Unveiling Secrets Hidden Behind Andy Warhol's Shot Marilyn Through a Statistical Perspective

Clayton Chan
clwchan@ucdavis.edu

Wesley Tat
wrtat@ucdavis.edu

Wendy Zhai
wyzhai@ucdavis.edu

Air Singh
ansing@ucdavis.edu

Abstract

Andy Warhol's Marilyn Monroe collection was a cultural phenomenon in the 1960s due to the artworks' color combinations and the artist's innovative approach. In this report, five of the most significant Marilyn Monroe painting were used and analyzed through a statistical perspective. Each pixel from each painting was broken down into RBG values and analyzed utilizing illustration such as RGB Density Plot Analysis, RGB correlation heatmap, and conditional entropy heatmap. Clustering techniques such as hierarchical clustering and K-means clustering were used to identify the key colors in a painting. Clustering pixels in the image is also a key step to recoloring Andy Warhol paintings. Each colored painting is recolored with two approaches, one attempts to recreate the painting to illustrate a more natural appearance, the other recreates the painting to illustrate a more artistic appearance. Finally, color transferring was attempted using the Andy Warhol painting and Monroe's original portrait.

1 Introduction

Andy Warhol, a leading figure in the pop culture movement, painted five images of Marilyn Monroe (MM) based on an image of Marilyn Monroe for the 1953 film Niagara known as Shot Marilyn.[1] Pop art is an art movement that emerged in the late 1950s and '60s that was inspired by commercial and popular culture.[2] Pop art is known for the vivid and iconic colors that define a painting. The five paintings of MM consist of one iconic color which were red, orange, green, aqua, and blue. Andy Warhol wanted to immortalize MM after her young death at 36 years old through these various paintings.[3]

The objective is to apply data science techniques to investigate the color composition of these paintings and explore the red, green, and blue color values. Extraction of data done on each painting will be collected for the purpose of color analysis. After the data analysis, the project will shed some light on Andy Warhol's artistic choices and reveal new perspectives on the five Marilyn Monroe paintings. To gain a deep comprehensive understanding can be reached by the use of data science techniques on Andy Warhol's Shot Marilyn Monroe paintings.

By employing hierarchical clustering and k-means clustering algorithms, we aim to find underlying patterns and groupings within the artworks based on their color characteristics. Hierarchical clustering is a powerful tool that allows for the examination of the relationship between the five paintings and determining their most prevalent figures while splitting them into layers. By constructing a hierarchical tree-like structure, we can identify groupings of paintings that share similar color profiles. Additionally, k-means clustering allows us to partition the dataset into a predetermined number of clusters, enabling a more granular analysis of color patterns within the Shot Marilyn Monroe paintings.

A comparison between hierarchical clustering and k-means clustering will also be accomplished by processing the five paintings to see the differences in how both techniques perform.

In the following sections, the outline of exploratory data analysis, clustering algorithms used, artistic interpretation, and the presentation of findings and insight will be shown while concluding an overall analysis of Warhol's iconic Shot Marilyn Monroe paintings.

2 Exploratory Data Analysis

In the exploratory data analysis, RGB Density Plot Analysis, RGB correlation heatmap, and conditional entropy heatmap will be explored. The purpose of each technique is written below:

The purpose of the RGB density plot for this project is to provide insights into the distribution of pixel values across the red, green, and blue color channels in the image. The RGB density plots will allow for the assessment of similarities and differences between all five Marilyn Monroe paintings created by Andy Warhol. Additionally, discernible patterns can be identified, and the relative intensities of all the images can be analyzed. The relative intensities can then be used to see the most dominant color that could help with color correction or image enhancement. In the context of Andy Warhol's Shot Marilyn Monroe paintings, the RGB density plot can show some insight into how Andy Warhol chose colors and his artistic style.

In this section, the purpose of 2-Dimensional (2D) Projection Plots was to see the geometry and to visualize the color distribution in an image. The RGB 2D Projection Plots have two colors plotted against one another with the third color on each point to show the intensity of the third color. Thus, the 2D Projection Plots have all three colors shown for color analysis and to observe patterns and relationships between the colors. Additionally, for this project, the 2D Projection Plots will help with understanding color composition and color intensity within each Marilyn Monroe painting.

The purpose of an RGB correlation heatmap plot is to visualize the relationships and correlations between red, green, and blue color channels in the five Marilyn Monroe paintings. The correlations and relationships will help identify any underlying patterns and trends between them. The color composition and color distribution can be found through the correlation plots. In the context of analyzing artwork, such as the Shot Marilyn Monroe paintings by Andy Warhol, an RGB correlation heatmap plot can help us explore the relationships between the color channels used by the artist. This visualization can provide valuable insights into the artist's color choices and preferences which will allow for the understanding of Warhol's artistic style and visual impact. Overall, an RGB correlation heatmap plot will uncover patterns within each painting of the Shot Marilyn Monroe.

The Shot Marilyn Monroe paintings will be shown below.

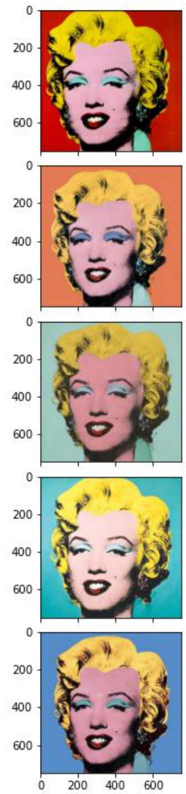

Figure 1: The Paintings of Red, Orange, Green, Aqua, and Blue Shot Marilyn

2.1 Red Shot Marilyn

2.1.1 RGB Density Plot

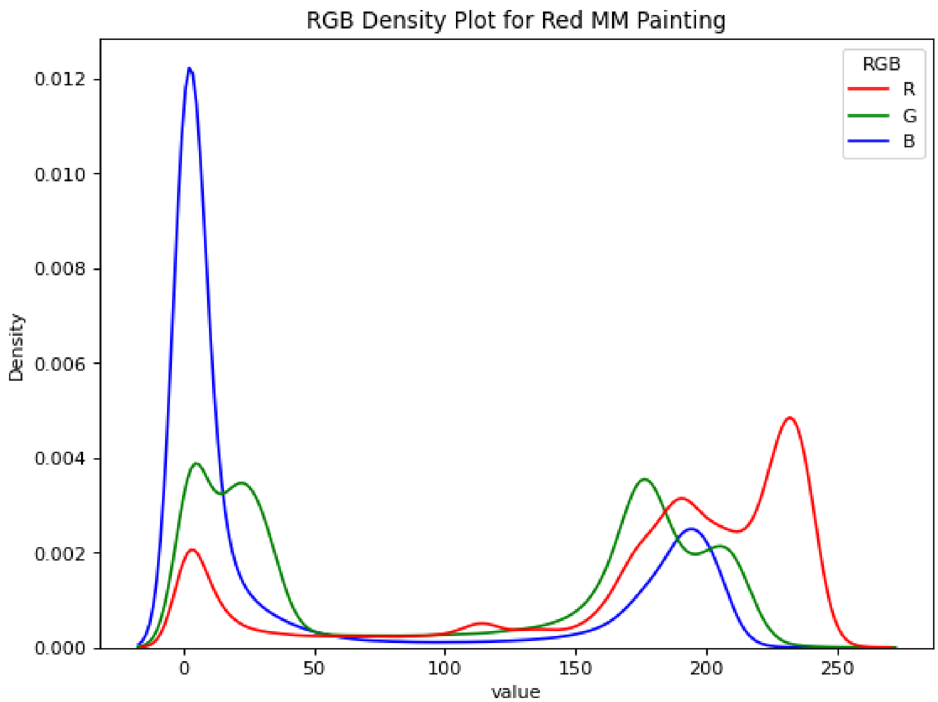

Figure 2: RGB Density Plot of Red Shot Marilyn

When looking at the RGB density plot for the Red Marilyn Monroe painting, there is a very large peak at 0 for blue that indicates how there are not as many blue tones or many blue colors in this painting. In the mid-range of the RGB values, the red and green curves are more uniform in their distribution. However, in the lower range of the density plot, green is more prevalent than red [discuss why it is that]. And in the upper range of the density plot, red is more prevalent than green. Blue stays low the entire range except for the end where it slightly peaks. The green slightly peaking at zero makes a lot of sense as the primary color in the Red MM painting is red and the painting does incorporate some green to help make the yellow/golden color hair. The large peak at 0 for blue is clear as this is the Red MM painting that primarily focuses on the color red.

2.1.2 RGB 2D Projection Plot

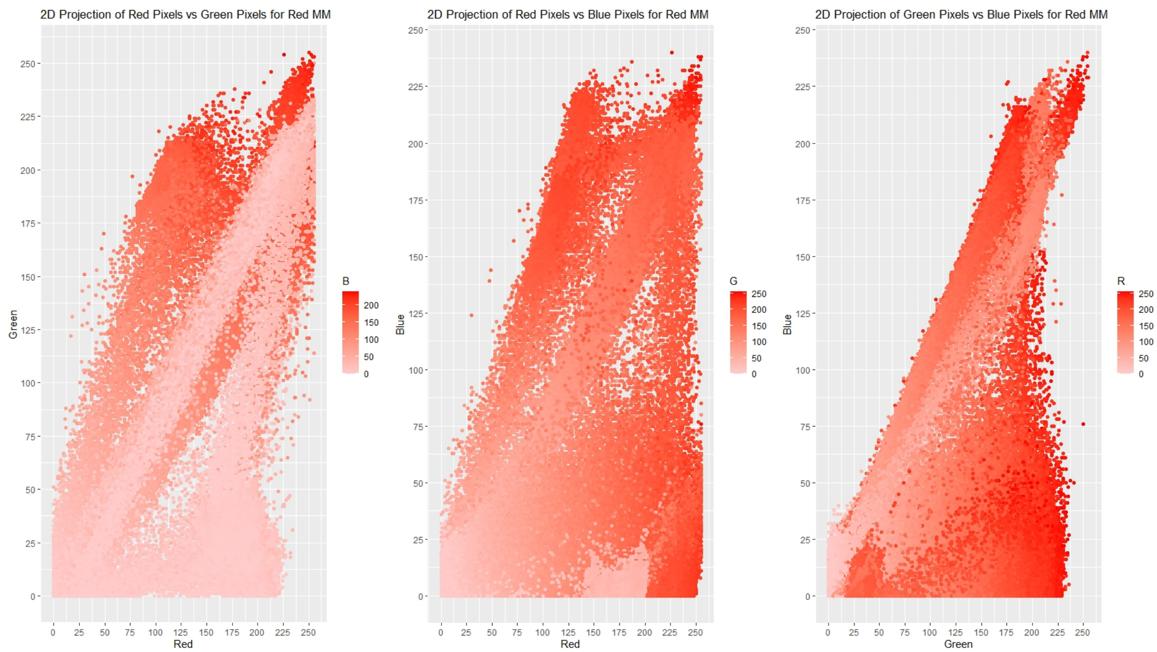

Figure 3: 2D Projection Plot of the Red Shot Marilyn

When looking at the 2D projection plot for the Red Marilyn Monroe Painting with Red plotted against Green and blue for intensity, the majority of the points on the projection are shown that blue is valued at 0 as there is not much blue in the painting. Additionally, there is a somewhat linear relationship between Red and Green where at the higher range of red and green, there is more intensity of blue. There is also a gap between red and green between the ranges at red from 75 to 175 and at green from 25 to 100. The distribution of colors across the plot does not seem uniform as there many clusters and a unique shape between the two.

When looking at the 2D projection plot for the Red Marilyn Monroe Painting with Red plotted versus Blue with green as the intensity, at most points in this graph, green is in the higher range of RGB values. Green is used more than blue based on the image. Additionally, there is somewhat of a cluster of points in a linear direction between red and blue, but there are a lot of points where red is at high intensity, and blue is in the lower range from 00 to 75. The points are scattered everything, there is no distinct distribution between red and blue. But for a red Marilyn Monroe painting, it is more logical for blue to not be as intense as the main focus is on blue.

When looking at the 2D projection plot for the Red Marilyn Monroe Painting with Green plotted versus Blue with red as the intensity, there is a very distinct linear relationship based on the cluster of points between blue and green. The intensity of red is in the lower range. The majority of the points fall under this certain distinct linear line. This makes sense as the eyeshadow is somewhat of a turquoise color along with the color on the back of Marilyn Monroe's back creating this distinct linear relationship between the green and blue. Green and blue do not reach the high end of intensity. The red is intense at most values between green and blue but at the lower ranges of green and blue where it is also low intense. As green increases in intensity and blue stays relatively low in intensity, there is an increase in red intensity. This shows that this growth of intensity of green and red with low intensity of blue is for the somewhat yellow hair.

2.1.3 RGB Correlation Heatmap

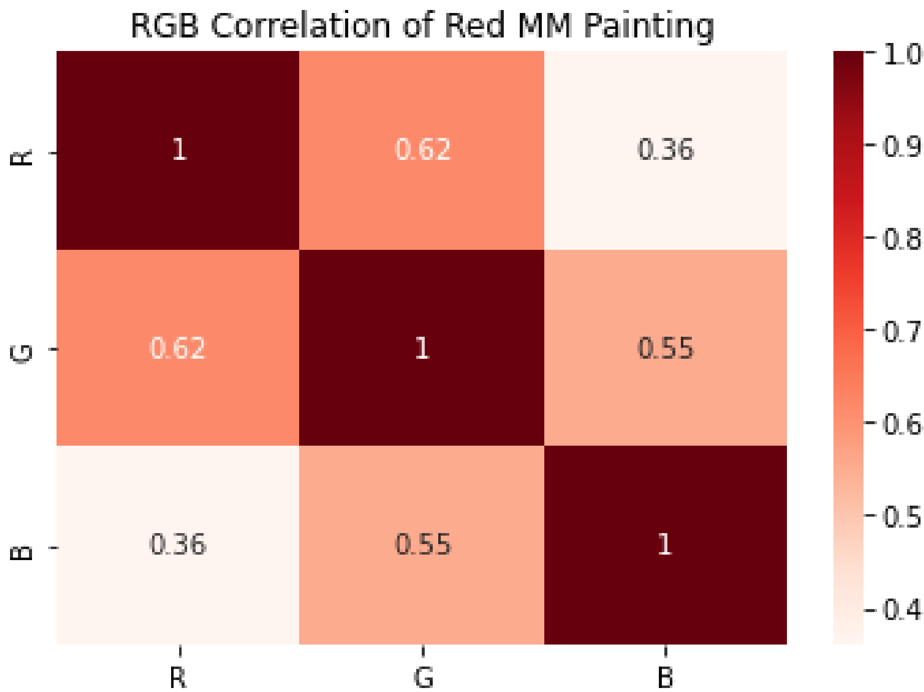

Figure 4: RBG Correlation of Red Shot Marilyn

When looking at the correlation heatmap between red, blue, and green for the Red Marilyn Monroe Painting, we see that the most highly correlated colors are blue and green. But shockingly, we see that blue and red are 36% correlated with one another meaning that if red increases, there is a chance of blue color visibility. This is shocking as the primary color of the red MM painting is red. Red being the most correlated with green is due to the color of the hair as it is yellow/golden which red and green make when combined.

2.1.4 RGB Conditional Entropy

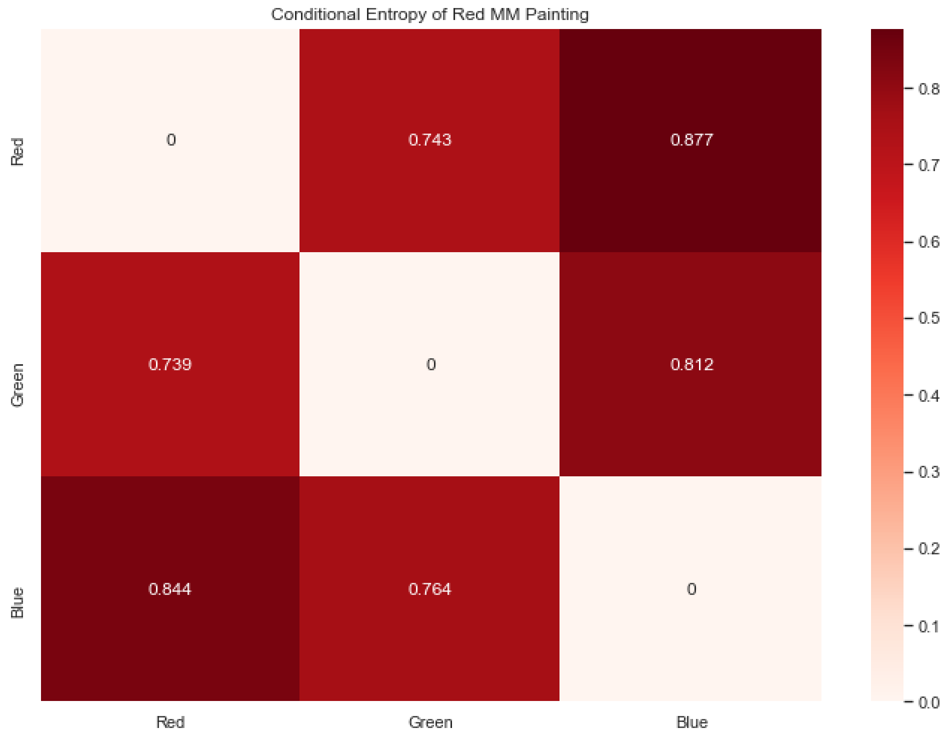

Figure 5: Conditional Entropy of Red Shot Marilyn

When looking at the Conditional Entropy heatmap for the Red MM Painting, noticeably all the combinations of RGB are relatively high (above 0.5) meaning that they are somewhat independent of each other and not determined by one another. When comparing the red color channel to the green color channel, it is at 0.743 meaning that red cannot predict green as they are more independent from one another. The highest conditional entropy is between red and blue and blue and red indicating that they are very independent from one another and cannot predict one another with high accuracy. This makes sense as they also have a low correlation with one another and red and blue do not go hand in hand in colors as red should be the more prominent color in the Red MM painting. However, green and red and red and green have one of the lowest conditional entropies for the red MM painting and they have a relatively mild correlation with one another. The only time they are used together is for the yellow/golden hair which allows the red/green combination to be predictable with one another.

2.2 Orange Shot Marilyn

2.2.1 RGB Density Plot

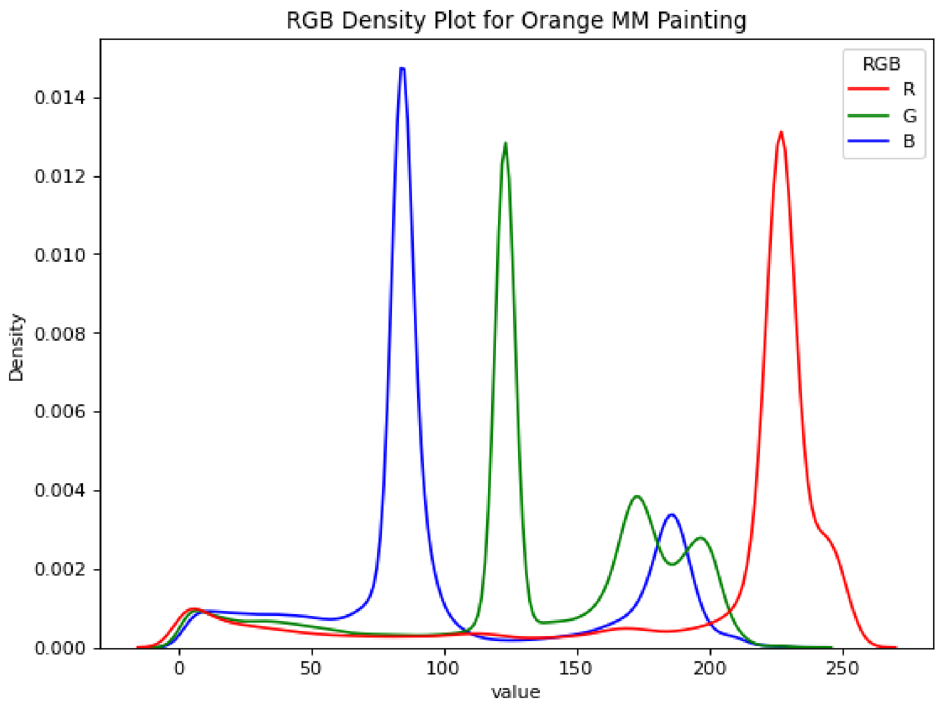

Figure 6: RGB Density Plot of Orange Shot Marilyn

For the orange density plot, we can see that each of the RGB colors each peaked once throughout different parts of the RGB range. Blue peaked first at a value of around 90, followed by green which peaked at a value of around 120. 2 mini-peaks of green and blue followed at the range from 160 to 200. Finally, it was followed by red which peaked at around 230. We can see that the major peaks of RGB are all around the same density, which is from 0.012 to 0.014. All three colors had a slight increase in density at 0 however the density will maintain below 0.002 until each color's first peak. Blue and green's density returns to 0 at around 220 on the RGB range, and red's density returns to 0 after its last and only peak, which is approximately around 227.

