

## COM5961 DATA DRIVEN PRODUCTS & SERVICES DESIGN: LESSON 4 - USE OF ANALYTICS IN PM LIFE CYCLE

Bernard Suen  
Center for Entrepreneurship  
Chinese University of Hong Kong

**Today's agenda.**

1. Useful **SQL** commands for analysing data patterns
2. Use of data in **product management life cycle**
3. The **CRISP-DM** data cycle
4. Most common **sources** of data
5. Web scraping in the context of **CRISP-DM** cycle
6. Scraping **techniques** and tools
7. Data **cleaning, transformation** and **analysis**

# **Download Data Tools for Tonight**



Features Download Pricing Help Blog Log in Sign up

## A web scraping tool that is easy to use

ParseHub is a free web scraping tool. With our advanced web scraper, extracting data is as easy as clicking the data you need.

[Download our free app](#)



<https://www.parsehub.com/>



Open a website

Download our [desktop app](#). Choose a site to scrape data from.



Click to select data

Get data from multiple pages. Interact with AJAX, forms, dropdowns, etc.



Download results

Access data via JSON, Excel and [API](#). Data is collected by our servers.



 Projects

 Runs

 My Account

 Integrations

 Plans & Billing

 Tutorials

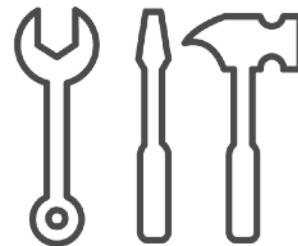
 Documentation

 API

 Contact

 Log Out

## Recent projects



 New Project

● covid19 hkej.com Project  

● travelchinaguide.com Project  

● startupbeat.hkej.com Project  

basketball-reference.com P...  

 See more

## Recent runs

## Interactive Tutorials

 Learn the Basics (8 min)

 Select and Download Data (3 min)

 Group data with Relative Select (3 min)

 Click and Navigate to Links (4 min)

## Written Tutorials

 Parsehub 101

 Pagination ('next' page buttons)

 Scrape product details

 Scrape leads from directories

# **Case Study: Use of the Northwind dataset**



## graphql-compose / graphql-compose-examples

Public

Sponsor

Notifications

Fork 283

Star 172

Code

Issues 2

Pull requests 9

Actions

Projects

Security

Insights

master

graphql-compose-examples / examples / northwind / data / csv /

Go to file



nodkz Preparation for big example: added sample data from Northwind.

6ab72eb on 20 Aug 2016

History

..

README.md

Preparation for big example: added sample data from Northwind.

6 years ago

categories.csv

Preparation for big example: added sample data from Northwind.

6 years ago

customers.csv

Preparation for big example: added sample data from Northwind.

6 years ago

employee\_territories.csv

Preparation for big example: added sample data from Northwind.

6 years ago

employees.csv

Preparation for big example: added sample data from Northwind.

6 years ago

order\_details.csv

Preparation for big example: added sample data from Northwind.

6 years ago

orders.csv

Preparation for big example: added sample data from Northwind.

6 years ago

products.csv

Preparation for big example: added sample data from Northwind.

6 years ago

regions.csv

Preparation for big example: added sample data from Northwind.

6 years ago

shippers.csv

Preparation for big example: added sample data from Northwind.

6 years ago

<https://github.com/graphql-compose/graphql-compose-examples/tree/master/examples/northwind/data/csv>

# Retrieve Records by the **SELECT** Command

**SELECT** (field name or wild card ‘\*’)  
**FROM** table name

**SELECT DISTINCT** (field name)  
**FROM** table name

**SELECT** (field name or wild card ‘\*’)  
**FROM** table name **WHERE** condition **AND** condition

**SELECT** (field name or wild card ‘\*’)  
**FROM** table name **WHERE** condition **OR** condition

Database Structure | Browse Data | Edit Pragmas | Execute SQL

X SQL 1 X SQL 2 X SQL 3

1 **SELECT \* FROM customers**

	customerID	companyName	contactName	contactTitle	address
86	WANDK	Die Wandernde Kuh	Rita Müller	Sales Representative	Adenauerallee 900
87	WARTH	Wartian Herkku	Pirkko Koskitalo	Accounting Manager	Torikatu 38
88	WELLI	Wellington Importadora	Paula Parente	Sales Manager	Rua do Mercado 12
89	WHITC	White Clover Markets	Karl Jablonski	Owner	305 – 14th Ave. S. Suite 3B
90	WILMK	Wilman Kala	Matti Karttunen	Owner/Marketing Assistant	Keskuskatu 45
91	WOLZA	Wolski Zajazz	Zbyszek Piestrzeniewicz	Owner	ul. Filtrowa 68

**Result: 91 rows returned in 64ms**

At line 1:  
**SELECT \* FROM customers**

Database Structure | Browse Data | Edit Pragmas | Execute SQL

X SQL 1

1 SELECT customerID, companyName, address FROM customers

	customerID	companyName	address
1	ALFKI	Alfreds Futterkiste	Obere Str. 57
2	ANATR	Ana Trujillo Emparedados y helados	Avda. de la Constitución 2222
3	ANTON	Antonio Moreno Taquería	Mataderos 2312
4	AROUT	Around the Horn	120 Hanover Sq.
5	BERGS	Berglunds snabbköp	Berguvsvägen 8
6	BLAUS	Blauer See Delikatessen	Forsterstr. 57
7	BLONP	Blondesddsl père et fils	24 place Kléber

Result: 91 rows returned in 29ms  
At line 1:  
SELECT customerID, companyName, address FROM customers

Database Structure | Browse Data | Edit Pragmas | Execute SQL

X SQL 1 | X SQL 2 | X SQL 3

1 SELECT DISTINCT city FROM customers

	city
64	Reims
65	Stuttgart
66	Oulu
67	Resende
68	Seattle
69	Helsinki
70	Warszawa

Result: 70 rows returned in 9ms

At line 1:  
SELECT DISTINCT city FROM customers

Database Structure | Browse Data | Edit Pragmas | Execute SQL

SQL 1 | SQL 2 | SQL 3 | SQL 4

1 `SELECT customerID, companyName, address, city FROM customers WHERE city = 'London'`

	customerID	companyName	address	city
1	AROUT	Around the Horn	120 Hanover Sq.	London
2	BSBEV	Bs Beverages	Fauntleroy Circus	London
3	CONSH	Consolidated Holdings	Berkeley Gardens 12 Brewery	London
4	EASTC	Eastern Connection	35 King George	London
5	NORTS	North/South	South House 300 Queensbridge	London
6	SEVES	Seven Seas Imports	90 Wadhurst Rd.	London

Result: 6 rows returned in 11ms  
At line 1:  
`SELECT customerID, companyName, address, city FROM customers WHERE city = 'London'`

Database Structure | Browse Data | Edit Pragmas | Execute SQL

X SQL 1 | X SQL 2 | X SQL 3 | X SQL 4 | X SQL 5 | X SQL 6 | X SQL 7

1 `SELECT customerID, companyName, address, city FROM customers WHERE city = "London" OR city = "Vancouver"`

	customerID	companyName	address	city
1	AROUT	Around the Horn	120 Hanover Sq.	London
2	BSBEV	Bs Beverages	Fauntleroy Circus	London
3	CONSH	Consolidated Holdings	Berkeley Gardens 12 Brewery	London
4	EASTC	Eastern Connection	35 King George	London
5	LAUGB	Laughing Bacchus Wine Cellars	1900 Oak St.	Vancouver
6	NORTS	North/South	South House 300 Queensbridge	London
7	SEVES	Seven Seas Imports	90 Wadhurst Rd.	London

Result: 7 rows returned in 25ms  
At line 1:  
`SELECT customerID, companyName, address, city FROM customers WHERE city = "London" OR city = "Vancouver"`

**Use of wildcards using the “**LIKE**“ clause.**

Database Structure | Browse Data | Edit Pragmas | Execute SQL

X SQL 1 | X SQL 2 | X SQL 3 | X SQL 4 | X SQL 5 | X SQL 6

```
1 SELECT customerID, companyName, address, city FROM customers WHERE city LIKE "B%"
```

	customerID	companyName	address	city
1	ALFKI	Alfreds Futterkiste	Obere Str. 57	Berlin
2	CACTU	Cactus Comidas para llevar	Cerrito 333	Buenos Aires
3	CHOPS	Chop suey Chinese	Hauptstr. 29	Bern
4	FOLKO	Folk och fā HB	Åkergratan 24	Bräcke
5	GALED	Galería del gastrónomo	Rambla de Cataluña 23	Barcelona
6	KOENE	Königlich Essen	Maubelstr. 90	Brandenburg
7	LILAS	LILA-Supermercado	Carrera 52 con Ave. Bolívar #65-98 Llano Largo	Barquisimeto

Result: 13 rows returned in 28ms  
At line 1:  
SELECT customerID, companyName, address, city FROM customers WHERE city LIKE "B%"

HTML CSS JAVASCRIPT SQL PYTHON PHP BOOTSTRAP HOW TO W3.CSS JAVA JQUERY C++ C# R React Kotlin

 w3schools  
CERTIFIED . 2022

Get started

CODE GAME 

Play Game

ADVERTISEMENT 

**SQL Tutorial**

SQL HOME  
SQL Intro  
SQL Syntax  
SQL Select  
SQL Select Distinct  
SQL Where  
SQL And, Or, Not  
SQL Order By  
SQL Insert Into  
SQL Null Values  
SQL Update  
SQL Delete  
SQL Select Top  
SQL Min and Max  
SQL Count, Avg, Sum  
SQL Like  
**SQL Wildcards**  
SQL In  
SQL Between  
SQL Aliases  
SQL Joins  
SQL Inner Join  
SQL Left Join  
SQL Right Join  
SQL Full Join  
SQL Self Join  
SQL Union  
SQL Group By  
SQL Having  
SQL Exists  
SQL Any/All

[]	Represents any single character within the brackets	h[oa]t finds hot and hat, but not hit
^	Represents any character not in the brackets	h[^oa]t finds hit, but not hot and hat
-	Represents any single character within the specified range	c[a-b]t finds cat and cbt

All the wildcards can also be used in combinations!

Here are some examples showing different **LIKE** operators with '%' and '\_' wildcards:

**"Like' is used as wildcard for filling in the blank.**

LIKE Operator	Description
WHERE CustomerName LIKE 'a%'	Finds any values that starts with "a"
WHERE CustomerName LIKE '%a'	Finds any values that ends with "a"
WHERE CustomerName LIKE '%or%'	Finds any values that have "or" in any position
WHERE CustomerName LIKE '_r%'	Finds any values that have "r" in the second position
WHERE CustomerName LIKE 'a__%'	Finds any values that starts with "a" and are at least 3 characters in length
WHERE ContactName LIKE 'a%o'	Finds any values that starts with "a" and ends with "o"

## Demo Database

The table below shows the complete "Customers" table from the Northwind sample database:

CustomerID	CustomerName	ContactName	Address	City	PostalCode	Country
1	Alfreds Futterkiste	Maria Anders	Obere Str. 57	Berlin	12209	Germany
2	Ana Trujillo Emparedados y helados	Ana Trujillo	Avda. de la Constitución 2222	México D.F.	05021	Mexico
3	Antonio Moreno Taquería	Antonio Moreno	Mataderos 2312	México D.F.	05023	Mexico

[https://www.w3schools.com/sql/sql\\_like.asp](https://www.w3schools.com/sql/sql_like.asp)

Database Structure | Browse Data | Edit Pragmas | Execute SQL

X SQL 1 | X SQL 2 | X SQL 3 | X SQL 4 | X SQL 5 | X SQL 6 | X SQL 7 | X SQL 10

```
1 SELECT customerID, companyName, address, city FROM customers WHERE city LIKE "%er%"
```

	customerID	companyName	address	city
1	ALFKI	Alfreds Futterkiste	Obere Str. 57	Berlin
2	CHOPS	Chop-suey Chinese	Hauptstr. 29	Bern
3	LACOR	La corne d'abondance	67 avenue de l'Europe	Versailles
4	LAUGB	Laughing Bacchus Wine Cellars	1900 Oak St.	Vancouver
5	MAGAA	Magazzini Alimentari Riuniti	Via Ludovico il Moro 22	Bergamo
6	RATTC	Rattlesnake Canyon Grocery	2817 Milton Dr.	Albuquerque
7	SANTG	Santé Gourmet	Erling Skakkes gate 78	Stavern

Result: 10 rows returned in 26ms  
At line 1:  
SELECT customerID, companyName, address, city FROM customers WHERE city LIKE "%er%"

**Use of the “IN” clause.**

Database Structure | Browse Data | Edit Pragmas **Execute SQL**

SQL 1 | SQL 2 | SQL 3 | SQL 4 | SQL 5 | SQL 6 | SQL 7 | SQL 8 | **SQL 9**

```
1 SELECT customerID, companyName, address, city FROM customers WHERE city IN ('London', 'Paris', 'Madrid');
```

	customerID	companyName	address	city
5	EASTC	Eastern Connection	35 King George	London
6	FISSA	FISSA Fabrica Inter. Salchichas S.A.	C/ Moralzarzal 86	Madrid
7	NORTS	North/South	South House 300 Queensbridge	London
8	PARIS	Paris spécialités	265 boulevard Charonne	Paris
9	ROMEY	Romero y tomillo	Gran Vía 1	Madrid
10	SEVES	Seven Seas Imports	90 Wadhurst Rd.	London
11	SPECD	Spécialités du monde	25 rue Lauriston	Paris

Result: 11 rows returned in 26ms  
At line 1:  
SELECT customerID, companyName, address, city FROM customers WHERE city IN ('London', 'Paris', 'Madrid');

**Use of the “>”, “>=”, “<=”  
clause.**

Database Structure | Browse Data | Edit Pragmas | Execute SQL

SQL 1 | SQL 2 | SQL 3 | SQL 4 | SQL 5 | SQL 6 | SQL 7 | SQL 8 | SQL 9 | SQL 10

1 SELECT \* FROM product WHERE unitPrice > 50

	productID	productName	supplierID	categoryID	quantityPerUnit	unitPrice	unitsInStock	unitsOnOrder	reorderLevel	discontinued
1	9	Mishi Kobe Niku	4	6	18 500 g pkgs.	97.0	29	0	0	1
2	18	Carnarvon Tigers	7	8	16 kg pkg.	62.5	42	0	0	0
3	20	Sir Rodney's Marmalade	8	3	30 gift boxes	81.0	40	0	0	0
4	29	Thüringer Rostbratwurst	12	6	50 bags x30 sausgs	123.79	0	0	0	1
5	38	Côte de Blaye	18	1	12 75 cl bottles	263.5	17	0	15	0
6	51	Manjimup Dried Apples	24	7	50 -300 g pkgs.	53.0	20	0	10	0
7	59	Raclette Courdavault	28	4	5 kg pkg.	55.0	79	0	0	0

Result: 7 rows returned in 34ms

At line 1:

SELECT \* FROM products WHERE unitPrice > 50

# **Use of the “BETWEEN” clause.**

Database Structure | Browse Data | Edit Pragmas **Execute SQL**

X SQL 1 | X SQL 2 | X SQL 3 | X SQL 4 | X SQL 5 | X SQL 6 | X SQL 7 | X SQL 8 | X SQL 9 | X SQL 10 | X SQL 11

1    `SELECT * FROM products WHERE unitPrice BETWEEN 19 AND 20`

	productID	productName	supplierID	categoryID	quantityPerUnit	unitPrice	unitsInStock	unitsOnOrder	reorderLevel	discontinued
1	2	Chang	1	1	24 - 12 oz bottles	19	17	40	25	0
2	36	Inlagd Sill	17	8	24 - 250 g jars	19	112	0	20	0
3	44	Gula Malacca	20	2	20 - 2 kg bags	19.45	27	0	15	0
4	49	Maxilaku	23	3	24 - 50 g pkgs.	20	10	60	15	0
5	57	Ravioli Angelo	26	5	24 - 250 g pkgs.	19.5	36	0	20	0

Result: 5 rows returned in 22ms  
At line 1:  
`SELECT * FROM products WHERE unitPrice BETWEEN 19 AND 20`

**Select of SELECT**

Database Structure | Browse Data | Edit Pragmas | Execute SQL

SQL 1 | SQL 2 | SQL 3 | SQL 4 | SQL 5 | SQL 6 | SQL 7 | SQL 8 | SQL 9 | SQL 10 | SQL 11 | SQL 12

```
1 SELECT *
2 FROM (SELECT customerID, companyName, city, address FROM customers ORDER BY companyName)
3 WHERE city = "London"
```

	customerID	companyName	city	address
1	AROUT	Around the Horn	London	120 Hanover Sq.
2	BSBEV	Bs Beverages	London	Fauntleroy Circus
3	CONSH	Consolidated Holdings	London	Berkeley Gardens 12 Brewery
4	EASTC	Eastern Connection	London	35 King George
5	NORTS	North/South	London	South House 300 Queensbridge
6	SEVES	Seven Seas Imports	London	90 Wadhurst Rd.

Result: 6 rows returned in 26ms  
At line 1:  
SELECT \*  
FROM (SELECT customerID, companyName, city, address FROM customers ORDER BY companyName)  
WHERE city = "London"

# **Aggregate Functions**

## **(AVG, MIN, MAX, COUNT, SUM)**

# **Retrieve Records by Aggregate/Statistic Functions**

**SELECT COUNT (field name)**  
**FROM table name GROUP BY field name WHERE condition**

**SELECT SUM (field name)**  
**FROM table name GROUP BY field name WHERE condition**

**SELECT AVG (field name)**  
**FROM table name GROUP BY field name WHERE condition**

**SELECT MAX I MIN (field name)**  
**FROM table name GROUP BY field name WHERE condition**

Database Structure | Browse Data | Edit Pragmas | **Execute SQL**

SQL 1 | SQL 2 | SQL 3 | SQL 4 | SQL 5 | SQL 6 | SQL 7 | SQL 8 | SQL 9 | SQL 10 | SQL 11 | SQL 12 | **SQL 13** | SQL 14 | SQL 15

```
1 SELECT categoryId, COUNT(*) FROM products GROUP BY categoryId
```

	categoryId	COUNT(*)
1	1	12
2	2	12
3	3	13
4	4	10
5	8	12

Result: 5 rows returned in 20ms  
At line 1:  
-- SELECT categoryId, COUNT(\*) FROM products GROUP BY categoryId HAVING COUNT(\*) >= 10  
SELECT categoryId, COUNT(\*) FROM products GROUP BY categoryId HAVING COUNT(\*) >= 10

Database Structure | Browse Data | Edit Pragmas | Execute SQL

SQL 1 | SQL 2 | SQL 3 | SQL 4 | SQL 5 | SQL 6 | SQL 7 | SQL 8 | SQL 9 | SQL 10 | SQL 11 | SQL 12 | SQL 13 | SQL 14 | SQL 15

1 | `SELECT categoryId, COUNT(*) FROM products GROUP BY categoryId HAVING COUNT(*) >= 10`

	categoryId	COUNT(*)
1	1	12
2	2	12
3	3	13
4	4	10
5	8	12

Result: 5 rows returned in 19ms  
At line 1:  
`SELECT categoryId, COUNT(*) FROM products GROUP BY categoryId HAVING COUNT(*) >= 10`

Database Structure | Browse Data | Edit Pragmas | Execute SQL

SQL 1 | SQL 2 | SQL 3 | SQL 4 | SQL 5 | SQL 6 | SQL 7 | SQL 8 | SQL 9 | SQL 10 | SQL 11 | SQL 12 | SQL 13 | SQL 14 | SQL 15

```
1 SELECT categoryId, ROUND(sum(unitsInStock),2) AS [Total Units in Stock] FROM products GROUP BY categoryId
2 --SELECT categoryId, ROUND(avg(unitsInStock),2) AS [AVG Units in Stock] FROM products GROUP BY categoryId
3 --SELECT categoryId, ROUND(max(unitPrice),2) AS [Max Price] FROM products GROUP BY categoryId
4 --SELECT categoryId, ROUND(min(unitPrice),2) AS [Min Price] FROM products GROUP BY categoryId
5
6
```

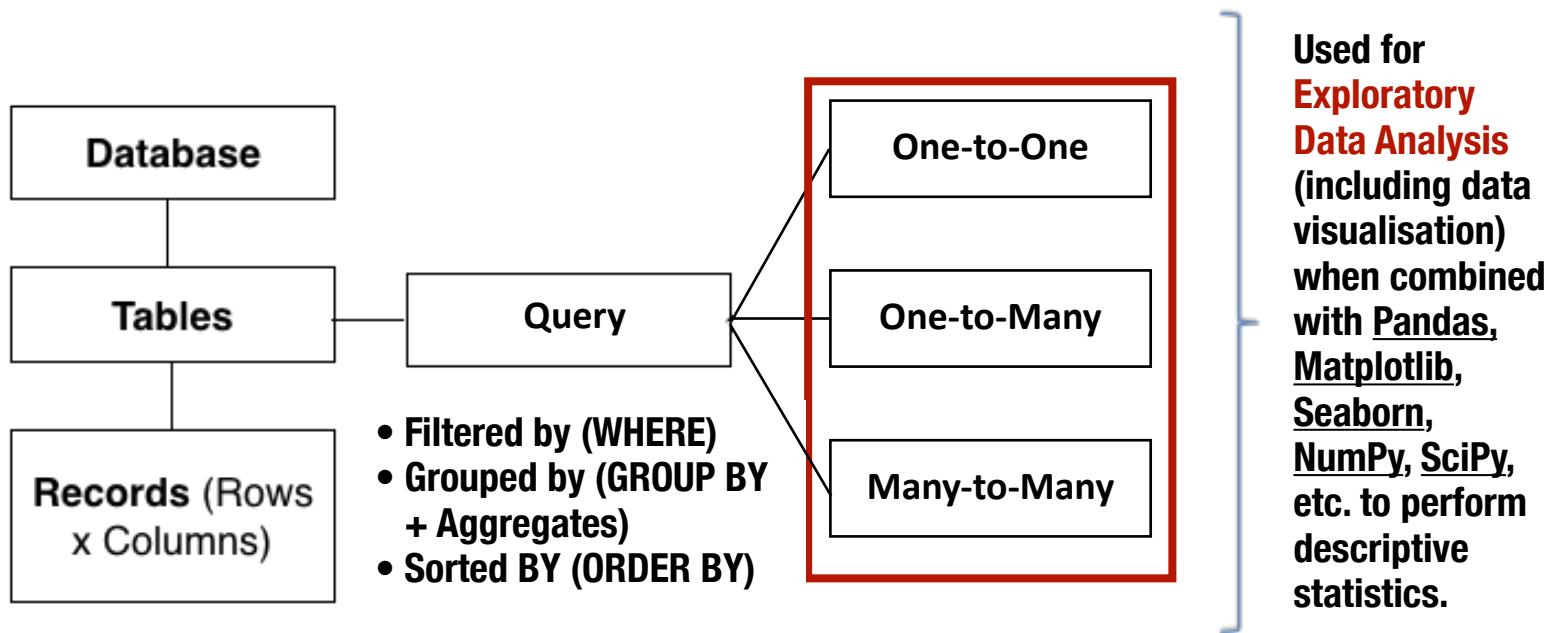
categoryId	Total Units in Stock
1	559.0
2	507.0
3	386.0
4	393.0
5	308.0
6	165.0
7	100.0

Result: 8 rows returned in 22ms

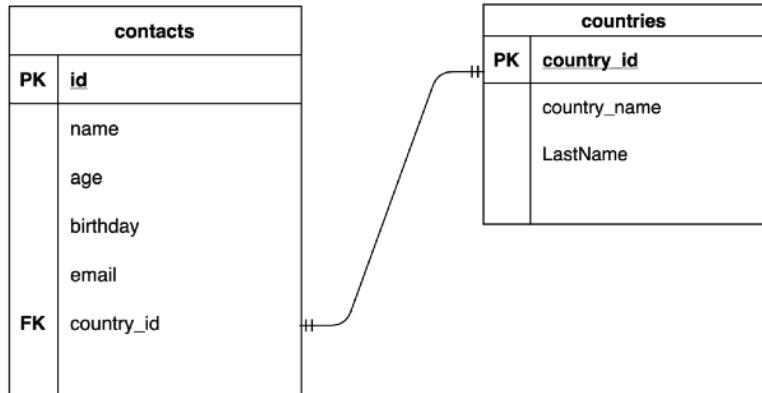
At line 1:

```
SELECT categoryId, ROUND(sum(unitsInStock),2) AS [Total Units in Stock] FROM products GROUP BY categoryId
--SELECT categoryId, ROUND(avg(unitPrice),2) AS [AVG Price] FROM products GROUP BY categoryId
--SELECT categoryId, ROUND(max(unitPrice),2) AS [Max Price] FROM products GROUP BY categoryId
```

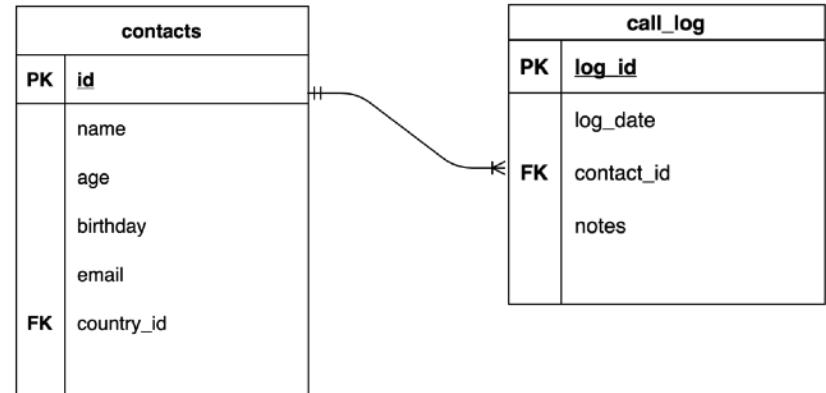
# **Relating Tables with “Join”**



## • One-to-One

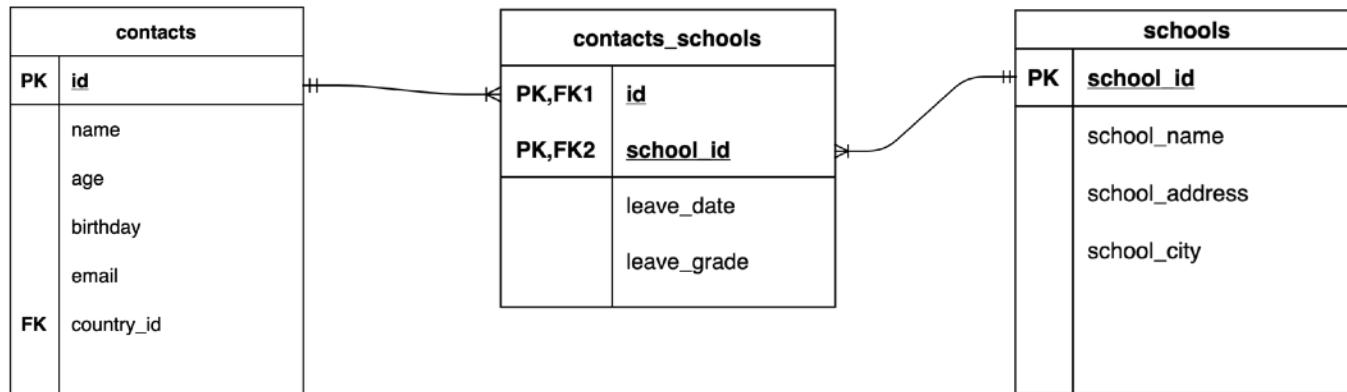


## • One-to-Many



<https://app.diagrams.net/>

## • Many-to-Many



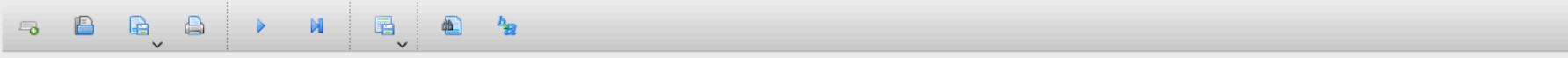
# **Retrieve Records Through SQL Join (e.g. inner, outer, left, right)**

```
SELECT tableName1.fieldName1,tableName2.fieldName2,  
tableName1.fieldName3, tableName2,fieldName4  
FROM tableName1  
INNER JOIN tableName2 ON tableName1.fieldName =  
tableName2.fieldName;
```

## **Example:**

```
SELECT Orders.OrderID, Customers.contactName, Customers.contactTitle,  
Customers.companyName, Orders.OrderDate  
FROM Orders  
INNER JOIN Customers ON Orders.CustomerID=Customers.CustomerID;
```

Database Structure | Browse Data | Edit Pragmas | Execute SQL



x SQL 1n2 | x SQL 3 | x SQL 4 | x SQL 5 | x SQL 6 | x SQL 7 | x SQL 8 | x SQL 9 | x SQL 10 | x SQL 11 | x SQL 12 | x SQL 13 | x SQL 14 | x SQL 15 | x SQL 16

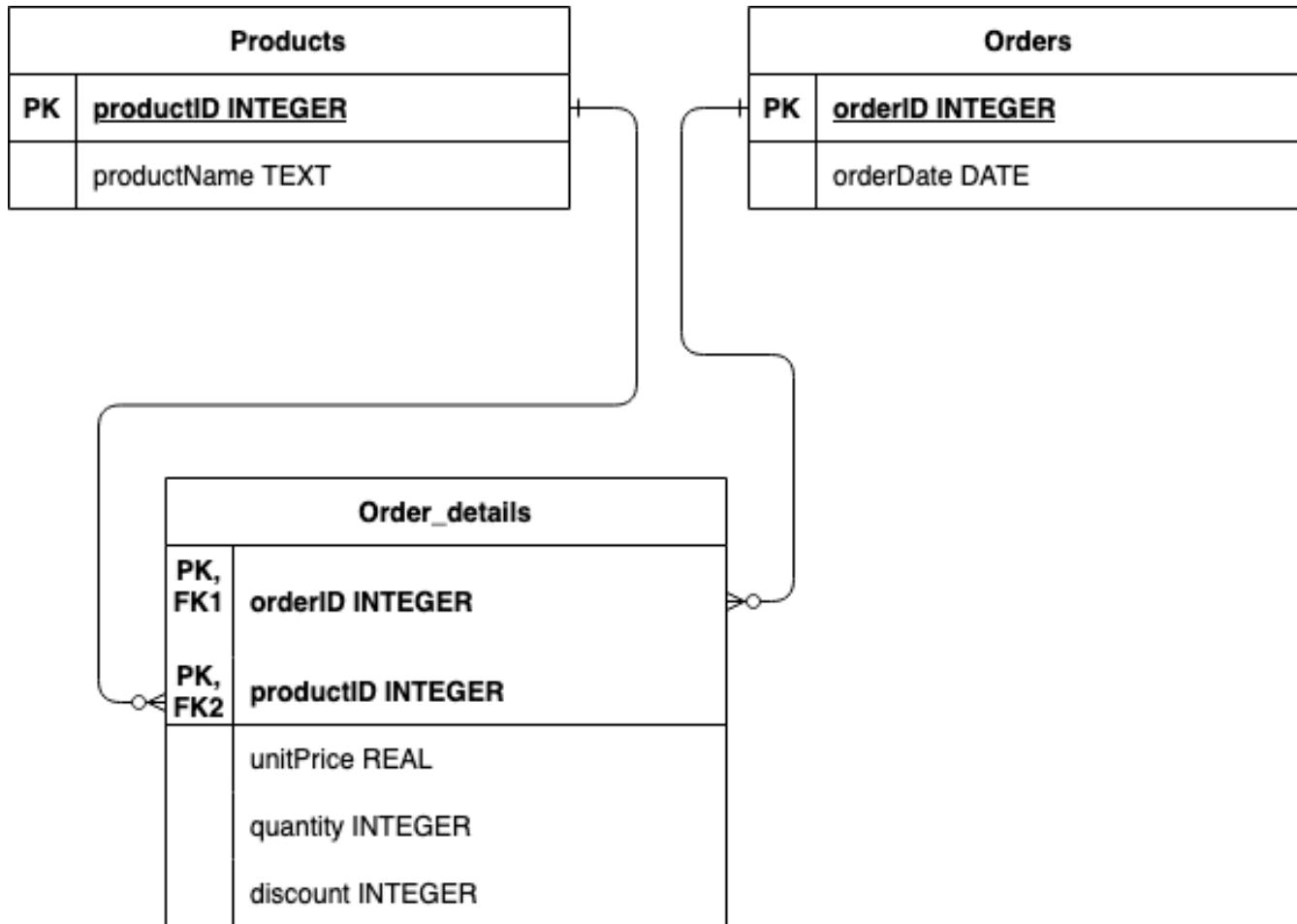
```
1  --SELECT * FROM Orders
2  --SELECT * FROM order_details
3  --SELECT * FROM products
4
5  --SELECT Orders.OrderID, Customers.contactName, Customers.contactTitle, Customers.companyName, Orders.OrderDate
6  --FROM Orders
7  --INNER JOIN Customers ON Orders.CustomerID=Customers.CustomerID;
8
9  SELECT Orders.OrderID, order_details.productID, products.productName, order_details.quantity, order_details.unitPrice
10 FROM orders
11 INNER JOIN order_details ON Orders.orderID = order_details.orderID
12 INNER JOIN products ON order_details.productID = products.productID
```

	orderID	productID	productName	quantity	unitPrice
1	10248	11	Queso Cabrales	12	14.0
2	10248	42	Singaporean Hokkien Fried Mee	10	9.8
3	10248	72	Mozzarella di Giovanni	5	34.8
4	10249	14	Tofu	9	18.6

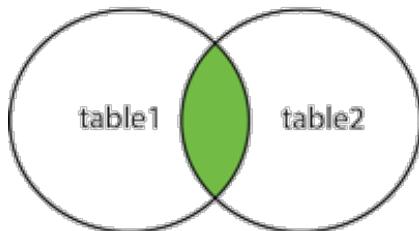
Result: 2155 rows returned in 194ms

At line 1:

```
--SELECT * FROM Orders
--SELECT * FROM order_details
--SELECT * FROM products
```

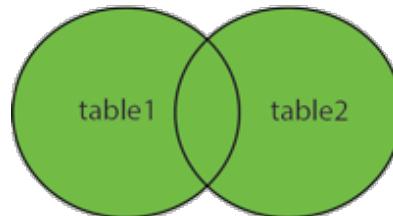


### INNER JOIN



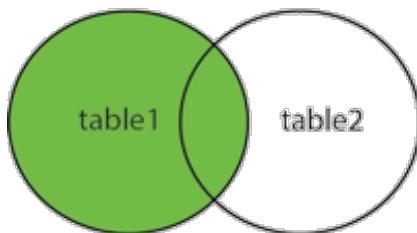
```
SELECT column_name(s)  
FROM table1  
INNER JOIN table2  
ON table1.column_name = table2.column_name;
```

### FULL OUTER JOIN



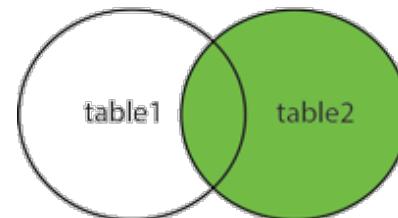
```
SELECT column_name(s)  
FROM table1  
FULL OUTER JOIN table2  
ON table1.column_name = table2.column_name  
WHERE condition;
```

### LEFT JOIN



```
SELECT column_name(s)  
FROM table1  
LEFT JOIN table2  
ON table1.column_name = table2.column_name;
```

### RIGHT JOIN



```
SELECT column_name(s)  
FROM table1  
RIGHT JOIN table2  
ON table1.column_name = table2.column_name;
```

