Implications for construction quality and sustainability

STAT 31631 – Statistical Modelling

Department of Statistics & Computer Science

University of Kelaniya

Academic Year 2022/2023

By

Group 03

Group Details

PS/2020/018 – W.D.Sandamali

PS/2020/110 – D.D.A.K.Indeewari

PS/2020/130 – H.P.I.Sewwandi

PS/2020/132 – H.A.H.Dimalsha

PS/2020/135 – W.T.N.Perera

PS/2020/238 – D.W.P.S.Gunathilaka

PS/2020/302 – E.H.A.Thathsarani

Content

01.Introduction

This study employs a quantitative research design to analyse how various concrete mix components affect construction quality and sustainability, with a focus on the impact of material proportions and the age of the concrete on its compressive strength. Compressive strength is a critical measure of concrete's load bearing capacity. The study examines the following independent variables: Cement, Fly Ash, Blast Furnace Slag, Water, Superplasticizer, Coarse Aggregate, Fine Aggregate, and Age (measured in days).

Cement binds the mix, while Fly Ash and Blast Furnace Slag improve durability and sustainability. Water is necessary for hydration but must be controlled to maintain strength. Superplasticizers enhance workability without extra water, and aggregates provide structural bulk. The curing age is crucial, as strength increases over time due to ongoing hydration.

The research aims to determine optimal mix proportions and curing periods to maximize compressive strength while promoting sustainability with supplementary materials like Fly Ash and Blast Furnace Slag. The study's findings will contribute to more effective and eco-friendly concrete mix designs, enhancing construction quality and reducing the environmental impact of concrete production.

Purpose and goals of the study

The overarching goal of this study is to develop concrete mix designs that enhance both compressive strength and sustainability by optimizing the proportions of Cement, Fly Ash, Blast Furnace Slag, Water, Superplasticizers, Coarse Aggregates, Fine Aggregates, and curing time. Specifically, the study has the following objectives:

⁕. To analyze the influence of Cement, Fly Ash, Blast Furnace Slag, Water, Superplasticizer, Coarse Aggregate, and Fine Aggregate on the compressive strength of concrete.

⁕. To examine how the age of concrete (measured in days) affects its compressive strength.

⁕. To identify optimal mix proportions that maximize compressive strength while ensuring sustainability.

⁕. To provide recommendations for concrete mix designs that can be applied in the construction industry to improve quality and sustainability.

Significance of the study

This research is significant because it addresses two major concerns in modern construction: the need for high-strength materials and the demand for environmentally sustainable building practices. Although many studies have examined the strength of concrete, few have focused on how the use of supplementary materials like Fly Ash and Blast Furnace Slag can both reduce environmental impacts and contribute to long-term strength development.

The novelty of this study lies in its holistic approach. By considering the effects of multiple concrete mix components and the curing process simultaneously, the research provides insights into the complex relationships between material proportions, curing age, and compressive strength. Additionally, it advances the understanding of how sustainable materials like Fly Ash and Blast Furnace Slag can replace traditional components without compromising performance, paving the way for more eco-friendly construction practices.

02. Problem Statement

In our project, we are focused on evaluating the strength of cement used in construction. for this, we will use some datasets taken from previous researches. These tests are designed to assess how well the cement can withstand different types of stress and strain, providing crucial information about its performance.

The primary goal of our evaluation is to ensure that the cement meets established industry standards and specifications. By conducting these tests, we aim to identify any potential weaknesses or inconsistencies in the cement's strength, which is essential for making informed decisions about its suitability for different construction applications.

Our findings will play a vital role in improving the overall quality and durability of construction projects.

03. Methodology

1.Design

This study employs a quantitative research design to analyze the implications of various concrete mix components on construction quality and sustainability. The primary focus is on understanding how the proportions of different materials and the age of the concrete influence its compressive strength, which is a critical measure of construction quality.

2. Data Collection

• Sample Selection: The study will use a dataset containing records of different concrete mixtures with varying proportions of components and their respective compressive strengths.

