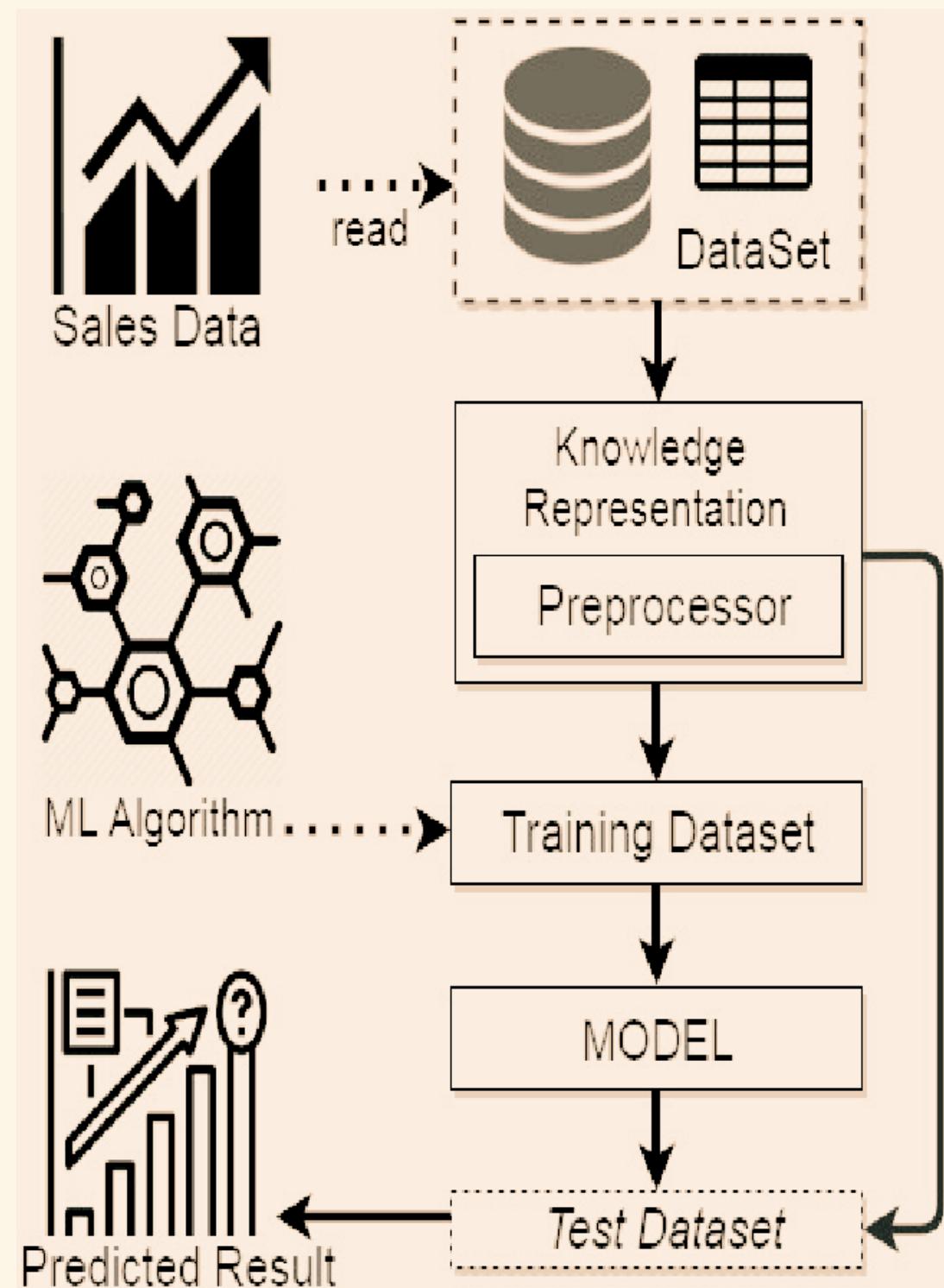


