



INF8808 - Visualisation de données

Hiver 2025

Mockup

Équipe 12

2244082 – Omar Benzekri

2213673 – Ireina Hedad

2068795 – Samuel Kouakou

2077446 – Félix Lamarche

2075826- Mohamed Lamine Gning

Soumis à : Thomas Hurtut

Mardi 18 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
4. Mockup	12
4.1 Vue générale	12
4.2 Visualisation 1	13
Description	14
Interactivité	14
4.3 Visualisation 2	15
Description	16
Interactivité	16
4.4 Visualisation 3	17
Description	18
Interactivité	18
4.5 Visualisation 4	19
Description	19
Interactivité	20
4.6 Visualisation 5	21
Description	21
Interactivité	22

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 s'adressent à un large public sans nécessiter de connaissances techniques préalables. Bien que nos trois groupes cibles aient des besoins spécifiques, ils partagent un objectif commun : mieux comprendre les facteurs influençant la réussite au-delà des notes académiques.

- **Étudiants et jeunes diplômés** : Ils souhaitent identifier les compétences et expériences qui favorisent leur employabilité et leur réussite professionnelle.
- **Conseillers d'orientation et responsables pédagogiques** : Ils cherchent à adapter leurs recommandations pour guider les étudiants vers des parcours mieux alignés avec les tendances du marché du travail.
- **Recruteurs et professionnels des RH** : Ils veulent mieux évaluer les candidats en prenant en compte des critères pertinents souvent ignorés par les indicateurs académiques classiques.

Ainsi, nos visualisations mettront en avant des indicateurs transversaux utiles aux trois groupes (ex. : compétences clés, expériences significatives, liens entre parcours et insertion professionnelle). Cependant, certaines analyses plus spécifiques seront différenciées pour répondre aux besoins particuliers de chaque audience.

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

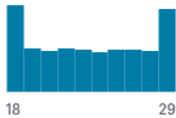
Variable	Description	Type de donnée	Intervalle / Valeurs	Distribution	Conclusion
<i>Student_ID</i>	Identifiant unique par étudiant	string	[S00001, S05000]	N/A	Il y a une entrée unique par étudiant
<i>Age</i>	Âge de l'étudiant	int	[18, 29]		La distribution de l'âge est uniforme (la première et dernière colonne contienne 17 et 18 et pour la dernière 28 et 29)
<i>Gender</i>	Genre	string	{Male, Female, Other} <u>3 Valeurs</u>	Male: 49% Female: 47% Other: 4%	La distribution est plutôt égale par rapport aux hommes et aux femmes

Tableau 2: Informations sur les performances académiques des étudiants

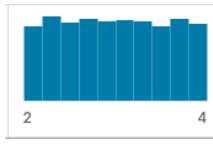
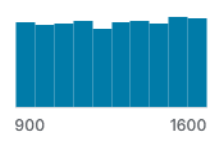
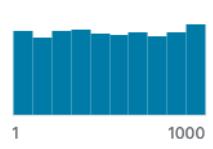
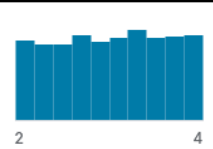

Variable	Description	Type de donnée	Intervalle / Valeurs	Distribution	Conclusion
<i>High_School_GPA</i>	Moyenne au lycée	float	[2.0, 4.0]		La tendance suit une distribution uniforme
<i>SAT_Score</i>	Score aux examens standardisés	int	[900, 1600]		La tendance suit une distribution uniforme
<i>University_Ranking</i>	Classement de l'université fréquentée	int	[1, 1000]		La tendance suit une distribution uniforme
<i>University_GPA</i>	Moyenne universitaire	float	[2.0, 4.0]		La tendance suit une distribution uniforme
<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%	C'est une distribution presque uniforme pour le domaine d'études

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]		La tendance suit une distribution uniforme




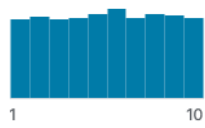
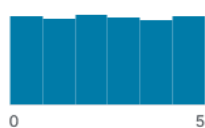
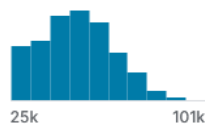
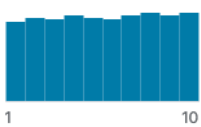


