

# Cahier des charges – Projet Fil Rouge Data Analyst

**« Analyse comparative des pays pour l'investissement et la qualité de vie »**

**Formation : Data Analyst - Simplon**

**Étudiant : Hajar Mamdouh**

**Formateur : Yassine**

## 1. Contexte

Dans le cadre de la formation **Data Analyst**, les investisseurs, les entreprises et les particuliers cherchent à comparer différents pays pour prendre des décisions d'investissement ou de mobilité internationale. Ces décisions reposent sur des indicateurs économiques, sociaux et de qualité de vie, souvent disponibles mais dispersés sur différentes sources open data.

Ce projet fil rouge vise à concevoir une **solution data complète, automatisée et orientée décisionnel**, permettant de transformer ces données brutes en indicateurs exploitables par des décideurs ou des particuliers.

Le projet respecte le référentiel Data Analyst Simplon et couvre l'ensemble du cycle de vie de la donnée : cadrage du besoin, collecte multi-sources, préparation, analyse, visualisation et restitution.

## 2. Problématique

Comment exploiter efficacement des données économiques, sociales et de qualité de vie afin de comparer objectivement les pays et fournir un support clair pour la prise de décision d'investissement ou de mobilité internationale ?

## 3. Objectifs du projet

### 3.1 Objectifs métier

- Aider les investisseurs à identifier les pays les plus attractifs selon différents indicateurs économiques et sociaux.
- Faciliter la comparaison des pays selon le coût de la vie, le revenu moyen et la stabilité économique.
- Fournir une vision synthétique et exploitable des indicateurs économiques et sociaux.

### 3.2 Objectifs analytiques

- Identifier les pays les plus attractifs grâce à un score global calculé à partir des données open source.
- Mettre en évidence des corrélations entre coût de la vie, revenus et stabilité économique.
- Segmenter les pays en groupes homogènes selon leur attractivité.
- Analyser les tendances économiques sur les dernières années.

### 3.3 Objectifs techniques

- Mettre en place un pipeline de données automatisé (ETL) de bout en bout.
- Structurer les données dans une base relationnelle normalisée.
- Développer un tableau de bord décisionnel interactif (Power BI).
- Assurer la reproductibilité et la documentation complète du projet.
- Mettre en place des tests de qualité des données.

## 4. Périmètre du Projet

### Périmètre Fonctionnel

Le projet inclut :

- Données économiques, sociales et qualité de vie issues de sources open data.
- Analyses descriptives, exploratoires et statistiques.
- Dashboard décisionnel interactif avec filtres par pays, région et indicateurs.
- Documentation technique et fonctionnelle complète.
- Gestion de versions avec Git / GitHub.

### Périmètre Temporel

Les données analysées couvrent une période de 5 à 10 ans selon disponibilité.

## 5. Questions Stratégiques & KPI Associés

- **Quels sont les pays les plus heureux au monde ?**

**KPI associés :**

- Score moyen de bonheur par pays
- Classement mondial par score de bonheur
- Évolution du score par année

### Analyse apportée :

Permet d'identifier les pays les mieux classés et d'observer les tendances globales.

- **Quels facteurs influencent le plus le niveau de bonheur ?**

**KPI associés :**

- Corrélation entre PIB et score de bonheur
- Corrélation entre soutien social et score
- Impact de l'espérance de vie sur le score
- Impact du niveau de corruption perçue

### Analyse apportée :

Permet d'identifier les variables ayant le plus fort impact statistique sur le bonheur.

- Existe-t-il une relation entre richesse économique et bonheur ?

**KPI associés :**

- PIB moyen par pays
- Score de bonheur moyen
- Indice de corrélation PIB / Bonheur
- Analyse comparative par région

**Analyse apportée :**

Permet de vérifier si la croissance économique est systématiquement liée à une amélioration du bien-être.

- Comment évolue le niveau de bonheur dans le temps ?

**KPI associés :**

- Évolution annuelle du score de bonheur
- Variation en pourcentage par année
- Tendance (croissance / stagnation / baisse)

**Analyse apportée :**

Permet d'identifier les tendances positives ou négatives sur plusieurs années.

- Quelles régions du monde présentent les meilleurs niveaux de bien-être ?

