Amy Jung
GoodRx — Take Home Report

Analysis of Provider Terminations & Recommendations

Summary

Analysis of the “Terminated Providers” spreadsheet revealed that Provider **Group B was the key driver of high provider turnover rate** (termination). Therefore this report recommends further investigation into the differences between Group B and the other provider groups to identify specific reasons via interviews and follow-up surveys of current and terminated providers from each provider group. In addition, the report identified higher termination rates for physicians who had not been at GoodRx for **longer than 200 days/~6 months**, and therefore recommends piloting a tiered promotion system to improve provider retention.

Exploratory Data Analysis

Reading the .xlsx spreadsheet into Jupyter Notebook, I saw that the dataset had 5 columns: Provider ID, Provider Group, Start Date, Termination Date, and Termination Reason. Quick informational analysis about the dataset revealed that there were 128 entries, no null entries, and data types datetime, integers, and objects.

Provider ID	Provider Group	Start Date	Termination Date	Termination Reason
ID1	Group A	2020-04-28	2021-05-17	Provider chose to leave
ID2	Group A	2019-09-16	2021-05-17	Medical error
ID3	Group A	2020-03-18	2021-05-17	Provider chose to leave
ID4	Group B	2020-02-27	2021-05-16	Not a good fit
ID5	Group B	2020-11-02	2021-05-16	Provider chose to leave