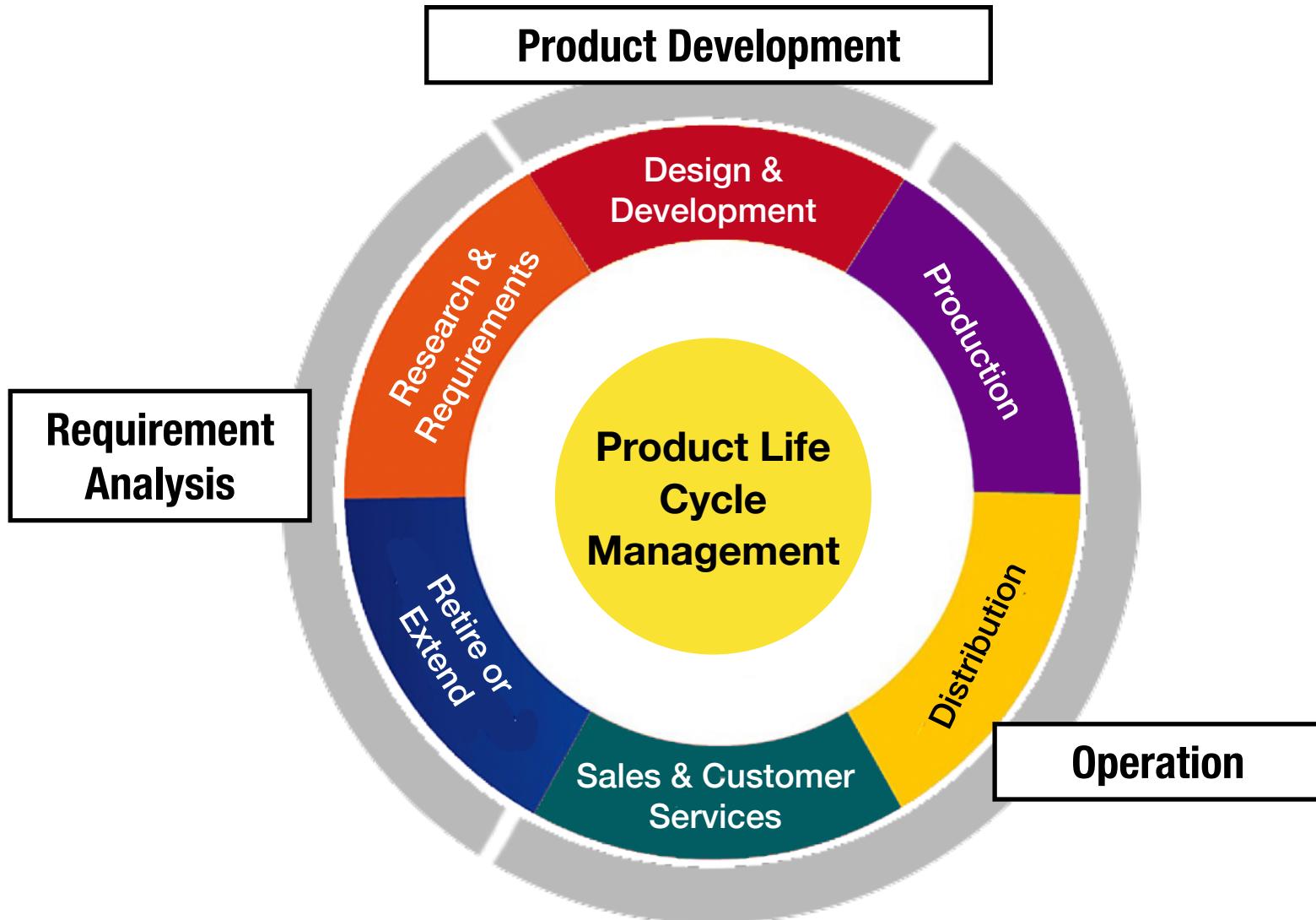
Source: <https://www.w3schools.com/sql/>

# **Executing SQL Through Python Codes**

```
import pandas as pd          # Import Pandas library
import sqlite3                # Import SQLite3 library
con = sqlite3.connect('northwind.db') # Connect to SQLite database
sql = "SELECT customerID, companyName, address, city, postalCode  
FROM customers WHERE city = 'London'" # Assign SQL variable
df = pd.read_sql_query(sql,con)        # Read result into dataframe
con.close()                      # Close connection
df                                # Display data frame
```

**Please refer to Jupyter Notebook example.**

# **Product Life Cycle Management**



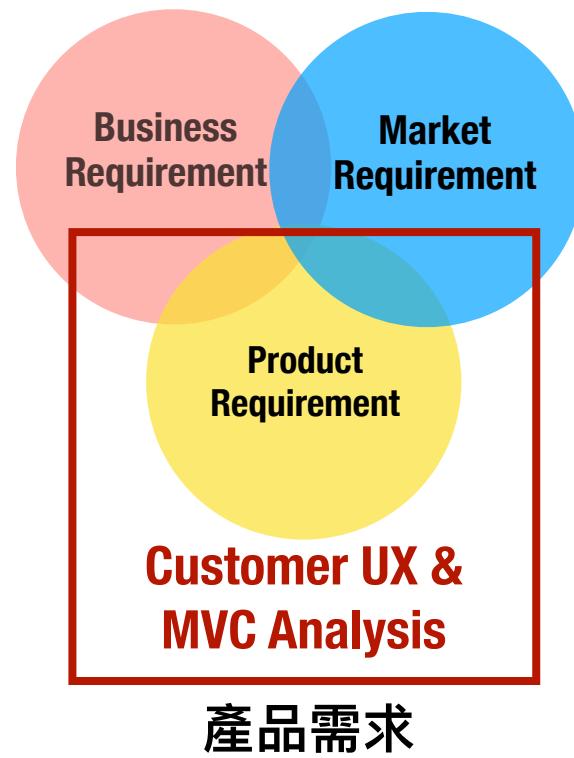
# **Essential Roles in Product Management**

# Essential Roles of a Product Manager



# Requirement Analysis 需求分析

**Company  
Strategy &  
Business Model  
Analysis**  
業務需求



**Competitors  
& Market  
Analysis**  
市場需求

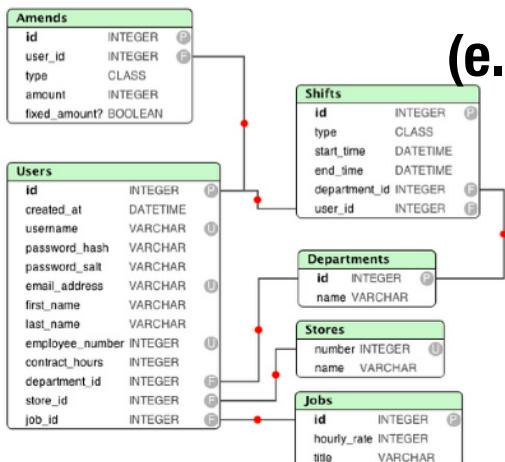
產品需求

# The **Technology Stack** Concept

Source: [commons.wikimedia.org](https://commons.wikimedia.org)



Source: [Caius Durling \(Flickr\)](https://flic.kr)

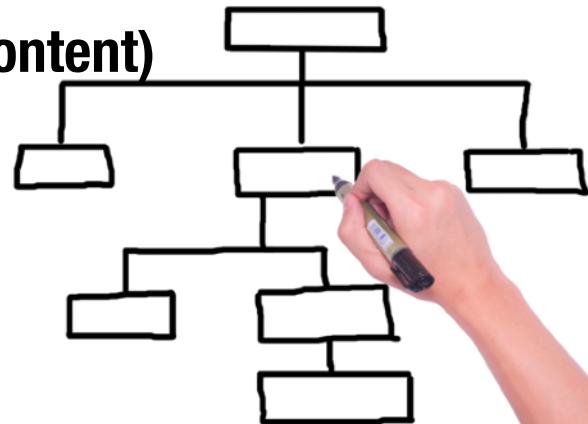


**M(odel)**  
(e.g. database/tables, relations, and queries)

**C(ontroller)**  
(e.g. sitemap, navigation, routes)

Source: [pexels.com](https://pexels.com)

**V(iew)**  
(e.g. user interface, media content)



## Browser/ Mobile

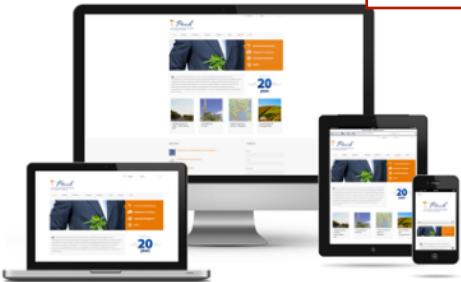


## Web Server

HTML/CSS/JS



## View



## Frontend

## App Server

NodeJS/Python  
Php/C#/Java



## Controller

A terminal window showing Python code for a command-line application. The code includes operations like addition, subtraction, multiplication, division, and power calculations, along with user input handling.

```
<!DOCTYPE html>
<html lang="en">
<head>
<meta charset="utf-8">
<title> A Tiny HTML Document </title>
<link href = "styles.css" rel="stylesheet">
<script src="scripts.js"></script>
</head>

<body>
<p>Let's rock the browser, HTML5 style.</p>
</body>
</html>
```

```

# take input from the user
print("Select operation:")
print("1.Add")
print("2.Subtract")
print("3.Multiply")
print("4.Divide")
print("5.Power")
choice = input("Enter choice(1/2/3/4/5):")

num1 = int(input("Enter first number:"))
num2 = int(input("Enter the second number:"))

if choice == '1':
    print(num1, "+", num2, "=", num1+num2)
    print(value, ..., sep=' ', end='\n', file=sys.stdout, flush=True)
else:
    print("Invalid input")
```

## Backend

## Database Server

SQL/NoSQL  
Data



## Model

A screenshot of a MySQL Workbench interface showing a table named 'popularity'. The table has columns for ID, PHONE, POPULARNAME, PREFERREDNAME, and LATITUDE. The data shows multiple rows of popularity data.

ID	PHONE	POPULARNAME	PREFERREDNAME	LATITUDE
1194020	00994414	preferred_name_00994414	00994414	23.789675
1194021	00994415	preferred_name_00994415	00994415	23.789675
1194022	00994416	preferred_name_00994416	00994416	23.789675
1194023	00994417	preferred_name_00994417	00994417	23.789675
1194024	00994418	preferred_name_00994418	00994418	23.789675
1194025	00994419	preferred_name_00994419	00994419	23.789675
1194026	00994420	preferred_name_00994420	00994420	23.789675
1194027	00994421	preferred_name_00994421	00994421	23.789675
1194028	00994422	preferred_name_00994422	00994422	23.789675
1194029	00994423	preferred_name_00994423	00994423	23.789675
1194030	00994424	preferred_name_00994424	00994424	23.789675
1194031	00994425	preferred_name_00994425	00994425	23.789675
1194032	00994426	preferred_name_00994426	00994426	23.789675

## “CLOUD” AND “STACK” STRATEGY?

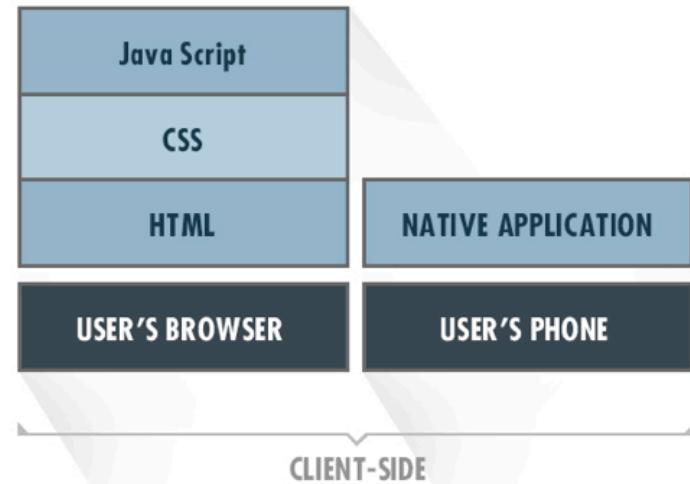
云架构及前端和后端的全栈策略



**BACK-END TECHNOLOGY**  
后台



THE INTERNET



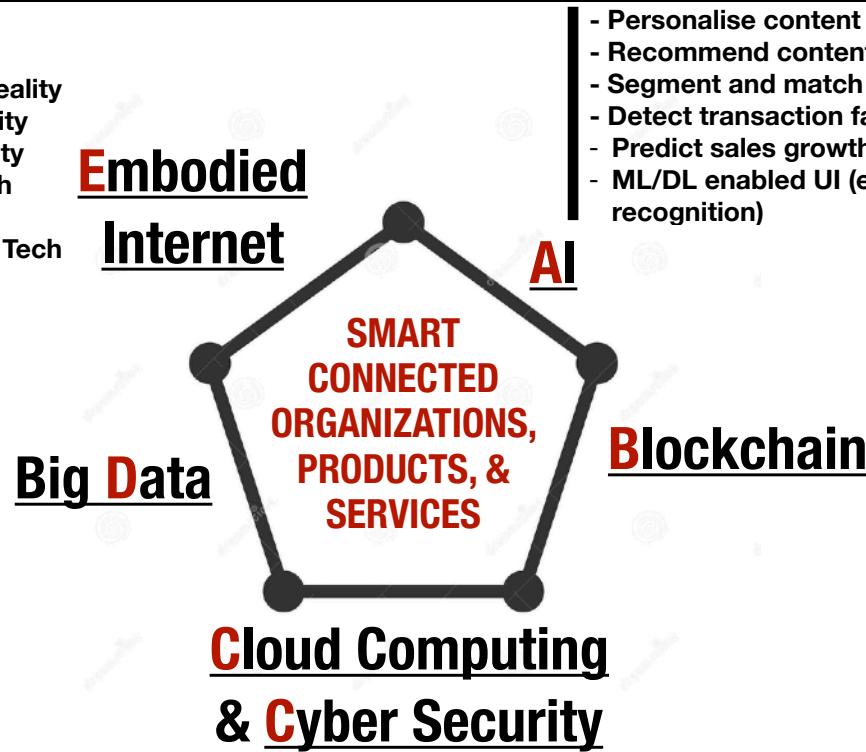
**FRONT-END TECHNOLOGY**  
前台

# Digital Infrastructure

- Assemble structured & unstructured customer data for profiling trends
- Add smart devices & data to network
- Build enterprise data warehouse and data lake
- Enable streaming services

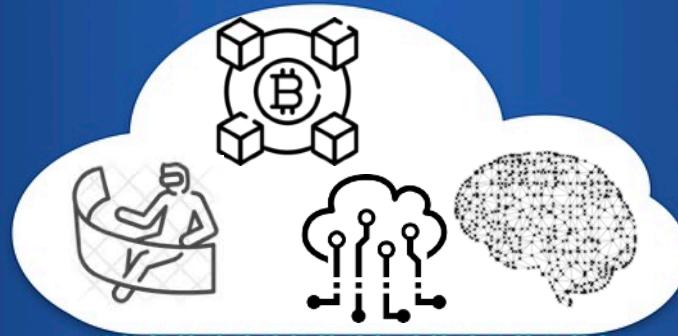
- Augment Reality
- Virtual Reality
- Mixed Reality
- Avatar Tech
- Embient & Experience Tech
- Metaverse

- Personalise content
- Recommend content and products
- Segment and match members
- Detect transaction fault and business risks
- Predict sales growth and optimise inventory
- ML/DL enabled UI (e.g. face, voice, and fingerprint recognition)



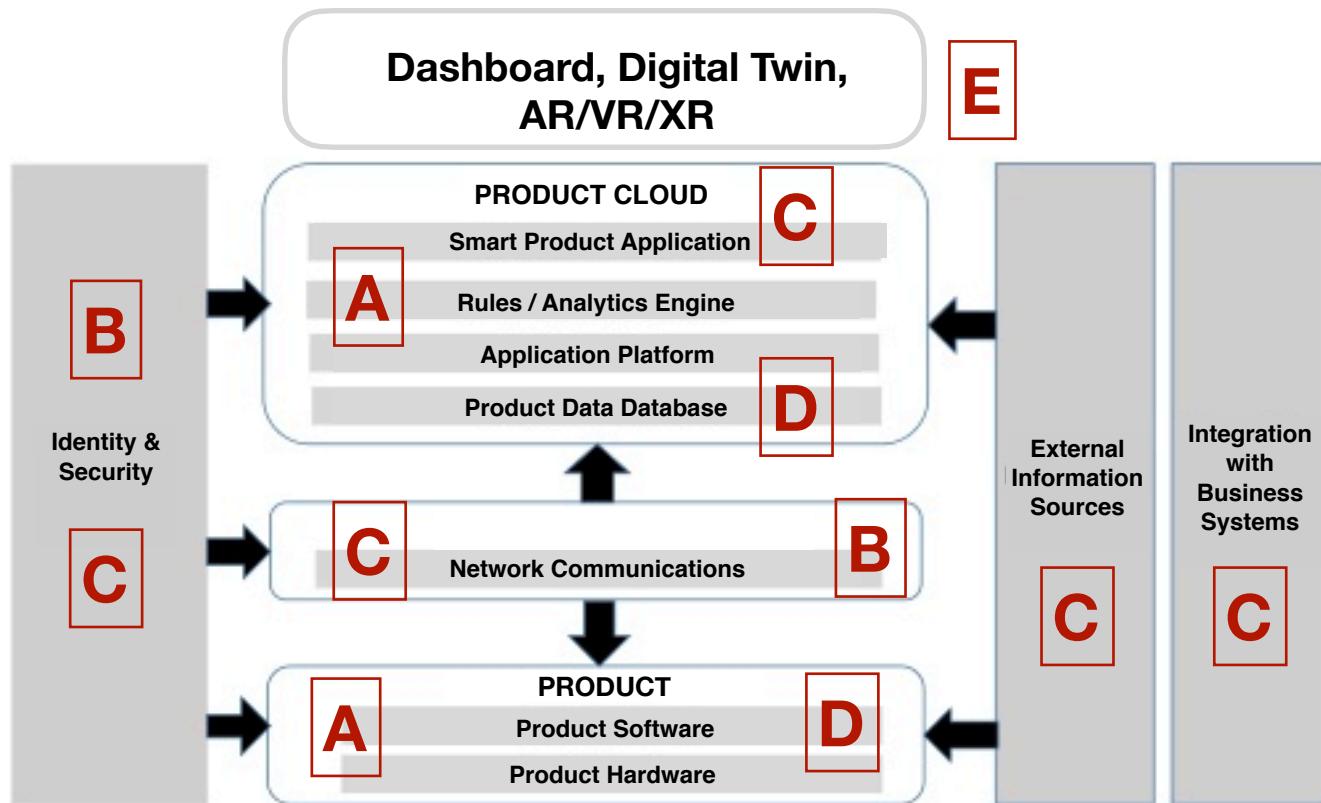
- Outsource data storage and computing
- Outsource hardware maintenance and upgrade
- Outsource IoT, big data and machine learning infrastructure
- Scale up flexibility
- Leverage container/serverless tech
- Encryption, authentication, authorization, and firewall
- Infrastructure and biometric security
- Cyber security audit & disaster recovery

- Token economics
- Disintermediated, decentralized, and distributed architecture
- Smart contracts
- Immutable and traceable
- Crypto-currency
- Web 3.0



```
1 01 01 00 10 10 01 00 10 01 10 00 10 11 00 11 00 0  
1 01 11 11 10 11 00 11 10 10 01 10 10 11 01 10 11 10 0  
0 10 10 10 10 10 00 00 11 00 00 00 00 00 01 01 11 10 0  
0 11 01 11 00 00 01 00 11 11 00 11 11 01 10 00 10 00 1  
0 00 11 00 01 11 01 10 10 01 00 11 11 11 01 01 01 11 0  
1 11 10 01 01 01 00 01 01 00 01 00 11 00 00 01 10 0  
1 10 01 01 01 11 01 00 10 00 00 10 00 01 10 11 00 10 1  
1 00 10 00 01 00 10 11 01 00 00 10 10 01 10 01 10 0  
1 10 11 00 01 11 00 10 01 11 11 11 11 10 11 01 01 1  
0 00 11 00 11 10 10 00 00 11 01 11 10 11 01 10 01 01 0  
1 01 00 01 01 00 01 10 01 00 00 11 10 00 11 00 10 01 0  
1 10 10 00 10 10 10 11 10 01 10 01 11 00 01 01 11 00 1  
0 00 01 10 01 00 10 00 11 00 10 00 10 00 11 00 00 10 1  
1 01 11 01 01 10 11 00 00 11 10 01 10 00 00 01 10 01 1  
0 11 00 11 01 00 01 11 11 10 10 10 11 11 11 00 00 10 0  
1 11 01 10 11 11 01 00 01 00 01 11 00 00 00 11 00 00 10 0  
1 11 01 00 10 11 01 11 11 10 11 00 11 10 01 01 11 10 01 1  
0 01 01 00 10 11 11 00 11 01 00 10 10 00 00 00 11 11 1  
0 10 11 01 00 01 11 10 11 10 01 01 11 10 00 01 00 00 1  
0 01 11 01 10 10 01 11 00 00 11 00 00 01 10 01 00 01 10 1  
1 00 01 00 11 11 10 10 10 11 11 11 01 11 10 11 10 11 1  
1 10 11 11 10 00 10 01 01 00 11 11 00 11 01 10 10 00 0  
1 11 10 11 11 10 00 10 10 00 10 01 11 01 11 10 00 11 0  
1 10 01 00 10 01 11 00 10 00 10 01 01 01 01 11 10 10 00 1  
0 10 00 11 10 10 01 11 00 00 10 10 11 00 10 01 11 11 01 1  
0 00 01 11 00 01 11 00 10 00 10 00 00 01 01 11 01 01 11 00 0  
1 11 11 01 00 10 00 11 10 10 00 11 00 00 10 11 10 10 0  
1 01 11 11 11 11 01 00 11 01 11 00 11 01 10 00 00 01 1  
0 11 00 00 11 01 01 11 01 11 01 00 11 10 11 01 10 10 1  
1 10 10 11 00 00 00 10 00 10 10 00 11 00 11 10 11 11 10 1  
1 10 01 10 11 00 01 01 01 10 10 10 11 11 11 10 01 01 0  
1 11 10 01 01 00 11 01 01 11 00 10 11 00 11 01 10 00 1  
0 01 10 00 01 10 11 10 01 00 00 00 11 00 11 01 00 01 0  
1 11 10 11 10 00 00 11 00 00 11 11 00 11 01 11 01 11 0  
0 00 00 01 10 01 01 10 11 11 10 10 01 01 00 11 11 00 0  
0 11 00 00 00 11 10 01 00 00 00 10 11 11 11 01 01 0  
0 11 11 10 11 11 10 00 10 00 01 11 10 11 00 01 00 11 0  
0 11 01 00 01 00 11 10 10 10 10 00 10 11 11 00 00 01 11 1
```

# Technology Stack



Source: Connected Product Transform Business by Michael E. Porter & James E. Heppelmann. Harvard Business Review

# Why Software Is Eating the World

by Marc Andreessen

how innovation happens +  
nature of the firm & evolving institutions



Software is eating the world.

More than 10 years after the peak of the 1990s dot-com bubble, a dozen or so new Internet companies like Facebook and Twitter are sparking controversy in Silicon Valley, due to their rapidly growing private market valuations, and even the occasional successful IPO. With scars from the heyday of Webvan and Pets.com still fresh in the investor psyche, people are asking, "Isn't this just a dangerous new bubble?"

I, along with others, have been arguing the other side of the case. (I am co-founder and general partner of venture capital firm Andreessen-Horowitz, which has invested in Facebook, Groupon, Skype, Twitter, Zynga, and Foursquare, among others. I am also personally an investor in LinkedIn.) We believe that many of the prominent new Internet companies are building real, high-growth, high-margin, highly defensible businesses.

Today's stock market actually hates technology, as shown by all-time low price/earnings ratios for major public technology companies. Apple, for example, has a P/E ratio of around 15.2 — about the same as the broader stock market, despite Apple's immense profitability and dominant market position (Apple in the last couple weeks became the biggest company in America, judged by market capitalization, surpassing Exxon Mobil). And, perhaps most telling, you can't have a bubble when people are constantly screaming "Bubble!"

But too much of the debate is still around financial valuation, as opposed to the underlying intrinsic value of the best of Silicon Valley's new companies. My own theory is that we are in the middle of a dramatic and broad technological and economic shift in which software companies are poised to take over large swathes of the economy.

More and more major businesses and industries are being run on software and delivered as online services — from movies to agriculture to national defense. Many of the winners are Silicon Valley-style entrepreneurial technology companies that are invading and overturning established industry structures. Over the next 10 years, I expect many more industries to be disrupted by software, with new world-beating Silicon Valley companies doing the disruption in more cases than not.

Why is this happening now?

<https://a16z.com/2011/08/20/why-software-is-eating-the-world/>

## No, Data Is Eating the World

Netflix and Google are still disruptive companies -- but no longer primarily because of their software.

By Dries Buytaert | Jan 7, 2015, 12:27pm EST

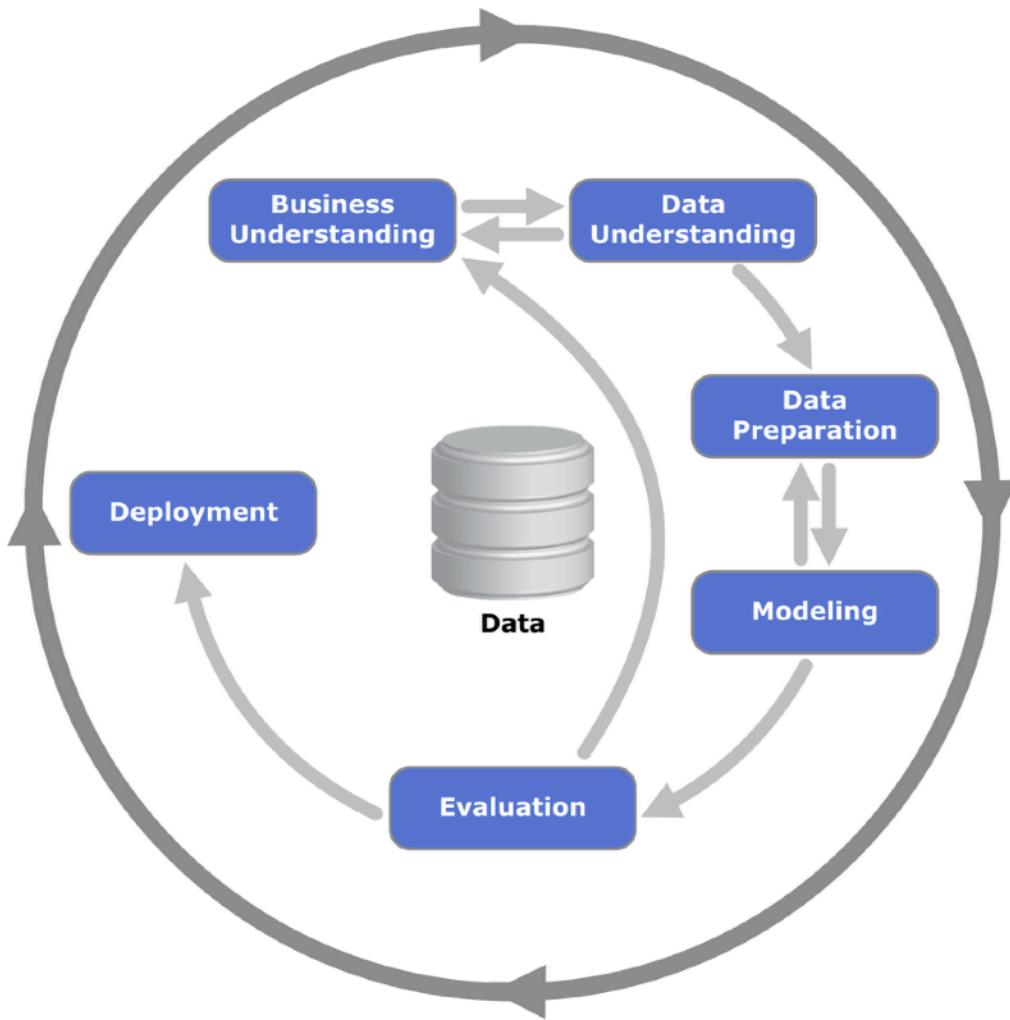
f t  SHARE



Creativemarc/Shutterstock

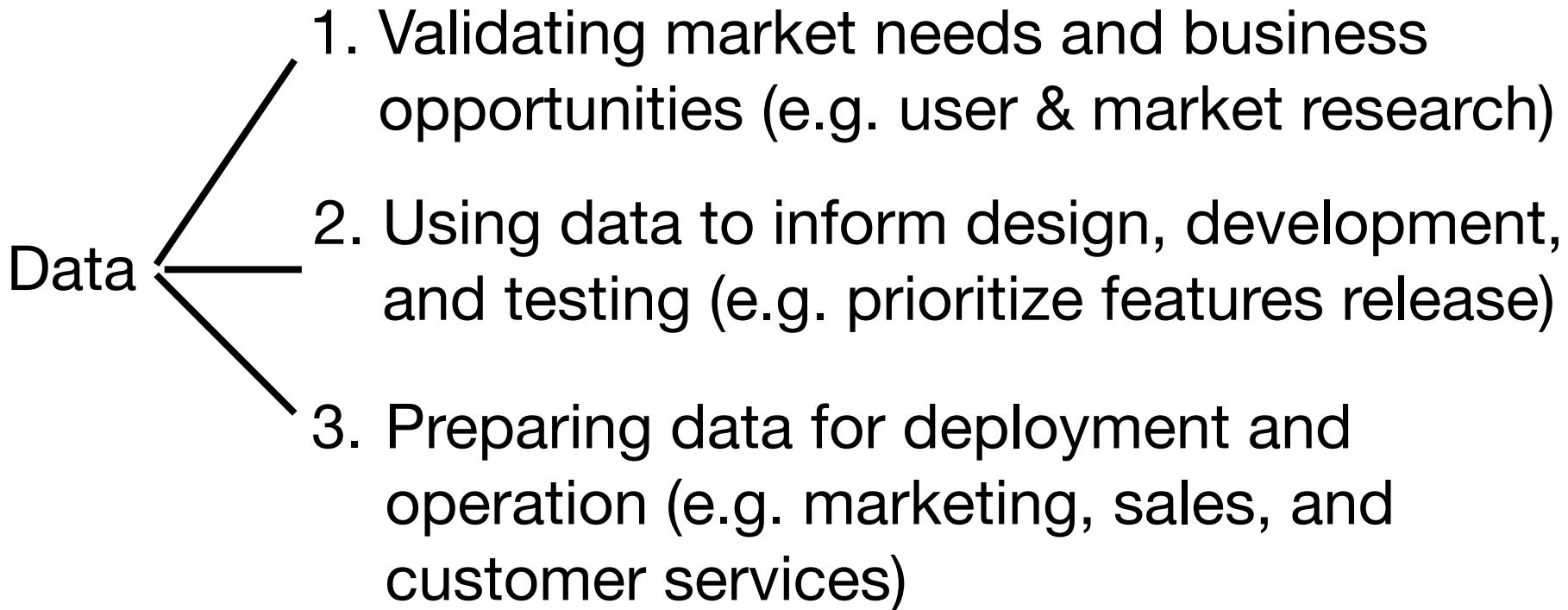
<https://www.vox.com/2015/1/7/11557562/no-data-is-eating-the-world>

# The CRISP-DM Model



# CRoss Industry Standard Process for Data Mining

## **Three Major Functions of Data Use**

- 
- Data
1. Validating market needs and business opportunities (e.g. user & market research)
  2. Using data to inform design, development, and testing (e.g. prioritize features release)
  3. Preparing data for deployment and operation (e.g. marketing, sales, and customer services)

BTA#1  Sushi in, codes out? Build my own page – HKDesign	120
Assignment1: Say Goodbye to Monday Blue with Coding! – HKDesign	110
A Battle Against Myself – HKDesign	96
New Post – HKDesign	87
Having fun with HTML and CSS – HKDesign	85
My first page using HTML and CSS – HKDesign	79
Making webpages with html&css – HKDesign	79
All Journals – Page 3 – HKDesign	76
First attempt at HTML and CSS – HKDesign	65
My first HTML experience – HKDesign	63
My First Webpage Created with HTML/CSS: Struggling While Having Fun – HKDesign	62
The First Step in Web Design: From Getting Started to Giving Up – HKDesign	61
My First Webpage – HKDesign	60
My first web-page: A struggling battle with HTML & CSS – HKDesign	60
Use of Analytics in Product Management Life Cycle	58
1st Assignment: This would be my first personal site! – HKDesign	56
Lambert's First HTML Webpage – HKDesign	56
Finally Survived ! Creating my first page using HTML and CSS – HKDesign	53
My First Attempt Towards the Technical Skill Achievement – HKDesign	48
Summer Island Adventure: My first page made in Hong Kong – HKDesign	45
First Try of HTML and CSS – HKDesign	35
A difficult but interesting journey – My first web page design – HKDesign	33
Thoughts on the first attempt to make a web page with HTML – HKDesign	31



### 1st Hand HK Property Data

The application provides the latest property news and property data on the first hand Hong Kong residential property market.



### 28Hse

28Hse provides information for the Hong Kong property market, including latest property listings by agents and landlords, property transaction data, market...



### Third Street

Third Street Hong Kong provides information for the Hong Kong Property Market, including latest property listings by agents and landlord, property...



### Badatu

Taking bus is a daily routine for many Hong Kong people. "Badatu" is a WeChat Mini Program. Users can create a...



### Blockchain.News

Our Hong Kong Blockchain Event Checklist covers all major FinTech/Blockchain related conferences and meetups held in Hong Kong.



### Citymapper

Citymapper is an iPhone/Android/Web application with the best and most comprehensive source of information for navigating Hong Kong, including information on...



### Datatone

We will gradually add more and more government open data to help everyone making good use of government resources. Besides government...



### Ding Ding

Ding Ding is the ultimate guide for taking ding ding - the 100+ years old tram fleet in Hong Kong.



### Drug Database

An app that provide quick search for registered drugs information in Hong Kong. It includes both registered western and chinese proprietary...

