



Programlama olmadan Veri kazıma

Pınar Dağ

İçerik



- Kazıma Nedir?
- Gazeteciler veriye nasıl erişiyor?

Kazıma nedir?



Kazıma demek...

Web'den veri toplamak, araştırmak, bir araya getirmek.

Veriyi kazıyanlara ise çoğunlukla:

Bir siteden veri
alıyorlarsa:

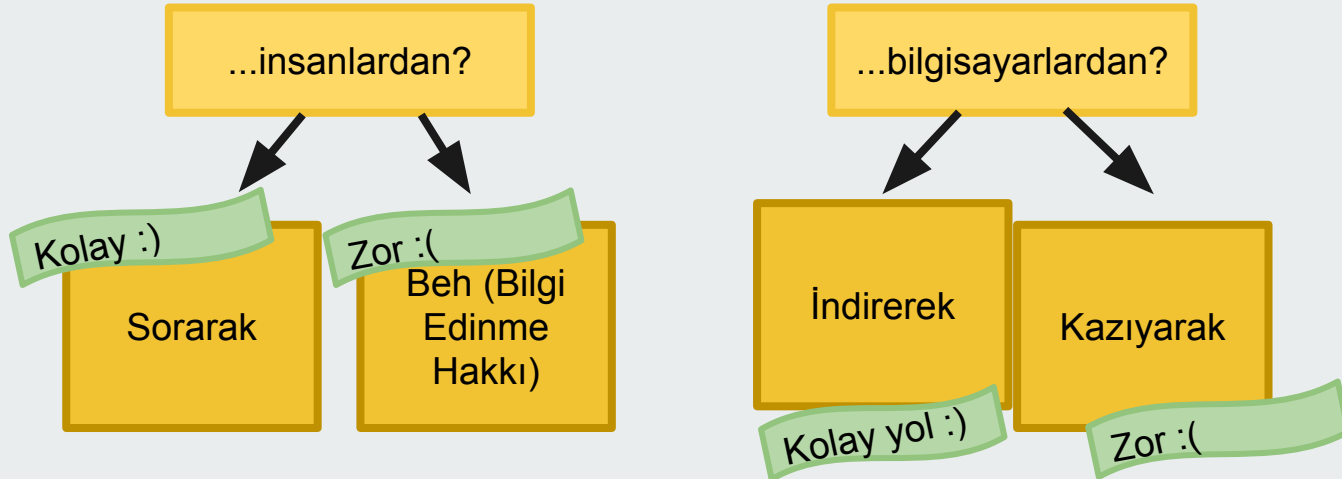
Kazıyıcı

Birden fazla siteden
veri alıyorlarsa:

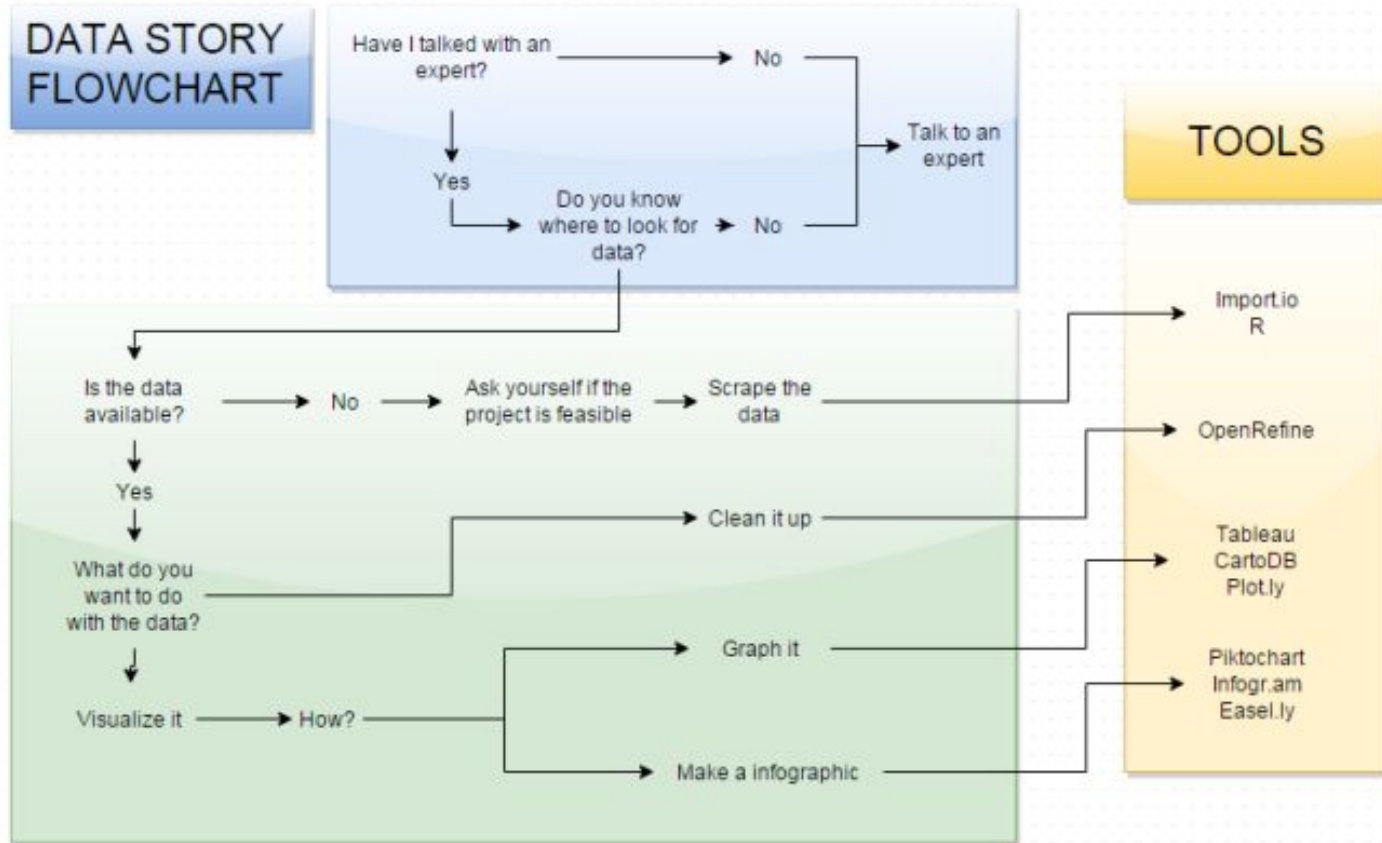
**Tarayıcı/
Böcek**

..denir.

Gazeteciler veriye nasıl erişiyor?



DATA STORY FLOWCHART





Temel HTML (Hiper Metin İşaretleme Dili)

