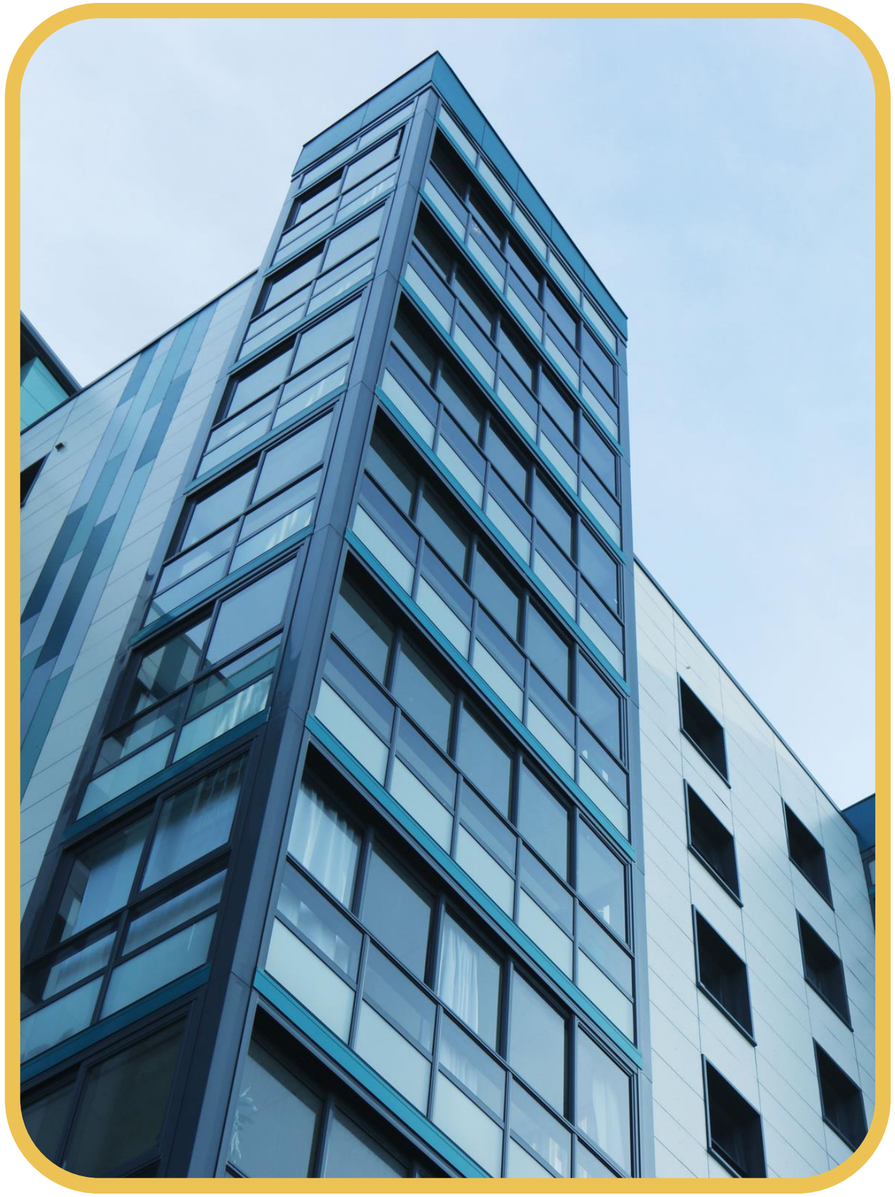
Power BI

Data Cleaning

Learn to clean the data in power BI

Data Preparation

50-60% of overall time



Tools for Data Preparation in Power BI

Power Query Editor

Data Modelling

DAX – Data Analysis Expression

Tool 1

Power Query Editor

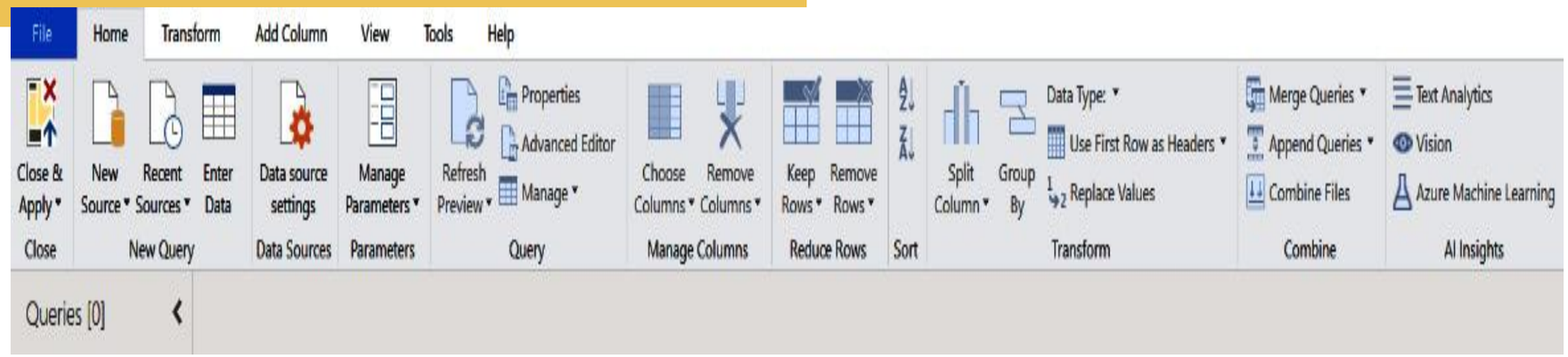
Powerful tool to clean and transform data

Splitting column

Merging tables

Filtering Data

Removing duplicates



Tool 2

Data Modelling

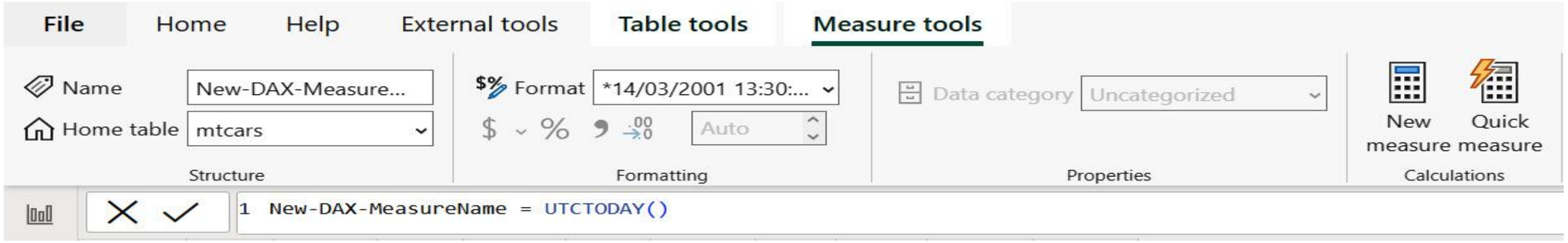
Helps users to build meaningful connections(relationships) between tables and establish hierarchies that provide structured view of their data.

Tool 3

DAX – Data Analysis Expression

Formula language used in power BI to create custom calculations and measures.

Used to created calculated columns, calculated tables and measures.



Understanding Data Quality & Need of Data Cleaning

What is Data Quality?

Where do data quality issues come from?

The role of data cleaning in improving data quality

Best practices for data quality overall

What is Data Quality?

Data Quality is essential for accurate analysis, informed decision making & successful business outcomes.

Factors Make up Data Quality:

The role of data cleaning in improving data quality

Data accuracy – data represents true values & attributes it intended to capture.

Data incompleteness

Data Consistency – uniformity & coherence of data across different sources or datasets.

Data validity – data conforms to defined rules, constraints or standards.

Data timeliness – relevancy & currency of data in relation to analysis or reporting timeframe.

Where do Data Quality issue come from?

Data Quality is essential for accurate analysis, informed decision making & successful business outcomes.

Data Entry errors - typos, misspellings, or incorrect values

Incomplete or missing data – failure in data collection process, data entry omissions or system limitations

Data integration challenges – inconsistency can occur while combining data from different sources.

Data transformation & manipulation – aggregations, calculations or data conversions can introduce errors if not implemented correctly

Where do Data Quality issue come from?

Data storage & transfer – unreliable storage systems may lead to data corruption, loss or unauthorized access.

Data governance & Documentation – inadequate data governance practices, lack of data standards, data definitions, or metadata documentation.

Data Changes & updates - changes in business rules, system updates or modifications to data sources.

External data sources – inaccurate or unreliable data from external sources

The role of data cleaning in improving data quality

Data integrity & accuracy

Decision-making and business outcomes

Data Ownership & accountability

A holistic view of the data ecosystem

Early detection of issues

Continuous improvement and learning

Best Practices for data quality

Establishing data quality standards

Developing data governance framework

Implementing data validation & verification techniques

Standardize data entry processes

Leveraging data cleaning tools & technologies

Collaborating across departments

Continuous data monitoring

Data education & training

The most common Data Cleaning operations

Removing duplicates

Removing missing data

Splitting columns

Merging columns

Replacing outliers

Creating calculated columns versus measures

Who is Responsible for Data Cleaning

Data Steward – overseeing the overall quality & integrity of data, ensures data governance

Data Analyst – hand-on work of cleaning & preparing data for analysis

IT professionals – DBA & system architects – maintaining infrastructure that hosts the data

Business users & Subject Matter expert – identify anomalies with business/domain knwoledge

Data Quality Managers – in large organizations.

Process for Cleaning Data

Data assessment – assess the data before cleaning – (Data stewards & business users)

Data profiling – Exploratory Data Analysis (EDA) – deeper understanding of data – Data Analyst

Data validation – implement techniques to validate data against predefined use cases. – DA

Data Cleaning strategies – designing & implementing approaches to address specific data issues

Data transformation - apply operations & functions to clean and enhance the data

Data Quality assurance – data integrity checks, comparing cleaned & original data for accuracy

Documentation

THANK YOU