<i>Projects_Comp</i> <i>leted</i>	Nombre de projets académiques et personnels complétés	int	[0, 9]		La tendance suit une distribution uniforme
<i>Certifications</i>	Nombre de certifications obtenus	int	[0, 5]		La tendance suit une distribution uniforme
<i>Soft_Skills_Score</i>	Évaluation des compétences interpersonnelles	int	[0, 10]		La tendance suit une distribution uniforme
<i>Networking_Score</i>	Évaluation sur le réseautage et les connexions de l'individu	int	[0, 10]		La tendance suit une distribution uniforme

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]		La tendance suit une distribution uniforme
<i>Starting_Salary</i>	Salaire de départ en USD	int	[25000, 101000]		La tendance semble suivre une distribution binomiale

<i>Career_Satisfaction</i>	Niveau de satisfaction professionnelle	int	[1, 10]		La tendance suit une distribution uniforme
<i>Years_to_Promotion</i>	Nombre d'années avant la première promotion	int	[1, 5]		La tendance suit une distribution uniforme
<i>Current_Job_Level</i>	Niveau du poste actuel	String	{Entry, Mid, Senior, Executive} <u>4 Valeurs</u>	Entry: 49% Mid: 31% Other: 20%	La majorité ont une position d'Entry level et chaque niveau successif est moins probable que le précédent et peu sont des seniors ou des exécutifs
<i>Work-Life_Balance</i>	Évaluation de l'équilibre entre la vie professionnelle et personnelle	int	[1, 10]		La tendance suit une distribution uniforme
<i>Entrepreneurship</i>	Indique si l'étudiant a fait de l'entrepreneuriat	string	{No, Yes} <u>2 Valeurs</u>	No: 80% Yes: 20%	Les données ont été choisies pour avoir cette distribution de 80% non et 20% oui

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 le salaire initial et la trajectoire 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.

4. Mockup

4.1. Vue générale

Nos visualisations seront présentées sous une forme de *scrollytelling*, où nos visualisations se succéderont pour permettre à notre audience d'aboutir à une conclusion adéquate à propos de nos données. Les différentes visualisations et les questions auxquelles elles répondent seront divisées en sections, ainsi il sera plus facile pour nos lecteurs d'identifier quelles visualisations répondent le plus à leur questionnements. Cette structure permet de comprendre simplement les points que l'on communique. Voici la mise en page prévue de notre site, telle que le contenu soit montré dans une seule page (les visualisations montrées n'étant que des exemples, et ne sont pas nos visualisations prévues qui suivent dans ce rapport). Cette première version de notre site est déjà déployée et disponible [ici](#).



Figure 1: Interface de *BeyondGPA*

4.2. Visualisation 1

La première visualisation permet de répondre aux questions suivantes:

A3.2 (★★☆) : Les variables démographiques, telles que l'âge et le genre, influencent t-elles la performance académique?

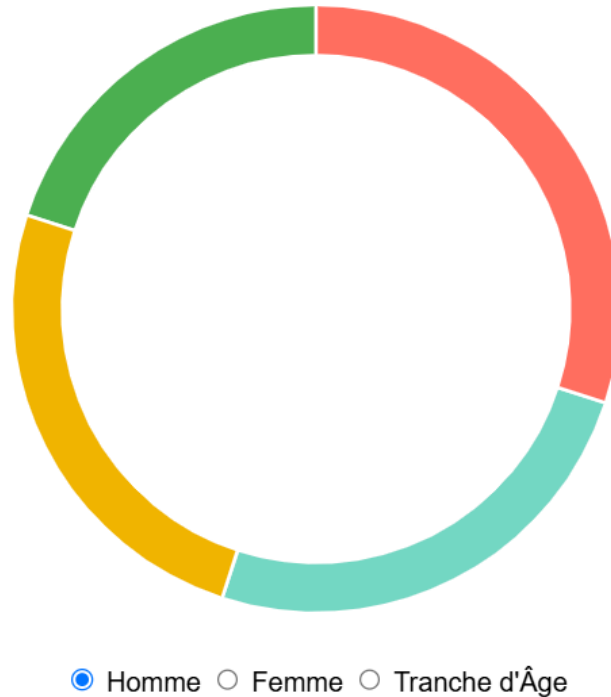


Figure 2 : Diagramme en anneau montrant la distribution des GPAs selon le genre

Description

- **Deux versions du diagramme en anneau** : une pour les hommes, une pour les femmes, et on a l'option tranche d'âge
- L'utilisateur peut basculer entre hommes, femmes et tranches pour comparer la distribution des GPAs.

