

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Analýza dat pomocí Power Pivot, Power Query a Power Map

1

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Datová analýza v MS Excel

Power Query	Úprava nahrávaných dat před importem do Power Pivot (filtrování, nahrazování a úprava hodnot)
Power Pivot	Uložení nahraných dat do operační paměti a následné výpočty
Power View	Vizualizace dat systémovými vizuály či vizuály staženými z Marketplace

www.ictpro.cz

2

2

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Power Query

www.ictpro.cz 3

3

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Power Query

- Slouží pro úpravu dat z datového zdroje před jejich importem do datového modelu
- ETL operace:
 - filtrování dat
 - dopočítávání chybných dat , ošetřování chyb, duplicit apod.
 - změna struktury dat
 - spojování dat z více zdrojů
- Dopočítávání nových sloupců, řádků

www.ictpro.cz 4

4

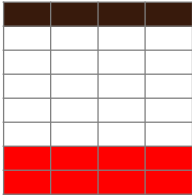
Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

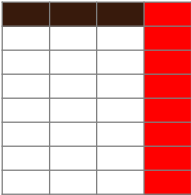
Základní operace na úrovni řádků a sloupců

- Příkazy v rámci uživatelského rozhraní
- Umělá inteligence (sloupce z příkladů)
- Jazyk M

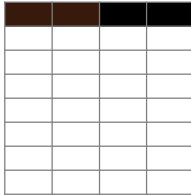
Filtrování nadbytečných řádků



Filtrování nadbytečných sloupců



Dopočtení chybějících sloupců



www.ictpro.cz
5

5

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Použitý postup

- Lze o něm přemýšlet jako o „záznamu“ našich kroků nad daty
- V rámci použitého postupu lze:
 - měnit pořadí
 - přejmenovávat
 - odstranit jednotlivé kroky
 - vkládat nový krok mezi již vytvořené kroky
- Pokud je k danému kroku přidán komentář na úrovni zdrojového kódu, tento komentář se objeví v popisu kroku

www.ictpro.cz
6

6

Anatomie použitého postupu

Klíčové slovo uvozující začátek postupu

Jednotlivé kroky zpravidla navazují na sebe, přičemž prvním argumentem metody nového kroku je název předchozího kroku

```
let
    #excel data = Csv.Document(File.Contents("C:\Users\data1.csv"),[Delimiter=";", Columns=7, Encoding=1250, QuoteStyle=QuoteStyle.None]),
    #Headers / zhlavi = Table.PromoteHeaders(#excel data), [PromoteAllScalars=true],
    // Zdroj obsahuje nekolik sloupce, ktere nejsou treba pro tvorbu reportu, byly tedy odstraneny, zachovany zustanou pouze nize uvedene sloupce
    // The datasource contains multiple columns that are unnecessary for the report so they are removed, only below mentioned columns are kept
    #vyber sloupce / keep columns = Table.SelectColumns(#Headers / zhlavi,{"TIME", "GEO", "Value"}),
    #nahrad hodnoty / replace values = Table.ReplaceValue(#vyber sloupce / keep columns, ":", " ", Replacer.ReplaceValue, {"Value"}),
    #zmen typ / change types = Table.TransformColumnTypes(#nahrad hodnoty / replace values,{{"TIME", Int64.Type}, {"GEO", type text}, {"Value", Int64.Type}})
in
    #zmen typ / change types
```

Výpočet
neovlivňující
komentář

Klíčové slovo uvozující konec postupu, za slovem „in“ následuje návratová hodnota

Jednotlivé kroky použitého postupu jsou odděleny čárkou s výjimkou kroku předcházejícímu klíčovému slovu in

Odebrání řádků

- V záhlaví sloupce jednoho či více sloupců tabulky (podobně jako v MS Excel)
- V hlavní navigaci v kartě Domů / Home části Odebrat řádky / Remove rows
- Možnosti odebrání řádků:
 - na základě toho, jestli hodnota splňuje podmínku
 - odebrání horních x či spodních x řádků
 - odebrání střídavých řádků
 - odebrání duplicit
 - odebrání prázdných řádků
 - odebrání chyb

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Odebrání sloupců

- Po označení kurzorem a vyvolání kontextové nabídky
- V části hlavní navigace Domů / Home Odebrat sloupce / Remove columns

