

©2016

Jingkun Zhuang

ALL RIGHTS RESERVED

EMPIRICAL STUDY OF WINE CONSUMER CHARACTERISTICS AND  
MARKETING STRATEGIES IN MID-ATLANTIC REGION

By

JINGKUN ZHUANG

A thesis submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Food and Business Economics

Written under the direction of

Ramu Govindasamy

And approved by

---

---

---

---

New Brunswick, New Jersey

October, 2016

## **ABSTRACT OF THE THESIS**

### **EMPIRICAL STUDY OF WINE CONSUMER CHARACTERISTICS AND MARKETING STRATEGIES IN MID-ATLANTIC REGION**

by JINGKUN ZHUANG

Thesis Director:

Dr. Ramu Govindasamy

In the past 15 years, the U.S. wine market has been growing very fast. The number of wineries has increased from 2688 in 1999 to 8862 in 2016 (Wines Vines Analytics, 2016). About 7% of all those wineries are located in the Mid-Atlantic region which includes New Jersey, New York and Pennsylvania. However, competition has been rising as the market grows. Many foreign wine companies from Europe, South America and Oceania are either selling or planning to sell their products to the fast growing U.S. wine market. These new market situations and changes in purchasing behavior demand that the Mid-Atlantic wineries revisit the preferences of wine consumers and consider the factors that affect the buying choice. In this research, we would like to investigate how wine drinking behavior is related to the demographic status of the residents in the three states. We expect that people with different age, gender, marital status, family income, and education background

will have different wine drinking behaviors due to their differing life styles. The study results will help the Mid-Atlantic wineries to develop a more efficient marketing strategy. This study is based on data from an online survey that was conducted by Penn State University in 2009. 1246 Mid-Atlantic wine drinkers participated in this survey. First, we summarized the characteristics of the Mid-Atlantic wine market by looking into the descriptive statistics of our survey questions. Then we employed Logistic Regression to answer the question of what kind of people are more likely to purchase locally produced wine. In addition, we used Cluster Analysis to segment the Mid-Atlantic wine market. Marketing strategies are based on the 4Ps Marketing Mix model that were developed for Mid-Atlantic wineries.

Keywords: Wine, Purchase Behavior, Consumer Behavior, Logistic Regression, Cluster Analysis, Market Segmentation, Marketing Strategy, Decision Making, Mid-Atlantic, NY, NJ, PA

## **ACKNOWLEDGEMENT**

Throughout my graduate education Dr. Govindasamy has been an excellent mentor. I won't be able to finish this thesis without his help. I'd also like to thank Dr. Gal Hochman and Dr. Isaac Vellangany for their valuable advice. Thanks to my family for supporting me all the time!

## Table of Contents

<b>ABSTRACT OF THE THESIS .....</b>	<b>ii</b>
<b>ACKNOWLEDGEMENT .....</b>	<b>iv</b>
<b>List of Figures .....</b>	<b>vii</b>
<b>1 INTRODUCTION .....</b>	<b>1</b>
1.1 Mid-Atlantic Wineries and Production.....	1
1.2 Mid-Atlantic Wine Consumption.....	2
<b>2 LITERATURE REVIEW .....</b>	<b>4</b>
2.1 Model of Consumer Behavior .....	4
2.2 Demographic Characteristics Affects Wine Consuming Decision.....	4
2.3 Marital Status Affects Alcohol Consumption.....	5
<b>3 METHODS.....</b>	<b>6</b>
<b>4 DESCRIPTIVE STATISTICS .....</b>	<b>7</b>
<b>5 LOGISTIC REGRESSION .....</b>	<b>36</b>
5.1 Dependent variable BUY .....	37
5.2 Independent variables.....	37
5.3 Model Tweak.....	43
5.4 Logistic Regression Output.....	45
<b>6 MARKET SEGMENTATION USING CLUSTER ANALYSIS .....</b>	<b>53</b>
6.1 Two-Way Contingency and Chi-Square Independence Test of Wine Consumer Clusters .....	57
6.2 Results of Cluster Analysis.....	61
<b>7 CONCLUSIONS .....</b>	<b>62</b>
7.1 What Kind of People Are More Likely to Buy Local Wine? .....	62
7.2 Mid-Atlantic Wine Market Segmentation.....	64
7.3 Marketing Strategies based on 4Ps Marketing Mix.....	65
<b>REFERENCE .....</b>	<b>67</b>

## **List of Tables**

Table 1 List of independent variables for logistic regression .....	40
Table 2 Logistic Regression Output.....	52
Table 3 Cross table of BUY variable and clusters .....	56
Table 4 Contingency Table and Independence Test of Wine Consumer Clusters.....	58
Table 5 Profile of Wine Consumer Clusters .....	60

## List of Figures

Figure 1. State where respondent resides.....	7
Figure 2. Gender of survey respondents .....	7
Figure 3. Age categories stacked by state .....	8
Figure 4. Education Level.....	9
Figure 5. Annual Family Income .....	10
Figure 6. Job Occupations.....	11
Figure 7. Marital Status.....	11
Figure 8. Q1a How often do you drink wine during an average year? Percentage of total respondents .....	12
Figure 9. Drinking frequency by gender. Counts are number of respondents. ....	13
Figure 11. Purchasing frequency stacked by state .....	15
Figure 10. Estimated Wine consumption of survey respondents per month in liters .....	15
Figure 12. Purchasing frequency stacked by age group .....	16
Figure 13. Purchasing frequency stacked by gender .....	16
Figure 14. How you purchase wine .....	17
Figure 15. What's your first alcoholic drink?.....	18
Figure 16. Do you buy different wine for everyday consumption and special occasions?19	
Figure 17. Price difference of wine you purchased for everyday and special occasion ...	19
Figure 18. Price ranges of wine you pay for everyday and special occasions.....	20
Figure 19. Price range break down by gender .....	21
Figure 20. Price range break down by state.....	21
Figure 21 Difference in sweetness/dryness.....	23
Figure 22 Difference in closure type .....	23
Figure 23 Difference in bottle size.....	23



Figure 24 Difference in Packaging .....	23
Figure 25 How your wine consumption has changed over the past 3 years .....	24
Figure 26. Reasons of decreased wine consumption .....	26
Figure 27. Reasons of increased wine consumption .....	26
Figure 28 Have visited wineries in the region .....	28
Figure 29 Have drank wine from the region .....	28
Figure 30 Neither purchased nor drank but interested .....	28
Figure 31 Have purchased wine from the region .....	28
Figure 32. Which outlet do people purchase wine from? .....	29
Figure 33. Wine consuming occasions (y-axis are ratio of total respondents, 0.5 means 50% of total respondents chose this option) .....	30
Figure 34 Drinking frequency for different wine varietal (x-axis is the ratio of total respondents) .....	31
Figure 35 Monthly Spending grouped by state and gender .....	32
Figure 36 Spending on wine during an average month .....	32
Figure 37 Monthly spending on different wine varieties .....	33
Figure 38 Which components is mandatory for a winery to offer or implement? .....	34
Figure 39 Do you buy wine made from fruits other than grapes? .....	35
Figure 40 Histogram of age .....	43
Figure 41 Regrouping Family Income .....	44
Figure 42 Dendrogram of Cluster Analysis .....	54
Figure 43 Elbow plot of optimal number of clusters .....	54
Figure 44 Dendrogram of 4 clusters .....	55

# **EMPIRICAL STUDY OF WINE CONSUMER CHARACTERISTICS AND MARKETING STRATEGIES IN MID-ATLANTIC REGION**

## **1 INTRODUCTION**

Wine is one of the most important drinks in people's daily life in the United States. It is also considered as a part of American culture. In the past few years, wine consumption in the U.S. Market has grown, although some consumers who used to consume wines at restaurants, began to purchase wine through retail stores in this down economy (RNCOS, 2011). This significant change in consumer behavior suggests that a new marketing strategy needs to be developed. Wine suppliers need to better understand their consumers in the retail segment, something they may have not done in the past. There are a lot of new questions that need to be answered, such as occasions for consuming wine, varietal preferences, purchasing frequency, drinking frequency and so forth. By uncovering these and other questions, wine suppliers can make their marketing and promotional efforts much more efficient. This research focuses on the Mid-Atlantic wineries and market.

### **1.1 Mid-Atlantic Wineries and Production**

By the end of June 2016, the number of wineries in the U.S. was 8862 (Wines Vines Analytics, 2016), which were only 2688 in 1999 (Fisher, 2011). About 7% of all wineries are located in these three Mid-Atlantic States: New Jersey (52 by June 2016), New York (367 by June 2016), and Pennsylvania (220 by June 2016). Though the total number of wineries in Mid-Atlantic area is relatively small, the growth has matched the U.S trend.

New York ranked 4<sup>th</sup> out of the 50 states in term of the number of wineries, with Pennsylvania ranked 7<sup>th</sup>, and New Jersey ranking 20<sup>th</sup> (Fisher, 2011). Grape and wine productions have more advantages for these states. New York and Pennsylvania ranked 3rd (Whetstone, 2011) and 7th, respectively. According to data, by the end of 2010, bulk wine production in these three states was just under 4% in which New York produced 93% of the total number. Of the remaining 96 percentage points, almost 90 percentage points were produced in California. In states other than California, New York, New Jersey and Pennsylvania shares the remaining 6 percentage points (Storchmann, 2010).

## **1.2 Mid-Atlantic Wine Consumption**

The wine consumption of the U.S. has been continually increasing since 1994 (Nichols, 2011), and has grown up to 330 million cases in 2010. From 2001 to 2012, the growth of the total consumption had outgrown the growth of per-capita consumption. More and more people had started to drink wine in the U.S. In 2010, the total volume of wine consumed overrides that of France. Nevertheless, the per capita consumption is still behind that of France. That also suggests that the U.S. wine market still has huge potential. Competition from within the U.S. and abroad for market share in the U.S. is intense. Sixty-one percent of wines consumed in the U.S. are produced in California (Marshall, Akoorie, Hamann, & Sinha, 2010) and imported wine shipments into the U.S. increased in 2011 by 4.9% compared to 2010 data (U.S. International Trade Association, 2011). Several countries including Italy, France, Chile, Spain, Argentina, and New Zealand reported gains in the U.S. market. In more recent years, groups of foreign wineries have joined forces to implement more concerted efforts to market their wine in the U.S. With another 10%

increase in consumption prediction for the U.S. between 2011 and 2015, a continued front of foreign winery groups that can targeting the U.S. markets is highly possible. The U.S market holds great promise for wine consumption for international companies and is a real opportunity and an equally compelling threat for smaller, independent local wineries (Lockshin, Spawton, & Macintosh, 1997).

## **2 LITERATURE REVIEW**

### **2.1 Model of Consumer Behavior**

Assael's (2005) model of consumer behavior exhibits different aspects of an individual which influence the consumer's final choice in the decision making process. A consumer's purchasing decision is influenced by their perceptions, attitudes, characteristics, lifestyle, and personality (Assael, 2005). Perceptions of risk have been identified by some researchers as the most influential factor in making wine buying decisions (Hall, Binney, & O'Mahony, 2004 2004). A wine consumer's level of knowledge and experience in purchasing wine can also affect their choice. (Mitchell & Greatedorex, 1989).

### **2.2 Demographic Characteristics Affects Wine Consuming Decision**

The demographic characteristics of consumers are considered to play a significant role in the wine consuming decision (Dodd, Laverie, Wilcox, & Duhan, 2005). Research has demonstrated that the number of information sources used by wine tourists vary based on the level of product involvement, the number of previous winery visits, and attitude (Dodd, 1995). A study about Australian wine purchasing and consumption has shown that the demographic characteristics of wine consumers such as their age, gender, education level, income, occupation and wine consumption habits are highly correlated with their wine purchasing behavior and preferences (Johnson & Bastian, 2007). The research results from Johnson and Bastian indicates a) 50.8% female and 49.2% male from their wine consumers' demographic data, b) the average age of respondents was younger than the general population of Australia, c) the education level of respondents was also higher than the

general, d) 72% of the respondents reported household incomes of AUD\$100,000 per year or less. The median household income of Australia is AUD\$91,624 in 2007 (Johnson & Bastian, 2007).

### **2.3 Marital Status Affects Alcohol Consumption**

People of different marital status have differences in their alcohol consumption. The alcohol consumption either increases or decreases as people's marital status varies (Power, Rodgers, & Hope, 1999). In Power, Rodgers and Hope's research, they found that the alcohol consumption was greater in men than women at the same age. Divorced people are most likely to have a heavy drinking problem and those married have the lowest. Single and those who have remarried are in the middle (Power et al., 1999). Men who drink more than 35 units (1 unit equivalent to 1 glass of wine) per week are considered to have a heavy drinking problem, and 20 units for women. The authors also point out that the increase in drinking associated with divorce is a short-term effect. However, alcohol-related health problems may occur in the immediate period around divorce (Power et al., 1999).

### **3 METHODS**

The main research question is that, in the Mid-Atlantic region, what kind of people are more likely to purchase locally produced wine, and how to target this market segment? The question can be defined into several small objectives. Identify the demographics and behaviors that describe Mid-Atlantic wine buyers. Identify wine consumers' preferences on different wine attributes. Segment wine consumers into several groups, and study the characteristics of each group. Understand how consumers learn about wine and the role of social media.

The data used in this study is from an online survey performed by the Penn State University in 2009. This survey helped us to quantify consumer wine purchases and preferred varieties, identify the demographics and behaviors that describe Mid-Atlantic wine buyers.

First we did descriptive statistics to describe the finds from each survey question, as well as some bi-variate analysis (Put two or more variables together to draw more insights). Then, we identified the characteristics and attributes of the most likely local wine buyers by doing Logistic Regression. After that, we looked into consumer segmentation by employing Cluster Analysis. More discussions were made on how to maintain business with current buyers, as well as how to target other less likely buyers given an understanding of their preferences.

## 4 DESCRIPTIVE STATISTICS

The survey was originally conducted by Penn State University in 2009. 1246 qualified wine consumers participated in this survey online. 41 questions were asked regarding demographics, drinking behaviors and preferences. Please see the Appendix for the full survey.

### 4.1 Demographics

#### *State, Gender*

From these 1246 survey respondents, 597 are from New York State, 407 are from Pennsylvania, and the remaining 242 respondents are from New Jersey, as shown in Figure 1. 63% of the total respondents are female as shown in Figure 2. In order to make sure that all the responses are unbiased, we eliminated respondents who are a member of the wine industry or trade such as a retailer, distributor or wine grape grower. Also, we want to make sure that our respondents are aged between 21 to 65 years old, which is the target market of local wineries.

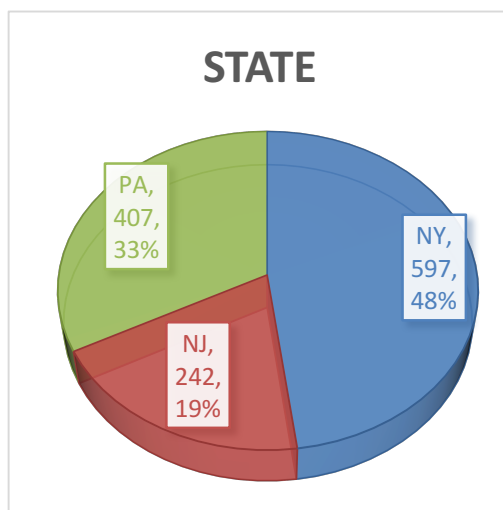


Figure 1. State where respondent resides

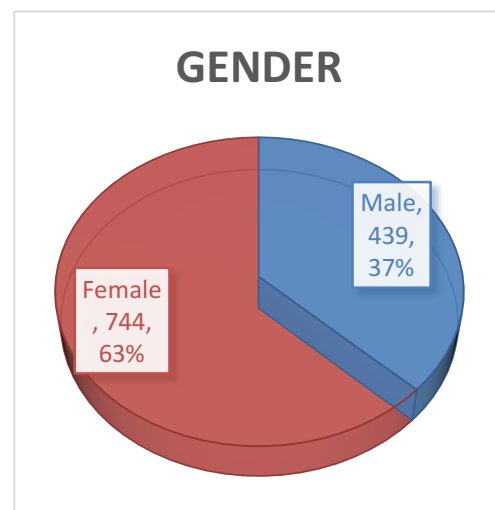
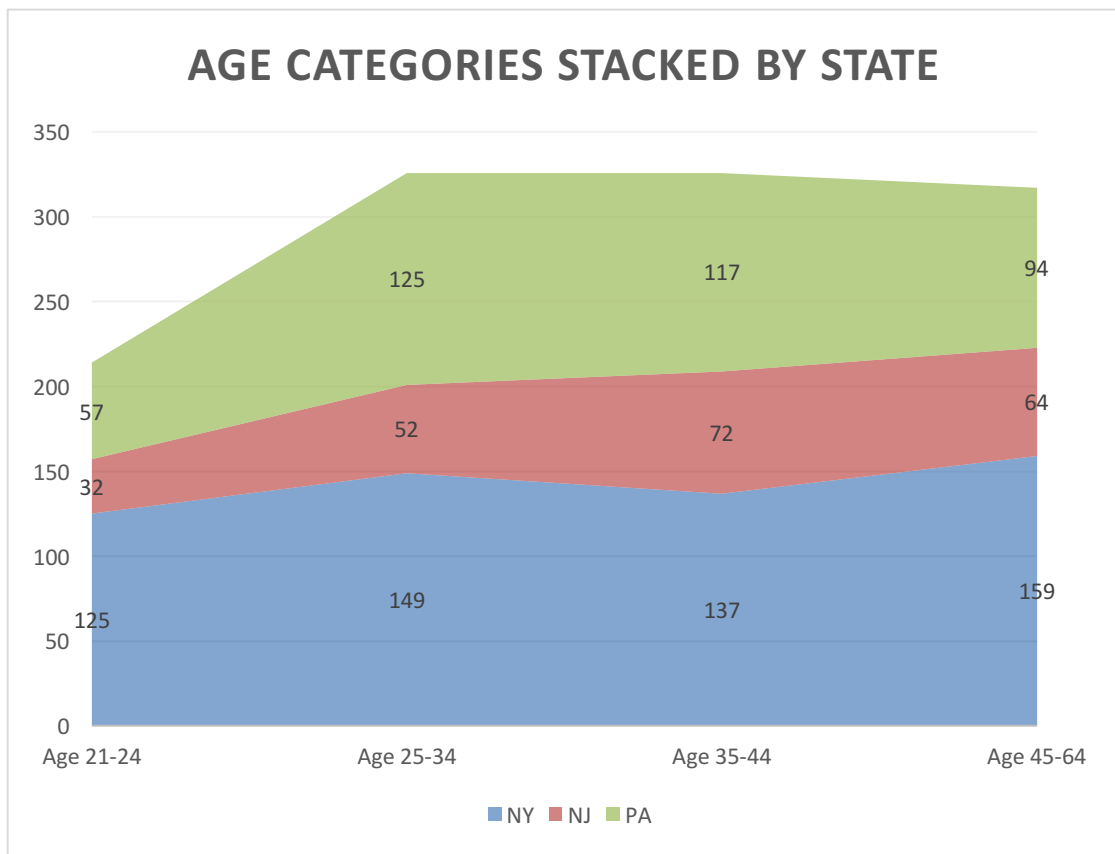