## Visitors

### OPERATING SYSTEM VERSION

### ▼ VISITS

Mac 10.15	423
Windows 10	295
iOS iOS 15.6	34
iOS iOS 15.2	15
iPadOS iOS 14.4	14
Mac 10.13	13
iOS iOS 16.0	12
iOS iOS 14.8	10
Android 12	4
Android 10	3
Android 11	2
iOS iOS 15.4	2
Android 0	1
GNU/Linux	1
HarmonyOS	1
iOS iOS 11.0	1
Mac 10.14	1
Windows 7	1

1–18 of 18

Related report: Operating System families

## Configurations

### CONFIGURATION

### ▼ VISITS

### BROWSER

### ▼ VISITS

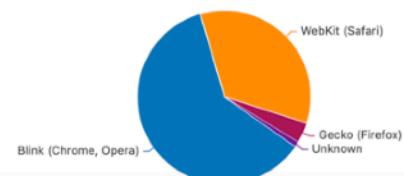
Chrome	480
Safari	198
Chrome Mobile iOS	43
Mobile Safari	42
Firefox	32
Microsoft Edge	22
WeChat	8
Chrome Mobile	3
Huawei Browser Mobile	1
MIUI Browser	1
QQ Browser	1
Quark	1
Vivo Browser	1

1–13 of 13

Related report: Browser version

## Browser engines

Visits



Q Search

FROM 2022-09-12 TO 2022-10-03

ALL VISITS



Dashboard

Visitors

Overview

Visits Log

Real-time

Real-time Map

Locations

Devices

Software

Times

User IDs

Behaviour

Acquisition

Goals

## Visits Over Time



## Visits Overview

819 visits + -10.8%

6,594 pageviews, 3,918 unique pageviews +109.9%

10 min 24s average visit duration +90.2%

0 total searches on your website, 0 unique keywords 0 %

23% visits have bounced (left the website after one page) -43.9%

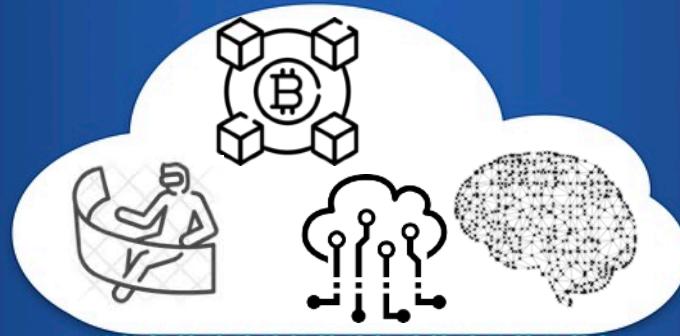
151 downloads, 134 unique downloads +62.4%

9.1 actions (page views, downloads, outlinks and internal site searches) per visit +160%

706 outlinks, 501 unique outlinks +70,500%

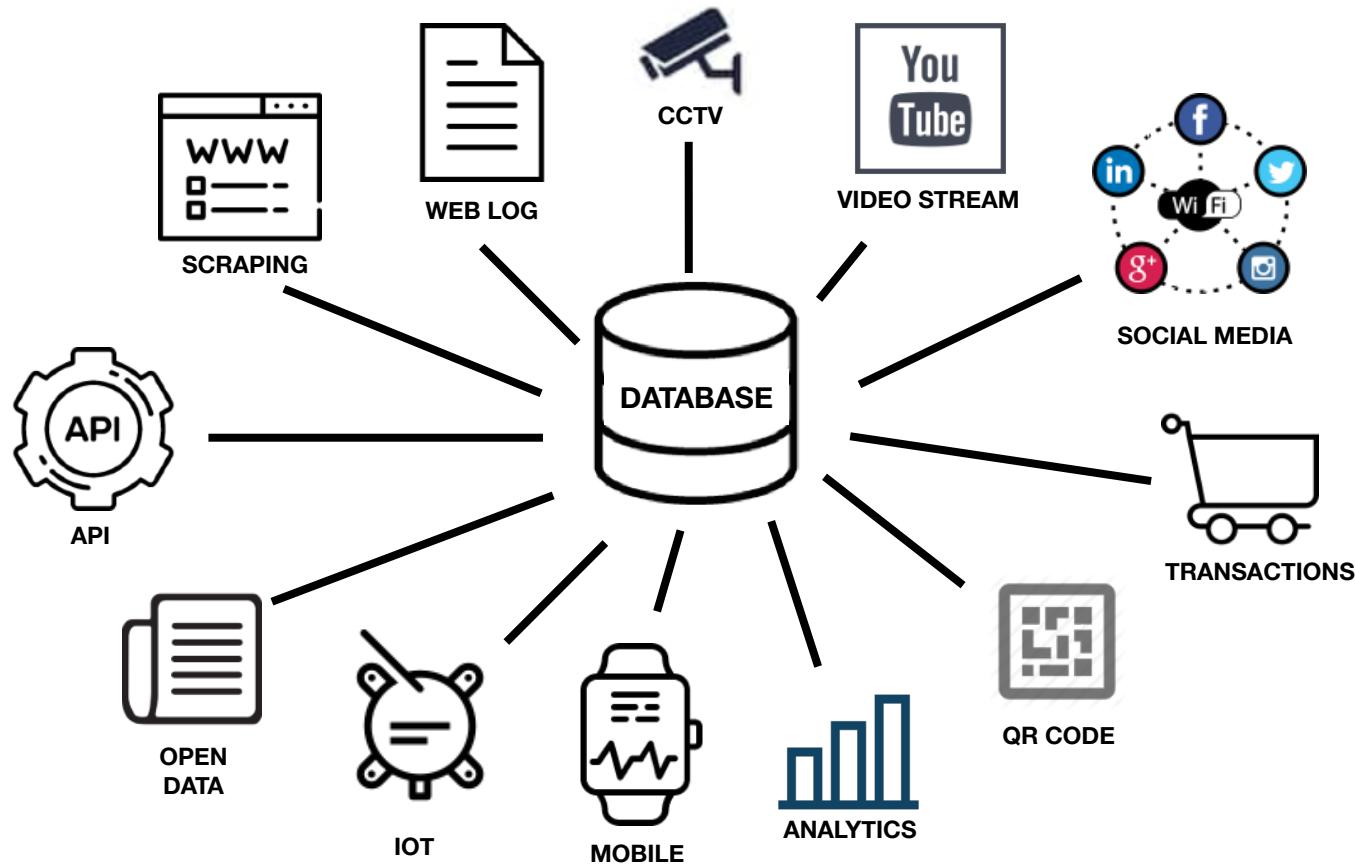
214 max actions in one visit +409.5%

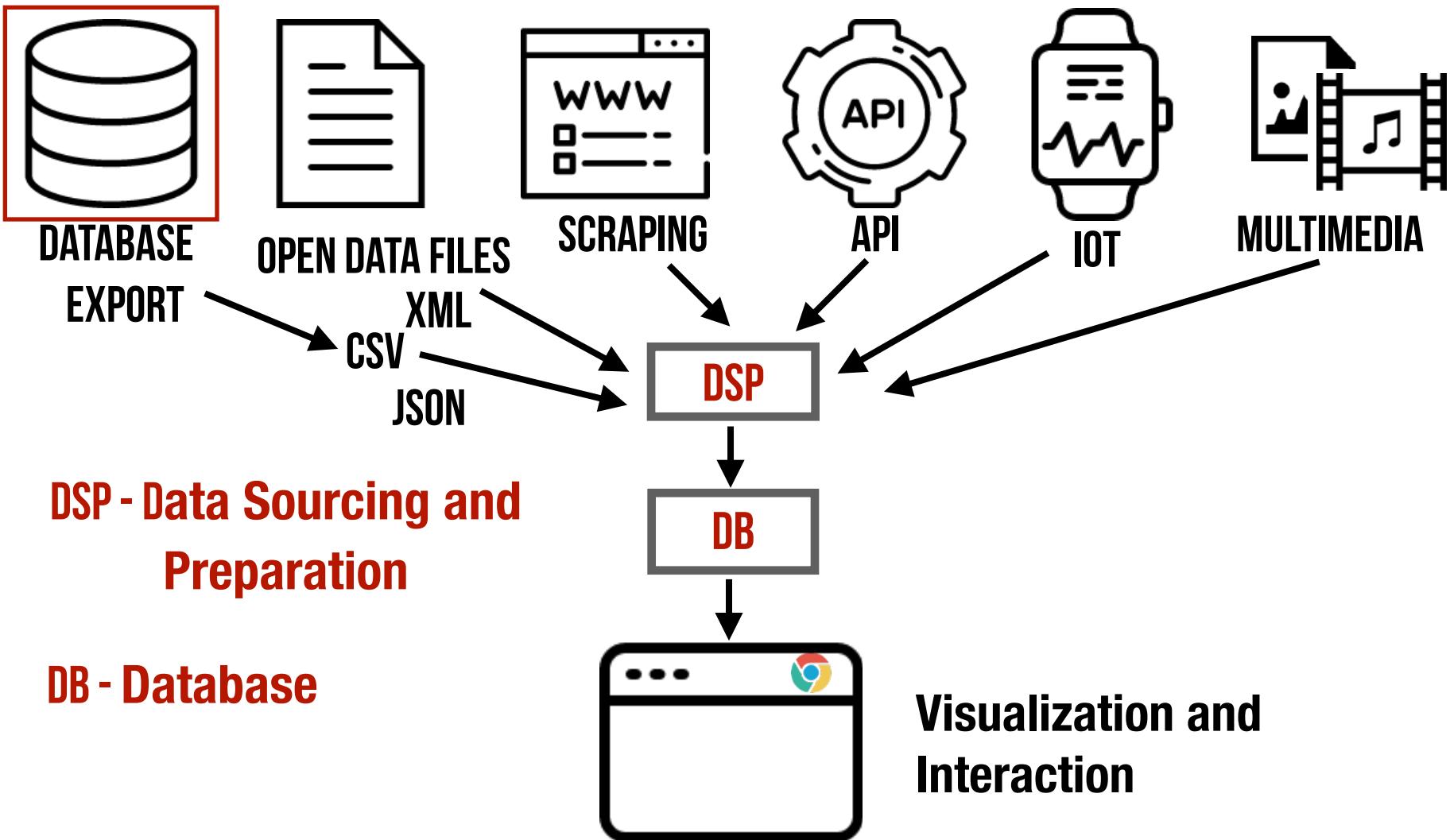
# **Sources of Data Demonstration**



```
1 01 01 00 10 10 01 00 10 01 10 00 10 11 00 11 00 0  
1 01 11 11 10 11 00 11 10 10 01 10 10 11 01 10 11 10 0  
0 10 10 10 10 10 00 00 11 00 00 00 00 00 01 01 11 10 0  
0 11 01 11 00 00 01 00 11 11 00 11 11 01 10 00 10 00 1  
0 00 11 00 01 11 01 10 10 01 00 11 11 11 01 01 01 11 0  
1 11 10 01 01 01 00 01 01 00 01 00 11 00 00 01 10 0  
1 10 01 01 01 11 01 00 10 00 00 10 00 01 10 11 00 10 1  
1 00 10 00 01 00 10 11 01 00 00 10 10 01 10 01 10 0  
1 10 11 00 01 11 00 10 01 11 11 11 11 10 11 01 01 1  
0 00 11 00 11 10 10 00 00 11 01 11 10 11 01 10 01 01 0  
1 01 00 01 01 00 01 10 01 00 00 11 10 00 11 00 10 01 0  
1 10 10 00 10 10 10 11 10 01 10 01 11 00 01 01 11 00 1  
0 00 01 10 01 00 10 00 11 00 10 00 10 00 11 00 00 10 1  
1 01 11 01 01 10 11 00 00 11 10 01 10 00 00 01 10 01 1  
0 11 00 11 01 00 01 11 11 10 10 10 11 11 11 00 00 10 0  
1 11 01 10 11 11 01 00 01 00 01 11 00 00 00 11 00 00 10 0  
1 11 01 00 10 11 01 11 11 10 11 00 11 10 01 01 11 10 01 1  
0 01 01 00 10 11 11 00 11 01 00 10 10 00 00 00 11 11 1  
0 10 11 01 00 01 11 10 11 10 01 01 11 10 00 01 00 00 1  
0 01 11 01 10 10 01 11 00 00 11 00 00 01 10 01 00 01 10 1  
1 00 01 00 11 11 10 10 10 11 11 11 01 11 10 11 10 11 1  
1 10 11 11 10 00 10 01 01 00 11 11 00 11 01 10 10 00 0  
1 11 10 11 11 10 00 10 10 00 10 01 11 01 11 10 00 11 0  
1 10 01 00 10 01 11 00 10 00 10 01 01 01 01 11 10 10 00 1  
0 10 00 11 10 10 01 11 00 10 10 11 00 10 01 11 11 01 1  
0 00 01 11 00 01 11 00 10 00 10 00 00 01 01 11 01 11 00 0  
1 11 11 01 00 10 00 11 10 10 00 11 00 00 10 11 10 10 0  
1 01 11 11 11 11 01 00 11 01 11 00 11 01 10 00 00 01 1  
0 11 00 00 11 01 01 11 01 11 01 00 11 10 11 01 10 10 1  
1 10 10 11 00 00 00 10 00 10 10 00 11 00 11 10 11 11 10 1  
1 10 01 10 11 00 01 01 01 10 10 10 11 11 10 01 01 01 0  
1 11 10 01 01 00 11 01 01 11 00 10 11 00 11 01 10 00 1  
0 01 10 00 01 10 11 10 01 00 00 00 11 00 11 01 00 01 0  
1 11 10 11 10 00 00 11 00 00 11 11 00 11 01 11 01 11 0  
0 00 00 01 10 01 01 10 11 11 10 10 01 01 00 11 11 00 0  
0 11 00 00 00 11 10 01 00 00 00 10 11 11 11 01 01 01 0  
0 11 11 10 11 11 10 00 10 00 01 11 10 11 00 01 00 11 0  
0 11 01 00 01 00 11 10 10 10 10 00 10 11 11 00 00 01 11 1
```

# Structured and Un-structured Data Sources





## Re: Pre-course Training workshop of COMM5961 Topical Studies in New Media: The design of Data Driven Products and Services (30 Aug or 6 Sep 2020)

AM

Fri 8/28/2020 9:53 AM

To: Bernard Suen (MGT)



COMM5961 class list (with U...  
19 KB

Dear Bernard,

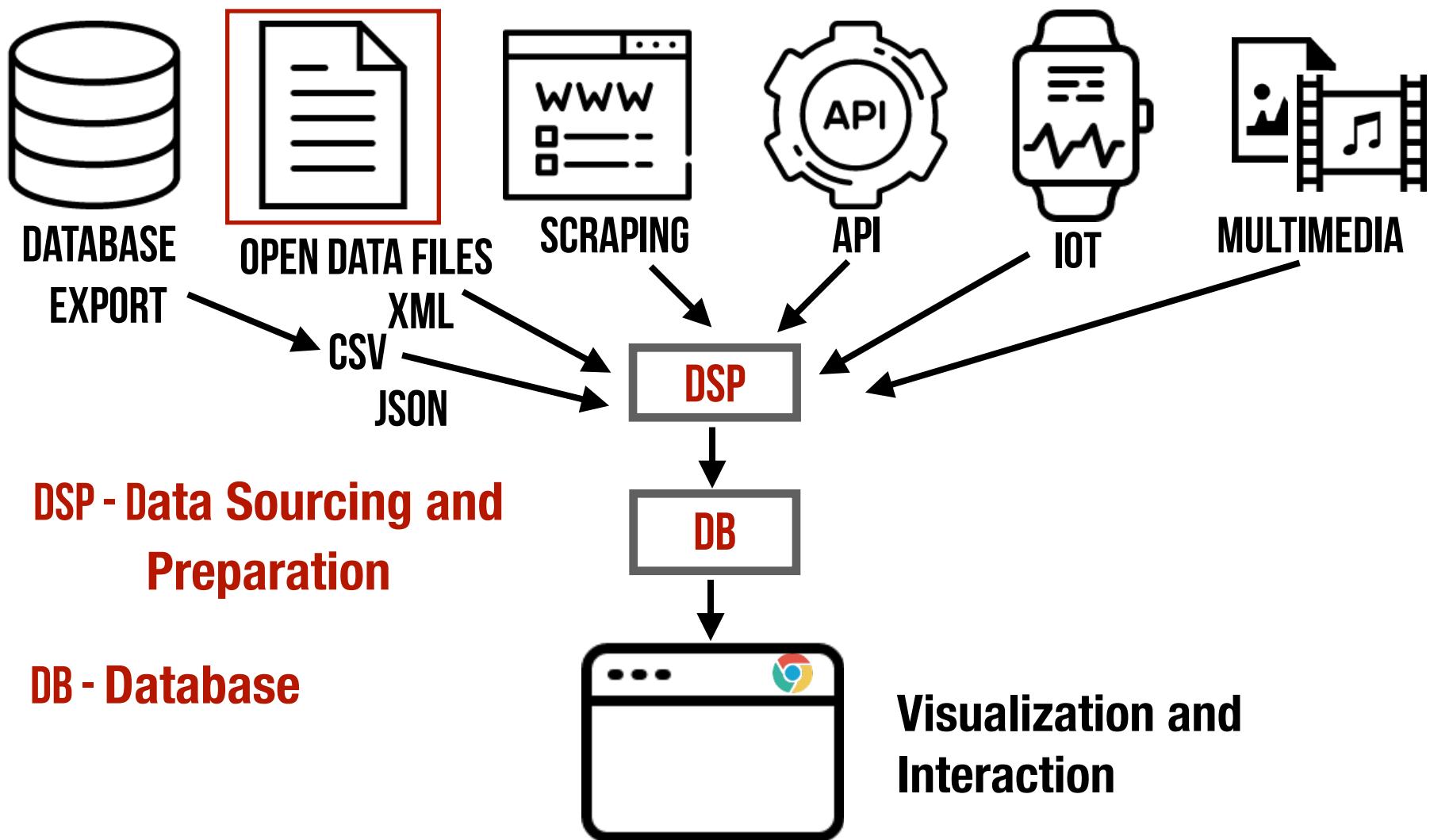
Please find the attached class list of COMM5961 with students' Ug background for your information. Thank you.

Best,

XYZ

Viewing

E25	A	B	C	D	E	F	G
1	No.	ID		Study Mode	Ug_School	Major/Minor	Admit Year
2	1	1155148		F	GUANGDONG UNIVERSITY OF FOREIGN STUDIES - 廣東外語外貿大學 (China)	Culture and Communications	2020
3	2	1155145		F	SHENZHEN UNIVERSITY - 深圳大學 (China)	Internet and New Media	2020
4	3	1155148		F	JINAN UNIVERSITY - 暨南大學 (China)	Journalism	2020
5	4	1155145		P	CITY UNIVERSITY OF HONG KONG (Hong Kong SAR, China)	Media and Communication / Global Spanish Studies	2020
6	5	1155148		F	CENTRAL CHINA NORMAL UNIVERSITY - 華中師範大學 (China)	Network and New Media	2020
7	6	1155145		F	SUN YAT-SEN UNIVERSITY - 中山大學 (China)	English	2020
8	7	1155155		F	JINAN UNIVERSITY - 暨南大學 (China)	Cultural Industry Management / Finance	2020
9	8	1155151		F	JINAN UNIVERSITY - 暨南大學 (China)	Broadcasting and Announcing	2020
10	9	1155152				Italian Language and Literature /TV Editing and Directing (TV Directing)	
11	10	1155148		F	COMMUNICATION UNIVERSITY OF CHINA	Editing and Directing (TV Directing)	2020
12	11	1155145		F	UNIVERSITY OF LIVERPOOL (United Kingdom)	English and Communication Studies	2020
13	12	1155148		F	WUHAN UNIVERSITY - 武漢大學 (China)	Archival Science	2020
14	13	1155151		F	NANKAI UNIVERSITY - 南開大學 (China)	Computer Science	2020
				F	University of Kansas (USA)	Journalism (Strategic Communication)	2020



## 2020-2022 Annual Open Data Plans Are Now Published



New to The Site? [START HERE](#)

Search Data e.g. population 

BROWSE DATASETS

<https://data.gov.hk/en/>

DATASETS

GEOSPATIAL DATA



# CITY DASHBOARD

City dynamics at a glance

New to The Site? **START HERE**

Search Data e.g. population

**BROWSE DATASETS**<https://data.gov.hk/en-data/dataset/hk-lcsd-facility-facility-bkbc>

- Convert JSON to CSV (<https://codebeautify.org/jsonviewer>)
- Convert XML to CSV (<https://www.convertcsv.com/xml-to-csv.htm>)
- Import CSV into Excel or Numbers



HOME DATA LEARN COMMUNITY



ENG ▾

Home > Datasets > Basketball Courts (Free Outdoor Pitches/Courts)

## Basketball Courts (Free Outdoor Pitches/Courts)

LEISURE AND CULTURAL SERVICES DEPARTMENT |

Recreation and Culture



UPDATE FREQUENCY: AS AND WHEN NEW FACILITY IS ADDED OR AMENDMENT IS MADE

Location of Basketball Courts (Free Outdoor Pitches/Courts)

---

Data Dictionary : [https://www.lcsd.gov.hk/datagovhk/facility/facility-bkbc\\_data\\_dictionary.pdf](https://www.lcsd.gov.hk/datagovhk/facility/facility-bkbc_data_dictionary.pdf)

---

1 JSON file(s)



Add All to Queue

[{"District\_en": "Kowloon City", "District\_cn": "九龍城區", "Name\_en": "Carpenter Road Park", "Name\_cn": "賈炳達道公園", "Address\_en": "Carpenter Road, Kowloon City, Kowloon.", "Address\_cn": "九龍城賈炳達道", "GIHS": "9G5i7NFpXL", "Court\_no\_en": "4", "Court\_no\_cn": "4", "Ancillary\_facilities\_en": "<li>Toilets are provided in the Park.<br><li>Other facilities include a jogging track, 7 fitness stations, a cycling track, 2 sets of children's playground, an elderly fitness station, a 7-a-side hard-surface soccer pitch and 2 volleyball courts.<br><li>Barrier Free Facilities: Accessible Toilet, Tactile Guide Path, Braille Directory Map\\Floor Plan<br>", "Ancillary\_facilities\_cn": "<li>公園內設有洗手間<br><li>其他設施包括1條緩跑徑、7個健身站、1條單車徑、2個兒童遊樂場、1個長者健身站、1個7人硬地足球場和2個排球場<br><li>無障礙設施: 暢通易達洗手間、觸覺引路帶、觸覺點字及觸覺平面圖。<br>", "Opening\_hours\_en": "7 am to 11 pm daily", "Opening\_hours\_cn": "每日上午7時至晚上11時", "Phone": "2716 9962", "Remarks\_en": "", "Remarks\_cn": "", "Longitude": "114-11-27", "Latitude": "22-19-52"}, {"District\_en": "Kowloon City", "District\_cn": "九龍城區", "Name\_en": "Ho Man Tin Park", "Name\_cn": "何文田公園", "Address\_en": "No.1 Chung Yee Street, Ho Man Tin, Kowloon.", "Address\_cn": "九龍何文田忠義街一號", "GIHS": "MSKgwBPmtd", "Court\_no\_en": "2", "Court\_no\_cn": "2", "Ancillary\_facilities\_en": "<li>Men's and ladies' changing rooms and toilets<br><li>A fee-charging car park (including 1 designated disabled parking space)<br><li>Other facilities include a hard-surface 7-a-side soccer pitch cum handball court, a children's playground and a jogging track with 6 fitness stations.<br><li>Barrier Free Facilities: Accessible Toilet, Tactile Guide Path, Braille Directory Map\\Floor Plan", "Ancillary\_facilities\_cn": "<li>男、女更衣室及洗手間<br><li>1個收費停車場 (設有1個殘疾人士專用車位)<br><li>其他設施包括1個硬地7人足球場兼手球場、1個兒童遊樂場、1條緩跑徑和6個健身站<br><li>無障礙設施: 暢通易達洗手間、觸覺引路帶、觸覺點字及觸覺平面圖", "Opening\_hours\_en": "7 am to 11 pm daily", "Opening\_hours\_cn": "每日上午7時至晚上11時", "Phone": "2762 7837", "Remarks\_en": "", "Remarks\_cn": "", "Longitude": "114-10-50", "Latitude": "22-18-44"}, {"District\_en": "Kowloon City", "District\_cn": "九龍城區", "Name\_en": "Hoi Sham Park", "Name\_cn": "海心公園", "Address\_en": "Yuk Yat Street, Tokwawan, Kowloon.", "Address\_cn": "九龍土瓜灣旭日街", "GIHS": "nZ2deVMDpF", "Court\_no\_en": "1", "Court\_no\_cn": "1", "Ancillary\_facilities\_en": "<li>Men's and ladies' toilets<br><li>Other facilities include 2 hard-surface 5-a-side soccer, a children's playground and elderly fitness equipment.<br><li>Barrier Free Facilities: Accessible Toilet, Tactile Guide Path, Braille Directory Map\\Floor Plan", "Ancillary\_facilities\_cn": "<li>男、女洗手間<br><li>其他設施包括2個硬地5人足球場、兒童遊樂場及長者健體設施<br><li>無障礙設施: 暢通易達洗手間、觸覺引路帶、觸覺點字及觸覺平面圖", "Opening\_hours\_en": "7 am to 11 pm daily", "Opening\_hours\_cn": "每日上午7時至晚上11時", "Phone": "2334 3576 \ 2762 2083", "Remarks\_en": "", "Remarks\_cn": "", "Longitude": "114-11-30", "Latitude": "22-18-54"}, {"District\_en": "Kowloon City", "District\_cn": "九龍城區", "Name\_en": "Junction Road Park", "Name\_cn": "聯合道公園", "Address\_en": "Junction Road, Kowloon City, Kowloon.", "Address\_cn": "九龍聯合道", "GIHS": "9G5i7NFpXL", "Court\_no\_en": "4", "Court\_no\_cn": "4", "Ancillary\_facilities\_en": "<li>Toilets are provided in the Park.<br><li>Other facilities include a jogging track, 7 fitness stations, a cycling track, 2 sets of children's playground, an elderly fitness station, a 7-a-side hard-surface soccer pitch and 2 volleyball courts.<br><li>Barrier Free Facilities: Accessible Toilet, Tactile Guide Path, Braille Directory Map\\Floor Plan", "Ancillary\_facilities\_cn": "<li>公園內設有洗手間<br><li>其他設施包括1條緩跑徑、7個健身站、1條單車徑、2個兒童遊樂場、1個長者健身站、1個7人硬地足球場和2個排球場<br><li>無障礙設施: 暢通易達洗手間、觸覺引路帶、觸覺點字及觸覺平面圖。<br>", "Opening\_hours\_en": "7 am to 11 pm daily", "Opening\_hours\_cn": "每日上午7時至晚上11時", "Phone": "2762 7837", "Remarks\_en": "", "Remarks\_cn": "", "Longitude": "114-10-50", "Latitude": "22-18-44"}]

# JSON Viewer★

Save & Share

```

1 [{"District_en":"Kowloon City","District_cn":"九龍城區",
  "Name_en":"Carpenter Road Park","Name_cn":"賈炳達道公園",
  "Address_en":"Carpenter Road, Kowloon City, Kowloon."
  , "Address_cn":"九龍城賈炳達道","GIHS":"9G5i7NFpXL"
  , "Court_no_en": "4", "Court_no_cn": "4"
  , "Ancillary_facilities_en": "<li>Toilets are provided in the
Park.<br><li>Other facilities include a jogging track, 7
fitness stations, a cycling track, 2 sets of children's
playground, an elderly fitness station, a 7-a-side hard
-surface soccer pitch and 2 volleyball courts.<br><li>
Barrier Free Facilities: Accessible Toilet, Tactile Guide
Path, Braille Directory Map\\Floor Plan<br>"
  , "Ancillary_facilities_cn": "<li>公園內設有洗手間<br><li>
其他設施包括1條緩跑徑、7個健身站、1條單車徑、2個兒童遊樂場、
1個長者健身站、1個7人硬地足球場和2個排球場<br><li>
無障礙設施：暢通易達洗手間、觸覺引路帶、觸覺點字及觸覺平面
圖。<br>","Opening_hours_en": "7 am to 11 pm daily"
  , "Opening_hours_cn": "每日上午7時至晚上11時", "Phone": "2716
9962", "Remarks_en": "", "Remarks_cn": "", "Longitude": "114-11
-27", "Latitude": "22-19-52"}, {"District_en": "Kowloon City"
  , "District_cn": "九龍城區", "Name_en": "Ho Man Tin Park"
  , "Name_cn": "何文田公園", "Address_en": "No.1 Chung Yee Street,
Ho Man Tin, Kowloon.", "Address_cn": "九龍何文田忠義街一號"
  , "GIHS": "MSKgwBPmtd", "Court_no_en": "2", "Court_no_cn": "2"
  , "Ancillary_facilities_en": "<li>Men's and ladies' changing
rooms and toilets<br><li>A fee-charging car park (including
1 designated disabled parking space)<br><li>Other facilities
include a hard-surface 7-a-side soccer pitch cum handball
court, a children's playground and a jogging track with 6
fitness stations.<br><li>Barrier Free Facilities: Accessible
Toilet, Tactile Guide Path, Braille Directory Map\\Floor
Plan" , "Ancillary_facilities_cn": "<li>公園內設有洗手間<br><li>
其他設施包括1條緩跑徑、7個健身站、1條單車
徑、2個兒童遊樂場、1個長者健身站、1個
7人硬地足球場和2個排球場<br><li>
無障
礙設施：暢通易達洗手間、觸覺引路帶、觸
覺點字及觸覺平面圖。<br>"}]

```

Ln: 1 Col: 281

**Result mode:**

view ▾

Load Url

Browse

Tree Viewer

2 Tab Space ▾

Beautify

① ×

**Website Designer Near You**

Stallions SEO LLC

We want to make you businesses phone ring!

**OPEN**

Minify

Validate

array ▷ 0 ▶

array [228]

▼ 0 {18}

District\_en : Kowloon City  
 District\_cn : 九龍城區  
 Name\_en : Carpenter Road Park  
 Name\_cn : 賈炳達道公園  
 Address\_en : Carpenter Road, Kowloon City, Kowloon.  
 Address\_cn : 九龍城賈炳達道  
 GIHS : 9G5i7NFpXL  
 Court\_no\_en : 4  
 Court\_no\_cn : 4  
 Ancillary\_facilities\_en : <li>Toilets are provided in the
 Park.<br><li>Other facilities
 include a jogging track, 7
 fitness stations, a cycling
 track, 2 sets of children's
 playground, an elderly fitness
 station, a 7-a-side hard-
 surface soccer pitch and 2
 volleyball courts.<br><li>
 Barrier Free Facilities:
 Accessible Toilet, Tactile
 Guide Path, Braille Directory
 Map\\Floor Plan<br>
 Ancillary\_facilities\_cn : <li>公園內設有洗手間<br><li>
 其他設
 施包括1條緩跑徑、7個健身站、1條單車
 級、2個兒童遊樂場、1個長者健身站、1個
 7人硬地足球場和2個排球場<br><li>
 無障
 礙設施：暢通易達洗手間、觸覺引路帶、觸
 觉點字及觸覺平面
 圖。<br>

<https://codebeautify.org/jsonviewer>

District_en	District_cn	Name_en	Name_cn	Address_en
Kowloon City	九龍城區	Carpenter Road Park	賈炳達道公園	Carpenter Road, Kowloon City, Kowloon.
Kowloon City	九龍城區	Ho Man Tin Park	何文田公園	No.1 Chung Yee Street, Ho Man Tin, Kowloon.
Kowloon City	九龍城區	Hoi Sham Park	海心公園	Yuk Yat Street, Tokwawan, Kowloon.
Kowloon City	九龍城區	Junction Road Park	聯合道公園	195 Junction Road, Kowloon City.
Kowloon City	九龍城區	Kam Shing Road Recreation Ground	金城道遊樂場	Kam Shing Road, Kowloon.
Kowloon City	九龍城區	Kau Pui Lung Road Playground	靠背壘道遊樂場	Kau Pui Lung Road, Kowloon.
Kowloon City	九龍城區	Kent Road Garden	根德道花園	Kent Road, Kowloon.
Kowloon City	九龍城區	King Wan Street Playground	景雲街遊樂場	King Wan Street, To Kwa Wan, Kowloon.
Kowloon City	九龍城區	King's Park High Level Service Reservoir Playground	京士柏上配水庫遊樂場	Chung Hau Street, Ho Man Tin, Kowloon.
Kowloon City	九龍城區	Kowloon Tsai Park	九龍仔公園	13 Inverness Road, Kowloon City, Kowloon.
Kowloon City	九龍城區	Lung Cheung Road Playground	龍翔道遊樂場	Beacon Hill Road, Kowloon.
Kowloon City	九龍城區	Ma Tau Wai Road Playground	馬頭圍道遊樂場	Ma Tau Wai Road, Tokwawan, Kowloon.
Kowloon City	九龍城區	Oxford Road Playground	牛津道遊樂場	Oxford Road, Kowloon.
Kowloon City	九龍城區	Peace Avenue Playground	太平道遊樂場	Peace Avenue, Ho Man Tin, Kowloon.
Kowloon City	九龍城區	Perth Street Sports Ground	巴富街運動場	Shek Ku Street , Ho Man Tin, Kowloon.
Kowloon City	九龍城區	Pui Ching Road Playground	培正道遊樂場	Pui Ching Road, Ho Man Tin, Kowloon.
Kowloon City	九龍城區	Rutland Quadrant Children's Playground	律倫街兒童遊樂場	Rutland Quadrant, Kowloon.
Kowloon City	九龍城區	Sung Wong Toi Playground	宋王臺遊樂場	Sung Wong Toi Road, Kowloon.
Kowloon City	九龍城區	Tai Wan Road Playground	大環道遊樂場	Tai Wan Road, Hung Hom, Kowloon.
Kowloon City	九龍城區	Tai Wan Shan Park	大環山公園	Wan Hoi Street, Hung Hom, Kowloon.
Kowloon City	九龍城區	To Kwa Wan Recreation Ground	土瓜灣遊樂場	66 Ha Ha Ling Road, To Kwa Wan, Kowloon



Home > Datasets > FEHD facility and service locations > FEHD facility and service locations (English)

## FEHD facility and service locations (English)

FOOD AND ENVIRONMENTAL HYGIENE DEPARTMENT

Health

XML

API Available



LAST UPDATED ON: 11/09/2020

UPDATE FREQUENCY: AS AND WHEN NECESSARY

Locations and details about different facilities and services provided by the Food and Environmental Hygiene Department.

This dataset provides geo-referenced information (including district, type of facility/service, name, address, contact numbers, opening hours and coordinates) about different types of facilities/services provided by the Food and Environmental Hygiene Department (including public toilets, public refuse collection points, public markets, cooked food markets, etc).

