Contents

[SECTION ONE 2](#_Toc124117318)

[Introduction 2](#_Toc124117319)

[APPLICATION OF BUSINESS INTELLIGENCE AND BIGDATA ANALYTICS STACK 4](#_Toc124117320)

[DATA COLLECTION/EXPLANATION 4](#_Toc124117321)

[DATA PREPROCESSING 5](#_Toc124117322)

[DATA CLEANING 5](#_Toc124117323)

[RATINGCOUNT AND AVG\_RATING 10](#_Toc124117324)

[COLOUR AND BRAND 10](#_Toc124117325)

[P\_ATTRIBUTES 10](#_Toc124117326)

[DATA INTEGRATION 13](#_Toc124117327)

[DATA EXPLORATION 14](#_Toc124117328)

[HOW OCCASION CONTRIBUTED TO REVENUE 15](#_Toc124117329)

[**CUSTOMER’S PREFERENCE BASED ON COLOUR** 19](#_Toc124117330)

[CUSTOMER BEHAVIOUR BASE ON CULTURE. 20](#_Toc124117331)

[DATA ANALYSIS 23](#_Toc124117332)

[MACHINE LEARNING 25](#_Toc124117333)

[DATA REDUCTION 26](#_Toc124117334)

[DATA MINNING AND PATTERN EVALUATION 27](#_Toc124117335)

[RANDOM FOREST 30](#_Toc124117336)

[KNOWLEDGE PRESENTATION 32](#_Toc124117337)

[VISUALIZATION 36](#_Toc124117338)

[SECTION TWO 37](#_Toc124117339)

[Current Issues in the Use of Big Data Analytics in The Fashion Retail Industry. 37](#_Toc124117340)

[Introduction 37](#_Toc124117341)

[What Are the Identified Issues? 37](#_Toc124117342)

[Effective Solution Using Various Technologies 39](#_Toc124117343)

[Conclusion 41](#_Toc124117344)

[References 42](#_Toc124117345)

# 

# **SECTION ONE**

# **Introduction**

Business intelligence is essential to the long-term success of any firm in today's period of rapid technological growth. Business intelligence (BI) is a term used to describe a collection of methods, systems, and concepts that might enhance the accuracy of fact-based business decisions. Raw, irrelevant data is transformed into something usable and informative by the architecture and technologies underlying this process. The future direction of a corporation can be greatly influenced by data like this. It helps with the creation of new strategies, the growth of operational excel insights, and the ability to make wise judgements.(Khan et al, 2019.)

Howard Dresner, a Gartner Group analyst, invented the term "business intelligence" (BI) in the middle of the 1990s to refer to a collection of interconnected theories and methods for enhancing decision-making through data mining, analysis, and presentation. (Azeroual & Theel 2018)

There are continual challenges for the fast fashion industry in keeping up with consumers' requests for new items (Cook and Yurchisin, 2017), hence the industry relies heavily on constant market analysis to quickly understand consumer preferences and change production accordingly.

As stated by (Silva et al., no date), When it comes to design, purchasing, and retailing, the fashion industry has traditionally depended primarily on intuition and imagination. However, in recent years, the industry has also started experimenting with advanced data analytics. (Silva et al., no date).

New Gen Apps (2017) claims that the fashion industry may leverage Big Data for a variety of goals, such as spotting markets, analysing trends, learning about consumers, increasing sales of expensive items, promoting emerging designers, gauging the influence of influential figures, and more. According to research (Silva et al., n.d.),

In this study, I introduce the big data analytics stack, machine learning, and visualisation to enable business intelligence in the fashion industry draw actionable insights from their data. The machine learning component, which adhered to the KDD concept, is split apart from the big data analytics stack, which adhered to the business intelligence methodology.

# **Application Of Business Intelligence And Bigdata Analytics Stack**

Big Data is comprised of a number of different significant technologies, including Hadoop, HDFS, NoSQL, Map Reduce, MongoDB, Cassandra, PIG, HIVE, and HBASE. These technologies work together to achieve a single goal, such as obtaining value from data that was conventionally considered to be obsolete or irrelevant (Zakir et al, 2015)

The following processes will be distinguished in applying business intelligence to the :

1. Data Collection
2. Data Preprocessing
3. Data Integration
4. Data Exploration
5. Data Analysis

# **Data Collection/Explanation**

VARIABLE DESCRIPTIONS: The records are stored in a comma-separated values (CSV) file. Variable names are listed in a header line.

This report's presumption is that the dataset was gathered from a fashion firm's online application or website.

This report makes use of two datasets: a fashion data collection considered to have originated in an OLAP environment, and information about various fashion brands. Detailed descriptions of the column formats can be found below..

|  |  |  |
| --- | --- | --- |
| COLUMN NAME | COLUMN DEFINITION | COLUMN ANNOTATION |
| p\_id | Product Identification | This column consists of a unique number assigned to each fashion item |
| Name | Product Name | This column consists of the name of the fashion item purchased |
| Price | Product Price | The monetary value of the fashion item purchased |
| Colour | Product Colour | This column tells the color of each fashion item purchased |
| Brand | Product Brand | This column specifies the brand name of the fashion item |
| ratingCount | Product Rating | This column shows the ratings count given to the column |
| Avg\_rating | Average Ratings | This column shows the rating average given of each product |
| Description | Product Description | This column gives a broader description of the fashion item |
| P\_attributes | Product Attributes | This column is in JSON format and it contains multiple columns that further give details about the product. |

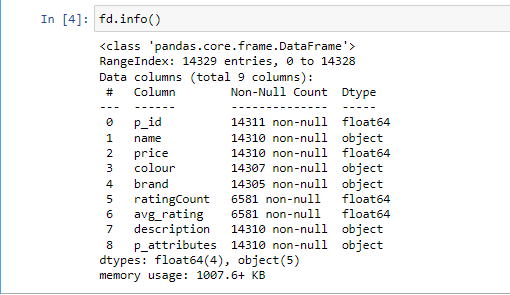
**Table 1**: Detailed Description of Column Names

# DATA PREPROCESSING

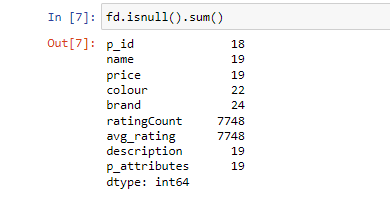
Data preparation is the process of transforming raw data into a more usable and understandable format. Raw data, which is derived from the real world, might suffer from poor presentation, human error, and incompleteness. Preprocessing allows for the correction of these issues and the enhancement of the quality and productivity of datasets. Cleaning, integrating, reducing, and transforming are the four phases of data processing.

# DATA CLEANING

Data cleaning, also known as data cleansing, aims to enhance the quality of data by removing irrelevant data, filling in blanks, and resolving anomalies. One must take into account missing values and noisy values for effective data cleaning.

Prior to cleaning, the fashion dataset (aliased as fd) included 9 columns and 14328 rows, although the number of columns was inconsistent.

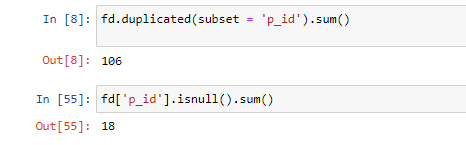
**Figure 1**: Showing info about the columns

There are a number of null values in the column, indicating some kind of inaccuracy.

**Figure 2**: Showing sum of null values in each columns

To properly clean the data, each column is processed independently or all together when necessary to ensure the data is clean.

P\_ID: The p\_id column is treated as a primary key, thus it must contain only unique IDs and never have any nulls. Due to the prevalence of duplicates and nulls in this column, they had to be removed before normalisation could take place.



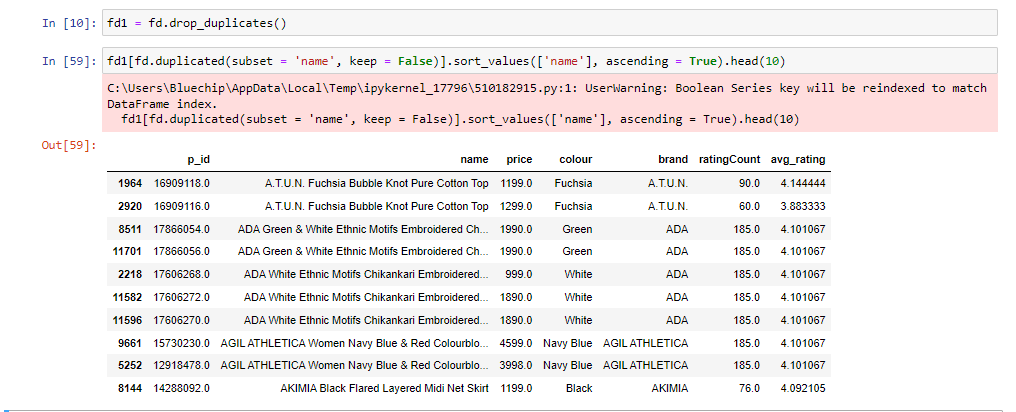
**Figure 3**: Showing sum of duplicated and null rows

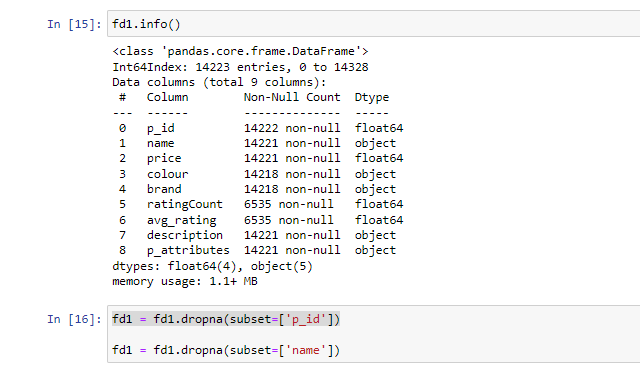
As can be seen in the preceding diagram, there are a total of 106 duplicates and 18 blanks within p id. Duplicate rows can't be removed until the underlying pattern is known.

**Figure 4**: Showing pattern of duplicated columns

In order to see the pattern of the duplication, the code above sorts based on the p id column, since some name columns occurred twice with the same p id, which is the primary key, but duplicated values for other columns. As a result, unnecessary repetition of data is removed.

NAME: In cases when the price is different but the name is the same, perhaps because prices were different at different times, we removed the duplicated value from the full dataset, and this trickled down to normalise the name column.

Assumption: The database was changed to reflect the new price and a new p\_id because it is presumed that the clothing store may adjust its pricing for reasons like sales. **Figure 5**: Showing pattern of dataset after dropping duplicates

The updated dataset details can be seen below after duplicates have been removed, and before the null values of the two columns (p id and name) are removed.

**Figure 6**: Showing information of the Data Setafter dropping duplicates and null

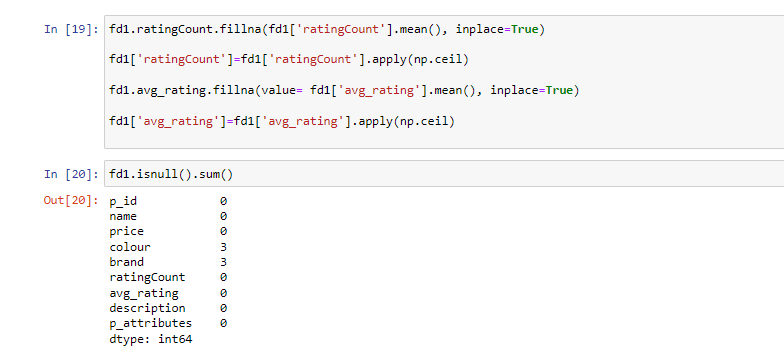
## RATINGCOUNT AND AVG\_RATING

Ratingcount and avg rating both have the same number of missing values and non-missing values, illustrating the common occurrence of missing values in datasets.

