The Power BI service may introduce terminology that is unfamiliar or confusing. Microsoft maintains a community [Glossary](https://docs.microsoft.com/en-us/power-bi/consumer/end-user-glossary) of commonly used terminology to assist users. Another great resource for learning about the building blocks that make up Power BI service is [Basic Concepts](https://docs.microsoft.com/en-us/power-bi/consumer/end-user-basic-concepts). Basic Concepts gives a high level overview of the Power BI *pieces* and how they're connected.

The [Glossary](https://docs.microsoft.com/en-us/power-bi/consumer/end-user-glossary) terms can be found below:

**A**

**account** There are different types of Power BI accounts, also called *licenses*. The accounts are either: per-user or organizational. Per-user accounts are available as *free* or *Pro*. A Power BI *free* user is either using stand-alone Power BI Desktop or is using Power BI service within a Premium organizational account. The Power BI per-user *Pro* account is a paid monthly subscription that allows for collaboration and sharing of content with other *Pro* users.

The organizational *Premium* (also known as *Premium capacity*) account adds a layer of features on top of per-user licenses. For example, *free* per-user account holders within an organization that has a *Premium* license, are able to do much more with Power BI than *free* users without *Premium*. For example, *free* users in *Premium* organizational accounts, can collaborate with colleagues and can view content that’s hosted on Power BI Premium capacity.

**admin portal** The location where Power BI admins manage users, features, and settings for Power BI in their organization. (Note: Office 365, Azure, and PowerApps use admin center.)

**aggregates** When the values of multiple rows are grouped together as input on criteria to form a single value of more significant meaning or measurement.

**aggregation** The reduction of rows in underlying data sources to fit in a model. The result is an aggregate.

**alert, alerts** A feature that notifies users of changes in the data based on limits they set. Alerts can be set on tiles pinned from report visuals. Users receive alerts on the service and on their mobile app.

**annotate** To write lines, text, or stamps on a snapshot copy of a tile, report, or visual on the Power BI mobile app for iOS and Android devices.

**app, apps**

**AppSource** Centralized online repository where you can browse and discover dashboards, reports, datasets, and apps to download.

**ArcGIS for Power BI** ArcGIS is a mapping and analytics platform created by the company ESRI. The name of the visual included in the Power BI visuals library is called ArcGIS for Power BI.

**assign** [Power BI Premium] Verb for moving a workspace into or out of dedicated capacity. Don't use promote for moving a workspace to dedicated capacity because this has the connotation of dedicated being superior to shared capacity.

**Auto Insights** Are now called *Quick Insights*.

**B**

**BI, business intelligence**

**bookmark** A view of data captured in the Bookmarks pane of a report in Power BI Desktop or service. In Desktop, the bookmarks are saved in the .pbix report file for sharing on the Power BI service

**breadcrumbs** The navigation at the top left to quickly navigate between reports and dashboards.

**C**

**calculation** A mathematical determination of the size or number of something.

**card** (visual type) A Power BI [visual type](https://docs.microsoft.com/en-us/power-bi/consumer/end-user-visual-type).

**card** (Power BI Home) Power BI Home displays rectangular and square pictures that represent dashboards, reports, apps, and more. These pictures are referred to as *cards*.

**certified custom visual** A Power BI custom visual that has met requirements and passed strict security testing.

**classification** Also called data classification. This option can be turned on by Power BI admins to raise awareness of the level of security required for a dashboard. The classification types can be customized by an organization.

**connect live** A method of connecting to SQL Server Analysis Services data models. Also called a live connection.

**container** The areas on the navigation pane are *containers*. In the nav pane you'll find containers for: Favorites, Recent, Apps, Shared with me, and Home.

**content** Content for the Power BI service is generally dashboards, reports, and apps. It can also include workbooks and datasets.

**content list** The content index for an app.

**content pack** A collection of pre-packaged datasets, reports, and dashboards. These are being replaced with **apps**, learn [about the new workspace experience](https://docs.microsoft.com/power-bi/service-create-the-new-workspaces).

**content view**

**Continuous variable** - a continuous variable can be any value between its minimum and maximum limits, otherwise it is a discrete variable. Examples are temperature, weight, age, and time. Continuous variables can include fractions or portions of the value. The total number of blue skateboards sold is a discrete variable since we can't sell half a skateboard.

**Correlation** - a correlation tells us how the behavior of things are related. If their patterns of increase and decrease are similar, then they are positively correlated. And if their patterns are opposite, then they are negatively correlated. For example, if sales of our red skateboard increase each time we run a tv marketing campaign, then sales of the red skateboard and the tv campaign are positively correlated.

**cross-filter** Applies to visual interactions. Cross-filtering removes data that doesn't apply. For example, selecting **Moderation** in the doughnut chart cross-filters the line chart. The line chart now only displays data points that apply to the Moderation segment.

**cross-highlight** Applies to visual interactions. Cross-highlighting retains all the original data points but dims the portion that does not apply to your selection. For example, selecting **Moderation** in the doughnut chart cross-highlights the column chart. The column chart dims all the data that does not apply to the Moderation segment and highlights all the data that does apply to the Moderation segment.

**custom visual** Visuals that are created by the community and Microsoft. They can be downloaded from the Office store for use in Power BI reports.

**D**

**dashboard** A selection of visuals that communicate status and metrics. A dashboard organizes and presents data in an easy-to-understand way.

**data classification** See classification

**data connector** See connectors

**data model, Excel Data Model** In Power BI content, a data model refers to a map of data structures in a table format. The data model shows the relationships that are being used to build databases. Report designers, administrators, and developers create and work with data models to create Power BI content.

**dataflow** Dataflows ingest, transform, integrate, and enrich big data by defining data source connections, ETL logic, refresh schedules, and more. Formerly data pool.

**dataset** vs **data source** A **dataset** is a collection of data used to create visualizations and reports.

**dedicated capacity** [Power BI Premium] Data models running on dedicated hardware in Microsoft cloud data centers to help ensure consistent performance at scale. BI solutions are delivered to the entire organization regardless of Power BI license.

**Desktop** or **Power BI Desktop** Free Power BI tool used primarily be report designers, admins, and developers.

**diamond** Power BI Premium. The shape of the icon that signifies a workspace is a Premium capacity workspace.

**Dimension** - dimensions are categorical (text) data. A dimension describes a person, object, item, products, place, and time. In a a dataset, dimensions are a way to group *measures* into useful categories. For our skateboard company, some dimensions might include looking at sales (a measure) by model, color, country, or marketing campaign.

**drill up, drill down, drillthrough** In Power BI, use drill down and drill up to refer to the ability to explore the next level of detail in a report or visual. Use drillthrough to refer to the ability to select a part of a visual and be taken to another page in the report, filtered to the data that relates to the part of the visual you selected on the original page. Drill to details commonly means to show the underlying records.

**E**

**Editing View** The mode in which report *designers* can explore, design, build, and share a report.

**ellipsis** ...  
Selecting an ellipsis displays additional menu options. Also referred to as the **More actions** menu.

