Forecasting of Staffing Needs in Health Care

Team Members:

- Luo (Iris) Yang
- Marcelle Chiriboga
- Patrick Tung
- Weifeng (Davy) Guo

Agenda

- Introduction
- The Analysis
  - Predicted Number of Exceptions
  - Predicted Number of Urgent Exception Groups
  - Exceptions Classification
- The Dashboard

Introduction

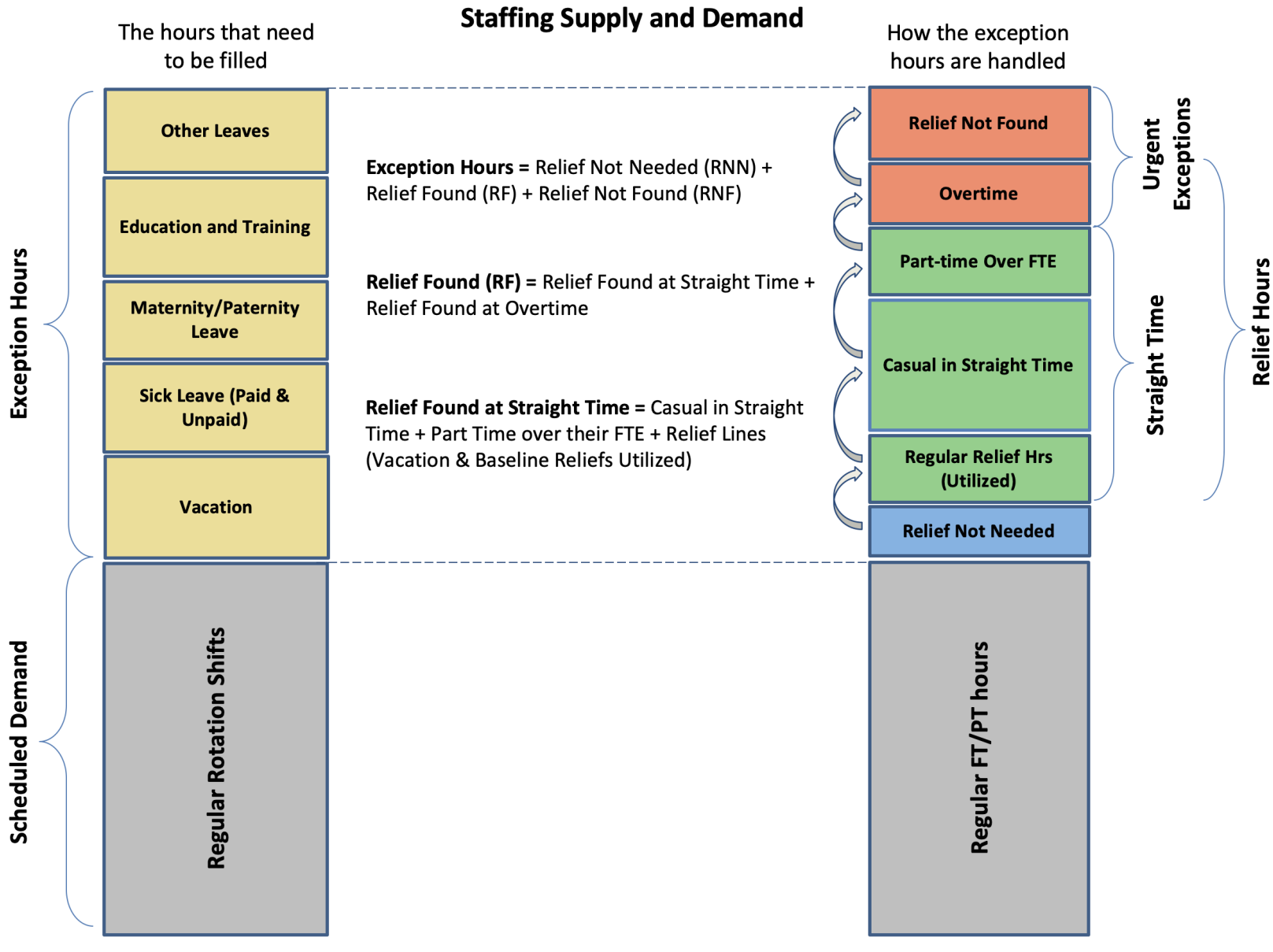
The Partner - Providence Health Care

- Providence Health Care (PHC) is a non-profit organization.
- Almost 9,000 people working at their 16 facilities - 6,000 staff, 1,000 medical staff/physicians, 200 researchers, 1,600 volunteers.
- PHC is the provincial centre for the care of six groups of people with often-intensive health needs.

Providence
HEALTH CARE
How you want to be treated.

The Problem

- In the healthcare business, staff absences must always be backfilled.
- These absences, expected or not, are called **exceptions**.
- One way to minimize their impact is to predict future exceptions based on historical data.


Staffing Supply and Demand
The hours that need to be filled
How the exception hours are handled
Exception Hours
Other Leaves
Education and Training
Maternity/Paternity Leave
Sick Leave (Paid & Unpaid)
Vacation
Scheduled Demand
Regular Rotation Shifts
Exception Hours = Relief Not Needed (RNN) + Relief Found (RF) + Relief Not Found (RNF)
Relief Found (RF) = Relief Found at Straight Time + Relief Found at Overtime
Relief Found at Straight Time = Casual in Straight Time + Part Time over their FTE + Relief Lines (Vacation & Baseline Reliefs Utilized)
Relief Not Found
Overtime
Part-time Over FTE
Casual in Straight Time
Regular Relief Hrs (Utilized)
Relief Not Needed
Regular FT/PT hours
Urgent Exceptions
Straight Time
Relief Hours

Objective

The purpose of this project was to predict the short-term staff needs in order to provide PHC some insight into unexpected potential costs and staff shortages.

Specifically we focused on building models for:

- Forecasting staffing needs on a weekly basis, allowing PHC to estimate how many back up staff are needed;
- Forecasting how many exceptions will fall under the urgent exception groups (i.e. overtime and relief not found);
- Forecasting possible outcome for each exceptions submitted.

