preferences and are also capable of making purchasing decisions independently. By asking questions about whether the respondents are interested in VR gameplay, we can segment them into groups. We believe keeping the target group relatively broad is justified due to the large potential market, the probing nature of the study, and the fact that targeting people with extensive experience with VR equipment risks resulting in an insufficient sample size with our available resources.

We then chose to develop a series of products based on Pico 4 under ByteDance, as we are focusing mainly on the mainland China market. Oculus headsets might be more successful in terms of global market share, but Facebook does not sell this type of headset in the Chinese market. We have tentatively decided to enhance the users' game experience by decreasing the weight (which is not satisfied by Pico 4's customers) to only 100g and increasing the number of qualified games we give away. Other than that, increasing the resolution or making game sets tailored for a certain group (like people with myopia) can be our future direction.

The sampling method is designed to be convenient sampling and snowballing sampling. The rationale behind this is that, firstly, we think that probability sampling like simple random sampling and stratified sampling is not feasible for our study. For example, the simple random sample requires a sampling frame with all target populations in the population, which might not be possible for us. Also, our survey range could be relatively small. It is not particularly meaningful to conduct methods like systematic sampling. Then, we thought of non-probability sampling, which is practicable with limited resources. Among these, we find convenient sampling and snowballing sampling the most accessible sampling methods for us. They are cheaper and less time-consuming, although we must assume that our sample is representative of the population. Our survey does include questions that allow us to observe the representativeness and relevance of the respondents, i.e., interest in VR gaming and ownership of gaming platforms.

The data collection procedures start with designing and launching an online questionnaire through our social media accounts (via WeChat friend circle) so that people we know might want to help us fill up the questionnaire. The rationale is that the online questionnaire can be easier to spread and more convenient for respondents to answer. Meanwhile, since our questions might include sensitive questions about income and age, this kind of data collection method can help us avoid biased responses better than interviews.

The data collection duration is around a month. Due to limited staff and budget, we only collected 60 questionnaires from our target population to decrease the margin of error as much as possible. For better illustration, a sample questionnaire is attached in the appendix of this report.

After setting the sampling design, we consider the statistical models that should be used to achieve our objectives. First, we will use factor analysis to reduce the dimensions for further analysis. We currently have many variables and some of them like monthly income and willingness to pay or refresh rate and resolution might be correlated. These more correlated variables might be removed to make our regression result more stable and allow us to understand customer needs better.

Next, we will apply cluster analysis to the factor scores (importance rating of the headset attributes) and relate the result to other cluster questions like gender. By adopting this analysis, we attempt to improve our understanding of the customers and their needs, which would benefit our further marketing efforts.

Moreover, logistic regression would be adopted to analyze the binary variables. The derived factors from the factor analysis are the independent variable, while the outcome variable will be predicted accordingly. By doing so, we will gain a more quantitative and accurate understanding of our potential target group in the future. Thus we can observe which features of the product the respondents place the highest actual monetary value on. For example, people who value graphics the most might be willing to buy the most, possibly indicating that graphics are perceived as a key feature that influences purchasing decisions.

3. Results

We then identify a series of variables for our study, and they are categorized as follows.

Quantitative variables:

1. Monthly available income
   This variable could help us to cluster our respondents. Since VR sets are generally relatively expensive, the assumption is that people with higher incomes will be more willing to enhance their game experience.
2. Willingness to pay
   Since the income is not necessarily how much the customers are willing to pay for the VR headset, directly asking questions about this variable might give us more accurate data for our future pricing strategy.
3. Age
   We might think that only young people would use VR, but we would like to find out.

Ordinal variables:

The perceived importance of the following attributes of a VR headset are summarised and labeled in the below table for better illustration.

Variable	Description
x1	Price
x2	Available games and quality
x3	Type of VR headset
x4	Weight
x5	Refresh rate
x6	Resolution
x7	Storage
x8	Screen type
x9	Battery life
x10	CPU (Central Processing Unit)
x11	Brand
x12	Position tracking
x13	Tactile feedback
x14	Compatibility of myopia lens

These are the typical attributes of a VR headset. By conducting later factor and cluster analysis, we aim to identify which sets of features are correlated and thus could be assessed to fulfill similar customer needs.

Nominal variables:

1. Gender

This can be an important cluster characteristic because we typically assume that the male would care more about game and game experience than the female; we can later find out if this is a stereotype or produce VR sets tailored for each group.

2. Wearing glasses or not

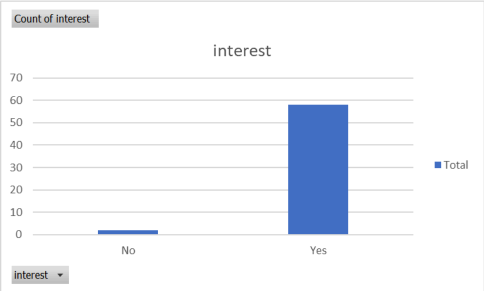
   For people with myopia, if the VR headset is not properly made, the game experience will be horrible. We can use it to cluster the respondents and to see if we should make VR sets improving the game experience for people with myopia.

After identifying the variables for our study focus, we then look at the basic statistics of our data.

Table 1 & Figure 1

Whether the respondents are interested in VR gaming

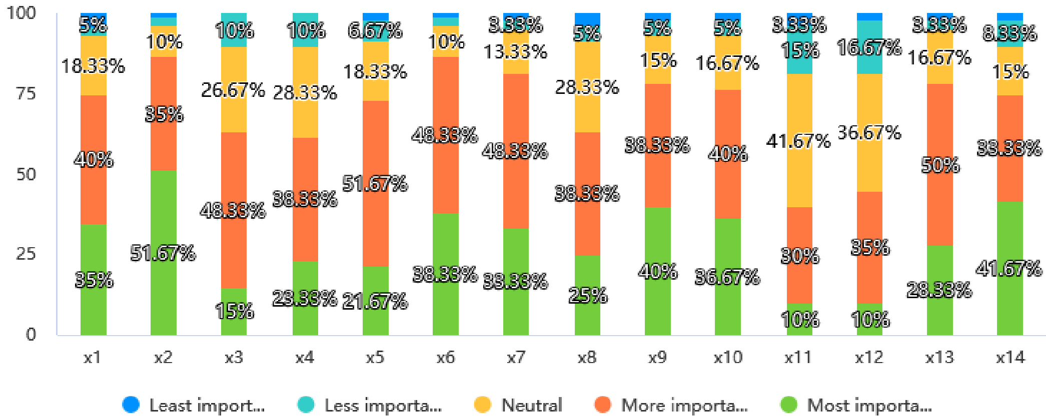
Row Labels	Count of interest
No	2
Yes	58
Grand Total	**60**



The response results show that almost every respondent is interested in VR equipment, while only 2 of them are not.

