Week 1 Deliverables:

Internship on Exhaustive Analysis of Indian Agriculture Sector Using Power BI

Overview of Week 1 Learning

The primary aim of this internship is to create an exhaustive analysis dashboard for the Indian agriculture sector using Power BI. In the first mentoring session, we explored the foundational features of Power BI and performed basic data operations step by step. Below is a summary of the key concepts and tasks covered:

Key Concepts Learned:

1. **Power BI Interface Overview**:

   o We started with a blank Power BI report interface.

   o Explored the visualization pane on the right, which contains various chart types for creating visualizations.

   o Understood the functionality of the home tab for accessing core options like loading datasets.

2. **ETL Process in Power BI**:

   o **Extract**: Pulling data from sources like CSV files. This involves identifying the correct dataset and ensuring it is in a compatible format for Power BI.

   o **Transform**: Preprocessing data, including cleaning, handling missing values, and removing unnecessary information. During this step:

      ▪ Unnecessary columns, such as identifiers or metadata not useful for analysis, were removed.

      ▪ Blank columns and those deemed irrelevant to the dashboard objectives were also eliminated.

      ▪ Data quality issues, such as null values or duplicate entries, were addressed to ensure consistency.

- o **Load**: Loading cleaned and processed data into Power BI for further analysis and visualization.

3. **Data Views in Power BI**:

    - o **Report View**: For creating and interacting with visualizations.

    - o **Table View**: For reviewing raw data and understanding its structure.

    - o **Model View**: For building relationships among datasets to enable advanced analytics.

4. **Power Query Editor**:

    - o Accessed via the Transform Data button in the Home tab.

    - o Used for data cleaning, removing duplicate or null values, and transforming data types.

Tasks Performed in Week 1:

1. **Loading Data**:

    - o Opened a CSV file using the "Get Data" option from the Home tab.

    - o Selected the CSV file and explored options to Load or Transform the data.

2. **Data Transformation**:

    - o Opened Power Query Editor to preprocess the dataset.

    - o Removed unnecessary columns, including those with blank values or data irrelevant to the analysis objectives, by selecting and clicking "Remove."

    - o Checked for null values using the "Column Quality" feature under the View tab.

    - o Handled issues like duplicates, blank rows, and errors using options like Remove Duplicates and Remove Errors.

    - o Adjusted data types for columns to ensure consistency and compatibility with Power BI visualizations.