**embed code** A common standard across the internet. In Power BI, the customer can generate an embed code and copy it to place content such as a report visual on a website or blog.

**Embedded** See Power BI Embedded.

**embedding** In the Power BI developer offering, the process of integrating analytics into apps using the Power BI REST APIs and the Power BI SDK.

**environment** [Power BI Desktop, Power BI Mobile, the Power BI service, etc.] Another way to refer to one of the Power BI tools. It's okay to use Power BI environment (tenant) in documentation where it may help business analysts who are familiar with the term tenant to know it's the same thing.

**F**

**favorite, unfavorite** Verb meaning to add to the Favorites list for quick access to frequently visited dashboards and reports in Power BI. When you no longer want them as a favorite, you unfavorite them.

**filter** versus **highlight** A filter removes data that does not apply. A highlight grays out the data that does not apply.

**focus mode** Use focus mode to pop out a visual or tile to see more detail. You can still interact with the visual or tile while in focus mode.

**Free account** See *account*

**full screen, full-screen mode** Use full screen mode to view Power BI content without the distraction of menus and navigation panes. This is sometimes referred to as TV mode.

**G**

**gateways** or **on-premises data gateways** A bridge to underlying data sources. It provides quick and secure data transfer between the Power BI service and on-premises data sources that support refresh. Usually managed by IT.

**H**

**high-density visuals** Visuals with more data points than Power BI can render. Power BI samples the data to show the shape and outliers.

**Home** The default landing age for Power BI service users. Doesn't modify anything. Can be called Power BI Home or simply Home.

**I**

**insights** See **quick insights**.

**J**

**K**

**KPIs** Key performance indicators. A type of visual.

**L**

**left navigation (left nav)** This has been replaced with **nav pane** but may still appear in some documentation.  
The controls along the left edge of Power BI service.

First instance: navigation pane Subsequent mentions or tight spaces: nav pane

**license** See *account*.

**list page** or **content list** One of the section pages for the elements in the nav pane. For example, Favorites, Recents, My workspace, etc.

**M**

**Measure** - a measure is a quantitative (numeric) field that can be used to do calculations. Common calculations are sum, average, and minimum. For example, if our company makes and sells skateboards, our measures might be number of skateboards sold and average profit per year.

**Mobile app** Apps that allow you to run Power BI on iOS, Android, and Windows devices.

**modeling** [Power BI Desktop] Getting the data you've connected to ready for use in Power BI. This includes creating relationships between tables in multiple data sources, creating measures, and assigning metrics.

**My workspace** The workspace for each Power BI customer to use to create content. If they want to bundle anything created here into an app, and they have *designer* permissions, they upload it to the appropriate workspace or create a new one.

**N**

**native** Included with the product. For example, Power BI comes with a set of *native* visualization types. But you can also import other types, such as Power BI visuals.

**navigation pane** or **nav pane** The controls along the left edge of the Power BI service.

First instance: navigation pane Subsequent mentions or tight spaces: nav pane

**notification** Messages sent by and to the Power BI Notification Center.

**Notification Center** The location in the service where messages are delivered to users, such as notice of sunsetting certain features.

**O**

**OneDrive for Business** vs **OneDrive** OneDrive is a personal account and OneDrive for Business is for work accounts.

**On-premises** The term used to distinguish local computing (in which computing resources are located on a customer’s own facilities) from cloud computing.

**P**

**PaaS** PaaS stands for platform as a service. For example, Power BI Embedded.

**page** Reports have one or more pages. Each tab on the report canvas represents a page.

**pbiviz** The file extension for a Power BI custom visual.

**pbix** The file extension for a Power BI Desktop file.

**permissions** What a user can and can't do in Power BI is based on permissions. As a *consumer* you won't have the same permissions as a *designer*, administrator, or developer.

**phone report** The name for a Power BI report that's been formatted for viewing on the phone.

**phone view** The user interface in the Power BI service for laying out a phone report.

**pin, unpin** The action a report *designer* takes of placing a visual, usually from a report, onto a dashboard.

**Power BI**, **Power BI service**, **Power BI Desktop**, **Power BI mobile**, Some of the Power BI offerings. *Power BI* is the general term. It is often used in place of a full product name, such as *Power BI service* and *Power BI mobile*, after the first mention of the full product name.

**Power BI Desktop** Also referred to as *Desktop*. The free Windows application of Power BI you can install on your local computer that lets you connect to, transform, and visualize your data. Used by report designers and admins. For more information, see [What is Power BI](https://docs.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview).

**Power BI Embedded** A product used by developers to embed Power BI dashboards and reports into their own apps, sites, and tools.

**Power BI Premium** An add-on to the Power BI Pro license that enables organizations to predictably scale BI solutions through the purchasing of dedicated hardware. See *account*.

**Power BI Pro** A monthly per-user license that provides the ability to build reports and dashboards, collaborate on shared data, keep data up-to date automatically, audit and govern how data is accessed and used, and the ability to package content to distribute (Power BI apps). See *account*.

**Power BI Report Server** A new branded offering of SSRS launching with Power BI Premium to allow organizations to build distributed, hybrid BI systems (a mix of cloud and on-premises deployments). Embedded provides a localized way to store and manage Power BI reports.

**Power BI service** An online SaaS (Software as a service) service. For more information, see [What is Power BI](https://docs.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview).

**Premium workspace** A workspace running in dedicated capacity, signified to customers by a diamond icon.

**Pro** license or **Pro** account See *account*.

**publish** Power BI service report *designers* bundle the contents of a Power BI workspace to make it available to others as a Power BI app. Power BI Desktop report *designers* use publish to refer to sending a Power BI Desktop report in .pbix format to the Power BI service so that they can build dashboards from it and easily share it with others.

**Q**

**Q&A** The ability to type natural language questions about a dataset and get responses in the form of visualizations. Appears in the Power BI service and Desktop.

**Q&A virtual analyst**  
[Power BI Mobile] For iOS, the conversational UI for Q&A.

**QR codes** [Power BI Mobile] A matrix barcode that can be generated for dashboards or tiles in the Power BI service to identify products. QR codes can be scanned with a QR code reader, or with the Power BI Mobile app on iOS or Android, to link directly to the dashboard or tile.

**query string parameter**  
Add to a URL to pre-filter the results seen in a Power BI report. In the broadest sense, a query string recovers information from a database.

**Quick Insights** Quick Insights refer to automatically generated insights that reveal trends and patterns in data.

**R**

**R, Microsoft R** R is a programming language and software environment for statistical computing and graphics.

**Reading View** Read-only view for reports (as opposed to Editing View).

**real-time streaming** The ability to stream data and update dashboards in real time from sources such as sensors, social media, usage metrics, and anything else from which time-sensitive data can be collected or transmitted.

**Recent** The container in the nav pane that holds all the individual artifacts that were accessed last.

**related content** Shows the individual pieces of content that contribute to the current content. For example, for a dashboard, you can see the reports and datasets providing the data and visualizations on the dashboard.

