

# SUSTAINABLE SMART CITY, ESG, AND THE FUTURE

Young Hoon KWAK, Ph.D.

Director, International Center for Project Management

School of Business

The George Washington University

May 9, 2023

# Global Mega Trends (Frost & Sullivan, 2022)

## Global Mega Trends



Urbanization –  
City as a  
Customer



Bricks and Clicks



Future  
Infrastructure  
Development



Smart is the New  
Green



Innovating to Zero



Health, Wellness  
and Well Being



Social Trends: Gen Y,  
Middle Bulge, She-  
economy,  
Geosocialization



Future of Energy



Future of Mobility



Connectivity and  
Convergence

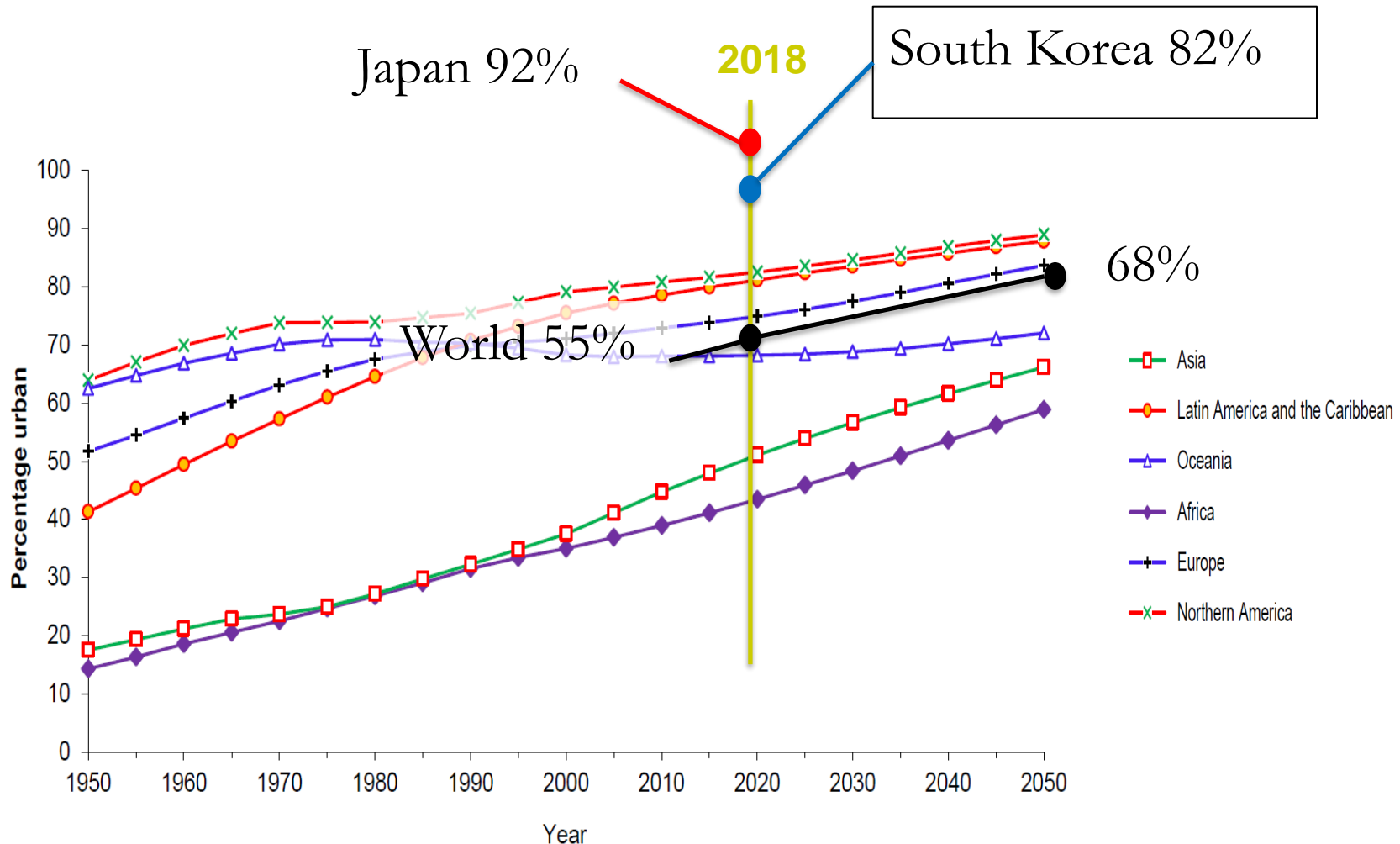


Economy: Beyond  
BRIC: The Next  
Game Changers



New Business  
Models: Value for  
Many

# World Urbanization (United Nation, 2018)



# Urban Problems



- Lack of infrastructure, land supply, energy problems
- Man-made pollutions including garbage, sanitary, fire, terrorism, traffic congestion, crime, viruses, etc.

# 17 Sustainable Development Goals

(United Nations, 2015)

THE GEORGE  
WASHINGTON  
UNIVERSITY

WASHINGTON, DC



UN in collaboration with Project Everyone (2015)



# Solution to Urban Problems

Application of Technology could solve Urban Problems



Technology

City

Smart city

# What is Smart City?







(International Telecommunication Union, 2016)



- **A smart sustainable city** is an innovative city with ICTs
- **To improve quality of life** for efficiency and competitiveness
- **Economic, Social, Environmental** for the future generations

# Smart City Layers

(Kwak and Lee 2022)

Layers		Contents
<b>Digital Space</b> Services	 	Traffic control, water treatment, garbage disposal, energy solution, security, healthcare, Car sharing, etc.
<b>Media</b> ICT Infrastructure	 	Internet service, Wi-fi networks, CCTV, Fiber optic networks, multipurpose sensor networks, service-oriented information system, integrated control center, Communication network, etc.
<b>Physical Space</b> Physical Infrastructure	 	Buildings, residences, roads, public transportation, electrical grid, etc.



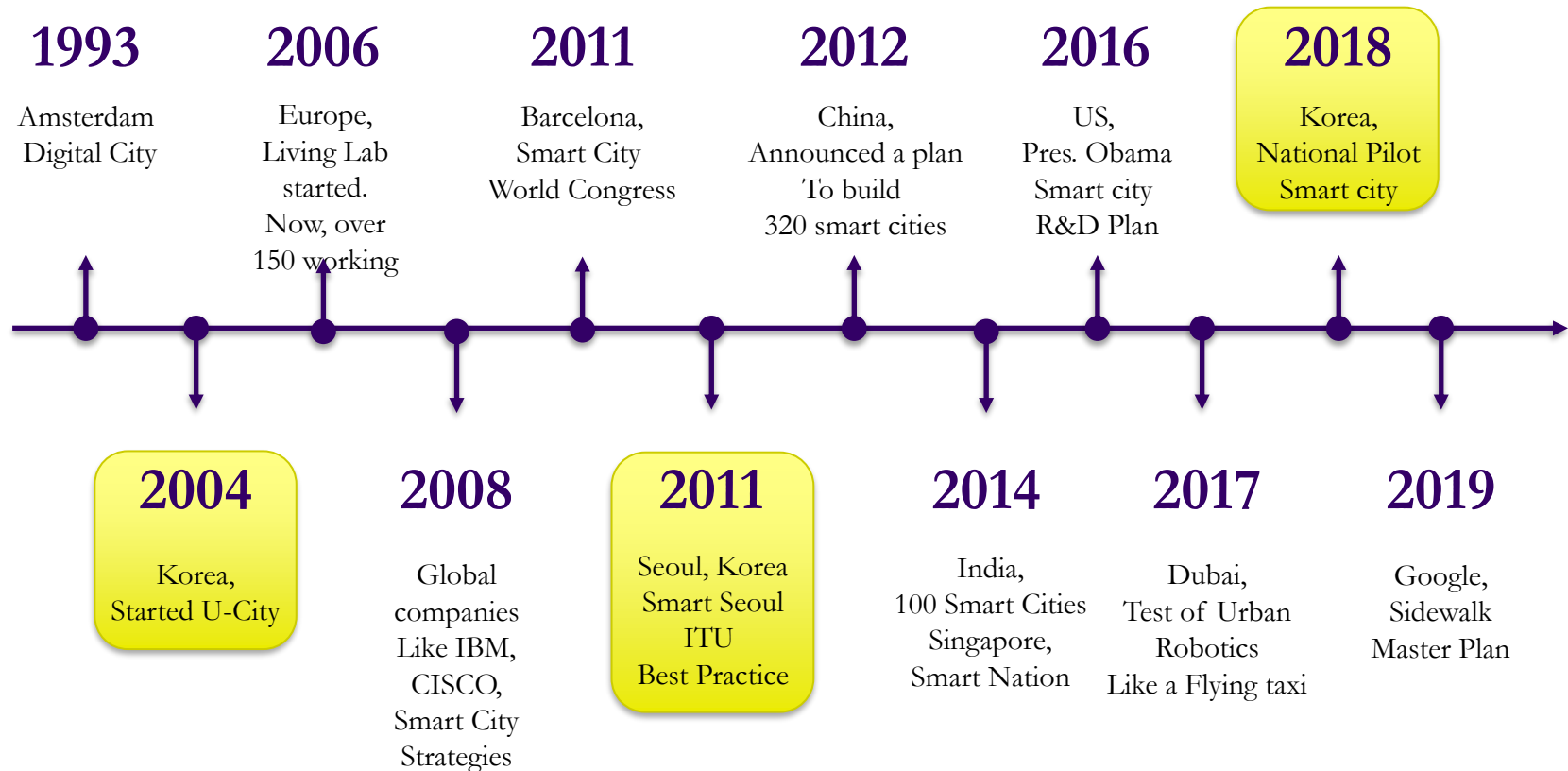
# Technologies are “NOT” Smart City Challenges

- Governance & leadership
- Politics & Government relations
- Citizen Engagement
- Business infrastructure
- Long-term life cycle
- Multi Stakeholders
- Complexity



World Bank Group, Korea Innovation Week (2020)

# History of Worldwide Smart Cities



Adopted from World Bank Group, Korea Innovation Week (2020)

# Korean Smart City is in its 3<sup>rd</sup> Generation!

Classification	1st generation (2009~2013)	2nd generation (2014~2018)	3rd generation (2019~)
Information	Vertical data integration	Horizontal data integration	Interoperability
Platform	Closed type (silo)	Partially open	Partially open (expendable)
Subject	Central government	Central government + Local government	Central government + Local government
Business character	Building infrastructure - ICT infrastructure - Construction infrastructure - Public Service	Platform building Standardization - Integration - Organization - Institution - Function - Public Service	Innovation system Public Service+ Private Service

# Examples of Korean Smart City Costs

(Based on 5 five local governments)

