Task 4: Analysis, Comparison, and Future Steps

1. Presenting Results

• Deep Learning Model:
- AUC: **0.7327**
- F1-Score: **0.3401**

• Offline RL Agent:
- Estimated Policy Value: **0.0**

- We defined rewards as profit/loss per loan as required.
- Then we attempted to train an offline Q-network, but training was numerically unstable , so the learned policy effectively degenerated, giving an estimated average reward close to **0.0.**

This highlights practical challenges of offline RL on large, skewed financial data, and you propose better RL algorithms and reward normalization as future work.

2. Why These Metrics?

AUC and F1 tell how well the DL model predicts defaults. AUC shows how well the model separates good and bad borrowers, and F1 shows the balance between precision and recall.

The RL agent uses Estimated Policy Value because RL focuses on decisions (approve or deny) based on expected profit, not prediction accuracy.

3. Comparing the Policies

• DL Model Policy: Approves loans only if the predicted default chance is low.
• RL Policy: Approves loans if expected profit is positive, even if the borrower is risky.

Example difference:
A risky borrower with high interest rate may be rejected by the DL model but approved by the RL agent because the RL agent focuses on potential profit.

4. Future Steps

• Improve data quality and add more borrower financial history.
• Use better offline RL algorithms like CQL or IQL.
• Perform real-world A/B testing before deployment.
• Add economic factors such as recovery rate and servicing cost.