    - o Ensured that the cleaned data was structured properly for analysis by reorganizing or renaming columns as needed.
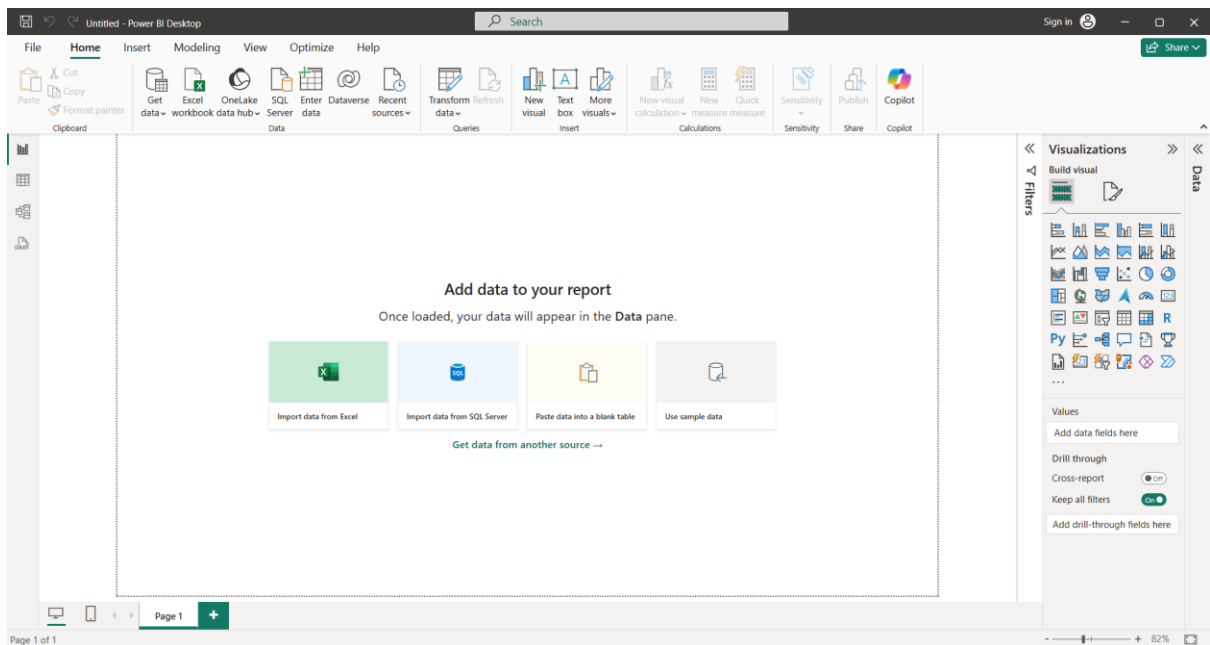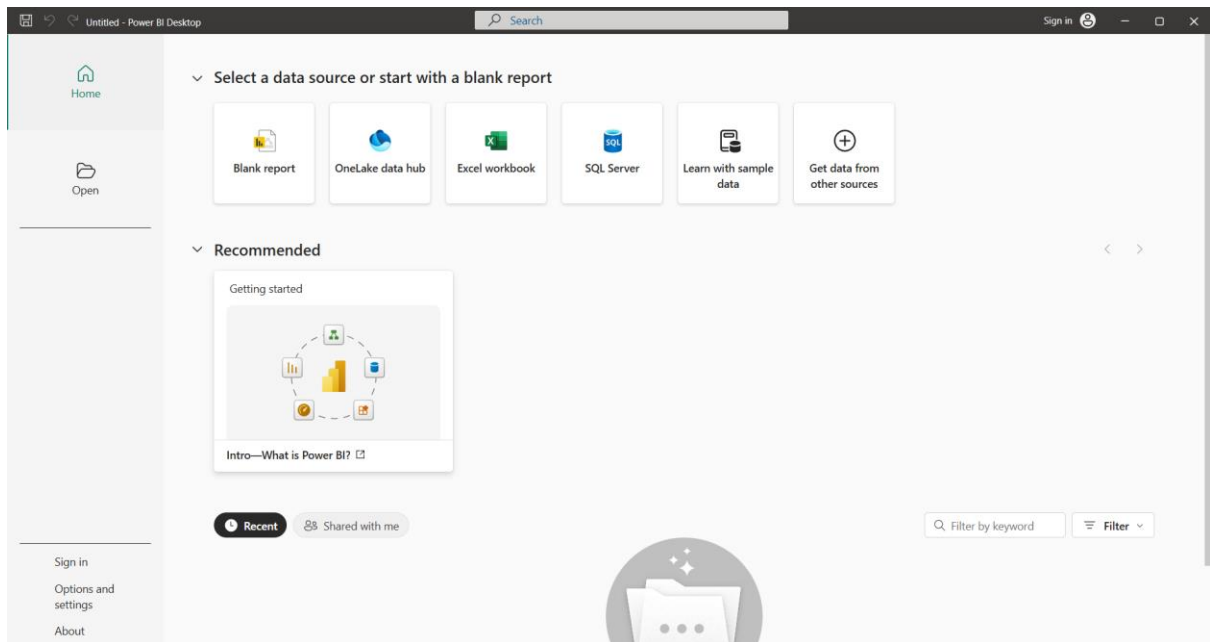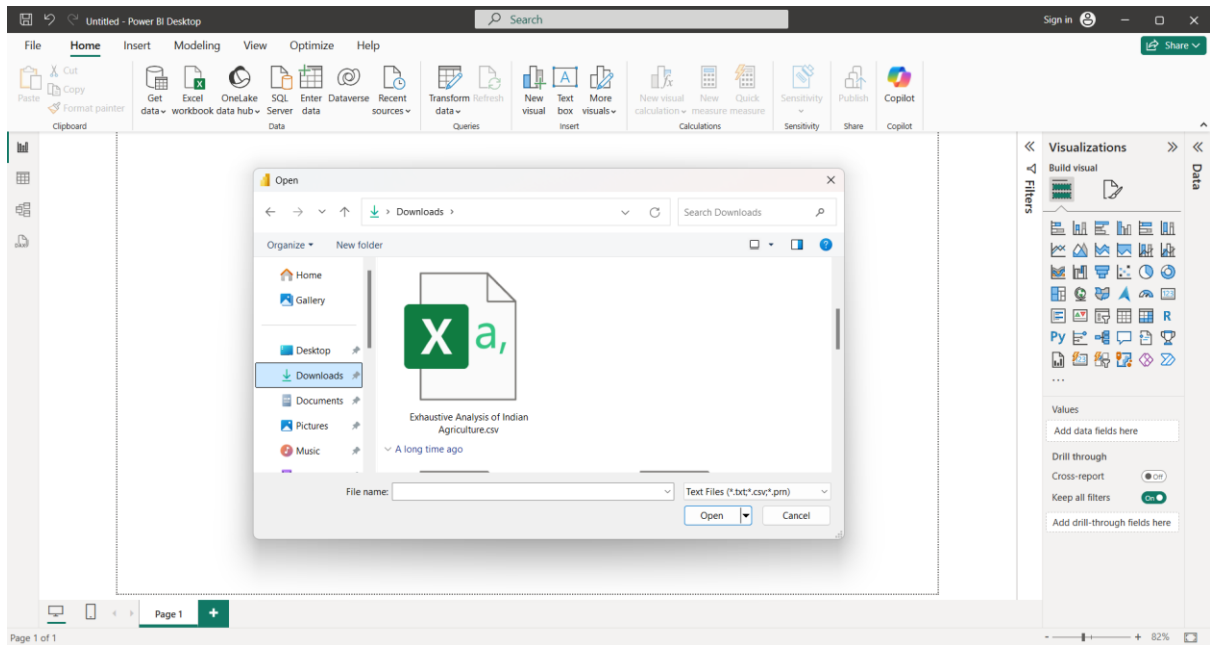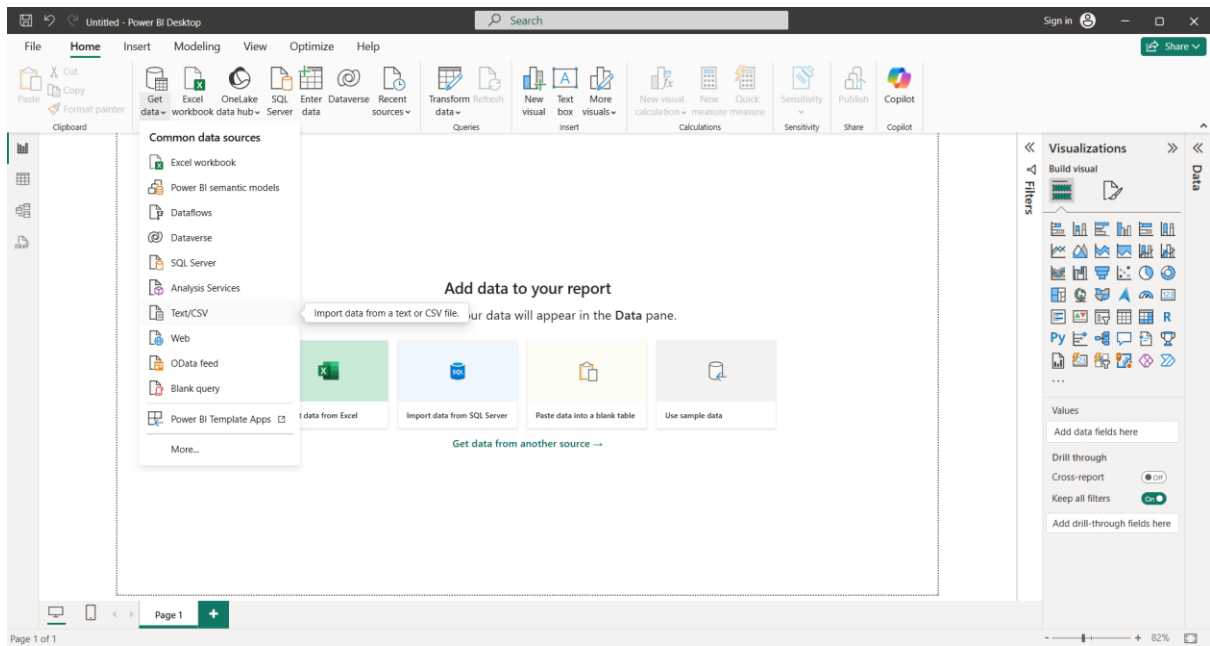
3. **Saving the Transformed Data**:

   o After completing the data cleaning process, saved the changes and loaded the transformed data into Power BI for visualization.

Screenshots:

Include screenshots of the following tasks performed:

1. Loading a CSV file in Power BI.

2. Power Query Editor interface showing data cleaning steps.

3. Column Quality feature in Power Query Editor.

4. Final dataset after transformation.

File   Home   Insert   Modeling   View   Optimize   Help

Exhaustive Analysis of Indian Agriculture.csv

File Origin: 1252: Western European (Windows)   Delimiter: Comma   Data Type Detection: Based on first 200 rows

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	Total production	A
3	Karnataka	UTTAR KANNAD	2013	Rabi	Groundnut	2872	4572	45168275000	8
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	JANJGIR-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	KASGANJ	2019	Rabi	Tobacco	5247	28554		
16	Haryana	MAHENDRAGARH	2006	Rabi	Wheat	45074	186000		
17	Assam	DHEMAJI	2017	Whole Year	Turmeric	321	211		
18	Assam	BAKSA	2015	Kharif	Small millets	284	127		
19	Kerala	PATHANAMTHITTA	2008	Whole Year	Sugarcane	224	10950		

Extract Table Using Examples   Load   Transform Data   Cancel

Visualizations — Build visual

Values — Add data fields here

Drill through
Cross-report — Off
Keep all filters — On
Add drill-through fields here

Page 1 of 1   Page 1   82%

File   Home   Insert   Modeling   View   Optimize   Help

Get data | Excel workbook | OneLake data hub | SQL Server | Enter data | Dataverse | Recent sources | Transform data | Refresh | New visual | Text box | More visuals | New visual calculation | New measure | Quick measure | Sensitivity | Publish | Copilot