• Variables:

- o Independent Variables:
  - ▪ Cement (kg in a m^3 mixture)
  - ▪ Blast Furnace Slag (kg in a m^3 mixture)
  - ▪ Fly Ash (kg in a m^3 mixture)
  - ▪ Water (kg in a m^3 mixture)
  - ▪ Superplasticizer (kg in a m^3 mixture)
  - ▪ Coarse Aggregate (kg in a m^3 mixture)
  - ▪ Fine Aggregate (kg in a m^3 mixture)
  - ▪ Age (days)
- o Dependent Variable:
  - ▪ Concrete Compressive Strength (MPa)

3. Experimental Procedure

• Mix Design Preparation: Prepare different concrete mixtures by varying the amounts of Cement, Blast Furnace Slag, Fly Ash, Water, Superplasticizer, Coarse Aggregate, and Fine Aggregate.

• Curing Process: Allow the concrete mixtures to cure over specified time periods to measure the effect of age on compressive strength.

4. Data Analysis

Statistical analysis will be performed using R

Descriptive Analysis

Initially, a descriptive analysis will be performed to understand the basic characteristics of the data. This includes:

- Summary statistics (mean, median, mode, standard deviation, etc.)
- Distribution plots for each variable (histograms, box plots, etc.)
- Correlation matrix to identify relationships between variables

Univariate Analysis

Each variable will be analyzed individually to understand its distribution and basic properties. This includes:

- Plotting histograms and density plots
- Calculating measures of central tendency and dispersion
- Identifying any outliers or anomalies

Multiple Linear Regression Analysis

Full Model

A multiple linear regression model will be developed with Concrete Compressive Strength as the response variable and all other variables as predictors. The steps include:

- Fitting the full model
- Checking the overall model fit (R^2, Adjusted R^2, etc.)
- Analyzing the significance of individual predictors using p-values

Residual Analysis

- Examining residual plots to check for homoscedasticity
- Analyzing the normality of residuals using Q-Q plots
- Identifying any patterns or trends in the residuals

Variable Selection

To improve the model, variable selection techniques will be employed. This includes:

- Stepwise regression (both forward and backward selection)
- Comparing the models using criteria such as AIC, BIC, and Adjusted R^2

Model Comparison and Selection

Models with different subsets of variables will be compared to select the one with the highest predictive accuracy. This involves:

- Calculating prediction errors (RMSE, MAE, etc.)
- Performing cross-validation to evaluate model stability
- Selecting the best model based on predictive performance and interpretability

Residual Analysis of the Final Model

A detailed residual analysis will be conducted for the selected model to ensure it meets all the assumptions of linear regression:

- Checking for homoscedasticity, normality, and independence of residuals
- Identifying points with high leverage or influence using Cook's distance

Model Validation

The final model will be validated using an independent dataset (if available) or by splitting the original dataset into training and testing sets. This includes:

- Evaluating the model's performance on the testing set
- Comparing the predictions with actual values to assess accuracy

5. Limitations

- The study is limited to the available data on concrete mix proportions and compressive strength.
- The generalization of findings may be restricted due to variations in concrete production processes and environmental conditions.

6. Implications for Construction Quality and Sustainability

Finally, the results will be interpreted in the context of construction quality and sustainability. This includes:

- Discussing the impact of each significant predictor on concrete compressive strength
- Providing recommendations for optimizing concrete mixtures for better quality and sustainability
- Highlighting any limitations of the study and suggesting areas for future research

04. Results

This is the **Descriptive analysis** of the whole dataset

Variable	Mean	Median	Max	Min	1st Quantile	3rd Quantile	St. Deviation
Cement	281.2	272.9	540.0	102.0	192.4	350.0	104.51
Blast furnace slag	73.9	22.0	359.4	0.0	0.0	142.9	86.28
Fly ash	54.19	0.0	200.1	0.0	0.0	118.3	63.99
Water	181.6	185.0	247.0	121.8	164.9	192.0	21.35
Superplasticizer	6.21	6.4	32.2	0.0	0.0	10.2	5.97
Coarse aggregate	972.9	968.0	1145.0	801.0	932.0	1029.4	77.75
Fine aggregate	773.6	779.5	992.6	594.0	731.0	824.0	80.18
Age (day)	45.6	28.0	365.0	1.0	7.0	56.0	63.17
Concrete compressive strength	35.82	34.45	82.6	2.33	23.71	46.13	16.71

