Analytical Marketing
Homework 3

Q1: Price Elasticities and Optimal Prices compared with the actual prices for Products 1, 4, 5, and 11 for week #93 of store #2:

Product	(a) Price Elasticity	(b) Optimal Price	(c) Actual Price
Product 1	-2.686801	3.272483	3.59
Product 4	-3.894563	2.058718	2.49
Product 5	-3.377983	2.206889	2.39
Product 11	-3.375459	1.445179	1.29

Q1c: The optimal prices are lower for products 1,4 and 5 as compared to actual prices. On the contrary, the optimal price for product 11 is higher than the actual price which shows that product 11 is not profitable.

Q2a: Pooled Model

Calculation	Estimation	Hold-Out Sample
Mean Absolute Error	0.47	0.47
Mean-error	0.00015	0.00015
Root Mean-Squared Error	0.61	0.61
Mean-Absolute Deviation	0.47	0.47

The errors are low but not as low as the other models, thus we can state that the pooled approach might not capture the complexities in the data as effectively.

Q2b: Store Model

Calculation	Estimation	Hold-Out Sample
Mean Absolute Error	0.28	0.31
Mean-error	-0.0015	-0.00015
Root Mean-Squared Error	0.4	0.4
Mean-Absolute Deviation	0.29	0.29

MAE, RMSE and MAD are relatively lower for the store model than the pooled model, hence implying that the store-specific model might be performing well if it is tailored to capture variations specific to each store but might not be sufficient for achieving the best accuracy.

Q2c: Mixed Model

Calculation	Estimation	Hold-Out Sample
Mean Absolute Error	0.01	-0.03
Mean-error	-0.0059	-0.01
Root Mean-Squared Error	1.51	1.51
Mean-Absolute Deviation	1.23	1.23

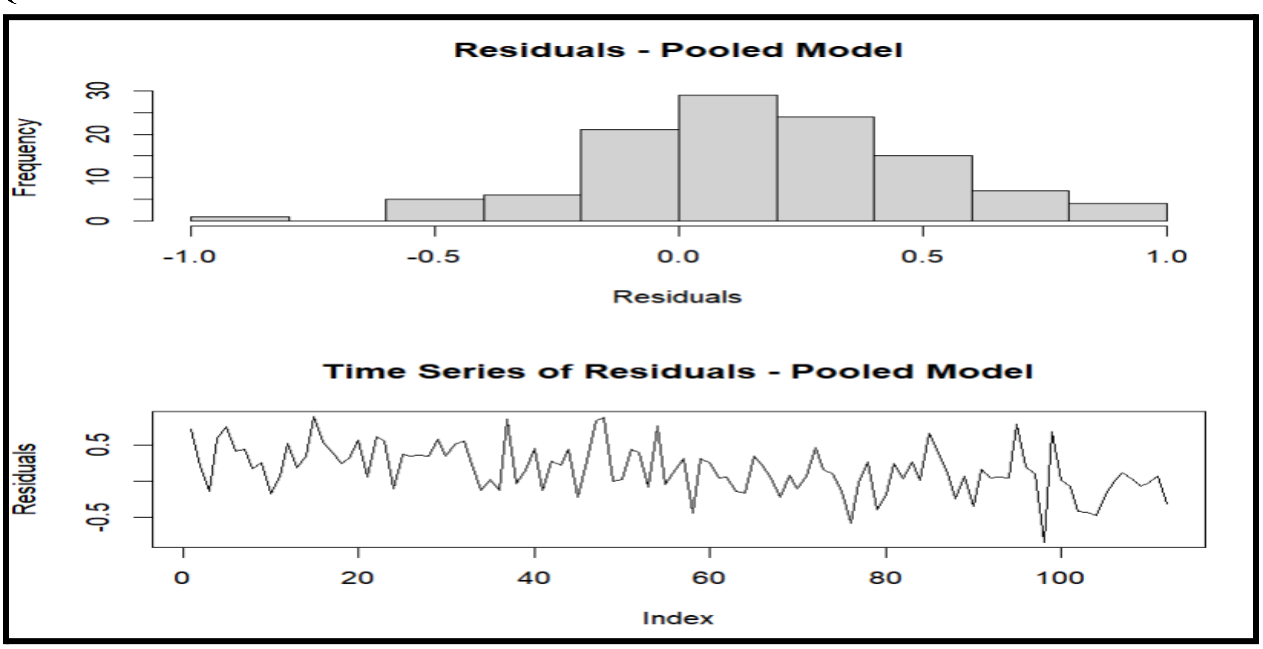
The Mixed model combines elements of both the pooled and store-specific approaches. Its accuracy depends on how well it balances global patterns and store-specific nuances but as we can see that the RMSE and MAD are higher than the individual models, therefore we can say that the accuracy is not good.

