1. Review of Lecture 3

2. Tutorial Week 4

3. Attendance & Tutorial Questions - Recognising student participation and engagement specifically identifying those who are most actively involved!

Lecturer/Tutor: Dr. Farshid Keivanian

Tutorial Week 4

The document "Tutor_Week4_Business Analytics Fundamentals" ....

1. Summary of Lecture 3

Extraction Transformation and Load?

Extract, transform, and load (ETL) is a process in data warehousing that involves
- extracting data from outside sources,
- transforming it to fit business needs, and ultimately
- loading it into the data warehouse.

ETL is important, as it is the way data actually gets loaded into the warehouse.

1. Summary of Lecture 3

Extract

HOLMES INSTITUTE

- The first part of an ETL process is to extract the data from the source systems.

- Most Business Intelligence environments consolidate data from different source systems. Each separate system may also use a different data organization / format.

- Common data source formats are relational databases and flat files.

- Extraction converts the data into a format for transformation processing
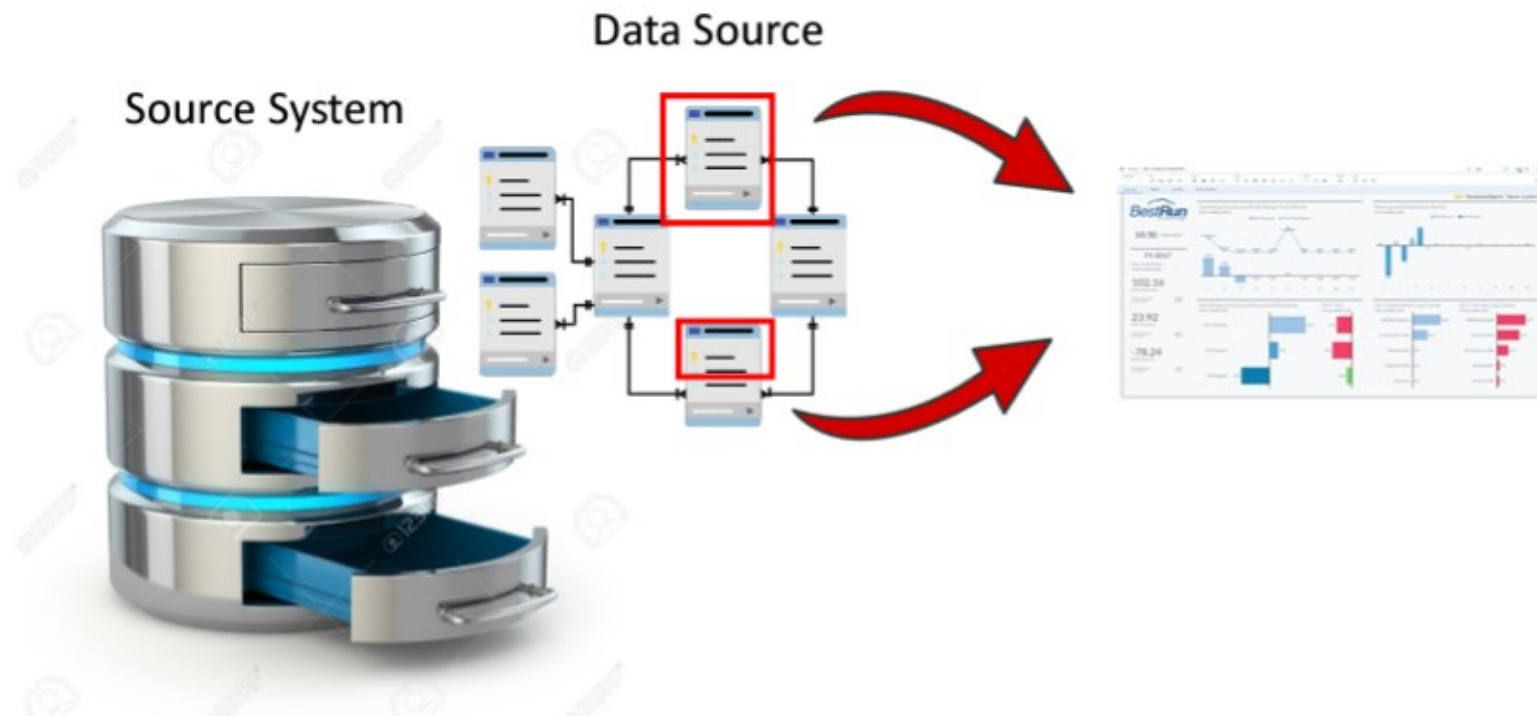
1. Summary of Lecture 3

Sources of data

Data Sources

Data Source = Source System > selected Tables > selected Fields > selected Data

1. Summary of Lecture 3

Let's consider some practical examples:

1. Australian Census Data Extraction:

- **Source System:** The Australian Bureau of Statistics (ABS) online database.

- **Extraction:** Data is pulled from the census, which includes demographic, economic, and housing characteristics of the population.

- **Transformation:** The data might be anonymized and summarized to ensure privacy and to provide insights at different geographic levels, such as states or local government areas.

- **Load:** Data is loaded into an analytical platform used by policymakers and researchers to monitor trends and make decisions.

1. Summary of Lecture 3

2. Retail Sales Data Integration:

- **Source Systems:** Various POS (Point of Sale) systems across retail chains in Australia.

- **Extraction:** Sales data, including transaction details, product information, and customer interactions, are extracted daily from different retail stores.

- **Transformation:** Data is cleaned and aggregated, product IDs are normalized, and sales metrics are calculated (e.g., total daily sales per product).

- **Load:** Consolidated data is loaded into a Business Intelligence tool to track performance, forecast demand, and optimize stock levels.

1. Summary of Lecture 3

3. Healthcare Patient Records System:

- **Source Systems:** Hospital and clinic electronic health record (EHR) systems.

- **Extraction:** Patient records including diagnoses, treatment information, and patient outcomes are extracted.

- **Transformation:** Identifiable information is encrypted or removed, records are standardized to a common format, and health outcomes are coded according to international standards.

