

Analysis of VDI System Usage in 2015

Prepared by:
Munem Shariar Shams
Independent Researcher and Analyst
Central Michigan University

Executive Summary

This report provides a comprehensive analysis of the Virtual Desktop Infrastructure (VDI) usage for the year 2015, focusing exclusively on VDI machines. The analysis reveals a robustly used system with 3,013 users, with clear patterns of academic engagement. Daily usage averaged 139 users, peaking at 540 on April 28th. The average number however was very different from the median number of users which was 62. The reason behind this is because from the months May till the end of August the school activities decline due to summer break. The usage of VDI was much higher during the regular school semester.

Crucially, this analysis is framed by significant data quality issues. Approximately 16% of application sessions were either incomplete or had invalid application stop times, and 5.88% CPU usage data contains impossible negative values, affecting the accuracy of application duration and CPU usage metrics. The data set is also missing data for the fall semester of the academic year and only contains data covering January to August. Despite these issues, the data reliably reveals strong seasonal and weekly usage patterns tied to the academic calendar, providing valuable insights for strategic planning.

1. Data Preparation and Quality Assessment

Prior to analysis, the server logs and application statistics datasets were joined and rigorously examined. The following key steps and data quality findings form the foundation for all subsequent insights.

Data Processing Steps

The raw data from two separate CSV files was loaded and joined to create a comprehensive usage dataset. A critical step involved standardizing date formats; the original two-digit year format ("4/15/11") was converted to a proper four-digit format ("2011-04-15") to ensure accurate filtering for the year 2015.

32	CMULABA55	userid1	2011-04-15 16:36:00	4/15/11 16:40					NA	16	WINWORD
35	CMULABA34	userid4	2011-04-15 16:58:00	4/15/11 17:01					NA	17	iexplore
38	CMULIB11	userid6	2011-04-15 18:19:00	4/15/11 18:35					NA	18	iexplore
39	CMULABA63	userid7	2011-04-15 20:07:00	4/16/11 0:15					NA	19	scrnssave.scr
39	CMULABA63	userid7	2011-04-15 20:07:00	4/16/11 0:15					NA	20	scrnssave.scr
39	CMULABA63	userid7	2011-04-15 20:07:00	4/16/11 0:15					NA	21	firefox
39	CMULABA63	userid7	2011-04-15 20:07:00	4/16/11 0:15					NA	22	firefox
39	CMULABA63	userid7	2011-04-15 20:07:00	4/16/11 0:15					NA	23	WinZip
39	CMULABA63	userid7	2011-04-15 20:07:00	4/16/11 0:15					NA	24	VISIO
39	CMULABA63	userid7	2011-04-15 20:07:00	4/16/11 0:15					NA	25	scrnssave.scr
39	CMULABA63	userid7	2011-04-15 20:07:00	4/16/11 0:15					NA	26	explorer
42	CMULABA8	userid10	2011-04-16 12:04:00	4/16/11 12:29					NA	27	EXCEL
43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	41	devenv
43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	42	coffee order.vhost
43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	43	coffee order.vhost
43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	44	coffee order.vhost
43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	45	explorer
43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	46	scrnssave.scr
43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	47	scrnssave.scr
43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	48	scrnssave.scr
43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	49	devenv
43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	50	firefox
43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	51	firefox
43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	52	Adobe Reader X

Figure 1.1: Wrong Date format

1	779026	CMUVDI17	userid23004	2015-01-07 13:42:00	2015-01-07 13:50:00	https://virtuallab.cmich.edu:443	10.137.93.52	Windows	10	2883489	WINWORD	2015-
2	779027	CMUVDI6	userid580	2015-01-07 14:51:00	2015-01-07 16:02:00	unknown	unknown	unknown	7	2883492	cmd	2015-
3	779027	CMUVDI6	userid580	2015-01-07 14:51:00	2015-01-07 16:02:00	unknown	unknown	unknown	7	2883493	WINWORD	2015-
4	779027	CMUVDI6	userid580	2015-01-07 14:51:00	2015-01-07 16:02:00	unknown	unknown	unknown	7	2883494	firefox	2015-
5	779027	CMUVDI6	userid580	2015-01-07 14:51:00	2015-01-07 16:02:00	unknown	unknown	unknown	7	2883495	explorer	2015-
6	779027	CMUVDI6	userid580	2015-01-07 14:51:00	2015-01-07 16:02:00	unknown	unknown	unknown	7	2883496	WINWORD	2015-
7	779027	CMUVDI6	userid580	2015-01-07 14:51:00	2015-01-07 16:02:00	unknown	unknown	unknown	7	2883497	WINWORD	2015-
8	779027	CMUVDI6	userid580	2015-01-07 14:51:00	2015-01-07 16:02:00	unknown	unknown	unknown	7	2883498	WINWORD	2015-
9	779028	CMUVDI18	userid23004	2015-01-07 14:01:00	2015-01-07 15:13:00	https://virtuallab.cmich.edu:443	10.137.93.52	Windows	5	2883499	WINWORD	2015-
0	779029	CMUVDI3	userid18286	2015-01-07 15:11:00	2015-01-07 15:26:00	https://virtuallab.cmich.edu:443	192.168.1.148	Windows	3	2883490	cmd	2015-
1	779029	CMUVDI3	userid18286	2015-01-07 15:11:00	2015-01-07 15:26:00	https://virtuallab.cmich.edu:443	192.168.1.148	Windows	3	2883491	dllhost	2015-
2	779030	CMU-LIB9	userid23439	2015-01-07 15:13:00	2015-01-07 16:59:00	https://libdesktop.vdi.cmich.edu:443	141.209.29.64	Windows	18	2883500	firefox	2015-
3	779030	CMU-LIB9	userid23439	2015-01-07 15:13:00	2015-01-07 16:59:00	https://libdesktop.vdi.cmich.edu:443	141.209.29.64	Windows	18	2883501	plugin-hang-ui	2015-
4	779030	CMU-LIB9	userid23439	2015-01-07 15:13:00	2015-01-07 16:59:00	https://libdesktop.vdi.cmich.edu:443	141.209.29.64	Windows	18	2883502	WINWORD	2015-
5	779030	CMU-LIB9	userid23439	2015-01-07 15:13:00	2015-01-07 16:59:00	https://libdesktop.vdi.cmich.edu:443	141.209.29.64	Windows	18	2883503	explorer	2015-
6	779030	CMU-LIB9	userid23439	2015-01-07 15:13:00	2015-01-07 16:59:00	https://libdesktop.vdi.cmich.edu:443	141.209.29.64	Windows	18	2883504	rundll32	2015-
7	779030	CMU-LIB9	userid23439	2015-01-07 15:13:00	2015-01-07 16:59:00	https://libdesktop.vdi.cmich.edu:443	141.209.29.64	Windows	18	2883505	iexplore	2015-
8	779030	CMU-LIB9	userid23439	2015-01-07 15:13:00	2015-01-07 16:59:00	https://libdesktop.vdi.cmich.edu:443	141.209.29.64	Windows	18	2883506	rundll32	2015-
9	779030	CMU-LIB9	userid23439	2015-01-07 15:13:00	2015-01-07 16:59:00	https://libdesktop.vdi.cmich.edu:443	141.209.29.64	Windows	18	2883507	iexplore	2015-
0	779030	CMU-LIB9	userid23439	2015-01-07 15:13:00	2015-01-07 16:59:00	https://libdesktop.vdi.cmich.edu:443	141.209.29.64	Windows	18	2883508	WINWORD	2015-
1	779030	CMU-LIB9	userid23439	2015-01-07 15:13:00	2015-01-07 16:59:00	https://libdesktop.vdi.cmich.edu:443	141.209.29.64	Windows	18	2883509	MSACCESS	2015-
2	779030	CMU-LIB9	userid23439	2015-01-07 15:13:00	2015-01-07 16:59:00	https://libdesktop.vdi.cmich.edu:443	141.209.29.64	Windows	18	2883510	WINWORD	2015-
3	779030	CMU-LIB9	userid23439	2015-01-07 15:13:00	2015-01-07 16:59:00	https://libdesktop.vdi.cmich.edu:443	141.209.29.64	Windows	18	2883511	AcroRd32	2015-
4	779030	CMU-LIB9	userid23439	2015-01-07 15:13:00	2015-01-07 16:59:00	https://libdesktop.vdi.cmich.edu:443	141.209.29.64	Windows	18	2883512	AcroRd32	2015-

