Information Technology Fundamentals

Mohammad Hossein Manshaei
manshaei@gmail.com

Database: Business Intelligence and Analytics
Module 6: Part 1

MODULE 6: DATABASE

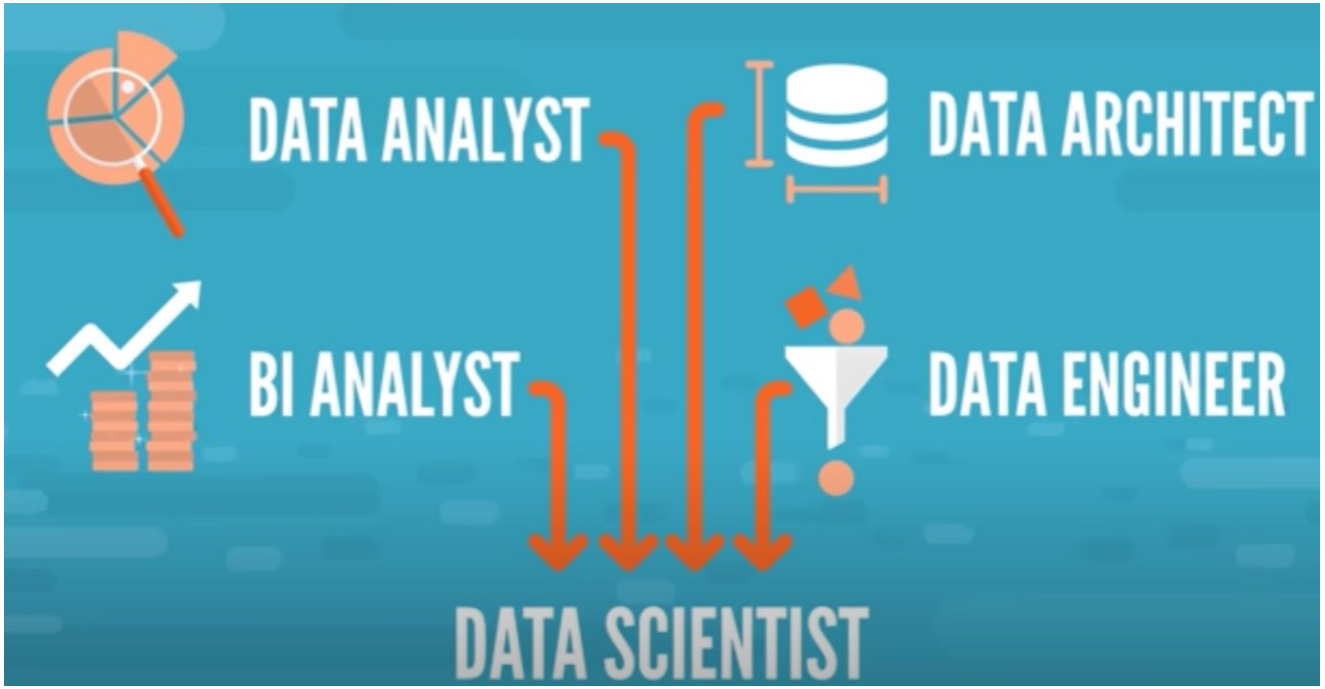
Data Generation: Big Data

- At the beginning of **2020**, the digital universe was estimated to consist of **44 zettabytes** of data.
- By **2025**, approximately **463 exabytes** would be created **every 24 hours** worldwide.
- In **2019**, Google processed **3.7 million queries**, **Facebook** saw **one million logins**, and **YouTube recorded 4.5 million videos** viewed every **60 seconds**.
- **Netflix's** content volume in **2019 outnumbered** that of the US TV industry in 2005.
- By **2025**, there would be **75 billion Internet-of-Things (IoT)** devices in the world.
- By **2030**, nine in every ten people aged six and above would be digitally active.
- **As of 2013, experts believed that 90% of the world's data was generated from 2011 to 2012.**

Why Data is Important?

1. Improve People's Lives
2. Make Informed Decisions
3. Get The Results You Want
4. Find Solutions To Problems
5. Stop The Guessing Game
6. Be Strategic In Your Approaches
7. Know What You Are Doing Well
8. Keep Track Of It All
9. Make The Most Of Your Money
10. Access The Resources Around You

How do We Work with Data?



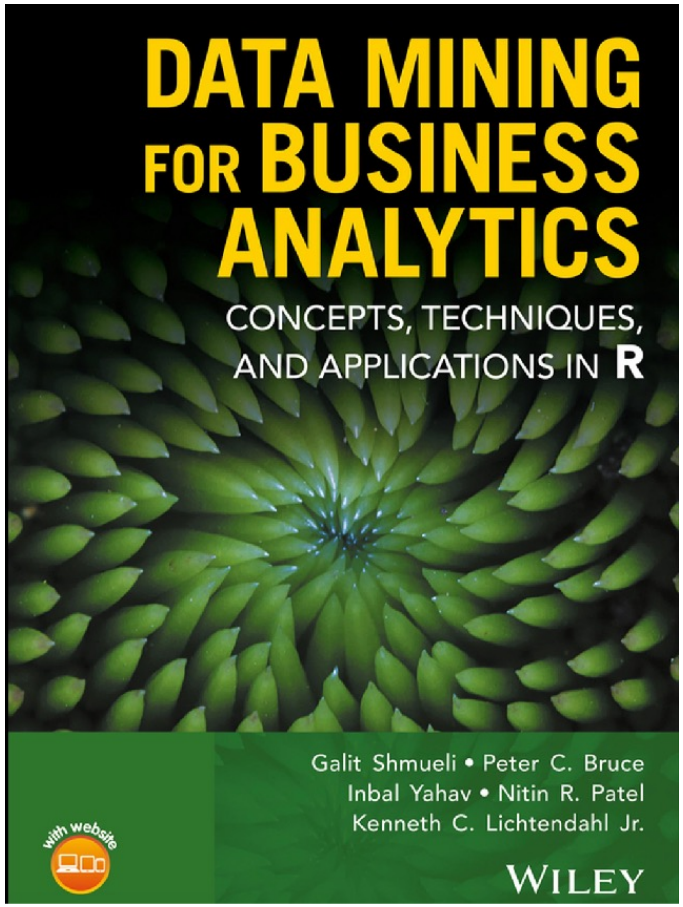
Module 6. Main Objectives

1. **Review Business Intelligence Concepts**
2. Explain Data Warehouse
3. Describe Data Mining

Introduction to Business Intelligence

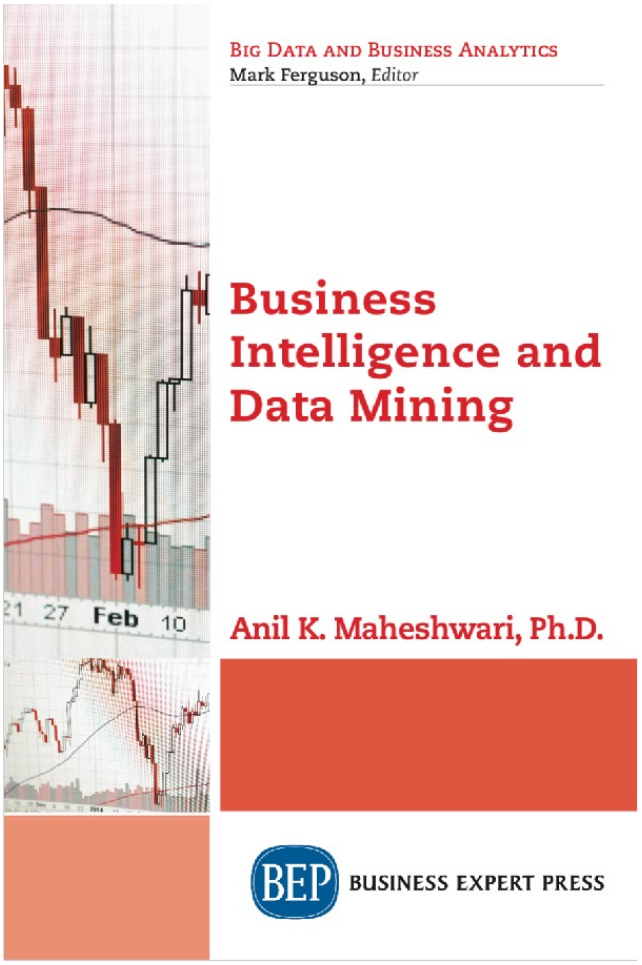
Main Reference

- Shmueli, G., Bruce, P.C., Yahav, I., Patel, N.R. and Lichtendahl Jr, K.C., 2017. **Data Mining for Business Analytics: Concepts, Techniques, and Applications in R**. John Wiley & Sons.



Main Reference

- Maheshwari, A., 2014. **Business Intelligence and Data Mining**. Business Expert Press.



Main Reference

- C. Howson, **Successful Business Intelligence: Unlock the Value of BI & Big Data,** Mc Graw Hill, 2014.

Main Reference

- Mishra, B.K., Hazra, D., Tarannum, K. and Kumar, M., 2016, November. **Business Intelligence using Data Mining Techniques and Business Analytics**. In *2016 International Conference System Modeling & Advancement in Research Trends (SMART)* (pp. 84-89). IEEE.

Proceedings of the SMART-2016, IEEE Conference ID: 39669
5th International Conference on System Modeling & Advancement in Research Trends, 25th-27th November, 2016
College of Computing Sciences & Information Technology, Teerthanker Mahaveer University, Moradabad, India

Business Intelligence using Data Mining Techniques and Business Analytics

Brojo Kishore Mishra[1], Deepannita Hazra[2], Kahkashan Tarannum[3] and Manas Kumar[4]
[1,2,3,4]Department of IT, C.V. Raman College of Engineering, Bhubaneswer, India
E-mail: [1]bkmishra@cvrgi.edu.in, [2]hazra.deepannita@gmail.com, [3]tara.shifa@gmail.com, [4]manasbit2712@yahoo.com

***Abstract*—The objective of this paper is to present a review literature on what are impacts of Data Mining (DM) in Business Intelligence (BI). The paper highlights various features of DM. It involves three steps: explorations, pattern identification and deployment. BI is the hot topic among all industries aiming for relevance. BI emphasizes on detail integration and or organizing of data. DM and BI work together to process and analyse data to lighten workload for the user and organization and hence in understanding discovered materials. It also explains Business Analytics (BA) as a part of BI which is again dependent on BI. There are various sectors in business to which BA has proved to be a powerful tool to obtain effective results.**

Keywords: Business Intelligence, Data Mining, Business Analytics

I. Introduction

The task of studying through and finding out certain patterns in business data is not new. The author of the paper [10] spreads the awareness that the business community of today is suffering information overload and business source analysis shows that:

- 61% of managers believe that information overload at their own workplace.
- 80% believe the situation will get worse.
- Over 50% of managers ignore data in current decision-making process because of the information overload.
- 84% of managers store this information for the future use, it is not used for current analysis.
- 60% believe that the cost of gathering information outweighs its value.

