BIODIVERSITY

TECHNICAL REPORT

TEAM 1 :

AMIR, ALIYAH, ZEYNAB, KARIM, NAIF, CATALIN

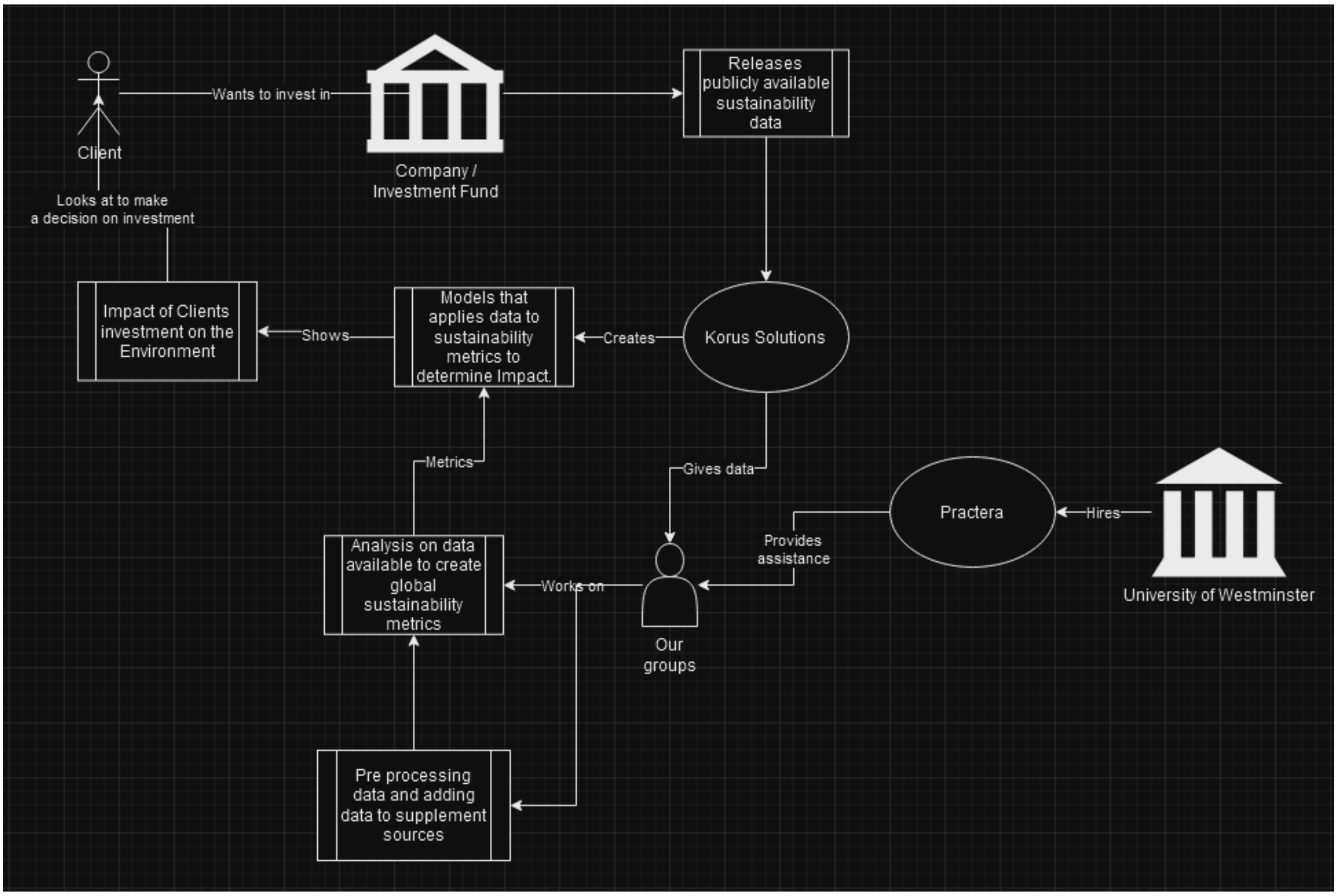
PROBLEM STRUCTURING METHOD

In order to analyse and formulate our problem, we turned to Soft Systems Methodology (SSM). This problem-structuring method allows us to work on and solve the various issues related to working on a technical project.

- **Evidence Gathering:**

No interviews were held, but we spoke to the stakeholders (Korus, Practera, Module Leader) to understand the system operation our team is adding on. Plus we researched and reviewed various literature on the background of our topic (Biodiversity) and on how our metrics relate to it.

- **Rich Picture:**



- Root Definitions:

CATWOE	Definitions
Client	Korus Solutions
Actor	Our Team

Transformation	Various publicly available datasets -> Global Sustainability Metric (related to our topic)
World View	Metrics given serve has correct jump off point to create Global Metrics
Owner	Practera, Module Leader
Environment	University

METHODOLOGY

1. We needed to understand our datasets, therefore we started off with understanding the background of each dataset. Below shows you the information we have gathered about the datasets that we had received from the client.
2. Below shows you the background information of the datasets that we received from Koru and the external dataset that we found.

OVERVIEW OF THE DATA

AGRICULTURAL LAND DATASET	
BACKGROUND INFORMATION	According to The World Bank, “the agricultural land indicator along with land use indicators can also elucidate the environmental sustainability of countries’ agricultural practices”. These variables together show how biodiversity could be affected, since if agricultural practices aren’t sustainable then this could lead to more land having to be used for agricultural purposes, and biodiversity being reduced since this land would then be used to grow a large amount of a few crops vs. the variety that could have been produced otherwise. “50% of all species in the EU rely upon agricultural habitats”, therefore agriculture has a large impact on biodiversity (at least in Europe) and is beneficial to maintaining biodiversity. The reverse is also true as, “the production of food and fibre depend upon a variety of genetic resources and the services they provide”. (European Commission)

<table>
<tr><td>ABOUT THE VARIABLE</td><td>The "Agricultural Land" dataset includes information on the amount of land area used for agriculture in different countries during a period, ranging from 1961 to 2023. A country is represented by each row, which has columns for the name of the series ("Agricultural land (% of land area)"), the country code, and annual data points.</td></tr>
<tr><td>POSSIBLE BIASES & LIMITATIONS</td><td>The “Agricultural Land” dataset comes from The World Bank, and covers 192 countries, only excluding Monaco (since there is no data for it). Possible biases include the FAO not being completely able to impose standardised reporting methods and definitions, this is shown through the varying degrees of accuracy present throughout the dataset. (Source: The World Bank) Another limitation is the fact that 31 of the 193 countries do not have data provided until at least 1990.

Some countries have repeated values over years to very extreme degrees of accuracy, this could be because these values weren’t remeasured or there was no change.</td></tr>
<tr><td>OBSERVATIONS, ROWS AND COLUMNS</td><td><table>
<tr><td>No. of Observations (including null values)</td><td>15,296 (i only included actual measurements and sums including those measurements)</td></tr>
<tr><td>No. of (non-null) Observations</td><td>13,524</td></tr>
<tr><td>No. of Rows (not including headers)</td><td>240</td></tr>
<tr><td>No. of Columns</td><td>68</td></tr>
</table></td></tr>
</table>