- **Load:** Data is loaded into a central data warehouse for analysis, used to improve patient care and for medical research.
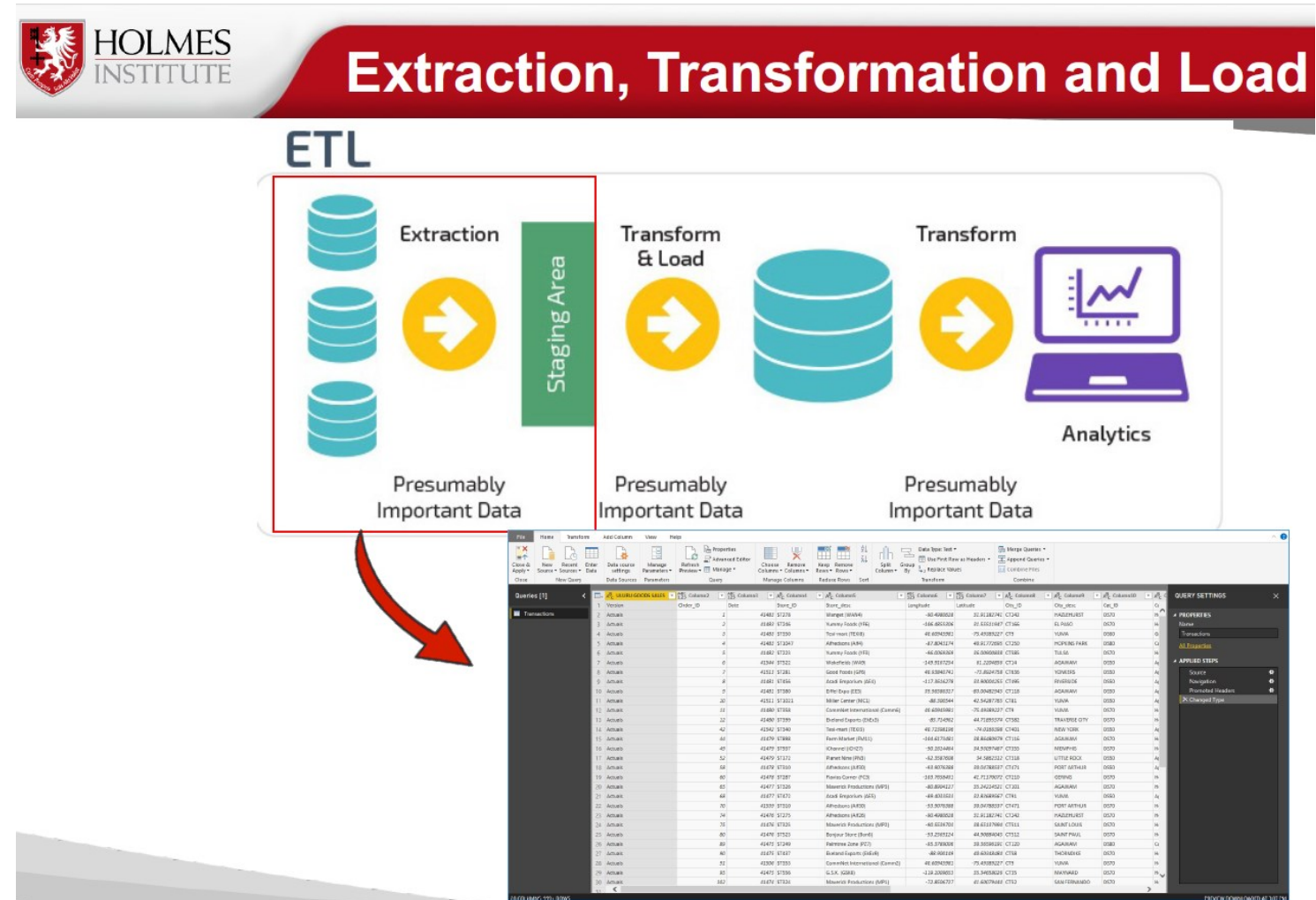
1. Summary of Lecture 3

4. Financial Data for Credit Risk Assessment:

- **Source Systems:** Banking systems that contain account information, transaction details, and customer profiles.

- **Extraction:** Data on loan applications, repayment histories, and account balances are extracted from multiple banking systems.

- **Transformation:** Data is normalized, credit scores are recalculated, and risk profiles are updated.

- **Load:** The transformed data is loaded into a risk management system to assist with real-time credit risk assessment and decision-making.
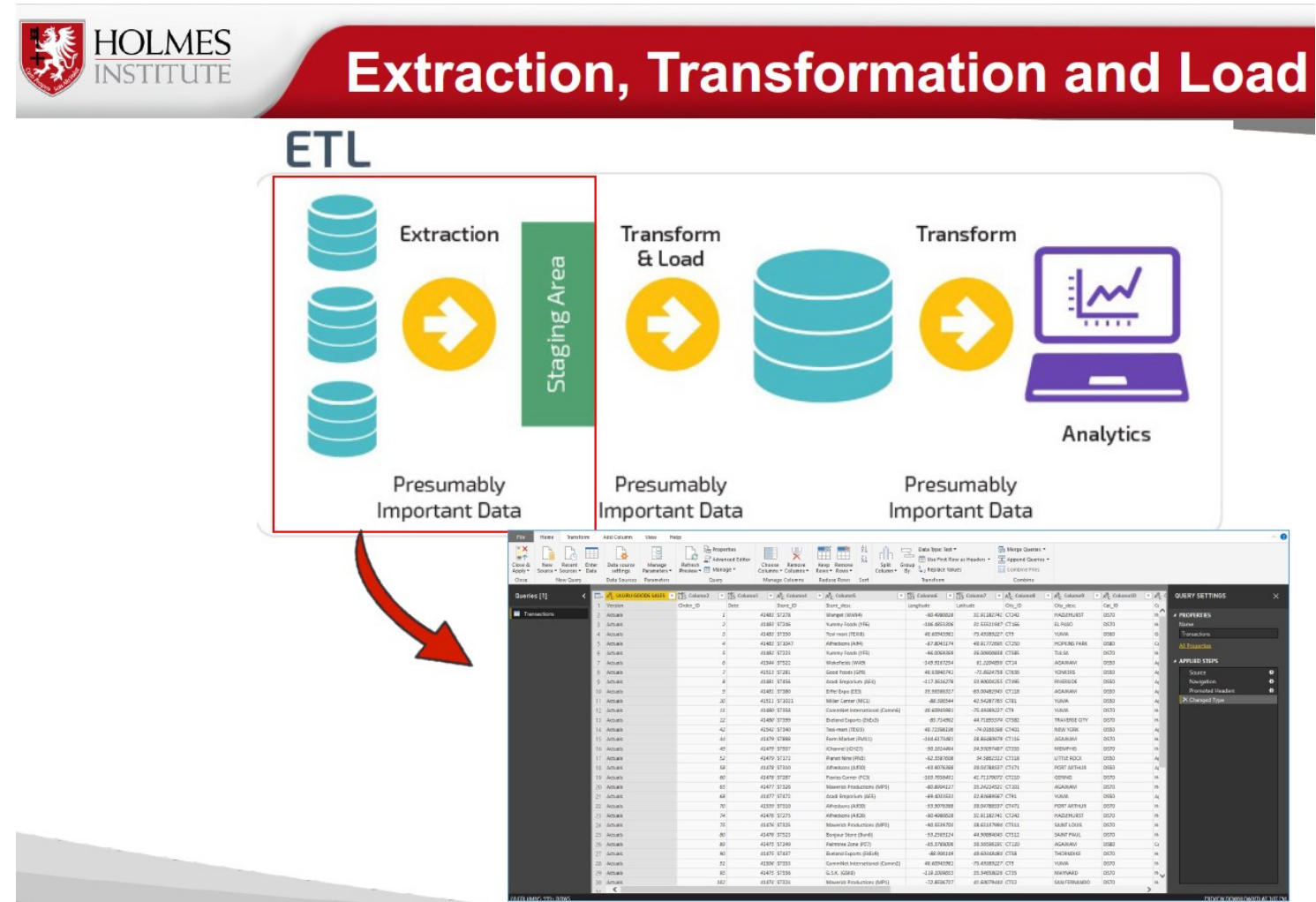
1. Summary of Lecture 3

## 5. Utility Usage Data for Energy Management:

- **Source Systems:** Smart meters and IoT devices in homes and businesses across Australia.

- **Extraction:** Usage data for electricity, water, and gas are extracted at regular intervals.

- **Transformation:** Data is aggregated by region and time of use, anomalies are detected and corrected, and consumption patterns are analyzed.

- **Load:** Processed data is loaded into an energy management system to optimize grid performance, predict demand, and support sustainable energy use initiatives.

1. Summary of Lecture 3 - Practical Applications of the Transformation Techniques

1. Selecting Columns for Healthcare Analytics:

- In a national healthcare database, analysts might select only certain columns relevant to public health studies, such as age, region, diagnosis, and outcome, while omitting personally identifiable information to ensure patient privacy.

1. Summary of Lecture 3 - Practical Applications of the Transformation Techniques

2. Translating Coded Values in Wildlife Studies:

- When studying the population of a particular species, such as koalas, researchers might need to translate coded values from field data. For instance, "M" and "F" recorded in the field might be converted to "1" for male and "2" for female in the analytical dataset to standardize the data before analysis.

1. Summary of Lecture 3 - Practical Applications of the Transformation Techniques

3. Encoding Free-form Values in Customer Feedback:

- In customer service data, various terms indicating customer satisfaction, such as "satisfied," "happy," or "content," might be encoded into a numerical satisfaction score (e.g., "1" for positive feedback) to allow for quantitative analysis.