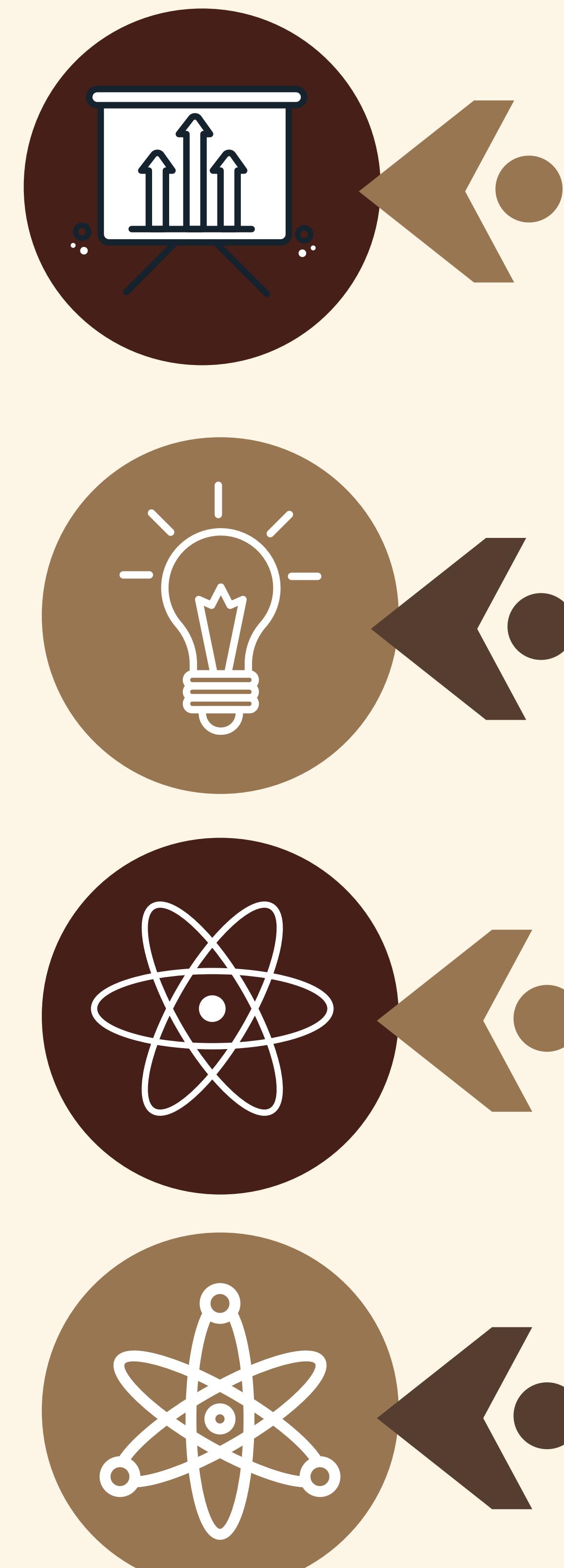
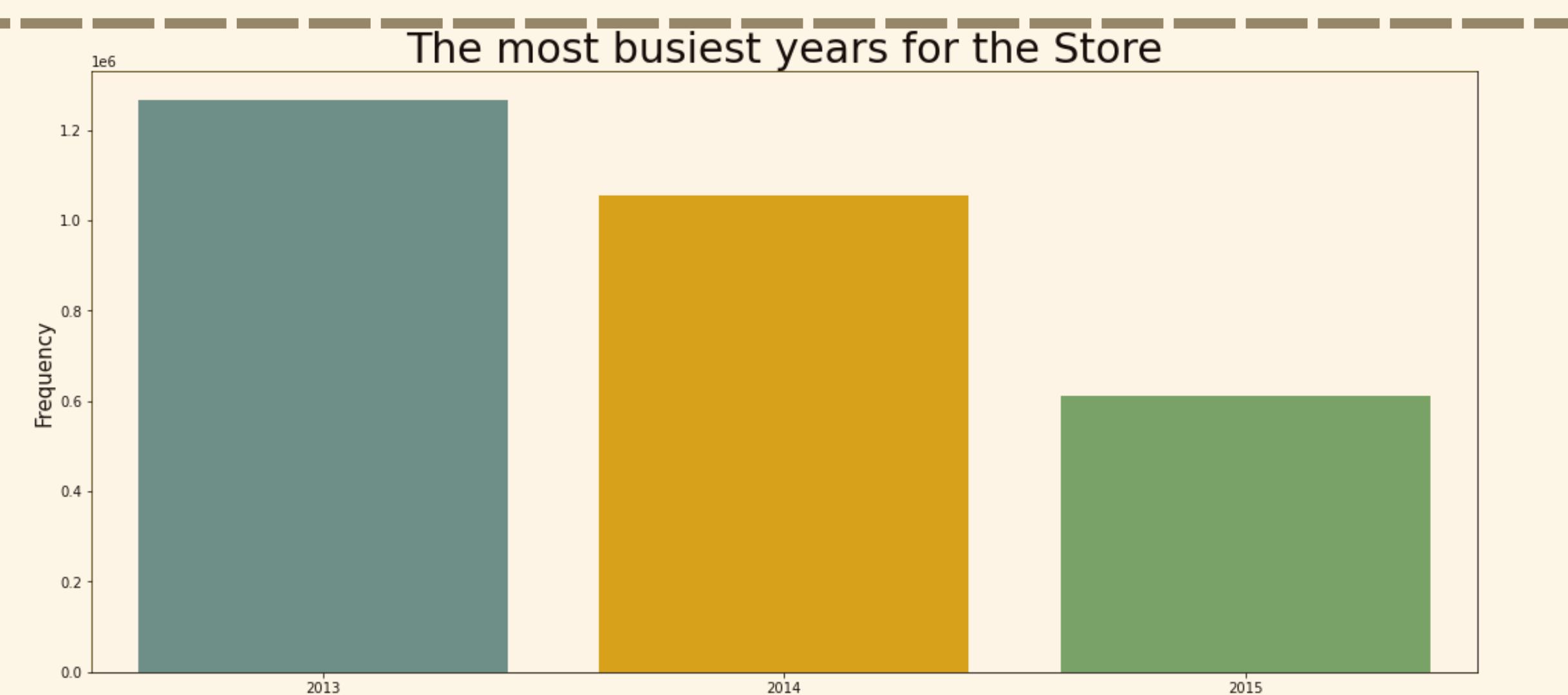
SALES PREDICTION: A MULTIVARIATE APPROACH