C02 EMISSIONS DATASET

BACKGROUND INFORMATION	CO_2 is widely known to affect climate change, which then affects biodiversity since “the response of some species to climate change may constitute an indirect impact on the species that depend on them” (Bellard et al., 2012).
ABOUT THE VARIABLE	The "Deforestation-CO2-Trade-by-Product" dataset provides data on the carbon dioxide (CO2) emissions due to deforestation in 2013 across several agricultural product production countries. Cattle meat, oilseed goods, rice, fruits, vegetables, nuts, wheat, sugar, plant fibres, other meat products, and other cereals are among the items. The values, indicating the environmental effect of creating these goods, are expressed in CO2 per tonne.
POSSIBLE BIASES & LIMITATIONS	The “Deforestation-CO2-Trade-by-Product” dataset only represents 44 of the 194 countries on Earth and only one of these countries is in Africa (South Africa) and only one from South America as well (Brazil). The Middle East is a notable exclusion as well. There is no reason provided for why these countries were chosen, so we cannot make judgements on whether these were the right choices, only that these choices are not representative of each part of the world. There are varying degrees of accuracy which suggests a lack of standardisation when it comes to testing and reporting, even within the same countries across different sectors. Possible bias: Methane has more of an impact on the environment than CO2 so the measurement of CO2 only for cattle makes it seem as though there is less impact than there is in a way? “Methane emissions are the primary factor influencing the carbon footprint of cattle. They have been reported to account for 55% to 92% of the carbon footprint, with the majority of the CH4 emissions arising from enteric fermentation.” (Desjardins et al., 2012) Possible Bias/Limitation: There is no indicated reasons for

	the chosen agricultural products.
OBSERVATIONS, ROWS AND COLUMNS	<table><tr><td>No. of Observations (including null values)</td><td>572 (all observations)</td></tr><tr><td>No. of (non-null) Observations</td><td>528</td></tr><tr><td>No. of Rows</td><td>44</td></tr><tr><td>No. of Columns</td><td>13</td></tr></table>

FOREST AREA DATASET	
BACKGROUND INFORMATION	This factor is important to biodiversity, since "forests contain 60,000 different tree species, 80 percent of amphibian species, 75 percent of bird species, and 68 percent of the world's mammal species" (United Nations Environmental Program, 2020). Reducing that forest space will shrink the area in which these species can live, which could lead to more competition for resources and food chain disruption. Thereby leading to loss of species through no longer having what is needed to survive (such as space, food etc.) and therefore loss of biodiversity.
ABOUT THE VARIABLE	The "Forest Area" dataset shows the percentage of land that is covered by forests in different nations between 1960 and 2021. With columns for the nation name, nation code, series name ("Forest area (% of land area)"), and yearly information points, each row belongs to a distinct country.
POSSIBLE BIASES & LIMITATIONS	The "Forest Area" dataset also covers every country, though most countries have no data given from 1960-1989, though about 40 countries are missing data after that, though mostly

	from 1990-1991.
OBSERVATIONS, ROWS AND COLUMNS	*No. of Observations (including null values)*: 16,009 *No. of (non-null) Observations*: 8,458 *No. of Rows*: 239 *No. of Columns*: 67

No. of Observations (including null values)	16,009
No. of (non-null) Observations	8,458
No. of Rows	239
No. of Columns	67

GHG EMISSIONS DATASET

BACKGROUND INFORMATION	Greenhouse gases make climate change worse and by doing so, reduce biodiversity since reduced habitat makes it difficult for species to sustain themselves due to increased competition, among other factors. Climate change (which greenhouse gas emissions are known to exacerbate) is one of three parts of the triple planetary crisis, alongside biodiversity loss and air pollution. These parts are "interlinked" as climate change can affect biodiversity through desertification and a decline in biodiversity. (United Nations Climate Change, 2022)
ABOUT THE VARIABLE	This dataset gives details on the greenhouse gas (GHG) emissions, as of 2010, related to the production of one kilogram of different food products. Products including apples, bananas, barley, and several breeds of beef are included in the dataset, and their CO_2 equivalent emissions represent the environmental impact of their production.
POSSIBLE BIASES & LIMITATIONS	This dataset chooses these "40 products to represent ~90%

	of global protein and calorie consumption" (Poore & Nemecek, 2018). In the dataset we received, there are 38 products instead of 40. These choices could reduce biases by being representative of the overwhelming majority of consumption and therefore including the most impactful data. A limitation is the fact that unlike a lot of the other datasets given, the dataset only covers a single year. This means there's nothing to compare it to and the year 2010 could have been an outlier.
OBSERVATIONS, ROWS AND COLUMNS	<table><tr><td>*No. of Observations (including null values)*</td><td>*114*</td></tr><tr><td>*No. of (non-null) Observations*</td><td>*76*</td></tr><tr><td>*No. of Rows*</td><td>*38*</td></tr><tr><td>*No. of Columns*</td><td>*4*</td></tr></table>

GLOBAL LIVING PLANET INDEX DATASET	
BACKGROUND INFORMATION	Living Planet Index links to biodiversity through "measuring the changing state of the world's biodiversity over time". (Westveer et al., 2022)
ABOUT THE VARIABLE	Data on the Living Planet Index (LPI) for several regions, from 1970 to the most current year available is provided by the "Global Living Planet Index" dataset. Based on patterns in the vertebrate populations of various species from throughout the world, the LPI is a measure of the condition of biodiversity worldwide. The index's initial value in 1970 is 100, and values after that indicate variations in abundance with respect to this baseline. The dataset additionally shows

	the extent of uncertainty surrounding the index values through upper and lower confidence intervals.
POSSIBLE BIASES & LIMITATIONS	The living planet index is extremely sensitive to outliers since it is an average of all the declines of population size of every species it measures and includes every number without outlier consideration. So if one species had a much larger or smaller decline than all the others, the reported decrease in population size overall would be distorted heavily. (Ritchie, 2022)
OBSERVATIONS, ROWS AND COLUMNS	*No. of Observations (including null values, null values in this case are only where collections of countries do not have Country Codes)*: *1,759* *No. of (non-null) Observations*: *1,467* *No. of Rows*: *343* *No. of Columns*: *6*

No. of Observations (including null values, null values in this case are only where collections of countries do not have Country Codes)	*1,759*
No. of (non-null) Observations	*1,467*
No. of Rows	*343*
No. of Columns	*6*

TREE COVER LOSS DATASET

BACKGROUND INFORMATION	In 2022,” forest loss produced 2.7 gigatonnes (Gt) of carbon dioxide emissions.” CO2 emissions are known to affect the climate and thus biodiversity. The results of reversing Tree Cover Loss are evident. By “restoring only 15 percent of ecosystems in priority areas can cut extinctions by 60 per cent by improving habitats.” (United Nations Environment Program) This will lessen biodiversity loss and the inverse will

<table>
<tr><td></td><td>exacerbate it. In 2022,” forest loss produced 2.7 gigatonnes (Gt) of carbon dioxide emissions.”(Weisse et al.) CO2 emissions are known to affect the climate and thus biodiversity.</td></tr>
<tr><td>ABOUT THE VARIABLE</td><td>The "Tree Cover Loss" dataset: Although the detailed data appears to start from more recent years, includes annual statistics on the loss of tree cover in hectares for several nations starting in 1960. The removal or death of trees "tree cover loss," can be caused by fire, natural catastrophes, deforestation, and other factors.</td></tr>
<tr><td>POSSIBLE BIASES & LIMITATIONS</td><td>This dataset is missing data from Eswatini, Marshall Islands, Samoa, Tonga and Zimbabwe and is missing data from 1960-2000 and 2022-2023 for all countries.</td></tr>
<tr><td>OBSERVATIONS, ROWS AND COLUMNS</td><td><table>
<tr><td>No. of Observations (including null values)</td><td>16,252</td></tr>
<tr><td>No. of (non-null) Observations</td><td>4,716</td></tr>
<tr><td>No. of Rows</td><td>239</td></tr>
<tr><td>No. of Columns</td><td>68</td></tr>
</table></td></tr>
</table>

<table>
<tr><td colspan="2">WHEAT YIELDS DATASET</td></tr>
<tr><td>BACKGROUND INFORMATION</td><td>Wheat yield may affect biodiversity due to more deforestation taking place in places with higher wheat yield where all the species that lived in the forests beforehand are now displaced, leading to loss of biodiversity in the area.</td></tr>
</table>

