

## PARTIE 1 — STATISTIQUES & FINANCE

### Question 1.1 : Interprétation des statistiques descriptives

- **La moyenne mensuelle** mesure le rendement moyen obtenu chaque mois.
- **L'écart-type** mesure la volatilité (le risque). Plus il est élevé, plus les rendements sont instables.
- **La médiane** indique le rendement central, moins sensible aux valeurs extrêmes.
- **Le rendement annualisé** permet de projeter la performance sur un an en tenant compte de la capitalisation.
- **La volatilité annualisée** représente le risque annuel total.

En général :

- Le portefeuille A (conservatif) a une volatilité plus faible.
- Le portefeuille B (agressif) a un rendement plus élevé mais un risque beaucoup plus important.

### Question 1.2 — Interprétation des graphiques

- L'histogramme montre que :
  - A est concentré autour de valeurs positives faibles.
  - B présente de fortes fluctuations (valeurs positives et négatives importantes).
- Le boxplot montre :
  - B a des outliers plus extrêmes.
  - L'intervalle interquartile de B est plus large → plus de risque.

Conclusion : B est plus risqué que A.

### Question 1.3 — VaR 95%

#### Définition

La **Value at Risk (VaR 95%)** mesure la perte maximale attendue avec 95% de confiance.

Formule :

$$\begin{aligned} &[ \\ \text{VaR} &= \mu - 1.65 \sigma \\ &] \end{aligned}$$

- 1.65 = quantile 5% loi normale
- $\mu$  = moyenne

- $\sigma$  = écart-type

### Interprétation

- Si VaR annuelle < -10% → contrainte client NON respectée.
- Le portefeuille B dépasse généralement la tolérance de perte.
- Le portefeuille A est plus compatible avec le profil prudent.

### Test Shapiro

- Si p-value > 0.05 → distribution normale acceptable.
- Si p-value < 0.05 → hypothèse normalité rejetée → VaR paramétrique moins fiable.

## ◆ Question 1.4 — Ratio de Sharpe

### Définition

$$\left[ \text{Sharpe} = \frac{R_{\text{annuel}} - r_f}{\sigma_{\text{annuel}}} \right]$$

Il mesure le rendement par unité de risque.

- Sharpe > 1 → excellent
- Sharpe entre 0.5 et 1 → correct
- Sharpe < 0.5 → faible

### 🔗 Recommandation

On recommande le portefeuille :

- Qui respecte la contrainte VaR
- Qui a le Sharpe le plus élevé
- Dont la distribution est proche de la normalité

👉 Pour un client prudent : **Portefeuille A recommandé**

## ✓ PARTIE 2 — THÉORÈME DE BAYES

## ◆ Question 2.1 — Bayes Manuel

### 🔗 Formule

$$\left[ P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)} \right]$$

Interprétation :

- Le retard de paiement multiplie fortement le risque de défaut.
- Le risque passe d'un niveau faible à un niveau significatif.

👉 Décision métier :

- Surveillance renforcée si >15%
- Restriction crédit si >30%

---

### ◆ Question 2.2 — Mise à jour séquentielle

Bayes permet une mise à jour dynamique :

Posterior 1 devient nouveau prior.

Chaque événement augmente progressivement le risque.

👉 Cela reflète la réalité bancaire : le scoring crédit évolue dans le temps.

---

### ◆ Question 2.3 — Fonction Bayes

La fonction générique permet :

- D'automatiser la mise à jour probabiliste
- D'appliquer Bayes à tout événement
- De créer un système de scoring dynamique

---

### ◆ Question 2.4 — Lien Bayes & Precision

Precision =

$$\left[ \frac{TP}{TP + FP} \right]$$

Cela représente :

$$\left[ P(\text{Défaut} \mid \text{Retard}) \right]$$

👉 Donc :

Precision ML = Probabilité bayésienne a posteriori.

Le classificateur Naive Bayes repose directement sur ce principe.

---

## ✓ PARTIE 3 — KNN & ÉVALUATION

---

### ◆ Question 3.1 — Analyse Dataset

- Dataset déséquilibré ( $\approx 15\%$  défaut)
- Variables les plus corrélées :
  - ratio\_dette\_revenu
  - historique\_retards
  - score\_credit

👉 Ces variables expliquent le risque client.

---

### ◆ Question 3.2 — Preprocessing

- Split stratifié conserve proportion défaut.
  - StandardScaler nécessaire car KNN basé sur distance.
  - Sans normalisation  $\rightarrow$  variables à grande échelle dominant.
- 

### ◆ Question 3.3 — Choix K optimal

- Petit K  $\rightarrow$  overfitting.
- Grand K  $\rightarrow$  underfitting.
- On choisit K maximisant AUC via validation croisée.

AUC mesure capacité à séparer défaut / non-défaut.

---

### ◆ Question 3.4 — Métriques

- Accuracy : performance globale.
- Precision : fiabilité des alertes.
- Recall : capacité à détecter les défauts.
- F1 : équilibre Precision/Recall.
- AUC : performance globale indépendante du seuil.
- Specificity : capacité à détecter les bons clients.

👉 Objectif métier : Recall  $\geq 80\%$ .

---

### ◆ Question 3.5 — Courbe ROC

ROC = TPR vs FPR.

AUC :

- 0.5 → aléatoire
- 0.7-0.8 → bon modèle
- 0.8 → très bon

Indice Youden :

$$J = \text{Sensibilité} + \text{Spécificité} - 1$$

Permet choisir seuil optimal.

---

### ◆ Question 3.6 — ROI

ROI =

Gains (TP)

– Coûts analyses (FP)

– Coûts opportunité

– Pertes FN

On choisit le seuil qui :

- Maximise ROI
- Respecte Recall  $\geq 80\%$