

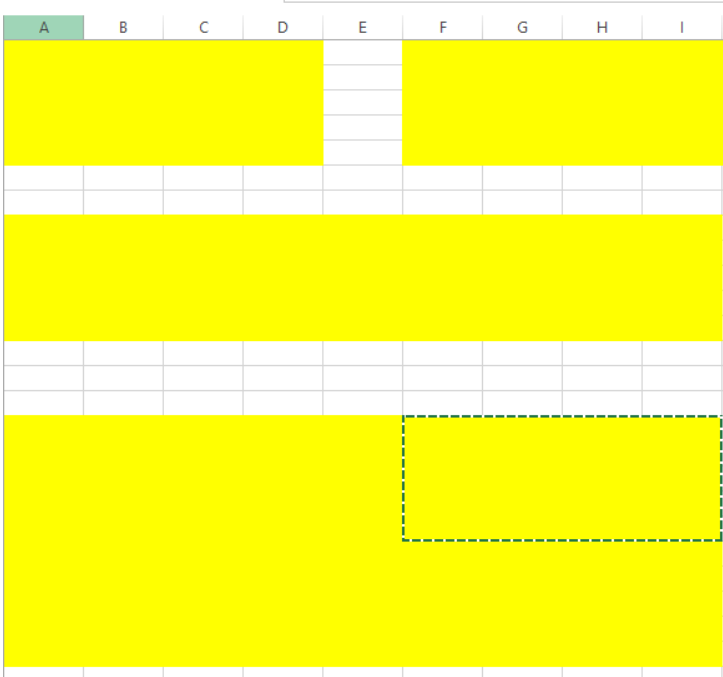
## Углубляемся в функционал Tableau Desktop

Итак, добро пожаловать на второй урок Tableau Desktop. На предыдущем занятии мы начали знакомство с этим инструментом. Изучили его базовый функционал. Сегодня мы углубимся в его возможности. Параллельно более детально затронем вопрос построения дашборда.

Для нашего семинара я решил использовать датасет, в котором представлена статистика преступлений в городе Бостон. Давайте попробуем узнать, где безопаснее всего жить в этом городе. А также отследим пики активности преступников.

Что ж для начала откроем TDPE, далее - на стартовой странице во вкладке подключений выберем CSV or Text file. Выбираем файл Crimes. После подключения к этому файлу мы видим, что текстовый разделитель задан некорректно. Для его правильной настройки заходим в text file properties нажав правой кнопкой мыши на источник данных, находящийся на панели подключения источника данных. Устанавливаем в качестве разделителя запятую. В качестве text qualifier - кавычки.

Теперь давайте посмотрим как работает функция объединения JOIN при объединении нескольких источников данных. Что такое JOIN - это объединение двух и более таблиц по определенному ключу. Есть разные типы JOIN. А также разные условия объединения таблиц.



A	B	C	D	E	F	G	H	I

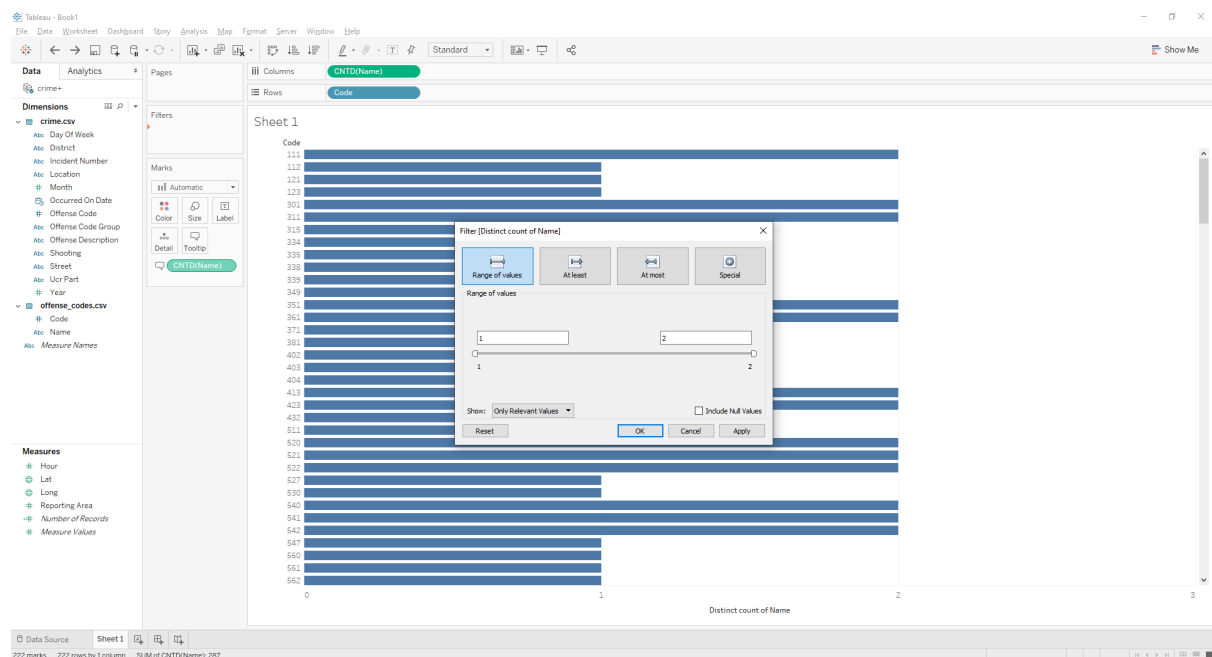
Для того, чтобы осуществить объединение данных из нескольких источников в одну витрину необходимо: из раздела Files берем OFFENSIVE\_CODES и перетаскиваем на

панель подключения источника данных. Указываем, какой тип JOIN хотим использовать. Указываем ключ для обоих источников данных, по которому будем происходить их объединение, а также его условия. В качестве ключа можно выбрать одно из полей, имеющихся в источниках данных, из выпадающего списка или создать свое вычисляемое поле.

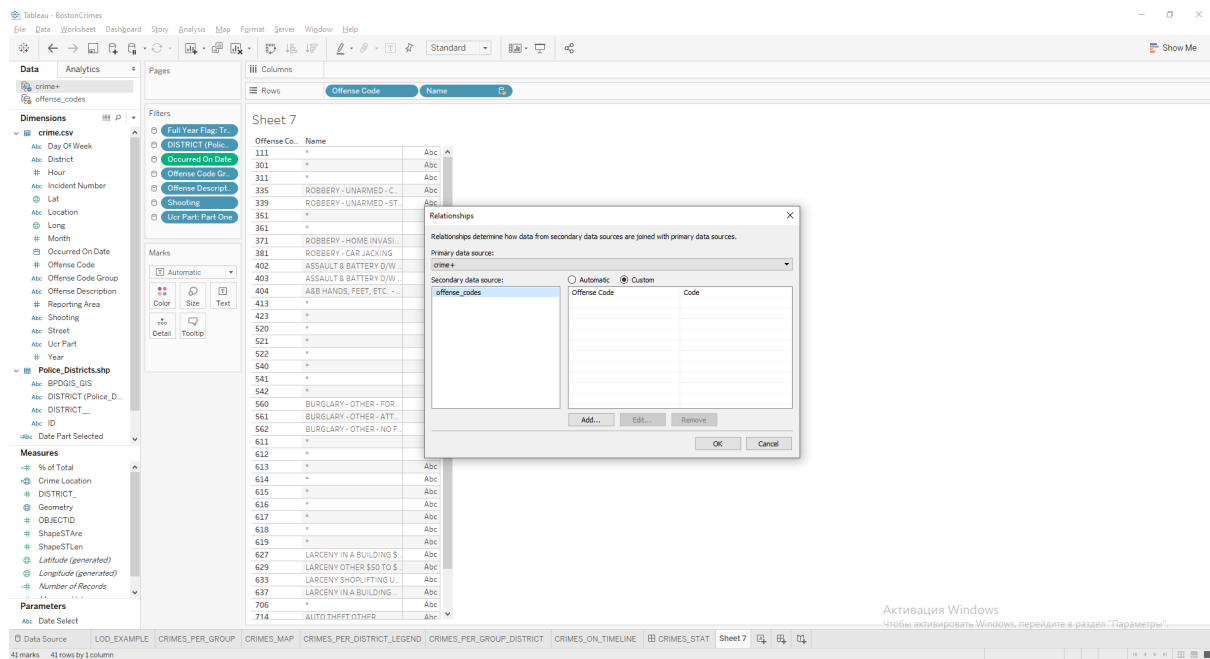
The screenshot shows the Tableau interface with a join operation between two CSV files: 'crime.csv' and 'offense\_codes.csv'. The join is configured as an Inner Join on the 'OFFENSE\_CODE' field. The resulting data table is displayed below the join configuration.