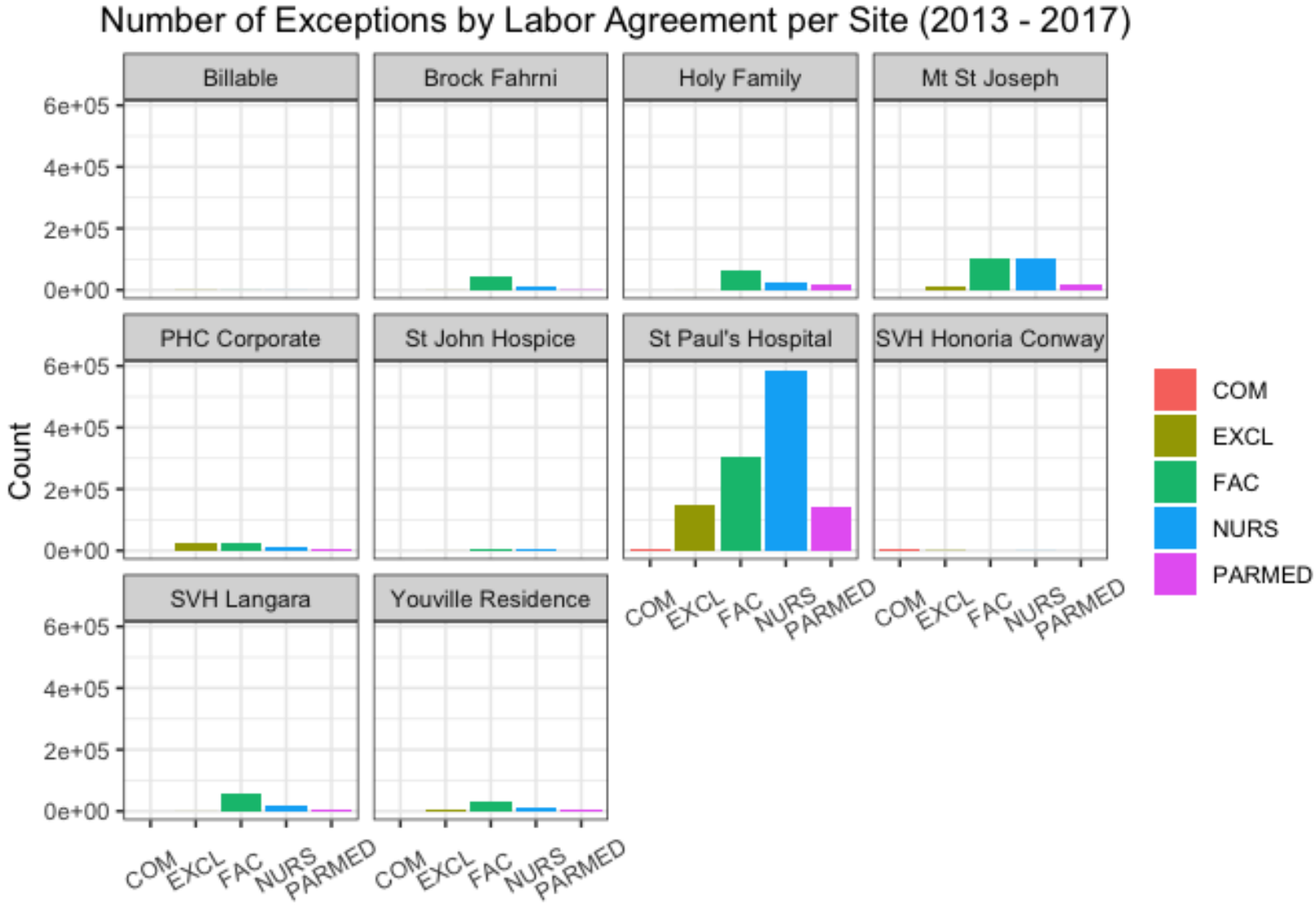
The Analysis

EDA

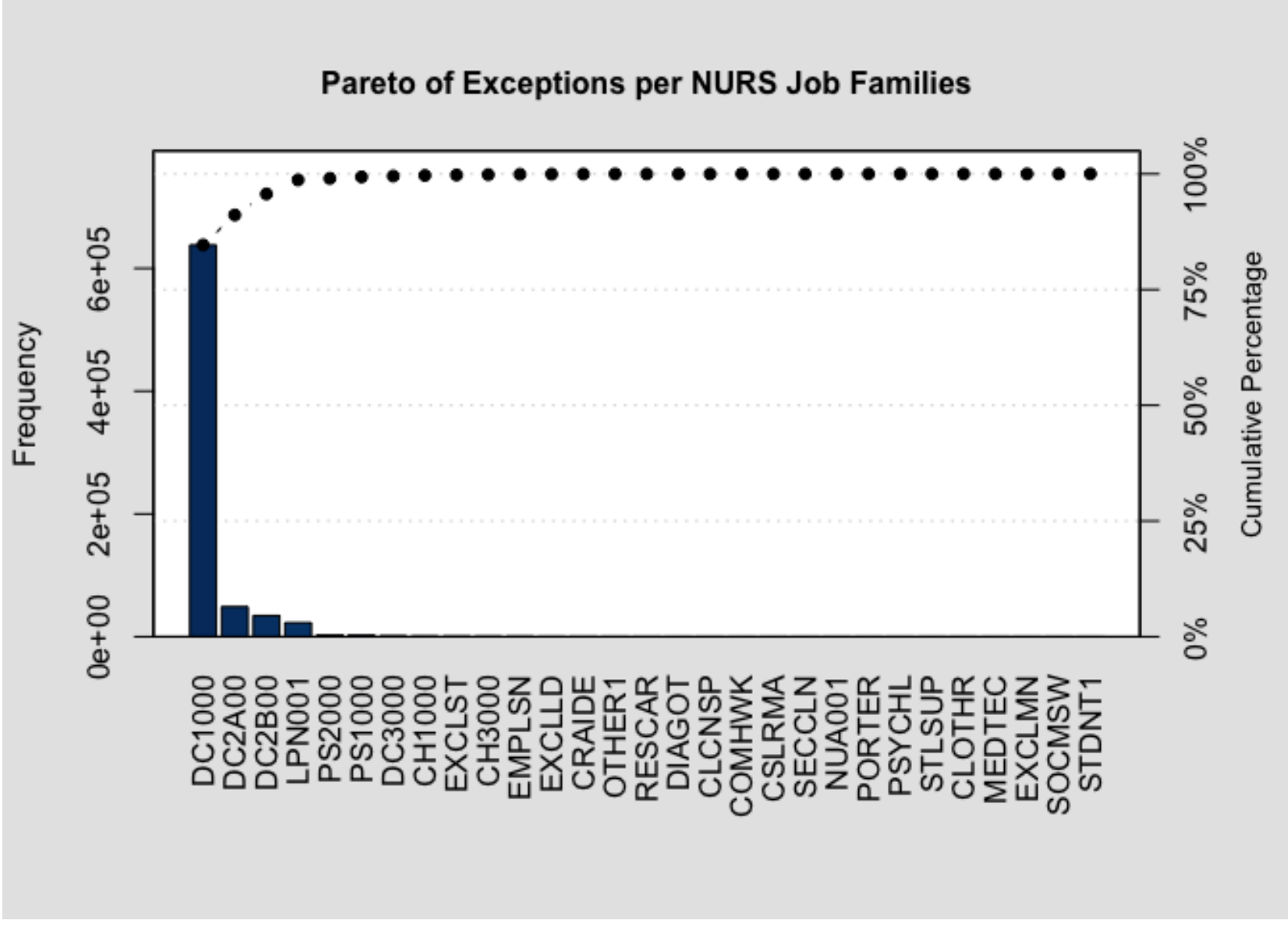
We performed an EDA to indentify the facilities, labour agreement and job families we should focus on.

```
## # A tibble: 10 x 2
##    SITE                count
##    <chr>               <int>
##  1 St Paul's Hospital 420961
##  2 Mt St Joseph        83590
##  3 Holy Family         37197
##  4 SVH Langara         29193
##  5 PHC Corporate       24002
##  6 Brock Fahrni        19530
##  7 Youville Residence  15678
##  8 SVH Honoria Conway   2799
##  9 St John Hospice      2154
## 10 Billable              555
```

