CIS 5270

Tableau Project

**BUILDING PERMITS**

Hemamalini Madhanguru

**Data set URL’s**

The data set that has been taken deals with the building permits of a certain area. The data set is of 130MB size which has a total of 131 columns and more than 450000 columns. The data set has the following columns - permit number column which has all the numbers of the permits that got issued, Permit type column deals with the type of permits, Issue date has the dates on which the permit type got issued, Estimated cost is the column which holds the estimated amount that need to be paid per permit, Amount Waived columns holds the amount that is waived off, Amount Paid columns contains the amount that is actually paid, Total fee column contains the sum of amount waived column and amount paid column and the work description column which tells about the type of work that has been done. The data set also contains the columns which hold the details regarding the location like street name, street number, street direction, latitude, longitude and Pin.

The dataset also has the columns which hold the details of the contractor. The column Contractor type has the details about the type of contractor he is. The contractor name column has the names of the contractor and the dataset also has the columns filled with the details regarding the address, city, zip code, phone number. The dataset has the details of ten contractors who are numbered as contractor 1 to contractor 10. This dataset is big enough to analyze the data using almost all the available techniques.

I am using the Tableau software in order to analyze the data and given useful information.

**Data Set Link:**

<https://data.cityofchicago.org/Buildings/Building-Permits/ydr8-5enu>

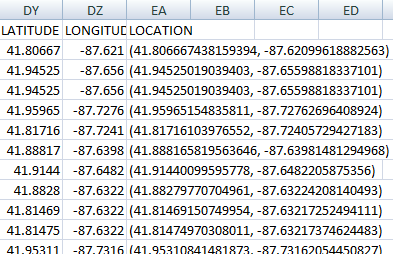
**Data Cleaning:**

The dataset has been cleaned using the below techniques:

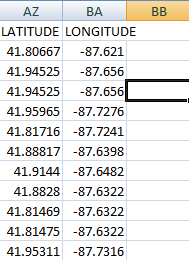
**Duplicate data:**

The dataset has same data in two columns. The column location has the data which is already present in latitude and longitude column. So, the dataset has been cleaned by removing the duplicate data.

**Before Data Cleaning:**



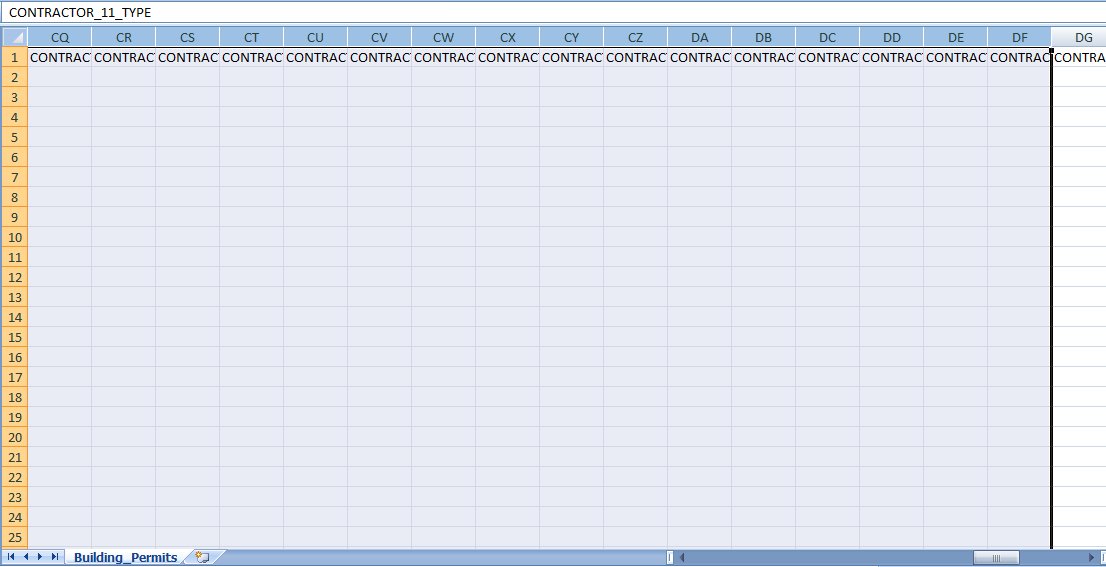
**After Data Cleaning:**

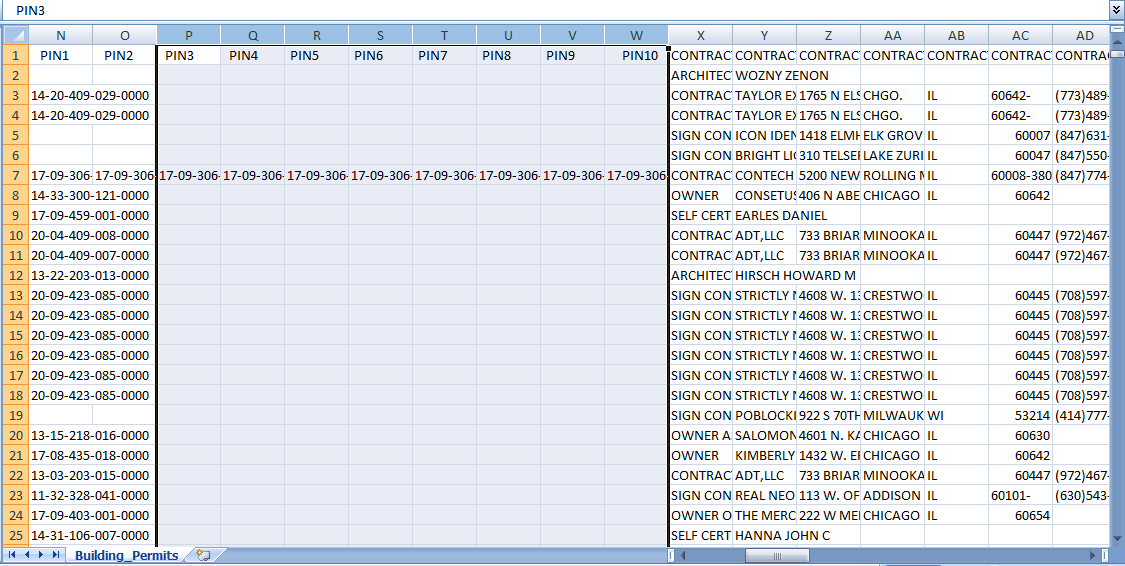


**Missing Values:**

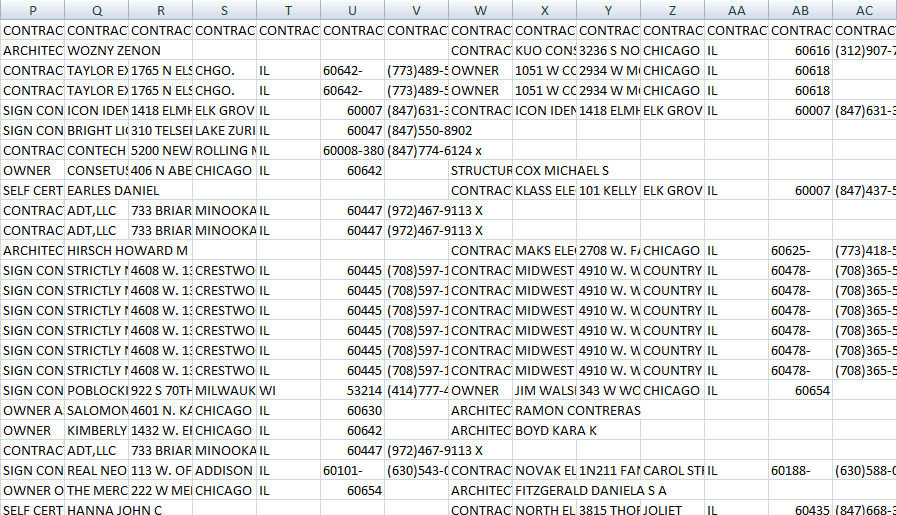
The data set has the columns with missing values in it. So, the columns are deleted as the columns are of no use.

