Municipal Water Network Infrastructure Database Management

Repository for Asset Management and Pipeline Replacement Scoring

Priyadarshini
Civil Engineering
Carnegie Mellon University
Pittsburgh, PA USA
priyada2@andrew.cmu.edu

Devashri Karve
Civil Engineering
Carnegie Mellon University
Pittsburgh, PA USA
dkarve@andrew.cmu.edu

Guillermo Gutierrez
Civil Engineering
Carnegie Mellon University
Pittsburgh, PA USA
ggutierr@andrew.cmu.edu

Tanay Kulkarni
Civil Engineering
Carnegie Mellon University
Pittsburgh, PA USA
tskulkar@andrew.cmu.edu

MOTIVATION

Water infrastructure is one of the most critical urban infrastructure systems. This system consists of clean water, wastewater, and stormwater infrastructure systems of pipes, valves, pumps, tanks, sensors, treatment plants, and reservoirs. These interconnected assets form a vast network system that ensures water transportation from the source to consumers. Every year millions of dollars are spent on the maintenance, repairs, replacement, and rehabilitation of these assets whose lives depend on multiple inter-related parameters. Hypothetically, Coquitlam, Canada, is looking for a pipeline maintenance company to manage the water infrastructure and maintain installed equipment. They have provided detailed data of installed water infrastructure and each equipment's descriptors. We will submit a bid and offer an optimized pipeline maintenance schedule. The goal will be to establish the decision-making framework and zonal prioritization based on a linear model and an independent scoring system created for the city-data. The deliverable will rank all the city zones in order of priority that city officials can use to coordinate budget allocation.

KEYWORDS

Water distribution network, Asset management, Database management systems, Network infrastructure, Pipe replacement, Break prediction, Machine Learning

1 Data

The water network elements have attribute metadata associated with them. These metadata help better understand the physical properties and functions of these assets. This data can be available in various formats like Spreadsheets, PDFs, CAD, Shapefiles, etc. The information for the following vital assets has been collected and cleaned for this project:

- Hydrant Feed - water pipe that feeds water to hydrants
- Hydrants - hydrants installed throughout the city
- Main - main water pipelines that feed certain areas of the zones
- Service connection - lateral main (usually of smaller diameters) connected to the mains that serve consumers
- Valves - used to control water flow

2 Data Collection and Processing

The GIS data for the water distribution network for Coquitlam is available1. GIS Shapefiles for Mains, Valves, Hydrants, Hydrant Feeds, Service Connections, out of all the available shapefiles, are used for this project. The GIS tool QGIS Software (https://qgis.org/) is used to access the geospatial (Figure 1). Unique key identifiers for individual elements are missing in the imported dataset and are defined explicitly in QGIS. The finalized attribute data table headers as shown in Tables A.4 - A.9. There are a total of 6744 Mains, 2856 Hydrants, 2828 Hydrant Feeds, 7883 Valves, and 19167 Service Connections assets considered for this data.

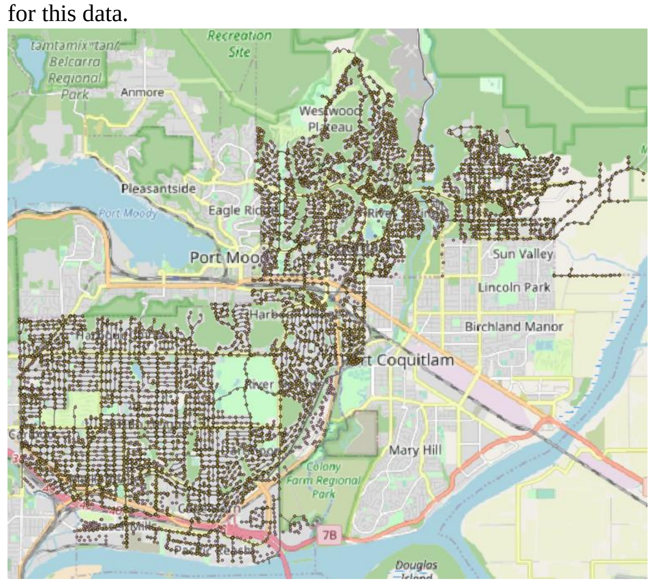

Figure1: The geospatial assets in QGIS

The geospatial joining of the spatial data establishes the spatial relationships. E.g., hydrants connect to the mains via hydrant feeders, the valves connect to the mains, and the service connections branch out from the Mains. Associating a linear network relation is done from within the QGIS tool.

