A STUDY ON ANALYSIS AND APPLICATION OF WATER MANAGEMENT BASED ON BIG DATA: FOCUSING ON PREVENTION OF WATER RATES DELINQUENCY IN LOCAL WATERWORKS

By

KIM, Seonju

CAPSTONE PROJECT

Submitted to

KDI School of Public Policy and Management

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF PUBLIC MANAGEMENT

2018

A STUDY ON ANALYSIS AND APPLICATION OF WATER MANAGEMENT BASED ON BIG DATA: FOCUSING ON PREVENTION OF WATER RATES DELINQUENCY IN LOCAL WATERWORKS

By

KIM, Seonju

CAPSTONE PROJECT

Submitted to

KDI School of Public Policy and Management

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF PUBLIC MANAGEMENT

2018

Professor Jongyearn LEE

A STUDY ON ANALYSIS AND APPLICATION OF WATER MANAGEMENT BASED ON BIG DATA: FOCUSING ON PREVENTION OF WATER RATES DELINQUENCY IN LOCAL WATERWORKS

By

KIM, Seonju

CAPSTONE PROJECT

Submitted to

KDI School of Public Policy and Management

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF PUBLIC MANAGEMENT

Committee in charge:

Professor Jongyearn LEE, Supervisor _____

Professor Junesoo LEE _____

Professor Taejun LEE _____

Approval as of December, 2018

ACKNOWLEDGEMENT

I have completed my first year of master's program and have completed my study. There are many people who have helped me over the past year. I am grateful to them for completing my capstone project. First and foremost, I would like to thank my company K-water and colleagues for giving me the opportunity to continue my study.

I would like to express my deepest gratitude to my supervisor, Prof. Jong Yern Lee who encouraged and supported me to grow as an independent researcher. His guidance and valuable comments were the milestone in every step of developing my study.

Above all, I especially wish to express to my dearest husband and my two sons to whom all I am enormously indebted for their endless love and support during my study.

ABSTRACT

A STUDY ON ANALYSIS AND APPLICATION OF WATER MANAGEMENT BASED ON BIG DATA: FOCUSING ON PREVENTION OF WATER RATES DELINQUENCY IN LOCAL WATERWORKS

By

Seonju Kim

Recently, there has been a growing interest in intelligent technology such as Big Data. It has become an emerging technology as a key issue in major trends in domestic and abroad. The purpose of this research is to study and propose Big Data analysis performance plan for minimizing the occurrence of delinquent customers by analyzing the past pattern of them in local waterworks operation efficiency project. In order to accomplish the purpose of this study, this research will be established analysis performance plan including infrastructure POC (that is, proof of concept) as a preliminary step for data collection and extraction, data refinement and transformation, data analysis and verification in the local waterworks. This paper will serve as an initial guide for Big Data analysis in this area. This research will be of interest to policy makers including K-water and municipal officers. This research will make use of existing databases, Korean public data, Korean Statistical Information Service data, and other working papers and journals.

Table of Contents

I. Introduction

Recently, there has been a growing interest in intelligent technology such as Big Data. It has become an emerging technology as a key issue in major trends in domestic and abroad, and the Fourth Industrial Revolution has been triggered by the development of vast amounts of data and ICT technology to analyze and link it and has also affected in various types of industries (NIA, 2016). It has also affected water industry including K-water. K-water, which is aiming to become integrated water management and global water company, needs preemptive response such as Big Data technology and analysis using it, and this would be the key to survival in the water industry. However, only about 5% of Korean companies is reported to have applied it, while approximately 30% of global companies have adopted Big Data technology (NIA, 2016). K-water has been implementing 'Integrated Water Management' according to one of the company's strategies in the meantime, but it has had limitations in its response due to lack of internal experiences for Big Data analysis. The integrated water management business through 'local waterworks operation efficiency project', which has been promoted as a growth engine in K-water management strategy, has not produced as much achievement as originally aimed (K-water, 2017).

The study on the analysis and application of water management based on Big Data is a matter of significant interest for policymakers and local governments; therefore, there is growing literature on the number of papers on new technologies in addition to the research papers on existing predictive technique theories (Lee, 2013; Lee e*t al.*, 2004; Yu *et al.*, 2013). The previous research has shown that the concept and fundamentals of data mining techniques to predict future from past patterns (Lee, 2013). Existing research has confirmed the analysis of how the

independent variables affect premium delinquency (Lee *et al.,* 2004). It has confirmed that the size of the data set at the time of data extraction and experiment has a great influence on the reliability of the research analysis (Yu *et al.,* 2013).

Generally, these seem to have focused on conventional understanding and introduction to Big Data technologies, as well as deep research into technology theory; however, I will take a different stance. I will intend to focus on utilizing and analyzing the huge amounts of data accumulated by K-water and open government data providing Korean government. Previous research has paid relatively little attention to the perspectives of practical value which is applying to real organization. Further research is required in order to improve the gap between academia and industries; therefore, this paper will attempt to conduct a study that includes a practical implementing project to find correlations between various types of variables. In addition, this paper also will propose ways to build a K-water Big Data Infrastructure and build data for predictive simulation such as preventing water rates delinquency. This research will be of interest to policy makers including K-water, professionals and government officers as well as municipal officers.

The purpose of this paper is to study and propose a plan to build a K-water Big Data Analysis Environment as a preliminary step for predicting and managing customers who are expected to switch to a long-term delinquent status on water rates. In this paper, I will also highlight why previous policies have failed to collect water rates. In particular, it will be argued that simple collection system by person in charge does not have an effect. I aim to shed light on the existing data that management has not been interested in. This paper illustrates that organizations may gain interesting insights if it extracts meaningful information and trends from

a Big Data analysis point of view. It is also expected that new business opportunities will be created in accordance with the government's policy of integrating the quantity and quality of water.

This research paper will attempt to answer the following research questions: Are government and K-water aware of water rates delinquency? Then, how can K-water contribute to improving the water rates delinquency? In order to solve the questions of this study, this paper will include the followings. it will be proposed a way to build K-water Big Data performance plan in local waterworks for ultimate data analysis. At this point, it should be possible to obtain a list of available data for analyzing past patterns of delinquent customers and to determine which data will be used as analysis variables. And it will be presented to implement K-water’s Big Data Infrastructure (H/W, S/W).

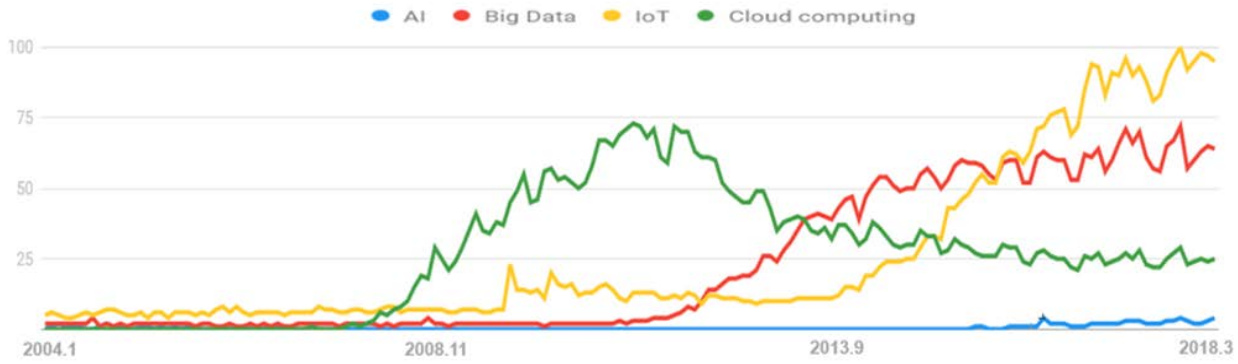
Having provided a context for this research paper, then the next step of this paper will proceed to review secondary literature on the topic of analysis and application of water rates management based on Big Data focusing on prevention of delinquency in local waterworks.

II. Literature Review

The empirical analysis of the prediction for the prevention of water rates delinquency, which is directly related to this study, is very limited. As a result of reviewing similar previous researches, most of them have focused mainly on technical aspects such as mining techniques. The major direction of the literature review on this research paper is as follows. First, it is necessary to understand the trends and concepts of intelligent technology, including big data, which is the main concept of the research topic. Second, it is necessary to comprehend predictive techniques, including data mining techniques to predict future delinquencies of water rates from past patterns of delinquent customers.

To begin with, I will take a look at the historical background of Big Data technologies. The development of technology leads the transformation of the paradigm of economic society in each age and leads social innovation. Among the key intelligent technology keywords, interest in 'Big Data' which is the core resource of intelligent information society is steadily increasing from 2012 until now can confirm (see Figure 1). Reflecting this trend, the World Economic Forum has selected Big Data as one of the top 10 emerging technologies, and Korea has been selected as one of the best IT technologies since 2012.

Figure 1. Changes in interest of major intelligent technologies over time.



Source: Google Trend (2004. 1 ~ 2018. 3)
Note: The figures represent the search interest for a specific time period, and the number 100 represents the highest popularity of the search.

In recent issues of Big Data, economists, engineers, and policy makers have presented diverse opinions (Lee, 2013; Ham and Chai, 2012; Kitcin, 2014; Gartner, 2017; NIA, 2016). On the one hand, there are skepticisms that many of the revolutionary new technologies that have been in the meantime may be on the brink of "hype-cycle," as soon as they are, and soon become fashionable (Gartner, 2017). In other words, there is a mixture of vague curiosity and expectation of new things and expansion inference based on some success stories. On the other hand, however, it is also widely anticipated that Big Data will be able to gather and analyze data on a scale that could not have been imagined until recently (NIA, 2016; Ham and Chai, 2012).

Before proceeding further, it is necessary to define clearly the key terminology referred to in this research paper. In particular, it is imperative to clarify what we mean when we talk about “Big Data”. The definition of Big Data can be expressed slightly different depending on the scholars. In order to understand previous concepts similar to Big Data, awareness of a problem, which is about the growing of the rapid increase in information volume and how they are managed and used it can be traced back to the early 1940s when the term “infor mation explosion” began to be used (Gleick, 2011). On the other hand, the recent trend toward being increased the amount of information called Big Data is quite different from the existing one in its size and speed to be called a mere increasement. Whereupon it is so different from the existing one in terms of its extent and velocity that it has become a tsunami of data, so it cannot be afforded by conventional tools (Berry, 2011).

To put it more concretely, it is clear the fact that information is produced, measured, recorded and stored in all aspects of life is itself a new and important phenomenon. In addition, most scholars have argued 3Vs (Volume, Variety, Velocity), which is commonly used to define Big Data, are abbreviating these quantitative and qualitative changes (Ham and Chai, 2012). It is clear that the fact that such data is being produced is a new and important phenomenon. It is also natural that the question about how we can capture, analyze and interpret quantitatively and qualitatively different data from existing data. The position to understand and use the concept of Big Data more broadly is based on this context. Similarly, Big Data is not simply a new data source, but rather an answer to the complex questions of what, how and why caused by its emergence, a new framework, including a systematic methodology in a broad sense, can be viewed as a new paradigm (Kitcin, 2014).

According to the Industrial Development Corporation (IDC), Big Data is defined as “next generation technology and architecture designed to extract value from economy through fast capture, data exploration and analysis from massive volume of data specified” (p.11) (NIA, 2016). That is, the process of finding business value from diverse and complex data is a Big Data analysis.

To sum up the various opinions of the academics and industry experts on the Big Data and the analysis harnessing this, it is fair to say that if there is no big business value in analyzing the data using difficult analysis methods from a corporate perspective, this paper adopts the view that it is not a Big Data analysis. It also seems reasonable to argue that these Big Data technologies are building the basis for optimized policies and contributing to specific data-based administration and policy development. At this, developed countries such as the United States

and the United Kingdom including Korea have been making various attempts to use Big Data in various fields such as traffic, safety, welfare, and so on.

Another term that needs definition in addition to Big Data is a delinquent customer. In order to predict customers who are expected to be delinquent, we need to define the term 'delinquency'. The attribute of delinquency, which indicates the state of payment of the water rates, will play a key factor in the analysis of Big Data in this paper. Whereupon, how does K-water define delinquent customer? K-water, which has been commissioned by the local government to manage local waterworks, is using the definition which refer from municipal ordinance as follows. If the previous month's water rates are not paid by the due date, it is classified as an unpaid customer. If the state of being unpaid continues for three consecutive months, it is regarded as a delinquent customer. The definition of a delinquent customer may be somehow different for each local government, and K-water should be in accordance with the standards of the contracted local government, however, this paper will adhere to this definition within the scope of this study.

Thus far, this paper has focused on the trends, terms and concepts of Big Data. The following section will address the concept and examples of data mining techniques in order to predict future delinquencies in patterns of past delinquent customers through reviews of various papers. What is the difference between Big Data from analytical technology perspective and Big Data, which it has just learned so far? Data is not just big but there are various data types and streaming data. Therefore, Big Data researchers argue that Big Data Research would be much more plausible to call Big Data Analysis rather than simply Big Data Research, and Big Data

Analytics uses this Big Data Set to interpret the semantics, which is a high-level analytical technology that gives value to people (Lee, 2013).

In order to predict the future delinquency rate of this research topic, it is an inquiry about the factors that cause it and the characteristics of delinquent customers. A similar case, Lee *et al.* (2004), confirmed the analysis of how the independent variables affect premium delinquency. Therefore, as in this paper, it is considered that independent variables can be extracted by setting the dependent variable and analyzing the correlation with the expected variables.

From another point of view, by examining the paper by Yu, E. et al (2013), it was confirmed that the size of the data set at the time of data extraction and experiment has a great influence on the reliability of future research results analysis. It was determined that sufficient data needed for future experiments of this study should be secured.

Decision tree analysis technique is an analysis method that classifies image grouping into several small groups or performs prediction by tabulating decision rules. Because the analysis process is represented by the tree structure, it is advantageous that the researcher can understand and explain the analysis process more easily than methods such as Discriminant Analysis, Regression Analysis, Neural Networks have. In this context, the following papers have been reviewed for the case studies on decision tree techniques. It is found that the decision tree analysis is used to identify the relationship between the termination characteristic of the personal mobile communication customer and the attribute variable of the subscriber, and the logistic regression model is used to score the termination possibility (Berry, *et al*., 1997). It is considered that the decision tree analysis is performed to find out the pattern of the termination

characteristics of the mobile communication subscriber through the customer segment based on the customer DB.

In order for this study, after reviewing the internal K-water environment, it was confirmed that there is no information infrastructure environment for Big Data analysis and there is no program to use for extraction through refinement. Therefore, it is considered that the pre-execution process for preparing the environment should take considerable resources and time before performing the delinquent forecast analysis. This should be recognized as a main factor in carrying out this research project and management. However, it is necessary to review the profitability improvement of 'Local waterworks operation efficiency project' internally. Analysis through forecasting modeling based on Big Data that has never been tried after commissioning is worth researching do. In addition, if this study is successfully conducted, it is expected that the local water rates charge officer will be able to provide a chance to prepare a preliminary response strategy based on the forecast data. Through the literature review part, we have looked at the background such as the definition and concept of Big Data and examined previous studies on analysis techniques. The research method will be discussed in the next part.

III. Method

1. procedure

The purpose of this research is to study and propose Big Data analysis performance plan for minimizing the occurrence of delinquent customers by analyzing the past pattern of them in local waterworks operation efficiency project. In order to accomplish the purpose of this study, this research will be established analysis performance plan including infrastructure POC as a preliminary step for data collection and extraction, data refinement and transformation, data analysis and verification in the local waterworks.

