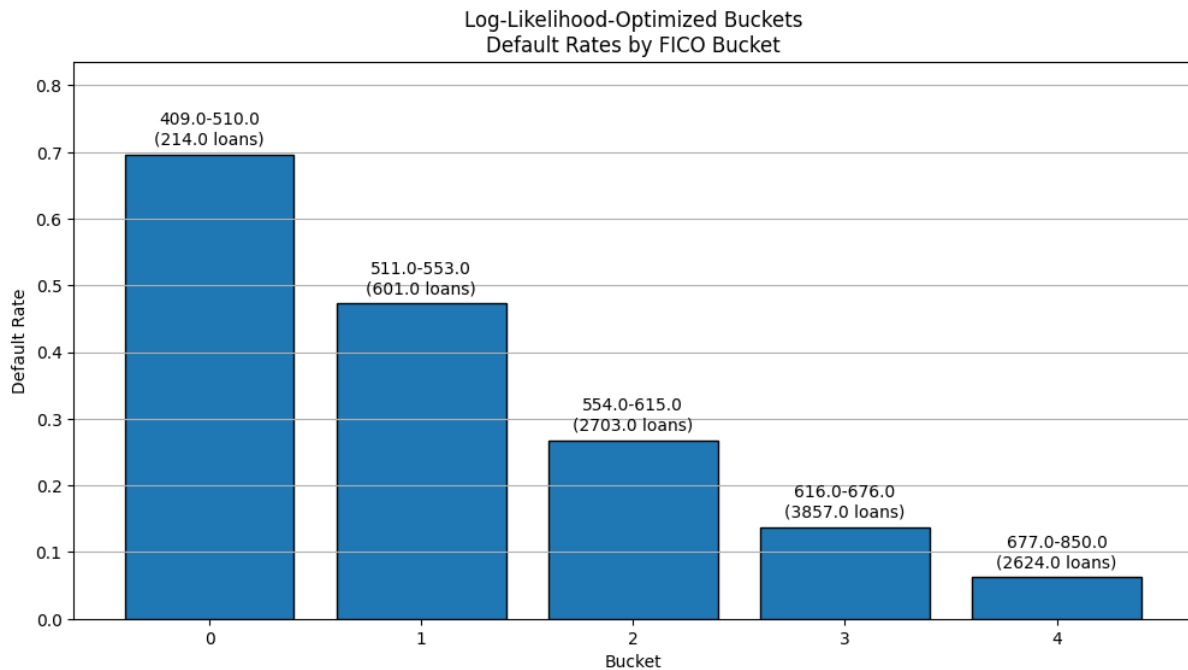# Task: Bucket FICO Scores Report

## 1. Log-Likelihood-Optimized Buckets & Default Rates by FICO Bucket

This section displays the FICO score ranges (buckets) optimized using the **log-likelihood method**, along with the number of loans in each bucket:

Interpretation:

- The log-likelihood method prioritizes **separating borrowers with significantly different default probabilities**. This results in uneven bucket sizes (e.g., 214 loans in the first bucket vs. 3857 in the fourth).

- The first bucket (409–510) likely has the **highest default rate** (0.6963, as seen above), while the last bucket (677–850) has the **lowest default rate** (0.0625).

