☐ Python 기반 데이터 분석 프로젝트

및 유튜브 프리미엄은 못참조

김대찬, 김재은 , 정하늘

동영상 시청과 환경 오염 사이

누워서 영상 보기, 환경을 파괴한다?





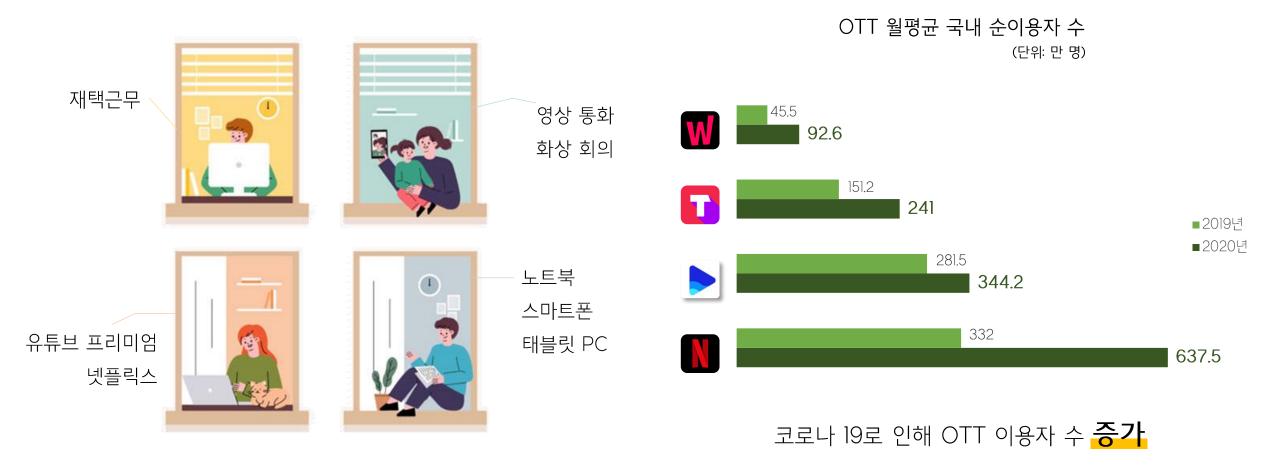
INDEX

- □ 주제 선정 배경
- □ 변수 설명
- □ 분석 결과 및 내용
- □ 결론 및 개선 방안
- □ 소감



주제 선정 배경 & 주제 소개









☑ 인터넷 검색

0.2g

Co.

☑ 이메일 전송

4g

☑ 동영상 시청 30분

1.6kg

(= 6.3km 자동차 운전 시 발생량)







실시간 검색과 영상 재생



모든 데이터 보관 • 전송

- ☑ IT기업, 통신사, 포털 사이트 별 데이터센터 운영
- ☑ 365일, 24시간 전력 사용
- ☑ 발생하는 열을 낮추기 위한 냉각 장치 사용
- 이산화탄소를 포함한 온실가스 배출



☑ 2020년 세계 데이터센터 에너지 사용량 <u>1조 9730억kWh</u> = 우리나라 1년 전기 사용량의 **4배**

✓ 전체 산업용 전기 소모량 7~8%의 전력 사용= <u>춘천시 2년 사용 전기 소모량</u>



- ☑ 전력효율지수 = 데이터센터의 총 전력소모량 / IT 장비 한 대당 전력소모량
 - 친환경 데이터센터의 전력효율지수: 1.1~1.3 정도
 - 미국 데이터센터: 1.5~1.8 정도, <u>한국 데이터센터: 2.66 측정</u>