Figure 2. Gender of survey respondents



### *Age Category*

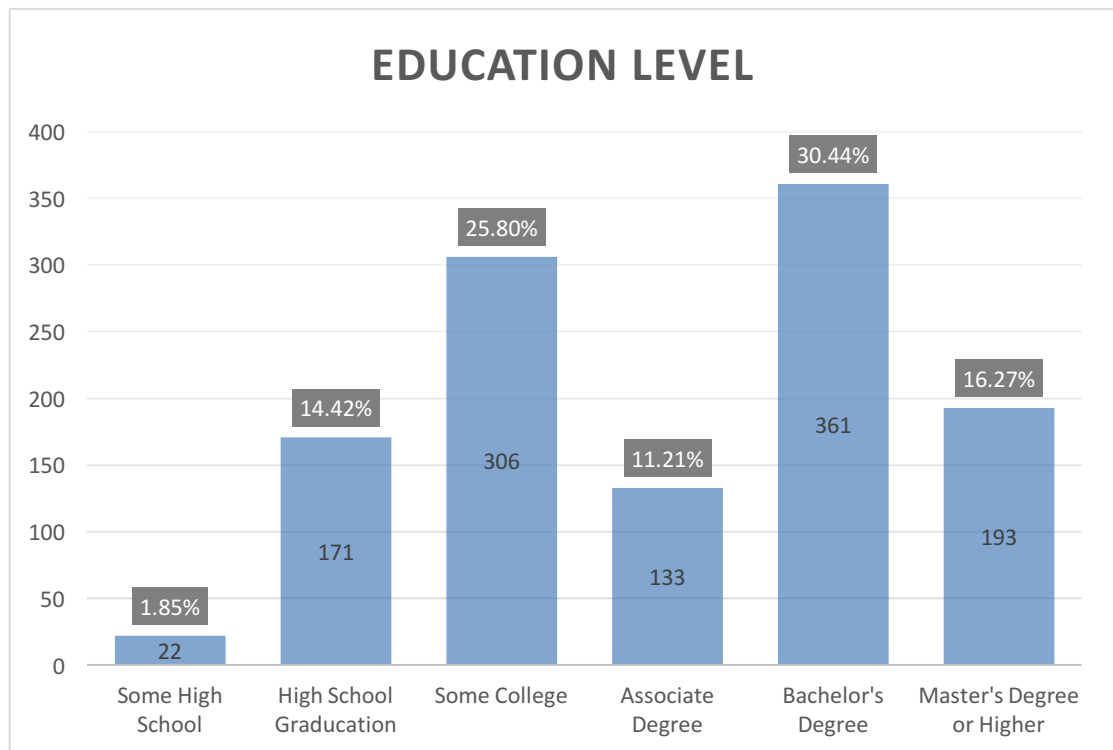
Our respondents are all aged between 21 – 64 years old. They were regrouped in to four different categories. Those categories are 21-24 years old, 25-34 years old, 35-44 years old and 45-64 years old. Figure 3 shows the age categories stacked by state. There are 214 respondents aged between 21-24; 326 respondents aged between 25-34; 326 respondents aged between 35-44; and 317 respondents aged between 45-64.



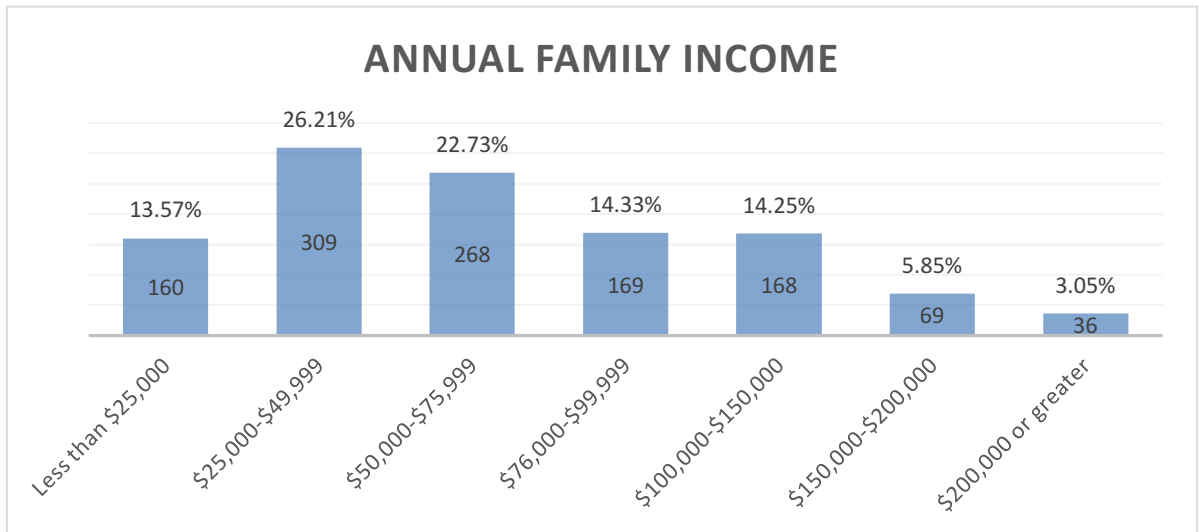
*Figure 3. Age categories stacked by state*

### *Education, Family Income*

As shown in Figure 4, most of our survey respondents are high school graduates or have higher education levels. Almost 50% of the total respondents have a Bachelor's degree or higher. 25.8% of total respondents indicated that their education level is some college. This may be caused by the number of 21-23 years old in our sample, who were currently attending college when they participated in this survey. Figure 5 shows the annual family income of our respondents. Most respondents fell into the \$25,000 to \$49,999 and \$50,000 to \$75,999 categories.



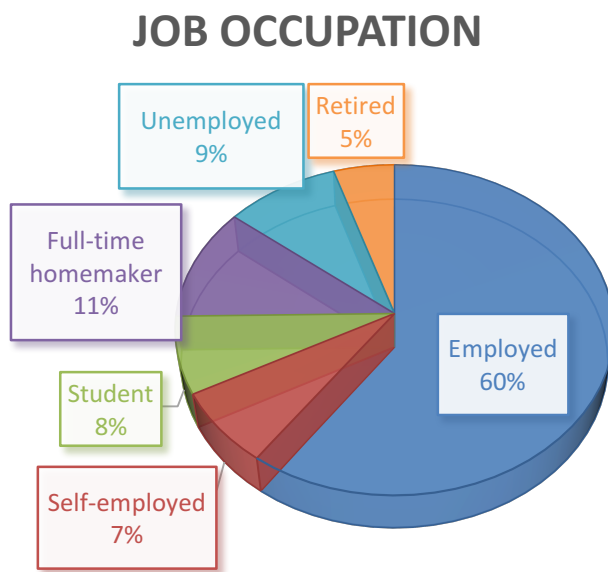
*Figure 4. Education Level*



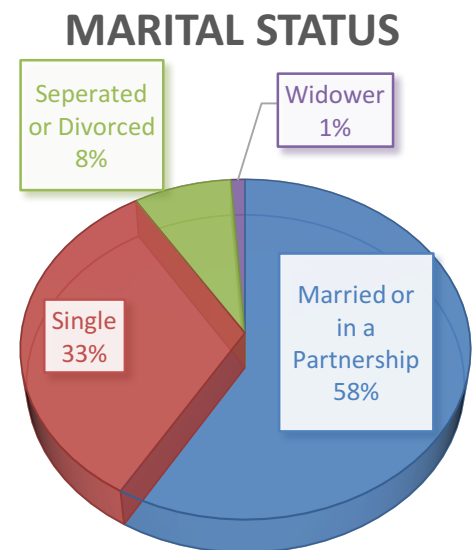
*Figure 5. Annual Family Income*

### *Job Occupation & Marital Status*

As shown in Figure 6, 60% of our respondents are employed by someone else; 7% is self-employed; 8% is student; 11% is full-time homemaker; 9% is unemployed; and 5% is retired. As shown in Figure 7, 58% of our respondents are married or in a partnership; 33% is single; 8% is separated or divorced, and 1% is widower.



*Figure 6. Job Occupations*



*Figure 7. Marital Status*

## 4.2 Wine Consuming Behavior

Figure 8 shows that the responses to the question “How often do you drink wine during an average year?” The percentages shown are the percentages of total observations. The options are from low drinking frequency (a few times a year) to high drinking frequency(Daily). Only about 7% people are intensive wine drinkers who drink wine daily. About 68% people are moderate wine drinkers who drink wine more than once a month. The remaining 25% people are leisure wine drinkers. Besides these, outliers who drinks wine more than 31 days a month (which is impossible) are dropped from our analyses. Figure 9 shows the same data, which is broken down by gender(1=Male, 2=Female).

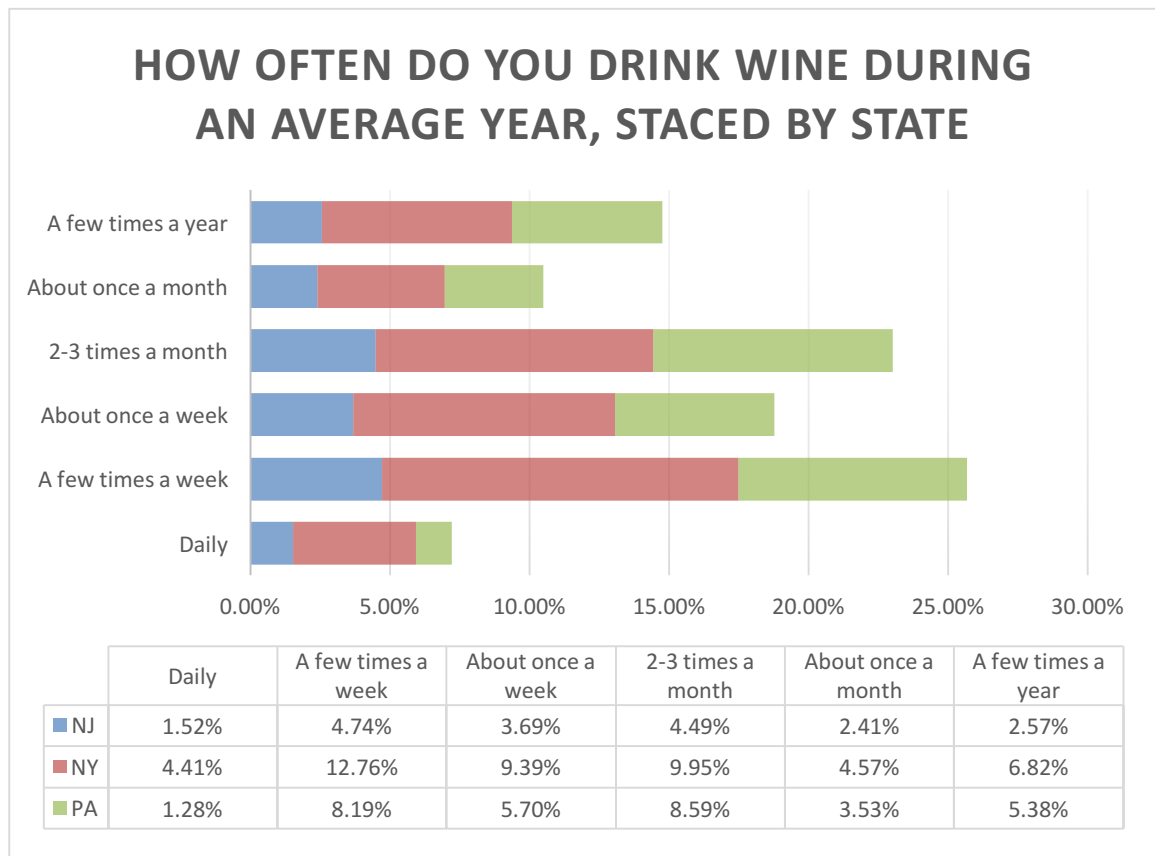


Figure 8. Q1a How often do you drink wine during an average year? Percentage of total respondents

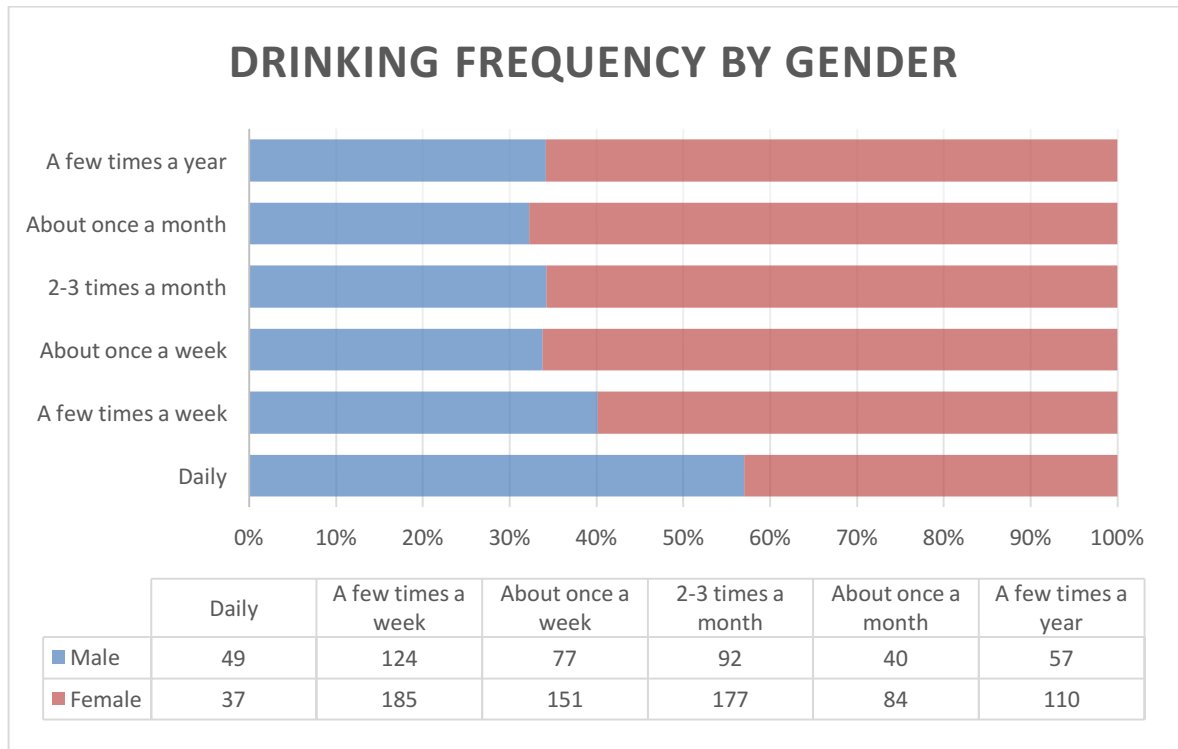


Figure 9. Drinking frequency by gender. Counts are number of respondents.

Now we have answered the question of drinking frequency, but what about the quantities? A person may not drink often, but drink a lot every time he drinks. In order to address this question, we asked how many days they drink wine during a month and how many glasses they consumed on the days they consume wine. An average glasses of wine is 150ml.

*Monthly Consumption in liters = (Drinking days during a month) \* (number of glasses of wine consumed on the days) \* 150ml / 1000*

The above function is how the monthly consumption quantity was computed.

Figure 11 below shows the Monthly Consumption in liters. It was broken down by state (y axis) and age categories (stacked color bars). The exact consumption number doesn't make too much sense here. However, the relationship across states and age categories does provide insights. We had 1246 respondents. 597 are from New York State, 407 are from Pennsylvania, and the remaining 242 respondents are from New Jersey (NJ). New York (NY) residents drink about 2 times and 2.5 times more than Pennsylvania (PA) and New Jersey residents, respectively. Figure 3 shows that PA has a smaller sampling weight in the 21-24 years old category than New York state. But their consumption is much more than the same category from NJ and NY. Younger drinkers in PA consume a big share of the total consumption of PA. In NY, 25-34 years olds drink more than other age groups.

### **4.3 Wine Purchasing Behavior**

#### *Purchasing Frequency*

As shown in Figure 10, only 28 out of 1246 people purchase wine daily. 71 out of 1246 people purchase wine a few times a week. 174 out of 1246 people purchase wine about once a week. 269 out of 1246 people purchase wine two to three times a month. 279 out of 1246 people purchase wine about once a month. 425 out of 1246 people purchase wine a few times a year. 100% respondents said they purchase wine more than once a month. Most of them purchase wine on a weekly or monthly basis. As shown in Figure 10, there is not much difference in purchasing frequency across states. As shown in Figure 12, most of the daily wine buyers are male. As shown in Figure 13, the 4<sup>th</sup> and 5<sup>th</sup> age groups, who are more than 34 years old are less likely to be daily wine buyers. So we concluded that males between 21-34 years old are more likely to be high frequent wine buyers.