To begin with, steps will be established for Big Data analysis performance plan. In order to do this, the following three steps are performed. First, to investigate the status of internal data. Second, to investigate the status of external data, and finally, to derive a PoC system construction environment. It will ask for obtaining access rights and examine the data to be analyzed in order to internal data survey.

2. preliminary step for data collection, refinement and generation

The data gathering and extracting steps will be described in more detail as follows. The data are largely classified into internal data stored in K-water and external data to be provided from outside. The data analysis is planned by selecting the data to be used as variables and securing a list of data that can be utilized in the delinquent customer prediction analysis is expected to be clearly defined.

First of all, the internal data should be organized in DB tables and lists related to the delinquency history based on the data of 21 local governments. First, check the tables and fields

related to local water supply customer information, charge adjustment, charge storage details, and so on. Second, we check the DB table field for public data connection to be provided from the outside. Third, we organize the DB table and field list based on the requirements collection interview results. Fourth, check tables and fields related to automatic transfer change, delinquent payment, household information, etc.

In order to investigate the status of external data, public data available from the public open data portal (www.data.go.kr) or Korean Statistical Information Service (KOSIS, http://kosis.kr) is used. First, investigate the status of public data to be linked and analyzed. Second, the attribute to be associated with the internal data is defined. At this time, an attribute definition and derivation attribute for each public data are also defined.

Acquired analytical data and infrastructure through this procedure will be possible to proceed next steps. 'Data Collection and Extraction Phase', data analysis should be planned and the scope of analysis performed should be clearly defined by securing original data that can be used for forecasting analysis of delinquent customers and selecting the data to be used as a variable. In the step of 'data refinement and transformation', it is necessary to check whether data attributes include noise, outliers, missing values, etc., and remove the corresponding data or refine attribute values. Then, the collected data is collected according to the purpose of analysis, classified, or converted into attribute values to generate new analytical data suitable for the purpose of analysis. In the last three steps, 'Analysis and Verification Phase', we design predictive modeling to determine payment type from analytical data and then design payment type decision process based on predictive model.

3. Study design

3.1. Definition of promblems

By reminding the ultimate purpose of this study, it is possible to analyze the past patterns of the delinquent customers of local governments subject to the 'local waterworks operation efficiency project' to minimize the 'delinquency rate' that may arise in the future. In addition, the analysis results predict monthly delinquency for each customer. Therefore, the dependent variable will be 'delinquency'. The purpose of the analysis is to provide the opportunity to reduce the possibility of delinquency of the local government by informing the customer of the possibility of delinquent payment at the time of adjusting the bill to issue the bill every month.

3.2 Analytical data list

Since it will take a considerable time for final analytical data to be confirmed through internal / external data collection, the list of analytical data (see Table 1), which was examined based on the scenario information among the internal data that can be accessed first.

Table 1. List of analytical data

Internal Data Related to Scenario		**DB Table for Analysis (Corresponding column)**	**Variables**	**Derived Variable (Y/N)**
Customer List	caliber	RateAdjustment(CustomerNo, RateAdjustmentYearMonth, WaterworksCode, UndergroundwaterCode)	WaterworksCaliber, UndergroundwaterCaliber	N

	Business Type	RateAdjustment(CustomerNo, RateAdjustmentYearMonth, WaterworksRepresentativeBizType, WaterworksConcurrentBizType, SewageRepresentativeBizType, SewageConcurrentBizType, UndergroundwaterRepresentativeBizType, UndergroundwaterConcurrentBizType)	WaterworksRepresentativeBizType, WaterworksConcurrentBizType, SewageRepresentativeBusinessType, SewageConcurrentBizType, UndergroundwaterRepresentativeBizType, UndergroundwaterConcurrentBizType	N
	Water Supply Status	WaterworksMeterReadingInfo (MeterReadingYearandMonth, CustomerNo, WaterworksState) UndergroundwaterMeterReadingInfo (MeterReadingYearandMonth, CustomerNo, UndergroundwaterMeterState)	WaterworksState, UndergroundwaterMeterState	N
	Bill Notification Method	RateAdjustment(CustomerNo, RateAdjustmentYearMonth, BillNotificationMethod)	BillNotificationMethod	N
Unpaid Customer List	Receipt Type	RateAdjustment (CustomerNo, RateAdjustmentYearMonth, DueDate, AfterDueDate, ReceiptType) ReceiptHistory(CustomerNo, RateAdjustmentYearMonth, ReceiptDate)	LastMonthoReceiptType	N
	Receipt Way	RateAdjustment(CustomerNo, RateAdjustmentYearMonth, DueDate, AfterDueDate, ReceiptType) ReceiptHistory(CustomerNo, RateAdjustmentYearMonth, ReceiptDate)	LastMonthoReceiptWay	N
Delinquent Customer List	Business Type	RateAdjustment(CustomerNo, RateAdjustmentYearMonth, WaterworksRepresentativeBizType, Waterworks ConcurrentBizType, SewageRepresentativeBizType, SewageConcurrentBizType, UndergroundwaterRepresentativeBizType, UndergroundwaterConcurrentBizType)	Waterworks RepresentativeBizType, Waterworks ConcurrentBizType, SewageRepresentativeBizType, SewageConcurrentBizType, UndergroundwaterRepresentativeBizType, UndergroundwaterConcurrentBizType	N

	DelinquentAmount	RateAdjustment(CustomerNo, RateAdjustmentYearMonth, AfterDueDateAmount) ReceiptHistory(CustomerNo, RateAdjustmentYearMonth)	AfterDueDateAmount	N
	SuccessionYN	RateAdjustment(CustomerNo, RateAdjustmentYearMonth, DelinquencySuccessionYN)	DelinquencySuccessionYN	N
DeficitDisposalList		DeficitInfo (RateAdjustmentYearMonth, CustomerNo)	LastDeficitYN	Y

The independent variables were further extracted and secured according to the above-mentioned independent variable selection criteria and accumulated in the list of analytical data.

3.3. Selection of generation algorithm of prediction model and Data analysis

In order to analyze the secured data ultimately, 'predictive model generation algorithm' should be designed. In this case, we will choose which methodology to apply and how to determine its suitability. Firstly, it is necessary to be able to process dependent variables which are qualitative variables. Secondly, the analytical data should be able to be processed in both of these cases since the qualitative and the quantitative variables are mixed. Thirdly, the generated prediction model (i.e., the pattern extracted from the data) needs to provide readability to the user.
Accordingly, a decision tree or a logistic regression analysis among the data mining algorithms commonly satisfying the above elements is considered appropriate.

Data mining is a special step algorithm that applies elaborated algorithms to detect structure, association, and hidden patterns from a large amount of stored data. However, as a disadvantage, it relies on data to interpret and improve the phenomenon. Therefore, if a model

that extracts information from a state in which the data does not sufficiently reflect the reality is developed, an error may be caused to construct a wrong model. Therefore, as data sampling and experiment for the prediction model generation, classification prediction analysis algorithms such as decision tree are generally influenced by the distribution of dependent variables (frequency of values). We will perform experiments using accumulated data so that the data can fully reflect the reality for the data experiment for the prediction model generation.

4. Utility and Limitations of the Proposed Research

The ultimate purpose of this study is to analyze the past patterns of the delinquent customers of local governments which are subject to the 'local waterworks operation efficiency project' to minimize the 'delinquency rate' that may arise in the future. This research project started with the question of what kind of new business area can apply the big data technology, except for the water quantity and quality analysis, which was once tried in the integrated water management field. As a result of the specifics of the prospectus, it was confirmed that there is no information infrastructure environment for processing Big Data analysis, and there are no programs to be used for program extraction through refining.

Ultimately, Big Data information infrastructure is required to perform predictive analysis, and the pre-execution process for building data for Big Data analysis is considered to take considerable resources. However, based on this study, it is necessary to review the profitability improvement of 'Local waterworks operation efficiency project' internally. analysis through forecasting modeling based on Big Data that has never been tried after commissioning is worth researching do. In addition, if this study is successfully conducted, it is expected that the local water rate charge officer will be able to provide a chance to prepare a preliminary response strategy based on the forecast data. The research Analysis and findings will be presented in the

next part.

IV. Analysis and findings: Analysis Performance Plan

In this part, 'Big Data analysis performance plan' will be created as a result of this study as well as the first step to perform Big Data analysis in local water works. This paper will serve as an initial guide for Big Data analysis in this area.

1. Definitions
2. Overview
    - 2.1 Purpose
    - 2.2 Goals
    - 2.3 Definition of Task
    - 2.4 Outline of analytical data
3. Internal Data Status related delinquency
    - 3.1 Linking internal data to generate analytical data
    - 3.2 Utilization of Internal Data Related to Scenario
        - 3.2.1 Criteria for Utilization of Scenario-Related Data
        - 3.2.2 Utilization of internal data
    - 3.3 Utilization of interview-related data
    - 3.4 Additional Analysis of Internal Data
4. Status of external data to be linked with internal data
    - 4.1 Korean Statistical Information Service (KOSIS)
        - 4.1.1 Population Census
        - 4.1.2 Economically Active Population Survey

1. Definition of terminology

In this part, the meanings of the terminologies that will be covered frequently in analysis performance plan will be defined. This will reduce confusion among K-water, government officers and municipal officers who will refer to it.

1) Internal data: Retained data within the organization.

2) External data: Data held by other organizations. E.g., Open Government Data

3) Data requested of analysis: Data presented in the task scenario or an interview etc.

4) Analytical data: Original data for analyzing the delinquent customer patterns.

5) Variables: The attributes that represent the data.

6) Derived variables: Variables generated by combining variables. E.g., Delinquency: to generate delinquent status compare to the due date, deadline, and payment date

7) Qualitative variables: Variables represent data as symbols. E.g., gender: Male, Female

8) Quantitative variables: variables represent data in numerical terms. E.g., Water rate:1,200

9) Independent variables: variables that affect dependent variables. E.g., Whether encouraging delinquency in the previous month or not.

10) Dependent variable: variable that changes according to change of independent variable. E.g., customer type: within due date, overdue, delinquency, deficit

11) Database (DB): A set of information that is integrated and managed for the purpose of sharing and using by many people. E.g., Local waterworks DB

12) Database Management System (DBMS): A program that manages the database. E.g., ORACLE, MSSQL, mySQL, cubrid

13) DB table: A database table containing the same type of data. E.g., Consumer information

2. Outline of analysis performance plan

This part will provide an overview of the tasks before analyzing the selected tasks. This will include the purpose of the task, goals, definition of the analysis problem, an overview of the metadata to be analyzed, and data selection criteria.

2.1 Purpose

It is possible to plan the data analysis and determine the scope of the analysis by obtaining a list of available data for delinquency prevention analysis and selecting the data to use as a variable. In addition, based on this, it is possible to suggest the construction of the final POC system to apply the preventive analysis model of delinquent customers.

2.2 Goals

First, it is needed to examine tables and related columns of internal database to be used for analysis.

Second, it is needed to investigate external data to be linked and to prepare a link with internal data.

Third, it is needed to prepare a method of generating variables and derived variables of analytical data.

Fourth, it is needed to propose a plan about an infrastructure of POC system.

2.3 Definition of task

1) Purpose of analysis: To minimize the delinquent rate of local waterworks rates

2) Analysis result: To predict monthly delinquent rate by customer. (Dependent variable: delinquent)

3) How to use the analysis results

   It is possible to pre-analyze the prospects of the customer's delinquency before payment of the bill every month, so that it is possible to provide an opportunity to lower the proportion of water rate delinquency in advance.

4) Selection criteria for independent variables

   It is needed to use the following data which may affect ‘delinquent’ which is dependent variable.

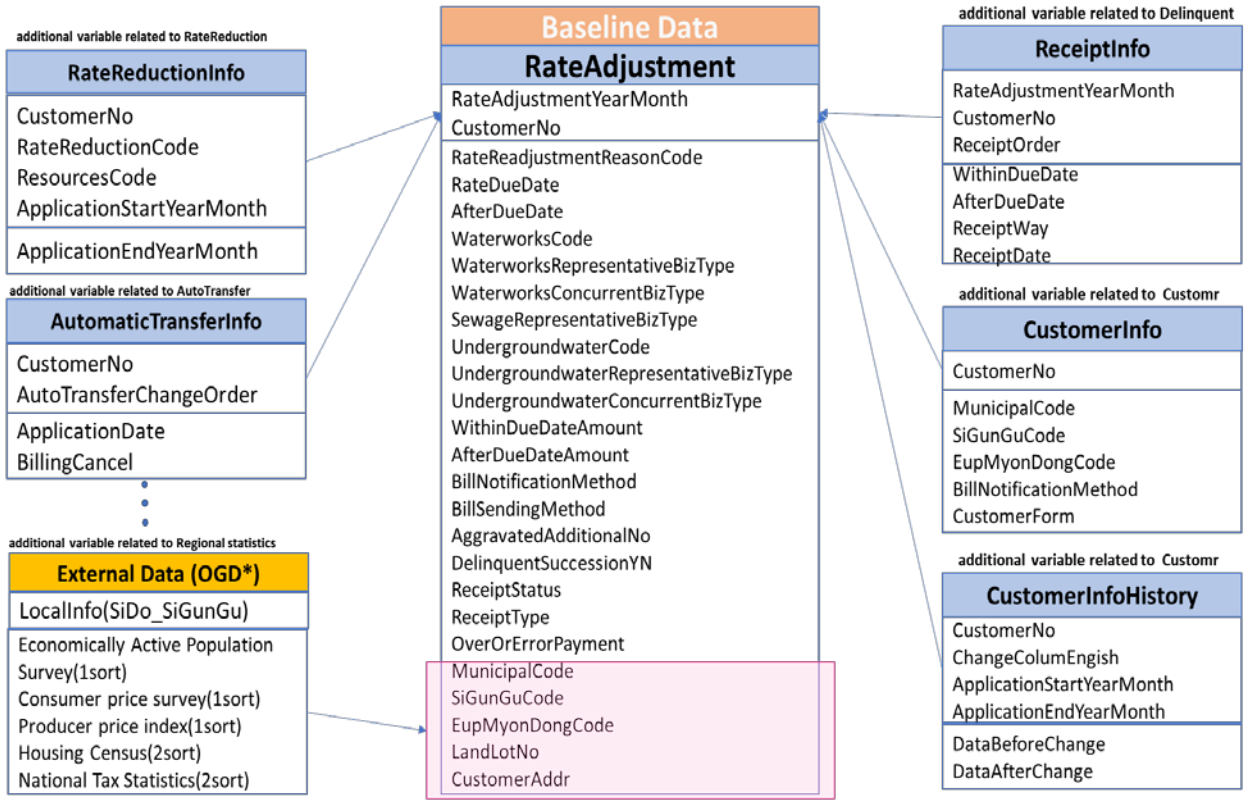
   - Data which are presented from a scenario of selected task
   - Data which are observed from an interview of Non-San Waterworks Center
   - Statistical data on the regional characteristics of the customer
   - Data associated with the above data

2.4 Outline of analytical data

First, analytical data which signify original data for analyzing the delinquent customer patterns are constructed based on ‘rate adjustment’ of DB table in the local waterworks (see Figure 2).

Next, it is needed to derive the variable of analytical data for preventing delinquency of water rate by using internal data i.e. 'RateReductionInfo', 'AutomaticTransferInfo', ‘ReceiptInfo', 'Customer Info', 'CustomerInfoHistory' and external data such as open government data (OGD*).

