

# Group Project

## Analytical Modelling

Juan Rebolledo – Sebastian Mendoza

### Milestone 1:

#### Dataset Selection: Access to electricity (% of population) – World Bank

World Bank Global Electrification Database from "Tracking SDG 7: The Energy Progress Report" led jointly by the custodian agencies: the International Energy Agency ( IEA ), the International Renewable Energy Agency ( IRENA ), the United Nations Statistics Division ( UNSD ), the World Bank and the World Health Organization ( WHO ).

#### Milestone 1-a: Information Table

The dataset we chose was created by the World Bank Global Electrification Database. There are several organizations who are responsible for the validity of this data, such as IEA, IRENA, UNSD, the World Bank and WHO. This database has two sheets, the first one has the percentage of each country's population that has access to electricity. The year with the first data recorded is 1990, and the most recent year that was updated is 2020. The second sheet also has Country Code, but with two additional variables, Region, and Income Group. With both sheets, we could analyze if the percentage of Access to Electricity is linked with the Income group, among other factors. The percentages are weighed, and there are multiple sources where the data came from, such as nationally representative household surveys, Demographic and Health Surveys (DHS) and Living Standards Measurement Surveys (LSMS), Multi-Indicator Cluster Surveys (MICS), the World Health Survey (WHS), among other sources. A simple modeling approach was adopted to fill in the missing data points - around 1990, around 2000, and around 2010. It is important to notice that these percentages are therefore estimates.

#### Milestone 1-b: Data Variable Description

##### Sheet 1: Electricity Data By Country

	count	unique		top	freq	mean	std	min	25%	50%	75%	max
Country Name	266	266		Aruba	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Country Code	266	266		ABW	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Indicator Name	266	1	Access to electricity (% of population)	266		NaN	NaN	NaN	NaN	NaN	NaN	NaN
Indicator Code	266	1	EG.ELC.ACCS.ZS	266		NaN	NaN	NaN	NaN	NaN	NaN	NaN
1960	0.0	NaN		NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...		...	...	...	...	...	...	...	...	...
2017	264.0	NaN		NaN	NaN	84.575392	24.692523	4.2	79.149998	99.300003	100.0	100.0
2018	263.0	NaN		NaN	NaN	85.274649	23.806922	6.17317	80.051456	99.52121	100.0	100.0
2019	264.0	NaN		NaN	NaN	85.804919	23.606509	6.707007	81.66817	99.704998	100.0	100.0
2020	263.0	NaN		NaN	NaN	86.41162	23.050572	7.241338	82.81064	99.989196	100.0	100.0
2021	0.0	NaN		NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

66 rows x 11 columns

Sheet 2: Country by Region and Income group

	count	unique		top	freq
Country Code	265	265		ABW	1
Region	217	7		Europe & Central Asia	58
IncomeGroup	216	4		High income	80
SpecialNotes	126	111	The reporting period for national accounts dat...		7
TableName	265	265		Aruba	1