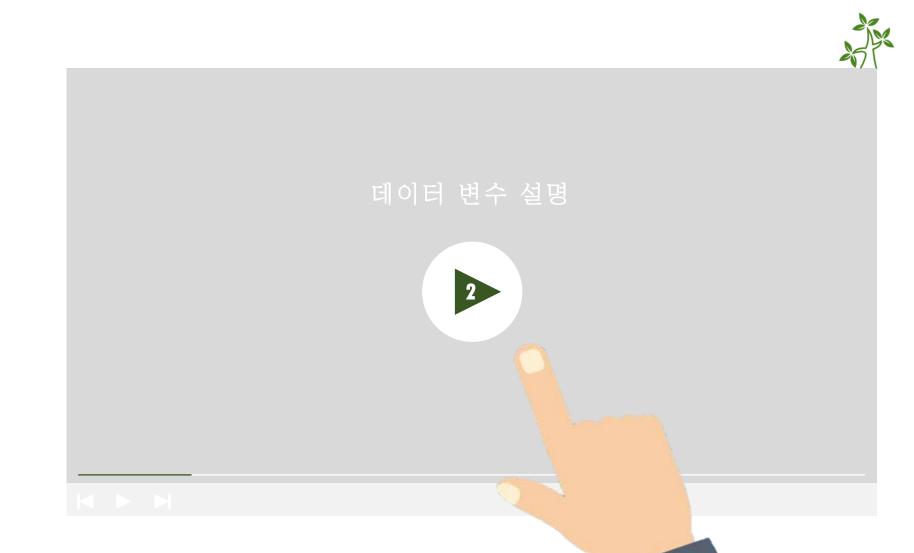




동영상 플랫폼 사용량과 환경오염의 연관성 분석으로

개인과 기업에 환경 문제에 대한 경각심 제고





□ 2. 변수 설명

데이터 변수 소개

☑ 온라인 동영상과 환경오염의 연관성



온라인 동영상 이용자 수

- OTT 이용자의 이용빈도
- OTT 시장 규모

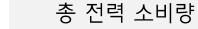




데이터센터 전력 소비량

- 데이터센터 전력 소비량
- 데이터센터 수 및 분포





- 1인당 전기소비량
- GDP 대비 전기소비량





온실가스 배출량

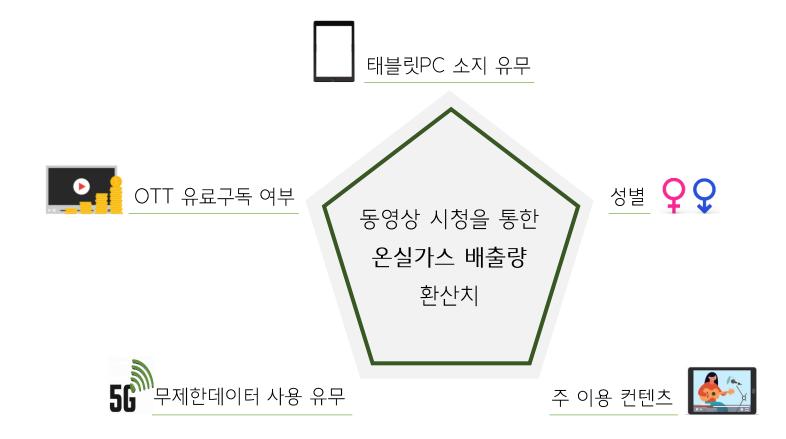
- 연도별 온실가스 배출량
- 행정별 온실가스 배출량



□ 2. 변수 설명

데이터 변수 소개

☑ 개인의 온실가스배출량 기여 요인







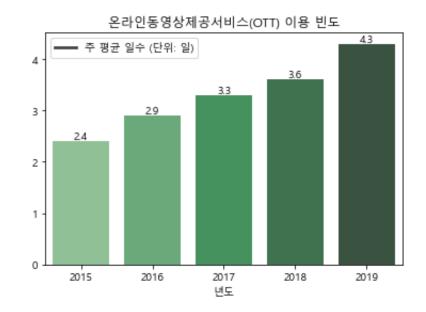
데이터 분석 결과 및 내용



OTT 사용량 추이

☑ 국내 온라인 동영상 제공 서비스(OTT) 이용 빈도 추이 (2015년~2019년)





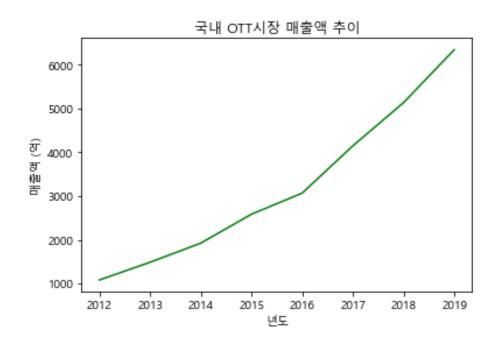
☑ 주 평균 OTT 이용자의 사용 일수

 $=\frac{7\times[\text{mg}(\%)]+5.5\times[\text{l}\text{Fg}]\text{d} 5-6\text{g}(\%)]+3.5\times[\text{l}\text{Fg}]\text{d} 3-4\text{g}(\%)]+1.5\times[\text{l}\text{Fg}]\text{d} 1-2\text{g}(\%)]+0.5\times[\text{l}\text{Fg}]}{100}$



