



UNIVERSITÀ | DIPARTIMENTO
degli STUDI | di ECONOMIA
di CATANIA | e IMPRESA

MASTER IN DATA SCIENCE FOR MANAGEMENT

SUBJECT : ANALYSIS OF QUESTIONNAIRE

PROF. : G.COMMIS

ETNA WINERY SURVEY



Students:

GIANLUIGI MAZZAGLIA

ANAHITA ESFANDIARYFARD

BIAGIO INCARDONA

THAMIRE DE SOUZA OLIVEIRA

Academic year 2021-2022

Table of Contents

INTRODUCTION	4
1. Questionnaire Structure	4
1.1. Dataset	7
2. Univariate Analysis	7
3. Missing values Handling	26
3.1. Handling Missing Value	27
3.1.1. Numeric Variables.....	27
3.1.2. Categorical Variables	27
4. Multivariate Analysis	27
4.1. Analysis of Correspondence	27
4.1.1. Is there a connection between gender and Etna Wine Preference?	27
4.1.2. Is there a connection between Age (over and under 30) and Etna Wine Preference?	
29	
4.1.3. Is there a connection between education and Etna Wine Preference?	30
4.1.4. Is there a connection between location and Etna Wine Preference?	32
4.1.5. Is there a connection between job and Etna Wine Preference?.....	33
4.2. Latent Class Analysis(LCA).....	35
4.2.1 Exploratory analysis	36
4.2.2 Class analysis	37
4.3. Principal Component Analysis(PCA).....	42
4.4. Classical Test Theory	47
4.4.1. Pre-Analysis: Cronbach's Alpha – Correlation matrix	48
4.5. Factor Analysis	51
4.5.1. Kaiser's MSA and Factor Extraction Results.....	51
4.5.2. Select the Number of Factors	52
4.5.3 Rotation of factor.....	57
4.5.4. Interpretation	58
4.5.5. Standardization and Plot.....	60
4.6. IRT- Item Response Theory	62
4.6.1 Dataset level IRT.....	62
4.6.2 IRT on topics	71
4.6.2.1 Wine preferences, expertise and experience.....	72
4.6.2.2 Wine buying experience	76
4.6.2.3 Factor1	79

4.6.2.4	Factor2	82
4.6.2.3	Etna DOC Wine	86
4.6.3	IRT on factors.....	90
Conclusion		91
Appendix		93

INTRODUCTION

The present study, with reference to the wine, was carried out between December and January 2020/21. The purpose of the survey was to discover winery market trends and customer preferences, focusing on Etna wines. All data have been collected anonymously in line with the privacy policy. Our project was done using the statistical software SAS, in particular after a firstly description of the Questionnaire Structure we did the Univariate Analysis in order to understand the distribution of the answer. Then we handled the missing values and we performed Multivariate Analysis, in particular we focus on Item Response Theory, Factor Analysis and Latent Class Analysis.

The Questionnaire was organized in four sections, the first of which was dedicated to the wine preferences, expertise and consumers'experience (drinking frequency, type of drinks / wine, winery experience). The second part is referred to the buying experience (buying frequency, consumed quantities, consumption time, factors that influence the purchasing moment). The third part of the questionnaire allowed us to collect information about the knowledge and consumption of Etna DOC (where the respondents will give an answer just in case they heard about Etna wine). The fourth section allowed us to obtain information on the socio-economic characteristics of the respondent (age, gender, level of education, occupation).

1. Questionnaire Structure

The Questionnaire is structured in 23 questions, that contains in total 42 elements.

Section 1: Wine preferences, expertise and experience

Question 1 : How much do you like the following drinks?

The question contains 4 items and asks the respondent how much do they like drinks like wine, beer, soft drinks and cocktails. Every item is placed on a semantic scale from 1 (not at all) to 4 (a lot). This question can be useful to understand the drinking preference of the respondent.

Question 2: How much do you like the following kinds of wine?

The question contains 5 items from 1 to 4 adding the possibility to answer “Never tasted”. As before every item is placed on a semantic scale from 1 (not at all) to 4 (a lot). This question allow us to understand which kind of wine the respondent prefers within red, white, rose', sparkling and sweet.

Question 3 : Have you ever tried a wine tasting experience?

This question asks to the respondent if he never tried a wine tasting experience, it's a categorical variable with binary choice between yes or not. Answer this question can help us with the analysis in order to categoryze the typology of respondent, likely a person who is interested in things like this can have an high at least curiosity to improve his knowledge of wine.

Question 4: Have you ever visited a winery?

Also this question it's a categorical variable with binary choice between yes or not. Like the previous question, it can be useful to understand if the respondent is used to going in a winery.

Question 5: Have you ever attended an in-depth wine course?

This question asks to the respondent if he never attended a wine course, it's a categorical variable with binary choice yes or not. This allow us to better understand the respondent in terms of technical expertise.

Question 6: Which is your level of knowledge around wines?

It's a closed item that asks about the level of knowledge around wines of the respondent, in particular it is an ordinal scale ordered from "None" (no knowledge) to "High" (professional knowledge level). This answer can also useful to understand if matches with the previous questions, a respondent who has followed a depth course it's supposed to have an high level of knowledge around wines.

Section 2: Wine buying Experience

Question 7: How often do you buy wine on average in a month?

The question asks the respondent how often he buy wine on average in a month, it's an ordinal variable and only one answer is allowed. The utility of this question can be to better understand if the respondent is a frequently user of wine.

Question 8: How many bottles of wine do you buy on average per month?

The question can be similar to the previous one, It's an ordinal variable and only one answer is allowed. The difference to the previous question is to understand the quantity of bottles that the respondent buy monthly (he can choose in a range from 1-3 bottles, 4-6, to more than 12 bottles) and it can be useful to highlight if he is a consumer during meals or maybe he just want to keep some bottles to share with the others.

Question 9: How often do you buy wine in the following store?

This question contains 3 items and each of them is placed on a semantic scale from 1 (never) to 4 (every time) in order to understand if the respondent used to buy wine in a supermarket, in a winery or online, so we can better understand the typology of consumer.

Question 10: How relevant are the following features when you buy a wine?

This question contains 8 items and it's asked to the respondent to answer a list of features that he consider important during the buying transaction. Every item is placed on a semantic scale from 1 (not at all relevant) to 4 (extremely relevant) and we find items like grape origin, grape variety, budget friendly, brand awarness, vintage, label detailed, attractive packaging, eventual promotions. This question could be extremely important to understand which of the following features can move on our respondent when he buy a wine and so we can realize the kind of person we are referring to.

Question 11: How much do you spend on a bottle of wine on average?

It's a closed item that can be placed in an ordinal scale and ask to the respondent how much he is spend on average for a bottle of wine. The answer can be just one and this information can be useful to understand if an higher quality of wine is important for the respondent and how he can afford to spend on it.

Question 12: During the pandemic, did the frequency with which you buy wine change?

The question contains 1 item and asks to the respondent if during the pandemic the frequency with which you buy wine is changed. Every item is placed on a semantic scale with 5 points and a middle point 0. This question can be useful to understand if the pandemic period has changed the habits of the consumer.

Question 13: For what reason have you bought wine in the last 3 months?

The question asks to the respondent the reason why he bought wine in the last 3 months. The question is a multi select Multiple choice question and allows to choice more than one option (within Home consumption, to buy a gift, for a special event, to try a new wine) and also allow using an opened answer to insert “Other” if the option doesn’t belong to the alternative. The utility of this question allow us to define a reason why the respondent used to buy wine.

Section 3: Etna DOC wine

Question 14: Have you ever heard about Etna DOC wine before?

This question is mandatory and according to the answer that the respondent will give, the next question will be different. In particular the question asks if the respondent has never heard about Etna DOC. It is a categorical variable with binary choice between yes or not. This question can be useful to understand if the respondent knows the Etna DOC or not. This item performs a branching and who answers no is skipped to the final section.

Question 15: Have you ever bought Etna wine?

It is a closed item and it asks the respondent if he never bought Etna wine. It’s a categorical variable with 3 choice (Yes, No, I don’t know) and it is useful to understand if the respondent has never testef Etna wine.

Question 16: How much do you like Etna wines more than other wines?

This question asks to the respondent how much he likes Etna wine more than other wines. Every item is placed on a semantic scale from 1 (not at all) to 4 (a lot). This question allow us to understand if the respondent appreciates the taste of Etna wine.

Question 17: How much do you agree with the following statements about Etna wine?

This question contains 4 closed items and asks to the respondent how much he agrees with certain statements. Every item is placed on a Likert scale with 4 modes from 1 (not at all in agreement) to 4 (completely in agreement). The utility of this question is to understand respondents’ opinion about Etna wine.

Question 18: How likely are you to recommend Etna wine to your family and friends?

This question asks to the respondent how likely they would recommend Etna wine to family and friends. It is on a Likert scale with 4 modes from 1 (not likely at all) to 4 (extremely likely). This question is useful to understand if respondents consider Etna wine as one of the top quality wine and they also provide to recommend it to closers.

Section 4: Socio – demographic data

Question 19: Please enter your gender.

This is a closed item and the question asks to the respondent his gender. It is a categorical variable with 3 choices (man, woman, prefer not to say).

Question 20: Please enter your age:

This question is an open item on an interval scale and the answer is required. It asks the respondent his/her age and give us a discrete variable in a set that is finite as a lower bound, so the respondent must have an age not less than 18, while the upper bound is infinite.

Question 21: Please enter your level of education.

This question allow us to understand the level of education of the respondent. We have here an ordinal scale (from primary and Medium school to University degree). This question can be useful to

understand if could be a correlation between people who have an higher level of education and the knowledge of Etna wine and if in general the choices during shopping time can be different.

Question 22: Where are you from?

This question is a closed item. It asks the respondent the place where he/she comes from. It is a categorical variable with multiple modes. The respondent can be Sicilian, Italian, from an EU country or not EU.

Question 23: What is your occupation?

It is a closed item and asks to the respondent what is her/his occupation. It is a categorical variable with multiple modes. If the occupation is not in the list the respondent can also add his own job applying an open item in the space next to the voice “Other”.

1.1. Dataset

Since we had to form in two languages Italian and English in order to target majority of people(Italians and non Italians), then we initially had two dataset in both languages. So we started by translating italian dataset. The next step is to convert variables from character to numeric. and then we merged the two dataset together to create a unique dataset.

In the next step we transformed categorical variables “Etna buying”, “wine knowledge”, “bottle budget”, “buying experience” and “wine bottles” to numerical values.

For question related to reason of buying wine we decided to transform answers into columns so we added four columns to dataset: party, gift, home, taste and we fill the data with 0 and 1 based on the respondents answer. We did this so we will be able to analyze this part better.

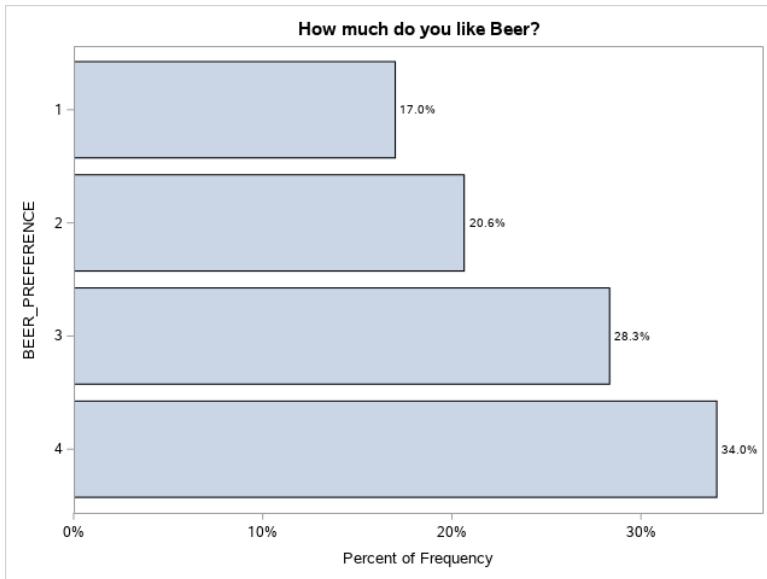
Also we classified age into groups for later analyse.

2. Univariate Analysis

Section 1: Wine preferences, expertise and experience

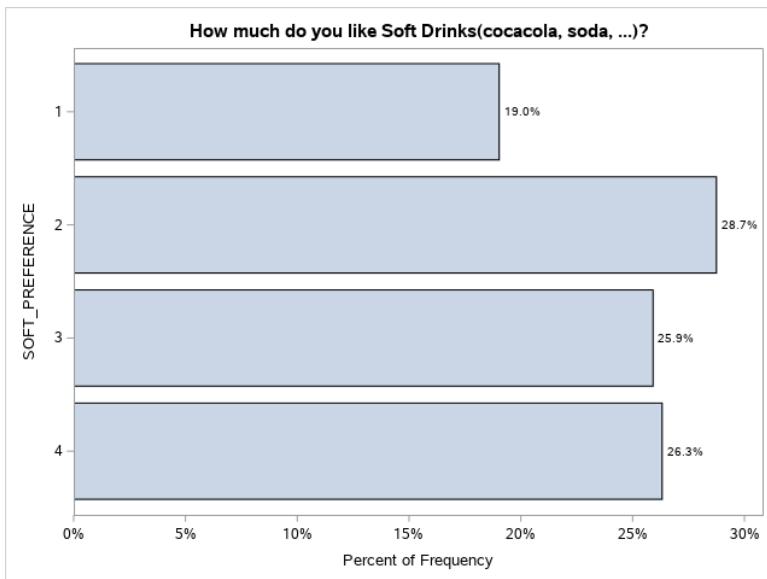
Question 1 : How much do you like the following drinks?

This question is a scale of preference to each type of drink, because of that we divided the graphs of barplot and table of frequency in each kind of drink to have a better view of the results, we will analyse the 247 responses.



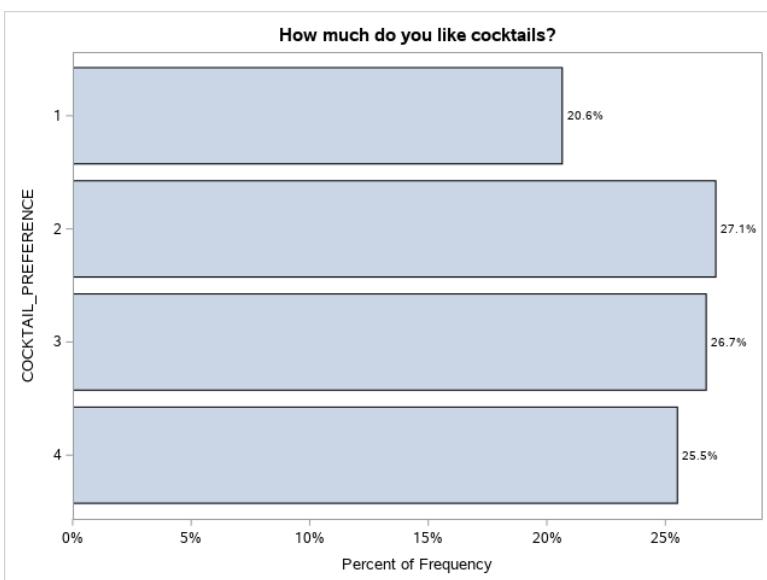
BEER_PREFERENCE				
BEER_PREFERENCE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	42	17.00	42	17.00
2	51	20.65	93	37.65
3	70	28.34	163	65.99
4	84	34.01	247	100.00

The first analysis is over beer preference, where 34% of the people declared that they like beer in a level 4 (a lot) and only 17% declared that the preference of beer is “not at all”. The pattern of the graph is the decrease of responses in the order from 4 to 1 in the level of preference.



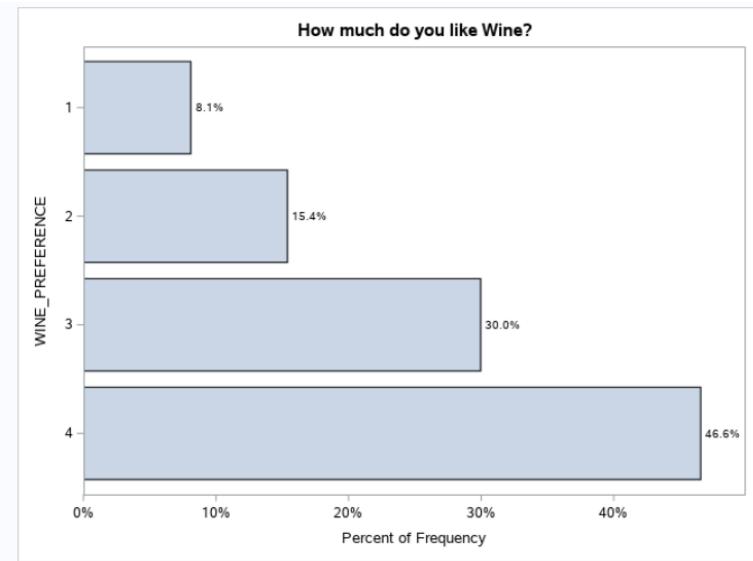
SOFT_PREFERENCE				
SOFT_PREFERENCE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	47	19.03	47	19.03
2	71	28.74	118	47.77
3	64	25.91	182	73.68
4	65	26.32	247	100.00

The second drink to be analysed is soft drinks preferences, like coca-cola, soda,etc. In this case the major part of the people that chose level 2 for their preference for this type of drink, which is considered a low level. Just over 26% declared that they like this kind of beverage a lot.



COCKTAIL_PREFERENCE				
COCKTAIL_PREFERENCE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	51	20.65	51	20.65
2	67	27.13	118	47.77
3	66	26.72	184	74.49
4	63	25.51	247	100.00

The third drink to be analysed is cocktail, where we can check through the graph that almost the same amount of people chose both level 2, 3 or 4 preference in this type of drink, and the minority chose option 1.

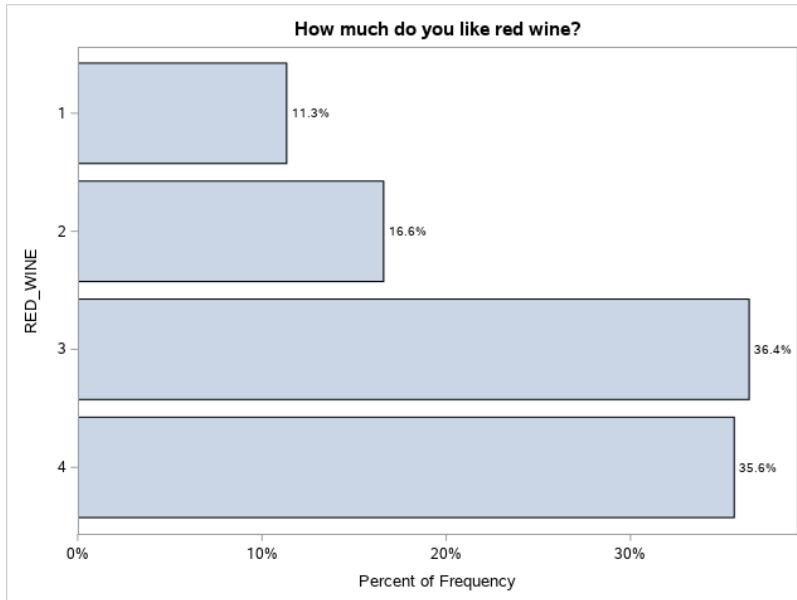


WINE_PREFERENCE				
WINE_PREFERENCE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	20	8.10	20	8.10
2	38	15.38	58	23.48
3	74	29.96	132	53.44
4	115	46.56	247	100.00

The fourth analysis is over wine preference, where 46.56% of the people declared that they like wine in a level 4 (a lot) and only 8.10% declared that the preference of wine is “not at all”. The pattern of the graph is the decrease of responses in the order from 4 to 1 in the level of

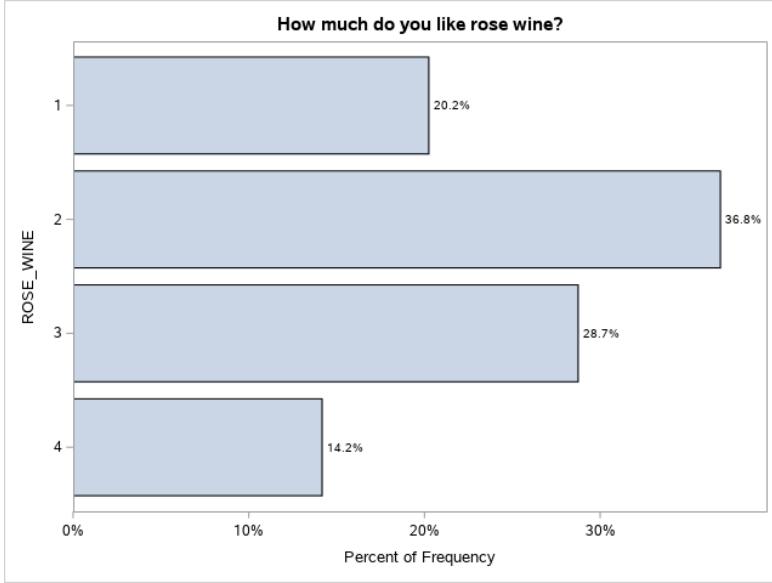
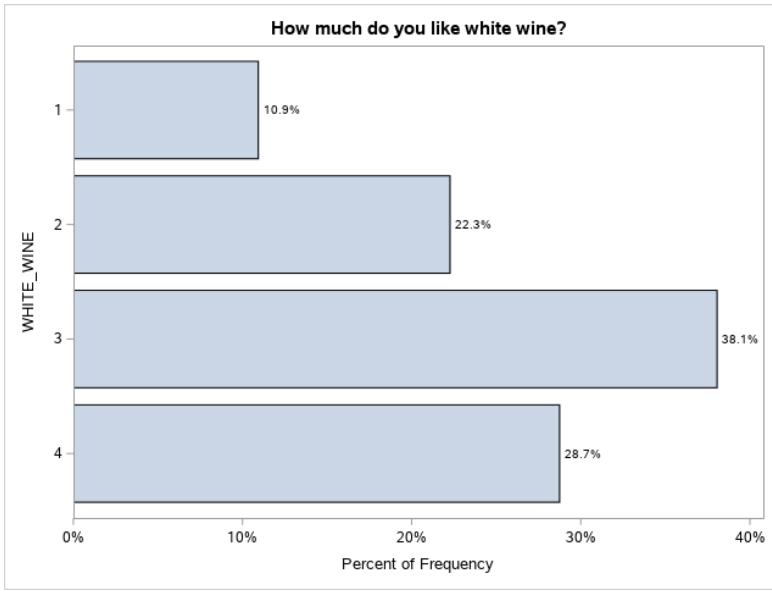
Question 2: How much do you like the following kinds of wine?

This question is another scale of preference with 247 answers. We will show the results by barplot and table of frequency to each option of wine and analyse it separately:



RED_WINE				
RED_WINE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	28	11.34	28	11.34
2	41	16.60	69	27.94
3	90	36.44	159	64.37
4	88	35.63	247	100.00

In red wine observation 27.94% of people are accumulated on levels 1 and 2 of preference.

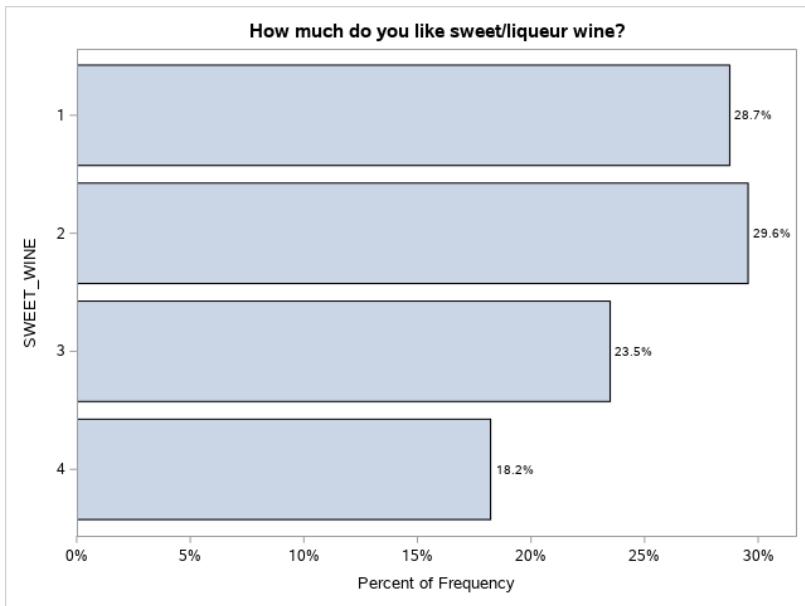


WHITE_WINE				
WHITE_WINE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	27	10.93	27	10.93
2	55	22.27	82	33.20
3	94	38.06	176	71.26
4	71	28.74	247	100.00

The second wine to be analysed is whine wine preference, in this case its possible to see that the majority chose the level 3 of preference in this kind of drink and almost 66.8% of all observations are concentrated on levels 3 and 4.

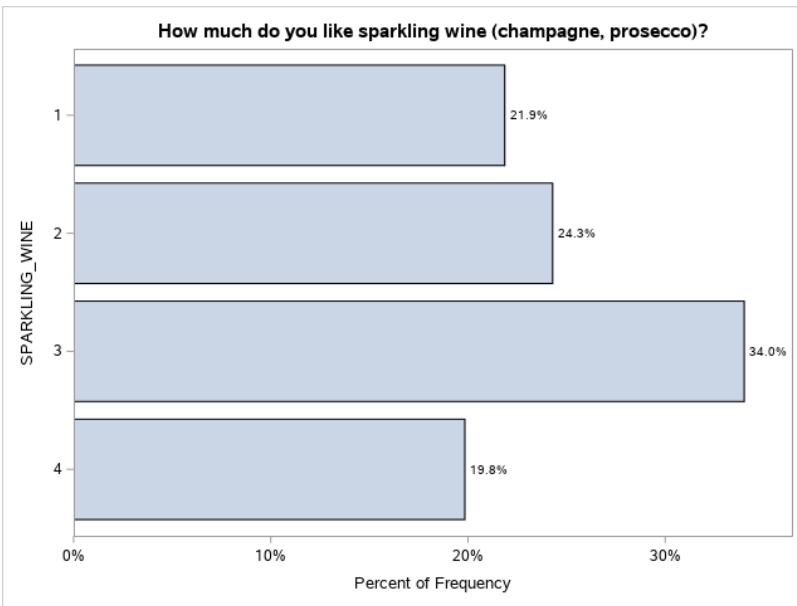
ROSE_WINE				
ROSE_WINE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	50	20.24	50	20.24
2	91	36.84	141	57.09
3	71	28.74	212	85.83
4	35	14.17	247	100.00

The other wine preference to be analysed is rose wine which has its majority at level 2 of preference and only 36.84% at level for of preference.



SWEET_WINE				
SWEET_WINE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	71	28.74	71	28.74
2	73	29.55	144	58.30
3	58	23.48	202	81.78
4	45	18.22	247	100.00

At the analysis of sweet wine, we can see that we have its minority on level 2 and the majority of people, almost 60%, are concentrated on the first 2 levels.

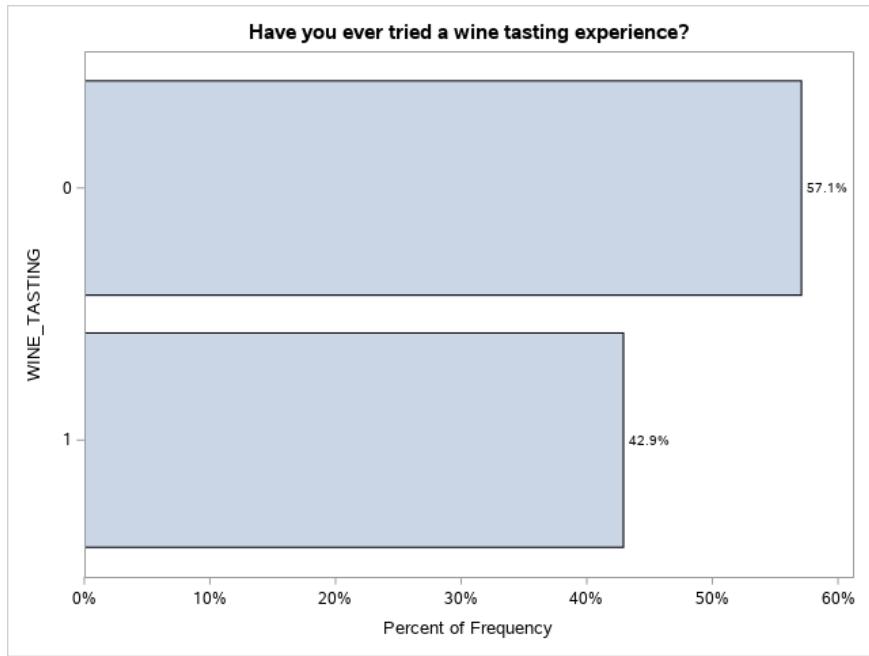


SPARKLING_WINE				
SPARKLING_WINE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	54	21.86	54	21.86
2	60	24.29	114	46.15
3	84	34.01	198	80.16
4	49	19.84	247	100.00

We can see that on sparkling wine preferences, we have less people on level 4 and we have a cumulative percent of 80,16% on levels 1, 2 and 3.

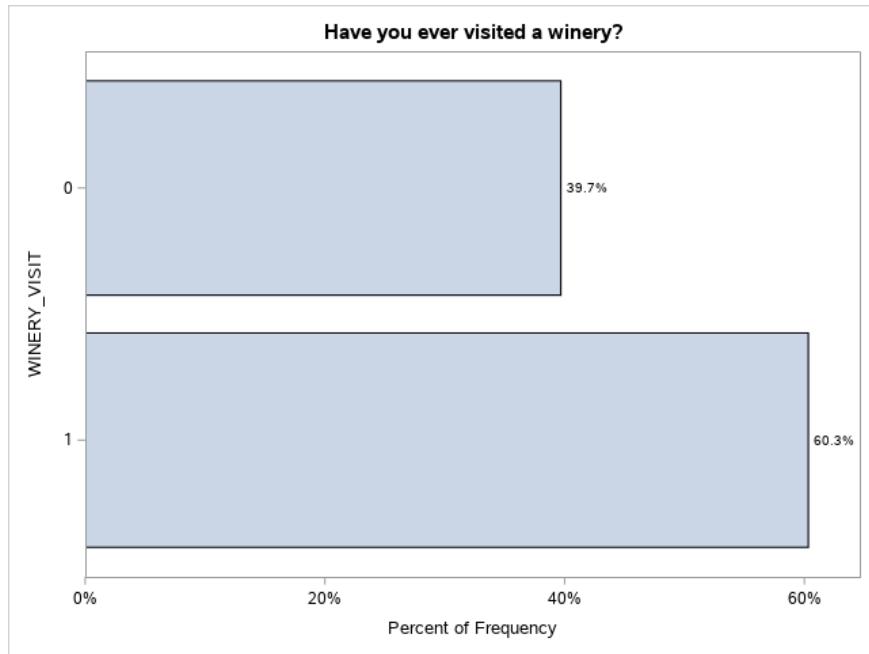
Question 3 : Have you ever tried a wine tasting experience?

This is a simple categorical question with binary choice, where 0 means no and 1 means yes. In this case we can check that more than the half (57,1%) never tried a wine taste experience.



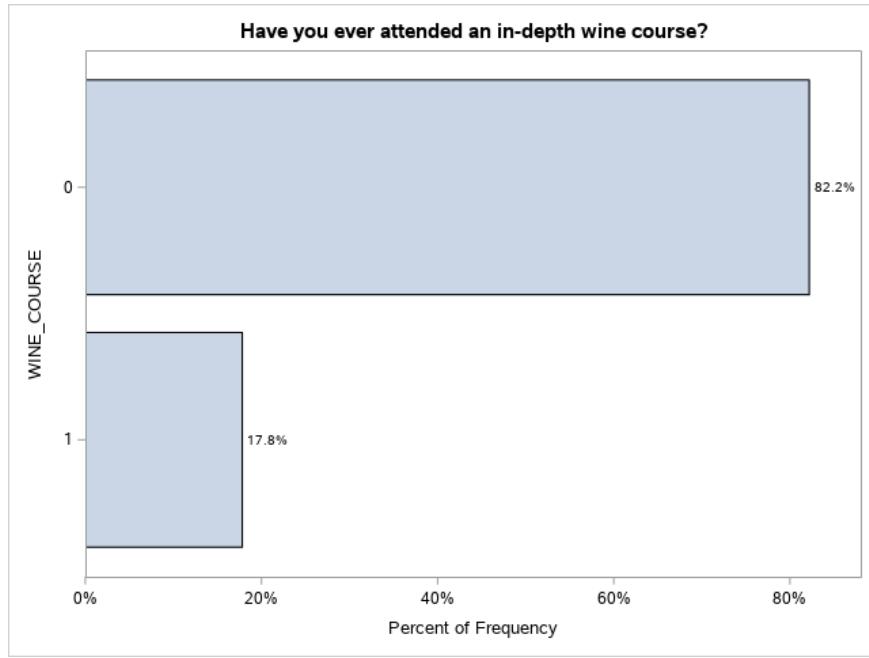
Question 4: Have you ever visited a winery?

Like the previous one is also a binary choice with yes and no options, in this case we can check that 60,3% of people already visited a winery.



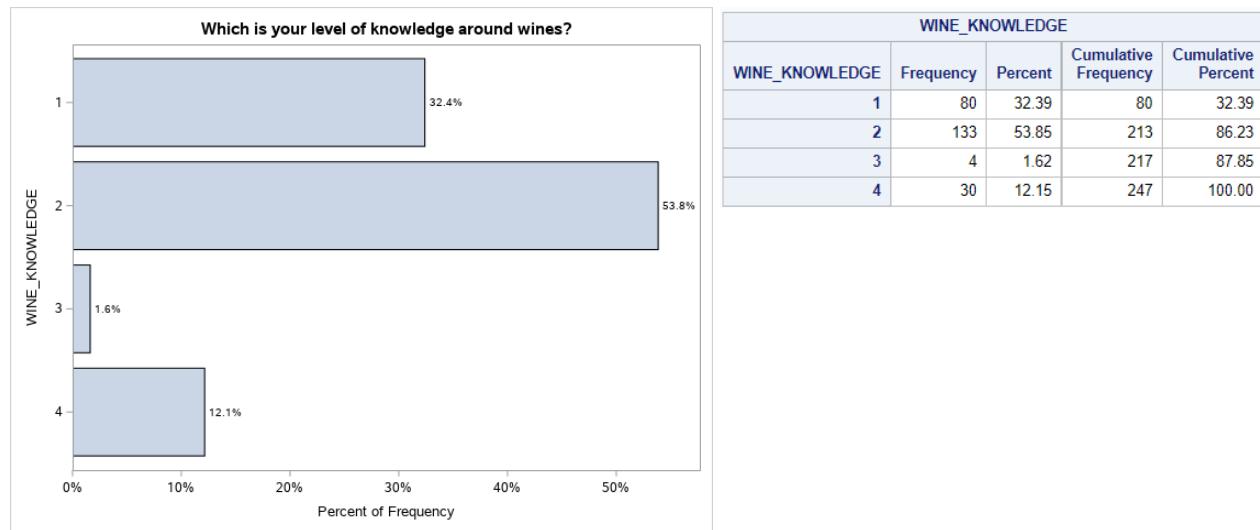
Question 5: Have you ever attended an in-depth wine course?

Another binary question, where we can check that 82,2% never had any depth wine course.



Question 6: Which is your level of knowledge around wines?

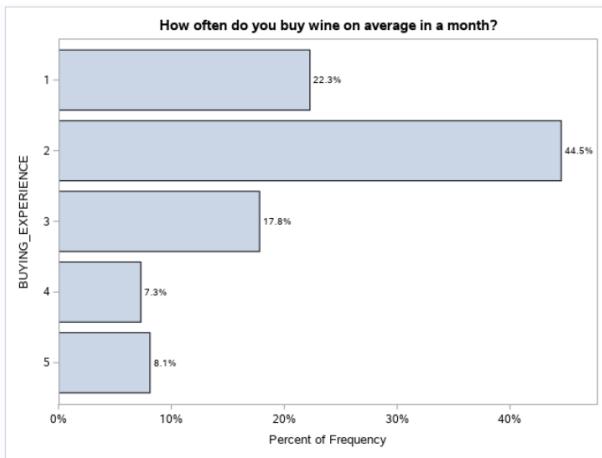
It's a closed question where we can check that 87,85% of people are accumulated on levels 1,2 and 3 of knowledge around wine, and most of it are concentrated on level 2.



Section 2: Wine buying Experience

Question 7: How often do you buy wine on average in a month?

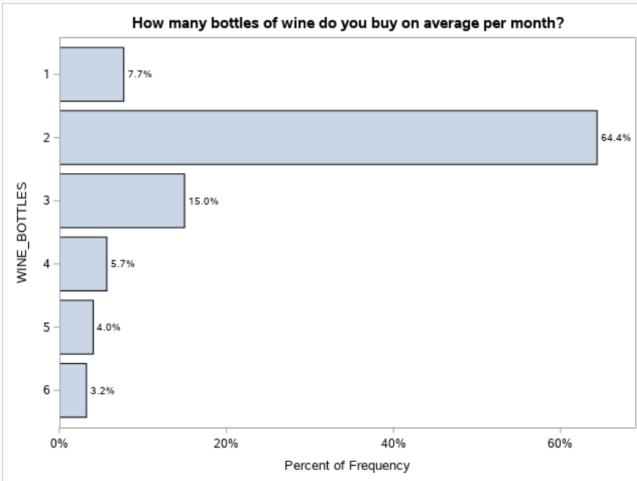
As we can check, most of the people (44.53%) declared to buy wine once a month and 32.3% of the observations were accumulated until 2 times a month.



BUYING_EXPERIENCE				
BUYING_EXPERIENCE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	55	22.27	55	22.27
2	110	44.53	165	66.80
3	44	17.81	209	84.62
4	18	7.29	227	91.90
5	20	8.10	247	100.00

Question 8: How many bottles of wine do you buy on average per month?

As we can check 64.37% of people declared to buy between 1 to 3 bottles of wine per month and 27.94% of the observations are accumulated that they buy more than 3 bottles per month.



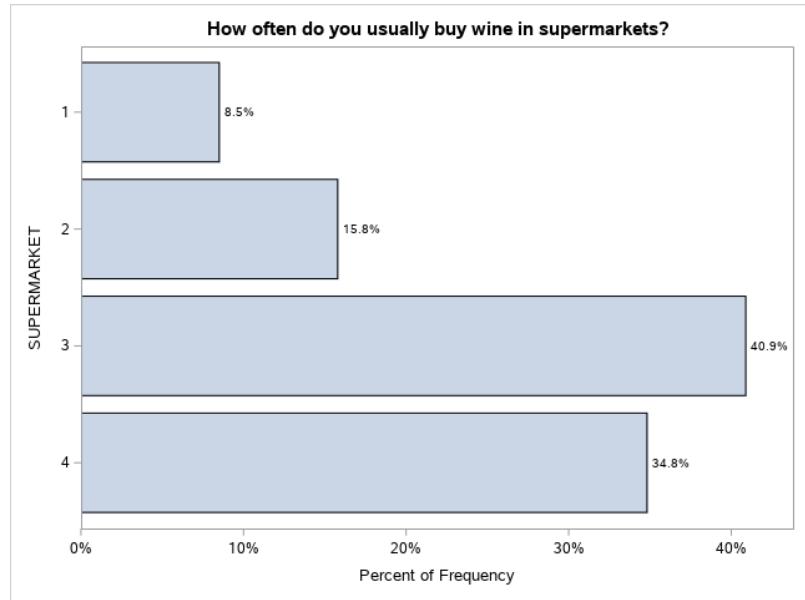
WINE_BOTTLES				
WINE_BOTTLES	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	19	7.69	19	7.69
2	159	64.37	178	72.06
3	37	14.98	215	87.04
4	14	5.67	229	92.71
5	10	4.05	239	96.76
6	8	3.24	247	100.00

Question 9: How often do you buy wine in the following store?

This question is another scale of preference from 1 (never) to 4 (every time) with 247 answers. We will show the results by barplot and table of frequency to each option of store and analyse it separately:

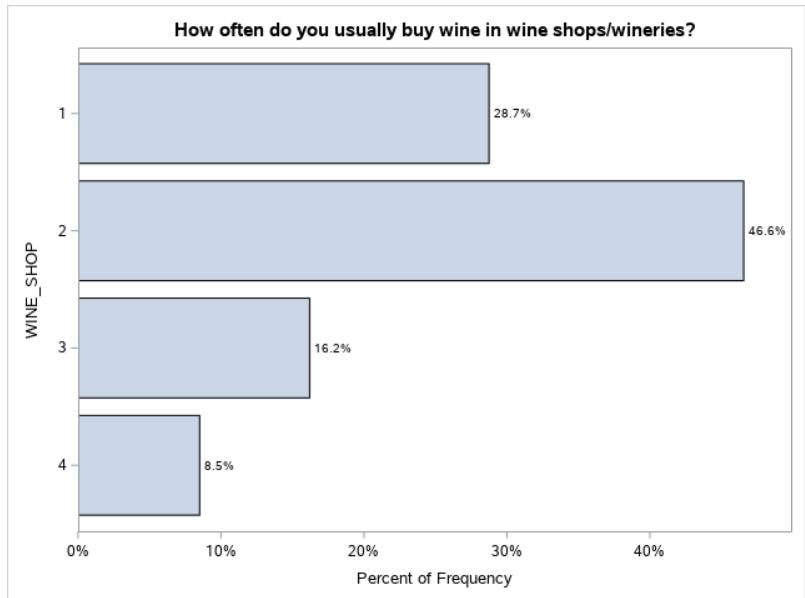
SUPERMARKET				
SUPERMARKET	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	21	8.50	21	8.50
2	39	15.79	60	24.29
3	101	40.89	161	65.18
4	86	34.82	247	100.00

In supermarket observation almost 75% of people buy wine on supermarket o levels 3 and 4, that are the highest levels of frequency of buying.



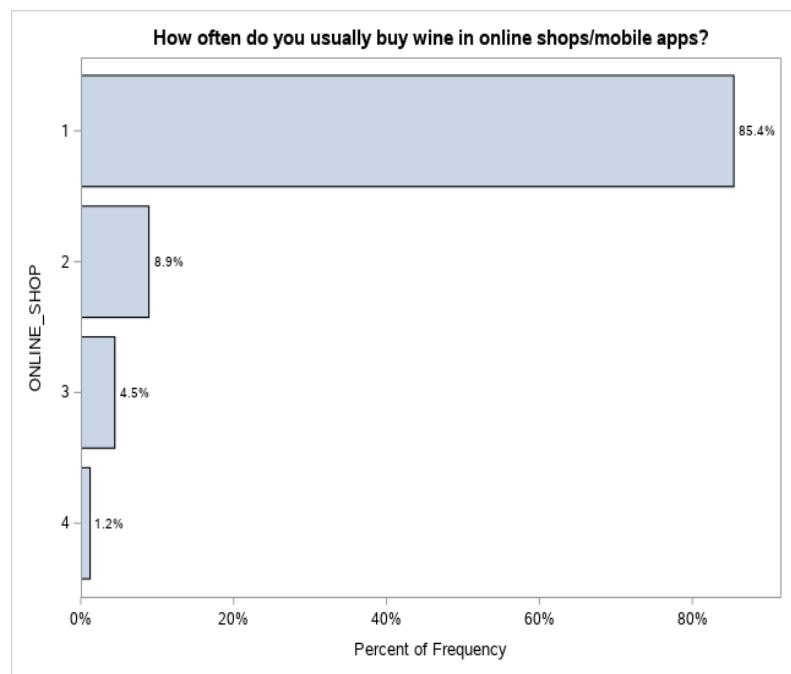
WINE_SHOP				
WINE_SHOP	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	71	28.74	71	28.74
2	115	46.56	186	75.30
3	40	16.19	226	91.50
4	21	8.50	247	100.00

Differently from the previous one, people who declared to buy wine in wine shops are accumulated almost 76% at first 2 levels, that are the lowest ones, and 46,6% buy on wine shops on level 2 of frequency.