Figure 1.2: Fixed Date format

Applying Necessary Filters

The dataset was filtered to include only records from the year 2015 and from VDI machines whose names start with "CMU". At first we printed a summary output to find out that there are 299,998 records in the year 2015. After applying the filter to only find entries from VDI machines, the dataset that was used for our final analysis contained 227,386 records spanning from January to August 2015, representing 3,013 unique users.

Addressing Duplicate User IDs

A important consideration in this analysis was handling duplicate user entries. When a user logs into a VDI session and launches multiple applications, the dataset creates multiple records with the same user ID, login time, and logout time, but different application information. To ensure accurate user counts, the n_distinct function was used, which correctly counts unique users regardless of how many applications they used during their session.

These application-specific records are necessary for analyzing application usage patterns and should not be filtered out arbitrarily. The goal of applying filters is to make analysis more accurate and relevant, not simply to reduce dataset size.

4	39	CMULABA63	userid7	2011-04-15 20:07:00	4/16/11 0:15					NA	19	scrnsave.scr
5	39	CMULABA63	userid7	2011-04-15 20:07:00	4/16/11 0:15					NA	20	scrnsave.scr
6	39	CMULABA63	userid7	2011-04-15 20:07:00	4/16/11 0:15					NA	21	firefox
7	39	CMULABA63	userid7	2011-04-15 20:07:00	4/16/11 0:15					NA	22	firefox
8	39	CMULABA63	userid7	2011-04-15 20:07:00	4/16/11 0:15					NA	23	WinZip
9	39	CMULABA63	userid7	2011-04-15 20:07:00	4/16/11 0:15					NA	24	VISIO
10	39	CMULABA63	userid7	2011-04-15 20:07:00	4/16/11 0:15					NA	25	scrnsave.scr
11	39	CMULABA63	userid7	2011-04-15 20:07:00	4/16/11 0:15					NA	26	explorer
13	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	41	devenv
14	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	42	coffee order.vhost
15	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	43	coffee order.vhost
16	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	44	coffee order.vhost
17	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	45	explorer
18	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	46	scrnsave.scr
19	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	47	scrnsave.scr
20	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	48	scrnsave.scr
21	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	49	devenv
22	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	50	firefox
23	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	51	firefox
24	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	52	Adobe Reader X
25	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	53	WindowsApplication1.vhost
26	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	54	MSACCESS
27	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	55	WindowsApplication1.vhost
28	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	56	WindowsApplication1.vhost
29	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	57	WindowsApplication1.vhost
30	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	58	WindowsApplication1.vhost
31	43	CMULIB8	userid11	2011-04-16 13:45:00	4/16/11 16:59					NA	59	scrnsave.scr

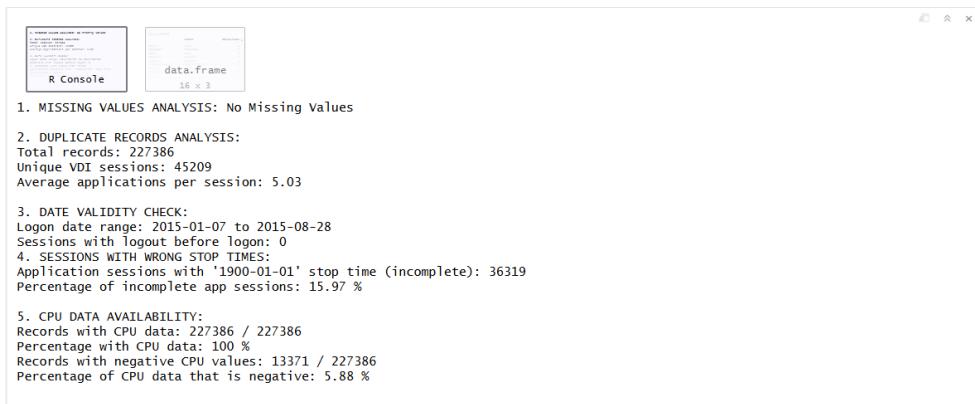
Figure 1.3: Example of duplicate User ID entries

Data Quality Evaluation

A thorough quality check was conducted on the filtered dataset. The key findings are summarized below:

The Positives:

- No Missing Values:** There are no missing values in any columns of the filtered dataset.
- Logical Session Times:** No sessions were found where the logout time occurred before the login time.
- Good Temporal Coverage:** The data provides good coverage from January to August 2015.



```

R Console
data.frame
16 x 3

1. MISSING VALUES ANALYSIS: No Missing Values
2. DUPLICATE RECORDS ANALYSIS:
Total records: 227386
Unique VDI sessions: 45209
Average applications per session: 5.03

3. DATE VALIDITY CHECK:
Logon date range: 2015-01-07 to 2015-08-28
Sessions with logout before logon: 0
4. SESSIONS WITH WRONG STOP TIMES:
Application sessions with '1900-01-01' stop time (incomplete): 36319
Percentage of incomplete app sessions: 15.97 %

5. CPU DATA AVAILABILITY:
Records with CPU data: 227386 / 227386
Percentage with CPU data: 100 %
Records with negative CPU values: 13371 / 227386
Percentage of CPU data that is negative: 5.88 %

```

Figure 1.4: Summary of Data Quality Check

The Negatives (Data Caveats):

- **Wrong Stop Time entries for Applications:** 36,319 records (15.97%) have an application stop time of "1900-01-01", indicating they were not properly closed or there was an error in data entry. This prevents accurate calculation of application usage duration and affects any analysis of how long applications were used.

logout_DTS	connection_server	remote_ip	remote_os	avg_cpu	app_id	app_name	start	stop	logon_date
8/27/15 22:21	https://virtuallab.cmich.edu:443	192.168.43.236	Windows	7	3188475	explorer	2015-08-27 20:26:00	1900-01-01 00:00:00	2015-08-28
8/27/15 21:15	https://virtuallab.cmich.edu:443	192.168.2.4	Mac	5	3188463	iexplore	2015-08-27 20:31:00	1900-01-01 00:00:00	2015-08-28
8/27/15 22:57	https://virtuallab.cmich.edu:443	192.168.2.6	Windows	6	3188485	firefox	2015-08-27 21:15:00	1900-01-01 00:00:00	2015-08-28
8/27/15 22:57	https://virtuallab.cmich.edu:443	192.168.2.6	Windows	6	3188487	explorer	2015-08-27 22:07:00	1900-01-01 00:00:00	2015-08-28
8/27/15 21:12	https://virtuallab.cmich.edu:443	141.209.228.153	Mac	28	3188461	firefox	2015-08-27 21:06:00	1900-01-01 00:00:00	2015-08-28
8/27/15 22:04	https://virtuallab.cmich.edu:443	107.181.196.21	Mac	7	3188469	AcroRd32	2015-08-27 21:58:00	1900-01-01 00:00:00	2015-08-28
8/27/15 22:09	https://virtuallab.cmich.edu:443	192.168.1.2	Windows	7	3188470	EXCEL	2015-08-27 22:05:00	1900-01-01 00:00:00	2015-08-28
8/28/15 8:04	https://virtuallab.cmich.edu:443	35.32.226.96	Windows	-2	3188495	EXCEL	2015-08-27 22:16:00	1900-01-01 00:00:00	2015-08-28
8/27/15 22:38	https://virtuallab.cmich.edu:443	192.168.1.97	Windows	3	3188484	saplogon	2015-08-27 22:25:00	1900-01-01 00:00:00	2015-08-28
8/28/15 3:05	https://virtuallab.cmich.edu:443	192.168.2.6	Mac	3	3188492	firefox	2015-08-28 00:54:00	1900-01-01 00:00:00	2015-08-28
8/28/15 12:41	https://virtuallab.cmich.edu:443	192.168.1.72	Windows	2	3188507	stats	2015-08-28 10:15:00	1900-01-01 00:00:00	2015-08-28
8/28/15 12:25	https://virtuallab.cmich.edu:443	35.32.232.90	Mac	6	3188506	iexplore	2015-08-28 12:18:00	1900-01-01 00:00:00	2015-08-28
8/28/15 14:09	https://virtuallab.cmich.edu:443	192.168.1.148	Windows	2	3188508	wssm	2015-08-28 13:48:00	1900-01-01 00:00:00	2015-08-28
8/27/15 5:03	https://virtuallab.cmich.edu:443	192.168.1.4	Mac	1	3188354	firefox	2015-08-27 04:48:00	1900-01-01 00:00:00	2015-08-27
8/27/15 4:04	https://virtuallab.cmich.edu:443	192.168.0.15	Windows	-2	3188387	VISIO	2015-08-26 20:43:00	1900-01-01 00:00:00	2015-08-27
8/27/15 4:04	https://virtuallab.cmich.edu:443	192.168.0.15	Windows	-2	3188389	explorer	2015-08-27 01:59:00	1900-01-01 00:00:00	2015-08-27
8/26/15 23:31	https://virtuallab.cmich.edu:443	10.0.0.16	Mac	4	3188365	devenv	2015-08-26 21:37:00	1900-01-01 00:00:00	2015-08-27
8/27/15 11:09	https://virtuallab.cmich.edu:443	10.20.58.26	Windows	7	3188416	explorer	2015-08-27 09:55:00	1900-01-01 00:00:00	2015-08-27
8/27/15 10:18	https://virtuallab.cmich.edu:443	141.209.228.153	Mac	15	3188406	POWERPNT	2015-08-27 10:12:00	1900-01-01 00:00:00	2015-08-27
8/27/15 10:18	https://virtuallab.cmich.edu:443	141.209.228.153	Mac	15	3188407	firefox	2015-08-27 10:15:00	1900-01-01 00:00:00	2015-08-27
8/27/15 13:09	https://virtuallab.cmich.edu:443	192.168.1.116	Windows	1	3188426	POWERPNT	2015-08-27 11:11:00	1900-01-01 00:00:00	2015-08-27
8/27/15 17:41	https://virtuallab.cmich.edu:443	10.20.58.26	Windows	2	3188443	explorer	2015-08-27 11:20:00	1900-01-01 00:00:00	2015-08-27
8/27/15 17:41	https://virtuallab.cmich.edu:443	10.20.58.26	Windows	2	3188446	EXCEL	2015-08-27 12:50:00	1900-01-01 00:00:00	2015-08-27
8/27/15 12:05	https://libdesktop.vdi.cmich.edu:443	141.209.34.169	Windows	7	3188419	explorer	2015-08-27 11:51:00	1900-01-01 00:00:00	2015-08-27
8/27/15 12:05	https://libdesktop.vdi.cmich.edu:443	141.209.34.169	Windows	7	3188422	AcroRd32	2015-08-27 12:01:00	1900-01-01 00:00:00	2015-08-27

Figure 1.4: Wrong Stop time data for applications

- Invalid CPU Usage Data:** The avg_cpu column contains negative values, which is impossible for CPU utilization percentages. 5.88% of the entries in average CPU usage are negative. This questions the credibility of the data and severely affects the reliability of any CPU Usage related analysis which we will be conducting in this project.