**KPI associés :**

- Score moyen par région
- Classement des régions
- Comparaison interrégionale

**Analyse apportée :**

Permet une analyse macroéconomique et géographique.

- Quel est l'impact de la corruption perçue sur le bonheur ?

**KPI associés :**

- Indice de perception de la corruption
- Corrélation corruption / score bonheur
- Analyse comparative pays à faible vs forte corruption

**Analyse apportée :**

Permet de mesurer l'effet de la gouvernance sur le bien-être.

- **Le soutien social influence-t-il significativement le bonheur ?**

**KPI associés :**

- Indice de soutien social
- Corrélation soutien social / score
- Moyenne par région

**Analyse apportée :**

Permet d'évaluer l'importance du capital social.

- **Quels pays ont connu la plus forte progression du niveau de bonheur ?**

**KPI associés :**

- Taux de croissance annuel du score de bonheur
- Variation absolue du score entre deux années
- Classement des pays par évolution positive

**Analyse apportée :**

Permet d'identifier les pays en amélioration rapide et d'analyser les facteurs associés à cette progression.

- **Quels pays présentent une baisse significative du bonheur ?**

**KPI associés :**

- Taux de décroissance du score
- Variation négative annuelle
- Analyse comparative des facteurs explicatifs

**Analyse apportée :**

Permet d'identifier les zones en difficulté socio-économique.

- **Existe-t-il un seuil minimal de PIB à partir duquel le bonheur augmente significativement ?**

**KPI associés :**

- Distribution du PIB par tranche
- Moyenne du score par tranche de PIB
- Analyse de corrélation non linéaire

**Analyse apportée :**

Permet d'analyser la notion de "rendement décroissant" de la richesse.

- Les pays avec une forte espérance de vie sont-ils nécessairement plus heureux ?

**KPI associés :**

- Espérance de vie moyenne
- Corrélation espérance de vie / score bonheur
- Comparaison pays à haute vs basse espérance de vie

**Analyse apportée :**

Permet d'évaluer l'impact des conditions sanitaires sur le bien-être.

- Quel est le facteur le plus déterminant dans le score de bonheur ?

**KPI associés :**

- Importance relative des variables (analyse statistique)
- Poids de chaque indicateur dans le modèle
- Analyse comparative des coefficients

**Analyse apportée :**

Permet d'identifier le levier principal d'amélioration du bien-être.

- Existe-t-il des disparités importantes entre régions ?

**KPI associés :**

- Écart-type du score par région
- Indice d'inégalité régionale
- Score moyen régional

**Analyse apportée :**

Permet d'identifier les déséquilibres géographiques.

- Quels pays combinent forte croissance économique et haut niveau de bonheur ?

**KPI associés :**

- PIB élevé + Score élevé
- Classement croisé PIB / Bonheur
- Matrice de performance économique et sociale

**Analyse apportée :**

Permet d'identifier les modèles de réussite économique durable.

- Le niveau de liberté individuelle influence-t-il le bonheur ?

**KPI associés :**

- Indice de liberté
- Corrélation liberté / score
- Comparaison interrégionale

**Analyse apportée :**

Permet d'évaluer l'impact des libertés civiles sur le bien-être.

- Les pays à faible corruption présentent-ils un meilleur score de bonheur ?

**KPI associés :**

- Indice de corruption
- Corrélation corruption / score
- Classement des pays les moins corrompus

**Analyse apportée :**

Permet d'analyser l'impact de la gouvernance sur la satisfaction de vie.

## 6. Sources de données

Le projet repose sur des sources de données différentes

### Sources principales

1. **World Happiness Report** – CSV/API
  - Données de satisfaction de vie et qualité de vie par pays.
  - Lien : <https://worldhappiness.report>
2. **Banque Mondiale – World Bank Open Data** – API / CSV
  - Indicateurs économiques : PIB, inflation, revenu par habitant, population.
  - Lien : <https://data.worldbank.org/>
3. **Numbeo – Cost of Living** – API ou CSV
  - Indices de coût de la vie et pouvoir d'achat par pays.
  - Lien : <https://www.numbeo.com/cost-of-living/>

### Caractéristiques des données

- Volume : 200-250 pays
- Période : 2015-2025 (selon disponibilité)
- Types d'indicateurs : économiques, sociaux, qualité de vie, stabilité économique

## 7. Qualité des Données

La gestion de la qualité repose sur :

- La détection des valeurs manquantes
- Le traitement des doublons
- L'harmonisation des formats
- La validation des types de données
- Le contrôle des incohérences

## 8. Indicateurs Clés (KPI)

### ◆ PIB par habitant

Mesure la richesse moyenne produite par personne dans un pays.