Construction costs account for a large portion of Korea's 1st-gen. smart city



**Construction  
Infrastructure**  
96.8%



**ICT  
Infrastructure**  
3.0%



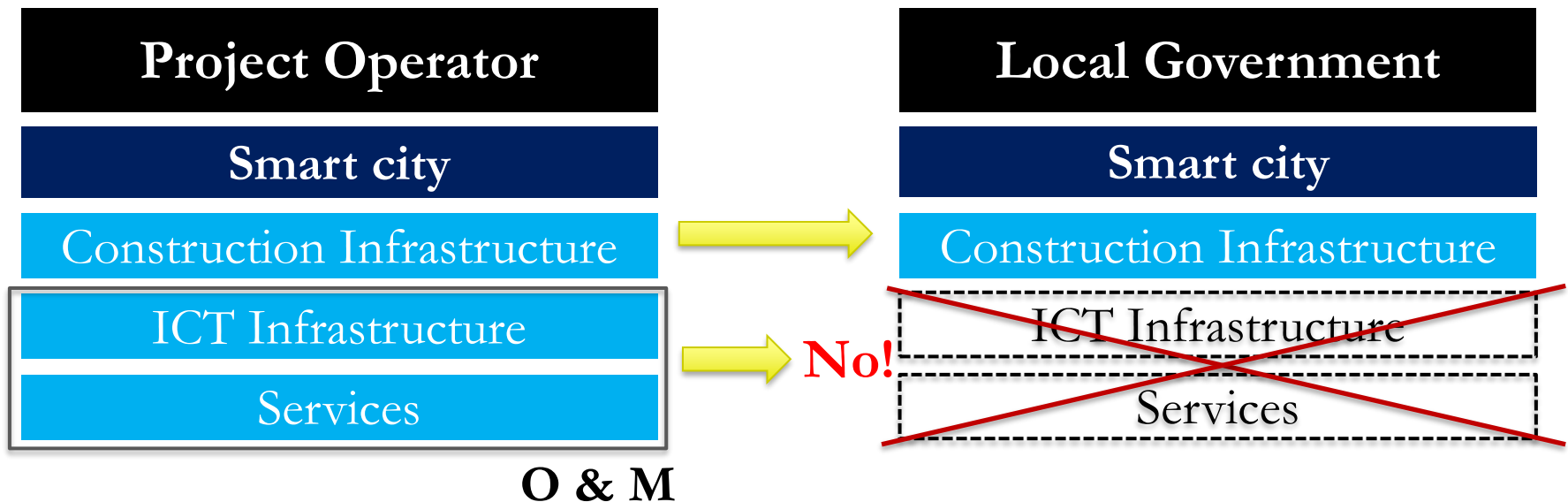
**Annual  
O&M**  
0.2% (million KRW)

District	Project duration	Population	Construction Infrastructure costs	ICT Infrastructure costs	Annual service O&M costs	Annual service operating revenue
Hwaseong Dongtan	2001~2008	120,000	3,263,000	45,000	3,170	-
Paju Gyoha	2006~2009	126,000	5,069,000	90,000	5,900	1,400
Suwon Gwanggyo	2005~2011	77,000	5,088,100	100,000	5,400	1,000
Sejong	2006~2030	200,000	7,600,000	409,800	26,500	6,500
Seongnam Pangyo	2005~2010	87,000	2,224,300	80,000	3,900	-

# Operations & Maintenance (Service)

## Project Lifecycle Perspective

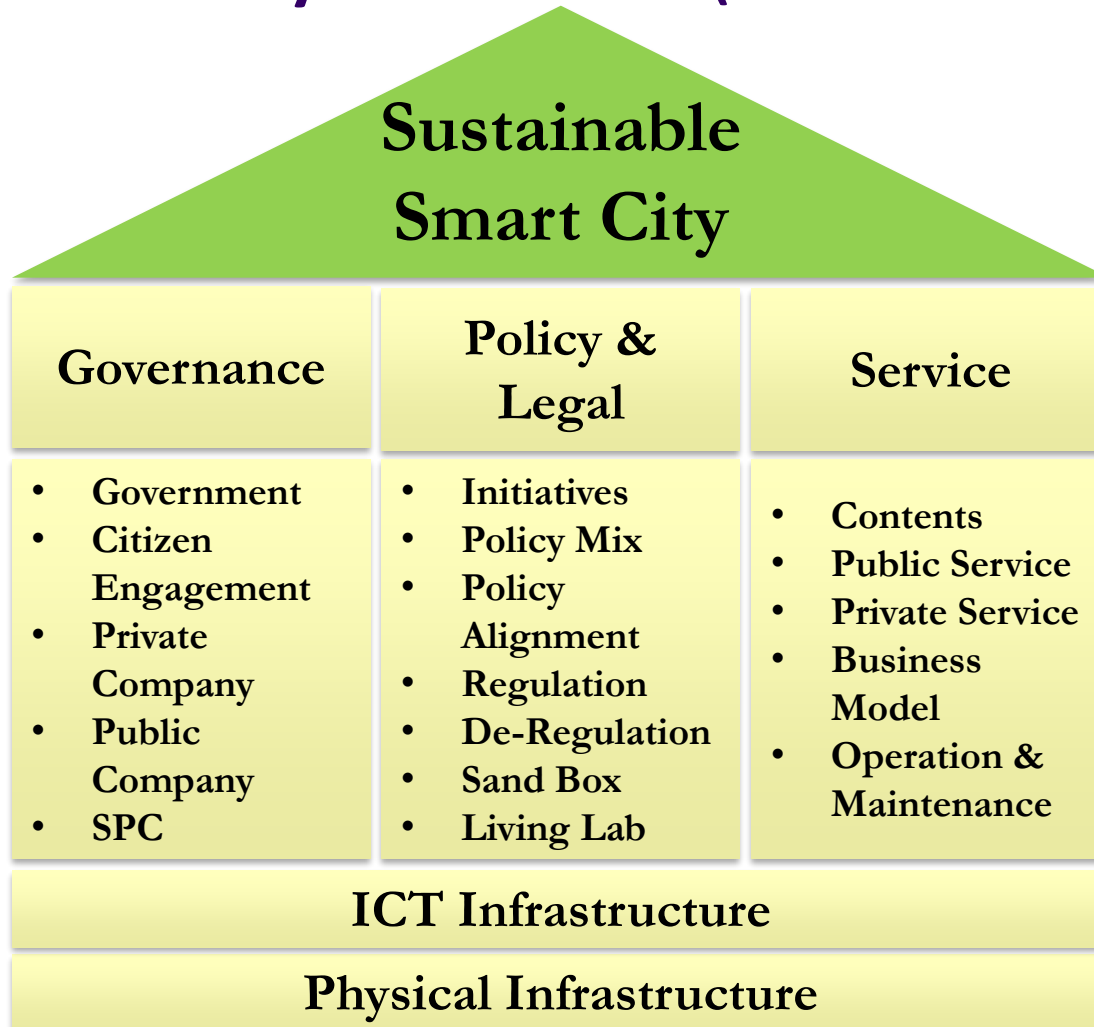
After completion of construction.



- Unplanned O&M costs
- Low revenues from the services operation
- Lack of Institutional support



# Smart City Framework (Kwak and Lee, 2022)

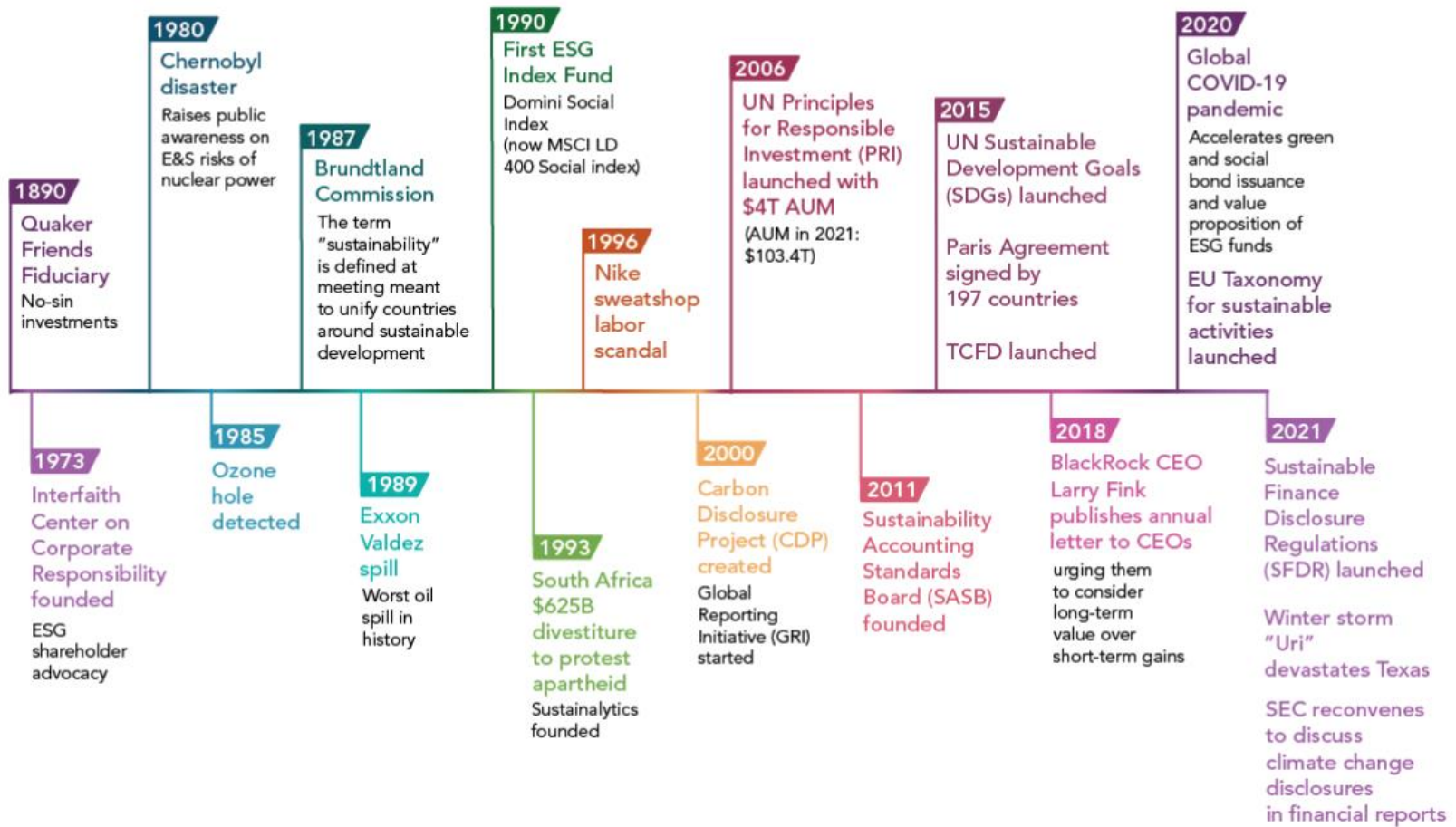


Y. H. Kwak and J. Lee, "Toward Sustainable Smart City: Lessons From 20 Years of Korean Programs," in *IEEE Transactions on Engineering Management*, vol. 70, no. 2, pp. 740-754, Feb. 2023, doi: 10.1109/TEM.2021.3060956.