La visualisation est un diagramme en anneau interactif qui permet d'analyser la répartition des niveaux de GPA en fonction du genre ou l'âge. Il y a quatre sections dans lesquelles chaque étudiant a été mis. Chaque section est répartie également sur l'intervalle de 2.0 à 4.0 pour un GPA, donc il y a une tranche pour les intervalles de 2.0 à 2.5, de 2.5 à 3.0, de 3.0 à 3.5 et de 3.5 à 4.0. Nous pouvons observer que les données sont distribuées d'une façon relativement égale entre les quatre différentes tranches. Ce diagramme met en évidence les différences entre les hommes et les femmes en termes de performance académique. Il permet de fournir le contexte des données au lecteur afin qu'il puisse avoir une meilleure compréhension des diagrammes présentés par la suite. Ainsi, il pourra comprendre plus facilement aisément les conclusions que nous désirons.

Interactivité

L'utilisateur peut basculer entre les hommes et les femmes pour comparer les distributions et observer les différences éventuelles dans les tendances académiques. Et il peut filtrer par âge pour voir si la distribution change selon l'âge de l'étudiant. De plus, si l'utilisateur survole une des tranches du diagramme, un onglet apparaît exposant les données en détails en fournissant le pourcentage d'étudiants se retrouvant dans cette section spécifique et quel est l'intervalle donné de cette tranche du diagramme.

4.3. Visualisation 2

La deuxième visualisation permet de répondre aux questions suivantes:

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 ?

A1.5 (★★) : Dans quelle mesure la performance académique prédit-elle la stabilité de la carrière sur le long terme ?

Diagramme de Sankey: Influence des Performances Académiques sur la Carrière



Figure 3 : Diagramme de Sankey

Description

La visualisation est un diagramme de Sankey interactif qui met en évidence l'influence des performances académiques sur la carrière des étudiants. Il illustre les relations entre les variables suivantes :

- **Performances académiques** : GPA universitaire, classement de l'université, score aux examens standardisés
- **Évolution professionnelle** : Salaire initial, Niveau du poste actuel, nombre d'années avant la première promotion, Niveau de satisfaction professionnelle

Cette visualisation est segmentée par domaine d'études, permettant d'observer des tendances spécifiques selon la discipline académique. On pourra donc observer la répartition des étudiants et voir les tendances globales. La visualisation montrera implicitement les proportions présentes selon les différentes variables.

Plusieurs variables sont continues et pour les représenter, il est plus intéressant de regrouper leurs valeurs en sections. Les valeurs du GPA sont donc divisées en quatre sections: 2.0 à 2.5, 2.5 à 3.0, 3.0 à 3.5 et 3.5 à 4.0. Les classements universitaires sont regroupés en quatre intervalles: 1 à 250, 251 à 500, 501 à 750 et 751 à 1000. Les scores SAT sont répartis en quatre sections également: 900 à 1100, 1101 à 1300, 1301 à 1500 et 1501 à 1600. Le salaire de départ sera aussi divisé en quatre tranches: 25 000 à 40 000, 40 000 à 60 000, 60 000 à 80 000 et 80 000 à 100 000. Finalement, la satisfaction professionnelle est segmentée en cinq intervalles: 1 à 2, 2 à 4, 4 à 6, 7 à 8 et 9 à 10.

Interactivité

L'utilisateur pourra sélectionner les indicateurs académiques à analyser parmi le GPA, le classement universitaire et le SAT pour le premier axe. Pour le deuxième, il pourra choisir d'analyser l'évolution professionnelle en ciblant le salaire initial, le niveau du poste actuel, le nombre d'années avant la première promotion ou le niveau de satisfaction professionnelle. L'utilisateur pourra choisir un domaine d'étude pour une analyse ciblée (ex. : Informatique, Médecine, Ingénierie, etc.). Également, l'utilisateur peut survoler les données dans le diagramme pour voir quelles sont les tranches desquelles l'ensemble de données survolé se retrouve.

