GREENVEST: INVESTING IN GREEN COMPANIES

Yassine Bouanani
B.S. Computer Science and Economics, Northeastern University, Boston, United States

Oleksandr Litus
B.S. Computer Science, Northeastern University, Boston, United States

Shreya Mitra
B.S. Computer Science and Business Admin, Northeastern University, Boston, United States

Jingwei (Darcy) Zhang
B.S. Mathematics and Business Admin, Northeastern University, Boston, United States

Jacqueline Zhao
B.S. Mathematics and Business Admin, Northeastern University, Boston, United States

Greenvest Website Link
https://greenvest-app.herokuapp.com/

June 2021

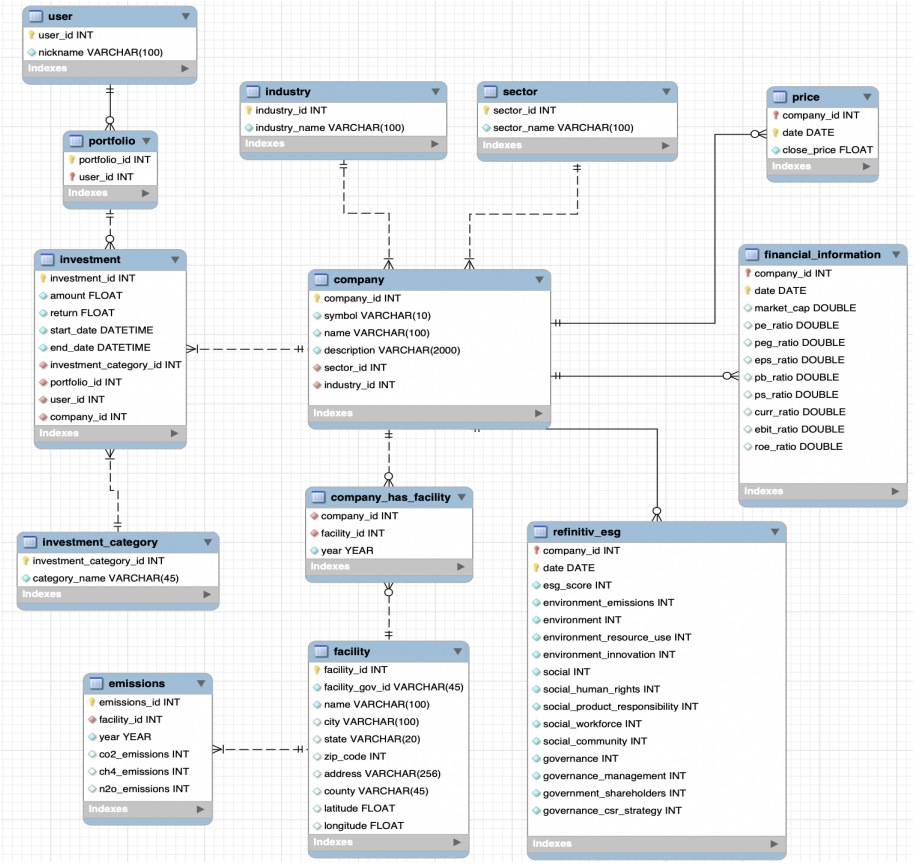
Introduction

Climate change is a global challenge. We see an unprecedented opportunity for our clients to apply sustainable finance to make a global impact as changes occur. Total issuance of sustainable debt including green, social and sustainable bonds has surpassed one trillion dollars. Renewable Energy now accounts for a third of global power capacity. And environmentally conscious Millennials inherit more than fifty trillion dollars in the incoming decades.

For this project, we aim to investigate the correlation between specific corporate investments into energy consumption and pollution and their financial returns coupled with their impact on the socioeconomic status of the regions they are in. This correlation would help inform investors interested in sustainable investing of which companies are good investments for them based on their specific criterias for sustainability as well as the financial performance of the companies.

Our goal is to develop an app that can expose fake rebranded sustainable companies and facilitate sustainable investing by providing data-driven insights on companies' environmental, social and governance initiatives. We want to be able to highlight the fact that young investors can make investment decisions that both align with their environmental values and make good returns in the long-term, pointing out trends that indicate an increased significance in a sustainable future.

Database Design



We represented our data using 13 tables and wanted to design our database in a way that also represents the functionality of an app. At the center of our database is the company entity that is represented by a ticker symbol, name, and the description as well as the company, sector, and industry ID. The company is connected to investments through a one to many relationship, which contains fields that represent the amount, return, start date, and end date of an investment. An investment is connected to the investment category through a many-to-one relationship to represent the type of investment, such as whether the investment is a stock, ETF, bond or, derivative. We also created entities to be able to represent a user and their ability to create an investment portfolio, which is connected to the investment entity. The company is also connected to the industry and sector, which both include a name and an ID, as well as the price entity.

