Low Level Design

Crop Production Data Analysis-India

Author: Nivetha N

Contents

1. Introduction

1.1 Why this Low-Level design document?

The goal of the LDD or Low-level design document (LLDD) is to give the internal logical design of the actual program code for the crop production Data Analysis- India. 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

Power BI is a collection of software services, apps, and connectors that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights.. Power BI lets you easily connect to your data sources, visualize and discover what's important, and share that with anyone or everyone you want.

- A Windows desktop application called **Power BI Desktop.**
- An online SaaS (*Software as a Service*) service called the **Power BI service.**
- Power BI mobile apps for Windows, iOS, and **Android devices.**

The following diagram shows Power BI Desktop architecture:

Power BI Architecture

Power BI Architecture
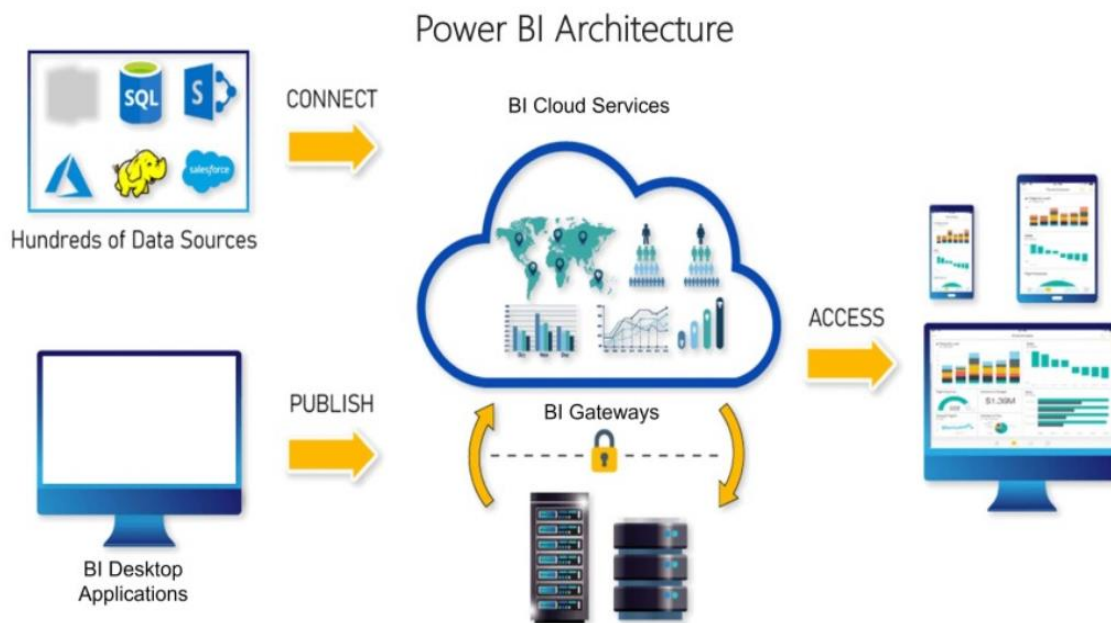
i. Data Preparation

Power BI Supports large Range of data Sources. We can click and get data and it shows us all the available data connections. It allows you to connect to different flat files, SQL Database, and Azure Cloud or even web platforms such as Facebook, Google Analytics, and Salesforce objects.

ii. ETL (Extract, Transform, Load) Process

Power BI follows ETL Process. During the ETL Process, data is extracted from a data source, then transformed, validated, standardized, corrected, quality checked and ultimately loaded into a data warehouse.

iii. Components of power bi

Power query: power query is a tool in Microsoft excel that simplifies the process of importing data from different source files and sorting them into an excel sheet in the most convenient and usable format. It is a tool where we can **transform** the data as per our need

Power pivot: Power pivot is an in-memory **data modelling** component that provides highly compressed data storage and extremely fast aggregation and calculation. It is also available as part of excel and can be used to create a data model in an Excel Workbook.

