Contents

[Summary 2](#_Toc207160486)

[A. Explore the dataset and conduct any EDA that you feel will help you understand. 3](#_Toc207160487)

[a. Dataset Overview 3](#_Toc207160488)

[b. Missing Values Treatment (EDA: Table 1) 3](#_Toc207160489)

[c. Visualizations (EDA) 3](#_Toc207160490)

[1. Policy Acceptance Distribution - Pie Chart 3](#_Toc207160491)

[2. Offered Premium Distribution – Histogram 3](#_Toc207160492)

[3. Policy Sales Distribution by Pricing Strategy - Grouped Bar Chart 3](#_Toc207160493)

[4. Correlation Heatmap 3](#_Toc207160494)

[d. Descriptive Statistics (EDA: Table 2) 4](#_Toc207160495)

[B. Compare the pricing performance of all strategies in a summary table with: 4](#_Toc207160496)

[a. Summary Table (Pricing Performance Summary - Table 1) 4](#_Toc207160497)

[b. Visual Insights (Pricing Performance Summary) 4](#_Toc207160498)

[c. How would you test (statistically) if they are different? 5](#_Toc207160499)

[1. Do conversion rates differ? (Pricing Performance Test - Conversion Rate Differences) 5](#_Toc207160500)

[2. Do Average Premiums Differ? (Pricing Performance Test - Offered Premium Differences) 5](#_Toc207160501)

[3. Do revenues (sold premiums) differ? (Pricing Performance Test - Revenue Differences) 5](#_Toc207160502)

[C. Compare price elasticity of the strategies in the dataset: 5](#_Toc207160503)

[a. Price Elasticity Across Group (Price Elasticity – Table 1) 5](#_Toc207160504)

[b. Price Elasticity - Withing Group. 6](#_Toc207160505)

[1. Customer Related (Price Elasticity – Table 2) 6](#_Toc207160506)

[2. Appliance Related (Price Elasticity – Table 3) 6](#_Toc207160507)

[D. Assuming each pricing point is generated by a model at a different global constraint, would you be able to indicate: 7](#_Toc207160508)

[a. Which one is the best model from the statistical point of view? (Pricing Model – Table 1) 7](#_Toc207160509)

[b. And from pricing point of view(revenue)? (Pricing Model – Table 2) 7](#_Toc207160510)

[c. Would you say these models are biased? (Pricing Model – Table 3) 8](#_Toc207160511)

[E. Claims Information are also provided, is there any direct relationship between conversion and claims? 8](#_Toc207160512)

[a. Loss Ratio on the Sold Cohort (Claims & Conversions – Table 1) 8](#_Toc207160513)

[b. Direct Correlation of Claims with Conversion (Claims & Conversions – Table 2) 8](#_Toc207160514)

[c. Profit/Loss Value on Average per Offer (Claims & Conversions – Table 3) 9](#_Toc207160515)

[F. Final Recommendation based on findings? 9](#_Toc207160516)

# Summary

This report evaluates D+G’s pricing strategies (@22%, @23%, and ASIS FEE) using customer, premium, and claims data to identify the most sustainable approach balancing revenue, profitability, and risk.

**Key Findings**

* **Conversions:** All strategies yield similar conversion rates (~22%), indicating acceptance volumes are stable across pricing.
* **Revenue:** The @22% strategy generates the highest expected revenue but is undermined by unprofitable loss ratios (>100%).
* **Profitability:** @23% delivers the best balance — strong revenue, positive profitability per offer, and a sustainable loss ratio (<100%).
* **Model Performance:** Statistically, @22% and @23% show strong predictive power (AUC ≈ 0.71, well-calibrated). ASIS FEE is unreliable (AUC = 0.50, equivalent to random).
* **Claims Relationship:** No strong correlation exists between conversions and claims. Strategy choice affects both differently — higher conversions do not necessarily mean higher claims.
* **Elasticity:** Pricing is generally inelastic (–0.4 to –0.5). Customers are not highly sensitive to moderate price changes, suggesting room for margin optimization.

**Recommendation**

* **Adopt @23% strategy** as the preferred pricing approach — it provides the most robust and sustainable balance of conversion, revenue, and claims risk.
* **Avoid @22%** despite higher revenue, as underwriting losses make it unsustainable.
* **Discard ASIS FEE** due to poor statistical performance and heavy prediction bias.

**Conclusion:**  
The D+G pricing strategy should prioritise **@23%**, it offers the strongest long-term profitability and operational stability while maintaining competitive conversion rates.

**Note:** In this report, any content enclosed in parentheses refers to a corresponding tab & table in the shared Excel file.  
For example, *(EDA, Table 1)* indicates that the values for the following section can be found in the **EDA** tab, **Table 1** of the Excel file.