Since a long time it is done with the help of statistical techniques. But now, only to make the task easier improved techniques like "Data Mining" is used. Data mining is the process of "knowledge discovery" in database which can be used in decision making. It is a fast expanding and dynamic field that uses artificial intelligence, machine learning, database systems and statistics to apply the advanced techniques of data analysis [1]. In this [5], the research stated that the process which is designed and used for the purpose of exploring data is called as data mining. This process is very much similar to the real life process of mining out nuggets of gold from the Earth. More specifically it is like taking out non-trivial nuggets from the huge volumes of available data. This paper gives a view about how data mining assists business intelligence to find out patterns and gain knowledge from existing data.

In the research [2], it was explained that it is because of the intense competition that the companies are compelled to find out innovative ideas in which they can capture and enhance their market shares while reducing their costs too. Implementation of the data analysis techniques can help the companies to find such solutions like finding out some unexpected patterns from the large volumes of the data present in the database or data warehouse. These patterns can provide information which could help in predicting future outcomes [1].

A. Why do we Need to Manage Data?

- For faster decision making: Almost 77% of the executives complain of not having real time information so that they can take decisions. Data needs to be managed and kept in an organized way so that it would be easy and quick to be referred to when taking the decisions.
- Limited insight due to large volumes of information: About 6 out of 10 respondents agree to the statistics that almost all of the organizations have more volumes of data than they can handle and use effectively. Since the organizations cannot handle so much of data, their working procedures and insight gets restricted and ultimately they function inefficiently.
- New emerging varieties: Emails, audio, video, documents, and images are responsible for generating 80% of the new data. Due to this newly generated data there emerges another problem of the storage. Data should be stored in such a way that they can be identified and segregated ant any point of time.

Contents

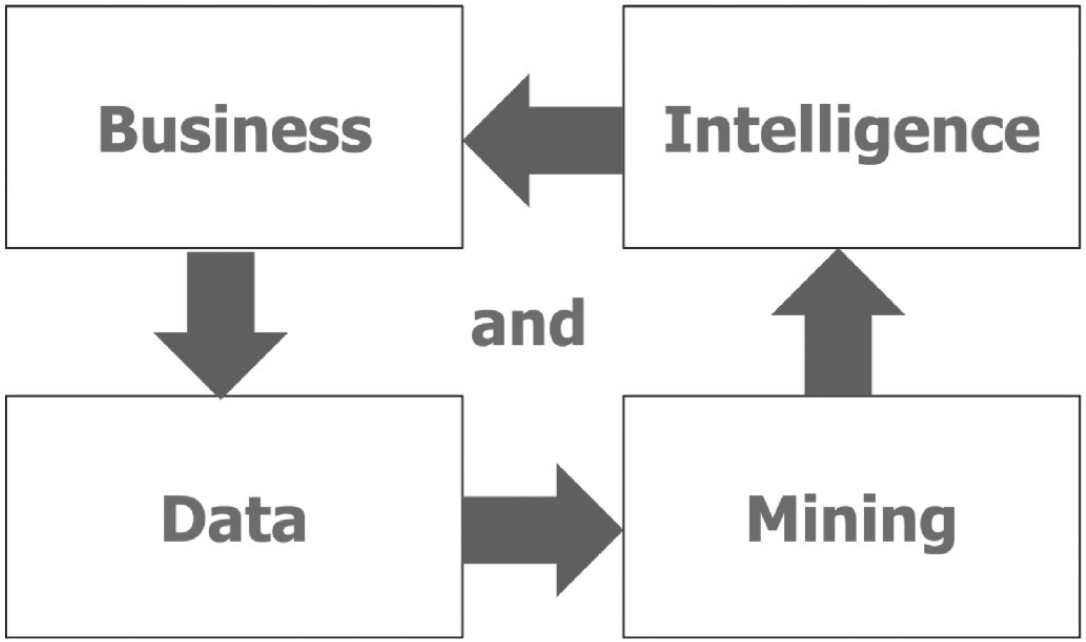
- Business Intelligence and Analytics: Definition and Benefits
- Case Studies: Applications, Challenges, and Tools
- BI Future Trends
- Data and Business Intelligence

Business Intelligence

Business intelligence (BI) is a broad set of information technology (IT) solutions that includes **tools** for **gathering**, **analyzing**, and **reporting information** to the users about **performance of the organization** and **its environment**.

Business Intelligence

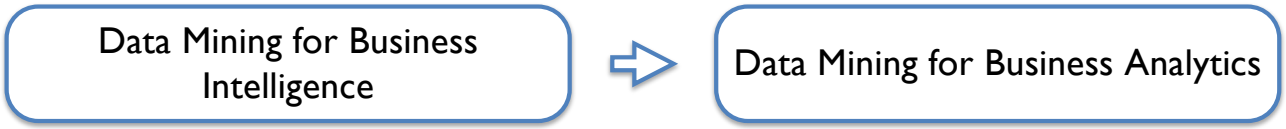
- BI comprises the strategies and technologies used by enterprises for the **data analysis** of business information. BI technologies provide **historical**, **current**, and **predictive** views of business operations. *[Wikipedia]*
- BI refers to data visualization and reporting for understanding **"what happened and what is happening."**



Business intelligence and data mining cycle

What Is Business Analytics?

- **Business Analytics (BA)** is the practice and art of bringing quantitative data to bear on decision-making. The term means different things to different organizations.
- BA or **Analytics**, include a range of data analysis methods. Many powerful applications involve little more than counting, rule-checking, and basic arithmetic. For some organizations, this is what is meant by analytics.



How Business Intelligence Provides Value

- Touches everyone in a company and beyond to customers, suppliers, and with public data, to citizens
- Can only provide value when it is used effectively by people
- There is a correlation between the effective use of business intelligence and company performance
- Having better access to data does not improve performance; the difference is in what companies do with the data

BI for Management and Control

- Without business intelligence, managers may talk about how they are **"flying blind"** with no insight until quarterly financial numbers are published.
- With BI, information is accessible on a timelier and more flexible basis to provide a view of:
  - Sales in various regions and by various product lines
  - Expenses compared to budget
  - Warehouse inventory for a particular product or raw materials
  - Sales pipeline versus forecast

BI for Improving Performance

- Business intelligence allows organizations to improve **performance**
- **Business performance** is measured by a number of financial indicators, such as revenue, margin, profitability, cost to serve, and so on

Contents

- Business Intelligence and Analytics: Definition and Benefits
- Case Studies: Applications, Challenges, and Tools
- BI Future Trends
- Data and Business Intelligence

Operational BI

- Early BI deployments focused more on strategic decisions and performance, BI increasingly plays a critical role in the daily operations
- Examples of operational BI include the following:

1. **Travel agents and airlines** use operational BI to monitor flight delays so they can proactively reaccommodate passengers with connections.
2. **Hospitals and emergency rooms** use business intelligence to determine optimum staffing levels during peak periods.
3. **Restaurants** use BI to estimate the wait time for a table based on the number of current patrons and average length to dine.
4. **Walt Disney World's Magic Kingdom** uses business intelligence for its service that issues park visitors FastPass tickets to avoid standing in long lines for rides.The business intelligence tools monitor waiting

BI to Improve Customer Service

- BI can help companies deliver high customer service levels by providing timely order processing, loan approvals, problem handling, and so on. For example:

1. **Whirlpool** uses business intelligence to monitor its warranty program to understand root causes for warranty problems and improve customer satisfaction with its products.
2. **United Airlines** uses business intelligence to monitor how full business-class cabins are and to ensure its most valued customers receive complimentary upgrades when space permits.
3. **FlightStats** provides real-time travel information on delays so that if a passenger is en route and might miss a connecting flight, the travel agent can automatically rebook them.
4. **Netflix** tracks how often a customer gets their first-choice DVD.

Challenges of the Business Community

- The awareness that the business community of today is suffering information overload and business source analysis shows that:
  - 61% of managers believe that information overload at their own workplace
  - 80% believe the situation will get worse
  - Over 50% of managers ignore data in current decision-making process because of the information overload
  - 84% of managers store this information for the future use; it is not used for current analysis
  - 60% believe that the cost of gathering information outweighs its value

Business Intelligence Tools

- BI includes a variety of software tools and techniques to provide the managers with the information and insights needed to run the business.
- BI tools include

1. Data warehousing
2. Online analytical processing
3. Social media analytics
4. Reporting
5. Dashboards
6. Querying
7. Data mining

Contents

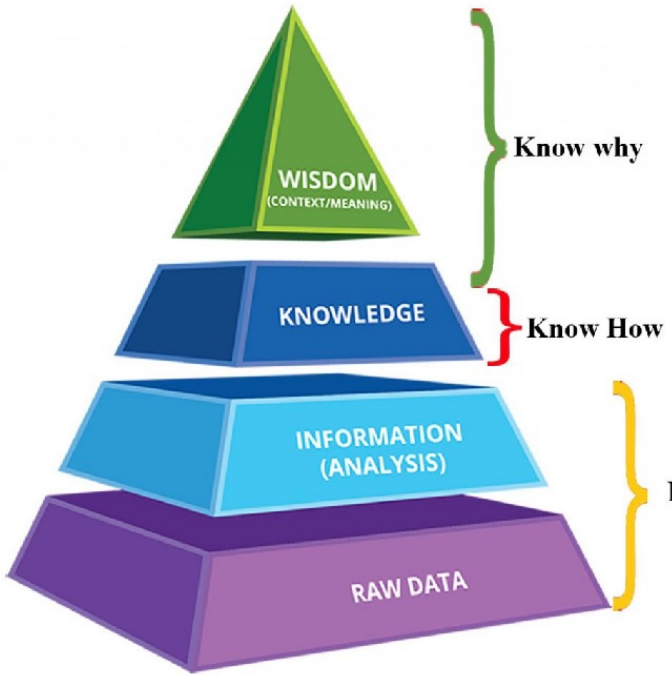
- Business Intelligence and Analytics: Definition and Benefits
- Case Studies: Applications, Challenges, and Tools
- BI Future Trends
- Data and Business Intelligence

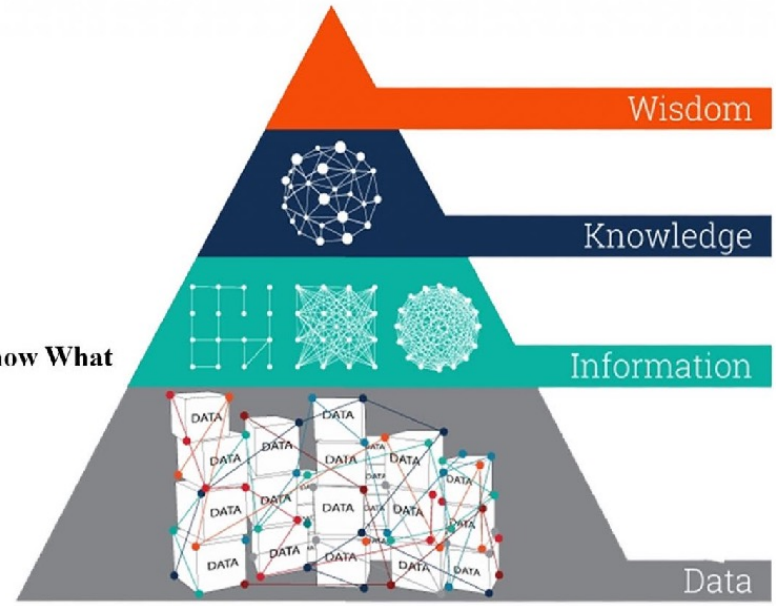
Trends and Predictions for the Future of Business Intelligence

- Data Governance
- Self-Service BI
- Prescriptive Analytics
- BI Through Natural Language Processing (NLP)
- Business-Intelligence-as-a-Service

Contents

- Business Intelligence and Analytics: Definition and Benefits
- Case Studies: Applications, Challenges, and Tools
- BI Future Trends
- Data and Business Intelligence





What is the DIKW Pyramid?

