

Low Level Design

Expenditure Data Analysis

Written By	Ujjawal Jani
Document Version	1.1
Last Revised Date	

Contents

1 Introduction.....	03
1.1 What is Low-Level Design Document?	03
1.2 Scope	03
2 Architecture	04
3 Architecture Description	06
3.1 Data Description	06
3.2 Web Scrapping	07
3.3 Data Preparation	08
3.4 Data cleaning	13
3.5 Connection with Tableau	14
3.6 Deployment	14

1. Introduction

1.1 What is Low-Level design document?

The goal of the LDD or Low-level design document (LLDD) is to give the internal logic design of the actual program code for the Expenditure Data Analysis dashboard. LDD describes the class diagrams with the methods and relations between classes and programs specs. It describes the modules so that the programmer can directly code the program from the document.

1.2 Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

2. Architecture

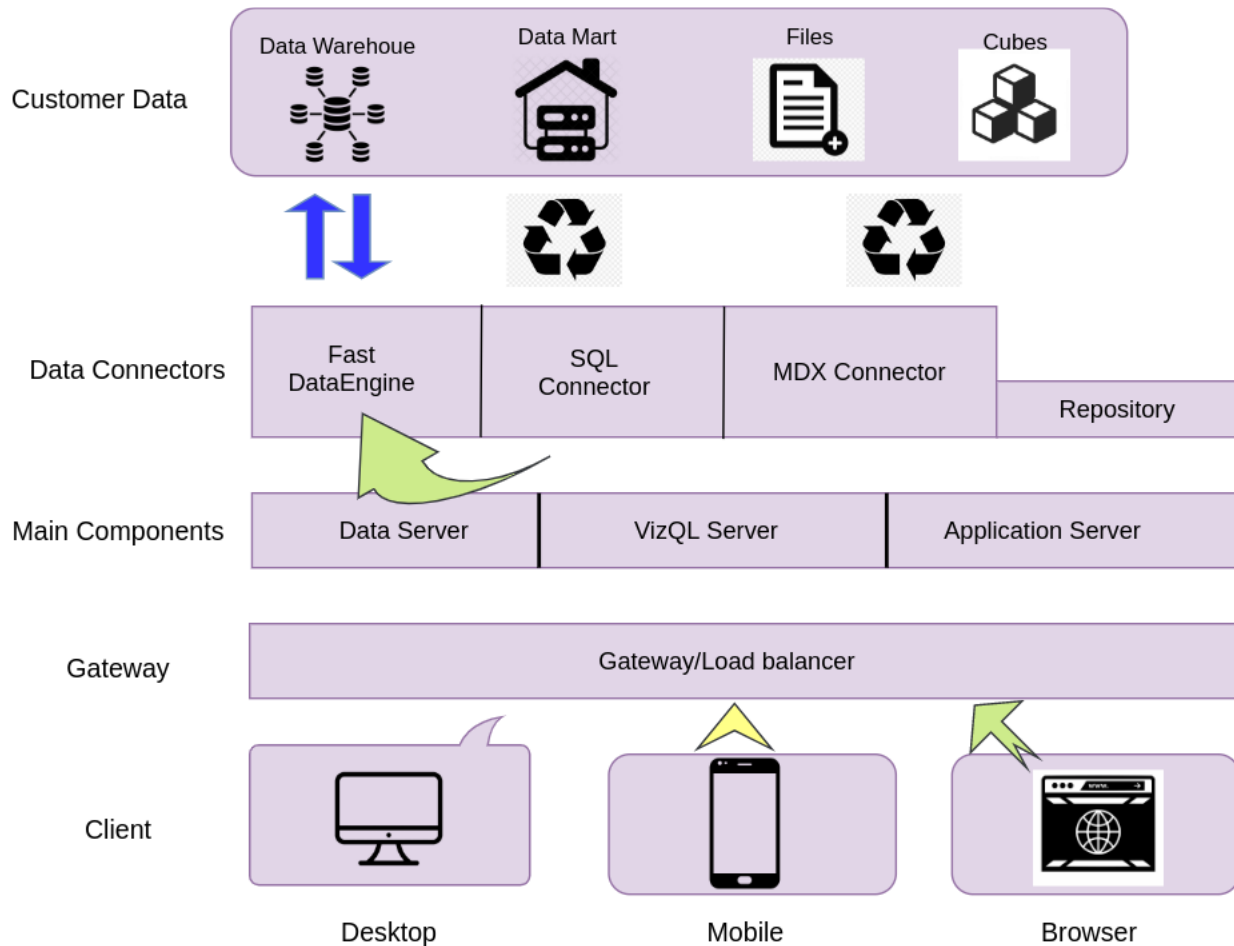


Tableau Server Architecture

Tableau has a highly scalable, n-tier client-server architecture that serves mobile clients, web clients and desktop-installed software. Tableau Server architecture supports fast and flexible deployments.