Figure 2. Outline of analytical data for prediction water rate delinquency



* *Open Government Data (OGD) is a philosophy- and increasingly a set of policies - that promotes transparency, accountability and value creation by making government data available to all. Public bodies produce and commission huge quantities of data and information. By making their datasets available, public institutions become more transparent and accountable to citizens. By encouraging the use, reuse and free distribution of datasets, governments promote business creation and innovative, citizen-centric services.*

sources: http://www.oecd.org/gov/digital-government/open-government-data.htm

3. Status of internal data related to delinquency

3.1 Linking internal data to generate analytical data

The criterion for linking internal data is determined by customer number and adjusted year and month. This is because the basic unit of the delinquent forecast data is the delinquency by the customer's year. In the case of a DB table without adjusted year and month, use the application, application, and registration date of the corresponding data.

3.2 Utilization of Internal Data Related to Scenario

3.2.1 Criteria for Utilization of Scenario-Related Data

1) List of customers by local government
    - Data requested of analysis: caliber, business type, waterworks condition, billing method
    - Data excluded of analysis: Customer number, customer name, business name
    - The reason why Customer number, customer name, business name is excluded from the target of analysis for the study is this is identification codes in this case, therefore, they can be used only for data linkage.
2) List of delinquent customers by local government
    - Data requested of analysis: receipt type, receipt method
    - Data excluded of analysis: Customer number, adjusted year
    - The reason why Customer number, adjusted year are excluded from the target of analysis for the study is this is identification codes in this case, therefore, they can be used only for data linkage.

3) List of complaints by local government

- Data requested of analysis: civil large, civil medium
- Data excluded of analysis: Customer number, application date
- The reason why Customer number, application date is excluded from the target of analysis for the study is this is identification codes in this case, therefore, they can be used only for data linkage.

4) List of observer information by local government

- Data requested of analysis: business type, delinquency amount, succession(Y/N)
- Data excluded of analysis: Customer number, adjusted year and month
- The reason why Customer number and adjusted year and month are excluded from the target of analysis for the study is this is identification codes in this case, therefore, they can be used only for data linkage.

5) List of deficit disposal for local government

3.2.2 Utilization of internal data

The internal data for analysis is extracted by using described method above and written in the table (see table 2) below.

Table 2. List of internal analytical data related to scenario.

Internal Data Related to Scenario		DB Table for Analysis (Corresponding column)	Variables	Derived Variable (Y/N)
Customer List	caliber	RateAdjustment(CustomerNo, RateAdjustmentYearMonth, WaterworksCode, UndergroundwaterCode)	WaterworksCaliber, UndergroundwaterCaliber	N

	Business Type	RateAdjustment(CustomerNo, RateAdjustmentYearMonth, WaterworksRepresentativeBizType, WaterworksConcurrentBizType, SewageRepresentativeBizType, SewageConcurrentBizType, UndergroundwaterRepresentativeBizType, UndergroundwaterConcurrentBizType)	WaterworksRepresentativeBizType, WaterworksConcurrentBizType, SewageRepresentativeBusinessType, SewageConcurrentBizType, UndergroundwaterRepresentativeBizType, UndergroundwaterConcurrentBizType	N
	Water Supply Status	WaterworksMeterReadingInfo (MeterReadingYearandMonth, CustomerNo, WaterworksState) UndergroundwaterMeterReadingInfo (MeterReadingYearandMonth, CustomerNo, UndergroundwaterMeterState)	WaterworksState, UndergroundwaterMeterState	N
	Bill Notification Method	RateAdjustment(CustomerNo, RateAdjustmentYearMonth, BillNotificationMethod)	BillNotificationMethod	N
Unpaid Customer List	Receipt Type	RateAdjustment (CustomerNo, RateAdjustmentYearMonth, DueDate, AfterDueDate, ReceiptType) ReceiptHistory(CustomerNo, RateAdjustmentYearMonth, ReceiptDate)	LastMonthoReceiptType	N
	Receipt Way	RateAdjustment(CustomerNo, RateAdjustmentYearMonth, DueDate, AfterDueDate, ReceiptType) ReceiptHistory(CustomerNo, RateAdjustmentYearMonth, ReceiptDate)	LastMonthoReceiptWay	N
CivilComplaintList	CivilComplaint CatL	CivilComplaintBaseInfo（CustomerNo, CivilComplaintApplicationDate, CivilComplaintCatL, CivilComplaintOccurence)	CivilComplaintCatL	N
	CivilComplaint CatM	CivilComplaintBaseInfo(CustomerNo, CivilComplaintApplicationDate, CivilComplaintCatM,	CivilComplaintCatM	N

		CivilComplaintOccurence)		
Delinquent Customer List	Business Type	RateAdjustment(CustomerNo, RateAdjustmentYearMonth, WaterworksRepresentativeBizType, Waterworks ConcurrentBizType, SewageRepresentativeBizType, SewageConcurrentBizType, UndergroundwaterRepresentativeBizType, UndergroundwaterConcurrentBizType)	Waterworks RepresentativeBizType, Waterworks ConcurrentBizType, SewageRepresentativeBizType, SewageConcurrentBizType, UndergroundwaterRepresentativeBizType, UndergroundwaterConcurrentBizType	N
	DelinquentAmount	RateAdjustment(CustomerNo, RateAdjustmentYearMonth, AfterDueDateAmount) ReceiptHistory(CustomerNo, RateAdjustmentYearMonth)	AfterDueDateAmount	N
	Succession YN	RateAdjustment(CustomerNo, RateAdjustmentYearMonth, Delinquency SuccessionYN)	DelinquencySuccessionYN	N
Deficit DisposalList		DeficitInfo (RateAdjustmentYearMonth, CustomerNo)	LastDeficitYN	Y

3.3 Utilization of interview-related data

This part records the extraction of the related DB table and data through the interview of charge worker in Non-san waterworks center (see Table 3).

Table 3. List of internal analytical data related to interview.

Major Contents of Interview	DB Table for Analysis (Corresponding column)	Variables	Derived Variable (Y/N)
If the customer does not trust the results of the meter reading, he or she will file a civil complaint. (For example, leakage accidents, abrupt water usage, carelessness by customers who could not lock the	CivilComplaintBaseInfo(CustomerNo, CivilComplaintApplicationDate, CivilComplaintCatL, CivilComplaintCatM, CivilComplaintOccurence)	CivilComplaintCatL, CivilComplaintCatM, CivilComplaintOccurence	N

faucet)			
A tenant sometimes escapes after water rate delinquency.	RateAdjustment(CustomerNo, RateAdjustmentYearMonth, FinalBillAmount) RateAdjustment(CustomerNo, RateAdjustmentYearMonth) DeficitInfo(RateAdjustmentYearMonth, CustomerNo)	LastmonthDelinquency	Y
In the case of apartment houses and malls, there may be cases where responsibility is transferred to other tenants for leaks and a large amount of water usage.	CustomerInfo(CustomerNo, RateAdjustmentYearMonth, HouseholeCat)	HouseholdCat	N
Excessive charges may occur due to the replacement of new meters.	MeterReplacementCostInfo(CustomerNo, RateAdjustmentYearMonth)	MeterReplacement	Y
When changing the notification method from automated transfer to general giro, the incidence of long-term delinquencies is often high.	CustomerInfo(CustomerNo, RateAdjustmentYearMonth, BillNotificationMethod) CustomerInfo(CustomerNo, ApplicationStartYearMonth ApplicationEndYearMonth ChangeColumEngish DataBeforeChange DataAfterChange)	BillNotificationMethodChange	Y
The amount of delinquency varies depending on the activities to prevent delinquency.	CustomerInfo(CustomerNo, RateAdjustmentYearMonth) DelinquencySpurHistoryMgt(CustomerNo, SpurDate)	SuccessiveDelinquencySpurNo	Y
If the customer does not recognize the date of the automatic debit cancellation, arrears may arise.	AutomaticTransferInfo(CustomerNo, ApplicationDate, CancelDate)	AutomaticTransferCancel	Y
If tap water is cut off, the payment is of course made quickly.	ReadMeterWaterworks(ReadMeterDate , CustomerNo, WatersupplyStatus) ReadMeterGroundwater(ReadMeterDate,	CutoffWater	Y

	CustomerNo,GroundwaterMeterStatus)		
Complaints about the automatic withdrawal method are often received. For example, the water rate was expected to be transferred first, but the water rate was overdue because it was not.	CivilComplaintBaseInfo(CustomerNo, CivilComplaintApplicationDate, CivilComplaintCatL, CivilComplaintCatM)	CivilComplaintCatL, CivilComplaintCatM, CivilComplaintOccurence	N

3.4 Additional Analysis of Internal Data

This part describes the data that is considered to be related to the task, in addition to the selected task scenario, the data set extracted from the interview (see Table 4).

Table 4. List of additional analytical data

DB Table for Analysis (Corresponding column)	Example of Column value (or Code name)	Variables	Derived Variable (Y/N)
RateAdjustment(RateReadjustmentReasonCode)	NULL, DiscountAmountChange, MovementCal	RateReadjustmentReason Code	N
RateAdjustment(BillSendingMethod)	Meterman	BillSendingMethod	N
RateAdjustment(AggravatedAdditionalNo)	5	AggravatedAdditionalNo	N
RateAdjustment(ReceiptType)	Receipt, NoReceipt, Deficit	LastmonthReceiptStatus	N
RateAdjustment(OverOrErrorPayment)	NULL, ReceiptLittle	LastmonthOverOrErrorPa yment	N
CustomerInfo(CustomerType)	GeneralHousehold	CustomerType	N
CustomerInfo(CustomerNo,MasterCustomerNo)	NULL, 451800632274	CustomerMasterSlaveCor respond	Y
CustomerInfo(CustomerNo, RateReductionTargetCode,Resource, ApplicationStartYearMonth, ApplicationEndYearMonth)	(483100807997, recipient of basic living, Waterworks, 201001, 201208)	ReductionTarget	N

4. Status of external data to be linked with internal data

4.1 Korean Statistical Information Service (KOSIS)

The scope of the data provided by Korean statistical Information Service (KOSIS*) portal is to analyze the following data, which is considered to be related to the analysis of water rates in local water works.

As a result, the data which is requested of analysis and excluded data are considered to be as follows.

- Data requested of analysis: Income level by region, utility bills, delinquency status
- Data excluded of analysis: Utility bill
- The reason why 'Utility bill' is excluded from target of analysis for the study is the data regarding utility bills provided by government services does not include statistics related to local information, therefore it is technically difficult to link with internal data.

4.1.1 Population Census

Data are available only in 2015, and 34 species can be selected from a total of 99 species (see Table 5). Selection criteria is whether to provide city or district information.

4.1.2 Economically Active Population Survey

Data are available from June 1999 to May 2017, with a total of 12 species selected from 49 species (see Table 6). Selection criteria is whether to provide trial information.

4.1.3 Housing Census

* *sources: http://kosis.kr/news/news_03List.jsp, http://kosis.kr/statisticsList/ statisticsList_01List.jsp*

This provides data with RSS which stands for rich site summary.

Only data is available in 2015, and 3 species are selected from 3 species (see Table 7). Selection criteria is whether to provide local information.

4.1.4 Consumer Price Survey

Data are available from February 1965 to May 2017, with a total of 4 species selected (see Table 8). Selection criteria is whether to provide local information.

4.1.5 Producer Price Index

Data is available from January 1965 to May 2017, no local information exists, and a total of 4 species are selected (see Table 9). Selection criteria is information on waterworks availability.

4.1.6 Water Supply Status

Data are available from 2008 to 2015, with a total of four to four (see Table 10). Optional criteria are availability of local information.

4.1.7 Sewage Statistics

Data are available from 1997 to 2015, and a total of 2 species are selected (see Table 11). Selection criteria is availability of local information.

4.1.8 Korea Urbanization Statistics

Data are available from 2009 to 2014, with a total of 9 species selected from 59 species (see Table 12). Selection criteria is whether to provide local information and waterworks.

4.1.9 National Housing Price Trend Survey

Collection period is different for each data, and 12 kinds are selected from a total of 23 kinds (see Table 13). Selection criteria is only for sale, rent, and rent information.

4.1.10 Current State of National Tax Statistics

Select 2 species from a total of 12 species (see Table 14). Selection criteria is whether to provide local information.

4.2 Korea Open Government Data Portal

The data provided by the National Statistical Office is provided in the form of an OPEN API [*], and specific data is requested and received through URL generation. https://www.data.go.kr/ (see Table 15).

Table 15. List of data extracted from Korean Open Government Data related to task

Institution	Service ID	Service Provision Type	Format	Technical Doc	Link URL
KOSIS	KOSIS statisticsList	REST	JSON +XML	Y	
	KOSIS statisticsData	REST	JSON +XML	Y	
	KOSIS statisticsMeta	REST	JSON +XML	Y	
	KOSIS statisticsExplData	REST	JSON +XML	Y	

[*] *open API (often referred to as a public API) is a publicly available application programming interface that provides developers with programmatic access to a proprietary software application or web service. APIs are sets of requirements that govern how one application can communicate and interact with another.*

sources: http://www.oecd.org/gov/digital-government/open-government-data.htm

	Korean Statistics Database as of 2013	N/A	N/A	Y	http://kosis.kr/openapi/index/index.jsp
KOSTAT	KOSTAT	N/A	JSON	N/A	https://sgis.kostat.go.kr/developer/html/home.html
Major Indicator of Korean Statistics	IndicaterService	REST	XML	Y	
	IndicatorListService	REST	XML	Y	
	PkNumberService	REST	XML	Y	
	IndExplanationService	REST	XML	Y	

The acquisition method of open government data which is listed in Table 15 will be provided as an integrated appendix 2 and there are listed primary contents of appendix 2 as follows.

4.2.1 KOSIS Statistical List Service

4.2.2 KOSIS statistics service

4.2.3 KOSIS Statistical Information Service

4.2.4 KOSIS Statistical Description Service

4.2.5 Indicator information inquiry service

4.2.6 List Indicator Search Service

4.2.7 Display description data by indicator number

4.2.8 Displaying Explanation Data by Indicator Name

5. Defining Derived Variables

In this part, it is aimed to present the derived variables which are generated by the combination of variables (see Table 16). In this research, internal and external data compiled in the previous part were used to extract next derived variables. Readers can be referred to the preceding chapter for the data set extracted from the local waterworks areas of K-water.

If a municipal officer would like to get a derived variable, it is needed to refer to the next manner. For example, when comparing or mixing data such as ‘WithinDueDate’, ‘AfterDueDate’, and ‘ReceiptDate’ from the customer waterworks rates information of DB Tables, one can found out whether the customer is 'unpaid' in the part of waterworks rate. This status of 'unpaid' or not is what is called 'the derived variable'.