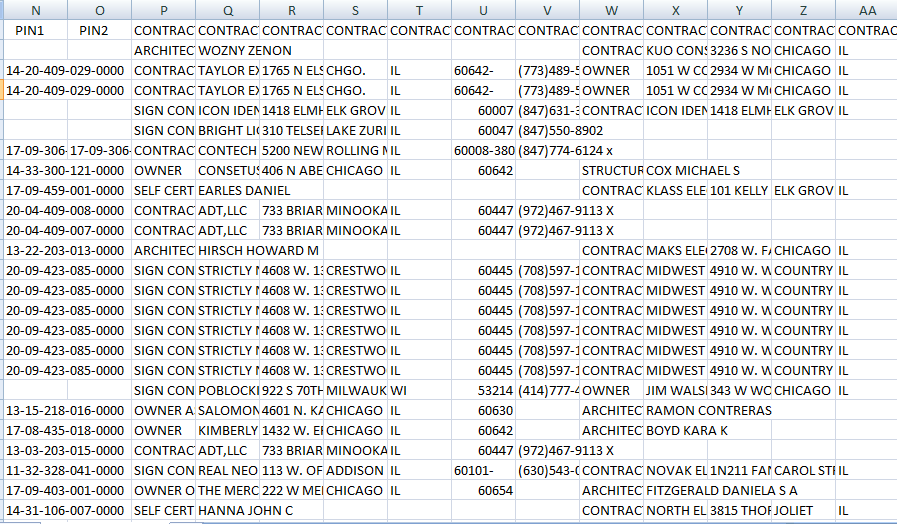
**Before Data Cleaning:**





**After Data Cleaning:**

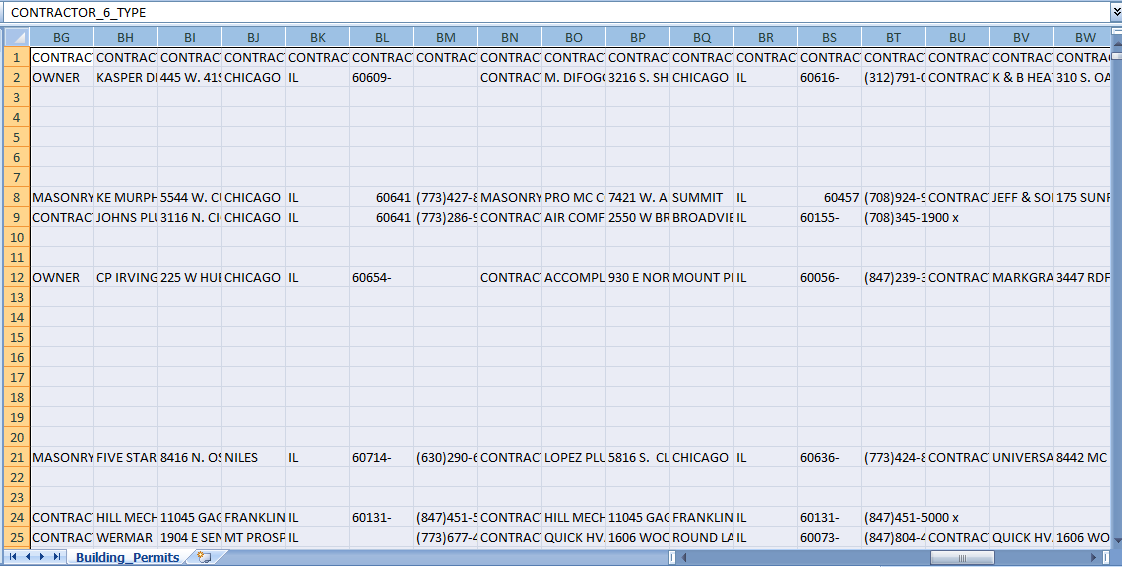




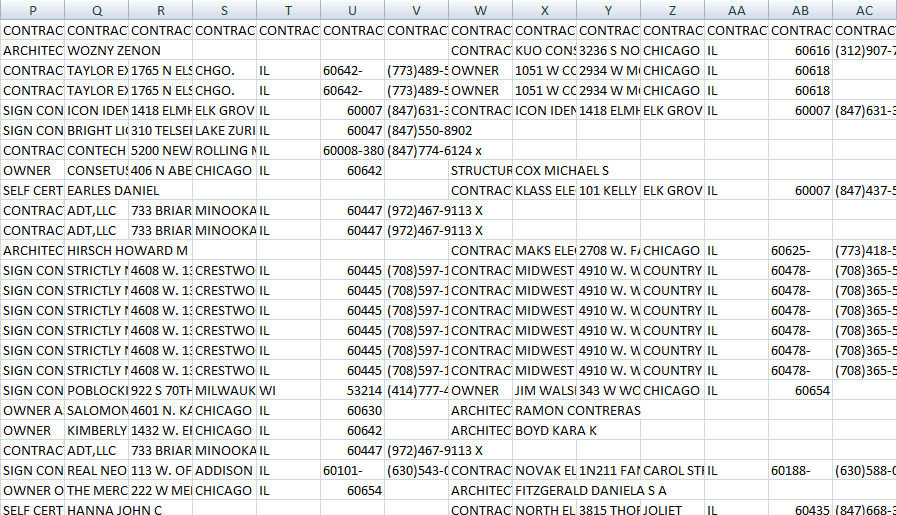
**Irrelevant Data:**

The dataset has the columns which contains data which is not useful. So, the columns have been removed from the data set

**Before Data Cleaning:**



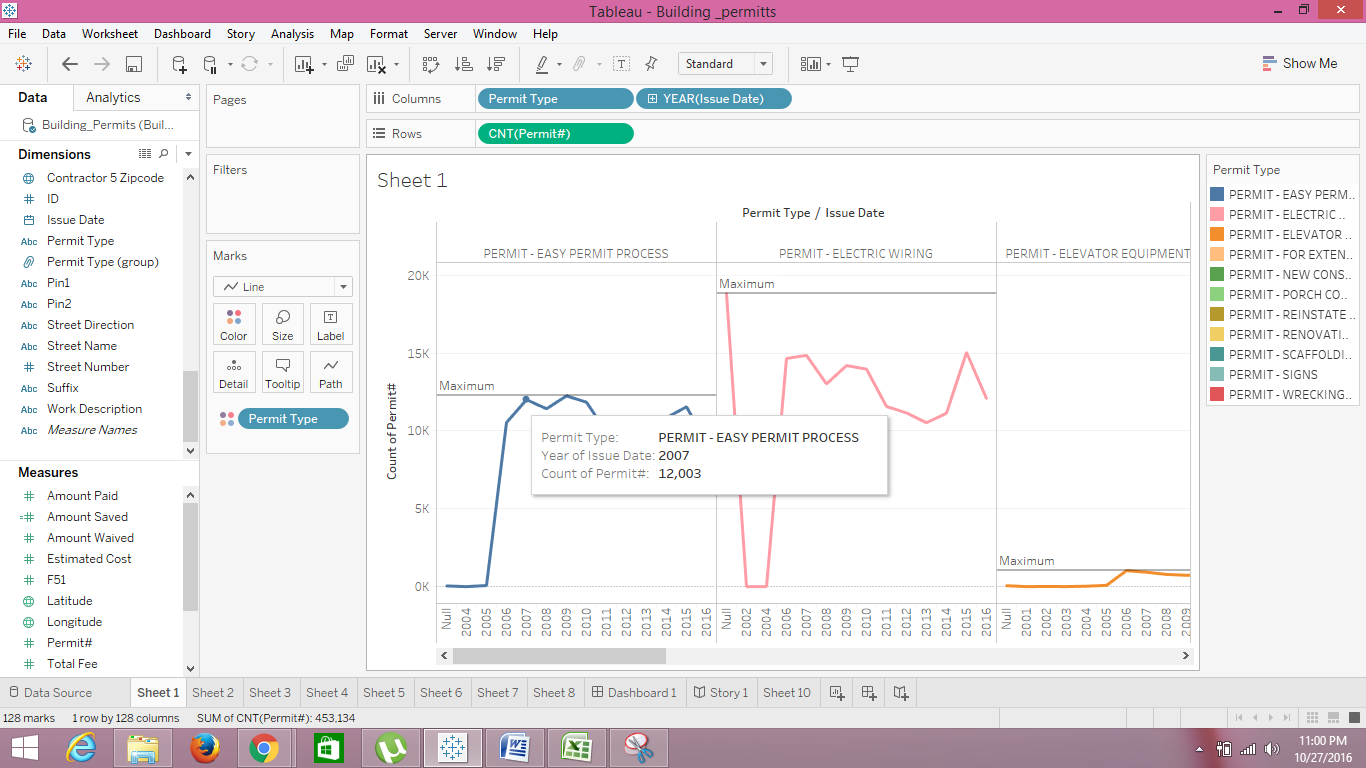
**After Data Cleaning:**



**Data Visualizations:**

1. Show a graph that has the flow of permits issued and also find in which year there are maximum permits issued for each permit type.