The following diagram shows Tableau Server's architecture:

Tableau Communication Flow

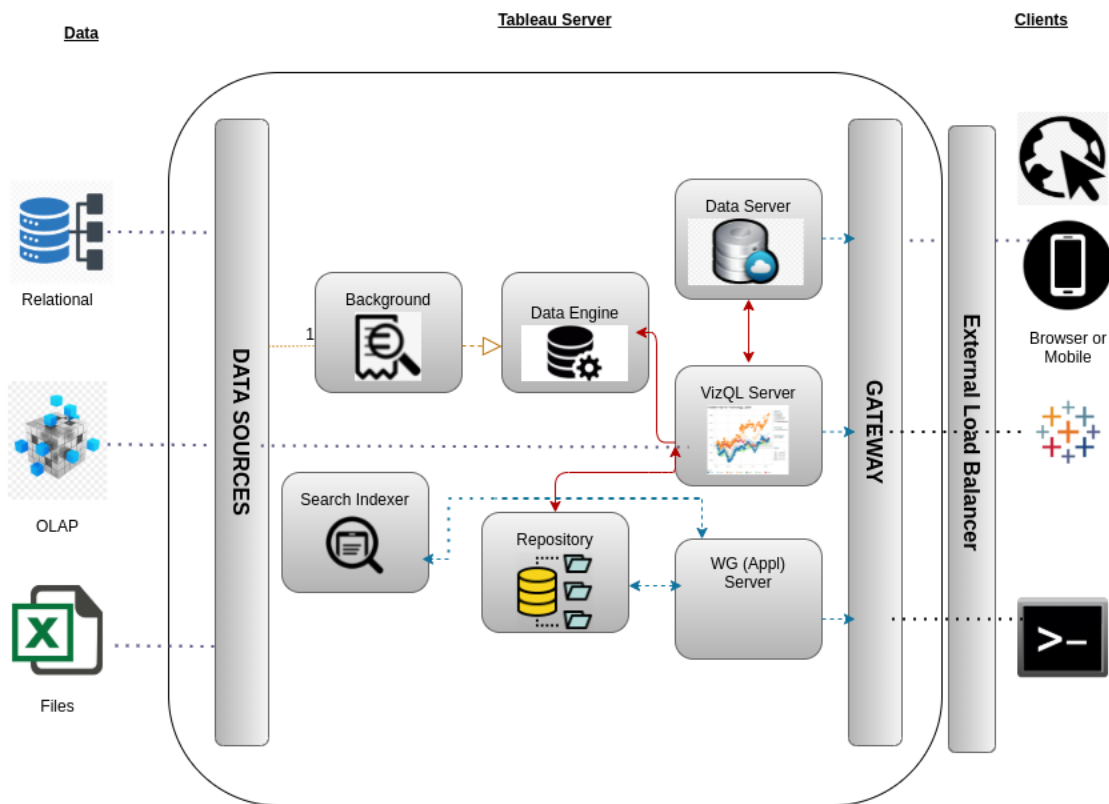


Tableau Server is internally managed by the multiple server processes.

1] Gateway/Load Balancer

It acts as an Entry gate to the Tableau Server and also balances the load to the Server if multiple Processes are configured.

2) Application Server

Application Server processes (wgserver.exe) handle browsing and permissions for the Tableau Server web and mobile interfaces. When a user opens a view in a client device, that user starts a session on Tableau Server. This means that an Application Server thread starts and checks the permissions for that user and that view.

3) Repository

Tableau Server Repository is a PostgreSQL database that stores server data. This data includes information about Tableau Server users, groups and group assignments, permissions, projects, data sources, and extract metadata and refresh information.

4) VIZQL Server

Once a view is opened, the client sends a request to the VizQL process (vizqlserver.exe). The VizQL process then sends queries directly to the data source, returning a result set that is rendered as images and presented to the user. Each VizQL Server has its own cache that can be shared across multiple users

5) Data Engine

It Stores data extracts and answers queries.

6) Backgrounder

The backgrounder Executes server tasks which includes refreshes scheduled extracts, tasks initiated from tab cmd and manages other background tasks.

7) Data Server

Data Server Manages connections to Tableau Server data sources

It also maintains metadata from Tableau Desktop, such as calculations, definitions, and groups.

3. Architecture Description

3.1. Data Description

The Dataset contains year wise distribution of all the states of India for the following parameters:

1) Aggregate Expenditure:

Aggregate expenditure is a measure of national income. Aggregate expenditure is defined as the current value of all the finished goods and services in the economy. The Aggregate expenditure is thus the sum total of all the expenditures undertaken in the economy by the factors during a given time period

2) Capital Expenditure:

Capital expenditure or capital expense is the money an organization or corporate entity spends to buy, maintain, or improve its fixed assets, such as buildings, vehicles, equipment, or land.