Table 16. List of derived variables and creation methods from DB Table for analysis

Category	Variables	Value	DB Table for Analysis (Corresponding column)	Creation Method of derived variable
Dependent Variable	Customer Type	payby	RateAdjustment(CustomerNo, RateAdjustmentYearMonth, ReceiptType)	IF ReceiptType = 'WithinDueDate' , THEN CustomerType = ' WithinDueDate'
		unpaid	RateAdjustment(CustomerNo, RateAdjustmentYearMonth, DueDate, AfterDueDate)	IF ReceiptType = ' AfterDueDate' , THEN CustomerType = ' Unpaid '
		delinquency	RateAdjustment(CustomerNo, RateAdjustmentYearMonth) RateAdjustment(CustomerNo, RateAdjustmentYearMonth, DueDate)	IF RateAdjustment and ReceiptInfo outer join , and then there are no data of ReceiptInfo. THEN CustomerType = 'Delinquent’
		deficit	DeficitInfo(RateAdjustmentYearMonth, CustomerNo)	IF theer is DeficiInfo, THEN CustomerType = 'deficit'

Independent Variable	Successive Delinquency SpurNo	Y/N	RateAdjustment(CustomerNo, RateAdjustmentYearMonth) DelinquencySpurHistoryMgt(CustomerNo,SpurDate)	Count 'DelinquencySpurHistory' Until no occurrence of DelinquencySpurHistory from RateAdjustmentYearMonth
Independent Variable	Lastmonth Deficit	Y/N	DeficitInfo(RateAdjustmentYearMonth, CustomerNo)	IF there is LastmonthDeficit, THEN LastmonthDeficit ='y' ELSE LastmonthDeficit ='n'
Independent Variable	Meter Replacement	Y/N	MeterReplacementAmountInfo(CustomerNo, RateAdjustmentYearMonth)	IF there is MeterReplacement, THEN MeterReplacement = 'y' ELSE MeterReplacement = 'n'
Independent Variable	BillNotificationMethod Change	Y/N	RateAdjustment(CustomerNo, RateAdjustmentYearMonth) CustomerInfoHistory(CustomerNo, ApplicationStartYearMonth, ApplicationEndYearMonth, ChangeColumEngish, DataBeforeChange, DataAfterChange)	IF RateAdjustment. RateAdjustmentYearMonth >= ApplicationStartYearMonth AND RateAdjustment. RateAdjustmentYearMonth <= ApplicationEndYearMonth, THEN BillNotificationMethodChange= 'y' ELSE BillNotificationMethodChange = 'y'
Independent Variable	Lastmonth_ Delinquency Spur	Y/N	DelinquencySpurHistoryMgt(CustomerNo, SpurDate)	IF there is LastmonthDelinquencySpurHistoryMgt, THEN LastmonthDelinquencySpur = 'y' ELSE LastmonthDelinquencySpur = n'
Independent Variable	Automatic_ Transfer_ Cancel	Y/N	RateAdjustment(CustomerNo, RateAdjustmentYearMonth) AutomaticTransferInfo(CustomerNo, ApplicationDate, BillingCancelDate)	IF AutomaticTransferInfo. BillingCancelDate<= RateAdjustment. RateAdjustmentYearMonth, THEN AutomaticTransferCancel = 'y' ELSE AutomaticTransferCancel = 'y'
Independent Variable	CutoffWater	Y/N	ReadMeterWaterworksInfo(ReadMeterDate, CustomerNo, WatersupplyStatus) ReadMeterGroundwater Info(ReadMeterDate,	IF ReadMeterWaterworksInfo. WatersupplyStatus = 'CutoffMeter' OR ReadMeterGroundwater. GroundwaterMeterStatus = "CutoffMeter' ' THEN Cutoff = 'y' ELSE Cutoff = 'n'

			CustomerNo, GroundwaterMeterStatus)	

6. Designing linkage method between internal and external data

6.1 Scheme of Data Linkage

In this part, it is aimed to link the local information of external data based on the address information of the internal data such as local government code, city/town code (name), Eup-Myon-Dong code (name) in order to utilize internal and external data for analysis.

6.2 Configuring Analytical Data

In this part, it attempts to select and present the analytical data to be analyzed. 6.2.1 shows the internal analytical data (see Table 17) and 6.2.2 summarizes the external analytical data (see Table 18, 19).

6.2.1 Internal data

Table 17. List of internal analytical data

Category	Variable Type	Variable	Value (E.g., Number of kind of Qualitative variable value)
Dependent	Qualitative	CustomerType	Payby, unpaid, delinquency, deficit (4 sorts)
Independent	Qualitative	WaterworksCaliber	NULL, 13 mm, 50 mm, Unification[68sorts]
Independent	Qualitative	UndergroundwaterCaliber	NULL, 13 mm, 50 mm, 200mm [25sorts]
Independent	Qualitative	WaterworksRepresentativeBusinessType	NULL, general, biz, biz(dist) [65sorts]
Independent	Qualitative	WaterworksConcurrentBusinessType	NULL, general, biz, biz(dist) [65sorts]
Independent	Qualitative	SewageRepresentativeBusinessType	NULL, household, temporary, publicbath [62sorts]
Independent	Qualitative	SewageConcurrentBusinessType	NULL, household, temporary, publicbath [62sorts]

Independent	Qualitative	UndergroundwaterRepresentativeBusinessType	NULL, sales, temporary, lossfault [61sorts]
Independent	Qualitative	UndergroundwaterConcurrentBusinesType	NULL, sales, temporary, lossfault [61sorts]
Independent	Qualitative	WaterworksState	StopHault(etc), StopCutoff(DelinquencyCutoff), StopHault(EmptyHouseCutoff), StopHault(CustomerRequest), StopCutoff(etc), TerminateMeter[7sorts]
Independent	Qualitative	UndergroundwaterMeterState	9999, OpenMeter, StopCutoff(etc), StopCutoff(DelinquencyCutoff), StopHault(EmptyHouseCutoff), StopHault(etc), StopHault(CustomerRequest), TerminateMeter[8sorts]
Independent	Qualitative	BillNotificationMethod	giro, AutomaticTransfer [2sorts]
Independent	Qualitative	LastMonthoReceiptType	NULL, payby, AfterDueDate, Urge[4sorts]
Independent	Qualitative	LastMonthoReceiptType	Telebanking, giro, virtualAccount, AutomaticTransfer, InternetGiro, HandInput, creditcard, simpleEpayment[8sorts]
Independent	Qualitative	CivilComplaintCatL	waterworksFacility, Meter, Cutomer, Waterworks, MeterReading, WaterRate, BillNotification, leak, waterQ, VillageWaterworks, CivilComplaint(ect)[11sorts]
Independent	Qualitative	CivilComplaintCatM	householdSplitApplication, IndividualReadMeterApplication, ReadMeter(etc)[70sorts]
Independent	Qualitative	CivilComplaintOccurence	General, Complaint, Compliment [3sorts]
Independent	Qualitative	Delinquency SuccessionYN	Y, N [2 sorts]
Independent	Qualitative	LastDeficitYN	Y, N [2 sorts]
Independent	Qualitative	LastDelinquencyYN	Y, N [2 sorts]
Independent	Qualitative	HouseholdCat	NULL, Singlehouse, Apt, apthouse, multiple-dwelling, multigenerationalhousing, dormitory, lowincomeapthouse, etc, dockyard[10 sorts]
Independent	Qualitative	MeterReplacement	Y, N [2 sorts]
Independent	Qualitative	BillNotificationMethodChange	Y, N [2 sorts]
Independent	Qualitative	AutomaticTransferCancel	Y, N [2 sorts]
Independent	Qualitative	CutoffWater	Y, N [2 sorts]

Independent	Qualitative	RateReadjustmentReason	NULL, DiscountAmountChange, CalAmountReflect, , leakReduction, MeterReadingInfoChange, BizChange, householdSplit, taxexemption, additionalChargeExemption, calibChange, etc, BillNotificationMethodChange, MovementCal [13sorts]
Independent	Qualitative	BillSendingMethod	Meterman, mail, e-mail, separateway, nosending [5sorts]
Independent	Qualitative	LastmonthReceiptStatus	Receipt, NoReceipt, Deficit [3sorts]
Independent	Qualitative	LastmonthOverOrErrorPayment	NULL, DoubleReceipt, OverReceipt, UnderReceipt [4 sorts]
Independent	Qualitative	CustomerType	NULL, household(S), household(G), manlessfire/policesubstation [56sorts]
Independent	Qualitative	CustomerMasterSlaveCorrespond	Y, N [2 sorts]
Independent	Qualitative	RateReductionTarget	recipientbasicliving, SeniorcitizencenterrRateReduction, IndustrialReduction [59sorts]
Independent	Quantitative	AfterDueDateAmount	10020
Independent	Quantitative	SuccessiveDelinquencySpurNo	5
Independent	Quantitative	AggravatedAdditionalNo	3

6.2.2 External data

This part aims to be the final destination of external data generated monthly or annually for 53 months from January 2013 to May 2017 (see Table 18). Information that does not exist such as irregular data or local information to be linked to internal data is excluded in final external analytical data. When domestic statistical data are used, waterworks, sewer statistical data, etc. are published after aggregation. E.g., 2016 statistical data will be available in December 2017. When using domestic statistical data, Korean city statistical data will be provided with a delay of about 2 years, therefore, reference should be made. E.g., 2015 statistical data will be available in December 2017).

Table 18. Determining whether external data is used

Statistics	Range of collecting	Provision cycle	Provision institution	Local unit	Whether use
Population Census	2015	irregularity	National Statistical Office	Si/Gun/Gu	X
Economically Active Population Survey	1999.06~2017.05	monthly	National Statistical Office	Si/Do	O
Housing Census	2015	irregularity	National Statistical Office	Si/Do	X
Consumer Price Survey	1965.02~2017.05	monthly	National Statistical Office	Si/Do	O
Producer Price Index	1965.02~2017.05	monthly	National Statistical Office	-	O
Water Supply Status	2008~2015	yearly	Ministry of Environment	Si/Do	X
Sewage Statistics	1997~2015	yearly	Ministry of Environment	Si/Do	X
Korea Urbanization Statistics	2009~2014	yearly	Ministry of Interior and safety	Si/Gun/Gu	X
National Housing Price Trend Survey	2013.01~2017.06	monthly	Korea Appraisal Board	Si/Gun/Gu	O
Current State of National Tax Statistics4.2 Public Data Portal	2006 ~ 2015	yearly	National Tax Service	Si/Do	O

Table 19. List of external analytical data

Statistics name	data	No of Variables	Independent variable
Economically Active Population Survey	Economically Active Population data by administrative districts (Si/Do)	4	▪ Ratio of not economically active population (%) = (not economically active population/ population of over 15 years) * 100 % ▪ Activity rate (%) = (Economically Active Population/ population of over 15 years) * 100 % ▪ Unemployment rate (%) = (unemployment / Economically Active Population) * 100% ▪ employment-population ratio (%) = (employee/ population of over 15 years) *100% ** Among population of over 15 years, soldiers, combat police and public service personnel are excluded.

Consumer Price Survey	Consumer Price Index data by item	6	▪ Price index of agricultural, livestock and fisheries products ▪ Price index of industrial products ▪ Electric hydro gas price index ▪ Housing tax price index ▪ Public service price index ▪ Individual service price index
Producer Price Index	Producer Price Index data by item	3	▪ Household water price index ▪ Sales Water price Index ▪ Business water price index
National Housing Price Trend Survey	Housing sale price by item	3	▪ Household water price index ▪ Sales water Price Index ▪ Business water price index ▪ Single House Sales Price Index ▪ Apartment sales price index ▪ Allied multi-generational trading price index
	Lease price data by item	3	▪ Singlefamily housing rent price index ▪ Apartment rent price index ▪ Alliance multigenerational rent price index
Current State of National Tax Statistics4.2 Public Data Portal	Arrange performance of delinquency amount I (Local/item)	5	▪ Total number of delinquency ▪ Total amount of delinquency ▪ The number of delinquent taxpayers during arranging ▪ The number of delinquency during arranging ▪ The amount of delinquency during arranging
	Release of results of large amount /habitual delinquent taxpayer	2	▪ The number of new large amount /habitual delinquent taxpayer ▪ The amount of new large amount /habitual delinquent taxpayer
Total		27	

7. Proof of Concept

In this part, the aim of this research paper is to suggest a minimum level of information infrastructure building to perform Big Data analysis to prevent the delinquent of local

waterworks customers. The POC system building plan will present the server configuration specification (see Table 20), the detailed software configurations for each server (see Table 21), and lastly the data collection plan.

7.1 Detailed System configuration

Table 20. System configuration (minimum specification)

Server	units	Standards and specifications.		
		CPU	Memory	SSD
Main server (name node)	1EA	i7 3.0Ghz	8GB	128 GB
Data server (data node) #1	1EA	i7 3.0Ghz	8GB	128 GB
Data server (data node) #2	1EA	i7 3.0Ghz	8GB	128 GB

7.2 Detailed Software Configuration

Table 21. Detailed Software Configuration by Servers

Server	Descript	Standards and specifications.	Units
Main server (name node)	Operating System	CentOS 6.5	1
	Execution environment	Oracle JDK 8	1
	Big data analysis storage and processing (File / data storage)	Hadoop 2.6	1
	Analysis Solution	Analyzer v1.1	1
	Algorithm library	Library v1.1	1
	Algorithm library	Weka 3.8	1
	Web Application Server (WAS)	Tomcat 7.0	1
	Database	Cubrid 9.3.6	1
Data server (data node) #1	Operating System	CentOS 6.5	1
	Execution environment	Oracle JDK 8	1
	Big data analysis storage and processing (File / data storage)	Hadoop 2.6	1
Data server (data	Operating System	CentOS 6.5	1

node) #2	Execution environment	Oracle JDK 8	1
	Big data analysis storage and processing (File / data storage)	Hadoop 2.6	1

7.3 Data collection plan

7.3.1 Collection of Internal Data

The objective of this part is to present the way in order to gather data to ultimately analyze data in the minimum level of Big Data Information Infrastructure environment as described in 7.1 and 7.2 above. In the case of local waterworks, it is necessary to develop a module (or program) for transferring and storing data stored in Oracle, which is an existing RDBMS, to Cubrid, a DBMS for Big Data analysis, in order to analyze internal data. The program should be also designed to be executed by setting job scheduler.

7.3.2 Collection of External Data

In this part, it is aimed to suggest ways to collect external data by K-water. At this point, it should be considered collecting the files in a dump rather than collecting the data manually.

- Import external data dump file downloaded from external data portal site (KOSIS) into K-water.
- or an external data dump module using Open API of OGD portal to store the dump file of the external data in the analysis infrastructure.
- Load the external data dump file into the analysis DB (cubicle) using the data file loading module.

V. Conclusion and suggestion

This paper started with the question of how to utilize the accumulated data related to the 'Local waterworks operation efficiency project' which was started as a part of securing new growth engine in K-water in early 2000. This paper emphasizes the necessity of preliminary preparation of actual Big Data construction project by finding correlation between various variables.

In this paper, it is proposed that 'K-water Big Data Analysis Implementation Plan' which is needed to predict and manage customers who are expected to switch to long-term delinquent state in advance. In order to successfully analyze under the Big Data base, it is necessary to establish a systematic pre-performance plan for the target analysis task. Therefore, it is adopted a guide form that can be followed sequentially. The reason for this is that it was necessary to be able to present a series of processes in a concrete way until obtaining the related data list and extracting and finalizing the data to be used as the final Big Data analysis target variable.

After selecting the analytical task, detailed steps were included. At this time, definitions of terms, summaries of analytical tasks, selection of dependent variables of internal data and extraction of independent variables, selection of external data using open government data, extraction of independent variables, and interlinkage method between internal and external data are presented. Finally, it was proposed a method to build a POC of Big Data Information Infrastructure needed to analyze actual final data.

This paper confirms that interesting insights can be obtained when extracting meaningful information and trends from the perspective of Big Data analysis. Therefore, this study will be of interest to the K-water local waterworks manager, government and local officials as well as related policy makers. Also, it can be used for employees who are interested in Big Data-based analytical work, even if it is not the analysis subject presented in this article.

However, due to the unusual background of the local waterworks project in the process of this study, it was recognized as a weak point in operation.

To begin with, on the supplier side, unlike K-water's own business, the local waterworks operation efficiency project has an unusual background that it has been operated by individual municipalities operating independently of each other and by inconsistent forms of data from each municipality have. Of course, in the process of merging and acquiring the data, the data is refined and the integrated information system and DBMS are constructed. However, according to the regulations of each local government, the operation information of unique method is still managed and operated according to the needs of local governments. This is not a big deal in terms of general operations, but it emphasizes that these issues can be a stumbling block in recent Big Data utilization and analysis. Therefore, as a follow-up measure of this study, it is suggested that data quality management policy for data consistency management should be accompanied. It is expected that data analysis work will increase in various ways in the future. Further efforts should be made to prevent the project cost from increasing due to the time and manpower required for data extraction and refinement.

