**DEPARTMENT OF DATA SCIENCE -**

**-FAST NUCES ISLAMABAD-**

****

**BDS-3A**

**IDS Project**

**Submitted to:**

Mam Saira Qamar / Sir Talha Tariq

**Submitted by:**

**Name**: Ibrahim Kiyani/Muhammad Noor

**Registration** **No:** 23i-2536/23i-2520

**Class / Section:** Ds-A

**Date: 04/12/2024**

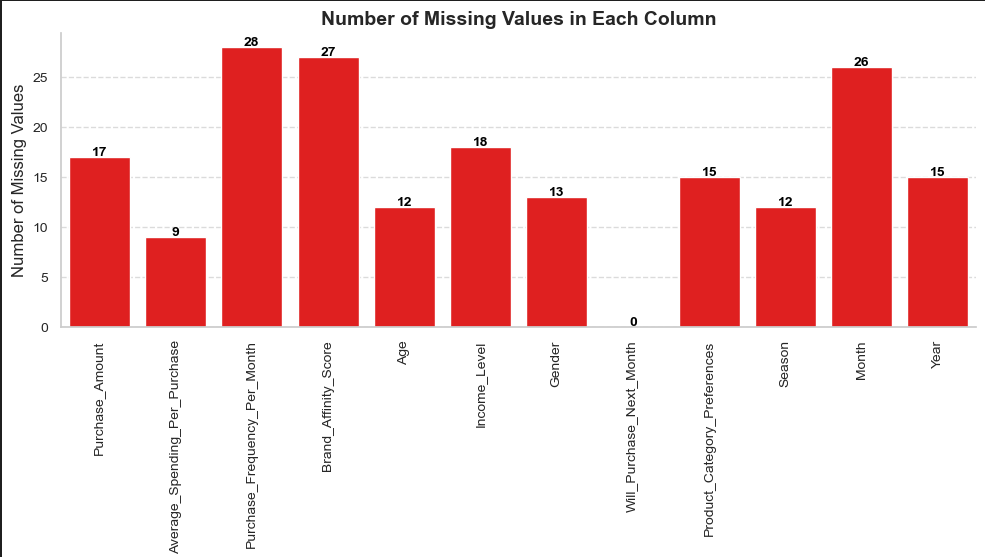
PURPOSE: Our project focuses on analyzing the product dataset from Imtiaz Mall, with the mission of uncovering valuable insights and finding the best solution to improve marketing strategies. We, Muhammad Noor (23i-2520) and Ibrahim Kiani(23i-2536), performed thorough data cleaning and preprocessing to ensure the quality of the dataset. Our approach towards this mission involved various techniques including regression.K-means clustering, decision trees. By leveraging these methods, we sought out patterns and trends that could guide better decision making for better customer engagements.

**MODULE 1- PREPROCESSING**

**SECTION 1- DATA ACQUISITION**

**1.4 CHECKING COLUMN-WISE NULL VALUES**

On Average, 4% of data is missing in each column with the most missing column being only 8% missing. No need to drop columns on the basis of missingness.



