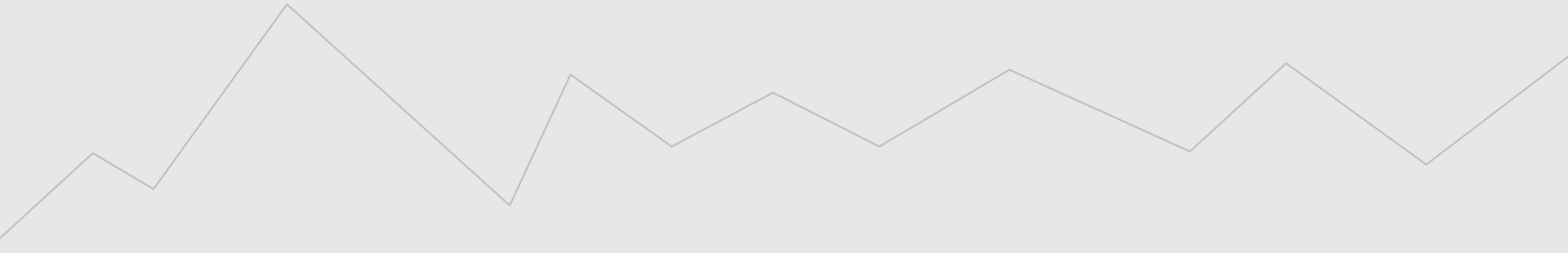
Study LSTM

DSC 180B - Group 3 - 2023/02/03



Data Preparation

User 1

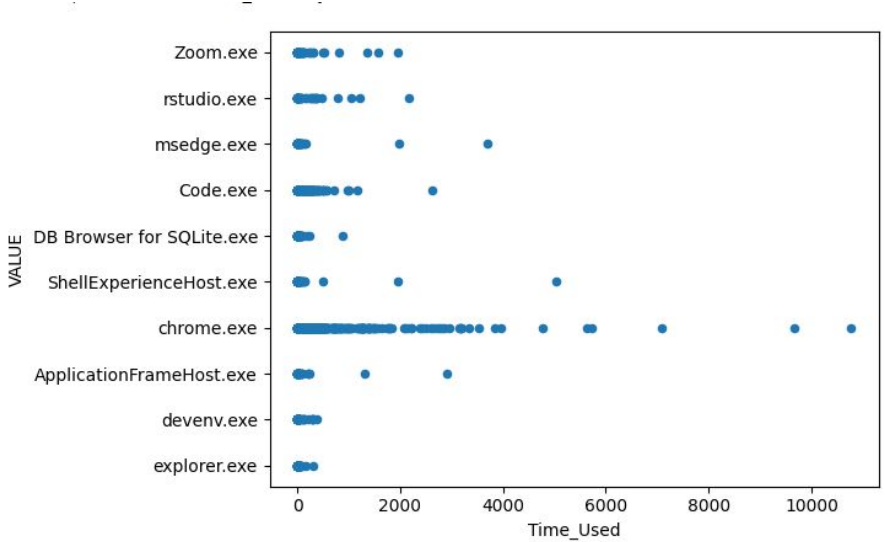
	MEASUREMENT_TIME	ID_INPUT	VALUE	PRIVATE_DATA	Time_Used
640	2023-01-17 06:01:02.341	4	chrome.exe	0	10768.311
1733	2023-01-23 00:42:20.235	4	chrome.exe	0	9673.942
796	2023-01-18 07:44:57.400	4	chrome.exe	0	7089.715
1630	2023-01-21 06:18:44.711	4	chrome.exe	0	5721.912
1653	2023-01-22 00:03:13.554	4	chrome.exe	0	5625.497
2575	2023-01-27 03:29:11.800	4	ShellExperienceHost.exe	0	5040.236
1642	2023-01-21 17:58:49.178	4	chrome.exe	0	4782.524
2288	2023-01-25 19:25:15.942	4	chrome.exe	0	3959.602
1680	2023-01-22 18:05:45.815	4	chrome.exe	0	3832.000
307	2023-01-16 02:18:08.317	4	msedge.exe	0	3698.763
1269	2023-01-20 05:16:54.972	4	chrome.exe	0	3520.850
123	2023-01-15 21:36:30.521	4	chrome.exe	0	3348.958
1003	2023-01-19 05:18:25.091	4	chrome.exe	0	3197.018
2528	2023-01-27 02:05:43.849	4	chrome.exe	0	3183.326
494	2023-01-16 21:36:58.888	4	chrome.exe	0	3160.566
1646	2023-01-21 22:22:01.224	4	chrome.exe	0	3137.300
2270	2023-01-25 18:08:37.536	4	chrome.exe	0	2965.804

