Problem description:

In agriculture sector where farmers and agribusinesses have to make innumerable decisions every day and intricate complexities involves the various factors influencing them. An essential issue for agricultural planning intention is the accurate yield estimation for the numerous crops involved in the planning. Data mining techniques are necessary approach for accomplishing practical and effective solutions for this problem. Agriculture has been an obvious target for big data. Environmental conditions, variability in soil, input levels, combinations and commodity prices have made it all the more relevant for farmers to use information and get help to make critical farming decisions. Mining the large amount of existing crop, soil and climatic data, and analysing new, non-experimental data optimizes the production and makes agriculture more resilient to climatic change.

Dataset Used: crop_production.csv

State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254	2000
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2	1
Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102	321
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176	641
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720	165
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Coconut	18168	65100000
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Dry ginger	36	100
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sugarcane	1	2
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sweet potato	5	15
Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Tapioca	40	169
Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Arecanut	1254	2061
Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Other Kharif pulses	2	1

Pig commands:

pig -x local

agriculture= load '/home/hadoop/Documents/miniproject/crop_production.csv' using PigStorage(',') as (State_Name:chararray,District_Name:chararray,Crop_Year:int,Season:chararray,Crop:chararray,Area:int,Production:int);

describe agriculture;

dump agriculture;

statewise_group = group agriculture by State_Name;

dump statewise_group;

store statewise_group into '/home/hadoop/Documents/miniproject/statewise_output';

filter_district = filter agriculture by District_Name == 'MADURAI';

dump filter_district;

store filter_district into '/home/hadoop/Documents/miniproject/filter_output;

order_year = order agriculture by Crop_Year desc;

dump order_year;

foreach_crop = foreach agriculture generate Crop;

dump foreach_crop;