Table 2 & Figure 2

Summary statistics of the perceived importance of the attributes of a VR headset



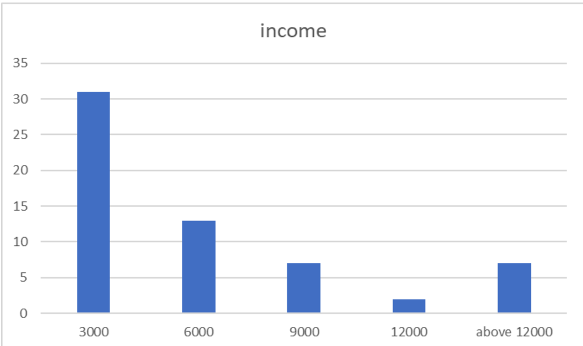
Vaiable\Score	1	2	3	4	5	Mean score	Standard deviation
x1	3	1	11	24	21	3.98	1.0332
x2	1	1	6	21	31	4.33	0.8570
x3	0	6	16	29	9	3.68	0.8535
x4	0	6	17	23	14	3.75	0.9320
x5	1	4	11	31	13	3.85	0.8987
x6	1	1	6	29	23	4.20	0.8193
x7	1	2	8	29	20	4.08	0.8693
x8	2	3	17	23	15	3.77	0.9977
x9	1	3	9	23	24	4.10	0.9514
x10	1	3	10	24	22	4.05	0.9464
x11	2	9	25	18	6	3.28	0.9583
x12	1	10	22	21	6	3.35	0.9356
x13	1	2	10	30	17	4.00	0.8636
x14	1	5	9	20	25	4.05	1.0321
Grand total	16	56	177	345	246	3.89	12.9482

Every feature we propose of the VR equipment can be considered significant since the means of the importance of features are all greater than 3. By observing the above visuals, x2 (Available games and quality) and x6 (Resolution) are considered the most important (highest mean scores) with the least disagreements (least standard deviation). Therefore, they should be valued in further analysis. Whereas x11 (brand) of the equipment is the least important and x12 (position tracking) is the second least important feature of the equipment.

Table 3 & Figure 4

Monthly income (RMB) of the respondents

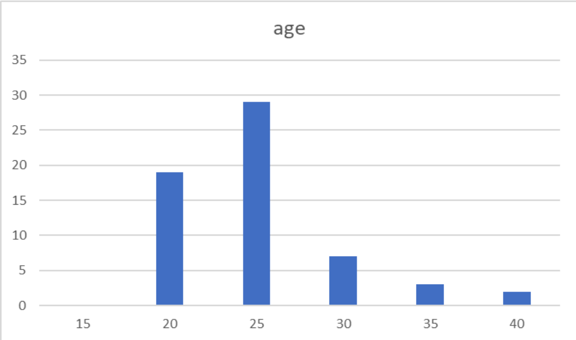
income range	frequency	average	sd
3000	31	3000	0
6000	13	6000	0
9000	7	9000	0
12000	2	12000	0
above 12000	7	above 12000	0
grand total	60	28290.8333	154828.4



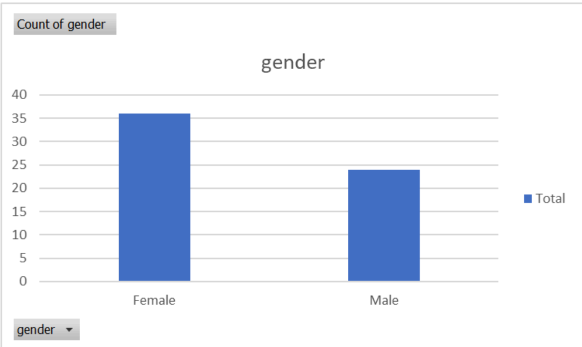
Although the average income of respondents is around 28000, most of them have a monthly income that is less than 6000.

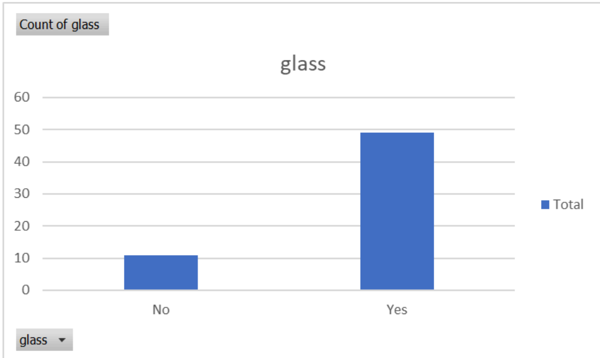
Figure 5

Demographics (gender, age, wearing glasses or not) of the respondents



age range	frequency	average	sd
15	0	15	0
20	19	20	0
25	29	25	0
30	7	30	0
35	3	35	0
40	2	40	0
grand total	60	23.15	4.664416

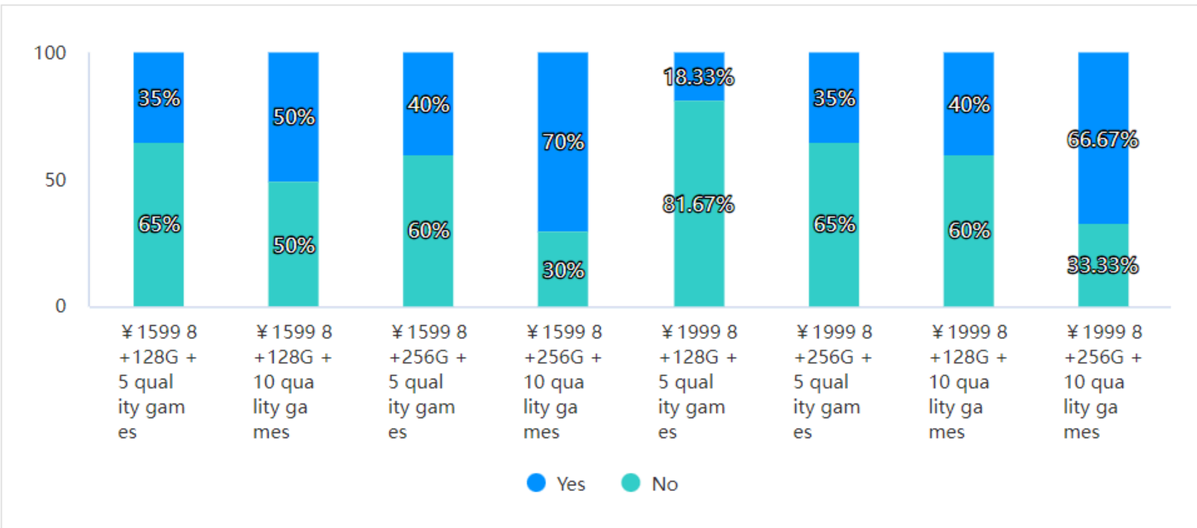




The age range is mostly from 20 to 30. We have more female respondents than males. Around 80% of respondents wear glasses, we infer that they will value the attribute (compatibility of myopia lens) more than the others.