2. Optimal Boundaries (MSE vs. Log-Likelihood)

MSE-Optimized Buckets
Default Rates by FICO Bucket
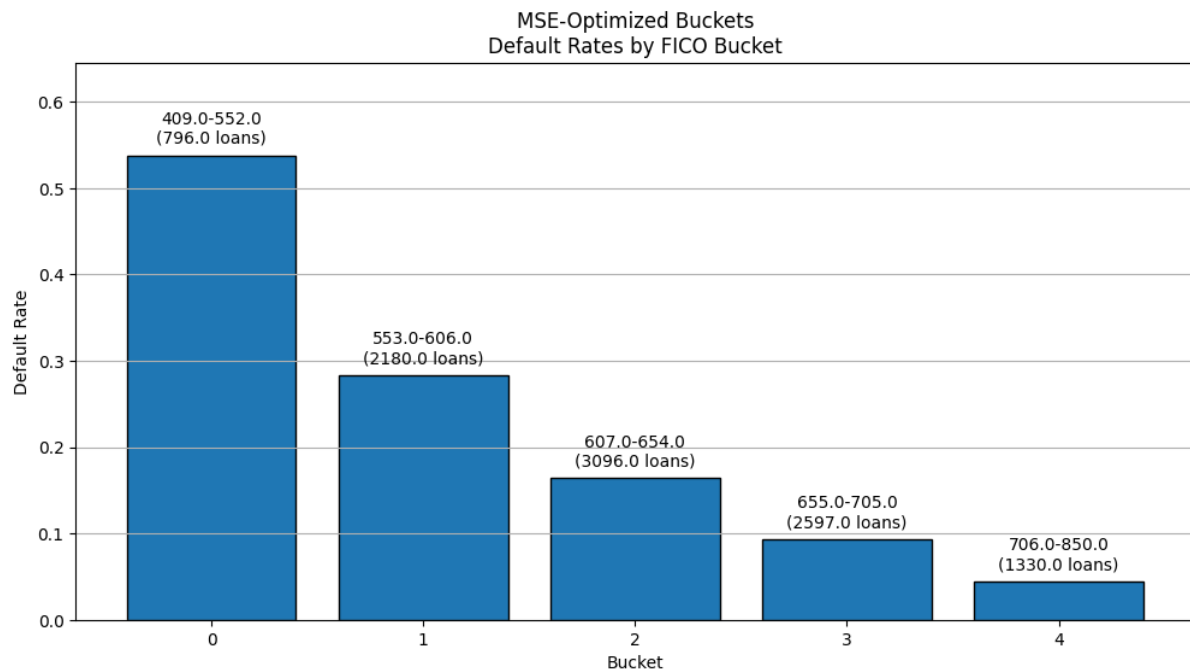
Two sets of boundaries were computed:

- **MSE-Optimized Boundaries**: [408, 552, 606, 654, 705, 851]

- **Log-Likelihood-Optimized Boundaries**: [408, 510, 553, 615, 676, 851]

Interpretation:

- **MSE Optimization**: Focuses on minimizing the squared error between actual FICO scores and their bucket representatives. This leads to more **uniform bucket sizes** (average ~2000 loans per bucket) but may not distinguish default rates as sharply.

- **Log-Likelihood Optimization**: Focuses on **maximizing the discriminative power** of default probabilities. Buckets are smaller where default rates change sharply (e.g., near low FICO scores).

3. Rating Maps

Two rating maps were generated, assigning ratings (1=best, 5=worst) to FICO ranges:

MSE-Optimized Rating Map

Rating	Range
5	408–551
4	552–605
3	606–653
2	654–704
1	705–851

Log-Likelihood-Optimized Rating Map

Rating	Range
5	408–509
4	510–552
3	553–614
2	615–675
1	676–851

Interpretation:

- The log-likelihood map **more aggressively segments low FICO scores** (e.g., 408–509 for Rating 5) because defaults are highly concentrated here.
- The MSE map spreads ranges more evenly, which may dilute risk differentiation.

---

4. Example Bucketing (Log-Likelihood)

Examples of FICO scores mapped to buckets and ratings:

- FICO 450 → Bucket 0 (Rating 5)
- FICO 580 → Bucket 2 (Rating 3)
- FICO 650 → Bucket 3 (Rating 2)
- FICO 720+ → Bucket 4 (Rating 1)

Interpretation:

- Low FICO scores (e.g., 450) are assigned the worst rating (5), reflecting higher default risk.
- High FICO scores (≥676) are assigned the best rating (1), reflecting lower risk.

5. Method Comparison

Metric	MSE-Optimized	Log-Likelihood-Optimized
Avg. Bucket Size	~2000 loans	~2000 loans
Default Rate Range	0.0444–0.5377	**0.0625–0.6963**

Key Takeaways:

- **Log-likelihood** achieves a **wider default rate range**, better separating high-risk and low-risk borrowers.

- **MSE** produces more balanced buckets but with less risk differentiation.

- **Recommendation**: Use **log-likelihood** for risk modeling (prioritizes default discrimination), and **MSE** for general-purpose bucketing (prioritizes uniformity).

Summary of Findings

- **Log-likelihood optimization** is superior to **MSE optimization** for **risk segmentation**, as it creates buckets with stark differences in default rates (e.g., 0.06 vs. 0.69).

- **MSE optimization** is simpler and more uniform but less effective for predicting defaults.

- The **rating maps** align with intuition: lower FICO → worse rating (higher default risk).

Business Implications:
For Charlie's goal of predicting defaults, the **log-likelihood buckets** are preferable because they maximize the model's ability to distinguish risk levels.