The city data does not provide historic maintenance performed or the number of pipe breaks; thus, the city's data will not develop a prediction model. Using the data from the research study, "Remaining Useful Life Prediction of Water Pipes Using Artificial Neural Network and Adaptive Neuro-Fuzzy Inference System Model" by Razieh Tavakoli, we developed a linear model that we could apply to the city's data for predicting the number of breaks per pipe using the following predictor variables: Age as of 2021, Log Length of pipe, pipe diameter, and pipe material[2]. Since the hydraulic data was not available for both sample and project data, the three physical properties of pipes are responsible for the change in pipe's condition over time.

The processing of the GIS data was done in conjunction with the collection and model building of the research sample data to ultimately be applied to the project data for analysis and results, as is shown in Figure 2.

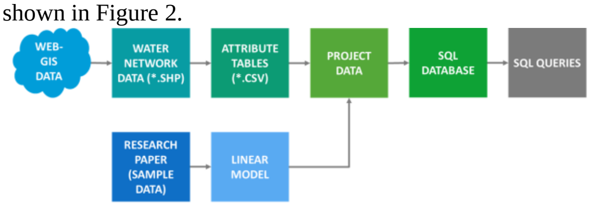

Figure 2: shows the data processing before model building and analyzing.

3 Database Development

Organizing all the data from various sources within a single SQL database is necessary. MySQL is used to develop such a database within the MySQL Workbench tool. Detailed SQL queries are in the Appendix. Initially, tables are created for each asset, viz. main, valve, service connection, hydrant feed, hydrant. All the attributes are imported from excel spreadsheets that were from GIS. Tables for Age, Material, Diameter Replacement Factors are imported within the database. Each table has Primary keys.

The ERD diagram in the Appendix shows the organization of the table data and the types of relationships each has. The assets viz. main, service connection, valve, hydrant feed, and hydrant are identified as the entities. These entities are related to the 'main' entity. The Replacement Factor tables, i.e., Age, Material, Diameter Replacement Tables, are identified as individual entities. These are related to the main entity via a weak entity called 'joined_factors.' The relation between these entities is explained in table A.10 and the ERD in Figure A.1 in Appendix.

The ARF, DRF, and MRF tables join with the MAIN table based on the mains' calculated age, diameter, and material. A normalized score, called Pipe Replacement Factor (PRF), is calculated for every pipe in the main table. The average values of these replacement factors (ARF, MRF, DRF, PRF) are identified grouped by 'WZone_ID' for all the water zones.

To identify the most critical Zone, queries are written to present Zone-wise scores and the predicted number of breaks at the 0th year and at the 30th year. Subsequently, Top Priority pipes in that Zone are identified. For improved understanding of the impact during the replacement of critical assets, queries are written to help the user understand the other assets such as service connections, hydrant feeds, hydrants that are connected to the critical pipes.

4 Model Development

The regression model is described in equation 1. The sample data was split 70%/30% as a train/test dataset. After finding the number of breaks predicted per pipe, we were able to get a cumulative number of predicted pipe breaks per Zone that would be used as part of the decision-making process.

$$NBreak = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \quad \ldots (1)$$

Where:
β_0 = constant intercept derived from regression
β_1 = coefficient for Age as of 2021, X1
β_2 = coefficient for Log Length in meters, X2
β_3 = coefficient for Metal (indicator variable), X3
β_4 = coefficient for Diameter in mm, X4

A scoring-based decision criterion is devised for robust decision-making. This will give the city information on where to deploy the budget for higher capital efficiency by identifying zonal pipe replacement scores. Each attribute (Age, Diameter, and Material) will have a score of 0-100, where 100 is the highest priority to replace infrastructure equipment. In the Appendix, Table A.1 has the Age Replacement Factor (arf), Table A.2 has the Diameter Replacement Factor (drf), and Table A.3 has the Material Replacement Factor (mrf). A Pipe Replacement Factor (PRF) for each pipe was calculated by taking the average of all replacement factors as in equation 2.

$$PRF_i = \frac{arf_i + drf_i + mrf_i}{3} \quad \ldots (2)$$

Where i is the pipe number, after calculating the PRF for each pipe installed in each Zone, we found the average of the PRF per Zone.

