

TP (L2) Premiers pas avec pandas

Base : part du revenu du top 1% (France)

Enseignant : Etienne Dagorn

etienne.dagorn@univ-lille.fr

Fichier : `top1_pretax.csv` Colonnes : `Entity, Year, Top_1_pretax` (en %).

Contexte : évolution de la part des 1% les plus riches en France (Our World in Data).

Objectifs : charger un CSV, vérifier la qualité, calculer des indicateurs, créer des variables dérivées, tracer des graphiques, exporter un jeu propre.

A. Démarrage & aperçu des données

Exercice 1 Chargement et premier aperçu

- a) Importer `pandas` et lire le CSV dans `df` (`pd.read_csv`).
- b) Afficher les 5 premières lignes (`df.head()`).
- c) Afficher `df.info()` et `df.columns`.

Exercice 2 Renommage et types

- a) Renommer les colonnes en `entity`, `year`, `top1`.
- b) Convertir `year` en entier et `top1` en flottant.
- c) Vérifier avec `df.dtypes`.

Exercice 3 Filtrage France & tri chronologique

- a) Filtrer `df` sur `entity == "France"` pour obtenir `fr`.
- b) Trier `fr` par `year` croissant et `reset_index()` si besoin.
- c) Afficher `fr["year"].min()` et `fr["year"].max()`.

B. Qualité des données & statistiques de base

Exercice 4 Valeurs manquantes et bornes plausibles

- a) Compter les valeurs manquantes (`fr.isna().sum()`).
- b) Vérifier que `top1` est dans $[0, 100]$; afficher les lignes hors borne si y en a.
- c) vérifier lunicité des années (`fr["year"].duplicated().sum()`).

Exercice 5 Statistiques descriptives de base

- a) Calculer moyenne, médiane, min, max de `top1`.
- b) Retrouver lannée du maximum et du minimum (`idxmax/idxmin` puis `loc`).

C. Variables dérivées & comparaisons

Exercice 6 Indice base 100 (année de référence)

- a) Prendre la première valeur observée de `top1` comme base.
- b) Créer `indice_100 = 100 * top1 / base`.
- c) Vérifier que `indice` vaut 100 à l'année de base.

Exercice 7 Variations : en points vs en pourcentage

- a) Créer `delta_pts = top1.diff()` (variation en points de pourcentage).
- b) Créer `delta_pct = 100 * top1.pct_change()` (variation en %).
- c) Identifier l'année de plus forte hausse/baisse selon chaque mesure.
- d) Interpréter en 2 lignes la différence points de % vs %.

Exercice 8 Détection de trous dans les données

- a) Créer `annee_suiv = year.shift(-1)` et `gap = annee_suiv - year`.
- b) Lister les lignes où `gap > 1` (années manquantes entre deux observations).
- c) *Option A (reco L2)* : garder tel quel et le mentionner dans le commentaire de fin.
Option B (challenge) : réindexer sur toutes les années (`set_index("year")`, `reindex(range(...))`) pour visualiser le trou (valeurs manquantes explicites).

Exercice 9 Lissage (moyenne mobile courte)

- a) Créer `mm3 = top1.rolling(window=3, center=True).mean()`.
- b) Garder `top1` et `mm3` pour le tracé (section Graphiques).
- c) En 12 lignes : que change le lissage dans la lecture ?

Exercice 10 Découpage par périodes (comparaisons)

- a) Construire `période` par `pd.cut` (ex. 18201899 / 19001945 / 19461980 / 19812025).
- b) Calculer la moyenne de `top1` par période (`groupby`).
- c) Interpréter : période la plus haute en concentration ?

D. Graphiques

Exercice 11 Séries temporelles

- a) Tracer la courbe `top1` en fonction de `year` (points + ligne).
- b) Ajouter sur le même graphe la moyenne mobile `mm3`.
- c) Soigner : titre, axes (unité %), grille légère, `plt.tight_layout()`.

Exercice 12 Indice base 100

- a) Tracer `indice_100` en fonction de `year`.
- b) *Option* : ajouter une 2^e indexation avec une autre année de base et comparer.

Exercice 13 Barres des variations

- Tracer un diagramme en barres de `delta_pts` par année.
- Option* : colorier positifs/négatifs (sans complexifier).

Exercice 14 Export des figures (reproductibilité)

- Créer un dossier `out/` (`os.makedirs("out", exist_ok=True)`).
- Enregistrer la série `top1+mm3` (`.png`, `dpi=150`, `bbox_inches="tight"`).
- Enregistrer le graphe de `indice_100` (`.png`, mêmes options).

E. Export de données & miniinterprétation

Exercice 15 Export CSV propre

- Conserver dans `fr` les colonnes utiles : `year`, `top1`, `indice_100`, `delta_pts`, `delta_pct`, `mm3`, `periode`.
- Exporter en CSV sous le nom `out/france_top1_L2.csv`.

Exercice 16 Interprétation (68 lignes)**Guidage :**

- Sur la période observée, la part du top 1% a globalement (augmenté/baissé/stagné) : préciser l'ordre de grandeur (en points).
- Repérer une phase haute et une phase basse (via `periode` ou `mm3`).
- Mentionner si il existe des *trous* dans les données et en quoi cela affecte (ou non) la lecture.
- Donner une piste d'explication économique *plausible* (12 lignes).

Conseils pratiques

- Toujours préciser les unités (%) et l'axe des temps (années).
- Ne pas supprimer silencieusement des valeurs manquantes : les repérer et documenter.
- Donner un nom explicite aux fichiers exportés (`out/`).

Barème indicatif

- Exactitude des calculs et variables dérivées : **40 %**
- Qualité des graphiques (lisibilité, titres, axes, export) : **30 %**
- Interprétation écrite (clarté, justesse, mobilisation des résultats) : **30 %**