**relative links** Links from dashboard tiles to other dashboards and reports that have been shared directly or distributed through a Power BI app. This enables richer dashboards that support drillthrough.

**report**  
A multi-perspective view into a single dataset, with visualizations that represent different findings and insights from that dataset. Can have a single visualization or many, a single page or many pages.

**report editor** The report editor is where new reports are created and changes are made to existing reports by report *designers*.

**report measures** Also called custom calculations. Excel calls these *calculated fields*. See also *measures*.

**responsive visuals** Visuals that change dynamically to display the maximum amount of data and insights, no matter the screen size.

**row-level security, RLS** Power BI feature that enables database administrators to control access to rows in a database table based on the characteristics of the user executing a query (for example, group membership).

Administrators can configure RLS for data models imported into Power BI with Power BI Desktop.

**S**

**SaaS**  
Software as a service (or SaaS) is a way of delivering applications over the internet—as a web-based service. Also referred to as: web-based software, on-demand software, or hosted software.

**screenshot** Simple screenshots of a report can be emailed using the send a screenshot feature.

**service** See *Power BI service* A standalone resource available to customers by subscription or license. A service is a product offering delivered exclusively via the cloud.

**Settings** The location for Power BI users to manage their own general settings, such as whether to preview new features, set the default language, close their account, etc. Also, users manage individual settings for content assets, alerts, and subscriptions. Represented by a cog icon.

**share, sharing** In Power BI, sharing typically means directly sharing an individual artifact (a dashboard or report) with one or more people by using their email address. Requires a Power BI Pro license for sender and recipient. On mobile devices, share can refer to native OS share functionality, such as "annotate and share."

**Shared with me** The container in the nav pane that holds all the individual artifacts that were directly shared by another Power BI user.

**snapshot** In Power BI, a snapshot is a static image vs. a live image of a tile, dashboard, or report.

**SQL Server Analysis Services (SSAS)**

**SQL Server Reporting Services (SSRS)**

**streaming data** See *real-time streaming*.

**subscriptions**, **Subscribe**  
You can subscribe to report pages, apps, and dashboards and receive emails containing a snapshot. Requires a Power BI Pro license.

**summarization** [Power BI Desktop] The operation being applied to the values in one column.

**T**

**tiles**  
Power BI dashboards contain report tiles.

**Time series** - a time series is a way of displaying time as successive data points. Those data points could be increments such as seconds, hours, months, or years.

**U**

**V**

**value, values** Numerical data to be visualized.

**visual, visualization** A chart. Some visuals are: bar chart, treemap, doughnut chart, map.

**Visual interaction** One of the great features of Power BI is the way all visuals on a report page are interconnected. If you select a data point on one of the visuals, all the other visuals on the page that contain that data change, based on that selection.

**Visualizations pane** Name for the visualization templates that ship in the shared report canvas for Power BI Desktop and the Power BI service. Contains small templates, also called icons, for each native visualization type.

**W**

**workbook** An Excel workbook to be used as a data source. Workbooks can contain a data model with one or more tables of data loaded into it by using linked tables, Power Query, or Power Pivot.

**workspace** Containers for dashboards, reports, and datasets in Power BI. Users can collaborate on the content in any workspace except My workspace. The contents can be bundled into a Power BI app. Those stored in Premium capacity can be shared with Free users. Personal workspaces (under My workspace) can be hosted in Premium capacity.

**X**

**x-axis**  
The axis along the bottom, the horizontal axis.

**Y**

**y-axis** The axis along the side, the vertical axis.

**Z**



#### Your goals:

* Create a Power BI report that analyzes wine review data. The data was scraped from WineEnthusiast and contains scores, prices, and other contextual information for thousands of wines from across the world.

#### Techniques you’ll need to use:

* Import data from CSV file
* Create explicit measures with DAX using formulas like [AVERAGE](https://docs.microsoft.com/en-us/dax/average-function-dax), [COUNT](https://docs.microsoft.com/en-us/dax/count-function-dax), [MEDIAN](https://docs.microsoft.com/en-us/dax/median-function-dax), [MIN](https://docs.microsoft.com/en-us/dax/min-function-dax), and [MAX](https://docs.microsoft.com/en-us/dax/max-function-dax).
* Add charts and other visuals

#### Data sources:

* Link: [Week 1 Wrap-Up Exercise (ZIP)](https://courses.edx.org/assets/courseware/v1/10a8b2748808bf889a946bff061cacc6/asset-v1:DavidsonX+DavidsonX.D005+1T2022a+type@asset+block/Week_1_Wrap-Up_Exercise.zip)
* Source: Kaggle (https://www.kaggle.com/zynicide/wine-reviews)

#### Tips:

* Try to find out which wineries, countries, or regions have the highest points rating, or which varieties have the highest cost.
* Try using a map visual to show the data, such as average points per country.

### Week 2: Exercise Files

Please download the necessary files below to get ready for Week 2.

* Stephen Curry basketball: [CurryShots.csv](https://courses.edx.org/assets/courseware/v1/01d141ccfd0e031287fbe869c7dd497c/asset-v1:DavidsonX+DavidsonX.D005+1T2022a+type@asset+block/CurryShots.csv), [nba\_court.jpg](https://courses.edx.org/assets/courseware/v1/ff9be4a6ebdeee5ca203ea04101c2dde/asset-v1:DavidsonX+DavidsonX.D005+1T2022a+type@asset+block/nba_court.jpg), [NBA\_teamlist.csv](https://courses.edx.org/assets/courseware/v1/95b1421820074daf0604eafe3dad9252/asset-v1:DavidsonX+DavidsonX.D005+1T2022a+type@asset+block/NBA_teamlist.csv)
* World Development Indicators: [countries.csv](https://courses.edx.org/assets/courseware/v1/ee12a324887a80d0d898ffd2e1a3d055/asset-v1:DavidsonX+DavidsonX.D005+1T2022a+type@asset+block/countries.csv), [Country-GDP.csv](https://courses.edx.org/assets/courseware/v1/de411bb9061639298e2b030da97769ec/asset-v1:DavidsonX+DavidsonX.D005+1T2022a+type@asset+block/CountryGDP2.csv), [Country-Population.csv](https://courses.edx.org/asset-v1:DavidsonX+DavidsonX.D005+1T2021+type@asset+block@Country-Population_1_.csv)

GDP = CALCULATE(SUM(Indicators[Value]), Indicators[Indicator] = "NY.GDP.MKTP.CD")

Population = CALCULATE(SUM(Indicators[Value]), Indicators[Indicator] = "SP.POP.TOTL")

GDP Per Capita = DIVIDE([GDP], [Population], 0)

GDP Growth =

VAR FirstYear =

    CALCULATE ( MIN ( Indicators[Year] ), Indicators[Indicator] = "NY.GDP.MKTP.CD" )

VAR LatestYear =

    CALCULATE ( MAX ( Indicators[Year] ), Indicators[Indicator] = "NY.GDP.MKTP.CD" )