bilgisine ihtiyacınız olacak

HTML, bilgisayarlar içindir, ancak bunların bir kısmı insanlar tarafından da anlaşılır.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.o../xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<title>Azerbaijan</title>
  <meta name="KEYWORDS" content="Azerbaijan, Azərbaycan, Heydar Aliyev Foundation, Heyda
Aliyev, president, Azerbaijan news, Karabakh, Nagorny Karabakh Confilict, Baku, weather
currency, Politics, Society, Culture, Sport, Diaspora, Official chronicle">
  <link href="/css/global.css" rel="stylesheet" type="text/css" />
  <link href="/css/menu.css" rel="stylesheet" media="screen">
  <link href="/css/section.css" rel="stylesheet" media="screen">
```

HTML

Tarayıcı /Browser

Number of female to 1000 male by age group							
	2008	2009	2010	2011	2012	2013	2014
Всего населения	1025	1021	1019	1017	1015	1013	1011
of which:							
0-4	875	870	867	863	860	859	857
5-9	886	878	874	874	876	875	873
10-14	920	913	904	897	888	885	883
15-19	984	973	955	938	931	921	919
20-24	1028	1020	1018	1013	1000	988	986
25-29	1009	1014	1020	1024	1030	1029	1027
30-34	1033	1021	1012	1008	1005	1009	1007
35-39	1076	1070	1068	1060	1053	1038	1036
40-44	1087	1092	1088	1085	1081	1080	1078
45-49	1058	1067	1080	1092	1098	1099	1097
50-54	1070	1073	1071	1067	1071	1078	1076
55-59	1113	1107	1105	1106	1107	1102	1101
60-64	1177	1182	1173	1167	1161	1161	1158
65-69	1264	1262	1270	1269	1259	1251	1251
70 -	1440	1426	1425	1427	1435	1450	1447

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/xhtml/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<title>Azerbaijan</title>
<meta name="KEYWORDS" content="Azerbaijan, Azərbaycan, Heydar Aliyev Foundation, Heydar
Aliyev, president, Azerbaijan news, Karabakh, Nagorny Karabakh Conflict, Baku, weather,
currency, Politics, Society, Culture, Sport, Diaspora, Official chronicle">
<link href="/css/global.css" rel="stylesheet" type="text/css" />
<link href="/css/menu.css" rel="stylesheet" media="screen">
<link href="/css/section.css" rel="stylesheet" media="screen">
```

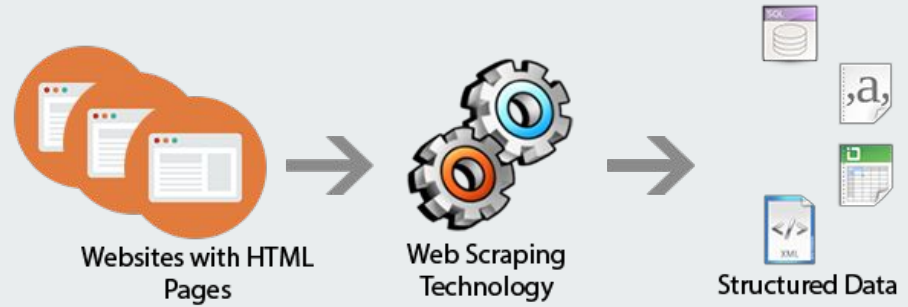
http://azerbaijan.az/portal/General/Population/population_e.html

Görmek için sağ tıklamalısınız.



DATA MINER İLE UYGULAMA

DATA MINER

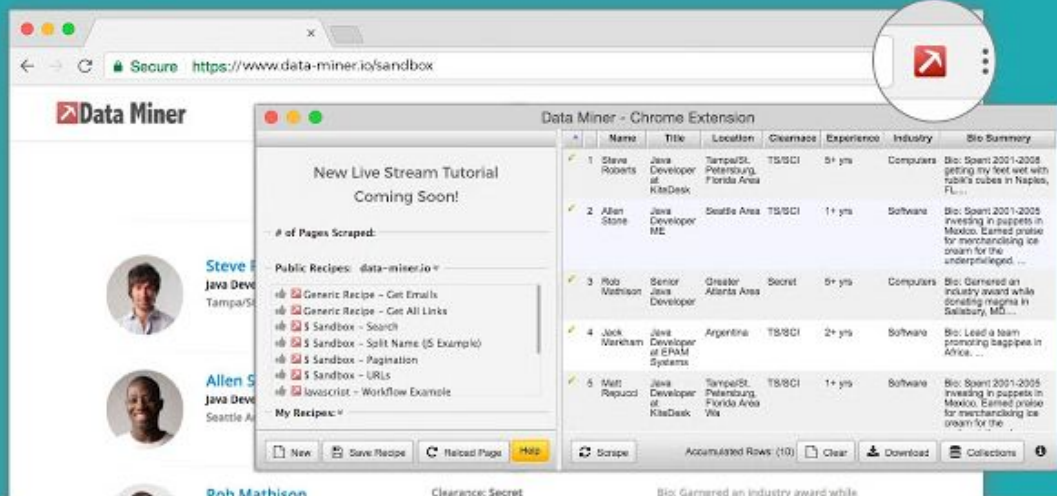


Eklentiyi kurunuz:

<https://chrome.google.com/webstore/detail/data-scraper-easy-web-scr/nndkne-pjnlbdbepjfgmncbggmopgden>

Instant Results

Extract Data From any Website Into Excel
in Seconds With Premade Extraction Recipes



The screenshot displays the Data Miner Chrome extension interface. The main window shows a list of extracted data from a website, with columns for Name, Title, Location, Clearance, Experience, Industry, and Bio Summary. The data is presented in a table format, with each row representing a person's profile. The interface also includes a sidebar with navigation options like 'Public Recipes', 'Generic Recipe', 'Sandbox', and 'My Recipes'. A 'New Live Stream Tutorial Coming Soon!' message is visible in the top left of the extension window. The background shows a web browser window with the URL 'https://www.data-miner.io/sandbox'.

	Name	Title	Location	Clearance	Experience	Industry	Bio Summary
1	Steve Roberts	Java Developer at KiteDesk	Tampa/St. Petersburg, Florida Area	TS/SCI	5+ yrs	Computers	Bio: Spent 2001-2005 getting my feet wet with rubik's cubes in Naples, FL...
2	Allen Stone	Java Developer at KiteDesk	Seattle Area	TS/SCI	1+ yrs	Software	Bio: Spent 2001-2005 investing in puppets in Mexico. Earned praise for merchandising ice cream for the underprivileged...
3	Roh Mathison	Senior Java Developer	Greater Atlanta Area	Secret	5+ yrs	Computers	Bio: Garnered an industry award while scolding magpies in Salisbury, MD...
4	Jack Warkham	Java Developer at EPAM Systems	Argentina	TS/SCI	2+ yrs	Software	Bio: Lead a team promoting bagpipes in Africa...
5	Matt Repucci	Java Developer at KiteDesk	Tampa/St. Petersburg, Florida Area	TS/SCI	1+ yrs	Software	Bio: Spent 2001-2005 investing in puppets in Mexico. Earned praise for merchandising ice cream for the...

