

Projet : programmation Python pour la datascience

A. MySQL :

A.1. Création de la base de données **Projets** dont le schéma est présenté ci-dessous

- **Projet**(code, nom, budget, dateDebut, Region)
- **Employe**(numero, nom, prenom, adresse, age, dateEmbauche, tel, profession, salaire, #projet)

A.2. Peuplement de la base de données

A.2.1. Création d'un dataframe **projet_df** contenant les données ci-dessous.

code	nom	Budget	dateDébut	Région
10	PUMA	125.000.000	12/11/2013	Ziguinchor
20	GOANA	5.000.000.000	06/05/2007	Diourbel
30	ACACIA	300.000.000	20/02/2006	Kaolack
40	MIROR	25.000.000	31/03/2013	
50	BRT	44.000.000.000	11/10/2019	Dakar

A.2.2. Exportation du dataframe **projet_df** vers la table **projet** de la base données **projets**

A.2.3. Importation de données à partir de fichier excel

a. Création d'un fichier Excel **Employes** avec les données ci-dessous.

N°	nom	Prénom	adresse	age	dateEmbauche	Tél	Profession	salaire	Projet
11	DIAHAM E	Moussa	Fann Hock	20	30/04/2013	781234455	Commercial	300000	10
22	DIALLO	Mouhamadou Moustapha	Mermoz	35	21/05/2013	778796656	Technicien	1400000	30
33	DIANE	Fatou Ndiaye	Fass		06/05/2007	765432112	Informaticien	1200000	60
44	LY	Niassa		40	20/02/2006		Gestionnaire	850000	20
55	DIALLO	Abdou Rahmane	Fann Hock	23	06/05/2007		Informaticien	1100000	10
66	MBALLO	Alassane	Colobane	19	01/04/2013		Sécretaire	30000	40
77	DIALLO	Adama	Mermoz	17	06/05/2007	779876543	Sécretaire	300000	60
88	DIOUF	Philippe Jordan	Colobane	45	20/02/2006	784532156	Comptable	700000	30
99	DIOUF	Papa Amadou	Fass	55	20/02/2007		Comptable	300000	10
101	FALL	Ndèye Fatou	Fass	48	11/10/2019		Sécretaire	300000	
202	FAYE	Abdoulaye		30	25/03/2014	786543278		300000	60
303	DIOP	Awa		29	01/12/2019	767891023	Gestionnaire	850000	20
404	HANE	Mahamadou	Fass	49	20/02/2006		Comptable	900000	40
505	LO	Mariama	Mermoz	19	25/04/2014		Sécretaire	300000	50
606	DIOP	Ndèye Awa	Fann Hock	30	06/05/2007		Comptable	900000	50
707	DIOP	Cheikh	Colobane	25	06/05/2007	761234567	Gestionnaire	850000	60

b. Importation des données du fichier vers un dataframe **employe_df**

c. Afficher le format (nombres de ligne et de colonnes) du dataframe importé

d. Exportation des données du dataframe *employe_df* vers la table *Employe* de la base de données *Projet*.

B. Interrogation de données de MySQL à partir Python

B.1. Ecrire et mettre requêtes suivantes à partir de Python.

- a. Afficher les noms, prénoms, âges, professions des employés habitant à Fass.
- b. Afficher les salaires moyens par profession
- c. Afficher les projets où travaillent des informaticiens
- d. Pour chaque projet, afficher le salaire moyen et l'âge moyen de ses employés

B.2. Ecrire et exécuter les programmes Python :

B.1. pour se connecter au serveur MySQL et charger les données des tables *employe* et *projet* dans des dataframes pandas ;

B.2. connaître les formats (nombres de ligne et de colonnes) des dataframes importés.

C. Préparation des données

- a. Gérer les valeurs manquantes des dataframes : remplacement ou suppression,

D. Extraction, filtre, concaténation, merge et combinaison :

Ecrire et exécuter les programmes Python pour créer les dataframes contenant les réponses des requêtes suivantes sur les données importées :

1. Noms et prénom des employés
2. Noms et prénom des employés, ordre croissant selon le nom
3. Noms et prénom des employés de moins 25 ans, ordre croissant selon le nom
4. Nom, budget, date de début et localisation des projets
5. Nom, budget, date de début et localisation des projets de Kaolack
6. Nom, budget, date de début et localisation des projets de Kaolack et Dakar
7. Professions et salaires des employés en commençant par les salaires les plus élevés
8. Noms, prénoms et salaires annuels des employés avec les budgets leurs projets
9. Noms, prénoms et salaires augmentés de 20% des employés
10. Montant total des salaires de tous les employés
11. Budget moyen des projets
12. Salaires le plus petit et le plus élevé
13. Salaire moyen par projet
14. Ages moyens des employés par profession
15. Salaire moyen par projet et par profession des projets localisés à Dakar, ou à Kaolack ou à Ziguinchor

E. Sauvegarde sur Excel

Ecrire et exécuter les programmes Python pour sauvegarder les résultats des questions précédentes dans des fichiers Excel et CSV d'un sous dossier *resultats*.

F. Visualisation des données

Ecrire et exécuter les programmes Python pour visualiser les graphiques :

1. Evolution du salaire en fonction de l'âge,

2. Distribution des salaires
3. Budgets moyen, minimal et maximal des projets

A rendre : le 17/01/2024

Rapport répondant à toutes les questions ci-dessus avec :

- Tous les scripts Python (notebooks)
- Fichiers exportés
- Base de données MySQL

Composition des groupes : deux étudiants par groupe