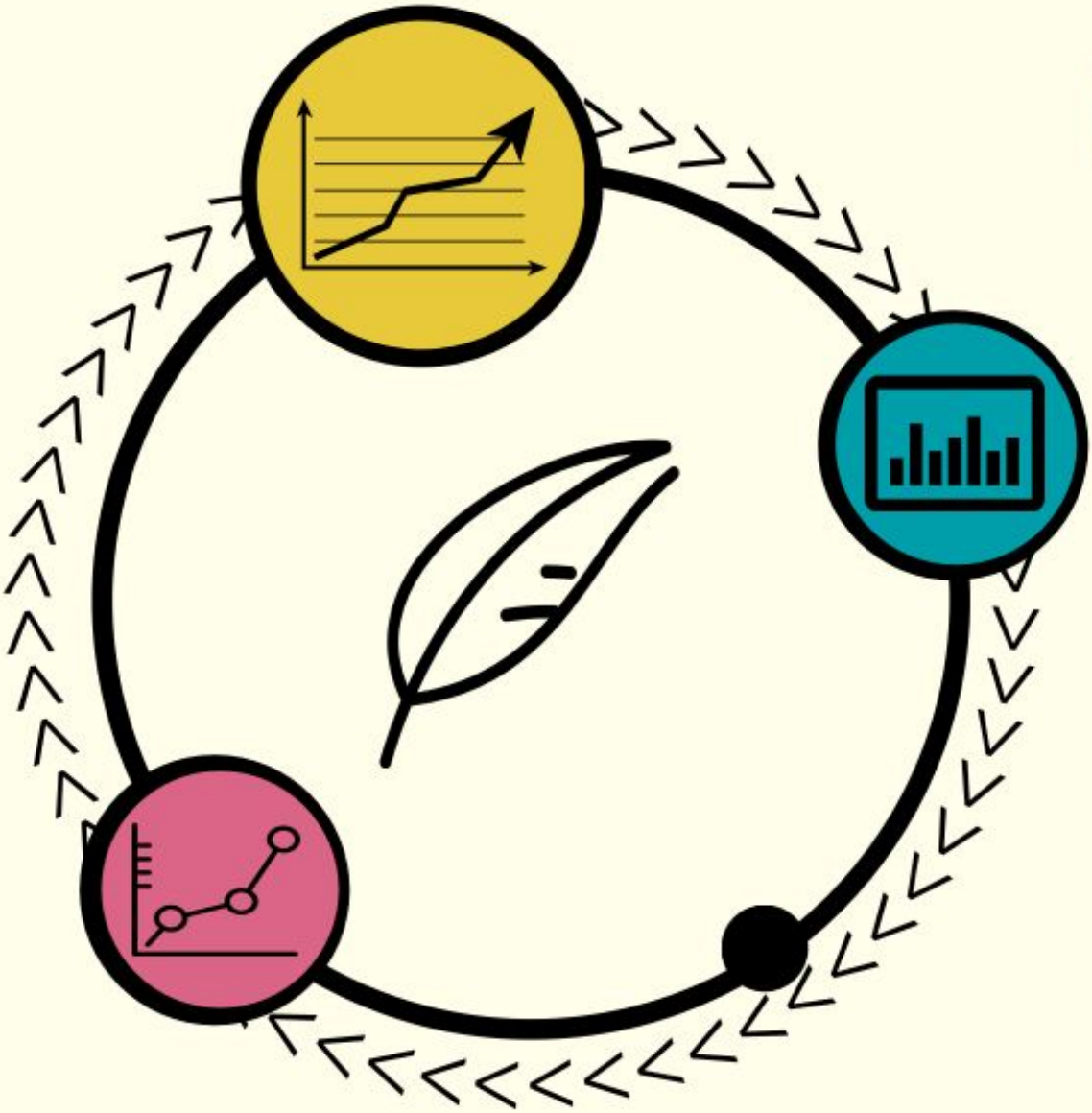
# История с данни



Институт за развитие на публичната среда

**Съдържание:**


Вместо въведение

Могат ли данните да разкажат позната история по нов начин?

Къде да открием данните?

Как може да се разкаже „история с данни”?

Инструменти за създаване на “история с данни”



“История с данни” е част от проект “Когато данните срещнат гражданите”, изпълняван от Институт за развитие на публичната среда, осъществен с финансовата подкрепа на "Програмата за подкрепа на неправителствени организации в България по Финансовия механизъм на Европейското икономическо пространство".



Цялата отговорност за съдържанието на документа се носи от сдружение “Институт за развитие на публичната среда” и при никакви обстоятелства не може да се приема, че този документ отразява официалното становище на Финансовия механизъм на Европейското икономическо пространство и Оператора на Програмата за подкрепа на неправителствени организации в България.

# Вместо въведение

Какво представляват "историите с данни"? Може би разкази, които се базират на данните? Това ли обаче е пълният отговор на въпроса, защото и „данните", и „историите", не са еднозначни понятия.

Какво може да направи една история, изпълнена с данни, по-интересна и полезна за публиката?

Настоящият текст ще се опита да представи примери, които да послужат за лесен и бърз анализ и визуализация на данни при тяхното използване за успешно провеждане на застъпнически кампании, усилване на прозрачността на работата на институциите, ангажираността на гражданите, ефективността в избора на начин за реализация на политики. Той ще очертае някои методи за визуализация на данните и представянето им по достъпен и разбираем начин.

Надяваме се "История с данни" да подпомогне дейността на гражданските активисти, НПО и заинтересованите в усилията им да придобият умения за изследване и използване на публични данни за застъпничество и да усилят посланията си към публиката.

## Данните

Могат ли данните да спасят света? Сами по себе си, едва ли. В световен мащаб е налице тенденция на прозрачност, която по всичко личи, ще засегне и различните бизнес сфери, както и правителствения сектор.

Потокът от данни може да бъде анализиран и използван от медии, неправителствени организации, академични среди и граждански активисти, които да се насочат към търсене на най-необходимата и вярна информация, за да я представят в най-кратък срок и в най-разбираема форма.

В тяхна подкрепа могат да бъдат „историите с данни" и отвореното управление.

„Историите с данни" например могат да свързват най-различни теми от обществения живот. Да ги комбинират и да представят темите по качествено нов начин - големи обеми от информация стават лесно

разбираеми и достъпни за широк кръг ползватели, а често пъти данните показват смущаващи тенденции. Именно взаимовръзката, която „историята с данни“ може да направи между различните източници на информация, е една от безценните й страни.

“Историята с данни” дава възможност за комбиниране на теми от социалната, икономическата, политическата и пр. сфера и вплитането им в цялостна картина. За да успеем да опишем коректно съответната история обаче, трябва да открием данните, да им придадем смисъл, разяснявайки ги коректно, да визуализираме събраната информация и да я представим нагледно.

Все по-често активисти и журналисти са принудени да се справят с масивите от информация, използвайки техники за визуализация на данните, прилагани от компютърни специалисти, изследователи и хора на изкуството.

# Могат ли данните да разкажат позната история по нов начин?

Откъде да започне „историята с данни”? Въпреки че очевидният отговор следва да е *„От данните*”, толкова логично звучи и *„От въпроса, който иска да постави*”.

Факт е например, че в миналото разследващите журналисти са имали сериозни трудности поради недостиг на информация относно проблемът, по който пишат. В днешни дни е налице информационно претоварване и предизвикателството пред журналистите е да успеят да отделят верния сигнал и по този начин да предоставят ценна информация.

Журналисти и активисти вече имат реалната възможност не просто да се доверят на един-единствен източник на информация, за да опишат и анализират казаното от него, а и да се разровят из достъпните масиви от данни и да съпоставят различните твърдения, докато в пресечната им точка се открои истинският факт. Ако съумеят да придобият знание за новите инструменти и техники за работа с данни, ще успеят да извлекат максималната стойност и да я предоставят на публиката и заинтересованите страни.

Търсенето на прозрачност със средствата на новите технологии подпомага борбата с корупцията и измамите и дава възможност да се разкрият влиянията зад политиката. Ако журналистите и активистите успеят да се възползват от тези средства, то те биха могли успешно да представят и обобщават реалните действия на правителствата и регулаторните органи в своите държави например.

В резултат на цифровизацията на информацията, вече има механизъм, с който да се “извадят от архива” позабравени данни, да се сравняват, актуализират и анализират, за да бъде информирана публиката в развитие.

Пример от работата на Институт за развитие на публичната среда:

## Зад страниците на един отчет

Пример по темата, е работата на Институт за развитие на публичната среда за проследяването в процес на финансирането на предизборните кампании.

Така по време на едномесечната кампания преди Евроизбори 2014 г. блогът ни "Отворен парламент" публикува редица статии относно финансирането на участниците в предизборния период. В рамките на този мониторинг стана ясно, че един от политическите субекти в надпреварата разходва значителни средства за медийно отразяване, които не съответстват на получените постъпления по време на кампанията му.