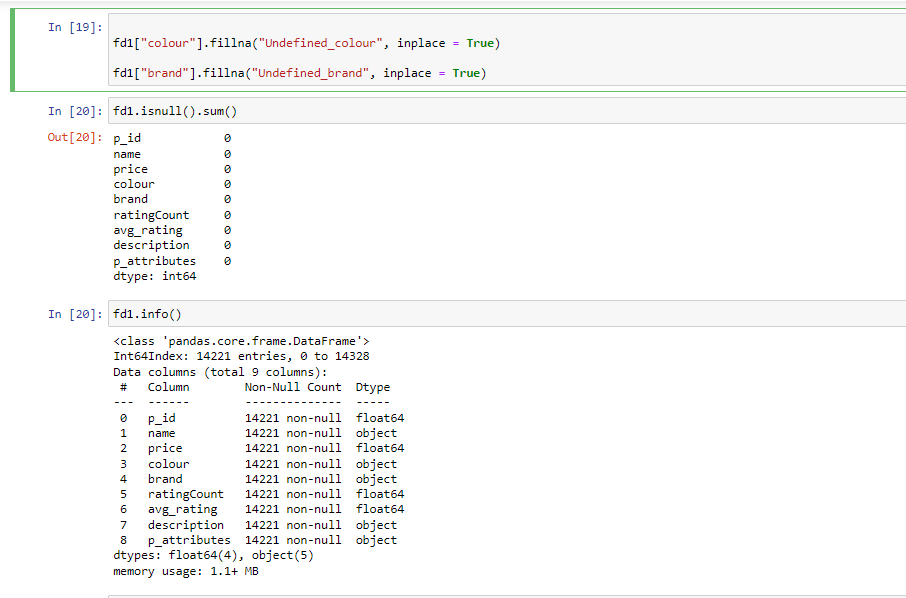
Mostafa 2019 describes imputation as "the process of replacing a missing value with an accurate assessment based on the available information." I decided to use this method to account for any missing values in my data. Both transitive and intransitive forms of this procedure exist. The mean, median, and mode are examples of intransitive methods because their calculation relies only on the target variable itself, rather than on any other variables.

Since the mean displays the average value in the column, it is utilized to represent the possible ratings for the missing variables.

Assumption: Based on the assumption that users would only ever provide ratings in whole numbers or to a single decimal place, the total number of ratings is rounded up using ceil to the closest whole number.

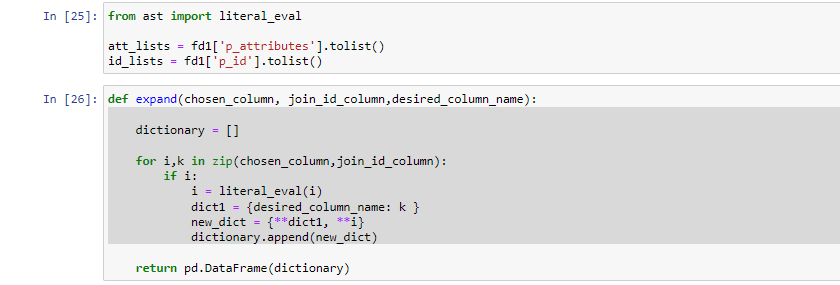


**Figure 7**: Showing how missing values of avg\_rating and ratingCount are filled

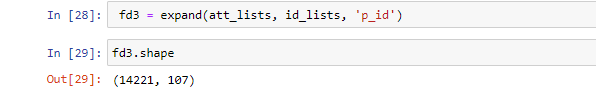
COLOUR AND BRAND: The colour and brand columns, which each have three null values, were each filled with "Undefined\_colour" and "Undefined\_brand" since they include information that is significant to other columns. 

**Figure 8**: Showing how missing values of brand and colour are filled

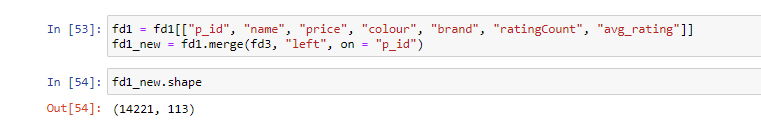
P\_ATTRIBUTES: The p attributes column is a key-value pair stored as a Java Script Object Notation (JSON) string. The literal eval function, included in the ast class, is used to transform this column into a format that can be read and processed by the database. Without needing to interpret the contents, the literal eval method can securely evaluate strings containing Python values from unknown sources.



**Figure 9**: Showing parsing of p\_attributes column

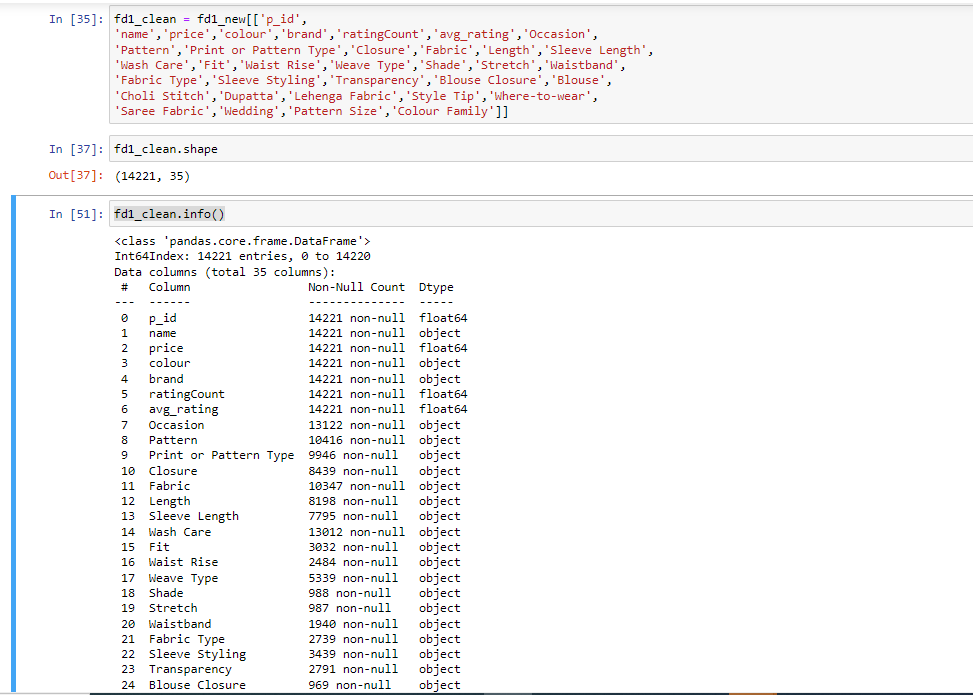
Afterwards, 107 additional columns, including p id, are produced from the original p\_attributes column.

**Figure 10**: Showing expansion of p\_attributes column

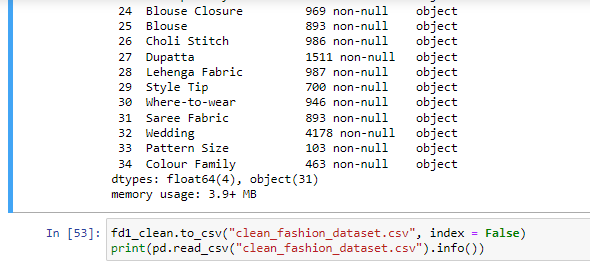
The p\_id, a unique identifier, is used to combine the splitted p attributes column with the full dataset. Following the consolidation, there were a total of 113 columns and 14221 rows.

**Figure 11**: Merging of the splitted column with the entire dataset

By splitting the data, new insights were unveiled. A few of the columns in the fashion dataset reveal its origin as data from an Indian fashion website (Saree Fabric, Lehenga Fabric, Dupatta, and Choli Stitch).



**Figure 12**: Showing selected columns needed columns for the report



**Figure 13**: Showing selected columns needed for the report

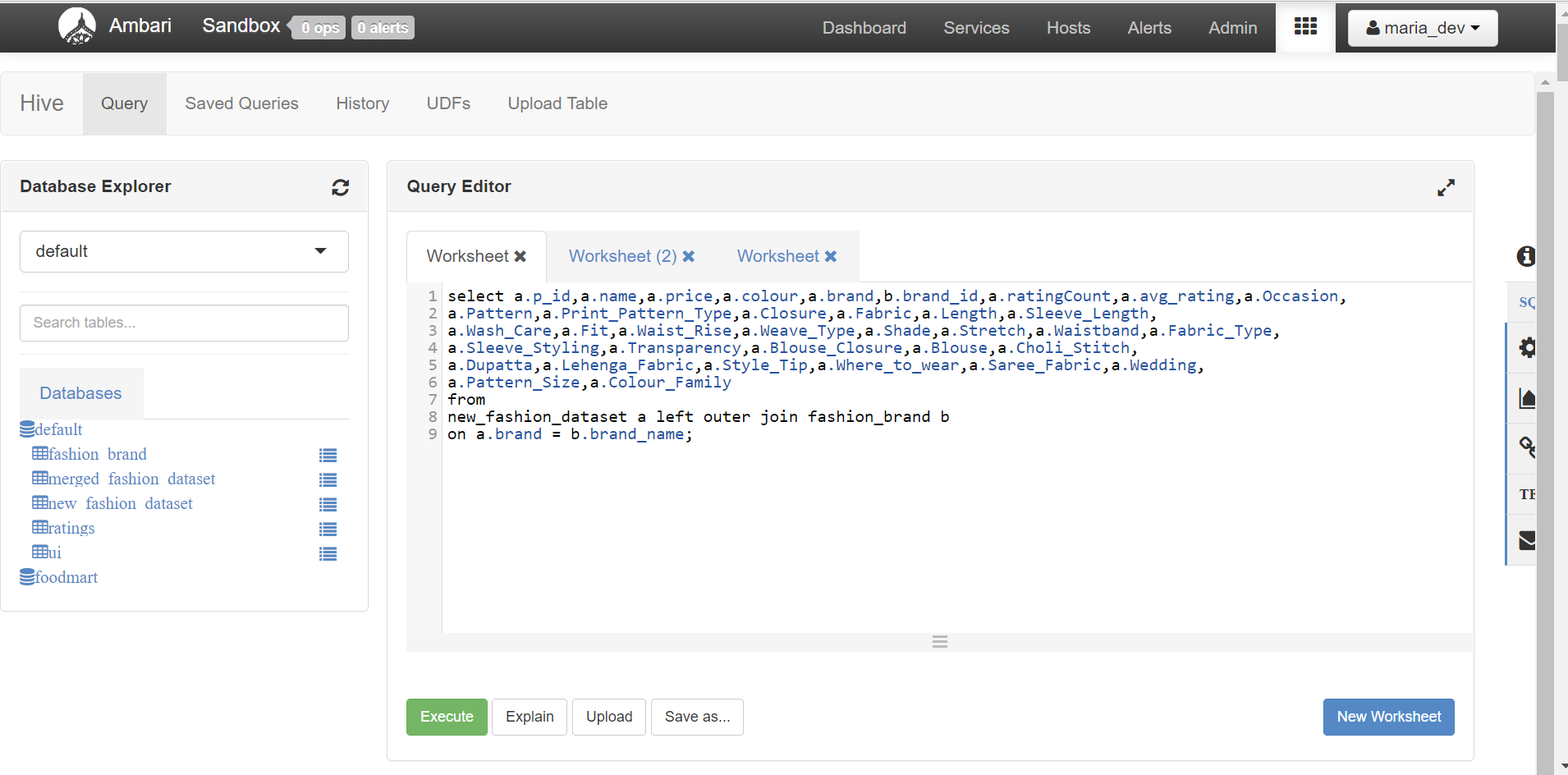
The 113 columns were cut to 35 and read into a CSV for further analysis because they contained a huge amount of useless data (known as "noise").

# DATA INTEGRATION

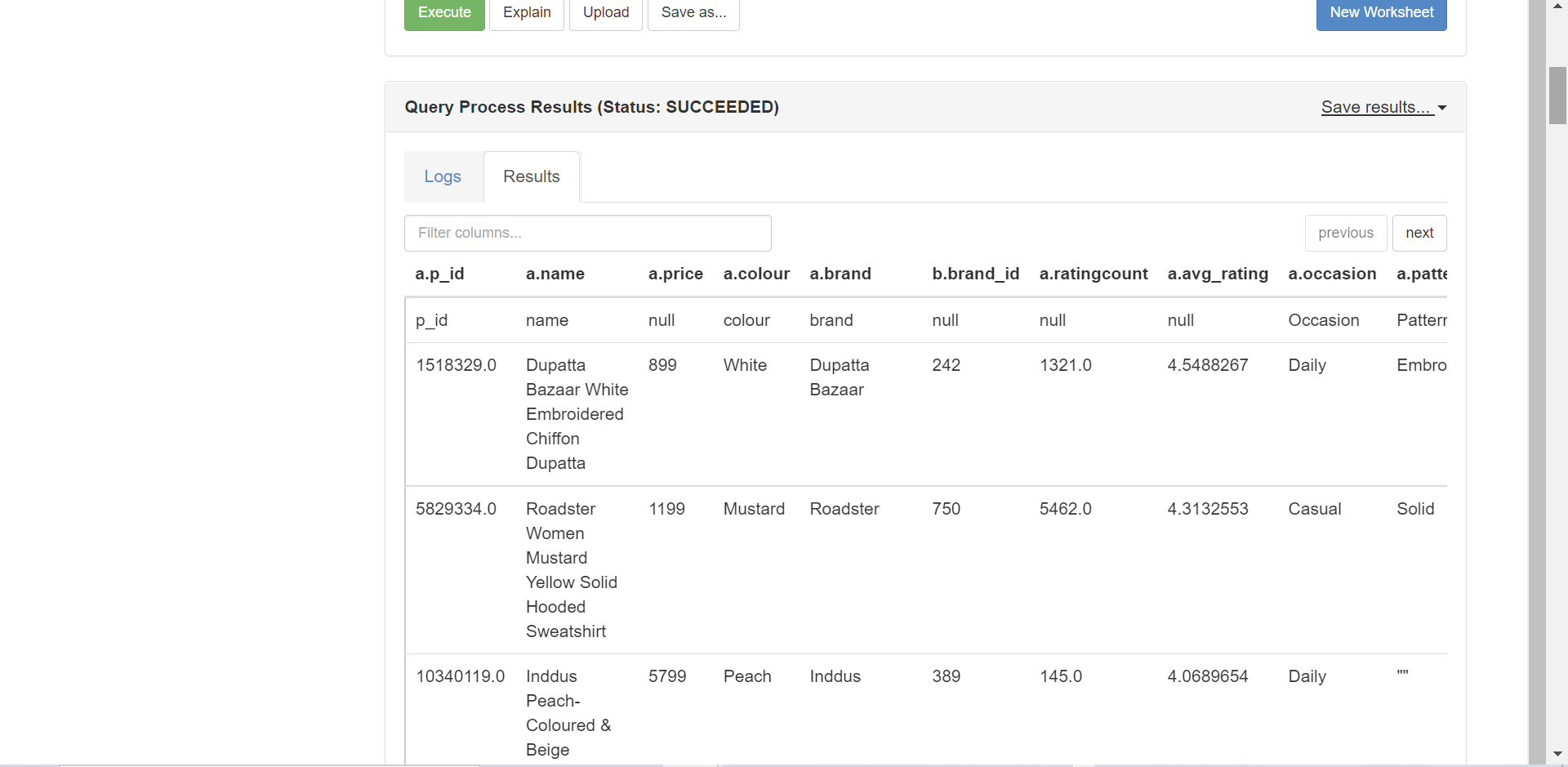
In data processing, data integration refers to the act of bringing together disparate data sets from many sources into a unified database. The Extract, Transform, and Load (ETL) procedure is a portion of data integration that ensures the transmission of all required data from various source systems once it has been processed and consolidated. Information is collected and kept in one convenient location. Having all information in one location is a productivity boon as it improves workflow and output.