2.2.2 RGB 2D Projection Plot

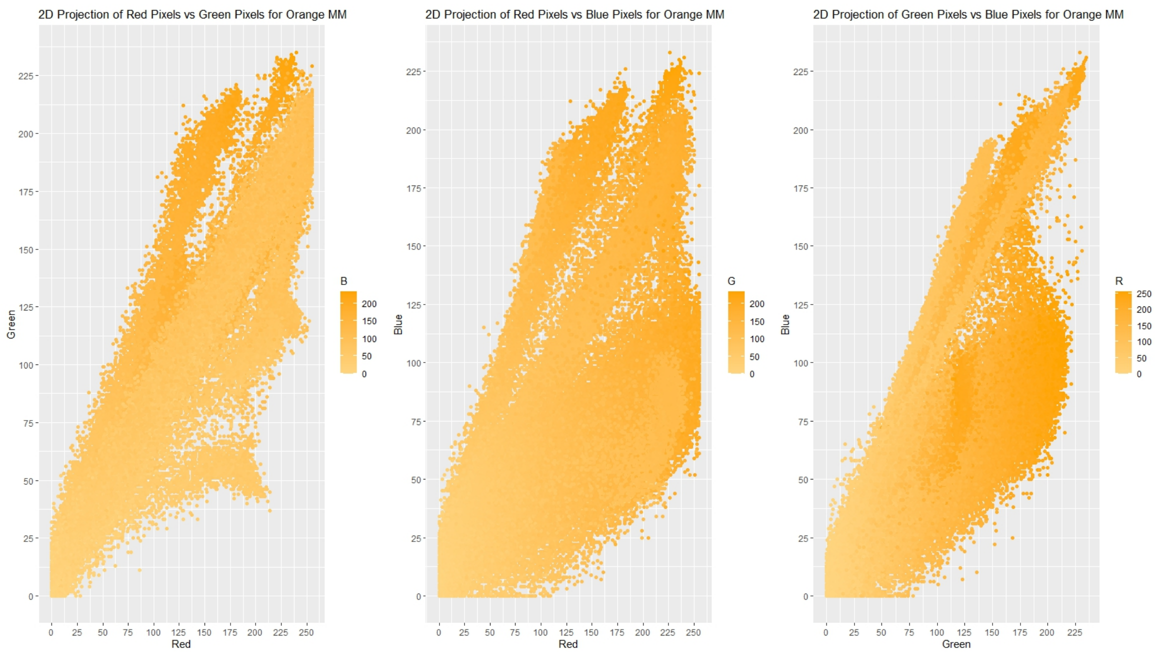

Figure 7: 2D Projection Plot of the Orange Shot Marilyn

When looking at the 2D projection plot for the Orange Marilyn Monroe Painting with Red plotted versus Green with Blue as the intensity, we can see a robust linear relationship between the red and the slope is around 1, and the overall trend is approximately y=x. When the pixel value of red is less than the pixel value of green (when the points are above the y=x line), the value of blue is mostly above 100, resulting in a more intense use of the color blue. When the pixel value of red is greater than the pixel value of green (when the points are below the y=x line), the value of blue is mostly below 150, resulting in less intense use of the color blue.

When looking at the 2D projection plot for the Orange Marilyn Monroe Painting with Red plotted versus Blue with green as the intensity, the points of red and blue also follows a linear relationship. As the pixel value of red and blue increases, the cluster of points spread out more around the trend line. When the value of red and blue are lower, the value of green also tends to be lower. Similarly, when the values of red and blue are higher, the value of green also tends to be higher.

When looking at the 2D projection plot for the Orange Marilyn Monroe Painting with Green plotted versus Blue with red as the intensity, the overall trend is linear but there's a gap in the cluster of points where blue ranges from 130 to 180, and where green ranges from 180 to 230. For a green value of less than 100, the pixel value of red is on the low end. For a green value greater than 100, the red value is less intensified above the y=x line. For points of green and blue below the y=x line, the red values are relatively higher.

2.2.3 RGB Correlation Heatmap

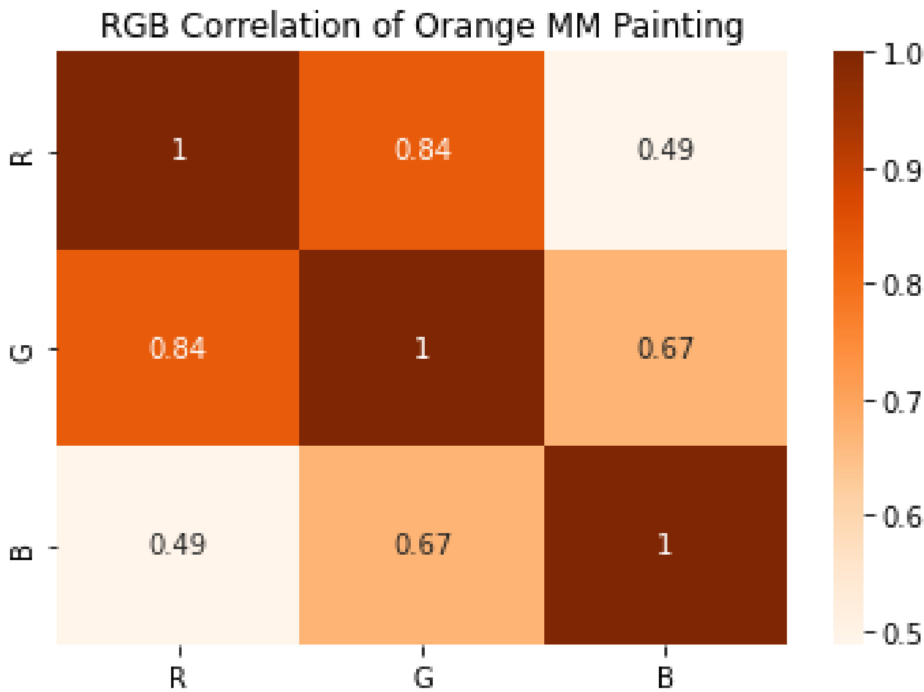

Figure 8: RBG Correlation of Orange Shot Marilyn

When looking at the correlation heatmap between red, blue, and green for the Orange Marilyn Monroe Painting, the correlation between red and green are very high at 84% while the correlation between blue and red is at 49%. The correlation between green and blue is at 67%. The red and green must have a high correlation due to two reasons which are how the background is orange and the color of the hair. The color of the hair is yellow/golden which is made by combining red and green to make yellow which increases their correlation with one another. The background is orange, the color orange is made by combining yellow and red whereas yellow is made by combining green and red. More green is used when making both yellow and orange and thus they have a much closer relationship in terms of color use with one another.

2.2.4 RGB Conditional Entropy

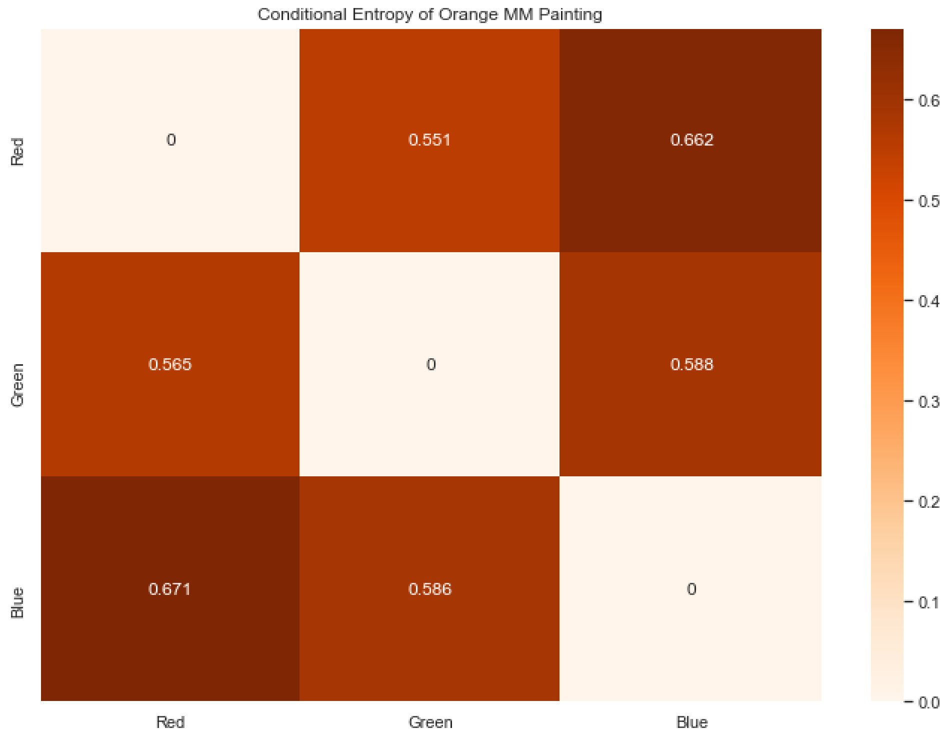

Figure 9: Conditional Entropy of Orange Shot Marilyn

When looking at the Conditional Entropy heatmap for the orange MM painting, noticeably all the conditional entropies of the colors are relatively high meaning that they are somewhat independent of each other and one color cannot give information about another color. However, since orange is a mix of colors, the conditional entropies for all the colors will be somewhat lower than the other conditional entropies plot for the other painting as orange is a mixture of green and red. Red and green and green and red have a conditional entropy of 0.551 and 0.565 respectively indicating that they are somewhat independent but the lower entropy could be because of the combination of red and green to make yellow and then the red and yellow to make the orange background. This can be supported by the orange MM painting correlation plot as red and green and green and red correlations are very high. However, red and blue and blue and red have the highest conditional entropies at 0.662 and 0.671 which make sense as red and blue should not be able to predict one another since the primary color being used here is orange that does not involve blue and red and blue combinations are not a majority proponent of the painting.

2.3 Green Shot Marilyn

2.3.1 RGB Density Plot

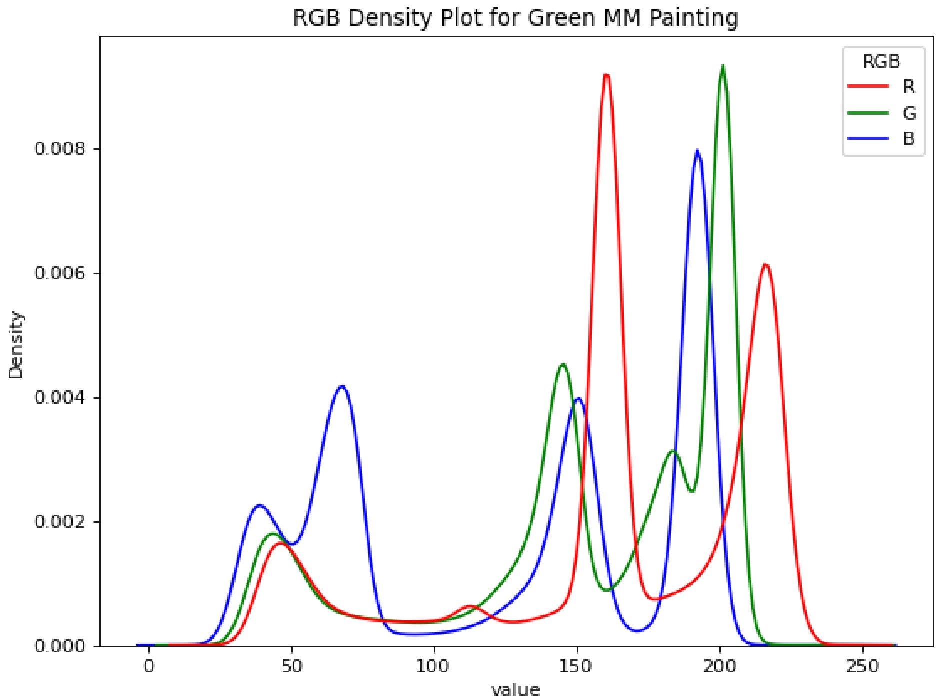

Figure 10: RGB Density Plot of Green Shot Marilyn

When looking at the RGB density plot for the Green Marilyn Monroe painting, there are multiple peaks throughout the RGB range. There are no peaks at 0 for any of the three colors. From range 0 to 100, there is a peak at blue along with slight peaks of red and green (primarily around 25 to 60). The highest intensity for red is in the range of 145 to 175 and the range of 200 to 225. There are multiple peaks for green throughout the mid to high range of RGB values, primarily the highest peak is at the range of 175 to 225. There are 4 peaks for blue that occur throughout the RGB value range. From the range 25 to 80, there are two peaks of blue. From the range of 125 to the range of 160 there is one peak of blue. Lastly, the highest peak of blue is the range of 175 to 210. All the peaks represent the intensity of where that color is used the most. Blue and green being the primary highest values in the Green MM painting is the most logical as the background is turquoise and green which incorporates both blue and green to make.

2.3.2 RGB 2D Projection Plot

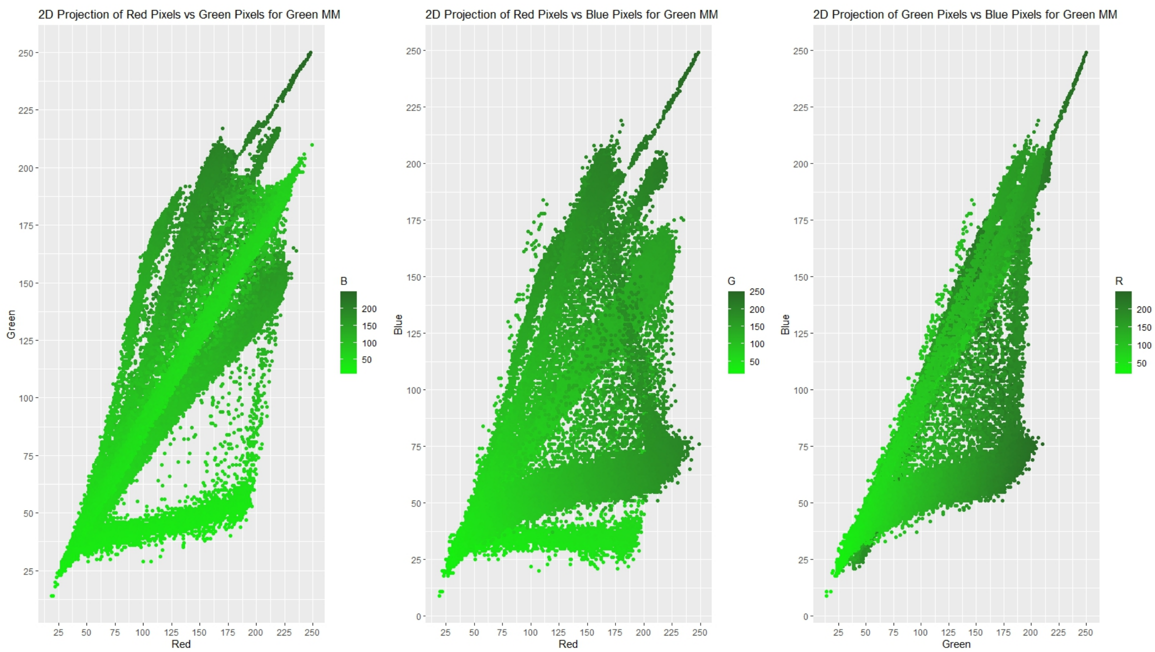

Figure 11: 2D Projection Plot of the Green Shot Marilyn

When looking at the 2D projection plot for the Green Marilyn Monroe Painting with Red plotted against Green and Blue for intensity, there is a gap of no to a few points between red from the ranges of 75 to 200 and green from the ranges of 50 and 125. The intensity of blue is only occurring at the most values that are not in the linear cluster of points between red and green. Higher values of green and lower values of red correlate with darker blue to make that turquoise background color. The shape is very distinct as it is not linear and has no unique distribution.

When looking at the 2D projection plot for the Green Marilyn Monroe Painting with red versus blue with green for the intensity has a similar distribution shape as the last 2D projection plot where red is plotted against green but in this case, the ranges of no to few points have a cluster of multiple points where the green is highly intense (at higher ranges of values). It seems that a majority of points on this 2D projection plot have green at somewhat high intensity except for ranges of red from 25 to 200 and ranges of blue from 25 to 50. This collection of points may only correlate to one portion of the painting which is the red lipstick that Marilyn Monroe has in the painting to create the red while everything else, red is not a necessary color for the Green Marilyn Monroe painting. The cluster of points as a whole is a cone shape and does not have a uniform distribution.

When looking at the 2D projection plot for the Green Marilyn Monroe painting with green plotted against blue with red as the intensity, the majority of the points have a low intensity of red as red is not the main color in the Green Marilyn Monroe painting. There is a linear line between green and blue where most of the points fall. Green and blue are correlated as we see that as green increases, so does blue which helps with the color spread of the Green Marilyn Monroe painting.

2.3.3 RGB Correlation Heatmap

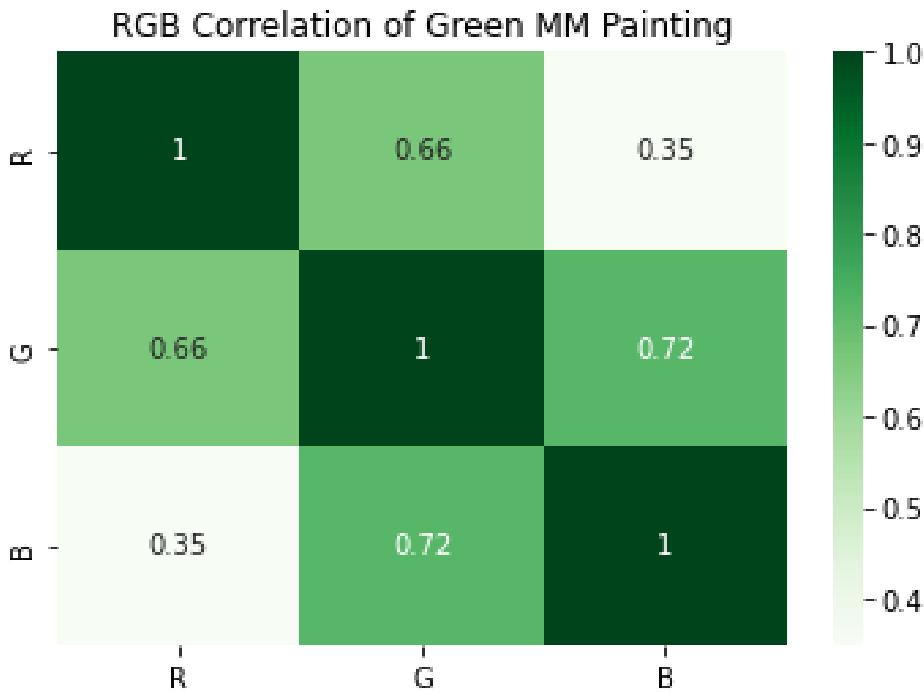

Figure 12: RBG Correlation of Green Shot Marilyn

When looking at the correlation heatmap between red, blue, and green for the Green Marilyn Monroe painting, we see that the most highly correlated colors are blue and green while blue and red are the lowest correlated colors. The correlation between red and green is 66%, this is mostly due to the yellow/golden hair. Additionally, the high correlation between blue and green is most likely determined by the background of the green MM painting as the background is turquoise. Turquoise is made by combining blue and green.

