

DOCUMENTATION FOR CHUBB CAPSTONE PROJECT ON CROP PRODUCTION DATASET

DATASET DESCRIPTION

The "Crop Production" dataset provides valuable insights into the agricultural landscape of India. It includes detailed information on crop production across various states, regions, and seasons over multiple years. The dataset offers comprehensive data for analysis and modeling of crop yield trends, crop diversity, and regional agricultural practices in India.

Features:

1. **State_Name:** The state or union territory in India where the crop was produced.
2. **District_Name:** The specific district within the state where the crop production data was collected.
3. **Season:** The agricultural season during which the crop was grown. This could include seasons like Kharif, Rabi, Summer, Autumn and Winter, as well as a whole year overview.
4. **Crop_Year:** The year of crop production data, spanning from 1997 to 2015.
5. **Crop:** The name of the crop grown in the respective state/district during the specific year and season. There are 124 distinct crops.
6. **Area:** The area (in hectares) where the crop was cultivated.
7. **Production:** The total quantity of the crop harvested, typically measured in tons or kilograms.

Columns: 7

Records: 246091

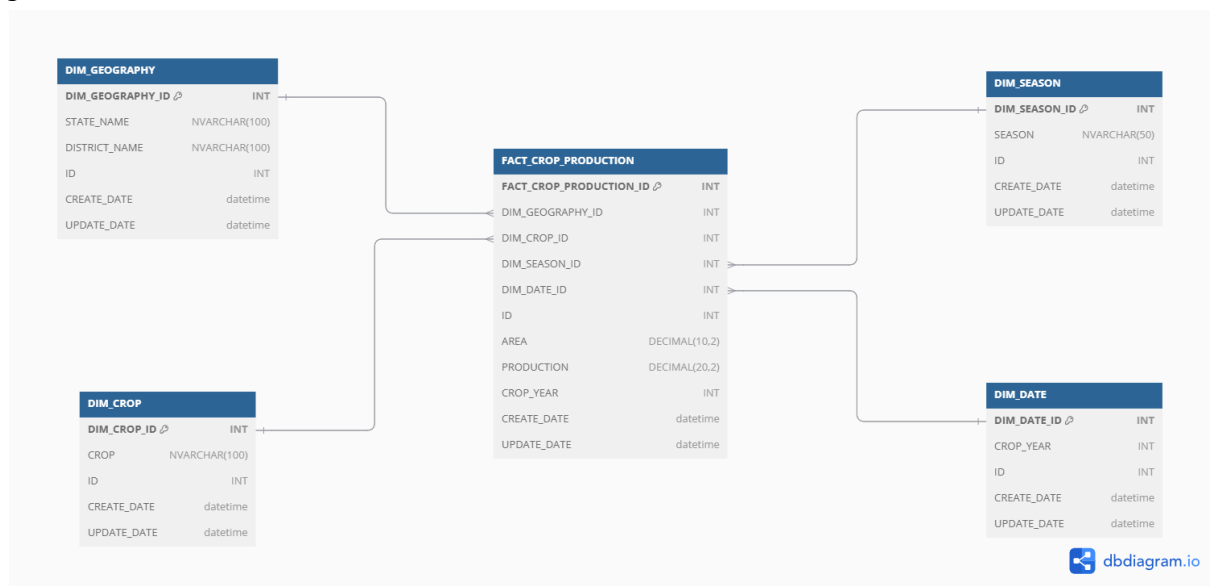
The database in the MSSQL server after Staging from the given CROP_PRODUCTION.csv flat file looks like

The screenshot shows a SQL Server query results window with the following data:

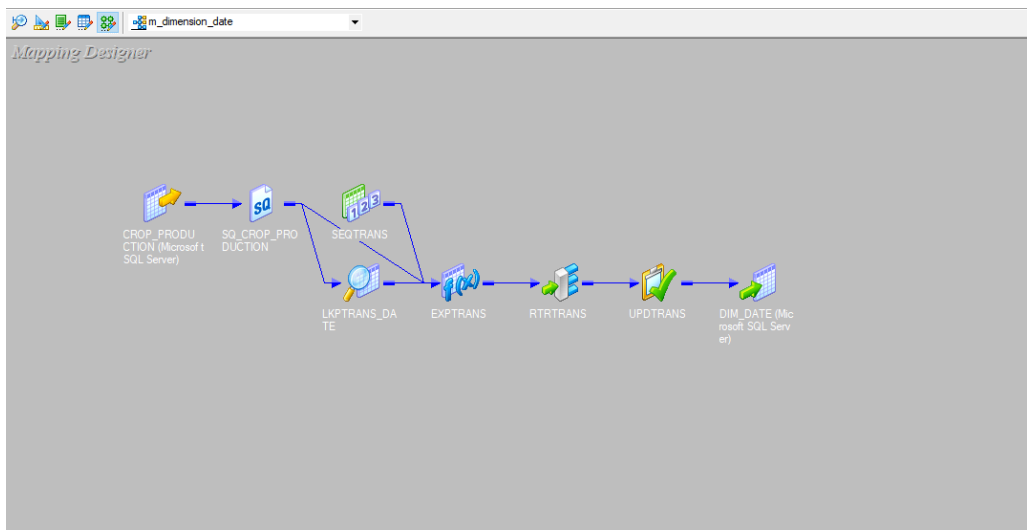
	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production	ID
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Areca nut	1254	2000	1
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2	1	2
3	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102	321	3
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176	641	4
5	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720	165	5
6	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Coconut	18168	65100000	6
7	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Dry ginger	36	100	7
8	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sugarcane	1	2	8
9	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sweet potato	5	15	9
10	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Tapioca	40	169	10
11	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Areca nut	1254	2061	11
12	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Other Kharif pulses	2	1	12
13	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Rice	83	300	13
14	Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Cashewnut	719	192	14
15	Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Coconut	18190	64430000	15
16	Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Dry ginger	46	100	16
17	Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Sugarcane	1	1	17
18	Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Sweet potato	11	33	18
19	Andaman and Nicobar Islands	NICOBARS	2002	Kharif	Rice	189.2	510.84	19
20	Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Areca nut	1258	2083	20
21	Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Banana	213	1278	21
22	Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Black pepper	63	13.5	22
23	Andaman and Nicobar Islands	NICOBARS	2002	Whole Year	Cashewnut	719	208	23

