CBH Market Analysis

Nichada Wongrassamee

2025-02-25

Contents

1 Overview

Understanding Unclaimed Shifts and Strategies to Reach 100% Claim Rate We will analyze shift booking patterns on Clipboard Health's marketplace, focusing on the **root causes of unclaimed shift offers** and strategies to **maximize claim rates**. Clipboard Health operates as a two-sided marketplace, where workers book per diem shifts posted by workplaces. Ensuring a **high claim rate** is crucial to driving revenue, improving worker satisfaction, and maintaining workplace efficiency.

Unclaimed shifts represent missed opportunities for both sides of the marketplace: workplaces face staffing shortages, while workers miss potential earnings. By examining **transactional data,** we aim to identify key **factors influencing unclaimed shifts,** such as **pay rate, lead time, and workerplace preferences.**

Our objective is to uncover actionable insights that can help drive the claim rate **closer to 100%.** This may involve **improving job-matching algorithms, optimizing pay structures, adjusting lead times, and enhancing workplace reputation**. By addressing the root causes of unclaimed shifts, we can help Clipboard Health strengthen its marketplace efficiency and increase engagement for both workers and workplaces.

```r
library(bit64)
library(data.table)
library(ggplot2)
library(knitr)
library(reshape2)

data_folder    = "/Users/nichada/MyCode/MPCS/Job/_Interview/CBH"
```

```r
# Load the CSV file using read.csv()
shift_offers <- fread(file = paste0(data_folder, "/Shift Offers v3.csv"), stringsAsFactors = FALSE)
```

1.1 Observe why some shift offers are not claimed

- As a product team, we aim to analyze why some shift offers remain unclaimed and identify strategies to achieve a 100% claim rate. Improving this metric will enhance worker and workplace satisfaction while driving revenue growth.

1.1.1 Data cleaning

- I will created new column called 'CLAIMED' to indicate offer that is 0 = unclaimed, 1 claimed. I'll be removing offers that are created by mistakes ie. canceled or deleted

```r
# Clean the canceled and deleted shifts
# create claimed column
shift_offers <- shift_offers[is.na(CANCELED_AT) & is.na(DELETED_AT)]
shift_offers[, c("CANCELED_AT", "DELETED_AT") := NULL]
shift_offers[, CLAIMED := ifelse(is.na(CLAIMED_AT), 0, 1)]

# Check the first few rows of unclaimed shifts
head(shift_offers)
```

```
                   SHIFT_ID                WORKER_ID             WORKPLACE_ID
                     <char>                   <char>                   <char>
1: 6757580b1e2d97752fd69167 65b01f2e46c0645699081cbe 5e7e45243bfbb200165914ae
2: 67550bddd79613f860549322 6696d1c1d0200bf317ee5d3c 626b0b89596c0601c2c39642
3: 66f5d05de01fd3697b18c206 66b285d5d0200bf317738e59 5cb9f07135163900163f532c
4: 66ee3848e62bb5f43e3baee5 620c6429e2ceb601ad203920 611af67795f4c501662edb31
5: 67254195894c05fb0e389089 62310b576d05fc01b11922ec 612e88a59e065601668a13a4
6: 677b553df0e33d9606282ec6 632565c79603d78083c25520 6081f3fc667fa6016195942c
        SHIFT_START_AT    SHIFT_CREATED_AT     OFFER_VIEWED_AT DURATION    SLOT
```

```
                <POSc>              <POSc>              <POSc>    <int> <char>
1: 2024-12-09 23:00:00 2024-12-09 20:50:19 2024-12-09 21:18:42        8     pm
2: 2024-12-08 15:00:00 2024-12-08 03:00:46 2024-12-08 04:04:14        6     am
3: 2024-09-27 14:00:00 2024-09-26 21:21:34 2024-09-27 04:19:45        8     am
4: 2024-10-08 21:30:00 2024-09-21 03:06:48 2024-10-06 00:46:37        8     pm
5: 2024-11-02 14:00:00 2024-11-01 21:01:09 2024-11-02 00:04:48        8     am
6: 2025-01-06 06:00:00 2025-01-06 03:59:58 2025-01-06 04:13:02        8    noc
   CLAIMED_AT IS_VERIFIED IS_NCNS PAY_RATE CHARGE_RATE CLAIMED
       <POSc>      <lgcl>  <lgcl>    <num>       <int>   <num>
1:       <NA>       FALSE   FALSE    21.29          29       0
2:       <NA>       FALSE   FALSE    21.97          30       0
3:       <NA>       FALSE   FALSE    19.05          28       0
4:       <NA>       FALSE   FALSE    22.13          24       0
5:       <NA>       FALSE   FALSE    21.73          28       0
6:       <NA>       FALSE   FALSE    34.02          33       0
```

- Create a new variable `LEAD_TIME_HOUR`

```r
# Create a new variable for lead time (in hours) between posting and shift start
shift_offers[, SHIFT_CREATED_AT := as.POSIXct(SHIFT_CREATED_AT)]
shift_offers[, SHIFT_START_AT := as.POSIXct(SHIFT_START_AT)]
shift_offers[, LEAD_TIME_HOUR := as.numeric(difftime(SHIFT_START_AT,
                                            SHIFT_CREATED_AT, units = "hours"))]
```

1.1.2 Initial comprison between claimed and unclaimed pay rate

```r
claimed_shifts <- shift_offers[CLAIMED == 1]
unclaimed_shifts <- shift_offers[CLAIMED == 0]
shift_offers$CLAIMED <- as.factor(shift_offers$CLAIMED)

cat("unclaimed counts: ", nrow(unclaimed_shifts), "\n")
```