⚠ There are pending changes in your queries that haven't been applied.   Apply changes   Discard changes

Add
Once loaded, yo

Load
⟳ Exhaustive Analysis of Indian Agriculture
Creating connection in model...
                                    Cancel

Import data from Excel   Import data from SQL Server   Paste data into a blank table   Use sample data
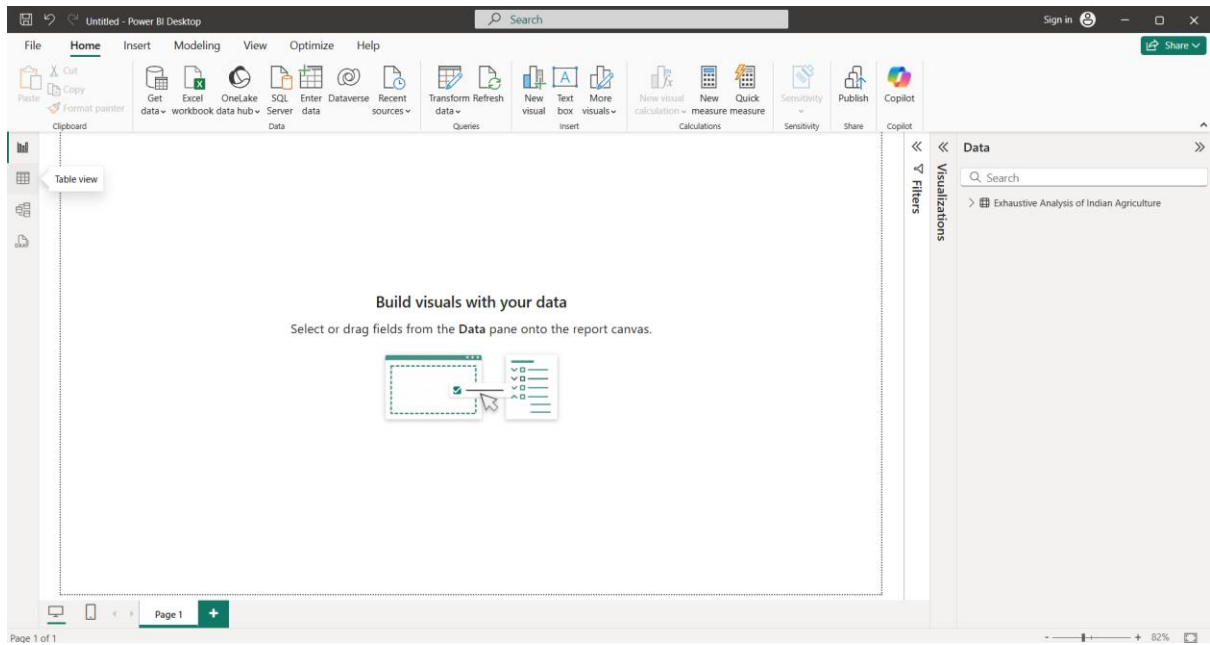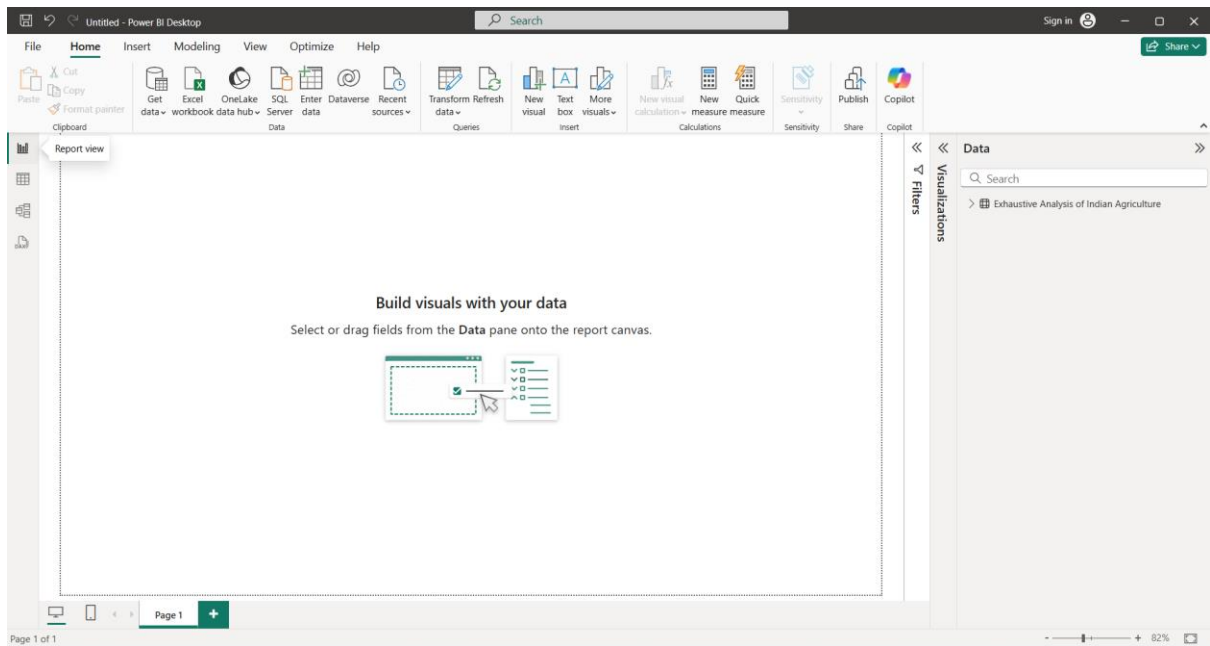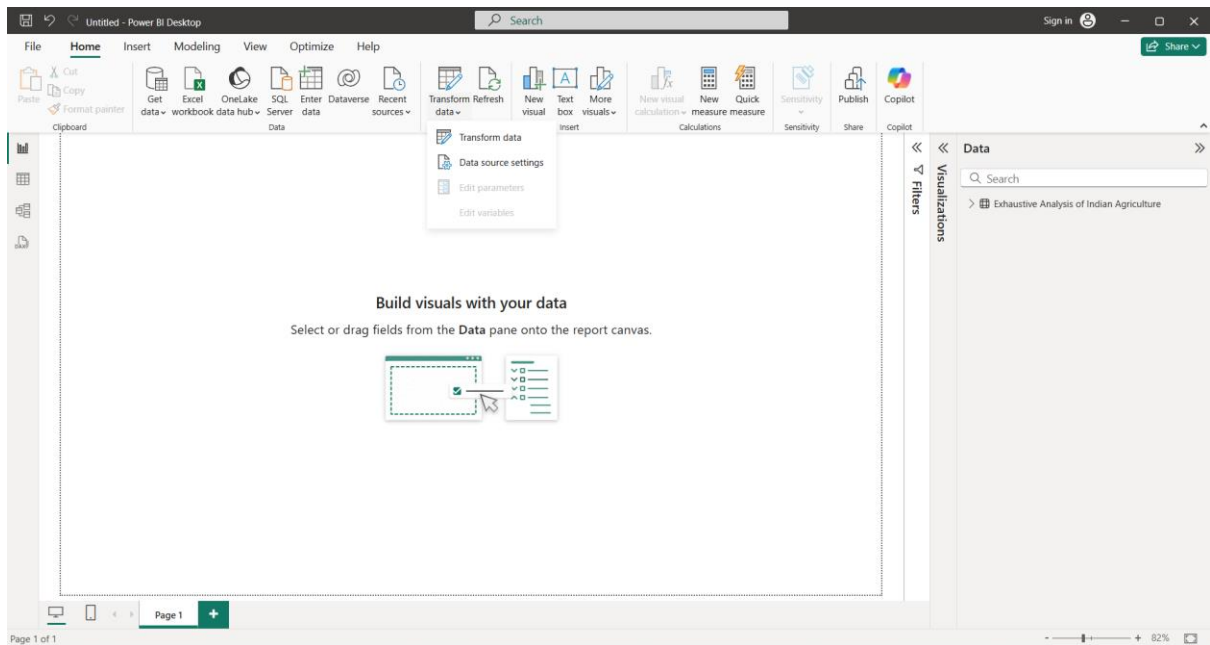
Get data from another source →

