



INF8808 - Visualisation de données

Hiver 2025

Plan de projet

Équipe 12

2244082 – Omar Benzekri

2213673 – Ireina Hedad

2068795 – Samuel Kouakou

2077446 – Félix Lamarche

2075826- Mohamed Lamine Gning

Soumis à : Thomas Hurtut

Mercredi 26 Février 2025

Table des matières	2
1. Mise en contexte	3
1.1 Contexte	3
1.2 But global	3
1.3 Public cible	4
2. Les données	5
2.1 Source et description générale	5
2.2 Contenu et structure des données	6
3. Les questions cibles	9

1. Mise en contexte

1.1 Contexte

Dans notre société, la réussite professionnelle est souvent associée à de forts indicateurs académiques. Pourtant, l'expérience pratique, les compétences interpersonnelles et le networking semblent jouer un rôle déterminant dans l'évolution de carrière. À travers ce projet, nous souhaitons explorer en profondeur cette problématique : dans quelle mesure la performance académique détermine-t-elle réellement le succès professionnel et quelles autres dimensions (stages, projets, certifications, soft skills, networking) pourraient compenser ou renforcer cet impact ? L'enjeu est d'identifier les leviers essentiels permettant de mieux comprendre les facteurs influençant l'employabilité et l'évolution de carrière.

1.2 Objectif

Notre objectif principal est de créer un ensemble de visualisations interactives qui permettront aux utilisateurs de découvrir et d'analyser les relations entre les performances académiques et les résultats professionnels. Nous souhaitons ainsi :

- Mettre en évidence la corrélation entre les indicateurs académiques (GPA, score SAT, classement universitaire) et les résultats tels que le salaire de départ, le nombre d'offres d'emploi ou le temps avant la promotion.
- Évaluer l'impact des expériences pratiques (stages, projets, certifications) et des compétences non académiques (soft skills, networking) sur la progression professionnelle.
- Permettre à différents publics – étudiants, conseillers d'orientation, recruteurs – d'explorer ces dynamiques afin de leur fournir une vision plus informée appropriée à leurs besoins spécifiques.

1.3 Public cible

Les visualisations seront conçues pour un grand public varié ne nécessitant pas de connaissances préalables spécifiques :

- **Étudiants et jeunes diplômés**, souhaitant s'informer sur les facteurs pertinents par rapport à la réussite de leur parcours professionnel futur.
- **Conseillers d'orientation et responsables pédagogiques**, souhaitant adapter leurs recommandations en fonction des facteurs identifiés.
- **Recruteurs et professionnels des ressources humaines**, pour mieux comprendre le profil des candidats et valoriser des compétences souvent négligées par les indicateurs académiques classiques.

2. Jeux de données

2.1 Source et description générale

Le dataset utilisé pour ce projet provient de Kaggle, intitulé « Education and Career Success ». Il comporte 5000 points de données, chacun correspondant à un étudiant, et offre une vue d'ensemble sur divers aspects de leur parcours éducatif, leurs compétences et leurs résultats professionnels.

Ce jeu de données a été généré synthétiquement, en se basant sur des tendances réelles observées dans le domaine de l'éducation et du succès professionnel. La source indique que les valeurs générées s'inspirent de données publiques, d'études de recherche et de rapports statistiques portant sur la performance académique et les trajectoires de carrière. Il a été conçu pour des utilisations dans l'éducation, l'apprentissage machine, l'analyse statistique et pour simuler des schémas. L'absence de données réelles permet d'éviter des problèmes de confidentialité. Les différentes distributions des données sont fournies et les paramètres sur lesquelles elles sont établies, cependant certains ont des explications superficielles.

Processus de Génération et Transformations

Les données ont été générées afin de représenter une variété uniforme d'étudiants provenant de différents parcours.