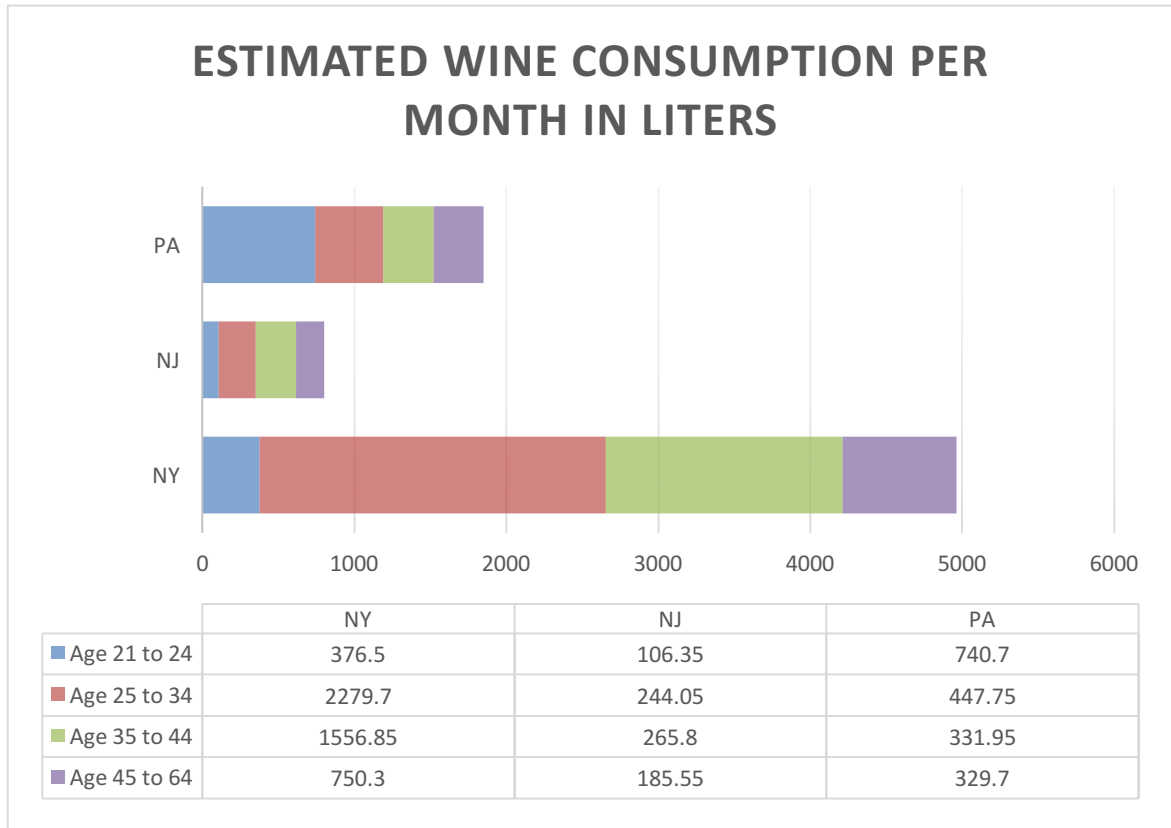


Figure 11. Estimated Wine consumption of survey respondents per month in liters

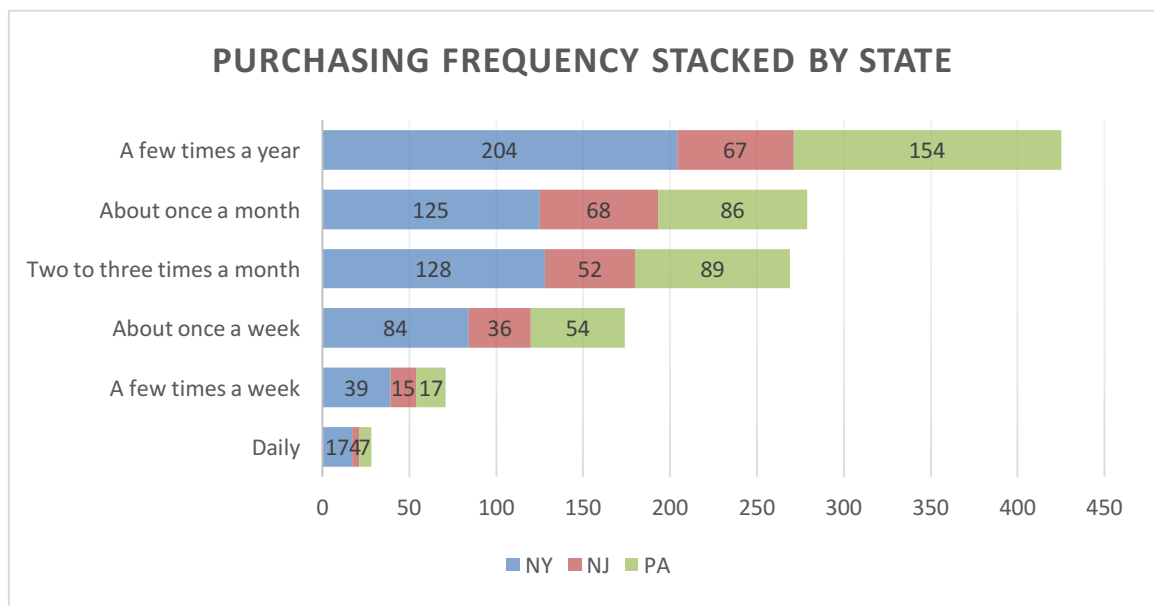


Figure 10. Purchasing frequency stacked by state



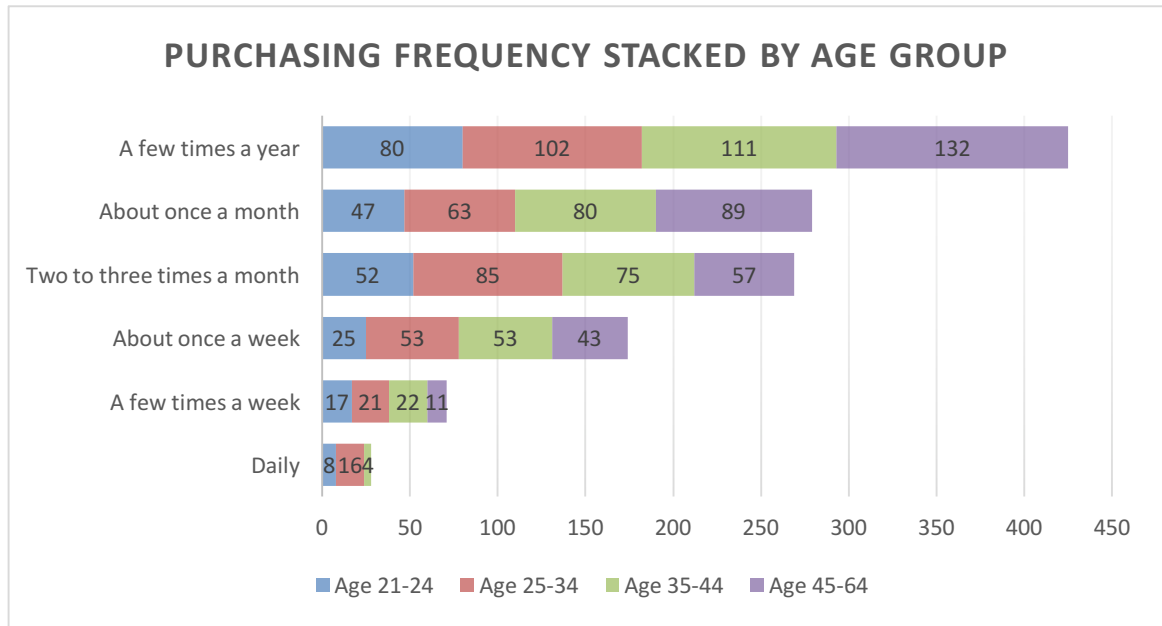


Figure 12. Purchasing frequency stacked by age group

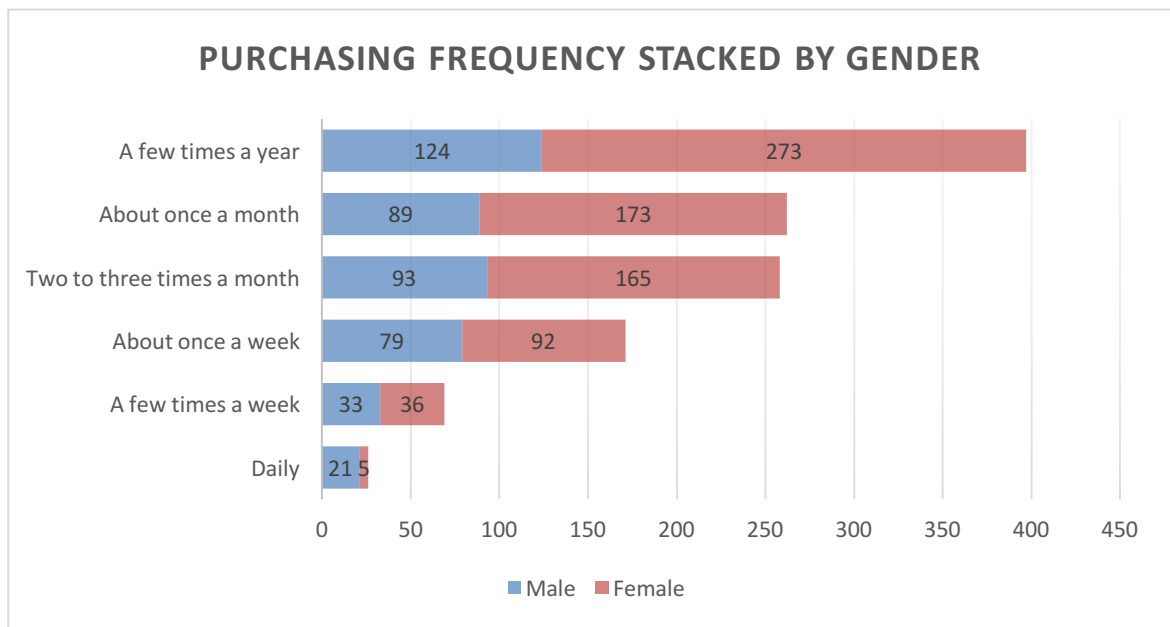
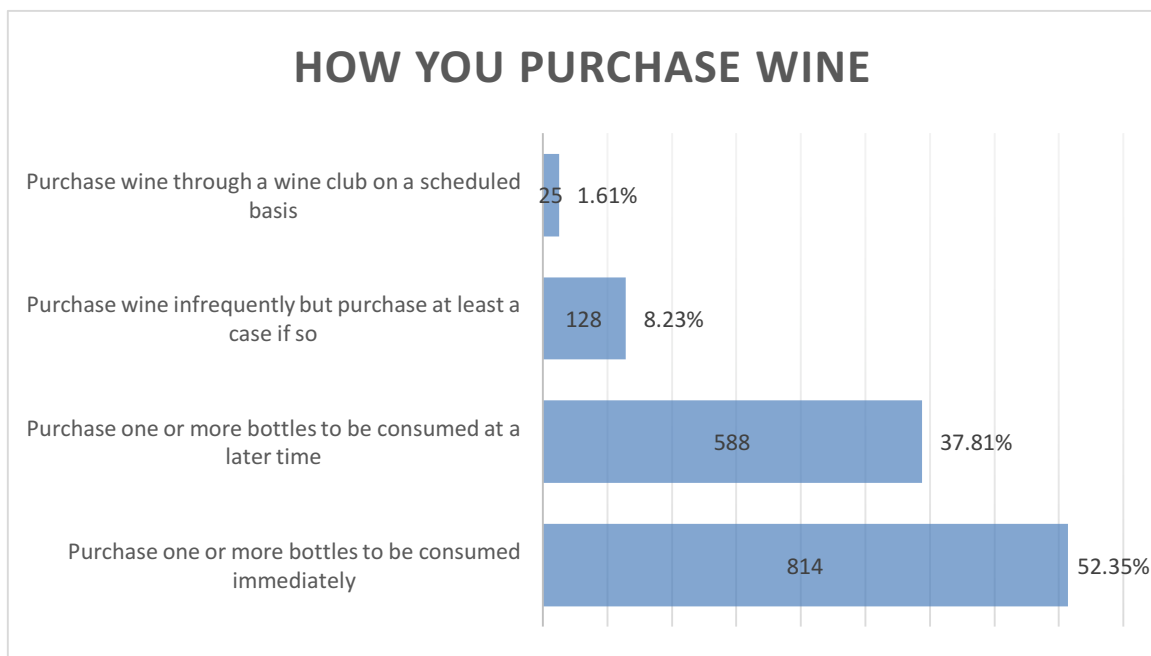


Figure 13. Purchasing frequency stacked by gender

### *How you purchase wine*

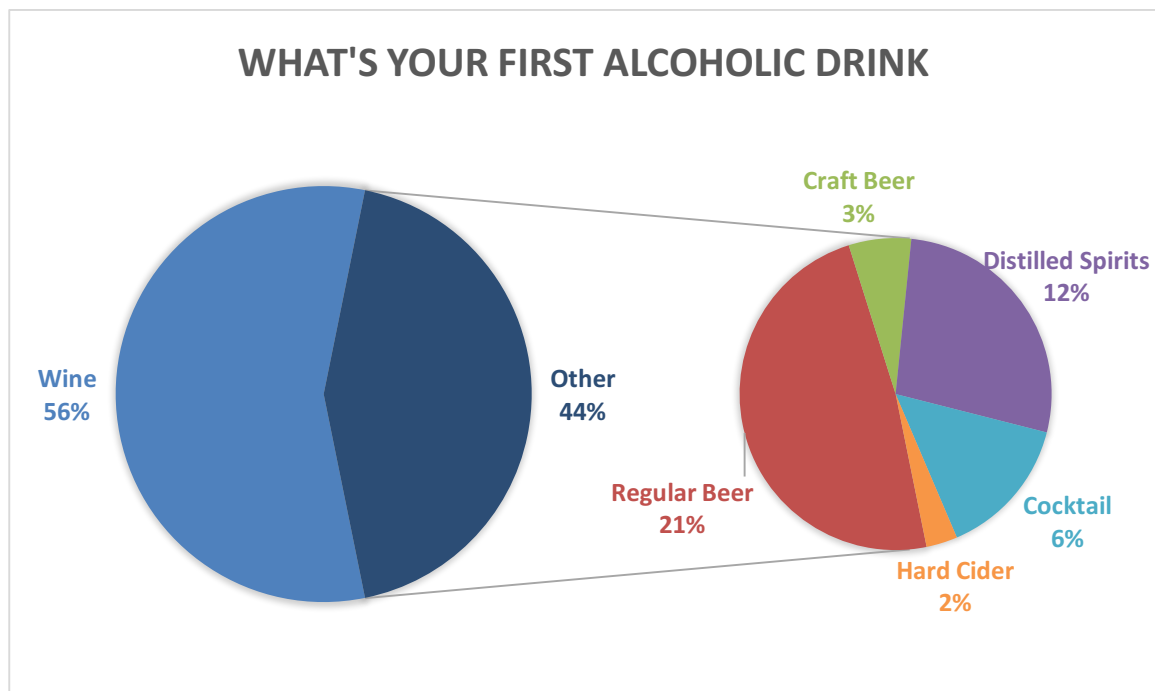
As shown in Figure 14, about 65% of respondents purchase one or more bottles to be consumed immediately. About 50% of respondents purchase one or more bottles to be consumed a later time. About 10% of respondents purchase wine infrequently but if they do, they purchase at least a case. Very few respondents purchase wine through a wine club on a scheduled basis. Most people purchase one or more bottles of wine each time for immediate or later need.



*Figure 14. How you purchase wine*

*What's your first alcoholic drink?*

As shown in Figure 15, 33.5% people's first alcoholic drink was wine, while 56.6% was not. The remaining 10% don't know or don't remember. In later Logistic Regression, we will produce a dummy variable that takes YES as 1, and other responses as 0 so we can exam the effect of wine as a first drink. In addition, in the people whose first drink was not wine, about 37% people's first drink was beer, about 20% was hard liquor such as whisky and rum, and about 12% was cocktails.



*Figure 15. What's your first alcoholic drink?*

#### 4.4 Different wine for everyday consumption and special occasions.

As shown in Figure 16, 72% respondents agree that they purchase different wines for everyday consumption and for consuming on special occasions or entertaining, in terms of price, varietal, container type, and/or other characteristic.

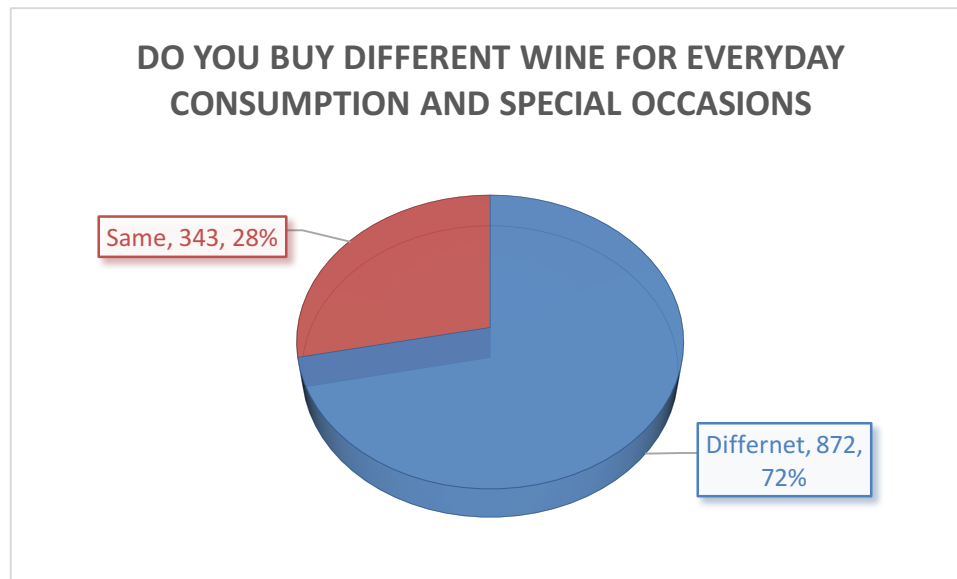


Figure 16. Do you buy different wine for everyday consumption and special occasions?