In addition, the use of external data provided by the government for research has been found to be marginal. At the time of the study plan, it was decided that the correlation between local waterworks tax delinquencies could be analyzed according to the distribution of income by region. However, if there is no regional income distribution data in the national statistical portal, it cannot be revealed. This study suggests that it is urgent to construct detailed statistical data such as distribution of income by region in Korea. Recently, public income data and housing

information have been disclosed in open government data portal, but they are provided in a limited form due to the sensitivity of information and inaccuracy of information. For subsequent studies, I expect that relevant data will be provided for minimal post-treatment research purposes.

As the Big Data based policy is expected to increase, the 'Big Data Analysis Plan' presented in this paper will be a useful reference book at the early stage of the analysis project. In addition, I expect that it will be a guide that can be continuously utilized by continuing supplementation through subsequent studies.

References

Berry, D. (2011), “The computational turn: thinking about the digital humanities.” Culture Machine 12. http://www.culturemachine.net/index.php/cm/ article/view/ 440/470 (5 April 2018).

Berry, M. J. A. and Linoff, G. S. (1997), Data Mining Techniques, New York: John Wiley & Sons, Inc.

Gartner IT Glossary, Big data [http://www.gartner.com/it-glossary/big-data (2018. 4. 7 visit)].

Google Trend, [https://trends.google.com/trends/explore?date=2004-01-01%202018-04-01&q=%EC%9D%B8%EA%B3%B5%EC%A7%80%EB%8A%A5,%2Fm%2F0bs2j8q,%2Fm%2F02vnd10,Cloud%20computing (2018. 4. 14 visit)]

Gleick, James. (2004). The Information: A History, A Theory, A Flood. New York: Pantheon Books.

Ham, Y., & Chai, S. (2012). 빅데이터, 경영을 바꾸다. [Big Data, Changing Management] Samsung Economics Research Institute. 27-54.

Kitcin, Rob. (2014). Big Data, New Epistemologies and Paradigm Shifts. Big Data & Society 1(1):1-12.

Korean Statistical Information Service (KOSIS), 2017, “KOSIS Shared Services Technical Document V1.0”

Korean Statistical Information Service (KOSIS), 2017, “Usage Guide _1 Metadata Information Query Service v2.3”

Korean Statistical Information Service (KOSIS), 2017, “Usage Guide_2 Explanation by Indicator Name Document Search Service v2.3”

Korean Statistical Information Service (KOSIS), 2017, “Utilization Guide_3 Explanation by Indicator No. Explanation Resource Search Service v2.3”

Korean Statistical Information Service (KOSIS), 2017, “Utilization Guide_4 List Indicator Search Service v2.3”

KOSIS Shared Services, Open API Code List [http://kosis.kr/openapi/introduce/introduce_01List.jsp (2018. 5. 7 visit)].

Lee, Ae Kyoung et al., (2004). Categorized the Contribution evasion through Health Insurance contribution evasion expected model. (2004). Korean Journal Of Health Policy And Administration, 14(2), 78-98. http://dx.doi.org/10.4332/kjhpa.2004.14.2.078

Lee, J. (2013). Understanding Big Data and Utilizing its Analysis into Library and Information Services. *Journal Of The Korean BIBLIA Society For Library And Information Science*, *24*(4), 53-73. http://dx.doi.org/10.14699/kbiblia.2013.24.4.053

Laney, Douglas. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. Application Delivery Strategies 949 (6 Feb 2001).

Min, J. H. and Jeong, C. W. (2009), A GA-based Classification Model for Predicting Consumer Choice, Journal of the Korean Operations Research and Management Science Society, 34(3), 29-41

National Information Society Agency(NIA), 2016, “Paradigm Shift and Response Strategies of Intelligent Age”, IT & Future Strategy, Vol.11

Yu, E., Kim, Y., Kim, N., & Jeong, S. (2013). Predicting the Direction of the Stock Index by Using a Domain-Specific Sentiment Dictionary. *Journal Of Intelligence And Information Systems*, *19*(1), 95-110. http://dx.doi.org/10.13088/jiis.2013.19.1.095

APPENDICES

APPENDIX 1

In this section, data names can be presented in Korean. This is because the vast amounts of data provided by the Korean Statistical Information Service portal are all in Korean, and the translation of these data is not consistent with the direction of the research paper. These are included from 4.1.1 to 4.1.10 and also Table 5 to Table 14.

4.1.1 Population Census

Data are available only in 2015, and 34 species can be selected from a total of 99 species (see Table 5). Selection criteria is whether to provide city or district information.

Table 5. List of population census

Institution Name	The latest recorded time/ period	Data
Korean Statistical Information Service	2015/ occasionality	□ 가구주의 성별/연령별/거주기간별 가구(일반가구) -시군구 □ 가구주의 성별/세대구성별/교육정도별 가구(일반가구) - 시군구 □ 거주기간별 가구(일반가구) - 시군구 □ 거주층별 가구(일반가구) - 시군구 □ 거처의 종류별/가구주의 종사상지위별 가구(15 세이상, 일반가구)-시군구 □ 거처의 종류별/사용방수별/점유형태별 1 인가구(일반가구)-시군구 □ 거처의 종류별/주거시설 형태별 1 인가구(일반가구)-시군구 □ 거처의 종류별/주거시설 형태별 가구(일반가구)-시군구 □ 거처의 종류별/평균사용방수별 가구(일반가구)-시군구 □ 경제활동상태별/추가자녀계획수별 기혼여성인구(15 세이상)-시군구 □ 경제활동상태별/출생자녀수별 기혼여성인구(15 세이상)-시군구 □ 산업별/출생자녀수별 기혼여성인구(15 세이상) - 시군구 □ 성별/경제활동상태별/연령별 고령자(60 세이상)-시군구 □ 성별/연령별/경제활동상태별/사회활동별 인구(20 세이상)-시군구 □ 성별/연령별/교육정도별 1 인가구(일반가구)-시군구 □ 성별/연령별/혼인상태별 1 인가구(15 세이상, 일반가구)-시군구 □ 성별/혼인상태별/종사상지위별 취업인구(15 세이상)-시군구 □ 세대구성별/거처의 종류별/점유형태별 가구(일반가구)-시군구 □ 세대구성별/연령별/거처의 종류별 고령자 거주가구(60 세 이상, 일반가구)-시군구 □ 연령별/경제활동상태별/초혼연령별 기혼여성인구(15 세이상)-시군구 □ 점유형태별/연령별 고령자 거주가구(60 세 이상, 일반가구)-시군구 □ 종사상지위별/출생자녀수별 기혼여성인구(15 세이상) - 시군구 □ 직업별/출생자녀수별 기혼여성인구(15 세이상)-시군구 □ 타지 주택 소유별/거처의 점유형태별 가구(일반가구)-시군구 □ 현 거주지별/성별/통근통학지별 통근통학 인구(강원)(12 세이상)-시군구 □ 현 거주지별/성별/통근통학지별 통근통학 인구(광주, 전남)(12 세이상)-시군구 □ 현 거주지별/성별/통근통학지별 통근통학 인구(대구, 경북)(12 세이상)-시군구 □ 현거주지별/성별/통근통학지별 통근통학 인구(대전, 세종, 충북, 충남)(12 세이상)-시군구 □ 현거주지별/성별/통근통학지별 통근통학 인구(부산, 울산,

		경남)(12 세이상)-시군구 □ 현거주지별/성별/통근통학지별 통근통학 인구(서울, 인천, 경기)(12 세이상)-시군구 □ 현거주지별/성별/통근통학지별 통근통학 인구(전북)(12 세이상)-시군구 □ 현거주지별/성별/통근통학지별 통근통학 인구(제주)(12 세이상)-시군구 □ 현거주지별/근무지별/산업별 취업인구 (15 세이상)-시군구 □ 현거주지별/근무지별/직업별 취업인구 (15 세이상)-시군구

4.1.2 Economically Active Population Survey

Data are available from June 1999 to May 2017, with a total of 12 species selected from 49 species (see Table 6). Selection criteria is whether to provide trial information.

Table 6. List of Economically Active Population Survey

Institution Name	**The latest recorded time/ period**	**Data**
Korean Statistical Information Service	201705	□ 행정구역(시도)/교육정도별 취업자 □ 행정구역(시도)/산업별 취업자 □ 행정구역(시도)/성별 경제활동인구 □ 행정구역(시도)/성별 실업률 □ 행정구역(시도)/성별 실업자 □ 행정구역(시도)/성별 취업자 □ 행정구역(시도)/연령별 취업자 □ 행정구역(시도)/종사상지위별 취업자 □ 행정구역(시도)/직업별 취업자 □ 행정구역(시도)/취업시간별 취업자 □ 행정구역(시도)/활동 상태별 비경제활동인구 □ 행정구역(시도)별 경제활동인구

- Economically active population by administrative district (province)

□ **15 세이상 인구 (천명)** = 경제활동인구 + 비경제활동인구
(15 세이상 인구 중 군인, 전투경찰, 공익근무요원, 형이 확정된 교도수 수감자 등은 제외됨)
□ **경제활동인구 (천명)** = 취업자 + 실업자
□ **취업자 (천명)**
□ **실업자 (천명)**
□ **비경제활동인구 (천명)**
□ **경제활동참가율(%)** = (경제활동인구/15 세이상 인구)×100%
□ **실업률(%)** = (실업자/경제활동인구)×100%
□ **고용률 (%)** = (취업자/15 세이상 인구)×100%

- Defining qualitative variables value

Variables	Number of values	Variable values
교육정도	5	{초졸이하, 중졸, 고졸, 전문대졸, 대학교졸이상}
산업	6	{농업_임업_어업, 광공업, 건설업, 도소매□숙박음식점업, 사업 · 개인 · 공공서비스및기타, 전기 · 운수 · 통신 · 금융}
경제활동인구	6	{15 세이상인구, 경제활동인구, 비경제활동인구, 경제활동참가율, 실업률, 고용률}
성	2	{남,여}
연령	11	{15~19, 20~24, 25~29, 30~34, 35~39, 40~44, 45~49, 50~54, 55~59, 60~64, 65 세이상}
종사상지위	5	{자영업자, 무급가족종사자, 상용근로자, 임시근로자, 일용근로자}
직업	9	{관리자, 전문가 및 관련 종사자, 사무 종사자, 서비스 종사자, 판매 종사자, 농림어업 숙련 종사자, 기능원 및 관련 기능종사자, 장치_기계조작 및 조립종사자, 단순노무 종사자}
활동상태	6	{육아, 가사, 통학, 연로, 심신장애, 그 외}

4.1.3 Housing Census

Only data is available in 2015, and 3 species are selected from 3 species (see Table 7).

Selection criteria is whether to provide local information.

Table 7. List of housing Census

Institution Name	The latest recorded time/ period	Data
Korean Statistical Information Service	201705 occasioanality	□ 거주가구수별/총방수별 단독주택 - 시군구 □ 주택의 종류별/총방수별 주택 – 시군구 □ 주택종류별/빈집 사유별/기간별/파손정도별 빈집 - 시도

- Defining qualitative variables value

Variables	Number of values	Variable values
주택종류	5	{단독주택, 아파트, 연립주택, 다세대주택, 비주거용 건물 내 주택}
총방수	6	{1 개, 2 개, 3 개, 4 개, 5 개, 6 개, 7 개이상}
빈집사유	7	{매매임대이사, 미분양미입주, 현재 수리중, 일시적 이용, 폐가, 영업용, 기타}
비어있는 기간	4	{3 개월 미만, 3~6 개월 미만, 6~12 개월 미만, 12 개월 이상}
파손정도	3	{파손없음, 일부파손, 반 이상 파손}

- Defining variables to use

□ **단독주택 수**
□ **아파트 수**
□ **연립주택 수**
□ **다세대주택 수**
□ **비주거용 건물 내 주택 수**
□ **단독주택 빈집 비율** = (단독주택 빈집 수/단독주택 수)⍰100%
□ **아파트 빈집 비율** = (아파트 빈집 수/아파트 수)⍰100%
□ **연립주택 빈집 비율** = (연립주택 빈집 수/연립주택 수)⍰100%
□ **다세대주택 빈집 비율** = (빈집 수/다세대주택 수)⍰100%
□ **비주거용 건물 내 주택 빈집 비율** = (비주거용 건물 내 주택 빈집 수/비주거용 건물 내 주택 수)⍰100%

4.1.4 Consumer Price Survey

Data are available from February 1965 to May 2017, with a total of 4 species selected (see Table 8). Selection criteria is whether to provide local information.

Table 8. List of Consumer price survey

Institution Name	The latest recorded time/ period	Data
Korean Statistical Information Service	201705 occasioanality	□ 품목별 소비자물가지수 (시도 정보 포함)

- consumer price index by items

□ 농축수산물 물가지수
□ 공업제품 물가지수
□ 전기수도가스 물가지수
□ 집세 물가지수
□ 공공서비스 물가지수
□ 개인서비스 물가지수

- Defining qualitative variables value

Variables	Number of values	Variable values
품목	5	{농축수산물, 공업제품, 전기수도가스, 집세, 공공서비스, 개인서비스}

4.1.5 Producer Price Index

Data is available from January 1965 to May 2017, no local information exists, and a total of 4 species are selected (see Table 9). Selection criteria is information on waterworks availability.

Table 9. List of Producer price index

Institution Name	The latest recorded time/ period	Data

The bank of Korea	2017.05/ 월	□ 생산자물가지수(품목별)(2010=100) - 가정용수도, 영업용수도, 업무용수도

- Defining variables to use

□ 가정용수도 물가지수 □ 영업용수도 물가지수 □ 업무용수도 물가지수

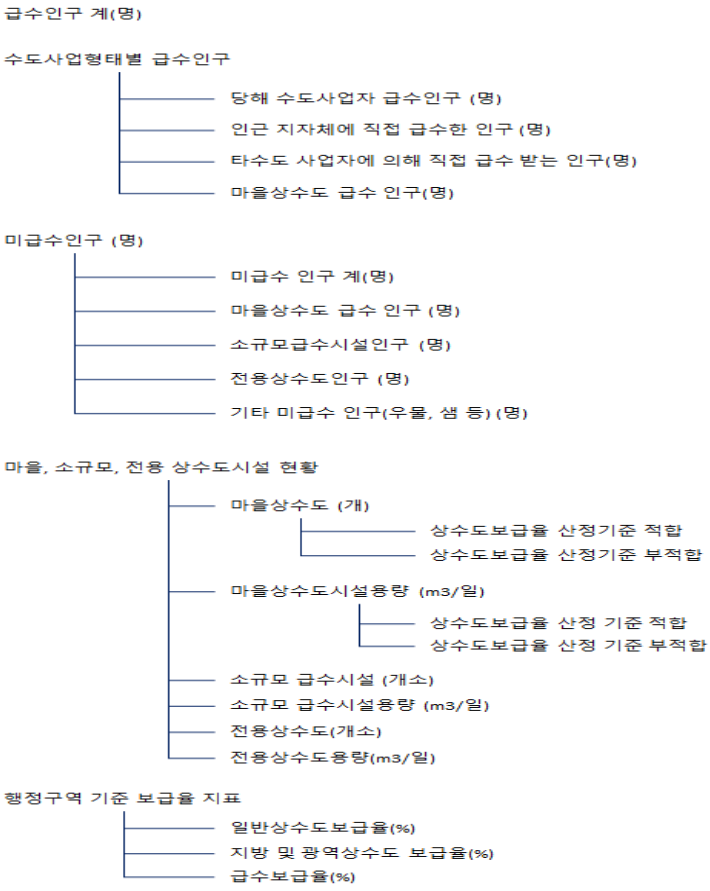
4.1.6 Water Supply Status

Data are available from 2008 to 2015, with a total of four to four (see Table 10). Optional criteria is availability of local information.