→ Permet d'évaluer le niveau de développement économique.

### ◆ Inflation annuelle

Indique l'augmentation générale des prix sur une année.

→ Permet d'analyser la stabilité économique et le pouvoir d'achat.

### ◆ Croissance économique

Taux d'évolution annuel du PIB.

→ Permet d'identifier les pays en expansion ou en ralentissement.

### ◆ Revenu National Brut (RNB)

Revenus totaux perçus par les résidents d'un pays.

→ Permet d'évaluer le niveau global de revenus disponibles.

### ◆ Indice de bonheur

Score mesurant le niveau de satisfaction de vie des habitants.

→ Indicateur principal du bien-être global.

### ◆ Coût de la vie

Indice mesurant le niveau général des prix dans un pays.

→ Permet de comparer l'accessibilité financière entre pays.

### ◆ Pouvoir d'achat

Capacité des habitants à consommer en fonction de leurs revenus et des prix.

→ Indique le niveau de vie réel.

### ◆ Score d'attractivité global

Indicateur composite calculé à partir des KPI économiques et sociaux normalisés.

→ Permet de classer les pays selon leur attractivité globale.

## 9. Méthodologie Analytique

### Phase 1 : Cadrage

1. Validation du cahier des charges
2. Identification précise des sources de données

### Phase 2 : Extraction et exploration

1. Extraction automatisée des données (CSV, API)
2. Exploration initiale (EDA)
3. Profiling et identification des problèmes de qualité

### Phase 3 : Préparation

1. Nettoyage et traitement des valeurs manquantes
2. Normalisation et standardisation des données
3. Harmonisation des noms de pays et formats

### Phase 4 : Structuration

1. Conception du modèle relationnel (MCD/MLD / schéma en étoile)
2. Création de la base PostgreSQL / MySQL
3. Chargement et contrôle qualité

### Phase 5 : Analyse

1. Analyses statistiques descriptives
2. Segmentation des pays (clustering)
3. Corrélations entre indicateurs
4. Calcul du score d'attractivité

## Phase 6 : Visualisation

1. Conception du dashboard Power BI
2. Tests utilisateurs et ajustements

## Phase 7 : Documentation et restitution

1. Rédaction du rapport analytique (15-25 pages)
2. Préparation du support de présentation (15-20 slides)

## 10. Contraintes et risques

Risque	Probabilité	Impact	Mitigation
Qualité et hétérogénéité des données	Haute	Moyen	Contrôles qualité et documentation des anomalies
Limites des données open source	Haute	Moyen	Simulation ou compléments de données
Disponibilité des APIs	Moyenne	Faible	Mise en cache et fallback CSV
Complexité technique	Moyenne	Moyen	Documentation et support formateur

## 11. User Stories

**US1 – Investisseur** : Visualiser les pays les plus attractifs selon les KPI et scores globaux.

**US2 – Analyste économique** : Explorer les corrélations entre indicateurs économiques et qualité de vie.

**US3 – Décideur RH / mobilité** : Comparer les pays pour planifier la mobilité des employés.

**US4 – Data Analyst** : Avoir un pipeline automatisé pour mettre à jour facilement les analyses avec de nouvelles données.

## 12. Outils et Technologies

- Python (pandas, numpy, requests, matplotlib, SciPy, Matplotlib, Seaborn)
- SQL (PostgreSQL / MySQL)
- Databricks
- Talend / airflow
- Power BI
- Git / GitHub
- Jira
- Confluence

## 13. Livrables

- Notebooks Jupyter : EDA, analyses statistiques, visualisations exploratoires
- Base de données structurée
- Modèle analytique
- Dashboard interactif
- Rapport final
- Présentation pptx

## 14. Bénéficiaires

- Investisseurs
- Entreprises
- Services RH
- Analystes

## 15. Limites

- Données agrégées
- Variabilité méthodologique
- Disponibilité variable selon pays