Wi-Fi Session Data

Time Period	3 months
Nº of Files	1
Nº of records	13983117
Size	3,36 GB

Origin

This data is collected from Wi-Fi sessions of users connected to Porto Digital's Wi-Fi access points (APs) distributed across the Porto municipality.

A session is essentially a period where the user connects to the network, actively uses the network and then disconnects. When a user connects to a Porto Digital's AP, ongoing session information is accounted by its respective wireless controller, such as the start time of a session, the number of bytes downloaded and uploaded bytes, the AP the user connected to, and others, and stored in a database.

Data Dictionary

There is only one csv containing all the 3 months Wi-Fi session data. Each record in the dataset refers to accounting data of a user connection and network usage during a specific period.

To protect user privacy, some variables were removed, and transformations were applied to the original data.

Name	Description	Type	Example
acctsessionid	Pseudo-identifier that identifies each session. It is not a unique identifier	String	fbfe95a31b975faa4b79968f95d834f2204ff050c4cb17
acctstarttime	Indicates the start time of the connection	Timestamp	2024-09-18 14:00:00

	in a specific hour		
acctsessiontime	The total duration of the connection in minutes	Int	30
upload	Total upload of the session in MB	Int	3
download	Total download of the session in MB	Int	20
calledstationid	Pseudo-identifier corresponding to the MAC address of the AP the user connected to	String	e0c8b4c72e48ff3726f36a987f29bab9b3d718646b3a6e
callingstationid	Pseudo-identifier corresponding to the MAC address of the user	String	412a9bf588f4a609cd8e4c8fc8b69482d6f69079f274de
calledstation_ssid	Indicates the network the user accessed to. It can be either Eduroam or Porto.FreeWiFi or NaN	String	eduroam

Anonymization process

As previously mentioned, for this data to be shared in this event, an anonymization process was applied. The following techniques were performed:

- Pseudo-anonymization
- Suppression
- Generalization
- Noise

Important Notes

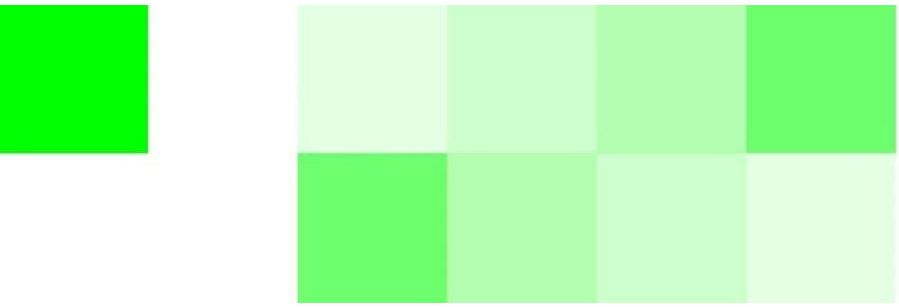
It is important to note that **a session can be split into multiple records** (connections). If a user moves around the city while using the network, his session will have multiple records in the database, as a change in AP during the same session causes a new record to be generated. This introduces some important details about the data – as only the most recent record for each user and session id combination gets updated.

For the reason described above, download and upload are not cumulative within a session. Thus, if a session is split into multiple records (same session id and user mac address), the max value of download and upload should be considered.

You will also find a large number of records with upload and download equal to 0. This is caused by the already mentioned record splitting, as a record may not get updated before a new one is generated. The same happens to session time, although for this variable the value will be NaN.

One way you can extract trajectories of users from the dataset is by utilizing the callingstationid and any time period you define, such as hour, morning, day, etc, which will give you the movements of a user for the defined time period.

You can merge this data with the Access Point data, which contains detailed information about each AP, although you have to perform an arbitrary link between the two sources.



Hack a City