Query executed successfully. DESKTOP-DK4MRUA\SQLEXPRESS ... | INFA_REP (64) | INF_METADATA | 00:00:03 | 2,46,091 rows

The database has been modeled into 4 dimension tables namely dim_season, dim_crop, dim_geogrpahy, dim_date which are then used to make a fact table as shown in the DB model given below.



Mapping and data of DIM_DATE table



select * from tgt.DIM_date

100 %

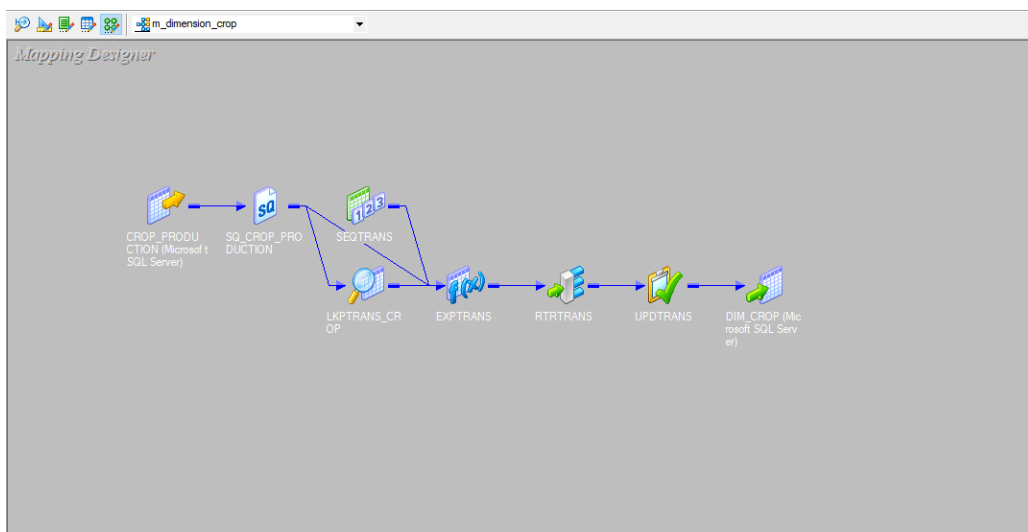
Results Messages

	DIM_DATE_ID	CROP_YEAR	ID	CREATE_DATE	UPDATE_DATE
1	1	1998	NULL	2024-12-10 00:42:46.737	NULL
2	2	2004	NULL	2024-12-10 00:42:46.737	NULL
3	3	2010	NULL	2024-12-10 00:42:46.737	NULL
4	4	2001	NULL	2024-12-10 00:42:46.737	NULL
5	5	2007	NULL	2024-12-10 00:42:46.737	NULL
6	6	2013	NULL	2024-12-10 00:42:46.737	NULL
7	7	2002	NULL	2024-12-10 00:42:46.737	NULL
8	8	2008	NULL	2024-12-10 00:42:46.737	NULL
9	9	1999	NULL	2024-12-10 00:42:46.737	NULL
10	10	2005	NULL	2024-12-10 00:42:46.737	NULL
11	11	2000	NULL	2024-12-10 00:42:46.737	NULL
12	12	2014	NULL	2024-12-10 00:42:46.737	NULL
13	13	1997	NULL	2024-12-10 00:42:46.737	NULL
14	14	2003	NULL	2024-12-10 00:42:46.737	NULL
15	15	2011	NULL	2024-12-10 00:42:46.737	NULL
16	16	2006	NULL	2024-12-10 00:42:46.737	NULL
17	17	2012	NULL	2024-12-10 00:42:46.737	NULL
18	18	2009	NULL	2024-12-10 00:42:46.737	NULL
19	19	2015	NULL	2024-12-10 00:42:46.737	NULL

Query executed successfully.

