**Model Performance and Fairness Analysis for Home Credit**

**Prediction Setup**

In this study, the **target variable is default risk**:

* **Label 1 = Default (customer fails to repay)**
* **Label 0 = Non-default (customer repays successfully)**

Thus, when the model predicts 1, it identifies a customer as **high risk (likely to default)**.

**Overall Model Performance**

Two baseline models were evaluated:

| **Model** | **AUC** | **Accuracy** | **Precision** | **Recall** |
| --- | --- | --- | --- | --- |
| Logistic Regression | 0.767 | 0.707 | 0.172 | 0.691 |
| LightGBM (baseline) | 0.778 | 0.757 | 0.196 | 0.647 |

* **LightGBM** slightly outperforms Logistic Regression across metrics, achieving a higher **AUC (0.778)** and **accuracy (0.757)**.
* However, both models show **low precision (<0.20)**, meaning a large share of predicted defaults are false alarms.
* Recall is decent (~0.65–0.69), indicating the models are able to capture a fair portion of actual defaults.

**Feature Importance (Global Explainability)**

The **most influential features** in predicting default are external credit scores and loan-related variables:

1. EXT\_SOURCE\_2 and EXT\_SOURCE\_3 – strong external credit scores that help discriminate between good and risky customers.
2. EXT\_SOURCE\_1 – another credit score indicator.
3. CODE\_GENDER\_M – gender plays a role, though care must be taken to evaluate fairness.
4. AMT\_GOODS\_PRICE\_x – size of the goods financed.

**Fairness Evaluation**

**1. Gender Groups**

| **Group** | **Selection Rate (default=1)** | **TPR (Equal Opportunity)** | **FPR (False Alarms)** |
| --- | --- | --- | --- |
| Base | 0.221 | 0.583 | 0.194 |
| Male | 0.356 | 0.732 | 0.313 |
| Missing | 0.500 | 0.000 | 0.500 |

* **Selection rate** (predicted default at threshold 0.5) is much higher for **males (35.6%)** than the base (22.1%).
* **Equal Opportunity**: Males have a higher TPR (73%) than the base (58%), meaning defaults among men are caught more often.
* **False Alarms**: However, men also suffer higher FPR (31.3% vs. 19.4%), i.e., more good male customers are wrongly flagged as risky.
* For missing gender values, performance collapses (TPR=0, FPR=50%), suggesting data issues.

**2. Age Groups**

| **Age Group** | **Selection Rate** | **TPR** | **FPR** |
| --- | --- | --- | --- |
| Base | 0.442 | 0.759 | 0.402 |
| 30–40 | 0.329 | 0.721 | 0.288 |
| 40–50 | 0.244 | 0.629 | 0.213 |
| 50–60 | 0.180 | 0.514 | 0.157 |
| 60–70 | 0.118 | 0.381 | 0.104 |

* **Younger customers (30–40)** face higher selection rates (33%) compared to older groups (e.g., only 12% for 60–70).
* **Equal Opportunity** declines with age: TPR drops from ~72% (30–40) to ~38% (60–70), meaning defaults among older customers are less likely to be detected.
* **False Alarms (FPR)** also fall with age, suggesting older groups are less frequently misclassified as risky.

**Interpretation**

* The model is **more aggressive on younger and male customers**, flagging them as higher risk more often.
* This results in **better detection (higher TPR)** but also **more false alarms (higher FPR)** for those groups.
* Conversely, older customers are less likely to be flagged as risky, but defaults among them are missed more often.

This highlights a trade-off between **catching more defaults (Equal Opportunity)** and **avoiding unfair false alarms (FPR)** across demographic groups.