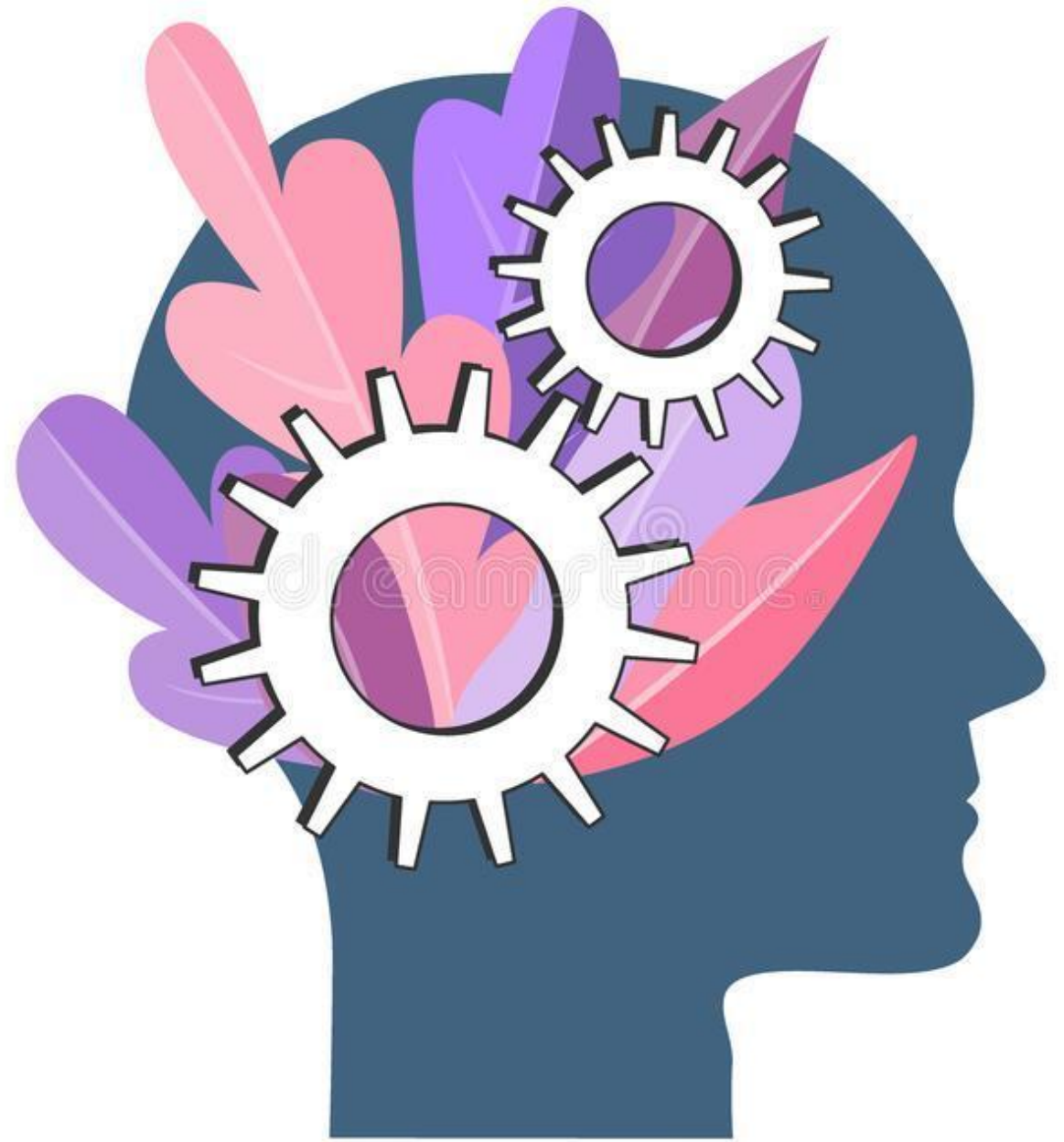


# Work Culture Project

Feb 24 2023

**Presenter:**

Mitanshi Vyas (MS Student)



# Overview

- Focused on getting accurate results for Case 2
- Case 2 is where a worker has attempted and finished the task within the first attempt
- Created more data to evaluate the Case 2 to make sure the results are consistent
- Worked towards getting accurate results for Case 3
- Case 3 is where a worker has attempted a task multiple times



