

Giới thiệu PowerBI

Bài 1 – Phân tích mô tả với Power BI

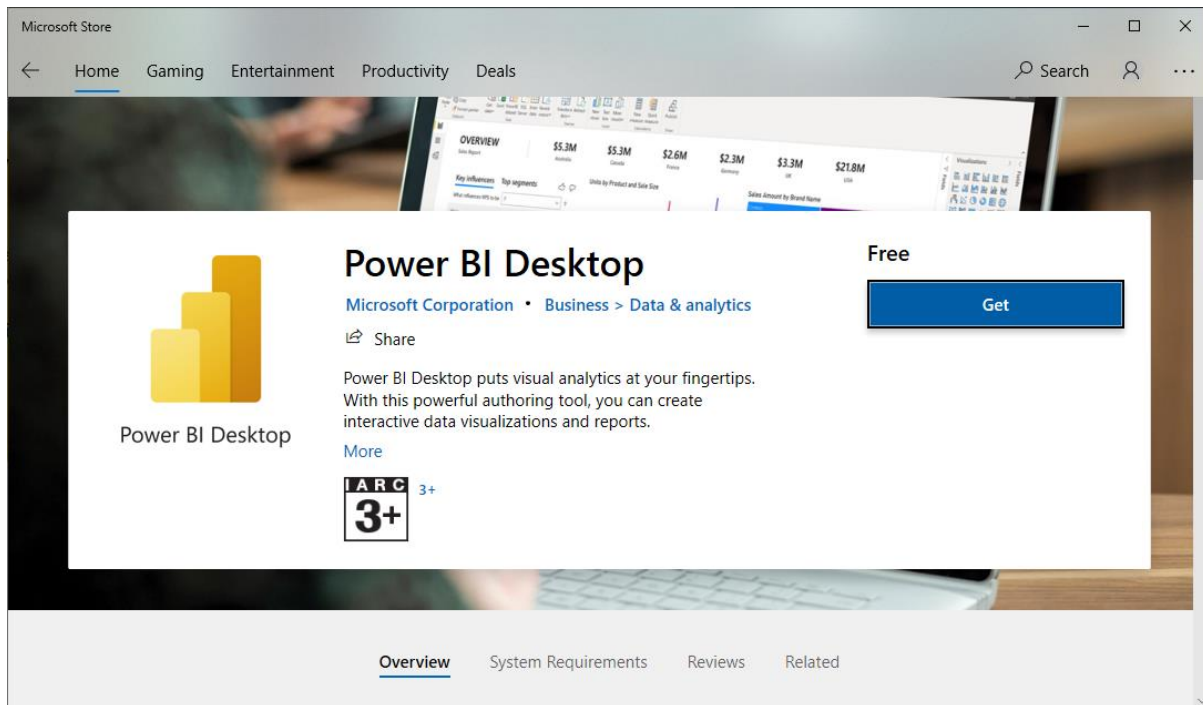
Tải và cài đặt PowerBI phiên bản Desktop tại:

<https://powerbi.microsoft.com/en-us/downloads>

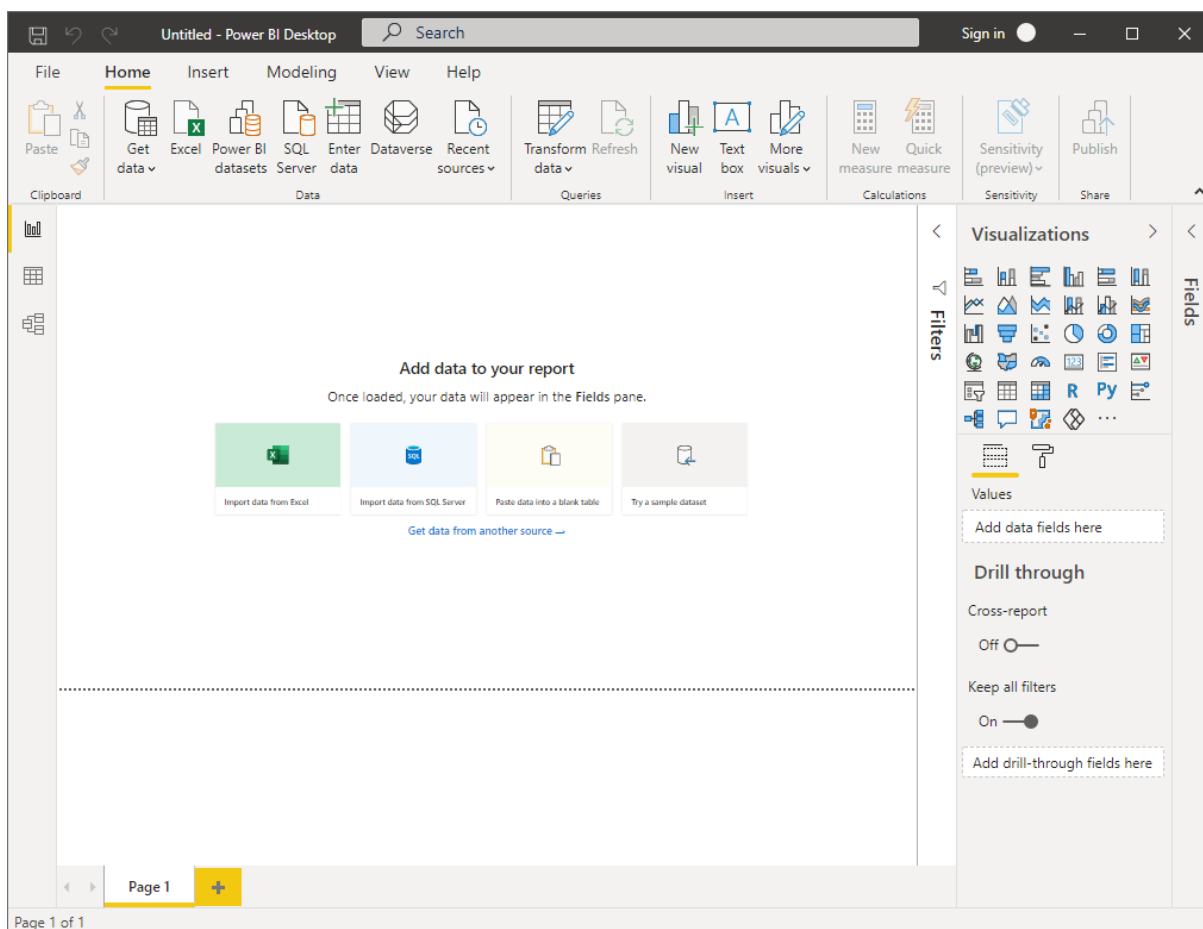
Xem chỗ mũi tên trong hình bên dưới:

The screenshot shows the 'Download Power BI tools and apps' page. It features a yellow header with the title and a description of Power BI's capabilities. Below the header, there are three main sections: Microsoft Power BI Desktop, Microsoft Power BI Mobile, and Microsoft on-premises data gateway. Each section includes a brief description and a 'Download' button. The Desktop section also has an 'Advanced download options' link. The Mobile section shows download links for the Microsoft Store, App Store, and Google Play. The on-premises gateway section has links for 'Download standard mode' and 'Download personal mode'.

Product	Description	Download Link
Microsoft Power BI Desktop	With the Power BI Desktop you can visually explore your data through a free-form drag-and-drop canvas, a broad range of modern data visualizations, and an easy-to-use report authoring experience.	Download > Advanced download options >
Microsoft Power BI Mobile	Access your data anywhere, anytime. These native apps provide live, interactive, mobile access to your important business information.	Get it from Microsoft Download on the App Store GET IT ON Google Play
Microsoft on-premises data gateway	Keep your dashboards and reports up to date by connecting to your on-premises data sources—without the need to move the data.	Download standard mode > Download personal mode >



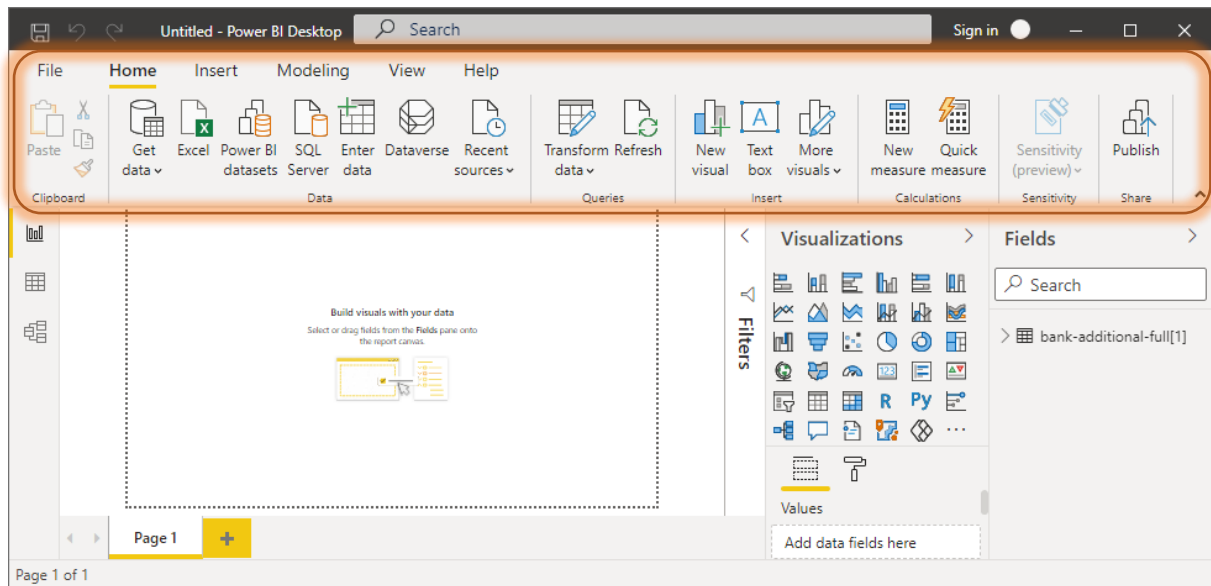
Khởi động PowerBI sau khi cài xong, bỏ qua màn hình giới thiệu ban đầu thì giao diện chính của PowerBI như sau:



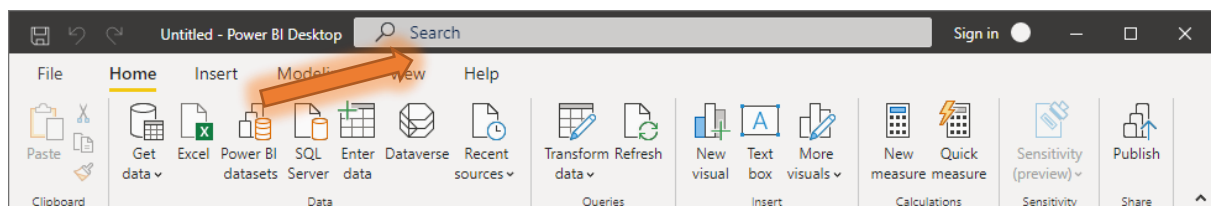
Khái niệm

Ribbon

Menu và thanh công cụ trình bày các chức năng của PowerBI gọi là Ribbon.