URL: [http://www.fehd.gov.hk/english/map/fehd\\_map\\_e.xml](http://www.fehd.gov.hk/english/map/fehd_map_e.xml)

Download

Add to Queue

<https://data.gov.hk/en-data/dataset/hk-fehd-fehdlocatn-fehd-facility-and-service-locations>

[Enter Data](#)[Choose File](#)[Enter URL](#)

Choose File

[Choose file fehd\\_map\\_e.xml](#)

Encoding

-Default-

[Clear Input](#)

Examples:

1 2

## Step 2: Choose output options (optional) ▾

## Step 3: Generate output

[Convert XML To CSV](#)[XML To Excel](#)

```
mapID,districtID,map_type,name_e,address_e,contact1,contact2,fax,openHr_e,google_coordinate,remarks_e,n  
1,CW,toilet,Wa Hing Lane Public Toilet and Bathhouse,"Junction of Shing Wong Street andamp; Wa Hing Lane, S  
7,E,toilet,Sheung On Street Public Toilet,Opposite to New World Bus Depot Sheung On Street,N/A,N/A,N/A,24 ho  
8,E,toilet,Tung Hei Road Public Toilet,No. 28 Tung Hei Road,N/A,N/A,N/A,24 hours,"22.281325,114.230214",N/A  
9,Wch,toilet,Hing Fat Street Public Toilet,Hing Fat Street near the Entrance of Victoria Park,N/A,N/A,N/A,24 hours  
10,Wch,toilet,Wing Hing Street Public Toilet,No. 15 Wing Hing Street,N/A,N/A,N/A,24 hours,"22.285273,114.1916  
11,E,toilet,Oil Street Public Toilet,At side of No. 3 Oil Street,N/A,N/A,N/A,24 hours,"22.288196,114.193093",N/A,,  
13,E,toilet,Ning Foo Street Public Toilet,Ning Foo Street near Bus Terminus,N/A,N/A,N/A,24 hours,"22.264876,11  
14,E,toilet,Kam Wa Street Public Toilet,Junction of Kam Wa Street and Wang Wa Street,N/A,N/A,N/A,24 hours,"2  
15,E,toilet,A Kung Ngam Village Road Temporary Public Toilet,Junction of A Kung Ngam Village Road and Tung V  
16,E,toilet,Nam On Street Public Toilet,No. 81 Nam On Street,N/A,N/A,N/A,24 hours,"22.279074,114.226717",N/A,  
17,E,toilet,Hoi Ning Street Public Toilet,Junction of Hoi Ning Street and Hing Man Street,N/A,N/A,N/A,24 hours,"2  
18,E,toilet,Quarry Bay Market Public Toilet,No. 38 Quarry Bay Street,N/A,N/A,N/A,24 hours,"22.283692,114.2118  
19,E,toilet,Pak Fuk Road Public Toilet,Junction of Pak Fuk Road and Healthy Street Central,N/A,N/A,N/A,24 hour  
20,E,toilet,Java Road Public Toilet,Java Road outside North Point Vehicular Ferry Pier,N/A,N/A,N/A,24 hours,"22.
```

Save your result: 

.csv

[Save to Disk](#)

EOL: CRLF ▾

<https://www.convertcsv.com/xml-to-csv.htm>

## Use VLOOKUP to lookup district names

	A	B	C	D	E	F	G	H	I	J	K	L	M
mapID	districtID	map_type	name_e	address	contact1	contact2	fax	openHr_e	google_coordinate	remarks	name_e	name_e	
1	CW	toilet	Wa Hing Junction	N/A	N/A	N/A	24 hours	22.283733,114.151492	N/A				
7	E	toilet	Sheung Opposite	N/A	N/A	N/A	24 hours	22.27220,114.24361	N/A				
8	E	toilet	Tung Hei No. 28 T	N/A	N/A	N/A	24 hours	22.281325,114.230214	N/A				
9	Wch	toilet	Hing Fat Hing Fat	N/A	N/A	N/A	24 hours	22.282368,114.191001	N/A				
10	Wch	toilet	Wing Hir No. 15 W	N/A	N/A	N/A	24 hours	22.285273,114.191676	Suspension Period: 1 April				
11	E	toilet	Oil Street At side c	N/A	N/A	N/A	24 hours	22.288196,114.193093	N/A				
13	E	toilet	Ning Foc Ning Foc	N/A	N/A	N/A	24 hours	22.264876,114.237199	N/A				
14	E	toilet	Kam Wa Junction	N/A	N/A	N/A	24 hours	22.277604,114.230554	N/A				
15	E	toilet	A Kung N Junction	N/A	N/A	N/A	24 hours	22.282208,114.232496	N/A				
16	E	toilet	Nam On No. 81 N	N/A	N/A	N/A	24 hours	22.279074,114.226717	N/A				
17	E	toilet	Hoi Ning Junction	N/A	N/A	N/A	24 hours	22.281421,114.223226	N/A				
18	E	toilet	Quarry E No. 38 Q	N/A	N/A	N/A	24 hours	22.283692,114.211811	Suspension Period: 7 Septe				
19	E	toilet	Pak Fuk Junction	N/A	N/A	N/A	24 hours	22.291136,114.205547	N/A				
20	E	toilet	Java Roa Java Roa	N/A	N/A	N/A	24 hours	22.293075,114.203447	Suspension Period: 4 May 2				
21	E	toilet	Tong Shu Junction	N/A	N/A	N/A	24 hours	22.29197,114.19864	N/A				
22	Wch	toilet	Lin Fa Ki Lily Stre	N/A	N/A	N/A	24 hours	22.2792,114.1928	N/A				
24	Wch	toilet	Bowen R Bowen R	N/A	N/A	N/A	24 hours	22.269939,114.171939	N/A				
26	Wch	toilet	Sing Wo Junction	N/A	N/A	N/A	24 hours	22.269382,114.18519	Suspension Period: 11 May				
27	Wch	toilet	Tai Hang Junction	N/A	N/A	N/A	24 hours	22.276699,114.19488	N/A				
29	Wch	toilet	Wan Cha No. 258	N/A	N/A	N/A	24 hours	22.275004,114.17383	Suspension Period: 30 June				
30	Wch	toilet	Warren S Junction	N/A	N/A	N/A	24 hours	22.278300,114.19175	Suspension Period: 17 Augu				
31	Wch	toilet	Moreton Tung Lo	N/A	N/A	N/A	24 hours	22.278955,114.18993	N/A				
32	Wch	toilet	Wong Na Opposite	N/A	N/A	N/A	24 hours	22.274314,114.18326	N/A				
33	Wch	toilet	Morrison Morrison	N/A	N/A	N/A	24 hours	22.274882,114.17999	N/A				

_ID	__text			
CW	Central and Western			
E	Eastern			
S	Southern			
Wch	Wan Chai			
Is	Islands			
YT	Yau Tsim			
MK	Mong Kok			
SSP	Sham Shui Po			
KC	Kowloon City			
WTS	Wong Tai Sin			
KT	Kwun Tong			
TW	Tsuen Wan			
KwT	Kwai Tsing			
N	North			
TP	Tai Po			
SK	Sai Kung			
ST	Sha Tin			
TM	Tuen Mun			
YL	Yuen Long			

<https://www.ablebits.com/office-addins-blog/vlookup-google-sheets-example/>

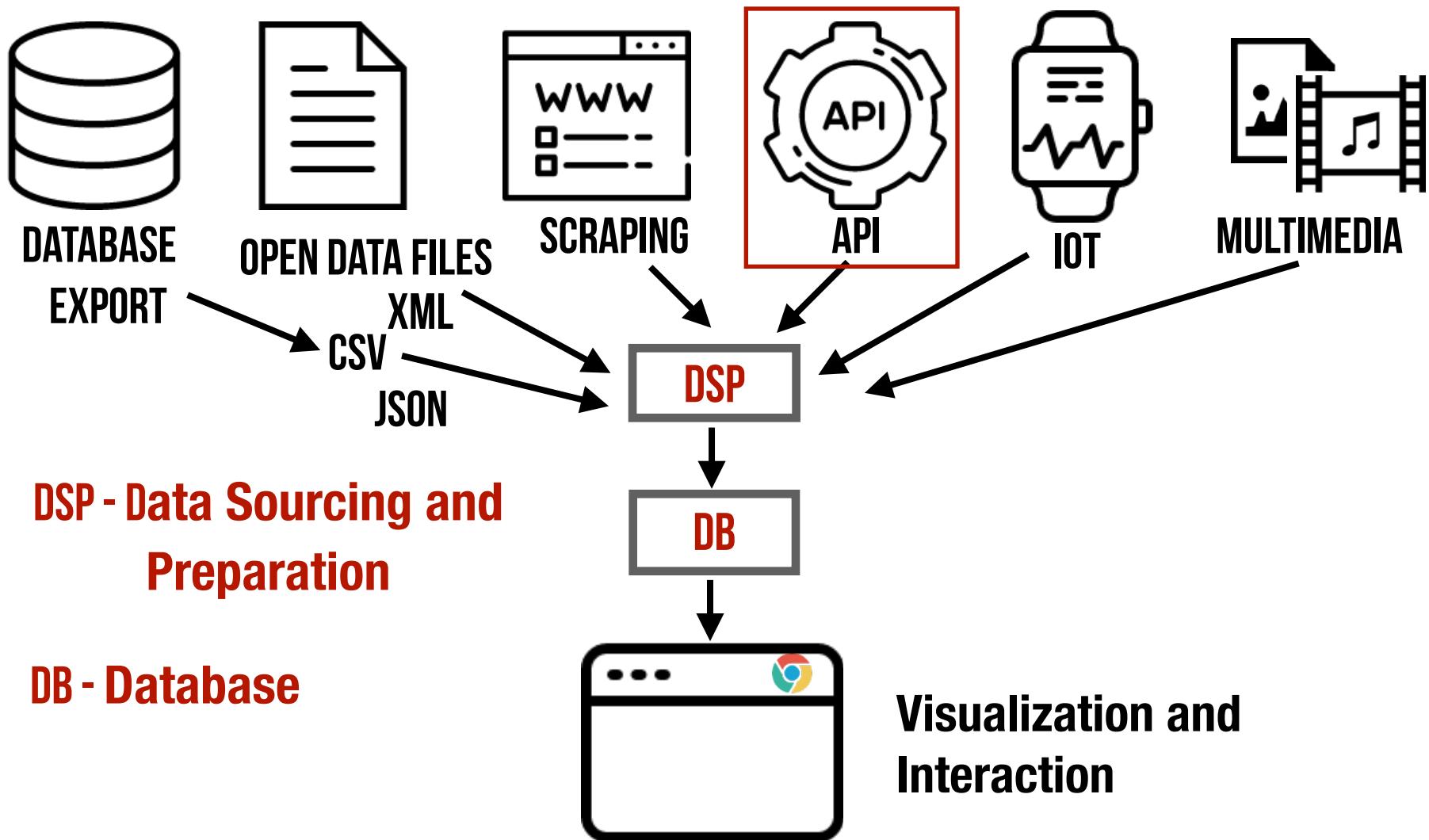
fehd\_map\_district

File Edit View Insert Format Data Tools Extensions Help Last edit was seconds ago

=VLOOKUP(A2, \$P\$2:\$Q\$20, 2, false)

M2	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	districtID	districtName	map_type	name_c	address_c	contact1	contact_fax	openHr_c	map_coordinate	updateDate	remarks_c	中西區	東區	東區	東區	南區	中西區
2	CW	中西區	toilet	華興里公廁及浴	上環城皇街及華	N/A	N/A	24 小時	22.283733,114.1	2020-07-15 15:3	N/A	中西區	東區	東區	東區	中西區	CW
3	E	東區	toilet	常安街公廁	常安街新巴車廠	N/A	N/A	24 小時	22.27220,114.2	2022-03-28 8:58	N/A	東區	東區	東區	東區	東區	E
4	E	東區	toilet	東喜道公廁	東喜道 28 號	N/A	N/A	24 小時	22.281325,114.2	2015-01-24 1:10	N/A	東區	東區	東區	東區	東區	S
5	Wch	灣仔區	toilet	興發街公廁	興發街近維多利	N/A	N/A	24 小時	22.282368,114.1	2016-03-30 22:1	N/A	灣仔區	灣仔區	灣仔區	灣仔區	灣仔區	Wch
6	Wch	灣仔區	toilet	永興街公廁	永興街 15 號	N/A	N/A	24 小時	22.285273,114.1	2020-12-15 15:5	N/A	灣仔區	灣仔區	灣仔區	灣仔區	灣仔區	Is
7	E	東區	toilet	油街公廁	油街3號側	N/A	N/A	24 小時	22.288196,114.1	2020-08-24 9:26	N/A	東區	東區	東區	東區	東區	YT
8	E	東區	toilet	寧富街公廁	寧富街近柴灣站	N/A	N/A	24 小時	22.264876,114.2	2021-06-28 15:0	N/A	東區	東區	東區	東區	東區	MK
9	E	東區	toilet	金華街公廁	金華街與宏華街	N/A	N/A	24 小時	22.277604,114.2	2011-09-05 16:0	N/A	東區	東區	東區	東區	東區	SSP
10	E	東區	toilet	阿公岩村道臨時	阿公岩村道與東	N/A	N/A	24 小時	22.282208,114.2	2018-03-15 14:4	N/A	東區	東區	東區	東區	東區	KC
11	E	東區	toilet	南安街公廁	南安街 81 號	N/A	N/A	24 小時	22.279074,114.2	2020-07-07 12:1	N/A	東區	東區	東區	東區	東區	WTS
12	E	東區	toilet	海寧街公廁	海寧街與興民街	N/A	N/A	24 小時	22.281421,114.2	2011-09-05 16:0	N/A	東區	東區	東區	東區	東區	KT
13	E	東區	toilet	鰂魚涌街公廁	鰂魚涌街 38 號	N/A	N/A	24 小時	22.283692,114.2	2021-01-11 11:0	N/A	東區	東區	東區	東區	東區	TW
14	E	東區	toilet	百福道公廁	百福道與健康中	N/A	N/A	24 小時	22.291136,114.2	2021-02-18 9:19	N/A	東區	東區	東區	東區	東區	KwT
15	E	東區	toilet	渣華道公廁	渣華道北角汽車	N/A	N/A	24 小時	22.293075,114.2	2020-11-26 10:3	N/A	東區	東區	東區	東區	東區	N
16	E	東區	toilet	糖水道公廁	糖水道與春秧街	N/A	N/A	24 小時	22.29197,114.1	2020-08-24 9:26	N/A	東區	東區	東區	東區	東區	TP
17	Wch	灣仔區	toilet	蓮花宮公廁	大坑蓮花宮蓮花	N/A	N/A	24 小時	22.2792,114.192	2011-08-25 9:48	N/A	灣仔區	灣仔區	灣仔區	灣仔區	灣仔區	SK
18	Wch	灣仔區	toilet	寶雲道公廁	寶雲道近灣仔峽	N/A	N/A	24 小時	22.269939,114.1	2017-10-19 1:19	N/A	灣仔區	灣仔區	灣仔區	灣仔區	灣仔區	ST
19	Wch	灣仔區	toilet	成和道公廁	成和道與奕蔭街	N/A	N/A	24 小時	22.269382,114.1	2021-01-08 10:1	N/A	灣仔區	灣仔區	灣仔區	灣仔區	灣仔區	TM
20	Wch	灣仔區	toilet	大坑道公廁	大坑道與勵德邨	N/A	N/A	24 小時	22.276699,114.1	2021-12-07 17:0	N/A	灣仔區	灣仔區	灣仔區	灣仔區	灣仔區	YL
21	Wch	灣仔區	toilet	灣仔街市公廁	皇后大道東 258 號	N/A	N/A	24 小時	22.275004,114.1	2021-07-06 13:0	N/A	灣仔區	灣仔區	灣仔區	灣仔區	灣仔區	
22	Wch	灣仔區	toilet	華倫街公廁	華倫街與施弼街	N/A	N/A	24 小時	22.278300,114.1	2021-04-08 9:30	N/A	灣仔區	灣仔區	灣仔區	灣仔區	灣仔區	
23	Wch	灣仔區	toilet	摩頓臺公廁	銅鑼灣徑巴士總	N/A	N/A	24 小時	22.278955,114.1	2011-09-05 16:0	N/A	灣仔區	灣仔區	灣仔區	灣仔區	灣仔區	
24	Wch	灣仔區	toilet	黃泥涌道公廁	黃泥涌道 95 號對	N/A	N/A	24 小時	22.274314,114.1	2020-07-02 17:4	N/A	灣仔區	灣仔區	灣仔區	灣仔區	灣仔區	
25	Wch	灣仔區	toilet	摩利臣山道公廁	摩利臣山道近體	N/A	N/A	24 小時	22.274882,114.1	2022-05-17 9:57	暫停服務	灣仔區	灣仔區	灣仔區	灣仔區	灣仔區	
26	Wch	灣仔區	toilet	堅拿道公廁	堅拿道與駱克道	N/A	N/A	24 小時	22.280177,114.1	2014-06-04 23:4	N/A	灣仔區	灣仔區	灣仔區	灣仔區	灣仔區	
27	Wch	灣仔區	toilet	交加里公廁	交加里與普樂里	N/A	N/A	24 小時	22.276349,114.1	2011-09-05 16:0	N/A	灣仔區	灣仔區	灣仔區	灣仔區	灣仔區	
28	Wch	灣仔區	toilet	修頓中心公廁	軒尼詩道 130 號附	N/A	N/A	24 小時	22.277268,114.1	2020-12-07 12:5	N/A	灣仔區	灣仔區	灣仔區	灣仔區	灣仔區	
29	Wch	灣仔區	toilet	吉安街公廁	吉安街與堅彌地	N/A	N/A	24 小時	22.273225,114.1	2011-09-27 11:4	N/A	灣仔區	灣仔區	灣仔區	灣仔區	灣仔區	
30	Wch	灣仔區	toilet	博覽道東公廁	博覽道東 (金紫荊	N/A	N/A	24 小時	22.284493,114.1	2015-01-28 22:0	N/A	灣仔區	灣仔區	灣仔區	灣仔區	灣仔區	

Lookup Data Range





百度地图  
开放平台

首页

功能与服务

解决方案

开发文档

合作咨询

控制台

登录

# 百度地图发布全球API试用版

拉开海外地图开发市场序幕 · 全球POI超过1.5亿 · 覆盖209个国家和地区



Android定位SDK v7.4 **NEW!**  
优化离线定位内部策略



2018年春运出行仪表盘 **NEW!**  
春运综合出行信息播报



iOS地图SDK v3.4.4  
适配 iOS 11 定位



iOS定位SDK v1.1  
适配 iOS 11 永久定位的设置



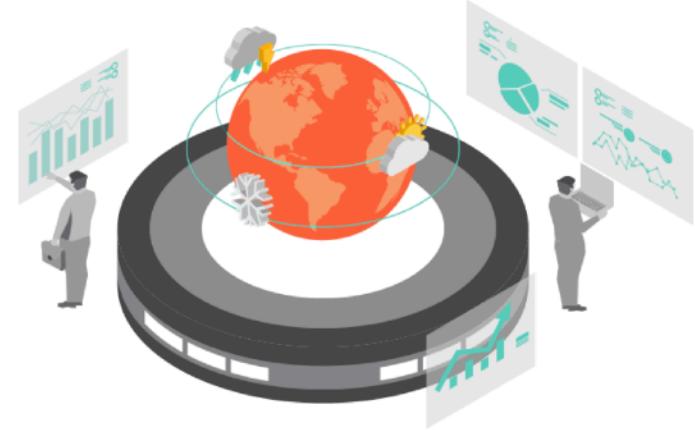
# **API Demonstration (in JSON format)**

# OpenWeather global services

Weather forecasts, nowcasts and history in fast and elegant way

2 Billion Forecasts Per Day  
2,500 new subscribers a day

2,600,000 customers  
20+ weather APIs



Search city

Search



Different Weather?

Metric: °C, m/s

Imperial: °F, mph

3:39pm, Sep 12

London, GB



<https://openweathermap.org/>



Weather in your city

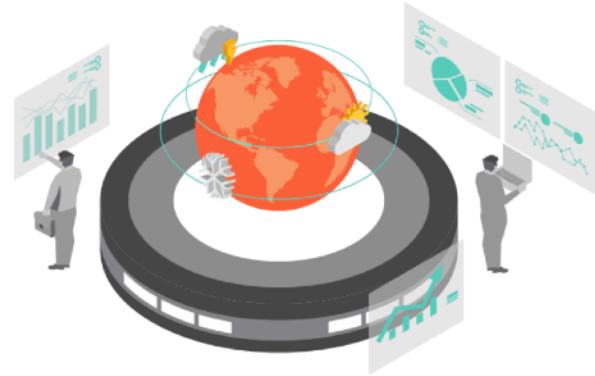
Get Started API Pricing Maps FAQ Partners Blog Marketplace Sign in Support

# OpenWeather global services

Weather forecasts, nowcasts and history in fast and elegant way

2 Billion Forecasts Per Day  
2,500 new subscribers a day

2,600,000 customers  
20+ weather APIs



Search city

Search



Different Weather?

Metric: °C, m/s

Imperial: °F, mph

3:39pm, Sep 12

London, GB

21°C



## By city name

### Description:

You can call by city name or city name, state code and country code. API responds with a list of weather parameters that match a search request.

### API call:

```
api.openweathermap.org/data/2.5/weather?q={city name}&appid={your api key}
```

```
api.openweathermap.org/data/2.5/weather?q={city name},{state code}&appid={your api key}
```

```
api.openweathermap.org/data/2.5/weather?q={city name},{state code},{country code}&appid={your api key}
```

### Parameters:

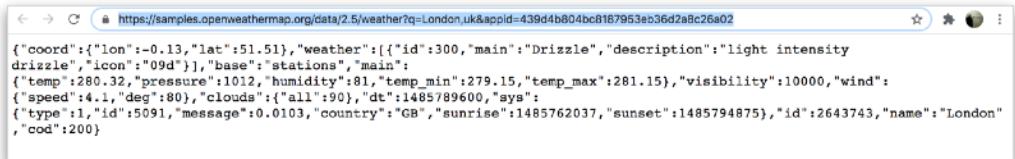
**q** city name, state code and country code divided by comma, use ISO 3166 country codes. You can specify the parameter not only in English. In this case, the API response should be returned in the same language as the language of requested location name if the location is in our predefined list of more than 200,000 locations.

### Examples of API calls:

```
api.openweathermap.org/data/2.5/weather?q=London
```

```
api.openweathermap.org/data/2.5/weather?q=London,uk
```

- List of condition codes
- Min/max temperature in current weather API and forecast API
- Other features
  - Format
  - Units format
  - Multilingual support
  - Call back function for JavaScript code



Searching by states available only for the USA locations.

There is a possibility to receive a central district of the city/town with its own parameters

<https://openweathermap.org/current>

**Code Beautify**

JSON viewer

Sample

```
1 [{"coord": {"lon": -0.13, "lat": 51.51}, "weather": [{"id": 300, "main": "Drizzle", "description": "light intensity drizzle", "icon": "09d"}], "base": "stations", "main": {"temp": 280.32, "pressure": 1012, "humidity": 81, "temp_min": 279.15, "temp_max": 281.15}, "visibility": 10000, "wind": {"speed": 4.1, "deg": 80}, "clouds": {"all": 90}, "dt": 1485789600, "sys": {"type": 1, "id": 5091}, "message": 0.0103, "country": "GB", "sunrise": 1485762037, "sunset": 1485794875}, "id": 2643743, "name": "London", "cod": 200}]
```

Ln: 1 Col: 460

**Result mode:**

**Website Designer Near You**  
Stallions SEO LLC  
We want to make you businesses phone ring!

object ► coord ▾  
    ► weather [1]  
        ▼ 0 {4}  
            id : 300  
            main : Drizzle  
            description : light intensity drizzle  
            icon : 09d  
            base : stations  
    ► main {5}  
        temp : 280.32  
        pressure : 1012  
        humidity : 81  
        temp\_min : 279.15  
        temp\_max : 281.15  
        visibility : 10000  
    ► wind {2}  
    ► clouds {1}  
        dt : 1485789600  
    ► sys {6}  
        id : 2643743  
        name : London  
        cod : 200

<https://codebeautify.org/jsonviewer>

## Weather API

City:

overcast clouds

**17°**

Paris

## Weather API

City:

light rain

**25°**

Shanghai

## Weather API

City:

light rain

**30°**

Hong Kong



# Meituan API Endpoints Example

<https://developer.meituan.com/ka/docs/wmapi-1>

到餐API文档

> 团购API

> 茶饮版 API

智能版 API

门店-获取营业日起...

获取授权资源ID列表

门店-获取物品（分...

总部-获取物品（分...

门店-获取物品类别...

总部-获取物品类别...

门店-获取供应商（...

总部-获取供应商（...

门店-获取物品单位...

总部-获取物品单位 ...

### HTTP请求示例

```
(提示: 实际请求需要进行 urlencode 编码)
POST /rms/base/v1/auth/resources/get HTTP/1.1
Host: api-open-cater.meituan.com
Content-Type: application/x-www-form-urlencoded; charset=utf-8

appAuthToken=qwe&
timestamp=123&
sign=sdsadfsfd&
charset=utf-8&
developerId=100&
version=1&
biz={}
```

### REQUEST

### 响应示例

#### Json示例

```
{
  "code": "OP_SUCCESS",
  "msg": "成功",
  "traceId": 123,
  "data": {
    "resources": {
      "orgType": "1 总部, 5 门店",
      "rootOrgId": 123,
      "orgId": 123
    }
  }
}
```

### RESPONSE

评价管理

外卖门店在延迟发配送名单内才能设置延迟时间

活动管理

订单管理

订单推送

取消订单推送

退款信息推送

图片上传

门店装修

安心卡

在线联系IM

接口code码集合

&gt; 到餐API文档

&gt; 茶饮版 API

API接口	接口描述
poi/save	创建或更新门店信息
poi/getids	获取门店ID
poi/mget	批量获取门店详细信息
poi/close	门店设置为休息状态
poi/online	门店设置为上线状态
poi/updatepromoteinfo	更改门店公告信息
poiTag/list	获取门店品类列表
shippingtime/update	更新门店营业时间
poi/logistics/isDelayPush	查询门店是否延迟发配送
poi/logistics/setDelayPush	设置门店延迟发配送时间

## 产品经理，你该懂点API知识了！



大房同学的产品之路

关注

1 2019.06.16 20:06:41

字数 2,989

阅读 971

13赞

赏

赞赏

回复

下载App

还记得5月，人生第一个产品经理面试，面试官问你知道api吗，当时脑袋一懵。事后，决定一定要把它弄清楚。在查询资料后发现，API确实对于产品经理来说那是相当重要：1.对技术理解更深刻，结合已有新的接口，让产品有更丰富的可行性，并降低工作量。2.更好地预估开发工作量，制定合适产品方案

文章将从API介绍，分类，API文档内容和案例四个方面进行介绍。

### 1.API是什么

全名：Application Programming Interface 中文名：应用程序编程接口

百度百科释义：API是一些预先定义的函数，目的是提供应用程序与开发人员得以访问一组例程的能力，而又无需访问源码，或理解内部工作机制的细节。

看到这你是不是和我一样有点懵，这也太抽象了。于是我又找了一些相关文

[https://www.jianshu.com/p/c1b107be9879?utm\\_campaign=maleskine&utm\\_content=note&utm\\_medium=seo\\_notes&utm\\_source=recommendation](https://www.jianshu.com/p/c1b107be9879?utm_campaign=maleskine&utm_content=note&utm_medium=seo_notes&utm_source=recommendation)

## 产品经理，你要懂点API接口知识！



LinKID

关注作者

2019-04-11

40评论 · 2.9万浏览 · 292收藏 · 10分钟

业内首套系统教学课程，从0-1快速掌握移动APP核心设计模块和逻辑，学完打造专属APP，了解一下 >>



释放双眼，带上耳机，听听看~!

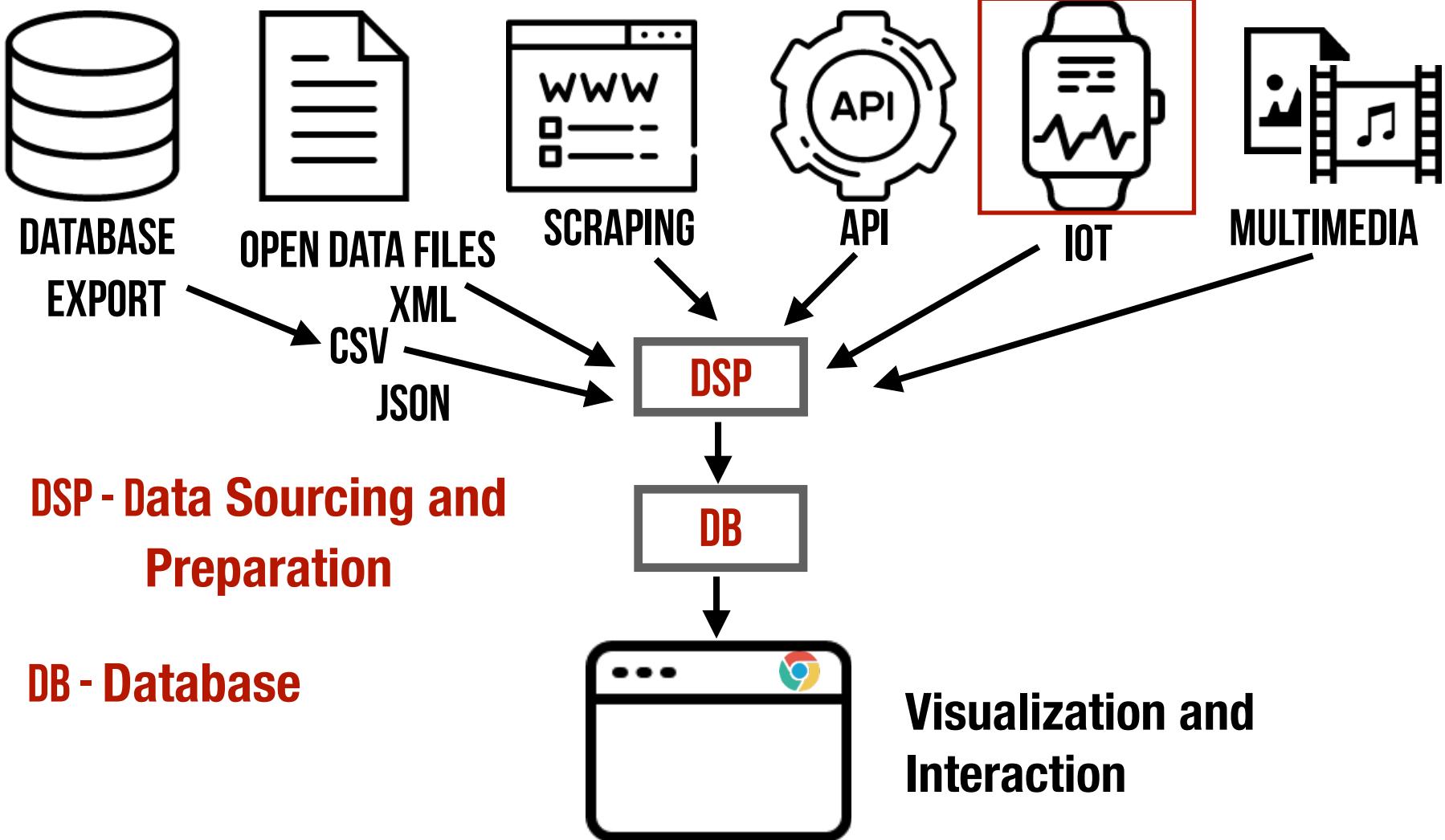
00:00

00:00



产品经理不需要深入地去了解各个接口的实现原理，毕竟术业有专攻，但是了解什么场景应该使用什么样的接口还是很有必要的，可以方便更好地对外提供数据服务。

<http://www.woshipm.com/pmd/2203403.html>



# Popular IoT Prototyping Boards



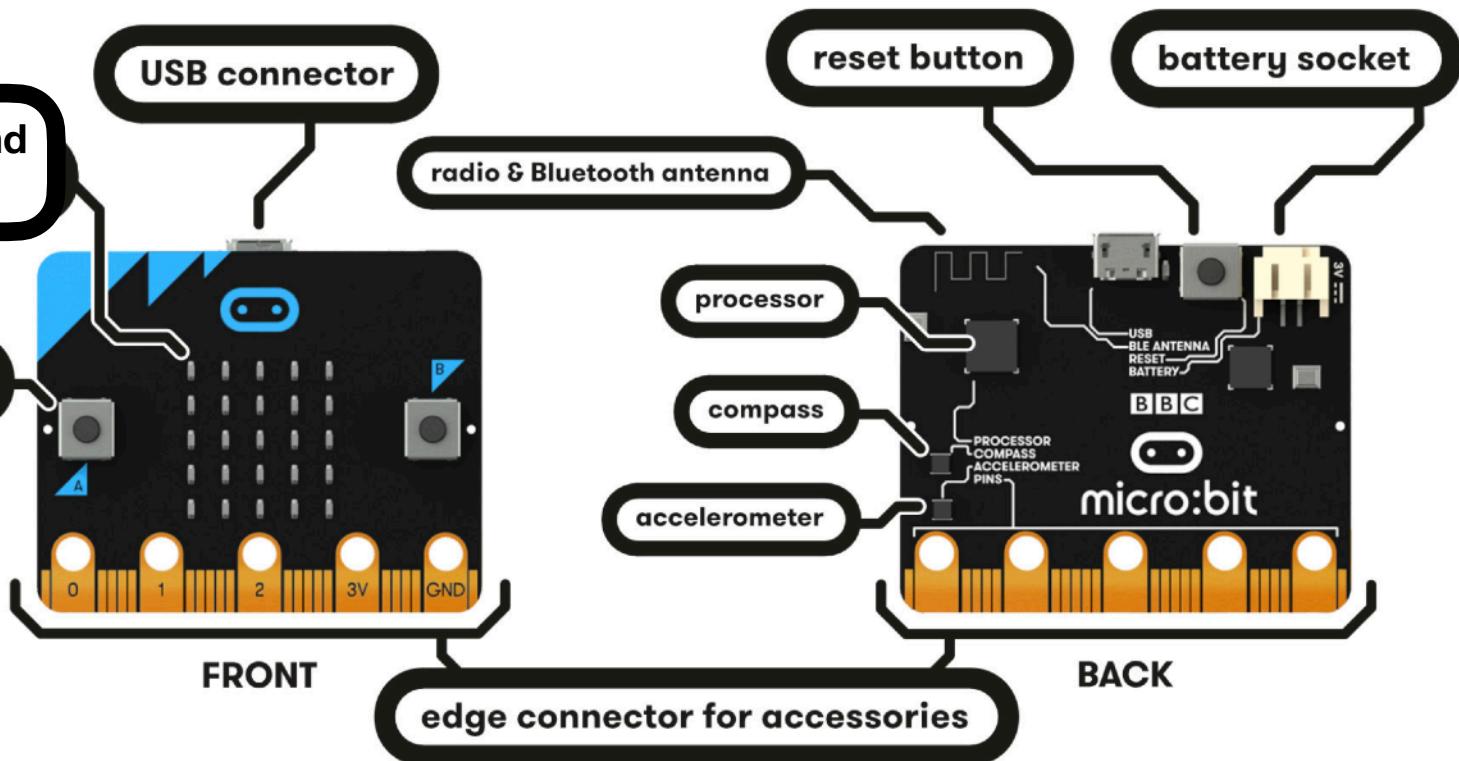
Adruino



Raspberry Pi



Micro:bit



## Basic Structure & Components of Micro:bit

source: <https://microbit.org/guide/features>

Server: localhost:3306 » Database: abc » Table: logs

Browse Structure SQL Search Insert Export Import Privileges Operations Triggers

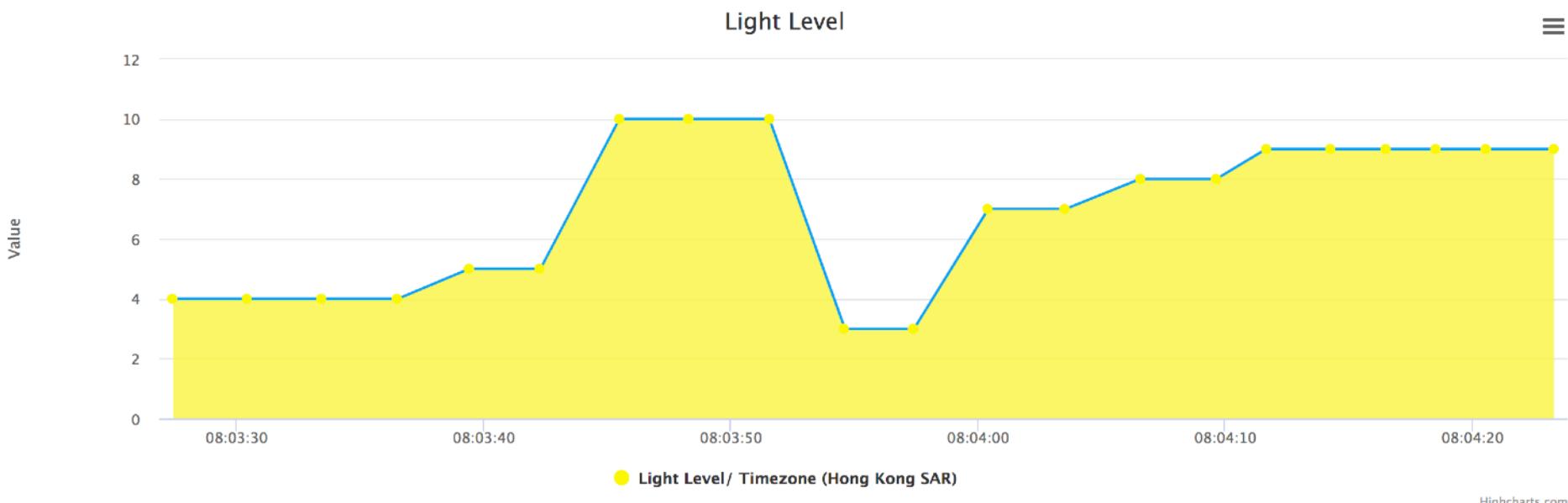
Number of rows: 25 Filter rows: Search this table Sort by key: None

+ Options

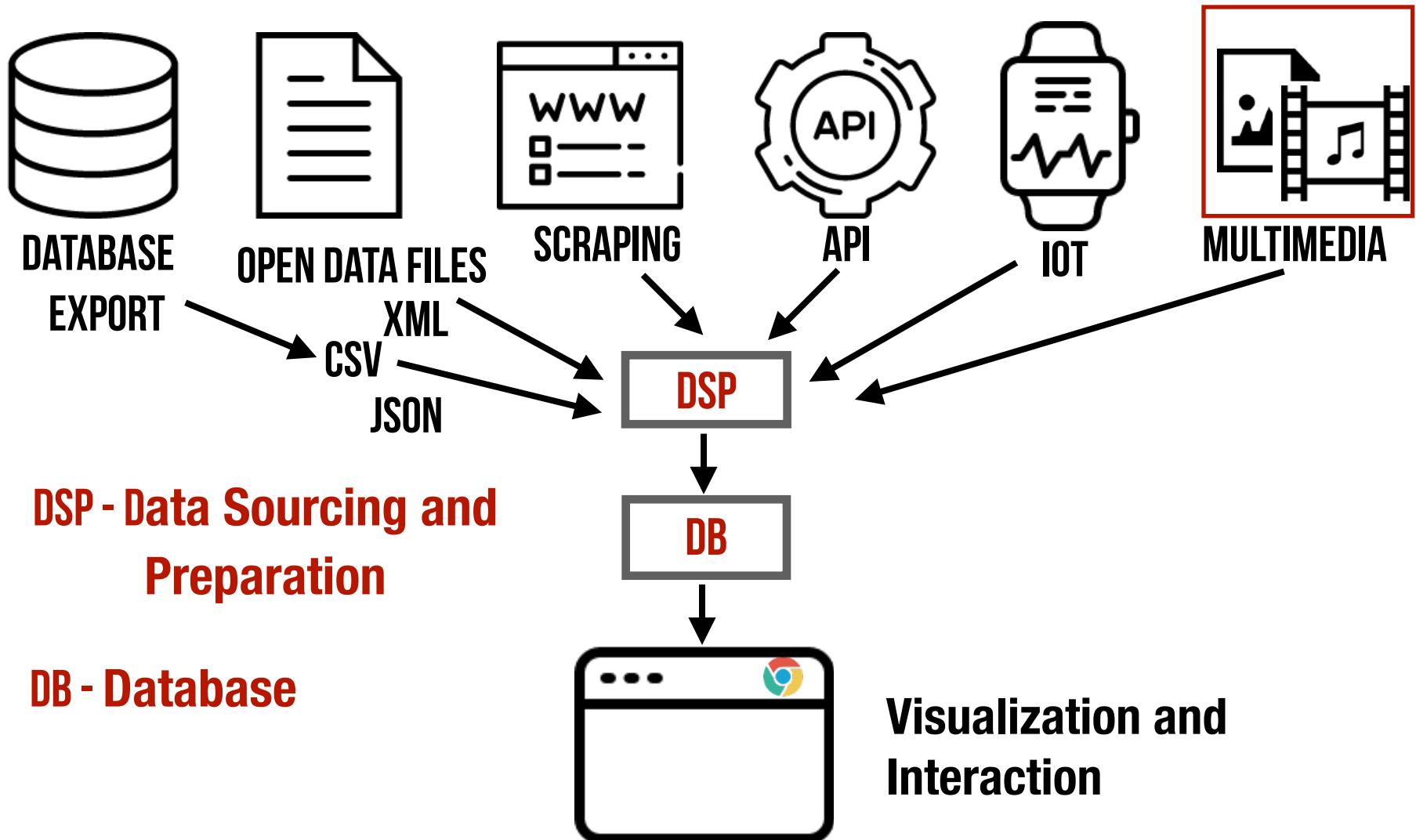
				logtID	lightLevel	timeStamp
<input type="checkbox"/>		<a href="#">Edit</a>		15181	4	2020-09-11 22:55:24
<input type="checkbox"/>		<a href="#">Edit</a>		15182	5	2020-09-11 22:55:30
<input type="checkbox"/>		<a href="#">Edit</a>		15183	4	2020-09-11 22:55:36
<input type="checkbox"/>		<a href="#">Edit</a>		15184	7	2020-09-11 22:55:42
<input type="checkbox"/>		<a href="#">Edit</a>		15185	5	2020-09-11 22:55:48
<input type="checkbox"/>		<a href="#">Edit</a>		15186	7	2020-09-11 22:55:54
<input type="checkbox"/>		<a href="#">Edit</a>		15187	7	2020-09-11 22:56:00
<input type="checkbox"/>		<a href="#">Edit</a>		15188	5	2020-09-11 22:56:06
<input type="checkbox"/>		<a href="#">Edit</a>		15189	9	2020-09-11 22:56:12
<input type="checkbox"/>		<a href="#">Edit</a>		15190	4	2020-09-12 19:50:56
<input type="checkbox"/>		<a href="#">Edit</a>		15191	9	2020-09-12 19:51:02
<input type="checkbox"/>		<a href="#">Edit</a>		15192	10	2020-09-12 19:51:08
<input type="checkbox"/>		<a href="#">Edit</a>		15193	3	2020-09-12 19:51:17
<input type="checkbox"/>		<a href="#">Edit</a>		15194	9	2020-09-12 19:51:23
<input type="checkbox"/>		<a href="#">Edit</a>		15195	9	2020-09-12 19:51:29

Check all With selected: Edit Copy Delete Export

# Light Level : 9

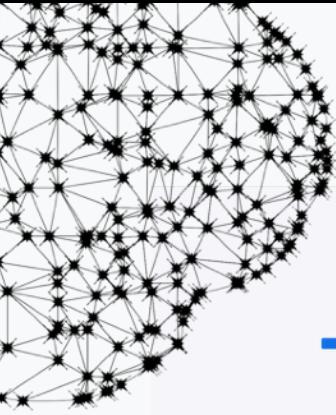


Thanks to [Soumil Nitin Shah](#) and [Peter Kazarinoff](#) for inspiring me to create this Python Flask data logger.