5 Analysis

Initially, we looked at the distribution of the sample data found in the research paper. We found that length in meters was right-skewed while every other factor followed a normal distribution, as shown in figure 3.

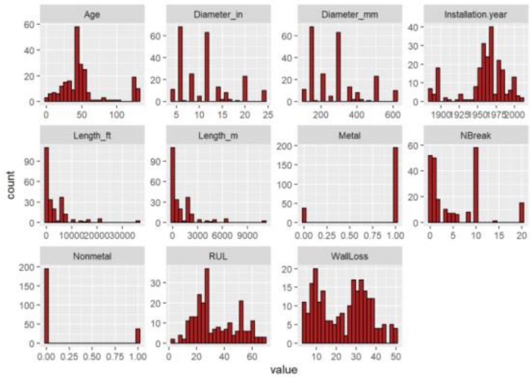

Figure 3 shows the distributions of all the different variables out of which Length_m, Age, Diameter_mm is used for this study.

Because the length data were right-skewed, we decided to take the log of the dataset and use that as the variable in the regression model for better representation of the variable in describing the dataset. This resulted in the length data taking a distribution closer to a normal distribution per figure 4 below.

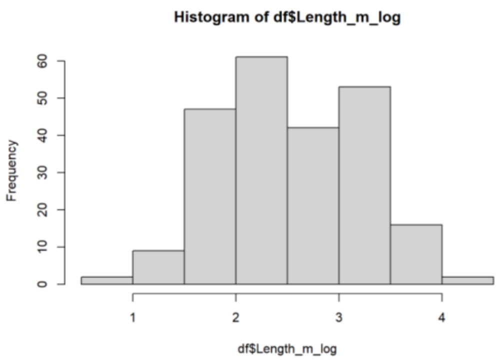

Figure 4 shows the distribution of the log of pipe length in meters. We chose to make the "Metal" variable as a binary indicator variable that would tell us whether the pipe was using a metal-based material because the data in the research paper used to develop the predictor model did not have the same amount of materials in its dataset as the city data found. While this comparison introduces a bias, given that we used it to make predictions of a dataset that does not contain the same types of data for the material, it made sense as the non-metals in the research sample data holistically had the same physical properties as the non-metals in the city data set. This meant that if the variable is binary, the differences for each type of material would be part of the error of the variable derived in the regression. You can view the research paper sample data set in reference number 2, from pages 154 - 161.

6 Results

The regression resulted in an R^2= 0.356, meaning that the model can describe ~35.6% of the dataset. This is a strong correlation for this type of data. One of our references to understand the breaking of pipe showed a correlation between Break Rate (breaks/(100 mi-year)) and Water Loss Percentage to have an R^2= 0.0794 [3]. While these are different data, it portrays that a breaking pipe can have so many contributing factors in addition to those analyzed in this study that having such a considerable measure of fit by the model means it is a robust representation of the data in this use. The results of the regression are in table 1 below.

Coefficients:

	Estimate	Std. Error	t value	Pr (>| t|)
(Intercept)	-8.050946	2.517819	-3.198	0.00168
Age	0.008912	0.014225	0.627	0.53189
Length_m_log	5.070169	0.589385	8.602	7.57E-15
Metal	0.915564	1.057249	0.866	0.38782
Diameter_mm	-0.002988	0.003149	-0.949	0.34426

Table 1 shows the estimates for the coefficients and their statistical parameters and descriptors.