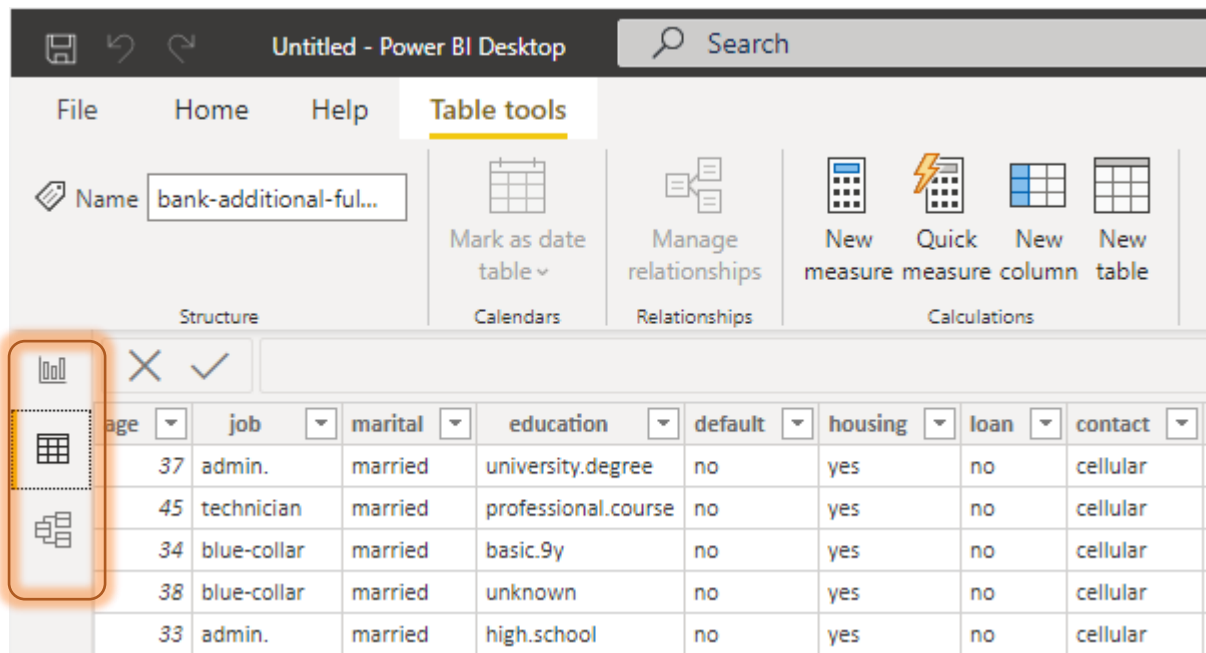
Search bar



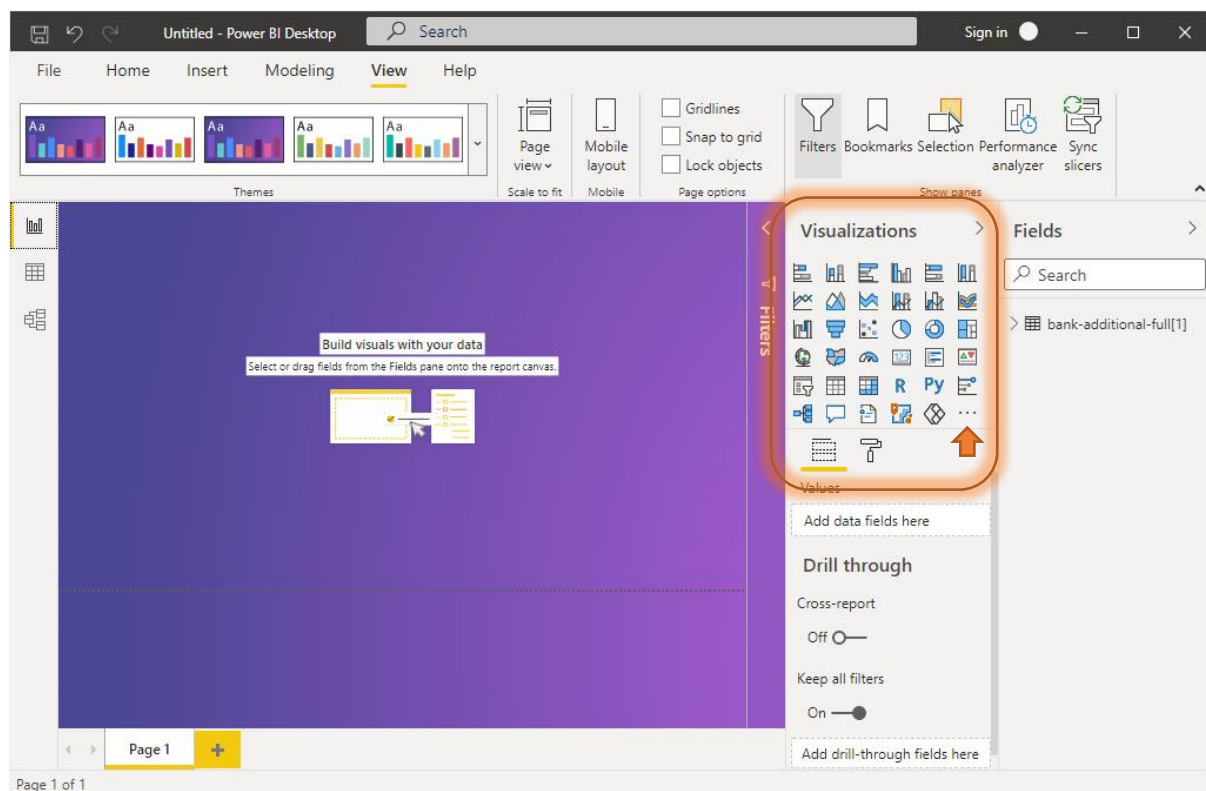
Chế độ xem (view)

Có 3 biểu tượng ở cạnh bên trái giúp bạn xem tương ứng theo 3 dạng:

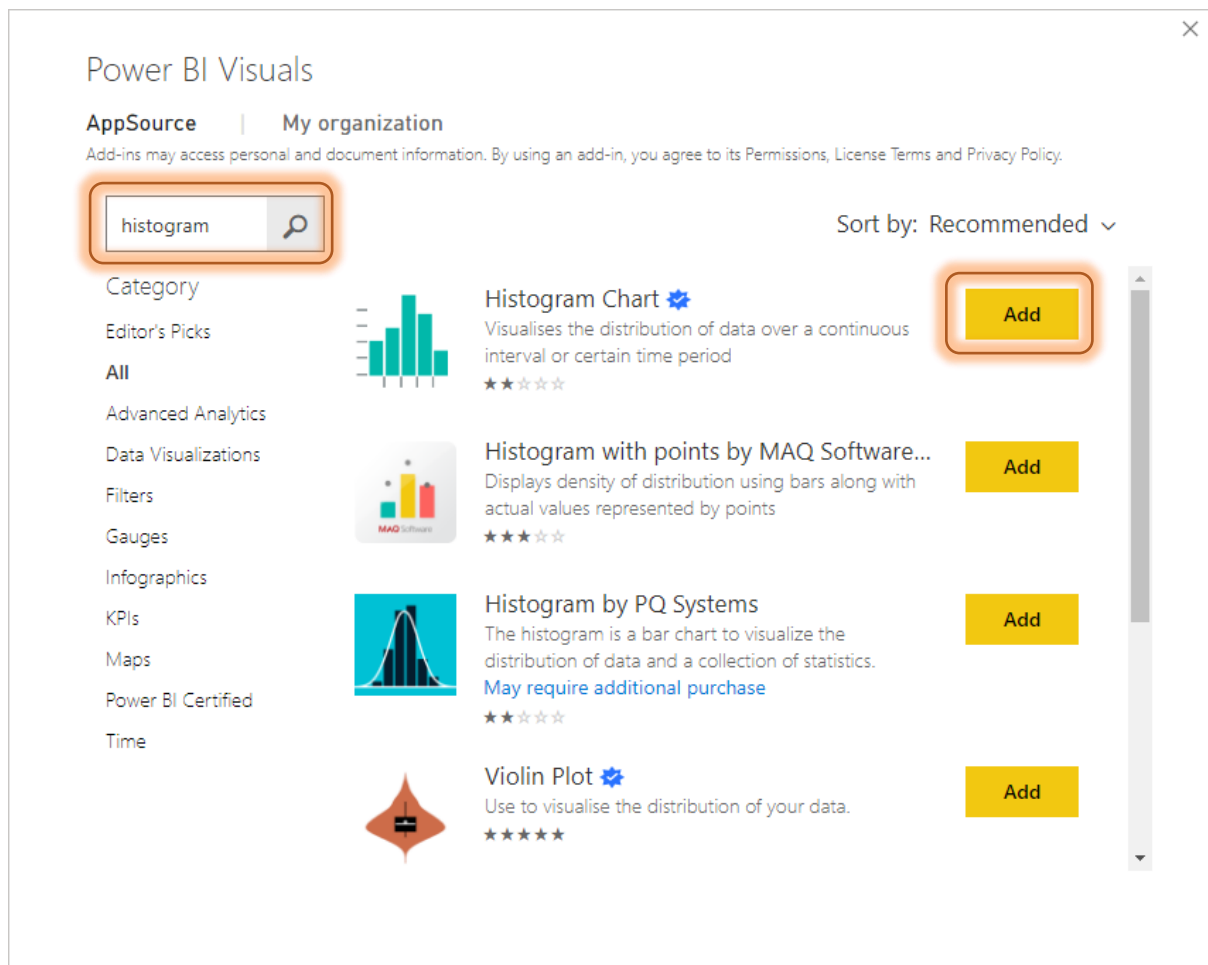
- Report: xem dạng biểu đồ
- Data: Xem dạng dữ liệu dạng bảng như là Excel.
- Model: Xem mô hình của dữ liệu



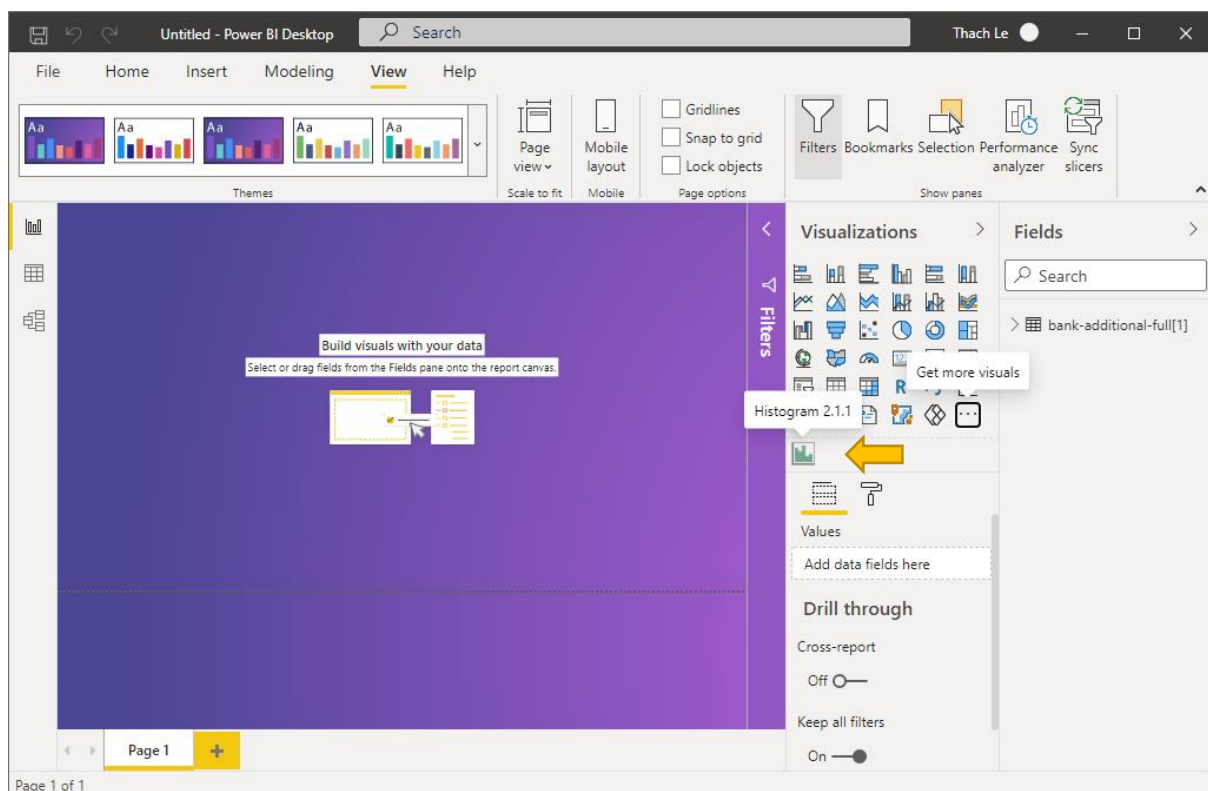
Các loại report



Mục Visualizations chỉ hiển thị vài biểu đồ. Để dùng thêm các biểu đồ khác thì bấm vào biểu tượng ba chấm (chỗ mũi tên).



Thử tìm biểu đồ histogram và bấm nút Add. Kết quả là có thêm biểu tượng Histogram.



Dữ liệu minh họa - Bank Marketing

Để minh họa cho bài này tôi dùng dữ liệu về Bank Marketing của UCI ³². Tải 2 file bank.zip và bank-additional.zip trong mục Data Folder.



