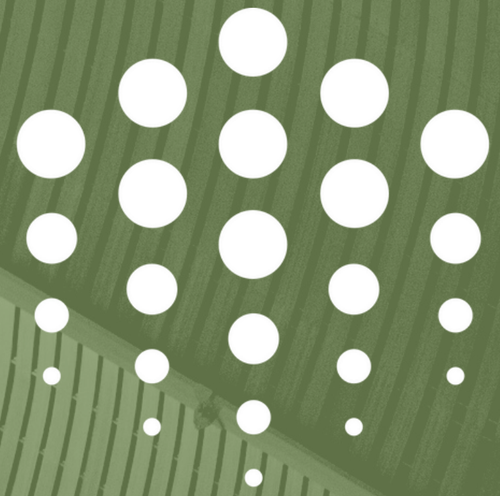


OCEAN PROTOCOL

STORAGE NETWORK ENERGY CONSUMPTION DATA CHALLENGE

by Nicolas Landry

OUTLINE OF THE REPORT

1. CLEANING AND SHAPING THE DATA
2. GLOBAL ANALYSIS
3. CORRELATIONS
4. MACHINE LEARNING ALGORITHM
5. ADDITIONAL DATA SET

Filecoin

1

CLEANING AND SHAPING THE DATA

DEFINITION OF THE "PRICE" OF $FIL

The "price" of filecoin that we use in our studies is based on the closing price for each day. This is because the high and low prices for the day are not relevant to our analysis. We focus on the closing price because it represents the most up-to-date information about the value of filecoin at the end of the trading day. By using only the closing price, we can accurately track the performance of filecoin over time and make more informed decisions about its value.

CHOICE OF THE ESTIMATE FOR THE ALGORITHMS

In the context of data related to energy, using estimates rather than lower and upper bounds can provide a more accurate and representative picture of the global situation. This is because estimates take into account a wide range of datacenters, providing a more comprehensive view of the data. Additionally, using estimates allows for better comparison and analysis of the data, as it eliminates the uncertainty and variability introduced by using lower and upper bounds. Overall, using estimates can provide a more accurate and meaningful representation of the data, making it a more useful input for algorithms and other analysis.

SYNCHRONISING THE DATA ON THE SAME TIMEFRAME

For the correlations and the global analysis, we had to remove some values from the $FIL price dataset to have a total dataset on the same timeframe.

GLOBAL ANALYSIS APPROACH

CONTEXT

Energy consumption in storage networks is a global issue because of the constant increase of digital transactions. With this in mind, Filecoin has chosen to use renewable energy for all its data centers and to create a measurement tool verifying the sustainability of IT infrastructures under the name of Filecoin Green.

Filecoin storage will be analysed following two axis : the energy consumption of data centers and its impact on the $fil token price.

INDICATORS

For the global analysis, 8 main indicators provided by the Filecoin Green Energy Consumption Dataset will be used find interesting insights related to the data storage capacity, share of renewable energy and energy consuption.