Output:

```
hadoop@ab1-cse107: ~

hadoop@ab1-cse107:~$ pig -x local
2022-11-05 09:35:56,630 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2022-11-05 09:35:56,631 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
2022-11-05 09:35:56,665 [main] INFO  org.apache.pig.Main - Apache Pig version 0.17.0 (r1797386) compiled Jun 02 2017, 15:41:58
2022-11-05 09:35:56,665 [main] INFO  org.apache.pig.Main - Logging error messages to: /home/hadoop/pig_1667621156663.log
2022-11-05 09:35:56,679 [main] INFO  org.apache.pig.impl.util.Utils - Default bootup file /home/hadoop/.pigbootup not found
2022-11-05 09:35:56,740 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapredu
ce.jobtracker.address
2022-11-05 09:35:56,742 [main] INFO  org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: fil
e:///
2022-11-05 09:35:56,804 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.
bytes-per-checksum
2022-11-05 09:35:56,817 [main] INFO  org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-92cc4cb0-d19c-45f0-9874-cad10aa0fd5
4
2022-11-05 09:35:56,817 [main] WARN  org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
grunt> agriculture= load '/home/hadoop/Documents/miniproject/crop_production.csv' using PigStorage(',') as (State_Name:chararray,District_Name
:chararray,Crop_Year:int,Season:chararray,Crop:chararray,Area:int,Production:int);
2022-11-05 09:36:05,908 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.
bytes-per-checksum
grunt> describe agriculture;
agriculture: {State_Name: chararray,District_Name: chararray,Crop_Year: int,Season: chararray,Crop: chararray,Area: int,Production: int}
grunt> dump agriculture;
```

```
hadoop@ab1-cse107: ~

(Uttarakhand,UTTAR KASHI,2010,Rabi        ,Barley,257,355)
(Uttarakhand,UTTAR KASHI,2010,Rabi        ,Masoor,249,303)
(Uttarakhand,UTTAR KASHI,2010,Rabi        ,Peas & beans (Pulses),539,367)
(Uttarakhand,UTTAR KASHI,2010,Rabi        ,Potato,278,1679)
(Uttarakhand,UTTAR KASHI,2010,Rabi        ,Rapeseed &Mustard,861,491)
(Uttarakhand,UTTAR KASHI,2010,Rabi        ,Wheat,12126,20489)
(Uttarakhand,UTTAR KASHI,2010,Whole Year ,Dry ginger,3,29)
(Uttarakhand,UTTAR KASHI,2010,Whole Year ,Garlic,18,28)
(Uttarakhand,UTTAR KASHI,2010,Whole Year ,Onion,47,306)
(Uttarakhand,UTTAR KASHI,2010,Whole Year ,Tobacco,1,3)
(Uttarakhand,UTTAR KASHI,2011,Kharif     ,Arhar/Tur,380,312)
(Uttarakhand,UTTAR KASHI,2011,Kharif     ,Horse-gram,861,832)
(Uttarakhand,UTTAR KASHI,2011,Kharif     ,Maize,462,577)
(Uttarakhand,UTTAR KASHI,2011,Kharif     ,Other Kharif pulses,1818,1330)
(Uttarakhand,UTTAR KASHI,2011,Kharif     ,Potato,2449,17437)
(Uttarakhand,UTTAR KASHI,2011,Kharif     ,Ragi,5388,9990)
(Uttarakhand,UTTAR KASHI,2011,Kharif     ,Rice,10619,17165)
(Uttarakhand,UTTAR KASHI,2011,Kharif     ,Sesamum,673,169)
(Uttarakhand,UTTAR KASHI,2011,Kharif     ,Small millets,3741,4693)
(Uttarakhand,UTTAR KASHI,2011,Kharif     ,Soyabean,414,422)
(Uttarakhand,UTTAR KASHI,2011,Kharif     ,Sunflower,1,)
(Uttarakhand,UTTAR KASHI,2011,Kharif     ,Urad,816,851)
(Uttarakhand,UTTAR KASHI,2011,Rabi       ,Barley,174,246)
(Uttarakhand,UTTAR KASHI,2011,Rabi       ,Gram,1,1)
(Uttarakhand,UTTAR KASHI,2011,Rabi       ,Masoor,244,178)
(Uttarakhand,UTTAR KASHI,2011,Rabi       ,Peas & beans (Pulses),347,228)
(Uttarakhand,UTTAR KASHI,2011,Rabi       ,Potato,86,543)
(Uttarakhand,UTTAR KASHI,2011,Rabi       ,Rapeseed &Mustard,904,497)
(Uttarakhand,UTTAR KASHI,2011,Rabi       ,Wheat,11503,18907)
(Uttarakhand,UTTAR KASHI,2011,Whole Year ,Garlic,21,74)
(Uttarakhand,UTTAR KASHI,2011,Whole Year ,Ginger,2,20)
(Uttarakhand,UTTAR KASHI,2011,Whole Year ,Onion,8,52)
(Uttarakhand,UTTAR KASHI,2011,Whole Year ,Tobacco,1,1)
(Uttarakhand,UTTAR KASHI,2012,Kharif     ,Arhar/Tur,360,316)
(Uttarakhand,UTTAR KASHI,2012,Kharif     ,Horse-gram,692,623)
(Uttarakhand,UTTAR KASHI,2012,Kharif     ,Maize,588,609)
(Uttarakhand,UTTAR KASHI,2012,Kharif     ,Other Kharif pulses,1575,1098)
(Uttarakhand,UTTAR KASHI,2012,Kharif     ,other oilseeds,5,1)
(Uttarakhand,UTTAR KASHI,2012,Kharif     ,Potato,2055,25661)
```