We can see that the most statistically significant coefficients are the intercept and the log length of the pipe. This means that there are many other factors that we do not include in the model that may also have a correlation. It also means that there is a high level of variance in 3 of our four attributes, and thus they are not highly correlated to the number of breaks at a 95% confidence interval. Essentially, we could remove those that did not show correlation at this level of confidence.

Because of the lack of data and knowledge of pipe failure, we know that these are some of the most critical factors in a pipe being compromised; thus, we decided to continue using this model as the predictor. Additionally, we must assume that the current variables carry all the weight in describing the number of breaks because we don't have situational or physical data like water pressure and pipe stresses. Given that our goal is to make an optimized schedule rather than predict the break, it serves a good purpose in providing context for the data.

Table 2 below shows the data ranked by number of breaks.

WZONE_ID	SUM (Number of breaks)	AVG (AVG_PRFi)
WZONE_8	3664.50	72.22
WZONE_24	1340.60	63.11
WZONE_4	1325.60	68.73
WZONE_13	1187.80	60.31
WZONE_7	571.50	68.43
WZONE_23	529.90	62.61
WZONE_3	512.50	58.19
WZONE_20	505.80	54.60
WZONE_12	458.70	51.54
WZONE_19	358.30	58.96
WZONE_16	337.40	50.03

WZONE_17	331.20	57.52
WZONE_21	203.90	57.43
WZONE_18	179.20	59.97
WZONE_10	124.50	70.47
WZONE_1	122.50	65.40
WZONE_6	109.90	66.93
WZONE_22	106.70	55.68
WZONE_15	97.40	63.04
WZONE_11	68.20	74.09
WZONE_2	33.70	57.33
WZONE_5	26.10	70.18
WZONE_14	24.60	56.23
WZONE_9	12.60	77.10

The highest priority zone will be W_Zone8. We found that our ranking is somewhat correlated to the number of breaks. As you can see, a couple of zones have high PRF but a low predicted number of breaks, while numbers 2-6 have a high PRF number of >60. This means that they do not expect as many breaks because they have short sections even though their material and age may place them at a higher risk of failure. We recommend using the number of breaks as the 1st pass while using the scoring system to fine-tune the schedule.

ACKNOWLEDGMENTS

The authors thank the City of Coquitlam, Canada, for providing their data to perform this academic study.

REFERENCES

[1] Water Utility https://data.coquitlam.ca/maps/83ea9a808f6b455299997545a2e946ab/about (accessed 2021 -12 -06).
[2] Tavakoli, R. REMAINING USEFUL LIFE PREDICTION OF WATER PIPES USING ARTIFICIAL NEURAL NETWORK AND ADAPTIVE NEURO FUZZY INFERENCE SYSTEM MODELS. 169.
[3] Water Main Break Rates in the USA and Canada: A Comprehensive Study. 2018, 48.

Appendix

Year	Age Replacement Factor
1945	100.00
1950	98.61
1951	97.22
1952	95.83
1953	94.44
1954	93.05
1955	91.66
1956	90.27
1957	88.88
1958	87.50
1959	86.11
1960	84.72
1961	83.33
1962	81.94
1963	80.55
1964	79.16
1965	77.77
1966	76.38
1967	75.00
1968	73.61
1969	72.22
1970	70.83
1971	69.44
1972	68.05
1973	66.66
1974	65.27
1975	63.88
1976	62.50
1977	61.11
1978	59.72
1979	58.33
1980	56.94
1981	55.55
1982	54.16
1983	52.77
1984	51.38

1985	50.00
1986	48.61
1987	47.22
1988	45.83
1989	44.44
1990	43.05
1991	41.66
1992	40.27
1993	38.88
1994	37.50
1995	36.11
1996	34.72
1997	33.33
1998	31.94
1999	30.55
2000	29.16
2001	27.77
2002	26.38
2003	25.00
2004	23.61
2005	22.22
2006	20.83
2007	19.44
2008	18.05
2009	16.66
2010	15.27
2011	13.88
2012	12.50
2013	11.11
2014	9.72
2015	8.33
2016	6.94
2017	5.55
2018	4.16
2019	2.77
2020	1.38

Table A.1: shows the scores, 0-100, given to the age of the file based on the installation data of the pipe, or how old the pipe is. The higher the score the higher the priority to replace.