Power view: Power view is a **data Visualization** technology that lets you create interactive charts, graphs, maps, and other visuals that bring your data to life. It is available in excel, in sharepoint, sql server, and Power BI.

Power map: power map is for visualizing geo-spatial information in 3D mode. You can highlight data based on the geo-graphical location such as country, city, state, and street address information.

Power BI Desktop: power bi desktop is the newest component in power bi suit.

Power bi desktop is a holistic development tool for power query, power pivot and power view. It is easier to develop BI and data analysis experience with that. Power BI Desktop updates frequently and regularly.

Power BI Website : power bi solution can be published to power BI Website. In power bi website the data source can be scheduled to refresh. We can built report and visualizations directly on power bi site as well.

Power Q & A : power Q&A is a natural language engine for questions and answers to your data model. once you have built your data model and deployed that into power bi website, then you or your users can ask questions and get answers easily.

Power BI Mobile Apps: power BI Mobile apps have three mobile operating system providers: Android, ios, and windows. These applications provide you an interactive view of reports and dashboards on the power bi site.

3. Architecture Description

Microsoft SQL Server

Microsoft SQL Server is a relational database management system developed by Microsoft. As a database server, it is a software product with the primary function of **storing and retrieving data** as requested by other software applications-which may run either on the same computer or on another computer across a network.

3.1. Data Description

The Dataset contains India crop production. The data refers to district wise, crop wise, season wise and year wise data on crop covered area (Hectare) and production (Tonnes). The data is being used to study and analyse crop production, production contribution to district/State/country,

Dimensionality   - 7 columns

Discrete   - State_Name, District_Name, Crop_year, Season, Crop

Continous   - Area, Production

Independent features - State_Name,District_Name,Crop_year,Season,Crop, Area

Dependent features   - Production

1. State_Name : States of India

2. Crop   : Different kind of crops

3. Crop_year   : 1997-2015

5. District   : Districts of India

6. Production : crop production (Tonnes)

7. Area   : Area distribution to the crop (hectares)

Data Quality

- outliers – Handled outliers using Grouping, Standard deviation in sql server.
- Missing Values – There are 3730 missing values in the production column.
- Duplicate Data – No duplicate present in the Data set.

3.2 Data Transformation

- Transform the data into **power Query**. In the Transformation Process, we have to transform our original data for remove the blank values and error values using power Query .
- we have to change the data type of crop_year column text data type into date data type in power bi

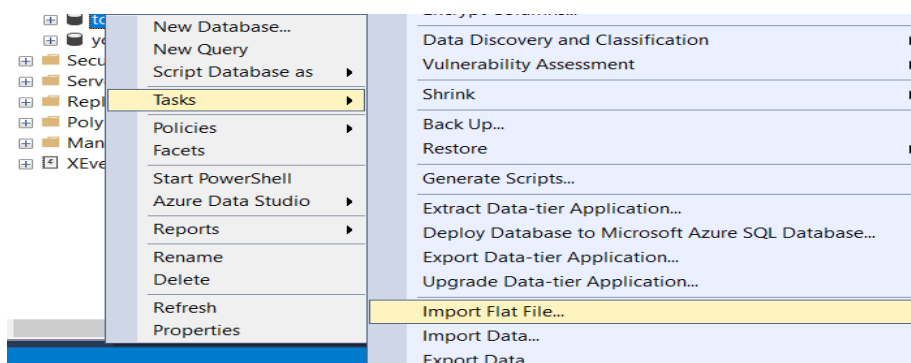## 3.3 Data Insertion into Database
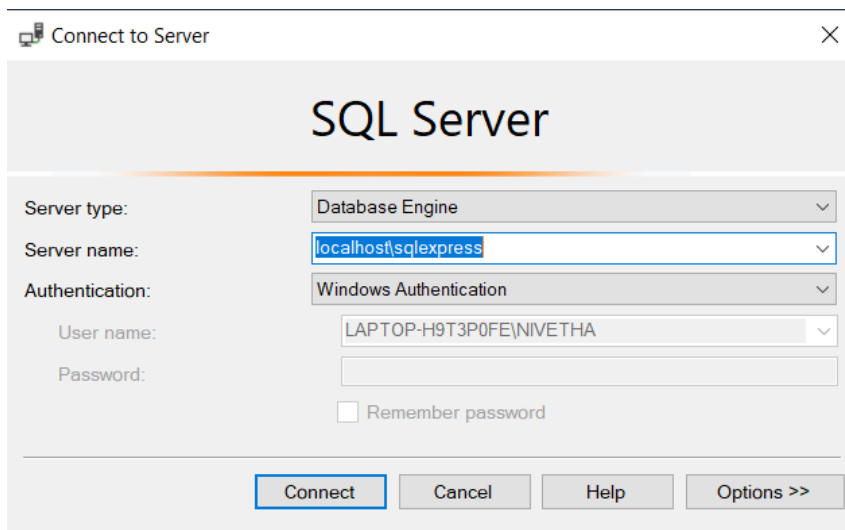
- Once we open the sql server the pop up window shows like this picture
- Select connect option

