

ОСНОВНІ ПОНЯТТЯ

Інформація та дані



Дані

Це інформація, якою можна поділитись і яку можна обробити, в першу чергу машинним (автоматичним) способом



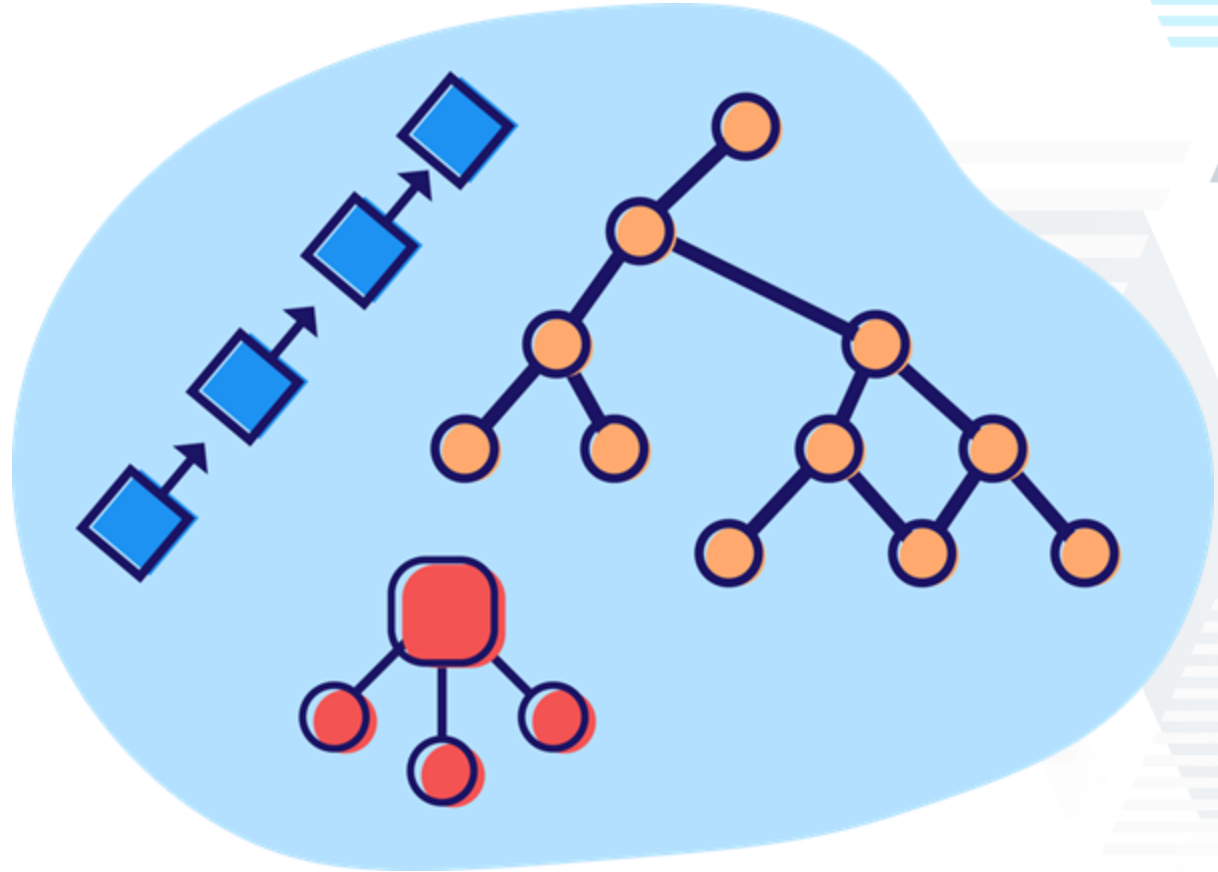
Набір даних

Це сукупність однорідних значень (записів) даних та метаданих, що їх описують



Дані бувають:

- Структуровані
- Неструктуровані
- Напівструктуровані



Неструктуровані дані



Текстові файли та документи



Медичні записи



Веб-сайти



Зображення



Відео



Аудіо



Email



Соціальні мережі

**Деякі з наведених даних можуть бути напівструктурованими*

Неструктуровані дані можна перетворити у структуровані

«Сьогодні, 17 лютого 2021 року, депутати нашого міста прийняли кілька важливих рішень. Зокрема, за проєкт ухвали про перейменування вулиці Леніна на вулицю Тараса Шевченка проголосували 37 депутатів, за проєкт рішення №123 “Про перелік відкритих даних” – 45. Також, міська рада 38 голосами “за” ухвалила рішення щодо створення скверу на перетині вулиць Хмельницького та Івана Франка».

Дата	ID	Назва голосування	“за”
2018-02-17	122	Щодо перейменування вулиці Леніна	37
2018-02-17	123	Про перелік наборів відкритих даних	45
2018-02-17	124	Щодо створення скверу у Франківському районі	38

Не всі дані, що містяться у таблицях є структурованими!

