Capstone Project Report

Overview of energy start-ups and prediction for the future trend of energy usage

By Poon Ka Nok Anson (a1818504)

Abstract

Understanding factors associated with energy production and consumption can highlight issues in global warming, with the increase in awareness of environmental protection, people advocate using clean energy. However, the research on clean energy start-up sector and public energy consumption pattern of clean energy are lacking. Here we show that machine learning successfully explore the lifetime of energy start-ups, from their creation to their exit, following with examination of energy consumption. The foundation of clean energy start-ups has an upward trending over the year, assisting with higher chance of receive funding and success for clean energy start-ups, the number of clean start-up foundation is expected to increase in the future and consumption of energy will lean toward green energy sources. Our results suggest that clean energy start-up have more advantage than traditional energy start-up and the public tend to use clean energy on their daily basis.

Introduction

Project Aim

This report inspects the energy start-up environment as a whole and explores the future trend of start-up foundation and consumption of energy by undertaking a comprehensive analysis on energy start-up sectors and global energy consumption data.

Key Hypotheses

1: Clean energy start-ups has superior advantage in secure funding and exit start-up stage earlier compared to traditional energy.

2: The use of clean energy is increasing in recent years and is expected to surpass traditional energy sources, potentially dominating the energy consumption sector.

Background

This report presents valuable insights for individuals interested in the advancement of clean energy sources, especially in a modern society that emphasize on sustainability. It serves as an incentive for governments and large corporations to offer investments into clean energy start-ups, leading to the growth of foundation and, most importantly, facilitating their success. By exploring into the historical context of energy start-ups and consumption data, this report also aims to make prediction regarding the future number of clean-energy start-ups and energy consumption.

Three stages of a start-up

The establishment of an energy start-up entails several stages, as previously mentioned, including creation, funding and success. Each of these stages is equally essential to the growth of the start-up. In the creation stage, the entrepreneur has to consider which energy is the business going to put resource in. During the funding stage, start-up tends to find funding from other parties with a larger scale, for example, receive funding from local government and more established firms. After receiving the financial support, the start-up can use the funding to cover operating expenses and continue running the business. The final stage of a start-up is to perform an initial public offering (IPO). IPO is the process of offering a company's shares to the public through a new stock listing. The purpose of this action is to raise capital and increase the scale of the start-up. The other way of exiting the start-up phrase is to being acquired by other company (Acquisition). Acquisition is a corporate action in which one company purchases most or even all another company's shares to gain full control of that company. Usually, a bigger, more successful company purchases a smaller start-up and gain the ownership of the start-up.

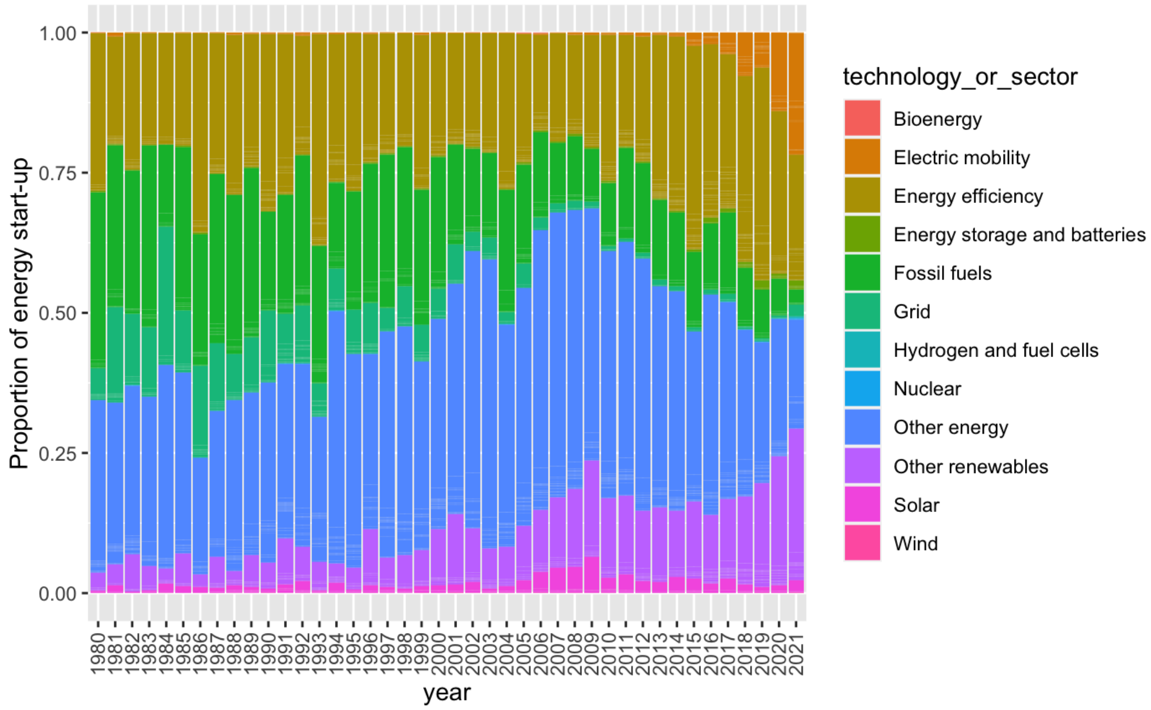


Figure 1: Filled barplot showing the distribution of energy start-up's types from 1980 to 2021

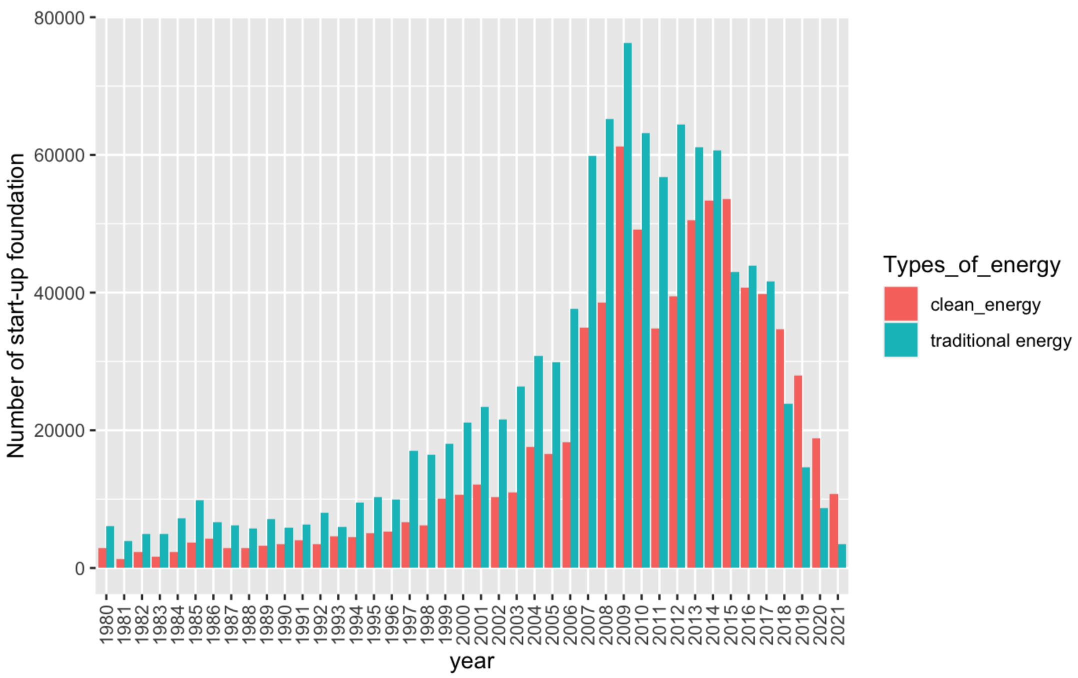


Figure 2: Barplot illustrating the distribution of clean and traditional energy start-ups from 1980 to 2021

Since 1980, the energy start-up sector has seen a growing interest in fossil fuels and energy efficiency. However, by 2000, there is a notable decline in fossil fuel, with an increase in number of energy start-ups shifting their focus toward clean energy sources. In 2021, the advent of electric vehicles provides incentive for entrepreneurs to invest **(Figure 1)**. While more entrepreneurs are willing to invest in renewable energy, it is important to note that traditional energy start-ups have dominated the industry until 2015. The period from 2007 to 2010 marks the peak in the founding of energy start-ups, with a subsequent decline starting 2018. In recent years, the number of clean energy start-ups being founded has surpassed those in traditional energy sector **(Figure 2)**.