DESKTOP-DK4MRUA\SQLEXPRESS ... | INF_A_REP (64) | INF_METADATA | 00:00:00 | 19 rows

Mapping and data of DIM_CROP table

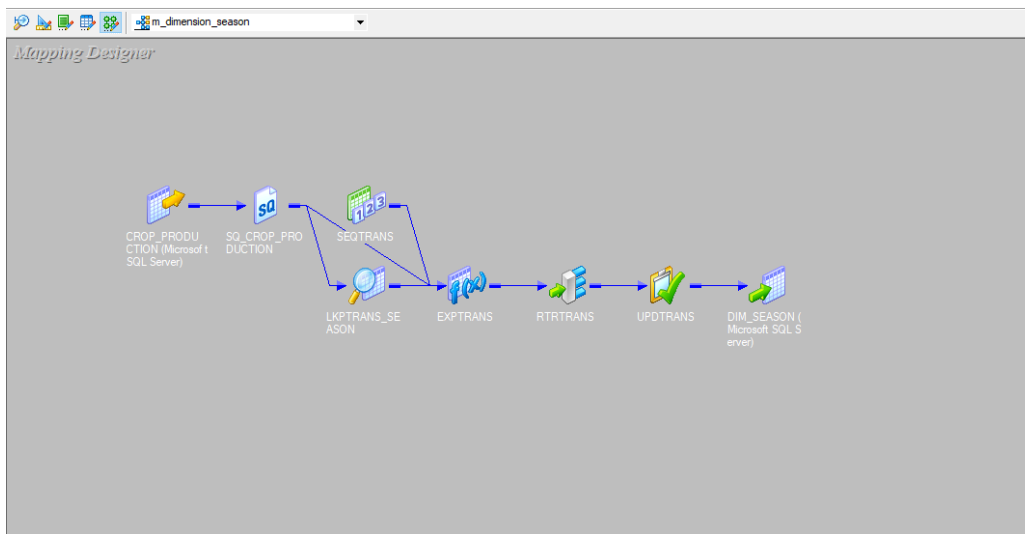


select * from tgt_DIM_CROP

	DIM_CROP_ID	CROP	ID	CREATE_DATE	UPDATE_DATE
1	1	Tapoca	NULL	2024-12-10 00:41:58.557	NULL
2	2	Soyabean	NULL	2024-12-10 00:41:58.557	NULL
3	3	Kapas	NULL	2024-12-10 00:41:58.557	NULL
4	4	Bhindi	NULL	2024-12-10 00:41:58.557	NULL
5	5	Cashewnut	NULL	2024-12-10 00:41:58.557	NULL
6	6	Ahar/Tur	NULL	2024-12-10 00:41:58.557	NULL
7	7	Beans & Mutter(Vegetable)	NULL	2024-12-10 00:41:58.557	NULL
8	8	Coffee	NULL	2024-12-10 00:41:58.557	NULL
9	9	Snak Guard	NULL	2024-12-10 00:41:58.557	NULL
10	10	Horse-gram	NULL	2024-12-10 00:41:58.557	NULL
11	11	Pome Fruit	NULL	2024-12-10 00:41:58.557	NULL
12	12	Varagu	NULL	2024-12-10 00:41:58.557	NULL
13	13	Arecanut	NULL	2024-12-10 00:41:58.557	NULL
14	14	Papaya	NULL	2024-12-10 00:41:58.557	NULL
15	15	Ricebean (magada)	NULL	2024-12-10 00:41:58.557	NULL
16	16	Caulflower	NULL	2024-12-10 00:41:58.557	NULL
17	17	Peas (vegetable)	NULL	2024-12-10 00:41:58.557	NULL
18	18	Total foodgran	NULL	2024-12-10 00:41:58.557	NULL
19	19	Potato	NULL	2024-12-10 00:41:58.557	NULL
20	20	Sannhamp	NULL	2024-12-10 00:41:58.557	NULL
21	21	Jack Fruit	NULL	2024-12-10 00:41:58.557	NULL
22	22	Onion	NULL	2024-12-10 00:41:58.557	NULL
23	23	Dry chillies	NULL	2024-12-10 00:41:58.557	NULL

Query executed successfully. DESKTOP-DK4MRUA\SQLEXPRESS ... INF_REP (64) INF_METADATA 00:00:00 124 rows