3) Gross Fiscal Deficits:

The gross fiscal deficit (GFD) is the excess of total expenditure including loans net of recovery over revenue receipts (including external grants) and

non-debt capital receipts. Generally fiscal deficit takes place either due to revenue deficit or a major hike in capital expenditure.

4) Nominal GDP Series:

Nominal GDP is an assessment of economic production in an economy that includes current prices in its calculation. In other words, it doesn't strip out inflation or the pace of rising prices, which can inflate the growth figure.

5) Own Tax Revenues:

The income generated by states for various activities include revenue receipts like taxes & grants and capital receipts like loans. States which are able to generate more revenue on their own are less dependent on the devolution & central grants.

6) Revenue Deficits:

A revenue deficit occurs when realized net income is less than the projected net income. This happens when the actual amount of revenue and/or the actual amount of expenditures do not correspond with budgeted revenue and expenditures.

7) Revenue Expenditure:

Revenue expenditures are short-term expenses used in the current period or typically within one year. Revenue expenditures include the expenses required to meet the ongoing operational costs of running a business, and thus are essentially the same as operating expenses (OPEX).

8) Social Sector Expenditure:

Social sector expenditure has been defined as the total of all expenditures incurred by the central and the state governments on promotional and protective measures.

3.2. Web Scrapping

Web scrapping is a technique to automatically extract content and data from websites using bots. It is also known as web data extraction or web harvesting. Web scrapping is made simple now days, many tools are used for web scrapping. Some of python libraries used for web scrapping are BeautifulSoup, Scrapy, Selenium, etc.

3.3. Data Preparation

In the Preparation Process, we will convert our original datasets with other necessary attributes format. And will merge it with the Scrapped dataset.



All the 8 datasets are of same format as shown below:

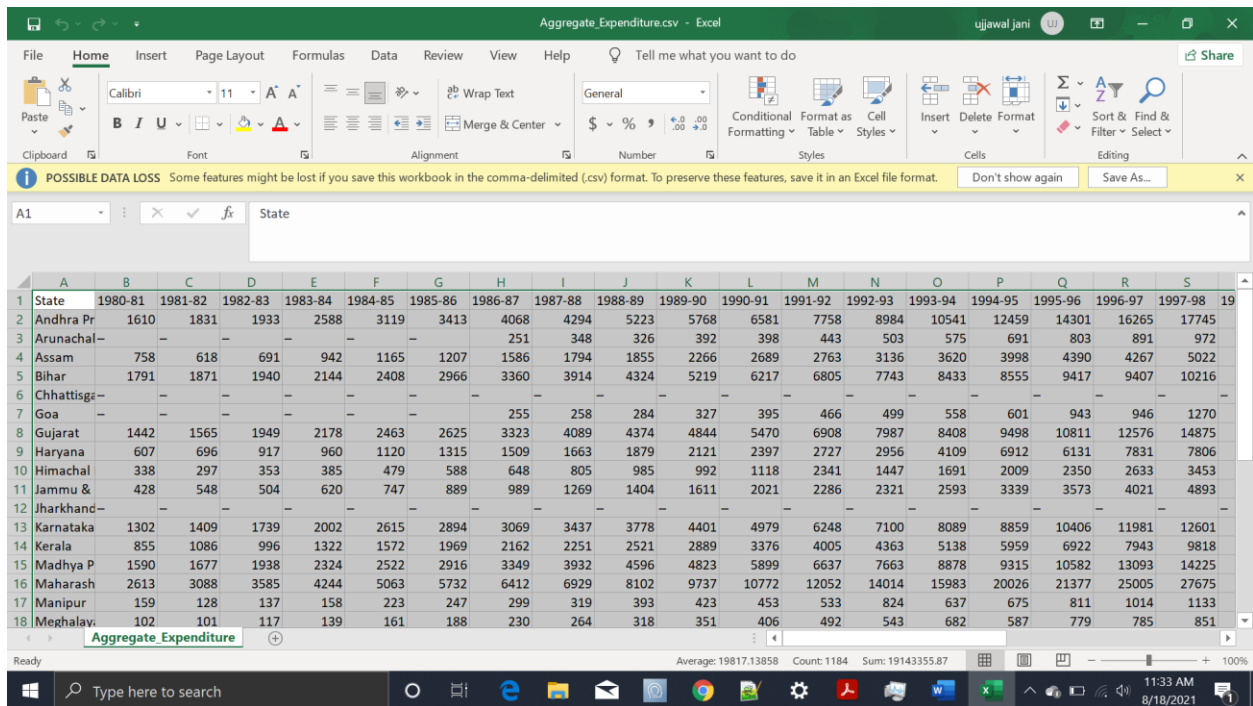
State	1980-81	1981-82	1982-83	1983-84	1984-85	1985-86	1986-87	1987-88	1988-89	1989-90	1990-91	1991-92	1992-93	1993-94	1994-95	1995-96	1996-97	1997-98
Andhra Pr	1610	1831	1933	2588	3119	3413	4068	4294	5223	5768	6581	7758	8984	10541	12459	14301	16265	17745
Arunachal	-	-	-	-	-	-	251	348	326	392	398	443	503	575	691	803	891	972
Assam	758	618	691	942	1165	1207	1586	1794	1855	2266	2689	2763	3136	3620	3998	4390	4267	5022
Bihar	1791	1871	1940	2144	2408	2966	3360	3914	4324	5219	6217	6805	7743	8433	8555	9417	9407	10216
Chhattisgarh	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Goa	-	-	-	-	-	-	255	258	284	327	395	466	499	558	601	943	946	1270
Gujarat	1442	1565	1949	2178	2463	2625	3323	4089	4374	4844	5470	6908	7987	8408	9498	10811	12576	14875
Haryana	607	696	917	960	1120	1315	1509	1663	1879	2121	2397	2727	2956	4109	6912	6131	7831	7806
Himachal	338	297	353	385	479	588	648	805	985	992	1118	2341	1447	1691	2009	2350	2633	3453
Jammu & Kashmir	428	548	504	620	747	889	989	1269	1404	1611	2021	2286	2321	2593	3339	3573	4021	4893
Jharkhand	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Karnataka	1302	1409	1739	2002	2615	2894	3069	3437	3778	4401	4979	6248	7100	8089	8859	10406	11981	12601
Kerala	855	1086	996	1322	1572	1969	2162	2251	2521	2889	3376	4005	4363	5138	5959	6922	7943	9818
Madhya Pradesh	1590	1677	1938	2324	2522	2916	3349	3932	4596	4823	5899	6637	7663	8878	9315	10582	13093	14225
Maharashtra	2613	3088	3585	4244	5063	5732	6412	6929	8102	9737	10772	12052	14014	15983	20026	21377	25005	27675
Manipur	159	128	137	158	223	247	299	319	393	423	453	533	824	637	675	811	1014	1133
Meghalaya	102	101	117	139	161	188	230	264	318	351	406	492	543	682	587	779	785	851