#	crime.csv MONTH	crime.csv DAY_OF_WEEK	crime.csv HOUR	crime.csv UCR_PART	crime.csv STREET	crime.csv Lat	crime.csv Long	crime.csv Location	offense_codes.csv CODE	offense_codes.csv NAME
8	1	Wednesday	18	Part Three	METROPOLITAN AVE	42,2808	-71,1325	(42.28075962, -71.13...	3 625	DANGEROUS C
8	1	Wednesday	18	Part Three	METROPOLITAN AVE	42,2808	-71,1325	(42.28075962, -71.13...	3 625	DANGEROUS C
8	1	Wednesday	19	Part Three	AUSTIN ST	42,3752	-71,0649	(42.37518695, -71.06...	3 160	FIRE REPORT -
8	1	Wednesday	19	Part Three	AUSTIN ST	42,3752	-71,0649	(42.37518695, -71.06...	3 160	FIRE REPORT -
8	1	Wednesday	19	Part Three	CYPRESS ST	42,2748	-71,1686	(42.27475809, -71.16...	3 803	M/V ACCIDENT
8	1	Wednesday	18	Part Two	DORCHESTER AVE	42,2784	-71,0664	(42.27839920, -71.06...	802	ASSAULT & BA
8	1	Wednesday	18	Part Two	DORCHESTER AVE	42,2784	-71,0664	(42.27839920, -71.06...	802	ASSAULT SIMF
8	1	Wednesday	19	Part One	MASSACHUSETTS AVE	42,3315	-71,0709	(42.33152148, -71.07...	619	LARCENY OTH

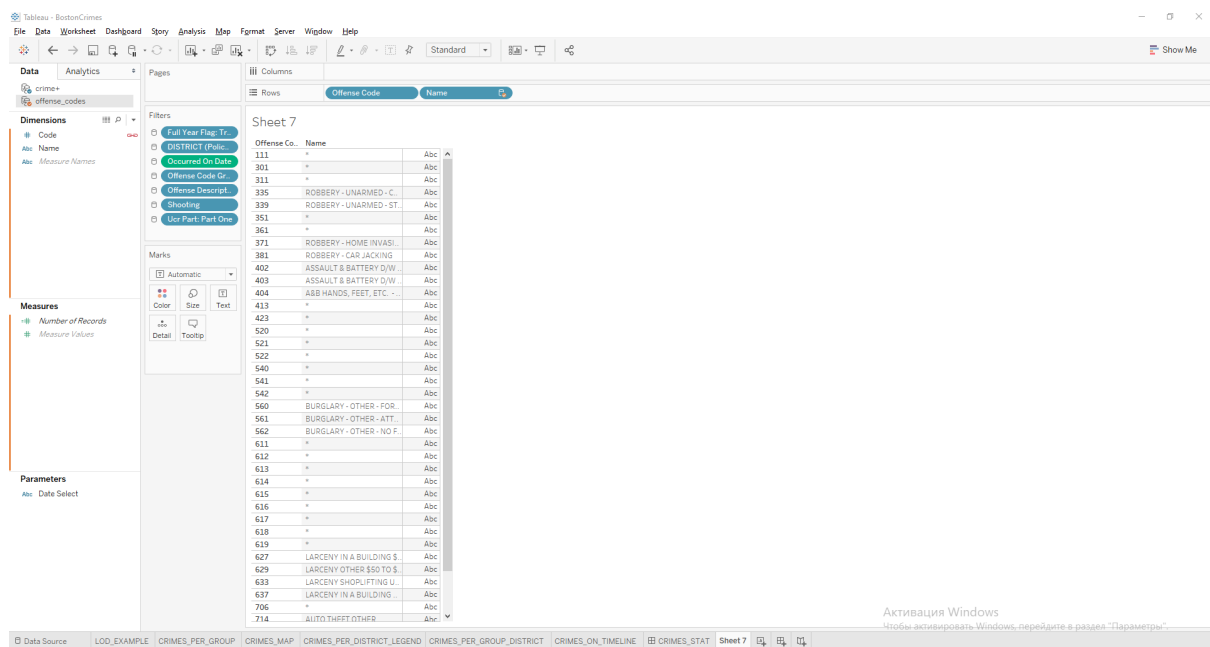
Далее переходим на наше рабочее пространство, нажав Sheet1. Попробуем проверить, действительно ли мы объединили данные нормально. Представим что мы видим этот датасет впервые и решили провести его непосредственный уже в самом инструменте. Кладем Code и Names на ROWS и видим что одному Code может соответствовать несколько Names. Такую проверку можно сделать различными способами. Это один из них. Второй - это трансформировать поле Names из Dimensions в Measures. Посчитать количество количество уникальных Names для одного Code. Для этого мы в разделе Data выбираем поле Names, нажимаем на него правой кнопкой мыши и ждем convert to measure. Это поле автоматически переместилось в раздел Measures и к нему была применена функция COUNT DISTINCT. Второй способ - это переместить поле Names в TOOLTIP MARK. Далее нажимаем на эту пиллюлю правой кнопкой мыши. Выбираем MEASURE и указываем COUNT DISTINCT. Можно также просто перетащить полученную пиллюлю в фильтры и указать условие по значению COUNT DISTINCT.



Второй - использовать DATA BLENDING. Это такая фишка у TD, когда данные из разных источников не объединяются в одну витрину. Таким образом, не будут появляться дубликаты. Такой способ хорошо подходит, когда используются различные источники данных, которые не надо объединять в витрину или для которых это сделать проблематично. Для это мы добавляем новый источник данных Data > New Data Source. Далее выбираем Data > Edit Relationship и указываем, по каким полям хотим производить объединение.