The two dataset fashion\_brand and new\_fashion\_dataset which was the pre-processed data is integrated into a new table using hive.

The new\_fashion\_dataset which is identified as the transactional dataset is left outer join with the fashion\_brand which is identified as the dimensional table and follows the first normal form (1NF) of normalization.

As brand name is shared by both datasets, we can use a left outer join to combine the two into a single table that contains both the transactional and dimensional data for easier exploration.

**Figure 14**: Showing Integration of the two datasets



**Figure 15**: Result of the integrated dataset

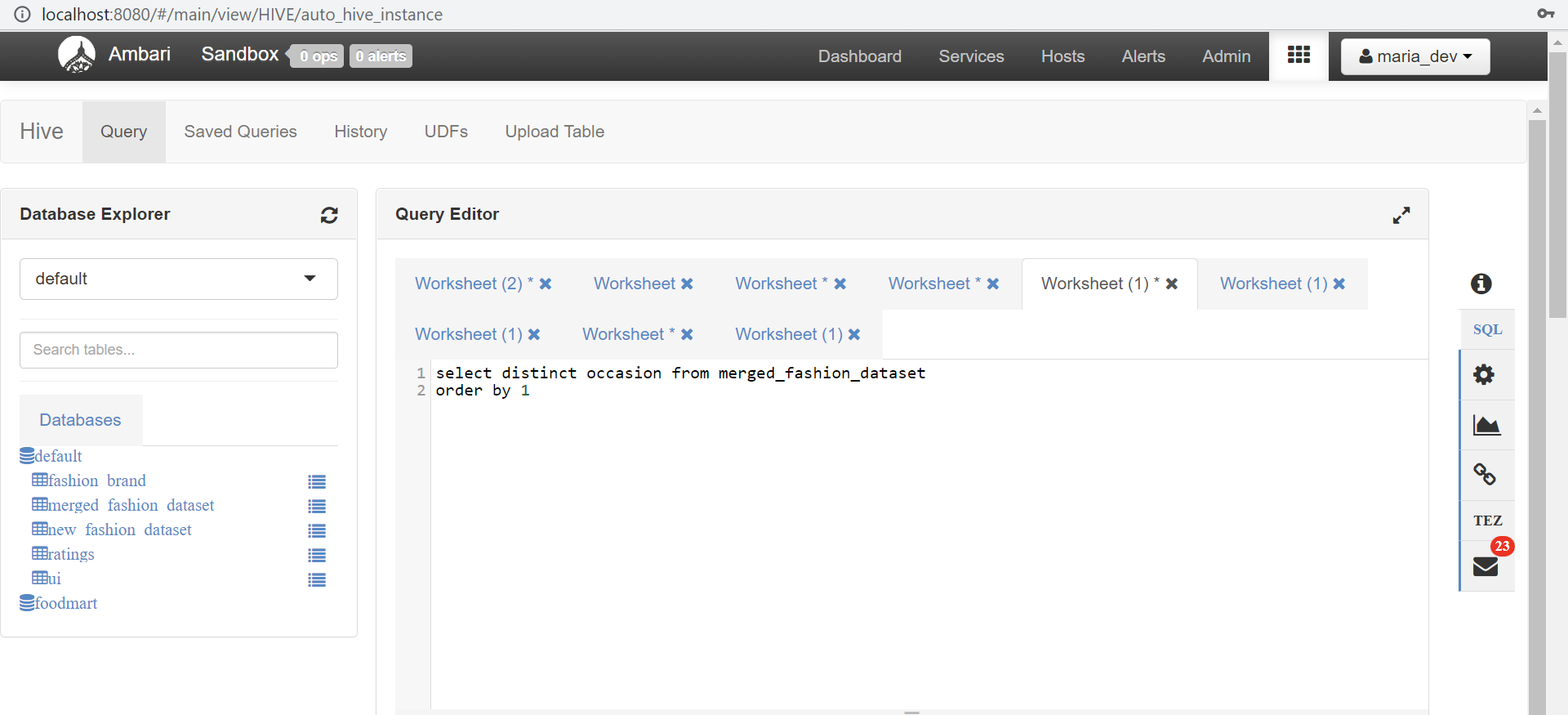
# DATA EXPLORATION

Exploratory Data Analysis (EDA), a method of descriptive data analytics, was used to sift through the information in this dataset. EDA is the process by which one can learn what occurred from a set of available historical records.

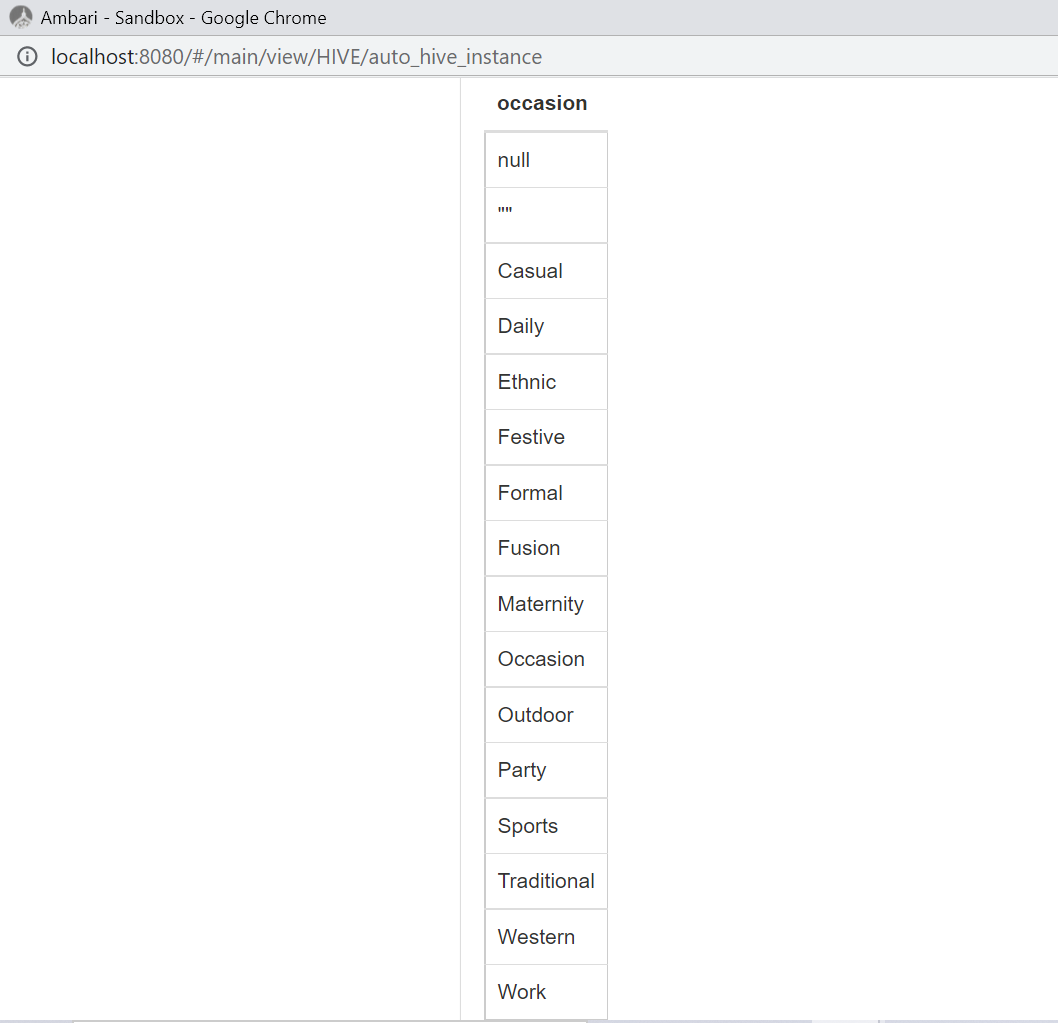
Yafooz et al. 2020 claim that big data analytics can examine historical data to produce future forecasts. As a result, firms may both improve their decisions for the now and plan for the future.

By analysing the fashion company's data, I was able to gain insight into the factors that contribute to their revenue, their customers' preferences, and their behaviour.

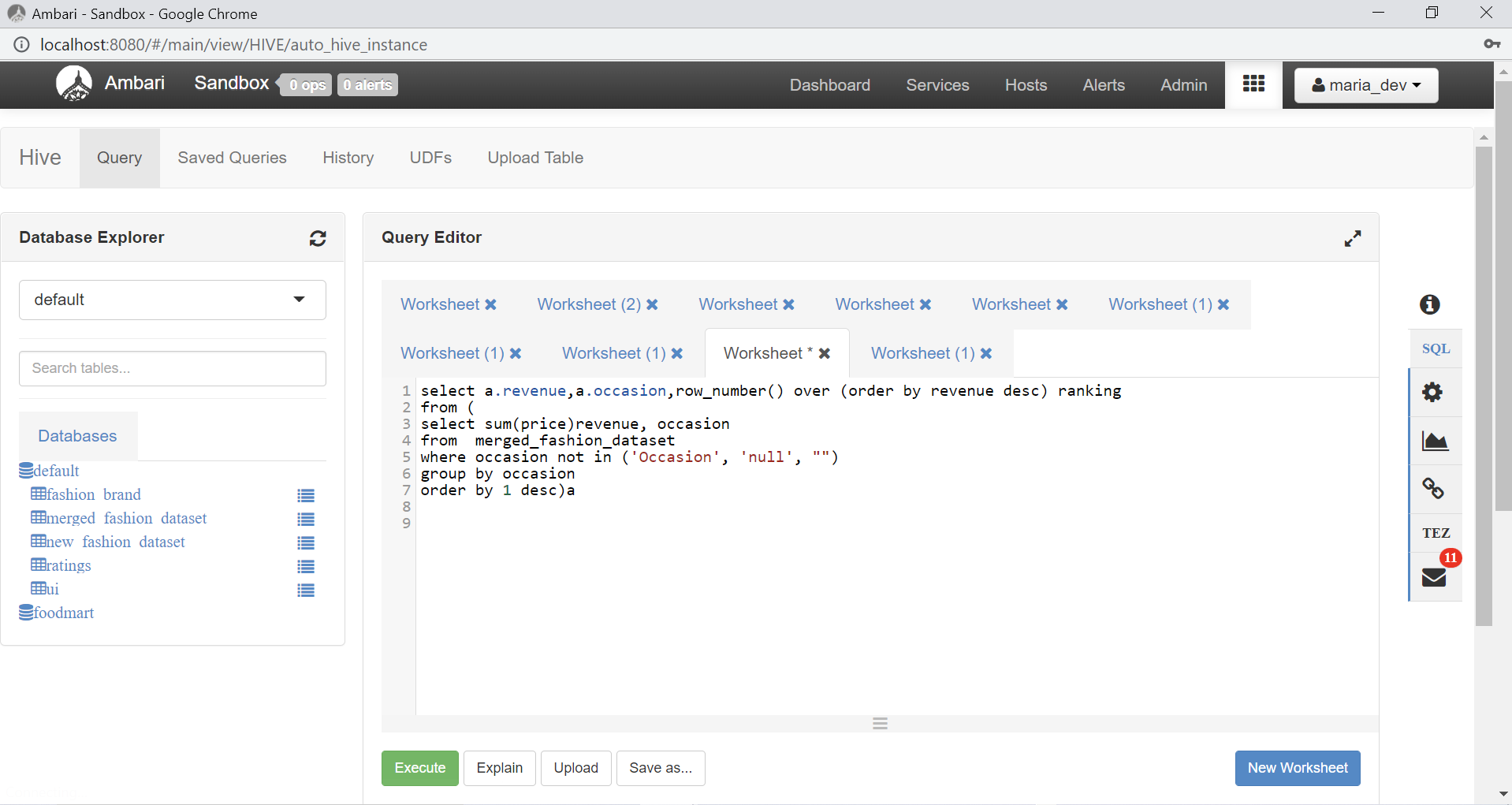
# HOW OCCASION CONTRIBUTED TO REVENUE

I explored the various events for which people purchase fabric and found that there are 12 distinct events, as shown in the table below.

