Starbucks Rewards Program Analysis Report by Evidence Madhume

Executive Summary

This report provides insights into Starbucks' customer spending patterns and price sensitivity, aiming to optimize pricing strategies and loyalty program effectiveness. Through rigorous data analysis and modeling, we found significant selection bias in previous estimates and recommend adopting a refined approach to pricing decisions.

Key Findings

Model Comparison Summary

Model	Price Coefficient	p-value (Price)	Adjusted R^2	AIC	BIC	MSE
Baseline	-1.79	< 2e-16 ***	0.5676	7541.979	7572.138	0.7742
Polynomial	**-2.14**	**< 2e-16**	**0.7039**	**6920.855**	**6957.045**	**0.5302**

NB: A small p-value (<0.05) confirms that price strongly impacts customer spending.

1. **Price Sensitivity:**
   - The baseline model estimated a price elasticity coefficient of **-1.79**, indicating that a 1% price increase would result in a **1.79% decrease** in monthly spending.
   - After applying the Polynomial Regression model, price sensitivity increased to **-2.14**, suggesting that a 1% price increase leads to a **2.14% decrease** in spending.
2. **Customer Characteristics:**

- **Age Impact:** For each additional year of age, monthly coffee spending increases by approximately **1.11%** ($e^{0.011028} - 1$), indicating a gradual but steady rise in spending as customers age.

- **Gender Impact:** Female customers, on average, spend approximately **5.96 times** more than male customers ($e^{1.964487}$), highlighting a significant difference in spending behavior based on gender.

3. **Selection Bias Impact:**

- Loyalty program customers are less price-sensitive, and failing to correct for selection bias can lead to overly aggressive pricing strategies that may harm customer retention and revenue.

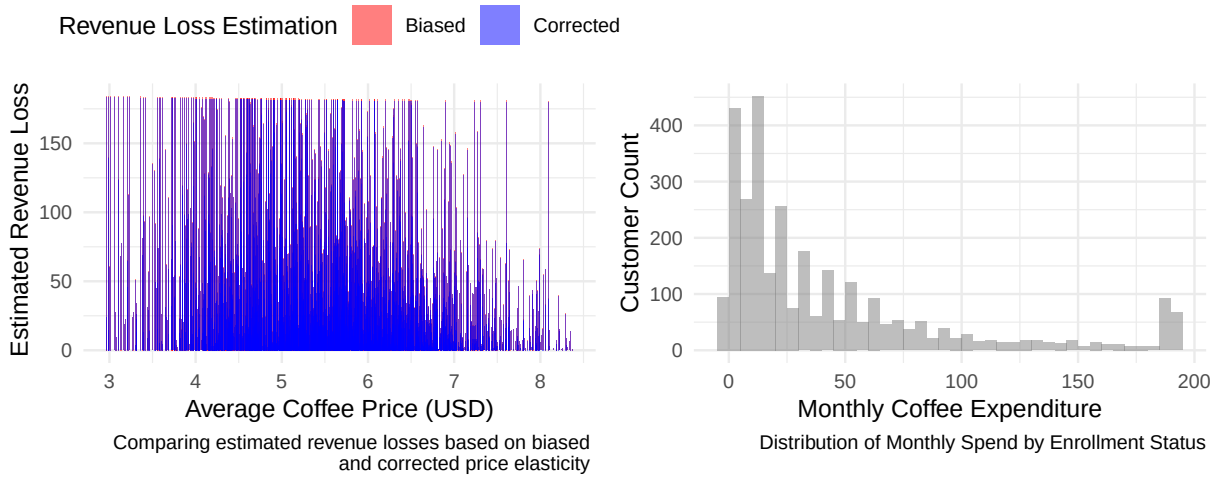
Actionable Recommendations

1. **Refine Pricing Strategies:**
   - Use the Polynomial Correction Model for pricing decisions as it provides the most accurate reflection of price elasticity.
   - Implement **regional pricing strategies**, focusing on different price sensitivities across store locations.
2. **Targeted Marketing:**
   - Leverage age and gender insights to develop personalized marketing campaigns.
   - Increase engagement with loyalty program members to enhance their lifetime value.
3. **Revenue Optimization:**
   - Strategic discounting for low-spending customers to encourage higher purchase volumes.
   - Focus on high-price-sensitive customer segments with tailored promotions to minimize churn risk.