Table 4 & Figure 6

Respondents' willingness to buy the eight VR items



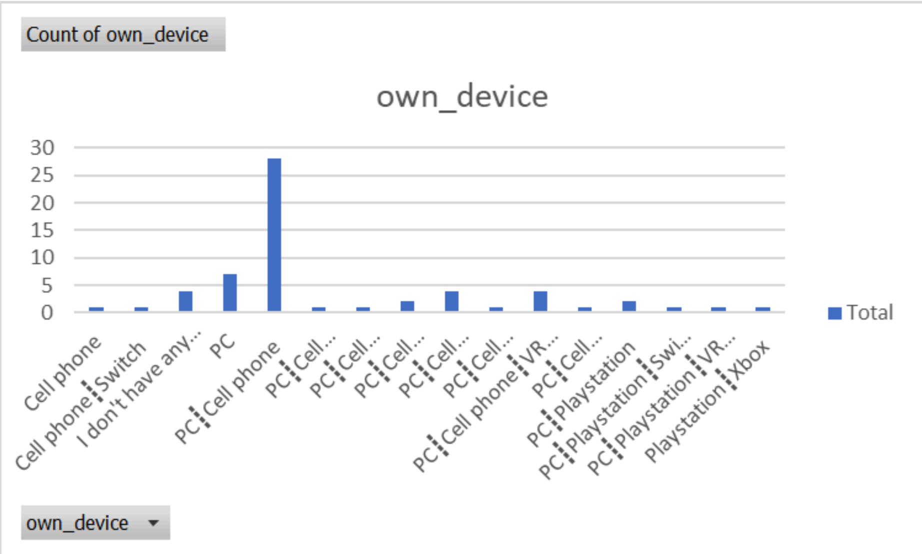
Items\will buy or not	Yes	No
￥1599 8+128G + 5 quality games	21(35%)	39(65%)
￥1599 8+128G + 10 quality games	30(50%)	30(50%)
￥1599 8+256G + 5 quality games	24(40%)	36(60%)
￥1599 8+256G + 10 quality games	42(70%)	18(30%)
￥1999 8+128G + 5 quality games	11(18.33%)	49(81.67%)
￥1999 8+256G + 5 quality games	21(35%)	39(65%)
￥1999 8+128G + 10 quality games	24(40%)	36(60%)
￥1999 8+256G + 10 quality games	40(66.67%)	20(33.33%)

If we set the criteria of a potential product that more than half of the respondents are willing to buy, then the models that meet those criteria are 1999_8_256G_10 (40 Yes), 1599_8_256G_10 (42 Yes), 1599_8_128G_10 (30 Yes). We should set the price of the 8_256_G_10 to 1999 definitely (2 less number of sales trading 400 more margins).

Table 5 & Figure 7

Respondents' own gaming sets

Row Labels	Count of own_device
Cell phone	1
Cell phone ┆ Switch	1
I don't have any gaming devices	4
PC	7
PC ┆ Cell phone	28
PC ┆ Cell phone ┆ Others	1
PC ┆ Cell phone ┆ Playstation ┆ Switch ┆ VR equipment	1
PC ┆ Cell phone ┆ Playstation ┆ Switch ┆ Xbox	2
PC ┆ Cell phone ┆ Switch	4
PC ┆ Cell phone ┆ Switch ┆ VR equipment	1
PC ┆ Cell phone ┆ VR equipment	4
PC ┆ Cell phone ┆ Xbox ┆ VR equipment	1
PC ┆ Playstation	2
PC ┆ Playstation ┆ Switch	1
PC ┆ Playstation ┆ VR equipment	1
Playstation ┆ Xbox	1
Grand Total	**60**



Most of the respondents have only personal computers and cell phones as their gaming devices. While PCs are actually the most common gaming device according to our response (53 out of 60 respondents own a PC as a gaming device).

Table 6

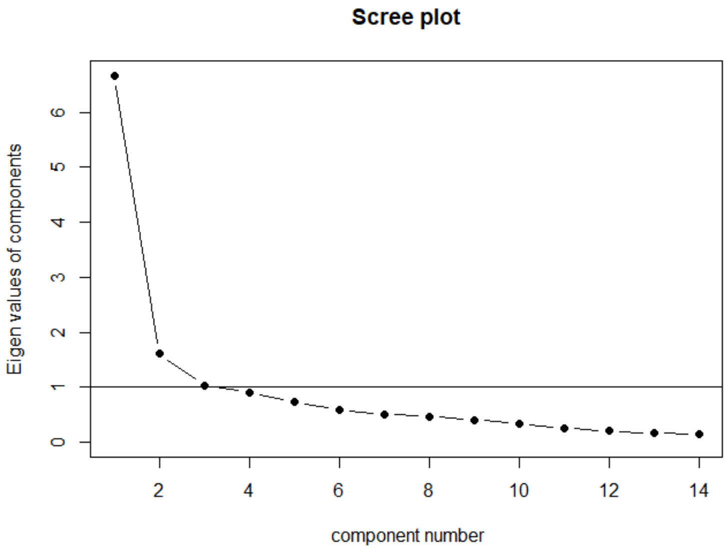
Correlation between features

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14
x1	1.000	0.581	0.398	0.383	0.435	0.384	0.304	0.424	0.433	0.330	0.279	0.041	0.304	0.398
x2	0.581	1.000	0.541	0.403	0.462	0.435	0.485	0.489	0.624	0.606	0.193	0.169	0.366	0.383
x3	0.398	0.541	1.000	0.517	0.445	0.407	0.447	0.509	0.520	0.502	0.298	0.162	0.391	0.365
x4	0.383	0.403	0.517	1.000	0.501	0.488	0.591	0.428	0.545	0.437	0.213	0.102	0.253	0.225
x5	0.435	0.462	0.445	0.501	1.000	0.709	0.515	0.698	0.672	0.507	0.247	0.043	0.328	0.447
x6	0.384	0.435	0.407	0.488	0.709	1.000	0.547	0.722	0.648	0.424	0.272	0.084	0.455	0.649
x7	0.304	0.485	0.447	0.591	0.515	0.547	1.000	0.668	0.666	0.551	0.358	0.297	0.429	0.298
x8	0.424	0.489	0.509	0.428	0.698	0.722	0.668	1.000	0.721	0.569	0.531	0.252	0.551	0.555
x9	0.433	0.624	0.520	0.545	0.672	0.648	0.666	0.721	1.000	0.634	0.284	0.074	0.351	0.513
x10	0.330	0.606	0.502	0.437	0.507	0.424	0.551	0.569	0.634	1.000	0.264	0.324	0.456	0.396
x11	0.279	0.193	0.298	0.213	0.247	0.272	0.358	0.531	0.284	0.264	1.000	0.587	0.430	0.242
x12	0.041	0.169	0.162	0.102	0.043	0.084	0.297	0.252	0.074	0.324	0.587	1.000	0.441	0.140
x13	0.304	0.366	0.391	0.253	0.328	0.455	0.429	0.551	0.351	0.456	0.430	0.441	1.000	0.418
x14	0.398	0.383	0.365	0.225	0.447	0.649	0.298	0.555	0.513	0.396	0.242	0.140	0.418	1.000

If we consider two features are correlated when the correlation is greater than 0.5, and highly correlated when the correlation is greater than 0.7. Then the price is correlated with the quality and number of the games(0.581). The quality and number of the games are correlated with the type of VR equipment(0.541), type of screen(0.624), and battery life(0.606). The type of VR equipment is correlated with the weight of the equipment(0.517), the type of screen(0.509), the battery life(0.520), and the CPU(0.502). The weight is correlated with the refresh rate(0.501), the device memory(0.591), and the battery life(0.545). The refresh rate is highly correlated with the resolution(0.709). The refresh rate is correlated with the device memory(0.515), the type of screen(0.698), the battery life(0.672), and the CPU(0.507). The resolution is correlated with the device memory(0.547), the battery life(0.648), and the compatibility to the myopia lens(0.649). The resolution is highly correlated with the type of screen(0.722). The device memory is correlated with the type of screen(0.668), the battery life(0.666), and the CPU(0.551). The type of screen is highly correlated with the battery life(0.721) and is correlated with the CPU(0.569) and the brand of the device(0.531). The battery life is correlated with the CPU(0.634) and the compatibility with the myopia lens(0.513). The device brand is correlated with the location tracking(0.587).

