An Analysis of Scotch Whisky Prices

Brian Detweiler

Thursday, May 5th, 2016

Abstract

Single malt scotch whisky is a varied and complex spirit that is aged in barrels for anywhere between three to 75 years. Prices range from $20 to hundreds of thousands of dollars. In this paper, we collect data from 17 online retailers and examine their prices of single malt scotch. We will investigate the independent distributions of age and price and determine if there is any correlation. We also test hypotheses of claims made by the industry, and finally, we devise a ranking algorithm of online retailers that will determine the overall lowest cost options.

Contents

List of Figures

I love scotch. Scotchy, scotch, scotch. Here it goes down, down into my belly...

Ron Burgundy, *Anchorman*

1 Introduction

It's easy to break a budget on alcohol, as many college students know. But it becomes exponentially easier with scotch whisky[1] with the rarest of breeds fetching upwards of $460,000.[1] Even middling bottles can easily run over $100, leaving many to wonder why anyone would spend their hard earned money on something that is sure to have the same net result as a $10 bottle of cheap swill that can be mixed with cola.

Scotch drinkers tend to see this from a different angle. Scotch collecting and tasting can be a hobby (albeit a high-end one), and should not be equated with just drinking. Connoisseurs appreciate the arduous time spent producing the whisky, from the distillation to the aging and even the bottling and packaging. Scotch takes much longer to age than bourbon, due to the climate in Scotland, and produces a vastly different flavor profile. In the past, the typical age for scotch has been twelve years of rest in a barrel, but due to rising demands, many distilleries have been experimenting with younger batches.

So called "no age statement" varieties have been rising in popularity, in part to meet market demand amid surging popularity and dwindling stock of longer aged whiskies. Scotch must be aged for three years by law, but for most scotches, desirability increases with the age. Is scotch *better* with age? It's a subjective question, and it depends on the scotch, but suffice it to say, it is definitely *different*. That topic is not discussed here. *Whisky-*

[1] Scotch whisky is spelled without the "e"

Analysis.com does a detailed meta-critic analysis if you are interested in a statistical analysis of tasting notes. [6]

When it comes to the question of age and quality, perception may not match reality. Chivas Brothers (makers of *The Glenlivet* and other whiskies) ran a survey in 2010 examining consumer perception of age pertaining to scotch. They found that the vast majority of consumers actively look for an age statement and see it as an indicator of quality. In fact, 94% believe that the older a whisky is, the better it must be. [3] Of course, this is not necessarily true, but could this perception have an affect on price?

In this paper, we examine the prices of scotch and how it correlates to age. We will also look at various online retailers and compare them with each other. We also present an open source database of prices for over two thousand whiskies, which may serve as guide for making informed consumer decisions.

In an effort to keep this project manageable, several restrictions must be put in place to avoid scope creep. These restrictions include the following:

1. Must be single malt scotch whisky

   Blends will not be considered

   Grain whiskies will not be considered

   Only whiskies from Scotland will be considered

2. Only 750ml bottles available for retail purchase will be considered

   Gift sets will not be considered

   Bottles of smaller or larger sizes will not be considered

   UK standard sizes of 700ml were not considered

   Inventory is not considered (the bottle may be out of stock, so long as the price is available)

3. Purchase minimums or maximums are not considered

4. Shipping, handling, customs, or other fees are not considered [2]

[2]These would obviously play a big role in a purchase decision, but building a robust pricing schema is beyond the scope of this project

2 Findings

2.1 Environment Setup

2.1.1 Gathering Data

The most tedious part of a data scientist's job is data wrangling. It is a necessary evil, because often, interesting data does not exist as a clean, ready to use dataset. If we want the data, we capture and clean it.

To this end, the data we want is scotch, complete with name, distilleries, producers, ages, and prices. There are some closed databases at [4] and [5], and an interesting open database of tasting notes at [6]. Each of these probably have more information than what is gathered here, but this is produced as open source and allows and encourages others to build on this work.

The data wrangling method used here was quite unsophisticated, and not at all scalable. After a web search of online scotch retailers, twenty candidates were selected. We filtered each retailer's inventory by scotch whisky and scraped the results into text files. From there, it was mostly a manual cleaning effort, using Vim and regular expressions. Once the data had been organized into simple CSVs, a Python script was used to insert the data into a normalized SQLite database.

After the data entry was complete, I meticulously went through and consolidated duplicate whiskies, distilleries, and producers by hand. The entire process took over a week to complete. In order to scale this up, automated methods would be necessary. Naive web scraping may not pass

muster here, and more sophisticated methods, including Natural Language Processing may be necessary. For this project, the one-time manual effort was sufficient.

2.1.2 Technical Notes

All statistical analysis was done in a Jupyter Notebook with Python. Plots were created using Seaborn and Matplotlib.

The CSVs were inserted into a SQLite database.

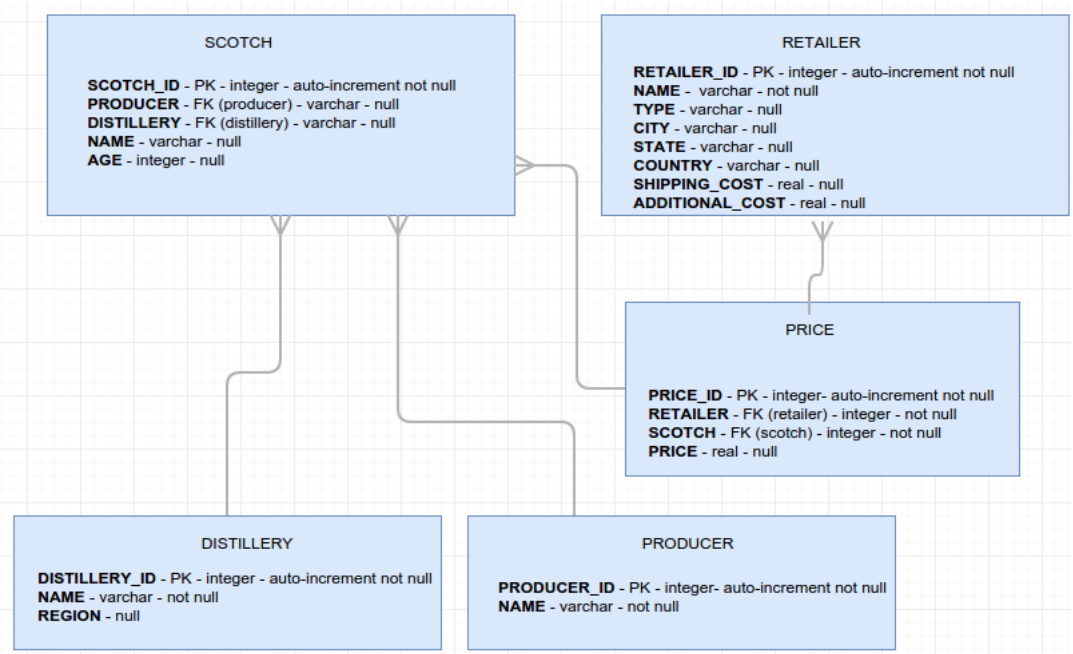

Figure 2.1: Scotch Database Schema

This entire project, complete with code, database, and LaTeXsource can be found at [13].