Table 10. List of water supply status

Institution Name	**The latest recorded time/ period**	**Data**	
Ministry of Environment	201705	□ 상수도 보급현황 – 시군구	□ 수도시설 현황_정수시설 – 시도
		□ 수도시설 현황_취수시설 – 시도	□ 수도요금 현황 - 시도

- Water Supply Status



- Current state of water rate

□ **총급수량(천 m3/년)**
□ **연간부과량(천 m3)**
□ **부과액(백만원)**
□ **유수율(%) =** (연간부과량 / 연간생산량)×100%
□ **요금(원/m3)** = (부과액(백만원) / 연간부과량(천 m3))×1,000
□ **생산원가(원/톤)**
□ **현실화율(%)**

4.1.7 Sewage Statistics

Data are available from 1997 to 2015, and a total of 2 species are selected (see Table 11). Selection criteria is availability of local information.

Table 11. List of sewage Statistics

Institution Name	The latest recorded time/ period	Data
Ministry of Environment	201705	□ 하수도 요금 - 시도

- Sewage rates

□ 연간부과량(천톤)
□ 부과액(백마원)
□ 평균단가(원/톤)
□ 총괄원가(백만원)
□ 총괄단가원가(원/톤)
□ 현실화율(%)

- The value of the processing cost item in 1997 was extracted from the production cost item in Sewage Statistics 1998, and the value of the processing cost item in 2002 was extracted from the total cost item per year in Sewage Statistics 2003
- The value of the Gwangju city classification from 1997 to 2000 is extracted from the Gwangju group of the 1998 ~ 2001 sewer statistics.
- The value of Pocheon classification from 1999 to 2002 is extracted from the Pocheon classification of 2000 ~ 2003 sewer statistics
- The values of the Yangzhou classification of 1999 ~ 2003 are extracted from the Yangzhou classification of the 2000 ~ 2004 sewer statistics.
- The value of the Hwaseong City classification in 2000 is extracted from the Mars classification of the 2001 Sewage Statistics.
- Values of Damyang group from 2002 to 2003 are extracted from classification of Damyang city of 2003 ~ 2004 sewer statistics

4.1.8 Korea Urbanization Statistics

Data are available from 2009 to 2014, with a total of 9 species selected from 59 species (see Table 12). Selection criteria is whether to provide local information and waterworks.

Table 12. List of Korea urbanization statistics

Institution Name	The latest recorded time/ period	Data
행정 자치부	2015 / 년	□ 1 일 1 인당 급수량 및 상수도 보급률 – 시군구 (최신수록시점: 2014) □ 1 일 1 인당 폐수 발생량 – 시군구 (최신수록시점: 2014) □ 규모별 사업체수 – 시군구 (최신수록시점: 2014) □ 기초생활보장 대상자 및 장애인 등록률 - 시군구 (최신수록시점: 2014) □ 사회복지시설수 - 시군구 (최신수록시점: 2014) □ 성 및 연령별 인구와 인구밀도 – 시군구 (최신수록시점: 2014) □ 인구이동 – 시군구 (최신수록시점: 2014) □ 주택가격 – 시도 (최신수록시점: 2014) □ 주택보급률 – 시도 (최신수록시점: 2014)

- water supply per person per day and water supply rate

□ 주민등록인구수(등록외국인포함) (명)
□ 급수인구수 (명)
□ 보급률(%)
□ 시설용량 (m3/일)
□ 급수량 (m3/일)
□ 1 일 1 인당 급수량(리터)

- Incidence of wastewater per person per day

□ 주민등록인구수(등록외국인포함) (명)
□ 1 일 1 인당 총계수발생량 (m3)
□ 1 일 총폐수발생량 (m3)

- Number of businesses by size

□ 합계
□ 1~299 명
□ 300~499 명
□ 500~999 명
□ 1,000 명 이상

- Basic Livelihood Guarantee and Disabled Registration Rate

□ 주민등록인구수(등록외국인제외) (명)
□ 국민기초생활보장 수급자수 (명)
□ 등록 장애인수 (명)
□ 남자 등록 장애인수 (명)
□ 여자 등록 장애인수 (명)
□ 국민기초생활보장 수급자비율 (%)
□ 등록 장애인율 (%)

- Number of social welfare facilities

□ 주민등록인구수(등록외국인포함) (명)
□ 시설수 (개소)
□ 생활인원수 (명)

- Defining variables to use

Variables	Number of values	Variable values
복지시설	6	{아동복지시설, 노인복지시설, 장애인복지시설, 여성복비시설, 정신질환자요양시설, 부랑인시설}

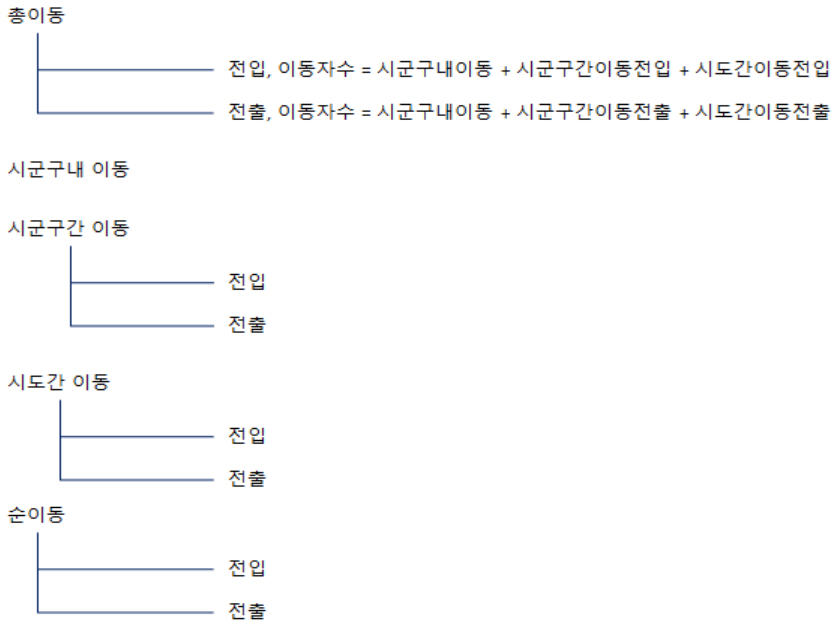
- Population and population density by gender and age

□ 면적(km2)
□ 전 년말 주민등록 인구수(명)
□ 성별 (명) = {남자, 여자}
□ 연령별 (명) = {0~14 세(유소년), 15~64 세(생산가능), 65 세이상(고령)}
□ 외국인등록인구수 (명) = {남자, 여자}
□ 인구밀도 (명/km2)
□ 인구증가율 (%)

- Population mobility

□ 이동자수 (명)
□ 이동률 (%)

- **Category of mobility of population**



- House price

□ **주택매매가격지수**
□ **주택전세가격지수**

- Diffusion Ratio of House

□ 신주택수 (천호)
□ 신주택가구수 (천가구)
□ 신주택보급률 (%)
□ 구주택수 (천호)
□ 구주택가구수 (천가구)
□ 구주택보급률 (%)

4.1.9 National Housing Price Trend Survey

Collection period is different for each data, and 12 kinds are selected from a total of 23 kinds (see Table 13). Selection criteria is only for sale, rent, and rent information.

Table 13. List of national housing price trend survey

Institution Name	The latest recorded time/ period	Data
한국 감정원	2015 / 월	□ 규모별 매매가격지수 – 시도 / 2012.01 ~ 2017.06 □ 규모별 월세가격지수 – 시도 / 2015.06 ~ 2017.06 □ 규모별 전세가격지수 – 시도 / 2012.01 ~ 2017.06 □ 연령별 매매가격지수 – 시도 / 2012.01 ~ 2017.06 □ 연령별 월세가격지수 – 시도 / 2015.06 ~ 2017.06 □ 연령별 전세가격지수 – 시도 / 2012.01 ~ 2017.06 □ 유형별 매매가격지수 – 시군구 / 2003.11 ~ 2017.06 □ 유형별 월세가격지수 – 시군구 / 2015.06 ~ 2017.06 □ 유형별 전세가격지수 – 시군구 / 2003.11 ~ 2017.06 □ 주요지역별 매매거래동향 – 시도 / 2013.01 ~ 2017.06 □ 주요지역별 월세거래동향 – 시도 / 2015.07 ~ 2017.06 □ 주요지역별 전제거래동향 – 시도 / 2013.01 ~ 2017.06

- Defining variables to use

Variables	Number of values	Variable values
주택유형	3	{아파트, 연립다세대, 단독주택}
주택규모	5	{규모 1, 규모 2, 규모 3, 규모 4, 규모 5}(1)
건축연령	5	{연령 1, 연령 2, 연령 3, 연령 4, 연령 5}(2)

4.1.10 Current State of National Tax Statistics

Select 2 species from a total of 12 species (see Table 14). Selection criteria is whether to provide local information.

Table 14. List of Current state national tax statistics

Institution Name	The latest recorded time/ period	Data
국세청	2015 / 월	□ 체납액 정리 실적 I (지역□세목) – 시도 □ 고액/상습체납자 명단공개 실적 – 시도

- Performance of the settlement of delinquency I (area, tax)

□ 전년도 이월 건수
□ 전년도 이월액 (억원)
□ 당해연도 발생 건수
□ 당해연도 발생액 (억원)
□ 체납발생 총 건수
□ 체납발생 총액 (억원)
□ 정리실적 건수
□ 정리실적 발생액 (억원)
□ 정리중체납(미정리)자 수(명)
□ 정리중체납(미정리) 건수
□ 정리중체납(미정리)액 (억원)

- making public the names of large amount of habitual delinquent taxpayer

□ 신규 고액상습체납자 수
□ 신규 고액상습체납세액(억원)

APPENDIX 2

The full detailed acquisition method of open government data which is listed in Table 15 will be given as below.

4.2.1 KOSIS Statistical List Service

4.2.2 KOSIS statistics service

4.2.3 KOSIS Statistical Information Service

4.2.4 KOSIS Statistical Description Service

4.2.5 Indicator information inquiry service

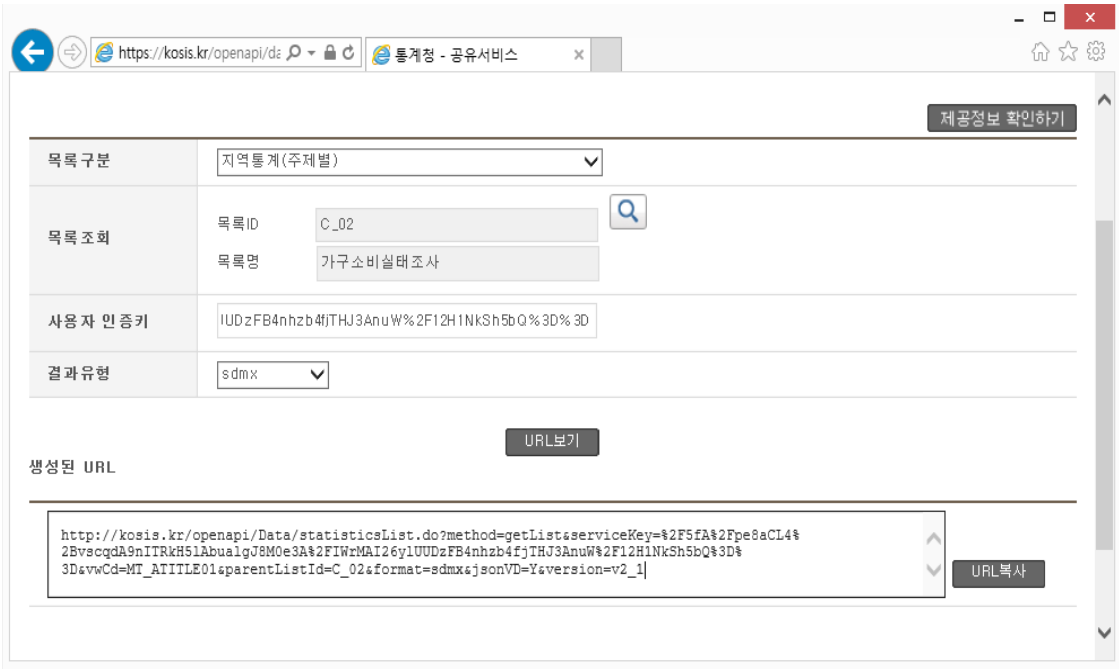
4.2.6 List Indicator Search Service

4.2.7 Display description data by indicator number

4.2.8 Displaying Explanation Data by Indicator Name

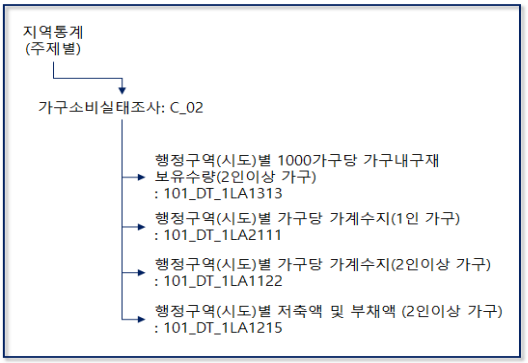
4.2.1 KOSIS Statistical List Service

o Create Data Request URL



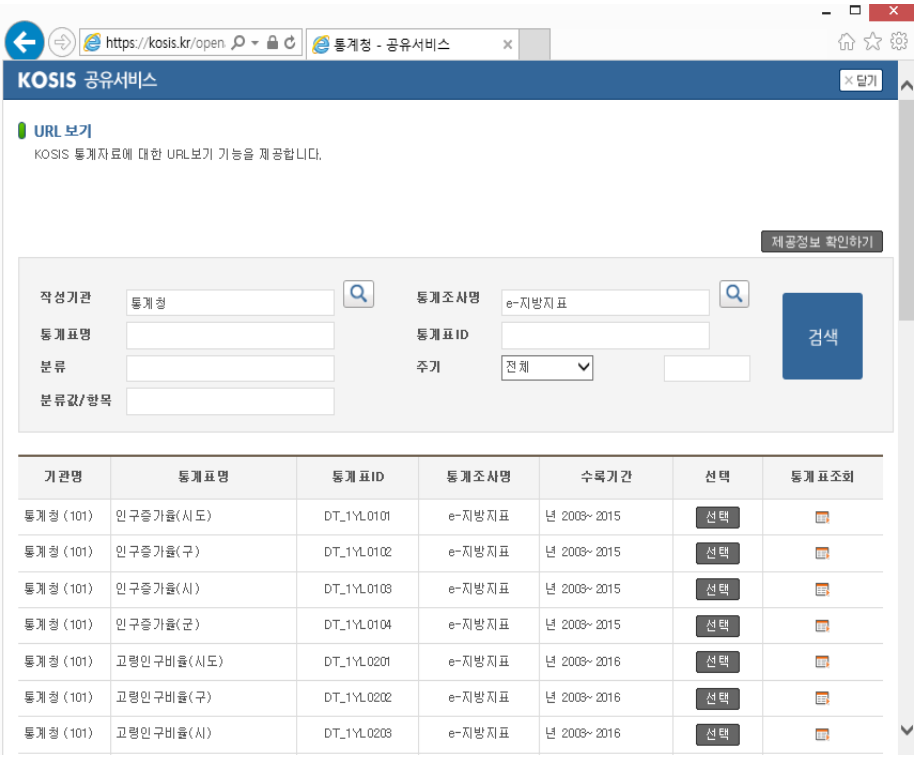
o Statistics list data generated by the request URL