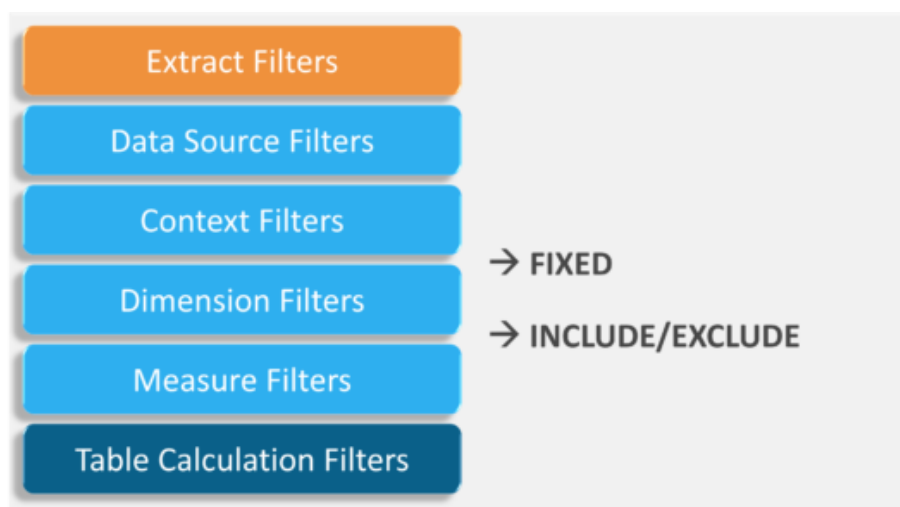


Однако, у Data Blending есть и свои минусы. Например, если одной строке из одного источника данных соответствует несколько строк из другого источника, то положив их на дашборд мы увидим \*. При этом часть функций TD может быть недоступна при Data Blending. Так у него есть проблемы работе с большими массивами данных. Таким образом, если позволяет время и есть возможность, то лучший вариант - это подготовить витрину данных заранее.



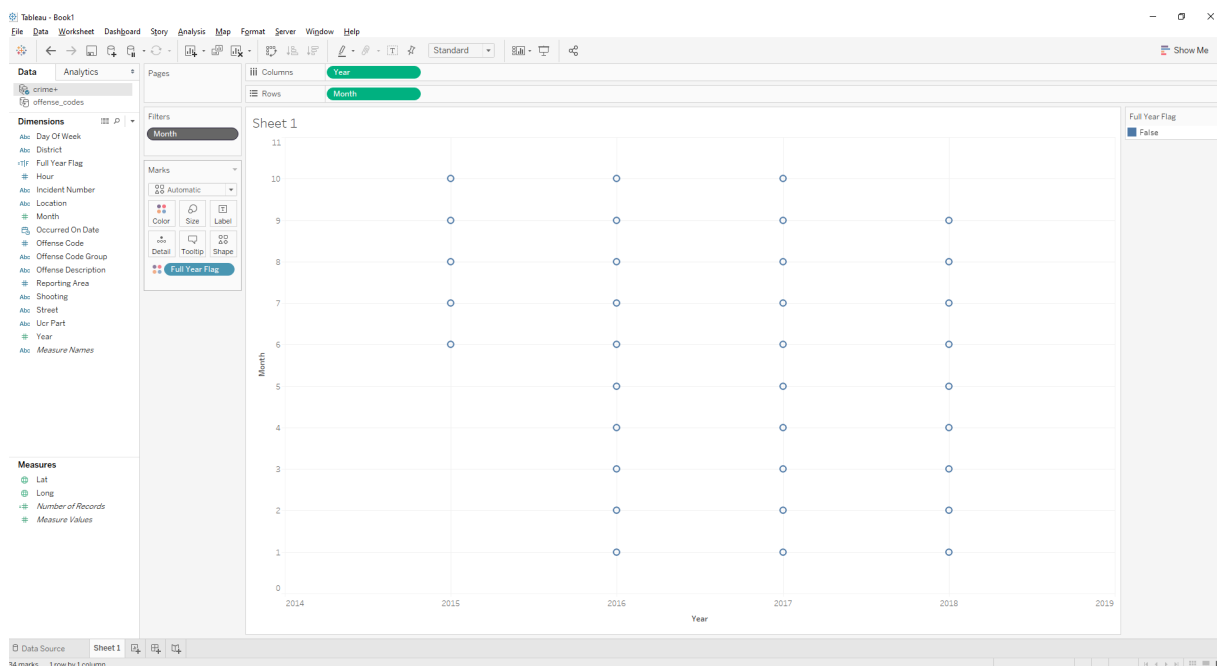
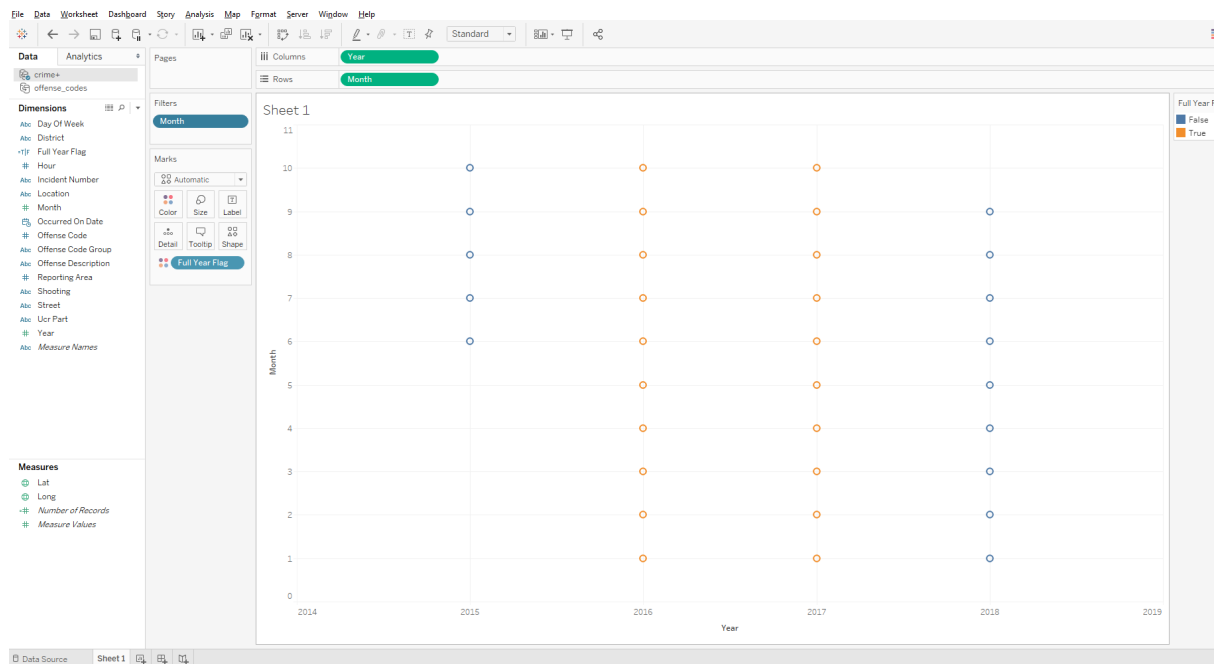
Третий - использовать LOD (LEVEL OF DETAILS EXPRESSION). Четвертый - это ACTION FILTER (последний подходит только для фильтрации данных). Обо всем этом мы поговорим на сегодняшнем занятии.

Итак, вернемся к созданию нашего дашборда. Сначала в DATA SOURCE FILTERS установим UCR = 1. UCR - Uniform Crime Reporting - это такая система по категоризации преступлений в США. Таким образом, мы будем рассматривать наиболее тяжкие преступления. Хочу сейчас немного подробнее рассказать про иерархию фильтров в TD. Итак, есть несколько видов фильтров. У каждого из них есть свой приоритет. Сначала идет extract filter, то есть фильтр, который применяется во время создания экстракта из источника данных. Далее идет data source filter - это фильтр, который применяется к нашему источнику данных (к экстракту или напрямую если экстракт не был создан). Далее идут Context Filters - они применяются к используемой витрине и влияют на все вычисления. В том числе на все LOD. Далее идут Dimensions Filters которые влияют на все вычисления за исключением LOD (Fixed). И завершают всю эту иерархию Table Calc Filters.