Filecoin

GRAPHICAL ANALYSIS

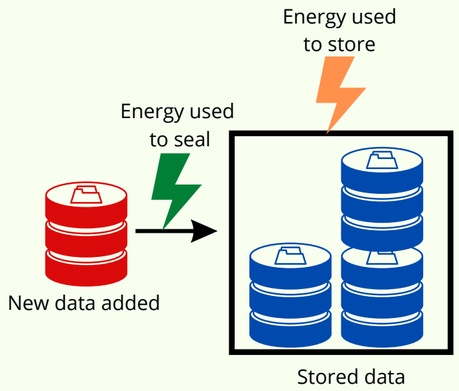
ENERGY PERFORMANCE & EVOLUTION OF FILECOIN'S STORAGE

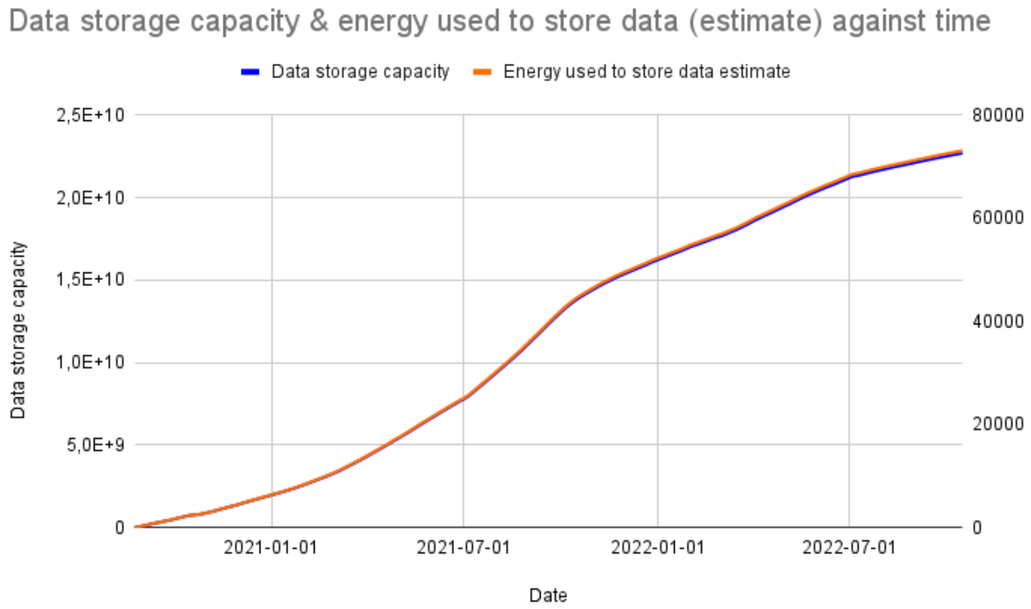
The energy consumption of Filecoin's installations has two components : the energy used to store the data, and the energy used to seal the data once it arrives into the storage center.

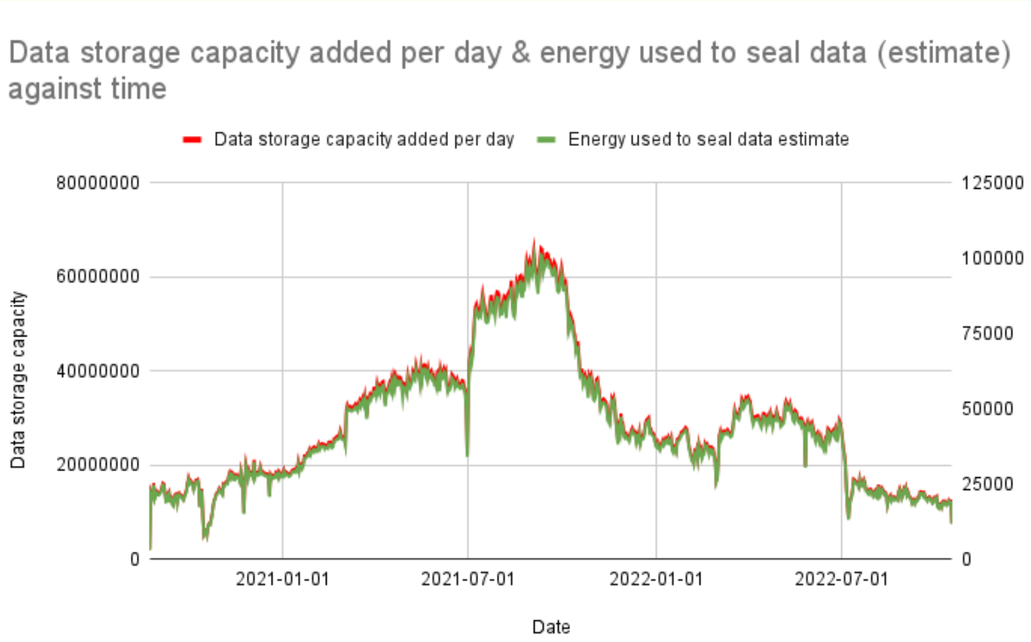
The energy used to store the data is completely proportional to the data storage capacity (graph 1). Indeed, each kW of stored data uses a certain amount of energy. As the amount of data stored is added day by day, the storage energy is multiplied by the total amount of data.