User 2

	MEASUREMENT_TIME	ID_INPUT	VALUE	PRIVATE_DATA	Time_Used
1263	2023-01-24 21:00:22.469	4	chrome.exe	0	19982.078
1005	2023-01-23 18:55:05.496	4	explorer.exe	0	11530.199
1727	2023-01-31 00:29:09.836	4	ApplicationFrameHost.exe	0	8935.665
1329	2023-01-26 00:37:02.316	4	explorer.exe	0	8611.865
1827	2023-02-01 00:48:18.343	4	ApplicationFrameHost.exe	0	8448.389
842	2023-01-22 15:39:14.564	4	chrome.exe	0	8370.358
1731	2023-01-31 03:17:15.001	4	chrome.exe	0	7864.547
1535	2023-01-27 03:34:56.038	4	Spotify.exe	0	7438.509
886	2023-01-23 01:42:40.339	4	chrome.exe	0	7291.960
1465	2023-01-27 01:03:21.880	4	ApplicationFrameHost.exe	0	6801.482
298	2023-01-20 00:56:02.975	4	msedge.exe	0	5761.546
858	2023-01-22 18:07:26.948	4	chrome.exe	0	4703.125
1705	2023-01-27 22:16:15.424	4	ApplicationFrameHost.exe	0	4254.468
1753	2023-01-31 06:23:40.426	4	chrome.exe	0	3870.576
870	2023-01-22 20:47:51.901	4	chrome.exe	0	3757.708
1400	2023-01-26 03:49:42.756	4	chrome.exe	0	3493.871
1181	2023-01-24 17:09:44.825	4	chrome.exe	0	3493.001
1414	2023-01-26 05:15:48.820	4	chrome.exe	0	3452.686
1434	2023-01-26 23:44:12.833	4	chrome.exe	0	3238.186

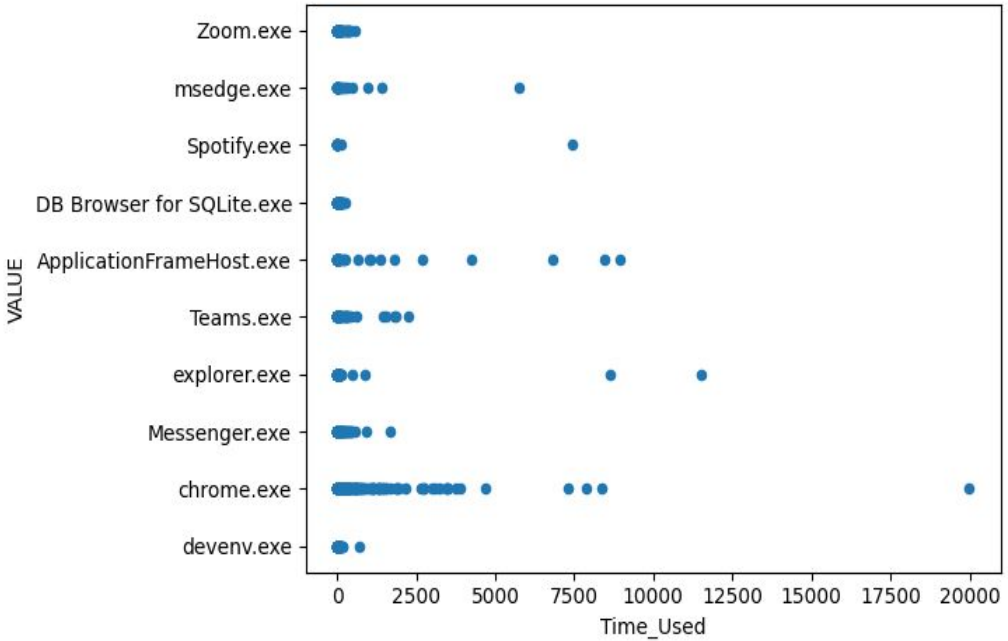
Data Preparation (User 1)

VALUE	
chrome.exe	264347.415
Code.exe	32619.039
Zoom.exe	10050.404
rstudio.exe	9281.864
ShellExperienceHost.exe	9026.440
msedge.exe	6833.118
ApplicationFrameHost.exe	5519.452
devenv.exe	3277.051
DB Browser for SQLite.exe	2827.803
explorer.exe	2212.040



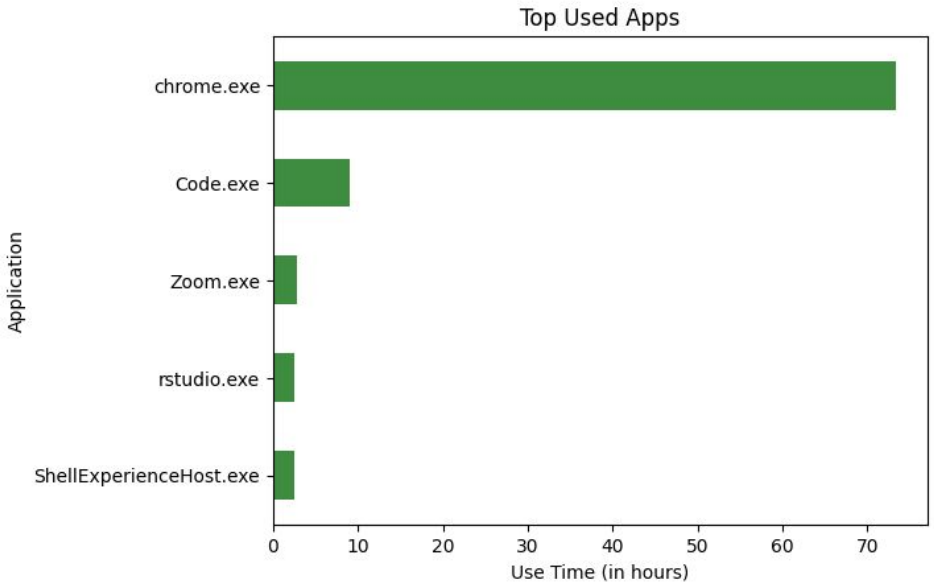
Data Preparation (User 2)

