

TD - Pandas - Matplotlib

A partir des données airbnb, répondez à ces questions en proposant à chaque fois une visualisation avec seaborn ou matplotlib:

1. Calculez le nombre d'host et le nombre de review par quartier. (neighbourhood_cleansed)?
2. Quels sont les pourcentage d'acceptation et de réponse moyen des hosts?
3. Quel pourcentage des hosts on une vérification:
 - par téléphone
 - par email
 - par email pro
4. En fonction du room type, calculez le nombre moyen d' "amenities" son écart type.
5. Quel est le prix median par type de room, le maximum, le minimum, le premier et le troisième quartile?
6. Combien y-a-t-il d'annonces par valeur de salle de bain? (0.5 / 1 / 1.5)?
7. Si on considère qu'un "bath" vaut 1, qu'une bath shared et une half-bath vaut 0.5 et qu'un private bath vaut 2. (2 half bath vaut donc 1)
8. Il y a t'il une corrélation entre le nombre de reviews et la taille de la description en termes de nombre de caractères?
9. On suspecte des faux commentaires. Pour identifier cela on regarde les séjours où le nom de l'host est le même que le nom du reviewer. Quel est ce pourcentage général de séjour où le nom de l'host est identique au nom du reviewer?