# Snippet of the Mini Log Data

	id	uid	event	timestamp	url	tab_id	scroll_count	blur_count	focus_count
0	191654	1685129673	TABUPDATED	1676579992909	https://toloka.yandex.com/signup	556	0.0	0.0	0.0
1	191655	1685129673	TABUPDATED	1676579994735	https://toloka.yandex.com/	556	0.0	0.0	0.0
2	191656	1685129673	WEBREQUEST	1676579994914	https://toloka.yandex.com/	556	NaN	NaN	NaN
3	191657	1685129673	TABUPDATED	1676579995999	https://toloka.yandex.com/tasks	556	0.0	2.0	2.0
4	191658	1685129673	WEBREQUEST	1676579996060	https://toloka.yandex.com/tasks	556	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...
155	192180	1685129673	TABACTIVATED	1676585897582	https://toloka.yandex.com/task/37226291/000238...	590	0.0	1.0	1.0
156	192181	1685129673	TABACTIVATED	1676585899191	https://docs.google.com/document/d/1vouWd54ln-...	530	NaN	NaN	NaN
157	192186	1685129673	TABACTIVATED	1676585925769	https://toloka.yandex.com/task/37226291/000238...	590	0.0	0.0	1.0
158	192187	1685129673	TABCLOSED	1676585929544	https://toloka.yandex.com/task/37226291/000238...	590	NaN	NaN	NaN
159	192188	1685129673	TABACTIVATED	1676585929556	https://northeastern-csm.symplicity.com/studen...	560	NaN	NaN	NaN

# Last Week

**Function Name:** `task_duration(pool_id)`

**CASE 1:** Task was attempted just once

```
[ ] task_duration('36703863')
```

The time took for the given task to get completed is: **6.441666666666666 minutes.**

Original duration is 7 mins.

```
[ ] task_duration('37062507')
```

The time took for the given task to get completed is: **8.984466666666666 minutes.**

Original duration is 9 minutes.

# Inaccurate Results for Case 2

**Function Name:** `task_duration(pool_id)`

**CASE 2:** Task was attempted and finished

```
[ ] task_duration('37225409')
```

The time took for the given task to get completed is: **2.2997 minutes.**

Original duration is 8 mins.

Our old method does not produce very accurate results for methods that are completed.

# Inaccurate Results for Case 2

**Function Name:** `task_duration(pool_id)`

```
[30] task_duration('37237050')
```

The time took for the given task to get completed is: **0.9453833333333334 minutes.**

Original duration of the task is 5 minutes.

```
[31] task_duration('37238015')
```

The time took for the given task to get completed is: **2.9362666666666666 minutes.**

Original duration of the task is 7 minutes.

---

# Modifying the Method

Interesting insight I figured from the log data to help improve the accuracy:

- When the worker finishes the task, the tab doesn't stay the same until we close it
- The tab goes back to the Task browsing tab in Toloka
- Also because it is not a new tab, the tab 'event' is recorded as 'TABUPDATED'
- So considering these two conditions I was able to grab accurate timestamps
- I grabbed the timestamp corresponding to the Toloka task tab URL and where the event is also recorded as 'TABUPDATED'
- The third condition used: the pool\_id of the index right before the timestamp must be pool\_id of the task that we are inquiring for.

# Final Results

**CASE 2:** Task was attempted and finished in a single attempt

```
[213] task_duration_modified(37225409)
```

```
1685129673
```

```
The time took for the given task to get completed is: 7.45195 minutes.
```

Original duration of the task is 8 minutes.

```
[95] task_duration_modified(37237050)
```

```
The time took for the given task to get completed is: 5.418966666666667 minutes.
```

Original duration of the task is 5 minutes.



# Final Results

**CASE 2:** Task was attempted and finished in a single attempt

```
[96] task_duration_modified(37238015)
```

```
The time took for the given task to get completed is: 7.155416666666665 minutes.
```

```
Original duration of the task is 7 minutes.
```

Finally, we are obtaining accurate results for Case 2. So, the Case 2 is resolved.

# Case 3 Previous Results

**Function Name:** `task_duration(pool_id)`

**CASE 3:** Task was attempted in multiple attempts

```
[ ] task_duration('37314992')
```

The time took for the given task to get completed is: **23.507483333333333 minutes.**

Original time took is 14 minutes.

The duration calculated by our old method is significantly larger than the actual duration, hence it indicates that the function is probably not excluding the time for which the worker isn't working on the task after initiating it.

# Final Results for Case 3

**Function Name:**

`task_duration_case3(user_id)`

```
[212] task_duration_case3(37314992)
```

```
The time took for the given task to get completed is: 15.3256 minutes.
```

Original duration of the task is 14 minutes.

The method does work for all three cases now!

# Appendix

## Python Scripts:

<https://colab.research.google.com/drive/1YfOclKYLtXxW2mifYwOf5ypUFTKOeld9?usp=sharing>

## Toloka Mini Data Activity Tracker:

<https://docs.google.com/document/d/1vouWd54ln-vl2mkF7VR6yjlR3VPxEFnH1Gf8877-lHg/edit?usp=sharing>

**Thank You!**

