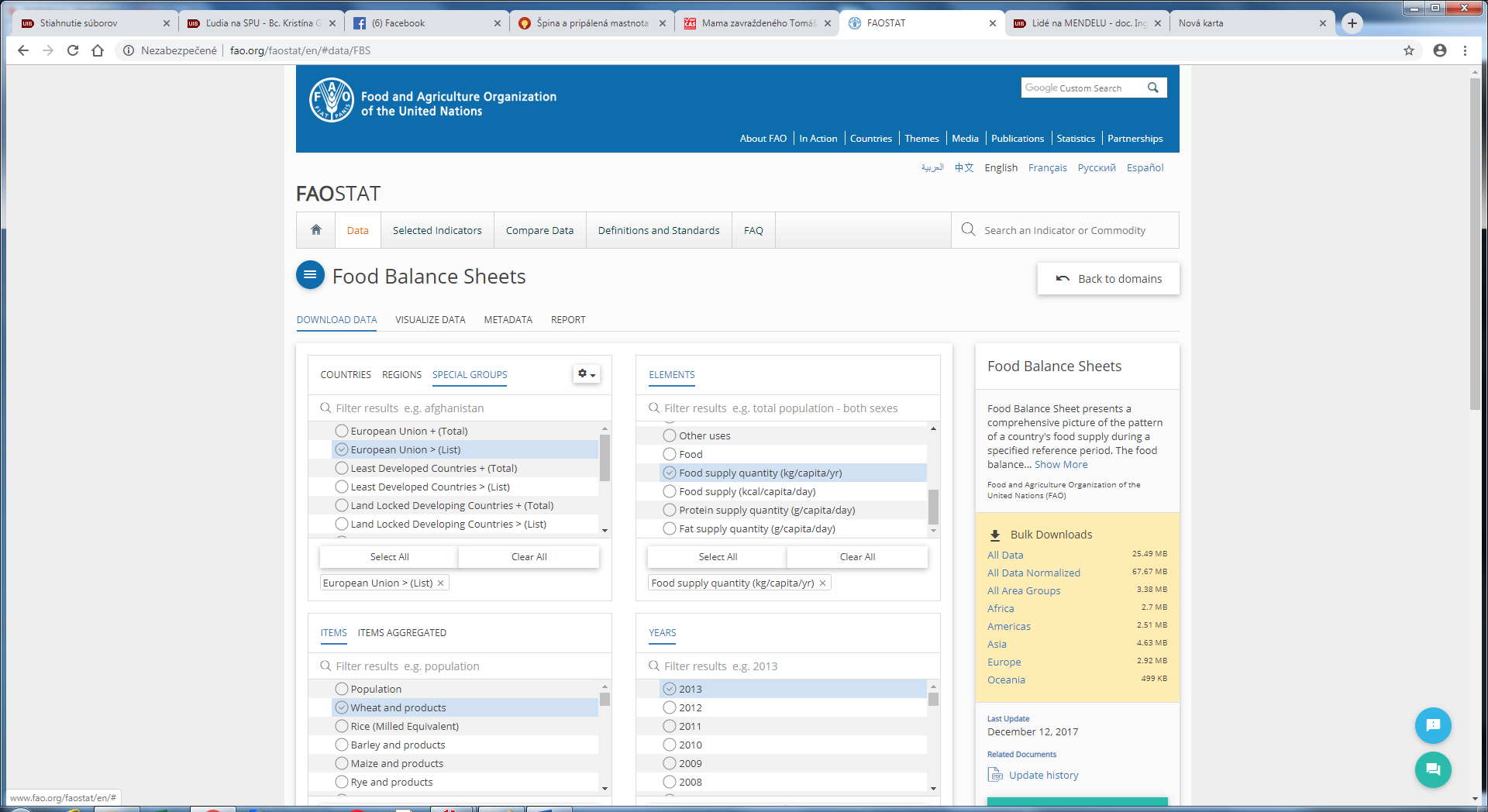
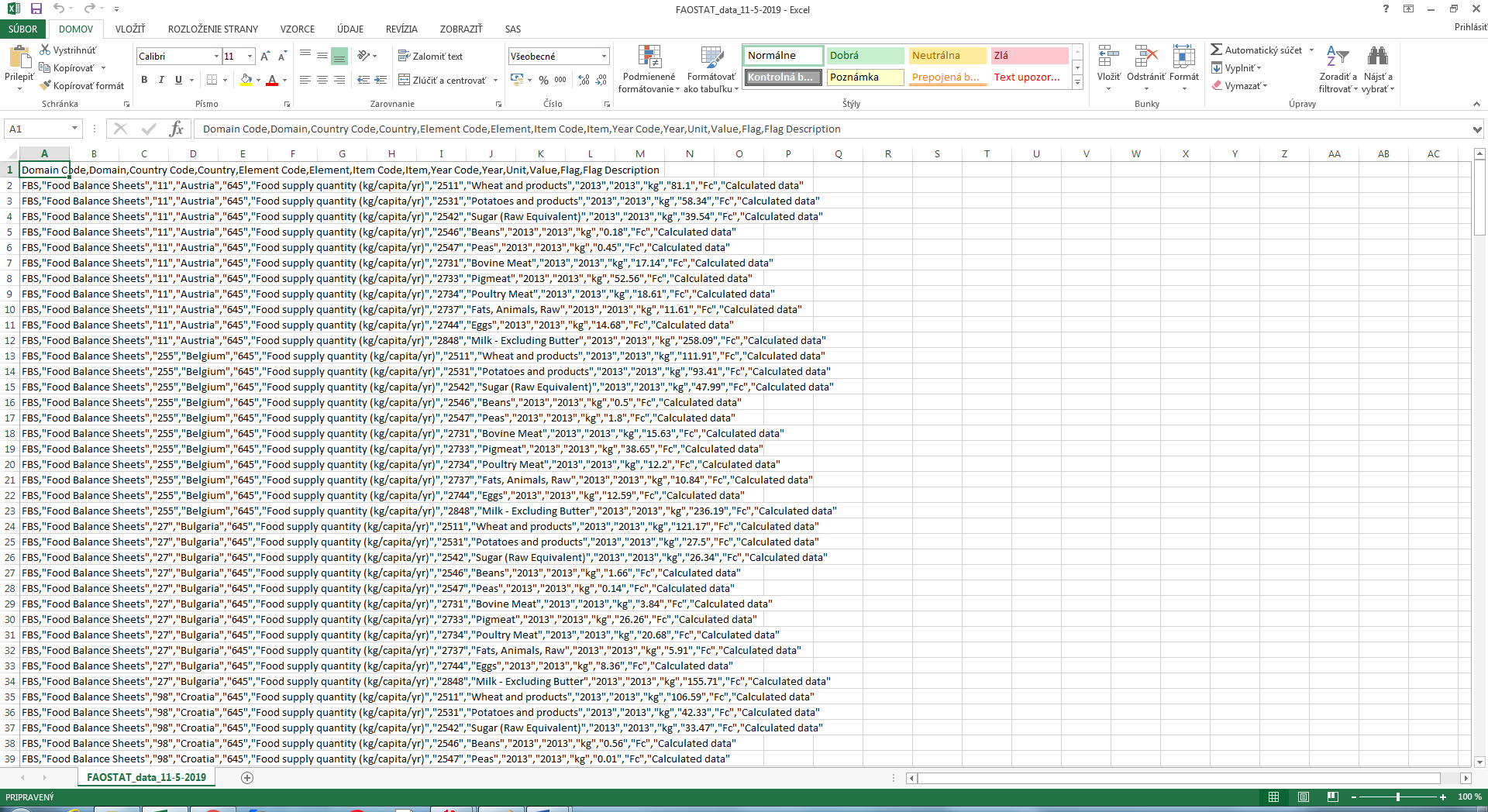
**Zadanie:** Aplikovať PCA analýzu na reálnych dátach o spotrebe potravín vo vybranom región (napr. krajiny Európy, krajiny EÚ, krajiny sveta, najvyspelejšie krajiny atď). Spotrebu potravín hodnotiť na základe vybraných min. 10 potravinových skupín (napr. spotreba mäsa, mlieka, syrov, vajec, zeleniny, ovocia, ....).Použiť databázu organizácie FAO.

