

Cahier d'Avancement - Projet de Data Visualisation

1. Informations Générales

- **Titre du Projet** : Exploration et Visualisation des données Netflix
 - **Date de Début** : 18 décembre 2024
 - **Membres de l'équipe** : Moussaoui Youssef / RKIBA HICHAM
-

2. Objectifs du Projet

L'objectif de ce projet est d'explorer et de visualiser de manière interactive les données du dataset Kaggle "[Netflix Shows](#)" en utilisant **D3.js**. L'analyse porte sur les tendances clés du contenu Netflix, notamment la répartition des films et séries par année de sortie, genre et type de contenu. Des visualisations dynamiques, telles que des diagrammes à barres, maps et graphiques interactifs, permettent aux utilisateurs de filtrer les données selon différents critères (année, genre, pays d'origine) et de mieux comprendre l'évolution du catalogue au fil du temps. Ce projet vise ainsi à communiquer efficacement les insights sur la diversité et les tendances du contenu Netflix.

3. Progression Hebdomadaire

Semaine 1 (16/12) : Exploration des Données

- **Avancement** :
 - Faire une demande à Netflix pour récupérer les données personnelles.
 - Chargement et nettoyage initial des données.
 - Analyse exploratoire des différentes tables et colonnes.
 - Détection et gestion des valeurs manquantes.

Semaine 2 (23/12): Création de Document de cadrage

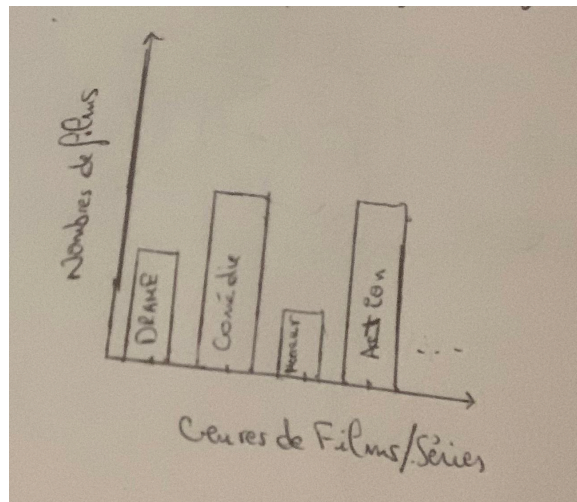
Avancement :

- Définir les premières pistes de conception
- Finaliser le cadrage du projet.
- Création de premières esquisses.

Esquisses (papier ou capture d'écran) :

- Esquisse d'un graphe en barre : Répartition des genres regardés

- Papier (photo) :



Heatmap temporelle : Activité de visionnage selon les heures et les jours.

- Papier (photo) :

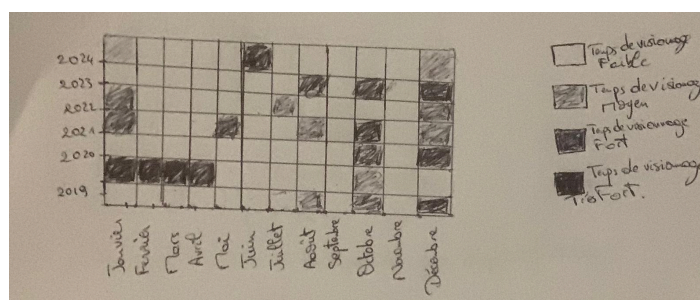
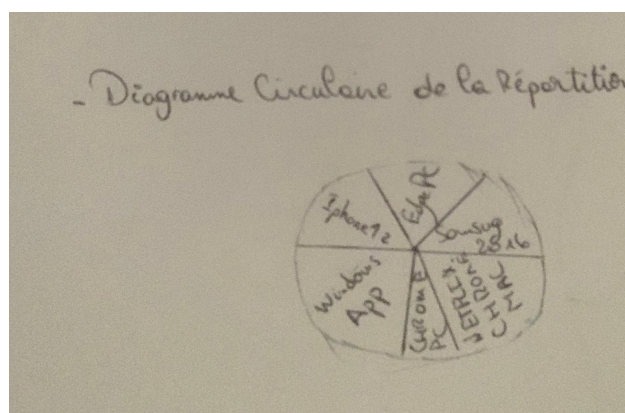


Diagramme circulaire : Répartition par dispositifs utilisés.