Visualizations — Build visual

Data — Search

Values — Add data fields here

Drill through
Cross-report — Off
Keep all filters — On
Add drill-through fields here

Page 1 of 1   Page 1   82%

Screenshot 1
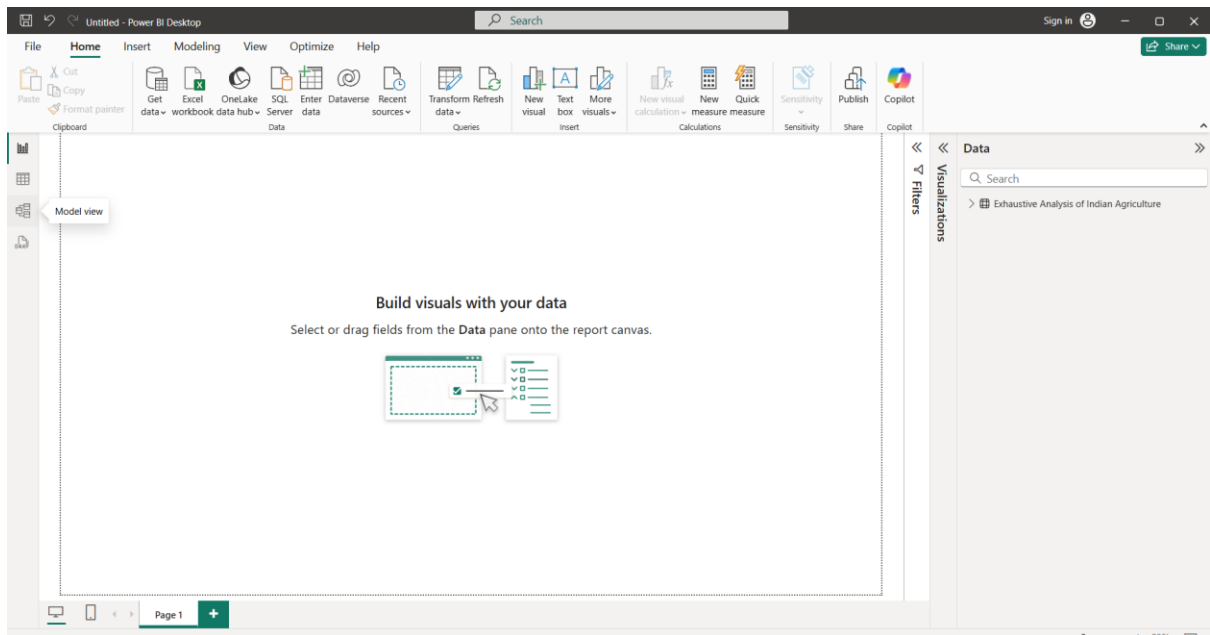
File | Home | Transform | Add Column | View | Tools | Help