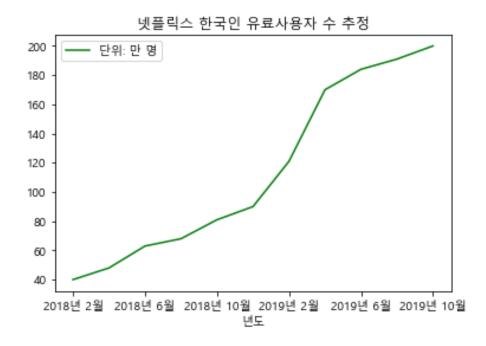
OTT 사용량 추이

☑ 국내 OTT 시장 매출액 추이



2012년 → 2019년 484% 성장

☑ 국내 넷플릭스 유료 사용자 수

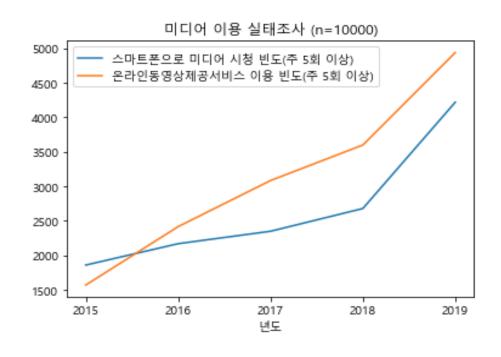


2018년 2월 → 2019년 10월 400% 증가



OTT 사용량 추이

☑ 미디어 이용 실태 조사



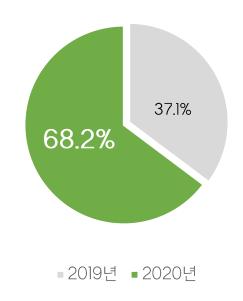
- 주 이용 매체: 스마트폰
- 스마트폰으로 미디어 시청하는 빈도 증가



OTT 사용량 추이

☑ 2019년 대비 2020년 OTT 이용자 분석

"최근 3개월 이내 OTT 이용해본 적 있다"



OTT 이용자 중 <mark>유료 결제 이용자</mark> (비율)

2019년	8.7%	
2020년		12.4%

OTT 이용자의 I회 평균 이용시간 (단위: 분)

2019	1	36.4	
2020			48.6

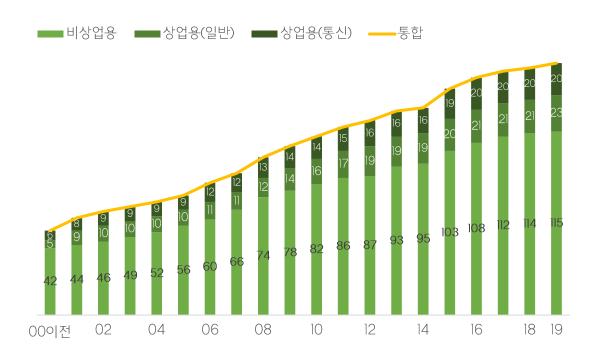
동영상 서비스 1주 평균 이용 시간 (단위: 시간)

2019년	4.5	
2020년		6

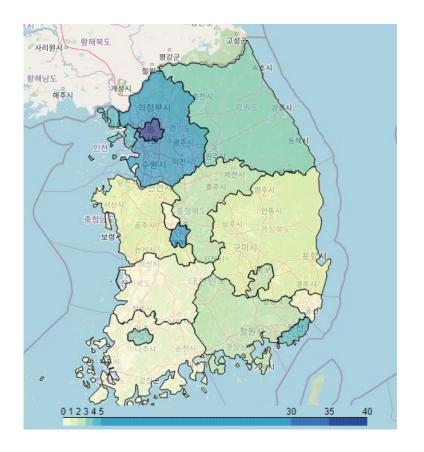


데이터센터 전력 소비량

☑ 국내 데이터센터 분포



☑ 상업용(민간) 데이터센터 지리적 분포도

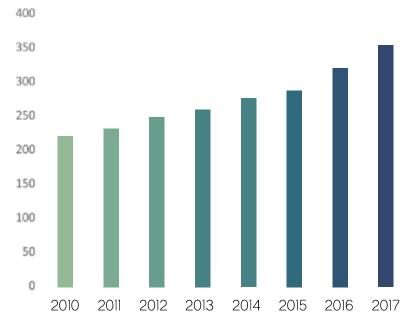






데이터센터 전력 소비량

☑ 글로벌 데이터센터의 에너지 소비량 현황 및 예측치



2010년 - 2017년 전 세계 데이터센터 에너지 소비량

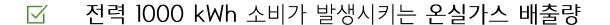
15	2016	2017	

구분 2014년 2015년 2016년 연평균증가율 264,618 1.0% 산업용 265,633 269,975 5.5% 데이터센터 2,511 2,731 2,797 데이터센터 산업용 X100 (%) 0.95% 1.03% 1.04%



온실가스와 전력 소비량 관계

- ☑ 온실가스 배출량 = 활동 자료 X 배출 계수 X 지구 온난화 지수
 - 온실가스: 직접/간접 배출량, CO₂ 환산 배출량, 6대 온실가스
 - 활동 자료: 연료 사용량, 제품 생산량, 폐기물 매립량 등
 - 매출 계수: 연료별 연소 배출계수, 발열량, 산화율 등
 - 지구 온난화 지수: CO2 (1), Ch4 (21), N2O (310), HFCs (140~11,700), PFCs (6,500~9,200), SF6 (23,900)



= 1000 kWh [활동 자료] X 0.466 kgCO₂eq/kwh [배출 계수] X 1 [지구 온난화 지수] = 0.466 tCO₂eq





온실가스와 전력 소비량 관계

- ☑ 종단적 분석과 횡단적 분석 동시 진행
 - 종단적 분석: 2006년부터 2019년까지의 시계열 자료 활용
 - 횡단적 분석: 2018년 지역별 자료 활용

온실가스 배출량 (단위: Gg CO₂eq.)