# ESG (Environmental, Social, and Governance)

- Environmental
  - 탄소배출량과 기후변화
  - 친환경 경영
- Social
  - 인적 자본
  - 작업 조건
  - 회사의 명성, 인권, 안전 등
- Governance
  - 기업 지배 구조 및 기업 행동 표준
  - 이사회와 다양화와 성별, 임금격차, 기업 윤리 및 투명성

# Evolution of ESG



<https://www.dnb.co.uk/perspectives/supply-chain/esg-data-valuable-for-supply-management-procurement.html>

## ESG Practices (USA)

### BlackRock

“투자 의사결정시 기후변화로 인한 위기 대응에 동참하지 않거나 수익의 25% 이상을 석탄 생산으로 창출하는 기업은(석탄 에 많이 의존하는 기업) 향후 직접 투자를 하지 않을 것이며 investment portfolios에서 삭제하는 절차를 밟을 것” -출처: Black Rock Letter to CEOs (2020.01)



"모든 기업이 기후변화와의 싸움에 동참해야 하며, 녹색혁신이 가져올 수 있는 모든 기회를 위해 공급업체와 지역사회는 함께 해야 한다. 우리는 긴급하게, 함께 대처하고 있다. 재생 가능한 자원, 더 친환경적이고 공평해지기 위해 신속하게 행동해야 한다." -출처: Apple Newsroom (2022.08)

### Alphabet

(Google, Youtube와 같은  
주요 기업의 모회사)

“기후 변화에 대처하는 것부터 인종 및 경제적 형평성 등 모든 배경의 사람들이 보다 지속 가능한 세상에서 살 수 있도록 돕는 데 전념해 왔다. 우리는 환경, 사회 및 거버넌스(ESG) 이니셔티브에 57억 5천만 달러를 기부할 계획이다.” -출처: Sustainability magazine (2022.08)

# ESG Practices (Korea)

## SK(주), 그룹 차원 ESG 전략 담은 '지속가능경영보고서' 발간

넷제로 달성 위한 비즈니스 혁신 정리  
SK그룹 ESG 플랫폼으로 확대 예정

조재범 기자 입력 2022-08-19 09:50 | 수정 2022-08-19 10:34

## 삼성엔지니어링, 올해 그린솔루션·환경인프라 신사업 780억 투자

## 포스코, 철강사업 ESG 경영성과 담은 '2021 기업시민보고서' 발간

철강 사업 전문회사 전환 후 첫 '기업시민보고서' 발간...100년 기업 향한 의지 담아

박영재 기자 기사입력 2022/07/08 [15:07] 본문듣기

## 한화, 계열사에도 'ESG 위원회' 설치... 분기마다 경영 점검

김유라 기자 승인 2022.07.07 14:34 댓글 0

- 한화그룹 'ESG위원회' 출범 1주년

## 'ESG 경영' 속도내는 LG...엔솔·이노텍·전자 잇따라 RE100 합류

서민지 기자 입력 2022.08.17 14:59

## GS리테일, 통합 후 첫 지속가능 경영보고서 발간...ESG 전략 담아

황병우 기자 승인 2022.07.04 14:23 댓글 0

지난해 이어 2번째, 통합 이후 첫 발간, 첫 여성 사외 이사 선임 등 변화 선보여  
ESG전략 및 실행 체계, 각 분야별 실행결과 등 7대 전략 과제와 함께 담겨

PEOF

## KT, 2022년 ESG 보고서 발간...전사 차원의 ESG 경영 실천 성과 집대성

최명진 기자 ugala7@fetv.co.kr 등록 2022.07.25 09:00:57 수정 2022.07.25 09:01:02

본문듣기

## 현대건설, ESG 정보 공개 확대 '2022 지속가능경영보고서' 발간

매일경제 메인 추가

정의선 GO, 현대차 GO...지속가능 미래 화두 'ESG'

입력 2021.05.21, 오후 1:47



## ESG Practices (Korean Financial Sector)

- **금융위원회:** ESG 공시를 단계적으로 **의무화** 하기 위한 방안을 2021년 발표. (지속가능경영 보고서)

적용시기	'19년	'22년	'24년	'26년
자산총액	2조원 이상	1조원 이상	5천억원 이상	전 코스피 상장사

출처: 금융위원회(2021)

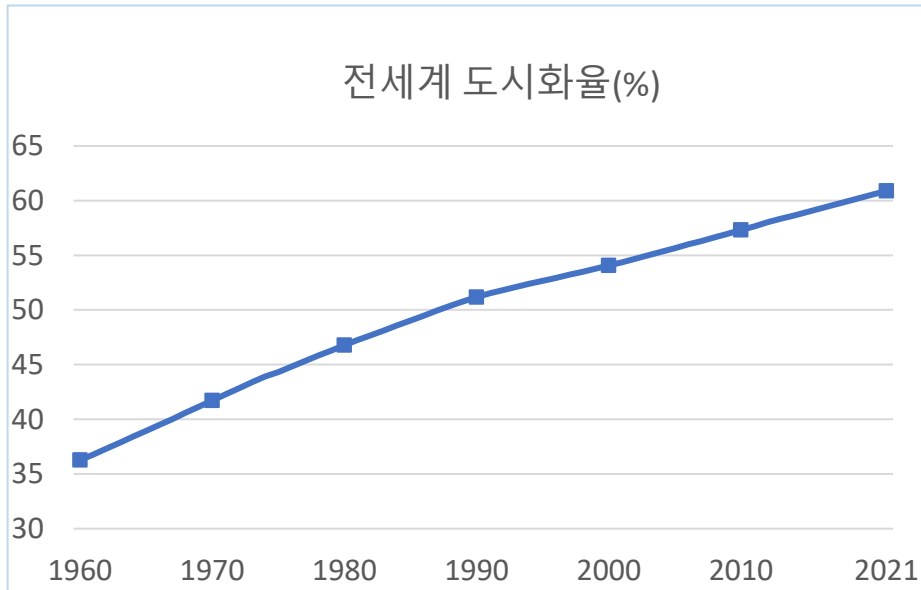
- **한국거래소:** 'ESG 정보공개 가이드스' 단계별 의무화 방안 발표.

단계	1단계	2단계	3단계
적용시기	~'25년	'25~'30년	'30년~
내용	'ESG정보 가이드스' 를 제시하여 '지속가능경영 보고서'자율공시를 활성화	일정규모 이상 코스피 상장사에 대하여 공시 의무화	전체 코스피 상장사에 대하여 공시 <b>의무화</b>

출처: 한국거래소(2021)

# Strong Needs for Smart City

2050년까지  
세계인구 70%가 도시에 몰릴 것으로 예상



데이터 출처: The World Bank  
([Urban population \(% of total population\)](#) | [Data \(worldbank.org\)](#))

도시화로 인한 문제들 발생

- 인구 밀도 증가
- 자원 소비 증가
- 온실가스 배출량 증가
- 고품질 공공서비스의 필요성 대두

부정적 측면의 대응전략으로서  
'Smart City' 관심

# Can ESG concept apply to Smart City?

스마트시티의 필수 목표:  
환경, 시민의 삶 등  
지속가능한 도시

‘지속가능도시’의 개념을 담고  
있어 ‘지속가능경영활동’인  
ESG가 필요

ESG는 지속가능한 미래를  
만드는데 필요한 척도



스마트시티를 구축하고, 더 나은 삶의 질과 깨끗한 환경 및 지속가능한 스마트 시티의 계획, 건설 및 운영을  
제공하려면 스마트 시티 참여기업의 ESG 성과를 아는것이 필요



**그렇다면, Smart City의 ESG수준을 측정해볼 수 있는가?**

# Can we elicit Smart City ESG score from participating firms' ESG score?

## AS-IS

- 스마트시티의 ESG 수준을 살펴보거나 이를 이해하기 위한 특정 지표, 가이드라인에 대한 연구 **전무**
- 스마트시티 관점에서 ESG수준을 **탐색한 연구는 없었음**



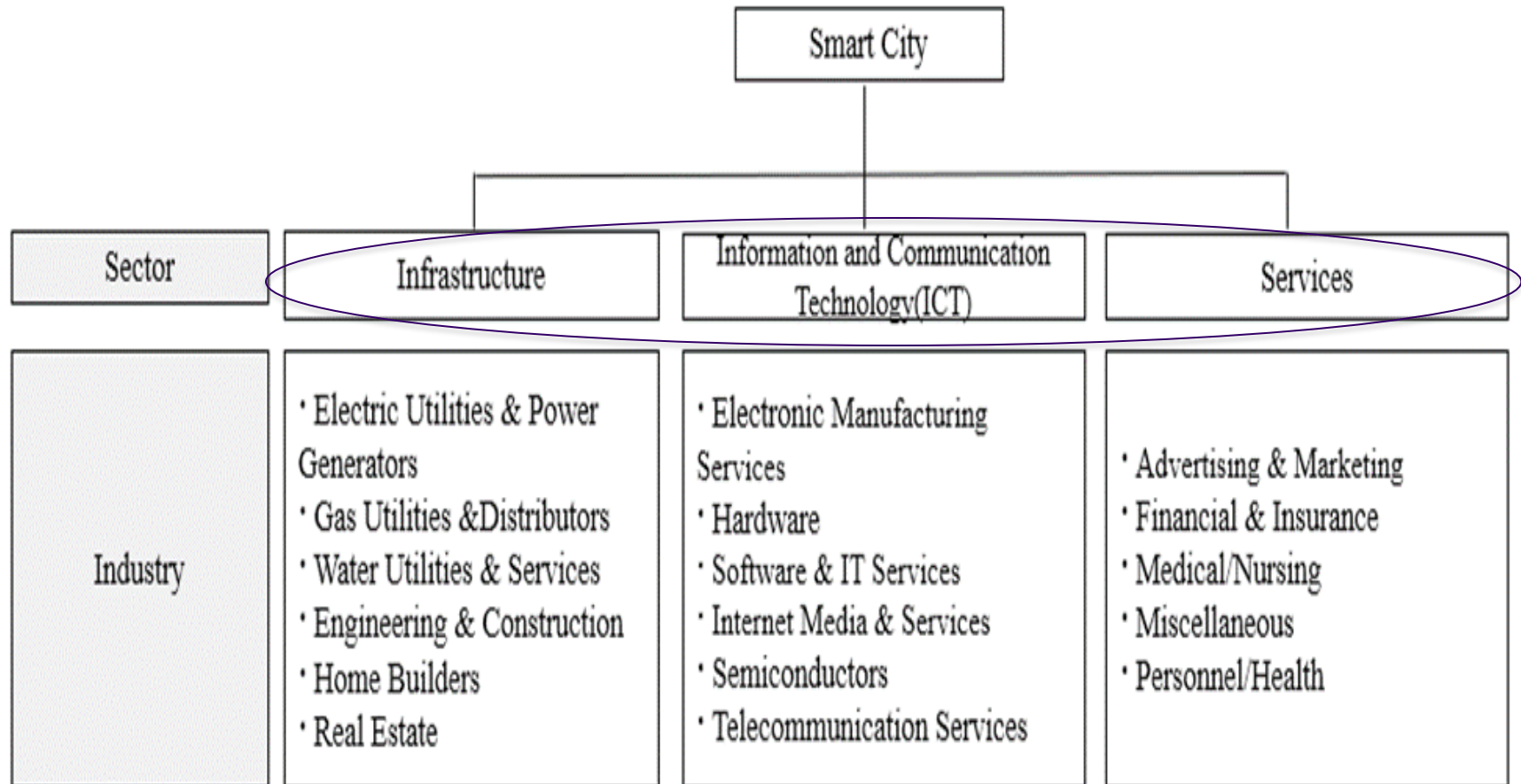
## TO-BE

- **현재 스마트시티의 ESG 수준 제공**  
: 3가지 산업부문의 데이터를 활용하여 ESG수준에 대한 포괄적 개요 제공
- **기존에는 없던 내용에 관해 시도**  
: 스마트시티의 ESG 수준을 추론할 수 있는 기본적 모델식 개발