Итак, давайте ограничим наш анализ годами в которых есть статистика за 12 месяцев. Для этого можно посмотреть сколько месяцев есть в каждом году. Кидаем Year на COLUMNS и Months на ROWS. У нас есть три полных года. Эти года можно оставить при помощи фильтров. А можно создать вычисляемое поле, по которому будут исключаться месяцы, в которых меньше 12 месяцев. И тут нам как раз на помощь приходит LOD. Что он позволяет делать? Используя LOD можно задавать уровень грануляции. Их три вида: Exclude, Include, Fixed. Соответственно исключаем какой-либо из уровней грануляции, добавляем какой-либо из уровней грануляции, или фиксируем уровень грануляции, указывая необходимые поля. Попробуем использовать FIXED и оставить только нужные нам года. Для этого создаем новое вычисляемое поле Full Year Flag. Прописываем в нем { FIXED [Year]:COUNTD([Month]) } = 12. Кидаем на COLOR MARK. Работает корректно. Возвращаемся к иерархии фильтров. Они применяются последовательно в зависимости от уровня, к которому принадлежат. Для того, что бы сделать из DIMENSION FILTER CONTEXT FILTER необходимо в рабочем пространстве нажать правой кнопкой мыши на DIMENSION FILTER и выбрать ADD TO CONTEXT. Так как FIXED невосприимчив к DIMENSION AND TABLE CALC FILTERS, то, если мы хотим применить фильтры к нашей LOD калькуляции с FIXED - фильтр должен быть контекстным. Давайте применим фильтр Month к нашей визуализации. Исключим несколько месяцев. Применим. Нужные года по прежнему корректно выделены цветом. Как только мы сделаем этот фильтр контекстным, то выделение цветом станет некорректным, потому что условия вычисляемого поля стали применяться после применения фильтра. Также стоит

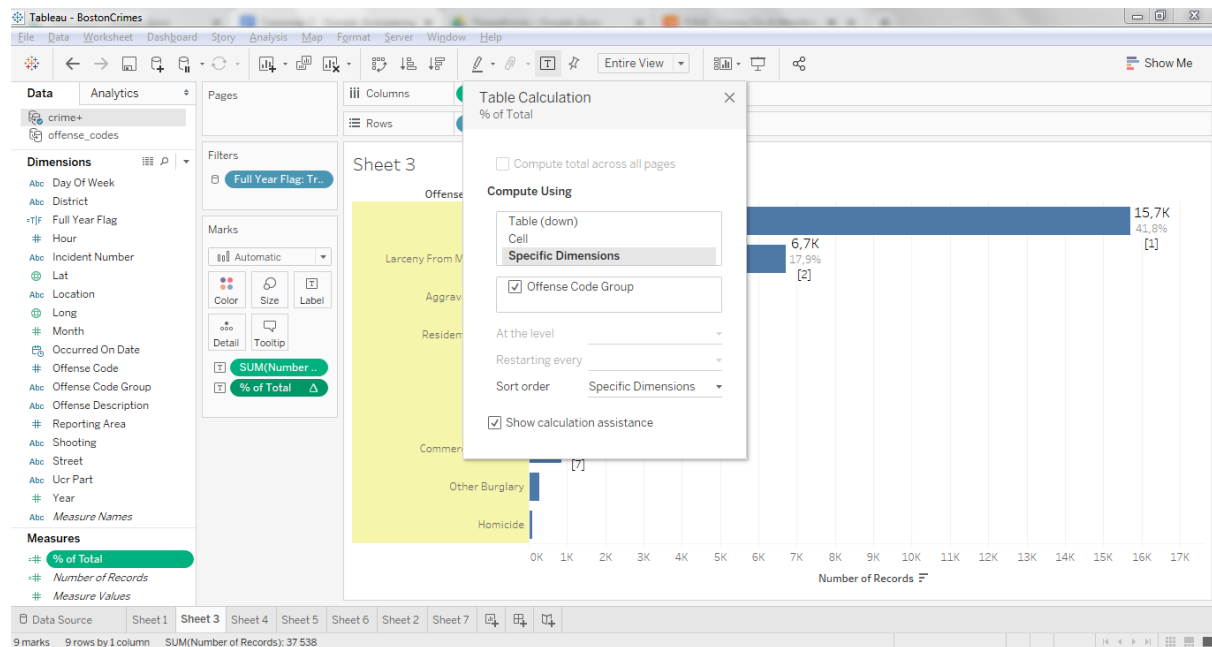
сказать что LOD не является сам по себе агрегированным значение. То есть он применяется построчно. При это сам результат LOD можно агрегировать. Что дает еще больше возможностей для различных операций.



Теперь давайте применим наше поле Full Year Flag ко всему источнику данных.

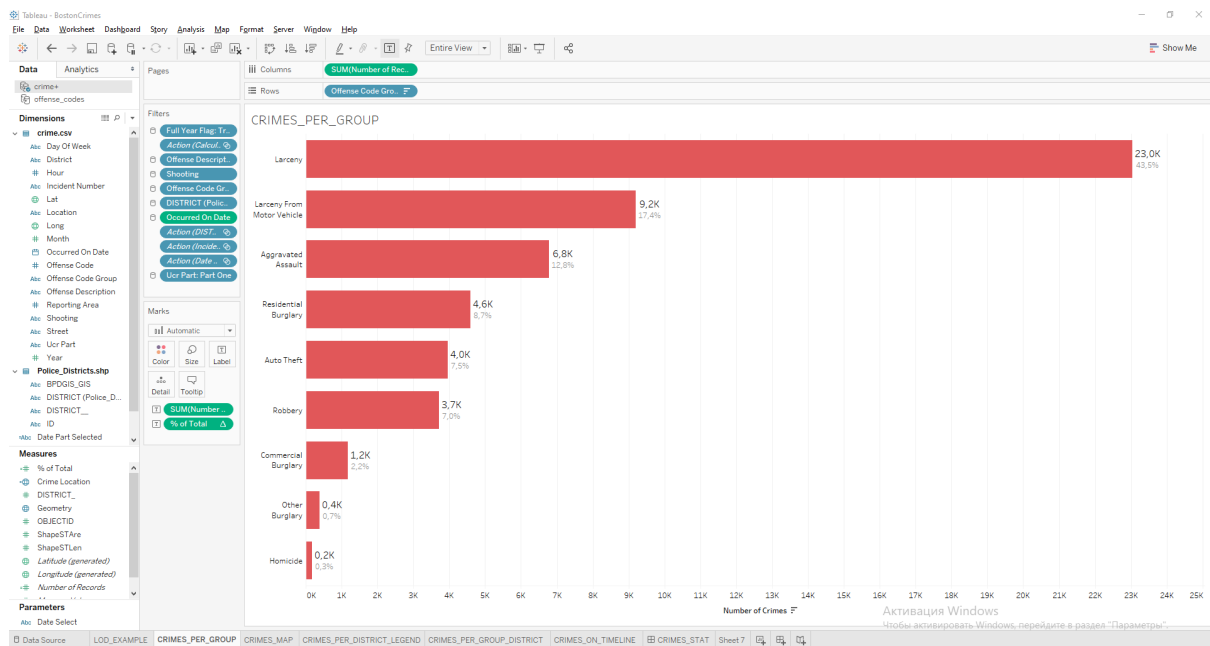
После того как мы узнали про LOD у нас появилось решение нашей обозначенной выше проблемы. Мы можем определить уникальный набор полей для каждой строчки и вычислять по нему, например, минимальное значение, таким образом не смотря на наличие дублей мы будем получать корректные значения.