Odebrání ostatních sloupců

C	D

Odebrání sloupců

A	D

www.ictpro.cz
9

9

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Agregace a transformace dat

- Agregace data
 - Na základě hodnot jednoho či více sloupců
 - Výsledek obsahuje agregované hodnoty, přičemž lze nad jedním sloupcem použít více agregačních funkcí
- Transformace dat
 - Převedení sloupců do řádků (tzv. unpivotování)
 - Převedení sloupců do řádků (tzv. pivotování), kdy výsledná tabulka obsahuje agregované hodnoty tabulky před pivotováním (podobně jako v kontingenční tabulce v MS Excel)

→

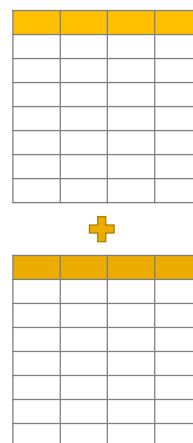
↔

www.ictpro.cz
10

10

Spojení dat z více zdrojů – spojení pod sebe

- Dva a více zdrojů jsou spojeny pod sebe
- Nezáleží na pořadí sloupců a jejich typu
- Záleží na názvu sloupců
- Lze připojovat data do již existující tabulky
- Výsledek je rovněž možné nechat vypsát jako novou tabulku



Spojení dat z více zdrojů – spojení vedle sebe

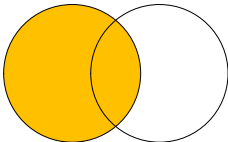
- Spojení tabulek „vedle sebe“ (lze chápat jako dopočítávání sloupců do MS Excel tabulky)
- Výsledek je rovněž možné nechat vypsát jako novou tabulku
- V obou spojovaných tabulkách musí existovat sloupec či kombinace sloupců, přes které lze data dohledat
- Existuje několik možností toho, jaká data jsou na výstupu, viz níže

Analýza dat pomocí Power Pivot, Power Query a Power Map

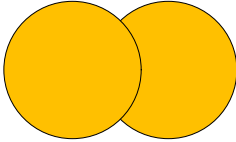
ictPRO

Spojení dat z více zdrojů – spojení vedle sebe

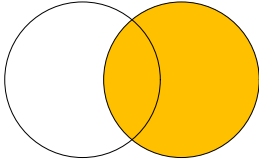
Spojení zleva



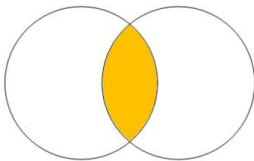
Vnější spojení



Spojení zprava



Vnitřní spojení



www.ictpro.cz

13

13

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Základní operace nad datovými typy

- V rámci již existujících sloupců, tj. upravujeme existující hodnoty a počet sloupců zůstává zachován (část hlavní navigace Transformace / Transform)
- V rámci přidání nového sloupce, tj. výsledné hodnoty jsou přidány do nového sloupce (část hlavní navigace Přidání nového sloupce / Add columns)

A
1.1.2023
2.1.2023
3.1.2023
4.1.2023
5.1.2023

→

A
2023
2023
2023
2023
2023

A	B
1.1.2023	2023
2.1.2023	2023
3.1.2023	2023
4.1.2023	2023
5.1.2023	2023

www.ictpro.cz

14

14

7

Základní operace nad datovými typy

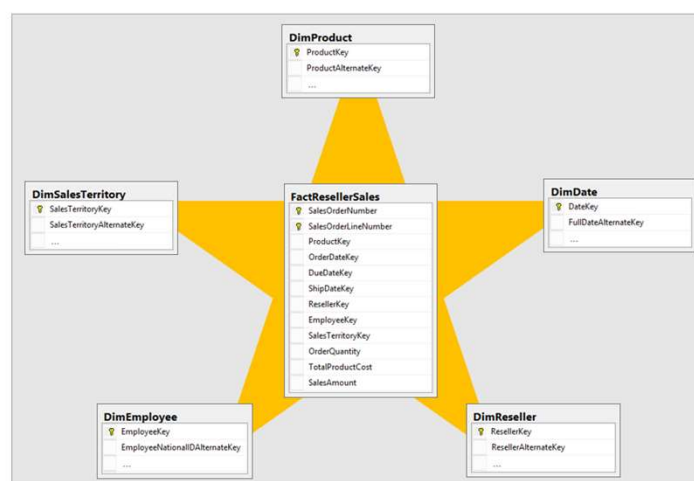
- Text
 - Odstranění nadbytečných znaků (mezery, netisknutelné znaky)
 - Úprava velikosti písmen (vše malé, vše velké, velká počáteční písmena slov)
 - Extrakce znaků dle pozice či oddělovače
- Čísla
 - Základní matematické operace (sčítání, odečítání, násobení, dělení apod.)
 - Zaokrouhlování
 - Datum
 - Dopočtení hodnoty z datumu (poslední datum či první datum v měsíci, čtvrtletí, roce apod.)
 - Extrakce hodnoty z datumu (rok, čtvrtletí, měsíc, den)

