

Evaluation des dossiers sur la Programmation en R

Les « 12 travaux d'AstéRix »

Daif Hakim

31/01/2021

Package Janitor

1. Critères d'évaluation

Dans ce dossier, nous allons parcourir et faire une évaluation des travaux de mes camarades sur la Programmation en R:

Nos critères seront :

1. Critères d'évaluation

Dans ce dossier, nous allons parcourir et faire une évaluation des travaux de mes camarades sur la Programmation en R:

Nos critères seront :

- 1. Visuel et l'organisation du dossier:** Es ce que la mise en page est propre ? Le travail est-il organisé (chapitres, sections, sous-sections, Parties) ?
- 2. Qualité de la rédaction du dossier:** le document est-il agréable à lire? es ce que l'auteur a respecté les consignes de rédaction d'un article scientifique ou d'une thèse... ?
- 3. Compréhension de l'idée générale:**Les Auteurs arrivent-ils à transmettre l'idée générale ?
- 4. lisibilité et fonctionnement du code (explication et commentaires):** Es ce que le code fonctionne parfaitement sans erreurs ? Es ce que l'auteur a mis des commentaires pour expliquer chaque ligne de commande ?
- 5. L'utilisation d'exemples et de cas concrets (cas pratique):** l'Auteur a t'il utilisé des exemples et des cas pratique lors de sa rédaction ?

2. Lien vers le document commenté

Pour ce dossier, nous avons étudié le travail de Imen Derrouiche qui porte sur le Package Janitor
le lien (github) vers son dossier est :

<https://github.com/imenderrouiche/PSBX/blob/main/JANITOR%20DOCUMENTATION.pdf>

Le dossier en question est le dossier **JANITOR DOCUMENTATION.pdf**

3. Synthèse du document

Dans ce dossier, IMENE nous a présenté un package nommé JANITOR conçu pour les utilisateurs R débutants à intermédiaires, optimisé pour la convivialité et dont la fonctionnalité est d'examiner et de nettoyer des jeux de données de façon simple et rapide. Ce travail a bien détaillé le fonctionnement de ce package sous R avec une application sur un exemple concret de jeu de données où elle a commencé par la récupération du jeu afin de le charger, le nettoyer en supprimant les colonnes vides et les doublons ; avec une dernière astuce de conversion des numériques en date.

4. Extrait commenté :

Comme tout projet traitant les packages, on y trouve le chargement des données au tout début avec la fonction:

```
mymsa = read_excel("/cloud/project/mymsa.xlsx")
```

NB: cloud/project/mymsa.xlsx l'emplacement des données utilisées pour ces exemples.

Pour entamer ensuite le nettoyage des noms avec la fonction :«clean.names()»:

```
x = janitor::clean_names(mymsa)
data.frame(mymsa = colnames(mymsa), x = colnames(x))
```

Après cette opération, elle a procédé à la suppression des colonnes vides via cette fonction :

```
x = remove_empty(x, which = c("rows", "cols"))
```

l'auteur précise que c'était possible de le faire directement au début de l'opération au chargement des données avec la fonction suivante:

```
x = read_excel("/cloud/project/mymsa.xlsx") %>%
clean_names() %>% remove_empty()
```

En fin, il est indispensable de vérifier l'existence des doublons pour optimiser le résultat tiré de cette base de données et ce, avec la fonction « get_dupes() »

```
x %>% get_dupes(rfid)
```

5. Evaluation du travail suivant les 5 critères précités

1. Visuel et l'organisation du dossier

Le visuel du document manque d'une touche finale, les sous-titres ont la même police que le corps du document et reste donc, moins visible.

2. Qualité de la rédaction du dossier

La qualité rédactionnelle ne peut être jugée faute de manque de rédaction, de courtes lignes introductives de chaque partie sont rédigées, on ne trouve pas aussi le format Rmd pour juger sa compréhension en programmation en R et l'utilisation des chunks

3. Compréhension de l'idée générale:

L'idée générale est assez bien expliquée avec l'utilisation d'exemple pour faciliter la compréhension. 4. lisibilité et fonctionnement du code (explication et commentaires):

Le code est bien lisible dans les cadres grisés, son exécution est enregistrée sans aucune erreur donnant un résultat similaire (utilisation d'une autre base de données que celle utilisé par l'auteur)

5. L'utilisation d'exemples et de cas concrets (cas pratique) : Le premier argument de ce package est son utilisation dédiée aux débutants et intermédiaires en R, donc ne nécessite pas de grandes connaissances comme il est indiqué tout au long des étapes de traitement de la base de données assez simplifiées.

On constate aussi le manque d'un cas pratique qui permettra d'appliquer toutes fonctionnalités de ce package dans le cadre d'une étude (cas pratique), néanmoins IMEN n'a pas manqué d'utiliser le maximum d'exemples simples pour expliquer le fonctionnement du code et de l'algorithme, ce qui est bien !

6. Conclusion

Le choix du thème est excellent avec une démonstration simplifiée pour obtenir des résultats pertinents. Le travail est en gros bon, a bien cerné l'utilisation du package JANITOR avec des conseils à prendre en considération lors de notre utilisation ultérieure, cependant on aurait aimé voir un cas pratique pour mieux illustrer son utilisation.