**1.5 VISUALIZING MISSING VALUES**

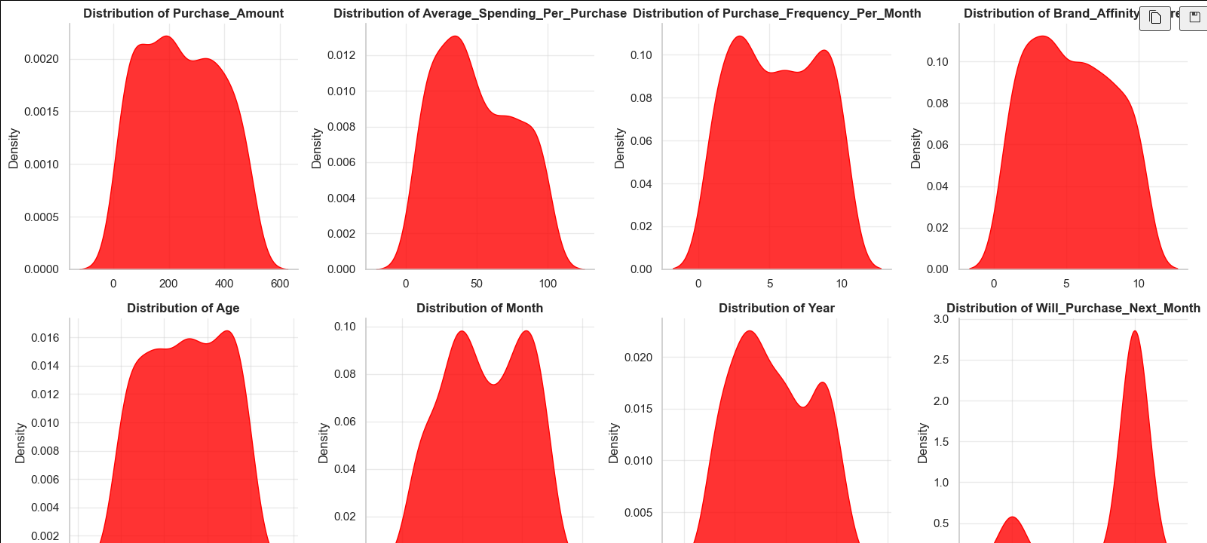
Plot for visualizing missing values per row is shown below. We can remove the entries with more than 2 missing values as they only contribute as only 1% of the data.



**SECTION 2- CLEANING**

**2.1 DETERMINING SKEWNESS AND CHECKING OUTLIERS**

Working on it, we got to know the data shows normal distribution.There are no extreme outliers with minimal skewness and kurtosis.



**MODULE 2- EXPLORATORY DATA ANALYSIS**

**SECTION 1- Univariate Data Analysis**

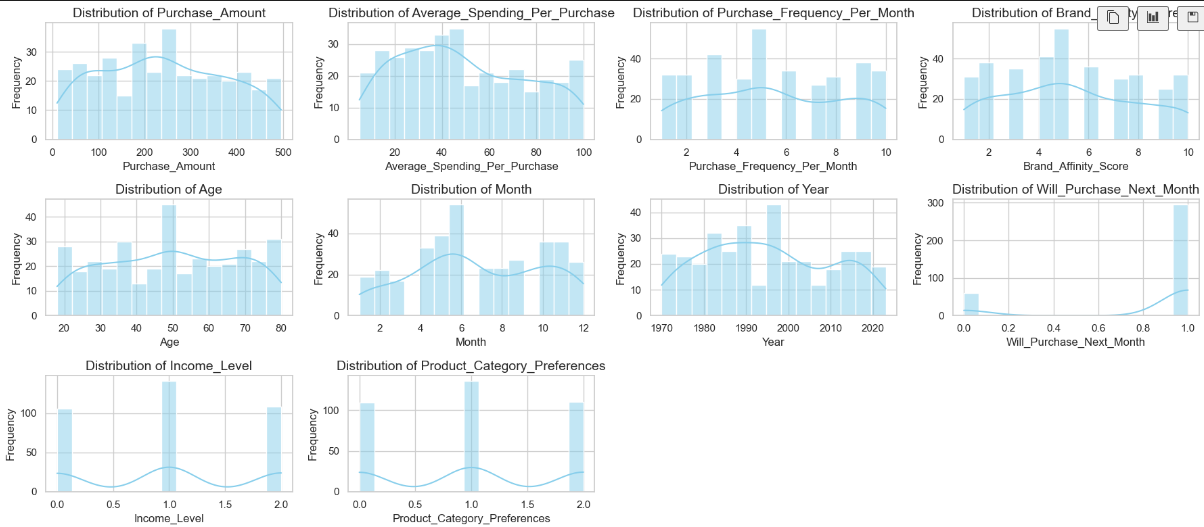
**1.2 HISTOGRAMS**

1. On plotting, we get to know average spending per month is more inclined to the lower end. This implies less of an inclination of customers to spend more money.

2. Year of purchase is also inclined more towards the left. This implies a decrease of customers over time, as aforementioned in the task provided.