Table 01

To improve the accuracy of the data set we **remove the outliers** using a function

```
####Outliers


# create detect outlier function
detect_outlier <- function(x) {

  # calculate first quantile
  Quantile1 <- quantile(x, probs=.25)

  # calculate third quantile
  Quantile3 <- quantile(x, probs=.75)

  # calculate interquartile range
  IQR = Quantile3 - Quantile1

  # return true or false
  x > Quantile3 + (IQR * 1.5) | x < Quantile1 - (IQR * 1.5)
}

# create remove outlier function
remove_outlier <- function(dataframe, columns = names(dataframe)) {

  # for loop to traverse in columns vector
  for (col in columns) {

    # remove observation if it satisfies outlier function
    dataframe <- dataframe[!detect_outlier(dataframe[[col]]), ]
  }

  print("Remove outliers")
  print(dataframe)
}
```

Figure 01

Also we consider about a **pair scatterplot matrix** for the cleaned data set

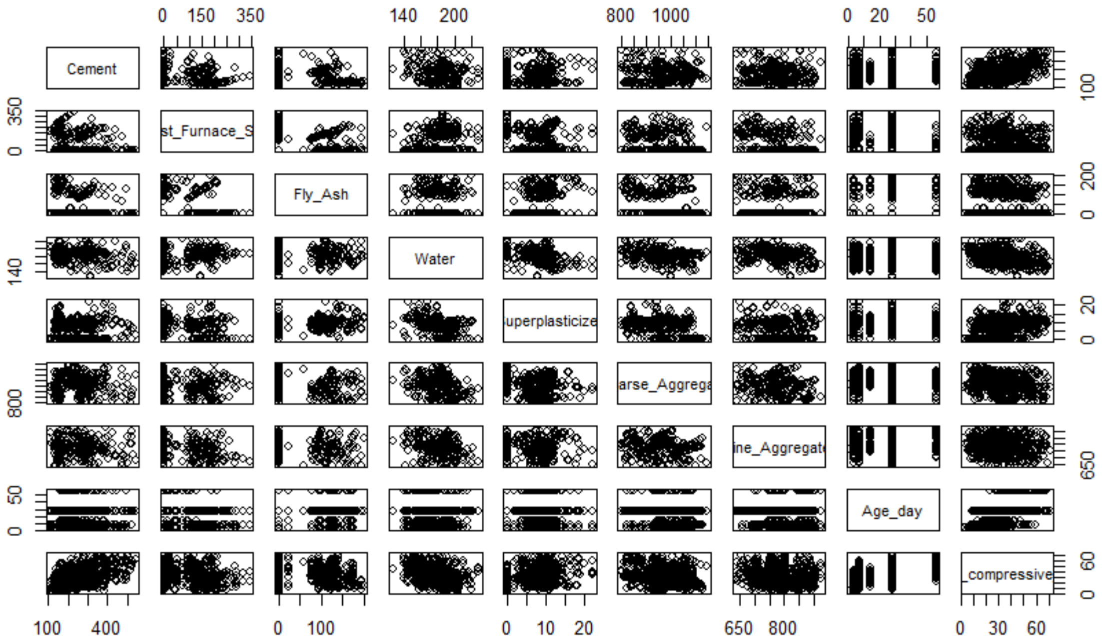

Figure 02

*.Strong Positive Influences: Cement content and curing time (Age) are key drivers of compressive strength, showing strong positive correlations.

*.Negative Influences: Water content negatively impacts compressive strength, indicating that controlling water levels is crucial for achieving higher strength.

*.Complex Influences: The roles of Fly Ash and Blast Furnace Slag are more complex, suggesting that their contributions to strength development depend on other factors such as curing time and mix proportions.

This matrix provides a foundation for further analysis, potentially using statistical techniques like regression or machine learning to better understand the interactions between these variables and to optimize concrete mix designs for strength and sustainability

Then we get **ANOVA results for full model**

```
Analysis of Variance Table

Response: Concrete_compressive_strength
                    Df Sum Sq Mean Sq   F value    Pr(>F)    
Cement               1  38081   38081  805.3194 < 2.2e-16 ***
Blast_Furnace_Slag   1  12941   12941  273.6626 < 2.2e-16 ***
Fly_Ash              1  21051   21051  445.1727 < 2.2e-16 ***
Water                1  11025   11025  233.1594 < 2.2e-16 ***
Superplasticizer     1   1699    1699   35.9398 3.191e-09 ***
Coarse_Aggregate     1      2       2    0.0448    0.8324    
Fine_Aggregate       1     31      31    0.6569    0.4179    
Age_day              1  47859   47859 1012.0896 < 2.2e-16 ***
Residuals          734  34709      47                        
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 03

*.Cement, Blast Furnace Slag, Fly Ash, Water, Superplasticizers, and Age are all statistically significant contributors to compressive strength, with Age and Cement having the most substantial effects.

*.Coarse and Fine Aggregates do not have a statistically significant impact on compressive strength, as indicated by their low F values and high p-values.

This ANOVA analysis shows that the mix components related to binders (Cement, Fly Ash, Blast Furnace Slag) and curing time (Age) play key roles in determining compressive strength, while aggregates do not significantly affect the outcome in this dataset.

After checking the significance to find the better model used the **forward selection method**