EDA


Number of Exceptions by Labor Agreement per Site (2013 - 2017)
Billable
Brock Fahrni
Holy Family
Mt St Joseph
PHC Corporate
St John Hospice
St Paul's Hospital
SVH Honoria Conway
SVH Langara
Youville Residence
Count
6e+05
4e+05
2e+05
0e+00
COM
EXCL
FAC
NURS
PARMED

EDA


Pareto of Exceptions per NURS Job Families
Frequency
0e+00
2e+05
4e+05
6e+05
Cumulative Percentage
0%
25%
50%
75%
100%
DC1000
DC2A00
DC2B00
LPN001
PS2000
PS1000
DC3000
CH1000
EXCLST
CH3000
EMPLSN
EXCLLD
CRAIDE
OTHER1
RESCAR
DIAGOT
CLCNSP
COMHWK
CSLRMA
SECCLN
NUA001
PORTER
PSYCHL
STLSUP
CLOTHR
MEDTEC
EXCLMN
SOCMSW
STDNT1

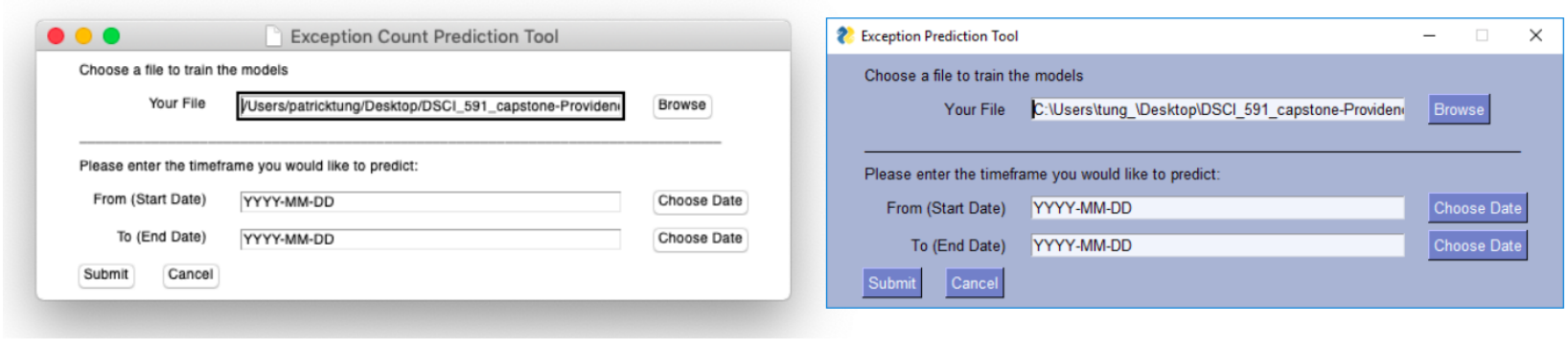
Exception Count Prediction

Forecasting the number of exceptions for Providence Health Care

Method

- Data
  - Training: 2013 - 2016
  - Validation: 2017
  - Testing: 2018
- Data Wrangling
  - Split data by SITE, JOB_FAMILY, and SUB_PROGRAM
  - e.g. St Paul's Hospital, Registered Nurse - DC1, Emergency
- Fit time series model for each “combination”
  - Facebook Prophet
- Predict the number of exceptions for the combinations
- Adjusted models based on Mean Absolute Error
- Output a .csv file containing the forecasts

Product/Interface



Output file

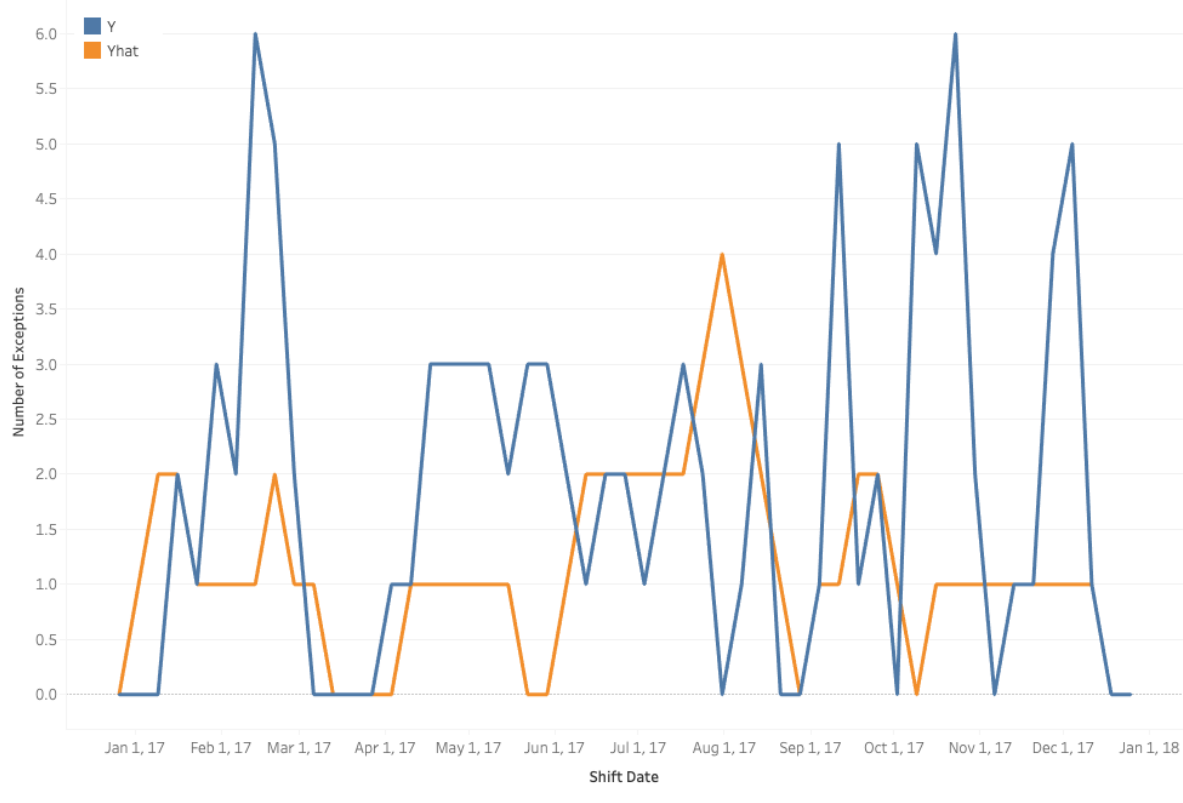
- `.csv` file containing all the predictions (on a weekly basis)