After summarizing the basic statistics of our data, we begin to apply factor analysis and cluster analysis. We consider a factor analysis using the principal component method to the values of ‘x1’ – ‘x14’, but first, we preliminarily decide the number of factors we should use with multiple criteria.

Figure 7
Scree plot



By the scree plot, we will suggest 2 factors as an elbow is observed at the 3rd eigenvalue.

Figure 8
Statistics of the factor models

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14
SS loadings	6.67	1.60	1.02	0.91	0.73	0.59	0.51	0.46	0.40	0.33	0.26	0.20	0.16	0.15
Proportion Var	0.48	0.11	0.07	0.06	0.05	0.04	0.04	0.03	0.03	0.02	0.02	0.01	0.01	0.01
Cumulative Var	0.48	0.59	0.66	0.73	0.78	0.82	0.86	0.89	0.92	0.94	0.96	0.98	0.99	1.00
Proportion Explained	0.48	0.11	0.07	0.06	0.05	0.04	0.04	0.03	0.03	0.02	0.02	0.01	0.01	0.01
Cumulative Proportion	0.48	0.59	0.66	0.73	0.78	0.82	0.86	0.89	0.92	0.94	0.96	0.98	0.99	1.00

By the latent root criterion, the eigenvalues for the 3rd factor > 1 while that of 4 the factors are <1. Therefore, 3 factors are suggested. Then we consider at least 80% of the cumulative proportion of variances explained. By observing the above table, we can see that up to factor 6, the cumulative proportion of variances explained is 82%, which is greater than 80%, while in factor 5, the cumulative proportion of variances explained is 78% <80%. Therefore, we should consider 6 factors.

We also apply a factor analysis with the maximum likelihood method to the values of ‘x1’ – ‘x14’ to decide the number of factors we should use. For the 2-factor model, the p-value

(0.0298) is smaller than 0.05, so we might conclude that the 2 factors are not sufficient. Meanwhile, three factors are sufficient as the p-value (0.241) here is greater than 0.05.

To furtherly decide the number of factors that we should use, we apply both the 2-factor and 3-factor model using the maximum likelihood method to ‘x1’ –‘x17’ and rotate the factor by the varimax method.

Figure 9

The rotated factor loadings of factor 2

```
Factor Analysis using method =  ml
Call: fa(r = vr[, 4:17], nfactors = 2, rotate = "varimax", scores = "regression",
    fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
         ML1  ML2   h2   u2 com
price    0.53 0.09 0.29 0.71 1.1
game     0.63 0.19 0.44 0.56 1.2
type     0.58 0.22 0.39 0.61 1.3
w        0.60 0.12 0.38 0.62 1.1
rate     0.80 0.07 0.64 0.36 1.0
resol    0.79 0.12 0.64 0.36 1.0
stora    0.67 0.34 0.56 0.44 1.5
screen   0.80 0.34 0.76 0.24 1.4
battery  0.86 0.12 0.75 0.25 1.0
cpu      0.62 0.35 0.50 0.50 1.6
brand    0.25 0.66 0.50 0.50 1.3
pos     -0.03 0.89 0.80 0.20 1.0
tac      0.42 0.52 0.44 0.56 1.9
myopia   0.59 0.17 0.37 0.63 1.2


                ML1   ML2
SS loadings    5.426 2.036
Proportion Var 0.388 0.145
Cumulative Var 0.388 0.533
```

We can observe that the total variance explained for the 2-factor model is 53.3% (=38.8% + 14.5%).

Figure 10

The rotated factor loadings of factor 3

```
Factor Analysis using method =  ml
Call: fa(r = vr[, 4:17], nfactors = 3, rotate = "varimax", scores = "regression",
    fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
     ML3   ML1  ML2   h2   u2 com
x1  0.48  0.29 0.07 0.32 0.68 1.7
x2  0.77  0.20 0.11 0.64 0.36 1.2
x3  0.58  0.28 0.18 0.45 0.55 1.7
x4  0.53  0.34 0.07 0.40 0.60 1.8
x5  0.44  0.68 0.05 0.66 0.34 1.7
x6  0.32  0.79 0.11 0.74 0.26 1.4
x7  0.53  0.44 0.31 0.56 0.44 2.6
x8  0.40  0.73 0.36 0.83 0.17 2.1
x9  0.65  0.57 0.08 0.76 0.24 2.0
x10 0.68  0.25 0.29 0.61 0.39 1.6
x11 0.09  0.27 0.71 0.58 0.42 1.3
x12 0.11 -0.09 0.83 0.72 0.28 1.1
x13 0.28  0.34 0.52 0.46 0.54 2.3
x14 0.27  0.56 0.16 0.41 0.59 1.6


                      ML3  ML1  ML2
SS loadings           3.23 3.03 1.88
Proportion Var        0.23 0.22 0.13
Cumulative Var        0.23 0.45 0.58
```

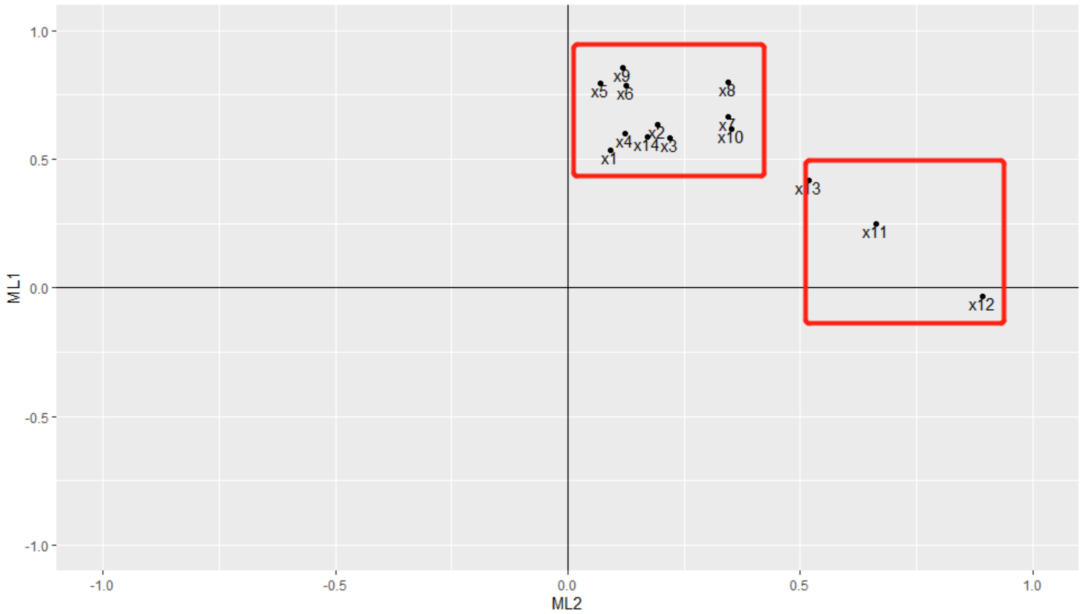
Similarly, the total variance explained for the 3-factor model is 58% (=23% + 22% + 13%).