a. Database Creation and connection

- Create a database with database name
- if the database is already created, open the connection to the database

b. Insertion of files in the table

MS SQL SERVER

- Select database, right-click on it -> Tasks ->Select "Import flat file".
- Browse the file and give **table name** for the data set.
- Preview data before saving it.

- Check Data-type and map it properly, to successfully import csv.

SQL Technique for analysis

I was Begin my analysis with the simplest query

1. Counting Rows and Items

2. Aggregation Functions

3. Extreme value Identification

4. Slicing Data

5. Top 5 (using this we can see the first 5 rows)

6. Sorting Data

7. Filtering Patterns

8. Grouping Data and Filtering in Groups

```sql
select count(*)as rows from crop_production

select avg(production)as average from crop_production

select min(production)as minimum from crop_production

select max(production)as maximum from crop_production

select sum(production)as totalproduction from crop_production

select stdev(production)as standarddeviation from crop_production
```
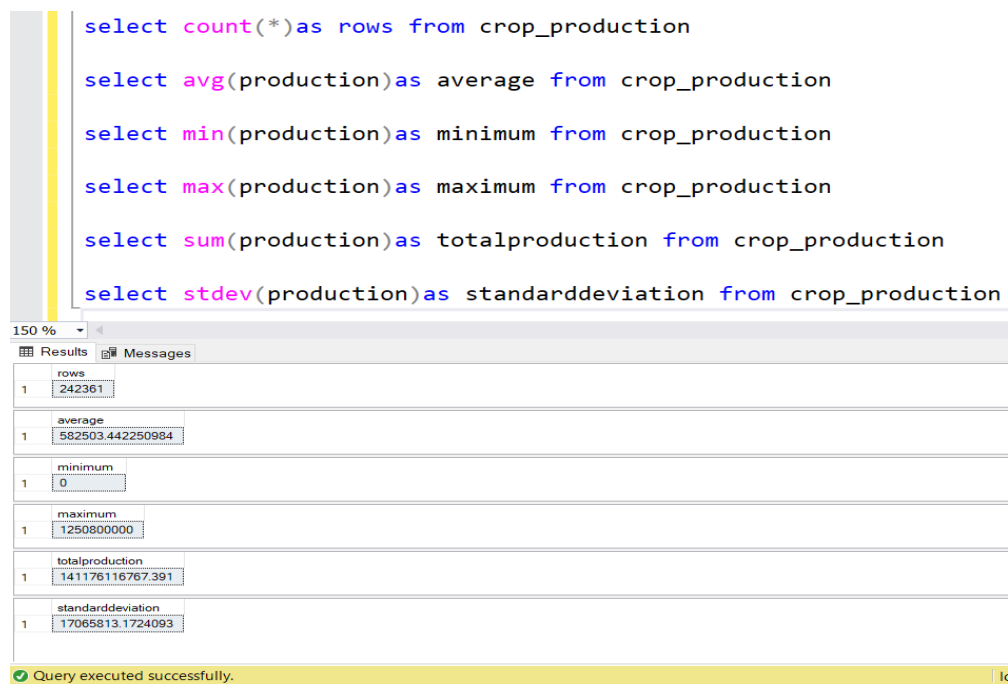
150 % ▼ ◀