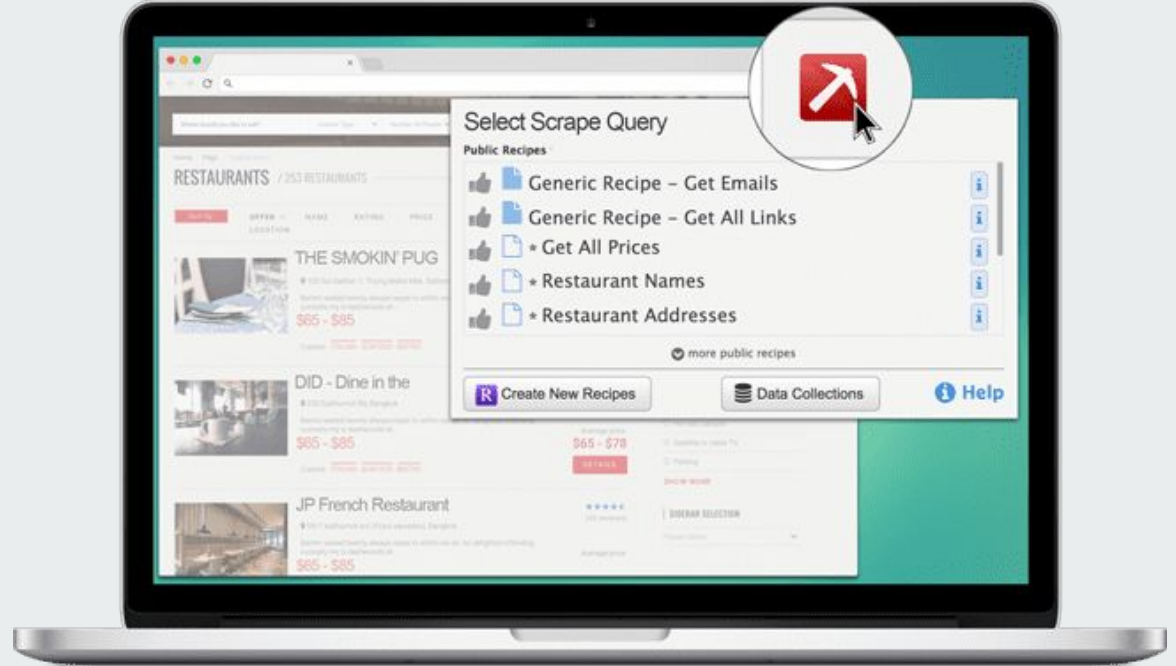


Data Miner Google Chrome eklentisi. Web sitelerinden veri kazıma ve kazınan verinin CSV dosyası ya da Excel, e-tabloya aktarılmasını sağlar.



Koda ihtiyacınız yok!

- 50,000+ üstü ücretsiz ön sorgu yapabilirsiniz.
- 15,000 + üstü popüler site için kullanılabilir.



Kaygı duymadan kazıma yapabilirsiniz

- Data Miner bot deęil
- Botlar site tarafından bloklanır.
- Data Miner kendi tarayıcınızı tıklar gibi davranır, alıřır.



Verinizi Gizli Tutar



- Data Miner verinizi satmaz
- Data Miner verinizi biriyle paylaşmaz
- Veriniz koruma altındadır ve güvendedir.
- Data Miner, gizliliğinizi ve güvenliğinizi korumak için HTTPS şifrelemesini kullanır.



Data Miner is the most powerful scraper around



One Click Scraping

Use one of 50,000 publicly available extraction queries to extract data with one click.



Custom Scraping

Make custom extraction queries to scrape any data from any site.



Automate Scrapes

Run bulk scrape jobs base off a list of URLs.



Fastest Table Scrapes

Extract basic table data with right click.



Pagination

Automatically click to the next page and scrape using Auto



Form Filling Automation

Data Miner can automatically fill forms for you using a pre-



Recipes by Other Users

Recipes Bookmarked

Recipes Made by You

Public Bookmarks My recipes

Search [Q] + New Recipe

Recipe Name

Data Extracted

* URL List

Columns (1) : Restaurant URL

Generic Recipe - Get Emails

Columns (1) : email

Page Details (Restaurant)

Columns (14) : Business Name | phone | website | Address

Form Filling

Columns (12) : Business Name | Address | City, State Zip | Neighborhood | Phone

Search Results

Columns (12) : Business Name | Address | City, State Zip | Neighborhood | Phone

Recipe Creator

Run Recipe

Bookmark

Edit Recipe

Delete Recipe

Recipe Indicators:
Actions
Javascript
Next Page

Data Miner Website → **View Account** → **Import CSV** → **Data Collections** → **Jobs**

Pagination → **Download Data** → **Help From a Human** → **Clear Scrape**

Data Miner

Public Bookmarks My recipes

Search [] Q + New Recipe

Generic Recipe - Get All Links
Columns (2) : URL, Link

Generic Recipe - Get Emails
Columns (1) : email

*** Biz Page Details (Generic)**
Columns (16) : Business Name, Rating, Reviews, Price Range, phone, website, Categories

*** Form Filling**
Columns (12) : Business Name, Address, City, State Zip, Neighborhood, Phone, Category 1

Details (Restaurant)
Columns (8) : Business Name, Link, Address, City/State/Zip, Neighborhood, Phone, Category 1

