LOW LEVEL DESIGN DOCUMENT

(Crop Production Analysis in India)

Akshay Salve

VERSION: 1.0

DATED: 28/12/2021

Document Version Control:

Crop Production Analysis in India – Business Intelligence Project

Version	Date	Author	Change
1.0	13/06/2022	Akshay Salve	First Version of Complete LLD

Abstract

The agriculture business domain, as a vital part of the overall supply chain, is expected to highly evolve in the upcoming years via the developments, which are taking place on the side of the Future Internet. This paper presents a novel business-to-business collaboration platform from the agri-food sector perspective, which aims to facilitate the collaboration of numerous stakeholders belonging to associated business domains, in an effective and flexible manner.

This dataset provides a huge amount of information on crop production in India ranging from several years. Based on the Information the ultimate goal would be to predict crop production and find important insights highlighting key indicators and metrics that influence the crop production.

Contents:

1.Introduction:

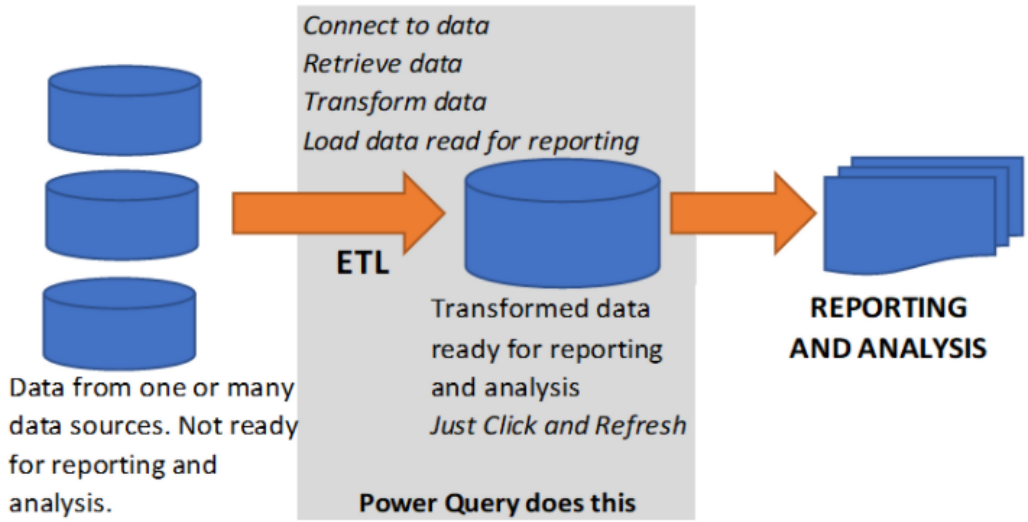
1.1. Why this Low-Level Design Document?

The goal of the LDD or Low-level design document (LLDD) is to give the internal logic design of the actual program code for the Bank Marketing Campaign Analysis. 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:



ETL (extract, transform and load) in Power BI uses preparation of data sets for analysis by removing irregularities in the data. It also involves data visualization to draw meaningful patterns and insights.

Based on the results of ETL, companies also make business decisions, which can have repercussions later.

- If ETL is not done properly then it can damage the business a lot in many ways such as loss of client which we are working for, the decision making will go completely wrong and many more issues.
- If done well, it may improve the efficacy of everything we do next.

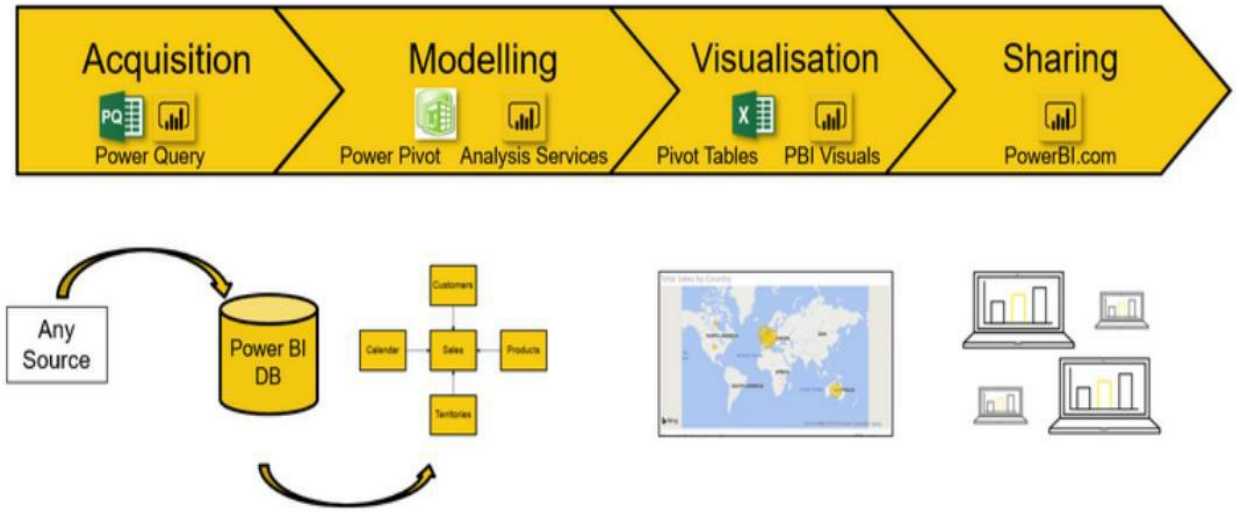
Below are following steps to follow for ETL:

1. Data Sourcing

2. Data Cleaning

3. Data Modelling

4. Data Visualization



3. Architecture Description:

3.1 Data Sourcing:

The dataset is in csv (comma separated values) format. MS Excel is used to load the data.

Low Level Design (LLD)

Citation Request:

This Dataset is publicly available for research, Available at https://data.world/thatzprem/agriculture-india named as crop_production.csv.

1. Title - India Crop Production - State wise
2. Source - https://data.world/thatzprem/agriculture-india

3.2. Data Overview –

- The Data includes single .csv file with all examples, ordered by date (Year 1997 to Year 2015).
- The Number of Instance - 246091 for crop_production.csv
- Number of attributes – 7 attributes

3.2 Date Description –

- State_name = Name of States in India (categorical : 'Andaman and Nicobar Islands', 'Andhra Pradesh', 'Arunachal Pradesh', 'Assam', 'Bihar', 'Chandigarh', 'Chhattisgarh', 'Dadra and Nagar Haveli', 'Goa', 'Gujarat', 'Haryana', 'Himachal Pradesh', 'Jammu and Kashmir ', 'Jharkhand', 'Karnataka', 'Kerala', 'Madhya Pradesh', 'Maharashtra', 'Manipur' , 'Meghalaya', 'Mizoram', 'Nagaland', 'Odisha', 'Puducherry', 'Punjab', 'Rajasthan', 'Sikkim', 'Tamil Nadu', 'Telangana ', 'Tripura', 'Uttar Pradesh', 'Uttarakhand', 'West Bengal')

- Dsitrict_Name – Name of Districts in India (categorical: 'NICOBARS', 'NORTH AND MIDDLE ANDAMAN', 'SOUTH ANDAMANS', 'ANANTAPUR', 'CHITTOOR', 'EAST GODAVARI', 'GUNTUR', 'KADAPA','KRISHNA', 'KURNOOL', 'PRAKASAM', 'SPSR NELLORE', 'SRIKAKULAM', 'VISAKHAPATANAM', 'VIZIANAGARAM', 'WEST GODAVARI', 'ANJAW', 'CHANGLANG', 'DIBANG VALLEY', 'EAST KAMENG', 'EAST SIANG', 'KURUCropNG KUMEY', 'LOHIT', 'LONGDING', 'LOWER DIBANG VALLEY', Etc)

- Crop_Year – Year of Crop Production (Numerical: 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2010, 1997, 1998, 1999, 2007, 2008, 2009, 2011, 2012, 2013, 2014, 2015)

- Season – Season of the Crops (Categorical: 'Kharif', 'Whole Year ', 'Autumn', 'Rabi', 'Summer', 'Winter')

- Crop – Name of the Crop Sown (Categorical: 'Arecanut', 'Other Kharif pulses', 'Rice', 'Banana', 'Cashew', 'Coconut ', 'Dry ginger', 'Sugarcane', 'Sweet potato', 'Tapioca', 'Black pepper', 'Dry chillies', 'other oilseeds' , Etc)

