# Data Science Project Scoping Guide

1. Goal

Using Python, MATLAB, Java or some things else to analysis data and then visualize them in-group (5 students in a group).

2. Data set

The attendance table for each class, totally 12. Each of tables is in excel format. The overall goals are to give remarks for every student according to the attendance rec ord.

3. Manual Text

1. Step 1 Integrate the 12 tables according to yours names used in our class, which will lead to a 28\*12 matrix

2. Step 2. Clean the data set, for example, deleting the repetitive record or filling the missing values

3. Step 3 (Optimal)To do some EDA for your data set e.g., mean variance, store your data in a data base e.g., sqlite3

4. Step 4. Clustering (k means) you’re the rows (your performance) into 4 groups (A, B, C, D)

5. Step5. Visualize your result.

Group

|  |  |  |  |
| --- | --- | --- | --- |
| 学号 | 中文名 | 英文名 | group |
| IS17170201 | 王辰 | AL-AWADHI HUSAM TAHA MOHAMMED | 1 |
| IS17170210 | 卫斯 | FAROOQ AWAIS | 1 |
| IS17170223 | 可力得 | MASOOD MUHAMMAD QASIM | 1 |
| IS17170224 | 齐飞 | QURESHI UMAR | 1 |
| IS17170225 | 王子 | UMAR MUHAMMAD | 1 |
| IS17170226 | 阿力 | ALI ASIF | 2 |
| IS17170203 | 赵雷 | AHSAN MUHAMMAD | 2 |
| IS17170215 | 麦家 | KAZMI SYED REHMAN RAZA | 2 |
| IS17170216 | 刘佳令 | NOOR NUKHBA | 2 |
| IS17170220 | 安思宁 | ASHIQ MAHER MUHAMMAD | 2 |
| IS17170204 |  | AJMAL HAMZA | 3 |
| IS17170208 | 张磊 | BISWAS DIPALOKE | 3 |
| IS17170209 | 李斯诺 | CHOWDHURY SATYEN ROY | 3 |
| IS17170212 | 萨娜 | GOHAR SANA | 3 |
| IS17170228 | 马龙 | SAYFIDDINOV JAVLON | 3 |
| IS17170219 | 吴思远 | ALI TAIMUR | 4 |
| IS17170202 |  | AL-KATEB ELYAS KHALED MOHAMMED | 4 |
| IS17170205 | 诺曼 | ARSHAD NOMAN | 4 |
| IS17170218 | 步豹 | ABU BAKAR | 4 |
| IS17170227 | 夏米大 | SHAHID ASFANDYAR | 4 |
| IS17170206 | 吴越寺 | ASHFAQ HAMZA | 5 |
| IS17170217 | 李飞 | RAHMAN MOHAMMAD HABIBUR | 5 |
| IS17170213 |  | HASSAN SYED MUBASHAR | 5 |
| IS17170214 | 马云 | IFTIKHAR MUHAMMAD TALHA | 5 |
| IS17170222 | 沈易 | KHAN SHARJEEL | 5 |