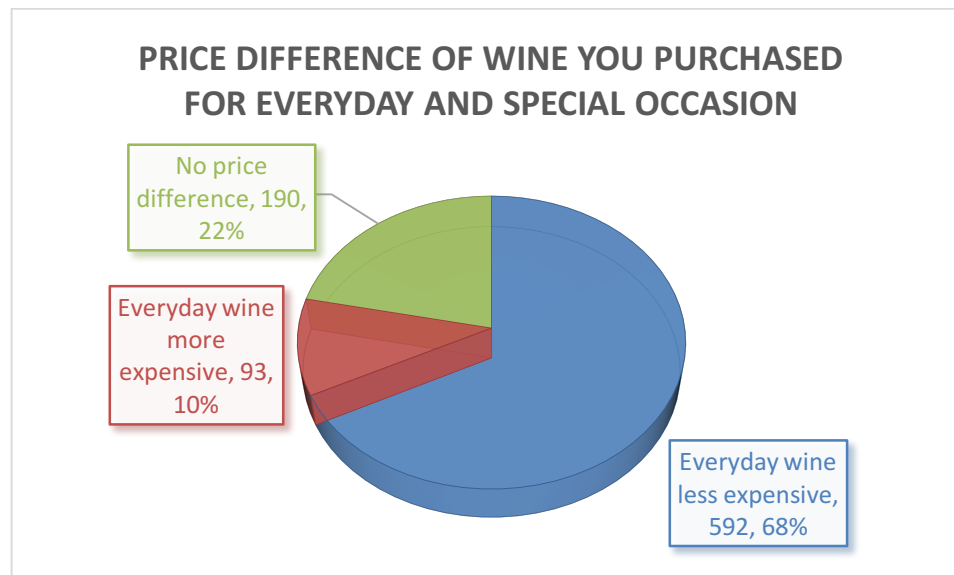


Figure 17. Price difference of wine you purchased for everyday and special occasion

### *Price Difference*

As shown in Figure 17, 68% of respondents who indicate that they buy different wine for everyday consumption and special occasions, agree that they pay more for special occasion wines. Only 10% indicate that they pay more for everyday wines. 22% indicate that there is no price difference. In order to understand more about people's willingness to pay for everyday consumption and special occasions, we asked survey respondents to indicate the ranges that correspond to what they pay for everyday wines and special occasion wines, as shown in Figure 18. The responses have been broken down by gender and state.



*Figure 18. Price ranges of wine you pay for everyday and special occasions*

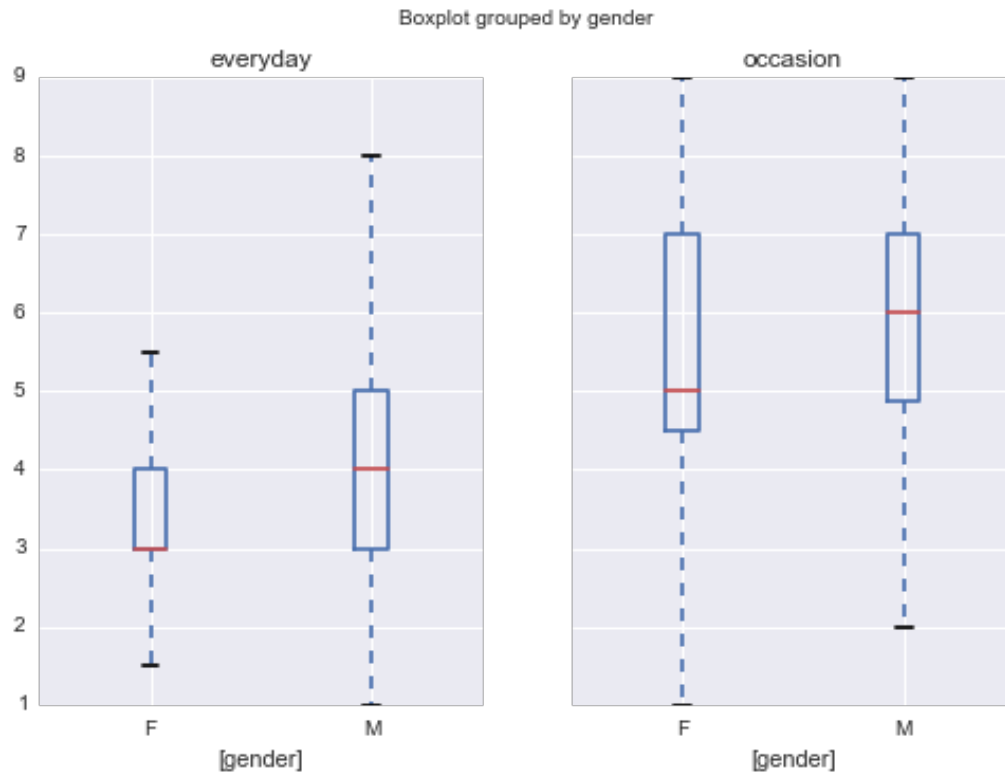


Figure 19. Price range break down by gender

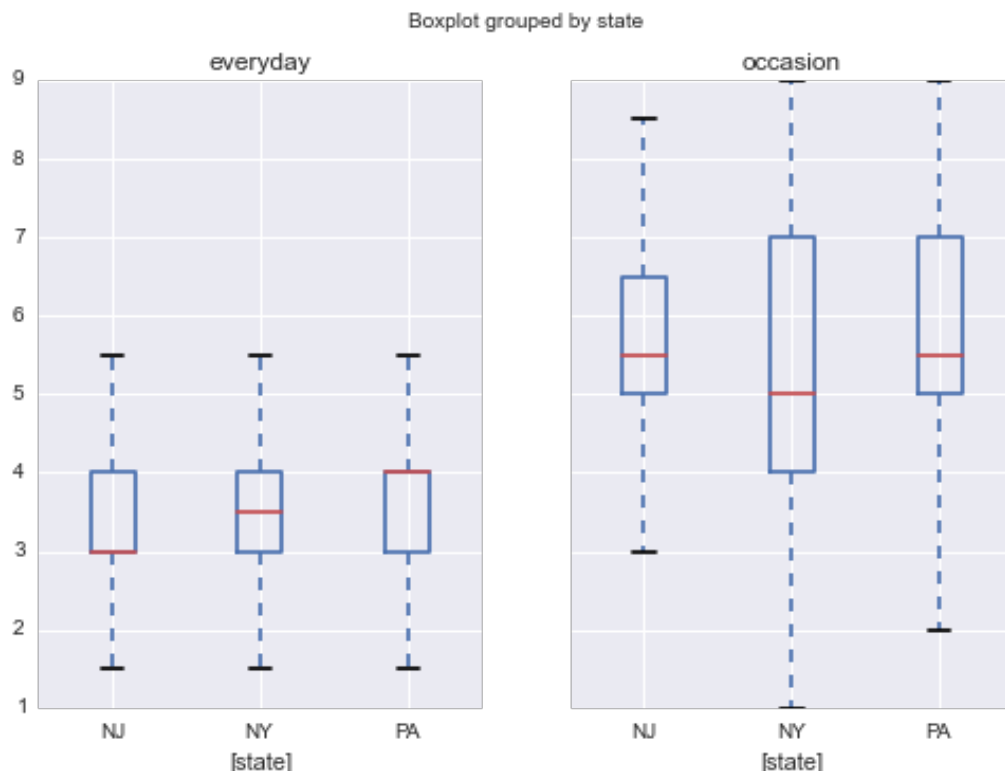


Figure 20. Price range break down by state

### *Break down by gender*

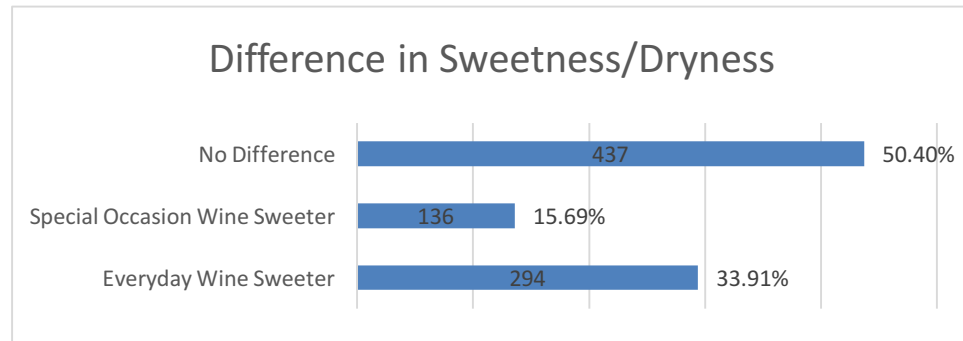
In Figure 19, data was broken down by gender. Both male and female tend to pay more for special occasion wines than everyday wine. However, generally females pay less on wine than males do. Most females are willing to pay \$8 to \$10.99 for everyday wines and \$15 to \$20 for special occasion wines. Most males are willing to pay \$11 to \$15 for everyday wines and \$20 to \$25 for special occasion wines. Males are willing to pay about 3.5 dollars more for a bottle of wine than females in general. Both males and females are willing to pay about 8 dollars more for special occasions than for everyday consumption.

### *Break down by state*

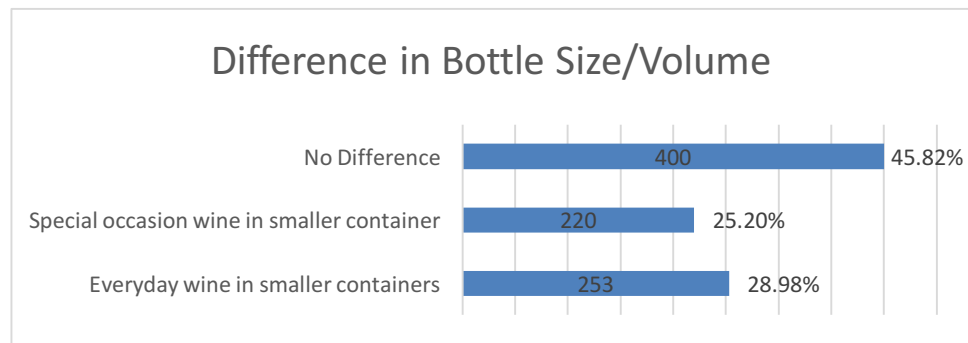
In Figure 20, data was broken down by state. As you can see, the price range for everyday wines is between \$8 to \$15. Taking New York residents as a reference group, New Jersey residents tend to spend slightly less whereas Pennsylvania residents tend to spend slightly more. On the right hand of Figure 20, the price range for special occasion wines is between \$15 to \$25. New York residents tend to spend slightly less for special occasion wine than residents from the other two states.

### *Other different attributes for everyday wine and special occasion wine*

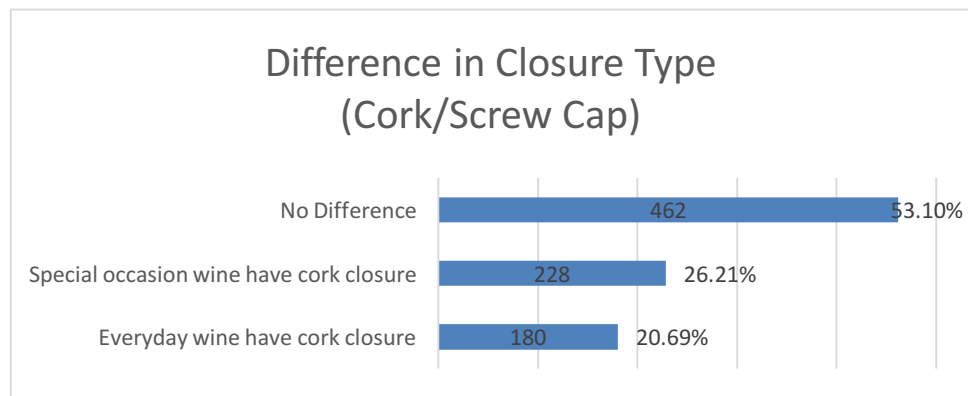
We also looked into other different attributes for everyday wine and special occasion wine. We already know most people are willing to pay a higher price for everyday wine over special occasion wine, but what about other attributes? As shown in Figure 21 to Figure 24, about 50% of respondents indicates that attributes such as sweetness/dryness, bottle size/volume, closure type and packaging materials doesn't affect their decisions on everyday wine or special occasion wine.



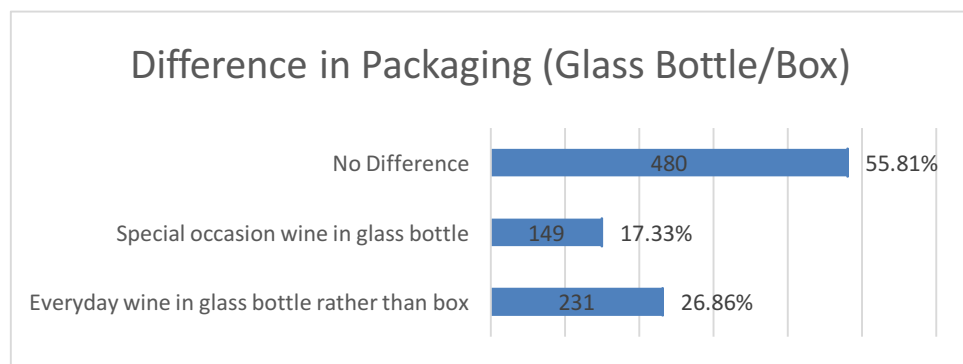
*Figure 21 Difference in sweetness/dryness*



*Figure 23 Difference in bottle size*



*Figure 22 Difference in closure type*

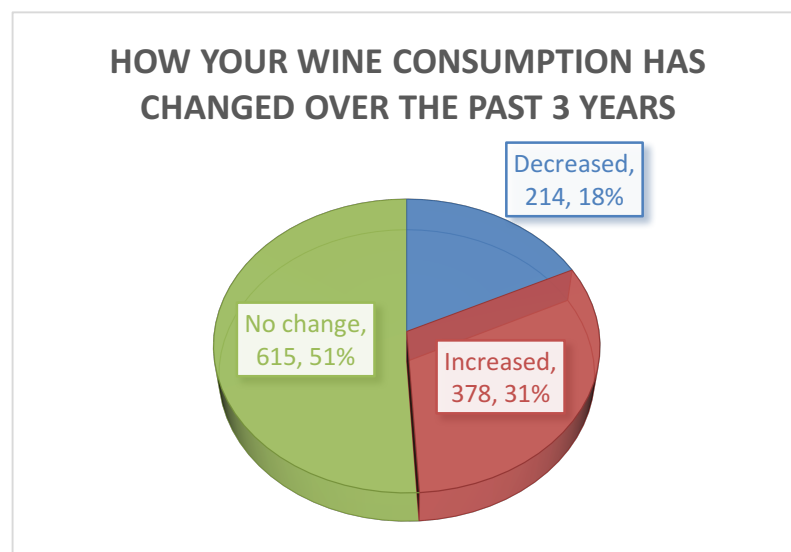


*Figure 24 Difference in Packaging*



#### 4.5 How wine consumption has changed over the past three year.

As shown in Figure 25, 31% of the total respondents indicated that their wine consumption has increased over the past three year; 51% indicated that their wine consumption has not changed during the past three years; 18% indicated that their wine consumption decreased over the past three years. The increased group has about two times more people than the decreased group. There is without a doubt, an upward trend in the Mid-Atlantic wine market.



*Figure 25 How your wine consumption has changed over the past 3 years*

*Reasons of wine consumption change*

First, let's take a look at the reasons why it has decreased. As shown in Figure 26, the top two reasons are price and spending money on other things. Spending money on other things can have two different interpretations. If someone cuts spending on wine because he spends more on other things, it could be that he has more important things to spend his money on. Or it could be that he switched to a new hobby other than drinking wine. The third reason is about weight gain. The fourth reason is about health concerns.

As shown in Figure 27, the top one reason is that one became more interested in drinking wine than drinking other beverages. The second reason is about health benefits of drinking wine. It is very interesting that health concern is one of the most important reasons for both decreasing and increasing in wine consumption. The third reason is that people learned more about wine and became more interesting in drinking wine. Weight control is also one reason for an increase in consumption. 6% respondents indicated that they increased wine consumption since they learned moderate wine drinking helps weight control. As you can see, the same reason can affect wine consumption in many different directions. These are very important things to note for a marketing manager to formulate their marketing messages.