1. Summary of Lecture 3 - Practical Applications of the Transformation Techniques

4. Deriving Calculated Values in E-commerce:

- An Australian online retailer might derive new calculated values, like total sales per transaction, by multiplying the quantity of items purchased by the unit price to analyze sales patterns and product performance.
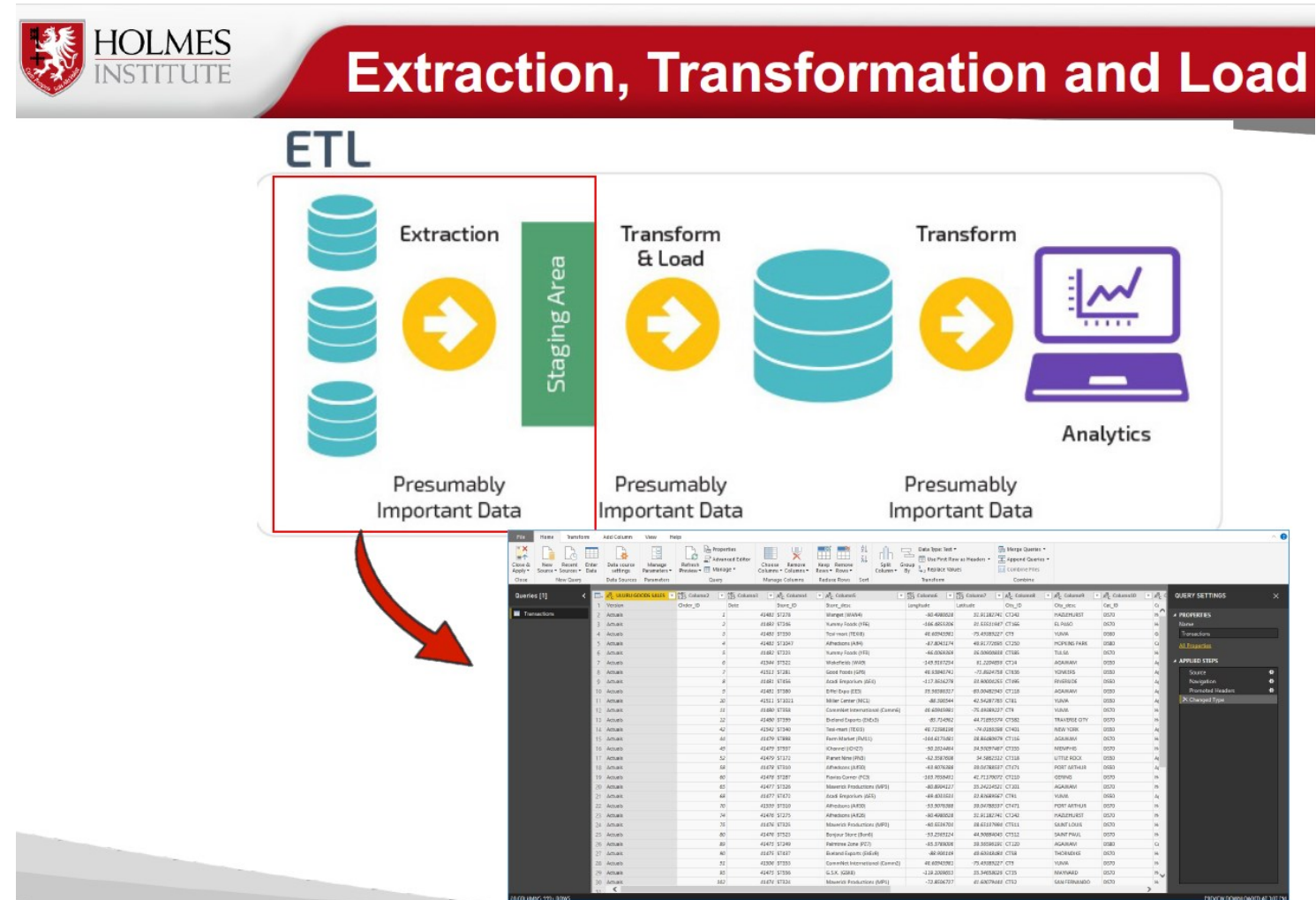
1. Summary of Lecture 3 - Practical Applications of the Transformation Techniques

5. Joining Data from Multiple Sources in Tourism Analytics:

- Tourism data might be merged from various sources, such as hotel bookings, attraction ticket sales, and transport usage, to get a comprehensive view of tourism trends and spending behaviors in different Australian regions.