ONLINE_SHOP				
ONLINE_SHOP	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	211	85.43	211	85.43
2	22	8.91	233	94.33
3	11	4.45	244	98.79
4	3	1.21	247	100.00

Most of the people, exactly 85,4%, never buy wine at online shop and only 3 declared to buy by that method every time.

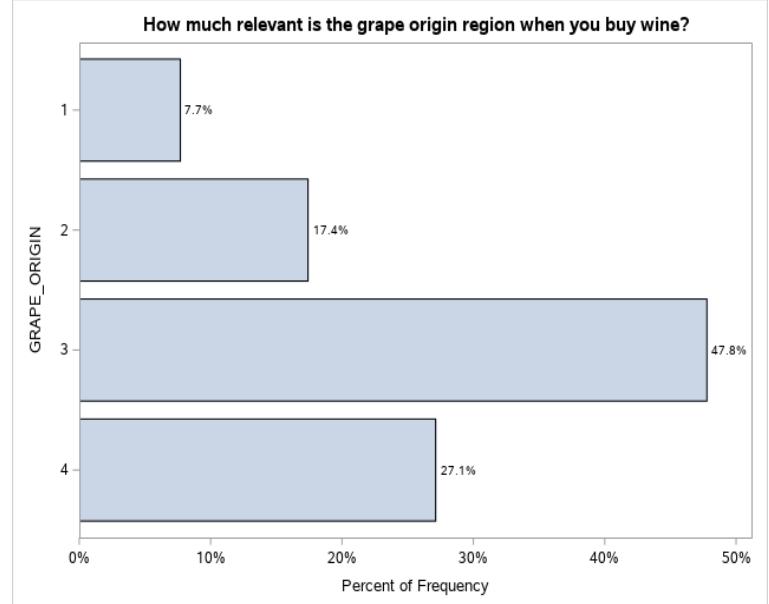


Question 10: How relevant are the following features when you buy a wine?

This question is another scale of preference from 1 (not all relevant) to 4 (extremely relevant) with 247 answers. We will show the results by bar plot and table of frequency to each option of features and analyse it separately:

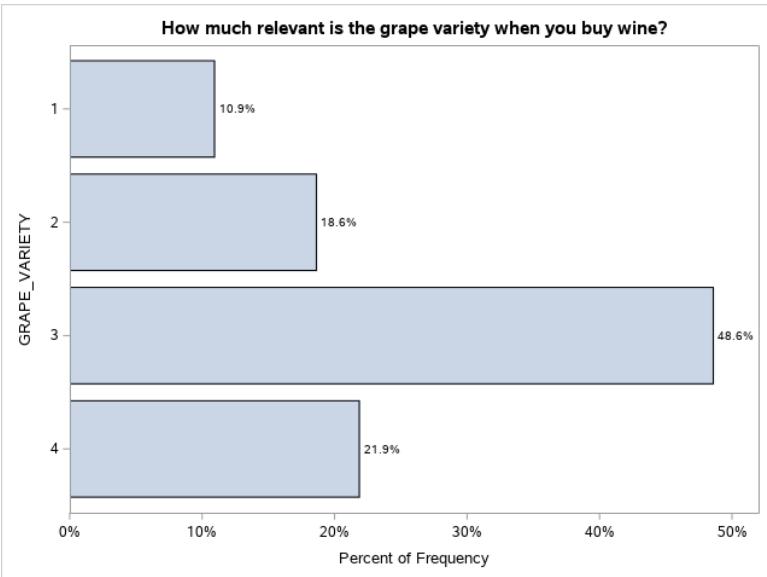
GRAPE_ORIGIN				
GRAPE_ORIGIN	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	19	7.69	19	7.69
2	43	17.41	62	25.10
3	118	47.77	180	72.87
4	67	27.13	247	100.00

For most of the people the grape origin has the relevance of level 3 and the cumulative with level 4 is for 74,9% what means is a really relevant feature.



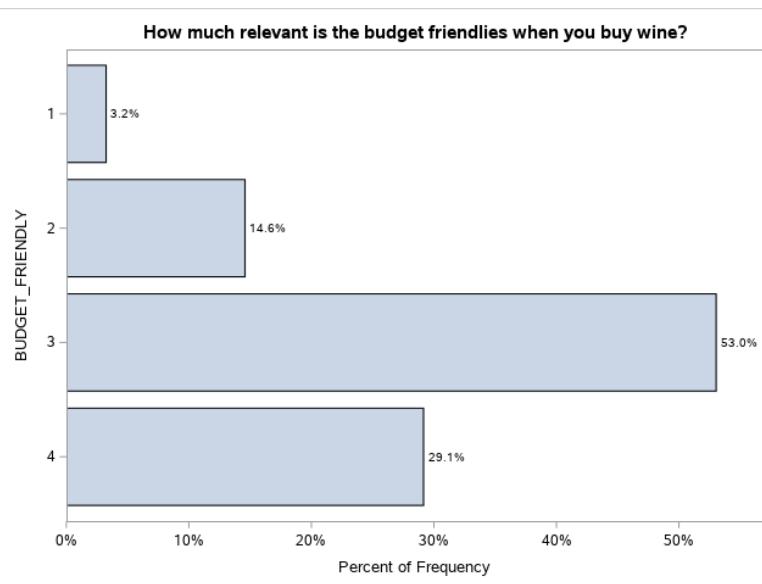
GRAPE_VARIETY				
GRAPE_VARIETY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	27	10.93	27	10.93
2	46	18.62	73	29.55
3	120	48.58	193	78.14
4	54	21.86	247	100.00

For most of the people the grape variety has the relevance of level 3 and the cumulative with level 4 is for 70,5% what means is a really relevant feature.



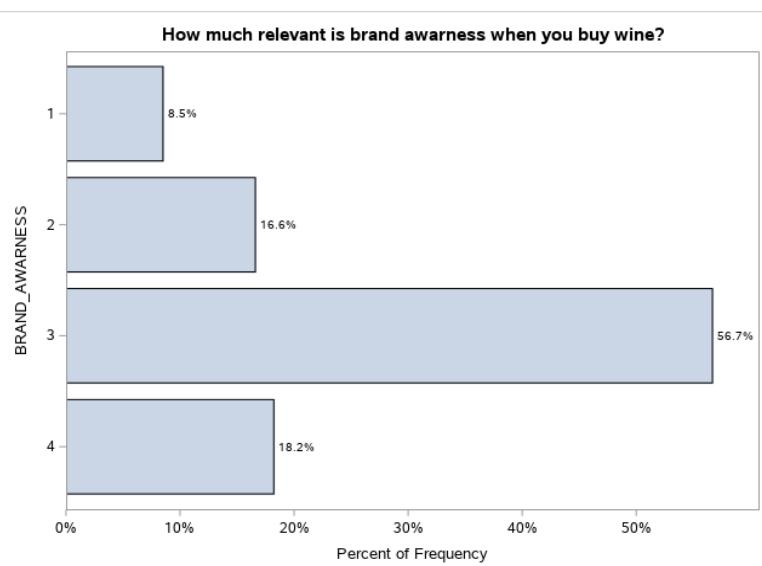
BUDGET_FRIENDLY				
BUDGET_FRIENDLY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	8	3.24	8	3.24
2	36	14.57	44	17.81
3	131	53.04	175	70.85
4	72	29.15	247	100.00

The feature of being budget friendly has the relevance of 3 for the most of people and the cumulative with level 4 is of 82,1% what means is also a relevant feature.



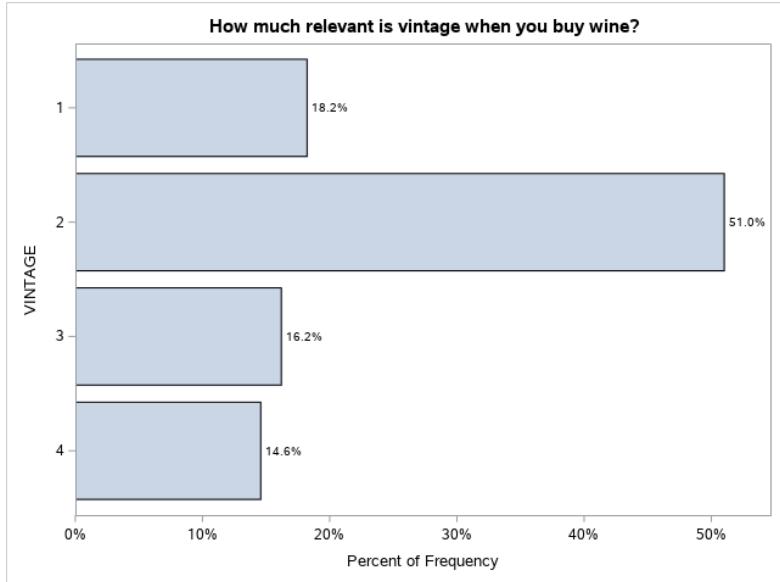
BRAND_AWARNESS				
BRAND_AWARNESS	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	21	8.50	21	8.50
2	41	16.60	62	25.10
3	140	56.68	202	81.78
4	45	18.22	247	100.00

Like the other ones most of the people consider the brand awarness of level 3, and the relevance of level 2 and 4 are pretty much the same.



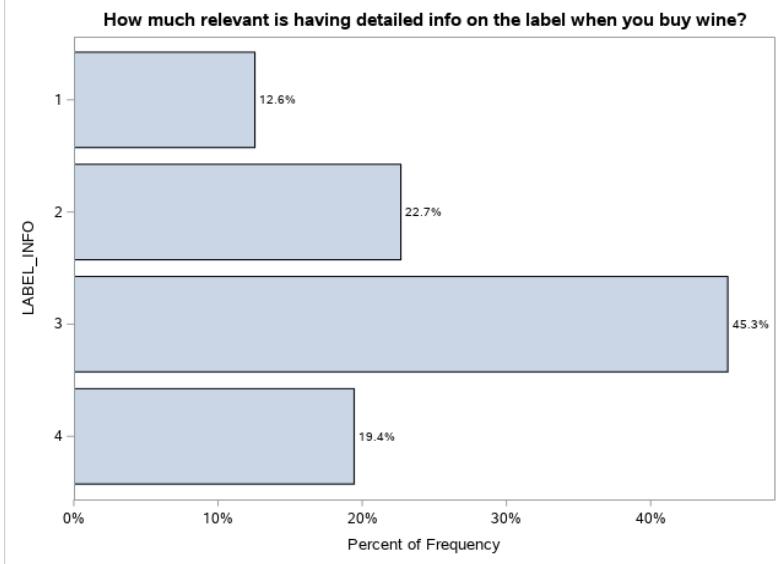
VINTAGE				
VINTAGE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	45	18.22	45	18.22
2	126	51.01	171	69.23
3	40	16.19	211	85.43
4	36	14.57	247	100.00

Unlike the other features, vintage has the relevance of 2 for more than half of the people what shows that is one of the less relevant features so far.



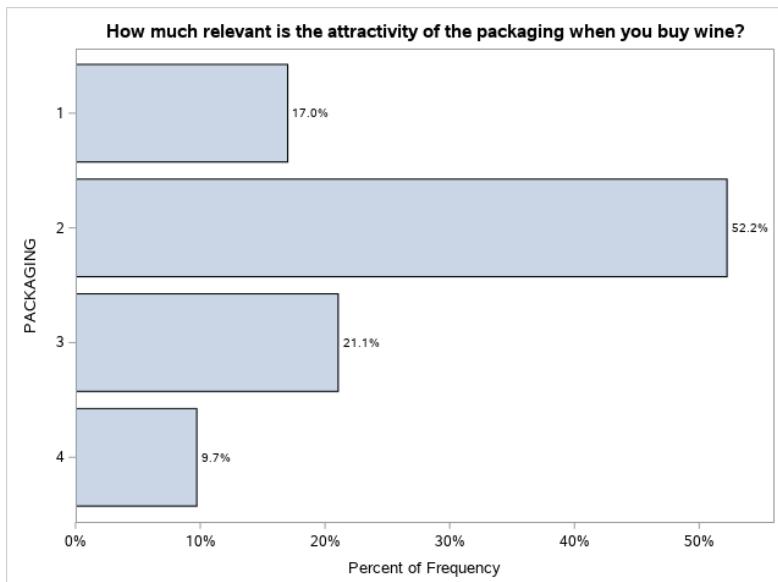
LABEL_INFO				
LABEL_INFO	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	31	12.55	31	12.55
2	56	22.67	87	35.22
3	112	45.34	199	80.57
4	48	19.43	247	100.00

The feature of having detailed label has the relevance of 3 for the most of people and the cumulative with level 2 is of 68%.



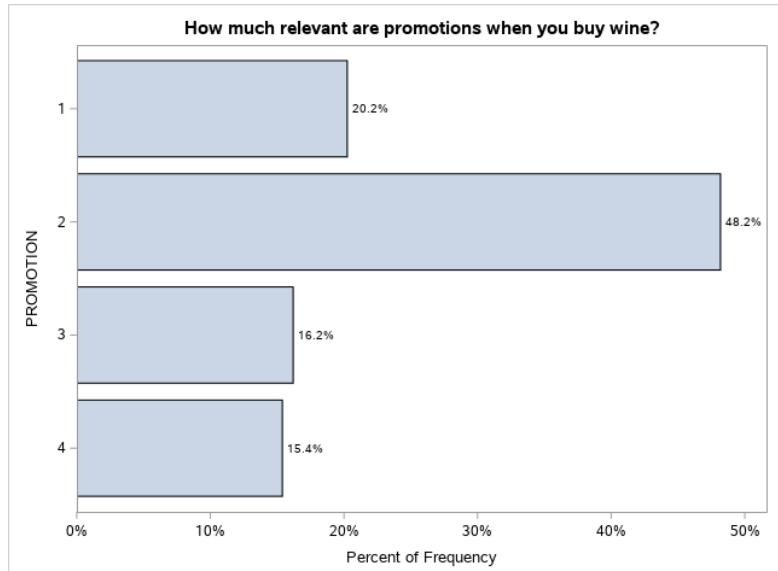
PACKAGING				
PACKAGING	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	42	17.00	42	17.00
2	129	52.23	171	69.23
3	52	21.05	223	90.28
4	24	9.72	247	100.00

More than half of people, exactly 52,2%, declared that packaging has a relevance of 2 for them.



PROMOTION				
PROMOTION	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	50	20.24	50	20.24
2	119	48.18	169	68.42
3	40	16.19	209	84.62
4	38	15.38	247	100.00

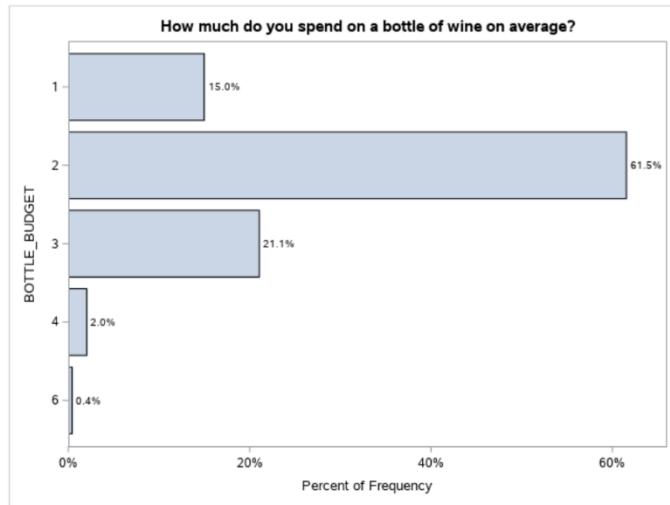
Almost half of people, exactly 48,2%, declared that packaging has a relevance of 2 for them.



Question 11: How much do you spend on a bottle of wine on average?

As we can check, more than half of the people, exactly 61.54% spend around 5 to 15 euro, which if we look at the occupation of our respondents we see majority are students which it makes sense. And only 0.40% of our sample buy wines costs more than 50 euro.

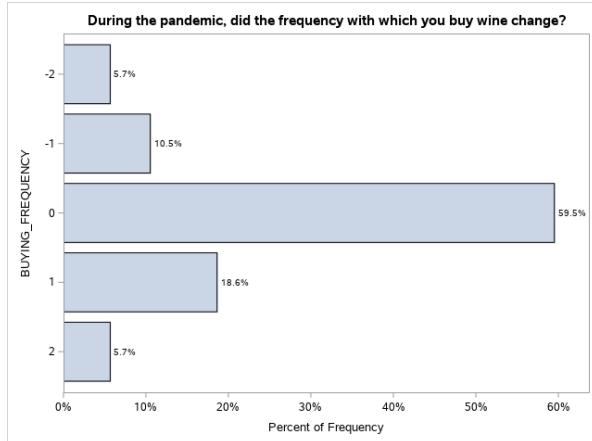
BOTTLE_BUDGET				
BOTTLE_BUDGET	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	37	14.98	37	14.98
2	152	61.54	189	76.52
3	52	21.05	241	97.57
4	5	2.02	246	99.60
6	1	0.40	247	100.00



Question 12: During the pandemic, did the frequency with which you buy wine change?

As we can check, more than half of the people, exactly 59,5%, during the pandemic, didn't change any habits related to buying wine. Instead, 24.3% increased the number of wine purchases and 16.2% decreased the amount of wines purchased.

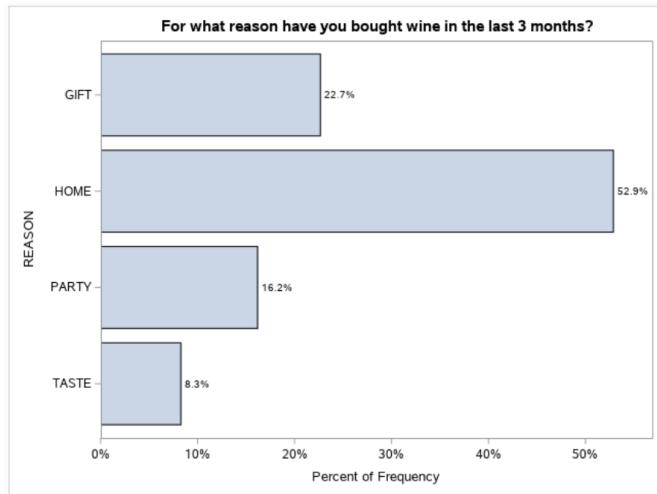
BUYING_FREQUENCY				
BUYING_FREQUENCY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
-2	14	5.67	14	5.67
-1	26	10.53	40	16.19
0	147	59.51	187	75.71
1	46	18.62	233	94.33
2	14	5.67	247	100.00



Question 13: For what reason have you bought wine in the last 3 months?

As we can check, almost half of the people, exactly 52.88%, bouth the wine for home consuming. Instead, only 8.23% bouth wine just for test it.

REASON				
REASON	Frequency	Percent	Cumulative Frequency	Cumulative Percent
GIFT	63	22.66	63	22.66
HOME	147	52.88	210	75.54
PARTY	45	16.19	255	91.73
TASTE	23	8.27	278	100.00

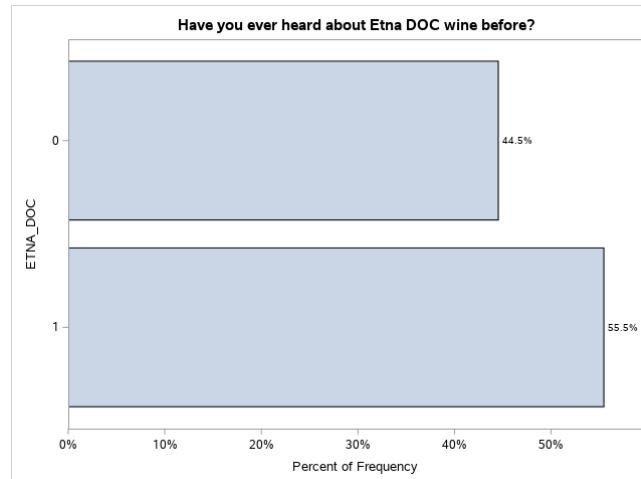


Section 3: Etna DOC wine

Question 14: Have you ever heard about Etna DOC wine before?

ETNA_DOC				
ETNA_DOC	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	110	44.53	110	44.53
1	137	55.47	247	100.00

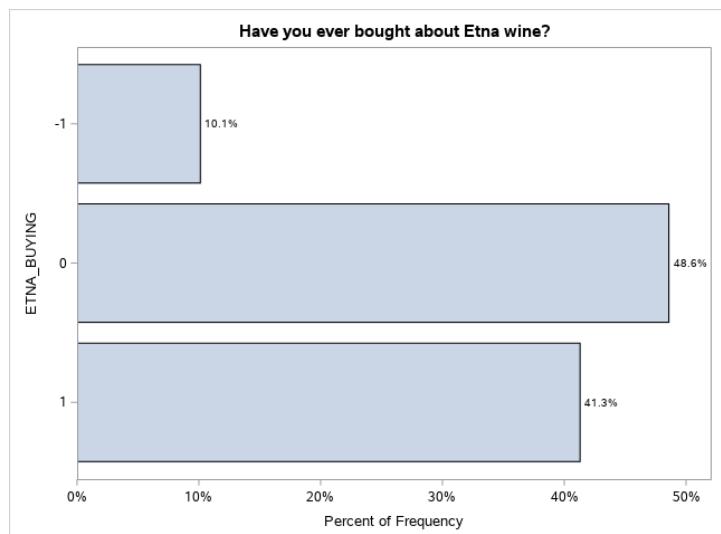
As we can observe 55,47% of people had heard about Etna Wine before.



Question 15: Have you ever bought Etna wine?

As we can observe more people declared that they never bought etna wine (48,6%) and we still have a amount of people (10,1%) that don't know if they bought Etna Wine or not.

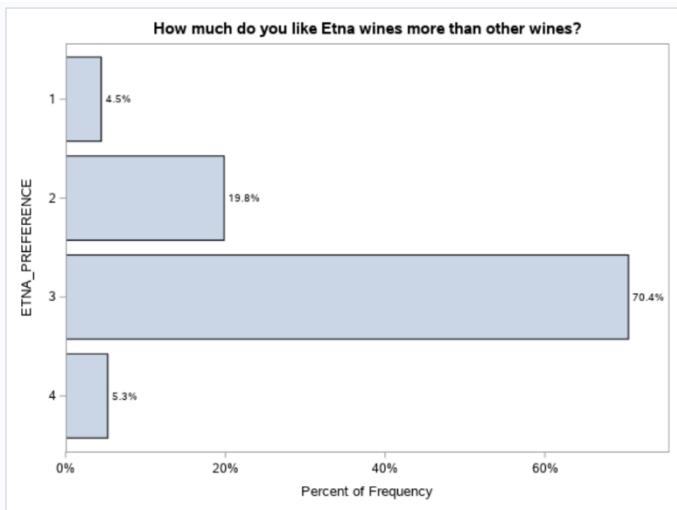
ETNA_BUYING				
ETNA_BUYING	Frequency	Percent	Cumulative Frequency	Cumulative Percent
-1	25	10.12	25	10.12
0	120	48.58	145	58.70
1	102	41.30	247	100.00



Question 16: How much do you like Etna wines more than other wines?

ETNA_PREFERENCE				
ETNA_PREFERENCE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	11	4.45	11	4.45
2	49	19.84	60	24.29
3	174	70.45	234	94.74
4	13	5.26	247	100.00

As we can observe more people (70%) declared that they prefer etna wine 3 out of 4 scale and almost 20% of people prefer it 2 out of 4 we can say majority of people has middle preference about this wine.

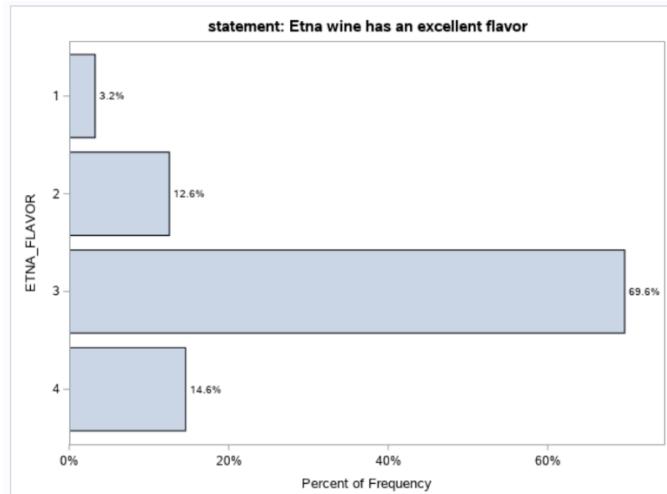


Question 17: How much do you agree with the following statements about Etna wine?

This question is another scale of preference from 1 (not all relevant) to 4 (extremely relevant) with 247 answers. We will show the results by bar plot and table of frequency to each option of features and analyse it separately:

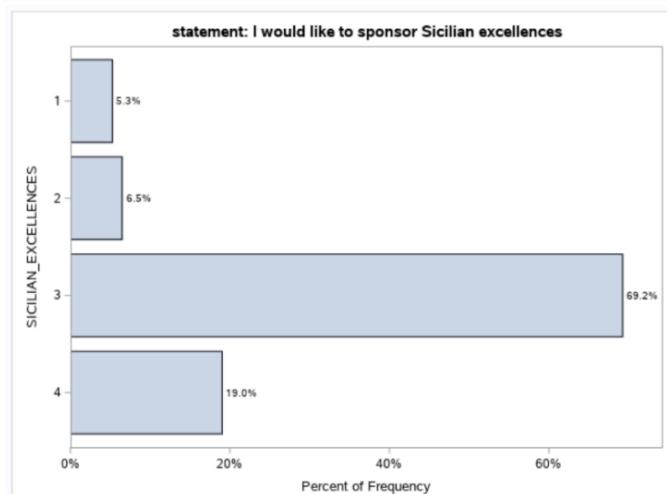
ETNA_FLAVOR				
ETNA_FLAVOR	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	8	3.24	8	3.24
2	31	12.55	39	15.79
3	172	69.64	211	85.43
4	36	14.57	247	100.00

As we can see more than half of people, exactly 69.64%, declared that etna flavor has a relevance of 3 for them.



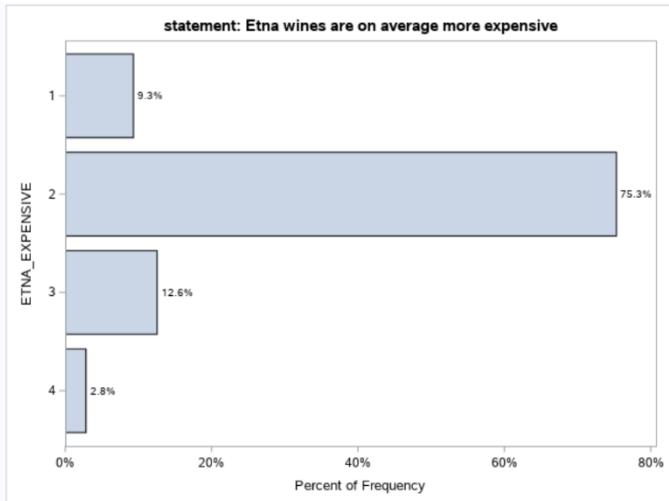
SICILIAN_EXCELLENCES				
SICILIAN_EXCELLENCES	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	13	5.26	13	5.26
2	16	6.48	29	11.74
3	171	69.23	200	80.97
4	47	19.03	247	100.00

As we can see more than half of people, exactly 69.23%, declared that sicilian excelences has a relevance of 3 for them.



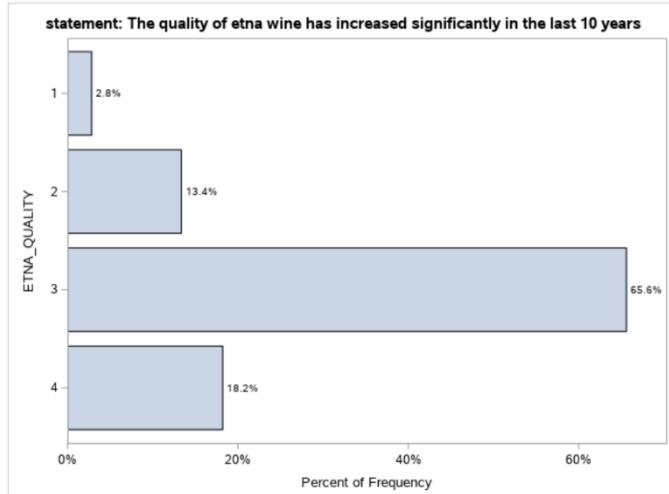
ETNA_EXPENSIVE				
ETNA_EXPENSIVE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	23	9.31	23	9.31
2	186	75.30	209	84.62
3	31	12.55	240	97.17
4	7	2.83	247	100.00

As we can see more than half of people, exactly 75.30%, declared that etna expensive has a relevance of 2 for them. So most people don't consider etna wine as an expensive wine.



ETNA_QUALITY				
ETNA_QUALITY	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	7	2.83	7	2.83
2	33	13.36	40	16.19
3	162	65.59	202	81.78
4	45	18.22	247	100.00

As we can see more than half of people, exactly 65.59%, declared that etna quality has a relevance of 3 for them. So most people consider etna wine as a quality wine.

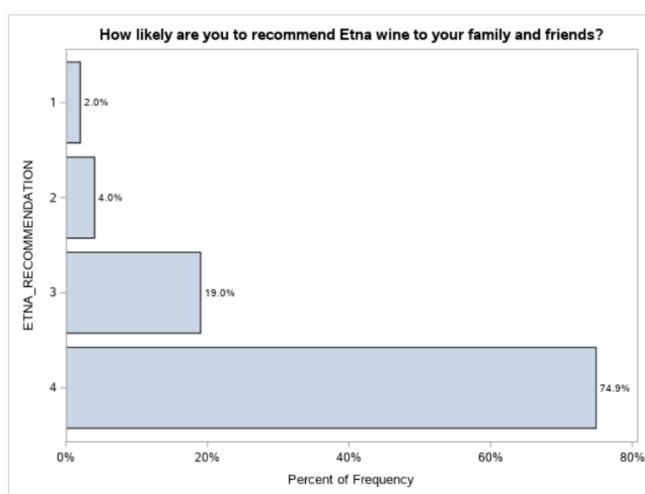


Question 18: How likely are you to recommend Etna wine to your family and friends?

This question is scale of people recommendation for etna wine from 1 (not all relevant) to 4 (extremely relevant) with 247 answers. We will show the results by bar plot and table of frequency:

ETNA_RECOMMENDATION				
ETNA_RECOMMENDATION	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	5	2.02	5	2.02
2	10	4.05	15	6.07
3	47	19.03	62	25.10
4	185	74.90	247	100.00

As we can see more than half of people, exactly 74.90%, declared that strongly recommend etna wine to the other people.

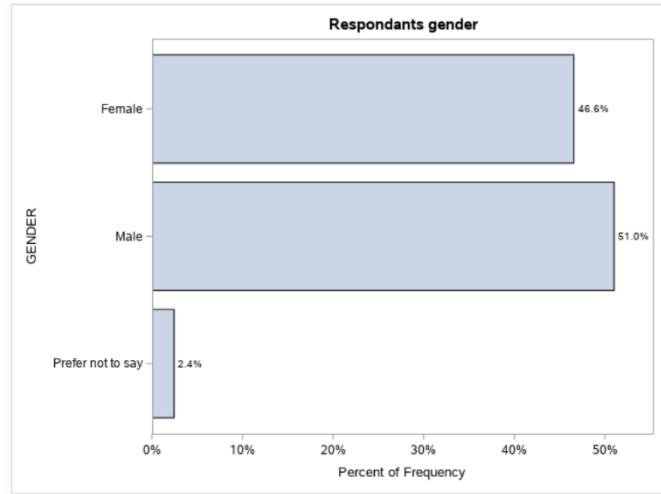


Question 19: Please enter you gender

This question to understand the gender of our respondents, we will show the results by bar plot and table of frequency:

GENDER				
GENDER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Female	115	46.56	115	46.56
Male	126	51.01	241	97.57
Prefer not to say	6	2.43	247	100.00

As we can see more than half of respondents, exactly 51.01%, are male.

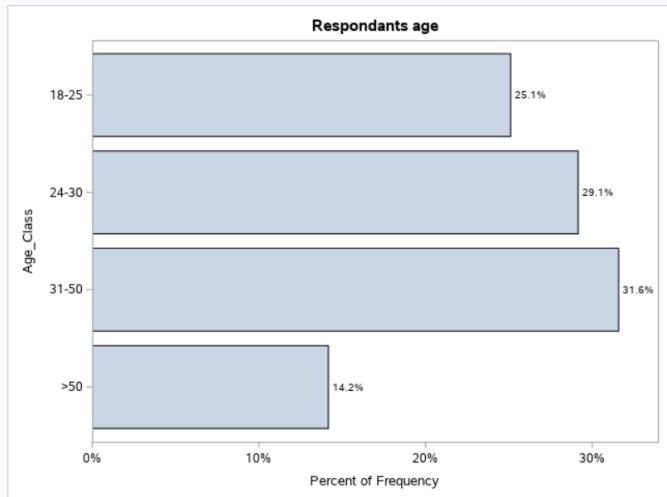


Question 20: Please enter your age

This question to understand the age of our respondents, we will show the results by bar plot and table of frequency:

Age_Class	Frequency	Percent	Cumulative Frequency	Cumulative Percent
18-25	62	25.10	62	25.10
24-30	72	29.15	134	54.25
31-50	78	31.58	212	85.83
>50	35	14.17	247	100.00

As we can see almost 32% of our respondents are between 31-50 and after that 29.15% are between 24-30.

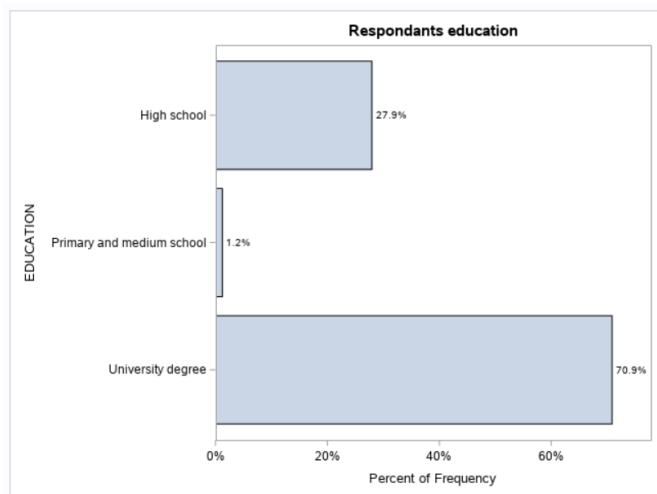


Question 21: Please enter your level of education

This question to understand the level of education of our respondents, we will show the results by bar plot and table of frequency

EDUCATION				
EDUCATION	Frequency	Percent	Cumulative Frequency	Cumulative Percent
High school	69	27.94	69	27.94
Primary and medium school	3	1.21	72	29.15
University degree	175	70.85	247	100.00

As we can see almost 70.85% of our respondents has university degree.

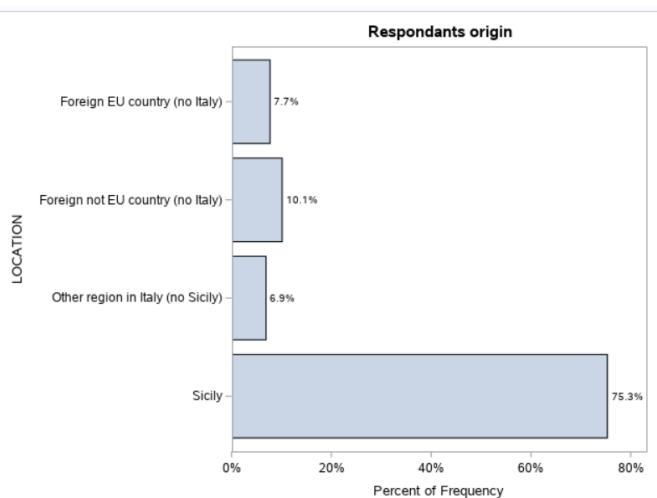


Question 22: Where are you from?

This question to understand the location of our respondents, we will show the results by bar plot and table of frequency:

LOCATION				
LOCATION	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Foreign EU country (no Italy)	19	7.69	19	7.69
Foreign not EU country (no Italy)	25	10.12	44	17.81
Other region in Italy (no Sicily)	17	6.88	61	24.70
Sicily	186	75.30	247	100.00

As we can see almost 75.30% of our respondents are from Sicily

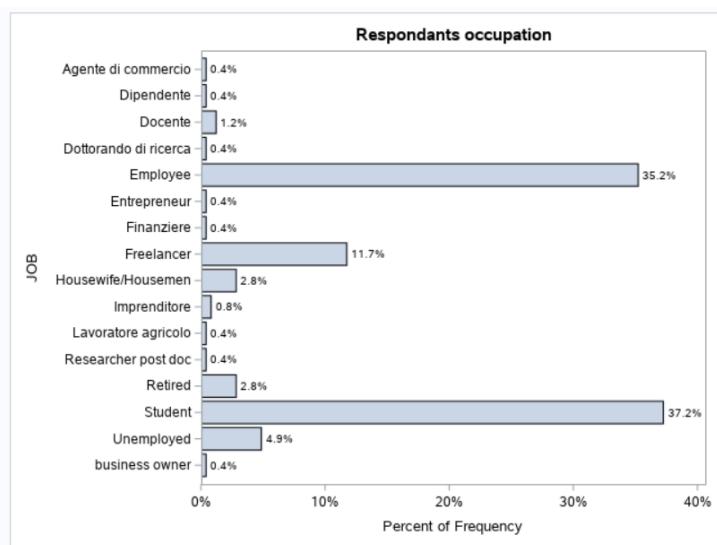


Question 23: What is your occupation?

This question is for understand the type of job our respondent. With 6 options and “other” option for writing the occupation if is not listed we have 247 answers. We will show the answers by barplot and table of frequency and analyse it

:

JOB				
JOB	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Agente di commercio	1	0.40	1	0.40
Dipendente	1	0.40	2	0.81
Docente	3	1.21	5	2.02
Dottorando di ricerca	1	0.40	6	2.43
Employee	87	35.22	93	37.65
Entrepreneur	1	0.40	94	38.06
Finanziere	1	0.40	95	38.46
Freelancer	29	11.74	124	50.20
Housewife/Housemen	7	2.83	131	53.04
Imprenditore	2	0.81	133	53.85
Lavoratore agricolo	1	0.40	134	54.25
Researcher post doc	1	0.40	135	54.66
Retired	7	2.83	142	57.49
Student	92	37.25	234	94.74
Unemployed	12	4.86	246	99.60
business owner	1	0.40	247	100.00



Most of the people who participated in our research are students with variaty of 37.25 percent and then employee with 35.22 percent.

3. Missing values Handling

As we observe from the univariate analysis the dataset contains missing values due to two reasons. One is due to branching of “buying experience” question and hearing about “etna doc” wine question.and Second due to choose “Never tested” in wine prefrence.The following table shows the percentage of missing values:

Percentage of missing values for numeric variables

Obs	_NAME_	_LABEL_	nvalues	nmiss
1	WHITE_WINE	WHITE_WINE	199	0.5
2	ROSE_WINE	ROSE_WINE	195	2.5
3	RED_WINE	RED_WINE	199	0.5
4	SPARKLING_WINE	SPARKLING_WINE	198	1.0
5	SWEET_WINE	SWEET_WINE	193	3.5
6	SUPERMARKET	SUPERMARKET	160	20.0
7	WINE_SHOP	WINE_SHOP	160	20.0
8	ONLINE_SHOP	ONLINE_SHOP	160	20.0
9	GRAPE_ORIGIN	GRAPE_ORIGIN	160	20.0
10	GRAPE_VARIETY	GRAPE_VARIETY	160	20.0
11	BUDGET_FRIENDLY	BUDGET_FRIENDLY	160	20.0
12	BRAND_AWARENESS	BRAND_AWARENESS	160	20.0
13	VINTAGE	VINTAGE	160	20.0
14	LABEL_INFO	LABEL_INFO	160	20.0
15	PACKAGING	PACKAGING	160	20.0
16	PROMOTION	PROMOTION	160	20.0
17	BUYING_FREQUENCY	BUYING_FREQUENCY	160	20.0
18	ETNA_PREFERENCE	ETNA_PREFERENCE	129	35.5
19	ETNA_FLAVOR	ETNA_FLAVOR	129	35.5

Obs	_NAME_	_LABEL_	nvalues	nmiss
20	SICILIAN_EXCELLENCES	SICILIAN_EXCELLENCES	129	35.5
21	ETNA_EXPENSIVE	ETNA_EXPENSIVE	129	35.5
22	ETNA_QUALITY	ETNA_QUALITY	129	35.5
23	ETNA_RECOMMENDATION	ETNA_RECOMMENDATION	129	35.5

3.1. Handling Missing Value

3.1.1. Numeric Variables

For handling missing value of numeric variables, Median is replaced with missing items.

3.1.2. Categorical Variables

For handling missing value of categorical variable Mode is replaced with missing items.

4. Multivariate Analysis

4.1. Analysis of Correspondence

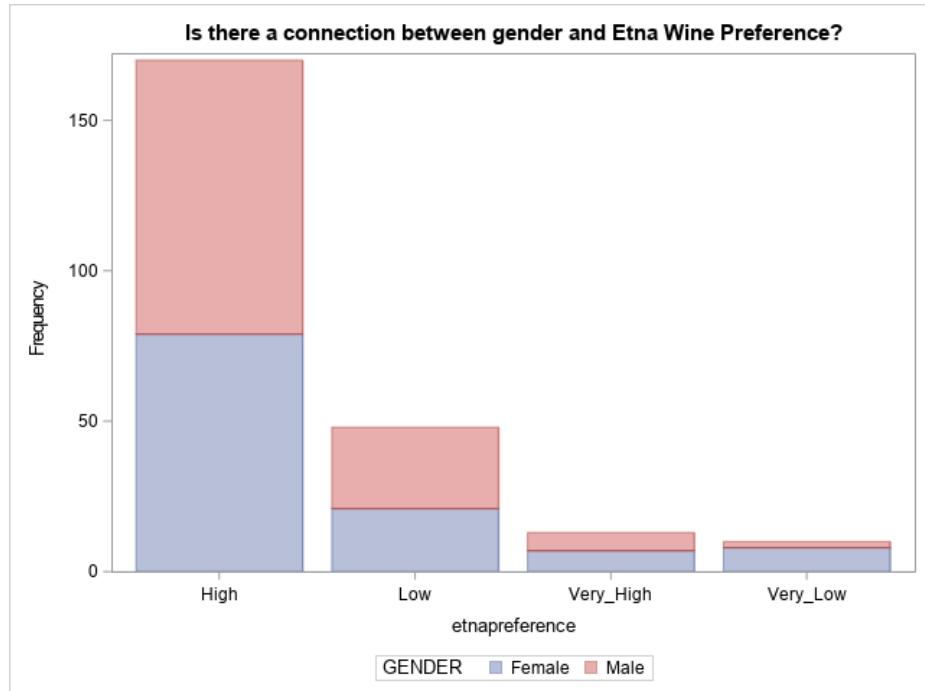
Correspondence Analysis a multivariate statistical technique that is conceptually similar to principal component analysis, but applies to categorical rather than continuous data. In a similar manner to principal component analysis, it provides a means of displaying or summarising a set of data in two-dimensional graphical form. The main goal is to reveal the relative relationships between and within two groups of variables, based on data given in a contingency table.