The energy used to seal the data is proportional to the data storage capacity added per day (graph2). Since data is only sealed once, the energy required to seal data per day is equivalent to the amount of data added per day multiplied by the energy required to seal one unit of data.

Please kindly see the "correlations" par for the calculation proof of these findings.







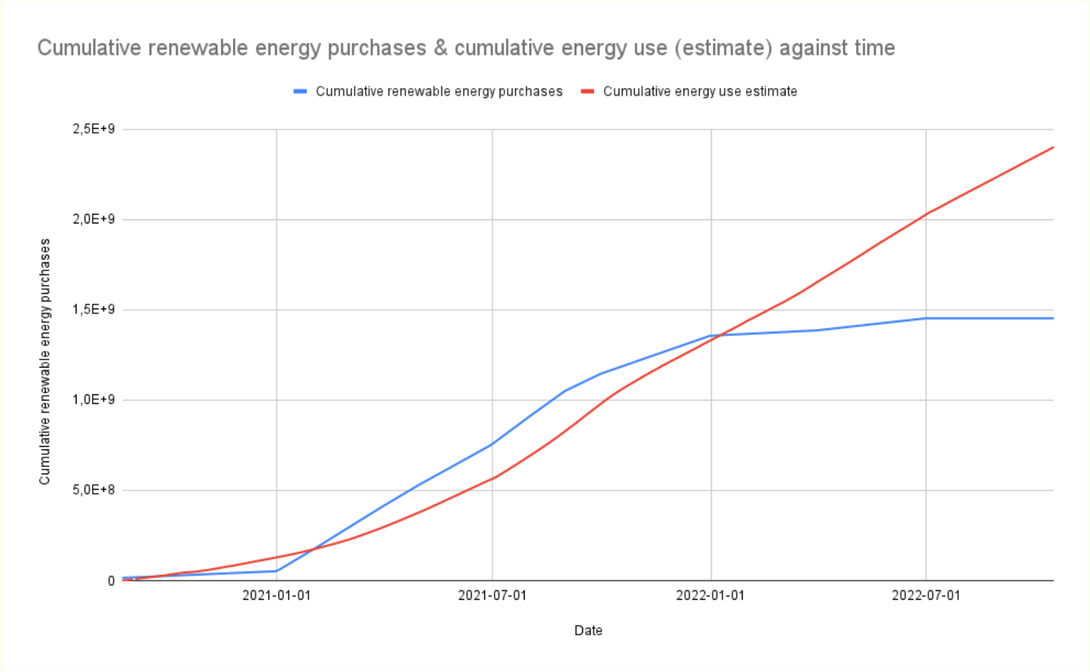
GRAPHICAL ANALYSIS

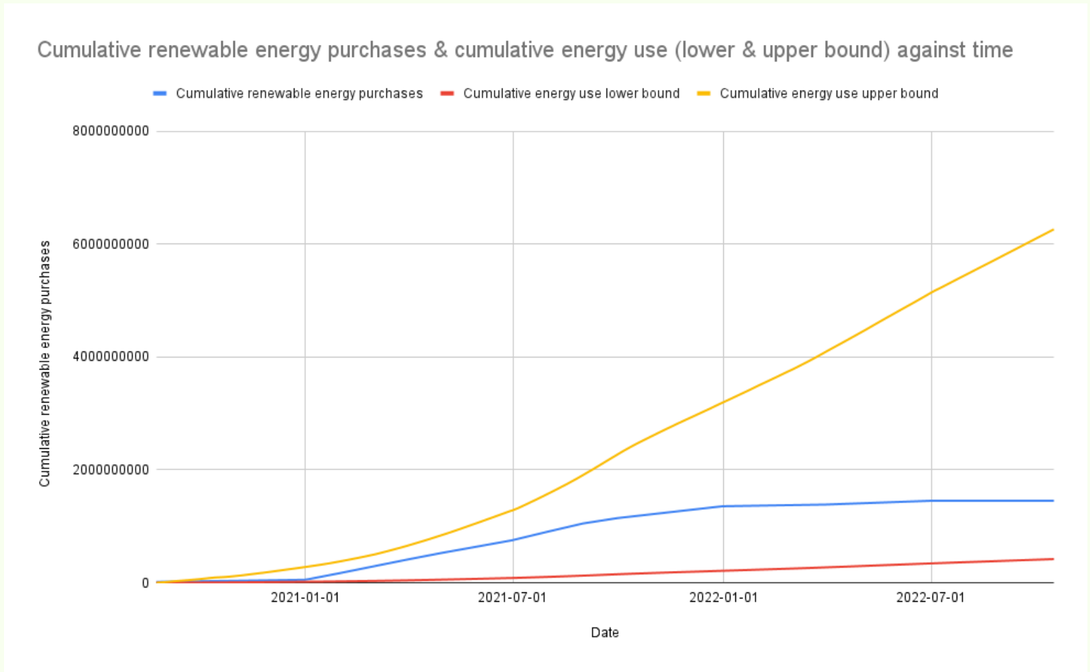
EVOLUTION OF THE SHARE OF RENEWABLE ENERGY USED BY FILECOIN

Filecoin buys renewable energy from generators in the region where the node is located in order to power all its data centers. It has also invested since 2021 to create its own green energy source : the built of new solar generation in the US.

The amount of energy used increases almost linearly (graph 1) as the amount of data stored also increases.However, the quantity of renewable energy purchased does not follow the same trend (graph 1) : from January 2022, its growth slows down sharply and becomes almost non-existent in July 2022. As a result, shortly after January 2022, the energy requirement exceeds the amount of renewable energy purchased. FIlecoin data centers must therefore be supplied with non-renewable energy.

It should be noted that the amount of energy purchased is always below the high bound and above the low bound of energy used(graph 2). This limits an over- or under-purchasing of energy.