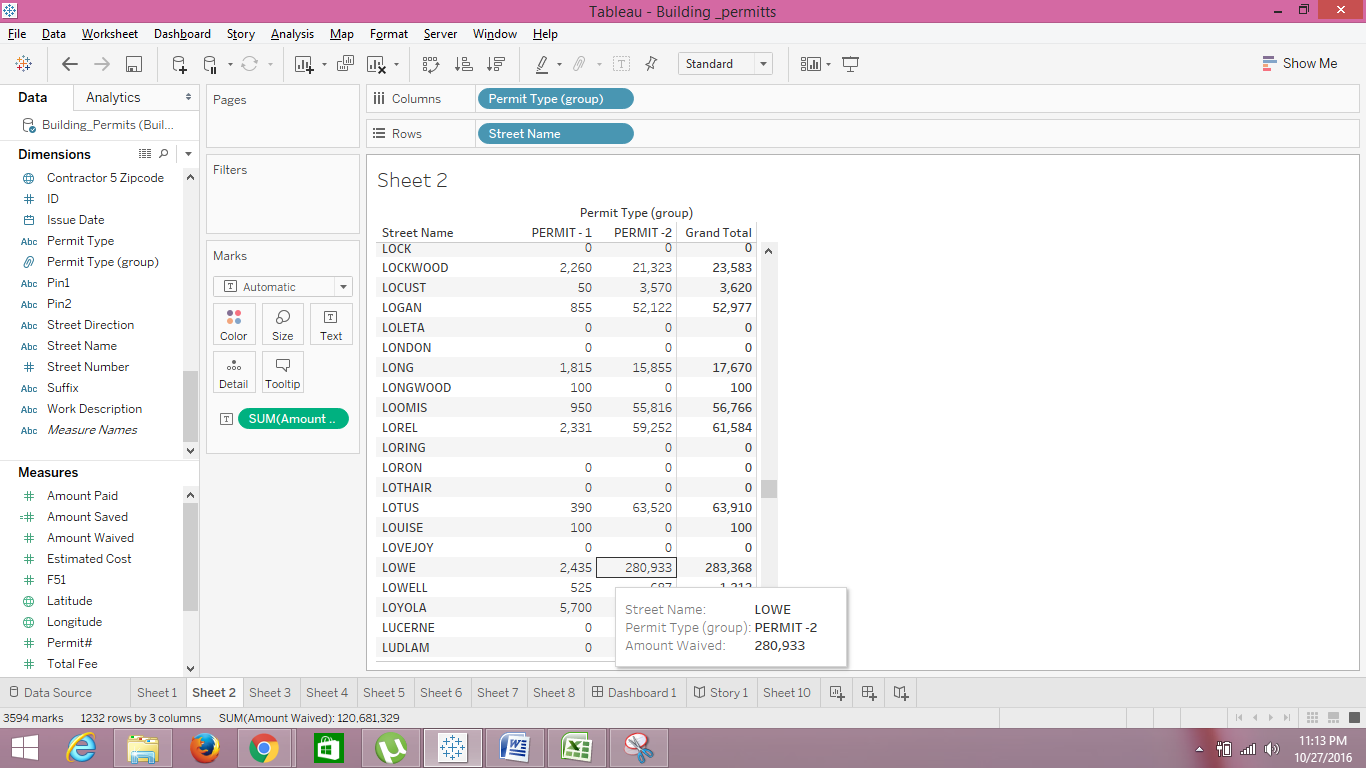
Solution:



The above visualization tells me the flow of permits that got issued for a range of years. I can see that the permit count has been taken from the years 2002 to the year 2006. Each Permit type is shown with different color and is separated with a vertical line. The reference line helps me to get the year which has maximum number of permit that are issued.

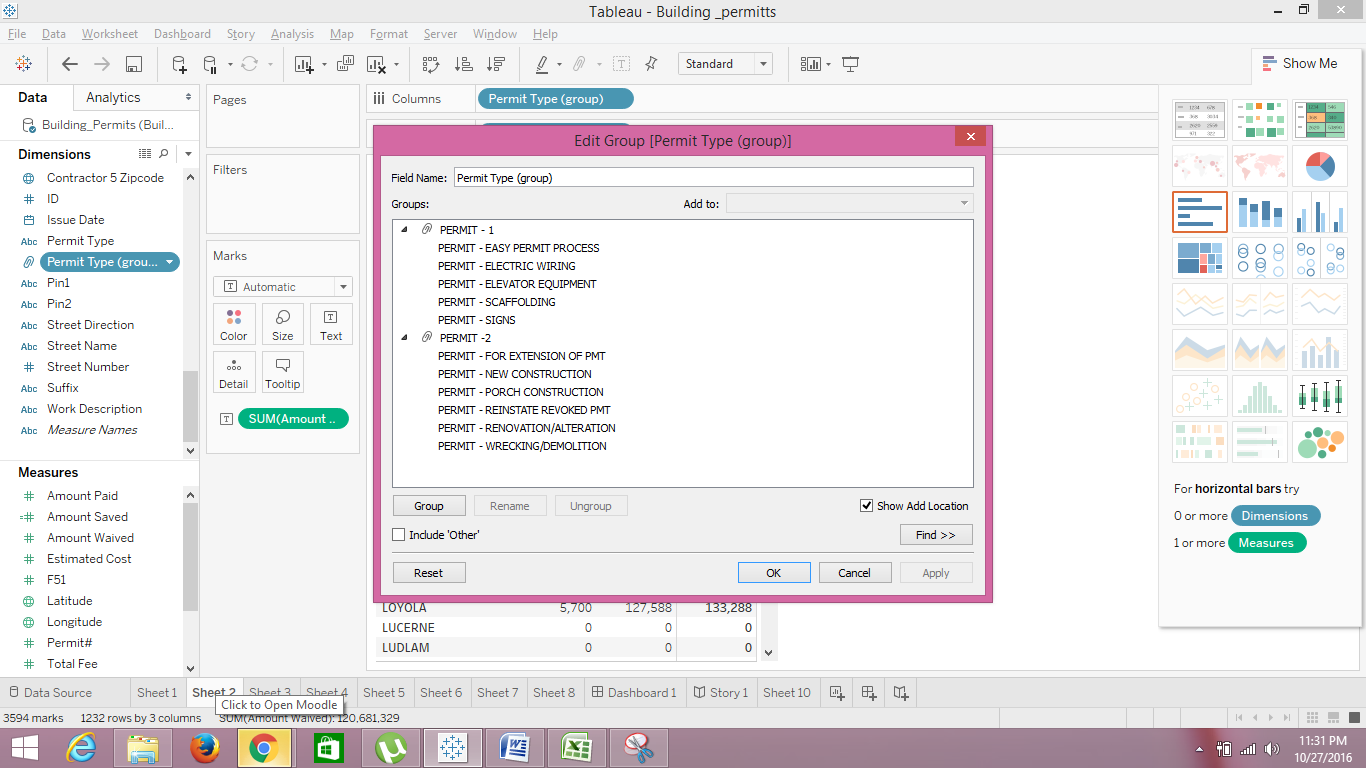
1. Compute the grand total of amount waived for all streets. Group the permit types which are related to as Permit -2 and others as Permit -1.

Solution:



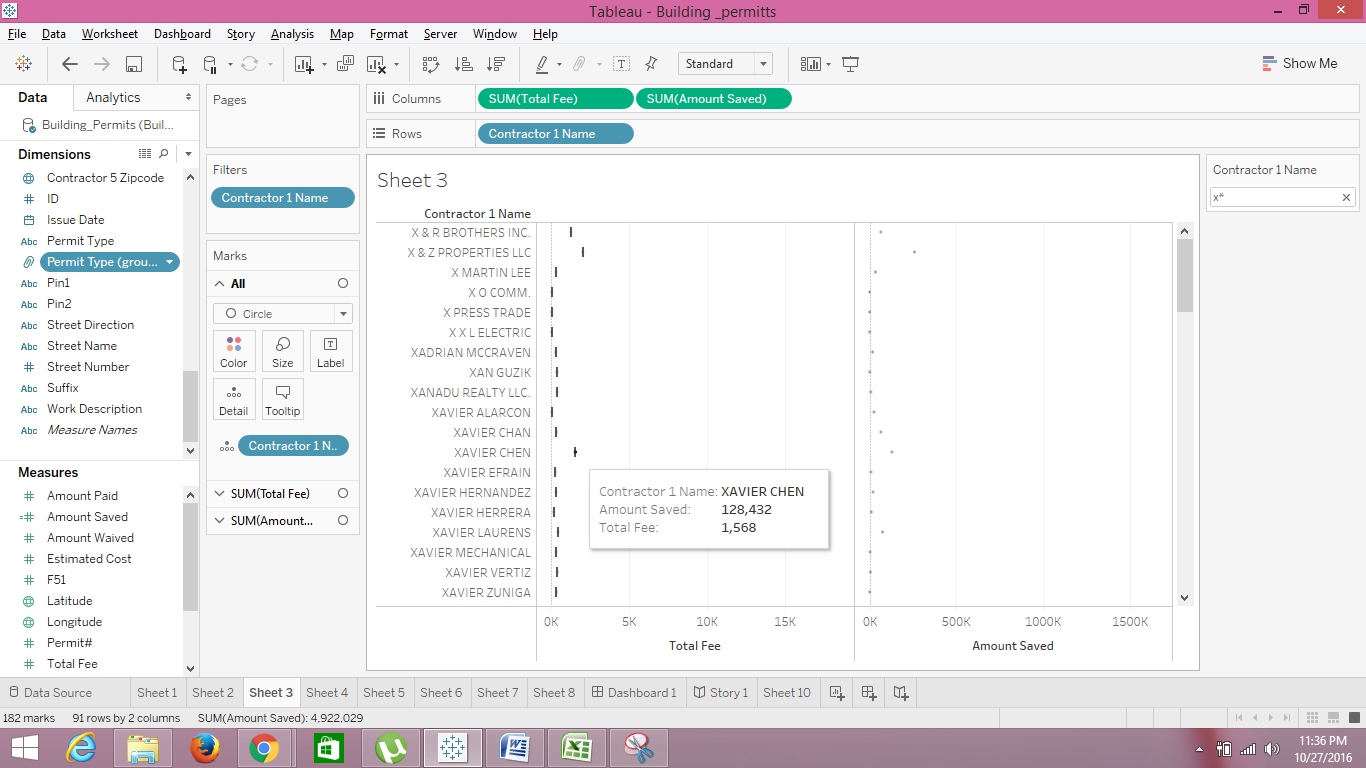
In the above visualization, I can see that the permit type has been grouped into Permit-1 and Permit-2. The Permit-2 group has the permits that come under construction and others are grouped as Permitt-1. The amount waived for each street has been shown clearly based on the groups that are created. I can also see the grand total of amount waived for each street.

Grouping:



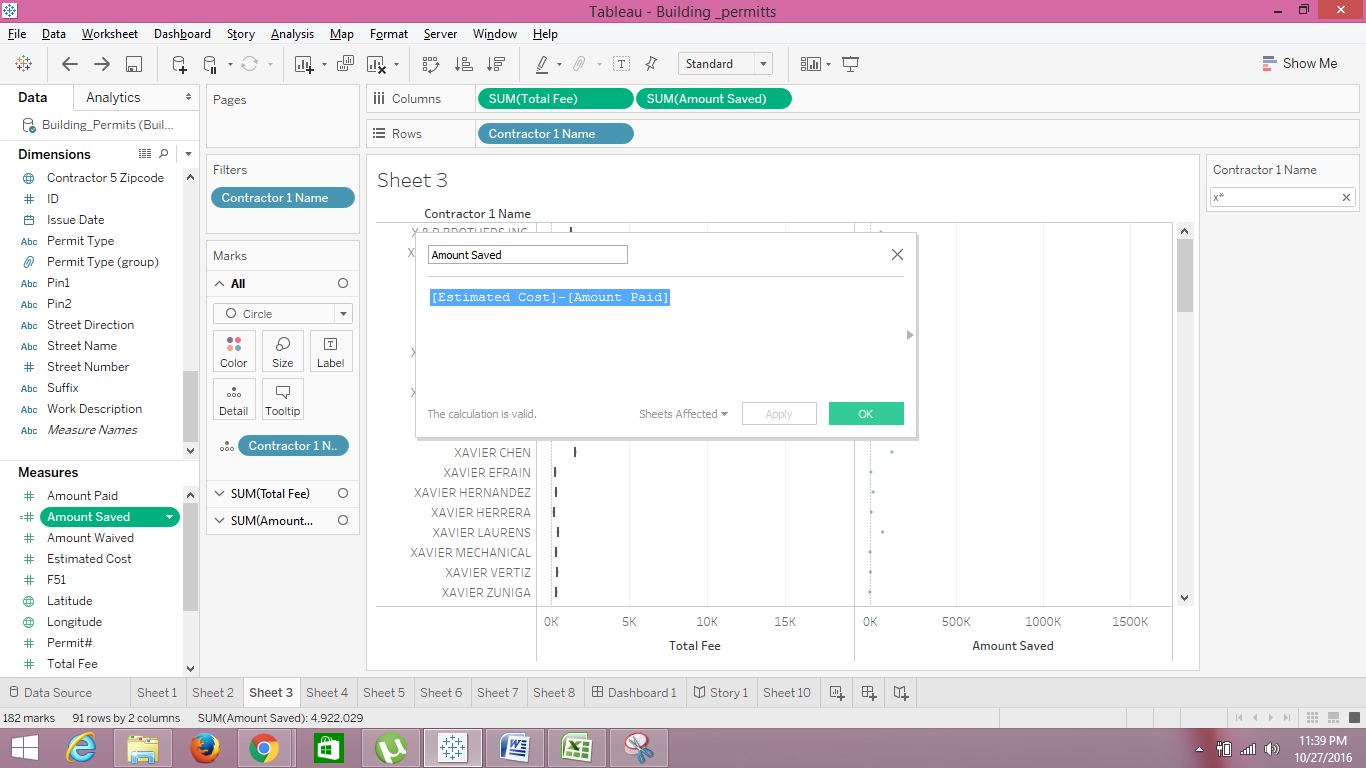
1. Compute the sum of total fee and amount saved of contractor-1 whose names starts with “x”. Show the result using box and whisker plots.

Solution:



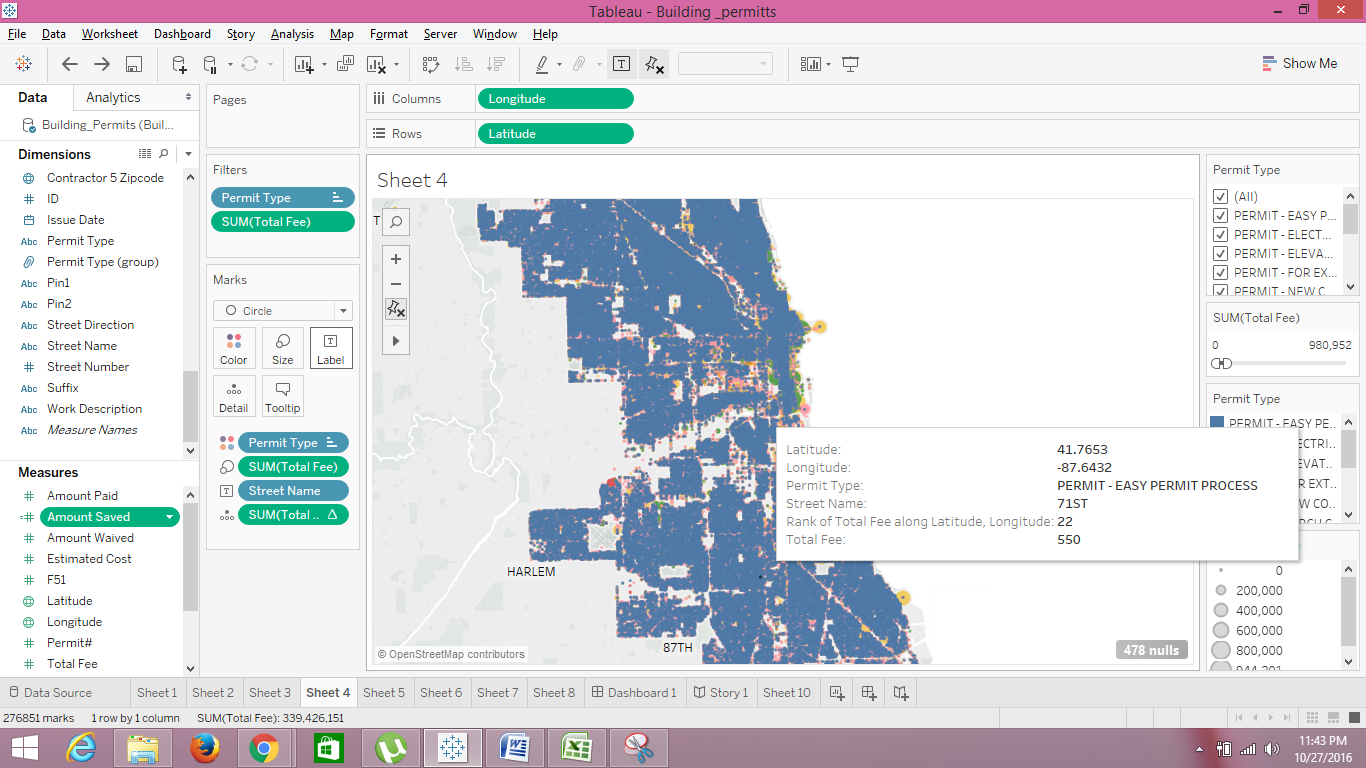
In the above visualization, I can see the contractor 1 name is filtered and only the names which start with “x” are shown. I can see the total fee and the amount saved for each contractor1. The values amount saved are generated using calculated field option. The formula used in order to get the amount saved is “[Estimated Cost]-[Amount Paid]”.