Q2d: Random Forests

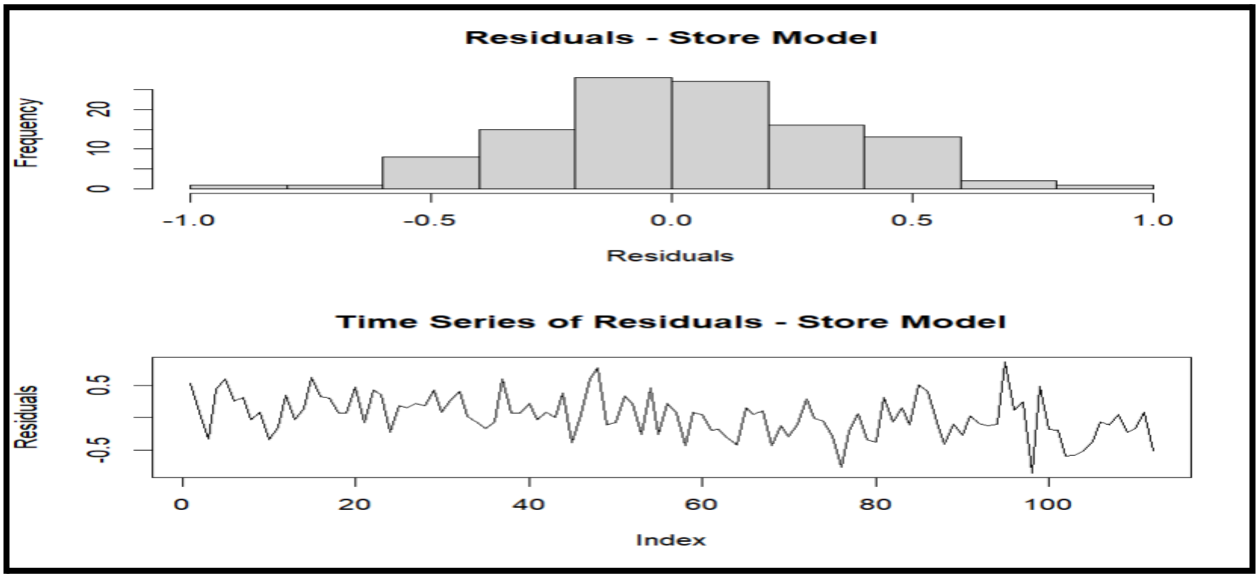
Calculation	Estimation	Hold-Out Sample
Mean Absolute Error	0.15	0.32
Mean-error	0.002	0.01
Root Mean-Squared Error	0.30	0.30
Mean-Absolute Deviation	0.22	0.22

The Random Forest model demonstrates the best accuracy among the models, as evidenced by its consistently lower MAE, RMSE, and MAD. This suggests that the model is effective in capturing the underlying patterns in the data, resulting in more accurate predictions.