VAR FirstYearGDP =

    CALCULATE ( [GDP], Indicators[Year] = FirstYear )

VAR LatestYearGDP =

    CALCULATE ( [GDP], Indicators[Year] = LatestYear )

VAR Diff =

    LatestYearGDP - FirstYearGDP

RETURN

DIVIDE ( Diff, FirstYearGDP, 0 )

Population Growth =

VAR FirstYear =

    CALCULATE ( MIN ( Indicators[Year] ), Indicators[Indicator] = "SP.POP.TOTL" )

VAR LatestYear =

    CALCULATE ( MAX ( Indicators[Year] ), Indicators[Indicator] = "SP.POP.TOTL" )

VAR FirstYearPop =

    CALCULATE ( [Population], Indicators[Year] = FirstYear )

VAR LatestYearPop =

    CALCULATE ( [Population], Indicators[Year] = LatestYear )

VAR Diff =

    LatestYearPop - FirstYearPop

RETURN

DIVIDE ( Diff, FirstYearPop, 0 )

DP = CALCULATE(SUM(Indicators[Value]), Indicators[Indicator] = "NY.GDP.MKTP.CD")

Population = CALCULATE(SUM(Indicators[Value]), Indicators[Indicator] = "SP.POP.TOTL")

GDP Per Capita = DIVIDE([GDP], [Population], 0)

GDP Growth =

VAR FirstYear =

    CALCULATE ( MIN ( Indicators[Year] ), Indicators[Indicator] = "NY.GDP.MKTP.CD" )

VAR LatestYear =

    CALCULATE ( MAX ( Indicators[Year] ), Indicators[Indicator] = "NY.GDP.MKTP.CD" )

VAR FirstYearGDP =

    CALCULATE ( [GDP], Indicators[Year] = FirstYear )

VAR LatestYearGDP =

    CALCULATE ( [GDP], Indicators[Year] = LatestYear )

VAR Diff =

    LatestYearGDP - FirstYearGDP

RETURN

DIVIDE ( Diff, FirstYearGDP, 0 )

Population Growth =

VAR FirstYear =

    CALCULATE ( MIN ( Indicators[Year] ), Indicators[Indicator] = "SP.POP.TOTL" )

VAR LatestYear =

    CALCULATE ( MAX ( Indicators[Year] ), Indicators[Indicator] = "SP.POP.TOTL" )

VAR FirstYearPop =

    CALCULATE ( [Population], Indicators[Year] = FirstYear )

VAR LatestYearPop =

    CALCULATE ( [Population], Indicators[Year] = LatestYear )

VAR Diff =

    LatestYearPop - FirstYearPop

RETURN

DIVIDE ( Diff, FirstYearPop, 0 )

Graduate

#### Your goals:

* Create a Power BI report analyzing population and GDP for US states since 1997. Clean and filter the data using Power Query, then create relationships between your tables. Use measures in DAX to calculate growth rates and GDP per capita similar to our approach in the Week 2 videos. Finally, visualize and explore the data in Power BI Desktop.

#### Techniques you’ll need to use:

* Power Query
  + Filtering
  + Replace values
  + Unpivoting
* Power BI Desktop
  + Map visual
  + Slicers and/or filters

#### Data sources:

You’ll find the following files in Week 2 Exercise Files.zip:

* Link: [US States Population by Year (Excel)](https://courses.edx.org/assets/courseware/v1/ac1ca3ce36416e8d17796cdebb3cc337/asset-v1:DavidsonX+DavidsonX.D005+1T2022a+type@asset+block/Week_2_Wrap-Up_-_US_States_Population_by_Year.xlsx)
  + Source: Federal Reserve Bank of St. Louis
* Link: [US States GDP by Year (Excel)](https://courses.edx.org/assets/courseware/v1/20fedee2f45ad9f7f26ca9f488f8eb4e/asset-v1:DavidsonX+DavidsonX.D005+1T2022a+type@asset+block/Week_2_Wrap-Up_-_US_States_GDP_by_Year.xlsx)
  + Source: U.S. Bureau of Economic Analysis
* Link: [US States (Excel .xls)](https://courses.edx.org/assets/courseware/v1/c746c3adaa817dc1f945e3f8277f4342/asset-v1:DavidsonX+DavidsonX.D005+1T2022a+type@asset+block/Week_2_Wrap-Up_-_US_States.csv)
* Link: [US States (Excel .xlsx)](https://courses.edx.org/asset-v1:DavidsonX+DavidsonX.D005+1T2021+type@asset+block@Week_2_Wrap-Up_-_US_States.xlsx)
  + Source: data.world (https://data.world/markmarkoh/us-state-table)

#### Tips:

* Notice how the state GDP figures in your data source are represented in the millions of dollars. You will need to account for this in your report by either naming your measures appropriately, or by multiplying the numeric values by 1,000,000.
  + For example, Alabama’s GDP shows 104087.4 millions of dollars for the year 1997. To get the correct GDP in dollars, multiply 104087.4 by 1,000,000, which will return the correct GDP amount of 104,087,400,000.

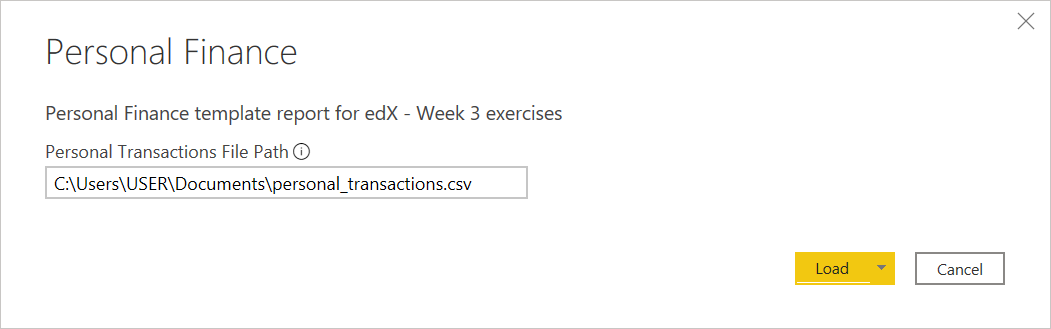
### Week 3: Exercise Files

Please download the necessary files below to get ready for Week 3.