# Explore the dataset and conduct any EDA that you feel will help you understand.

## Dataset Overview

* The dataset contains **8,867 rows and 21 columns** initially.
* After handling duplicates and missing values, the final dataset has **8,864 rows**.
* Two exact duplicate rows were removed.

## Missing Values Treatment (EDA: Table 1)

* **Sold Premium**: Left as missing when sale\_flag = 0, since no premium is paid if the policy is not purchased.
* **Purchase Price**: One record with missing value was dropped due to lack of reliable imputation.
* **Plan Count**: Missing values occurred only when plan\_flag = 0. These were imputed as **0**, reflecting no active plans.
* **Plans Active/Cancelled Last Year**:

If plan\_flag = 0 and sold\_premium exists → assumed one active plan last year that was later cancelled.

Otherwise → both set to **0**.

* **Claims Count & Claim Amount**: Missing values imputed as **0**, assuming no claims were made (with caveat that this could understate claims if missingness is due to data entry).

## Visualizations (EDA)

### Policy Acceptance Distribution - Pie Chart

* Majority of customers **did not purchase** a policy.
* Around **22% policies sold** vs **78% not sold**, indicating low conversion rates overall.

### Offered Premium Distribution – Histogram

* Premiums are **right-skewed**, with most offers clustered in the lower premium ranges.
* A few high-premium outliers exist, potentially impacting customer acceptance.

### Policy Sales Distribution by Pricing Strategy - Grouped Bar Chart

* Acceptance rates vary significantly across pricing strategies.
* Some strategies achieve visibly better conversions, highlighting sensitivity of customers to premium levels.

### Correlation Heatmap

* **Sold Premium and Offered Premium** show strong positive correlation.
* **Claims Count and Claim Amount** are highly correlated, as expected.
* No severe multicollinearity across most other features, though pricing-related variables cluster together.

## Descriptive Statistics (EDA: Table 2)

* **Average Offered Premium** is around mid-range of distribution, with large spread due to high-value items.
* **Claim Amounts** are mostly zero, but when present, they can be large outliers.
* **Plan-related counts** generally small, reflecting few customers with multiple plans.

The dataset is mostly clean after handling duplicates and missing values. Policy acceptance is relatively low (~22%), with strong variation across pricing strategies. Offered premiums are skewed, and claim-related data shows sparse but high-impact events. These insights highlight the importance of pricing in driving conversion and the need to carefully handle claim-related features for modelling.

# Compare the pricing performance of all strategies in a summary table with:

## Summary Table (Pricing Performance Summary - Table 1)

It shows for each pricing strategy (@22%, @23%, ASIS FEE):

* **Number of Offers**: Ranged between ~1,100 and ~3,900 per strategy, with @22% having the highest volume.
* **Average Base vs Offered vs Sold Premium**:
  + @22% and @23% involve uplift over the base rate, while ASIS FEE aligns closely with the base.
  + Sold premiums were slightly higher than offered premiums, suggesting some rounding or adjustments post-sale.
* **Average Price Increase**: Highest for @22% (≈ £0.12 uplift), smaller for @23% (≈ £0.06), and none for ASIS FEE.
* **Conversion Rate**: All strategies had broadly similar conversion rates (~22–23%), with ASIS FEE slightly ahead.

## Visual Insights (Pricing Performance Summary)

* **Number of Offers by Pricing Point:** @22% and @23% dominate volume, while ASIS FEE is less common.
* **Average Rates by Pricing point:** Shows clear differences in pricing levels across strategies, with ASIS FEE lowest.
* **Average Price Increase by Pricing Point:** Confirms incremental loading in @22% and @23% vs no uplift in ASIS FEE.
* **Conversion vs No. of offers by Pricing Point**: Conversion rates visually similar across all strategies (all just above 20%).
* **Offered Premium by Pricing Point:** Distribution of offers varies across strategies, showing different risk/pricing mixes. Too many extreme values can be seen in @23%

## How would you test (statistically) if they are different?

### Do conversion rates differ? (Pricing Performance Test - Conversion Rate Differences)

* **Chi-Square Test of Independence:**
  + χ² = 1.02, p = 0.60.
  + Decision: Fail to reject H₀.
  + Conclusion: No statistical evidence that conversion differs across pricing strategies**.**

### Do Average Premiums Differ? (Pricing Performance Test - Offered Premium Differences)

* **Normality & Variance Tests**: KS test rejected normality; Levene’s test showed unequal variances.
* **Kruskal–Wallis Test:** H = 188.36, p < 0.001.
* **Conclusion:** Offered premium distributions are significantly different across strategies.

### Do revenues (sold premiums) differ? (Pricing Performance Test - Revenue Differences)

* **Normality Test (Shapiro–Wilk)**: All groups non-normal (p < 0.05).
* **Kruskal–Wallis Test**: H = 62.13, p < 0.001.
* **Conclusion**: Revenue outcomes differ significantly across pricing strategies.