Smart City의 ESG 수준을 추정하기 위한 기본적 가이드라인 제공

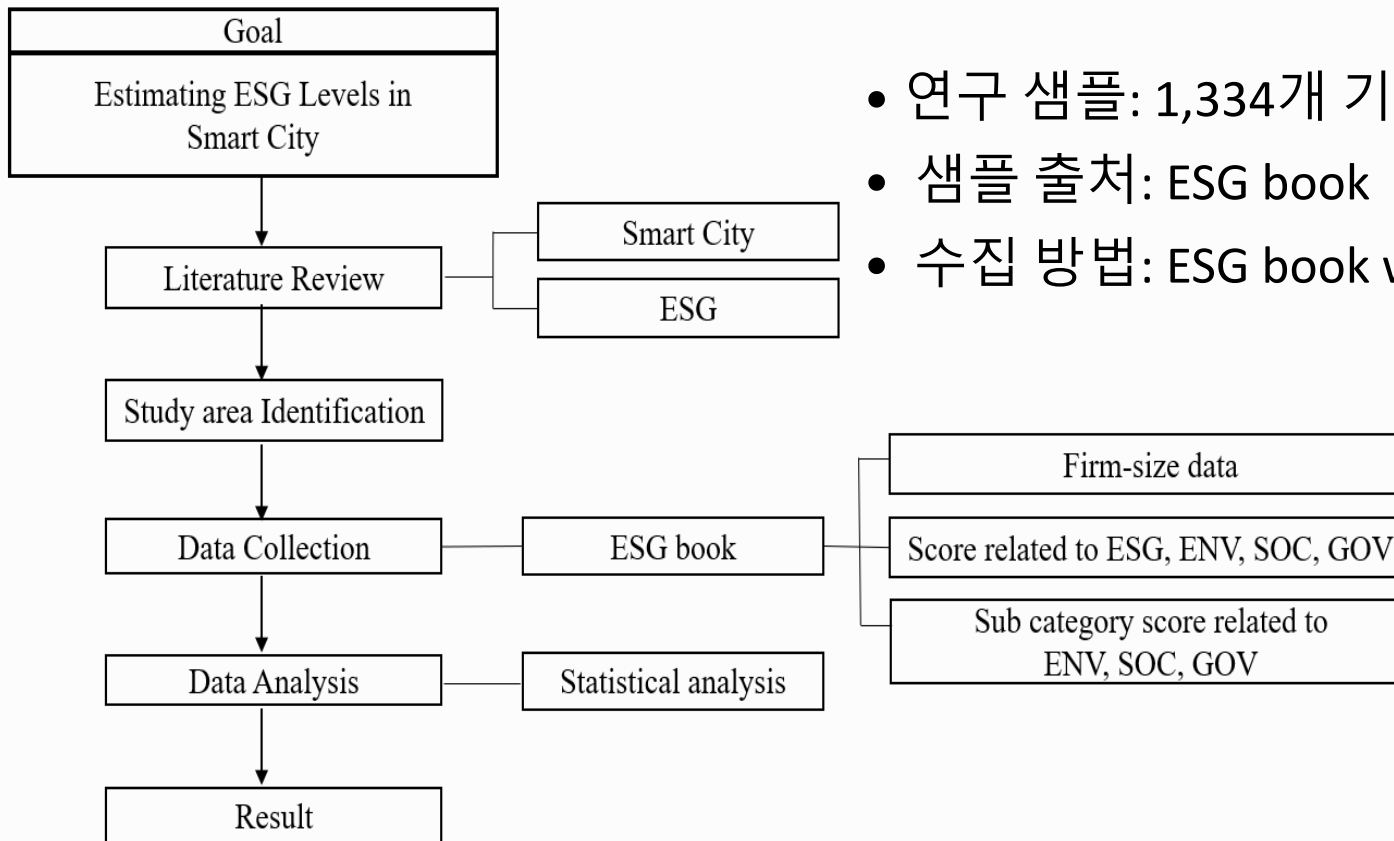
# Integrating Three sectors' ESG for Smart City

## Sustainability Accounting Standards Boards (SASB)





# Research Approach



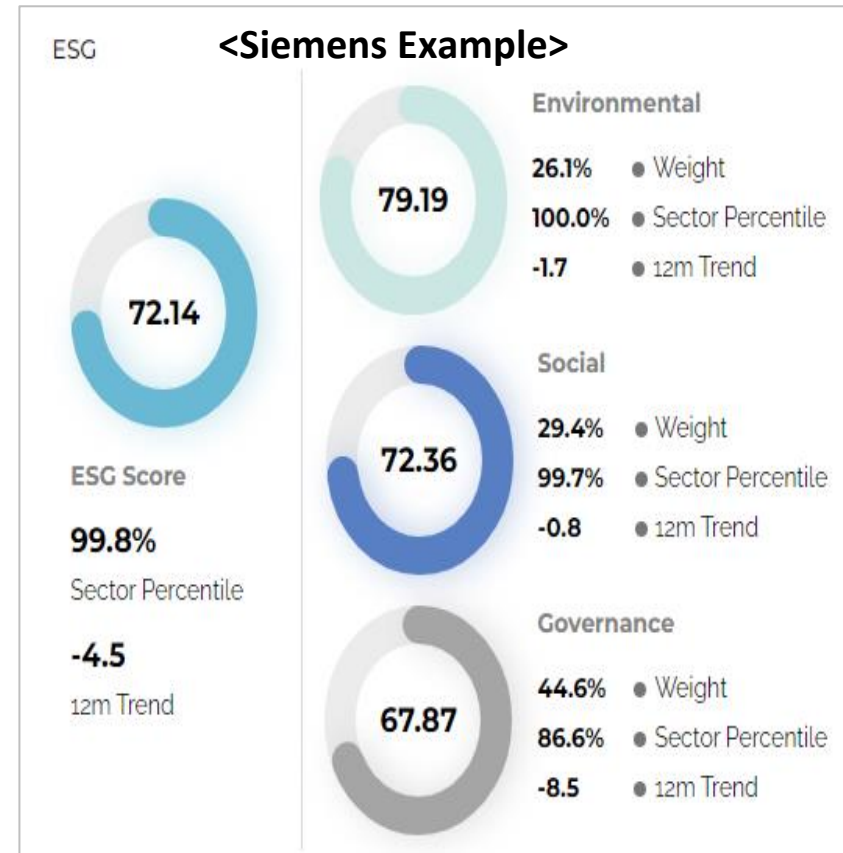
- 연구 샘플: 1,334개 기업 (3개 산업)
- 샘플 출처: ESG book
- 수집 방법: ESG book web scraping

# Data from ESG Book

## (December 2021 data)

Size	Total \$
Micro	50 million to 300 million
Small	300 million to 2 billion
Mid	2 billion to 10 billion
Large	10 billion to 200 billion

Criteria used for our measurement		Criteria used for our measurement		Criteria used for our measurement	
Waste	80.33	Compensation	49.57	Business Ethics	47.93
Water	84.82	Diversity	78.26	Corporate Governance	62.11
Emissions	79.43	Employment Quality	80.59	Transparency	65.36
Resource Use	72.96	Human Rights	87.09	Capital Structure	91.87
Environmental Management	75.12	Labour Rights	71.01	Forensic Accounting	56.60
Environmental Stewardship	88.16	Occupational Health and Safety	72.34		
Environmental Solutions	79.57	Product Access	78.25		
		Product Quality and Safety	65.87		
		Training and Development	71.72		
		Community Relations	79.09		



# Descriptive Statistics of 1334 Infrastructure + ICT + Service Firms

- Infrastructure 42%(n=554), ICT 34%(n=454), Service 24%(n=326).
- Average ESG score was 53.5 (ranging from 29.7-72.9).
- ENV average score was the highest (56.3); SOC was 56.1
- GOV was the lowest (50.2) with 60 out of 1334 (~5%) were below 30.

