Predicting the Next Fashion Color Trend with Machine Learning

Lorena Estrada   
*230149445*Dr. Christopher Phillips   
MSc Computing and Information Systems

*Abstract*— Machine learning improves prediction accuracy and data analysis. Many evaluation techniques can be used, including K-Means and Hierarchical clustering. Machine Learning has evolved into many sectors, including the fashion industry. The aim of this analysis is to use machine learning techniques to predict new fashion color trends.

Keywords—Machine Learning, Unsupervised Learning, K-Means, Hierarchical Clustering, Prediction

# Introduction

A group of mathematical symbols

Description automatically generatedThe digital world has changed everyone's way of working and living. From shopping in person to shopping online, from looking at magazines for inspiration for new styles to seeing your favorite celebrity online and copying their style, the way we once used to do has, in some way or form, become an online or digital presence. One of the industries that digital technology has completely changed is the fashion industry. Consumers have changed the way fashion is portrayed and bought because of the high influence of technology. Users have created a constant need for fashion trends by constantly searching for new inspiration. Causing the industry to develop a system that facilitates product searching and product recommendations(Lee, et al., 2020). Thus, there is constant change not only in the fashion industry but in different sectors that have gone fully digitalized. In order for industries to be on top of their consumer goals, they must be precise in what they are portraying. Industries such as fashion and apparel must make the best predictions of trends, styles, and colors because, therefore, their businesses will be the best consumed by people. Being able to predict accurately has a high impact on this industry, as the global apparel market is estimated at USD 3 Trillion (Al-Halah, et al., 2017). This allows us to see how significant the fashion industry is. Measuring and analyzing trends is a complicated task. Machine learning has influenced many aspects of the fashion industry since trend cycles are unexpected and short-lasting; machine learning technology enables analyzing trends quantitatively and quickly with accurate findings. As a result, there is now an extensive use of machine learning and big data (Han, et al., 2021).

The purpose of this study aims is to test how machine learning techniques can help the fashion industry make precise predictions. The research will involve a) the collection of data from online fashion runway shows done in the Summer of 2024, b) the application of correct machine learning techniques to access what could be the next color trend for Fall 2024, c) the evaluation of the techniques and performance of the models created.

# Literature review

## Machine Learning

Machine learning starts not with a technological approach but started with psychology. Psychologist Franklin Rosen proposed the first ideas of what is now known as machine learning through his creation of The “Perception,” which derived from perceiving and recognizing automation of the human nervous system. The “Perceptron” would work by accepting several inputs x i, i = 1, . . . ,N, and intended to calculate the sum of the inputs and resulted in a fixed weight “w” that could only be +1 or -1. The sum would then be set with a threshold of 0, and a Y output would result in 0 or 1 (Kanal, 2003).

*Figure 1. Perceptron Structure*

Leading forward to the 21st century, three moments moved what is known today as machine learning. Through the advances of data in the past years, new ways to use and build data were being discovered. This includes the first significant point in machine learning, which is big data. As a result of the development of searching for new data came the use of more memory, which would cost companies revenue. Companies like Google and Hadoop became aware of this trend and introduced technology to improve the massive volumes of data within simple processors. Google introduced MapReduce in 2004, followed by Hadoop in 2006, which had an open-source counterpart. During this time the cost of RAM decreased making it possible for more corporations to work with large amounts of data. Later, NoSQL and the Apache framework emerged, becoming helpful in implementing machine learning algorithms. Machine learning was derived from these concepts and years of research. Today, what is known of machine learning is due to this (Fradkov, 2020).

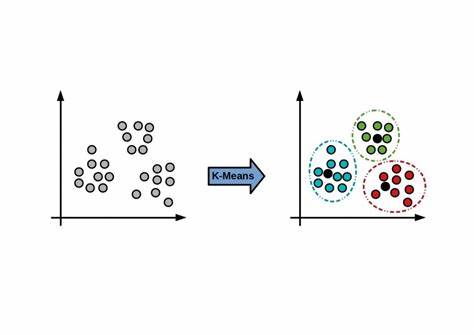
The use of machine learning is defined by the use of computational methods to enhance output and the ability to correlate predictions. Information used in machine learning can be given by data collected electronically or physically inserted by the user. Data gathered can be in the form of training sets with user-made labels or by research interacting with data. The algorithms in machine learning are made to enhance performance or to develop predictions from the data given. For most analyses, large data sets are easier to access, but the labels and columns of these data sets equally form heavily on performance and prediction. The effectiveness of success depends on the data used, which is why machine learning correlates with statistics and data analysis. These techniques connect computer science with mathematical uses such as statistics and probability to better illustrate data-driven techniques. (Kanal,2003) (Fradkov,2020)

## Applications of Machine Learning

Machine learning covers two areas: supervised learning and unsupervised learning. Utilizing these two machine learning algorithms helps identify sequences and patterns by using a data set.

During supervised learning the dataset is known and includes labeled input data, trained data that can be used to form predictions. The labels in supervised add value to the performance of the algorithms. This is because the dataset is more accurately identifiable for the machine to predict. There are two distinct algorithms in supervised learning: classification algorithms where the data is divided into separate “classes”; this type of data is used for categorical response values (Gutierrez Portela, el al., 2019). In unsupervised learning, the approach is different from that of supervised learning. The unsupervised data set doesn’t involve labeling, which lets the algorithm interpret the data patterns or distinctive features by analyzing tags or color repetition. Evaluation techniques include clustering of data, which analyzes data clusters based on where it is most repeated, and performance metrics involving accuracy, F1 score, ROC, and precision. During this research, unsupervised learning will be used. This comes as an advantage as unsupervised learning provides a new perspective and uncovers patterns in the data, which leads to meaningful groups of information. (Patel,2018)