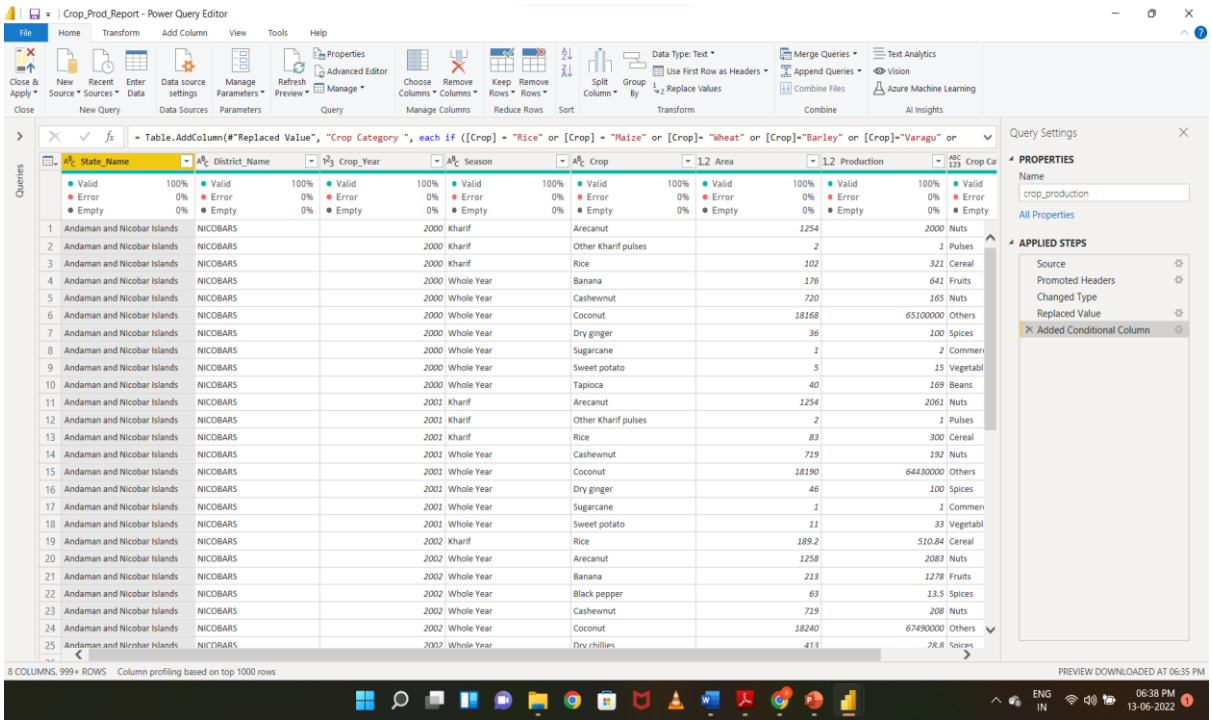
- Area – Area Under cultivation (Numerical)

- Production – Production of the crops (Numerical)

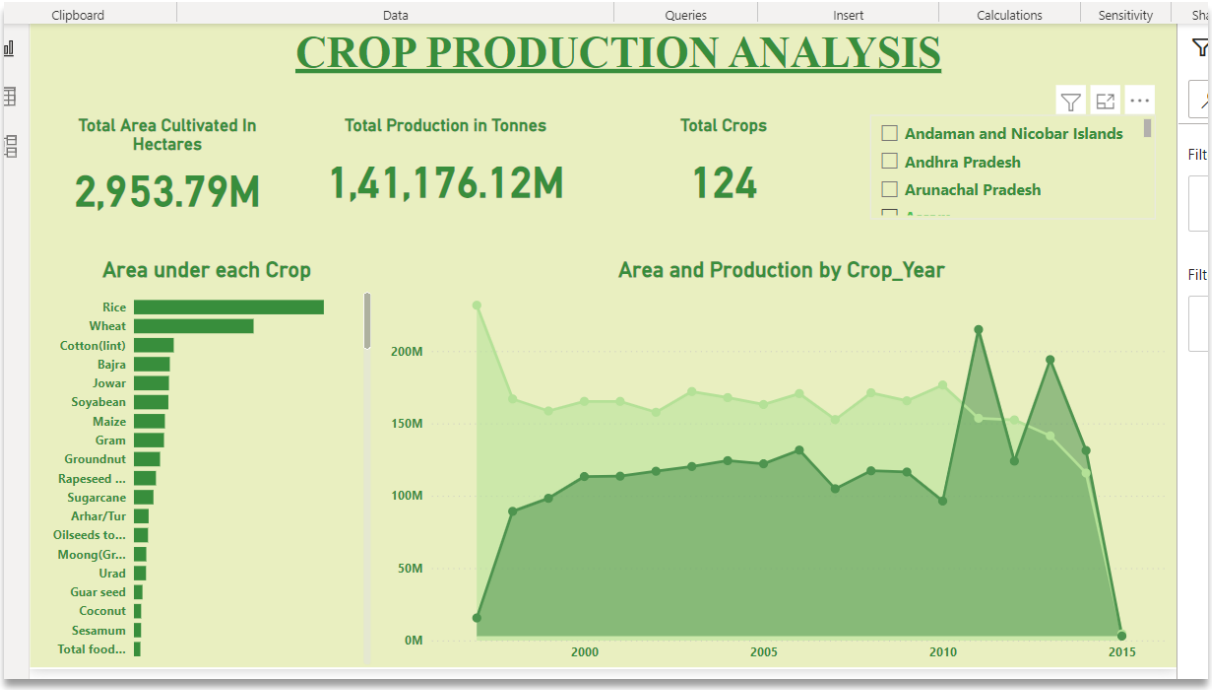
3.4 Data loading in Power BI Query Editor

