Syracuse University

Memo

|  |  |
| --- | --- |
| To: | Dr. Landowski |
| From: | Your name(s) |
| Date: | February 18, 2022 |
| Re: | Project Proposal |

**Topic: Analyze public workstation usage**

**Data Description:**

The main data set consists of 18 Excel Spreadsheets that contain information for each user login event on a public workstation at the University of Rochester main campus from July 2015 to July 2018.

|  |  |  |
| --- | --- | --- |
| Fields | Description | Example |
| Station | workstation id | GGN431-9020-04 |
| Group | Location group | Goergen 431 |
| User | user’s netid | fakeusr2 |
| Start Time | time the user logged on the workstation | 2018-04-11 10:57 AM |
| End Time | time the user logged off the workstation | 2018-05-08 08:15 PM |
| Duration (Seconds) | total number of seconds the user was logged on | 2366316 |
| Adjusted Duration (Seconds) | number of seconds from when the user logged on until the close time of the location | 677741 |

I am trying to get supplemental data to identify the status of the user (undergraduate, graduate, staff, faculty) to be able to do some analysis of who is using the workstations when. I am also waiting for some data on wifi usage in the same spaces, but I don’t know if this will be available in the time frame of this project.

**Research Questions:**

* What are the busiest times of the year, least busy?
* How busy is each location? How often are all the workstations in use?
* Are there any longer-term trends in the use of public workstations? Is usage declining?
* Have the number of unique users changed over the 3-year period covered by the data?

**Data Preparation Plan**

1. Load the data into Python
2. Concatenate the data from each file into one data set
3. Filter the data set to only records in Groups we are interested in analyzing
4. Explore the data for anomalies, missing data, adjusted duration, etc.
5. Resolve any issues related to the previous step, including deciding how to handle records with adjusted duration and how to handle records for IT and library staff using the workstations.
6. Convert the data from start and end time duration to hourly records to count logins in each hour, day, month.
7. Add features such as day of the week, day, month, hour (see above).