Давайте взглянем на то, какой тип преступлений лидирует. Отправляем [Offense Code Group] на ROWS и Number of Records на COLUMNS. Дальше сортируем группы преступлений по количеству совершенных в порядке убывания. Добавляем количество преступлений на текст. При этом хотелось бы видеть еще долю в процентном выражении количества преступлений по группе от общего количества преступлений. Для этого давайте воспользуемся функцией Table Calculations. Что они делают. В Table Calculations вы выбираете уровень детализации относительно данных которые представлены на визуализации. Есть несколько способов сделать Table Calculations. Первый - создать новое вычисляемое поле. Назовем его % of Total Crimes. Table Calculation работает только с агрегированными функциями. Поэтому мы сначала создаем агрегированное вычисление и дальше заворачиваем его в Table Calculation.  $SUM([Number of Records]) / TOTAL(SUM([Number of Records]))$  - вот примерно так. Что мы делаем дальше - кладем эту пилюлю на текст. Затем жмем на нее правой кнопкой мыши и выбираем Edit Table Calculation. Тут мы определяем поля, по которым будет применяться наша функция Total. Выбираем Offset code groups.



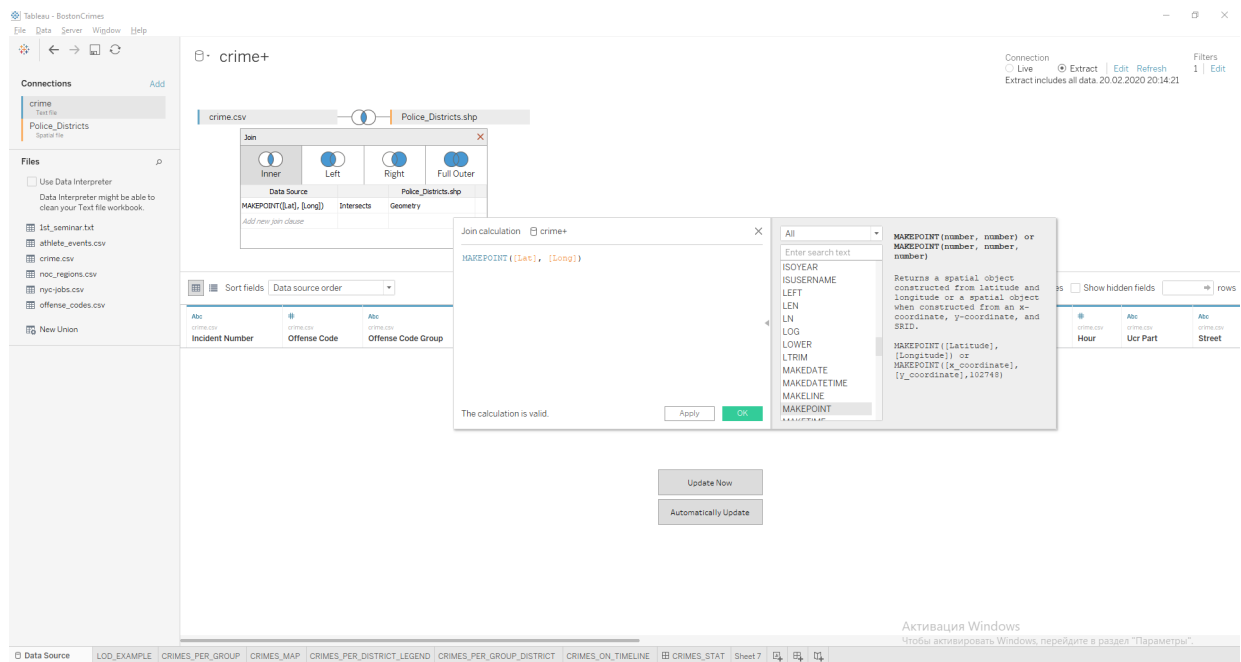
После чего закрываем наше окно и получаем нужную калькуляцию. Также есть возможность автоматически трансформировать вычисление в Table Calculation. Для этого нажимаем правой кнопкой мыши на нашу пилюлю и выбираем либо Add Table Calculation, либо Quick Table Calculation. Дальше выбираем Percent of Total и получаем аналогичный расчет. Соответственно если по какой-то причине расчет предложенный Tableau нам не подходит, то мы всегда можем его скорректировать. Чтобы это сделать необходимо нажать правой кнопкой мыши на пилюлю и выбрать Edit Table Calculation.

Далее открываем Text Mark. Редактируем цвет и размер наших подписей.

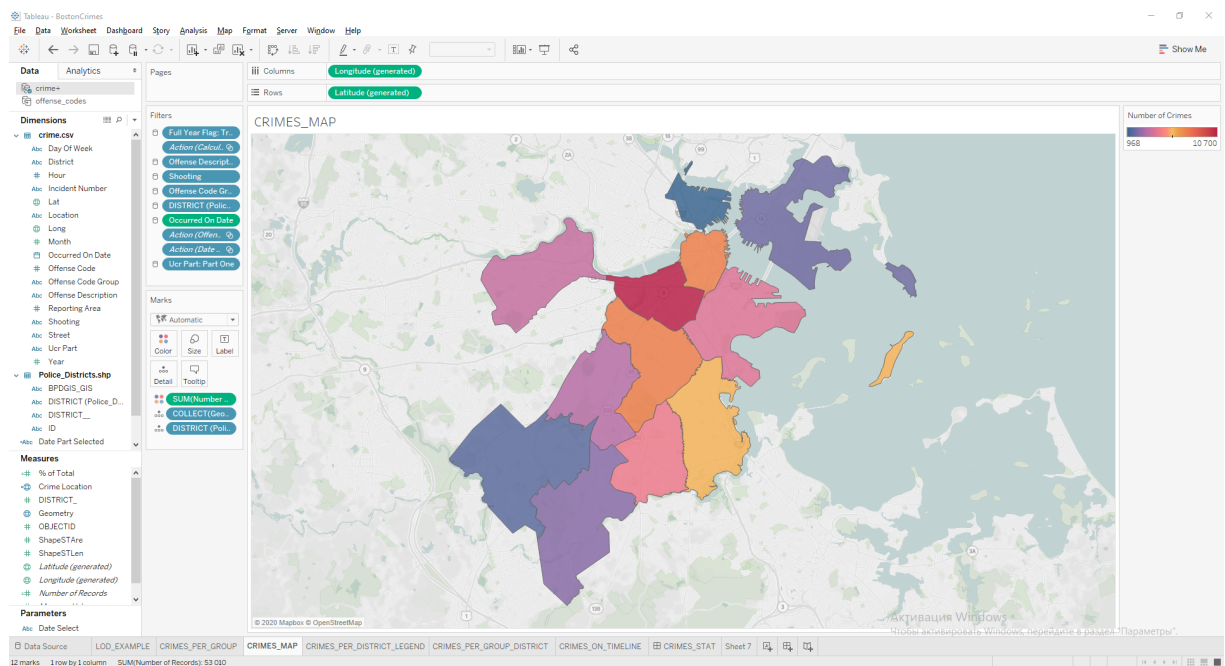


Теперь попробуем создать карту города, на которой будут представлены очаги по количеству преступлений с цветовой раскраской районов в зависимости от количества этих преступлений. TD довольно хорошо работает с геофайлами. Давайте сначала проверим все ли записи имеют территориальную принадлежность. Смотрим какие есть записи в поле District. Там есть Null. При этом видно, что точки находятся в черте города. Давайте добавим в нашу витрину данные из еще одного источника. Переходим на страницу источника данных. Выбираем add connection > spatial file. Далее как я уже говорил можно выбирать по какому ключу объединять данные из двух источников. Можно использовать одно из уже имеющихся полей. А можно создать свое вычисляемое поле. Давайте создадим точку при помощи функции MAKEPOINT и будем объединять по принципу вхождения созданной координаты в полигон. Для этого в качестве условия объединения указываем intersect.