Đây là dữ liệu thu thập được từ một dự án tìm kiếm khách hàng (direct marketing campaign) bằng điện thoại để chào một dịch vụ hoặc sản phẩm của ngân hàng cho khách hàng ở Bồ Đào Nha. Dịch vụ hoặc sản phẩm theo thuật ngữ ngân hàng ở đây là **deposit**.

Dữ liệu gồm các biến đầu vào (input variables):

#	Biến	Ý nghĩa (kiểu dữ liệu, giải thích)
Nhóm thông tin về khách hàng (bank client data)		
1	age	tuổi (numeric)
2	job	nghề nghiệp
3	marital	tình trạng hôn nhân. 'divorced' (ly dị hoặc góa phụ), 'married' (đã lập gia đình), 'single' (độc thân), 'unknown' (không biết).
4	education	Trình độ học vấn
5	default	Có quan hệ tín dụng hay không: 'no', 'yes', 'unknown'. Được hiểu là có đang vay tiền ngân hàng không?
6	housing	Có khoản vay nhà ở không? 'no', 'yes', 'unknown'
7	loan	Có khoản vay cá nhân không? 'no', 'yes', 'unknown'
Nhóm thông tin về lần liên lạc gần đây trong chiến dịch		
8	contact	Kênh liên lạc: 'cellular', 'telephone'

³² Link: <https://archive.ics.uci.edu/ml/datasets/bank+marketing>

9	month	Tháng liên lạc gần đây nhất: 'jan', 'feb', 'mar', ..., 'nov', 'dec'
10	day_of_week	Ngày liên lạc gần đây nhất: 'mon', 'tue', 'wed', 'thu', 'fri'
11	duration	Thời lượng cuộc gọi gần đây: tính bằng giây.
12	campain	Số lần liên lạc với khách hàng này. Giá trị số có nghĩa là lần liên lạc gần nhất)
13	pdays	Số ngày tính từ lần liên lạc trong chiến dịch gần đây. Giá trị 999 có nghĩa là không có liên lạc trong chiến dịch trước.
14	previous	Số lần liên lạc trước khi chiến dịch này xảy ra.
15	poutcome	Kết quả của chiến dịch marketing lần trước: 'failure', 'nonexistent', 'success'
Nhóm thông tin về kinh tế và xã hội		
16	emp.var.rate	Tỉ lệ công ăn việc làm - employment variation rate - quarterly indicator (numeric)
17	cons.prive.idx	Chỉ số giá tiêu dùng - consumer price index - monthly indicator (numeric)
18	cons.conf.idx	Chỉ số niềm tin tiêu dùng - consumer confidence index - monthly indicator (numeric)
19	euribor3m	Tỉ lệ chào bán liên ngân hàng Euro - euribor 3 month rate - daily indicator (numeric). Là tỉ lệ tham chiếu được xây dựng từ lãi suất trung bình mà các ngân hàng Châu Âu cung cấp cho vay ngắn hạn không có tài sản bảo đảm trên thị trường liên ngân hàng
20	nr.employed	Số người lao động - number of employees - quarterly indicator (numeric)
Biến kết quả		
21	y	biến nhị phân: 'yes', 'no'. Khách hàng đồng ý sử dụng dịch vụ yes, không đồng ý là no.

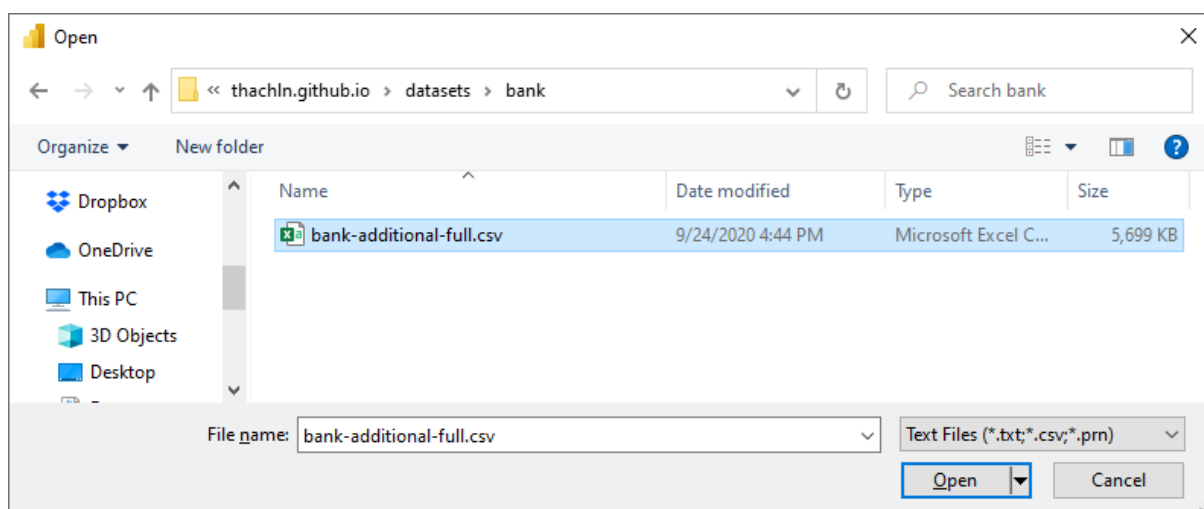
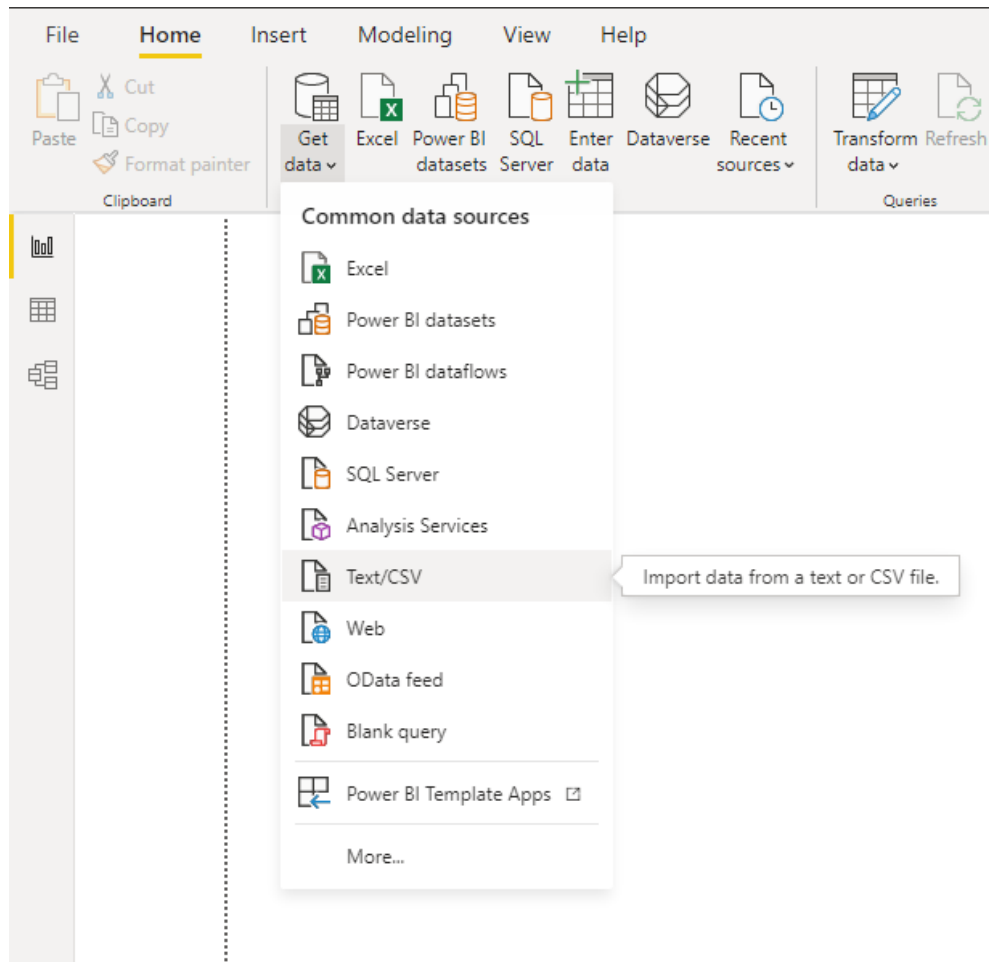
Dữ liệu minh họa có lưu tại:

<https://thachln.github.io/datasets/bank/bank-additional-full.csv>

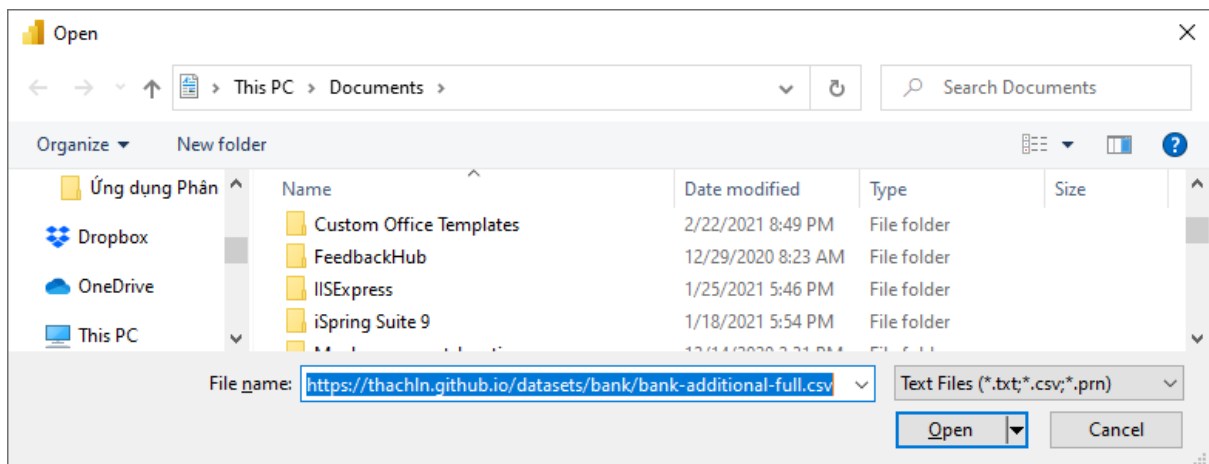
Có thể import vào PowerBI Desktop bằng 2 cách sau:

- 1) Tải file .csv ở trên về máy. Sau đó import từ file đã tải.
- 2) Import trực tiếp từ link trên Internet.

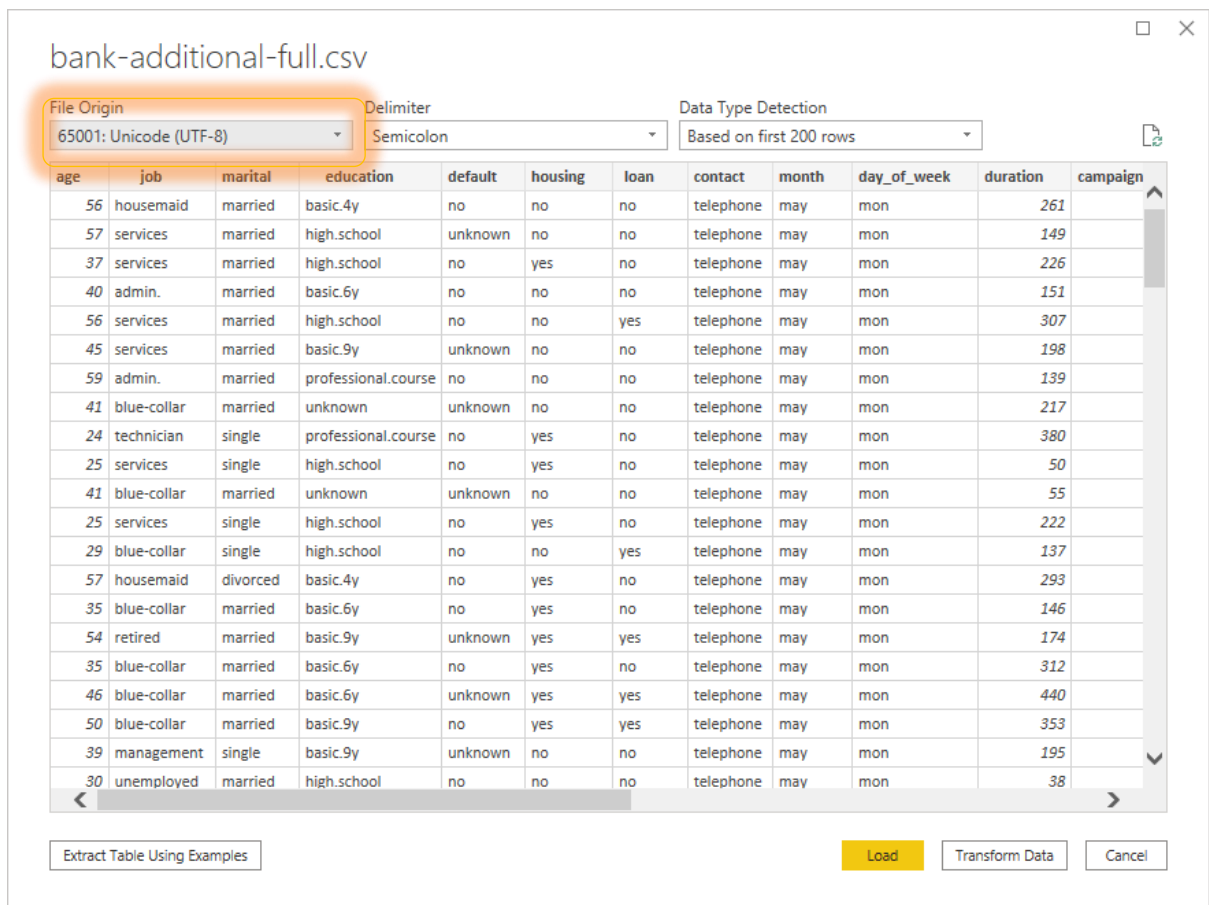
Vào chức năng Get data > Text/CSV để import dữ liệu như hình bên dưới:



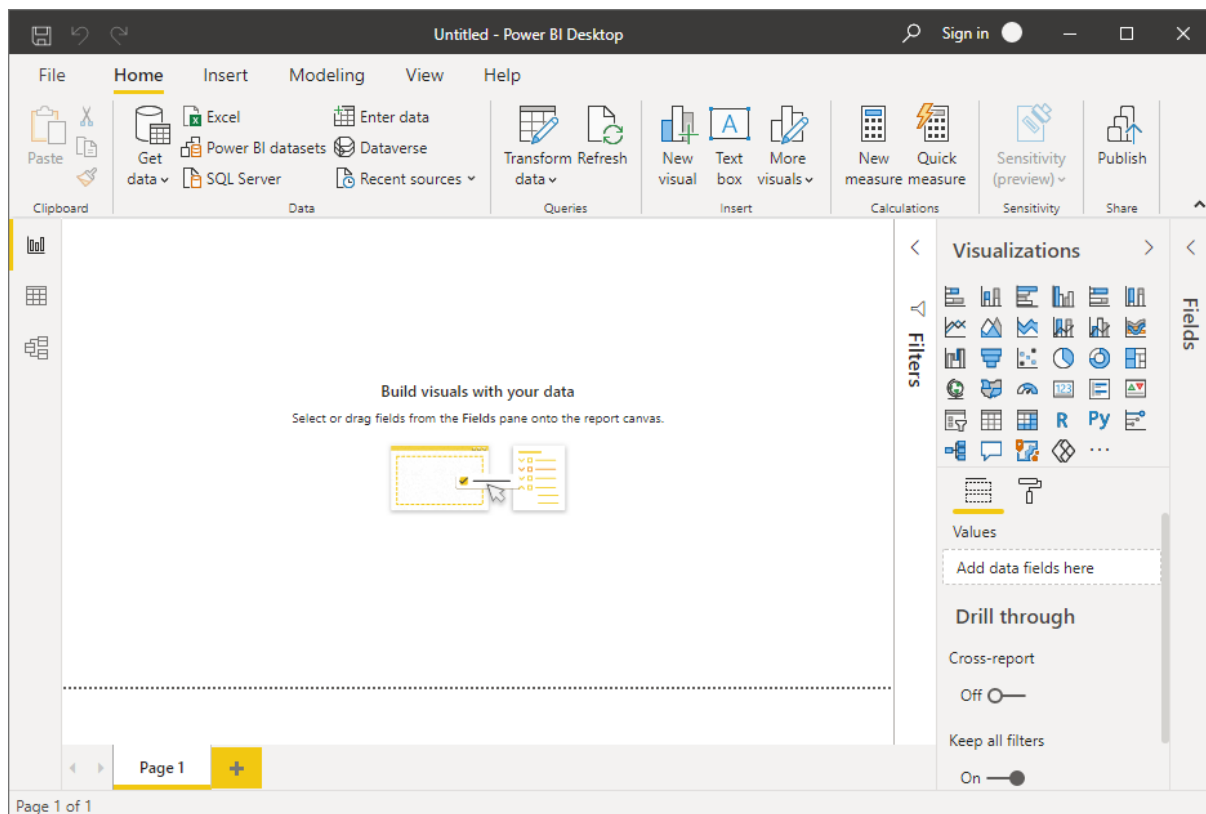
Trong hộp thoại Open, bạn có thể chọn file hoặc dán đường dẫn link Internet như sau:



Vào Chức năng **Get data** trên thanh công cụ, chọn mục More... ở dòng cuối cùng. Trong trường hợp dữ liệu của bạn có tiếng Việt thì chọn dạng mã tương ứng.



Bấm nút Load để tiếp tục.



Tạo biểu đồ

Bước 1: Bấm chọn biểu đồ.

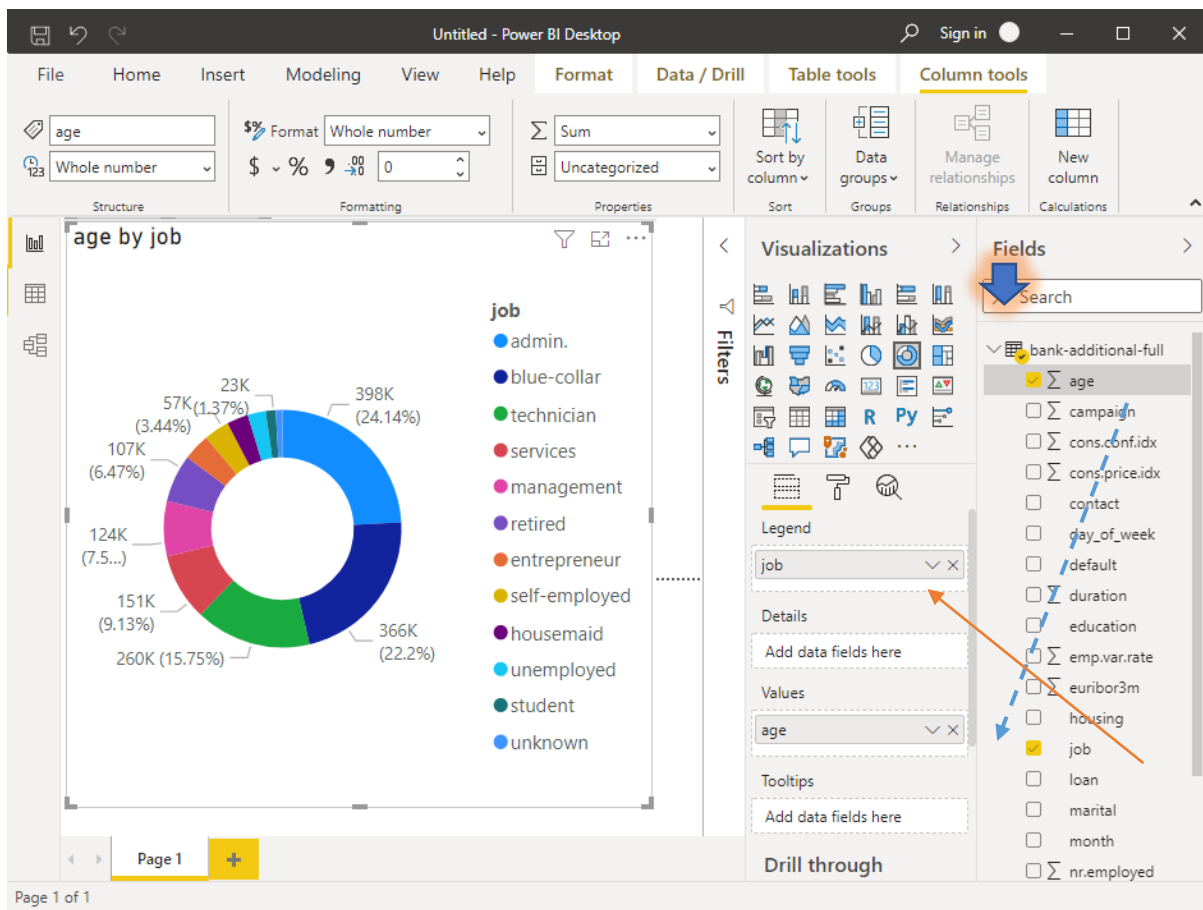
Ví dụ chọn biểu đồ Donut chart.

Bước 2: Chọn cột dữ liệu cần vẽ.

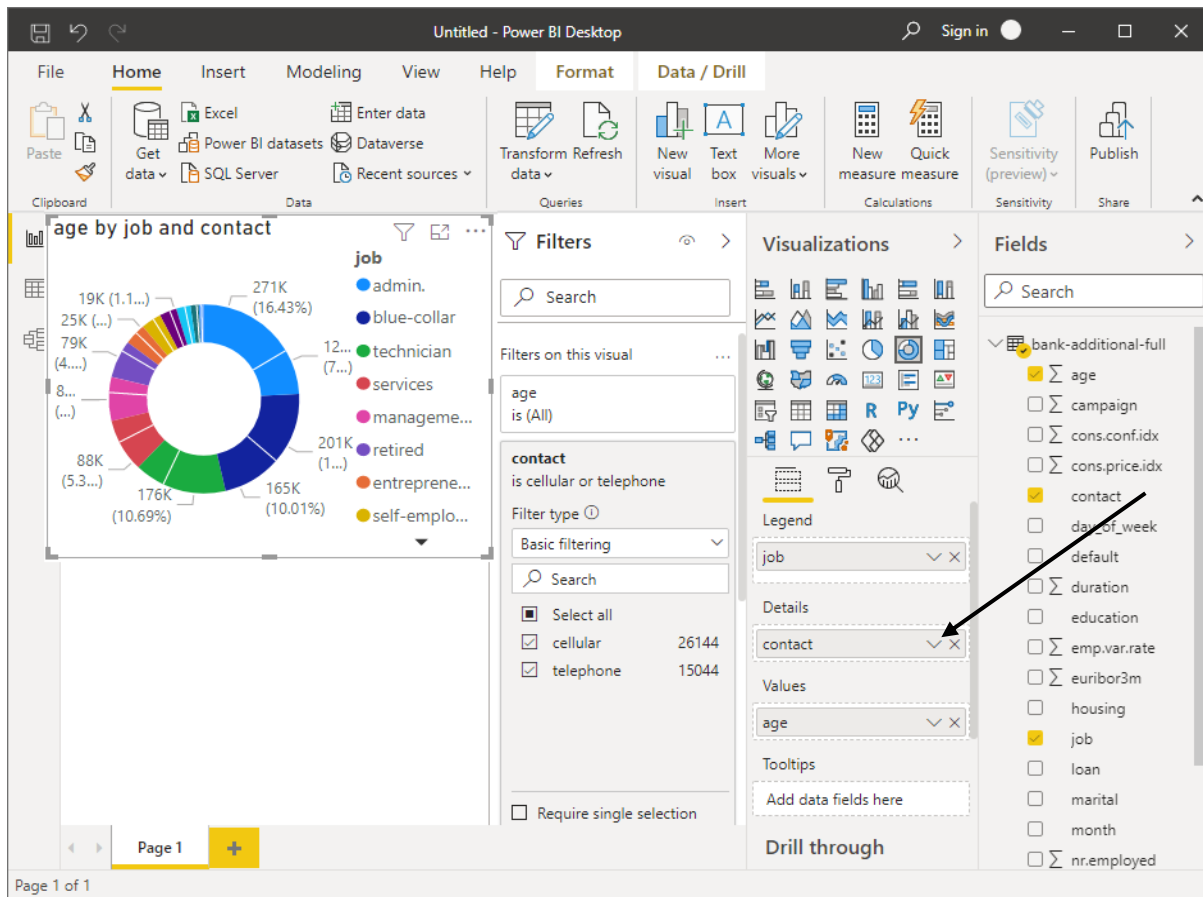
Ví dụ kéo trường dữ liệu **age** vào vùng **Values** (xem mũi tên đứt nét).

