1. Набор от данни. Избиране и изтегляне на свободно достъпен набор от данни; Описване на набора от данни – предназначение, структура, адрес за достъп.

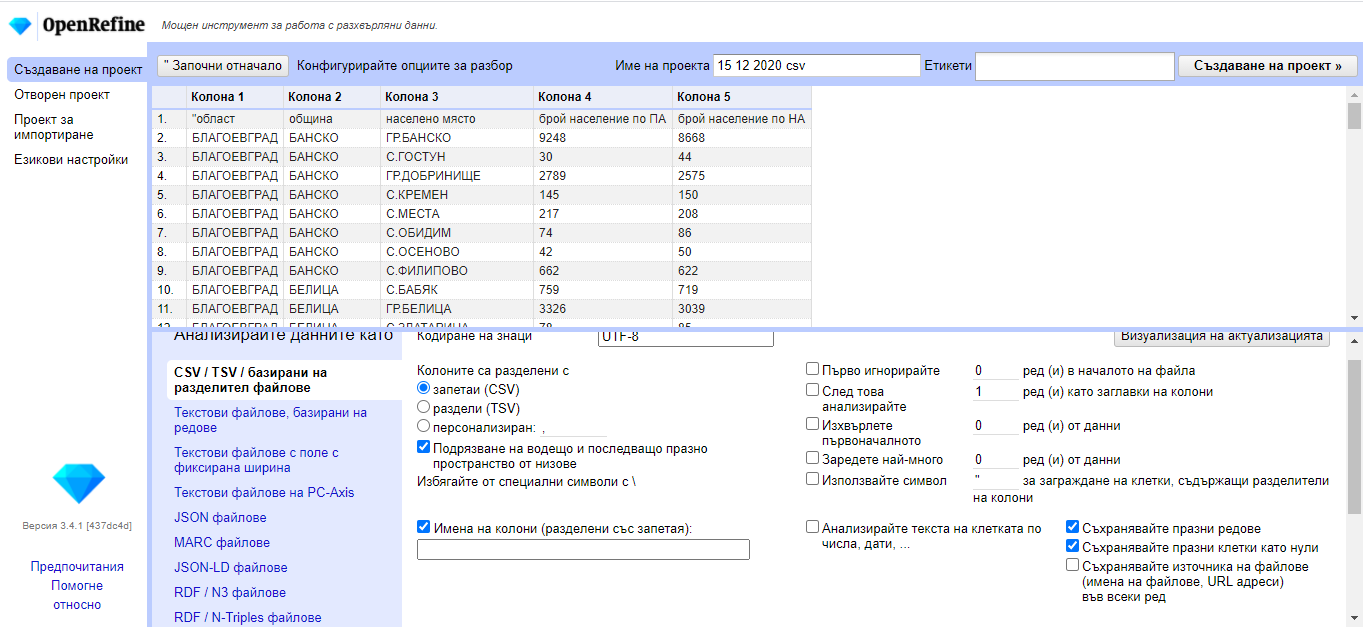
Изтеглих набора от данни от отвореният портал на Република България. opendata.government.bg. Набора от данни дава информация за броя на населението по населени места в Република България по постоянен и настоящ адрес към 15.12.2020г. Предназначението им е да дава информация с точност какъв е броят на жителите, по постоянен и по настоящ адрес, в градовете и селата.

Структурата на данните е разпределена в 5 колони: име на областта – за да може да се търси и по област и да се конкретизира населеното място, име на общината – за да може да се търси по община, населено място – за да се знае кой град или село, брой по постоянен адрес, брой по настоящ адрес – за да се направи разликата колко са тези които само прибивават и какъв е истинският брой на живущи.

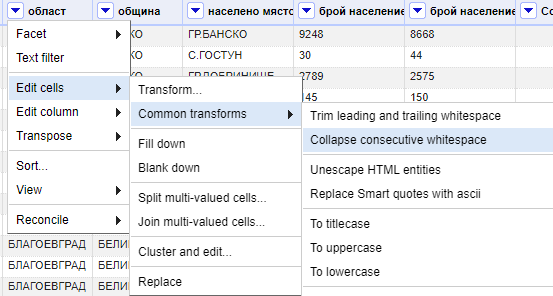
Адресът за достъп до набора от данни: <https://data.egov.bg/data/resourceView/2784b798-0ea3-49cb-9d4a-360f0b074e10?rpage=1> Данните за броят на жителите се обновяват ежемесечно на 15-то число.

2. Пречистване на данни. Отваряне или импортиране на набора от данни с Google Таблици и/или Google Refine; Прилагане на операции за пречистване на данни; Описване на приложените операции за пречистване на данни – предназначение и примерен резултат.

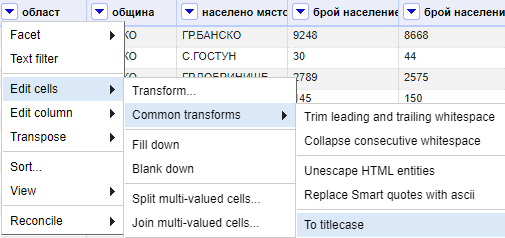
Отваряне на набора от данни с Google Refine.