```
Subset selection object
Call: regsubsets.formula(Concrete_compressive_strength ~ ., data = cleaned_data, 
    nvmax = 8, method = "forward")
8 Variables  (and intercept)
                   Forced in Forced out
Cement                 FALSE      FALSE
Blast_Furnace_Slag     FALSE      FALSE
Fly_Ash                FALSE      FALSE
Water                  FALSE      FALSE
Superplasticizer       FALSE      FALSE
Coarse_Aggregate       FALSE      FALSE
Fine_Aggregate         FALSE      FALSE
Age_day                FALSE      FALSE
1 subsets of each size up to 8
Selection Algorithm: forward
         Cement Blast_Furnace_Slag Fly_Ash Water Superplasticizer Coarse_Aggregate Fine_Aggregate Age_day
1  ( 1 ) " "    " "                " "     " "   " "              " "              " "            "*"    
2  ( 1 ) "*"    " "                " "     " "   " "              " "              " "            "*"    
3  ( 1 ) "*"    "*"                " "     " "   " "              " "              " "            "*"    
4  ( 1 ) "*"    "*"                " "     "*"   " "              " "              " "            "*"    
5  ( 1 ) "*"    "*"                "*"     "*"   " "              " "              " "            "*"    
6  ( 1 ) "*"    "*"                "*"     "*"   "*"              " "              " "            "*"    
7  ( 1 ) "*"    "*"                "*"     "*"   "*"              " "              "*"            "*"    
8  ( 1 ) "*"    "*"                "*"     "*"   "*"              "*"              "*"            "*"    
```

Figure 04

Here the criterian values (Adj R^2, Cp, BIC, RSS) in graphically

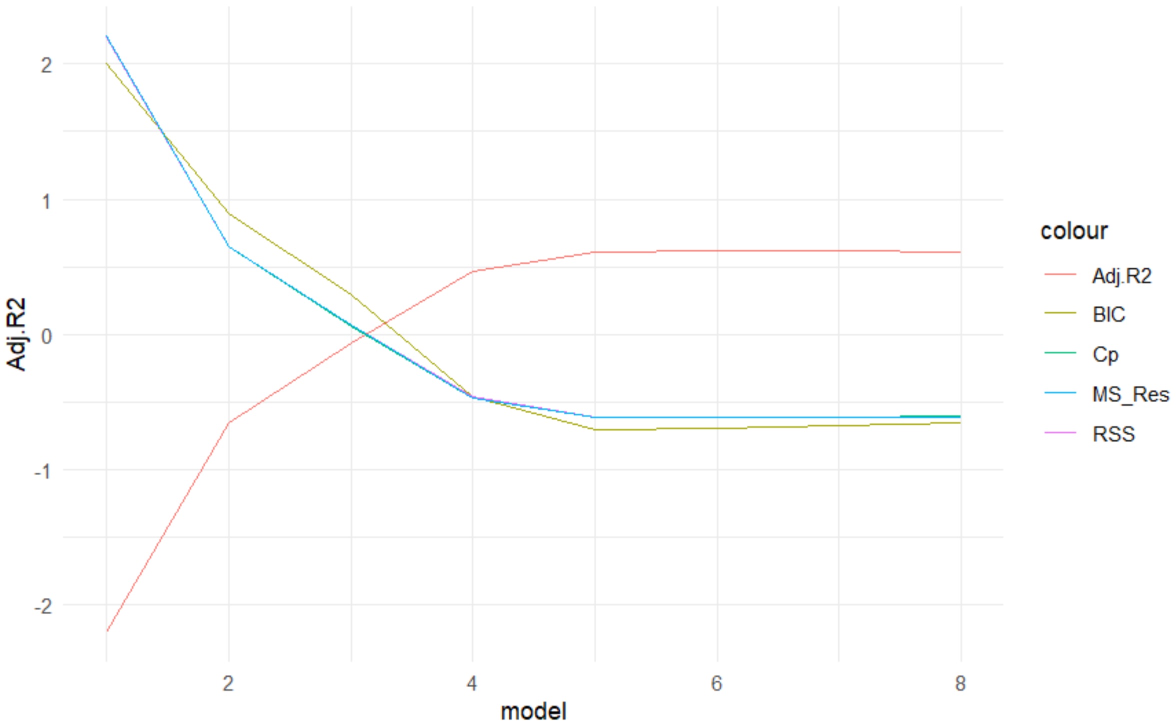

Figure 05

*. All the metrics show a rapid improvement (increased Adjusted R^2, decreased BIC, Cp, MS Residual, and RSS) up to Model 5 or 6, after which additional predictors add little to no improvement.

*. Based on these metrics, Model 5 seems to be the optimal model. It balances model complexity (as indicated by BIC and Cp) while still providing a good fit (low MS Residual and RSS, high Adjusted R^2).

Therefore we can suggest the 5th model is the better one based on this figure 05. So, Cement, Blast Furnace Slag, Fly Ash, Water, and Age are contributors to compressive strength.

Then Check the **Leverage points** in the better model and remove them for more accurate this dataset.