Bước 3: Chọn dữ liệu cần phân loại (Legend)

Ví dụ kéo cột dữ liệu **job** vào mục **Legend** (xem mũi tên liền nét).



Biểu đồ trên sẽ cho biết

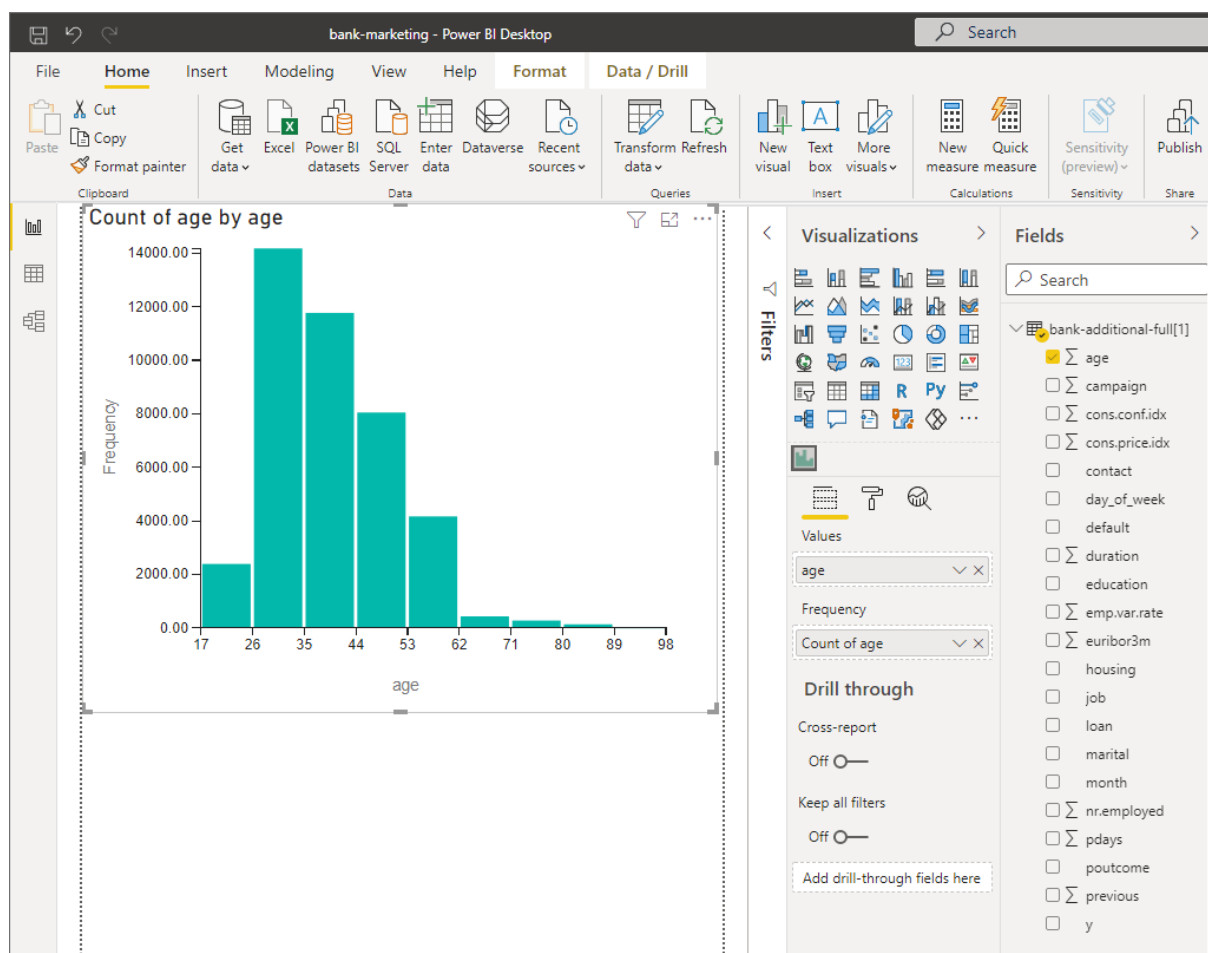


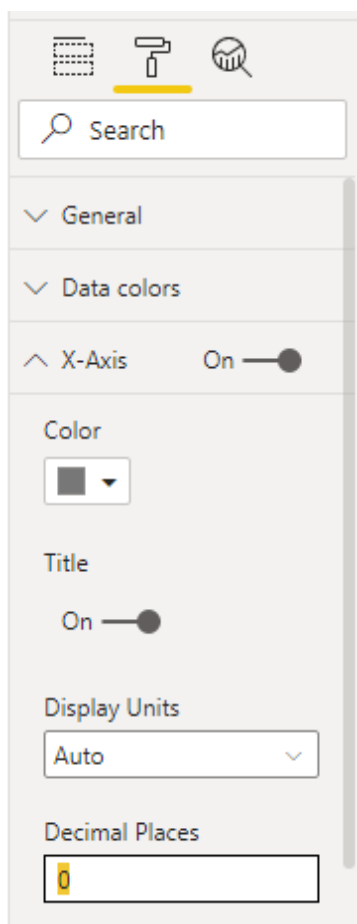
Bài 2 - Phân tích biểu đồ với dữ liệu Bank Marketing

Xem bức tranh tổng thể về dữ liệu

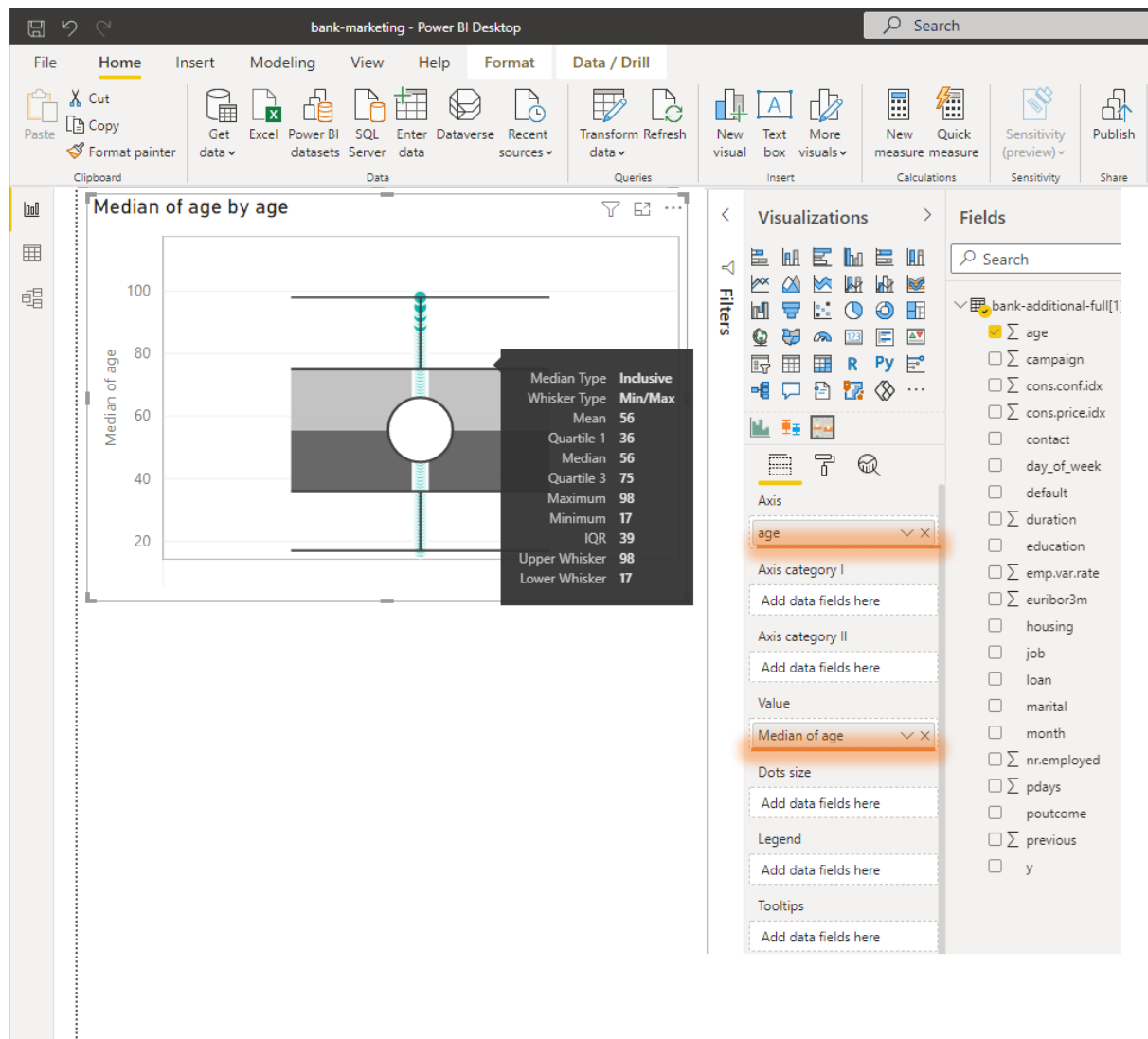
Xem phân bố dữ liệu “tuổi” với histogram

Chart	Histogram
Values	age
Frequency	Count of age
Format	
X-axis	Decimal Places: 0