2.3.4 RGB Conditional Entropy Plot

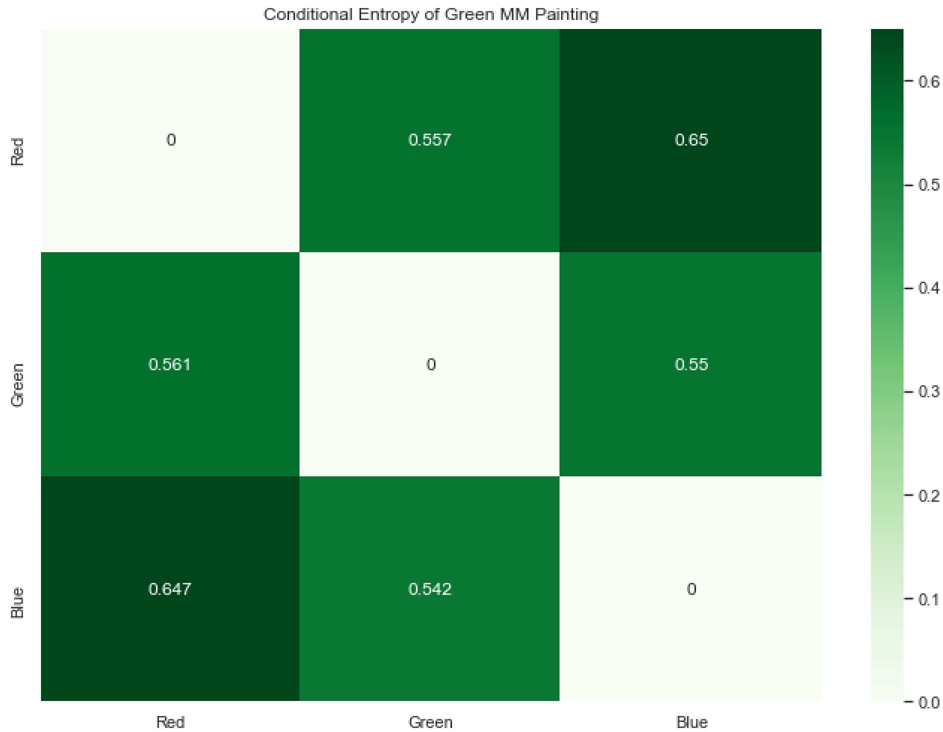

Figure 13: Conditional Entropy of Green Shot Marilyn

When looking at the conditional entropy heatmap for the green MM painting, green and blue and blue and green should be the lowest conditional entropy out of all color combinations since the background of the green MM painting is turquoise. Since the background is turquoise, the colors of blue and green combined to make turquoise indicating that they can somewhat predict one another as when blue is used, green could potentially be used to make the background. Regardless of this information, green and blue and blue and green have a conditional entropy of 0.55 and 0.542, respectively. They are the lowest conditional entropy indicating that they are the most related to one another but they are still independent as it is closer to 1 than 0. Red and blue and blue and red have the highest conditional entropy indicating independence and not much prediction can be made by knowing one of the color channels in this case, they are at 0.65 and 0.647, respectively. Red and green and green and red are between the two conditional entropy sets of red and blue and green and blue, as they are at 0.557 and 0.561 respectively. The values of red and green cannot be predicted as well comparatively to blue and green but are not as unpredictable as red and blue as noted by their larger conditional entropies.

2.4 Aqua Shot Marilyn

2.4.1 RGB Density Plot

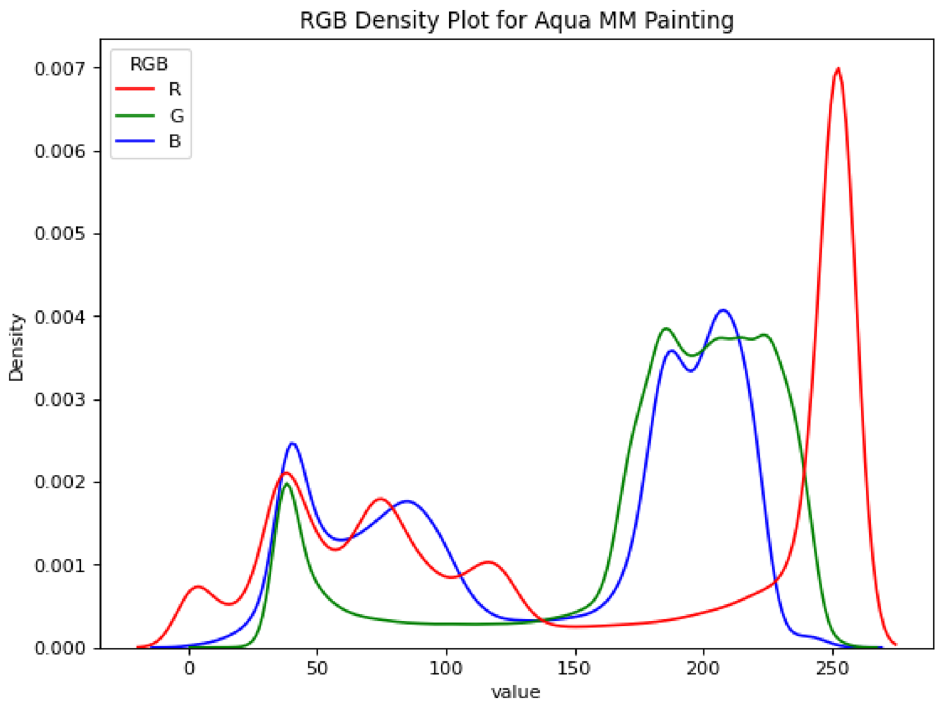

Figure 14: RGB Density Plot of Aqua Shot Marilyn

For the aqua shot Marilyn, we can see that the largest amount of red pixels are at the value of 255 (8.52% specifically). There is a large proportion of red pixels in the value range of 249-255 making up 31.33% of the red pixels. There are a couple value areas that the red pixels like to cluster around, such as the local maximums at values in the 30s and 70s. For the green pixels, the majority of them are clustered around the values around 170-230. For the blue pixels, there are two areas where they seem to cluster around. The larger area is around the values of 180-215 and the second cluster is near the values in the 30s.

2.4.2 RGB 2D Projection Plot

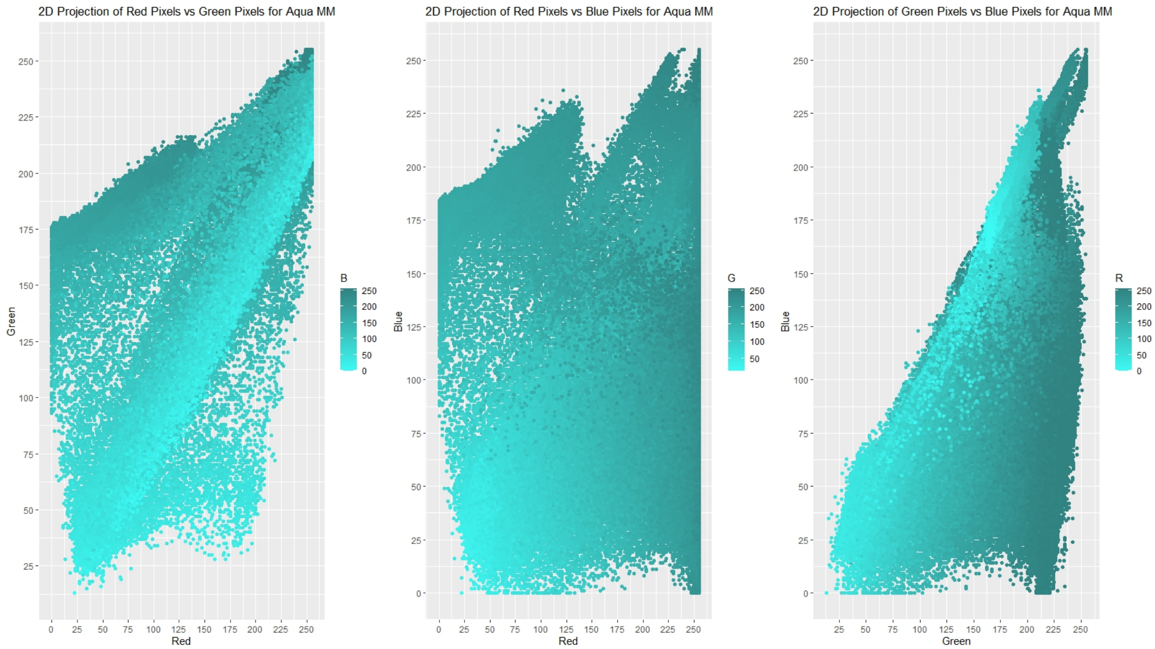

Figure 15: 2D Projection Plot of the Aqua Shot Marilyn

When looking at the 2D projection plot for the Aqua Shot Marilyn with green plotted against blue, we can see that the points around the y = x line has very little red as evidenced by the lighter points while points far from the y = x line tends to be darker and have more red. Also notice how this plot is mostly empty above the y = x line and how the possible combinations of green and blue only forms a pattern that's like a triangle/half a square.

Meanwhile, plotting red against blue, we don't see much empty space indicating that the aqua picture utilizes a lot of combinations between the red and blue values. Also, if we carve out a rectangle where blue < 75 and red <125, we can see the points in this area tends to have less green compared to the other points.

Plotting red against green, we can see that points near the y = x line has less blue while points that are far above or below the line tends to have more blue.

2.4.3 RGB Correlation Heatmap

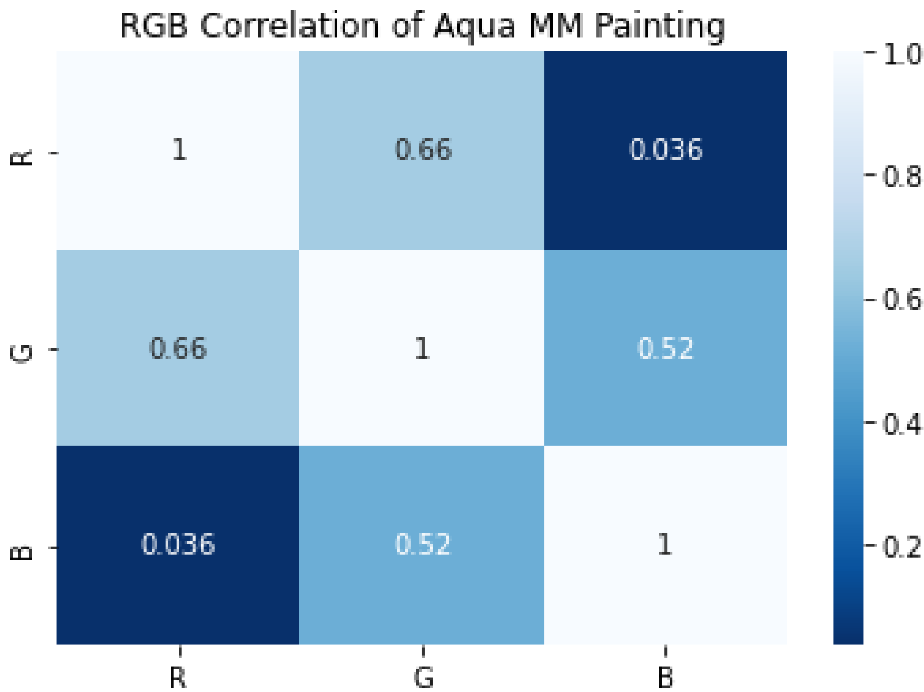

Figure 16: RBG Correlation of Aqua Shot Marilyn

When looking at the correlation between red, blue, and green for the Aqua Marilyn Monroe Painting, we see that red and blue are the least correlated among the colors. The green is the most correlated with red. Blue is most correlated with green. The correlation between red and green make sense as when there the painting is blue, there should not be any red as the primary color of the blue MM painting is blue and not any colors mixed with red.

2.4.4 RGB Conditional Entropy Plot

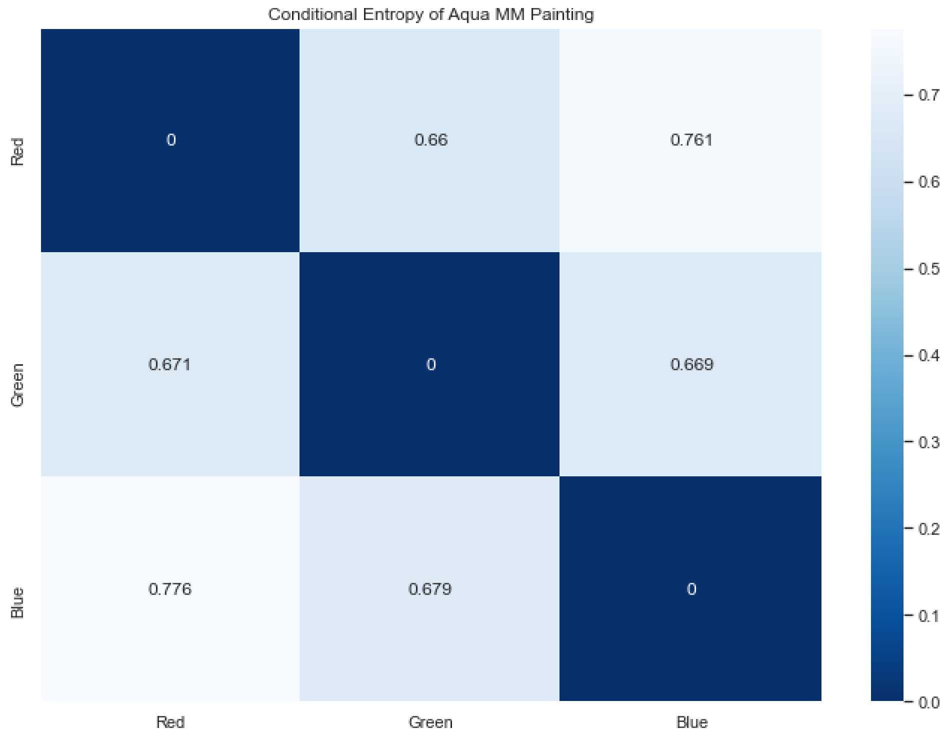

Figure 17: Conditional Entropy of Aqua Shot Marilyn

When looking at the conditional entropy heatmap for the aqua MM painting, red and blue and blue and red have the highest conditional entropy meaning they are the most independent colors in the painting. They have a conditional entropy of 0.761 and 0.776, respectively. The conditional entropy between the red and blue combination is logical because in the aqua paint, the color that is prominent is blue and to keep it blue, red is not needed thus not being able to predict when blue and is independent of red. While red and green and green and red have a conditional entropy of 0.66 and 0.671, respectively. They are the midpoint of conditional entropies of all color combinations and they are not correlated with one another well and are not able to predict one another at high accuracies. The green and blue and blue and green have a conditional entropy of 0.669 and 0.679, respectively. The overall color composition and relationships within the image are not predictable at high accuracy since all the conditional entropies are above 0.5 meaning somewhat independent.

2.5 Blue Shot Marilyn

2.5.1 RGB Density Plot

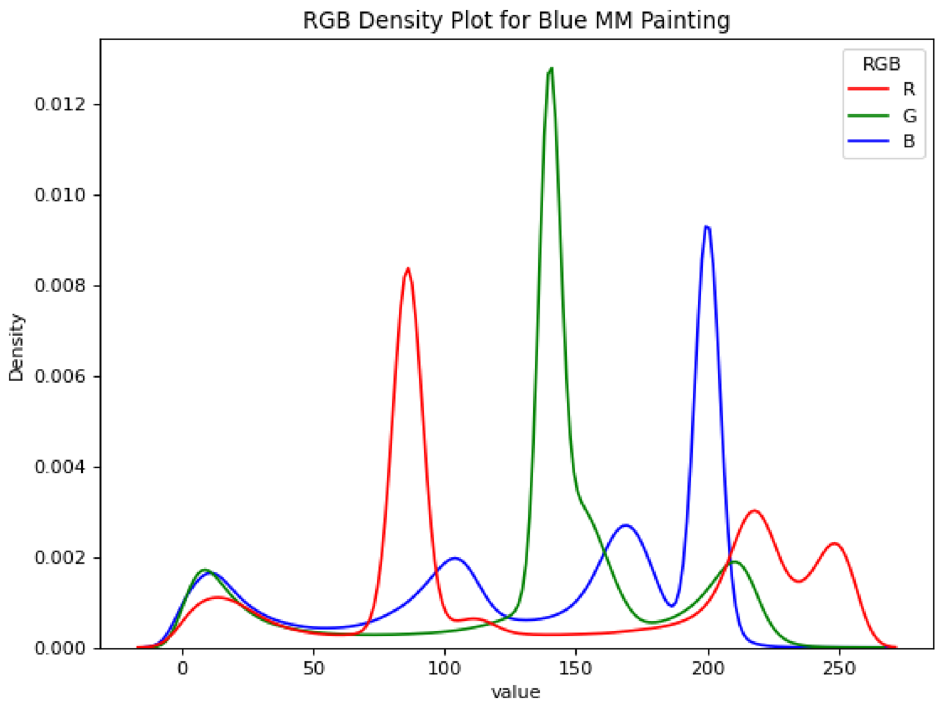

Figure 18: RGB Density Plot of Blue Shot Marilyn

For the blue shot Marilyn, we can see that most of the pixels have a blue value of 200. Also, surprisingly it seems that there are very few pixels that have a blue value greater than 200 judging by the extremely low density beyond the 200 points. For the green values, most of the pixels have a green value of 140. We can also see that there's an extremely low density beyond 225 for the green value and that hardly any pixels have a green value above 225. As for red, the most frequent red value is 86. Also based on the density plot, there doesn't seem to be certain red values that have an exceptionally low amount of pixels.

2.5.2 RGB 2D Projection Plot

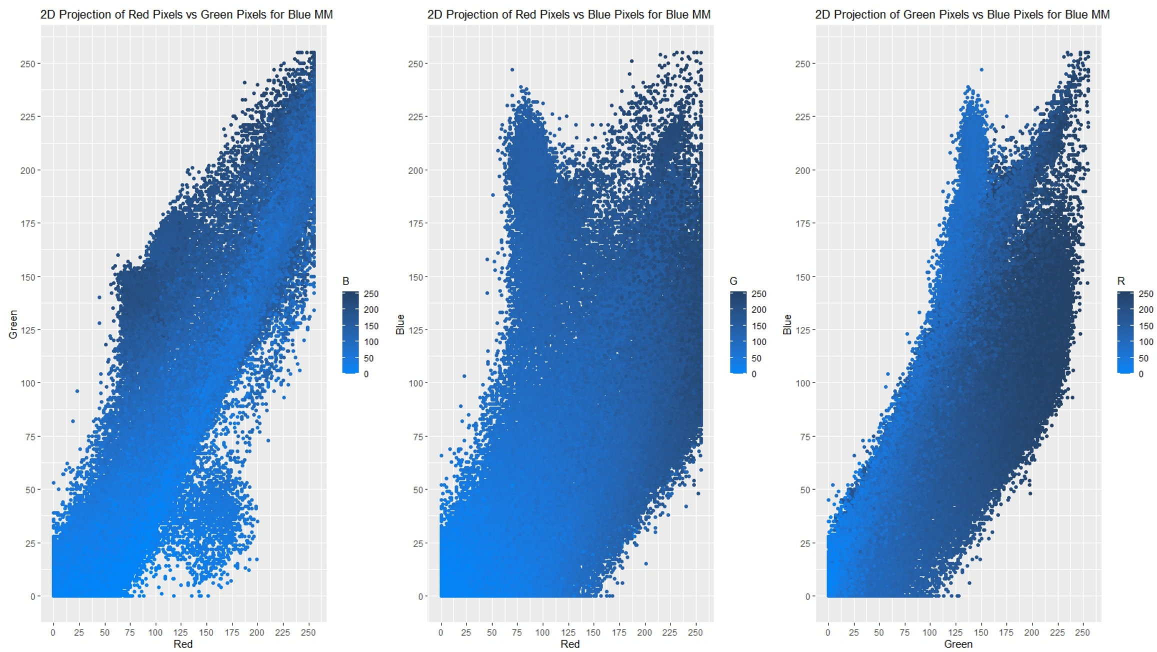

Figure 19: 2D Projection plot of the Blue Shot Marilyn

Judging by this projection plot for red against green for the blue Shot Marilyn painting, we can see that the red and green values are fairly correlated and seem to follow a trend similar to y = x. For example, when the green value is 225, we can be 100% certain that the red value is always greater than 175. In fact, this plot has a lot of empty space and the dots don't fill the entire quadrant meaning a lot of combinations of red and green are not possible. Then, if we split the points in a pattern similar to y = x, we can see that above the line, the points tend to have a higher blue value compared to the points below the line as evidenced by the legend. In other words, given a certain value of red, the more green a pixel has, the more blue it is.

