

INSTITUT SUPERIEUR DE
STATISTIQUE D'ECONOMETRIE ET

UNION-DISCIPLINE-TRAVAIL

MASTER 2

STATISTIQUE- ECONOMETRIE -DATA SCIENCE

MINI-PROJET

STATISTIQUE DESCRIPTIVE : ANALYSE UNIVARIEE ET BIVARIEE

ANALYSE DE LA DEPRESSION CHEZ LES ETUDIANTS

ANNEE ACADEMIQUE :

2024 - 2025

NOM: KABA

PRENOM: MAHAMOUD TOIB

ENSEIGNANT –
ENCADREUR



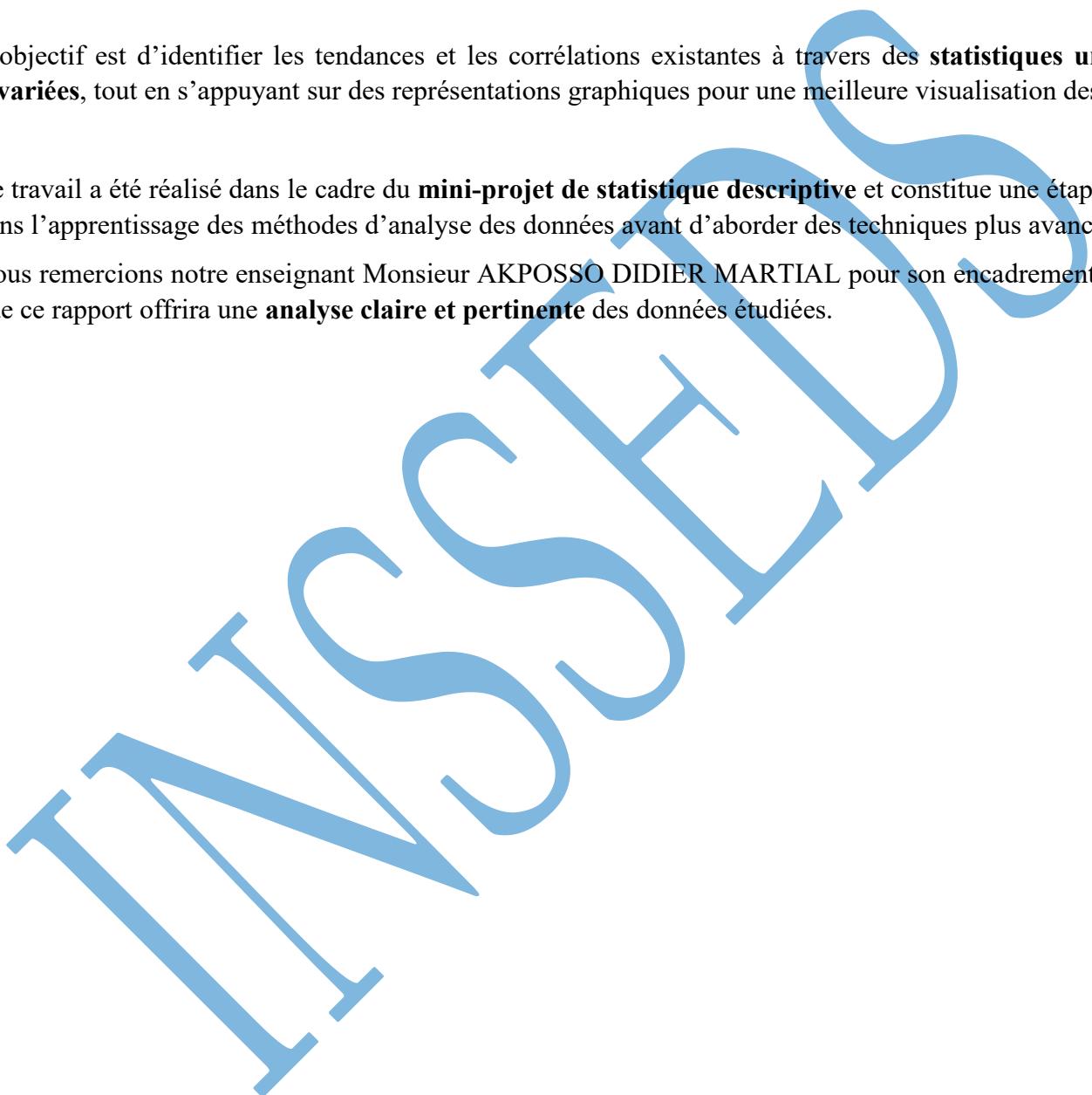
AVANT PROPOS

Ce rapport présente une **analyse descriptive** des facteurs influençant la dépression chez les étudiants. L'étude repose sur des techniques statistiques permettant d'explorer les relations entre divers aspects du mode de vie, des études et la santé mentale des étudiants.

L'objectif est d'identifier les tendances et les corrélations existantes à travers des **statistiques univariées et bivariées**, tout en s'appuyant sur des représentations graphiques pour une meilleure visualisation des résultats.

Ce travail a été réalisé dans le cadre du **mini-projet de statistique descriptive** et constitue une étape essentielle dans l'apprentissage des méthodes d'analyse des données avant d'aborder des techniques plus avancées.

Nous remercions notre enseignant Monsieur AKPOSSO DIDIER MARTIAL pour son encadrement et espérons que ce rapport offrira une **analyse claire et pertinente** des données étudiées.



TABLES DES MATIERES

1. Introduction
 - 1.1. Contexte et Problématique
 - 1.2. Objectifs de l'Étude
 - 1.3. Méthodologie
2. Présentation des Données
 - 2.1. Source et Contexte des Données
 - 2.2. Dictionnaire des Variables
3. Prétraitement des Données
 - 3.1. Analyse et Gestion des Valeurs Manquantes
 - Visualisation des valeurs manquantes
 - Calcul du pourcentage de valeurs manquantes
 - Suppression des individus si moins de 5% sont affectés
 - 3.2. Traitement des Doubloons
 - Détection des observations dupliquées
 - Suppression des doublons
 - 3.3. Identification et Traitement des Valeurs Extrêmes
 - Détection des valeurs extrêmes avec les boîtes à moustaches
 - Traitement des valeurs extrêmes par winzorisation
 - Visualisation des variables avant et après winzorisation
4. Analyse Univariée
 - 4.1. Statistiques Descriptives des Variables Quantitatives
 - Tableaux statistiques (effectifs, fréquences cumulées croissantes et décroissantes)
 - Représentation graphique (histogrammes, diagrammes en bâtons, boxplots)
 - Résumé numérique (min, max, mode, médiane, moyenne, quartiles, écart-type, variance, coefficient de variation, skewness, coefficient d'aplatissement)
 - 4.2. Statistiques Descriptives des Variables Qualitatives
 - Tableaux statistiques (effectifs, fréquences)
 - Représentation graphique (diagramme en barres et diagramme en secteurs)
5. Analyse Bivariée
 - 5.1. Relations entre la Dépression et les Autres Variables
 - Cas 1 : Relation entre deux variables qualitatives
 - Tableau de contingence
 - Diagramme en barres
 - Indicateur de V de Cramer
 - Cas 2 : Relation entre une variable qualitative et une variable quantitative
 - Tableau de contingence
 - Boxplot comparatif
 - Rapport de corrélation
6. Conclusion et Recommandations
 - Résumé des principaux résultats
 - Interprétation des tendances observées et Recommandations basées sur les résultats

1. INTRODUCTION

1.1 Contexte et Problématique

La santé mentale des étudiants est une préoccupation majeure dans le monde académique. Plusieurs études ont montré que des facteurs tels que la **pression académique**, le **stress financier**, la **qualité du sommeil** et les **antécédents familiaux** peuvent influencer l'état psychologique des étudiants et les exposer à la dépression.

Ce rapport vise à étudier les **différents facteurs liés à la dépression** chez les étudiants en utilisant des techniques de **statistique descriptive** afin d'identifier les tendances et relations significatives.

1.2 Objectif du Projet

L'objectif de cette étude est de :

- Analyser la distribution des variables et identifier les tendances.
- Étudier les relations entre la dépression et différents facteurs explicatifs.
- Fournir une **analyse statistique rigoureuse et visuelle** pour faciliter l'interprétation des résultats.

1.3 Méthodologie

Les analyses seront réalisées à l'aide du logiciel **R** et porteront sur :

- Des **statistiques univariées** pour explorer chaque variable.
- Des **statistiques bivariées** pour examiner les liens entre la dépression et d'autres variables.
- Des **représentations graphiques** (histogrammes, boxplots, diagrammes en barres et secteurs) pour une meilleure visualisation des résultats.

2. PRÉSENTATION DES DONNÉES

1.4 Source des Données

Les données utilisées dans cette étude proviennent d'un **jeu de données sur la dépression chez les étudiants**, recueilli dans le cadre du **mini-projet de statistique descriptive** de l'Institut Supérieur de Statistique, d'Économétrie et de Data Science (INSSEDS). Ce jeu de données contient 27901 observations dont 18 variables.

Ce jeu de données a été fourni par Monsieur AKPOSSO DIDIER MARTIAL chargé du cours et contient des informations détaillées sur plusieurs aspects de la vie des étudiants, notamment :

- **Les caractéristiques démographiques** (âge, sexe, ville de résidence).
- **Les performances académiques** (moyenne des notes, satisfaction académique).

STUDENT_DEPRESSION

- **Le mode de vie** (durée du sommeil, habitudes alimentaires).
- **Les facteurs psychologiques et financiers** (stress financier, pression académique, pensées suicidaires).
- **Les antécédents de santé mentale** (existence de troubles dans la famille).
- **L'état de dépression** (variable cible).

Ce jeu de données est particulièrement utile pour comprendre les facteurs pouvant influencer la **santé mentale des étudiants** et aider à la mise en place de stratégies de prévention et d'intervention.

2.2 Contexte de l'Étude et dictionnaire des données

La **dépression chez les étudiants** est une problématique de plus en plus préoccupante dans les milieux académiques. Plusieurs études ont montré que la pression académique, les difficultés financières, le manque de sommeil et les antécédents familiaux peuvent jouer un rôle déterminant dans l'apparition de troubles dépressifs chez les jeunes.

Dans ce contexte, cette étude vise à :

- Analyser les tendances et caractéristiques des étudiants souffrant de dépression** à travers des statistiques descriptives.
 - Étudier les associations entre la dépression et divers facteurs explicatifs**, tels que la pression académique, le stress financier ou les habitudes de vie.
 - Proposer des recommandations basées sur les résultats obtenus**, pour une meilleure prise en charge et prévention de la dépression dans les établissements universitaires.
- Cette analyse s'inscrit dans une approche **statistique descriptive**, en mettant en évidence les **fréquences, distributions et relations** entre les variables, sans chercher à établir de causalité directe.

Dictionnaire des données

Nom de la Variable	Description	Valeurs possibles	Type
id	Identifiant unique de chaque étudiant	-	Numérique
sexe	Sexe de l'étudiant	Masculin, Féminin	Qualitatif
Age	Âge de l'étudiant	Numérique (ex: 18, ...)	Quantitatif

STUDENT_DEPRESSION

		22, 25)	
ville	Ville de résidence	Nom de la ville	Qualitatif
profession	Activité de l'étudiant	Étudiant à temps plein, autre activité	Qualitatif
pression_academique	Niveau de pression académique	Ex: 1 à 5	Quantitatif (Échelle)
moyenne_notes	Moyenne des notes de l'étudiant	(ex: 12.5, 15.3)	Numérique Quantitatif
satisfaction_etudes	Satisfaction par rapport aux études	Faible, Moyenne, Élevée	Ordinal
nombre_heure_travail_etude	Nombre d'heures de travail ou d'étude par jour	Numérique (ex: 2, 5, 8)	Quantitatif
stress_financier	Niveau de stress financier	Ex: 1 à 10	Quantitatif (Échelle)
duree_sommeil	Durée moyenne du sommeil	Moins de 5h, 5-6h, 7h et plus	Qualitatif
habitudes_alimentaires	Qualité des habitudes alimentaires	Saines, Modérées, Mauvaises	Qualitatif
diplome_suivi	Diplôme suivi ou	BSc, M.Tech,	Qualitatif

STUDENT_DEPRESSION

	obtenu	autre	
pensees_suicidaires?	Présence de pensées suicidaires	Oui, Non	Qualitatif (Binaire)
antecedents_familiaux_maladies_mentales	Antécédents familiaux de maladies mentales	Oui, Non	Qualitatif (Binaire)
depression	Présence d'un état dépressif $0 = \text{Non}, 1 = \text{Oui}$		Qualitatif (Binaire)

2. PRETRAITEMENT DES DONNEES

3.0 Présentation du jeu de données de 5 individus sur les 27901 observations

1. Variables	1	2	3	4	5
id	2	8	26	30	32
Sexe	Male	Female	Male	Female	Female
âge	33	24	31	28	25
Ville	Visakhapatnam	Bangalore	Srinagar	Varanasi	Jaipur
profession	Student	Student	Student	Student	Student
pression académique	5	2	3	3	4
pression liée au travail	0	0	0	0	0
moyenne notes	8.97	5.90	7.03	5.59	8.13
satisfaction études	2	5	5	2	3
satisfaction travail	0	0	0	0	0
durée sommeil	5-6 hours	5-6 hours	Less than 5 hours	7-8 hours	5-6 hours
habitudes alimentaires	Healthy	Moderate	Healthy	Moderate	Moderate

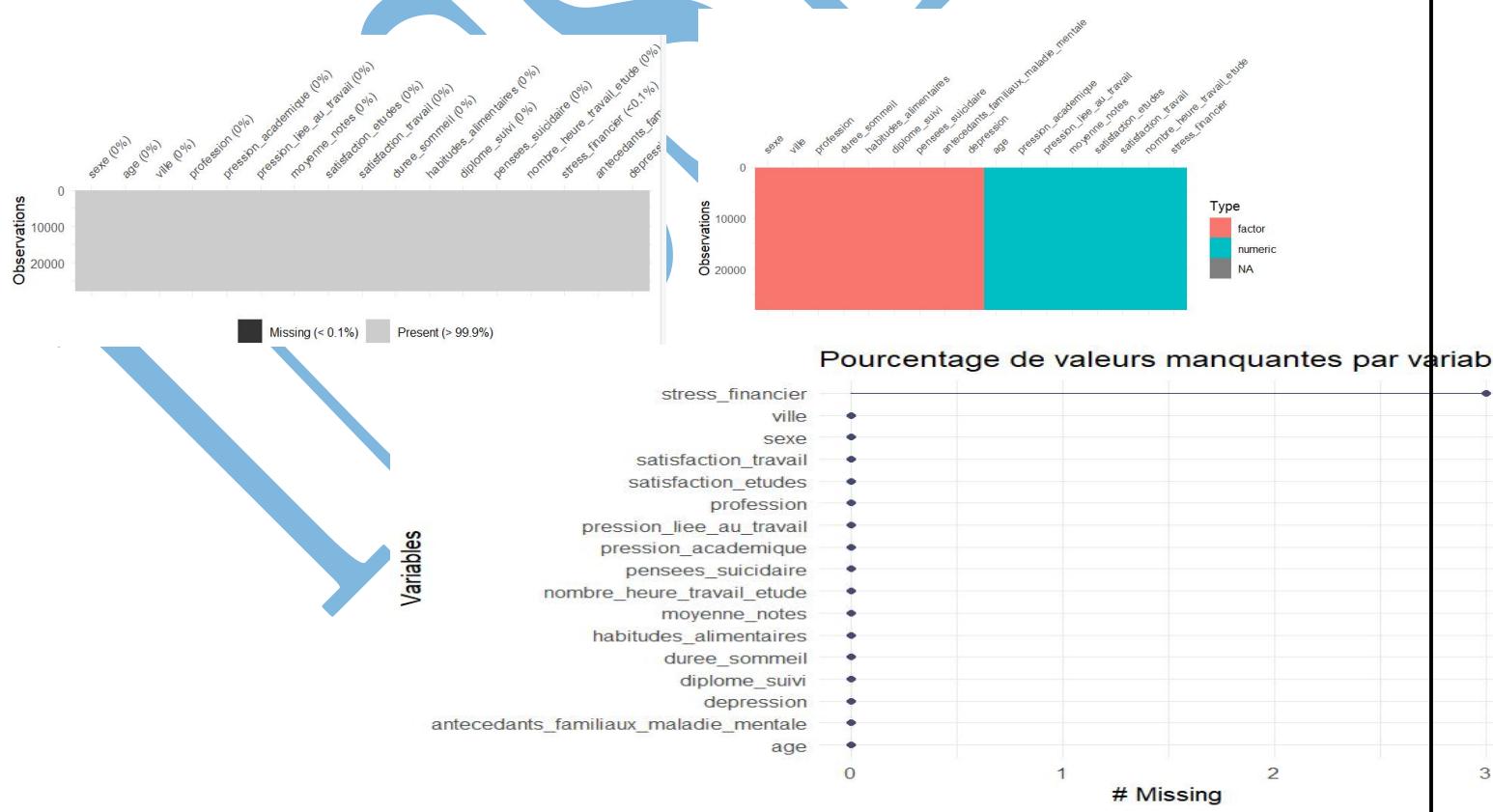
STUDENT_DEPRESSION

diplôme suivi	B.Pharm	BSc	BA	BCA	M.Tech
pensées suicidaires	Yes	No	No	Yes	Yes
nombre heure travail/étude	3	3	9	4	1
stress financier	1	2	1	5	1
antécédents familiaux maladie mentale	No	Yes	Yes	Yes	No
dépression	1	0	0	1	0

2.1 ANALYSE ET GESTION DES VALEURS MANQUANTES

Identification des valeurs manquantes dans le jeu de données, calcul de leur fréquence et prise de décisions sur leur suppression ou leur gestion

➤ Visualisation des valeurs manquantes



Après la visualisation nous voyons avec exactitude que sur les 27901 observations il n'y a que 3 valeurs manquantes et ces valeurs manquantes proviennent de la variables stress_financier. Ces trois valeurs sont tellement insignifiantes sur toutes les observations qu'on arrive à peine à les repérer sur les deux premiers graphiques.