	A	B	C	D	E	F	G	H	I
1		yhat	yhat_lower	yhat_upper	ds	week	site	job_family	sub_program
2	0	8	0	19	12/31/18	1	St Paul's Hos	DC2B00	RENAL 6AB
3	1	6	0	18	12/31/18	1	St Paul's Hos	DC2B00	RENAL 6AB
4	2	5	0	17	12/31/18	1	St Paul's Hos	DC2B00	RENAL 6AB
5	3	5	0	16	12/31/18	1	St Paul's Hos	DC2B00	RENAL 6AB
6	4	5	0	16	12/31/18	1	St Paul's Hos	DC2B00	RENAL 6AB
7	5	5	0	16	12/31/18	1	St Paul's Hos	DC2B00	RENAL 6AB
8	6	4	0	16	12/31/18	1	St Paul's Hos	DC2B00	RENAL 6AB
9	7	4	0	15	1/7/19	2	St Paul's Hos	DC2B00	RENAL 6AB
10	8	4	0	16	1/7/19	2	St Paul's Hos	DC2B00	RENAL 6AB
11	9	5	0	16	1/7/19	2	St Paul's Hos	DC2B00	RENAL 6AB
12	10	7	0	18	1/7/19	2	St Paul's Hos	DC2B00	RENAL 6AB
13	11	7	0	19	1/7/19	2	St Paul's Hos	DC2B00	RENAL 6AB
14	12	7	0	19	1/7/19	2	St Paul's Hos	DC2B00	RENAL 6AB
15	13	7	0	19	1/7/19	2	St Paul's Hos	DC2B00	RENAL 6AB
16	14	7	0	18	1/14/19	3	St Paul's Hos	DC2B00	RENAL 6AB
17	15	7	0	18	1/14/19	3	St Paul's Hos	DC2B00	RENAL 6AB
18	16	7	0	18	1/14/19	3	St Paul's Hos	DC2B00	RENAL 6AB
19	17	6	0	17	1/14/19	3	St Paul's Hos	DC2B00	RENAL 6AB
20	18	5	0	17	1/14/19	3	St Paul's Hos	DC2B00	RENAL 6AB
21	19	5	0	16	1/14/19	3	St Paul's Hos	DC2B00	RENAL 6AB
22	20	5	0	17	1/14/19	3	St Paul's Hos	DC2B00	RENAL 6AB

Difficulties

- Certain combinations of data had very little exceptions
  - Little to no pattern
  - Predictions are not meaningful

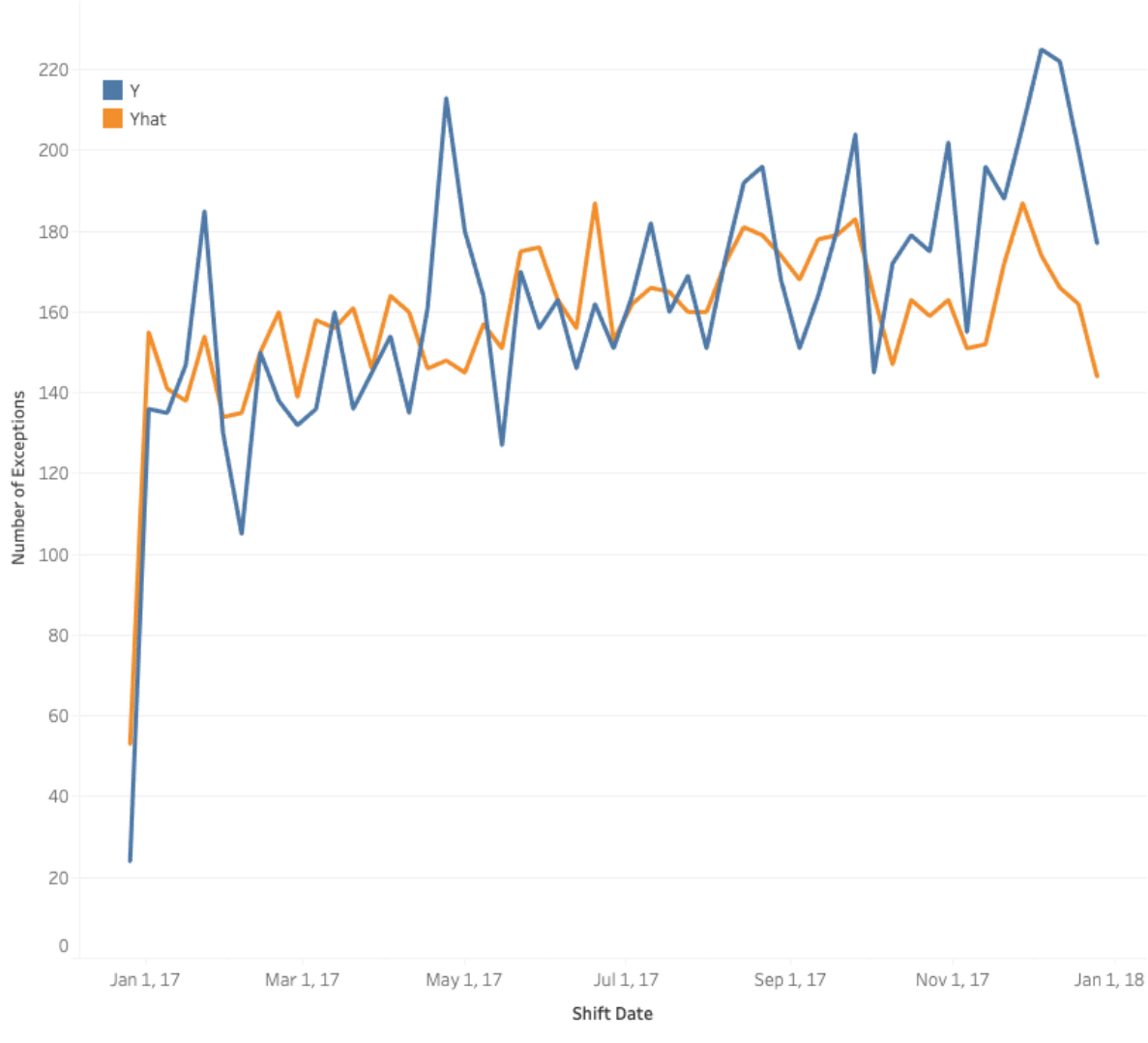
- e.g. Youville Residence, Registered Nurse - DC2B - Parkview