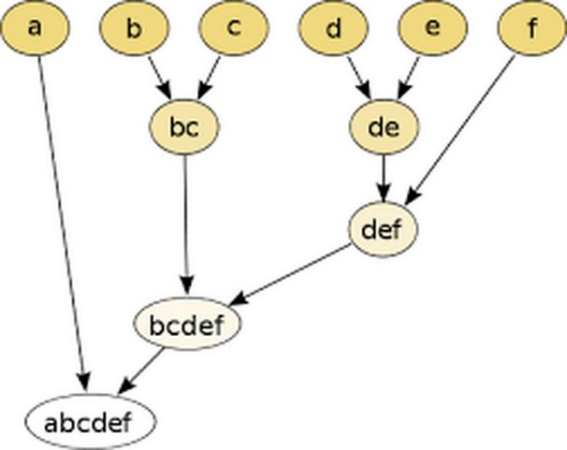
The clustering methods applied in this study will be K-Means clustering and Hierarchical clustering. Clustering is used to identify patterns found in unsupervised learning algorithms. In K-Means clustering, the algorithm will choose a specific amount of K-clusters, provided that the number of clusters is defined prior. Clustering will optimize the within-cluster variation and reduce it as much as possible. This is referred to as inertia; it will average the within-cluster throughout all the K-clusters (Patel,2018). The K-Means algorithm is commonly referenced to Llyod’s algorithm, which is a common factor for solving K-means. Lloyd’s algorithm works by setting k centers for each center “ci”, letting “vi” denote the set of data points for which ci is the nearest neighbor. For the next iteration of the algorithm, replace “ci” with the centroid of “vi” and update “vi” accordingly. (Kanungo et al., 2000)

**

*Figure 2: K-Means Clustering*

Hierarchical clustering is an alternative technique to K-Means. This method explores unsupervised learning datasets by creating a binary merge tree and builds what is called a dendrogram (Patel,2018) (Nielsen,2016). In K-Means, we see a predetermined amount of clusters. Hierarchical clustering begins in what can be described as an upside-down tree by depicting individual data at the leaves. It incrementally merges the closest subset based on their encage distance. The clusters are determined in hierarchical clustering by selecting a horizontal cut level; the smaller the cuts, the more clusters; the more significant the cuts, the fewer clusters produced. This clustering technique offers greater customization than K-Means or other methods that require predefined clusters. (Patel,2018).

*Figure 3: Hierarchical Clustering*



Machine Learning has brought a new perspective to decision-making and predictions. Surprisingly, an algorithm can detect faces, colors, shapes, and even written essays. Although human intelligence is still predominantly the most effective way to be certain of decision-making or predictions, machine learning helps give human tests a new perspective and improve the goal of what is being evaluated. (Kleinberg, et al., 2017)

## Machine learning in Fashion

Since the 18th century, the garment industry has been predominant in industrial transformation, especially in Europe. As time passes, the industry changes, trying to adapt to new technologies, innovative production, social cost, and sustainability. Historically, fashion has played a large part in the world's economic stability. In the late 20th century, the introduction of outsourcing created a significant demand for global supply chains. Fast forward to 2024 in the age of technology, the industry has moved forward, trying to incorporate the tech industry into its culture to continue with the times. Implementing technologies like machine learning, artificial intelligence, and online shopping creates a technological environment to continue with the fashion movement. (Bertola, et al., n.d)

Machine Learning in many industries has become a rising topic. In fashion, data analysis has become popular in the industry and in top fashion academics, like the University of arts London offering fashion analytics and forecasting courses. The increase of interest in the data world has come from the effectiveness results from data driven decision making. WGSN, Stylumia, and GoFind.AI are some of the major companies that have been using machine learning to analyze consumer, market, and product trends by using web scraping, social media, and online images. The objective of these companies is to analyze massive amounts of data in relation to fashion. The field called “Fashion Informatics” uses social media, image recognition, and machine learning to aid the fashion industry with data supporting trend forecasting, product recommendation, and strategic sales plans. These studies show how image data analysis can be helpful by classifying clothing and examining the color in the images. Companies like WGSN, which has just recently launched an AI driven platform for buyers, which uses machine learning to classify trends (Kennedy,2024). A fundamental area of fashion data analysis image recognition which distinguishes between clothing and unnecessary background images. Although much research has gone into this field, much more is needed to advance the full benefits of machine learning in a fashion centric approach.(Han, et al., 2021) (Kennedy,2024)

Image recognition technology has been widely used in life. For example: The scene recognition through deep learning algorithm which can automatically recognize some common scenes in images such as sky, grass, people, and so on. Based on that function, the client’s application can easily realize the automatic management, grouping and search images, complete the intelligent management of large image library, and save a lot of time. The significance of image recognition technology is that it frees people from heavy and mechanical repetitive work and gives them more time to deal with other more meaningful things, thus greatly increasing the efficiency of working. The traditional machine learning algorithmic system has been very mature, and a lot of them can be applied to image recognition. For the present computer processing and computing ability, it is easy to use traditional machine learning methods to solve some problems. (Lai, 2019)

# METHODOLOGY

## Data Gathering

From this analysis, the objective is to gather enough data from web scraping Vogue Runway Fall 2024 Ready-to-Wear designer highlights. Together with a correct evaluation of collecting cultivable runway, we show an application of machine learning techniques that can be utilized to determine a color class prediction using our analysis. The analysis would provide insight into what would be seen in color trends in Fall 2024. Vogue Runway counts with hundreds of designer runway highlights to choose from. Utilizing research, a justifiable amount of designers can be web-scraped and used to create a dataset. Following data collection, an exploration of the data can take part. Since the data is being made and has not been an altered dataset, this would be considered an unsupervised dataset since it does not count with labels or continuous data that can be used like a supervised dataset. This being done an application of machine learning techniques can be applied. Hierarchical and K-Means clustering will be implemented to further the analysis. Based on the results of the clusters of K-Means and Hierarchical clustering, an evaluation of the findings can be used to come up with an educated analysis of what colors could be seen most in trend in the Fall of 2024. The research should give us a general approximation of the color trends in Fall 2024 from the use of the algorithms and data presented for analysis.