Incorporating ML techniques and ensemble models to enhance sales forecast leading to optimized inventory management and more effective sales strategies

Advayta Zadoo (22201078)
Devansh (22201475)



Architecture



MOTIVATION

Sales forecasting remains crucial for effective decision-making and it enables businesses to adapt to evolving customer behaviors, optimize resource allocation, manage inventory efficiently to stay ahead of the competition. Thus, accurate sales predictions are essential for navigating uncertainties. Machine learning has become increasingly valuable in sales forecasting due to its ability to analyze large volumes of data and uncover complex patterns. It can capture non-linear relationships, seasonal trends etc. to generate more accurate predictions.

METHODOLOGY

After acquiring the dataset, we would follow these steps:

1. Data understanding
2. Exploratory Data Analysis
3. Model Training and Fitting (Time Series Forecasting, Decision Tree)
4. Model Evaluation based on the predictive performance
5. Model Selection and future predictions

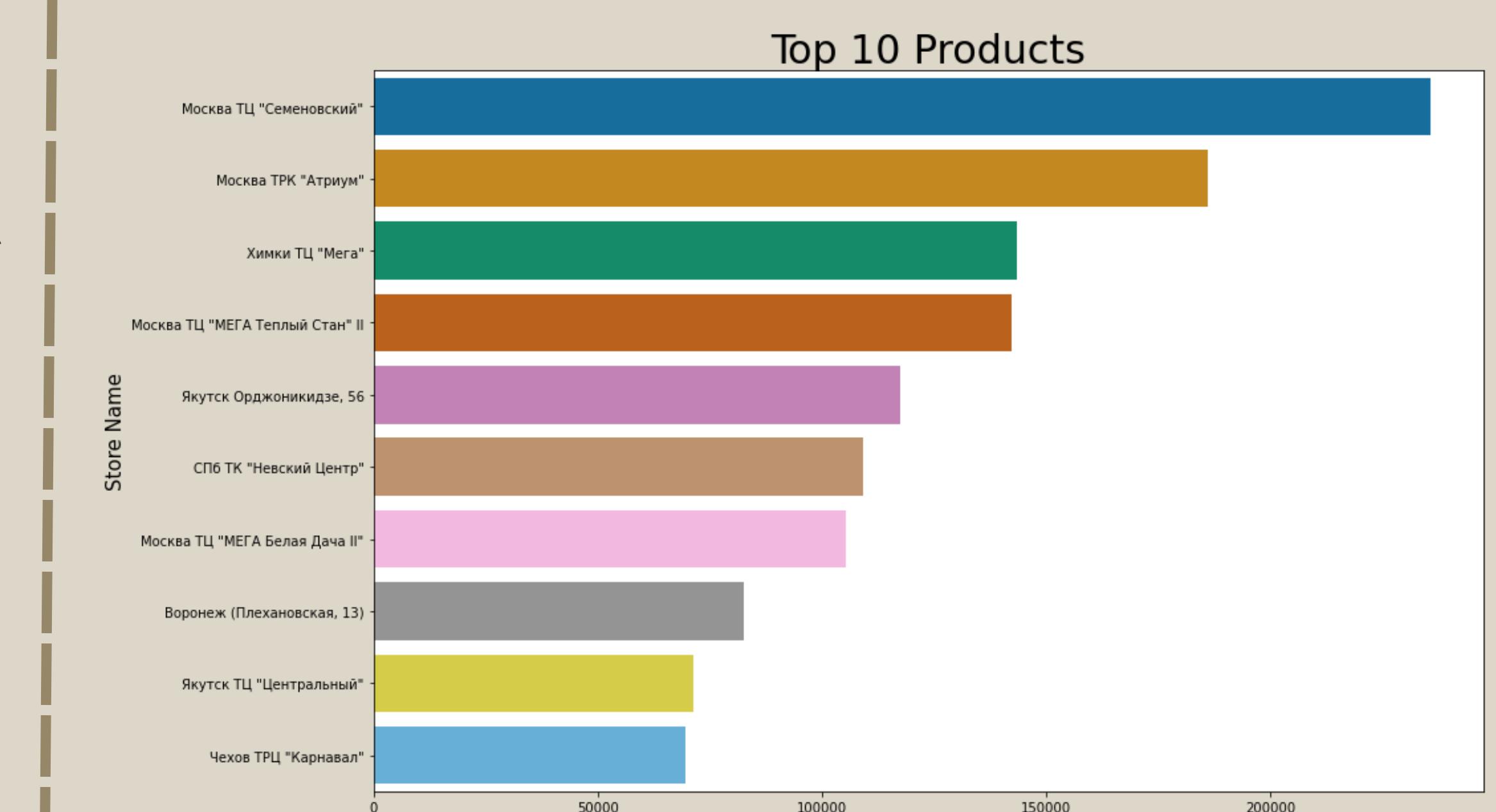
MODEL1: XGBOOST

XGBoost is a gradient boosting algorithm that is known for its accuracy and speed, making it an ideal choice for demand forecasting tasks.

By using the XGBoost algorithm for demand forecasting, we can leverage the power of advanced machine learning techniques to make accurate predictions about future demand.