```
##Check the leverage points


leverage<-hatvalues(model_1)
high_leverage<-which(leverage>2*(6+1)/nrow(cleaned_data))
high_leverage

cleaned_dataset <- cleaned_data[-high_leverage, ]



nrow(cleaned_dataset)

```
```

```
  74  91 114 142 143 147 150 378 380 381 446 450 500 501 502 503 504 505 506 507 564 572 584 585 828 878 891 909 924 934 941 954
  15  28  46  66  67  68  70 224 225 226 283 287 324 325 326 327 328 329 330 331 376 384 396 397 550 597 610 628 643 651 655 668
 972
 686
 [1] 710
```

Figure 06

Better Model's Summary become,

```
Call:
lm(formula = Concrete_compressive_strength ~ Cement +
Blast_Furnace_Slag +
    Fly_Ash + Water + Age_day, data = cleaned_dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-20.1263  -4.0706  -0.2408   3.7504  22.9424

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        20.769914   3.615727   5.744 1.36e-08 ***
Cement              0.108082   0.003508  30.813  < 2e-16 ***
Blast_Furnace_Slag  0.080753   0.003657  22.082  < 2e-16 ***
Fly_Ash             0.050592   0.005417   9.340  < 2e-16 ***
Water              -0.218714   0.016234 -13.473  < 2e-16 ***
Age_day             0.552769   0.017342  31.875  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.87 on 727 degrees of freedom
Multiple R-squared:  0.7912,	Adjusted R-squared:  0.7897
F-statistic: 550.8 on 5 and 727 DF,  p-value: < 2.2e-16
```