За финал на мониторинга "Отворен парламент" проучи и отчетите, подадени от политическите партии, коалиции и инициативни комитети до Сметната палата.

Така стана ясно, че има очевидни разминавания между случвалото се по време на кампанията за описания по-горе политически участник и това, което е декларирано на хартия пред компетентния държавен орган.

Предоставянето на тези данни на обществото от своя страна води до изискването на все повече, по-нова и коректна последваща информация.

Така и разказващите "истории с данни" са реално принудени да добавят нови измерения на информацията, която предоставят, да я визуализират все по-запомнящо се и да я представят по максимално достъпен начин.

## Предизвикателството

Начините, по които една "история с данни" може да бъде разказана ще продължат все повече да се развиват. Журналисти, активисти и граждани се сблъскват с безпрецедентно количество информация от практически неограничени новинарски източници, в това число и социалните мрежи. Новинарските редакции, които са традиционните канали за разпространение на информацията, са засегнати в много голяма степен от новосъздалата се среда. Често новините се раждат в социалните мрежи и после биват използвани от професионалисти и любители, желаещи първи да ги анализират и синтезират, за да ги покажат в журналистически контекст.

Така данните в редица случаи се генерират от потребители с интерес в съответната област (социална, политическа, икономическа и пр.), като в последствие може да се превърнат в основа за журналистически материали и разследвания. Същевременно съществува и риск публикуваните данни да останат неизползвани. Именно в това се изразява предизвикателството пред всяка „история с данни" - да съумее да се възползва от новите платформи за събиране, обработка, анализиране и представяне на данни, да ги превърне в тема, предоставяща полезна за потребителите информация.

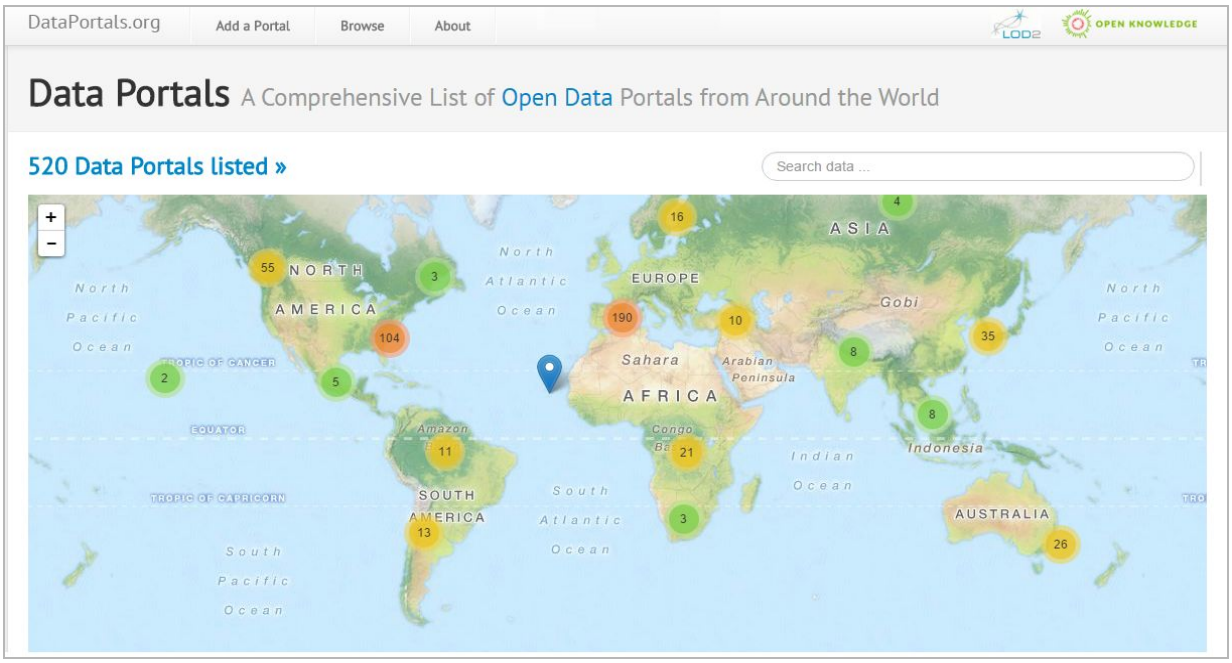
# Къде да открием данните?

**Къде могат да бъдат открити данни** по конкретна тема или въпрос? Налична ли е дадена информация и къде може да бъде потърсена тя? Откъде да се започне?

В тази част ще се опитаме да покажем някои от публичните източници на данни в интернет.

В последните няколко години се появиха редица специализирани международни портали за данни. Те са добра отправна точка за желаещите да се запознаят с публикуваната информация и възможностите за тяхна последваща употреба. Някои от тези портали са:

- Евростат - статистическата служба към Европейската Комисия (ЕК), която публикува данни за страните в Европейския Съюз (ЕС), като по този начин цели да разшири публичния достъп до информация за политиките на ЕС. Евростат не събира данни, те са предоставяни от държавите членки и техните статистически институти, които верифицират и анализират националните данни преди да ги изпратят на Евростат. Ролята на Евростат е да консолидира данните и да ги предоставя във вид позволяващ тяхното сравнение, съгласно уеднаквена методология. По същество Евростат е единствения източник на статистически данни на европейско ниво.
- Официални държавни портали за данни (като тук предоставяните от правителствените власти данни варират за различните страни) - редица държави публикуват данни на специални портали по примера на правителствата на САЩ и Обединеното Кралство (data.gov за САЩ и data.gov.uk за Обединеното Кралство). Информация, съдържаща връзки към подобни портали, може да откриете тук - Data Portals.

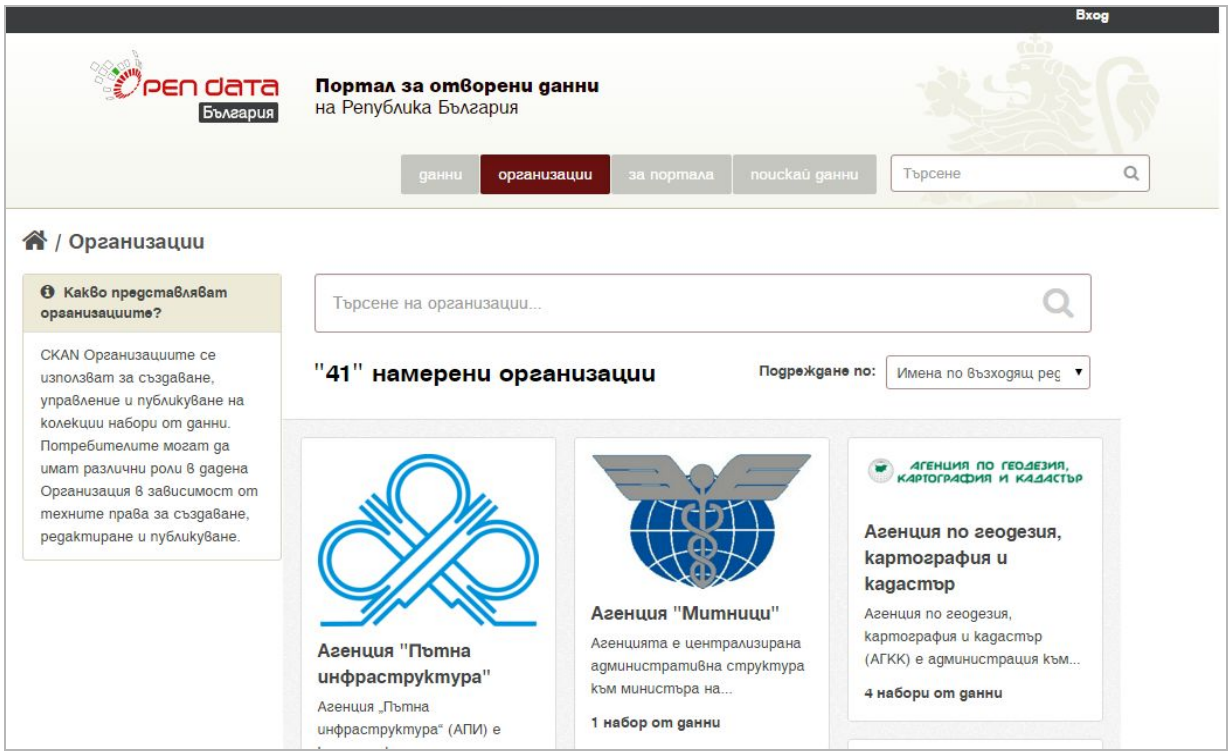



- Open Knowledge Foundation управлява платформа за споделяне на международни бази данни. Това е The Data Hub, който публикува информация във формат, улесняващ повторната й обработка.
- Порталите на Световната банка (http://data.worldbank.org/) и Организацията на Обединените Нации (ООН) (http://data.un.org/) също предоставят данни за индикатори (икономически, социални, демографски и пр.) по различни критерии за всички държави.