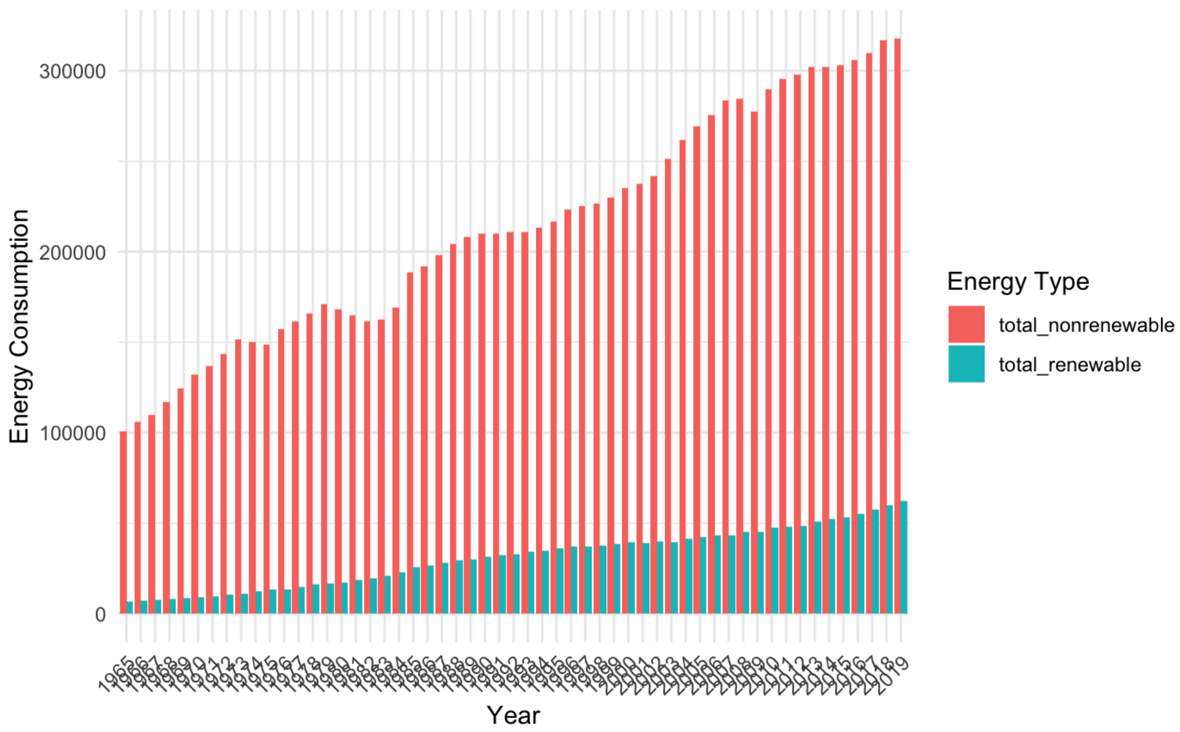


Figure 3: Barplot showing the total energy consumption from 1965 to 2019

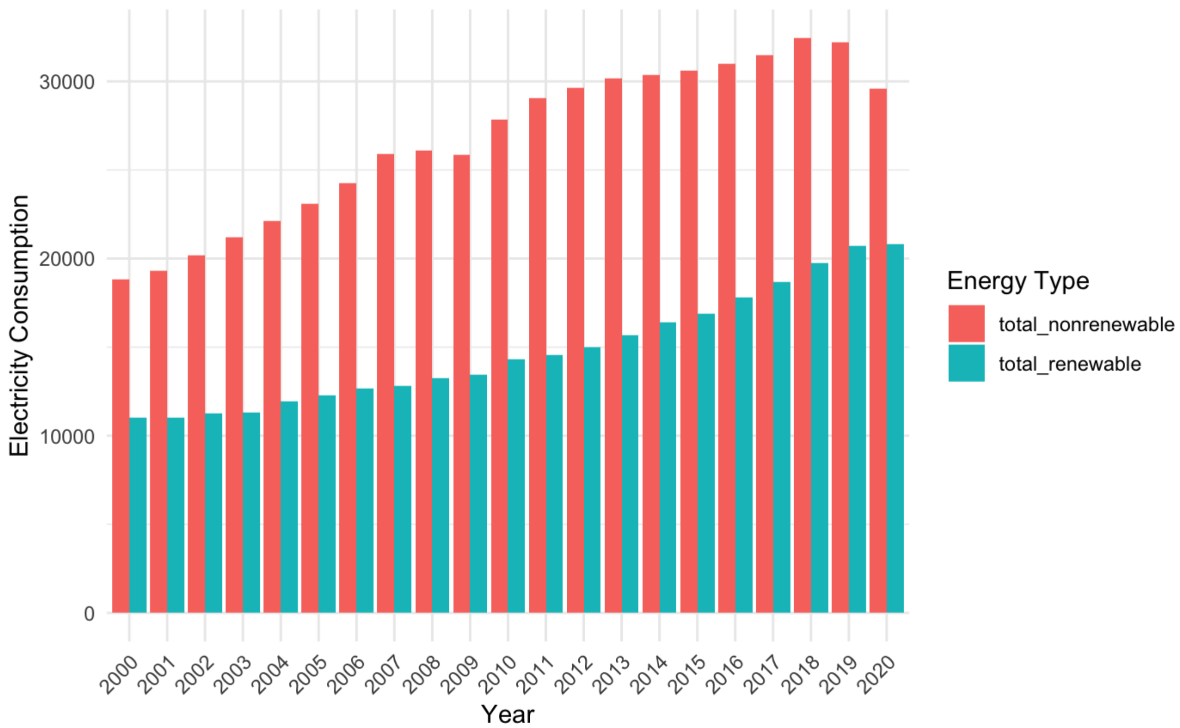


Figure 4: Barplot showing the total electricity consumption from 2000 to 2020

Energy start-ups implies the potential production of energy whereas energy consumption reflects current and future energy usage trends. It is noteworthy that the consumption of both renewable and non-renewable energy sources has been on the rise **(Figure 3)**. Additionally, electricity consumption serves as a significant indicator for energy consumption, as electricity is considered a secondary energy source generated from primary energy source. There is a ideal upward trend in the utilization of renewable energy in the electric sector with non-renewable energy shows a stable trend **(Figure 4)**. Following the increase in world population, energy demand is doubtlessly rising. To protect the environment, it is increasingly important to replace non-renewable energy sources with renewable over time.

This poses the question: what caused this increase in clean energy start-ups creation and what will. The hypotheses of this report introduced in the introduction section provides a basis for answering this question. Further evidence and analyzing the findings of this report are provided in the discussion section.

Methods

Four open source dataset were used in this report. The creation, funding and success dataset, which represented three stages of start-ups, from the IDA and the world energy consumption dataset from Kaggle. These datasets were imported to R studio to perform an Exploratory Data Analysis (EDA) and further machine learning technique.

An initial EDA was conducted on the datasets to explore the general patterns in the data, especially identifying outliers and features. This include visualizing the distribution of each energy source to provide a potential insight for the fitted model and examining the result of the prediction.

For all three energy start-up datasets, there were some common approaches of data cleaning that needed to be done before EDA. First, convert the data types of all the categorical variables into factor, then clean the dataset by omitting “NA” value and rename the column. Select all the types of energy that related to clean and traditional energy, the types were as following: "Bioenergy", "Electric mobility", "Energy efficiency", "Energy storage and batteries", "Fossil fuels", "Grid", "Hydrogen and fuel cells", "Nuclear", "Other energy", "Other renewables", "Solar" and "Wind".

There were two major indicators in the funding datasets, percentage of receive funding and the amount of funding a start-up had received. Visualize a boxplot to illustrate a relationship between a categorical variable, “technology or sector” and a numeric variable, “probability of funding”. Following with a dotted scatterplot to show the amount of funding received by each energy start-up.

There were also two indicators in the success datasets, probability of exiting the start-up stage in different channels and the year a start-up took to exit. Showing the distributions of numeric data value, “Probability of exit”, between categorical data, “Indictor”.

There were 122 columns in the world energy dataset with a lot of unnecessary variables. First, select the variable that could potentially be significant for regression. The following possible predictors had been selected: “country”, “year”, “population”, “gdp”, “biofuel_consumption”, “hydro_consumption”, “nuclear_consumption”, “solar_consumption”, “wind_consumption”, “other_renewable_consumption”, “coal_consumption”, “gas_consumption”, “oil_consumption”. Distinguish the clean and non-clean energy consumption and create a new column to sum the total amount of both energy consumption. Train the data and use recipe() in machine learning to perform a multiple linear regression and evaluate the prediction using adjected R-squared.

Linear Regression Analysis was undertaken to explore the potential factor influencing the chance of receive funding, while another linear regression was used to assess the potential factor that affect the probability of being success and the year it takes to success. Thus, these models were used to determine the accuracy of the first hypothesis. Multiple Linear Regression was used after some EDA on the distribution of the variable.

Two models were used to test the first hypothesis: H1M1, H1M2

H1M1: *Prob of funding = Country + technology and sector + funding stage + year + number of observation*

H1M2: *Years of success = Country + technology and sector*

Three models were used to assess the relationship between the numeric data of energy start-ups and the categorical variable and therefore regression was used instead of classification. To secure the accuracy of the model, all predictors were test at the 5% significance level and only predictors with p-value lower than 0.05 would be considered as an effective predictor and be included in the model.