Calculated Field:



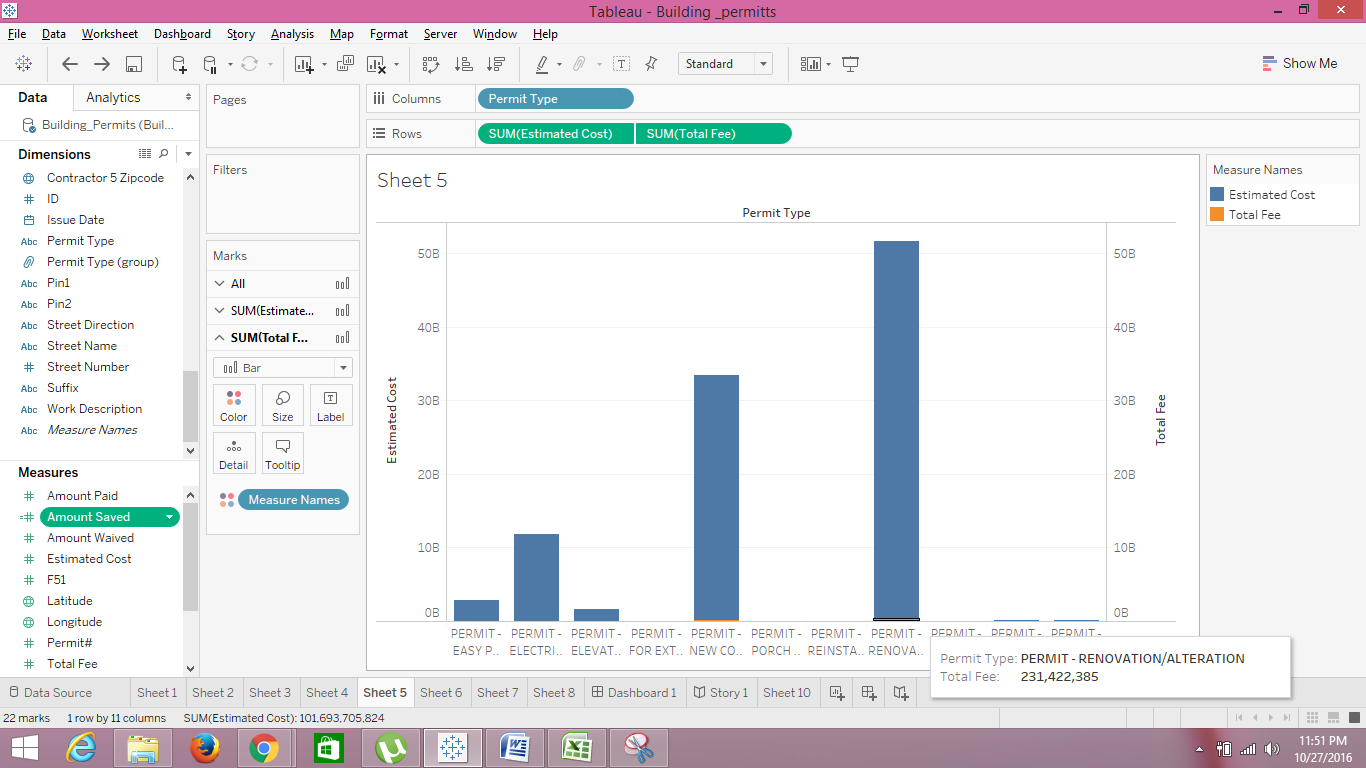
1. Show the geographical area where the permits issued. Also show the sum of total fee ranges from 0- 980952 and rank the state based on total fee.

Solution:



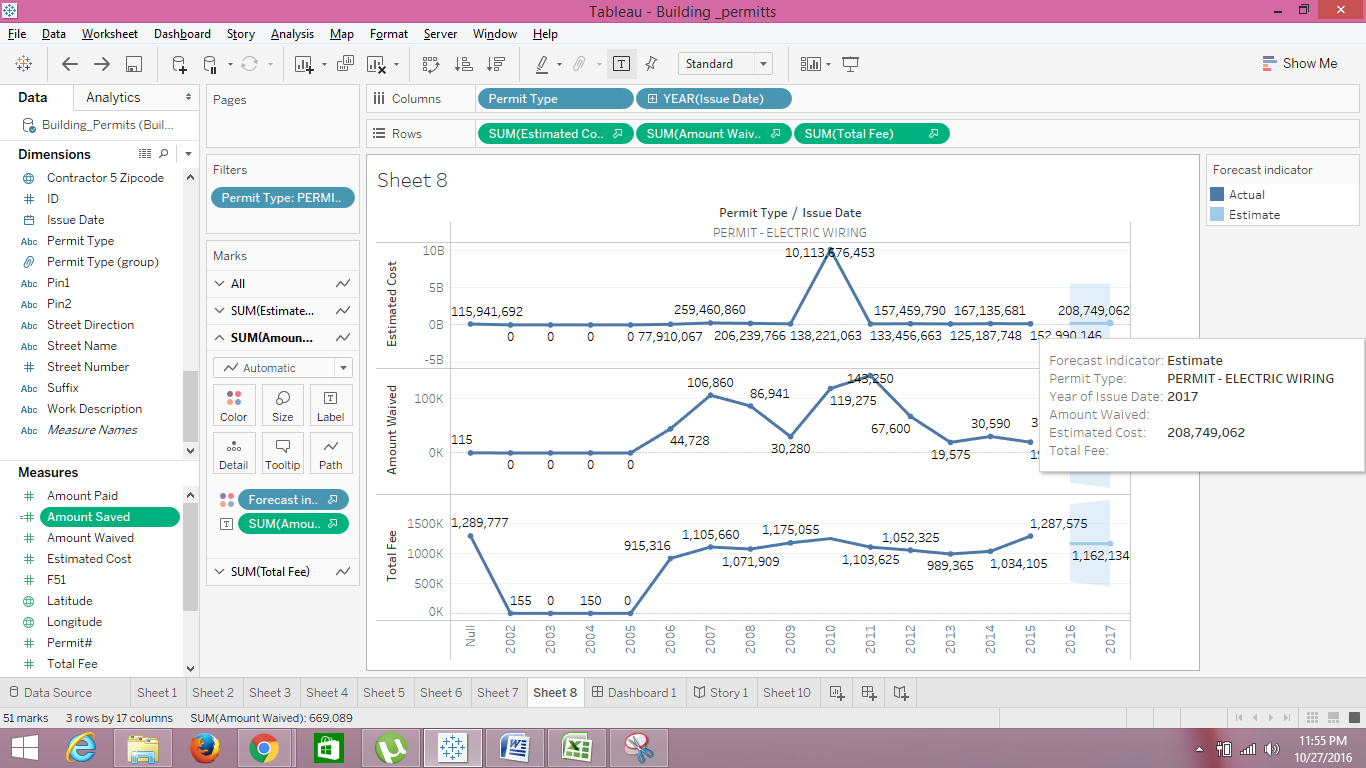
In the above visualization, I can see the geographical area where the issues are permitted. I can see the permit type in filters that helps to select only a particular permit type. I can also see the sum of total fare on the right side that helps to see the sum of total fare for different ranges. In the label, I can see the rank of each state based in the total fee.

1. Compute the sum of estimated cost and total fee for each permit type. Show me the solution using dual axis.



In the above visualization, I can see the sum of estimated cost and total fee for each permit type. The estimated cost and total fee are synchronized showing that amount which is estimated is too high when compared to the actual total fee. The estimated cost is shown in blue color and the total fee is in orange color. The usage of contrast colors helps to differentiate the sum of estimated cost and total fee.