Mapping and data of DIM_SEASON table

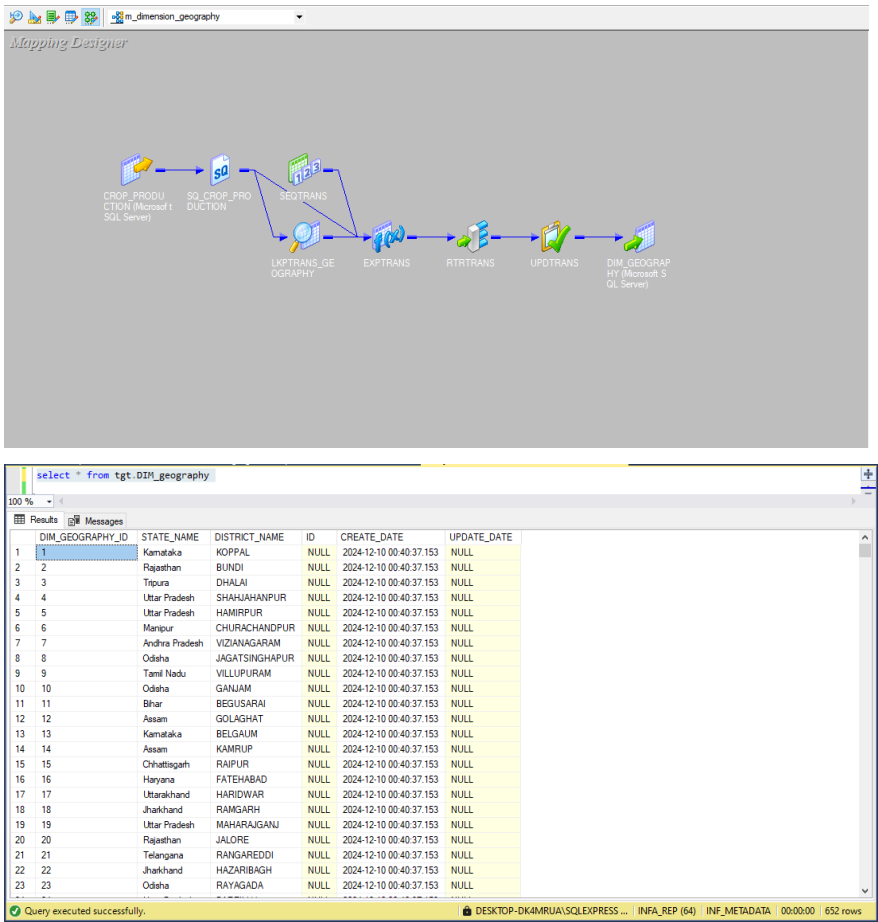


select * from tgt_DIM_season

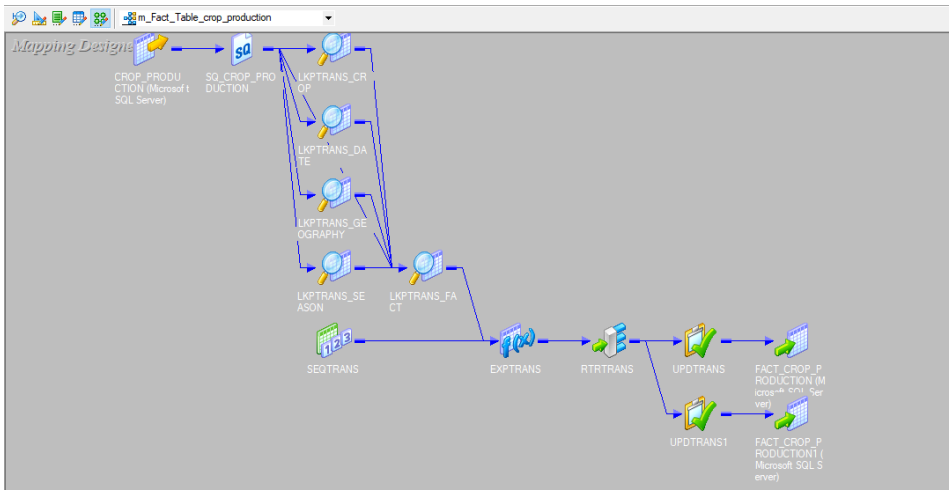
	DIM_SEASON_ID	SEASON	ID	CREATE_DATE	UPDATE_DATE
1	1	Whole Year	NULL	2024-12-10 00:15:42.897	NULL
2	2	Summer	NULL	2024-12-10 00:15:42.897	NULL
3	3	Rabi	NULL	2024-12-10 00:15:42.897	NULL
4	4	Autumn	NULL	2024-12-10 00:15:42.897	NULL
5	5	Kharif	NULL	2024-12-10 00:15:42.897	NULL
6	6	Winter	NULL	2024-12-10 00:15:42.897	NULL

Query executed successfully. DESKTOP-DK4MRUA\SQLEXPRESS ... INF_REP (64) INF_METADATA 00:00:00 6 rows

