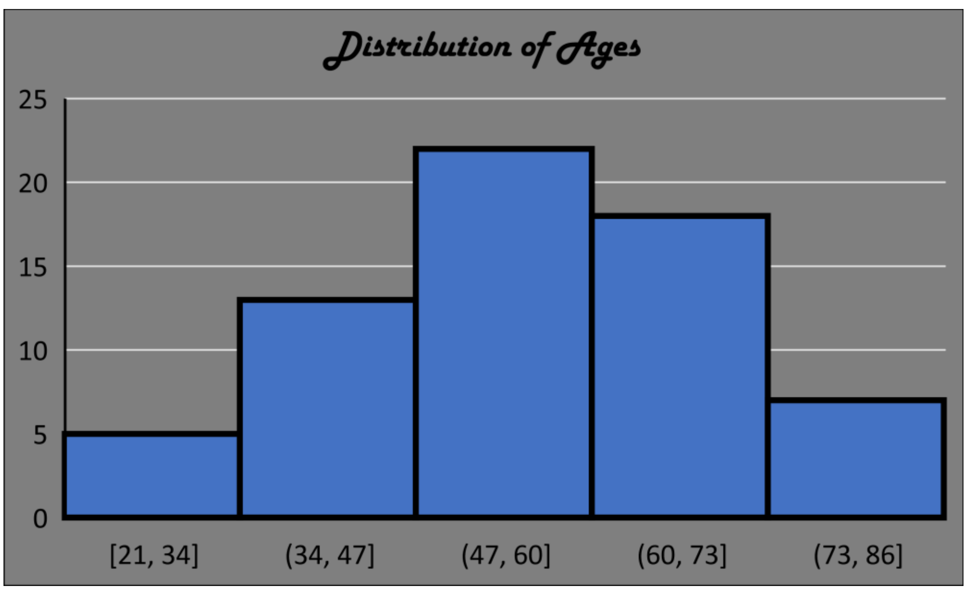
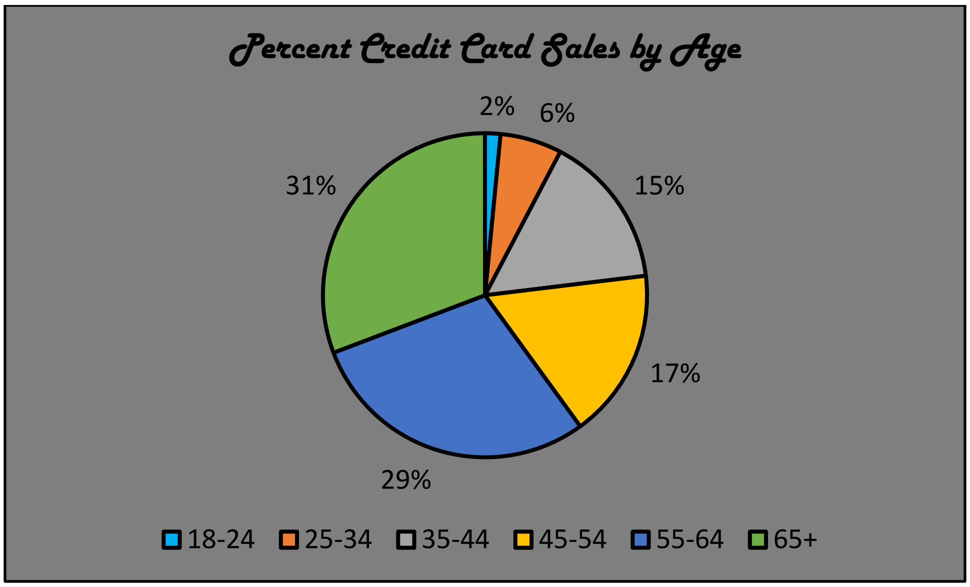
Credit Card Sales Analysis

Age

To start, we are going to look at the ages of our clients at time of credit card application. Our sample spans a range of ***64.87*** years with a minimum value of ***21.29*** and a maximum of ***86.16***. The distribution of ages is approximately normally distributed. Meaning that the data is roughly symmetrical, following a bell-shaped curve. A skewness of ***-0.12*** suggests that the data is very slightly left skewed, implying that there are outliers present in the younger age groups. Our median ***(58.25)*** being larger than our mean ***(56.84)*** further provides evidence for this. However, when calculating our lower outlier fence, we get a value of ***16*** which is much lower than our minimum value of ***21.29***. This implies that there are no true outliers and that our data is roughly symmetrical but with a slightly higher concentration of applicants in the older age ranges.