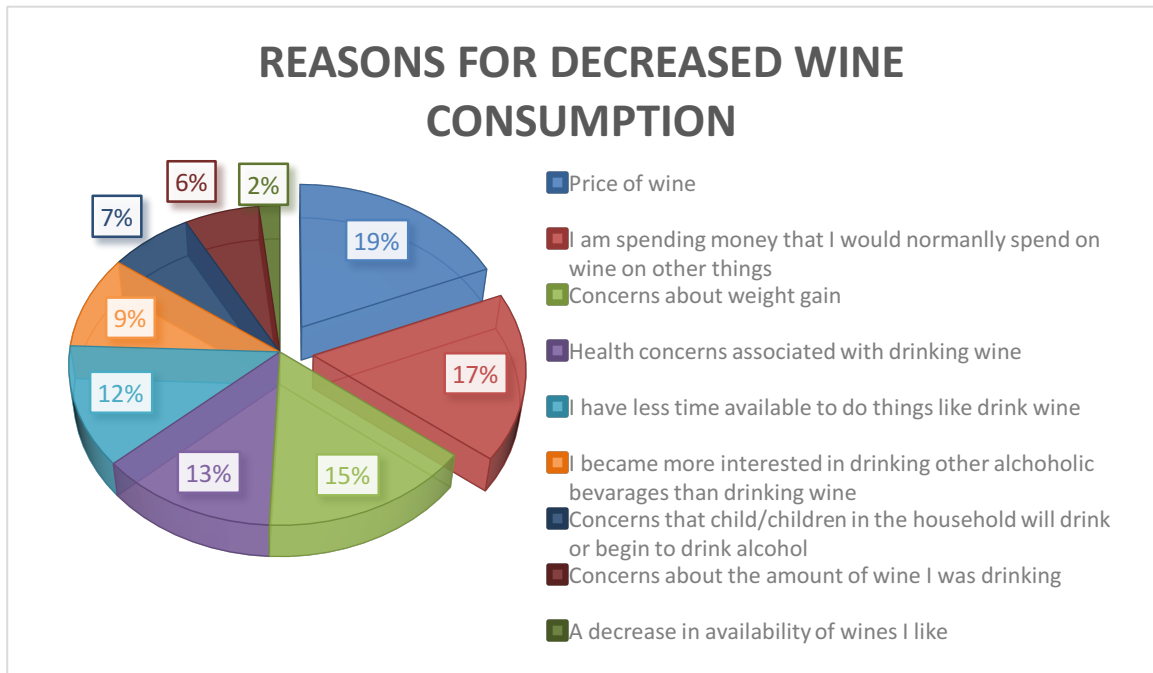


Figure 26. Reasons for decreased wine consumption

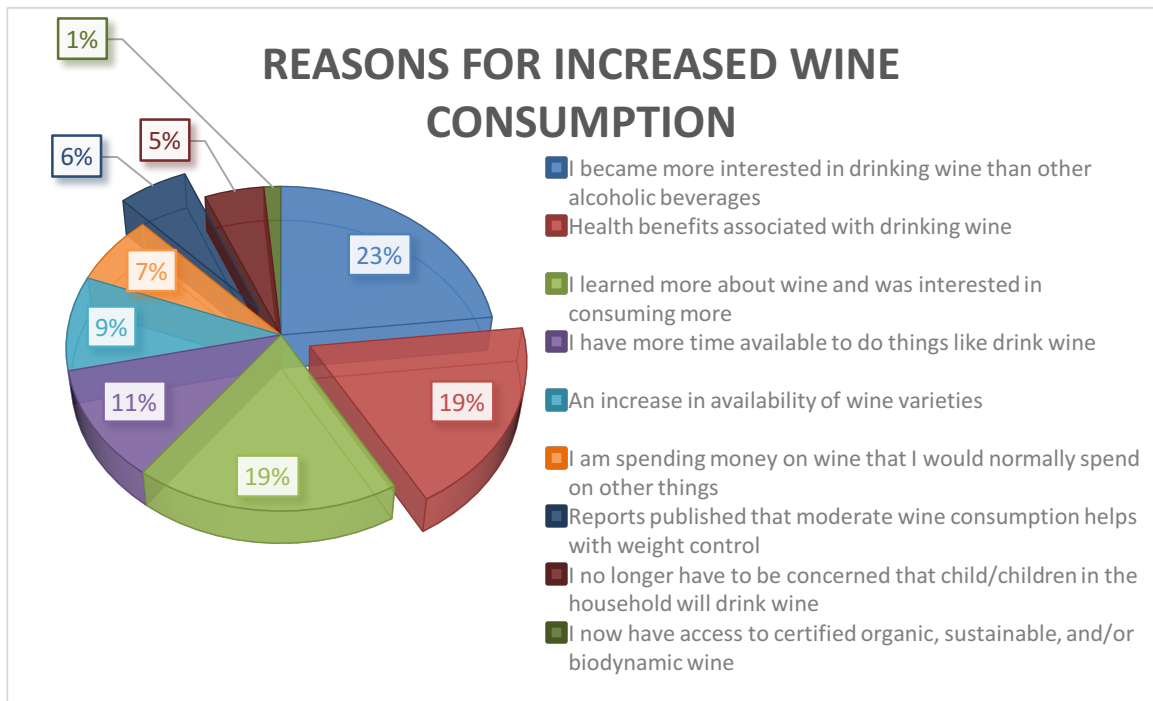


Figure 27. Reasons for increased wine consumption

#### **4.6 Popularity of wine from different wine regions**

As shown in Figure 28, the top four most visited wine regions are New York, Pennsylvania, California and New Jersey. However, while our respondents are all from the Mid-Atlantic area, only about 10% respondents indicated that they had visited a winery in the New York region. Even less respondents indicated that they had visited a winery in Pennsylvania or New Jersey.

Figure 29 is about whether people have drunk wine from the region, whereas Figure 31 is about whether people have purchased wine from the region. If people answered yes to both questions, we consider this wine region as a popular wine region. From the data, New York, California and France are the most popular wine regions for mid-Atlantic wine consumers. New Jersey and Pennsylvania are relatively not very popular compared to New York, California and France. The most unpopular wine regions are South Africa, New Zealand, Austria and Canada. However, in Figure 30, respondents also indicated that they are interested in purchasing and drinking wine from these unpopular regions.



Figure 28 Have visited wineries in the region

Figure 29 Have drunk wine from the region

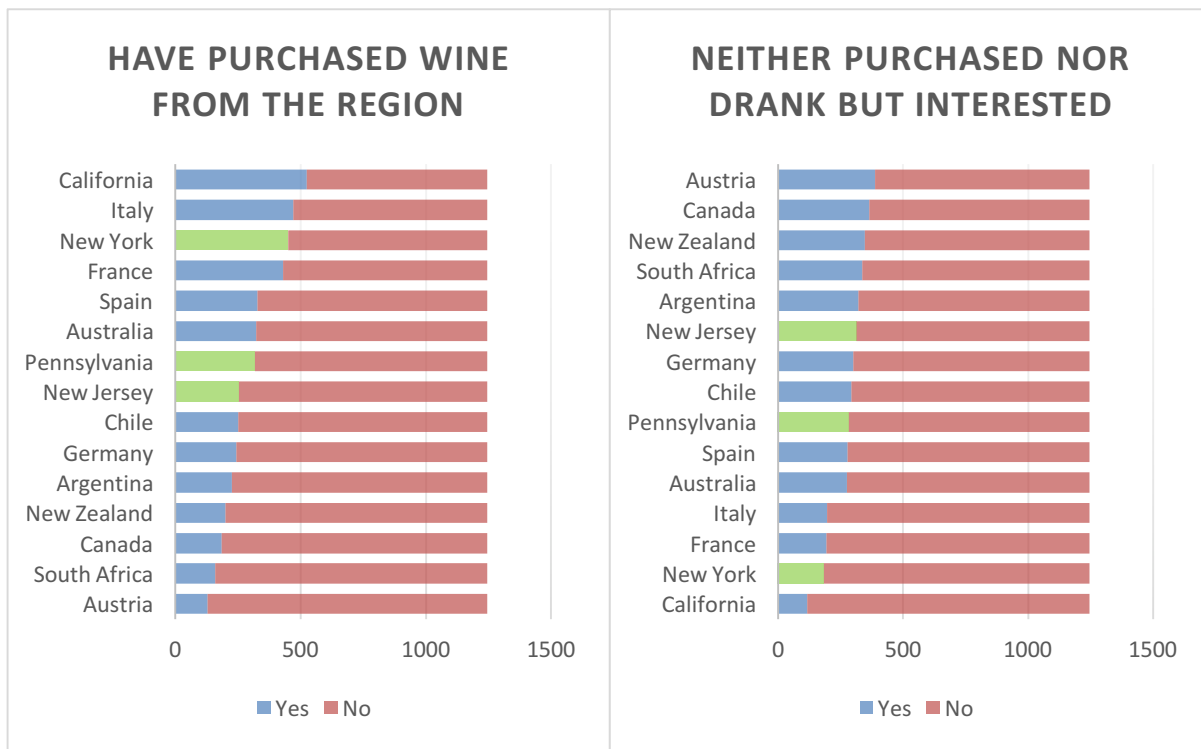
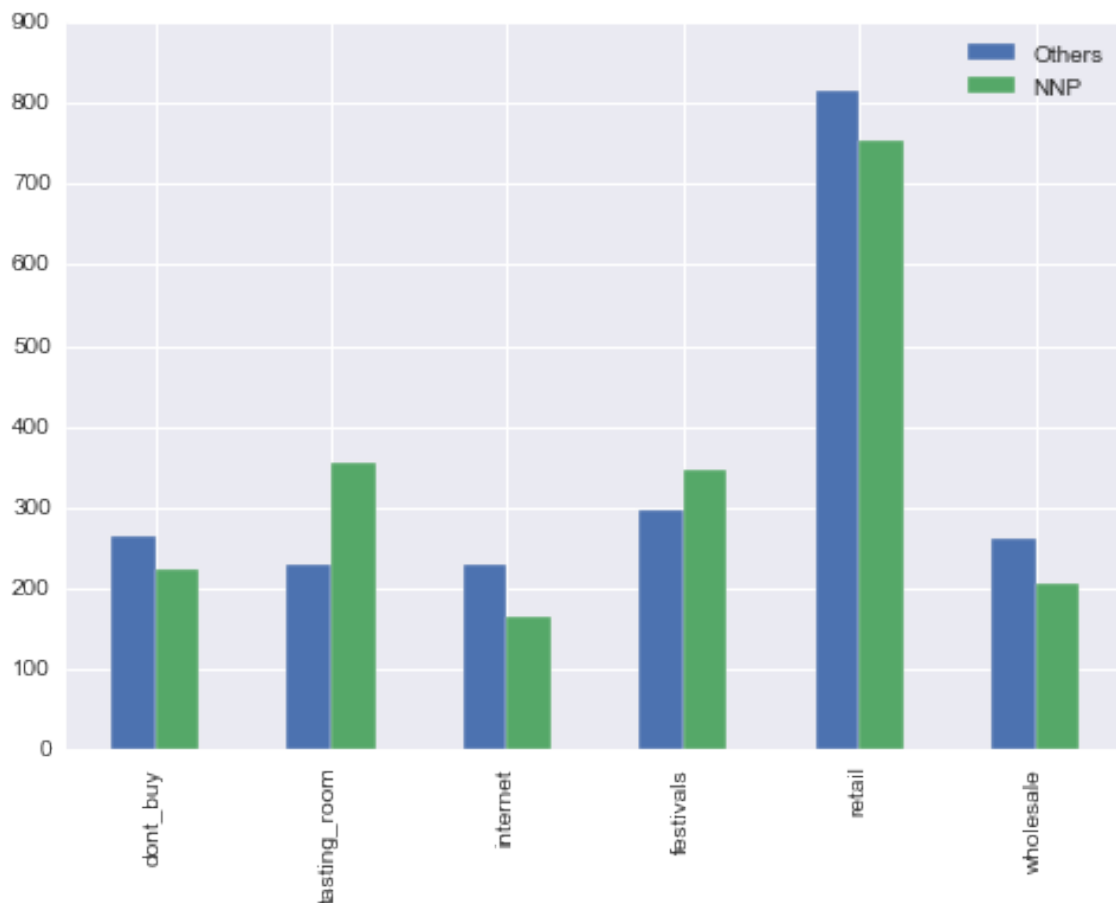


Figure 31 Have purchased wine from the region

Figure 30 Neither purchased nor drank but interested

#### 4.7 Where do people purchase wine from?

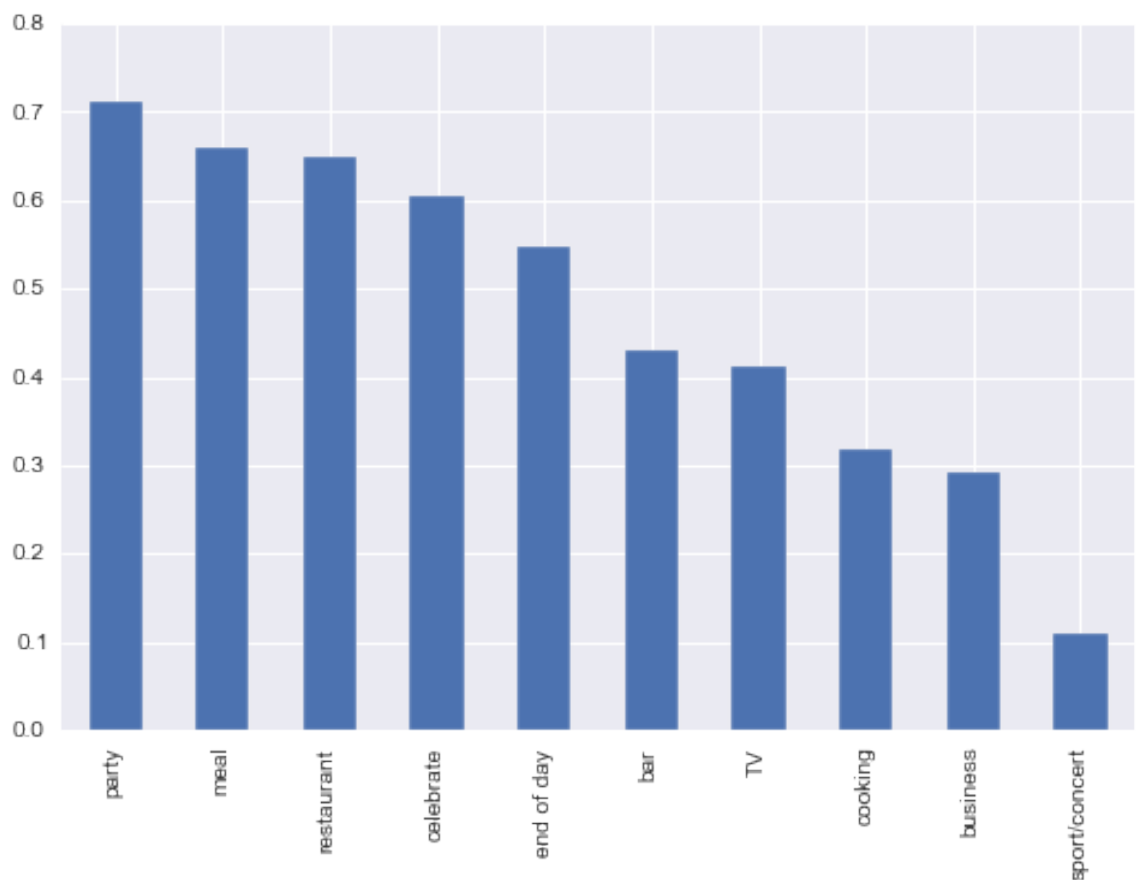
As shown in Figure 32, we asked survey participants at which outlet they purchase wine from. Green bars are responses of wine purchased from New York, New Jersey and Pennsylvania. Blue bars are responses of wine purchased from other wine regions. You can see that the most popular outlet is still retail liquor stores. Tasting rooms and festivals are also important buying channels for mid-Atlantic wine region compared to other wine regions. Mid-Atlantic wineries could put more effort on wine tasting events and wine festivals when they develop their marketing strategies for local market.



*Figure 32. At which outlet do people purchase wine from?  
(Green=wine from NJ, NY and PA, Blue=wine from other regions)*

#### 4.8 Wine consuming occasions

We asked survey participants at which occasions they consume wine. Y-axis are ratio of total respondents, 0.5 means 50% of total respondents chose this option. As shown in Figure 33, five occasions have received over 50% votes. Over 70% of total respondents indicated that they consume wine when at a party or gathering with family or friends. About 65% indicated that they consume wine during meals. About 65% indicated that they consume wine when dining out at a restaurant. About 60% indicated that they consume wine when celebrating holidays or other special occasions. About 55% indicated that they consume wine at the end of the day to relax.



*Figure 33. Wine consuming occasions (y-axis are ratio of total respondents, 0.5 means 50% of total respondents chose this option)*

#### 4.9 Drinking frequency for different wine varietal

As shown in Figure 34, the top three popular varieties are 1<sup>st</sup> Chardonnay, 2<sup>nd</sup> Pinot, 3<sup>rd</sup> Merlot. Chardonnay is a green-skinned grape variety used to make white wine. Pinot is a red wine grape variety. Merlot is a dark blue-colored wine grape variety that is used as both a blending grape and for varietal wines. The top three least popular varieties are 1<sup>st</sup> Traminette, 2<sup>nd</sup> Chambourcin, 3<sup>rd</sup> Vidal Blanc. In the top three popular varieties, more people consume Merlot as everyday wine than Chardonnay and Pinot.

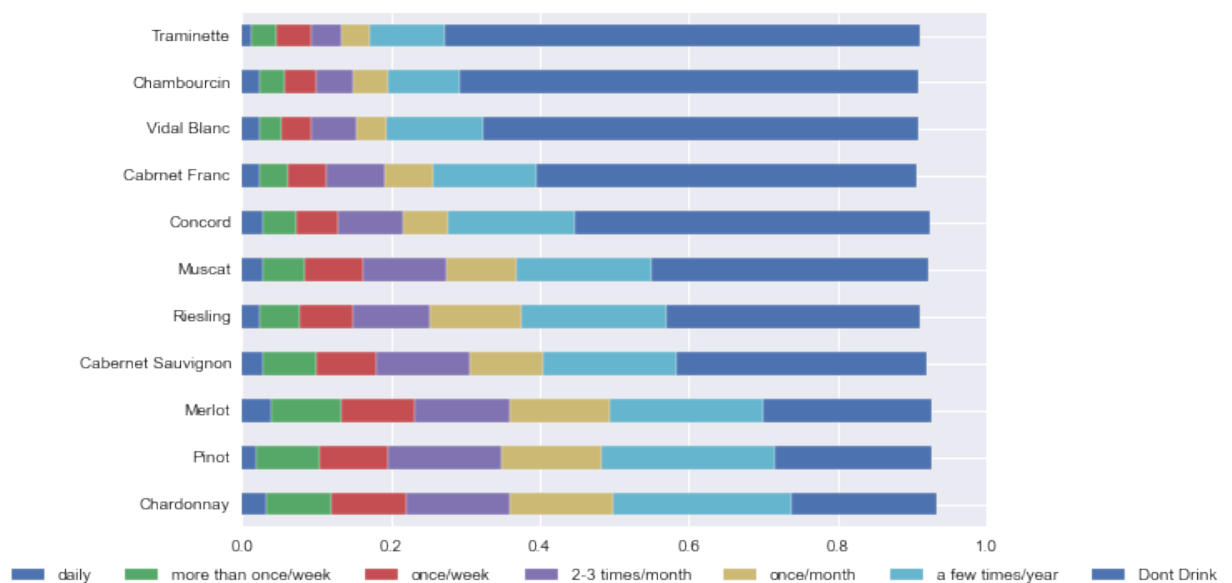


Figure 34 Drinking frequency for different wine varietal (x-axis is the ratio of total respondents)



#### 4.10 Monthly spending on wine purchasing

As shown in Figure 36, the average spending on wine purchasing is 75.9 dollars per month. However, there are three main spending groups. They are groups spending around 20 dollars, 50 dollars and 100 dollars per month. In Figure 35, we grouped the data by state and gender. No significant differences were found across gender or states. The lower band of Pennsylvania is slightly lower than New Jersey and New York. Females indicated that they spend less on wine than males do.

