

London Bike Hire Lab

Boris Bikes were introduced in London in 2010 and for some people are the only way to travel around London. This dataset has the number of bikes hired every day from 30th July 2010 to this year. It comes from the [London DataStore](#), a free open public data portal run by the GLA and the Mayor of London

Dataset

London Daily Bike Hire Data.xlsx – a tidied up version of the dataset on the London DataStore with just the data we need for the exercise. It has two tabs

- Bike Hire – the number of bikes hired each day
- Average Hire Time in minutes by month
- Calendar – a date table, useful for aggregating by year, quarter, month or day of week. It contains a row for every date in 2010 through 2018

Objectives

Import the data. Calculate some basic stats. Visualise the data in various ways. What insight can we draw?

Instructions

Launch Power BI Desktop

Set up to use our own calendar

When Power BI see a date in an imported dataset, it will helpfully create a year-quarter-month hierarchy for it. This is useful in some cases but in this exercise, we also want to analyse by day of week (Monday – Sunday) so we are going to import our own calendar tables including a Day of Week column. So, we'll disable the automatic date hierarchy in Power BI

Go to File | Options

In Current File | Data Load, uncheck the Auto Date/Time box

Options

The screenshot shows the 'Options' dialog box in Power BI Desktop. The 'CURRENT FILE' tab is selected, and the 'Data Load' option is highlighted. The settings for 'Data Load' are as follows:

- GLOBAL**
 - Data Load
 - Power Query Editor
 - DirectQuery
 - R scripting
 - Security
 - Privacy
 - Updates
 - Usage Data
 - Diagnostics
 - Preview features
 - Auto recovery
- CURRENT FILE**
 - Data Load
 - Regional Settings

The 'Data Load' settings are:

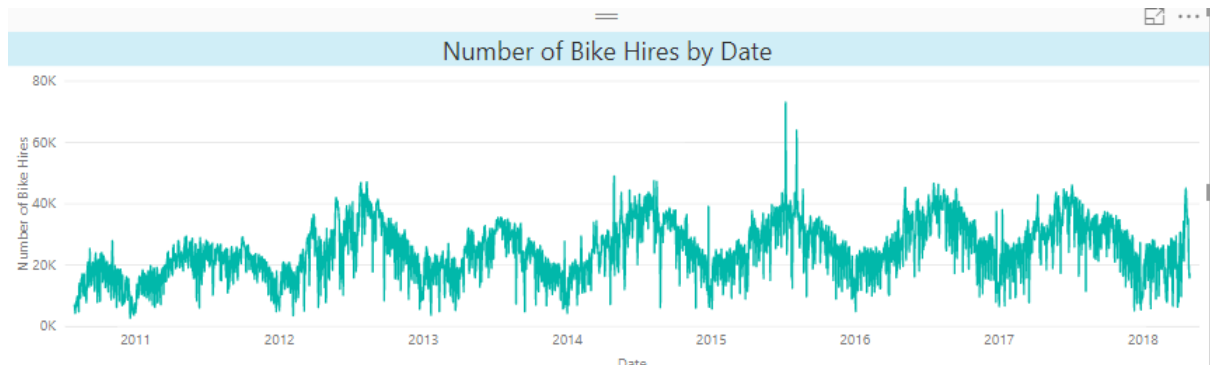
- Type Detection**
 - ☒ Automatically detect column types and headers for unstructured sources
- Relationships**
 - ☒ Import relationships from data sources ⓘ
 - ☐ Update relationships when refreshing queries ⓘ
 - ☒ Autodetect new relationships after data is loaded ⓘ
- Time intelligence**
 - ☐ Auto Date/Time ⓘ
- Background Data**
 - ☒ Allow data preview to download in the background
- Parallel loading of tables**
 - ☒ Enable parallel loading of tables ⓘ
- Natural language**
 - ☒ Turn on natural language questions with Q&A ⓘ

Import the data

Choose Get Data | Excel then import the data from the 'Bike Hire' tab of the Excel file.