Selecting the data begins in Vogue Runway. Relevant data surrounding fashion could be found in various pre-made datasets. Since the approach is to find the color trend of Fall 2024, no data is known since it is the current year, and what we are aiming to find is only relevant if it is made from web scraping. Different paths can be taken to discover current fashion data; in this instance, the investigation starts with online research of fashion magazines. This approach was taken because fashion magazines have participated in the affluent trends since the 1900s (Cox, et al., 2012).The decision to choose Vogue was made from research concluding that the magazine has been around since the 1930s and continues to show the world’s best designers, entertainers, and actors (Borrelli-Persson, 2017).Although other magazines like Harper Bazar, Glamour, and many others do the same, Vogue began and continues this type of advertising; therefore, this research will take from here (Cox, et al., 2012).

Considering the data used for this analysis of the color trends for the Fall of 2024, the designer's collections for Fall 2024 is conducted in the Summer of 2024 to be presented to the public and fellow fashion enthusiasts worldwide. Each house of designers comes out with a collection of around 30-80 looks; this can be found in the Vogue Runway Fall 2024 Ready to Wear collection (Nast, 2024). Since the Vogue Runway website offers several hundreds of designers, how designers would be in the dataset was to take some of the biggest names in designer fashion, according to organizations like Forbes, Who What Wear, Marie Clarie, and Vogue Magazine, some of the biggest names include Gucci, Louis Vuitton, Chanel, and many others (Nast, 2023) (S.H,2024) (Nichols,2023) (Loeb,n.d) The dataset will show 30 designer houses in Vogue Runway that align with the top designer brands. The images in Vogue Runway include the designer's name, followed by a slideshow of the Fall Ready-to-Wear 2024 collection. The runway shows during this time were conducted in several locations, including New York, London, and Paris (WWD,2024).

*A screenshot of a fashion show

Description automatically generatedFigure 4: Vogue Runway Fall 2024 Ready-to-Wear Website*

## Data Implementation

As clearly stated, this analysis aims to extract the Vogue Runway Ready-to-Wear 2024 images from various designers. Given that information, the goal is to create a dataset including the name of the designer, image URL, dominant color extracted, and color classification. Following obtaining that data and creating data, data exploration is applied to see if any data must be dropped or altered, and data is studied to understand better what is being perceived. Subsequently, the data being examined will be prepared for the application of machine learning clustering. Data is prepared for feature selection so that it may be used for clustering. Ultimately, the data is separated to introduce K-Means clustering and Hierarchical clustering. The examination of which colors are the most dominant will be determined during clustering. Finalizing our analysis with visualization of the clustering and how it concludes which are the most dominant colors.

The process of obtaining the basis of this analysis of the next color trends will be conducted in Jupyter. To be capable of obtaining the images from Vogue Runway, we must use the HTTP pathway for each distinct designer. Python can get images via HTTP, APIs, HTML, and XML; it is only necessary for the correct libraries to be imported for testing.

A close-up of a text

Description automatically generatedThe primary libraries needed for the start are Requests, Beautiful Soup, Image, BytesIO, Counter, and Pandas. The Request library is used for HTTP requests, making it possible to connect with APIs and web-based services (Reitz,2023). Subsequently, Beautiful Soup is applied for parsing HTML or XML files; in this case, we use it for web scraping the path of Vogue Runway designer images. BytesIO allows binary data to be used like a file and works for image processing (docs.python, n.d). Counter is used to count objects in a dictionary subclass (docs. python.org, n.d). Image is used for the application of opening, manipulating, and saving image files (Python,n.d). Lastly Pandas which is a widely used library in data analytics and machine learning. In this instance, it is used for data manipulation to provide a data structure like a data frame. (Pandas,n.d)

## Once the application of imported Python libraries is made, the creation of web scraping and image processing can occur. For the application of machine learning a dataset is needed. This is why web scraping and image processing are applied to create a data set since we are collecting current data from images. Web scraping can be used to count the colors of images being extracted. The dataset was first implemented by web scraping to get the HTML information from the Vogue Runway URLs. Thirty designers or companies were chosen, and each designer or company URL was placed to scrape. Alongside scraping the URLs, the company's or designers' names were also scraped to comprehend better which designers' images were alongside the colors in the data frame. Following this, the URLs were extracted and set in the data frame as well. Next, the image's dominant colors are fetched and set as RGB values, making it easier for Python to get the colors this way. Additionally, color classification was collected by calculating the average brightness of the RGB value. Ultimately, the data frame is assembled and set as company, image URL, dominant color, and color class. The data frame would be saved and named ‘Fashion\_Show\_Colors\_Filitered.csv’.

During the first attempts to create the data frame, the CSV would be produced as expected. The data was explored during this course, and companies or designers who did not portray the collected ones were shown in the CSV. Further exploration was done, and we saw that advertising and magazine editors' images were also being scraped. Given this filtering was added, the URLs of these specific outcomes were dropped. It was observed that there was a pattern in the data set, and code was implemented to filter that data out. Without filtering these URLs, the analysis would be less accurate since it would be scraping colors that are not in the designer's collection.