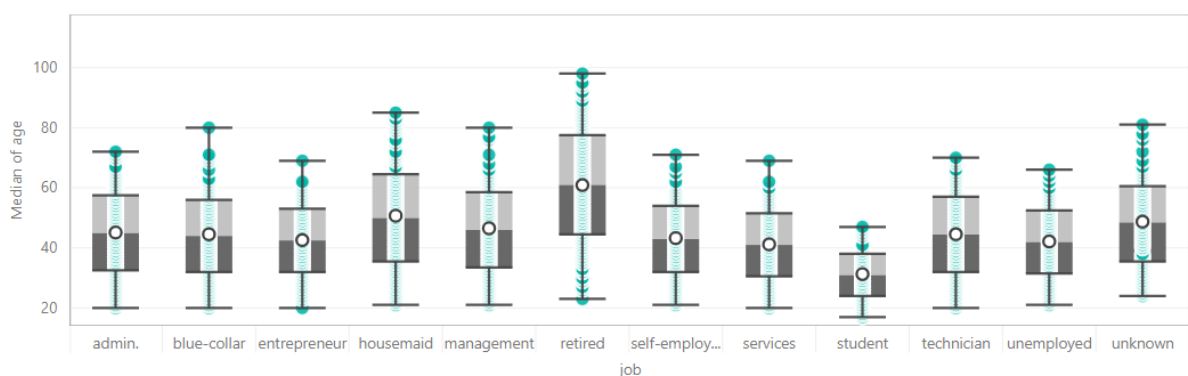


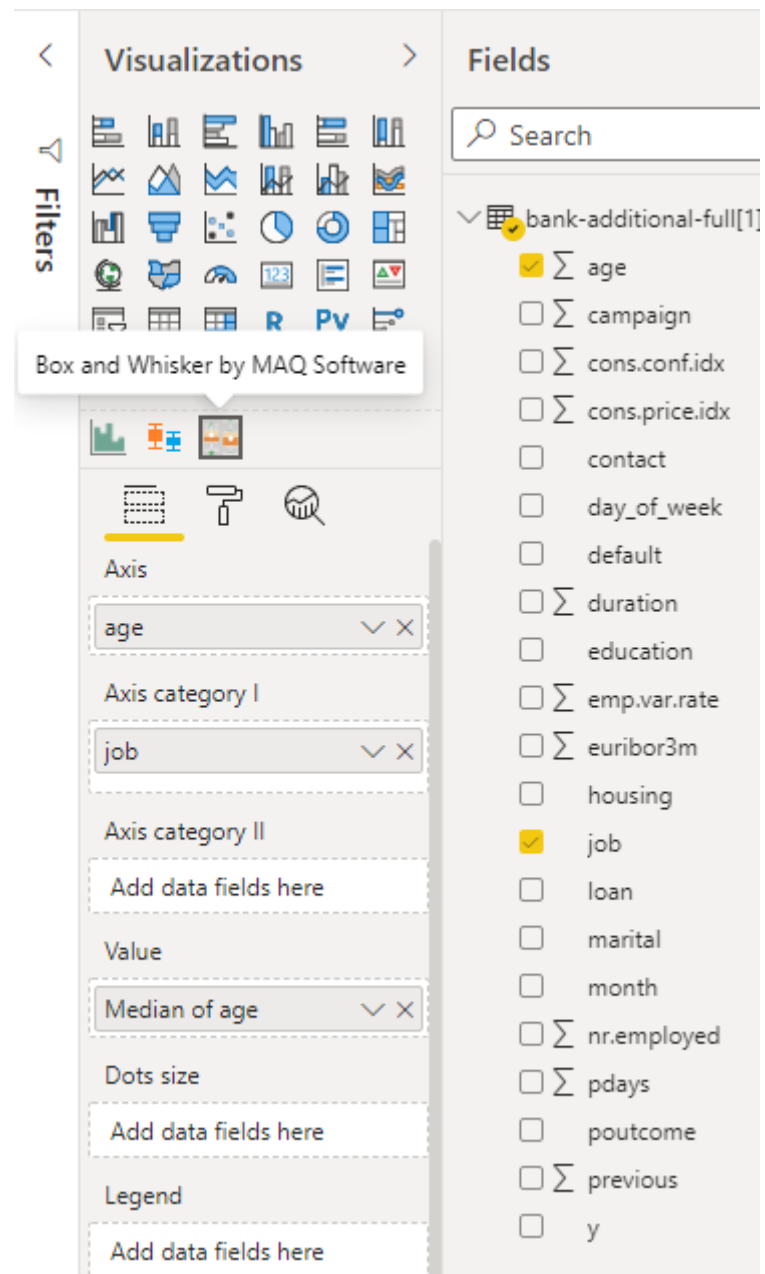
Quan sát tuổi bằng biểu đồ hộp (box)



Xem phân bố tuổi theo công việc

Median of age by age and job

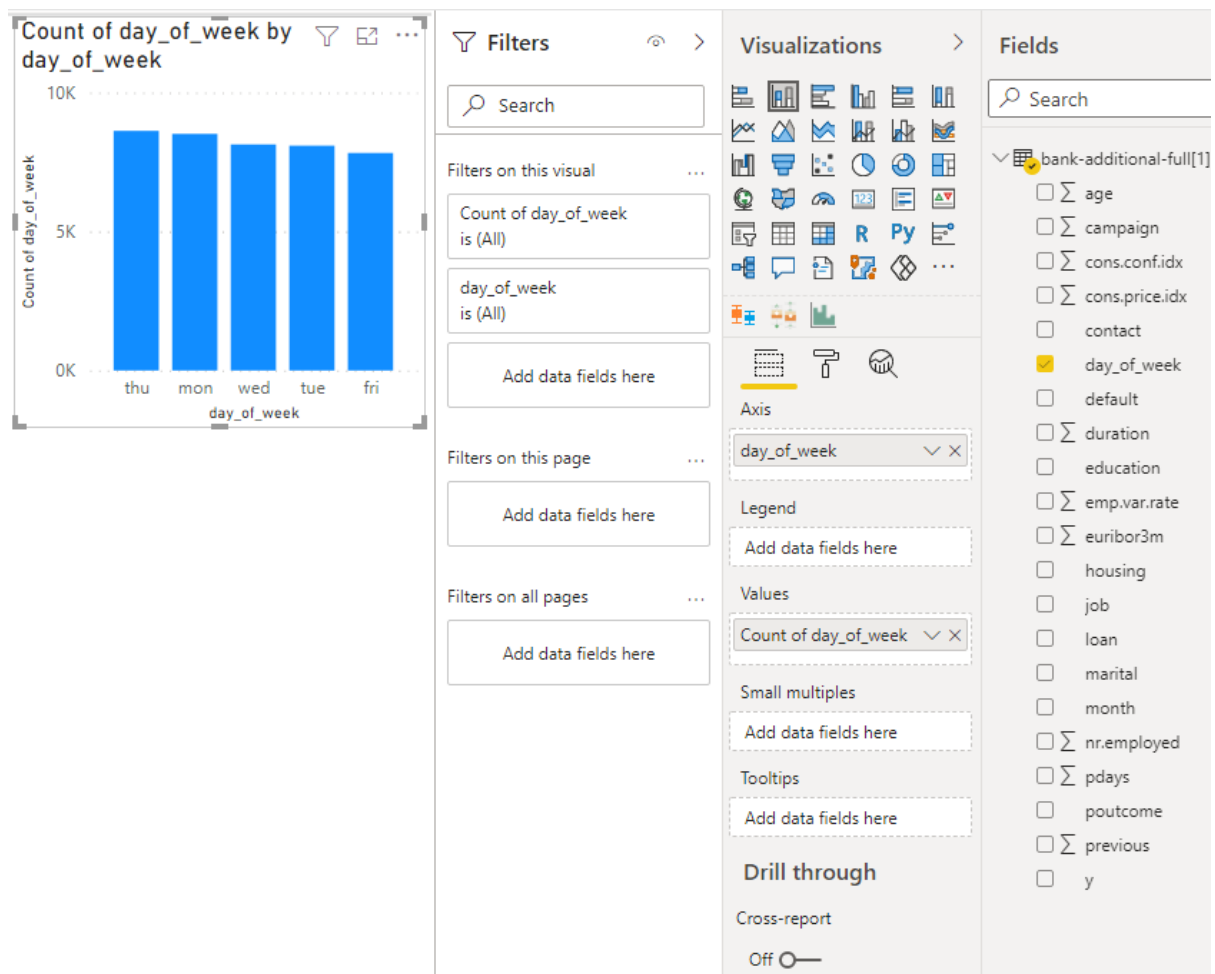




So sánh giá trị của 1 thuộc tính theo các yếu tố khác nhau

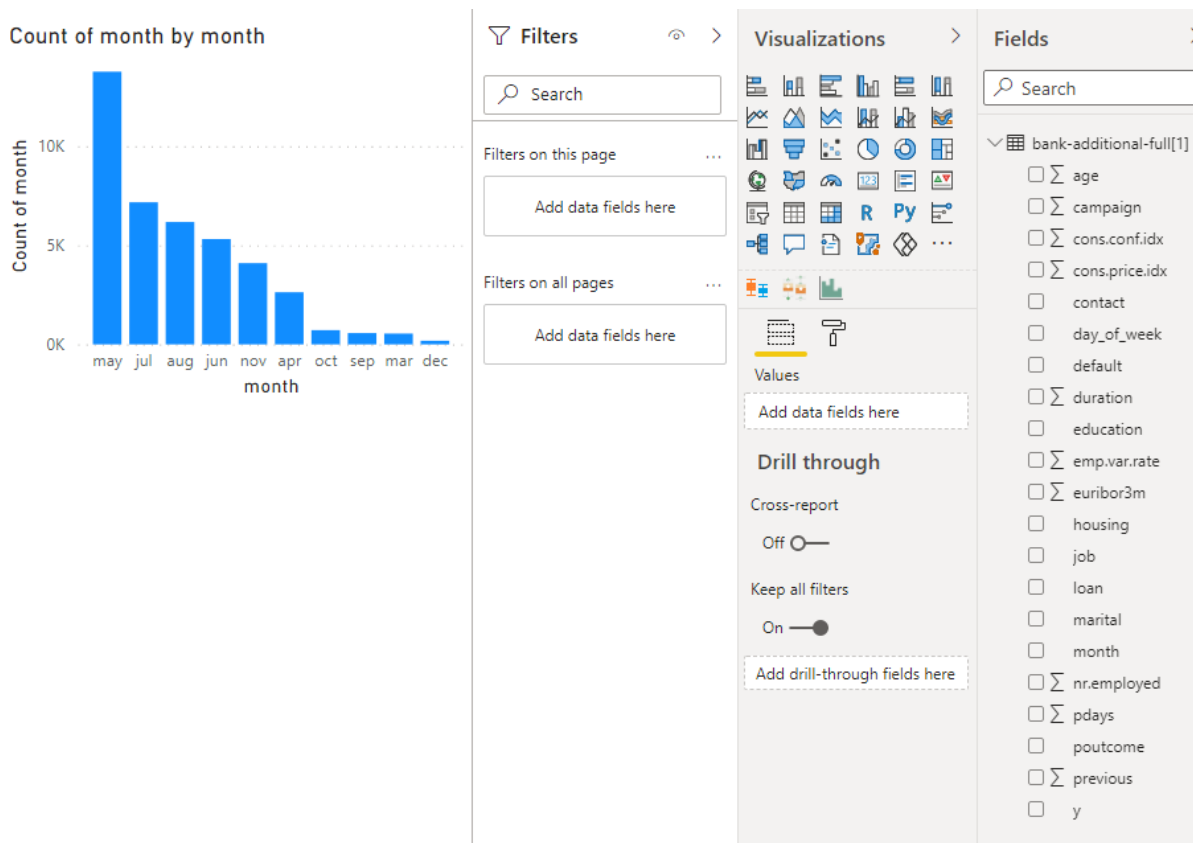
Quan sát số lượng cuộc gọi theo ngày trong tuần

Dùng Biểu đồ thanh để đếm số dòng dữ liệu (hay còn gọi là số quan sát – observation) theo thuộc tính **day_of_week**.



Quan sát số lượng cuộc gọi theo tháng

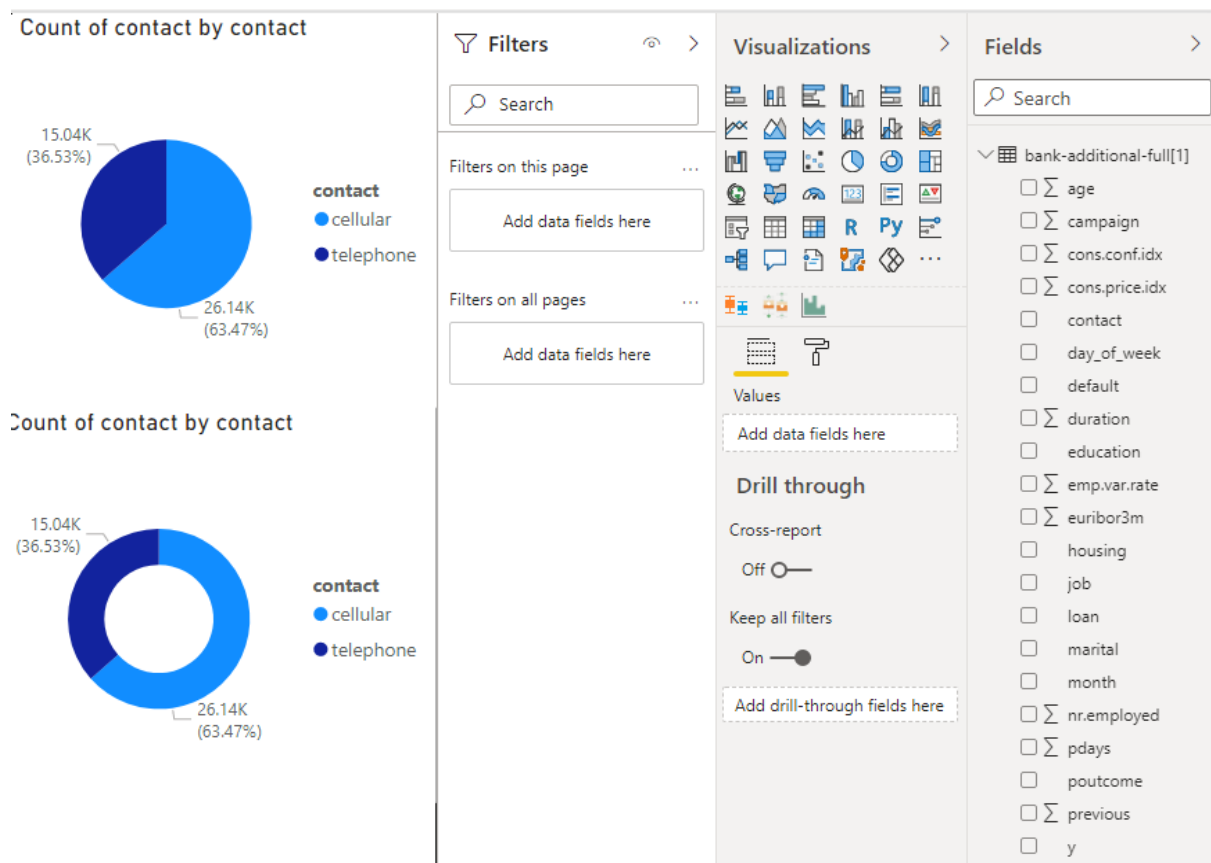
Tương tự, dùng Biểu đồ thanh để đếm số dòng dữ liệu theo thuộc tính **month**.