As you all can notice that format of the data we have is not good to analyze and visualize. So, we need to reconstruct the structure of the dataset.

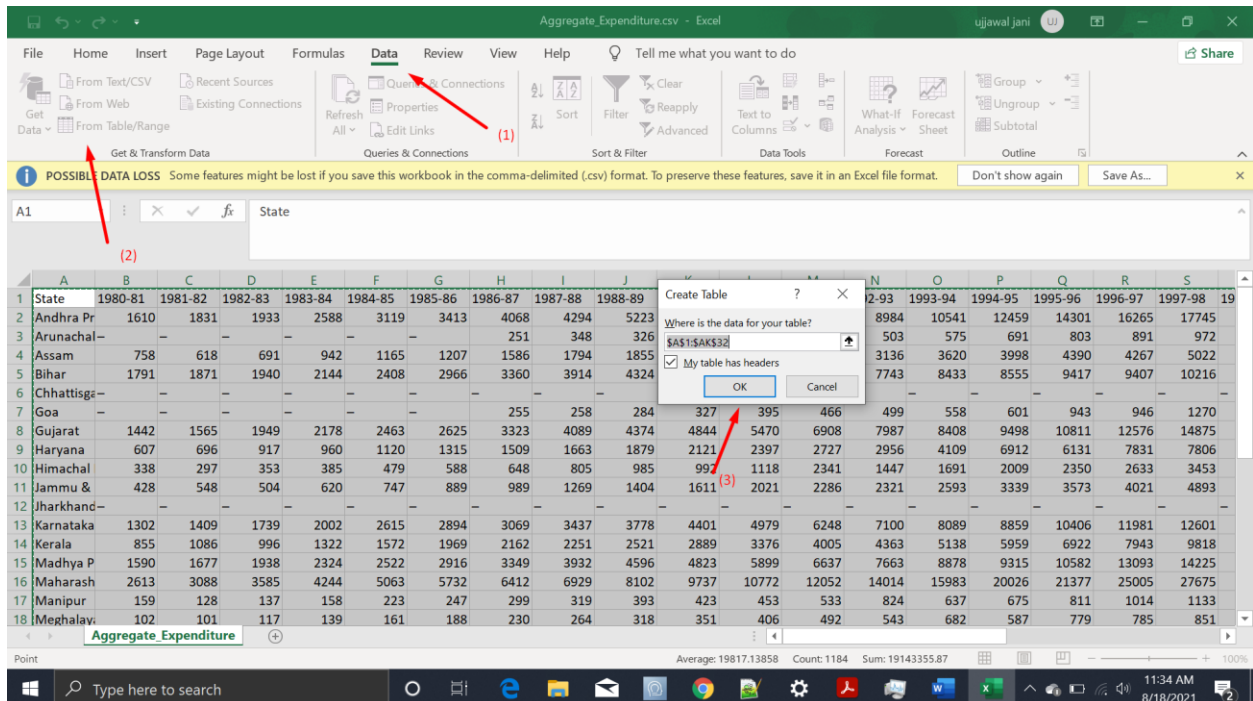
As this is a pivoted data set, we need to unpivot it. That's the only way by which we can make meaningful insights from it. This process is known as converting wide data to long data.