Mapping and data of DIM_GEOGRAPHY table



Mapping and data of FACT_CROP_PRODUCTION table



select * from tgt.FACT_CROP_PRODUCTION

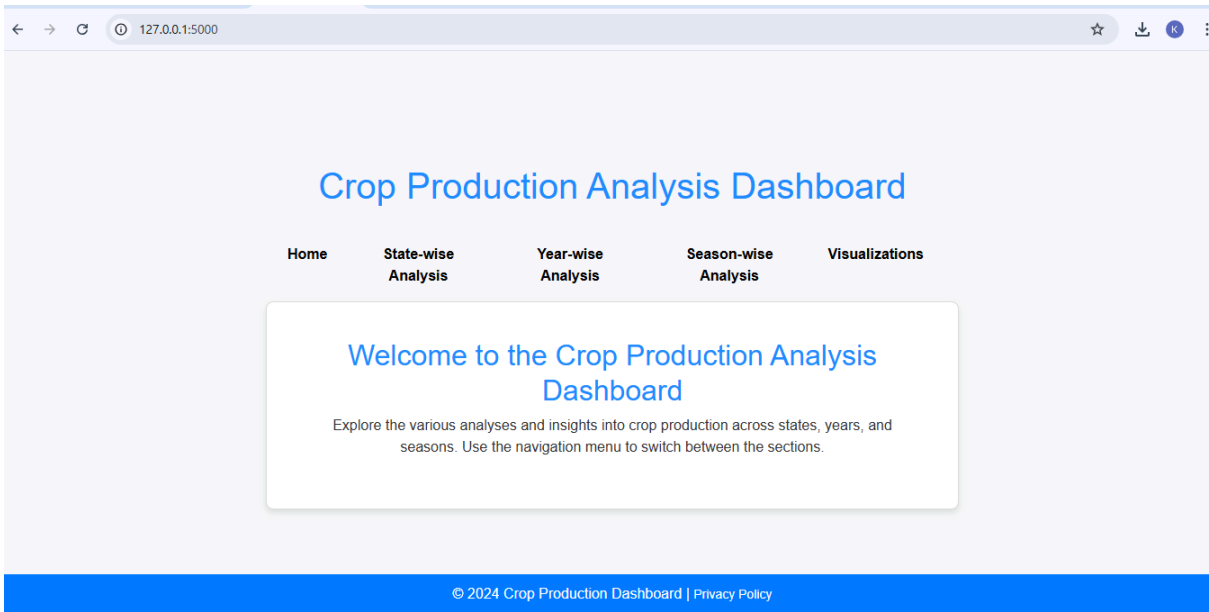
	FACT_CROP_PRODUCTION_ID	DIM_GEOGRAPHY_ID	DIM_CROP_ID	DIM_SEASON_ID	DIM_DATE_ID	ID	AREA	PRODUCTION	CROP_YEAR	CREATE_DATE	UPDAT
1	1	180	13	5	11	NULL	1254.00	2000.00	2000	2024-12-10 00:44:17.667	NULL
2	2	180	44	5	11	NULL	2.00	1.00	2000	2024-12-10 00:44:17.667	NULL
3	3	180	102	5	11	NULL	102.00	321.00	2000	2024-12-10 00:44:17.667	NULL
4	4	180	81	1	11	NULL	176.00	641.00	2000	2024-12-10 00:44:17.667	NULL
5	5	180	5	1	11	NULL	720.00	165.00	2000	2024-12-10 00:44:17.667	NULL
6	6	180	36	1	11	NULL	18168.00	65100000.00	2000	2024-12-10 00:44:17.667	NULL
7	7	180	101	1	11	NULL	36.00	100.00	2000	2024-12-10 00:44:17.667	NULL
8	8	180	27	1	11	NULL	1.00	2.00	2000	2024-12-10 00:44:17.667	NULL
9	9	180	62	1	11	NULL	5.00	15.00	2000	2024-12-10 00:44:17.667	NULL
10	10	180	1	1	11	NULL	40.00	169.00	2000	2024-12-10 00:44:17.667	NULL
11	11	180	13	5	4	NULL	1254.00	2061.00	2001	2024-12-10 00:44:17.667	NULL
12	12	180	44	5	4	NULL	2.00	1.00	2001	2024-12-10 00:44:17.667	NULL
13	13	180	102	5	4	NULL	83.00	300.00	2001	2024-12-10 00:44:17.667	NULL
14	14	180	5	1	4	NULL	719.00	192.00	2001	2024-12-10 00:44:17.667	NULL
15	15	180	36	1	4	NULL	18190.00	64430000.00	2001	2024-12-10 00:44:17.667	NULL
16	16	180	101	1	4	NULL	46.00	100.00	2001	2024-12-10 00:44:17.667	NULL
17	17	180	27	1	4	NULL	1.00	1.00	2001	2024-12-10 00:44:17.667	NULL
18	18	180	62	1	4	NULL	11.00	33.00	2001	2024-12-10 00:44:17.667	NULL
19	19	180	102	5	7	NULL	189.20	510.84	2002	2024-12-10 00:44:17.667	NULL
20	20	180	13	1	7	NULL	1258.00	2083.00	2002	2024-12-10 00:44:17.667	NULL
21	21	180	81	1	7	NULL	213.00	1278.00	2002	2024-12-10 00:44:17.667	NULL
22	22	180	99	1	7	NULL	63.00	13.50	2002	2024-12-10 00:44:17.667	NULL

Query executed successfully. DESKTOP-DK4MRUJA\SQLEXPRESS ... INF_A_REP (64) INF_METADATA 00:00:03 2,46,091 rows

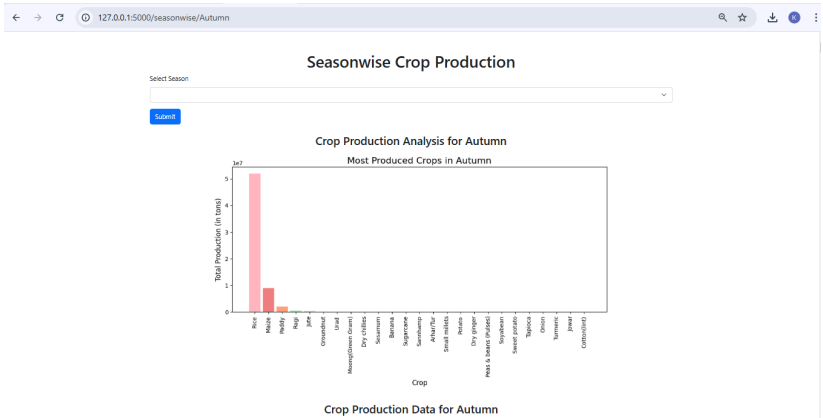
A comprehensive dashboard was developed using Flask to facilitate the visualization and analysis of crop production data. The dashboard enables users to explore the data through interactive plots and tables with the following key features:

- 1. **State-wise Analysis:** Gain insights into crop production trends across different states.
- 2. **Year-wise Analysis:** Explore changes in production over the years to identify historical patterns and growth trajectories.
- 3. **Season-wise Analysis:** Examine production data categorized by agricultural seasons, such as Rabi, Kharif, and others.

The home screen for the application.