1. Summary of Lecture 3 - Practical Applications of the Transformation Techniques

6. Summarizing Rows in Real Estate:
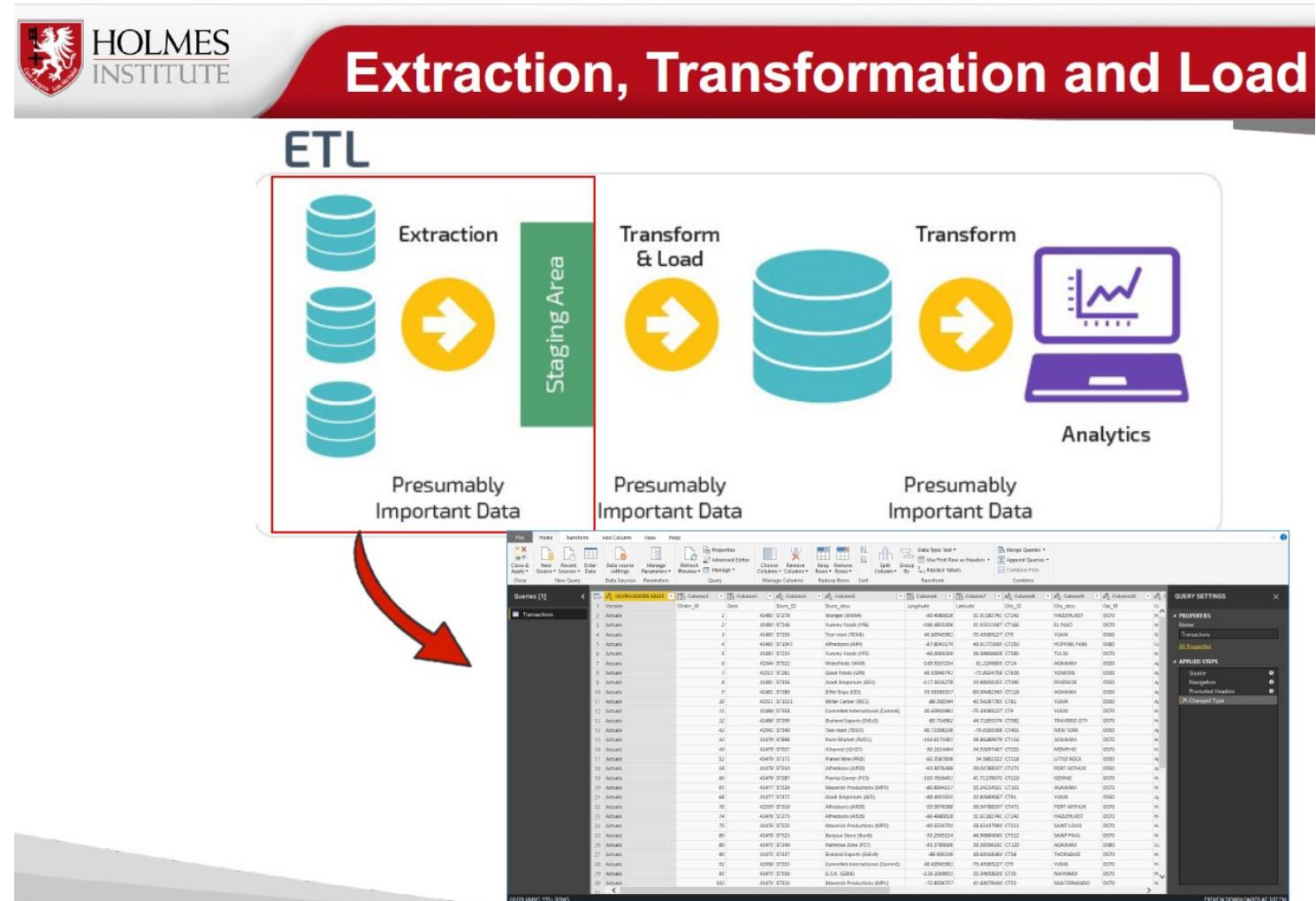
Real estate agencies could summarize multiple rows of data to see the total number of sales per region or the average selling price of properties in different suburbs of cities like Sydney or Melbourne.
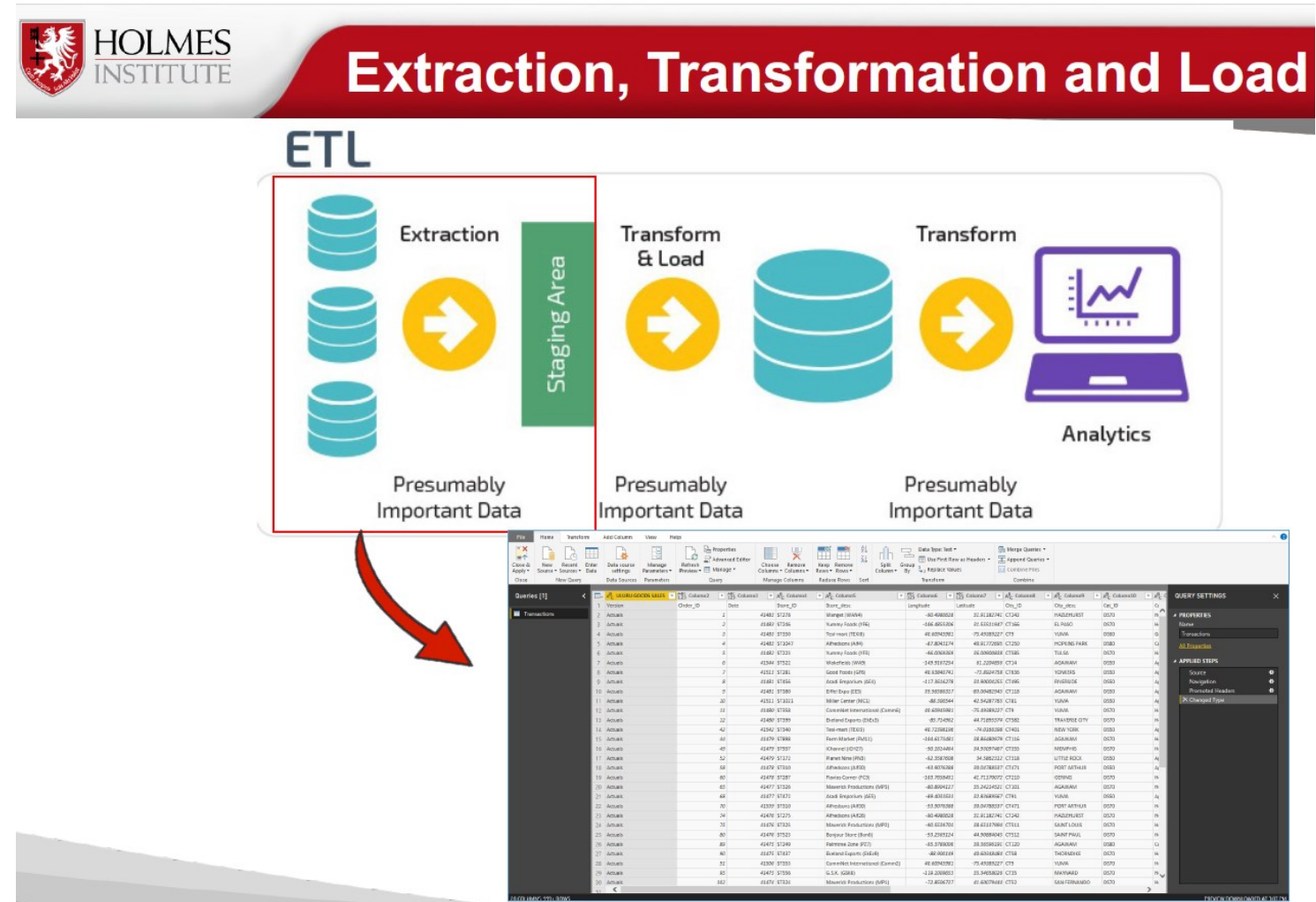
1. Summary of Lecture 3 - Practical Applications of the Transformation Techniques

7. Transposing or Pivoting Education Data:

- In the education sector, schools might transpose yearly student performance data to compare subjects across different years or to track the performance of individual students over time.
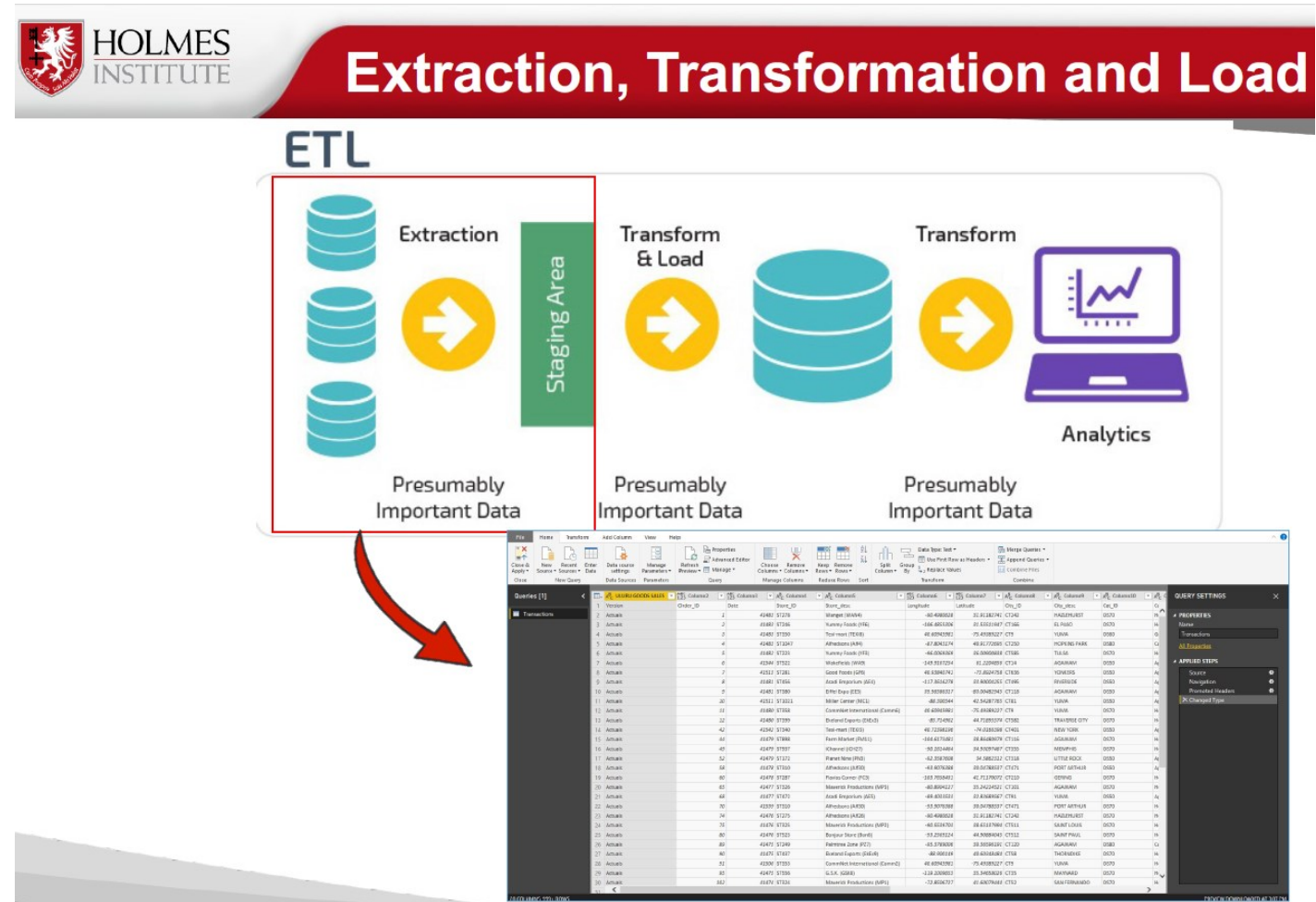
These examples show how data transformation is applied in various Australian industries, supporting the notion that the ability to effectively manipulate and prepare data is crucial in deriving valuable insights for business analytics.