The model H2M1, H2M2, H2M3 were used to test the second hypothesis:

H2M1: *Consumption of renewable energy = year + population + gdp + total*
H2M2: *Consumption of renewable energy = country + gdp + non renewable*
H2M3: *Consumption of renewable energy = year + population + country + gdp + total + non-renewable*

The model were tested by Adjusted R-squared (adj-R2) and Root-mean-square deviation (RMSE). R-squared is a statistical measurement in regression that determines the proportion of variance in the dependent variable that can be explained by the independent variable. It explains how well the data fit the regression model, where adj-R2 greater than 0.7 or 0.8 is considered good fits but not a hard rule. RMSE is a main performance indicator for a regression model which measures the average difference between value predicted by a model and the actual value. It provides estimation of how well the model is able to predict the outcome with lower RMSE is considered as the better model.

The following is an explanation of the key variables used in the energy start-up dataset:

- **Prob_of_funding**: The probability to receive funding
- **Country**: Include data of 213 different countries
- **Technology_or_sector**: Types of energy that the start-up chooses to form
- **Funding_stage**: Include early stage, later stage and all stage of receive funding
- **Year**: Record data from 2000 to 2022
- **Number_of_observation**: The total number of observation by multiplying value and observation

The following is an explanation of the key variables used in the energy start-up dataset:

- **Consumption_of_renewable_energy**: The total number of clean energy consumption for each subject observed
- **Year**: Include data from 1965 to 2019
- **Population**: The number of people given the country
- **Gdp**: The gross domestic product given the country
- **Country**: Include data of 83 different countries
- **Total**: The total amount of energy consumption for that observed country
- **Non-renewable**: The amount of non-renewable consumption for that observed country

Observations & Results

Initial EDA

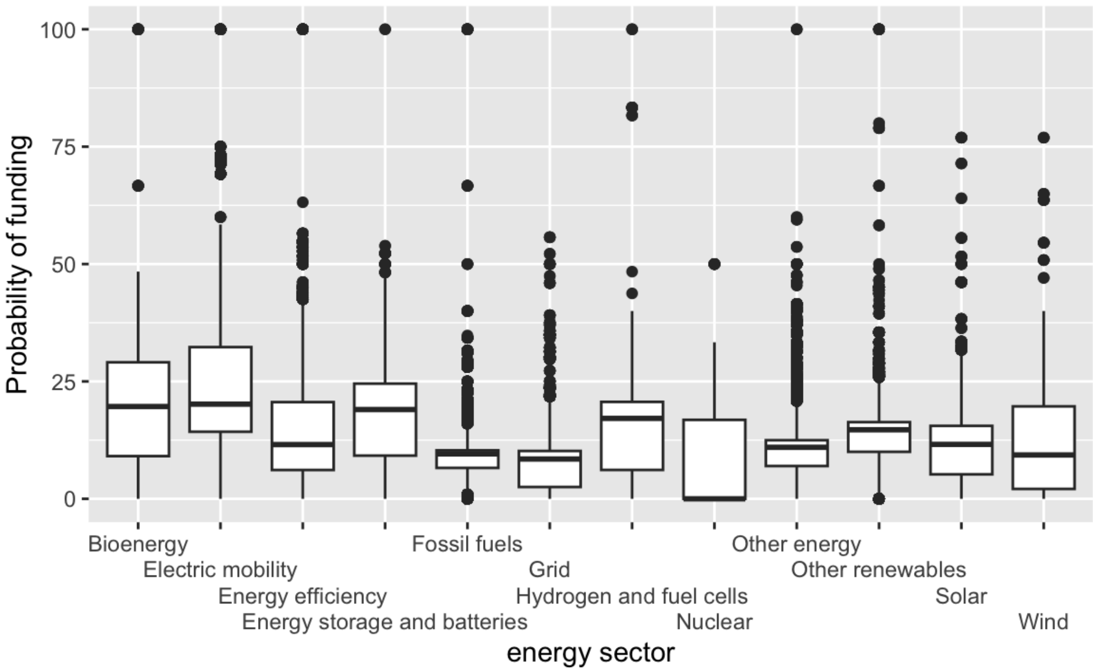


Figure 5: Side-By-Side boxplot showing the probability of secure funding for each energy start-ups

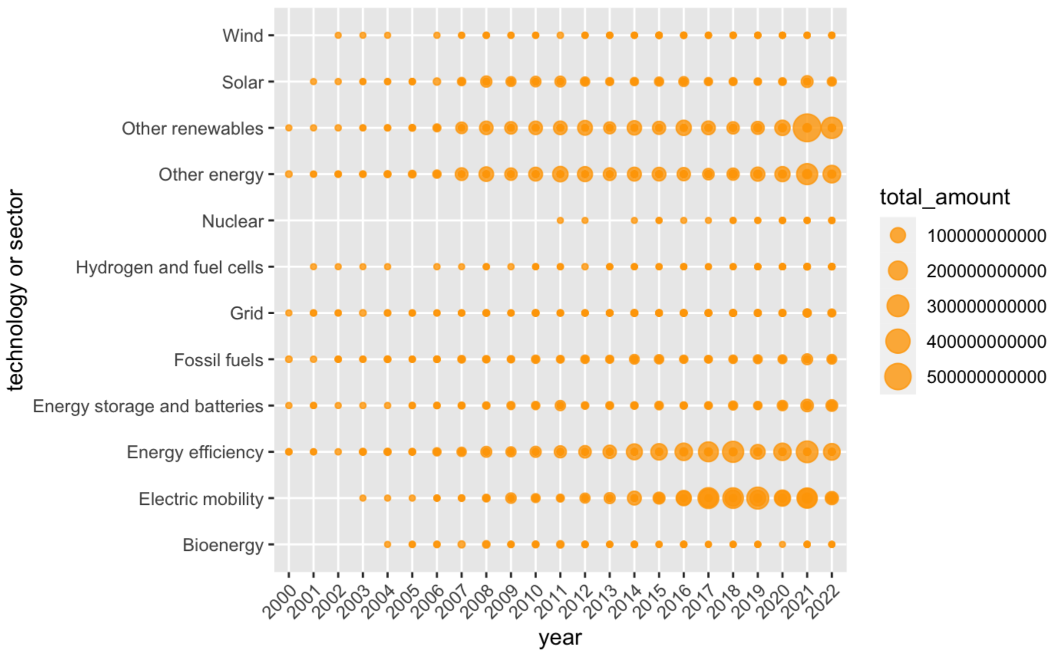


Figure 6: A dotted plot showing the amount of funding received by each energy start-ups from 2000 to 2022

Obtain funding for a start-up is crucial for growing the network and expanding the scale of the business. Usually, governments and larger firms provide funding for promising start-ups with a considerable business future. Investing in clean energy start-up helps cater the demand of promoting a green future. Generally, energy start-ups in the form of bioenergy, hydropower and electric vehicle can have a better chance of secure funding and renewable energy can receive a larger funding comparing to traditional energy. **(Figure 5 and Figure 6)**

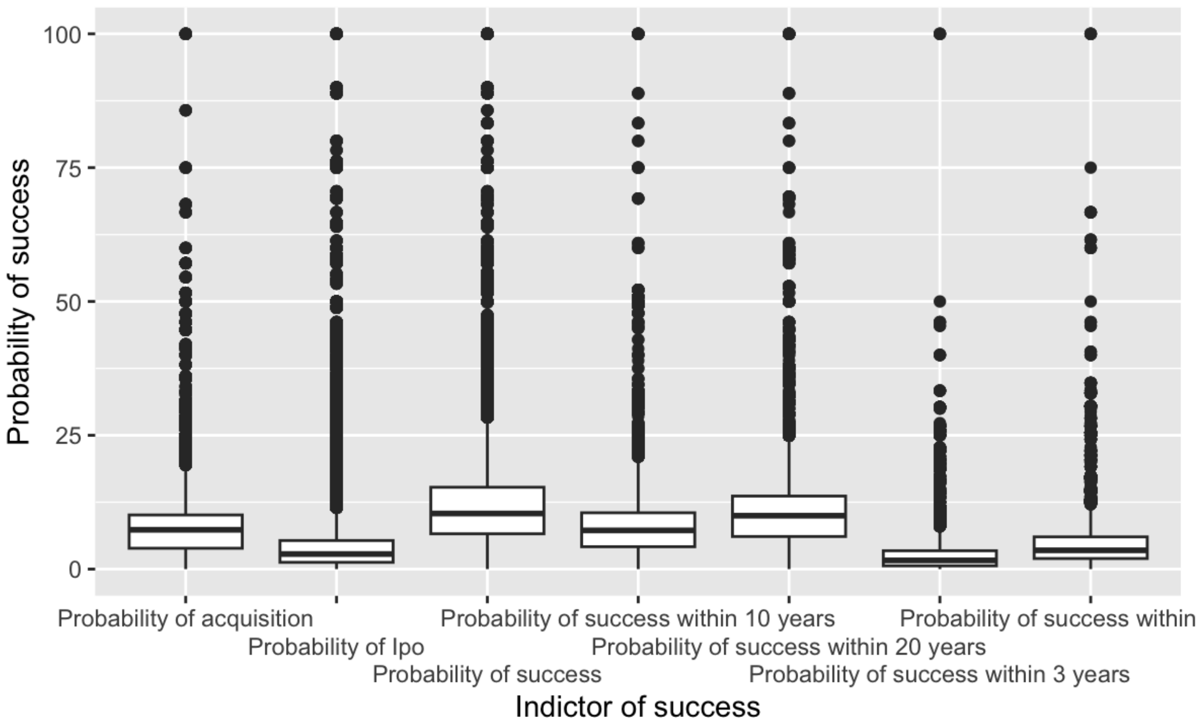


Figure 7: Side-By-Side boxplot showing the probability of succussing in different ways