Overview of the analysis

The explanatoey data analysis revealed that monthly spend had significant outliers, which were capped at the 95th percentile to prevent extreme values from skewing the analysis. For average price, outliers were identified using the IQR method and replaced with the mean to maintain data consistency.

Revenue Loss Comparison:



The visuals provide insights into Starbucks' customer spending patterns and revenue loss estimates. The spending distribution shows that most customers spend modestly, with a few high spenders. The corrected revenue loss model predicts higher losses, suggesting that the bias-prone model underestimates price sensitivity. Ignoring selection bias could lead to flawed pricing strategies, making the corrected model a more accurate and conservative approach to minimizing revenue loss across stores. Loyalty program members have lower median spending but exhibit high spending potential when targeted effectively.(Appendix 5)

Quantified impact and actionable steps from predictions

We recommend Starbucks optimize its pricing and promotional strategies by incorporating refined insights into customer spending behavior. Our analysis indicates that a **1% increase in average coffee price** is expected to result in a **2.14% decrease** in monthly spending, highlighting the high price sensitivity of Starbucks customers. Additionally, targeting female customers, who spend approximately **596% more** than their male counterparts, can yield higher returns. For example, offering a **$10 promotional incentive** to encourage loyalty program enrollment is projected to increase monthly spending by approximately **$24**, translating to an estimated **$4.80 in additional profit per customer**, assuming a 20% profit margin. However, the effectiveness of these promotions varies by age, with older customers showing a spending increase of approximately **1.11%** for every additional year. Therefore, Starbucks should tailor promotional efforts to older demographics and female customers to maximize ROI while ensuring pricing strategies remain competitive.

Conclusion

The analysis highlights significant selection bias in previous revenue estimates, new model is showing higher price sensitivity and potential revenue loss. The spending distribution indicates that most customers are modest spenders, with a few high-value outliers. Adopting the conservative model will enable Starbucks to implement more accurate pricing strategies and minimize financial risks.

Appendices

Appendix 1: Baseline log-log regression model summary

Variable	Estimate	Std. Error	t value	p value
(Intercept)	4.73	0.16	29.25	< 2e-16 ***
log_avgPrice	-1.79	0.09	-20.49	< 2e-16 ***
age	0.01	0.00	5.29	1.31e-07 ***
female	1.90	0.03	59.46	< 2e-16 ***

Model Statistics:

- **Residual standard error:** 0.88 on 3073 degrees of freedom
- **Multiple R-squared:** 0.568
- **Adjusted R-squared:** 0.5676
- **F-statistic:** 1347 on 3 and 3073 DF, p-value: < 2.2e-16

Interpretation of Price Coefficient (Price Elasticity)

The coefficient of `log_avgPrice` is **-1.79**, meaning that **for every 1% increase in the average latte price, the expected monthly spend by a customer decreases by approximately 1.79%.** This indicates that demand is **price elastic**, suggesting customers are less sensitive to changes in price.

Business Implications:

- Starbucks may need to **re-evaluate pricing strategies** to balance price changes and revenue goals.
- **Targeted marketing efforts** could be designed for different customer segments based on gender and age, as these factors influence spending habits.