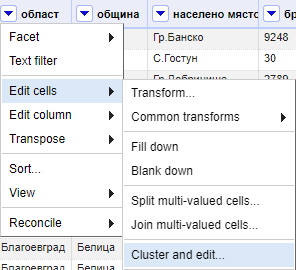
След това отстраних допълнителните интервали между думите:

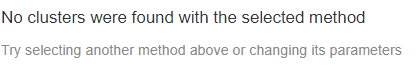


След това направих всички редове да започват с главна буква, а не да са всички главни:

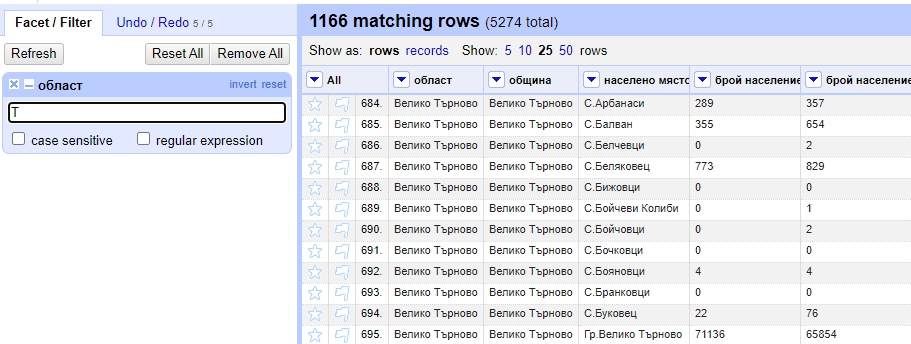


След това използвах функцията за намиране и уеднаквяване :



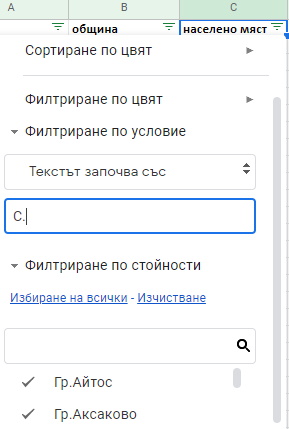
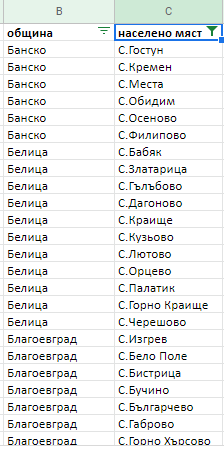


- Филтриране на данните с Text Filter – например търсене по колоната „област“:



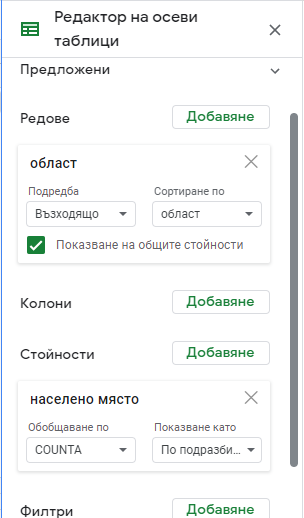
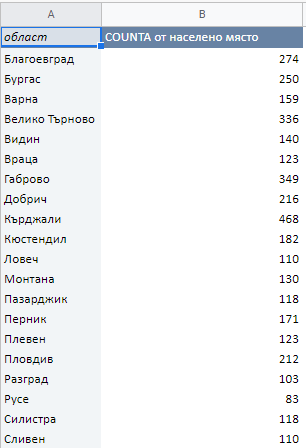
3. Филтриране, обобщаване, групиране. Прилагане на филтриране, обобщаване, групиране с помощта на Google Таблици (осеви таблици; pivot tables) или Microsoft Access (заявки; queries); Описване на създадените осеви таблици в Google Таблици и/или заявки в Microsoft Access – предназначение и примерен резултат.