2.1.3 Caveats

One late-term regret was waiting until I had my full dataset to explore the data. This can introduce bias. A better approach would have been to load one retailer, explore the dataset, make hypotheses, load the rest, and see how my hypotheses hold up. [12] However, this was a tricky dataset to obtain, so it comes inherent with some imperfections.

2.2 Exploring the Data

2.2.1 Distribution of Age

Age of a whisky is the truncated amount of time in years that the youngest malt in the blend has spent in a barrel. This is a discrete random variable X such that $X \in \{3, 4, \ldots, 75\}$. By law, scotch must be aged at least 3 years, or else it is called a *spirit*. There is nothing saying that X cannot exceed 75, however, as of this publication, the oldest known whisky is a *Mortlach*, aged 75 years, produced by independent bottlers *Gordon & MachPhail*. Whisky evaporates as it is aged, so there is some upper bound - determined by physics - on how long it can remain in the barrel before it simply disappears. We will use the oldest known whisky as an upper bound on the population.

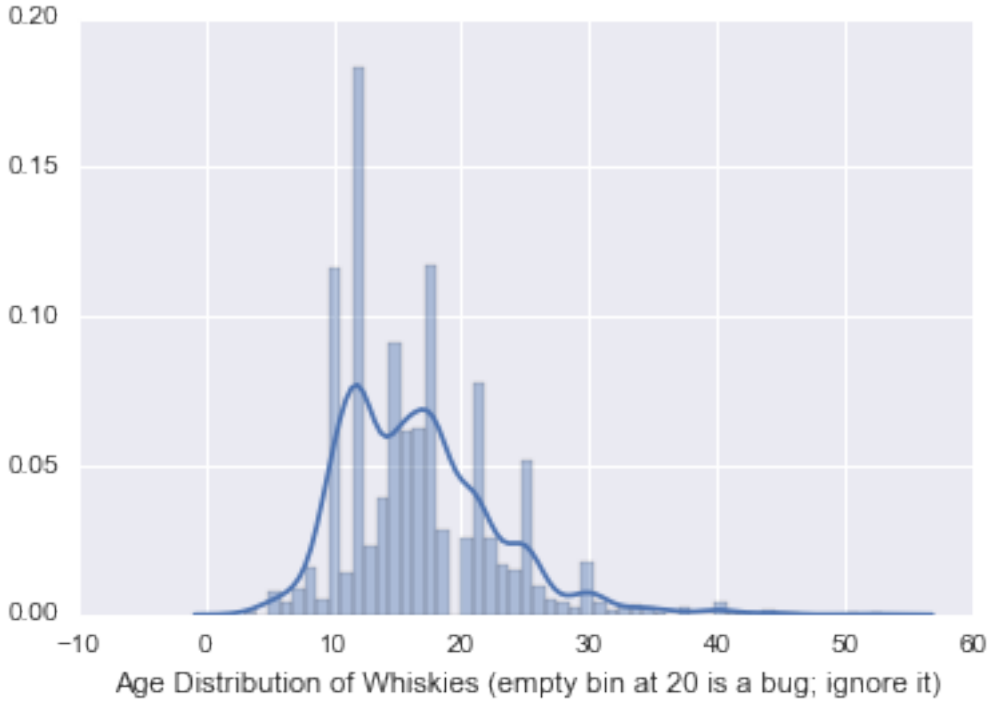

Figure 2.2: Age distribution of whiskies with age statements

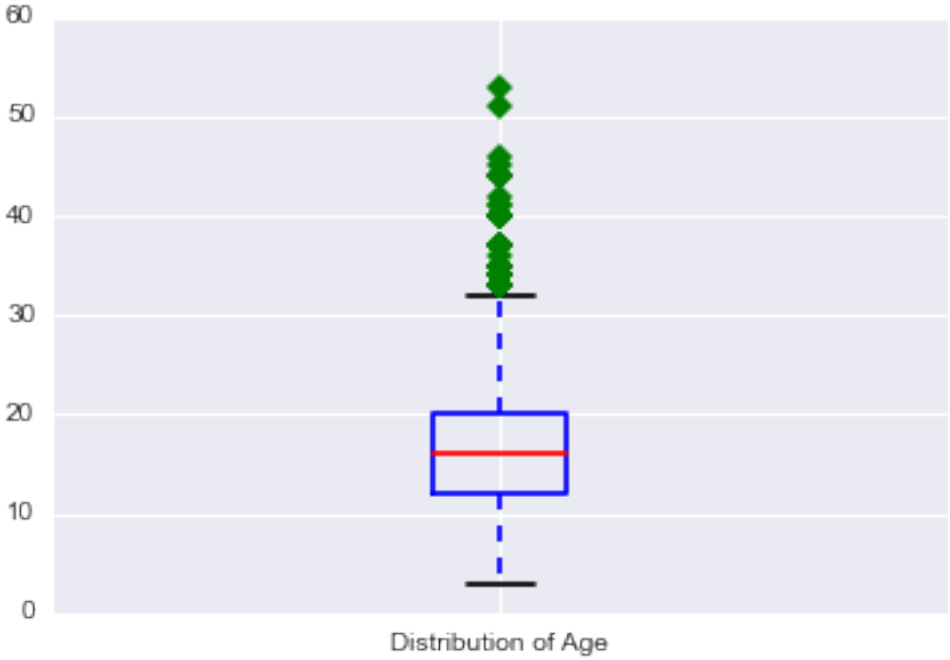

Figure 2.3: Age distribution of whiskies with age statements

From the density plot, it quite clear that our distribution is non-normal.

From the boxplot, we can see that most ages fall between 12 and 21 years old, but there are a good amount of outliers that will skew the distribution.

It is interesting to note that twelve year old scotches dominate, accounting for over 17% of the market. This is likely the "sweet spot" for whiskies with age statements, where distillers are able to minimize age and maximize quality and flavor. This is not true for all whiskies, of course, as flavor profiles differ for each whisky. However, in this quantitative analysis, we are interested generalizing and estimating the population. Randomness is inherent.

Year	Frequency (percent)
10 year olds	10.889292196
12 year olds	17.0961887477
15 year olds	8.42105263158
18 year olds	11.0344827586
21 year olds	7.36842105263
25 year olds	5.00907441016
30 year olds	1.70598911071

We obtain the following statistics from our sample distribution of ages:

$$\begin{aligned} n &= 2755 \\ \overline{x} &= 15.6134946314 \\ \widetilde{x} &= 16.0 \\ s^2 &= 36.682608369 \\ s &= 6.05551760361 \\ \text{Skewness} &= 1.15380146462 \end{aligned}$$

2.2.2 Distribution of Price

We could reasonably expect most of the prices to be in the more affordable region, as that's what sells best. After all, there are far more *consumers* of whisky than collectors. There is a reason comparatively cheaper blends like Johnny Walker are the best selling scotches in the world. The economics dictate that the ROI will be stronger with quality products in the affordable region, and a few flashy showcase products to drum up envy and garner attention to the brand. We will treat price as a non-negative continuous random variable.