On the other hand, for the red against blue plot, we see that there's less empty space in this plot meaning there are more combinations of blue and red possible compared to red and green. If we divide the points with a vertical line somewhere around 200, we can see that the green values are higher for the points on the right as opposed to the left. In other words, the more red a pixel is, the more green it tends to be.

This green vs blue projection plot has the 2nd least empty space meaning it has the 2 most combinations available (between blue and green). Dividing the points with the line shown in the figure. We can see that the points above tend to have a lower red value compared to the points below where the red value is pretty high. In other words, given a certain green value a pixel has, the more blue a pixel has, the less red it tends to be.

2.5.3 RGB Correlation Heatmap

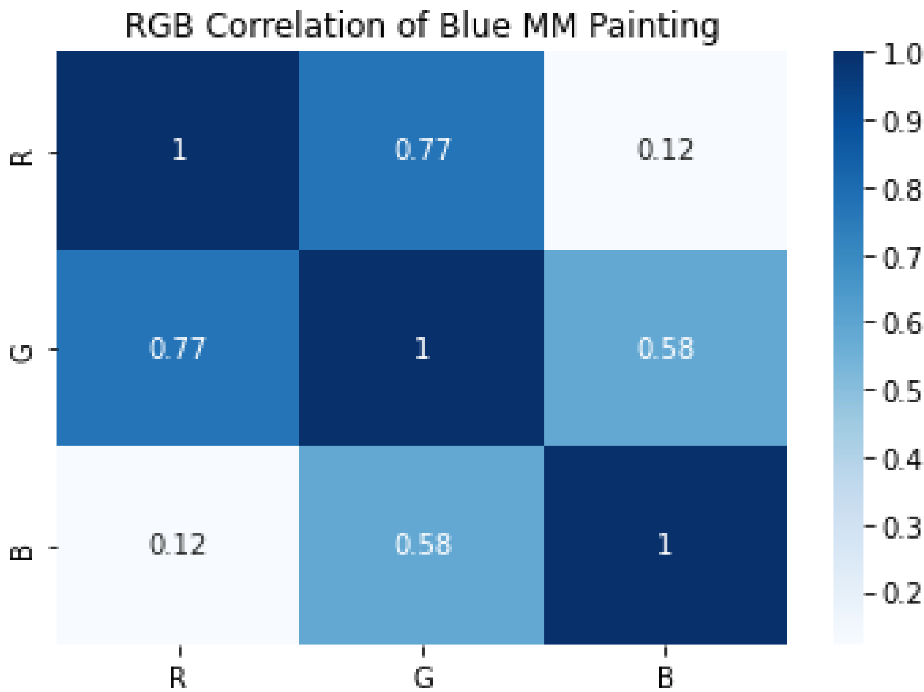

Figure 20: RBG Correlation of Blue Shot Marilyn

When looking at the correlation heatmap between red, blue, and green for the Blue Marilyn Monroe Painting, we see that red and blue are the least correlated colors, and green and red are the most correlated colors. The correlation between blue and green is 58%. When comparing to the Aqua Marilyn Monroe Painting, the correlation of red and blue was 3.6% which is much lower than the correlation of red and blue in this painting at 12%. Aqua Marilyn Monroe painting must have a lower correlation between red and blue as the color of the skin is much less magenta while the blue MM painting has a darker magenta hue to it. Magneta is made by combining red and blue together, hence why the blue MM painting has some correlation between red and green more so than Aqua MM painting even though the primary color is blue in both paintings.

2.5.4 RGB Conditional Entropy Plot

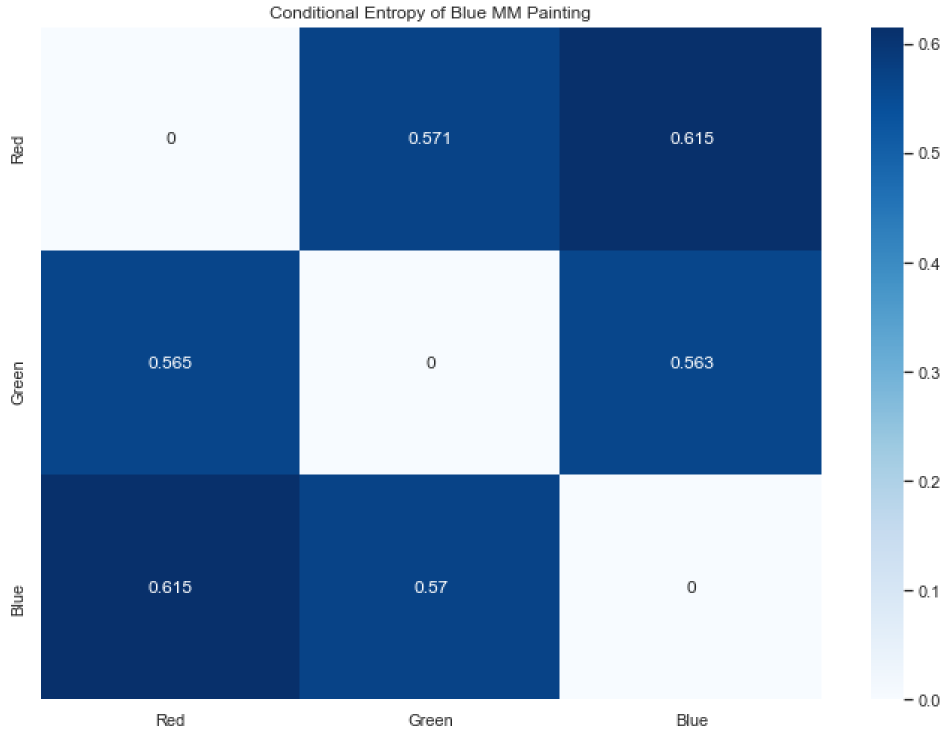

Figure 21: Conditional Entropy of Blue Shot Marilyn

When looking at the conditional entropy heatmap for the blue MM painting, shockingly the red and blue and blue and red have one of the lower conditional entropies of all paintings as they have a conditional entropy of 0.615 and 0.615, respectively. The reason why in the blue MM painting the conditional entropies of red and blue and blue and red are the lowest out of all other paintings is because of the color of the skin in this MM painting as it is a darker shade of magenta. Magneta is a combination of blue and red and thus allows for some correlation between the two. However, the conditional entropy is still above 0.5 and thus somewhat independent and cannot be used to predict one another. The conditional entropy for red and green, green and red, green and blue, and blue and green all have similar conditional entropy at 0.571, 0.565, 0.563, and 0.57 respectively. All that indicates is independence from one another and low accuracy for predictability among the color channels since all the conditional entropies are above 0.5 and independent.

2.6 Comparison and Contrast of all paintings

2.6.1 RGB Density Plot

Based on the density plots, red has the least blue out of all paintings. Red also has the least green out of all paintings judging by the higher density in the region around 0.

Meanwhile, it seems like blue has the least red.

On the other hand, aqua surprisingly has the most red most likely due to the dark magenta face color. Green unsurprisingly has the most green and blue has the most blue.

Orange, green, and blue tend to peak at specific places while the RGB values are more spread out for red and aqua.

2.6.2 RGB 2D Projection Plot

It seems like out of all the projection plots, green has the least amount of color combinations possible judging by the empty space. On the other hand, red and aqua have the least empty space signifying more color combinations present.

Also, for all the projection plots, most of them have a pattern around y = x where points above/below the line tend to have a higher/lower value of the 3rd color as discussed earlier.

All the 2D Projection plots for all paintings have a distinct and unique shape that is shared among the specific painting. It can be seen that blue and aqua has the greatest spread among the RGB color comparison while green is the most narrow in terms of spread among the RGB color comparisons.

2.6.3 RGB Correlation Heatmap

Among all the correlation heatmaps, one thing that is constant is how red and green are usually highly correlated with one another. This may be due to the fact that when red and green are mixed, it makes yellow which is the color hairs that MM in all of the 5 paintings share in common. Additionally, the Orange Marilyn Monroe has the highest correlation in terms of red and green and red and blue. Mixed colors most likely than a primary color must have a higher correlation with RGB because mixed colors require more colors to be used with one another increasing their relationship with one another and thus a correlation.

2.6.4 RGB Conditional Entropy Plot

The most noticeable thing about the conditional entropy in each painting is how red and blue and blue and red are the conditional entropies that are the highest compared to the other combinations of color channel indicating that red and blue does not provide enough information about each other. Additionally, the red MM painting had the highest conditional entropies for all combinations of colors due to the fact that red is the most prominent color and does not necessarily need blue or green when making that painting. It can be seen that is there some relation between correlation and conditional entropy between two colors. If the conditional entropy is higher, the lower correlation between the two indicates independence for the two colors compared to one another.

3 Clustering

Another motivation for this project was determining the main colors of each shot Marilyn and we have decided clustering would be a great way to do that. Since each image has thousands of colors we can't really distinguish well from the human eye, we decided to cluster colors that look relatively similar together.

There were two methods we used, Hierarchical with average linkage and K-Means clustering. Before we fit the models, we did a little bit of pre-processing. Because the data is extremely large with over 500,000 data points/pixels, we decided to center the pixels to reduce the size. That means that pixels 0-7 become 4, 8-15 become 12, ... all the way to 248-255 becomes 252. That reduces the number of colors we have to deal with significantly since 256/8 gives us 32 values. $32^3 = 32768 << 256^3 = 16,777,216$. Since we only care about color, we will ignore the x and y coordinates and drop duplicate pixels. Also, we do not have to worry about scaling the data since the data points are on the same scale. Here is a before (left) and after (right) centering and you can see that there are no perceivable differences.

Figure 22: Before and After Centering

To determine how good our clusters are, we need to visualize how well the pixels of similar colors have been separated and how many minor details the clusters captured which will discuss in the results. Also, when determining the number of clusters, we looked at the dendrogram for hierarchical clustering and cut the tree when the clusters started to get very far from each other distance-wise and then fitted a k-means with the same amount of clusters from hierarchical clustering.

We also visualized the percentage of data in each cluster. Since the clusters may have multiple colors within them but are still very similar, we used the median RGB values to represent that color/cluster in the donut charts we created.

3.1 Red Shot Marilyn

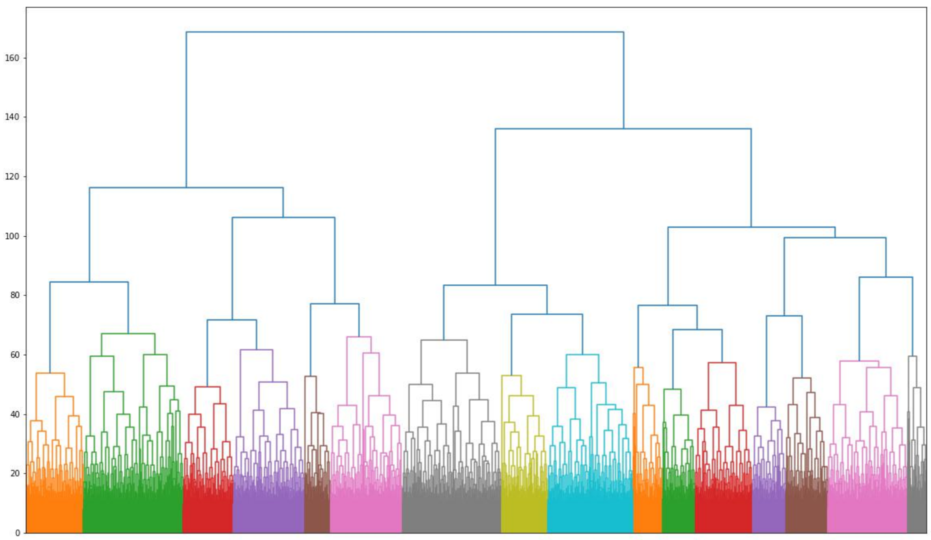

Figure 23: Dendrogram for Red Shot Marilyn

In figure 23, we cut the tree at roughly height 68 as that's when the clusters started getting really far from each other resulting in 16 clusters.

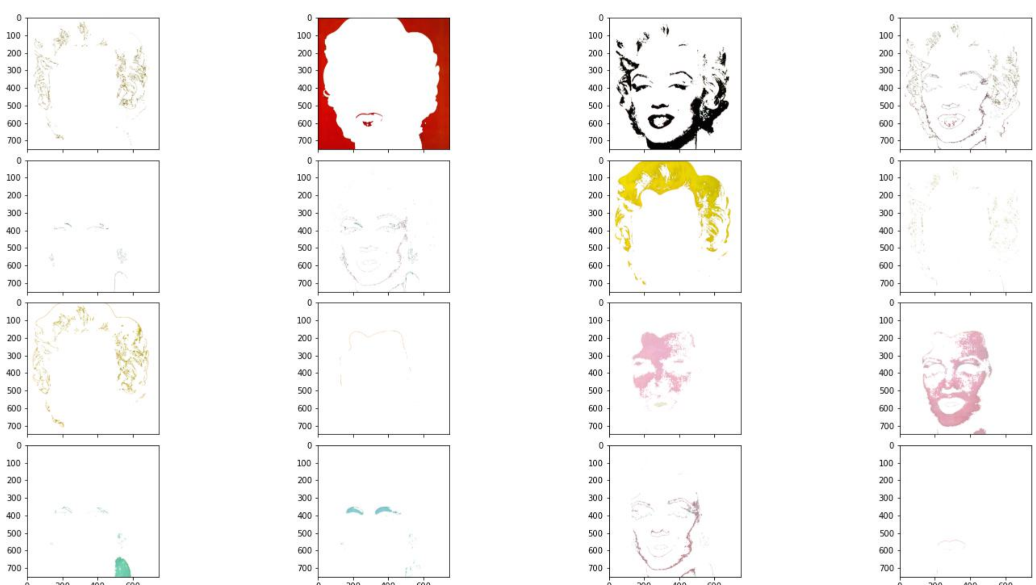

Figure 24: Resulting Clusters from HC for Red Shot Marilyn

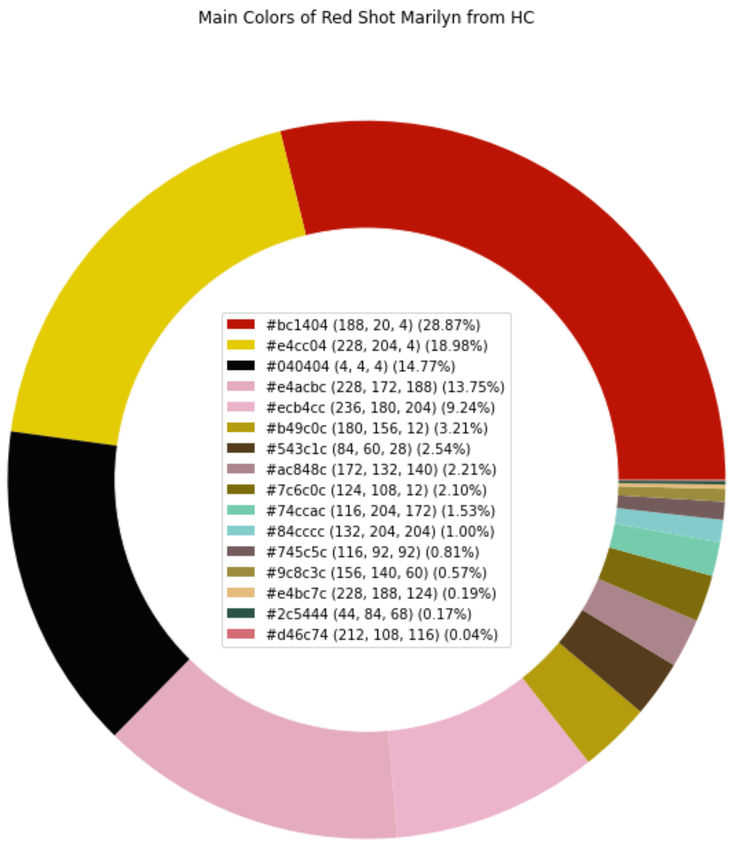

Figure 25: Donut Chart for HC Clusters for Red Shot Marilyn

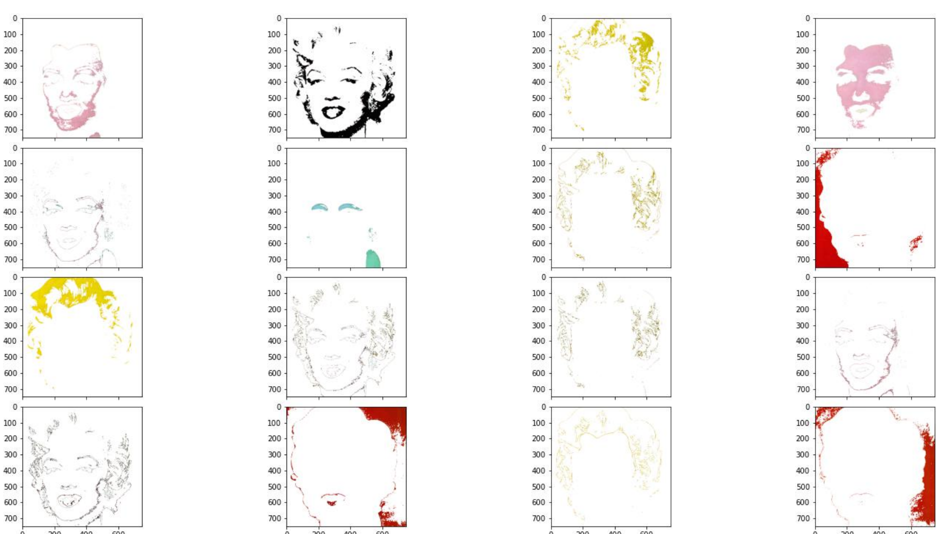

Figure 26: Resulting K-Means Clusters for Red Shot Marilyn

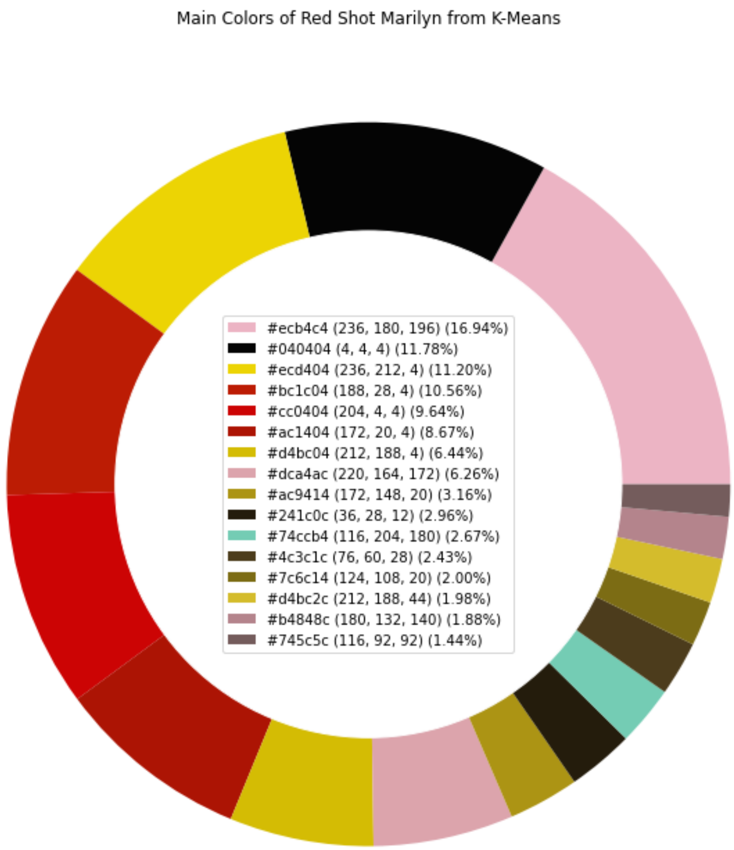

Figure 27: Donut Chart for K-Means Clusters for Red Shot Marilyn

Figure 25 and figure 26 are the main colors from Hierarchical clustering and K-means represented by donut charts.

In figure 24 and figure 26, we can see the resulting clusters from K-Means and Hierarchical clustering for the red shot Marilyn. We can see that both models separate the main parts of the Marilyn shot fairly well with some clusters depicting the skin, hair, subtle shading, or the red background.