The average age of our applicants is ***56.84*** with a standard deviation of ***14.84***. This means that on average the data falls within about ***15*** years above or below the mean. When calculating a ***95% confidence interval*** for the mean, we get an interval of ***53 to 61*** years of age. Meaning that we can be 95% sure that our population parameter (μ) falls within this interval. When compared to the average national retirement age of 63, these numbers are astonishingly high. It appears that customers who are closer to retirement age are the ones applying more often than those in their teens, twenties, thirties, and forties. This is made even more apparent when looking at the age group breakdown as shown in the graph below. ***60%*** of credit card sales came from people aged 55 and older compared to a measly ***8%*** from the 18 to 34 age groups. Even more surprising is the ***31%*** coming from customers aged 65 and older, a count that includes ***five 80+ year olds.***



Let's now test one of our assumptions. Our assumption is that on average, older customers do not apply for credit cards. We'll set the cut off for an "older" customer equal to the national retirement age. This gives us the following hypotheses:

$$H_0: \mu \leq 63$$

$$H_a: \mu > 63$$

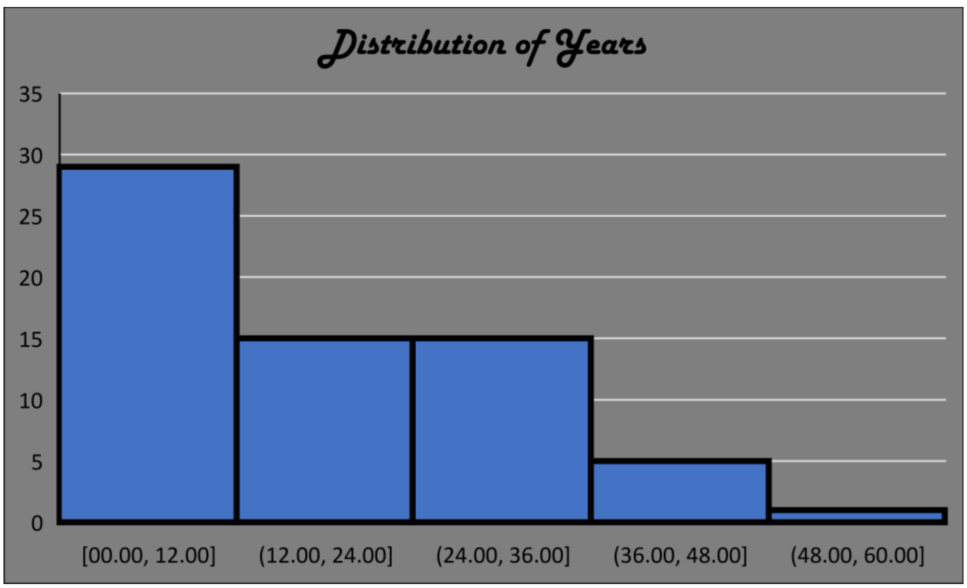
Our null hypothesis (aka the one we are testing or the status quo) is that the average age of our credit card customers is less than or equal to 63. Meanwhile our alternative hypothesis is that the average age is greater than 63. One important thing to note is that the result of our hypothesis test is in reference to the ***population*** mean rather than the sample mean. So just because our sample mean is less than 63 doesn't mean that the population mean can't be or isn't 63 or larger. We'll test our claim at the ***5% significance level***. We get the following results:

T-Score / Test Statistic	-3.3496818
Critical Value	1.669
P-Value	0.9993

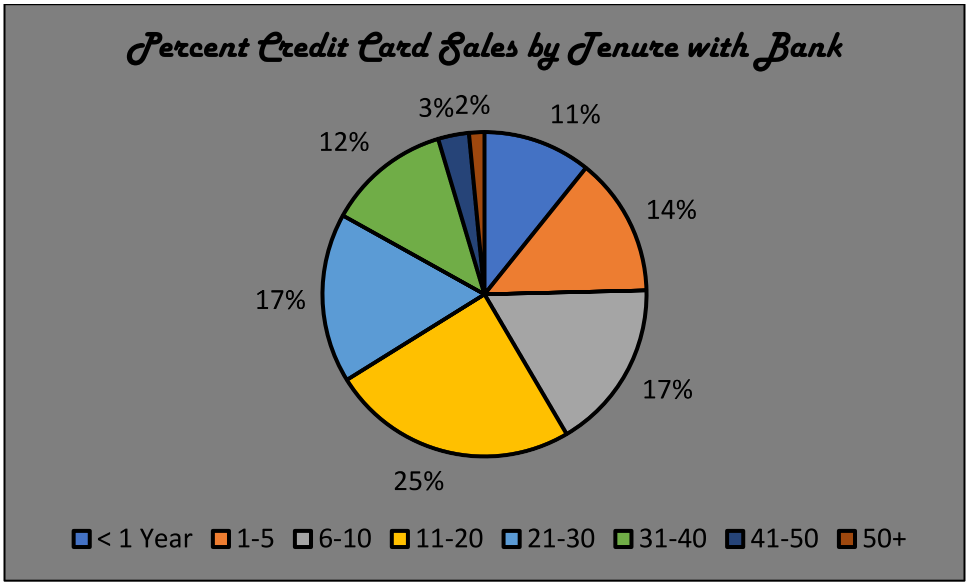
It appears that we do not have sufficient evidence to conclude that the population mean age is greater than 63. In fact, we have overwhelming evidence that the opposite is true. A hypothesis test of greater or less than is what is known as a one tailed test. During a one tailed test we are testing for statistical significance in only one direction (right/positive in this scenario) and completely disregarding the effect in the opposite direction. For us to be able to reject the null hypothesis then our test statistic would have had to have been greater than our critical value. In this case, our test statistic is significantly lower than our critical value ($T = -3.35 < t = 1.669$) so we fail to reject the null hypothesis. This makes sense as, if you remember back to our confidence interval, we calculated an upper bound of 61 years which is two years lower than our hypothesized mean. So, it would only make sense that when testing at the 5% significance level we would be unable to prove that the mean age is greater than 63.

Tenure with Bank

We will now look at the tenure of customers with the bank. Tenure is defined as the fractional number of years between when the customer established their relationship with the bank and when they applied for their credit card. This is calculated by subtracting their application date from their establishment date. Unlike the distribution of ages, the distribution of years is moderately right skewed. From looking at the graph, we observe that a decent proportion of customers are concentrated in the mid to lower year bins. A skewness of ***0.66*** supports this claim. However, like the distribution of ages, when calculating our upper outlier fence ***(60.37)*** we once again see that there are no true outliers. Instead, the data is just distributed in such a way that there is a higher concentration of applicants in the mid to lower bins. Our distribution covers a range of ***52.49*** years which is exactly equal to our maximum value. This makes sense as some customers choose to apply for a credit card at the time of their relationship establishment date. The average tenure of our applicants is ***16.59*** years with a standard deviation of ***13.23***. This converts to a ***95 % confidence interval*** of ***13 to 20 years***. However, seeing as our distribution exhibits a moderate amount of skew, it may be more beneficial to look at the median which, unlike the mean, is not affected by skew. By looking at the median in conjunction with the mean we can get a better sense of the central tendency of our data set. Our median value is ***13.88*** which, as expected, is lower than our mean value. When a data set is skewed, the median does a better job than the mean at capturing a more accurate/representative estimate of center. This is because the mean can be “pulled” by extreme values in either tail. So, when we compare the mean to the median, not only do we get a better idea of what the typical value in our data set is, but we also get an idea of by just how much the mean is being pulled and, as a result, what direction the distribution is skewed.



