

Regression Project - Group 6

Team Members

Ahuja, Kunaal
Sarat, Parvathy
Sweet, Will

Dataset and Source

We're looking to explore the mortality dataset published by CDC (Centers for Disease Control and Prevention)

The data source for this dataset is Kaggle.

<https://www.kaggle.com/cdc/mortality>

This mortality dataset is a record of every death in the United States for 2005 through 2015, including detailed information about causes of death and the demographic background of the deceased.

It contains 77 quantitative and qualitative features.

There has been some research on the suicide cases, Gun v/s Vehicle deaths, season mortality, etc based on this dataset.

Questions

Cause of death predicted by different demographics

We would control different features such as age, gender, etc and run the multivariate regression model.

Tasks

Data Collection and stitching

Data Cleaning and Preparation

- Handling Nulls and missing values

- Handling Qualitative variables

Build Initial Model

Check for the Model assumption

Model Performance and accuracy

Hypothesis testing

Inference and Analysis Report