В България също са достъпни публични регистри, които част от държавните институции поддържат. По-долу ще се опитаме да ви представим някои от тях:

- Портал за отворени данни на Република България (https://opendata.government.bg/) - С Решение №103 на Министерския съвет от 2015 г. е приет Списък от набори от данни по приоритетни области (119 набора от данни), които да се публикуват в отворен формат. Отварянето и публикуването на тази информация на платформата се извършва съгласно График, приет на заседанието на Съвета за административната реформа на 5 март 2015 г. Порталът за отворени данни представлява единна, централна, публична уеб-базирана информационна система, която осигурява публикуването и управлението на информация за

повторно използване в отворен, машинно-четим формат заедно със съответните метаданни.




- Информационна система за управление и наблюдение на средствата от ЕС в България 2020 (ИСУН 2020 - https://eumis2020.government.bg) - разработена във връзка със задължението на държавата да използва своя собствена информационна система за управление, изпълнение и наблюдение на одобрените Оперативни програми, съфинансирани от Кохезионния и Структурните фондове на ЕС.
- Национален статистически институт (http://www.nsi.bg/bg/content/766/статистически данни) - на страницата на НСИ се предоставят резултати от провежданите **статистически** изследвания, демографски данни, класификации, номенклатури и регистри.
- Централна избирателна комисия (www.cik.bg) - на официалната интернет страница на комисията може да се намери информация за изборните резултати по видове избори от 2003 г. насам, като базите данни съдържат файлове, улесняващи повторната употреба на данни.

- Данни за финансирането на предизборните кампании на политическите субекти се публикуват на страницата на Сметната палата на Република България. Към този момент те обаче са във формат (pdf), който не предполага машинна употреба - http://erik.bulnao.government.bg/RegDonors/
- Информация за обществените поръчки в страната може да бъде търсена на страницата на Агенция по обществени поръчки - http://rop3-app1.aop.bg:7778/portal/page?_pageid=93,1&_dad=portal&_schema=PORTAL
- Информация може да бъде търсена и в Търговския регистър към Агенцията по вписванията на Министерство на правосъдието - http://www.brra.bg/
- Информация за българските народни представители и дейността на Народното събрание се публикува на страницата на парламента - www.parliament.bg (част от тази информация е в машинообработваем вид - напр. съдържащата се в профилите на депутатите).

## За необходимостта от повече публично достъпни данни:

В рамките на ООН през септември 2011 г. в Бразилия стартира Партньорството за открито управление (Open Government Partnership - OGP). Инициативата определя основните стандарти на откритост в управлението. Такъв стандарт например е публикуването на проекта на държавния бюджет. Всяка страна, която отговаря на стандартите, може да се присъедини към Партньорството.

България прави това още в самото начало на инициативата. Основният ангажимент на страната ни се съдържа в съставянето на план за действие по темите, свързани с отвореното управление – бюджетна прозрачност, достъп до информация, гражданско участие, почтеност и антикорупция. С втория план за действие на Република България по инициативата "Партньорство за открито управление", приет с Решение на Министерския съвет № 530 от 21 юли 2014 г., може да се запознаете тук.

В края на април 2015 г. в Народното събрание е депозиран законопроект за изменение и допълнение на Закона за достъп до обществена

информация /ЗДОИ/. Вносител е Министерският съвет. Със законопроекта се транспонират разпоредбите на Директива 2013/37/ЕС на Европейския парламент и на Съвета от 26 юни 2013 г. за изменение на Директива 2003/98/ЕО относно повторната употреба на информацията в обществения сектор.

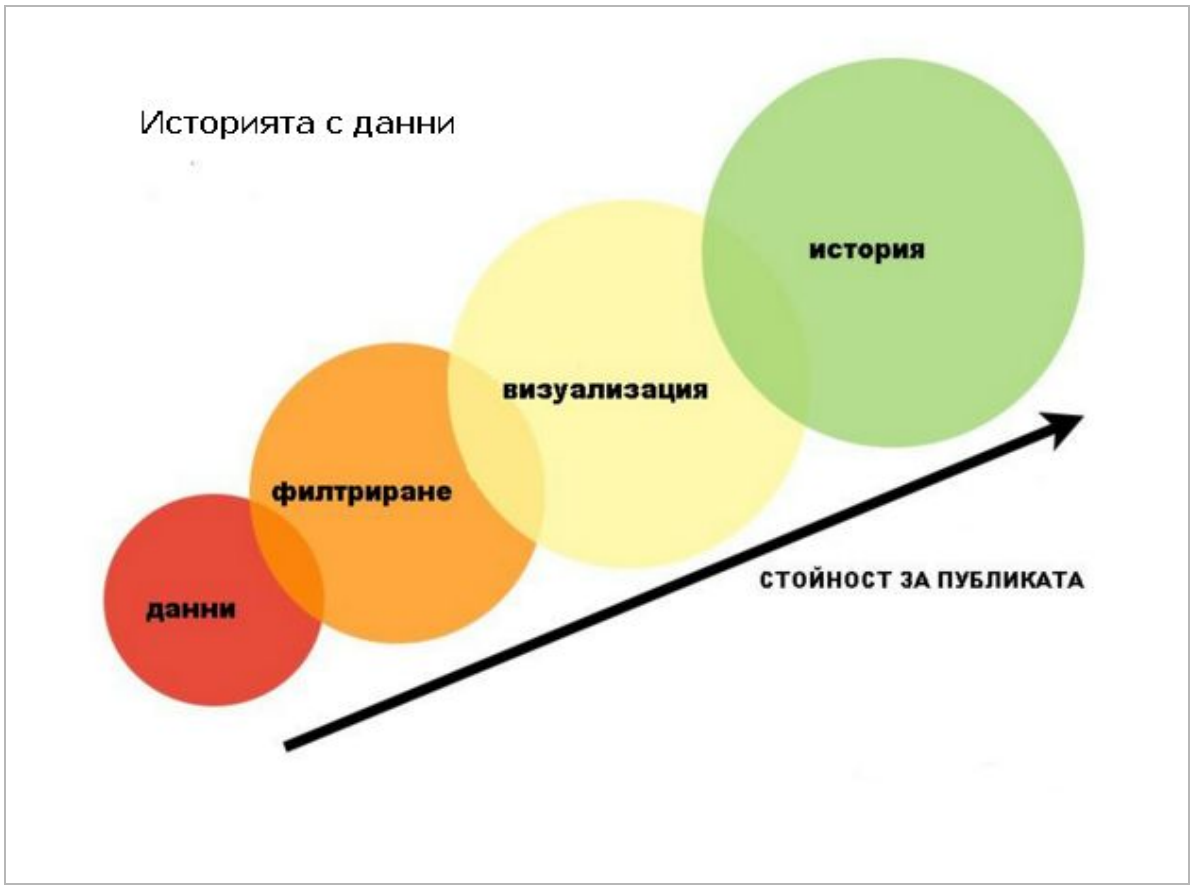
Проектът на нормативен акт е депозиран в изпълнение на мерките от Втория план за действие на Република България за участие в Инициативата „Партньорство за открито управление" и представлява основна стъпка за въвеждането на все повече "отворени данни" в България.

Предвижда се всяка информация от обществения сектор, създадена след 1 октомври 2015 г., да се поддържа в електронен вид.

Разширява се и кръгът на актуалната обществена информация, която администрацията следва да публикува – в срок от три месеца от влизането в сила на закона.

# Как може да се разкаже „история с данни”?

След като вече описахме част от източниците, от които може да бъдат ползвани данни, в тази част ще се представим няколко интересни примера за визуалното оформление на историите, които искаме да разкажем.




Фигурата по-горе показва стъпките в написването на “история с данни”.