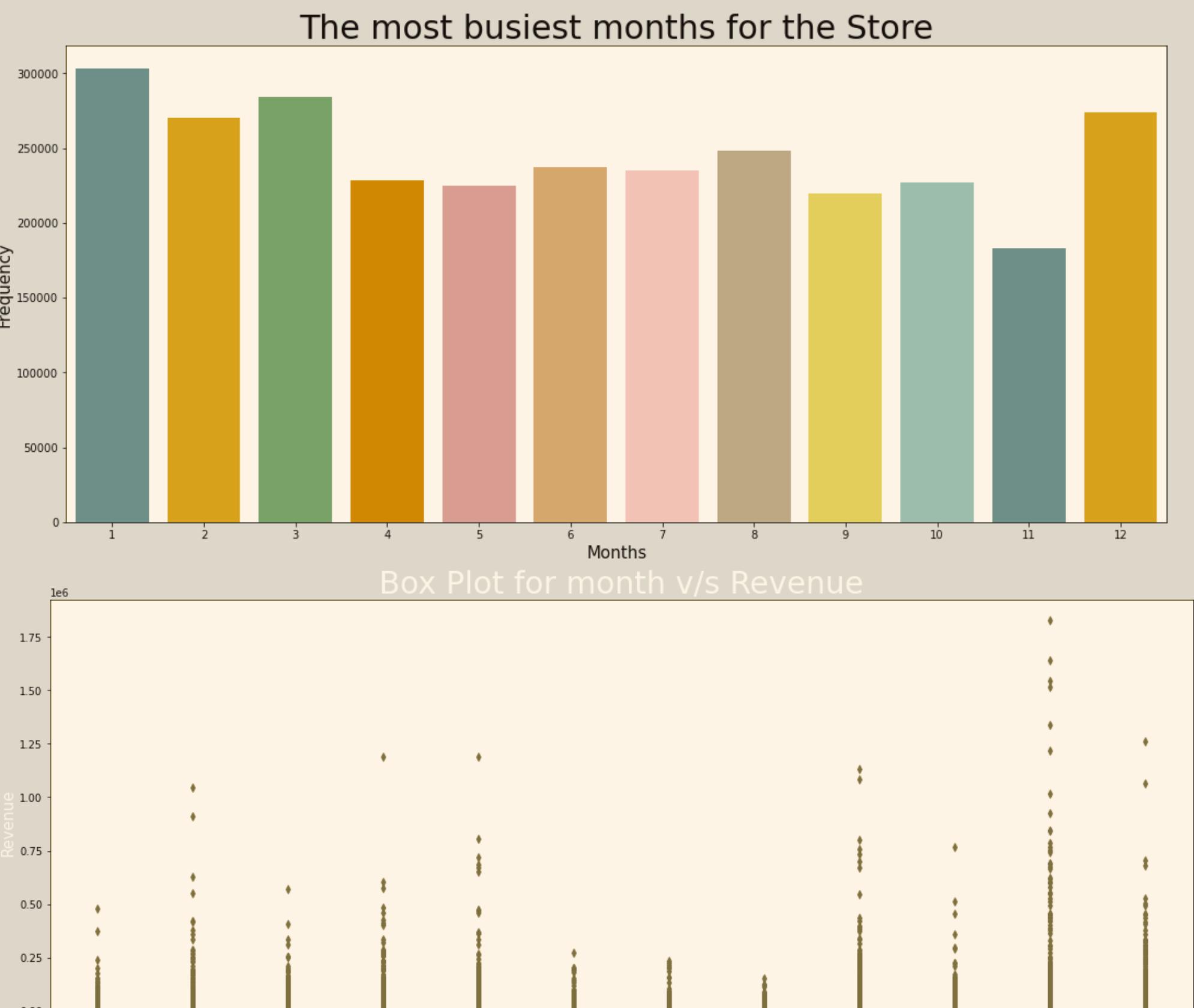
MODEL 2: LIGHT GBM

The light gradient boosting machine algorithm – also known as LGBM or LightGBM – is an open-source technique created by Microsoft for machine learning tasks like classification and regression. It is quite similar to XGBoost as it too uses decision trees to classify data.



The plot shows the distribution of year wise foot fall across all the stores. It is observed that the busiest year was 2013 and the count has decreased over the years.

The plot shows the distribution of month wise foot fall across all the stores. It is observed that the busiest month was January and the least busiest was November. The count fluctuates for the remaining months.



The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.



The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the boxplot better, this distribution plot depicts the revenue share. It is clear that the price of items are widely distributed from \$0-5000 and very few above \$20000

The boxplot here describes how the revenue is spread across months. As the y-limit is very large, the box seems like a line which is mostly around 0-5000\$. Also, there are months with very high revenue like in November.

To visualize the box