The ultimate objective for a start-up is to transition from the initial start-up stage to either being acquired by other cooperation or launching an IPO. In the case of an energy start-up, achieving success often occurs within 20 years. **(Figure 7)**

Hypothesis 1:

Hypothesis 1 insists that clean energy start-ups often have significant advantages when it comes to secure funding with a higher funding amount and enjoy lower time for reaching success. H1M1 examined the estimated effect that each variable had on the probability of secure funding, following all the variable were significant in influencing the chance of receive funding. Meanwhile H1M2 examined the estimated effect that each variable had on the year that an energy start-up had to take to reach success, knowing some countries and energy sectors were not significant.

H1M1:

```
Call:
lm(formula = value ~ ., data = dk_percent)

Residuals:
   Min     1Q Median     3Q    Max
-47.69  -2.77  -0.09   1.98  96.78

Coefficients:
```

```
country_nameFINLAND                          13.573919   1.062644   12.77 < 0.0000000000000002 ***
country_nameFRANCE                            7.066261   1.054400    6.70   0.00000000002063510 ***
country_nameGEORGIA                          -4.321791   1.489360   -2.90               0.00371 **
country_nameGERMANY                           3.835482   1.054483    3.64               0.00028 ***
```

```
 country_nameSTATE OF PALESTINE               43.886176   2.978256   14.74 < 0.0000000000000002 ***
```

```
country_nameUNITED KINGDOM                    7.440136   1.053770    7.06   0.00000000000166236 ***
country_nameUNITED STATES                    13.451158   1.054187   12.76 < 0.0000000000000002 ***
```

```
 technology_or_sectorElectric mobility          4.532186   0.156108   29.03 < 0.0000000000000002 ***
 technology_or_sectorEnergy efficiency         -5.109183   0.152886  -33.42 < 0.0000000000000002 ***
 technology_or_sectorEnergy storage and batteries  -1.846235   0.160009  -11.54 < 0.0000000000000002 ***
 technology_or_sectorFossil fuels             -10.809627   0.152893  -70.70 < 0.0000000000000002 ***
 technology_or_sectorGrid                     -11.395600   0.154961  -73.54 < 0.0000000000000002 ***
 technology_or_sectorHydrogen and fuel cells   -4.906816   0.236663  -20.73 < 0.0000000000000002 ***
 technology_or_sectorNuclear                  -11.178193   0.255817  -43.70 < 0.0000000000000002 ***
 technology_or_sectorOther energy              -7.868489   0.154219  -51.02 < 0.0000000000000002 ***
 technology_or_sectorOther renewables          -5.061789   0.152382  -33.22 < 0.0000000000000002 ***
 technology_or_sectorSolar                     -8.653945   0.153472  -56.39 < 0.0000000000000002 ***
 technology_or_sectorWind                      -5.726715   0.194081  -29.51 < 0.0000000000000002 ***
 funding_stageLater stage                      -5.502329   0.019238 -286.02 < 0.0000000000000002 ***
 year2016                                       0.369893   0.038691    9.56 < 0.0000000000000002 ***
 year2017                                       0.725459   0.038102   19.04 < 0.0000000000000002 ***
 year2018                                       0.868110   0.037702   23.03 < 0.0000000000000002 ***
 year2019                                       0.997415   0.037456   26.63 < 0.0000000000000002 ***
 year2020                                       1.138902   0.037322   30.52 < 0.0000000000000002 ***
 year2021                                       1.199179   0.037240   32.20 < 0.0000000000000002 ***
 number_of_observations                        -0.001334   0.000028  -47.61 < 0.0000000000000002 ***
 ---
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Residual standard error: 5.57 on 335334 degrees of freedom
 Multiple R-squared:  0.563,	Adjusted R-squared:  0.563
 F-statistic: 2.83e+03 on 153 and 335334 DF,  p-value: <0.0000000000000002
```

Figure 8: Regression output for model H1M1

H1M2:

```
Call:
lm(formula = value ~ technology_or_sector + country_name, data = dt_years)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.4463 -0.4085  0.0409  0.3629  6.6905 
```

```
Coefficients:
                                              Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                    5.12895    0.14062  36.474  < 2e-16 ***
technology_or_sectorElectric mobility         -0.01784    0.13699  -0.130  0.89636    
technology_or_sectorEnergy efficiency         -0.03146    0.13389  -0.235  0.81422    
technology_or_sectorEnergy storage and batteries  0.11370    0.14020   0.811  0.41738    
technology_or_sectorFossil fuels              -1.08008    0.13461  -8.024 1.07e-15 ***
technology_or_sectorGrid                      -0.05304    0.13723  -0.386  0.69915    
technology_or_sectorHydrogen and fuel cells    0.53674    0.19367   2.771  0.00559 ** 
technology_or_sectorNuclear                   -2.84927    0.25922 -10.991  < 2e-16 ***
technology_or_sectorOther energy              -0.64103    0.13371  -4.794 1.64e-06 ***
technology_or_sectorOther renewables           0.19823    0.13424   1.477  0.13977    
technology_or_sectorSolar                      0.05055    0.13591   0.372  0.70993    
technology_or_sectorWind                       1.31404    0.17524   7.498 6.69e-14 ***
country_nameaustria                            1.76721    0.17899   9.873  < 2e-16 ***
country_namebangladesh                         2.99219    0.38095   7.854 4.18e-15 ***
country_namebelgium                            4.33725    0.08982  48.290  < 2e-16 ***
country_namebermuda                           -1.40574    0.65670  -2.141  0.03232 *  
country_namebrazil                             0.52310    0.08449   6.191 6.06e-10 ***
country_namecanada                             0.43590    0.05360   8.132 4.43e-16 ***
country_namechile                             -0.92360    0.18727  -4.932 8.20e-07 ***
country_namedenmark                            2.92763    0.09589  30.532  < 2e-16 ***
country_nameegypt                              4.14495    0.31809  13.031  < 2e-16 ***
country_nameestonia                           -0.08712    0.30682  -0.284  0.77647    
country_namefinland                            1.17649    0.11467  10.259  < 2e-16 ***
country_namefrance                             2.67811    0.05780  46.332  < 2e-16 ***
country_namegermany                            1.83014    0.05794  31.587  < 2e-16 ***
country_namegreece                             6.72303    0.25793  26.066  < 2e-16 ***
country_namehong kong (china)                 -2.04832    0.13624 -15.034  < 2e-16 ***
country_nameindia                              2.32584    0.05885  39.522  < 2e-16 ***
country_nameindonesia                         -4.62900    0.38113 -12.145  < 2e-16 ***
country_nameireland                            2.94461    0.10480  28.096  < 2e-16 ***
country_nameisrael                             0.73287    0.07683   9.539  < 2e-16 ***
country_nameitaly                              0.87207    0.07160  12.180  < 2e-16 ***
country_namejapan                              0.05634    0.07398   0.762  0.44634    
country_namekenya                              3.43694    0.20588  16.694  < 2e-16 ***
country_namekorea                              3.70763    0.15347  24.159  < 2e-16 ***
country_namemexico                            -0.84840    0.12915  -6.569 5.16e-11 ***
country_namenetherlands                        1.78544    0.05909  30.214  < 2e-16 ***
country_namenew zealand                        1.29563    0.22732   5.700 1.21e-08 ***
country_namenigeria                            4.17125    0.21940  19.012  < 2e-16 ***
country_namenorway                             1.59679    0.08427  18.948  < 2e-16 ***
country_namepeople's republic of china         2.68294    0.05716  46.935  < 2e-16 ***
country_namepoland                             1.22819    0.11040  11.124  < 2e-16 ***
country_namerussian federation                 2.04377    0.12017  17.007  < 2e-16 ***
country_namesingapore                          1.77630    0.09264  19.175  < 2e-16 ***
country_namesouth africa                      -1.66130    0.16246 -10.226  < 2e-16 ***
country_namespain                              4.88988    0.06095  80.225  < 2e-16 ***
country_namesweden                             1.82713    0.07636  23.927  < 2e-16 ***
country_nameswitzerland                        1.47506    0.06759  21.824  < 2e-16 ***
country_nametaiwan                             3.10632    0.14607  21.266  < 2e-16 ***
country_nameturkey                            -1.14473    0.20603  -5.556 2.79e-08 ***
country_nameunited arab emirates               1.69802    0.13486  12.591  < 2e-16 ***
country_nameunited kingdom                     1.99487    0.05159  38.668  < 2e-16 ***
country_nameunited states                      1.90969    0.04656  41.017  < 2e-16 ***
country_nameviet nam                           0.11939    0.43152   0.277  0.78203    
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.135 on 24028 degrees of freedom
Multiple R-squared:  0.4838,	Adjusted R-squared:  0.4826 
F-statistic: 424.8 on 53 and 24028 DF,  p-value: < 2.2e-16
```