In order to start with the analysis, data has been processed again since it is necessary to have only categorical variables. The variable we will try to connect with other is "Etna Wine Preference" so we transform it in four levels of preference: low, very low, high and very high.

4.1.1. Is there a connection between gender and Etna Wine Preference?

To analyze this connection in a statistically valid way, people who chose "prefer not to say" as their gender were taken out of the sample.

Thereby, in this first correspondence analysis, we output a graph grouped by the 4 levels of preference in Etna Wine and divided by color in each gender. Visually, it is possible to notice that there is not a big difference between the gender preference for Etna wine, except when the preference is very low, where we have clearly a high number of women than men.



Frequency	Table of etnapreference by GENDER			Statistic	DF	Value	Prob
	etnapreference	GENDER(GENDER)					
		Female	Male	Total			
High	79	91	170	Chi-Square	3	4.7819	0.1885
Low	21	27	48	Likelihood Ratio Chi-Square	3	5.0294	0.1697
Very_High	7	6	13	Mantel-Haenszel Chi-Square	1	2.4167	0.1200
Very_Low	8	2	10	Phi Coefficient		0.1409	
Total	115	126	241	Contingency Coefficient		0.1395	
				Cramer's V		0.1409	
				Sample Size = 241			

We have a table of frequency as a output to demonstrate these numbers more clearly, where we can check that high, low and very high levels has, proportionally, good divided opinions.

In the other table, its possible to see that the p-value is 0.1885, so we can conclude that there is no connection between the gender and the Etna Wine preference.

Contingency Table				Chi-Square Statistic Expected Values		
Percents	Female	Male	Sum	Percents	Female	Male
High	32.780	37.759	70.539	High	33.6599	36.8795
Low	8.714	11.203	19.917	Low	9.5040	10.4130
Very_High	2.905	2.490	5.394	Very_High	2.5740	2.8202
Very_Low	3.320	0.830	4.149	Very_Low	1.9800	2.1694
Sum	47.718	52.282	100.000			

The difference between values in the contingency table and the Chi-Square Statistic Expected Values explain the level of p-value in chi-square value. In this case, since gender was not connected to Etna Wine Preference, the difference between the expected values and the actual values are smaller than it would be if there was a valid connection.

We have just small differences, like we can see on the tables above, it was expected a higher number of men who has very low level of etna preference, but the real number was less. The opposite happened when it comes to women, it was expected a lower number of this gender choosing a very low level, but the number was way higher.

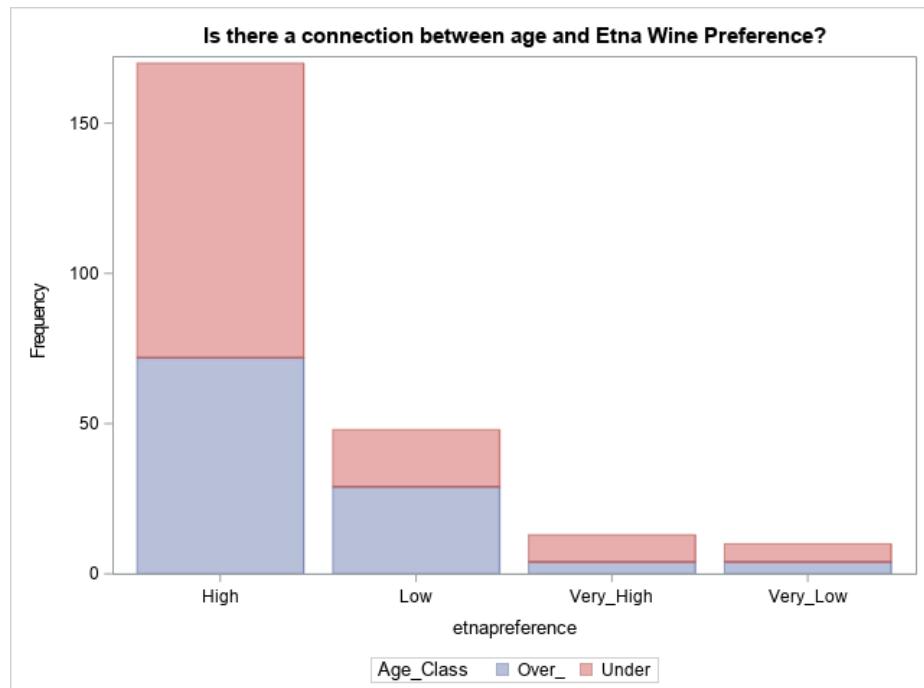
Contributions to the Total Chi-Square Statistic			
Percents	Female	Male	Sum
High	1.159	1.058	2.217
Low	3.312	3.023	6.335
Very_High	2.140	1.953	4.093
Very_Low	45.672	41.684	87.356
Sum	52.282	47.718	100.000

The result of total chi-square did not confirm the connection, and we can see that the category “very-low” had more power in this result of chi-square statistic.

4.1.2. Is there a connection between Age (over and under 30) and Etna Wine Preference?

To analyze this connection in a statistically valid way, we divided all the ages in only two groups: over 30 and under 30.

After that, in this second correspondence analysis we output the graph below grouped by the 4 levels of preference in Etna Wine and divided by color in each group of age. Visually, we can observe a higher number of under 30 in high level of preference, and the opposite happens in low level of preference.



							Statistic			DF	Value	Prob	
							Chi-Square			3	6.2447	0.1003	
							Likelihood Ratio Chi-Square			3	6.2687	0.0992	
							Mantel-Haenszel Chi-Square			1	0.1067	0.7439	
							Phi Coefficient				0.1610		
							Contingency Coefficient				0.1589		
							Cramer's V				0.1610		
Frequency		Table of Age_Class by etnapreference											
		etnapreference											
		Age_Class	High	Low	Very_High	Very_Low	Total						
		Over_	72	29	4	4	109						
		Under	98	19	9	6	132						
		Total	170	48	13	10	241						

Sample Size = 241

We have a table of frequency as a output to demonstrate these numbers more clearly, so we can see that on high level of preference we have 98 responses of under 30 against 72 of over 30.

In the other table, its possible to see that the p-value is 0.1003, so we can conclude that there is no connection between the age and the Etna Wine preference.

Contingency Table					
Percents	High	Low	Very_High	Very_Low	Sum
Over_	29.876	12.033	1.660	1.660	45.228
Under	40.664	7.884	3.734	2.490	54.772
Sum	70.539	19.917	5.394	4.149	100.000

Chi-Square Statistic Expected Values				
Percents	High	Low	Very_High	Very_Low
Over_	31.9037	9.0081	2.4397	1.8767
Under	38.6357	10.9089	2.9545	2.2727

Since the p-value explained that there is no valid connection between age and Etna Wine preference, the differences between the contingency table and expected values are not so significant. We can only observe small differences: it was expected to have less over 30 having low preference in Etna Wine than what really happened and it was expected to have less under 30 to have high preference in Etna Wine, but the numbers were not that different.

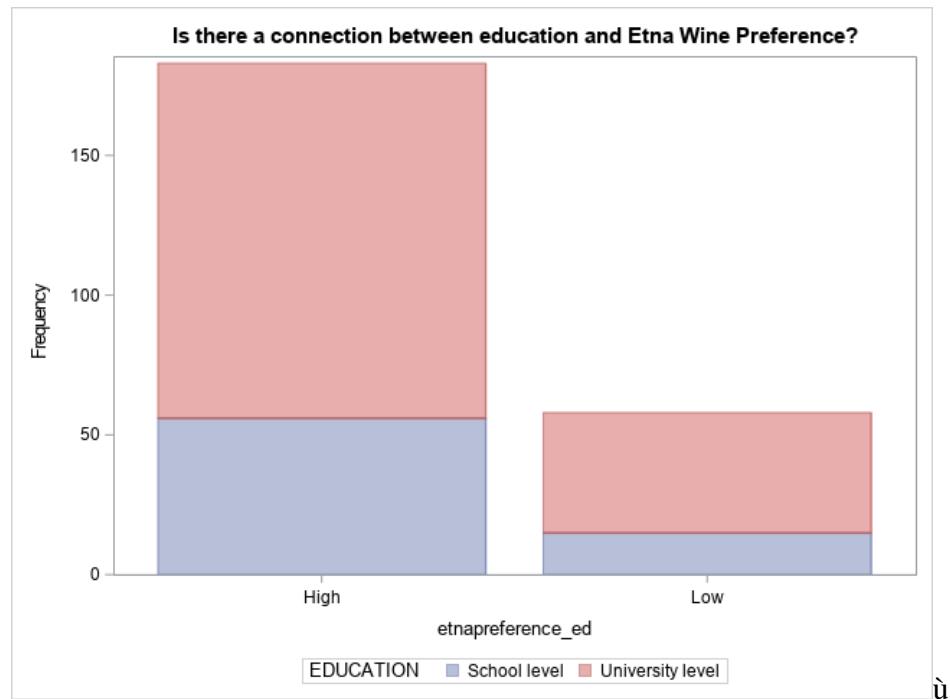
Contributions to the Total Chi-Square Statistic					
Percents	High	Low	Very_High	Very_Low	Sum
Over_	4.976	39.205	9.623	0.968	54.772
Under	4.109	32.374	7.946	0.799	45.228
Sum	9.085	71.579	17.569	1.767	100.000

The result of total chi-square did not confirm the connection, and we can see that the category "low" had more power in this result of chi-square statistic.

4.1.3. Is there a connection between education and Etna Wine Preference?

To analyze this connection in a statistically valid way, we divided all the levels of education in 2 groups: school level and university level. It was also necessary to group the Etna Wine preferences in only low and high preference, because otherwise we had an invalid test.

Thereby, in this third correspondence analysis we output the graph below grouped by the 2 levels of preference in Etna Wine and divided by color in each group of education. Visually, we can observe a higher number of university level in both of preferences.



		Statistic			DF	Value	Prob
Frequency		Chi-Square			1	0.4760	0.4903
		Likelihood Ratio Chi-Square			1	0.4844	0.4864
		Continuity Adj. Chi-Square			1	0.2752	0.5998
		Mantel-Haenszel Chi-Square			1	0.4740	0.4912
		Phi Coefficient				0.0444	
		Contingency Coefficient				0.0444	
		Cramer's V				0.0444	

In the table of frequency above, it is possible to demonstrate these numbers more clearly. There is 127 responses of university level in high preference against 56 of school level. And 43 responses of university level in low preference agains 15 of school level.

In the other table, its possible to see that the p-value is 0.4903, so we can conclude that there is no connection between education and Etna Wine preference.

Contingency Table				Chi-Square Statistic Expected Values		
Percents	High	Low	Sum	Percents	High	Low
School level	23.237	6.224	29.461	School level	22.3705	7.0901
University level	52.697	17.842	70.539	University level	53.5631	16.9763
Sum	75.934	24.066	100.000			

Since the p-value explained that there is no valid connection between education and Etna Wine preference, the differences between the contingency table and expected values are not so significant. We can only observe small differences: it was expected to have more university level in high preference than it happened, but the difference between it and the reality was very low.

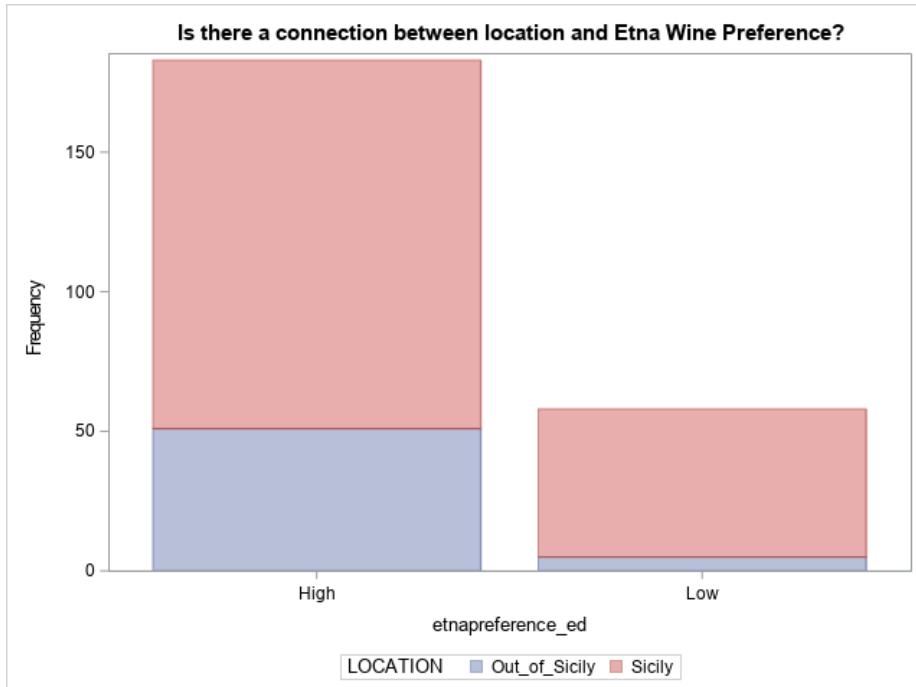
Contributions to the Total Chi-Square Statistic			
Percents	High	Low	Sum
School level	16.976	53.563	70.539
University level	7.090	22.370	29.461
Sum	24.066	75.934	100.000

The result of total chi-square did not confirm the connection, and we can see that the category “low” had more power in this result of chi-square statistic.

4.1.4. Is there a connection between location and Etna Wine Preference?

To analyze this connection in a statistically valid way, we divided all the locations in 2 groups: Sicily and out of Sicily. It was also necessary to group the Etna Wine preferences in only low and high preference, because otherwise we had an invalid test.

In this correspondence analysis we output the graph below grouped by the 2 levels of preference in Etna Wine and divided by color in each group of location. Visually, we can observe a higher number of people from Sicily in both of it.



Frequency	Table of LOCATION by etnapreference_ed			Statistic	DF	Value	Prob	
	LOCATION(LOCATION)	etnapreference_ed						
		High	Low	Total				
	Out_of_Sicily	51	5	56	Chi-Square	1	9.1478	0.0025
	Sicily	132	53	185	Likelihood Ratio Chi-Square	1	10.6690	0.0011
	Total	183	58	241	Continuity Adj. Chi-Square	1	8.1005	0.0044
				Mantel-Haenszel Chi-Square	1	9.1098	0.0025	
				Phi Coefficient		0.1948		
				Contingency Coefficient		0.1912		
				Cramer's V		0.1948		

In the table of frequency above, it is possible to demonstrate these numbers more clearly. There is 132 responses of Sicily people in High level of preference, and 51 from out of Sicily. And in low preference we have 53 responses from Sicily and 5 from out of Sicily.

In the other table, its possible to see that the p-value is 0.0025, so we can conclude that there is connection between the location and Etna Wine preference. It can be explained because people who were born or live in Sicily tend to like more the wines that are made in this region, for different reasons that come from tradition and emotional connection with the place. It also can be explained because people usually buy and know more wines from their cities or region because of lowest price when it compared to others. So its easier to know and even prefer these kind of wines.

Contingency Table				Chi-Square Statistic Expected Values		
Percents	High	Low	Sum	Percents	High	Low
Out_of_Sicily	21.162	2.075	23.237	Out_of_Sicily	17.6443	5.5922
Sicily	54.772	21.992	76.763	Sicily	58.2893	18.4742
Sum	75.934	24.066	100.000			

The difference between values in the contingency table and the Chi-Square Statistic Expected Values explain why the Chi-Square value has such a low p-value. If the location was not connected to Etna Wine preferences, the difference between the expected values and the actual values would have been smaller.

In this case we have a higher expectation that people from Sicily would have a high preference in Etna Wine, and we also has the expectation that people Out of Sicily would have a lower preference in Etna Wine.

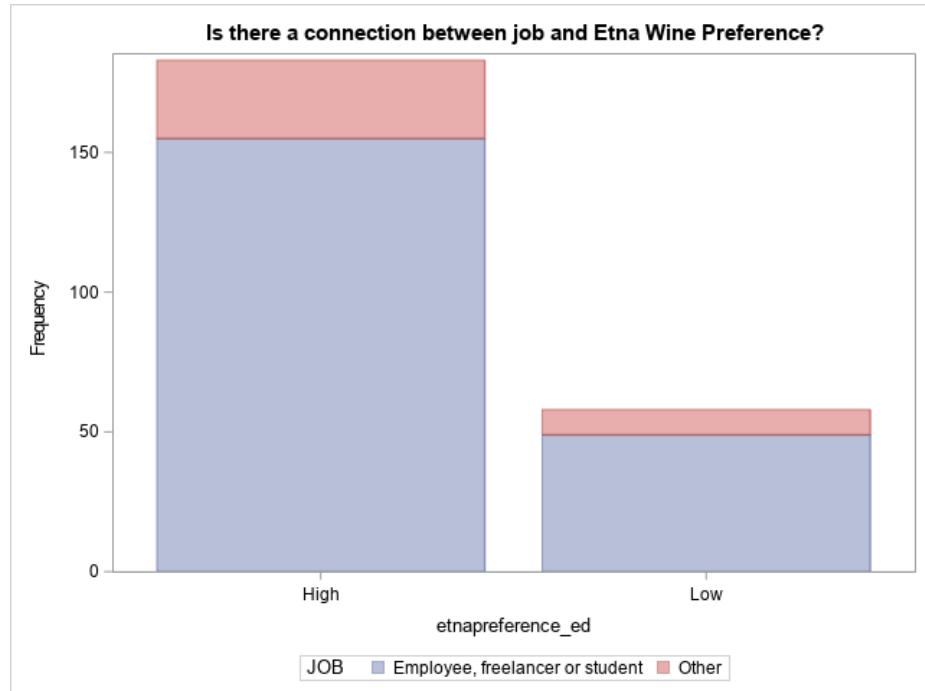
Contributions to the Total Chi-Square Statistic			
Percents	High	Low	Sum
Out_of_Sicily	18.474	58.289	76.763
Sicily	5.592	17.644	23.237
Sum	24.066	75.934	100.000

The category that has more importance than the others to this Chi-square result is out of Sicily, and low preference seems to be more important in establishing connection between the location and Etna Wine preference.

4.1.5. Is there a connection between job and Etna Wine Preference?

To analyze this connection in a statistically valid way, we divided all the jobs in 2 groups: employee, freelancer or students, and others. It was also necessary to group the Etna Wine preferences in only low and high preference, because otherwise we had an invalid test.

In this correspondence analysis we output the graph below grouped by the 2 levels of preference in Etna Wine and divided by color in each group of location. Visually, we can observe a higher number of employee, freelancer or student in both of it.



Frequency	Table of JOB by etnapreference_ed				Statistic	DF	Value	Prob				
	JOB(JOB)	etnapreference_ed										
		High	Low	Total								
	Employee, freelancer or student	155	49	204	Chi-Square	1	0.0016	0.9682				
	Other	28	9	37	Likelihood Ratio Chi-Square	1	0.0016	0.9682				
	Total	183	58	241	Continuity Adj. Chi-Square	1	0.0000	1.0000				
					Mantel-Haenszel Chi-Square	1	0.0016	0.9682				
					Phi Coefficient		0.0026					
					Contingency Coefficient		0.0026					
					Cramer's V		0.0026					

In the table of frequency above, it is possible to check these numbers. There 155 responses from employee, freelancer or student in high level of preference and 28 from others job in high preference.

In the other table, its possible to see that the p-value is 0.9682, so we can conclude that there is no connection between the location and Etna Wine preference.

Contingency Table			
Percents	High	Low	Sum
Employee, freelancer or student	64.315	20.332	84.647
Other	11.618	3.734	15.353
Sum	75.934	24.066	100.000

Chi-Square Statistic Expected Values		
Percents	High	Low
Employee, freelancer or student	64.2758	20.3716
Other	11.6579	3.6948

Since the p-value explained that there is no valid connection between job and Etna Wine preference, the differences between the contingency table and expected values are not so significant at all. We don't even need to point the smaller differences, because the values are almost the same.

Contributions to the Total Chi-Square Statistic			
Percents	High	Low	Sum
Employee, freelancer or student	3.695	11.658	15.353
Other	20.372	64.276	84.647
Sum	24.066	75.934	100.000

The result of total chi-square did not confirm the connection, and we can see that the category "other" and "low" had more power in this result of chi-square statistic.

4.2. Latent Class Analysis(LCA)

Latent Class Analysis (LCA) is a statistical method that allows to identify hidden population subgroups. These subgroups are unique from each other and individuals within a subgroup are similar to each other. Our final goal with Latent Class Analysis will be trying to identify groups of consumers in order to develop marketing decisions.

In order to achieve that we will not consider the demographic questions as determinant for the class assignments, but we will use them to understand more of the demographic composition of the subjects, this can allow to make specific decision based on a certain latent class (i.e. if we find out that young male subjects tend to prefer rose wine more than others, it can be the case to design a specific set of advertising, or other marketing strategies, specific aimed to male young subjects), instead if we decided to allow the demographic responses to contribute to the class creation, then more classes with potentially the same underlying behavior can be generated that differ for the demographic composition, hence creating a marketing decision specific for one of these classes can cut out a potential slice of subjects (that instead will be taken into account using the method described above).

Initially all the chosen items of the dataset will be transformed in order to make them dichotomous variables.

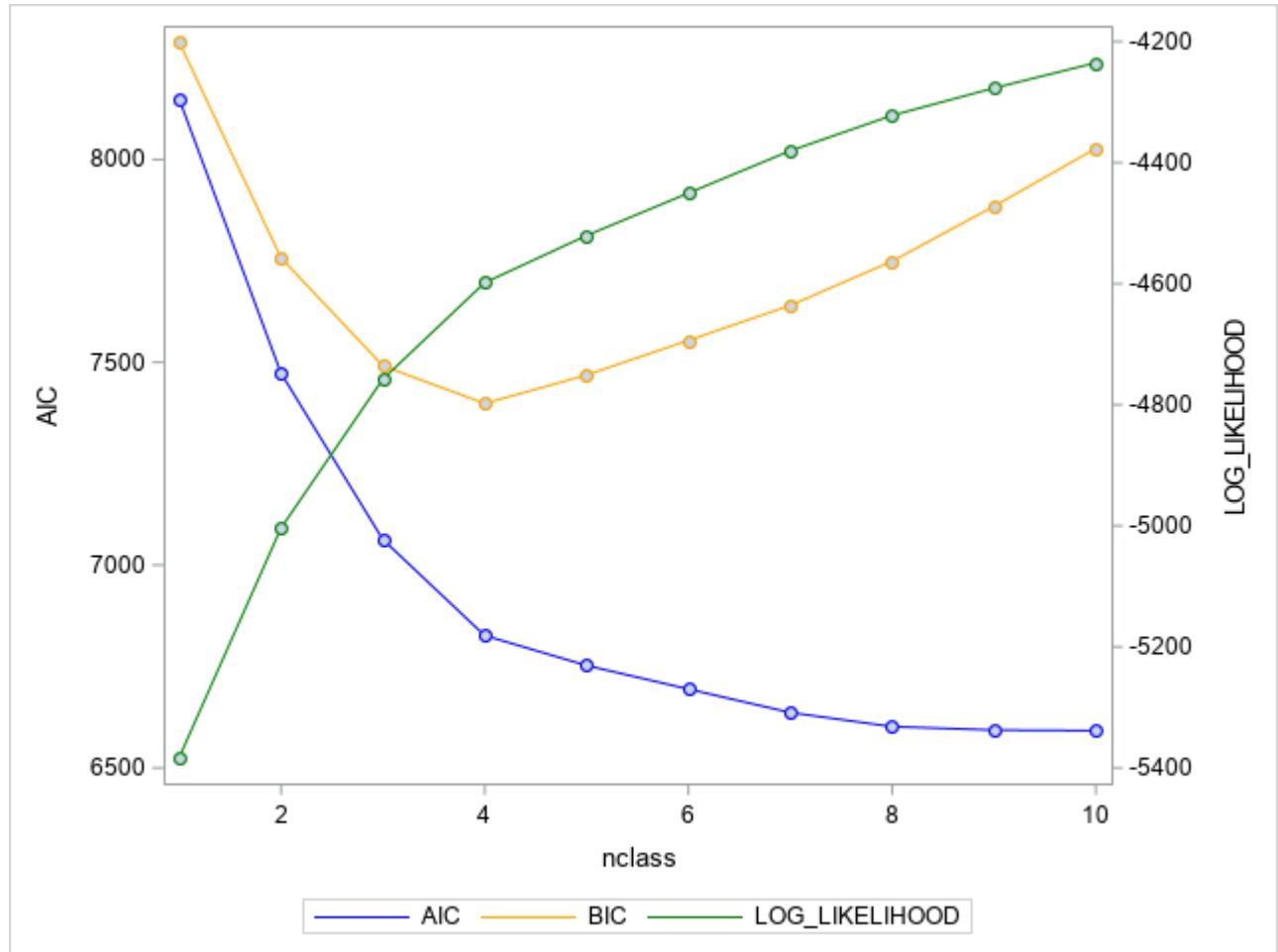
They have been modified in the following ways:

- *binary variables*: the values 0 have been mapped into 1 and the values 1 have been mapped into 2.
- *likert variables*: the values lower than or equal the central point (if any) have been mapped into 1 and the higher values into 2.
- *semantic variables*: the values lower than or equal the central point (if any) have been mapped into 1 and the higher values into 2.
- *Net promoter score variables*: the values lower than or equal the central point (if any) have been mapped into 1 and the higher values into 2.

4.2.1 Exploratory analysis

The first step is to identify how many latent classes are present in the respondents. Since the number of variables is relatively small, we will assume that there are at least 1 class and at most 10 different latent classes.

Let's see which number of classes gives the best splitting of the subjects.

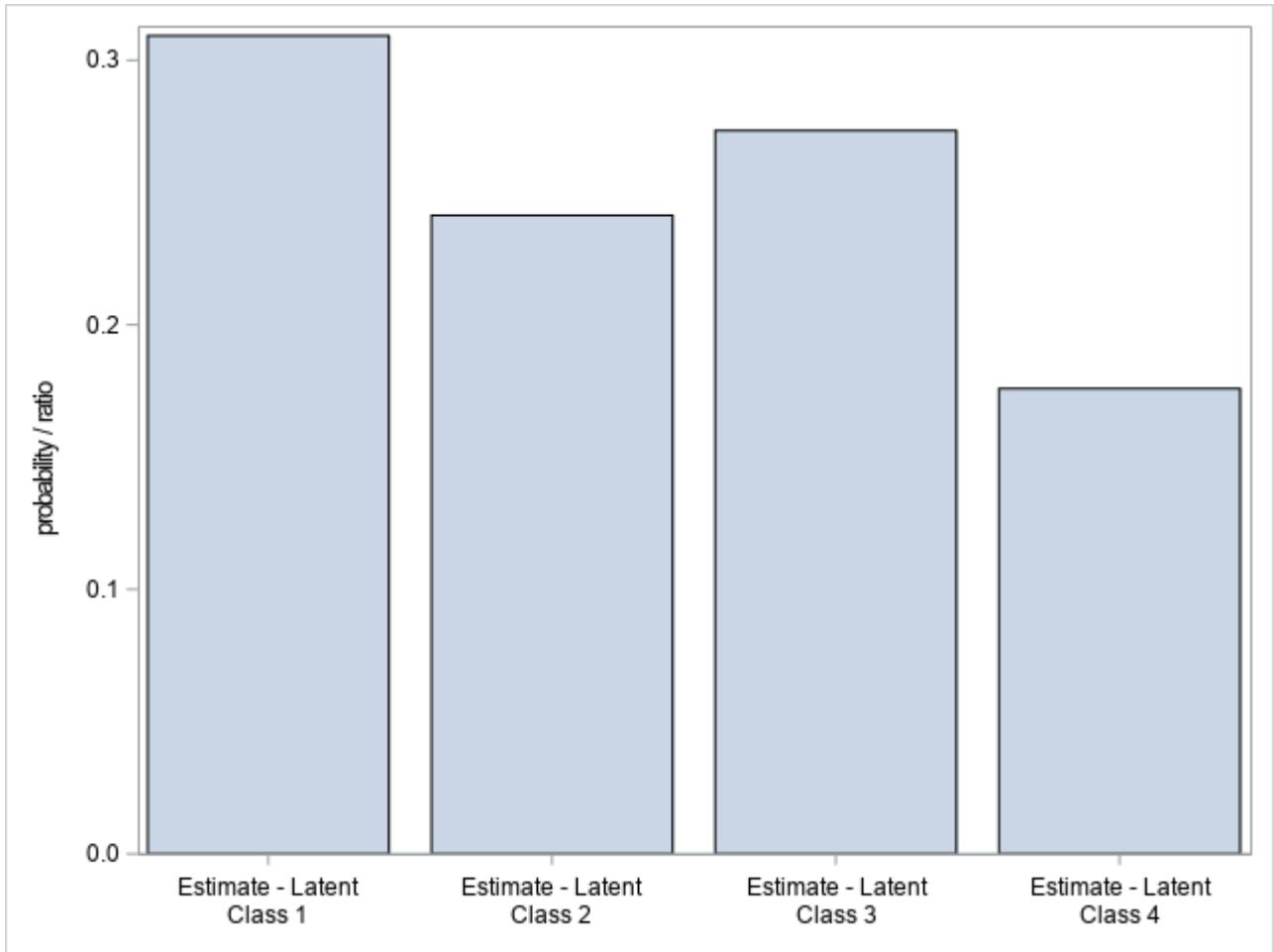


As can be seen from the plot above (AIC and BIC follow the left y-axis, instead Log Likelihood follows the one on the right), according to BIC the optimal number of classes is 4. The AIC curve is always decreasing, but as we can see the difference from moving from 3 classes to 4 classes is way bigger than the difference in moving from 4 to 5, hence using the 5th class does not provide a significative improvement.

As we can see the log_likelihood gives us the same conclusion as BIC did, adding a 5th class apport a much lower improvement than adding the 4th class.

As a conclusion we will use 4 different latent classes to describe the respondents.

The following plot shows how the 4 latent classes are distributed in the population (also could be interpreted as the a-priori probability for each of the 4 classes)



4.2.2 Class analysis

At this point we recomputed the latent classes only for the case with 4 classes we use “outparam” output which contains the conditional probability to give an answer belonging to a given class. The probability to belong to the classes are:

Estimate – Latent class 1	Estimate – Latent class 2	Estimate – Latent class 3	Estimate – Latent class 4
0.3093	0.2414	0.2734	0.1759

The model has identified 4 different classes which are fairly balanced: 30.93% of respondents were categorized in class 1, 24.14% categorized in class 2, 27.34% categorized in class 3 and 17.59% in class 4.

We now analyze the class membership probabilities to try to understand the unique characteristics of respondents in each class:

Variable name	Response Category 1				Response Category 2			
	Estimate-LC 1	Estimate-LC 2	Estimate-LC 3	Estimate-LC 4	Estimate-LC 1	Estimate-LC 2	Estimate-LC 3	Estimate-LC 4
wine_prefere	0.0012	0.3010	0.1484	0.6880	0.9988	0.6990	0.8516	0.3120
beer_prefere	0.1575	0.5262	0.3831	0.5454	0.8425	0.4738	0.6169	0.4546
soft_prefere	0.6865	0.5088	0.3916	0.2025	0.3135	0.4912	0.6084	0.7975
cocktail_pre	0.5984	0.4647	0.3365	0.5030	0.4016	0.5353	0.6635	0.4970
white_wine	0.2277	0.3428	0.2994	0.5505	0.7723	0.6572	0.7006	0.4495
rose_wine	0.5432	0.6712	0.4662	0.6441	0.4568	0.3288	0.5338	0.3559
red_wine	0.0443	0.3499	0.2647	0.6177	0.9557	0.6501	0.7353	0.3823
sparkling_wi	0.3268	0.6139	0.4258	0.5446	0.6732	0.3861	0.5742	0.4554
sweet_wine	0.5469	0.6212	0.5244	0.6848	0.4531	0.3788	0.4756	0.3152
wine_tasting	0.1052	0.7121	0.7993	0.8396	0.8948	0.2879	0.2007	0.1604
winery_visit	0.0454	0.4015	0.5933	0.7017	0.9546	0.5985	0.4067	0.2983
wine_course	0.4937	0.9620	0.9991	0.9304	0.5063	0.0380	0.0009	0.0696
wine_knowled	0.6468	0.9670	0.9696	0.9306	0.3532	0.0330	0.0304	0.0694
buying_exper	0.6573	0.9657	0.8555	0.9991	0.3427	0.0343	0.1445	0.0009
wine_bottles	0.6584	0.9994	0.9133	0.9993	0.3416	0.0006	0.0867	0.0007
supermarket	0.4705	0.2085	0.1717	0.0014	0.5295	0.7915	0.8283	0.9986
wine_shop	0.4212	0.8668	0.8696	0.9986	0.5788	0.1332	0.1304	0.0014
online_shop	0.8566	0.9833	0.9697	0.9997	0.1434	0.0167	0.0303	0.0003
grape_origin	0.1039	0.4029	0.4440	0.0014	0.8961	0.5971	0.5560	0.9986
grape_variet	0.1286	0.4190	0.5647	0.0017	0.8714	0.5810	0.4353	0.9983
budget_frien	0.2449	0.1764	0.2183	0.0010	0.7551	0.8236	0.7817	0.9990
brand_awarne	0.3779	0.2514	0.2680	0.0014	0.6221	0.7486	0.7320	0.9986
vintage	0.4455	0.7762	0.7001	0.9982	0.5545	0.2238	0.2999	0.0018
label_info	0.4476	0.4350	0.3970	0.0021	0.5524	0.5650	0.6030	0.9979
packaging	0.6647	0.7889	0.4410	0.9982	0.3353	0.2111	0.5590	0.0018
promotion	0.7112	0.6002	0.5255	0.9982	0.2888	0.3998	0.4745	0.0018
bottle_budge	0.9664	0.9757	0.9707	0.9999	0.0336	0.0243	0.0293	0.0001
buying_frequ	0.7329	0.7878	0.6016	0.9986	0.2671	0.2122	0.3984	0.0014
party	0.5124	0.5489	0.4553	0.9977	0.4876	0.4511	0.5447	0.0023
gift	0.5691	0.8203	0.7137	0.9985	0.4309	0.1797	0.2863	0.0015
home	0.1471	0.4863	0.2430	0.9965	0.8529	0.5137	0.7570	0.0035
taste	0.5120	0.8981	0.7688	0.9986	0.4880	0.1019	0.2312	0.0014
etna_doc	0.0907	0.0019	0.9947	0.8227	0.9093	0.9981	0.0053	0.1773
etna_buying	0.1313	0.4873	0.9983	0.8852	0.8687	0.5127	0.0017	0.1148
etna_prefere	0.3011	0.5904	0.0009	0.0402	0.6989	0.4096	0.9991	0.9598
etna_flavor	0.1821	0.4195	0.0006	0.0010	0.8179	0.5805	0.9994	0.9990
sicilian_exc	0.1177	0.3347	0.0004	0.0007	0.8823	0.6653	0.9996	0.9993
etna_expensi	0.6848	0.8311	0.9994	0.9120	0.3152	0.1689	0.0006	0.0880
etna_quality	0.0729	0.5759	0.0006	0.0012	0.9271	0.4241	0.9994	0.9988
etna_recomme	0.0247	0.2194	0.0002	0.0003	0.9753	0.7806	0.9998	0.9997

We highlighted in red the probability greater than 50% to belong to a certain class and give as answer 1 (reduction) and in green the probability for answer 2 (growth). It can be noticed that there are several wine buying habits which are shared between the 4 classes. In particular, for all classes following activities are the activities people do more:

- Buying wine from supermarket
- Buying wines that they know about grape origins
- Buying wines that is budget frindly
- Buying known brand wines
- Buying etna wine because of its flavor
- Buying etna wine because they like to sponser sicilian excellences
- And recommending etna wine to friend and family

Instead the following activities are the one people do less:

- like sweet wine
- Having knowledge about wine

- Buying lots of wine per month
- Buying lots of bottle of wine
- Buying wine online
- Buying wine because of promotion
- Spending lots of money on bottle of wine
- Buying wine frequently
- Buying wine as a gift
- Buying wine just for taste
- Believing wine is expensive

Thanks to the table above we can identify some specific patterns in the 4 classes.

Class 1 is composed by those subjects who are more expert in wine, they have taken winery classes and take into account several aspects when they buy wine. It's interesting to see that this type of subjects seems to don't like cocktails, soft drinks and sweet wines and tend to consume wine at home.

Class 2 subjects share some patterns with class 1 items, but they seem to have less interest in the quality of wine they consume, focusing more also on other types of drinks.

Class 3 subjects seem to enjoy almost every type of drinks, almost not caring about the quality of the wine but more interested in how wine is presented. Consider that they are the only ones who expressed consumption in parties it could make sense to start assuming that in this class there are the youngest subjects.

Class 4 subjects are the least interested in wine or any other drinks.

In the following you can see number of respondent in each class:

Class	Frequency	Percent
1	76	30.77
2	59	23.89
3	68	27.53
4	44	17.81

Gender

For the variable Gender we can see that class 1 is not really balanced and there is no data from "prefer not to say" in this class, while other classes are almost balanced and have a larger presence of Female.

Gender	Class 1		Class 2		Class 3		Class 4	
	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
Female	25	32.89	31	52.54	39	57.35	20	45.45
Male	51	67.11	26	44.07	28	41.18	21	47.73
Prefer Not To Say			2	3.39	1	1.47	3	6.82

Age

For the variable age we can see that in class 2 is balanced between groups of ages. Can be noted that the assumption that class 3 is mainly composed by young people is actually true.

Age_Class	Class 1		Class 2		Class 3		Class 4	
	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
18-24	7	9.21	16	27.12	23	33.82	16	36.36
25-30	14	18.42	19	32.20	24	35.29	15	34.09
31-50	39	51.32	14	23.73	15	22.06	10	22.73
>50	16	21.05	10	16.95	6	8.82	3	6.82

Location

In general in all 4 classes the majority of respondents are from Sicily. Also most of EU people libing in Italy are in class 3 and most of non EU people are in class 4. Also people from other part of Italy moslty are in class 3.

LOCATION	Class 1		Class 2		Class 3		Class 4	
	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
Foreign EU country (no Italy)	4	5.26	1	1.69	9	13.24	5	11.36
Foreign not EU country (no Italy)	4	5.26	1	1.69	10	14.71	10	22.73
Other region in Italy (no Sicily)	6	7.89	2	3.39	8	11.76	1	2.27
Sicily	62	81.58	55	93.22	41	60.29	28	63.64

Education

In general in all 4 classes the majority of respondents has university degree. Also number of primary school and medium school education level in all 4 groups are low and actually there is nothing in class 4.

Education	Class 1		Class 2		Class 3		Class 4	
	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
High School	13	17.11	21	35.59	23	33.82	12	27.27
Primary And Medium School	1	1.32	1	1.69	1	1.47		
University Degree	62	81.58	37	62.71	44	64.71	32	72.73

Job

As we can see more employee classified in class 1, more students are in class 2,3,4. Also we have one farmer which classified in class 3, one financier in class 4, 3 teachers in class 2 and one business owner in class 4.

