DspytAI

Storage Network Energy Consumption Data Challenge

Pavel Fedotov (pfed-prog)

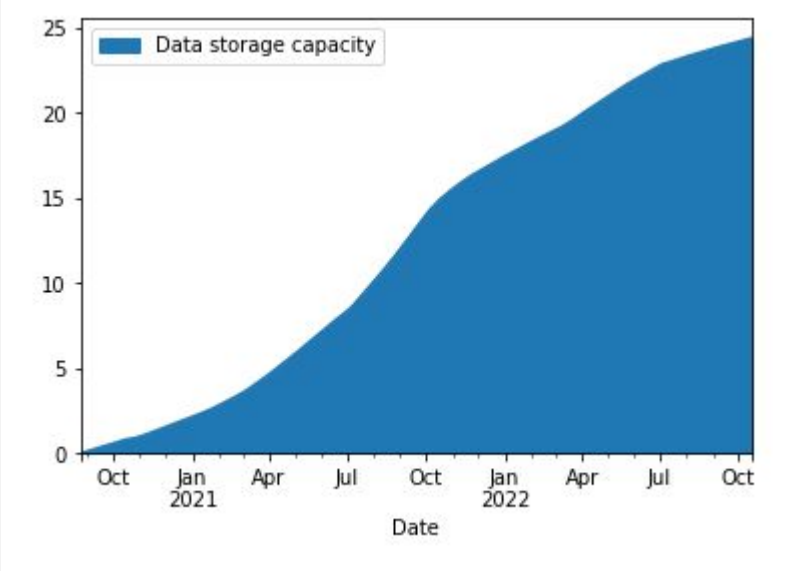
Introduction

The global data center industry has been rapidly expanding to meet the growing demand for digital transactions. This exponential growth has put a strain on energy supplies, with the data center industry accounting for 4% of global electricity consumption and 1% of global greenhouse gas emissions. Filecoin Green is an initiative from Protocol Labs that is aiming to address this issue. It is a decentralized storage network that will enable storage providers to use 100% renewable energy and to measure and reduce environmental impacts. Filecoin Green will provide the world's best tools to measure and reduce environmental impacts, creating a future of Web3 powered by verifiably clean energy. This report will explore the challenges associated with the global data center industry and Filecoin Green's potential to create a sustainable digital future.

Energy performance and the evolution of Filecoin's storage

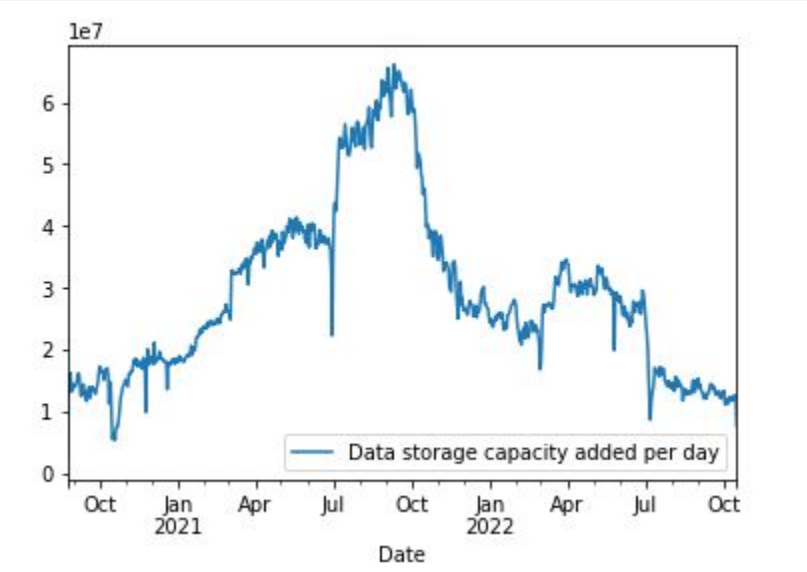
Filecoin is a decentralized storage network that is aiming to use 100% renewable energy.

As of 2022, Filecoin is responsible for over 24 exabytes of storage and has seen a steady increase in storage capacity since its launch in 2020. This growth is indicative of the increasing demand for cloud storage and Filecoin's ability to meet this demand.