Variable	N	Mean	SD	Minimum	Maximum
ESG	1334	53.5	7.30	29.7	72.9
ENV	1334	56.3	12.68	28.6	86.8
SOC	1334	56.1	7.76	30.9	74.2
GOV	1334	50.2	12.21	16.2	80.8

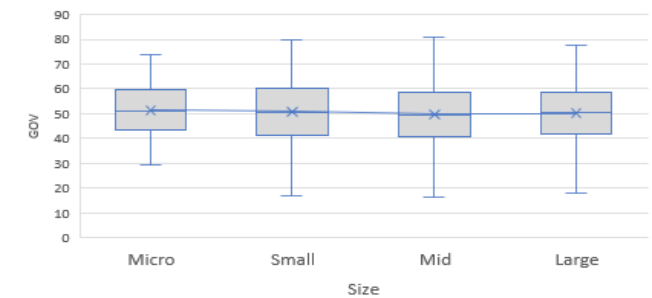
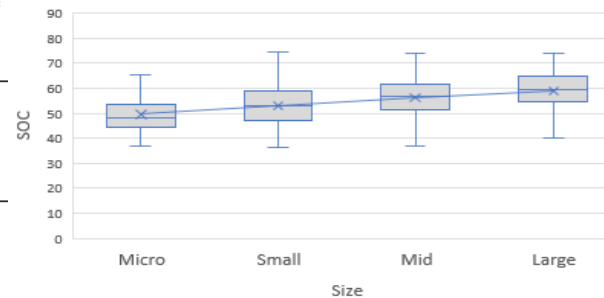
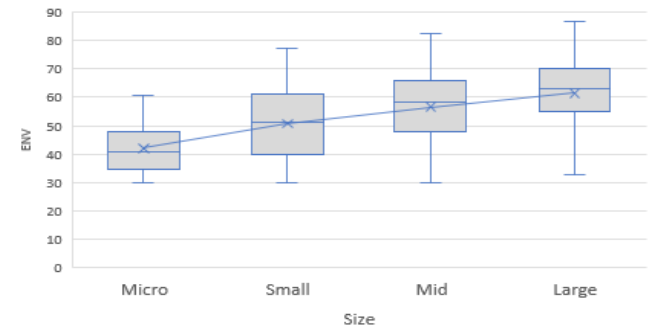
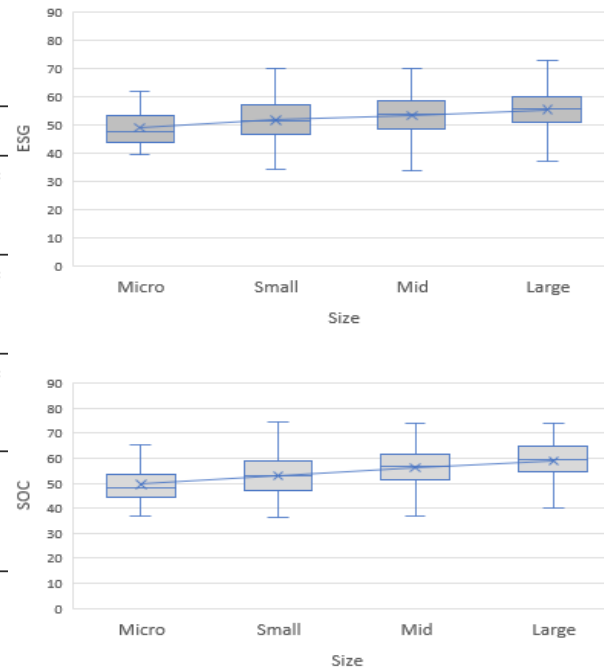
# ESG scores of Infrastructure, ICT, and Service based on the Firms' Sizes

Size	ESG			ENV		
	Infra.	ICT	Service	Infra.	ICT	Service
Micro	49.25	51.26	46.59	44.85	41.39	38.99
Small	52.60	51.35	50.91	55.76	48.41	45.75
Mid	54.26	53.65	51.27	60.01	56.84	49.19
Large	55.23	56.15	54.14	63.11	62.24	57.39
Size	SOC			GOV		
	Infra.	ICT	Service	Infra.	ICT	Service
Micro	50.72	48.41	49.18	50.83	56.50	47.36
Small	54.48	51.96	51.94	49.50	51.97	51.72
Mid	56.48	57.18	54.55	49.38	50.04	50.19
Large	58.58	59.09	54.14	48.36	51.59	50.58

# Smart City ESG vs. Firm Size

- ESG, ENV, SOC has positive correlation between companies' size and relative ESG score. GOV has no correlation between companies' size and GOV score.

Variable		Size
ESG	Pearson's r	0.208***
	p-value	<.001
ENV	Pearson's r	0.356***
	p-value	<.001
SOC	Pearson's r	0.309***
	p-value	<.001
GOV	Pearson's r	-0.022
	p-value	0.426





## Smart City ESG vs. Return on Assets (ROA)

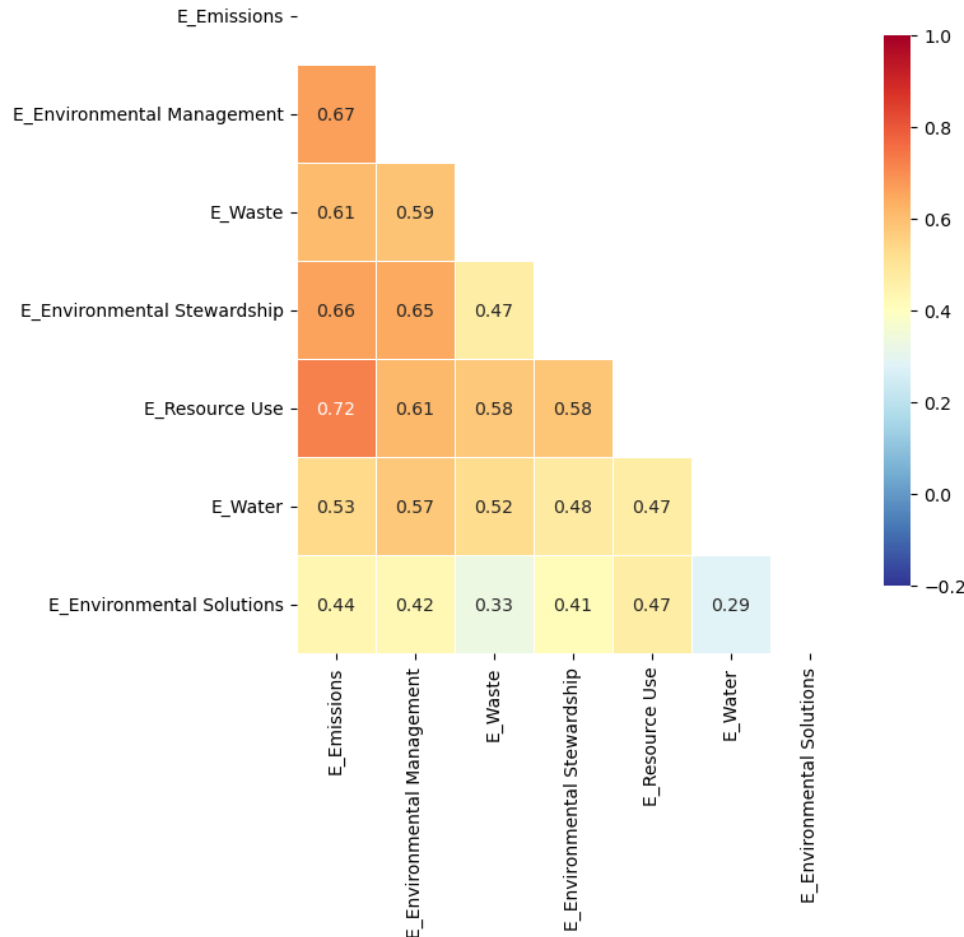
- Looking into financial significance (Investment in Smart City)
- GOV only had the statistical significance to Return on Assets (ROA)!

Variable		ROA
ENV	Pearson's r	0.064
	p-value	0.146
	N	510
SOC	Pearson's r	0.037
	p-value	0.400
	N	510
GOV	Pearson's r	0.227***
	p-value	<0.001
	N	510

## Sub-Categories of ESG Measurement

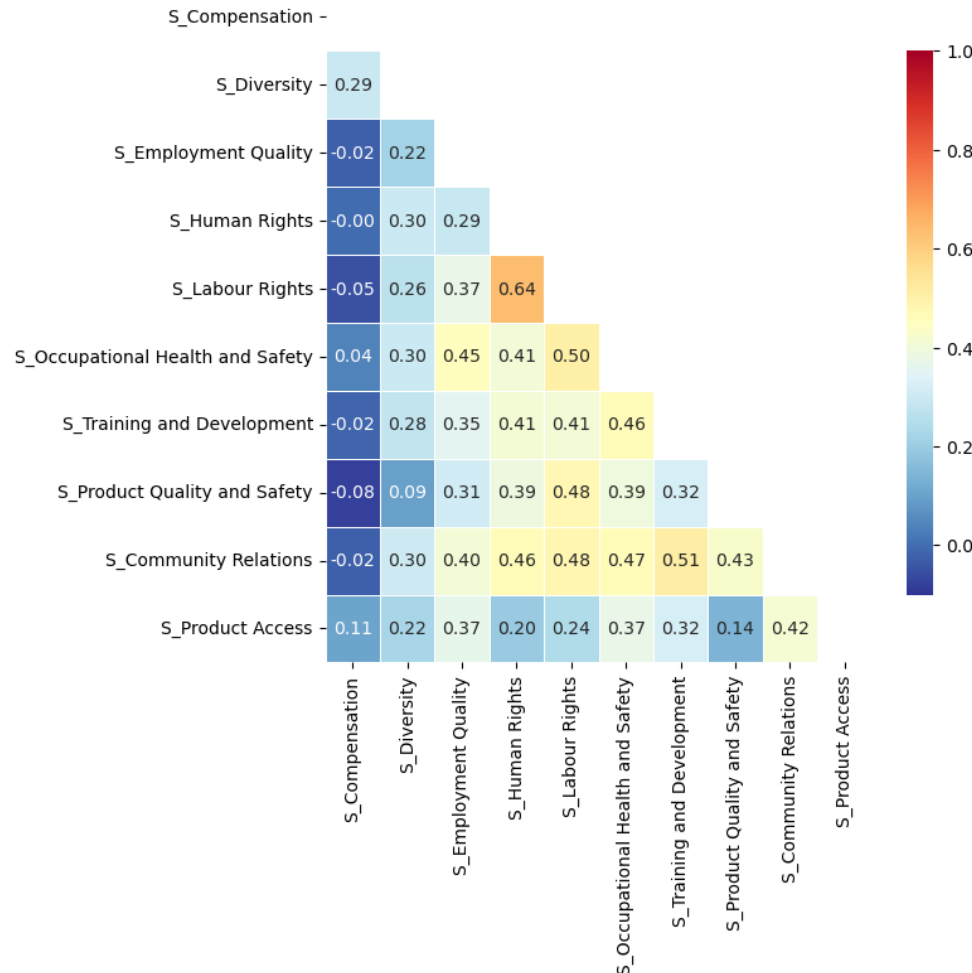
ENV	SOC	GOV
Emissions	Compensation	Business Ethics
Environmental Management	Diversity	Capital Structure
Waste	Employment Quality	Corporate Governance
Environmental Stewardship	Human Rights	Transparency
Resource Use	Labor Rights	Forensic Accounting
Water	Health and Safety	
Environmental Solution	Training and Development	
	Product Quality and Safety	
	Community Relationship	
	Product Access	

