**Engeto - Projekt z SQL**

Zpracování tohoto projektu je jedním z úkolů kurzu Datový analytik s Pythonem od Engeto akademie. Týká se zpracování dat ze zadaných datových sad a na tomto základě odpovědět na výzkumné otázky.

Je třeba připravit dva datové podklady, ze kterých se bude následně čerpat:

1. Porovnání dostupnosti potravin na základě průměrných příjmů v České republice za určité časové období.
2. Tabulku s HDP, GINI koeficientem a populací dalších evropských států ve stejném období, jako primární přehled pro ČR.

Následně pomocí těchto sad zodpovědět na následující výzkumné otázky:

1. Rostou v průběhu let mzdy ve všech odvětvích, nebo v některých klesají?
2. Kolik je možné si koupit litrů mléka a kilogramů chleba za první a poslední srovnatelné období v dostupných datech cen a mezd?
3. Která kategorie potravin zdražuje nejpomaleji (je u ní nejnižší percentuální meziroční nárůst)?
4. Existuje rok, ve kterém byl meziroční nárůst cen potravin výrazně vyšší než růst mezd (větší než 10 %)?
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?

**Příprava datových podkladů**

**Vytvoření tabulky project\_SQL\_primary\_final:** Pro vytvoření tohoto datového podkladu jsem využil primární tabulky *czechia\_payroll* a *czechia\_price*. Každou jsem seskupil na stejné období (stejné roky) a pomocí příkazu UNION je spojil do jedné tabulky. Obsahuje hodnoty pro jednotlivé roky, filtrování bude prováděno pomocí sloupce value\_type\_code.

**Vytvoření tabulky project\_SQL\_secondary\_final:** Jako základ tohoto podkladu byla použita tabulka *countries*, filtrovaná pouze na země v Evropě. Ke každému státu jsem pomocí JOIN připojil hodnoty HDP, GINI a populace z tabulky *economies* pro stejné roky, jako v první tabulce.

**Odpovědi na výzkumné otázky**

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

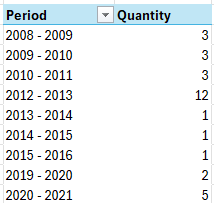
Ano, v průběhu let mzdy v některých odvětvích klesaly.

Záznamy obsahují údaje o mzdách mezi lety 2000 a 2021 (21 období). Jsou rozděleny podle různých odvětví (20 odvětví). Celkem se tedy jedná o 420 záznamů, z nichž byl u 30 zaznamenán meziroční pokles mezd.

## Detaily:

Nejvíce období zaznamenalo pokles v období 2012 – 2013 (12 odvětví). Patrně do způsobila finanční krize v předchozích letech.

Největší pokles byl zaznamenán ve stejném období v oboru Peněžnictví a pojišťovnictví – meziroční pokles 8,91%



## Postup získání dat:

Vytvoření tabulky mezd zobrazující v jednom řádku mzdy pro každý rok včetně roku následujícího (pomocí LEAD).

K tabulce jsem pomocí JOIN připojil názvy jednotlivých odvětví.

Vytvořil dva nové sloupce:

* payroll\_difference: rozdíl mezi mzdami.
* difference\_percent: procentuální změna mzdy

Pro výslednou tabulku jsem vyfiltroval pouze záporné hodnoty (poklesy mezd).

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

První srovnatelné období je rok 2006, poslední rok 2018.

V roce 2006 bylo možné za průměrnou mzdu koupit 1192 kilogramů chleba, v roce 2018 1300 kilogramů.

Za průměrnou mzdu bylo možné v roce 2006 kopit 1330 litrů mléka. V roce 2018 už 1590 litrů.

Obsah obrázku text, snímek obrazovky, Písmo, číslo

Popis byl vytvořen automaticky

## Postup získání dat:

Využitím Common Table Expression jsem si vytvořil pomocnou tabulku obsahující ceny chleba a mléka v jednotlivých letech.

Následně jsem pomocí Window Functions MIN a MAX z této pomocné tabulky vybral jen hodnoty pro první a poslední srovnatelné období a připojil názvy zboží (mléko, chleba).

Pomocí JOIN jsem k těmto obdobím připojil výši průměrné mzdy.

Ve finální tabulce jsem mimo zobrazení nezbytných hodnot doplnil hodnotu podílu průměrné mzdy a ceny zboží zaokrouhlenou dolů na celé číslo.

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

Ve sledovaném období nejméně meziročně zdražila Rajská jablka červená kulatá – mezi roky 2006 a 2007 došlo ke snížení ceny o 30,28%. V roce 2006 byla průměrná cena 57,83 Kč/kg a roce 2007 se prodávala průměrně za 40,32 Kč/kg.

## Postup získání dat:

Využitím Common Table Expression jsem si vytvořil tabulku obsahující meziroční rozdíly cen jednotlivých kategorií potravin.

Z těchto hodnot jsem vypočítal meziroční cenový rozdíl a výsledek seřadil od nejnižšího.

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

Ano, takový rok existuje. Na základě sledovaných dat k tomuto došlo ve 33 případech. Největší rozdíl byl mezi průměrnou cenou papriky mezi roky 2006 a 2007 a změnou průměrné mzdy ve stejném období. Cena paprik vzrostla o 94,82 %, ale průměrná mzda vzrostla pouze o 7,23 %.

## Postup získání dat:

Nejprve jsem zjistil meziroční změnu průměrných cen jednotlivých potravin. Následně pro meziroční změnu průměrných mezd. Obě hodnoty jsem vyjádřil jako procentní změnu.

Pro jednotlivé roky a potraviny jsem vytvořil sloupec s rozdílem výše uvedených procentních změn a seřadil od nejnižšího rozdílu.

# 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?

Pro řešení úkolu byly k dispozici data z let 2006 – 2018. Pro zjištění meziročních změn cen potravin byl k dispozici pouze datový model poskytující ceny potravin, ale už ne jejich vážený průměr (množství nakupovaných potravin).

Meziroční změna HDP má určitý vliv na meziroční změnu cen a mez ve stejném nebo následujícím roce. Nicméně tento vliv není zcela jednoznačný a neprojevuje se pokaždé. Zejména ceny jdou v některých letech proti směru vývoje HDP. Mzdy se více přibližují změnám v HDP – nejsou tak rozkolísané.

Obecně tedy lze konstatovat, že HDP má na ceny a mzdy vliv spíše z delšího časového období, kdy se obě veličiny pokouší přibližovat k trendu změn HDP. Toto ilustruje níže uvedený graf vytvořený na základě získaných dat.

## Postup získání dat:

S využitím scriptů z úkolů, jsem vytvořil dotazy pro získání procentních rozdílů HDP, cen a mezd přepočítaná na stejná období. Tyto dotazy jsme spojit a získal tak výsledek, který zobrazuje všechny potřebné hodnoty v jedné tabulce.

Pro lepší vyhodnocení jsem data překopíroval do MS Excel a vytvořil z nich graf.