We will be using only MS Excel for data restructuring and cleaning purpose.

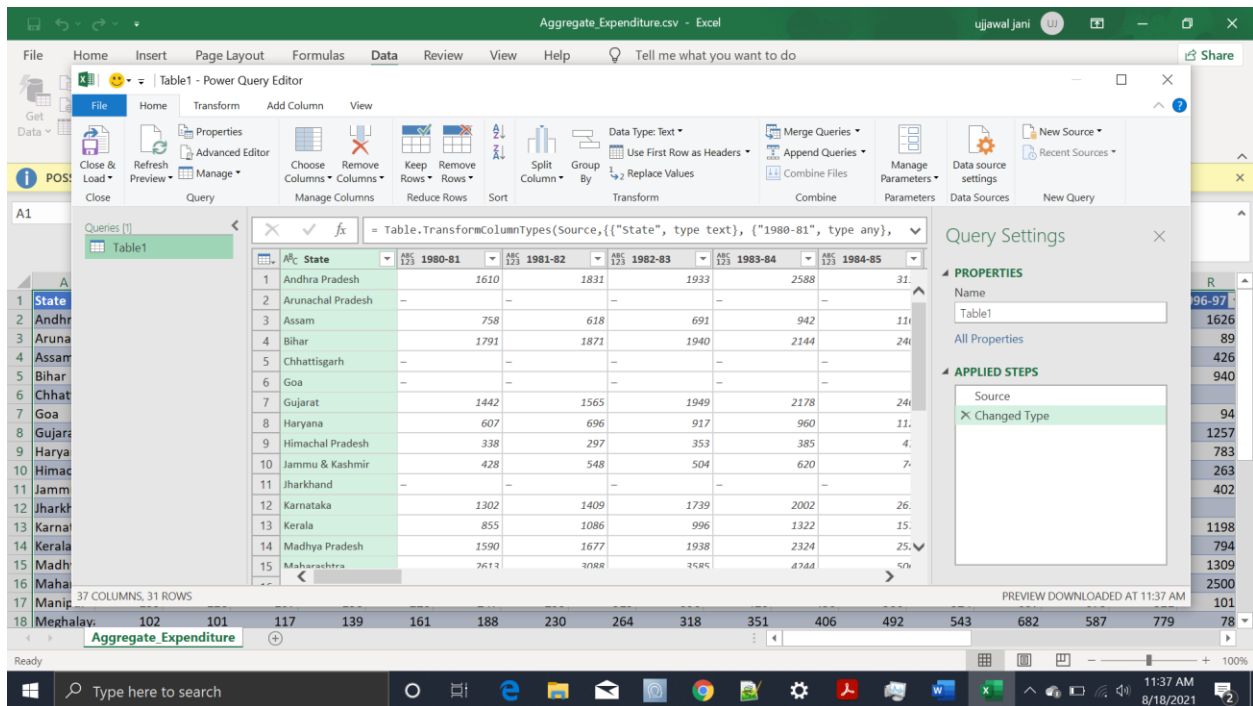
Step 1) Select all rows and columns.