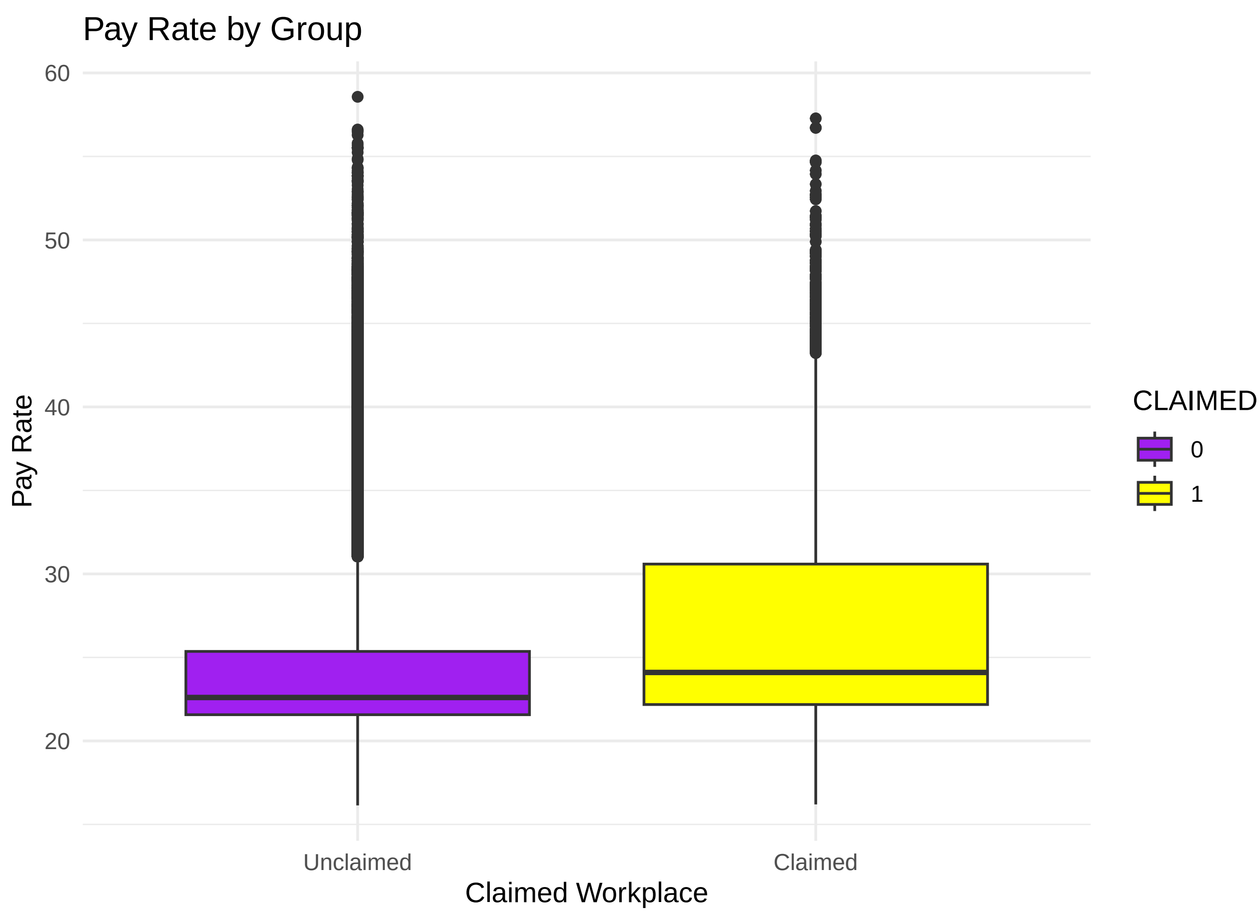
```
unclaimed counts:  197829
```

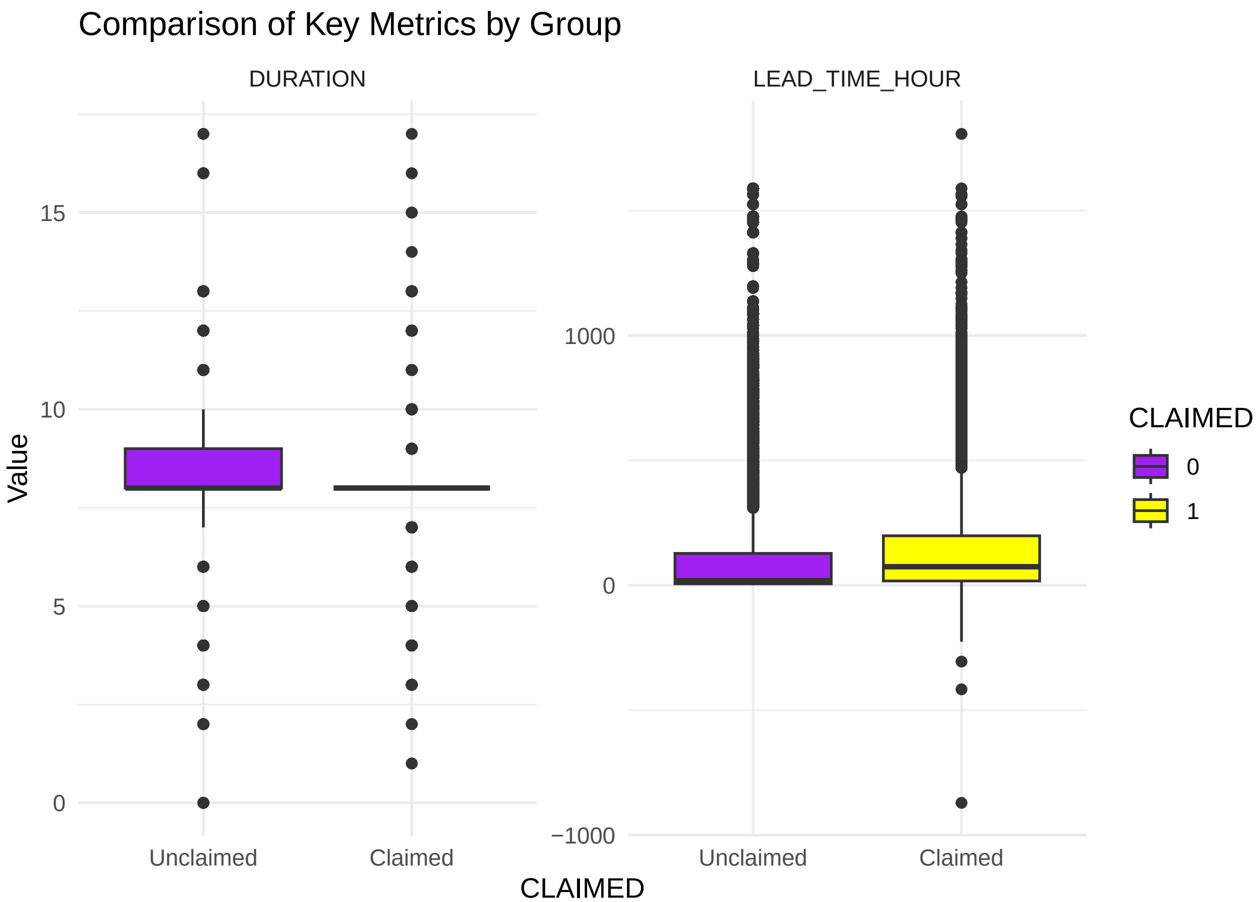
```r
cat("claimed counts: ", nrow(claimed_shifts), "\n")
```

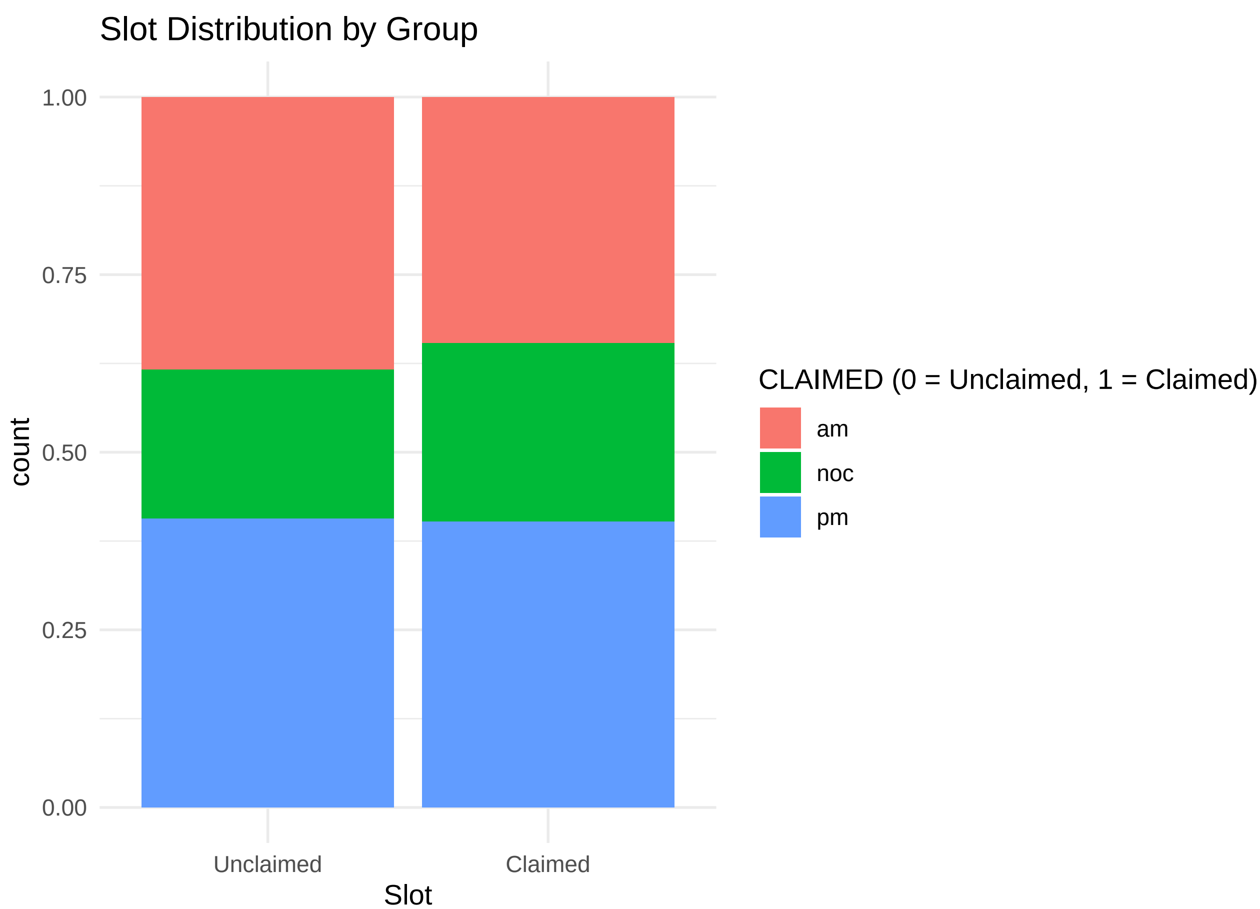
```
claimed counts:  12583
```

```r
cat("ave unclaimed rate: ", nrow(unclaimed_shifts)/
      (nrow(unclaimed_shifts) + nrow(claimed_shifts)), "% \n")
```

```
ave unclaimed rate:  0.9401983 %
```