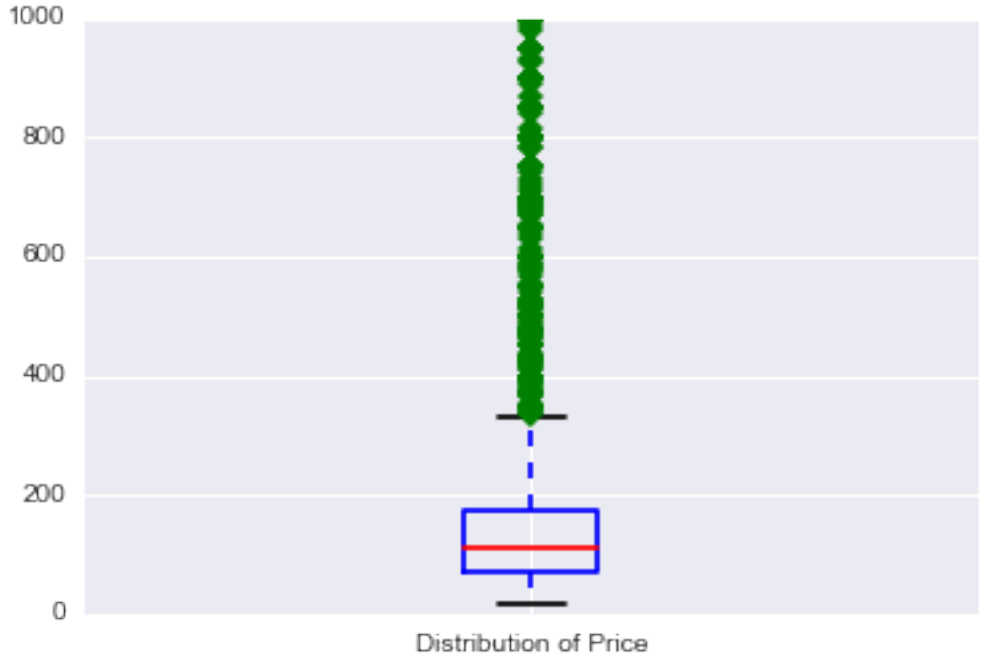

Figure 2.4: Price distribution of whiskies with age statements

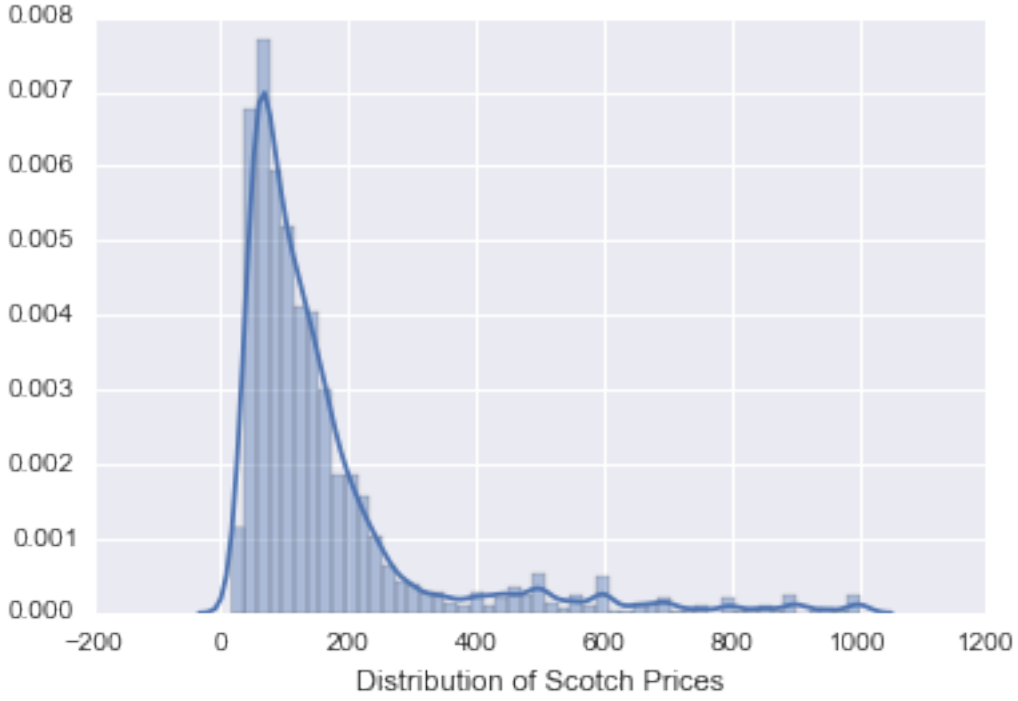

Figure 2.5: Whisky price distribution

We have the following statistics from the price distribution:

$$
\begin{aligned}
n &= 2755 \\
\overline{x} &= 119.165987296 \\
\widetilde{x} &= 110.0 \\
s^2 &= 26827.9806971 \\
s &= 163.762763695 \\
\text{Skewness} &= 2.72715134235
\end{aligned}
$$

Skewness is greater than 1, which indicates that this is heavily skewed to the right, so we cannot assume a normal distribution. Instead, what we have is closer to a **log-normal distribution**. This still has some skew to the right, but more closely resembles a bell curve.

Figure 2.6: Whisky price distribution as Log-Normal

The log-normal distribution is a transformation of the normal, written as $\ln N(\mu, \sigma^2)$.

The log-normal has the following moments:

$$\mu_{LN} = e^{\mu + \frac{\sigma^2}{2}}$$
$$\sigma^2_{LN} = (e^{\sigma^2} - 1)(e^{2\mu + \sigma^2})$$

Using a P-P plot, we can see how closely the log-normal follows a normal distribution. With an R-squared of 0.9662, it indicates a close goodness of fit. The bump in the tail at the high end is due to the bump in our pdf, where the prices have a slight increase in frequency in the \$600-\$700 range. No real world model matches a distribution 100%, so we will accept this small imperfection.