БАЛАНС
на 01 січня 2019 року

Форма №1-дс

АКТИВ	Код рядка	На початок звітного періоду	На кінець звітного періоду
1	2	3	4
I. НЕФІНАНСОВІ АКТИВИ			
<i>Основні засоби:</i>	1000	53314477	52977102
<i>первісна вартість</i>	1001	55468732	55506450
<i>знос</i>	1002	2154255	2529348
<i>Інвестиційна нерухомість:</i>	1010	-	-
<i>первісна вартість</i>	1011	-	-
<i>знос</i>	1012	-	-
<i>Нематеріальні активи:</i>	1020	-	-
<i>первісна вартість</i>	1021	-	-
<i>накопичена амортизація</i>	1022	-	-
Незавершені капітальні інвестиції	1030	-	-
<i>Довгострокові біологічні активи:</i>	1040	-	-
<i>первісна вартість</i>	1041	-	-
<i>накопичена амортизація</i>	1042	-	-
Запаси	1050	110003	85879
Виробництво	1060	-	-
Поточні біологічні активи	1090	-	-
Усього за розділом I	1095	53424480	53062981

Правила впорядкованих даних

- Кожна змінна, яку ви вимірюєте, повинна бути в одному стовпці
- Кожне окреме спостереження цієї змінної – в окремому рядку
- Для кожного “виду” змінної має бути одна таблиця
- Якщо у вас є декілька таблиць – вони повинні включати ідентифікатор

country	year	cases	population
Afghanistan	1999	3775	19987071
Afghanistan	2000	5666	20593360
Brazil	1999	37737	17206362
Brazil	2000	80488	174504898
China	1999	212258	1272015272
China	2000	218766	1280428583

ЗМІННІ

country	year	cases	population
Afghanistan	1999	3775	19987071
Afghanistan	2000	5666	20593360
Brazil	1999	37737	17206362
Brazil	2000	80488	174504898
China	1999	212258	1272015272
China	2000	218766	1280428583

СПОСТЕРЕЖЕННЯ

country	year	cases	population
Afghanistan	99	75	987071
Afghanistan	00	66	593360
Brazil	99	737	17206362
Brazil	00	488	174504898
China	99	212258	1272015272
China	00	18766	1280428583

ЗНАЧЕННЯ

Порушення структури рядків та стовпців

неправильно

рік	чисельність наявного населення		
	всього	у тому числі	
		міське	сільське
1990	51838,5	34869,2	16969,3
1991	51944,4	35085,2	16859,2
1992	52056,6	35296,9	16759,7
1993	52244,1	35471,0	16773,1
1994	52114,4	35400,7	16713,7
1995	51728,4	35118,8	16609,6
1996	51297,1	34767,9	16529,2
1997	50818,4	34387,5	16430,9
1998	50370,8	34048,2	16322,6
1999	49918,1	33702,1	16216,0
2000	49429,8	33338,6	16091,2

правильно

рік	тип населення	кількість
1990	міське	34869.2
1990	сільське	16969.3
1991	міське	35085.2
1991	сільське	16859.2
1992	міське	35296.9
1992	сільське	16759.7
1993	міське	35471
1993	сільське	16773.1
1994	міське	35400.7
1994	сільське	16713.7
1995	міське	35118.8
1995	сільське	16609.6

В заголовках стовпців знаходяться значення,
а не назви змінних

неправильно

місто	2016	2017	2018
Київ	8585	11134	13547
Львів	5858	7521	11855
Одеса	6967	8469	11909
Харків	6365	8311	11546

правильно

місто	рік	середня зарплата
Київ	2016	8585
Київ	2017	11134
Київ	2018	13547
Львів	2016	5858
Львів	2017	7521
Львів	2018	11855
Одеса	2016	6967
Одеса	2017	8469
Одеса	2018	11909
Харків	2016	6365
Харків	2017	8311
Харків	2018	11546

Кілька змінних зберігаються в одному стовпці

неправильно

місто	2017_аудит	2017_дизайн	2018_аудит	2018_дизайн
Київ	9169	11141	11892	14449
Львів	8087	9889	10383	12697
Харків	8253	9208	10033	11194
Одеса	7239	10319	9453	13474

правильно

місто	рік	професія	середня зарплата
Київ	2017	аудит	9169
Львів	2017	аудит	8087
Харків	2017	аудит	8253
Одеса	2017	аудит	7239
Київ	2017	дизайн	11141
Львів	2017	дизайн	9889
Харків	2017	дизайн	9208
Одеса	2017	дизайн	10319
Київ	2018	аудит	11892
Львів	2018	аудит	10383
Харків	2018	аудит	10033
Одеса	2018	аудит	9453
Київ	2018	дизайн	14449

Змінні у таблиці не є уніфікованими, "брудні дані"

неправильно

місто	підприємство	дата	прибуток
Київ	Комунсервіс	2 червня 2017	14439934
Київ	ДП "Комунсервіс"	2016-04-01	1 млн122 тис 14
м. Київ	"Комун-сервіс"	23.01.2018	2 000 000
Львів	Львівкартон	1 лист 2016 р.	1,567,890
Львів	Львів картон	23.08.2019	768431
Харків	Льодова арена	2015-03-11	+2300943
Харьков	Льодова арена	14/2/2012	5487643

правильно

місто	підприємство	дата	прибуток
Київ	ДП "Комунсервіс"	2017-06-02	14439934
Київ	ДП "Комунсервіс"	2016-04-01	1122014
Київ	ДП "Комунсервіс"	2018-01-23	2000000
Львів	ДП "Львівкартон"	2016-11-01	1567890
Львів	ДП "Львівкартон"	2019-08-23	768431
Харків	ДП "Льодова арена"	2015-03-11	2300943
Харків	ДП "Льодова арена"	2012-02-14	5487643