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: Whole Number / Use First Row as Headers / Replace Values | Merge Queries / Append Queries / Combine Files | Text Analytics / Vision / Azure Machine Learning

Queries [1] — Exhaustive Analysis of In...

`= Table.TransformColumnTypes(#"Promoted Headers",{{"RowID", Int64.Type}, {"State_Name", type text}, {"District_Name", type text},`

	RowID	State_Name	District_Name	Crop_Year	Season	Crop	Area
1	0	Bihar	NALANDA	2005	Rabi	Wheat	
2	1	Assam	KARBI ANGLONG	2019	Whole Year	Onion	
3	2	Gujarat	ANAND	2020	Summer	Maize	
4	3	Karnataka	UTTAR KANNAD	2013	Rabi	Groundnut	
5	4	Uttar Pradesh	JAUNPUR	2016	Rabi	Onion	
6	5	Assam	MARIGAON	2014	Rabi	Rapeseed &Mustard	
7	6	Odisha	SONEPUR	2006	Winter	Rapeseed &Mustard	
8	7	Rajasthan	DHOLPUR	2017	Whole Year	Garlic	
9	8	Karnataka	BELGAUM	2018	Whole Year	Coconut	
10	9	Bihar	MUNGER	2020	Summer	Moong(Green Gram)	
11	10	Chhattisgarh	JANJGIR-CHAMPA	2013	Kharif	Other Kharif pulses	
12	11	Assam	KARBI ANGLONG	2019	Rabi	Rapeseed &Mustard	
13	12	Uttar Pradesh	SHRAVASTI	2005	Kharif	Groundnut	
14	13	Gujarat	PATAN	2019	Kharif	Moong(Green Gram)	
15	14	Tamil Nadu	KARUR	2008	Whole Year	Sweet potato	
16	15	Uttar Pradesh	KASGANJ	2019	Rabi	Tobacco	
17	16	Haryana	MAHENDRAGARH	2006	Rabi	Wheat	
18	17	Assam	DHEMAJI	2017	Whole Year	Turmeric	
19	18	Assam	BAKSA	2015	Kharif	Small millets	
20	19	Kerala	PATHANAMTHITTA	2008	Whole Year	Sugarcane	
21	20	Chhattisgarh	JANJGIR-CHAMPA	2018	Rabi	Linseed	
22	21	Chhattisgarh	DHAMTARI	2020	Whole Year	Banana	
23	22	Karnataka	BELLARY	2016	Rabi	Maize	
24	23	Assam	TINSUKIA	2020	Autumn	Rice	
25	24	Uttar Pradesh	RAE BARELI	2016	Kharif	Maize	
26	25	Karnataka	KOLAR	2020	Rabi	Cowpea(Lobia)	
27	26	Madhya Pradesh	NEEMUCH	2019	Whole Year	Sesamum	

PROPERTIES
Name: Exhaustive Analysis of Indian Agriculture
All Properties

APPLIED STEPS
Source
Promoted Headers
Changed Type

12 COLUMNS, 999+ ROWS — Column profiling based on top 1000 rows — PREVIEW DOWNLOADED AT 06:56 PM

Screenshot 2

`= Table.TransformColumnTypes(#"Promoted Headers",{{"RowID", Int64.Type}, {"State_Name", type text}, {"District_Name", type text},`

Context menu:
- Copy
- Remove
- Remove Other Columns
- Duplicate Column
- Add Column From Examples...
- Remove Duplicates
- Remove Errors
- Change Type
- Transform
- Replace Values...
- Replace Errors...
- Group By...
- Fill
- Unpivot Columns
- Unpivot Other Columns
- Unpivot Only Selected Columns
- Rename...
- Move
- Drill Down
- Add as New Query