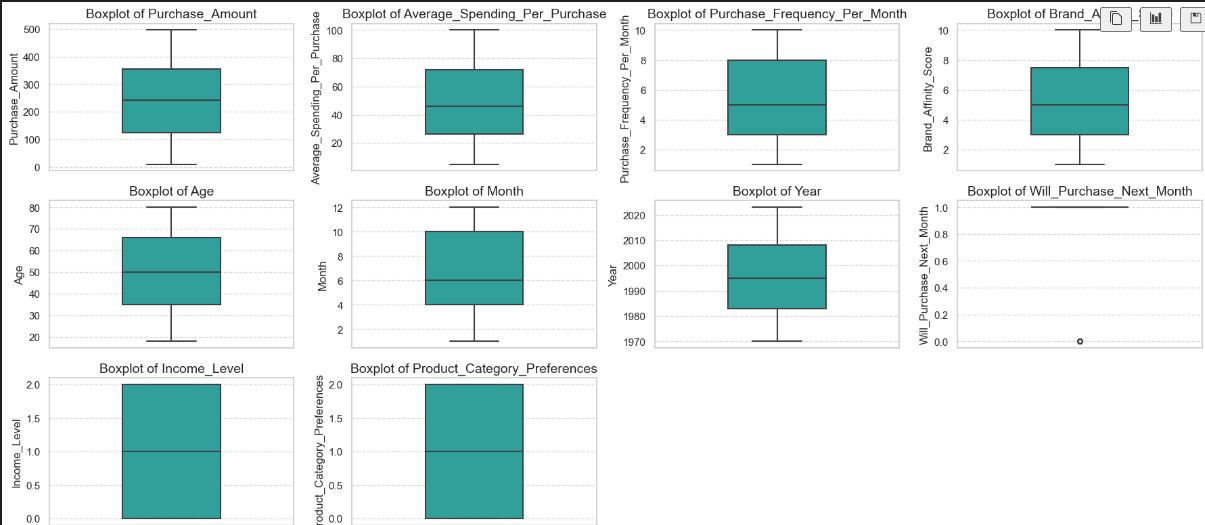
3. Majority of the purchases occur during the months of peak summer(4-6) and peak winter(10-11).

4. The rest of the attributes display normal distribution.



**1.3 BOXPLOTS**

On plotting, we get to know that no outliers were found.