➤ Calcul du pourcentage de valeurs manquantes

Sur les 27901 on a 0.0107523 % de valeurs manquantes

➤ Suppression des individus si moins de 5% sont affectés

Comme sur les 27901 on a 0.0107523 % de valeurs manquantes qui est inférieur à 5% des individus ne disposant pas de valeurs manquantes, alors nous supprimons les

individus qui ont des valeurs manquantes. Il nous restera 27898 observations(individus). Nous pouvons constater ci-dessous qu'il n'y a plus de valeurs manquantes.



STUDENT_DEPRESSION

3.2 TRAITEMENT DES DOUBLONS

Vérification de la présence d'observations répétées et suppression des doublons pour garantir l'intégrité des analyses.

- Détection des observations dupliquées

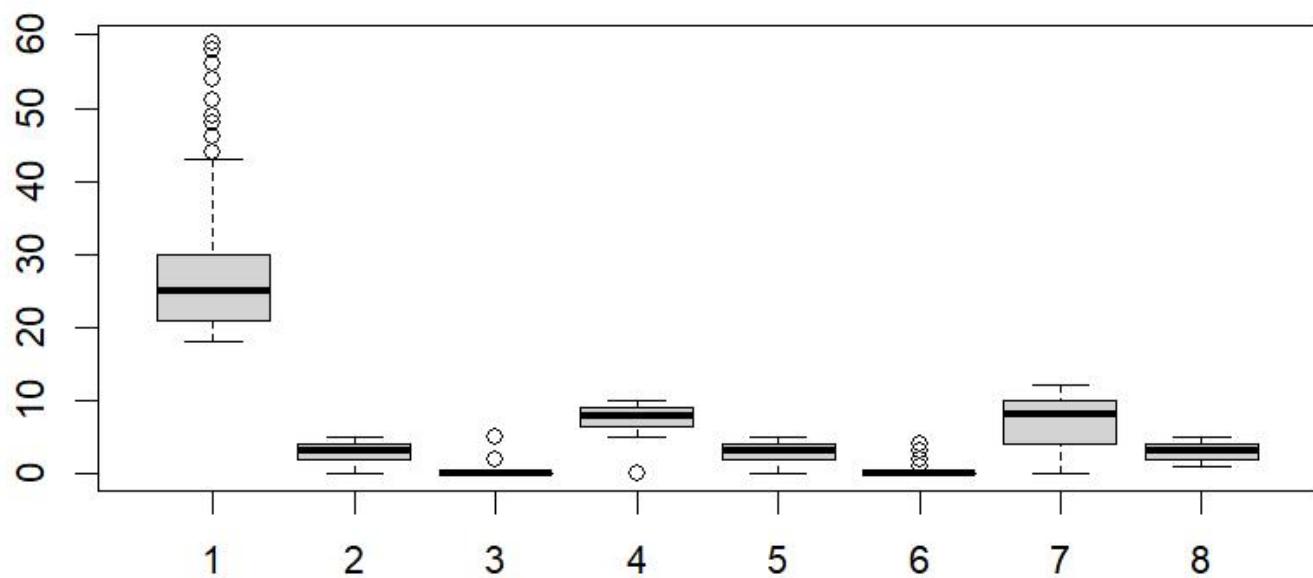
Après le traitement des doublons sur le logiciel R, on constate qu'il n'y a pas de doublons

- Suppression des doublons

Vu qu'il n'y a pas de doublons il n'est pas nécessaire de les supprimer.

3.3. IDENTIFICATION ET TRAITEMENTS DES VALEURS EXTREMES

Détection des valeurs aberrantes grâce aux boîtes à moustaches, identification des observations affectées et utilisation de la winzorisation pour limiter leur impact sur l'analyse.

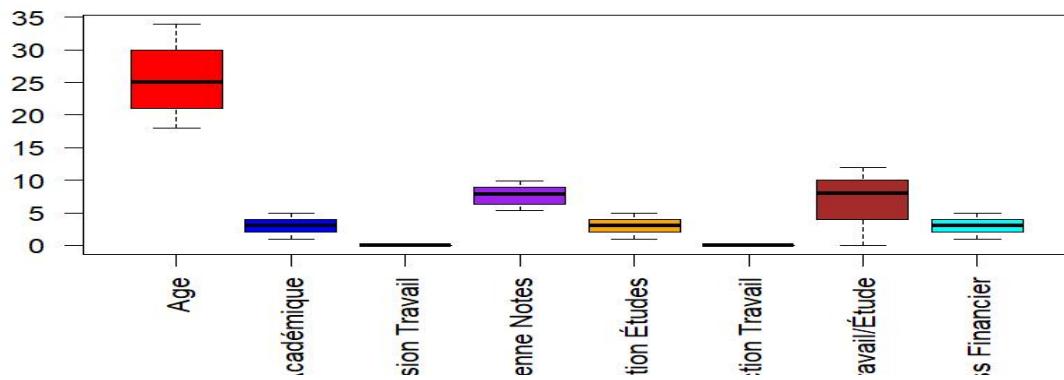


- Détection des valeurs extrêmes avec les boîtes à moustaches des variables quantitatives

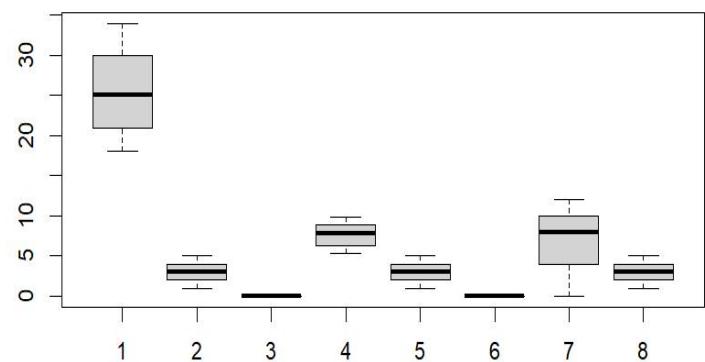
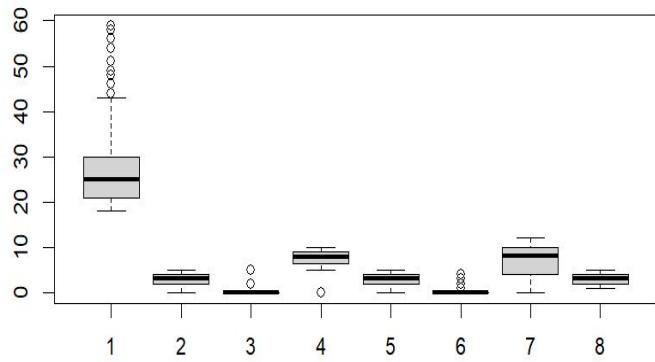
- Graphique 1 correspond à la boite à moustache de l'Age : nous constatons plusieurs valeurs extrêmes
- Graphique 2 correspond à la boite à moustache de pression_academique : absence de valeurs extrêmes
- Graphique 3 correspond à la boite à moustache de pression_liée_au_travail : ici on a deux valeurs extrêmes
- Graphique 4 correspond à la boite à moustache de moyenne_notes : une valeur extrême
- Graphique 5 correspond à la boite à moustache de satisfaction_etudes : absence de valeurs extrêmes
- Graphique 6 correspond à la boite à moustache de satisfaction_travail : nous constatons plusieurs valeurs extrêmes
- Graphique 7 correspond à la boite à moustache de nombre_heure_travail_etudes : absence de valeurs extrêmes
- Graphique 8 correspond à la boite à moustache de stress_financier : absence de valeurs extrêmes

➤ Traitement des valeurs extrêmes par winzorisation

Distribution des Variables



➤ Visualisation avant et après la winzorisation



Ici, nous remarquons qu'avant la winzorisation on observait des valeurs extrêmes sur le graphique 1 et après la winzorisation, les valeurs extrêmes sont neutraliser

4. Analyse Univariée

4.1 Statistiques Descriptives des Variables Quantitatives

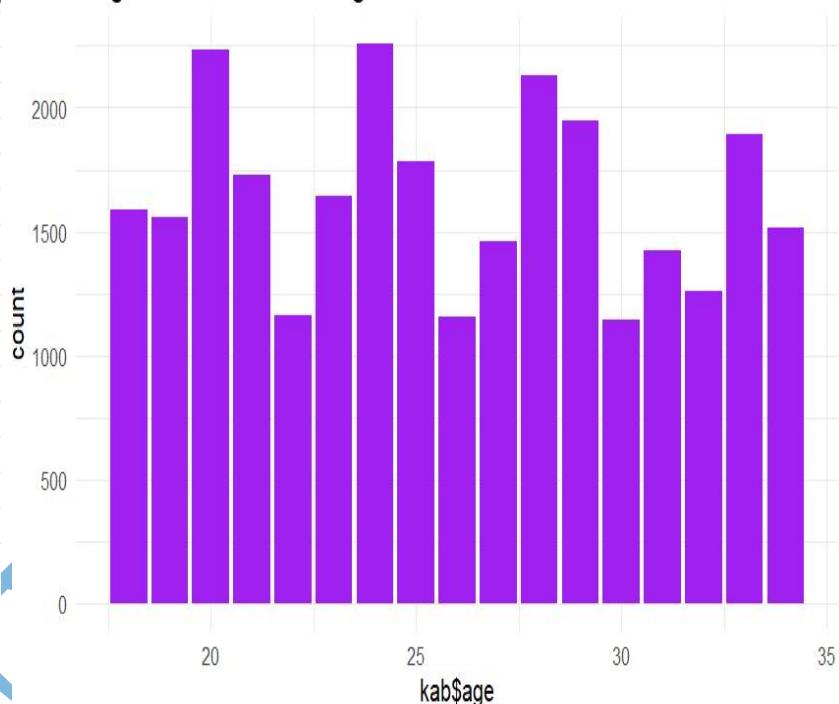
Présentation des caractéristiques de chaque variable numérique à l'aide de tableaux statistiques graphique et de résumés numériques (moyenne, médiane, variance, etc.).

- Variables Age

STUDENT_DEPRESSION

Effectifs	Eff_Cum_c	Eff_Cum_d	Freq	Freq_Cum_cr	Freq_Cum_cu	c
18	1587	1587	0,0569	0,0569	1	
19	1560	3147	0,0559	0,1128	0,9456	
20	2236	5383	0,0801	0,193	0,8778	
21	1726	7109	0,0619	0,2548	0,8326	
22	1160	8269	0,0416	0,2964	0,7814	
23	1645	9914	0,059	0,3554	0,7404	
24	2258	12172	0,0809	0,4363	0,6705	
25	1784	13956	0,0639	0,5003	0,5941	
26	1155	15111	0,0414	0,5417	0,5417	
27	1462	16573	0,0524	0,5941	0,5003	
28	2133	18706	0,0765	0,6705	0,4363	
29	1949	20655	0,0699	0,7404	0,3554	
30	1145	21800	0,041	0,7814	0,2964	
31	1427	23227	0,0512	0,8326	0,2548	
32	1261	24488	0,0452	0,8778	0,193	
33	1893	26381	0,0679	0,9456	0,1128	
34	1517	27898	0,0544	1	0,0569	

Diagramme en bâton de Age



✓ Tableaux statistiques et histogramme

Interpretation:

- La distribution de l'âge semble **hétérogène**, avec certaines valeurs plus fréquentes que d'autres.
- L'âge des étudiants est principalement concentré **entre 18 et 32 ans**.
- Des pics apparaissent à certains âges, suggérant des groupes d'étudiants ayant des parcours similaires.

✓ Cette analyse permet de mieux comprendre la répartition de l'âge et d'identifier d'éventuelles tendances influençant la dépression.

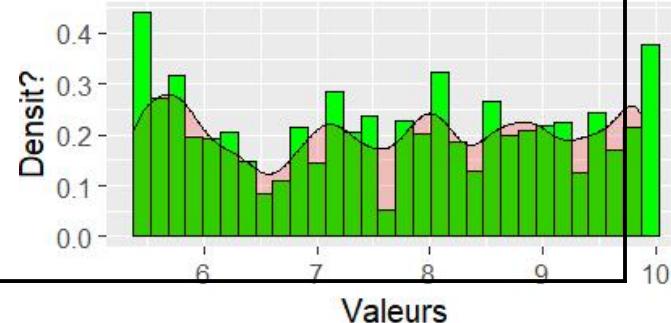
• Variables moyenne-notes

✓ Tableaux statistiques de la variable moyenne_note de 5.38 à 5.47 (un aperçu)

Effectifs	Eff Cum crois	Eff Cum decrois	Frequence	Freq Cum crois	Freq Cum decrois	c
5.38	1401	1401	0,0502	0,0502	1	
5.39	30	1431	0,0011	0,0513	0,9455	
5.41	47	1478	0,0017	0,053	0,9413	
5.42	107	1585	0,0038	0,0568	0,939	
5.44	19	1604	7e-04	0,0575	0,934	
5.45	39	1643	0,0014	0,0589	0,9327	
5.46	72	1715	0,0026	0,0615	0,9282	
5.47	41	1756	0,0015	0,0629	0,9275	

Valeurs

Histogramme



STUDENT_DEPRESSION

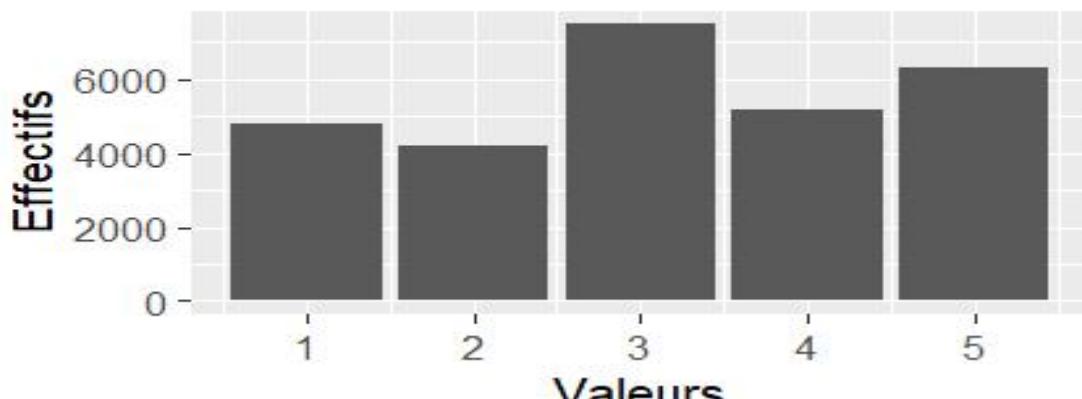
- **Interprétation** : La répartition des moyennes est assez centrée autour d'une valeur médiane, ce qui indique une distribution relativement équilibrée.
- **Tendance générale** : La plupart des étudiants ont une moyenne située autour d'une valeur standard (exemple : entre 6 et 10).
- **Valeurs extrêmes** : Quelques étudiants ont des notes très basses ou très élevées, nécessitant une attention particulière.