VALUE	
chrome.exe	155944.585
ApplicationFrameHost.exe	38215.832
explorer.exe	24667.769
Teams.exe	14558.154
Messenger.exe	11137.530
msedge.exe	9883.659
Spotify.exe	7634.397
Zoom.exe	6477.202
DB Browser for SQLite.exe	3794.490
devenv.exe	2632.721

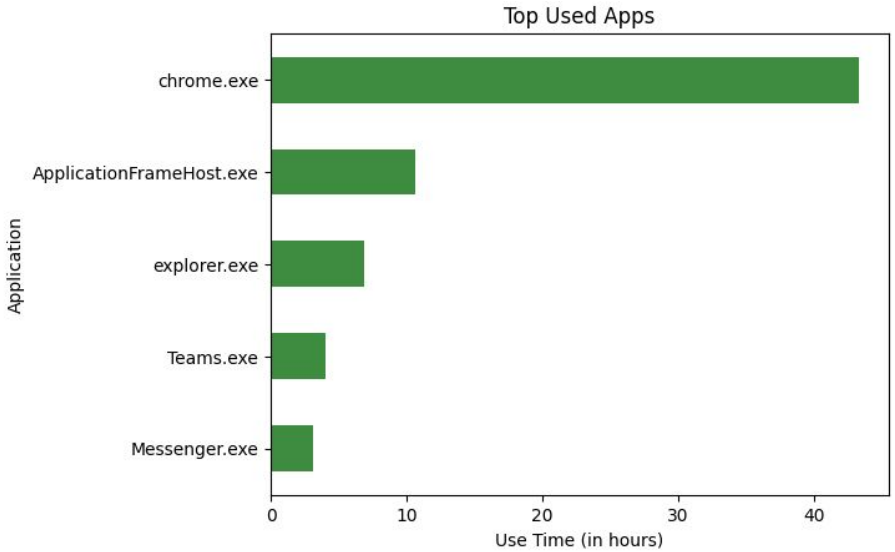


Data Visualization

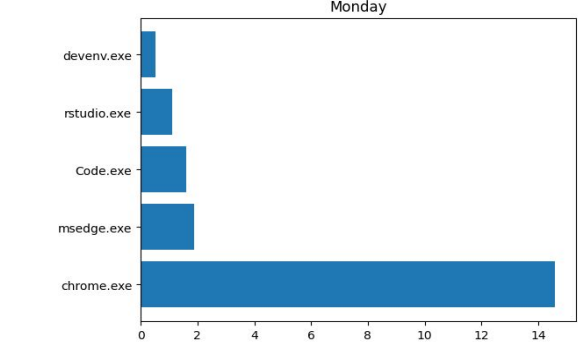
User 1

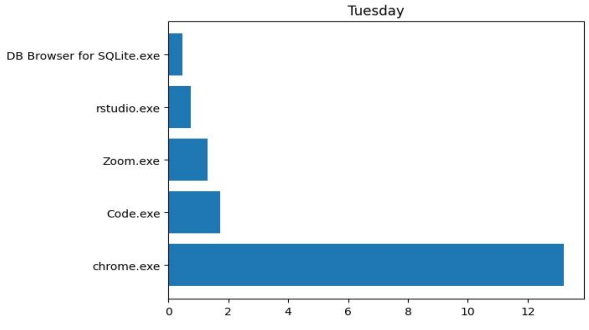


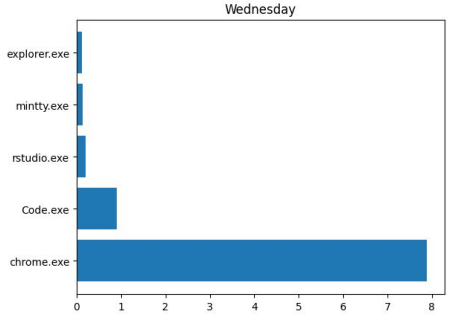
User 2

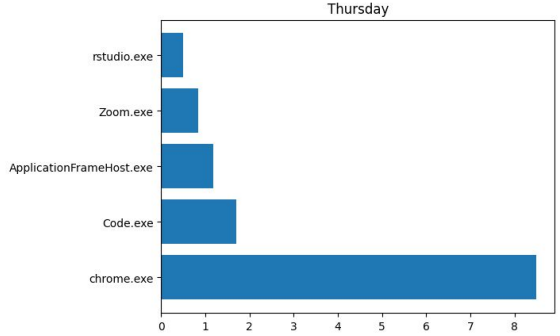


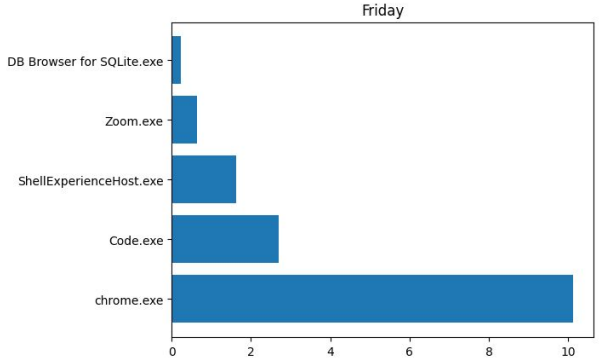
User 1

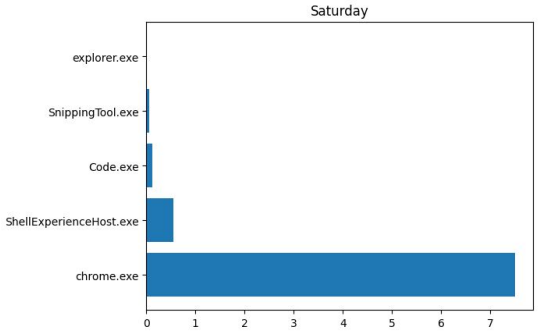

Monday
devenv.exe
rstudio.exe
Code.exe
msedge.exe
chrome.exe
0
2
4
6
8
10
12
14

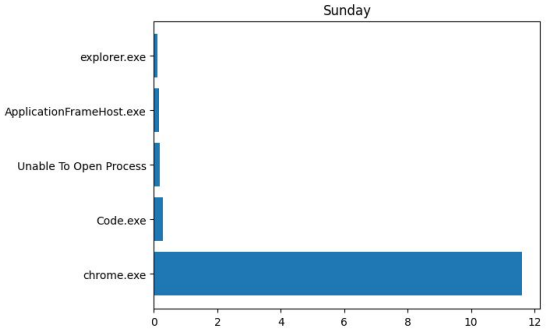

Tuesday