	VDI_ID	comp_name	userid	logon_DTS	logout_DTS	connection_server	remote_ip	remote_od	avg_cpu	app_id	app_name
187685	846470	CMUVDI31	userid15567	2015-05-12 10:23:00	5/15/15 9:58	https://virtuallab.cmich.edu:443	141.209.230.50	Mac	-256	314_218	jmp
15302	787465	CMU-LAB38	userid23598	2015-01-26 15:19:00	1/27/15 1:22	https://labdesktop.cmich.edu:443	141.209.11.195	Windows	-64	291_383	Xcelsius
15303	787465	CMU-LAB38	userid23598	2015-01-26 15:19:00	1/27/15 1:22	https://labdesktop.cmich.edu:443	141.209.11.195	Windows	-64	291_384	EXCEL
15304	787465	CMU-LAB38	userid23598	2015-01-26 15:19:00	1/27/15 1:22	https://labdesktop.cmich.edu:443	141.209.11.195	Windows	-64	291_385	AcroRd32
15305	787465	CMU-LAB38	userid23598	2015-01-26 15:19:00	1/27/15 1:22	https://labdesktop.cmich.edu:443	141.209.11.195	Windows	-64	291_386	AcroRd32
176591	844412	CMUVDI50	userid18334	2015-05-01 12:00:00	5/2/15 8:02	https://virtuallab.cmich.edu:443	35.32.220.133	Windows	-60	313_808	WINWORD
176592	844412	CMUVDI50	userid18334	2015-05-01 12:00:00	5/2/15 8:02	https://virtuallab.cmich.edu:443	35.32.220.133	Windows	-60	313_809	firefox
176593	844412	CMUVDI50	userid18334	2015-05-01 12:00:00	5/2/15 8:02	https://virtuallab.cmich.edu:443	35.32.220.133	Windows	-60	313_810	AcroRd32
176594	844412	CMUVDI50	userid18334	2015-05-01 12:00:00	5/2/15 8:02	https://virtuallab.cmich.edu:443	35.32.220.133	Windows	-60	313_811	SnippingTool
176595	844412	CMUVDI50	userid18334	2015-05-01 12:00:00	5/2/15 8:02	https://virtuallab.cmich.edu:443	35.32.220.133	Windows	-60	313_812	explorer
160220	840987	CMUVDI9	userid25941	2015-04-26 16:54:00	4/27/15 7:11	https://virtuallab.cmich.edu:443	35.32.245.138	Mac	-57	312_000	firefox
160221	840987	CMUVDI9	userid25941	2015-04-26 16:54:00	4/27/15 7:11	https://virtuallab.cmich.edu:443	35.32.245.138	Mac	-57	312_001	EXCEL
160222	840987	CMUVDI9	userid25941	2015-04-26 16:54:00	4/27/15 7:11	https://virtuallab.cmich.edu:443	35.32.245.138	Mac	-57	312_002	EXCEL
160223	840987	CMUVDI9	userid25941	2015-04-26 16:54:00	4/27/15 7:11	https://virtuallab.cmich.edu:443	35.32.245.138	Mac	-57	312_003	ieExplore
160224	840987	CMUVDI9	userid25941	2015-04-26 16:54:00	4/27/15 7:11	https://virtuallab.cmich.edu:443	35.32.245.138	Mac	-57	312_004	explorer
160225	840987	CMUVDI9	userid25941	2015-04-26 16:54:00	4/27/15 7:11	https://virtuallab.cmich.edu:443	35.32.245.138	Mac	-57	312_005	CLVIEW
177746	844618	CMUVDI43	userid18334	2015-05-02 09:19:00	5/3/15 2:09	https://virtuallab.cmich.edu:443	35.32.220.133	Windows	-34	313_521	AcroRd32
177747	844618	CMUVDI43	userid18334	2015-05-02 09:19:00	5/3/15 2:09	https://virtuallab.cmich.edu:443	35.32.220.133	Windows	-34	313_522	notepad
177748	844618	CMUVDI43	userid18334	2015-05-02 09:19:00	5/3/15 2:09	https://virtuallab.cmich.edu:443	35.32.220.133	Windows	-34	313_523	firefox
177749	844618	CMUVDI43	userid18334	2015-05-02 09:19:00	5/3/15 2:09	https://virtuallab.cmich.edu:443	35.32.220.133	Windows	-34	313_524	WINWORD
177750	844618	CMUVDI43	userid18334	2015-05-02 09:19:00	5/3/15 2:09	https://virtuallab.cmich.edu:443	35.32.220.133	Windows	-34	313_525	saplogon
177751	844618	CMUVDI43	userid18334	2015-05-02 09:19:00	5/3/15 2:09	https://virtuallab.cmich.edu:443	35.32.220.133	Windows	-34	313_526	explorer
177752	844618	CMUVDI43	userid18334	2015-05-02 09:19:00	5/3/15 2:09	https://virtuallab.cmich.edu:443	35.32.220.133	Windows	-34	313_527	SnippingTool
177753	844618	CMUVDI43	userid18334	2015-05-02 09:19:00	5/3/15 2:09	https://virtuallab.cmich.edu:443	35.32.220.133	Windows	-34	313_528	EXCEL
177754	844618	CMUVDI43	userid18334	2015-05-02 09:19:00	5/3/15 2:09	https://virtuallab.cmich.edu:443	35.32.220.133	Windows	-34	313_529	SnippingTool

Figure 1.5: Wrong CPU Usage data

- Incomplete Year Coverage:** The dataset lacks records for the Fall semester (September to December), meaning the analysis does not represent the entire academic year.

These caveats are critical for the reader to understand the limitations of certain parts of this analysis.

2. Total Number of Unique Users

In 2015, the VDI system served **3,013 total users**. This count was calculated using the n_distinct() function to ensure accuracy, as a single user's session can generate multiple records when different applications are launched. This approach correctly counts unique users regardless of how many applications they used during their sessions.



Figure 2.1: Total number of Users in the system

The visualization contrasts this total user count with the usage frequency of the top 10 most launched applications such as Internet explorer(Iexplorer.exe) or Microsoft Visual Studio (Devenv.exe). This comparison highlights that while the user base is broad, actual system usage is concentrated around a specific set of core applications, suggesting the VDI environment primarily supports specialized academic or administrative workflows.