**Figure 16**: code showing distinct occasion

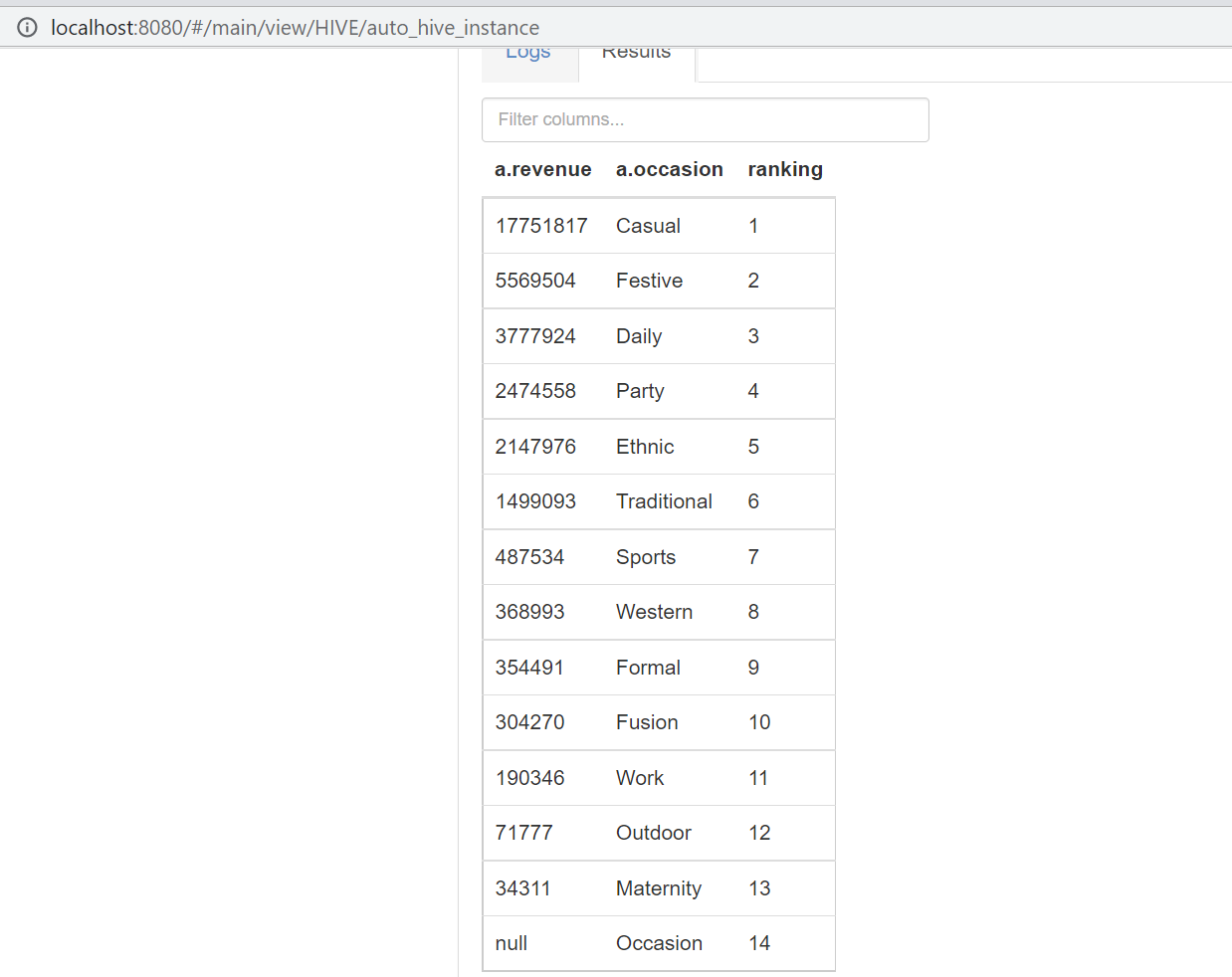


**Figure 17**: Showing listed occasion

In addition, I looked into the event that brought in the most revenue by adding up the prices and ranking them from 1 to 13 from highest to lowest. The code does not include the irrelevant words "occasion", "", or "".

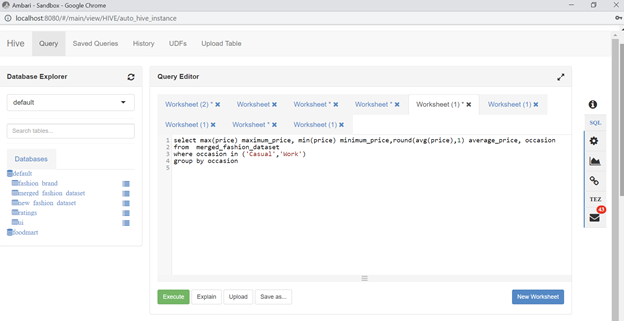
**Figure 18**: code for ranking revenue generated based on occasion

Result



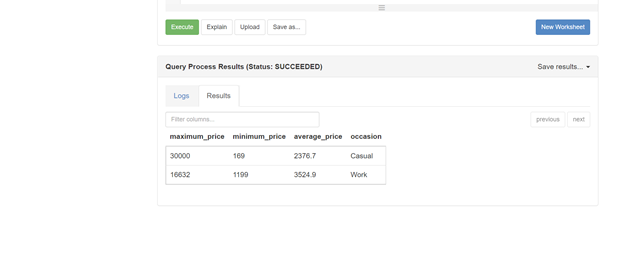
**Figure 19**: Showing result of the ranked occasion based on revenue

This analysis led me to conclude that casual wear was the most financially significant category for fashion retailers, while maternity wear ranked last (13th). Because only pregnant women shop for maternity clothes, the market for these items is lower.

I compared the cheapest, most expensive, and average prices of casual and work clotings for the sake of comparison.

**Figure 20**: Showing code for comparing casual and work ocassion

Result:



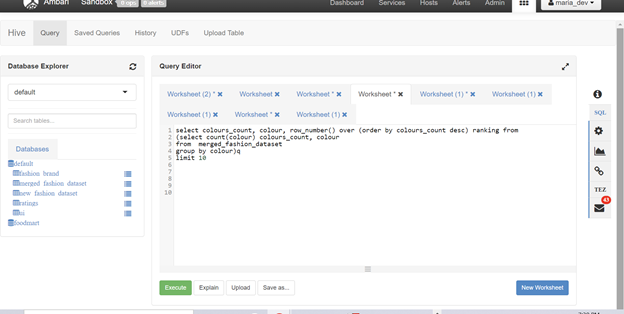
**Figure 21**: Showing result of comparing casual and work occasion

Moreover, the analysis reveals why people prefer to buy casual clothes rather than work clothes: the former's minimum price is more than 100 percent lower.

# **CUSTOMER’S PREFERENCE BASED ON COLOUR**

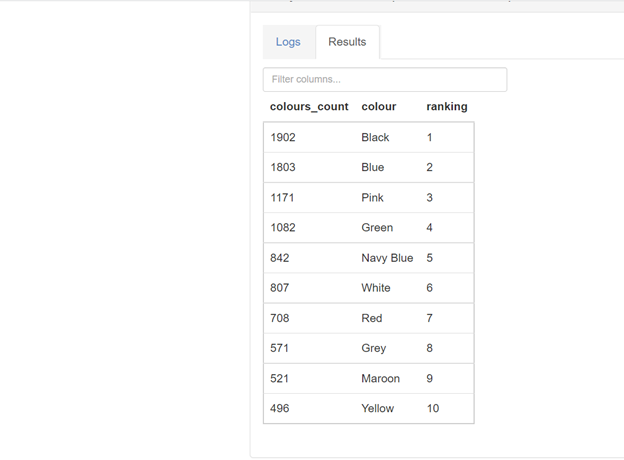
I looked at the top 10 colours purchased by customers to get a sense of what they like, and it turns out that black is the most popular colour while yellow is the least.

It's safe to say that everyone has at least one black clothing item in their closet.



**Figure 22**: Showing code for selecting top 10 colours

Result:



**Figure 23**: Showing Data Cleaning of the Data Set

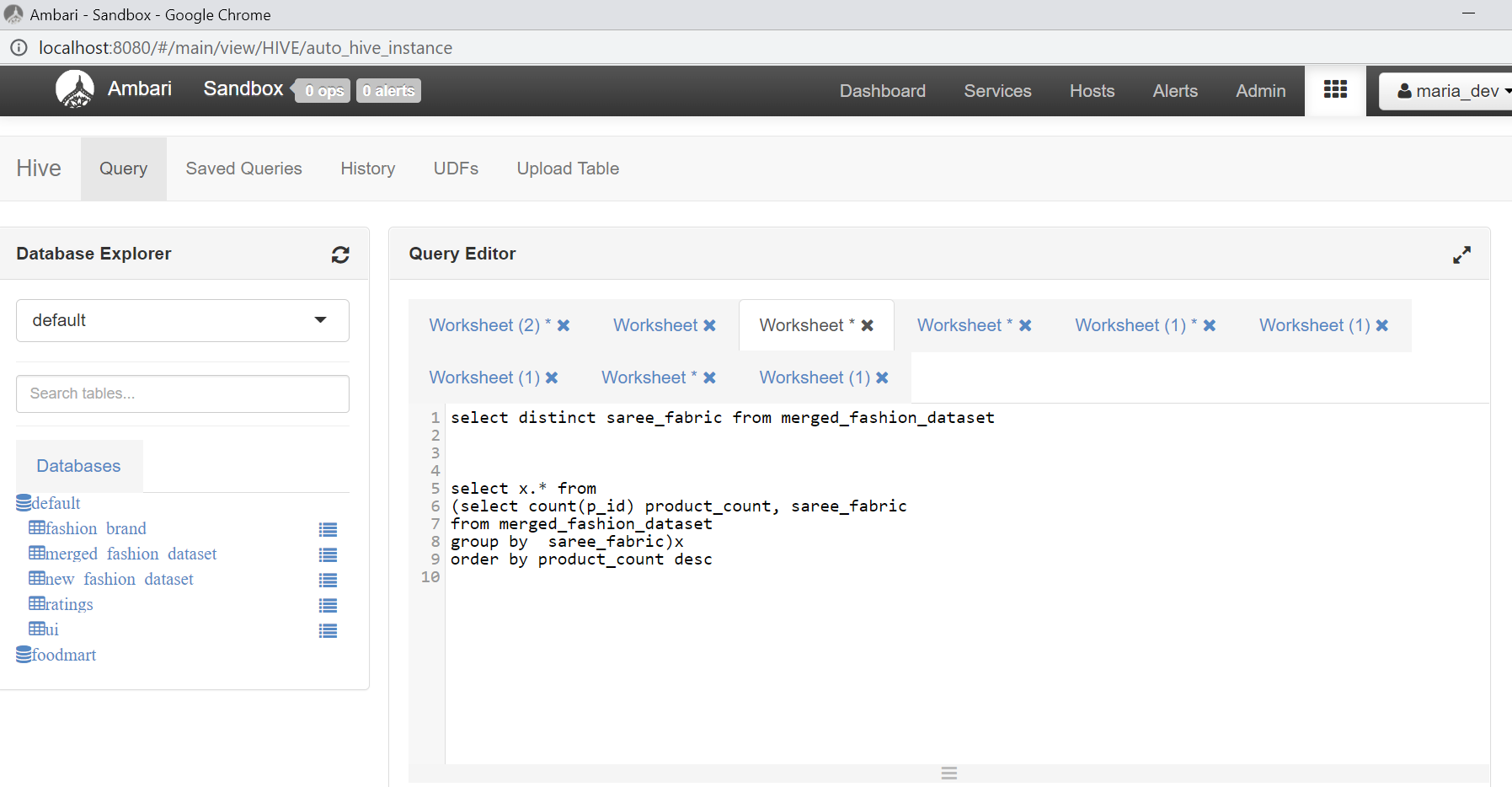
# CUSTOMER BEHAVIOUR BASE ON CULTURE.