2. Tutorial Week 4

In the last session, we transformed the data in the queries by editing the headings, data types, deleted columns, created custom columns and removed null values. These activities can be referred to as Data Preparation. Data Preparation is usually the first step in the Business Intelligence workflow. The diagram below shows the Business Intelligence workflow and the associated tools in Power BI.

Business Intelligence Workflow		
Data Preparation ➡	Data Modelling ➡	Data Visualisation

Power BI		
Query Editor ➡	Data View + Relationship View ➡	Report View
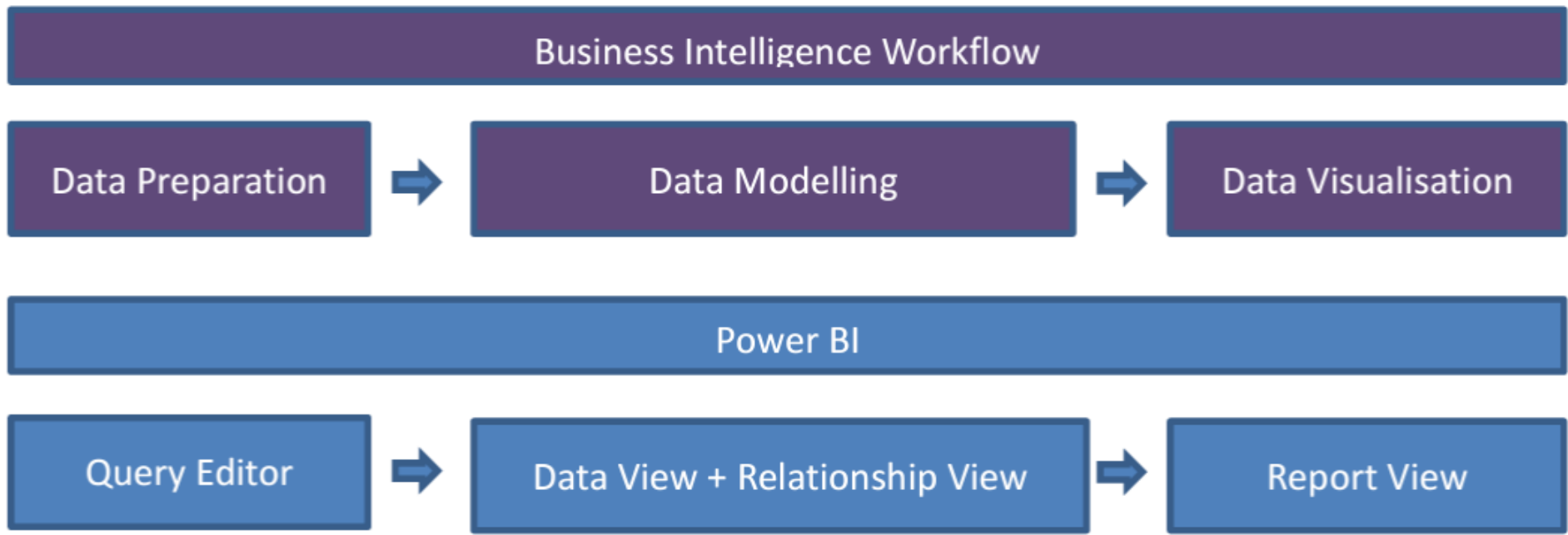
- Open '**Tutor Week 3.pbix**'.

- Select Model View

- Collapse Properties and Data

- Select Fit to Page
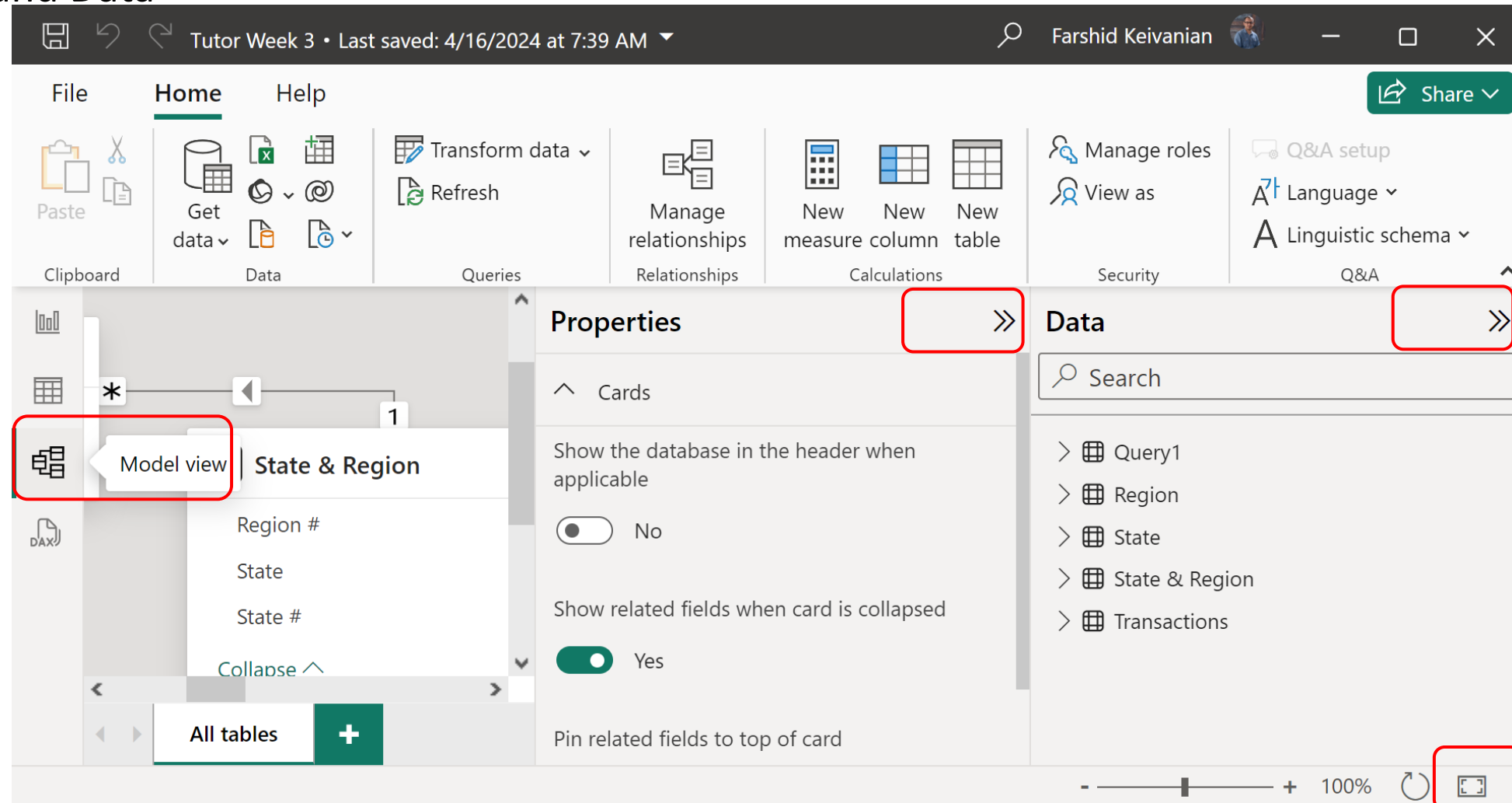
- All the datasets appear on the screen and the datasets have automatically been related.