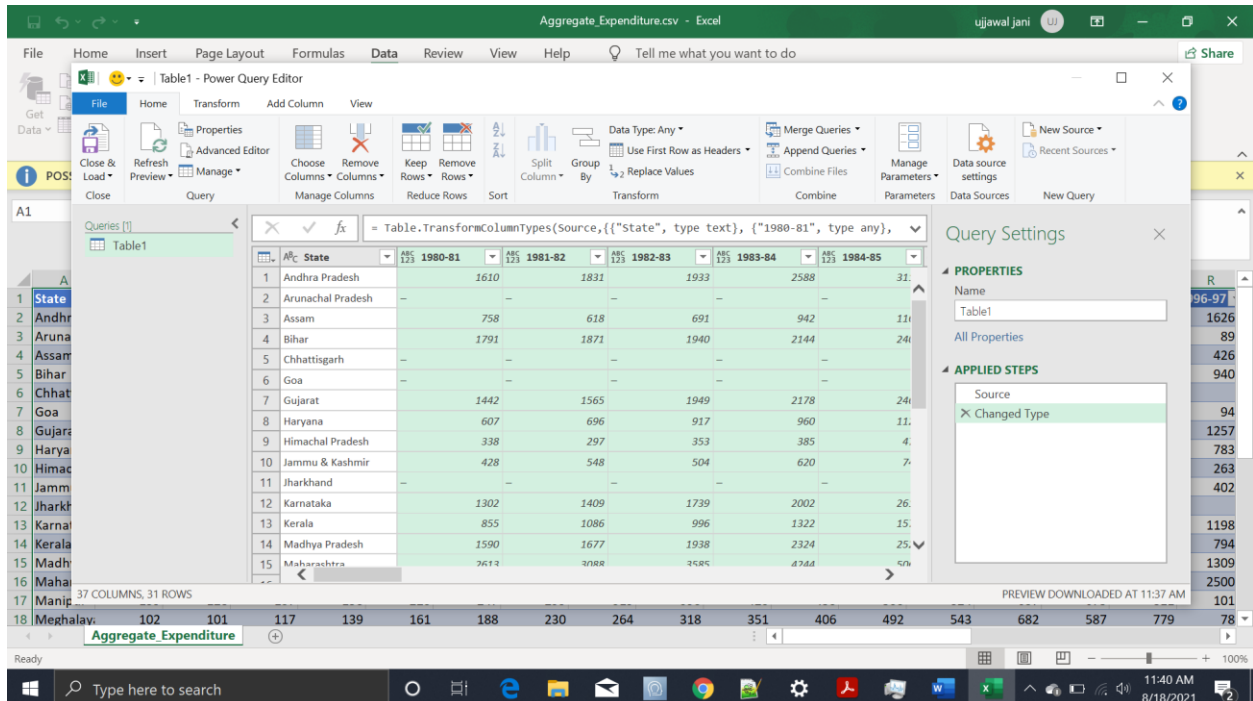
Step 2) Go to data section: click From Table/Range: then click OK.



Then a Power Query Editor window will get popped up.



Step 3) Select all the columns(i.e. all the years columns) that we have to unpivot.



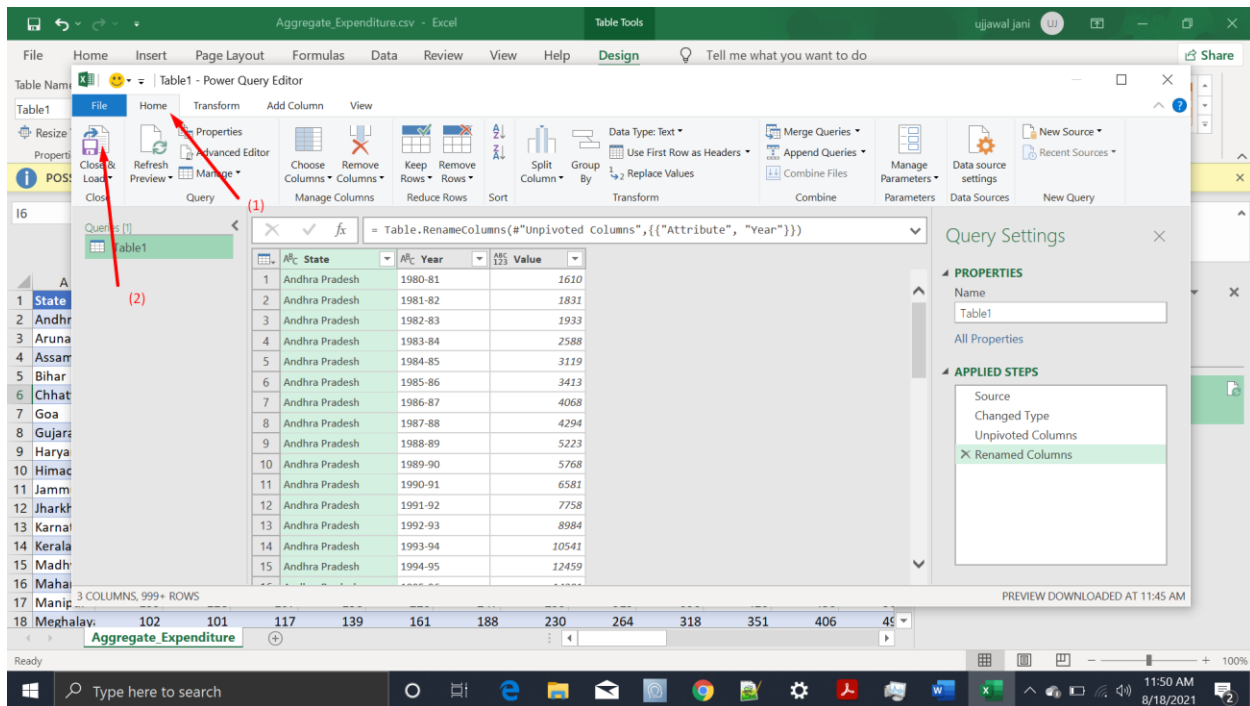
Step 4) Go to transform section and click unpivot columns.