4.4. Visualisation 3

La troisième visualisation permet de répondre aux questions suivantes:

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 ?

A2.3 (★★☆) : 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 ?

Analyse interactive des facteurs influençant le succès professionnel



Figure 4 : Nuage de points interactif analysant les facteurs influençant le succès professionnel

Description

Ce diagramme met en évidence l'influence des expériences pratiques et des compétences complémentaires sur l'accès au marché du travail et le salaire de départ. Il permet d'analyser la contribution de plusieurs variables relatant à la capacité d'un étudiant à propos de ses résultats professionnels:

- Expériences pratiques : Nombre de stages réalisés, projets complétés, certifications obtenues.
- Compétences transversales : Soft skills, score de réseautage.
- Expériences extracurriculaires : Participation à des clubs et associations.

La visualisation met en évidence les facteurs clés du succès professionnel au-delà des performances académiques traditionnelles. Elle permet de voir si un élève performant moins bien à l'école peut compenser grâce à d'autres facteurs que ceux académiques telles que les compétences de réseautage et d'observer si elles ont réellement une influence sur la carrière d'un individu.

Le diagramme est un nuage de points, où l'axe X et l'axe Y peuvent être choisies parmi différents facteurs, n'étant pas des caractéristiques académiques, pouvant affecter la recherche d'un emploi. Les données pouvant être observées grâce à la légende, étant un gradient de couleur, sont le nombre d'offres d'emplois reçus à la graduation allant de 0 à 5 et le salaire annuel initial qui est dans l'intervalle de 25000\$ à 101 000\$.

Interactivité

L'utilisateur pourra modifier les axes X et Y pour analyser l'interaction entre deux variables parmi : le score de réseautage, le score de soft skills, le nombre de projets réalisés, le nombre de stages accomplis, le nombre d'expériences extracurriculaires et le GPA. Il pourra également modifier les données représenter par la légende pour soit le salaire initial, soit le nombre d'offres d'emploi reçues. Finalement, on pourra filtrer les résultats en fonction du domaine d'étude pour observer des tendances spécifiques.

4.5 Visualisation 4

La quatrième visualisation permet de répondre aux questions suivantes:

A3.1 (★★★) : Existe-t-il des différences dans les résultats professionnels selon le domaine d'études choisi ?

A3.3 (★★) : Les variables démographiques, telles que l'âge et le genre, influencent-elles le salaire initial et la trajectoire professionnelle ?

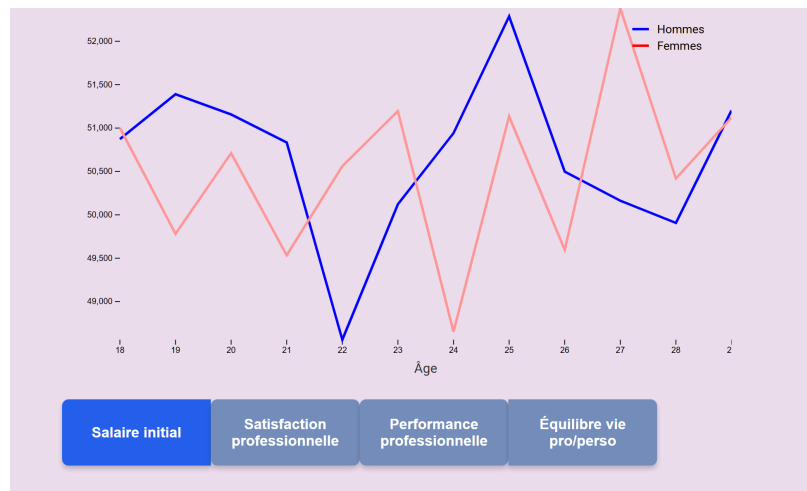


Figure 5 : Diagramme d'un graphique linéaire montrant une de quatre variables professionnels selon l'âge et le genre

Description

Le graphique nous permettra d'examiner les différences dans les résultats professionnels en fonction des variables démographiques telles que l'âge et le genre sur la performance professionnelle des étudiants. La courbe bleue représente les hommes et la courbe rose représente les femmes.