DB Browser for SQLite.exe
rstudio.exe
Zoom.exe
Code.exe
chrome.exe
0
2
4
6
8
10
12


Wednesday
explorer.exe
mintty.exe
rstudio.exe
Code.exe
chrome.exe
0
1
2
3
4
5
6
7
8

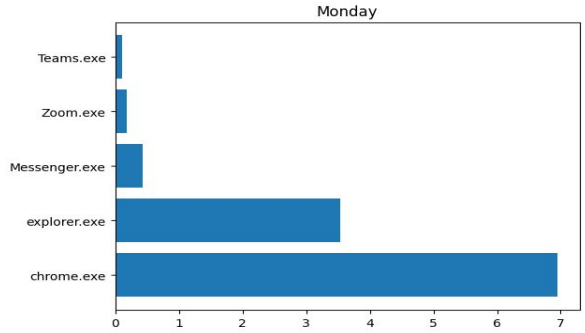

Thursday
rstudio.exe
Zoom.exe
ApplicationFrameHost.exe
Code.exe
chrome.exe
0
1
2
3
4
5
6
7
8

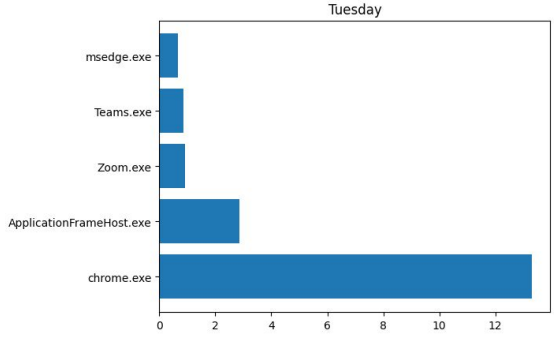

Friday
DB Browser for SQLite.exe
Zoom.exe
ShellExperienceHost.exe
Code.exe
chrome.exe
0
2
4
6
8
10

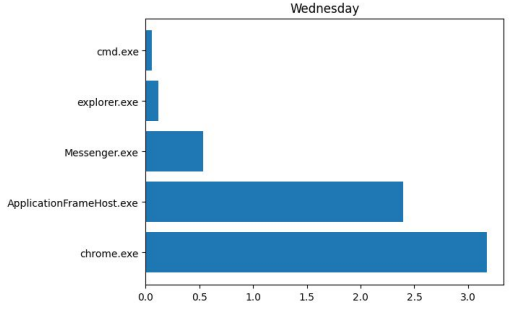

Saturday
explorer.exe
SnippingTool.exe
Code.exe
ShellExperienceHost.exe
chrome.exe
0
1
2
3
4
5
6
7

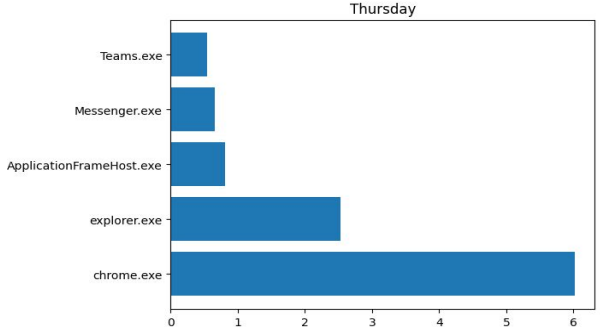

Sunday
explorer.exe
ApplicationFrameHost.exe
Unable To Open Process
Code.exe
chrome.exe
0
2
4
6
8
10
12