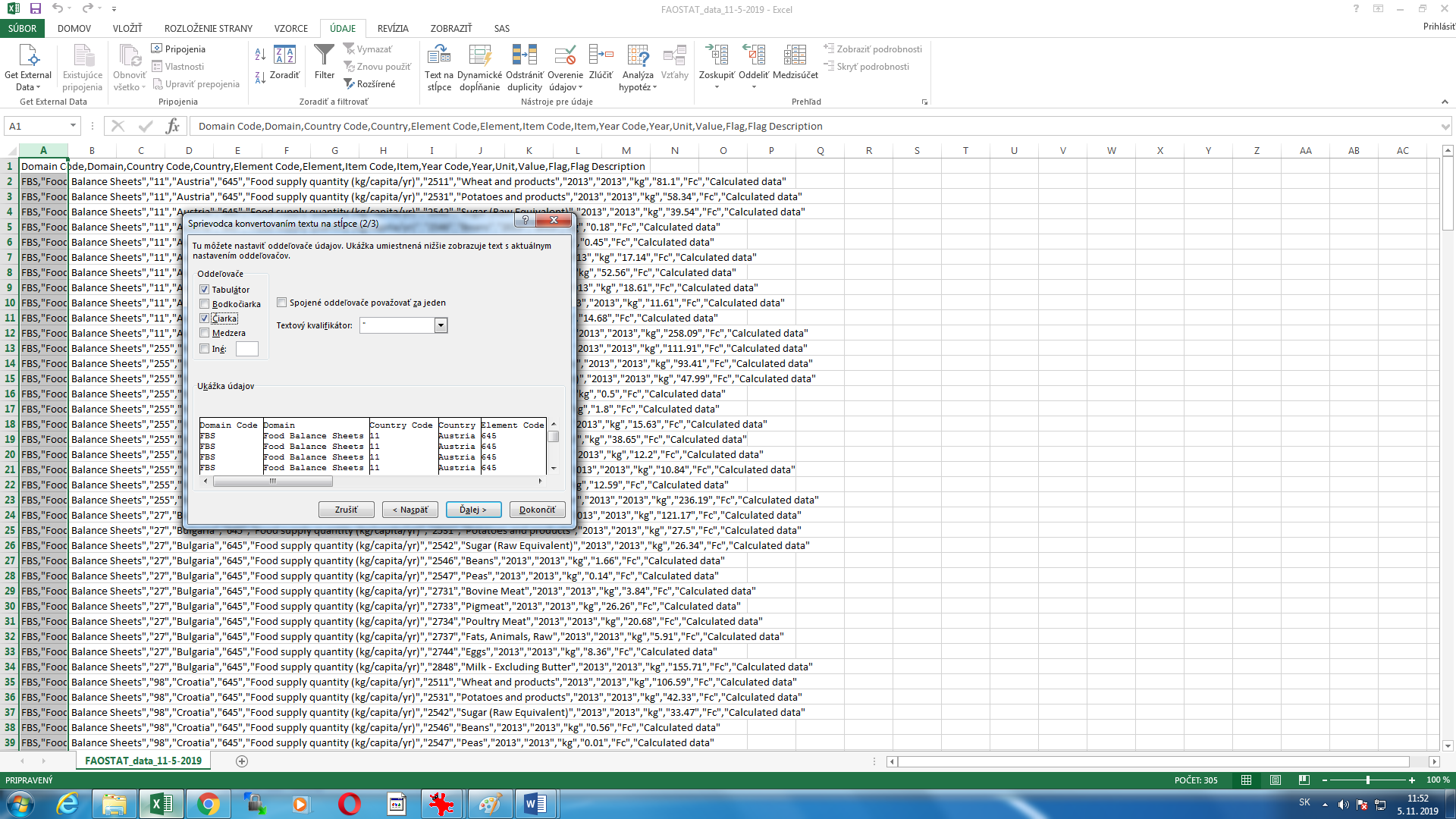
**POSTUP získania databázy:**

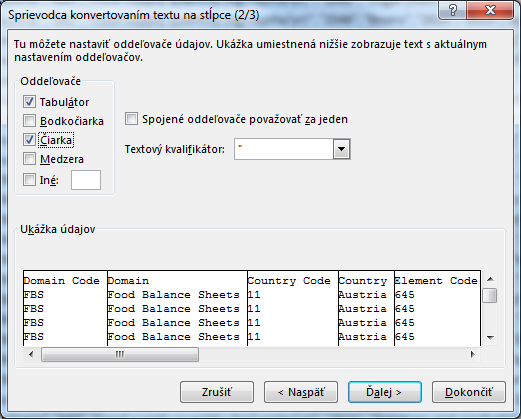
1. Použiť databázu FAO – faostat <http://www.fao.org/faostat/en/#home> 
2. Vybrať DATA/FOOD BALANCE/FOOD BALANCE SHEETS (údaje o spotrebe potravín)
3. Vyfiltrovať si údaje podľa vlastného uváženia (v položke Elements vybrať Food supply quantitiy (kg/capita/yr), ostatné je možnosť zvoliť si, t.j. napr. vybrať krajiny EU (EU total List), vybrať si min. 10 potravinových skupín (ITEMS) a zvoliť si jeden rok (YEARS). Pri výbere komodít vyberať také, medzi ktorými predpokladáme existenciu vzťahu.
4. Stiahnuť si údaje do excelu Download Data (údaje sú stiahnuté ako csv subor, t.j. textový súbor, ktorý je potrebné následne si upraviť na vstupnú databázu).