Pay Rate by Group
60
50
40
30
20
Pay Rate
Unclaimed
Claimed
Claimed Workplace
CLAIMED
0
1


Comparison of Key Metrics by Group
DURATION
LEAD_TIME_HOUR
Value
15
10
5
0
1000
0
−1000
Unclaimed
Claimed
Unclaimed
Claimed
CLAIMED
CLAIMED
0
1




```
# Export to csv
fwrite(unclaimed_shifts, file = paste0(data_folder, "/unclaimed_shifts.csv"))
fwrite(claimed_shifts, file = paste0(data_folder, "/claimed_shifts.csv"))
```

Key Findings & Insights from Initial analysis

- 94% of shift offers are unclaimed. Only 12,583 out of 210,412 were successfully claimed.
- Pay rate alone does not strongly predict claim rates.
  - Claimed shifts have a slightly higher median pay but still overlap with unclaimed shifts.
- Similarly slot time does not provide significant impact. While lead time suggests some impact. However, the workplace ID has product strongest impact (see next section).
- Next step, I will perform **Deeper Workplace Analysis**, identifying which workplaces have the lowest claim rates and understand why.

2 Data Analysis

To better understand why some shifts remain unclaimed, we start by analyzing unclaimed rates across different **workplaces**. This provides a high-level view of which workplaces struggle the most to fill shifts.

Next, we apply logistic regression to determine whether these patterns hold after controlling for key factors like:

- Pay rate (Does higher pay increase claim rates?)
- Shift timing (Are certain shift slots harder to fill?)
- Lead time (Do shifts posted too late remain unclaimed?)

By combining direct statistical analysis with machine learning predictions, we aim to identify actionable strategies to reduce unclaimed shifts.

2.0.1 Identify the Bad Workplace

To analyze how specific workplaces correlate with unclaimed shift offers, I use two complementary methods:

(1) **Unclaimed Rate Analysis:**

- I calculate the **percentage of shifts left unclaimed per workplace** to identify consistently low-performing locations.

(2) **Logistic Regression Model on Workplace Id:**

(3) **Logistic Regression Model Expanded:**

- I build a predictive model to determine which factors contribute most to unclaimed shifts while controlling for pay rate, shift type, and lead time.
- This helps separate true workplace effects from other confounding variables.

2.1 Method 1: Unclaimed Rate Method

Objective:

- This method calculates the percentage of unclaimed shifts per workplace by dividing the number of unclaimed shifts by the total shifts.

Strengths:

- Simple and easy to interpret. - Provides a direct view of the proportion of unclaimed shifts without relying on complex modeling.

Limitations:

- Does not account for differences in the total number of shifts between workplaces.

- Lacks statistical significance testing to determine if observed differences between each workplace unclaimed rate are meaningful or due to random variation. (because if data is small it could be due to data fluctuations).

- I created a simplified index WORKER_IDX to easily identify WORKER_ID.

```r
# Convert workplace_id to a factor, then to integer
shift_offers[, WORKPLACE_IDX := as.integer(as.factor(WORKPLACE_ID))]
shift_offers[, WORKPLACE_IDX := factor(WORKPLACE_IDX)]
workplace_map <- unique(shift_offers[, .(WORKPLACE_IDX, WORKPLACE_ID)])
```

- Calculate Unclaimed Rate

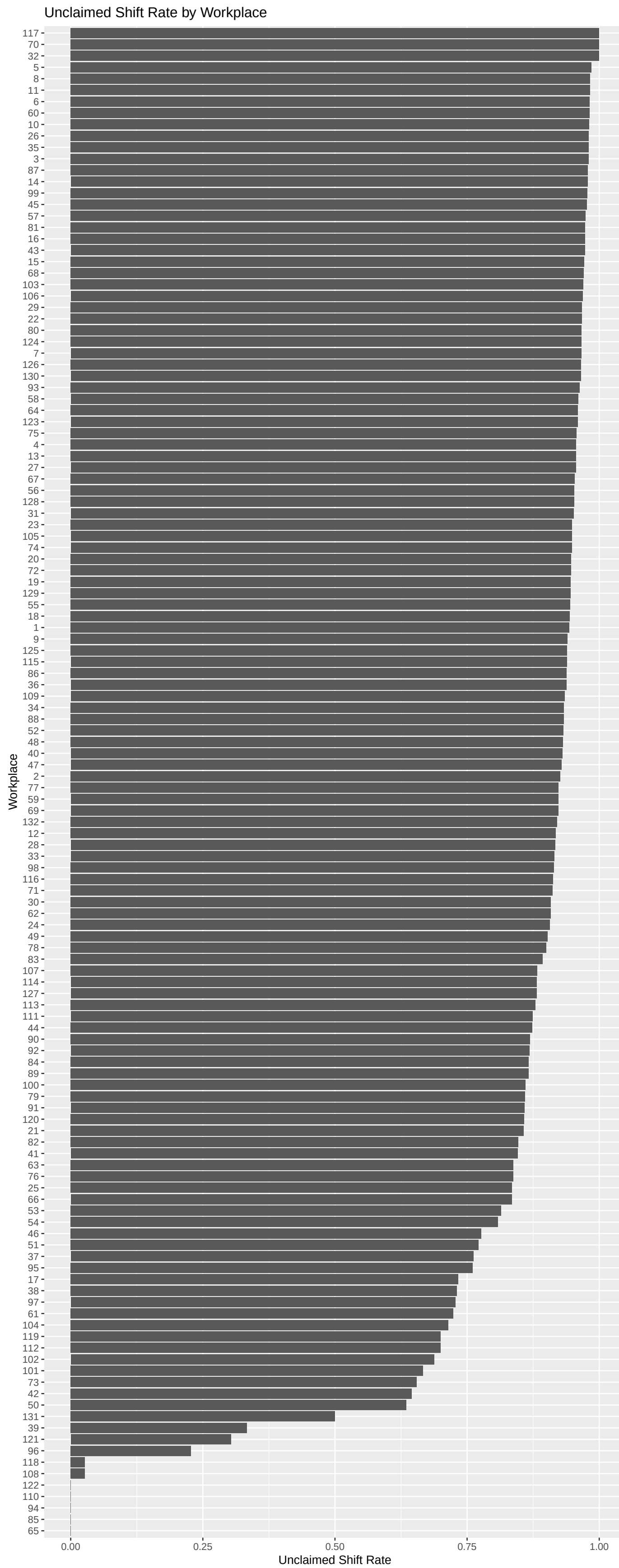
```r
# Group by workplace_id and calculate total shifts and unclaimed shifts per workplace
workplace_stats <- shift_offers[, .(
  total_shifts = .N,
  unclaimed_count = sum(CLAIMED == 0),
  claimed_count = sum(CLAIMED == 1)
), by = WORKPLACE_IDX]

# Calculate the unclaimed rate per workplace
workplace_stats[, unclaimed_rate := unclaimed_count / total_shifts]

# Order the results by the highest unclaimed rate
setorder(workplace_stats, -unclaimed_rate)
```