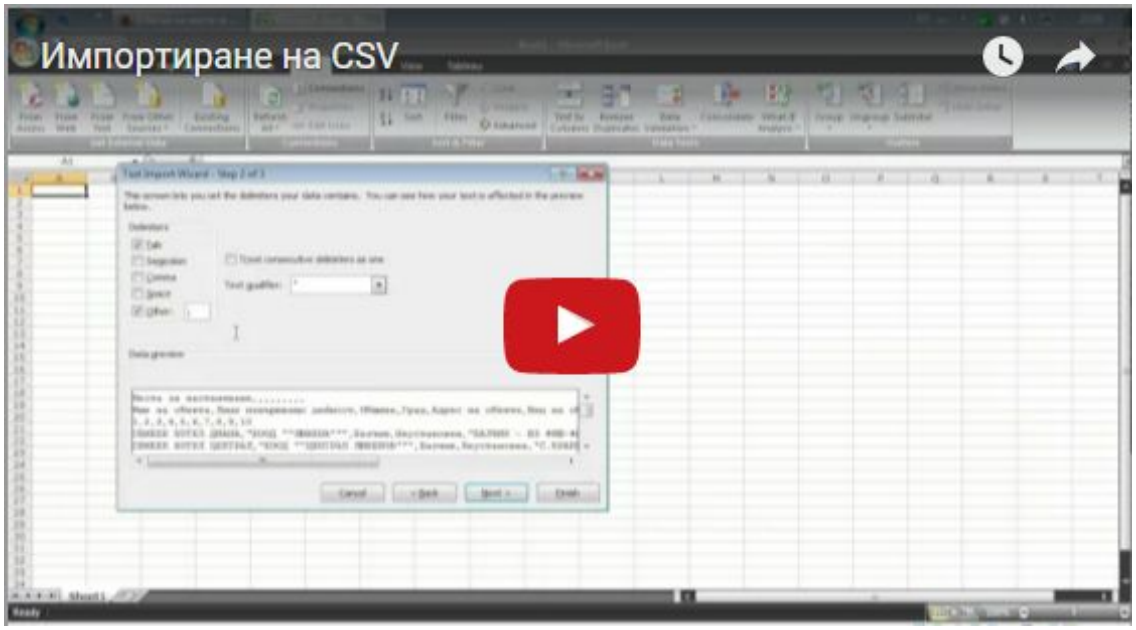
## “Извличане” и обработка на необходимите данни

Следващите линкове представят нагледно как база данни може да бъде обработена от CSV формат в Excel, за да се продължи с тяхната последваща употреба и анализ с цел разказване на история.

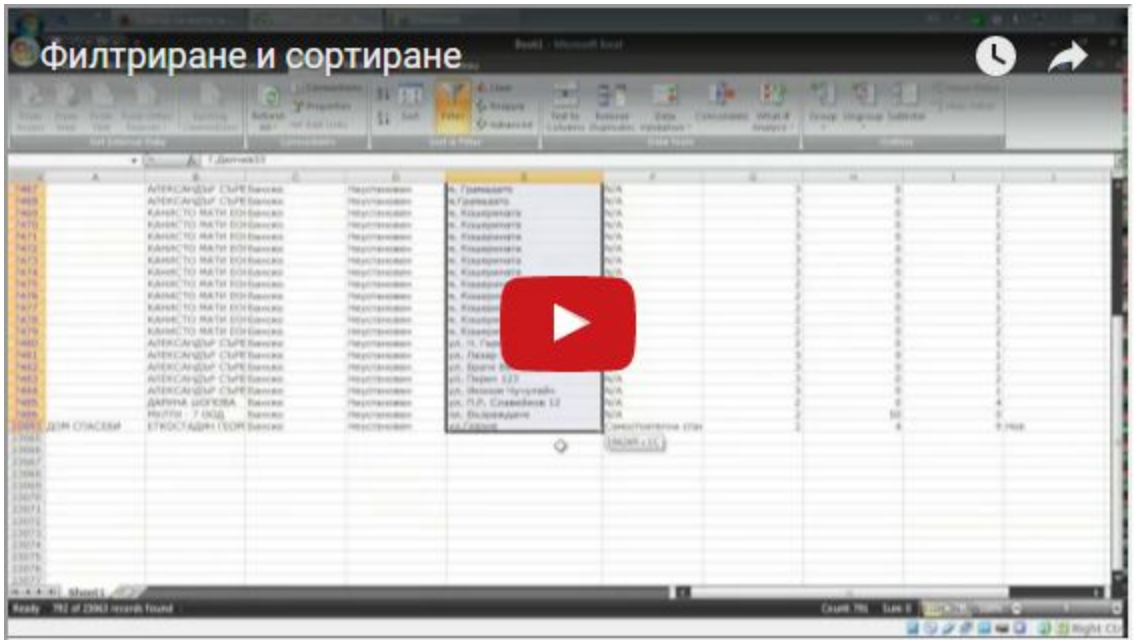
За пример е използван регистър на местата за настаняване според данни на Министерството на туризма на Република България. Обучителният модул е разработен от екипа на Институт за развитие на публичната

среда в рамките на семинар “Журналистиката среща данните - сезон 1”, проведен на 29-и май, 2015 г.

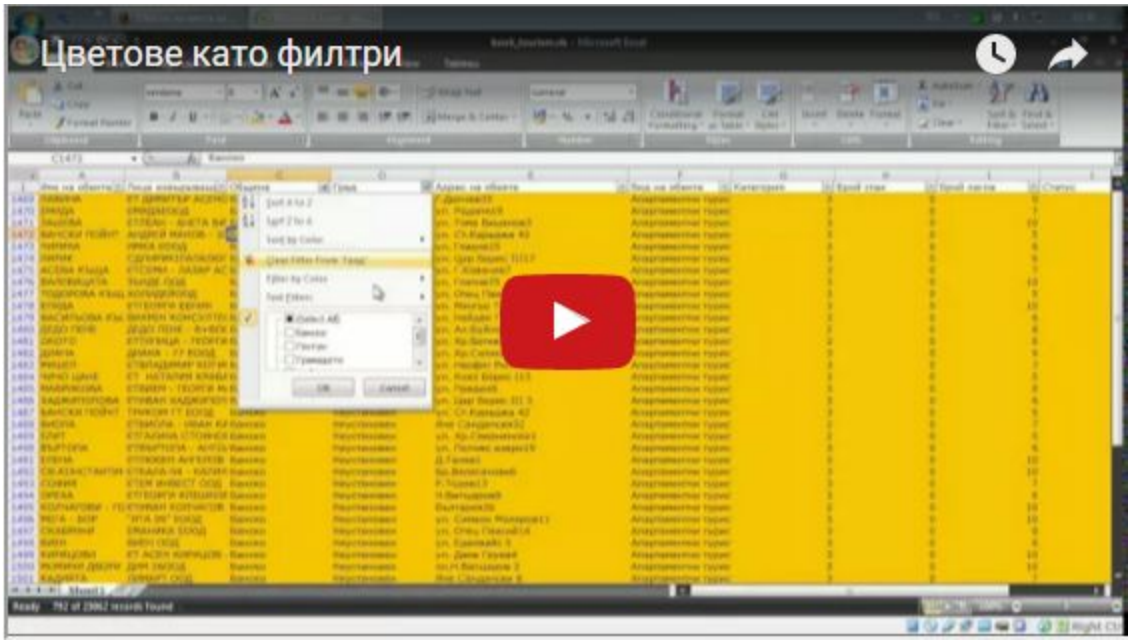
1. Първа стъпка - Представяне на Портал за отворени данни на Република България, даунлоуд на примерна база и импортиране на данните (CSV) в Excel.



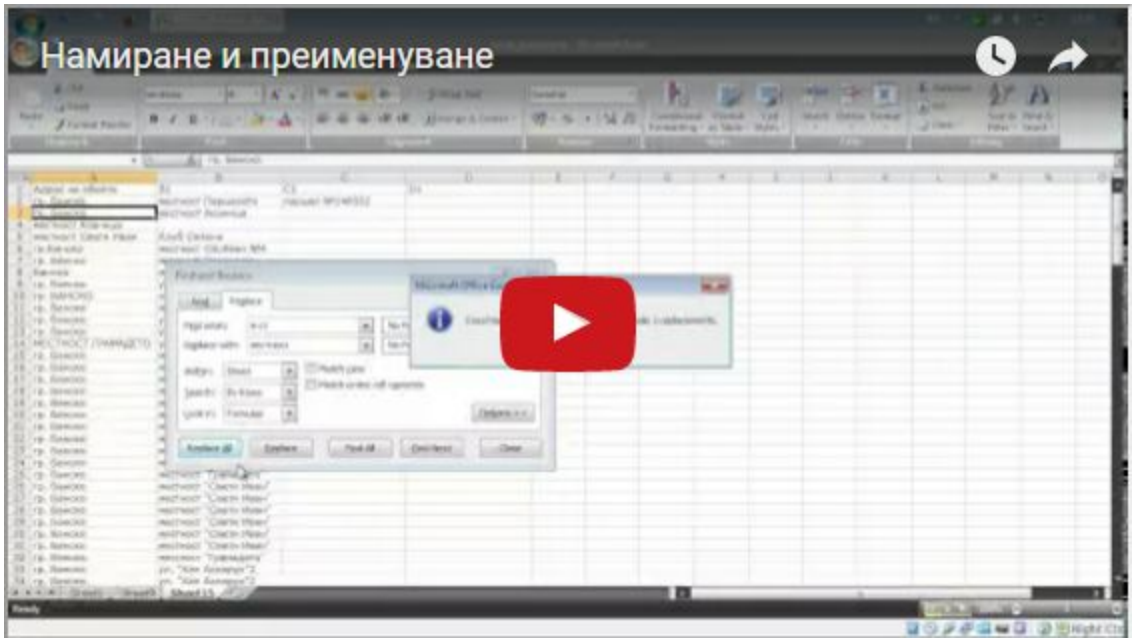

2. Втора стъпка - Методи за филтриране и сортиране на данни. Основна цел е да се извлекат конкретни данни и да бъдат подредени според нуждите на потребителя.