Figure 07
```

After construct the better model & consider the model diagonastics.

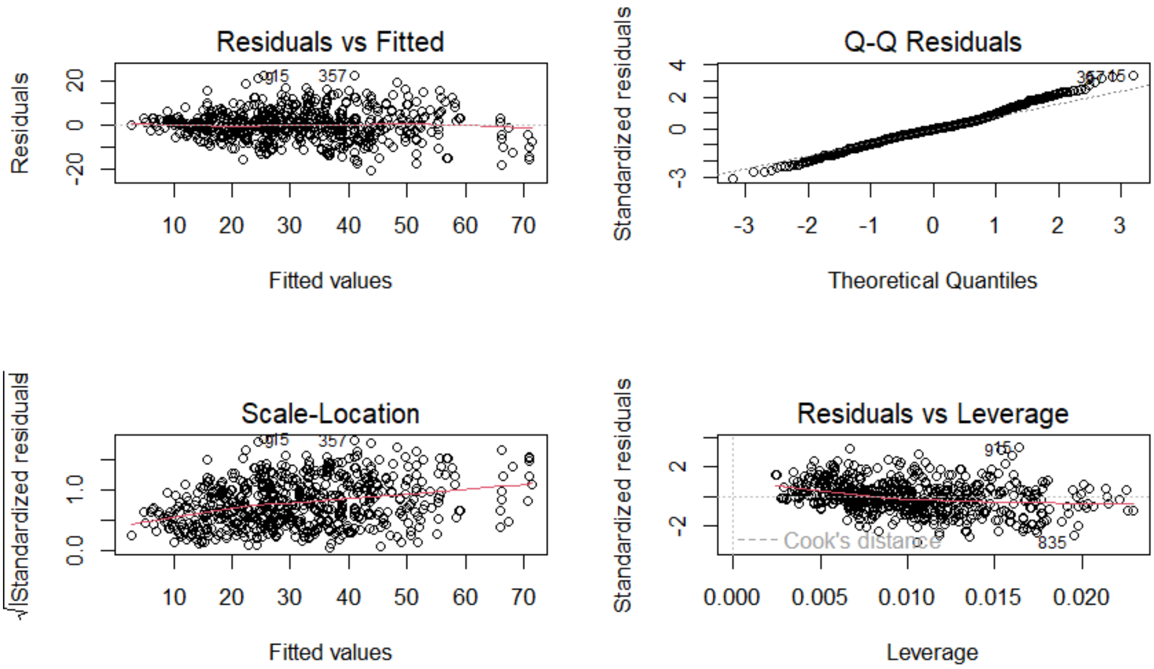

Figure 08

Residuals vs Fitted

The residuals appear to be scattered randomly around 0, which suggests that the relationship is likely linear. However, there seems to be some minor clustering, indicating potential non-constant variance (heteroscedasticity), though it's not severe.

Q-Q Residuals

The points mostly follow the straight line, but there are slight deviations at the tails (especially at higher quantiles). This suggests that the residuals are approximately normal, with potential mild issues at the extremes.

Scale Location

The residuals seem fairly scattered, but there is a slight upward trend in the spread of residuals, which indicates potential heteroscedasticity (variance of residuals increasing with fitted values).

Residuals vs Leverage

Most points are clustered near the bottom, indicating that no data points exert an undue amount of leverage on the model. However, some points like 915 and 835 might be influential as they are close to the Cook's distance line.

05. Discussion

The findings of this study provide crucial insights into the correlations between various concrete mix components and compressive strength, which have a direct impact on building quality and sustainability. Key findings suggest that the age of the concrete and its cement content are the key determinants of compressive strength, with substantial positive connections. Water content, on the other hand, has a detrimental impact on compressive strength, highlighting the need of controlling water proportions to improve the structural quality of concrete.

Supplementary materials like as fly ash and blast furnace slag play complex but valuable functions in promoting both sustainability and strength growth under certain situations. While their individual influence is less obvious than cement and curing time, they provide environmental benefits by lowering reliance on traditional materials while maintaining concrete performance.

Notably, neither coarse nor fine aggregates have a statistically significant effect on compressive strength in the dataset investigated. This observation challenges common thinking in the sector, where aggregates are frequently considered to give structural bulk. However, it emphasizes the importance of binders and curing methods in maximizing concrete strength.

The results of the ANOVA analysis highlight the significance of cement, fly ash, blast furnace slag, and curing time as key variables in affecting compression strength. The best model produced using stepwise regression, which includes these factors, strikes a balance between model complexity and predictive accuracy, implying potential uses in mix design.

06. Conclusion

This study successfully identifies the key elements impacting concrete compressive strength, highlighting the need of regulating cement amount, curing time, and additional materials such as fly ash and blast furnace slag for long-term construction practices. The findings show that cement and curing age are the most important factors in strength, with water content requiring careful control to minimize detrimental consequences.

From a sustainability standpoint, using Fly Ash and Blast Furnace Slag into concrete mixes provides a potential approach to minimize environmental impacts while maintaining strength. These materials improve long-term durability and offer an environmentally responsible alternative to typical concrete components.

07. Individual Contribution

Task Completed	PS/2020/018	PS/2020/110	PS/2020/130	PS/2020/132	PS/2020/135	PS/2020/238	PS/2020/302	PS/2020/318
Data Collecting & Cleaning		■	■			■		
Introduction with problem statement						■		
Descriptive Analysis	■							■
Regression Analysis with residual Analysis				■	■			
Results, Discussion & conclusions		■					■	
Accuracy checking					■			■
Progress reports creating			■				■	
Final Report Creating	■			■				