Data Science Group Project Proposal (Group 10)

1. Proposed datasets to be used.

Data set	Link	Reason
Crime data 2010-2012	http://www.data.gov.lk/dataset/crime-data-2010-2012	Public importance District categorization Year-wise data
Employment by Industrial Sector 2012	http://www.data.gov.lk/dataset/employees-industrial-sector-2012	District categorization Female and Male classification 3 industrial sectors
GDB distribution by Province	http://www.data.gov.lk/dataset/province-wise-gross-domestic-production-2006-2012	Income of people related to crimes.
Accommodation information for tourists	http://www.data.gov.lk/dataset/accommodation-information-tourists	Crime can relate to the tourism industry

2. Group members and their index numbers

Name	Index Number
Prageeth Anjula	199304P
Waruna Wickramasingha	199371P
Hesitha Wijayasinghe	199372U
Pumudinee Kumarasiri	199338X

3. Git Repo Link: https://github.com/PraAnj/gov.lk.data.analysis

4. Research question to be explored or hypothesis to be validated

   ● We are trying to analyze the relationship between crimes happening in various parts of the country and the industrial sector distribution of male and female.
   ● Government and community can make use of this analysis to understand the reason for different crimes and make preventive actions to mitigate possible types of crimes
   ● Identification of correlation in each district between types of crime and industry sector distribution is part of this work.
   ● Predict crime rates of each district with changes to each industrial sector/ GDP or tourism will be of interest.