*Figure 5: Fashion\_Show\_Colors\_Filtered.csv*

   Having obtained the dataset, we cannot immediately go into machine learning clustering. The dataset must be examined and understood to comprehend the final clustering analysis fully. Multiple steps have been taken to learn more about the dataset. Initially observing what is in the data, which we now know is the company, image URL, dominant color, and color classification. The comprehension of how the dominant colors were displayed, which are in an RBG value. RGB values stand for red, green, and blue, which are the primary colors of light; every color displayed in the RGB model is shown in sets of three values, e.g. (255,255,255). Every color has an intensity that ranges from 0 to 255; 0 indicates no intensity, and 255 indicates the fullest intensity (W3schools,2019). The ‘dominant colors’ column in the dataset would be represented like this from the extraction of the images depending on the intensity of the color in the image. Additional exploration of the dataset was done; the data types in the dataset are all object types and not null values, and the shape of the dataset, which is 854 rows and 4 columns, was learned.

  To further see visually how the number of dominant colors, including dark, light, or brown/beige, were classified as belonging to the color class displayed. Further examination of data was done to visualize this. A new Python library called matplotlib.pyplot was imported to plot the graphs. matplotlib.pyplotlib is a state-based matplotlib interface that offers plotting methods similar to MATLAB (matplotlib,n.d). Python libraries also used were counter and pandas for color occurrences and data manipulation. Firstly, the CSV file ‘Fashion\_Show\_Colors\_Filtered’ is loaded to filter out dominant colors and color classification. The dominant colors must be converted to integers as they were in the string representation of RGB values. Converting the RGB values into tuples to count the frequency of each color. The tuples would be converted into hexadecimal color code to plot in Matplotlib. As mentioned, RGB values have high and low intensity to classify the colors dark, light, and brown/beige; less than or equal to 100 would be dark, greater than or equal to 600 would be light, and red, green, or blue values between 60 and 200 would be brown/beige. Each color would be classified and counted on the graphs. Resulting in plotting the three graphs as dark colors, light colors, and brown/beige colors. (Full image in appendix)

A chart of colors with numbers

Description automatically generated with medium confidence

 Continued investigation to gain a better understanding was to import the Python library seaborn. Seaborn is used in Python to give a high-level visualization of graphical scenarios (seaborn,2012). After calculating the different color classes, implementation of how many dark, light, brown/beige, and others were plotted to seaborn to gain a deeper understanding of how the color classification was distributed among various designers/companies. Three graphs were produced: a heatmap of the color classification counts by company, a bar plot of each color class by company, and a count plot displaying the number of occurrences of the color classification per company. These visualizations offer an in-depth understanding of the distribution and relationship of color classification among the different fashion companies, which permits us to thoroughly analyze the color trends in the Fall 2024 Ready-to-Wear collections. Ultimately, with the results of the graphs, the calculation of exactly how many dark, light, brown/beige, and other color classes were in total. By applying value counts and count plots to visualize, we can see the total number of dark colors: 341, light colors:258, brown/beige:18, and other:241. Overall, the data exploration gives us insight into what we might see as a result of our final testing phases.

*Figure 8: Heatmap of Color Class Counts by Company*

A graph of different colored columns

Description automatically generated*Figure 6: Proportion of Each Color Class by Company*

A graph of different colored lines

Description automatically generated*Figure 7: Count of Each Color Class by Company*

In the final step, feature selection is performed on the dataset's dominant colors before entering the clustering phases. Feature selection is the process of improving the accuracy scores of estimators or enhancing functionality on very high dimensional datasets. To use the feature selection method, Python libraries are imported, including Pandas and Matplotlib. Python, which was previously used. The new libraries shown are StandardScaler and ast. StandardScaler is commonly used in data analytics projects. The process standardizes features by eliminating the mean and scaling the unit variance. The use of StandardScaler is quite helpful in machine learning algorithms that are sensitive to high amounts of data (Scikit-Learn,2019). Abstract syntax tree – ast helps parse Python code literals and create actual tuple objects (docs.python.org.,n.d).

The dataset is then loaded to read the CSV. The dominant colors are filtered to get the RGB values as R, G, B and can be filtered separately instead of the three as a string into a tuple. The separate R, G, B values are then normalized for the purpose of ensuring it has a mean of 0 and a standard deviation of 1. Using StandardScaler, the R, G, B values are standardized; this is important for the use of machine learning algorithms, especially if using K-Means clustering, as it guarantees features contribute precisely to the distance calculations.

Finally, after exploring, visualizing, and featuring the dataset, we can use the first machine learning technique, K-Means clustering. Normalized filtered data is first introduced as a feature of the clustering algorithm. Before applying K-Means clustering, we must find the optimal number of clusters in the K-value; if the K-value is not selected correctly, the analysis can be incorrect. Different approaches can be taken, the input of various numbers can be inserted, or the Elbow method can be applied. The elbow method helps select the best K-value possible for specific analysis. This method is best for small K-values and calculates the squared difference between different K-values (Cui,2020). Inertia is the total squared distance between all points and the cluster centroid assigned to it. Lower inertia indicates closer clusters, which typically means better clustering (Gupta,2019). The K-values range from 1 to 10 and are tested and returned with inertia. The results are shown with length K, length of inertia, and inertia values. The results can then be plotted using matplotlib.pyplot, to better visualize the results of the elbow method.

Given the results of the Elbow method, the application of K-Means clustering can be done. The Elbow method gave K=3 as the number of clusters. Knowing this, K-Means is fitted to the number of clusters and labeled clusters in the data frame. The dataset will show the clusters to each R, G, B value color. The clusters can then be analyzed in normalized and original format. Both are applied to interpret better the dominant colors presented by each cluster. Visualization of the normalized cluster demonstrates the dominant RGB value colors. The final findings will be further investigated in the research.

