

Analýza dat II

Číselné formáty v Excelu

- Obecné – výchozí formát; program může datový typ odhadnout
- Text – obsah buňky se zobrazí přesně tak, jak jste jej zapsali
- Číslo (a další formáty) – program chápe obsah jako číselnou hodnotu; můžete ale upravit způsob zobrazení
- Datum/čas – i to jsou formáty typu čísla

- Správné nastavení -> možnost práce s daty
- Cesta: Domů->číslo / Domů->číslo->formát->formát buněk /
Pravá myš->formát buněk

Funkce

- Zápis pomocí = na začátku obsahu buňky
- Základní matematické operace pomocí + - * /
- Spojovací textových řetězců pomocí &
- Složitější operace pomocí nabídky f_x vedle řádku vzorců
- Odkazy A1 vs. R1C1 (Cesta: Soubor->možnosti->vzorci->práce se vzorci->styl odkazu R1C1)
- Kopírování vzorců dodržuje relativní pozici buněk
- Možnost fixace odkazů pomocí \$ (přepínač F4)
- Při zapnutém filtru možnost odkazovat na celý sloupec [sloupec], případně psát vzorec pro celý sloupec aktualizovaný podle řádků [@sloupec] (ekvivalent zkopírovaného vzorce ze řádku do celého sloupce)

Funkce: příklady

- Načtěte datové soubory psrk.csv a psvolkr.csv ze studijních materiálů
- Funkce SUMA (sčítá hodnoty v dané oblasti)
 - `=SUMA(psrk[POCHLASU]) / =SUMA(psrk!P2:P5263)` – zápisy jsou ekvivalentní; funkce vrátí počet preferenčních hlasů udělených ve volbách do PS PČR v roce 2021
- Funkce SUMIF (sčítá hodnoty v dané oblasti, pokud buňky splňují zadané kritérium; kritérium se v tomto případě ověřuje v jiné oblasti, než je oblast hodnot)
 - `=SUMIF(psrk[MANDAT];"A";psrk[POCHLASU])` – vrátí počet preferenčních hlasů udělených zvoleným poslancům

Funkce: příklady

- Funkce KDYŽ (ověří splnění podmínky a vrátí buď hodnotu PRAVDA, nebo NEPRAVDA)
 - =KDYŽ([@MANDAT]="A";KDYŽ([@PORADIMAND]>[@PORCISLO];"SKOKAN";"POSLANEC");"KANDIDÁT") – do nového sloupce doplníte tímto vzorcem status politiků a političek podle toho, jak uspěli ve volbách; skokani jsou ti, kteří získali mandát a zároveň se díky preferenčním hlasům posunuli na kandidátce vzhůru (omlouvám se za generické maskulinum)
- Funkce INDEX (vrátí hodnotu v řádku a sloupci v dané oblasti)
 - =INDEX(psvolkr[#Vše];[@VOLKRAJ]+1;2) – do nového sloupce doplníte tímto vzorcem názvy krajů na základě jejich číselného kódu

Podmíněné formátování

- Grafické zvýraznění dat v tabulce na základě hodnot
- Cesta: Domů->styly->podmíněné formátování
- Lze kombinovat
- U podmínění vzorcem možnost ovlivnit celý řádek
 - oblast platnosti zapsaná bez souřadnic řádků
 - zamykání buněk v podmínce pomocí \$
- Logické operátory pro psaní podmínek A(vzorec) a NEBO(vzorec)
- Příklad
 - vzorec: `=A($R2="A";$T2>$D2)`, oblast: `=$2:$5263`
 - označuje řádky se skokany z předchozího příkladu (nepoužívá nově vzniklý sloupec)

Grafy

- Minigrafy
 - Cesta: Vložení->minigrafy
 - Vejdou se do jedné buňky; nemnoho nastavení
- Kontingenční graf
 - Cesta: Vložení->grafy->kontingenční graf (buď výběrem zdrojových dat, nebo existující kontingenční tabulky)
 - Logika stejná jako u kontingenční tabulky
- Jinak obecně přes kartu Vložení -> grafy
 - Přibudou záložky *návrh grafu, formát*
 - Možnost editovat legendu, název, popisky os apod.
 - Převrácení os, vypnutí zobrazení části dat

Pokročilé zdroje dat

získávání dat z webu

Pomocí Power Query

- Cesta: Data->načíst a transformovat data->z webu
- Možnost aktualizovat
- Url adresa datového souboru (csv, json apod.)
 - Stejně jako offline soubory
- Url adresa webové stránky
 - Program vyhledá zobrazené tabulky a nabídne je k importu

Datové typy

- Součást MS Excel
- Cesta: Data->datové typy
- Doplnující informace k vybraným heslům (zvířata, místa, akcie, apod.)
 - Hesla potřeba vepsat do buněk, poté přiřadit datový typ
 - Po najetí myši se objeví nabídka *Vložit data*
- Spíše anglicky
- Syntaxe podobná funkcím (lze kopírovat)

Web Scraping

- Když si data jdete vzít
- Etika?
 - veřejné informace
 - zatížení serveru
- Programovací jazyky (například Python, R)
- Simulace procházení celými weby
- Uživatelské alternativy (většinou placené)
 - např. zdarma doplněk prohlížeče www.webscraper.io
- Struktura HTML, classes, ids, XPath

Domácí úkol

- Načtěte online externí data (možnost aktualizace)
- Vizualizujte data pomocí podmíněného formátování za využití vzorce, aby se měnil formát celého řádku
- Vytvořte na základě dat smysluplný graf
 - zvolte vhodně typ, popisky, měřítko
- Odevzdejte do 16. listopadu 2021