**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**

*ALTERNATIVE 1:*

Das Ziel des Projekts ist, Zusammenhänge und Einflüsse zwischen einer Auswahl an Indikatoren der World Bank Datenbank zu untersuchen. Dies erfolgt unter Betrachtung von mehreren Fragestellungen, kategorisiert in fünf Themenbereiche: (i) der Zugang zur Elektrizität und das nationale Einkommen, (ii) die Bildung, (iii) das menschliche Abwehrschwächevirus („HIV“), (iv) der Tabakkonsum sowie (v) die Landwirtschaft.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*ALTERNATIVE 2:*

Das Ziel des Projekts ist, Zusammenhänge zwischen den folgenden Indikatoren zu untersuchen: (i) der Zugang zu Elektrizität und das nationale Einkommen sowie welche Rolle die Landesfläche und/oder Einwohnerzahl hierbei spielt, (ii) der Anteil der Erwerbsbevölkerung mit Grundausbildung, die zentrale Staatsverschuldung und die Schüler-Lehrer-Relation, (iii) die HIV-Prävalenz, der Alkoholkonsum pro Kopf und welche Bedeutung dem Anteil der Erwerbsbevölkerung mit Grundausbildung auf die HIV-Prävalenz zufällt, (iv) das BIP pro Kopf und die Tabakkonsum-Prävalenz sowie (v) der Anteil an landwirtschaftlicher Nutzfläche und die CO2 Emissionen pro Kopf und welchen Einfluss die allgemeine Landesfläche auf diese Beziehung hat.

**Problematik und Methodik**

Die Längsschnittdaten sehen sich mit Herausforderungen wie Erhebungs-schwierigkeiten und daraus resultierenden arbiträren und strukturellen Datenlücken konfrontiert.

*[PROBLEME DURCH FEHLENDE DATEN UND WIE DAMIT UMGEGANGEN WIRD*

* *ARBITRÄR: FEHLENDE DATEN OHNE ERKENNBARES MUSTER*
* *STRUKTURELL: FEHL. DATEN FÜR ALLE LÄNDER IN GLEICHEN JAHREN]*

Außerdem liegen für die CO2 Emissionen bei rund zehn Prozent der vorliegenden Daten fehlerhafte Werte vor, die um das Zehnfache (bei zwei Einträgen das Hundertfache) kleiner ausfallen als in der Realität. Nach entsprechender Rücksprache werden diese zwar in die Analyse mitaufgenommen, allerdings gleichzeitig als ausreißende Fehler berücksichtigt und damit die Aussagekraft auf die wahrhaftigen Daten beschränkt. Die Zusammenhänge zwischen den einzelnen Indikatoren werden mit einer Vielzahl von unterschiedlichen Visualisierungen (z.B. Punkt-, Linien-, Säulen- und Boxdiagramme) ausgewertet. Aufgrund der mannigfaltigen Charakteristika und geringen Vergleichbarkeit der untersuchten Länder wird hierfür im Laufe der Analyse innerhalb der Fragestellungen zudem eine Kategorisierung der Länder in äquidistante Quantile unternommen, um aussagekräftigere Datengruppen zu generieren - je nach vorliegender Datenmenge in drei oder fünf Gruppen.

*[FEHLT ETWAS ZUR METHODIK, VGL. BEISPIELABGABE (DETAILLIERTER)?]*

**Ergebnisse**

[VGL. BEISPIELABGABE UND FÜGE ERGEBNISSE EIN]

**Ausblick**

Methodische Herausforderungen in der Analyse umfassen Datenlücken in den Längs-schnittdaten sowie die Vergleichbarkeit der inkongruenten Länder.

* BESSERE VERGLEICHBARKEIT INNERHALB DER STICHPROBE
* BESSERE ERHEBUNGSART UM DATENLÜCKEN ZU VERHINDERN

Die Ergebnisse betonen die Wichtigkeit von präziseren Datenerfassungsmethoden. 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.