We can see that both the 2 factors and the 3 factors account for a moderate to a large percentage of the total variance, so we might not be so satisfied with the fitness. However, the variance explained doesn't increase a lot when adding one more factor (4.7% only). Considering that the purpose of factor analysis is dimension reduction, we chose to apply the 2-factor model for further analysis instead of the 3-factor model.

Then, by constructing the loading plot, we can actually observe 2 groups of variables, although most of the variables are only explained moderately well (as the 2-factor model can only explain 53.3% of the variances).

Figure 11

The loading plot of factor 2



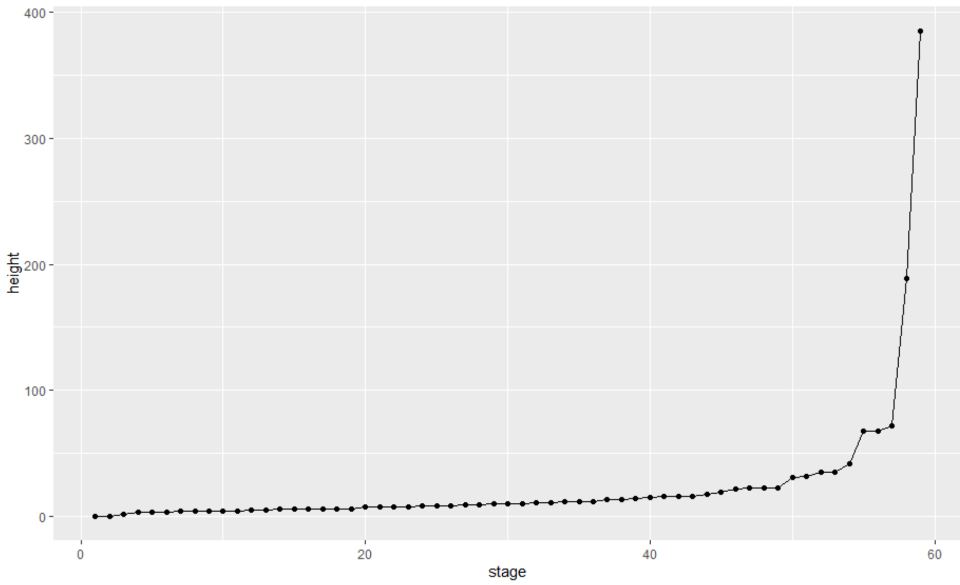
For factor ML2:
Brand (x11), Position tracking (x12), and Tactile feedback (x13) define factor ML2 (moderate to high loadings on factor ML2, small or negligible loadings on factor ML1). The last two are very related to the higher level of gaming experience, while the brand to some extent could also determine the experience. Therefore, we would summarize this ML2 factor as the high level of the gaming experience factor.

For factor ML1:
Price (x1), Available games and quality (x2), Type of VR headset (x3), Weight (x4), Refresh rate (x5), Resolution (x6), Storage (x7), Screen type (x8), Battery life (x9), CPU (x10), Compatibility of myopia lens (x14) defines factor ML1 (moderate to high loadings on factor ML1, small or negligible loadings on factor ML2). These variables are more related to the basic/essential aspects of the VR equipment, so ML1 is the basic level of the gaming experience factor.

We are ready to perform the cluster analysis, but we want first to identify the appropriate number of clusters we should use.

Figure 12

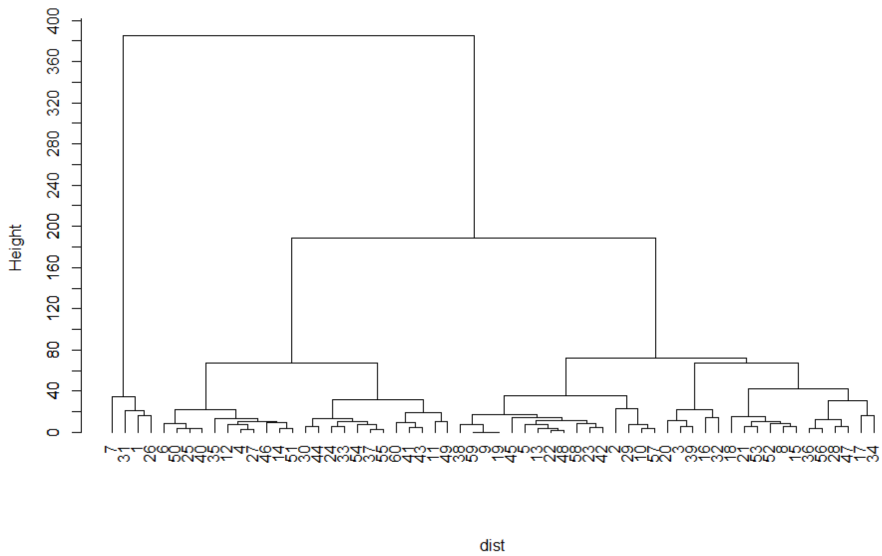
The between-group sum of squares (height)



By looking at the height of the data points from right to left, we can observe a relatively large increase from 3 to 2 clusters, so a 3-cluster solution is suggested.

Figure 13

The dendrogram



According to the dendrogram, it also appears that the first 3 clusters are formed at very low heights, and the further merging of these clusters results in a large merging height. Therefore, we suggest having 3 clusters.

Based on the values of 'x1' – 'x14', we apply Ward's method for a 3-cluster solution. Then we use the cluster means as the initial seeds for a K-means method with the 'Hartigan-Wong' algorithm.