- Plot the workplace_stats unclaimed rate

```r
ggplot(workplace_stats, aes(x = reorder(WORKPLACE_IDX, unclaimed_rate), y = unclaimed_rate)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  labs(title = "Unclaimed Shift Rate by Workplace", x = "Workplace",
       y = "Unclaimed Shift Rate")
```


Unclaimed Shift Rate by Workplace
Workplace
117
70
32
5
8
11
6
60
10
26
35
3
87
14
99
45
57
81
16
43
15
68
103
106
29
22
80
124
7
126
130
93
58
64
123
75
4
13
27
67
56
128
31
23
105
74
20
72
19
129
55
18
1
9
125
115
86
36
109
34
88
52
48
40
47
2
77
59
69
132
12
28
33
98
116
71
30
62
24
49
78
83
107
114
127
113
111
44
90
92
84
89
100
79
91
120
21
82
41
63
76
25
66
53
54
46
51
37
95
17
38
97
61
104
119
112
102
101
73
42
50
131
39
121
96
118
108
122
110
94
85
65
0.00
0.25
0.50
0.75
1.00
Unclaimed Shift Rate

Observation:

- The data reveals that a few workplaces exhibit a high unclaimed rate. I will conduct further analysis to investigate the underlying causes.
- Since workplaces with high unclaimed rates align with those identified as problematic in Method 2: Logistic Regression, I will rely on this method for a more accurate classification of problematic workplaces.

2.2 Method 2: Logistic Regression Method

Objective :

- This method models the probability of a shift being claimed (or unclaimed) for each workplace while testing whether the differences are statistically significant.

How It Works :

- The regression estimates coefficients (in log-odds) for each workplace compared to a reference category.

- A **negative and significant coefficient** for a workplace means that, after controlling for other factors (if any are included), shifts at that workplace are less likely to be claimed.

- The model provides p-values and confidence intervals, allowing for a statistically robust comparison between workplaces.

Strengths :

- Provides formal statistical testing (p-values) to determine if a workplace's unclaimed rate is significantly different from the reference.

- Can expand to incorporate additional variables (e.g., shift timing, pay rate) to control for confounding factors and improve accuracy.

Limitations :

- If only workplace ID is included, it might largely reflect the raw differences; however, it still accounts for sample size variability and provides significance testing.

- The choice of reference category and dummy coding can influence interpretation.

```r
# Create a binary variable: 1 = Claimed, 0 = Unclaimed
# Run a logistic regression with workplace_id as a predictor.
# Note: This will create dummy variables for each workplace_id.
model <- glm(CLAIMED ~ WORKPLACE_IDX, data = shift_offers, family = binomial)
```

```r
model_summary <- summary(model)

# Convert the coefficients matrix into a data frame
coefs <- as.data.frame(model_summary$coefficients)

# Name the columns for clarity
# Typically, the columns are: Estimate, Std. Error, z value, and Pr(>|z|)
colnames(coefs) <- c("Estimate", "Std.Error", "z.value", "p.value")

# Create a column for the term names (row names)
coefs$Term <- rownames(coefs)

# Exclude the intercept from the sorting:
coefs <- coefs[coefs$Term != "(Intercept)", ]

# Sort by p-value in ascending order (lowest p-value = most significant)
```

```r
coefs_sorted <- coefs[order(coefs$Estimate), ]
coefs_significant <- coefs_sorted[coefs_sorted$p.value < 0.05, ]
```

2.2.1 Investigate the worst workplaces

- Investigate the worst workplaces: The most negative coeficients, statistically significant workplaces ($p < 0.05$) are likely causing unclaimed shifts.
- Explore factors like pay rates, shift duration, or lead time at these locations.

```r
problematic_workplaces <- setDT(coefs_significant[coefs_significant$Estimate < 0, ])

setnames(problematic_workplaces, "Term", "WORKPLACE_IDX")
problematic_workplaces[, WORKPLACE_IDX := sub("^WORKPLACE_IDX", "", WORKPLACE_IDX)]

problematic_workplaces <- merge(
  problematic_workplaces,
  workplace_map,
  by = "WORKPLACE_IDX",    # common column
  all.x = TRUE             # left join; keep all rows from 'problematic_workplaces'
)

problematic_workplaces <- problematic_workplaces[, .(WORKPLACE_IDX, WORKPLACE_ID)]

kable(data.frame(problematic_workplaces), caption = "Problematic Workplaces - Coeficient < 0 ")
```

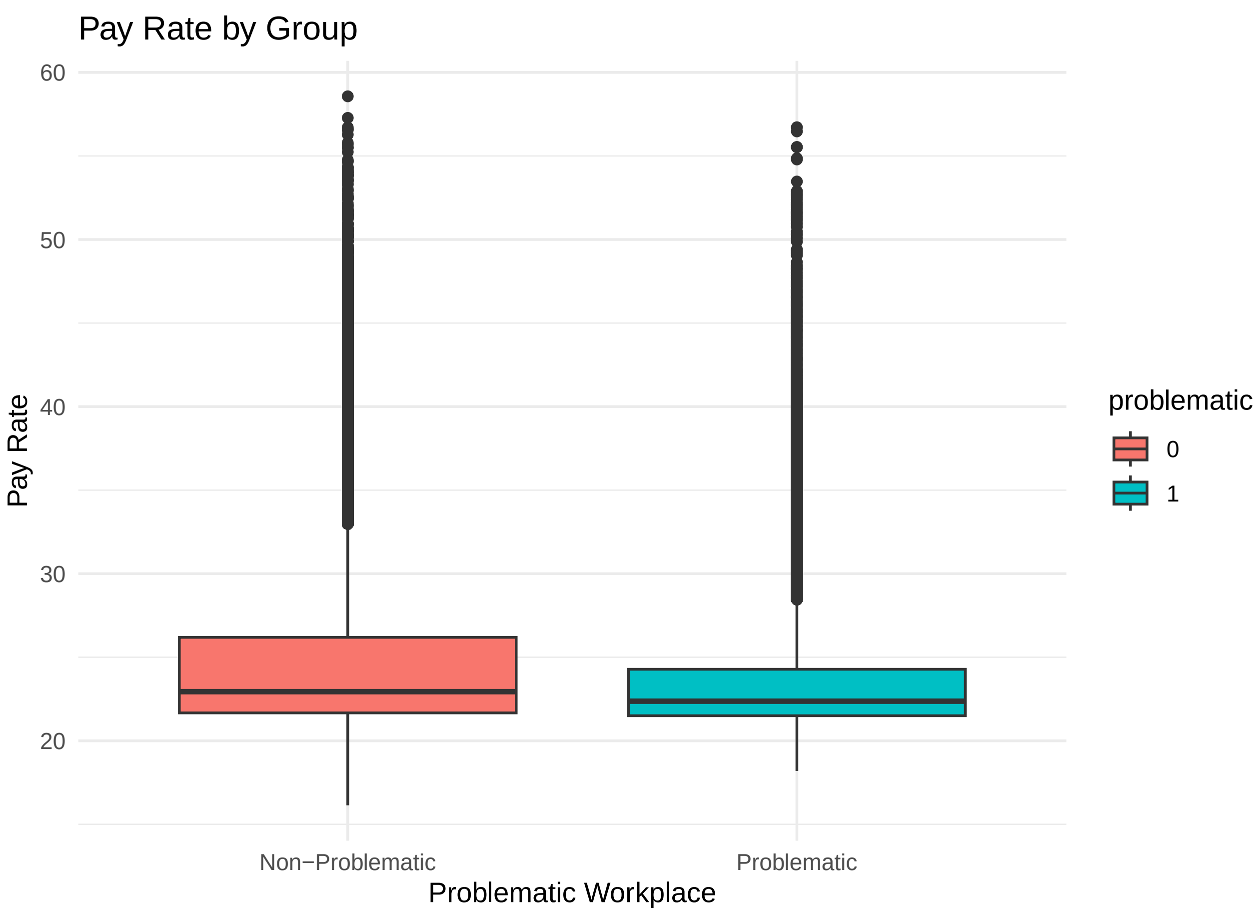
Table 1: Problematic Workplaces - Coeficient < 0

