# Power BI and City Data Tutorial

Welcome to this Power BI tutorial. My name is Adam Hecktman Microsoft’s director of technology in civic innovation for Chicago.

This is the first in a series of videos on how to use Microsoft's Power BI with open data. Specifically, we are going to use open data from the City of Chicago's data portal.

In this first section, I am going to take you the various key features of Power BI. We will learn how to use the Power BI Desktop, the power BI web service, you will learn how to quickly access city data through in PBI, use the data to create interactive reports with useful visualizations.

Then we will use the Power BI web service to share those reports with a dashboard, as well as publish them directly to the web.

Now the data we use in this tutorial is available on the city of Chicago’s data portal. But I have already downloaded them for convenience and put them on the thumb drive.

*[open portal]*

We’ll focus 311 data but I'm also to show you later how to bring data from multiple datasets together to create useful reports using the power of queries, and how to create visualizations from those queries.

Now let's just start with the end in mind. By the end of this video you're going to create this dashboard.

*[show dashboard]*

There are reports that make up the components of this dashboard. And you can see that we have some fairly simple visualizations that we will create.

At the top, we have a map showing the requests for rodent baitment and some data associated with it.

Beneath that, we have gauges that show us, using 311 data, how many potholes have been repaired, graffiti requests completed, tree trimming requests addressed, all against a metric of goals.

To the right, we have a slightly more complex visualization – a stacked column chart showing us the food inspection results by year, broken down by result.

Beneath that, we will create a line chart showing us sanitation code service requests, and then we will see how to add media components like this video.

## Setup Power BI

First, let’s make sure you have what you need to get started.

*[go to poweBI.com]*

First, you will want to sign in with a Microsoft account. If you don’t have one, it takes just a minute or so to create one using your existing email address. I have one, so I’ll sign in.

I have the option of downloading the Power BI desktop, so do that, and once you do, you will want to open it.

The reason you will want the desktop as well as the web service is that the desktop allows you to access a very large variety of data sources. I will show you the variety of sources soon.

*[open Power BI Desktop]*

This is the Power BI desktop. The middle is something of a pallet where you will drag your visualizations and attach them to data.

To the right is the visualization pane, where you can specify what you will visualize and how.

And to the right of that is the field pane which will contain the fields from every dataset you bring in.

At the top, we have the familiar ribbon to access our commands.

## Accessing the data:

The first thing you will notice is that it is very obvious where to pull in the data you need.

It first lists the most common data types including everything from Excel data to CSV data, to oData.

But if I click on more you can see that there's a wide variety of places I can pull my data from including…

Let’s jump over to the Chicago City data portal for a moment. The nice thing about the open data stored here is the variety of formats that are available to be used.

*[Search on 311]*

*[Show 311 datasets]*

*[Open Sanitation Code Complaints]*

I can click on Export and see….

Notice over here I can export using the oData format. OData is a way to use a URL to subscribe to data, meaning that it is always going to be fresh when you refresh the data in PowerBI. You don’t have to keep downloading data, it will pull in the changes. Let’s subscribe to this data.

*[If you are time constrained, you can also use the downloaded copy of the data]*

I simply copy this URL.

*[open Power BI Desktop]*

And back in Power BI, I tell it I want to get my data from an oData feed and paste that URL.

Now depending on how old big that data said is it could take some time for it to load the first time.

*[If you are time constrained, you can also use the downloaded copy of the data]*

This is something you are going to see a number of times in this tutorial. Power BI will give you the option to i either load this data as it is or I can edit it before I actually load it.

This is useful if, for example, you do not want the entire data set, and there are some columns that can be removed. Or when you want to make sure that certain columns are brought in using a specific data type, such as a whole number or text.

*[Remove Location column]*

*[load]*

Now I can see that the data's been loaded into power BI and if I go over here on the Right under Fields, I can see it's called query 1. In fact, I can expand it and see all the exact same fields that I saw in the city data portal.

All open sanitation code complaints made to 311 and all requests completed since January 1, 2011 are contained in this data set. The Department of Streets and Sanitation investigates and remedies reported violations of Chicago’s sanitation code, and that data is all captured here.

## Transform and visualize

But now it's in my data model. I am going to be referring to what I create with the data here as a data model. A data model is like a spreadsheet in that it holds data in rows and columns, but being in a data model means that it has more of the attributes and usefulness of a database.

*[rename Query 1 to something useful]*

The first thing I like to do when I bring in data is give it a real name. Right now this is just called query one but we're going to rename this “sanitation code requests”.

Before we even bring in any more data, let’s play around a bit with this data set. For example, let’s just create a very quick and easy visualization.

I can go over to my visualization pane, and click on a line chart and that puts a placeholder on my report. From here, I can just start dragging on fields to chart.

For example, I want to track the number of requests made over time by years, so I can drag “service requests made” as my value, which shows the count of service requests.