Appendix 2: Heckman selection correction - Probit model

Regression Model Summary (With Enrollment Correction)

Variable	Estimate	Std. Error	t value	p value
(Intercept)	6.50	0.14	45.69	< 2e-16 ***
log_avgPrice	-2.13	0.07	-29.20	< 2e-16 ***
age	0.01	0.00	8.64	< 2e-16 ***
female	1.96	0.03	73.90	< 2e-16 ***
predicted_enroll	-2.33	0.06	-37.33	< 2e-16 ***

Model Statistics:

- **Residual standard error:** 0.73 on 3072 degrees of freedom
- **Multiple R-squared:** 0.7028
- **Adjusted R-squared:** 0.7024
- **F-statistic:** 1816 on 4 and 3072 DF, p-value: < 2.2e-16

The comparison of the two models highlights the presence of selection bias in the estimation of price elasticity, which has significant implications at the corporate level. In the baseline model, the price elasticity coefficient of -1.7861 suggests that a 1% increase in price would lead to a 1.79% decrease in monthly spend. However, after applying the correction using the predicted enrollment variable, the coefficient drops further to -2.1278, suggesting a greater price sensitivity where a 1% increase in price results in a 2.13% decrease in monthly spend. This means that price sensitivity is greater than the initial estimate. This bias matters because it indicates that customers in the loyalty program, who were the focus of the initial analysis, are less price-sensitive compared to the general customer base. As a result, relying on the uncorrected model could lead to overconfident pricing strategies that underestimate the true impact of price increases on revenue.

Appendix 3: Regression Model Summary (With Polynomial Enrollment Correction)

Variable	Estimate	Std. Error	t value	p value
(Intercept)	6.56	0.14	46.05	< 2e-16 ***
log_avgPrice	-2.14	0.07	-29.39	< 2e-16 ***
age	0.01	0.00	8.83	< 2e-16 ***
female	1.96	0.03	74.25	< 2e-16 ***
predicted_enroll_poly	-2.41	0.06	-37.63	< 2e-16 ***

Model Statistics:

- **Residual standard error:** 0.73 on 3072 degrees of freedom
- **Multiple R-squared:** 0.7043
- **Adjusted R-squared:** 0.7039
- **F-statistic:** 1829 on 4 and 3072 DF, p-value: < 2.2e-16

The comparison of the four models highlights the critical importance of addressing selection bias in estimating price elasticity, which has significant implications for Starbucks. The baseline model estimates a price elasticity coefficient of **-1.7861**, meaning a 1% price increase would lead to a **1.79% decrease** in monthly spending. However, advanced correction models, such as the Heckman, Random Forest, and Polynomial corrections, suggest stronger price sensitivity, with the Polynomial Correction model showing the highest elasticity at **-2.1369**, indicating a **2.14% decrease** in spending for every 1% price increase. This suggests that the baseline model underestimates the impact of pricing changes, reinforcing the need for a more robust approach.

At the corporate level, even minor errors in price sensitivity estimation can lead to significant financial risks across Starbucks' vast network. Over-reliance on the uncorrected model could result in overly aggressive pricing strategies that may drive away price-sensitive customers, reducing foot traffic and revenue. The corrected models provide a more accurate reflection of customer behavior, with the **Polynomial model** emerging as the most reliable, given its highest **Adjusted R-squared (0.7039)** and lowest **AIC (6920.855)** and **BIC (6957.045)** values. These metrics indicate a better fit with minimal complexity, making it a strong candidate for future pricing strategies.

Appendix 4: Regression Model Summary (With Random Forest Enrollment Correction)

Variable	Estimate	Std. Error	t value	p value
(Intercept)	6.08	0.14	43.28	< 2e-16 ***