Results   Messages

rows
242361

average
582503.442250984

minimum
0

maximum
1250800000

totalproduction
141176116767.391

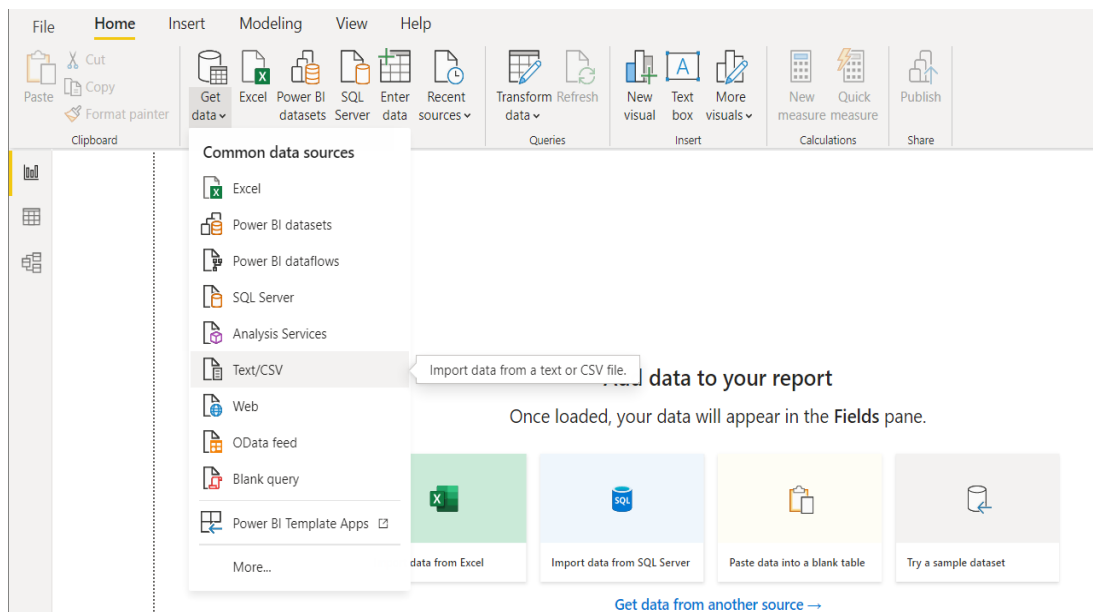| standarddeviation |
| --- |
| 17065813.1724093 |

✓ Query executed successfully.

- This queries shows us Basic Understanding of the Dataset.
- We have created difficult queries as well to analyze the data.

3.4 Import the CSV file data into Power BI

- Click Get the data from Text/CSV file and load the data set into power bi

- The data set will load into the power bi

- We don't need to clean the data. Because we have already clean the data using **Power Query** in **Excel.**

3.5 Development

For development process we have used power bi desktop

Use power bi desktop to edit reports and data sets:

- Consider **power bi desktop** as our **local development environment**. Power bi desktop allows us to try, explore, and review updates to our reports and datasets. Once the work is done, you can upload your new version to the development stage. Due to the following reasons, it's recommended to edit PBIX files in the desktop (and not in power BI Service).

- It is easier to collaborate with fellow creators on the same PBIX file, if all changes are being done on the same tool.