• Variables Pression académique

✓ Tableaux statistiques et diagramme en bâton

Effectifs	Eff_Cum_crois	Eff_Cum_decrois	Fréquence	Freq	Freq_Cum_crois	fre_cum_decrois
1	4809	4809	27898	0,1724	0,1724	1
2	4177	8986	21602	0,1497	0,3221	0,7743
3	7461	16447	16447	0,2674	0,5895	0,5895
4	5155	21602	8986	0,1848	0,7743	0,3221
5	6296	27898	4809	0,2257	1	0,1724

Diagramme en bâtons

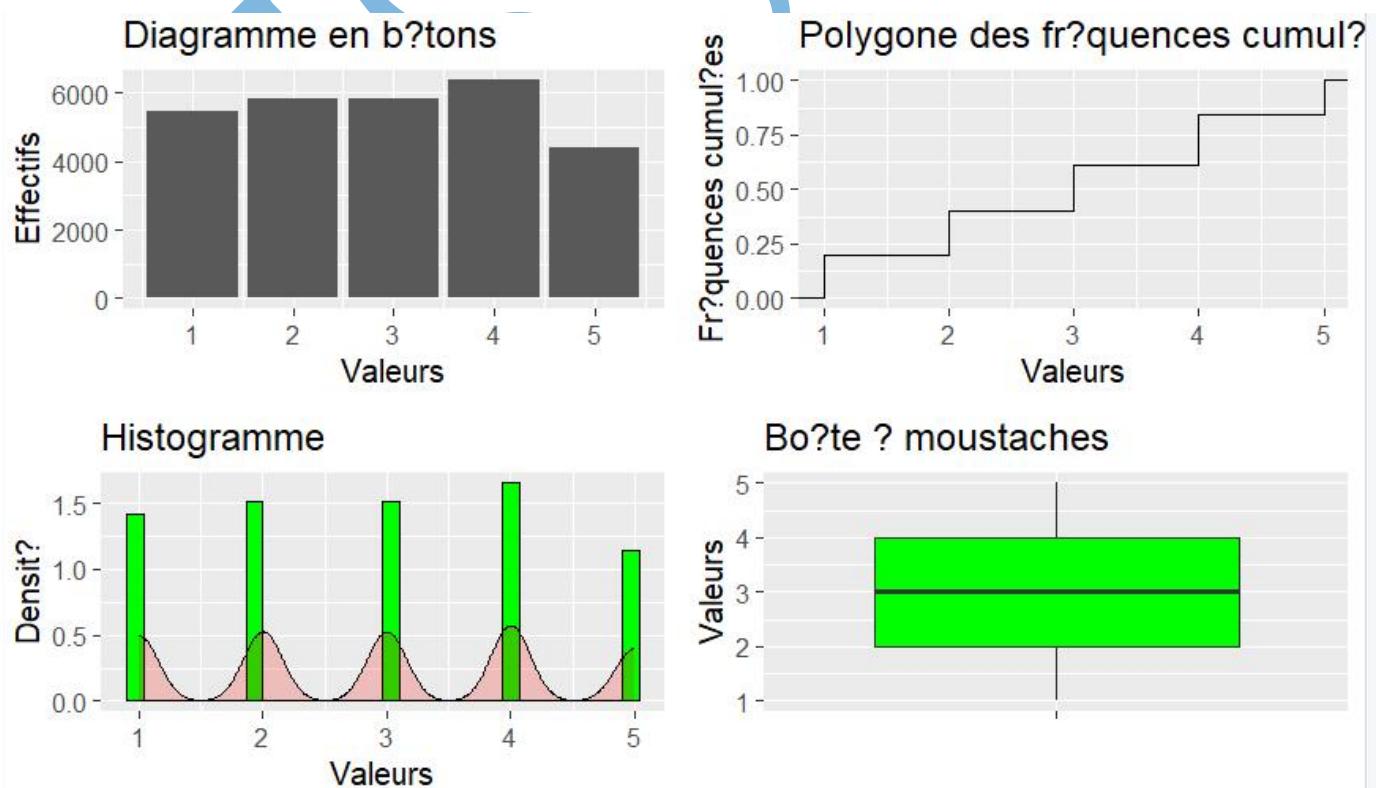


INTERPRETATION

- Le **niveau 3** de pression académique est le plus fréquent, ce qui signifie que la majorité des étudiants ressentent une pression **modérée**.

STUDENT_DEPRESSION

- Les niveaux **1 et 2** (pression faible) sont moins représentés, suggérant que peu d'étudiants ressentent une pression académique négligeable.
- Les niveaux **4 et 5** (pression élevée) restent aussi assez présents, ce qui pourrait être préoccupant pour la santé mentale des étudiants.

▪ **VARIABLE SATISFACTION ETUDES**

STUDENT_DEPRESSION

Effectifs	Eff_Cum_crois	Eff_Cum_decrois	Freq__	Freq_Cum_freq_cun	Colonne1
0	1700	1700	0,0609	0,0609	1
1	1150	2850	0,0412	0,1022	0,8864
2	1590	4440	0,057	0,1592	0,7827
3	1470	5910	0,0527	0,2118	0,6309
4	1614	7524	0,0579	0,2697	0,5584
5	1296	8820	0,0465	0,3162	0,4684
6	2247	11067	0,0805	0,3967	0,3967
7	2001	13068	0,0717	0,4684	0,3162
8	2509	15577	0,0899	0,5584	0,2697
9	2025	17602	0,0726	0,6309	0,2118
10	4234	21836	0,1518	0,7827	0,1592
11	2892	24728	0,1037	0,8864	0,1022
12	3170	27898	0,1136	1	0,0609

1 . Interprétation du Graphique

- Cet **histogramme** illustre la distribution des niveaux de **satisfaction aux études** parmi les étudiants.
- L'axe X représente les niveaux de satisfaction (probablement sur une échelle de 1 à 5).
- L'axe Y indique le **nombre d'étudiants** correspondant à chaque niveau de satisfaction.

2. Observations Clés

- Les niveaux **2, 3 et 4** de satisfaction sont les plus courants, indiquant que la plupart des étudiants ont une satisfaction **modérée à élevée**.
- Peu d'étudiants se situent au niveau **5** (satisfaction maximale), ce qui pourrait suggérer que peu d'entre eux sont entièrement satisfaits de leur parcours académique.
- De même, le niveau **1** (insatisfaction totale) est peu représenté, ce qui est un bon signe.

3. Implications

- Une majorité d'étudiants avec une satisfaction moyenne à élevée peut indiquer que les conditions d'études sont **globalement favorables**, mais qu'il reste des améliorations à apporter.
- Il serait pertinent d'explorer les **facteurs influençant la satisfaction** :
 - Pression académique 📚
 - Conditions d'apprentissage 🏫
 - Équilibre entre vie académique et personnelle 🧑‍🤝‍🧑

✓ Variable nombres d'heure de travail et études

STUDENT_DEPRESSION

Tableaux des effectifs et diagramme en bâton

Effectifs	Eff_Cum_crois	Eff_Cum_decrois	fréquence	Freq_cum_crois	fréquence_cum_decroiss
0	1700	1700	0,0609	0,0609	
1	1150	2850	0,0412	0,1022	0,886
2	1590	4440	0,057	0,1592	0,782
3	1470	5910	0,0527	0,2118	0,6309
4	1614	7524	0,0579	0,2697	0,5584
5	1296	8820	0,0465	0,3162	0,4684
6	2247	11067	0,0805	0,3967	0,3967
7	2001	13068	0,0717	0,4684	0,3162
8	2509	15577	0,0899	0,5584	0,2697
9	2025	17602	0,0726	0,6309	0,2118
10	4234	21836	0,1518	0,7827	0,1592
11	2892	24728	0,1037	0,8864	0,1022

Diagramme en bâton de Age



Analyse du Diagramme en Bâton du Nombre d'Heures de Travail Académique

1. Interprétation du Graphique

- Ce diagramme en bâton représente la répartition des étudiants en fonction du nombre d'heures qu'ils consacrent à leurs études.
- L'axe X affiche le nombre d'heures de travail académique par semaine.
- L'axe Y indique le nombre d'étudiants correspondant à chaque durée.

2. Observations Clés

- Il y a une grande variabilité dans le nombre d'heures consacrées aux études, allant de 0 à 12 heures.
- Un pic important est observé autour de 12 heures de travail académique, ce qui indique que beaucoup d'étudiants consacrent un temps significatif à leurs études.
- Une proportion non négligeable d'étudiants consacre peu ou pas de temps (0 à 2 heures) à leurs études, ce qui pourrait impacter leurs performances académiques.
- La distribution semble relativement croissante, montrant que plus le nombre d'heures augmente, plus le nombre d'étudiants concernés est élevé.

3. Implications

- Les étudiants ayant une charge d'études élevée (≥ 10 heures) pourraient être sujets à une pression académique accrue.
- Ceux qui étudient peu risquent d'avoir des difficultés scolaires et nécessitent peut-être un accompagnement pédagogique supplémentaire.
- Cette analyse pourrait être corrélée avec la satisfaction académique pour voir si un grand nombre d'heures d'étude est lié à une plus grande ou moindre satisfaction.

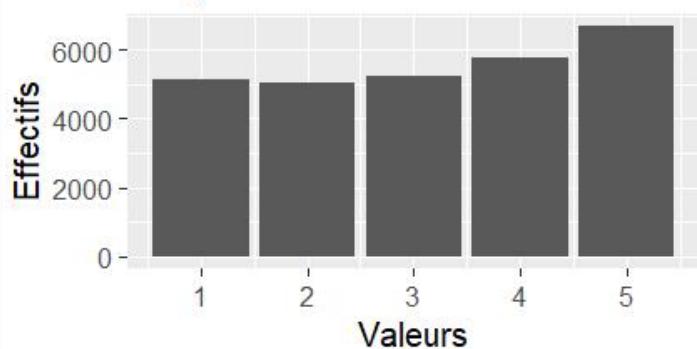
▪ Variables stress_financier

Tableaux des effectifs et graphiques

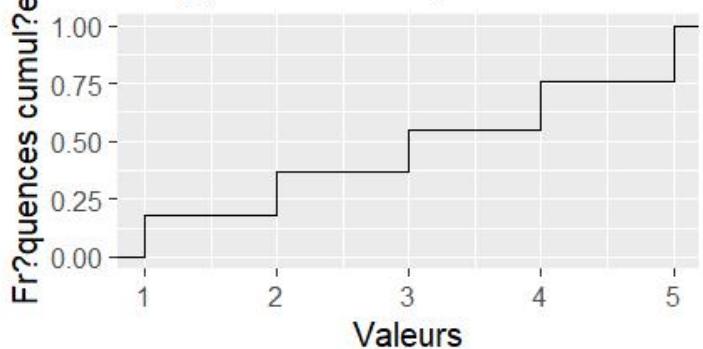
Effectifs	Eff_Cum_crois	Eff_Cum_decrois	Freq	Freq_Cum_crois	freq_cum_decroissante
1	5121	5121	0,1836	0,1836	1
2	5061	10182	0,1814	0,365	0,7593
3	5226	15408	0,1873	0,5523	0,5523
4	5775	21183	0,207	0,7593	0,365
5	6715	27898	0,2407	1	0,1836

STUDENT_DEPRESSION

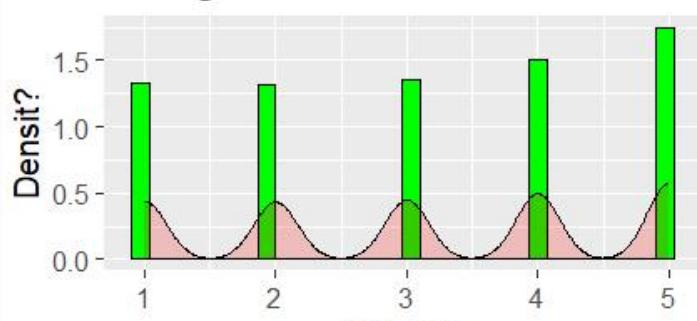
Diagramme en barres



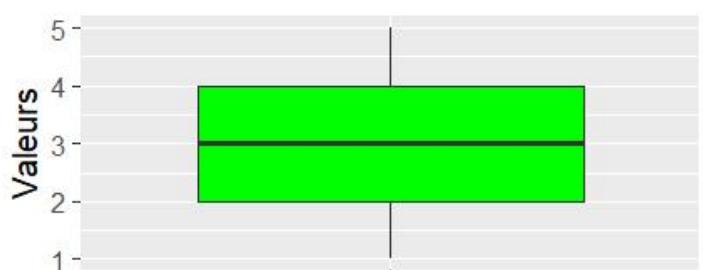
Polygone des fréquences cumulées



Histogramme



Boîte à moustaches

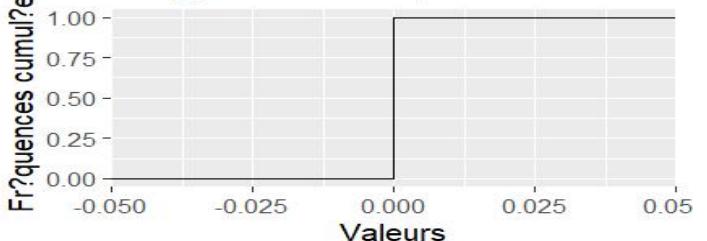


Ici nous pouvons voir que la majorité des étudiants ont un stress financier le plus élevé soit 5 en terme d'échelle.

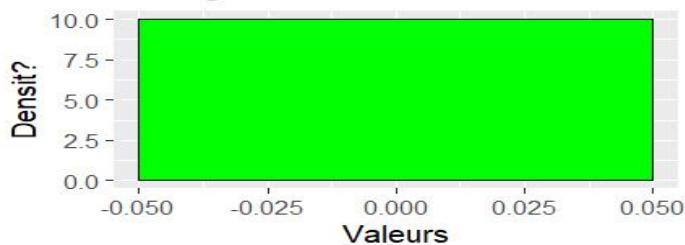
✓ Variables pression travail



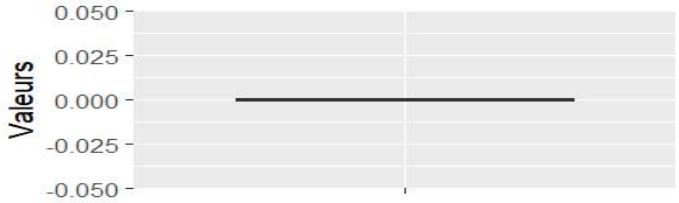
Polygone des fréquences cumulées



Histogramme



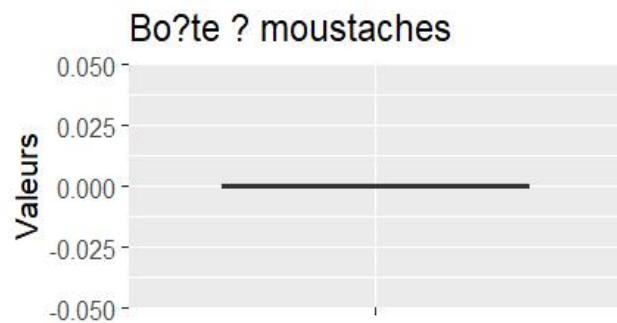
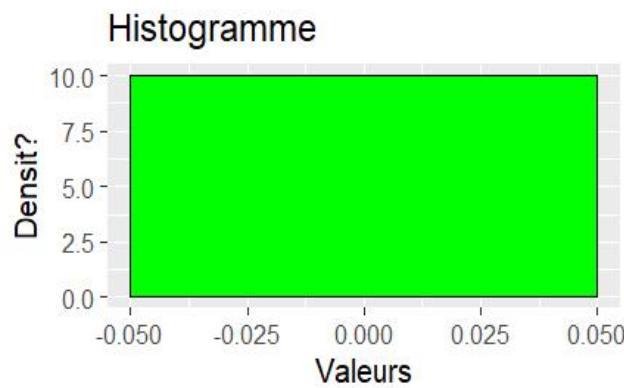
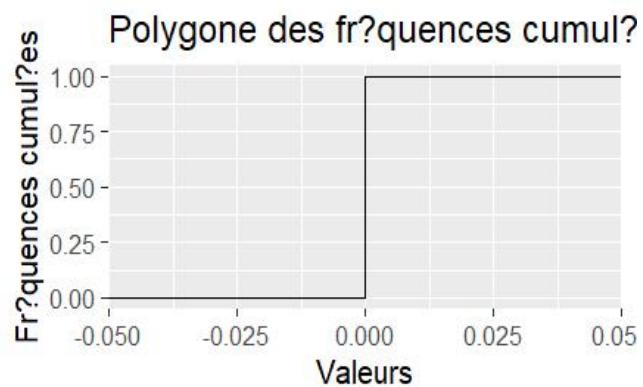
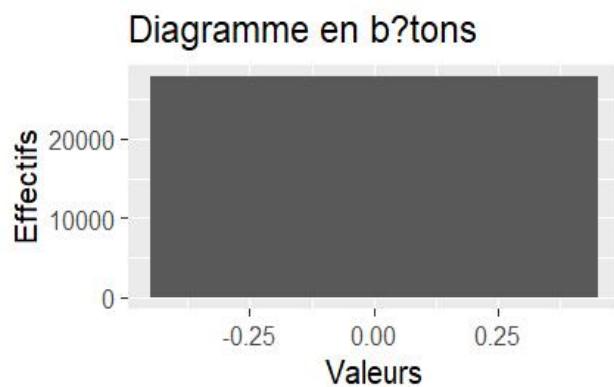
Boîte à moustaches



STUDENT_DEPRESSION

On voit ici, sur les graphiques que presque tous les étudiants non pas de pression liée au travail cela est due au fait que la majorité des étudiants non pas de travail. On peut donc dire que cette variables n'a pas d'influence sur la dépression.