Energy performance and the evolution of Filecoin's storage

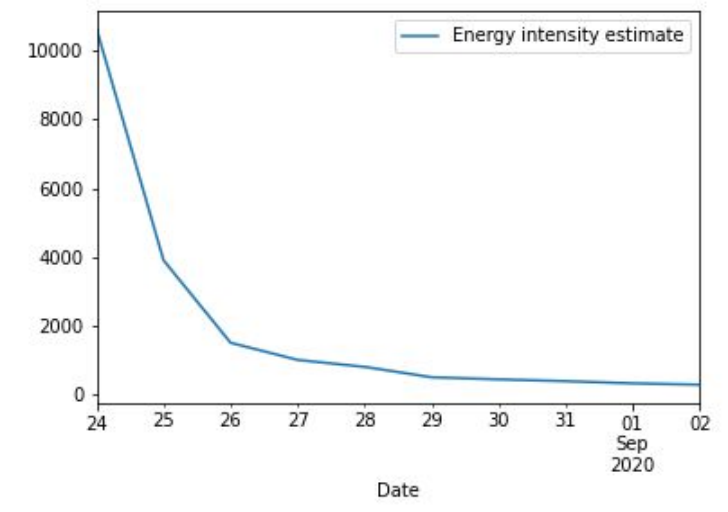
The network has performed incredibly well in 2021, reaching a peak for Data storage capacity added per day on September 10th 2021 of 66155072 GiB.

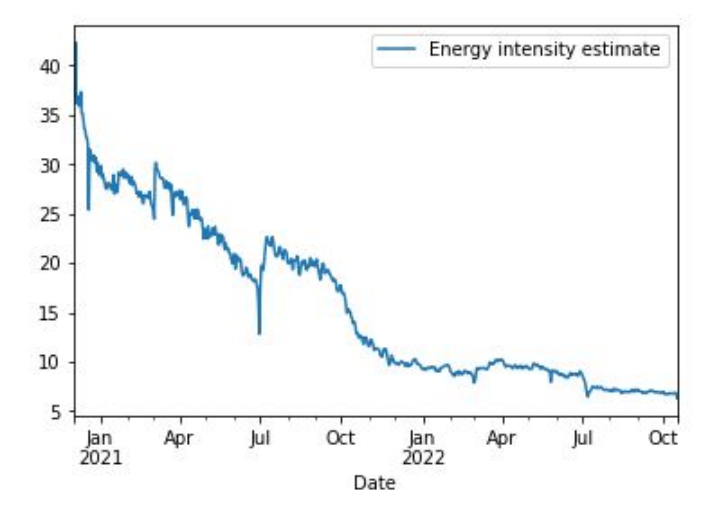


Energy performance and the evolution of Filecoin's storage

Filecoin has also seen an increase in energy efficiency since its launch. Filecoin's storage network estimated to use only 6.298506 MW/EiB on October 16th 2022, compared to August 24th 2020 10618.24561 MW/EiB. This is indicative of the efficiency gains achieved by Filecoin's decentralized storage network.

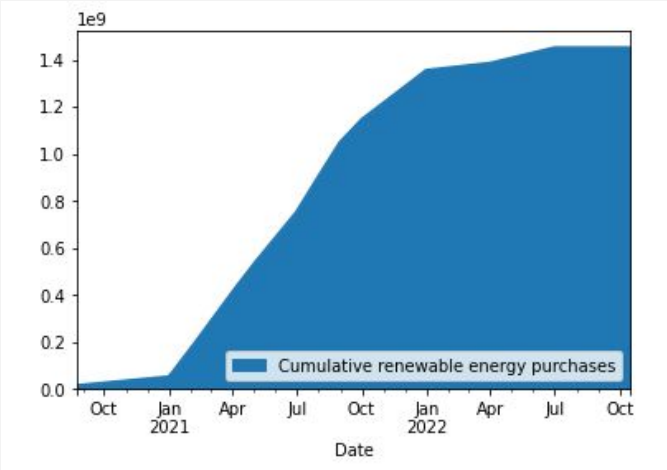
The network has also seen an increase in the number of nodes and miners, allowing it to become more resilient and fault-tolerant. This has led to fewer storage failures and reduced downtime, providing a superior user experience.