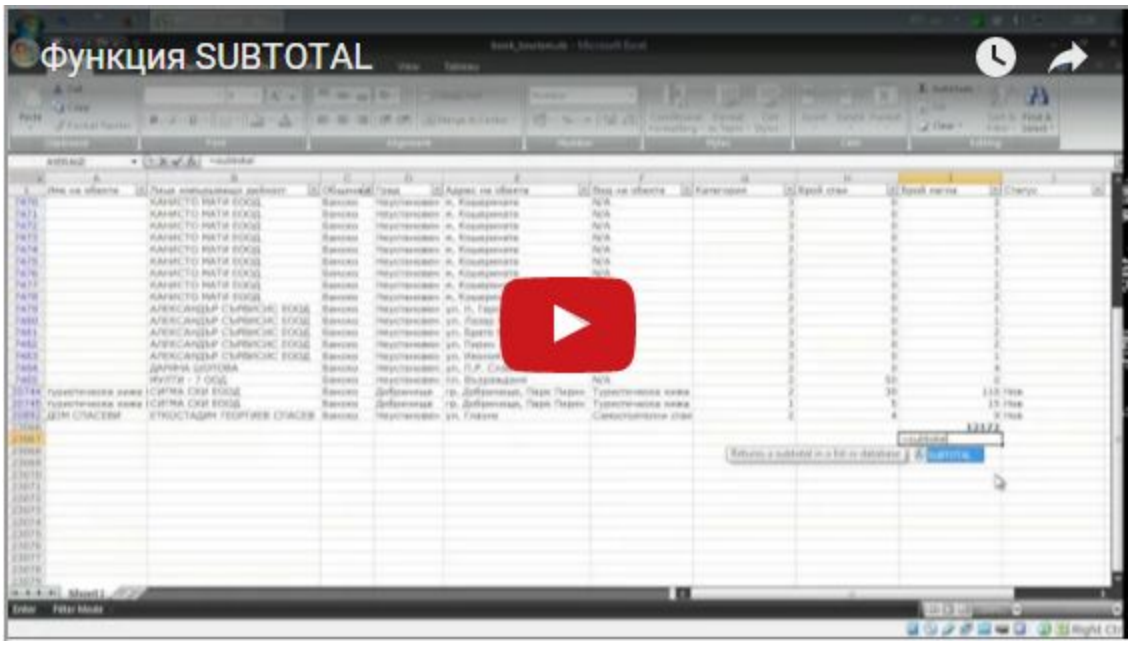
3. Трета стъпка - Използване на цветове като филтри.



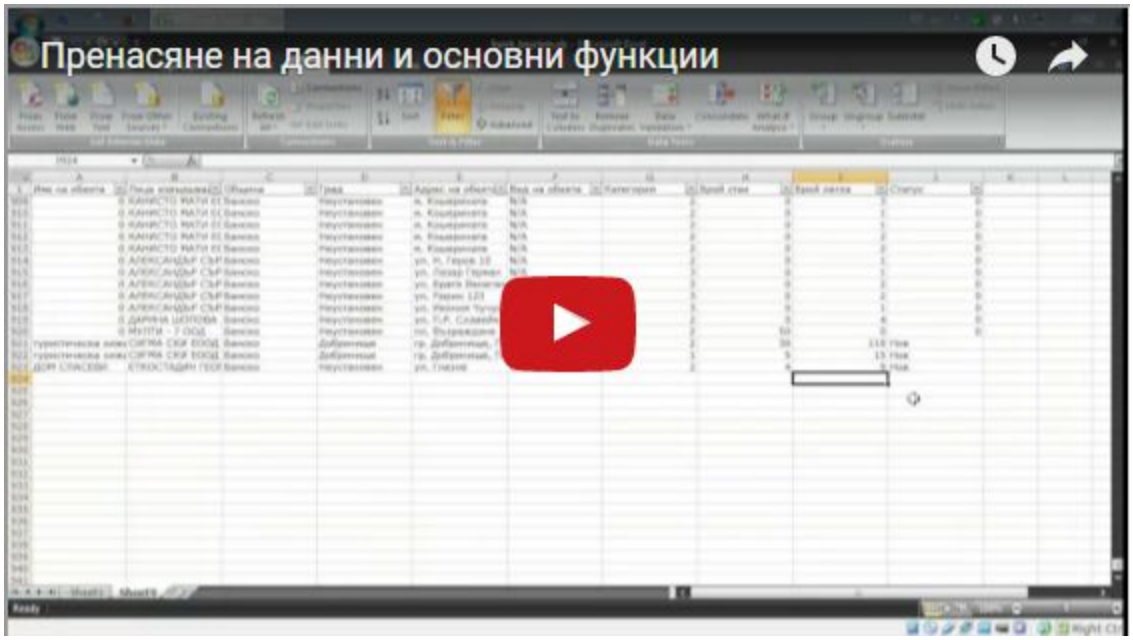

4. Четвърта стъпка - "Намиране и преименуване" е функция, която позволява да бъдат намерени елементи от данните и да бъдат приведени в общ вид.



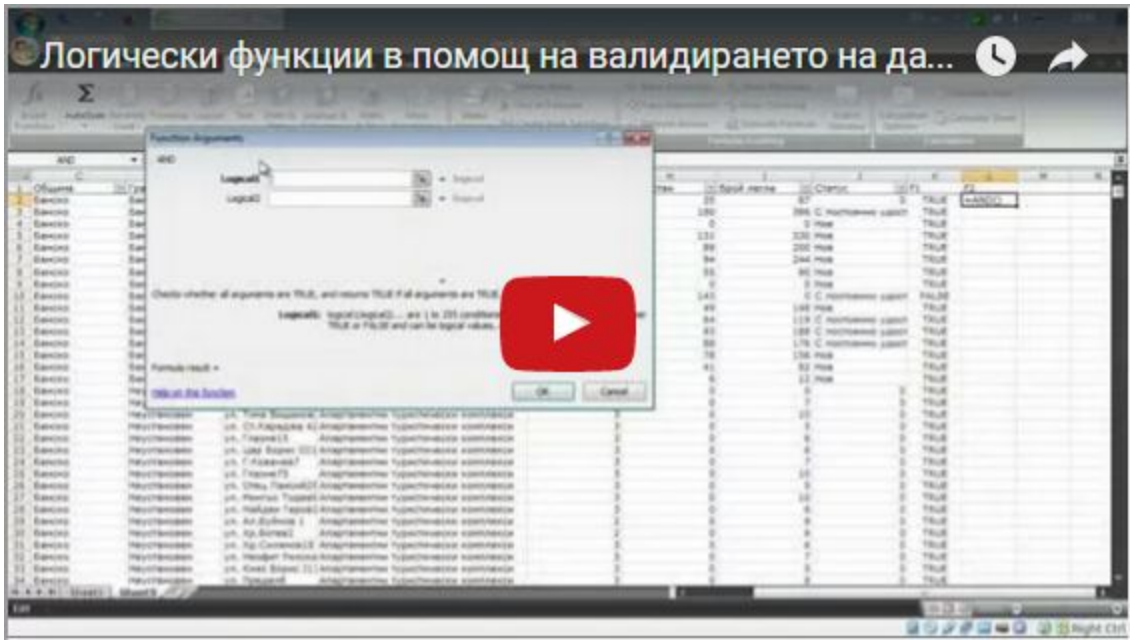

5. Пета стъпка - Функцията SUBTOTAL ни показва междинна сума. Функцията е полезна, когато се работи с филтриран списък - SUBTOTAL сумира само видимите величини.