Figure 9: Regression output for model H1M2

The p-value of both linear regressions were lower than 0.05 which indicated the null hypothesis had to be rejected and the claim could be accepted. Since only the outcome variable was numeric data, fitted equation was not ideal. The interpretation was:
From H1M1, the chance of funding a fossil fuel start-up is lower than those clean energy especially the eclectic vehicle category.
From H1M2, the time it takes to reach success are not necessarily faster for clean energy start-up compared to fossil fuel start-up.
The adjusted r squared for H1M1 is 0.563 and 0.4826 for H1M2. The interpretation of the model is clean energy start-ups can have a higher chance of secure funding and also receive a larger amount that allow them to continue expanding them business **(Figure 6)**. Although the time to success may not be faster as a clean energy start-up compared to a traditional energy start-up, the overall rate of success as an energy start-up is not high, despite the type of energy **(Figure 7)**. Therefore, Hypothesis 1 was not supported as there were zero evidence that clean energy was easier to achieve success in any means.

Hypothesis 2

After performing linear regression on the model H2M1, H2M2 and H2M3 for Hypothesis 2., the best model was determined based on the adj-R2 value.

A tibble: 3 x 13

r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC	deviance	df.residual	nobs	model
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<int>	<chr>
0.9737032	0.9736755	2.488733e+02	3.515754e+04	0	4	-26374.66	52761.32	52798.78	2.352401e+08	3798	3803	Model 1
0.9816159	0.9812157	2.102306e+02	2.452862e+03	0	81	-25694.00	51554.00	52072.21	1.644566e+08	3721	3803	Model 2
1.0000000	1.0000000	5.890722e-12	3.068965e+30	0	84	92983.37	-185794.75	-185257.80	1.290169e-19	3718	3803	Model 3

Figure 10: Model Performance matrics in Hypothesis 2

As shown in Figure 10, the model with highest adj-R2 value was model 3, which represented H2M3, therefore, H2M3 was chosen as the best model for prediction.

```
Call:
lm(formula = renewable ~ year + population + gdp + total + not_renewable +
    country, data = dg_consumption)

Residuals:
       Min         1Q     Median         3Q        Max
-1.743e-10 -1.580e-13  6.000e-15  1.650e-13  2.928e-11
```

```
Coefficients:
                          Estimate Std. Error    t value Pr(>|t|)    
(Intercept)             -5.905e-11  7.723e-12 -7.646e+00 2.62e-14 ***
year                     2.743e-14  3.867e-15  7.094e+00 1.56e-12 ***
population              -5.587e-21  1.355e-21 -4.123e+00 3.82e-05 ***
gdp                     -2.388e-24  7.073e-26 -3.377e+01  < 2e-16 ***
total                    1.000e+00  2.977e-16  3.359e+15  < 2e-16 ***
not_renewable           -1.000e+00  3.144e-16 -3.181e+15  < 2e-16 ***
countryArgentina         5.266e-12  6.687e-13  7.875e+00 4.44e-15 ***
countryAustralia         5.434e-12  6.706e-13  8.103e+00 7.18e-16 ***
countryAustria           5.245e-12  6.702e-13  7.826e+00 6.51e-15 ***
countryAzerbaijan        5.242e-12  7.622e-13  6.878e+00 7.11e-12 ***
countryBangladesh        4.928e-12  6.992e-13  7.049e+00 2.14e-12 ***
countryBelarus           5.265e-12  7.622e-13  6.907e+00 5.81e-12 ***
countryBelgium           5.328e-12  6.703e-13  7.948e+00 2.49e-15 ***
countryBrazil            4.973e-12  6.826e-13  7.286e+00 3.87e-13 ***
countryBulgaria          5.292e-12  6.694e-13  7.906e+00 3.48e-15 ***
countryCanada            5.571e-12  7.235e-13  7.700e+00 1.74e-14 ***
countryChile             5.284e-12  6.691e-13  7.896e+00 3.76e-15 ***
countryChina             3.503e-12  1.340e-12  2.615e+00  0.00897 ** 
countryColombia          5.240e-12  6.689e-13  7.833e+00 6.15e-15 ***
countryCroatia           5.180e-12  8.036e-13  6.446e+00 1.30e-10 ***
countryCyprus            5.293e-12  6.693e-13  7.908e+00 3.41e-15 ***
countryCzechia           5.317e-12  6.863e-13  7.747e+00 1.20e-14 ***
countryDenmark           5.322e-12  6.693e-13  7.952e+00 2.41e-15 ***
countryEcuador           5.271e-12  6.690e-13  7.879e+00 4.29e-15 ***
countryEgypt             5.234e-12  6.695e-13  7.817e+00 6.97e-15 ***
countryEstonia           5.302e-12  7.624e-13  6.954e+00 4.16e-12 ***
countryFinland           5.310e-12  6.704e-13  7.921e+00 3.09e-15 ***
countryFrance            5.373e-12  6.981e-13  7.697e+00 1.77e-14 ***
countryGermany           5.677e-12  6.794e-13  8.356e+00  < 2e-16 ***
countryGreece            5.307e-12  6.690e-13  7.932e+00 2.82e-15 ***
countryHong Kong         5.311e-12  6.691e-13  7.938e+00 2.70e-15 ***
countryHungary           5.313e-12  6.692e-13  7.939e+00 2.67e-15 ***
countryIceland           5.294e-12  6.696e-13  7.906e+00 3.46e-15 ***
countryIndia             3.043e-12  1.261e-12  2.413e+00  0.01589 *  
countryIndonesia         4.866e-12  6.953e-13  6.998e+00 3.08e-12 ***
countryIran              5.310e-12  6.700e-13  7.926e+00 2.98e-15 ***
countryIraq              5.282e-12  6.687e-13  7.899e+00 3.67e-15 ***
countryIreland           5.362e-12  6.692e-13  8.012e+00 1.50e-15 ***
countryIsrael            5.296e-12  6.691e-13  7.914e+00 3.26e-15 ***
countryItaly             5.371e-12  6.700e-13  8.017e+00 1.44e-15 ***
countryJapan             5.653e-12  6.790e-13  8.326e+00  < 2e-16 ***
countryKazakhstan        5.311e-12  7.626e-13  6.964e+00 3.90e-12 ***
countryKuwait            5.332e-12  7.017e-13  7.599e+00 3.76e-14 ***
countryLatvia            5.250e-12  7.624e-13  6.886e+00 6.70e-12 ***
countryLithuania         5.280e-12  7.626e-13  6.924e+00 5.16e-12 ***
countryLuxembourg        5.306e-12  6.693e-13  7.928e+00 2.92e-15 ***
countryMalaysia          5.268e-12  6.688e-13  7.876e+00 4.39e-15 ***
countryMexico            5.192e-12  6.702e-13  7.746e+00 1.21e-14 ***
countryMorocco           5.249e-12  6.687e-13  7.850e+00 5.39e-15 ***
countryNetherlands       5.369e-12  6.702e-13  8.012e+00 1.50e-15 ***
countryNew Zealand       5.320e-12  6.701e-13  7.939e+00 2.68e-15 ***
countryNorth Macedonia   5.210e-12  8.036e-13  6.483e+00 1.02e-10 ***
countryNorway            5.250e-12  6.767e-13  7.759e+00 1.10e-14 ***
countryOman              5.302e-12  6.693e-13  7.922e+00 3.05e-15 ***
countryPakistan          5.020e-12  6.779e-13  7.406e+00 1.60e-13 ***
```

```
countryPeru                  5.291e-12  6.689e-13  7.911e+00 3.33e-15 ***
countryPhilippines           5.152e-12  6.705e-13  7.684e+00 1.96e-14 ***
countryPoland                5.341e-12  6.705e-13  7.966e+00 2.16e-15 ***
countryPortugal              5.301e-12  6.692e-13  7.922e+00 3.06e-15 ***
countryQatar                 5.278e-12  7.018e-13  7.521e+00 6.77e-14 ***
countryRomania               5.256e-12  6.691e-13  7.855e+00 5.20e-15 ***
countryRussia                6.847e-12  8.385e-13  8.166e+00 4.33e-16 ***
countrySaudi Arabia          5.428e-12  6.714e-13  8.084e+00 8.39e-16 ***
countrySingapore             5.359e-12  6.695e-13  8.003e+00 1.60e-15 ***
countrySlovakia              5.247e-12  7.627e-13  6.879e+00 7.02e-12 ***
countrySlovenia              5.279e-12  8.038e-13  6.567e+00 5.84e-11 ***
countrySouth Africa          5.316e-12  6.696e-13  7.939e+00 2.68e-15 ***
countrySouth Korea           5.357e-12  6.702e-13  7.992e+00 1.75e-15 ***
countrySpain                 5.288e-12  6.700e-13  7.893e+00 3.84e-15 ***
countrySri Lanka             5.246e-12  6.688e-13  7.844e+00 5.64e-15 ***
countrySweden                5.248e-12  6.773e-13  7.749e+00 1.19e-14 ***
countrySwitzerland           5.277e-12  6.713e-13  7.860e+00 4.98e-15 ***
countryTaiwan                5.330e-12  6.694e-13  7.962e+00 2.23e-15 ***
countryThailand              5.219e-12  6.692e-13  7.799e+00 8.03e-15 ***
countryTrinidad and Tobago   5.332e-12  6.693e-13  7.966e+00 2.17e-15 ***
countryTurkey                5.151e-12  6.689e-13  7.701e+00 1.73e-14 ***
countryTurkmenistan          5.290e-12  7.624e-13  6.939e+00 4.65e-12 ***
countryUkraine               5.365e-12  7.667e-13  6.997e+00 3.09e-12 ***
countryUnited Arab Emirates  5.208e-12  8.359e-13  6.230e+00 5.17e-10 ***
countryUnited Kingdom        5.638e-12  6.730e-13  8.378e+00  < 2e-16 ***
countryUnited States         6.122e-12  1.121e-12  5.459e+00 5.10e-08 ***
countryUzbekistan            5.208e-12  7.622e-13  6.832e+00 9.71e-12 ***
countryVenezuela             5.250e-12  6.700e-13  7.836e+00 6.02e-15 ***
countryVietnam               5.133e-12  6.708e-13  7.652e+00 2.50e-14 ***
countryWorld                 4.423e-12  4.040e-12  1.095e+00  0.27370
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.322e-12 on 3718 degrees of freedom
  (481 observations deleted due to missingness)
Multiple R-squared:      1,     Adjusted R-squared:      1
F-statistic: 9.653e+30 on 84 and 3718 DF,  p-value: < 2.2e-16
```