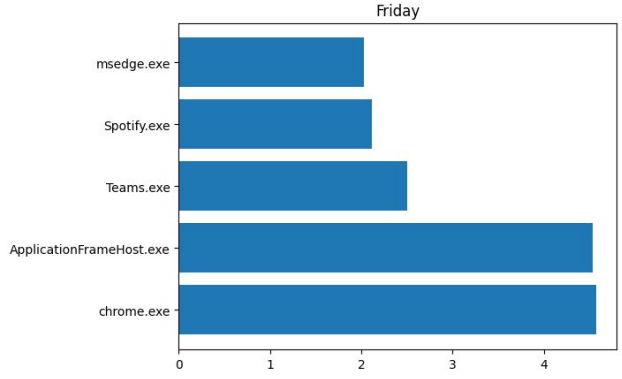
User 2

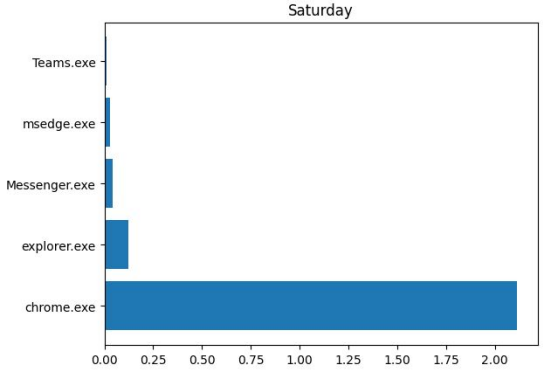

Monday
Teams.exe
Zoom.exe
Messenger.exe
explorer.exe
chrome.exe
0
1
2
3
4
5
6
7

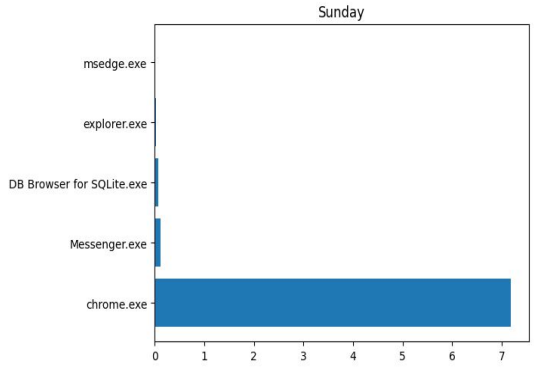

Tuesday
msedge.exe
Teams.exe
Zoom.exe
ApplicationFrameHost.exe
chrome.exe
0
2
4
6
8
10
12


Wednesday
cmd.exe
explorer.exe
Messenger.exe
ApplicationFrameHost.exe
chrome.exe
0.0
0.5
1.0
1.5
2.0
2.5
3.0


Thursday
Teams.exe
Messenger.exe
ApplicationFrameHost.exe
explorer.exe
chrome.exe
0
1
2
3
4
5
6


Friday
msedge.exe
Spotify.exe
Teams.exe
ApplicationFrameHost.exe
chrome.exe
0
1
2
3
4


Saturday
Teams.exe
msedge.exe
Messenger.exe
explorer.exe
chrome.exe
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

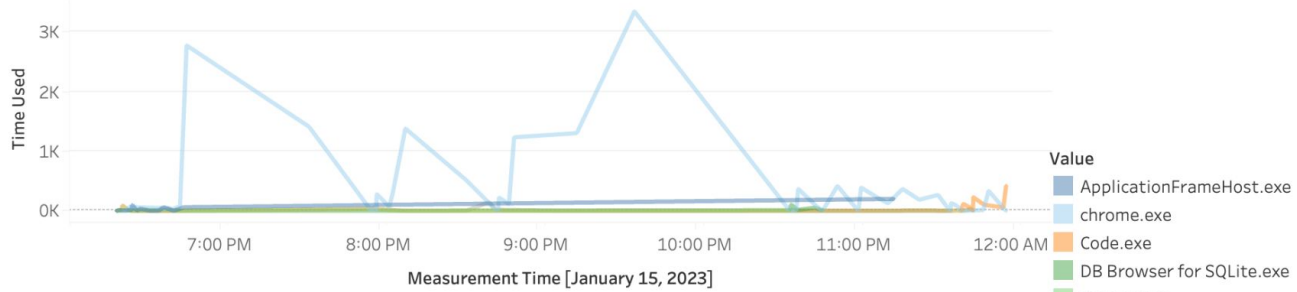

Sunday
msedge.exe
explorer.exe
DB Browser for SQLite.exe
Messenger.exe
chrome.exe
0
1
2
3
4
5
6
7

User 1

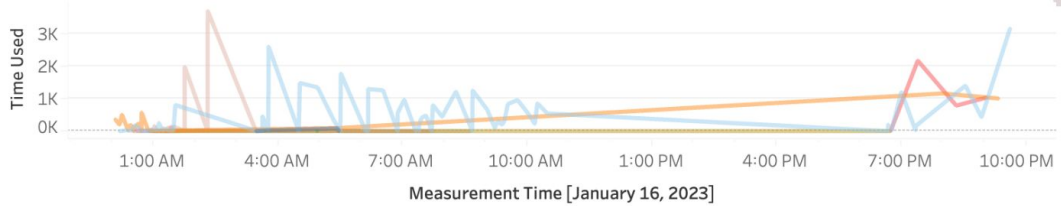
User 1

- A visual representation of the time spent on multiple processes from Jan 15 - Jan 27, 2023
- A calendar-like representation / break down of top 10 most used processes by user 1