- If Power BI Desktop cannot determine with a high-level of confidence there is a match, it will not automatically create the relationship. You can manually create relationships between table.
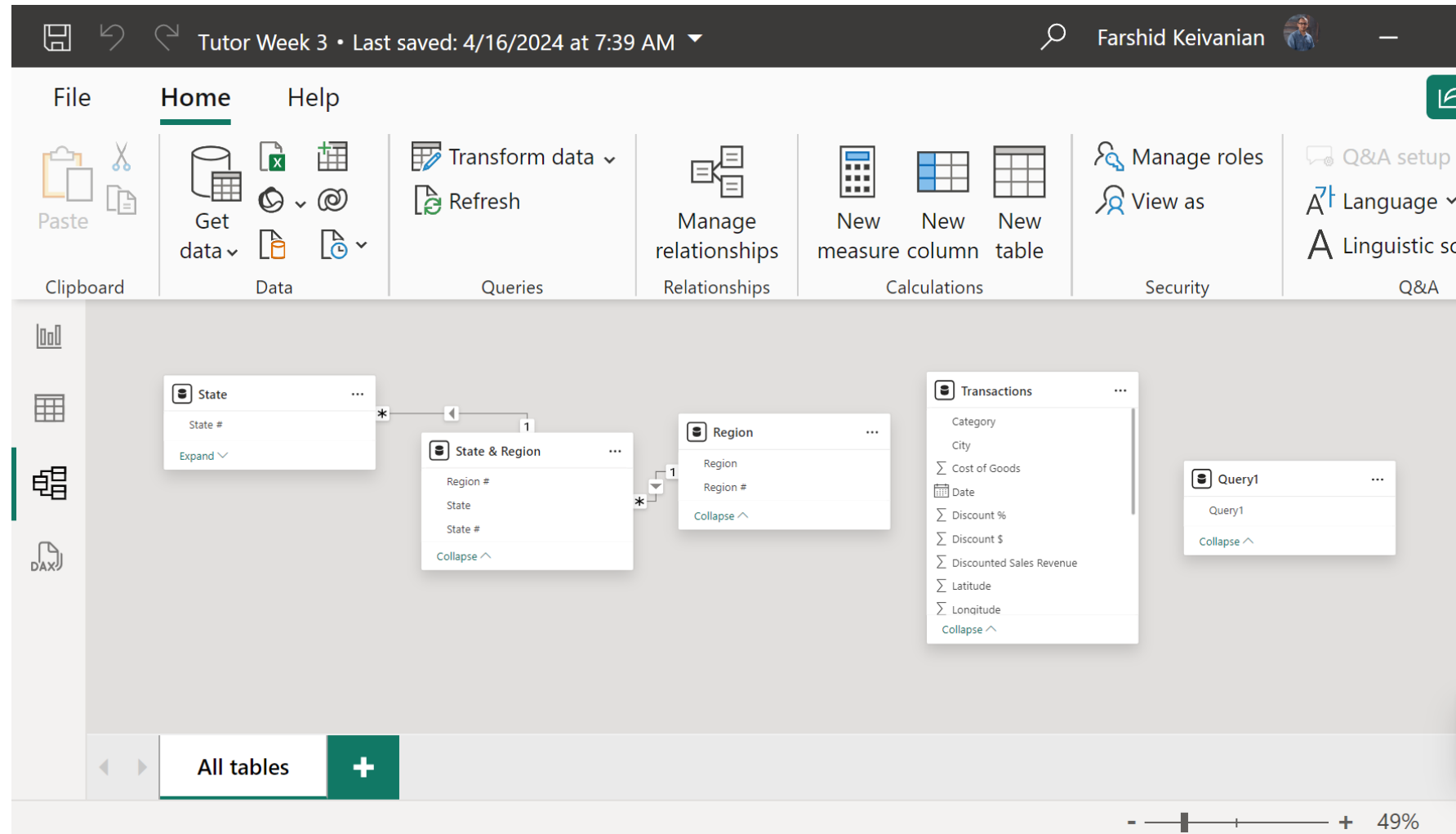
- Click on . . . at **Transactions**

- Select **Manage Relationships**

- Select **New**
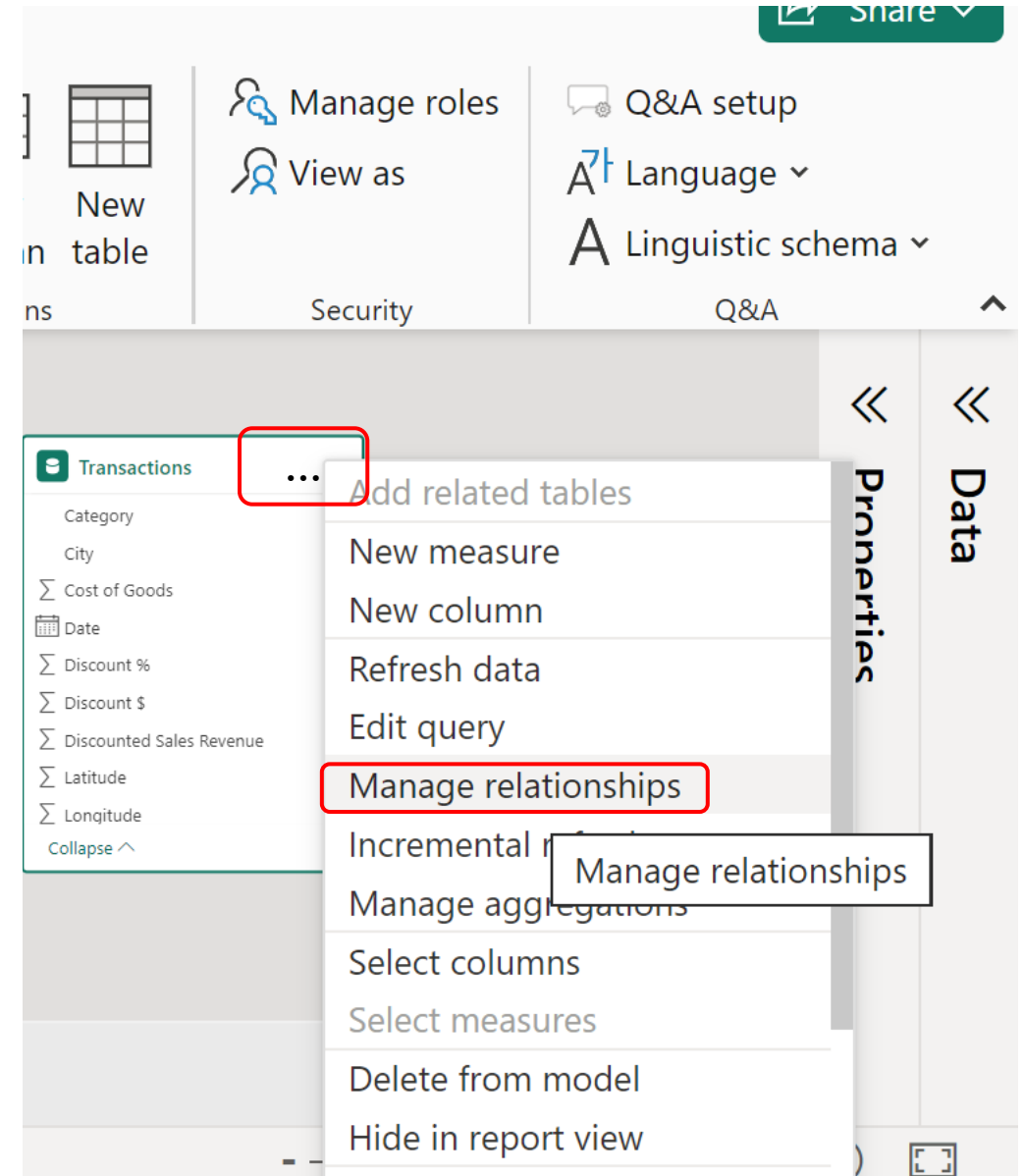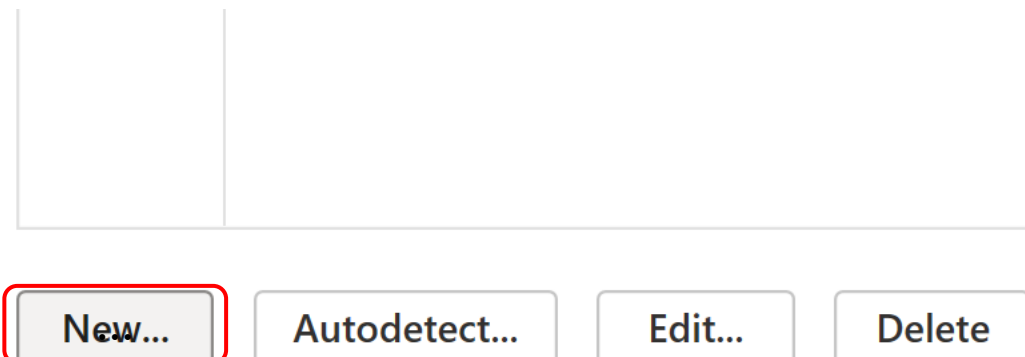
- **Select Transactions** and **State**