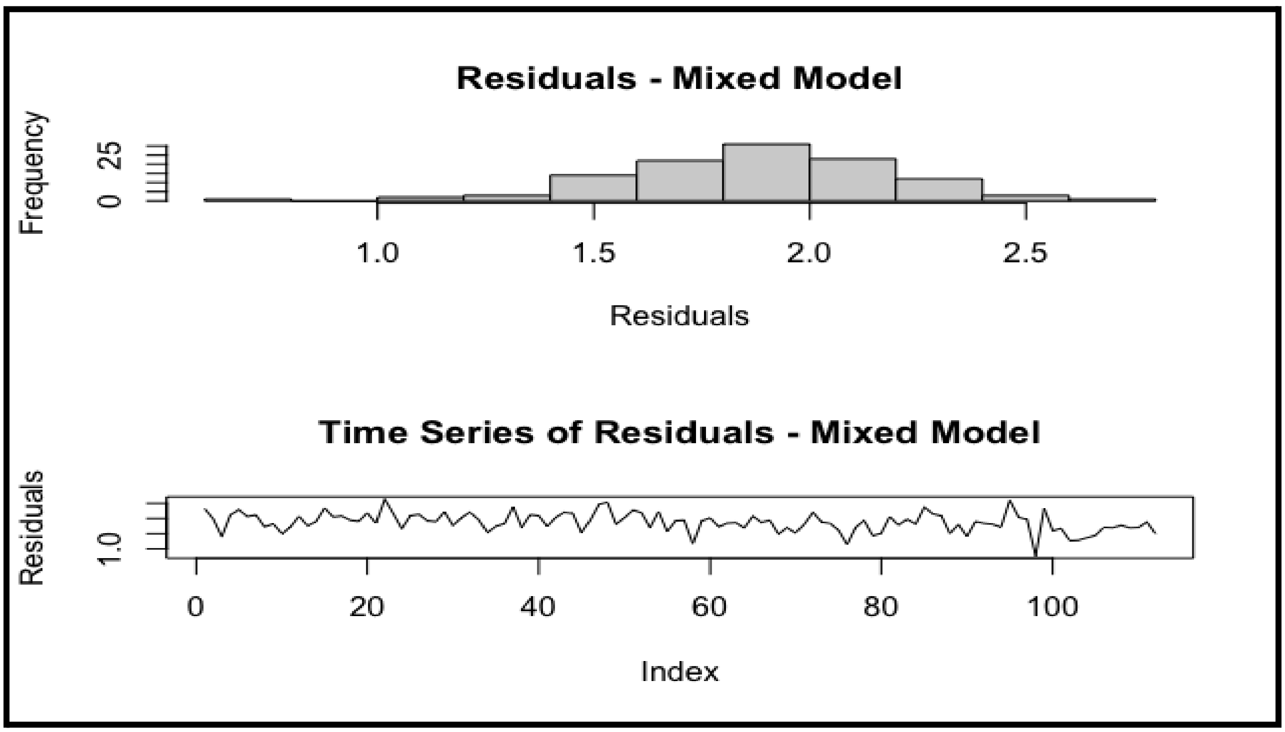
Q3:



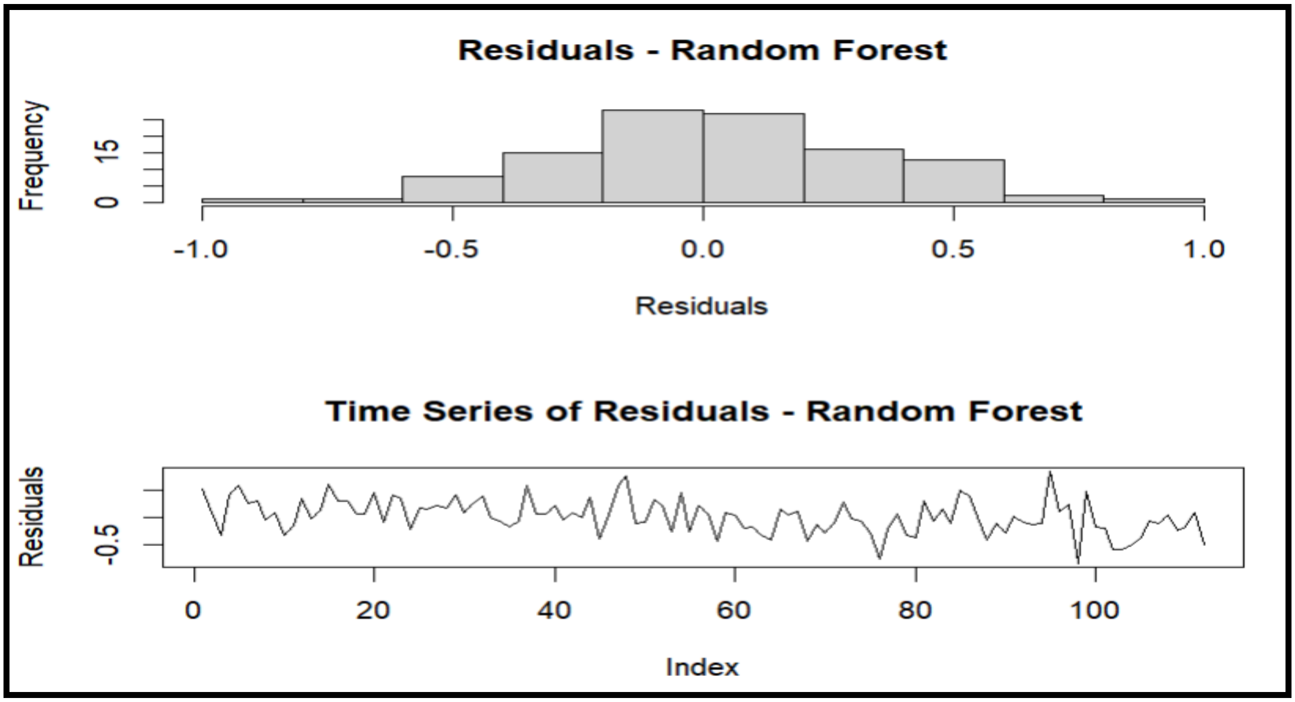
It is not normally distributed, it is slightly skewed towards the right. Residuals do not align well with the assumptions made during model construction.