However, K-Means seems to be good at picking up on the fact that the red background has 3 different shades of red while hierarchical clustering failed to do so and put them in one cluster. K-Means also separated different parts of Marilyn's hair much better compared to hierarchical clustering as evidenced by K-means having more clusters for Marilyn's hair.

3.2 Orange Shot Marilyn

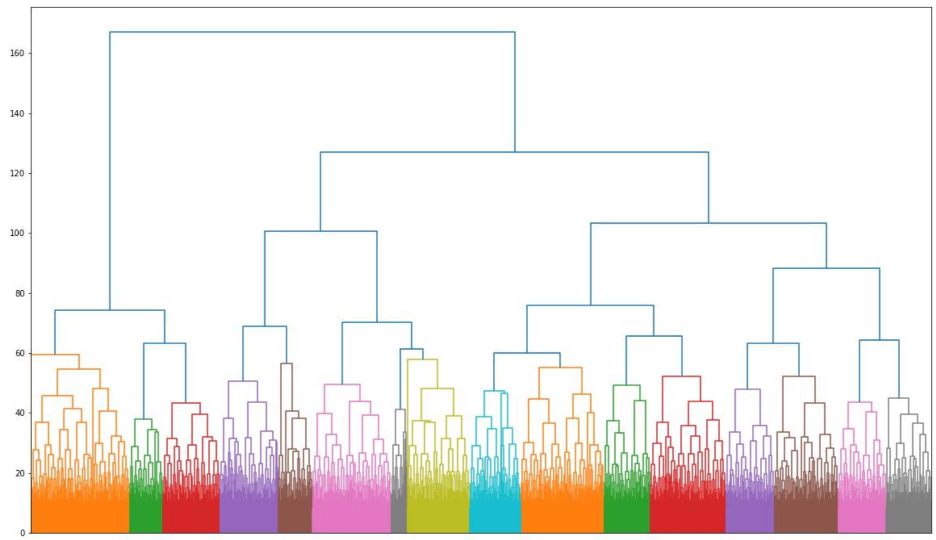

Figure 28: Dendrogram for Orange Shot Marilyn

Meanwhile, for the orange, we cut the tree at a height of 60 as shown in figure 28.

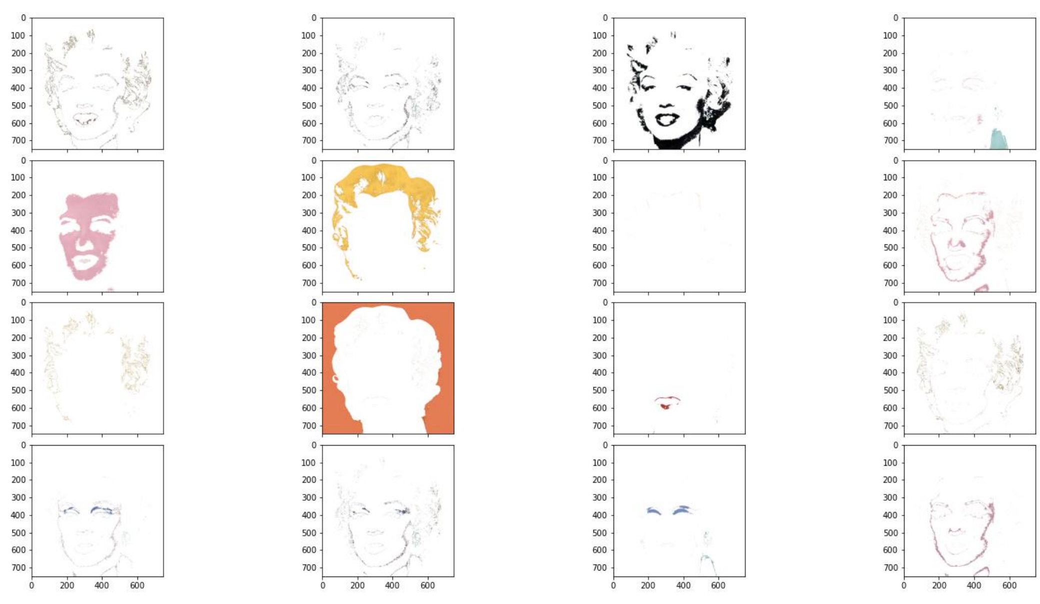

Figure 29: Resulting Clusters from HC for Orange Shot Marilyn

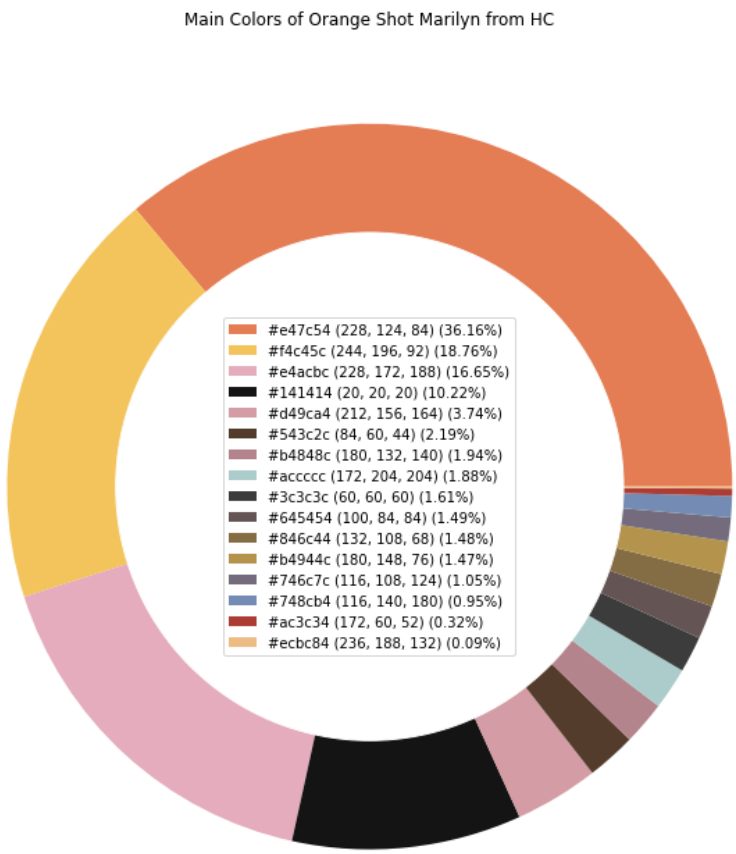

Figure 30: Donut Chart for HC Clusters for Orange Shot Marilyn

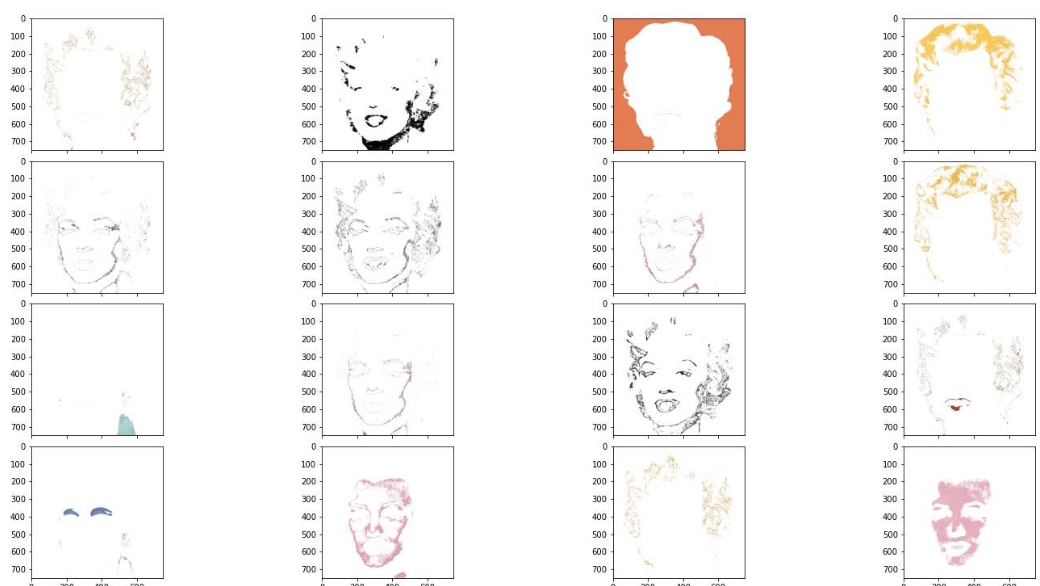

Figure 31: Resulting K-Means Clusters for Orange Shot Marilyn

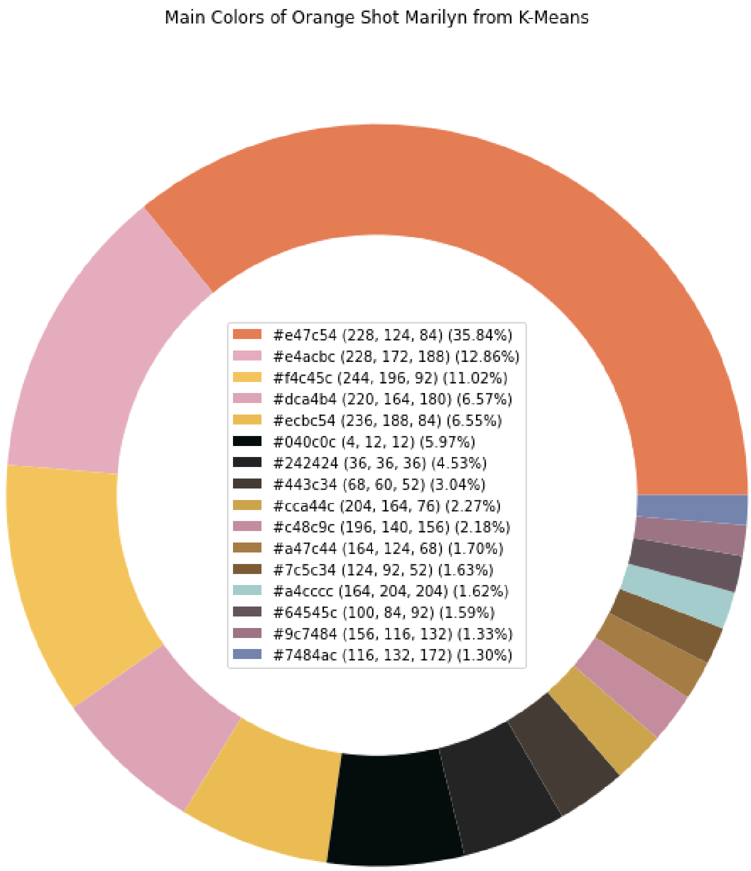

Figure 32: Donut Chart for K-Means Clusters for Orange Shot Marilyn

While both models did separate the main parts of the picture fairly well, it was not 100% clear which model was absolutely better. From both methods of clustering, we can see that the orange background was very simple and monotone as it was one cluster. Hierarchical clustering was better at detecting the fact that the eyeshadow had multiple layers as there were 2 clusters for eyeshadow. However, K-Means was much better at detecting the fact that the orange Marilyn had multiple layers to the hair and skin as evidenced by more clusters for these parts.

3.3 Green Shot Marilyn

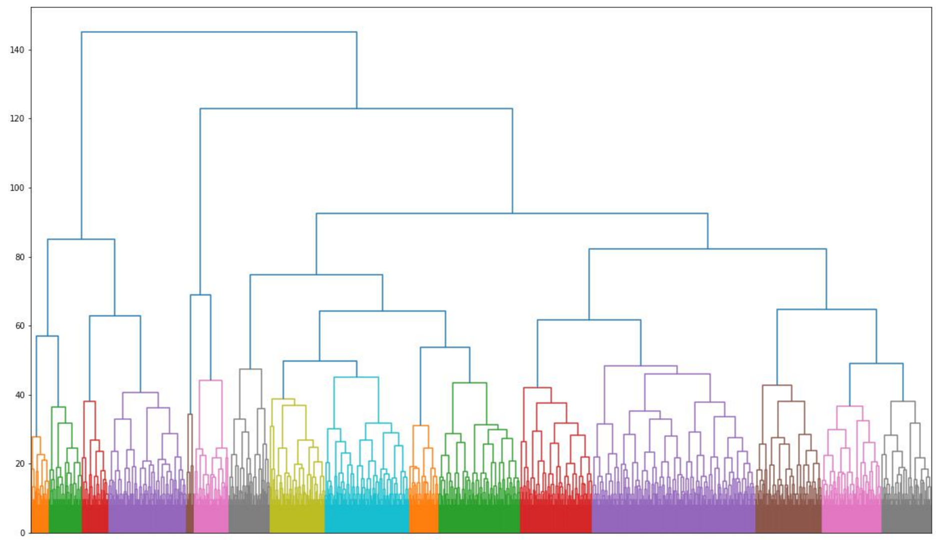

Figure 33: Dendrogram for Green Shot Marilyn

As for the green Marilyn, we cut the tree at 48.5 as shown in figure 33.

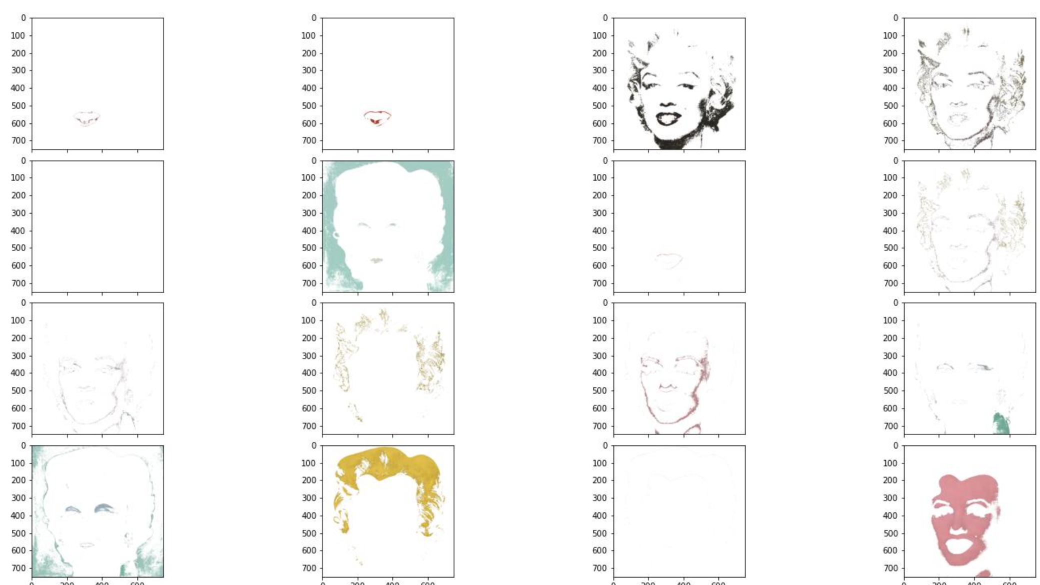

Figure 34: Resulting Clusters from HC for Green Shot Marilyn

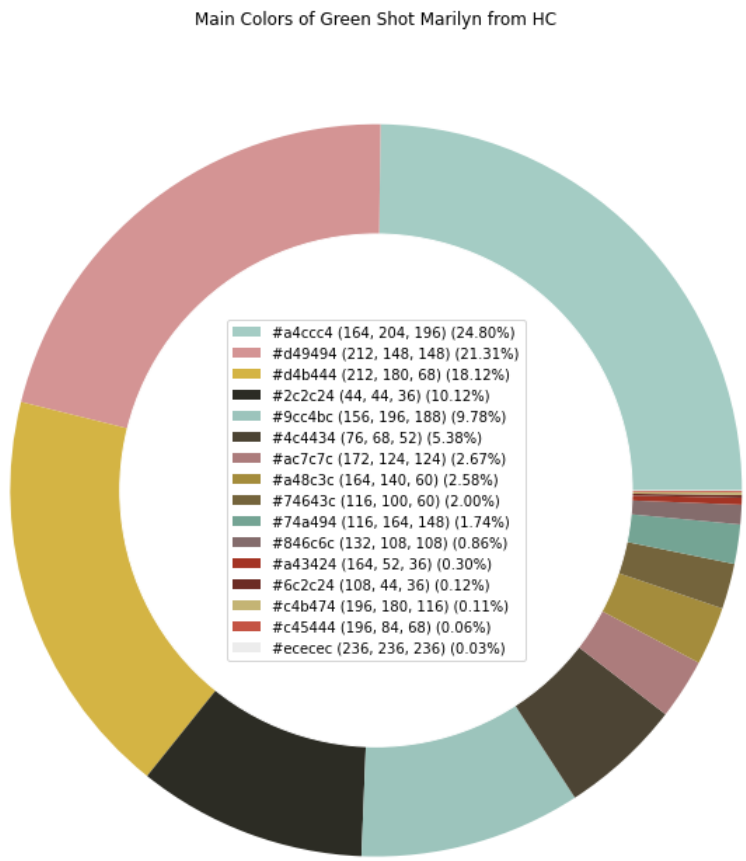

Figure 35: Donut Chart for HC Clusters for Green Shot Marilyn

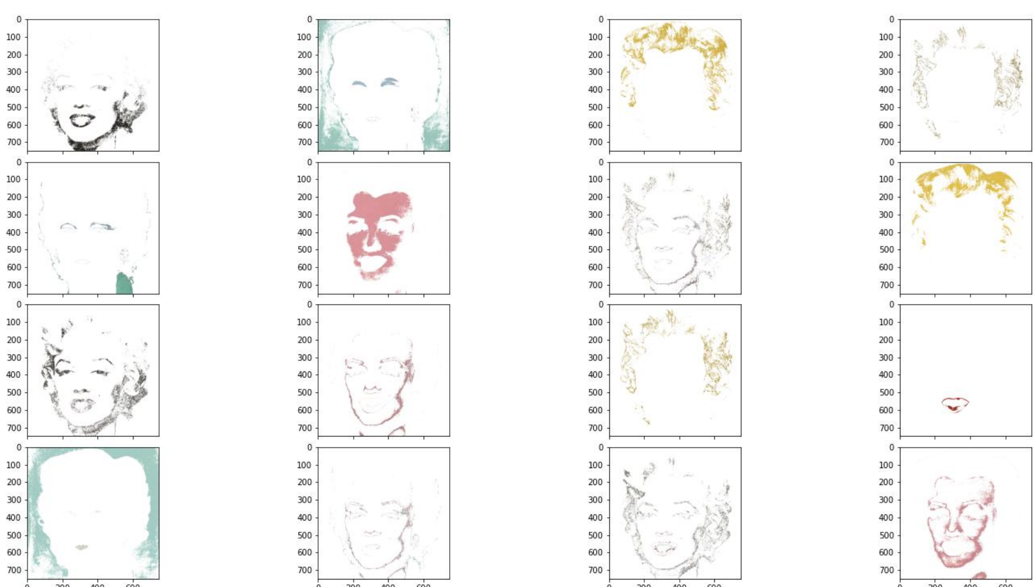

Figure 36: Resulting K-Means Clusters for Green Shot Marilyn

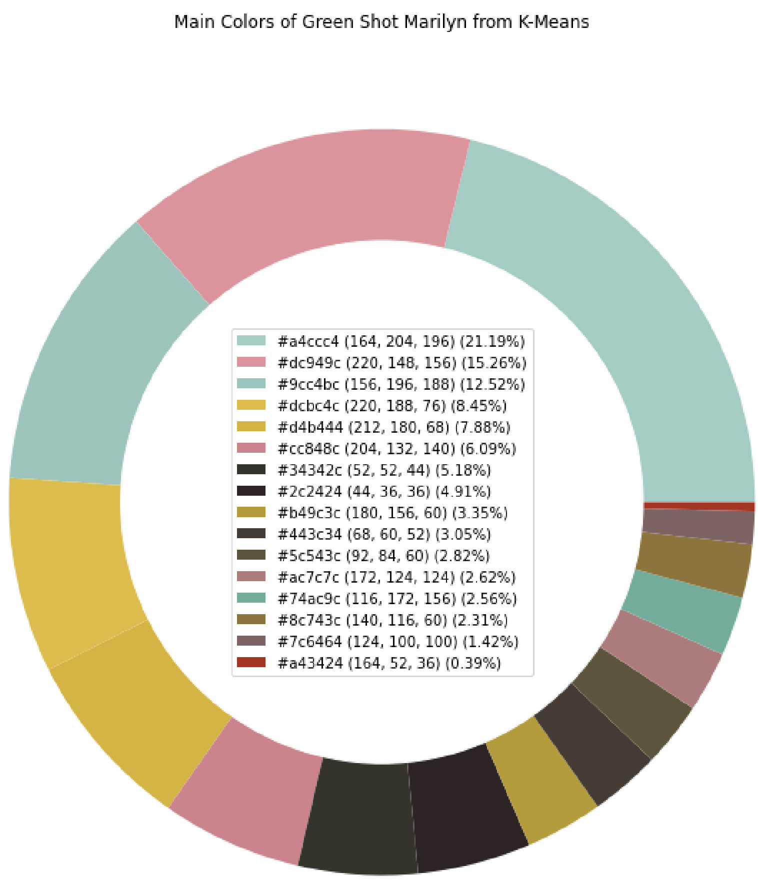

Figure 37: Donut Chart for K-Means Clusters for Green Shot Marilyn

Just like the other pictures, K-Means outperforms hierarchical clustering in terms of detecting the subtleties in Marilyn's hair and skin as it was able to extract more layers from these parts.