- **Démographie Étudiante :**
L'âge et le genre ont été attribués aléatoirement en suivant des distributions typiques dans l'enseignement supérieur.
- **Performance Académique :**
La moyenne au lycée et à l'université a été générée via une distribution normale entre 2.0 et 4.0. Les scores SAT ont été assignés aléatoirement dans une plage de 900 à 1600, respectant la répartition usuelle de ces résultats. Le classement des universités a été fixé aléatoirement entre 1 et 1000.
- **Compétences et Activités Extracurriculaires :**
Le nombre de stages, de projets et de certifications a été attribué de manière aléatoire, en tenant compte d'une corrélation avec le classement universitaire et le domaine d'études. De plus, les scores de compétences interpersonnelles (soft skills) et de networking ont été établis sur une échelle de 1 à 10 de façon aléatoire.
- **Résultats Professionnels :**
Les offres d'emploi et les salaires de départ ont été définis à partir d'une combinaison d'éléments tels que le classement universitaire, les moyennes (GPA), le nombre de stages réalisés et le score de networking. La satisfaction professionnelle, l'équilibre entre vie professionnelle et personnelle, ainsi que le délai avant promotion ont été simulés en s'appuyant sur des tendances de carrière estimées. Enfin, le critère d'entrepreneuriat a été fixé à « Oui » dans 20 % des cas et « Non » dans 80 % des cas.

2.2 Contenu et structure des données

Le jeu de données se divise en quatre grandes catégories : Informations personnelles, performance académique, expériences et compétences, résultats professionnels.

Tableau 1: Informations sur les données personnelles des étudiants


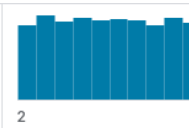


Variable	Description	Type de donnée	Intervalle / Valeurs	Distribution	Conclusion
<i>Student_ID</i>	Identifiant unique par étudiant	string	[S00001, S05000]	N/A	
<i>Age</i>	Âge de l'étudiant	int	[18, 29]		
<i>Gender</i>	Genre	string	{Male, Female, Other} <u>3 Valeurs</u>	Male: 49% Female: 47% Other: 4%	

Tableau 2: Informations sur les performances académiques des étudiants

Variable	Description	Type de donnée	Intervalle / Valeurs	Distribution	Conclusion
<i>High_School_GPA</i>	Moyenne au lycée	float	[2.0, 4.0]		
<i>SAT_Score</i>	Score aux examens standardisés	int	[900, 1600]		
<i>University_Ranking</i>	Classement de l'université fréquentée	int	[1, 1000]		

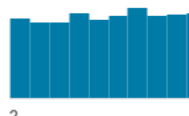
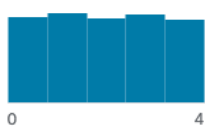



<i>University_GPA</i>	Moyenne universitaire	float	[2.0, 4.0]	 2	
<i>Field_of_Study</i>	Domaine d'études	string	{ Computer Science, Medicine, Business, Engineering, Arts, Law, Mathematics } <u>7 Valeurs</u>	Arts: 15% Mathematics: 15% Other: 70%	

Tableau 3: Informations sur les expériences et compétences des étudiants

Variable	Description	Type de donnée	Intervalle / Valeurs	Distribution	Conclusion
<i>Internships_Completed</i>	Nombre de stages complétés	int	[0, 4]	 0 4	
<i>Projects_Completed</i>	Nombre de projets académiques et personnels complétés	int	[0, 9]	 0 9	
<i>Certifications</i>	Nombre de certifications obtenus	int	[0, 5]	 0 5	
<i>Soft_Skills_Score</i>	Évaluation des compétences interpersonnelles	int	[0, 10]	 1 10	



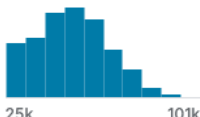
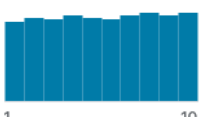
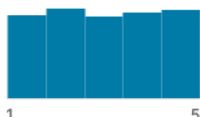
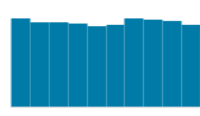
<i>Networking_Score</i>	Évaluation sur le réseautage et les connexions de l'individu	int	[0, 10]		
-------------------------	--	-----	---------	---	--