Figure 11: Regression output of selected model H2M3

The p-value of the selected regression model H3M3 is smaller than 0.05 which indicated the reject of null hypothesis and accept the claim. With the Adjusted R-squared equaled to 1, it indicated a perfectly predicts value in the model. The model can be interpreted as: increase the year by one, the number of renewable energy consumption is increased by 2.743e-14 terawatt-hours. Indeed, the utility of clean energy consumption had increased in recent years **(Figure 3)**, M2M3 also suggested an upward trending of renewable energy consumption in the future, it is believed that surpassing traditional energy consumption was only the matter of time, supporting Hypothesis 2.

Discussions

Hypothesis 1 demonstrated that from **H1M1**, clean energy start-ups have a relatively higher chance of receive funding. In countries like, Finland, France, Palestine, United Kingdom and United States, energy start-up can have a better opportunity of being funded. It is contributed to the attitude and the resource that these countries willing to offer to the energy start-up in supporting the grow of the business. Moreover, with the past of the year, the probability of funding is having an stable increase which also implies the environment of start-up funding is indeed getting better with parties injecting capital to the industry. Moreover, the ultimate goal for a start-up is not only exit the start-up stage, most of the time, is to make profit to maintain the running of the business. The cost of opening a renewable energy start-up is disperse with the expectation to spend only $12 to even $27, 209 (Wall, 2023). Apart from the opening cost, which is a one-time cost, the ongoing cost needed to be taken into account for an entrepreneur. With the sky-rocketing demand for electric Vehicle, the probability of secure funding is also the highest among all the energy sectors, marking the only energy sector with a positive relationship with the probability of getting funded.

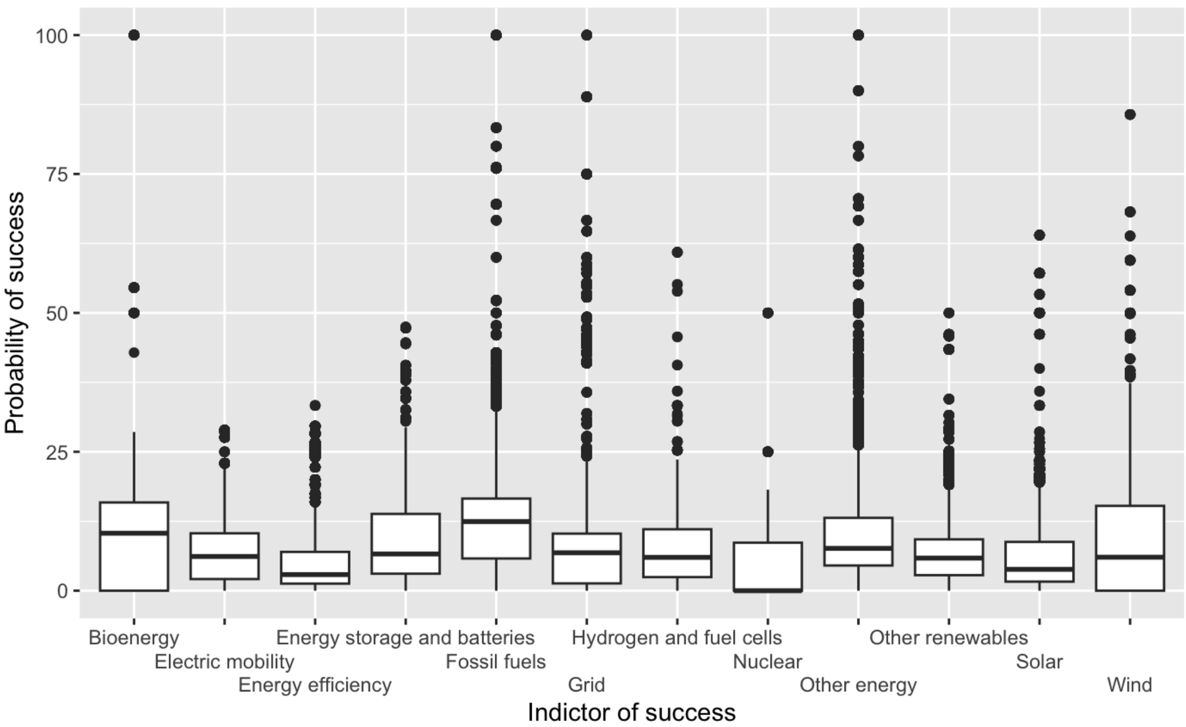


Figure 12: Side-by-side boxplot showing the probability of success in different energy sectors

However, on **H1M2**, the year of success are not guarantee by founding a clean energy start-up. By the Startup Failure Rate Statistics (2023), the failure rate for new startups is currently 90%, following with nearly 10% of new businesses do not survive the first year and the first-time start-up founders only have a success rate of 18%. It is never an easy task to attract investor to perform an acquisition or initial public offering, despite the energy type of the start-up **(Figure 12)**. Due to the limitation of the dataset, the relationship between receive funding and become success cannot be tested. However, according to Bill Gross, a legendary investor, there are five factors for start-up success: timing, team, idea, business model and funding. With an emphasis on funding, he suggested that funding should occur once an entrepreneur have determined the other factors, which make funding become an important predictor for the success of a start-up.

Hypothesis 2 recommended an increase in the usage of renewable energy in the future, with predicting the number with historical data. Notably, a big proportion of energy is used to produce electricity, and electricity is basically the major source of energy that an individual used every day. The upward trending in the renewable energy consumption has given a heart-strengthening shot for environmental protection **(Figure 4)**. Furthermore, all the countries are having a positive relationship with the renewable, that indicates a continue increase in renewable energy consumption for the future **(H2M3)**. According to the executive director of IEA, there are some foreseeable trends that are going to happen in the future: China continues to lead the way in renewable energy, rising cost can potentially hold back the future of renewable energy and energy storage is a growing market, etc. Some of these prediction match with the finding in this report, with the lack in category of the original data, there are lots of potential factors that have not been included, such as the total number of energy produced by each energy source.