WORKPLACE_IDX	WORKPLACE_ID
10	5e47201390bc130016bbe03f
103	63c6f41d89b71301b666627f
106	641deef039646601c0df8cea
11	5e5e85ea43a2390016e3d65c
124	65fde23ff14f0eebdc8178d2
130	66b3ab3c0e400bcd45f4ae16
14	5e7e2c903bfbb20016504128
15	5e7e411c3bfbb200165725c8
16	5e7e45243bfbb200165914ae
22	5ebf09a7fe8b200017aeb9eb
26	5ecc20d1a451e60017ab56f7
29	5f2305c10626e100169c6000
3	5c06fe1f61d521000488a0f2
35	5f5bca132f95160016b723ff
43	5ff4f626909f7a00160d06fd
45	60257cdc1cd3a100160c67cf
5	5c82b3dfa08cb800166dc04d
57	611bc90b8a1a140166c0a051
6	5cb9f07135163900163f532c
60	613a636b4be49701666cf108
68	61b0d0ed0a1f9a01805c0531
7	5d011e7757d52f00162c9fe1
8	5e1ce78827ff480016e9133e
80	622fd7d048ade501b1ab4b15
81	6233963a9f423f01aa0b8e0c

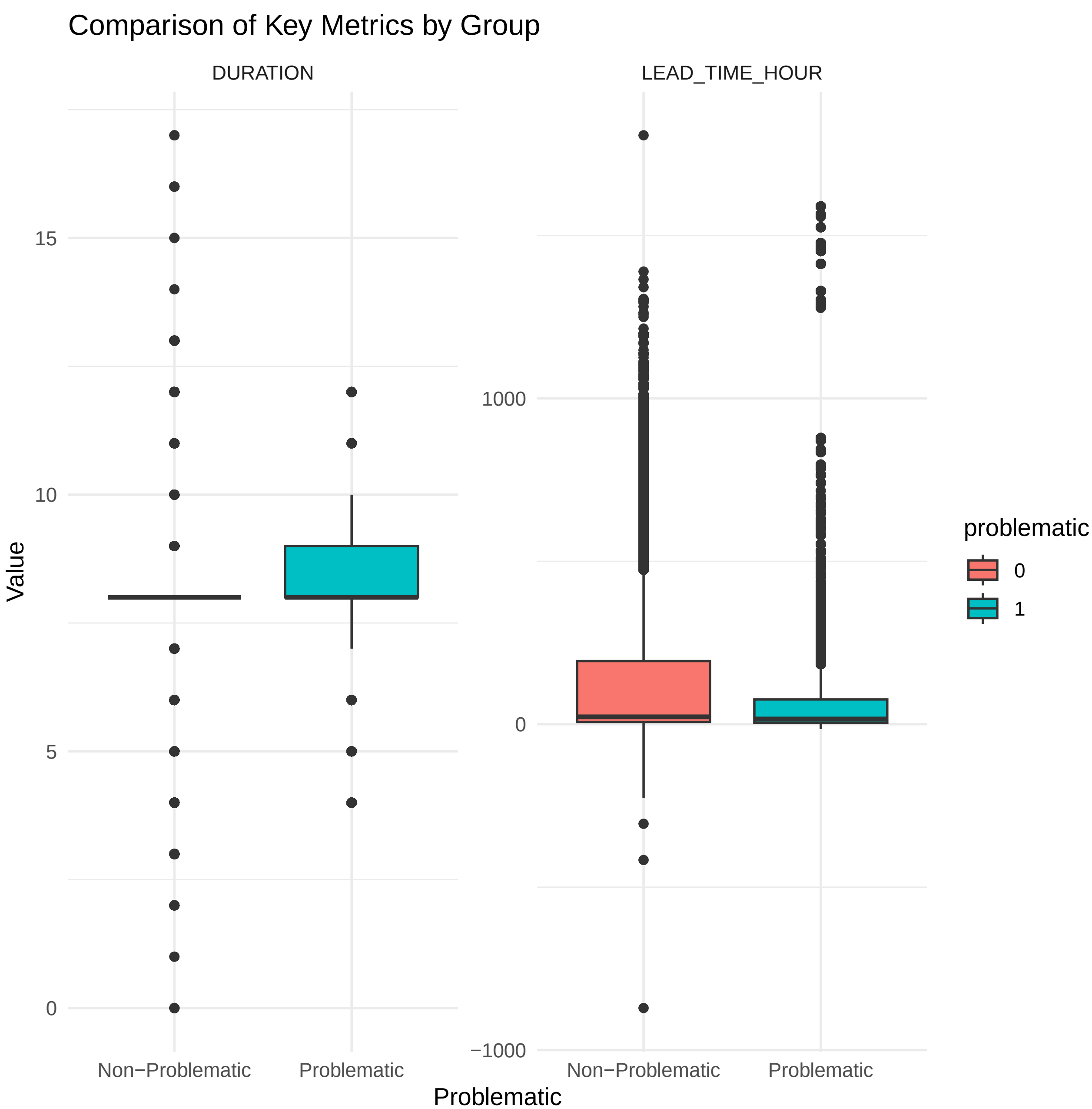
- Incorporate problematic workplaces indicator on the shift offers data.

```
# Mark problematic workplace in the shift_offers data set
shift_offers[, problematic := factor(ifelse(WORKPLACE_IDX %in% problematic_workplaces$WORKPLACE_IDX, 1,
```