분야·부문/연도	2016	2017	2018
A. 연료연소	598,726.70	611,641.89	627,909.14
1. 에너지산업	263,657.08	270,986.26	287,614.10
a. 공공전기 및 열 생산	244,035.22	252,603.47	269,570.99
b. 석유정제	17,916.02	15,847.60	15,858.29
c. 고체연료 제조 및 기타 에너지 산업	1,705.84	2,535.19	2,184.82
2. 제조업 및 건설업	181,427.79	186,545.17	186,596.10
a. 철강	93,360.11	99,444.87	95,287.62
b. 비철금속	2,674.08	2,631.22	2,979.01
c. 화학	39,167.26	40,626.58	45,952.81
d. 펄프, 제지 및 인쇄	646.30	645.38	662.19
e. 식음료품 가공 및 담배 제조	1,849.96	1,832.40	1,955.08
f. 기타	43,730.07	41,364.73	39,759.39
서울 부산 대구 인천 광주 대전 울산 세종 경기 강원 충북	북 충남 전북 전남	경북 경남 제주	+ : •

시구	사용량(kWh)	전기요금(원)	평균판매단가(원/kWh)
경기도	48,915,491,883	5,380,266,922,343	110
경상북도	20,262,966,500	2,068,349,814,211	102.1
광주광역시	3,201,074,212	373,794,852,722	116.8
대구광역시	5,906,931,423	684,919,946,409	116
대전광역시	3,580,144,784	409,588,126,683	114.4
부산광역시	7,817,499,493	928,260,555,633	118.7
서울특별시	16,003,895,073	1,954,823,215,105	122.1
세종특별자치시	1,214,498,083	133,440,019,244	109.9

