Insurance Fraud

# The Problem

Fraud costs insurance companies billions of dollars every year. The senior leadership at our company wants to better understand fraud, its impact on our company and how we can better control it using advanced analytics.

# The Data

IT and the business have provided the data from four different databases: Customers, Policy Details, Claims and Fraud (each a tab in the attached Excel spreadsheet). These datasets can be linked together using the IDs in each database. IT has also provided the data dictionary to help you understand the data better.

# The Task

You met with the Chief Data Officer and Data Scientist who agreed that the first step should be to identify some of the predictors of fraud. Since Life insurance and Disability insurance are two distinct fraud problems, we have decided that focusing on our disability policies would be best for now, so please make sure you exclude Life insurance policies in your analysis.

Using the data provided to you, analyze the data to determine what variables are correlated with fraud, if any. Build models using the existing data to predict fraud in the future.