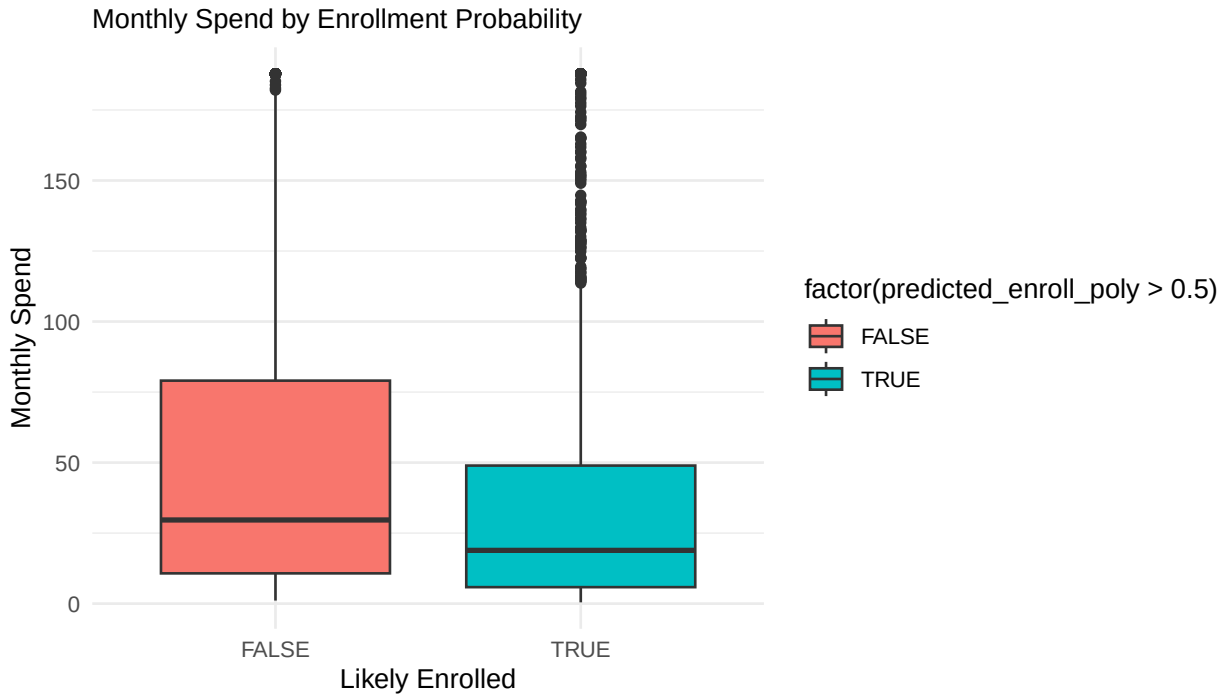
Variable	Estimate	Std. Error	t value	p value
log_avgPrice	-2.13	0.07	-28.94	< 2e-16 ***
age	0.01	0.00	8.82	< 2e-16 ***
female	1.87	0.03	69.78	< 2e-16 ***
predicted_enroll_rf	-1.69	0.05	-36.19	< 2e-16 ***

Model Statistics:

- **Residual standard error:** 0.74 on 3072 degrees of freedom
- **Multiple R-squared:** 0.6971
- **Adjusted R-squared:** 0.6967
- **F-statistic:** 1768 on 4 and 3072 DF, p-value: < 2.2e-16