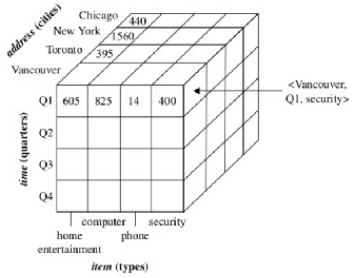
Reasons Why Data Management Leads To Business Success

1. Immediate Availability of Information
2. Protection of business interests
3. Keeping Up With Changing Times And Technological Advancements
4. Information is Always Valuable
5. Enhances Efficiency by Saving Costs and Time

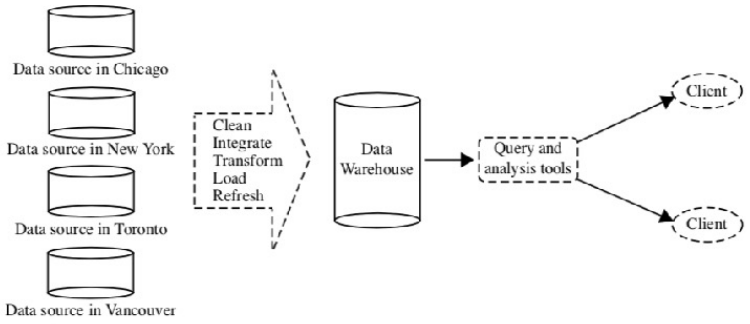
What Kinds of Data Can Be Mined?

1. Database Data
2. Data Warehouses
3. Transactional Data
4. Other Kinds of Data

customer	(*cust_ID, name, address, age, occupation, annual_income, credit_information, category,* ...)
item	(*item_ID, brand, category, type, price, place_made, supplier, cost,* ...)
employee	(*empl_ID, name, category, group, salary, commission,* ...)
branch	(*branch_ID, name, address,* ...)
purchases	(*trans_ID, cust_ID, empl_ID, date, time, method_paid, amount*)
items_sold	(*trans_ID, item_ID, qty*)
works_at	(*empl_ID, branch_ID*)



trans_ID	*list_of_item_IDs*
T100	I1, I3, I8, I16
T200	I2, I8
...	...

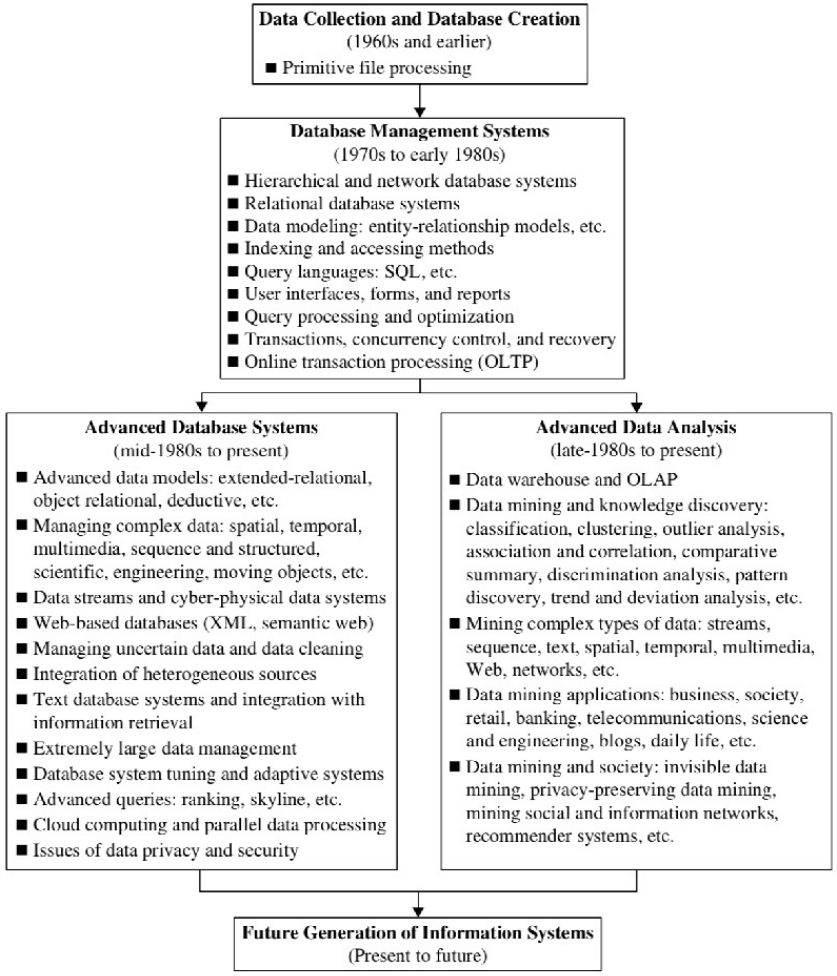


Data Processing Chain

1. Anything that is recorded is **data**. Facts and observations are data.
2. A **database** is a modeled collection of data that is accessible in many ways.
3. A **data warehouse** is an organized store of data from all over the organization, specially designed to help make management decisions.
4. **Data Mining** is the art and science of discovering useful innovative patterns from data.
5. **Data visualization**: As data and insights grow in number, a new requirement is the ability of the executives and decision makers to absorb this information in real time.

Big Data

- The challenge Big Data present is often characterized by the four V's-volume, velocity, variety, and veracity.
  - **Volume** refers to the amount of data.
  - **Velocity** refers to the flow rate- the speed at which it is being generated and changed.
  - **Variety** refers to the different types of data being generated (currency, dates, numbers, texts, etc.).
  - **Veracity** refers to the fact that data is being generated by organic distributed process (e.g. millions of people signing up for a or free downloads) and not subject to the controls or quality checks that apply to data collected for a study.



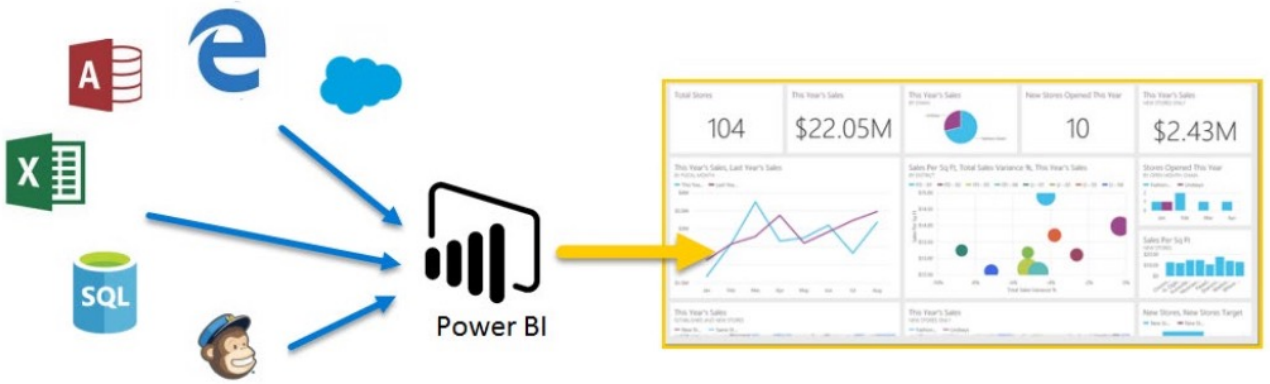
The Evolution of Database System Technology

Power BI

- **Microsoft Power BI**
- **Interactive data visualization software**
- **Primary focus on business intellige**

Power BI

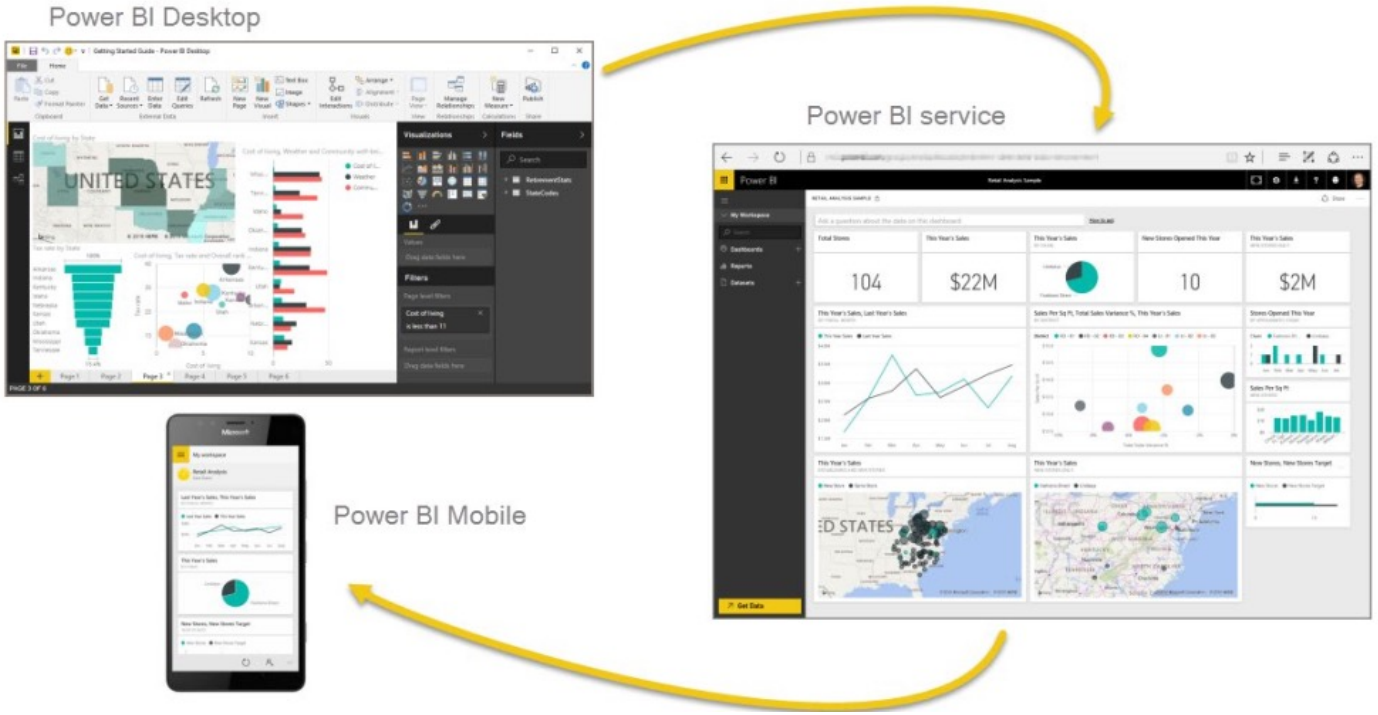
- Power BI is a suite of business analytics tools which connects to different data sources to analyze data and share insights throughout your organization.