# **Image and Voice Recognition and Modelling**

## **Using Machine/Deep Learning**



# Teachable Machine

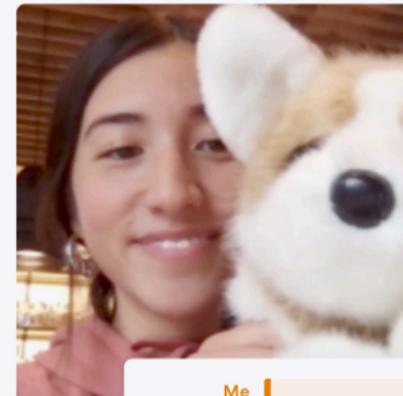
**Train a computer to recognize your own images, sounds, & poses.**

A fast, easy way to create machine learning models for your sites, apps, and more – no expertise or coding required.

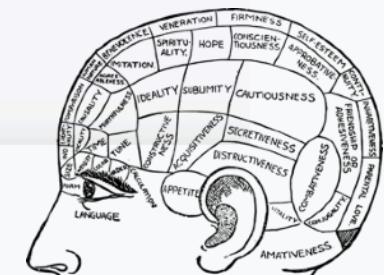
[Get Started](#)



<https://teachablemachine.withgoogle.com/>



Me  
Me + Dog <3  
98%



# ≡ Teachable Machine

## Mask

Add Image Samples:



## No-Mask

Add Image Samples:



⋮

## Training

Train Model

Advanced 

⋮

## Preview

 Export Model

You must train a model on the left before you can preview it here.

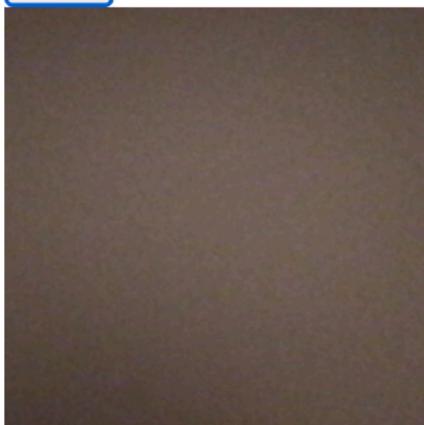
 Add a class

 English (U.S.) 

release-2-2-2 - 2.2.2#062770 - 83

Teachable Machine Image Model

Start



Mask: 0.36

No Mask: 0.64

# Teachable Machine

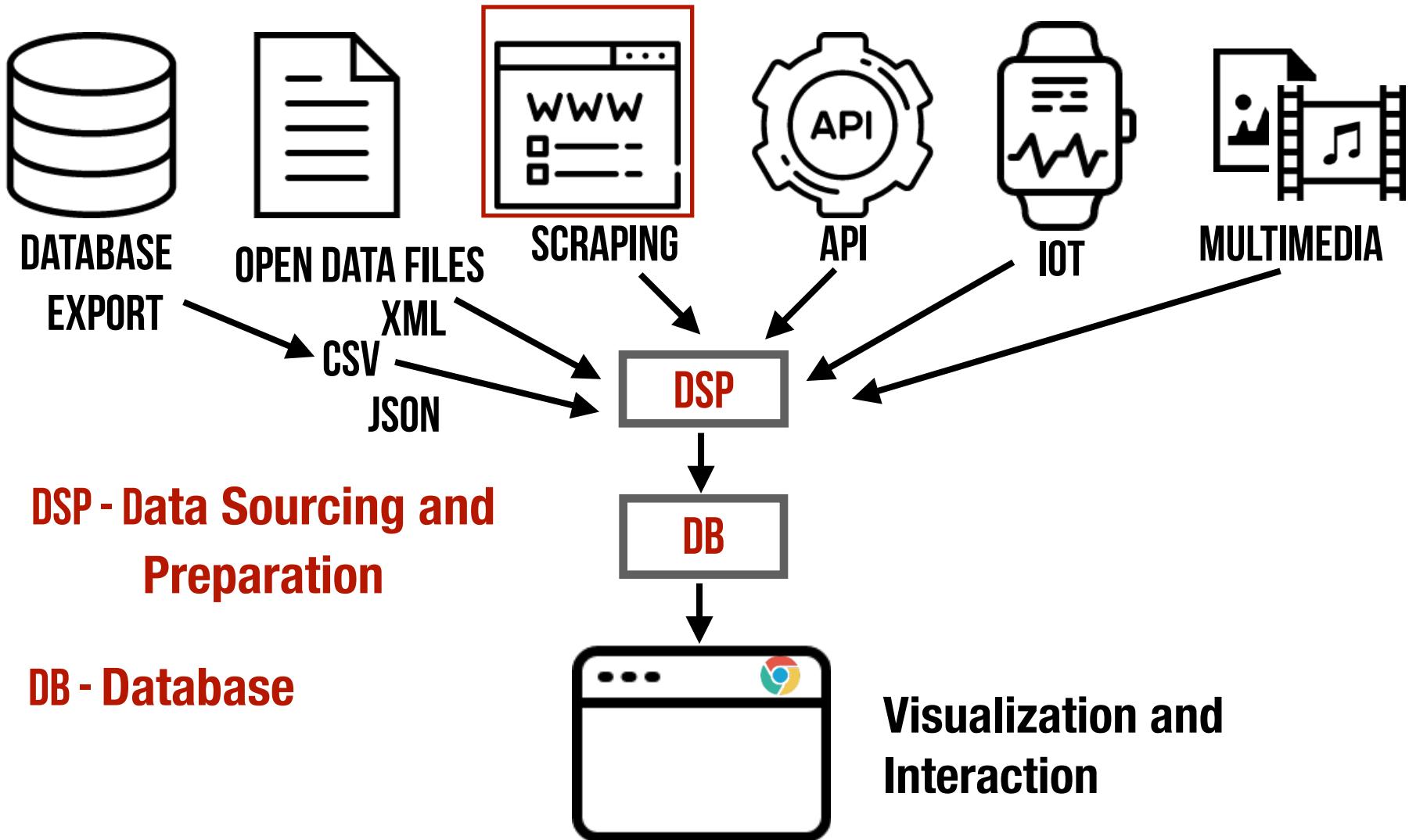
Teachable Machine is an online platform for ML beginners to experience Google's powerful TensorFlow deep learning tech. You can use it to recognize:

- Images
- Audios
- Poses

The simple user interface allows you to train the ML model to recognize images, audios, and poses with ease.

The result can be downloaded for storage or stored online.

You can easily embed the trained model into web pages, with just a few lines of JavaScript codes. The result can be integrated with HTML/CSS codes for more engaging user experience.



# **What is web scraping?**

**Web scraping is a process of fetching web pages from a website and extracting specific data from it.**

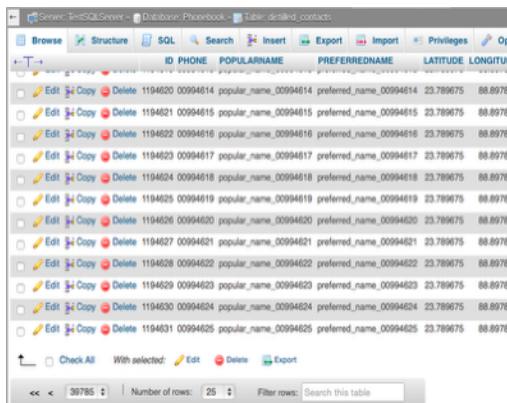
- ✓ Scrap stock prices
- ✓ Scrap product information
- ✓ Scrap race scores
- ✓ Scrap market trend data
- ✓ Scrap demographics and census data
- ✓ Social media and web articles, etc.



# **How does it work?**

# Data

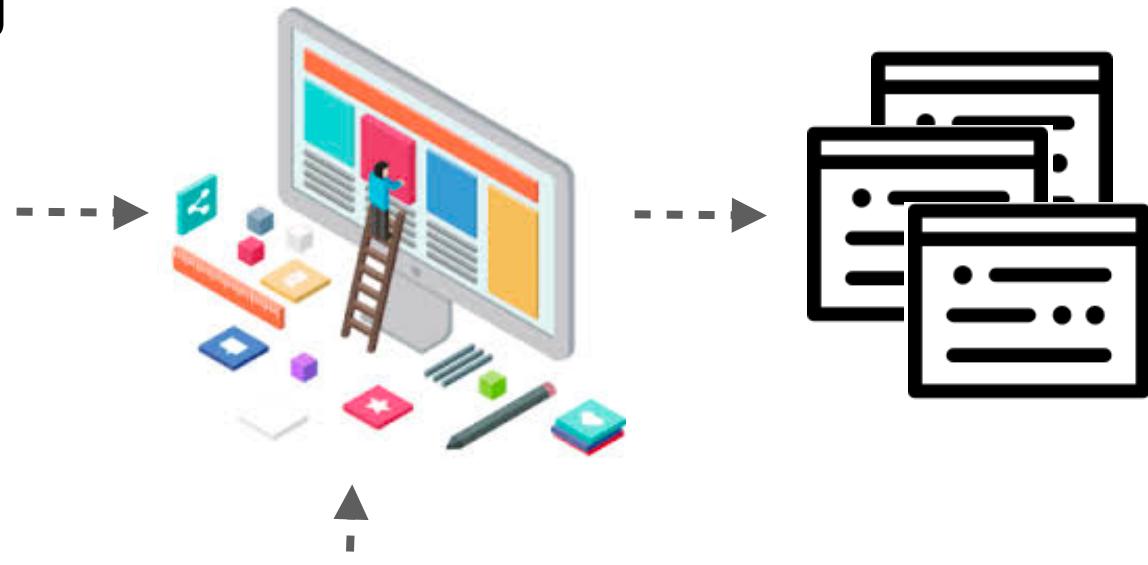
## Dynamic Web Building



ID	PHONE	POPULARNAME	PREFERREDNAME	LATITUDE	LONGITUDE
1194620	00994614	popular_name_00994614	preferred_name_00994614	23.789675	88.897665
1194621	00994615	popular_name_00994615	preferred_name_00994615	23.789675	88.897665
1194622	00994616	popular_name_00994616	preferred_name_00994616	23.789675	88.897665
1194623	00994617	popular_name_00994617	preferred_name_00994617	23.789675	88.897665
1194624	00994618	popular_name_00994618	preferred_name_00994618	23.789675	88.897665
1194625	00994619	popular_name_00994619	preferred_name_00994619	23.789675	88.897665
1194626	00994620	popular_name_00994620	preferred_name_00994620	23.789675	88.897665
1194627	00994621	popular_name_00994621	preferred_name_00994621	23.789675	88.897665
1194628	00994622	popular_name_00994622	preferred_name_00994622	23.789675	88.897665
1194629	00994623	popular_name_00994623	preferred_name_00994623	23.789675	88.897665
1194630	00994624	popular_name_00994624	preferred_name_00994624	23.789675	88.897665
1194631	00994625	popular_name_00994625	preferred_name_00994625	23.789675	88.897665

# Web Publishing

# Web



## Static Web Building

# Data

# Web Scraping

# Web



# **Scraping in Practice**

# Why scraping?

1. In the absence of open data, API and other means of sourcing data, scraping is a viable alternative.
2. Useful for collecting market data to conduct competitive analysis (e.g. price comparison)
3. As part of a data understanding exercise, gather data from multiple sources, including scraping, can help to see the big picture.
4. In exploratory data investigation, the more data the better for discovering patterns
5. Machine learning requires enormous amount of training and testing data for building predictive models.



## A web scraping tool that is easy to use

ParseHub is a free web scraping tool. With our advanced web scraper, extracting data is as easy as clicking the data you need.

Download our free app



<https://www.parsehub.com/>



Open a website

Download our [desktop app](#). Choose a site to scrape data from.



Click to select data

Get data from multiple pages. Interact with AJAX, forms, dropdowns, etc.



Download results

Access data via JSON, Excel and [API](#). Data is collected by our servers.



 Projects

 Runs

 My Account

 Integrations

 Plans & Billing

 Tutorials

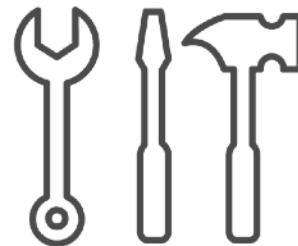
 Documentation

 API

 Contact

 Log Out

## Recent projects



 + New Project

● covid19 hkej.com Project  

● travelchinaguide.com Project  

● startupbeat.hkej.com Project  

basketball-reference.com P...  

 + See more

## Recent runs

## Interactive Tutorials

 Learn the Basics (8 min)

 Select and Download Data (3 min)

 Group data with Relative Select (3 min)

 Click and Navigate to Links (4 min)

## Written Tutorials

 Parsehub 101

 Pagination ('next' page buttons)

 Scrape product details

 Scrape leads from directories



WIKIPEDIA  
The Free Encyclopedia

Article Talk

Not logged in Talk Contributions Create account Log in

Read

Edit

View history

Search Wikipedia

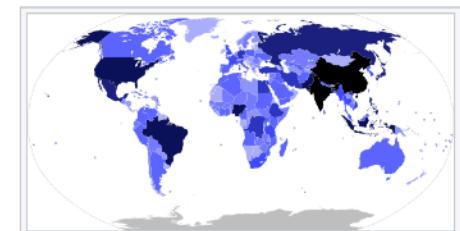


# List of countries and dependencies by population

From Wikipedia, the free encyclopedia

This is a **list of countries and dependencies by population**. It includes **sovereign states**, inhabited **dependent territories** and, in some cases, **constituent countries** of sovereign states, with inclusion within the list being primarily based on the ISO standard **ISO 3166-1**. For instance, the **United Kingdom** is considered as a single entity, while the constituent countries of the **Kingdom of the Netherlands** are considered separately. In addition, this list includes certain **states with limited recognition** not found in ISO 3166-1.

Also given in percent is each country's population compared with the **world population**, which the **United Nations** estimates at 7.82 billion as of today.



Map of countries and territories by population in 2019



A cartogram of the world population in 2018

## Contents [hide]

- 1 Method
- 2 Sovereign states and dependencies by population
- 3 See also
  - 3.1 Lists of countries by population
    - 3.1.1 Continental

[https://en.wikipedia.org/wiki/List\\_of\\_countries\\_and\\_dependencies\\_by\\_population](https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population)

## Sovereign states and dependencies by population [\[ edit \]](#)

Note: A number rank is assigned to the 193 member states of the United Nations and the two observer states to the United Nations General Assembly. Dependent territories and constituent countries that are parts of sovereign states are shown in *italics* and not assigned a numbered rank. In addition, sovereign states with limited recognition are not assigned a number rank.

Rank	Country (or dependent territory)	Population	% of world	Date	Source (official or UN)
1	China <sup>[b]</sup>	1,405,239,080	18.0%	9 Nov 2020	National population clock <sup>[3]</sup>
2	India <sup>[c]</sup>	1,369,434,247	17.5%	9 Nov 2020	National population clock <sup>[4]</sup>
3	United States <sup>[d]</sup>	330,617,844	4.22%	9 Nov 2020	National population clock <sup>[5]</sup>
4	Indonesia	269,603,400	3.45%	1 Jul 2020	National annual projection <sup>[6]</sup>
5	Pakistan <sup>[e]</sup>	220,892,331	2.82%	1 Jul 2020	UN Projection <sup>[2]</sup>
6	Brazil	212,311,986	2.71%	9 Nov 2020	National population clock <sup>[7]</sup>
7	Nigeria	206,139,587	2.63%	1 Jul 2020	UN Projection <sup>[2]</sup>
8	Bangladesh	160,615,066	2.17%	9 Nov 2020	National population

◀ bernard@intechni...

Projects

Runs

My Account

Integrations

Plans & Billing

Tutorials

Documentation

API

Contact

Log Out

## My Projects

+ New Project

Import Project

Export All

Click 'New Project' to start.

catalog.hkstore.com demo 1



basketball-reference.com Proj...



OpenRice



startupbeat.hkj.com Project



## Untitled Project

Enter a website you'd like to extract data from

**Start project on this URL**

Enter URL and click  
'Start project on this  
URL'.

The screenshot shows the ParseHub landing page. At the top, there's a search bar with the URL "countries\_and\_dependencies\_by\_population" and a green button below it labeled "Start project on this URL". A red box highlights this button and the URL input field. To the right, the ParseHub logo and the URL "https://www.parsehub.com/landing" are visible. The main content area starts with a "Welcome!" message. Below it, there are two sections: "Beginner Tutorials" and "Advanced Tutorials", each containing five numbered steps. The "Beginner Tutorials" section includes steps: 1. Start Here: Create your first project; 2. Extract text from a web page; 3. Extract data from many pages (pagination); 4. Run project & download Excel & JSON data; and 5. Use the REST API. The "Advanced Tutorials" section includes steps: 1. Collect data on a schedule; 2. Enter text into a search box; 3. Get data from behind a log-in; 4. Infinite scrolling pages; and 5. Enter URLs for ParseHub to crawl. At the bottom, there's a preview area with tabs for "CSV/Excel", "JSON", and "CSV/Excel Wide (beta)", and a message stating "A preview of your data will appear here". There are also checkboxes for "Show more data" and "Visuals enabled (advanced)". A blue circular icon with a white message icon is in the bottom right corner.

Welcome!

**Beginner Tutorials**

1. Start Here: Create your first project
2. Extract text from a web page
3. Extract data from many pages (pagination)
4. Run project & download Excel & JSON data
5. Use the REST API

**Advanced Tutorials**

1. Collect data on a schedule
2. Enter text into a search box
3. Get data from behind a log-in
4. Infinite scrolling pages
5. Enter URLs for ParseHub to crawl

CSV/Excel   JSON   CSV/Excel Wide (beta)

A preview of your data will appear here

Show more data ?    Visuals enabled (advanced) ?

API

Tutorials

Contact

Select page

Select selection1 (202)

Extract name

Extract url

Selection Node: 

All td's  
> All a's

Wait up to 60 seconds for elements to appear.

W List of countries and dependencies

https://en.wikipedia.org/wiki/List\_of\_countries\_and\_dependencies\_by\_population

	Rank	Country (or dependent territory)	Population	% of world	Date	Source (official or UN)
1	China[b]	 China[b]			2020	National population clock[3]
2	India[c]	 India[c]			2020	National population clock[4]
3	United States[d]	 United States[d]	330,544,466	4.23%	27 Oct 2020	National population clock[5]
4	Indonesia	 Indonesia	269,603,400	3.45%	1 Jul 2020	National annual

Click first target as anchor column.

CSV/Excel JSON CSV/Excel Wide (beta)

selection1_name	selection1_url
China	<a href="https://en.wikipedia.org/wiki/Demographics_of_China">https://en.wikipedia.org/wiki/Demographics_of_China</a>
India	<a href="https://en.wikipedia.org/wiki/Demographics_of_India">https://en.wikipedia.org/wiki/Demographics_of_India</a>
United States	<a href="https://en.wikipedia.org/wiki/Demographics_of_United_States">https://en.wikipedia.org/wiki/Demographics_of_United_States</a>
Indonesia	<a href="https://en.wikipedia.org/wiki/Demographics_of_Indonesia">https://en.wikipedia.org/wiki/Demographics_of_Indonesia</a>
Pakistan	<a href="https://en.wikipedia.org/wiki/Demographics_of_Pakistan">https://en.wikipedia.org/wiki/Demographics_of_Pakistan</a>
Brazil	<a href="https://en.wikipedia.org/wiki/Demographics_of_Brazil">https://en.wikipedia.org/wiki/Demographics_of_Brazil</a>
Nigeria	<a href="https://en.wikipedia.org/wiki/Demographics_of_Nigeria">https://en.wikipedia.org/wiki/Demographics_of_Nigeria</a>
Bangladesh	<a href="https://en.wikipedia.org/wiki/Demographics_of_Bangladesh">https://en.wikipedia.org/wiki/Demographics_of_Bangladesh</a>

This is a live preview. When you are ready to run your project, click Get Data.



en.wikip...

BROWSE

Select page

Select country (202) +

Extract name

Extract url

Get Data

main\_template

Enter 'Select' name and 'Extract' column name as you wish.

Wait up to 60 seconds for elements to appear.

W List of countries and depende X +

https://en.wikipedia.org/wiki/List\_of\_countries\_and\_dependencies\_by\_population

	Rank	Country (or dependent territory)	Population	% of world	Date	Source (official or UN)
1	China <sup>b</sup>	1,405,050,320	18.0%	27 Oct 2020	National population clock <sup>[3]</sup>	
2	India <sup>c</sup>	1,368,910,880	17.5%	27 Oct 2020	National population clock <sup>[4]</sup>	
3	United States <sup>d</sup>	330,544,466	4.23%	27 Oct 2020	National population clock <sup>[5]</sup>	
4	Indonesia	269,603,400	3.45%	1 Jul 2020	National annual	

CSV/Excel JSON CSV/Excel Wide (beta)

country_name	country_url
China	https://en.wikipedia.org/wiki/Demographics_of_China
India	https://en.wikipedia.org/wiki/Demographics_of_India
United States	https://en.wikipedia.org/wiki/Demographics_of_United_States
Indonesia	https://en.wikipedia.org/wiki/Demographics_of_Indonesia
Pakistan	https://en.wikipedia.org/wiki/Demographics_of_Pakistan
Brazil	https://en.wikipedia.org/wiki/Demographics_of_Brazil
Nigeria	https://en.wikipedia.org/wiki/Demographics_of_Nigeria
Bangladesh	https://en.wikipedia.org/wiki/Demographics_of_Bangladesh

This a live preview. When you are ready to run your project, click Get Data.

Show more data ?  Visuals enabled (advanced) ?

API Tutorials Contact

main\_template

en.wikip...

BROWSE

Attach already-selected elements to related elements (auto-extracts data if possible). Good for grouping different fields together in your data.

Relative Select

Click

Advanced

Select page

Select country (202)

Extract name

Extract url

Get Data

Use 'Relative Select' to select remaining columns with 1st column as the anchor.

W List of countries and dependencies

Relative Select

country\_name

China

India

United States

Indonesia

Pakistan

Brazil

Nigeria

Bangladesh

Source (official or UN)

Date

of world

18.0%

27 Oct 2020

National population clock<sup>[3]</sup>

17.5%

27 Oct 2020

National population clock<sup>[4]</sup>

4.23%

27 Oct 2020

National population clock<sup>[5]</sup>

3.45%

1 Jul 2020

National annual

This is a live preview. When you are ready to run your project, click Get Data.

Show more data

Visuals enabled (advanced)

API Tutorials Contact

main\_template

country_name	Source	Date	Value
China	National population clock <sup>[3]</sup>	27 Oct 2020	18.0%
India	National population clock <sup>[4]</sup>	27 Oct 2020	17.5%
United States	National population clock <sup>[5]</sup>	27 Oct 2020	4.23%
Indonesia	National annual	1 Jul 2020	3.45%
Pakistan			
Brazil			
Nigeria			
Bangladesh			

en.wikip...

BROWSE

Select page +

Select country +

Extract name

Extract url

Relative selection1 (1) +

> End to

Wait up to 60 seconds for elements to appear.

main\_template

API Tutorials Contact

List of countries and dependencies

https://en.wikipedia.org/wiki/List\_of\_countries\_and\_dependencies\_by\_population

	Rank	Country (or dependent territory)	Population	% of world	Date	Source (official or UN)
	1	China <sup>b</sup>	1,405,050,320	18.0%	27 Oct 2020	National population clock <sup>[3]</sup>
	2	India <sup>c</sup>	1,368,910,880	17.5%	27 Oct 2020	National population clock <sup>[4]</sup>
	3	United States <sup>d</sup>	330,544,466	4.23%	27 Oct 2020	National population clock <sup>[5]</sup>
	4	Indonesia	269,603,400	3.45%	1 Jul 2020	National annual

CSV/Excel JSON CSV/Excel Wide (beta)

country_name	country_url	country_selection1
China	https://en.wikipedia.org/wiki/Demographics_of_China	1,405,050,320
India	https://en.wikipedia.org/wiki/Demographics_of_India	
United States	https://en.wikipedia.org/wiki/Demographics_of_United_States	
Indonesia	https://en.wikipedia.org/wiki/Demographics_of_Indonesia	
Pakistan	https://en.wikipedia.org/wiki/Demographics_of_Pakistan	

This is a live preview. When you are ready to run your project, click Get Data.

Show more data  Visuals enabled (advanced)

en.wikip...

BROWSE

main\_template

Select page +  
Select country +  
Extract name  
Extract url  
Relative population (202) ↕ - +  
Get Data

W List of countries and dependencies

https://en.wikipedia.org/wiki/List\_of\_countries\_and\_dependencies\_by\_population

	Country (or dependent territory)	Rank	Population	% of world	Date	Source Select Mode (official or UN)
1	China <sup>b</sup>	1	1,405,050,320	18.0%	27 Oct 2020	National population clock <sup>[3]</sup>
2	India <sup>c</sup>	2	1,368,910,880	17.5%	27 Oct 2020	National population clock <sup>[4]</sup>
3	United States <sup>d</sup>	3	330,544,466	4.23%	27 Oct 2020	National population clock <sup>[5]</sup>
4	Indonesia	4	269,603,400	3.45%	1 Jul 2020	National annual

Rename the column name as you wish.  
Make sure no blank space and odd symbols.

CSV/Excel JSON CSV/Excel Wide (beta)

country_name	country_url	country_population
China	https://en.wikipedia.org/wiki/Demographics_of_China	1,405,050,320
India	https://en.wikipedia.org/wiki/Demographics_of_India	1,368,910,880
United States	https://en.wikipedia.org/wiki/Demographics_of_United_States	330,544,466
Indonesia	https://en.wikipedia.org/wiki/Demographics_of_Indonesia	269,603,400
Pakistan	https://en.wikipedia.org/wiki/Demographics_of_Pakistan	220,892,331

This is a live preview. When you are ready to run your project, click Get Data.



en.wikip...

BROWSE

Select page

Select country

Extract name

Extract url

Relative population

Relative percent\_of\_world

Relative date (202)

Get Data

When done defining the columns, click 'Get Data'.

main\_template

List of countries and dependencies by population

Country

Rank (or dependent territory) Population % of world Date

Source: Select Mode (official or UN)

	Rank (or dependent territory)	Population	% of world	Date	Source: Select Mode (official or UN)
1	China <sup>[b]</sup>	1,405,050,320	18.0%	27 Oct 2020	National population clock <sup>[3]</sup>
2	India <sup>[c]</sup>	1,368,910,880	17.5%	27 Oct 2020	National population clock <sup>[4]</sup>
3	United States <sup>[d]</sup>	330,544,466	4.23%	27 Oct 2020	National population clock <sup>[5]</sup>
4	Indonesia	269,603,400	3.45%	1 Jul 2020	National annual

CSV/Excel JSON CSV/Excel Wide (beta)

country_name	country_url	country_population	country_percent_of...	country_date
China	https://en.wikipedia.org/wiki/Demographics_of_China	1,405,050,320	18.0%	27 Oct 2020
India	https://en.wikipedia.org/wiki/Demographics_of_India	1,368,910,880	17.5%	27 Oct 2020
United States	https://en.wikipedia.org/wiki/Demographics_of_the_United_States	330,544,466	4.23%	27 Oct 2020
Indonesia	https://en.wikipedia.org/wiki/Demographics_of_Indonesia	269,603,400	3.45%	1 Jul 2020
Pakistan	https://en.wikipedia.org/wiki/Demographics_of_Pakistan	220,892,331	2.82%	1 Jul 2020

This is a live preview. When you are ready to run your project, click Get Data.

Show more data ?  Visuals enabled (advanced) ?



← ⌂ en.wikipedia.org Project

Run your project on ParseHub's  
servers, once

Test Run



Run



Schedule

Edit project

Previous Runs

Click 'Run' to start scraping.

Your results will appear here after you've run your project.



 Edit project

Data is being collected. Please wait.

Starting up...

Refreshing status in **2** second(s). [Refresh now](#)

Download Data

CSV/Excel



JSON

API

 Cancel Run

All dates and times are in UTC +0000.

Empty file with no results? [Click here](#) to fix.

CSV file too big? Save the JSON file and [click here to convert to CSV](#).

Run Details

Settings

Started job



 Edit project

When done scraping, download the data as CSV file and import into a spreadsheet or SQL database.

Your data is ready! Click on the green buttons to download.

#### Download Data

CSV/Excel

JSON

API

Template Name  
main\_template

Pages Scrapped  
1 

All dates and times are in UTC +0000.

Empty file with no results? [Click here](#) to fix.

CSV file too big? Save the JSON file and [click here](#) to convert to CSV.

#### Run Details

Status	complete
Pages	1 collected
Initialized	2020-10-28T14:43:31

#### Settings

URL

[https://en.wikipedia.org  
/wiki/List\\_of\\_countries\\_and\\_dependencies\\_by\\_population](https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population)



country_name	country_url	country_population	country_percent_of_world	country
China	<a href="https://en.wikipedia.org/wiki/Demographics_of_China">https://en.wikipedia.org/wiki/Demographics_of_China</a>	1,405,050,320	18.0%	27 Oct 2
India	<a href="https://en.wikipedia.org/wiki/Demographics_of_India">https://en.wikipedia.org/wiki/Demographics_of_India</a>	1,368,910,880	17.5%	27 Oct 2
United States	<a href="https://en.wikipedia.org/wiki/Demographics_of_United_States">https://en.wikipedia.org/wiki/Demographics_of_United_States</a>	330,544,466	4.23%	27 Oct 2
Indonesia	<a href="https://en.wikipedia.org/wiki/Demographics_of_Indonesia">https://en.wikipedia.org/wiki/Demographics_of_Indonesia</a>	269,603,400	3.45%	1 Jul 202
Pakistan	<a href="https://en.wikipedia.org/wiki/Demographics_of_Pakistan">https://en.wikipedia.org/wiki/Demographics_of_Pakistan</a>	220,892,331	2.82%	1 Jul 202
Brazil	<a href="https://en.wikipedia.org/wiki/Demographics_of_Brazil">https://en.wikipedia.org/wiki/Demographics_of_Brazil</a>	212,254,617	2.71%	27 Oct 2
Nigeria	<a href="https://en.wikipedia.org/wiki/Demographics_of_Nigeria">https://en.wikipedia.org/wiki/Demographics_of_Nigeria</a>	206,139,587	2.64%	1 Jul 202
Bangladesh	<a href="https://en.wikipedia.org/wiki/Demographics_of_Bangladesh">https://en.wikipedia.org/wiki/Demographics_of_Bangladesh</a>	169,542,128	2.17%	27 Oct 2
Russia	<a href="https://en.wikipedia.org/wiki/Demographics_of_Russia">https://en.wikipedia.org/wiki/Demographics_of_Russia</a>	146,748,590	1.88%	1 Jan 202
Mexico	<a href="https://en.wikipedia.org/wiki/Demographics_of_Mexico">https://en.wikipedia.org/wiki/Demographics_of_Mexico</a>	127,792,286	1.63%	1 Jul 202
Japan	<a href="https://en.wikipedia.org/wiki/Demographics_of_Japan">https://en.wikipedia.org/wiki/Demographics_of_Japan</a>	125,880,000	1.61%	1 Oct 20
Philippines	<a href="https://en.wikipedia.org/wiki/Demographics_of_Philippines">https://en.wikipedia.org/wiki/Demographics_of_Philippines</a>	109,349,790	1.40%	27 Oct 2
DR Congo	<a href="https://en.wikipedia.org/wiki/Demographics_of_Democratic_Republic_of_the_Congo">https://en.wikipedia.org/wiki/Demographics_of_Democratic_Republic_of_the_Congo</a>	101,935,800	1.30%	1 Jul 202
Egypt	<a href="https://en.wikipedia.org/wiki/Demographics_of_Egypt">https://en.wikipedia.org/wiki/Demographics_of_Egypt</a>	101,100,567	1.29%	27 Oct 2
Ethiopia	<a href="https://en.wikipedia.org/wiki/Demographics_of_Ethiopia">https://en.wikipedia.org/wiki/Demographics_of_Ethiopia</a>	100,829,000	1.29%	1 Jul 202
Vietnam	<a href="https://en.wikipedia.org/wiki/Demographics_of_Vietnam">https://en.wikipedia.org/wiki/Demographics_of_Vietnam</a>	96,483,981	1.23%	1 Jul 202
Iran	<a href="https://en.wikipedia.org/wiki/Demographics_of_Iran">https://en.wikipedia.org/wiki/Demographics_of_Iran</a>	83,898,529	1.07%	27 Oct 2
Turkey	<a href="https://en.wikipedia.org/wiki/Demographics_of_Turkey">https://en.wikipedia.org/wiki/Demographics_of_Turkey</a>	83,154,997	1.06%	31 Dec 2
Germany	<a href="https://en.wikipedia.org/wiki/Demographics_of_Germany">https://en.wikipedia.org/wiki/Demographics_of_Germany</a>	83,122,889	1.06%	30 Jun 2
France	<a href="https://en.wikipedia.org/wiki/Demographics_of_France">https://en.wikipedia.org/wiki/Demographics_of_France</a>	67,132,000	0.858%	1 Sep 20

**Scraping multiple pages.**

◀ bernard@intechni...

