

Intégration des données

Sokhna Awa Falilou SOW

April 5, 2024



Workflow

Voici le workflow demandé

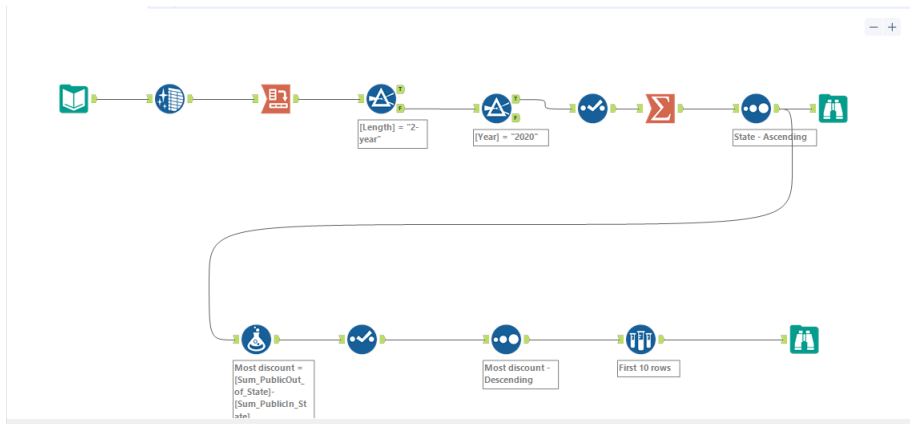


Figure: Workflow

Table 1: Workflow 2020 College

Pour Créer une liste d'établissements d'enseignement supérieur de quatre ans par État qui comprend le paiement total (droits et frais de scolarité plus chambre et pension) en 2020 pour les établissements privés et les établissements publics pour les étudiants de l'État et les étudiants hors de l'État. j'ai:

- nettoyé les données avec l'outil **Data Cleansing**
- transformé les lignes de la colonne Type en colonnes(PublicInState, PublicOutState) avec l'outil **Cross Tab**
- posé 2 outils **Filter** l'un pour séparer les 2year au 4year et l'autre pour prendre les universités à partir de 2020
- utilisé l'outil **Select** pour afficher State, Private, PublicInstate, PublicOutstate
- utilisé l'outil **Summarize** nous permet de regrouper les Etats et additionner les valeurs des colonnes
- utilisé l'outil **Sort** nous permet organiser les State par ordre alphabétique.

Table 2: 10 schools with most discount

Pour trouver les 10 premiers établissements qui accordent les réductions les plus importantes aux étudiants de l'État. Nous avons utilisé la sortie des données de l'outil **Sort** du premier Workflow et nous avons ajoutés:

- l'outil **Formula** qui nous permet de trouver le Most Discount en faisant une soustraction entre le PublicOutstate et le PublicInstate
- l'outil **Select** pour nous permettre d'afficher le State et le Most Discount
- l'outil **Sort** pour afficher le Most Discount par ordre décroissant
- l'outil **Sample** pour afficher les 10 premières lignes de la table.