



Pflichten- und Lastenheft

Alt, arm, abgehängt? Datengetriebene Perspektiven auf Altersarmut und Rentensysteme in Europa

Version 0.1

Autor des Dokuments	Retirement Analytics EU	Erstellt am	21.05.2025
Dateiname	PflichtenUndLastenheft		
Seitenanzahl	13		

Historie der Dokumentversionen

Version	Datum	Autor	Änderungsgrund / Bemerkungen
0.1	22.05.2025	Retirement Analytics EU	Ersterstellung

Inhaltsverzeichnis

1 Einleitung.....	3
1.1 Allgemeines.....	3
2 Konzept und Rahmenbedingungen.....	4
2.1 Ziele des Anbieters.....	4
2.2 Ziele und Nutzen des Anwenders.....	4
2.3 Zielgruppe.....	4
2.4 Systemvoraussetzungen.....	4
2.5 Ressourcen.....	4
3 Beschreibung der Anforderungen.....	5
3.1 Anforderung 1: Datenerhebung und Datenvalidierung.....	5
3.2 Anforderung 2: Analyse der Altersarmutsverteilung.....	5
3.3 Anforderung 3: Visualisierung der Ergebnisse.....	6
3.4 Anforderung 4: Korrelation Rentenausgaben und Altersarmut.....	7
3.5 Anforderung 5: Politische Handlungsempfehlungen.....	7
4 Technische Umsetzung der Anforderungen.....	9
4.1 Anforderung 1: Datenerhebung und Datenvalidierung.....	9
4.2 Anforderung 2: Analyse der Altersarmutsverteilung.....	10
4.3 Anforderung 3: Visualisierung der Ergebnisse.....	11
4.4 Anforderung 4: Korrelation Rentenausgaben und Altersarmut.....	12
4.5 Anforderung 5: Politische Handlungsempfehlungen.....	12

1 Einleitung

1.1 Allgemeines

1.1.1 Zweck und Ziel dieses Dokuments

Dieses Pflichten- und Lastenheft beschreibt die Pflichten und Lasten die im Projekt „Alt, arm, abgehängt? Datengetriebene Perspektiven auf Altersarmut und Rentensysteme in Europa“ zu beachten sind. Das Pflichten- und Lastenheft dient als Grundlage zur Planung und Umsetzung des Projektes.

1.1.2 Projektbezug

Dieses Projekt bezieht sich auf die Analyse der Altersarmut und deren Entwicklung innerhalb der europäischen Ländern mit dem Fokus auf die EU, basierend auf den Daten der EU und der verschiedenen Ländern. Dabei werden verschiedene Aspekte, Kriterien untersucht, sowie die verschiedenen Lösungsansätze

1.1.3 Abkürzungen

API = Application Programming Interface
BIP = Brutto Inlandsprodukt
CSV = Comma-Separated Values
EU = Europäische Union

1.1.4 Ablage und Bezüge zu anderen Dokumenten

Ablage aller Ressourcen, Dokumentation, Präsentationen in Github: [Elpidahope/Projektrealisierung](https://github.com/Elpidahope/Projektrealisierung) / „<https://github.com/Elpidahope/Projektrealisierung>“

Dieses Lasten- und Pflichtenheft hängt mit dem Projektauftrag und dem Projektstrukturplan zusammen.

2 Konzept und Rahmenbedingungen

2.1 Ziele des Anbieters

Das Ziel dieses Projekts ist es die Altersarmut und deren Entwicklung innerhalb Europas zu analysieren. Konkret ist die Analyse von Mustern der Altersarmut in der EU auf Basis offizieller EU-Daten, die Visualisierung und Interpretation von Daten zu Rentenausgaben und Armut, sowie die Entwicklung von datenbasierten Empfehlungen an die EU-Politik das Ziel dieses Projekts.

Insbesondere soll auch untersucht werden, welche Bevölkerungsgruppen und welche Länder stark von Altersarmut betroffen sind und was die Unterschiede der verschiedenen Rentensysteme der Länder sind.

