

# Forecasting of Staffing Needs

April, 2019

Contributors: Luo (Iris) Yang, Marcelle Chiriboga, Patrick Tung, Weifeng (Davy) Guo

Mentor: Rodolfo Lourenzutti

## Executive Summary

The purpose of the project was to help the People Analytics and Innovation Team from Providence Health Care (PHC) to predict the short-term staff needs in order to prepare for potential cost rising and staff shortage.

## Introduction

Providence Health Care is a government agency that operates 16 healthcare facilities in British Columbia. For most positions in the healthcare business, any staffing leaves must always be filled in by another person. More than 70% of the operation costs in health care is for staffing, and the substitute caused by short notice leave usually costs more than regular staff. Hence, predicting the short-term staffing needs could help PHC to prepare for potential leave with short notice. Suggested by our partner, we will focus our target on the *operation level*, which has the time horizon of that less than a month. *Operational* means that we must answer the question “how many back up staff do we need on a daily basis to maintain 100% staffing?”

Based on the data provided by Providence Health, our identified question of this project is “How many exceptions will happen for the next 30 days and what are the types of exceptions?”. We plan to explore the data first to identify potential features and labels. Then, using the sample data, we plan to build and train the model to predict and test for accuracy. Once we have a model prototype, we will bring new details and update the data and the model until we have reached a certain level.

The final product will consist of three components:

- a dashboard (developed in R Shiny or Tableau);
- the scripts containing the code used to proceed with the analysis
- a report outlining the methodologies and findings.

## Data Science Techniques

The dataset consists of more than 2 millions records of exceptions since 2012, and we intend to split the raw data based on years. This way, not only will we have a smaller dataset to generate some insight from, but we will also be able to tell the difference caused by time (facility opening, system develop, etc).

We will start from simple question with simple data and move forward once we achieve certain goal. The milestones for our problems are:

Step 1: How many exceptions will happen next month?

Step 2: For those exceptions predicted, what kind of exception group they are?

Step 3: For those exception predicted, what kind of position they are?

We intend to approach the problem with 3 methods, applying some/all of them based on further study and process. The following approaches are:

- **Time Series:** We assume for every nature year, there is a pattern for exception occurrences. We want to explore the trend throughout the years, both as a whole and for separate groups, to make predictions as a base. This will provide a general idea based on historical data.
- **Linear Regression:** We assume that the partner would like to know the exception situation for next month based on a certain period of data (for example, one week from today using 1 year's data before today). In this case, we will generate a linear regression to see the relationship among certain predictors, such as the information of exception records for the next month that are already created.
- **Neural Network:** Because our problem is to forecast time series data, neural networks may potentially be very helpful. We have seen that neural networks can be used in many applications in the world for problems ranging from driverless cars, to image recognition, and even for forecasting. The applications of neural networks as machine learning models has been dramatically increasing. Especially for time series data, perhaps using LSTMs can allow use to learn the history of a sequence of data to “effectively” predict sequence of future data.

### Timeline and Evaluation

We have the following time schedule based on group discussion, and the actual process may be updated due to the reality situation:

Time Period	Milestone
Week 1	Review documentation, and finalize the proposal reports to mentor and partner
Week 2	Data wrangling, feature selection, EDA and implement baseline model
Weeks 3 - 4	Explore different approaches to fit the models
Week 5	Build algorithms, testing, adjusting
Week 6	Improve the dashboard, wrapping up
Week 7	Presentations and reports