We also created a financial information entity that is connected to the company through a one to many relationship, containing all the ratios that are important for determining the market value, long-term profitability of a company, and the overall stock value relative to competing stocks. The company entity is also connected to the ESG data, containing an ESG score and the specific environmental, social, and governance scores that make up each sub-category. The company entity is also connected to the facility through a many to many relationship with the facility entity containing fields to represent the name and specific location of the facility. The emissions entity is connected to this facility, and it contains the year of the recorded emissions as well as the specific carbon, methane, and ammonia emissions.

Triggers and Queries

```
drop trigger if exists ESGRankingTrigger;

delimiter //
create trigger ESGRankingTrigger after insert on facility
for each row
begin
select RANK() OVER (
    ORDER BY esg_score DESC
    ) company_esg_ranking
    FROM
    refinitiv_esg;
end;

drop trigger if exists ROERankingTrigger;

drop trigger if exists EmissionsRankingTrigger;
delimiter //
create trigger EmissionsRankingTrigger after insert on facility
for each row
begin
select RANK() OVER (
    ORDER BY co2_emissions + ch4_emissions + n2o_emissions DESC
    ) company_emissions_ranking
    FROM
    emissions;
end;
```

```
-- retrieves company name, ticker, esg score
select c.name company_name, c.symbol company_symbol, esg_score
from company c inner join refinitiv_esg using company_id;

-- retrieves company name, ticker, and all financial data for company
select c.name company_name, c.symbol company_symbol, f.date, market_cap,
e_ratio, peg_ratio, eps_ratio, pb_ratio, es_ratio,
 curr_ratio current_ratio, ebit_ratio, roe_ratio
from company c inner join financial_information f using company_id;

-- retrieves company name, ticker, esg, and roe
select c.name company_name, c.symbol company_symbol, esg_score, f.roe_ratio
from company c inner join refinitiv_esg re on c.company_id = re.company_id
join financial_information f on f.company_id = c.company_id;

-- retrieves the company symbol, facility name, year,
-- and all three types of emissions
select c.symbol company_symbol, f.name facility_name, e.year,
 e.co2_emissions, e.ch4_emissions, e.n2o_emissions
from company c join company_has_facility chf using company_id
join facility f on f.facility_id = chf.facility_id
join emissions e on f.facility_id = e.facility_id;
```

We also worked on creating triggers that can rank companies based on specific data.This included creating triggers that ranked companies based on their ESG scores, total emissions produced, and return on equity ratios to determine overall profitability of the company. We also created many queries for data retrieval. For example, we produced queries that retrieve the general information about the company and what their ESG score is. We created queries that

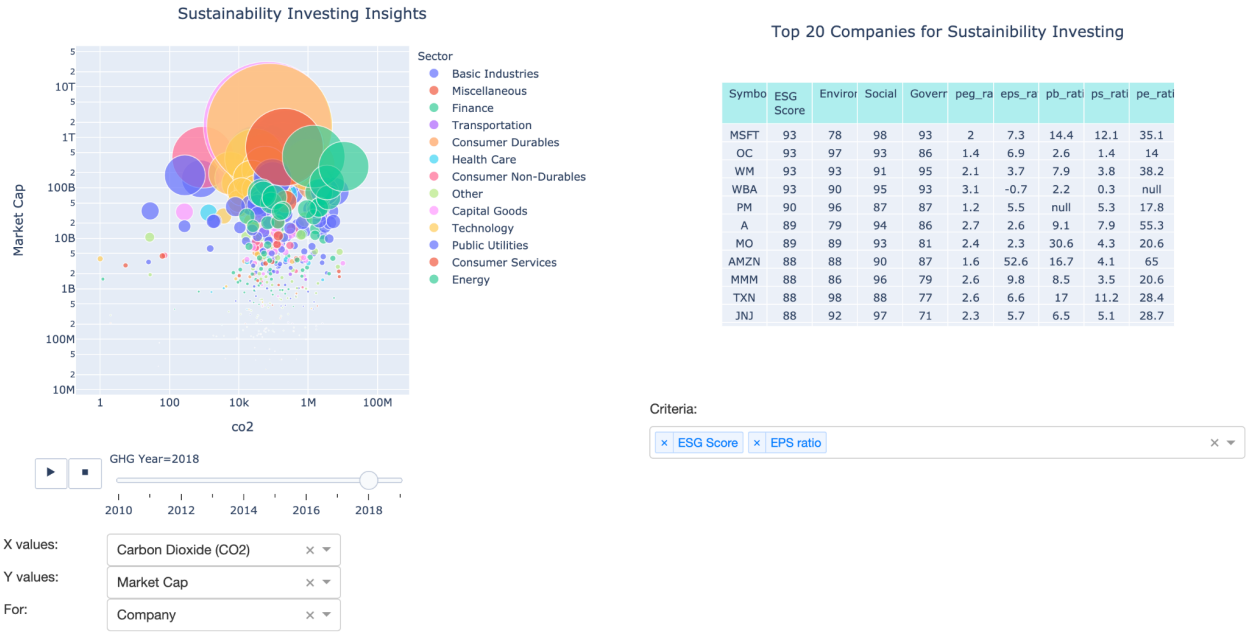
retrieve a company and get all the ratios necessary to analyze the financial standing for a company. We also created a query that gets a company's facility names and the level of carbon, methane, and ammonia emissions that each facility produces.

