# Medical Appointment - No Show Analysis

Capstone Project

# Executive Summary

## Data Set

- The data set I used to analyze and predict the no shows for medical appointments came from the Kaggle website
- The data was collected in Brazil and was a large enough sample to study this issue.

## Data Wrangling

There were a few changes that needed to be made to the dataset. Below is a list of the high-level transformations

- Take absolute value to change handful of negative values for age to positive
- Converting variables with binary values to factors to help analyze them
- Checking for outliers

# Executive Summary

## Exploratory Data Analysis

Through exploratory analysis, understanding the effects of the different variables on the dependent variable, Status. The following variables were analyzed

- No Shows by Day of week
- No Shows by Gender
- No Shows by Age
- No Shows by SMS Reminder
- Effect of Alcoholism on No Show Status
- Effect of Government Assistance on No Show Status

## Data Mining

- Split data set into test and train sections.
- Logistic Regression as the algorithm
- Analyse the Results using the confusion matrix at different threshold levels
- Understand False and True Positives and False and True Negatives
- Look for ways to improve the accuracy of the model

## Executive Summary - Conclusion & Further Analysis

Based on my analysis of the data, it looks like the three factors that seem to influence No-shows are ***Age, Alcoholism and Government assistance***. The younger the patient, the more likely they are to skip their appointment. And patient struggling with Alcoholism and that are on government subsidies tend to be more likely to be no-shows for their appointments.

Further analysis is needed to understand how the different variables in the data set might help us better understand the no show issue. And can help us reduce the number of no shows in the healthcare space.