

PROJET FINAL 1

Analyse de Performance Étudiante

Module BSC-ADD-201 | BDC-ADD-2

Dr. John Benson LOUIS

Décembre 2024

[DEADLINE] Date Limite

DATE LIMITE : Jeudi 5 Décembre 2025 à 23h59

1 Contexte du Projet

1.1 Présentation

Vous êtes data analyst pour une université qui souhaite comprendre les facteurs influençant la performance académique de ses étudiants. L'université a collecté des données sur 200 étudiants sur deux semestres consécutifs.

1.2 Objectifs

Vos objectifs sont de :

1. Caractériser les profils démographiques et comportementaux des étudiants
2. Analyser l'évolution des performances entre les deux semestres
3. Identifier les facteurs associés à la réussite académique
4. Formuler des recommandations concrètes pour améliorer les résultats

1.3 Questions de Recherche

Votre analyse doit répondre aux questions suivantes :

- Quels sont les facteurs les plus fortement associés à la performance académique ?
- Existe-t-il des différences significatives de performance entre les filières ?
- Comment les performances évoluent-elles du semestre 1 au semestre 2 ?
- Quel impact ont les heures d'étude et les absences sur les résultats ?
- Peut-on identifier des profils d'étudiants à risque ?

2 Description des Données

2.1 Localisation

[INFO] Information Importante

Les données sont disponibles dans le dossier : lien :
<https://drive.google.com/drive/folders/1dlqVwVjaC0EMeRC37QXUVETWNwICG3Ku?usp=sharing>
 /data/projet1/

Fichier principal : `etudiants_performance.csv`

2.2 Structure des Données

Le dataset contient **200 observations** (étudiants) et **24 variables** :

2.2.1 Variables Démographiques

Variable	Type	Description
id_etudiant	Character	Identifiant unique (ETU001-ETU200)
nom	Character	Nom de l'étudiant
age	Numeric	Âge (18-26 ans)
genre	Factor	Homme / Femme
ville	Character	Ville d'origine
filiere	Factor	Info / Math / Physique / Eco / Bio

Table 1: Variables démographiques

2.2.2 Variables Comportementales

Variable	Type	Description
heures_etude_semaine	Numeric	Heures d'étude par semaine (5-40h)
nb_absences_s1	Numeric	Nombre d'absences semestre 1
nb_absences_s2	Numeric	Nombre d'absences semestre 2

Table 2: Variables comportementales

2.2.3 Notes Académiques

Semestre 1 : 5 matières

- `note_math_s1` : Note de mathématiques (0-20)
- `note_info_s1` : Note d'informatique (0-20)
- `note_physique_s1` : Note de physique (0-20)
- `note_economie_s1` : Note d'économie (0-20)
- `note_anglais_s1` : Note d'anglais (0-20)

Semestre 2 : 5 matières

- **note_math_s2** : Note de mathématiques (0-20)
- **note_info_s2** : Note d'informatique (0-20)
- **note_physique_s2** : Note de physique (0-20)
- **note_economie_s2** : Note d'économie (0-20)
- **note_anglais_s2** : Note d'anglais (0-20)

2.2.4 Variables Calculées

Variable	Type	Description
moyenne_s1	Numeric	Moyenne générale semestre 1
moyenne_s2	Numeric	Moyenne générale semestre 2

Table 3: Moyennes calculées

3 Analyses Requises**3.1 Analyse 1 : Statistiques Descriptives (25 points)****3.1.1 Démographie (8 points)**

- Distribution de l'âge (moyenne, médiane, écart-type)
- Répartition par genre (effectifs et pourcentages)
- Répartition par filière (effectifs et pourcentages)
- Distribution géographique (villes principales)

3.1.2 Comportement Académique (8 points)

- Statistiques sur les heures d'étude
- Statistiques sur les absences (S1 et S2)
- Relation heures d'étude / absences

3.1.3 Performance Académique (9 points)

- Statistiques complètes pour chaque matière (S1 et S2)
- Distribution des moyennes générales (S1 et S2)
- Taux de réussite global ($\geq 10/20$)
- Identification des matières les plus difficiles

3.2 Analyse 2 : Évolution S1 → S2 (20 points)**3.2.1 Évolution Globale (10 points)**

- a) Comparaison des moyennes S1 vs S2
- b) Distribution des variations de notes
- c) Identification des progressions et régressions
- d) Top 10 des progressions
- e) Top 10 des régressions

3.2.2 Évolution par Matière (10 points)

- a) Comparaison S1 vs S2 pour chaque matière
- b) Matières en progression / régression
- c) Tests de comparaison (bonus si fait)

3.3 Analyse 3 : Performance par Groupes (25 points)**3.3.1 Par Filière (10 points)**

- a) Statistiques descriptives par filière
- b) Comparaison des moyennes entre filières
- c) Identification de la filière la plus performante
- d) Taux de réussite par filière

3.3.2 Par Genre (8 points)

- a) Comparaison des performances H/F
- b) Différences par matière
- c) Évolution S1 → S2 par genre

3.3.3 Par Catégorie d'Âge (7 points)

- a) Créer des catégories d'âge pertinentes
- b) Comparer les performances
- c) Analyser les différences

3.4 Analyse 4 : Facteurs de Réussite (15 points)**3.4.1 Impact des Heures d'Étude (8 points)**

- a) Corrélation heures d'étude / performance
- b) Analyse par tranche d'heures d'étude
- c) Identification d'un seuil optimal