Methodology

The data used in this project consisted of raw csv files and json returns from various sources which included Refinitive, EPA, yFinance, and financial modeling prep. The raw csv files were downloaded straight from the source sites while various python scripts were created to collect data from sources that used APIs to access their data. These python scripts ran url requests to retrieve company data and company ticker/stock data, and were stored in csv format to enable local access to the data. Although the raw data collected was not formatted in a convenient manner, the data contained yearly company financials data, yearly company emissions data, company stock data, yearly company facility data, facility locations and sector/industry, yearly facility emissions data, and general company data such as company descriptions, industry, and sector. These categories were scattered amongst numerous csv files without much uniformity between them, thus making it extremely difficult to load the csv files into MySQL and perform SQL queries on the tables. This issue was solved by using a multitude of python scripts that iterated through the csv files in question to filter out unwanted data and gather the interesting data to one location. An inconsistency in the naming convention of the same companies across different data files caused an issue while trying to aggregate data from the csv files, so a script performing partial matching of companies in one csv file to companies in another csv file was also created to solve this issue. The script iterations were repeated until an orderly amount of csv files containing the information necessary to create GreenVest's database easily were created. Each csv file created represents a table shown in the database design above. MySQL load statements were used to load the csv files into the MySQL database. Later visualizations of the data in the database were created with the python libraries Plotly and Dash.

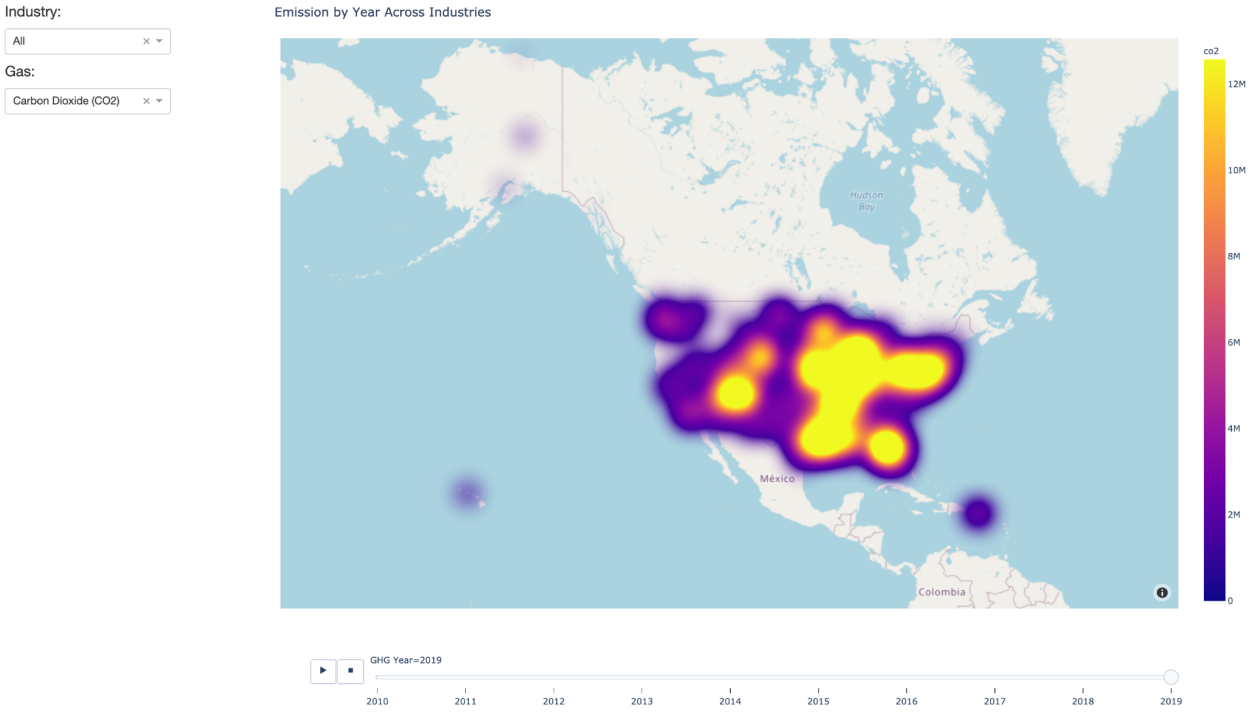
Analysis, Visualizations, and Insights

Below is our app's main page where users can gather investment insights from interactive market and greenhouse gas emission graph and Top 20 companies' ESG score chart. Users can input different values for X values for filtering based on different types of greenhouse gases, Y values for market cap and volume, with conditioning on company, sector, or industry. Our users also have the ability to filter Top 20 companies for sustainable investing based on their preferences.