EXTERNAL DATASET FOUND

RED LIST INDEX DATASET	
BACKGROUND INFORMATION	The Red List Index doesn't have a direct impact on biodiversity on its own since it is only a reflection of trends, however it has been used as an indicator in policies regarding conservation (e.g., UN Sustainability Development Goals) (International Union for Conservation of Nature and

	Natural Resources).
ABOUT THE VARIABLE	The Red List Index is a measure of change of risk of extinction across "different groups of species". The observations are measured between a value of 0 (the species are thought to be extinct) and 1 (the species are thought to be of least concern). (Organisation for Economic Co-Operation and Development)
POSSIBLE BIASES & LIMITATIONS	Includes both upper bound and lower bound of each RLI measurement. Exporting the dataset causes the dataset to be displayed differently from the website in such a way that there were variables left entirely empty (Flag Codes and Flags) and two variables conveying the same piece of information differently (VAR and Variable - Red List). Includes regions that technically aren't even countries (Martinique, Hong Kong, Wallis and Futuna, Bermuda, Guadeloupe etc.).
OBSERVATIONS, ROWS AND COLUMNS	*No. of Observations (including null values)*: 136080 *No. of (non-null) Observations*: 105840 *No. of Rows*: 15120 *No. of Columns*: 9

EXPOSURE TO PM2.5 IN COUNTRIES & REGIONS DATASET	
BACKGROUND	Exposure to PM2.5 can be dangerous to both humans and

<table>
<tr><td>INFORMATION</td><td>animals. In fact, an article explains that howler monkeys affected by PM2.5 pollution from smoke produced by wildfires suffered a “25% lower survival rate due to respiratory issues and impaired immunity, rendering them more susceptible to disease”. The same article says that this type of pollution “is harming Amazonian biodiversity”. (AQI India, 2023)

Exposure to PM2.5 manifests quite similarly in human beings. “Long-term (months to years) exposure to PM2.5 has been linked to premature death, particularly in people who have chronic heart or lung diseases, and reduced lung function growth in children.” (California Air Resources Board)</td></tr>
<tr><td>ABOUT THE VARIABLE</td><td>The dependent variable is “Mean population exposure to PM2.5” measured in micrograms per cubic metre. The observations estimate how much a population is exposed to PM2.5, a group of particulates sized 2.5 micrometres and smaller. This group of particulates are categorised as “fine particulate matter” and are “inhalable”. (California Air Resources Board)</td></tr>
<tr><td>POSSIBLE BIASES & LIMITATIONS</td><td>Every value is estimated. There are two estimates: the “underlying PM2.5 concentration estimates” from the Global Burden of Disease and “concentration estimates… from the Joint Research Centre Global Human Settlement project.”. (Organisation For Economic Co-Operation and Development, 2020)</td></tr>
<tr><td>OBSERVATIONS, ROWS AND COLUMNS</td><td><table>
<tr><td>No. of Observations (including null values)</td><td>93189</td></tr>
<tr><td>No. of (non-null) Observations</td><td>81915</td></tr>
<tr><td>No. of Rows</td><td>5497</td></tr>
<tr><td>No. of Columns</td><td>19</td></tr>
</table></td></tr>
</table>

PRE-PROCESSING STEPS:

1. We tried different methods of code to clean the data.
    a. The first method was to apply a code to all the datasets to clean them. - this did not work.
        i. As the data did not have the same attributes, for example one dataset worked with products, some of the datasets did not have the right countries etc. (this is all explained in the code)
    b. The second method which is shown from the work was to go through each dataset file individually and change them from wide format into long format.
2. After converting the datasets into clean files, we started to remove the outliers and make our visualisations.

MARGINS OF ERROR

	Error Margins	In Percentage
AGRICULTURAL LAND	37.74 +- 0.418	1.11 %
DEFORESTATION	437417.47 +- 68682.34	15.7 %
FOREST AREA	33.08 +- 0.519	1.57 %
LIVING PLANET INDEX	67.38 +- 3.239	4.81 %
TREE COVER LOSS	20910.55 +- 1207.988	5.78 %
WHEAT YIELDS	2.26 +- 0.034	1.49 %
GHG EMISSIONS	2.66 +- 1.175	44.14 %

STATISTICAL MODEL

- We then looked at the datasets to confirm which would be used in the regression analysis.
    - Although we did have research about the metric weighting (in the appendices) for the datasets to see what would be used for the regression analysis, this did change.

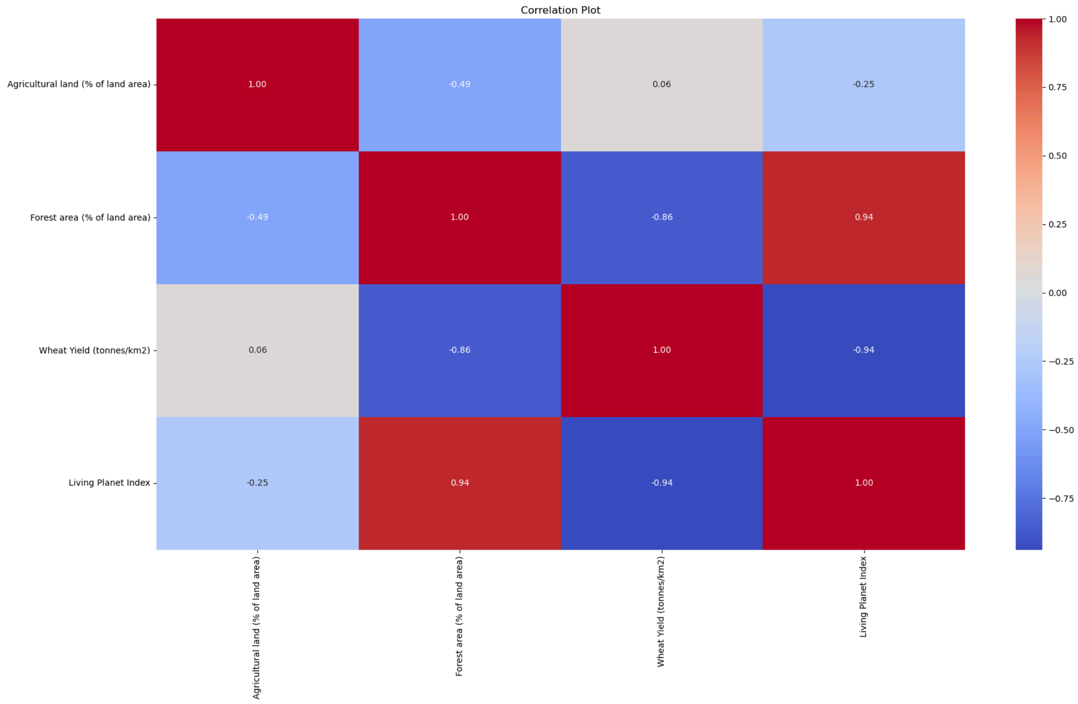
- Out of our 7 metrics:
    - GHG emissions and CO2 emissions were for 1 year only and such were not included in the overall model (plus they did not fit well with the other metrics).
    - Tree Cover Loss only had data that spread across 8 years, and due to this did not show particularly interesting results
- The metrics we will use for our statistical regression model:
    - Agricultural Land, Forest Area and Wheat Yields shall be our dependent variables
    - Living Planet Index our independent variable.