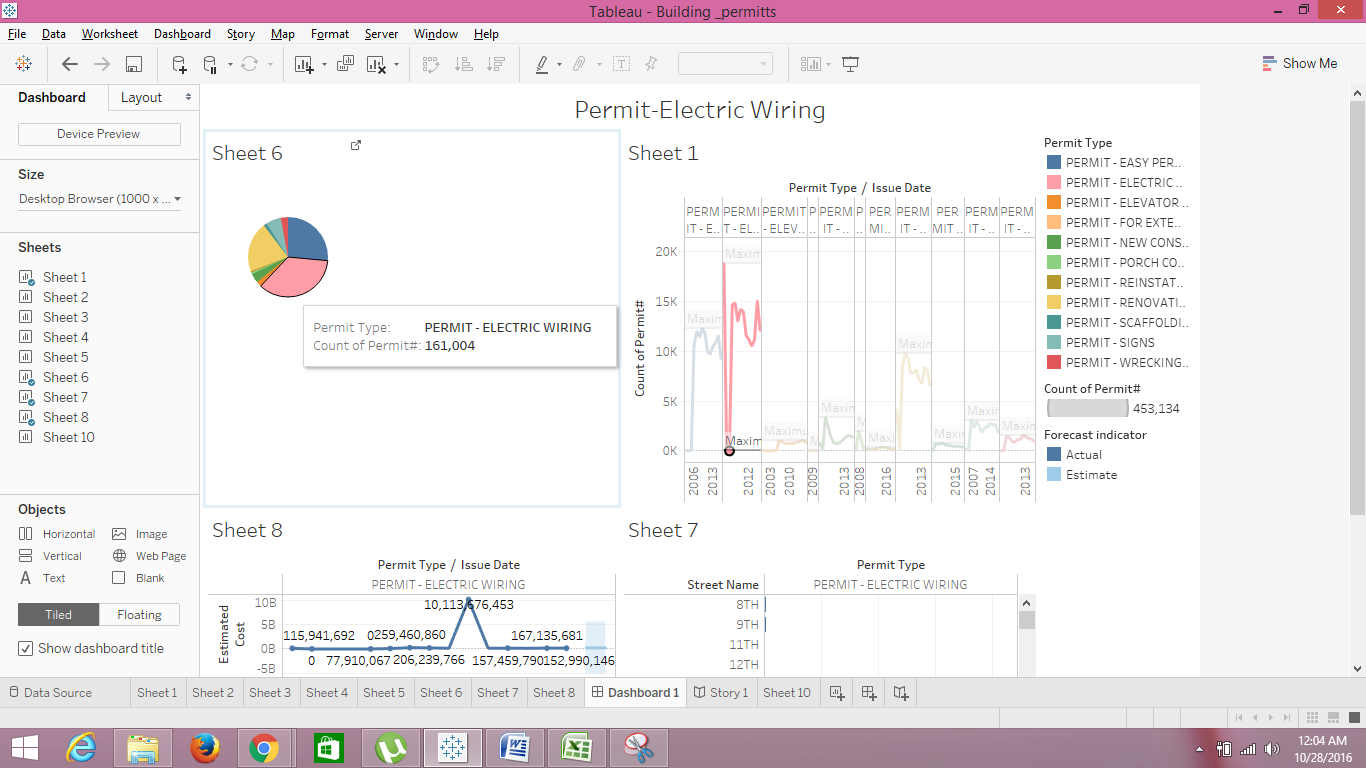
1. Forecast the sum of estimated cost, amount waived and total fee for the permit type “Permit-electric wiring”.

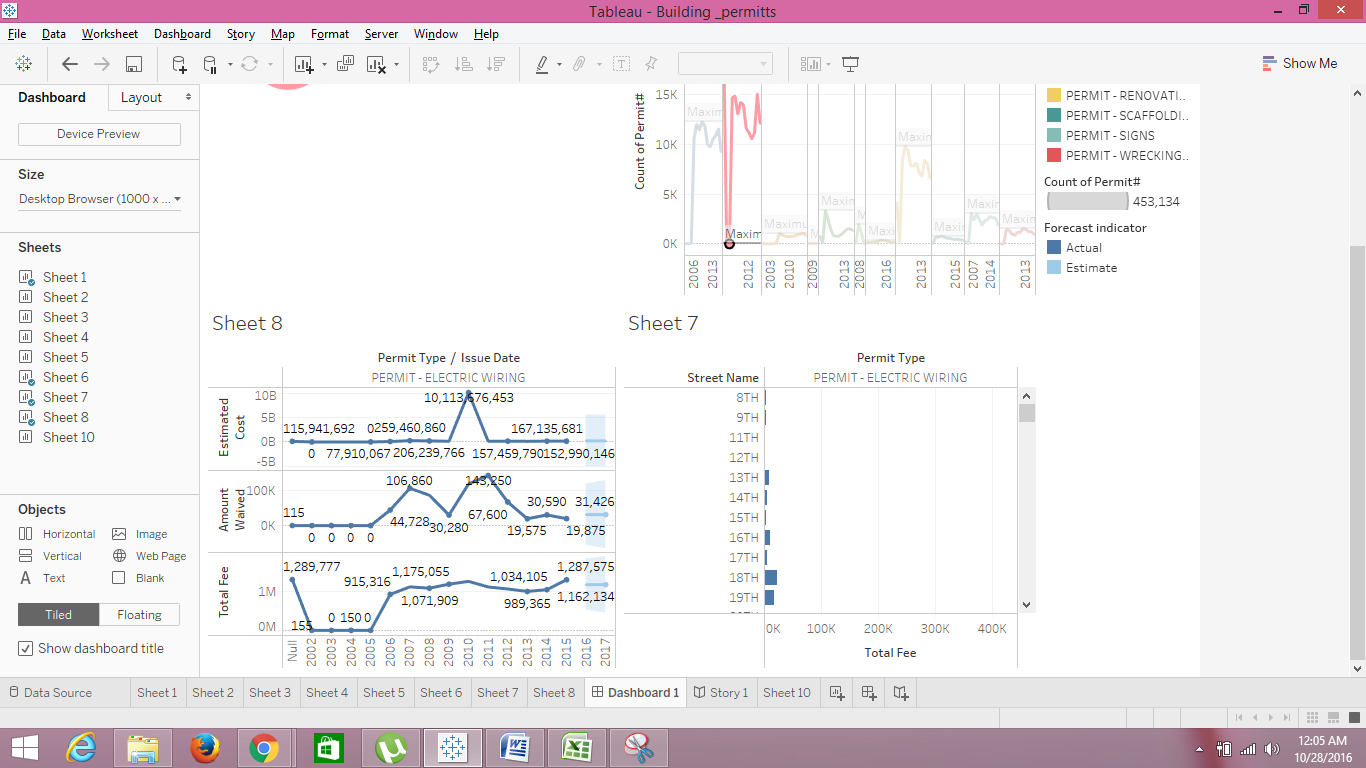


In the above visualization, I can see the future prediction of permit type “Permit-Electric wiring”. The forecast has been shown for the sum of estimated cost and amount waived and total fee. From the above visualization, I can see that the total fee is too low when compared to the estimated cost.

**Dashboard:**

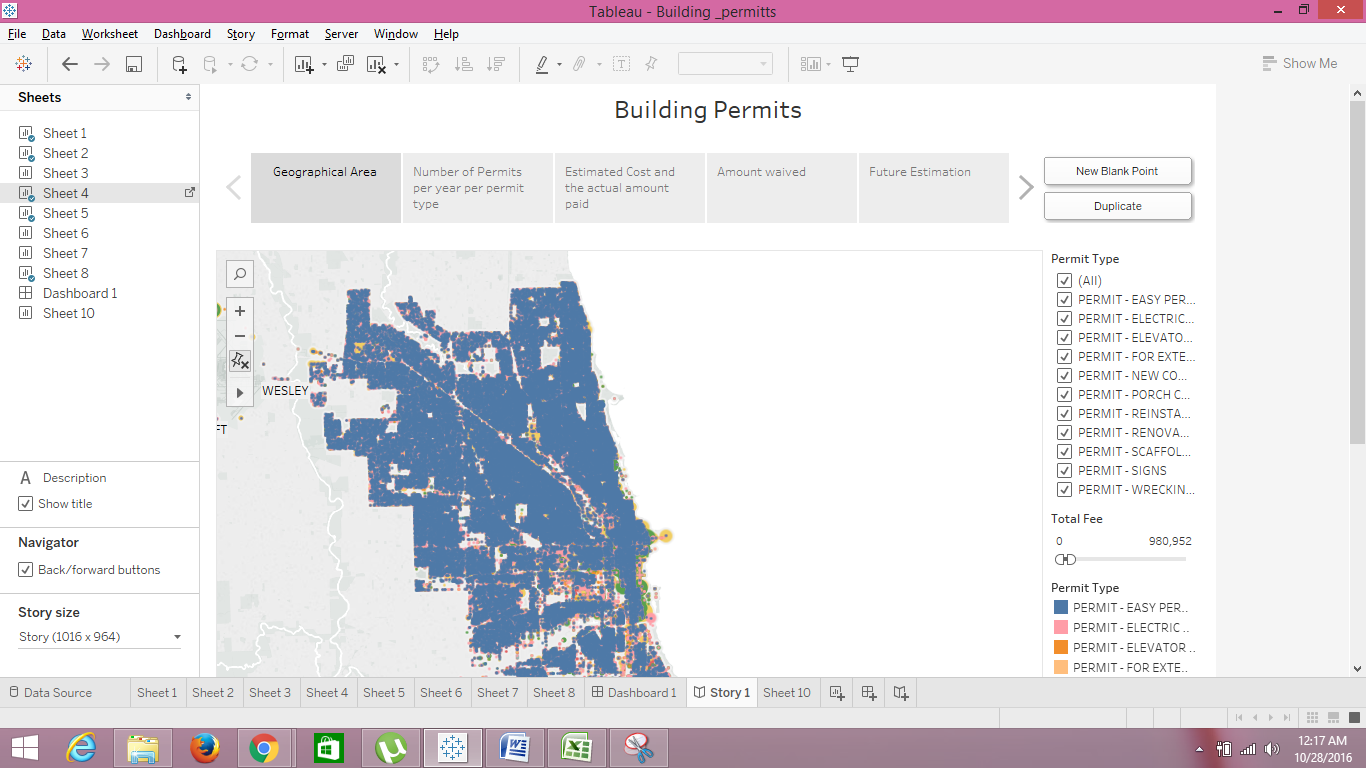
In the dashboard, I have shown the details for the permit type “Permit- Electric Wiring”.