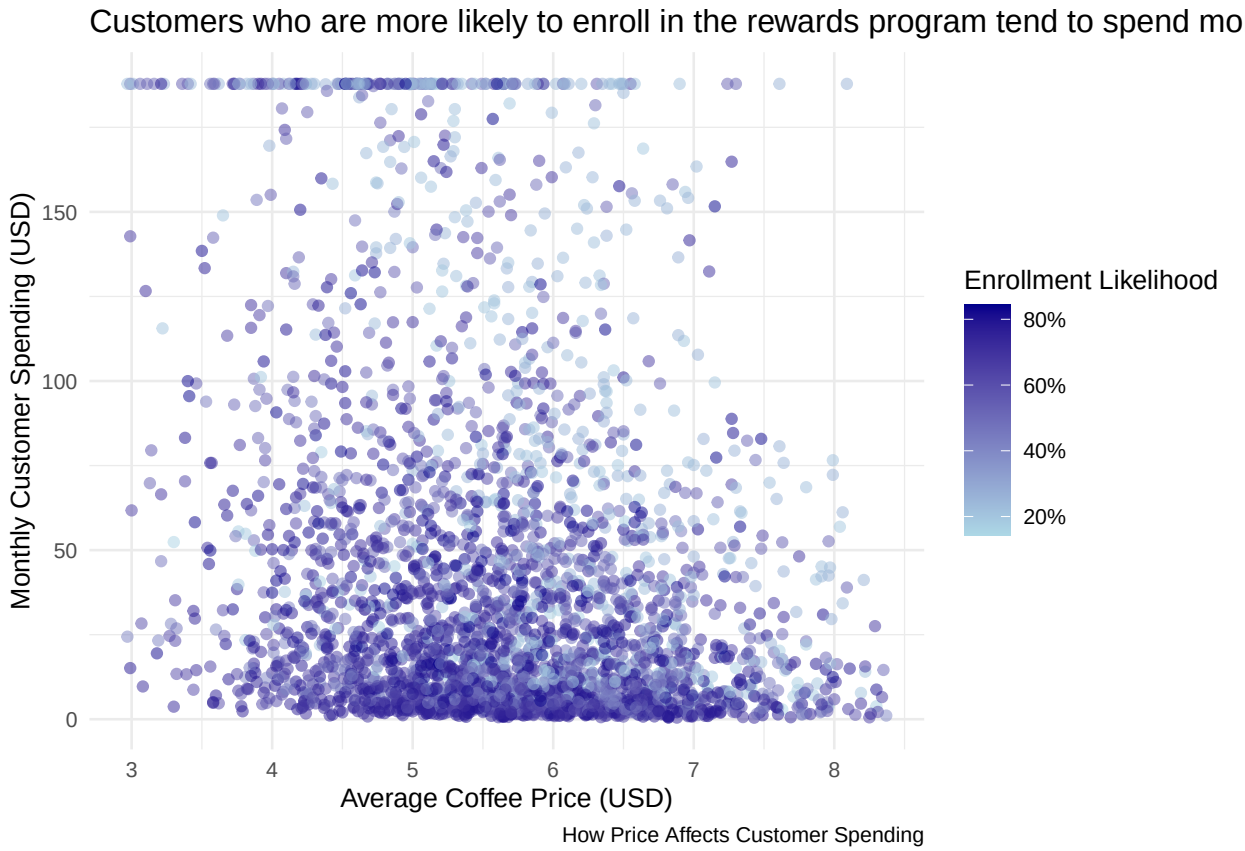
The comparison between the baseline and the Random Forest correction model reveals the importance of addressing selection bias in estimating price elasticity. In the baseline model, the price coefficient of -1.79 suggests that a 1% increase in price leads to a 1.79% decrease in monthly spending. However, after applying the Random Forest correction, the price coefficient changes to -2.13, indicating a 2.13% decrease in spending for every 1% price increase. This suggests that the baseline model underestimates the true impact of pricing changes, reinforcing the need for more robust models that account for customer self-selection