The screenshot shows the Microsoft Excel interface with the Power Query Editor open. The 'Transform' tab is selected in the ribbon. The 'Unpivot Columns' button is highlighted with a red arrow and labeled (2). Another red arrow labeled (1) points to the 'Table' button in the 'Home' tab. The data table shows columns for State, Year, and Value. The 'Query Settings' pane on the right shows the 'APPLIED STEPS' list with 'Changed Type' selected.

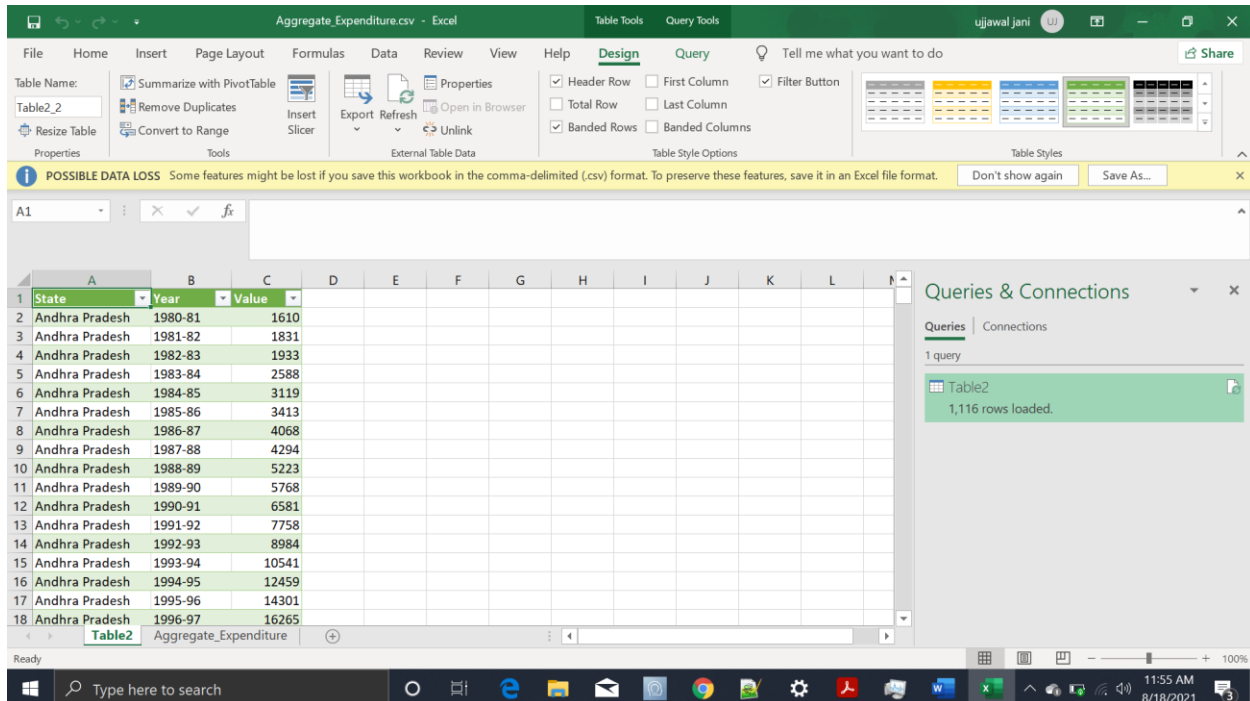
Now an unpivoted dataset will be shown, change column name of attribute to year.

The screenshot shows the Microsoft Excel interface with the Power Query Editor open. The 'Transform' tab is selected in the ribbon. The 'Unpivot Columns' button is highlighted with a red arrow and labeled (2). Another red arrow labeled (1) points to the 'Table' button in the 'Home' tab. The data table shows columns for State, Year, and Value. The 'Query Settings' pane on the right shows the 'APPLIED STEPS' list with 'Renamed Columns' selected.

Step 5) Now go to home section and click close and load.



Now another table has been created with unpivoted columns.



Now we need to all these steps for all the datasets we have.

We can even combine all the datasets by having a column which specifies the dataset name.