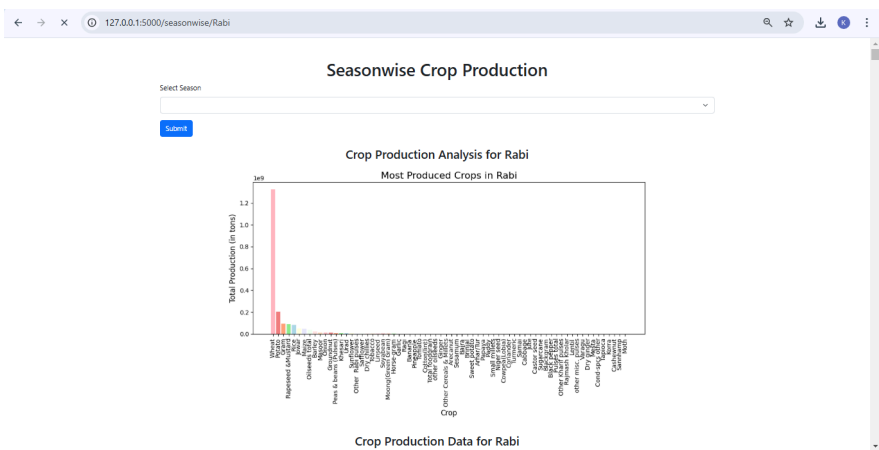
The analysis generated for Autumn season shows that Rice is the most produced crop during Autumn season.



Crop Production Data for Autumn

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
1	Andaman and Nicobar Islands	NICOBARS	2010	Autumn	Rice	4	10
2	Andaman and Nicobar Islands	NICOBARS	2010	Autumn	Sugarcane	13	42
3	Andaman and Nicobar Islands	NORTH AND MIDDLE ANDAMAN	2010	Autumn	Rice	6791	20118
4	Andaman and Nicobar Islands	NORTH AND MIDDLE ANDAMAN	2010	Autumn	Sugarcane	73	889
5	Andaman and Nicobar Islands	SOUTH ANDAMAN	2010	Autumn	Rice	1596	3788
6	Andaman and Nicobar Islands	SOUTH ANDAMAN	2010	Autumn	Sugarcane	68	402
7	West Bengal	24 PARAGANAS NORTH	1997	Autumn	Rice	15017	33840
8	Assam	CACHAR	2002	Autumn	Rice	9010	17330
9	Assam	CACHAR	2003	Autumn	Rice	9230	18799
10	Assam	CACHAR	2004	Autumn	Rice	10174	2140
11	Assam	CACHAR	2005	Autumn	Rice	9411	18157
12	Assam	CACHAR	2006	Autumn	Rice	4070	6344
13	Assam	CACHAR	2007	Autumn	Rice	10130	16238
14	Assam	CACHAR	2008	Autumn	Rice	8350	13008
15	Assam	CACHAR	2009	Autumn	Rice	8276	16457
16	Assam	CACHAR	2010	Autumn	Rice	8359	16087
17	Assam	CACHAR	2011	Autumn	Paddy	7700	13116
18	Assam	CACHAR	2011	Autumn	Rice	7700	13116
19	Bihar	PASHCHIM CHAMPARAN	2014	Autumn	Maize	2073	4874
20	Bihar	PASHCHIM CHAMPARAN	2014	Autumn	Rice	43085	78558
21	Bihar	PAITIA	1997	Autumn	Rice	1179	1594

The analysis generated for Rabi season shows that Wheat is the most produced crop during Rabi season.