2.2 Ziele und Nutzen des Anwenders

Das Nutzen für den Anwender ist es die Problematik zu verstehen, sowie die verschiedenen Lösungsansätze, des Problems der Altersarmut. Dabei sind vor allem für den Anwender relevant, wie das Problem aufgebaut ist und sich widerspiegelt in den Daten bei den verschiedenen Länder. Ziel ist es auch dem Anwender dieses Projektes die verschiedenen Lösungsansätze und Handlungsoptionen aufzuweisen und wie diese sich auswirken.

2.3 Zielgruppe

Dieses Projekt ist für alle geeignet, die sich mit dem Thema der Altersarmut interessieren, sowie diejenigen die verschiedene Lösungsansätze untersuchen wollen. Dazu gehören:

- Institutionen und Gremien der Europäischen Union
- Sozial- und wirtschaftspolitische Entscheidungsträger
- Wissenschaftliche und zivilgesellschaftliche Akteure im Bereich Armutsforschung
- Allgemein interessierte Personen

2.4 Systemvoraussetzungen

Da es sich um eine analytische Darstellung handelt liegen keine Systemvoraussetzungen vor.

2.5 Ressourcen

Das Projekt bezieht sich hauptsächlich auf die Daten, die von der EU gezogen werden unter „data.europa.eu“ sowie von den verschiedenen staatlichen Behörden, wie zum Beispiel das „Statistische Bundesamt“.

3 Beschreibung der Anforderungen

3.1 Anforderung 1: Datenerhebung und Datenvalidierung

Priorität: Muss

3.1.1 Beschreibung:

Zur Durchführung der datenbasierten Analyse müssen relevante, qualitätsgesicherte Daten zur Altersarmut und zu Rentenausgaben in den EU-Mitgliedstaaten erhoben werden. Die Daten sind über vertrauenswürdige Quellen wie Eurostat und data.europa.eu zu beziehen. Mindestens folgende Indikatoren müssen abgedeckt werden: Armutsgefährdungsquote älterer Menschen (ab 65 Jahren), öffentliche Rentenausgaben in % des BIP sowie soziodemografische Merkmale (z. B. Geschlecht, Bildungsniveau, Haushaltsform). Es sind Daten für mindestens 20 EU-Staaten über einen Zeitraum von 5 Jahren zu erfassen und auf Vollständigkeit sowie Konsistenz zu prüfen.

3.1.2 Wechselwirkungen:

Diese Anforderung bildet die Grundlage für alle weiteren Analyseschritte. Fehlerhafte, unvollständige oder nicht vergleichbare Daten würden die Aussagekraft der Analysen erheblich einschränken.

3.1.3 Risiken:

- Uneinheitliche Datenformate oder Definitionsunterschiede
- Fehlende Werte für bestimmte Jahre oder Staaten
- Zeitliche Verzögerungen in der Datenveröffentlichung

3.1.4 Vergleich mit bestehenden Lösungen:

In wissenschaftlichen Studien werden üblicherweise ebenfalls Daten von Eurostat verwendet. Das Vorgehen entspricht dem aktuellen methodischen Standard der empirischen Sozialforschung.

3.1.5 Grobschätzung des Aufwands:

- Recherche und Download: 8 Stunden
- Sichtung und Validierung: 6 Stunden
- Dokumentation der Quellen und Definitionen: 4 Stunden

3.2 Anforderung 2: Analyse der Altersarmutsverteilung

Priorität: Muss

3.2.1 Beschreibung:

Die erhobenen Daten sollen hinsichtlich ihrer Verteilung analysiert werden, um besonders betroffene Staaten und Gruppen zu identifizieren. Berücksichtigt werden:

- Länder mit hoher bzw. niedriger Altersarmutsquote
- Entwicklungen über einen Zeitraum von fünf Jahren
- Subgruppen (z. B. Geschlecht, Bildung, Haushaltsform)

3.2.2 Wechselwirkungen:

Aufbauend auf Anforderung 1. Die Ergebnisse fließen in Visualisierungen und politische Empfehlungen ein.

3.2.3 Risiken:

- Unvollständige Subgruppendaten
- Verzerrungen durch Erhebungsmethodik

3.2.4 Vergleich mit bestehenden Lösungen:

Die geplante Detailtiefe übertrifft viele bestehende Studien, die meist aggregierte Werte betrachten.