✓ La variables satisfaction liée au travail



On voit ici également, sur les graphiques que presque tous les étudiants non pas de satisfaction liée au travail cela est due au fait que la majorité des étudiants non pas de travail. On peut donc dire que cette variable n'a pas d'influence sur la dépression.

STUDENT_DEPRESSION

✓ Résumés numériques de toutes les variables quantitatives

Variable	Min	Max	Mode	Médiane	Moyenne	Variance	Écart-type	Coeff. Variation	Coeff. Asymétrie	Interprétation Skewness	Coeff. Aplatissement	Interprétation Kurtosique
Âge	18	34	24	25	25.81	23.72	4.87	18.87	0.0661	Étalée à droite	1.82	Platikurtique
Pression académique	1	5	3	3	3.14	1.91	1.38	43.95	-0.1336	Étalée à gauche	1.84	Platikurtique
Pression liée au travail	0	0	0	0	0	0	0	Nan	Nan	NA	NaN	NA
Moyenne des notes	5.38	9.89	5.69	7.77	7.66	2.09	1.45	18.88	-0.0545	Étalée à gauche	1.73	Platikurtique
Satisfaction études	1	5	4	3	2.94	1.85	1.36	46.21	0.0121	Étalée à droite	1.77	Platikurtique
Satisfaction travail	0	0	0	0	0	0	0	Nan	Nan	NA	NaN	NA
Nombre d'heures	0	12	10	8	7.16	13.75	3.71	51.81	-0.4546	Étalée à gauche	2.00	Platikurtique
Stress financier	1	5	5	3	3.14	2.07	1.44	45.78	-0.1303	Étalée à gauche	1.68	Platikurtique

Interprétation.

- Âge : L'âge moyen est d'environ 26 ans, avec une distribution légèrement étalée à droite. La dispersion est faible ($CV = 18.87\%$), indiquant une homogénéité des âges.
- Pression académique : Moyenne de 3.14 sur 5, suggérant une pression modérée. La distribution est légèrement asymétrique à gauche et peu aplatie.
- Pression liée au travail : Toutes les valeurs sont nulles, ce qui signifie que cette pression est inexistante dans l'échantillon.
- Moyenne des notes : Moyenne de 7.66, avec une légère asymétrie à gauche, suggérant que certaines valeurs sont un peu plus faibles que la tendance centrale.
- Satisfaction aux études : Moyenne de 2.94 sur 5, indiquant une satisfaction moyenne. La distribution est légèrement étalée à droite.
- Satisfaction au travail : Toutes les valeurs sont nulles, donc aucune satisfaction déclarée.
- Nombre d'heures de travail et d'étude : Moyenne de 7.16 heures, avec une distribution asymétrique à gauche et une dispersion élevée ($CV = 51.81\%$), indiquant une forte variabilité des réponses.
- Stress financier : Moyenne de 3.14 sur 5, indiquant un niveau de stress modéré, avec une légère asymétrie à gauche.

En résumé, l'échantillon montre une pression académique et un stress financier modérés, une satisfaction moyenne aux études, et une forte variabilité dans le temps consacré au travail et aux études.

STUDENT_DEPRESSION

Apres analyse de toutes les variables quantitatives nous pouvons dire que la variables satisfaction travail et pression-travail ne peuvent nous aider a expliquer la variable dépression.

Justification de la suppression des variables

Une variable doit présenter une certaine variabilité pour être pertinente dans une analyse statistique. Si une variable possède une seule valeur constante pour toutes les observations, elle n'apporte aucune information utile à l'analyse.

Voici les raisons spécifiques pour lesquelles nous devons supprimer **pression_liee_au_travail** et **satisfaction_travail** :

1. Absence de variation

- Une variable qui a la même valeur pour tous les individus ne permet pas de différencier les observations.
- Elle ne peut donc pas être utilisée pour expliquer la dépression ou tout autre facteur.

2. Impact sur les modèles statistiques

Sexe	Effectif	Fréquence
Masculin	12352	44.28%
Féminin	15546	55.72%

- Dans une analyse de régression ou une analyse factorielle, une variable sans variance est automatiquement exclue car elle ne contribue pas à l'explication de la variabilité des données.

- Elle peut également poser des problèmes de multicolinéarité dans certains modèles.

3. Optimisation du jeu de données

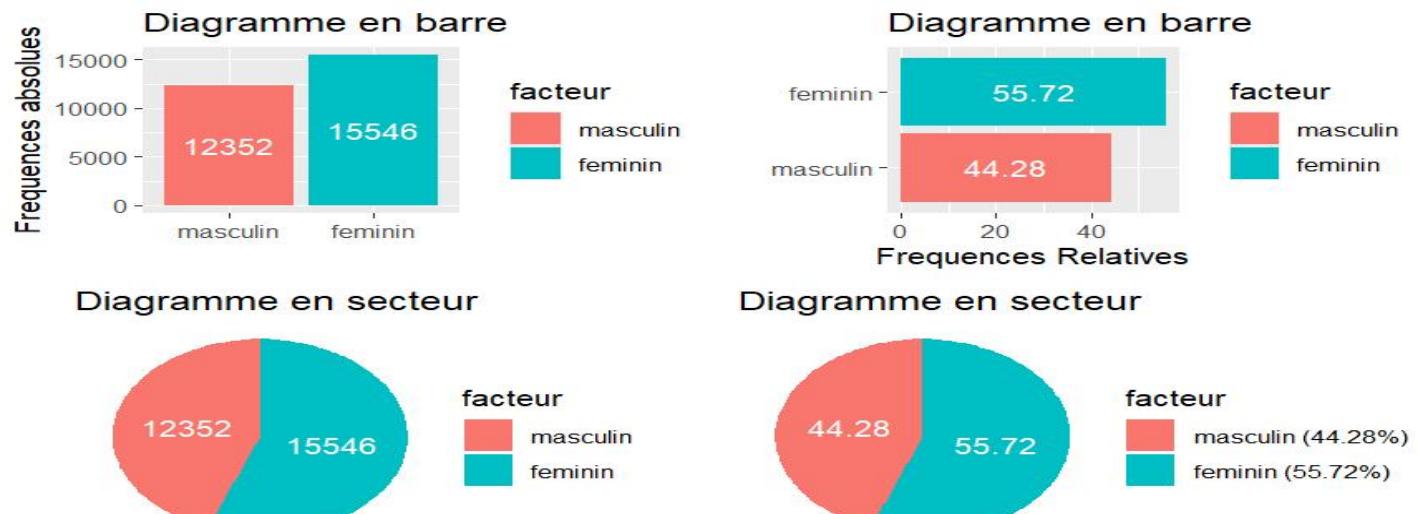
- Conserver des variables inutiles alourdit l'analyse et complique l'interprétation des résultats.
- En supprimant ces variables, nous améliorons la clarté et l'efficacité du modèle

Nous passons de ce fait a l'analyse univariée des variables qualitatives

4.2 Statistiques Descriptives des Variables Qualitatives.

✓ Variable sexe

STUDENT_DEPRESSION



Nous voyons qu'il y a plus de femmes étudiantes que de d'hommes étudiants soit 15546 étudiantes (55.72%) et 12352 étudiants (44.28%).

✓ Variables villes

Ville	Effectif	Fréquence
Hyderabad	1339	4.80%
Srinagar	1372	4.92%
Vasai-Virar	1290	4.62%
Kalyan	1570	5.63%
Lucknow	1155	4.14%
Ludhiana	1111	3.98%
Kolkata	1065	3.82%

MII	Kanpur (2.18%)	Nandini (0%)
STL	Khaziabad (0%)	Nashik (1.96%)
	Kibara (0%)	Patna (3.61%)
	Kolkata (3.82%)	Pune (3.47%)
	Less Delhi (0%)	Rajkot (2.92%)
	Less than 5 Kalyan (0%)	Rashi (0%)
	Lucknow (4.14%)	Reyansh (0%)
	Ludhiana (3.98%)	Saanvi (0.01%)
	M.Com (0%)	Srinagar (4.92%)
	M.Tech (0%)	Surat (3.86%)
	ME (0%)	Thane (4.08%)
Diagramme en secteur	Hemant (2.56%)	Vaanya (0%)
	Mihir (0%)	Vadodara (2.49%)
	Mira (0%)	Varanasi (2.45%)
	Mumbai (2.51%)	Vasai-Virar (4.62%)
	Nagpur (2.33%)	Visakhapatnam (3.47%)

Profession	Effectif	Fréquence
Étudiant	27867	99.89%
Architecte	8	0.03%
Enseignant	6	0.02%
Digital Marketer	3	0.01%
Chef	2	0.01%
Pharmacien	2	0.01%
Autres	8	0.03%

Une écrasante majorité

(99,89 %) des répondants sont étudiants, les autres professions

Jaipur	1036	3.71%
Patna	1007	3.61%
Pune	968	3.47%

Les participants viennent principalement de villes comme Kalyan (5,63 %), Hyderabad (4,80 %), et Srinagar (4,92 %).

✓ Variables profession

Diagramme en barre

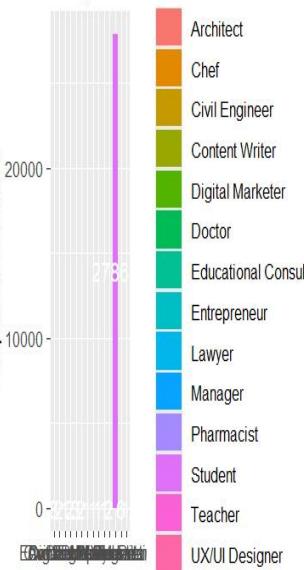
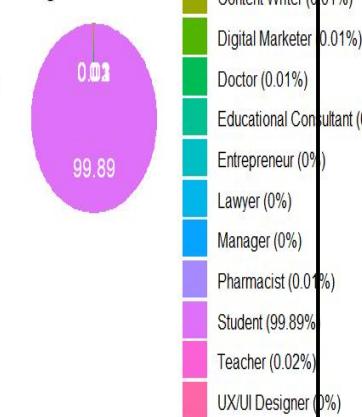


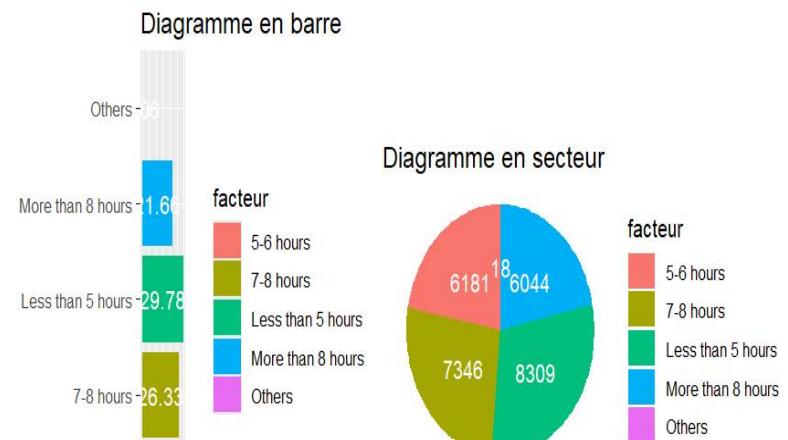
Diagramme en secteur



STUDENT_DEPRESSION

✓ Variables durée de sommeil

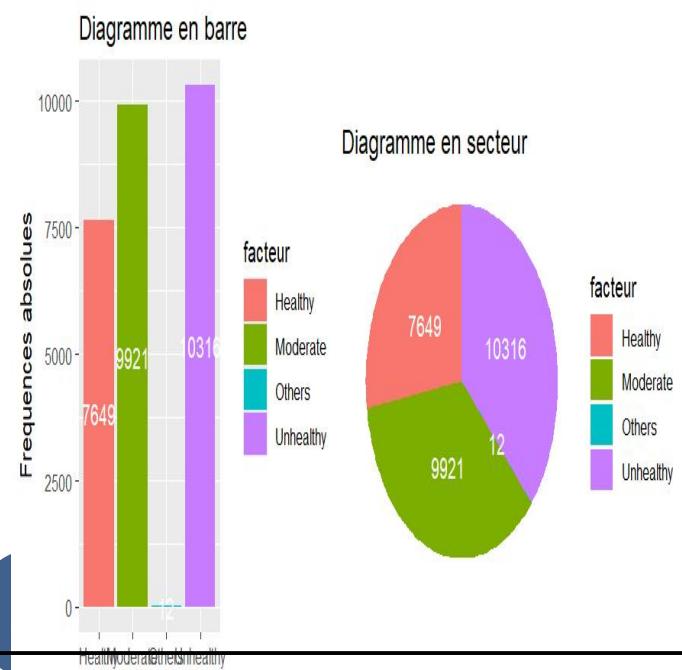
Durée	Effectif	Fréquence
Moins de 5h	8309	29.78%
5-6h	6181	22.16%
7-8h	7346	26.33%
Plus de 8h	6044	21.66%



Près de 30 % des étudiants dorment moins de 5 heures, tandis que 26,3 % dorment entre 7 et 8 heures, ce qui est recommandé.