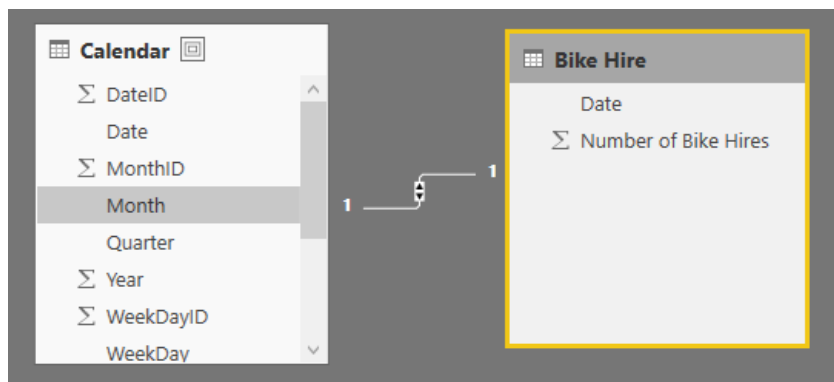
First look – build a time-series chart

Create a page named First Look. Add a line chart. Place the Date field in the Axis well and the 'Number of Bike Hires' in the Values well. Your visual make look like this.



import the data from the 'Calendar' tab of the Excel file

In the relationships window, join the two tables on the date column. (Notice that Power BI creates a 1-1 bidirectional join – we will discuss this later in the course)



Build a column chart with a year-quarter-month-date hierarchy and drill capabilities

Create a column chart. Place four fields in the Axis well; Year, Quarter, Month and Date. Place the 'Number of Bike Hires' field in the Values well. Since you have placed more than one field on the axis the drill buttons appear. Explore the drill capabilities

- Expand all level in the hierarchy to get more detail
- Go to next level in the hierarchy to see a seasonal analysis
- Drill down into a particular year -> quarter -> month

Note that the month level, the months are ordered alphabetically rather than chronologically. Fix this by:

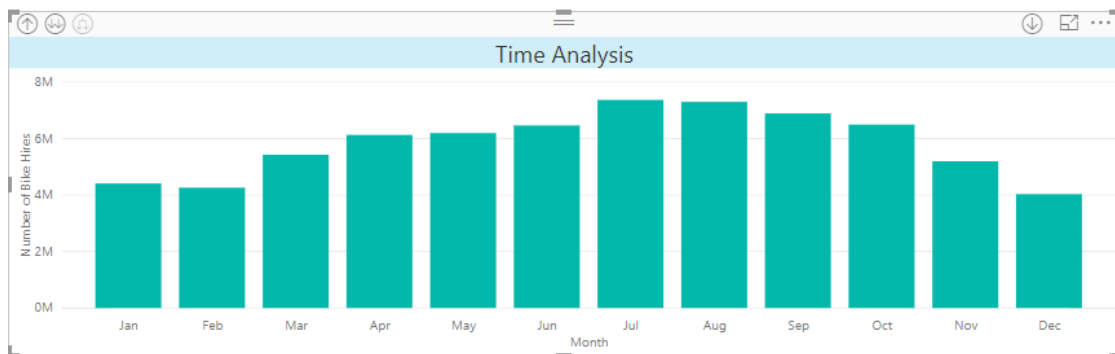
1. Go to the data pane
2. Select the DateTable
3. Select the Month Column

4. Select the Modelling tab and click on 'Sort by Column'
5. Sort Month by the Month ID column

While we are tidying up the data model, let's do the following

- Sort WeekDay by the WeekDayID column
- Sort Year-Month by the Year-MonthID column
- Hide in Report View these columns:
 - In DateTable – DateID, MonthID, WeekDayID, Year-MonthID, DateID
 - In 'Bike Hire' – Date

Your visual make look like this at the month level



Improve the model

Create a calculation which is simply the count of the number of observations (dates). On the modelling tab, click on New Measure and enter the line below into the formula bar.

Number of Observations = COUNTROWS('Bike Hire')

Visualise with a box-whisker chart

Create a new page named Box & Whisker.