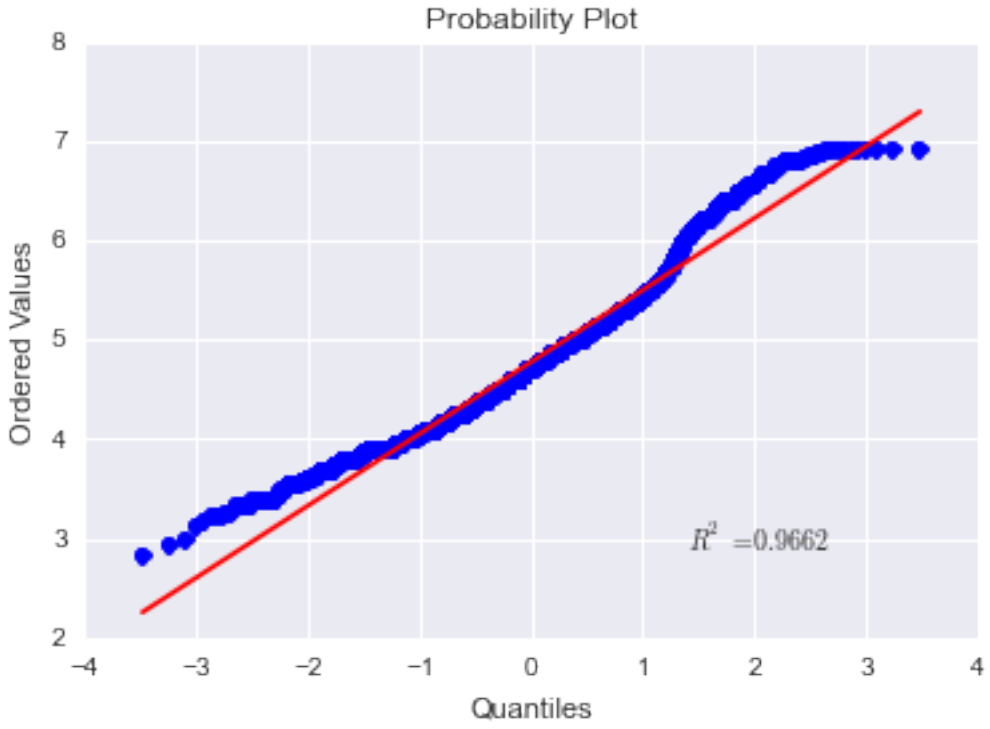

Figure 2.7: P-P Plot of Whisky Prices vs. Log-Normal

2.2.3 Point Estimates

Let $X_1 \dots X_n$ be independent random variables having the log-normal distribution, $Y_i \sim \ln N(\theta_1, \theta_2^2)$. Then the transformation $Y_i = \ln X_i, i = 1, \dots, n$ has the normal distribution, $N(\mu, \sigma^2)$ with mean μ and variance σ^2, such that

$$\theta_1 = e^{\mu + \frac{\sigma^2}{2}}$$
$$\theta_2^2 = e^{2\mu+\sigma^2} \cdot (e^{\sigma^2} - 1)$$

Because of its relation to the normal, it is clear that $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} \ln X_i$ and $S_Y^2 = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$.

Using the result from Finney (1941), as mentioned by Shen (1998), we cat get unbiased MVUE estimators for θ_1 and θ_2^2 by first defining the infinite series,

$$f(t) = 1 + t + \frac{n-1}{n+1}\frac{t^2}{2!} + \frac{(n-1)^2}{(n+1)(n+3)}\frac{t^3}{3!} + \dots \tag{2.1}$$

The adjusted unbiased MLEs are

$$\hat{\theta}_{MLE} = e^{\overline{Y}} \cdot f\left(\frac{1}{2n} S_Y^2\right) \tag{2.2}$$

$$\hat{\theta^2}_{2MLE} = e^{\overline{2Y}} \cdot \left[f\left(\frac{2}{n} S_Y^2\right) - f\left(\frac{n-2}{n(n-1)} S_Y^2\right) \right] \tag{2.3}$$

Their variances are

$$Var(\hat{\theta}_{1MLE}) = \frac{1}{n}\left(\sigma^2 + \frac{1}{2}\sigma^4\right) e^{2\mu+\sigma^2} \tag{2.4}$$

$$Var(\hat{\theta^2}_{2MLE}) = \frac{2\sigma^2}{n} e^{4\mu+2\sigma^2} \left[2(e^{\sigma^2} - 1)^2 + \sigma^2(2e^{\sigma^2} - 1)^2 \right] \tag{2.5}$$

And since $Var(\hat{\theta}_{1MLE}) \to 0$ and $Var(\hat{\theta^2}_{2MLE}) \to 0$ as $n \to \infty$, the estimators are consistent. [8] [9]

2.2.4 Confidence Intervals of Particular Scotch Prices

As what I will call "research" for this project, I purchased a few bottles of scotch and took note of their prices. I wondered if I had gotten a fair deal on my purchases.

Among the acquisitions were an Auchentoshan Three Wood (no age statement) for $63.00, a McClelland's Islay (also no age statement) for $24.99, and a Laphroaig 10 Year Old, for $50.00.

We will compute 90% confidence intervals to see if my purchases were within the expected mean. We will use the t-distribution, since our dataset only contains up to at most, 17 sample prices per scotch. [1]

$$\frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim T(n-1)$$

We have the following values

Auchentoshan Three Wood	McClelland's Islay	Laphroaig 10 Year Old
55.99	24.99	39.98
64.98	37.95	39.99
64.99		42.99
66.99		44.98
68.95		44.99
69.99		45.09
69.99		48.99
69.99		49.99
74.99		49.99
79.99		50.95
79.99		64.99

Computing each separately, we find the mean and variance for the Auchentoshan to be

$$\begin{aligned} n_{\text{Auchentoshan}} &= 11 \\ \overline{x}_{\text{Auchentoshan}} &= 69.71 \\ s_{\text{Auchentoshan}} &= 6.93 \\ \alpha &= 0.10 \\ t_{0.05,10} &= 1.812 \end{aligned}$$

[1] Due to some data cleaning issues, some scotch prices have been misplaced or incorrectly attributed. These three have been thoroughly inspected and cleaned, however.

The 90% confidence interval for the mean price of Auchentoshan Three Wood is

$$\left(\overline{X} - t_{\frac{\alpha}{2},n-1} \cdot \frac{S}{\sqrt{n}}, \overline{X} + t_{\frac{\alpha}{2},n-1} \cdot \frac{S}{\sqrt{n}}\right)$$
$$\left(69.71 - 1.812 \cdot \frac{6.93}{\sqrt{11}}, 69.71 + 1.812 \cdot \frac{6.93}{\sqrt{11}}\right)$$
$$(69.71 - 3.786, 69.71 + 3.786)$$
$$(65.92, 73.50)$$

Since I paid \$63.00, I can feel confident that I got a below average price for my bottle, and I am likely to return to that particular vendor.