Parts of Power BI

- Power BI Desktop: It is a Windows desktop application (Report Authoring Tool) which Lets you build queries, models and reports that visualize data.
- Power BI Service: Power BI Service is cloud based Software as Service Application which allows us to create dashboards, Setup schedule data refreshes, Share the reports securely in the organization.
- Power BI Mobile: It is an application (App) on mobile devices which allows you to interact with the reports and dashboard from Power BI Service.

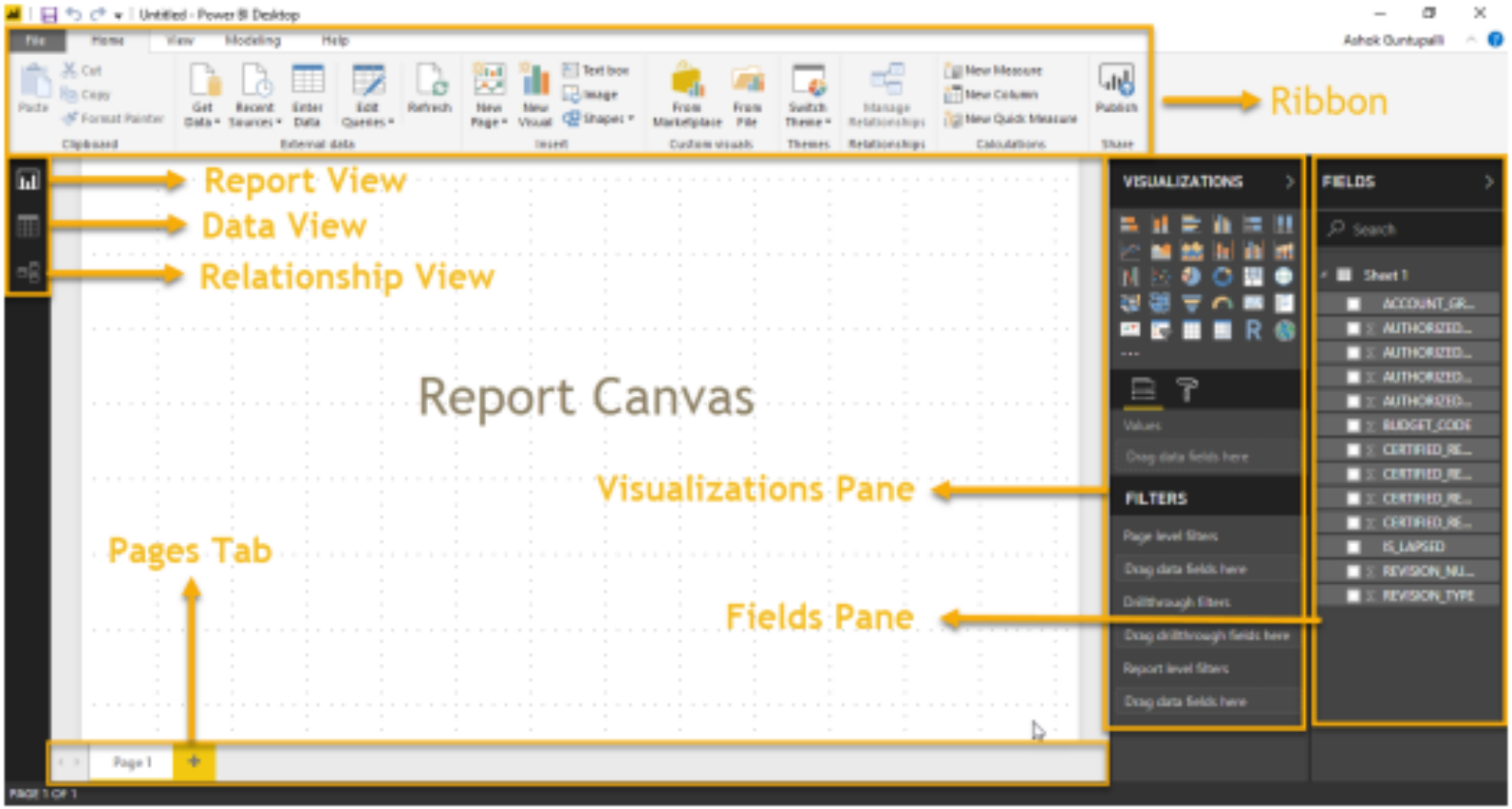
The flow of work in Power BI



Power BI Desktop

- Power BI Desktop is report authoring tool that allows you to create reports, queries, Extract Transform and Load the data from data sources and model the queries.

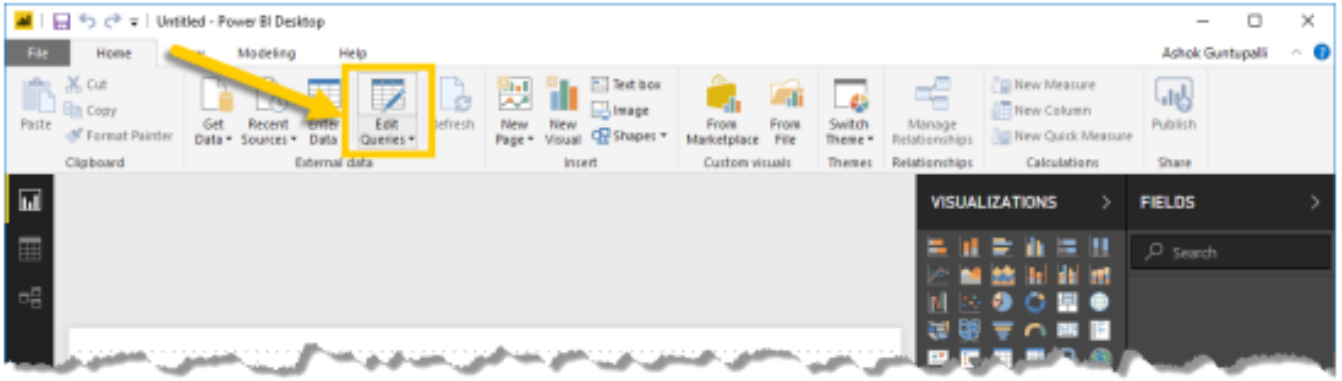
Power BI Desktop Interface



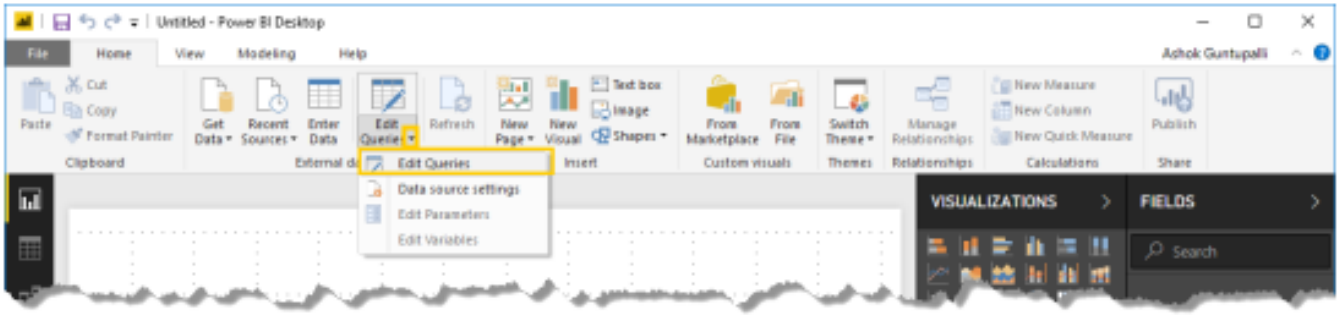
Querying Data from CSV

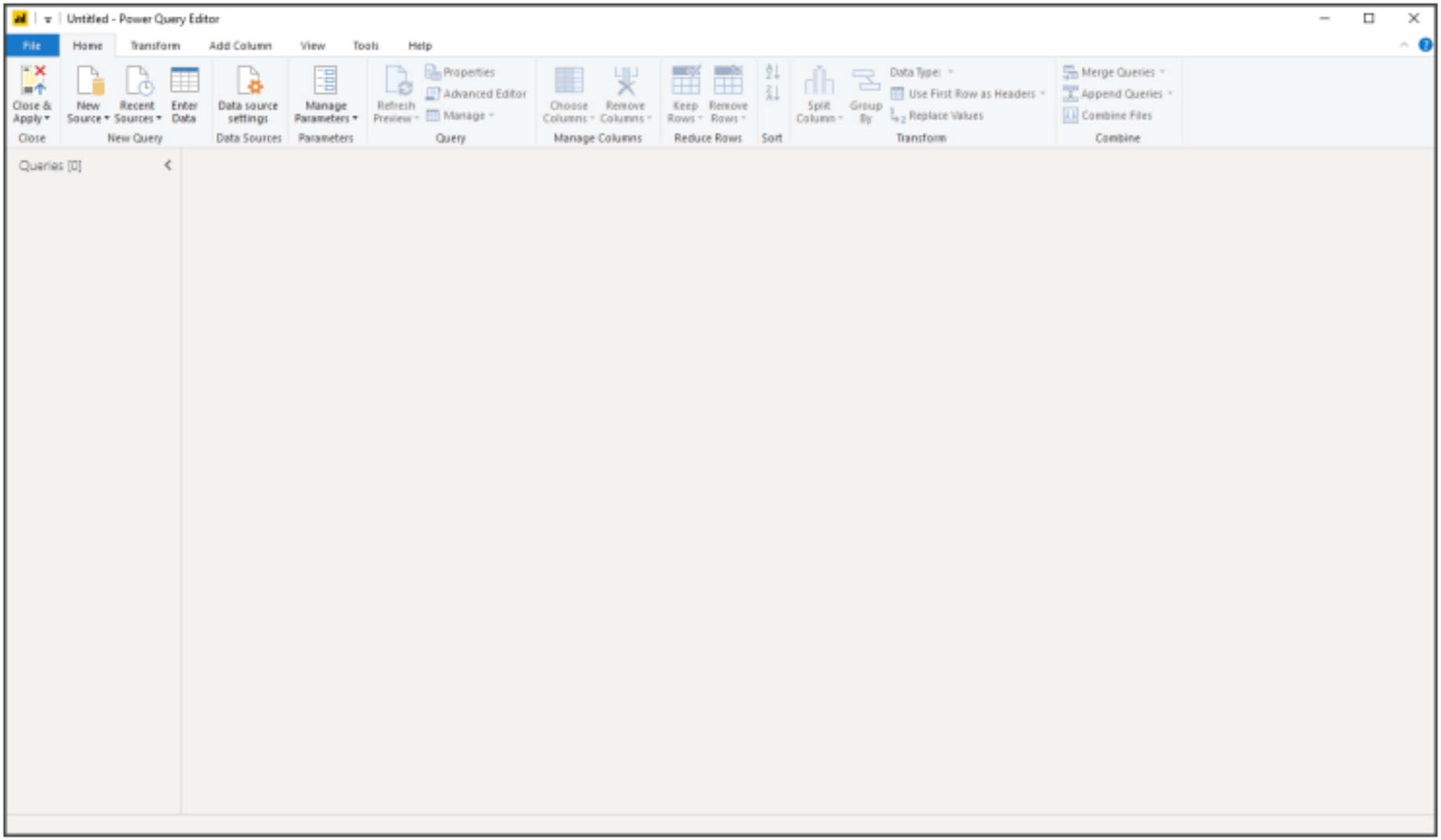
Exercise 1: Get Started with Query Editor