12 COLUMNS, 999+ ROWS — Column profiling based on top 1000 rows — PREVIEW DOWNLOADED AT 06:56 PM

Top window - Power Query Editor

Formula: `= Table.RemoveColumns(#"Changed Type",{"RowID"})`

	1.2 Area	1.2 Production	ABC _1	ABC _2	ABC _3
1	81934	160425			
2	257	514			
3	100	100			Average
4	2872	4572			89488.975
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			

Context menu:
Copy / Remove / Remove Other Columns / Duplicate Column / Add Column From Examples... / Remove Duplicates / Remove Errors / Change Type / Transform / Replace Values... / Replace Errors... / Split Column / Group By... / Fill / Unpivot Columns / Unpivot Other Columns / Unpivot Only Selected Columns / Rename... / Move / Drill Down / Add as New Query

Query Settings — PROPERTIES — Name: Exhaustive Analysis of Indian Agriculture — All Properties — APPLIED STEPS: Source / Promoted Headers / Changed Type / Removed Columns

11 COLUMNS, 999+ ROWS   Column profiling based on top 1000 rows   PREVIEW DOWNLOADED AT 07:06 PM

Bottom window - Power Query Editor

Formula: `= Table.RemoveColumns(#"Changed Type",{"RowID", "", "_1", "_2", "_3"})`

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
1	Bihar	NALANDA	2005	Rabi	Wheat	81934	
2	Assam	KARBI ANGLONG	2019	Whole Year	Onion	257	
3	Gujarat	ANAND	2020	Summer	Maize	100	
4	Karnataka	UTTAR KANNAD	2013	Rabi	Groundnut	2872	
5	Uttar Pradesh	JAUNPUR	2016	Rabi	Onion	110	
6	Assam	MARIGAON	2014	Rabi	Rapeseed &Mustard	6535	
7	Odisha	SONEPUR	2006	Winter	Rapeseed &Mustard	91	
8	Rajasthan	DHOLPUR	2017	Whole Year	Garlic	1	
9	Karnataka	BELGAUM	2018	Whole Year	Coconut	336	
10	Bihar	MUNGER	2020	Summer	Moong(Green Gram)	125	
11	Chhattisgarh	JANJGIR-CHAMPA	2013	Kharif	Other Kharif pulses	223	
12	Assam	KARBI ANGLONG	2019	Rabi	Rapeseed &Mustard	19337	
13	Uttar Pradesh	SHRAVASTI	2005	Rabi	Groundnut	72	
14	Gujarat	PATAN	2019	Kharif	Moong(Green Gram)	9100	
15	Tamil Nadu	KARUR	2008	Whole Year	Sweet potato	20	
16	Uttar Pradesh	KASGANJ	2019	Rabi	Tobacco	5247	
17	Haryana	MAHENDRAGARH	2006	Rabi	Wheat	45074	
18	Assam	DHEMAJI	2017	Whole Year	Turmeric	321	
19	Assam	BAKSA	2015	Kharif	Small millets	284	
20	Kerala	PATHANAMTHITTA	2008	Whole Year	Sugarcane	224	
21	Chhattisgarh	JANJGIR-CHAMPA	2018	Rabi	Linseed	2497	
22	Chhattisgarh	DHAMTARI	2020	Whole Year	Banana	46	
23	Karnataka	BELLARY	2016	Rabi	Maize	3418	
24	Assam	TINSUKIA	2020	Autumn	Rice	5806	
25	Uttar Pradesh	RAE BARELI	2016	Kharif	Maize	11	
26	Karnataka	KOLAR	2020	Rabi	Cowpea(Lobia)	14	
27	Madhya Pradesh	NEEMUCH	2019	Whole Year	Sesamum	780	

Query Settings — PROPERTIES — Name: Exhaustive Analysis of Indian Agriculture — All Properties — APPLIED STEPS: Source / Promoted Headers / Changed Type / Removed Columns

7 COLUMNS, 999+ ROWS   Column profiling based on top 1000 rows   PREVIEW DOWNLOADED AT 07:06 PM

Top window — Power Query Editor

Formula bar: `= Table.RemoveColumns(#"Changed Type",{"RowID", "", "_1", "_2", "_3"})`

Queries [1]: Exhaustive Analysis of In...