Get All Links
Columns (2) : URL, Link

*** Form Filling - Get Emails**
Columns (1) : email

*** Biz Page Details**
Columns (16) : Business Name, Rating, Reviews, Price Range, phone, website, Categories

*** Form Filling**
Columns (12) : Business Name, Address, City, State Zip, Neighborhood, Phone, Category 1

(Restaurant)
Columns (8) : Business Name, Link, Address, City/State/Zip, Neighborhood, Phone, Category 1
Found 2 public recipes for: demo.dataminer.io

Recipe Details:
Name: Restaurant Scrape

Next Page

Download

Live Support

Clear Scrape

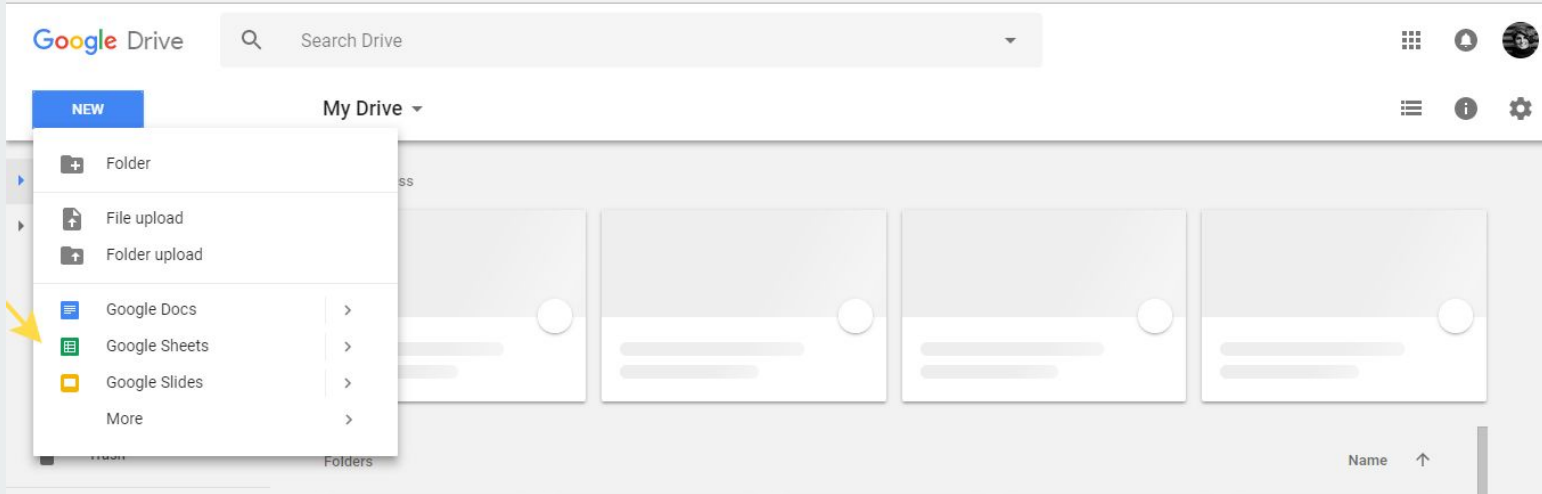
Name	Cuisine Type	Average Price	Reviews
Radiant Lounge Details	Italian	\$65 - \$78	(65 reviews)
Torque Details	Italian	\$65 - \$78	(143 reviews)
Café Cosmo Diner Details	Italian	\$65 - \$78	(65 reviews)
La Pelé Bistro Spot Details	Italian	\$65 - \$78	(27 reviews)
The Dining Car Café Details	Italian	\$65 - \$78	(27 reviews)
Velveteen Restaurant Details	Italian	\$65 - \$78	(43 reviews)
Radiant Lounge Details	Italian	\$65 - \$78	(65 reviews)
Torque Details	Italian	\$65 - \$78	(143 reviews)
Café Cosmo Diner Details	Italian	\$65 - \$78	(65 reviews)
Café Octopus Garden Tap D...	Italian	\$65 - \$78	(27 reviews)
Epicuria Café Details	Italian	\$65 - \$78	(27 reviews)
Oulu French Details	Italian	\$65 - \$78	(43 reviews)
Le le Wagon Details	Italian	\$65 - \$78	(65 reviews)
Tivondra Grill Details	Italian	\$65 - \$78	(143 reviews)
Le Budlatia Eru Details	Italian	\$65 - \$78	(65 reviews)
Coffee Patch Bistro Details	Italian	\$65 - \$78	(27 reviews)
Corks Cantina Details	Italian	\$65 - \$78	(27 reviews)
Olympio Bistro Details	Italian	\$65 - \$78	(43 reviews)
Angel Walk Grill Details	Italian	\$65 - \$78	(65 reviews)
Tibisi Details	Italian	\$65 - \$78	(143 reviews)
Floq Fish Chips Details	Italian	\$65 - \$78	(65 reviews)