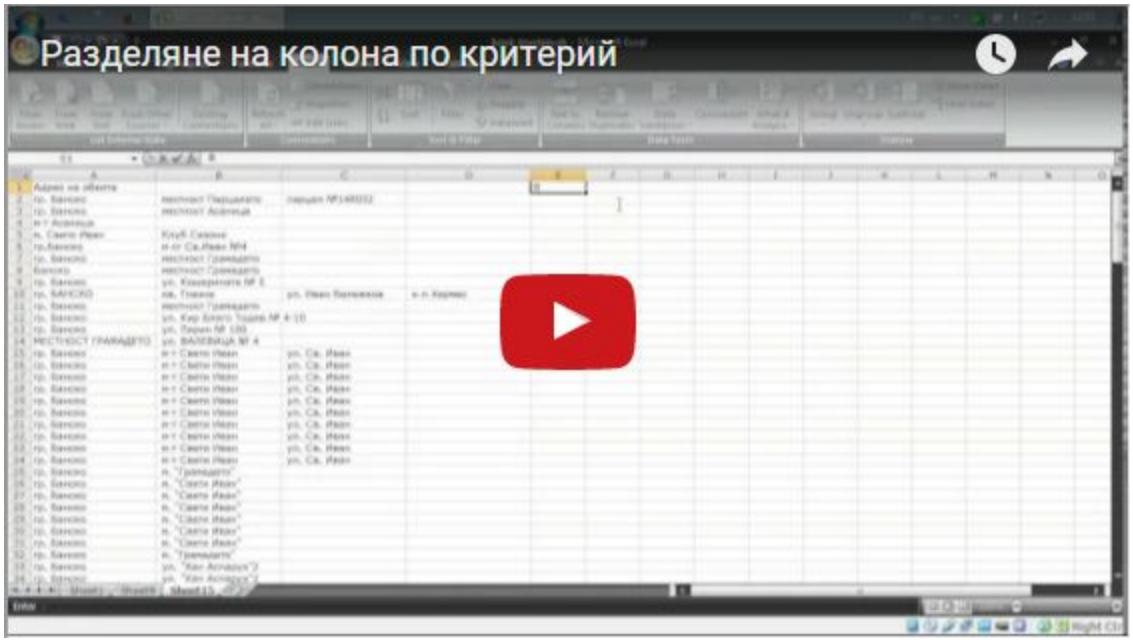

6. Шеста стъпка - Пренасяне на данни от едно работно поле в друго. Промяна, направена в първото работно поле, се отразява и във второто. Функции за събиране и изичсляване на средна стойност.



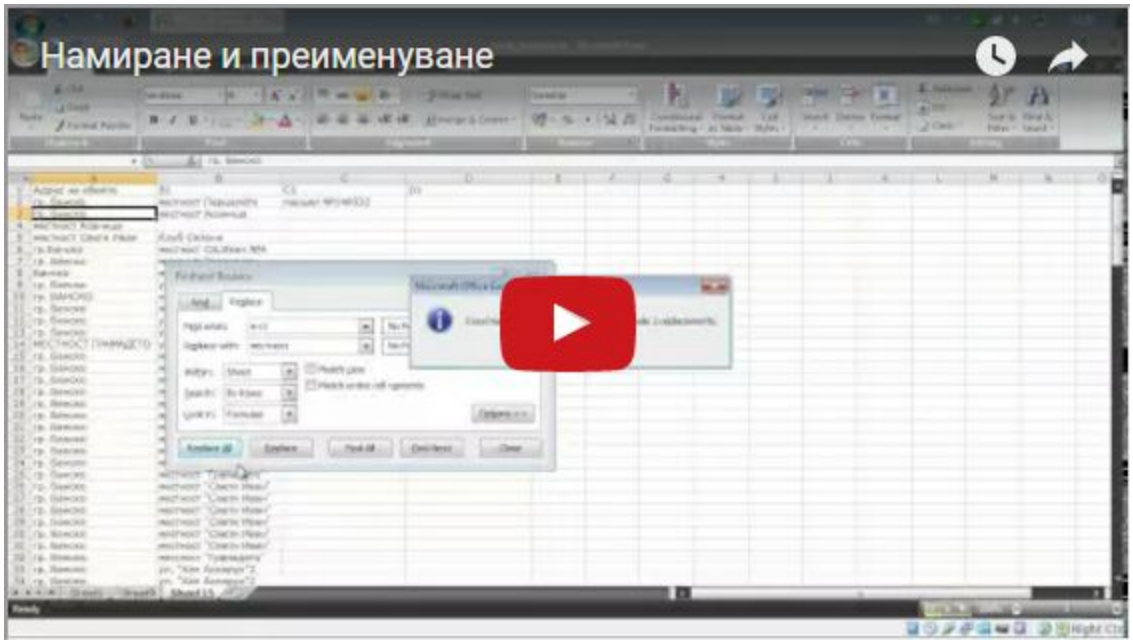

7. Логически функции в помощ на валидирането на данни. IF и AND функции като част от методите за проверка коректността на предоставените данни.



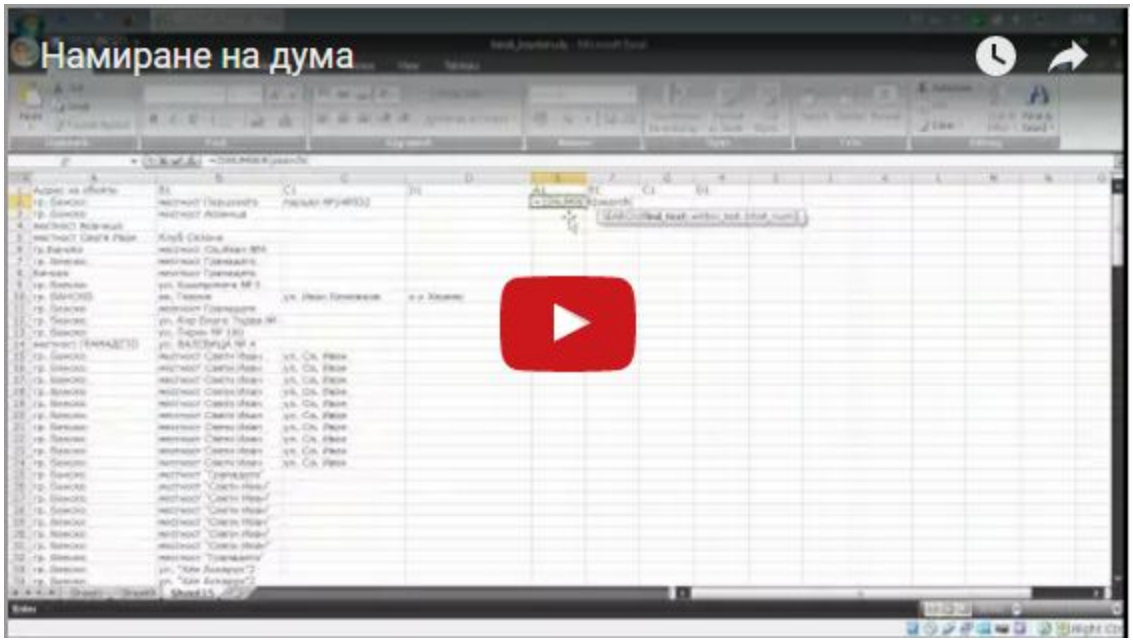

8. Разделяне на данните от дадена колона по зададен каритерий.



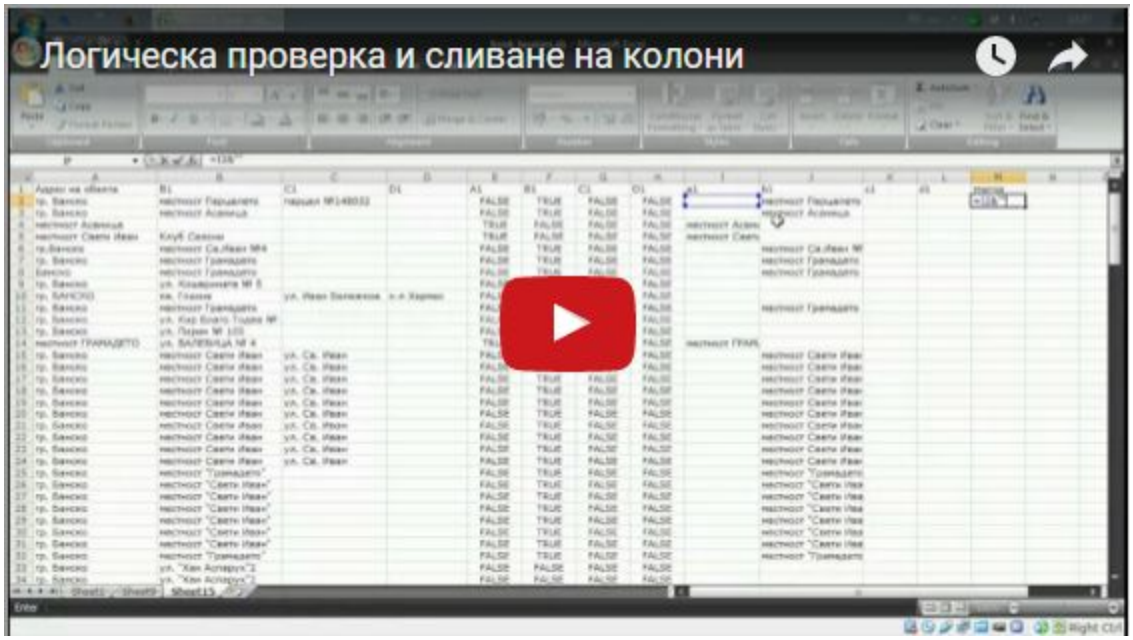

9. “Намиране и преименуване” - функция, която позволява да бъдат намерени елементи от данните и да бъдат приведени в общ вид.




10. Намиране на дума. Функция ISNUMBER, в комбинация със SEARCH за проверка дали дадена дума съществува в конкретна клетка. Функцията връща резултат TRUE/FALSE.




11. Логическа проверка и сливане на колони.