Conclusion

The research findings offer several implications. Analyzing the visualization of energy start-up indicators derived from the energy start-up dataset, it has proven that the type of energy start-up significantly influences the likelihood of securing funding and plays a role in determining the level of financial support to some extent. Unfortunately, clean energy does not necessarily become an advantage in the success of an energy start-up compared to traditional energy. As a result, the hypothesis has been rejected unless there are more indicators explaining the relationship.

Undoubtedly, there is a clear shift in world's energy consumption toward renewable sources, with the public actively endorsing clean energy. This is not only reflected in the larger proportion of renewable energy utilization in the electricity consumption sector but also in the increasing number of clean energy start-ups, which could play a pivotal role in the future energy production sector. Hopefully the increase in clean energy start-ups and growing adoption of renewable energy can contribute to addressing present environment challenges and pave the way for a more sustainable future for the next generation.

References

IEA (2022). *Energy Start-up Data Explorer – Data Tools*. [online] IEA. Available at: https://www.iea.org/data-and-statistics/data-tools/energy-start-up-data-explorer [Accessed 6 Nov. 2023].

PRALABH , P. (2021). *World Energy Consumption*. [online] www.kaggle.com. Available at: https://www.kaggle.com/datasets/pralabhpoudel/world-energy-consumption/data [Accessed 6 Nov. 2023].

Howarth, J. (2022). *What Percentage of Startups Fail? 80+ Statistics (2022)*. [online] Exploding Topics. Available at: https://explodingtopics.com/blog/startup-failure-stats.

Walls, P. (2022). *How Much Does It Cost To Start A Renewable Energy Equipment And Devices Selling Business? (In 2023) - Starter Story*. [online] www.starterstory.com. Available at: https://www.starterstory.com/ideas/renewable-energy-equipment-and-devices-selling-business/startup-costs?successful_subscribe=true [Accessed 6 Nov. 2023].

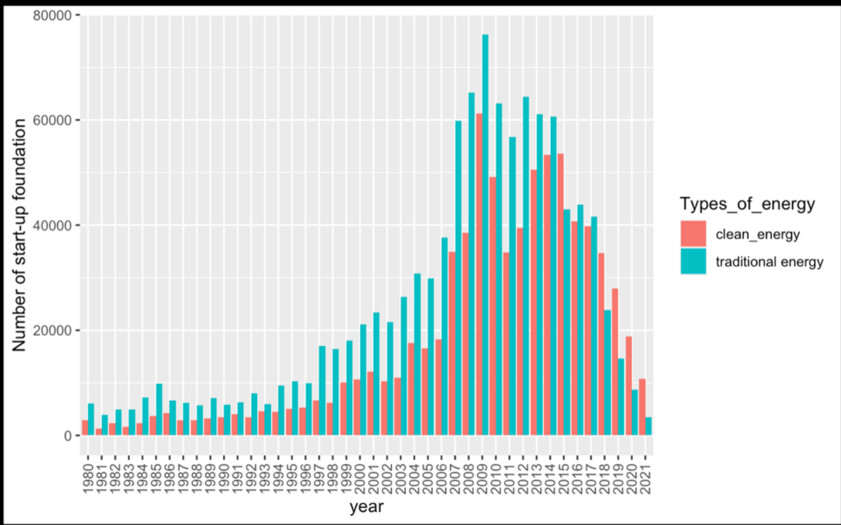
Bensley, T. (2022). *Startup success factors*. [online] Qonto - Blog. Available at: https://qonto.com/en/blog/creators/tools-tips/startup-success-factors.

profile (2022). *7 Predictions For The Future Of Renewable Energy*. [online] www.nesfircroft.com. Available at: https://www.nesfircroft.com/resources/blog/7-predictions-for-the-future-of-renewable-energy/.

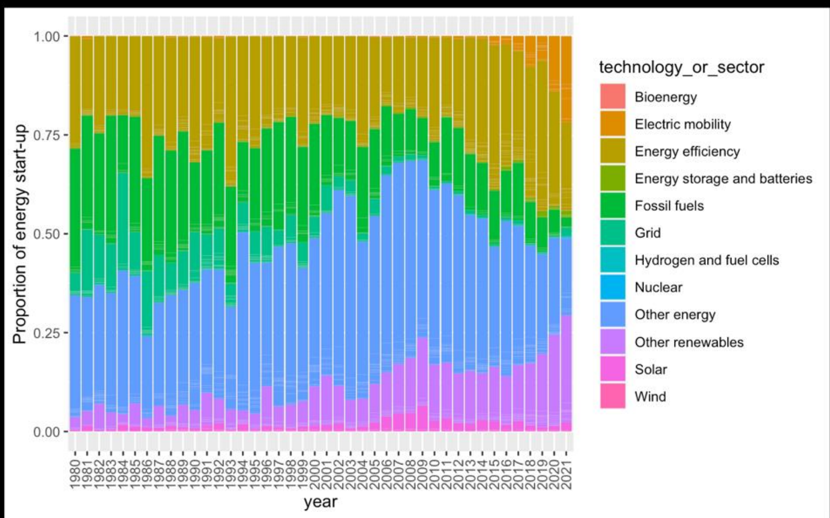
Appendix - Project code

Code for creation dataset

````
```{r}
sum_no |>
  ggplot(aes(x = year, y = Total_year, fill = Types_of_energy)) +
  geom_bar(stat = "identity", position = "dodge")+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  labs(y = "Number of start-up foundation")
```
````

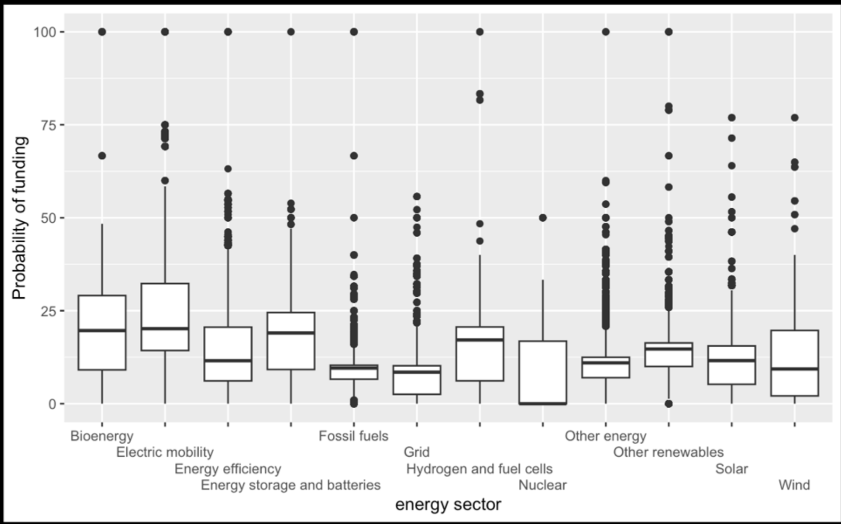



````
```{r}
df_f |>
  ggplot(aes(fill=technology_or_sector, y=total_observation, x=year)) +
  geom_bar(position = "fill", stat = "identity") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  labs(y = "Proportion of energy start-up")
```
````

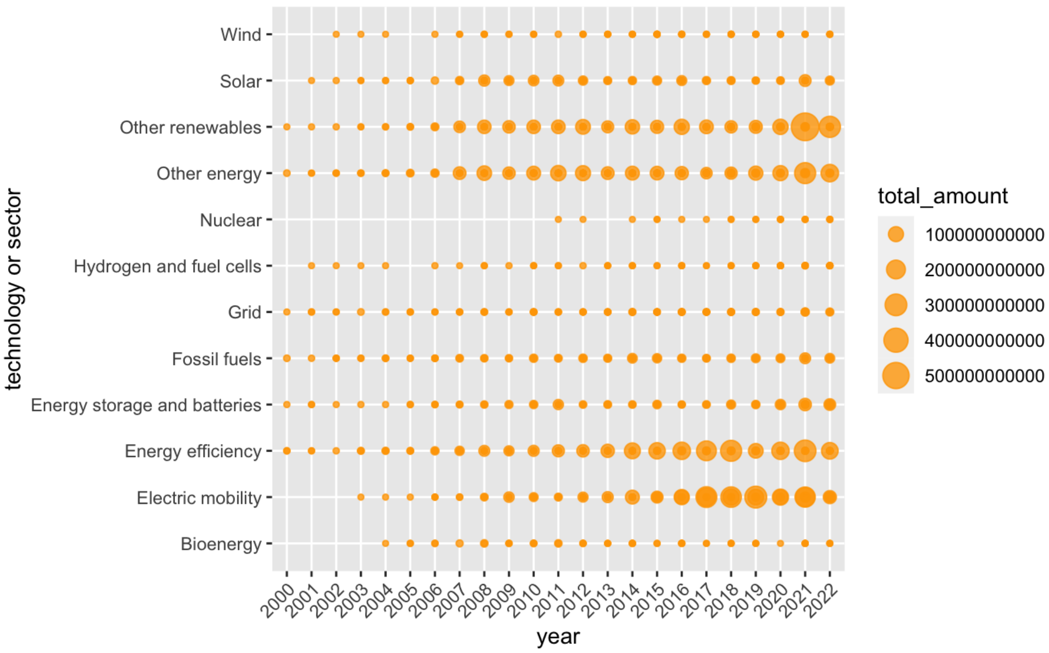


Code for funding dataset

```{r}
dk_percent |>
  ggplot(aes(x = technology_or_sector, y = value)) +
  geom_boxplot() +
  labs( y = "Probability of funding", x = "energy sector") +
  scale_x_discrete(guide = guide_axis(n.dodge=4))
```