3.4 Aqua Shot Marilyn

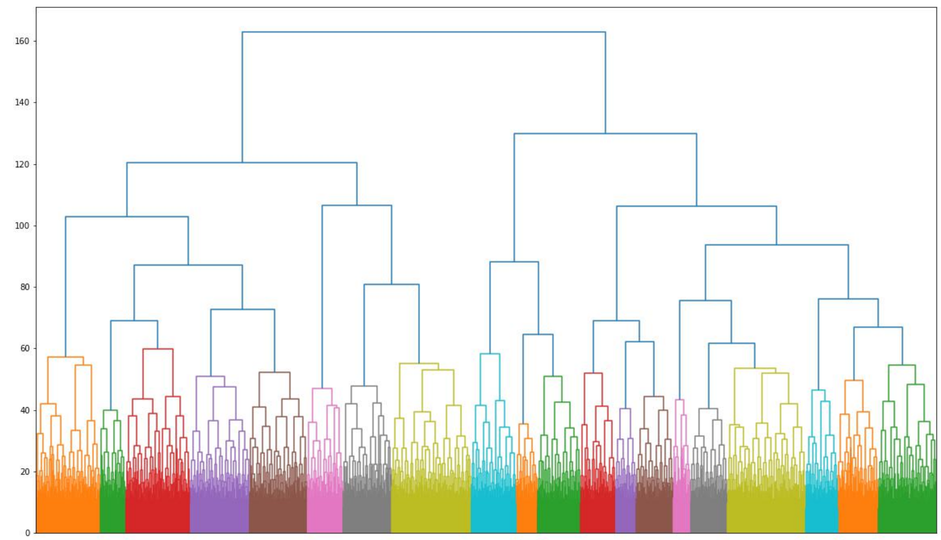

Figure 38: Dendrogram for Aqua Shot Marilyn

As for the aqua Marilyn image, we cut the tree at a height of 60 and resulting in 20 clusters as shown in figure 38.

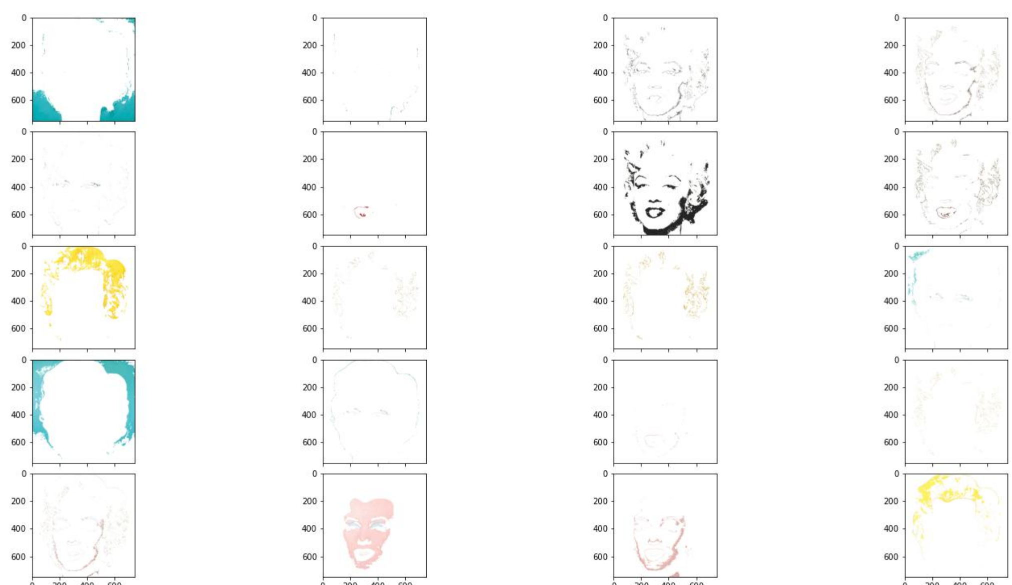

Figure 39: Resulting Clusters from HC for Aqua Shot Marilyn

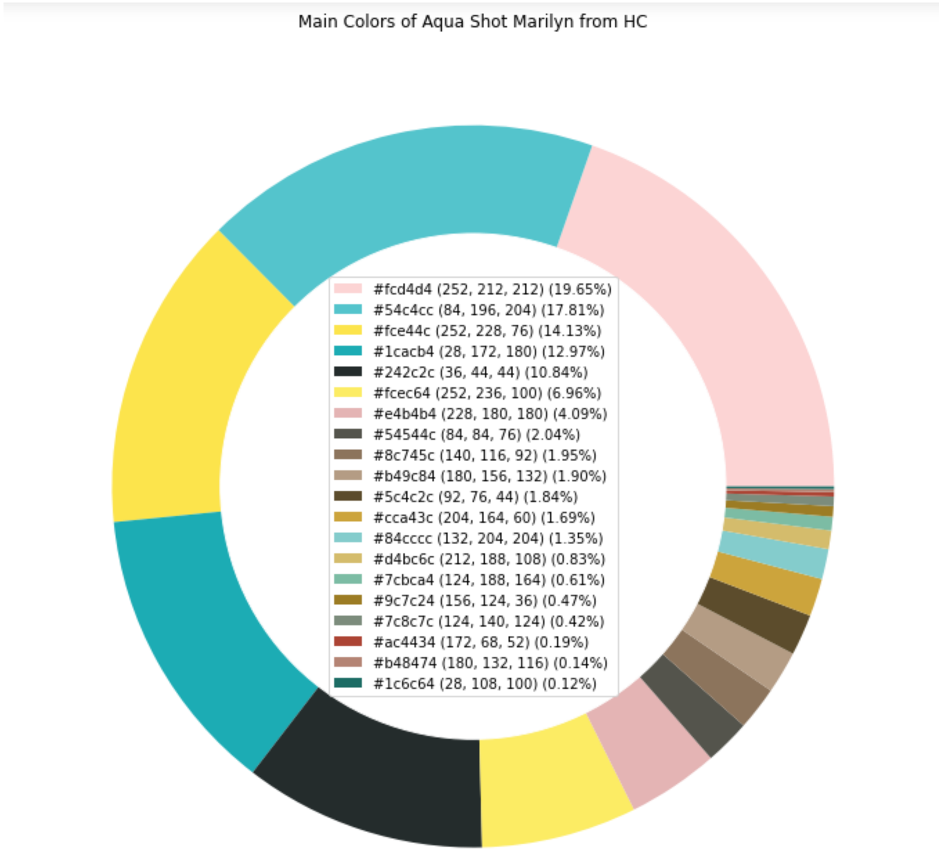

Figure 40: Donut Chart for HC Clusters for Aqua Shot Marilyn

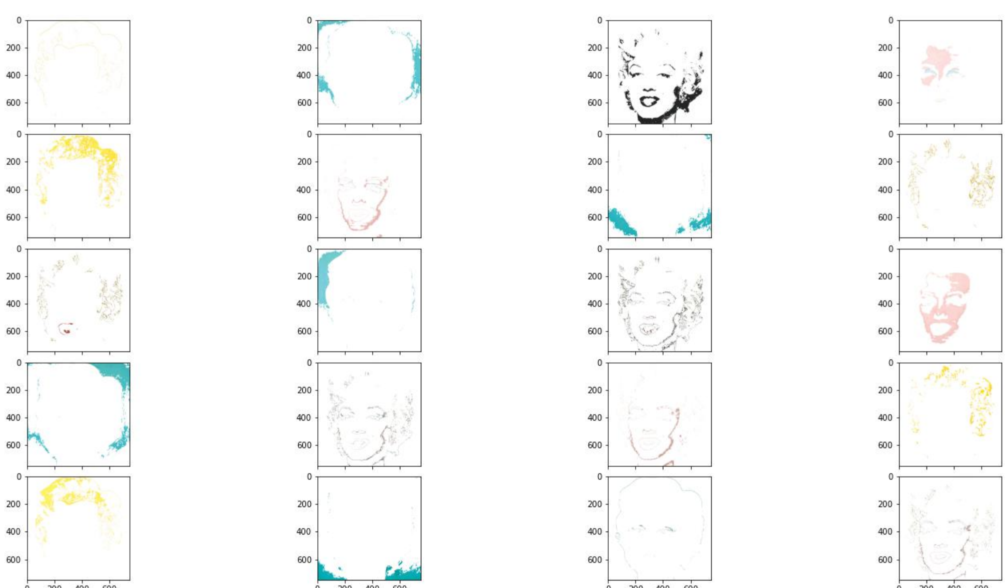

Figure 41: Resulting K-Means Clusters for Aqua Shot Marilyn

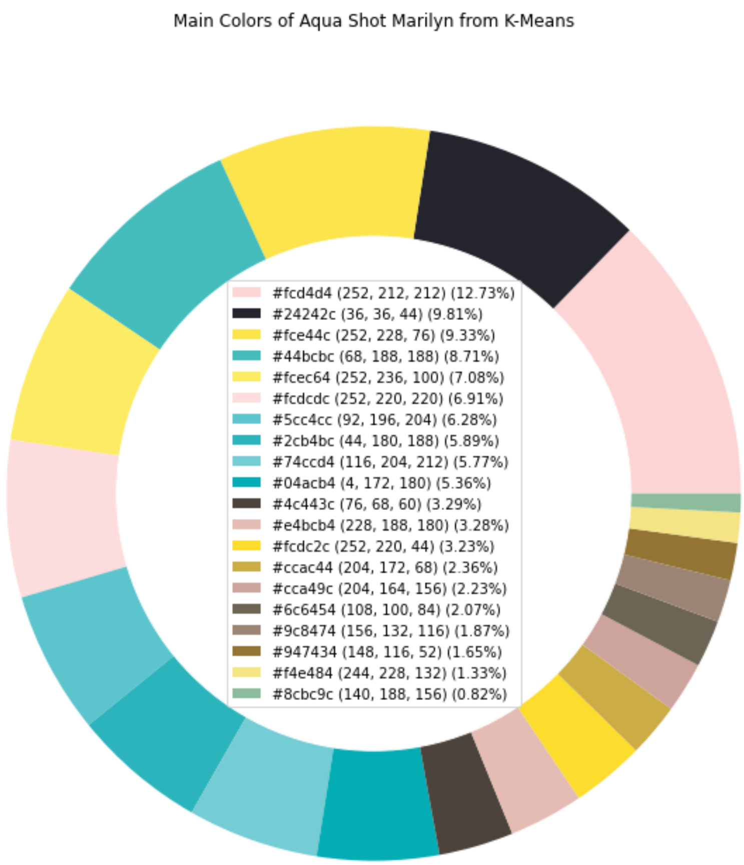

Figure 42: Donut Chart for K-Means Clusters for Aqua Shot Marilyn

The aqua shot Marilyn seems to be the most complex out of all the Marilyn images as evidenced by having more clusters after cutting the dendrogram at an appropriate height. As for separating the main parts of Marilyn, k-means was able to detect that the aqua background had 5 main components while hierarchical clustering only had 2. Also, k-means did well again in extracting the subtle complexities in Marilyn's hair and skin.

3.5 Blue Shot Marilyn

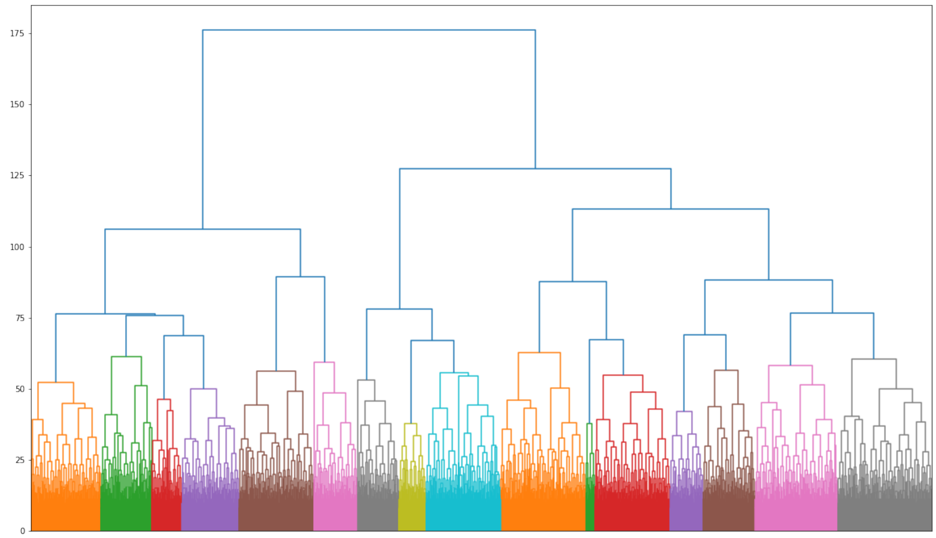

Figure 43: Dendrogram for Blue Shot Marilyn

For the blue Marilyn, we cut the tree at 65 as seen in figure 43.

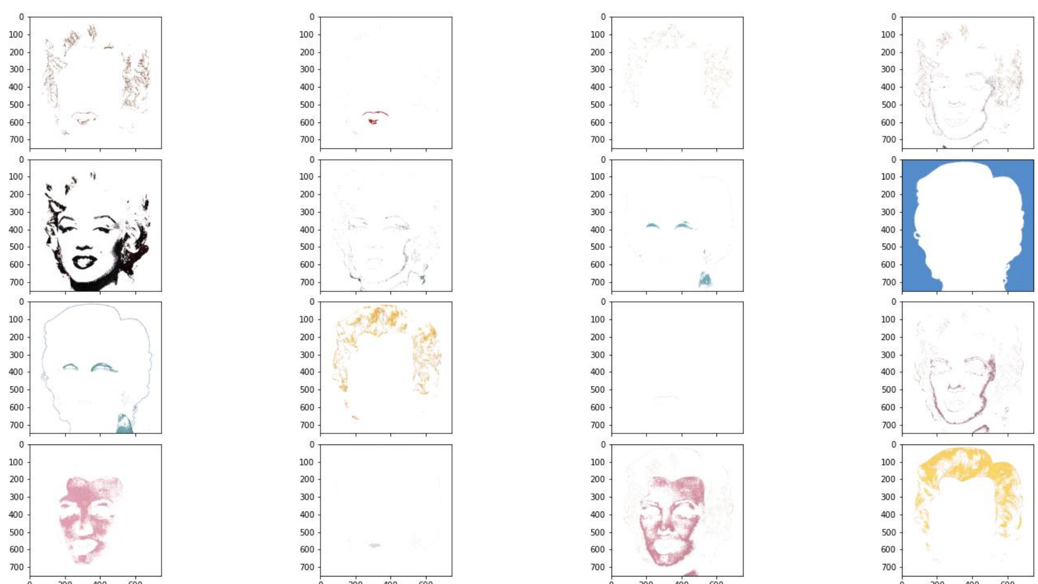

Figure 44: Resulting Clusters from HC for Blue Shot Marilyn

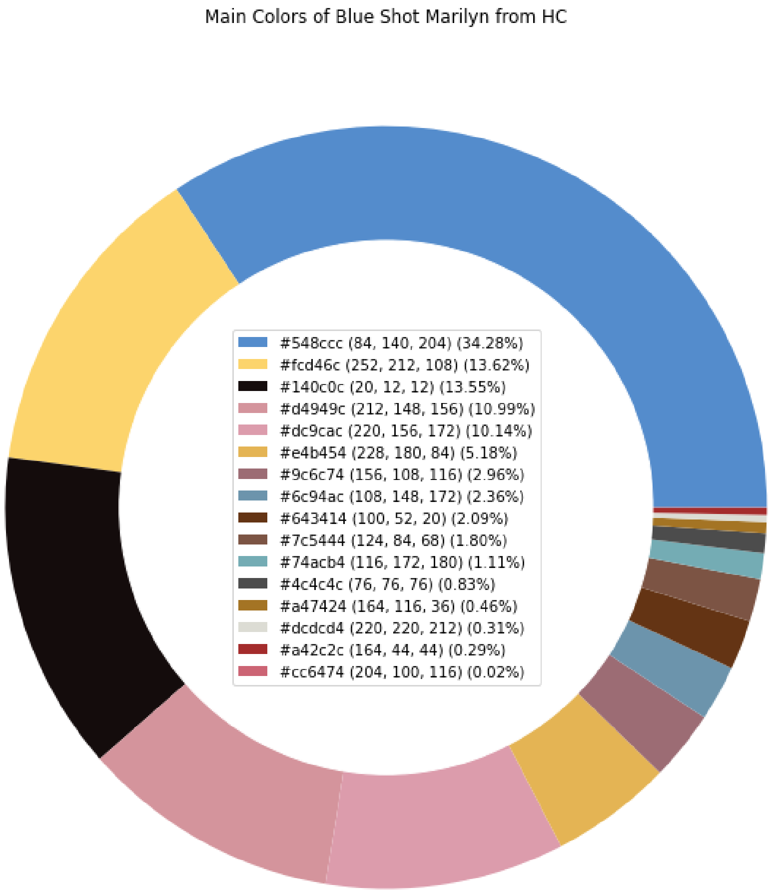

Figure 45: Donut Chart for HC Clusters for Blue Shot Marilyn

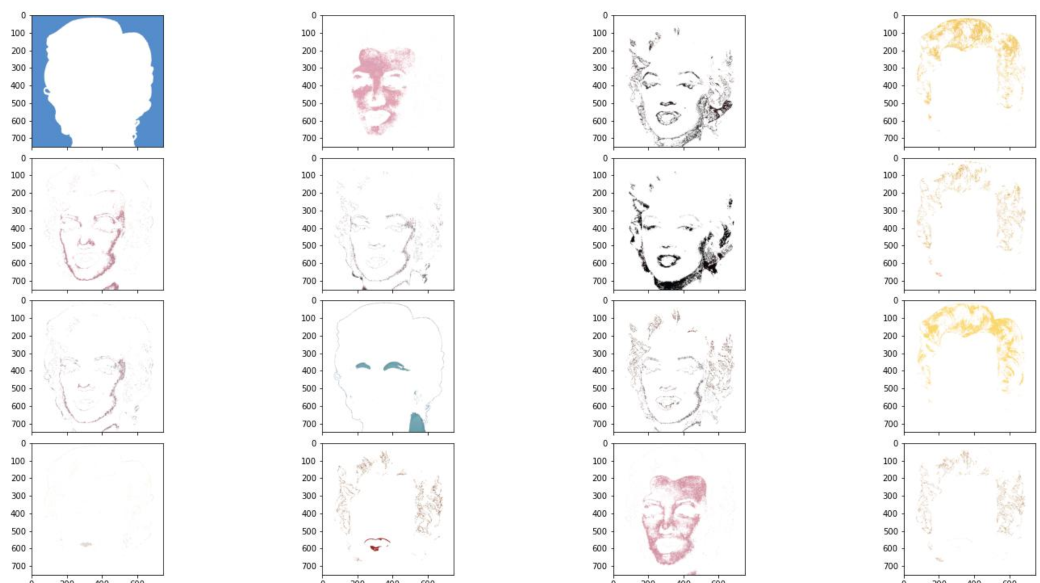

Figure 46: Resulting K-Means Clusters for Blue Shot Marilyn

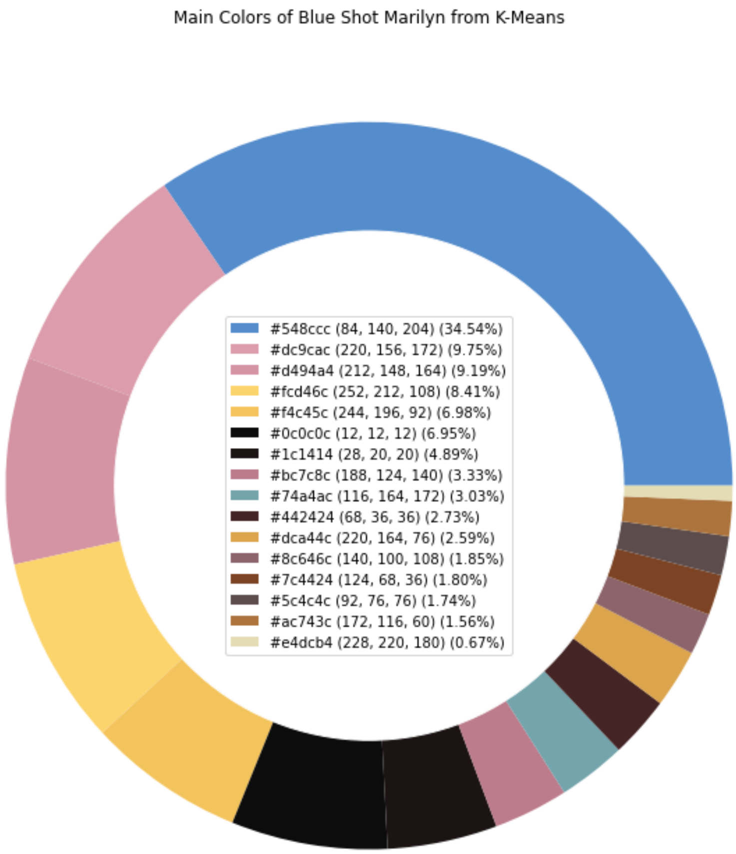

Figure 47: Donut Chart for K-Means Clusters for Blue Shot Marilyn

Surprisingly, performance was relatively the same for both models in terms of separating the layers of the hair and skin. However hierarchical clustering really outperformed k-means in finding the layers in the eyeshadow. Hierarchical clustering was able to detect that Marilyn had two layers to the eyeshadow as opposed to just one from k-means.