Our results are somewhat surprising. Seeing as a customer who has been with the bank for many years prior would likely have had many opportunities in the past to apply for a credit card, one might assume that if they hadn’t gotten one up until this point, they likely don’t want one. However, it appears this is not the case. Looking at the pie chart below, while we see a decent number of sales coming from our lower tenure groups, (***25%*** coming our customers in the less than 1- and 1-5-year groups) we also see a much larger amount coming from our customers in the longer tenured groups (***59%*** from 6 to 30 and ***17%*** from 31 to 50+). This includes a single category leading ***25%*** from the 11-20 group. So, while newer customers certainly do apply, it seems that being with the bank for several years prior has a negligible effect on a customer's decisions to apply.



Earlier we stated that we were surprised by the number of customers who had been with the bank for years prior to applying. Let's see if this is in fact normal. We will classify a "new" customer as a customer who has been with the bank for a maximum of 5 years. Giving us the following hypotheses:

$$H_0: \mu \leq 5$$

$$H_a: \mu > 5$$

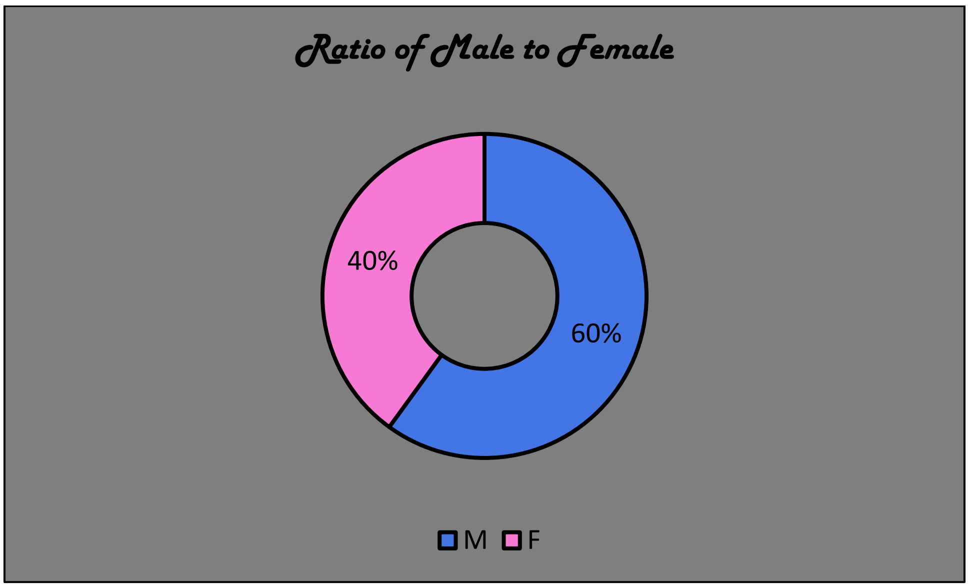
In our null hypothesis we are testing if the population average is less than or equal to 5 years and in our alternative hypothesis we are testing if the population mean is greater than 5 years. Once again, we will test at the ***5% significance level***. We get the following results:

T-Score / Test Statistic	7.07
Critical Value	1.669
P-Value	0.00001

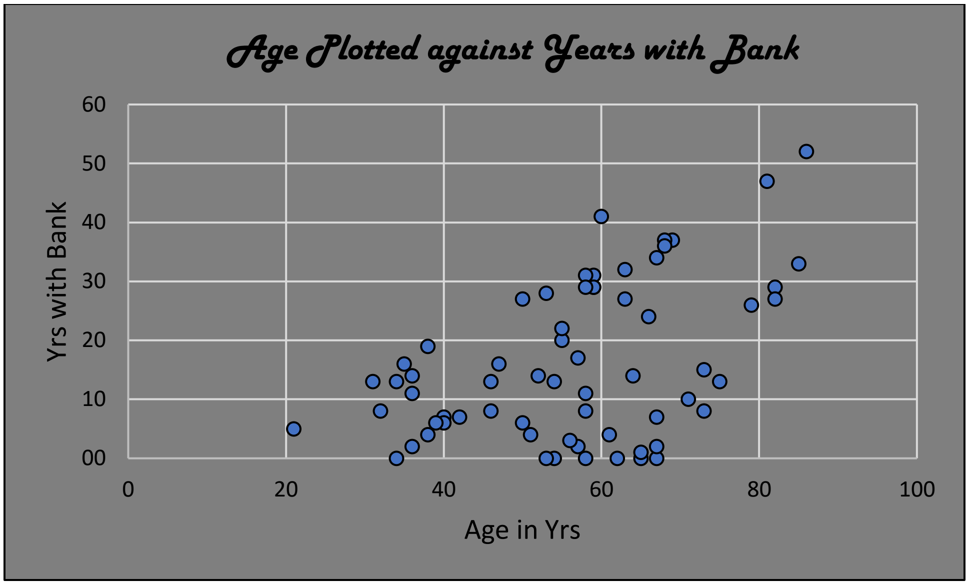
Our results are highly significant in favor of the alternative hypothesis. The p-value is the probability of obtaining a test statistic at least as extreme as the result observed. In other words, it is the probability that the data we got could be observed given that the null hypothesis is true. In our test we got a p-value that is near 0. Meaning that it is highly unlikely that our data could have occurred if the null hypothesis were true. With this in mind, we can safely reject the null in favor of the alternative hypothesis. ***At the 5% significance level there is no statistical evidence that the population mean tenure is less than or equal to 5 years.*** It appears that our original assumption that long time customers do not want credit cards is unsubstantiated.

