|  |  |  |
| --- | --- | --- |
| LogoUPPAnoir.png | Formation : | **L1 informatique NEC** |
| Module : | **Business Analysis 1** |
| Thématique : | **Révisions de Business Intelligence 1** |
| Nom du fichier : | **Attrition** |



|  |  |
| --- | --- |
| Source | <https://www.kaggle.com/datasets/colearninglounge/employee-attrition?select=employee_attrition_train.csv> |
| Présentation | Les données contiennent un ensemble de données sur les employés qui travaillent actuellement au sein de l'entreprise (colonne Attrition = non) ou qui ont démissionné (colonne Attrition = oui). |
| Format | Fichier au format texte (pour ouvrir sur Excel 🡪 onglet Données 🡪 Convertir |



|  |  |  |
| --- | --- | --- |
| **Description** | | **Références** |
| **Nom statistique du tableau** | Tableau de données individuelles |  |
| **Ingrédients** | * En lignes : unités statistiques ou individus statistiques   + Qui ? 1 employé   + Combien ? 1029   + Population / échantillon ? Population   + (indépendance) * En colonnes : ID + variables statistiques   + Combien ? 34   + ID ? EmployeeNumber   + Dictionnaire de données  |  |  |  |  | | --- | --- | --- | --- | | 1 QL\_n | 1 QL\_o | 1 QT\_d | 1 QT\_c | | Department | BusinessTravel | DistanceFromHome | Age |   Y = Attrition (QL\_n) |  |



|  |  |  |  |
| --- | --- | --- | --- |
| **Manipulations préliminaires éventuelles** | | **Références** | **Résultats** |
| **Nettoyage** | * Replacer les « . » par des « , » * Valeurs manquantes = cases vides * Pas de lignes entièrement vides ni de colonnes entièrement vides | TD2\_récupération des données.docx |  |
| **Codage** | * RechercheV() * Si() qui gère les valeurs manquantes | TD3\_retravailler ses données.docx | Exemple Department   1. Tableau de codage :      1. Résultat obtenu : Department\_cod |
| **Recodage** |  | TD3\_retravailler ses données.docx | Exemple Monthly Income   1. Tableau de recodage      1. Résultat obtenu : Monthly\_incom\_recod     Remarque : le choix des classes a été imposé. On pourrait faire ses propres choix ou suivre un critère mathématique (règle de Sturges, calcul de l’amplitude, construction du tableau avec les bornes…) |
| **Calculs** | * On peut utiliser : +,-,\*,/ entre plusieurs colonnes (QT) * On peut utiliser : concatener() entre plusieurs colonnes (QL, QT) | TD3\_retravailler ses données.docx | Exemple Satisfaction  On concatène Job Satisfaction avec RelationShip Satisfaction    Exemple %\_time\_current\_role  YearsInCurrentRole / YearsAtCompany en % |



|  |  |  |  |
| --- | --- | --- | --- |
| **Analyse statistique descriptive** | | **Références** | **Résultats** |
| **1 QT\_c** | * **Tableau synthétique :** * **Graphique(s) : Histogramme + Boxplot** * **Indicateurs :** | TD4\_analyse statistique descriptive univariée.docx | Exemple de Monthly Income |
| **1 QT\_d** |  |  |
| **1 QL\_n :** |  |  |
| **1 QL\_o** |  |  |

|  |
| --- |
| **REMARQUES** |
|  |