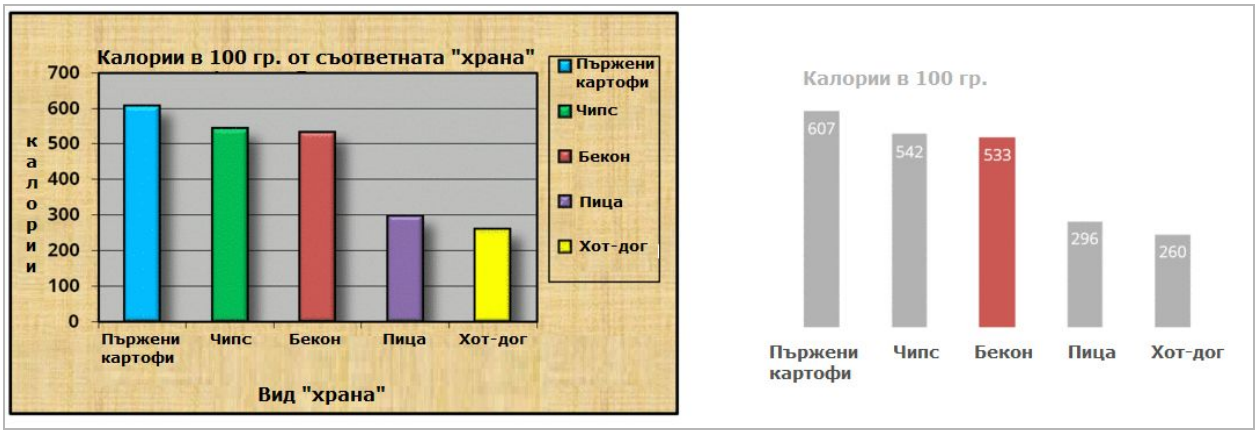

## Препоръки за визуализация на данните

След като вече разполагаме с данните, необходими за разказването на история, пристъпваме към тяхното оформление по най-подходящ начин.

Водещ принцип е разбирането, че по-опростената графика предава по-ефективно и въздействащо информацията. Основни стъпки за по-доброто й графично представяне са:

- премахване на фона (remove backgrounds)
- премахване на излишния текст (remove redundant labels)
- премахване на очертанията (remove borders)
- намаляване на използваните цветове (reduce colors)
- премахване на допълнителните 'специални' ефекти (remove special effects)
- премахване на удебеления шрифт (remove bolding)
- изсветляване на текста (lighten labels)
- изсветляване или премахване на линиите (lighten or remove lines)
- преместване на етикетите вътре в самата графика (direct label)

Следващите две графики показват информацията преди и след използването на “стъпките”, описани по-горе:




***изт.*[*darkhorseanalytics.com*](darkhorseanalytics.com)**

От съществено значение за визуализацията е тя да е подбрана по начин, който да представя информацията по най-разбираем начин.

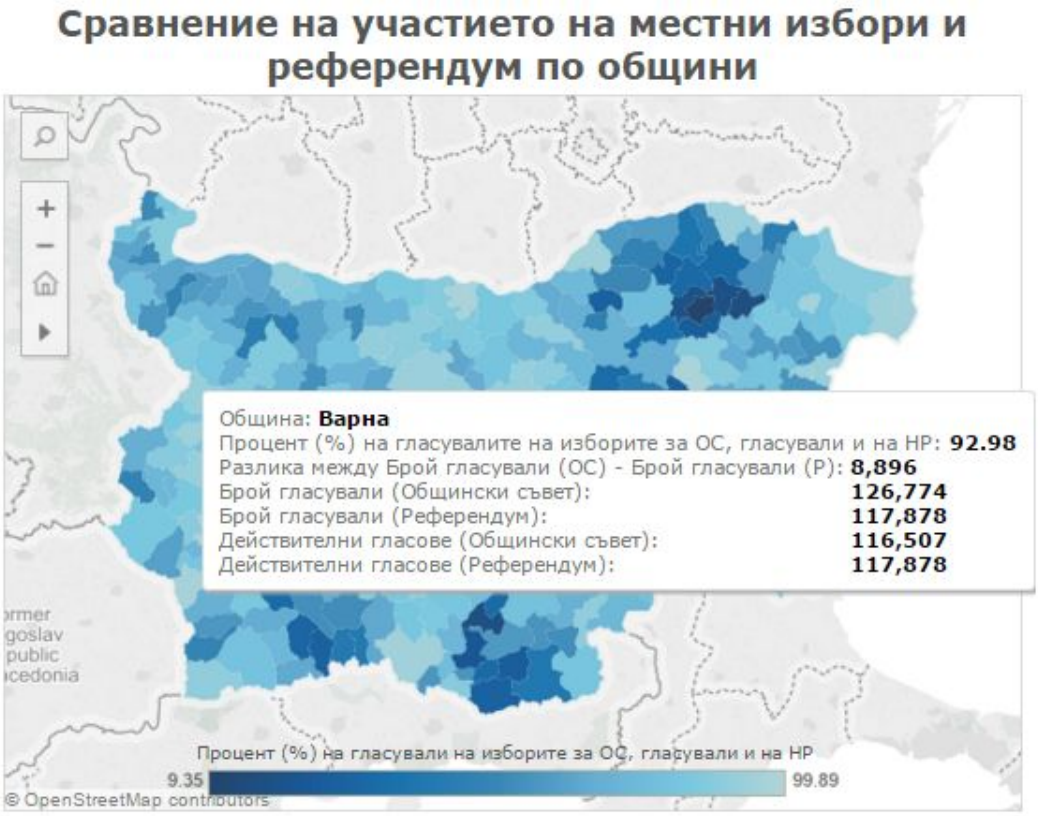
## Пример от работата на Институт за развитие на публичната среда:

### Сравнение на участието на местни избори и референдум по общини

Пример за визуализация на голям обем от информация на картова основа е работата на Институт за развитие на публичната среда за сравнението на гласувалите в страната по време на местните избори и националния референдум (проведени на 25-ти октомври, 2015 г.). Картата визуализира данните от протоколите на 265 общински избирателни комисии в страната, публикувани на страницата на Централната избирателна комисия. Визуализацията дава информация за всяка една от общините на територията на България за следните индикатори:

- процент гласували на изборите за общински съвет, гласували и на националния референдум;
- разлика между брой гласували за общински съвет и брой гласували за референдума;
- общият брой гласували за общински съвет и общият брой гласували за референдум;

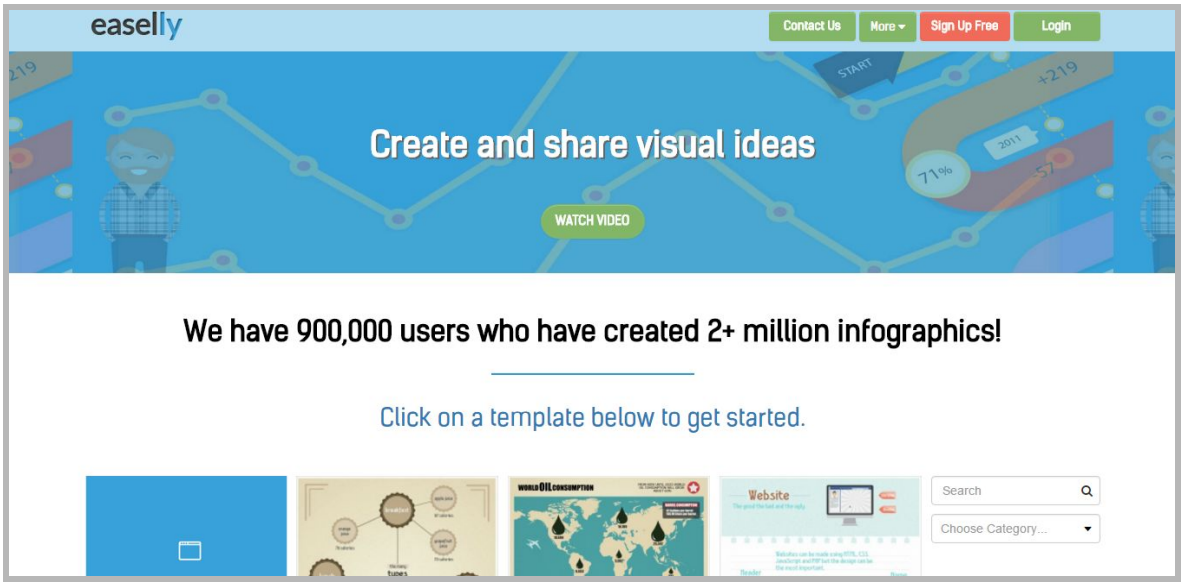
➢ действителни гласове за общински съвет и действителни гласове за референдум.