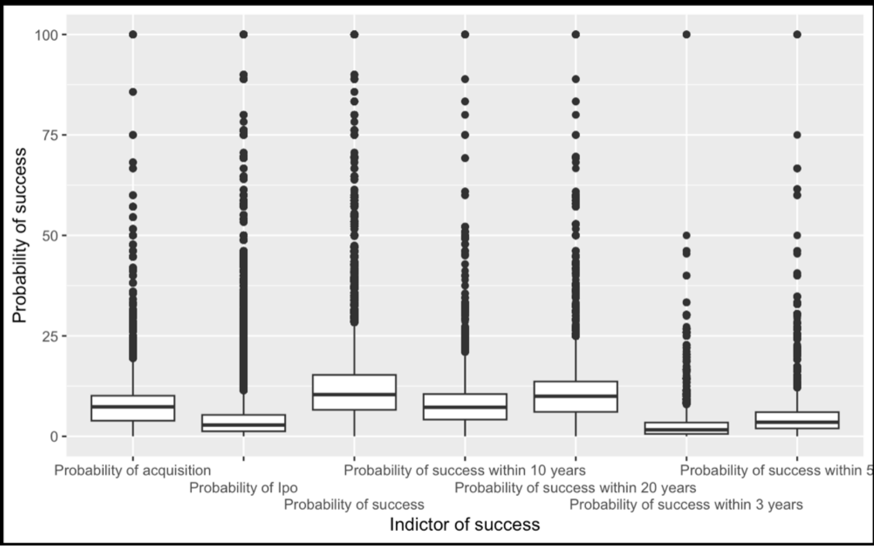



```{r}
dk_usd |>
  ggplot(aes(x = year, y = technology_or_sector, size = total_amount)) +
  geom_point(color = "orange", alpha = 0.8) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

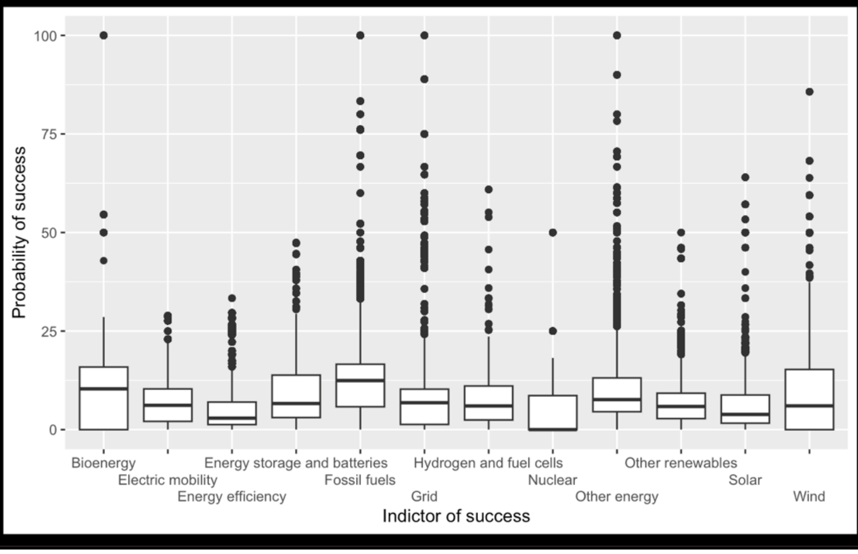


Code for success dataset

```{r}
dt_percent |>
  ggplot(aes(x = indicator, y = value)) +
  geom_boxplot() +
  scale_x_discrete(guide = guide_axis(n.dodge=3)) +
  labs( y = "Probability of success", x = "Indictor of success")
```

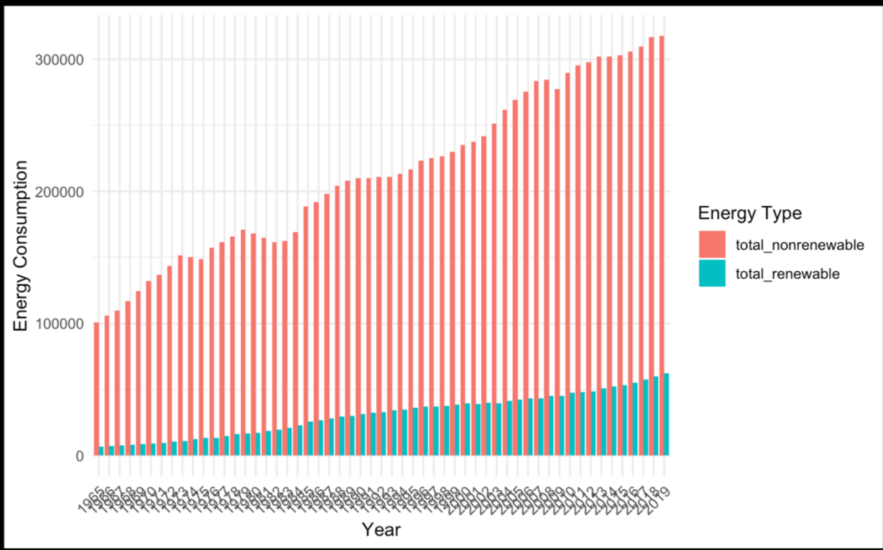



```{r}
dt_percent |>
  ggplot(aes(x = technology_or_sector, y = value)) +
  geom_boxplot() +
  scale_x_discrete(guide = guide_axis(n.dodge=3)) +
  labs( y = "Probability of success", x = "Indictor of success")
```

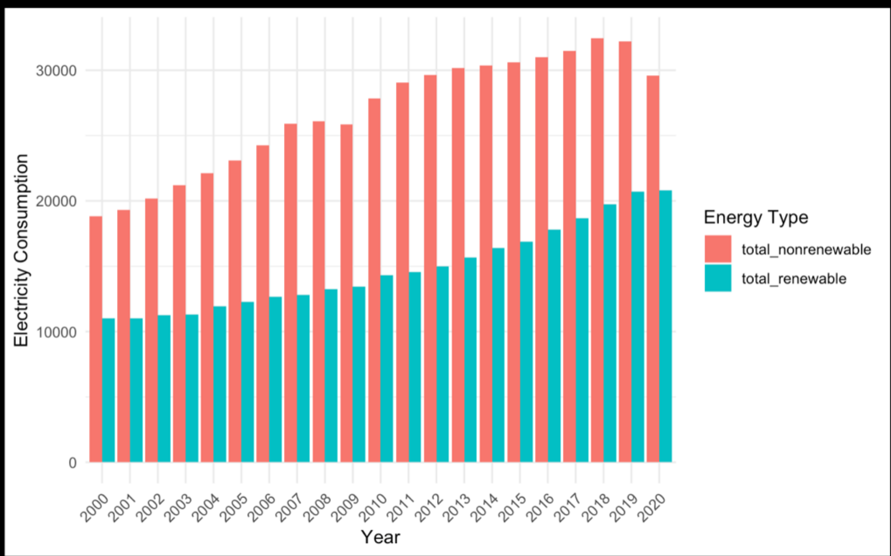


Code for World consumption dataset

```{r}
dg_consumption_1 |>
  pivot_longer(cols = 10:11, names_to = "indicator", values_to = "value") |>
  ggplot(aes(x = as.factor(year), y = value, fill = indicator)) +
        geom_bar(stat = "identity", position = "dodge") +
  labs(x = "Year", y = "Energy Consumption", fill = "Energy Type") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```{r}
dg_elec_1 |>
  pivot_longer(cols = 10:11, names_to = "indicator", values_to = "value") |>
  ggplot(aes(x = as.factor(year), y = value, fill = indicator)) +
        geom_bar(stat = "identity", position = "dodge") +
  labs(x = "Year", y = "Electricity Consumption", fill = "Energy Type") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```{r}
dg |>
  ggplot(aes(x = as.factor(year), y = population)) +
        geom_bar(stat = "identity", position = "dodge") +
  labs(x = "Year", y = "World population") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

⚠ Warning: [38;5;252mRemoved 1756 rows containing missing values (`geom_bar()`). [39m