



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

TECHNOLOGIE PRO BIG DATA CVIČENÍ II. OPAKOVÁNÍ – PYTHON

Lukáš Matějů
04.10.2024 | TPB



CVIČENÍ

- analyzér dat z webu iDNES.cz
 - vypište počet článků
 - vypište počet duplicitních článků
 - vypište datum nejstaršího článku
 - vypište jméno článku s nejvíce komentáři
 - vypište nejvyšší počet přidanych fotek u článku
 - vypište počty článků podle roku publikace
 - vypište počet unikátních kategorií a počet článků v každé kategorii
 - vypište 5 nejčastějších slov v názvu článků z roku 2021
 - vypište celkový počet komentářů
 - vypište celkový počet slov ve všech člancích

CVIČENÍ

- bonusové úlohy
 - vypište 8 nejčastějších slov v článcích, odfiltrujte krátká slova (< 6 písmen)
 - vypište 3 články s nejvyšším počtem výskytů slova Covid-19
 - vypište články s nejvyšším a nejnižším počtem slov
 - vypište průměrnou délku slova přes všechny články
 - vypište měsíce s nejvíce a nejméně publikovanými články