



TECHNICAL UNIVERSITY OF LIBEREC
Faculty of Mechatronics, Informatics
and Interdisciplinary Studies ■

TECHNOLOGIE PRO BIG DATA

CVIČENÍ III.

OPAKOVÁNÍ – PYTHON

Lukáš Matějů
25.10.2023 | TPB



CVIČENÍ

- vizualizace dat z webu [iDNES.cz](https://idnes.cz)
 - volba knihovny pro vizualizaci je na vás
 - matplotlib, Seaborn, Plotly, ggplot, ...
 - vykreslete křivku zobrazující přidávání článků v čase
 - vykreslete sloupcový graf zobrazující počet článků v jednotlivých rocích
 - vykreslete scatter graf zobrazující vztah mezi délkou článku a počtem komentářů
 - vykreslete koláčový graf zobrazující podíl článků v jednotlivých kategoriích
 - vykreslete histogram pro počet slov v člancích
 - vykreslete histogram pro délku slov v člancích
 - vykreslete časovou osu zobrazující výskyt slova koronavirus v názvu článků
 - přidejte druhou křivku pro výraz vakcína
 - vykreslete histogram pro počet článků v jednotlivých dnech týdne

CVIČENÍ

- BONUS: NoSQL databáze pro data z webu [iDNES.cz](https://idnes.cz)
 - načtěte data do libovolné NoSQL databáze a pomocí dotazů
 - vypište jeden náhodný článek
 - vypište celkový počet článků
 - vypište průměrný počet fotek na článek
 - vypište počet článků s více než 100 komentáři
 - pro každou kategorii vypište počet článků z roku 2022