Calculating the mean and variance for the McClelland's, we get

$$\begin{aligned} n_{\text{McLellands}} &= 2 \\ \overline{x}_{\text{McLellands}} &= 31.47 \\ s_{\text{McLellands}} &= 2.920 \\ \alpha &= 0.10 \\ t_{0.05,1} &= 6.314 \end{aligned}$$

The 90% confidence interval for the mean price of McClelland's Islay is

$$\left(\overline{X} - t_{\frac{\alpha}{2},n-1} \cdot \frac{S}{\sqrt{n}}, \overline{X} + t_{\frac{\alpha}{2},n-1} \cdot \frac{S}{\sqrt{n}}\right)$$
$$\left(31.47 - 6.314 \cdot \frac{2.920}{\sqrt{2}}, 31.47 + 6.314 \cdot \frac{2.920}{\sqrt{2}}\right)$$
$$(31.47 - 13.037, 31.47 + 13.037)$$
$$(18.43, 44.50)$$

I paid \$25.00 for my bottle of McClelland's Islay (not a particularly good bottle either - a far-too-young Bowmore that could use a few more years in the barrel), and I can feel comfortable that I paid about average for my bottle, though more samples could tighten up that interval to be sure.

Finally, calculating the mean and variance for the Laphroaig, we find

$$\begin{aligned} n_{\text{Laphroaig}} &= 11 \\ \overline{x}_{\text{Laphroaig}} &= 47.54 \\ s_{\text{Laphroaig}} &= 6.96 \\ \alpha &= 0.10 \\ t_{0.05,10} &= 1.812 \end{aligned}$$

The 90% confidence interval for the mean price of Laphroaig 10 Year Old is

$$\left(\overline{X} - t_{\frac{\alpha}{2},n-1} \cdot \frac{S}{\sqrt{n}}, \overline{X} + t_{\frac{\alpha}{2},n-1} \cdot \frac{S}{\sqrt{n}}\right)$$
$$\left(47.54 - 1.812 \cdot \frac{6.96}{\sqrt{11}}, 47.54 + 1.812 \cdot \frac{6.96}{\sqrt{11}}\right)$$
$$(47.54 - 3.80, 47.54 + 3.80)$$
$$(43.74, 51.34)$$

I paid \$50.00 for my bottle, which lands a little on the high side of the confidence interval, but still within it, so I can be confident I paid about average for it. And while it is nice to find bargains, we should be content knowing that we are not overpaying.

2.2.5 Regression and Correlation

Figure 2.8: Whiskies plotted by age vs. price showing a few significant outliers

It is instantly clear that the majority of whiskies are priced below $5000, but that there are some significant outliers. The majority of the outliers can be attributed to Macallan, which accounts for half of all whiskies over $5000 in our samplings. The next closest distillery in priciness was Mortlach, which only accounted for 11% of all whiskies over $5000. Macallan is widely regarded as the "Rolls Royce" of whiskies and it shows in their prices.

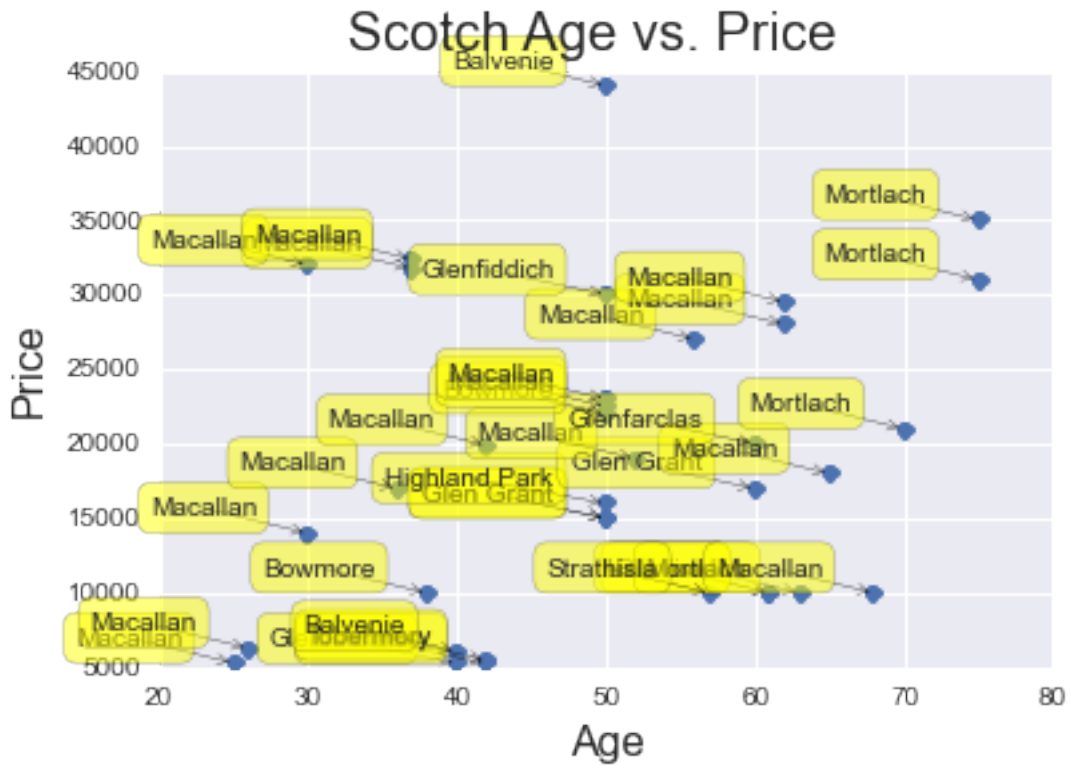

Figure 2.9: Age vs. Price - Outliers

Since the more digestible range lies in the sub-$5k category, let's zoom in on that.

We can start to see a clear pattern here, though we are still very scattershot, particularly as the age progresses into the 20s and beyond. Notice, however, that there are some very old whiskies that are cheaper than some much younger whiskies. Just from this chart we can pick out an almost-60 year old whisky that is cheaper than a few late-teens whiskies. Right away, we can say that age does not imply price. It should start to become clear, that age is only one of many factors that are taken into account when pricing a whisky.

Figure 2.10: Whiskies plotted by age vs. price as a pattern starts to emerge

We can see a very solid grouping below $1000, so let's zoom in on that and add some linear regression.

The regression line is given by

$$\hat{y} = 19.7547958954 \cdot x - 165.573087176$$

I will once again caution against buying a whisky based solely on its age description, but this regression line would certainly be a good litmus test when purchasing a whisky that is not rare or collectible. If I'm buying a 21 year old whisky, I could expect my price to be around $250. If I can find one for under $200, it's worth considering. And yet, this chart shows there are a good number of sub-$200 scotches in the 21 year old category.

At this point, I wondered, "What are the best bargain whiskies?" That is, what whiskies are under $50 and over 12 years old? The full results are listed in Table I.