- Payrate between problematic vs. non-problematic workplaces



- Duration, Lead Time Hour

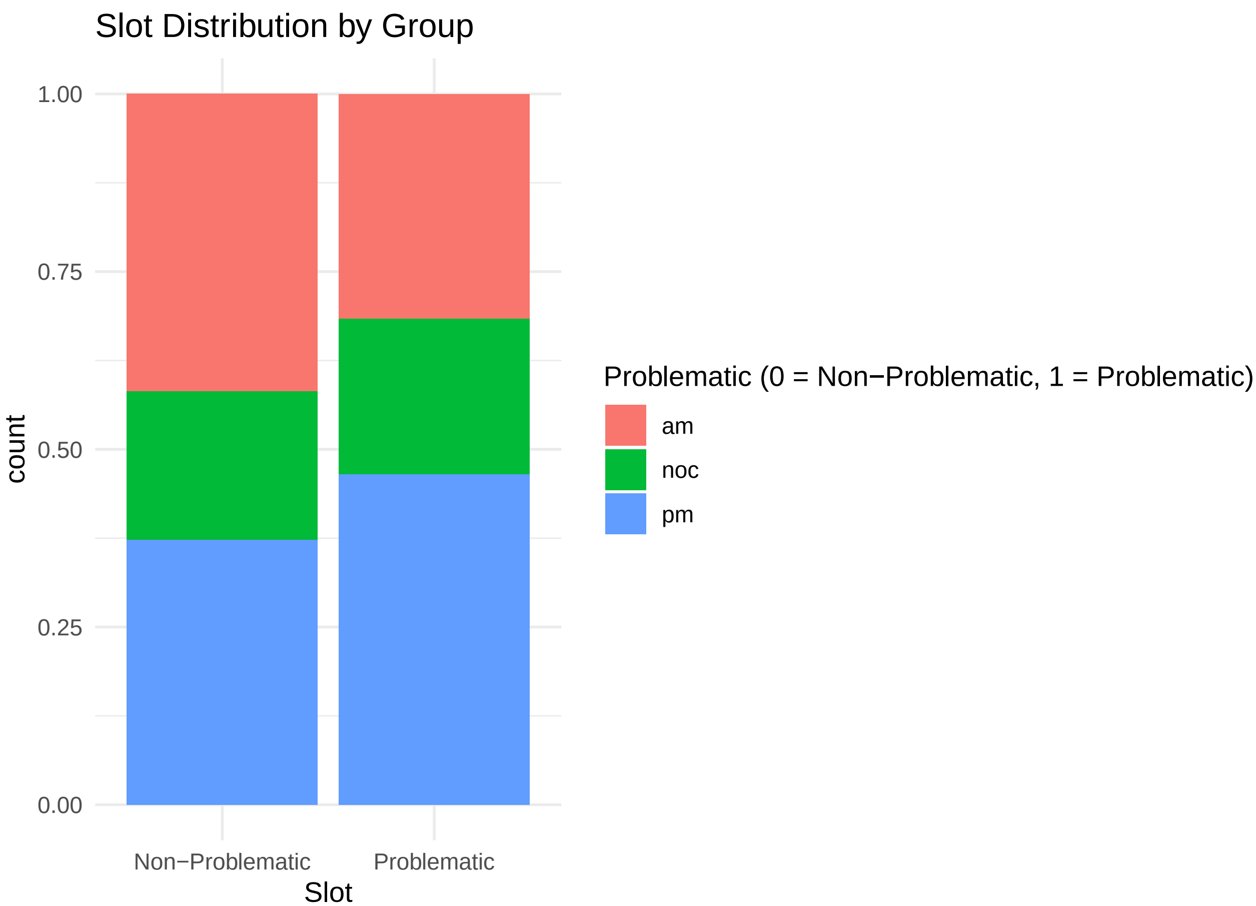


- Calculate stats for problematic (1) vs non-problematic workplace (0)

```
# Compute summary statistics for each group
shift_offers[, .(
  count = .N,
  mean = mean(LEAD_TIME_HOUR, na.rm = TRUE),
  median = median(LEAD_TIME_HOUR, na.rm = TRUE),
  sd = sd(LEAD_TIME_HOUR, na.rm = TRUE),
  min = min(LEAD_TIME_HOUR, na.rm = TRUE),
  max = max(LEAD_TIME_HOUR, na.rm = TRUE),
  q25 = quantile(LEAD_TIME_HOUR, 0.25, na.rm = TRUE),
  q75 = quantile(LEAD_TIME_HOUR, 0.75, na.rm = TRUE)
), by = problematic]
```

```
   problematic  count     mean   median        sd         min        max        q25
```

```
        <fctr>  <int>    <num>    <num>    <num>      <num>    <num>    <num>
1:           1  76142  86.8405 16.07167 204.1351  -14.75889 1588.878 4.895833
2:           0 134270 150.5896 22.63444 235.8678 -870.93278 1807.198 7.059444
         q75
       <num>
1:  76.00944
2: 193.87056
```



2.2.2 Problematic vs Non-Problematic workplace characteristics analysis

1. **Pay Rate**: No significant difference between problematic and non-problematic workplaces.
2. **Shift Duration**: No significant impact observed.
3. **Lead Time Differences**: Problematic workplaces have a shorter lead time than non-problematic ones.
    1. Median Lead Time:
        - Non-Problematic: 22.63 hours
        - Problematic: 16.07 hours
    2. Interquartile Range (IQR):
        - Problematic: Q25 = 4.89, Q75 = 76.00 (Range ≈ 71 hours)
        - Non-Problematic: Q25 = 7.05, Q75 = 193.87 (Range ≈ 186 hours)
    3. Negative Lead Time:
        - Non-Problematic minimum = -870.93 hours, possibly indicating data errors or shifts posted after they started (suggesting backfilled shifts).
4. **Shift Timing**: Problematic workplaces tend to have more PM shifts than AM shifts, but this difference

is not statistically significant.

5. **Workplace Name Influence**: There may be an issue with the workplace name itself making it less desirable to workers. Further investigation is needed.

2.3 Method 3 : Expand Models

- Explore other factors that are contributing to the unclaimed offers
- Use log regression
- Remove n/a data from the variables

```r
sum(is.na(shift_offers$CHARGE_RATE))
```

```
[1] 0
```

```r
sum(is.na(shift_offers$PAY_RATE))
```

```
[1] 0
```

```r
sum(is.na(shift_offers$DURATION))
```

```
[1] 0
```

- Normalize linear data : PAY_RATE, DURATION, LEAD_TIME_HOUR

```r
# Before normalizing, save the mean and standard deviation of each variable:

pay_rate_mean <- mean(shift_offers$PAY_RATE, na.rm = TRUE)
pay_rate_sd <- sd(shift_offers$PAY_RATE, na.rm = TRUE)

duration_mean <- mean(shift_offers$DURATION, na.rm = TRUE)
duration_sd <- sd(shift_offers$DURATION, na.rm = TRUE)

lead_time_mean <- mean(shift_offers$LEAD_TIME_HOUR, na.rm = TRUE)
lead_time_sd <- sd(shift_offers$LEAD_TIME_HOUR, na.rm = TRUE)
```

```r
# Automatically handled one hot encoding by glm() in R when using factor().
shift_offers$SLOT <- factor(shift_offers$SLOT, levels = c("am", "pm", "noc"))

# Normalizing Linear Data
shift_offers$PAY_RATE <- scale(shift_offers$PAY_RATE)
shift_offers$DURATION <- scale(shift_offers$DURATION)
shift_offers$LEAD_TIME_HOUR <- scale(shift_offers$LEAD_TIME_HOUR)
```

- exclude workplace_id from variable to avoid overfitting to the workplace_id

```r
model_full <- glm(CLAIMED ~ PAY_RATE + DURATION + SLOT + LEAD_TIME_HOUR,
                  data = shift_offers, family = binomial)
summary(model_full)
```

```

Call:
glm(formula = CLAIMED ~ PAY_RATE + DURATION + SLOT + LEAD_TIME_HOUR,
    family = binomial, data = shift_offers)

Coefficients:
                Estimate Std. Error  z value Pr(>|z|)
(Intercept)    -2.948694   0.016066 -183.536  < 2e-16 ***
PAY_RATE        0.355733   0.007187   49.500  < 2e-16 ***
DURATION        0.033926   0.008716    3.892 9.93e-05 ***
SLOTpm          0.138039   0.021555    6.404 1.51e-10 ***
SLOTnoc         0.282108   0.024395   11.564  < 2e-16 ***
LEAD_TIME_HOUR  0.105045   0.008396   12.511  < 2e-16 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 95284  on 210411  degrees of freedom
Residual deviance: 92715  on 210406  degrees of freedom
AIC: 92727

Number of Fisher Scoring iterations: 5
```

- duration has a low impact due to low coefficient < 0.05 so i'll remove.

```r
model_full <- glm(CLAIMED ~ PAY_RATE + SLOT + LEAD_TIME_HOUR,
                  data = shift_offers, family = binomial)
summary(model_full)
```

```
Call:
glm(formula = CLAIMED ~ PAY_RATE + SLOT + LEAD_TIME_HOUR, family = binomial,
    data = shift_offers)

Coefficients:
                Estimate Std. Error  z value Pr(>|z|)
(Intercept)    -2.944446   0.016010 -183.914  < 2e-16 ***
PAY_RATE        0.359050   0.007126   50.389  < 2e-16 ***
SLOTpm          0.131671   0.021482    6.129 8.83e-10 ***
SLOTnoc         0.276527   0.024348   11.357  < 2e-16 ***
LEAD_TIME_HOUR  0.114280   0.008017   14.254  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 95284  on 210411  degrees of freedom
Residual deviance: 92730  on 210407  degrees of freedom
AIC: 92740

Number of Fisher Scoring iterations: 5
```

- Separate lines for PAY_RATE, CHARGE_RATE, DURATION, etc. Each of these has its own coefficient and p-value, indicating how they affect the odds of a shift being claimed after controlling for workplace and other variables.
- Note: exclude CHARGE_RATE because it may not be a factor of consideration for worker's decision.