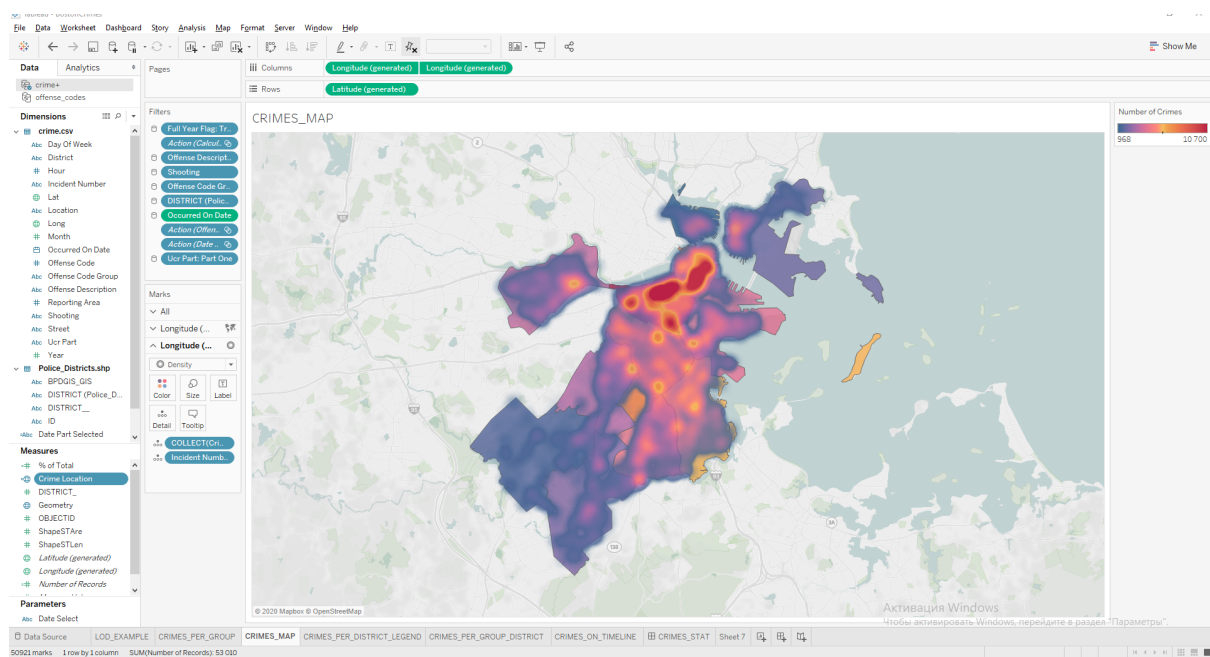




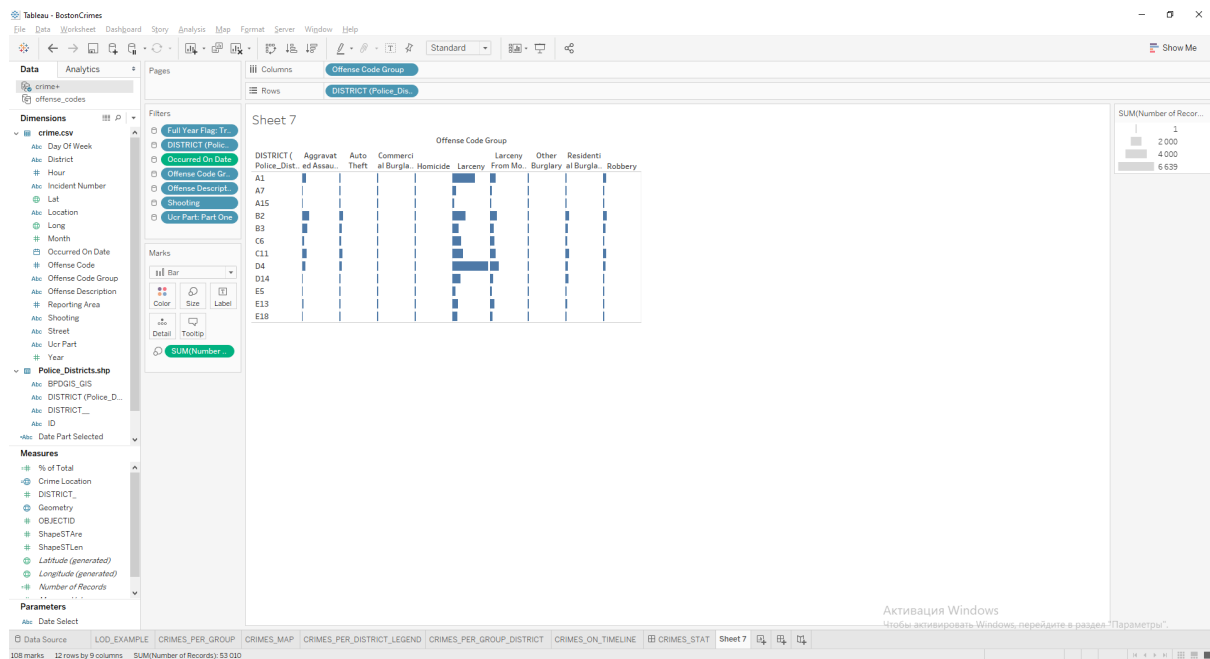
Теперь возвращаемся в наше рабочее пространство. Добавляем наши полигоны (GEOMETRY) на Details Mark, Добавляем количество записей на цвет. Выбираем соответствующую цветовую раскраску, выбрав Edit Colors.