*A graph with a line

Description automatically generatedFigure 9: Elbow Method*

To move on to the second clustering, hierarchical clustering is used in this analysis to test the same thing using a distinct approach. Although different, this type of clustering involves similar factors. Python libraries used are pandas, ast, StandardScaler, matplotlib, and SciPy. Cluster. Hierarchy. SciPy. Cluster. Hierarchy functions for hierarchical clustering are helpful for calculating statistics on clusters, linkages to create flat clusters, hierarchical clusters from distance metrics, and visualization of clusters using dendrograms (docs.scipy.org,n.d). Once again, the dataset is loaded and filtered into R, G, B normalized using StandardScaler. The filtered normalized R, G, B values are preprocessed for the input of hierarchical clustering. The hierarchical clustering method is then applied using the Ward method. The Ward method is used when the clusters generated have a minimum within-cluster variance. Clustering uses variance analysis instead of the distance metrics approach, as in K-Means. The process is based on the error sum of squares, the total Squared Euclidean Distance between points and clusters means from a given cluster (Dexlab,2018).

  From the Ward method, a linkage is created for the clusters. The results can then be plotted into a Dendrogram. A Dendrogram is used in hierarchical clustering to represent the data. Dendrograms are a type of tree diagram showing the clusters and their relationships. The branches or clades show how similar or dissimilar the clusters are; the higher the height of the branch, the more dissimilar they are (Stephanie,2016). From the Dendrogram, we can form an idea of the clusters and create flat clusters from the hierarchical clustering specified by linkage. Lastly, the visualization of the grouped hierarchical clusters is shown, and can be observed which are the dominant RGB values.

A white background with green and orange lines

Description automatically generated

#### Figure 10 : Dendrogram

*A graph of a box

Description automatically generated*

*Figure 11: K-Means Clusters*

# Results

After clustering testing is put into action, we can start evaluating the output of the implemented data. At the start of our exploration phase, we came to see the total amount of each color classification. This gave us an opportunity to formulate an idea of what the final results could be. 341 dark colors were calculated for dark colors, 258 for light colors, 241 for other, and 18 for brown/beige. This infers that the most dominant colors could be a light or dark. This gives us a start to build a hypothesis of which colors the clustering testing will predict. In our clustering phases, K-Means is first applied. To use K-Means, one must determine the number of clusters. For this reason, the Elbow method is used. In the results of the Elbow method, the length of K equaled 10, the length of inertia was 10, and inertia values of 2064.0, 312.15, 151.80, 95.3, 66.7, 56.7, 49.3, 36.97, and 31.7. The length of K demonstrates that it tried to test ten different values within a range of 1 to 10 clusters every run. The value of inertia was also ten; this means that for every ten values of K, inertia was measured. The inertia values demonstrated allow us to interpret that for every K value inserted, the total amount of squared distance across each data point, and the centroid of its chosen cluster represents these values. The benefit of this outcome is that the lower the inertia, the more compact the clusters, which means well-fitted clusters for the data.

*A graph of color clusters

Description automatically generated*Starting with the first value of inertia 2064.0, it can be seen how it starts high but continues to go down in each inertia. When it hits inertia 151.80, after that, the inertia does not go down significantly. This shows that the number of clusters should be between three or four. For this time, the value of K will be three, and K-Means can be done. Three clusters will be applied to the normalized RGB values data. The output of the normalized clusters centroids would be [0.69080734, 0.50401471, 0.37019529], [1.06119281, 1.28028494,1.38171076], and [-1.04710791, -1.03013431, -0.98876532]. To make more sense of the clusters, we convert them back to the RGB values to identify the colors. The original cluster

*Figure 12: Hierarchical Clusters*

output would be [214.00458716,173.43577982, 151.67889908],[253.11111111,252.58641975,252.14197531] ,[30.50974026, 17.00974026, 16.70779221] which can be interpreted as [214, 173, 151] , [253, 253, 252] , [30, 17, 17] .This gives us the result of our analysis using the dataset created by the Vogue Runway Ready-to-wear collections.

     Proceeding into the second testing phase, the hierarchical clustering. With hierarchical clustering, we form a dendrogram using the ward method to get the clusters. The Dendrogram is visualized, and a cutoff point of 10 is determined. For example, in K-Means, 10 clusters are returned regardless of how defined they are. All of the points that form a cluster are separated from each other by a dissimilarity of not greater than ten. The clusters are then visualized, and the dominant colors are shown. Normalized values would be [-1.197213, -1.139369, -1.084678],[-0.490350,0.617594,0622616],[1.077845,1.303956,1.401537],[0.702929,0.525415,0.401754]. To better understand, we convert back to the original RGB values, which are [14,5,7], [89,59,53], [254,255,254], [215,175,154].

    Having now learned what the most dominant RGB values are and what they mean. The RGB colors can be visualized using an RGB to HEX color converter. This process may also be done in Jupyter, but it was done with an online RBG to HEX converter for visualization purposes. Starting with the K-Means results, RGB value [214, 173, 151] returned a light beige color called Cameo, [253, 253, 252] returned a white shade color called Pampas, and [30, 17, 17] returned a dark brown color called Gondola. For the Hierarchical clustering results, RGB value [14,5,7] returned a dark brown color called Asphalt, [89,59,53] returned a light brown color called Congo Brown, [254,255,254] returned a white color called Sugar cane, and [215,175,154] returned a light beige color called Cameo. All of these colors can allow us to predict what is the next color trend from the Vogue Runway Fall 2024 Ready-To-Wear collections. Since this is an unsupervised machine learning analysis the prediction is not perfect, but it is an approach to what could most probably be the next Fall color trend. Given said that the dominant colors shades are beige, dark brown, and white.