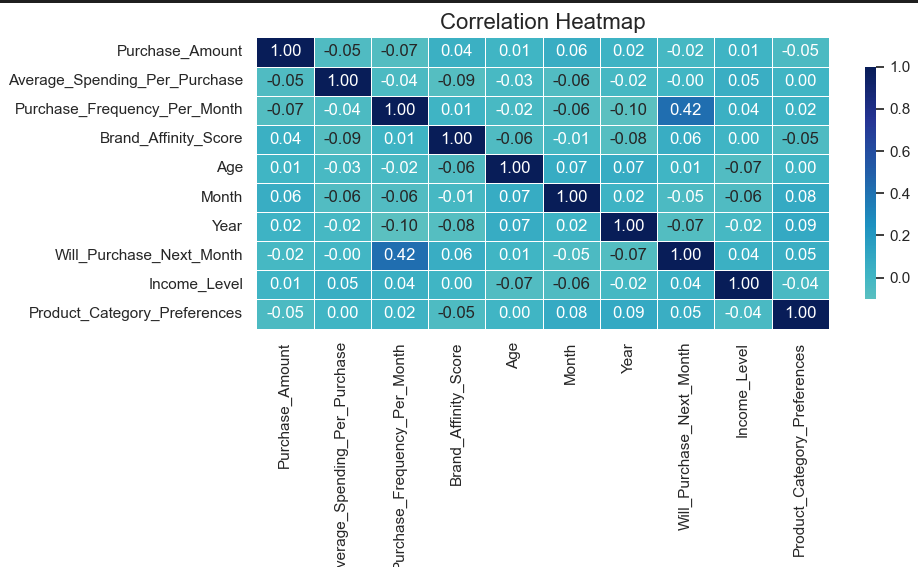


**SECTION 2- BIVARIATE DATA ANALYSIS**

**2.1 HEATMAPS**

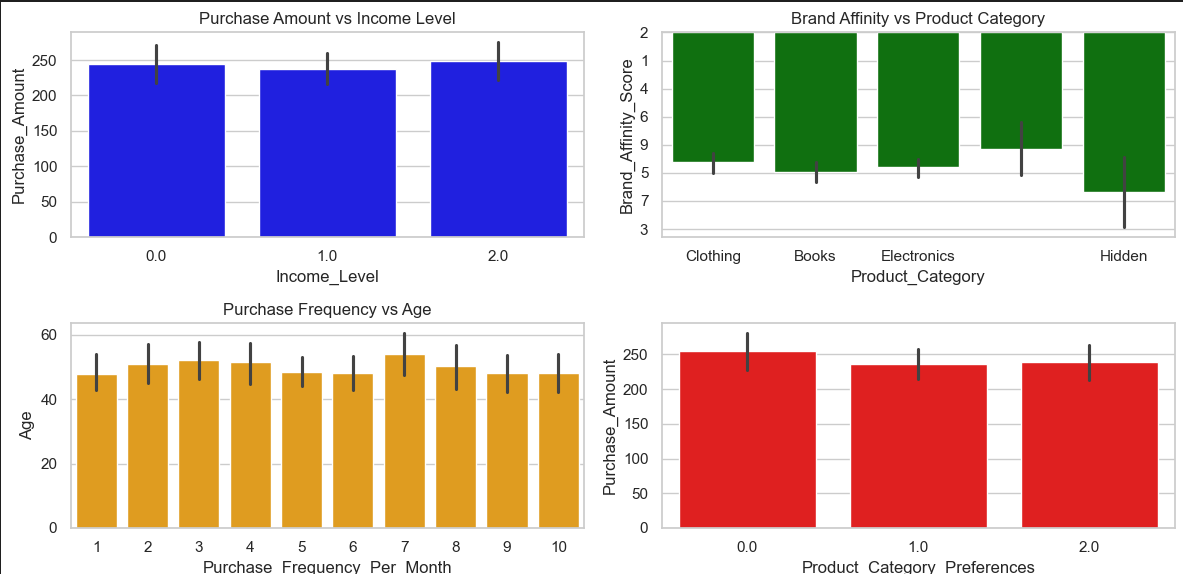
1. There is a high correlation between monthly purchase frequency and whether someone will purchase next month or not.

2. Brand affinity score plays some role on whether someone will purchase next month too.

3. Unfortunately,the likelihood of purchasing next month goes down monthly/yearly. Same with monthly purchase frequency and brand affinity.

**2.2 BOX PLOTS**

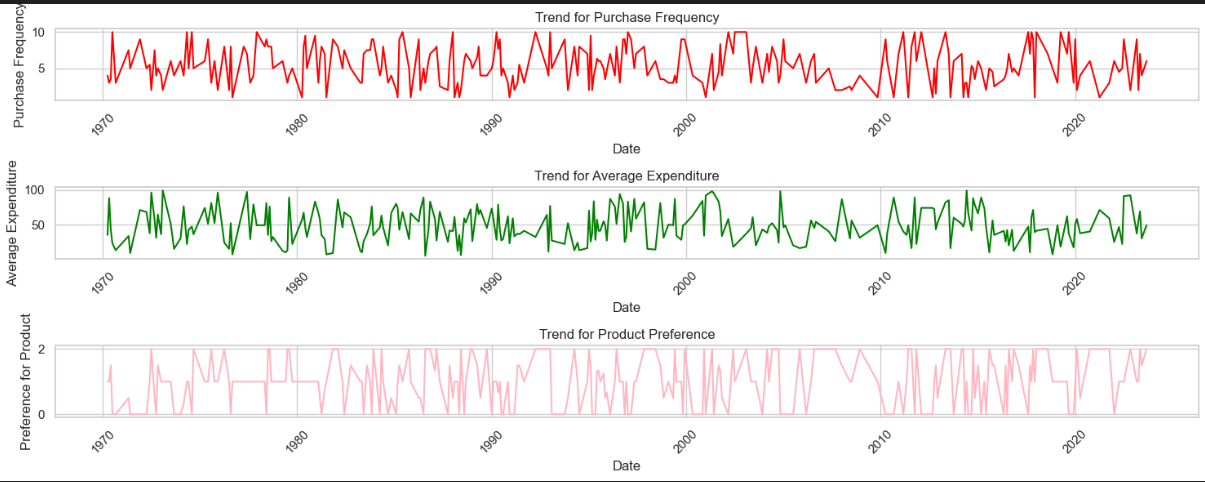
In these, we get to know the relation between different variables and the underlying trends in the dataset. Different lots are shown with different colours establishing distinct relations in the dataset.