* 출처: 환경부, 한국전력공사



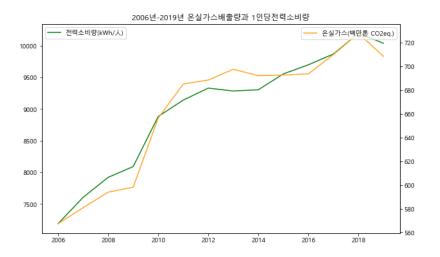
온실가스와 전력 소비량 관계

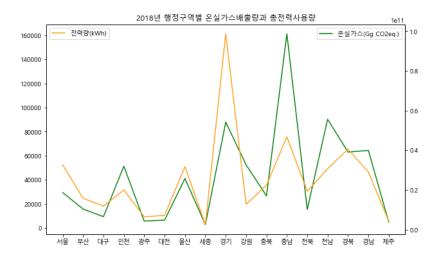
☑ 종단적 분석



☑ 횡단적 분석







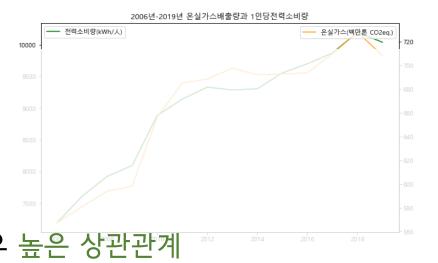


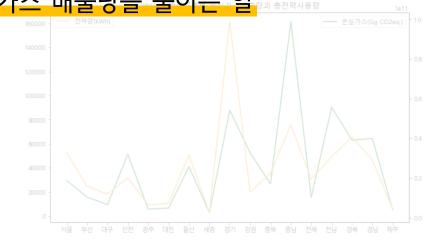
3. 분석 결과 및 내용

온실가스와 전력 소비량 관계

종단적 분석



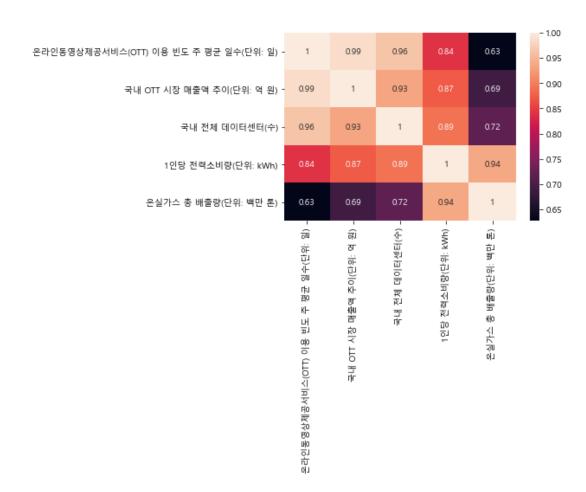


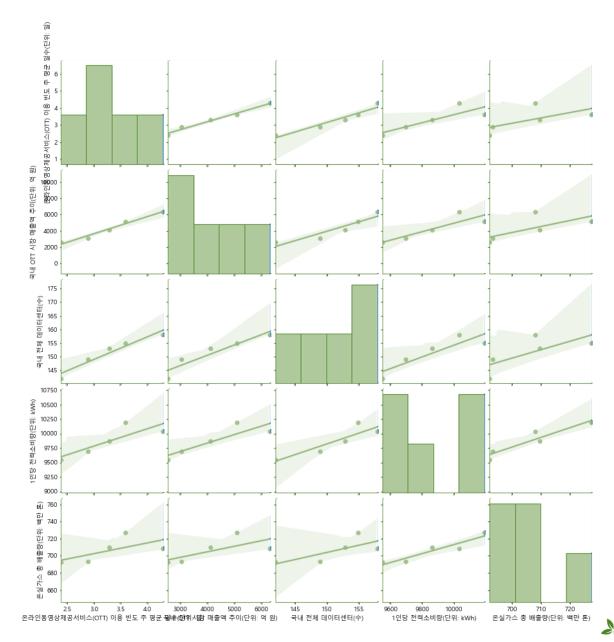




데이터 변수 간의 종합 비교

☑ 2015년 - 2019년 상관관계





데이터 변수 간의 종합 비교

☑ 2015년 - 2019년 상관관계

	OTT 이용빈도 주 평균일수	국내 OTT 시장 매출액 추이	국내 전체 데이터센터(수)	1인당 전력소비량	온실가스 총 배출량
OTT이용빈도 주평균일수	1.000000	0.986895***	0.962572**	0.840723	0.627269
국내 OTT 시장 매출액 추이	0.986895***	1.000000	0.933550*	0.873138*	0.690691
국내 전체 데이터센터(수)	0.962572**	0.933550*	1.000000	0.891529*	0.718363
1인당 전력소비량	0.840723	0.873138*	0.891529*	1.000000	0.936921**
온실가스 총 배출량	0.627269	0.690691	0.718363	0.936921**	1.000000

*p<0.05 **p<0.01 ***p<0.001



OTT 이용자의 온실가스 배출량 분석

☑ Step 1. 주평균 이용 시간 환산

주평균 이용시간: 6.0 시간 = 360분



주평균 이용횟수 (상수): $c = 360 \div 48.6 = 7.4$ (회)



| 회 평균 이용 시간 m (분)인 사람의 주평균 이용시간

$$t = \frac{c \times m}{60}$$
 (시간)

 $= t_{\text{-}}$ $+ t_{\text{-}}$ $+ t_{\text{-}}$ $+ t_{\text{-}}$ $+ t_{\text{-}}$ $+ t_{\text{-}}$

☑ Step 2. 온실가스 배출량 산정

 t_i : 매체별(해상도별) 이용 시간

 R_i : 해상도별 데이터트래픽 전송량 (GB/h)

표준 해상도	비트레이트 (GB/h)
480p	0.45
720p	1.2
1080p	1.8
	480p 720p

 γ : 데이터센터 에너지강도 ($\gamma = 0.049 \text{ kWh/GB}$)

ho : 배출 계수 ($ho=0.466~{
m kgCO_2eq/kwh}$)

$$G = \sum_{i=1}^3 t_i imes R_i imes \gamma imes
ho$$
 (kgCO₂eq)



OTT 이용자의 온실가스 배출량 분석

☑ 기술통계 및 T-test & ANOVA

'한국미디어패널조사'응답자 10302중 OTT 이용 경험이 있는 7027명의 데이터 활용

기술 통계랑	온실가스배출량 (추정치)
count	7027
mean	0.070707
std	0.101938
min	0.002535
25%	0.025346
50%	0.038019
75%	0.076037
max	2.433191

Q1. 성별에 따른 차이?

Q2. 태블렛 보유 유무에 따른 차이?

Q3. OTT 유료결제 이용 여부에 따른 차이?

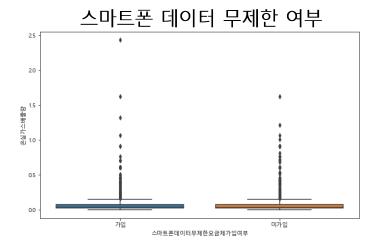
Q4. OTT 이용 컨텐츠에 따른 차이?