Figure 12

The means and the sizes of the clusters by the K-means method

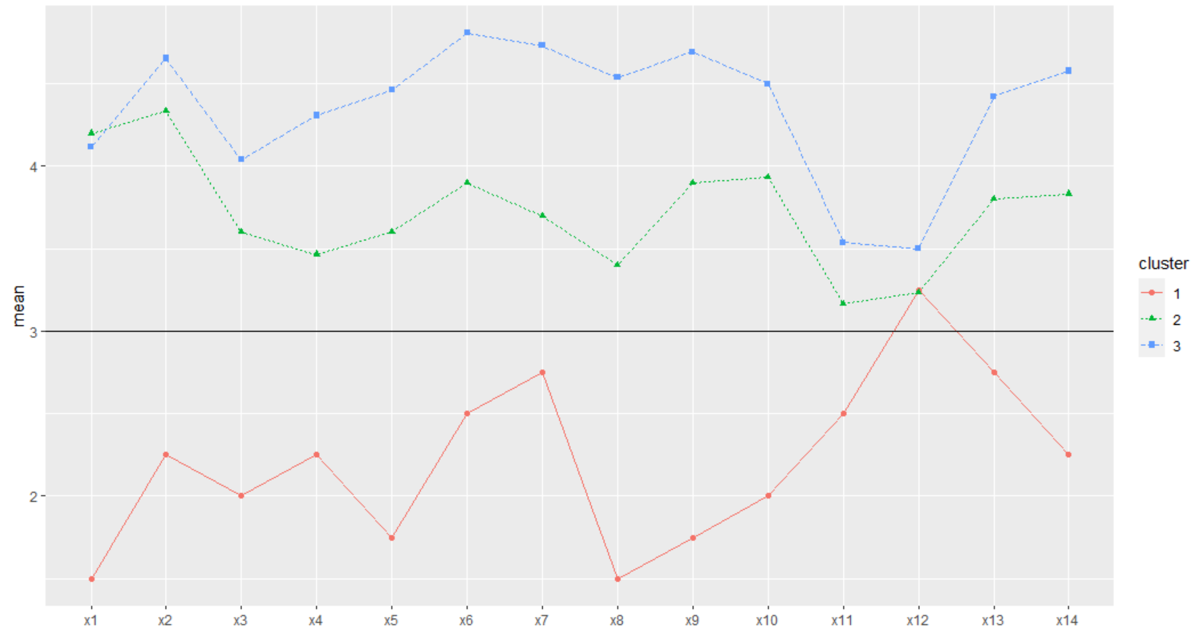
```
K-means clustering with 3 clusters of sizes 4, 30, 26

Cluster means:
        x1       x2       x3       x4       x5       x6       x7       x8       x9      x10
1 1.500000 2.250000 2.000000 2.250000 1.750000 2.500000 2.750000 1.500000 1.750000 2.000000
2 4.200000 4.333333 3.600000 3.466667 3.600000 3.900000 3.700000 3.400000 3.900000 3.933333
3 4.115385 4.653846 4.038462 4.307692 4.461538 4.807692 4.730769 4.538462 4.692308 4.500000
       x11      x12      x13      x14
1 2.500000 3.250000 2.750000 2.250000
2 3.166667 3.233333 3.800000 3.833333
3 3.538462 3.500000 4.423077 4.576923
```

In terms of cluster sizes, cluster 1 only has 4 respondents compared to the other 2 groups, which is the minority. When observing the cluster means, cluster 1 tends to have a lower rating (<3) except for x12 (3.25). For respondents in clusters 2 and 3, they give higher ratings (>3) to all variables, so generally, they consider the attributes as important and are very conscious of the attributes of the VR equipment.

By looking at the clusters, we can preliminarily observe the characteristics of clusters, now we plot the profiles of the clusters, to look in more detail for each cluster.

Figure 13

Profile pot of the 3 clusters

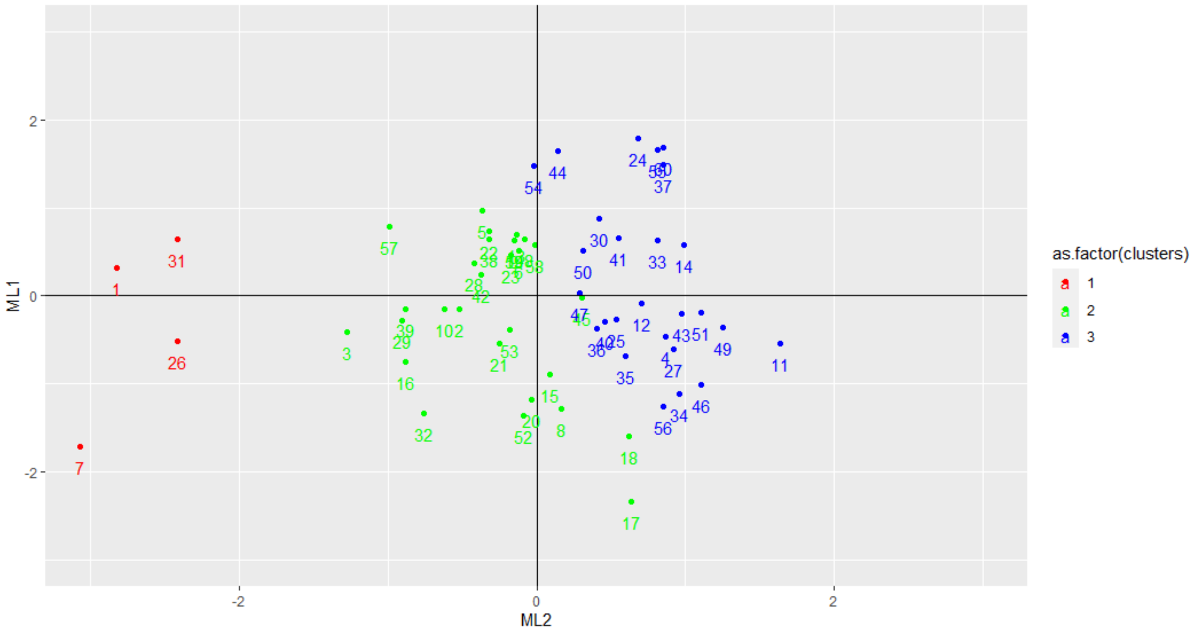


Basically, there are three clusters of respondents. Cluster 1 has a low level of importance rating (<3) on all attributes except for x12 (position tracking), while cluster 2 has a high level of importance rating (>3) on all attributes. Cluster 3 has a high level of importance rating (>3) on all attributes. We also noted that x12 (position tracking) is a relatively less important attribute for clusters 2 and 3, but it is the most important attribute for cluster one, which is interesting to observe.

We then plot the factor scores obtained from the prior factor analysis labeled and colored by the clustering in the cluster analysis.

Figure 14

The plot of factor scores colored by the corresponding clusters



Recall that ML1 is the basic level of the gaming experience factor and ML2 factor is the high level of the gaming experience factor. We can observe in general that most people think that both the basic and the higher level of the gaming experience are important, while most people in cluster 1 don't think that they are important (most of the cluster 1 respondents are the extreme cases in the plot). Then for the specific clusters:

Cluster 1: low on ML1 and low on ML2. People here don't value both the basic and higher level of the gaming experience factor and have different opinions about the basic level of the gaming experience factor. We infer that they might be non-gamers.

Cluster 2: high on ML1 and low on ML2. We infer that they might be entry-level gamers.

Cluster 3: high on ML1 and high on ML2. We think that they might be senior gamers.

Therefore, we could observe that generally, the customers can be grouped by their gaming profile.

After finding the dimensions of the factors and the customer clustering, we would like to apply the logistic regression model further to determine the exact variables that are crucial in the purchasing decision of customers.

Moreover, we will use logistic regression to observe which variables influence the consumers' decision to buy different products. We gave the respondents a set of eight binary choices to buy or not to buy a given product. The eight binary choices were made for eight combinations of three features: price, memory, and number of available high-quality games. Thus, we gained data to employ eight binary logit choice models.