**

*Figure 13: Dominant Color Shades*

# Conclusion

In this analysis, the aim was to collect data from online runway shows from Summer of 2024 and make a reasonable estimate of what could be the following Fall 2024 color trends. The study was done via Vogue Runway Fall Ready-to-Wear with several designers or companies. Furthermore, the aim was to use machine learning techniques to help the fashion industry make correct predictions. We achieved this using K-Means and Hierarchical clustering using machine learning techniques that are used more frequently on unsupervised datasets. Lastly, to evaluate the results given by these machine learning techniques with the results of the K-Means and Hierarchical clustering, we achieved the most proximate conclusion to what our study was all about, finding the color trend of Fall 2024. This being done, the results were as proximate as possible with tools used and time frame. The results provided three color possibilities that will likely be seen in color trends 2024.

    This study was conducted not only to advance predictive analysis in fashion but also to demonstrate the potential of machine learning in a wide range of industries. Companies like WGSN, Gartner, and many others focus on trend analysis in diverse sectors such as travel, lifestyle, beauty, and food. The insights from this type of work can empower millions of companies to better understand and respond to global trends, fostering a sense of hope and excitement about possibilities that lie ahead.

     The results we saw were white, beige, and dark brown. If the various designer’s images are checked individually, the dominance of the predicted colors can be seen quite often. Machine learning algorithms in fashion have a long way to go, and it is a field perhaps not many think of in the engineering or data science world, but it shows how much of the world is changing in terms of technology. This type of work of analyzing color trends in the fashion world is usually done manually by experts studying the runway shows individually. With the help of technology, experts can make more precise predictions along with knowledge of the runway. This can and is changing the fashion industry in terms of trends analysis and can help companies make more revenue if done correctly. Once again, there is still much to study and there are different methods of approach to exist. Still, it is a study that moves color trend analysis with machine learning a step further and can move the awareness of machine learning to another sector that is not just scientifically based.

# Future work

The study created is just a small portion of what machine learning can do within the fashion industry. This analysis focused on a specific magazine to see an outcome from a season collection. Data analysis has the potential go much deeper into analyzing fashion color trends. It could even analyze specific trends, for example, searching words in trend in social media like "quiet luxury," scraping what images come up, and making an analysis of what fashion trends are involved in this phrase. Social media can play a large part in fashion trend analysis as it can be used to scrape words and images. It would come up with color trends and calculate if a trend is rising or falling in popularity. As well as doing large quantities of research from different magazines instead of just a certain number of designers from a magazine.

Trend analysis can go beyond just colors. Instead of just using images, a combination of images and specific words can be scraped. The analysis built could be about colors, fashion terms, prints, and percentages. In addition to fashion magazines, fashion websites, and social media posts, websites such as Google Trends and Pinterest Trends can see the percentages of how an item or term is being used.

If we were to continue this research, the data collection could be scaled up significantly. By scraping data from a wider range of magazines, we could make predictions on a much larger scale. More data equals more potential for more results. While our current study is a significant step forward, there's still so much more we can explore in the realm of color trends. This is  a promising opportunity to delve deeper into the fashion color trends of specific collections, and the potential for further research is truly exciting.

##### References

Al-Halah, Z.A.-H., Stiefelhagen, R. and Grauman, K. (n.d.). Fashion Forward: Forecasting Visual Style in Fashion. [online] Available at: <https://openaccess.thecvf.com/content_ICCV_2017/papers/Al-Halah_Fashion_Forward_Forecasting_ICCV_2017_paper.pdf>.

Bertola, P. and Teunissen, J. eds., (n.d.). Fashion 4.0. Innovating Fashion Industry Through Digital Transformation . [online] Available at: <https://ualresearchonline.arts.ac.uk/id/eprint/13190/7/Fashion%204.0%20final%20submission%5b1%5d.pdf>.

Cox, H. and Mowatt, S. (2012). Voguein Britain: Authenticity and the creation of competitive advantage in the UK magazine industry. *Business History*, 54(1), pp.67–87. doi:https://doi.org/10.1080/00076791.2011.61720.

Cui, M. (2020). Introduction to the K-Means Clustering Algorithm Based on the Elbow Method. [online] doi:https://doi.org/10.23977/accaf.2020.010102.

Dexlab (2018). *Rudiments of Hierarchical Clustering: Ward’s Method and Divisive Clustering*. [online] DexLab Analytics | Big Data Hadoop SAS R Analytics Predictive Modeling & Excel VBA. Available at: https://m.dexlabanalytics.com/blog/rudiments-of-hierarchical-clustering-wards-method-and-divisive-clustering [Accessed 16 Aug. 2024].

docs.python.org. (n.d.). *ast — Abstract Syntax Trees — Python 3.8.3 documentation*. [online] Available at: <https://docs.python.org/3/library/ast.html>.

docs.python.org. (n.d.). *io — Core tools for working with streams — Python 3.8.3rc1 documentation*. [online] Available at: https://docs.python.org/3/library/io.html.35. [Python's Counter: The Pythonic Way to Count Objects – Real Python](https://realpython.com/python-counter/)

docs.python.org. (n.d.). *io — Core tools for working with streams — Python 3.8.3rc1 documentation*. [online] Available at: https://docs.python.org/3/library/io.html.

docs.scipy.org. (n.d.). *Hierarchical clustering (scipy.cluster.hierarchy) — SciPy v1.9.1 Manual*. [online] Available at: <https://docs.scipy.org/doc/scipy/reference/cluster.hierarchy.html>.

Fernando Gutierrez Portela, Florina Almenares Mendoza and Liliana Calderon Benavides (2019). Evaluation of the performance of supervised and unsupervised Machine learning techniques for intrusion detection. doi:https://doi.org/10.1109/icasat48251.2019.9069538.