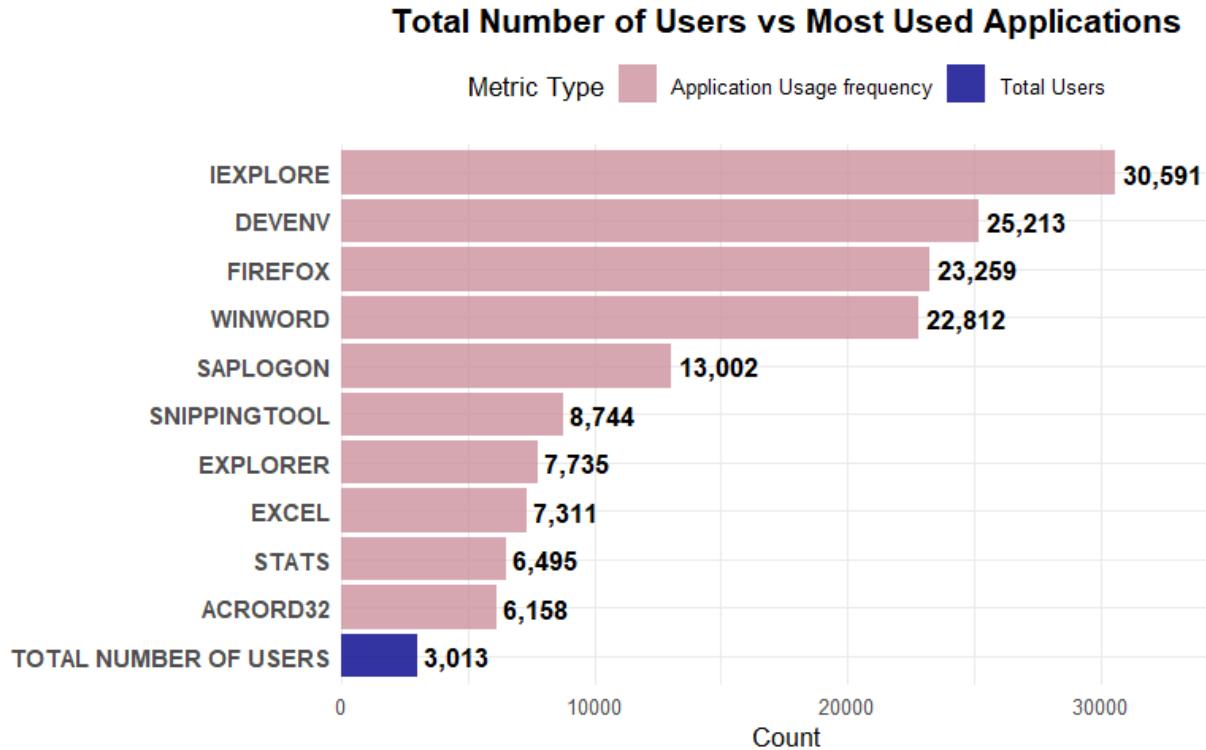


Figure 2.2: Comparison of total users and application usage

3. Average Number of Users Per Day

The average number of unique users per day was **139**. However, this figure is highly influenced by seasonal academic patterns. The median value of 62 users per day is less than half the average, indicating that the mean is heavily skewed by periods of very high usage during the regular semesters.

A tibble: 1 × 5	min_users	max_users	median_users	avg_users	total_days
	<int>	<int>	<int>	<dbl>	<int>
1 row	4	540	62	139	223

Figure 3.1: Average number of users per day

A time-series analysis clearly shows intense activity during the Spring semester (January to April), followed by a dramatic decline beginning in mid-May as the summer break began. This pattern strongly correlates the system's load with the academic term schedule. During the Summer the usage declines

significantly and daily usage goes much lower than average usage. We can see that the usage from May to the end of August is well below the average line.

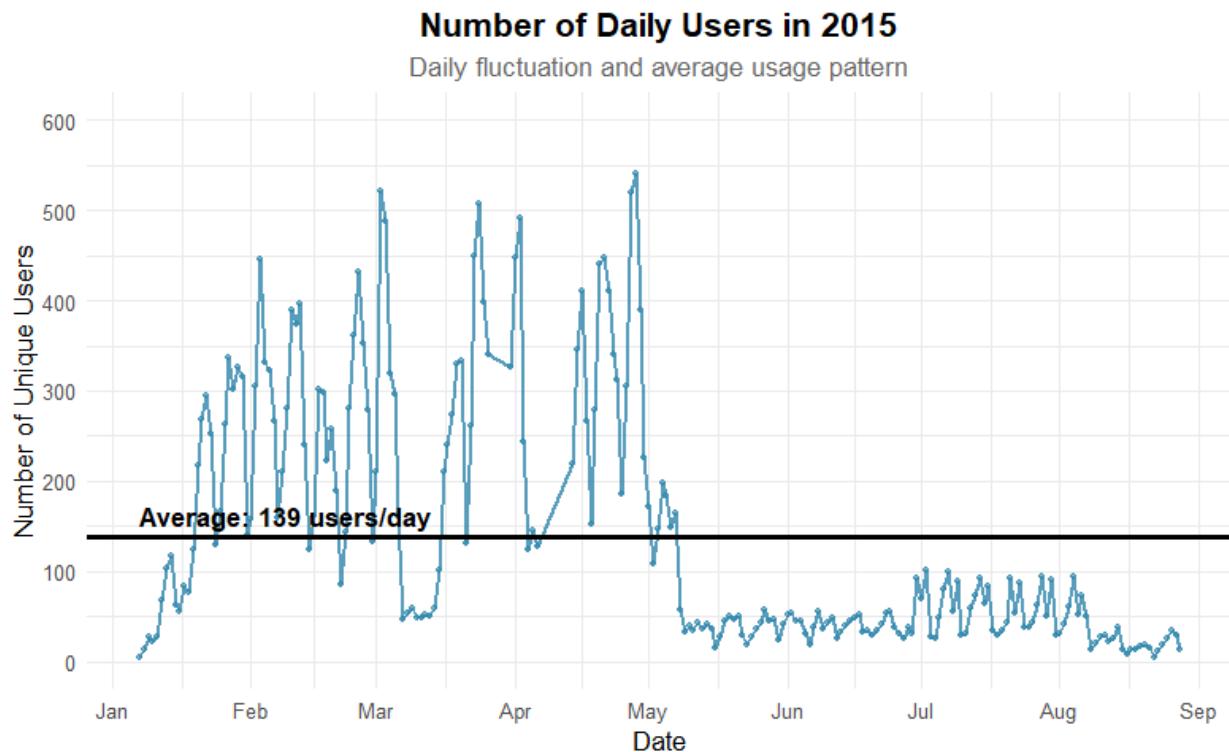


Figure 3.2: Visual Representation for Number of Daily Users

The average was also broken down by day of the week, revealing predictable, workflow-driven engagement:

- **Tuesday** was the peak day with an average of **191 users**.
- Usage remained high on **Monday (157)** and **Wednesday (169)** and **Thursday (169)**.
- A noticeable decline occurred on **Friday (117)**.
- Weekend usage was less than half of the weekday usage, with **Sunday (84)** and **Saturday (71)** seeing the lowest traffic. So the traffic declined more than 50% on weekends.

Average VDI Users by Day of Week - 2015

Tuesday is the busiest day; weekends show significantly lower usage

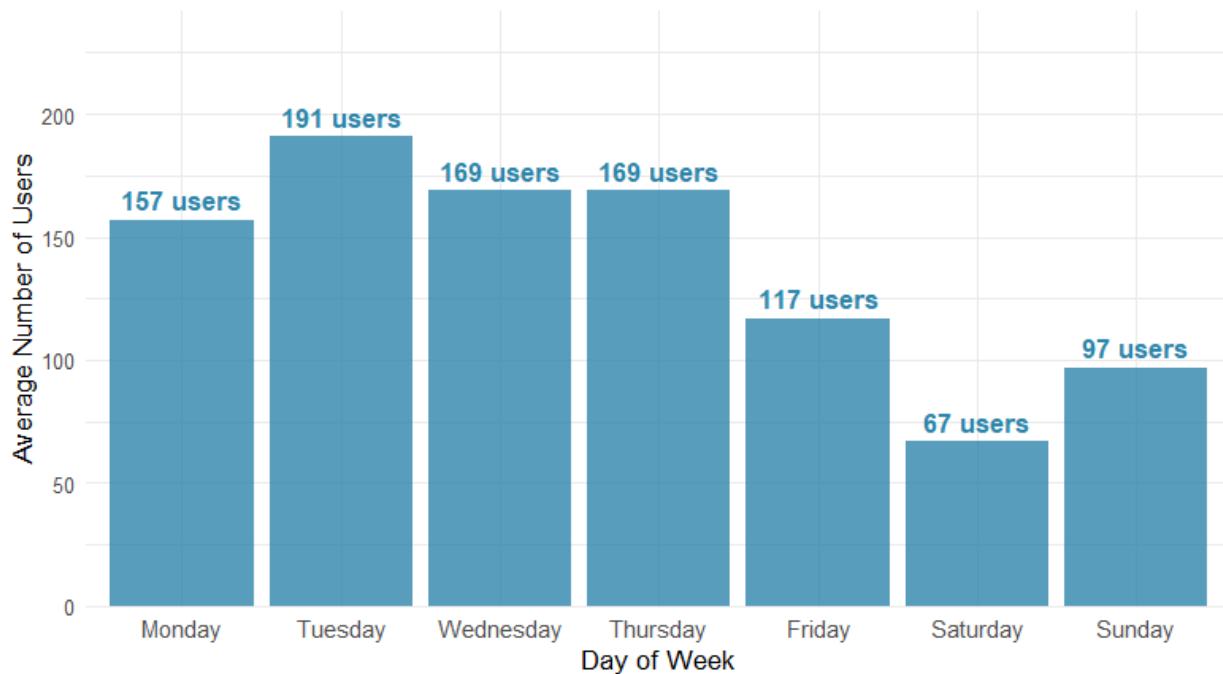


Figure 3.3: Average Number of Users each day of the week

4. Highest Number of Users Per Day

The single day with the highest system load was **Tuesday, April 28, 2015**, with **540 unique users**.

A tibble: 1 × 5				
min_users	max_users	median_users	avg_users	total_days
4	540	62	139	223

Figure 3.1: Highest number of Users on a single day

Afterwards, the highest number of users on every day of the week like Sunday or Monday was calculated a visual representation was provided as well. The bar chart shows resemblance to the previous bar chart that showed average number of users on each day of the week.

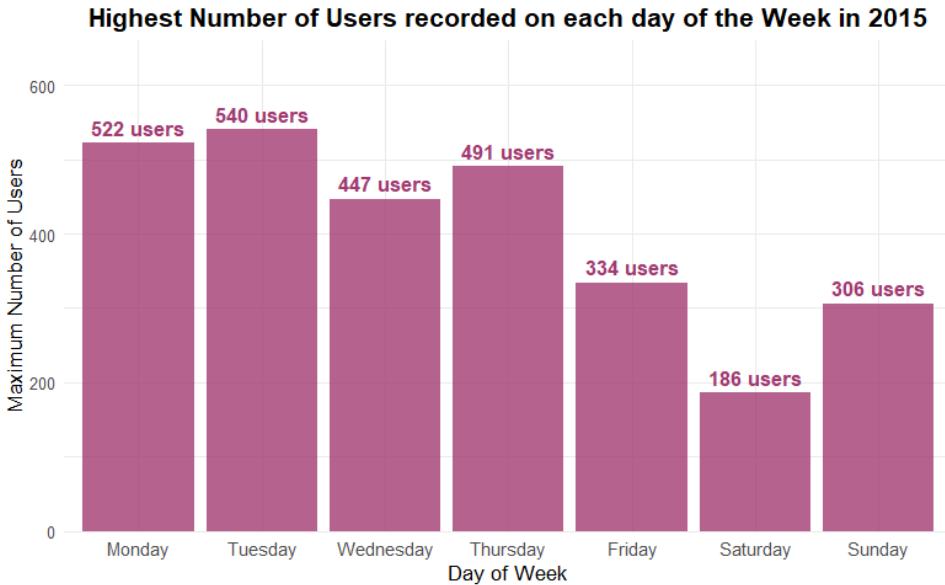


Figure 3.2: Highest number of users on each day of the week

An analysis of the top 10 busiest days reveals a strong concentration in the Spring semester, with 8 of the 10 days falling on a Monday or Tuesday. This clustering of peak usage days suggests that the beginning of the week is when demand is highest. The maximum users by day of week further confirms this pattern, with Tuesday (540) and Monday (528) recording the highest potential loads.