The decision to utilize eight binary choices instead of one multinomial choice model was made because of the expected small sample size: only choosing one product out of eight alternatives hides the data of whether the respondent would have also been willing to buy another combination of features, and with the expected small sample size, we did not want to lose this information. Our model is also still quite easy to interpret from a business perspective, as we get direct data on how many people would be willing to buy each product, and we still get the information from the logit model to understand how different features of the respondents influence their decisions, improving our understanding of the target group.

Our study found that neither the income level, the perceived importance of different features nor other collected data of the respondents were able to predict willingness to buy on a statistically significant level.

When predicting willingness to buy only based on respondents' level of income (8 models of the form "willingness to buy product i ~ income" for each product i), income was not statistically significant for any of the combinations. In terms of the perceived importance of VR features (8 models of the form "willingness to buy product i ~ x1+x2+...+x14"), only for products 1, 4, and 8 was any single variable statistically significant with a p-value <0.05, price for combinations 1 and 4, and screen type for combination 8. For this many variables, we do expect some variables to be statistically significant just by random chance.

P-values below 0.05 are bolded. These were the only variables with p-values <0.05 out of all eight logit models for the eight product combinations, with all variables x1-x14.

Table 7

P-values of variables "price" and "screen type" for product combinations

Product/Variable	Price	Screen type
1	**0.0468**	0.8335
4	**0.0146**	0.1272
8	0.1992	**0.0155**

For the feature preferences, we did additional analysis searching for statistically significant effects of variables by utilizing the built-in step()-function in R. With it, we performed backward selection based on Akaike Information Criterion (AIC). However, the results do not seem to be stable: assuming that respondents have similar decision-making criteria across all products, we would expect the same preferences to be significant over multiple products. However, even with feature selection to decrease the number of variables and increase the significance of the variables left, many features have seemingly varying significance depending on the product. These results are presented in the table below, with rows corresponding to product combinations and columns corresponding to variables on the perceived importance of different VR features. I = variables included in the final set of selected features by the step-function, S = variable is also significant with p-value <0.05 in the final selection of the step-function.

Table 8

Indication of variables in the products that are selected by the step-function or significant

Product/X	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	I		S		I	I					S			
2	I				S		I			I				
3			I		I				I		S			I
4	S		S	I		I		I						
5		S						I			I			
6			S	S								I		
7				S							I			
8	I					I	S	S					I	

From the table, we can try to draw two inferences: first, the perceived importance of features do not have a clear effect on willingness to buy: this can be due to either this study having missed some important decision-making variables, due to the small sample size relative to the number of variables or due to the fact that perceived importance of some features does not significantly influence willingness to buy. Secondly, we can still observe that some variables are not significant in almost any situation, i.e. variables 6, 9, 10, 12, 13, and 14 do not have a significant power of prediction for any of the products even if other variables are removed via backward selection by the step() function. This could be interpreted as follows: whether people do find resolution, battery, CPU, position tracking, tactile feedback, and compatibility of myopia lenses important or not, this does not significantly affect their willingness to buy. Consumers' decision to buy is almost never dependent on the perceived importance of these factors, so while this does not mean that these features themselves would be unimportant, it means that a marketing strategy based on promoting these features would likely be less efficient.

We also calibrated eight binary logit choice models to predict buying decisions based on previously owned gaming devices, interest in VR, gender, having eyeglasses, and age (in R, models were of the form "willingness to buy product i ~ PC + Cellphone + PS + Xbox + Switch + NGD + interest + gender + glasses + age), where PC, Cellphone…NGD were binary factor variables for whether each gaming device was owned or not. Again, none of the models identified statistically significant effects on buying decisions, even implementing backward selection based on AIC with the step() function in R.

4. Limitation

One of the limitations is the response rate. Due to the nature of an online survey spread through social media groups, we do not know who has seen our questionnaire and does not want to fill out the questionnaire. Therefore, we are unable to access the exact response rate, which might bring some uncertainty to our research. Clearly, the sample might also not be representative of the entire market since we were not able to perform a random sample through the entire population. However, for market analysis, we believe the population we were able to reach (mostly university students) to have a higher concentration of respondents in our intended target group.

Preliminarily, the factor analysis and the cluster analysis have shown that there are two factors influencing the customers' perception of the importance of the attributes of the VR sets and they are basically three types of customers (53.3% of the variance explained). However, we still notice that most customers think that most of the attributes are important. Sometimes equally important might mean equally not important. Without the key point that urges the customers to buy, the product might be too neglectable and confuse the customers.

In addition, the result of the logistic regression suggests that none of the variables (including variables like age and income) are significant. Our interpretation might be that, firstly, our questionnaire is designed mainly to survey the perceived importance of the attributes of a VR set. The importance is not preference. A customer might think that a particular attribute is important to the VR set without being intrigued. And even preference doesn't guarantee purchasing behaviors. For example, if the customers are too busy with working or prefer other leisure activities. Even if they value all attributes of the VR sets, they might still not buy. Moreover, it is also possible that our respondents might not know the VR equipment very well. If a customer doesn't have much experience, they might be unable to pinpoint why they would buy it. For instance, some customers might have motion sickness, which hinders them from pursuing a higher level of the gaming experience if the gaming sets are not properly made. Even though some VR sets like Pico 4 can already reduce sickness to a large extent, customers without experience will still assume that the VR sets are not for them. For this kind of situation, we might want to launch promotions aiming at getting customers to familiarize themselves with the user experience (like a VR trial in the shopping mall, youtube video about the product trial experience).

5. Conclusion

To summarize, we recommend releasing the model that has 256G memory, 8 storage(RAM), and 10 free games when buying at the price of 1999 RMB based on the basic statistics.

Table 4

Respondents' willingness to buy the eight VR items

Items\will buy or not	Yes	No
¥1599 8+128G + 5 quality games	21(35%)	39(65%)
¥1599 8+128G + 10 quality games	30(50%)	30(50%)
¥1599 8+256G + 5 quality games	24(40%)	36(60%)
¥1599 8+256G + 10 quality games	42(70%)	18(30%)
¥1999 8+128G + 5 quality games	11(18.33%)	49(81.67%)
¥1999 8+256G + 5 quality games	21(35%)	39(65%)
¥1999 8+128G + 10 quality games	24(40%)	36(60%)
¥1999 8+256G + 10 quality games	40(66.67%)	20(33.33%)

Sales_1999= 1999*40 =79960 > Sales_1599 = 67158

Then considering the factor analysis, we may design a premium that is more suitable for people who value the high level of the gaming experience factor more. For some features of the equipment, they could choose whether the equipment has it with extra pay or without it with the lowest price which would be 1999.