A saree represents a deity in Indian culture and belief, and when worn, a woman is seen as more powerful and confident. When an Indian woman wears her traditional saree, she is reminded of the importance of honouring and preserving her country's rich cultural heritage.

Beautiful interpretations of India's vibrant culture, vivid colours, and exotic hand crafted techniques of embroidery and weaving have been produced by international designers over the past few decades. The saree is one of the oldest and possibly the only surviving unstitched garment from the past. The saree is at the heart of culture and heritage. A select number of Indian designers understand the significance of this drape and are working to revive and modernize the saree for their upcoming seasons' collections (Kaur et al., 2019)

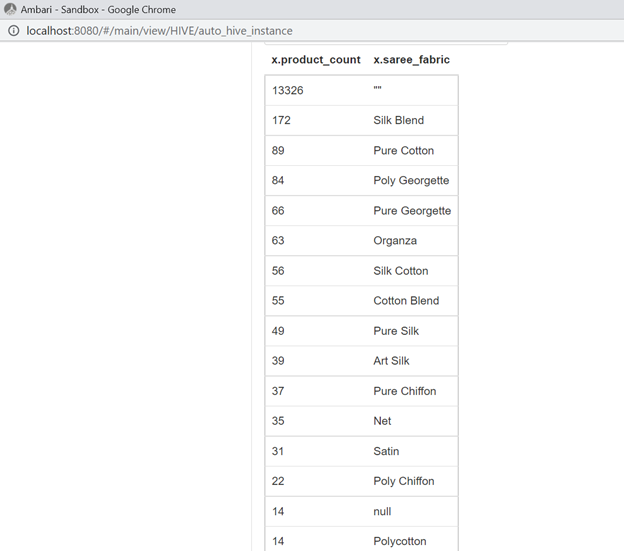
It is said that the saree was originally woven from pure silk, as stated by Rajput 2020. With the passage of time, sarees were crafted both entirely of cotton and with a silk-cotton blend (silk yarn in the warp and cotton in the weft). Wool is also being used in the modern saree industry.

I explored different type of saree fabric purchased by different customers and silk blend is the most preferred.

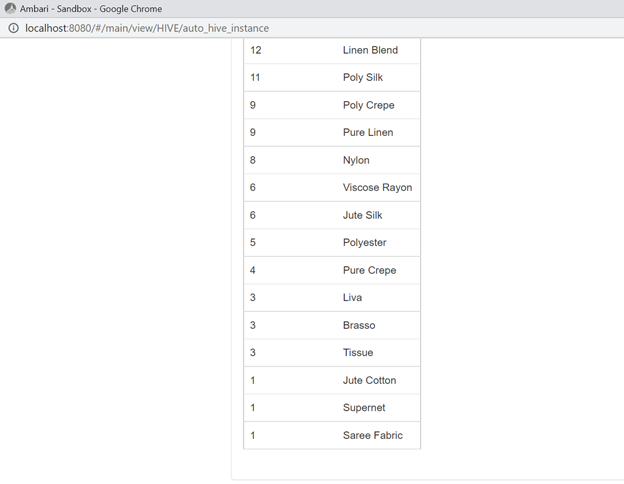


**Figure 24**: Showing code for selecting distinct saree\_fabric

Result:



**Figure 25**: Showing result of saree fabrics count



**Figure 26**: Showing result of saree fabrics count

# DATA ANALYSIS

MapReduce's ability to process large data sets in parallel makes it an excellent tool for data analysis due to its speed and efficiency. In the event that a node in the cluster fails, the job can be redeployed to a healthy node without losing any information.



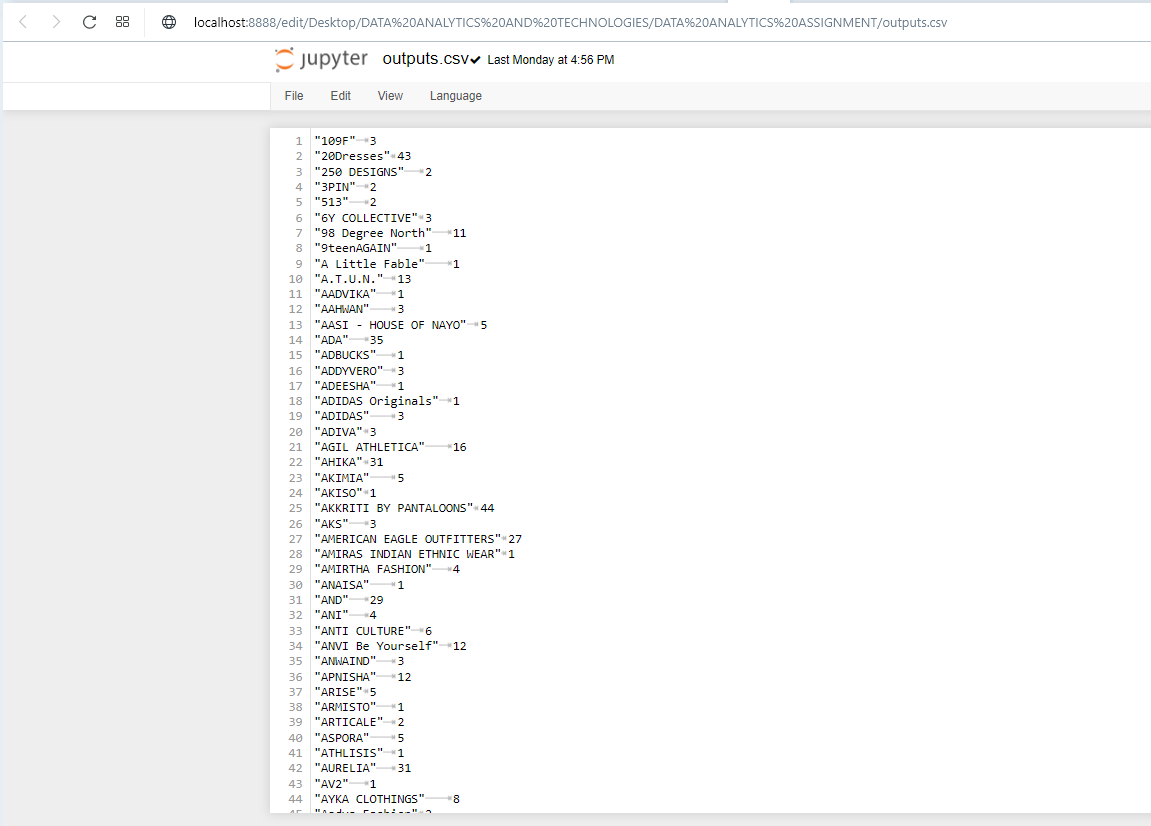
**Figure 27**: Showing Mapper and reducer function code of the MRJob

The provided MRJob code establishes an "assignment" class which extends the MRJob class. There are two methods in this class: a mapper and a reducer.

In order to process a line of input, the mapper method calls the split function, which divides the line at the ',' character. Once the line is split, if it has more than 5 items, the mapper method uses the first 6 items to determine the key and the value 1, respectively. When a line is split into five or fewer parts, the mapper method uses the first five parts to create a key and a value of 1 based on the fifth part (the brand).

A brand and its associated values are input into the reducer method (counts). As output, you'll receive both the key and the total of the values.

The MRJob code specifies a MapReduce task to count how often each brand name appears in the given data. Data is parsed by the mapper method, then outputs fields with the brand name as the key and a value of 1 for each field. The reducer procedure accumulates the numbers (values) for each brand and returns both the brand and the sum as output.



**Figure 28** Showing the amount of time a brand appear in the dataset

# MACHINE LEARNING

One of the most important tasks in the machine learning process is selecting the appropriate technique. Of the many machine learning method methods available, Association Rules, Classification Trees, Cluster Analysis, Decision Trees, and Neural Networks stand out as particularly well-suited to real-world applications. (li et al 2019)

Plotnikova et al 2022 mentioned that there are three models which are widely adopted by professionals and researchers for the machine learning process and these types are; Knowledge Discovery Databases (KDD) process model, CRISP-DM and SEMMA.

I adopted the KDD method. Nine steps make up the Knowledge Discovery Databases (KDD) paradigm for understanding data and highlighting the significance of individual data points.( Yafooz et al. 2020)

The nine steps of the Knowledge Discovery in Databases (KDD) process are:

1. Data selection: selecting the data that will be used for analysis.

2. Data preprocessing: cleaning and preparing the data for analysis.

3. Data transformation: transforming the data into a form that is more suitable for analysis.

4. Data reduction: reducing the size of the data while maintaining its integrity and usefulness.

5. Data mining: applying algorithms and statistical techniques to the data to discover patterns and relationships.

6. Pattern evaluation: evaluating the discovered patterns to determine their usefulness and validity.

7. Knowledge presentation: presenting the discovered patterns in a way that is understandable and useful to the user.

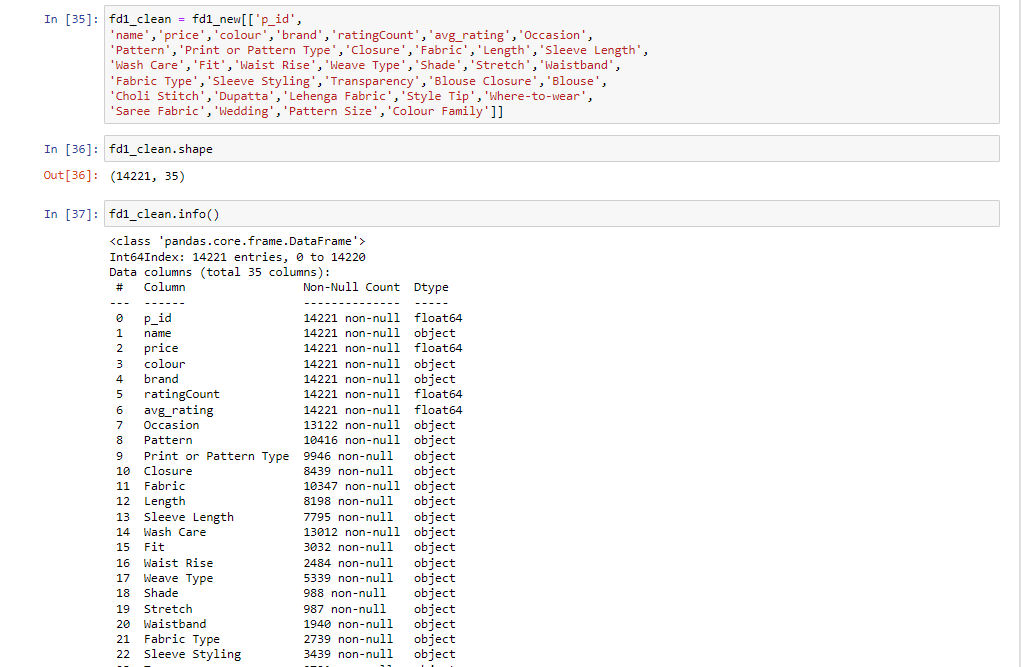
8. Knowledge deployment: implementing the discovered patterns in the real world.

9. Evaluation: evaluating the effectiveness of the deployed patterns.

From the steps above, step 1 – 4 has been deployed at the beginning of this report.

To begin the machine learning, the clean pre-processed dataset was used.

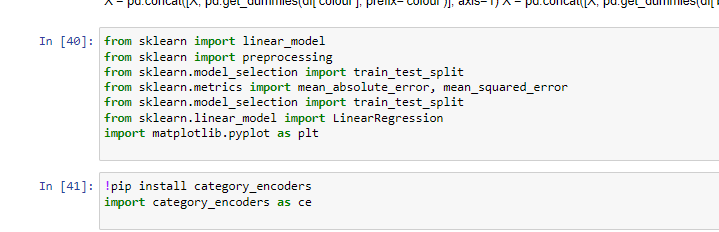
# DATA REDUCTION



**Figure 29** Showing the reduced dataset to be used for machine learning

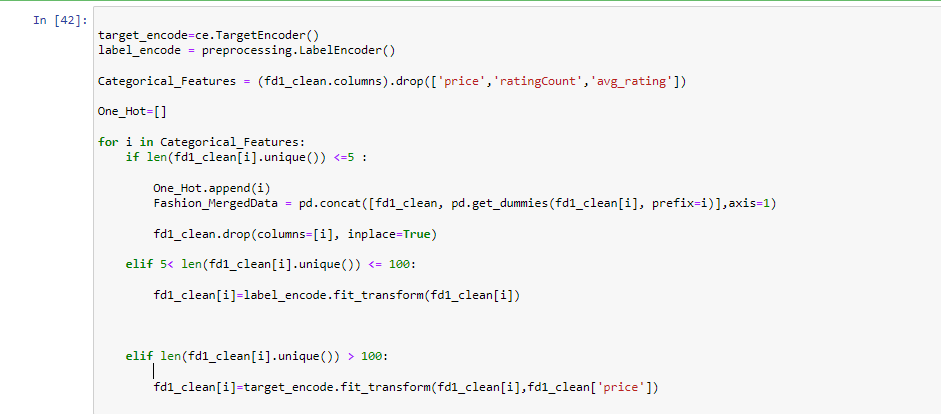
# DATA MINNING AND PATTERN EVALUATION

Next was to import all the necessary libraries



**Figure 30** Showing the imported libraries

I encoded the descriptive variables using one-hot encoding, target encoding and label encoding where necessary in order to perform different machine learning analysis on it.