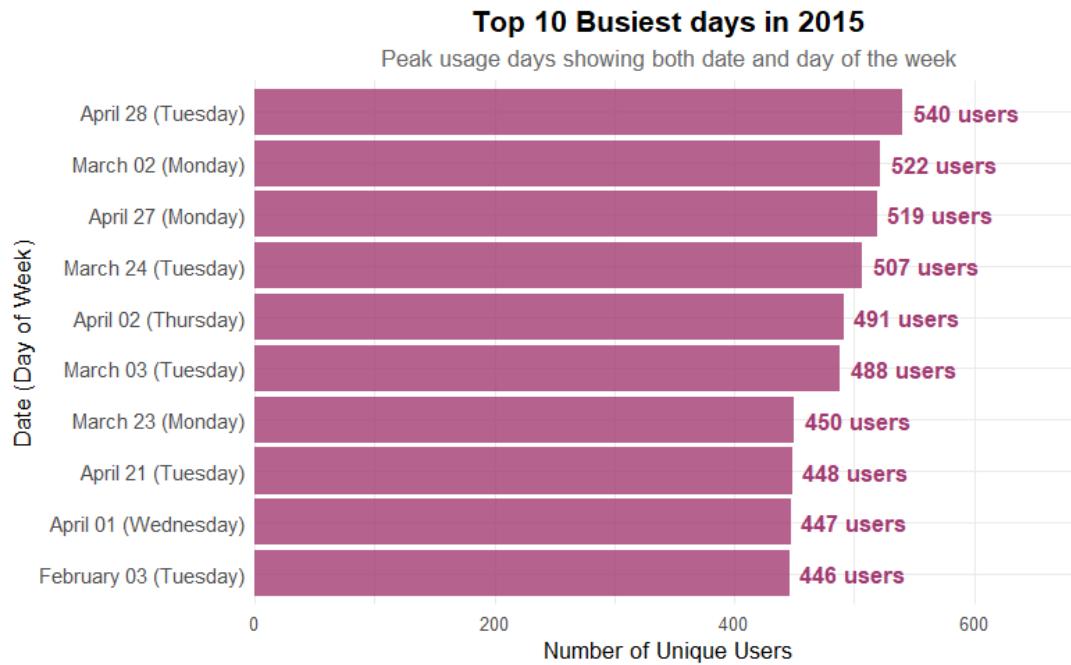


Figure 3.3: The busiest days in 2015

5. Top 5 Users by Time Logged In & CPU Usage

Top 5 Users by Length of Time Logged In

For this analysis, session duration was calculated by converting both login and logout timestamps and finding the difference for each distinct VDI session. The top 5 users by total cumulative login time are:

1. **User id 18508:** 1,261 hours (across 271 sessions)
2. **User id 23439:** 738 hours (across 176 sessions)
3. **User id 18334:** 544 hours (across 128 sessions)
4. **User id 25630:** 520 hours (across 251 sessions)
5. **User id 20875:** 459 hours (across 167 sessions)

userid	total_hours_logged	total_sessions	avg_session_length
userid18508	1261.2	271	4.7
userid23439	737.8	176	4.2
userid18334	544.4	128	4.3
userid25630	519.6	251	2.1
userid20875	459.3	167	2.8

Figure 5.1: The top 5 users and details of their sessions

These users represent the system's "power users" with the top user averaging 4.7 hours per session over 271 separate logins, indicating a high dependency on VDI usage.

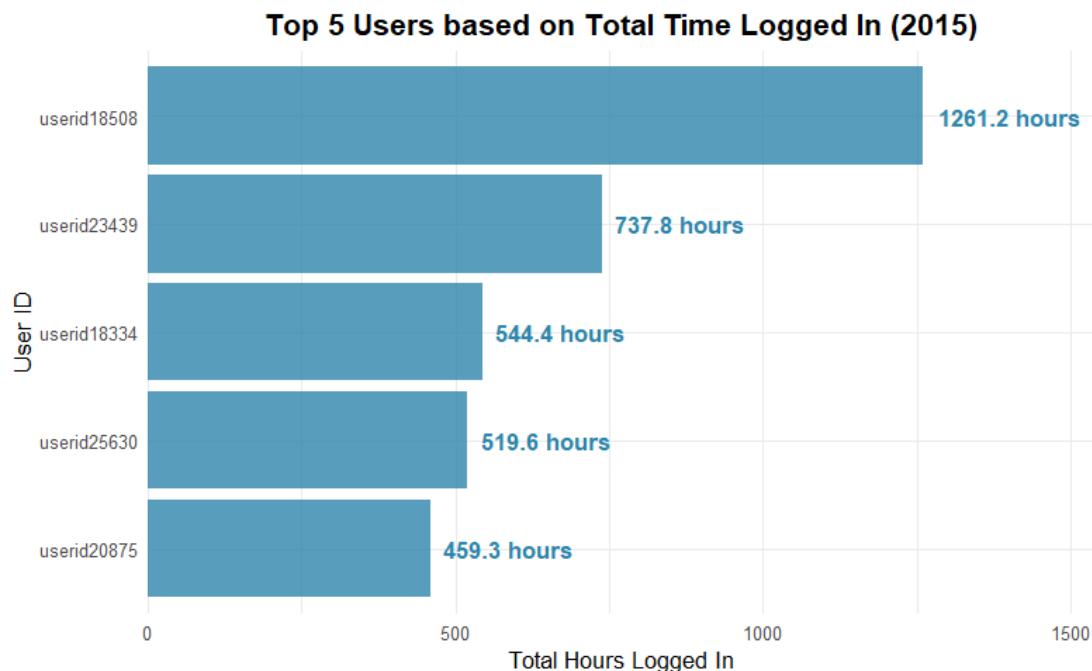


Figure 5.2: Visual representation of the usage of top 5 users

Average CPU Usage for Top Users

As noted in the data quality section, the CPU usage data contains significant issues. We already know that 5.88% (total 13,371) of the entries are negative in this column which is a data quality issue since CPU usage percentage cannot be negative. This affected our analysis of the average CPU usage of the top 5 users as well. So, for this section we conducted two separate analysis, one with the given data which contains negative values and one with a filtered data set with only positive values.