This lead to the creation of the following file (which includes the global values for each metric for the years 1990-2018):

	A	B	C	D	E
1	Year	Agricultural land (% of land area)	Forest area (% of land area)	Wheat Yield (tonnes/km2)	Living Planet Index
2	1990	37.5123125	33.62035897	2.2828125	62.68
3	1991	37.19533742	34.19247475	2.253020833	61.61
4	1992	38.48120879	33.61475771	2.331896552	60.12
5	1993	38.71118919	33.50030435	2.285086207	58.11
6	1994	38.63502703	33.49230435	2.262478632	56.2
7	1995	38.57140541	33.44195652	2.297777778	54.04
8	1996	38.55237838	33.391	2.270603448	53.04
9	1997	38.60783784	33.33978261	2.356551724	51.75
10	1998	38.648	33.31578261	2.3925	50.47
11	1999	38.776	33.26495652	2.257478261	48.66
12	2000	38.7731016	33.16133621	2.335	47.22
13	2001	38.71828877	33.12034483	2.407217391	45.67
14	2002	38.68475936	33.0825	2.465309735	44.5
15	2003	38.61513369	33.03956897	2.355	43.52
16	2004	38.36935829	32.98633621	2.570175439	42.93
17	2005	38.26877005	32.94508621	2.554741379	42.2
18	2006	38.25597884	32.98824786	2.516869565	41.3
19	2007	38.15470899	32.95692308	2.581810345	40.14
20	2008	38.26359788	32.92487179	2.642212389	38.72
21	2009	38.31248677	33.16910638	2.608468468	36.88
22	2010	38.35148148	33.1347234	2.68377193	35.15
23	2011	38.37529101	33.09076596	2.790619469	33.44
24	2012	38.37336842	32.94394068	2.752	32.37
25	2013	38.32436842	32.89550847	2.751415929	31.74
26	2014	38.26273684	32.85050847	2.752363636	31.67
27	2015	38.46510526	32.80728814	2.808392857	31.63
28	2016	38.47821053	32.75817797	2.834122807	31.15
29	2017	38.53757895	32.71025424	2.825309735	30.89
30	2018	38.53273684	32.65461864	2.804273504	30.9

The global values were calculated by either taking the average of the values through that year (Agri Land, Forest Area) or by taking its sum for that year (Wheat Yield). The LPI already had a global value for the years. With the years (1990-2018) being the years these 4 metrics had in common.

Subsequently we created a correlation matrix of these metrics:



Through this we understand that LPI is very positively correlated to Forest Area, while being negatively correlated to Agricultural Land and very negatively correlated to Wheat Yield.

