📊 Project Insights – Loan Default Risk Analysis

🔹 1️⃣ Credit Score is the Strongest Risk Indicator

Customers with **Credit Score < 600** show significantly higher probability of default compared to customers with higher scores.

👉 Insight:
Bank should classify low credit score customers as **high-risk segment** and apply stricter approval criteria.

🔹 2️⃣ Default Probability is Conditional, Not Random

From conditional probability analysis:

$$P(\text{Default} \mid \text{Low Credit Score})$$

is much higher than overall default probability.

👉 Insight:
Risk assessment should be **conditional-based**, not overall average-based.

🔹 3️⃣ Loan Amount Shows High Dispersion

Standard deviation of Loan_Amount indicates significant variability.

👉 Insight:
Customers borrow very different loan amounts, so risk exposure is uneven.
Bank must apply **risk-based pricing model**.

🔹 4️⃣ Income Distribution Shows Moderate Variation

Q-Q plot indicates income distribution may not be perfectly normal.

👉 Insight:
Loan approval models should not assume strict normality of income.

🔹 5️⃣ Vector Similarity Shows Financial Profile Clusters

Using dot product & angle between vectors:

- Smaller angle → Similar financial profile
- Larger angle → Different financial strength

👉 Insight:
Bank can use vector similarity for **customer segmentation** and targeted lending strategies.

🎯 Final Overall Conclusion

Credit Score is the most influential factor in predicting loan default.
Customers with low credit score significantly increase bank's financial risk.
Therefore, a probability-based and data-driven loan approval system is essential for minimizing default losses.