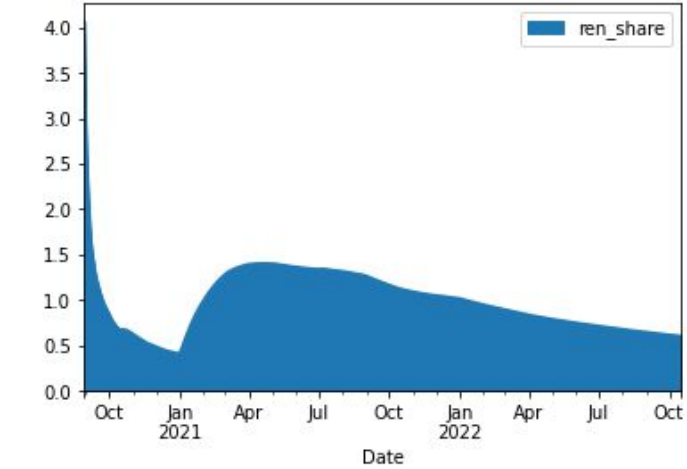
The evolution of the share of renewable energy used by Filecoin

Filecoin has been continuously working towards using 100% renewable energy since its launch in 2020. In fact, Filecoin has made over 1452033144 kWh in renewable energy certificate (REC) purchases over time.



The evolution of the share of renewable energy used by Filecoin

The share of renewable energy has sharply risen over Spring in 2021 and has been on a slight decline trend in 2022. Meanwhile the mean estimate of the renewable share of the total energy is at 1.197482%.



Summary Statistics per year

[154]:
```
df[df.year==2020].describe()
```

[15…

	Data storage capacity	Data storage capacity added per day	Cumulative renewable energy purchases	Energy intensity lower bound	Energy intensity estimate	Energy intensity upper bound	Energy consumption rate lower bound	Energy consumption rate estimate	Energy consumption rate upper bound	Energy used to seal data lower bound	Energy used to seal data estimate	Energy used to seal data upper bound	Cumulative energy use lower bound	Cumulative energy use estimate	Cumulative energy use upper bound	Energy used to store data lower bound	Energy used to store data estimate	Energy used to store data upper bound
count	1.300000e+02	1.300000e+02	1.300000e+02	130.000000	130.000000	130.000000	130.000000	130.000000	130.000000	130.000000	130.000000	130.000000	1.300000e+02	1.300000e+02	1.300000e+02	130.000000	130.000000	130.000000
mean	9.172850e+08	1.519408e+07	3.419768e+07	28.533072	211.998913	434.262171	5868.304266	41100.180833	88867.157630	4084.473512	23223.700725	38067.256666	8.112800e+06	5.800796e+07	1.230695e+08	888.665697	2954.758404	7977.902728
std	5.639737e+08	3.425864e+06	1.078490e+07	131.554397	995.218304	2005.377628	1541.942892	10092.433410	23212.937590	920.940831	5236.330746	8583.160318	5.257421e+06	3.676172e+07	7.960536e+07	546.377724	1816.672094	4905.048493
min	4.984480e+05	2.053024e+06	1.573466e+07	3.600785	23.384402	54.201516	651.804634	4929.158175	9935.604259	551.893912	3137.985593	5143.646330	1.564318e+04	1.182999e+05	2.384543e+05	0.482896	1.605601	4.335152
25%	4.399166e+08	1.364505e+07	2.496173e+07	5.288415	36.151441	79.927263	4687.715063	34469.718200	71351.837405	3668.061803	20856.046520	34186.303262	3.601446e+06	2.661061e+07	5.478655e+07	426.191188	1417.059305	3826.086517
50%	8.448714e+08	1.602704e+07	3.419701e+07	6.727845	47.040835	101.869532	5792.980918	41429.845125	87799.931120	4308.388893	24496.849825	40154.146020	7.279943e+06	5.257128e+07	1.105277e+08	818.511382	2721.499654	7348.099758
75%	1.380350e+09	1.787266e+07	4.343102e+07	12.153555	88.086710	184.577494	7360.989879	50639.833890	111289.579775	4804.529536	27317.830743	44778.172390	1.235879e+07	8.790267e+07	1.873973e+08	1337.282860	4446.382689	12005.316080
max	1.966270e+09	2.114131e+07	5.266258e+07	1404.098113	10618.245610	21402.982540	8395.849442	58207.598740	127037.128800	5683.207492	32313.861150	52967.443080	1.830898e+07	1.282327e+08	2.772653e+08	1904.922699	6333.749997	17101.242970

```
df[df.year==2021].describe()
```