Once you've completed your dashboard, follow these steps:-

File → save as →filename →power BI file(*.pbix) format
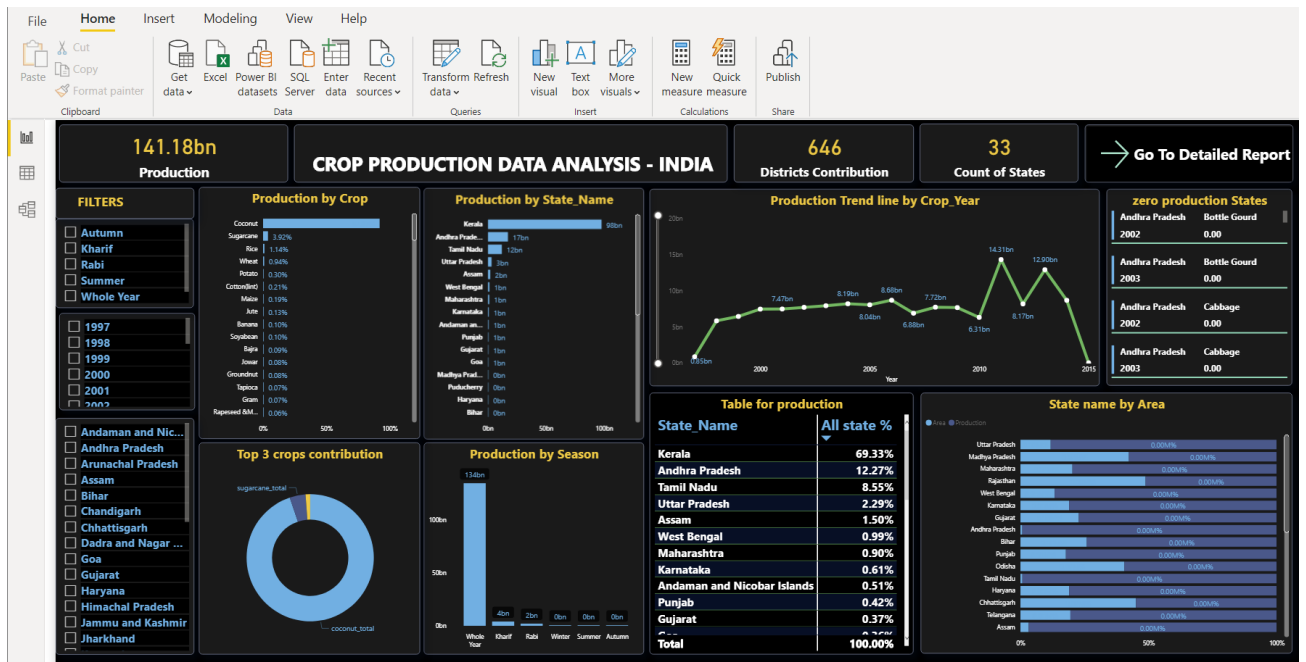
3.6 Explanation of the dashboard

Slicers

I have created a **3 Slicers** to filters the other charts and visuals

1. States_Name

2. crop_year

3. Season

Clustered Bar charts:

1. Production by Crop

- It gives us the results of crop wise production

2. Production by State_Name

- How the production vary state to state and state wise Production

Clustered Column chart:

Production by season

- How the production is vary to season to season and shows the each season's contribution.

Donut Chart

Top 3 crops production

- what is the top 3 crop production contribution compare to other crops in different years .

Line Chart and Table

Production trend line by crop_year

- production is improving or not year by year Trend.
- The **Table** shows us different states in india and their contribution in Percentage.

Stacked Column Chart

State_name by Area

- Area distribution for each states and their production.

Multi-row Card

Zero production states

- which states gets zero production in which year and which crop.

DAX (Data Analysis Expressions)

Measures

All crop % = SUM (crop_production [Production]) / [All crop total]

All state % = SUM (crop_production [Production]) / [All state total]

coconut_total = CALCULATE (SUM (crop_production [Production]),
                crop_production [Crop] = "coconut")

rice_total = CALCULATE (SUM (crop_production [Production]),
crop_production [Crop] = "Rice")

sugarcane_total = CALCULATE (SUM (crop_production [Production]),
crop_production [crop] = "sugarcane")

zero production = CALCULATE (SUM (crop_production [Production]),
crop_production [Production] = 0)