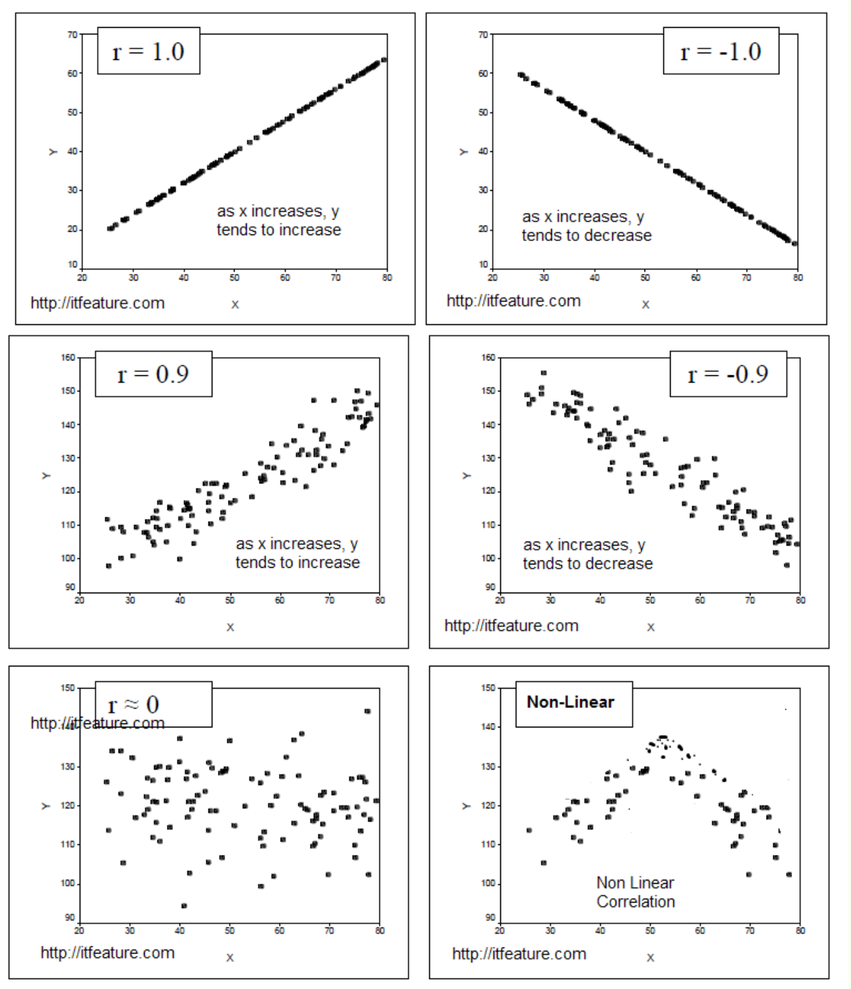


CORRELATIONS CALCULATIONS

PEARSON CORRELATION

Pearson correlation is a measure of the linear relationship between two variables. A Pearson correlation coefficient of 1 means that there is a perfect positive linear relationship between the two variables, which means that as one variable increases, the other variable also increases at a constant rate. A coefficient of 0 means that there is no linear relationship between the two variables. A coefficient of -1 means that there is a perfect negative linear relationship between the two variables, which means that as one variable increases, the other variable decreases at a constant rate.

The p-value shows how much this correlation can be the result of luck. Here we have big datasets so the p-value is always very small, and show no randomness is involved.



```python
import pandas as pd
from scipy.stats import pearsonr

# read the CSV file into a DataFrame
df = pd.read_csv("https://raw.githubusercontent.com/Sawcrate0/FilecoinDataStorageBounty/main/fullDataCorrelations.csv")

# extract the energy used and the filecoin price columns from the DataFrame
energyUsedSealDataEstimate = df["energyUsedSealDataEstimate"]
storageCapacityAdded = df["storageCapacityAdded"]

# calculate the Pearson correlation coefficient
corr, p_value = pearsonr(energyUsedSealDataEstimate, storageCapacityAdded)

# print the Pearson correlation coefficient and the p-value
print(f"Pearson correlation coefficient: {corr}")
print(f"p-value: {p_value}")

corr_matrix = df.corr()
energySeal_storage_corr = corr_matrix["energyUsedSealDataEstimate"]["storageCapacityAdded"]
print(energySeal_storage_corr)
```

[15] ✓ 0.2s

```
Pearson correlation coefficient: 1.0
p-value: 0.0
```

As we saw in the last part that the energy used to seal data was what decide of the storage capacity added everyday, we wanted to calculate the Pearson correlation between the two. The algorithm returns 1, which is the strongest correlation possible.

We claculated the same for the energy used to store data and the total data stroage also have a Pearson correlation coeficient of 1, which is also proving our last point. It is explained by the technology behind.

```
corr_matrix = df.corr()
energySeal_price_corr = corr_matrix["energyUsedSealDataEstimate"]["close"]
print(energySeal_price_corr)
```

✓ 0.2s

```
Pearson correlation coefficient: 0.5836671865903879
p-value: 9.075065379221996e-73
0.5836671865903875
```

For the full code for the correlations calculations, please kindly check the GitHub repository.

CORRELATION BETWEEN THE ENERGY USED TO SEAL THE FILECOIN DATA AND THE PRICE OF ITS $FIL TOKEN

A Pearson correlation coefficient of 0.583 means that there is a moderate positive linear relationship between the two variables. This means that as one variable increases, the other variable also tends to increase, but the relationship is not as strong as it would be with a coefficient of 1.