1. To get to Query Editor, select Edit Queries from the Home tab of Power BI Desktop.

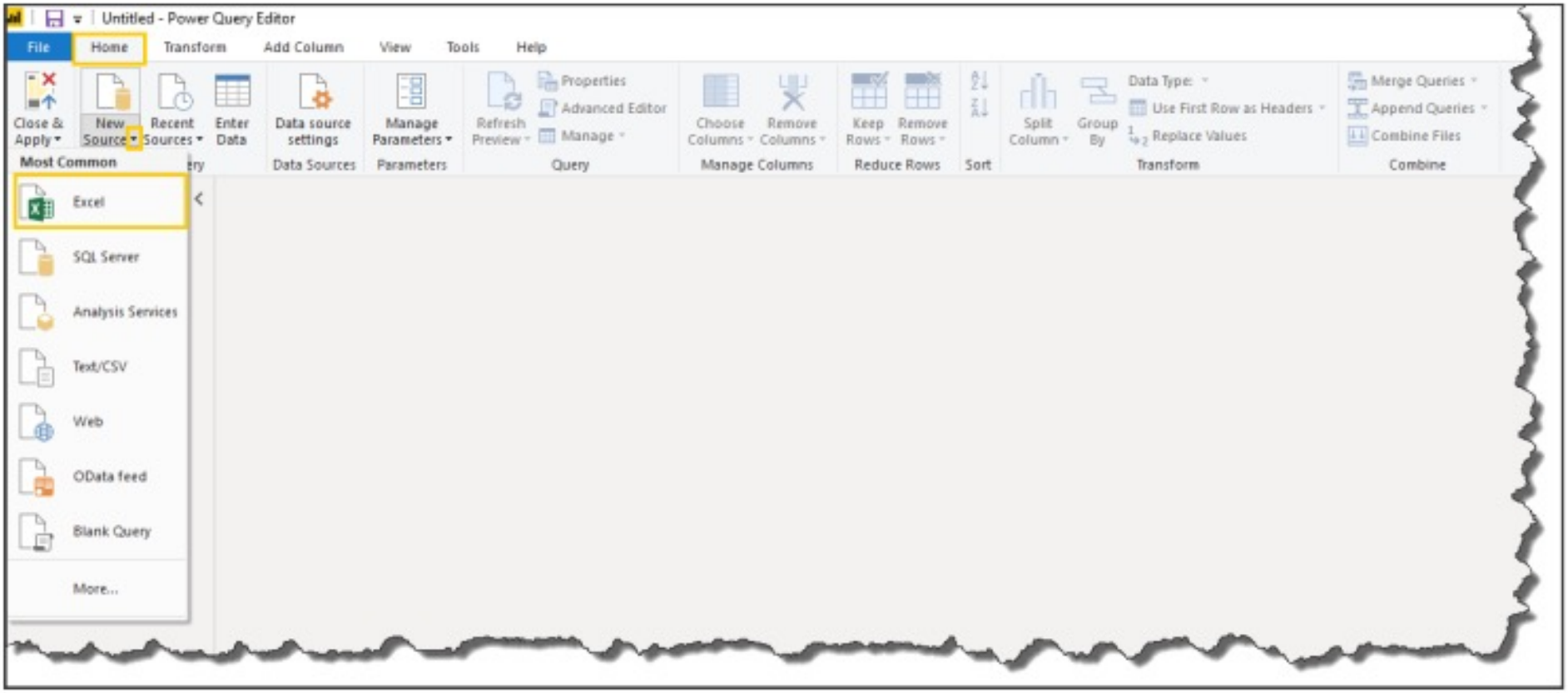


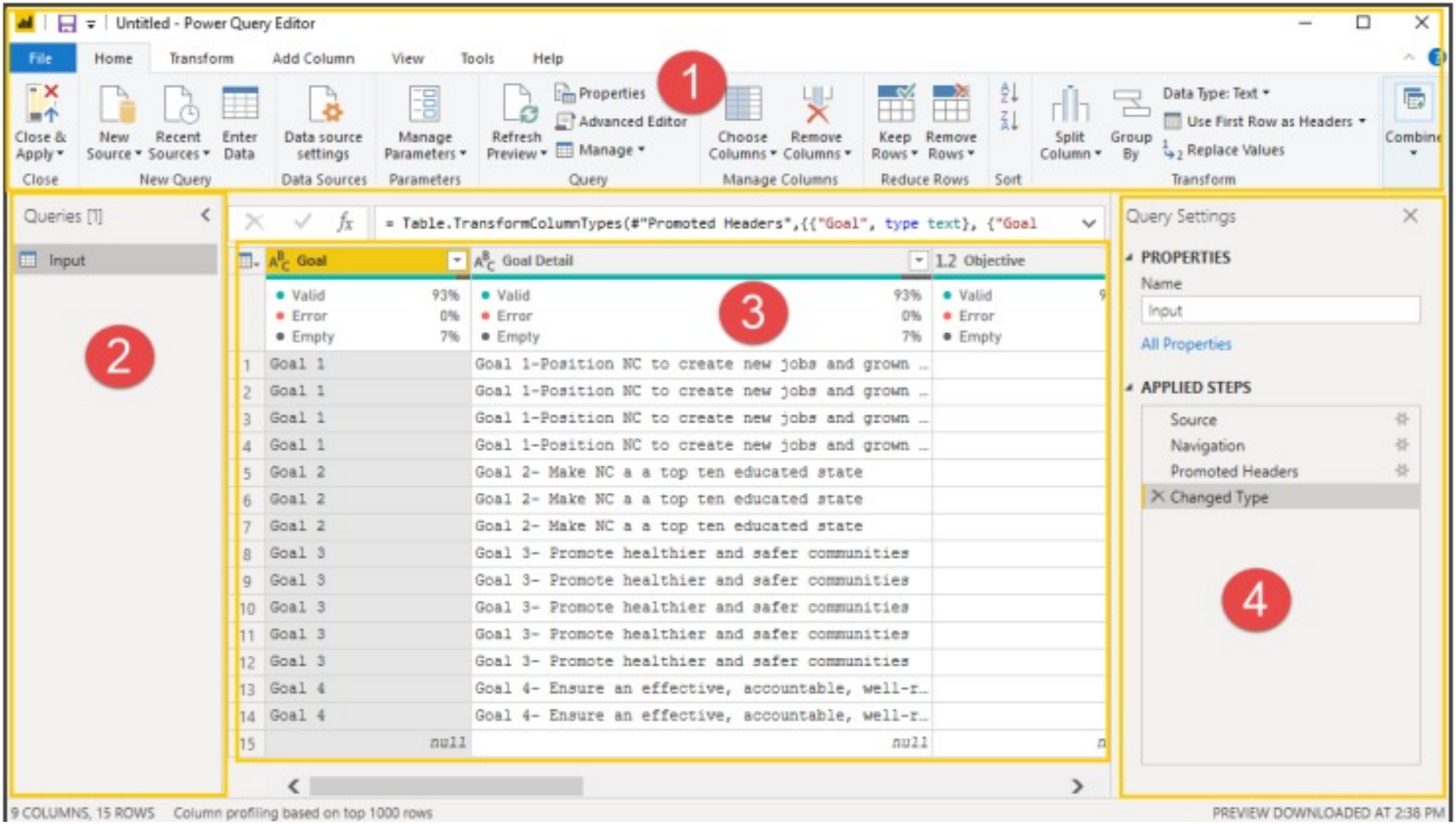
2. Click on the drop down of the Edit Queries on the bottom right corner, click on Edit Queries



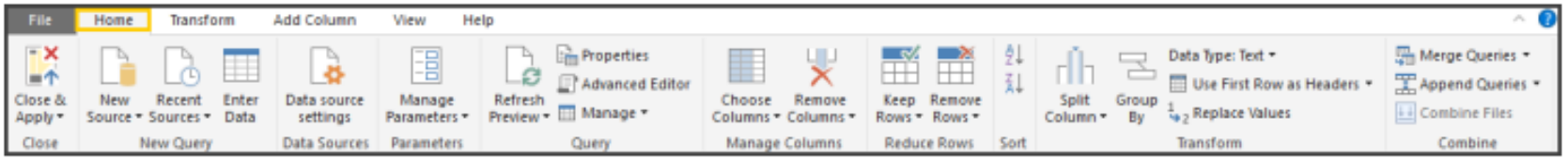

Untitled - Power Query Editor
File
Home
Transform
Add Column
View
Tools
Help
Close & Apply
Close
New Source
Recent Sources
Enter Data
New Query
Data source settings
Data Sources
Manage Parameters
Parameters
Refresh Preview
Properties
Advanced Editor
Manage
Query
Choose Columns
Remove Columns
Manage Columns
Keep Rows
Remove Rows
Reduce Rows
Sort
Split Column
Group By
Data Type:
Use First Row as Headers
Replace Values
Transform
Merge Queries
Append Queries
Combine Files
Combine
Queries [0]