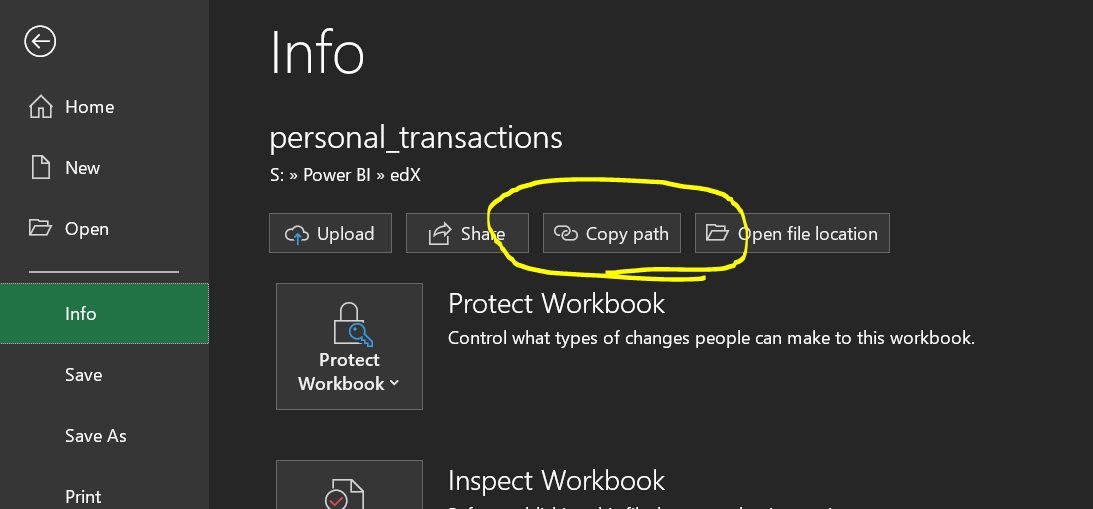
* Dataset: [personal\_transactions.csv](https://courses.edx.org/asset-v1:DavidsonX+DavidsonX.D005+1T2021+type@asset+block@personal_transactions.csv)
* Background image: [Week 3 Background (JPG)](https://courses.edx.org/assets/courseware/v1/6a0378580404c676d5a982b1f507fc66/asset-v1:DavidsonX+DavidsonX.D005+1T2022a+type@asset+block/Week_3_background.jpg)
* Power BI Template file: [Personal\_Finance.pbit](https://courses.edx.org/asset-v1:DavidsonX+DavidsonX.D005+1T2021+type@asset+block@Personal_Finance.pbit" \t "_blank)
* Budget data file: [Budget.csv](https://courses.edx.org/assets/courseware/v1/b804d8678df37d413ee00a0e7bdc2c11/asset-v1:DavidsonX+DavidsonX.D005+1T2022a+type@asset+block/Budget.csv)

#### To Use the Power BI Template file:

1. Open the PBIT (Power BI Template) file in Power BI Desktop.
2. When the file opens, you will be prompted to enter the file path for the personal\_transactions.csv file. Copy and paste the full location into the box for "Personal Transactions File Path" (see screenshot below). For example, your file path may be **C:\Users\{your username}\Downloads\personal\_transactions.csv**.
3. Click "Load" to load the data into your data model.
4. When you're done building the report, you can save the file as a normal PBIX file.



#### Need help finding your CSV file path?

1. Open personal\_transactions.csv in Microsoft Excel.
2. Go to File >> Info, and click the button at the top of the screen which says "Copy path." This will copy the file's complete path to your computer's clipboard, allowing you to paste it directly into Power BI Desktop.
3. 

#### Your goals:

* Create a Power BI report that visualizes city-wide budgets and actuals. Find out which departments are generating the most revenue or spending the most money, and which departments have gone furthest over their budgets.

#### Techniques you’ll need to use:

* Duplicate queries and merge queries
* Hierarchies & Drill down
* Drillthrough (BONUS)

#### Data sources:

* Link: [Week 3 Wrap-Up File (Excel)](https://courses.edx.org/assets/courseware/v1/74245fda4966c94008a57b5e9c226217/asset-v1:DavidsonX+DavidsonX.D005+1T2022a+type@asset+block/Week_3_Wrap-Up_File.xlsx): contains budget amounts and actuals for the city of Houston, TX.
  + Source: data.world (https://data.world/houston).

#### Tips:

* Note how the transactions are divided between Expenditures and Revenues using the 14th column. Use this column to write measures in DAX that distinguish between revenue and expenditures (similar to how we differentiated between income and expenses in the Week 3 exercises).
* Use the hierarchies functionality to create and analyze hierarchies in this data:
  + Hierarchy #1
    - Business Area (parent)
    - Fund Center (child)
  + Hierarchy #2
    - Category (parent)
    - Account (child)
* Use the drill up/drill down functionality to navigate through the different hierarchies.

