**Executive Summary – World Bank Datenanalyse**

Diese Executive Summary bietet einen Überblick über die Datenanalyse der offen zugänglichen Weltentwicklungsdatenbank der World Bank Group. Der für diese Analyse zur Verfügung stehende Auszug beinhaltet gesellschaftliche und sozioökonomische Daten im Rahmen von jeweils 18 Indikatoren pro Land für 25 Länder der Welt und umfasst eine jährliche Datenerhebung von 2000 bis 2021.

**Fragestellungen**

Das Ziel des Projekts ist, Zusammenhänge zwischen den folgenden Indikatoren zu untersuchen: (i) die HIV-Prävalenz, der Alkoholkonsum pro Kopf und welche Bedeutung dem Anteil der Erwerbsbevölkerung mit Grundbildung auf die HIV-Prävalenz zufällt, (ii) der Anteil der Erwerbsbevölkerung mit Grundbildung, die Zentralstaatsverschuldung und die Schüler-Lehrer-Relation, (iii) das BIP pro Kopf und die Tabakkonsum-Prävalenz sowie (iv) der Anteil an landwirtschaftlicher Nutzfläche und die CO2 Emissionen pro Kopf und welche Rolle die Landesfläche in dieser Beziehung spielt.

**Problematik und Methodik**

Die Längsschnittdaten sehen sich mit Herausforderungen wie Datenlücken konfrontiert. Außerdem liegen für die CO2 Emissionen bei rund zehn Prozent der vorliegenden Daten fehlerhafte Werte vor, die um das Zehn- oder Hundertfache niedriger ausfallen. Nach Rücksprache wurden diese zwar in die Analyse mitaufgenommen, allerdings als Fehler berücksichtigt. Die Zusammenhänge zwischen den einzelnen Indikatoren werden mit einer Vielzahl von Visualisierungen (z.B. Säulen-, Linien-, Punkt- und Boxdiagramme) ausgewertet. Aufgrund der vielfältigen Charakteristika und geringen Vergleichbarkeit der untersuchten Länder wird im Laufe der Analyse innerhalb der Fragestellungen zudem eine Kategorisierung der Länder in äquidistante Quantile nach thematisch gewählten Indikatoren unternommen, um aussagekräftigere Datengruppen zu generieren - je nach vorliegender Datenmenge in drei oder fünf Gruppen.

**Ergebnisse**

Für die Betrachtung der Beziehungen zwischen den einzelnen Indikatoren unter ungruppierten Umständen konnten für alle jeweils tendenziell positive Zusammenhänge im schwachen bis mittelstarken Intensitätsbereich festgestellt werden.

#? Diese zusätzlichen Anmerkungen, haben sich durch die Gruppierung ergeben:)

Zu (i): Mit Gruppierung der Daten ist erkennbar, dass die HIV-Werte für Länder mit sehr großem relativem Alkoholkonsum pro Kopf eine viel größere Streuung aufweisen, aber nicht unbedingt höhere HIV-Prävalenzen. (Der positive Zusammenhang von Anteil an Erwerbstätigen mit grundlegender Bildung und der HIV-Prävalenz wird durch die Gruppierung bestätigt.)

Zu (ii):

Zu (iii): Mit der Gruppierung zeigt sich, dass Länder mit sehr geringem relativem Tabakkonsum ein sehr niedriges BIP pro Kopf aufweisen, während sich für andere Kategorien ein negativer Zusammenhang zeigt.

Zu (iv): der Zusammenhang ist auch im gruppierten Zustand für fast alle Ländergrößen wiederzufinden, einzig bei relativer Betrachtung haben die sehr großen Länder eine gegenläufige Entwicklung.

**Ausblick**

Methodische Herausforderungen in der Analyse umfassen Datenlücken in den Längsschnittdaten sowie die Vergleichbarkeit der inkongruenten Länder. Die Ergebnisse betonen die Wichtigkeit von präziseren Datenerfassungsmethoden, um dem geringwertigen Datenaufkommen entgegenzuwirken. Zudem werden für einen zukünftigen Rückschluss aus der Stichprobe auf die restlichen Länder der Welt repräsentativere Untersuchungseinheiten für die weltweite Lage benötigt, um außerhalb eines vakuumierten Umfelds schlussfolgern zu können. Auch Informationen zu weiteren Indikatoren können dabei helfen, gewonnene Erkenntnisse zu bestätigen oder zu widerlegen.