Clipboard Health – Problems we tackle: Shift Offers Analysis

Analysis:

Before creating a problem statement, I conducted Exploratory Data Analysis* on the entire dataset. I created multiple tables and visualizations to understand the main characteristics of the dataset. A basic summary of the data is below:

Total workplaces	132
Total shift offers	266340
Total number of workers	10291
Total number of unique shifts created	19900
Total number of shifts claimed	13064
Total number of shifts deleted	3668
Total number of shifts cancelled	321
Total number of no-shows	31
Number of workers cancelled more than once	60

Table 1: Sorted by duration of shifts (0 to 18 hrs.)		
Most unique shifts created	8-hour shifts	13384
Most claimed shifts	8-hour shifts	9140
Most unclaimed shifts	8-hour shifts	4244
Most deleted shifts	8-hour shifts	2358
Most cancelled shifts	8-hour shifts	210

Table 2: Sorted by starting hours of shits (0:00 to 23:00)		
Most unique shifts created	22:00	4099
Most claimed shifts	22:00	2777
Most unclaimed shifts	22:00	1322
Most deleted shifts	22:00	776
Most cancelled shifts	22:00	73

Table 3: Sorted by slots (AM, PM & NOC)		
Most unique shifts created	PM	8218
Most claimed shifts	PM	5254
Most unclaimed shifts	PM	2964
Most deleted shifts	NOC	1726
Most cancelled shifts	PM	119

Table 4: Sorted by shifts created by workplaces		
Most unique shifts created	611af67795f4c501662edb31	1779
Most viewed shifts	611af67795f4c501662edb31	29671
Most claimed shifts	611af67795f4c501662edb31	980

(Please refer to the EDA sheet in the attached excel document)

Additionally, in Table 4, I created 2 more columns:
Views/Created Ratio – average views each created offer received
Views/Claimed Ratio – average views each claimed offer received

For both these ratios, Workplace ID **5e5e85ea43a2390016e3d65c** had the highest numbers:
Views/Created – 42 (*indicating high views rate)*
Views/Claimed – 90 *(indicating low claim rate)*

At this point, a few questions arose:

1. Why are there so many shifts with 0 durations and why are the deleted shifts with 0 hours so much higher than created?
2. Why are so many shifts deleted after they are viewed? Is it because workplaces are waiting too long for shifts to be claimed? What is the time between creation and viewing for the offers that were deleted?

To investigate this, I extracted data to compare various time averages before and after a shift is deleted or cancelled with respect to creating, viewing, claiming, and starting times. I created multiple combinations and produced heat maps to explore what the data is indicating*.

(Please refer to the Shifts Deleted sheet)

Highlights:

- 18.04% shifts (3591 out of 19900 unique shifts created) were deleted before they started
- 18.44% shifts (3670 out of 19900 unique shifts created) were deleted after they were viewed

As mentioned in the case document, workplaces can determine they no longer have a staffing need and delete the shift. Therefore, ~18% deletion in both cases does not stand out as an anomaly. Additionally, this could also support the theory that the 0 duration shifts may have been erroneous and hence were deleted.

- In the 'Shifts that are Claimed' heat map *(Figure. 6)*, offers are deleted 6.85 days on average after they are claimed, however, the start time for those shifts is 4.93 days. Similarly, in the 'Shifts that are Viewed' heat map *(Figure. 5)*, offers are deleted 4.06 days on average after they are viewed, however, the start time for those shifts is 3.29 days
- In the 'Shifts that are Viewed' heat map, offers are viewed 3.11 days on average after being created

From this analysis, I could come up with two problems to explore:

1. Why are offers deleted after they have crossed the start date? Are the workplaces conducting hygiene clean-up to remove redundant entries? If they are, why is the number of shifts deleted after starting only 80 and not significantly higher? Or did these offers not get claimed?
2. Why does it take so long for workers to view shift offers? Are there technical issues with notifications, or are there product-related reasons (e.g., poor matching algorithm, offers sent at inconvenient times)? Is there any correlation between charge rate and number or posts or timing of posts? Is there a correlation between higher pay rates and a faster claim time?

Since the number of deleted shifts after starting were only 80/19900 (~0.4%), whereas the number of claimed shifts were 13064/19900 (~65.65%) indicating that viewed shifts would have an equal percentage if not more. Hence, pursuing the first problem statement did not seem beneficial and I proceeded to solve for the second problem statement.