Diameter (mm)	**Diameter Replacement Factor**

25	100
50	95
75	90
100	85
150	80
200	75
250	70
300	65
350	60
400	55
450	50
500	45
600	40
650	35
750	30
900	25
1200	20
1500	15
1800	10
1950	5

Table A.2: shows the scores, 0-100, given to the diameter data. The higher the score the higher the priority to replace.

Material	**Material Replacement Factor**
Steel	11
Galvanized Steel	22
Copper	33
HDPE	44
Ductile Iron	56
PVC	67
Concrete	78
Cast Iron	89
Asbestos Cement	100

Table A.3: shows the scores 0-100, given to the material data. The scores for the materials were derived based on experience.

STATUS	**INSTALL_DT**	**MAIN_SIZE**	**LENGTH**	**TYPE**	**WM_ID**	**WZONE_ID**

Table A.4: Attribute Fields for Main(s)

STATUS	**INSTALL_DT**	**DIAMETER**	**MAT_TYPE**	**WSC_ID**	**WZONE_ID**	**WM_ID**

Table A.5: Attribute Fields for Service Connection(s)

STATUS	TYPE	VALVE_STS	VALVE_SIZE	WVAL_ID	WZONE_ID	WM_ID

Table A.6: Attribute Fields for Valve(s)

STATUS	INSTALL_DT	MODEL	FLOW_CLASS	WH_ID	WZONE_ID	WHF_ID

Table A.7 : Attribute Fields for Hydrant(s)

STATUS	WHF_ID	WZONE_ID	WM_ID

Table A.8: Attribute Fields for Hydrant Feed(s)

Asset	Unique Identifier
Main	WM_ID
Service Connection	WSC_ID
Valve	WVAL_ID
Hydrant	WH_ID
Hydrant Feed	WHF_ID
Zone	WZONE_ID

Table A.9 : Unique Identifiers for Asset(s)