3.2.5 Grobschätzung des Aufwands:

- Datenanalyse gesamt: 10 Stunden
- Subgruppenvergleich: 8 Stunden

3.3 Anforderung 3: Visualisierung der Ergebnisse

Priorität: Muss

3.3.1 Beschreibung:

Die Analyseergebnisse sind anschaulich und zielgruppengerecht aufzubereiten. Geplant sind:

- Heatmaps für Ländervergleiche
- Zeitreihen für Armutsentwicklungen
- Scatterplots zur Darstellung von Zusammenhängen

Die Grafiken sollen exportierbar und barrierearm gestaltet sein.

3.3.2 Wechselwirkungen:

Abhängig von Anforderung 2. Die Visualisierung ist Voraussetzung für das Data Storytelling.

3.3.3 Risiken:

- Missverständliche Darstellungen
- Geringe Datenqualität beeinträchtigt Visualisierungen

3.3.4 Vergleich mit bestehenden Lösungen:

Im Vergleich zu öffentlichen Berichten soll ein höherer Fokus auf analytische Genauigkeit und optische Klarheit gelegt werden.

3.3.5 Grobschätzung des Aufwands:

- Design & Erstellung: 8 Stunden
- Export & Beschriftung: 4 Stunden

3.4 Anforderung 4: Korrelation Rentenausgaben und Altersarmut

Priorität: Muss

3.4.1 Beschreibung:

Es ist zu untersuchen, ob ein statistischer Zusammenhang zwischen öffentlichen Rentenausgaben (% BIP) und der Altersarmutsquote besteht. Hierzu werden:

- Korrelationen (z. B. Pearson) berechnet
- Regressionsanalysen durchgeführt

3.4.2 Wechselwirkungen:

Verknüpft Datensätze aus Anforderung 1 und Analyseergebnisse aus Anforderung 2.

3.4.3 Risiken:

- Verzerrungen durch Drittvariablen
- Unvollständige Finanzdaten

3.4.4 Vergleich mit bestehenden Lösungen:

Bestehende Studien beleuchten häufig nur qualitative Zusammenhänge. Das Projekt bietet quantitative Belege.

3.4.5 Grobschätzung des Aufwands:

- Berechnung: 6 Stunden
- Interpretation: 4 Stunden

3.5 Anforderung 5: Politische Handlungsempfehlungen

Priorität: Muss

3.5.1 Beschreibung:

Basierend auf den Analysen sind datenbasierte Empfehlungen zu formulieren. Ziel ist die Unterstützung politischer Entscheidungsprozesse durch:

- Identifikation sozialpolitischer Stellschrauben
- Praxisnahe Vorschläge für Monitoring- oder Ausgleichsmechanismen

3.5.2 Wechselwirkungen:

Direkt abhängig von den Ergebnissen aller vorangegangenen Anforderungen.

3.5.3 Risiken:

- Fehlende Datenbasis zur Generalisierung
- Politisch sensible Interpretationen

3.5.4 Vergleich mit bestehenden Lösungen:

EU-Strategien enthalten oft allgemeine Ziele. Dieses Projekt bietet konkrete, datenfundierte Ansätze.

3.5.5 Grobschätzung des Aufwands:

- Auswertung & Ableitung: 5 Stunden
- Redaktionelle Formulierung: 3 Stunden

4 Technische Umsetzung der Anforderungen

4.1 Anforderung 1: Datenerhebung und Datenvalidierung

Priorität: Muss

4.1.1 Technische Lösung

Die Datenerhebung erfolgt über data.europa.eu und Eurostat. Python-Skripte mit der Bibliothek Pandas werden entwickelt, um Daten im CSV- oder Excel-Format über APIs oder manuelle Downloads zu beziehen. Die Validierung umfasst:

- Download: Automatisierter Import von Datensätzen (z. B. Armutsgefährdungsquote ab 65 Jahren, Rentenausgaben in % des BIP, soziodemografische Merkmale) für mindestens 20 EU-Staaten über 5 Jahre (2018–2022).
- Konsistenzprüfung: Überprüfung auf einheitliche Formate (z. B. Währung, Zeiteinheiten) mit Pandas.isna() und Vergleich mit Eurostat-Metadaten.
- Fehlwertbehandlung: Imputation durch Mittelwerte oder Interpolation, falls methodisch vertretbar.
- Dokumentation: Erstellung eines Datenwörterbuchs in CSV-Format, das Quellen und Anpassungen dokumentiert.