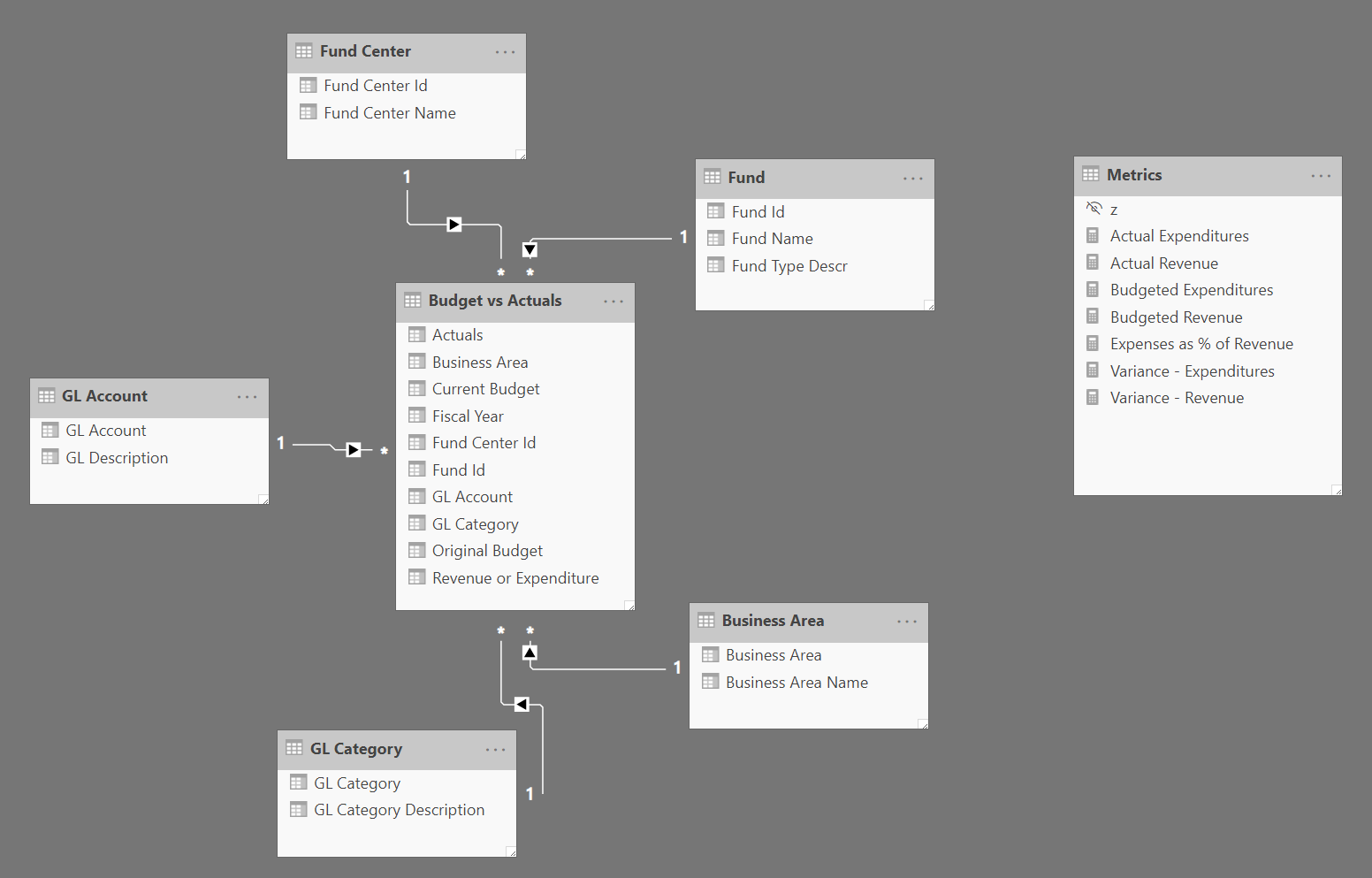
#### Solutions

This dataset contains accounting data. Each transaction belongs to a fund, business area, fund center, account, and category. These attributes are how data is typically represented inside of the operating ledger of a business or government agency.

There are two general approaches for completing this week's exercise:

In the first approach, you can transform each of these attributes into its own table using the Duplicate Queries functionality in Power Query, remove all columns except the ones pertaining to each attribute, then remove duplicate values. When removing duplicates, you want to find the column that has the most unique values, then right-click and select "Remove Duplicates." You would then use the Merge Queries functionality in your original table to merge each attribute's ID column back into the main table. This approach helps to normalize the data, cutting down on duplicates and allowing for faster and more efficient filtering of the data within Power BI.

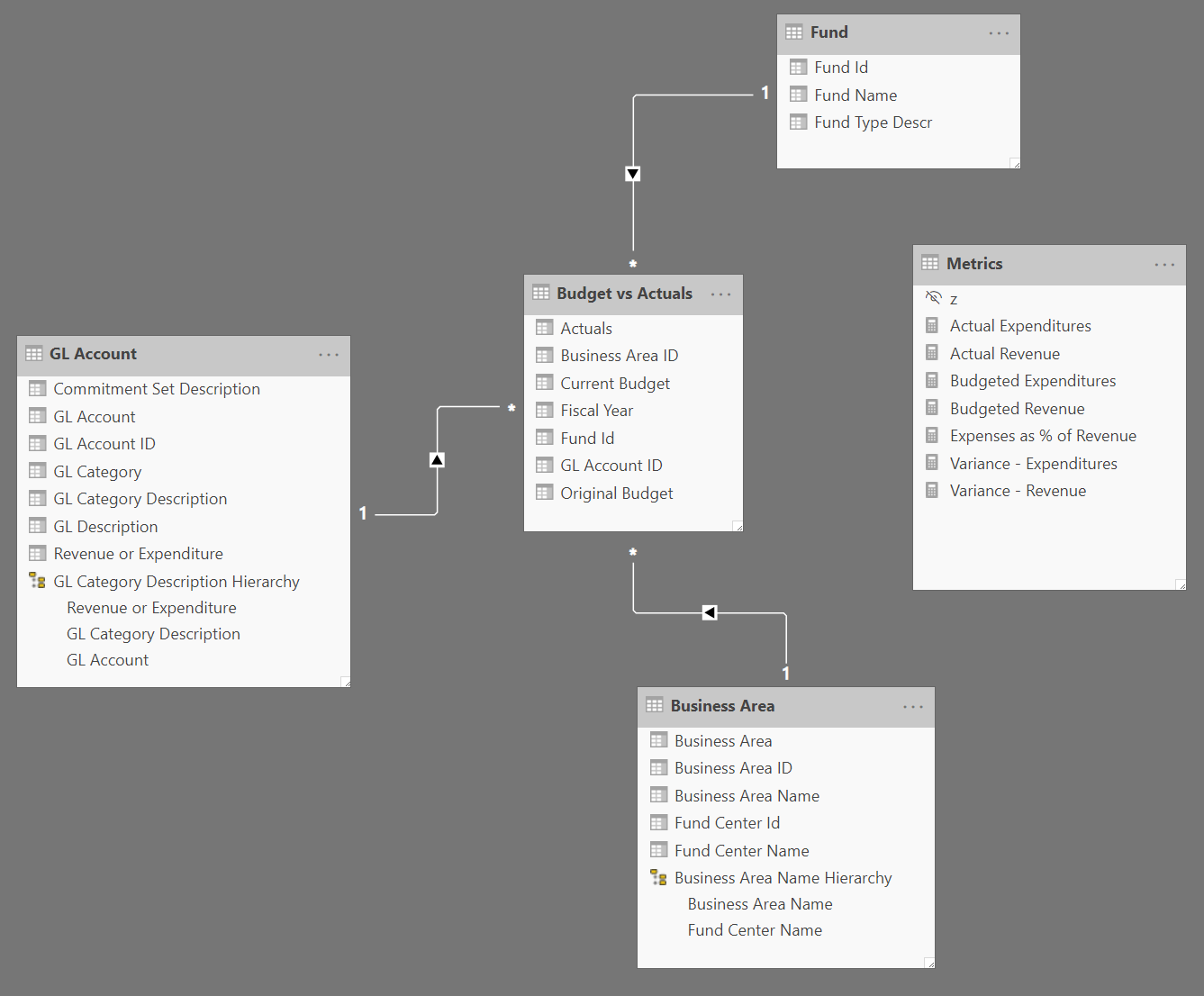
[You can download a Power BI template showing the first approach here](https://courses.edx.org/assets/courseware/v1/066d72ae8e88404fead807e304e718e0/asset-v1:DavidsonX+DavidsonX.D005+1T2022a+type@asset+block/Week_3_Exercise_-_Approach_1.pbit). The screenshot below shows the proposed data model.



The second approach is my preferred approach. In the second approach, instead of transforming each attribute into a separate table, you will instead group the attributes that go together. For example, Business Area and Fund Center would go into one table together, and the GL Account and GL Category columns would go into a different table together. This is because these columns all naturally relate to each other, which you can determine by a close analysis of the data. In this approach, you will still use Duplicate Queries to break the data out into different tables, and you'll use Remove Duplicates and Merge Queries to bring everything back together in the main table. The main difference is that you will have fewer tables than before, which means fewer table relationships, resulting in a more efficient data model. Consider using Index Columns to create a primary key on each table, and bring those into the main table ("Budget vs Actuals") to create the table relationships.

Another benefit of this approach is you can create explicit hierarchies. For example, right-click on Business Area Name and select "Create Hierarchy" to create the new hierarchy, then drag "Fund Center Name" on top of the new hierarchy. Then you can use the hierarchy in your charts and tables. [Here's a video from Patrick LeBlanc at "Guy in a Cube" on how to create hierarchies.](https://guyinacube.com/2019/08/07/how-to-create-and-use-a-power-bi-hierarchy/)

[You can download a Power BI template showing the second approach here](https://courses.edx.org/assets/courseware/v1/ff0a6d8bd59ecc98e7123d9dcdf8936a/asset-v1:DavidsonX+DavidsonX.D005+1T2022a+type@asset+block/Week_3_Exercise_-_Approach_2.pbit). The screenshot below shows the proposed data model.



