

Manuel de Révision - Études de Cas Data Analyst

Vue d'Ensemble des Études de Cas

Ce dossier contient des études de cas complètes couvrant tous les aspects de la préparation Data Analyst pour UniBank Haiti.

Liste des Études de Cas

1. DataViz_FULL/etude_cas_dataviz_intermediaire.md

Thème: Visualisation de Données

Niveau: Intermédiaire

Compétences testées: - Choix approprié des graphiques - Design de dashboards exécutifs - Critique et amélioration de visualisations - Justification des choix visuels

Points clés à retenir: - Un graphique = Un message - Histogramme pour distributions continues - Box plot pour comparer des groupes - Line chart pour tendances temporelles - Éviter les pie charts > 5 catégories

2. BI_FULL/etude_cas_bi_bancaire.md

Thème: Business Intelligence Bancaire

Niveau: Intermédiaire-Avancé

Compétences testées: - Calcul des KPIs bancaires (ROE, ROA, NIM, NPL, CAR) - Analyse de portefeuille de crédit - Requêtes SQL pour analyse de risque - Segmentation et analyse client - Recommandations business

Formules essentielles:

ROE = Résultat Net / Capitaux Propres × 100

ROA = Résultat Net / Total Actifs × 100

NIM = (Revenus Int. - Charges Int.) / Actifs Productifs × 100

NPL Ratio = Prêts > 90j / Total Prêts × 100

CAR = Fonds Propres / RWAs × 100

3. SQL_Analyst_FULL/etude_cas_sql_analytics.md

Thème: SQL Analytique

Niveau: Intermédiaire-Avancé

Compétences testées: - Window functions (ROW_NUMBER, RANK, LAG, LEAD) - CTEs simples et multiples - Analyses de cohortes - Optimisation de requêtes - Création d'index

Patterns SQL à maîtriser:

-- *Top N par groupe*

ROW_NUMBER() OVER (PARTITION BY groupe ORDER BY valeur DESC)

-- *Cumul courant*

SUM(montant) OVER (ORDER BY date)

-- Variation période sur période
LAG(valeur) OVER (ORDER BY période)

4. Stats_FULL/etude_cas_statistiques_complete.md

Thème: Statistiques et Analyse de Données

Niveau: Intermédiaire-Avancé

Compétences testées: - EDA complète - Tests d'hypothèses (t-test, Chi-carré, ANOVA) - Corrélation et régression - Modélisation du risque - Interprétation business

Tests statistiques clés: | Situation | Test | |-----|---| | Comparer 2 moyennes | t-test | | Comparer 3+ moyennes | ANOVA | | Variables catégorielles | Chi-carré | | Corrélation | Pearson/Spearman |

5. ML_FULLSCREEN/etude_cas_machine_learning.md (À VENIR)

Thème: Machine Learning Bancaire

Niveau: Intermédiaire-Avancé

Compétences testées: - Préparation des données pour ML - Modèles de scoring de crédit - Détection de fraude - Prédiction de churn - Évaluation et interprétation des modèles

Algorithmes clés: | Cas d'usage | Algorithme recommandé | |-----|-----| | Scoring crédit | Régression logistique, Gradient Boosting | | Détection fraude | Random Forest, Isolation Forest | | Segmentation | K-Means, Clustering hiérarchique | | Churn | XGBoost, Random Forest |

Méthodologie Recommandée pour les Études de Cas

Étape 1: Comprendre le Problème (5 min)

- Lire attentivement l'énoncé
- Identifier les questions clés
- Noter les contraintes et données disponibles

Étape 2: Planifier l'Approche (5 min)

- Lister les étapes nécessaires
- Identifier les outils/techniques appropriés
- Estimer le temps par section

Étape 3: Exécuter (60-80% du temps)

- Suivre le plan établi
- Documenter chaque étape
- Vérifier les résultats intermédiaires

Étape 4: Interpréter et Conclure (15-20% du temps)

- Synthétiser les findings
- Formuler des recommandations actionables

- Relier aux objectifs business
-

Conseils pour Réussir les Études de Cas

1. Structure de la Réponse

1. Contexte et compréhension du problème
2. Données utilisées et préparation
3. Méthodologie et justification
4. Résultats avec visualisations
5. Interprétation et limites
6. Recommandations et next steps

2. Pièges à Éviter

- Ne pas commencer sans plan
- Oublier de vérifier la qualité des données
- Ignorer les valeurs manquantes/outliers
- Confondre corrélation et causalité
- Oublier l'interprétation business

3. Éléments Différenciants

- Mentionner les limites de l'analyse
 - Proposer des analyses complémentaires
 - Contextualiser pour le secteur bancaire
 - Quantifier l'impact business
-

Checklist Avant de Soumettre

- Toutes les questions sont répondues
 - Le code est propre et commenté
 - Les visualisations sont claires et titrées
 - Les résultats sont interprétés
 - Les recommandations sont actionables
 - Les sources et hypothèses sont mentionnées
 - Le format est professionnel
-

Révision Express (30 min avant l'examen)

Formules à connaître par cœur

Moyenne: $\bar{x} = \sum x_i / n$

Variance: $s^2 = \sum (x_i - \bar{x})^2 / (n-1)$

CV: $(s / \bar{x}) \times 100$

IC 95%: $\bar{x} \pm 1.96 \times (s/\sqrt{n})$

p-value < 0.05 → Significatif

Graphique approprié

Distribution → Histogramme
Comparaison groupes → Box plot / Bar chart
Tendance → Line chart
Relation 2 variables → Scatter plot
Composition → Pie (si < 6 cat.)

SQL essentiels

-- Ranking
`ROW_NUMBER() OVER (PARTITION BY ... ORDER BY ...)`

-- Cumul
`SUM() OVER (ORDER BY ...)`

-- Période précédente
`LAG(col, 1) OVER (ORDER BY ...)`

ML essentiels

```
# Régression logistique pour scoring
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
model.fit(X_train, y_train)
proba = model.predict_proba(X_test)[:, 1]

# Métriques
from sklearn.metrics import roc_auc_score
auc = roc_auc_score(y_test, proba)
gini = 2 * auc - 1

# K-Means pour segmentation
from sklearn.cluster import KMeans
kmeans = KMeans(n_clusters=5)
labels = kmeans.fit_predict(X_scaled)
```

Ressources Complémentaires

- /Foundational/ - Manuels de préparation par sujet
 - /revisions/ - Fiches de synthèse
 - /tests/ - Questions d'entraînement
 - /out_of_scope/ - Documents hors périmètre (PowerBI, DAX)
-

Bonne préparation et bonne chance!