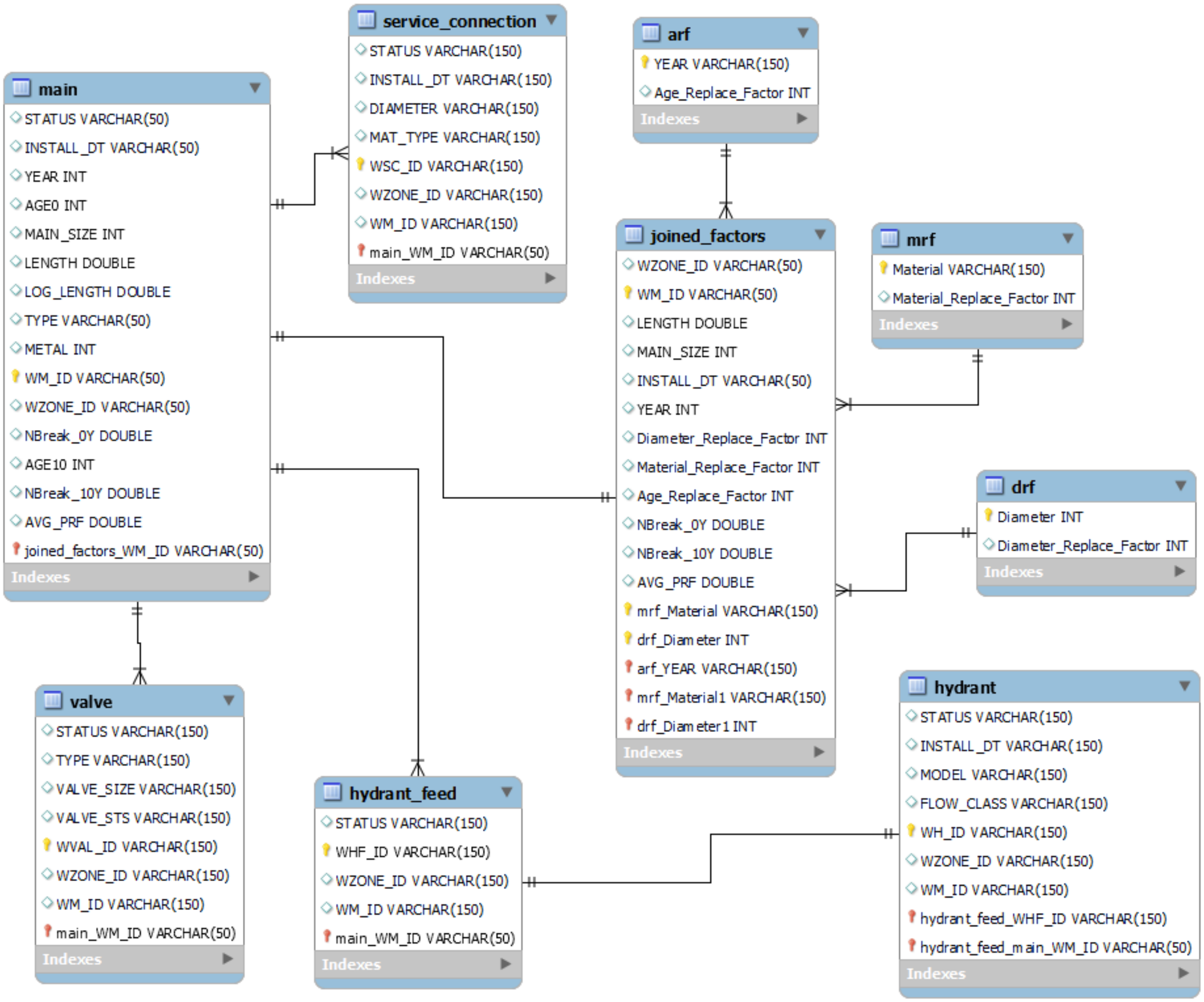

Figure A.1: shows the ERD we designed for the data set and the relationships between the entities. As you can see, we are saying that each pipe is unique and may feed many different hydrants and have different valves. Additionally, we will have many different joined_factors that will take only one arf, mrf, drf, as each pipe will have their own corresponding factor that is unique to it.

Entity 1	Relation	Entity 2
Valve	Many to One	Main
Service Connection	Many to One	Main
Hydrant Feed	Many to One	Main
Hydrant	One to One	Hydrant Feed
ARF	One to Many	joined_factors
MRF	One to Many	joined_factors
DRF	One to Many	joined_factors
joined_factors	One to One	Main

Table A.10: shows the relationships between the different entities in our model.

WZONE_ID	AVG (Diameter Replacement Factor = drfi)	AVG (Material Replacement Factor = mrfi)	AVG (Age Replacement Factor = arfi)	SUM (Number of breaks)	SUM (Number of Breaks at 10Y from today)	AVG (AVG_PRFi)
WZONE_8	77.19	74.37	65.10	3664.50	3965.80	72.22
WZONE_24	73.23	63.69	52.41	1340.60	1472.10	63.11
WZONE_4	76.78	69.58	59.82	1325.60	1435.90	68.73
WZONE_13	74.19	60.21	46.54	1187.80	1317.00	60.31
WZONE_7	76.41	70.23	58.64	571.50	621.50	68.43
WZONE_23	77.69	61.62	48.52	529.90	585.20	62.61
WZONE_3	74.23	58.60	41.74	512.50	568.40	58.19
WZONE_20	75.00	58.21	30.58	505.80	559.70	54.60
WZONE_12	72.80	57.50	24.33	458.70	514.10	51.54
WZONE_19	72.78	62.80	41.30	358.30	410.40	58.96
WZONE_16	70.63	57.01	22.46	337.40	379.70	50.03
WZONE_17	71.98	61.71	38.85	331.20	375.90	57.52
WZONE_21	74.14	61.70	36.46	203.90	232.00	57.43
WZONE_18	73.09	60.57	46.25	179.20	197.70	59.97
WZONE_10	80.31	67.52	63.59	124.50	135.10	70.47
WZONE_1	77.82	64.36	54.01	122.50	140.90	65.40
WZONE_6	76.09	70.35	54.35	109.90	120.40	66.93
WZONE_22	71.34	61.56	34.13	106.70	118.30	55.68
WZONE_15	72.68	67.11	49.33	97.40	109.70	63.04
WZONE_11	79.85	80.00	62.42	68.20	73.10	74.09
WZONE_2	81.67	45.89	44.44	33.70	35.40	57.33
WZONE_5	80.00	72.00	58.55	26.10	28.80	70.18
WZONE_14	67.41	58.79	42.48	24.60	27.50	56.23
WZONE_9	81.43	84.29	65.57	12.60	13.30	77.10