Zorlanıyorsanız video izleyiniz



Uygulama

Pınar'ın çalışma dosyası



Tıklayınız



- 1)https://akom.ibb.istanbul/Akomas/Akom_Sorgu.aspx?sp=1&au=-1&as=-1&ai=-1&at=1&t1=01.01.2015&t2=28.07.2018
- 2)<http://www.ibb.gov.tr/tr-TR/Pages/Ihalelanlari.aspx?&gun1=&gun2=&ay1=&ay2=&yil1=&yil2=&kelime=&birim=&tur=ARSA%20SATI%C5%9EI&durum=&sr=>
- 2)<http://icisleri.gov.ct.tr/Bakanl%C4%B1k/Belediyeler>
- 3)<http://icisleri.gov.ct.tr/KURUL-KOM%C4%B0SYON-VE-B%C4%B0R%C4%B0MLER/Ta%C5%9F%C4%B1nmaz-Mal-Birimi>
- 4)https://www.tbmm.gov.tr/develop/owa/milletvekillerimiz_sd.liste
- 5)<http://www.cm.gov.nc.tr/Milletvekillerimiz1.aspx>
- 6)<http://www.maliye.gov.ct.tr/maliye/f?p=100:1105:0::NO::>
- 7)<http://www.mebnet.net/?q=node/7>

Pınar'ın çalışma dosyası



Farklı konu kazıma

<http://vdc-sy.net/en/>

https://docs.google.com/spreadsheets/d/1Ns65lY5pB_AoRojoG_ZCVLtqSC4aUwnGqMYiw4K6Lec/edit?usp=drivesdk

Pınar'ın çalışma dosyası

Filter by

Province Sex Marital Status Status Display

Start Date End Date This Date

Cause of Death

- ☒ Explosion ☐ Shelling ☐ Field Execution ☐ Shooting ☐ Kidnapping - Execution ☐ Kidnapping - Torture ☐ Kidnapping - Torture - Execution
☐ Detention - Execution ☐ Detention - Torture ☐ Detention - Torture - Execution ☐ Un-allowed to seek Medical help ☐ Warplane shelling
☐ Other ☐ Chemical and toxic gases ☐ Siege ☐ Unknown

Actors

Name Mother's Name Area \ Place of birth Occupation Notes

Martyrdom location Age Rank / Military situation Type of Weapon

Submit

Latest Killed (23)

Beginning

Name



Status



Sex



Province

Area \ Place
of birthDate of
deathCause of
Death


Actors

Kaynak: <http://vdc-sy.net/en/>



FARKLI VERİ KAZIMA

<https://www.barobirlik.org.tr/Haberler/avukat-sayilari-31122016-75022>



***Lütfen siz de e-tablo oluşturunuz ve tüm kazımaları oraya ekleyiniz ve Slack üzerinden paylaşınız**

<https://datajournalismteam.slack.com/messages/C2KG07KJP/details/>