3. From **Home tab > New Source > Choose Excel**

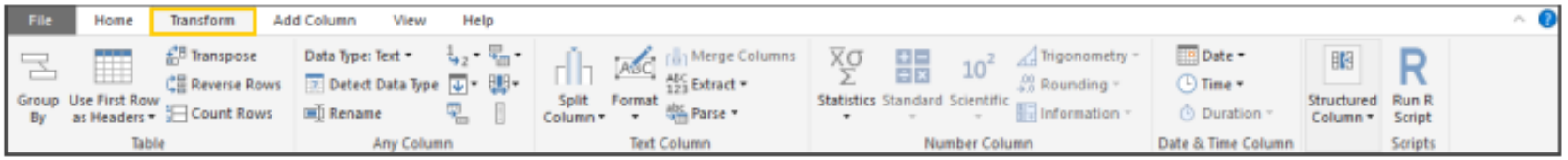



Untitled - Power Query Editor
File
Home
Transform
Add Column
View
Tools
Help
1
Close & Apply
New Source
Recent Sources
Enter Data
Data source settings
Manage Parameters
Refresh Preview
Properties
Advanced Editor
Manage
Choose Columns
Remove Columns
Keep Rows
Remove Rows
Split Column
Group By
Data Type: Text
Use First Row as Headers
Replace Values
Combine
Close
New Query
Data Sources
Parameters
Query
Manage Columns
Reduce Rows
Sort
Transform
Queries [1]
Input
2
= Table.TransformColumnTypes(#"Promoted Headers",{{"Goal", type text}, {"Goal
Goal
Goal Detail
Objective
Valid 93%
Error 0%
Empty 7%
3
1 Goal 1 Goal 1-Position NC to create new jobs and grown …
2 Goal 1 Goal 1-Position NC to create new jobs and grown …
3 Goal 1 Goal 1-Position NC to create new jobs and grown …
4 Goal 1 Goal 1-Position NC to create new jobs and grown …
5 Goal 2 Goal 2- Make NC a a top ten educated state
6 Goal 2 Goal 2- Make NC a a top ten educated state
7 Goal 2 Goal 2- Make NC a a top ten educated state
8 Goal 3 Goal 3- Promote healthier and safer communities
9 Goal 3 Goal 3- Promote healthier and safer communities
10 Goal 3 Goal 3- Promote healthier and safer communities
11 Goal 3 Goal 3- Promote healthier and safer communities
12 Goal 3 Goal 3- Promote healthier and safer communities
13 Goal 4 Goal 4- Ensure an effective, accountable, well-r…
14 Goal 4 Goal 4- Ensure an effective, accountable, well-r…
15 null null
Query Settings
PROPERTIES
Name
Input
All Properties
APPLIED STEPS
Source
Navigation
Promoted Headers
Changed Type
4
9 COLUMNS, 15 ROWS
Column profiling based on top 1000 rows
PREVIEW DOWNLOADED AT 2:38 PM

Home Tab: The **Home** tab contains the common query tasks, including the first step in any query, which is **Get Data.**

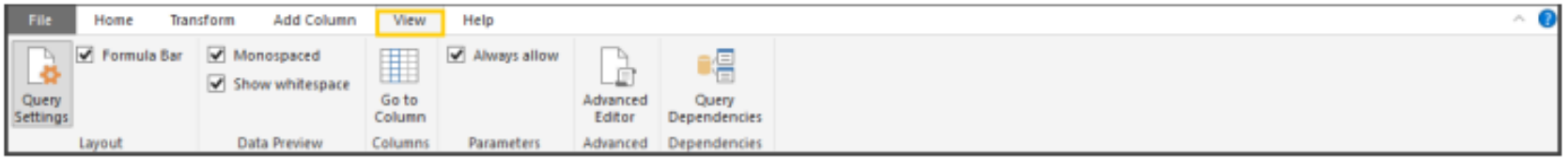


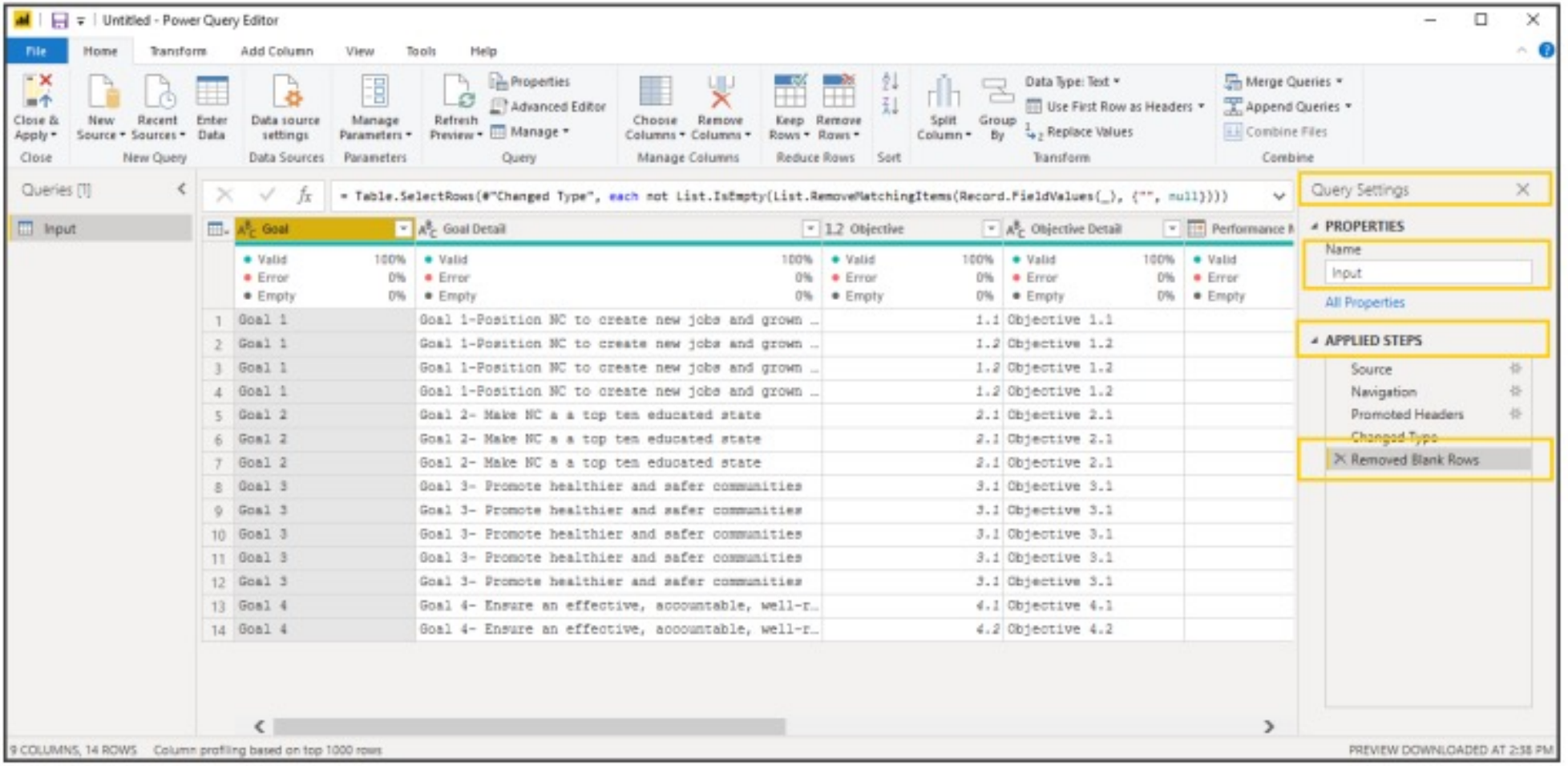
Transform: The **Transform** tab provides access to common data transformation tasks, such as adding or removing columns, changing data types, splitting columns, and other data-driven tasks.



Add Column: The **Add Column** tab provides additional tasks associated with adding a column, formatting column data, and adding custom columns. The following image shows the **Add Column** tab.

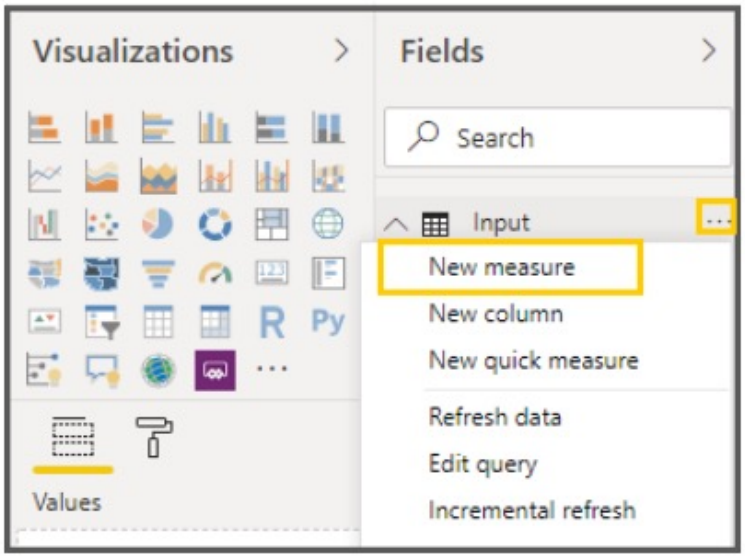
View Tab: The **View** tab on the ribbon is used to toggle whether certain panes or windows are displayed. It's also used to display the Advanced Editor. The following image shows the **View** tab.




Untitled - Power Query Editor
File
Home
Transform
Add Column
View
Tools
Help
Close & Apply
New Source
Recent Sources
Enter Data
Data source settings
Manage Parameters
Refresh Preview
Properties
Advanced Editor
Manage
Choose Columns
Remove Columns
Keep Rows
Remove Rows
Split Column
Group By
Data Type: Text
Use First Row as Headers
Replace Values
Merge Queries
Append Queries
Combine Files
Close
New Query
Data Sources
Parameters
Query
Manage Columns
Reduce Rows
Sort
Transform
Combine
Queries [1]
Input
= Table.SelectRows(#"Changed Type", each not List.IsEmpty(List.RemoveMatchingItems(Record.FieldValues(_), {"", null})))
Goal
Goal Detail
Objective
Objective Detail
Performance
Valid 100%
Error 0%
Empty 0%
1 Goal 1 Goal 1-Position NC to create new jobs and grown … 1.1 Objective 1.1
2 Goal 1 Goal 1-Position NC to create new jobs and grown … 1.2 Objective 1.2
3 Goal 1 Goal 1-Position NC to create new jobs and grown … 1.2 Objective 1.2
4 Goal 1 Goal 1-Position NC to create new jobs and grown … 1.2 Objective 1.2
5 Goal 2 Goal 2- Make NC a a top ten educated state 2.1 Objective 2.1
6 Goal 2 Goal 2- Make NC a a top ten educated state 2.1 Objective 2.1
7 Goal 2 Goal 2- Make NC a a top ten educated state 2.1 Objective 2.1
8 Goal 3 Goal 3- Promote healthier and safer communities 3.1 Objective 3.1
9 Goal 3 Goal 3- Promote healthier and safer communities 3.1 Objective 3.1
10 Goal 3 Goal 3- Promote healthier and safer communities 3.1 Objective 3.1
11 Goal 3 Goal 3- Promote healthier and safer communities 3.1 Objective 3.1
12 Goal 3 Goal 3- Promote healthier and safer communities 3.1 Objective 3.1
13 Goal 4 Goal 4- Ensure an effective, accountable, well-r… 4.1 Objective 4.1
14 Goal 4 Goal 4- Ensure an effective, accountable, well-r… 4.2 Objective 4.2
Query Settings
PROPERTIES
Name
Input
All Properties
APPLIED STEPS
Source
Navigation
Promoted Headers
Changed Type
Removed Blank Rows
9 COLUMNS, 14 ROWS
Column profiling based on top 1000 rows
PREVIEW DOWNLOADED AT 2:38 PM

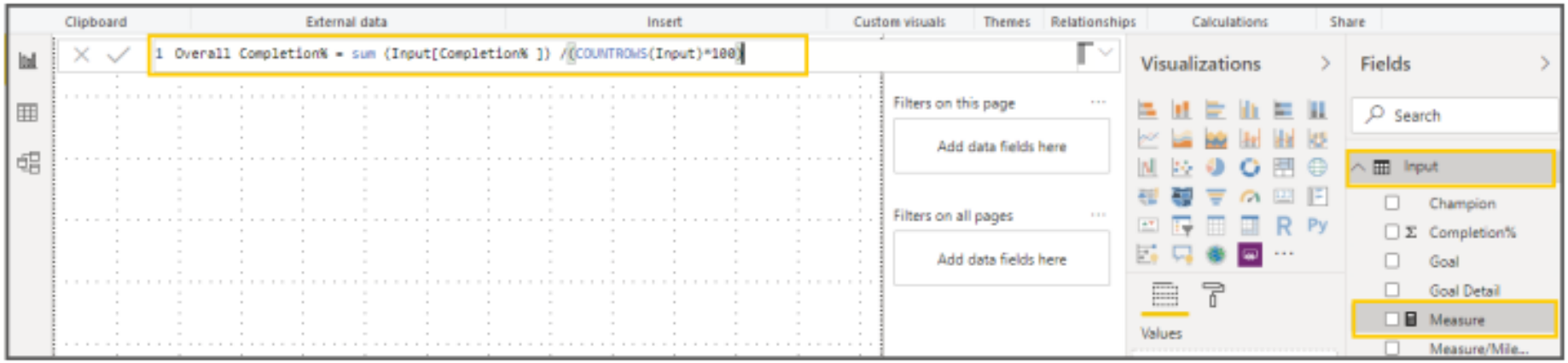
Measures

- In general, Measures are used to calculate aggregates, such as the sum or average of a column



Expression Bar

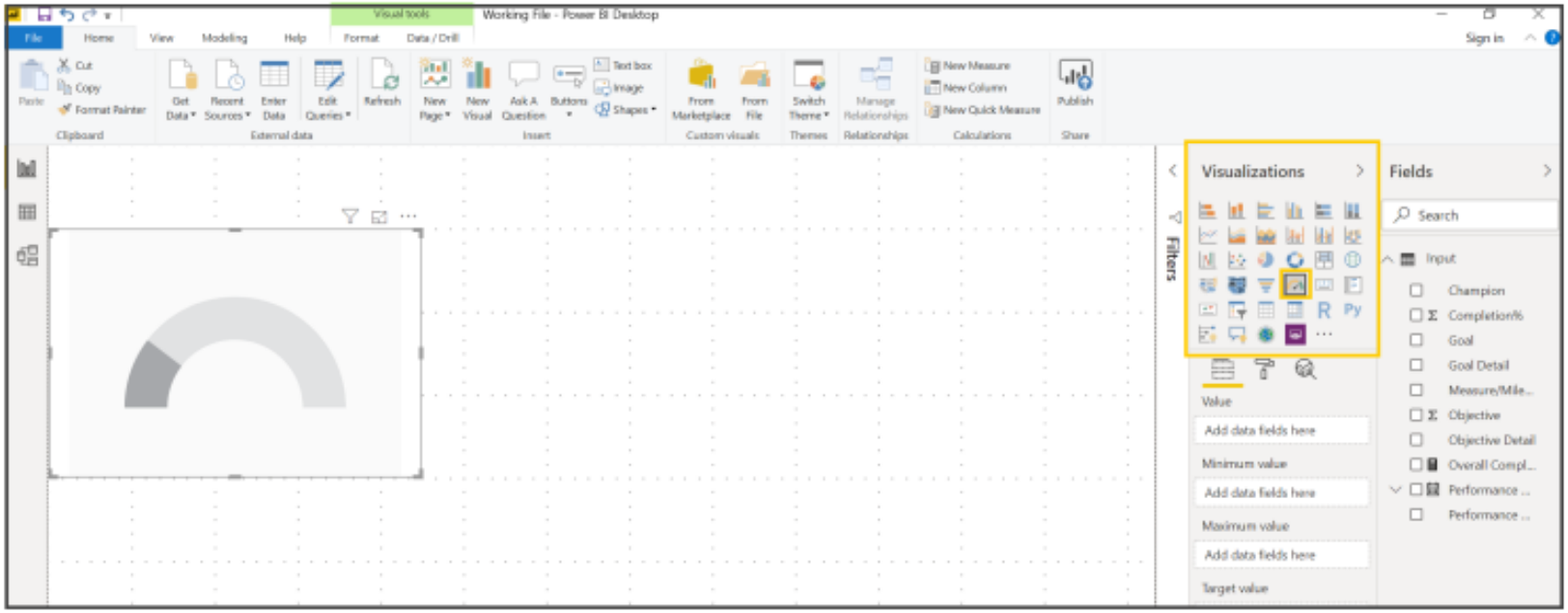
Overall Completion% = sum (Input [Completion%]) /(COUNTROWS(Input)*100)



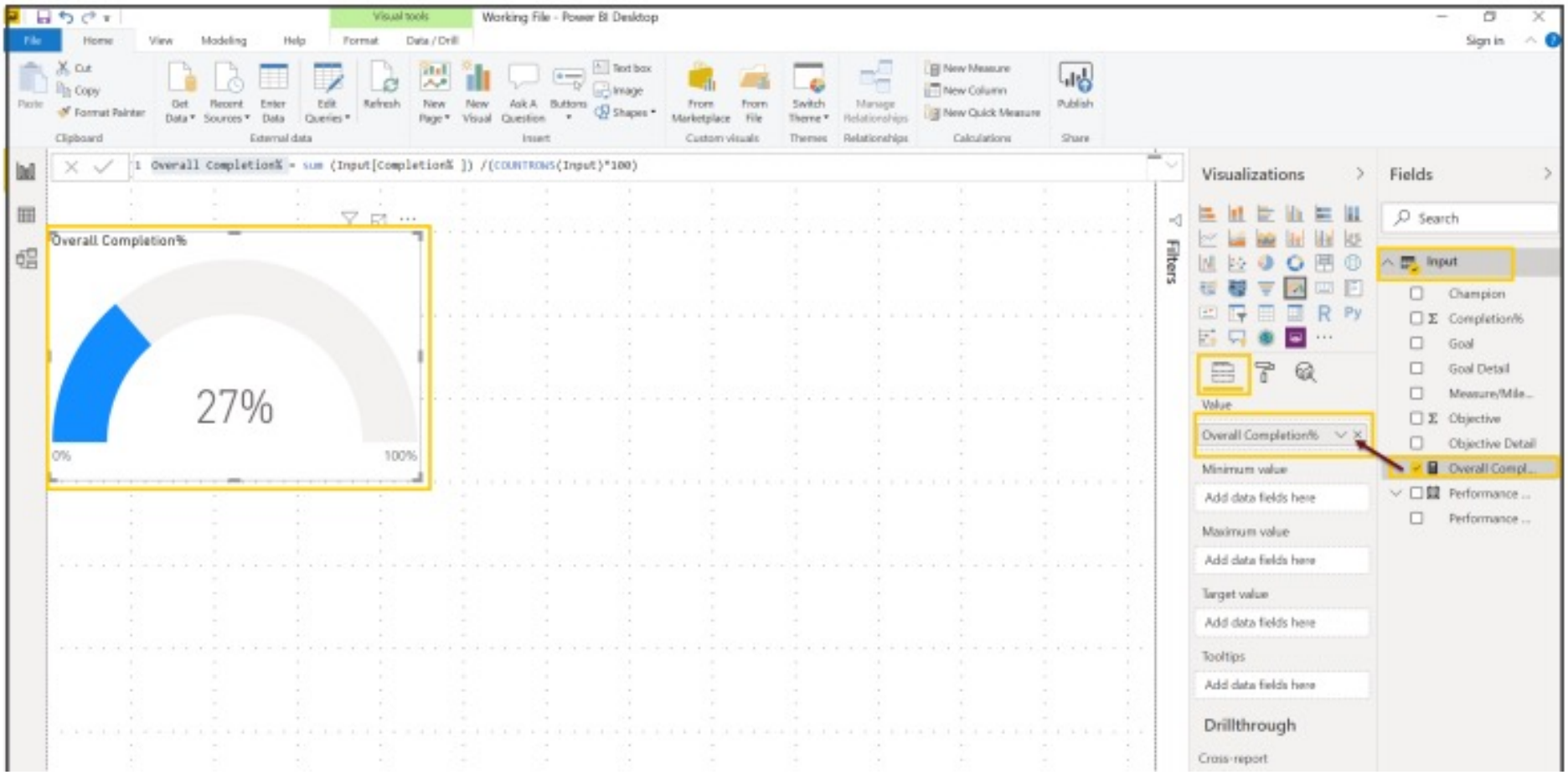
Creating Reports & Visualizations

Exercise 6: Creating your first visualization (Completion % of All Goals) Gauge Chart

