Formula Design and Rationale

1. Enhanced Risk Score Formula

The **enhanced risk score** was designed as a **composite index (0–100 scale)** that reflects multiple aspects of driving behavior and vehicle risk.

Formula (simplified):

$$\text{Enhanced Risk Score} = \sum_i w_i \cdot \text{Normalized Feature}_i + w_v \cdot \text{Vehicle Risk}$$

- **Weights (sum = 100):**
    - Total Harsh Brakes → 20%
    - Total Harsh Accels → 15%
    - Max Speed Overall → 20%
    - Night Trip Percentage → 10%
    - Claims Weighted Score → 25%
    - Vehicle Risk → 10%
- **Why these weights?**
    - **Harsh Brakes (20%)**: Aggressive braking strongly correlates with near-miss events and accident likelihood.
    - **Harsh Accelerations (15%)**: Captures risky driving but is slightly less predictive than braking.
    - **Max Speed (20%)**: Overspeeding is a leading indicator of accident severity, hence weighted heavily.
    - **Night Trips (10%)**: Driving at night carries higher accident risk, but applies only to some drivers.
    - **Claims Weighted Score (25%)**: Historical claims are the strongest risk indicator, so given the highest weight.

    - **Vehicle Risk (10%)**: Adjusts for inherent differences in vehicle type (e.g., Sports Cars riskier than Sedans, Electric safer).

- **Normalization:** Each feature is scaled 0–1 before weighting, ensuring comparability across different units.

---

2. Claims Weighted Score

The **claims weighted score** was designed to represent **driver claim history** in a fair but impactful way.

$$\text{Claims Weighted Score} = \alpha \cdot \text{Claim Frequency} + \beta \cdot \text{Claim Severity}$$

Formula (conceptual):

- **Why combine frequency and severity?**
    - A driver who files many small claims and a driver who files one catastrophic claim are both high-risk, but in different ways.
    - Balancing both ensures neither group is under- or over-penalized.
- **Choice of weights (α, β):**
    - Heavier emphasis placed on **severity** (larger payouts are more costly to insurers).
    - Frequency still matters because consistent small claims suggest risky habits.

---

3. Premium Cost Formula

The premium formula connected predicted **risk scores** to actual **insurance pricing**.

Formula:

$$\text{Premium} = \text{Base Premium} \times (1 + 0.5 \times \frac{\text{Risk Score}}{100})$$

- **Base Premium = $2,285/year**
    - This figure reflects the **U.S. average full coverage insurance cost**, ensuring industry realism.
- **Scaling Factor = up to 50% increase**
    - Chosen so that **safe drivers** pay close to the baseline.
    - **High-risk drivers** can see premiums rise significantly, but not unrealistically (max ~$3,425/year).
- **Monthly Premium**: Simply annual divided by 12 for affordability context.

Why This Formula Design Works

- **Balanced:** Risk score integrates behavior, claims, and vehicle factors without over-relying on any single dimension.
- **Transparent:** Each component has a clear rationale that can be explained to insurers, regulators, or customers.
- **Realistic:** Premiums are grounded in industry averages and scaled in a way that reflects real-world insurer adjustments.
- **Flexible:** Weights, base premium, and scaling factor can be tuned for different geographies or business goals.