Table A.11: shows the results of the zonal average of each factor, the predicted number of breaks, predicted number of breaks after 10 years, and the zonal average PRF.

SQL Script and Queries:

```
# clear the existing database
#create schema `new_db` ;
# create a new schema
#drop schema `new_db`;
use new_db;
###################################
# create 'main' table.no.1
create table main(
STATUS    varchar(50),
INSTALL_DT    varchar(50),
YEAR int,
AGE0 int,
MAIN_SIZE    int,
LENGTH    double,
LOG_LENGTH double,
TYPE    varchar(50),
METAL int,
```

```
WM_ID    varchar(50),
WZONE_ID varchar(50),
NBreak_0Y double,
AGE10 int,
NBreak_10Y double,
AVG_PRF double);

# IGNORE THIS LINE // SET GLOBAL local_infile=1;
# IGNORE THIS LINE // SHOW VARIABLES LIKE 'secure_file_priv';
# populate table.no.1
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/main.csv'
INTO TABLE main
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;

#change date varchar to date format
UPDATE `main`
SET `INSTALL_DT` = str_to_date( `INSTALL_DT`, '%m/%d/%YYYY' );

################################
# create 'hydrant' table.no.2
create table hydrant(
STATUS    varchar(150),
INSTALL_DT    varchar(150),
MODEL    varchar(150),
FLOW_CLASS    varchar(150),
WH_ID varchar(150),
WZONE_ID varchar(150),
WM_ID    varchar(150)
);
#truncate table hydrant;
#drop table hydrant;

# populate table.no.2
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/Hydrant.csv'
INTO TABLE hydrant
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;

#change date varchar to date format
UPDATE `hydrant`
SET `INSTALL_DT` = str_to_date( `INSTALL_DT`, '%m/%d/%YYYY' );

################################
# create 'hydrant_feed' table.no.3
create table hydrant_feed (
STATUS    varchar(150),
WHF_ID varchar(150),
WZONE_ID varchar(150),
WM_ID    varchar(150)
);
#truncate table hydrant_feed;
#drop table hydrant_feed;

# populate table.no.3
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/Hydrant_Feed.csv'
INTO TABLE hydrant_feed
```

```
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;

################################
# create 'service_connection' table.no.4
create table service_connection (
STATUS    varchar(150),
INSTALL_DT varchar(150),
DIAMETER varchar(150),
MAT_TYPE varchar(150),
WSC_ID varchar(150),
WZONE_ID varchar(150),
WM_ID    varchar(150)
);
#truncate table service_connection;
#drop table service_connection;

# populate table.no.4
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/Service_Connection.csv'
INTO TABLE service_connection
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;

#change date varchar to date format
UPDATE `service_connection`
SET `INSTALL_DT` = str_to_date( `INSTALL_DT`, '%m/%d/%YYYY' );

################################
# create 'valve' table.no.5
create table valve (
STATUS    varchar(150),
TYPE varchar(150),
VALVE_SIZE varchar(150),
VALVE_STS varchar(150),
WVAL_ID varchar(150),
WZONE_ID varchar(150),
WM_ID    varchar(150)
);
#truncate table valve;
#drop table valve;

# populate table.no.5
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/Valve.csv'
INTO TABLE valve
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;


################################
# create 'ARF' table.no.6
create table ARF (
YEAR    varchar(150),