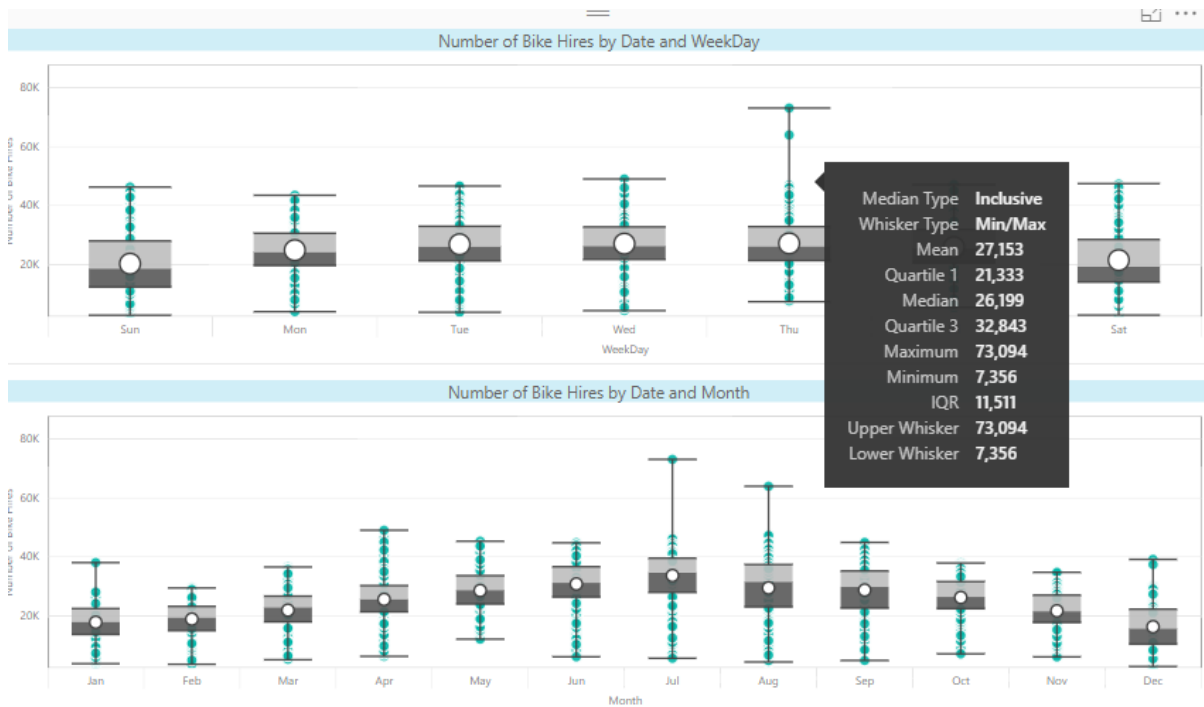
We will make two versions of the box & whisker, one with day of week on the axis, another with month on the axis.

The box and whisker chart is a custom visual – it does not come as part of the standard palette. To use this, click the 'From File' button of the 'Custom Visuals' section on the Home tab of the ribbon and import from the 'Custom Visuals' folder.

Place the Date field in the Axis well, the WeekDay field in the 'Axis Category I' well and the 'Number of Bike Hires' in the value well.

Build a similar chart but for month rather than weekday.

Your final page may look like this



Visualise the distribution of values with a histogram

Firstly we need to bin the number of bike hires – this will become the axis of our bars. To do this, right-click on the 'Daily Bike Hire' and choose 'New Group'. You can take the defaults or choose a round number like 2,000 for the bin size

Groups

Name	Number of Bike Hires (bins)	Field	Number of Bike Hires
Group type	Bin	Min value	2,764
Bin Type	Size of bins	Max value	73,094

Binning splits numeric or date/time data into equally sized groups. The default bin size is calculated based on your data.