Solution

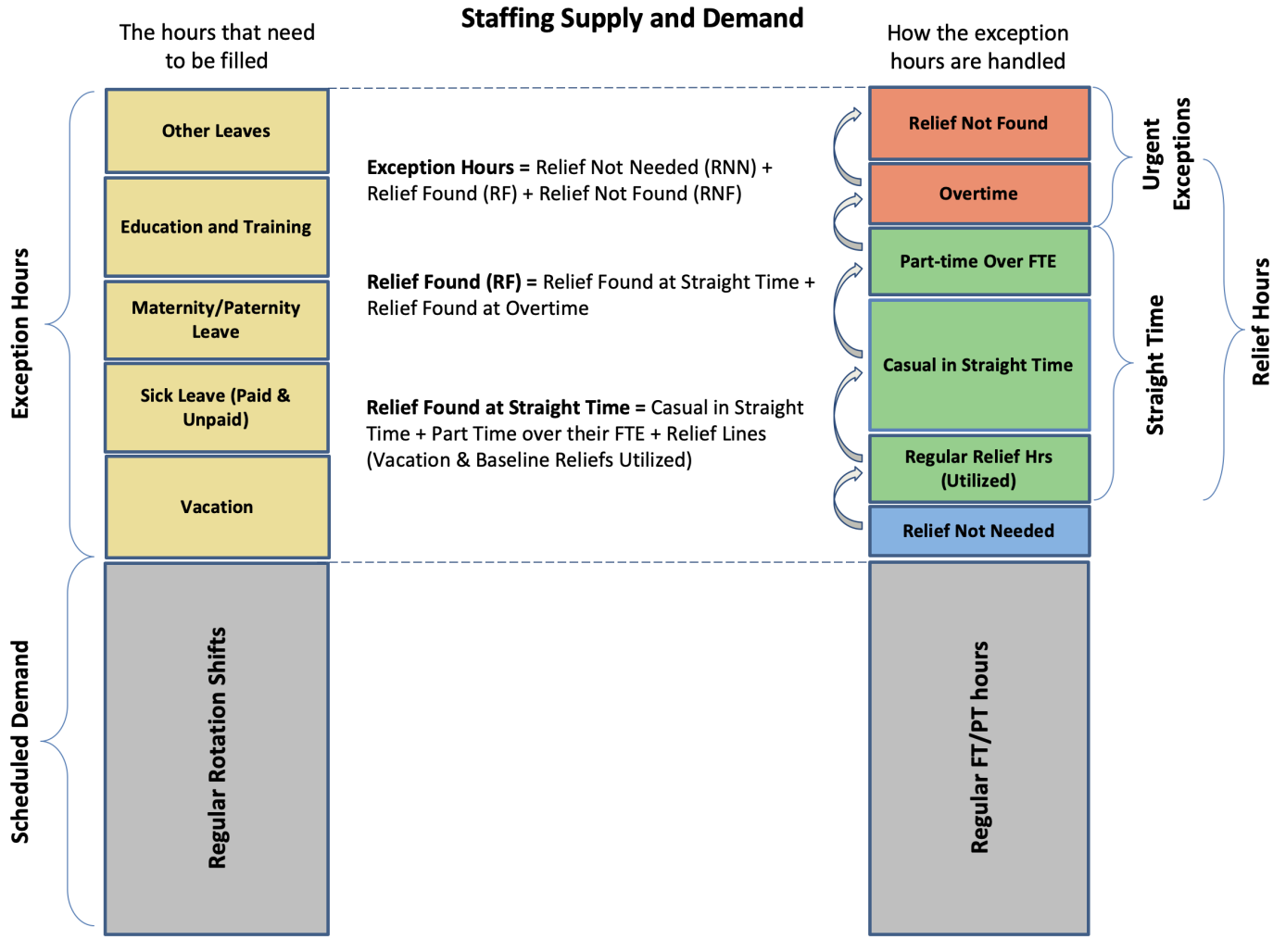
- Fit meaningful data using a threshold
  - Must have 300 exceptions within the past 4 years

- e.g. St Paul's Hospital, Registered Nurse - DC1, EMERG
  - MAE: 55.22



Urgent Exception Prediction

Predicting the number of urgent exceptions


Staffing Supply and Demand
The hours that need to be filled
How the exception hours are handled
Other Leaves
Education and Training
Maternity/Paternity Leave
Sick Leave (Paid & Unpaid)
Vacation
Exception Hours
Regular Rotation Shifts
Scheduled Demand
Exception Hours = Relief Not Needed (RNN) + Relief Found (RF) + Relief Not Found (RNF)
Relief Found (RF) = Relief Found at Straight Time + Relief Found at Overtime
Relief Found at Straight Time = Casual in Straight Time + Part Time over their FTE + Relief Lines (Vacation & Baseline Reliefs Utilized)
Relief Not Found
Overtime
Part-time Over FTE
Casual in Straight Time
Regular Relief Hrs (Utilized)
Relief Not Needed
Regular FT/PT hours
Urgent Exceptions
Straight Time
Relief Hours

Urgent Exception

- Exceptions backfilled by **Overtime** and **Relief Not Found**
  - Overtime: high cost that need to minimize
  - Relief Not Found: need to avoid
- Give a insight so HR can arrange on-call and other backfills

Difficulties

- Not too many features
- Randomness in daily basis

Method

- Linear Regression

Data

- Dates: Until 2018, excluding 2014
- Job Family: DC1000, DC2A00, DC2B00
- Earning Category: Overtime & Relief Not Found

Variables

- Dates (One Hot Encoding)
  - Day of week, day of month, week of year, month of year
- Productive hours