2.3.1 Find Cut-off points

If the logistic regression model predicts that a shift is unlikely to be claimed when LEAD_TIME_HOUR_scaled < -0.5, convert it back:

- Codes for convert normalized data into actual data

```r
lead_time_actual <- (7 * lead_time_sd) + lead_time_mean
lead_time_actual
```

```
[1] 1716.37
```

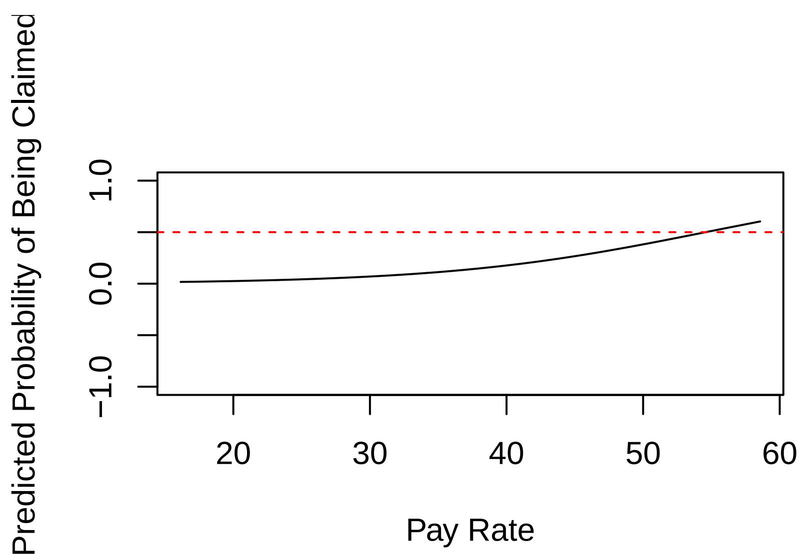
```
pay_rate_actual <- (7 * pay_rate_sd) + pay_rate_mean
pay_rate_actual
```

```
[1] 58.48083
```

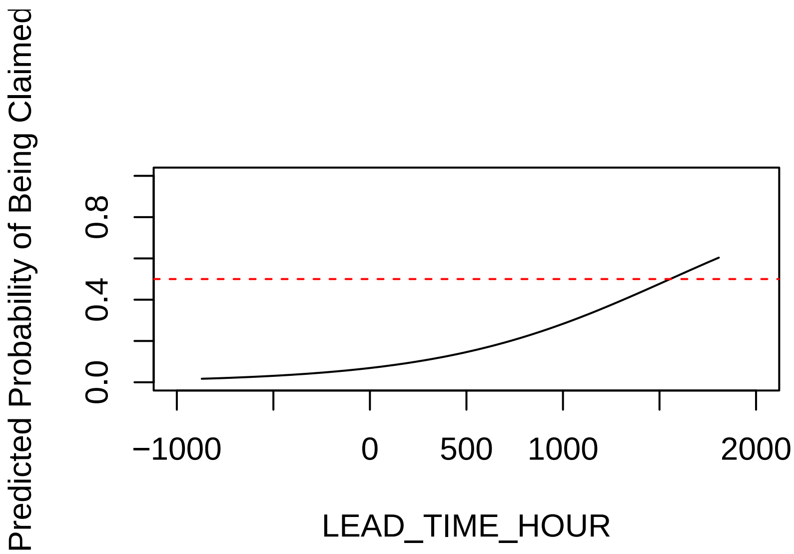
```
# Create a new data table spanning the range of rate values
new_data <- data.table(
  PAY_RATE = seq(min(shift_offers$PAY_RATE), max(shift_offers$PAY_RATE),
                 length.out = 100),
  SLOT = factor("am", levels = unique(shift_offers$SLOT)),  # Ensure it matches existing levels
  LEAD_TIME_HOUR = seq(min(shift_offers$LEAD_TIME_HOUR), max(shift_offers$LEAD_TIME_HOUR), length.out =
)

# Compute the predicted probabilities using the expanded model
new_data$pred_probs <- predict(model_full, newdata = new_data, type = "response")

# Plot predicted probability vs. rate
plot((new_data$PAY_RATE * pay_rate_sd ) + pay_rate_mean, new_data$pred_probs,
     type = "l", xlab = "Pay Rate",
     ylab = "Predicted Probability of Being Claimed", ylim = c(-1, 1))
abline(h = 0.5, col = "red", lty = 2)  # a reference line for 50% probability
```




```
# Plot predicted probability vs. rate
plot((new_data$LEAD_TIME_HOUR * lead_time_sd ) + lead_time_mean, new_data$pred_probs,
     type = "l", xlab = "LEAD_TIME_HOUR", ylab = "Predicted Probability of Being Claimed",
     xlim = c(-1000, 2000),  # Extend x-axis range
     ylim = c(0, 1))
abline(h = 0.5, col = "red", lty = 2)  # a reference line for 50% probability
```



Analysis :

- The baseline probability is low (~4.8%), meaning most shifts are unlikely to be claimed unless certain factors (like pay rate) increase the odds.

PAY_RATE cut-off point

- My model suggests that to reach a 50% claim probability, the pay rate needs to be around $55, given that every factors remai the same.
- However, the median pay rate of successfully claimed shifts is only $27. This means that lower pay rates ($27) can still result in claims, but under different conditions that may not be included in this data i.e. workplace reputation, workplace distance, negative reviews etc.

LEAD_TIME_HOUR cut-off point

- Similarly with lead time cut-off, the P=0.5 cut-off at 1500hr lead time, which is on the higher end of the median successful leadtime.
- The extreme value indicate that we cannot resolve the unclaim problems only by varying the lead time.

3 Monetary Value

- Calculate Margin of each shift offer

```
claimed_shifts$MARGIN <- claimed_shifts$CHARGE_RATE - claimed_shifts$PAY_RATE

# Compute the average margin
average_margin <- mean(claimed_shifts$MARGIN, na.rm = TRUE)

# Print the result
print(paste("Average Margin for Claimed Shifts ($):", round(average_margin, 2)))
```

```
[1] "Average Margin for Claimed Shifts ($): 5.75"
```

```
print(paste("Period of Time:", min(shift_offers$SHIFT_CREATED_AT),
            "-", max(shift_offers$SHIFT_CREATED_AT)))
```

```
[1] "Period of Time: 2024-07-29 17:36:11 - 2025-01-21 23:45:56"
```

4 Conclusion

The analysis reveals that **94%** of shift offers remain unclaimed, highlighting a critical gap in fulfillment. Several factors influence claim rates, with workplace ID having the strongest impact, followed by lead time, while pay rate and slot time show limited effects.

By addressing these factors, we can significantly reduce unclaimed shifts, improve worker engagement, and increase revenue. Below, I outline key insights and corresponding recommendations:

4.0.1 Key Insights

4.0.2 (1) Workplace Reputation Has the Strongest Impact

Finding: Some workplaces consistently struggle with shift fulfillment, even when controlling for pay rate, shift timing, and lead time. This suggests that workplace reputation, worker experiences, or management issues play a major role.

Recommendation:

- **Introduce a workplace review & rating system** to allow workers to share feedback on workplaces.
- **Analyze worker reviews & complaints** to identify common themes and intervene where necessary.
- **Improve workplace transparency** by providing insights on workplace reliability, shift completion rates, and worker satisfaction.

Business Impact:

- Improving workplace ratings and reputation can increase worker trust and shift acceptance rates, leading to fewer unclaimed shifts and higher retention.

4.0.3 (2) Shorter Lead Times Reduce Claim Rates

Finding: Shifts withshorter lead times (median 16 hours) are significantly less likely to be claimed, especially in problematic workplaces.