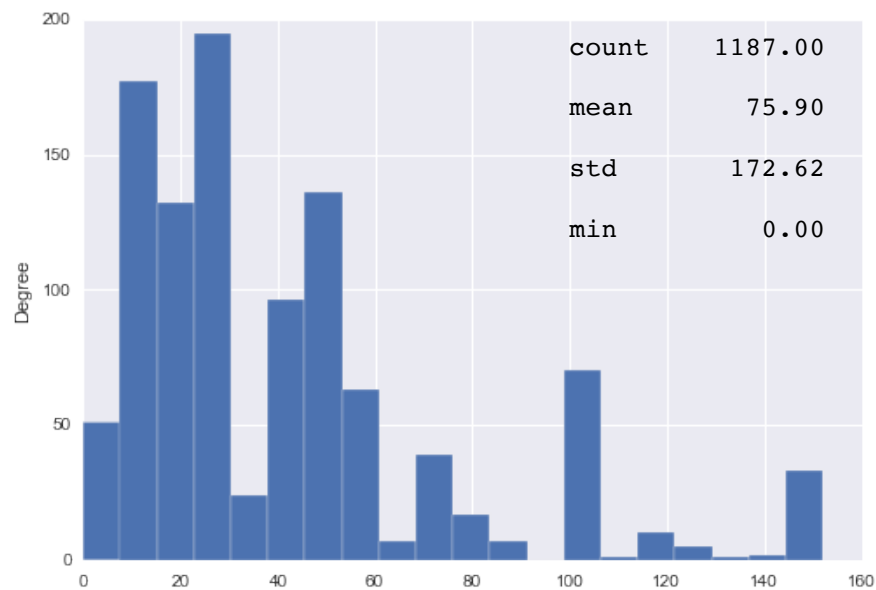


Figure 36 Spending on wine during an average month

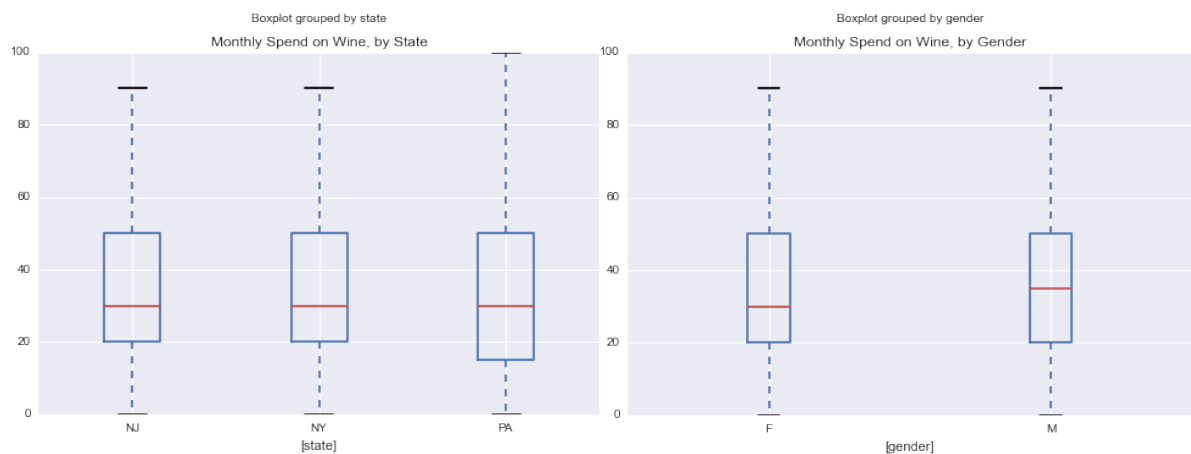


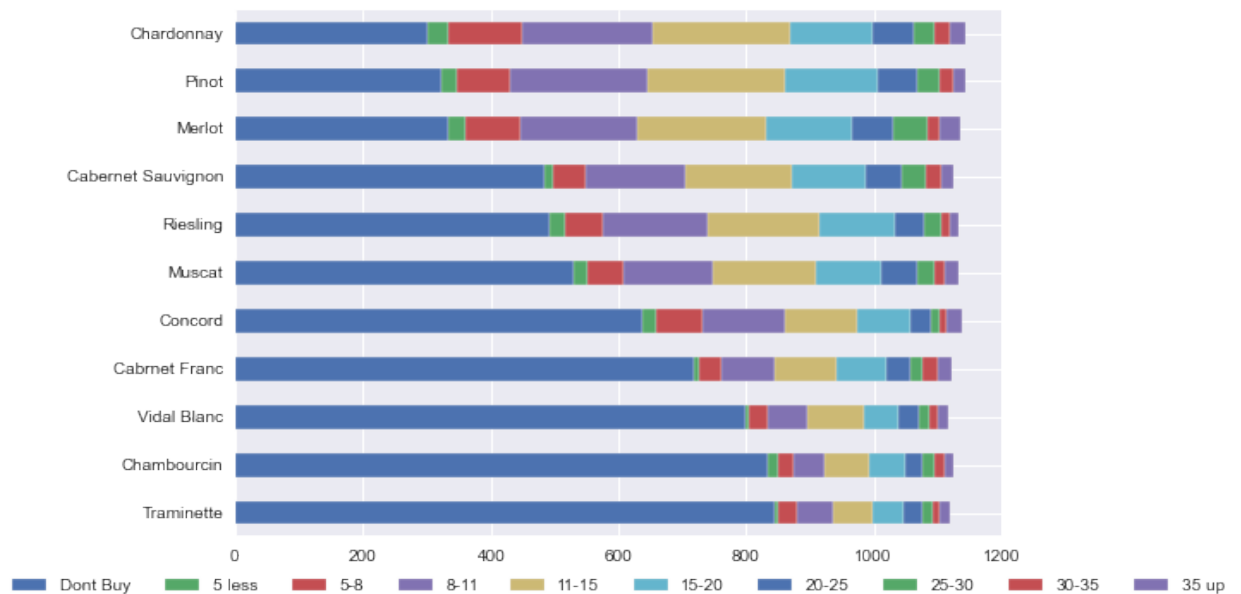
Figure 35 Monthly Spending grouped by state and gender

#### 4.11 Monthly spending on different wine varieties

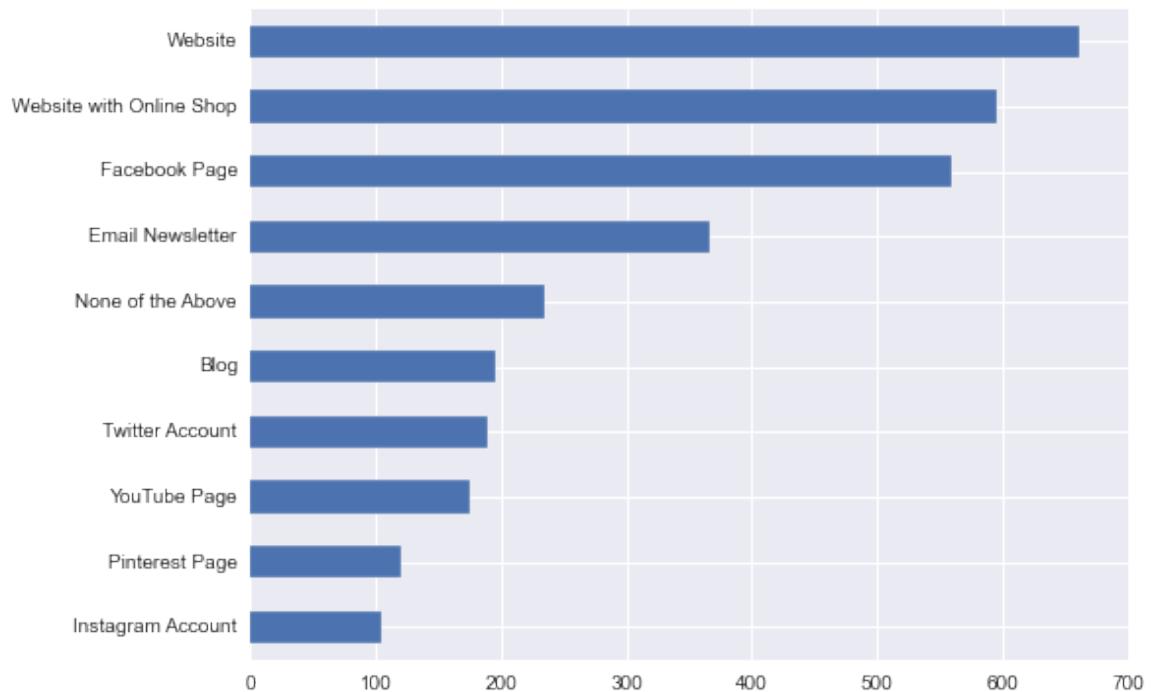
As shown in Figure 37, most money was spent on Chardonnay, Pinot and Merlot. These three are also found as the mostly consumed wine varieties. The most acceptable price categories fell into \$8 to \$20.

#### 4.12 Social Media Preference

As shown in Figure 38, survey respondents indicated that website, website with online shop and Facebook page are mandatory for a winery to offer or implement. An email newsletter is also somewhat important. However, Instagram, Pinterest Page, YouTube Page, Twitter and blog are not so important compared to the website, online shop and Facebook page.



*Figure 37 Monthly spending on different wine varieties  
(x-axis is the number of respondents; the unit of legends is U.S dollar)*



*Figure 38 Which components are mandatory for a winery to offer or implement?*

Survey respondents also stated that the following information on a winery's website or social networking site would best appeal to them.

1. *Wine serving and pairing suggestions*
2. *Notice of coupons, promotions, and discounts for wine and related products sold at the winery*
3. *Notice of events and special occasions held at the winery*
4. *Recipe/link to a recipe using wine as an ingredient*
5. *Information that educates the reader about wine*

#### 4.13 Wine made from fruits but not made primarily from grapes

As shown in Figure 39, more than 50% of total participants indicated that they purchase wine made from fruits but not made primarily from grapes.



*Figure 39 Do you buy wine made from fruits other than grapes?*

## 5 LOGISTIC REGRESSION

After the descriptive statistics, we have already got an adequate understanding of our survey data. However, we are not satisfied with that. We still want to examine how different attributes can affect the purchasing decisions of consumers. We want to answer the question about what kind of people are more likely to buy local wine. A Binomial Logistic Regression was deployed to find the answers we are looking for. The dependent variable of Binomial Logistic Regression is binary. In our case, our dependent variable is BUY which has only two values 1 and 0. Whereas 1 means ‘buy local wine’ and 0 means ‘not buy local wine’. The logistic regression model is to predict the probability that a consumer falls into ‘buy’ or ‘not buy’.

Below is how the Logistic Regression model was specified.

*Model*

$$\begin{aligned}
 \text{logit}(p_{buy}) &= \log\left(\frac{p_{buy}}{1 - p_{buy}}\right) \\
 &= \beta_0 + \beta_1 \text{stateNY} + \beta_2 \text{statePA} + \beta_3 \text{age45to64} + \beta_4 Q1a_2 + \beta_5 Q1a_3 \\
 &\quad + \beta_6 Q1a_4 + \beta_7 Q1a_5 + \beta_8 Q1a_6 + \beta_9 Q3b + \beta_{10} Q3e + \beta_{11} Q4a \\
 &\quad + \beta_{12} Q4b + \beta_{13} Q4c + \beta_{14} Q7_1 + \beta_{15} Q7_3 + \beta_{16} Q7_4 + \beta_{17} Q7_5 \\
 &\quad + \beta_{18} Q11a + \beta_{19} Q11b + \beta_{20} Q11c + \beta_{21} Q11d + \beta_{22} Q11e + \beta_{23} Q11i \\
 &\quad + \beta_{24} Q15 + \beta_{25} \text{gender} + \beta_{26} \text{educ} + \beta_{27} \text{fam}_{inc1} + \beta_{28} \text{fam}_{inc3} \\
 &\quad + \beta_{29} \text{job}_2 + \beta_{30} \text{job}_3 + \beta_{31} \text{job}_4 + \beta_{32} \text{job}_5 + \beta_{33} \text{job}_6 + \beta_{34} \text{marital}_2 \\
 &\quad + \beta_{35} \text{marital}_3 + \beta_{36} \text{marital}_4
 \end{aligned}$$

### **5.1 Dependent variable BUY**

The dependent variable BUY was derived from one of our survey questions. In the survey, one of the questions (Q9) asked participants to indicate whether they have purchased wine from the regions given in the question. Consumers who indicated they have purchased wine from any one of New Jersey, New York and Pennsylvania regions are coded as 1 in BUY. Those who have not purchased wine from the Mid-Atlantic regions are coded as 0 in BUY. 613 out of 1246 consumers indicated that they have purchased local wine.

### **5.2 Independent variables**

The questions from our survey have covered pretty much every aspect of information. However, not all of variables were used in the Logistic Regression. Some variables are removed due to excessive missing values.

Table 1 has the list of independent variables that were used in our Logistic Regression. This table contains information of the variable names, definitions and how they were recoded before we dumped them into the regression.

There are 48 variables in

Table 1. Some are about demographics, some are about consumer behavior and preferences. Details of these variables have already been covered when we were discussing descriptive statistics in Section 4. If you'd like to know more about the survey questions and answer options, please refer to the questionnaire itself which has been appended to the end of the thesis as Appendix. For the state variable, observations that are from states other than NJ, NY and PA are dropped out. Participants who age younger than 21 or older than 64 are also removed. Dummy variables were created for categorical variables such as state, age\_cat (age category), job and marital (marital status).



Table 1 List of independent variables for logistic regression

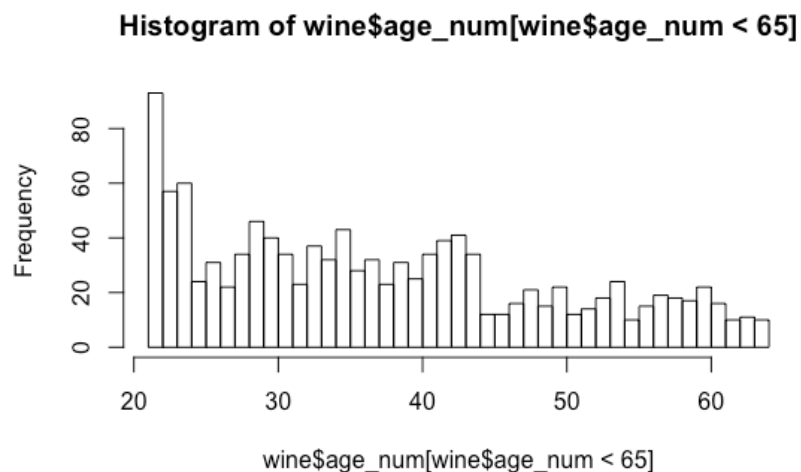
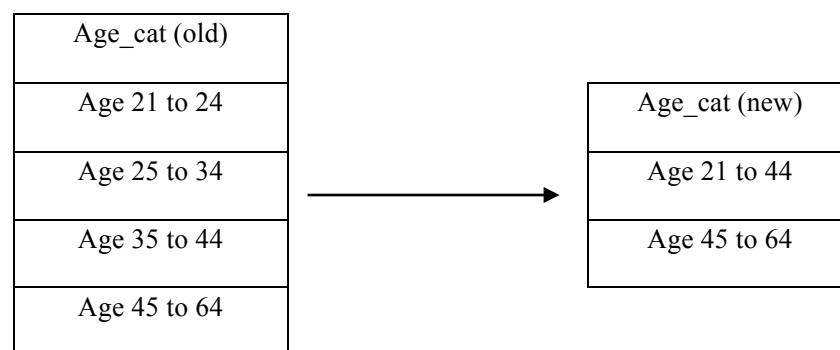
NAME	DEFINITION	RECODING
state	State	Categorical variable
age_cat	Age category	Categorical
Q1a	During an average year, how often do you drink wine?	Categorical variable Frequency from (daily) to (about once a year)
Q2	During an average year, how often do you purchase bottles of wine?	Categorical variable Frequency from (daily) to (about once a year)
Q3a	Your involvement in the wine purchased for your household	1 = YES 0 = NO
Q3b		
Q3c		
Q3d		
Q3e		
Q4a	Which statements describe how often you purchase 750ml bottles of wine?	1 = YES 0 = NO
Q4b		
Q4c		
Q4d		
Q5	Was wine(including sparkling wine, champagne, port, sherry etc.) the first alcoholic beverage you ever drank? (1)YES, (2)NO, (3)DON'T KNOW/DON'T REMEMBER	Categorical variable
Q6a	How often do you consume TABLE WINE?	

NAME	DEFINITION	RECODING
Q6b	How often do you consume SPARKLING WINE AND CHAMPAGNE?	
Q6c	How often do you consume Fortified wine	
Q6d	How often do you consume Regular Beer	
Q6e	How often do you consume Craft Beer	
Q6f	How often do you consume Distilled Spirits	
Q6g	How often do you consume Ready-to-drink Cocktails	
Q6h	How often do you consume Hard Cider	
Q7	We purchase different wine s for everyday consumption than special occasions or when entertaining	1 = YES 0 = NO
Q7_1	price difference, NA=no diff	
Q7_3	sweetness/dryness differ, recode NA= NO DIFF	Categorical variable
Q7_4	bottle size/volume differ, recode NA=NO DIFF	
Q7_5	closure type differ, recode NA= NODIFF	
Q7_6	container material differ, recode NA= nodiff	
Q8	Wine consumption change over the past three years	Categorical variable using level 2 as reference , level 2 means no change level 1 means decreased level 3 means increased
Q11a	Consume wine during meals	
Q11b	~ when dinning out at a restaurant	
Q11c	~ when at a party or gathering with family and/or friends	

NAME	DEFINITION	RECODING
Q11d	~ at a bar or lounge	
Q11e	~ at a sporting event or concert	
Q11f	~ when at a business dinner or event	
Q11g	~ when cooking	
Q11h	~ when watching TV or related activity	
Q11i	~ at the end of the day to relax	
Q11j	~ when celebrating holidays or other special occasions	
Q13	Average amount spent on wine each month, in dollar	continuous variable
Q15	Do you purchase fruit wine? Fruit wine not made primarily from grapes (1=yes, 2=no)	
gender	1=male, 2= female	
Q100_21	Excluding yourself, the number of adults age 21 and older in your household who drink wine	
Q100_17	The number of children, age 17 and younger, in your household	
educ	some high school(1) ~ master or higher(6)	Categorical variable. Regrouped into 2 categories. 0 = Lower than Bachelor's Degree 1 = Bachelor or Higher
fam_inc	less than \$25000(1) ~ \$200000 or greater(7)	Categorical variable
job		Categorical variable
marital		Categorical variable

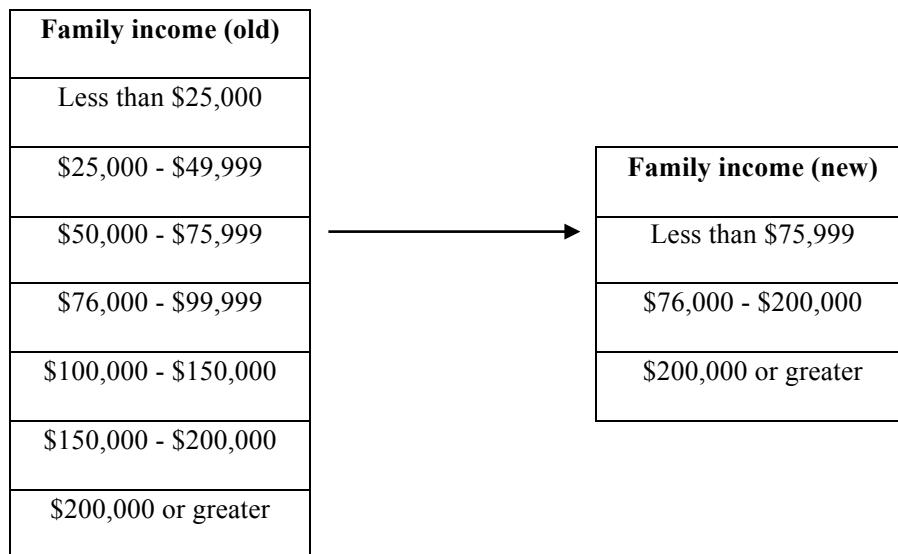
### 5.3 Model Tweak

After fitting our first Logistic Regression model with dependent variable BUY and independent variables from Table 1, we dropped out variables that are clearly not helping explaining the BUY. The age category variable and family income variable were also not statistically significant in the first model, but they are important demographic specs that we don't want to drop easily. So we regrouped the age category and family income in a different way. Figure 40 is a histogram of age, where age 44 is a clear divider for two age groups. So we regrouped age into two new groups. One from age 21 to 44, another from age 45 to 64. It turned out that Age 45 to 64 became significant in the following regression output.