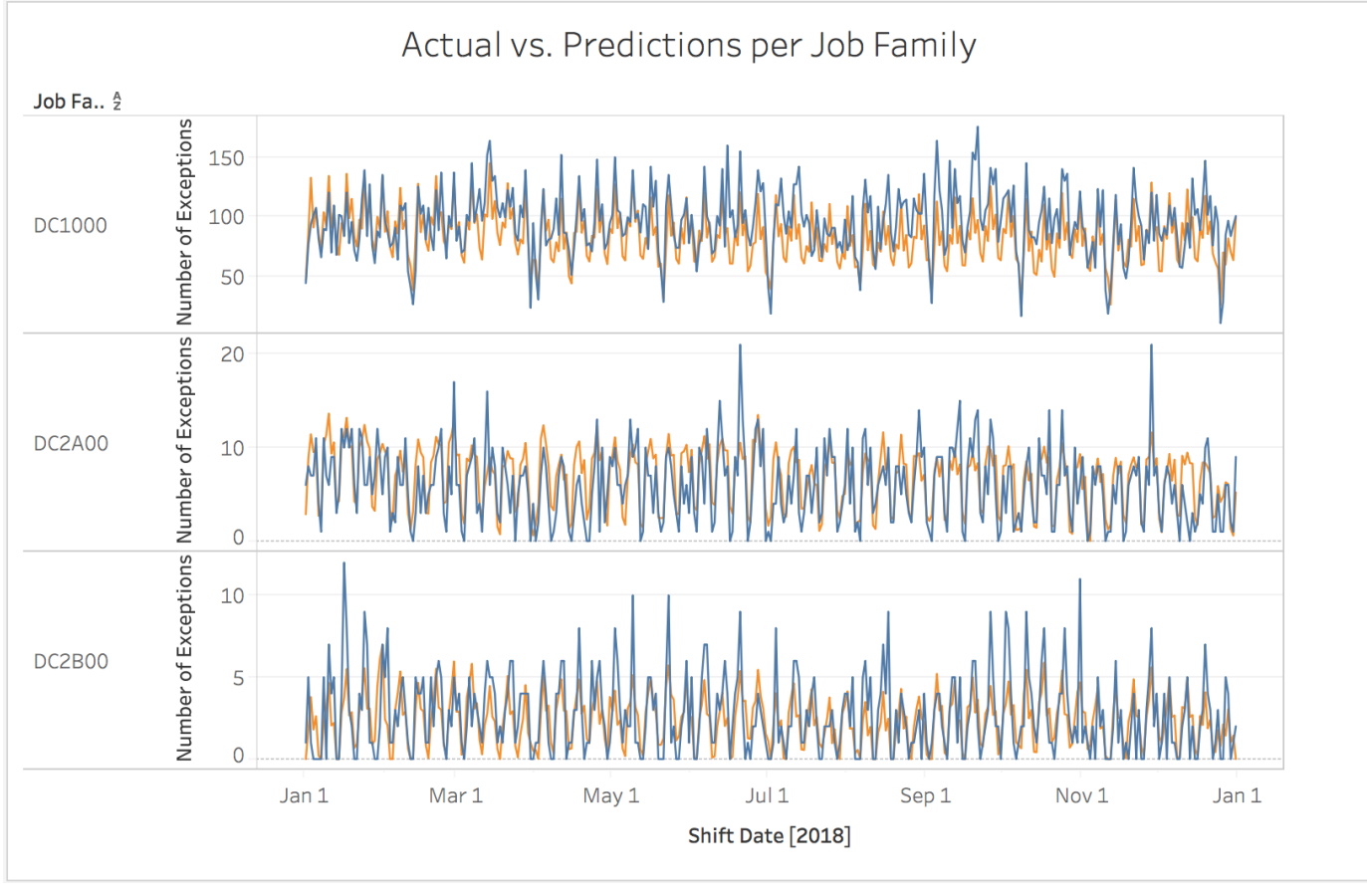
Input file

- Exception Hours for past years
- Productive Hours for past years
- Productive Hours for the period you want to predict (estimation)

Output file

- `.csv` file with dates, job family, predicted count

	A	B	C	D
1		yhat	ds	job_family
2	0	46.5760008	1/1/18	DC1000
3	1	86.868815	1/2/18	DC1000
4	2	133.55888	1/3/18	DC1000
5	3	91.3956964	1/4/18	DC1000
6	4	106.572805	1/5/18	DC1000
7	5	85.9658619	1/6/18	DC1000
8	6	72.0211963	1/7/18	DC1000
9	7	104.243195	1/8/18	DC1000


Actual vs. Predictions per Job Family
Job Fa..
DC1000
DC2A00
DC2B00
Number of Exceptions
150
100
50
20
10
0
10
5
0
Jan 1
Mar 1
May 1
Jul 1
Sep 1
Nov 1
Jan 1
Shift Date [2018]
Shift Date
1/1/2018
12/31/2018
Actual Values
Predictions

Exception Classification

Forecasting possible outcome for each exception submitted

Label Grouping

- EARNING_CATEGORY is the final outcome for an exception
- Original EARNING_CATEGORY has 12 values which is too much for classification
- 3 labels is more reasonable for classifation:
  - Straight Time: Regular Relief Utilized, Casual at Straight-Time, PT Over FTE, Miscellaneous Straight-Time, PT Employee Moved - Straight-Time, FT Employee Moved - Straight-Time
  - Overtime and Beyond: Overtime, Agency, Insufficient Notice, On-Call, Relief Not Found
  - Relief Not Needed: Relief Not Needed.

Feature Selection

- Using `EXCEPTION_HOURS`, `EXCEPTION_CREATION_TO_SHIFTSTART_MINUTES`,`NOTICE` as accuracy baseline.
- Using forward selection, adding `SITE`, `PROGRAM`, `SUB_PROGRAM`, `EXCEPTION_GROUP`, `MONTH`, `DEPARTMENT`, `SHIFT`.