The histogram is widely spread as more residual is distributed around the mean. The model fails to capture recurring patterns as seen from the time series graph, thus we can say that this model is not accurate.



Histogram is not at all normally distributed, it is skewed towards the right. it deviates significantly, indicating issues with the assumption of normally distributed errors. Cyclic patterns are absent in the model residual of the time series graph, indicating that the model is not very accurate as compared to the random forest model.



The histogram is normally distributed, indicating that the residuals are symmetrically distributed around zero and recurring patterns can be seen within the time series graph.

Q4: We experimented with different models, validated their performance, and chose the one that aligns best with the business objectives. Considering the complexity of pricing dynamics, potential variations across stores, and the desire to understand the impact of a pricing strategy shift, **a Random Forest Mode**l would be the best choice in this scenario because of the following reasons:

- Features Considered: We tried multiple combination of features and found that model performance is significantly better when we consider all features related to product 1 along with related store prices which are - lpr1, lpr4, lpr5, lpr11, afeat1,adisp1, profm1. Having the correct features for this model helped random forest to learn the underlying patterns in the data and is able to predict more accurately.
- Hyper Parameter Tuning: We selected the number of trees here to 300 by setting the hyperparameter ntree = 300 while training the algorithm.
- Ensemble Learning: The most important feature of RF model's is that it is based on ensemble learning, which makes it less prone to overfitting and increases robustness. In this scenario, since we did not have enough data, there are high chances that the model can overfit but random forest rules that out. They can provide stable and reliable predictions, especially when there are multiple factors influencing pricing decisions.
- Complex and Non linear Relationships: Random Forest models are effective in capturing complex, non linear relationships within the data, making them suitable for scenarios with intricate pricing dynamics. Hence, we decided to move forward with Random forest as it can capture these patterns better than linear models.

The random forest model makes sense in a way that as economists and marketers believe that if price increases then quantity of the product drops. The model fits well in this scenario as we checked how the model impacts the quantity of the product by increasing and decreasing the price. Currently, the quantity of product 1 is 518597.4. If the price drops from \$3.8 to \$2.0, the quantity increases to 924905.7 and if the price increases to \$10.0 then the quantity drops to 464538.2. This accurately aligns with the Law of Demand.