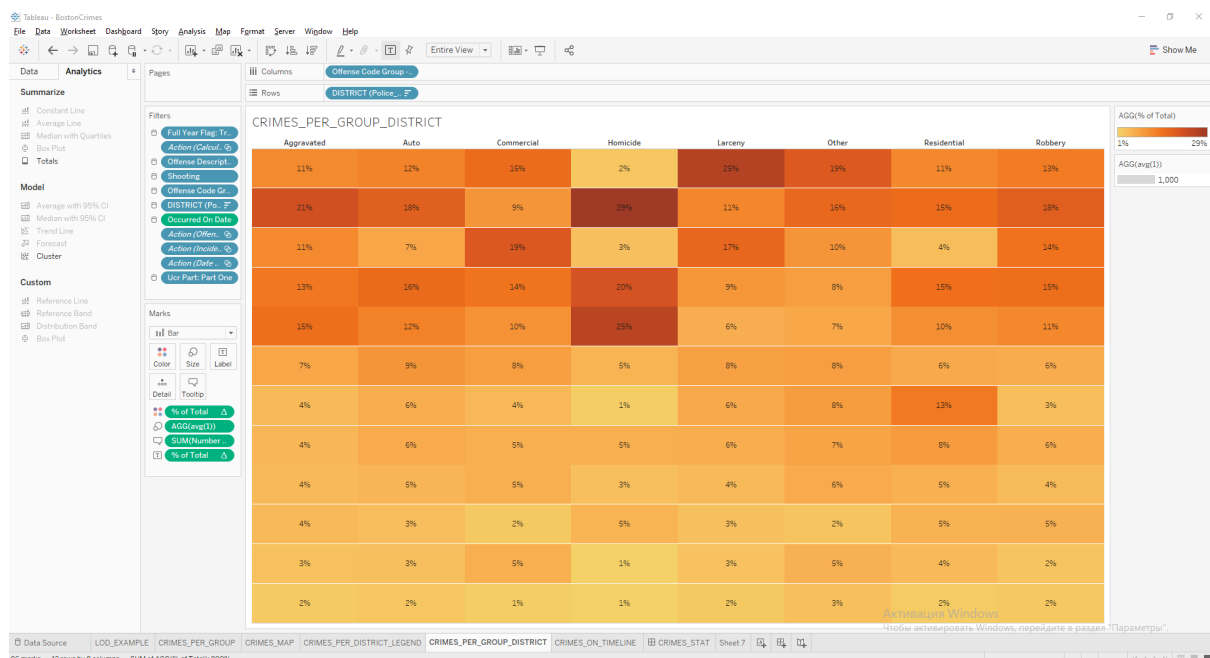
Также давайте отобразим плотность преступлений. В TD можно создавать многослойные карты. Поэтому давайте сделаем второй слой, на котором будет отображаться плотность. Кидаем еще одни longitude generated и вместо map в marks выбираем density. Для details давайте создадим вычисляемое поле crime location с координатами места преступления. MAKEPOINT([Lat],[Long]). В Marks вместо maps выбираем density и кладем crime location на карту. Далее настраиваем цвет, прозрачность и саму карту. Первое и второе можно сделать в color marks. Третье в панели меню сверху в разделе map > map layers.



Для большей наглядности распределения видов преступлений по районам давайте сделаем heatmap. Что там будет цветовая легенда районов в качестве заголовка. Она будет совпадать по цветам с той, что представлена на карте. И доля каждого из типов преступлений, которая приходится на этот район. Также добавим колонку тотал, что видеть долю от общего количества преступлений, который приходится на отдельно взятый район. Для этого нам надо сделать два вкешита. Один - с заголовками таблицы (районами). Второй - с распределением. Чтобы построить такую историю есть два варианта. Первый - когда наша таблица должна быть полностью закрашена. Второй - когда должны быть пробелы. Для первого способа мы выбираем в Marks Square, кидаем District on Rows, кол-во записей на цвет. У нас автоматически закрашивается все пространство. Второй вариант - это выбрать Bars и положить некое значение на size. Например, так можно сделать Bars Table. Так, давайте возьмем и положим группы преступлений на колонки, районы на поля. Выберем Bars в Marks и положим сейчас кол-во записей на Size.



Как видим закрашка происходит в соответствии с количеством записей. То есть если мы хотим закрасить эти поля одинаково, нам надо создать такое поле, которое будет одинаковым для всех наших групп. Это поле - AVG(1), например. Дальше повторяем настройку table calculation. Выбираем District, поскольку нас интересует распределение по районам. Кстати, TD нам подсказывает, каким образом будет происходить вычисление, подсвечивая визуализацию во время настройки table calculation.



Теперь давайте добавим распределение всех преступлений между районами. Это можно сделать при помощи одной из функции во вкладке аналитикс под названием тотал. Соответственно эта функция считает итоги по заданному одному из заданных принципов. Для этого надо нажать во вкладке аналитикс на тотал и перенести его на

визуализацию. Указать соответственно row grand total потому что мы хотим посмотреть распределение по районам. Добавляем текстовые подписи и наша визуализация готова.

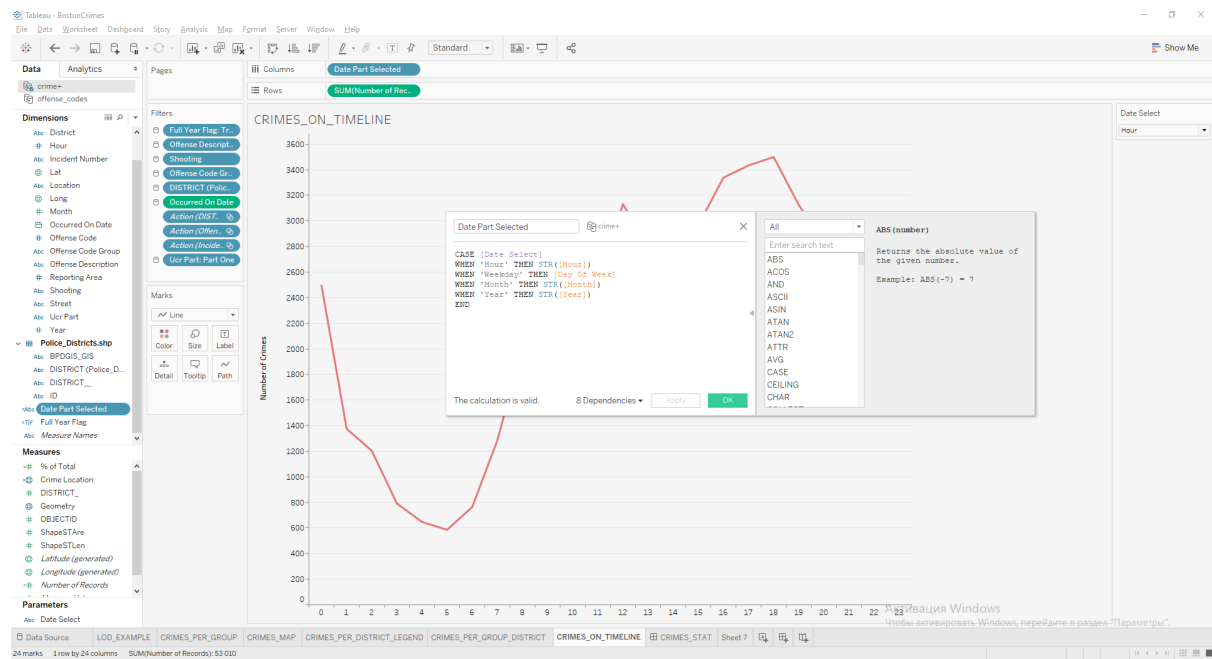


Теперь нам желательно посмотреть распределение преступлений по времени и датам. Для этого я предлагаю воспользоваться параметром. Параметр - созданное поле, которое применяется для всего источника данных. В этом поле вы можете самостоятельно выбирать и задавать значения и указывать тип данных. Укажем необходимые нам срезы данных.