4.3. Principal Component Analysis(PCA)

The Principal Component Analysis, helps us to understand if it is possible to reduce the number of variables. First we find the correlation between numeric variables:

Pearson Correlation Coefficients, N = 247 Prob > r under H0: Rho=0													
	WINE_PREFERENCE	BEER_PREFERENCE	SOFT_PREFERENCE	COCKTAIL_PREFERENCE	WHITE_WINE	ROSE_WINE	RED_WINE	SPARKLING_WINE	SWEET_WINE	WINE_TASTING	WINERY_VISIT	WINE_COURSE	
WINE_PREFERENCE	1.00000	0.43672 <.0001	-0.20879 0.0010	0.15177 0.0170	0.45518 <.0001	0.27658 <.0001	0.73662 0.0042	0.18167 0.0049	0.17857 0.0049	0.37608 <.0001	0.23008 0.0003	0.23606 0.0002	
BEER_PREFERENCE	0.43672 <.0001	1.00000	-0.16894 0.0078	0.08301 0.1935	0.22125 0.0005	0.15039 0.0160	0.38122 0.0005	0.13735 0.0309	0.10356 0.0144	0.10439 0.0107	0.05905 0.3554	0.07862 0.2182	
SOFT_PREFERENCE	-0.20879 0.0010	-0.16894 0.0078	1.00000	0.20659 0.0011	-0.02510 0.6946	0.11357 0.0748	-0.13262 0.0373	0.01750 0.5085	0.08153 0.0485	-0.23733 0.0115	-0.25240 0.0850	-0.16973 0.0075	
COCKTAIL_PREFERENCE	0.15177 0.0170	0.08301 0.1935	0.20659 0.0011	1.00000	0.25994 <.0001	0.25393 <.0001	0.04227 0.5085	0.12570 0.0485	0.16058 0.0115	-0.10982 0.0850	-0.07700 0.2279	-0.18716 0.0031	
WHITE_WINE	0.45518 <.0001	0.22125 0.0005	-0.02510 0.6946	0.25994 <.0001	1.00000 0.0000	0.46968 0.0001	0.31842 0.0001	0.39532 0.0001	0.24254 0.0001	0.09624 0.1315	-0.00066 0.9917	0.05251 0.4113	
ROSE_WINE	0.27858 <.0001	0.15039 0.0160	0.11357 0.0748	0.25393 0.0001	0.46968 0.0001	1.00000 0.0005	0.21932 0.0005	0.39246 0.0001	0.37166 0.0001	0.02512 0.6944	-0.05063 0.4264	0.06383 0.3177	
RED_WINE	0.73662 <.0001	0.38122 0.0373	-0.13262 0.0373	0.04227 0.5085	0.31842 0.0001	0.21932 0.0005	1.00000 0.0002	0.23507 0.0002	0.28945 0.0001	0.29717 0.0001	0.20480 0.0012	0.18869 0.0029	
SPARKLING_WINE	0.18167 0.0042	0.13735 0.0309	0.01750 0.7843	0.12570 0.0485	0.39532 0.0001	0.39246 0.0001	0.23507 0.0002	1.00000 0.0000	0.38052 0.0001	0.07126 0.2646	0.06189 0.3327	0.12399 0.0516	
SWEET_WINE	0.17857 0.0049	0.10356 0.1044	0.08153 0.2016	0.16058 0.0115	0.24254 0.0001	0.37166 0.0001	0.28945 0.0001	0.38052 0.0001	1.00000 0.0000	-0.00795 0.9010	0.01195 0.8518	0.07174 0.2613	
WINE_TASTING	0.37608 <.0001	0.10439 0.1017	-0.23733 0.0002	0.10982 0.0650	0.09624 0.1315	0.02512 0.6944	0.29717 0.0264	0.07126 0.0910	-0.00795 0.0001	1.00000 0.0000	0.56614 0.8518	0.51557 0.0001	
WINERY_VISIT	0.23008 0.0003	0.05905 0.3554	-0.25240 0.0001	-0.07700 0.2279	-0.00066 0.9917	-0.05083 0.4264	0.020480 0.0012	0.06189 0.3327	0.01195 0.0516	0.58614 0.2613	0.00000 0.0001	0.33431 0.0001	
WINE_COURSE	0.23606 0.0002	0.07862 0.2162	-0.16973 0.0075	-0.18716 0.0031	0.05251 0.4113	0.06383 0.3177	0.18869 0.0029	0.12399 0.0016	0.07174 0.0516	0.51557 0.2613	0.33431 0.0001	1.00000 0.0001	

Pearson Correlation Coefficients, N = 247 Prob > r under H0: Rho=0													
	SUPERMARKET	WINE_SHOP	ONLINE_SHOP	GRAPE_ORIGIN	GRAPE_VARIETY	BUDGET_FRIENDLY	BRAND_AWARENESS	VINTAGE	LABEL_INFO	PACKAGING	PROMOTION		
SUPERMARKET	1.00000 <.0001	-0.49223 0.0001	-0.24561 0.0004	-0.22225 0.0003	-0.22570 0.0003	0.28582 0.0003	0.17696 0.0053	-0.15895 0.0124	0.06914 0.2791	0.25436 0.0001	0.23864 0.0002		
WINE_SHOP	-0.49223 0.0001	1.00000 0.0013	0.20348 0.0013	0.36155 0.0001	0.40643 0.0003	-0.23095 0.0003	-0.13603 0.0326	0.23711 0.0002	-0.02923 0.6475	-0.18671 0.0032	-0.32522 0.0001		
ONLINE_SHOP	-0.24561 0.0001	0.20348 0.0013	1.00000 0.0000	0.16269 0.0104	0.15574 0.0143	-0.03097 0.0014	-0.13689 0.0315	0.18770 0.0031	-0.05355 0.4021	0.00462 0.9424	0.05792 0.3647		
GRAPE_ORIGIN	-0.22225 0.0004	0.36155 0.0001	0.16269 0.0104	1.00000 0.0000	0.53277 0.0001	0.04456 0.4857	0.01059 0.1783	0.19128 0.0025	0.11730 0.0057	-0.05374 0.4004	-0.24655 0.0001		
GRAPE_VARIETY	-0.22570 0.0003	0.40643 0.0143	0.15574 0.0143	0.53277 0.0001	1.00000 0.0000	-0.08593 0.1783	-0.07780 0.1783	0.12425 0.2231	-0.05907 0.0511	-0.18234 0.3552	-0.21136 0.0004		
BUDGET_FRIENDLY	0.28582 0.0001	-0.23095 0.0003	-0.03097 0.6281	0.04456 0.4857	-0.08593 0.1783	1.00000 0.0000	0.25927 0.0001	0.01508 0.8136	0.03340 0.6014	0.20677 0.0011	0.44072 0.0001		
BRAND_AWARENESS	0.17656 0.0053	-0.13603 0.0326	-0.13669 0.0315	0.01059 0.6685	-0.07780 0.2231	0.25927 0.0001	1.00000 0.0002	0.19517 0.0021	0.26091 0.0001	0.21699 0.0006	0.16233 0.0106		
VINTAGE	-0.15895 0.0124	0.23711 0.0002	0.18770 0.0031	0.19128 0.0025	0.12425 0.0025	0.01508 0.0511	0.19517 0.8136	1.00000 0.0001	0.41046 0.0013	0.20363 0.0001	-0.01335 0.8347		
LABEL_INFO	0.06914 0.2791	-0.02923 0.6475	-0.05355 0.4021	0.11730 0.0657	-0.05907 0.3552	0.03340 0.6014	0.26091 0.0001	0.41046 0.0001	1.00000 0.0001	0.19016 0.0027	-0.01988 0.7559		
PACKAGING	0.25436 0.0001	-0.18671 0.0032	0.00462 0.9424	-0.05374 0.4004	-0.18234 0.0040	0.20677 0.0011	0.21699 0.0006	0.20363 0.0013	0.19016 0.0027	1.00000 0.0001	0.37954 0.0001		
PROMOTION	0.23864 0.0002	-0.32522 0.0001	0.05792 0.3647	-0.24655 0.0001	-0.21136 0.0008	0.44072 0.0001	0.16233 0.0106	-0.01335 0.8347	-0.01988 0.7559	0.37954 0.0001	1.00000 0.0001		

Pearson Correlation Coefficients, N = 247 Prob > r under H0: Rho=0													
	PARTY	GIFT	HOME	TASTE	ETNA_DOC	ETNA_BUYING	ETNA_PREFERENCE	ETNA_FLAVOR	SICILIAN_EXCELLENCES	ETNA_EXPENSIVE	ETNA_QUALITY	ETNA_RECOMMENDATION	
PARTY	1.00000 0.0001	0.30679 0.1987	0.08205 0.1987	0.11346 0.0751	0.08251 0.1962	0.10858 0.1386	0.09044 0.0518	0.03257 0.1141	-0.04883 0.0930	-0.01441 0.4448	0.03542 0.5795	-0.00594 0.9260	
GIFT	0.30679 0.0001	1.00000 0.0001	0.31232 0.0001	0.36925 0.0001	0.09450 0.1366	0.17778 0.0501	0.05761 0.3763	-0.00286 0.9644	0.07796 0.2221	-0.02629 0.6810	0.10649 0.0950	0.01305 0.8383	
HOME	0.08205 0.1987	1.00000 0.0001	0.28796 0.0001	0.12389 0.0518	0.19381 0.0222	-0.01998 0.7547	0.02020 0.7521	0.06076 0.7521	0.08558 0.3416	0.10313 0.1059	0.01015 0.0739		
TASTE	0.11346 0.0751	0.36925 0.0001	0.28796 0.0001	1.00000 0.1141	0.10078 0.0007	0.21478 0.0007	0.01326 0.0930	-0.00560 0.9302	-0.07231 0.2575	0.04578 0.4738	0.09390 0.1412	-0.03518 0.5822	
ETNA_DOC	0.08251 0.1962	0.09450 0.1386	0.12389 0.0518	0.10078 0.1141	1.00000 0.0001	0.43258 0.0226	-0.34405 0.0322	-0.06319 0.0276	0.02659 0.6775	0.14020 0.0276	-0.01108 0.8625	-0.45674 0.0001	
ETNABuying	0.10858 0.0866	0.17778 0.0002	0.19381 0.0002	0.21478 0.0837	0.43258 0.0326	1.00000 0.0001	0.29813 0.0001	0.34180 0.0001	0.13275 0.0371	0.06764 0.2896	0.39839 0.0001	0.17869 0.0049	
ETNA_PREFERENCE	0.09044 0.1565	0.05761 0.3673	-0.01998 0.7547	0.01326 0.0837	-0.34405 0.0001	0.29813 0.0001	1.00000 0.0001	0.62285 0.0001	0.26376 0.0001	0.11822 0.0636	0.48035 0.0001	0.62759 0.0001	
ETNA_FLAVOR	0.03257 0.6104	-0.02866 0.9644	0.02020 0.7521	-0.00560 0.8625	-0.06319 0.0276	0.34180 0.0001	0.62285 0.0001	1.00000 0.0001	0.40627 0.0001	0.37141 0.0001	0.57655 0.0001	0.57432 0.0001	
SICILIAN_EXCELLENCES	-0.04883 0.4448	0.07796 0.2221	0.06076 0.1412	-0.07231 0.0575	0.02659 0.0276	0.13275 0.0371	0.26376 0.0001	0.40627 0.0001	1.00000 0.0001	0.25605 0.0001	0.56221 0.0001	0.43437 0.0001	
ETNA_EXPENSIVE	-0.01441 0.8217	-0.02829 0.6810	0.08558 0.1800	0.04578 0.4738	0.14020 0.0276	0.06764 0.0276	0.11622 0.0636	0.37141 0.0001	0.25605 0.0001	1.00000 0.0001	0.29508 0.0001	0.07975 0.2116	
ETNA_QUALITY	0.03542 0.5795	0.10649 0.0950	0.10133 0.1059	0.09390 0.1412	-0.01108 0.0625	0.39839 0.0001	0.48035 0.0001	0.57655 0.0001	0.56221 0.0001	0.29508 0.0001	1.00000 0.0001	0.56307 0.0001	
ETNA_RECOMMENDATION	-0.00594 0.9260	0.01305 0.8383	0.01015 0.8739	-0.03516 0.5822	-0.45674 0.0001	0.17869 0.0001	0.62759 0.0001	0.43437 0.0001	0.07975 0.2116	0.56307 0.0001	1.00000 0.0001		

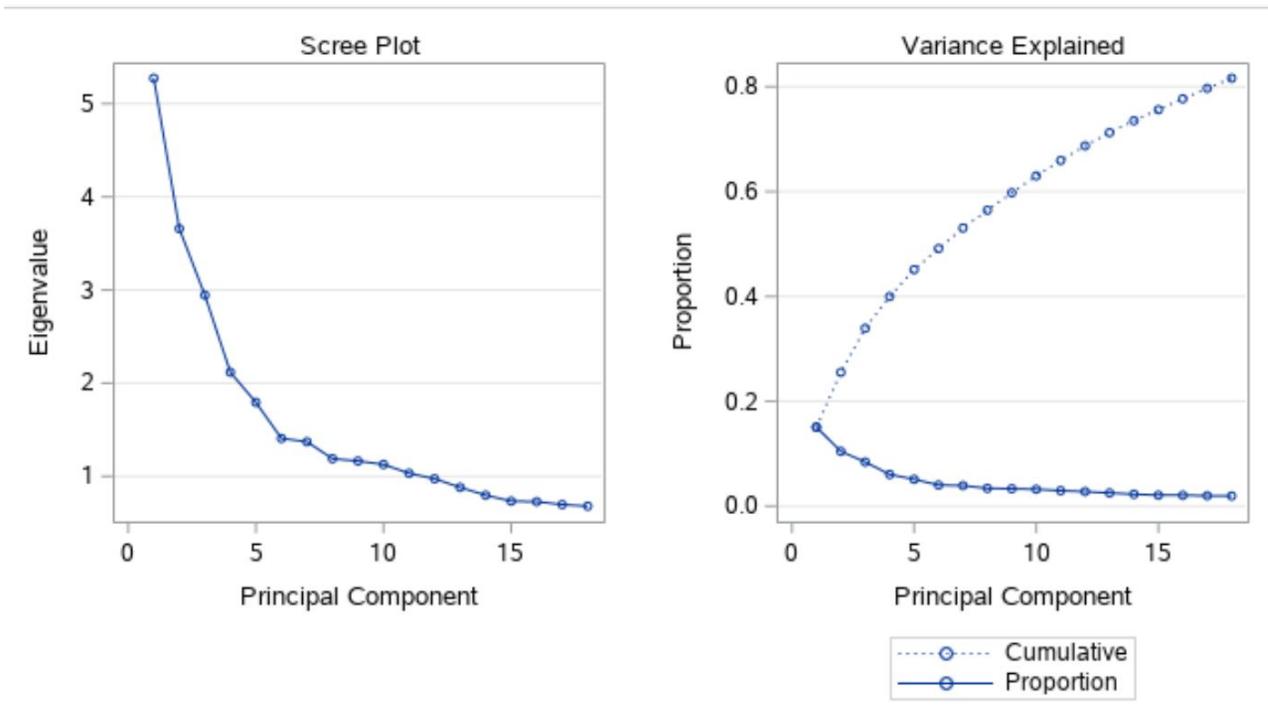
Then we need to find eigenvalue of each variable.

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	5.27349812	1.61486128	0.1507	0.1507
2	3.65863685	0.71333647	0.1045	0.2552
3	2.94530037	0.82861878	0.0842	0.3394
4	2.11668159	0.32297449	0.0605	0.3998
5	1.79370710	0.38679808	0.0512	0.4511
6	1.40690903	0.03590388	0.0402	0.4913
7	1.37100515	0.18187587	0.0392	0.5304
8	1.18912928	0.02488012	0.0340	0.5644
9	1.16424916	0.03653097	0.0333	0.5977
10	1.12771819	0.09546558	0.0322	0.6299
11	1.03225260	0.05858681	0.0295	0.6594
12	0.97366580	0.09236518	0.0278	0.6872
13	0.88130061	0.08398007	0.0252	0.7124
14	0.79732054	0.06422712	0.0228	0.7352
15	0.73309342	0.00552923	0.0209	0.7561
16	0.72756419	0.03096631	0.0208	0.7769
17	0.69659788	0.01684502	0.0199	0.7968
18	0.67975285	0.06094944	0.0194	0.8162
19	0.61880341	0.05652422	0.0177	0.8339
20	0.56227919	0.03459215	0.0161	0.8500
21	0.52768704	0.01162482	0.0151	0.8651
22	0.51606222	0.04200639	0.0147	0.8798
23	0.47405582	0.02172551	0.0135	0.8934
24	0.45233032	0.01112790	0.0129	0.9063
25	0.44120242	0.05369039	0.0126	0.9189
26	0.38751203	0.01230462	0.0111	0.9300
27	0.37520741	0.01926518	0.0107	0.9407
28	0.35594222	0.03986878	0.0102	0.9508
29	0.31607344	0.01908119	0.0090	0.9599
30	0.29699225	0.02407783	0.0085	0.9684
31	0.27291442	0.01490798	0.0078	0.9762
32	0.25800644	0.00893657	0.0074	0.9835
33	0.24906987	0.06804728	0.0071	0.9906
34	0.18102259	0.03456641	0.0052	0.9958
35	0.14645618		0.0042	1.0000

In order to choose how many Principal Components to keep, the eigenvalue associated with each component is considered. Only components with an eigenvalue > 1 will be kept. Which it will be 11 but since it just explain 65% of variance So we decide to keep the 18 principal components to explain atleast 80% of total variance

The eigenvector of these 18 principal components are as follow:

	Eigenvectors																	
	Print1	Print2	Print3	Print4	Print5	Print6	Print7	Print8	Print9	Print10	Print11	Print12	Print13	Print14	Print15	Print16	Print17	Print18
WINE_PREFERENCE	0.290674	0.024502	0.253179	0.033388	-0.100290	-0.038368	-0.267710	-0.137680	-0.120737	-0.020797	-0.190420	-0.013942	-0.074430	-0.089298	0.034356	0.051480	-0.064471	-0.072035
BEER_PREFERENCE	0.146659	0.038319	0.153101	-0.029073	-0.260343	-0.017717	-0.352376	-0.056184	-0.083971	-0.250527	-0.037068	-0.223084	0.081684	0.264014	0.079148	0.079452	0.217625	0.398574
SOFT_PREFERENCE	-1.133341	0.125287	0.018596	-0.187802	0.010434	0.030196	0.191027	0.130524	-0.095629	0.271452	-0.320510	0.428109	-0.060125	0.181643	0.451893	0.287202	0.105009	0.142070
COCKTAIL_PREFERENCE	-0.011445	0.174489	0.161250	-0.277341	-0.005613	0.028016	-0.020648	-0.270050	-0.308756	0.232147	-0.030442	0.073784	0.463372	0.078943	-0.012669	-0.202901	-0.131512	0.054666
WHITE_WINE	0.142293	0.081790	0.332820	-0.144428	-0.108842	-0.153475	0.039833	-0.123668	0.077368	0.175553	0.111554	-0.011139	-0.070528	-0.196393	-0.091870	0.083559	-0.200854	0.219479
ROSE_WINE	0.101607	0.146152	0.272063	-0.201271	-0.150761	-0.088853	0.263069	0.056049	-0.056534	0.170787	0.042631	-0.084348	-0.147494	-0.134284	-0.020564	-0.052335	-0.041783	-0.015359
RED_WINE	0.253831	0.045334	0.243285	-0.020801	-0.037540	0.011137	-0.300212	0.092621	-0.068804	-0.148481	-0.199220	0.032408	-0.201058	0.047245	0.000316	0.049949	-0.028398	-0.105038
SPARKLING_WINE	0.116789	0.051665	0.260313	-0.200800	-0.085448	0.006094	0.352245	0.110434	0.128402	-0.061817	0.222064	-0.023109	-0.102506	0.144905	-0.297303	0.178462	0.262092	0.090512
SWEET_WINE	0.100461	0.119608	0.215583	-0.241250	-0.007180	0.177738	0.148125	0.340986	0.164964	-0.018986	-0.105586	-0.183308	-0.061095	0.048245	0.087158	-0.372099	0.144760	-0.359993
WINE_TASTING	0.293334	-0.126151	-0.045961	0.123388	0.084564	0.018565	0.047442	-0.020645	-0.223318	0.061216	0.159075	0.133718	-0.178355	0.212547	0.028807	-0.172952	0.069006	0.078162
WINERY_VISIT	0.232518	-0.169014	-0.042709	0.083667	0.143673	0.129647	-0.035861	0.095342	-0.180268	0.191090	0.287056	0.069212	-0.027343	0.356897	-0.153189	-0.214499	-0.094464	-0.082028
WINE_COURSE	0.266294	-0.080443	-0.074629	0.069973	0.173833	-0.103825	0.260168	-0.063927	-0.004383	-0.072194	-0.049833	-0.137406	-0.086652	0.155453	0.315672	-0.091416	0.357602	0.050530
SUPERMARKET	-1.146026	0.163464	0.265681	0.147527	-0.022245	-0.051254	-0.041541	0.227043	0.026164	0.024986	0.387901	0.043636	0.226591	-0.074324	0.202765	-0.026117	-0.025111	0.123076
WINE_SHOP	0.234978	-0.097933	-0.128909	-0.280496	-0.048847	0.123386	-0.028461	-0.077288	0.091645	0.030554	-0.205395	0.135574	-0.088646	0.168320	-0.283446	0.211689	-0.148468	-0.109519
ONLINE_SHOP	0.204623	-0.040284	-0.021240	0.036827	0.140079	-0.133431	0.321339	-0.279286	0.052812	-0.140317	-0.262198	0.068393	0.069613	-0.291726	0.029880	-0.353169	-0.136466	0.203286
GRAPE_ORIGIN	0.201579	0.021998	-0.135155	-0.239179	0.072511	-0.060169	-0.241481	0.036057	0.358189	-0.106898	0.197719	0.198756	0.230331	-0.166993	0.010532	-0.107024	0.170511	-0.108583
GRAPE_VARIETY	0.198990	-0.062274	-0.134776	-0.265557	-0.028430	-0.171957	-0.102483	-0.045121	0.231303	-0.191875	0.215000	0.375825	0.119678	-0.029299	0.139088	0.029650	0.023059	-0.133552
BUDGET_FRIENDLY	-0.030786	0.191045	0.184945	0.112657	0.228658	-0.351103	-0.139220	-0.052025	0.310311	-0.103810	0.022384	-0.056803	0.022235	0.136240	0.069279	-0.053118	-0.035336	
BRAND_AWARENESS	-0.061851	0.240411	0.045716	-0.09997	0.324134	0.041262	-0.243924	0.060541	0.046002	0.190450	0.026978	-0.123198	-0.127779	0.243116	0.152572	-0.262825	-0.261443	-0.027137
VINTAGE	0.162728	0.115676	0.006455	-0.157638	0.312360	0.321598	0.006399	-0.025587	-0.145210	-0.079873	-0.075524	-0.036286	0.042583	-0.151494	0.062463	0.228839	-0.058696	-0.107864
LABEL_INFO	-0.006399	0.187163	-0.060671	-0.143858	0.315895	0.346364	-0.046505	0.124438	-0.075753	-0.286808	0.191767	-0.135031	-0.152720	0.039993	0.208457	0.002129	0.208374	
PACKAGING	-0.026556	0.196704	0.188585	0.121764	0.249087	0.126483	0.047409	-0.267345	-0.202656	-0.094563	0.197344	0.306875	0.033846	0.007694	-0.251230	-0.004979	0.237705	-0.026630
PROMOTION	-0.065127	0.139518	0.211331	0.307921	0.248300	-0.240736	0.077725	-0.101988	0.037620	-0.087707	-0.132973	0.234632	-0.194690	-0.067799	-0.163876	0.077653	-0.006225	-0.205572
PARTY	0.050781	0.048471	0.134083	0.230572	0.133174	0.259899	-0.145582	0.104736	0.310924	0.249863	-0.362902	0.061495	0.297622	-0.048051	-0.185401	-0.110875	-0.354894	0.196076
GIFT	0.147966	-0.053095	0.077957	0.305975	-0.129897	0.348672	0.054507	0.226133	0.222004	0.103987	0.016745	0.220700	-0.027229	-0.116118	-0.059047	-0.008997	-0.274291	0.135576
HOME	0.170810	-0.041022	0.159312	0.256311	-0.216579	0.074191	0.043649	0.077998	-0.265185	-0.126111	0.035039	0.038759	0.302510	-0.104008	0.222674	0.120300	0.050775	-0.487328
TASTE	0.217011	-0.081012	0.095076	0.145120	-0.029622	0.192904	0.188301	-0.151791	0.259067	-0.185411	0.070211	0.048613	0.089595	0.193393	0.278927	0.087459	-0.383079	0.128764
ETNA_DOC	0.197411	-0.220157	0.034021	0.016245	-0.264208	-0.259430	-0.117656	0.268018	-0.086592	0.261393	0.035122	0.068409	-0.028111	-0.227099	0.083430	0.016785	-0.025230	0.192577
ETNA_BUYING	0.262038	0.034400	-0.129429	0.147663	0.064604	-0.094233	0.018841	-0.100214	0.016733	0.430155	0.126648	-0.224475	0.046726	-0.106222	0.124966	0.236577	0.141159	-0.012620
ETNA_PREFERENCE	0.055284	0.350016	-0.182136	0.132430	-0.159345	0.1015750	0.052592	-0.142262	0.135919	0.161160	0.056894	-0.121388	0.002017	0.183125	0.048613	0.186089	0.019098	-0.082341
ETNA_FLAVOR	0.121922	0.364073	-0.192529	0.083418	-0.002446	0.088708	0.029566	0.093051	0.035122	0.068409	-0.028111	-0.227099	0.083430	0.016785	-0.102043	-0.102043	-0.184969	-0.167560
SICILIAN_EXCELLENCES	0.123626	0.224929	-0.188538	0.043910	-0.153322	-0.138117	-0.102098	0.318503	-0.207031	-0.067151	0.0308417	-0.198841	-0.16477	-0.062630	-0.192669	-0.031364	0.169883	
ETNA_EXPENSIVE	0.118280	0.121007	-0.110423	-0.013372	0.130710	-0.243850	0.211169	0.364529	-0.117237	-0.281765	-0.163899	0.000162	0.433853	0.243176	-0.206439	0.104290	-0.159645	0.112207
ETNA_QUALITY	0.229993	0.293316	-0.195292	0.045298	-0.067530	0.045848	0.018757	0.042858	-0.056348	0.020790	0.060150	-0.055797	-0.109582	0.019462	-0.069108	0.077900	0.131576	
ETNA_RECOMMENDATION	0.050584	0.375764	-0.194688	0.092803	-0.225787	0.068122	0.011915	-0.171787	0.005720	-0.022645	0.002304	0.049342	-0.089984	0.000580	-0.115299	-0.126964	0.079373	-0.021603

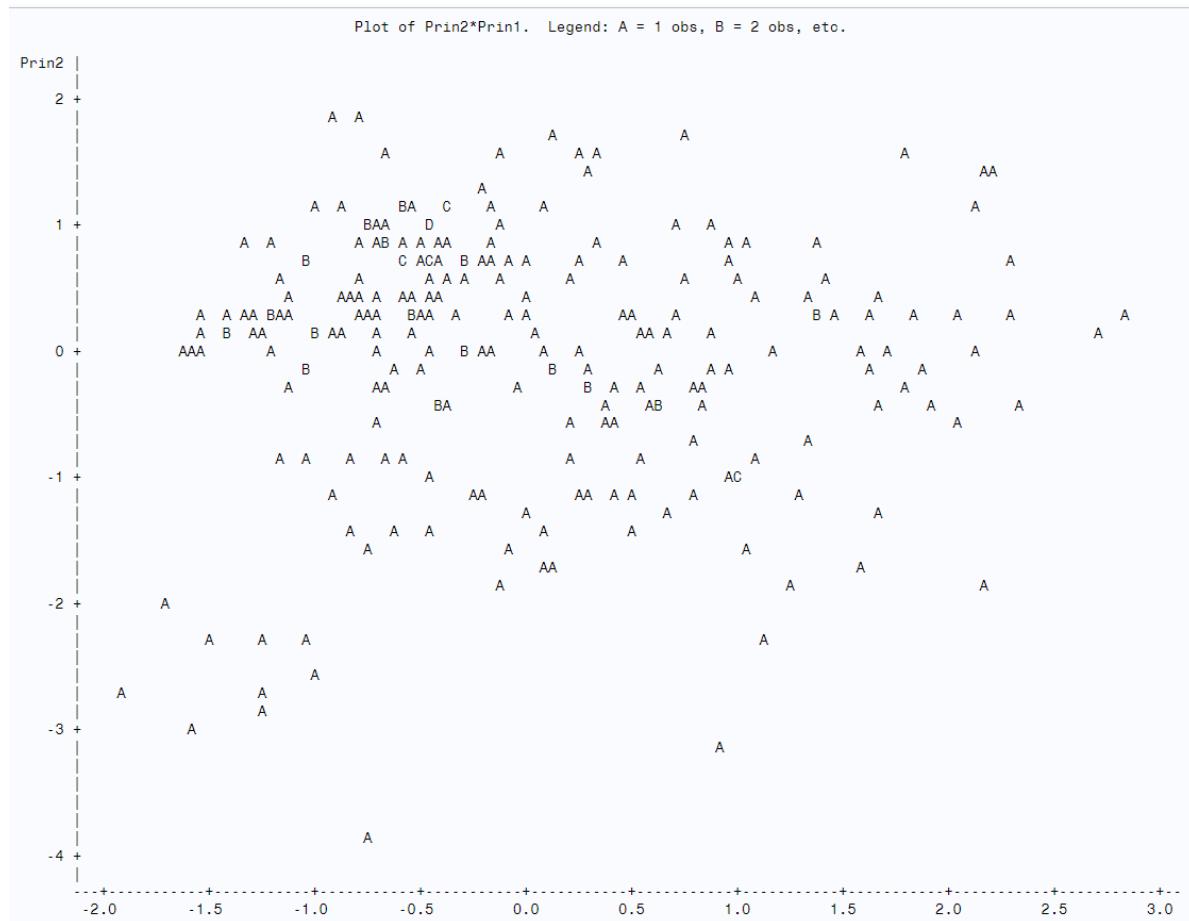


From the correlation matrix of Principal Components and the original variables, it seems that the first PC is more positively correlated with wine_preference, wine_tasting, wine_course. The second principal component is correlated with etna_recommendation, etna_flavor, etna_preference. The third principal component is correlated with white_wine, rose_wine, sparkling_wine. Even if the first three principal components make sense, it is difficult to interpret the others.

Obs	Type	Name	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9	Prin10	Prin11	Prin12	Prin13	Prin14	Prin15	Prin16	Prin17	Prin18
1	CORR	WINE_PREFERENCE	0.67	0.05	0.43	0.05	-0.13	-0.05	-0.31	-0.15	-0.13	-0.02	-0.19	-0.01	-0.07	-0.08	0.03	0.04	-0.05	-0.06
2	CORR	BEER_PREFERENCE	0.34	0.07	0.26	-0.04	-0.35	-0.02	-0.41	-0.06	-0.09	-0.27	-0.04	-0.22	0.08	0.24	0.07	0.07	0.18	0.33
3	CORR	SOFT_PREFERENCE	-0.31	0.24	0.03	-0.27	0.01	0.04	0.22	0.14	-0.10	0.29	-0.33	0.42	-0.06	0.16	0.39	0.24	0.09	0.12
4	CORR	COCKTAIL_PREFERENCE	-0.03	0.33	0.28	-0.40	-0.01	0.03	-0.02	-0.29	-0.33	0.25	-0.03	0.07	0.44	0.07	-0.01	-0.17	-0.11	0.05
5	CORR	WHITE_WINE	0.33	0.16	0.58	-0.21	-0.15	-0.18	0.05	-0.13	0.08	0.19	0.12	-0.01	-0.07	-0.18	-0.08	0.07	-0.17	0.18
6	CORR	ROSE_WINE	0.23	0.28	0.47	-0.29	-0.20	-0.11	0.31	0.06	-0.06	0.18	0.04	-0.08	-0.14	-0.12	-0.02	-0.04	-0.03	-0.01
7	CORR	RED_WINE	0.58	0.09	0.42	-0.03	-0.05	0.01	-0.35	0.10	-0.07	-0.16	-0.20	0.03	-0.19	0.04	0.00	0.04	-0.02	-0.09
8	CORR	SPARKLING_WINE	0.27	0.10	0.45	-0.29	-0.11	0.01	0.41	0.12	0.14	-0.07	0.23	-0.02	-0.10	0.13	-0.25	0.15	0.22	0.07
9	CORR	SWEET_WINE	0.23	0.23	0.37	-0.35	-0.01	0.21	0.17	0.37	0.18	-0.02	-0.11	-0.18	-0.06	0.04	0.07	-0.32	0.12	-0.30
10	CORR	WINE_TASTING	0.67	-0.24	-0.08	0.18	0.11	0.02	0.06	-0.02	-0.24	0.07	0.16	0.13	-0.17	0.19	0.02	-0.15	0.06	0.06
11	CORR	WINERY_VISIT	0.53	-0.32	-0.07	0.13	0.19	0.15	-0.04	0.10	-0.19	0.20	0.29	0.07	-0.03	0.32	-0.13	-0.18	-0.08	-0.01
12	CORR	WINE_COURSE	0.61	-0.15	-0.13	0.10	0.23	-0.12	0.30	-0.07	0.00	-0.08	-0.05	-0.14	-0.08	0.14	0.27	-0.08	0.30	0.04
13	CORR	SUPERMARKET	-0.34	0.31	0.46	0.21	-0.03	-0.06	0.02	0.25	0.03	0.03	0.39	0.04	0.21	-0.07	0.17	-0.02	-0.02	0.10
14	CORR	WINE_SHOP	0.54	-0.19	-0.22	-0.41	-0.07	0.15	-0.03	-0.08	0.10	0.03	-0.21	0.14	-0.08	0.15	-0.24	0.18	-0.12	-0.09
15	CORR	ONLINE_SHOP	0.47	-0.08	-0.04	0.05	0.19	-0.16	0.38	-0.30	0.06	-0.15	-0.27	-0.07	0.07	-0.26	0.03	-0.30	-0.11	0.17
16	CORR	GRAPE_ORIGIN	0.46	0.04	-0.23	-0.35	0.10	-0.07	-0.25	-0.04	0.39	-0.11	0.20	0.20	0.22	-0.15	0.01	-0.09	0.14	-0.09
17	CORR	GRAPE_VARIETY	0.45	-0.12	-0.23	-0.39	-0.04	-0.20	-0.12	-0.05	0.25	-0.02	0.22	0.37	0.11	-0.03	0.12	0.03	0.02	-0.11
18	CORR	BUDGET_FRIENDLY	-0.07	0.37	0.32	0.16	0.31	-0.42	-0.16	-0.06	0.33	-0.11	0.02	0.02	-0.06	0.20	0.12	0.06	-0.04	-0.03
19	CORR	BRAND_AWARENESS	-0.14	0.46	0.08	-0.14	0.43	0.05	-0.29	0.07	0.05	0.20	0.03	-0.12	-0.12	0.22	0.13	-0.22	-0.22	-0.02
20	CORR	VINTAGE	0.37	0.22	0.01	-0.23	0.50	0.38	0.01	-0.10	-0.16	-0.08	-0.08	-0.04	0.04	-0.14	0.05	0.20	-0.05	-0.09
21	CORR	LABEL_INFO	-0.01	0.36	-0.10	-0.21	0.42	0.41	-0.05	0.14	-0.08	0.30	0.19	-0.13	-0.11	-0.23	0.09	0.18	0.00	0.17
22	CORR	PACKAGING	-0.06	0.38	0.32	0.18	0.33	0.15	0.06	-0.29	-0.22	-0.10	0.20	0.30	0.03	0.01	-0.22	0.00	0.20	-0.02
23	CORR	PROMOTION	-0.15	0.27	0.36	0.45	0.33	-0.29	0.09	-0.11	0.04	-0.09	-0.14	0.23	-0.18	-0.01	-0.14	0.07	-0.01	-0.17
24	CORR	PARTY	0.13	0.09	0.23	0.34	0.18	0.31	-0.17	0.11	0.34	0.27	-0.37	0.06	0.28	-0.04	-0.16	-0.09	0.30	0.16
25	CORR	GIFT	0.34	-0.10	0.13	0.44	-0.17	0.41	0.06	0.25	0.24	0.11	0.02	0.22	-0.03	-0.10	-0.02	-0.01	-0.23	0.11
26	CORR	HOME	0.39	-0.08	0.27	0.37	-0.29	0.09	0.05	0.09	-0.29	-0.13	0.04	0.04	0.28	-0.09	0.19	0.10	0.04	-0.40
27	CORR	TASTE	0.50	-0.15	0.16	0.21	-0.04	0.23	0.23	-0.17	0.28	-0.20	0.07	0.05	0.08	0.17	0.24	0.07	-0.32	0.11
28	CORR	ETNA_DOC	0.45	-0.42	0.06	0.02	0.35	-0.31	-0.14	0.29	-0.10	0.28	0.01	-0.10	0.04	-0.18	-0.02	0.16	-0.01	0.05
29	CORR	ETNABuying	0.60	0.07	-0.22	0.21	0.09	-0.11	0.02	-0.11	0.02	0.46	0.13	-0.22	0.04	-0.09	0.11	0.20	0.12	-0.01
30	CORR	ETNA_PREFERENCE	0.13	0.67	-0.31	0.19	-0.21	0.13	0.06	-0.16	0.15	0.17	0.06	-0.13	0.00	0.16	-0.13	0.16	0.02	-0.07
31	CORR	ETNA_FLAVOR	0.28	0.70	-0.33	0.12	0.00	-0.12	0.03	0.10	0.04	0.07	-0.03	-0.22	0.08	0.01	-0.08	0.09	-0.15	-0.14
32	CORR	SICILIAN_EXCELLENCES	0.28	0.43	-0.32	0.06	-0.18	-0.16	0.12	0.35	-0.22	-0.07	-0.03	0.30	-0.19	-0.18	-0.05	-0.16	-0.03	0.14
33	CORR	ETNA_EXPENSIVE	0.27	0.23	-0.19	-0.02	0.18	-0.29	0.25	0.40	-0.13	-0.30	-0.17	0.00	0.41	0.22	-0.18	0.09	-0.13	0.09
34	CORR	ETNA_QUALITY	0.53	0.56	-0.34	0.07	-0.09	-0.05	0.02	0.05	-0.06	-0.01	0.02	0.06	-0.05	-0.10	0.02	-0.06	0.07	0.11
35	CORR	ETNA_RECOMMENDATION	0.12	0.72	-0.33	0.14	-0.30	0.08	0.01	-0.19	0.01	-0.02	0.00	0.05	-0.08	0.00	0.10	-0.11	0.07	-0.02

Following we can see the explanation of the variables by each pcas:

Obs	Type	Name	Expl_1	Expl_2	Expl_3	Expl_4	Expl_5	Expl_6	Expl_7	Expl_8	Expl_9	Expl_10	Expl_11	Expl_12	Expl_13	Expl_14	Expl_15	Expl_16	Expl_17	Expl_18
1	CORR	WINE_PREFERENCE	0.45	0.45	0.64	0.64	0.66	0.66	0.76	0.78	0.80	0.80	0.83	0.83	0.84	0.85	0.85	0.85	0.85	0.86
2	CORR	BEER_PREFERENCE	0.11	0.12	0.19	0.19	0.31	0.31	0.48	0.49	0.49	0.56	0.57	0.61	0.62	0.68	0.68	0.69	0.72	0.83
3	CORR	SOFT_PREFERENCE	0.09	0.15	0.15	0.23	0.23	0.28	0.30	0.31	0.39	0.50	0.68	0.68	0.71	0.86	0.92	0.92	0.94	
4	CORR	COCKTAIL_PREFERENCE	0.00	0.11	0.19	0.35	0.35	0.35	0.35	0.44	0.55	0.61	0.61	0.62	0.81	0.81	0.84	0.85	0.86	
5	CORR	WHITE_WINE	0.11	0.13	0.47	0.51	0.53	0.57	0.57	0.59	0.59	0.63	0.64	0.64	0.65	0.68	0.68	0.69	0.72	0.75
6	CORR	ROSE_WINE	0.05	0.13	0.35	0.44	0.48	0.49	0.58	0.59	0.59	0.62	0.63	0.63	0.65	0.67	0.67	0.67	0.67	0.67
7	CORR	RED_WINE	0.34	0.35	0.52	0.52	0.53	0.53	0.65	0.66	0.66	0.73	0.73	0.73	0.77	0.77	0.77	0.77	0.77	0.78
8	CORR	SPARKLING_WINE	0.07	0.08	0.28	0.37	0.38	0.38	0.55	0.56	0.58	0.59	0.64	0.64	0.65	0.67	0.73	0.75	0.80	0.81
9	CORR	SWEET_WINE	0.05	0.11	0.24	0.37	0.37	0.41	0.44	0.58	0.61	0.61	0.62	0.65	0.66	0.66	0.67	0.77	0.78	0.87
10	CORR	WINE_TASTING	0.45	0.51	0.52	0.55	0.56	0.56	0.57	0.57	0.63	0.63	0.66	0.67	0.70	0.74	0.74	0.76	0.76	0.77
11	CORR	WINERY_VISIT	0.29	0.39	0.39	0.41	0.45	0.47	0.47	0.48	0.52	0.56	0.65	0.65	0.65	0.76	0.77	0.81	0.81	0.81
12	CORR	WINE_COURSE	0.37	0.40	0.41	0.42	0.48	0.49	0.59	0.59	0.59	0.60	0.60	0.62	0.62	0.64	0.72	0.72	0.81	0.81
13	CORR	SUPERMARKET	0.11	0.21	0.42	0.46	0.47	0.47	0.53	0.53	0.53	0.69	0.69	0.69	0.73	0.74	0.77	0.77	0.78	0.78
14	CORR	WINE_SHOP	0.29	0.33	0.38	0.54	0.55	0.57	0.57	0.58	0.59	0.59	0.63	0.65	0.66	0.68	0.68	0.74	0.77	0.79
15	CORR	ONLINE_SHOP	0.22	0.23	0.23	0.27	0.29	0.43	0.53	0.53	0.55	0.62	0.63	0.63	0.63	0.70	0.70	0.79	0.80	0.83
16	CORR	GRAPE_ORIGIN	0.21	0.22	0.27	0.39	0.40	0.41	0.47	0.47	0.62	0.63	0.67	0.71	0.76	0.78	0.79	0.81	0.82	
17	CORR	GRAPE_VARIETY	0.21	0.22	0.27	0.42	0.43	0.47	0.48	0.48	0.55	0.55	0.59	0.73	0.74	0.74	0.76	0.76	0.76	0.77
18	CORR	BUDGET_FRIENDLY	0.00	0.14	0.24	0.27	0.36	0.53	0.56	0.56	0.68	0.69	0.69	0.69	0.69	0.73	0.74	0.75	0.75	
19																				



4.4. Classical Test Theory

Classical Test Theory has been developed to quantify measurement error and to solve related problems such as correcting observed dependencies between variables. We firstly speaks about the concept of reliability, that is dependent on the group used to develop the test. If the group has a wide range of skill or abilities, then the **reliability** will be higher than if the group has a narrow range of skill or abilities. CTT (Classical Test Theory) indicates that

$$X = T + E$$

Where X is the observed score, T is the true score and E is the error. So, reliability is : $r = 1 - E$.

The type of reliability that we assess for a measure depends on the type of error that we seek to evaluate.

4.4.1. Pre-Analysis: Cronbach's Alpha – Correlation matrix

Internal consistency indicates the extent to which the responses on the items within a measure are consistent. Coefficient alpha (or **Cronbach's Alpha**) is the most widely used reliability measure of internal consistency. In our survey we want to make the analysis on question 17 and 18 (Likert scale) and then we applied it on all the numerical scale. Using on SAS the following code:

La procedura CORR								
5 Variabili: ETNA_FLAVOR SICILIAN_EXCELLENCES ETNA_EXPENSIVE ETNA_QUALITY ETNA_RECOMMENDATION								
Statistiche semplici								
Variabile	N	Media	Dev std	Somma	Minimo	Massimo	Etichetta	
ETNA_FLAVOR	247	2.95547	0.63281	730.00000	1.00000	4.00000	ETNA_FLAVOR	
SICILIAN_EXCELLENCES	247	3.02024	0.68342	746.00000	1.00000	4.00000	SICILIAN_EXCELLENCES	
ETNA_EXPENSIVE	247	2.08907	0.57041	516.00000	1.00000	4.00000	ETNA_EXPENSIVE	
ETNA_QUALITY	247	2.99190	0.65638	739.00000	1.00000	4.00000	ETNA_QUALITY	
ETNA_RECOMMENDATION	247	3.66802	0.65263	906.00000	1.00000	4.00000	ETNA_RECOMMENDATION	

Coefficiente alfa di Cronbach								
Variabili		Alfa						
Raw		0.782196						
Standardizzate		0.777880						

Coefficiente alfa di Cronbach con var eliminata								
Variabile eliminata	Variabili grezze		Variabili standardizzate		Etichetta			
	Correlazione con totale	Alfa	Correlazione con totale	Alfa				
ETNA_FLAVOR	0.663956	0.705688	0.666162	0.696979	ETNA_FLAVOR			
SICILIAN_EXCELLENCES	0.558961	0.741736	0.555428	0.735451	SICILIAN_EXCELLENCES			
ETNA_EXPENSIVE	0.312503	0.811217	0.313315	0.812173	ETNA_EXPENSIVE			
ETNA_QUALITY	0.701036	0.690797	0.695475	0.686425	ETNA_QUALITY			
ETNA_RECOMMENDATION	0.563094	0.739738	0.552501	0.736439	ETNA_RECOMMENDATION			

We are considering 5 variables that answering 2 questions : “How much do you agree with the following statements about Etna wine?” This question is based on a Likert scale from 1 to 4 (from completely disagree to totally agree), and “How likely are you recommend Etna wine to your family and friends?” that it is based on a scale fro 1 (not likely at all) to 4 (extremely likely).

What we can see from the image above is that the overall Cronbach's Alpha is around 0.78. If the average correlation is = 1, all items are maximally correlated, have no error components and therefore all measure the true latent ability. The close the Alfa is close to 1, the higher the reliability it is. Conventionally a value of alpha greater than 0.70 is considered acceptable, and this basically means that our value of 0.78 can be considered as a valid value.