# Environment Score Heatmap (Correlation check included)

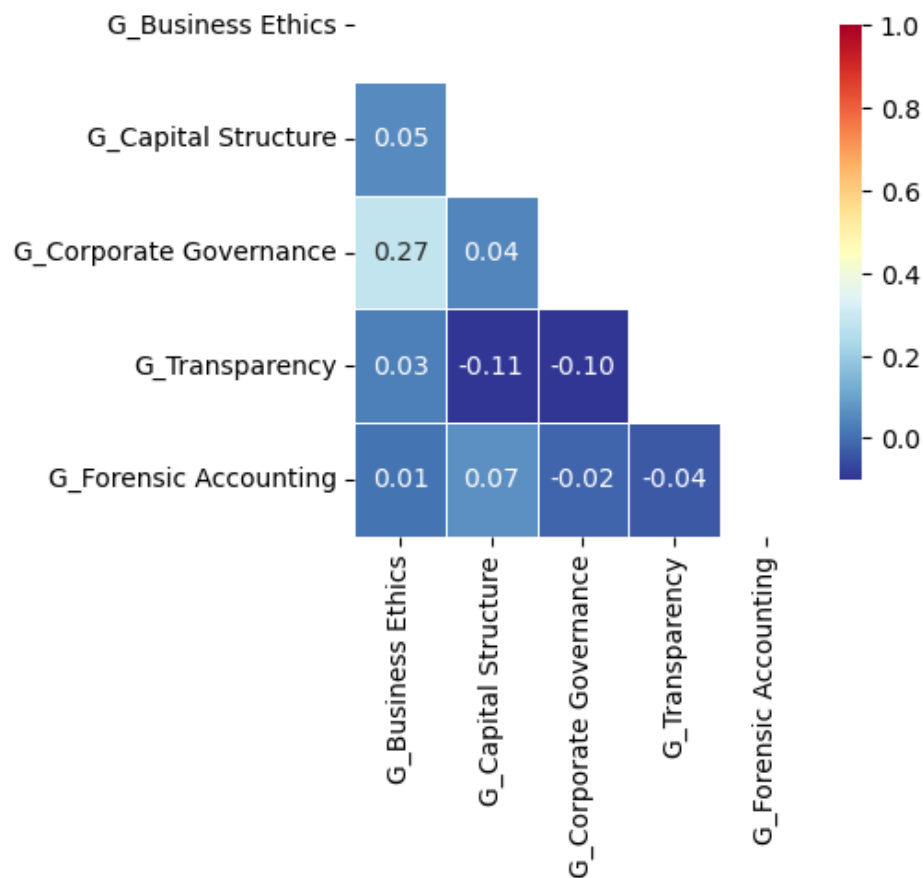


Absolute Value of r	Strength of Relationship
$r < 0.3$	None or very weak
$0.3 < r < 0.5$	Weak
$0.5 < r < 0.7$	Moderate
$r > 0.7$	Strong

# Social Score Heatmap (Correlation check included)



# Governance Score Heatmap (Correlation check included)



# Deriving Smart City ESG Scores

## Multiple Linear Regression Analysis

$$\rightarrow \text{ENV}(Y) = -0.295 + 0.123X_1(\text{Emissions}) + 0.162X_2(\text{Environmental Management}) + 0.168X_3(\text{Waste}) + 0.103X_4(\text{Environmental Stewardship}) + 0.188X_5(\text{Resource Use}) + 0.124X_6(\text{Water}) + 0.141X_7(\text{Environmental Solution})$$

$$\rightarrow \text{SOC}(Y) = 0.735 + 0.066X_1(\text{Compensation}) + 0.129X_2(\text{Diversity}) + 0.122X_3(\text{Employment Quality}) + 0.080X_4(\text{Human Rights}) + 0.067X_5(\text{Labour Rights}) + 0.121X_6(\text{Health and Safety}) + 0.066X_7(\text{Training and Development}) + 0.175X_8(\text{Product Quality and Safety}) + 0.085X_9(\text{Community Relationship}) + 0.082X_{10}(\text{Product Access})$$

$$\rightarrow \text{GOV}(Y) = 0.068 + 0.156X_1(\text{Business Ethics}) + 0.312X_2(\text{Capital Structure}) + 0.143X_3(\text{Corporate Governance}) + 0.069X_4(\text{Transparency}) + 0.320X_5(\text{Forensic Accounting})$$

# Is there a better way to analyze this by using Machine Learning?

- 기존 MLR 방식의 특징은 ‘규칙 또는 메커니즘’에 대한 이해를 바탕으로 모델을 도출
  - 현재 ESG 데이터의 특징은 산정 기준이 기관마다 다름.
  - 또한 산정 기준 계산식을 공개하지 않음
  - 따라서, 공개된 데이터에서 특징을 도출해내야 함
  - Machine Learning은 규칙 또는 메커니즘보다는 ‘성능’에 중점을 두고 있음.
  - 이에 데이터 검증을 통한 모델 생성이 가능함.



# Why Machine Learning?

- P-value 검증방식의 한계
  - 데이터의 수는 점점 커짐
  - 데이터의 수가 크면 p-value의 값이 항상 유의함으로 도출됨
  - 이에 p-value를 통한 기존 회귀 방식은 의미가 떨어짐

## Why Machine Learning? (cont'd)

- **Machine Learning Model은 Sustainable함**
  - Machine Learning은 새로운 데이터가 추가되면 재학습을 통해 모델이 자동조정 가능
  - Smart City는 미래지향적 도시이며, 미래에 중요하다고 생각하는 가치가 계속 변화될 것임. 새로운 데이터가 계속해서 추가될 것임.
  - 기존 방식으로 진행시 새로운 메커니즘 및 규칙을 확인하여 다른 모델 생성을 해야 하는데 시간과 비용이 높음
  - 데이터가 모델의 기준이 되는 Machine Learning Model이 적합

# Machine Learning Methods and Metrics

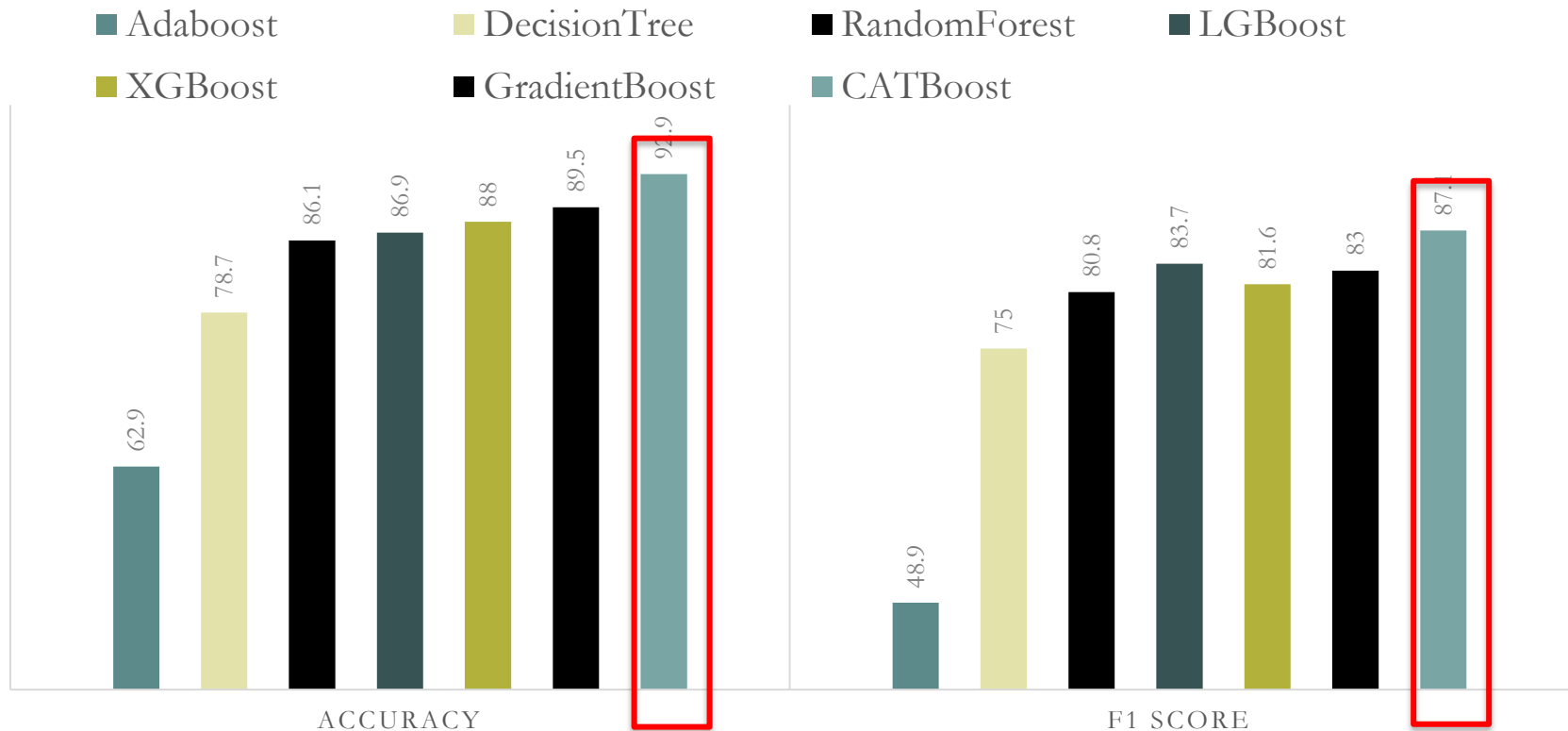
- **F1 Score**
  - Recall과 Precision의 조화평균으로 모델을 평가
  - 0과 1사이의 값을 가지며 1에 가까울수록 우수한 모델로 평가(본 평가에서는 0 – 100으로 환산하여 표현)
- **Accuracy**
  - 예측값과 실제값의 정확도로 모델을 평가

# Machine Learning Methodologies

No	Classifier Model	Classifier Type
1	Decision Tree	Tree-based
2	AdaBoost	Tree-based ensemble
3	XGBoost	Tree-based ensemble
4	LightGBM	Tree-based ensemble
5	Random Forest	Tree-based ensemble
6	Extremely Randomized trees	Tree-based ensemble
7	CATBoost	Tree-based ensemble