Using the statsmodels python package, we conducted Ordinary Least Squares (OLS) Regression:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     Living Planet Index   R-squared:                       0.938
Model:                             OLS   Adj. R-squared:                  0.930
Method:                  Least Squares   F-statistic:                     125.6
Date:                 Tue, 26 Mar 2024   Prob (F-statistic):           3.35e-15
Time:                         01:28:20   Log-Likelihood:                -68.046
No. Observations:                   29   AIC:                             144.1
Df Residuals:                       25   BIC:                             149.6
Df Model:                            3
Covariance Type:             nonrobust
======================================================================================================
                                         coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------------
const                               -187.3967    191.742     -0.977      0.338    -582.298     207.504
Agricultural land (% of land area)    -1.3534      2.047     -0.661      0.515      -5.570       2.863
Forest area (% of land area)          10.9765      3.559      3.084      0.005       3.646      18.307
Wheat Yield (tonnes/km2)             -32.0721      5.175     -6.197      0.000     -42.731     -21.413
==============================================================================
Omnibus:                        0.160   Durbin-Watson:                   1.770
Prob(Omnibus):                  0.923   Jarque-Bera (JB):                0.354
Skew:                           0.117   Prob(JB):                        0.838
Kurtosis:                       2.512   Cond. No.                     1.93e+04
==============================================================================
```

Using these results we can figure out a formula to create a Biodiversity Score:

BS = -187.3967 - 1.3534x1 + 10.9765x2 - 32.0721x3

Where :

- x_1 is the value for Agricultural land (% of land area)
- x_2 is the value for Forest area (% of land area)
- x_3 is the value for Wheat Yield (tonnes/km2)

For example, the BS for the United States (in 2021) is:

-187.3967 - 1.3534(44.36) + 10.9765(33.87) - 32.0721(2.98)

Which would be = 28.76, this value would have a range of -187

Four Lenses (Quantitative)

In order to find a way to quantify each of the 4 lenses, a dataset related to the lens was considered:

- For Health & Environment, two new data sets were found. Those being “Exposure to PM2.5 in countries and regions” and “The Red List Index”, respectively.
- For Societal, GHG emissions caused by food production were considered as the corresponding dataset, specifically using the GHG emissions caused by wheat production. However, there was only data for 2010.
- For Carbon Impact, CO_2 emissions were considered specifically using the CO_2 emissions caused by wheat production. However, there was only data for 2013.

In the cases for GHG and CO_2 emissions, since we only had data from one year we had to extrapolate from the wheat yield the values for the other years (i.e. if 3.2472 wy = 11603387.97 tonnes of CO_2 then we configure out the CO_2 for the other years based on this, the same was done for GHG).

Subsequently, all this data was collated into one excel file in order to see what fit and what did not. Note: This data contains the global values of each metric from 1992 to 2021 (depending on whether they had the data for it).

World					carbon	environment	society	health
	Agricultral land %	Forest area %	wheat yield	Biodiversity score	wheat co2	red lst score	food	PM2.5
1992	37.11346504	32.51101914	2.534	37.9604366	9054873.5		10.02395	
1993	37.00124251	32.45485214	2.537	37.39958517	9065593.5		10.02395	
1994	37.17622275	32.39715325	2.4777	38.43131065	8853693.8		10.02395	
1995	37.16720184	32.33725784	2.4958	37.2055725	8918371.4		10.02395	
1996	37.22874218	32.27672239	2.564	34.27049921	9162074		10.02395	
1997	37.36490696	32.19881906	2.7157	28.36577038	9704151.5		10.02395	
1998	37.47590814	32.14466568	2.7063	27.92260452	9670562		10.02395	
1999	37.49655141	32.0823544	2.7461998	25.93103575	9813138		10.02395	
2000	37.52018411	32.00484633	2.7316	25.51653016	9760967.8		10.02395	26.25875
2001	37.53586019	31.96488769	2.7417	24.73278001	9797058.6	0.794	10.22288	25.1144
2002	37.47500405	31.92517773	2.7551	23.9495002	9844941.5	0.791	10.2063	25.61636
2003	37.33186008	31.88317137	2.6513999	27.00802839	9474384.6	0.787	10.16732	25.56613
2004	36.97053504	31.80694126	2.9426	17.32090717	10514945	0.784	10.06891	25.36098
2005	37.02644679	31.76606492	2.8286998	20.44957565	10107940	0.781	10.08414	24.92201
2006	37.03744216	31.72734467	2.8904	18.03082668	10328416	0.777	10.08713	24.64818
2007	36.97291801	31.68757733	2.8155	20.08384779	10060772	0.774	10.06956	25.42769
2008	36.85593543	31.65162673	3.0626	11.92254435	10943747	0.771	10.0377	25.08418
2009	36.81078171	31.61217853	3.0356998	12.41339807	10847623	0.768	10.0254	24.17933
2010	36.80543401	31.57177176	2.9721	14.01688997	10620359	0.765	10.02395	21.84841
2011	36.94991875	31.53696347	3.1639	7.28784233	11305728	0.761	10.0633	24.49121
2012	36.867753	31.47231384	3.0916998	9.005030822	11047731	0.758	10.04092	25.003
2013	36.79023248	31.43820022	3.2472	3.748281005	11603388	0.754	10.01981	24.09162
2014	36.78905099	31.40380692	3.3162	1.159387061	11849949	0.75	10.01949	23.32275
2015	36.62049552	31.36838645	3.3216999	0.822323919	11869602	0.747	9.973579	25.43931
2016	36.58701415	31.34091596	3.4150999	-2.429426423	12203353	0.744	9.964461	24.25884
2017	36.83400084	31.28999701	3.5379999	-7.264271175	12642518	0.74	10.03173	23.63849
2018	36.73845807	31.25263574	3.4243999	-3.901668955	12236586	0.737	10.00571	23.61346
2019	36.76264776	31.21375455	3.5424	-8.14569768	12658241	0.732	10.01229	22.21145
2020	36.73092029	31.17690459	3.4738998	-6.310296095	12413466	0.729	10.00365	21.95369
2021	36.8416655	31.17704902	3.4919	-7.035897474	12477787	0.726	10.03382	

As we can see our data only fit for the years between 2001 and 2020. Afterwards, the values we are using for the lenses where normalised using Z-score Normalisation:

World					carbon			environment			society			health		
	Agricultral land %	Forest area %	wheat yeild	Biodiversity score	wheat co2	wheat co2 normalized	wheat co2 Impact	red lst score	red lst score normalized	red lst score Impact	food	food normalized	food Impact	PM2.5	PM2.5 normalized	PM2.5 Impact
2001	37.53586019	31.96488769	2.7417	**24.73278001**	9,760,967.78	-1.22	-21.82	0.794	1.58	26.57	10.22	2.38	40.00	25.11	0.69	11.61
2002	37.47500405	31.92517773	2.7551	**23.9495002**	9,797,058.63	-1.18	-21.07	0.791	1.43	24.12	10.21	2.14	36.10	25.62	1.11	18.71
2003	37.33186008	31.88317137	2.6513999	**27.00802839**	9,844,941.54	-1.14	-29.35	0.787	1.23	22.41	10.17	1.58	28.82	25.57	1.07	19.43
2004	36.97053504	31.80694126	2.9426	**17.32090717**	9,474,384.61	-1.49	-21.54	0.784	1.08	39.37	10.07	0.18	6.49	25.36	0.90	32.59
2005	37.02644679	31.76606492	2.8286998	**20.44957565**	10,514,945.01	-0.51	-34.79	0.781	0.93	32.75	10.08	0.40	13.89	24.92	0.53	18.56
2006	37.03744216	31.72734467	2.8904	**18.03082668**	10,107,939.52	-0.89	-34.47	0.777	0.74	32.68	10.09	0.44	19.51	24.65	0.30	13.34
2007	36.97291801	31.68757733	2.8155	**20.08384779**	10,328,416.04	-0.68	-37.56	0.774	0.59	21.20	10.07	0.19	6.80	25.43	0.95	34.44
2008	36.85593543	31.65162673	3.0626	**11.92254435**	10,060,771.99	-0.93	-11.13	0.771	0.44	28.37	10.04	-0.27	-17.35	25.08	0.66	43.15
2009	36.81078171	31.61217853	3.0356998	**12.41339807**	10,943,747.22	-0.10	-24.41	0.768	0.29	26.45	10.03	-0.44	-40.67	24.18	-0.09	-8.47
2010	36.80543401	31.57177176	2.9721	**14.01688997**	10,847,623.35	-0.19	-15.60	0.765	0.14	4.44	10.02	-0.46	-14.80	21.85	-2.04	-65.16
2011	36.94991875	31.53696347	3.1639	**7.28784233**	10,620,358.89	-0.41	36.00	0.761	-0.06	-11.68	10.06	0.10	19.27	24.49	0.17	33.05
2012	36.867753	31.47231384	3.0916998	**9.005030822**	11,305,727.76	0.24	-6.34	0.758	-0.21	-19.03	10.04	-0.22	-20.19	25.00	0.60	54.44
2013	36.79023248	31.43820022	3.2472	**3.748281005**	11,603,387.97	0.52	30.28	0.754	-0.41	-25.91	10.02	-0.52	-33.27	24.09	-0.17	-10.54
2014	36.78905099	31.40380692	3.3162	**1.159387061**	11,849,949.24	0.75	26.99	0.75	-0.61	-22.78	10.02	-0.53	-19.83	23.32	-0.81	-30.41
2015	36.62049552	31.36838645	3.3216999	**0.822323919**	11,869,602.32	0.77	20.27	0.747	-0.75	-20.76	9.97	-1.18	-32.53	25.44	0.96	26.44
2016	36.58701415	31.34091596	3.4150999	**-2.429426423**	12,203,353.38	1.08	32.19	0.744	-0.90	-27.33	9.96	-1.31	-39.71	24.26	-0.03	-0.78
2017	36.83400084	31.28999701	3.5379999	**-7.264271175**	12,642,518.31	1.49	42.79	0.74	-1.10	-31.54	10.03	-0.35	-10.09	23.64	-0.54	-15.58