Looking at the second table the value increases to 0.81 when item “ETNA_EXPENSIVE” is deleted. This indicates stronger internal consistency among the other items (etna_flavor, sicilian_excellences, etna_quality, etna_recommendation) than all five. As we said the value of 0.78 is higher with the desired threshold of 0.70 which suggests that there would still be sufficient internal consistency if ETNA_EXPENSIVE remained in the measure. What we can also see is that looking at items like

etna_flavour and etna_quality the value of Alpha drastically decreases if one of the 2 variables is deleted and this means that the items have strong internal consistency with the other items and so these items must stay in the questionnaire. We can also see that considering all the numerical items the value of the alpha still above the critical value of 0.70, so we can maintain all the items in our survey and looking at the second table below we can see that there are no particular item in which the value of alpha increase if deleted:

Coefficiente alfa di Cronbach	
Variabili	Alfa
Raw	0.729392
Standardizzate	0.742212

Variabile eliminata	Coefficiente alfa di Cronbach con var eliminata			Etichetta
	Variabili grezze	Variabili standardizzate	Alfa	
	Correlazione con totale	Alfa	Correlazione con totale	Alfa
WINE_PREFERENCE	0.548096	0.701913	0.505318	0.720939 WINE_PREFERENCE
BEER_PREFERENCE	0.280429	0.720278	0.263081	0.735144 BEER_PREFERENCE
SOFT_PREFERENCE	-0.046889	0.743751	-0.044397	0.752262 SOFT_PREFERENCE
COCKTAIL_PREFERENCE	0.260403	0.721772	0.244123	0.736229 COCKTAIL_PREFERENCE
WHITE_WINE	0.421282	0.710722	0.375118	0.728654 WHITE_WINE
ROSE_WINE	0.378318	0.713676	0.346878	0.730303 ROSE_WINE
RED_WINE	0.514945	0.703737	0.477939	0.722577 RED_WINE
SPARKLING_WINE	0.309104	0.718152	0.275146	0.734452 SPARKLING_WINE
SWEET_WINE	0.299051	0.718866	0.282432	0.734033 SWEET_WINE
WINE KNOWLEDGE	0.267968	0.721719	0.268319	0.734844 WINE KNOWLEDGE
BUYING_EXPERIENCE	0.310935	0.717949	0.291780	0.733495 BUYING_EXPERIENCE
WINE_BOTTLES	0.192552	0.726734	0.176545	0.740064 WINE_BOTTLES
SUPERMARKET	-0.003429	0.737966	-0.025347	0.751231 SUPERMARKET
WINE_SHOP	0.206298	0.724954	0.207336	0.738322 WINE_SHOP
ONLINE_SHOP	0.221786	0.724545	0.225760	0.737276 ONLINE_SHOP
GRAPE_ORIGIN	0.261688	0.721608	0.283828	0.733953 GRAPE_ORIGIN
GRAPE_VARIETY	0.205598	0.725025	0.215884	0.737837 GRAPE_VARIETY
BUDGET_FRIENDLY	0.184367	0.725931	0.181244	0.739799 BUDGET_FRIENDLY
BRAND_AWARENESS	0.119890	0.729614	0.134236	0.742440 BRAND_AWARENESS
VINTAGE	0.338929	0.716556	0.352266	0.729989 VINTAGE
LABEL_INFO	0.085059	0.732562	0.116309	0.743441 LABEL_INFO
PACKAGING	0.197252	0.725393	0.191796	0.739202 PACKAGING
PROMOTION	0.037426	0.735978	0.046622	0.747299 PROMOTION
BOTTLE_BUDGET	0.164877	0.727021	0.182672	0.739718 BOTTLE_BUDGET
BUYING_FREQUENCY	0.071328	0.732730	0.056073	0.746779 BUYING_FREQUENCY
ETNA_PREFERENCE	0.258182	0.722860	0.310086	0.732438 ETNA_PREFERENCE
ETNA_FLAVOR	0.358313	0.718355	0.430946	0.725369 ETNA_FLAVOR
SICILIAN_EXCELLENCES	0.255727	0.722562	0.298991	0.733079 SICILIAN_EXCELLENCES
ETNA_EXPENSIVE	0.202933	0.725304	0.235838	0.736702 ETNA_EXPENSIVE
ETNA_QUALITY	0.503398	0.711346	0.566676	0.717237 ETNA_QUALITY
ETNA_RECOMMENDATION	0.281608	0.721564	0.336872	0.730885 ETNA_RECOMMENDATION

Now a **correlation matrix** should be used in order to understand the relationship between the variables. The suggestion is to inspect correlation coefficients over 0.30. So this value can be considered as the minimal threshold even if +/- 0.30 is considered “minimal”, +/- 0.40 is considered “important” and +/-0.50 is practically significant. If no correlations go beyond 0.30, then we should reconsider if factor analysis is the appropriate method to utilise. How we have seen in the Pearson matrix above there are several items with a very high correlation, for example between “Wine_Preference” and “Red_Wine” the value of 0.736 means that most of respondents that have wine as a preference choose the Red one. Can be interesting having a look to the negative correlation between “Wine_Shop” and “Promotion” with a value of -0.36, people that goes to buy wine in a “Wine shop” are more interested in the quality of wine to respect to the price and this is confirmed by the negative correlation between “Wine_shop” and “budget_friendly” with a value of -0.26. About people that have tried the Etna Doc can be interesting the relation between “Etna_Quality” and “Etna_Recommendation” with an high value of 0.57 basically most of them will suggest Etna wine judging it as high quality wine.

4.5. Factor Analysis

Factor analysis is a way to take our amount of data and shrinking it to a smaller data set that is more manageable and of more clear interpretation from a theoretical point of view. This technique is generally used to investigate variable relationship that are complex by collapsing several variables into a small number of interpretable factors. A factor refers to a set of observed variables that have similar response patterns, these variables are associated with an hidden variable known as a latent variable.

Factor Analysis is performed on all the scale that are in an interval scale so, we are considering items from “wine_preference --sweet_preference”, “wine_knowledge -- buying_frequency” and “etna_preference – etna_recommendation”.

4.5.1. Kaiser's MSA and Factor Extraction Results

Partial Correlation Matrix

La procedura FACTOR Metodo fattoriale iniziale: Fattori principali															Correlazioni parziali che controllano tutte le			
	VINE_PREFERENCE	BEER_PREFERENCE	SOFT_PREFERENCE	COCKTAIL_PREFERENCE	VWHITE_VINE	ROSE_VINE	RED_WINE	SPARKLING_WINE	SWEET_VINE	VINE KNOWLEDGE	BUYING_EXPERIENCE	WINE_BOTTLES	SUPERMARKET	VINE_SHOP	ONLINE_SHO			
WINE_PREFERENCE	1.00000	0.19655	-0.08967	0.14042	0.28875	0.08875	0.61083	-0.09279	-0.06801	0.07847	0.27266	-0.09233	-0.20440	-0.10774	0.0346			
BEER_PREFERENCE	0.19655	1.00000	-0.11104	0.03751	0.00813	0.03146	0.07170	0.04666	-0.01258	0.12500	-0.00791	0.02902	-0.00268	-0.01683	-0.1096			
SOFT_PREFERENCE	-0.08967	-0.11104	1.00000	0.16542	-0.00970	0.10651	0.05857	-0.01704	0.08475	-0.09098	-0.02754	0.03551	-0.00268	0.06305	-0.0969			
COCKTAIL_PREFERENCE	0.14042	0.03751	0.16542	1.00000	0.12802	0.06503	-0.15430	0.02139	0.05914	-0.10972	0.09468	-0.07937	0.06376	0.00810	-0.0168			
WHITE_WINE	0.28875	0.00813	-0.00970	0.12802	1.00000	0.27389	-0.08516	0.21604	-0.01591	0.03184	-0.09164	0.04865	0.16546	0.08632	0.0361			
ROSE_WINE	0.08875	0.03146	0.06503	-0.00970	0.27389	1.00000	-0.04716	0.18317	0.20818	-0.07280	0.04263	-0.07951	0.04468	0.02015	0.0769			
RED_WINE	0.61083	0.07170	0.05857	0.12802	-0.08516	-0.04716	1.00000	0.08664	0.18109	-0.0616	0.01542	0.01899	0.15208	0.14961	-0.0460			
SPARKLING_WINE	-0.09279	0.04666	-0.01704	0.02139	0.21604	0.18317	0.08664	1.00000	0.26369	0.01041	-0.07709	0.10574	0.07660	0.05481	0.0115			
SWEET_WINE	-0.06801	-0.01258	0.08475	0.05914	-0.01591	0.20818	0.15109	0.26369	1.00000	0.10078	0.02419	-0.13259	0.04011	0.01434	0.0161			
VINE KNOWLEDGE	0.07847	0.12500	-0.09098	-0.10972	0.03184	-0.07280	-0.00616	0.01041	0.10078	1.00000	0.15761	-0.04258	-0.09083	0.03103	0.1908			
BUYING_EXPERIENCE	0.27266	-0.00791	-0.02754	0.09468	-0.09164	0.04263	0.01542	-0.07079	0.02419	0.19761	1.00000	0.67569	0.12987	-0.05776	0.0343			
WINE_BOTTLES	-0.09233	0.02902	0.03551	-0.07987	0.04865	-0.07951	0.01899	0.10574	-0.13289	-0.04258	0.67569	1.00000	0.02202	0.12233	0.0180			
SUPERMARKET	-0.20440	-0.00268	0.06376	0.16546	0.04468	0.18208	0.07660	0.04011	-0.09083	0.12367	0.02202	1.00000	-0.38219	-0.1478				
WINE_SHOP	-0.10774	-0.01683	0.06305	0.00810	0.08632	0.02015	0.24961	0.05481	0.01434	0.03103	-0.05776	0.13233	-0.38219	1.00000	0.0016			
ONLINE_SHOP	0.03469	-0.10963	-0.09697	-0.01683	0.03611	0.07690	-0.04603	0.01150	0.01614	0.19081	0.03431	0.01803	-0.14785	0.00167	1.0000			
GRAPE_ORIGIN	0.02629	-0.01562	-0.14578	-0.02106	0.04975	-0.05296	0.03023	-0.06011	0.11699	-0.02954	-0.03662	0.05587	-0.06683	0.09458	0.0518			
GRAPE_VARIETY	0.03883	-0.00502	0.03904	0.02631	-0.01193	0.01234	0.04011	0.10011	-0.09845	0.11103	-0.12371	0.07805	0.07719	0.14864	-0.0047			
BUDGET_FRIENDLY	0.06068	0.08875	0.03304	-0.04880	0.07230	-0.01909	-0.00328	-0.01780	0.02164	-0.09337	-0.04015	-0.02779	0.13856	-0.08133	-0.0269			
BRAND_AWARENESS	-0.10305	0.05004	0.05431	0.12503	0.05512	-0.02687	0.16121	-0.15336	0.10369	-0.05958	-0.05830	-0.07773	0.06005	-0.05218	-0.0469			
VINTAGE	0.16861	-0.08420	0.02447	0.08328	-0.06611	-0.01677	-0.03905	0.03541	0.07007	0.04211	0.06383	-0.07199	-0.07550	0.12002	0.0847			
LABEL_INFO	-0.09016	0.02327	-0.00232	-0.03180	-0.04750	-0.05368	0.11194	0.05913	0.01052	0.01046	-0.13206	-0.08411	0.10664	-0.05421	-0.0431			
PACKAGING	-0.03310	0.03071	-0.02338	0.12276	0.00683	0.04151	0.03752	0.08283	-0.10724	-0.08322	0.08165	-0.10224	0.12283	-0.01584	0.0249			
PROMOTION	0.04628	-0.17020	0.03834	-0.11234	0.04632	-0.04157	0.04223	0.05081	-0.06545	-0.00189	0.04050	-0.02228	-0.05571	-0.15280	0.1193			
BOTTLE_BUDGET	-0.01121	-0.01752	0.04800	0.07758	0.00216	-0.00327	-0.08027	-0.04974	-0.06577	-0.01347	-0.04254	0.05884	-0.10444	0.07220	0.0282			
BUYING_FREQUENCY	0.07656	0.03223	0.01781	0.08639	-0.07649	0.01590	0.03394	-0.05966	-0.03377	0.02491	-0.12043	0.15204	0.13611	0.07146	0.0928			
ETNA_PREFERENCE	-0.05706	0.03665	0.02985	-0.04755	-0.01564	0.03018	-0.07167	0.10582	-0.07799	0.07408	0.07527	-0.04903	-0.02594	0.05706	-0.1441			
ETNA_FLAVOR	0.09600	-0.08389	-0.12938	-0.00181	-0.04390	0.07660	-0.06529	-0.07941	0.05659	-0.00813	-0.00556	-0.01729	0.07401	0.06500	0.0173			
SICILIAN_EXCELLENCES	-0.01218	0.06065	0.07429	-0.07155	-0.06357	0.06939	0.08242	-0.04548	-0.03536	-0.17410	0.04753	-0.02953	-0.03069	-0.02525	-0.0684			
ETNA_EXPENSIVE	-0.16232	0.06584	0.08619	0.08435	-0.07202	-0.00086	0.10027	0.07892	-0.0496	0.13787	-0.01069	-0.03802	0.00204	0.00605	0.1243			
ETNA_QUALITY	0.06798	-0.03935	-0.02491	-0.02263	0.07111	-0.04391	-0.03002	0.03502	0.03998	0.15078	0.01572	0.06493	-0.01172	0.01376	0.0869			
ETNA_RECOMMENDATION	-0.05163	0.06567	0.09931	0.11211	-0.02018	-0.01102	0.05550	-0.08719	0.00200	-0.04264	-0.04907	0.06256	-0.01656	-0.09115	0.0681			

Le altre variabili																
HOP	GRAPE_ORIGIN	GRAPE_VARIETY	BUDGET_FRIENDLY	BRAND_AWARENESS	VINTAGE	LABEL_INFO	PACKAGING	PROMOTION	BOTTLE_BUDGET	BUYING_FREQUENCY	ETNA_PREFERENCE	ETNA_FLAVOR	SICILIAN_EXCELLENCES	ETNA_EXPENSIVE	ETNA_QUALITY	ETNA_RECOMMENDATION
3489	0.02629	0.03883	0.06068	-0.10305	0.16861	-0.09016	-0.03310	0.04628	-0.01121	0.07636	-0.05706	0.09600	-0.01218	-0.16232	0.06798	-0.05163
9863	-0.01162	-0.00502	0.08575	0.05004	-0.08420	0.02327	0.03071	-0.17020	-0.01752	0.03223	0.03685	-0.08359	0.06065	0.06584	-0.03935	0.06567
9697	-0.14573	0.03304	0.05431	0.02447	-0.00232	-0.02338	0.03834	0.04800	0.01781	0.02985	-0.12938	0.07429	0.08619	-0.02491	0.09931	
1683	-0.02106	0.02531	-0.04880	0.12503	0.08228	-0.03180	0.12276	-0.11234	0.07758	0.08639	-0.04755	-0.00181	-0.07155	0.08435	-0.02263	0.11211
6611	0.04975	-0.01193	0.07230	0.05512	-0.06611	-0.04750	0.00063	0.04632	0.00216	-0.07649	-0.01564	-0.04390	-0.06357	-0.07202	0.07111	-0.02018
7690	-0.05296	0.01234	-0.01909	-0.02687	-0.01677	-0.03638	0.04151	-0.04157	-0.00327	0.01590	0.03018	0.07660	0.06939	-0.00096	-0.04391	-0.01102
4603	0.00323	0.00411	-0.00328	0.16121	-0.03905	0.11194	0.03752	0.04223	-0.08027	0.03394	-0.07167	-0.06529	0.08242	0.10027	-0.03002	0.05550
1150	-0.06011	0.10211	-0.01780	-0.15336	0.03541	0.03913	0.08283	0.05081	-0.04974	-0.05966	0.10582	-0.07941	-0.04448	0.07892	0.03502	-0.08719
1614	0.11699	-0.09945	0.02164	0.10369	0.07007	0.01052	-0.10724	-0.06545	-0.06577	-0.03377	-0.07799	0.05659	-0.03336	-0.04996	0.03998	0.00200
9081	-0.02954	0.11103	-0.09337	-0.05988	0.04211	0.01046	-0.08322	-0.00189	-0.01347	0.02491	0.07408	-0.00813	-0.17410	0.13787	0.15078	-0.04264
3431	-0.03662	-0.12371	-0.04015	-0.05830	0.06383	-0.13206	0.08165	0.04050	-0.04254	0.12043	0.07527	-0.05556	0.04753	-0.01669	0.01872	-0.04807
1803	0.05587	0.07305	-0.00279	-0.07773	-0.07199	-0.05411	-0.10224	-0.02228	0.05884	0.15204	-0.04903	-0.01729	-0.02953	-0.03002	0.06493	0.06256
4785	-0.06683	0.07719	0.13656	0.06005	-0.07580	0.10664	0.12283	-0.05871	-0.11044	0.13611	-0.02594	0.07401	-0.03069	0.00204	-0.01172	-0.01656
0167	0.05458	0.14364	-0.08133	-0.05218	0.12002	-0.05421	-0.01584	-0.15280	0.07220	0.07146	-0.05706	0.06500	-0.02525	0.00605	0.01376	-0.09115
0000	0.05188	-0.00473	0.02691	-0.04697	0.08473	-0.04318	0.02493	0.11935	0.02820	0.09288	-0.14413	0.01734	-0.06844	0.12436	0.08691	0.06819
5188	1.00000	0.41469	0.19993	-0.00556	-0.01969	0.10707	0.10986	-0.20098	0.06728	-0.06935	0.02239	0.05953	0.04954	0.01463	0.00301	
0473	0.41489	1.00000	-0.04728	-0.01410	0.01697	-0.13668	-0.12998	0.05439	0.10082	-0.06167	-0.05845	-0.03473	0.06578	-0.00290	0.05600	0.00588
6691	0.19993	-0.04718	1.00000	0.12795	0.00035	-0.04918	-0.04483	0.34622	0.00561	0.05960	-0.04983	0.05654	-0.11780	0.07481	0.05313	0.00713
4697	-0.00556	-0.01110	0.12795	1.00000	0.07307	0.03714	0.05859	0.08223	0.20063	-0.11942	0.04889	0.15594	0.00332	-0.08535	-0.02356	-0.05269
5473	-0.01969	0.01597	0.00535	0.07307	1.00000	0.34844	0.15391	0.05853	0.15398	-0.07787	0.00351	0.02939	-0.04134	0.03415	0.03725	-0.06038
4318	0.10707	-0.13568	-0.04018	0.03714	0.34844	1.00000	0.07003	-0.02882	0.08316	-0.07731	-0.02599	0.02715	0.00132	-0.01097	0.10500	0.02644
2493	0.10986	-0.12988	-0.04483	0.05889	0.15391	0.07003	1.00000	0.31555	0.06887	0.07566	0.05384	-0.12133	-0.06009	0.00402	0.04288	0.10286
1935	-0.20088	0.05439	0.36222	0.08223	0.00583	-0.02852	0.31555	1.00000	-0.14201	0.07566	0.00294	0.07730	0.07917	0.01637	-0.06306	-0.05173
2820	0.06728	0.18362	0.00561	0.20063	0.18395	0.08316	0.06887	-0.14201	1.00000	0.14208	-0.03104	0.02790	-0.06819	-0.04668	0.12136	-0.04846
9288	-0.06095	-0.06167	0.05960	-0.11942	-0.07787	-0.07731	0.07566	0.07566	0.14208	1.00000	0.11936	-0.12126	-0.00845	-0.01255	0.03294	-0.04804
4413	0.02239	-0.05345	-0.00493	0.04889	0.00351	-0.02599	0.05384	0.02924	-0.03104	0.11936	1.00000	0.38656	-0.11682	-0.08622	0.13040	0.35554
1734	-0.02220	-0.03473	0.05554	0.15984	0.02939	0.02715	-0.12133	0.07730	0.02790	-0.12126	0.00000	0.07510	0.32348	0.16438	0.21241	
5844	0.05953	0.06578	-0.11790	0.00332	-0.04134	0.00132	-0.06009	0.07917	-0.06819	-0.00845	-0.11852	0.07510	1.00000	0.12500	0.39527	0.16176
2436	0.04954	-0.00290	0.07481	-0.08535	0.03415	-0.01087	0.00402	0.16137	-0.04668	-0.01255	-0.08622	0.32348	0.12500	1.00000	0.09136	-0.19171
6691	0.01463	0.05500	0.05313	-0.02356	0.03725	0.10500	0.04288	-0.06806	0.12136	0.03294	0.13040	0.16438	0.39527	0.09136	1.00000	0.24174
5819	0.00301	0.00585	0.00713	-0.05269	-0.06038	0.02644	0.10286	-0.05173	-0.04804	0.35554	0.21241	0.16176	-0.19171	0.24174	1.00000	

Kaiser's MSA

WINE_PREFERENCE	BEER_PREFERENCE	SOFT_PREFERENCE	COCKTAIL_PREFERENCE	WHITE_WINE	ROSE_WINE	RED_WINE	SPARKLING_WINE	SWEET_WINE	WINE KNOWLEDGE	BUYING_EXPERIENCE	WINE_BOTTLES
0.71939019	0.81829030	0.69552132	0.66769905	0.76982248	0.79248946	0.71577162	0.71231807	0.71983940	0.83445135	0.68403882	0.68063079
Misura di adeguatezza campionaria di Kaiser: MSA totale = 0.73825162											
SUPERMARKET	WINE_SHOP	ONLINE_SHOP	GRAPE_ORIGIN	GRAPE_VARIETY	BUDGET_FRIENDLY	BRAND_AWARENESS	VINTAGE	LABEL_INFO	PACKAGING	PROMOTION	BOTTLE_BUDGET
0.69501129	0.78565155	0.78400989	0.74106667	0.74061566	0.68016695	0.76052067	0.74804041	0.73110239	0.68524619	0.64049834	0.74497647
BUYING_FREQUENCY	ETNA_PREFERENCE	ETNA_FLAVOR	SICILIAN_EXCELLENCES	ETNA_EXPENSIVE	ETNA_QUALITY	ETNA_RECOMMENDATION					
0.18599216	0.55403633	0.62172618	0.43183863	0.29675281	0.62228277	0.57252891					

As we can see from the figure above , if the data are appropriate for the common factor model, the **partial correlations** should be smaller than the original correlations. And this is confirmed, for example the correlation between Wine_Preference and Beer_Preference goes from 0.43 to 0.19. **Kaiser's Measure of Sampling Adequacy (MSA)** is a summary, for each variable and for all variables together, of how much smaller the partial correlations are than the original correlations. Value of 0.8 or 0.9 are considered good, while MSAs below 0.5 are unacceptable. In our case we can see that variables like "Beer_Preference", "Wine_Knowledge", "Etna_Quality" for example have a value higher than 0.8. The overall MSA is 0.745 that can be considered acceptable.

4.5.2. Select the Number of Factors

Using the rule of eigenvalues greater than 1 we can take into account the first **4 factors** which together explain 82.92% of the total variance. Also choosing **5 factors** can be a good solution by the way it explains 90% of the total variance, while if we want to reach the 100% of total variance explained we should choose the first **7 factors**.

Autovalori della matrice di correlazione ridotta: Totale = 13.3367475 Media = 0.43021766

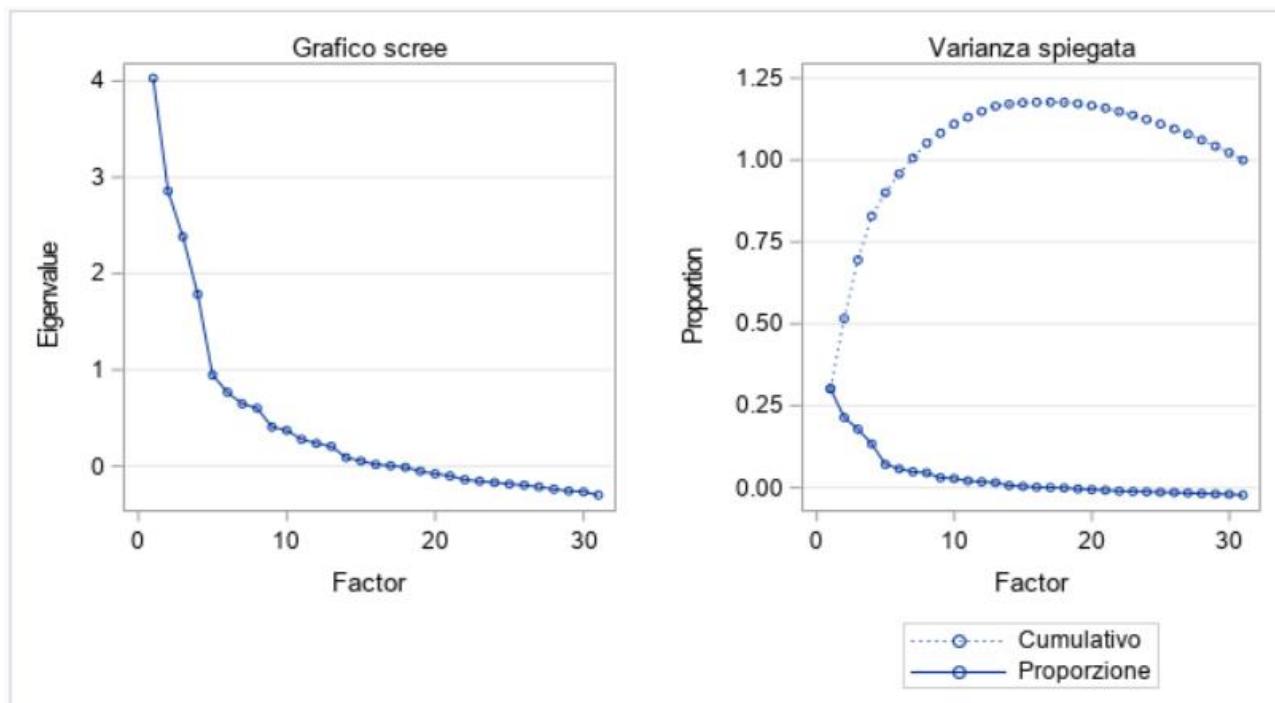
	Autovalore	Differenza	Proporzione	Cumulativa
1	4.03058198	1.17177982	0.3022	0.3022
2	2.85880215	0.47403863	0.2144	0.5166
3	2.38476352	0.59958495	0.1788	0.6954
4	1.78517857	0.83759358	0.1339	0.8292
5	0.94758500	0.17860976	0.0711	0.9003
6	0.76897524	0.12116639	0.0577	0.9579
7	0.64780884	0.04432727	0.0486	1.0065
8	0.60348157	0.19604285	0.0452	1.0518
9	0.40743872	0.03468061	0.0306	1.0823
10	0.37275811	0.09340999	0.0279	1.1103
11	0.27934813	0.04199258	0.0209	1.1312
12	0.23735555	0.03061423	0.0178	1.1490
13	0.20674132	0.11687203	0.0155	1.1645
14	0.08986929	0.03528653	0.0067	1.1713
15	0.05458276	0.03597784	0.0041	1.1753
16	0.01860492	0.01351400	0.0014	1.1767
17	0.00509092	0.01622364	0.0004	1.1771
18	-.01113272	0.04074996	-0.0008	1.1763
19	-.05188268	0.02685535	-0.0039	1.1724
20	-.07873803	0.02173928	-0.0059	1.1665
21	-.10047731	0.03881070	-0.0075	1.1590
22	-.13928801	0.01643242	-0.0104	1.1485
23	-.15572043	0.01339421	-0.0117	1.1368

Autovalori della matrice di correlazione ridotta: Totale = 13.3367475 Media = 0.43021766

	Autovalore	Differenza	Proporzione	Cumulativa
24	-0.16911464	0.01629416	-0.0127	1.1242
25	-0.18540881	0.01109302	-0.0139	1.1103
26	-0.19650183	0.01776614	-0.0147	1.0955
27	-0.21426797	0.02387668	-0.0161	1.0795
28	-0.23814465	0.01865506	-0.0179	1.0616
29	-0.25679971	0.00878516	-0.0193	1.0423
30	-0.26558487	0.03357252	-0.0199	1.0224
31	-0.29915739		-0.0224	1.0000

7 fattori saranno mantenuti dal criterio PROPORTION.

Looking at the Scree plot below we can see that clearest elbow is in correspondence of **5 factors**, it's plausible to hypothesize the existence of a latent structure, if the shape is almost rectilinear it means that the factors are only a transformation of the manifest variables. The relevant factors are those above the elbow. We can also see that looking on the plot of variance explained in correspondence of **4 factors**, we can have a good solution by the time it is explains more 82% of the total variance.



Initial Factor Pattern Matrix and Communalities

The elements in the loading matrix are called factor loadings. We can see the correlation between the variables and the factor, for example here we can see that “Wine_Preference” has the largest loading on the first factor, also “Red_wine”, and “Etna_Quality” have a large positive loading on the first factor. Looking at the second factor we can see a negative loading of the variables “Buying_Experience” and “Wine_Bottles” while variables like “Etna_Flavour”, “Etna_Reccommendation” have a large positive loadings on the second factors. Adding on SAS the functions **Fuzzy=0.3** and **Reorder** we didn’t consider the loadings lowers than 0.3 and we sorted all the variables in order to facilitate the reading of the loadings.

Pattern fattoriale								
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
WINE_PREFERENCE	WINE_PREFERENCE	0.70027	0	0.36706	0	0	0	0
ETNA_QUALITY	ETNA_QUALITY	0.63261	0.42690	0	0	0	0	0
RED_WINE	RED_WINE	0.60770	0	0.33771	0	0	0	-0.30877
WINE KNOWLEDGE	WINE KNOWLEDGE	0.51628	0	0	0	0	0	0
BUYING_EXPERIENCE	BUYING_EXPERIENCE	0.50126	-0.34633	0	-0.45384	0	0	0
WINE_SHOP	WINE_SHOP	0.47581	0	-0.37092	0	0	0	0
GRAPE_ORIGIN	GRAPE_ORIGIN	0.43309	0	-0.31369	0	0	0	0
GRAPE_VARIETY	GRAPE_VARIETY	0.40100	0	-0.33193	0	0	0	0
BEER_PREFERENCE	BEER_PREFERENCE	0.38247	0	0	0	0	0	0
ONLINE_SHOP	ONLINE_SHOP	0.36208	0	0	0	0	0	0.34693
SICILIAN_EXCELLENCE	SICILIAN_EXCELLENCE	0.36167	0.35165	0	0	0	0	0
ETNA_FLAVOR	ETNA_FLAVOR	0.41861	0.63041	0	0	0	0	0
ETNA_RECOMMENDATION	ETNA_RECOMMENDATION	0.32660	0.58079	0	-0.33518	0	0	0
ETNA_PREFERENCE	ETNA_PREFERENCE	0	0.55572	0	-0.31269	0	0	0
BRAND_AWARENESS	BRAND_AWARENESS	0	0.47925	0	0	0	0	0
LABEL_INFO	LABEL_INFO	0	0.41930	0	0.31966	0	0	0
SOFT_PREFERENCE	SOFT_PREFERENCE	0	0	0	0	0	0	0
SUPERMARKET	SUPERMARKET	0	0	0.53030	0	0	0	0
WHITE_WINE	WHITE_WINE	0.37509	0	0.46895	0	0	0	0
PROMOTION	PROMOTION	0	0	0.46517	0	0.34059	0	0
ROSE_WINE	ROSE_WINE	0.30007	0	0.39744	0	-0.33246	0	0
BUDGET_FRIENDLY	BUDGET_FRIENDLY	0	0	0.38652	0	0	0	0
PACKAGING	PACKAGING	0	0.31272	0.36269	0	0	0	0
SPARKLING_WINE	SPARKLING_WINE	0	0	0.32245	0	0	0	0
VINTAGE	VINTAGE	0.34362	0	0	0.38702	0.31980	0	0
SWEET_WINE	SWEET_WINE	0	0	0	0.38019	0	0	0
WINE_BOTTLES	WINE_BOTTLES	0.42310	-0.42277	0	-0.45114	0	0	0
COCKTAIL_PREFERENCE	COCKTAIL_PREFERENCE	0	0	0	0	0	0.34944	0
BOTTLE_BUDGET	BOTTLE_BUDGET	0	0	-0.32878	0.30021	0	0.34580	0
BUYING_FREQUENCY	BUYING_FREQUENCY	0	0	0	0	0	0	0
ETNA_EXPENSIVE	ETNA_EXPENSIVE	0	0	0	0	0	0	0

I valori minori di 0.3 non sono stampati.

Varianza spiegata da ogni fattore							
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	
	4.0305820	2.8588022	2.3847635	1.7851786	0.9475850	0.7689752	0.6478088

In the table above there is the amount of explained variance by each factor, and we can see that the first 2 factors explain more than half of the total amount of explained variance (the total has a value of 13.13).

Now we have a look to the communalities, so the percentage of variance explained, and so we are going to see what happens adding a new factor, if basically there is a consistent (higher 0.15%) improvement in any variables, looking at the table below:

- The first factor has too many factors, probably because 7 factors are too much but it is not satisfactory at all.
- The 2-factor solution increase in a substantial way the communalities of 8 variables but there still too many variables with low communalities.
- The 3-factor increase in a substantial way the communalities of other 8 variables.
- The 4-factor increase in a substantial way the communalities of other 10 variables.
- The 5-factor increase in a good way the communalities of other 2 variables.
- The 6-factor increase in a substantial way the communalities of other 2 variables.
- The 7-factor increase in a substantial way the communalities of 2 variables.

Oss	_NAME_	Factor 1	Factor2	Factor 3	Factor4	Factor5	Factor6	Factor7	EXPL_1	EXPL_2	EXPL_3	EXPL_4	EXPL_5	EXPL_6	EXPL_7		
1	WINE_PREFERENC E	0.73713	-0.28907	0.40599	0.01344	0.21313	-	0.03729	-0.30750	0.54336	0.62692	0.79175	0.79194	0.83736	0.83875	0.93331	
2	BEER_PREFERENC E	0.40260	-0.15380	0.19980	-	0.02950	-0.06309	-	0.04455	-0.37108	0.16209	0.18574	0.22566	0.22653	0.23051	0.23250	0.37020
3	SOFT_PREFERENCE	-0.1787	0.24411	0.07882	0.12408	-0.2270	0.29428	0.11212	0.03195	0.09154	0.09775	0.11314	0.16469	0.25129	0.26387		
4	COCKTAI_L_PREFER ENCE	0.14467	0.18836	0.26931	0.24531	-0.1227	0.46252	0.06123	0.02093	0.05641	0.12894	0.18912	0.20418	0.41811	0.42186		
5	WHITE_WINE	0.39483	-0.08319	0.51870	0.24542	-0.2320	0.08307	0.10810	0.15589	0.16281	0.43186	0.49209	0.54594	0.55284	0.56452		
6	ROSE_WINE	0.31586	0.08672	0.43961	0.26355	-0.4238	0.02579	0.18879	0.09977	0.10729	0.30054	0.37000	0.54967	0.55034	0.58598		
7	RED_WINE	0.63969	-0.18222	0.37354	0.11454	0.16792	-0.2010	-0.4183	0.40921	0.44241	0.58194	0.59506	0.62326	0.66368	0.83873		
8	SPARKLING_WINE	0.28675	-0.06450	0.35665	0.30556	-0.3649	-0.1152	0.29462	0.08223	0.08639	0.21359	0.30696	0.44012	0.45340	0.54021		
9	SWEET_WINE	0.27581	0.11480	0.26045	0.43049	-0.3047	-0.2160	0.00769	0.07607	0.08925	0.15709	0.34241	0.43525	0.48193	0.48199		
10	WINE_KNOWLEDGE	0.54346	-0.28126	-0.1487	-0.0555	0.10913	-0.2514	0.18221	0.29535	0.37445	0.39659	0.39968	0.41159	0.47481	0.50801		
11	BUYING_EXPERIENCE	0.52764	-0.37354	0.27854	-0.5138	0.09903	0.23775	0.04263	0.27841	0.41794	0.49553	0.75960	0.76941	0.82594	0.82775		
12	WINE_BOTTLES	0.44537	-0.45599	0.03517	-0.5108	-0.0076	0.35272	0.11670	0.19835	0.40628	0.40752	0.66846	0.66852	0.79292	0.80654		
13	SUPERMARKET	-0.2276	0.26337	0.58656	-0.1235	-0.1202	0.07457	-0.0621	0.05184	0.12120	0.46525	0.48053	0.49500	0.50056	0.50442		
14	WINE_SHOP	0.50086	-0.27283	-0.4102	0.27967	-0.0662	0.10283	0.05108	0.25086	0.32530	0.49362	0.57183	0.57622	0.58680	0.58941		

Oss	_NAME_	Factor 1	Factor2	Factor 3	Factor4	Factor5	Factor6	Factor7	EXPL_1	EXPL_2	EXPL_3	EXPL_4	EXPL_5	EXPL_6	EXPL_7
15	ONLINE_SHOP	0.38114	-0.15751	-0.0822	0.01329	0.23944	-0.1616	0.47010	0.14527	0.17008	0.17684	0.17702	0.23435	0.26047	0.48146
16	GRAPE_ORIGIN	0.45589	-0.00841	-0.3469	0.27603	0.06617	0.02899	-0.0692	0.20783	0.20790	0.32829	0.40448	0.40886	0.40970	0.41450
17	GRAPE_VARIETY	0.42211	-0.16803	-0.3671	0.23802	-0.0572	0.10812	0.04825	0.17818	0.20641	0.34120	0.39786	0.40113	0.41282	0.41515
18	BUDGET_FRIENDLY	0.01958	0.29244	0.42753	-0.0312	0.28337	-0.1191	0.04367	0.00038	0.08591	0.26868	0.26966	0.34996	0.36416	0.36606
19	BRAND_AWARNESS	-0.0429	0.51691	0.15190	0.27425	0.20979	0.11724	-0.2235	0.00185	0.26904	0.29211	0.36733	0.41134	0.42508	0.47504
20	VINTAGE	0.36171	0.21928	-0.0740	0.43822	0.40773	0.12037	0.08400	0.13083	0.17892	0.18441	0.37645	0.54269	0.55718	0.56424
21	LABEL_INFO	0.03420	0.45224	-0.0972	0.36195	0.24200	-0.0362	-0.1358	0.00117	0.20569	0.21515	0.34616	0.40472	0.40603	0.42448
22	PACKAGING	0.00196	0.33729	0.40117	0.05982	0.35812	0.23247	0.15855	0.00000	0.11377	0.27470	0.27828	0.40654	0.46058	0.48572
23	PROMOTION	-0.1144	0.23158	0.51452	-0.2013	0.43424	-0.1885	0.29549	0.01310	0.06673	0.33146	0.37199	0.56056	0.59610	0.68342
24	BOTTLE_BUDGET	0.23735	0.12063	-0.3636	0.33992	0.23977	0.45769	0.06129	0.05633	0.07089	0.20314	0.31868	0.37617	0.58566	0.58941
25	BUYING_FREQUENCY	0.06876	-0.15525	0.17336	-0.1886	0.13514	0.26837	0.14914	0.00473	0.02883	0.05888	0.09445	0.11272	0.18474	0.20698
26	ETNA_REFERENCE	0.30918	0.59938	-0.1092	-0.3540	-0.1657	0.11730	0.02400	0.09559	0.45483	0.46679	0.59215	0.61960	0.63336	0.63394
27	ETNA_FLAVOR	0.44064	0.67995	-0.1485	-0.2379	-0.0564	-0.1800	0.05431	0.19417	0.65650	0.67856	0.73519	0.73838	0.77080	0.77375
28	SICILIAN_EXCELLENCE	0.38071	0.37928	-0.1841	-0.2565	-0.1488	-0.1239	-0.1496	0.14494	0.28879	0.32270	0.38849	0.41064	0.42600	0.44838
29	ETNA_EXPENSIVE	0.27167	0.23177	-0.1442	-0.0008	0.02484	-0.3663	0.30089	0.07381	0.12752	0.14833	0.14895	0.28316	0.37369	
30	ETNA_QUALITY	0.66591	0.46044	-0.2230	-0.2230	-0.0409	-0.0228	0.05964	0.44344	0.65543	0.70522	0.75495	0.75663	0.75715	0.76071
31	ETNA_RESPONSE	0.34379	0.62643	-0.1212	-0.3795	-0.1849	0.1589	-0.1121	0.11819	0.51060	0.52530	0.66933	0.70352	0.72880	0.74138

What we can see from the table above is that considering 5, 6 or 7 – factors solution, of course increase the explanatory capacity but only on 6 couple of variables, while considering the 4-factor solution improve the communalities in a lot of variables. So what seems to suggest our analysis is a 4-factors solution.

4.5.3 Rotation of factor

Now we apply the rotation of factors using the Varimax method where the variance explained by the factors are more evenly distributed as compared with that of the unrotated solution. Before the varimax rotation the four factors explain the values in the table below of the common variance:

Varianza spiegata da ogni fattore			
Factor1	Factor2	Factor3	Factor4
4.0305820	2.8588022	2.3847635	1.7851786

After the varimax rotation the four rotated factors explain the common variance :

Varianza spiegata da ogni fattore			
Factor1	Factor2	Factor3	Factor4
3.0020947	2.7015394	2.6864987	2.6691934

4.5.4. Interpretation

Pattern fattoriale ruotato		Factor1	Factor2	Factor3	Factor4
ETNA_FLAVOR	ETNA_FLAVOR	79 *	0	0	0
ETNA_QUALITY	ETNA_QUALITY	76 *			0
ETNA_RECOMMENDATION	ETNA_RECOMMENDATION	75 *			0
ETNA_PREFERENCE	ETNA_PREFERENCE	70 *			0
SICILIAN_EXCELLENCES	SICILIAN_EXCELLENCES	57 *			0
ETNA_EXPENSIVE	ETNA_EXPENSIVE	32 *	0	0	0
WHITE_WINE	WHITE_WINE	0	62 *	0	0
WINE_PREFERENCE	WINE_PREFERENCE	0	60 *		
RED_WINE	RED_WINE	0	59 *		
ROSE_WINE	ROSE_WINE	0	55 *		
SWEET_WINE	SWEET_WINE	0	51 *		
SPARKLING_WINE	SPARKLING_WINE	0	50 *	0	0
VINTAGE	VINTAGE	0	37 *	31 *	0
COCKTAIL_PREFERENCE	COCKTAIL_PREFERENCE	0	36 *	0	0
WINE_SHOP	WINE_SHOP	0	0	67	0
GRAPE_VARIETY	GRAPE_VARIETY	0	0	56 *	0
GRAPE_ORIGIN	GRAPE_ORIGIN	0	0	53 *	
BOTTLE_BUDGET	BOTTLE_BUDGET	0	0	43 *	
WINE KNOWLEDGE	WINE KNOWLEDGE	0	0	40 *	
ONLINE_SHOP	ONLINE_SHOP	0	0	0	
PACKAGING	PACKAGING	0	0	-33 *	0
BUDGET_FRIENDLY	BUDGET_FRIENDLY	0	0	-36 *	0
PROMOTION	PROMOTION	0	0	-52 *	0
SUPERMARKET	SUPERMARKET	0	0	-60 *	0
BUYING_EXPERIENCE	BUYING_EXPERIENCE	0	0	0	77 *
WINE_BOTTLES	WINE_BOTTLES	0	0	0	74 *
BEER_PREFERENCE	BEER_PREFERENCE	0	0	0	31 *
BUYING_FREQUENCY	BUYING_FREQUENCY	0	0	0	0
SOFT_PREFERENCE	SOFT_PREFERENCE	0	0	0	0
BRAND_AWARENESS	BRAND_AWARENESS	0	0	0	-43 *
LABEL_INFO	LABEL_INFO	0	0	0	-45 *

I valori stampati sono moltiplicati per 100 e arrotondati all'intero più vicino. I valori maggiori di 0.3 sono contrassegnati da un **. I valori minori di 0.3 non sono stampati.

All items are loaded into factors, except to “Online_Shop” (How often do you buy wine in the following store?), “Buying_Frequency” (During the pandemic did the frequency with which you buy wine changed?) and “Soft_Preference” (How much do you like the following drinks?). Those 3 items will be removed in our analysis. The rest of the items are grouped as:

1. The first factor can be interpreted as the “Overall perception of Etna wine”, all the variables belong to it are inherent to the Etna wine and if the respondents feel to suggest this as Sicilian excellence. High score indicates a good opinion.
2. The second factor is an indicator of wine preference, in general almost all the items are referred to the wine preference based on the typology of wine, the only two items different are vintage which still referred to the wine choice so we can consider as an appropriate indicator, while the other item is Cocktail_preference, this is the only item not inherent to what this factor is going to consider, so we can drop it in factor analysis. We can consider the score as higher as the wine preference.
3. The third factor is based on the reasons that encourage people to buy a specific bottle of wine, like grape_origin, grape_variety, wine_knowledge, packaging, the habits that the consumers tend to have during the buying experience, considering also promotions, budget_friendly, bottle_budget and according to this also the location, supermarket, wine_shop. We can see that some items (especially Supermarket, Promotion) have a negative impact. High value indicates great attention by the consumer to the reference of wine.
4. The fourth factor Brand prestigious and Buying Experience is referred to the focus that people have when they make their choice. A part from Soft_Preference and Buying_frequency that are not loaded into factors, there is another item, Beer_Preference that has a very low contribute and moreover is not inherent to this interpretation, so we can do not consider in our factor. Brand_impact and label_info have a negative impact. High score indicates an high frequency of rate buying.