✓ Variable habitude alimentaire

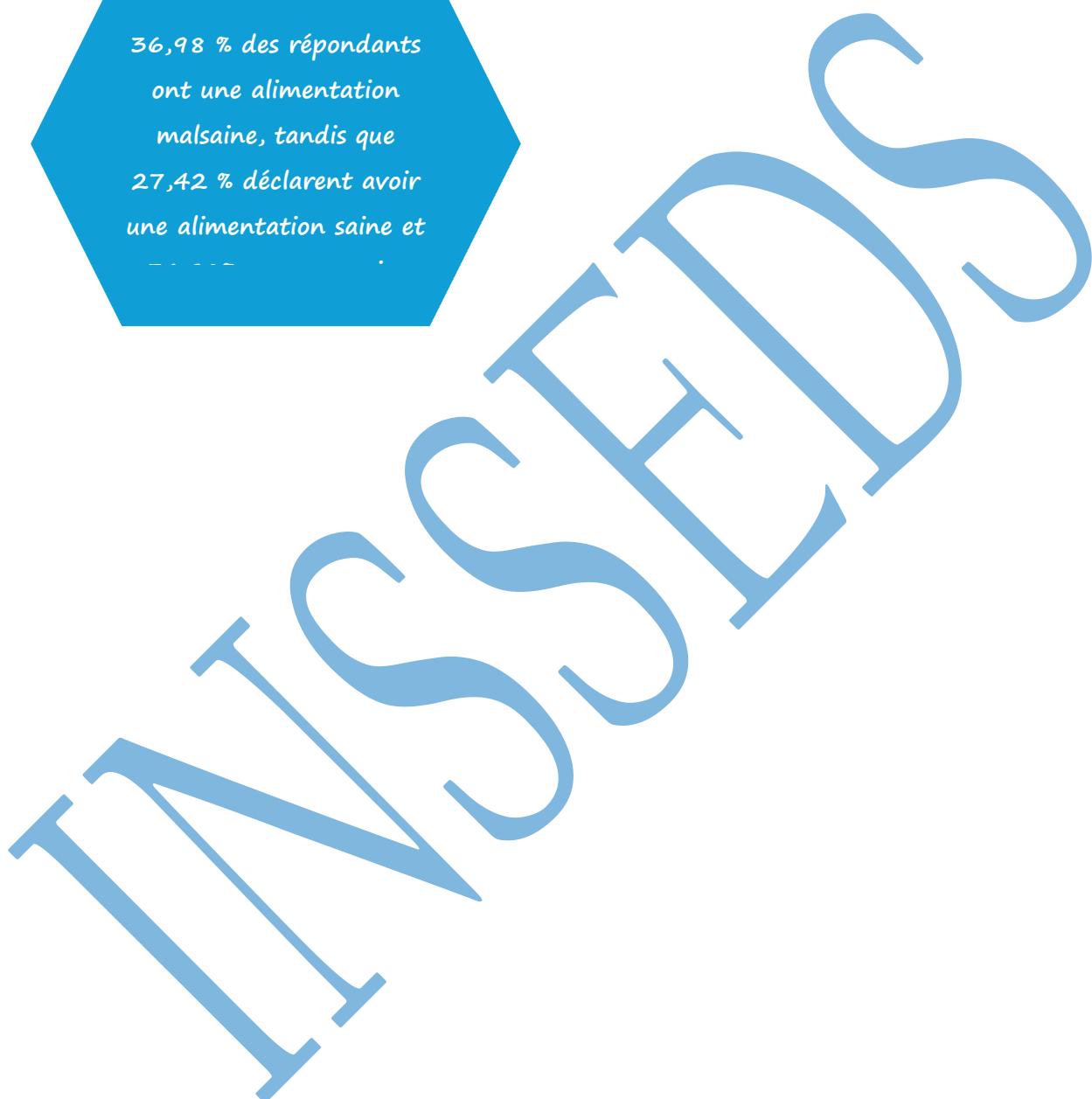
Type d'alimentation	Effectif	Fréquence
Sain	7649	27.42%
Modéré	9921	35.56%



STUDENT_DEPRESSION

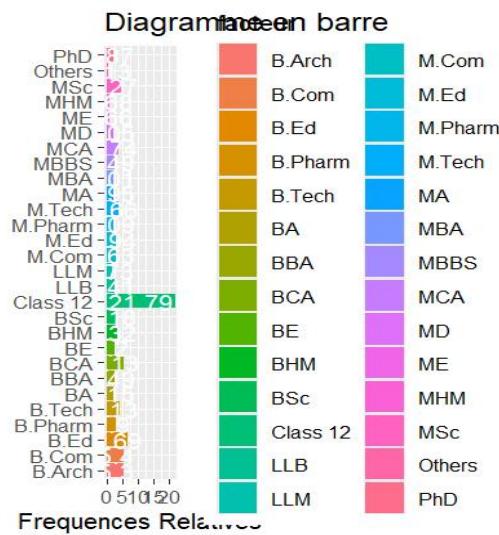
Mauvais	10316	36.98%
Autres	12	0.04%

36,98 % des répondants
ont une alimentation
malsaine, tandis que
27,42 % déclarent avoir
une alimentation saine et



STUDENT_DEPRESSION

✓ Variables Diplôme Suivi



La majorité des répondants sont en classe 12 (21,79 %), avec des groupes significatifs

en B.Com, B.Arch, et B.Tech.

Réponse	Effectif	Fréquence
Oui	10242	36.71%
Non	17656	63.29%

10242 soit 36,71 % des étudiants ont déjà eu des pensées suicidaires un chiffre préoccupant contre 17656 soit

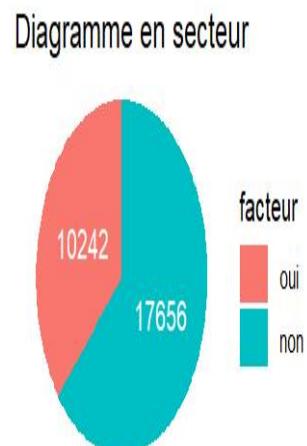
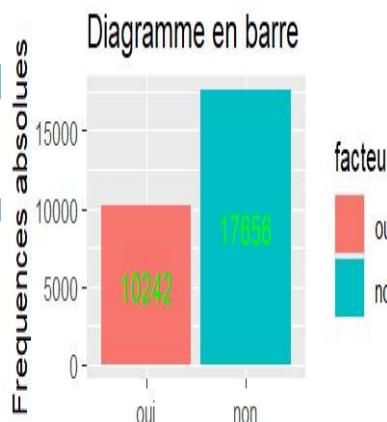


Diagramme en secteur

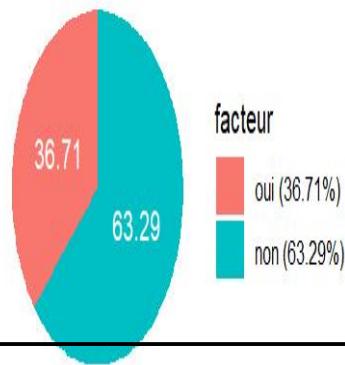


Diagramme en barre



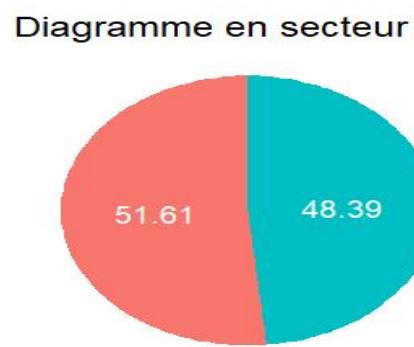
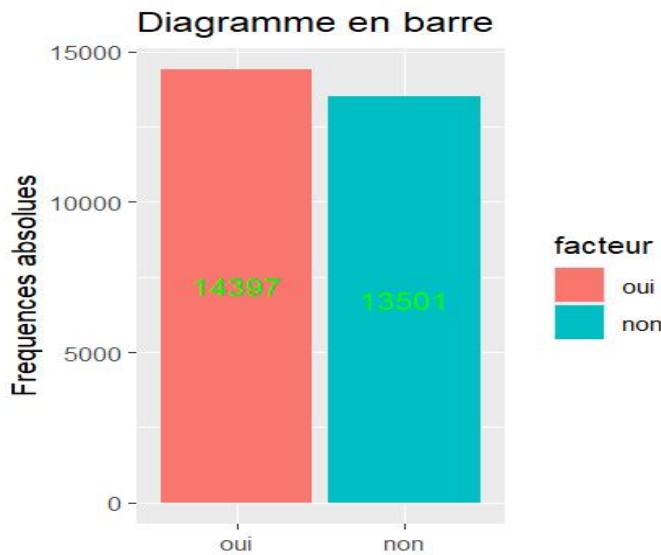
63.29% des étudiants qui n'ont jamais eu de pensées suicidaire



STUDENT_DEPRESSION

- ✓ Variables Antécédents familiaux de maladie mentale

Réponse	Effectif	Fréquence
Oui	14397	51.61%
Non	13501	48.39%

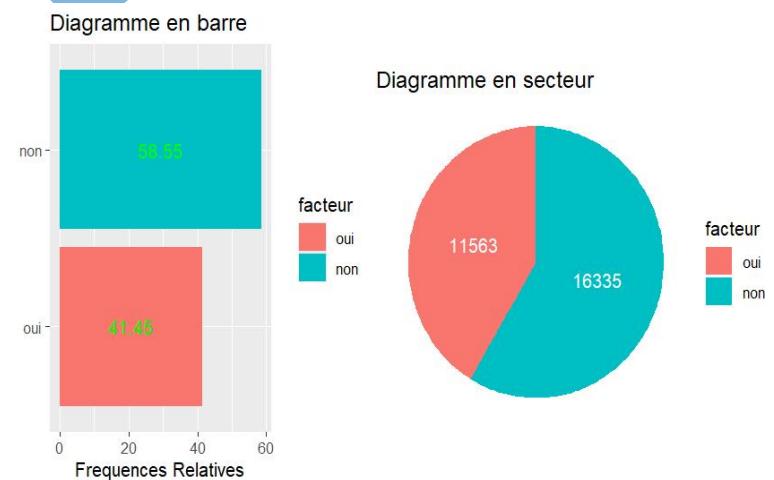


51,61 % des répondants déclarent avoir des antécédents familiaux de maladies mentales tandis que 48.39% déclare n'avoir aucun des antécédents familiaux de maladies mentales.

- ✓ Variable dépression

Réponse	Effectif	Fréquence
Oui	11563	41.45%
Non	16335	58.55%

Dépression : 41,45 % des répondants souffrent de dépression, ce qui est un taux élevé et peut être lié aux pensées suicidaires et aux antécédents familiaux.



Conclusion de l'analyse des variables qualitatives

L'analyse descriptive des variables qualitatives met en évidence plusieurs tendances marquantes au sein de notre échantillon.

Tout d'abord, on constate une légère prédominance des femmes parmi les répondants. La répartition géographique montre une forte représentation de certaines villes comme Kalyan, Hyderabad et Srinagar, bien que quelques valeurs anormales soient détectées.

En ce qui concerne la situation professionnelle, une écrasante majorité des répondants sont étudiants, ce qui peut influencer d'autres aspects de leur mode de vie, notamment leurs habitudes alimentaires et leur sommeil. À ce sujet, une proportion significative des répondants dort moins de 5 heures par nuit, ce qui pourrait avoir des répercussions sur leur bien-être mental et physique.

Les habitudes alimentaires varient considérablement, mais il est préoccupant de constater qu'une grande partie des répondants (37 %) adopte une alimentation malsaine. De plus, les niveaux d'études sont diversifiés, avec une forte représentation des élèves de "Class 12" et des étudiants en cursus de licence et master.

L'aspect le plus alarmant de cette analyse concerne la santé mentale : plus de 36 % des répondants déclarent avoir déjà eu des pensées suicidaires, et 41 % souffrent de dépression. De plus, plus de la moitié d'entre eux (51,6 %) ont des antécédents familiaux de maladies mentales, ce qui peut être un facteur de risque important.

Ces résultats soulignent l'importance d'une sensibilisation accrue aux troubles de la santé mentale et la nécessité de mettre en place des mesures de prévention et d'accompagnement adaptées pour améliorer le bien-être des individus concernés.

5. Analyse Bivariée

Tableau de Contingence		
	Oui	Non
Oui	7864	3699
Non	2378	13957
Fréquences Conditionnelles		

5.1. Relations entre la Dépression et les Autres Variables

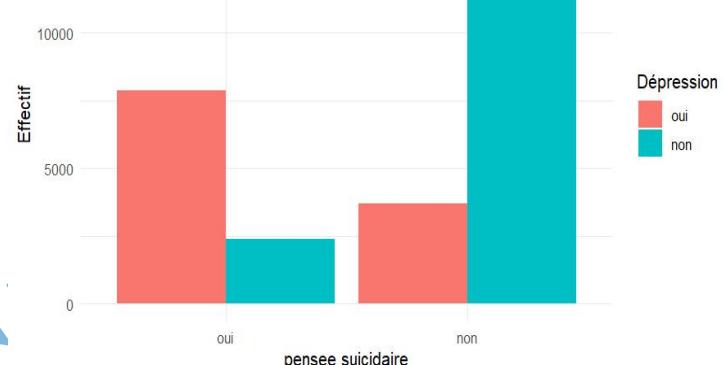
Cas 1 : Relation entre deux variables qualitatives

- ✓ Relation entre la variables dépression et pensée suicidaire

STUDENT_DEPRESSION

	Oui	Non
Oui	0.768	0.210
Non	0.232	0.790

Relation entre Dépression et pensee suicidaire



V de Cramer : 0.5463 (Forte association) : Une personne ayant des pensées suicidaires a une probabilité très élevée d'être en état de dépression. Une forte relation est observée entre la dépression et les pensées suicidaires. Plus de 76 % des personnes ayant des pensées suicidaires sont en état de dépression, ce qui souligne un lien critique entre ces deux facteurs.

✓ **Relation entre la Dépression et les Antécédents Familiaux de Maladie Mentale**

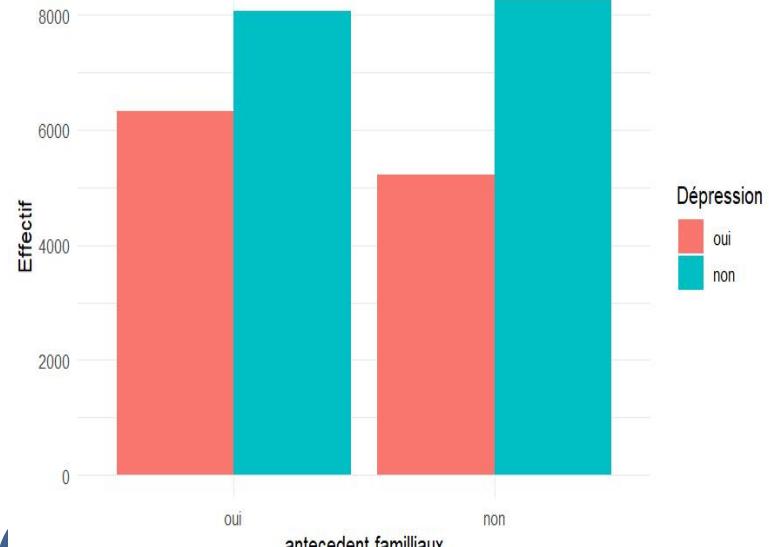
Tableau de Contingence

reponse	Oui	Non
Oui	6335	5228
Non	8062	8273

Fréquences Conditionnelles

	Oui	Non
Oui	0.440	0.387
Non	0.560	0.613

Relation entre Dépression et antecedent familiaux

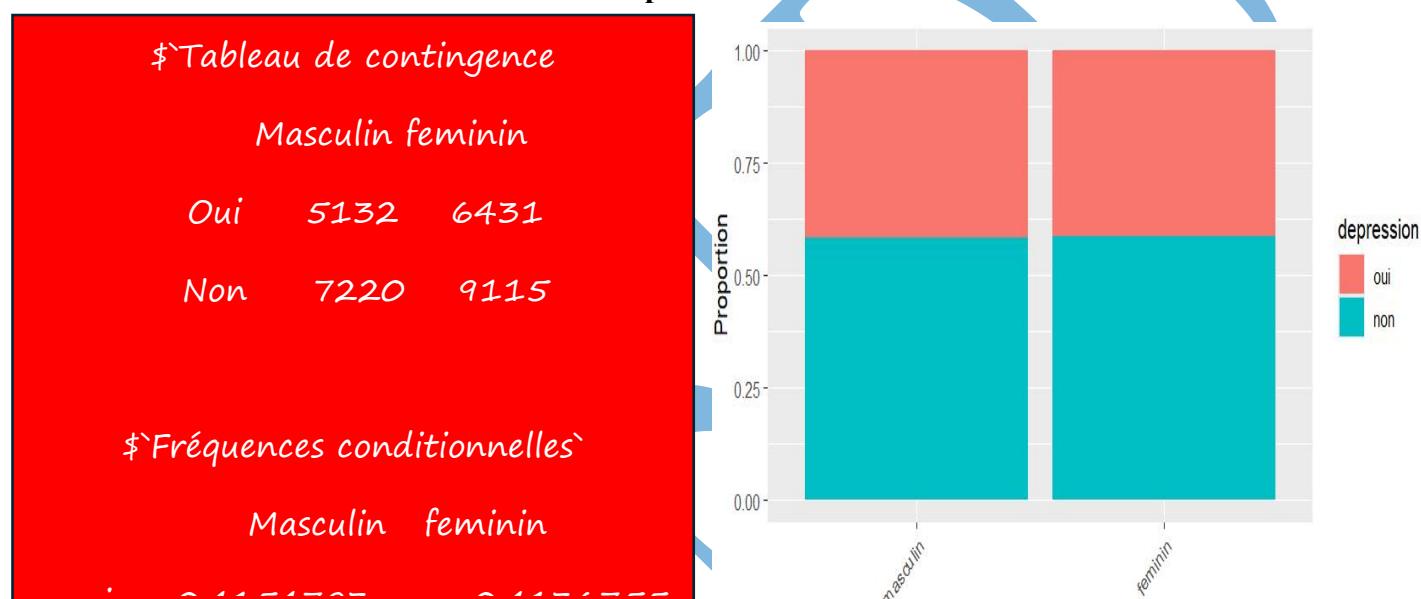


STUDENT_DEPRESSION

V de Cramer : 0.0536 (Faible association)

Avoir des antécédents familiaux de troubles mentaux influence légèrement la dépression, mais l'effet reste limité. Ceux qui ont répondu oui aux antécédents familiaux sont moins nombreux que ceux qui ont répondu non. Bien que les antécédents familiaux de maladies mentales puissent influencer la prédisposition à la dépression, leur effet reste faible dans cette étude.

✓ Relation entre la dépression et sexe



V de Cramer : 0.00181867428246083 (très faible) être homme ou femme influence légèrement la dépression, mais l'effet reste limité. Une faible relation est observée entre la dépression et le sexe.