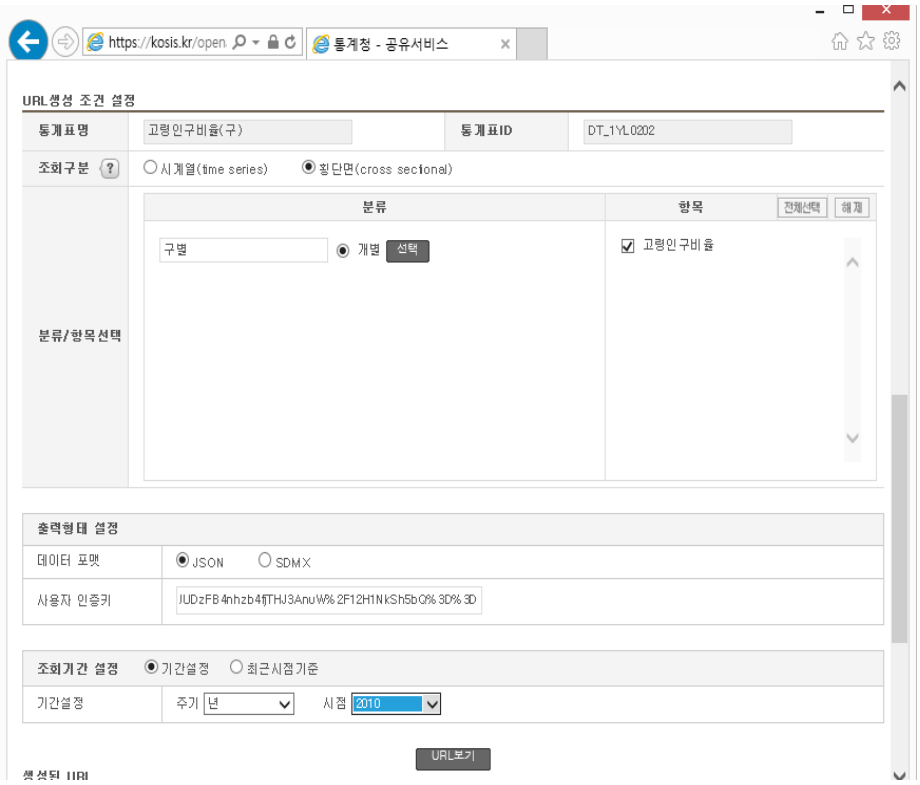

```
<?xml version="1.0" encoding="UTF-8"?>
- <message:Structure xsi:schemaLocation="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/message ../schemas/SDMXmessage.xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:common="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common" xmlns:structure="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/structure"
xmlns:message="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/message">
  - <message:Header>
      <message:resultCode>00</message:resultCode>
      <message:resultMsg>NORMAL SERVICE.</message:resultMsg>
      <message:ID>CATEGORY_MT_ATTLE01</message:ID>
      <message:Test>false</message:Test>
      <message:Prepared>2017-06-27T13:25:05</message:Prepared>
    - <message:Sender id="KOSIS">
        <common:Name>통계청</common:Name>
      </message:Sender>
      <common:Name>지역통계(주제별)</common:Name>
    </message:Header>
  - <message:Structures>
    - <structure:OrganisationSchemes>
      - <structure:AgencyScheme id="AGENCIES" version="1.0" agencyID="SDMX">
          <common:Name xml:lang="en">SDMX Agency Scheme</common:Name>
        - <structure:Agency id="SDMX">
            <common:Name xml:lang="en">SDMX</common:Name>
          </structure:Agency>
        - <structure:Agency id="KOSIS">
            <common:Name xml:lang="en">KOSIS</common:Name>
          </structure:Agency>
        </structure:AgencyScheme>
      </structure:OrganisationSchemes>
    - <structure:CategorySchemes>
      - <structure:CategoryScheme id="C_02" agencyID="KOSIS">
          <common:Name>가구소비실태조사</common:Name>
        - <structure:Category id="101_DT_1LA1313">
            <common:Name>행정구역(시도)별 1000가구당 가구내구재 보유수량(2인이상 가구)</common:Name>
            <common:Description xml:lang="en">101_DT_1LA1313</common:Description>
          </structure:Category>
        - <structure:Category id="101_DT_1LA2111">
            <common:Name>행정구역(시도)별 가구당 가계수지(1인 가구)</common:Name>
            <common:Description xml:lang="en">101_DT_1LA2111</common:Description>
          </structure:Category>
        - <structure:Category id="101_DT_1LA1122">
            <common:Name>행정구역(시도)별 가구당 가계수지(2인이상 가구)</common:Name>
            <common:Description xml:lang="en">101_DT_1LA1122</common:Description>
          </structure:Category>
        - <structure:Category id="101_DT_1LA1215">
            <common:Name>행정구역(시도)별 저축액 및 부채액(2인이상 가구)</common:Name>
            <common:Description xml:lang="en">101_DT_1LA1215</common:Description>
          </structure:Category>
        </structure:CategoryScheme>
      </structure:CategorySchemes>
    </message:Structures>
</message:Structure>
```

4.2.2 KOSIS statistics service

o Create Data Request URL





o Statistics list data generated by the request URL


```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <message:StructureSpecificTimeSeriesData xsi:schemaLocation="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/message ../../../../schemas/SDMXMessage.xsd
urn:sdmx:org.sdmx.infomodel.datastructure.DataStructure=101_DT_1YL0202:ObsLevelDim:TIME_PERIOD ecb_exr_ng_ts.xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xmlns:common="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common"
xmlns:data="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/data/structurespecific" xmlns:message="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/message"
xmlns:exr="urn:sdmx:org.sdmx.infomodel.datastructure.DataStructure=101_DT_1YL0202:ObsLevelDim:TIME_PERIOD">
  - <message:Header>
      <message:resultCode>00</message:resultCode>
      <message:resultMsg>NORMAL SERVICE.</message:resultMsg>
      <message:ID>101_DT_1YL0202</message:ID>
      <message:Test>false</message:Test>
      <message:Prepared>2017-06-27T14:13:51</message:Prepared>
    - <message:Sender id="KOSIS">
        - <message:Contact>
            <message:Department>통계청 주석참조</message:Department>
            <message:Telephone/>
          </message:Contact>
      </message:Sender>
    - <message:Structure dimensionAtObservation="TIME_PERIOD" structureID="101_DT_1YL0202">
        - <message:StructureUsage>
            <Ref id="101_DT_1YL0202" agencyID="KOSIS"/>
          </message:StructureUsage>
      </message:Structure>
      <common:Name>고령인구비율(구) (년 2003~2016)</common:Name>
      <message:Source>KOSIS(주석참조)</message:Source>
  </message:Header>
  - <message:DataSet data:dataScope="DataStructure" data:structureRef="101_DT_1YL0202">
    - <Series UNIT="14STD00018" C_A="11010" ITEM="T10" FREQ="A">
        <Obs TIME_PERIOD="2010" OBS_VALUE="13.08"/>
      </Series>
    - <Series UNIT="14STD00018" C_A="11020" ITEM="T10" FREQ="A">
        <Obs TIME_PERIOD="2010" OBS_VALUE="13.04"/>
      </Series>
    - <Series UNIT="14STD00018" C_A="11030" ITEM="T10" FREQ="A">
        <Obs TIME_PERIOD="2010" OBS_VALUE="12.57"/>
      </Series>
  </message:DataSet>
</message:StructureSpecificTimeSeriesData>
```


2010년 고령인구비율(구)
- 서울 종로구(11010) 13.08
- 서울 중구(11020) 13.04
- 서울 용산구(11030) 12.57

4.2.3 KOSIS Statistical Information Service

o Create Data Request URL

- https://kosis.kr/openapi/dataLink/dataLink_050201Form.jsp#

o Statistics table description data generated by the request URL

Prameter: orgId(기관코드) = 101 , tblId(통계표 ID) = DT_1IN0001

```
<?xml version="1.0" encoding="UTF-8"?>
- <response>
  - <Header>
      <resultCode>00</resultCode>
      <resultMsg>NORMAL SERVICE.</resultMsg>
      <id>_</id>
      <Prepared/>
    - <Sender>
        <id>KOSIS</id>
        <Name>통계청</Name>
      </Sender>
    </Header>
  - <Structures>
      <TBL_NM>총조사인구 총괄(읍면동/성/연령별)</TBL_NM>
      <TBL_NM_ENG>Summary of Census Population(by administrative district/sex/age)</TBL_NM_ENG>
    </Structures>
</response>
```

eg. orgId(기관코드) = 101 , tblId(통계표 ID) = DT_1IN0001

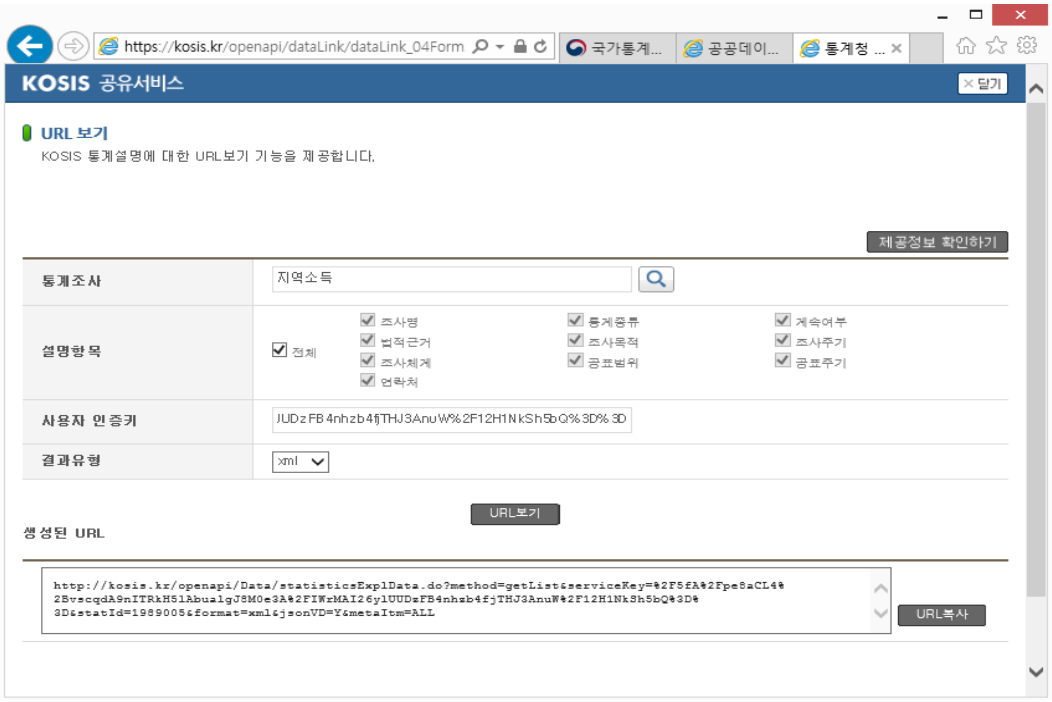
구분	결과변수	설명	예시
통계표	tblNm	통계표 국문명	총조사인구 총괄(읍면동/성/연령별)

명칭	tblNmEng	통계표 영문명	Summary of Census Population(by administrative district/sex/age)
기관 명칭	orgNm	기관 국문명	통계청
	orgNmEng	기관 영문명	Statistics Korea
수록정보	prdSe	수록주기	5 년
	strPrdDe	시작 수록시점	1925
	endPrdDe	종료 수록시점	2010
분류/항목	objId	분류 ID	A
	objNm	분류 국문명	행정구역별
	objNmEng	분류 영문명	By administrative divisions
	itmld	자료코드 ID	00
	itmNm	자료코드 국문명	전국
	itmNmEng	자료코드 영문명	Whole country
	upItmld	상위 자료코드	-
주석	CMMT_NM	주석유형	항목
	CMMT_DC	주석	성비는 여자 100 명당 남자의 수
	OBJ_ID	분류 ID	ITEM
	OBJ_NM	분류명	항목
	ITM_ID	자료코드 ID	T30
	ITM_NM	자료코드명	성비
단위	unitNm	단위 국문명	
	unitNmEng	단위 영문명	
출처	josaNm	조사명	인구총조사
	deptNm	통계표 담당부서	조사관리국 인구총조사과
	deptPhone	통계표 담당부서, 전화번호	042-481-3756(전수),042-481-3735(표본)
가중치	C1 ~ C8	분류값 ID1 ~ ID8	
	C1_NM ~ C8_NM	분류값 명 1 ~ 분류값 명 8	
	ITM_ID	항목 ID	
	ITM_NM	항목명	
	WGT_CO	가중치	

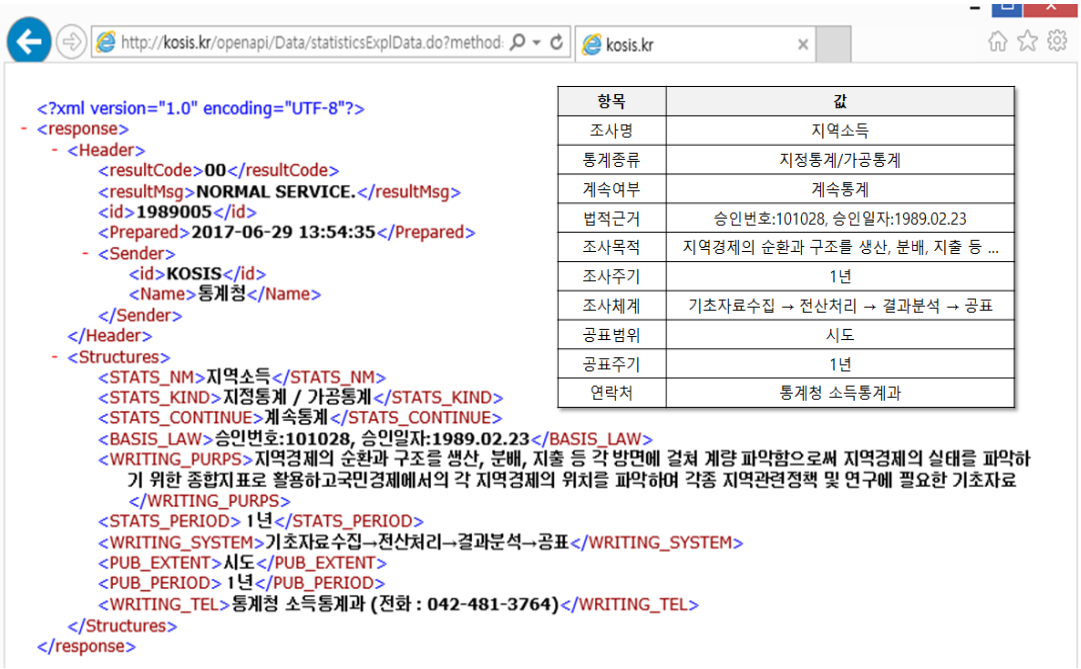
4.2.4 KOSIS Statistical Description Service

o Create Data Request URL

https://kosis.kr/openapi/dataLink/dataLink_04Form.jsp#



o Statistics table description data generated by the request URL



<?xml version="1.0" encoding="UTF-8"?>
<response>
<Header>
<resultCode>00</resultCode>
<resultMsg>NORMAL SERVICE.</resultMsg>
<id>1989005</id>
<Prepared>2017-06-29 13:54:35</Prepared>
<Sender>
<id>KOSIS</id>
<Name>통계청</Name>
</Sender>
</Header>
<Structures>
<STATS_NM>지역소득</STATS_NM>
<STATS_KIND>지정통계 / 가공통계</STATS_KIND>
<STATS_CONTINUE>계속통계</STATS_CONTINUE>
<BASIS_LAW>승인번호:101028, 승인일자:1989.02.23</BASIS_LAW>
<WRITING_PURPS>지역경제의 순환과 구조를 생산, 분배, 지출 등 각 방면에 걸쳐 계량 파악함으로써 지역경제의 실태를 파악하기 위한 종합지표로 활용하고국민경제에서의 각 지역경제의 위치를 파악하여 각종 지역관련정책 및 연구에 필요한 기초자료</WRITING_PURPS>
<STATS_PERIOD> 1년</STATS_PERIOD>
<WRITING_SYSTEM>기초자료수집→전산처리→결과분석→공표</WRITING_SYSTEM>
<PUB_EXTENT>시도</PUB_EXTENT>
<PUB_PERIOD> 1년</PUB_PERIOD>
<WRITING_TEL>통계청 소득통계과 (전화 : 042-481-3764)</WRITING_TEL>
</Structures>
</response>