2018	36.73845807	31.25263574	3.4243999	**-3.901668955**	12,236,585.55	1.11	30.16	0.737	-1.25	-34.40	10.01	-0.72	-19.91	23.61	-0.57	-15.54
2019	36.76264776	31.21375455	3.5424	**-8.14569768**	12,658,241.42	1.51	28.09	0.732	-1.50	-27.88	10.01	-0.63	-11.71	22.21	-1.74	-32.31
2020	36.73092029	31.17690459	3.4738998	**-6.310296095**	12,413,466.10	1.28	22.58	0.729	-1.65	-29.30	10.00	-0.75	-13.39	21.95	-1.95	-34.72

The Impact columns represent the percentage by which the Biodiversity score affects that particular lense:

- A positive impact score says the lense is trending upwards.
- A negative impact score says the lense is trending downwards.
- An impact score close to 0 says the Biodiversity score has less effect on the lense.

This is represents our scores for each lens:

A	B	C	D	E	H	K	N	Q
WORLD					CARBON	ENVIRONMENT	SOCIETY	HEALTH
	AGRICULTURAL LAND %	FOREST AREA %	WHEAT YIELD	**BIODIVERSITY SCORE**	WHEAT CO2 IMPACT	RED LIST SCORE IMPACT	FOOD IMPACT	PM2.5 Impact
2001	37.53586019	31.96488769	2.7417	**24.73278001**	- 21.82	26.57	40.00	11.61
2002	37.47500405	31.92517773	2.7551	**23.9495002**	- 21.07	24.12	36.10	18.71
2003	37.33186008	31.88317137	2.6513999	**27.00802839**	- 29.35	22.41	28.82	19.43
2004	36.97053504	31.80694126	2.9426	**17.32090717**	- 21.54	39.37	6.49	32.59
2005	37.02644679	31.76606492	2.8286998	**20.44957565**	- 34.79	32.75	13.89	18.56
2006	37.03744216	31.72734467	2.8904	**18.03082668**	- 34.47	32.68	19.51	13.34
2007	36.97291801	31.68757733	2.8155	**20.08384779**	- 37.56	21.20	6.80	34.44
2008	36.85593543	31.65162673	3.0626	**11.92254435**	- 11.13	28.37	- 17.35	43.15
2009	36.81078171	31.61217853	3.0356998	**12.41339807**	- 24.41	26.45	- 40.67	- 8.47
2010	36.80543401	31.57177176	2.9721	**14.01688997**	- 15.60	4.44	- 14.80	- 65.16
2011	36.94991875	31.53696347	3.1639	**7.28784233**	36.00	- 11.68	19.27	33.05
2012	36.867753	31.47231384	3.0916998	**9.005030822**	- 6.34	- 19.03	- 20.19	54.44
2013	36.79023248	31.43820022	3.2472	**3.748281005**	30.28	- 25.91	- 33.27	- 10.54
2014	36.78905099	31.40380692	3.3162	**1.159387061**	26.99	- 22.78	- 19.83	- 30.41
2015	36.62049552	31.36838645	3.3216999	**0.8223239191**	20.27	- 20.76	- 32.53	26.44
2016	36.58701415	31.34091596	3.4150999	**-2.429426423**	32.19	- 27.33	- 39.71	- 0.78
2017	36.83400084	31.28999701	3.5379999	**-7.264271175**	42.79	- 31.54	- 10.09	- 15.58
2018	36.73845807	31.25263574	3.4243999	**-3.901668955**	30.16	- 34.40	- 19.91	- 15.54
2019	36.76264776	31.21375455	3.5424	**-8.14569768**	28.09	- 27.88	- 11.71	- 32.31
2020	36.73092029	31.17690459	3.4738998	**-6.310296095**	22.58	- 29.30	- 13.39	- 34.72

APPENDICES

METRIC WEIGHTING (PROPOSED)

Forest Area

Weighting: 20%

- Forests are essential for all types of wildlife found on the planet and contribute a significant portion towards the world's biodiversity.
- Our background research states that forests can contain:

➔ up to 60,000 different species of trees
➔ 80% of all different types of amphibians
➔ 75% of bird species
➔ Almost 68% of the worlds mammal species

- Due to this dataset containing specifically forest cover, and it being an essential factor in maintaining global biodiversity, a higher weighting of 20% is assigned.

Global Living Planet Index

Weighting: 20%

- This index was independently created by organisations to specifically measure countries' specific biodiversity over time.
- As we discovered in our research, the information links biodiversity through "measuring the changing state of the world's biodiversity over time.
- Because it directly measures biodiversity, we also assigned it a higher weighting of 20%.

Tree Cover Loss
Weighting: 15%

- Tree cover loss is directly and closely linked to fragmentation and habitat loss, which are both major drivers of biodiversity loss globally.
- Our research indicates that the forest loss produced an estimated 2.7 gigatonnes of carbon dioxide emissions and by only restoring 15% of our ecosystems in significant areas can cut extinctions down by 60%.
- This dataset specification in tree cover loss is valuable as it can be used to estimate its potential global impact on biodiversity.
- Hence a weighting of 15% has been assigned to this dataset.

Agricultural land
Weighting: 15%

- The expansion of agricultural land often comes from the removal of natural habitat, which consequently leads to biodiversity loss.
- Based on our research, we can conclude that if agricultural practices aren't made sustainable then, inevitably more land will be dedicated to agriculture.
- This will result in the displacement of more animals and destruction of their natural habitat.
- This dataset helps us assess the potential impact that agricultural practices might have on biodiversity.
- This is the reason we have weighted this dataset at 15%.

Deforestation C02 by Trade Product
Weighting: 10%

- While this dataset primarily focuses on C02 emissions, deforestation is a major driver of biodiversity loss.
- The products listed in the dataset can be indicative of the pressure that industry has on certain ecosystems and habitats.
- However, due to the fact that the data doesn't directly measure biodiversity, we have given it a lower weighting of 10%.

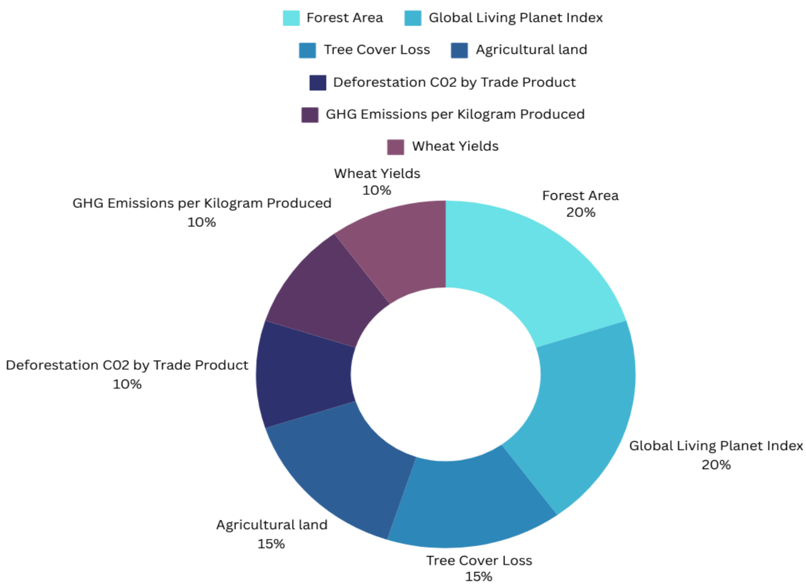
GHG Emissions per Kilogram Produced
Weighting: 10%

- Greenhouse gas emissions contribute greatly to climate change, which can have a direct impact on our global biodiversity; through habitat alterations and rising sea levels.
- Our background information states that "climate change (which greenhouse gas emissions are known to exacerbate) is one of three parts of the triple planetary crisis, alongside biodiversity loss and air pollution."
- Due to the link being indirect, however still relevant we assigned a lower weighting of 10%.

Wheat Yields

Weighting: 10%

- Higher wheat yields usually indicate a larger surface area of agricultural land being used to produce it.
- Similar to the agricultural land dataset, we can assume that compromising increasing amounts of land area for agriculture, takes away more land from nature.
- This land can only be used to cultivate wheat or its designated product, and oftentimes requires levelling the surrounding land to accommodate.
- Due to the link to agriculture land being weaker, we assigned this dataset with a lower weighting of 10%



**This was our speculated ideas on what the biodiversity weightings could be. After carefully analysing the data, this did change drastically as we did not include some of these datasets due to the various factors. One of them being some of the datasets counting for a year only and another dataset including products.

DATA SOURCING

File Name: statistic_id1227391_annual-tree-cover-loss-worldwide-2001-2022.xls - [4]

- This small dataset contains the global aggregate tree cover loss in hectares (1000 ha).
- It has data ranging from 2001-2022
- Despite our given data ranging through a much longer period of time, this aggregate data will help determine the accuracy of our data within the 2001-2022 range.
- If this data is similar, we can assume that our initial dataset is within a range of decent accuracy as a 16 year timeframe is still relevant to our context.
- Our tree cover data is based on tree cover loss per country per year, as opposed to the aggregate global loss per year. This just means we need to do some data manipulation to validate our aggregate data from our given dataset.
- The source of this dataset seems to be credible due to the source's recognized status.

Attribute Name	Data Type	Number of Entries
Year	Integer	22

TreeCoverLoss	Integer	22

File Name: statistic_id636341_forest-area-as-a-percentage-of-land-area-apac-2020-by-country.xls - [2]

- This dataset contains the forest area as a percentage by countries in the Asian-Pacific.
- The data was sourced in 2020.
- Compared to the dataset we were given, this list contains significantly less countries, however it still has enough records to be used for analysis.
- The source of this dataset seems to be credible due to the source's recognized status.

Attribute Name	Data Type	Number of Entries
Country	String	34
ForestAreaPercentage	Decimal	34

<u>Side Note: For GHG Emissions per kg, no alternative data could be found. Only multiple sources leading back to the same dataset we were provided with.</u>

File Name: statistic_id527989_global-change-in-forest-area-1990-2020.xls - [3]

- This dataset contains the global change in forest are from 1990-2020
- This dataset will be used to validate our forest area dataset by calculating our own percentage change and measuring its reliability.
- The source of this dataset seems to be credible due to the source's recognized status.

Attribute Name	Data Type	Number of Entries
Year	Integer	4