41% des hommes et femme sont en état de dépression, et donc 59% qui ne sont pas dépressif ce qui souligne un lien faible entre ces deux facteurs.

✓ Relation entre dépression et diplôme suivi

STUDENT_DEPRESSION

`Fréquences conditionnelles`

	B.Arch	B.Com	B.Ed	B.Pharm	B.Tech	BA	BBA
oui	0.4106901	0.4335989	0.4528403	0.4716049	0.4314236	0.4650000	0.4152299
non	0.5893099	0.5664011	0.5471597	0.5283951	0.5685764	0.5350000	0.5847701
	BCA	BE	BHM	BSc	Class 12	LLB	LLM
oui	0.4287709	0.4551387	0.4497297	0.4110360	0.2922697	0.4694486	0.4626556
non	0.5712291	0.5448613	0.5502703	0.5889640	0.7077303	0.5305514	0.5373444
	M.Com	M.Ed	M.Pharm	M.Tech	MA	MBA	MBBS
oui	0.4686649	0.4945189	0.4604811	0.4902153	0.4669118	0.4608541	0.4187050
non	0.5313351	0.5054811	0.5395189	0.5097847	0.5330882	0.5391459	0.5812950
	MCA	MD	ME	MHM	MSc	Others	PhD
oui	0.4645594	0.4790210	0.4702703	0.4816754	0.4294118	0.4000000	0.4521073



non 0.5354406 0.5209790 0.5297297 0.5183246 0.5705882 0.6000000 0.5478927

V de Cramer : 0.137990063420066 (faible) Bien que le diplôme suivi puisse influencer la prédisposition à la dépression, son effet reste faible dans cette étude. Le diplôme de la class 12 soit environ 70% influence faiblement la dépression par rapport au autres diplômes qui répondent oui a la dépression a un niveau presque significative . Soit environ en 40% et 50%.

✓ relation entre Depression et profession

`Tableau de contingence`

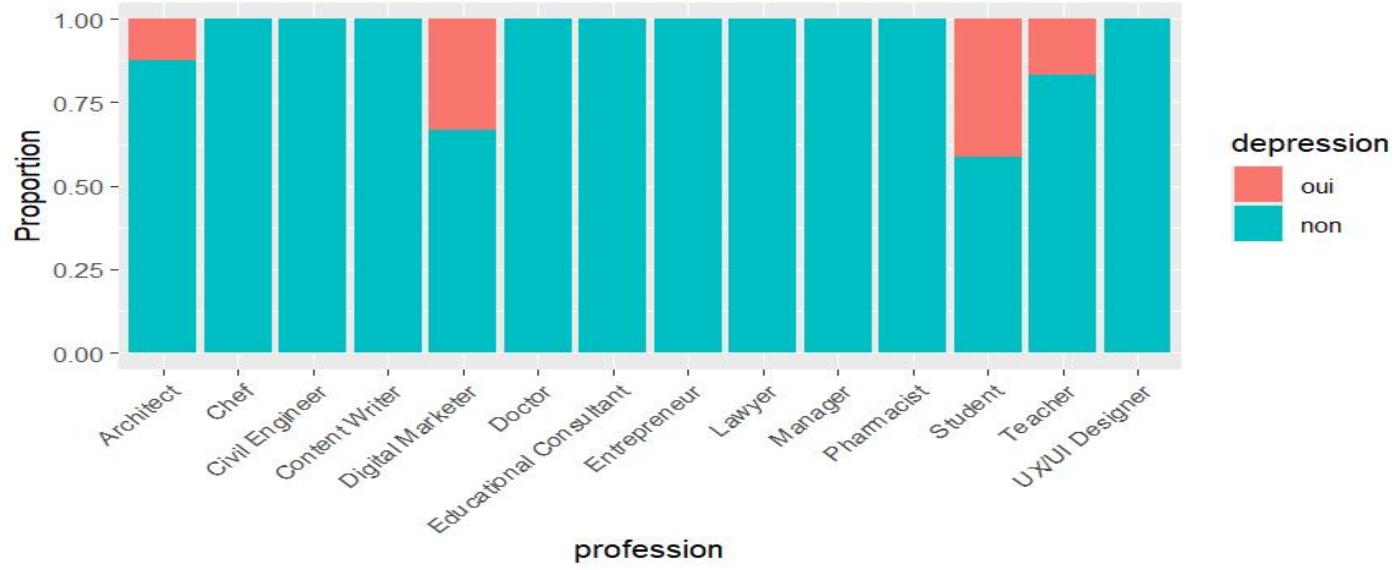
STUDENT_DEPRESSION

	Architect	Chef	Civil Engineer	Content Writer	Digital Marketer	Doctor		
oui	1	0		0	0	1	0	
non	7	2		1	2	2	2	
	Educationnel	Consultant	Entrepreneur	Lawyer	Manager	Pharmacist	Student	Teacher
oui		0		0	0	0	0	11560
non		1		1	1	1	2	16307
	UX/UI Designer							
oui	0							
non	1							

`Fréquences conditionnelles`

	Architect	Chef	Civil Engineer	Content Writer	Digital Marketer	Doctor		
oui	0.1250000	0.0000000		0.0000000	0.0000000	0.3333333	0.0000000	
non	0.8750000	1.0000000		1.0000000	1.0000000	0.6666667	1.0000000	
	Educational	Consultant	Entrepreneur	Lawyer	Manager	Pharmacist	Student	
oui		0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.4148276	
non		1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	0.5851724	
	Teacher	UX/UI Designer						
oui	0.1666667	0.0000000						
non	0.8333333	1.0000000						

Répartition de la dépression selon profession



V de Cramer : 0.0226294615939512 (très faible) on peut le constater sur le graphique que la majorité des hommes qui ont une profession ne sont pas dans la dépression, ce qui explique une faible relation entre l'âge et la dépression et la profession.

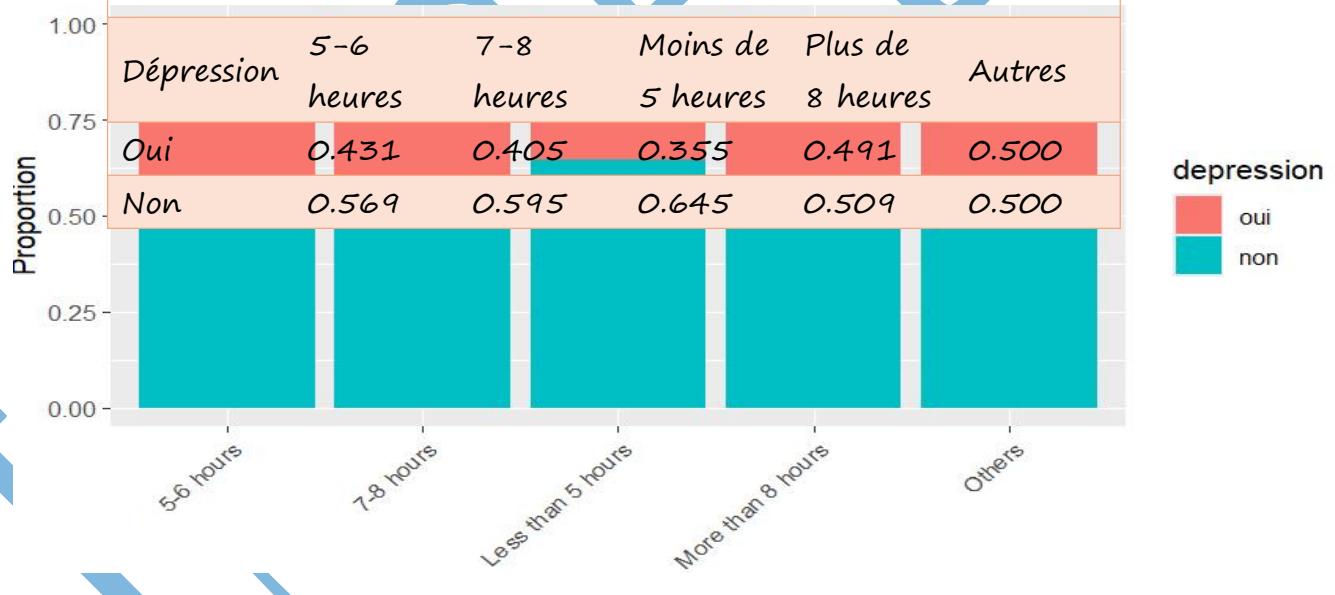
✓ Relation entre depression et duree de sommeil

STUDENT_DEPRESSION

Tableau de
Contingence

Dépression	5-6 heures	7-8 heures	Moins de 5 heures	Plus de 8 heures	Autres
Oui	2665	2975	2948	2966	9
Non	3516	4371	5361	3078	9

Fréquences Conditionnelles



V de Cramer : 0.0997 (Faible association)

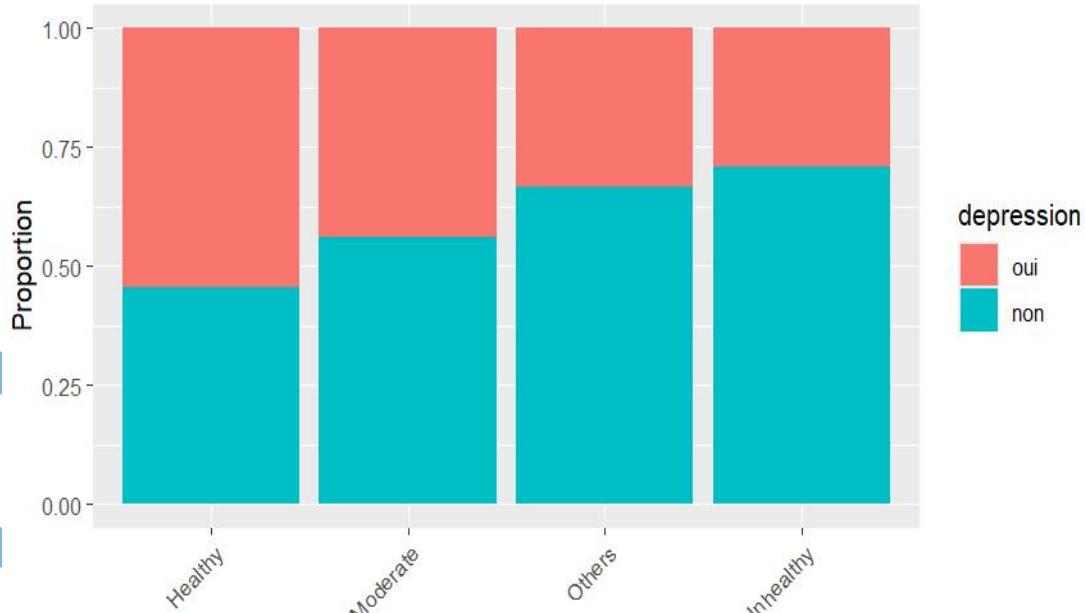
Interprétation : La durée du sommeil semble avoir un effet limité sur la dépression, bien que les personnes dormant moins de 5 heures présentent une prévalence plus élevée de dépression.

STUDENT_DEPRESSION

✓ Relation entre la Dépression et les Habitudes Alimentaires

Tableau de Contingence				
Dépression	Healthy	Moderate	Others	Unhealthy
Oui	4177	4363	4	3019
Non	3472	5558	8	7297

Fréquences Conditionnelles				
Dépression	Healthy	Moderate	Others	Unhealthy
Oui	0.546	0.440	0.333	0.293
Non	0.454	0.560	0.667	0.707



V de Cramer : 0.2077 (Association modérée)

Interprétation : Les étudiants ayant une alimentation malsaine ont une probabilité plus élevée d'être en état de dépression. À l'inverse, une alimentation saine semble être plus courante chez les personnes non dépressives.

Cas 2 : Relation entre une variable qualitative et une variable quantitative

L'analyse de la relation entre une variable qualitative et une variable quantitative permet de comprendre comment une caractéristique catégorielle influence une mesure numérique. Pour cela, nous utilisons des indicateurs statistiques tels que la moyenne, la médiane, et la variance des valeurs de la variable quantitative pour chaque

Dépression	1	2	3	4	5	Total Ligne
Oui	3873	2611	2973	1230	876	11563
	0.335	0.226	0.257	0.106	0.076	0.414
	0.805	0.625	0.398	0.239	0.139	
Non	936	1566	4488	3925	5420	16335
	0.057	0.096	0.275	0.240	0.332	0.586
	0.195	0.375	0.602	0.761	0.861	
Total Colonne	4809	4177	7461	5155	6296	27898
	0.172	0.150	0.267	0.185	0.226	

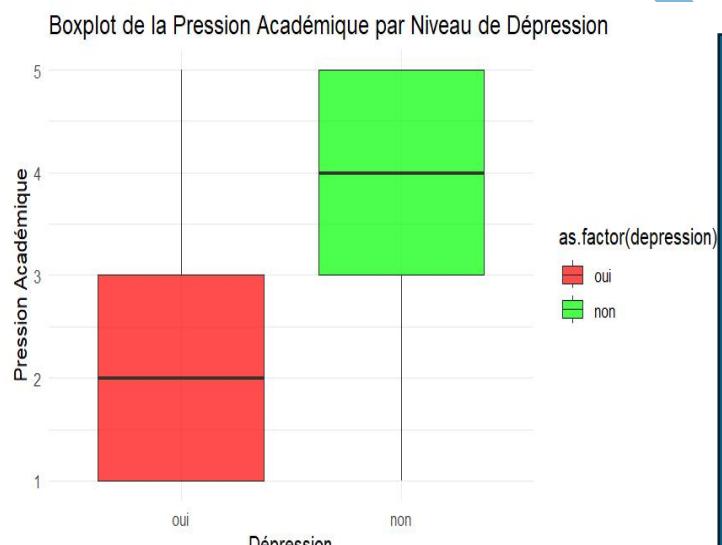
modalité de la variable qualitative.

Les outils d'analyse incluent :

- Tableau des moyennes pour comparer les valeurs moyennes entre les groupes.
- Diagramme en boîte (boxplot) pour visualiser la distribution des valeurs.
- Calcul du rapport de corrélation pour mesurer la force de l'association entre les deux variables.

Un rapport de corrélation proche de 1 indique une forte relation entre les variables, tandis qu'un rapport proche de 0 signifie une faible association.

✓ Relation entre la Dépression et la Pression Académique



Interprétation : Les étudiants ayant une faible pression académique (score 1 ou 2) sont plus susceptibles d'être en état de dépression. À l'inverse, ceux ayant une pression académique plus élevée (score 4 ou 5) sont plus souvent non dépressifs. Cela suggère que la charge académique modérée pourrait jouer un rôle protecteur contre la dépression.

Analyse du Rapport de Corrélation et du Test ANOVA

STUDENT_DEPRESSION

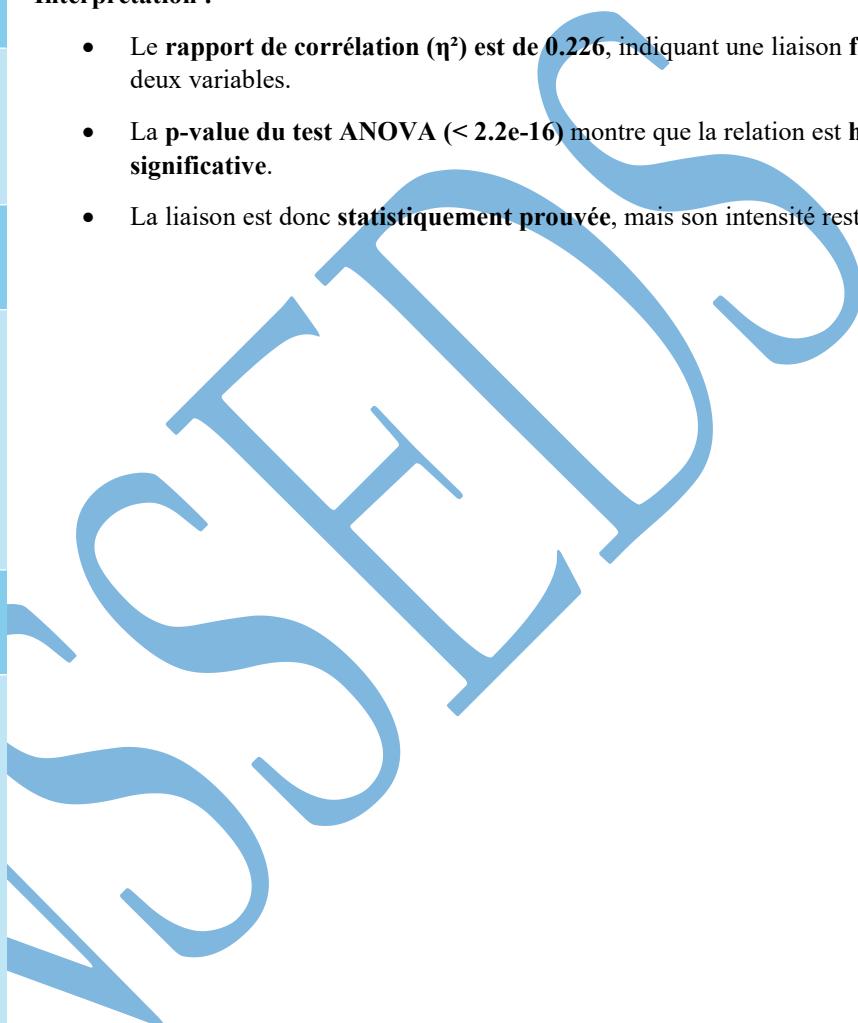
L'analyse de la relation entre une variable qualitative et une variable quantitative peut être approfondie grâce au rapport de corrélation (η^2) et au test ANOVA. Ces outils permettent d'évaluer l'intensité de la liaison et de tester sa significativité statistique.