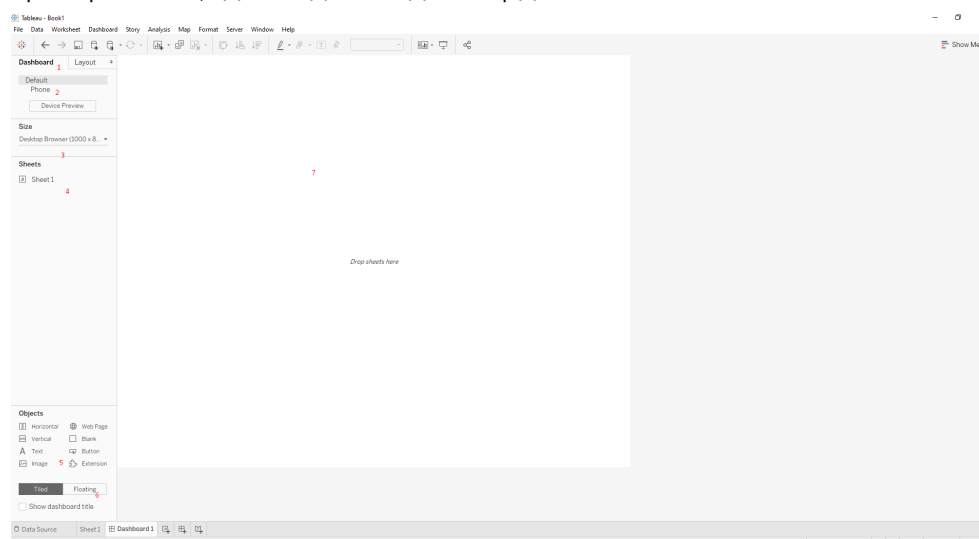
Далее создаем вычисляемое поле

```
CASE [Date Select]
WHEN 'Hour' THEN STR([Hour])
WHEN 'Weekday' THEN [Day Of Week]
WHEN 'Month' THEN STR([Month])
WHEN 'Year' THEN STR([Year])
END
```

Далее создаем line chart и наш график готов.



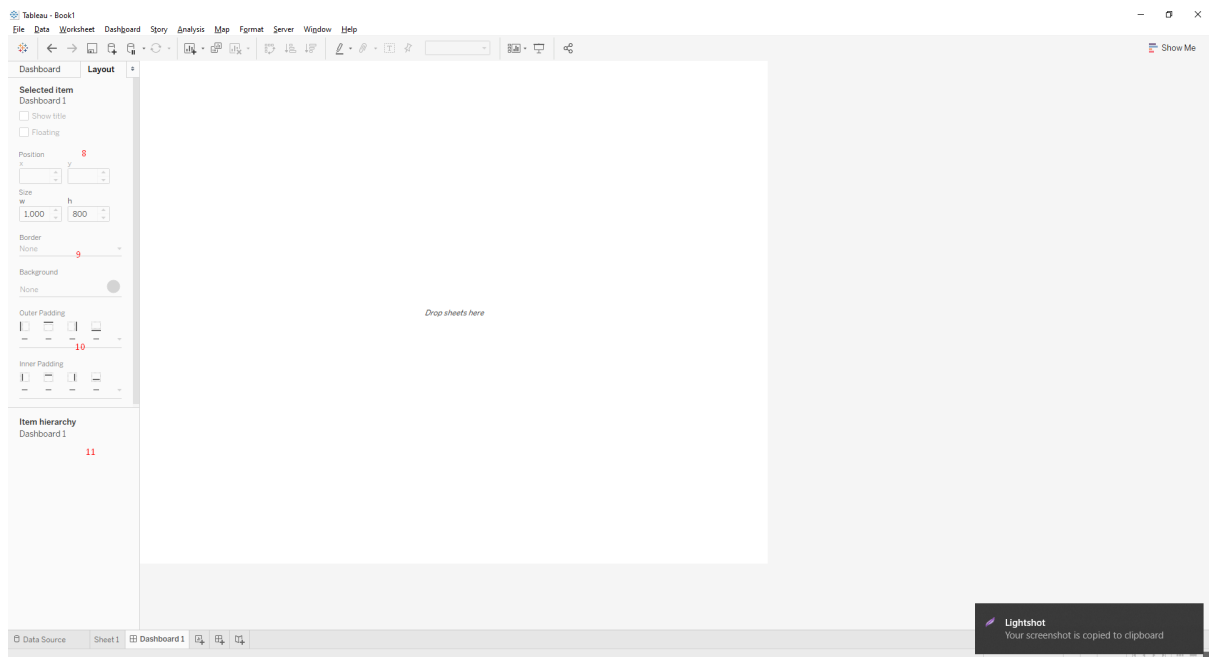
Теперь давайте подробнее затронем тему создания дашборда. Итак, в рабочем пространстве, где создается дашборд есть:



1. боковая панель слева . Она поделена на две части: dashboard и layout. Первая отвечает за создание элементов и размер самого дашборда, вторая менеджерит структуру, расположения элементов, размер отступов. Справа находится пространство, в котором будет создаваться дашборд.
2. настройка представления, при помощи которого вы можете посмотреть как будут представлен дашборд на каждом устройстве, задав его параметры
3. настройка длины и ширины дашборда. можно задать определенные диапазоны изменения масштаба, фиксированный масштаб или автоматический.
4. перечень воркшопов, которые мы можем использовать для создания дашборда
5. Перечень дополнительных объектов которые мы также можем использовать во время создания дашборда.
6. Определение типа объекта будет ли он плавающим или привязан к структуре
7. Пространство, на котором мы будем размещать наши объекты, и которое впоследствии станет нашим дашбордом

После переключения со вкладки дашборд на вкладку лэйаут

8. Детальная настройка позиции элемента на дашборде
9. Настройка границ и заливки
10. Настройка отступов (внешних и внутренних)
11. Структура дашборда



Итак, что важно знать при создании дашборда:

1. Показывать только необходимую информацию, то есть ту, которая прямо отвечает на поставленные вопросы и решает конкретную задачу. Наиболее просто и доступно. Отсюда мы постепенно переходим в п. 2
2. Показывать ограниченное количество информации на одном визуальном пространстве. Как правило не более 5 визуализаций, с которыми ведется работа. Визуальное пространство - это пространство, на котором

[illegible]

- [illegible]

Вооружившись этими знаниями перейдем к самому процессу создания дашборда. У нас есть общая структура. Легче всего ее представить в качестве контейнера, который мы наполняем разными объектами. В том числе, мы можем вложить в нее один и более контейнеров. Объекты, которые перемещена в дашборд в качестве tiled занимают какую-то конкретную часть от общего пространства. Объекты floating накладываются сверху и далее друг на друга свободно перемещаясь по площади

дашборда. При этом в самом объекте можно также произвести распределение по его площади.

Давайте сначала зададим общий масштаб дашборда и определим количество элементов, которые будут на нем представлены. Там будет заголовок, четыре графика, у каждого из которых также будут заголовки. Соответственно нам нужно сделать отделение общего заголовка для дашборда, заголовки графиков от графиков и самих визуальных блоков друг от друга.

Для этого мы будем использовать несколько контейнеров. В один мы положим наш заголовок дашборда и фильтры, в другие наши визуальные блоки. То есть вертикальный (заголовок + фильтры, визуальные блоки) Второй вертикальный будет разбит на 2 вертикальных, в которых будут находиться по 2 наших графика. Каждый из этих графиков состоит из заголовка и визуализации. В одном случае добавляется параметр. Проще всего создавать такие элементы, когда используются несколько контейнеров за счет промежуточных текстовых вставок, они помогают проще распределить объекты внутри контейнера.

Далее убираем стандартные отступы табло и настраиваем свои собственные. Какого универсального варианта нет, но желательно чтобы под заголовком можно было проследить некую прямую по отношению к объектам ниже. Это можно сделать выровняв его по линии оси или краю таблицы или карты. Далее настраиваем actions highlights. Так, чтобы наши графики фильтровались и подсвечивалась таблица и карта.

Вы можете вносить любые другие изменения и дорабатывать дашборд так, как посчитаете. Важно экспериментировать и пробовать новое. Все это - опыт и знания.

