| Point de contrôle qualité | Problème observé et commentaire | | | Correction : comment ? importance ? | | | | | |
|---|--|----------------|-----------------|-------------------------------------|--|--|------------|--|--|
| 1. le jeu de données est difficilement accessible (format "image", PDF, HTML), le fichier est mal formé | NON Il y a 3 sexes indiqués M, F et I ce qui nous a interpellé. Après vérification, il y a une « Sara » qui a comme sexe « I ». Nous l'avons détecté grâce à Breve. On a lu la documentation via le schéma associé (car ce jeu données a adopté le schéma « prénoms des nouveaux nés »). C'est bien précisé que dans la propriété « ENFANT_SEXE » 3 valeurs sont possibles (F pour féminin, M pour masculin, I pour indéterminé/interséxué). Par ailleurs, le jeu de données est téléchargeable au format CSV et XLS mais j'avais un | | | | Gravite Faible: Pour résoudre le problème des caractères accentués, j'ai visionné et téléchargé le fichier XLS puis je l'ai enregistré sous un format CSV et les caractères accentués apparaissent bien. Mais nous avons dû refaire une réinstallation de LibreOffice pour plus de sécurité (éviter les formats propriétaires) et cela a finalement réglé de problème d'accentuation. C'était juste un problème de mise à jour du logiciel LibreOffice. Donc nous n'avons plus besoin du fichier xls en fin de compte. | | | | |
| | problème d'affi fichier csv. | chage des cara | actères accentu | • | A 1 COMMUNE_N 2 Antibes 3 Antibes 4 Antibes 5 Antibes 6 Antibes 7 Antibes 10 Antibes 11 Antibes 12 Antibes 13 Antibes 14 Antibes 15 Antibes 16 Antibes 17 Antibes 18 Antibes 19 Antibes 20 Antibes 21 Antibes 22 Antibes 23 Antibes 24 Antibes 25 Antibes 26 Antibes | B Id COLL_INSEE 06004 | ENFANT_SEX | ENFANT_PRENOM Lina Hugo Maëlys Yassine Inès Léa Adam Sacha Chloé Maxime Théo Emma Manon Sarah Ayoub Enzo Jules Mattéo Nathan Kenza Lola Nour Adem Gabriel Kylian | E NOMBRE_OCCURF 14 11 10 10 10 8 8 8 8 7 7 7 6 6 6 6 6 6 6 6 6 5 5 5 5 5 5 5 5 |

| 2. la licence est absente ou inhabituelle, le jeu de données n'est pas "open data" | NON Licence ouverte V2: Le « Réutilisateur » est libre de réutiliser l' « Information » : de la reproduire, la copier, de l'adapter, la modifier, l'extraire et la transformer, pour créer des « Informations dérivées », des produits ou des services, de la communiquer, la diffuser, la redistribuer, la publier et la transmettre, de l'exploiter à titre commercial, par exemple en la combinant avec d'autres | IMPORTANT : Citer la source comme indiquée dans les termes de la licence en donnant le nom de la commune d'Antibes et en mettant le lien vers les jeux de données : https://trouver.datasud.fr/dataset/prenoms-des-nouveaux-nes-de-la-ville-antibes-juan-les-pins |
|---|---|---|
| | informations, ou en l'incluant dans son propre produit ou application. Sous réserve de : mentionner la paternité de l' « Information » : sa source (au moins le nom du « Concédant ») et la date de dernière mise à jour de l' « Information » réutilisée. Source https://www.etalab.gouv.fr/licence-ouverte-open-licence/ | |
| 3. le fichier fait peu appel aux standards répandu et aux données pivots | NON: Les données respectent le schéma des prénoms https://schema.data.gouv.fr/scdl/prenoms/ . Le code postale indiqué dans le fichier CSV fait bien référence aux Code de l'INSEE: https://www.insee.fr/fr/information/6800675 | N.A |
| 4. le fichier est mal documenté | NON Je n'ai pas de documentation directement sur la plateforme de la ville d'Antibes mais cela s'explique par le fait que ça respecte un schéma et dans le schéma les « Propriétés » sont expliquées ainsi que le caractère obligatoire ou non. On n'a pas directement la documentation sur la plateforme mais le schéma documente bien les champs et attentes. Ce serait bien de l'avoir directement sur le site de la ville d'Antibes dans l'onglet information du jeu de données pour avoir un raccourci et ne pas aller sur schema.data.gouv.fr mais ce n'est pas bloquant car la documentation existe. Il faut aller la chercher. | Gravité faible : Voir le schéma pour comprendre les différents intitulés https://schema.data.gouv.fr/scdl/prenoms/ |

| 5. il existe des problèmes de syntaxe | OUI D'après l'analyse faite sur https://publier.etalab.studio/fr/ il y a 2 erreurs aux lignes 7310, 7355. | Gravité Faible : on a rectifié la syntaxe des prénoms. Puis on l'a remis sur https://publier.etalab.studio/fr/ et c'était conforme. |
|--|--|---|
|--|--|---|

| Point de contrôle qualité | Problème observé et commentaire | Correction : comment ? importance ? |
|---|---|-------------------------------------|
| 6. Valeurs aberrantes, suspectes, inexplicables, pas crédibles | NON | N.A |
| 7. Il manque des données et cela n'est pas documenté (trous, données tronquées, valeurs vides, granularité / fréquence / maillage / fraîcheur) | NON | N.A |
| 8. Trop de données : doublons, inutilement vieilles, précision / fréquence / maillage / fraîcheur | NON Il y a des orthographes différentes pour un même prénom mais cela dépend de comment les parents épellent/écrivent le prénom de l'enfant donc ce n'est pas un doublon mais différentes manières d'écrire par exemple « Timothé » et « Timothée » (voir Lignes 2000 et 2001 du fichier CSV) | N.A |
| 9. Données posant problème avec la réglementation (données perso, relatives à la santé, la religion, propriété littéraire et artistique, etc.) | NON Il n'y a que les prénoms, s'il y avait les noms, ce serait une violation des données personnelles mais ce n'est pas le cas. | N.A |
| 10. Les contenus posent problèmes : synonymies, non traduits (USA), cryptique (DAECPP), utilisation du 0 au lieu du "null" | NON | N.A |

Sprint qualité du jeu de données des prénoms des nouveaux nés de la ville d'Antibes Juan-les-Pins de 2011 à 2022 par Dieynaba KOUYATE accessible sur https://trouver.datasud.fr/dataset/prenoms-des-nouveaux-nes-de-la-ville-antibes-juan-les-pins

Comme on peut le constater en partie dans l'analyse de Breve : ces données sont de bonnes qualités. D'ailleurs, avant de les choisir nous avons vérifié sur Open data la qualité des métadonnées et c'était vert. Ces données nous semblent tout à fait exploitables et répondent aux critères de l'Open Data.



Breve. See your data.