- Papier (photo) :



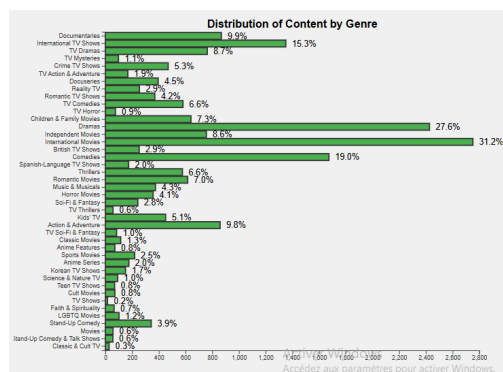
Semaine 3 (30/12): Changement de données

- **Avancement :**
 - Abandon des données Netflix au profit d'un dataset similaire trouvé sur Kaggle.
 - Création d'un tableau de bord dans Tableau pour centraliser les visualisations.

Semaine 4 (06/01) : Début de codage avec D3.js

- **Avancement :**
 - Développement d'une interface interactive pour les visualisations en se basant sur l'année de création des films (utilisation de **D3.js** pour des animations dynamiques).
 - Création d'une visualisation sous forme de **diagramme à barres dynamique** représentant la distribution de contenu par genre, avec des barres interactives affichant des informations supplémentaires au survol.
 - Conception d'une visualisation interactive pour la distribution des **films et des TV shows**, permettant de filtrer les résultats par année ou par genre grâce à des menus déroulants.

- **Esquisses (capture d'écran):**



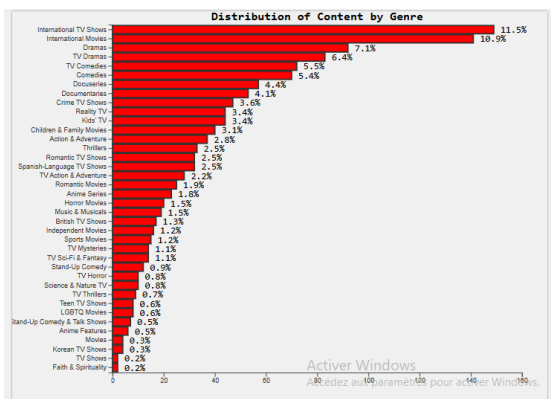
Semaine 5 (13/01) : Intégration et Ajustements

- **Avancement :**
 - Finalisation des différentes visualisations créées avec **D3.js**.
 - Intégration des visualisations dans un tableau de bord interactif et responsive.
 - Ajustements des couleurs et des échelles pour maximiser la lisibilité et s'assurer que les graphiques respectent les principes de design (contraste, hiérarchie visuelle).

- Tests sur différents navigateurs pour garantir la compatibilité des visualisations basées sur D3.js.

Semaine 6 (20/01) : Retouches demandées par le professeur

- **Avancement :**
 - Nettoyage et amélioration des **légendes graphiques** pour une meilleure compréhension des visualisations.
 - Réorganisation et tri des dates de sortie des films (axe des abscisses) pour assurer un affichage chronologique précis dans les diagrammes interactifs.
 - Ajustements mineurs pour optimiser les performances des visualisations D3.js sur de grands datasets.
- **Esquisses (capture d'écran): Modification après les remarques du professeur**



4. Erreurs et Choix Non Retenus

- Problèmes de données :**
 - **Problème :** Les données Netflix initiales contenaient un grand nombre de valeurs importantes manquantes, notamment sur des colonnes clés comme le temps de visionnage ou le genre des utilisateurs.
 - **Décision :** Abandon des données Netflix au profit d'un dataset similaire trouvé sur Kaggle.
 - **Justification :** Les données Kaggle étaient plus complètes et mieux documentées, facilitant le travail d'analyse.
- Problèmes techniques :**
 - Les fichiers CSV de Kaggle comportaient des colonnes avec des noms longs et complexes, ce qui compliquerait leur manipulation.
 - **Solution :** Renommer et simplifier les colonnes avec Python avant de les charger dans Tableau.
- Choix non retenus :**

- Utilisation initiale de graphiques circulaires pour représenter la répartition des genres → Trop confus pour des données avec plusieurs catégories.
- **Solution adoptée** : Passage à des barplots et des histogrammes pour plus de clarté.