



Datová Akademie



Obsah

1 ÚVOD DO DATOVÉ ANALÝZY	3
2 STRUKTURA SQL DOTAZU	3
3 AGREGACE	3
4 SPOJOVÁNÍ TABULEK	3
5 VIZUALIZAČNÍ NÁSTROJE	4
6 WINDOW FUNKCE	4
7 ZÁKLADY PANDASU	4
8 PANDAS VS. SQL, VIZUALIZACE	4
9 POKROČILEJŠÍ PANDAS	5
10 VIZUALIZACE POMOCÍ MATPLOTLIB	5
11 REGRESE A KLASIFIKACE	5
12 REKAPITULACE A PŘÍPADOVÁ STUDIE	5



1 Úvod do datové analýzy

V první lekci si představíme celý obsah kurzu, řekneme si něco o datech, databázích, datové analýze a její roli v současném světě. Dále se seznámíme s prostředím, ve kterém budeme pracovat v první polovině kurzu, a vyzkoušíme si první SQL dotazy.

2 Struktura SQL dotazu

Druhá lekce bude věnována základní formě SQL dotazu. Naučíme se vybírat data, řadit je, a také si vyzkoušíme jednoduché i složitější podmínky při dotazování. Pomocí těchto nástrojů se seznámíme s daty, se kterými budeme pracovat po většinu kurzu.

3 Agregace

Ve třetí lekci se budeme věnovat seskupování řádků v tabulkách. Ukážeme si, jak vybrat klíče pro seskupování. Zároveň se seznámíme s agregačními funkcemi, které můžeme při seskupování řádků používat. Dále si ukážeme složitější podmínkování.

4 Spojování tabulek

Čtvrtá lekce bude věnována spojování tabulek. Věnovat se budeme hlavně napojování přes sloupce tabulek. Představíme si druhy tzv. joinů a všechny si vyzkoušíme na příkladech. Krátce se budeme věnovat i druhému typu spojování, tzv. union.



5 Vizualizační nástroje

Pátou lekci věnujeme nástrojům vizualizace dat. Představíme si populární nástroje Tableau a Power BI. Pomocí znalostí z předchozích lekcí vytvoříme grafické reprezentace našich dat.

6 Window funkce

V šesté lekci se podíváme na pokročilejší funkce v SQL, tzv. window funkce. Tyto funkce jsou podobné jako seskupování tabulek, protože používají informace z více řádků tabulky zároveň. Na rozdíl od agregačních funkcí ale nezmenšují dimenzi dané tabulky. Window funkce si opět vyzkoušíme na praktických příkladech.

7 Základy pandasu

V sedmé lekci přejdeme od SQL ke druhému nejlepšímu příteli datového analytika, pandasu. Pandas je modul jazyka Python používaný k práci s daty. Na úvod se seznámíme s modulem numpy, na němž je pandas založen. Poté přejdeme k základnímu objektu pandasu - DataFrame. Naučíme se načítat data do DataFrame a ukážeme si, jak získat popisné statistiky dat.

8 Pandas vs. SQL, vizualizace

V osmé lekci si ukážeme, jak na DataFrame provádět operace, které už známe z SQL. Pandas přebírá celou řadu SQL funkcí, například seskupování tabulek nebo jejich spojování. Dále si ukážeme, jak z DataFrame snadno vytvářet grafy.



9 Pokročilejší pandas

V deváté lekci se zaměříme na další užitečné funkce pandasu. Pandas je velice obsáhlý modul disponující mnoha funkcemi, které ulehčují život datového analytika. Podíváme se, jak transponovat tabulky, jak pracovat s víceúrovňovými indexy nebo jak na DataFrame aplikovat vlastní funkce. Vše si ukážeme na praktických příkladech.

10 Vizualizace pomocí matplotlib

Obsahem desáté lekce bude vizualizační modul Pythonu, matplotlib. Ukážeme si, jak v matplotlibu vytvořit užitečné grafy, například spojnicový graf, koláčový graf nebo sloupcový graf. U každého grafu si řekneme, na jaký typ dat je vhodné jej použít. Poté motivujeme další možná použití matplotlibu.

11 Regrese a klasifikace

V jedenácté lekci si představíme některé nástroje pokročilejší datové analýzy. Pokročilí data analytici využívají matematické a statistické modely k získání informací z dat. Nejčastěji používané modely jsou lineární a logistická regrese. Vyjasníme si rozdíl mezi regresní a klasifikační úlohou a na názorných příkladech ilustrujeme oba typy úloh. Ukážeme si moduly v Pythonu, které implementují dané modely.

12 Rekapitulace a případová studie

V poslední, dvanácté lekci zopakujeme nejdůležitější části kurzu. Na základě vyřešených problémů si zopakujeme, co je úlohou analytika a co jsou jeho klíčové dovednosti. Na závěr si ukážeme case study z praxe, na které ilustrujeme, jak datový analytik přistupuje k řešení typických úkolů.