And then I can click on Completion Data, and Power BI figures out that I probably want that as my x-axis. And that took just a few seconds.

Let’s add a second visualization: this time a pie chart that shows by zip code how many service requests:

*Value: count of service requests*

*Legend: zip codes*

Now notice that without defining any relationships or creating any queries, Power BI figures that if I click on a specific zip code on my pie chart, I probably want the focus on my line chart to be that Zip code too! I can hover over the data points and get more info.

But for now, we are going to remove these two visualizations to make room for new ones that use a variety of data sets.

*[Delete them]*

Now if you remember the finished product, we had 311 data from many sources. Let’s pull in some of those sources. You saw already how to subscribe to the data using oData, and that is the best way to make sure what you have is current.

But you might have some files on your computer or on the network you want to access. And in fact, I have downloaded some data from the portal here to my machine in a few different formats.

You can see that these are pretty big files and the first time I pull them in, it is going to take some time. Let’s start with the Excel file that has the graffiti data in it.

*[Get Data]*

*[CSV Data]*

*[311 Graffiti Removal Requests the CSV version]*

Note that just like before I have the option of loading it or editing it first. Let’s see what happens when I edit it first.

*[Edit]*

This brings up the query editor, which allows you to do some fairly sophisticated things to the data before you bring it in. And as you edit the query, query editor is saving your steps so that you can repeat them later.

We are going to do something simple. We don’t need this “location” column, so we will get rid of it to make the data set a little smaller.

Note all the other things I can do here. For example, I can make sure that the data's in the right data type that I want, which is a common task.

But for now, let’s just load this data into the data model

*[Show field pane]*

On the right-hand side on the field pain I can see I have two queries: sanitation code requests and graffiti removal

Again, I like to queries real names here so we're going to call this: graffiti removal

Note that you can see which fields are numbers (Sigma) and which are text (like service request) - service request is an ID, not a number, so you want to treat it as text, and it does.

So at this point we have two datasets one that came from a CSV file and one we subscribed to with oData. Let’s now bring in an Excel file.

*[Get Data]*

*[Excel]*

*[311 Rodent Baitment Requests]*

If I go into my folder here I can see that I've downloaded the Rodent Baitment dataset from the data portal and I can bring that data into Power BI

I get a preview just like I did before and again I can load or edit it. I am going to load it directly.

Now I'm going to go away for a minute and bring in a number of other datasets, but won’t make you watch me do them all.

*[bring in the rest of the 311 spreadsheets]*

So I repeated the same steps to bring in data from other data sources. Now I have

*[read data sets]*

Let’s start to create some visualizations. Probably the simplest visualization is the gauge. It uses circular arc to display a single value that measures progress toward a goal. So let’s use it to measure pothole requests.

Let’s go into our pothole request data set. Each service request has a unique service request number. That is really useful because we can count the number of unique service requests numbers to see how many requests were made, and plot them on the arc.

*[value = service request]*

By itself, it doesn’t mean anything, so click on the format paintbrush to tailor how the gauge should look.

For example, I can say the maximum number that I expect in any given year is going to be

Max: 800,000 service requests.

And that the target that the department has set for itself is 0.5 million

Now I can see that the hash mark that tells me where my goal is and what my maximum is and where it stands today

*[Title]*

It's a good practice to give it a real title so instead of count of service request number

we can say “potholes repaired”

Now, this is for the entire dataset, but we can use the filter to make this gauge reflect just the current year or a given year.

*[start making it smaller and move it to top middle]*

So again, Gauges are a great choice when you

* show progress toward a goal.
* show the health of a single measure.
* Or display information that can be quickly scanned and understood.

Let’s create gauges so that we can quickly understand two other data sets.

The first is graffiti removal. We will click on a blank spot on our pallet. Click on gauge.

Make the value the number of service requests.

Note that I can choose the count of service requests and distinct service requests. We will want distinct requests so as not to duplicate.

*Max: 1.5m*

*Target: 1.0M*

*Title: Graffiti Removal Request*

And last, we will create a gauge for tree trimming.

Value: service requests – distinct count

Max: 500,000

Target: 400,000

Line them up in a column.

Note that I have tools that can help me with alignment. I am going to align them all to the right and then distribute them evenly.

*[show alignment tools from ribbon]*

Let’s create another type of visual. A line chart that shows you the sanitation code requests completed over multiple years.

*[Click on blank space towards the bottom of our pallet and then click line chart.]*

I’ll open my sanitation code request data set.

Again, I have a unique service request id. [Value]

Line charts are a good way to emphasize the magnitude of changes over time.

Now, my x axis will be completion date. And note that Power BI has made a good guess as to what I want to see here.

But instead of having an aggregated completion date, I may want to show a

much more granular level of detail.