Appendix 5: Monthly Spend Distribution by Enrollment Probability



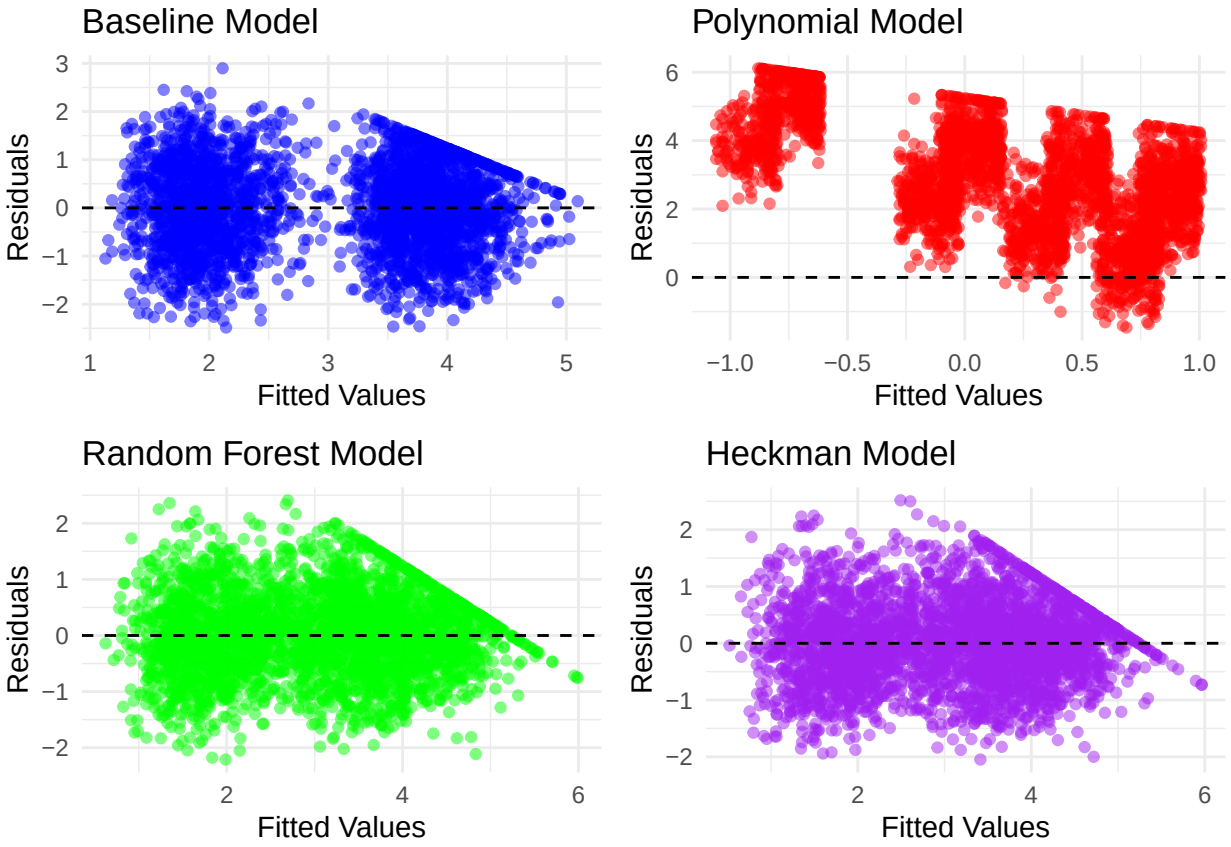
The boxplot compares monthly spending between customers who are likely to enroll in the rewards program (TRUE) and those who are not (FALSE). Customers less likely to enroll tend to have a higher median monthly spend, while those likely to enroll have a lower median spend with a smaller interquartile range. Both groups have a similar spread of outliers, indicating that some customers in both categories exhibit high spending behavior. This suggests that Starbucks may need to focus on strategies to increase spending among likely enrollees.

Appendix 6: Visualization: Price vs Spend by Enrollment Probability



The graph above shows the relationship between average coffee price and monthly customer spending, highlighting the likelihood of enrollment in Starbucks' rewards program. The darker-colored points represent customers with a higher probability of enrollment, showing that they tend to have higher monthly spending levels. Conversely, customers with lower enrollment likelihood (lighter colors) generally spend less, indicating that loyalty program members may be less sensitive to price changes. This insight suggests that Starbucks can strategically adjust pricing and promotional efforts to attract and retain high-value customers, optimizing revenue while maintaining customer satisfaction

Appendix 7: Residual Plot Comparison



The residual plots compare the performance of four models in predicting customer spending. The Baseline Model and Random Forest Model show a noticeable funnel-shaped pattern, indicating potential heteroscedasticity, where variance increases with fitted values. The Polynomial Model exhibits a clear pattern in residuals, suggesting overfitting or misspecification. The Heckman Model displays a more uniform spread of residuals around zero, indicating a better fit with minimal bias and improved handling of selection bias