Datový model a relace

Datové modelování

- Kritická část tvorby reportu, která rozhodne o tom:
 - jestli budeme vůbec schopni vytvořit dotaz vracející správné hodnoty
 - jestli budeme schopni výpočet dokončit s ohledem na dostupnost operační paměti

Datové modelování – schéma hvězdy



Relace

- Základní možnosti propojení tabulek relacemi:
 - tabulky nejsou propojeny
 - tabulky jsou propojeny neaktivní relací, která je zapínána v rámci výpočtu
 - tabulky jsou propojeny aktivní relací
- Rozdělení v závislosti na směru:
 - jednosměrné, kdy jedna tabulka filtruje druhou a ne naopak
 - obousměrné, kdy se tabulky ovlivňují navzájem

Relace

- Rozdělení relací v závislosti na kardinalitě:
 - **Relace M:1 (*:1):** Sloupec v dané tabulce může obsahovat více než jeden výskyt nějaké hodnoty a druhá související obsahuje jen jeden výskyt nějaké hodnoty.
 - **Jedna k jedné (1:1):** V relaci 1:1 obsahuje sloupec v jedné tabulce jen jeden výskyt konkrétní hodnoty a jiná související tabulka obsahuje jen jeden výskyt konkrétní hodnoty.
 - **Jedna k mnoha (1:*):** V relaci 1:N obsahuje sloupec v jedné tabulce jen jeden výskyt konkrétní hodnoty a jiná související tabulka může obsahovat více než jeden výskyt nějaké hodnoty.
 - **Mnoho až mnoho (*:*):** U složených modelů můžete mezi tabulkami vytvořit relaci M:N, která odstraňuje požadavky na jedinečné hodnoty v tabulkách.

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Power Pivot

www.ictpro.cz

21

21

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Power Pivot

- Nástroj pro práci s velkým množstvím dat
- Data jsou uložena v operační paměti
- Po importu z Power Query, tabulky jsou rozděleny na jednotlivé sloupce, komprimovány a následně uloženy
- K datové analýze slouží jazyk DAX
- Při analýze pracujeme se sloupci tabulek uložených v operační paměti, tyto objekty lze chápat jako databázové tabulky, případně jako Excel tabulky

www.ictpro.cz

22

22

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

DAX

www.ictpro.cz

23

23

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Syntaxe jazyka DAX

- Nástroj pro práci s velkým množstvím dat
- Data jsou uložena v operační paměti
- Po importu z Power Query, tabulky jsou rozděleny na jednotlivé sloupce, komprimovány a následně uloženy
- K datové analýze slouží jazyk DAX
- Při analýze pracujeme se sloupci tabulek uložených v operační paměti, tyto objekty lze chápat jako databázové tabulky, případně jako Excel tabulky

www.ictpro.cz

24

24

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Operátory

Typ operátoru	Symbol	Použití
Aritmetické operátory	+	Sčítání
	-	Odčítání
	*	Násobení
	/	Dělení
	^	Umocnění
Operátory porovnávání	=	je rovno
	>	Větší než
	<	Menší než
	>=	Větší či rovno než
	<=	Menší či rovno než
	<>	Nerovná se
Logické operátory	&&	A zároveň
		Nebo

www.ictpro.cz

25

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Použité funkce

Typ funkce	Funkce a syntaxe	Použití
Základní agregační funkce (Agregace při respektování všech filtrů)	AVERAGE()	Vrátí aritmetický průměr čísel ve sloupci
	COUNT()	Vrátí počet neprázdných řádků ve sloupci
	MAX()	Vrátí největší hodnotu ve sloupci či mezi dvěma skalárními výrazy
	MIN()	Vrátí nejmenší hodnotu ve sloupci či mezi dvěma skalárními výrazy
	SUM()	Vrátí součet všech čísel
Podmíněné agregační funkce (Agregace při ovlivnění filtrů, např. za pomoci funkce ALL(), FILTER())	AVERAGEX(tabulka,výraz)	Vrátí aritmetický průměr čísel ve sloupci, který je druhým argumentem funkce (výraz) tabulky, která je prvním argumentem funkce.
	COUNTX(tabulka,výraz)	Vrátí počet neprázdných řádků ve sloupci, který je druhým argumentem funkce (výraz) tabulky, která je prvním argumentem funkce.
	MAXX(tabulka,výraz)	Vrátí maximální hodnotu ve sloupci, který je druhým argumentem funkce (výraz) tabulky, která je prvním argumentem funkce.
	MINX(tabulka,výraz)	Vrátí minimální hodnotu ve sloupci, který je druhým argumentem funkce (výraz) tabulky, která je prvním argumentem funkce.
	SUMX(tabulka,výraz)	Vrátí součet číselných hodnot ve sloupci, který je druhým argumentem funkce (výraz) tabulky, která je prvním argumentem funkce.