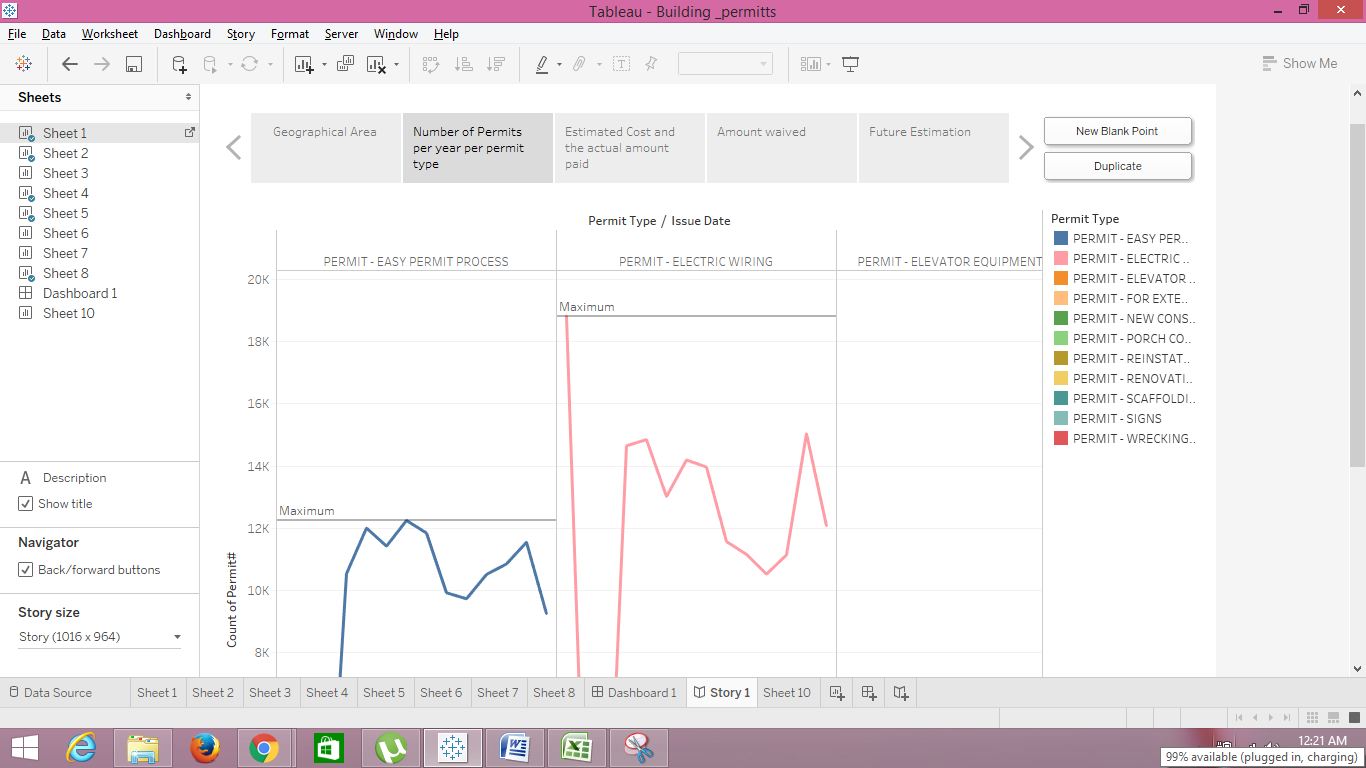


**Story Telling:**

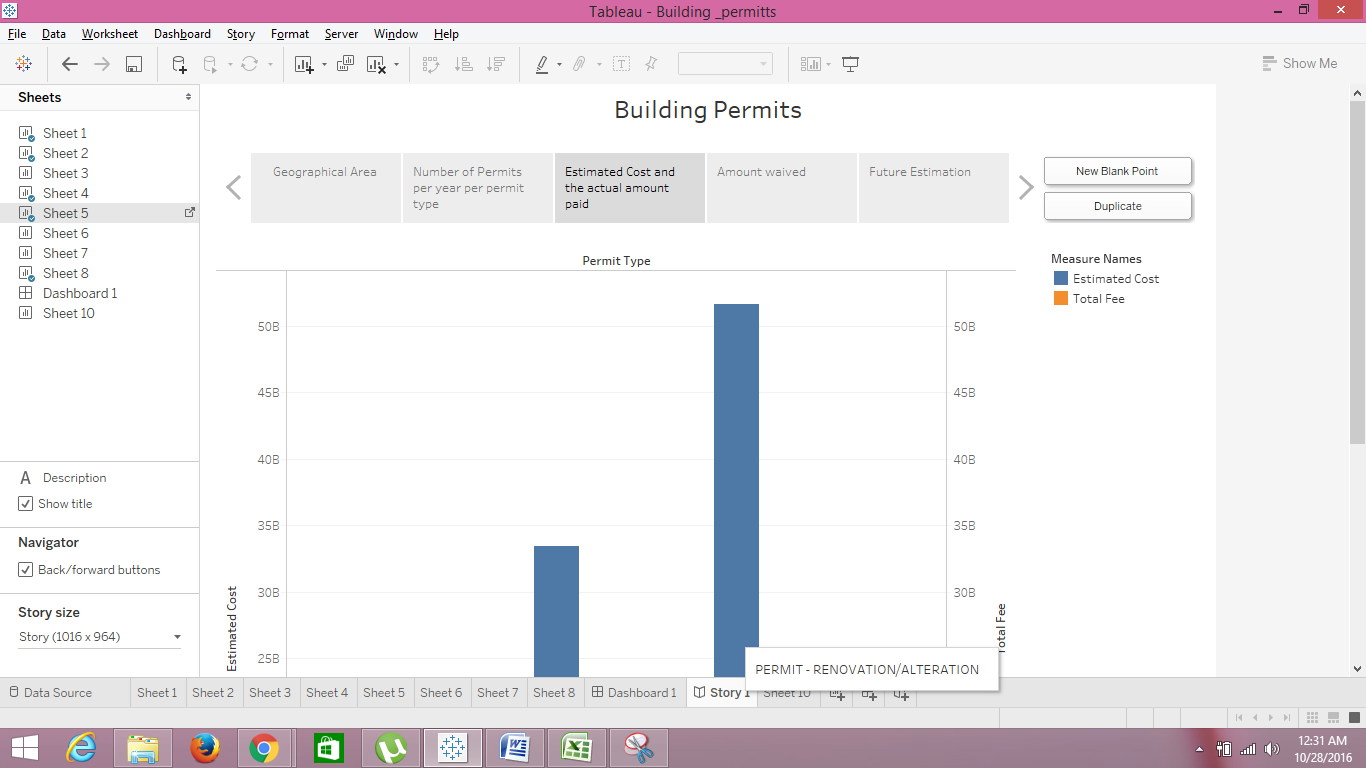
In modern days, apart from busy life the first priority is home. I have seen people spending a lots of money on constructions. So, I have taken a dataset that deals with building permits. I have selected a dataset that concentrates on only one part of the country. I have taken 11 types of permit types that cover construction, electric wiring, elevator, renovation, wrecking etc.