PercentChange	Decimal	4

File Name: statistic_id631346_share-of-agricultural-land-apac-2021-by-country.xls - [4]

- This dataset contains the share of agricultural land in the asian pacific region
- The data was sourced in 2021
- Its metrics will be used to compare the asian pacific countries from our dataset to theirs and determine its validity
- The source of this dataset seems to be credible due to the source's recognized status.

Attribute Name	Data Type	Number of Entries
Country	String	37
ShareOfAgricultureLand	Decimal	37

File Name: annual-change-forest-area [5]

- This dataset contains the global annual tree cover loss.
- The data was sourced in 2023.
- Its metrics will be used to compare the loss of forest area throughout the years.
- The source of this dataset seems to be credible due to the source's recognized status.

File name: forest-area-km. [6]

- This dataset contains the forest area in km.
- The data was sourced in 2023.
- Its metrics will be used to check the forest area and compare the past years with the current years.
- The source of this dataset seems to be credible due to the source's recognized status.

File Name : ghg-per-kg-poore.[7]

- This dataset contains the FGF emission per kilogram for food.
- The data was sourced in 2023.
- Its metrics will be used to compare the emission of ghg per food.
- The source of this dataset seems to be credible due to the source's recognized status.

File Name: statistic_id267268_wheat_-production-volume-worldwide1990-1991-2023-2024. [8]

- This dataset contains data about the wheat production over the years.
- The data was sourced in 2024.
- Its metrics will be used to compare the global wheat production over the years.
- The source of this dataset seems to be credible due to the source's recognized status.

ERROR MARGINS (NOT USED - BEGINNING STEPS) :

In our project the margin of error is calculated by taking the product of the critical value and the standard deviation of the sample, divided by the square root of the sample size. This value is then divided by the sample mean. The resulting margin of error represents the range within which the true population parameter is likely to lie, given a certain level of confidence.

- Agricultural Land :
  - Margin of Error: 1.11%
    - This indicates that, on average, the estimated value of agricultural land area is expected to vary by approximately 1.11% from the true population value.
- CO2 Emissions :
  - Margin of Error: 69.74%

        - The high margin of error suggests substantial variability in the estimated CO_2 emissions. This could be due to various factors such as measurement errors or significant differences in CO_2 emissions across different entities.
- Forest Area :
    - Margin of Error: 1.59%
        - This implies that the estimated forest area percentage is expected to vary by about 1.59% from the actual value.
- GHG Emissions :
    - Margin of Error: 60.81%
        - The relatively high margin of error suggests significant uncertainty in the estimated greenhouse gas emissions per kilogram of production. This could be due to variations in production methods or incomplete data.
- Living Planet Index:
    - Margin of Error: 4.81%
        - The margin of error indicates that the estimated Living Planet Index value is expected to deviate by approximately 4.81% from the true population value.

- Tree Cover Loss :
    - Margin of Error: 12.84%
        - This suggests that there is a moderate level of uncertainty in the estimated tree cover loss in hectares.
- Wheat Yield :
    - Margin of Error: 1.68%
        - The margin of error implies that the estimated wheat yield per square kilometre may vary by around 1.68% from the actual value.
- We can conclude that these results highlight the variability and uncertainty in the estimated values of the biodiversity metrics.

REFERENCES

[1] - GFW. (2023). Global annual tree cover loss from 2001 to 2022 (in 1,000 hectares). Statista. Statista Inc.. Accessed: March 17, 2024.
https://www.statista.com/statistics/1227391/annual-tree-cover-loss-global/

[2] - World Bank. (2023). Forest area as a percentage of land area in the Asia Pacific region in 2020, by country . Statista. Statista Inc.. Accessed: March 17, 2024.
https://www.statista.com/statistics/636341/asia-pacific-forest-share-by-country/

[3] - FAO. (2020). Percent change in forest area worldwide from 1990 to 2020. Statista. Statista Inc.. Accessed: March 17, 2024.
https://www.statista.com/statistics/527989/loss-and-gained-forest-area-by-region-worldwide/

[4] - World Bank. (2023). Share of agricultural land in the Asia-Pacific region in 2021, by country or territory. Statista. Statista Inc.. Accessed: March 17, 2024.
https://www.statista.com/statistics/631346/asia-pacific-size-of-agricultural-land-by-country/

[5] - Ritchie, H. (2021). Deforestation and Forest Loss. [online] Our World in Data. Available at: https://our worldindata.org/deforestation.
[6] - Ritchie, H. (2021). Forest area. [online] Our World in Data. Available at: https://ourworldindata.org/forest-area.
[7] - ur World in Data. (n.d.). Greenhouse gas emissions per kilogram of food product. [online] Available at: https://ourworldindata.org/grapher/ghg-per-kg-poore.
[8] - Statista (2018). Production of wheat worldwide, 2018/2019 | Statista. [online] Statista. Available at: https://www.statista.com/statistics/267268/production-of-wheat-worldwide-since-1990/.

PROJECT PLAN

Team Details

Name	Role/Responsibilities
Amirreza Nozari (AN)	**Project Manager** • **Evaluating the data source and research methods.** • **Working with the data quality issues.** • **Adding additional data sets.** • **Communicating with the rest of the team to see if the work is completed on time.**
Naif Alharthi (NA)	**Specialist** • **Communicating with the team to help them with the data if needed.** • **Evaluating the data in python to understand it closely.** • **Making visuals and converting the data into suitable forms for analysis.**
Catalin Soare (CS)	**Complete/Finisher** • **Communicating with the team to see if any changes need to be done.** • **Looking closely at the data.** • **Comparing different data sets to see how they are.** • **Taking the last look at the work to make sure it is fit to send.**

Karim Mokbel (KM)	**Co-Ordinator** ● **Communicating with everyone to make sure they are on track and understand.** ● **Checking the data sources of our work.** ● **Adding more data sources to improve our accuracy.**
Zeynab Sheikh-Ali (ZS)	**Plant** ● **Evaluating the data research methods.** ● **Creating models of our data to visualise the work.** ● **Comparing all the data sources we have found.** ● **Communicating with others if they need help with new ideas.**
Aliyah Aslam (AA)	**Resource Investigator -** ● **Evaluating the data research methods.** ● **Identifying the data quality** ● **Comparing the data sources and finding new data to help reduce error margins.** ● **Communicating with the team if research help is needed.** ● **Submits final pieces.**

Project Background

The main business challenge that our team is looking to answer is providing Koru Impact Solutions with accurate insights into the environmental impact of financial investments. Our main focus is the biodiversity metrics. The client has scores of publicly available data that they need processed and cleaned in order to perform analysis of the various environmental factors that companies affect, and how investment in these companies influence those factors.We face various challenges in the undertaking of this task. One such example is preparing the data, which involves researching its reliability; cleaning the data; handling missing values etc. We also need to create statistical models, use them to measure error margins of the data and compare the data to other datasets related to biodiversity, in addition to researching the various differences between data collection and measuring worldwide in these other datasets. This serves the goal of providing the client with accurate, standardised, broad/globalised data fit to be used by models that can determine environmental impact. Understanding the environmental impact of these financial investments is important for Koru Impact Solutions due to many reasons:

1. Koru's clients wish to gain a more in-depth understanding of the businesses or stocks that they are planning on investing in.
2. To ensure that the data Koru is working with is valid, it must be compared to other

similar globally available data.

3. The data being accurate and reliable is crucial for Koru Impact Solutions LTD as it is the only metric which they use to judge a company or stock.