*Figure 40 Histogram of age*

As mentioned before, the categorical variable family income was not significant in our first try, either. However, we still believe family income must have some explaining power on the BUY variable. We tried to regroup family income categories in a different way. The left side of Figure 41 is how family income grouped originally. The right side shows how it was regrouped in a new way. Basically, we divided the family income categories into three new groups. The first group is with annual family income less than \$75,999. The second group is with annual family income between \$76,000 to \$200,000. The third group is with annual family income \$200,000 or greater. You will see the second category become significant in the following regression output.



*Figure 41 Regrouping Family Income*

## 5.4 Logistic Regression Output

After manipulating our data in section 5.3, we fitted our model again. The output of our logistic regression is shown in Table 2. The definitions of variables can be referred to in Table 1 and the Appendix.

The dependent variable is still BUY. There are 49 independent variables in the output table. But we only used 33 variables from Table 1. The extra 16 are dummy variables derived from categorical variables such as state, job occupation and marital status. As you can see, many variables are not statistically significant, but we still keep them in the model. It is for interpretation purposes. For instance, we have four categories in marital status variable. They are (1) Married or in a Partnership, (2) Single, (3) Separated or Divorced, (4) Widower. We want to test whether single people are less likely to buy local wine than those who are married. The first category married or in a partnership was selected as the reference group. As you can see in Table 2, the second category Single is statistically significant at 5% level. The value of its coefficient is negative, which means that single people are significantly less likely to buy local wine than those who are married. Although the other two categories are not significant, if we removed them from the model, we're not comparing single towards married anymore. Instead, we're comparing single towards 'not-single' including married, divorced and widower and all other possible categories.

152 observations were deleted due to missing values, but we still have 1093 observations in this model. It is still a good sample size. The AIC of the whole model is bigger than the AIC of our first try, which means that our model is better after dropping some uninteresting variables. The following part of this section is about interpretation of the regression output. The *Margin* column of Table 2 shows the marginal effect of each variable.

### *State*

Let's start to interpret the regression output from the first variable state. The state variable means where the respondent resides. It has three levels: New Jersey, New York and Pennsylvania. From the regression output, people who live in New York state are more likely to buy local wine than people who live in New Jersey. The coefficient of PA is negative, but its p-value is 0.8. There is not enough evidence to prove that PA residents are less likely to buy local wine than NJ residents.

### *Age*

The age category variable has been regrouped into two categories. The new variable is named age45to64, it has value 1 means age from 45 to 64, and value 0 means age from 21 to 44. As you can see in the regression output, the p-value of age45to64 is 0.02. The null hypothesis is rejected soundly at 5% level. Consumers who are aged in the 45 to 64 range are more likely to buy local wine than those are younger.

### *Q1a. Wine Drinking Frequency*

Q1a is a categorical variable about wine drinking frequency. Figure 8 shows the details of this variable. It has six levels from (1) drinking daily to (6) drinking a few times a year. The first level drinking daily was selected as the reference group. In the remaining 5 levels, level 3 (drinking about once a week) and level 4 (drinking two to three times a month) are statistically significant. In addition, level 3 and level 4 also has the bigger log odds

compared to other levels of Q1a. The results are that people who drink about once a week or two to three times a month are more likely to buy local wine than those daily drinkers. Roughly speaking, moderate drinkers are more likely to buy local wine than heavy drinkers.

*Q3 Involvement in the wine purchased for one's household*

Five statements were given in this question. Respondents were asked to select all the statements that apply to themselves. Below is a list of those five statements.

*Q3a: Even though I do not purchase wine for the household I do suggest/select the wine that is purchased.*

*Q3b: I purchase the "everyday" wine that I/we consume in the home during an average day.*

*Q3c: I purchase the wine I/we serve during special occasions and when we entertain.*

*Q3d: I am the one who purchases wine to give as gifts for others or to bring to other's home when invited over.*

*Q3e: When at a restaurant, I am the one who selects the wine from the menu that I/we will drink.*

From the regression output, we can see that people who identify themselves as "everyday" wine buyers are more likely to buy local wine. No evidence shows that consumers who buy wine for special occasions, for gifts or when at a restaurant, has any effect on our dependent variable.



*Q4 Statements describe how often one purchases 750ml bottles of wine*

Four statements were given in this question. Respondents were asked to select all the statements that apply. Those four statements are as below.

*Q4a: I typically purchase one or more 750ml bottles to be consumed immediately (either in my home or for meals at other's home).*

*Q4b: I purchase one or more bottles to be added to my collection and/or be consumed at a later time.*

*Q4c: I purchase wine infrequently but when I do I purchase at least a case (12 or more 750ml bottles) so that I know I will have wine available when I needed.*

*Q4d: I purchase wine through a wine club with a fixed number of 750ml bottles purchased/delivered on a scheduled basis.*

Consumers who tend to purchase wine to be added to their collections or be consumed at a later time are more likely to buy local wine. These people may be considered as elegant wine drinkers. They may have a deeper understanding on wine. Instead of buying wine for immediate need or for a bulk discount, they buy bottles of wine to be added to their collection.

*Q7 Different wine attributes of everyday-wine and special occasion wine*

As shown in Figure 16, 72% survey respondents indicated that they purchase different wine for everyday drink and special occasions. In order to know what is different between them, we looked into the Q7 series variables. Q7 series variables are about the differences in wine price, sweetness, bottle size, closure type and container material. Details are shown in Figure 17, Figure 21, Figure 23, Figure 22 and Figure 24.

In our logistic regression model, Q7\_1 is about the price. The result shows that the willingness to pay more for everyday wine or special occasion wine doesn't have a significant effect on the BUY decision of local wine. Q7\_3 is about the flavor. Consumers who prefer everyday wine to be dryer are more likely to buy local wine. Q7\_4 is about the bottle size. Consumers who prefer everyday wine to be in smaller containers are more likely to buy local wine. Q7\_5 is about the closure type. Consumers who prefer everyday wine with cork closures are more likely to buy local wine.

#### *Q11 Occasions that people consume wine*

People consume wine on different occasions. Some people tend to consume wine during meals, while others tend to drink wine when celebrating holidays. We want to know on which occasions, the local wine buyers drink wine. Figure 33 shows the details. People who tend to drink wine during meals, when at a party or gathering with family/friends, and at the end of the day to relax, are more likely to purchase local wine. People who tend to consume wine when dining out at a restaurant, and at a sporting event or concert are less likely to buy local wine. More than 50% consumers indicated that they drink wine when celebrating holidays or other special occasions, but there is not enough evidence to prove its effect on buying decisions.

#### *Q15 Do you purchase fruit wine that is not made primarily from grapes?*

The p-value of Q15 variable is 0.1077. It is not small enough to be rejected at the 10% level, but it is already very close. We still decided to keep it in the model. Consumers who

purchase fruit wine that is not made primarily from grapes are more likely to buy local wine. It is statistically significant at the 15% level.

### *Gender*

In the gender variable, we have male as 1 and female as 0. The result shows that males are more likely to buy local wine than females.

### *Education*

As shown in Figure 4, the education (educ) variable has six categories in the beginning. They are 1) some high school, 2) high school graduate, 3) some college/technical school, 4) associate degree/tech. school grad., 5) bachelor's degree, 6) master's degree or higher. In the first try, the education categorical variable is not statistically significant at any level. Then we regrouped the education variable into two categories. One category is education level lower than bachelor's degree, the other category is bachelor's degree or higher. It turned out that wine consumers with a bachelor's degree or higher are more likely to buy local wine than those with a lower education level.

### *Family Income*

The annual family income (fam\_inc) variable was regrouped in to three categories. They are 1) less than \$75,999, 2) \$76,000 to \$200,000, 3) \$200,000 or greater. In the regression,

we use the second level as the reference group. The result shows that both lower income people or higher income people are less likely to buy local wine than middle income people.

Level 1 has p-value 0.03, whereas level 3 has p-value 0.1.

### *Job Occupation*

As shown in Figure 6, there are six categories in the job variable. They are 1) employed by someone else, 2) self-employed, 3) student, 4) full-time homemaker, 5) unemployed and 6) retired. The first category was selected as the reference group. In the results from logistic regression, full-time homemakers are more likely to buy local wine than people employed by someone else. Those unemployed are less likely to buy local wine than people employed by someone else.

### *Marital Status*

As shown in Figure 7, there are four categories in the marital status variable. They are 1) married or in a partnership, 2) single, 3) separated or divorced, 4) widower. The regression results show that, single consumers are less likely to buy local wine than those who are married or in a partnership.

*Table 2 Logistic Regression Output*

<b>Variable</b>	<b>Margin</b>	<b>Std. Err.</b>	<b>z-value</b>	<b>P&gt; z </b>	<b>Signif.</b>
stateNY	0.11339	0.04409	2.5720	0.0101	*
statePA	-0.01059	0.04747	-0.2231	0.8235	
age45to641	0.09370	0.04026	2.3272	0.0200	*
Q1a_F2	0.07880	0.06903	1.1415	0.2536	
Q1a_F3	0.20119	0.06652	3.0243	0.0025	**
Q1a_F4	0.17423	0.06803	2.5613	0.0104	*
Q1a_F5	0.12677	0.07747	1.6364	0.1018	
Q1a_F6	0.07690	0.08170	0.9412	0.3466	
Q3b	0.11219	0.03764	2.9803	0.0029	**
Q3e	0.03902	0.03696	1.0557	0.2911	
Q4a	0.04403	0.04249	1.0362	0.3001	
Q4b	0.10239	0.03884	2.6362	0.0084	**
Q4c	0.09794	0.05660	1.7304	0.0836	.
Q7_1new1	0.09653	0.06516	1.4815	0.1385	
Q7_3new1	-0.13442	0.05452	-2.4655	0.0137	*
Q7_4new1	0.13975	0.04349	3.2136	0.0013	**
Q7_5new1	-0.19749	0.04753	-4.1547	0.0000	***
Q11a	0.06540	0.03834	1.7060	0.0880	.
Q11b	-0.06189	0.03969	-1.5594	0.1189	
Q11c	0.08557	0.04096	2.0890	0.0367	*
Q11d	0.03871	0.03605	1.0737	0.2830	
Q11e	-0.09403	0.05247	-1.7919	0.0731	.
Q11i	0.05807	0.03546	1.6378	0.1015	
Q151	0.05431	0.03379	1.6071	0.1080	
gender1	0.12796	0.03611	3.5437	0.0004	***
educ_bachelor	0.08924	0.03530	2.5277	0.0115	*
fam_inc_new2	0.07916	0.03814	2.0754	0.0379	*
fam_inc_new3	-0.07604	0.09528	-0.7981	0.4248	
job2	0.02390	0.06386	0.3742	0.7082	
job3	0.03131	0.06569	0.4766	0.6336	
job4	0.14372	0.05163	2.7835	0.0054	**
job5	-0.11168	0.06049	-1.8464	0.0648	.
job6	0.04000	0.08126	0.4922	0.6225	
marital2	-0.06861	0.03889	-1.7643	0.0777	.
marital3	0.02454	0.06360	0.3859	0.6996	
marital4	0.09976	0.15664	0.6368	0.5242	

## **6 MARKET SEGMENTATION USING CLUSTER ANALYSIS**

As we discussed in the first section of this thesis, marketing cost is one of the concerns of local wineries. Local wineries cannot afford the cost if the marketing strategy is dependent upon targeting an entire mass market. The importance of market segmentation is that it allows a business to precisely reach a consumer with specific needs and wants. In the long run, this benefits the company because they are able to use their corporate resources more effectively and make better strategic marketing decisions. In this section, we employed Cluster Analysis to class wine consumers into several groups. Different groups will have different demographics and preferences.

Many cluster analysis methods are available out there. We used the `hclust` function in R to achieve the hierarchical clustering. Ward linkage was used when we applied the hierarchical clustering. The hierarchical clustering method defines the cluster distance between two clusters to be the maximum distance between their individual components. At every stage of the clustering process, the two nearest clusters are merged into a new cluster. The process is repeated until the whole data set is agglomerated into one single cluster.

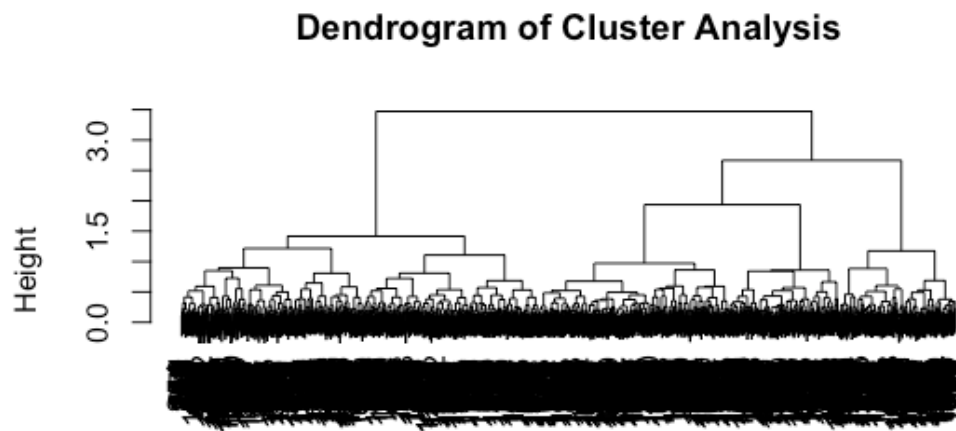


Figure 42 Dendrogram of Cluster Analysis

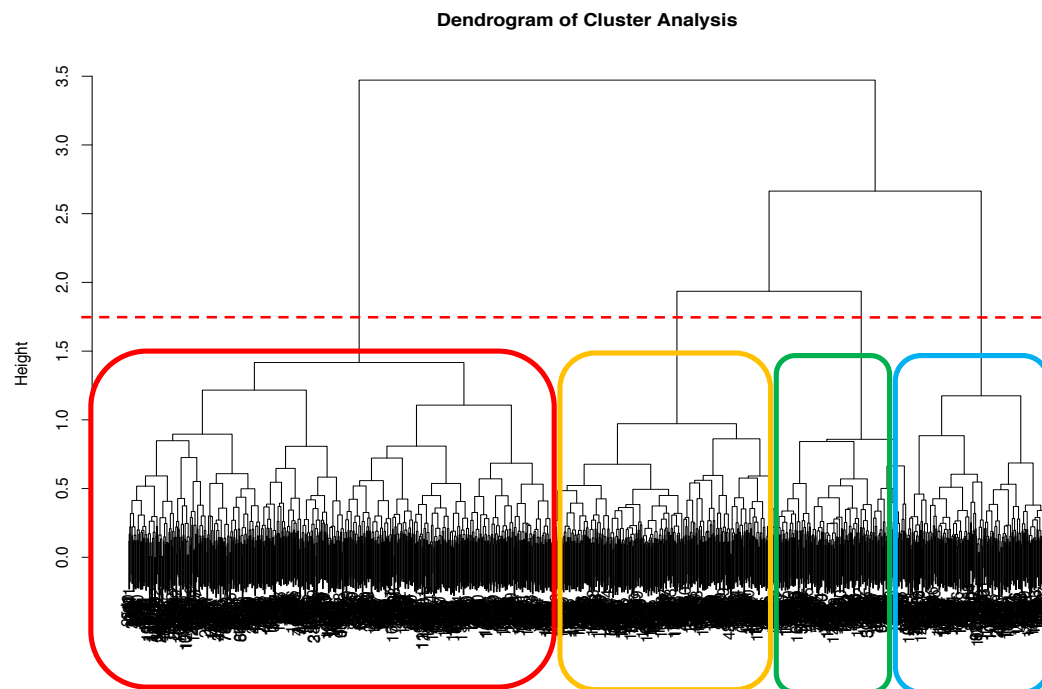


Figure 43 Elbow plot of optimal number of clusters

Figure 42 is the dendrogram of our cluster analysis. From the dendrogram, it is not very clear how many clusters we should choose. It can be cut at either 2, 3 or 4 clusters. In order to decide the optimal number of clusters, we plotted an elbow plot as shown in Figure 43. The elbow is very clear; it appears at the fourth cluster.

According to the elbow plot, we chose to keep four clusters. A simple annova was employed to test whether there were significant differences between any two classes. Annova results as below shows that there is significant difference between at least two classes.