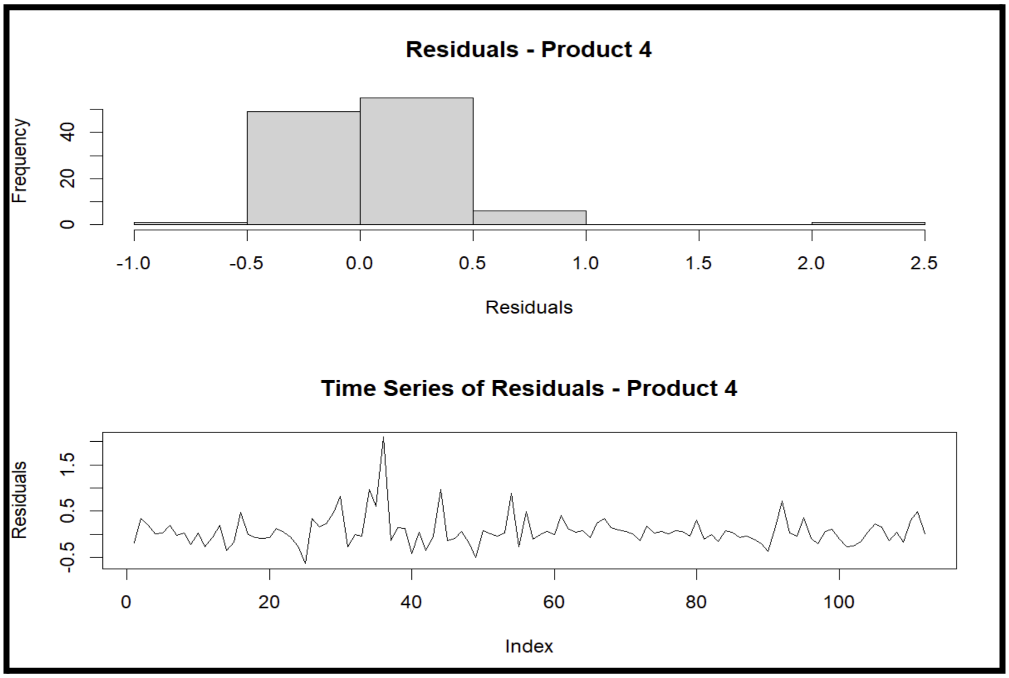
Additionally, as we know that if products are substitutes then if the price of a competing product goes up then the quantity of our product should go up. Our price is currently $3.89 and the number of products are 528383.7 and our competitor's (product 5) current price is $3.17. If the price of the competing good increased to $5.279701, our number of products also increased to 541919.5, thus holding the relationship between the price of substitute goods and quantity of goods true.

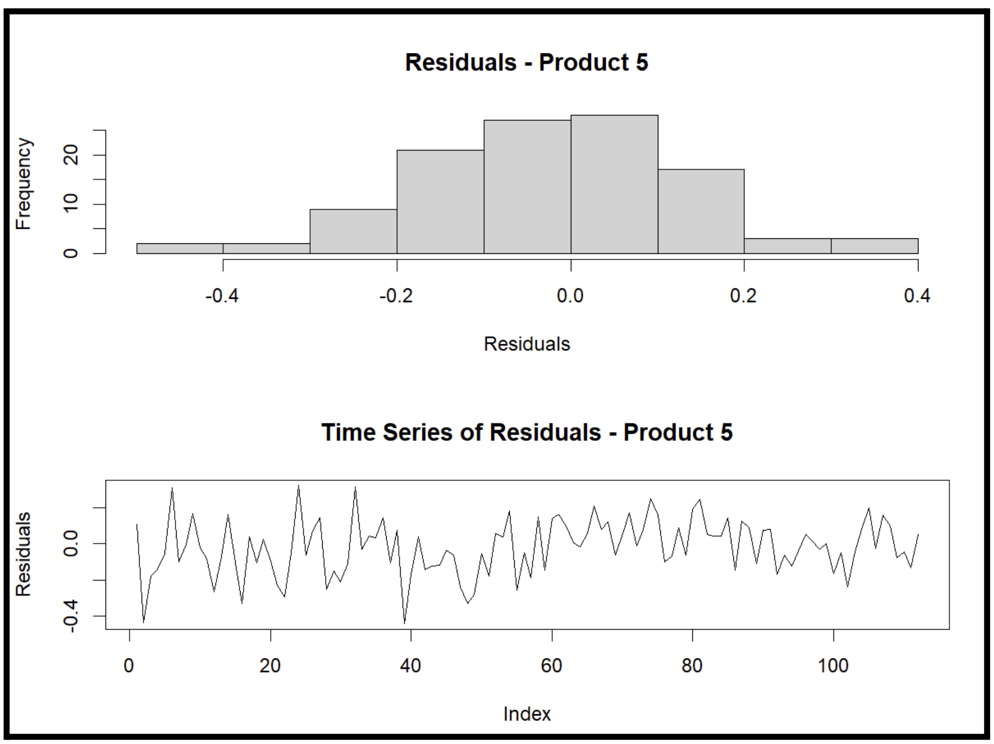
Q5: Optimal price for **Pooled Model is $4.091276** and optimal price for **Random Forest is $3.12.**
For Store model and Mixed model below is the table, below is the table that contains optimal values for specific stores:

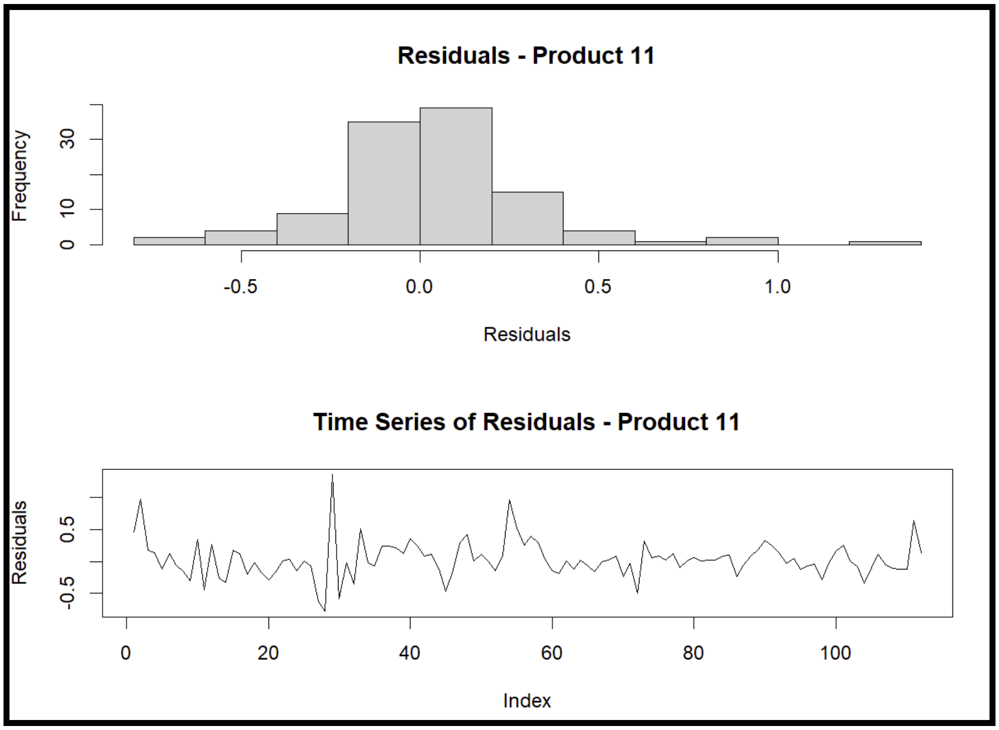
Model	**Store 2**	**Store 5**	**Store 8**	**Store 9**	**Store 12**
Store Model	3.844754	1.231216	6.200836	4.41574	3.242608
Mixed Model	1.910168	3.109026	3.132068	3.484173	3.72396

Q6: **Random Forest Model is Better than Cost-Plus Rule because of the following reasons:**
- Flexibility and Adaptability: The cost-plus rule often relies on fixed margins or rules of thumb, which may not adapt well to changing market conditions. The random forest model, on the other hand, can adapt to evolving patterns and incorporate new information, providing a more dynamic pricing strategy.
- Consideration of Multivariate Factors: The random forest model considers a multitude of factors simultaneously, including customer behavior, market trends, and various product attributes. This multivariate approach allows for a more nuanced understanding of pricing dynamics compared to a simplistic cost-plus approach.
- Optimization for Profitability: The random forest model can be explicitly designed to optimize for profitability, considering both costs and revenue factors. This contrasts with cost-plus rules that may not necessarily maximize profit but rather ensure a predefined margin.
- Dynamic Pricing Adjustments: Random forest models can adapt to changing circumstances in real-time, allowing for dynamic adjustments in pricing based on factors like demand fluctuations, competitor pricing changes, or supply chain disruptions. This adaptability is a key advantage over static cost-plus pricing.
- Data-Driven Decision-Making: Emphasize the importance of data-driven decision-making. The random forest model leverages historical data and patterns to make informed pricing decisions, enabling a more systematic and evidence-based approach compared to manual rule-setting.
- Competitive Edge: A sophisticated pricing model like random forest can provide a competitive edge by allowing the business to respond quickly to market changes and optimize pricing strategies to stay ahead of competitors relying on traditional pricing rules.