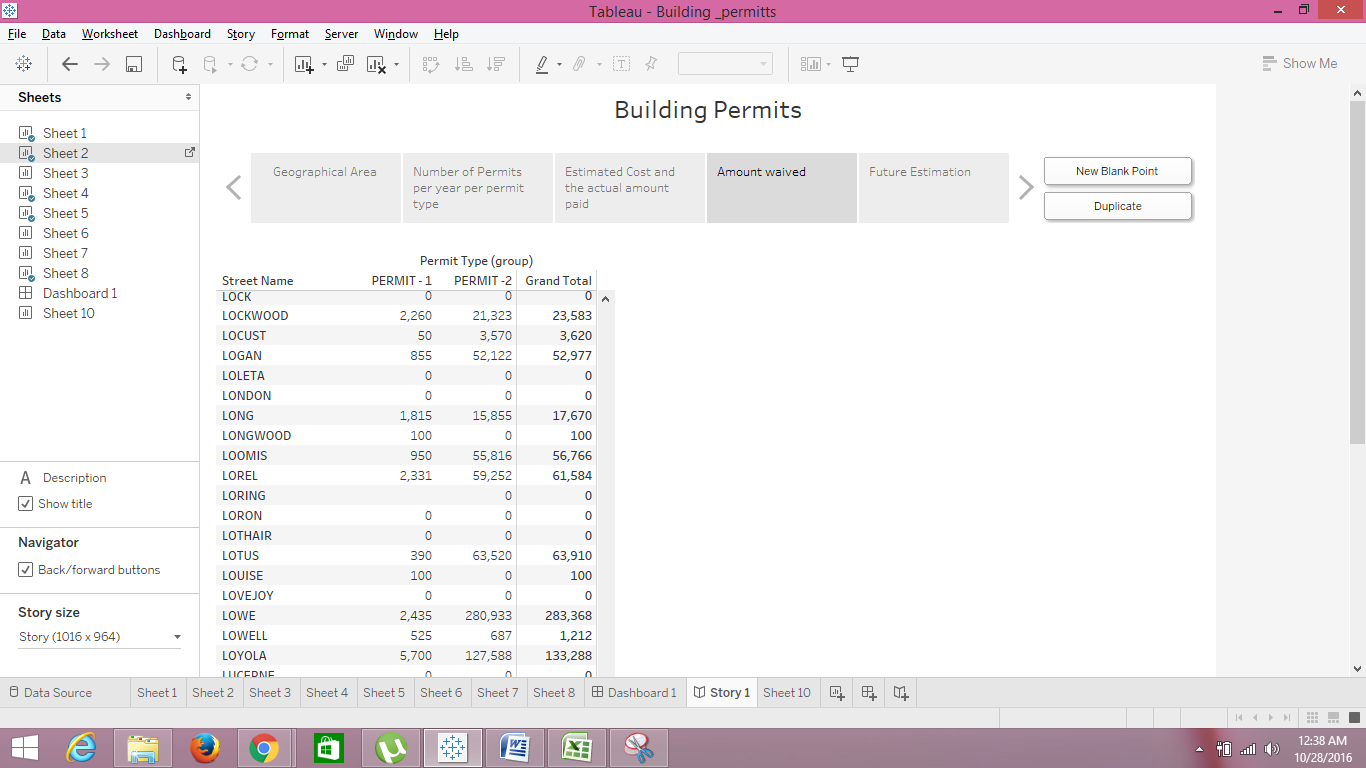
After selecting the geographical area, I have decided to get the data regarding number of permits that are issued over a range of period and also to know whether is an increase or decrease of issuing the permits over that range. After knowing the flow, I have decided to know the year where there is maximum number of permits got issued for each permit type. For few permit types, there is a huge increase of permits and few the issuing of permits rate has been declined. The only permit that has been issued so rarely is “Permit- For extension”. The maximum permits are issued in the year “2008” and the count is “30”.



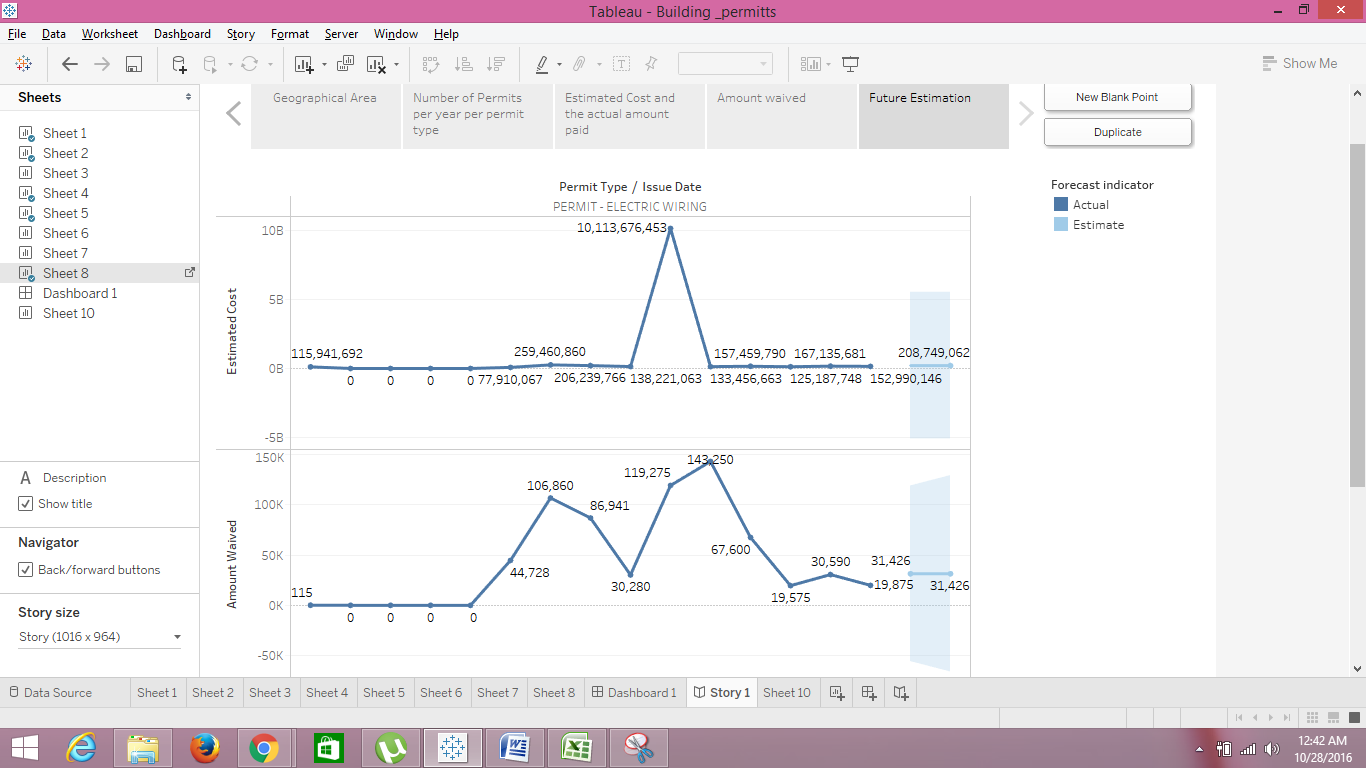
After knowing that there is huge increase in issuing permits, I became so eager to know the amount people spend on the buildings renovations and constructions. So, I have decided to create a bar graph that shows the estimated cost and actual amount they paid. So, I have used the dual axis technique in order to get both the estimated cost and actual amount paid by the people. But after see the bar graph visualization I came to know that, the people are paying so low when compared to estimated cost for almost all the permit types. I also can see that people are spending more money on construction related things than others.



As the total fee is too low when compared to estimated cost, I thought people are getting some amount waived. So, I decided to know the sum of amount waived for all streets. In before visualization, I can see that the amount used is more on construction. So, I grouped the construction related permits into one group and others in other permit group and calculated the grand total of amount waived per each street.



After seeing the amount people are spending on buildings, I thought of forecasting the sum of estimated cost, amount waived and the total fee that people going to spend on buildings. So, I have taken a single permit type “Permit-Electric Wiring” and have forecasted the sum of estimated cost, amount waived and the total fee.



**References:**

<https://www.cityofchicago.org/city/en/depts/bldgs/supp_info/a_guide_to_gettingstartedapermitoverview.html>

<https://www.cityofchicago.org/city/en/depts/bldgs/provdrs/permit_proc/svcs/applications.html>

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