*[stretch it out]*

Now, it starting to look a little bit bland with all of this blue, so I'm going to change some of the colors of the data label from blue to a shade of red consistent with the theme of this color pallet. Notice that my palette has preselected colors based on that theme.

And, again, let’s give this visualization a useful title

*[Sanitation Code Requests]*

I can go to the options for my X axis and display the title so that it is clear that this is the completion date field

*[X axis]*

And I can do the same with my Y axis

*[Y axis]*

That is good for now.

Now, something that may or may not be correlated to sanitation code requests are food inspections.

This is a good data set to subscribe to using oData because it is updated so frequently.

For this, we are going to use a stacked column chart. A stacked column chart is a good one when you want to show comparisons between each element in the categories, and comparing elements across categories. For example, here we will want to show how many food inspections were made each year, and then within that year, what was the breakdown in terms of inspection results.

In other words:

*[Add the stacked column chart]*

We will have the count of the unique inspection IDs as the value that represents the number of food inspections that were done in a year

*[add ID as value]*

And Inspection Data as our axis to show the number of inspections per year

*[add data as axis]*

And within each column we want different colors we want to represent the results of those inspections

So I will take result and drag it to the legend

*[result = legend]*

And now we can very quickly see the difference between inspections resulted in a pass vs. fail vs. pass with conditions, etc.

But very rarely does green mean fail, let’s play with the colors.

I’ll click on the format icon and click on Data Colors.

Title: Give a meaningful name: Food Inspections

Let’s move the legend position from the top to the left.

And we will place it next to our gauges.

## Mapping in Power BI:

There are several ways to create maps. There are good add-ins for a variety of maps available from the Power BI Visualization Gallery. There is also an ESRI visualization for using ArcGIS maps. We will create a basic map to map out where the Rodent Baitment is taking place.

Let's go back to the finished product dashboard so I can show you exactly what it is that we're going to be creating.

this is a map that is completely interactive. You can zoom in, you can hover over the data points to see the data underneath it, and you can even create a visual within a visual, note that each circle’s size represents the number of requests, and within the circle you can see how it breaks down by status. Let’s create this.

*[Go back to Power BI Desktop]*

We will click on the world map and drag it to the open space.

What are we tracking? We are tracking the number of service requests, just like we did before.

But since this is a map we want to tell it how to represent location. And notice that we have a lot of different fields that could represent location. We have Zip codes, address, longitude and latitude. You need to think through the level of granularity that is appropriate for your purpose. Sometimes you want to map at the street level, sometimes the neighborhood level, or very specifically at lat/long, which we will use here.

*[Drag lat to lat, and long to long].*

Now, the size of the circle represents the number of requests for that particular geography so I can see that I had of ball only one request here but I have a 14 requests over here

To make this even more meaningful we can do is we can take the status field and drag it over here to the legend. Now I can see relative to all of the requests for a place, what the status is.

And this is a fully interactive map.

Let’s give it a meaningful title: Rodent Batement

And put the legend on the left. And size it.

## Publishing and Sharing

That’s it! We’ve created a report. We can use them here or we can publish them to the organization.

*[file/Publish]*

Publishing this report is going to make it available to me and anybody else in my org that I want to share it with.

But often, what you will want to do is to either take reports from many sources and put them on a customized dashboard. And a second thing you may want to do is simply take these reports and publish them to the web for sharing on a blog, in an email, etc.

The sharing and publishing options live in the Power BI Web Service, so let’s open that.

You are in the power BI web service, which again, you got from going to powerBI.com and signing in. Let’s look at it a bit. It looks a lot like the desktop.

*[open Power BI Web Service]*

On the left, I can see the list of Datasets I have used. And reports that I have created and published.

This is called my Workspace.

It is also where I can take the reports and visuals and pin them to dashboards. Dashboards are a way to share reports throughout my organization, so that is what we are going to create, a dashboard.

To create a dashboard, you go through your reports, pick the visualizations you want to show, and pin them, by click on this pin icon. Let’s pin all of these visuals, and then add a few things of our own to create a nice looking dashboard.

*[pin all the visuals to my dashboard]*

Now, I can also add non-PowerBI content into my dashboard. For example, I know that there is a video where the DoIT CIO is talking about this data, so I can add that as a Widget. Widgets can include video. And it can also include images, and other web content.

*[Add the video and add a picture]*

And once I have created my dashboard, the user can drill into this content. Clicking on a visual brings me to the report that was used to create it!

Now that I am in this report, I can show you another way to share this, and that is publish to web.

*[File/Publish to Web]*

With Publish to Web, I can send this out as a standard URL, or, it also gives me an embed code so that I can embed this, with all the interactivity.

So congratulations, you have created some basic data visualizations, a report, and a dashboard, and you have learned how to publish and share.

In the next videos, we will go deeper into using queries against the city’s open data.