# ML Analyses & Data Comparison

## RANKINGS OF 5 ML MODEL RESULTS



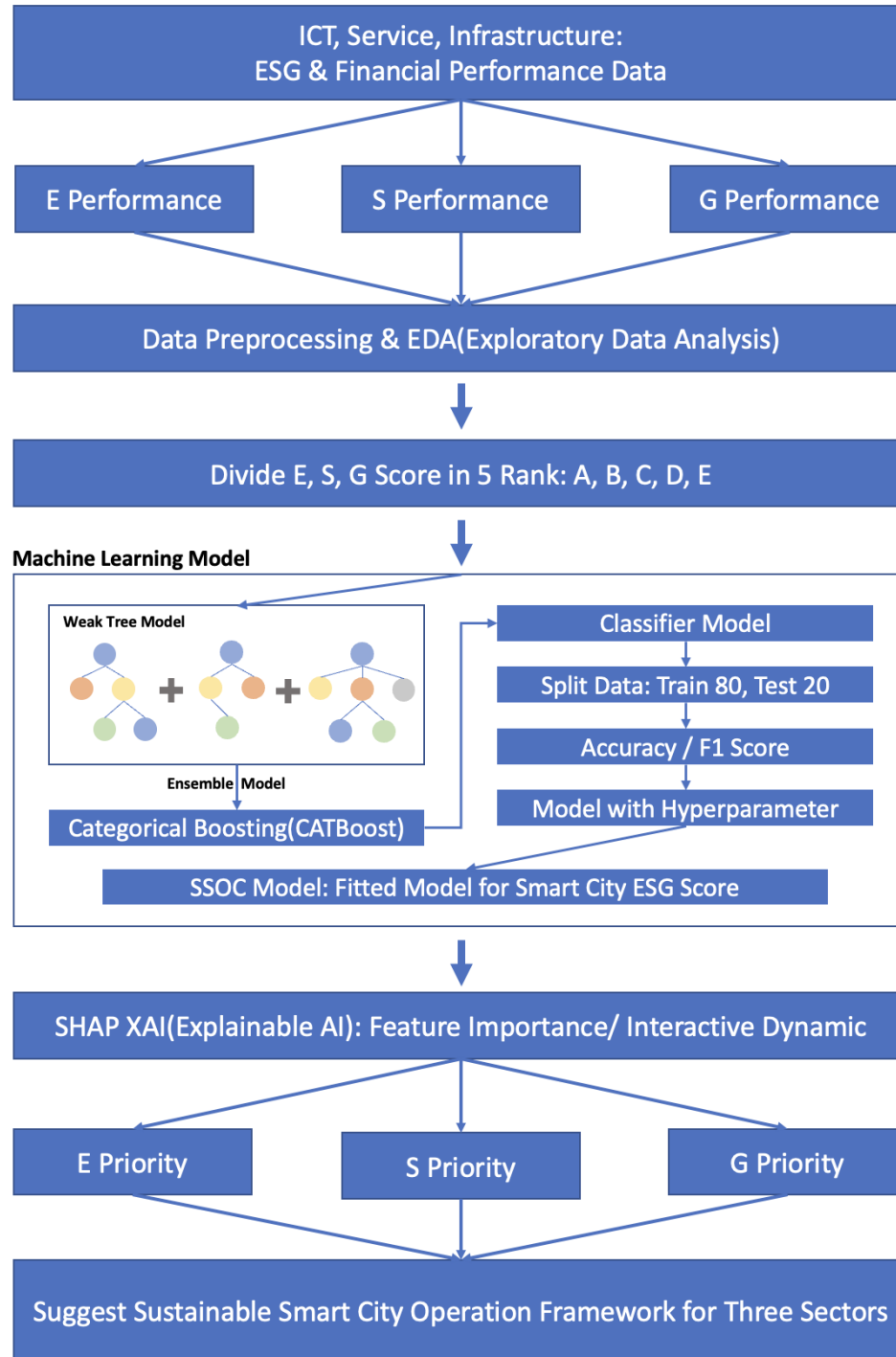
# ML Analyses & Data Comparison

No	Classifier Model	E score			S score			G score		
		Accuracy (%)	F1-score(x100)	Accuracy Rank	Accuracy (%)	F1-score(x100)	Accuracy Rank	Accuracy (%)	F1-score(x100)	Accuracy Rank
1	CATBoost	91.4	87.9	1	87.3	79.3	1	92.9	87.1	1
2	GradientBoost	85.8	79.2	5	81.6	72.1	5	89.5	83	2
3	XGboost	87.6	84.1	2	83.5	74.7	3	88	81.6	3
4	LGBost	87.6	83.7	2	82.4	78.8	4	86.9	83.7	4
5	Random Forest	87.3	80.4	4	83.9	72.6	2	86.1	80.8	5
6	Decision Tree	74.9	72.6	6	66.7	59.0	6	78.7	75	6
7	Adaboost	48.7	35.5	7	52.4	33.3	7	62.9	48.9	7

# CATBoost

- Categorical 변수로 이루어진 데이터셋에서 예측 성능이 우수한 알고리즘
- Boosting 기반의 모델로 약한 Classifier를 결합하여 성능이 더 좋은 Classifier을 생성하는 원리





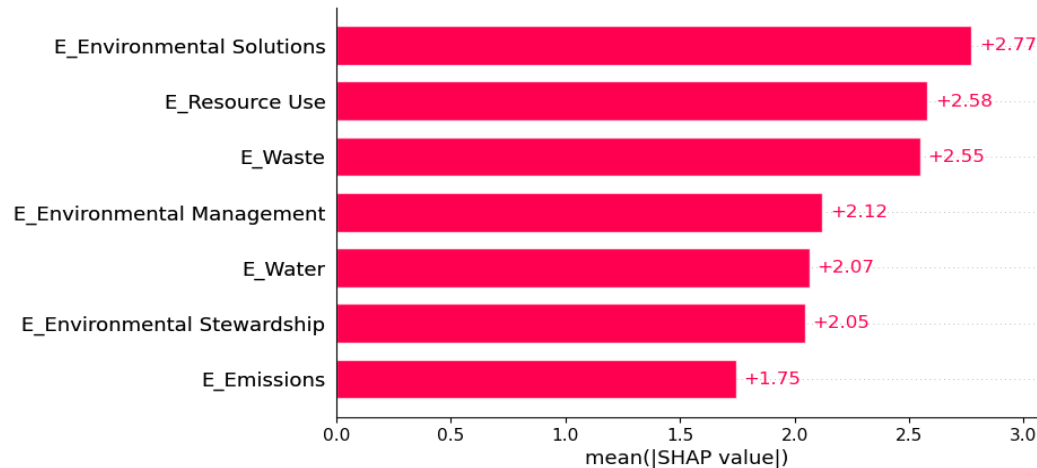
Research  
Approach  
using  
Machine  
Learning

## XAI: Feature Importance, P/N effects

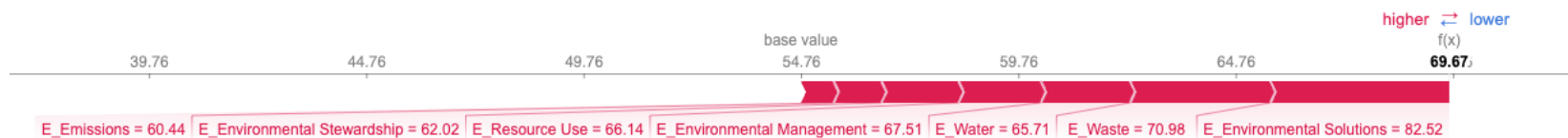
- Feature Importance Plots
  - Sub-category의 각 변수가 모델에 얼마만큼의 영향을 주는지 확인
  - 다중선형회귀에서의 회귀계수 숫자와 SHAP 방법을 활용한 Feature Importance의 중요도가 다름
- Sub-Category Positive & Negative Effects
  - Sub-category들이 모델에 미치는 영향의 크기뿐 아니라, positive한지 negative한지 확인 가능
  - 빨간색: Positive, 파란색: Negative

# XAI: Environment Feature Importance

- Feature importance plots of E-score

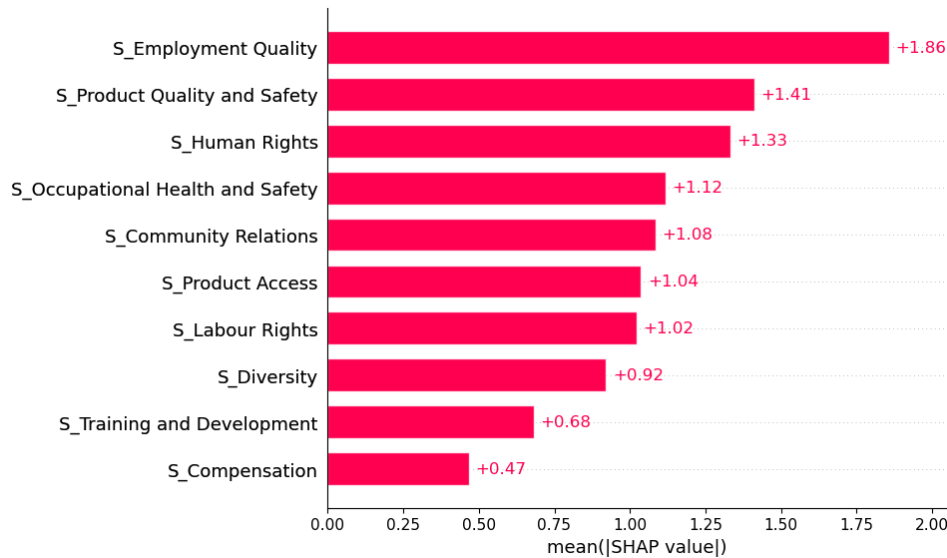


- Sub-Category Contributions

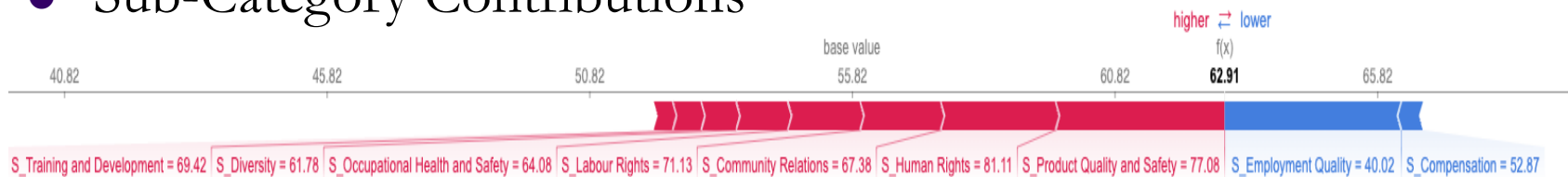


# XAI: Social Feature Importance

- Feature importance plots of S-score

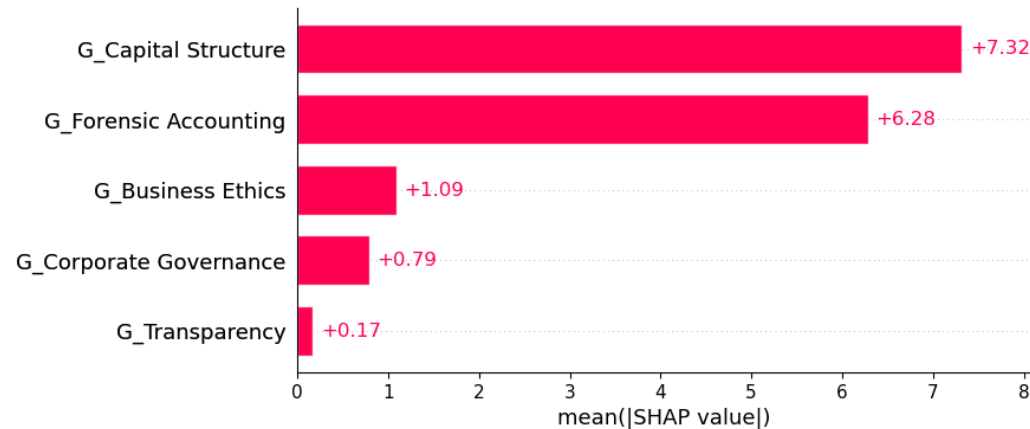


- Sub-Category Contributions



# XAI: Governance Feature Importance

- Feature importance plots of G-score



- Sub-Category Contributions



# Regression vs. Feature Importance

## Environment

Environment Score				
Regression Importance	Coefficient from MLR	ML Importance	Previous Rank	Updated Rank
Resource use	0.188	Environmental Solution	3	1
Waste	0.168	Resource use	1	2
Environmental Management	0.162	Waste	2	3
Environmental Solution	0.141	Environmental Management	3	4
Water	0.124	Water	5	5
Emission	0.123	Environment Stewardship	7	6
Environment Stewardship	0.103	Emission	6	7