‌Fradkov, A.L. (2020). Early History of Machine Learning. *IFAC-PapersOnLine*, 53(2), pp.1385–1390. doi:https://doi.org/10.1016/j.ifacol.2020.12.1888.

Gupta, A. (2019). *Elbow Method for optimal value of k in KMeans*. [online] GeeksforGeeks. Available at: <https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/>.

Han, A., Kim, J. and Ahn, J. (2021). Color Trend Analysis using Machine Learning with Fashion Collection Images. *Clothing and Textiles Research Journal*, 40(4), p.0887302X2199594. doi:https://doi.org/10.1177/0887302x21995948.

Jo, J., Lee, S., Lee, C., Lee, D. and Lim, H. (2020). Development of Fashion Product Retrieval and Recommendations Model Based on Deep Learning. *Electronics*, [online] 9(3), p.508. doi:https://doi.org/10.3390/electronics9030508.

Kanal, L.N. (2003). Perceptron. *Research Gate*. [online] Available at: <https://www.researchgate.net/publication/262288029_Perceptron/citations#fullTextFileContent>.

Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C., Silverman, R. and Wu, A.Y. (2000). The analysis of a simple k-means clustering algorithm. *Proceedings of the sixteenth annual symposium on Computational geometry - SCG ’00*. doi:https://doi.org/10.1145/336154.336189.

Kennedy, J.K. (2024). WGSN Launches AI-Driven Platform for Buyers. *Business of Fashion*. [online] Available at: <https://www.businessoffashion.com/news/retail/wgsn-launches-ai-driven-platform-for-buyers/>.

‌ Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. and Mullainathan, S. (2017). Human Decisions and Machine Predictions\*. *The Quarterly Journal of Economics*, [online] 133(1). doi:https://doi.org/10.1093/qje/qjx032.

Lai, Y. (2019). A Comparison of Traditional Machine Learning and Deep Learning in Image Recognition. *Journal of Physics: Conference Series*, 1314, p.012148. doi:https://doi.org/10.1088/1742-6596/1314/1/012148.

Laird Borrelli-Persson (2017). *Vogue Facts About Covers, Editors, Models, Stars*. [online] Vogue. Available at: <https://www.vogue.com/article/vogue-covers-models-facts-history>.

Loeb, W. (n.d.). *Who Are The World’s Top Fashion Designer’s And What Are Their Top Products*. [online] Forbes. Available at: <https://www.forbes.com/sites/walterloeb/2022/10/27/who-are-the-worlds-top-fashion-designers-and-what-are-their-top-products/>.

matplotlib.org. (n.d.). *matplotlib.pyplot — Matplotlib 3.5.2 documentation*. [online] Available at:https://matplotlib.org/stable/api/pyplot\_summary.html.

Nast, C. (2023). *Vogue Business Index*. [online] Vogue Business. Available at: https://www.voguebusiness.com/advanced/index#:~:text=Louis%20Vuitton%20leads%20and%20Saint%20Laurent%20returns%20to [Accessed 16 Aug. 2024].

Nast, C. (n.d.). *Fall 2024 Ready-to-Wear Fashion shows*. [online] Vogue. Available at: <https://www.vogue.com/fashion-shows/fall-2024-ready-to-wear>.

Nichols, K. (2023). *The 2023 Luxury Report: The Brands, Items, and Trends Defining Designer Fashion*. [online] Who What Wear. Available at: <https://www.whowhatwear.com/luxury-report-2023>.

Nielsen, F. (2016). Hierarchical Clustering. *Introduction to HPC with MPI for Data Science*, pp.195–211. doi:https://doi.org/10.1007/978-3-319-21903-5\_8.

Pandas (n.d.). *pandas - Python Data Analysis Library*. [online] pandas.pydata.org. Available at: <https://pandas.pydata.org/about/>.

Patel, A.A. (2018). *Hands-On Unsupervised Learning Using Python : How to Build Applied Machine Learning Solutions from Unlabeled Data*. Sebastopol: O’Reilly Media, Incorporated.

published, S.H. (2024). *10 Trends That Defined the Fall 2024 Runways*. [online] Marie Claire Magazine. Available at: <https://www.marieclaire.com/fashion/fall-2024-trends/>.

Python, R. (n.d.). *Image Processing With the Python Pillow Library – Real Python*. [online] realpython.com. Available at: https://realpython.com/image-processing-with-the-python-pillow-library/.

Reitz, K. (2023). *requests: Python HTTP for Humans.* [online] PyPI. Available at: <https://pypi.org/project/requests/>.

Scikit-Learn (2019). *sklearn.preprocessing.StandardScaler — scikit-learn 0.21.2 documentation*. [online] Scikit-learn.org. Available at: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>.

Scikit-learn.org. (2019). *1.13. Feature selection — scikit-learn 0.21.3 documentation*. [online] Available at: <https://scikit-learn.org/stable/modules/feature_selection.html>.

seaborn (2012). *seaborn: statistical data visualization — seaborn 0.9.0 documentation*. [online] Pydata.org. Available at: <https://seaborn.pydata.org/>.