4 Color Transfer/ Color Change/ Color Correction

4.1 Image Recoloring

Using hierarchical clustering, we were able to deconstruct the images into many different components. By having the different components of the images, we were able to recolor the individual components of the image. When recoloring the original images, we had two main goals. We wanted to render the images in more natural lighting and we also constructed another image with colors specifically chosen with a distinct vision for a more artistic rendition of the image.

4.1.1 Red Shot Marilyn

When recoloring the red shot Marilyn image, we created six clusters to separate the image (we didn't care much about the number of clusters and the proper height, we just needed enough clusters to separate the image into its major parts). We chose six clusters because when used more clusters, the clusters were separating components that belong in one group. In the processing of clustering and recoloring, the contours of the nose are lost along. In addition, the bow in the background is also lost.

Figure 48: Red Shot Marilyn Recolored Naturally

In the image above, we recolored the red shot Marilyn image to recreate a more natural appearance. Compared to the original image, the skin tone is much less pink and closer to a real skin tone. The yellow in her hair was muted to appear more similar to a real blond tone. The color of the eyeshadow was changed to a coral tone in order to match the palette of the recolored image. Because of the lack of definition in this image and the color palette, this image brings the feeling of a comic, similar to the original red shot Marilyn image.

Figure 49: Red Shot Marilyn Recolored Artistically

In this image, the color of the hair is chosen to be brown in order to replicate a natural, brunette hair color. The skin tone is chosen to have olive undertones and the eyeshadow is chosen to be a nude tone for a more natural eye makeup look. Since the makeup is chosen to look more natural, the background is closer to a wine color in order to bring a more sultry feeling to the image, also similar to the feeling of the original red shot Marilyn image.

4.1.2 Orange Shot Marilyn

When recoloring the orange shot Marilyn image, we created ten clusters to separate the image. We chose ten clusters because a lower number of clusters was grouping separate components into the same component. In the processing of clustering and recoloring, the contours of the nose are lost along. In addition, the bow in the background is also lost.

Figure 50: Orange Shot Marilyn Recolored Naturally

In this image, we recolored the orange shot Marilyn image to recreate a more natural appearance. Compared to the original image, the skin tone is much more neutral. The yellow in her hair, again, was muted to appear more similar to a real blond tone. Due to the clustering picking up the tones of the shadows and the eyeshadow tones to be in the same cluster, we can see more of a gradient on her eyes rather than a solid shade. Here we can see that the bow has the same tones as the shadows and eyeshadow. The rendition of this image is low quality and simpler than the original image, so the image seems as though it may appear in a comic panel.

Figure 51: Orange Shot Marilyn Recolored Artistically

This recoloring of the orange shot Marilyn image, the intention of the image is to be jarring. The contrasts are a focal point of this image. The backgrounds and some other colors are washed out, however the shadows are much deeper and stronger. This composition of colors creates the effect of there being lipliner applied along with the lipstick apparent in this image.

4.1.3 Green Shot Marilyn

When recoloring the green shot Marilyn image, we created ten clusters to separate the image. We chose ten clusters because a lower number of clusters was grouping separate components into the same component. In the processing of clustering and recoloring, again, the contours of the nose are lost. In the clusters of this image, some components that do not belong together, such as the teeth and the background in the same cluster, were getting placed in the same component, even after we tried clustering with a larger number of clusters.

Figure 52: Green Shot Marilyn Recolored Naturally

In this image, we recolored the green shot Marilyn image to recreate a more natural appearance. Compared to the original image, the skin tone is much less pink and sun-burnt looking. The yellow in her hair, again, was muted to appear more similar to a real blond tone. Due to the clustering, the eyeshadow is not one uniform color and rather has a feeling of makeup applied with the intention of delivering a statement. The pops of teal on the darker base make for a contemporary eye makeup look. Again, due to the low resolution and the lack of detail in the recoloring, this image is also visually similar to a comic panel.

Figure 53: Green Shot Marilyn Recolored Artistically

This image has colors chosen to be bold and to create a strong impression on the viewer. The colors chosen are deeper and more saturated than the natural recoloring of this image. The background has strong, complementary colors that go well together. The eyeshadow colors are the same colors as the background, but in different ratios. The hair was chosen to be a deep, but vivid green to play along with the background colors. This image is also low resolution, however since the colors are so strong, it has the feeling of pop art, similar to the themes the original artist was aiming for.

4.1.4 Aqua Shot Marilyn

The aqua shot of the MM painting used 17 layers to separate all the major colors. The clustering process employed hierarchical clustering, and although there are 17 layers, the number of major colors is less than the number of layers. The clustering result is less than ideal due to the fact that lots of solid color blocks, such as the aqua background, were separated into different clusters, making it harder for the recoloring process. When the number of clusters is set to a lower number for hierarchical clustering, phenomena such as the mixing of colors occur, making it unsuitable for the recoloring process in the future.

Figure 54: Aqua Shot Marilyn Recolored Naturally

Due to the mix of the color on Marilyn's lips and the black outline, it was difficult to recolor the lip to a more natural color. However, the color for skin tone and color were recolored to a natural color to the best effort. The separation of the eyeshadow was also not ideal due to the fact that it was spread out among different layers, making it impossible to recolor the eyeshadow only.

Figure 55: Aqua Shot Marilyn Recolored Artiscially

The inspiration was taken from the blonde colored hair that Marilyn Monroe had. It was one of the more iconic features of her and it would be interesting to give her a different hair color. After many trials and errors, the color was decided to be platinum blonde, which is a different but interesting look for Marilyn. The background was also changed slightly to match the hair color better.

4.1.5 Blue Shot Marilyn

The blue shot of MM painting has its layers clustered using hierarchical clustering. The number of clusters we chose is 16 as this number of clusters separated the Marilyn painting sufficiently. Clustering successfully separated the background, hair color, skin color, and shadings for different areas. The clustering also successfully separated out a clean black outline. The well-performed clustering built a strong foundation for an easier recoloring process in the future.

Figure 56: Blue Shot Marilyn Recolored Naturally

Natural color restoration for the blue MM painting is one of the more successful restorations that use natural color as if it was a picture taken. Instead of the peach-toned skin the original artwork had, the newly recolored skin tone has a more yellow tint to it, which is more realistic in real life. The color of the eyeshadow was also changed into nude coloring since it is one of the most popular shades of eyeshadow. Thanks to the clean separation of layers, the recoloring of the painting looks relatively clean.

Figure 57: Blue Shot Marilyn Recolored Artistically

For the artistic recoloring, the background was left untouched and the color tone of Marilyn Monroe is changed to a color complementary to the background: orange. This color combination is inspired by the 80's aesthetic and the contrast in color makes the painting's main object pops out. The skin tone is slightly redder than the natural look and Marilyn's hair is entirely recolored to a color of ginger, the most orange natural hair color. The black outline is also changed to a softer color to be more homogenous with the rest of the painting.

4.2 Histogram Matching

One technique that was employed for color transfer and color matching is histogram matching. Histogram matching, also known as histogram equalization, creates multiple histograms of an image about its color channels for both the reference image and the source image. The histograms of the color channels represent the frequency distribution of each pixel intensity for that particular color channel. Additionally, the cumulative distribution function is shown in each histogram which is the occurrence of each pixel intensity value for both images. Then, the histogram matching redistributes the pixel intensities based on the histograms of the source image to attempt to match the reference image and produces the third image which is the color transfer/color correction. The histogram of all the red, green, and blue color channels has a CDF line as well which is depicted by the orange line the source below are the images of each five MM paintings and the reference image is the original portrait of Marilyn Monroe that Andy Warhol based his art on.

The reference image was taken from the Hamilton-Selway Fine Art website and this is the image after Andy Warhol had written notes on what he wanted to do with the colors [4]. The histogram for all color channels will be shown in conjunction with the color transfer, however, this portion will focus on how well histogram matching did for the color transfer of the source image to the reference image.

4.2.1 Red Marilyn Monroe Painting

Figure 58: Color transfer of Red Marilyn Monroe Painting to Original Portrait

The histogram matching process for color transfer between the red MM painting and the original portrait was reasonably effective. Many of the main colors such as yellow, red, blue, and magenta were successfully transferred over. The hair looks fairly realistic as it is not golden but more of a natural blonde. The blue color was the most accurate color transfer as it is distinct and well-colored. While the background of the source image did not transfer over well, and only half of the background is colored red while the other half of the background looks more like the reference image.

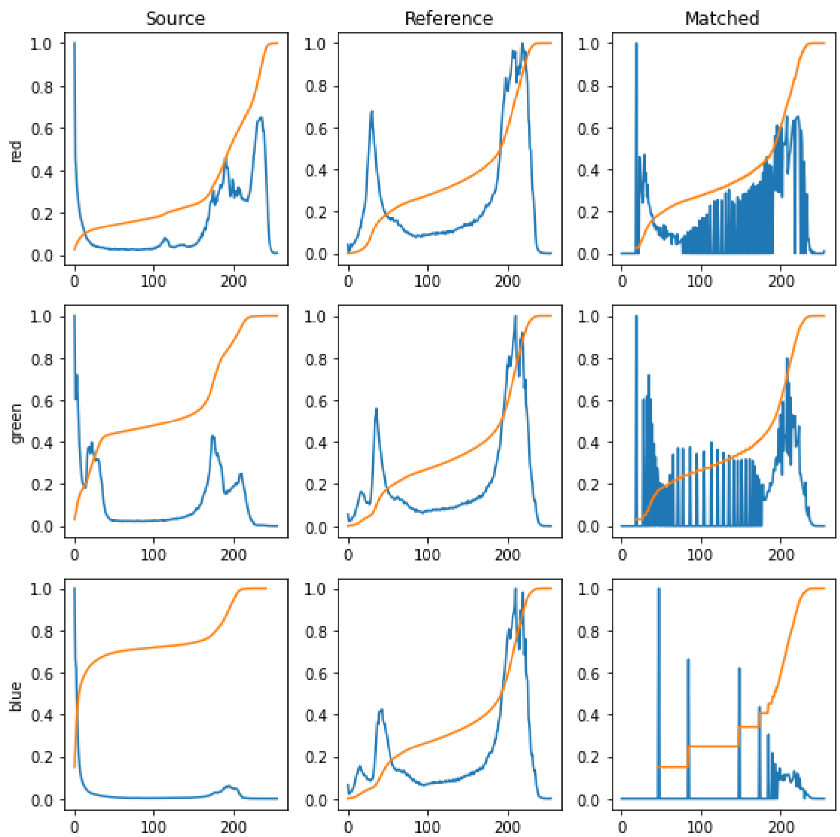

Figure 59: RGB Histograms of red Marilyn Monroe Painting to Original Portrait

4.2.2 Orange Marilyn Monroe Painting

Figure 60: Color transfer of Orange Marilyn Monroe Painting to Original Portrait

The histogram matching process for color transfer between the orange MM painting and the original portrait performance was a bit unsatisfactory. The color of the source image did not transfer as well because the orange background did not transfer over at all and mainly was left with the white color of the reference image. Additionally, the blue transferred over but it is not bold and distinct. While some of the yellow hair did transfer from the source to the reference image, the hair was not colored all the way. The only part of the source that was transferred over was the red lipstick well as it is bold and distinct.

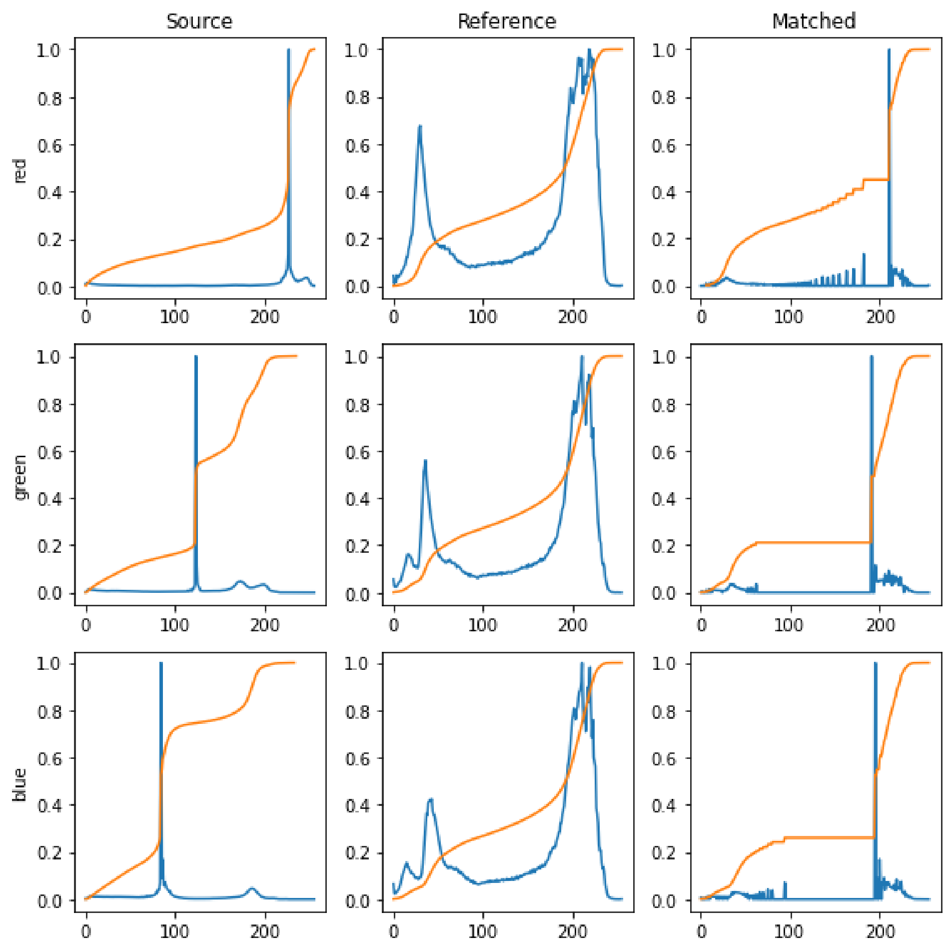

Figure 61: RGB Histograms of Orange Marilyn Monroe Painting to Original Portrait

4.2.3 Green Marilyn Monroe Painting

Figure 62: Color transfer of Green Marilyn Monroe Painting to Original Portrait

The histogram matching process for color transfer between the green MM painting and the original portrait performance was very successful because many of the main colors, such as yellow, magenta, red, blue, and turquoise, transferred over effectively. The turquoise background transferred well as can be seen as the majority of the background is now turquoise with small spots of white. The magenta face color transferred over along with the blue eye shadow and collar. The colors are not as strong as the original.

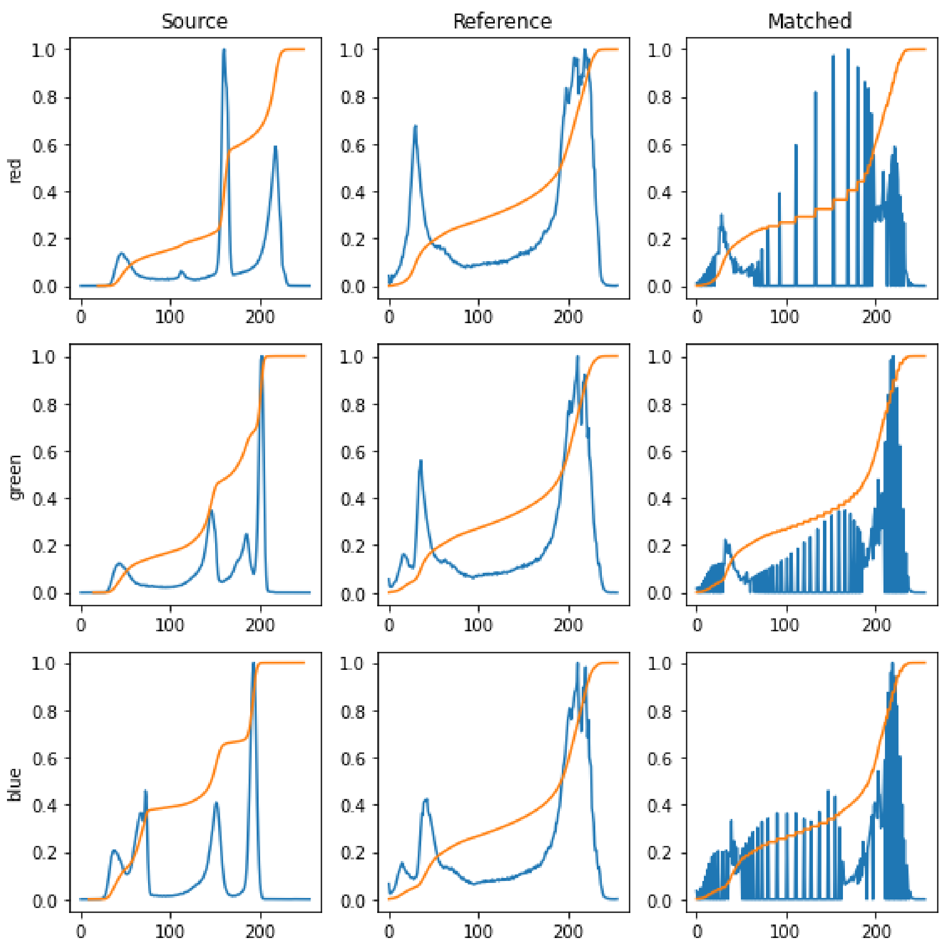

Figure 63: RGB Histograms of Green Marilyn Monroe Painting to Original Portrait

4.2.4 Aqua Marilyn Monroe Painting

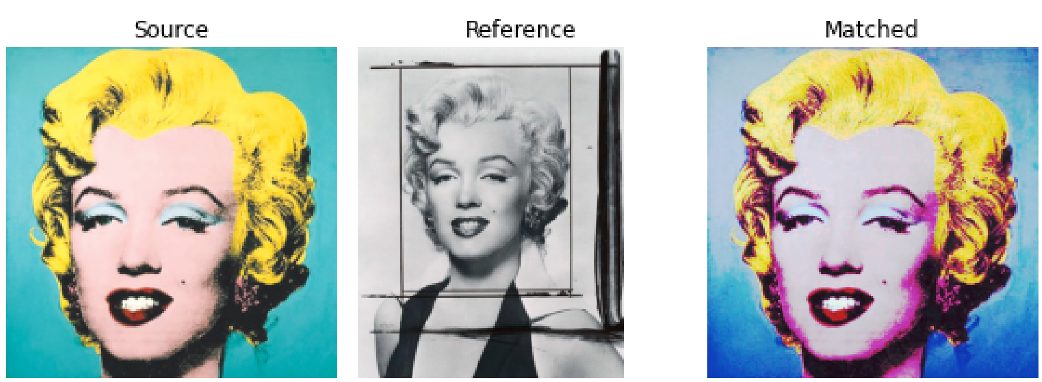

Figure 64: Color transfer of Aqua Marilyn Monroe Painting to Original Portrait

The histogram matching process for color transfer between the aqua MM painting and the original portrait performance was successful because many of the main colors, such as yellow, magenta and blue transferred over effectively. The hair color transferred super well as it looks very similar to the source image. The background color did not transfer as well as there are a lot of white spots. The face color of the matched looks similar to the source as it was supposed to be paler in this image. Overall, the color transfer could have been a bit better in terms of the background.