- Press **Tab**

- Select **OK**

New relationship

Select tables and columns that are related.

Transactions ▾

Version	Order_ID	Date	Store	Longitude	Latitude	City	Category
Actuals	310	Sunday, 30 October 2016	iChannel (iCH10)	-71.1036819	42.0970247	WEYMOUTH	Apparel
Actuals	372	Tuesday, 28 June 2016	Alfredsons (Alf32)	40.70941658	-74.0100709	NEW YORK	Apparel
Actuals	573	Tuesday, 7 June 2016	Finnys (Finn16)	40.71474749	-74.00042342	NEW YORK	Apparel

‹ ▬▬▬▬▬▬▬▬▬▬ ›

State ▾

City	State #	Description	Column4	Column5	Column6	Column7	Column8
CT212	8050456	GILLETTE	null	null	null	null	null
CT486	8050456	RAWLINS	null	null	null	null	null
CT498	8050456	ROCK SPRINGS	null	null	null	null	null

Cardinality

Many to one (*:1) ▾

Cross filter direction

Single ▾

- Press **Tab**

- Select **Close**

Active	From: Table (Column)	To: Table (Column)
☑	State (State #)	State & Region (State #)
☑	State & Region (Region #)	Region (Region #)
☑	Transactions (City)	State (City)

New...   Autodetect...   Edit...   Delete

Close

- Arrange them

- Move mouse over the join line between State and Transactions

- Notice that city in Transactions is joined to city in State
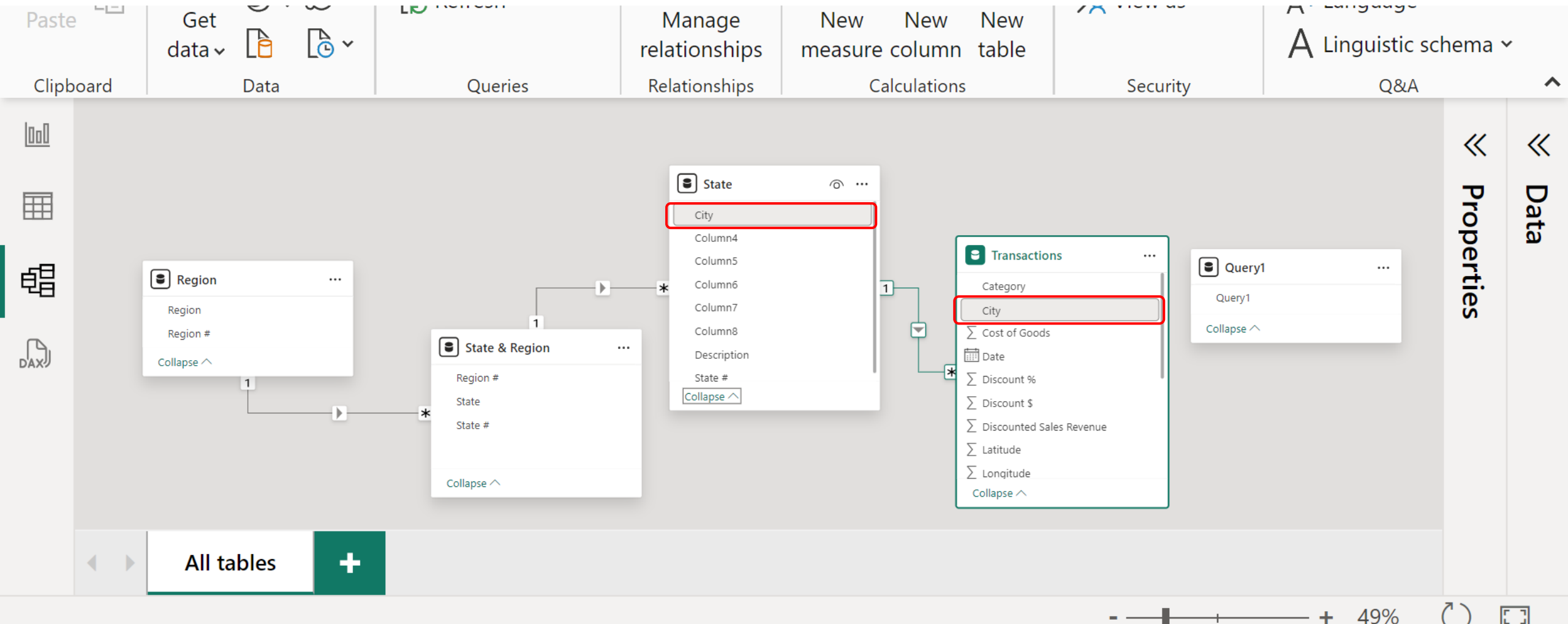
- Select the State-Transaction join line

- Right Click on it and choose **Delete**

- Re-create the relationship between city in Transactions and State

- Click **New**

- Select **Transactions** and **State**

- Press **Tab**

- Select **OK**

- Select **Close**

Version	Order_ID	Date	Store	Longitude	Latitude	City	Category
Actuals	310	Sunday, 30 October 2016	iChannel (iCH10)	-71.1036819	42.0970247	WEYMOUTH	Apparel
Actuals	372	Tuesday, 28 June 2016	Alfredsons (Alf32)	40.70941658	-74.0100709	NEW YORK	Apparel
Actuals	573	Tuesday, 7 June 2016	Finnys (Finn16)	40.71474749	-74.00042342	NEW YORK	Apparel

Active	From: Table (Column)	To: Table (Column)
☑	State (State #)	State & Region (State #)
☑	State & Region (Region #)	Region (Region #)
☑	Transactions (City)	State (City)

	ription	Column4	Column5	Column6	Column7	Column8
	ETTE	null	null	null	null	null
	LINS	null	null	null	null	null
	SPRINGS	null	null	null	null	null

Cross filter direction

Single

Apply security filter in both directions

o active

tegrity

New... Autodetect... Edit... Delete

Close

OK Cancel

- Move Mouse over to see the relationship between City in Transactions and City in State

2. Tutorial Week 4 – Visualising Data

The visualisations can be the basis for a report. You have been asked to create a report that lists the total Profit by Region, by State, by City.