4.1.2 Werkzeuge und Technologien

- Programmiersprache: Python 3.8+
- Bibliotheken: Pandas für Datenmanipulation und Strukturierung, Requests (für API-Zugriffe), Openpyxl (für Excel-Dateien).
- Speicherung: Lokale Verzeichnisse mit Backup auf GitHub.

4.1.3 Wechselwirkungen

Diese Anforderung ist die Grundlage für alle weiteren Schritte (Analyse, Visualisierung, Empfehlungen). Fehlerhafte Daten würden die Aussagekraft der Ergebnisse beeinträchtigen.

4.1.4 Risiken

- Uneinheitliche Datenformate: Lösung durch Transformation in ein Standardformat (CSV).
- Fehlende Daten: Fallback auf ältere Jahre oder andere Indikatoren.
- API-Probleme: Manuelle Downloads als Alternative.

4.1.5 Vergleich mit bestehenden Lösungen

Die automatisierte Verarbeitung mit Python ist effizienter als manuelle Excel-Analysen, die in vielen Studien üblich sind, und ermöglicht reproduzierbare Ergebnisse.

4.1.6 Grobschätzung des Aufwands

- Skriptentwicklung: 6 Stunden
- Datenvalidierung: 8 Stunden
- Dokumentation: 4 Stunden

- Gesamt: 18 Stunden

4.1.7 Teststrategie

- Unit-Tests für Download-Funktionen (z. B. pytest).
- Manuelle Stichprobenprüfung für 5 Länder und 2 Jahre.
- Plausibilitätsprüfung durch Vergleich mit Eurostat-Berichten.

4.2 Anforderung 2: Analyse der Altersarmutsverteilung

Priorität: Muss

4.2.1 Technische Lösung

Die Analyse wird mit Pandas, NumPy und SciPy durchgeführt:

- Deskriptive Statistik: Berechnung von Mittelwert, Median, Standardabweichung und Extremwerten mit `Pandas.describe()`.
- Zeitreihenanalyse: Glättung von Trends mit gleitendem Durchschnitt (`Pandas.rolling()`).
- Subgruppenanalyse: Aggregation nach Geschlecht, Bildung und Haushaltsform mit `Pandas.groupby()`.
- Clustering: Anwendung von k-Means (scikit-learn) zur Identifikation von Ländern mit ähnlichen Armutsmustern.

4.2.2 Werkzeuge und Technologien

- Bibliotheken: Pandas, NumPy, SciPy, scikit-learn.
- Entwicklungsumgebung: Jupyter Notebook für interaktive Analyse.

4.2.3 Wechselwirkungen

Basiert auf Anforderung 1; Ergebnisse fließen in Visualisierungen (3.3) und Empfehlungen (3.5) ein.

4.2.4 Risiken

- Unvollständige Subgruppen-Daten: Reduktion auf verfügbare Kategorien.
- Verzerrungen durch Erhebungsmethodik: Validierung mit Metadaten.
- Hoher Rechenaufwand: Optimierung durch Teil-Datensätze.

4.2.5 Vergleich mit bestehenden Lösungen

Die Kombination aus deskriptiver Statistik und Clustering bietet detailliertere Einblicke als einfache aggregierte Analysen in EU-Berichten.

4.2.6 Grobschätzung des Aufwands

- Deskriptive Analyse: 6 Stunden
- Zeitreihen- und Subgruppenanalyse: 8 Stunden
- Clustering: 6 Stunden
- Gesamt: 20 Stunden

4.2.7 Teststrategie

- Vergleich der Ergebnisse mit Eurostat-Berichten.
- Plausibilitätsprüfung (z. B. keine negativen Quoten).
- Kreuzvalidierung von Clustering-Ergebnissen.

4.3 Anforderung 3: Visualisierung der Ergebnisse

Priorität: Muss

4.3.1 Technische Lösung

Visualisierungen werden mit Matplotlib, Seaborn und Plotly erstellt:

- Heatmaps: Darstellung der Armutsquoten nach Ländern (Seaborn.heatmap()).
- Zeitreihen: Liniendiagramme für Trends (Matplotlib.plot()).
- Scatterplots: Zusammenhänge zwischen Rentenausgaben und Armutsquoten (Plotly.express.scatter()).
- Barrierefreiheit: Kontrastreiche Farben (WCAG 2.1), Schriftgröße ≥ 12 pt.