www.ictpro.cz

26

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Použité funkce

Typ funkce	Syntaxe	Použití
Filtrovací funkce	ALL(TabulkaNeboSloupec)	Vrátí všechny řádky v tabulce nebo všechny hodnoty ve sloupci a ignoruje přitom všechny použité filtry. Tato funkce je užitečná pro mazání filtrů a vytváření výpočtů na všech řádcích v tabulce.
	FILTER(tabulka,výraz)	Vrátí tabulku, která představuje podmnožinu jiné tabulky nebo výrazu.
	SELECTEDVALUE(sloupec, alternativní hodnota)	Vrátí hodnotu, pokud byl kontext pro columnName vyfiltrován pouze na jednu jedinečnou hodnotu. V opačném případě vrátí alternativní výsledek.
	VALUES(TabulkaNeboSloupec)	Pokud je vstupním parametrem název sloupce, vrátí tabulku s jedním sloupcem, která obsahuje jedinečné hodnoty ze zadaného sloupce. Duplicitní hodnoty se odeberou a vrátí se jen jedinečné hodnoty. Je možné přidat hodnotu BLANK. Pokud je vstupním parametrem název tabulky, vrátí řádky ze zadané tabulky. Duplicitní řádky jsou zachovány. Je možné přidat prázdný řádek.
Časové funkce	DAY(datum)	Vrátí den v měsíci, tedy číslo od 1 do 31.
	MONTH(datum)	Vrátí měsíc ve tvaru čísla od 1 (leden) do 12 (prosinec).
	QUARTER(datum)	Vrátí rok data jako čtyřmístné celé číslo v rozsahu 1900–9999.
	YEAR(datum)	Vrátí rok data jako čtyřmístné celé číslo v rozsahu 1900–9999.
Ostatní	BLANK()	Vrátí prázdnou hodnotu.
	IF(podmínka, hodnota pravdy, hodnota nepravdy)	Zkontroluje podmínku a vrátí první hodnotu, pokud se podmínka vyhodnotí jako TRUE. V opačném případě vrátí druhou hodnotu.

www.ictpro.cz

27

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Názvy parametrů funkcí

Pojem	Definice
expression	Libovolný výraz DAX, který vrací jednu skalární hodnotu a který se má vyhodnotit několikrát (pro každý řádek/kontext)
hodnota	Libovolný výraz DAX, který vrací jednu skalární hodnotu a který se má vyhodnotit právě jednou před všemi ostatními operacemi
tabulka	Libovolný výraz DAX, který vrací tabulku dat.
tableName	Název existující tabulky pomocí standardní syntaxe jazyka DAX. Nemůže se jednat o výraz.
columnName	Název existujícího sloupce pomocí standardní syntaxe jazyka DAX, obvykle plně kvalifikovaný. Nemůže se jednat o výraz.
name	Řetězcová konstanta, která se použije k zadání názvu nového objektu.
pořadí	Výčet sloužící k určení pořadí řazení.
vazby	Výčet sloužící k určení zpracování hodnot vazeb.
typ	Výčet sloužící k určení datového typu pro PathItem a PathItemReverse.

www.ictpro.cz

28

Logika práce s funkcemi

- Pokud je v závorkách uvozujících začátek a konec funkce uvedena hodnota, funkce vyžaduje vstupní hodnotu, tzv. parametr.
- Počet parametrů se může lišit.
- Parametry jsou povinné či volitelné (volitelné parametry jsou dále uvozeny hranatými závorkami)

Kalkulované sloupce

- Využívají tzv. kontextu řádku, kdy výpočet je proveden pro každý jeden řádek tabulky
- Data jsou uložena v operační paměti a aktualizována při aktualizaci datového zdroje
- Je vhodné je použít pouze v těchto případech:
 - výsledek chceme použít do průřezu či osu vizuálu
 - vytváříme segmenty dat
 - výsledek skutečně potřebujeme pro každý jeden řádek

Kalkulované míry

- Využívají tzv. kontextu filtru, kdy reagují na filtry aplikované vizuály
- K aktualizaci dochází při změně kontextu filtru, například interakci uživatele s průřezem