CORRELATION BETWEEN THE ENERGY CONSUMPTION RATE OF THE FILECOIN NETWORK AND ITS $FIL TOKEN PRICE

We obtain a Pearson correlation coefficient of 0.08 means that there is a very weak positive linear relationship between the two variables. This means that there is some relationship between the two variables, but it is not very strong. It's possible that other factors are having a greater effect on the variables than the relationship between the two variables.

CORRELATION BETWEEN THE CAPACITY OF STORAGE ADDED PER DAY AND THE PRICE OF ITS $FIL TOKEN

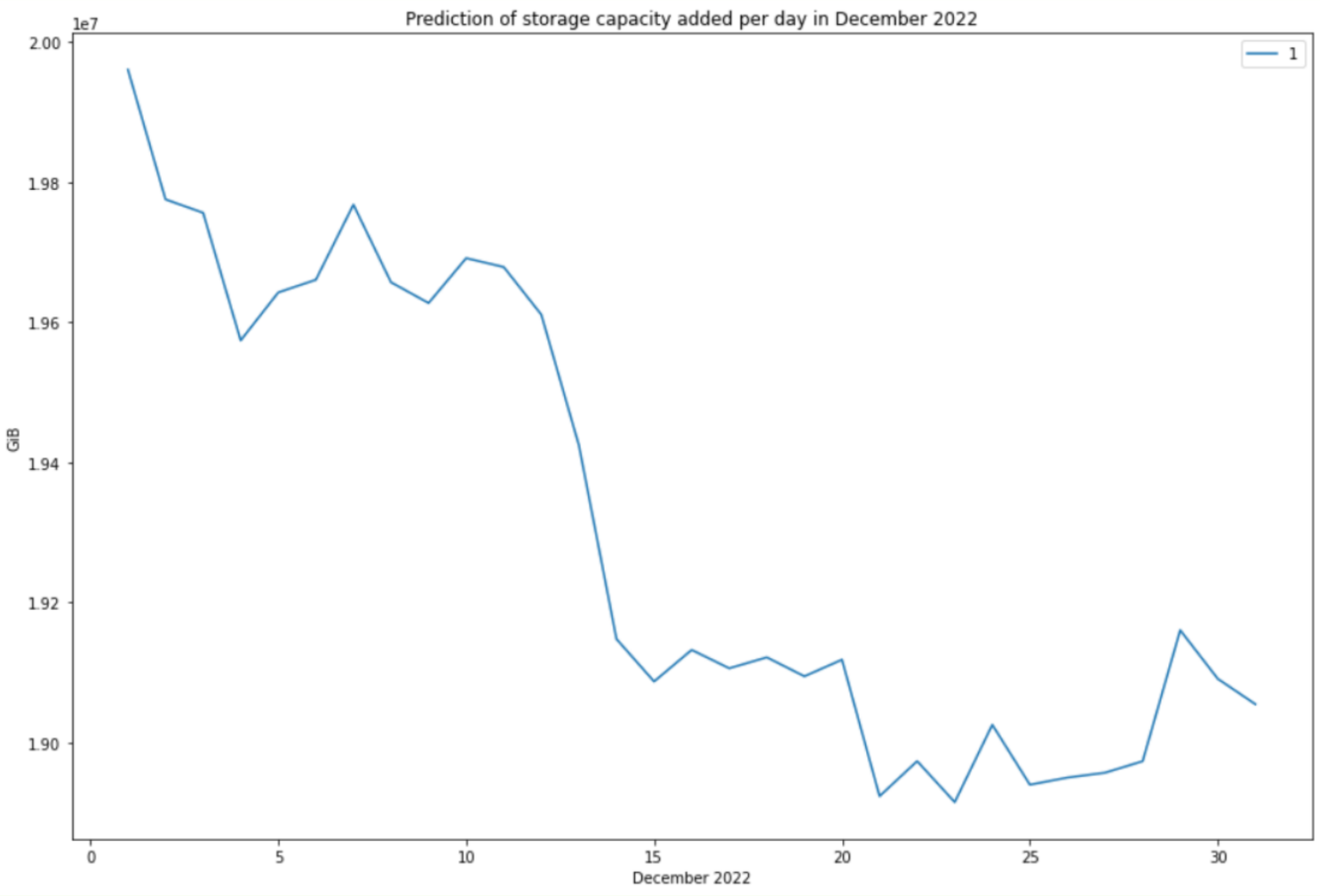
We have a Pearson correlation coefficient of 0.583, which is the same as the one between the energy used to seal the filecoin data and the price. It is quite logical, as we saw in the global analysis before that they are very very strongly correlated.