Symbo	ESG Score	Enviror	Social	Goverr	peg_ra	eps_ra	pb_rati	ps_rati	pe_rati
MSFT	93	78	98	93	2	7.3	14.4	12.1	35.1
OC	93	97	93	86	1.4	6.9	2.6	1.4	14
WM	93	93	91	95	2.1	3.7	7.9	3.8	38.2
WBA	93	90	95	93	3.1	-0.7	2.2	0.3	null
PM	90	96	87	87	1.2	5.5	null	5.3	17.8
A	89	79	94	86	2.7	2.6	9.1	7.9	55.3
MO	89	89	93	81	2.4	2.3	30.6	4.3	20.6
AMZN	88	88	90	87	1.6	52.6	16.7	4.1	65
MMM	88	86	96	79	2.6	9.8	8.5	3.5	20.6
TXN	88	98	88	77	2.6	6.6	17	11.2	28.4
JNJ	88	92	97	71	2.3	5.7	6.5	5.1	28.7

Our industry page has an embedded map indicating how different types of industry contribute to greenhouse gas within the US on a yearly basis. Users can choose the industry they are interested in making investments in and make a financial decision to see if their investment in the specific industry will make an impact on the environment.



Overall, our platform facilitates users' research on companies, sectors, and industries' level of sustainability coupled with their financial performance by displaying interactive visualizations that offer insights into sustainability investing.

Shortcomings and Next Steps

One of our major challenges is the lack of data. We weren't able to find greenhouse emission data at the company level. The free APIs that we used also limit requests and do not offer real time data. The raw csv files of emission data that we used also posed challenges as they did not contain the companies tickers and therefore we had to perform partial matching between the company names in our emission data and those in our financial data that had corresponding tickers and therefore data accuracy was compromised. SQL also posed limits as the data loading process takes too much time and data cannot be optimized to decent levels. NoSQL solutions might have facilitated the process and provided additional insights.

The next steps of this project would be to expand the investment category to include not only stocks but also bond and derivative markets, acquire real time data and a more comprehensive set of data crucial to the area of sustainability investing and improve user interface. The necessary steps involved in augmenting the interface include implementing AI for better financial advising, connecting with banks and stock and security exchange platforms, and allowing users to create portfolios and invest directly within our app.

Conclusion

There is no comprehensive tool on the market that would let users explore and compare both companies' financial information and its socio-ecological impact on the planet and community. Furthermore no dataset exists that would include all necessary green data associated with each publicly traded company. Based on the current limitations our project appeared to have high relevance, especially during the time when people are more worried than ever about their impact on the planet. Green investing has huge potential and requires more investigation. Our approach was not comprehensive and would be improved with more data. Based on the data we acquired we noticed that a lot of times not only companies that use facilities are responsible for the emissions, but also the owners of such facilities. You can see from the scatter plot on our main page, that the most impactful sector on greenhouse emissions is financial. Also additional interesting insights can be found. One of the examples is Tesla factories that produce significant carbon emissions footprint, while being marketed as a green alternative to non-electric cars. We believe that every user of our platform can investigate data easily and find additional insights relevant to their interests.

Author Contributions

Jacqueline contributed to the project by writing helper functions to extract data from yahoo finance API, producing visualizations for financial, esg and emission data and writing codes for plotly dash user interface to display the visualizations. Shreya's main focus on the project was the design of the database and mapping of the relationships between data as well as writing functions and queries on SQL to show specific trends in the data and retrieve information. Oleks' main focus during the project was data sources search and preprocessing, as well as final app design and code using previously made plotly figures. Darcy's focus was on raw

data collection, app building idea process, artistic design of GreenVest App, financial advising and analysis. Yassine's main focus was the processing of raw csv/xlsx files into acceptable csv files to prepare the data to be easily loaded into the MySQL database, the partial matching of companies across csv files, and the gathering of raw financial data from API access sources.

References

"ESG Scores." *Refinitiv*, www.refinitiv.com/en/sustainable-finance/esg-scores.

"Yfinance." *PyPI*, pypi.org/project/yfinance/.

"GHG Reporting Program Data Sets." *EPA*, Environmental Protection Agency, 19 Feb. 2021, www.epa.gov/ghgreporting/ghg-reporting-program-data-sets.