userid	total_hours_logged	total_sessions	total_cpu_records	positive_cpu_records	negative_cpu_records	negative_cpu_percent	avg_cpu_all	avg_cpu_positive
userid18508	1261.2	271	1235	791	444	36.0	4.3	10.6
userid23439	737.8	176	1016	968	48	4.7	16.4	17.5
userid18334	544.4	128	644	497	147	22.8	4.1	7.9
userid25630	519.6	251	916	899	17	1.9	5.7	5.8
userid20875	459.3	167	812	508	304	37.4	4.6	9.6

Figure 5.3: Negative CPU Usage records for the top 5 users

Comparison of CPU calculations with and without negative values

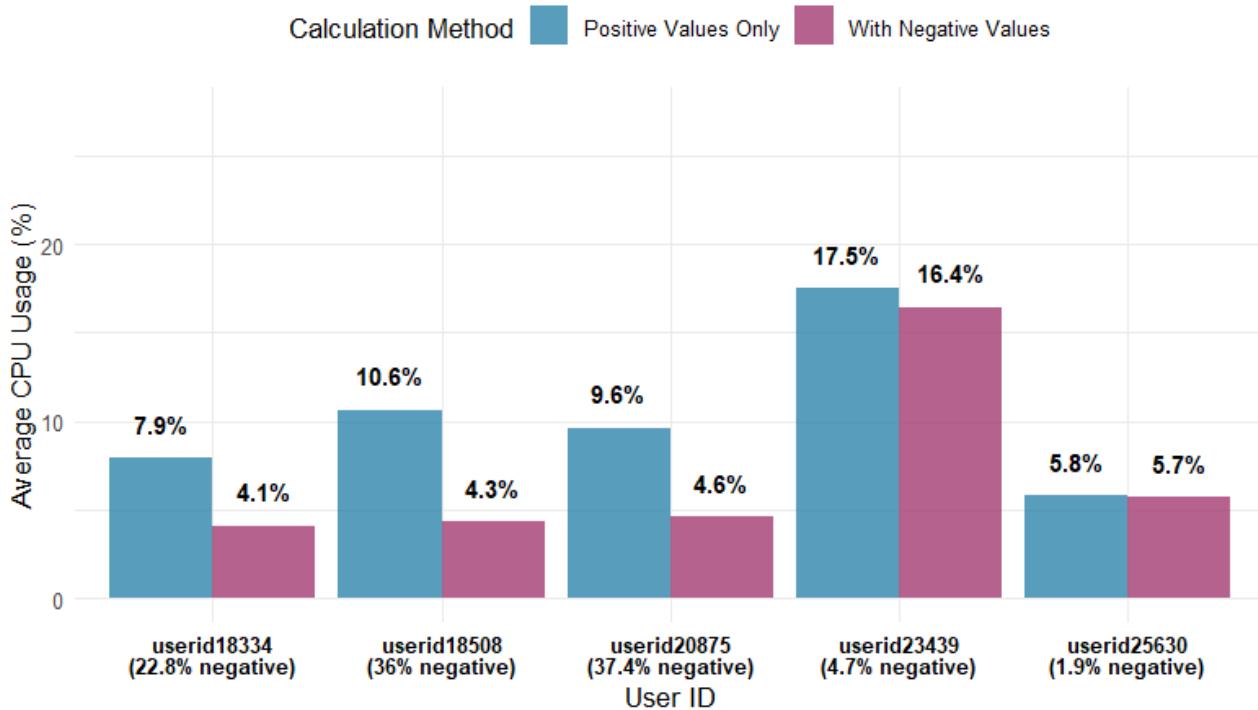


Figure 5.4: Comparison of CPU Usage of top 5 users with and without the negative values

Risk Analysis and Methodological Limitations

Data Reliability Concerns:

The analysis faces several reliability challenges that affect results and the interpretation:

1. **CPU Usage Data Integrity:** With 5.88% of CPU records containing invalid negative values and this severely affects any analysis related to CPU usage.
2. **Session Completion Bias:** 15.97% of application sessions lack proper application stop times, creating a systematic gap in usage duration analysis. This may skew our understanding of actual application engagement patterns.
3. **Temporal Coverage Limitations:** The missing Fall semester data (September-December) creates a significant gap in understanding full academic year patterns. This limits our understanding of annual usage.

External Factor Considerations:

Findings may be influenced by external factors not captured in the data:

1. Academic calendar variations (exam periods or breaks)
2. System upgrades or outages
3. Changes in user training or user interface
4. Competing technology adoption

Strategic Recommendations

1. **Infrastructure Optimization:** Implement dynamic scaling during summer months to reduce resource allocation by 40-50%, potentially saving thousands in operational costs.
2. **Power User Engagement:** Identify and recruit the top 5 users for a beta testing program to gather qualitative feedback and improve system features.
3. **Maintenance Scheduling:** Move all system maintenance to weekends when usage drops by more than 50%, minimizing disruption to academic activities.
4. **Data Quality Improvement:**
 - Fix CPU monitoring system calibration to eliminate negative values
 - Implement proper session termination tracking
 - Establish data validation protocols
5. **Capacity Planning Enhancement:**
 - Allocate 25% more resources for peak usage weekdays.
 - Develop seasonal staffing models aligned with academic calendar. Regular staffing is not needed during the Summer due to significantly low user traffic.

- Create automated scaling protocols for exam periods

6. User Support Optimization:

- Extend support hours during peak usage periods like Monday or Tuesday.
- Develop specialized training for high-engagement applications. These aren't needed for the basic applications like internet explorer or Microsoft Word but for the more advanced or technical applications like Visual Studio or SAP.
- Create a quick start guide for infrequent users (555 one-time users).

Conclusion

This analysis demonstrates that while the VDI system serves a substantial user base of 3,013 individuals with clear academic usage patterns, significant opportunities exist for optimization and improvement. The predictable seasonal and weekly fluctuations from peak Tuesday usage of 540 users to less than 100 users on weekends, also from that same peak of 540 users in the spring semester to below 100 users every day during the Summer which is well below the average of 139, provide a data-driven foundation for strategic resource allocation. However, the identified data quality issues, particularly the invalid CPU metrics and incomplete session records, highlight the need for improved system monitoring. By implementing the recommended strategies for infrastructure optimization, power user engagement, and data quality enhancement, the organization can transform these insights into tangible cost savings and improved service delivery, ensuring the VDI system continues to effectively support our academic missions while operating more efficiently.

Appendix

The complete R Markdown notebook containing all code used for data preparation, analysis, and visualization along with a knitted output file (word doc) has been submitted separately along with this document.