

Documentation for 11 Variables Data Cleaning and updation

Process Followed.....	2
Up to 2019	2
Adding 2020 data and updation	2
Work Done.....	3
Up to 2019	3
Adding 2020 data and updation	4
Common Mistakes Found.....	5
Data Collection	5

Process Followed

Up to 2019

Step1: Downloaded 2014-2019 data from Nordic compass website. ([Data sheet](#))

Step 2: Subsetting the 11 variables data from main data.

Step 3: Checked every data point that has dropped either by 30% or increased by 65% compared to previous year using online annual reports and sustainability reports. ([raw data sheet](#))

Step 4: Merged it with original data using R. ([Code here](#))

Step 5: Output is a file with original data and updated data side by side along with remaining variables. ([output file](#))

Adding 2020 data and updation

Step1: Downloaded 2014-2019 data from Nordic compass website. ([Data sheet](#))

Step 2: Downloaded the 2020 data and 2018 supplemental data from mail by Zihao. ([File 1](#), [File 2](#))

Step 3: Downloaded GvKeys from WRDS ([GVkey Data file](#))

Step 4: Subsetting 11 variables to check the data after merging in excel.

Step 5: Checked every data point that has dropped either by 30% or increased by 65% compared to previous year using online annual reports and sustainability reports. ([raw data sheet](#))

Step 6: Merged 2019 with 2020 files again using R to do data cleaning. ([Code here](#))

Step 7: Merged existing data with manually updated data. ([output file](#))

Work Done

Up to 2019

Variables	Corrections
sales	289
energy_consump	172
water_withdraw	126
ghg_emis	215
transport_emis	81
num_employees	38
tot_sal_exp	145
female_emp	59
female_exec	17
board_size	19
female_board	35
Total	1196

Adding 2020 data and updation

Variables	Changes
sales	45
energy_consump	54
water_withdraw	39
ghg_emis_updated	96
transport_emis	66
num_employees	32
tot_sal_exp	64
female_emp	33
female_exec	30
board_size	8
female_board	43
Total	510

Common Mistakes Found

Data Collection

Variables	Common Mistakes
Sales	<ol style="list-style-type: none"> 1. Identifying if it's million or thousands in currency 2. Companies like Investment groups having a portfolio whose revenue is picked up sometimes 3. Only segment or quarter figure is picked up.
Energy consumption	<ol style="list-style-type: none"> 1. Conversion of units 2. Change of data from year on year in the financial statements
Water_withdraw	
Ghg_emis	
Transport_emis	
Num_employee	<ol style="list-style-type: none"> 1. No clear data availability 2. Mostly average number reported than as year end.
Total_sal_exp	<ol style="list-style-type: none"> 1. No clear data availability
Female_emp	<ol style="list-style-type: none"> 1. No clear data availability 2. Mostly average number reported than as year-end.
Female_exec	One tip could be to check the picture page which is right after board of directors pictures which gives easy answers than looking into text which is usually confusing
Board size	One tip could be to check the picture page which has board of directors pictures which gives easy answers than looking into text which is usually confusing. This page is usually found in board of director's report
Female_board	