Now we will reconsider the analysis without the items not relevant, in particular dropping **Cocktail_Preference**, **Online_Shop**, **Buying_Frequency**, **Soft_Preference** and **Beer_Preference** and we are going to see if this confirms our previous result.

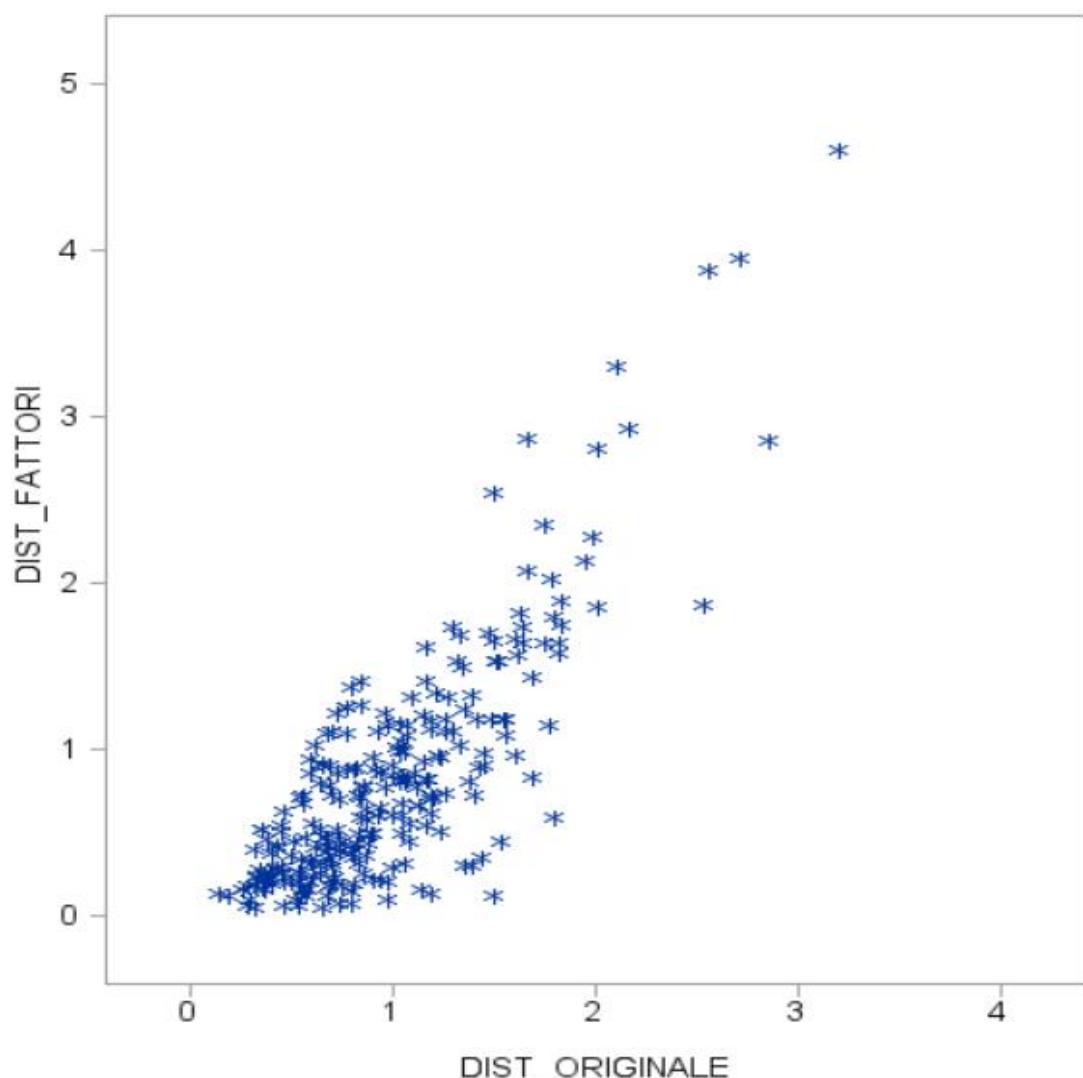
Pattern fattoriale ruotato					
		Factor1	Factor2	Factor3	Factor4
ETNA_FLAVOR	ETNA_FLAVOR	78 *	0	0	0
ETNA_QUALITY	ETNA_QUALITY	77 *	0	0	0
ETNA_RECOMMENDATION	ETNA_RECOMMENDATION	74 *	Overall perception of Etna Wine	0	0
ETNA_PREFERENCE	ETNA_PREFERENCE	70 *		0	0
SICILIAN_EXCELLENCES	SICILIAN_EXCELLENCES	57 *		0	0
ETNA_EXPENSIVE	ETNA_EXPENSIVE	31 *		0	0
WINE_PREFERENCE	WINE_PREFERENCE	0	69 *	0	40 *
RED_WINE	RED_WINE	0	67 *	0	0
WHITE_WINE	WHITE_WINE	0	62 *	Typology of wine and vintage	0
ROSE_WINE	ROSE_WINE	0	52 *		0
SPARKLING_WINE	SPARKLING_WINE	0	51 *		0
SWEET_WINE	SWEET_WINE	0	48 *		0
VINTAGE	VINTAGE	0	33 *	0	0
WINE_SHOP	WINE_SHOP	0	0	68 *	0
GRAPE_VARIETY	GRAPE_VARIETY	0	0	56 *	0
GRAPE_ORIGIN	GRAPE_ORIGIN	0	0	51 *	Wine reference and shop location
BOTTLE_BUDGET	BOTTLE_BUDGET	0	0	43 *	
WINE KNOWLEDGE	WINE KNOWLEDGE	0	0	37 *	
PACKAGING	PACKAGING	0	0	-35 *	
BUDGET_FRIENDLY	BUDGET_FRIENDLY	0	0	-38 *	0
PROMOTION	PROMOTION	0	0	-54 *	0
SUPERMARKET	SUPERMARKET	0	0	-60 *	Brand prestigious and buying experience
BUYING_EXPERIENCE	BUYING_EXPERIENCE	0	0	0	
WINE_BOTTLES	WINE_BOTTLES	0	0	0	
BRAND_AWARENESS	BRAND_AWARENESS	0	0	0	
LABEL_INFO	LABEL_INFO	0	0	0	-46 *
					-50 *

I valori stampati sono moltiplicati per 100 e arrotondati all'intero più vicino. I valori maggiori di 0.3 sono contrassegnati da un **. I valori minori di 0.3 non sono stampati.

4.5.5. Standardization and Plot

We now standardize the obtained factors and we can verify that the distance in the new space follow a linear relationship with the distance in the original space, meaning that the new space well represents the original space.

The SAS System



4.6. IRT- Item Response Theory

Item response theory (IRT) is a paradigm for the analysis, and scoring of tests, questionnaires, and similar instruments measuring abilities, attitudes, or other factors. IRT is a theory of testing based on the relationship between individuals' performances on a test item and the test takers' levels of performance on an overall measure of the ability that item was designed to measure. IRT is based on the idea that the probability of a correct/keyed response to an item is a mathematical function of person and item parameters. The first parameter affecting the item responses of an item is the respondent's level on the latent trait that item is trying to evaluate, other factors can be the difficulty level (how much of skills/satisfaction is needed to respond in a correct/positive way), the discrimination power (how good is the item in discriminating between individuals with different levels of the latent trait) and the easiness of guessing of a certain item.

Item Response Theory is based on 4 main assumptions:

1. Monotonicity: if the trait level increases, the probability of a correct/positive response increases as well
2. Unidimensionality: there is one dominant latent trait that is measured and this trait is the driving force for the responses observed for each item
3. Local Independence: the responses for an item don't affect the responses for other items

There are 3 main steps in applying an IRT analysis: model selection, estimation of the parameters and estimation of the abilities of the respondent.

Initially IRT will be applied to the whole dataset in order to have an overall look on the items' difficulty, then it will be applied onto the subgroups of items extrapolated from the original questionnaire.

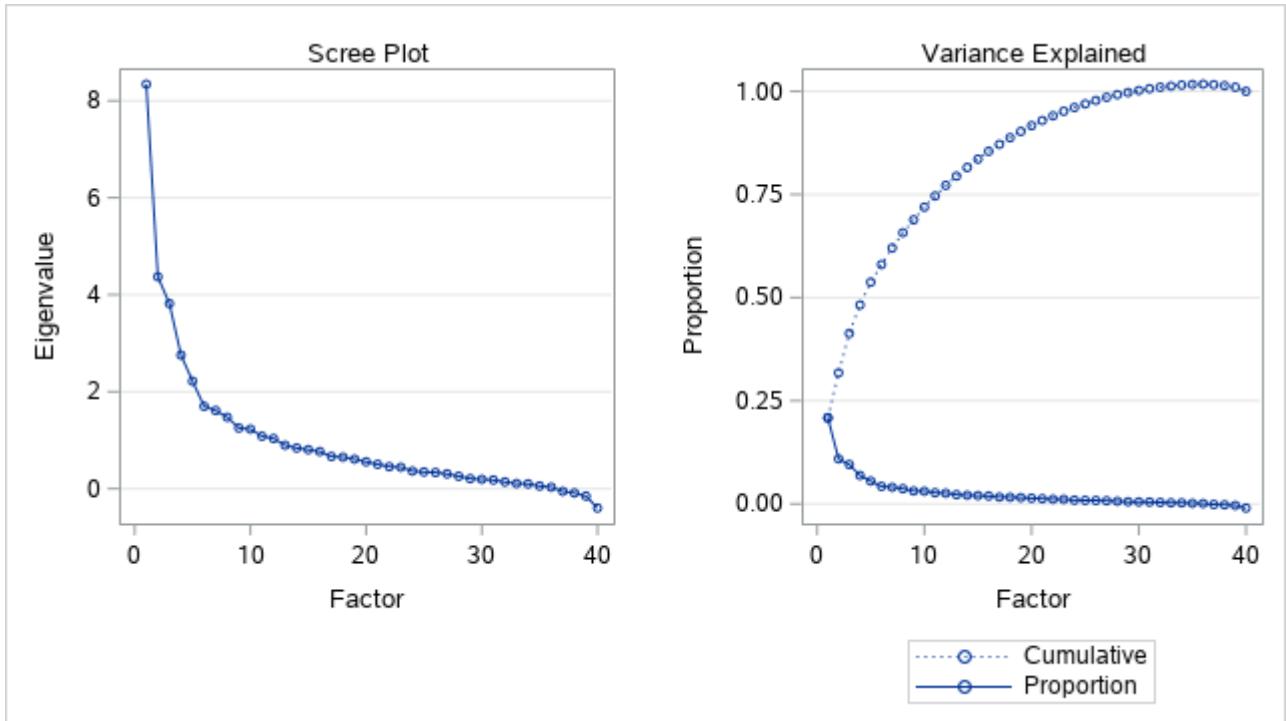
4.6.1 Dataset level IRT

In this first step, IRT will be applied to all the items together, except for the socio-demographic items (GENDER, AGE, EDUCATION, LOCATION, JOB) since for these items doesn't make any sense to try studying their difficulties (in this context).

In the following SAS procedure, all the default options will be used. The Marginal Maximum Likelihood estimation method will be used to estimate the parameters of the Graded Response Model (which, for binary items, became the two Parameter Logistic Model). This model expresses the probability of responding in a certain way as a function of the properties of both the item and the individual's latent trait.

Let's start by seeing if there is only one dominant latent trait that drives the responses observed for each factor. To do that we can use the following scree plot of the eigenvalues and the subsequent table:

Let's start by seeing if there is only one dominant latent trait that drives the responses observed for each factor. To do that we can use the following scree plot of the eigenvalues:



As we can see from the images above the first eigen value is much higher than the others and it explains almost the 25% of the total variability, hence the model hypothesis is satisfied and so a unidimensional model is reasonable in this example.

Now we can have a look at the parameter estimations, in doing this we will focus on the threshold parameter, which represent the difficulty of the items (the higher the threshold the higher the difficulty of that item) and also on the slope, which is the discrimination parameter (the higher the slope the better the item is in differentiating subjects), actually the value of the slope represents also the amount of information provided by a certain item, hence items with an higher slope are more informative than items with a lower slope.

To preserve the reader's eyes all the easier items will be highlighted in Green, the harder will be highlighted in Red, and the one with higher discrimination power will be highlighted in Yellow.

Item Parameter Estimates						
Response Model	Item	Label	Parameter	Estimate	Standard Error	Pr > t
Graded	WINE_PREFERENCE	WINE_PREFERENCE	Threshold 1	-1.58711	0.15619	<.0001
			Threshold 2	-0.86371	0.10685	<.0001
			Threshold 3	0.04665	0.09251	0.3071
			Slope	2.92696	0.45477	<.0001
	BEER_PREFERENCE	BEER_PREFERENCE	Threshold 1	-2.55658	0.54554	<.0001
			Threshold 2	-0.79869	0.25837	0.0010
			Threshold 3	1.11300	0.29999	0.0001
			Slope	0.67267	0.14175	<.0001
	SOFT_PREFERENCE	SOFT_PREFERENCE	Threshold 1	2.79867	0.66672	<.0001
			Threshold 2	0.20565	0.24637	0.2019
			Threshold 3	-1.96083	0.50162	<.0001
			Slope	-0.55822	0.13084	<.0001
	COCKTAIL_PREFERENCE	COCKTAIL_PREFERENCE	Threshold 1	66.23133	394.30221	0.4333

Item Parameter Estimates						
Response Model	Item	Label	Parameter	Estimate	Standard Error	Pr > t
			Threshold 2	4.42109	26.83875	0.4346
			Threshold 3	-52.68750	313.94766	0.4334
			Slope	-0.02034	0.12121	0.4334
	WHITE_WINE	WHITE_WINE	Threshold 1	-3.45556	0.76156	<.0001
			Threshold 2	-1.13634	0.31790	0.0002
			Threshold 3	1.55659	0.38234	<.0001
			Slope	0.64755	0.14281	<.0001
	ROSE_WINE	ROSE_WINE	Threshold 1	-3.59555	1.20305	0.0014
			Threshold 2	0.78198	0.41278	0.0291
			Threshold 3	4.75234	1.56039	0.0012
			Slope	0.39160	0.12837	0.0011
	RED_WINE	RED_WINE	Threshold 1	-1.77814	0.22123	<.0001
			Threshold 2	-0.83403	0.14041	<.0001
			Threshold 3	0.54504	0.13103	<.0001
			Slope	1.57909	0.22334	<.0001
	SPARKLING_WINE	SPARKLING_WINE	Threshold 1	-3.07513	0.94111	0.0005
			Threshold 2	-0.36996	0.32658	0.1286
			Threshold 3	3.38646	1.01978	0.0004
			Slope	0.42900	0.12692	0.0004
	SWEET_WINE	SWEET_WINE	Threshold 1	-2.64235	1.01945	0.0048
			Threshold 2	1.05576	0.51396	0.0200
			Threshold 3	4.46685	1.61839	0.0029
			Slope	0.34743	0.12542	0.0028
	WINE KNOWLEDGE	WINE KNOWLEDGE	Threshold 1	-0.65479	0.12000	<.0001
			Threshold 2	1.52932	0.17472	<.0001
			Threshold 3	2.68513	0.30856	<.0001
			Slope	1.81721	0.23808	<.0001
	BUYING_EXPERIENCE	BUYING_EXPERIENCE	Threshold 1	-1.20746	0.17734	<.0001
			Threshold 2	0.71608	0.14419	<.0001
			Threshold 3	1.64499	0.22109	<.0001
			Threshold 4	2.27163	0.29736	<.0001
			Slope	1.37768	0.18969	<.0001
	WINE_BOTTLES	WINE_BOTTLES	Threshold 1	-2.72146	0.40791	<.0001
			Threshold 2	1.04723	0.19672	<.0001
			Threshold 3	2.14768	0.31620	<.0001
			Threshold 4	2.82454	0.41473	<.0001
			Threshold 5	3.66520	0.57065	<.0001
			Slope	1.08667	0.17111	<.0001
	SUPERMARKET	SUPERMARKET	Threshold 1	6.63958	2.44807	0.0033
			Threshold 2	3.08258	1.19990	0.0051
			Threshold 3	-1.89407	0.73366	0.0049
			Slope	-0.36439	0.13426	0.0033
	WINE_SHOP	WINE_SHOP	Threshold 1	-1.37477	0.27794	<.0001
			Threshold 2	1.50708	0.30767	<.0001

Item Parameter Estimates						
Response Model	Item	Label	Parameter	Estimate	Standard Error	Pr > t
			Threshold 3	3.32173	0.58411	<.0001
			Slope	0.80352	0.14615	<.0001
	ONLINE_SHOP	ONLINE_SHOP	Threshold 1	1.68635	0.24295	<.0001
			Threshold 2	2.60619	0.38280	<.0001
			Threshold 3	3.86567	0.66413	<.0001
			Slope	1.40108	0.25992	<.0001
	GRAPE_ORIGIN	GRAPE_ORIGIN	Threshold 1	-3.65652	0.73269	<.0001
			Threshold 2	-1.73484	0.35590	<.0001
			Threshold 3	1.43634	0.32938	<.0001
			Slope	0.73516	0.14530	<.0001
	GRAPE_VARIETY	GRAPE_VARIETY	Threshold 1	-2.97751	0.58011	<.0001
			Threshold 2	-1.35298	0.28793	<.0001
			Threshold 3	1.78691	0.36825	<.0001
			Slope	0.77266	0.14932	<.0001
	BUDGET_FRIENDLY	BUDGET_FRIENDLY	Threshold 1	-7775	2291318	0.4986
			Threshold 2	-3500	1031304	0.4986
			Threshold 3	2033	599121	0.4986
			Slope	0.0004369	0.12874	0.4986
	BRAND_AWARENESS	BRAND_AWARENESS	Threshold 1	7.12640	2.77472	0.0051
			Threshold 2	3.24527	1.30659	0.0065
			Threshold 3	-4.58424	1.77050	0.0048
			Slope	-0.33995	0.13254	0.0052
	VINTAGE	VINTAGE	Threshold 1	-3.20272	0.82224	<.0001
			Threshold 2	1.64812	0.48911	0.0004
			Threshold 3	3.69335	0.94890	<.0001
			Slope	0.50468	0.12974	<.0001
	LABEL_INFO	LABEL_INFO	Threshold 1	12.05889	9.19212	0.0948
			Threshold 2	3.71140	2.97484	0.1061
			Threshold 3	-8.94454	6.74711	0.0925
			Slope	-0.16151	0.12277	0.0942
	PACKAGING	PACKAGING	Threshold 1	21.00334	35.54539	0.2773
			Threshold 2	-10.84054	18.23249	0.2761
			Threshold 3	-29.63437	49.97600	0.2766
			Slope	-0.07530	0.12697	0.2766
	PROMOTION	PROMOTION	Threshold 1	8.61957	6.88610	0.1053
			Threshold 2	-4.97124	3.93822	0.1034
			Threshold 3	-10.82063	8.55281	0.1029
			Slope	-0.15853	0.12530	0.1029
	BOTTLE_BUDGET	BOTTLE_BUDGET	Threshold 1	-7.05956	4.65571	0.0647
			Threshold 2	-5.09151	3.35895	0.0648
			Threshold 3	15.00197	9.93948	0.0656
			Threshold 4	22.30344	15.17737	0.0708
			Slope	0.24825	0.16472	0.0659
	BUYING_FREQUENCY	BUYING_FREQUENCY	Threshold 1	-10.03602	6.08351	0.0495

Item Parameter Estimates						
Response Model	Item	Label	Parameter	Estimate	Standard Error	Pr > t
			Threshold 2	-7.73974	4.67807	0.0490
			Threshold 3	5.29353	3.25710	0.0521
			Threshold 4	13.12933	8.00151	0.0504
			Slope	0.21574	0.13128	0.0501
	ETNA_BUYING	ETNA_BUYING	Threshold 1	-2.10463	0.29083	<.0001
			Threshold 2	0.29202	0.13656	0.0162
			Slope	1.33132	0.20426	<.0001
	ETNA_PREFERENCE	ETNA_PREFERENCE	Threshold 1	-196.78699	1869	0.4581
			Threshold 2	-72.95519	693.06905	0.4581
			Threshold 3	185.53602	1762	0.4581
			Slope	0.01558	0.14795	0.4581
	ETNA_FLAVOR	ETNA_FLAVOR	Threshold 1	-9.56804	4.18930	0.0112
			Threshold 2	-4.79657	2.05191	0.0097
			Threshold 3	4.96208	2.15970	0.0108
			Slope	0.36057	0.15729	0.0109
	SICILIAN_EXCELLENCES	SICILIAN_EXCELLENCES	Threshold 1	-6.76739	2.33273	0.0019
			Threshold 2	-4.77588	1.62078	0.0016
			Threshold 3	3.39077	1.16947	0.0019
			Slope	0.43871	0.15152	0.0019
	ETNA_EXPENSIVE	ETNA_EXPENSIVE	Threshold 1	-4.82441	1.51965	0.0007
			Threshold 2	3.57086	1.12891	0.0008
			Threshold 3	7.40761	2.37546	0.0009
			Slope	0.49564	0.16056	0.0010
	ETNA_QUALITY	ETNA_QUALITY	Threshold 1	-3.82652	0.69193	<.0001
			Threshold 2	-1.93375	0.32022	<.0001
			Threshold 3	1.72118	0.29324	<.0001
			Slope	1.02244	0.17847	<.0001
	ETNA_RECOMMENDATION	ETNA_RECOMMENDATION	Threshold 1	93.07903	353.79424	0.3962
			Threshold 2	65.72963	249.68120	0.3962
			Threshold 3	26.25594	99.66843	0.3961
			Slope	-0.04171	0.15854	0.3962
TwoP	WINE_TASTING	WINE_TASTING	Difficulty	0.19067	0.10737	0.0379
			Slope	2.04718	0.33126	<.0001
	WINERY_VISIT	WINERY_VISIT	Difficulty	-0.44763	0.13496	0.0005
			Slope	1.33973	0.24455	<.0001
	WINE.Course	WINE.Course	Difficulty	1.21475	0.15641	<.0001
			Slope	2.07097	0.37447	<.0001
	PARTY		Difficulty	1.26253	0.73306	0.0425
			Slope	0.29752	0.14249	0.0184
	GIFT		Difficulty	1.61551	0.38717	<.0001
			Slope	0.73692	0.17702	<.0001
	HOME		Difficulty	-0.44186	0.14682	0.0013
			Slope	1.17556	0.21447	<.0001
	TASTE		Difficulty	1.10422	0.18202	<.0001

Item Parameter Estimates						
Response Model	Item	Label	Parameter	Estimate	Standard Error	Pr > t
			Slope	1.41189	0.24603	<.0001
	ETNA_DOC	ETNA_DOC	Difficulty	-0.27505	0.14945	0.0329
			Slope	1.07433	0.20218	<.0001

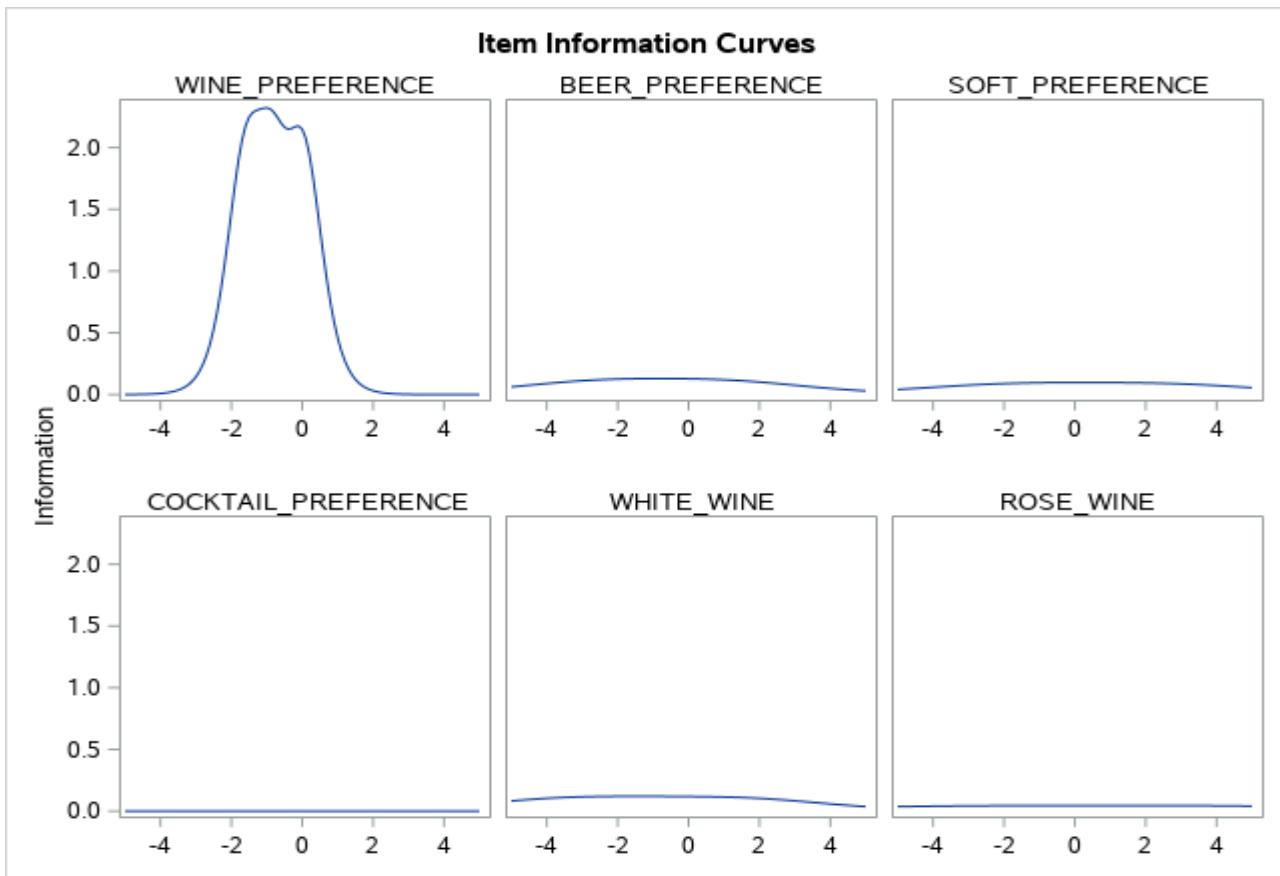
According to the table above we can identify the easiest and the hardest items. Let's start with the easiest ones:

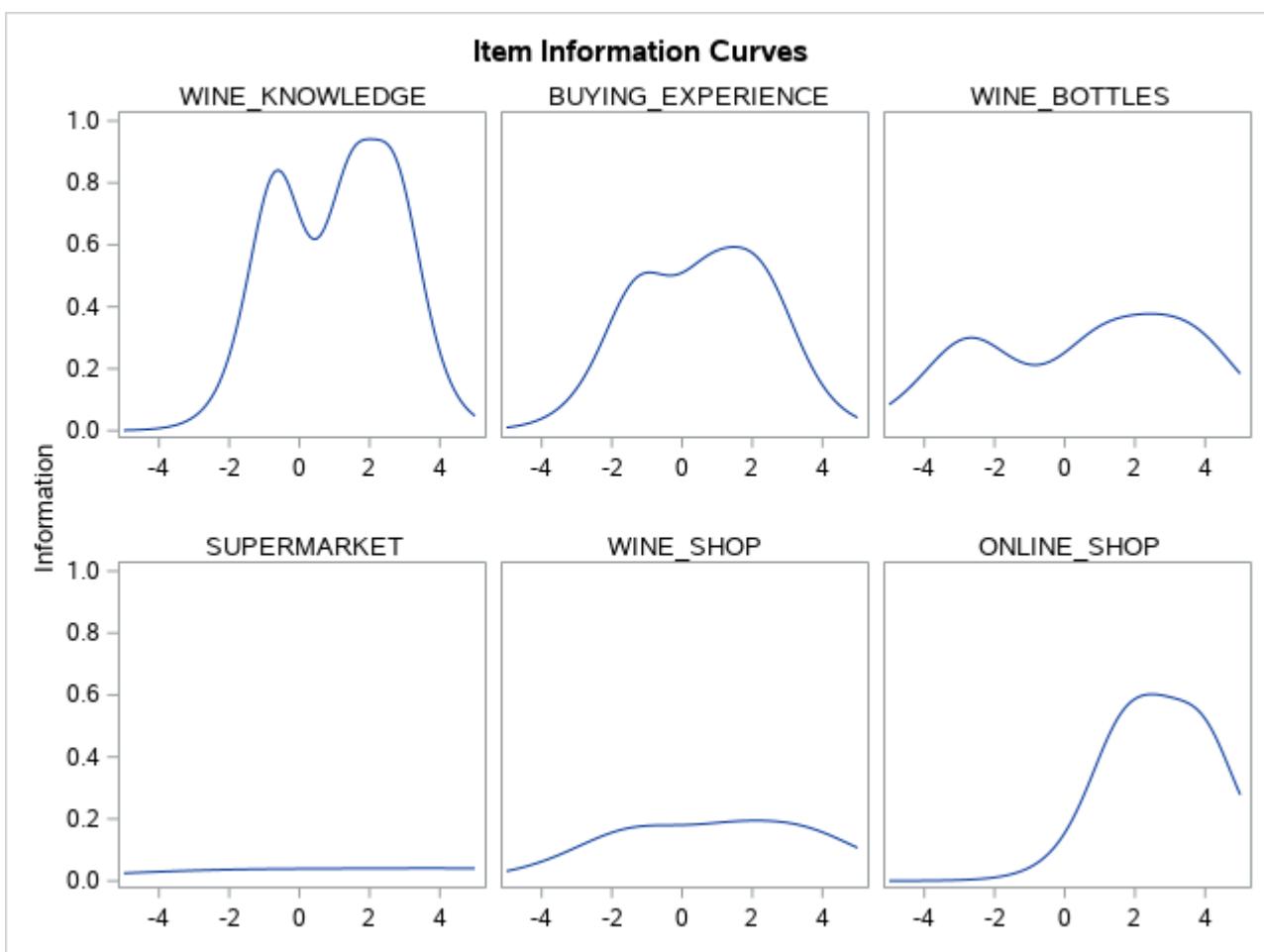
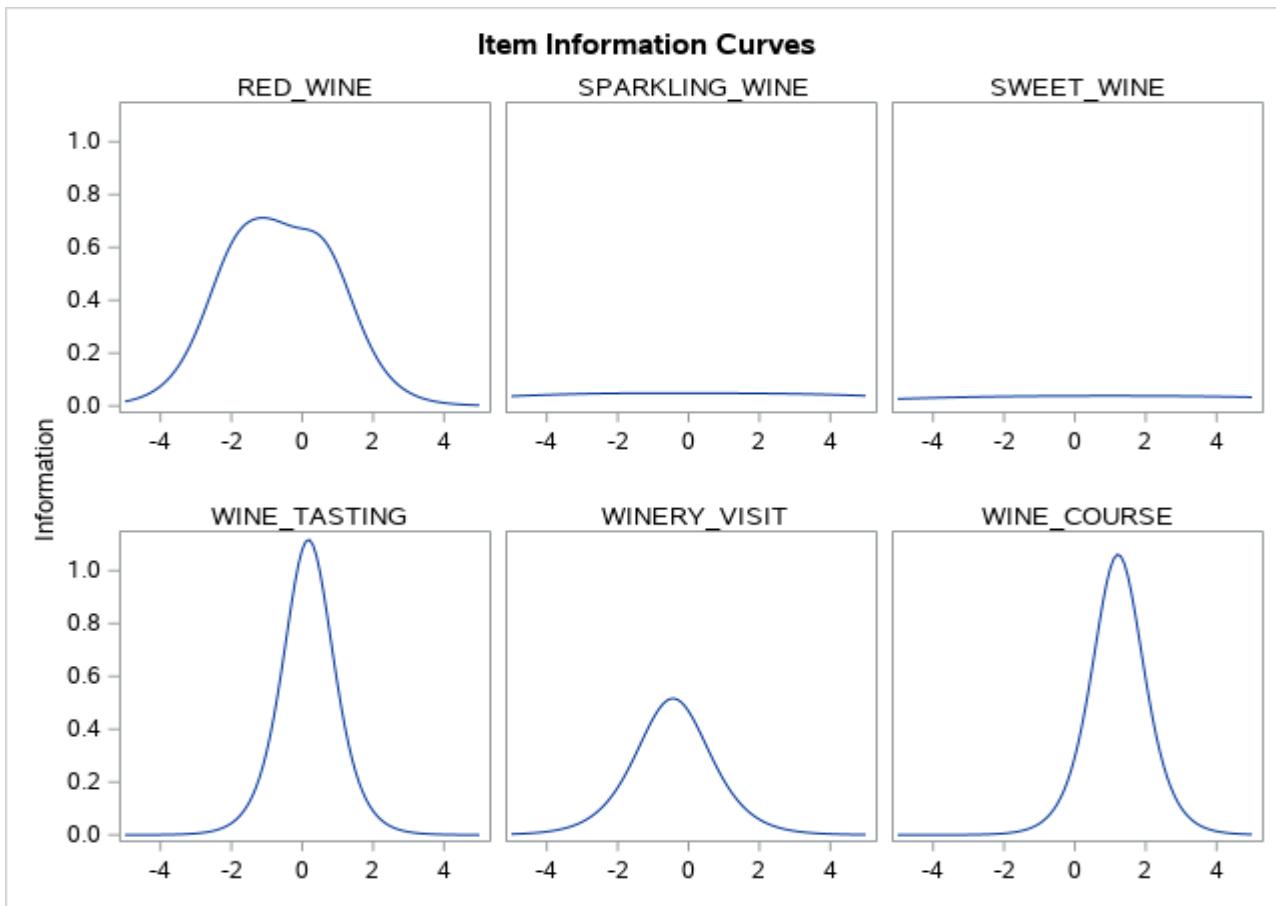
- Budget_Friendly
- Packaging
- Promotion
- Etna_Preference

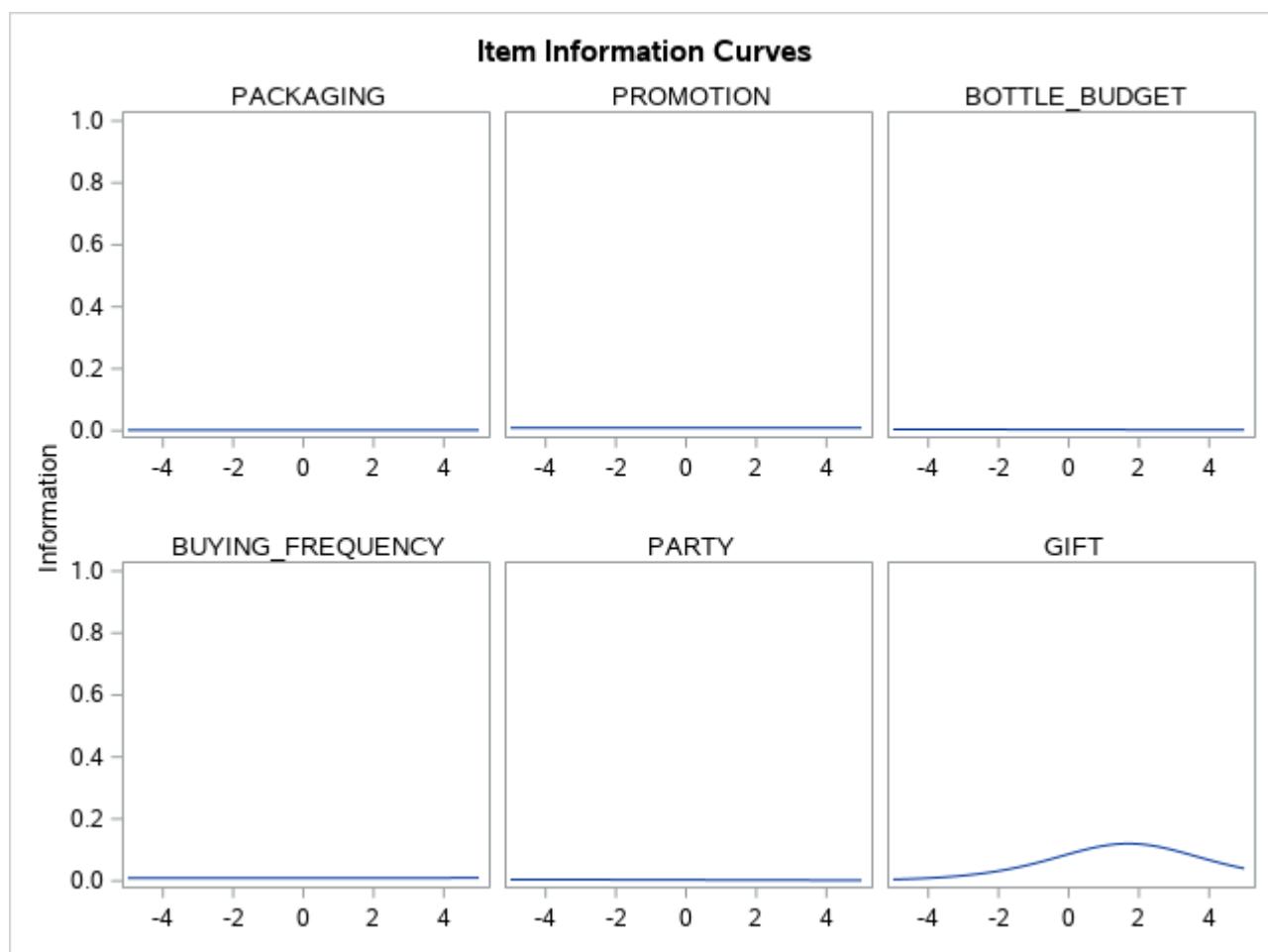
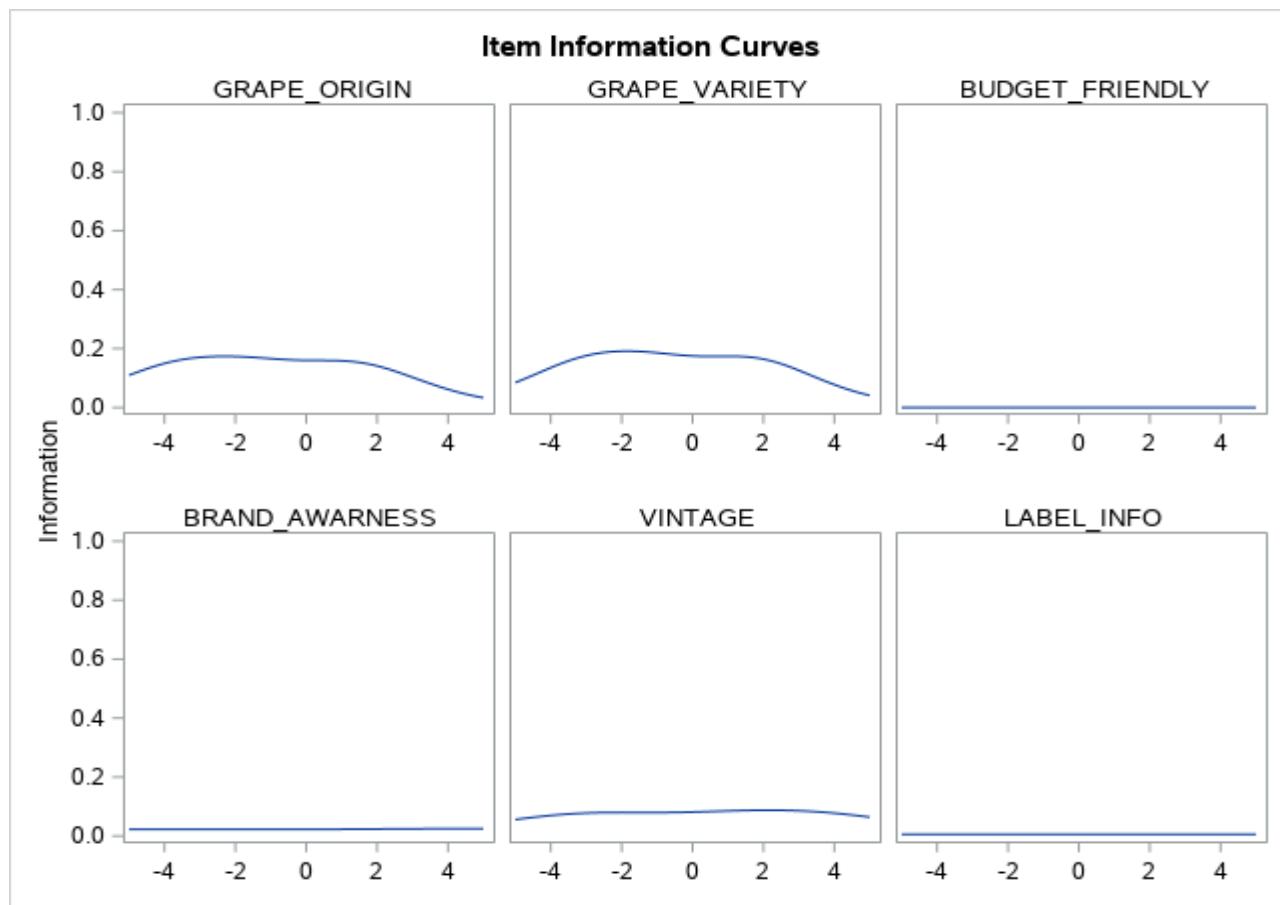
Now let's see the hardest ones:

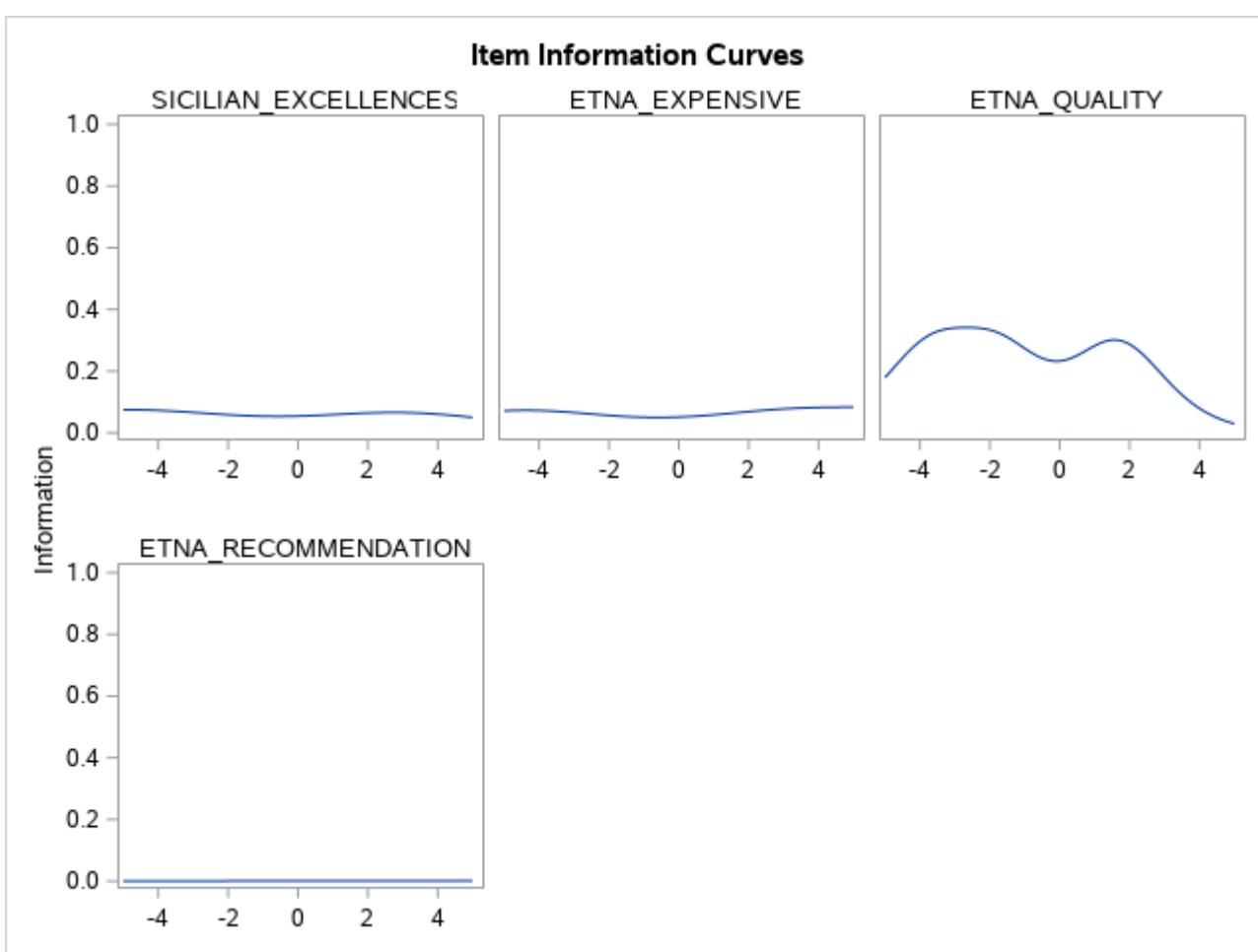
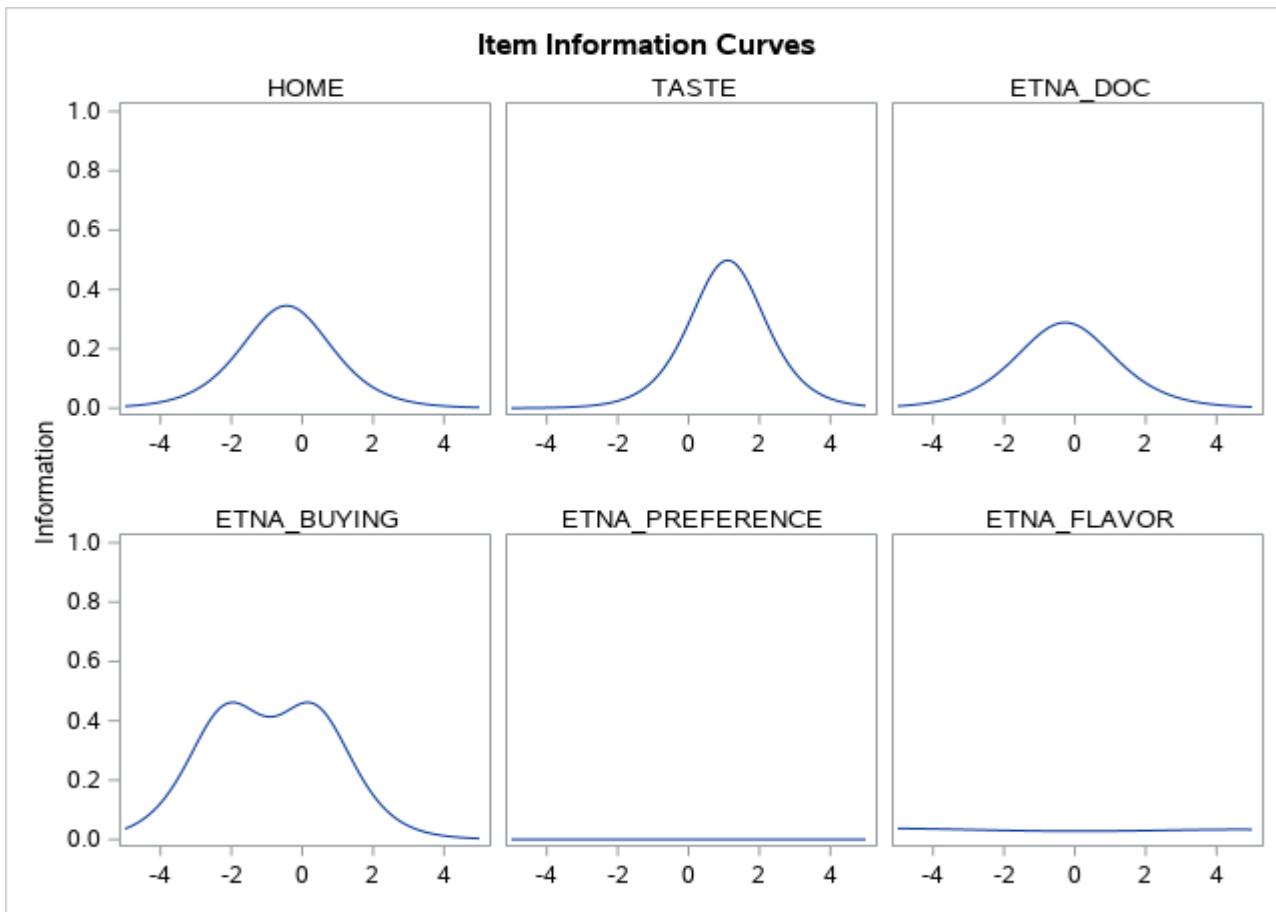
- Cocktail_Preference
- Label_Info
- Bottle_Budget
- Buying_Frequency
- Etna_Recommendation

Now let's analize also the slope parameter, which represents the discrimination parameter. Recalling that the slope parameter identifies the amount of information described by each item we can analize it looking at the following information curves:





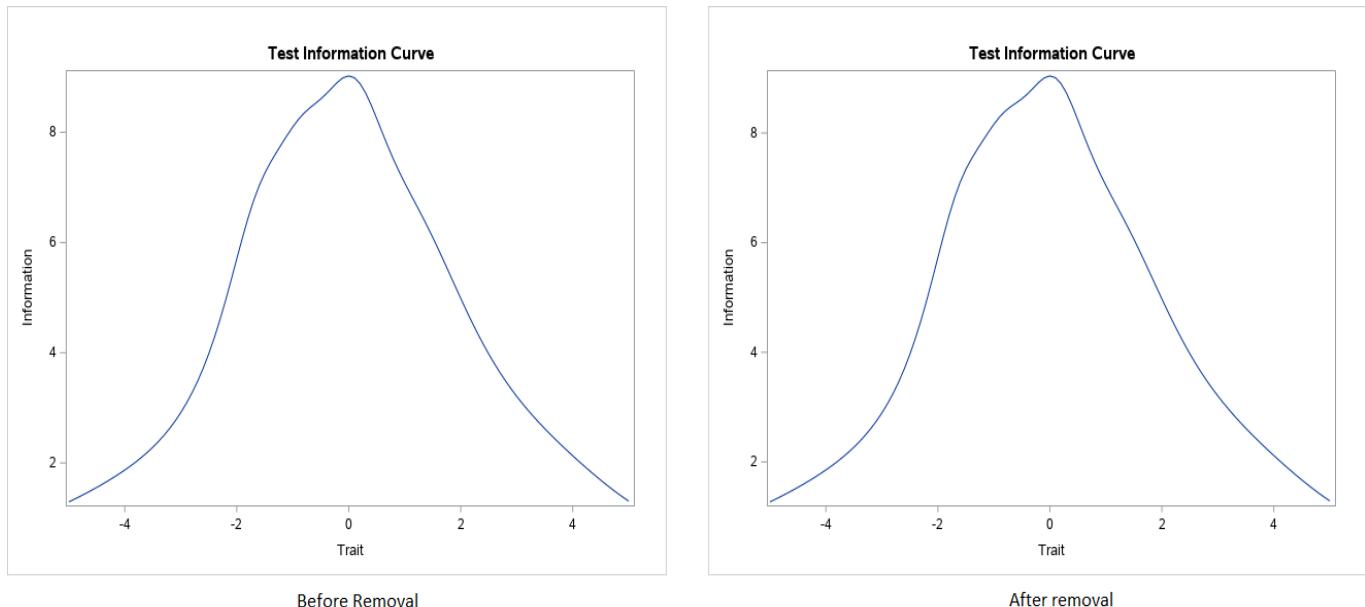




The first thing that we can notice from these plots is that very few items are informative for subjects with a really high or really low latent trait ability.

Another thing that can be seen by these plots is that all the items identified as easy and hard provide little to no information at all. We can try to remove all these items because they are not general enough and do not provide any sort of information.

As we can see in the images below the item test curve do not change at all after the removal of these items.



It comes straightforward to notice that the test is not informative for those subjects whose latent trait level is very distant from the difficulty level of the items

4.6.2 IRT on topics

Now that we have seen the properties of the various items at a global level, we can focus on each subgroup of items, identified by each topic of our questionnaire.

In the questionnaire there are 3 main topics:

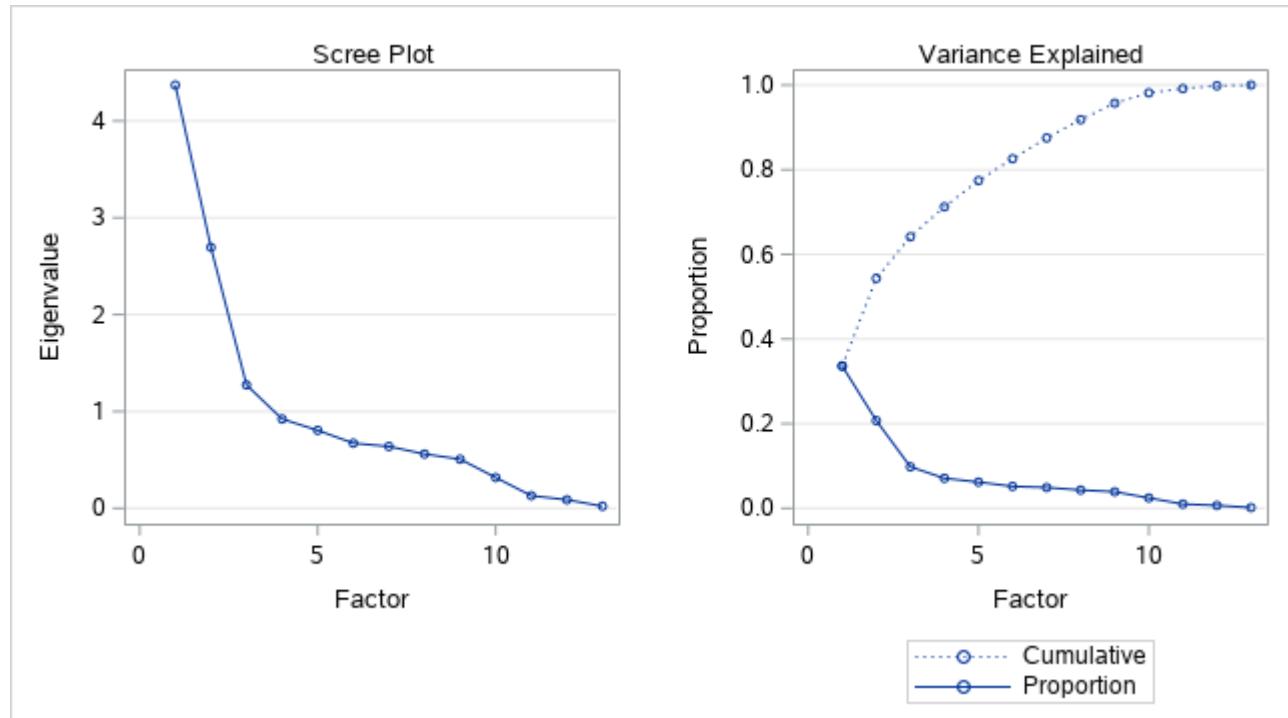
1. Wine preferences, expertise and experience
2. Wine buying experience
3. Etna DOC Wine

and 1 socio-demographic topic that we will ignore.

For each topic the same approach of the global level IRT will be used.

4.6.2.1 Wine preferences, expertise and experience

From the following scree plot we can identify that there is only one latent trait guiding the responses



Let's have a look at the estimated parameters

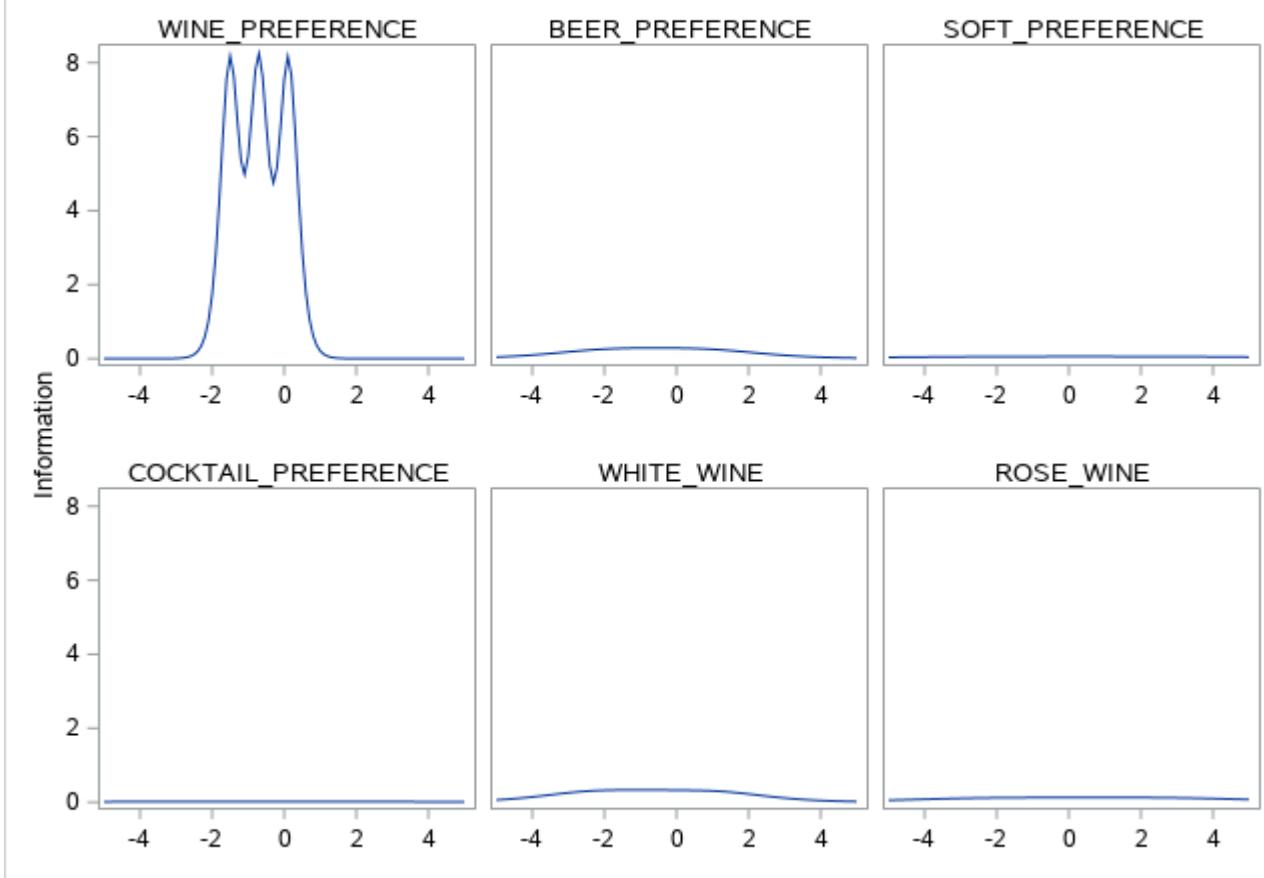
Item Parameter Estimates						
Response Model	Item	Label	Parameter	Estimate	Standard Error	Pr > t
Graded	WINE_PREFERENCE	WINE_PREFERENCE	Threshold 1	-1.50560	0.13448	<.0001
			Threshold 2	-0.70713	0.09534	<.0001
			Threshold 3	0.11362	0.08044	0.0789
	BEER_PREFERENCE	BEER_PREFERENCE	Slope	5.69107	1.66047	0.0003
			Threshold 1	-1.91935	0.32388	<.0001
			Threshold 2	-0.60688	0.17825	0.0003
			Threshold 3	0.82403	0.19782	<.0001
	SOFT_PREFERENCE	SOFT_PREFERENCE	Slope	0.96219	0.15966	<.0001
			Threshold 1	3.50717	1.06623	0.0005
			Threshold 2	0.24857	0.31266	0.2133
	COCKTAIL_PREFERENCE	COCKTAIL_PREFERENCE	Threshold 3	-2.47245	0.78526	0.0008
			Slope	-0.43257	0.13042	0.0005
			Threshold 1	-6.87997	4.51098	0.0636
	WHITE_WINE	WHITE_WINE	Threshold 2	-0.42631	0.71702	0.2761
			Threshold 3	5.51072	3.61579	0.0637
			Slope	0.19615	0.12740	0.0618
			Threshold 1	-2.35695	0.35912	<.0001
			Threshold 2	-0.76947	0.17923	<.0001
			Threshold 3	1.06039	0.20534	<.0001
			Slope	1.05185	0.16464	<.0001

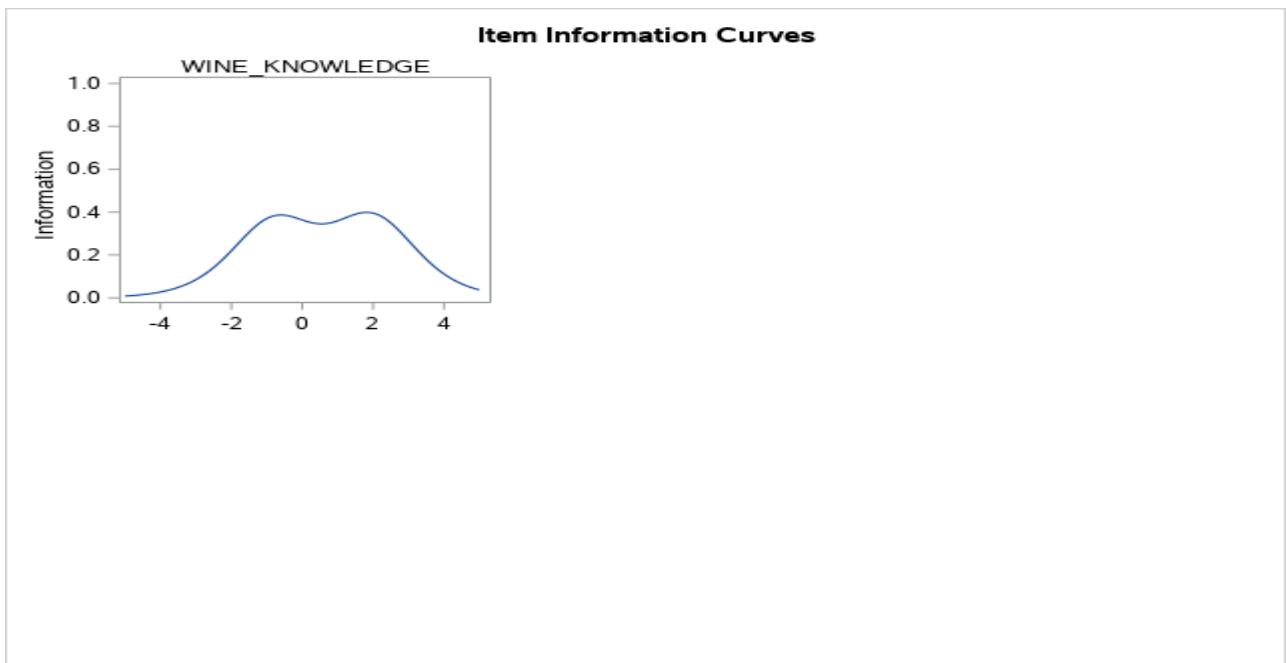
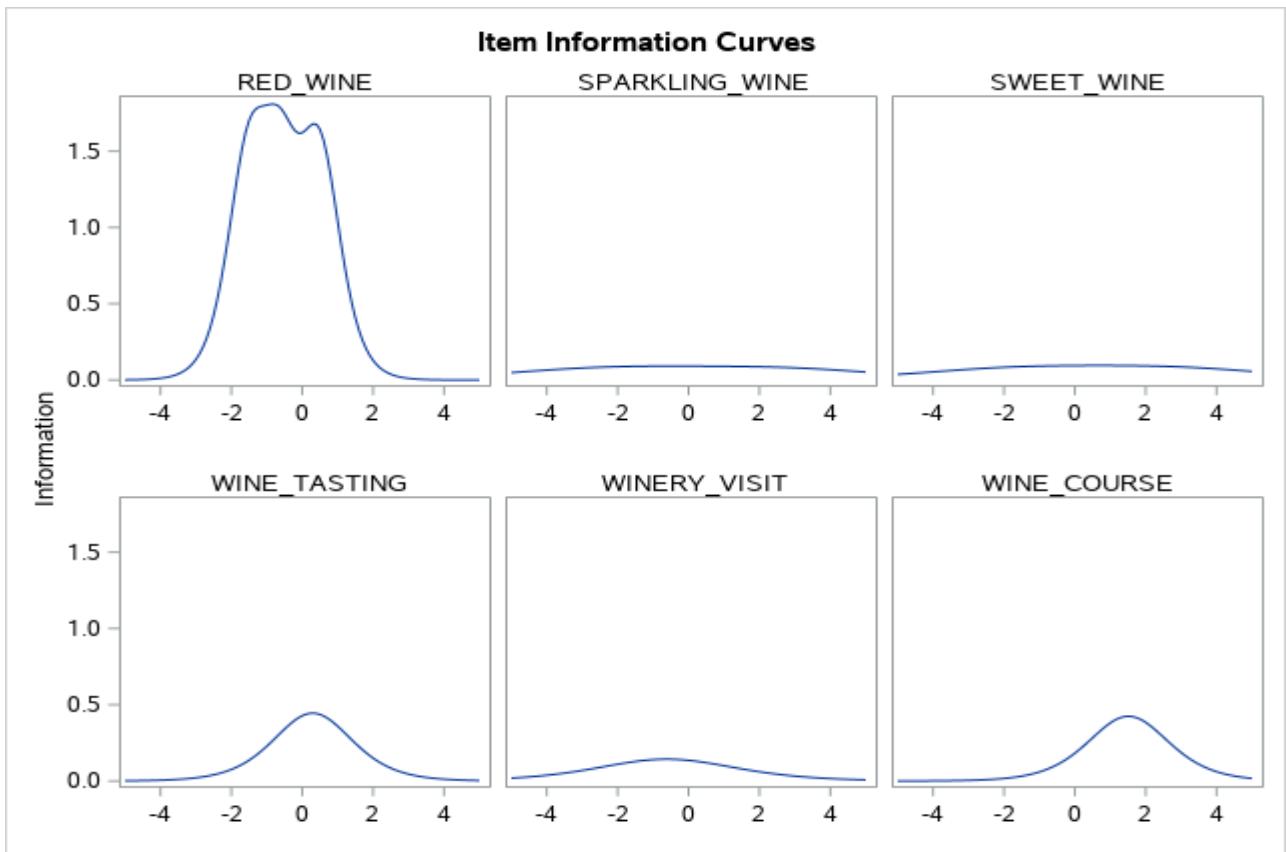
Item Parameter Estimates						
Response Model	Item	Label	Parameter	Estimate	Standard Error	Pr > t
	ROSE_WINE	ROSE_WINE	Threshold 1	-2.29161	0.50341	<.0001
			Threshold 2	0.50536	0.23352	0.0152
			Threshold 3	3.00009	0.64319	<.0001
			Slope	0.64874	0.13767	<.0001
	RED_WINE	RED_WINE	Threshold 1	-1.48574	0.15549	<.0001
			Threshold 2	-0.64289	0.10826	<.0001
			Threshold 3	0.46628	0.10121	<.0001
			Slope	2.51156	0.32044	<.0001
	SPARKLING_WINE	SPARKLING_WINE	Threshold 1	-2.48933	0.63269	<.0001
			Threshold 2	-0.30987	0.25859	0.1154
			Threshold 3	2.71771	0.68172	<.0001
			Slope	0.54458	0.13307	<.0001
	SWEET_WINE	SWEET_WINE	Threshold 1	-1.71241	0.47612	0.0002
			Threshold 2	0.70333	0.28704	0.0071
			Threshold 3	2.90880	0.72076	<.0001
			Slope	0.55342	0.13584	<.0001
	WINE KNOWLEDGE	WINE KNOWLEDGE	Threshold 1	-0.75610	0.16687	<.0001
			Threshold 2	1.87656	0.27827	<.0001
			Threshold 3	2.01532	0.29694	<.0001
			Slope	1.21819	0.19507	<.0001
TwoP	WINE_TASTING	WINE_TASTING	Difficulty	0.29448	0.13185	0.0128
			Slope	1.33426	0.24456	<.0001
	WINERY_VISIT	WINERY_VISIT	Difficulty	-0.61752	0.22668	0.0032
			Slope	0.75703	0.17666	<.0001
	WINE.Course	WINE.Course	Difficulty	1.51179	0.26747	<.0001
			Slope	1.30343	0.29295	<.0001

This section doesn't seem to have neither too easy items nor to difficult ones, the only item that have more "extreme" threshold values is Cocktail_Preference, but also here the values are relatively balanced.

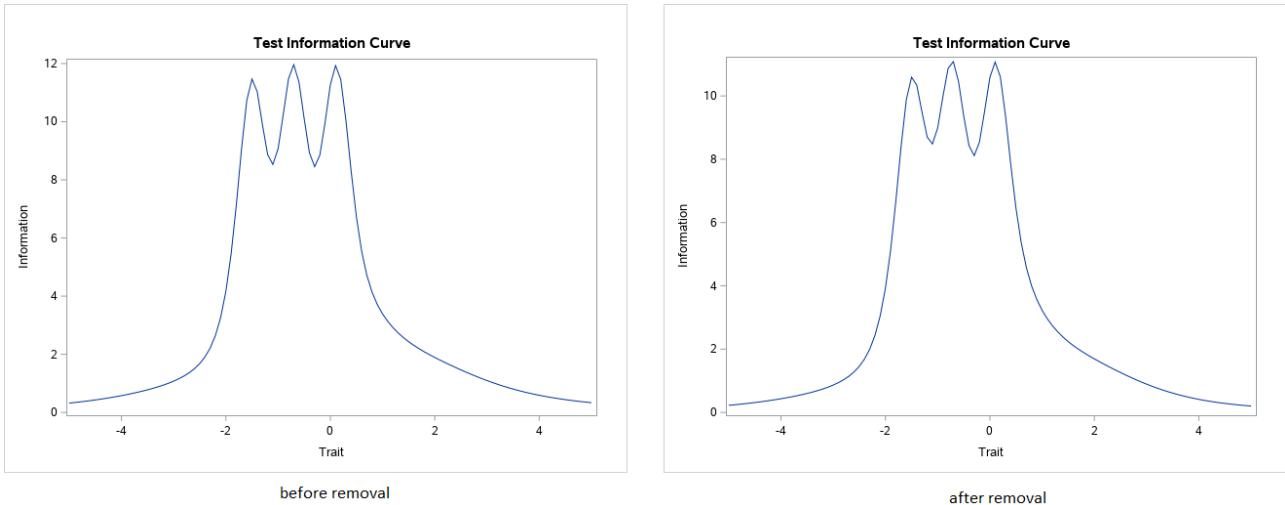
Let's see how the discrimination values behave

Item Information Curves





As we can see here half of the items provide a small contribution to the analysis and 3 of them apport litteraly 0 contribution. Let's see if we can remove those 3 items without losing information



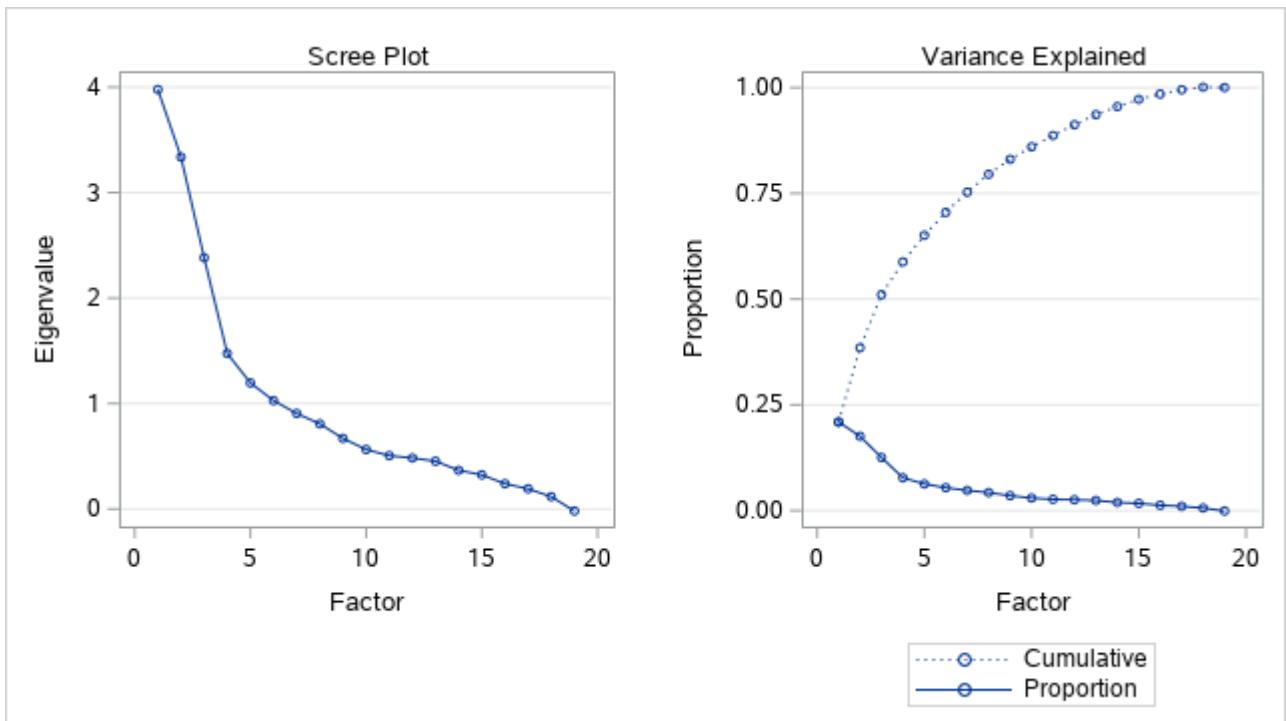
As we can clearly see, the shape remains almost the same, hence it can make sense to remove these items from this topic. Must be noted that with these items the test is almost not informative for the subjects with a latent trait far away from the difficulty level.

4.6.2.2 Wine buying experience

The items that fall into this topic are:

- BUYING_EXPERIENCE
- WINE_BOTTLES
- SUPERMARKET
- WINE_SHOP
- ONLINE_SHOP
- GRAPE_ORIGIN
- GRAPE_VARIETY
- BUDGET_FRIENDLY
- BRAND_AWARNESS
- VINTAGE
- LBEL_INFO
- PACKAGING
- PROMOTION
- BOTTLE_BUDGET
- BUYING_FREQUENCY
- PARTY
- GIFT
- HOME
- TASTE

Analyze the scree plot for this topic we find out that there are more than one driving force that influence the responses of the subjects (we can identify that considering that the first two eigenvalues are really close).



Let's check how the items relate to each of the 2 main factors.

	Factor1	Factor2
BUYING_EXPERIENCE	4.53496	-0.23805
WINE_BOTTLES	2.44893	0.67705
SUPERMARKET	0.11626	-1.58009
WINE_SHOP	0.36892	2.54220
ONLINE_SHOP	0.77575	0.70853
GRAPE_ORIGIN	0.18895	1.19539
GRAPE_VARIETY	0.20644	1.42104
BUDGET_FRIENDLY	-0.00366	-0.74577
BRAND_AWARENESS	-0.60841	-0.38655
VINTAGE	-0.01450	0.60222
LABEL_INFO	-0.53257	0.04486
PACKAGING	-0.00043	-0.63375
PROMOTION	0.04529	-1.09449
BOTTLE_BUDGET	-0.40587	1.15453
BUYING_FREQUENCY	0.47094	-0.18387
PARTY	0.50809	-0.33394
GIFT	0.74838	0.01602
HOME	2.25313	-0.36770
TASTE	1.08484	0.39723

In order to be sure of this division of the items let's apply a confirmatory model.

	Slope Matrix Estimate/StdErr/p-value	
	Factor1	Factor2
BUYING_EXPERIENCE	79.34910 27.19377 0.00176	0.00000
WINE_BOTTLES	2.18268 0.26621 <.00001	0.00000
SUPERMARKET	0.00000	-1.55373 0.23736 <.00001
WINE_SHOP	0.00000	2.57797 0.44087

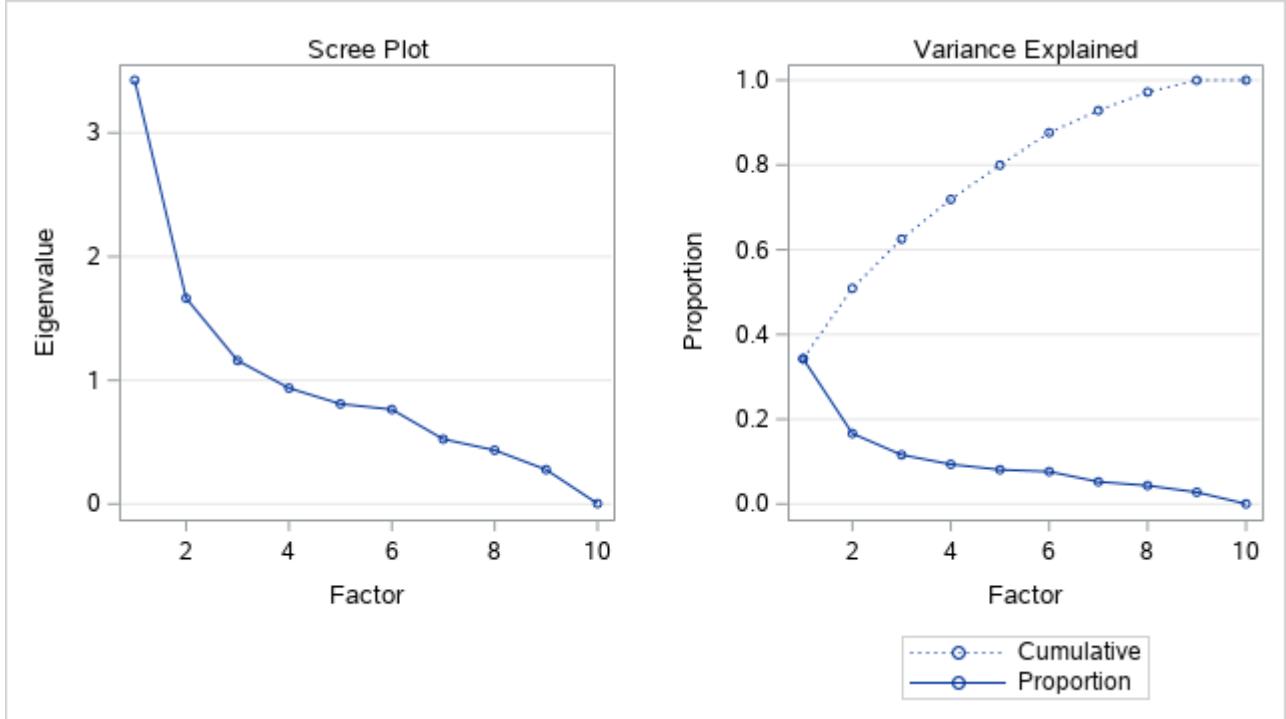
Slope Matrix Estimate/StdErr/p-value		
	Factor1	Factor2
		<.00001
ONLINE_SHOP	0.00000	0.82458 0.25045 0.00050
GRAPE_ORIGIN	0.00000	1.29075 0.20484 <.00001
GRAPE_VARIETY	0.00000	1.47098 0.22067 <.00001
BUDGET_FRIENDLY	0.00000	-0.68364 0.16641 0.00002
BRAND_AWARENESS	-0.51930 0.13975 0.00010	0.00000
VINTAGE	0.00000	0.70279 0.16017 <.00001
LABEL_INFO	-0.51551 0.13090 0.00004	0.00000
PACKAGING	0.00000	-0.55791 0.15332 0.00014
PROMOTION	0.00000	-1.02670 0.17793 <.00001
BOTTLE_BUDGET	0.00000	1.21632 0.22830 <.00001
BUYING_FREQUENCY	0.42222 0.13829 0.00113	0.00000
PARTY	0.58960 0.15762 0.00009	0.00000
GIFT	0.61259 0.17326 0.00020	0.00000
HOME	1.96227 0.28358 <.00001	0.00000
TASTE	0.94059 0.19632 <.00001	0.00000

As we can see, with a 0.05 confidence level, all our “classifications” are statistically significant.

Let's run IRT individually on each of these factors.

4.6.2.3 Factor1

Let's start by the classical check of the eigenvalues (even though we have already tested it)



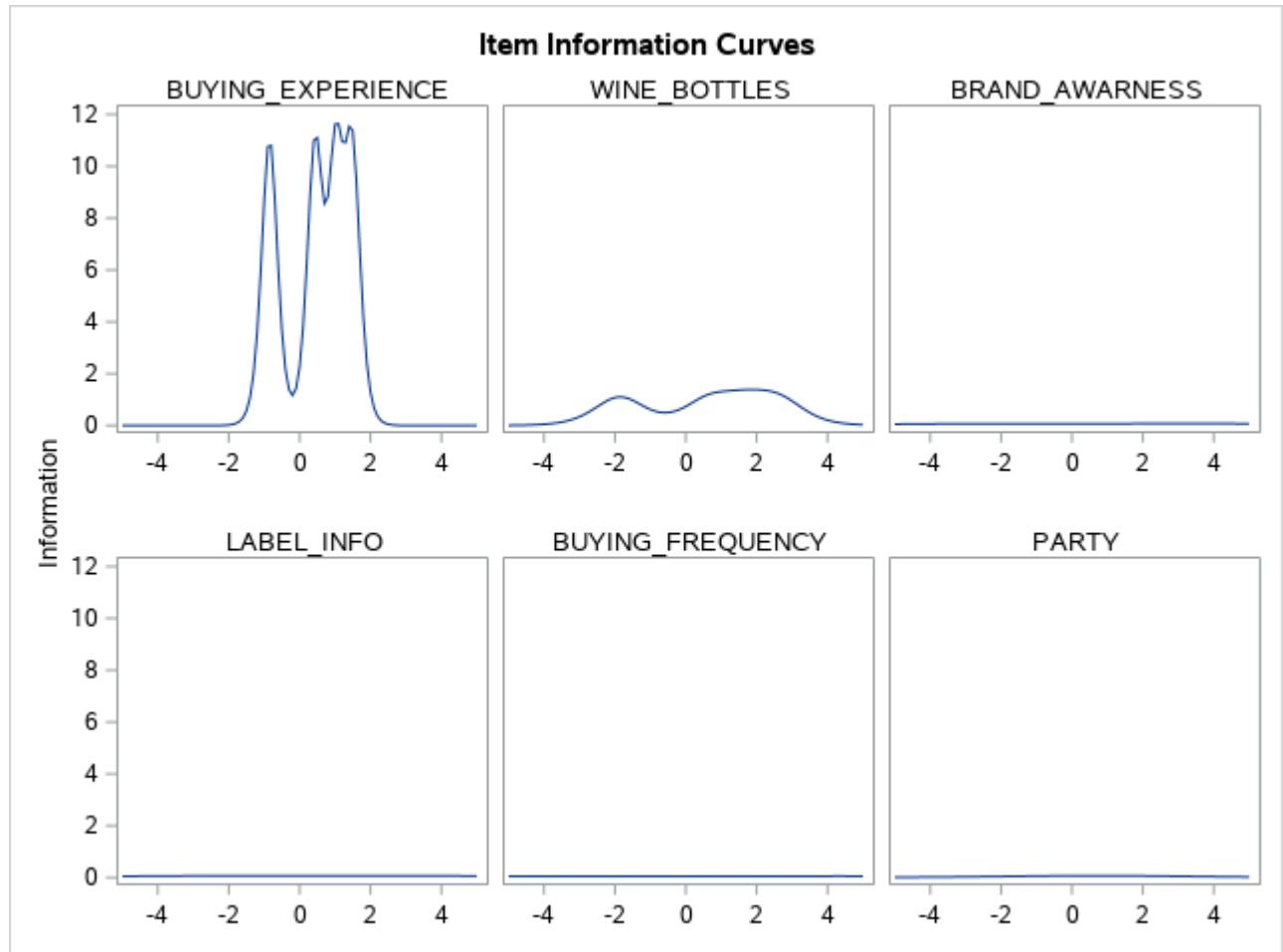
Clearly now one single latent trait drives all the answers in this subgroup.

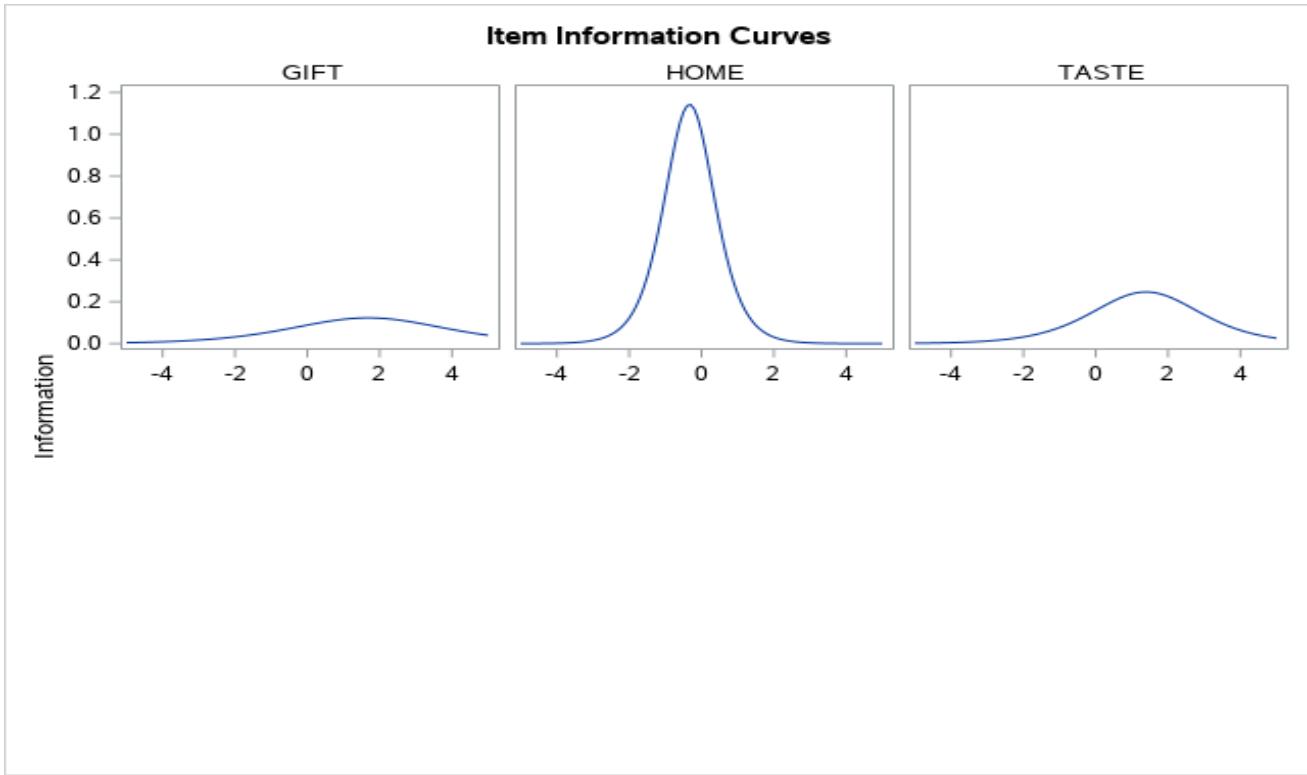
Let's have a look at the parameters.