Figure 1. Reading the excel data.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 128 entries, 0 to 127
Data columns (total 6 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Provider ID         128 non-null    object
 1   Provider Group      128 non-null    object
 2   Start Date          128 non-null    datetime64[ns]
 3   Termination Date    128 non-null    datetime64[ns]
 4   Termination Reason  128 non-null    object
 5   days                128 non-null    int64
dtypes: datetime64[ns](2), int64(1), object(3)
memory usage: 6.1+ KB
```

Figure 2. Informational analysis of data.

The Provider Group column had three unique groups: Group A, Group B, and Group C. The earliest and latest Start Date and Termination Date were 2019-01-02 and 2021-04-20, 2019-03-03 and 2021-05-17, respectively. This confirmed that there were no abnormal Date outliers. Therefore, I moved on to creating a new column called “days” to calculate how long each provider was with GoodRx. The “days” column was created by subtracting the Start Date from the Termination Date. In this new column, there were three negative days (shown below).

Provider ID	Provider Group	Start Date	Termination Date	Termination Reason	days
ID147	Group B	2020-04-07	2020-04-01	Provider chose to leave	-6
ID158	Group A	2019-12-27	2019-12-01	Not a good fit	-26
ID177	Group A	2019-12-18	2019-03-03	Medical error	-290

Figure 3. The three negative “days” columns that will be dropped from further analysis

These three rows may have been data entry errors, so they were dropped from the analysis until further notification.

Visualizations of Data

Visualizations of the dataset were performed to identify trends within “days” (length of provider employment).

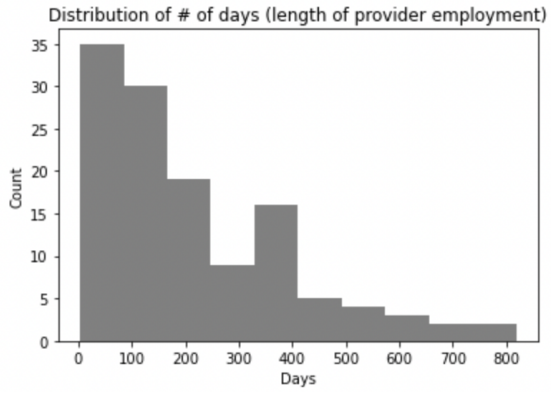

Figure 4. Histogram of number of “days”.

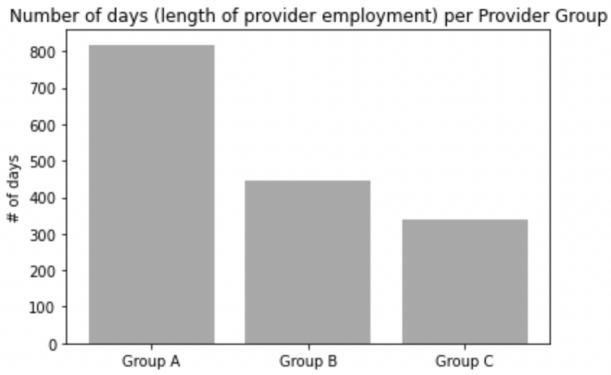

Figure 5. Bar chart of “days” per Provider Group.

The data shows that a large number of providers stay with GoodRx for about a year (365 days) or less. And those who stay for longer only make a small percentage (16.8% of providers stay for over a year). Segmenting this data into the three Provider Groups shows that Group A pulls the most days. This may be because Group A has the largest number of providers, but the data showed that Group A only had 36 providers, whereas Group B had 83 and Group C had 6.

Focus on High Turnover Group B

Group B had significantly more providers than Group A and Group C, but did not bring in similar numbers of days per provider. The turnover rate for Group B seems unusually high, therefore, I dove deeper into Group B’s data.

Subsetting the data to only Group B and creating a histogram on Termination Reason revealed that most providers in Group B (51.8%) left due to the providers choosing to leave. However, when comparing this phenomenon to the other Provider Groups, we see that the trends are similar across

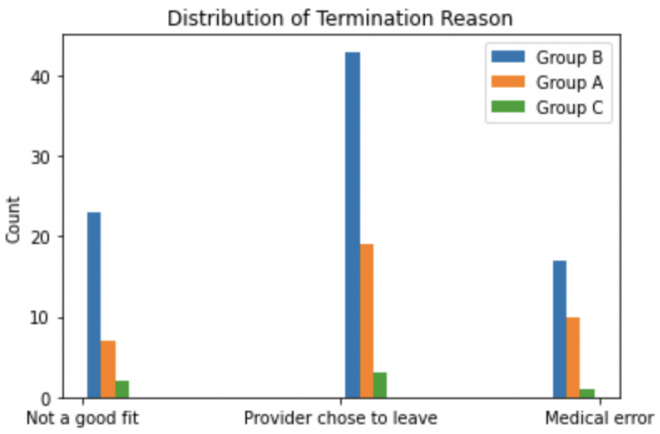

Figure 6. Histogram of Termination Reason segmented by Provider Groups.

the board. A majority of terminations due to providers choosing to leave is not a trend exclusive to Group B.

Therefore I looked further into the relationship between Termination Reason and days. From the number of counts alone, visualizations of this relationship showed that all three termination reasons were heavily positive/right skewed, indicating that the outliers are further out towards the right and closer to the mean on the left. In this case, the mean is greater than the median or mode – a lot of providers are terminated early (less than 200 days) in their partnership with GoodRx, regardless of termination reason. But once providers stayed with GoodRx longer than this time period, they tended to stay for longer.

But would this interpretation change when segmenting by each Provider Group? Comparing histograms of the relationship between Termination Reason and days for each Provider Group showed that Group B was the driver for the positive skew.

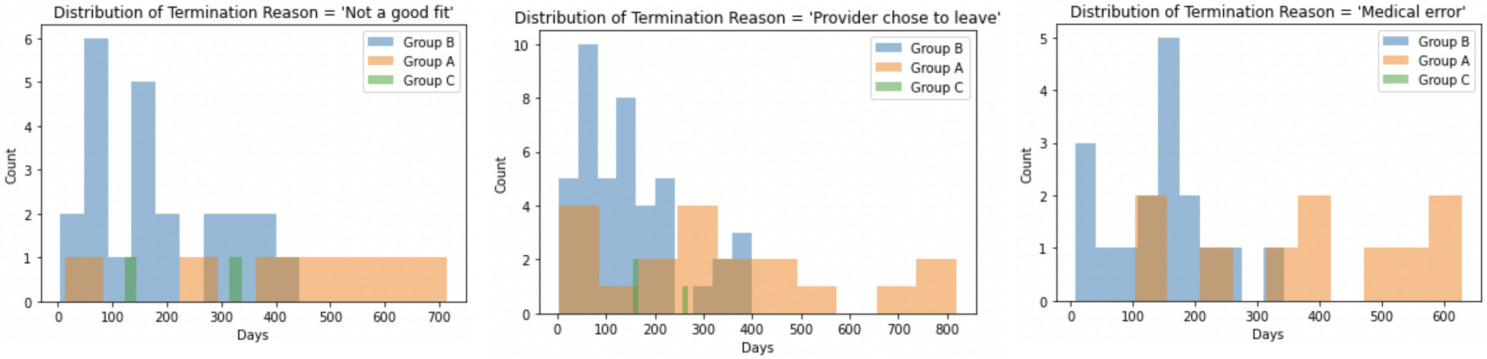

Figure 7. Histograms of each Termination Reason vs. Days (provider employment) segmented by Provider Groups.

Recommendations

From these data analyses, I would recommend looking further into the differences between Provider Group B and the other groups to identify reasons for Group B's high provider turnover rate (especially within the first ~6 months of partnership). Some additional data points may be collected through select interviews and follow-up surveys of current and terminated providers from each provider group. Also, I would look into Termination Reason "Not a good fit" to determine whether this type of termination was initiated by GoodRx or the provider.

In addition, I would recommend piloting a tiered promotion for providers to keep providers and decrease the risk of termination. The probability of termination decreases the longer providers are at GoodRx, therefore a tiered promotion system where providers receive XX if they stay for 6 months, XX if they stay for 1 year, etc. may help with provider retention.