4. Creating an accurate model of the company's sustainability factors, requires a large set of accurate and reliable data. Sourcing this data is a challenge, and ensuring its reliability is even more important. Making sure that the data is accurate and reliable is crucial as this would be closely tied with the reputation of Koru.
5. Koru Impact Solutions provides their clients with a processed report of a requested company's sustainability.

Understanding the environmental impacts of financial investments in relation to biodiversity is essential for Koru Impact Solutions to meet the clients expectations,create sustainability and keep a high reputation.

Project Scope

Project Outcomes	Key activities to achieve the outcome	Activity lead
Project Plan	1. Figure out which tasks must be done to complete the project. 2. Determine team roles and responsibilities 3. Create realistic timeline, with a meeting schedule 4. Identify risks and create mitigation plans. 5. Write & submit project plan. 6. Project Plan Feedback 7. Implementing feedback into PP	• AN (ALL) • AA (ALL) • CS (ALL BUT 5,7) • NA (ALL BUT 7) • KM (ALL BUT 5,7) • ZS (ALL)
Data Exploration & Preprocessing	1. Evaluate the sources of data and the research methods to determine inherent accuracy risks or biases in data 2. Explore the variables using visual and aggregation techniques 3. Identify data quality issues. 4. Handle missing values and outliers 5. Convert data into suitable format for analysis (using various data transformation techniques). 6. Status Update Week 3 7. Team 360 Review 8. Status Update Feedback	• AN (ALL BUT 2,5) • AA (ALL BUT 2,5) • CS (ALL BUT 3,4) • NA (ALL BUT 3,4) • KM (ALL BUT 3,4) • ZS (ALL BUT 2,5)
Data Analysis	1. Writing Technical Report 2. Submit Technical Report 3. Technical Report Feedback Review 4. Creating a statistical model to assess the error	• AN (ALL BUT 2,7) • AA (ALL BUT 4,5) • CS (ALL BUT 2,8) • NA (ALL BUT)

	margins for the data. 5. Evaluate each metric and compare their error margins per country. Determine which ones are more precise. 6. Compare current data sources with others. 7. Add additional datasets that improve the accuracy of our metrics and reduces our error margins 8. Apply the new outputs through 4 lenses (Health, Society, Environment and Carbon Intensity). Determine how changes in one affect the others 9. Midpoint Reflection Submission	• KM (1,3,6,7,8,9) • ZS (1,3,4,5,6,9)
Project Report	1. Status Update Week 5 2. Status Update Feedback 3. Formulating the preliminary outline for the draft report 4. Submitting the draft report 5. Assessing the critique provided on the preliminary draft report. 6. Engaging in the development of the live presentation. 7. Live presentation 8. Live Feedback / Finishing Final Report 9. Submit the final report	• AN (ALL BUT 9) • NA (ALL BUT 4,9) • CS (ALL BUT 4,9) • KM (ALL BUT 4,9) • ZS (ALL BUT 4,9) • AA (ALL)
Resources required	• Google collaborative workspace • Biodiversity Data provided by Katherine • Online repository of similar relevant data • (Github) • Computers • Python/R compatible IDE • Interview with Mrs. Katherine • Collaboration space on campus (ie. empty room for meetings) • Jupyter Notebook	

Project Constraints	• Team coordination
	    ○ Miscommunication or poor time management skills.
	    ○ Not attending meetings for personal reasons.
	• Sourcing reliable/valid data from 3rd parties
	• Individual collaboration (ie. equal contribution)	

Assumptions	• It has been assumed that :
	    ○ There will be many publicly available datasets for us to use which relate to biodiversity.
	        ■ They are easily accessible for analysis.
	    ○ Clients will be engaging with us actively.
	    ○ Team members have the skills they have said they possess.
	    ○ Team has access to the resources required above.	

	○ Activities that we go through in this report, will be in guidelines to data ethics. ○ We will follow the project plan as a whole for our timeline of this project.
Risks (+Mitigation Plans)	● Data Quality Issues ○ Incomplete datasets given/found. ■ Assessing the quality of the data and going through data validation checks to look at inconsistencies.for not meeting a deadline will result in exclusion from credit. ● Inability of finding datasets ○ Not finding relevant data relating to biodiversity which limits the depth of our analysis ○ Finding similar datasets to what has been provided. ■ Group collaboration on finding datasets. ■ Seeking guidance from an expert. ● Team coordination ○ Due to individual or personal circumstances it is possible for one or more members of the group to be unable to meet or do their work. ○ Poor communication and personal time availability may vary. ■ We must ensure open communication between the team. ■ We need to make sure to have more team meetings ● Data Loss ○ Losing data while preprocessing which cause the team damage ■ having frequent data checks and going through a data validation process to get rid of this fear. ● Scalability ○ High amounts of data can cause system errors and take a long time. ■ use certain amounts of data at a time, making sure to optimise data processing. ● Individual Work Incomplete ○ not completing individual work given ■ following the project plan and a consequence for not meeting deadlines will result in exclusion from credit.

Meeting Dates/Times/Venues

- Our meetings have been scheduled below. This is an outline of the dates we will meet up in our university.
- We have set up a whatsapp group chat to keep regular contact with each other.
- Meetings online if we miss our scheduled meeting below.
- Work done on google docs to track who's doing what.

Proposed Team Meetings:
Week 1 : (06/02/24 -12pm)

Week 2 : (13/02/24 - 11am) (14/02/24 - 12pm)
Week 3 : (20/02/24 - 11am) (22/02/24 - 2pm)
Week 4 : (27/02/23 - 12pm) (01/03/24 - 3pm)
Week 5 : (05/03/24 - 12pm) (07/03/24 - 12pm)
Week 6 : (12/03/24 - 12 pm) (13/03/24 - 12 pm)(16/03/24 - 6pm (17/03/24 - 7pm
Week 7 : (20/03/24 - 11am) (22/03/24 - 1:30pm) (25/03/24 - 1pm)
Week 8 : (26/02/24 - 11am) (27/03/24 - 3pm)

GANTT CHART

PROJECT TITLE	Biodiversity	COMPANY NAME	Koru Impact Solutions LTD
PROJECT MANAGER	Amirreza Nozari	DATE	13/02/24

PHASE	TASK TITLE	START DATE	DUE DATE
1	**ORIENTATION & TEAM FORMATION**		
1.1	Welcome & Orientation Session	05/02/24	08/02/24
1.2	Submission of the teamwork toolkit and team charter	05/02/24	08/02/24
1.3	Reviewing the project brief	05/02/24	08/02/24
2	**Project Plan**		
2.1	Determine team roles and responsibilities	13/02/24	15/02/24
2.2	Figure out which tasks must be done to complete the project.	13/02/24	15/02/24
2.3	Create realistic timeline, with a meeting schedule	13/02/24	15/02/24
2.4	Identifying risks and create mitigation plans	13/02/24	15/02/24
2.5	Writing & submit project plan	13/02/24	15/02/24
2.6	Project plan feedback	16/02/24	
2.7	Implementing the feedback into project plan	16/02/24	
3	**Data Exploration and Preprocessing**		
3.1	Evaluate the sources of data and the research methods to determine inherent accuracy risks or biases in data	16/02/24	26/02/24
3.2	Explore the variables usuing visual and aggregration techniques	16/02/24	26/02/24
3.3	Identify data quality issues	16/02/24	26/02/24
3.4	Handle missing values and outliers	16/02/24	26/02/24
3.5	Convert data into a suitable format for analysis	16/02/24	26/02/24
3.6	Status Update Week 3	16/02/24	26/02/24
3.7	Team 360 Review	16/02/24	26/02/24
3.8	Status Update Feedback	27/02/24	
4	**Data Analysis**		
4.1	Working on Technical Report	16/02/24	28/02/24
4.2	Submit Technical Report	16/02/24	28/02/24
4.3	Technical Report Feedback Review	29/02/24	
4.4	Creating a statistical model to assess the error margins for the data.	27/02/24	04/03/24
4.5	Evaluate each metric and compare their error margins per country. Determine which ones are more precise.	27/02/24	04/03/24
4.6	Compare current data sources with others.	27/02/24	04/03/24
4.7	Add additonal datasets that improve the accuracy of our metrics and reduces our error margins	27/02/24	04/03/24
4.8	Apply the new outputs through 4 lenses (Health, Society, Environment and Carbon Intensity). Determine how changes in one affect the others.	27/02/24	04/03/24
4.9	Midpoint Reflection Submission	27/02/24	04/03/24
5	**Project Report**		
5.1	Status Update Week 5	05/03/24	11/03/24
5.2	Status Update Feedback	12/03/24	
5.2	Formulating the preliminary outline for the draft report	05/03/24	18/03/24
5.3	Submitting the draft report	05/03/24	18/03/24
5.4	Assessing the critique provided on the preliminary draft report	19/03/24	26/02/24
5.5	Engaging in the development of the live presentation.	19/03/24	28/03/24
5.6	Live Presentation	26/03/24	26/03/24
5.7	Live Feedback/ Working on Report	26/02/24	28/03/24
5.8	Submit Final Report	26/03/24	28/03/24