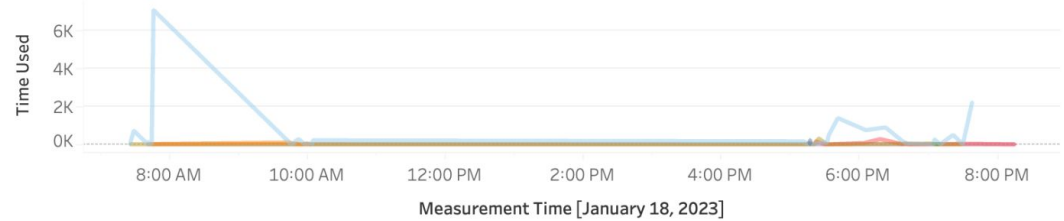
Daily Analysis of Process Use 01/15

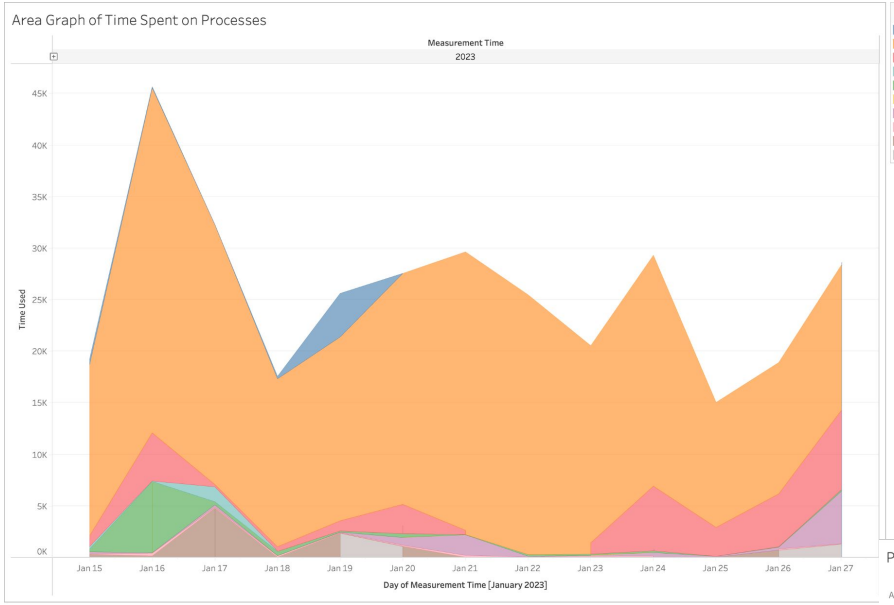


Daily Analysis of Process Use 01/16

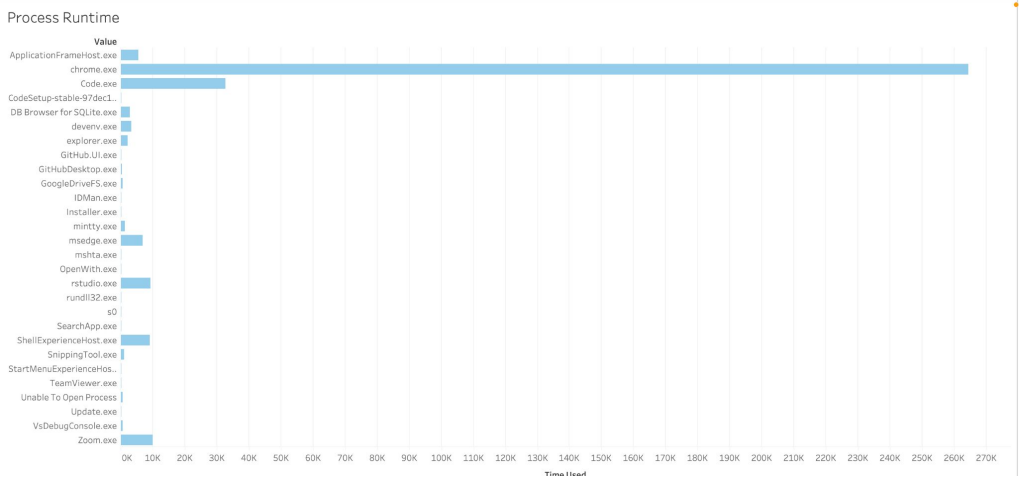


Daily Analysis of Process Use 01/18



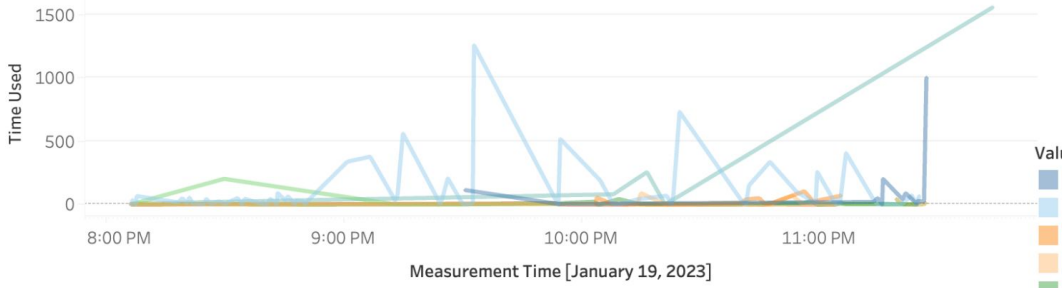


Average time spent actively using a particular application between 01/15/23 - 1/27/23

