

TP5-6-7

Données libres et visualisation

1. Prise en main des données et préparation

Q₁). Récupérez le fichier `vgsales.csv` sur Moodle. Il s'agit de données libres sur les ventes de jeux vidéos en Europe, aux États-Unis, au Japon et ailleurs dans le monde. Les valeurs sont indiquées en millions d'exemplaires. Analysez les données de ce fichier pour en déterminer le contenu et concevoir une table pouvant contenir ces informations. Faites attention au choix de votre clé primaire.

Vous pouvez ouvrir le fichier dans un éditeur de texte classique, dans un tableur (Libre Office Calc, par exemple), ou utiliser Python pour parcourir les données.

Q₂). Écrire un script python permettant de créer cette table dans votre base de données PostgreSQL et d'insérer les données de `vgsales.csv` dedans. Choisissez judicieusement les types de vos colonnes. Attention, le nom des jeux peut être composé de près de 150 caractères.

Q₃). Si vous exécutez votre script, vous allez obtenir des erreurs car les données ne sont pas "propres". Il y a des lignes qui sont dupliquées. Modifiez votre script pour éliminer ces lignes avant l'ajout en base.

Q₄). Sur votre base, vérifiez que la table a été créée en listant les tables et que l'insertion a fonctionné en comptant le nombre de lignes dans votre table.

Q₅). La colonne `Global_Sales` est redondante car elle peut être calculée grâce aux autres colonnes. Cela pose d'ailleurs problème puisque certaines données sont incohérentes. Écrire une requête SQL permettant de détecter les lignes incohérentes.

Q₆). Écrire une requête SQL permettant de recalculer la colonne `Global_sales` en fonction des autres colonnes pour ces lignes problématiques.

Q₇). Écrire un script SQL permettant de créer un DataFrame à partir de votre table Games mise à jour.

2. Analyse des ventes de jeux par année

Q₁). Écrire la requête SQL permettant de calculer les ventes moyennes par année de sortie des jeux, pour les jeux du genre '`Adventure`'.

Attention, pensez à exclure les jeux pour lesquels on ne connaît pas l'année de sortie.

Q₂). Afficher le résultat de la requête précédente sous forme de courbe.

Q₃). Même question pour le total des ventes par année de sortie des jeux.

Q₄). Même question pour l'ensemble des jeux.

3. Analyse des ventes de jeux par plateforme

- Q₁). Écrire la requête SQL permettant de calculer le total des ventes en Europe par plateforme.
- Q₂). Afficher sous forme de courbe les résultats de la requête précédente. Votre graphique devra respecter les contraintes suivantes :
- Les plateformes sont sur l'axe des abscisses et le nombre de ventes sur l'axe des ordonnées.
 - Les axes sont renommés : **'Plateforme'** pour l'axe des abscisses, **'Ventes (en millions)'** pour l'axe des ordonnées.
 - Les noms des plateformes doivent apparaître.
 - Il n'y a qu'une seule courbe, la légende n'est pas nécessaire.
 - L'axe des ordonnées doit commencer à 0.
- Q₃). Même question sous forme d'histogramme.
- Q₄). Écrire la requête SQL permettant de calculer le total des ventes aux Etats-Unis, en Europe, au Japon et ailleurs par plateforme.
- Q₅). Afficher sous forme de graphique les résultats de la requête précédente avec une courbe par zone géographique. Cette fois-ci la légende devra apparaître.
- Q₆). Même question mais en changeant le style des courbes, autrement dit le type de trait et de marqueurs (les points sur la courbe), pour que ce soit lisible y compris imprimé en noir et blanc.
- Q₇). Même question avec un diagramme en bâtons.

4. Analyse des ventes d'un jeu

- Q₁). Écrire la requête SQL de récupérer les ventes (Etats-unis, Europe, Japon, Autre) du jeu Mario Kart 64.
- Q₂). Afficher les résultats de la requête précédente sous forme de diagramme en bâtons. Pensez à enlever la légende et à adapter les labels des axes.
- Q₃). Même question sous forme de diagramme "camembert". Les noms des zones devront être à côté de la tranche concernée.
- Q₄). Même question mais le pourcentage de ventes dans chaque zone doit apparaître à côté de la tranche concernée et le nom de la zone doit apparaître dans la légende.
- Q₅). Faire la même chose avec le jeu Mario Kart Wii et comparez les résultats

5. Analyse des ventes de jeux par genre

- Q₁). Écrire la requête SQL permettant de calculer le total des ventes de jeux dans le monde par genre.
- Q₂). Écrire la requête SQL permettant de calculer le pourcentage des ventes dans le monde par genre.
- Q₃). Afficher les résultats sous forme de diagramme camembert.
- Q₄). (*Bonus*) Afficher le total des ventes par genre sous forme de diagrammes camembert de façon à voir sur la même figure :

- un camembert représentant les ventes avant l'année 1990 (exclue),
- un camembert représentant les ventes entre 1990 et 1999,
- un camembert représentant les ventes entre 2000 et 2009,
- un camembert représentant les ventes à partir de 2010.

Indice : regarder la fonction `subplots` de `matplotlib`.