Category	State	Year	Value
Aggregate_Expenditure	Andhra Pradesh	1980-81	1610
Aggregate_Expenditure	Andhra Pradesh	1981-82	1831
Aggregate_Expenditure	Andhra Pradesh	1982-83	1933
Aggregate_Expenditure	Andhra Pradesh	1983-84	2588
Aggregate_Expenditure	Andhra Pradesh	1984-85	3119
Aggregate_Expenditure	Andhra Pradesh	1985-86	3413
Aggregate_Expenditure	Andhra Pradesh	1986-87	4068
Aggregate_Expenditure	Andhra Pradesh	1987-88	4294
Aggregate_Expenditure	Andhra Pradesh	1988-89	5223
Aggregate_Expenditure	Andhra Pradesh	1989-90	5768
Aggregate_Expenditure	Andhra Pradesh	1990-91	6581
Aggregate_Expenditure	Andhra Pradesh	1991-92	7758
Aggregate_Expenditure	Andhra Pradesh	1992-93	8984
Aggregate_Expenditure	Andhra Pradesh	1993-94	10541
Aggregate_Expenditure	Andhra Pradesh	1994-95	12459
Aggregate_Expenditure	Andhra Pradesh	1995-96	14301
Aggregate_Expenditure	Andhra Pradesh	1996-97	16265
Aggregate_Expenditure	Andhra Pradesh	1997-98	17745

Thus, we have restructured and combined all the datasets into one file. This way we can visualize using filters of category, state or year.

3.4 Data Cleaning.

The data we have contains blank or “-” values. We have to fill them with Zero(0) to make values column a proper numerical column.

In Excel we can do that by selecting the whole column,

1. Click on “Find and select” and then “Find” to replace all “-” values with “0”.
2. Click on “Find and select” and then “Go To Special” to replace all Blanks with “0”.

Next, In the years column some data points have extra text. To remove them, we can create a new column with formula “=Right(C2,7)” and apply it for whole column to select only year range from Years column and replace it with original Years column.

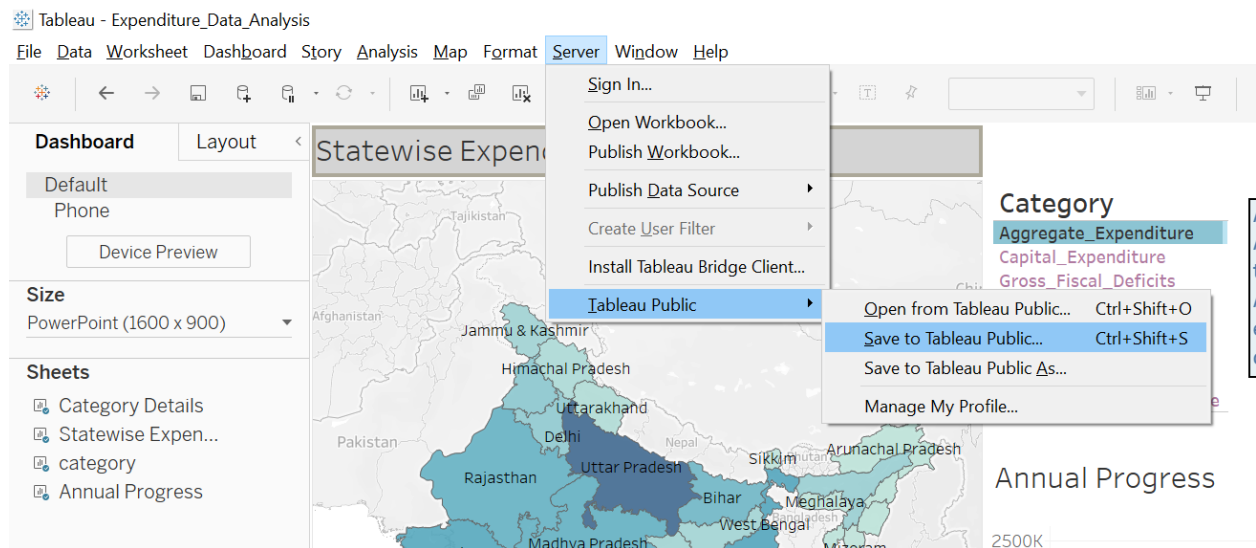
3.5 Make the MS Excel connection and set up the data source

Step 1: Configuring Tableau

1. After you open Tableau, under **Connect**, click **Excel**.
2. Select the Excel workbook you want to connect to, and then click **Open**.
3. On the Data Source page, click the sheet tab to start your analysis.

3.6 Deployment.

1. Once you've completed your dashboard, follow these steps: **Server, Tableau Public, Save to Tableau Public As**
2. You may be prompted to log into your Tableau Public profile first if this is your first-time publishing.



- Next, fill out the title you want your viz to have and click “save”.
- Here in the below screenshot, we can see that our workbook has been published to tableau public.