Item Parameter Estimates						
Response Model	Item	Label	Parameter	Estimate	Standard Error	Pr > t
Graded	BUYING_EXPERIENCE	BUYING_EXPERIENCE	Threshold 1	-0.84837	0.09525	<.0001
			Threshold 2	0.44134	0.08575	<.0001
			Threshold 3	1.02677	0.11230	<.0001
			Threshold 4	1.47781	0.14217	<.0001
			Slope	6.65605	3.97951	0.0472
	WINE_BOTTLES	WINE_BOTTLES	Threshold 1	-1.86157	0.20434	<.0001
			Threshold 2	0.69241	0.11599	<.0001
			Threshold 3	1.49533	0.16299	<.0001
			Threshold 4	2.00724	0.20605	<.0001
			Threshold 5	2.58909	0.27820	<.0001
	BRAND_AWARENESS	BRAND_AWARENESS	Slope	2.09483	0.29682	<.0001
			Threshold 1	4.85979	1.39640	0.0003
			Threshold 2	2.19160	0.66899	0.0005
			Threshold 3	-3.18248	0.90102	0.0002
	LABEL_INFO	LABEL_INFO	Slope	-0.50949	0.14862	0.0003
			Threshold 1	4.10875	1.09723	<.0001
			Threshold 2	1.18885	0.42543	0.0026
			Threshold 3	-3.16453	0.82423	<.0001
			Slope	-0.49170	0.13041	<.0001

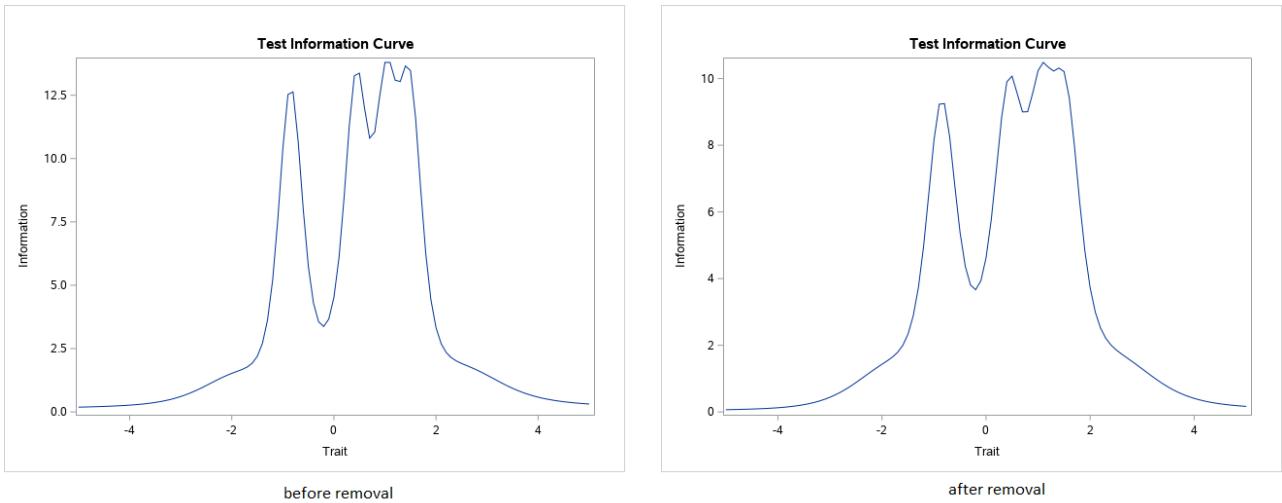
Item Parameter Estimates						
Response Model	Item	Label	Parameter	Estimate	Standard Error	Pr > t
	BUYING_FREQUENCY	BUYING_FREQUENCY	Threshold 1	-5.16670	1.65592	0.0009
			Threshold 2	-4.03148	1.27897	0.0008
			Threshold 3	2.67177	0.89882	0.0015
			Threshold 4	6.70733	2.16916	0.0010
			Slope	0.43248	0.14066	0.0011
TwoP	PARTY		Difficulty	0.72019	0.31930	0.0120
			Slope	0.54365	0.16319	0.0004
	GIFT		Difficulty	1.68423	0.44043	<.0001
			Slope	0.70145	0.18861	<.0001
	HOME		Difficulty	-0.33020	0.10563	0.0009
			Slope	2.13790	0.35278	<.0001
	TASTE		Difficulty	1.38902	0.28350	<.0001
			Slope	0.99340	0.21850	<.0001

Here we don't have neither extremely easy items nor extremely difficult items. Let's understand the contribution of each of these items.



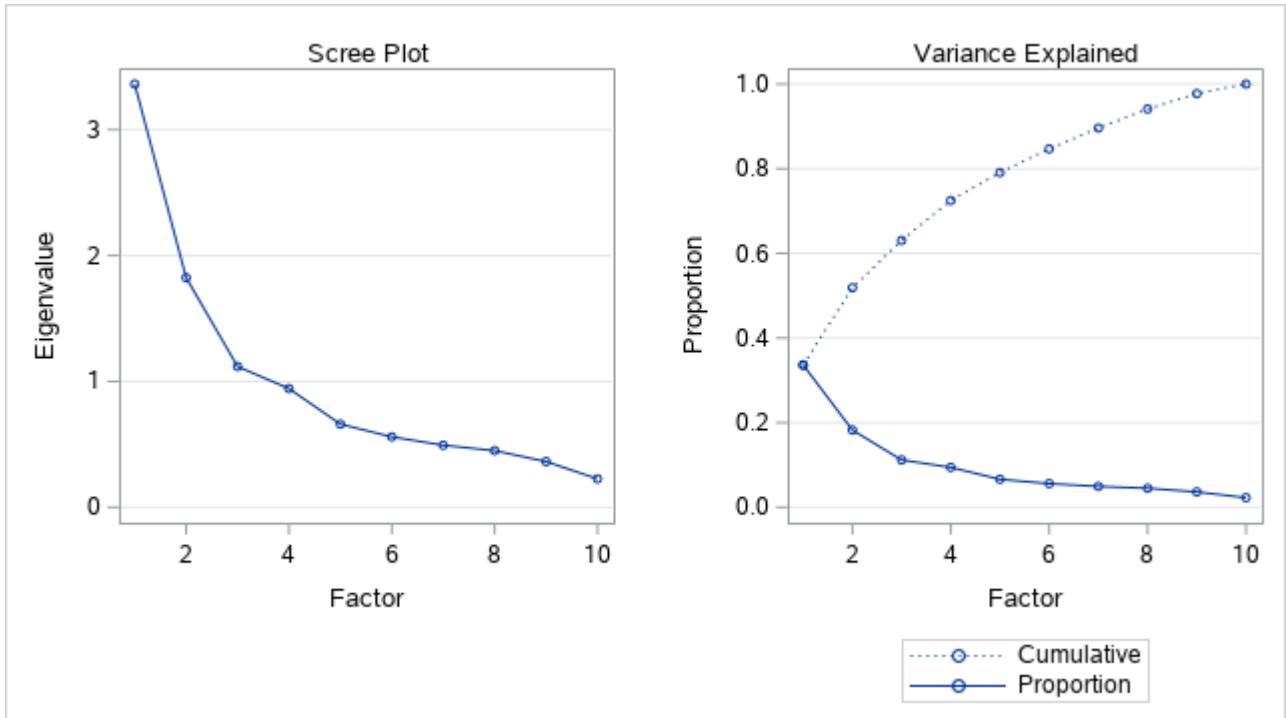


Not all the variables are informative, so let's remove those that don't apport any information. As we can see all the variables are more informative for those subjects whose latent trait is close to the difficulty level.



Also in this case the topic is not informative for those subjects with a really high or really low latent trait.

4.6.2.4 Factor2



Also for this factor we can see that our division was done in a good way. Let's analize the parameters.

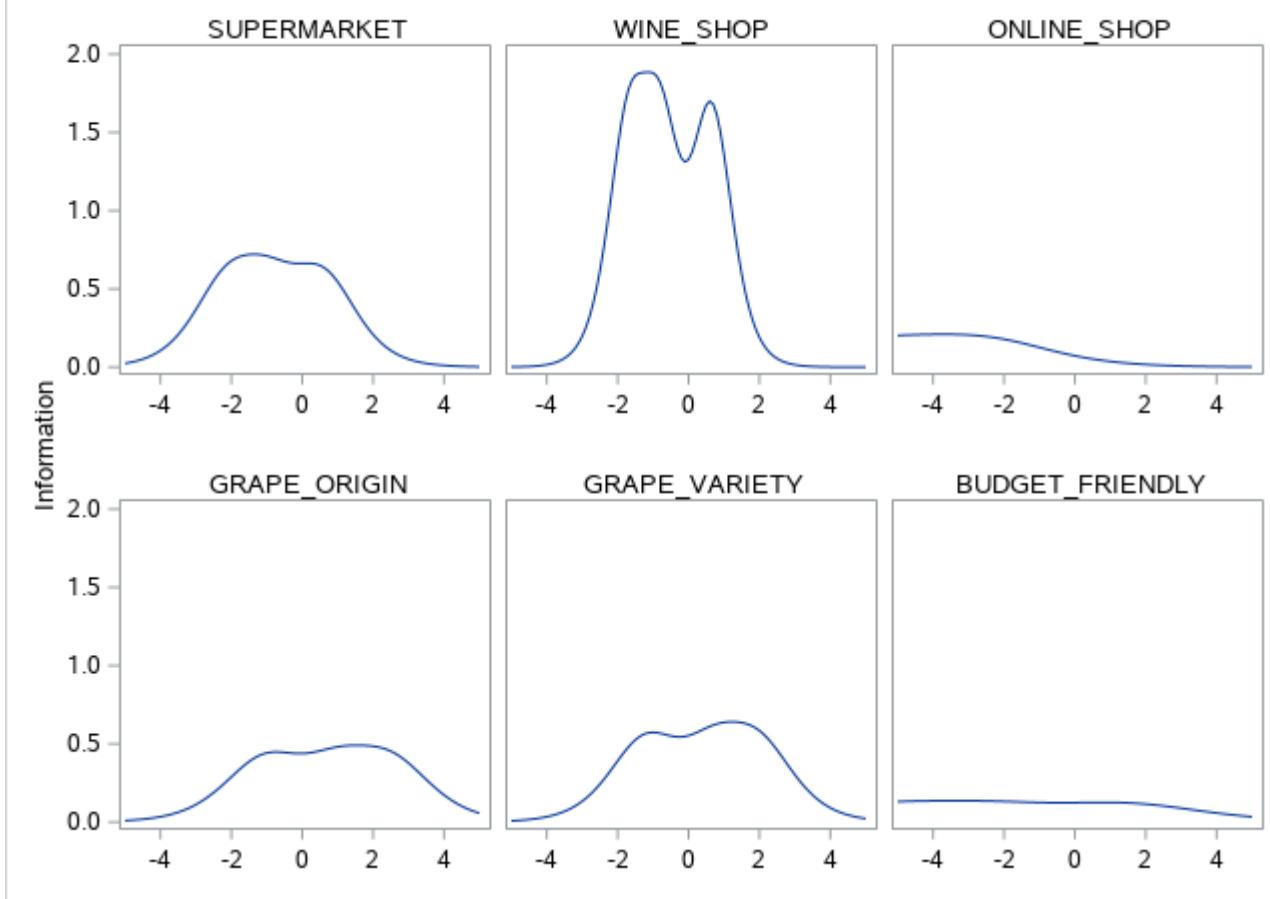
Item Parameter Estimates					
Item	Label	Parameter	Estimate	Standard Error	Pr > t
SUPERMARKET	SUPERMARKET	Threshold 1	-2.05084	0.26494	<.0001
		Threshold 2	-1.00194	0.15756	<.0001
		Threshold 3	0.55136	0.13460	<.0001
		Slope	1.55320	0.23721	<.0001
WINE_SHOP	WINE_SHOP	Threshold 1	0.64472	0.11225	<.0001
		Threshold 2	-0.82778	0.11538	<.0001
		Threshold 3	-1.66072	0.17776	<.0001
		Slope	-2.57776	0.44100	<.0001
ONLINE_SHOP	ONLINE_SHOP	Threshold 1	-2.39785	0.63111	<.0001
		Threshold 2	-3.74134	1.01889	0.0001
		Threshold 3	-5.71833	1.68898	0.0004
		Slope	-0.82452	0.25044	0.0005
GRAPE_ORIGIN	GRAPE_ORIGIN	Threshold 1	2.44037	0.33850	<.0001
		Threshold 2	1.03492	0.18377	<.0001
		Threshold 3	-1.02546	0.17883	<.0001
		Slope	-1.29103	0.20484	<.0001
GRAPE_VARIETY	GRAPE_VARIETY	Threshold 1	1.90895	0.25016	<.0001
		Threshold 2	0.75787	0.14696	<.0001
		Threshold 3	-1.19877	0.17881	<.0001
		Slope	-1.47078	0.22059	<.0001
BUDGET_FRIENDLY	BUDGET_FRIENDLY	Threshold 1	-5.22802	1.28977	<.0001
		Threshold 2	-2.42216	0.58137	<.0001
		Threshold 3	1.41284	0.37415	<.0001
		Slope	0.68339	0.16638	<.0001

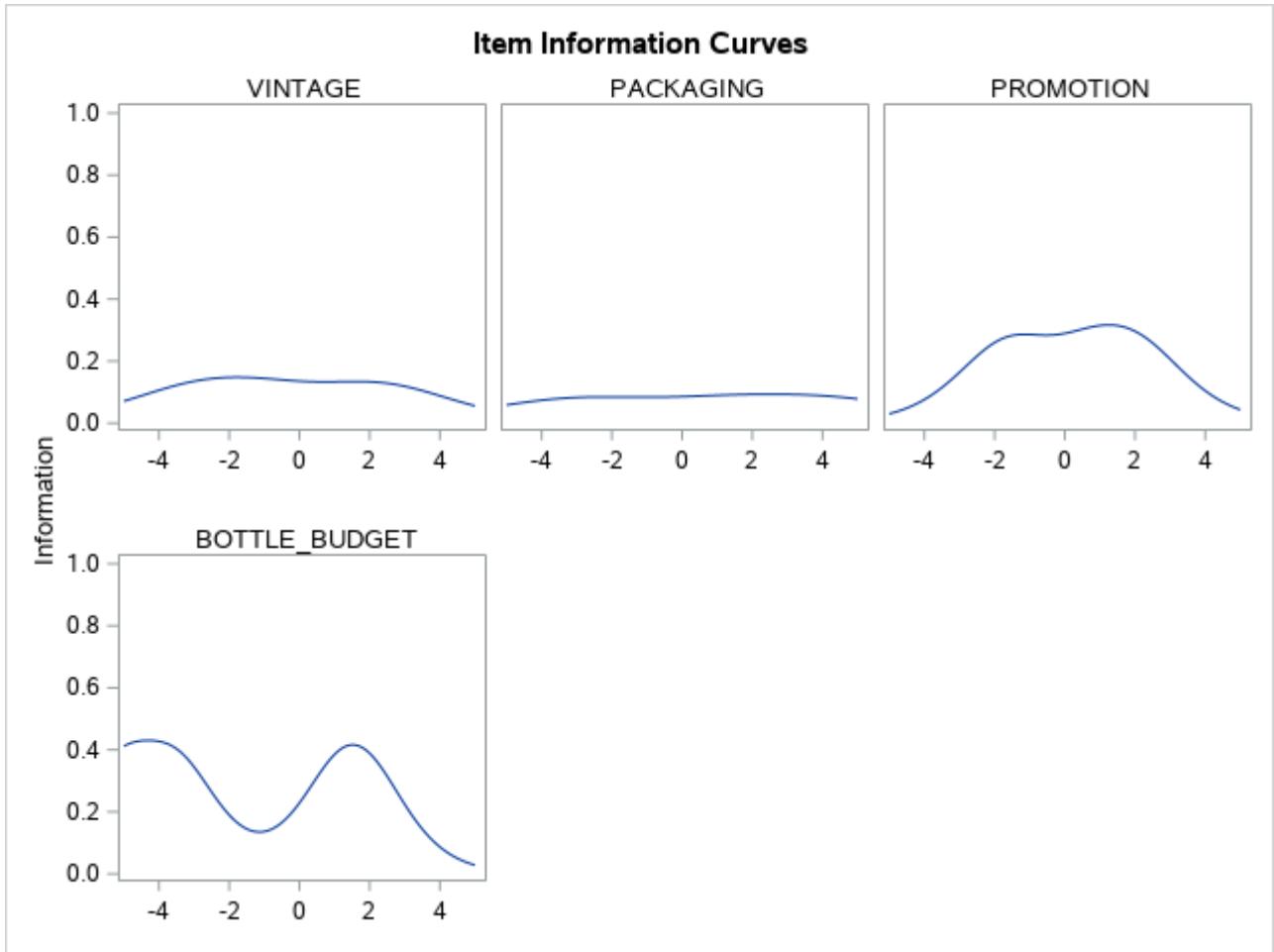
Item Parameter Estimates					
Item	Label	Parameter	Estimate	Standard Error	Pr > t
VINTAGE	VINTAGE	Threshold 1	2.28997	0.52675	<.0001
		Threshold 2	-1.31770	0.32701	<.0001
		Threshold 3	-2.76453	0.61062	<.0001
		Slope	-0.70275	0.16012	<.0001
PACKAGING	PACKAGING	Threshold 1	-3.02100	0.82093	0.0001
		Threshold 2	1.51175	0.46321	0.0005
		Threshold 3	4.20183	1.13141	0.0001
		Slope	0.55778	0.15329	0.0001
PROMOTION	PROMOTION	Threshold 1	-1.59804	0.27803	<.0001
		Threshold 2	0.87707	0.19726	<.0001
		Threshold 3	1.97581	0.32751	<.0001
		Slope	1.02667	0.17795	<.0001
BOTTLE_BUDGET	BOTTLE_BUDGET	Threshold 1	1.77444	0.29108	<.0001
		Threshold 2	1.27511	0.22409	<.0001
		Threshold 3	-3.54829	0.61508	<.0001
		Threshold 4	-5.07811	1.13654	<.0001
		Slope	-1.21608	0.22824	<.0001

Also in this case all the items' difficulties seems to be balanced, the only items that can be considered relatively difficult are “ONLINE_SHOP” and “BUDGET_FRIENDLY” and “BOTTLE_BUDGET”.

Let's analyze how the items contribute in providing information.

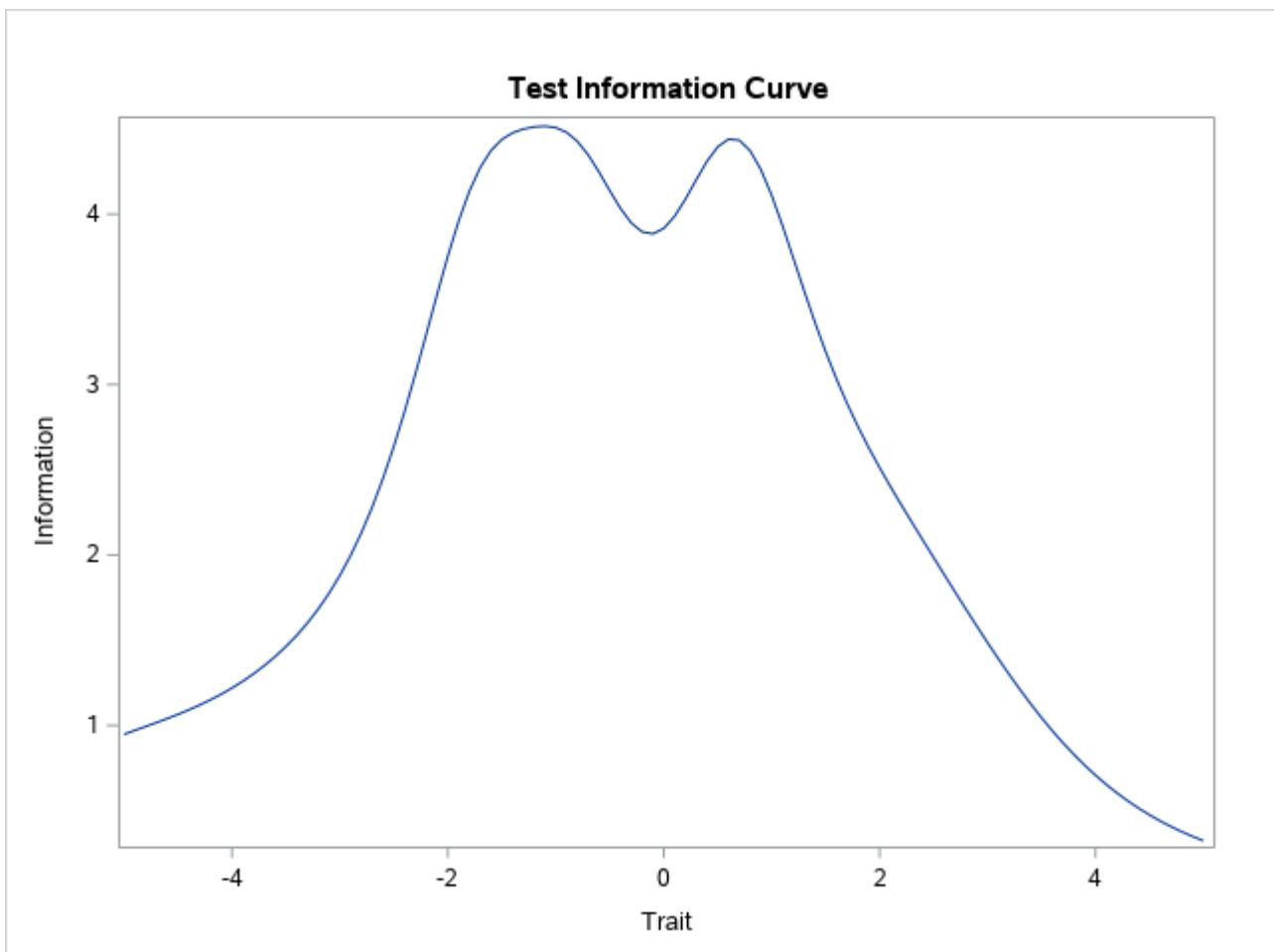
Item Information Curves





We can notice that also in this case we don't have any item that is able to be informative for those subjects with a very high latent trait but there are some of them that are informative for subjects with a low latent trait.

As we can see all the items provide some information, so we will not delete any of them.



As we can see the Test information curve remains the same and so it is safe to remove the item.

Also in this case the test is not accurate for subjects with a higher level of the latent trait.

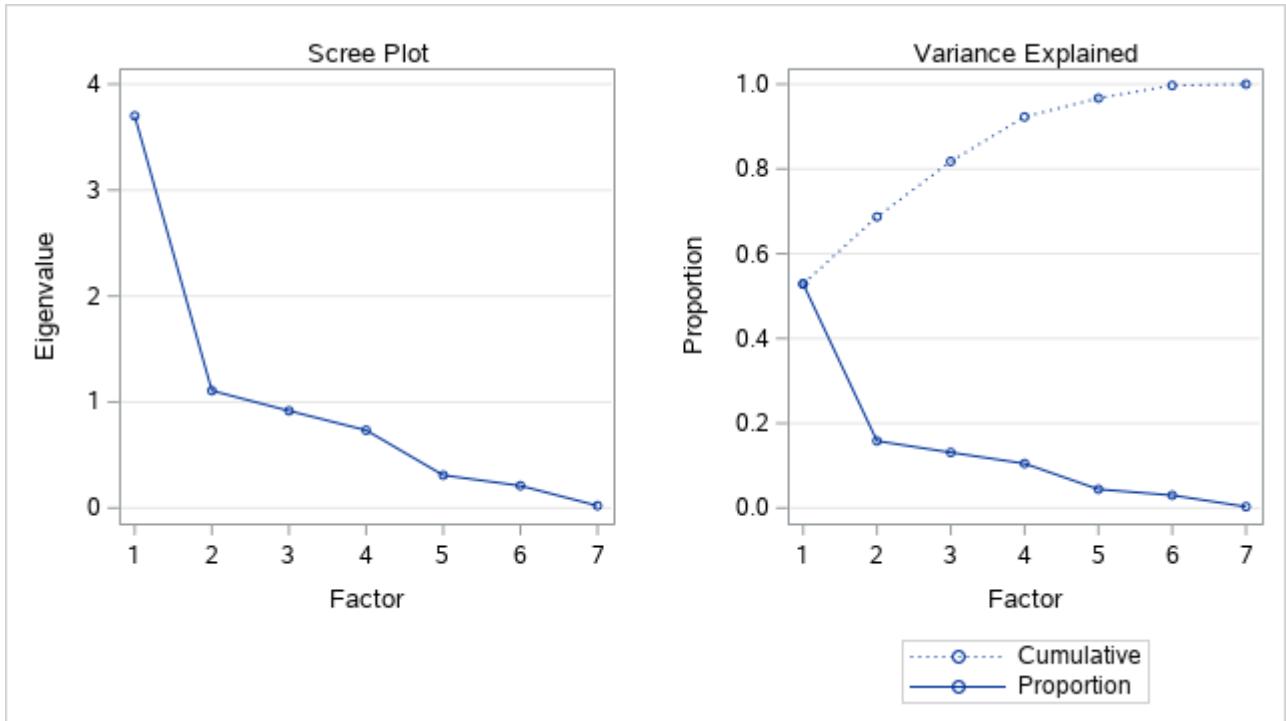
4.6.2.3 Etna DOC Wine

The items in this topic are:

- ETNA_DOC
- ETNA_BUYING
- ETNA_PREFERENCE
- ETNA_FLAVOR
- SICILIAN_EXCELLENCES
- ETNA_EXPENSIVE
- ETNA_QUALITY
- ETNA_RECOMMENDATION

Here the IRT model will be applied to all these items except for ETNA_DOC, considering that it was the discriminatory question to finish the questionnaire, in fact if we try running PROC IRT including also ETNA_DOC it will skip the calculation of the Pearson chi-square model and interrupt the computation.

Let's start looking at the scree plot



As we can see in this case we, thankfully, have that one single latent trait drives the responses.

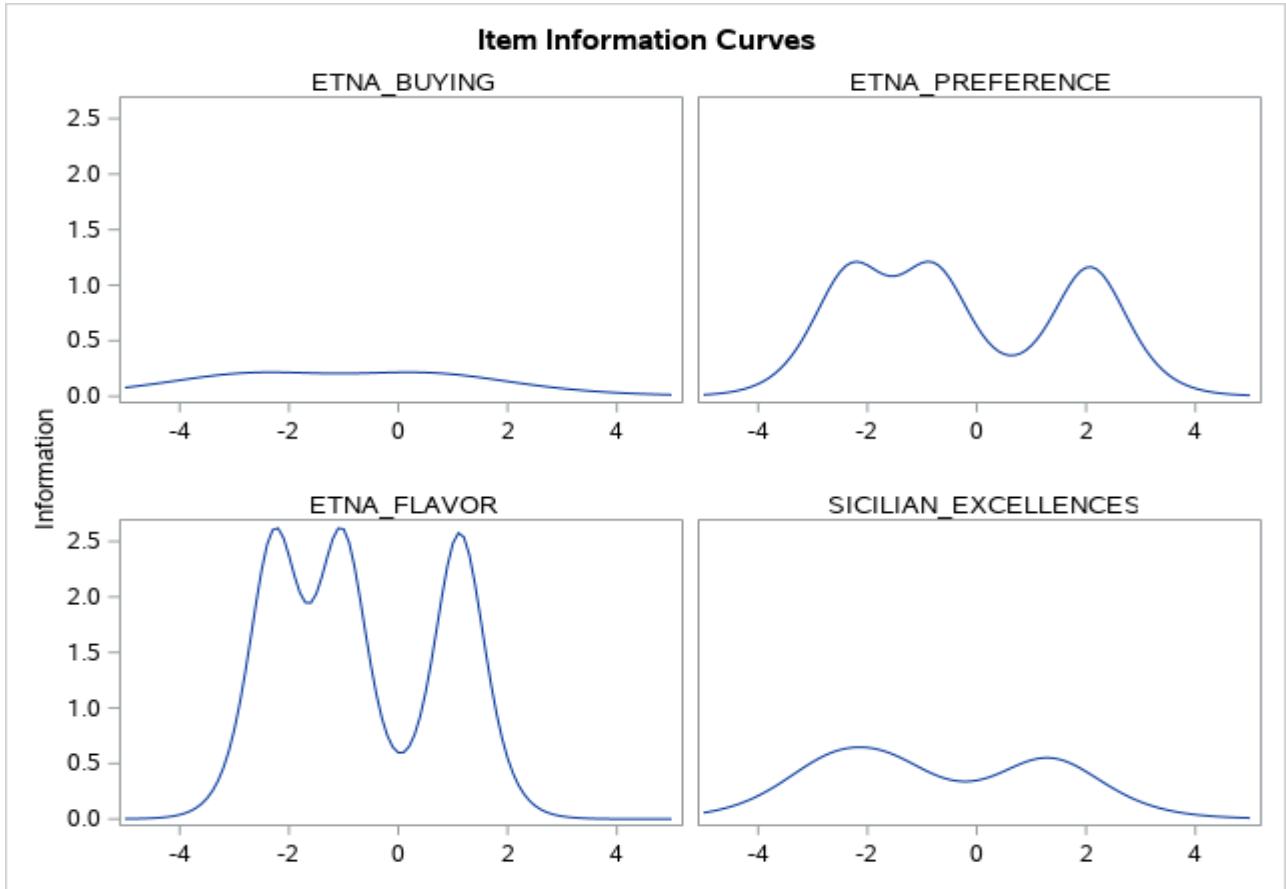
As usual, now we will have a look at the estimated parameters.

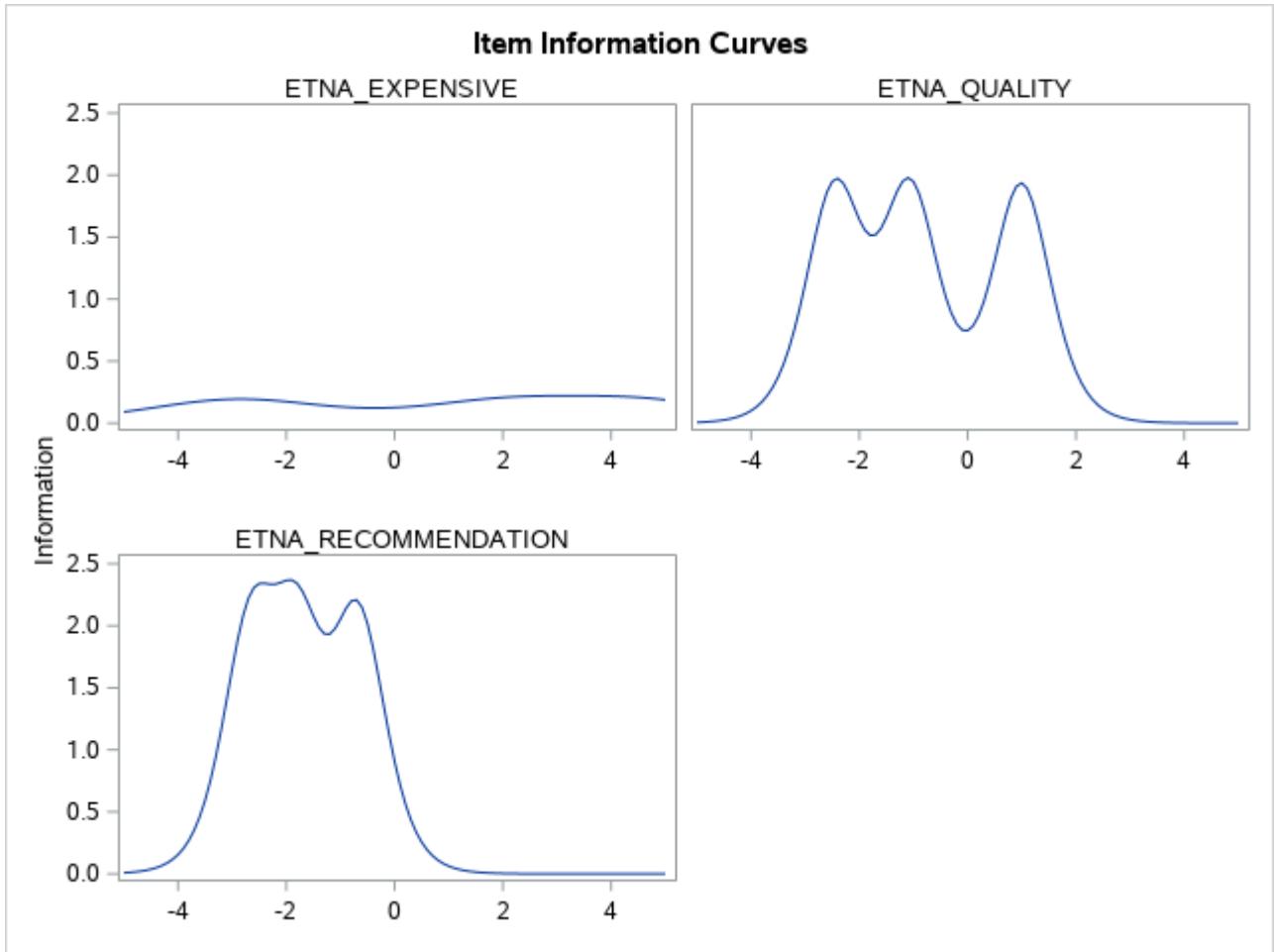
Item Parameter Estimates					
Item	Label	Parameter	Estimate	Standard Error	Pr > t
ETNA_BUYING	ETNA_BUYING	Threshold 1	-2.65466	0.50939	<.0001
		Threshold 2	0.50623	0.18294	0.0028
		Slope	0.90241	0.17655	<.0001
ETNA_PREFERENCE	ETNA_PREFERENCE	Threshold 1	-2.28931	0.25441	<.0001
		Threshold 2	-0.80289	0.12466	<.0001
		Threshold 3	2.07287	0.23130	<.0001
ETNA_FLAVOR	ETNA_FLAVOR	Slope	2.15628	0.30981	<.0001
		Threshold 1	-2.26216	0.23019	<.0001
		Threshold 2	-1.03141	0.12388	<.0001
SICILIAN_EXCELLENCES	SICILIAN_EXCELLENCES	Threshold 3	1.12235	0.12917	<.0001
		Slope	3.20976	0.52298	<.0001
		Threshold 1	-2.56419	0.33633	<.0001
ETNA_EXPENSIVE	ETNA_EXPENSIVE	Threshold 2	-1.74905	0.23848	<.0001
		Threshold 3	1.31425	0.19186	<.0001
		Slope	1.47788	0.21703	<.0001
ETNA_QUALITY	ETNA_QUALITY	Threshold 1	-2.44641	0.25699	<.0001
		Threshold 2	-1.07092	0.12956	<.0001
		Threshold 3	0.99348	0.12632	<.0001
ETNA_RECOMMENDATION	ETNA_RECOMMENDATION	Slope	2.77894	0.41064	<.0001
		Threshold 1	-2.64018	0.28968	<.0001
		Threshold 2	-1.81608	0.18965	<.0001

Item Parameter Estimates					
Item	Label	Parameter	Estimate	Standard Error	Pr > t
		Threshold 3	-0.67946	0.10659	<.0001
		Slope	2.92060	0.50387	<.0001

As can be seen from the table, there are no extremely easy nor extremely difficult items.

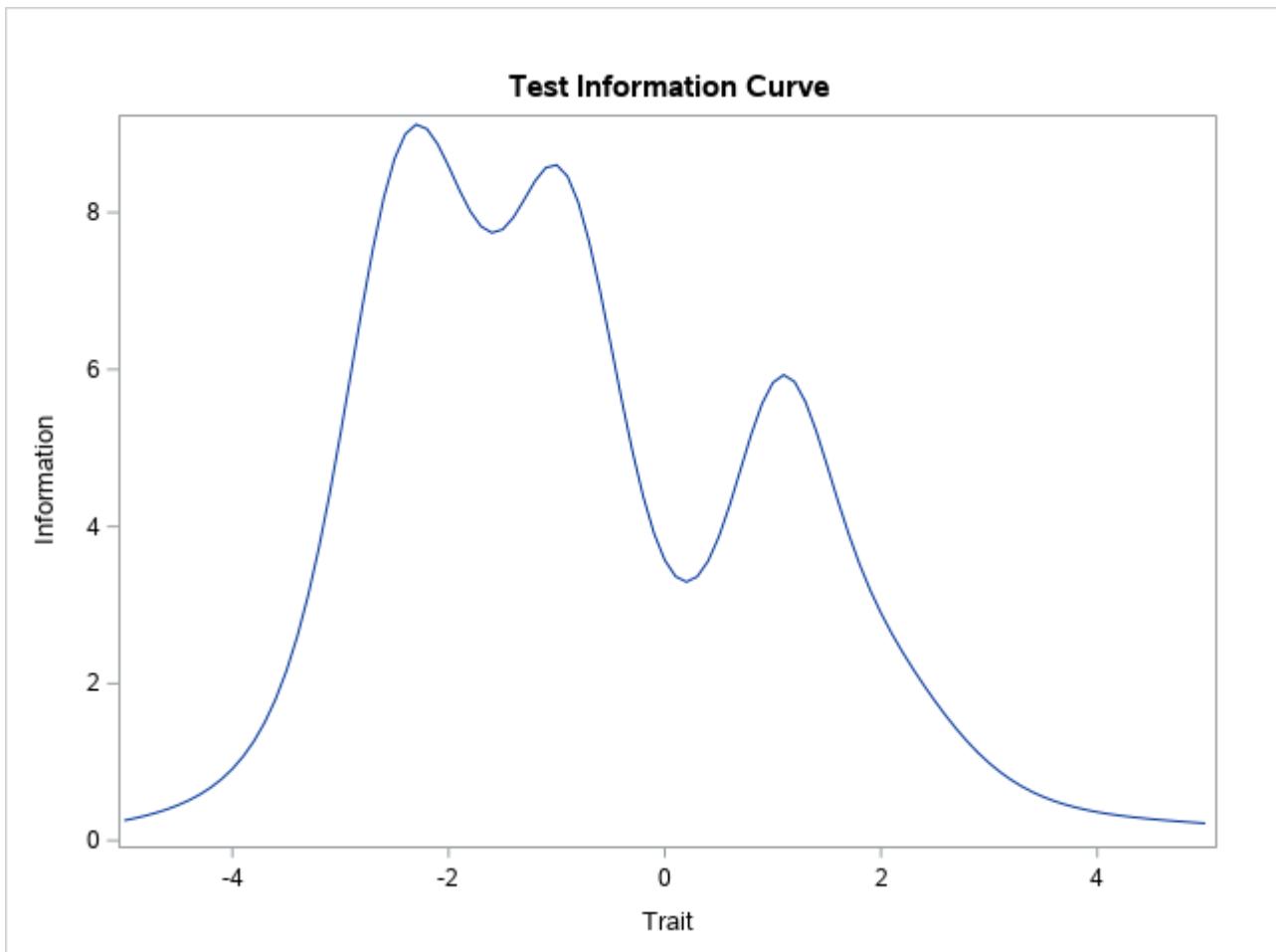
Let's see how much information each item provides.





Even though also here not the full spectrum for the latent rate is covered, this items are better distributed than the others covering a good part of the spectrum. All of them provide information so there is no way to remove items without losing information about the subjects.

Let's see how the test information curve behaves.



Differently from the other topics here less information are provided for subjects with a latent trait level close to the difficulty level of the items, instead more information are given for those with a lower latent trait level.

4.6.3 IRT on factors

Considering that there is almost a perfect match between the factors identified in the factor analysis and the topics identified by both the questionnaire design and the exploratory IRT in the second topic, we can affirm that the factors identified in the factor analysis are meaningful.

Conclusion

According to the Multivariate Analysis we have done, we found the presence of some items not relevant and that don't give particular insights in our survey.

Based to the Factor Analysis and IRT the following table describe the items can be removed and that are not relevant in the survey.

FACTOR ANALYSIS	ITEM RESPONSE THEORY
Cocktail_Preference	Cocktail_Preference
Beer_Preference	
Soft_Preference	
Online_Shop	
Buying_Preference	Buying_Preference
	Budget_friendly
	Packaging
	Promotion
	Etna_Preference
	Label_Info
	Bottle_Budget
	Etna_Recommendation

We can see that the only matching variables are Cocktail_Preference and Buying_Preference, this doesn't means that are the only items useless and not relevant. we have found that the questions:

- “During the pandemic, did the frequency with which you buy wine change?” can be removed from the Questionnaire, the respondents on average didn't change their habits.
- The items Cocktail_Preference doesn't provide any relevant information in our analysis, in addition it can be out of range to the main topic of the survey that is wine. According to this consideration Beer_Preference and Soft_Preference even if are out of topic give us some relevant statistical correlations and are useful to analyze some habits.
- There are some items like “Online_Shop” and “How often do you buy wine on average in a month?” which are almost empty and we can delete as possible option in our survey, there are just 3 out 247 respondents that buy wine online and just one respondent that buy more than 7 times per month wine.
- Question 13 : “For what reasons have you bought wine in the last 3 months” is a multiple choice question, to have a more precise answer could be better giving to the respondents the possibility to make a single choice in order to have on average a more precise indicator.

According to the results of IRT and Factor Analysis, the second topic, "Wine buying experience" could be splitted into two topics that here we will call "Wine Buying Experience" and "Wine consumption", composed respectively by the following items:

1. Wine Buying Experience :

- SUPERMARKET
- WINE_SHOP
- ONLINE_SHOP
- GRAPE_ORIGIN
- GRAPE_VARIETY
- BUDGET_FRIENDLY
- VINTAGE
- PACKAGING
- PROMOTION
- BOTTLE_BUDGET

2. Wine Consumption:

- BUYING_EXPERIENCE
- WINE_BOTTLES
- BRAND_AWARENESS
- LABEL_INFO
- BUYING_FREQUENCY
- PARTY
- GIFT
- HOME
- TASTE

From the Latent Class Analysis we have identified that there are 2 classes that can be target of marketing actions: Class 1 and Class 3.

Class 1 is mainly composed by mid-age highly educated subjects with a strong interest in wine. Considering they have taken wine classes and have an high knowledge in wine a marketing action should focus more on features that can help a particular type of wine to stand up from the others.

Class 3, instead, is mainly composed by young people with an high-school degree. Considering they enjoy wine in parties and that they can be more easily influenced by packagings and fancy labels in the bottle, a marketing action aimed to these subjects must transmit the so called "chill vibes" (enjoy the life and having fun).

Appendix

For future research we suggest:

- Add an option to insert the salary of the respondent. Combining this with the price ranges they buy can be interesting to actually evaluate how much attention they have towards wine.
- add control items to preventively remove subjects that responded randomly.
- According to IRT it should be necessary to design items able to be informative for subjects with a really high or really low latent trait.
- Be more specific on the winery course, a subject who randomly participated to a winery course must be treated differently from someone who has certifications.
- The middle value option considered in all the question in Likert scale should be avoided to not bias the sample, because people are not always completely indifferent to an argument so, respondents have to be forced to give an answer.
- ask items to identify if subjects could be interested in a free winery course and/or a free winery tasting experience. This could be used to make people of class 4 (see classes in Latent Class Analysis) closer to wine.
- Add item to see if subjects have ever seen a wine related advertising. A positive response can identify a subject who is more easily susceptible to advertising campaigns.