```
                            ( Start )
                               |
                               v
+--------------------------------------------------------------------+
| **Data Collection:**                                               |
|                                                                    |
|   • Annual sustainability reports in PDF are downloaded from each   |
|     company website                                                |
|   • If sustainability reports are not available, corporate         |
|     financial reports are considered instead                       |
|   • SDG reports in PDF are downloaded and used as benchmark         |
+--------------------------------------------------------------------+
                               |
                               v
+--------------------------------------------------------------------+
| **Text Preprocessing:**                                            |
|                                                                    |
|   • Extract blocks of text from reports                            |
|   • Split text into relevant sentences                             |
|   • Clean text and remove special characters that are not relevant |
+--------------------------------------------------------------------+
                               |
                               v
+--------------------------------------------------------------------+
| **Text Vectorization:**                                            |
|                                                                    |
|   • The extracted text is split into individual sentences          |
|   • The sentences are converted into vectors using a pre-trained   |
|     model                                                          |
+--------------------------------------------------------------------+
                    |                          |
                    v                          v
+-------------------------------------+  +-------------------------------------+
| **Thematic Analysis:**              |  | **Sentiment Analysis:**             |
|                                     |  |                                     |
|  • Cosine similarity is used to     |  |  • Classification of sentiment      |
|    calculate similarity score       |  |    scores is performed using a      |
|    between the vectors from the     |  |    pre-trained model                |
|    sentences from corporate reports |  |  • Calculate the ratio of positive  |
|    and SDG reports                  |  |    and negative sentiment scores    |
|  • Similarity scores are normalised |  |                                     |
|    between 0-100                    |  |                                     |
+-------------------------------------+  +-------------------------------------+
                    |                          |
                    +------------+-------------+
                                 v
                  +-------------------------------+
                  | **Visualisation:**            |
                  | Aggregate multi-level groups  |
                  +-------------------------------+
```