Projects

Runs

My Account

Integrations

Plans & Billing

Tutorials

Documentation

API

Contact

Log Out

## My Projects

+ New Project

Import Project

Export All

Search Projects

catalog.hkstore.com demo 1



basketball-reference.com Proj...



OpenRice



startupbeat.hkj.com Project



## Untitled Project

Enter a website you'd like to extract data from

https://catalog.hkstore.com/catalog\_tc\_hk/m

Start project on this URL

Enter URL and click  
'Start project on this  
URL'.

男裝

https://catalog.hkstore.com/catalog\_tc\_hk/m

註冊成為新會員即送\$100網店優惠券。 [立即註冊](#)

CATALOG MARATHON SPORTS

最新商品 品牌 聯乘系列 男裝 女裝 童裝 電子禮品卡 SALE

1-24 / 2507件 排序 最新上架

分類 +

熱賣系列 +

商品分類 +

鞋款 +

服裝類別 +

CSV/Excel JSON CSV/Excel Wide (beta)

This a live preview. When you are ready to run your project, click Get Data.

Show more data Visuals enabled (advanced)

catalog.... BROWSE

Select page +

Select & Extract selection1 (1) +

Get Data

Selection Node: Edit

All elements with class product-item-brand  
> All elements

Wait up to 60 seconds for elements to appear.

男裝 https://catalog.hkstorecom/catalog\_tc\_hk/men

註冊成為新會員即送\$100網店優惠券。立即註冊 Select Mode

CATALOG MARATHON SPORTS 馬拉松 搜索

最新商品 品牌 聯乘系列 男裝 女裝 童裝 電子禮品卡 SALE

配件類別 + 袋款 + 品牌 + 性別 + 活動類型 + 功能 + 顏色 + 價格 +

SPAN Adidas Stan Smith 中性球鞋 \$899

Click first target as anchor column.

Adidas Puffylette 中性無鞋帶 \$599

加入購物車 加入喜愛清單

CSV/Excel JSON CSV/Excel Wide (beta)

selection1

Adidas

This is a live preview. When you are ready to run your project, click Get Data.

Show more data Visuals enabled (advanced)

catalog....

BROWSE

Select page +

Select selection1 (24) - +

Extract name

Get Data

Selection Node: [Edit](#)

All elements with class `product-item-brand`  
> All elements s

Wait up to 60 seconds for elements to appear.

男裝

https://catalog.hkstore.com/catalog\_to\_hk/men

訂單淨值滿\$400即享免費門市取貨

Select Mode

CATALOG MARATHON SPORTS 馬拉松

最新商品 品牌 聯乘系列 男裝 女裝 童裝 電子禮品卡 SALE

搜索

鞋款 +

服裝類別 +

配件類別 +

袋款 +

品牌 +

性別 +

Adidas Stan Smith

Retropy E5

Puffylette

活動類型 + 中性球鞋

男裝球鞋

功能 +

顏色 + DIV

Make sure the first columns are properly defined with all the targets bounded by the green borders.

CSV/Excel JSON CSV/Excel Wide (beta)

selection1\_name

Adidas

Adidas

Adidas

PUMA

This a live preview. When you are ready to run your project, click Get Data.

Show more data ?  Visuals enabled (advanced) ?

catalog....

男裝

https://catalog.hkstore

Select Mode

Select page

Select selection1(24)

Extract name

Get Data

Relative Select

Click

Advanced

... main\_template

鞋款

Always use the first column as the anchor to ‘Relative Select’ each new column.

功能

顏色

CSV/Excel JSON CSV/Excel Wide (beta)

selection1\_name

Adidas

Adidas

Adidas

PUMA

This a live preview. When you are ready to run your project, click Get Data.

Show more data Visuals enabled (advanced)

The screenshot shows a web browser window with a scraping tool overlay. The tool has a sidebar on the left with sections for 'main\_template', 'catalog....', 'Select page', 'Select selection1(24)', 'Extract name', and a 'Get Data' button. A tooltip for 'Relative Select' is highlighted with a red box. The main area displays a shoe catalog with two columns of shoes. The first column is selected with orange boxes around its header and body. The second column is partially visible. Below the catalog, there's a table with rows for '功能' and '顏色'. At the bottom, there are CSV/Excel export options and a data preview table for 'selection1\_name' with rows for 'Adidas' and 'PUMA'. A message at the bottom says 'This a live preview. When you are ready to run your project, click Get Data.' There are also checkboxes for 'Show more data' and 'Visuals enabled (advanced)'.

訂單淨值滿\$400即享免費門市取貨

Select Mode

CATALOG MARATHON SPORTS 馬拉松

最新商品 品牌 聯乘系列 男裝 女裝 童裝 電子禮品卡 SALE

搜索

會員登入 | 註冊

Always use the first column as the anchor to 'Relative Select' each new column.

選項	Adidas Stan Smith	Adidas Retropy E5	Adidas Puffylette
鞋款	Stan Smith	Retropy E5	Puffylette
服裝類別	中性球鞋	男裝球鞋	中性無鞋帶
	\$899	\$999	\$599

https://catalog.hkstore.com/catalog\_tc\_hk/men/adidas-108gx-8

CSV/Excel JSON CSV/Excel Wide (beta)

selection1_name	selection1_series	selection1_series_url
Adidas	Stan Smith	https://catalog.hkstore.com/catalog_tc_hk/men/adidas-108gx-8
Adidas	Retropy E5	https://catalog.hkstore.com/catalog_tc_hk/men/adidas-108hq6460
Adidas	Puffylette	https://catalog.hkstore.com/catalog_tc_hk/men/adidas-108hr1481

This is a live preview. When you are ready to run your project, click Get Data.

Show more data ?    Visuals enabled (advanced) ?

Select page



Select selection1



Extract name

Relative series (24)



Extract series

Delete Command

Extract series\_url



Get Data

Remove URL  
link if you don't  
need it.

男裝 +

[https://catalog.hkstore.com/catalog\\_tc\\_hk/men](https://catalog.hkstore.com/catalog_tc_hk/men)

註冊成為新會員即送\$100網店優惠券。 [立即註冊](#)

Select Mode

CATALOG MARATHON SPORTS

最新商品 品牌 聯乘系列 男裝 女裝 童裝 電子禮品卡 SALE



PUMA Scuff	Hoka Bondi 8	Keen Uneek SNK Sneaker
中性涼鞋	男裝登山鞋	男裝球鞋
\$259	\$1,380	\$1,090

chrome://phapp/content/views/index.html

CSV/Excel JSON CSV/Excel Wide (beta)

selection1_name	selection1_series	selection1_series_url
Adidas	Stan Smith	<a href="https://catalog.hkstore.com/catalog_tc_hk/men/adidas-108gx-8">https://catalog.hkstore.com/catalog_tc_hk/men/adidas-108gx-8</a>
Adidas	Retropy E5	<a href="https://catalog.hkstore.com/catalog_tc_hk/men/adidas-108hq6460">https://catalog.hkstore.com/catalog_tc_hk/men/adidas-108hq6460</a>
Adidas	Puffylette	<a href="https://catalog.hkstore.com/catalog_tc_hk/men/adidas-108hr1481">https://catalog.hkstore.com/catalog_tc_hk/men/adidas-108hr1481</a>

This a live preview. When you are ready to run your project, click Get Data.

 Show more data 
 Visuals enabled (advanced) 


catalog... BROWSE

Select page (1)

Select selection1

Extract name

Relative series

Extract series

Relative style

Extract style

Relative price

Extract price

New Select Command

Get Data

Targets one or more elements for a command (auto-extracts data if possible).

Select

Relative Select

Click

Advanced

男裝

訂單淨值滿\$400即享免費門市取貨

Select Mode

會員登入 | 註冊

分類

熱賣系列

商品分類

鞋款

服裝類別

男裝 女裝 童裝 電子禮品卡 SALE

Adidas Stan Smith

Adidas Retropy E5

Adidas Puffylette

Click the 'Select page' + button to create the next page.

CSV/Excel JSON CSV/Excel Wide (beta)

selection1_name	selection1_series	selection1_style	selection1_price
Adidas	Stan Smith	中性球鞋	\$899
Adidas	Retropy E5	男裝球鞋	\$999
Adidas	Puffylette	中性無鞋帶	\$599
PUMA	Scuff	中性涼鞋	\$259

This is a live preview. When you are ready to run your project, click Get Data.

Show more data ?  Visuals enabled (advanced) ?

API Tutorials Contact

catalog... BROWSE

Select Mode

... main\_template

Select page +  
Select selection1 +  
Extract name  
Relative series  
Extract series  
Relative style  
Extract style  
Relative price  
Extract price  
Select & Extract selection2(1) - +  
New Click Command  
Get Data

**Relative Select**  
Click Advanced ▾

Click on an already-selected element. Good for dropdowns, popups or navigating to new pages.  
常见问题

關於我們  
關於Catalog  
關於送貨  
關於退貨  
聯絡我們

追蹤我們  
免費訂閱Catalog電郵通訊，收取有關Catalog最新優惠及產品資訊。  
輸入電郵地址 提交

马拉松 Sports App

Marathon Sports App

男裝

https://catalog.hkstore.com/catalog\_tc\_hk/men

訂單淨值滿\$400即享免費門市取貨

CATALOG MARATHON SPORTS 馬拉松

最新商品 品牌 聯乘系列 男裝 女裝 童裝 電子禮品卡 SALE

1 2 3 4 5 >

Click the 'Click' button to start defining the next page.

selection1_name	selection1_series	selection1_style	selection1_price	selection2	selection2_url
Adidas	Stan Smith	中性球鞋	\$899	頁 下一頁	https://catalog.hkstore.com/catalog_tc_hk/men?p=2
Adidas	Retropy E5	男裝球鞋	\$999	頁 下一頁	https://catalog.hkstore.com/catalog_tc_hk/men?p=2
Adidas	Puffylette	中性無鞋帶	\$599	頁 下一頁	https://catalog.hkstore.com/catalog_tc_hk/men?p=2

This a live preview. When you are ready to run your project, click Get Data.

Show more data ⓘ  Visuals enabled (advanced) ⓘ

catalog... BROWSE

Select page  
Select selection1  
Extract name  
Relative series  
Extract series  
Relative style  
Extract style  
Relative price  
Extract price  
Select selection2  
Click each selection... (1) and go to main\_template

Get Data

Loads a new page    Uses AJAX

Go to Existing Template main\_template

Go to Another Project

Wait up to 5 seconds for page to load

Repeat the Current Template 0 more time(s). (0 = ∞)

Click setup

Is a next page button?

**Click 'Yes' to confirm.**

Examples of next page buttons

Examples of non next page buttons

Director: David Leitch  
Writers: Rhett Reese, Paul Wernick [more credit >](#)

UNLOCKED Apple iPhone 6 16GB/64GB/128GB with Warranty

iPhone 6 16GB 64GB 128GB

Buy It Now

Free Shipping

Free Returns

30-day Refund

13 new & refurbished from \$199.99

追蹤我們

免費訂閱Catalog電郵通訊，收取有關Catalog最新優惠及產品資訊。

輸入電子郵件地址 提交

Marathon Sports App

Marathon Sports App

selection1_name	selection1_series	selection1_style	selection1_price
Adidas	Stan Smith	中性球鞋	\$899
Adidas	Retropy E5	男裝球鞋	\$999
Adidas	Puffylette	中性無鞋帶	\$599
PUMA	Scuff	中性涼鞋	\$259

This is a live preview. When you are ready to run your project, click Get Data.

Show more data [?](#)  Visuals enabled (advanced) [?](#)

**Click 'Repeat Current Template' to confirm.**

This live preview. When you are ready to run your project, click Get Data.

Loads a new page    Uses AJAX

Go to Existing Template   main\_template ▾

Go to Another Project

Wait up to 5 seconds for page to load

Repeat the Current Template 0 more time(s). (0 = ∞)

API   Tutorials   Contact

CSV/Excel   JSON   CSV/Excel Wide (beta)

selection1_name	selection1_series	selection1_style	selection1_price
Adidas	Stan Smith	中性球鞋	\$899
Adidas	Retropy E5	男裝球鞋	\$999
Adidas	Puffylette	中性無鞋帶	\$599
PUMA	Scuff	中性涼鞋	\$259

Click setup

Is [ ] a next page button?

Yes   No

This click takes you to the next page of results. It will repeat the current template (pagination help).

Repeat the Current Template 0 more time(s). (0 = ∞)

Repeat Current Template

男裝

catalog... BROWSE

Select page

Select selection1

Extract name

Relative series

Extract series

Relative style

Extract style

Relative price

Extract price

Select selection2

Click each selection... (1)

and go to main\_template

Get Data

VIP會員及積分  
電子禮品卡  
員工註冊

購物指南

立即註冊

會員登入 | 註冊

童裝 電子禮品卡 SALE

1 2 3 4 5 >

Select Mode

追蹤我們

免費訂閱Catalog電郵通訊，收取有關Catalog最新優惠及產品資訊。

輸入電郵地址 提交

Marathon Sports App

Marathon Sports App

API   Tutorials   Contact

main\_template

catalog.... 

...  
Select page +  
Select selection1 +  
Extract name  
Relative series  
Extract series  
Relative style  
Extract style  
Relative price  
Extract price  
Select selection2  
**Click each selection... (1) **  
**and go to main\_template **

**Get Data**

**Next page defined.**

Go to Existing Template **main\_template** ▾  
 Go to Another Project

Wait up to **5** seconds for page to load

Repeat the Current Template **0** more time(s). (0 = ∞)

[API](#) [Tutorials](#) [Contact](#)

男裝  男裝  男裝 

https://catalog.hkstore.com/catalog\_tc\_hk/men?p=2

Click each s... Select Mode

訂單淨值滿\$400即享免費門市取貨

CATALOG MARATHON SPORTS 馬拉松

最新商品 品牌 聯乘系列 男裝 女裝 童裝 電子禮品卡 SALE

主頁 > 男裝

男裝 25-48 / 2507件 排序 最新上架

過濾

分類 +  
熱賣系列 +  
商品分類 +  
鞋款 +  
服裝類別 +

CSV/Excel JSON CSV/Excel Wide (beta)

selection1_name	selection1_series	selection1_style	selection1_price
New Balance	2002R	中性球鞋	\$1,199
New Balance	2002R	中性球鞋	\$1,199
New Balance	2002R	中性球鞋	\$1,199
PLASTIC THING X CATALOG系列	Plastic Thing 為食妹刺繡圖案外套	中性外套	\$699

This a live preview. When you are ready to run your project, click Get Data.

Show more data   
 Visuals enabled (advanced) 

## catalog...

Settings

Get Data

Select

Save

Sele

Undo

E Redo

F Export

Delete Project

I Exit

Logout

Help

Extract price

Select selection2

Click each selection... (1)



and go to main\_template

Get Data

Click the  
'Settings'  
option to  
define max #  
of pages to  
be scraped.

API

Tutorials

Contact

catalog...

男裝 男裝 +

https://catalog.hkstore.com/catalog\_to\_hk/men?p=2

Click each s... 訂單淨值滿\$400即享免費門市取貨 Select Mode

CATALOG MARATHON SPORTS HK

最新商品 品牌 聯乘系列 男裝 女裝 童裝 電子禮品卡 SALE

分類 +

熱賣系列 +

商品分類 +

鞋款 +

服裝類別 +

配件類別 +

袋款 +

品牌 +

性別 +

New Balance 2002R 中性球鞋

New Balance 2002R 中性球鞋

New Balance 2002R 中性球鞋

CSV/Excel JSON CSV/Excel Wide (beta)

selection1_name	selection1_series	selection1_style	selection1_price
New Balance	2002R	中性球鞋	\$1,199
New Balance	2002R	中性球鞋	\$1,199
New Balance	2002R	中性球鞋	\$1,199
PLASTIC THING X CATALOG系列	Plastic Thing 為食妹刺繡圖案外套	中性外套	\$699

This a live preview. When you are ready to run your project, click Get Data.

Show more data  Visuals enabled (advanced)

**catalog... BROWSE**

**Back to Commands**

**Project title**: catalog.hkstore.com Project

**Starting Site**: https://catalog.hkstore.com/cat

**Starting Template**: main\_template

**Web Hook**:

**Project Token**: You must save your project first  
tkOX156BQ5W2

**Max Workers**: 0

**Max Pages**:  2

**Enable Email Notifications**:

**Load JavaScript & Images**:

**Rotate IP Addresses**:

**Starting Value**:

**Max # of pages to be scraped.**

男裝 男裝

https://catalog.hkstore.com/catalog\_to\_hk/men?p=2

Click each s... 標題

訂單淨值滿\$400即享免費門市取貨 Select Mode

CATALOG MARATHON SPORTS

最新商品 品牌 聯乘系列 男裝 女裝 童裝 電子禮品卡 SALE

分類 +

熱賣系列 +

商品分類 +

鞋款 +

服裝類別 +

配件類別 +

袋款 +

品牌 +

性別 +

New Balance 2002R 中性球鞋

New Balance 2002R 中性球鞋

New Balance 2002R 中性球鞋

CSV/Excel JSON CSV/Excel Wide (beta)

selection1_name	selection1_series	selection1_style	selection1_price
New Balance	2002R	中性球鞋	\$1,199
New Balance	2002R	中性球鞋	\$1,199
New Balance	2002R	中性球鞋	\$1,199
PLASTIC THING X CATALOG系列	Plastic Thing 為食妹刺繡圖案外套	中性外套	\$699

This a live preview. When you are ready to run your project, click Get Data.

Show more data ?  Visuals enabled (advanced) ?

catalog.hkstore.com Project

Edit project

Your data is ready! Click on the green buttons to download.

Download Data

CSV/Excel

JSON

API

Template Name  
main\_template

Pages Scrapped  
2

# of pages scraped.

All dates and times are in UTC +0000.

Empty file with no results? [Click here](#) to fix.

CSV file too big? Save the JSON file and [click here](#) to convert to CSV.

## Details

Status	complete
Pages	2 collected
Initialized	2022-10-03T02:55:52
Start Time	2022-10-03T02:55:53
Finished	2022-10-03T02:56:29
API Key	tkOX156BQ5W2
Project Token	t2hdrZsNoUGS
Run Token	tuSJ0HKZxiyo

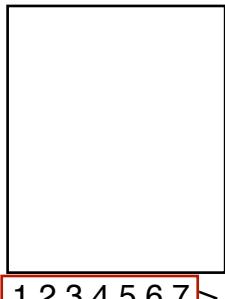
## Settings

URL	<input type="text" value="https://catalog.hkstore.com/catalog_tc_hk/men"/>
Starting Template	main_template
Starting Value	<input type="text" value="0"/>
Load Javascript	true
Rotate IPs	false

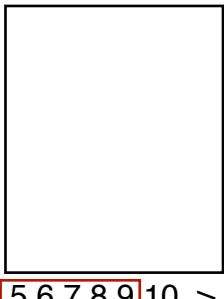


**Periodic scraping in smaller batches.**

**Monday  
morning**

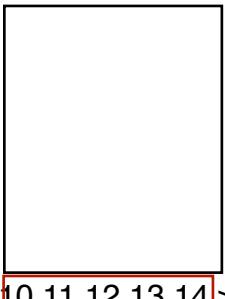


1 2 3 4 5 6 7 >

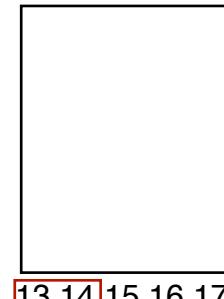


5 6 7 8 9 10 >

**Monday  
afternoon**

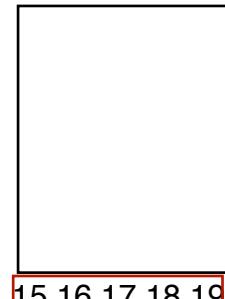


10 11 12 13 14 >

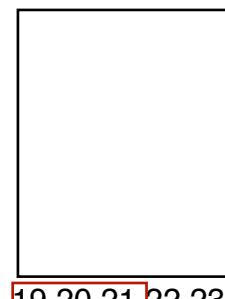


13 14 15 16 17 >

**Tuesday**



15 16 17 18 19 >



19 20 21 22 23 >

**More Tutorials Can be Found Below**

<https://www.youtube.com/channel/UCRfbcJQu9HZvc7uks2wmGzw/playlists>

## **Legal and Ethical Considerations.**

scraping and how they play out.

Share this



## Is Web Scraping Legal?

Your first thought might be to look at the legal side of things.

The truth is that the legality of web scraping is still relatively up in the air.

Meaning that there are currently no specific laws that refer to the legality of web scraping. So it is neither legal or illegal.

<https://www.parsehub.com/blog/web-scraping-ethical/>

# Scraping Publicly Available Information

Another factor to keep in mind is the type of data you'd be scraping. In our case, we always refer to publicly available data.

This is data that has been made public by the owner of said data. Private and leaked information is not considered publicly available information.

# Ethics in Web Scraping



James Densmore Jul 23, 2017 · 3 min read



We all scrape web data. Well, those of us who work with data do. Data scientists, marketers, data journalists, and the data curious alike. Lately, I've been thinking more about the ethics of the practice and have been dissatisfied by the lack of consensus on the topic.

Let me be clear that I'm talking **ethics** not the law. The law in regards to scraping web data is complex, fuzzy and ripe for reform, but that's another matter. It's not that no one is thinking, or writing, about the ethics in scraping but rather that both those scraping and those being scraped can't

<https://towardsdatascience.com/ethics-in-web-scraping-b96b18136f01>

WRITTEN BY

James Densmore

Data Science and  
Data Engineering  
Consultant at Data  
Liftoff  
<https://www.dataliftoff.com>

Follow

---

397

5

2

## The Ethical Scraper

I, the web scraper will live by the following principles:

- If you have a public API that provides the data I'm looking for, I'll use it and avoid scraping all together.
- I will always provide a User Agent string that makes my intentions clear and provides a way for you to contact me with questions or concerns.
- I will request data at a reasonable rate. I will strive to never be confused for a DDoS attack.
- I will only save the data I absolutely need from your page. If all I need is OpenGraph meta-data, that's all I'll keep.
- I will respect any content I do keep. I'll never pass it off as my own.
- I will look for ways to return value to you. Maybe I can drive some (real) traffic to your site or credit you in an article or post.
- I will respond in a timely fashion to your outreach and work with you towards a resolution.
- I will scrape for the **purpose of creating new value from the data**, not to duplicate it.

# Clean up data in **OpenRefine**



A free, open source,  
powerful tool for working  
with messy data



[Home](#)  
[Community](#)  
[Documentation](#)  
**[Download](#)**  
[Data Privacy](#)  
[Contact Us](#)  
[Blog](#)

## Download

On this page you will find a list of OpenRefine distributions and extensions available for download. Are we missing something? Want to fix a typo? You can [submit changes](#).

### Official Distribution

Read the [installation instructions](#).

You can also download all official releases and source from our [GitHub releases page](#)

#### OpenRefine 3.5.2

The latest stable release of OpenRefine 3.5, released on January 26, 2021. Please [backup your workspace directory](#) before installing and report any problems that you encounter. A change log is provided on [the release page](#).

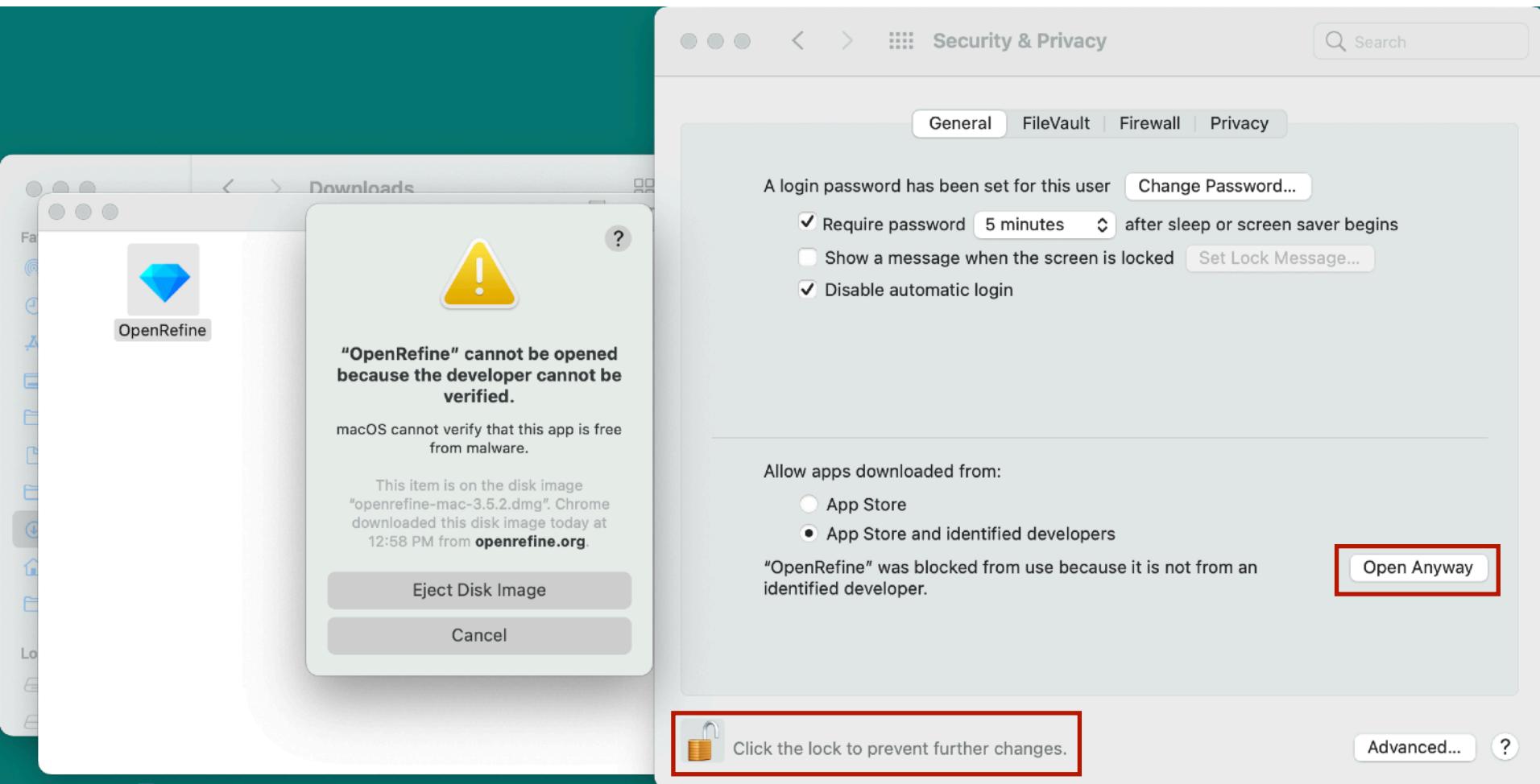
- **Windows kit**, This requires Java to be installed on your computer. Download, unzip, and double-click on `openrefine.exe` or `refine.bat` if the former does not work.
- **Windows kit with embedded Java**, includes [OpenJDK Java](#), available under the [GPLv2+CE](#) license. Download, unzip, and double-click on `openrefine.exe` or `refine.bat` if the former does not work.
- **Mac kit**, Download, open, drag icon into the Applications folder and double click on it. You do not need to install Java separately.
- **Linux kit**, Download, extract, then type `./refine` to start. This requires Java to be installed on your computer.

#### OpenRefine 3.4.1

The previous stable release of OpenRefine, released on September 24, 2020. Please [backup your workspace directory](#) before installing and report any problems that you encounter. A change log is provided on [the release page](#).

- **Windows kit**, This requires Java to be installed on your computer. Download,

<https://openrefine.org/download.html>



## **Most Basic Data Cleaning Tasks Part 1**

- 1. Resolve inconsistent entries.**
- 2. Rename, combine and split columns for making further data presentation and processing easier.**
- 3. Transform incorrect code (e.g. wrong data types such as text instead of numeric and calculations)**

# **Tutorial**

<http://d3-media.blogspot.hk/2013/11/how-to-refine-your-data.html>

# **Practice Dataset**

<https://raw.githubusercontent.com/suentze2020/workshop/main/lesson4/demo-refine-Sheet.csv>

## OpenRefine refine Sheet1 csv Permalink

Facet / Filter

Undo / Redo 0 / 0

### Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?  
[Watch these screencasts](#)

## Cleaning column 1: company names

Click on the drop down menu of header of column 1, company, and choose text facet.

25 rows

Show as: rows records

Show: 5 10 25 50 rows

All	company	Product code / n	address	city	country	name
1.	Facet		Text facet	arnhem	the netherlands	dhr p. jansen
2.		Text filter	Numeric facet	arnhem	the netherlands	dhr p. hansen
3.		Edit cells	Timeline facet	arnhem	the netherlands	dhr j. Gansen
4.		Edit column	Scatterplot facet	arnhem	the netherlands	dhr p. mansen
5.		Transpose	Custom text facet...	arnhem	the netherlands	dhr p. franssen
6.		Sort...	Custom Numeric Facet...	arnhem	the netherlands	dhr p. bansen
7.		View	Customized facets	arnhem	the netherlands	dhr p. vansen
8.			Leeuwardenweg 180	arnhem	the netherlands	dhr p. bransen
9.			Leeuwardenweg 181	arnhem	the netherlands	dhr p. janssen
10.		Reconcile	Leeuwardenweg 182	arnhem	the netherlands	mevr l. rokken
11.	ak zo	q-5	Leeuwardenweg 183	arnhem	the netherlands	mevr l. lokken
12.	akzo	q-9	Leeuwardenweg 184	arnhem	the netherlands	mevr l. mokken
13.	akzo	x-8	Delfzijlstraat 54	arnhem	the netherlands	mevr l. mokken
14.	phillips	p-56	Delfzijlstraat 55	arnhem	the netherlands	mevr l. mokken
15.	fillips	v-67	Delfzijlstraat 56	arnhem	the netherlands	mevr l. mokken
16.	philips	v-21	Delfzijlstraat 57	arnhem	the netherlands	mevr l. sokken
17.	Van Houten	x-45	Delfzijlstraat 58	arnhem	the netherlands	mevr l. wokken
18.	van Houten	v-56	Delfzijlstraat 59	arnhem	the netherlands	mevr l. kokken
19.	van houten	v-65	Delfzijlstraat 60	arnhem	the netherlands	mevr l. Bokken
20.	van houten	x-21	Delfzijlstraat 61	arnhem	the netherlands	mevr l. dokken
21.	Van Houten	p-23	Jourestraat 23	arnhem	the netherlands	mevr l. gokken
22.	unilver	x-3	Jourestraat 24	arnhem	the netherlands	mevr l. stokken
23.	unilever	q-4	Jourestraat 25	arnhem	the netherlands	mevr l. rokken
24.	Unilever	q-6	Jourestraat 26	arnhem	the netherlands	mevr l. rokken
25.	unilever	q-8				

The screenshot displays two panels of the OpenRefine interface. The left panel shows a list of companies with a 'Cluster' button highlighted. The right panel shows a 'Cluster & Edit column "company"' dialog with settings for 'Method: key collision' and 'Keying Function: fingerprint'. It lists clusters of names like 'akzo', 'Van Houten', 'philips', etc., with checkboxes for merging selected clusters.

On the left you see the companies listed. Click on cluster and choose 'key collision' and 'finger print' for clustering; that is correcting the names. Choose 'merge selected & re cluster'.

### Cluster & Edit column "company"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. [Find out more...](#)

Method key collision

Keying Function **fingerprint**

8 clusters found

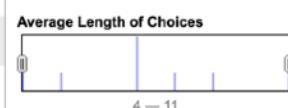
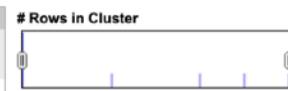
Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
1	6	• AKZO7 (6 rows)	<input type="checkbox"/>	AKZO7
1	3	• Unilever4 (3 rows)	<input type="checkbox"/>	Unilever4
1	7	• philips9 (7 rows)	<input type="checkbox"/>	philips9
1	1	• fillips (1 rows)	<input type="checkbox"/>	fillips
1	1	• philips (1 rows)	<input type="checkbox"/>	philips
1	5	• Van Houten5 (5 rows)	<input type="checkbox"/>	Van Houten5
1	1	• akz0 (1 rows)	<input type="checkbox"/>	akz0
1	1	• univer (1 rows)	<input type="checkbox"/>	univer

Select All

Unselect All

Export Clusters

Merge Selected & Re-Clu



### Cluster & Edit column "company"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. [Find out more...](#)

Method key collision

Keying Function **ngram-fingerprint**

Ngram Size 2

4 clusters found

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
1	9	• philips9 (9 rows)	<input type="checkbox"/>	philips9
1	5	• Van Houten5 (5 rows)	<input type="checkbox"/>	Van Houten5
1	4	• Unilever4 (4 rows)	<input type="checkbox"/>	Unilever4
1	7	• AKZO7 (7 rows)	<input type="checkbox"/>	AKZO7

Select All

Unselect All

Export Clusters

Merge Selected & Re-Cluster

Merge Selected & Close Close

Select All

Unselect All

Export Clusters

Merge Selected & Re-Cluster

Merge Selected & Close Close

Continue clustering with a new algorithm: ngram-fingerprint. Continue till you have the following names:

AKZO7  
philips9  
Unilever4  
Van Houten5

Facet / Filter   Undo / Redo 3 / 3

Refresh   Reset All   Remove All

company change

4 choices Sort by: name count Cluster

AKZO7 7  
philips9 9  
Unilever4 4  
Van Houten5 5  
Facet by choice counts

25 rows

Show as: rows records   Show: 5 10 25 50 rows

All	company	Product code / n	address	city	country	name	
1.	Facet		Groningen singel 147	arnhem	the netherlands	dhr p. jansen	
2.	Text filter		Groningen singel 148	arnhem	the netherlands	dhr p. hansen	
3.			Groningen singel 149	arnhem	the netherlands	dhr j. Gansen	
4.	Edit cells		Transform...				
5.	Edit column		Common transforms				
6.	Transpose						
7.	Sort...		Trim leading and trailing whitespace				
8.	View		Collapse consecutive whitespace				
9.			Unescape HTML entities				
10.	Reconcile		Replace Smart quotes with ascii				
11.	AKZO7	q-5	To titlecase				
12.	AKZO7	q-9	To uppercase				
13.	AKZO7	x-8	To lowercase				
14.	philips9	p-56	To number				
15.	philips9	v-67	To date				
16.	philips9	v-21	To text				
17.	Van Houten5	x-45	To null				
18.	Van Houten5	v-56	To empty string				
19.	Van Houten5	v-65					
20.	Van Houten5	x-21	Delfzijlstraat 54	arnhem	the netherlands	mevr l. Bokken	
21.	Van Houten5	p-23	Delfzijlstraat 55	arnhem	the netherlands	mevr l. dokken	
22.	Unilever4	x-3	Delfzijlstraat 56	arnhem	the netherlands	mevr l. gokken	
23.	Unilever4	q-4	Delfzijlstraat 57	arnhem	the netherlands	mevr l. stokken	
24.	Unilever4	q-6	Jourestraat 23	arnhem	the netherlands	mevr l. rokken	
25.	Unilever4	q-8	Jourestraat 24	arnhem	the netherlands	mevr l. rokken	
			Jourestraat 25	arnhem	the netherlands	mevr l. rokken	
			Jourestraat 26	arnhem	the netherlands	mevr l. rokken	

Now we have the correct spelling of the company names; let's turn them into uppercase. Choose from column company in the drop down menu edit cells, common transformations, into uppercase.