Q7)


Residuals - Product 4
Frequency
0
20
40
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
Residuals
Time Series of Residuals - Product 4
Residuals
-0.5
0.5
1.5
0
20
40
60
80
100
Index


Residuals - Product 5
Frequency
0
10
20
-0.4
-0.2
0.0
0.2
0.4
Residuals
Time Series of Residuals - Product 5
Residuals
-0.4
0.0
0
20
40
60
80
100
Index



The analysis of product sales using the Random Forest model has yielded insightful results, however, a closer examination of the residual plots for products 4, 5, and 11 suggests that the model fit is not ideal for all cases. Specifically, the residual distribution for Product 4 exhibits a leftward skew, indicating that the model may be overestimating the frequency of lower sales volumes. Conversely, Product 5's residuals show a rightward skew, implying an underestimation of higher sales volumes. Product 11's residuals also present a left skew, similar to Product 4, suggesting an overestimation of lower sales volumes.

Such skewness in the residuals indicates that the Random Forest model may not be capturing all the underlying patterns effectively for these particular products. While the model has shown promise in other scenarios, its applicability seems to vary on a case-by-case basis. It is essential to assess each product's data individually to determine the most suitable model. This nuanced approach to model selection ensures that the distinct characteristics of each product's sales data are accurately represented, leading to more reliable and actionable insights.

Q8)

Product	**Optimal Price**	**Cost**	**Optimal Profit**
1	3.272483	2.054498	1.217185
4	2.058718	1.530105	0.528613
5	2.206889	1.553573	0.653316
11	1.445179	1.017036	0.428143

In optimizing total profits for a product category, the calculated optimal prices and corresponding profits for products 1, 4, 5, and 11 illustrate a targeted pricing strategy aimed at maximizing the margins for individual items within the category. The aggregation of optimal profits for these selected products results in a total category profit of 2.827257, indicative of a strategy where each product is methodically priced to effectively balance cost against consumer demand. This methodology, while grounded in theoretical accuracy, presupposes independent demand for each item without accounting for potential cross-elasticity or the overall impact on consumer purchasing behaviors across the category. The provided data sets a solid foundation for immediate profit maximization strategies but also emphasizes the necessity for dynamic market analysis. This includes considerations around competitor pricing, market positioning, and customer perception, necessitating adaptive and refined pricing strategies over time. The analysis brings to light the challenge of optimizing individual product profits while managing the category's broader market performance, underscoring that achieving and maintaining profitability in a competitive environment demands ongoing monitoring and strategic adjustments.

Q9)

As Dominick's pricing managers, our decision to set the price of Product 11 (Dominick's 64 oz) needs to account for our own cost structure, the optimal profit margin, and the competitive pricing landscape.

1. Premium Brand Benchmark: The highest optimal price in the category is $3.272 for Product 1 (Tropicana Premium). This sets a benchmark for the premium end of the market.

2. Mid-tier Pricing Context: Products 4 (Tropicana) and 5 (Minute Maid), priced at $2.058 and $2.206 respectively, represent the mid-tier options. Their pricing helps to define the market expectations for quality and price.

3. Store Brand Strategy: As a store brand, Product 11 should be an affordable alternative to national brands. We aim to offer value to our customers while maintaining strong perceptions of quality.

Given that the highest optimal price is for the premium product, we would price Product 11 in a way that offers a clear value proposition. The cost for Product 11 is $1.017, with an optimal profit margin of $0.428 per unit at the suggested price.

We propose setting the price for Product 11 at **$1.99**. This strategy positions Dominick's 64 oz significantly lower than mid-tier options, solidifying our competitive advantage. Although it is less than the highest optimal price, it allows us to achieve a substantial profit margin. Furthermore, the $1.99 price leverages the psychological impact of being just under the **$2 threshold**, which can influence consumer purchase decisions.

We will closely monitor the market's response to this price point, ready to make adjustments if there's an opportunity to increase it without negatively impacting demand. Regularly reviewing this price against market and competitor changes, as well as our internal cost adjustments, will ensure it remains effective and relevant.