**SECTION 3- TEMPORAL ANALYSIS**

**3.1 ANALYSING TRENDS**

In it, we get to analyze the seasonal trends which uncover more hidden relationships lying in the dataset. Different plots are made which help us to identity changes in

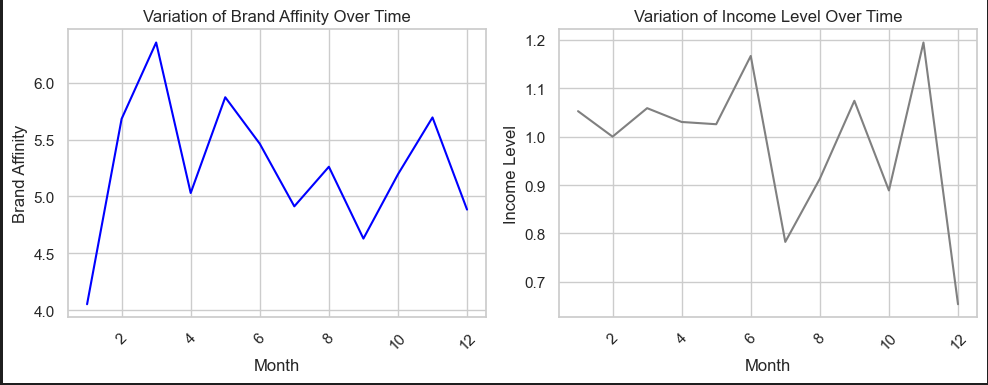
customer behaviour patterns over time.

**3.2 SEASONAL VARIATIONS**

1. Furthermore, more seasonal relationships have been found which show the customer behaviour pattern over months. Trend is more like a zigzag path with up and down trend over months.

2. Trend of Brand Affinity Over Time shows that most affinity was in March while the least was in January.

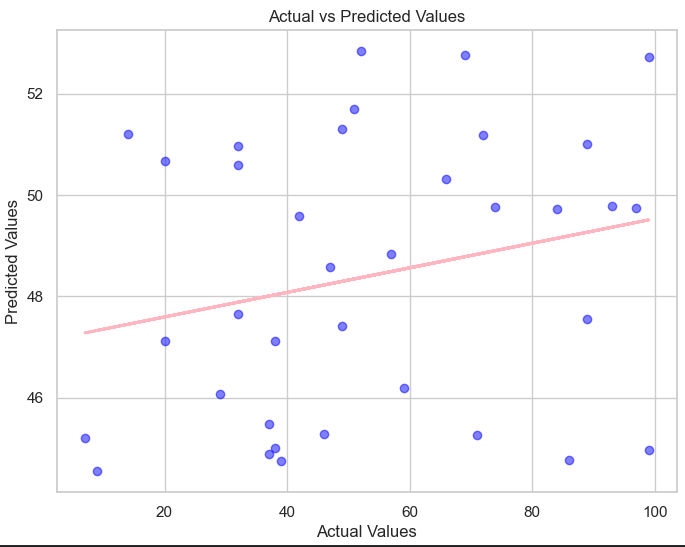
3. Trend of Income Level Over Time shows that customers had the most Income level in November and least in December.



MODULE 3- REGRESSION

**SECTION 1- LINEAR REGRESSION ANALYSIS**

Regression line shows the best approximation line between actual values and predicted values. Observing it, we can find the overall trend of the values in the dataset.



**SECTION 2- DECISION TREE ANALYSIS**

In this model, we split the data into training and testing sets and then make predictions on it.