Tableau 4: Informations sur les carrières professionnels des étudiants gradués

Variable	Description	Type de donnée	Intervalle / Valeurs	Distribution	Conclusion
<i>Job_Offers</i>	Nombre d'offres d'emploi reçues après la graduation	int	[0, 5]		
<i>Starting_Salary</i>	Salaire de départ en USD	int	[25000, 101000]		
<i>Career_Satisfaction</i>	Niveau de satisfaction professionnelle	int	[1, 10]		
<i>Years_to_Promotion</i>	Nombre d'années avant la première promotion	int	[1, 5]		
<i>Current_Job_Level</i>	Niveau du poste actuel	String	{Entry, Mid, Senior, Executive} <u>4 Valeurs</u>	Entry: 49% Mid: 31% Other: 20%	
<i>Work_Life_Balance</i>	Évaluation de l'équilibre entre la vie professionnelle et personnelle	int	[1, 10]		
<i>Entrepreneurship</i>	Indique si l'étudiant a fait de l'entrepreneuriat	string	{No, Yes} <u>2 Valeurs</u>	No: 80% Yes: 20%	

3. Questions cibles

Axe 1 : Corrélation entre performance académique et résultats professionnels (Priorité globale : ★★★)

- **A1.1 (★★★)** : Dans quelle mesure un GPA élevé à l'université corrèle avec un succès dans leur carrière ?
 - **A1.2 (★★★)** : Quel est l'impact du classement universitaire par rapport au salaire initial ?
 - **A1.3 (★★★)** : Existe-t-il une corrélation entre les indicateurs académiques (GPA, SAT, classement) et le niveau de satisfaction professionnelle ?
 - **A1.4 (★★☆)** : Comment les indicateurs académiques influencent-ils le temps moyen nécessaire pour obtenir une promotion ?
 - **A1.5 (★★)** : Dans quelle mesure la performance académique prédit-elle la stabilité de la carrière sur le long terme ?
-

Axe 2 : Influence des expériences et compétences complémentaires (Priorité globale : ★★★)

- **A2.1 (★★★)** : Est-ce que les stages, projets réalisés et certifications obtenus peuvent compenser pour un profil académique moins performant dans l'obtention d'opportunités professionnelles ?
 - **A2.2 (★★★)** : Comment les compétences interpersonnelles et le réseautage influencent-ils la satisfaction professionnelle et l'évolution de carrière ?
 - **A2.3 (★★☆)** : Quel est l'impact combiné du nombre de stages et de certifications sur le salaire de départ et la progression professionnelle ?
 - **A2.4 (★★☆)** : Y a-t-il une interaction significative entre la réalisation de projets et le développement des compétences non académiques dans l'obtention d'offres d'emploi ?
 - **A2.5 (★★)** : En quoi les expériences extracurriculaires (participation à des clubs, associations, etc.) impacte l'obtention d'emploi initial ?
-

Axe 3 : Analyse par segmentation et variables démographiques (Priorité globale : ★★☆)

- **A3.1 (★★★)** : Existe-t-il des différences dans les résultats professionnels selon le domaine d'études choisi ?
- **A3.2 (★★☆)** : Les variables démographiques, telles que l'âge et le genre, influencent-elles la performance académique ?
- **A3.3 (★★☆)** : Les variables démographiques, telles que l'âge et le genre, influencent-elles la performance académique ?
- **A3.4 (★★)** : Les variables démographiques, telles que l'âge et le genre, influencent-elles le salaire initial et la trajectoire professionnelle ?

Axe 4 : Analyse temporelle et évolutions (Priorité globale : ★★★)

- **A4.1 (★★★)** : Est-ce que l'âge de l'individu influe sur son début de carrière professionnelle ?
- **A4.2 (★★★)** : Est-ce que le nombre d'années avant la première promotion est relié à la satisfaction de l'équilibre vie privée et professionnelle ?

Chaque axe d'analyse sera approfondi à l'aide de visualisations interactives permettant aux utilisateurs de filtrer et d'explorer ces questions à différents niveaux de détail. La priorisation proposée permettra de concentrer les efforts sur les aspects ayant le plus grand impact sur la compréhension des liens entre performance académique et réussite professionnelle, tout en explorant les dimensions complémentaires et temporelles qui enrichiront l'analyse globale.