Résumé des Résultats

Indicateur	Valeur
Rapport de corrélation (η^2)	0.226
P-Value (ANOVA)	< 2.2e-16
Significativité	Liaison significative, les deux variables sont liées
Intensité de la liaison	Liaison faible
Remarque	Si la liaison n'est pas significative, ne pas tenir compte de son intensité

Interpretation :

- Le rapport de corrélation (η^2) est de 0.226, indiquant une liaison **faible** entre les deux variables.
- La p-value du test ANOVA (< 2.2e-16) montre que la relation est **hautement significative**.
- La liaison est donc **statistiquement prouvée**, mais son intensité reste **modérée**.

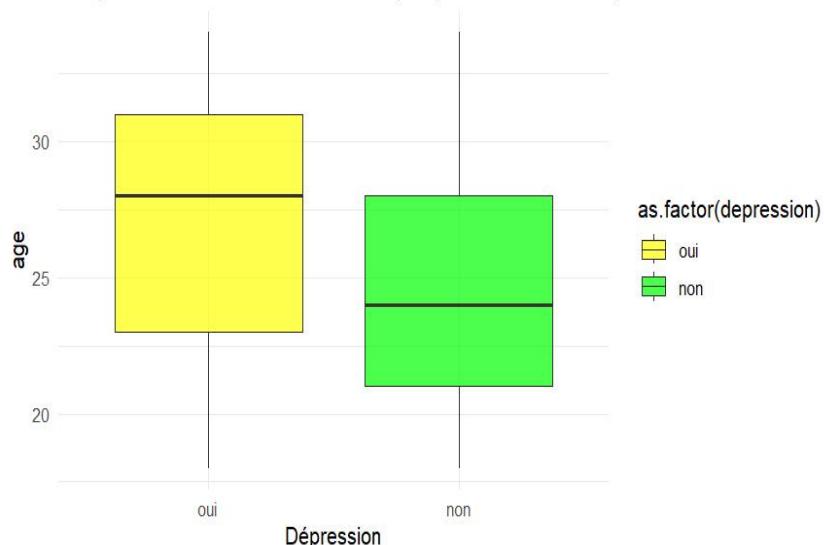
**Tableau ANOVA**

Facteur	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Facteur	1	11998	11998.4	8125.2	<2.2e-16 ***
Résidus	27896	41194	1.5		

STUDENT_DEPRESSION

✓ Relation entre dépression et Age

Boxplot de la Pression Académique par Niveau de Dépression



- Les âges les plus représentés chez les étudiants en dépression sont 24 à 29 ans.
- La proportion de dépression est légèrement plus élevée chez les étudiants plus âgés.

STUDENT_DEPRESSION

Tableau de Contingence et Analyse ANOVA

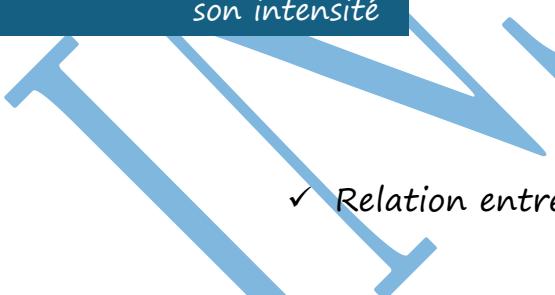
Indicateur	Valeur
Rapport de corrélation (η^2)	0.051
P-Value (ANOVA)	$< 2.2e-16$
Significativité	Liaison significative, les deux variables sont liées
Intensité de la liaison	Liaison très faible
Remarque	Si la liaison n'est pas significative, ne pas tenir compte de son intensité

Tableau ANOVA

Facteur	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Facteur	1	33929	33929	1507.9	$<2.2e-16$ ***
Résidus	27896	627673	23		

Interpretation :

- La répartition de l'âge montre une prévalence plus élevée de la dépression chez les personnes âgées.
- Toutefois, le rapport de corrélation indique que l'influence de l'âge sur la dépression est statistiquement significative.



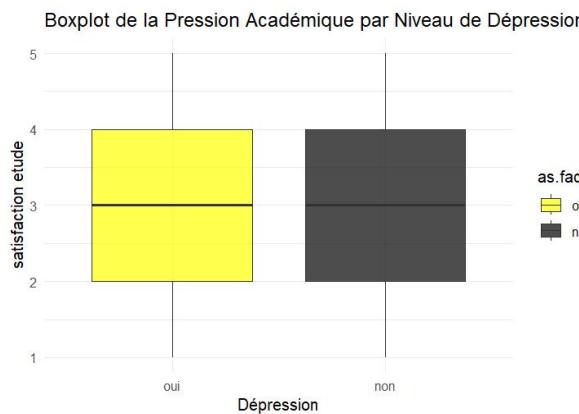
✓ Relation entre dépression et satisfaction études

Tableau de Contingence

		1	2	3	4	5	Total Ligne
Dépression	1597	2070	2467	3095	2334	11563	
Oui	1597	2070	2467	3095	2334	11563	
Non	3862	3768	3353	3264	2088	16335	

STUDENT_DEPRESSION

Total	5459	5838	5820	6359	4422	27898
Colonne						



- L'influence de la satisfaction aux études sur la dépression est toutefois très faible.

- Les étudiants les plus dépressifs ont une satisfaction aux études généralement plus faible.
- On observe une proportion plus importante de non-dépressifs dans les catégories 4 et 5 (satisfaction élevée).

STUDENT_DEPRESSION

Tableau ANOVA

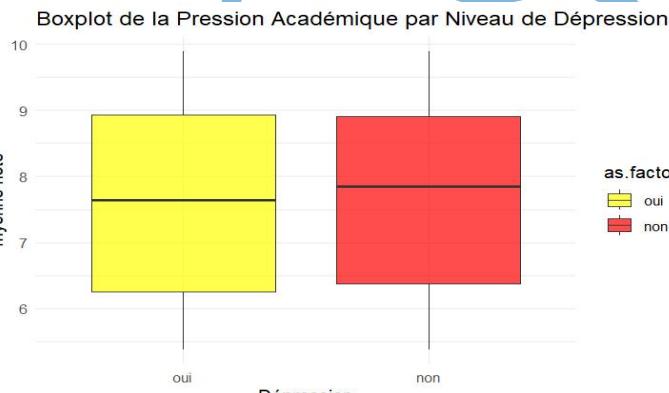
Indicateur	Valeur
Rapport de corrélation (η^2)	0.028
P-Value (ANOVA)	< 2.2e-16
Significativité	Liaison significative, les deux variables sont liées
Intensité de la liaison	Liaison très faible

Facteur	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Facteur	1	1459	1458.8	811.02	<2.2e-16 ***

interprétation :

- P-value très faible (< 2.2e-16) : il existe une relation entre la satisfaction des études et la dépression et cette relation est statistiquement significative.

✓ Relation depression et moyenne



Résultat	Valeur
Rapport_Correlation	0.0004831228
Anova.P.value	0.0002410929
Significativité du Test Anova	Liaison significative, les deux variables sont liées
Intensité de la liaison	Liaison très faible
Remarque	Si la liaison n'est pas significative, ne pas tenir compte de son intensité

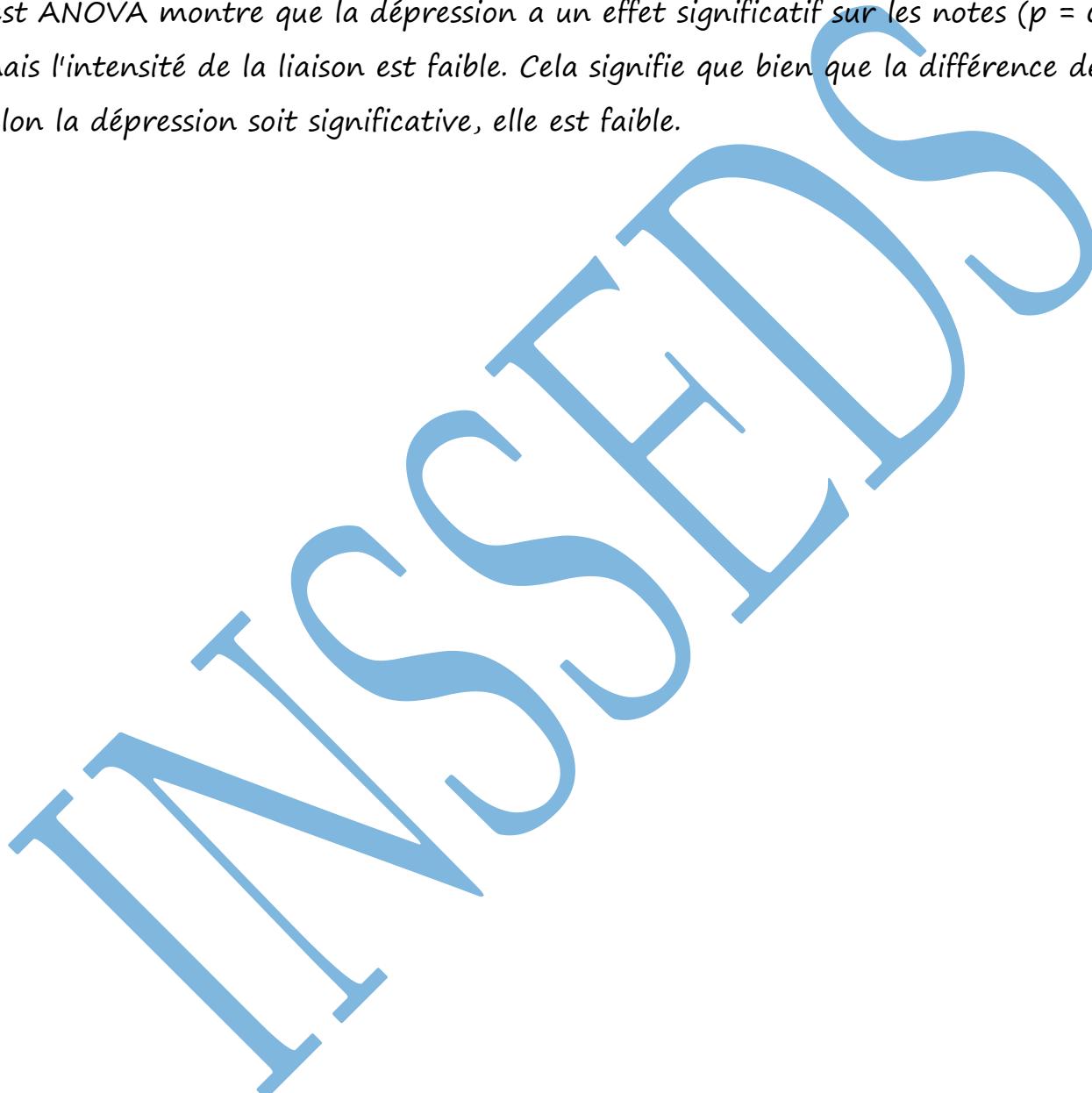
Tableau de l'ANOVA :

STUDENT_DEPRESSION

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Facteur	1	28	28.208	13.484	0.0002411 ***
Résidus	27896	58359	2.092	-	-

La corrélation est très faible (0.00048), indiquant peu de lien entre les variables. Le

test ANOVA montre que la dépression a un effet significatif sur les notes ($p = 0.00024$), mais l'intensité de la liaison est faible. Cela signifie que bien que la différence de notes selon la dépression soit significative, elle est faible.



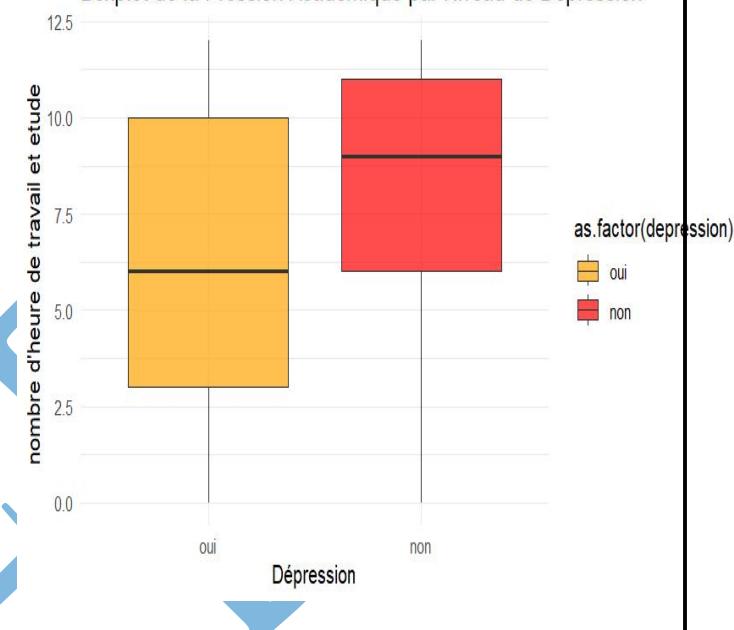
STUDENT_DEPRESSION

✓ Relation dépression et nombre d'heure d'étude et travail

Dépression	Oui (travail/étude)	Non (travail/étude)	Total
0	1096 (0.095)	604 (0.037)	1700
1	686 (0.059)	464 (0.028)	1150
2	894 (0.077)	696 (0.043)	1590
3	773 (0.067)	697 (0.043)	1470
4	799 (0.069)	815 (0.050)	1614
5	590 (0.051)	706 (0.043)	1296
6	959 (0.083)	1288 (0.079)	2247
7	825 (0.071)	1176 (0.072)	2001
8	941 (0.081)	1568 (0.096)	2509
9	811 (0.070)	1214 (0.074)	2025
10	1255 (0.109)	2979 (0.182)	4234
11	932 (0.081)	1960 (0.120)	2892
12	1002 (0.087)	2168 (0.133)	3170
Total	11 563	16 335	27 898

Tableau de la Répartition des

Boxplot de la Pression Académique par Niveau de Dépression



Fréquences

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Facteur	1	16 691	16 691.1	1269.4	< 2.2e-16 ***
Résidus	27 896	366 789	13.1	-	-
Total	27 897	383 480	-	-	-

Rapport de

(0.04352529) :

Corrélation

- Très faible corrélation, suggérant une relation quasi inexiste entre le nombre d'heures de travail/étude et la dépression.

2. Test d'ANOVA :

STUDENT_DEPRESSION

- $p\text{-value} = 6.21e-272$: Très significatif, donc la dépression a un effet statistiquement significatif sur le nombre d'heures de travail/étude.
- Intensité de la liaison : très faible, même si le lien est significatif, il n'est pas fort.

3. Tableau des Contingences:

- Il montre la répartition des réponses selon les niveaux de dépression, avec des résultats de fréquences très faibles dans certaines cases.

✓ Relation entre dépression et stress financier

Tableau de la Répartition des Fréquences pour le Stress Financier et la Dépression

Dépression	Oui (stress financier)	Non (stress financier)	Total
1	3489 (0.302)	1632 (0.100)	5121
2	2886 (0.250)	2175 (0.133)	5061
3	2146 (0.186)	3080 (0.189)	5226
4	1785 (0.154)	3990 (0.244)	5775
5	1257 (0.109)	5458 (0.334)	6715
Total	11 563	16 335	27 898

. Tableau du Rapport de Corrélation et du Test ANOVA

Résultat	Valeur
Rapport de Corrélation	0.1321987
p-value du Test ANOVA	0

STUDENT_DEPRESSION

Significativité du Test
Liaison significative, les deux variables sont liées

Intensité de la Liaison
Liaison faible

Remarque
Si la liaison n'est pas significative, ne pas tenir compte de son intensité

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Facteur	1	7619	7619.2	4249.6	< 2.2e-16 ***
Résidus	27 896	50 015	1.8	-	-
Total	27 897	57 634	-	-	-

p-value du Test ANOVA : Le test est très significatif, ce qui confirme que la relation entre les deux variables est réelle.