Bin size:

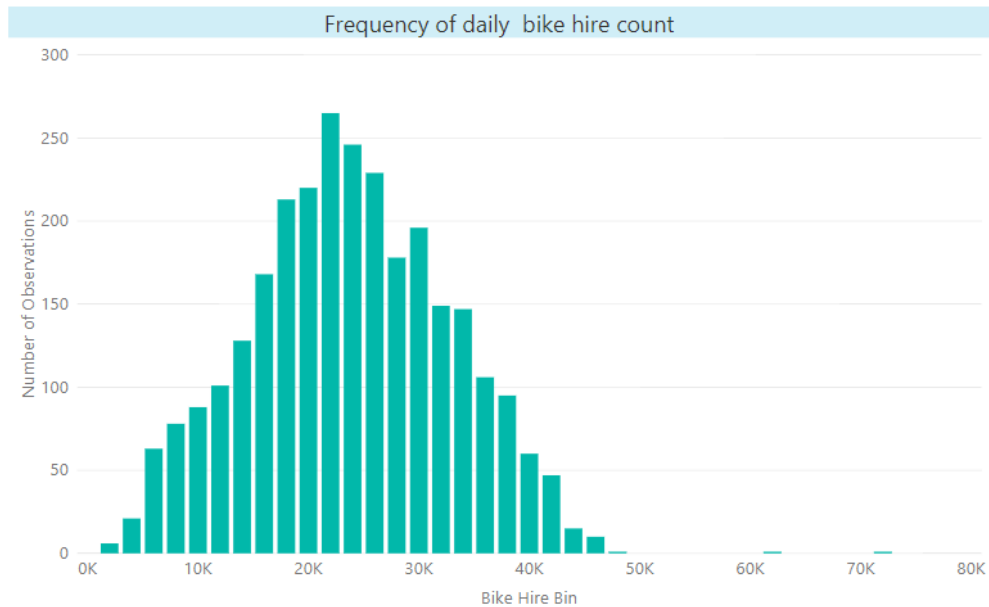
[Reset to default](#)

Rename the field to 'Bike Hire Bin'

Add a column chart.

- Place the 'Bike Hire Bin' field in the Axis well
- Place the 'Number of Observations' in the Value wells.

The final chart may look like this.



Looking at the histogram, do you think that the standard deviation we calculated is a useful measure?

Finally...

What insights can you draw from our results from the results and the visuals? What is your view on any outliers?

Optional - Create some basic descriptive stats

Create a new page named 'Summary Stats'

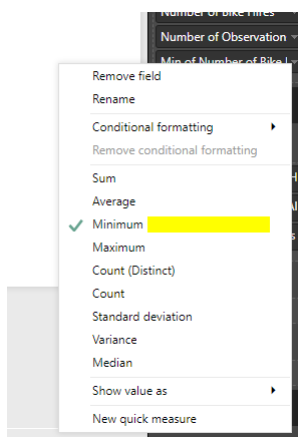
Check the box to the left of the field – Power BI will add a bar chart to the page.

Add a matrix. (Hint: set 'Show on Rows' On in the matrix to see the measures nicely laid out in rows rather than columns as below)

Place the 'Number of Bikes Hired' field in the Values well

Place the 'Number of Observations' field in the Values well

Place the 'Number Of Bikes Hired' field into the Values well and then set the aggregation to min



Let's do this a better way. Add a new measure

Min Hire = MIN('Bike Hire'[Number of Bike Hires])

and place this in the Values well instead

Similarly, create the following measures and add to the matrix

Mean Hire = AVERAGE('Bike Hire'[Number of Bike Hires])

Median Hire = MEDIAN('Bike Hire'[Number of Bike Hires])

Max Hire = MAX('Bike Hire'[Number of Bike Hires])

SD Hire = STDEV.P('Bike Hire'[Number of Bike Hires])

Your matrix may look like this.

Number of Bike Hires	70,185,405
Number of Observations	2,832
Min Hire	2,764
Mean Hire	24,783
Median Hire	24,527
Max Hire	73,094
SD Hire	9,049

Expand the matrix to show these stats by year and day of week. Add data bars or other conditional formatting if you wish.

Basic Stats by Week Day								
	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Total
Min Hire	2,764	3,971	3,763	4,327	7,356	5,402	2,805	2,764
Median Hire	18,624	24,268	26,138	26,293	26,199	25,567	19,349	24,527
Max Hire	46,276	43,497	46,625	49,025	73,094	47,102	47,454	73,094
Mean Hire	20,258	24,852	26,845	27,058	27,153	25,861	21,469	24,783
SD Hire	9,835	7,951	8,064	8,088	8,595	8,182	9,708	9,049