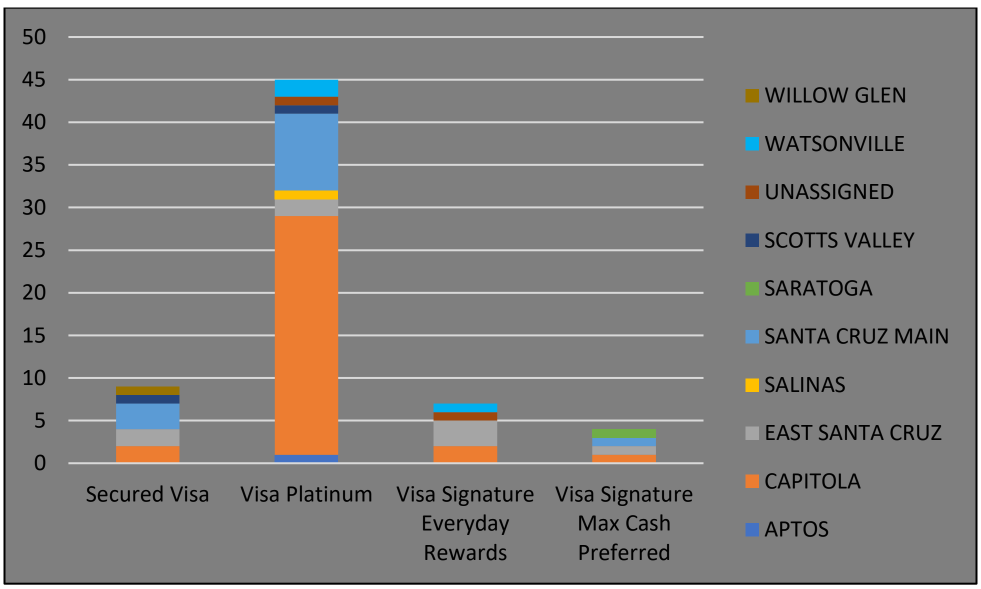
Other Miscellaneous Findings

Most of our applicants were male, equating to a 60-40 male-female split.



Age and tenure are positively correlated. Given that a customer must age in order to increase their tenure with the bank, this comes as no surprise. Looking at the graph below we see that as age increases, so does tenure. When calculating for the correlation coefficient we get positive ***0.47*** relationship. While the positive relationship certainly exists, it is not a very strong one. This is likely due to the fact that a person can establish an account at any age. Therefore, a person could theoretically be in their eighties but only be in their first year with bank.


Age Plotted against Years with Bank
Yrs with Bank
00
10
20
30
40
50
60
0
20
40
60
80
100
Age in Yrs


0
5
10
15
20
25
30
35
40
45
50
Secured Visa
Visa Platinum
Visa Signature Everyday Rewards
Visa Signature Max Cash Preferred
WILLOW GLEN
WATSONVILLE
UNASSIGNED
SCOTTS VALLEY
SARATOGA
SANTA CRUZ MAIN
SALINAS
EAST SANTA CRUZ
CAPITOLA
APTOS