- Select **Report View**

- **Table**
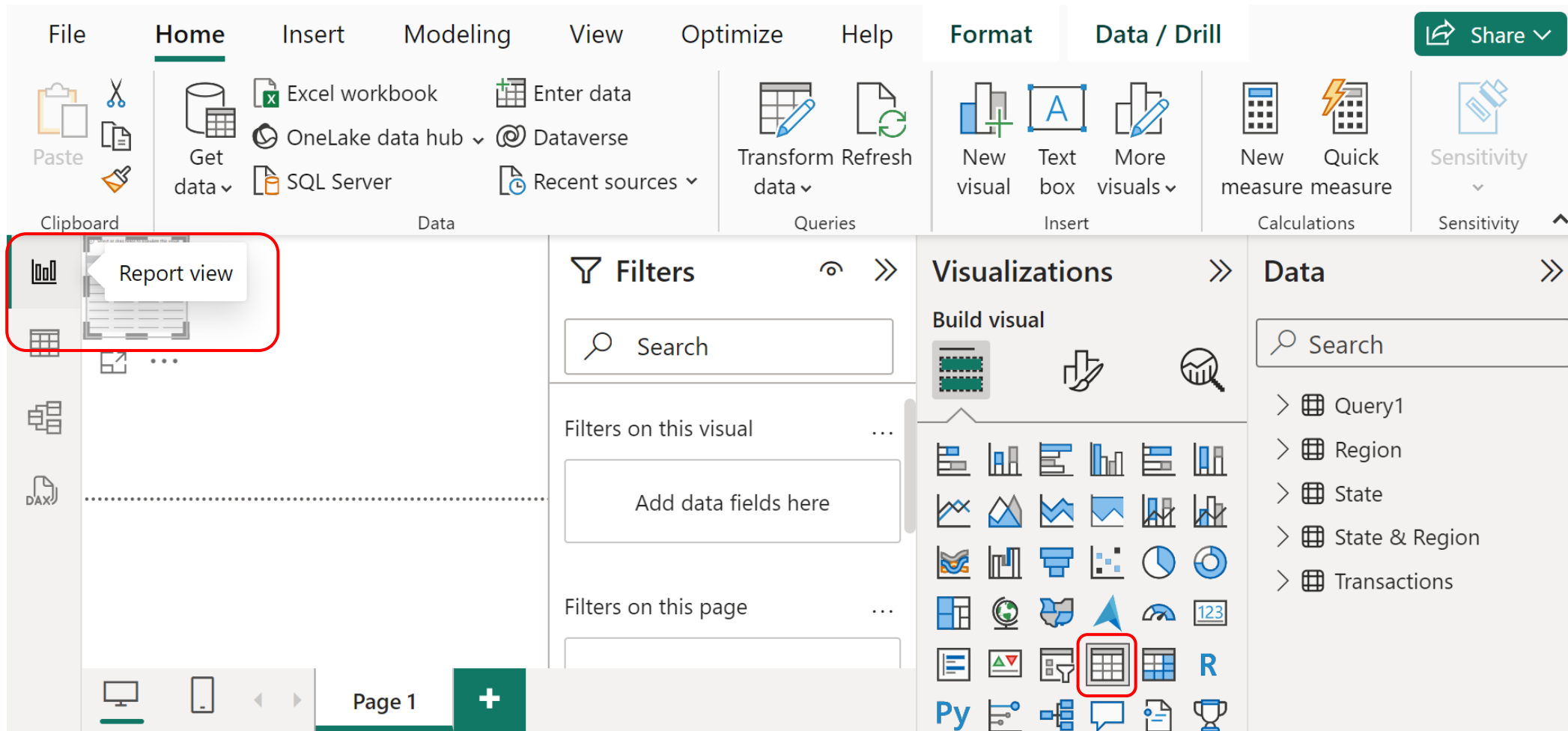
2. Tutorial Week 4 – Visualising Data

- Collapse Visualizations, and Filters.

- Expand Data

- From Region Select Region, then this field will be added to the Table and the data will be displayed

- From State & Region Select State

- From Transactions Select City

- From Transactions Select Profit

2. Tutorial Week 4 – Visualising Data

- The table displays the total *Profit* for each city and the total *Profit* for the company.

City	Sum of Profit
AGAWAM	$50,477.1526
EAU CLAIRE	$2,240.1514
EDDINGTON	$1,207.3952
EDENTON	$396.8537
EDGARD	$413.7355
EDISON	$670.5575
EL CAJON	$869.5025
EL DORADO	$1,623.8375
EL PASO	$390.4113
EL SEGUNDO	$981.5793
ELGIN	$715.7956
ELKHART	$287.6751
ELKINS	$226.431
ELTON	$199.4337
EMERSON	$36.6029
ENCINO	$248.6974
ENFIELD	$1,238.8749
ENGLAND	$676.8745
ENGLEWOOD	$1,383.1743
EUGENE	$2,563.8861
EUREKA	$1,155.4794
EUSTIS	$822.2896
EVANSVILLE	$217.5206
FAIRFAX	$550.1808
FARGO	$849.4144
FENTON	$498.9325
FLINT	$937.3397
FLORA	$315.6726
FLORENCE	$355.8647
FONDA	$499.5603
FOREST PARK	$502.5565
FORREST CITY	$560.0003
FORT COLLINS	$2,137.715
	$603,170.369

2. Tutorial Week 4 – Adding Columns

- Select Transactions

- Select Store

2. Tutorial Week 4 – Sorting

Your report displays the Profit data for each Store sorted by City then by State and then by Region. You can quickly sort your data by either clicking the corresponding column by which you want sort the data by or using the Table context menu. To the City with the highest (maximum) Profit:

- Click on sorting symbol to see the lowest and highest proft.

👍 *Test Your Skills*

- Which Store has the lowest profit?

2. Tutorial Week 4 – Filtering

At the moment the data for all Cities is displayed. There are a number of ways to restrict the values displayed in report through the use of Filters. Filters in a report limit the values displayed on reports, tables, and charts within a document. They do not modify the data that is retrieved from the data source. They simply hide the values not required on the report level. You have realised that the World Center Store in New Orleans is the most profitable. You want to know the Profit for the other World Center Stores throughout the country.

- Click on **Filters**

2. Tutorial Week 4 – Filtering

- Scroll Down to find **Store**

- Select **Advanced filtering**

- Type **World Center**

- Click on **Apply Filter**

- The Table now only displays data for the World Center stores and has calculated a total profit for these stores.

City	Sum of Profit	Store
AGAWAM	$1,614.9771	World Center (WC12)
EAU CLAIRE	$390.0378	World Center (WC12)
EDDINGTON	$403.9623	World Center (WC2)
GARDNER	$15.9751	World Center (WC3)
GARDNER	$14.3749	World Center (WC4)
HUTCHINSON	$371.1276	World Center (WC2)
ITASCA	$12.8573	World Center (WC1)
LITTLE ROCK	$356.4837	World Center (WC8)
MABELVALE	$168.8945	World Center (WC8)
MARRERO	$170.7242	World Center (WC10)
METAIRIE	$170.175	World Center (WC10)
MISSION VIEJO	$0.9839	World Center (WC9)

▽ Filters

Search

contains

World Center

● And ○ Or

Apply filter

2. Tutorial Week 4 – Saving

- Save your report 'Tutor Week 4.pbix' as you need to use it in the next tutorials.

Thank you,
Happy a Learning Day