Having personally tried both the *Glenlivet* 15 year old, and the *Glenfiddich* 15 year old, I would easily pay $50 for either. While these are some of the most accessible scotches in the world (you can probably find these

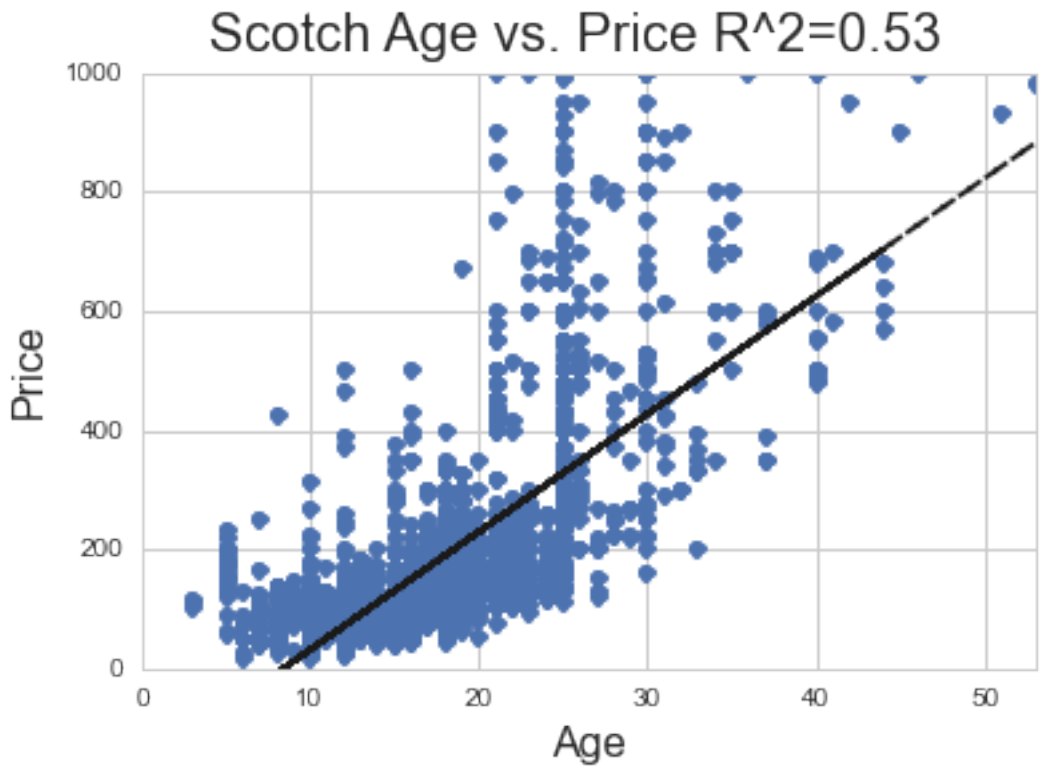

Figure 2.11: Whiskies plotted by age vs. price with linear regression

at your local dive bar), it is worth noting that collectability is not always the objective. You may want a decent scotch for more frequent drinking, or possibly giving away as a gift. These are solid choices.

The *Unknown* distilleries are typically store-brand bottles, by retailers such as Costco and Trader Joe's. While these can often be diamonds in the rough, it is prudent to proceed with caution.

One of the more interesting results from this study was that while the most expensive whiskies per age group were very noisy, the cheapest price per age fits a beautiful quadratic curve.

$$\hat{y} = 0.83003484 \cdot x^2 - 18.5544543 \cdot x + 120.39902851$$

2.2.6 Testing Claims

K&L Wines sells the ubiquitous Glenlivet 12 Year Old for $27.29. As we've seen so far, this is on the extreme low side, particularly for a 12 year old

Figure 2.12: Cheapest whiskies by age

whisky. They claim, however, that it is $37 elsewhere, implying that is the mean price of this particular whisky.[10]

We can easily test that claim against the alternative hypothesis that the mean price of The Glenlivet 12 Year Old is less than $37.00.

We pull the following eleven prices from various retailers in our database: $36.98, $27.99, $29.99, $39.99, $27.99, $27.99, $29.99, $32.95, $34.99, $39.99, and $44.99.

$$
\begin{aligned}
H_0 &: \mu = \mu_0 \\
H_1 &: \mu < \mu_0 \\
n &= 11 \\
\mu_0 &= \$37.00 \\
\overline{x} &= 33.98\overline{54} \\
s &= 5.59237689528 \\
\alpha &= 0.05 \\
t_{10,0.05} &= -1.812
\end{aligned}
$$

Using our test statistic, we find

$$\frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq -t_{10,0.05}$$
$$\frac{33.98\overline{54} - 37.00}{\frac{5.59237689528}{\sqrt{11}}} \leq -1.812$$
$$-1.78781158232 \leq -1.812 \boxtimes$$

We do not have enough evidence to reject their claim at the 0.05 level of significance. Their claim does not appear to be exaggerated.

I often come across "listicles" or reviews of scotch and I usually scoff at their claims of the prices of scotch. There aren't any liquor stores in my neighborhood, so the very small selection of bottles the local grocers have are extremely marked up when compared to prices claimed by various online articles. However, since we have a dataset, we can now test these claims statistically, rather than by assumption.

Thrillist, a lifestyle and leisure website published an article entitled *10 Scotches Under $50 That You Should Be Drinking*. One claim I found extraordinary, was that a Highland Park 12 Year Old can be had for just $46.

> This scotch is like a solar eclipse. One just can't help but stare at the color of Highland Park; it's a deep warm amber that you can almost taste. The 12-year-old's heather-honey taste is attributed to the time its spent in ex-sherry casks—which gives this bottle some individuality and depth. All the good for $46. [11]

I am not the only one dubious of this claim. In the comments, reader Noah Bieber says, *"Highland park for $46? Yea good luck with that..."*

Let's test this claim, against the alternative that Highland Park 12 Year Old is actual more expensive on average than $46.

We find the following prices in our database: $44.99, $44.99, $45.99, $46.99, $49.98, $49.99, $51.95, $52.99, $54.0, $59.99, $59.99, $59.99, and $64.99.

Surprisingly, Highland Park 12 Year Old can be found for as low as $44.99. If we are interpreting *Thrillist's* claims as the lowest price, then they have already been proven right. But the article does not link to the particular retailers selling at this price, so we will make the assumption that the claim is for the average market value.

$$
\begin{aligned}
H_0 &: \mu = \mu_0 \\
H_1 &: \mu > \mu_0 \\
\mu_0 &= \$46.00 \\
n &= 13 \\
\overline{x} &= 52.8330769231 \\
s &= 6.34787473171 \\
\alpha &= 0.05 \\
t_{12,0.05} &= 1.782
\end{aligned}
$$

Using our test statistic, we find

$$
\begin{aligned}
\frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} &\geq -t_{12,0.05} \\
\frac{52.8330769231 - 46.00}{\frac{6.34787473171}{\sqrt{13}}} &\geq 1.782 \\
3.88114294258 &\geq 1.782 \checkmark
\end{aligned}
$$

We can definitely reject the claim that on average, a Highland Park 12 Year Old can be found for \$46.00 with a p-value < 0.005.