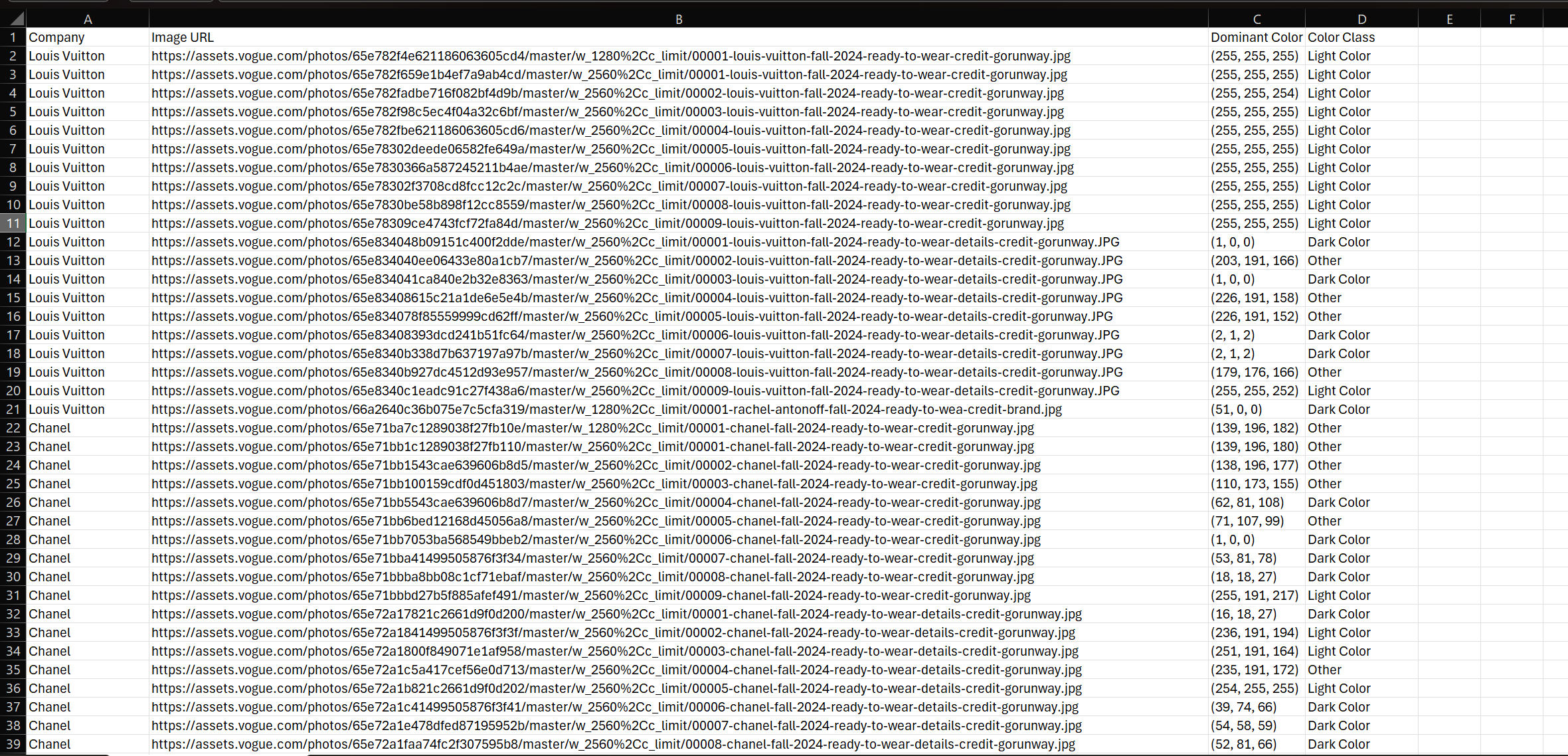
Stephanie (2016). *Hierarchical Clustering / Dendrogram: Simple Definition, Examples*. [online] Statistics How To. Available at: <https://www.statisticshowto.com/hierarchical-clustering/>.

‌W3schools.com. (2019). *Colors RGB*. [online] Available at: https://www.w3schools.com/colors/colors\_rgb.asp.

WWD. (2024). *2024 Fall Ready-to-Wear – WWD*. [online] Available at: https://wwd.com/runway/fall-2024/ [Accessed 16 Aug. 2024].

## Appendix A

*DATASET VISUAL*

**

*Figure 5: Fashion\_Show\_Colors\_Filtered.csv*

## Appendix B

*A graph of a number of different colored squares

Description automatically generated with medium confidence*

*DATA EXPLORATION VISUALS*

*Figure 14 : Frequency of Dominant Dark Colors in Runway Show Images*

A bar graph with different colored lines

Description automatically generated

*Figure 15: Frequency of Dominant Brown/Beige Colors in Runway Show Images*

A white background with black and red text

Description automatically generated

*Figure 16 : Frequency of Dominant Light Colors in Runway Show Images*

A close-up of a chart

Description automatically generated

*Figure 8: Heatmap of Color Class Counts by Company*

A graph of different colored bars

Description automatically generated

*Figure 6: Proportion of Each Color Class by Company*

A graph of different colored lines

Description automatically generated

*Figure 7: Count of Each Color Class by Company*

A graph of different colors

Description automatically generated

*Figure 17: Countplot of Color Class*

## Appendix C

MACHINE LEARNING VISUALS

A graph of a person with a blue line

Description automatically generated

*Figure 9: Elbow method*

A table with black text

Description automatically generated

*Figure 18: K-Means Dataframe of added normalized RGB values*

A close-up of numbers

Description automatically generated

*Figure 19: K-Means Normalized and Original Cluster Centroids*

A graph of a scatter plot

Description automatically generated

Figure 11: K-Means Clusters

A long shot of a white background

Description automatically generated

#### Figure 10 : Dendrogram

A table with numbers and symbols

Description automatically generated

*Figure 20: Hierarchical Dataframe of added normalized RGB values*

A white sheet with black text

Description automatically generated with medium confidence

*Figure 21: Hierarchical Normalized and Original Cluster Centroids*

A diagram of a graph

Description automatically generated with medium confidence

*Figure 12: Hierarchical Clusters*

## Appendix D

RBG TO HEX VISUALS

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

*Figure 22: RBG to HEX Hierarchical Results*

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

*Figure 23: RBG to HEX K-Means Results*