	Data storage capacity	Data storage capacity added per day	Cumulative renewable energy purchases	Energy intensity lower bound	Energy intensity estimate	Energy intensity upper bound	Energy consumption rate lower bound	Energy consumption rate estimate	Energy consumption rate upper bound	Energy used to seal data lower bound	Energy used to seal data estimate	Energy used to seal data upper bound	Cumulative energy use lower bound	Cumulative energy use estimate	Cumulative energy use upper bound	Energy used to store data lower bound	Energy used to store data estimate	Energy used to store data upper bound
count	3.650000e+02	3.650000e+02	3.650000e+02	365.000000	365.000000	365.000000	365.000000	365.000000	365.000000	365.000000	365.000000	365.000000	3.650000e+02	3.650000e+02	3.650000e+02	365.000000	365.000000	365.000000
mean	8.573107e+09	3.904893e+07	7.626805e+08	3.192743	20.297538	47.981446	22187.254965	137062.262803	332724.402647	10497.132212	59685.111496	97833.178500	9.854931e+07	6.414550e+08	1.483698e+09	8305.626233	27615.692837	74562.885046
std	4.594770e+09	1.302226e+07	3.976652e+08	0.809995	6.127666	12.347326	7724.244445	44931.468890	115209.659590	3500.644670	19904.137931	32625.977068	5.990165e+07	3.746910e+08	8.991130e+08	4451.413455	14800.674053	39962.095629
min	1.984233e+09	1.797949e+07	5.559982e+07	1.742985	9.330019	25.881753	8055.022281	53740.794780	121517.395200	4833.245964	27481.108020	45045.809240	1.850230e+07	1.295242e+08	2.801817e+08	1922.325089	6391.611868	17257.471100
25%	4.364039e+09	2.751014e+07	4.198907e+08	2.744859	16.909260	41.154021	16246.926750	107237.956900	244867.987100	7395.276910	42048.429800	68923.914780	4.428372e+07	3.006904e+08	6.689297e+08	4227.880619	14057.441220	37955.353120
50%	7.814262e+09	3.726186e+07	7.615898e+08	3.202048	20.367931	48.123288	22290.228700	136932.060300	334780.206800	10016.732130	56953.629040	93355.854020	8.582973e+07	5.670852e+08	1.293703e+09	7570.457141	25171.301140	67962.981930
75%	1.304972e+10	5.141136e+07	1.147426e+09	3.952146	26.042473	59.557563	28133.686260	172122.106600	420364.854100	13820.401800	78580.721420	128806.021300	1.515867e+08	9.833888e+08	2.281611e+09	12642.571410	42035.766970	113497.353800
max	1.621590e+10	6.615507e+07	1.355702e+09	4.494727	30.147135	67.828509	34539.414020	218115.448200	518189.013900	17783.806460	101116.042900	165744.917400	2.126684e+08	1.328899e+09	3.191929e+09	15709.968200	52234.671320	141034.585500