	Df	Sum Sq.	Mean Sq.	F value	Pr(>F)	
class	3	5.71	1.9049	7.739	4.01E-05	***
Residuals	1242	305.7	0.2461			



*Figure 44 Dendrogram of 4 clusters*



*Table 3 Cross table of BUY variable and clusters*

	Class1 <b>Detractors</b> (n=574)	Class2 <b>Enthusiasts</b> (n=306)	Class3 <b>Neutral</b> (n=189)	Class4 <b>Advocators</b> (n=177)	<b>Total</b> <b>(N=1246)</b>
<b>BUY</b>					
YES	32.6%	74.5%	49.2%	59.3%	<b>49.2%</b>
NO	67.4%	25.5%	50.8%	40.7%	<b>50.8%</b>

Table 3 is a cross table of the BUY variable and wine consumer clusters. 67.4% of Class1 don't buy local wine. That means Class1 is very unlikely to buy local wine, so this class was named Detractors. 74.5% of Class2 buy local. This is a very high percentage. We call Class2 Enthusiasts. The third class is considered as Neutral since about 50% of Class3 buy local wine. The last class has 59.3% buy local wine. It is not as high as Class2, but is still more likely to buy local wine compared to Class1 and Class3. The cluster names follow the convention of Kolyesnikova's research in 2009.

## 6.1 Two-Way Contingency and Chi-Square Independence Test of Wine Consumer Clusters

In order to study the differences between the four market segments we derived from Cluster Analysis, a two-way contingency table and Chi-square independence tests are performed. The results are presented in Table 4.

The Chi-square independence test is used to test whether two variables are associated or not. In the case of the state variable, our hypotheses are:

$H_0$  : State and Wine Consumer Clusters are not associated.

$H_1$  : State and Wine Consumer Clusters are associated.

The idea behind the chi-square independence test is to compare the observed frequencies with the frequencies we would expect if the null hypothesis of non-association is true. Equation (1) is the test statistic used for this comparison.  $E$  represents the expected frequencies whereas  $O$  refers to observed frequencies. Equation (2) was used to estimate  $E$ .

$$\chi^2 = \sum (O - E)^2 / E \quad (1)$$

$$E = \text{row} * \text{column} / n \quad (2)$$

The two-way contingency table shows us the distribution of the data in each group, which allows us to compare the difference of the levels in the categorical variables in each group. Based on the two-way contingency table, the Chi-square tests are conducted in order to test if each of the variables is associated with the response variable. In Table 4, we can see that the Chi-Squared test results for state, age, income, education, and marital status are significantly related with the wine consumer clusters.

Table 4 Contingency Table and Independence Test of Wine Consumer Clusters

	CLUSTER					Chi-Squared
	Class1 (n=574)	Class2 (n=306)	Class3 (n=189)	Class4 (n=177)	Total (n=1246)	$\chi^2$
<b>State</b>						
NJ	21.6%	20.6%	13.2%	16.9%	19.4%	F = 30.189 p = 3.618e-05 ***
NY	44.1%	42.8%	65.1%	50.8%	47.9%	
PA	34.3%	36.6%	21.7%	32.2%	32.7%	
<b>Age</b>						
21-24	20.2%	13.1%	22.2%	17.5%	18.4%	F=43.473 p = 1.765e-06 ***
25-34	23.9%	25.8%	41.3%	26.0%	27.3%	
35-44	27.9%	29.1%	24.3%	28.2%	27.7%	
45-64	28.0%	32.0%	12.2%	28.2%	26.6%	
<b>Q1a Wine Drink. Freq.</b>						
Daily	3.5%	2.6%	28.0%	5.1%	7.2%	F=266.81 p < 2.2e-16 ***
A few times a week	20.6%	25.5%	39.2%	28.2%	25.7%	
About once a week	15.2%	19.6%	18.5%	29.4%	18.8%	
2 to 3 times a month	23.7%	29.4%	11.6%	22.0%	23.0%	
About once a month	14.6%	9.2%	1.1%	9.6%	10.5%	
A few times a year	22.5%	13.7%	1.6%	5.6%	14.8%	
<b>Q2 Wine Buying Freq.</b>						
Daily	0.3%	0.0%	12.2%	1.7%	2.2%	F=358.15 p < 2.2e-16 ***
A few times a week	3.3%	1.6%	21.7%	3.4%	5.7%	
About once a week	9.2%	11.4%	30.2%	16.4%	14.0%	
2 to 3 times a month	17.2%	25.8%	22.2%	27.7%	21.6%	
About once a month	23.5%	24.8%	9.0%	28.8%	22.4%	
A few times a year	46.3%	36.3%	4.8%	22.0%	34.1%	
<b>Q5 First Alcohol Wine</b>						
YES	27.8%	25.8%	61.9%	36.0%	33.7%	F=88.828 p < 2.2e-16 ***
NO	62.3%	61.4%	35.4%	54.9%	56.9%	
Don't Remember	9.9%	12.7%	2.6%	9.1%	9.4%	

	CLUSTER					Chi-Squared
	Class1 (n=574)	Class2 (n=306)	Class3 (n=189)	Class4 (n=177)	Total (n=1246)	$\chi^2$
<b>Gender</b>						
Male	27.1%	39.9%	61.9%	37.0%	37.1%	F=71.304
Female	72.9%	60.1%	38.1%	63.0%	62.9%	p = 2.244e-15
<b>Education</b>						
Lower than Bachelor's Degree	62.6%	40.5%	51.6%	48.6%	53.3%	F=50.593
Bachelor's Degree or Higher	37.4%	59.4%	48.3%	51.5%	46.7%	p = 9.628e-06  ***
<b>Family Income</b>						
Less than \$75,999	69.6%	55.5%	56.1%	59.5%	62.5%	F=23.27
\$76,000-\$200,000	27.5%	41.8%	39.4%	37.6%	34.4%	p = 0.0007108
\$200,000 or greater	2.8%	2.7%	4.4%	2.9%	3.1%	***
<b>Job Occupation</b>						
Employed by someone else	54.2%	64.3%	65.9%	63.0%	59.9%	F=31.806
Self-employed	6.9%	5.4%	8.2%	11.0%	7.3%	p = 0.006842
Student	8.6%	6.1%	8.2%	6.4%	7.6%	***
Full-time homemaker	11.8%	11.4%	9.9%	9.2%	11.1%	
Unemployed	12.4%	6.7%	6.6%	6.9%	9.3%	
Retired	6.1%	6.1%	1.1%	3.5%	4.9%	
<b>Marital Status</b>						
Married or in a Partnership	55.3%	65.9%	60.2%	53.8%	58.5%	F = 19.583
Single	35.4%	24.4%	34.8%	36.3%	32.6%	p = 0.02067
Separated or Divorced	8.6%	8.4%	4.4%	7.6%	7.8%	**
Widower	0.8%	1.3%	0.6%	2.3%	1.1%	

*Table 5 Profile of Wine Consumer Clusters*

	<b>Class1 Detractors</b>	<b>Class2 Enthusiasts</b>	<b>Class3 Neutral</b>	<b>Class4 Advocators</b>
<b>State</b>		More PA, less NY	63% is from NY	53% is from NY, less PA than average
<b>Age</b>		Oldest	Youngest	mid-age
<b>Drinking Freq.</b>	Least frequent	Slightly less frequent than average	Most frequent, heavy wine drinkers	Moderate wine drinkers
<b>Buying Freq.</b>	Least frequent	Slightly less frequent than average	Most frequent.	Moderate
<b>First Drink was wine</b>	25%YES, lower than average (33%)	26% YES, lower than average	65% YES,	37% YES
<b>Gender</b>	75% female	65% female	37% female	58% female
<b>Education</b>	lower	higher	higher	higher
<b>Family Income</b>	Lowest	middle	middle	Highest
<b>Occupation</b>	More retired and unemployed	More retired and unemployed	Less retired and unemployed	Less retired and unemployed
<b>Marital Status</b>		Slightly more married	Slightly more single	Slightly more single

## 6.2 Results of Cluster Analysis

Only variables that tested significant were kept in Table 4. Demographics such as state, age, gender, education, income, occupation and marital status are all associated with the market segments we derived from Cluster Analysis. Consumer behavior in terms of drinking frequency and buying frequency are also correlated with our market clusters.

Most of Class1 the Detractors are infrequent wine drinkers. They tend to have lower education level and less income. Most of Class2 the Enthusiasts are in their 40's and 50's. They are moderate wine drinkers. Usually, they drink wine once a week or two to three times a month. They tend to have a higher education level and mid-level income. Many of them are married or in a partnership at least. Class3 Neutral are frequent wine drinkers. Most of them are in their 20's and 30's. They also have a higher education level and mid-level income like Class2. Many of them are male. Class2 Advocators are moderate wine drinkers. They tend to have a higher education level and the highest income compared to other classes. Many of them are single.

Class2 the Enthusiasts and Class4 the Advocators are the target market of local wineries. If local wineries want to target larger markets, Class3 the Neutral can also be a likely potential market. However, Class1 the Detractors is the market segment that local wineries should avoid.

## 7 CONCLUSIONS

Throughout this paper, we have employed different statistical techniques to address our research questions. First, we fitted a Logistic Regression model to answer what kind of people are more likely to buy local wine. Then we applied Cluster Analysis to segment the Mid-Atlantic local wine market into four clusters. In this conclusion section, we summarized the findings from Logistic Regression and Cluster Analysis. After that, a set of marketing strategies based on 4Ps Marketing Mix was provided to help Mid-Atlantic wineries to protect their local wine market.

### 7.1 What Kind of People Are More Likely to Buy Local Wine?

What kind of people are more likely to buy local wine? It's the most important question we wanted to answer through this research. This question was answered from three aspects. They are demographics, behavior and preferences.

#### *Demographics*

New York state residents are more likely to buy local wine than New Jersey residents; Consumers aged between 45 to 64 years old are more likely to buy local wine than those who are younger; Males are more likely to buy local wine than females; Middle and upper income level (family income from \$76,000 to \$200,000 per year) people are more likely to buy local wine than people with other income levels (either lower than \$76,000/year, or higher than \$200,000/year); People with Bachelor's degree or higher are more likely to buy local wine than those without Bachelor's degree; Full-time homemakers are more likely to buy local wine than people employed by someone else. People who are married or in a partnership are more likely to buy local wine than single people.

### *Behavior*

Consumers who drink wine once a week or 2 to 3 times a week are considered as moderate wine drinkers. They are more likely to buy local wine than those daily drinkers.

Consumers buy wine to serve different purposes. Sometimes people buy wine for everyday consumption, sometimes people buy wine for special occasions. Those everyday-consumption-wine buyers are more likely to buy local wine than those special-occasion-wine buyers.

Consumers who purchase wine to be added to their collections or to be consumed at a later time are more likely to buy local wine, compared to those buy wine for immediate need.

People drink wine on many different occasions. Those who often drink wine during meals, when at a party or gathering with family and friends, or at the end of day to relax, are more likely to purchase local wine.

### *Preferences*

The demographic and behavior aspects focused on the characteristics of consumers. The preference aspect is different. It focused on the attributes of wine. According to our study, Mid-Atlantic wineries should pay more attentions to the everyday-consumption-wine sector than the special-occasion-wine sector. Likely local wine buyers indicated that they prefer everyday-consumption-wine to be dry, in small containers (compared to boxed wine) and with cork closure (compared to screw cap type).



## 7.2 Mid-Atlantic Wine Market Segmentation

By using Cluster Analysis, the Mid-Atlantic wine market was segmented into four clusters.

They are Class1 Detractors, Class2 Enthusiasts, Class3 Neutral and Class4 Advocators.

### *Segment of Unlikely Buyers - Class1 Detractors*

Class1 Detractors is the cluster that is the most unlikely to buy local wine. Mid-Atlantic wineries should avoid this market segment when designing marketing strategies. 67.4% of Class1 Detractors indicated that they had never bought local wine before. Consumers in this cluster are infrequent wine drinkers. They buy and drink wine infrequently. Compared to other clusters, less of Class1 Detractors have a Bachelor's degree. In addition, they tend to have a lower income level.

### *Segments of Likely Buyers - Class2 Enthusiasts and Class4 Advocators*

Class2 Enthusiasts and Class4 Advocators are the target market of Mid-Atlantic local wineries. More attention should be paid to these two market segments. Class2 Enthusiasts are the most likely to purchase local wine compared to other three clusters. 74.5% of Class2 indicated that they had bought wine from the Mid-Atlantic wine region. Most of Class2 Enthusiasts are in their 40s or 50s. They are moderate wine drinkers (once a week, or 2 to 3 times a week). Many of them have bachelor's degree or a higher education level. They tend to have mid income level. About 60% of Class4 Advocators stated that they had bought local wine before. The characteristics of Class4 are very similar to Class2. Many of them are moderate wine drinkers with a Bachelor's degree. However, unlike many of Class2, they are married or in a partnership, while most of Class4 are single.

### *Segment of Neutral Buyers - Class3 Neutral*

The chance of Class3 Neutral to buy local wine is 50/50. Most consumers in Class3 are males in their 20s or 30s. They drink and buy wine more frequently than consumers in other clusters. Many of them hold a Bachelor's degree, and have a mid-income level. Typically, we don't recommend Mid-Atlantic wineries to target this market segment, unless they want to expand their market beyond Class2 and Class4.

### **7.3 Marketing Strategies based on 4Ps Marketing Mix**

After the discussions of our findings from descriptive statistics, Logistic Regression and Market Segmentation, we developed some marketing strategies based on 4Ps Marketing Mix for Mid-Atlantic wineries. The first P is Product. According to our study, consumers who mostly buy wine for everyday consumption are more likely to buy local wine, so Mid-Atlantic wineries should focus on the everyday-consumption-wine sector. One example is table wine that is consumed during a meal or casual occasion. In addition, consumers prefer local wine to be dry rather than sweet. Small container and cork closure is also important. The second P is Price. Although local wine buyers tend to buy local wine for everyday consumption, they also sometimes buy local wine for special occasions. The most popular price range for everyday-consumption-wine is from \$8 to \$15 per bottle. Typically, one bottle is 750ml. For special-occasion-wine, consumers are willing to pay from \$15 to \$25 per bottle. The third P is Place. Place refers to the channel where the product was sold. According to our descriptive statistics, retail liquor stores are still the most popular channel among all other channels. In addition, compared to wine from other regions, more local wine is sold in tasting rooms and festivals. Local wineries can utilize this advantage to attract more local consumers. The last P is Promotion. Wine consumers are the most

interested in the health benefits associated with drinking wine and other wine knowledge. There are many different ways to distribute information to the target market. According to our descriptive statistics, the top three popular digital marketing media is a website with online shop, Facebook page and email newsletter.

Additional studies should be conducted to research preferences of consumers on wine varietals. Other possible further research can investigate the consumers' response to different wine tasting activities and festivals.

## REFERENCE

- Assael, Henry. (2005). *Consumer Behavior A Strategic Approach* (2005 Indian: Dreamtech Press.
- Brown, T. A. (2012). *Confirmatory factor analysis for applied research*. Guilford Press.
- Chrea, C., Melo, L., Evans, G., Forde, C., Delahunty, C., & Cox, D. N. (2011). An investigation using three approaches to understand the influence of extrinsic product cues on consumer behavior: An example of Australian wines. *Journal of Sensory Studies*, 26(1), 13-24.
- Dodd, Tim H. (1995). Opportunities and pitfalls of tourism in a developing wine industry. *International Journal of Wine Marketing*, 7(1), 5-16.
- Dodd, Tim H, Laverie, Debra A, Wilcox, James F, & Duhan, Dale F. (2005). Differential effects of experience, subjective knowledge, and objective knowledge on sources of information used in consumer wine purchasing. *Journal of Hospitality & Tourism Research*, 29(1), 3-19.
- Fisher, D. (2011). U.S. wineries grow 9% to 6,785. *Wine Business Monthly*, 70-37.
- Hall, John, Binney, Wayne, & O'Mahony, G Barry. (2004). Age related motivational segmentation of wine consumption in a hospitality setting. *International journal of wine marketing*, 16(3), 29-43.
- Hosmer Jr, D. W., & Lemeshow, S. (2004). *Applied logistic regression*. John Wiley & Sons.

- Johnson, Trent E, & Bastian, Susan EP. (2007). A preliminary study of the relationship between Australian wine consumers' wine expertise and their wine purchasing and consumption behaviour. *Australian Journal of Grape and Wine Research*, 13(3), 186-197.
- Lockshin, Lawrence S, Spawton, Anthony L, & Macintosh, Gerrard. (1997). Using product, brand and purchasing involvement for retail segmentation. *Journal of Retailing and Consumer services*, 4(3), 171-183.
- Marshall, R Scott, Akoorie, Michèle EM, Hamann, Ralph, & Sinha, Paresha. (2010). Environmental practices in the wine industry: An empirical application of the theory of reasoned action and stakeholder theory in the United States and New Zealand. *Journal of World Business*, 45(4), 405-414.
- Mitchell, Vincent-Wayne, & Greator, Michael. (1989). Risk reducing strategies used in the purchase of wine in the UK. *European Journal of Marketing*, 23(9), 31-46.
- Nichols, R. (2011, Jan 24, 2011). U.S. wine consumer trends; Boomers' tastes evolve; Millennials continue to drive market growth. Retrieved Jan 12, 2012
- Power, Chris, Rodgers, Bryan, & Hope, Steven. (1999). Heavy alcohol consumption and marital status: disentangling the relationship in a national study of young adults. *Addiction*, 94(10), 1477-1487.
- Rao, J. N. K., & Scott, A. J. (1981). The analysis of categorical data from complex sample surveys: chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association*, 76(374), 221-230.

RNCOS. (2011). U.S. wine drinking habit remains stable despite economic downturn: RNCOS Industry Research Solutions.

Storchmann, Karl. (2010). The economic impact of the wine industry on hotels and restaurants: evidence from Washington State. *Journal of Wine Economics*, 5(01), 164-183.

U.S. International Trade Association. (2011). Wines Sparkling and Non-Sparkling: U.S. Imports for Consumption from Jan - Sep.

Whetstone, K.J., B.L. Smith, and B. Farley. (2011). Fruit Report: October 2011 (N. Y. D. o. A. a. Markets, Trans.) (Vol. No. 975-4-11): United States Department of agriculture, National Agricultural Statistics Service.