***Úprava stiahnutého súboru do databázy:***

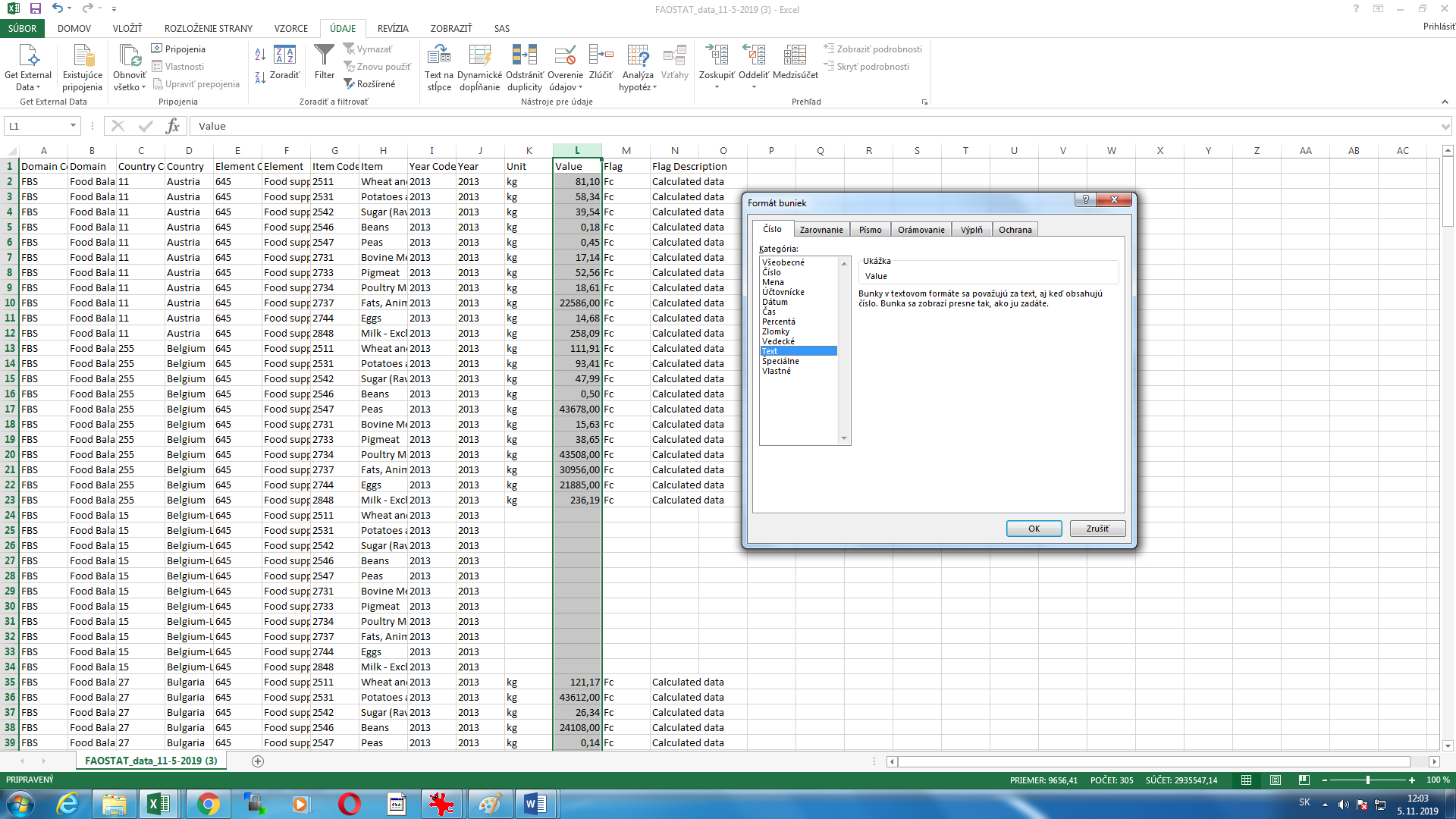
1. Vysvietiť v exceli celý prvý stĺpec a vybrať Údaje/text na stĺpce, označiť oddeľovač **čiarka**