2.3 Ranking Online Retailers

17 retailers were considered in this study. Three additional retailers were rejected because either they did not have the option of shipping out of state, or the customs import fees were too high to consider. Furthermore, UK based retailers deal primarily in 700 ml bottles, which is the standard volume for the UK, making comparisons slightly skewed. There are undoubtedly many other retailers out there, including local retailers that may have better prices, but this study gives a good feel for the competition and price spread.

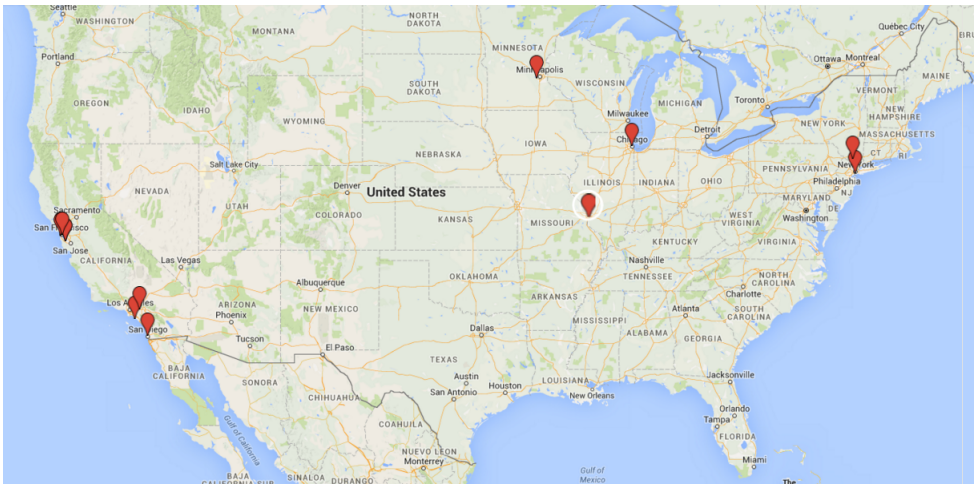

Figure 2.13: Brick and mortar locations of online retailers. Caskers, Love Scotch, Pacific Online Spirits, and Ultimate Wine Shop have no brick and mortar shops.

While the primary goal of this paper is to meet the requirements of MATH 8756 (Probability and Statistics II), a secondary personal goal was to find the cheapest online retailers of scotch in general. To do this, I developed an algorithm that makes use of the data and compares each online retailer with the others and gives each one an overall score. The smaller the score the better the prices when compared with the other retailers.

The algorithm uses two matrices; $\mathbf{P}$, the price matrix, which holds all of the prices we want to compare, and $\mathbf{C}$, the comparison matrix, which holds the scores of the retailers and is used to compute the final scores.

$$\mathbf{P} = \left[\begin{array}{c|cccc} s_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ s_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{m1} & x_{m2} & x_{m3} & \dots & x_{mn} \end{array}\right]$$

The first column of s variables is the scotch ID. The other columns are for the retailers, and the x variables are the prices for each retailer pertaining to the scotch ID.

$$\mathbf{C} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & \dots & x_{mn} \end{bmatrix}$$

We begin by initializing the $\mathbf{C}$ matrix to all zeroes,

$$\mathbf{C} := 0 \cdot \mathbf{C}$$

Then, we compare each retailer with the other retailers by scotch. If one retailer does not have a price for a particular scotch, we do not compare the two.

$j \leftarrow 1$
for $j \leq \text{Len}(\mathbf{P}_0)$ **do**
  $k \leftarrow j$
  for $k \leq \text{Len}(\mathbf{P}_0)$ **do**
    $i \leftarrow 0$
    for $i \leq \text{Len}(\mathbf{P})$ **do**
      $\mathbf{C}_{j-1,k-1} \leftarrow \mathbf{C}_{j-1,k-1} + (\mathbf{P}_{ij} - \mathbf{P}_{ik})$
      $\mathbf{C}_{k-1,j-1} \leftarrow -(\mathbf{C}_{k-1,j-1} + (\mathbf{P}_{ij} - \mathbf{P}_{ik}))$
      $i \leftarrow i + 1$
    end for
    $k \leftarrow k + 1$
  end for
  $j \leftarrow j + 1$
end for

What we end up with is the hollow $N \times N$ matrix $\mathbf{C}$, with

$$tr(C) = \sum_{i=1}^{N} a_{ii} = 0$$

By summing the rows, we end up with a rank for each retailer, $r_i \in \mathbb{R}$, such that $min(r_i)$ reflects the retailer with the lowest overall prices.

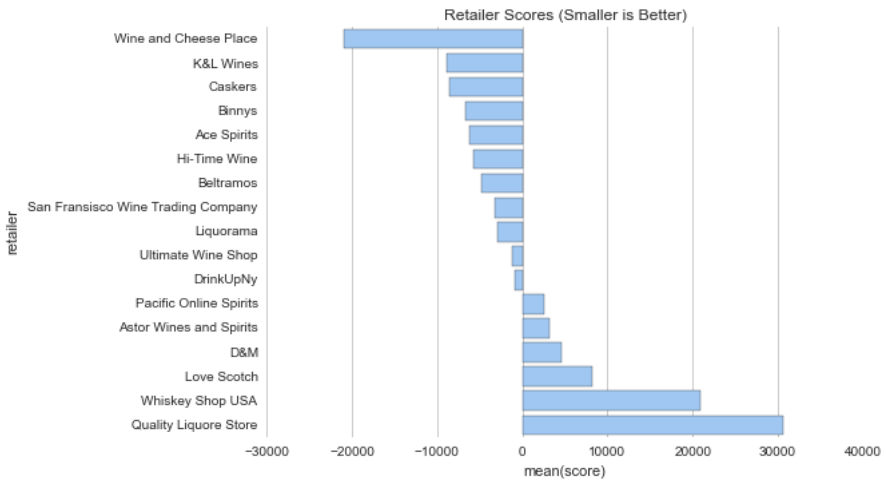

Figure 2.14: Results of retailer ranking algorithm

In this case, Wine and Cheese Place has the best all around prices, and Quality Liquore Store has the worst. Price is only one factor, however, as there are several opportunities for add-on charges that were not accounted for here, including taxes, shipping, discounts for larger orders, and specials. The winner here, for instance, charges $24.99 out of state FedEx shipping. An online-only retailer such as Caskers may end up beating them out in the all-in price. If you happen to live near St. Louis, however, selecting the in-store pickup option can save you a lot of money. The bottom line is, this covers one metric. Make sure to shop around to find the best deals.

3
Future Work

It should be clear by this point, that while age does not imply price, prices can vary wildly among retailers. This kind of information would be incredibly valuable to consumers in the form of either a web app or mobile app. The ability to scan a barcode in-store and get an instant price comparison or to type in a whisky and get a list of prices in ascending order would be very useful for consumers.

In order for this to happen, the infrastructure would need to be more sophisticated than what I did here. The data capture and cleaning methods were intended for one-time use, but for an app like this, prices would need to be updated much more frequently and in an automated fashion.