항목	값
조사명	지역소득
통계종류	지정통계/가공통계
계속여부	계속통계
법적근거	승인번호:101028, 승인일자:1989.02.23
조사목적	지역경제의 순환과 구조를 생산, 분배, 지출 등 ...
조사주기	1년
조사체계	기초자료수집 → 전산처리 → 결과분석 → 공표
공표범위	시도
공표주기	1년
연락처	통계청 소득통계과

4.2.5 Indicator information inquiry service

o Display list by index name

요청변수(Request Parameter) 초기화 닫기

항목명	샘플데이터	설명
STAT_JIPYO_NM	실업률	조회할 지표명(URL(UTF-8) encoding) 입력

미리보기 XLS XML JSON CSV RDF

```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <response>
  - <header>
      <resultCode>0000</resultCode>
      <resultMsg>OK</resultMsg>
    </header>
  - <body>
    - <items>
      - <item>
          <areaTypeName>시도</areaTypeName>
          <endPrdDe>201604</endPrdDe>
          <prdDe>2016년4분기</prdDe>
          <prdSeName>분기</prdSeName>
          <rn>69</rn>
          <statJipyoId>148</statJipyoId>
          <statJipyoNm>실업률</statJipyoNm>
          <strtPrdDe>199903</strtPrdDe>
          <unit>%</unit>
        </item>
      - <item>
          <areaTypeName>국가</areaTypeName>
          <endPrdDe>2015</endPrdDe>
          <prdDe>2015년</prdDe>
          <prdSeName>년</prdSeName>
          <rn>25</rn>
          <statJipyoId>708</statJipyoId>
          <statJipyoNm>실업률</statJipyoNm>
          <strtPrdDe>1990</strtPrdDe>
          <unit>%</unit>
        </item>
      - <item>
          <areaTypeName>전국</areaTypeName>
          <endPrdDe>201705</endPrdDe>
          <prdDe>2017년5월</prdDe>
          <prdSeName>월</prdSeName>
          <rn>215</rn>
          <statJipyoId>1376</statJipyoId>
          <statJipyoNm>청년층(15~29세)실업률</statJipyoNm>
```

statJipyoId	지표ID
statJipyoNm	지표명
unit	단위
areaTypeName	지역구분명
prdSeName	수록주기명
strtPrdDe	수록시작시점
endPrdDe	수록종료시점
rn	수록시점 개수

o List by unique number

요청변수(Request Parameter) 초기화 닫기

항목명	샘플데이터	설명
STAT_JIPYO_ID	148	지표ID

미리보기 XLS XML JSON CSV RDF

```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <response>
  - <header>
      <resultCode>0000</resultCode>
      <resultMsg>OK</resultMsg>
    </header>
  - <body>
    - <items>
      - <item>
          <areaTypeName>시도</areaTypeName>
          <endPrdDe>201604</endPrdDe>
          <prdDe>2016년4분기</prdDe>
          <prdSeName>분기</prdSeName>
          <rn>69</rn>
          <statJipyoId>148</statJipyoId>
          <statJipyoNm>실업률</statJipyoNm>
          <strtPrdDe>199903</strtPrdDe>
          <unit>%</unit>
        </item>
      </items>
      <numOfRows>10</numOfRows>
      <pageNo>1</pageNo>
      <totalCount>1</totalCount>
    </body>
</response>
```

statJipyoId	지표ID
statJipyoNm	지표명
unit	단위
areaTypeName	지역구분명
prdSeName	수록주기명
strtPrdDe	수록시작시점
endPrdDe	수록종료시점
rn	수록시점 개수

o Detailed index by unique number

요청변수(Request Parameter) 초기화 닫기

항목명	샘플데이터	설명
STAT_JIPYO_ID	148	지표ID
RN	0	조회 기준이 되는 시점
SRV_RN	1	조회 기준 시점으로부터 조회할 시점의 개수
STRT_PRD_DE	201601	조회 시작 시점
END_PRD_DE	201601	조회 종료 시점

미리보기 XLS XML JSON CSV RDF

```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <response>
  - <header>
      <resultCode>0000</resultCode>
      <resultMsg>OK</resultMsg>
    </header>
  - <body>
    - <items>
      - <item>
          <itmNm>서울</itmNm>
          <prdDe>2016.1/4</prdDe>
          <prdSe>Q</prdSe>
          <statJipyoId>148</statJipyoId>
          <statJipyoNm>실업률</statJipyoNm>
          <val>4.7</val>
        </item>
      - <item>
          <itmNm>부산</itmNm>
          <prdDe>2016.1/4</prdDe>
          <prdSe>Q</prdSe>
          <statJipyoId>148</statJipyoId>
          <statJipyoNm>실업률</statJipyoNm>
          <val>4.7</val>
        </item>
      - <item>
          <itmNm>대구</itmNm>
          <prdDe>2016.1/4</prdDe>
          <prdSe>Q</prdSe>
          <statJipyoId>148</statJipyoId>
```

statJipyoId	지표ID
statJipyoNm	지표명
prdSe	수록주기
prdDe	시점
itmNm	항목
val	통계수치

o List by period in which data is recorded

요청변수(Request Parameter) 초기화 닫기

항목명	샘플데이터	설명
PRD_SE	M	수록주기

미리보기 XLS XML JSON CSV RDF

```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <response>
  - <header>
      <resultCode>0000</resultCode>
      <resultMsg>OK</resultMsg>
    </header>
  - <body>
    - <items>
      - <item>
          <areaTypeName>시도</areaTypeName>
          <endPrdDe>201512</endPrdDe>
          <prdDe>2015년12월</prdDe>
          <prdSeName>월</prdSeName>
          <rn>95</rn>
          <statJipyoId>20</statJipyoId>
          <statJipyoNm>강수량</statJipyoNm>
          <strtPrdDe>200801</strtPrdDe>
          <unit>mm</unit>
        </item>
      - <item>
          <areaTypeName>시도</areaTypeName>
          <endPrdDe>201412</endPrdDe>
          <prdDe>2014년12월</prdDe>
          <prdSeName>월</prdSeName>
          <rn>203</rn>
          <statJipyoId>26</statJipyoId>
          <statJipyoNm>경제활동참가율@</statJipyoNm>
          <strtPrdDe>199801</strtPrdDe>
          <unit>%</unit>
        </item>
```

statJipyoId	지표ID
statJipyoNm	지표명
Unit	단위
areaTypeName	지역구분명
prdSeName	수록주기명
strtPrdDe	수록시작시점
endPrdDe	수록종료시점
rn	수록시점 개수

o Detailed view by indicator name


```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <response>
  - <header>
      <resultCode>0000</resultCode>
      <resultMsg>OK</resultMsg>
    </header>
  - <body>
    - <items>
      - <item>
          <itmNm>전국</itmNm>
          <prdDe>2017.05</prdDe>
          <prdSe>M</prdSe>
          <statJipyoId>1285</statJipyoId>
          <statJipyoNm>실업률</statJipyoNm>
          <val>3.6</val>
        </item>
      - <item>
          <itmNm>전국</itmNm>
          <prdDe>2017.05</prdDe>
          <prdSe>M</prdSe>
          <statJipyoId>2639</statJipyoId>
          <statJipyoNm>실업률</statJipyoNm>
          <val>3.6</val>
        </item>
      - <item>
          <itmNm>서울</itmNm>
          <prdDe>2017.05</prdDe>
          <prdSe>M</prdSe>
          <statJipyoId>1286</statJipyoId>
          <statJipyoNm>실업률</statJipyoNm>
          <val>4.2</val>
        </item>
```

statJipyoId	지표ID
statJipyoNm	지표명
prdSe	수록주기
prdDe	시점
itmNm	항목
val	통계수치

4.2.6 List Indicator Search Service

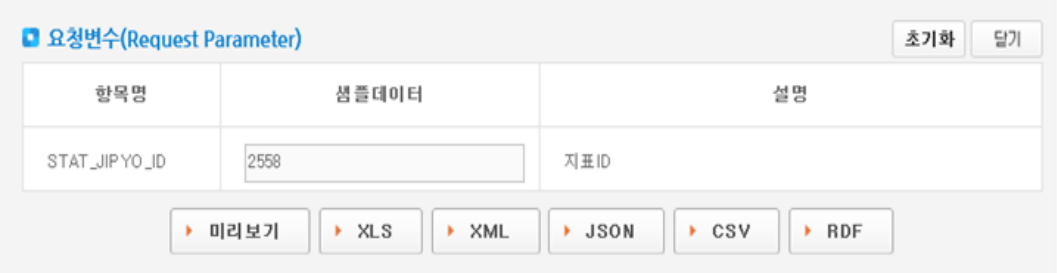
o View index by list


```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <response>
  - <header>
      <resultCode>0000</resultCode>
      <resultMsg>OK</resultMsg>
    </header>
  - <body>
    - <items>
      - <item>
          <areaTypeName>전국</areaTypeName>
          <endPrdDe>201706</endPrdDe>
          <listId>C_06</listId>
          <listNm>소비자물가지수(월)</listNm>
          <listSn1>2</listSn1>
          <listSn2>1</listSn2>
          <prdDe>2017년6월</prdDe>
          <prdSeName>월</prdSeName>
          <rn>629</rn>
          <statJipyoId>2558</statJipyoId>
          <strtPrdDe>196501</strtPrdDe>
          <unit>2015=100</unit>
        </item>
      - <item>
          <areaTypeName>시도</areaTypeName>
          <endPrdDe>201706</endPrdDe>
          <listId>C_06</listId>
          <listNm>소비자물가지수(월)</listNm>
          <listSn1>2</listSn1>
          <listSn2>2</listSn2>
          <prdDe>2017년6월</prdDe>
          <prdSeName>월</prdSeName>
          <rn>629</rn>
          <statJipyoId>2557</statJipyoId>
          <strtPrdDe>196501</strtPrdDe>
          <unit>2015=100</unit>
        </item>
```

listId	세부목록ID
listNm	세부부문명
statJipyoId	지표ID
statJipyoNm	지표명
unit	단위
areaTypeName	지역구분명
prdSeName	수록주기명
strtPrdDe	수록시작시점
endPrdDe	수록종료시점
rn	수록시점 개수
listSn1	목록순서1
listSn2	목록순서2
prdDe	시점

4.2.7 Display description data by indicator number

o Look up concepts by unique number



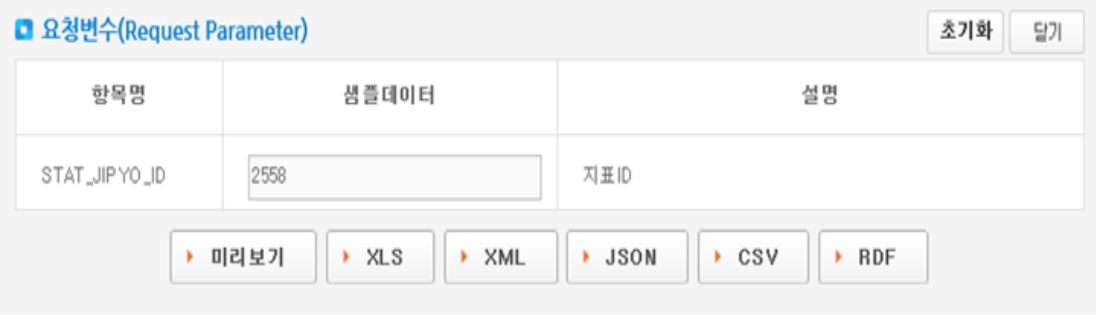

```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <response>
  - <header>
      <resultCode>0000</resultCode>
      <resultMsg>OK</resultMsg>
    </header>
  - <body>
    - <items>
      - <item>
          <jipyoExplan>소비자물가지수</jipyoExplan>
          <statExplan1>가구에서 소비생활을 영위하기 위해 구입하는 상품가격과 서비스 요금의 변동을 종합적으로 측정하기 위해 작성하는 지수</statExplan1>
        </item>
      </items>
      <numOfRows>10</numOfRows>
      <pageNo>1</pageNo>
      <totalCount>1</totalCount>
    </body>
</response>
```

o Calculation method and source search by unique number


```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <response>
  - <header>
      <resultCode>0000</resultCode>
      <resultMsg>OK</resultMsg>
    </header>
  - <body>
    - <items>
      - <item>
          <jipyoExplan>소비자물가지수</jipyoExplan>
          <statExplan2>매월 상품가격과 서비스 요금의 변동을 측정, 기준시 고정 가중산술평균 방식(라스파이레스 산식)을 적용하여 산정</statExplan2>
          <statExplan3>통계청, 소비자물가조사</statExplan3>
        </item>
      </items>
      <numOfRows>10</numOfRows>
      <pageNo>1</pageNo>
      <totalCount>1</totalCount>
    </body>
</response>
```

jipyoExplan	설명자료 제목
statExplan2	선정방법
statExplan3	출처정보

o Full description data by unique number




```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <response>
  - <header>
      <resultCode>0000</resultCode>
      <resultMsg>OK</resultMsg>
    </header>
  - <body>
    - <items>
      - <item>
          <jipyoExplan>소비자물가지수</jipyoExplan>
          <statExplan1>가구에서 소비생활을 영위하기 위해 구입하는 상품가격과 서비스 요금의 변동을 종합적으로 측정하기 위해 작성하는 지수</statExplan1>
          <statExplan2>매월 상품가격과 서비스 요금의 변동을 측정, 기준시 고정 가중산술평균 방식(라스파이레스 산식)을 적용하여 산정</statExplan2>
          <statExplan3>통계청, 소비자물가조사</statExplan3>
        </item>
      </items>
      <numOfRows>10</numOfRows>
      <pageNo>1</pageNo>
      <totalCount>1</totalCount>
    </body>
</response>
```

jipyoExplan	설명자료 제목
statExplan1	개념
statExplan2	선정방법
statExplan3	출처정보

4.2.8 Displaying Explanation Data by Indicator Name

o Full description data by indicator name


```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <response>
  - <header>
      <resultCode>0000</resultCode>
      <resultMsg>OK</resultMsg>
    </header>
  - <body>
    - <items>
      - <item>
          <jipyoExplan>고용률</jipyoExplan>
          <statExplan1>만 15세 이상 인구 중 취업자가 차지하는 비율</statExplan1>
          <statExplan2>고용률 = (취업자/만 15세이상 인구) × 100</statExplan2>
          <statExplan3>통계청, 경제활동인구조사</statExplan3>
          <statJipyoNm>고용률</statJipyoNm>
        </item>
```

jipyoExplan	설명자료 제목
statExplan1	개념
statExplan2	산정방법
statExplan3	출처