* BONUS:
  + Read how to set up “drill through” in Power BI reports here: https://docs.microsoft.com/en-us/power-bi/desktop-drillthrough
  + Create a report that allows you to drill through from an overview or summary page into a detail page that for a particular Business Area or Fund Center.

#### Your goals:

* Create a Power BI report showing the distribution of coronavirus cases by geography and over time.

#### Techniques you’ll need to use:

* Merge queries
* Unpivoting
* Dates and date table
* Geographic data (latitude/longitude or location)

#### Data Sources:

All three data sources are CSV files that can be downloaded from Johns Hopkins University’s public GitHub account using the links below. To import each of these into your data model, click “Get Data,” find the “Web” connector, and copy and paste each of the links below.

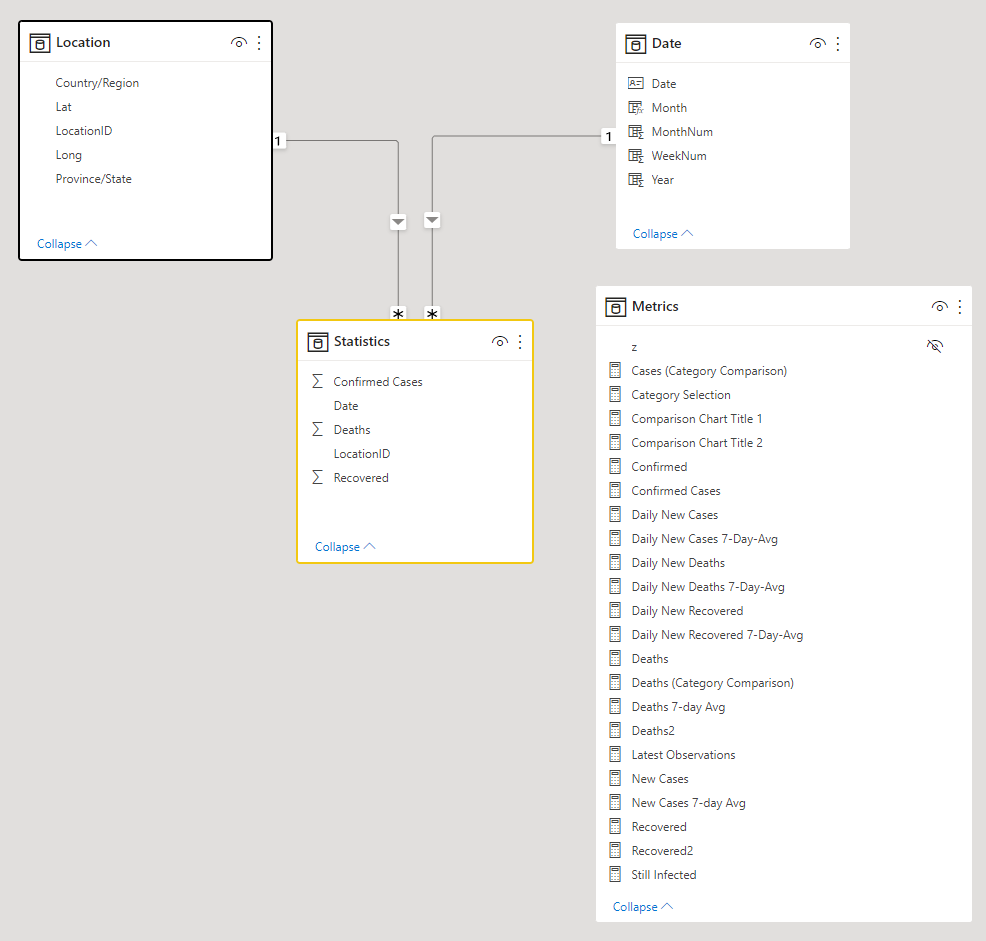
* Data source #1: Confirmed cases
  + Link: <https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv>
* Data source #2: Deaths
  + Link: <https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_deaths_global.csv>
* Data source #2: Recovered
  + Link: <https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_recovered_global.csv>

#### Tips:

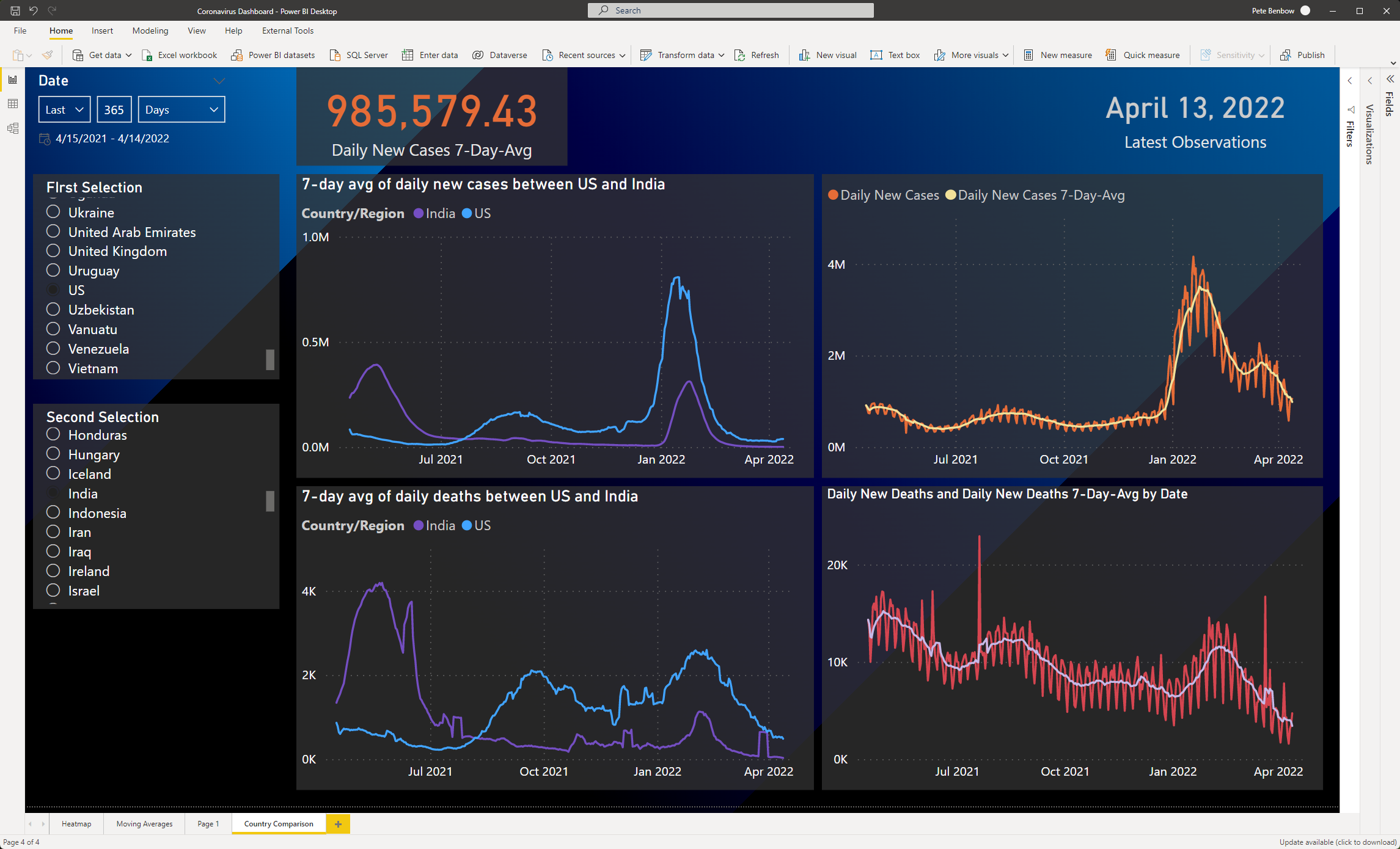
* The geographic data is nearly identical across all three data sources, so try creating a Location table in Power Query by duplicating one of your queries. Then remove everything except the location data and create an Index column named LocationID.
* Use the Merge Queries transformation to get all confirmed cases, deaths, and recoveries into a single table, then connect it to your Location table and your Date table. This will keep your model neat and tidy.
* See if you can use DAX to calculate the growth in cases over time, or use time-intelligence functions to compare week-over-week or month-over-month numbers.
* The numerical data for confirmed cases, deaths, and recoveries is already in a cumulative format, similar to what we created when we explored time-intelligence functions. For your measures, you’ll need to write a DAX formula that calculates the most recent date, and then filters your measure based on that most recent date. Try using a DAX variable to store the most recent date, then reference that variable in your calculation, like this:
* Explore the theme customization options this time, and try to find a style that suits you that is different from the default theme.

