Q&A to key questions (including code description and abnormalities found)

1 question: Wages are increasing over time in all industries (Agriculture, forestry and fishing, Mining and quarrying, Human health and social work activities, ect).
The data shows a positive trend across all sectors: from 2000 to 2021, the average gross wage has grown in each industry. While there may be small fluctuations in certain years, the overall direction is upward — indicating that all industries are thriving in terms of wage development.

Code description:

- The year ('payroll_year') and industry ('industry_name') are selected because we want to observe how wages develop over time across different sectors.
- The average wage is calculated from the values in the value column for each year and industry.
- The function 'ROUND(..., 2)' is used to round the result to two decimal places for clarity. However, AVG(value) returns a double precision, and PostgreSQL cannot round a double precision with a second argument (like 2) unless you cast it to numeric.
- The data is filtered to include only wage records, specifically those referring to the average gross wage per employee. It excludes food price records from the result.- Rows where the value is 'NULL' are excluded because we can't calculate an average with missing data.
- Lastly, the data is grouped by year and industry, and sorted the final results.

2 question: An empty table with the following column names: payroll_year, average_gross_wage, product_type, product_price_total, product_unit, units_affordable. Highly probably that the result is empty due to a mismatch in the assumption that both wage and food data (for milk and bread) exist in the same years. The final SQL query joined wages and food data by year, looking specifically for the:

First year: 'MIN(payroll_year)' with average wage.
Last year: 'MAX(payroll_year)' with average wage.

But upon inspection, the years in which wage data is available do not overlap with the years of the selected food items (milk and bread).
Additionally, no records were found where: value_type_name = 'Průměrná hrubá mzda na zaměstnance' AND payroll_year IN (first AND last year of wages) AND milk or bread exist for those years.

Code description:

- The following columns in 'SELECT' are chosen to: compare two specific years the first and last available, show how much food a person can buy for the wage, the product itself, total price in CZK, and the quantity.
- 'ROUND……' gives the result how many units of food can be bought with the wage.
- Since the primary table contains two types of data in one place (wage data and food price data) , they are joined on the same year using aliases. It lines up each year's wage with the food prices from the same year.

- ‘WHERE’ clauses ensure that the left side only includes wage data, the right side includes only milk and bread rows. The output is restricted to only the first and last available years in the dataset.

3 question: We analyzed the average year-over-year percentage price increase for each food product included in the dataset. The calculation was based on unit prices and compared across multiple years. The product with the slowest price increase over time is: Pork with bone (Vepřové maso s kostí). Average annual price growth: 0.10 %.

There is also a note that in some years (for example, for potatoes or onions), there was indeed a price decrease compared to the previous year. However, these decreases were not significant enough to affect the overall average year-on-year growth of the food category across the entire analyzed period.

Code description:

- The average percentage change in price per unit is calculated, and the final result is rounded to 2 decimal places.
- The clean food name by removing "Food: " from the beginning of the string is extracted, and the unit price of the food is calculated.
- ‘PARTITION BY’ is used to calculate the unit price from the previous year separately for each product.- ‘ORDER BY’ gets the previous year’s value in order.
- The function ‘LAG()’ gives the value from 2020 for the same product.
- The first year for each product is removed, because there’s no previous year to compare to (‘LAG()’ is ‘NULL’).

4 question: When comparing the average year-over-year growth in gross wages and food prices between 2006 and 2018, we found out that there is no year, where food prices grew more than 10 percentage points faster than wages. The result was an empty table with the following column names: payroll_year, wage_growth_percent, food_price_growth_percent, and difference_percent, confirming that the condition was not met in any year. Therefore, the conclusion is - such a year does not exist in the available data.

Code description:

- Only wage records are filtered. The average wage for each year is calculated and rounded to 2 decimal places.
- All food rows are taken: ‘value_type_name LIKE 'Food:%", and the unit price is calculated. As the result, we get the average food price per year. The outcome is rounded and grouped by year.This gives us one average food price per year.
-Wage and food price tables are joined by year using the function ‘LAG(...)’ to grab the value from the previous year. As the result year-over-year growth is calculated, and its done for both wages and food prices.
- ‘ROUND(((average_wage - wage_previous_year) / wage_previous_year) * 100, 2)’
- Wage growth %, food price growth %, and the difference between them are computed. This tells us if food prices rose faster than wages and by how much.
- The condition is set to food prices grew more than 10 percentage points faster than wages. Since we get an empty result, it means no such year exists in our data.

5 question: Based on the available datasets, the answer to the question is the following: the tables (neither primary nor secondary) do not contain the necessary data about all European countries. The tables 'countries' and 'economies' which are the source for the secondary table contain country name, region, continent, GDP and population per year. They do not include wages, prices or any other domestic economic indicators.
These tables are useful for comparing economic size and growth, but they do not provide the other data needed to correlate GDP with wages or food prices. Although our secondary dataset includes economic data for many countries, only the Czech Republic provides a complete picture including wages and food prices. Therefore, this analysis focuses for Czechia only. Based on the results GDP growth does sometimes correlate with wage and food price growth, but the relationship is not always direct or automatic. In some years, a strong GDP growth led to higher wages and higher food prices (year 2007: GDP grew significantly: +6%. Food prices also increased noticeably: +6%). Here we see a strong positive connection: growth of GDP → higher wages and higher food prices). In other years, even with GDP growing, wages or food prices grew only slightly or even stagnated (year 2010: GDP grew by +2% , wages increased very little — only about +2% . Food prices also barely increased by around +2%. Even though GDP increased, wages and food prices reacted only slightly). Thus, GDP growth alone is not a perfect predictor for immediate wage or food price changes.

Code description:

- Base GDP data and previous year's GDP calculations.
- GDP growth percentage calculation (rounds the result to 2 decimal places).
- Average gross wage for each year from the primary table is calculated, with the data grouped by year.
- 'LAG()' is used to get previous year's average wage.
- Wage growth percentage is calculated, rounds to 2 decimal places. The first year is excluded because there is no previous wage to compare.
- Average price per unit for food items is calculated.
- 'LAG()' is used to get previous year's average food price.
- Food growth is calculated, rounds to 2 decimal places excluding the first year with no previous price.
- GDP growth, wage growth, and food price growth for each year are taken. 'LEFT JOIN' connects GDP with wages and food prices, sorting the results by year.

Abnormalities:

- A lot of 'NULLs'. Real wage and food price data start around 2000 or later. Before 2000, there are no available records in the table for average gross wage or food items and prices. GDP data is available from 1990s, but the project only contains wages and food prices for roughly 2000–2021.