4

MACHINE LEARNING ALGORITHM

THE MACHINE LEARNING ALGORITHM THAT PREDICTS THE STORAGE CAPACITY ADDED PER DAY FOR THE MONTH OF DECEMBER 2022 CAN BE FOUND ON THE GITHUB UNDER THE NAME "MACHINE_LEARNING_MODELS_FILECOIN_DATA_STORAGE.IPYNB".

THIS CODE INCLUDES ALL INSIGHTS RELATED TO THIS PREDICTION.



THE MODEL PREDICTS AN AVERAGE DAILY ADDED STORAGE CAPACITY OF 17805245.70830726 GIB FOR DECEMBER 2022.

DIFFICULTY TO FIND AN ADDITIONAL DATASET

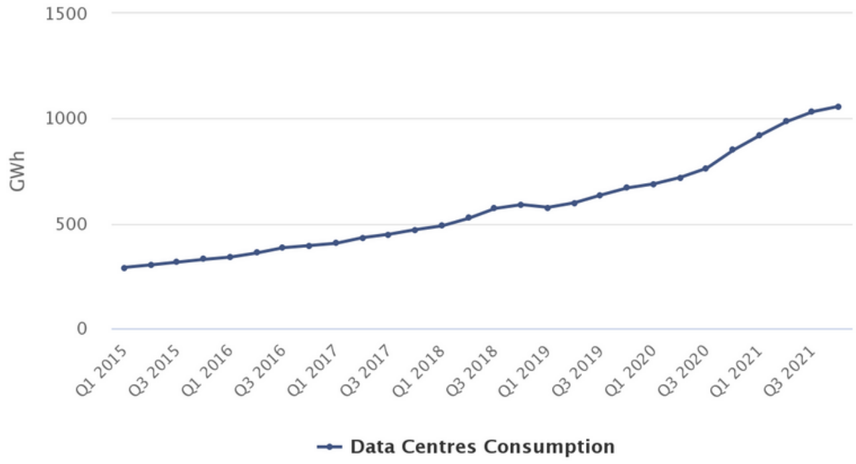
While Filecoin is very transparent on its consumption and the data is accesible very easily, it is not the fact for most datacenters. I found a table giving the Data Centres Metered Electricity Consumption for a few years from the Central Stastitics Office and published it on the Ocean Market.

WHAT DOES IT SHOW ?

Table 1 Metered Electricity Consumption 2015-2021

			Gigawatt hours	*% of Total*
Year	Data Centre	Other Metered Customers	Total	% Data Centre
2015	1 236	23 364	24 600	5
2016	1 477	23 879	25 356	6
2017	1 755	23 970	25 725	7
2018	2 172	24 558	26 730	8
2019	2 478	24 027	26 505	9
2020	3 019	24 037	27 056	11
2021	3 993	24 513	28 506	14

Figure 2 Data Centres Metered Electricity Consumption by Quarter 2015–2021 (GWh)



Source: CSO Ireland

The energy consumption of Filecoin's installations and the energy consumption of "traditionals" data centers is following the same trend, even though Filecoin would account for a gigantic datacenter (it is composed of a lot of them and is widely used).

CONTACT INFORMATION

MAILING ADDRESS
nicolaslandry01@gmail.com

THANK YOU FOR READING THIS REPORT, I AM LOOKING FORWARD TO THE NEXT DATA CHALLENGE.