|  |
| --- |
| Confirmed =  var LatestDate = MAX('Statistics'[Date])  return  CALCULATE(SUM('Statistics'[Confirmed Cases]), 'Statistics'[Date] = LatestDate) |

Example data model:



Example report:



### Calculating daily measurements

One of the tricks with this dataset is the fact that the case counts are additive. As the time series moves forward, the numbers will always go higher. So, in order to figure out how many new COVID-19 cases were diagnosed on a particular date, you will need to compare that date's case count with the previous day's case count, subtracting one from the other to arrive at the number of cases that were added from the previous day to the next.

In the measure below, we will use the current\_day variable to get each day's total of [Confirmed Cases]. Then we'll use the prev\_day variable to get the [Confirmed Cases] from the previous day. Finally, to get our result, we'll subtract the two variables.

Daily New Cases =

VAR current\_day = [Confirmed Cases]

VAR prev\_day =

CALCULATE (

[Confirmed Cases],

DATEADD ( 'Date'[Date], -1, DAY )

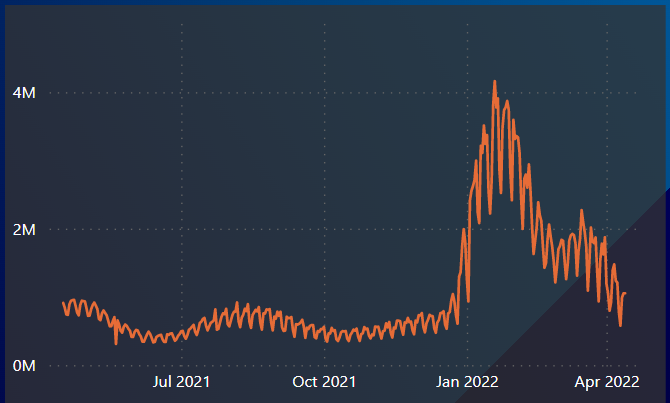
)

RETURN

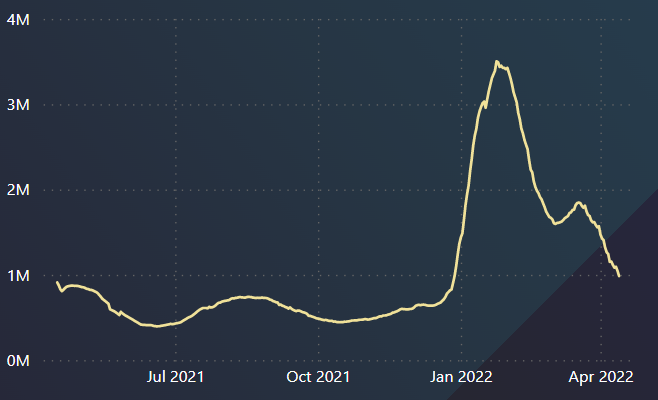
IF ( ISBLANK ( prev\_day ), BLANK(), current\_day - prev\_day )

### Calculating moving averages

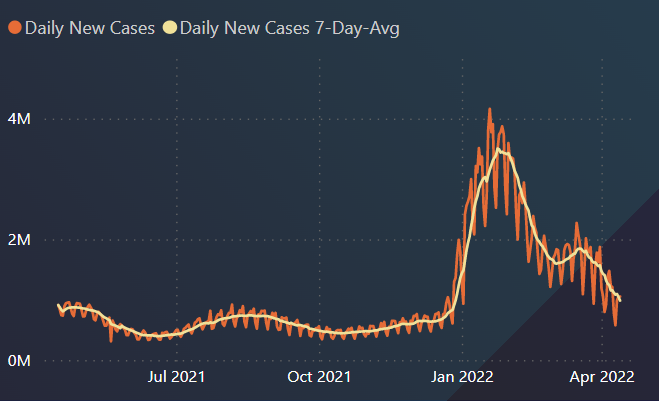
This is a time series dataset, and one thing you'll probably notice when you chart the Confirmed cases and Deaths is that there is a lot of daily volatility in the data. For instance, check out the screenshot below where you can see large periodic changes in the daily confirmed cases:



One mathematical technique we have for controlling this kind of fluctuation is through a simple moving average. The moving average can be used to smooth this volatility in time series data by averaging the observations over a given period of time. For instance, a 7-day moving average will average the observations over the last 7 days, which gives us a chart that looks like this:



Combining these two measurements together in Power BI helps you see both the original volatile data alongside the smoothed moving average:



Calculating a moving average is relatively simple with DAX. See the example below. We will take each date in our date table using the LatestDate variable. Then, we'll use the DATESBETWEEN function to get that date plus the six days prior to it, and we will calculate the average of the [Daily New Cases] measure using the AVERAGEX function.

Daily New Cases 7-Day-Avg =

VAR LatestDate = MAX ( 'Date'[Date] )

RETURN

AVERAGEX (

DATESBETWEEN ( 'Date'[Date], LatestDate - 6, LatestDate ),

[Daily New Cases]

)