Summary Statistics per year

```
df[df.year==2022].describe()
```

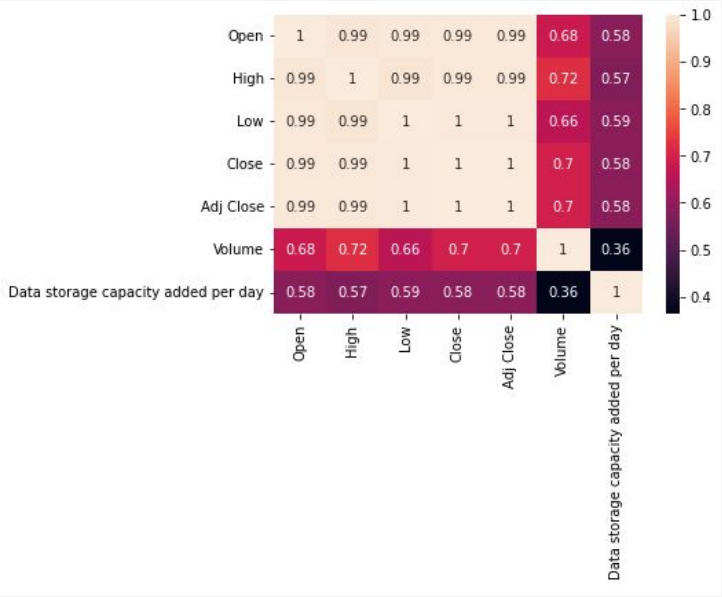
	Data storage capacity	Data storage capacity added per day	Cumulative renewable energy purchases	Energy intensity lower bound	Energy intensity estimate	Energy intensity upper bound	Energy consumption rate lower bound	Energy consumption rate estimate	Energy consumption rate upper bound	Energy used to seal data lower bound	Energy used to seal data estimate	Energy used to seal data upper bound	Cumulative energy use lower bound	Cumulative energy use estimate	Cumulative energy use upper bound	Energy used to store data lower bound	Energy used to store data estimate	Energy used to store data upper bound
count	2.890000e+02	2.890000e+02	2.890000e+02	289.000000	289.000000	289.000000	289.000000	289.000000	289.000000	289.000000	289.000000	289.000000	2.890000e+02	2.890000e+02	2.890000e+02	289.000000	289.000000	289.000000
mean	1.989474e+10	2.242077e+07	1.416390e+09	1.623427	8.425554	24.059245	29855.386647	154416.427071	442362.786531	6027.152491	34269.480573	56173.007405	3.152389e+08	1.870609e+09	4.713701e+09	19274.022634	64084.931573	173030.508933
std	2.020504e+09	7.252472e+06	3.560490e+07	0.148454	1.123066	2.262992	1913.962730	13199.453528	28657.505775	1949.609501	11085.185751	18170.343142	6.111920e+07	3.182913e+08	9.059729e+08	1957.464659	6508.448760	17572.932882
min	1.624013e+10	7.650816e+06	1.356034e+09	1.342260	6.298506	19.773216	25492.355260	128556.939700	377294.180200	2056.692357	11694.042730	19168.354410	2.132991e+08	1.332270e+09	3.201294e+09	15733.435360	52312.698160	141245.259400
25%	1.800405e+10	1.437181e+07	1.379935e+09	1.449733	7.111543	21.411496	29007.113250	143504.615000	429276.067200	3863.429427	21966.877370	36007.127760	2.605878e+08	1.583674e+09	3.903182e+09	17442.319730	57994.632870	156586.589000
50%	2.022467e+10	2.431549e+07	1.425743e+09	1.676417	8.826421	24.866998	29705.647130	147460.941500	438849.166100	6536.489484	37165.493940	60920.023640	3.152991e+08	1.878830e+09	4.716005e+09	19593.660820	65147.708740	175900.027100
75%	2.175660e+10	2.864445e+07	1.452033e+09	1.749087	9.376178	25.974766	31571.728450	168615.268900	469034.798800	7700.200511	43782.179430	71765.800020	3.686084e+08	2.151154e+09	5.505276e+09	21077.796590	70082.368270	189223.699700
max	2.270389e+10	3.458179e+07	1.452033e+09	1.862770	10.236194	27.707707	33541.436070	180134.763600	497762.033700	9296.277325	52857.231620	86641.221680	4.197450e+08	2.399932e+09	6.260157e+09	21995.524110	73133.755430	197462.501900

Correlation

We found a moderately positive correlation between daily price estimates such as Open, High, Low, Close and Adjusted Close and the energy used to seal the Filecoin data. We also found slightly weaker positive correlation between volume and the energy used to seal the Filecoin data. Hence, there is a positive relationship between higher price, volume and the energy used to seal the Filecoin data.

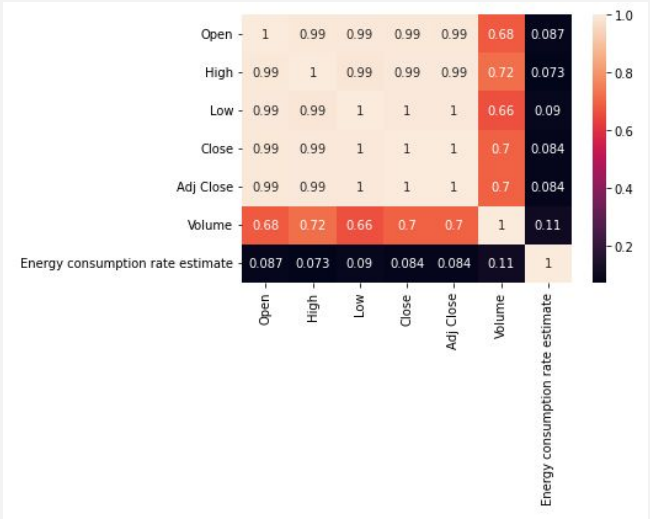
Correlation

We found a moderately positive correlation between daily price estimates such as Open, High, Low, Close and Adjusted Close and Data storage capacity added per day. We also found slightly weaker positive correlation between volume and Data storage capacity added per day. Hence, there is a positive relationship between higher price, volume and Data storage capacity added per day similar to the energy used to seal the Filecoin data.