Prediction Result Analysis

Validation Accuracy	0.841
Straight Time Accuracy	0.936
Overtime and Beyond Accuracy	0.638
Relief Not Needed Accuracy	0.308

Difficulties

- Imblanced Data

	Training Data Set
Number of Straight Time	262,608
Number of Overtime and Beyond	76,863
Number of Relief Not Needed	11,086

Accuracies

	Validation Accuracy		Test Accuracy
	Original Model	Adjusted Model	
Overall	0.841	0.794	0.800
Straight Time	0.936	0.823	0.830
Overtime & Beyond	0.638	0.735	0.756
Relief Not Needed	0.308	0.625	0.633

Output file

- .csv file with the prediction result

	AX	AY
1	SHIFT_	PREDICTION
2	1	Straight Time
3	1	Overtime and Beyond
4	2	Straight Time
5	3	Straight Time
6	1	Straight Time
7	1	Straight Time
8	1	Straight Time
9	2	Overtime and Beyond
10	1	Overtime and Beyond
11	3	Straight Time
12	1	Overtime and Beyond

Dashboard

Exception Predictions

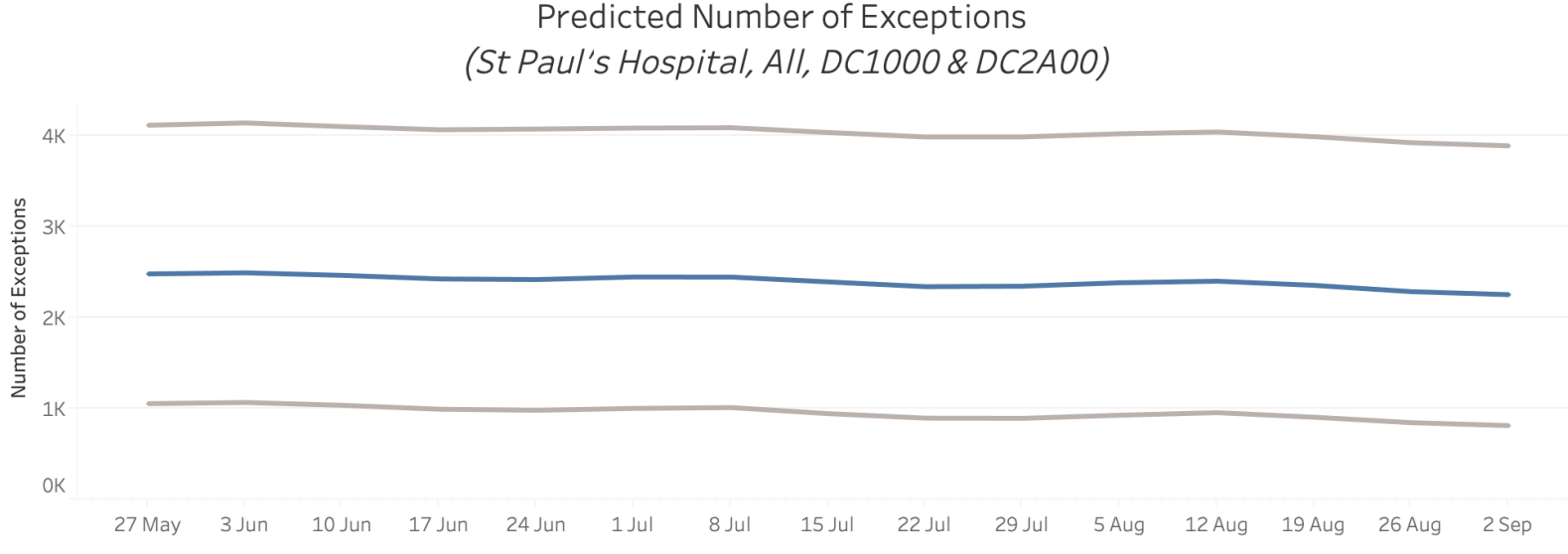
Shift Date: 2018-05-23 — 2018-09-05

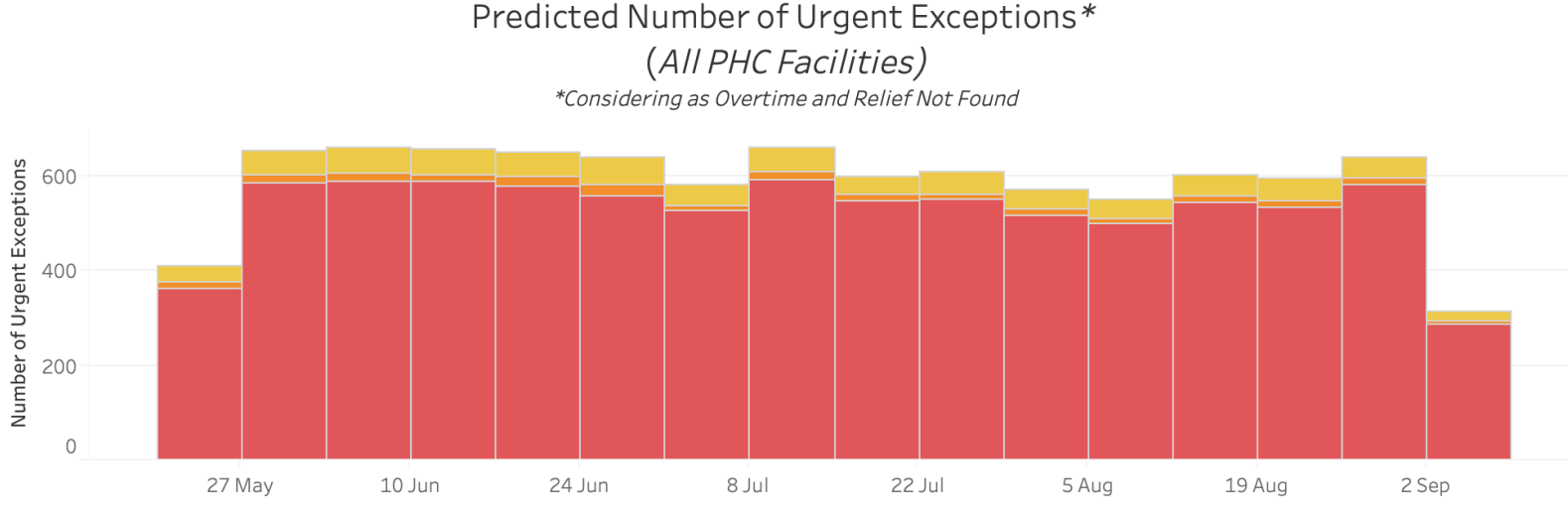
Site: St Paul's Hospital

Sub Program: (All)