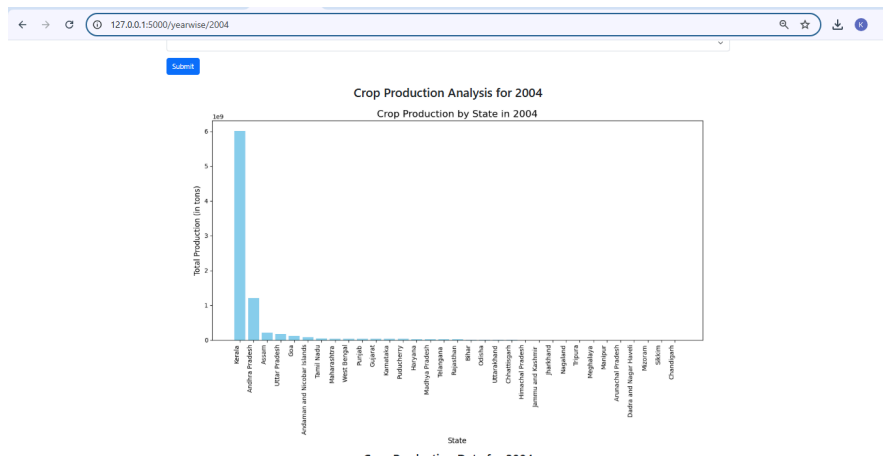


127.0.0.1:5000/seasonwise/Rabi

Crop Production Data for Rabi

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
1	Karnataka	GADAG	2008	Rabi	Horse-gram	1235	422
2	Karnataka	GADAG	2008	Rabi	Jowar	65046	62806
3	Karnataka	GADAG	2008	Rabi	Linseed	822	472
4	Karnataka	GADAG	2008	Rabi	Maise	2607	6013
5	Karnataka	GADAG	2008	Rabi	Onion	242	1616
6	Karnataka	GADAG	2008	Rabi	Other Rabi pulses	341	85
7	Karnataka	GADAG	2008	Rabi	Rapeseed & Mustard	161	65
8	Karnataka	GADAG	2008	Rabi	Safflower	4391	4292
9	Karnataka	GADAG	2008	Rabi	Sunflower	61430	22364
10	Karnataka	GADAG	2008	Rabi	Wheat	33917	22306
11	Karnataka	GADAG	2009	Rabi	Dry chilies	22	66
12	Karnataka	GADAG	2009	Rabi	Gram	1604	911
13	Karnataka	GADAG	2009	Rabi	Groundnut	9446	7394
14	Karnataka	GADAG	2009	Rabi	Horse-gram	1604	911
15	Karnataka	GADAG	2009	Rabi	Jowar	59056	39711
16	Karnataka	GADAG	2009	Rabi	Linseed	510	312
17	Karnataka	GADAG	2009	Rabi	Maise	4120	3801
18	Karnataka	GADAG	2009	Rabi	Onion	433	2565
19	Karnataka	GADAG	2009	Rabi	Other Rabi pulses	695	192
20	Karnataka	GADAG	2009	Rabi	Rapeseed & Mustard	158	66
21	Karnataka	GADAG	2009	Rabi	Safflower	4020	3762

The analysis generated for the year 2004 shows that Kerala produced the most crop in 2004.



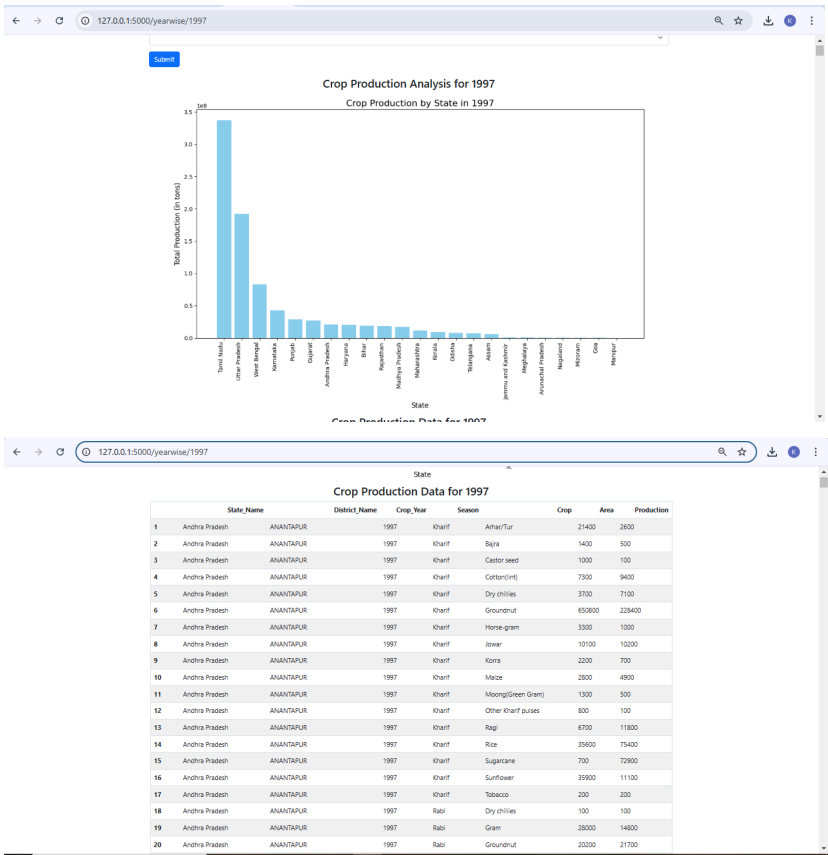
← → ↺ 127.0.0.1500/yearwise/2004 🔍 ⚙️

State

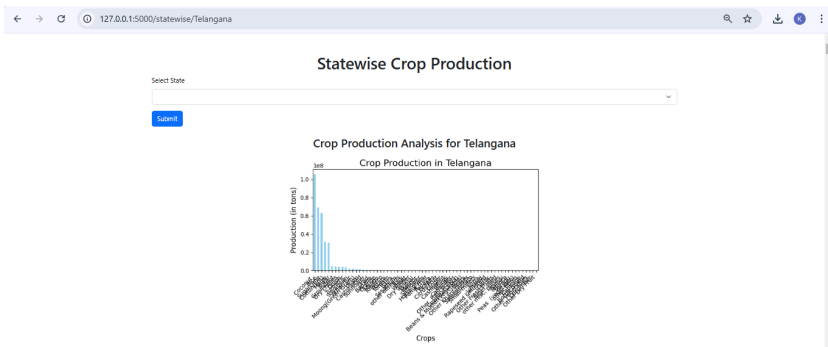
Crop Production Data for 2004

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
1	Maharashtra	GONDIA	2004	Kharif	Ashar/Tur	4500	2800
2	Kerala	THIRUVANANTHAPURAM	2004	Autumn	Rice	2621	6423
3	Kerala	THIRUVANANTHAPURAM	2004	Summer	Rice	2	3
4	Kerala	THIRUVANANTHAPURAM	2004	Whole Year	Arecanut	1263	547
5	Kerala	THIRUVANANTHAPURAM	2004	Whole Year	Banana	2492	15903
6	Kerala	THIRUVANANTHAPURAM	2004	Whole Year	Black pepper	7320	2028
7	Kerala	THIRUVANANTHAPURAM	2004	Whole Year	Cashewnut	2061	761
8	Kerala	THIRUVANANTHAPURAM	2004	Whole Year	Coconut	84073	56800000
9	Kerala	THIRUVANANTHAPURAM	2004	Whole Year	Dry ginger	103	306
10	Kerala	THIRUVANANTHAPURAM	2004	Whole Year	Sugarcane	3	nan
11	Kerala	THIRUVANANTHAPURAM	2004	Whole Year	Sweet potato	31	351
12	Kerala	THIRUVANANTHAPURAM	2004	Whole Year	Tapioca	19833	467006
13	Kerala	THIRUVANANTHAPURAM	2004	Whole Year	Turmeric	59	85
14	Kerala	THIRUVANANTHAPURAM	2004	Winter	Rice	2573	5930
15	Himachal Pradesh	KANGRA	2004	Kharif	Bajra	210	111
16	Himachal Pradesh	KANGRA	2004	Kharif	Dry ginger	28	200
17	Himachal Pradesh	KANGRA	2004	Kharif	Maze	58050	90940
18	Himachal Pradesh	KANGRA	2004	Kharif	Rice	37079	48660
19	Himachal Pradesh	KANGRA	2004	Rabi	Barley	3412	3200
20	Himachal Pradesh	KANGRA	2004	Rabi	Wheat	94424	170283