OpenRefine refine Sheet1 csv Permalink

Facet / Filter Undo / Redo 4 / 4

Text transform on 18 cells in column company:  
value.toUpperCase() Undo

Open... Export... Help

Extensions: Wikidata ▾

25 rows

Show as: rows records Show: 5 10 25 50 rows

« first < previous 1 - 25 next > last »

	company	product code / n	address	city	country	name
1.	PHILIPS9	p-5	Groningen singel 147	arnhem	the netherlands	dhr p. jansen
2.	PHILIPS9	p-43	Groningen singel 148	arnhem	the netherlands	dhr p. hansen
3.	PHILIPS9	x-3	Groningen singel 149	arnhem	the netherlands	dhr j. Gansen
4.	PHILIPS9	x-34	Groningen singel 150	arnhem	the netherlands	dhr p. mansen
5.	PHILIPS9	x-12	Groningen singel 151	arnhem	the netherlands	dhr p. fransen
6.	PHILIPS9	p-23	Groningen singel 152	arnhem	the netherlands	dhr p. franssen
7.	AKZO7	v-43	Leeuwardenweg 178	arnhem	the netherlands	dhr p. bansen
8.	AKZO7	v-12	Leeuwardenweg 179	arnhem	the netherlands	dhr p. vansen
9.	AKZO7	x-5	Leeuwardenweg 180	arnhem	the netherlands	dhr p. bransen
10.	AKZO7	p-34	Leeuwardenweg 181	arnhem	the netherlands	dhr p. janssen
11.	AKZO7	q-5	Leeuwardenweg 182	arnhem	the netherlands	mevr l. rokken
12.	AKZO7	q-9	Leeuwardenweg 183	arnhem	the netherlands	mevr l. lokken
13.	AKZO7	x-8	Leeuwardenweg 184	arnhem	the netherlands	mevr l. mokken
14.	PHILIPS9	p-56	Delfzijlstraat 54	arnhem	the netherlands	mevr l. mokken
15.	PHILIPS9	v-67	Delfzijlstraat 55	arnhem	the netherlands	mevr l. mokken
16.	PHILIPS9	v-21	Delfzijlstraat 56	arnhem	the netherlands	mevr l. mokken
17.	VAN HOUTEN5	x-45	Delfzijlstraat 57	arnhem	the netherlands	mevr l. sokken
18.	VAN HOUTEN5	v-56	Delfzijlstraat 58	arnhem	the netherlands	mevr l. wokken
19.	VAN HOUTEN5	v-65	Delfzijlstraat 59	arnhem	the netherlands	mevr l. kokken
20.	VAN HOUTEN5	x-21	Delfzijlstraat 60	arnhem	the netherlands	mevr l. Bokken
21.	VAN HOUTEN5	p-23	Delfzijlstraat 61	arnhem	the netherlands	mevr l. dokken
22.	UNILEVER4	x-3	Jourestraat 23	arnhem	the netherlands	mevr l. gokken
23.	UNILEVER4	q-4	Jourestraat 24	arnhem	the netherlands	mevr l. stokken
24.	UNILEVER4	q-6	Jourestraat 25	arnhem	the netherlands	mevr l. rokken
25.	UNILEVER4	q-8	Jourestraat 26	arnhem	the netherlands	mevr l. rokken

All company names are now capitalised.

Facet / Filter Undo / Redo 4 / 5 Extract... Apply...

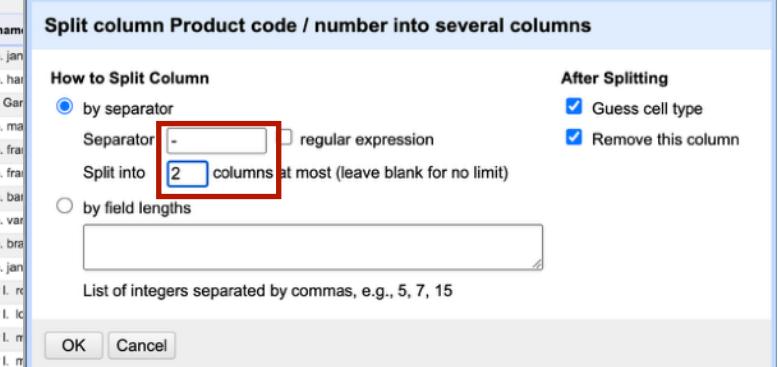
Facet / Filter Undo / Redo 7 / 7 Refresh Reset All Remove All

**company** change  
4 choices Sort by: name count Cluster

AZKO7 7  
PHILIPS9 9  
UNILEVER4 4  
VAN HOUTENS 5  
Facet by choice counts

25 rows Show as: rows records Show: 5 10 25 50 rows

		company	Product code / n	address	city	country	name
1.	PHILIPS9	Facet	Groningsingel 147	amhem	the netherlands	dhr p. jan	
2.	PHILIPS9	Text filter	Groningsingel 148	amhem	the netherlands	dhr p. han	
3.	PHILIPS9	Edit cells	Groningsingel 149	amhem	the netherlands	dhr j. Gar	
4.	PHILIPS9	Edit column	Groningsingel 150	amhem	the netherlands	dhr p. ma	
5.	PHILIPS9	Transpose				r p. fra	
6.	PHILIPS9	Sort...				r p. fra	
7.	AKZ07	View	Add column based on this column...			r p. bar	
8.	AKZ07	Reconcile	Add column by fetching URLs...			r p. var	
9.	AKZ07	q-5	Add columns from reconciled values...			r p. bra	
10.	AKZ07	Le	Rename this column			r p. jan	
11.	AKZ07	Remove this column				wvr l. ro	
12.	AKZ07	q-9	Move column to beginning			wvr l. id	
13.	AKZ07	x-8	Move column to end			wvr l. m	
14.	PHILIPS9	p-56	Move column left			wvr l. m	
15.	PHILIPS9	v-67	Move column right			wvr l. mokken	
16.	PHILIPS9	v-21				r p. malink	
17.	VAN HOUTENS	x-45	Delfzijlstraat 57	amhem	the netherlands	mevr l. st	
18.	VAN HOUTENS	v-56	Delfzijlstraat 58	amhem	the netherlands	mevr l. w	
19.	VAN HOUTENS	v-55	Delfzijlstraat 59	amhem	the netherlands	mevr l. lk	
20.	VAN HOUTENS	x-21	Delfzijlstraat 60	amhem	the netherlands	mevr l. B	
21.	VAN HOUTENS	p-23	Delfzijlstraat 61	amhem	the netherlands	mevr l. di	
22.	UNILEVER4	x-3	Jourestraat 23	amhem	the netherlands	mevr l. gr	
23.	UNILEVER4	q-4	Jourestraat 24	amhem	the netherlands	mevr l. st	
24.	UNILEVER4	q-6	Jourestraat 25	amhem	the netherlands	mevr l. rc	
25.	UNILEVER4	q-8	Jourestraat 26	amhem	the netherlands	mevr l. rc	



25 rows Show as: rows records Show: 5 10 25 50 rows

		company	Product code / n	address	city	country	name
1.	PHILIPS9	Facet	Groningsingel 147	amhem	the netherlands	dhr j. Gansen	
2.	PHILIPS9	Text filter	Groningsingel 148	amhem	the netherlands	dhr p. mansen	
3.	PHILIPS9	Edit cells	Groningsingel 149	amhem	the netherlands	dhr p. franssen	
4.	PHILIPS9	Edit column	Groningsingel 150	amhem	the netherlands	dhr p. bansen	
5.	PHILIPS9	Transpose				dhr p. vansen	
6.	PHILIPS9	Sort...				dhr p. bransen	
7.	AKZ07	View	Add column based on this column...			dhr p. janssen	
8.	AKZ07	Reconcile	Add column by fetching URLs...			mevr l. rokken	
9.	AKZ07	q	Add columns from reconciled values...			mevr l. lokken	
10.	AKZ07	Remove this column	Rename this column			mevr l. mokken	
11.	AKZ07	q	Move column to beginning			mevr l. mokken	
12.	AKZ07	x	Move column to end			mevr l. mokken	
13.	AKZ07	p	Move column left			mevr l. mokken	
14.	PHILIPS9	v	Move column right			mevr l. mokken	
15.	PHILIPS9	v				mevr l. mokken	
16.	PHILIPS9	v				mevr l. mokken	

**Split column 2 product name and number. Go to edit column in the drop down menu and choose split into several columns. Use – as the separator and choose 2 columns. Give column 3 a new name: choose edit column and rename: number. Change the name of column 2 in product, by choosing in the drop down menu: edit column and rename column.**

Facet / Filter   Undo / Redo 7 / 7

Refresh   Reset All   Remove All

company change

4 choices Sort by: name count Cluster

AKZO7 7  
PHILIPS9 9  
UNILEVER4 4  
VAN HOUTEN5 5  
Facet by choice counts

<input checked="" type="checkbox"/>	All	<input checked="" type="checkbox"/>	company	<input checked="" type="checkbox"/>	Product	<input checked="" type="checkbox"/>	Number	<input checked="" type="checkbox"/>	address	<input checked="" type="checkbox"/>	city	<input checked="" type="checkbox"/>	country	<input checked="" type="checkbox"/>	name
1.	PHILIPS9	p		5	Groningen singel 147	arnhem	the netherlands	dhr p. jansen							
2.	PHILIPS9	p		43	Groningen singel 148	arnhem	the netherlands	dhr p. hansen							
3.	PHILIPS9	x		3	Groningen singel 149	arnhem	the netherlands	dhr j. Gansen							
4.	PHILIPS9	x		34	Groningen singel 150	arnhem	the netherlands	dhr p. mansen							
5.	PHILIPS9	x		12	Groningen singel 151	arnhem	the netherlands	dhr p. franssen							
6.	PHILIPS9	x		23	Groningen singel 152	arnhem	the netherlands	dhr p. franssen							
7.	AKZO7	v		43	Leeuwardenweg 178	arnhem	the netherlands	dhr p. bansen							
8.	AKZO7	v		12	Leeuwardenweg 179	arnhem	the netherlands	dhr p. vansen							
9.	AKZO7	x		5	Leeuwardenweg 180	arnhem	the netherlands	dhr p. bransen							
10.	AKZO7	p		34	Leeuwardenweg 181	arnhem	the netherlands	dhr p. janssen							
11.	AKZO7	q		5	Leeuwardenweg 182	arnhem	the netherlands	mevr l. rokken							
12.	AKZO7	q		9	Leeuwardenweg 183	arnhem	the netherlands	mevr l. lokken							
13.	AKZO7	x		8	Leeuwardenweg 184	arnhem	the netherlands	mevr l. mokken							
14.	PHILIPS9	p		56	Delfzijlstraat 54	arnhem	the netherlands	mevr l. mokken							
15.	PHILIPS9	v		67	Delfzijlstraat 55	arnhem	the netherlands	mevr l. mokken							
16.	PHILIPS9	v		21	Delfzijlstraat 56	arnhem	the netherlands	mevr l. mokken							
17.	VAN HOUTEN5	x		45	Delfzijlstraat 57	arnhem	the netherlands	mevr l. sokken							
18.	VAN HOUTEN5	v		56	Delfzijlstraat 58	arnhem	the netherlands	mevr l. wokken							
19.	VAN HOUTEN5	v		65	Delfzijlstraat 59	arnhem	the netherlands	mevr l. kokken							
20.	VAN HOUTEN5	x		21	Delfzijlstraat 60	arnhem	the netherlands	mevr l. Bokken							
21.	VAN HOUTEN5	p		23	Delfzijlstraat 61	arnhem	the netherlands	mevr l. dokken							
22.	UNILEVER4	x		3	Jourestraat 23	arnhem	the netherlands	mevr l. gokken							
23.	UNILEVER4	q		4	Jourestraat 24	arnhem	the netherlands	mevr l. stokken							
24.	UNILEVER4	q		6	Jourestraat 25	arnhem	the netherlands	mevr l. rokken							
25.	UNILEVER4	q		8	Jourestraat 26	arnhem	the netherlands	mevr l. rokken							

Product code and number are now separated.

In column 2 we have a code for the product but not the name of the **product**. Here is the conversion:

p=radio

v=tv

x=computer

q=tablet

Go in column 2 to drop down menu, choose edit cells and transform. Now we have to write some code in order to the transformation:

Use the following expression:

value.replace("p", "radio")

The left column show the original and the right the result of the transformation.

Do this for all different products.

The screenshot shows the OpenRefine interface with a data grid titled '25 rows'. The columns are labeled 'Product', 'Number', 'address', 'city', 'country', and 'name'. The 'Product' column contains values like 'PHILIPS9', 'AKZO7', etc. A context menu is open over the 'Product' column, with 'Edit cells' and 'Transform...' highlighted. Below the grid, a 'Custom text transform on column Product' dialog is open. The 'Expression' field contains the code 'value.replace("p", "radio")'. The 'Preview' section shows the transformation results for rows 1 through 6. The 'On error' options are set to 'keep original'. The 'OK' button is visible at the bottom of the dialog.

row	value	value.replace("p", "radio")
1.	p	radio
2.	p	radio
3.	x	x
4.	x	x
5.	x	x
6.	p	radio

Text transform on 5 cells in column Product:  
grel:value.replace("q","tablet") [Undo](#)

Facet / Filter [Undo / Redo](#) 11 / 11

Refresh [Reset All](#) [Remove All](#)

company [change](#)

4 choices Sort by: name count [Cluster](#)

AKZO7 7  
PHILIPS9 9  
UNILEVER4 4  
VAN HOUTEN5 5  
Facet by choice counts

**25 rows**

Show as: [rows](#) [records](#) Show: [5](#) [10](#) [25](#) [50](#) rows

<input checked="" type="checkbox"/> All	<input checked="" type="checkbox"/> company	<input checked="" type="checkbox"/> Product	<input checked="" type="checkbox"/> Number	<input checked="" type="checkbox"/> address	<input checked="" type="checkbox"/> city	<input checked="" type="checkbox"/> country	<input checked="" type="checkbox"/> name
1.	PHILIPS9	radio	5	Groningen singel 147	arnhem	the netherlands	dhr p. jansen
2.	PHILIPS9	radio	43	Groningen singel 148	arnhem	the netherlands	dhr p. hansen
3.	PHILIPS9	computer	3	Groningen singel 149	arnhem	the netherlands	dhr j. gansen
4.	PHILIPS9	computer	34	Groningen singel 150	arnhem	the netherlands	dhr p. mansen
5.	PHILIPS9	computer	12	Groningen singel 151	arnhem	the netherlands	dhr p. fransen
6.	PHILIPS9	radio	23	Groningen singel 152	arnhem	the netherlands	dhr p. franssen
7.	AKZO7	tv	43	Leeuwardenweg 178	arnhem	the netherlands	dhr p. bansen
8.	AKZO7	tv	12	Leeuwardenweg 179	arnhem	the netherlands	dhr p. vansen
9.	AKZO7	computer	5	Leeuwardenweg 180	arnhem	the netherlands	dhr p. bransen
10.	AKZO7	radio	34	Leeuwardenweg 181	arnhem	the netherlands	dhr p. janssen
11.	AKZO7	tablet	5	Leeuwardenweg 182	arnhem	the netherlands	mevr l. rokken
12.	AKZO7	tablet	9	Leeuwardenweg 183	arnhem	the netherlands	mevr l. lokken
13.	AKZO7	computer	8	Leeuwardenweg 184	arnhem	the netherlands	mevr l. mokken
14.	PHILIPS9	radio	56	Delfzijlstraat 54	arnhem	the netherlands	mevr l. mokken
15.	PHILIPS9	tv	67	Delfzijlstraat 55	arnhem	the netherlands	mevr l. mokken
16.	PHILIPS9	tv	21	Delfzijlstraat 56	arnhem	the netherlands	mevr l. mokken
17.	VAN HOUTEN5	computer	45	Delfzijlstraat 57	arnhem	the netherlands	mevr l. sokken
18.	VAN HOUTEN5	tv	56	Delfzijlstraat 58	arnhem	the netherlands	mevr l. wokken
19.	VAN HOUTEN5	tv	65	Delfzijlstraat 59	arnhem	the netherlands	mevr l. kokken
20.	VAN HOUTEN5	computer	21	Delfzijlstraat 60	arnhem	the netherlands	mevr l. Bokken
21.	VAN HOUTEN5	radio	23	Delfzijlstraat 61	arnhem	the netherlands	mevr l. dokken
22.	UNILEVER4	computer	3	Jourestraat 23	arnhem	the netherlands	mevr l. gokken
23.	UNILEVER4	tablet	4	Jourestraat 24	arnhem	the netherlands	mevr l. stokken
24.	UNILEVER4	tablet	6	Jourestraat 25	arnhem	the netherlands	mevr l. rokken
25.	UNILEVER4	tablet	8	Jourestraat 26	arnhem	the netherlands	mevr l. rokken

Product renamed to more meaningful representation.

Facet / Filter Undo / Redo 11 / 11

Refresh Reset All Remove All

company change

4 choices Sort by: name count

Cluster

AKZO7 7

PHILIPS9 9

UNILEVER4 4

VAN HOUTEN5 5

Facet by choice counts

25 rows

Show as: rows records Show: 5 10 25 50 rows

All	company	Product	Number	address	city	country	name
1.	PHILIPS9	radio	5	Facet	amhem	the netherlands	dhr p. jansen
2.	PHILIPS9	radio	43	Text filter	amhem	the netherlands	dhr p. hansen
3.	PHILIPS9	computer	3				
4.	PHILIPS9	computer	34	Edit cells	Transform...		
5.	PHILIPS9	computer	12	Edit column	Common transforms		
6.	PHILIPS9	radio	23	Transpose	Fill down		
7.	AKZO7	tv	43	Sort...	Blank down		
8.	AKZO7	tv	12	View	Split multi-valued cells...		
9.	AKZO7	computer	5	Reconcile	Join multi-valued cells...		
10.	AKZO7	radio	34				
11.	AKZO7	tablet	5	Leeuwardenweg 182	ar	okken	
12.	AKZO7	tablet	9	Leeuwardenweg 183	ar	okken	
13.	AKZO7	computer	8	Leeuwardenweg 184	ar	okken	
14.	PHILIPS9	radio	56	Delfzijlstraat 54	amhem	the netherlands	mevr l. mokken
15.	PHILIPS9	tv	67	Delfzijls			
16.	PHILIPS9	tv	21	Delfzijls			
17.	VAN HOUTEN5	computer	45	Delfzijls			
18.	VAN HOUTEN5	tv	56	Delfzijls			
19.	VAN HOUTEN5	tv	65	Delfzijls			
20.	VAN HOUTEN5	computer	21	Delfzijls			
21.	VAN HOUTEN5	radio	23	Delfzijls			
22.	UNILEVER4	computer	3	Jourest			
23.	UNILEVER4	tablet	4	Jourest			
24.	UNILEVER4	tablet	6	Jourest			
25.	UNILEVER4	tablet	8	Jourest			

Custom text transform on column address

Expression: `cells["address"].value + ", " + cells["city"].value + ", " + cells["country"].value`

Language: General Refine Expression Language (GREL)

Preview History Starred Help

row value	cells["address"].value + "..."
1. Groningsingel 147	Groningsingel 147, amhem, the netherlands
2. Groningsingel 148	Groningsingel 148, amhem, the netherlands
3. Groningsingel 149	Groningsingel 149, amhem, the netherlands
4. Groningsingel 150	Groningsingel 150, amhem, the netherlands
5. Groningsingel 151	Groningsingel 151, amhem, the netherlands
6. Groningsingel 152	Groningsingel 152, amhem, the netherlands

On error:  keep original  set to blank  store error  Re-transform up to 10 times until no change

OK Cancel

In order to geocode we have to bring all address information into one column. First address and city. Go to address choose edit column, add column based on this column: use the following transformation:

```
cells["address"].value + ", " + cells["city"].value + ", " + cells["country"].value
```

25 rows

Show as: rows records Show: 5 10 25 50 rows

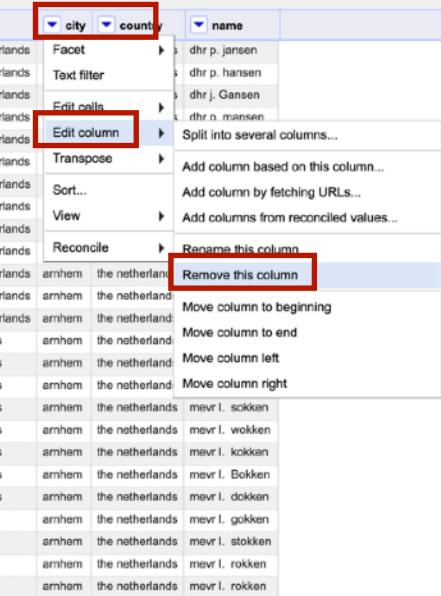
All	company	Product	Number	address	city	country	name
1.	PHILIPS9	radio	5	Groningsingel 147, arnhem, the netherlands	arnhem	the netherlands	dhr p. jansen
2.	PHILIPS9	radio	43	Groningsingel 148, arnhem, the netherlands	arnhem	the netherlands	dhr p. hansen
3.	PHILIPS9	computer	3	Groningsingel 149, arnhem, the netherlands	arnhem	the netherlands	dhr j. Gansen
4.	PHILIPS9	computer	34	Groningsingel 150, arnhem, the netherlands	arnhem	the netherlands	dhr p. mansen
5.	PHILIPS9	computer	12	Groningsingel 151, arnhem, the netherlands	arnhem	the netherlands	dhr p. franssen
6.	PHILIPS9	radio	23	Groningsingel 152, arnhem, the netherlands	arnhem	the netherlands	dhr p. franssen
7.	AKZO7	tv	43	Leeuwardenweg 178, arnhem, the netherlands	arnhem	the netherlands	dhr p. jansen
8.	AKZO7	tv	12	Leeuwardenweg 179, arnhem, the netherlands	arnhem	the netherlands	dhr p. hansen
9.	AKZO7	computer	5	Leeuwardenweg 180, arnhem, the netherlands	arnhem	the netherlands	dhr j. Gansen
10.	AKZO7	radio	34	Leeuwardenweg 181, arnhem, the netherlands	arnhem	the netherlands	dhr p. mansen
11.	AKZO7	tablet	5	Leeuwardenweg 182, arnhem, the netherlands	arnhem	the netherlands	dhr p. franssen
12.	AKZO7	tablet	9	Leeuwardenweg 183, arnhem, the netherlands	arnhem	the netherlands	dhr p. jansen
13.	AKZO7	computer	8	Leeuwardenweg 184, arnhem, the netherlands	arnhem	the netherlands	dhr p. hansen
14.	PHILIPS9	radio	56	Delfzijlstraat 54, arnhem, the netherlands	arnhem	the netherlands	dhr j. Gansen
15.	PHILIPS9	tv	67	Delfzijlstraat 55, arnhem, the netherlands	arnhem	the netherlands	dhr p. mansen
16.	PHILIPS9	tv	21	Delfzijlstraat 56, arnhem, the netherlands	arnhem	the netherlands	dhr p. franssen
17.	VAN HOUTEN5	computer	45	Delfzijlstraat 57, arnhem, the netherlands	arnhem	the netherlands	dhr p. jansen
18.	VAN HOUTEN5	tv	56	Delfzijlstraat 58, arnhem, the netherlands	arnhem	the netherlands	dhr p. hansen
19.	VAN HOUTEN5	tv	65	Delfzijlstraat 59, arnhem, the netherlands	arnhem	the netherlands	dhr j. Gansen
20.	VAN HOUTEN5	computer	21	Delfzijlstraat 60, arnhem, the netherlands	arnhem	the netherlands	dhr p. mansen
21.	VAN HOUTEN5	radio	23	Delfzijlstraat 61, arnhem, the netherlands	arnhem	the netherlands	dhr p. franssen
22.	UNILEVER4	computer	3	Jourestraat 23, arnhem, the netherlands	arnhem	the netherlands	dhr p. jansen
23.	UNILEVER4	tablet	4	Jourestraat 24, arnhem, the netherlands	arnhem	the netherlands	dhr p. hansen
24.	UNILEVER4	tablet	6	Jourestraat 25, arnhem, the netherlands	arnhem	the netherlands	dhr j. Gansen
25.	UNILEVER4	tablet	8	Jourestraat 26, arnhem, the netherlands	arnhem	the netherlands	dhr p. mansen

After combining the address, city, and country columns into one column, we can now remove the city and country columns.

25 rows

Show as: rows records Show: 5 10 25 50 rows

All	company	Product	Number	address	city	country	name
1.	PHILIPS9	radio	5	Groningsingel 147, arnhem, the netherlands	arnhem	the netherlands	dhr p. jansen
2.	PHILIPS9	radio	43	Groningsingel 148, arnhem, the netherlands	arnhem	the netherlands	dhr p. hansen
3.	PHILIPS9	computer	3	Groningsingel 149, arnhem, the netherlands	arnhem	the netherlands	dhr j. Gansen
4.	PHILIPS9	computer	34	Groningsingel 150, arnhem, the netherlands	arnhem	the netherlands	dhr p. mansen
5.	PHILIPS9	computer	12	Groningsingel 151, arnhem, the netherlands	arnhem	the netherlands	dhr p. franssen
6.	PHILIPS9	radio	23	Groningsingel 152, arnhem, the netherlands	arnhem	the netherlands	dhr p. franssen
7.	AKZO7	tv	43	Leeuwardenweg 178, arnhem, the netherlands	arnhem	the netherlands	dhr p. jansen
8.	AKZO7	tv	12	Leeuwardenweg 179, arnhem, the netherlands	arnhem	the netherlands	dhr p. hansen
9.	AKZO7	computer	5	Leeuwardenweg 180, arnhem, the netherlands	arnhem	the netherlands	dhr j. Gansen
10.	AKZO7	radio	34	Leeuwardenweg 181, arnhem, the netherlands	arnhem	the netherlands	dhr p. mansen
11.	AKZO7	tablet	5	Leeuwardenweg 182, arnhem, the netherlands	arnhem	the netherlands	dhr p. franssen
12.	AKZO7	tablet	9	Leeuwardenweg 183, arnhem, the netherlands	arnhem	the netherlands	dhr p. jansen
13.	AKZO7	computer	8	Leeuwardenweg 184, arnhem, the netherlands	arnhem	the netherlands	dhr p. hansen
14.	PHILIPS9	radio	56	Delfzijlstraat 54, arnhem, the netherlands	arnhem	the netherlands	dhr j. Gansen
15.	PHILIPS9	tv	67	Delfzijlstraat 55, arnhem, the netherlands	arnhem	the netherlands	dhr p. mansen
16.	PHILIPS9	tv	21	Delfzijlstraat 56, arnhem, the netherlands	arnhem	the netherlands	dhr p. franssen
17.	VAN HOUTEN5	computer	45	Delfzijlstraat 57, arnhem, the netherlands	arnhem	the netherlands	dhr p. jansen
18.	VAN HOUTEN5	tv	56	Delfzijlstraat 58, arnhem, the netherlands	arnhem	the netherlands	dhr p. hansen
19.	VAN HOUTEN5	tv	65	Delfzijlstraat 59, arnhem, the netherlands	arnhem	the netherlands	dhr j. Gansen
20.	VAN HOUTEN5	computer	21	Delfzijlstraat 60, arnhem, the netherlands	arnhem	the netherlands	dhr p. mansen
21.	VAN HOUTEN5	radio	23	Delfzijlstraat 61, arnhem, the netherlands	arnhem	the netherlands	dhr p. franssen
22.	UNILEVER4	computer	3	Jourestraat 23, arnhem, the netherlands	arnhem	the netherlands	dhr p. jansen
23.	UNILEVER4	tablet	4	Jourestraat 24, arnhem, the netherlands	arnhem	the netherlands	dhr p. hansen
24.	UNILEVER4	tablet	6	Jourestraat 25, arnhem, the netherlands	arnhem	the netherlands	dhr j. Gansen
25.	UNILEVER4	tablet	8	Jourestraat 26, arnhem, the netherlands	arnhem	the netherlands	dhr p. mansen



## **Most Basic Data Cleaning Part 2**

- 4. Remove duplicates.**
- 5. Remove blanks.**

Select: NBA\_Page()



Select Row



Extract Rank

Relative Pk

Relative Player

Relative College

Get Data

Selection Node:

p 1st body

2014 NBA Draft | Basketball-Reference.com

[https://www.basketball-reference.com/draft/NBA\\_2014.html](https://www.basketball-reference.com/draft/NBA_2014.html)

Sports Reference | Baseball | Football (college) | Basketball (college) | Hockey | Football | Blog | Stathead | Widgets

Login | Logout | Chat

Select Mode

Enter Person, Team, Section, etc

Search

BASKETBALL REFERENCE

Players Teams Seasons Leaders Scores (2) WNBA Draft Stathead Newsletter Full Site Menu Below ▾

## 2014 NBA Draft

« 2013 NBA Draft | 2015 NBA Draft »

Date: Thursday, June 26, 2014  
 Location: New York, New York  
 Number of Picks: 60 (53 played in NBA)  
 First Overall Pick: [Andrew Wiggins](#) (23.5 Win Shares)  
 Most Win Shares: [N. Jokić](#) (74.7), [C. Capela](#) (49.7) and [J. Embiid](#) (39.6)  
 All-Stars: 4 ([J. Embiid](#), [N. Jokić](#), [Z. LaVine](#) and [J. Randle](#))

 via Sports Logos.net About logos

HURRY, ENDS SOON 

Draft History Draft Years ▾

CSV/Excel	JSON	CSV/Excel Wide (beta)				
Row_Rank	Row_Pk	Row_Pk_url	Row_Player	Row_Player_url	Row_College	Row_College_url
1	1	<a href="https://stathead.com/basketball/draft_finder.cgi?request...year_min=&amp;year_max=&amp;round_min=&amp;round_max=&amp;order_by=rank&amp;sort_order=asc&amp;format=json&amp;rows=1000&amp;page=1">https://stathead.com/basketball/draft_finder.cgi?request...year_min=&amp;year_max=&amp;round_min=&amp;round_max=&amp;order_by=rank&amp;sort_order=asc&amp;format=json&amp;rows=1000&amp;page=1</a>	Andrew Wiggins	<a href="https://www.basketball-reference.com/players/w/wiggian01.html">https://www.basketball-reference.com/players/w/wiggian01.html</a>	Kansas	<a href="https://www.basketball-reference.com/friv/draft.fcgi?college=k...">https://www.basketball-reference.com/friv/draft.fcgi?college=k...</a>
2	2	<a href="https://stathead.com/basketball/draft_finder.cgi?request...year_min=&amp;year_max=&amp;round_min=&amp;round_max=&amp;order_by=rank&amp;sort_order=asc&amp;format=json&amp;rows=1000&amp;page=1">https://stathead.com/basketball/draft_finder.cgi?request...year_min=&amp;year_max=&amp;round_min=&amp;round_max=&amp;order_by=rank&amp;sort_order=asc&amp;format=json&amp;rows=1000&amp;page=1</a>	Zach LaVine	<a href="https://www.basketball-reference.com/players/l/lavine01.html">https://www.basketball-reference.com/players/l/lavine01.html</a>	Duke	<a href="https://www.basketball-reference.com/friv/draft.fcgi?college=d...">https://www.basketball-reference.com/friv/draft.fcgi?college=d...</a>

This is a live preview. When you are ready to run your project, click Get Data.

Show more data   Visuals enabled (advanced) 

	Round 1			Totals			Shooting			Per Game			Advanced								
Rk	Pk	Tm	Player	College	Yrs	G	MP	PTS	TRB	AST	FG%	3P%	FT%	MP	PTS	TRB	AST	WS	WS/48	BPM	VORP

Diagram illustrating the calculation of various basketball statistics from raw data:

- Team** (Rk, Pk, Tm) feeds into **Player**.
- Player** feeds into **College**.
- College** feeds into **Yrs**.
- Yrs** feeds into **G**.
- G** feeds into **Points**.
- G** feeds into **Total Rebounds**.
- Minutes Played** (MP) feeds into **Assists**.
- Assists** feeds into **Field Goal Percentage**.
- Field Goal Percentage** feeds into **3-Point Field Goal Percentage**.
- 3-Point Field Goal Percentage** feeds into **Free Throw Percentage**.
- Free Throw Percentage** feeds into **Win Shares**.
- Win Shares** (An estimate of the number of wins contributed by a player) feeds into **Win Shares Per 48 Minutes**.
- Win Shares Per 48 Minutes** (An estimate of the number of wins contributed by a player per 48 minutes (league average is approximately .100)) feeds into **Box Plus/Minus**.
- Box Plus/Minus** (A box score estimate of the points per 100 possessions a player contributed above a league-average player, translated to an average team) feeds into **Value over Replacement Player**.
- Value over Replacement Player** (A box score estimate of the points per 100 TEAM possessions that a player contributed above a replacement-level (-2.0) player, translated to an average team and prorated to an 82-game season) has a note: "Multiply by 2.70 to convert to wins over replacement."

Facet / Filter Undo / Redo 0 / 0

63 rows Extensions: Wikidata ▾

Show as: rows records Show: 5 10 25 50 rows « first < previous 1 - 10 next > last »

All	Column	Rk	Pk	Tm	Player	College	Yrs	G	MP	PTS	TRB	AST	FG%	FT%	3P%	STL	BLK	PF	TO	PER
1.	0	1	1	CLE	Andrew Wiggins	Kansas	8	536	18929	10454	2371	1248	.446	.341						
2.	1	2	2	MIL	Jabari Parker	Duke	8	302	8464	4351	1682	605	.495	.324						
3.	2	3	3	PHI	Joel Embiid	Kansas	6	269	8221	6649	3022	832	.485	.331						
4.	3	4	4	ORL	Aaron Gordon	Arizona	8	464	13232	5889	2939	1145	.451	.320						
5.	4	5	5	UTA	Dante Exum		6	245	4545	1389	441	515	.407	.305						
6.	5	6	6	BOS	Marcus Smart	Oklahoma State	8	459	13540	4667	1631	1943	.375	.319						
7.	6	7	7	LAL	Julius Randle	Kentucky	8	458	14032	8007	4224	1534	.483	.344						
8.	7	8	8	SAC	Nik Stauskas	Michigan	5	335	6662	2272	688	513	.389	.353						
9.	7	8	8	SAC	Nik Stauskas	Michigan	5	335	6662	2272	688	513	.389	.353						

1. Run the sort function on the column which contains duplicates. For instructions on sorting, see the [Sorting Data](#) section.
2. After you have sorted the column, choose "Sort" and then "reorder rows permanently."
3. Go to the column with duplicates and click on the arrow button in the column header.
4. Choose "Edit cells" and then select "Blank down."
  - a. "Blank down" will detect if two rows following each other have the same content. If they do, the second row will be "blanked out" and the cell values removed.

Source: library.illinois.edu

Facet / Filter Undo / Redo 0 / 0

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?  
[Watch these screencasts](#)

63 rows														Extensions: Wikidata					
		Show as: <a href="#">rows</a> <a href="#">records</a>	Show: <a href="#">5</a> <a href="#">10</a> <a href="#">25</a> <a href="#">50</a>	<a href="#">rows</a>														« first < previous <a href="#">1 - 10</a> next > last »	
All		<a href="#">Column</a>	<a href="#">Row</a>	Ck	Pk	Tm	<a href="#">Player</a>	College	<a href="#">Yrs</a>	G	MP	<a href="#">PTS</a>	TRB	<a href="#">AST</a>	<a href="#">FG%</a>				
1.	Facet			CLE			Andrew Wiggins	Kansas	8	536	18929	10454	2371	1248	.446	.341			
2.	Text filter			MIL			Jabari Parker	Duke	8	302	8464	4351	1682	605	.495	.324			
3.	Edit cells			PHI			Joel Embiid	Kansas	6	269	8221	6649	3022	832	.485	.331			
4.	Edit column			ORL			Aaron Gordon	Arizona	8	464	13232	5889	2939	1145	.451	.320			
5.	Transpose			UTA			Dante Exum		6	245	4545	1389	441	515	.407	.305			
6.	Sort...			BOS			Marcus Smart	Oklahoma State	8	459	13540	4667	1631	1943	.375	.319			
7.	View			LAL			Julius Randle	Kentucky	8	458	14032	8007	4224	1534	.483	.344			
8.	Reconcile			SAC	Nik Stauskas		Michigan	5	335	6662	2272	688	513	.389	.353				
9.				SAC	Nik Stauskas		Michigan	5	335	6662	2272	688	513	.389	.353				

- Run the sort function on the column which contains duplicates. For instructions on sorting, see the [Sorting Data](#) section.
- After you have sorted the column, choose "Sort" and then "reorder rows permanently."
- Go to the column with duplicates and click on the arrow button in the column header.
- Choose "Edit cells" and then select "Blank down."
  - "Blank down" will detect if two rows following each other have the same content. If they do, the second row will be "blanked out" and the cell values removed.