Query Settings — PROPERTIES
Name: Exhaustive Analysis of Indian Agriculture
All Properties

APPLIED STEPS
- Source
- Promoted Headers
- Changed Type
- Removed Columns

State_Name tooltip: 1000 (100%) Valid | 0 (0%) Error | 0 (0%) Empty

#	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
			2005	Rabi	Wheat		81934
		ANGLONG	2019	Whole Year	Onion		257
			2020	Summer	Maize		100
		KANNAD	2013	Rabi	Groundnut		2872
			2016	Rabi	Onion		110
6	Assam	MARIGAON	2014	Rabi	Rapeseed &Mustard		6535
7	Odisha	SONEPUR	2006	Winter	Rapeseed &Mustard		91
8	Rajasthan	DHOLPUR	2017	Whole Year	Garlic		1
9	Karnataka	BELGAUM	2018	Whole Year	Coconut		336
10	Bihar	MUNGER	2020	Summer	Moong(Green Gram)		125
11	Chhattisgarh	JANJGIR-CHAMPA	2013	Kharif	Other Kharif pulses		223
12	Assam	KARBI ANGLONG	2019	Rabi	Rapeseed &Mustard		19337
13	Uttar Pradesh	SHRAVASTI	2005	Kharif	Groundnut		72
14	Gujarat	PATAN	2019	Kharif	Moong(Green Gram)		9100
15	Tamil Nadu	KARUR	2008	Whole Year	Sweet potato		20
16	Uttar Pradesh	KASGANJ	2019	Rabi	Tobacco		5247
17	Haryana	MAHENDRAGARH	2006	Rabi	Wheat		45074
18	Assam	DHEMAJI	2017	Whole Year	Turmeric		321
19	Assam	BAKSA	2015	Kharif	Small millets		284
20	Kerala	PATHANAMTHITTA	2008	Whole Year	Sugarcane		224
21	Chhattisgarh	JANJGIR-CHAMPA	2018	Rabi	Linseed		2497
22	Chhattisgarh	DHAMTARI	2020	Whole Year	Banana		46
23	Karnataka	BELLARY	2016	Rabi	Maize		3418
24	Assam	TINSUKIA	2020	Autumn	Rice		5806
25			2016	Kharif	Maize		

7 COLUMNS, 999+ ROWS   Column profiling based on top 1000 rows   PREVIEW DOWNLOADED AT 07:06 PM

Bottom window — Power Query Editor

Formula bar: `= Table.RemoveRowsWithErrors(#"Removed Columns", {"S...`

Remove Rows menu:
- Remove Top Rows
- Remove Bottom Rows
- Remove Alternate Rows
- Remove Duplicates
- Remove Blank Rows
- Remove Errors

APPLIED STEPS
- Source
- Promoted Headers
- Changed Type
- Removed Columns
- Removed Errors

#	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
1	Bihar	NALANDA	2005	Rabi	Wheat		81934
2	Assam	KARBI ANGLONG	2019	Whole Year	Onion		257
3	Gujarat	ANAND	2020	Summer	Maize		100
4	Karnataka	UTTAR KANNAD	2013	Rabi	Groundnut		2872
5	Uttar Pradesh	JAUNPUR	2016	Rabi	Onion		110
6	Assam	MARIGAON	2014	Rabi	Rapeseed &Mustard		6535
7	Odisha	SONEPUR	2006	Winter	Rapeseed &Mustard		91
8	Rajasthan	DHOLPUR	2017	Whole Year	Garlic		1
9	Karnataka	BELGAUM	2018	Whole Year	Coconut		336
10	Bihar	MUNGER	2020	Summer	Moong(Green Gram)		125
11	Chhattisgarh	JANJGIR-CHAMPA	2013	Kharif	Other Kharif pulses		223
12	Assam	KARBI ANGLONG	2019	Rabi	Rapeseed &Mustard		19337
13	Uttar Pradesh	SHRAVASTI	2005	Kharif	Groundnut		72
14	Gujarat	PATAN	2019	Kharif	Moong(Green Gram)		9100
15	Tamil Nadu	KARUR	2008	Whole Year	Sweet potato		20
16	Uttar Pradesh	KASGANJ	2019	Rabi	Tobacco		5247
17	Haryana	MAHENDRAGARH	2006	Rabi	Wheat		45074
18	Assam	DHEMAJI	2017	Whole Year	Turmeric		321
19	Assam	BAKSA	2015	Kharif	Small millets		284
20	Kerala	PATHANAMTHITTA	2008	Whole Year	Sugarcane		224
21	Chhattisgarh	JANJGIR-CHAMPA	2018	Rabi	Linseed		2497
22	Chhattisgarh	DHAMTARI	2020	Whole Year	Banana		46
23	Karnataka	BELLARY	2016	Rabi	Maize		3418
24	Assam	TINSUKIA	2020	Autumn	Rice		5806
25			2016	Kharif	Maize		

7 COLUMNS, 999+ ROWS   Column profiling based on top 1000 rows   PREVIEW DOWNLOADED AT 07:08 PM

ETL : Extract - PULL DATA from data source: excel , csv , text , databse file
Transform - data processing , data cleaning
LOAD : for analysis

analysis : keep facts , insights via dashboards

load : when your data is cleaned
transformed : when you want to process the data.

report view : visualization
table view : see the data
model view : so that we can create relationship among data.

Conclusion: In the first week, we gained a solid understanding of the Power BI interface and the essential steps of the ETL process. These foundational skills will be instrumental as we progress toward creating the dashboard for the Indian agriculture sector.