15. Click on Visualizations Pane and Click on Gauge Chart

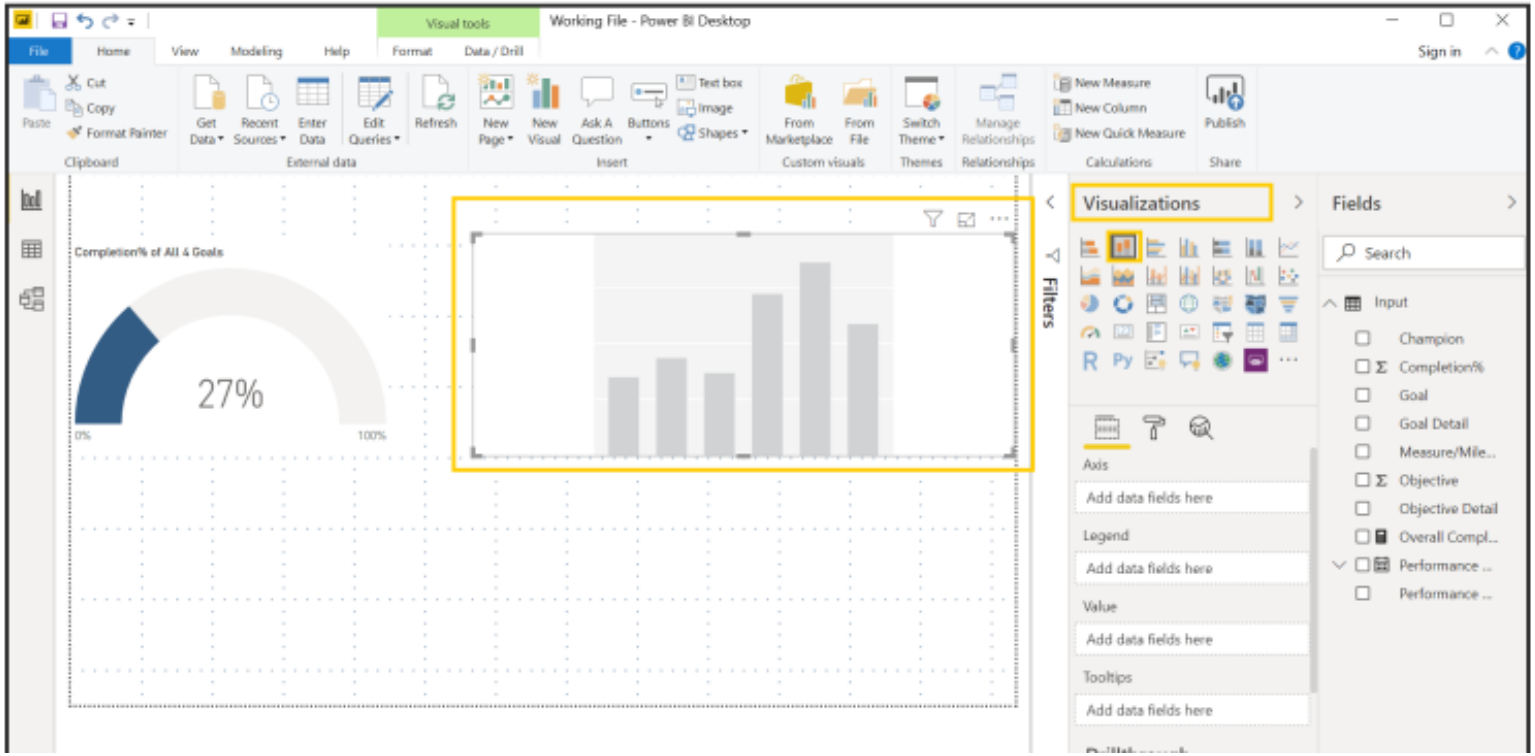


16. Expand Input Query, Drag Overall Completion% to the Value section of the Fields pane of the gauge Visual

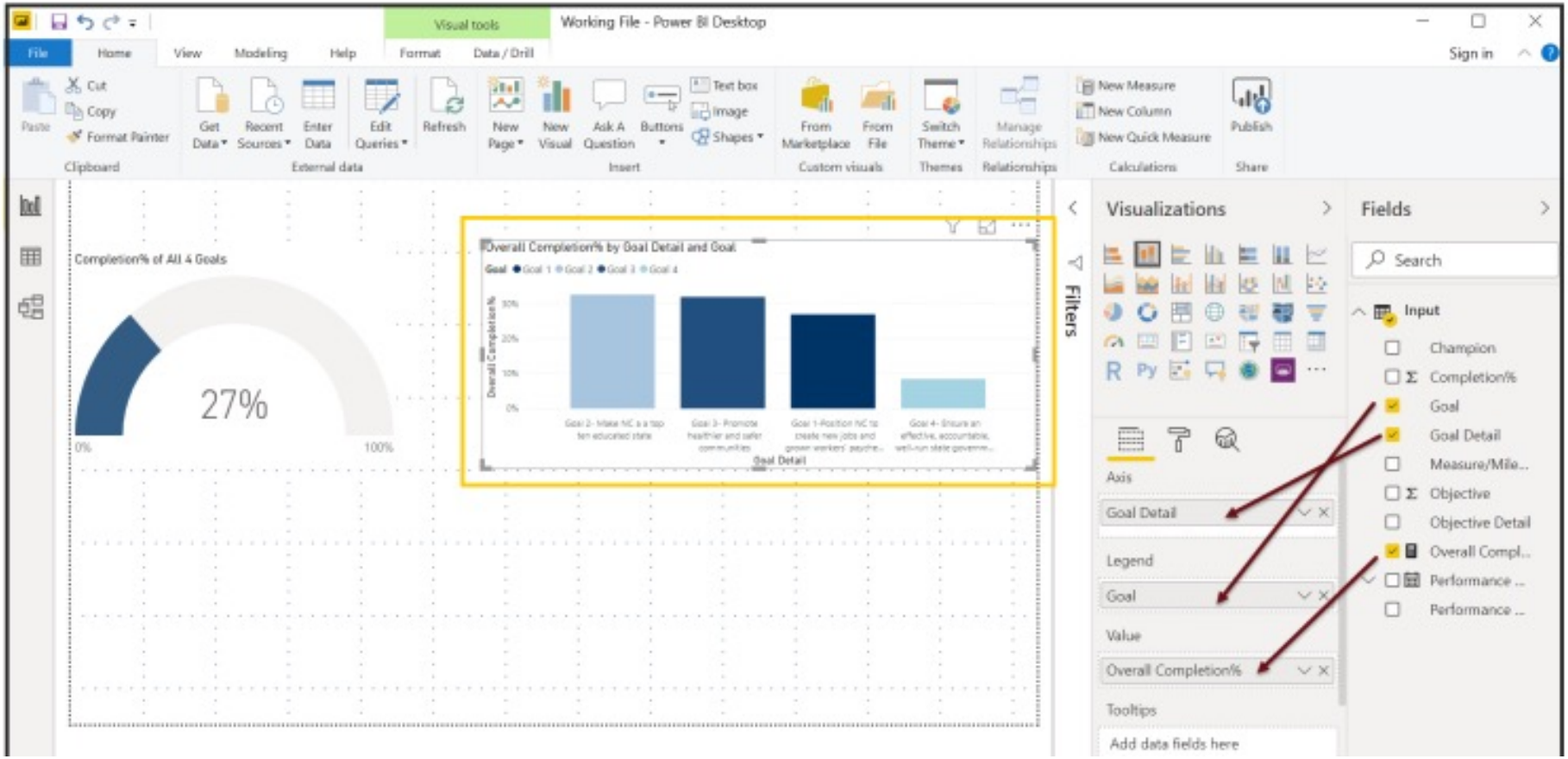


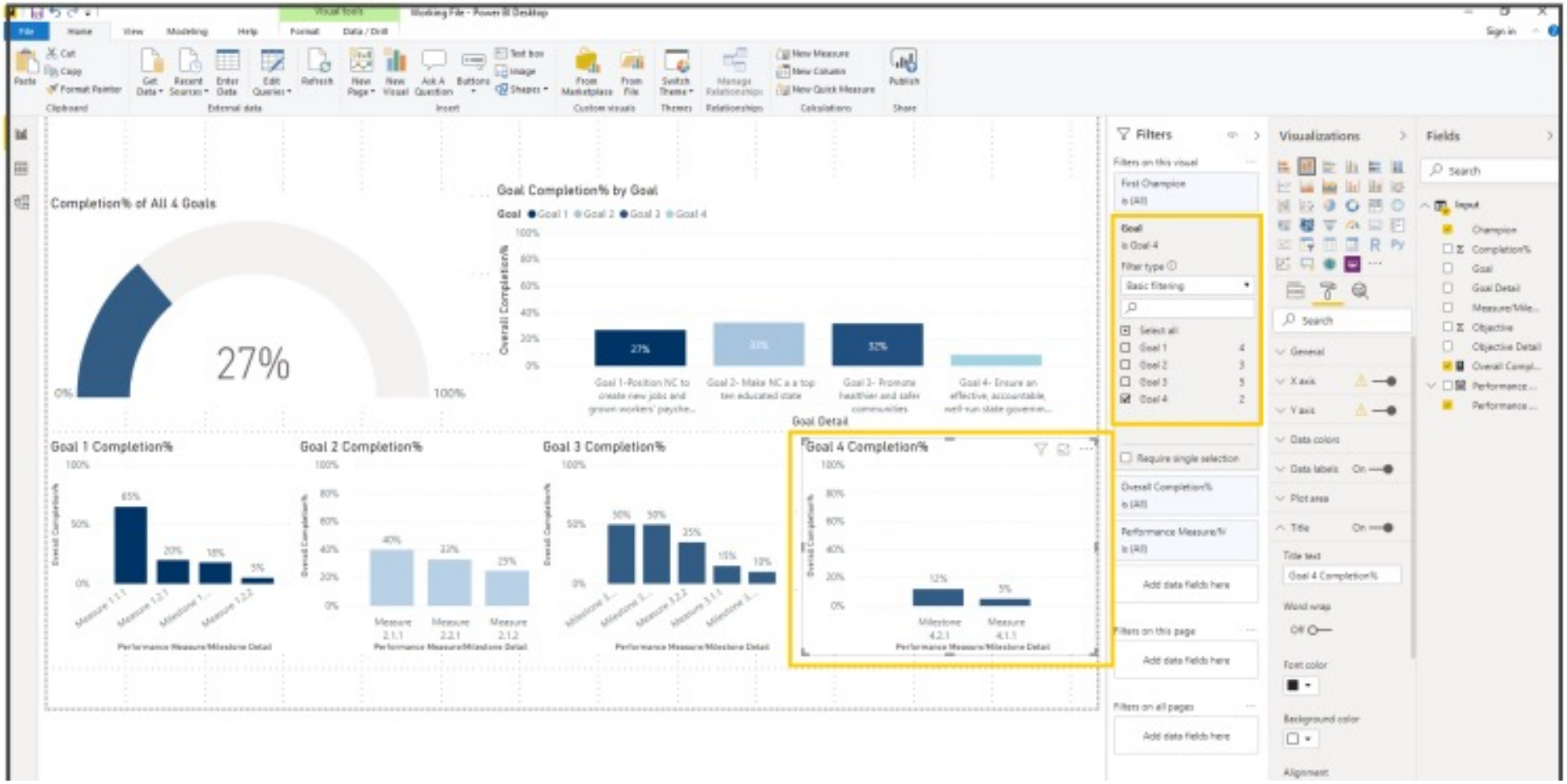
Add Stacked Column Chart

23. Click anywhere on the Canvas other than the visuals, select Stacked Column Chart and bring the visual next to the Donut Chart.



24. Expand Input, Drag **Overall Completion%** to the Value section, **Goal Detail** to the **Legend**, **Goal** to the **Axis** of the **Fields pane** of **the Stacked Column Visual.**




Completion% of All 4 Goals
27%
0%
100%
Goal Completion% by Goal
Goal
Goal 1
Goal 2
Goal 3
Goal 4
Overall Completion%
27%
32%
Goal 1-Position NC to create new jobs and grown workers' psyche...
Goal 2- Make NC a a top ten educated state
Goal 3- Promote healthier and safer communities
Goal 4- Ensure an effective, accountable, well-run state governm...
Goal 1 Completion%
Goal 2 Completion%
Goal 3 Completion%
Goal Detail
Goal 4 Completion%
Performance Measure/Milestone Detail
Filters
Visualizations
Fields
Search
Goal 4
Basic filtering
Select all
Goal 1 4
Goal 2 3
Goal 3 5
Goal 4 2
Require single selection
Add data fields here
Filters on this page
Filters on all pages
Title
Goal 4 Completion%
Word wrap
Font color
Background color
Alignment