All strategies yield similar conversion rates (~22%), with no statistical difference. However, pricing levels (offered premiums) and realized revenues (sold premiums) differ significantly across strategies. This means that while acceptance likelihood does not change materially by strategy, financial outcomes do, and hence strategy choice affects profitability rather than volume.

# Compare price elasticity of the strategies in the dataset:

## Price Elasticity Across Group (Price Elasticity – Table 1)

* **Reference Group:** “ASIS FEE” is treated as the baseline.
* **Elasticity Formula:** % change in conversion ÷ % change in price relative to ASIS FEE.
* **Findings**:
  + @22%: Elasticity ≈ –0.44
  + @23%: Elasticity ≈ –0.46
* Both are inelastic (between 0 and –1).
* A 1% increase in price leads to only a ~0.45% decrease in conversion.
* Negative sign confirms the expected inverse relationship (higher price → lower conversion).
* Conversion rates do not fall sharply with moderate price increases, meaning revenues can still grow despite higher premiums.

## Price Elasticity - Withing Group.

### Customer Related (Price Elasticity – Table 2)

**Elasticity by Plan Flag:**

* No Active Plan (Plan Flag = 0):
  + Conversions are low (~16–17%).
  + @22%: Elasticity ≈ –0.49 → moderately elastic.
  + @23%: Elasticity ≈ –0.02 → virtually insensitive to price.
* Active Plan (Plan Flag = 1):
  + Conversions are much higher (~46–49%).
  + @22%: Elasticity ≈ –0.28 → fairly inelastic.
  + @23%: Elasticity ≈ –1.29 → highly elastic (very sensitive to price).

**Insight:**

* Customers with existing plans convert more often, but their sensitivity depends on strategy: they are tolerant at @22% but very price-sensitive at @23%.
* Customers without plans are generally price-insensitive, except a moderate sensitivity at @22%.

### Appliance Related (Price Elasticity – Table 3)

**Elasticity by Item Age Group**

* **Pre-purchase & New items:** Low conversion (≈18–23%). Elasticities are negative and moderate to strong (e.g., Pre-purchase @23% ≈ –5.24), showing **high sensitivity**.
* **Young & Mid-age items:** Mixed results. Some unusual/positive elasticities (e.g., Young @23% ≈ +0.10, Mid-age @23% ≈ +0.85), suggesting factors beyond price (e.g., customer intent or item value).
* **Old items:** Very high conversions (>50%).
  + @22%: Elasticity ≈ +2.25 → unusually positive, indicating conversions may rise with higher price (possible signalling effect of value).
  + @23%: Extremely negative (≈ –28.9), showing extreme price sensitivity.
* **Very Old items:** High conversion at ASIS baseline (≈75%). Elasticities for @22% and @23% are negative (–0.44, –0.89), showing moderate price sensitivity.

**Insight:**

* Elasticity varies sharply by item age.
* Younger/pre-purchase items are more price-sensitive, while older items sometimes show atypical behaviour, possibly due to **small sample sizes** or **value perception effects**.
* Pricing strategies need to account for appliance lifecycle: early-stage products require competitive pricing, while mid/old products may sustain higher premiums.

**Between strategies**: Both @22% and @23% are inelastic relative to ASIS FEE, meaning moderate price rises don’t strongly hurt conversions.

**Customer segmentation**:

* With a plan: tolerant at @22%, but highly elastic at @23%.
* Without a plan: largely insensitive, except moderate elasticity at @22%.

**Appliance segmentation**:

* Pre-purchase and new items → highly price-sensitive.
* Mid/old items → mixed effects; some unusual positive elasticities suggest other drivers of purchase.
* Very old items → conversions are high, but elasticity shows moderate sensitivity.

# Assuming each pricing point is generated by a model at a different global constraint, would you be able to indicate:

## Which one is the best model from the statistical point of view? (Pricing Model – Table 1)

**Overall model metrics:**

* AUC = **0.666** → indicates modest discriminatory power (better than random, but not strong).
* Brier Score = **0.228** → reasonable calibration, but not perfect.

**By pricing point :**

* @22%: **AUC 0.715, Brier 0.149** → strongest discrimination and best calibration.
* @23%: **AUC 0.714, Brier 0.153** → also strong, slightly weaker than @22%.
* ASIS FEE: **AUC 0.500, Brier 0.767** → no discriminatory power (equivalent to random guessing), very poor calibration.