Q5. 스마트폰 데이터 무제한 요금제 가입여부에 따른 차이?

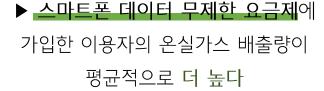


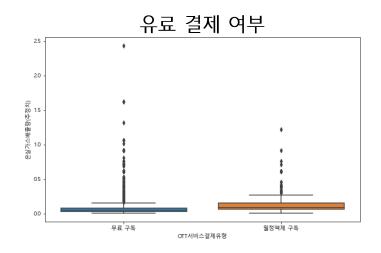
OTT 이용자의 온실가스 배출량 분석

☑ T-test (T 검정)



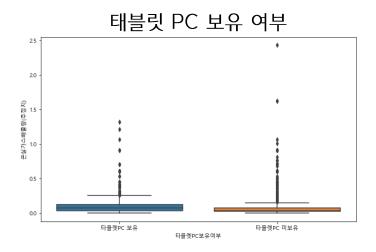
T statistic = 42.115 p value = 9.196e-11 < 0.01.





T statistic = 71.262 p value = 3.767e-17 < 0.01





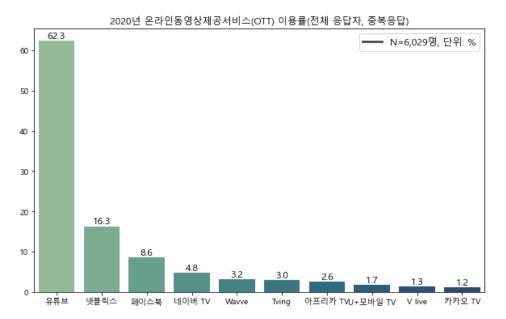
T statistic = 100.544 p value = 1.665e-23 < 0.01

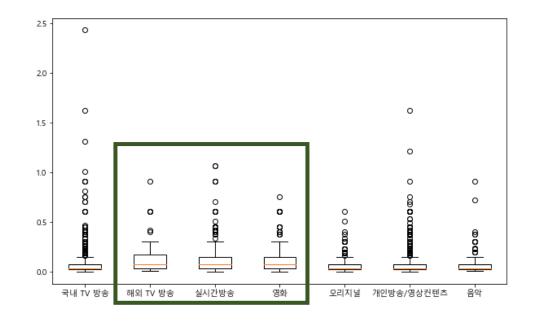
▶ 태블릿 PC를 보유한 이용자의 온실가스 배출량이 평균적으로 더 높다



OTT 이용자의 온실가스 배출량 분석

☑ ANOVA 분산분석, 다중 비교 및 사후분석





분산분석 결과 ▶분류간 평균의 차이가 있다

	df	sum_sq	mean_sq	F	PR(>F)
C(OTT컨텐츠I순위)	7.0	2.909981	0.415712	41.624796	7.208718e-58
Residual	7019.0	70.099546	0.009987	NaN	NaN



3. 분석 결과 및 내용

OTT 이용자의 온실가스 배출량 분석

ANOVA 분산분석, 다중 비교 및 사후분석

Multipl	Le	Col	mp	ar	1	50	n	О)Ť	P	1e	ar	าร	-		Ιu	Ke	y	Н	SL),	h	M	Ė	{=	0	. 6	15	
======		==:	==	==	=	==	=	==	=	==	==	=:		==	==		==	=	==	==	=	==	=	-	==	=:	-	=	=
1			- 2					2.4	ı				42			1 -										_	٠.		_

=====						
group1	group2	meandiff	p-adj	lower	upper	reject
1	2	0.0771	0.001	0.0427	0.1114	True
1	3	0.0549	0.001	0.0399	0.07	True
1	4	0.0565	0.001	0.0384	0.0746	True
1	5	-0.0048	0.9	-0.0195	0.01	False
1	6	-0.0057	0.3695	-0.0137	0.0024	False
1	7	0.0134	0.6054	-0.0099	0.0366	False
2	3	-0.0221	0.5462	-0.0586	0.0143	False
2	4	-0.0206	0.6564	-0.0584	0.0172	False
2	5	-0.0819	0.001	-0.1182	-0.0456	True
2	6	-0.0828	0.001	-0.1169	-0.0486	True
2	7	-0.0637	0.001	-0.1043	-0.0232	True
3	4	0.0016	0.9	-0.0203	0.0234	False
3	5	-0.0597	0.001	-0.0789	-0.0406	True
3	6	-0.0606	0.001	-0.0753	-0.046	True
3	7	-0.0416	0.001	-0.0679	-0.0153	True
4	5	-0.0613	0.001	-0.0829	-0.0397	True
4	6	-0.0622	0.001	-0.08	-0.0444	True
4	7	-0.0432	0.001	-0.0713	-0.015	True
5	6	-0.0009	0.9	-0.0152	0.0135	False
5	7	0.0181	0.3859	-0.008	0.0443	False
6	7	0.019	0.1834	-0.004	0.0421	False