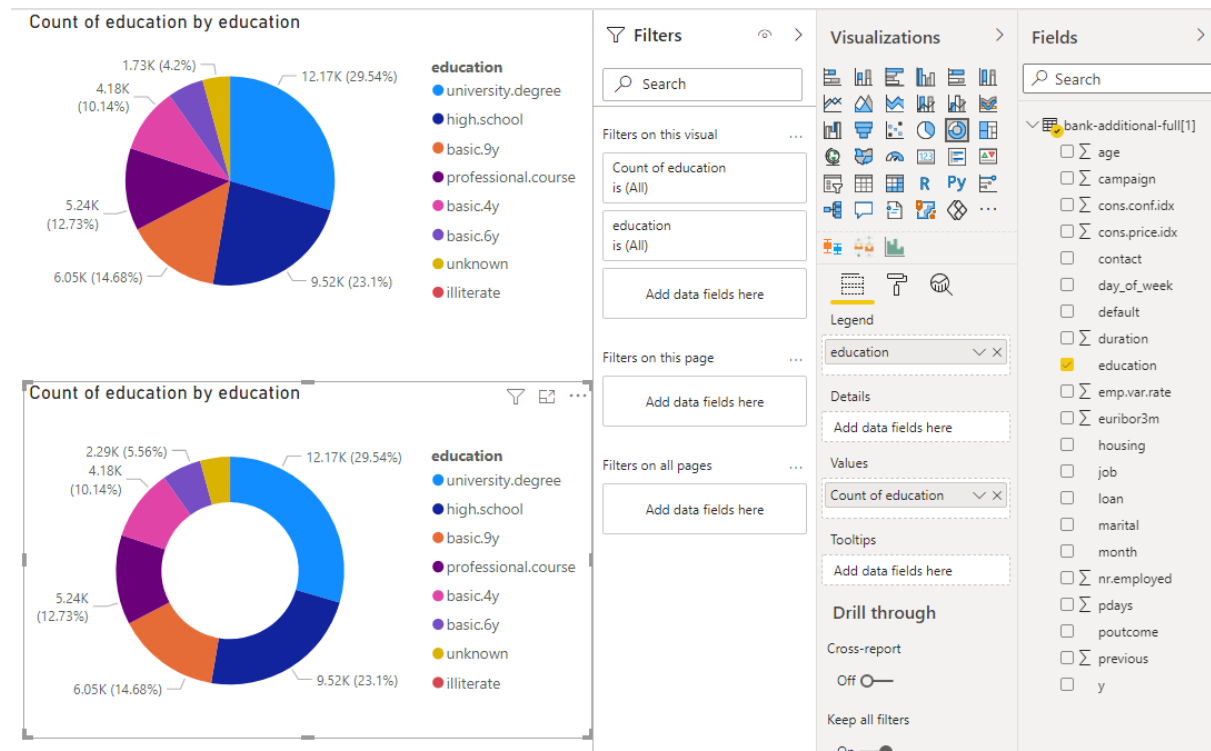
Xem phân bố giá trị của 1 thuộc tính

Tỉ lệ các hình thức liên lạc

Dùng biểu đồ Pie hoặc Donut để xem tỉ lệ các kênh liên lạc với khách hàng.

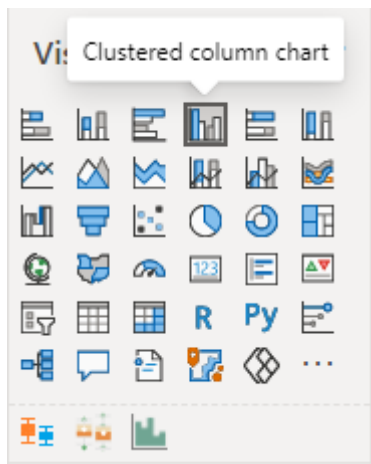


Tỉ lệ trình độ học vấn



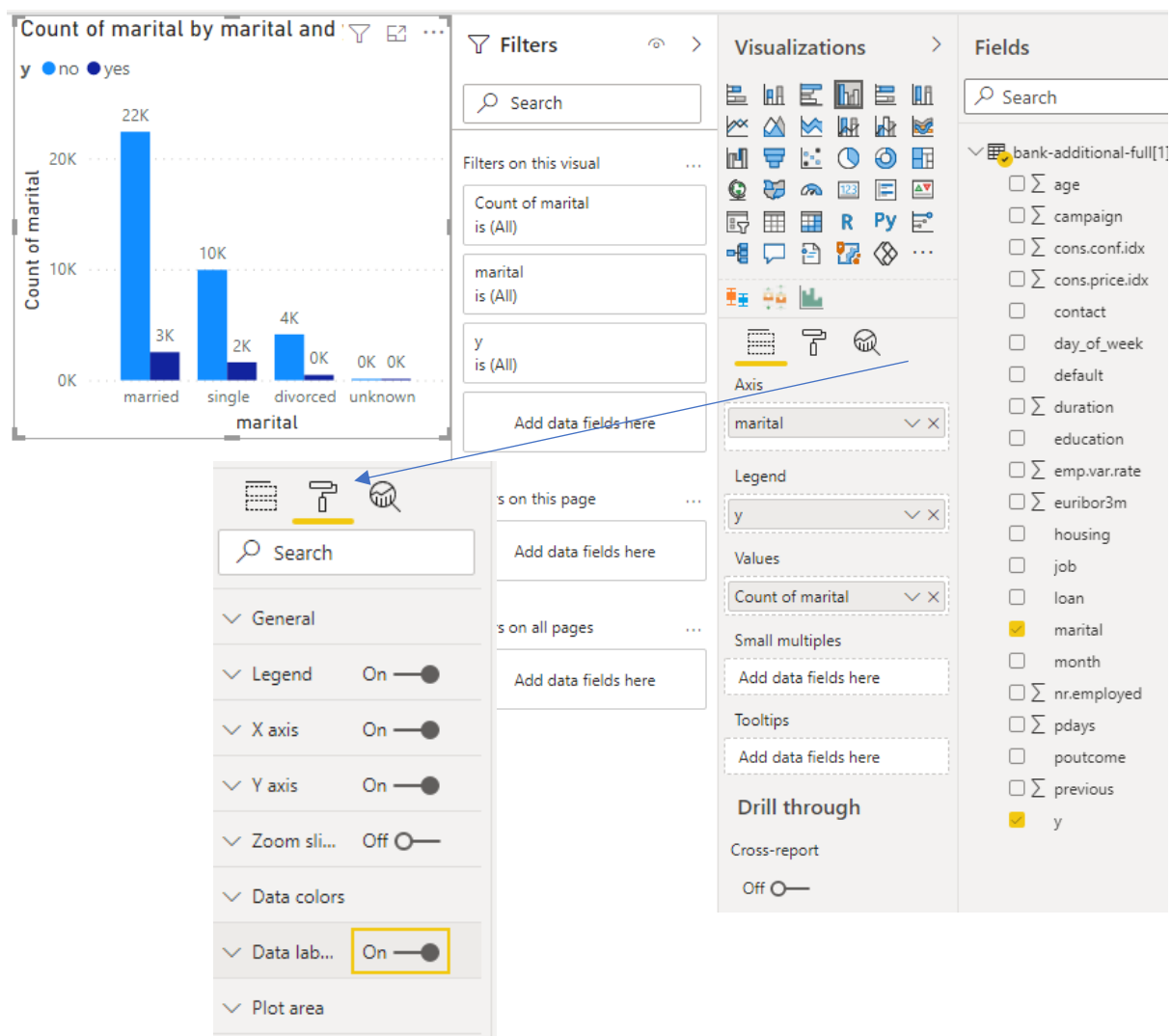
So sánh giá trị của 2 thuộc tính

Dùng clustered column chart



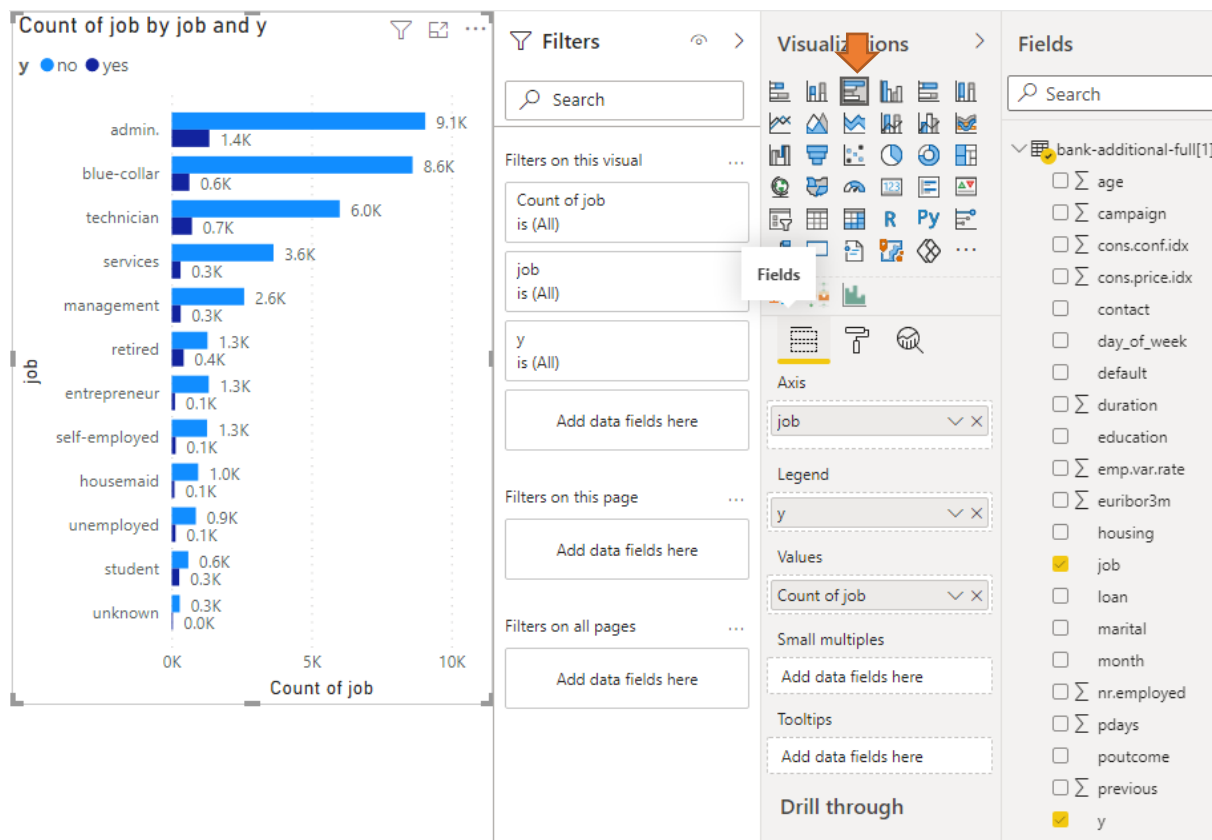
Ví dụ 1

So sánh “Số lượng khách hàng tiềm năng” được gọi và “Số lượng khách hàng sử dụng dịch vụ” theo tình trạng hôn nhân.