Problem Statement:

Improving Job Visibility and Worker Engagement: How can workplaces reduce the create-to-view time for new offers shown to workers?

Solution:

To begin solving this, I referred to the workplace data in Table 4. I noticed that workplaces with high viewership do not necessarily have a have a high claim rate. Alternatively, workplaces with low viewership do have a high claim rate.

I looked at the workplace with the highest Views/Created Ratio and Views/Claimed Ratio as a starting point. When I extracted its data*, I observed the following:

- Average duration of shifts and average pay was less than total average
- Average hour of day that shifts were claimed were significantly less than total average
- Start times of the shifts were less than total average
- Average time of shift creation was around midday (12.23 which is approximately 12:15 hrs.)
- Average time of shift was around early afternoon(14.17 which is approximately 14:10 hrs.)

These indicated a few things:

- Workers are most likely viewing and claiming new offers between 12:00 hrs. to 14:30 hrs.
- Lesser pay may have been a reason for offers getting low claims even with high views
- Workers may be inclined to shifts starting in the evenings

(Please refer to Sheet 4 for workplace summary and Figure.1 and Figure. 2 in EDA sheet for comparisons)

Since this was data from only one workplace, it now made sense to look at the data from all workplaces. Referring to Table. 4, I created a scatter plot *(Figure.3 in EDA sheet)* using Views/Created Ratio and Claimed/Created % (indicates what percentage of created shifts are being claimed) and distributed the companies into four quadrants* based on median values from these columns:

- Quadrant 1 (High Views – High Claims) – High performers
  Workplaces in this quadrant have high 'Views/Created' and high 'Claimed/Created %'
- Quadrant 2 (Low Views – High Claims)
  Workplaces in this quadrant have low 'Views/Created' and high 'Claimed/Created %'
- Quadrant 3 (Low Views – Low Claims) – Low performers
  Workplaces in this quadrant have low 'Views/Created' and low 'Claimed/Created %'
- Quadrant 4 (High Views – Low Claims)
  Workplaces in this quadrant have high 'Views/Created' and low 'Claimed/Created %'

(Please refer to Quadrant Analysis sheet)

After the categorization, I chose to focus on Quadrants 2 and 4* (each quadrant has 52 companies) as these workplaces represent unique performance profiles, so a comparative analysis can reveal valuable insights. The goal was to identify trends and outliers within each group.

I continued to work with the same metrics to compare workplaces. In Quadrant 2, the goal was to see which workplaces were the most successful despite low visibility. In Quadrant 4, the goal was to see which workplaces were struggling despite high visibility.

1. Views/Claimed
   a. For Quadrant 2 – Workplaces with exceptionally low ratio were the most efficient at converting views into claims
   b. For Quadrant 4 – Workplaces with exceptionally high ratio were the least efficient at converting views into claims
2. Claimed/Created %
   a. For Quadrant 2 – Workplaces with near 100% accuracy were most successful
   b. For Quadrant 4 - Workplaces with near 0% accuracy were least successful

After this broad categorization, I decided to make the sample size smaller by identifying the best and worst performers for each metric. I used median values to determine the thresholds for each metric. Additionally, I also chose to work with workplaces that have a positive match for both metrics, giving us the best-of-the-best in Quadrant 2 and worst-of-the-worst in Quadrant 4.

In Quadrant 2, a positive match will be the top performer:

- The best performers will have a above the median Claimed/Created % value
- The best performers will have a below the median Views/Claimed ratio

In Quadrant 4, a positive match will be the worst performer:

- The worst performers will have a below the median Claimed/Created % value
- The worst performers will have a above the median Views/Claimed ratio

After this classification, Quadrant 2 showed 16 workplaces which have both, above median high claims and below median low views and Quadrant 4 showed 17 workplaces which have both, below median low claims and above median high views.

(Please refer to Quadrant 2 and Quadrant 4 sheets)

Now was the time to deep dive and perform a root-cause analysis using other columns in the dataset, to identify potential reasons for their success of failure. I extracted the following additional details of each individual workplace across all its shifts to do this:

- Shifts created
- Most common slot

- Average duration
- Average pay rate
- Average charge rate
- Average created at hour
- Average viewed at hour (1st view for each unique worker)
- Average start hour
- Average time difference between offer viewed and claimed
- Average time difference between shift created and offer viewed (1st view for each unique worker)

Here are the highlights of this analysis:

Quadrant 2 (Low Views – High Claims)	
Shifts created	353
Average create-to-view time	1.44 hours
Average view-to-claim time	26.6 secs
Average created at hour	15.45
Average viewed at hour	13.91*
Average start hour	11.79
Average pay rate	$33.06
Average charge rate	$32.5
Average duration	8.71
Most Common Slot distribution	AM – 6 PM – 5 NOC – 5

Quadrant 4 (High Views – Low Claims)	
Shifts created	2747
Average create-to-view time	30.19 hours
Average view-to-claim time	53.2 secs
Average created at hour	13.94
Average viewed at hour	12.48*
Average start hour	15.57
Average pay rate	$23.51
Average charge rate	$32.71
Average duration	8.4
Most Common Slot distribution	AM – 6 PM – 10 NOC – 1

I also created histograms of a few metrics to isolate outliers and study general behavior.

As per the histograms for Quadrant 2:

- Window of highest shift creation is 16:00 – 18:00 hrs.
- Window of highest shift viewing is 14:00 – 16:00 hrs.
- Window of highest start hour of shifts is 11:00 – 13:00 hrs.
- Average create-to-view time is less than 2 hours

As per the histograms for Quadrant 4:

- Windows of highest shift creation are 13:00 – 15:00 hrs. and 17:00 – 19:00 hrs.
- Window of highest shift viewing is 14:00 – 16:00 hrs.
- Window of highest start hour of shifts is 14:50 – 16:50 hrs.
- Average create-to-view time is more than 2 hours

Conclusion:

1. Both quadrants have similar Average Duration values. This tells us that companies are targeting 8-hour shifts, which are claimed the most as seen in Table 1. This is corroborated by workers' data: Average Duration most viewed by workers is 8.46. *(Please refer to Workers Summary sheet)*
2. Both quadrants have similar Average Charge Rate. Since Clipboard Health charges a flat rate per hour to workplaces, similar Charge Rate numbers also make sense.
3. The view-to-claim time in Quadrant 2 is 26.6 seconds quicker than in Quadrant 4, indicating that workers possibly find those offers more attractive, hence the high claims.
4. The differences in average pay rates tell us that workers clearly prefer offers with higher-than-average pay rates, hence the high claims for Quadrant 2.
5. The average starting hour of shifts in Quadrant 2 is 11.79 (~12:00 hrs.), which is 3.5 hours before the average of shifts in Quadrant 4 (15.57 or ~15:30 hrs.). This is indicative that workers prefer shifts starting between 12:00 and 15:00 hrs. and can be supported by the data in Table 2.
6. There are three peaks in frequency of shift viewing across all workplaces, 04:00 – 06:00 hrs.,12:00 – 14:00 hrs., and 20:00 – 22:00 hrs. *(Figure. 2)*.
   a. Workplaces in Quadrant 4 have the highest windows of shift creation between 13:00 – 15:00 hrs. and 17:00 – 19:00 hrs. This ideal timing coupled with the high number of shifts created could explain why these offers have a high view rate.

b. Workplaces in Quadrant 2 have the highest windows of shift creation between 16:00 – 18:00 hrs. This off timing coupled with the low number of shifts created could explain why these offers have a low view rate.

7. Referring to the screenshot of the shift offer provided in the case document, four primary points are visible to workers: Duration, Pay Rate, Role, and Workplace. Here, pay rate stands out as an important contributing factor while viewing a shift. The low average pay rate of Quadrant 4 could be a reason why the create-to-view time is so high.

For both quadrants, average viewed at hour is less than average created at hour. Although the average create-to-view time for Quadrant 4 is 30.19 hours and hence this does not seem like an odd metric, the average create-to-view time for Quadrant 2 is 1.44 hours which shows there is something inconsistent with the data or there are significant outliers which are skewing the analysis.

Next Steps:

The analysis above is highly indicative but not conclusive. The next steps will be to answer the following questions to move from correlation to causation:

- What is the user journey a worker goes through when viewing shifts?
- Which workplaces lie in Quadrant 4? Does the reputation of a workplace play any role in claiming shifts?
- Do workers consider traveling distance as an important factor when looking at workplaces?
- If the data indicates that workers view offers between 12:00 – 14:00 hrs. (most likely at the end of AM shift or during lunch breaks), why do workplaces post offers in odd times? Are there any benefits in not creating shifts during those hours, or is it general strategy to post requirements as they come up? Does Clipboard Health's charge rates also include time of day of creating shifts in addition to duration?