I suggested Natural Language Processing earlier. This is because scotches don't have easy to categorize names like beer or wine. There are complications like independent bottlers, vintages, and collectible one-offs. There are cask strength versions and different cask type variations. Narrowing this down to a single SKU for each bottle becomes even more difficult when online retailers don't use a unified naming convention.

I was unable to successfully match many bottle listings by hand, so it would be interesting to see how something like a Bayesian text classifier could do.

The findings here open the door to more research, and since this project is open source, it is free to grow.

4
Conclusion

In this paper, first and foremost, we saw that age does not imply price. But we also saw that the lower bound of price by age can be fit to a quadratic.

We also saw that prices can vary wildly by age and by retailer, but using the algorithm presented in 2.3, we have a list of places to look in order of best to worst. Of course, only the prices of the bottles were considered, not shipping, sales, minimum orders, etc. So while the algorithm doesn't exactly make a decision for you, it does make your decision easier, and could, in the future, be made to make the decision entirely.

References

[1] The Macallan 64 Years Old in Lalique Cire Perdue, http://us.themacallan.com/the-whisky/limited-releases/the-macallan-64-years-old-in-lalique-cire-perdue/

[2] Whiskey Database, http://whiskyanalysis.com/index.php/methodology-introduction/methodology-biases-and-limitations/

[3] Whisky Advocate, http://whiskyadvocate.com/2010/06/28/what-does-a-whiskys-age-really-mean/

[4] Whisky.com, https://www.whisky.com/whisky-database/database.html

[5] Whiskeystats.net, http://www.whiskystats.net/

[6] WhiskyAnalysis.com http://whiskyanalysis.com/

[7] Whiskybase.com Maps, https://www.whiskybase.com/maps

[8] D. J. Finney, On the distribution of a variate whose logarithm is normally distributed, J. Roy. Statist. Soc. Ser. B, **7** (1941), pp. 155-161

[9] Wei-Hsiung Shen, Estimation of Parameters of a Lognormal Distribution, Taiwanese Journal of Mathematics, Vol.2, No. 2, pp. 243-250, June 1998

[10] K&L Wines,The Glenlivet 12 Year Old http://www.klwines.com/p/i?i=620004

[11] Thrillist,10 Scotches Under $50 That You Should Be Drinking https://www.thrillist.com/vice/best-scotch-whisky-under-50-johnnie-walker-glenlivet-laphroaig

[12] Reinhart, Alex (2015). *Statistics Done Wrong*. San Francisco: No Starch Press, Inc.

[13] Detweiler, Brian. MATH 8756 Final Project, https://github.com/bdetweiler/math-8756-final-project

Appendices

A

Whiskies 12 Years or Older Under $50

Producer	Distillery	Age	Price	Retailer
Lismore	Unknown Speyside	15	39.99	Liquorama
Tullibardine	Tullibardine	14	39.99	Binnys
Lismore	Unknown Speyside	15	39.99	Hi-Time Wine
Lismore	Unknown Speyside	15	41.95	Love Scotch
William Maxwell and Co.		18	41.98	Ace Spirits
Glenfiddich	Glenfiddich	14	44.98	Hi-Time Wine
Glen Moray	Glen Moray	16	44.99	Binnys
Glenlivet	Glenlivet	15	44.99	Hi-Time Wine
Tomintoul	Tomintoul	16	44.99	Ultimate Wine Shop
Lismore	Unknown Speyside	15	44.99	Wine and Cheese
William Lundie and Co.		15	48.99	Ace Spirits
Glenfiddich	Glenfiddich	14	49.99	Liquorama
Lismore	Unknown Speyside	18	49.99	Liquorama
Glenlivet	Glenlivet	15	49.99	Liquorama
Craigellachie	Craigellachie	13	49.99	Astor
Tomatin	Tomatin	15	49.99	Quality Liquore Store
Glenfiddich	Glenfiddich	14	49.99	Binnys
Tullibardine	Tullibardine	14	49.99	Binnys
Glenfiddich	Glenfiddich	14	49.99	Beltramos
Glenlivet	Glenlivet	15	49.99	Beltramos
Craigellachie	Craigellachie	13	49.99	Caskers
Tobermory	Tobermory	15	49.99	Caskers
Dalwhinnie	Dalwhinnie	15	49.99	K&L Wines
Craigellachie	Craigellachie	13	49.99	Hi-Time Wine
Glenfiddich	Glenfiddich	15	49.99	Hi-Time Wine
Craigellachie	Craigellachie	13	49.99	Wine and Cheese
Glenfiddich	Glenfiddich	15	49.99	Wine and Cheese

Whiskies over 12 years old under $50

B

Whiskies over $5,000

Producer	Distillery	Age	Price
Signatory Vintage	Glenfarclas	40 years old	$5,249.99
Tobermory	Tobermory	42 years old	$5,250.0
Macallan	Macallan	25 years old	$5,299.0
Tobermory	Tobermory	42 years old	$5,499.0
Balvenie	Balvenie	40 years old	$6,069.95
Macallan	Macallan	26 years old	$6,299.0
Gordon & MacPhail	Glenlivet	61 years old	$9,999.0
Gordon & MacPhail	Linkwood	61 years old	$9,999.0
Gordon & MacPhail	Mortlach	63 years old	$9,999.0
Gordon & MacPhail	Strathisla	57 years old	$9,999.0
Macallan	Macallan	68 years old	$9,999.0
Bowmore	Bowmore	38 years old	$9,999.99
Macallan	Macallan	30 years old	$13,999.0
Glen Grant	Glen Grant	50 years old	$14,999.0
Glen Grant	Glen Grant	50 years old	$14,999.0
Highland Park	Highland Park	50 years old	$15,999.0
Gordon & MacPhail	Glen Grant	60 years old	$16,999.0
Macallan	Macallan	36 years old	$16,999.0
Macallan	Macallan	65 years old	$17,999.0
Macallan	Macallan	52 years old	$19,000.0
Duncan Taylor	Macallan	42 years old	$19,899.99
Glenfarclas	Glenfarclas	60 years old	$19,999.99
Mortlach	Mortlach	70 years old	$20,999.99
Bowmore	Bowmore	50 years old	$21,999.0
Macallan	Macallan	50 years old	$22,499.0
Macallan	Macallan	50 years old	$22,999.99
Macallan	Macallan	56 years old	$26,999.0
Macallan	Macallan	62 years old	$27,999.99
Macallan	Macallan	62 years old	$29,500.0
Glenfiddich	Glenfiddich	50 years old	$29,999.0
Gordon & MacPhail	Mortlach	75 years old	$31,000.0
Macallan	Macallan	37 years old	$31,748.99
Macallan	Macallan	30 years old	$31,999.99
Macallan	Macallan	37 years old	$32,374.99
Gordon & MacPhail	Mortlach	75 years old	$34,999.0
Balvenie	Balvenie	50 years old	$43,999.0