	해외 TV 방송	실시간방송	영화	오리지널	개인방송/ 영상 컨텐츠	음악
국내 TV 방송	6.972e-10*	1.404e-26*	1.379e-35*	8.956e-01	4.267e-01	2.387e-01
해외 TV 방송		4.267e-01	8.962e-01	4.214e-08*	3.115e-11*	1.343e-04*
실시간방 송			2.336e-02	1.673e-14*	4.897e-33*	1.188e-04*
영화				7.677e-23*	1.440e-41*	1.566e-09*
오리지널					4.267e-01	4.328e-01
개인방송/ 영상 컨텐츠						4.4232e-02

▶ 해외방송, 실시간방송, 영화가 주 이용 컨텐츠인 이용자들이 그 외의 이용자들에 비해 온실가스 배출량이 평균적으로 높다



선형회귀분석을 이용한 분석

- ☑ 이전의 데이터로 구한 2020년 예측치
 - 1인당 전력 소비량: 11024.696 kwh
 - 1인당 에너지 소비량: 238.180 백만 TOE
 - 온실가스 총 배출량: 72376 만 톤
 - 국내 OTT 시장 매출액: 655.571 억 원
 - 국내 데이터센터 수: 169개

1인당 전력 소비량(1990, 2000, 2010-2019년)

0	1990	2202	<pre>line_fitter = LinearRegression()</pre>
1	2000	5067	line_fitter.fit(x.values.reshape(-1,1), y)
2	2010	8883	LinearRegression()
3	2011	9142	
4	2012	9331	line_fitter.predict([[2020]])
5	2013	9285	array([[11024.69603045]])
6	2014	9305	
7	2015	9555	line_fitter.coef_
8	2016	9699	array([[283.08907015]])
9	2017	9869	
10	2018	10195	line_fitter.intercept_
11	2019	10039	array([-560815.22566612])

1인당 에너지 소비량(2011-2019년)

0	2011	205.2	<pre>line_fitter = LinearRegression() line_fitter.fit(x.values.reshape(-1,1), y)</pre>	
1	2012	207.0	line_fitter.predict([[2020]])	
2	2013	208.4	array([[238.18055556]])	
3	2014	210.4	line_fitter.coef_	
4	2015	215.4	array([[3.99833333]])	
5	2016	221.9	line_fitter.intercept_	
6	2017	230.6	array([-7838.45277778])	
7	2018	233.4	plt.plot(x, y, 'o')	
8	2019	231.4	<pre>plt.plot(x,line_fitter.predict(x.values.reshape(-1,1)) plt.show()</pre>	

온실가스 총 배출량(2010-2019년) array([[723.76]]

국내 데이터센터 수(2002 -2019년) array([[6555.57142857]])

국내 OTT 시장 매출액 추이(2012년-2019년) array([[169.08496732]])





결론 및 개선 방안



CONCLUSION

온라인 동영상 서비스 이용 추이 및 데이터센터 전력 소비량 꾸준히 증가



- 온실가스 배출량에 직접적인 요인은 동영상 시청 시간(횟수), 영상의 해상도 등
- 무제한데이터 요금제 가입 여부, OTT서비스 유료구독 여부 등 개인의 행동에 따라서도 달라지기도 함



한계점 및 개선 방향

☑ 한계점

- 2020년 코로나19의 영향은 반영되지 않음
- 사용된 데이터의 수집된 기간이 짧음
- 온실 가스는 에너지 소비 외의 다른 요인에도 영향 有

☑ 개선 방향



개 인

- 잘 때는 영상 끄기(ASMR, 수면유도 영상 등)
- 가능한 다운 받아 오프라인으로 보기
- 고화질인 4K 대신 HD 선택 (데이터 낭비 30% 감소)

☑ 개선 방향

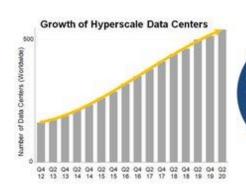


기 업

- 고지대, 고위도 지역 활용
- 친환경 에너지 활용
- 폐열을 지역난방에 활용
- 하이퍼스케일 데이터센터 설립



<폐열을 활용한 지역난방>





United

<하이페스케일 데이터센터>



친환경 데이터센터

☑ 재생 에너지 활용

- 석유나 석탄 같은 화석 연료로 만들어지는 전기의 사용을 줄이거나.
 전혀 사용하지 않는 형태
- 페어 네트웍스: 라스베이거스 사막에 데이터센터 설치
- 센터 외곽에 태양광 패널을 설치해서 태양력 에너지로만 운영
- 페이스북: 100% 풍력 발전으로 움직이는 친환경 데이터 센터
- '오픈 컴퓨트 프로젝트(OCP)'의 데이터센터 설계 기술을 도입해 자체적으로 냉각 효율 강화 (OCP: 외부업체와 함께 서버, 냉각 장비, 냉각 구조 등을 표준화)
- 저전력 저발열 서버를 도입해 100% 풍력 발전 가능