Age_Replace_Factor varchar(150)
);
#truncate table ARF;
```

```
#drop table ARF;

# populate table.no.6
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/AGE_Replacement_Factors.csv'
INTO TABLE ARF
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;
#change date varchar to date format
UPDATE `arf`
SET `YEAR` = str_to_date( `YEAR`, '%YYYY' );

################################
# create 'DRF' table.no.7
create table DRF (
Diameter    int,
Diameter_Replace_Factor int
);
#truncate table DRF;
#drop table DRF;

# populate table.no.7
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/DIAMETER_Replacement_Factors.csv'
INTO TABLE DRF
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;

################################
# create 'MRF' table.no.8
create table MRF (
Material varchar(150),
Material_Replace_Factor int
);
#truncate table MRF;
#drop table MRF;

# populate table.no.8
LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/MATERIAL_Replacement_Factors.csv'
INTO TABLE MRF
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\r\n'
IGNORE 1 ROWS;

#####CREATE A NEW TABLE "JOINED_FACTORS" BY JOINING MAIN+ARF+MRF+DRF
CREATE TABLE Joined_Factors
SELECT main.WZONE_ID, main.WM_ID, main.LENGTH, main.MAIN_SIZE, main.INSTALL_DT, main.YEAR,
drf.Diameter_Replace_Factor, mrf.Material_Replace_Factor, arf.Age_Replace_Factor,
main.NBreak_0Y, main.NBreak_10Y, main.AVG_PRF
FROM main
JOIN drf ON main.MAIN_SIZE=drf.Diameter
JOIN mrf ON main.TYPE=mrf.Material
JOIN arf ON main.YEAR=YEAR(arf.YEAR)
ORDER BY main.MAIN_SIZE ASC;

#####POPULATING AVE_PRF and PRESENTING AVG replacement factors grouped by ***ZONE***

## UPDATING AVG_PRF
```

```
SELECT * FROM new_db.joined_factors;
UPDATE joined_factors
SET joined_factors.AVG_PRF = (joined_factors.Diameter_Replace_Factor + joined_factors.Age_Replace_Factor +
joined_factors.Material_Replace_Factor)/3;

## Average FACTORS, and Sum of NBreaks grouped by ZONES
SELECT WZONE_ID,
AVG(Diameter_Replace_Factor), AVG(Material_Replace_Factor), AVG(Age_Replace_Factor),
SUM(NBreak_0Y), SUM(NBreak_10Y), AVG(AVG_PRF)
from joined_factors
GROUP BY WZONE_ID
ORDER BY SUM(NBreak_0Y) DESC;

## SELECT TOP 5 PIPES from the Highest Priority Zone that needs replacement based on NBreak
SELECT WM_ID, WZONE_ID,
(Diameter_Replace_Factor), (Material_Replace_Factor), (Age_Replace_Factor),
(NBreak_0Y), (NBreak_10Y), (AVG_PRF)
from joined_factors
WHERE WZONE_ID = 'WZONE_8'
ORDER BY (NBreak_0Y) DESC
LIMIT 5;

## SELECT TOP 5 PIPES from the Highest Priority Zone that needs replacement based on AVG_PRF
SELECT WM_ID, WZONE_ID,
(Diameter_Replace_Factor), (Material_Replace_Factor), (Age_Replace_Factor),
(NBreak_0Y), (NBreak_10Y), (AVG_PRF)
from joined_factors
WHERE WZONE_ID = 'WZONE_8'
ORDER BY (AVG_PRF) DESC
LIMIT 5;

## SELECT Hydrant_Feed connected to Highest Priority Pipe from Highest Priority Zone
SELECT *
FROM new_db.hydrant_feed
WHERE (WM_ID LIKE '%WM_6527%') OR (WM_ID LIKE '%WM_2052%');

## SELECT Service_Connection connected to Highest Priority Pipe from Highest Priority Zone
SELECT *
FROM new_db.service_connection
WHERE (WM_ID LIKE '%WM_6527%') OR (WM_ID LIKE '%WM_2052%');

## SELECT Valve connected to Highest Priority Pipe from Highest Priority Zone
SELECT *
FROM new_db.valve
WHERE (WM_ID LIKE '%WM_6527%') OR (WM_ID LIKE '%WM_2052%');
```