Recommendation:

- Encourage workplaces to post shifts earlier, aiming for a minimum 24-hour lead time.
- Implement automated early-posting reminders for workplaces with high unclaimed rates.
- Provide lead-time insights to workplaces, showing them how posting earlier improves their claim rates.

Business Impact:

- Increasing average lead time from 16 to 24 hours could significantly reduce unclaimed shifts, making Clipboard Health more efficient in shift fulfillment.

4.0.4 (3) Pay Rate Alone Does Not Guarantee Shift Claims

Finding: While claimed shifts have a slightly higher median pay rate, there is substantial overlap between claimed and unclaimed shifts, indicating that higher pay alone is not enough to drive claims.

Recommendation:

- Instead of blanket pay increases, target shifts with low claim rates for strategic pay adjustments.
- Test incentive structures (e.g., bonuses for workers who claim last-minute shifts).
- Explore non-monetary incentives, such as priority scheduling for workers who frequently accept shifts.

Business Impact:

- Optimizing pay strategy (rather than arbitrarily increasing pay) can reduce unclaimed shifts without unnecessary cost increases.

4.0.5 (4) Optimizing Shift Matching for Better Claim Rates

Finding: Neither slot time nor pay rate significantly predict shift claims, suggesting that job-worker matching **algorithms** need improvement.

Recommendation:

- **Enhance job-matching algorithms** to prioritize workers who are most likely to accept a shift.
- **Segment workers based on shift preferences, past claims, and availability** to better target offers.

Business Impact:

- A better job-matching system can increase claim rates without raising costs.

5 Recommendation to the Product Team

5.0.1 Short Term Actions

Objective

- Identify the root causes of unclaimed offers and implement quick solutions to improve claim rates.

Action Plan

- **Analyze Workplace Reviews**:
  - Review workplace feedback to identify recurring issues (e.g., poor working conditions, management concerns, lack of flexibility).
  - Investigate Worker Patterns: Identify problematic worker IDs and analyze common trends (e.g., industry type, skill mismatch, undesirable work types).
- User Research:
- **User Research (Workers)**: Conduct user interviews to understand key factors influencing their decision to accept or reject shift offers.
  - Mobile App Customer Reviews - See **Appendix**

    Customers complains are related to distance, pay rate, the amount of shifts and support team. With this regards relating to the analysis earlier it is worth emphasize on the location, type of work and the pay of the problematic companies in the earlier list.
- **User Research (Workplaces):** Interview problematic workplaces to uncover challenges they face in attracting workers.

Conclusion

- After collecting and analyzing insights from multiple sources, collaborate with the team to develop quick, iterative solutions. Evaluate the financial impact of long-term solutions, such as increasing claim rates by a specific percentage. Implement short-term fixes while planning for long-term strategic improvements.

5.0.2 Long Term Strategy

Objective

- Increase claim rates to boost revenue and improve user satisfaction.
- **Financial Impact**: If we reduce the unclaimed rate by 10%, we could generate: $5.75 \times (0.10 \times 210{,}412) = 120{,}986.9$ additional revenue per 6-month period.

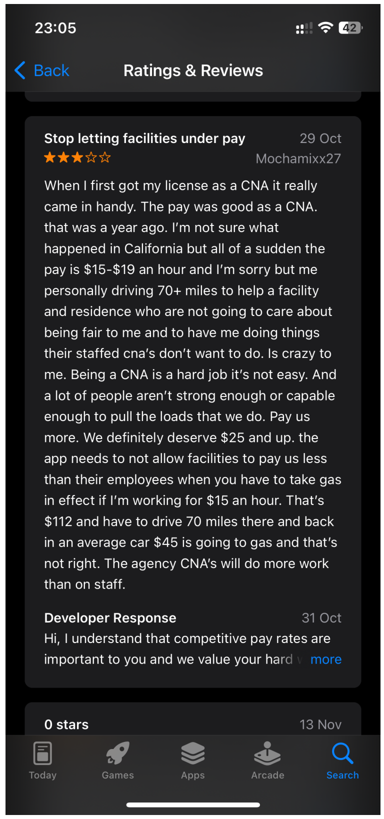
Action Plan

1. **Introduce Review & Rating Features:**
   - If not already implemented, add a **review system** where workers can **rate workplaces** and vice versa.
   - Use insights from reviews to improve matching between workers and workplaces.
2. **Improve Job Matching Algorithms:**
   - Leverage data on **worker preferences, past claims, and skills** to suggest better-matched shift offers.
   - Reduce offers being sent to workers who are unlikely to accept them.
3. **Incentivize Low-Claim Shifts:**

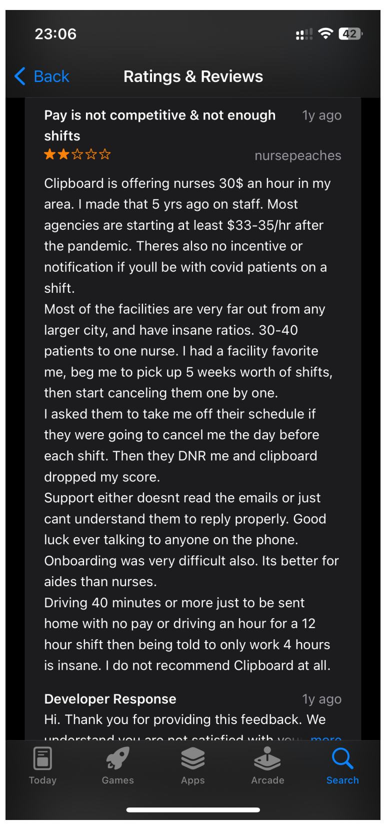
- Offer dynamic incentives (e.g., bonus pay, early access) for shifts that historically have a high unclaimed rate.

4. **Optimize Lead Time Management:**
   - Analyze whether shifts posted with short lead times are harder to claim and adjust posting strategies accordingly.

6 Appendix



23:05
Back
Ratings & Reviews
Stop letting facilities under pay
29 Oct
Mochamixx27
When I first got my license as a CNA it really came in handy. The pay was good as a CNA. that was a year ago. I'm not sure what happened in California but all of a sudden the pay is $15-$19 an hour and I'm sorry but me personally driving 70+ miles to help a facility and residence who are not going to care about being fair to me and to have me doing things their staffed cna's don't want to do. Is crazy to me. Being a CNA is a hard job it's not easy. And a lot of people aren't strong enough or capable enough to pull the loads that we do. Pay us more. We definitely deserve $25 and up. the app needs to not allow facilities to pay us less than their employees when you have to take gas in effect if I'm working for $15 an hour. That's $112 and have to drive 70 miles there and back in an average car $45 is going to gas and that's not right. The agency CNA's will do more work than on staff.
Developer Response
31 Oct
Hi, I understand that competitive pay rates are important to you and we value your hard
more
0 stars
13 Nov
Today
Games
Apps
Arcade
Search



23:06
Back
Ratings & Reviews
Pay is not competitive & not enough shifts
1y ago
nursepeaches
Clipboard is offering nurses 30$ an hour in my area. I made that 5 yrs ago on staff. Most agencies are starting at least $33-35/hr after the pandemic. Theres also no incentive or notification if youll be with covid patients on a shift.
Most of the facilities are very far out from any larger city, and have insane ratios. 30-40 patients to one nurse. I had a facility favorite me, beg me to pick up 5 weeks worth of shifts, then start canceling them one by one.
I asked them to take me off their schedule if they were going to cancel me the day before each shift. Then they DNR me and clipboard dropped my score.
Support either doesnt read the emails or just cant understand them to reply properly. Good luck ever talking to anyone on the phone.
Onboarding was very difficult also. Its better for aides than nurses.
Driving 40 minutes or more just to be sent home with no pay or driving an hour for a 12 hour shift then being told to only work 4 hours is insane. I do not recommend Clipboard at all.
Developer Response
1y ago
Hi. Thank you for providing this feedback. We
Today
Games
Apps
Arcade
Search