This document gives brief overview of the business problem and the programming approach to solve the same. The aim here is to predict the likelihood of an employee turnover and the potential pay-off in case of proactive retention strategy.

The dataset here consists of a rich combination of various employee related data points like percent hike, tenure, age, performance rating etc. The emphasis is to dig out any hidden reasons like distance to office, age difference of employee and manager, overtime etc. which are the potential factors influencing the attrition of employees.

We perform an extensive exploratory analysis to investigate any hidden relationship and consider various nuances like correlation, information strength before running a multivariable regression model. The effective usage of dplyr, lubridate, ggplot2, caret, car package yields a great predictive model which exhibits an accuracy of about 94% on unseen test data.

After evaluating the risk of leaving for individual active employee, we also simulate a model for gauging a probable payoff if we wish to retain set of employees from lower hike range, by offering higher average hike and potentially save on turnover cost like on-boarding, knowledge transfer, training etc. A well-informed & judicious choice of these two control parameters can possibly prove out to be an effective medium of achieving better rate of return on the investment.