

TD - Advanced Google Data Studio

- Thème : les 200 meilleurs Box-office au Cinéma en France
 - Source Allociné
 - Fichier BOXOFFICE comprenant les 200 films ayant eu le plus gros nombre d'entrées au cinéma en France, avec leurs années de sortie, genre, notes, budgets, distributeurs, réalisateurs, nationalités, durée, nominations
1. Ouvrir Google Data Studio et créer un nouveau rapport. Télécharger le fichier BOXOFFICE en utilisant **Google Sheets** et en veillant à mettre les bons types associés aux données.
Les différents types à changer sont : Nationalité (en données géographiques Pays), Année de sortie (en date Année AAAA), Budget (en Devises US\$).
 2. Présentation :
Créer votre propre présentation (fond, titre, couleurs, portrait/paysage etc.).
Ajouter le logo de l'université disponible sur le dossier drive (Bouton Image).
Grouper les éléments titre et logo (via le menu Organisation).
Et les 'placer au niveau du rapport', afin de le mettre en titre et le dupliquer sur toutes les pages.
 3. Classement des 20 meilleurs films :
Faire un Tableau donnant le classement des 20 meilleurs box-offices, leurs budgets et année de sortie. Aller dans Style pour ensuite améliorer votre présentation :
"compresser les nombres", "afficher la pagination", bordures etc.
 4. Provenance des films :
 - a. Ajouter une carte via l'objet 'Synthèse Géographique' avec les variables Nationalité et Box-Office. Créer une 2nd carte de la même manière en appliquant un zoom sur l'Europe. Que constatez-vous ?
 - b. Créer un nouveau champs 'Provenance' à 3 modalités (France, Etats-Unis, Autres) à l'aide de l'instruction CASE.
 - c. Ajouter un tableau pour avoir le nombre de films selon cette nouvelle variable provenance.
 5. Répartition des films par distributeurs :
Créer un diagramme (Pie Chart) qui affiche la répartition des films par distributeurs selon le box-office. Tester différents styles (cercle central, couleurs etc.). Augmenter le nombre de secteurs pour éviter d'avoir trop de modalités dans 'Autres'.

6. Répartition des films par genre :
 - a. Créer un autre diagramme (Pie Chart) qui affiche la répartition des films par genre selon le box-office.
 - b. Refaire 2 mêmes graphiques pour différencier la provenance : un pour la France et l'autre pour les Etats-Unis (via la création de filtres).
Penser à ajouter des titres aux graphiques.

7. Budget Moyen et Ecart-Type :
 - a. Créer un Tableau de données pour calculer le budget moyen par film.
Renommer le titre du champ en '*Budget moyen*' et Compresser les nombres.
Regarder bien le résultat et regarder aussi la colonne budget du jeu de données dans Google Sheets. **Que constatez-vous ?**
La variable budget contient plusieurs cellules vides. Le logiciel en tient compte pour le calcul de la moyenne.
 - b. Recréer cet indicateur deux fois pour les films français et américains (dupliquer et utiliser les filtres créés en question 5).
 - c. Créer un nouveau champ calculé permettant de définir l'écart-type du budget (Fonction STDDEV).

8. Réalisateurs et séries :
 - a. Réaliser une jointure avec la table REALISATEURS sur le nom des films.
 - b. Créer le classement des réalisateurs par nombre de films et par box-office.
 - c. Ajouter un 'tableau de données' comptant le nombre de films correspondant à des séries.
 - d. Faire un nouveau tableau de classement comme en question a) en prenant en dimension les films à séries (variable Nom Série). Penser à enlever la modalité 'nul' via un filtre.
 - e. Rajouter dans le tableau en Statistique le nombre de films (d'épisodes) par série de films ainsi que le box-office moyen.

9. Commande de filtrage :

Créer une 2e page (le titre devrait apparaître automatiquement).
Reprendre le Tableau de classement des meilleurs Box-Office de la 1^{ère} page (via un copier/coller).
Ajouter une commande de filtrage par provenance.
Tester son fonctionnement sur le mode Vue

10. Comparaison notes presse/spectateur :

- a. Créer un nuage de points sur la base de ces 2 champs. Adapter les échelles et afficher les libellés des données. **Que constatez-vous ?**
- b. Afficher une courbe de tendance sur le nuage de points.
- c. Ajouter ensuite une dimension 'Genre' et coloriser selon cette dimension.

11. Pour aller plus loin dans cette analyse

- a. Créer un nouveau champs 'Diff Notes' : $ABS(\text{note presse} - \text{note spectateur})$.
- b. Créer pour commencer un petit tableau avec cette nouvelle variable en statistique pour afficher les films avec la plus grosse différence de notes.
- c. Créer un indicateur (tableau de données) pour avoir la moyenne de cette différence.
- d. Puis un autre indicateur pour avoir le nombre de films où cette différence est supérieure à 1 (via l'ajout d'un filtre : $INCLUDE \text{Diff notes} \geq 1$).

12. Série temporelle :

- a. Créer un graphique 'série temporelle' en fonction du budget moyen.
- b. Refaire le même travail en utilisant un 'graphique combiné' : Evolution dans le temps (1938 à aujourd'hui) du budget global (en barre) et du budget moyen (courbe) par année. Penser à trier le graphique par année, et à adapter l'échelle.
- c. Créer un histogramme pour avoir le box-office moyen des films par année de sortie. Attention à bien adapter les échelles et le nombre de barres. **Que constatez-vous ?**

13. Genre et nominations :

- a. Faire un tableau du nombre d'entrées moyennes par genre.
- b. Ajouter un histogramme décroissant du nombre moyen de nominations par genre.

14. Durée ordinale :

- a. La durée en minutes est une variable peu utilisable sauf si on la découpe en tranches ordinales. Créer donc par un champ calculé une nouvelle variable 'Durée en classe' avec une découpe selon les modalités suivantes : +3h / 2h30-3h / 2H10-2H30 / 1H50-2h10 / 1h30-1h50 / -1h30
Utiliser les fonctions CASE et WHEN vues en question 4.
- b. Dans une nouvelle feuille, créer un tableau qui décompte le nombre de films, la moyenne du box-office et le budget moyen par durée en classe.
Que constatez-vous ?
- c. Recréer un Pie Chart sur le genre des films (comme en question 6).
Puis rajouter une commande de filtrage sur la durée en classe et observer les liaisons entre le genre des films et leurs durées (particulièrement pour les comédies, dessins-animées et films historiques).

15. Tableaux croisés dynamiques :

Réaliser 3 tableaux croisés dynamiques pour voir le nombre de films en fonction de la provenance, le genre et la durée en classe. Utiliser la notion de cartes de densité pour faire ressortir les éléments importants. Utiliser le total général éventuellement et compléter la présentation. **Que constatez-vous ?**

16. Classement nominations : Afficher le classement des nominations par film et rajouter une colonne pour les rangs.

17. Nuage de points :

- a. Créer graphique en nuage de points pour observer le nombre de nominations et le nombre de prix. Ajouter une courbe de tendance et adapter les échelles des axes.
- b. Dupliquer ce graphique et appliquer un changement de dimension par genre. Tester les couleurs et réadapter les échelles.

N'hésitez pas s'il vous reste du temps, à tester et découvrir d'autres fonctionnalités en se baladant sur les menus du logiciel. Eventuellement, importer un autre tableau de données Google Sheets et combiner les données par jointure (en utilisant la nationalité par exemple).

Bon courage !