



ENGETO

Datová Akademie

*Poslední aktualizace osnovy:
6. 11. 2020*



Obsah

1 Úvod do datové analýzy	3
2 Struktura SQL dotazu	3
3 Agregace	3
4 Spojování tabulek	3
5 Window funkce	4
6 Vizualizační nástroje	4
7 Základy pandas	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 ti představíme obsah kurzu, řekneme si něco o datech, databázích, datové analýze a její roli v současném světě. Seznámíš se s prostředím, ve kterém budeš pracovat v první polovině kurzu, a vyzkoušíš si jednoduché SQL dotazy.

2 Struktura SQL dotazu

Druhá lekce se bude věnovat základní formě SQL dotazu. Naučíš se vybírat data, řadit je a také si vyzkoušíš jednoduché i složitější podmínky při dotazování. Pomocí těchto nástrojů se seznámíš s daty, se kterými budeš pracovat většinu kurzu.

3 Agregace

Třetí lekce je zaměřená na seskupování řádků v tabulkách. Ukážeme ti, jak vybrat klíče pro seskupování. Zároveň se seznámíš s agregačními funkcemi, které můžeš při seskupování řádků používat. A ukážeme ti také složitější podmínkování.

4 Spojování tabulek

Se čtvrtou lekcí přichází na řadu spojování tabulek. Budeš řešit hlavně napojování přes sloupce tabulek. Představíme ti druhy tzv. joinů a všechny si vyzkoušíš na příkladech. Krátce se budeme věnovat i druhému typu spojování, tzv. union.



5 Window funkce

Podíváme se na pokročilejší funkce v SQL, tzv. Window funkce. Ty jsou podobné jako seskupování tabulek, protože používají informace z více řádků tabulky zároveň. Narozdíl od agregačních funkcí ale nezmenšují dimenzi dané tabulky. Window funkce si opět vyzkoušíš na příkladech.

6 Vizualizační nástroje

Šestou lekci věnujeme nástrojům na vizualizaci dat. Představíme ti populární nástroje Tableau a Power BI. Pomocí znalostí z předchozích lekcí si vytvoříš grafické reprezentace dat.

7 Základy pandas

V sedmé lekci tě čeká přechod od SQL ke druhému nejlepšímu příteli datového analytika – pandas. Pandas je modul jazyka Python používaný k práci s daty. Na úvod tě seznámíme s modulem numpy, na kterém je pandas založený. Poté přejdeme k základnímu objektu pandasu – DataFrame. Naučíš se načítat data do DataFrame a ukážeme ti, jak získat popisné statistiky dat.

8 Pandas vs. SQL, vizualizace

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



9 Pokročilejší pandas

V deváté lekci se zaměříme na další užitečné funkce pandasu. Pandas je velmi obsáhlý modul, který má spoustu funkcí. Ty ulehčují život datového analytika. Podíváme se na to, jak transponovat tabulky, pracovat s víceúrovňovými indexy nebo jak na DataFrame aplikovat vlastní funkce. Praktická cvičení nebudou chybět.

10 Vizualizace pomocí Matplotlib

Obsahem desáté lekce bude vizualizační modul Pythonu – Matplotlib. Zjistíš, jak v Matplotlibu vytvořit užitečné grafy, například spojnicový graf, koláčový graf nebo sloupcový graf. U každého grafu ti řekneme, na jaký typ dat je vhodné jej použít. A pak si povíme něco dalšího o použití Matplotlibu.

11 Regrese a klasifikace

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



12 Rekapitulace a případová studie

V poslední lekci zopakujeme nejdůležitější části kurz a co patří mezi klíčové dovednosti datového analytika. Na závěr ti ukážeme case study z praxe, na kterých uvidíš, jak datový analytik přistupuje k řešení typických úkolů.