3.4.2 Impact des Absences (7 points)

- a) Corrélation absences / performance
- b) Comparaison faibles vs fortes absences
- c) Évolution des absences S1 → S2

3.5 Analyse 5 : Visualisations (15 points)

Vous devez créer **au minimum 5 graphiques de qualité :**

1. **Histogramme** : Distribution des moyennes (3 pts)
2. **Boxplot** : Moyennes par filière (3 pts)
3. **Scatter plot** : Heures d'étude vs Performance (3 pts)
4. **Graphique d'évolution** : S1 vs S2 (3 pts)
5. **Graphique au choix** : Crétatif et pertinent (3 pts)

[!] Attention

Qualité requise pour les graphiques :

- Titres informatifs
- Axes labelisés correctement
- Légendes appropriées
- Couleurs lisibles
- Résolution ≥ 300 DPI

4 Livrables Attendus

4.1 Structure du Projet

Votre projet doit avoir cette structure :

```
mon_projet_1/
|-- data/
|   |-- raw/
|   |   '-- etudiants_performance.csv
|   '-- processed/
|       '-- donnees_nettoyees.csv
|-- scripts/
|   |-- 00_fonctions.R
|   |-- 01_import_nettoyage.R
|   |-- 02_analyses_descriptives.R
|   |-- 03_analyses_comparatives.R
|   |-- 04_visualisations.R
|   '-- main.R
|-- output/
```

```
|   |-- figures/
|   |   |-- 01_distribution_moyennes.png
|   |   |-- 02_comparaison_filières.png
|   |   '-- ...
|   '-- tables/
|       |-- stats_descriptives.csv
|       |-- comparaison_s1_s2.csv
|       '-- ...
|-- rapport.pdf (ou rapport.docx)
'-- README.txt
```

4.2 Fichiers à Soumettre

1. **Scripts R** : Tous les fichiers .R organisés
2. **Rapport simple** : PDF ou DOCX (2 ou 3 pages)
3. **Graphiques** : Tous les graphiques en PNG haute résolution
4. **Tableaux** : Résultats exportés en CSV
5. **README.txt** : Instructions d'exécution

4.3 Contenu du Rapport

Votre rapport doit contenir :

1. **Page de garde** : Titre, nom, date
2. **Résumé**
3. **Introduction** : Contexte et objectifs
4. **Méthodologie** : Description des données et méthodes
5. **Résultats** : Présentation des analyses avec tableaux et graphiques
6. **Discussion** : Interprétation des résultats
7. **Recommandations** : Au moins 3 recommandations concrètes
8. **Conclusion** : Synthèse et limites
9. **Annexes** : Code important si nécessaire

Critère	Points
1. Structure et Organisation	15
- Structure de dossiers correcte	5
- Fichiers bien nommés et organisés	5
- README clair et complet	5
2. Qualité du Code	30
- Code propre et commenté	10
- Fonctions personnalisées (min 3)	8
- Bonnes pratiques R	7
- Code exécutable sans erreur	5
3. Analyses	30
- Analyse 1 : Statistiques descriptives	8
- Analyse 2 : Évolution S1→S2	6
- Analyse 3 : Performance par groupes	8
- Analyse 4 : Facteurs de réussite	5
- Pertinence et exactitude	3
4. Visualisations	15
- graphiques minimum	8
- Qualité graphique	4
- Pertinence	3
5. Rapport	10
- Structure et rédaction	4
- Interprétations	3
- Recommandations	3
TOTAL	100

Table 4: Grille d'évaluation détaillée

5 Grille d'Évaluation

5.1 Barème Détaillé (100 points)

6 Conseils et Recommandations

6.1 Organisation du Travail

6.2 Bonnes Pratiques

[INFO] Information Importante

Pour réussir votre projet :

- Commencez TÔT
- Testez votre code régulièrement
- Documentez au fur et à mesure
- Sauvegardez fréquemment
- Demandez de l'aide si besoin
- Soignez la présentation

6.3 Import des Données

```
# Import des données
data <- read.csv("/data/projet1/etudiants_performance.csv")

# Vérification
str(data)
summary(data)
head(data)
```

7 Checklist Finale

Avant de soumettre, vérifiez :

7.1 Code

Structure de dossiers correcte

Tous les scripts présents

Code bien commenté

Les fonctions créées

`main.R` s'exécute sans erreur

Pas de chemins absous (uniquement relatifs)

7.2 Analyses

Toutes les analyses requises effectuées

Résultats exportés (CSV)

Calculs vérifiés

7.3 Visualisations

Haute résolution (≥ 300 DPI)

Titres et labels présents

Sauvegardés en PNG

7.4 Soumission

Tous les fichiers présents

Archive ZIP créée

Nom correct : **NOM_Prenom_Projet1.zip**

Soumis avant le 5 Décembre 23h59

lien de soumission :

<https://drive.google.com/drive/folders/1RGY5lnLwNtzXJS8yz8C56OdhaVb07AR2?usp=sharing>

8 Contact et Support

8.1 Enseignant

Dr. John Benson LOUIS Email : [john_benson.louis@ueh.edu.ht]

Bon courage et excellente réussite !

"L'analyse de données est autant un art qu'une science."