**Figure 31** Showing the code for encoding

When there are fewer than five possible outcomes for a categorical feature, the one-hot encoding strategy is implemented. For each distinct value of the categorical feature, it generates a new binary column, with a value of 1 denoting the presence of that value in the original data and a value of 0 denoting its absence. After that, we get rid of the first column.

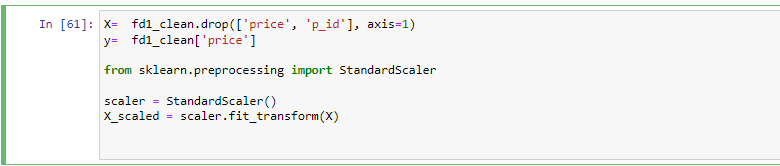
When working with categorical features that can take on between 5 and 100 different values, the label encoding technique is employed. It generates a distinct numeric value for each feature category.

Since the categorical features has more than 100 possible values, I used the target encoding technique. Each featured category is given a weight that is proportional to its relative importance, and the target variable (here, the price column) is distributed accordingly.

Both the LabelEncoder class from the sklearn.preprocessing library and the TargetEncoder class from the category encoders library are imported so that these encoding techniques can be used in the code. Then, it generates new instances of the label encode and target encode classes.

Using the resulting Categorical Features variable, which is constructed by selecting all columns in the fd1\_clean dataframe other than price, ratingCount, and avg rating, I can isolate the categorical features of the data.

The program then iterates over these categorical features, encoding them as either "one-hot" values (a single value) or "label" values (a set of possible labels) or "target" values (a set of possible targets) depending on the number of unique values in each feature. The encoded information is then saved in the fd1\_clean table.

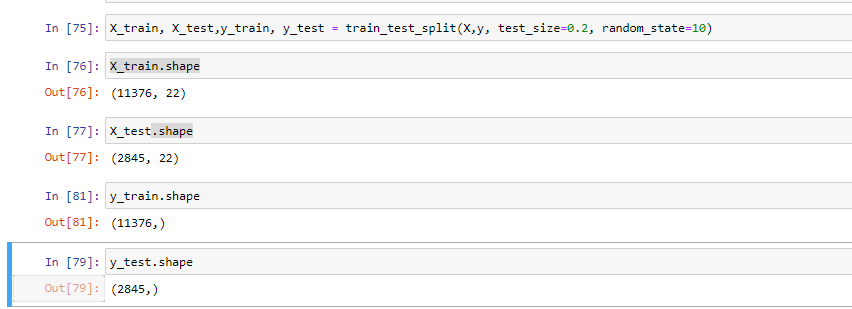


**Figure 32** showing standardizing the price column

This code shows a matrix of independent variable X and a dependent variable y from the fd1 clean dataframe.

In order to make X the independent variable, I first remove the price and p\_id columns from fd1\_clean, and then we make y dependent variable by selecting only the price column.To then normalise the features in X, the StandardScaler class is imported from the sklearn.preprocessing package.

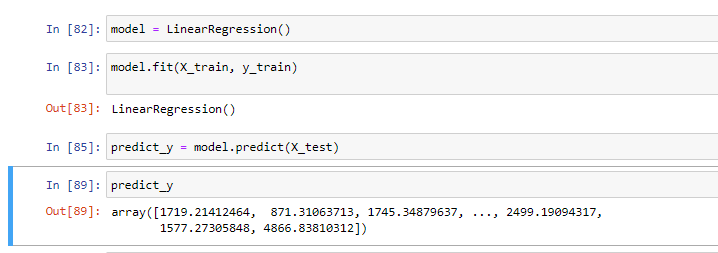
I splitted the data into train and test in order to perform regression on it



**Figure 33** Showing the test and trained dataset

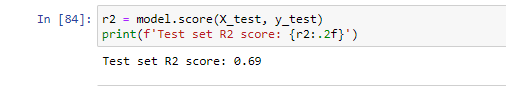
This code shows the splitting of the data set into train and test, with train size of 80% and test size of 20% using a random state of 10

I ran a linear regression on the model and show prediction based on the test set.



**Figure 34** Showing the linear regression and the prediction

Accuracy of my model showed 0.69 which indicates it is closer to being fit.



**Figure 35** Showing accuracy of themodel

How well a model fits the data is quantified by its R2 score, also called the coefficient of determination. A score of 1 indicates an excellent match, while a score of 0 indicates no such match exists.

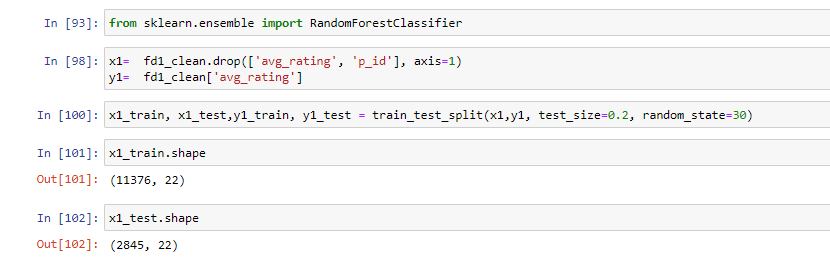
If the R2 value for the test set is 1.00 for a linear regression model, it means that the model is able to perfectly predict the target variable values given the features in the test set. This suggests the model does a good job of capturing the underlying relationships in the data and will generalise well to novel datasets.

# RANDOM FOREST

In the realm of machine learning, the ensemble learning family includes the widely used Random Forest algorithm. It is an algorithm for supervised learning, and it can be applied to both classification and regression problems.

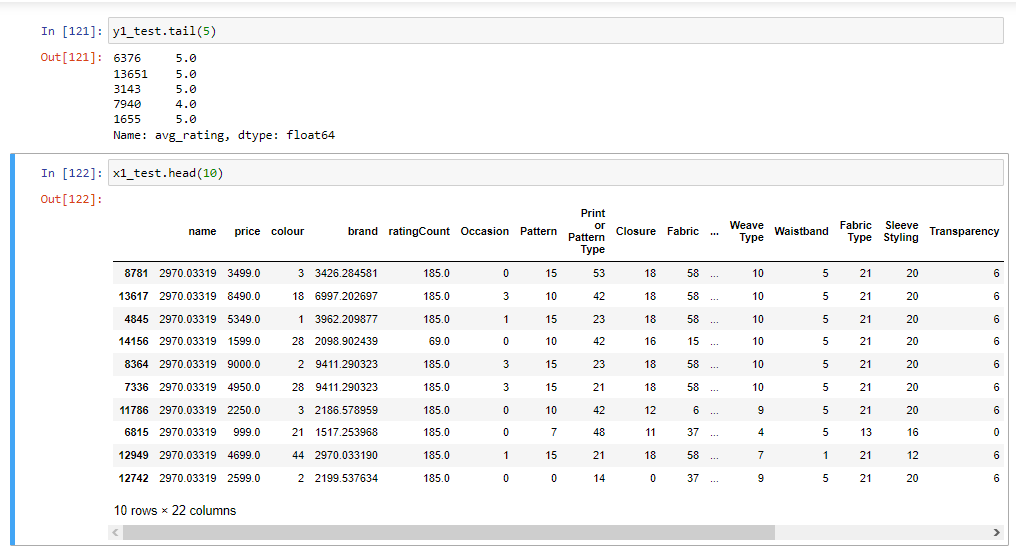
When it comes to high-dimensional and categorical data, Random Forest shines. It is also resistant to overfitting and has a high performance level across a variety of tasks. For classification or regression tasks respectively, the RandomForestClassifier and RandomForestRegressor classes in the sklearn.ensemble module of scikit-learn can be imported for use with a Random Forest model.

I adopted the RandomForestClassifier class to train a Random Forest model for the fashion dataset.



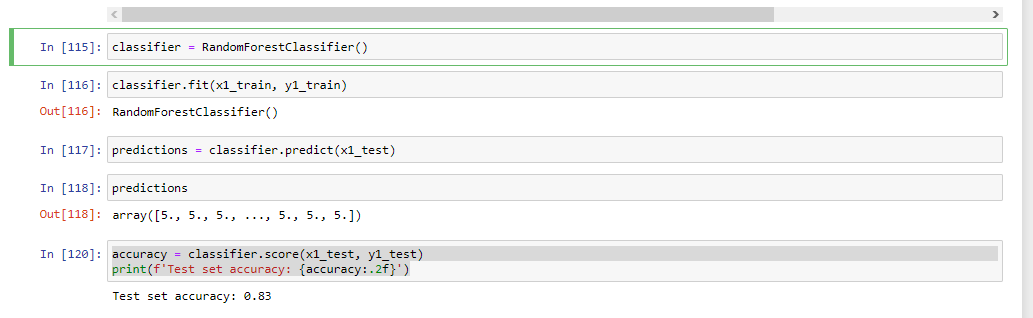
**Figure 36:** Train and test variable for RandomForestClassifier

The randomforestclassifier library was successfully imported. The fd1\_clean table's columns were all used for the independent variable x1, with the exception of the avg rating and p id columns, while the avg rating column represents the dependent variable y1



**Figure 37:** Showing pattern of the dataset

After splitting the dataset into train and test, I ran the randomforestclassifier on the avg\_rating to predict avg\_ratings based on the other columns and got accuracy of 0.83

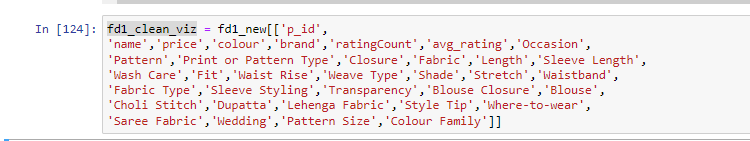


**Figure 38:** Showing accuracy of the model

In the above output, the accuracy of the model on the test set is 0.83, which means that the model made correct predictions for 83% of the samples in the test set.

It is common practice in machine learning to evaluate the performance of a model on a separate test set, in order to get a better estimate of the model's generalization performance. This helps to ensure that the model has not overfitted to the training data and is able to generalize to unseen data.

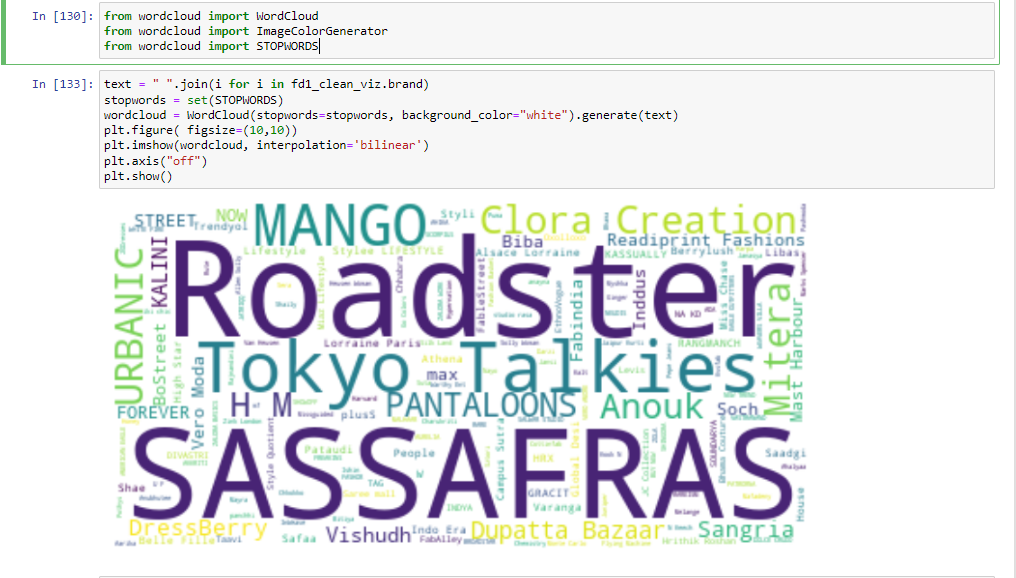
# KNOWLEDGE PRESENTATION



**Figure 39:** Showing List Of Columns To Be Used For Knowledge Presentation

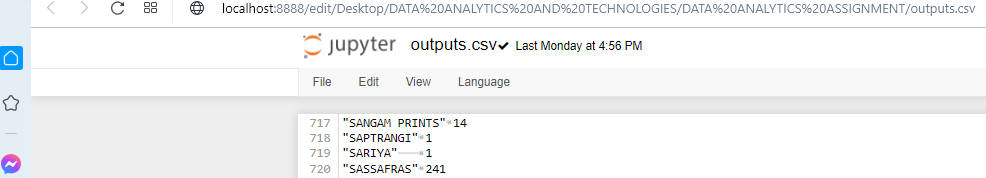
The author imported the wordcloud to graphically represent most occurring brands.