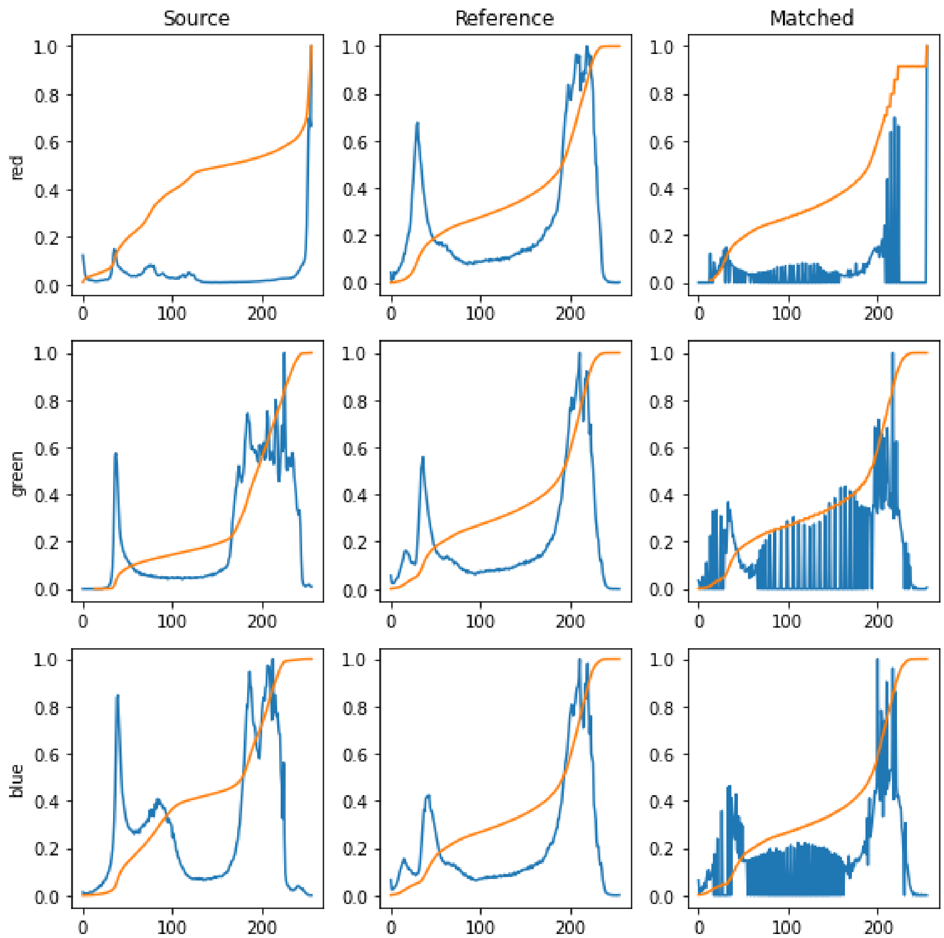

Figure 65: Histograms of Aqua Marilyn Monroe Painting to Original Portrait

4.2.5 Blue Marilyn Monroe Painting

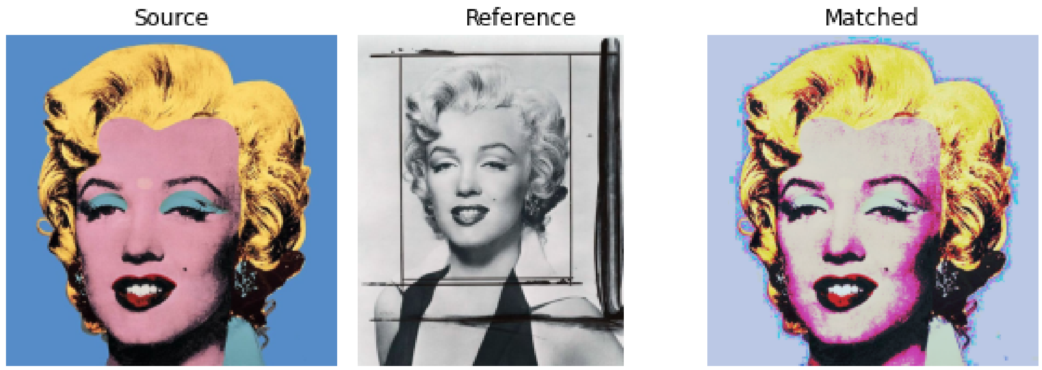

Figure 66: Color transfer of Blue Marilyn Monroe Painting to Original Portrait

The histogram matching process for color transfer between the aqua MM painting and the original portrait performance was successful because many of the main colors, such as yellow, magenta and blue transferred over effectively. The hair color transferred super well as it looks very similar to the source image. The background color did not transfer as well as there are a lot of white spots. The face color of the matched looks similar to the source as it was supposed to be paler in this image. Overall, the color transfer could have been a bit better in terms of the background.

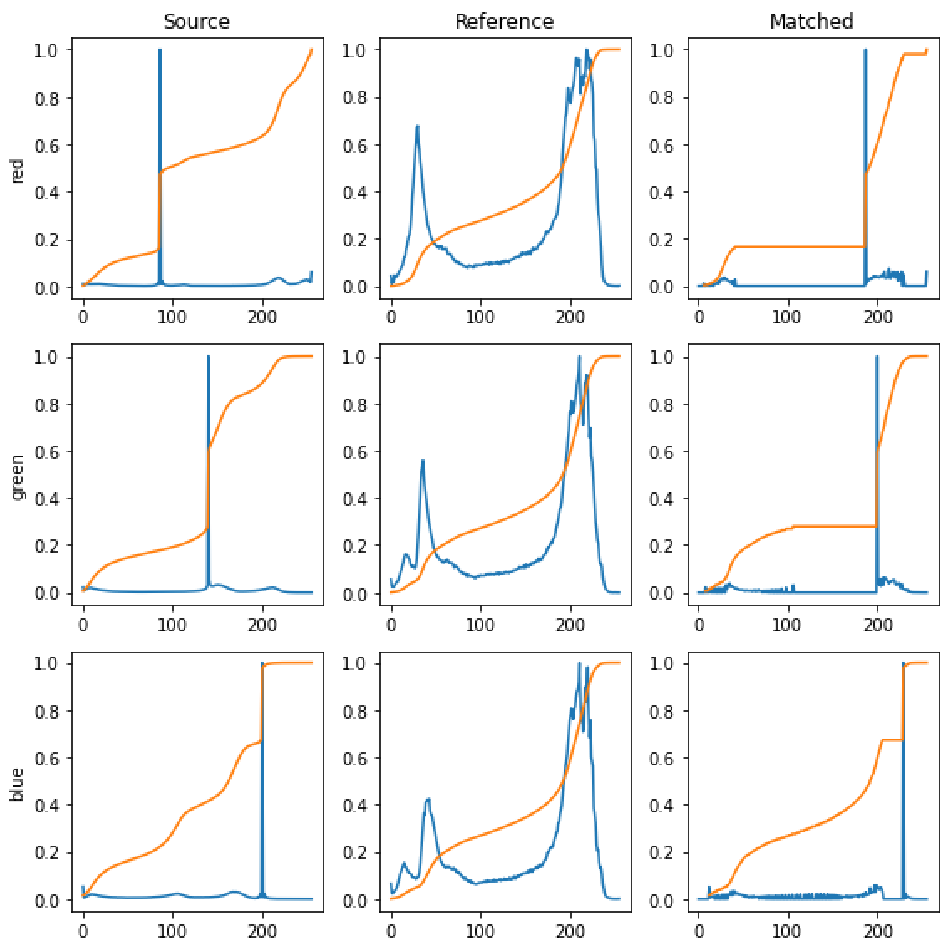

Figure 67: RGB Histograms of Blue Marilyn Monroe Painting to Original Portrait

The histogram matching process for color transfer between the blue MM painting and the original portrait was not as satisfactory compared to the past few color transfer. The background did not match at all and is a very, very light blue. Additionally, the darker magenta face color of the source was not shown in the matched image. The only color that was transferred successfully was the yellow hair as it is bold and distinct. The color of the collar and eye shadow did not transfer over well.

4.2.6 Compare and Contrast

For all images, the source-to-reference image matching created an effect where the original source image has aged and the colors have faded. Among all the color transfers, the green and aqua paintings exhibit the most successful color transfers as many of the main colors were reproduced accurately. For all five color transfers, the color that was most consistently transferred over was yellow most likely because it is the most unique color and how it is distributed in all five paintings. The orange color transfer performed the worse as many of the colors were reproduced and gave the appearance of the reference image of black and white with slight colors. This could be caused because of the way the colors are distributed in both images and how orange is collectively made by combining other colors. Overall, histogram matching worked fairly well between the source image of the painting and the reference image of the original Marilyn Monroe portrait.

5 Artistic Interpretation

In this portion, we will be exploring the five paintings of Marilyn Monroe from an artistic perspective. We will take a look at different elements of a painting such as color composition, lighting, saturation levels, the warmth of the colors, the emotions evoked from the images, and other aspects of the paintings.

5.1 General Notes

Just from taking a look at the images, it is visible that the colors have been added on top of a base image in black and white. We have prior knowledge of the artist retouching the original black and white images with color thought the silkscreen technique, but this is also evident visibly. We can also see that the coloring placement is not exactly the same for each image. For example, the line of where her eyeshadow extends to will differ in each image. In some images, the eyeshadow extends past the crease the contours of her eyes are visible and enhanced; however, in other images, her eyeshadow only reaches the crease, which does not emphasize the contours of her eyes. One more detail of the eyeshadow to note is that it appears to be one solid color rather than having a gradient. Typically when eyeshadow is applied, it appears as a gradient due to the shape and shadows of the eyes, adding darkness to more sunken areas and brightness to the higher points. This detail also supports the first impression of the colors being added later through digital tools. Another example of the image coloring being inconsistent is with her hair. The general shape of her hair differs in different images. We can also notice that there are different aspects of her hair being emphasized, highlighted, and contrasted in the different images. The next note on her makeup is that her lipstick is overlined in all the images. It is a common practice to overline the lips to achieve the appearance of larger, fuller lips. This choice in coloring may have been made because Marilyn is often seen as a sex symbol and this choice in coloring boosts her sex appeal.[5] The last note to be made is the inconsistent coloring of her clothing in the images. Some images emphasize that she is wearing clothing, but with other images, the color of her clothing is the same as the background which can make it appear as if she is not wearing clothing.

5.2 Red Marilyn Monroe Painting

The red painting is the most saturated images among the five recolored images. The color palette for this image includes warm tones. Because of this, the colors that stand out more are the pops of teal. The colors chosen in this image do not seem to simulate accurate lighting in a real environment. The vividness and intense saturation of the colors in this image paint a strong impression on a viewer. In combination with the lower resolution of this image, this image seems as if it belongs in a comic book. Because the image is low resolution, it also looks like it can be rendered with charcoal instead of a digital medium. The colors, specifically the deep red and vivid yellow, of the image, invokes the mood of consumerism and sex appeal.[6] Consumerism and sex appeal typically go hand-in-hand since companies tend to advertise by appealing to the senses and sex appeal will appeal to the senses.[6] The theme of consumerism is evident through the similarities of the colors of other major American brands such as McDonald's, Coca Cola, among other brightly colored brands. The sex appeal is apparent with the choice of the dark red background of the image. Because of the combination of consumerism and sex appeal present, this image feels the most objectifying and dehumanizing of Marilyn.

5.3 Orange Marilyn Monroe Painting

The balance of warm and cool tones in the orange image falls perfectly into the golden ratio of colors: a sea of warmth with splashes of cool and a counterpoint of saturated, bloody red lips. It seems that the lighting is not meant to be accurate to a real environment with lighting, but rather the artist meant to showcase their ability to create an emotional environment with color interactions and balances. This shows mastery of their ability to conjure a specific tone to contrast with the brashness of the red image. The background color is reminiscent of a simple orange, while the complementary blue eyeshadow is reminiscent of an academic setting such as a uniform or school color. This gives a sense that this version of color choice is meant to feel sophisticated and harmonious, thoughtful and nuanced.

5.4 Green Marilyn Monroe Painting

Compared to the red image, blue image, and aqua image, the green image has a much less saturated palette in regards to the background color, the hair, and the makeup. Compared to the blue image and red image, even the lips appear to be less saturated. Due to the desaturation and the skin appears to be as saturated as the other images, it makes the skin tone look much redder in comparison. Because of this, the skin tone is the highlight of the piece. Since the surrounding colors are desaturated, the impression the colors gives is a cool palette with the red skin tone being the focus, despite warm tones taking up a good deal of surface area. The image invokes the sensations of spring due to the way the background color interacts with the pink of the skin, reminiscent of cherry blossoms on water.

5.5 Aqua Marilyn Monroe Painting

The colors in the aqua image are saturated, however, it is less saturated than the colors in the red image. The choice of colors appears to be the inverted or opposite colors of the red image. The color palette of the aqua image is cool and her skin tone is a neutral color so as to not take away the focus from the yellows and reds. The most prominent color interaction is the background teal and the yellow of the hair because they are triadic colors. The lighting and the colors of the image seem to simulate real lighting and the colors chosen align with that lighting. The background color chosen is reminiscent of the ocean. Because of the choice of lighting and the background color choice, the overall theme of the image seems to be a day at the beach. The lighting, background, and color palette combined are reminiscent of a pinup model wearing a swimsuit at the beach.

5.6 Blue Marilyn Monroe Painting

For the blue image, the color palette is not quite as balanced. There is no immediately obvious winner between the cool and warm, but it seems that the hair is receiving the most initial focus, so it is a cool palette. The color of the background is reminiscent of the last moment of the sky being blue in the early evening. It feels like she is about to go out on the town and wearing furs and silk gloves to attend an award ceremony, bringing an elegant feeling. Her skin is much redder and the colors are much less accurate to a real lighting environment compared to the aqua image. However, there is still a feeling of environmental lighting, helped by the consistent early evening mood to give the audience a boost in imagining the scene. The yellow being the focus also draws the eye between the eyebrows where the painting was shot.

6 Conclusion

Andy Warhol, known as a leading figure in the pop art movement, was an artist bound to create greatness with every creation. One of Andy Warhol's goals was to immortalize Marilyn Monroe forever after her young death through these paintings.[3] The vibrant and bold colors that Andy Warhol chose would be a reminder that her legacy would live on. The purpose of the data science project was to create a deep understanding through the application of various data analysis techniques to gain insights and view patterns that were shown in each painting.

Firstly, exploratory data analysis techniques such as RGB density plot, 2D projection plot, correlation heatmap, and conditional entropy heatmap assisted us to discover the relationship between RGB colors in each painting. The RGB density plots were able to show the most frequent value of each color in a painting. Some results are easier to interpret, such as for the red MM painting, the most frequent value for blue is 0. This result is expected due to the overall tone of the painting. Some results are more complex and suggest a pattern in combinations of RGB values in the painting. The 2D projection plots displayed relationships between the value of 2 colors and used the value of the third color as the intensity of the data points. These plots show that all of the 2 color combinations reflect a positive linear relationship, meaning that when one color has a low value, the other color also tends to have a low value. The pattern of the intensity of the points is more variate and difficult to categorize. The correlation heatmap displays the relationship between each RGB color and the conditional entropy heatmap displays the dependencies between each color. Comparing two heatmaps of the same plot, it is observed that the values are an inverse relationship, meaning high correlation often means low conditional entropy, and vice versa. The correlation heatmaps presented the fact that red and green are typically highly correlated with each other, which could be explained by the fact that a large part of the painting is made up of MM's hair, which color is a combination of red and green. The conditional entropy between blue and red is among the highest in color combinations, indicating that these two colors provide relatively low information about each other.

Color Clustering was a crucial step in both identifying the key colors of a painting and future recoloring. Hierarchical with average linkage and K-Means clustering were applied to each painting and the results are visualized and compared for the better method. Overall, the two clustering techniques have similar performance but hierarchical clustering generally performs better in separate layers in the eyeshadow. On the other hand, K-Means clustering had a better performance separating layers of hair and skin.

Two forms of color transformation were done, one with clustering and the other with histogram matching. To recolor the images, a layer is observed each time and the layer's color will be changed to achieve the desired result. Different colored paintings have different numbers of optimal clusters, ranging from 6 to 18. Two recoloring images were created for each colored painting, one attempting to recreate a more natural look of MM, the other trying to create an art piece with uncommercial but aesthetic colors. Color transferring was performed to transfer color from the artwork to the black-and-white reference images. The performance for color transferring differs from color to color. For instance, the process of transferring color between the green MM painting and the reference image was very successful because all the major colors transferred over effectively. However, the transferring process between the orange MM painting and the reference image failed to transfer over either the orange background or the skin color, leaving the result unsatisfactory.

The findings of this analysis have contributed significantly to who Warhol was as an artist. A lot of colors were independent and seemed unpredictable from one another as shown in the Shot Marilyn Monroe paintings. The analysis backs up how he was an artist with unpredictable colors and strokes that were less about calculation between two RGB values and rather about what he believed was best. Even attempting to see how to "color correct" his images lead to inadequate paintings and trying different colors could not match up to the real versions. Unraveling his art through data science techniques only to find out his process was unique and iconic. The human mind behind these creations cannot be truly analyzed correctly as Warhol's artistry transcends data analysis as his art is composed of much more than logic. His art encompasses emotions, thoughts, and human empathy for his subject. Through his art, the essence of Marilyn Monroe was captured.

Acknowledgments

We would like to express our deepest gratitude to Professor Hsieh for the invaluable feedback on the exploratory data analysis our group conducted during each presentation. His knowledge and expertise on the subject of statistical practice on color extraction and color analysis were a great help. We are also thankful to the remarkable teacher assistants Shuting Liao and Abdul-Hakeem Omotayo for their assistance and insightful discussions. Their commitment to host office hours to answer various questions helped many groups, including ours. They have our sincere thanks.

References

[1] Andy Warhol. Marilyn Monroe. 1967: Moma. The Museum of Modern Art. (n.d.). https://www.moma.org/collection/works/61240

[2] Encyclopædia Britannica, inc. (2023, May 30). Pop art. Encyclopædia Britannica. https://www.britannica.com/art/Pop-art

[3] Person. (2022, April 21). Andy Warhol's Marilyn: An icon of beauty. Christie's. https://www.christies.com/features/warhol-shot-marilyn-12161-7.aspx

[4] Andy Warhol: Marilyn Monroe portfolio: 1967. Hamilton. (2022, February 10). https://hamiltonselway.com/portfolio-item/marilyn-monroe/

[5] HERTEL , H., amp; NEFF, D. (1962, August 6). From the archives: Marilyn Monroe Dies; Pills Blamed. Los Angeles Times. https://www.latimes.com/local/obituaries/archives/la-me-marilyn-monroe-19620806-story.html

[6] Hu, F., Wu, Q., Li, Y., Xu, W., Zhao, L., amp; Sun, Q. (2020, May 29). Love at first glance but not after deep consideration: The impact of sexually appealing advertising on product preferences. Frontiers in neuroscience. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7273180/

[7] Tom Leonard In New York For The Daily Mail. (2022, March 22). The bullet hole that turned Andy Warhol's Monroe into a £150million masterpiece, writes Tom Leonardnbsp;. Daily Mail Online. https://www.dailymail.co.uk/news/article-10641805/The-bullet-hole-turned-Andy-Warhols-Monroe-150million-masterpiece-writes-TOM-LEONARD.html