The analysis generated for the year 1997 shows that Tamilnadu produced the most crop in 1997.



The analysis generated for the state Telangana shows that Coconut is the most produced crop in 2004.

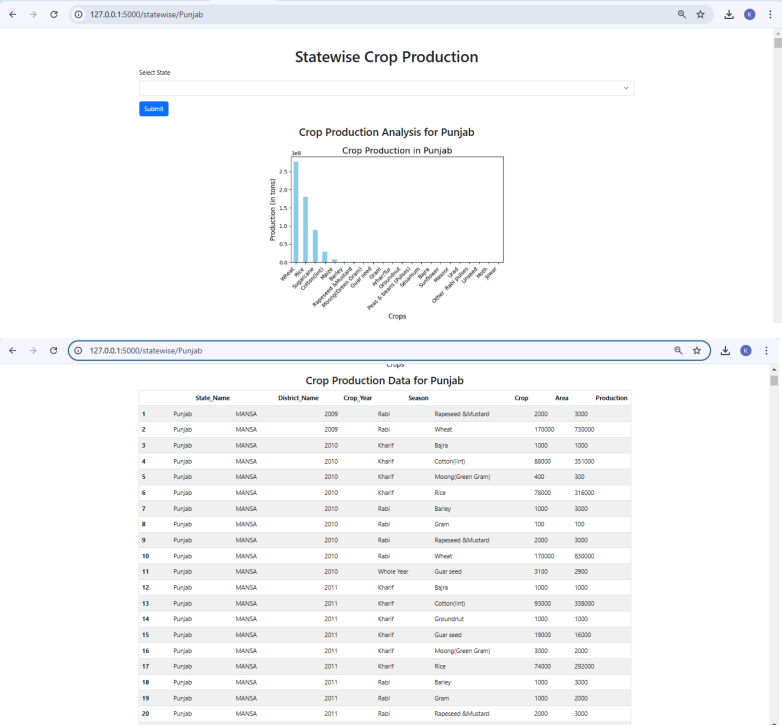


127.0.0.1:5000/statewise/Telangana

Crop Production Data for Telangana

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
1	Telangana	MEDAK	2013	Rabi	Wheat	2110	1711
2	Telangana	MEDAK	2013	Whole Year	Castor seed	1388	982
3	Telangana	MEDAK	2013	Whole Year	Coriander	471	337
4	Telangana	MEDAK	2013	Whole Year	Dry ginger	1713	12346
5	Telangana	MEDAK	2013	Whole Year	Potato	2997	38830
6	Telangana	MEDAK	2013	Whole Year	Safflower	3915	2388
7	Telangana	MEDAK	2013	Whole Year	Soyabean	15597	21758
8	Telangana	MEDAK	2014	Kharif	Arhar(Tur)	22001	6996
9	Telangana	MEDAK	2014	Kharif	Bajra	284	275
10	Telangana	MEDAK	2014	Kharif	Baranai	904	37703
11	Telangana	MEDAK	2014	Kharif	Castor seed	423	244
12	Telangana	MEDAK	2014	Kharif	Cotton(int)	133703	233587
13	Telangana	MEDAK	2014	Kharif	Cowpea(Lobia)	473	505
14	Telangana	MEDAK	2014	Kharif	Dry chilies	291	134
15	Telangana	MEDAK	2014	Kharif	Ginger	2164	16373
16	Telangana	MEDAK	2014	Kharif	Groundnut	35	31
17	Telangana	MEDAK	2014	Kharif	Jowar	7006	7714
18	Telangana	MEDAK	2014	Kharif	Millet	110476	170464
19	Telangana	MEDAK	2014	Kharif	Moong(Green Gram)	14940	8232
20	Telangana	MEDAK	2014	Kharif	Onion	148	2301
21	Telangana	MEDAK	2014	Kharif	Other Kharif Cereals	18	14

The analysis generated for the state Punjab shows that Wheat is the most produced crop in Punjab.



An additional page on the dashboard provides enhanced visualizations to deliver deeper insights into the crop production data. It includes visual representations of the distribution of crops across the six agricultural seasons, offering a clear understanding of seasonal trends. Summarized data highlights, such as the total crop count, provide an at-a-glance overview of overall production patterns. Furthermore, the page incorporates an interactive zoom feature, allowing users to closely examine specific visualizations for finer details. These additions enhance the dashboard's analytical capabilities and user experience, making it a valuable tool for exploring and interpreting data.