Word clouds generated with the Python WordCloud library are called "Python word clouds". A word cloud is a graphical depiction of the prevalence and importance of individual words within a body of text. Usually, it's made by showing a list of words, with the size of each word representing its frequency or importance. Customizable in terms of font, colour, and layout, word clouds are commonly used to illustrate the overarching concept or feeling of a piece of text.



**Figure 40** Showing wordcloud graphical representation of brand

The above diagram shows sassafras as the most occurring brand this could also be seen from our mapreduce result where sassafras occurred 241 times.



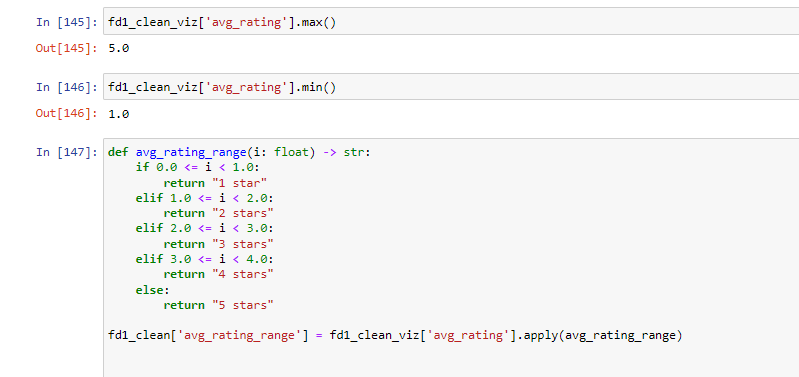
**Figure 41** Showing brandcount on mapreduce



**Figure 42** Showing wordcloud graphical representation of colour

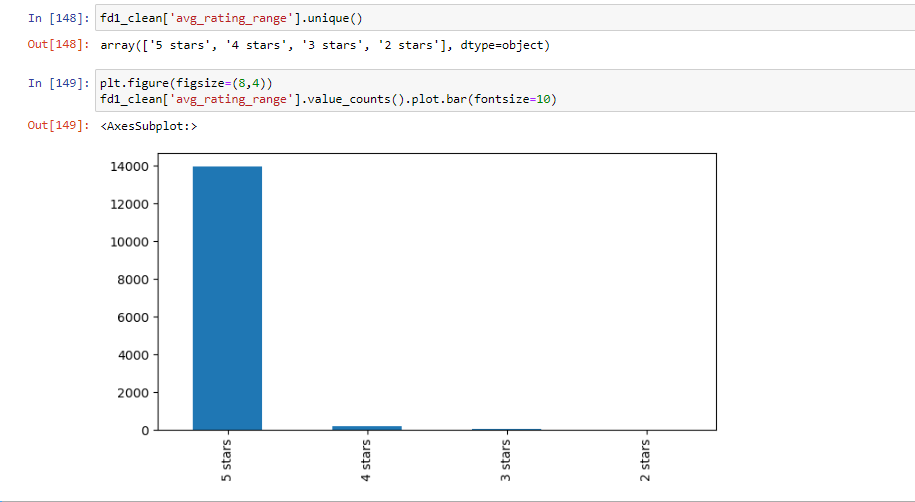
For colour, Black is seen as the most occurring colour which could be seen from our analysis on hive where black rank as number 1

To visualize the avg\_rating, I checked for the minimum and maximum number so as to define by stars



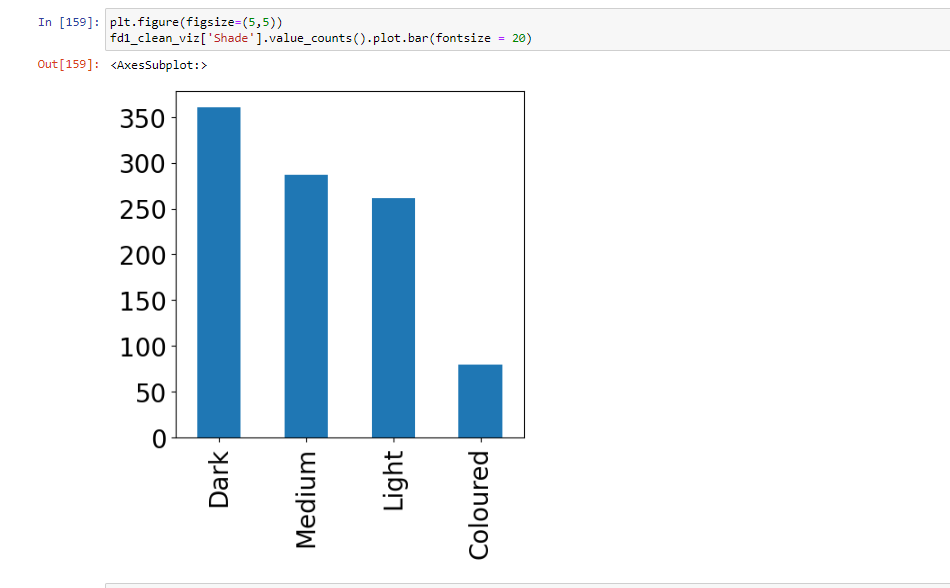
**Figure 43 Showing code for**

Then displaying it on a barchart

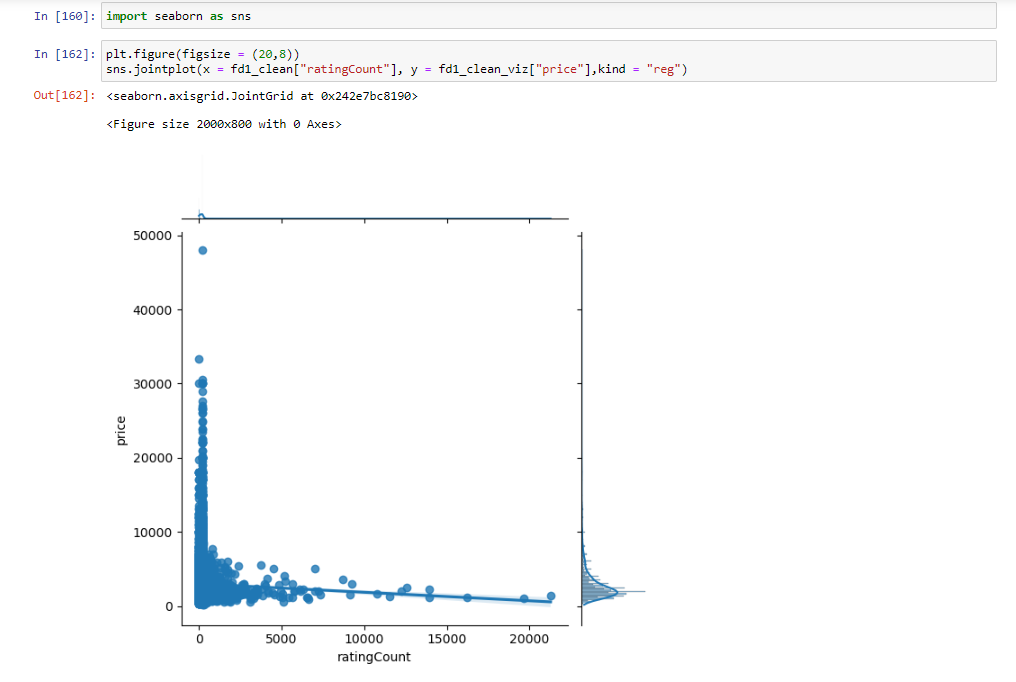


**Figure 44** Showing barchart of average rating on stars

I also visualized the shades in a barchart



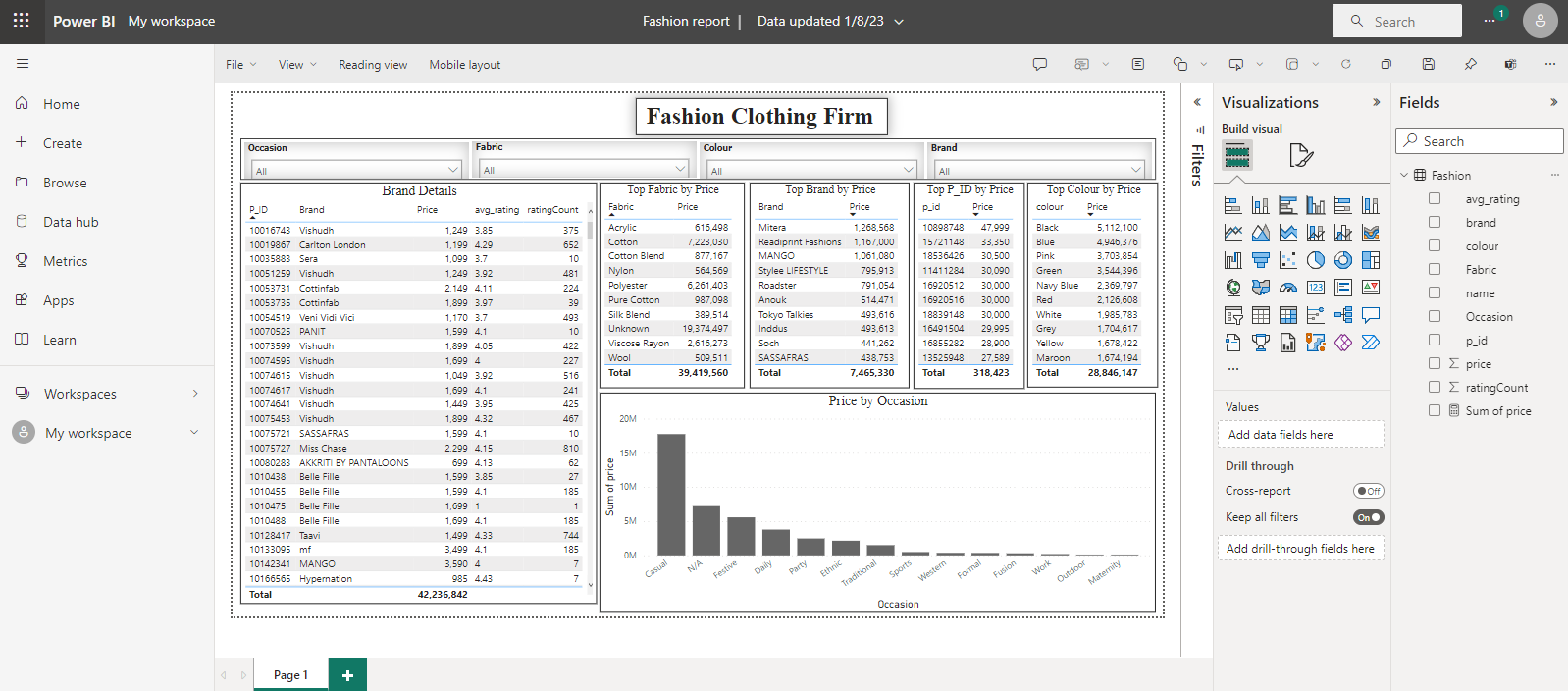
**Figure 45** showing barchart of colour shade



**Figure 46** Showing graphical representation of rating count based on price

# VISUALIZATION

I used PowerBi for my visualization, below is my visualization showing brand details, top fabric by price, top brand by price, top product ID by price and Top colour by price.



**Figure 47** Showing visualization

# SECTION TWO

# Current Issues in the Use of Big Data Analytics in The Fashion Retail Industry.

# Introduction

In recent years, technology is rapidly changing businesses in many ways, as it now allows businesses from diverse sectors to examine and analyze data growth (Russom, 2011). Further, with Big Data analytics, airline operators can now identify the number of passengers using their service, telecommunication companies can now monitor major trends in customer data usage, and doctors can now identify genes are responsible for the type of disease, ultimately social networking platforms can now tailor ads to a wide range of demography (Sazu and Jahan, 2022).

Big data experts often wonder about data analytics. Big Data analytics collects, organizes, and analyzes large amounts of data (called big data) to find schedules and other useful information. Big data analytics helps companies understand data and identify important data needed for future business and business decisions (Sazu and Jahan, 2022).