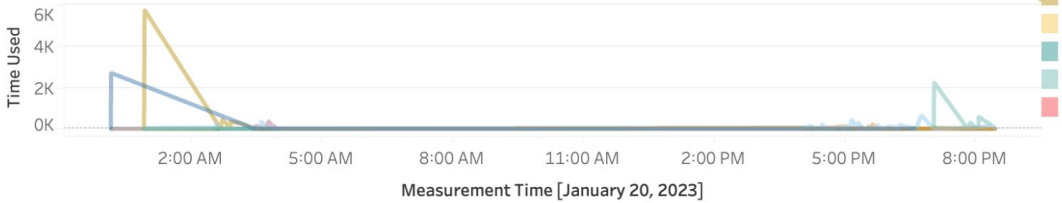


User 2

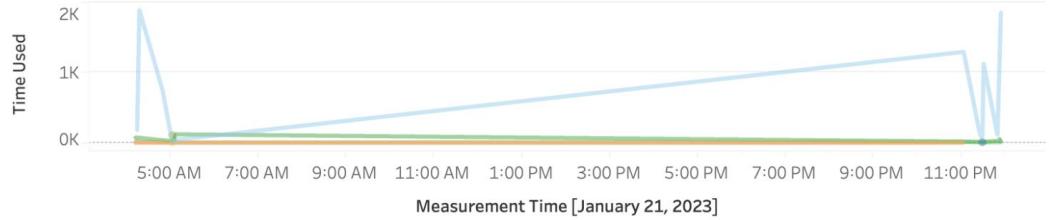
Daily Analysis of Process Use 01/19



Daily Analysis of Process Use 01/20

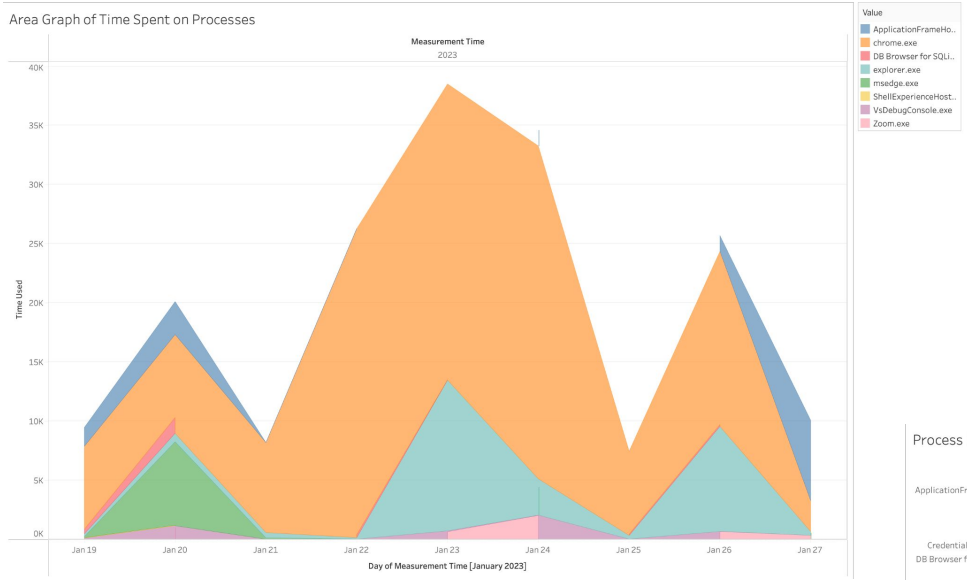


Daily Analysis of Process Use 01/21

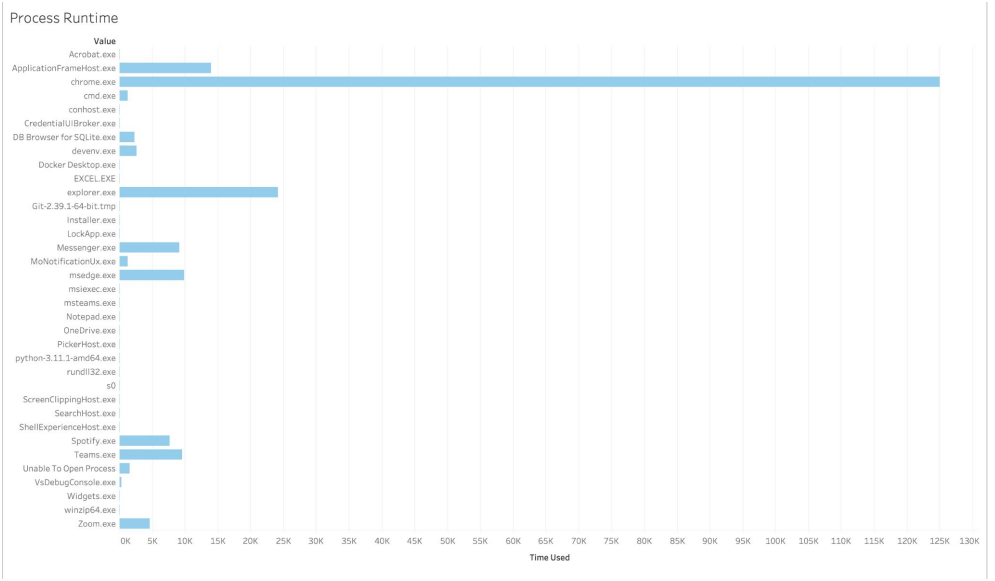


User 2

- A visual representation of the time spent on multiple processes from Jan 19 - Jan 27, 2023
- A calendar-like representation / break down of top 10 most used processes by user 2



Average time spent actively using a particular application between 01/19/23 - 1/27/23



Problem statement

- **Main task**:
  Predict the total time an app/tab is used in the foreground

  - **Subtask 1**: Predict #hrs an app is used during a day of a week
  - **Subtask 2**: Predict #secs an app is used during an hour of a day

  ⇒ Time-series forecasting ⇒ RNN/LSTM

Feature Selection

- MEASUREMENT_TIME
  ⇒ The ***datetime*** the app was used
  ⇒ ***Use time*** in secs (of an hr) or hrs (of a day)

Example

App 1	hr1 [12am, 1am)	hr2 [1am, 2am)	…	hr23 [10pm, 11pm)	hr24 [11pm, 12am)
day_1	1800 sec	0 sec		900 sec	450 sec
day_2	900 sec	0 sec		3600 sec	3600 sec
…	…	…		…	…
day_n	1800 sec	450 sec		1000 sec	200 sec

