

## Exhibit A.

Given tab delimited input file with four columns with logs from music streaming portal:

- PLAY\_ID – unique 16-byte identifier given in hexadecimal string
- SONG\_ID – Identifier of a song that a client played
- CLIENT\_ID – Identifier of a client
- PLAY\_TS – date time of the client playing a song (can be just date, or date and time)

Output required is a table that shows the distribution distinct song play counts per user on date of August 10, 2016, as per below example.

**Example.** input data is below table (highlighted lines with PLAY\_TS on August 10, 2016 with yellow background):

PLAY_ID	SONG_ID	CLIENT_ID	PLAY_TS
44BB190BC2493964E053CF0A000AB546	6164	1	09/08/2016 09:16:00
44BB190BC24A3964E053CF0A000AB546	544	3	10/08/2016 13:54:00
44BB190BC24B3964E053CF0A000AB546	9648	3	08/08/2016 06:08:00
44BB190BC24C3964E053CF0A000AB546	7565	2	10/08/2016 17:30:00
44BB190BC24D3964E053CF0A000AB546	8995	1	11/08/2016 02:40:00
44BB190BC24E3964E053CF0A000AB546	4407	1	08/08/2016 07:30:00
44BB190BC24F3964E053CF0A000AB546	5839	2	10/08/2016 02:40:00
44BB190BC2503964E053CF0A000AB546	548	3	09/08/2016 20:45:00
44BB190BC2513964E053CF0A000AB546	376	3	10/08/2016 04:57:00
44BB190BC2523964E053CF0A000AB546	3403	2	08/08/2016 21:14:00
44BB190BC2533964E053CF0A000AB546	7256	2	10/08/2016 06:29:00
44BB190BC2543964E053CF0A000AB546	4291	3	08/08/2016 09:26:00
44BB190BC2553964E053CF0A000AB546	5722	2	08/08/2016 23:33:00
44BB190BC2563964E053CF0A000AB546	9857	1	10/08/2016 22:05:00
44BB190BC2573964E053CF0A000AB546	3122	2	09/08/2016 08:35:00
44BB190BC2583964E053CF0A000AB546	217	2	10/08/2016 13:20:00
44BB190BC2593964E053CF0A000AB546	3022	1	10/08/2016 17:06:00
44BB190BC25A3964E053CF0A000AB546	9857	1	10/08/2016 15:06:00
44BB190BC25B3964E053CF0A000AB546	2168	3	11/08/2016 13:30:33

As it can be seen from above table, client 1 played songs 3 times during 10/08/2016, but 2 of them were same, so the client 1 played 2 distinct songs (9857 and 3022). Client 2 played 4 distinct songs, client 3 played 2 distinct songs. In overall there are 2 clients that played 2 distinct songs, and 1 client that played 4 distinct songs. Output result expected is as below (can be tab delimited file or comma delimited file)

DISTINCT_PLAY_COUNT	CLIENT_COUNT
2	2
4	1

Please use below link to get input file for assignment:

[https://piworks.sharepoint.com/hiring/\\_layouts/15/guestaccess.aspx?guestaccesstoken=2Y5ZUTQGncrK3PW9PvAyhXf1w9TCrWBxhlgx14Hhkiw%3d&docid=0ec903281584e4dbeb8b65e3ea055ebe3&rev=1](https://piworks.sharepoint.com/hiring/_layouts/15/guestaccess.aspx?guestaccesstoken=2Y5ZUTQGncrK3PW9PvAyhXf1w9TCrWBxhlgx14Hhkiw%3d&docid=0ec903281584e4dbeb8b65e3ea055ebe3&rev=1)