Across the globe, the retail and fashion industry are also witnessing change as the world brace up for the fourth industrial revolution with the adoption of Big Data technologies (Jain et al., 2017). In the fashion retail industry, big data usage is also gaining traction. With big data analytics, fashion retailers can now drive one-to-one engagement with online shoppers. Also, most of the systems that enable big data can detect false signals by analyzing real-time machine learning to accurately save users and illegal transactions and create alertness, thereby validating the importance of big data analytics in the fashion retail industry (Bertola, 2017). Thus, this report will not only identify current issues in the usage of Big Data Analytics in the fashion retail industry, it will also carefully proffer effective solution using various technologies.

# What Are the Identified Issues?

The fashion industry is timeless and will push its boundaries over time as it continues to grow (Jain et al., 2017). This global industry will see rapid growth in companies, suppliers and investors, and during this expansion there will be a growing need to generate a return on investment that suits and benefits one of the hierarchies. This is where Big Data Analytics comes in as an integral part of the industry innovative expansion (Kumar and Sikka, 2022).

More importantly, data gathering and turning it into business intelligence which has long helped fashion retailers like Zara to increase productivity, improve decision-making and gain a competitive advantage as a key part of their strategy is now a big thing among many other retailers (Pantano et al., 2018). However, the changes and adjustments to the standards were not without problems. Some of the issues identified by the author are:

1. **Ethical Issues about Customer Data**

Ethical issues related to the use of big data analytics in the fashion retail industry include algorithmic bias and the sale of user information for remarketing purposes. It's more common than people think that third parties will use their access to big data for commercial purposes, and they often do so without the customer's knowledge or consent (Zhao, 2022).

1. **Talent and Skills gap:**

Those with the knowledge and abilities to mine insights from large data sets are in high demand as they are essential for realising the benefits of big data (Russom, 2011). Experts with the technical knowledge and training to manage and analyse large datasets are in short supply and in high demand. Finding and retaining employees with the expertise needed to leverage large data sets is a major challenge for the retail industry (Silva et al., 2019).

There is a shortage of professionals with the necessary skills to analyse and interpret big data in the fashion industry. This can make it difficult for retailers to fully leverage the potential of big data analytics.

1. **Data Privacy and Security:**

The third challenge of big data analysis is protecting the privacy of individuals' information (Silva et al., 2019). There are many channels through which retailers can collect customer data, including social media and loyalty programmes (Jain et al., 2017). For retailers, the challenge will be to keep customers' personal information private in light of the fact that big data platforms can now collect massive amounts of data on a single person. In the fashion industry, where big data analytics is becoming increasingly prevalent, some people are worried about their privacy. Given the prevalence of data breaches in the fashion retail industry, customers may be wary of providing their personal information to these businesses.

1. **Analytical Algorithm Scalability Issue**

The collection, processing, and utilisation of big data by the fashion retail industry is not standardised. As a result, it can be challenging for retailers to compare and analyse data from various channels. One of the difficulties of data analysis is the scalability of the algorithms used for big data analysis, given that big data is concerned with massive amounts of data (Russom, 2011). A scalable algorithm is one that can easily accommodate growing data sets (Sazu and Jahan 2022). Because of the exponential increase in data volume, the efficiency of data retrieval processes degrades rapidly (Giri et al., 2019). In light of this, analytical algorithms should be developed to guarantee scalability as the size of datasets increases.

1. **Integration with traditional methods:**

Retailers can leverage a number of technologies, such as real time location data from smartphones, to collect information on customers‟ in-store behaviour (customer footpath and time spent in different parts of the store) (Giri et al., 2019). Many fashion retailers have been using traditional methods of data analysis for decades, therefore, it can be challenging for them to integrate big data analytics into their existing processes and systems.

# Effective Solution Using Various Technologies

No doubt, the fashion retail industry will continue to face difficulties and several identified issues due to increasing inflation, costly regulations, and a less robust economy if it fails to integrate into the new world order of big data. Thus, it is imperative than ever before for fashion retailers must focus on all possible strategies to reduce costs and increase sales through the adoption various technologies in big data analytics.

1. **Data Governance And Protection**: To safeguard customers' private information and forestall data breaches, brands in the fashion retail industry must, first, adopt stringent data governance policies and invest in cutting-edge security technology.
2. **Data Standardisation**: Businesses in the fashion retail sector can collaborate to develop common guidelines for data collection, processing, and analysis, making it simpler to combine data from various sources and apply analytical techniques.
3. **Ethical Considerations**: With regards to ethical consideration issues, retailers in the fashion industry must make sure that their data analysis practises are open and honest by doing things like removing any potential bias from their algorithms and obtaining customers' permission before using any personal information.
4. **Talent Development**: businesses should invest in their employees' professional growth by providing them with training and development opportunities related to big data analytics.
5. **Integration with Traditional Methods**: traditional methods can be integrated with big data analytics through the help of data science and analytics professionals, allowing stores to take advantage of both sets of insights and knowledge.

**Various Technologies** that can be used to address these issues are:

1. **Data and Warehouse Management Systems**

To store multiple copies of transaction data in a format suitable for query and analysis, a database called a data warehouse is used. Storage, retrieval, and analysis of structured large datasets are typically handled by data warehouses or data marts (a subset of a data warehouse) (Chen and Zhang, 2014). Processes of extraction, transformation, and loading are used in data warehouses to convert massive datasets into a more manageable, structured format for storage and retrieval (Giri et al., 2019).

Considering the increased number of suppliers fashion retailers must work with, each of which has different standards for managing their own product data, this presents a challenge for the industry as a whole. This creates the challenge of efficiently onboarding and standardising these data sources in order to reduce time to market. Those businesses that make the most of their highly automated capabilities, such as a variety of tools that make it simpler to unify the relationship between retailers and suppliers/vendors, will have a distinct advantage in the market.

Due to the large number of stores that have had to quickly adapt to compete in the online retail space, customer experiences on retail platforms can vary widely. On the other hand, this means that consumers have more options than ever before, and businesses whose offerings feature comprehensive, reliable details about their wares will undoubtedly succeed.

1. **Technologies For Data Analysis And Visualisation**

Through the use of data analytics, retailers are now able to comprehend the inner workings of their various retail divisions (logistics, stores, sales, etc.) by analysing customer, inventory, and sales information. By revealing the characteristics of all points, data visualisation in retail has fueled sales strategy, improved the customer experience, and streamlined the entire retail process.

As information gathered from social media, online shopping, geo-tagged smart interactions, etc., generates numerous recommendations, the retail industry has embraced data-driven customization. As a result, businesses will be able to better understand customer journeys and enhance the customer experience with the aid of data analytics and visual analytics (CX). Companies in the retail sector would be wise to implement a real-time tracking system, combining Big data analytics and data visualisation to anticipate future trends. Power BI, Tableau, and the programming language R are all examples of such tools (Acharya et al., 2018).

1. **Artificial Intelligence And Machine Learning**

The development of algorithms that teach computers to modify their behaviour in response to new or changing data is the focus of machine learning, a subfield of AI (Kumar and Sikka, 2022).

Discovery of information through the identification of patterns and the formation of well-informed judgements are two primary goals of machine learning (Chen and Zhang, 2014; Manyika et al., 2011). Predictions made by retailers can be more accurate and nuanced with the help of machine learning (Davenport, 2013). Natural language processing, supervised and unsupervised learning, and various types of ensemble learning are all types of machine learning (Giri et al., 2019, Banica and Hagiu, 2016).

Artificial intelligence's use in stock management is one area where the industry is optimistic about its future. AI appears to have infinite applications and outcomes. AI and ML can also be used by online clothing stores to make more tailored product suggestions.

You can use AI algorithms to improve your supply chain operations and shorten the time it takes to pack, stock, and ship customer orders. It can help you make sense of the mountains of data gleaned from your supply chain datasets, automate shipments and deliveries, and analyse and predict future orders. AI chatbots, also known as intelligent assistants, are another form of AI that finds application in the world of online clothing retail. These chatbots are designed to look and act like human customer service representatives in order to assist users in finding the information they need.

1. **Data Security And Governance Technologies**

Despite the doubts and concerns you may have had in the past, it's important to know that cloud storage is now more trustworthy and secure than storing data on-premises. Not only do cloud services offer more secure encryption, but they also remove the possibility of an employee stealing the server to gain access to confidential information in the fashion retail industry.

# **Conclusion**

For many fashion retailers like JD, Zara, SheIN, and Primark the driver behind their effective supply chain has been identified as the use of data and analytics for accurate forecasting and decision-making.

# **References**

Acharya, A., Singh, S.K., Pereira, V. and Singh, P., 2018. Big data, knowledge co-creation and decision making in fashion industry. *International Journal of Information Management*, *42*, pp.90-101.

Agrawal, D., Bernstein, P., Bertino, E., Davidson, S., Dayal, U., Franklin, M., Gehrke, J., Haas, L., Halevy, A., Han, J. and Jagadish, H.V., 2011. Challenges and opportunities with Big Data 2011-1.

Aktas, E. and Meng, Y., 2017. An exploration of big data practices in retail sector. *Logistics*, *1*(2), p.12.

Banica, L. and Hagiu, A., 2016. Using big data analytics to improve decision-making in apparel supply chains. In *Information Systems for the Fashion and Apparel Industry* (pp. 63-95). Woodhead Publishing.

Bertola, P., 2021, July. Fashion Within the Big Data Society: How can data enable fashion transition towards a more meaningful and sustainable paradigm?. In *CHItaly 2021: 14th Biannual Conference of the Italian SIGCHI Chapter* (pp. 1-8).

Cook, S.C. and Yurchisin, J. (2017), “Fast fashion environments: consumer’s heaven or retailer’s nightmare?”, International Journal of Retail and Distribution Management, Vol. 45 No. 2, pp. 143-157.

Giri, C., Thomassey, S. and Zeng, X., 2019. Customer analytics in fashion retail industry. In *Functional Textiles and Clothing* (pp. 349-361). Springer, Singapore.

Jain, S., Bruniaux, J., Zeng, X. and Bruniaux, P., 2017. Big data in fashion industry. In *IOP Conference Series: Materials Science and Engineering* (Vol. 254, No. 15, p. 152005). IOP Publishing.

Kaur, K. and Agrawal, A., 2019. Indian saree: a paradigm of global fashion influence. *Int. J. Home Sci*, *5*(2), pp.299-306.

Khan, M., 2019. Challenges with big data analytics in service supply chains in the UAE. *Management Decision*.

Kumar, J. and Sikka, S., 2022. BIG DATA IN FASHION INDUSTRY. Retrieved from <https://www.rajasthali.marudharacollege.ac.in/papers/Volume-1/Issue-3/03-23.pdf> in January 2023.

Li, J., Pan, S. and Huang, L., 2019. A machine learning based method for customer behavior prediction. *Tehnički vjesnik*, *26*(6), pp.1670-1676.

Madsen, D.Ø., Silva, E.S. and Hassani, H., 2020. The application of big data in fashion retailing: a narrative review. *International Journal of Management Concepts and Philosophy*, *13*(4), pp.247-274.

Mostafa, S.M., 2019. Imputing missing values using cumulative linear regression. *CAAI Transactions on Intelligence Technology*, *4*(3), pp.182-200.

Pantano, E., Giglio, S. and Dennis, C., 2018. Making sense of consumers’ tweets: Sentiment outcomes for fast fashion retailers through Big Data analytics. *International Journal of Retail & Distribution Management*.

Plotnikova, V., Dumas, M. and Milani, F.P., 2022. Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements. *Data & Knowledge Engineering*, *139*, p.102013.

Rajput, A., 2020. A royal representation of Indian culture.

Russom, P., 2011. Big data analytics. *TDWI best practices report, fourth quarter*, *19*(4), pp.1-34.

Sazu, M. and Jahan, S. (2022). How Big Data Analytics Impacts the Retail Management on the European and American Markets?. CECCAR Business Review. 3. 62-72. 10.37945/cbr.2022.06.07.

Silva, E.S., Hassani, H. and Madsen, D.Ø., 2019. Big Data in fashion: transforming the retail sector. *Journal of Business Strategy*.

Thomassey, S. and Zeng, X., 2018. Introduction: Artificial Intelligence for Fashion Industry in the Big Data Era. In *Artificial intelligence for fashion industry in the big data era* (pp. 1-6). Springer, Singapore.

Yafooz, W., Bakar, Z.B.A., Fahad, S.K. and Mithun, M., 2020. Business intelligence through big data analytics, data mining and machine learning. In *Data management, analytics and innovation* (pp. 217-230). Springer, Singapore.

Zhao, L., 2022. Understanding the Paradigm Shift to Fashion Big Data Analytics. In *International Textile and Apparel Association Annual Conference Proceedings* (Vol. 78, No. 1). Iowa State University Digital Press.