

# Exam: Atelier : Intégration de données

Anis Ghersbraham

05April 2024

## 1 Introduction

À l'approche de la remise des diplômes de la promotion 2022, la question de l'enseignement supérieur se pose pour de nombreux étudiants. Les décisions relatives au choix d'une université sont souvent influencées par des considérations financières. Dans ce rapport, nous examinons les frais de scolarité et les coûts de séjour des collèges privés et publics aux États-Unis pour l'année 2020, en mettant en lumière les réductions offertes aux étudiants résidents. Ces analyses ont été rendues possibles grâce à l'utilisation d'Alteryx et d'autres outils analytiques. L'objectif est de fournir des informations pertinentes pour aider les futurs étudiants et leurs familles à prendre des décisions éclairées.

## 2 Les outils utilisées pour ce projet

- 1- Filter
- 2-Select
- 3-Summarize
- 4-Semple
- 5-Sort

## 3 Le role de chaque outils

**1-Filter:** L'outil de filtrage est généralement utilisé pour sélectionner des données spécifiques en fonction de critères prédéfinis. Dans le contexte de ce projet, l'outil de filtrage à était utilisé pour extraire les données pertinentes concernant les collèges privés et publics aux États-Unis, ainsi que pour filtrer les données relatives aux frais de scolarité et aux coûts de séjour pour l'année 2020. Cela aurait permis de réduire le volume de données à analyser et de se concentrer uniquement sur les informations nécessaires pour répondre aux objectifs du projet.

- **2-*select*:** L'outil de sélection est utilisé pour choisir des colonnes spécifiques à inclure dans l'analyse ou à exclure de l'ensemble de données. Dans ce projet, l'outil de sélection est utilisé pour choisir les colonnes pertinentes contenant des informations telles que les noms des collèges, les frais de scolarité, les coûts de séjour, les États, les types d'établissements (privé ou public), et d'autres données nécessaires à votre analyse. Cela permet de focaliser l'attention sur les variables importantes pour répondre aux questions spécifiques posées dans votre projet, tout en réduisant la complexité des données.

**3-*Summarize*:** L'outil de résumé est comme un assistant qui réduit une grande quantité d'informations en quelques points clés. Il nous permet de comprendre rapidement ce qui se passe dans nos données. Par exemple, il peut calculer des statistiques comme la moyenne, la somme, le minimum, le maximum ou même le nombre d'éléments uniques dans une colonne. C'est très utile pour obtenir une vue d'ensemble de nos données et repérer les tendances importantes sans avoir à passer trop de temps à les examiner en détail. En gros, ça nous aide à faire la lumière sur ce qui compte vraiment dans nos données.

- **4-*Sample*:** Il est utilisé pour prélever un échantillon aléatoire de données à partir d'un ensemble de données plus large. Dans ce projet, l'outil "Sample" aurait été utilisé pour sélectionner aléatoirement une partie des données sur les collèges privés et publics aux États-Unis, ce qui pourrait être utile pour des analyses exploratoires ou pour vérifier la qualité des données sans avoir à traiter l'ensemble du jeu de données. Cela permet d'avoir une idée générale des données sans devoir les examiner dans leur totalité.
- **5-*Sort*:** Il est utilisé pour organiser les données des collèges selon un critère spécifique, tel que le nom du collège ou le montant des frais de scolarité. Cela a permis de structurer les données de manière à les rendre plus lisibles et plus faciles à analyser.

## 4 Dans ce projet

- **1-*Filter*:** Sélectionner les établissements des quatre dernières années.
- **2-*Select*:** On a modifié le type de la colonne Value en choisissant Int64 et une taille 8
- **3-*Summarize*:** Calculer les statistiques agrégées pour obtenir les paiements totaux
- **4-*Sample*:** prélever un petit échantillon aléatoire de données parmi l'ensemble des données des collèges privés et publics aux États-Unis.
- **5-*Sort*:** On choisit Sum value et on la met dans l'ordre croissant

## 5 Conclusion

En résumé, ce projet a utilisé des outils analytiques tels qu'Alteryx pour explorer les données des collèges privés et publics aux États-Unis. En se concentrant sur les frais de scolarité et les coûts de séjour, nous avons pu identifier des tendances significatives et mettre en évidence les différences entre les établissements. En intégrant davantage de données, telles que les performances académiques ou les taux de placement professionnel, ce projet pourrait fournir une vue encore plus détaillée des collèges américains. En fin de compte, ces analyses offrent des informations cruciales pour aider les futurs étudiants et leurs familles à prendre des décisions éclairées sur leur parcours éducatif.