4.3.2 Werkzeuge und Technologien

- Bibliotheken: Matplotlib, Seaborn, Plotly.
- Design: Canva für Feinabstimmung.

4.3.3 Wechselwirkungen

Abhängig von Anforderung 2; Voraussetzung für Data Story (3.5).

4.3.4 Risiken

- Missverständliche Darstellungen: Vereinfachung durch Reduktion der Datenpunkte.
- Export-Probleme: Test in mehreren Formaten (PNG, JPEG).
- Inkonsistentes Design: Einheitliches Farbschema (EU-Blau/Gelb).

4.3.5 Vergleich mit bestehenden Lösungen

Die Verwendung interaktiver Plotly-Grafiken hebt die Visualisierungen von statischen Berichten ab.

4.3.6 Grobschätzung des Aufwands

- Erstellung in Python: 6 Stunden
- Design-Optimierung: 4 Stunden
- Export und Barrierefreiheit: 3 Stunden
- Gesamt: 13 Stunden

4.3.7 Teststrategie

- Überprüfung der Lesbarkeit auf verschiedenen Geräten.
- Testexport in PNG/JPEG.

- Peer-Review auf Design-Klarheit.

4.4 Anforderung 4: Korrelation Rentenausgaben und Altersarmut

Priorität: Muss

4.4.1 Technische Lösung

- Korrelation: Pearson-Koeffizient mit `scipy.stats.pearsonr()`.
- Regression: Lineare Regression mit `statsmodels.OLS()`.
- Robustheit: Prüfung auf Drittvariablen (z. B. Bildung) mit multipler Regression.

4.4.2 Werkzeuge und Technologien

- Bibliotheken: SciPy, Statsmodels.
- Daten: Kombinierte Datensätze aus Anforderung 1.

4.4.3 Wechselwirkungen

Verknüpft Anforderungen 1 und 2; Ergebnisse fließen in Anforderung 5 ein.

4.4.4 Risiken

- Drittvariablen: Berücksichtigung durch multiple Regression.
- Fehlende Daten: Fallback auf aggregierte Werte.
- Überinterpretation: Fokus auf statistische Signifikanz ($p < 0.05$).

4.4.5 Vergleich mit bestehenden Lösungen

Die quantitative Analyse mit Regressionsmodellen ist präziser als qualitative Studien.

4.4.6 Grobschätzung des Aufwands

- Korrelation und Regression: 5 Stunden
- Validierung: 5 Stunden
- Visualisierung: 2 Stunden
- Gesamt: 12 Stunden

4.4.7 Teststrategie

- Kreuzvalidierung mit Teil-Datensätzen.
- Signifikanzprüfung ($p < 0.05$).
- Vergleich mit Literaturwerten.

4.5 Anforderung 5: Politische Handlungsempfehlungen

Priorität: Muss

4.5.1 Technische Lösung

- Ableitung: Synthese der Ergebnisse in einem Markdown-Dokument.

- Formulierung: Strukturierte Empfehlungen mit Verweis auf statistische Ergebnisse.
- Integration: Einbindung von Visualisierungen zur Untermauerung.

4.5.2 Werkzeuge und Technologien

- Text: Markdown, konvertiert zu PDF.
- Visualisierung: Integration von Grafiken aus Anforderung 3.

4.5.3 Wechselwirkungen

Abhängig von allen vorherigen Anforderungen; zentral für Data Story und Video.

4.5.4 Risiken

- Fehlende Generalisierbarkeit: Fokus auf spezifische Länder/Gruppen.
- Sensible Formulierungen: Team-Review zur Neutralität.

4.5.5 Vergleich mit bestehenden Lösungen

Konkrete, datenbasierte Vorschläge heben sich von allgemeinen EU-Strategien ab.

4.5.6 Grobschätzung des Aufwands

- Auswertung: 4 Stunden
- Formulierung: 4 Stunden
- Review: 2 Stunden
- Gesamt: 10 Stunden

4.5.7 Teststrategie

- Peer-Review auf Logik und Verständlichkeit.
- Abgleich mit Projektzielen.