**From a statistical standpoint,** @22% **is the best model**, followed closely by @23%. ASIS FEE clearly underperforms.

## And from pricing point of view(revenue)? (Pricing Model – Table 2)

**Expected Revenue:**

* @22%: EV revenue = 12.27 → highest, driven by higher premiums despite lower conversion.
* @23%: EV revenue = 11.66 → mid-level performance.
* ASIS FEE: EV revenue = 11.56 → lowest, as lower premiums offset slightly higher conversion**.**

**From a pricing perspective,** @22% **is the most profitable strategy**, balancing conversion and premium levels effectively.

## Would you say these models are biased? (Pricing Model – Table 3)

**Calibration Curve:**

* Deviations from the diagonal line indicate bias.

**Calibration Statistics:**

* Intercept = –1.67 (ideal ≈ 0) → model systematically underestimates probabilities.
* Slope = 1.24 (ideal ≈ 1) → over-dispersion: low probabilities underestimated, high probabilities overestimated.

**Bias by Pricing Point:**

* @22%: Slight underprediction (predicted 0.212 vs. actual 0.220).
* @23%: Very small underprediction (predicted 0.223 vs. actual 0.227).
* ASIS FEE: Severe overprediction (predicted 1.0 vs. actual 0.233).

**The models at** @22% **and** @23% **are reasonably well-calibrated with only mild bias, whereas ASIS FEE is heavily biased and unreliable.**

# Claims Information are also provided, is there any direct relationship between conversion and claims?

## Loss Ratio on the Sold Cohort (Claims & Conversions – Table 1)

**Definition:** Loss Ratio = (Expected Claim Cost ÷ Premium Earned) on policies actually sold.

**Findings:**

* @22%: Claim cost (58.46) slightly exceeds average premium (55.77) → Loss Ratio > 100%.
* @23%: Claim cost (49.17) is below premium (51.48) → Loss Ratio < 100%.
* ASIS FEE: Claim cost (46.40) below premium (49.57) → Loss Ratio < 100%.

**Only** @22% **shows an unprofitable sold cohort** (claims outweigh premiums). Both @23% and ASIS FEE are loss-ratio positive.

## Direct Correlation of Claims with Conversion (Claims & Conversions – Table 2)

**Pearson Correlation**

* Measures the strength and direction of a linear relationship between two continuous variables.
* Pearson correlation helps quantify whether higher conversions are associated with higher claims, and if the relationship is positive or negative.

**Range:** Values range from -1 to +1:

* +1 → perfect positive linear relationship
* 0 → no linear relationship
* -1 → perfect negative linear relationship

**Observation**:

* Higher conversion is not necessarily associated with higher or lower claims.
* For example, ASIS FEE has the highest conversion rate but moderate claims, while @22% has the lowest conversion but the highest claims cost.

**Conclusion:**

* There is no strong direct relationship between conversion rate and claim cost.
* Pricing strategy influences both conversion and risk differently — higher conversion does not always mean higher risk.

## Profit/Loss Value on Average per Offer (Claims & Conversions – Table 3)

**Definition:** Profit/Loss per offer = Conversion × (Premium – Claim Cost).

**Findings (Table 12: Profit/Loss per Offer):**

* @22%: Slight negative (due to higher claims than premiums).
* @23%: Positive, indicating sustainable profitability.
* ASIS FEE: Positive but smaller than @23%, since lower premiums reduce margins despite acceptable claims.

**On a per-offer basis,** @23% **delivers the healthiest balance of conversion and profitability.**

# Final Recommendation based on findings?

| **Pricing Strategy** | **Statistical Strength (AUC/Brier)** | **Expected Revenue** | **Loss Ratio (Sold Cohort)** | **Profitability per Offer** | **Overall Assessment** |
| --- | --- | --- | --- | --- | --- |
| **@22%** | Best (AUC 0.715, Brier 0.149) | Highest (12.27) | >100% (unprofitable sold cohort) | Slight negative | Strong model fit, but underwriting results are unprofitable |
| **@23%** | Strong (AUC 0.714, Brier 0.153) | Mid (11.66) | <100% (profitable) | Positive (best balance) | Most robust strategy overall |
| **ASIS FEE** | Weak (AUC 0.500, Brier 0.767) | Lowest (11.56) | <100% (profitable) | Positive but low | Statistically unreliable, not recommended |

**Key Insights:**

* Conversion rates are similar across all strategies (~22%), so revenue and profitability hinge more on pricing and claims than acceptance volume.
* @22% looks strongest statistically and for revenue, but its loss ratio >100% makes it unsustainable in practice.
* @23% achieves a healthier balance: strong predictive performance, profitable loss ratio, and positive profit per offer.
* ASIS FEE is financially acceptable but statistically invalid (random-like model performance).

**Recommendation**

* Adopt the @23% pricing strategy as the final choice: it is the most sustainable, balancing conversion, profitability, and claims risk.
* Avoid relying on @22% despite high revenue, as it leads to losses on sold policies.
* Discard ASIS FEE from operational consideration due to poor model discrimination and heavy bias.