Source: library.illinois.edu

OpenRefine nba csv Permalink

Facet / Filter Undo / Redo 0 / 0

63 rows

Show as: rows records Show: 5 10 25 50 rows

Extensions: Wikidata ▾

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?  
[Watch these screencasts](#)

**Sort by Column**

Sort cell values as

text  case-sensitive  
 numbers  
 dates  
 booleans

Position blanks and errors

Valid values  
 Errors  
 Blanks

Drag and drop to re-order

smallest first  largest first

	G	MP	PTS	TRB	AST	FG%
536	18929	10454	2371	1248	.446	.341
302	8464	4351	1682	605	.495	.324
269	8221	6649	3022	832	.485	.331
464	13232	5889	2939	1145	.451	.320
245	4545	1389	441	515	.407	.305
459	13540	4667	1631	1943	.375	.319
458	14032	8007	4224	1534	.483	.344
335	6662	2272	688	513	.389	.353
335	6662	2272	688	513	.389	.353
					.459	.308

1. Run the sort function on the column which contains duplicates. For instructions on sorting, see the [Sorting Data](#) section.
2. After you have sorted the column, choose "Sort" and then "reorder rows permanently."
3. Go to the column with duplicates and click on the arrow button in the column header.
4. Choose "Edit cells" and then select "Blank down."
  - a. "Blank down" will detect if two rows following each other have the same content. If they do, the second row will be "blanked out" and the cell values removed.

Source: library.illinois.edu

Facet / Filter Undo / Redo 0 / 0

Extensions: Wikidata

### Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?

[Watch these screencasts](#)

63 rows												Extensions: Wikidata				
Show as: <a href="#">rows</a> <a href="#">records</a>			Show: <a href="#">5</a> <a href="#">10</a> <a href="#">25</a> <a href="#">50</a> rows					Sort ▾		« first < previous <a href="#">1 - 10</a> next > last »						
All	Column	Rk	Pk	Tm	Player	Remove sort					TRB	AST	FG%	...		
1.	0	1	1	CLE	Andrew Wiggins	2371	1248	.446	.341							
2.	1	2	2	MIL	Jabari Parker	1682	605	.495	.324							
3.	2	3	3	PHI	Joel Embiid	Kansas	6	269	8221	6649	3022	832	.485	.331		
4.	3	4	4	ORL	Aaron Gordon	Arizona	8	464	13232	5889	2939	1145	.451	.320		
5.	4	5	5	UTA	Dante Exum		6	245	4545	1389	441	515	.407	.305		
6.	5	6	6	BOS	Marcus Smart	Oklahoma State	8	459	13540	4667	1631	1943	.375	.319		
7.	6	7	7	LAL	Julius Randle	Kentucky	8	458	14032	8007	4224	1534	.483	.344		
8.	7	8	8	SAC	Nik Stauskas	Michigan	5	335	6662	2272	688	513	.389	.353		
9.	7	8	8	SAC	Nik Stauskas	Michigan	5	335	6662	2272	688	513	.389	.353		

1. Run the sort function on the column which contains duplicates. For instructions on sorting, see the [Sorting Data](#) section
2. After you have sorted the column choose 'Sort' and then 'reorder rows permanently.'
3. Go to the column with duplicates and click on the arrow button in the column header.
4. Choose "Edit cells" and then select "Blank down."
  - a. "Blank down" will detect if two rows following each other have the same content. If they do, the second row will be "blanked out" and the cell values removed.

Source: library.illinois.edu

## 63 rows

Show as: rows records

Show: 5 10 25 50 rows

&lt;&lt; first &lt; previous 1 - 10 next &gt; last &gt;&gt;

Using facets and filters 

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?

[Watch these screencasts](#)

All	Column	Rk	Pk	Tm	Player	College	Yrs	G	MP	PTS	TRB	AST	FG%	
1.	Facet		CLE	Andrew Wiggins	Kansas	8	536	18929	10454	2371	1248	.446	.341	
2.	Text filter		MIL	Jabari Parker	Duke	8	302	8464	4351	1682	605	.495	.324	
3.	Edit cells				Transform...	sas	6	269	8221	6649	3022	832	.485	.331
4.	Edit column				Common transforms	ona	8	464	13232	5889	2939	1145	.451	.320
5.	Transpose				Fill down		6	245	4545	1389	441	515	.407	.305
6.	Sort...				Blank down	anoma	8	459	13540	4667	1631	1943	.375	.319
7.	View				Split multi-valued cells...	tucky	8	458	14032	8007	4224	1534	.483	.344
8.	Reconcile				Join multi-valued cells...	nigan	5	335	6662	2272	688	513	.389	.353
9.					Cluster and edit...	nigan	5	335	6662	2272	688	513	.389	.353
					Replace		7							

1. Run the sort function on the column which contains duplicates. For instructions on sorting, see the [Sorting Data](#) section.
2. After you have sorted the column, choose "Sort" and then "reorder rows permanently."
3. Go to the column with duplicates and click on the arrow button in the column header.
4. Choose "Edit cells" and then select "Blank down."
  - a. "Blank down" will detect if two rows following each other have the same content. If they do, the second row will be "blanked out" and the cell values removed.

Facet / Filter Undo / Redo 2 / 2

63 rows

Extensions: Wikidata

Show as: rows records

Show: 5 10 25 50 rows

&lt; first &lt; previous 1 - 10 next &gt; last &gt;

**Using facets and filters**

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?

[Watch these screencasts](#)

All	Column	Rk	Pk	Tm	Player	College	Yrs	G	MP	PTS	TRB	AST	FG%	
1. 0	1	1	1	CLE	Andrew Wiggins	Kansas	8	536	18929	10454	2371	1248	.446	.341
2. 1	2	2	2	MIL	Jabari Parker	Duke	8	302	8464	4351	1682	605	.495	.324
3. 2	3	3	3	PHI	Joel Embiid	Kansas	6	269	8221	6649	3022	832	.485	.331
4. 3	4	4	4	ORL	Aaron Gordon	Arizona	8	464	13232	5889	2939	1145	.451	.320
5. 4	5	5	5	UTA	Dante Exum		6	245	4545	1389	441	515	.407	.305
6. 5	6	6	6	BOS	Marcus Smart	Oklahoma State	8	459	13540	4667	1631	1943	.375	.319
7. 6	7	7	7	LAL	Julius Randle	Kentucky	8	458	14032	8007	4224	1534	.483	.344
8. 7	8	8	SAC	Nik Stauskas	Michigan	5	335	6662	2272	688	513	.389	.353	
9.		8	8	SAC	Nik Stauskas	Michigan	5	335	6662	2272	688	513	.389	.353

1. Run the sort function on the column which contains duplicates. For instructions on sorting, see the [Sorting Data](#) section.
2. After you have sorted the column, choose "Sort" and then "reorder rows permanently."
3. Go to the column with duplicates and click on the arrow button in the column header.
4. Choose "Edit cells" and then select "Blank down."
  - a. "Blank down" will detect if two rows following each other have the same content. If they do, the second row will be "blanked out" and the cell values removed.

Source: library.illinois.edu

Facet / Filter

Undo / Redo 2 / 2

### Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?

[Watch these screencasts](#)

5. After you have used the “Blank down” function, use the “[Facet by blank](#)” to identify rows with blank cell values for that column.
6. From the facet window, select the “true” option.
7. Go to the column labeled “All” and click on the arrow button, then select “Edit rows” and choose “Remove all matching rows.”
8. All rows with the identified duplicates will be removed. To restore the full data view, simply reset the facets

**63 rows**

Extensions: Wikidata

Show as: **rows** records

Show: 5 10 25 50 rows

< first < previous **1 - 10** next > last >

All	Column	Rk	Pk	Tm	Player	College	Yrs	G	MP	PTS	TRB	AST	FG%	...	
1.	Facet					Text facet	sas	8	536	18929	10454	2371	1248	.446	.341
2.						Numeric facet	e	8	302	8464	4351	1682	605	.495	.324
3.						Timeline facet	sas	6	269	8221	6649	3022	832	.485	.331
4.						Scatterplot facet	ona	8	464	13232	5889	2939	1145	.451	.320
5.						Custom text facet...		6	245	4545	1389	441	515	.407	.305
						Custom Numeric Facet...									
						Customized facets									
						Word facet	631	1943	.375	.319					
						Duplicates facet	224	1534	.483	.344					
						Numeric log facet	88	513	.389	.353					
						1-bounded numeric log facet	88	513	.389	.353					
						Text length facet	742	257	.459	.308					
						Log of text length facet									
						Unicode char-code facet									
						Facet by error									
						Facet by null									
						Facet by empty string									
						Facet by blank (null or empty string)									

Source: library.illinois.edu

Facet / Filter Undo / Redo 2 / 2

Refresh

Reset All Remove All

 Column change invert reset

2 choices Sort by: name count

false 62

true 1

exclude

Facet by choice counts

## 1 matching rows (63 total)

Extensions: Wikidata

Show as: rows records Show: 5 10 25 50 rows

&lt; first &lt; previous 1 - 1 next &gt; last &gt;

All	Column	Rk	Pk	Tm	Player	College	Yrs	G	MP	PTS	TRB	AST	FG%	3P%
9.		8	8	SAC	Nik Stauskas	Michigan	5	335	6662	2272	688	513	.389	.353

Facet / Filter Undo / Redo 2 / 2

Refresh Reset All Remove All

**62 matching rows (63 total)**

Show as: **rows** records Show: 5 10 25 50 rows « first < previous **1 - 10** next > last »

All  Column  Rk  Pk  Tm  Player  College  Yrs  G  MP  PTS  TRB  AST  FG%  :

	Rk	Pk	Tm	Player	College	Yrs	G	MP	PTS	TRB	AST	FG%
1.	0	1	1	CLE	Andrew Wiggins	Kansas	8	536	18929	10454	2371	.446 .341
2.	1	2	2	MIL	Jabari Parker	Duke	8	302	8464	4351	1682	.495 .324
3.	2	3	3	PHI	Joel Embiid	Kansas	6	269	8221	6649	3022	.485 .331
4.	3	4	4	ORL	Aaron Gordon	Arizona	8	464	13232	5889	2939	.451 .320
5.	4	5	5	UTA	Dante Exum		6	245	4545	1389	441	.407 .305
6.	5	6	6	BOS	Marcus Smart	Oklahoma State	8	459	13540	4667	1631	.375 .319
7.	6	7	7	LAL	Julius Randle	Kentucky	8	458	14032	8007	4224	.483 .344
8.	7	8	8	SAC	Nik Stauskas	Michigan	5	335	6662	2272	688	.389 .353
10.	8	9	9	CHH	Noah Vonleh	Indiana	7	339	5683	1660	1742	.459 .308
11.	9	10	10	PHI	Elfrid Payton	Louisiana	8	455	12909	4920	1936	.450 .288

Source: library.illinois.edu

**OpenRefine nba dataset csv Permalink**

Facet / Filter Undo / Redo 0 / 5 Refresh Reset All Remove All

62 matching rows (63 total)

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

Extensions: Wikidata ▾

Rk Column Player College Yrs G MP PTS TRB AST FG% 3P% FT% MP2 PTS2 TRB2 AST2 WS WS/48 BPM

1. 0 1 Facet Text facet Numeric facet

2. 1 2 Edit cells Timeline facet

3. 2 3 Edit column Scatterplot facet

4. 3 4 Transpose Custom text facet...

5. 4 5 Sort... Customized facets

6. 5 6 View Word facet

7. 6 7 Reconcile Duplicates facet

8. 7 8 auskas Michigan 5 Numeric log facet

9. 7 8 10 10 CHI Noah Vonleh Indiana 5 1-bounded numeric log facet

10. 8 9 11 11 PHI Elfrid Payton Louisiana 5 Text length facet

11. 9 10 12 12 DEN Doug McDermott Creighton 5 Log of text length facet

12. 10 11 13 13 ORL Dario Šarić UCLA 5 Unicode char-code facet

13. 11 12 14 14 MIN Zach LaVine T.J. Warren NC State 5 Facet by error

14. 12 13 15 15 PHO Adreian Payne Michigan State 4 Facet by null

15. 13 14 16 16 ATL James Young Kentucky 4 Facet by empty string

16. 14 15 17 17 CHI Jusuf Nurkić Syracuse 4 Facet by blank (null or empty string)

17. 15 16 18 18 BOS Tyler Ennis Michigan State 5

18. 16 17 19 19 CHI Gary Harris Michigan State 5

19. 17 18 20 20 TOR Bruno Caboclo Michigan State 5

20. 18 19 21 21 OKC Mitch McGary Michigan State 5

21. 19 20 22 22 MEM Jordan Adams UCLA Michigan State 5

22. 20 21 23 23 UTA Rodney Hood Duke Michigan State 5

23. 21 22 24 24 CHH Shabazz Napier UConn Michigan State 5

24. 22 23 25 25 HOU Clint Capela Michigan State 5

25. 23 24 26 26 MIA P.J. Hairston UNC Michigan State 5

26. 24 25 27 27 PHO Bogdan Bogdanović Michigan State 5

27. 25 26 28 28 LAC C.J. Wilcox Washington Michigan State 5

28. 26 27 29 29 OKC Josh Huestis Stanford Michigan State 5

29. 27 28 30 30 SAS Kyle Anderson UCLA Michigan State 5

30. 28 29 31 31 Rk Damien Inglis Michigan State 5

31. 29 30 32 32 MIL Damion Inglis Michigan State 5

1. To remove blank rows, just use facet by blank.

# **Most Basic Data Cleaning and Wrangling Tasks**

- 1. Remove blanks**
- 2. Remove duplicates**
- 3. Resolve inconsistent entries**
- 4. Transform incorrect data (e.g. wrong data types such as text instead of numeric and calculations)**
- 5. Rename, combine and split columns for making further data presentation and processing easier**
- 6. Impute missing entries (need expertise to judge usage instead of relying on statistics alone)**

# **Analysing Cleaned Data in Python Using SQL**

**OpenRefine** nba run results csv [Permalink](#)

Facet / Filter Undo / Redo 0 / 0 [Extensions: Wikidata](#)

### 60 rows

Show as: rows records Show: 5 10 25 50 100 500 1000 rows [first](#) [previous](#) [1 of 1 page](#) [next](#) [last](#)

**Using facets and filters** 

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? [Watch these screencasts](#)

	Col_Rank	Col_Player	Col_College	Col_Yrs	Col_Game	Col_Minutes_Played	Col_Points	Col_Total_Rebounds	Col_Assists	Col_Field_Goal_Percentage	Col_Three_Pt_Percentage	Col_
1.	1	Andrew Wiggins	Kansas	8	598	20935	11519	2649	1393	0.448	0.350	0.723
2.	2	Jabari Parker	Duke	8	310	8535	4380	1703	610	0.494	0.326	0.743
3.	3	Joel Embiid	Kansas	6	328	10240	8535	3732	1080	0.490	0.338	0.810
4.	4	Aaron Gordon	Arizona	8	528	15277	6887	3310	1313	0.460	0.323	0.702
5.	5	Dante Exum	[REDACTED]		245	4545	1389	441	515	0.407	0.305	0.764
6.	6	Marcus Smart	Oklahoma State	8	520	15484	5438	1867	2318	0.382	0.321	0.780
7.	7	Julius Randle	Kentucky	8	518	16145	9191	4814	1842	0.472	0.332	0.743
8.	8	Nik Stauskas	Michigan	6	343	6701	2290	691	515	0.389	0.354	0.812
9.	9	Noah Vonleh	Indiana	7	339	5683	1660	1742	257	0.459	0.308	0.691
10.	10	Elfrid Payton	Louisiana	8	500	13383	5036	2007	2868	0.447	0.287	0.623
11.	11	Doug McDermott	Creighton	8	527	11012	4846	1236	496	0.476	0.409	0.820
12.	12	Dario Šarić	[REDACTED]		356	8965	4180	2086	699	0.441	0.357	0.838
13.	13	Zach LaVine	UCLA	8	478	15219	9466	1876	1860	0.461	0.386	0.830
14.	14	T.J. Warren	NC State	7	332	9553	5142	1372	397	0.507	0.357	0.780
15.	15	Adreian Payne	Michigan State	4	107	1403	429	315	66	0.406	0.254	0.680
16.	16	Jusuf Nurkić	[REDACTED]		411	9815	5015	3513	903	0.500	0.232	0.672
17.	17	James Young	Kentucky	4	95	812	219	96	28	0.367	0.277	0.563
18.	18	Tyler Ennis	Syracuse	4	186	2336	779	250	359	0.419	0.317	0.768
19.	19	Gary Harris	Michigan State	8	468	13473	5526	1162	965	0.448	0.363	0.812
20.	20	Bruno Caboclo	[REDACTED]		105	1293	442	268	72	0.403	0.308	0.836
21.	21	Mitch McGary	Michigan	2	52	557	227	183	17	0.527	0.000	0.580
22.	22	Jordan Adams	UCLA	2	32	263	101	30	19	0.402	0.385	0.607
23.	23	Rodney Hood	Duke	8	448	10977	4656	1160	736	0.420	0.366	0.841
24.	24	Shabazz Napier	UConn	6	345	5986	2433	657	849	0.397	0.345	0.815
25.	25	Clint Capela	[REDACTED]		471	12614	5851	5023	465	0.624	0.000	0.527
26.	26	P.J. Hairston	UNC	2	111	2000	664	266	59	0.343	0.295	0.810
27.	27	Bogdan Bogdanović	[REDACTED]		316	9041	4501	1083	1073	0.439	0.384	0.822
28.	28	C.J. Wilcox	Washington	3	66	376	132	31	30	0.370	0.333	0.813
29.	29	Josh Huestis	Stanford	3	76	1068	187	180	23	0.346	0.312	0.240
30.	30	Kyle Anderson	UCLA	8	505	10583	3366	2220	1167	0.476	0.334	0.711
31.	31	Damien Ingles	[REDACTED]		20	156	36	31	10	0.351	0.231	0.875
32.	32	K.J. McDaniels	Clemson	3	148	2092	782	324	95	0.412	0.290	0.776
33.	33	Joe Harris	Virginia	8	414	10601	4584	1344	679	0.483	0.439	0.780
34.	34	Cleanthony Early	Wichita State	2	56	801	241	123	42	0.346	0.263	0.750
35.	35	Jarnell Stokes	Tennessee	3	28	151	67	40	7	0.581		0.531

[OpenRefine](#) nba run results csv Permalink

Facet / Filter Undo / Redo 0 / 0 < 60 rows Show as: rows records Show: 5 10 25 50 100 500 1000 rows < first < previous 1 of 1 page next > last >

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? Watch these screencasts

Extensions: Wikidata ▾

	All	Col_Rank	Col_Player	Col_College	Col_Yrs	Col_Game	Col_Minutes_Played	Col_Points	Col_Total_Rebounds	Col_Assists	Col_Field_Goal_Percentage	Col_Three_Pt_Percentage	Col_F
1.	1	Andrew Wiggins	Facet	Text facet				11519	2649	1393	0.448	0.350	0.723
2.	2	Jabari Parker	Text filter	Numeric facet				4380	1703	610	0.494	0.326	0.743
3.	3	Joel Embiid	Edit cells	Timeline facet				8535	3732	1080	0.490	0.338	0.810
4.	4	Aaron Gordon	Edit column	Scatterplot facet				6887	3310	1313	0.460	0.323	0.702
5.	5	Dante Exum	Transpose	Custom text facet...				1389	441	515	0.407	0.305	0.764
6.	6	Marcus Smart	Sort...	Custom Numeric Facet...				5438	1867	2318	0.382	0.321	0.780
7.	7	Julius Randle	View	Customized facets				9191	4814	1842	0.472	0.332	0.743
8.	8	Nik Stauskas	Reconcile	Word facet				515	389			0.354	0.812
9.	9	Noah Vonleh		Duplicates facet				257	459			0.308	0.691
10.	10	Elfrid Payton		Numeric log facet				2868	447			0.287	0.623
11.	11	Doug McDermott	Creighton	1-bounded numeric log facet				496	476			0.409	0.820
12.	12	Dario Šarić		Text length facet				699	441			0.357	0.838
13.	13	Zach LaVine	UCLA	Log of text length facet				1860	461			0.386	0.830
14.	14	T.J. Warren	NC State	Unicode char-code facet				397	507			0.357	0.780
15.	15	Adreian Payne	Michigan State	Facet by error				66	406			0.254	0.680
16.	16	Jusuf Nurkić		Facet by null				903	500			0.232	0.672
17.	17	James Young	Kentucky	Facet by empty string				28	367			0.277	0.563
18.	18	Tyler Ennis	Syracuse	Facet by blank (null or empty string)				359	419			0.317	0.768
19.	19	Gary Harris	Michigan State					965	448			0.363	0.812
20.	20	Bruno Caboclo						72	403			0.308	0.836
21.	21	Mitch McGary	Michigan					17	527			0.000	0.580
22.	22	Jordan Adams	UCLA					19	402			0.385	0.607
23.	23	Rodney Hood	Duke					736	420			0.366	0.841
24.	24	Shabazz Napier	UConn					849	397			0.345	0.815
25.	25	Clint Capela						4656	1160			0.000	0.527
26.	26	P.J. Hairston	UNC					2433	657			0.295	0.810
27.	27	Bogdan Bogdanović						4501	1083			0.384	0.822
28.	28	C.J. Wilcox	Washington					5851	5023			0.624	0.813
29.	29	Josh Huestis	Stanford					132	31			0.333	0.240
30.	30	Kyle Anderson	UCLA					187	180			0.312	0.711
31.	31	Damien Inglis						3366	2220			0.476	0.711
32.	32	K.J. McDaniels	Clemson					1167	1083			0.334	0.875
33.	33	Joe Harris	Virginia					782	324			0.412	0.776
34.	34	Cleanthony Early	Wichita State					4584	1344			0.483	0.780
35.	35	Jarnell Stokes	Tennessee					241	123			0.346	0.750
								67	40			0.581	0.531

javascript:{}

**OpenRefine** nba run results csv Permalink

Facet / Filter Undo / Redo 0 / 0 Refresh Reset All Remove All

Extensions: Wikidata ▾ Open... Export ▾ Help

**14 matching rows** (60 total)

Show as: rows records Show: 5 10 25 50 100 500 1000 rows « first < previous 1 of 1 page next » last »

	Col_Rank	Col_Player	Col_College	Col_Yrs	Col_Game	Col_Minutes_Played	Col_Points	Col_Total_Rebounds	Col_Assists	Col_Field_Goal_Percentage	Col_Three_Pt_Percentage	Col_F
5. 5	Dante Exum			6	245	4545	1389	441	515	0.407	0.305	0.764
12. 12	Dario Šarić			5	356	8965	4180	2086	699	0.441	0.357	0.838
16. 16	Jusuf Nurkić			8	411	9815	5015	3513	903	0.500	0.232	0.672
20. 20	Bruno Caboclo			7	105	1293	442	268	72	0.403	0.308	0.836
25. 25	Clint Capela			8	471	12614	5851	5023	465	0.624	0.000	0.527
27. 27	Bogdan Bogdanović			5	316	9041	4501	1083	1073	0.439	0.384	0.822
31. 31	Damien Inglis			1	20	156	36	31	10	0.351	0.231	0.875
41. 41	Nikola Jokić			7	527	16018	10364	5456	3281	0.542	0.345	0.830
43. 43	Edy Tavares			2	13	101	33	32	4	0.625		0.273
51. 51	Thanasis Antetokounmpo			4	127	1159	398	250	83	0.517		0.545
52. 52	Vasilije Micić											
53. 53	Alessandro Gentile											
54. 54	Nemanja Dangubić											
57. 57	Louis Labeyrie											

javascript:{}

**OpenRefine nba run results csv** [Permalink](#)

Facet / Filter Undo / Redo 0 / 0 < Extensions: Wikidata ▾

Refresh Reset All Remove All

**46 matching rows (60 total)**

Show as: rows records Show: 5 10 25 50 100 500 1000 rows « first < previous 1 of 1 page next » last »

**Col\_College** change invert reset

2 choices Sort by: name count

false 46 exclude

true 14

Facet by choice counts

All Col\_Rank Col\_Player Col\_College Col\_Yrs Col\_Game Col\_Minutes\_Played Col\_Points Col\_Total\_Rebounds Col\_Assists Col\_Field\_Goal\_Percentage Col\_Three\_Pt\_Percentage Col\_F

			Kansas	8	598	20935	11519	2649	1393	0.448	0.350	0.723
			Duke	8	310	8535	4380	1703	610	0.494	0.326	0.743
			Kansas	6	328	10240	8535	3732	1080	0.490	0.338	0.810
			Arizona	8	528	15277	6887	3310	1313	0.460	0.323	0.702
			Oklahoma State	8	520	15484	5438	1867	2318	0.382	0.321	0.780
			Kentucky	8	518	16145	9191	4814	1842	0.472	0.332	0.743
			Michigan	6	343	6701	2290	691	515	0.389	0.354	0.812
			Indiana	7	339	5683	1660	1742	257	0.459	0.308	0.691
			Louisiana	8	500	13383	5036	2007	2868	0.447	0.287	0.623
			Creighton	8	527	11012	4846	1236	496	0.476	0.409	0.820
			UCLA	8	478	15219	9466	1876	1860	0.461	0.386	0.830
			NC State	7	332	9553	5142	1372	397	0.507	0.357	0.780
			Michigan State	4	107	1403	429	315	66	0.406	0.254	0.680
			Kentucky	4	95	812	219	96	28	0.367	0.277	0.563
			Syracuse	4	186	2336	779	250	359	0.419	0.317	0.768
			Michigan State	8	468	13473	5526	1162	965	0.448	0.363	0.812
			Michigan	2	52	557	227	183	17	0.527	0.000	0.580
			UCLA	2	32	263	101	30	19	0.402	0.385	0.607
			Duke	8	448	10977	4656	1160	736	0.420	0.366	0.841
			UConn	6	345	5986	2433	657	849	0.397	0.345	0.815
			UNC	2	111	2000	664	266	59	0.343	0.295	0.810
			Washington	3	66	376	132	31	30	0.370	0.333	0.813
			Stanford	3	76	1068	187	180	23	0.346	0.312	0.240
			UCLA	8	505	10583	3366	2220	1167	0.476	0.334	0.711
			Clemson	3	148	2092	782	324	95	0.412	0.290	0.776
			Virginia	8	414	10601	4584	1344	679	0.483	0.439	0.780
			Wichita State	2	56	801	241	123	42	0.346	0.263	0.750
			Tennessee	3	28	151	67	40	7	0.581		0.531
			LSU	4	147	1684	509	352	69	0.402	0.360	0.663
			UConn									
			Colorado	8	387	10206	5042	1167	1933	0.411	0.322	0.791
			Syracuse	8	555	14454	6329	2189	747	0.452	0.349	0.723
			Michigan	7	304	5299	1804	776	238	0.457	0.373	0.779
			Nick Johnson	1	28	262	74	39	11	0.347	0.238	0.680

Filter Undo / Redo 0 / 0 Reset All Remove All

**46 matching rows (60 total)**

Show as: rows records Show: 5 10 25 50 100 500 1000 rows « first

_College	Col_Rank	Col_Player	Col_College	Col_Yrs	Col_Game	Col_Minutes_Played	Col_Points	Col_Total_Rebounds	Col_Assists	Col_Field_Go	Col_F
	1. 1	Andrew Wiggins	Kansas	8	598	20935	11519	2649	1393	0.448	
	2. 2	Jabari Parker	Duke	8	310	8535	4380	1703	610	0.494	
	3. 3	Joel Embiid	Kansas	6	328	10240	8535	3732	1080	0.490	
	4. 4	Aaron Gordon	Arizona	8	528	15277	6887	3310	1313	0.460	
	6. 6	Marcus Smart	Oklahoma State	8	520	15484	5438	1867	2318	0.382	
	7. 7	Julius Randle	Kentucky	8	518	16145	9191	4814	1842	0.472	
	8. 8	Nik Stauskas	Michigan	6	343	6701	2290	691	515	0.389	
	9. 9	Noah Vonleh	Indiana	7	339	5683	1680	1742	257	0.459	
	10. 10	Elfrid Payton	Louisiana	8	500	13383	5036	2007	2868	0.447	
	11. 11	Doug McDermott	Creighton	8	527	11012	4846	1236	496	0.476	
	13. 13	Zach LaVine	UCLA	8	478	15219	9466	1876	1860	0.461	
	14. 14	T.J. Warren	NC State	7	332	9553	5142	1372	397	0.507	
	15. 15	Adreian Payne	Michigan State	4	107	1403	429	315	66	0.406	
	17. 17	James Young	Kentucky	4	95	812	219	96	28	0.367	
	18. 18	Tyler Ennis	Syracuse	4	186	2336	779	250	359	0.419	0.317 0.768
	19. 19	Gary Harris	Michigan State	8	468	13473	5526	1162	965	0.448	0.363 0.812
	21. 21	Mitch McGary	Michigan	2	52	557	227	183	17	0.527	0.000 0.580
	22. 22	Jordan Adams	UCLA	2	32	263	101	30	19	0.402	0.385 0.607
	23. 23	Rodney Hood	Duke	8	448	10977	4656	1160	736	0.420	0.366 0.841
	24. 24	Shabazz Napier	UConn	6	345	5986	2433	657	849	0.397	0.345 0.815
	26. 26	P.J. Hairston	UNC	2	111	2000	664	266	59	0.343	0.295 0.810
	28. 28	C.J. Wilcox	Washington	3	66	376	132	31	30	0.370	0.333 0.813
	29. 29	Josh Huestis	Stanford	3	76	1068	187	180	23	0.346	0.312 0.240
	30. 30	Kyle Anderson	UCLA	8	505	10583	3366	2220	1167	0.476	0.334 0.711
	32. 32	K.J. McDaniels	Clemson	3	148	2092	782	324	95	0.412	0.290 0.776
	33. 33	Joe Harris	Virginia	8	414	10601	4584	1344	679	0.483	0.439 0.780
	34. 34	Cleanthony Early	Wichita State	2	56	801	241	123	42	0.346	0.263 0.750
	35. 35	Jarnell Stokes	Tennessee	3	28	151	67	40	7	0.581	
	36. 36	Johnny O'Bryant	LSU	4	147	1684	509	352	69	0.402	0.360 0.663
	37. 37	DeAndre Daniels	UConn								
	38. 38	Spencer Dinwiddie	Colorado	8	387	10206	5042	1167	1933	0.411	0.322 0.791
	39. 39	Jerami Grant	Syracuse	8	555	14454	6329	2189	747	0.452	0.349 0.723
	40. 40	Glenn Robinson III	Michigan	7	304	5299	1804	776	238	0.457	0.373 0.779
	42. 42	Nickeil Alexander-Walker	Arizona	1	26	260	74	26	11	0.247	0.226 0.600

```
In [3]: 1 import pandas as pd
2 import sqlite3
3 con = sqlite3.connect('nba_db.db')
4 sql = "SELECT * FROM nba run results WHERE Col Game > 400"
5 df = pd.read_sql_query(sql,con)
6 con.close()
7 df
```

Out[3]:

Col_Rank	Col_Player	Col_College	Col_Yrs	Col_Game	Col_Minutes_Played	Col_Points	Col_Total_Rebounds	Col_Assists	Col_Field_Goal_Percentage	Col_T
0	1	Andrew Wiggins	Kansas	8	598	20935	11519	2649	1393	0.448
1	4	Aaron Gordon	Arizona	8	528	15277	6887	3310	1313	0.460
2	6	Marcus Smart	Oklahoma State	8	520	15484	5438	1867	2318	0.382
3	7	Julius Randle	Kentucky	8	518	16145	9191	4814	1842	0.472
4	10	Elfrid Payton	Louisiana	8	500	13383	5036	2007	2868	0.447
5	11	Doug McDermott	Creighton	8	527	11012	4846	1236	496	0.476
6	13	Zach LaVine	UCLA	8	478	15219	9466	1876	1860	0.461
7	19	Gary Harris	Michigan State	8	468	13473	5526	1162	965	0.448
8	23	Rodney Hood	Duke	8	448	10977	4656	1160	736	0.420
9	30	Kyle Anderson	UCLA	8	505	10583	3366	2220	1167	0.476
10	33	Joe Harris	Virginia	8	414	10601	4584	1344	679	0.483
11	39	Jerami Grant	Syracuse	8	555	14454	6329	2189	747	0.452
12	45	Dwight Powell	Stanford	8	511	9720	3926	2349	524	0.585
13	46	Jordan Clarkson	Missouri	8	600	16191	9216	1971	1526	0.439

```
1 import pandas as pd
2 import sqlite3
3 con = sqlite3.connect('nba_db.db')
4 sql = SELECT Col_College, sum(Col_Game) AS [Total # of Games] FROM nba_run_results GROUP BY Col_College ORDER BY |
5 df = pd.read_sql_query(sql,con)
6 con.close()
7 df
```

	Col_College	Total # of Games
0	UCLA	1015.0
1	Kansas	926.0
2	Duke	758.0
3	Syracuse	741.0
4	Michigan	699.0
5	Oklahoma State	633.0
6	Kentucky	613.0
7	Missouri	600.0
8	Stanford	587.0
9	Michigan State	575.0
10	Arizona	556.0
11	Creighton	527.0
12	Louisiana	500.0
13	Virginia	414.0
14	Colorado	387.0
15	UConn	345.0
16	Indiana	339.0
17	NC State	332.0
18	Tennessee	151.0
19	Clemson	148.0
20	LSU	147.0

```

1 # CAST(ROUND(Column_Name, 2) AS DECIMAL(10,2)
2 import pandas as pd
3 import sqlite3
4 con = sqlite3.connect('nba_db.db')
5 # sql = "SELECT Col_Player, Col_College, Col_Game, Col_Yrs, ROUND((Col_Game*1.0/Col_Yrs),2) AS [Games/Years], ROUND(CAST(Col_Game AS FLOAT)/Col_Yrs,2) AS [Avg_Points]"
6 sql = "SELECT Col_Player, Col_College, Col_Game, Col_Yrs, ROUND(CAST(Col_Game AS FLOAT)/Col_Yrs,2) AS [Games/Years]"
7 df = pd.read_sql_query(sql,con)
8 con.close()
9 df

```

	Col_Player	Col_College	Col_Game	Col_Yrs	Games/Years	Avg Poits
0	Zach LaVine	UCLA	478.0	8.0	59.75	158.43
1	Joel Embiid	Kansas	328.0	6.0	54.67	156.13
2	Andrew Wiggins	Kansas	598.0	8.0	74.75	154.10
3	Julius Randle	Kentucky	518.0	8.0	64.75	141.95
4	Jordan Clarkson	Missouri	600.0	8.0	75.00	122.88
5	Jabari Parker	Duke	310.0	8.0	38.75	113.03
6	T.J. Warren	NC State	332.0	7.0	47.43	108.42
7	Aaron Gordon	Arizona	528.0	8.0	66.00	104.35
8	Spencer Dinwiddie	Colorado	387.0	8.0	48.38	104.23
9	Gary Harris	Michigan State	468.0	8.0	58.50	94.46
10	Jerami Grant	Syracuse	555.0	8.0	69.38	91.23
11	Joe Harris	Virginia	414.0	8.0	51.75	88.58
12	Marcus Smart	Oklahoma State	520.0	8.0	65.00	83.66
13	Rodney Hood	Duke	448.0	8.0	56.00	83.14
14	Elfrid Payton	Louisiana	500.0	8.0	62.50	80.58
15	Doug McDermott	Creighton	527.0	8.0	65.88	73.56
16	Dwight Powell	Stanford	511.0	8.0	63.88	61.46
17	Kyle Anderson	UCLA	505.0	8.0	63.13	53.32
18	Shabazz Napier	UConn	345.0	6.0	57.50	42.31
19	Glenn Robinson III	Michigan	304.0	7.0	43.43	41.54
20	Nik Stauskas	Michigan	343.0	6.0	57.17	40.06

## **Problem Set #2**

- 1. Pick a district in Open Rice for performing data scraping in ParseHub.**
- 2. Export the CSV file from ParseHub and import it into OpenRefine for cleaning.**
- 3. Import the CSV file into SQLite for SQL query analysis.**
- 4. Connect to the SQLite database containing food data and performing analysis using Python codes combined with SQL queries to present the result in a Pandas data frame.**
- 5. Write up your analysis and place the summary into your assignment #2 web page (Just the analysis result — not reflection).**
- 6. Provide a link from your journal and reflect on your learning.**