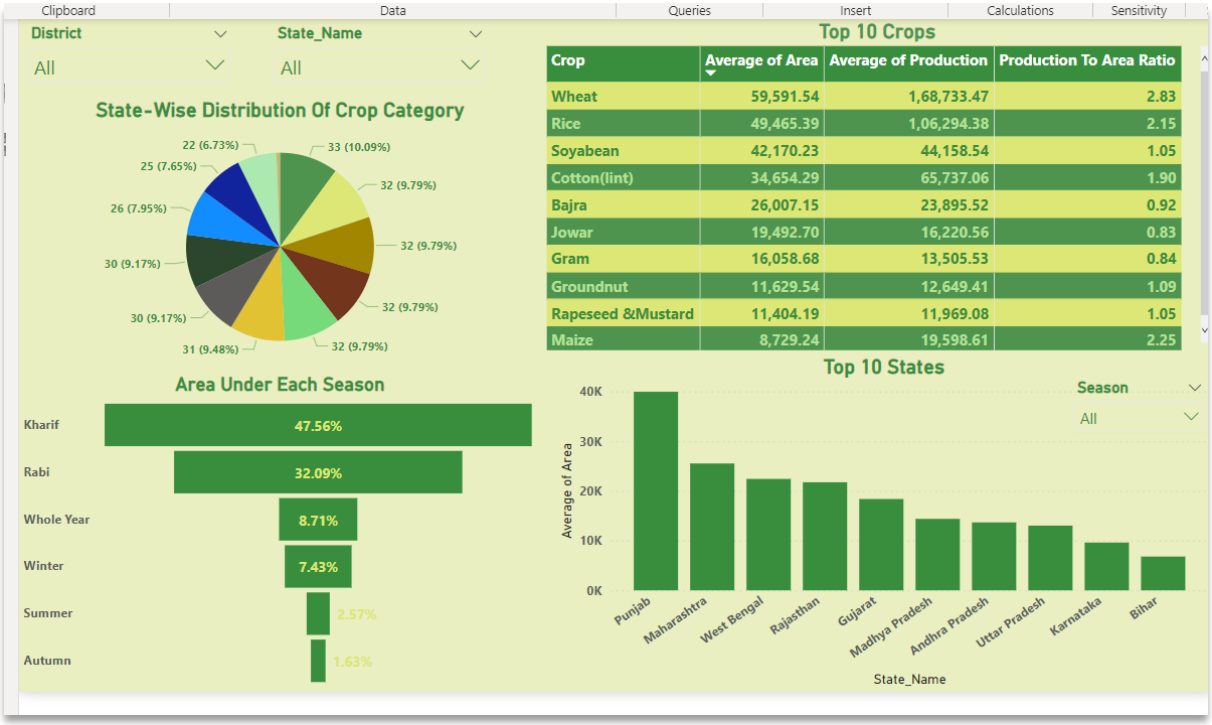
Power Query is the data connectivity and data preparation technology that enables end users to seamlessly import and reshape data from within a wide range of Microsoft products, including Excel, Power BI, Analysis Services, data verse, and more with the following characteristics:

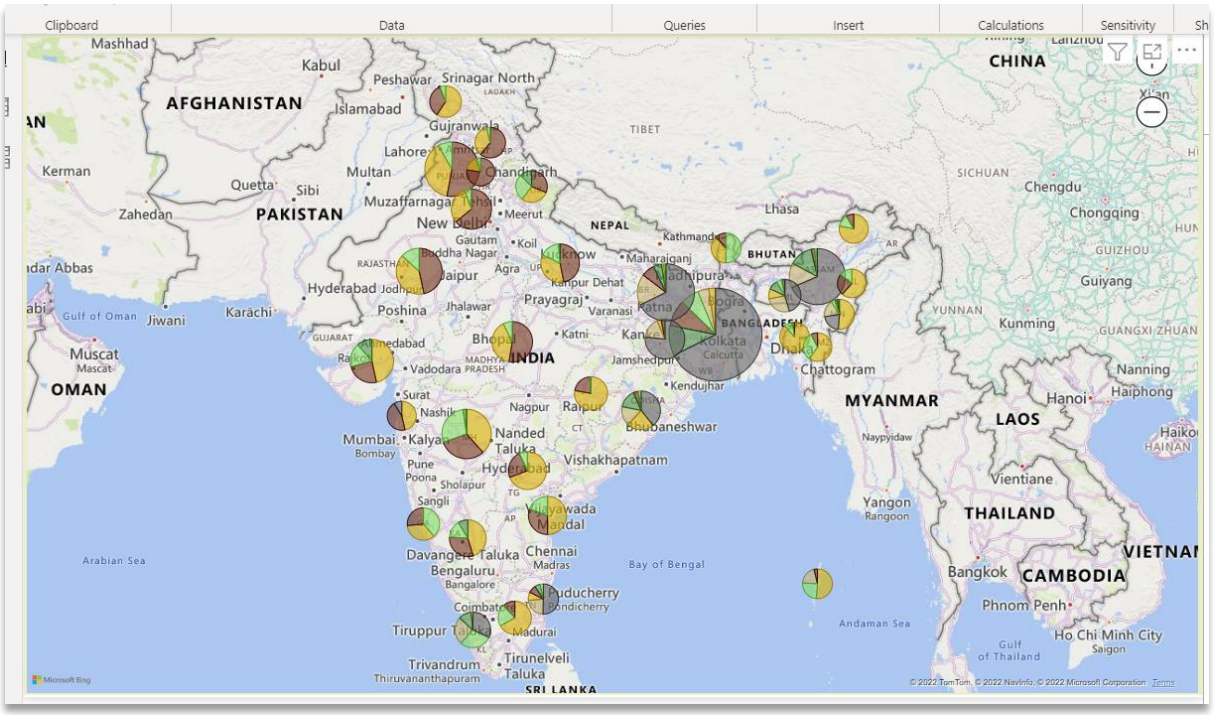
- There can be multiple rows and columns in the data.
- Each row represents a sample of data,
- Each column contains a different variable that describes the samples (rows).
- The data in every column can be a different type of data – e.g. numbers, strings, dates, Boolean etc.



3.5 Data to Insights through Visualizations and Excel Data Analysis




Clipboard
Data
Queries
Insert
Calculations
Sensitivity
District
All
State_Name
All
State-Wise Distribution Of Crop Category
22 (6.73%)
33 (10.09%)
25 (7.65%)
32 (9.79%)
26 (7.95%)
32 (9.79%)
30 (9.17%)
32 (9.79%)
30 (9.17%)
31 (9.48%)
32 (9.79%)
Area Under Each Season
Kharif
47.56%
Rabi
32.09%
Whole Year
8.71%
Winter
7.43%
Summer
2.57%
Autumn
1.63%
Top 10 Crops
Crop
Average of Area
Average of Production
Production To Area Ratio
Wheat 59,591.54 1,68,733.47 2.83
Rice 49,465.39 1,06,294.38 2.15
Soyabean 42,170.23 44,158.54 1.05
Cotton(lint) 34,654.29 65,737.06 1.90
Bajra 26,007.15 23,895.52 0.92
Jowar 19,492.70 16,220.56 0.83
Gram 16,058.68 13,505.53 0.84
Groundnut 11,629.54 12,649.41 1.09
Rapeseed &Mustard 11,404.19 11,969.08 1.05
Maize 8,729.24 19,598.61 2.25
Top 10 States
Season
All
40K
30K
20K
10K
0K
Average of Area
Punjab
Maharashtra
West Bengal
Rajasthan
Gujarat
Madhya Pradesh
Andhra Pradesh
Uttar Pradesh
Karnataka
Bihar
State_Name


Clipboard
Data
Queries
Insert
Calculations
Sensitivity
CHINA
AFGHANISTAN
PAKISTAN
INDIA
NEPAL
BHUTAN
BANGLADESH
MYANMAR
LAOS
THAILAND
CAMBODIA
VIETNAM
OMAN
SRI LANKA
TIBET
Kabul
Peshawar
Islamabad
Lahore
Multan
Quetta
Karachi
New Delhi
Jaipur
Hyderabad
Ahmedabad
Mumbai
Pune
Bengaluru
Chennai
Hyderabad
Kolkata
Dhaka
Lhasa
Chengdu
Chongqing
Kunming
Guiyang
Hanoi
Haiphong
Yangon
Bangkok
Phnom Penh
Ho Chi Minh City
Vientiane
Arabian Sea
Bay of Bengal
Andaman Sea
Gulf of Thailand

3. Deployment to Power BI Service