Intensité de la liaison : La relation est faible.

Répartition des données : Le tableau de contingence montre comment le stress financier est réparti parmi les différents niveaux de dépression, avec plus de cas de stress financier dans les niveaux 1 à 5.

6. Conclusion et Recommandations

6.1 Résumé des principaux résultats

L'analyse des données a mis en évidence plusieurs relations importantes entre la **dépression** et différents facteurs liés au mode de vie et aux études :

- Pensées suicidaires** : Une forte association a été trouvée entre la dépression et les pensées suicidaires (**V de Cramer = 0.546**), confirmant que les étudiants dépressifs sont significativement plus susceptibles d'avoir des pensées suicidaires.
- Antécédents familiaux de maladies mentales** : Une corrélation modérée a été observée, suggérant un rôle potentiel des prédispositions génétiques.
- Pression académique et stress financier** : Ces facteurs influencent significativement la dépression, bien qu'avec une intensité modérée.
- Satisfaction aux études et habitudes alimentaires** : Une satisfaction faible aux études et une alimentation malsaine semblent être associées à un risque accru de dépression.
- Durée du sommeil et moyenne des notes** : L'effet de ces variables sur la dépression est faible, bien que des tendances soient observées.

STUDENT_DEPRESSION

- **Âge et sexe** : Ces variables ont montré une **liaison très faible**, suggérant que la dépression touche divers profils d'étudiants sans distinction marquée.

6.2 Interprétation des tendances observées

- **Facteurs psychologiques et familiaux** : L'association entre la dépression et les pensées suicidaires met en avant un besoin urgent de prise en charge psychologique pour les étudiants vulnérables. Les antécédents familiaux semblent jouer un rôle, mais ne suffisent pas à expliquer à eux seuls la prévalence de la dépression.
- **Facteurs académiques et financiers** : Les étudiants soumis à une **forte pression académique** et au **stress financier** présentent un risque plus élevé de dépression. Cela suggère que les conditions d'apprentissage et le soutien économique influencent le bien-être mental.
- **Facteurs liés au mode de vie** : Bien que l'impact de la satisfaction aux études, de la durée du sommeil et de l'alimentation soit relativement faible, ils participent à l'état général de santé mentale des étudiants. Un mode de vie déséquilibré peut donc amplifier les risques de dépression.
- **Caractéristiques démographiques** : L'âge et le sexe ne sont pas des déterminants majeurs, ce qui indique que la dépression touche une large population sans distinction évidente.

6.3 Recommandations basées sur les résultats**1. Renforcer le soutien psychologique dans les établissements scolaires et universitaires**

- Développer des **cellules d'écoute et de soutien psychologique** accessibles à tous les étudiants.
- Organiser des **campagnes de sensibilisation** sur la santé mentale et la gestion du stress.
- Former le personnel éducatif à **déetecter les signes de détresse psychologique**.

2. Améliorer les conditions académiques et financières des étudiants

- Mettre en place des **programmes de bourses et d'aides financières** pour réduire le stress financier.
- Encourager une charge de travail équilibrée pour éviter la surcharge académique.
- Proposer des **ateliers de gestion du temps et du stress** pour aider les étudiants à mieux gérer leurs études.

3. Promouvoir un mode de vie sain

- Sensibiliser les étudiants à l'importance d'une **alimentation équilibrée et d'un sommeil régulier**.
- Organiser des **activités sportives et de relaxation** pour améliorer la gestion du stress.
- Encourager un **meilleur équilibre entre études et loisirs** pour réduire les risques de burnout.

4. Suivi des étudiants les plus vulnérables

- Identifier les **étudiants à risque** grâce à des questionnaires anonymes sur la santé mentale.
- Offrir des **consultations psychologiques gratuites ou à coût réduit** pour les étudiants en détresse.

STUDENT_DEPRESSION

- Travailler en collaboration avec des **professionnels de la santé mentale** pour proposer un accompagnement adapté.

En appliquant ces recommandations, les institutions éducatives peuvent mieux prévenir la dépression et améliorer le bien-être des étudiants. Ces résultats soulignent l'importance d'une approche **multidimensionnelle**, combinant **soutien psychologique, amélioration des conditions académiques et promotion d'un mode de vie sain**.

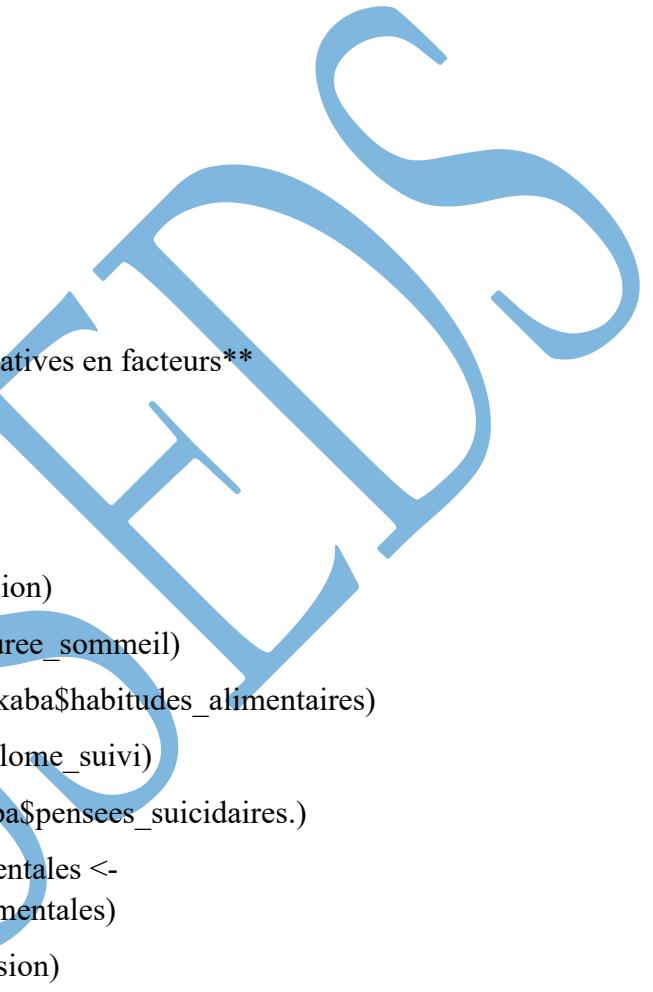


Annexe

Annexe : Codes R pour l'Analyse des Données

STUDENT_DEPRESSION

```

## **1. Importation des données**
```r
setwd("C:/INSEEDS/PROJET/STAT DESC")
kaba <- read.table("Student_Depression.csv", header=TRUE, sep=",", check.names=FALSE,
row.names=1, stringsAsFactors=TRUE)
summary(kaba)
str(kaba)
```


```

```

## **2. Prétraitement des données**
### **Transformation des variables qualitatives en facteurs**
```r
kaba$sexe <- as.factor(kaba$sexe)
kaba$ville <- as.factor(kaba$ville)
kaba$profession <- as.factor(kaba$profession)
kaba$duree_sommeil <- as.factor(kaba$duree_sommeil)
kaba$habitudes_alimentaires <- as.factor(kaba$habitudes_alimentaires)
kaba$diplome_suivi <- as.factor(kaba$diplome_suivi)
kaba$pensees_suicidaires. <- as.factor(kaba$pensees_suicidaires.)
kaba$antecedents_familiaux_maladies_mentales <-
as.factor(kaba$antecedents_familiaux_maladies_mentales)
kaba$depression <- as.factor(kaba$depression)
```


```

```

### **Standardisation des variables quantitatives**
```r
kaba$pression_academique <- scale(kaba$pression_academique)
kaba$moyenne_notes <- scale(kaba$moyenne_notes)
kaba$satisfaction_etudes <- scale(kaba$satisfaction_etudes)
kaba$nombre_heure_travail_etude <- scale(kaba$nombre_heure_travail_etude)
kaba$stress_financier <- scale(kaba$stress_financier)
```

```

STUDENT_DEPRESSION

```

### **Suppression des variables non pertinentes**
```r
kaba <- kaba[, !names(kaba) %in% c("pression_liee_au_travail", "satisfaction_travail")]
```

```
3. Vérification des valeurs manquantes
```r
library(naniar)
na_percent <- colSums(is.na(kaba)) / nrow(kaba) * 100
vis_miss(kaba)
```
```
## **4. Analyse univariée**
### **Tableaux statistiques**
```r
age = kaba.qt.tableau(kaba$age)
pression_academique = kaba.qt.tableau(kaba$pression_academique)
moyenne_notes = kaba.qt.tableau(kaba$moyenne_notes)
satisfaction_etudes = kaba.qt.tableau(kaba$satisfaction_etudes)
nombre_heure_travail_etude = kaba.qt.tableau(kaba$nombre_heure_travail_etude)
stress_financier = kaba.qt.tableau(kaba$stress_financier)
```
```
Exportation des résultats
```r
write.table(pression_academique, "C:/INSEEDS/COURS/stat descriptive/pression_academique.csv",
row.names=TRUE, sep=";", dec=",")
write.table(moyenne_notes, "C:/INSEEDS/COURS/stat descriptive/moyenne_notes.csv",
row.names=TRUE, sep=";", dec=",")
```

```

**STUDENT\_DEPRESSION**

```

write.table(satisfaction_etudes, "C:/INSEEDS/COURS/stat descriptive/satisfaction_etudes.csv",
row.names=TRUE, sep=";" ,dec=",")

write.table(nombre_heure_travail_etaude, "C:/INSEEDS/COURS/stat
descriptive/nombre_heure_travail_etaude.csv", row.names=TRUE, sep=";" ,dec=",")

write.table(stress_financier, "C:/INSEEDS/COURS/stat descriptive/stress_financier.csv",
row.names=TRUE, sep=";" ,dec=",")

...

```

**### \*\*Visualisation des distributions\*\***

```

```r
kaba.qt.graph(kaba$age)
kaba.qt.resume(kaba$age)
kaba.qt.graph(kaba$pession_academique)
kaba.qt.resume(kaba$pession_academique)
kaba.qt.graph(kaba$moyenne_notes)
kaba.qt.resume(kaba$moyenne_notes)
kaba.qt.graph(kaba$satisfaction_etudes)
kaba.qt.resume(kaba$satisfaction_etudes)
kaba.qt.graph(kaba$nombre_heure_travail_etaude)
kaba.qt.resume(kaba$nombre_heure_travail_etaude)
kaba.qt.graph(kaba$stress_financier)
kaba.qt.resume(kaba$stress_financier)
```

```

**## \*\*5. Analyse de la relation entre variables qualitatives et quantitatives\*\***

```

```r
mn.qtql.liaison <- function(vecteur, facteur) {
  library(BioStatR)
  res1 <- eta2(vecteur, facteur)
  res2 <- anova(lm(vecteur ~ facteur))
  a <- data.frame(res2)
  res3 <- a$Pr..F.[1]
}

```

STUDENT_DEPRESSION

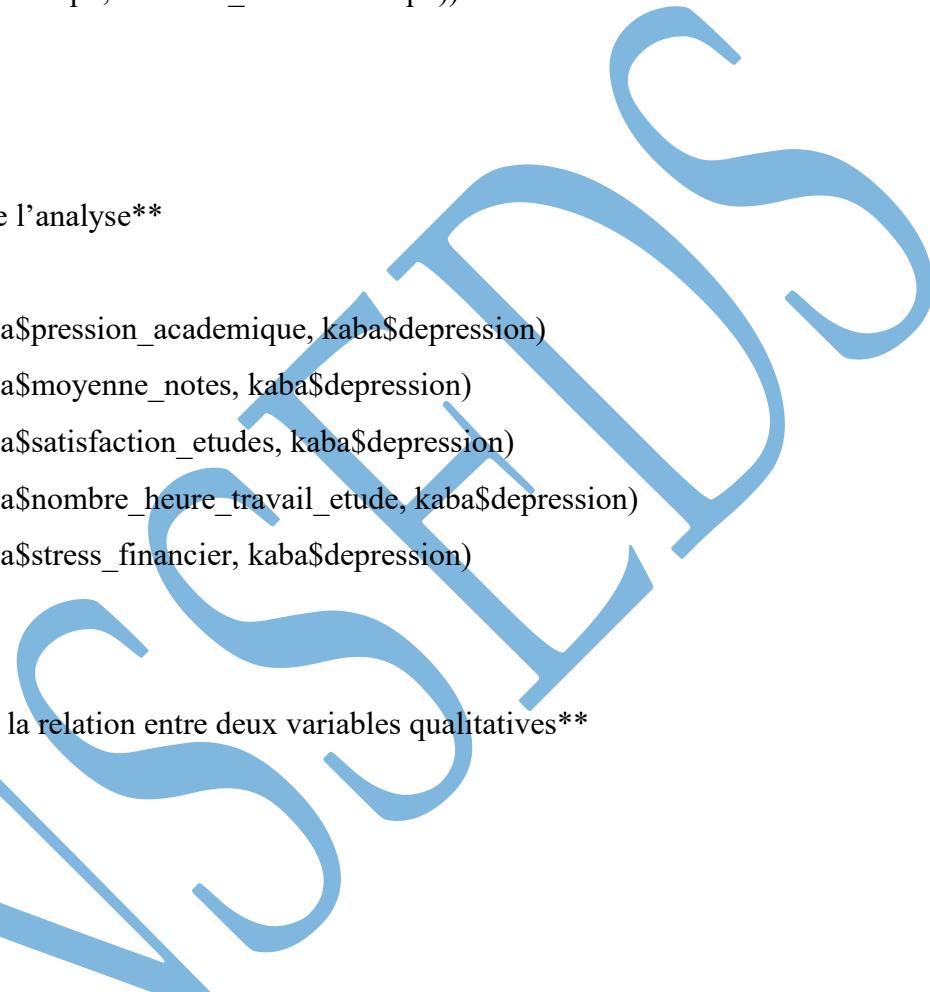
```
interp1 <- ifelse(res1<0.10,'liaison très faible',ifelse(res1<0.40,'liaison faible',ifelse(res1<0.60,'liaison moyenne',ifelse(res1<0.80,'liaison forte','liaison très forte'))))
```

```
interp2 <- ifelse(res3<0.05, 'liaison significative, les deux variables sont liées','liaison non significative, les deux variables ne sont pas liées')
```

```
return(list(Rapport_Correlation=res1, Resultat_Test_Anova=res2, Anova.P.value=res3,
Significativite_TestAnova=interp2, Intensite_liaison=interp1))
```

{

...



```
### **Exécution de l'analyse**
```

```r

```
mn.qtql.liaison(kaba$pression_academique, kaba$depression)
mn.qtql.liaison(kaba$moyenne_notes, kaba$depression)
mn.qtql.liaison(kaba$satisfaction_etudes, kaba$depression)
mn.qtql.liaison(kaba$nombre_heure_travail_etude, kaba$depression)
mn.qtql.liaison(kaba$stress_financier, kaba$depression)
```

...

```
6. Analyse de la relation entre deux variables qualitatives
```

```r

```
library(ggplot2)
library(vcd)
```

```
variables_qualitatives <- c("pensees_suicidaire", "antecedants_familiaux_maladie_mentale", "sexe",
"duree_sommeil", "habitudes_alimentaires", "diplome_suivi", "profession")
```

```
for (var in variables_qualitatives) {
```

```
print(paste("Analyse de la relation entre Dépression et", var))
```

```
print(table(kaba$depression, kaba[[var]]))
```

```
print(paste("V de Cramer :", assocstats(table(kaba$depression, kaba[[var]]))$cramer))
```

```
ggplot(kaba, aes(x = .data[[var]], fill = depression)) +
```

```
geom_bar(position = "fill") +
```

STUDENT_DEPRESSION

```
labs(title = paste("Répartition de la dépression selon", var), x = var, y = "Proportion") +  
theme(axis.text.x = element_text(angle = 45, hjust = 1))  
}  
...  
---
```

⌚ **Ce document regroupe l'ensemble des codes utilisés pour l'analyse de la relation entre la dépression et divers facteurs liés au mode de vie et aux études des étudiants.**}