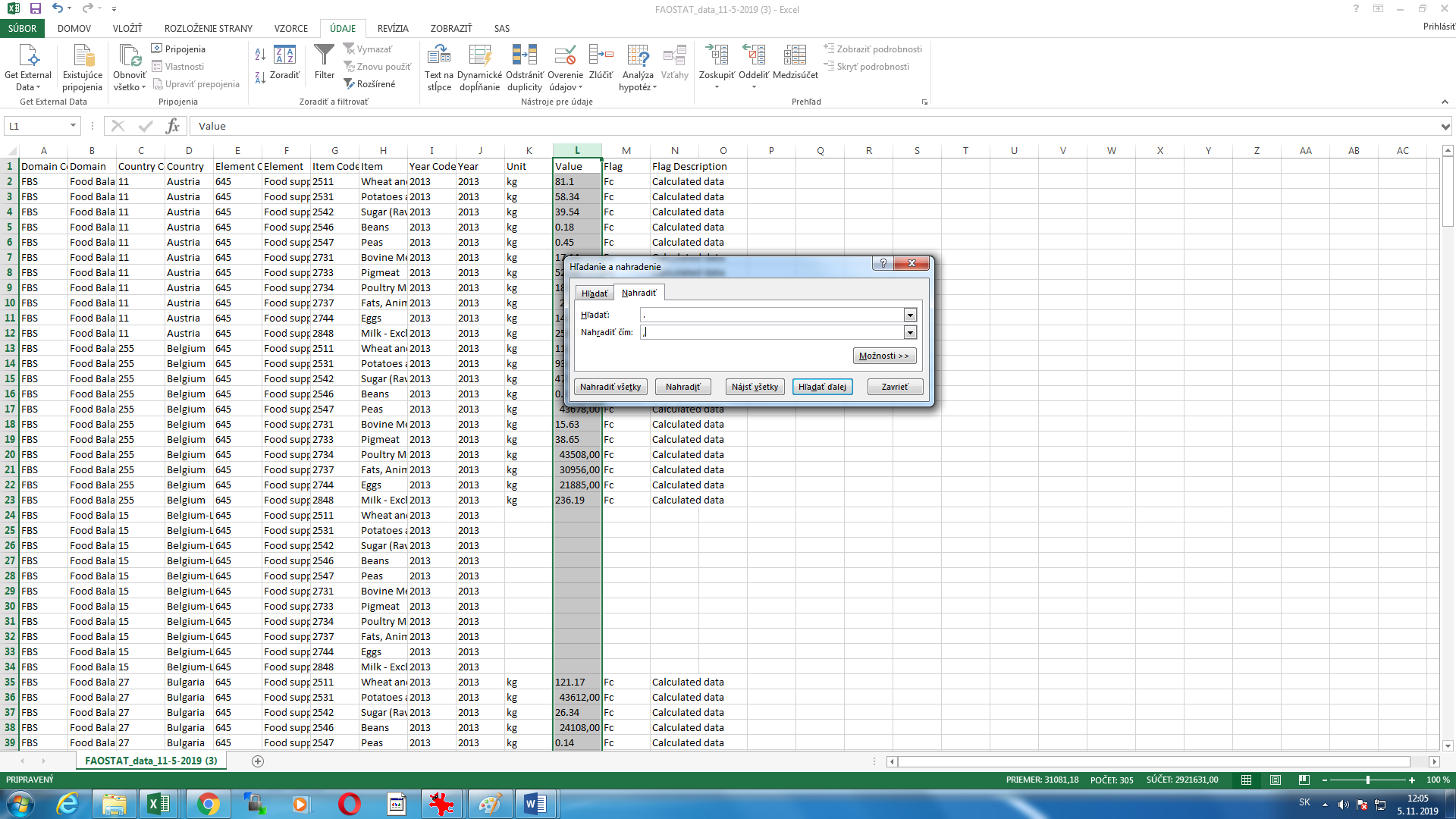


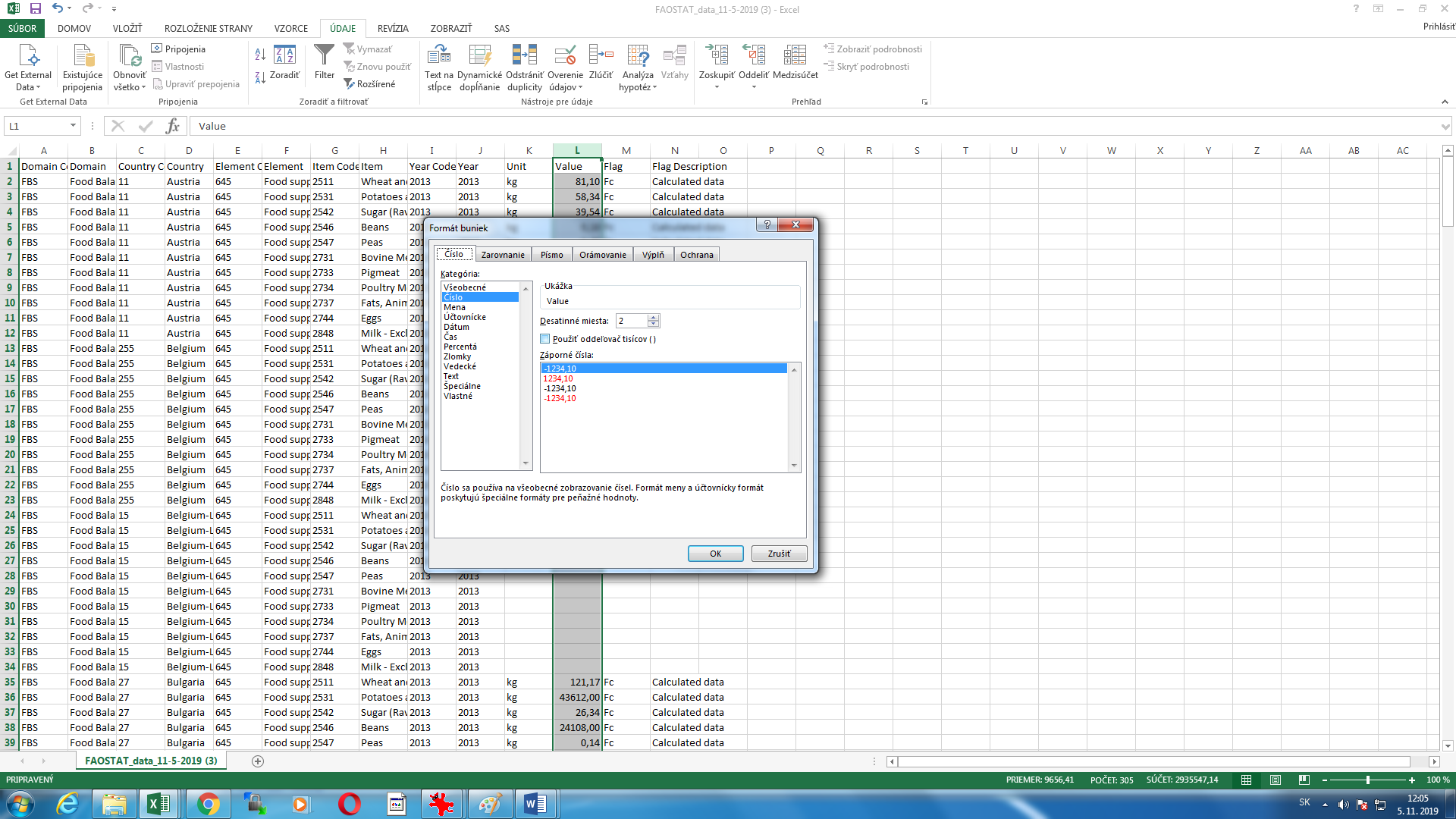


Pozn: Máte vytvorené samostatné texty v samostatných stĺpcoch, problém je však v tom, že Vám pri prenose ako oddeľovač desať. miesta dá bodku a nie čiarku, pozor – niektoré čísla môže transformovať na formát dátumu, preto je potrebné urobiť nasledovný postup:

* Vysvietiť stĺpec Value a cez formát bunky nastaviť najprv formát TEXT
* Nechať vysvietený celý stĺpec a použiť kombináciu klávec Ctrl+H (hľadať a nahradiť) – nechať vyhľadať bodku (.) a nahradiť ju čiarkou (,) – nahradiť všetky záznamy v celom riadku.
* Opäť nechať vysvietený celý stĺpce Value a zmeniť formát bunky na formát čísla.