Závorky v DAX

- Hranaté závorky pro identifikaci sloupce, např. zápis `data[Value]` odkazuje na tabulku `data` a její sloupec `[Value]`
- Kulaté závorky pro funkce, např. `SUM()` volá funkci pro součet hodnot ve sloupci
- Složené závorky pro vytvoření množiny, např. `{2022,2023}` vytváří množinu obsahující hodnoty 2022 a 2023

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Závorky v DAX - příklad

```
1 value 2019 cesko =
2
3 // krok4: Funkce SUMX() prochází řádky tabulky vracené funkcí FILTER() a následně
4 // agreguje hodnoty ve sloupci [Value]
5 SUMX(
6     //krok3: Funkce filtr vrací novou tabulku
7     FILTER(
8         // krok1: Funkce filtr přebírá tabulku data se všemi filtry, které jsou aplikovány
9         data,
10        // krok2: Funkce filtr aplikuje další filtry
11        data[TIME] = 2019
12        && data[Country] = "Česko"
13    ),
14    data[Value]
15 )
```

Čárky oddělují jednotlivé parametry funkcí SUMX(FILTER(table,condition),expression)

Každá funkce je uvozena levou kulatou závorkou a uzavřena pravou kulatou závorkou

www.ictpro.cz

33

33

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Příklad výpočtu v jazyku DAX

```
1 01 Oblast rok 2020 =
2
3
4 SUMX(
5     --krok3: funkce FILTER() vrací tabulku
6     FILTER(
7         --krok1: Funkce FILTER() převezme tabulku filtrovanou vizuály
8         cesko,
9         --krok2: Funkce FILTER() dále aplikuje filtr nad sloupcem
10        cesko[Rok] = 2020
11    ),
12    --krok4: Pro každý jeden řádek tabulky vracené
13    -- funkci FILTER() je použita funkce agregace
14    -- funkce suma
15    cesko[Hodnota]
16 )
```

Jednotlivé argumenty jsou odděleny čárkou (v USA lokalizaci středníkem)

Funkce začíná a končí kulatou závorkou

www.ictpro.cz

34

34

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Příklad výpočtu v jazyku DAX

```
4 SUMX(
5
6 --krok4: Funkce FILTER() filtruje tabulku vrácenou
7 -- funkci ALL() a vrací tabulku
8 FILTER(
9 --krok3: Funkce ALL() vrátil tabulku
10 -- se všemi jejími hodnotami
11 ALL(cesko),
12 cesko[Rok]=
13 -- krok2: Funkce MINX() najde minimální
14 -- hodnotu v tabulce vrácenou funkcí ALL()
15 MINX(
16 --krok1: Funkce ALL() vrátil tabulku
17 -- se všemi jejími hodnotami
18 ALL(cesko),
19 cesko[Rok])),
20 -- funkce SUMX() prochází tabulku vrácenou
21 -- funkcí FILTER() a počítá hodnoty ve sloupci
22 cesko[Hodnota])
```

	data	

	ALL(cesko)	

	ALL(cesko)	

Funkce ALL() vytváří kopii tabulky data se všemi jejími hodnotami. Původní tabulka data zůstává zachována

Funkce ALL() vytváří kopii tabulky data se všemi jejími hodnotami. Původní tabulka data zůstává zachována

www.ictpro.cz

35

35

Analýza dat pomocí Power Pivot, Power Query a Power Map

ictPRO

Proměnné

- Zlepšují výkon
- Zpřehledňují kód
- Zjednodušují ladění chyb
- Jsou definovány klíčovým slovem VAR
- Požadavky na název jsou:
 - podporovaná sada znaků: a–z, A–Z, 0–9
 - znaky 0–9 nejsou platné jako první znak
 - jako předpona názvu identifikátoru je povoleno _ (dvojitě podtržítko)
 - nejsou podporované žádné jiné speciální znaky
 - nejsou povolená rezervovaná klíčová slova
 - nejsou povolené názvy existujících tabulek
 - nejsou povolené prázdné mezery

www.ictpro.cz

36

36

Proměnné

- V rámci skriptu nelze měnit jejich hodnotu
- V rámci jednoho skriptu můžeme mít více proměnných
- Nemůže na ně být odkazováno z jiných kalkulovaných měř či sloupců

Pozn. Možné řešení je použití celé kalkulované míry jako proměnné

- Pokud pracujeme s proměnnými, musíme použít klíčové slovo RETURN pro určení návratové hodnoty