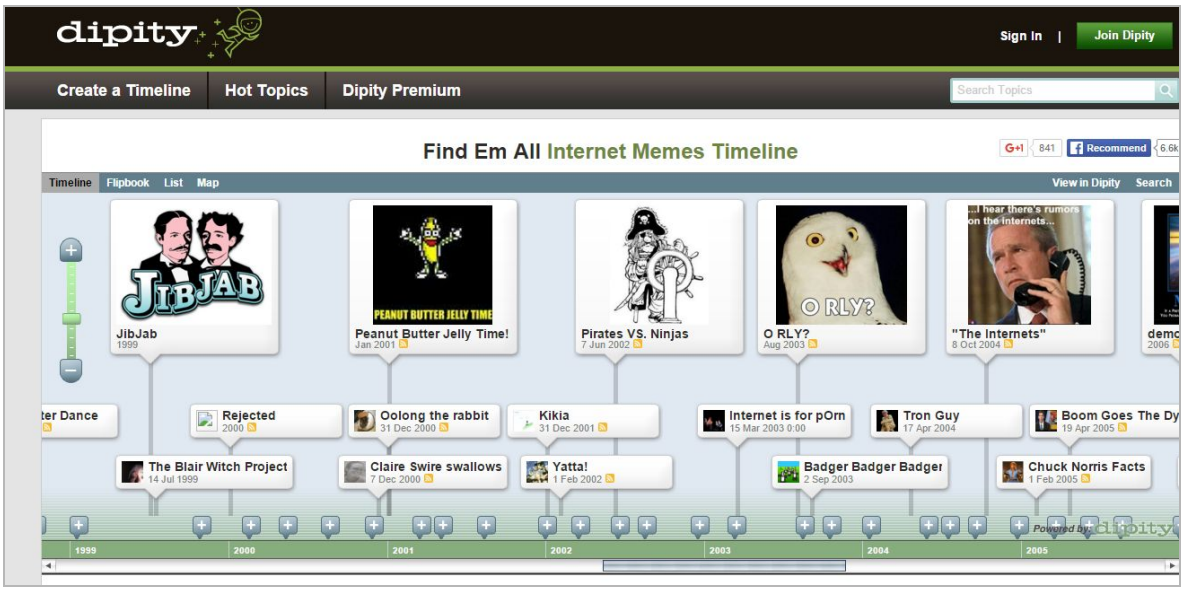
# Инструменти за създаване на "история с данни"

В тази част ще представим няколко онлайн инструмента, чрез които може да се създадат и визуализират различни истории:

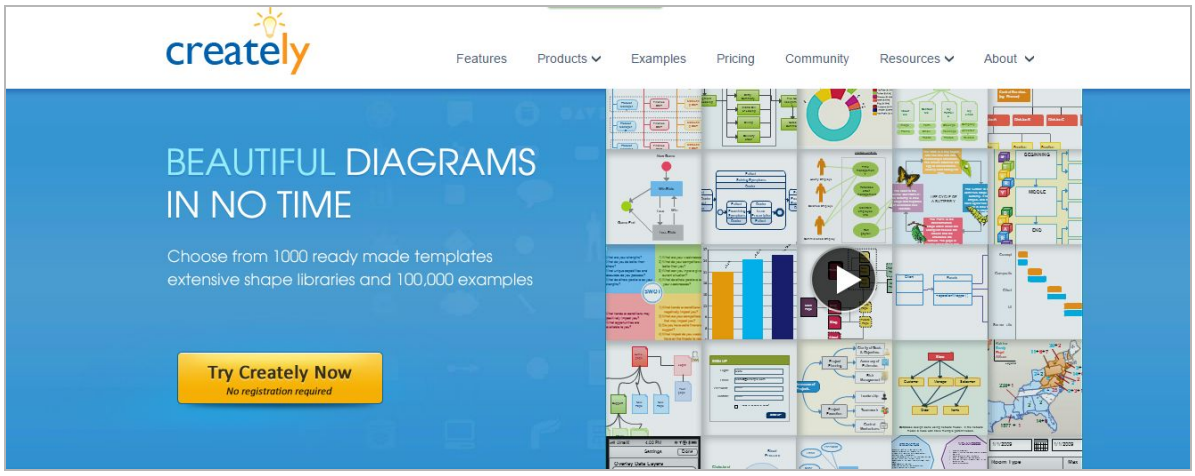
- Easel.ly - www.easel.ly - уебсайт, предоставящ голям брой безплатни теми за инфографики, които потребителите могат да персонализират с цел представяне на своите идеи онлайн (има безплатна версия с ограничени функционалности и платени варианти на инструмента).



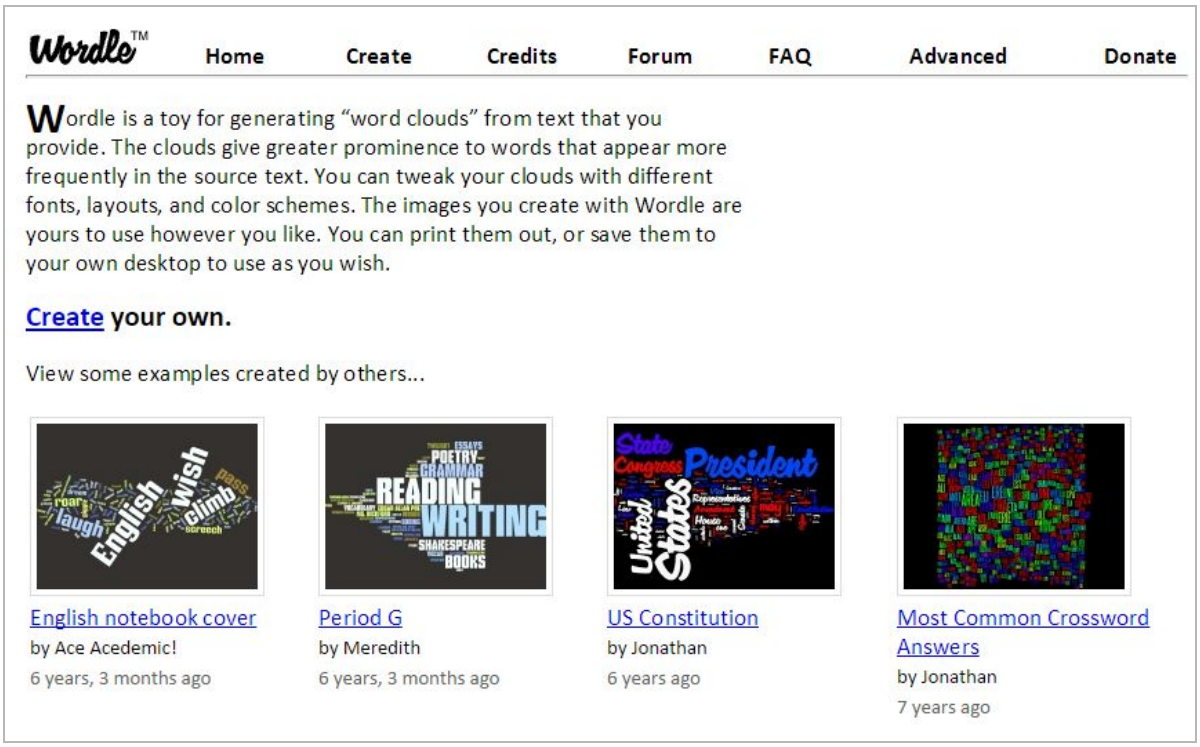

- Visme.co - www.visme.co - инструмент, който позволява създаването на интерактивни презентации, инфографики, банери, анимации и др. (има безплатна версия с ограничени функционалности и платени варианти на инструмента).
- Infogr.am - infogr.am - инструмент, чрез който може да се създаде инфографика в три "лесни стъпки" - избор на шаблон, визуализиране на данните и публикуване (има безплатна версия с ограничени функционалности и платени варианти на инструмента).
- Dipity - www.dipity.com - е инструмент, който дава възможност за създаване на дигитални "времеви линии" и инфографики, базирани на тях.




- Google Charts - [developers.google.com/chart/?hl=en](developers.google.com/chart/?hl=en) - вероятно е инструмент, за който много потребители знаят. Той изисква и известна предварителна подготовка и опит, за да може да бъде използван напълно функционално. Google Charts е безплатна възможност за изработване на различни видове таблици, карти, графики и др.
- Creately - [creately.com](creately.com) - интернет инструмент, който дава възможности за създаване на различни диаграми и графики (има вариант за безплатна версия с ограничени функционалности и платени варианти на инструмента).



- Charts Bin - chartsbin.com - безплатен онлайн инструмент, който позволява на потребителите да създават интерактивни карти, на които могат да публикуват различна информация.
- Wordle - http://www.wordle.net/ - безплатен инструмент за генериране на "облаци от думи" от текст, който е зададен от потребителя. "Облаците" октроят най-често използваните в дадения текст думи.




**Вместо заключение** бихме искали да използваме един цитат на Антоан дьо Сент Екзюпери:

*"Съвършенството е достигнато, не когато няма какво повече да се добави, а когато няма какво да се премахне."*