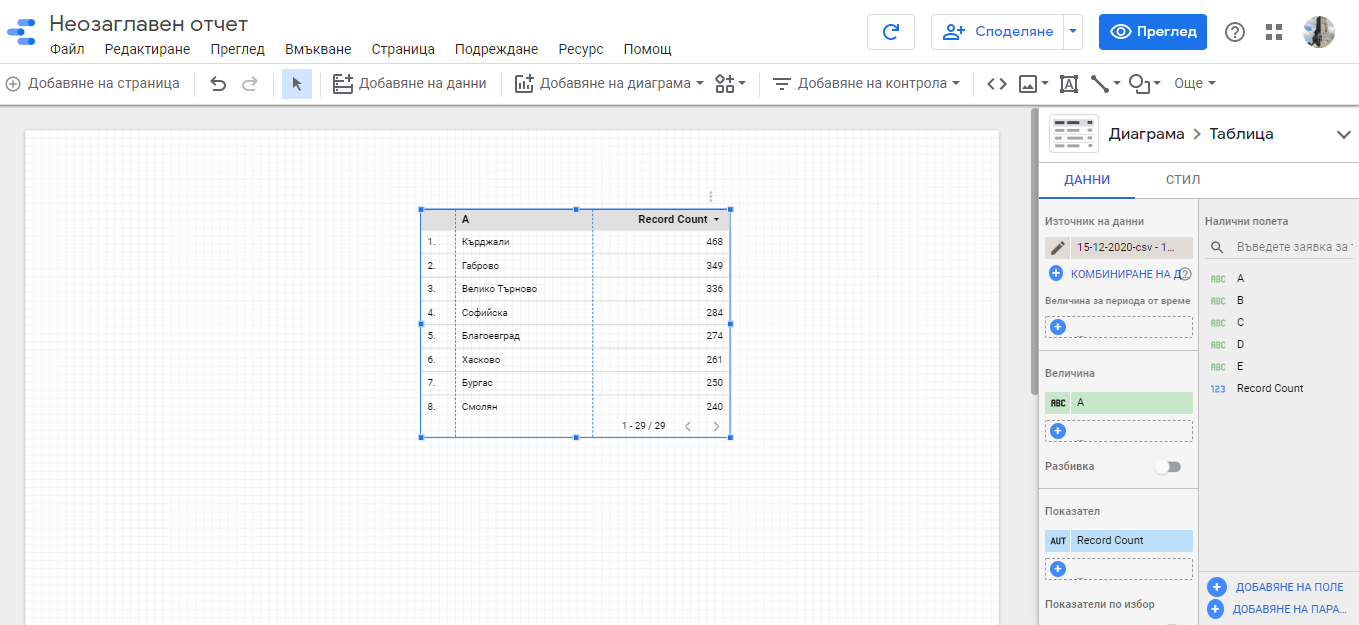
3.1 Филтриране на данни в Google Таблици чрез Данни->Създаване на филтър->Филтриране по условие->Текстър започва със-> Избрах да филтрира всички села в областите.

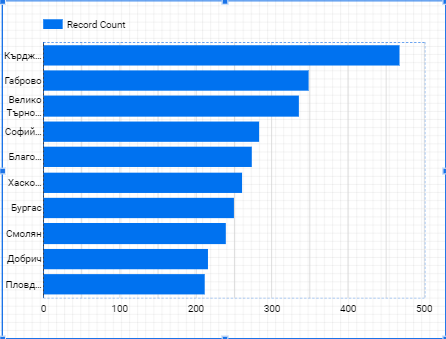
 

3.2 Обобщаване и групиране на данни с помощта на осеви таблици

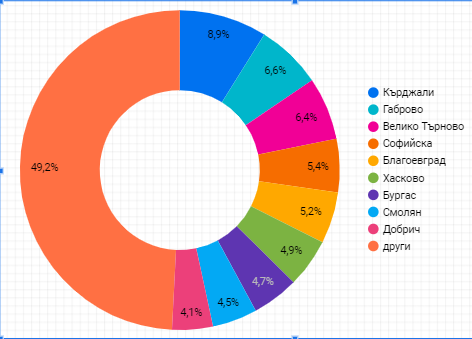
Чрез осева таблица в Google Таблици обобщаваме и групираме данни от данните за броя на населението по населени места в Република България. След като избера от менюто Данни->Осеви таблици избрах обобщаването да се създаде на нов лист. И зададох преброяване „COUNTA“ на броя населени места по различните области.

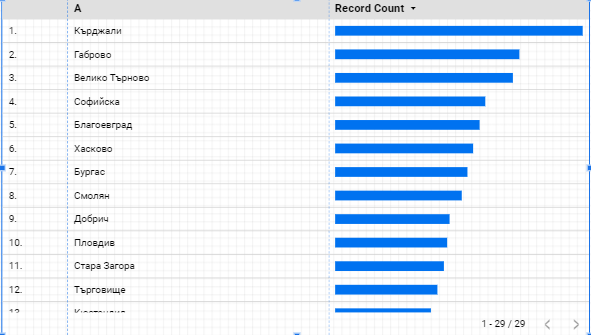
4. Визуализиране на данни. Използване на Google Data Studio за създаване на различни визуализации на набора от данни; Описване на създадените визуализации – предназначение и примерен резултат. След като импортираме в Google Data Studio създаваме различни визуализации на осевата таблица с обобщени данни за броя на населените места в областите:



Тази диаграма има числа в долната част, които показват приблизителната стойност на броя на населени места по различните области.



Диаграмата разделя областните градове в Република България на проценти, които съответстват на броят на населените места в тях.



Тази визуализация със стълбове не показва брой и проценти, но по нея можем да видим, кои области градове отговарят за най-много градове и села.

5. Споделяне на данни. Използване на Google Таблици, Google Data Studio или GitHub за споделяне на данни; Описване на споделянето – предназначение, обхват.

Споделяне чрез GitHub: https://github.com/vanyateneva/Project