Correlation

We found a little correlation between daily price estimates such as Open, High, Low, Close and Adjusted Close and the energy consumption rate of the Filecoin network. We also found a very weak positive correlation between volume and the energy consumption rate. Hence, there is a little relationship between higher price, volume and the energy consumption rate.



Correlation: Conclusion

The three correlations suggest that there is a positive relationship between higher price, volume, and the energy used to seal the Filecoin data, as well as the data storage capacity added per day. This suggests that as the price and volume of Filecoin increases, it is likely that the energy used to seal the Filecoin data and the data storage capacity added per day will also increase. Additionally, the weak correlation between price, volume and the energy consumption rate suggests that the energy consumption rate may not be as closely linked to the price and volume of Filecoin.

Algorithm

The algorithm should consider the following factors:

1. Historical energy consumption data

2. Historical $FIL/USD price data

3. Current estimated energy usage per Filecoin Storage Provider

4. Current estimated demand for Filecoin Storage Providers

The algorithm should take into account all of the above factors to accurately predict the storage capacity added per day for the month of December 2022. The algorithm should also be able to adjust its predictions if any of the factors change in the future.

Algorithm

First we connect, extract and merge the latest data with Filecoin green Energy Consumption API:

[56]:
```
# Energy Intensity

URL_0 = 'https://api.filgreen.d.interplanetary.one/models/model?id=0'

data_0 = pd.read_json(URL_0)

# Energy consumption rate

URL_2 = 'https://api.filgreen.d.interplanetary.one/models/model?id=2'

data_2 = pd.read_json(URL_2)

# Energy used to seal data

URL_3 = 'https://api.filgreen.d.interplanetary.one/models/model?id=3'

data_3 = pd.read_json(URL_3)

# Energy used to store data

URL_4 = 'https://api.filgreen.d.interplanetary.one/models/model?id=4'

data_4 = pd.read_json(URL_4)

# Cumulative Energy Use

URL_5 = 'https://api.filgreen.d.interplanetary.one/models/model?id=5'

data_5 = pd.read_json(URL_5)

# Data Storage capacity added per day

URL_7 = 'https://api.filgreen.d.interplanetary.one/models/model?id=7'

data_7 = pd.read_json(URL_7)
```

[65]:
```
df_0 = pd.json_normalize(data_0['data'][1]['data'])
df_2 = pd.json_normalize(data_2['data'][1]['data'])
df_3 = pd.json_normalize(data_3['data'][1]['data'])
df_4 = pd.json_normalize(data_4['data'][0]['data'])
df_5 = pd.json_normalize(data_5['data'][1]['data'])
df_7 = pd.json_normalize(data_7['data'][0]['data'])
```

[121]:
```
df = pd.DataFrame()

df['Date'] = pd.to_datetime(df_0[68:-9].start_date)
df['year'] = df.Date.dt.year
df['month'] = df.Date.dt.month
df.Date = df.Date.dt.strftime('%Y-%m-%d')
df['Energy Intensity'] = df_0[68:-9].value
df['Energy consumption rate'] = df_2[68:-9].value
df['Energy used to seal data'] = df_3[68:-9].value
df['Energy used to seal data'] = df_4[68:-9].value
df['Cumulative Energy Use'] = df_5[68:-9].value
df['Data Storage capacity added per day'] = df_7[68:-9].value
```

Algorithm

Next we obtain recent data from Binance API:

[126]:
```
binance_data
```

[12…

	open_time	open	high	low	close	volume	close_time	qav	num_trades	taker_base_vol	taker_quote_vol
0	2020-10-15	10.0000	430.0000	10.0000	62.5418	1083589.84	1602806399999	7.198688e+07	170124	530923.14	3.539159e+07
1	2020-10-16	62.5902	75.8725	38.8302	44.3480	1696876.76	1602892799999	9.461002e+07	240258	806748.36	4.522710e+07
2	2020-10-17	44.3512	50.6273	40.3325	42.2400	672950.78	1602979199999	3.028810e+07	107899	330452.82	1.487497e+07
3	2020-10-18	42.2364	42.3136	31.7722	32.4049	593629.16	1603065599999	2.171248e+07	76375	309192.97	1.129916e+07
4	2020-10-19	32.3569	35.0000	29.9321	33.9755	814098.46	1603151999999	2.599180e+07	83389	398467.17	1.273460e+07
...	...	...	...	...	...	...	...	...	...	...	...
781	2022-12-05	4.5800	4.6200	4.4600	4.5200	2931216.23	1670284799999	1.337093e+07	21879	1327670.02	6.059644e+06
782	2022-12-06	4.5200	4.6300	4.4900	4.5900	2612142.31	1670371199999	1.187845e+07	19626	1330963.55	6.055439e+06
783	2022-12-07	4.6000	4.6000	4.1900	4.3500	2978626.78	1670457599999	1.307734e+07	25774	1361821.25	5.974891e+06
784	2022-12-08	4.3400	4.4400	4.3000	4.4000	1911133.52	1670543999999	8.363459e+06	15397	948791.64	4.154511e+06
785	2022-12-09	4.4000	4.4700	4.3800	4.4200	1778601.72	1670630399999	7.881834e+06	12916	832439.68	3.691073e+06

Algorithm

Next we obtain recent data from Binance API and merge with Green Filecoin API Data:

```
[168]: df = df.merge(binance_data[binance_data_columns][17:-9], on='Date')
```

```
df
```

[16…

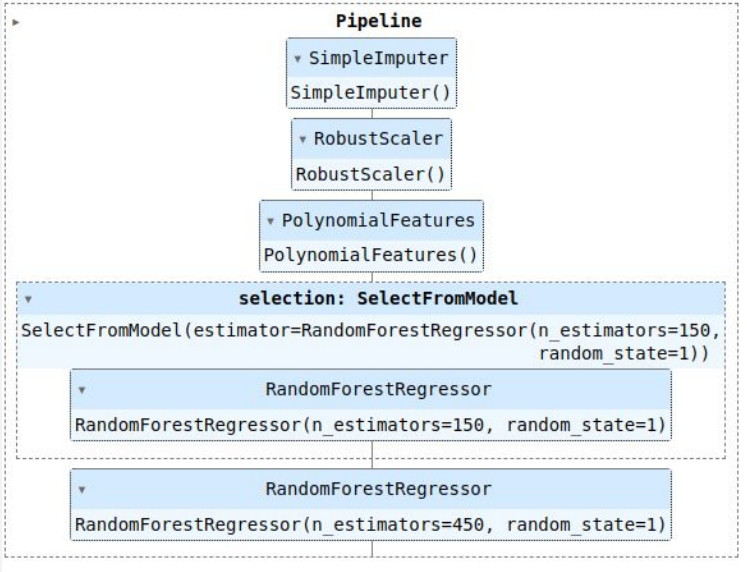
	Date	year	month	Energy Intensity	Energy consumption rate	Energy used to seal data	Cumulative Energy Use	Data Storage capacity added per day	open	high	low	close	volu
0	2020-11-01 00:00:00+00:00	2020	11	38933.407664764200	2859.5662643700	41065862.710857685807908000000	47.091205987586521433087410176	887733225	31.3196	31.8877	30.5975	31.6076	334260
1	2020-11-02 00:00:00+00:00	2020	11	37926.396615208056	2904.3553882308	41662012.306961273356890000000	45.165766552207194995342442496	901637709	31.5645	31.9782	29.3600	29.6401	661737
2	2020-11-03 00:00:00+00:00	2020	11	37092.864246096704	2947.9337854172	42258161.903064860905872000000	43.520132203146205815906500608	915166331	29.6254	29.6850	27.9081	28.8364	603170
3	2020-11-04 00:00:00+00:00	2020	11	41669.777472579436	2997.5513721148	42854311.499168448454854000000	48.080846679936673073103634432	930569779	28.8386	29.5325	28.1668	28.8967	549375
4	2020-11-05 00:00:00+00:00	2020	11	41388.572587312764	3046.6879337952	43450461.095272036003836000000	46.986169001018352867014082560	945823896	28.8979	30.2522	28.3853	29.3544	628207
…	…	…	…	…	…	…	…	…	…	…	…	…	
755	2022-11-26 00:00:00+00:00	2022	11	108447.169615971708	69074.6303286444	2995670431.548366920442474000000	5.430217346646016000000000000	21443757087	4.3000	4.4100	4.2700	4.3000	2374212
756	2022-11-27 00:00:00+00:00	2022	11	108447.169615971708	69074.6303286444	2995670431.548366920442474000000	5.430217346646016000000000000	21443757087	4.3000	4.5200	4.2900	4.3500	5384090
757	2022-11-28 00:00:00+00:00	2022	11	108447.169615971708	69074.6303286444	2995670431.548366920442474000000	5.430217346646016000000000000	21443757087	4.3500	4.3900	4.1300	4.2500	3927160
758	2022-11-29 00:00:00+00:00	2022	11	108447.169615971708	69074.6303286444	2995670431.548366920442474000000	5.430217346646016000000000000	21443757087	4.2500	4.4100	4.2100	4.3500	3665157
759	2022-11-30 00:00:00+00:00	2022	11	108447.169615971708	69074.6303286444	2995670431.548366920442474000000	5.430217346646016000000000000	21443757087	4.3500	4.5100	4.2300	4.4200	4057278