Figure 11

The loading plot of factor 2

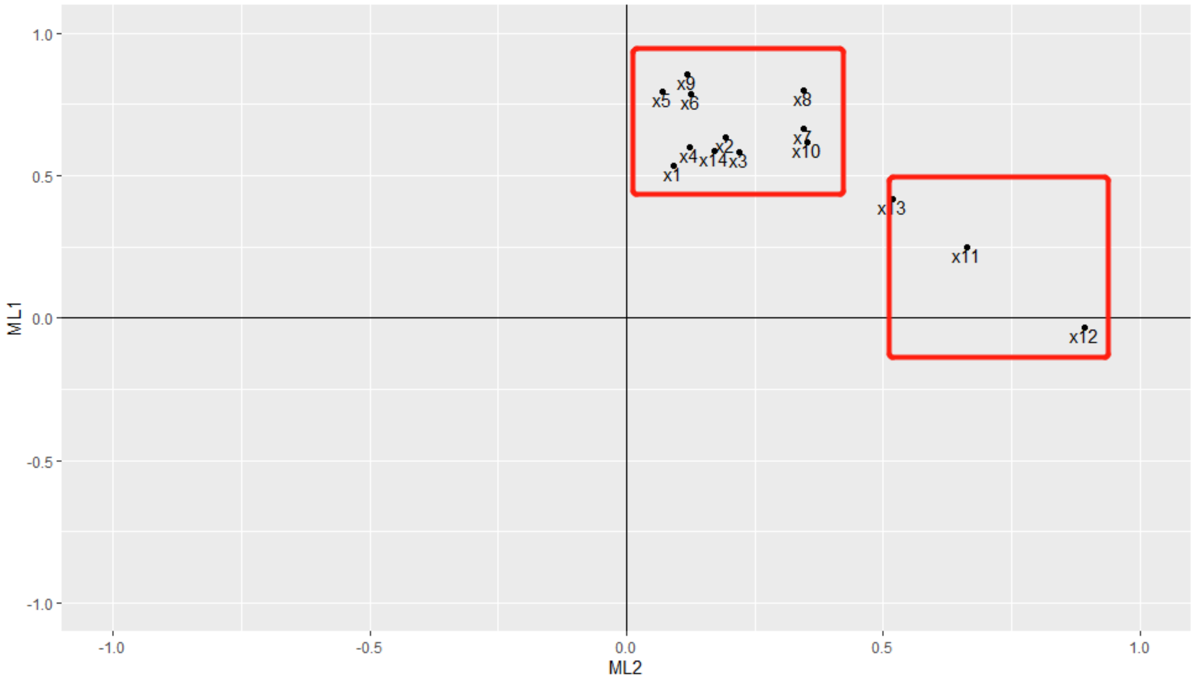
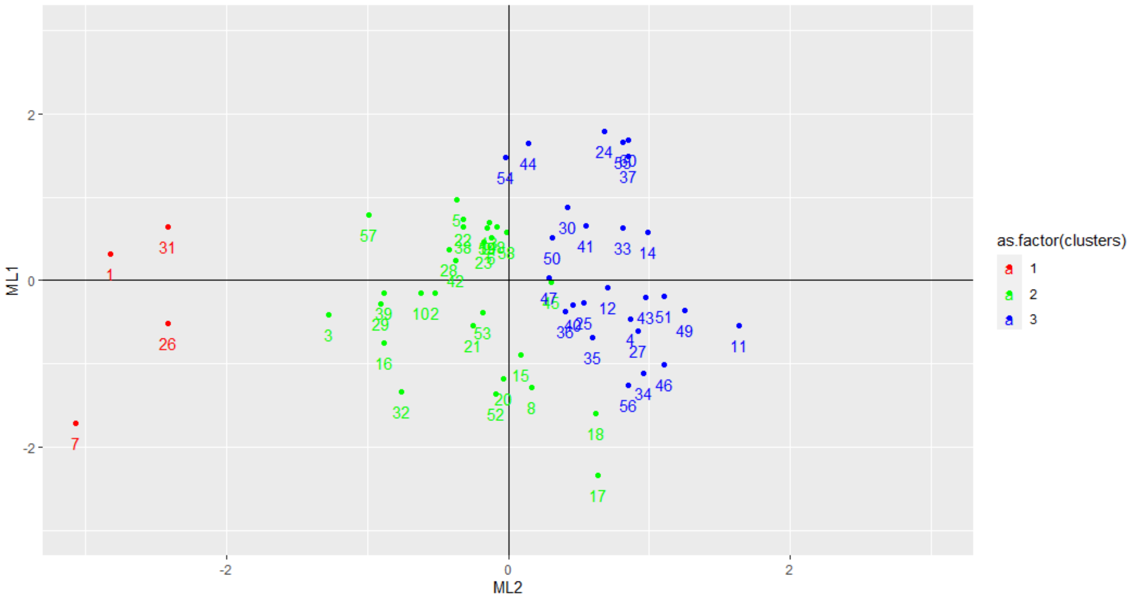


Figure 14

The plot of factor scores colored by the corresponding clusters



Our model should be designed mainly based on the needs of clusters 2 and 3 since they are the majority of our potential buyers.

Although the logistic regression shows results that are not quite satisfactory and likely should not be used as a basis for big marketing decisions, the results obtained from the data summary statistics, factor analysis, and cluster analysis are still usable. VR equipment needs more promotion for the population to acknowledge it fully. After that, we believe the results of the regression will become more clear and more reliable.

Appendix:

Questionnaire on VR headsets

This questionnaire is to investigate factors that the potential/existing customers of VR headsets value in their game experience and their characteristics. Some questions may concern your privacy, but all of your opinions will never be revealed to third parties and only be used as a reference for academic purposes. Please complete this questionnaire as soon as possible! Thanks for your support!

1. Which of the following types of devices do you currently own and use for gaming purposes (multiple choices)?

□ PC □ Cell phone □ Playstation

□ Switch □ Xbox □ VR equipment

□ Others □ I don't have any gaming devices

2. Are you interested in VR gaming?

○ Yes ○ No

3. Please rate the importance of the attributes of a VR headset (1=least important, …, 5=most important).

	Least important	Less important	Neutral	More important	Most important
1. Price	○	○	○	○	○
2. Available games and quality	○	○	○	○	○
3. Type of VR headset (e.g., PC VR)	○	○	○	○	○
4. Weight	○	○	○	○	○
5. Refresh rate	○	○	○	○	○
6. Resolution	○	○	○	○	○
7. Storage	○	○	○	○	○
8. Screen type	○	○	○	○	○
9. Battery life	○	○	○	○	○
10. CPU	○	○	○	○	○
11. Brand	○	○	○	○	○
12. Position tracking	○	○	○	○	○
13. Tactile feedback	○	○	○	○	○
14. Compatibility of myopia lens	○	○	○	○	○

4. What is your monthly income (RMB)?

____

5. Please indicate your gender.

○Male ○Female

6. Do you wear glasses?

○Yes ○No

7. Please tell us your age.

____

8. There are 8 VR headsets weighted only 100g that are described by the storage (8+128G/8+256G), price (HK$1399/HK$1999), and number of quality games (e.g., 3A releases) given away (5/10).

For each of the eight VR sets below, please indicate whether you would be willing to buy it or not.

	Yes	No
¥1599 8+128G + 5 quality games	○	○
¥1599 8+128G + 10 quality games	○	○
¥1599 8+256G + 5 quality games	○	○
¥1599 8+256G + 10 quality games	○	○
¥1999 8+128G + 5 quality games	○	○
¥1999 8+256G + 5 quality games	○	○
¥1999 8+128G + 10 quality games	○	○
¥1999 8+256G + 10 quality games	○	○