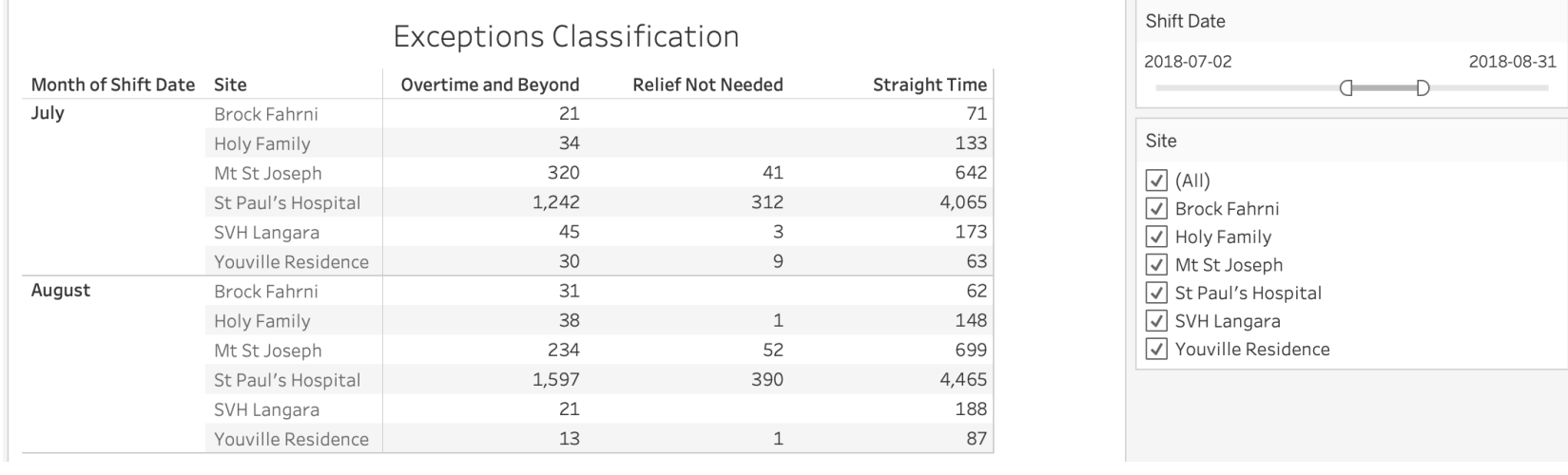
Job Family: (Multiple values)

Exceptions Predicted
95% CI - Lower Limit
95% CI - Upper Limit
DC2A00
DC2B00
DC1000

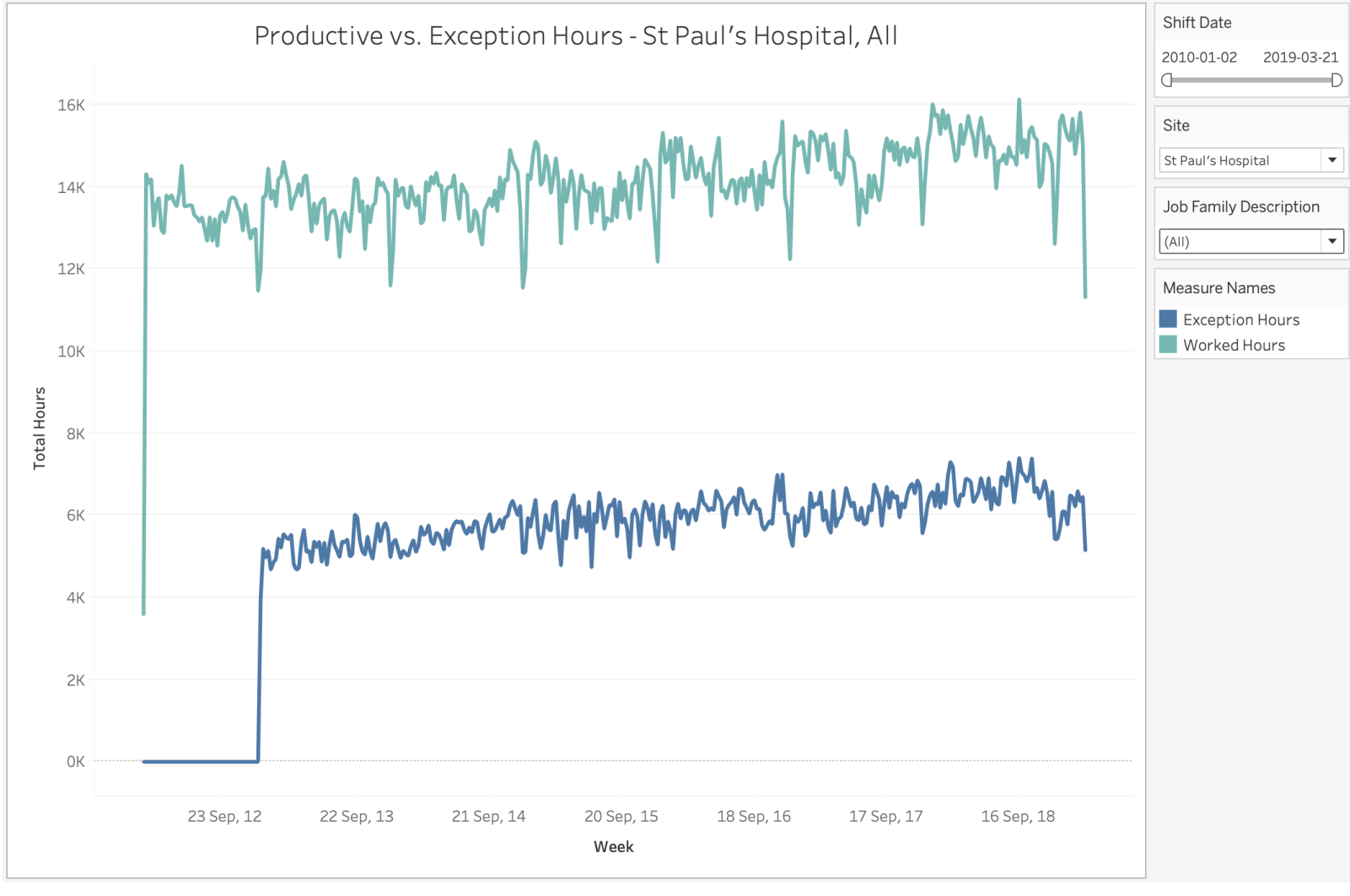




Exceptions Classification



Comparison of Productive and Exception Hours



Summary

- Data product contains the three models
- Results from the models can be shown in tableau dashboard
- HR can choose models based on the data they get
- Provide insights from the predictions to help decision making

Thank you!

Appendix

JOB_FAMILY <chr>	count <int>	cumsum <int>	freq <dbl>	cum_freq <dbl>
DC1000	638075	638075	0.846	0.846
DC2A00	48841	686916	0.065	0.911
DC2B00	34328	721244	0.046	0.957
LPN001	22792	744036	0.030	0.987
PS2000	2502	746538	0.003	0.990
PS1000	2390	748928	0.003	0.993
DC3000	1409	750337	0.002	0.995
CH1000	959	751296	0.001	0.996
EXCLST	769	752065	0.001	0.997
CH3000	608	752673	0.001	0.998

1–10 of 29 rows