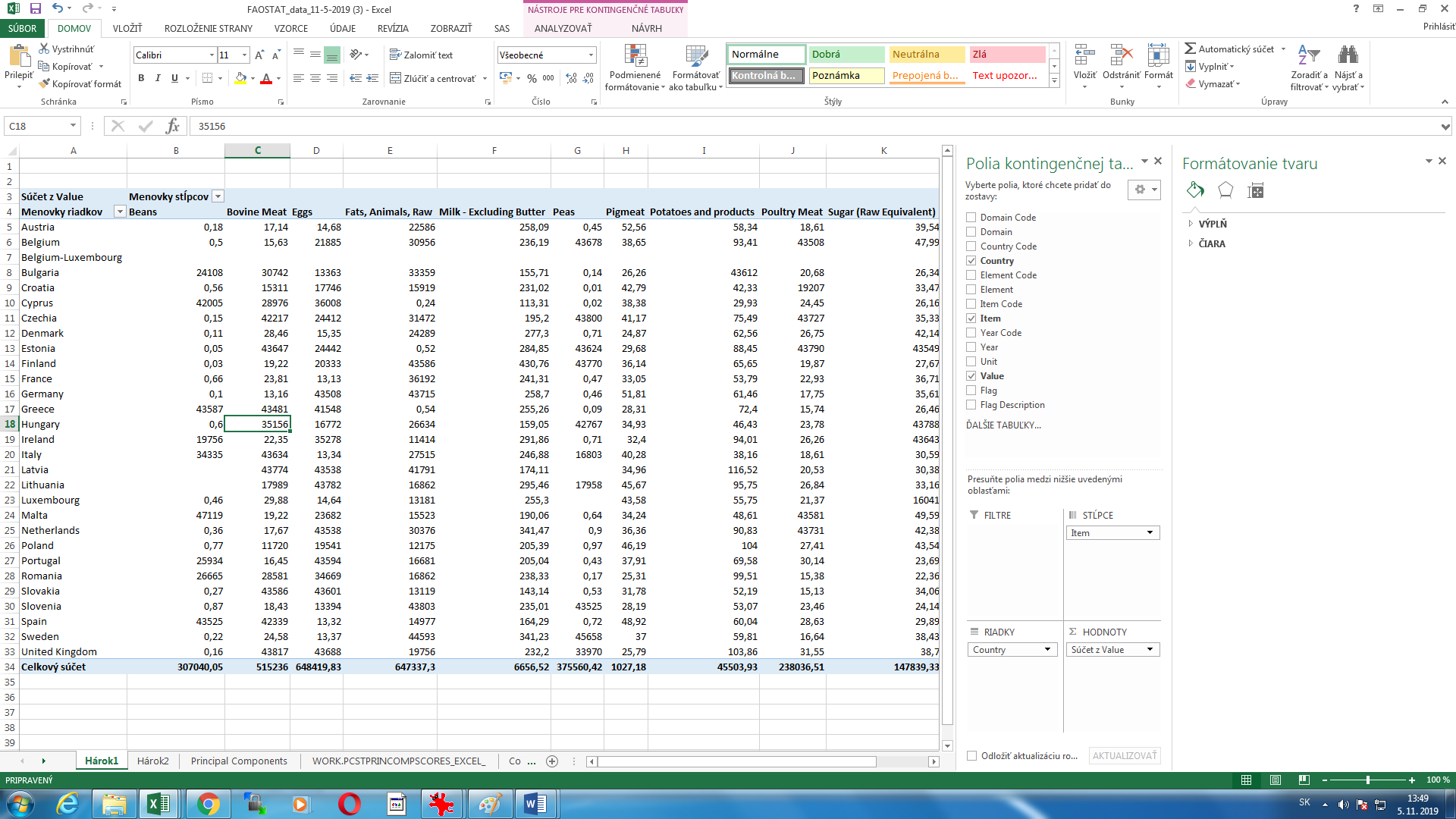






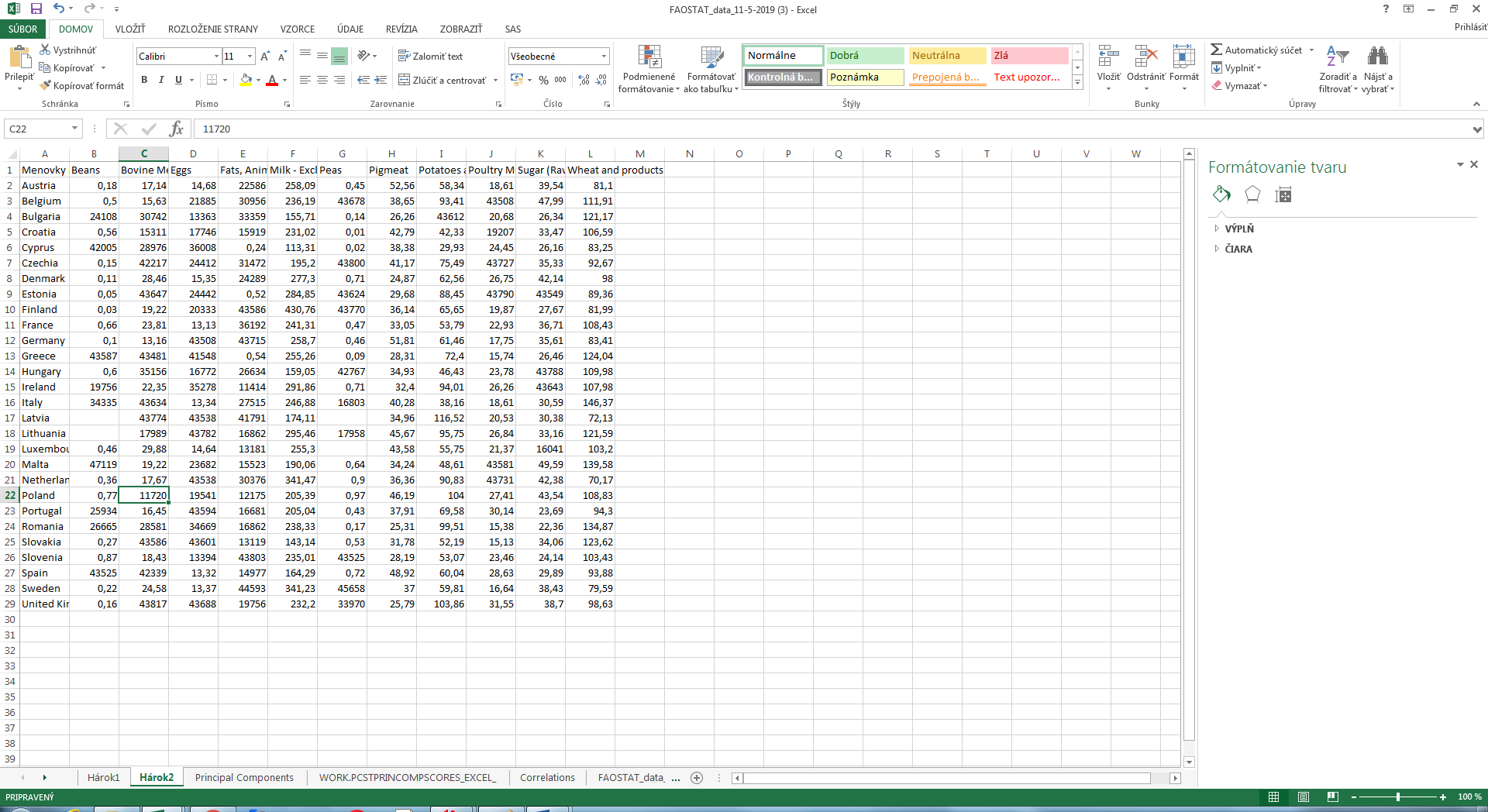
**POSLEDNÝ KROK:**

Nakoniec je potrebné upraviť si získané údaje do matice nxp, t.j. v riadku je potrebné mať krajiny a v stĺpci premenné. Najjednoduchšie je použiť Pivot table – do riadkov dáte country , do stĺpca items, a do poľa hodnoty dáte Value a nastavíte operáciu Súčet.



Následne si databázu prekopírujete – vysvietite databázu od bunky A4 (menovky riadkov) po celkový súčet po riadok 33, použijete Ctrl+C a možnosť prilepiť špeciálne – len hodnoty.

Výsledná databáza, ktorá bude vstupovať do PCA, vyzerá nasledovne:



Na vašich reálnych údajoch je potrebné urobiť PCA analýzu:

1. Urobiť korelačnú analýzu na overenie existencie závislosti
2. Urobiť PCA, určiť počet významných komponentov, interpretovať vlastné vektory aj cez tabuľku Eigenvector aj cez korelačné koeficienty medzi pôvodnými premennými a vlastnými číslami.
3. Vytvoriť grafy a okomentovať ich.
4. Vytvoriť stĺpcovné grafy podla PRIN1 z usporiadaných hodnôt komponentového skóre, na os y dať názvy krajín, okomentovať grafy.

(v skratke: urobiť to isté, čo sme spolu robili na cvičení, len na vašich dátach).