760 rows × 17 columns

Algorithm

Further, we aggregate the data columns by month and year with median, mean and sum function in NumPy. We shift the columns one period, in order to predict one step ahead period which is December 2022.

We Train the Model to estimate the median of the Data Storage capacity added per day. The sklearn pipeline looks as following:



Algorithm

We estimated the median of the Data Storage capacity added per day

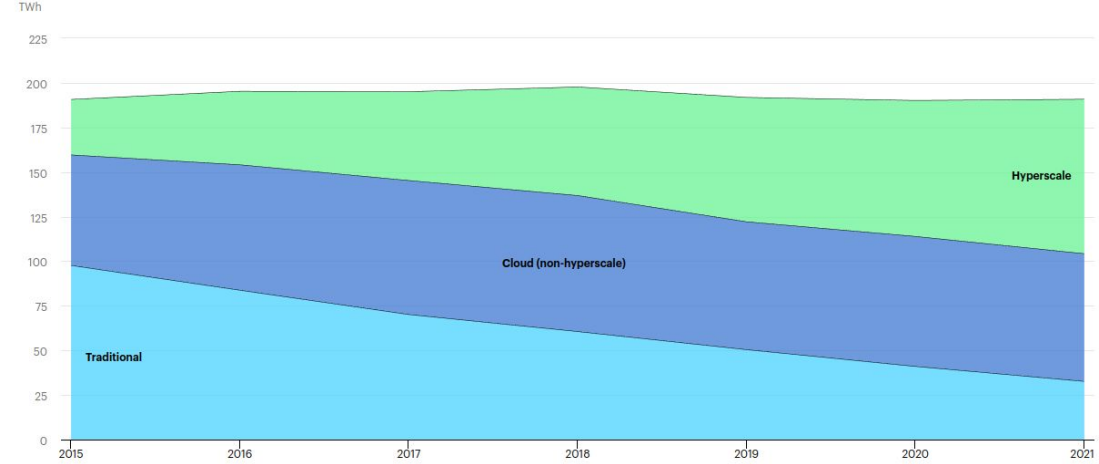
```
clf_xg.predict(final_df[cols][-1:].values)[0]
```

```
20432085538.684444
```

Bonus

We include Energy demand in data centers worldwide from 2015 to 2021, by type dataset which assists in determining the trends in data center storage industry.

The traditional data center industry has seen a reduction in energy demand since 2015, with estimates suggesting that this figure will reach nearly 33 terawatt-hours by 2021. However, the demand for energy by hyperscale data centers has doubled in the same period of time.



Conclusion

Traditional data center operators have been traditionally focused on maximizing output and performance, often overlooking the power implications. This has led to an increase in energy consumption as additional data centers were built to meet the increased demand for computing power. However, the availability of cloud computing resources, which require no floor space, has enabled organizations to shut down legacy data centers and move operations to the cloud. Non-hyperscale cloud data centers have kept their energy consumption steady, while large, hyperscale cloud data centers have steadily increased their energy usage and effectively managed it.

Filecoin storage has the potential to be significantly more energy-efficient than conventional data centers. The decentralized storage network is aiming to use 100% renewable energy, and has seen an increase in energy efficiency since its launch. It has also seen an increase in the number of nodes and miners, allowing it to become more resilient and fault-tolerant. In comparison, traditional data centers have seen a decrease in energy demand since 2015, but large, hyperscale cloud data centers have steadily increased their energy usage. Filecoin storage has the potential to be more energy-efficient, and to reduce its carbon footprint, while still meeting the growing demand for digital transactions.