Comme illustré dans l'image ci-dessus, l'axe horizontal représente l'âge des étudiants diplômés et l'axe vertical représente l'une des quatre catégories: le salaire initial annuel en dollars, la satisfaction professionnelle noté selon un score de 1 à 10, la performance professionnelle noté selon un score de 1 à 10 et l'équilibre entre la vie professionnelle et la vie personnelle selon un score de 1 à 10. Ainsi, le graphe permet de comparer facilement les différences des résultats professionnels selon l'âge et le genre sous différentes métriques désirées.

Interactivité

L'utilisateur aura quatre boutons permettant de passer d'un graphique à l'autre en fonction de la métrique sélectionnée. Le bouton sélectionné sera visuellement foncé afin qu'il puisse savoir quel graphe est actuellement représenté. De plus, lorsque l'utilisateur survole la courbe avec sa souris, cette courbe sélectionnée gardera une couleur foncée tandis que l'autre perdra de la transparence afin qu'ils puissent souligner une métrique afin de plus facilement l'analyser visuellement sans avoir à perdre de vue l'autre courbe.

4.6. Visualisation 5

A1.4 (★★☆) : Comment les indicateurs académiques influencent-ils le temps moyen nécessaire pour obtenir une promotion ?

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 ?

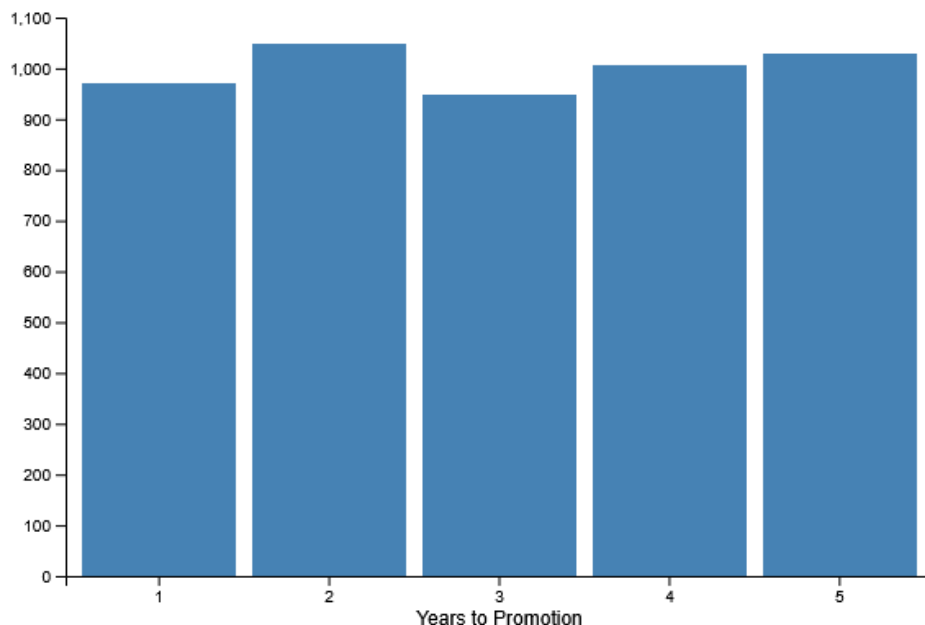


Figure 6 : Diagramme à barres

Description

Ce diagramme à barres présente la distribution des « années jusqu'à la première promotion » au sein de la population d'étudiants étudiée. L'axe des abscisses (x) représente le nombre d'années avant la première promotion obtenue et l'intervalle est de 1 à 5 ans. L'axe des ordonnées (y) indique le nombre d'étudiants correspondant à chaque intervalle, qui communément compose un ensemble de 5000 points de données. Ce visuel offre ainsi une vue d'ensemble sur le délai moyen d'avancement professionnel et permet d'explorer comment certaines variables académiques pourraient influencer ce délai.

Interactivité

Un survol des barres affichera des détails précis, le nombre exact de personnes et un rappel du nombre d'années avant la promotion. En option, un filtre additionnel pourra permettre de comparer ces distributions selon des tranches d'indicateurs académiques (par exemple, GPA ou classement), offrant une analyse plus fine de l'impact de ces facteurs sur le temps de promotion. L'utilisateur pourra trier les barres en ordre croissant ou décroissant pour mettre en évidence les tendances de la distribution.