|  |
| --- |
| Description: Description: RMIT logo 35mm col |
| **Title of Report**  **Project Proposal**  By  Rinku Bajaj (S3672522) Arion Evans  Deepika Joshi |

# Introduction

The main aim of this project is to identify the most influential factors affecting life expectancy in developed and developing countries from 2000-2015. The project considers factors traditionally studied for life expectancy predictions such as demographic variables, mortality rate and income composition as well as the effects of new factors such as immunization and human development index.

The data is sourced from Kaggle Repository and is a combination of life-expectancy and health data from Global Health Observatory (GHO) published by World Health Organisation (WHO) and economic data published by United Nations. It consists of data spanning over 15 years from 2000-2015 for 193 countries and can be broadly divided into Economic Factors, Social Factors, Immunization Factors and Mortality Factors. The dataset has 2938 observations for 20 predictor variables describing the life expectancy of each nation for that year.

# Proposed Data Analysis

The project aims to accurately predict the life – expectancy of nations over the years considering a holistic range of factors. Initially the research on life expectancy only involved mortality rates and demographic variables; however, it was observed that in the last 15 years the human mortality rates have declined greatly in developed nations due to development in the health care sector. Hence, we aim to study the effects of factors such as immunization and human development index on life expectancy.

Additionally, the following key questions will be studied –

1. Did the traditionally chosen predictors really affect Life Expectancy?
2. Which predictor variables are actually responsible for Life Expectancy?
3. What is the impact of increasing healthcare expenditure on a country with lower Life Expectancy?
4. How does Alcohol Consumption affect Life Expectancy?
5. What is the impact of Immunization Coverage on Life Expectancy?

# Proposed Methodology

The proposed methodology is to build multiple linear regression models for accurately predicting the Life Expectancy of over 193 nations from 2000-2015. Preliminary analysis will be conducted to make sure the models follow the regression assumptions such as homoskedasticity, multivariate normality and multicollinearity. The regression hypothesis will be tested for different regression models such as backward elimination, forward selection and stepwise regression to identify the most influential predictors. These results will then be used to answer key research questions identified in this research proposal.

Lastly, analyses will be conducted to study the relationships between the different predictor variables and their effect on Life Expectancy of the nation.