☑ 자연 환경 활용

- 구글: 하미나의 종이를 만드는 공장을 데이터센터로 개조
- 하미나의 풍부한 바닷물을 끌어들여 데이터센터를 식히는 데 사용
- 공기와 물을 모두 사용하는 하이브리드 냉각 방식을 적용

- 네이버: 2013년 강원도 춘천에 데이터센터 '각'설립
- 춘천: 전국에서 평균 기온이 가장 낮은 지역 중 하나 연평균 온도가 수도권보다 2℃가량 낮음
- 춘천 산지에서 내려오는 차가운 바람이 자연스럽게 서버실 내부의 열을 식히도록 설계



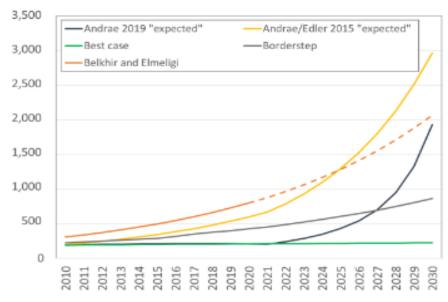
☑ 재생 에너지와 자연환경을 모두 이용한 혼합형

- 페이스북: 2013년 스웨덴 루레아(Luleå)에 데이터센터 설립
- 전체 전력 중 70%를 근처 루레아 강에 있는 수력발전소에서 공급
- 96km 거리의 북극의 바람으로 내부 장비의 열을 식혀 냉각 효율 극대화

- ☑ 2030년 전 세계 데이터센터 에너지 소비량 예측
- 2010년 2017년 데이터센터 에너지 소비량 기준

기술에 30분만 동영상 소비를 덜 했을 때, 기인당 연간83.2 kgCO₂eq 의 온실가스를 줄일 수 있다!











프로젝트 소감



□ 5. 프로젝트 소감

프로젝트 소감



분명 지난 프로젝트와 크게 다르지 않은 프로젝트였지만 부담도 크고 걱정도 큰 상태에서 시작한 프로젝트였습니다. 주제 선정부터 고생을 많이 해서 과연 무사히 끝낼 수 있을지 걱정도 많았지만 이렇게 잘 끝내게 되어 다행이라고 생각합니다. 시작할 때만 하더라도 시각화나 분석을 다시 공부하면서 잘할 수 있을까 했지만 팀원분들과 많은 의견 을 나누고 의논하며 즐겁고 가벼운 마음으로 진행할 수 있었 습니다. 배움도 많고 즐거움도 많았던 프로젝트였습니다.



저희는 몇 번이나 주제를 고쳐 생각하는 과정에서 우리는 수십개의 기사, 논문, 연구보고서 등을 읽어야 했습니다. 환경을 생각하는 우리 팀원들의 마음과, 셋 다 OTT 유료구독자라는 점에서, 또 SbS의 흥미로운 기사를 읽음으로써 겨우 저희 프로젝트 주제까지 도당할 수 있었습니다. 그 뒤에는 각자의 역할에 맞게 분석과 시각화를 하고 코드를 공유하며, 나름 가벼운 마음으로 또 즐겁게 임했어요. 코딩 기술을 넘어서, 데이터를 해석하고 그것을 통해 주장을 펼칠 수 있는 능력을 기르는 좋은 훈련이 되었습니다. 감사합니다!



데이터의 중요성과 명확한 기준의 필요성에 대해 알게 되었습니다. 다양한 주제가 논의되었지만 해당 데이터의 여부와 상태에 따라 프로젝트의 진행에 큰 영향을 끼치는 것을 알게 되었기 때문입니다. 수집한 데이터의 기간이나 속성이 더 많았더라면 좀 더 정확하고 다양한 분석 결과를 도출해냈을 것 입니다. 마지막 전공 프로젝트인 만큼 여러 가지의 기법을 활용해 분석하려고 했고, 많은 사람들이 관심 가지면 좋을 법한 주제를 선정했습니다. 모두 수고하셨습니다~



5. 프로젝트 소감

출처

- 환경부
- 한국전력공사
- 방송통신위원회
- 방송매체 이용행태 조사 참조
- 한국데이터센터연합회
- 에너지경제연구원, 과학기술정보통신부





