



INTERNATIONAL DATA SCIENCE INSTITUTE

Projet de visualisation des données

Mise en Place de tableau de bord d'analyse des resultats d'un sondage sur Cheikh Anta Diop

Auteurs :

Enseignant :

Année académique : 2023-2024

SOMMAIRE

INTRODUCTION

I - Présentation, Exploration et Traitement

II - Mise en forme et Analyse des données

III - Présentation de l'application Rshiny et Description du dashborad et des graphiques.

INTRODUCTION

L'intellectuel Sénégalais mort en 1923, Cheikh Anta Diop s'est érigé comme un restaurateur de l'histoire Africaine par la mise en lumière de la contribution de l'Afrique noire dans la culture et la civilisation mondiale. Il s'est distingué en tant qu'historien, anthropologue et homme politique sénégalais avec sa forte volonté d'écrire l'histoire africaine précédant la colonisation.

La rédaction de ce rapport fait l'objet de la présentation d'un résumé des résultats de nos travaux relatifs au cadre d'une série d'études des figures marquantes de l'Afrique et de sa Diaspora, un sondage visant à évaluer le niveau auquel est connu Cheikh Anta Diop a été réalisé à cet effet.

Nos différents axes méthodologiques s'articuleront autour de la présentation des données dans un premier temps, ensuite des traitements effectués sur ces données et enfin de la description des graphiques obtenus dans la présentation.

I - Présentation, Exploration et Traitement

- Présentation des données

Le jeu de données soumis à notre étude provient d'un sondage réalisé sur Cheikh Anta Diop et nous avons reçu 2 fichiers Excel l'un contenant les questions et l'autre les réponses du sondage.

Notre travail se fera essentiellement sur les données relatives aux réponses du sondage qui contient 236 lignes et 81 colonnes.

- Traitement des données

Notre ensemble de données contenait 42 colonnes qui ne contiennent que des valeurs manquantes que nous avons décidé de supprimer. Il contient en outre 39 autres colonnes dont les valeurs manquantes ont été traitées au cas par cas comme suit :

- Les colonnes entièrement vides et les 10 dernières colonnes ont été supprimées car elles n'auront pas d'influence sur notre étude.
- Les valeurs manquantes des colonnes âge des participants et l'âge de connaissance de notre auteur sont imputées par la moyenne arrondie des valeurs de chaque colonne. Les valeurs négatives de ces colonnes sont prises en valeurs absolues.
- Les valeurs manquantes des colonnes profession(I-3), œuvre lue(III-1), période historique(II-2), description de l'approche méthodologique(III-3), "enseignants_aut"(IV-4), "perception_gen",
- "influence", "autres_auteur", "pertinence_idees", "principales_critiques"(V-5), "critiques_analyses"(), "celebres_soutien_aut"(IV_4) sont imputées par "aucune réponse"

II - Mise en forme et Analyse des données

- Pour faciliter notre travail nous décidons de renommer respectivement les colonnes suivantes :

"I_1", "I_2a", "I_2b", "I_2c", "I_3", "I_3a", "I_4", "I_4a", "II_1", "II_2", "III_1", "III_2", "IV_1", "IV_2", "IV_3", "IV_4", "IV_5", "V_1", "V_2", "V_3", "VI_1", "VI_2", "VII_1", "VII_2", "VII_3", "VII_4", "VII_5", "VII_6", "VIII_1"

Comme suit :

"connaiss", "Pays_naiss_aut", "Period_naiss", "Period_deces", "nature_aut", "Preciser_nature", "moyens_connaiss", "Preciser_moyens_connaiss", "oeuvres_aut", "Period_hist", "oeuvres_lu", "description_app", "influence", "autres_auteur", "perception_gen", "enseignants_aut", "celebres_soutien_aut", "critiques_analyses", "principales_critiques", "pertinence_idees", "lieu_obtention_info", "Preciser_lieu", "education_user", "Pays_naiss_user", "Pays_resi_user", "age_user", "age_entente_aut", "Poursuivre_contribution", "commentaire_sup".

- Les réponses qui s'apparentait à des réponses négatives ("Aucune idée", "Je sais pas", "Ras" etc..) dans les colonnes suivantes description_app(II_2), enseignants_aut, celebres_soutien_aut, autres_auteur et les valeurs manquantes de la colonne oeuvres_aut(II_1) ont toutes été remplacées par "non".
- Tous les caractères de certaines réponses sont toutes transformés en minuscule
- Un regroupement en 4 modalités a été réaliser sur les variables age_user et age_entente_aut
Enfant=[2:10], Adolescent=[11:18], Jeune adulte[19:35], Adulte[36:70], Vieillard 71ans et plus
- Dans oeuvres_aut, perception_gen nous réalisons les regroupements suivants : Les réponses qui ne corresponde pas au œuvre écrite pas l'auteur sont encodé en mauvaise réponses et les réponses qui diffère par des fautes d'orthographe sont regrouper en modalité similaire
- Une matrice de similarité a été faites sur celebres_soutien_aut pour calculer la distance entre les mots les plus proche pour les regrouper
- Séparation des réponses multiples pour faciliter le regroupement en modalité

III - Présentation de l'application Rshiny et Description du dashborad et des graphiques.

Interface d'authentification avant accès au dashboard aux codes suivants :

Username : kadjassagou_angie

Password : 1234



Statistiques descriptives des données contenues dans notre base de données d'analyse.



L'interface suivante est un dashboard résumant les niveaux de perceptions des personnes interrogées par niveau de perception de générale, ceux approuvant la pertinence des idées de Cheikh Amidou Can par tranche d'âge, par fréquence et un nuage de mots d'apparition des mots clés par critiques apportées par personnes interrogées.

CONCLUSION

Nous retenons en somme des méthodes assez poussées en ce qui concerne les techniques de visualisations de données aux moyens des packages ggplot et shiny. Ce projet portant sur la conception d'une application Rshiny nous a été des plus enrichissant, résumant les niveaux d'influences et de popularité des travaux de Cheikh Anta Diop à travers le Monde. Nous nous rendons également compte au terme de notre analyse que les applications Rshiny sont de réelles solutions porteuses de valeurs en ce qui concerne la visualisation des données et qui font désormais partie de notre ensemble de compétences.