Example

App 1	Monday	Tuesday	…	Saturday	Sunday
week_1	3 hrs	2 hrs		5 hrs	4 hrs
week_2	4 hrs	4 hrs		2 hrs	3 hrs
…	…	…		…	…
week_n	2 hrs	5 hrs		2 hrs	4 hrs

Experiment

- Split data: train/validation/test sets: 70/20/10 %
  - No shuffle before splitting data to ensure
    - Consecutive windows of samples
    - Realistic results from the validation/test sets
      - Evaluated on data collected after training models

Experiment

- Normalize data:
  - Scale features before training
    - Sin/Cos transform the “time of day”, “time of week” signals
      - 12 am → 12 am (again)
      - Mon → Mon (again)
        - Periodically
    - Min-Max scale the use time
      - Preserve the original distribution
      - Smaller STD, reduce the effect of outliers

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

    - Z-score:

$$Z = \frac{x - \mu}{\sigma}$$

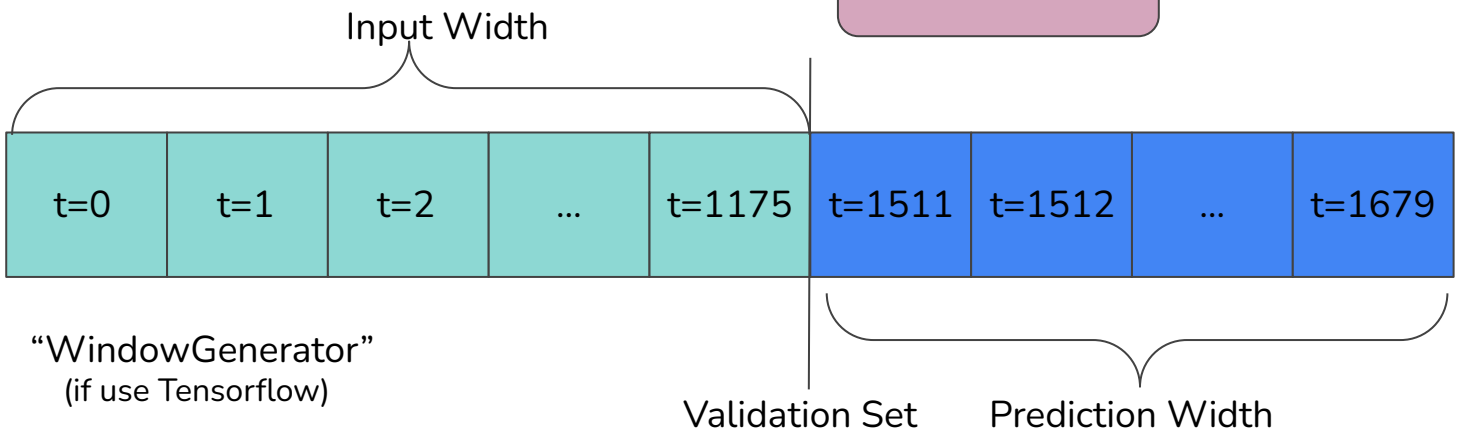
  - Do this separately for train/validation/test sets

Experiment

- "Window" data:
  - Make a prediction (24*7*1) hrs into the future, given (24*7*7) hrs of history
    ⇒ Our the window:

168 hrs

1176 hrs

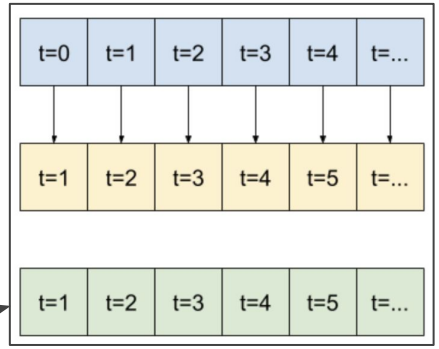


Experiment

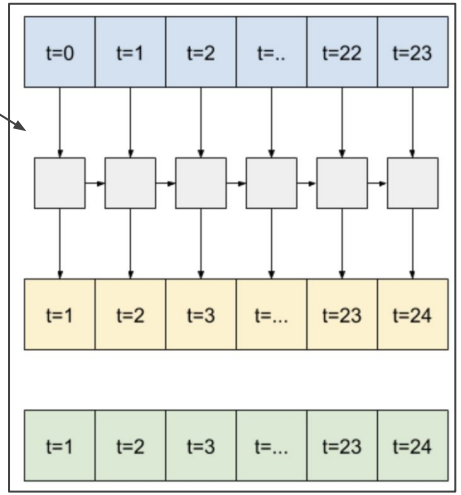
- Tune hyperparameters:
  - (Recurrent) Activation functions
    - tanh(x), relu, ...
  - Dropout; Regularization
    - Prevent overfitting
  - Bias for layers

Experiment

- Compare the models:
  - Baseline
  - Dense layers
  - LSTM layers



- ~ linear model
- + dense layers



Experiment

- Performance Metrics:
  - (Root) Mean Squared Errors
  - Mean Absolute Errors

$$RMS = \sqrt{\frac{1}{n}\sum_{i} x_i^2}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\boldsymbol{x}_i - \boldsymbol{x}|$$

Thank you!

- References:
  - https://www.tensorflow.org/tutorials/structured_data/time_series
  - https://keras.io/api/layers/recurrent_layers/lstm/