

TD2 – Partie B – Exploration de données

Préparation de données

1. Chargez dans un dataframe le fichier « français_etranger.csv ». (Attention à l'encodage !)

	Pays/année	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
0	AFGHANISTAN	NaN	NaN	NaN	NaN	206	226	189	177	303	319	334	286	241
1	AFRIQUE DU SUD	5 254	5 438	5 958	6 312	6 130	6 560	6 447	6 987	6 998	7 054	7 108	7209	7283
2	ALBANIE	54	55	56	56	79	96	100	90	86	121	172	198	201
3	ALGERIE	7 389	12 541	22 202	30 964	38 080	41 498	36 782	34 718	30 993	28 287	28 830	30344	31677
4	ALLEMAGNE	87 677	96 619	107 774	106 781	101 391	108 821	99 288	106 842	109 468	111 742	114 372	110881	112238

2. Quel était le nombre total de français à l'étranger en 2007 ? (Si vous obtenez systématiquement une erreur, repartez à la question 15 et renseignez le paramètre « thousands » de la fonction « read_csv »)
3. Dans quel pays y'avait-il le plus de français en 2013 ?
4. Combien de français y'avait-il en AUSTRALIE, au BURKINA et au LUXEMBOURG en 2005 et 2009 ?
5. Pour quels pays y'a t-il des valeurs manquantes ?
6. Quels sont les 10 pays où la variation du nombre de français au fil des années est la plus faible ? la plus forte ?
7. Supprimez la ligne « TOTAL » et renommez la colonne « Pays/année » en « pays ».
8. A l'aide de la fonction melt(), restructurez le dataframe en trois colonnes (pays, année, effectif)

	pays	annee	effectif
0	AFGHANISTAN	2001	NaN
1	AFRIQUE DU SUD	2001	5254.0
2	ALBANIE	2001	54.0
3	ALGERIE	2001	7389.0
4	ALLEMAGNE	2001	87677.0

9. Éliminez les lignes contenant des effectifs vides. Combien de lignes ont été supprimées ?
10. Depuis la base de données SQLite « covid19.db », chargez la table « stats_covid » dans un dataframe

11. Supprimez les colonnes de « HospConv », « SSR_USLD » et « autres »
Renommez les colonnes « dep » en « departement », « hosp » en « nombre_hosp », « rea » en « nombre_rea », « rad » en « nombre_gueris », et « dc » en « nombre_décès ».
12. Dans la colonne "sexe", remplacez les valeurs "0" par "hommes" et "1" par "femmes".
Supprimez les lignes où "sexe" = 2
13. Exportez le résultat dans un fichier csv.