

Engeto Data Academy – SQL projekt

Sprievodná listina

Kristína Majerová

Prípravu podkladových tabuliek a odpovedí na výskumné otázky začínam analýzou podkladových tabuliek a dôsledným prečítaním výskumných otázok. Na základe výskumných otázok bude pre ďalšiu prácu optimálne pracovať s dátami na ročnej báze.

Príprava primary table

Z tabuľky Czechia_payroll vyberiem dáta o mzde prepočítané na plný úväzok za jednotlivé odvetvia

(podľa informácií z <https://www.mpsv.cz/statistika-prumerne-mzdy>

Průměrná hrubá měsíční mzda je publikována ve dvou podobách - (i) na fyzické osoby, kdy v čitateli mohou být i příjmy z vedlejšího pracovního poměru, částečného pracovního poměru apod.; (ii) na přepočtené počty, kdy se výpočet průměrné mzdy koriguje o vliv různých pracovních úvazků a poměrů (výpočet by měl odrážet průměrnou mzdu za plný pracovní úvazek).

Dáta o mzde za kvartály spriemerujem a zjednotím na ročnú bázu ako u ostatných dát, do tabuľky si doplním názvy jednotlivých odvetví z číselníka.

Pomocou karteziánskeho súčinu k tabuľke pripojím dáta o priemerných cenách potravín v ČR v jednotlivých rokoch a kategóriách potravín, doplnené o názvy kategórie. Priemerné dáta o cene za ČR majú v stĺpci region_code hodnotu NULL . (dôkaz v SQL)

Dáta o počte zamestnancov v danom období a a odvetví nie sú známe (ani pre všetky odvetvia, ani pre všetky obdobia) takže štatistika neberie do úvahy vážený priemer mzdy podľa toho, koľko pracuje v danom odvetví obyvateľov...

Rozhodla som sa priamo v tejto tabuľke pracovať s názvami kategórií, aby som ich pri následnej práci nepotrebovala už dosadzovať.

Príprava secondary table

Z tabuľky Economies budú pre prípravu secondary_table potrebné len dáta za ČR od roku 2005, HDP, a počet obyvateľov v jednotlivých rokoch.

V ďalších krokoch bude potrebné porovnanie medziročného rastu HDP s inými metrikami, pripravím si stĺpec medziročného rastu.

(Keďže mzdy a ceny potravín nie sú závislé od počtu obyvateľstva a HDP áno, budem pracovať s údajom HDP na 1 obyvateľa – per_capita, ktorý mi pre tento účel príde ako presnejší.)

Otázka 1

Rostou v průběhu let mzdy ve všech odvětvích, nebo v některých klesají?

Postup:

1. Pre prípravu odpovede ma budú zaujímať dáta po jednotlivých odvetviach v jednotlivých rokoch (keďže je tabuľke je karteziánsky súčin a hodnoty pre odvetvie a rok sú v tabuľke viac krát, pre platy v odvetví v danom roku použijem priemer)
2. vypočítam tempo medziročného rastu.
3. Zaujímanú nás odvetvia, kde je medziročný pokles = rast je nižší ako nula
4. Ako výsledok dostanem zoznam rokov, odvetví a výšku poklesu :

Odpoveď:

Áno sú viaceré odvetvia ktoré majú medziročný pokles miezd, najvýznamnejšie sú :

- Těžba a dobývání v 4 obdobiach
- Výroba a rozvod elektřiny, plynu, tepla a klimatiz. Vzduchu v 3 obdobiach

Najhorší bol rok 2023, kedy poklesli mzdy v 11 sledovaných obdobiach

Otázka 2

Kolik je možné si koupit litrů mléka a kilogramů chleba za první a poslední srovnatelné období v dostupných datech cen a mezd?

Postup:

1. získam priemernú cenu za 2 sledované potraviny, v prvom a poslednom roku . Prvý a posledný rok získam ako min a max hodnoty stĺpca price_year
2. k získaným hodnotám dosadím priemernú mzdu (ako priemer zo všetkých odvetví) v danom roku
3. hodnoty priemernej mzdy a priemer ceny potraviny vydelím

Odpoveď:

V roku 2006 je možné kúpiť za priemernú mzdu 1 312 kg chleba alebo 1 465 l mlieka

V roku 2018 je možné kúpiť za priemernú mzdu 1 365 kg chleba alebo 1 669 l mlieka

(Var. 2 je riešenie je s výsledkami po jednotlivých odvetviach)

Otázka 3

Která kategorie potravin zdražuje nejpomaleji (je u ní nejnižší percentuální meziroční nárůst)?

Postup:

1. zistíme medziročné tempo rastu cien pre každú skupinu potravín
2. výsledok získame ako priemer hodnôt v jednotlivých rokoch

Odpoveď :

banány žlté majú najnižší medziročný rast ceny
U cukru a červených jablák cena medziročne klesá.

Otázka 4

Existuje rok, ve kterém byl meziroční nárůst cen potravin výrazně vyšší než růst mezd (větší než 10 %)?

Postup:

1. Vypočítam si medziročný rast miezd (počítaný z priemeru miezd všetkých sledovaných odvetví v jednotlivých rokoch)
2. Vypočítam si medziročný rast cien (počítaný z priemeru cien všetkých sledovaných potravín v jednotlivých rokoch)
3. Tieto 2 hodnoty od seba odčítam
4. Vyselektujem hodnoty spĺňajúce kritérium > 10

Odpoveď:

V žiadnom roku nebol medziročný rast cien potravín výrazne vyšší ako rast miezd (viac ako o 10%). Najvyšší rozdiel bol v roku 2013, kedy ceny potravín vzrástli o 6,66% viac ako mzdy

Otázka 5

Má výška HDP vliv na změny ve mzdách a cenách potravin? Neboli, pokud HDP vzroste výrazněji v jednom roce, projeví se to na cenách potravin či mzdách ve stejném nebo následujícím roce výraznějším růstem?

1. V rovnakom roku :

- 1.1 pripravím si view s hodnotami medziročných rastov cien potravín, miezd a HDP
- 1.2 riadky zoradím podľa rastu HDP, od najvyššej hodnoty, pretože nás zaujímajú hlavne roky s výrazným rastom HDP
- 1.3 vizuálnym porovnaním je možno povedať, že v rokoch, kedy HPD výraznejšie vzrástlo, vzrástli aj mzdy / ceny (2007, 2017, 2018)

2. V nasledujúcom roku

- 2.1 pre porovnanie súvislosti rastu HDP s ukazovateľmi v nasledujúcom roku, si „posuniem“ hodnotu rastu DPH do nasledujúceho roku a pozriem v vzťahy medzi číslami v „nasledujúcom roku“ od rastu HDP
- 2.2 riadky zoradím podľa rastu HDP, od najvyššej hodnoty, pretože nás zaujímajú hlavne roky s výrazným rastom HDP
- 2.3 vizuálnym porovnaním je možno povedať, ak HDP výraznejšie vzrástlo, (2006, 2007, 2017) vzrástli aj mzdy / ceny v nasledujúcich rokoch (nasledujúce roky = 2007, 2008, 2018)

Bez možnosti výpočtu korelácie je odpoveď len interpretáciou údajov zobrazených vo výstupe SQL

hodnoty korelace:

- rast HDP a medziročný rast cien v aktuálnom roku – 0,48 - stredná
- rast HDP a medziročný rast miezd v aktuálnom roku – 0,38 – slabá

- rast HDP a medziročný rast cien v nasledujúcom roku – 0,12 – slabá až žiadna
- rast HDP a medziročný rast miezd v nasledujúcom roku – 0,71 – vysoká