Loader – Krátká dokumentace

# Table overwriter

Table overwriter (dále TO) je základní funkce celého loader skriptu. TO funkce má za úkol načíst aktuální tabulku škol a novou tabulku škol, a následně je násilně spojit.

## Fungování

TO začíná načítáním:

* Proměnná **current\_schools** je inicializována jako prázdný dataframe, a následně pokud je nalezen soubor „schools.xlsx“ (obsahující aktuální školy), načte tyto školy. Jinak vytvoří prázdný dataframe se sloupečky které by měl výsledný dataframe obsahovat.
* Proměnná **new\_schools** obsahuje nové (hodnoty) škol(y). Soubor je vybrán kontextovým okénkem na začátku.
* Proměnná **fac** obsahuje název fakulty. Hodnota je načtena kontextovým okénkem na začátku.
* Proměnná **addresses** obsahuje nutné informace k získání lokací škol. Je využívána k přejmenování škol a získání lokací.

Nutno říci: Kontextová okénka jsou volána před začátkem samé funkce, v driveru na konci (za if \_\_name\_\_\_ …).

Následně jsou volány funkce upravující tabulku new\_schools. Tyto funkce jsou popsány o kapitolu níže.

Na konci skriptu je přidán sloupeček s fakultou daných škol. Tato část je na konci kvůli funkci přidávající obory, která podstatně mění délku tabulky, tudíž na konci dává konzistentnější výsledky. Vlastně by mohla bejt kdekoliv, protože by se délka měnila se zbytkem tabulky, but why fuckin bother. Poté jsou vybrány sloupečky relevantní pro current\_schools, a následně dochází k mergování tabulek.

První je aplikován anti-join na current\_schools. Zbytek skriptu je designován tak, aby new\_schools obsahoval všechny informace z current\_schools. Toto nefunguje úplně, protože teď tak tam budou samostatně stejné školy se smlouvama pod různejma fakultama, ale cest la vie. Anyway, potom jsou pod zbývající current\_schools řádky nalepeny new\_schools řády, a výsledek je zapsán do schools.xlsx.

## Funkce upravující tabulku

### Unite\_cols

### Extract\_dptmnts

### Rename\_subs\_hard

### Get\_url

### Get\_coords