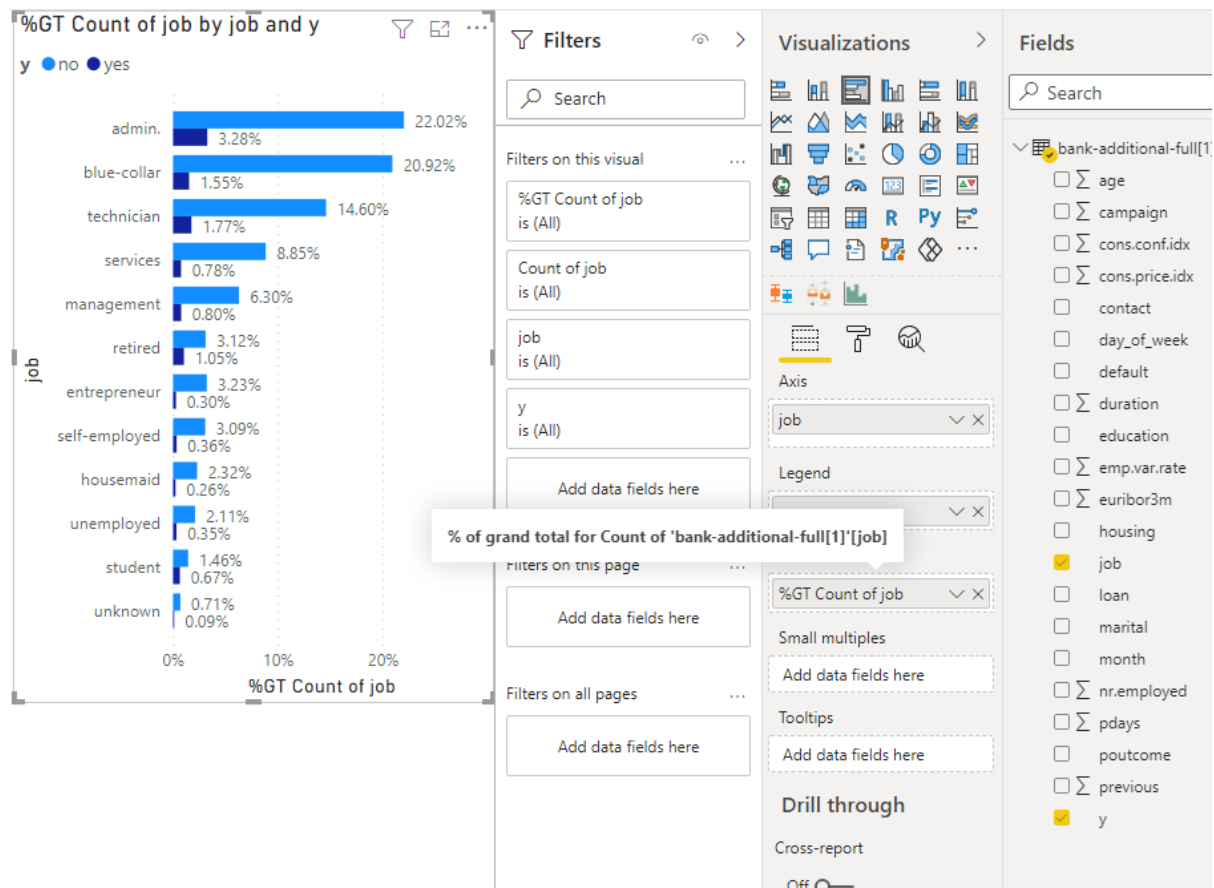


Ví dụ 2

So sánh “Số lượng khách hàng tiềm năng” được gọi và “Số lượng khách hàng sử dụng dịch vụ” phân nhóm theo nghề nghiệp.

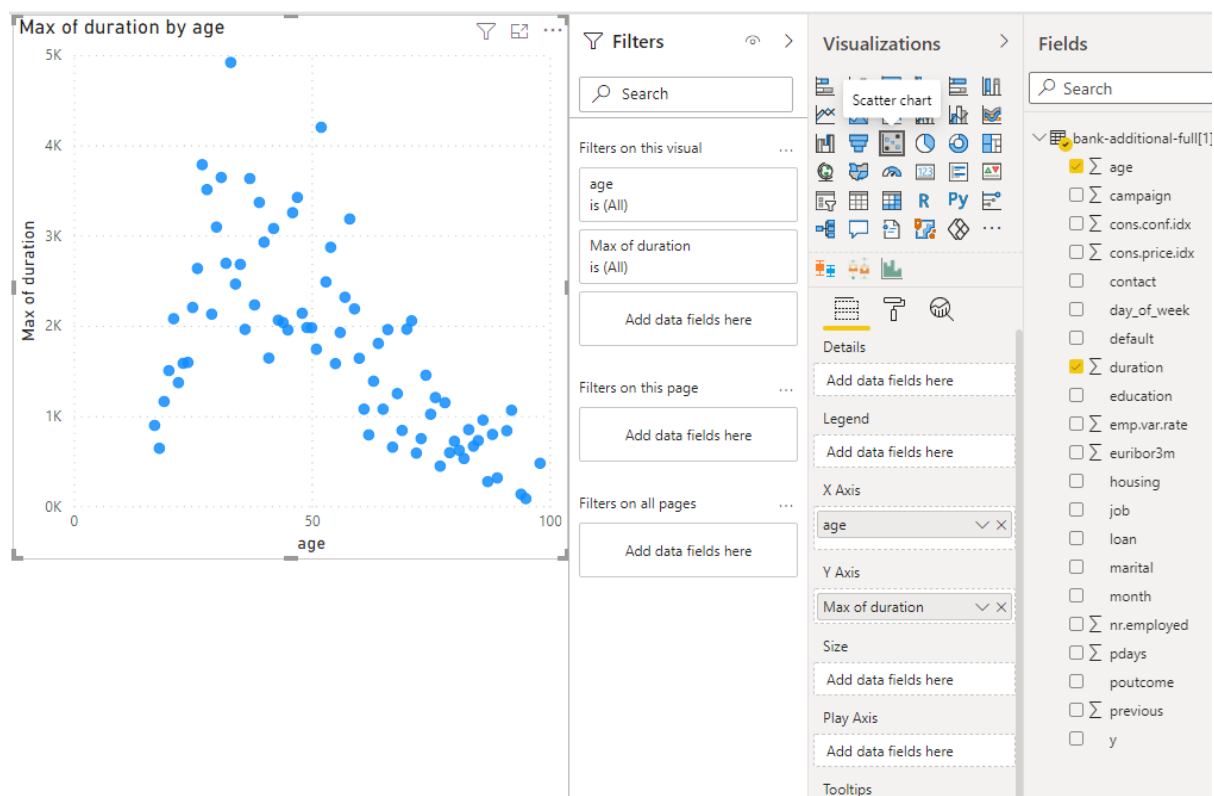


Xem theo phần trăm:



Tìm mối quan hệ giữa 2 thuộc tính

Khảo sát giữa độ tuổi và thời gian cuộc gọi tối đa



Bài 3 - Minh họa biểu đồ theo thời gian

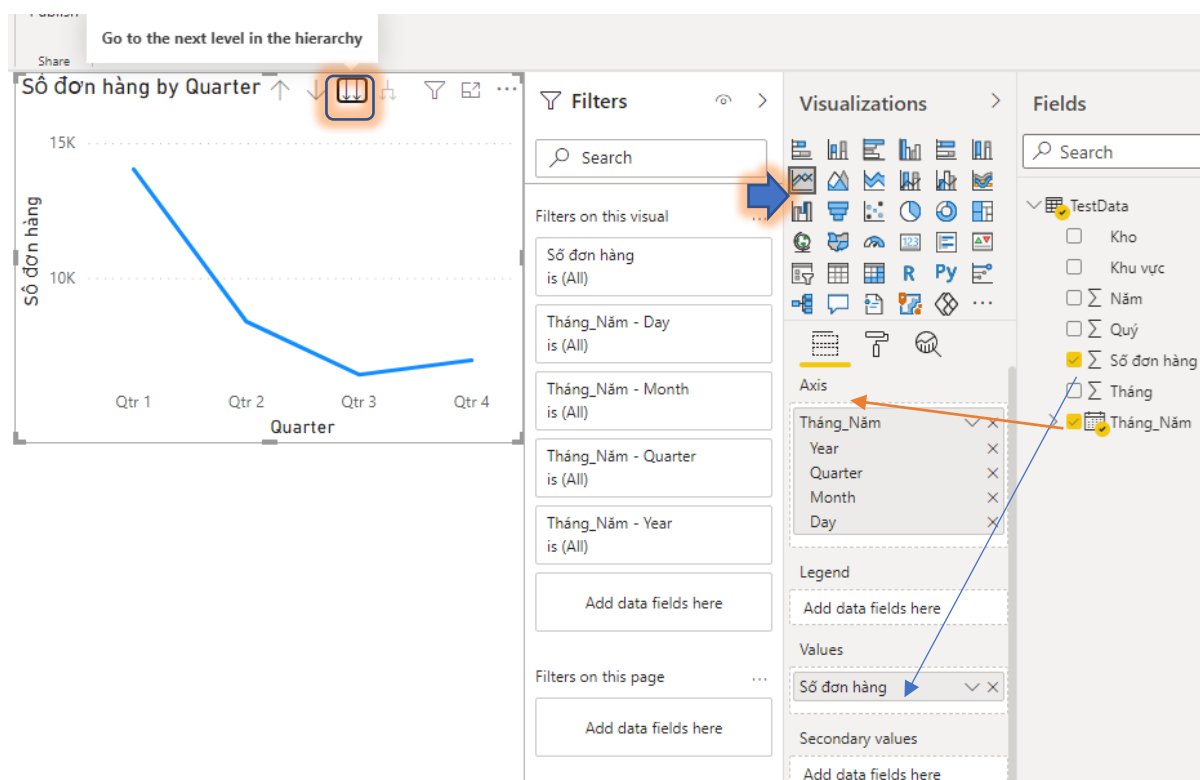
Dữ liệu mẫu tại:

https://thachln.github.io/datasets/sales_report.xlsx

Biểu đồ line theo thời gian

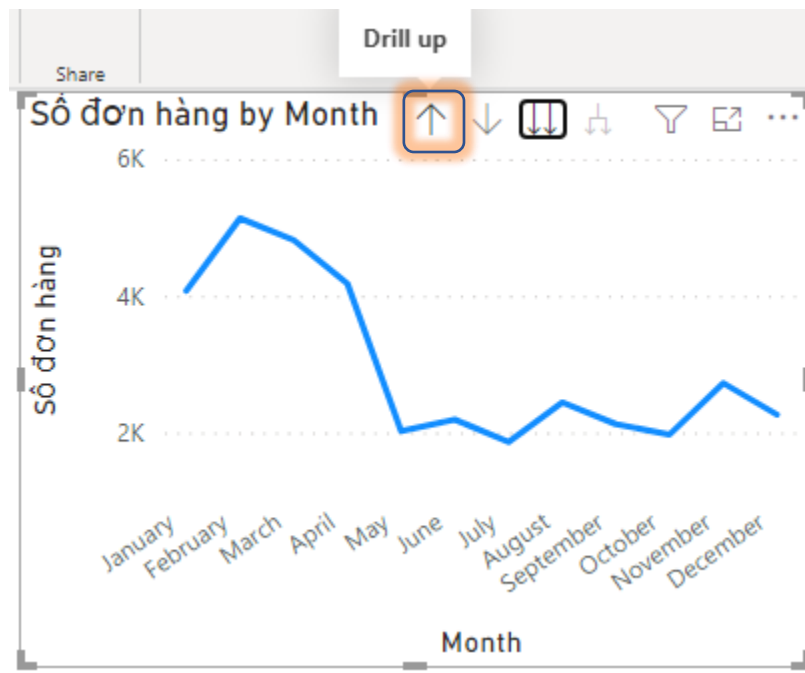
Xem số lượng đơn hàng theo thời gian (cột **Tháng_Năm**) bằng cách sử dụng biểu đồ Line (xem biểu tượng chỗ mũi tên).

Kéo thả trường dữ liệu **Số đơn hàng** vào mục Values; kéo thả trường **Tháng_Năm** vào mục Axis. Kết quả sẽ có biểu đồ như hình bên dưới.



Để xem chi tiết hơn về thời gian thì bấm vào mũi tên (có gợi ý: Click to the next level in the hierarchy).

Muốn xem dữ liệu biểu đồ theo thời gian cấp cao hơn (trong 3 cấp: Năm > Tháng > Ngày) thì bấm vào biểu tượng Drill up trong hình bên dưới (đang xem tổng số đơn hàng theo mức tháng)



Biểu đồ line theo thời gian có chia theo năm

Mở rộng biểu đồ ở phần trước bằng cách kéo thả trường dữ liệu **Năm** vào vùng Legend. Kết quả sẽ có thêm dòng thông tin Năm gồm 2020 và 2021:

