

ET in Power BI

1. Introduction

The ET (Extract, Transform) process is essential for preparing and analyzing data. This document details the step-by-step procedure of performing ET in Power BI.

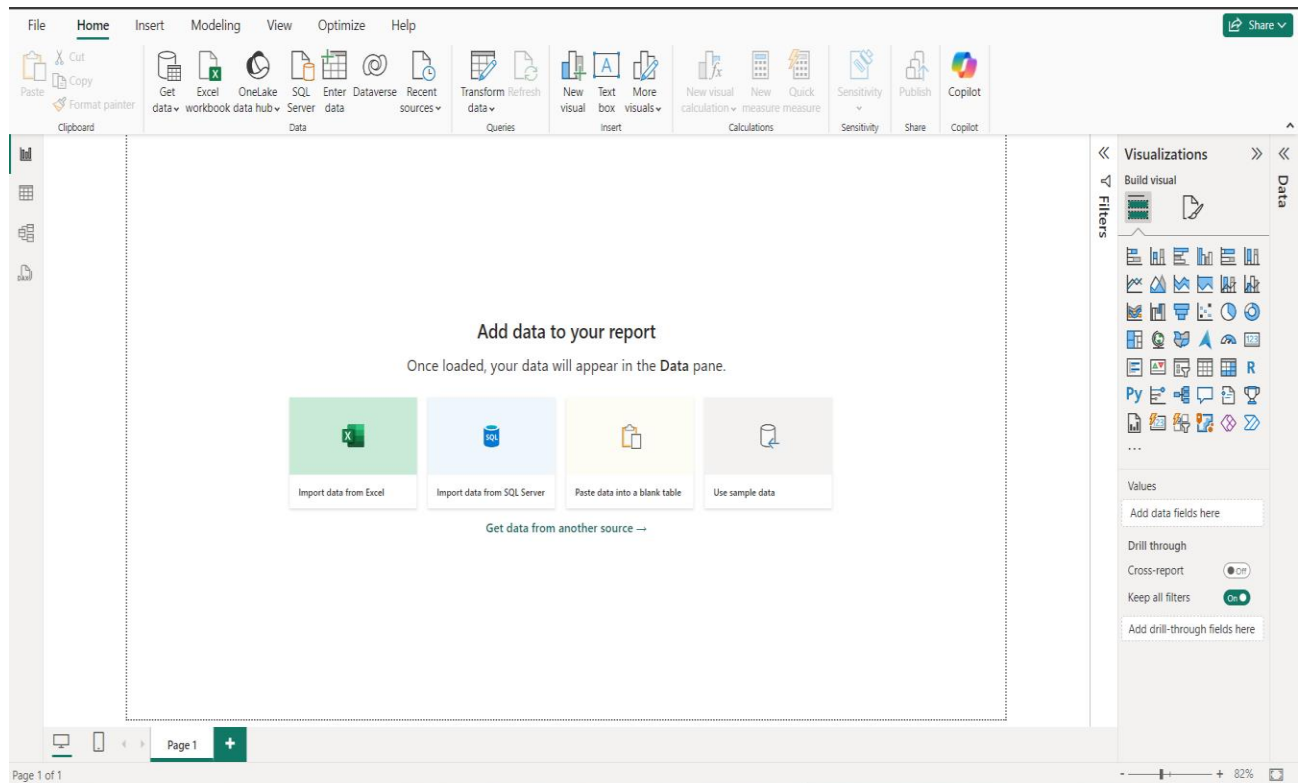
2. Steps Performed

Step 1: Extract

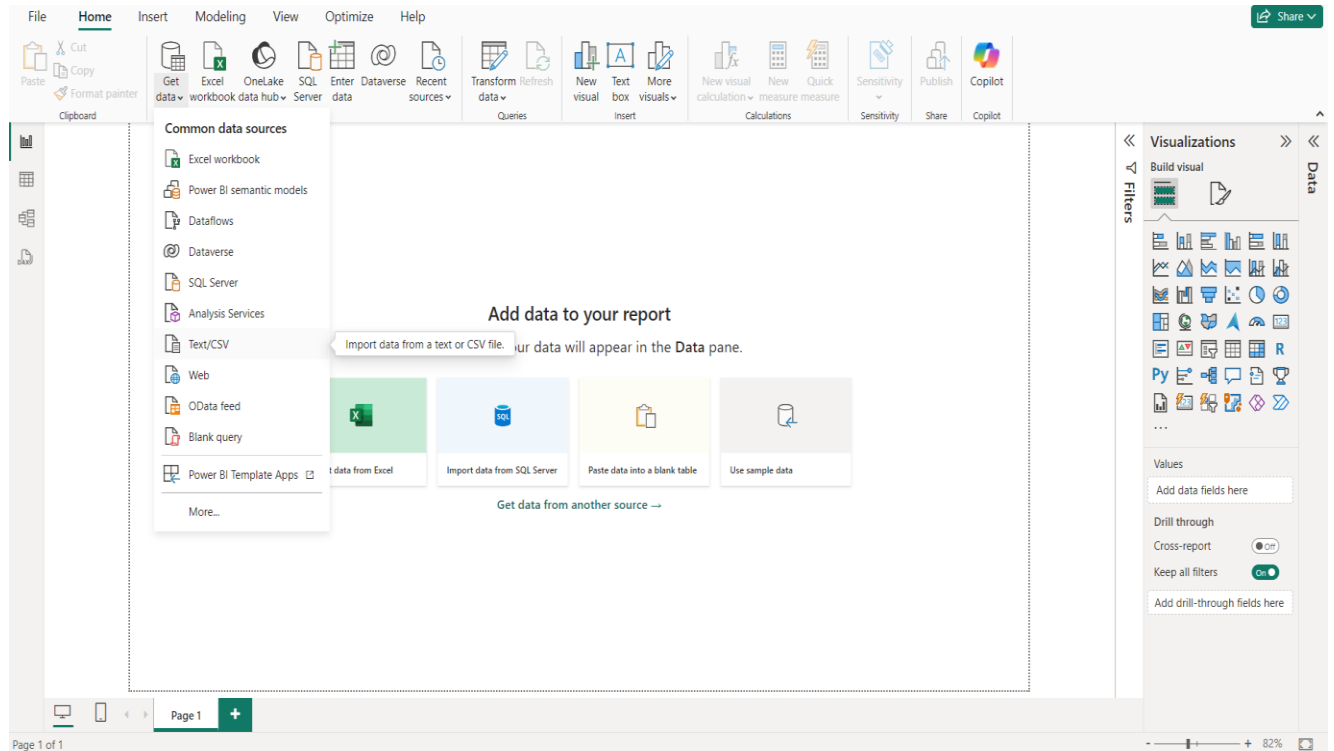
Objective: Pull data from various sources such as Excel, CSV, Text, or a database file.

Procedure:

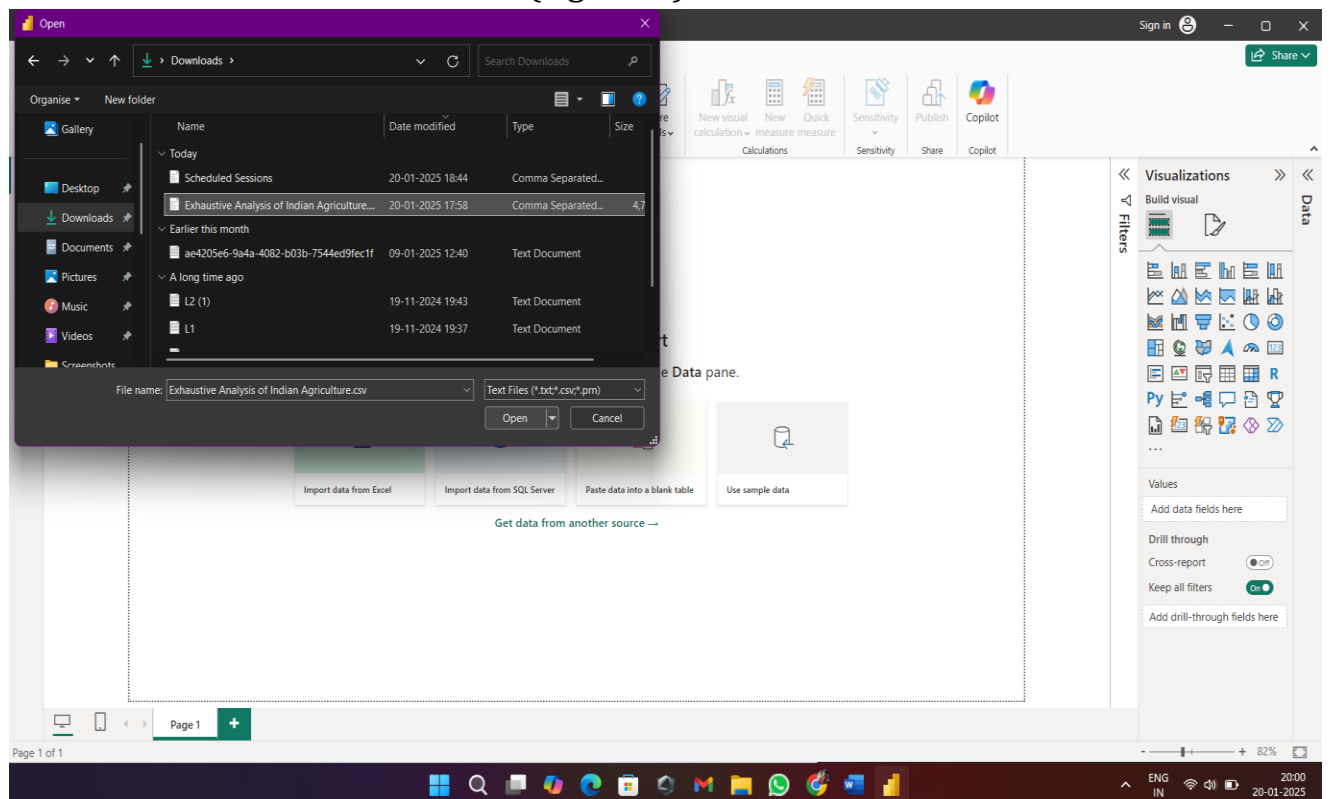
1. Open Power BI Desktop.



2. Navigate to Home > Get Data.



3. Select the desired data source (e.g., CSV).



4. Browse and load the file.

The screenshot shows the Power BI Desktop interface with a file selection dialog box open. The dialog box is titled 'Exhaustive Analysis of Indian Agriculture.csv.csv' and displays a table of agricultural data. The 'Load' button is highlighted in green.

RowID	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production	_1	_2
0	Bihar	NALANDA	2005	Rabi	Wheat	81934	160425		
1	Assam	KARBI ANGLONG	2019	Whole Year	Onion	257	514		
2	Gujarat	ANAND	2020	Summer	Maize	100	100		
3	Karnataka	UTTAR KANNAD	2013	Rabi	Groundnut	2872	4572	45168275000	
4	Uttar Pradesh	JAUNPUR	2016	Rabi	Onion	110	1290		
5	Assam	MARIGAON	2014	Rabi	Rapeseed & Mustard	6535	2719		
6	Odisha	SONEPUR	2006	Winter	Rapeseed & Mustard	91	6		
7	Rajasthan	DHOLPUR	2017	Whole Year	Garlic	1	1		
8	Karnataka	BELGAUM	2018	Whole Year	Coconut	336	3212		
9	Bihar	MUNGER	2020	Summer	Moong(Green Gram)	125	78		
10	Chhattisgarh	JANUGIR-CHAMPA	2013	Kharif	Other Kharif pulses	223	107		
11	Assam	KARBI ANGLONG	2019	Rabi	Rapeseed & Mustard	19337	8652		
12	Uttar Pradesh	SHRAVASTI	2005	Kharif	Groundnut	72	58		
13	Gujarat	PATAN	2019	Kharif	Moong(Green Gram)	9100	3300		
14	Tamil Nadu	KARUR	2008	Whole Year	Sweet potato	20	309		
15	Uttar Pradesh	KASGANI	2019	Rabi	Tobacco	5247	28554		
16	Haryana	MAHENDRAGARH	2006	Rabi	Wheat	45074	186000		
17	Assam	DHEMAJLI	2017	Whole Year	Turmeric	321	211		
18	Assam	BAKSA	2015	Kharif	Small millets	284	127		
19	Kerala	PATHANAMTHITTA	2008	Whole Year	Sugarcane	224	10950		

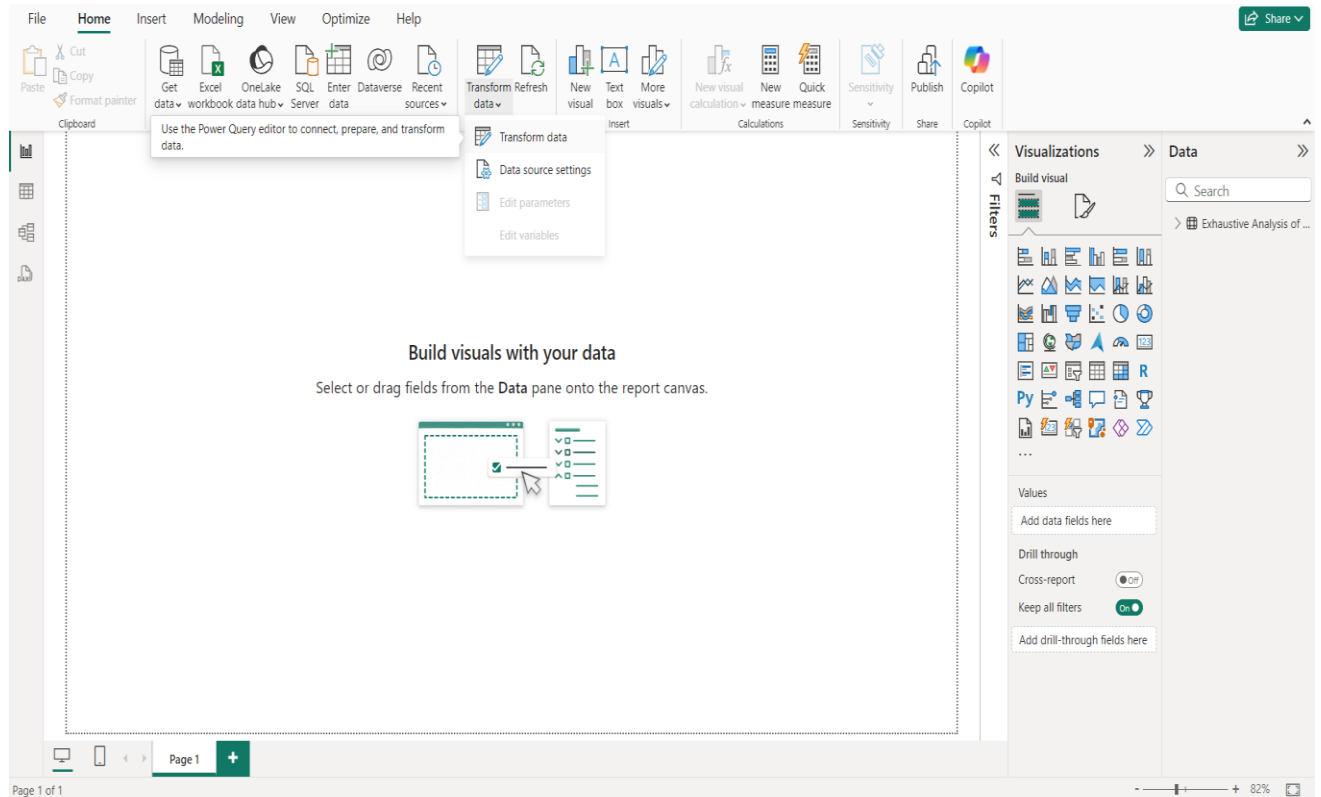
5. Preview the data in the Navigator window and load it.

Step 2: Transform

Objective: Clean and process the data to prepare it for analysis.

Procedure:

1. Open Power Query Editor : Home > Transform Data.



2. Perform the following tasks:

- Remove unnecessary columns.
- Rename columns for clarity.
- Handle missing values (e.g., replace nulls).
- Change data types (e.g., Text, Number, Date).
- Remove duplicates.

By right click on we will get all the transformation tasks

Given below

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

Group By Use First Row as Headers Count Rows

Table

Any Column

Data Type: Text

Replace Values

Unpivot Columns

Move

Convert to List

Split Column

Format

Parse

Merge Columns

Statistics

Standard Scientific

Information

Trigonometry

Rounding

Date

Time

Duration

Run R

Run Python

script

Scripts

Queries [1]

Exhaustive Analysis of In...

Table.TransformColumnTypes(#"Promoted Headers",({{"RowID", Int64.Type}, {"State_Name", type text}, {"District_Name", type text}, {"Area", Int64.Type}, {"Production", Int64.Type}, {"Crop", type text}, {"Season", type text}, {"Yield", Int64.Type}, {"Error", type text}, {"Empty", type text}}))

1 2 Area Production Crop Season Yield Error Empty

1 81934 190425

2 257 514

3 100 100

4 2872 4572

5 110 1290

6 rd 6535 2719

7 rd 91 6

8 1 1

9 336 3212

10 m) 125 78

11 s 223 107

12 rd 19337 8652

13 72 58

14 m) 9100 3300

15 20 309

16 5247 28554

17 45074 186000

18 321 211

19 284 127

20 224 10950

21 2497 658

22 46 1520

23 3418 7487

24 5806 7964

25 11 13

26 14 5

27 780 456

28

12 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 2001

Query Settings

PROPERTIES

Name

Exhaustive Analysis of Indian Agriculture

All Properties

APPLIED STEPS

Source

Promoted Headers

Changed Type

3. After cleaning data we can view the quality of data :

View > Column quality

Untitled - Power Query Editor

File Home Transform Add Column View Tools Help

Formula Bar

Monospaced

Column distribution

Show whitespace

Column profile

Column quality

Layout

Data Preview

Columns

Parameters

Advanced

Dependencies

Advanced Editor

Query Dependencies

Queries [1]

Exhaustive Analysis of In...

Table.RemoveColumns(#"Changed Type",{"", "_1", "_2", "_3"})

RowID State_Name District_Name Crop_Year Season Crop Area

Valid 100% Valid 100% Valid 100% Valid 100% Valid 100% Valid 100%

Error 0% Error 0% Error 0% Error 0% Error 0% Error 0%

Empty 0% Empty 0% Empty 0% Empty 0% Empty 0%

RowID 1000 (100%)

Valid 0 (0%)

Error 0 (0%)

Empty

1 NALANDA 2005 Rabi Wheat

2 KARBI ANGLONG 2019 Whole Year Onion

3 ANAND 2020 Summer Maize

4 UTTAR KANNAD 2013 Rabi Groundnut

5 JAUNPUR 2016 Rabi Onion

6 MARIAGAON 2014 Rabi Rapeseed & Mustard

7 SONEPUR 2006 Winter Rapeseed & Mustard

8 DHOLPUR 2017 Whole Year Garlic

9 BELGAUM 2018 Whole Year Coconut

10 MUNGER 2020 Summer Moong(Green Gram)

11 JANIGIR-CHAMPA 2013 Kharif Other Kharif pulses

12 KARBI ANGLONG 2019 Rabi Rapeseed & Mustard

13 SHRIVASTI 2005 Kharif Groundnut

14 PATAN 2019 Kharif Moong(Green Gram)

15 KARUR 2008 Whole Year Sweet potato

16 KASGANI 2019 Rabi Tobacco

17 MAHENDRAGARH 2006 Rabi Wheat

18 DHEMAJI 2017 Whole Year Turmeric

19 BAKSA 2015 Kharif Small millets

20 PATHANAMTHITTA 2008 Whole Year Sugarcane

21 JANIGIR-CHAMPA 2018 Rabi Linseed

22 DHAMTARI 2020 Whole Year Banana

23 BELLARY 2016 Rabi Maize

24 TINSUKIA 2020 Autumn Rice

25

8 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 2001

Query Settings

PROPERTIES

Name

Exhaustive Analysis of Indian Agriculture

All Properties

APPLIED STEPS

Source

Promoted Headers

Changed Type

Removed Columns

4. To save all the changes we will:

Home > Close and Apply

The screenshot displays the Power Query Editor window. The main area shows a table with the following columns: RowID, State, District, Crop Year, Season, Crop, and Area. The table contains 25 rows of data, including entries for various Indian states like Bihar, Assam, Gujarat, Karnataka, Uttar Pradesh, Odisha, Rajasthan, Karnataka, Chhattisgarh, Assam, Gujarat, Tamil Nadu, Uttar Pradesh, Haryana, Assam, Assam, Kerala, Chhattisgarh, Chhattisgarh, Karnataka, and Assam. The 'Applied Steps' pane on the right shows a list of steps: 'Source', 'Promoted Headers', 'Changed Type', and 'Removed Columns'. The status bar at the bottom indicates '8 COLUMNS, 999+ ROWS' and 'Column profiling based on top 1000 rows'.

RowID	State	District	Crop Year	Season	Crop	Area
1	Bihar	NALANDA	2005	Rabi	Wheat	
2	Assam	KARBI ANGLONG	2019	Whole Year	Onion	
3	Gujarat	ANAND	2020	Summer	Maize	
4	Karnataka	UTTAR KANNAD	2013	Rabi	Groundnut	
5	Uttar Pradesh	JALNPUR	2016	Rabi	Onion	
6	Assam	MARIGAON	2014	Rabi	Rapeseed & Mustard	
7	Odisha	SONEPUR	2006	Winter	Rapeseed & Mustard	
8	Rajasthan	DHOLPUR	2017	Whole Year	Garlic	
9	Karnataka	BELGAUM	2018	Whole Year	Coconut	
10	Bihar	MUNGER	2020	Summer	Moong(Green Gram)	
11	Chhattisgarh	JANIGIR-CHAMPA	2013	Kharif	Other Kharif pulses	
12	Assam	KARBI ANGLONG	2019	Rabi	Rapeseed & Mustard	
13	Uttar Pradesh	SHRAVASTI	2005	Kharif	Groundnut	
14	Gujarat	PATAN	2019	Kharif	Moong(Green Gram)	
15	Tamil Nadu	KARUR	2008	Whole Year	Sweet potato	
16	Uttar Pradesh	KASGANI	2019	Rabi	Tobacco	
17	Haryana	MAHENDRAGARH	2006	Rabi	Wheat	
18	Assam	DHEMAJI	2017	Whole Year	Turmeric	
19	Assam	BAKSA	2015	Kharif	Small millets	
20	Kerala	PATHANAMTHITTA	2008	Whole Year	Sugarcane	
21	Chhattisgarh	JANIGIR-CHAMPA	2018	Rabi	Linseed	
22	Chhattisgarh	DHAMTARI	2020	Whole Year	Banana	
23	Karnataka	BELLARY	2016	Rabi	Maize	
24	Assam	TINSUKIA	2020	Autumn	Rice	
25	Uttar Pradesh	BALABHAR	2017	Whole Year	Onion	