# Regression vs. Feature Importance

## Social

Social Score				
Regression Importance	Coefficient from MLR	ML Importance	Previous Rank	Updated Rank
Product Quality and Safety	0.175	Employment Quality	3	1
Diversity	0.129	Product Quality and Safety	1	2
Employment Quality	0.122	Human Rights	7	3
Health and Safety	0.121	Health and Safety	4	4
Community Relationship	0.085	Community Relationship	5	5
Product Access	0.082	Product Access	6	6
Human Rights	0.080	Labour Rights	8	7
Labour Rights	0.067	Diversity	2	8
Training and Development	0.066	Training and Development	9	9
Compensation	0.066	Compensation	10	10



# Regression vs. Feature Importance Governance

Governance Score				
Regression Importance	Coefficient from MLR	Machine Learning Rank	Previous Rank	Updated Rank
Forensic Accounting	0.320	Capital Structure	2	1
Capital Structure	0.312	Forensic Accounting	1	2
Business Ethics	0.156	Business Ethics	3	3
Corporate Governance	0.143	Corporate Governance	4	4
Transparency	0.069	Transparency	5	5

## Regression Coefficient vs Feature importance

- Regression coefficient는 종속 변수에 대한 각 독립 변수의 개별적 영향을 이해하는데 관련이 있음
- Feature Importance는 모델의 정확한 예측을 수행하기 위한 가장 중요한 독립 변수를 식별하는데 관련이 있음
- Machine Learning 모델에서는 선형 회귀 모델에 비해 상대적으로 독립변수와 종속변수의 비선형적인 관계를 더 잘 표현할 수 있음
- 선형 회귀 모델에서 모델의 예측도를 찾아내는 측면에서는 회귀계수 값의 크기에 집중할 필요성은 낮음

# Regression Coefficient vs Feature importance

## References

Title	Precedent studies on Research
Gelman, A., & Hill, J. (2007). On the interpretation of regression coefficients as standardized effects. <i>Journal of Educational and Behavioral Statistics</i> , 32(3), 283-289.	다중 회귀 모델에서 독립 변수의 중요성을 비교하기 위해 회귀 계수를 사용하는 한계에 대한 연구
Candes, E. J., & Wakin, M. B. (2020). On the interpretation and usefulness of linear regression models. <i>Foundations and Trends® in Machine Learning</i> , 12(4), 311-428	머신러닝 모델의 관점에서 regression coefficient와 feature importance의 관계에 대한 연구
Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In <i>Advances in neural information processing systems</i> (pp. 4765-4774)	머신러닝 모델의 예측에 기여하는 SHAP를 포함한 다양한 정량화 기법에 대한 연구
Efron, B., & Hastie, T. (2016). Variable importance in regression models. In <i>The Annals of Applied Statistics</i> (Vol. 10, No. 1, pp. 1-50).	Regression coefficient의 크기, hypothesis testing, 예측도를 포함하여 regression model에서 독립 변수의 중요성을 정량화하는 다양한 기법에 대한 연구
Molnar, C. (2019). <i>Interpretable Machine Learning: A Guide for Making Black Box Models Explainable</i> (Doctoral dissertation, Johannes Kepler University Linz).	머신러닝 기법에서의 블랙 박스 모델의 예측을 설명하는 다양한 방법에 대한 소개 및 연구

## Lessons #2: Smart City and ESG

- We can derive smart city's ESG level by looking into participating firms
- Company's ESG scores can affect sustainable smart city development and operation.
- GOV received the lowest average score.
  - GOV has the largest potential to increase the ESG scores
  - GOV needs to be better practiced and focused
  - GOV only had statistical significance to Return on Assets (ROA)

# Questions or Comments?

# EXTRAS

# Regression Coefficient

- 선형 회귀 모델을 모든 예측을 변수가 독립적이고 종속변수에 선형 효과가 있다고 가정
- 독립변수와 종속변수간의 관계의 강도와 방향을 정량화함
- 수학적으로는 R-squared 값으로 선형 회귀 모델이 데이터에 얼마나 잘 맞는지에 대한 척도를 제공함
- 다른 모든 독립변수를 일정하게 유지하면서 독립 변수의 단위 변화(unit change)에 대한 변화를 개별적으로 측정함

## Feature importance

- 종속 변수에 대한 여러 독립 변수의 결합된 상호 작용 효과를 특징을 추출해 낼 수 있음
- Feature importance는 feature들이 제거되거나 섞일 때 모델의 성능이 얼마나 감소하거나 증가하는지 확인하여 결정됨
- 정확도, 정밀도, 재현율, ROC, AUC와 같은 기준으로 중요도를 결정함
- Feature Importance는 모델의 예측력에 대한 각 독립 변수의 기여도를 측정함



# XAI: Social Feature Importance

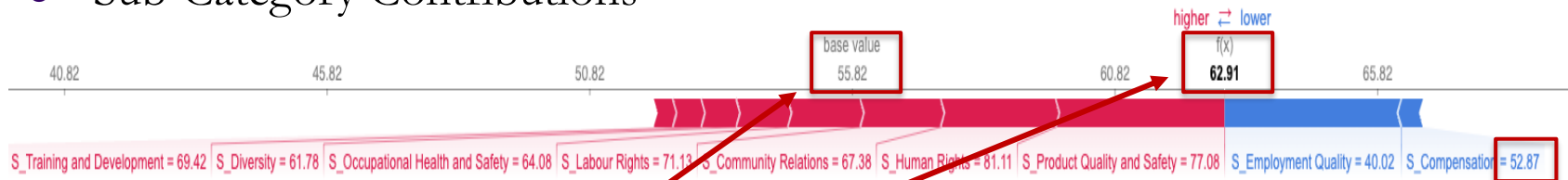
- Feature importance plots of S-score



- Feature importance using mean absolute value
- SHAP Value의 positive와 negative의 값에 대해 절대값을 활용하여 feature importance의 순위를 결정
- Positive와 Negative의 Contribution visualization 결과와 순위가 다른 이유는 shap value의 절대값을 활용한 순위결정이기 때문
- 그러나 값 자체의 크기를 활용한 분석결과이기 때문에 모델에 영향을 미치는 feature importance를 잘 확인할 수 있음

# XAI: Social Feature Importance

- Sub-Category Contributions



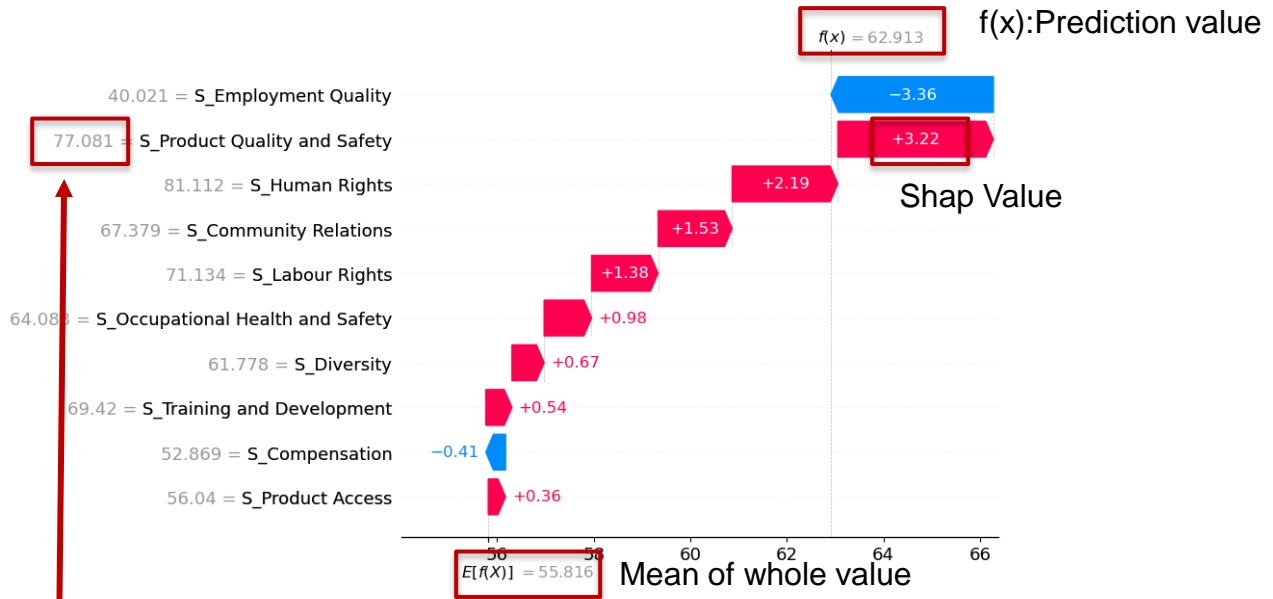
- 본 시각화 방법은 'force plot'이라고 불림

- Base value 은 7개의 sub-category들의 1,334개 전제 값의 평균값
- f(x) 은 Model의 prediction 최종 값
- 각 sub-category의 값은 데이터 중 한 개의 sample 값에 대한 actual 값 중 하나의 값을 display함
- Base Value 55.82부터 f(x)로 표현되는 최종 62.91 prediction 값까지 도출해내는데 각 sub-category의 contribution이 어떻게 이루어지는지 시각해

# XAI: Social Feature Importance

- Feature importance with contribution of S-score

One of the actual Sample value from a sub-category of S\_product Quality and Safety



- 본 시각화 방법은 'waterfall plot'이라고 불림
  - 앞의 Force plot와 동일한 원리이지만 positive와 negative를 feature importance의 순위를 같이 표현함
  - $E[f(x)]$ 의 값인 55.816는 전체 데이터의 평균 값
  - $f(x)$ 는 모델의 prediction 값으로,  $E[f(x)]$ 와 모든 shap value의 합으로 나타남(아래 식 참고)
    - ✓  $F(x) = 62.913 = E[f(x)] + 0.36 - 0.41 + 0.54 + 0.67 + 0.98 + 1.38 + 1.53 + 2.19 + 3.22 - 3.36$
  - 좌측 회색 값은 각 sub category의 값 중 한 개의 sample 값을 예시로 보여주는 값임