ARS,2010,Rabi   ,Cashewnut,1000,260),(Andaman and Nicobar Islands,NICOBARS,2010,Rabi   ,Black pepper,23,8),(Andaman and Nicobar Island
s,NICOBARS,2010,Rabi   ,Arecanut,944,1610),(Andaman and Nicobar Islands,NICOBARS,2010,Autumn   ,Sugarcane,13,41),(Andaman and Nicobar Is
lands,NICOBARS,2010,Autumn   ,Rice,3,10),(Andaman and Nicobar Islands,NICOBARS,2006,Whole Year ,Turmeric,2,1),(Andaman and Nicobar Islands,N
ICOBARS,2006,Whole Year ,Tapioca,52,450),(Andaman and Nicobar Islands,NICOBARS,2006,Whole Year ,Sweet potato,18,86),(Andaman and Nicobar Islan
ds,NICOBARS,2006,Whole Year ,Sugarcane,0,0),(Andaman and Nicobar Islands,NICOBARS,2006,Whole Year ,Dry ginger,6,21),(Andaman and Nicobar Islan
ds,NICOBARS,2006,Whole Year ,Dry chillies,3,9),(Andaman and Nicobar Islands,NICOBARS,2006,Whole Year ,Coconut ,14358,59370000),(Andaman and Ni
cobar Islands,NICOBARS,2006,Whole Year ,Cashewnut,450,85),(Andaman and Nicobar Islands,NICOBARS,2006,Whole Year ,Black pepper,22,0),(Andaman a
nd Nicobar Islands,NICOBARS,2006,Whole Year ,Banana,198,976),(Andaman and Nicobar Islands,NICOBARS,2006,Whole Year ,Arecanut,896,477),(Andaman
 and Nicobar Islands,NICOBARS,2005,Whole Year ,Turmeric,10,70),(Andaman and Nicobar Islands,NICOBARS,2005,Whole Year ,Dry chillies,17,),(Andam
an and Nicobar Islands,NICOBARS,2005,Whole Year ,Coconut ,13876,52030000),(Andaman and Nicobar Islands,NICOBARS,2005,Whole Year ,Cashewnut,450
,55),(Andaman and Nicobar Islands,NICOBARS,2005,Whole Year ,Black pepper,41,0),(Andaman and Nicobar Islands,NICOBARS,2005,Whole Year ,Banana,2
2,582),(Andaman and Nicobar Islands,NICOBARS,2005,Whole Year ,Arecanut,795,),(Andaman and Nicobar Islands,NICOBARS,2005,Kharif   ,Rice,2,12)
,(Andaman and Nicobar Islands,NICOBARS,2004,Whole Year ,other oilseeds,1,0),(Andaman and Nicobar Islands,NICOBARS,2004,Whole Year ,Dry ginger,
118,366),(Andaman and Nicobar Islands,NICOBARS,2004,Whole Year ,Dry chillies,60,145),(Andaman and Nicobar Islands,NICOBARS,2004,Whole Year ,Co
conut ,18394,52380000),(Andaman and Nicobar Islands,NICOBARS,2004,Whole Year ,Cashewnut,683,197),(Andaman and Nicobar Islands,NICOBARS,2004,Wh
ole Year ,Black pepper,85,15),(Andaman and Nicobar Islands,NICOBARS,2004,Whole Year ,Banana,66,404),(Andaman and Nicobar Islands,NICOBARS,2004
,Whole Year ,Arecanut,1264,805),(Andaman and Nicobar Islands,NICOBARS,2004,Kharif   ,Rice,52,72),(Andaman and Nicobar Islands,NICOBARS,2003,
Whole Year ,other oilseeds,4,0),(Andaman and Nicobar Islands,NICOBARS,2003,Whole Year ,Dry ginger,102,326),(Andaman and Nicobar Islands,NICOBA
RS,2003,Whole Year ,Dry chillies,60,102),(Andaman and Nicobar Islands,NICOBARS,2003,Whole Year ,Coconut ,18284,68580000),(Andaman and Nicobar
Islands,NICOBARS,2003,Whole Year ,Cashewnut,717,208),(Andaman and Nicobar Islands,NICOBARS,2003,Whole Year ,Black pepper,75,15),(Andaman and N
icobar Islands,NICOBARS,2003,Whole Year ,Banana,266,1763),(Andaman and Nicobar Islands,NICOBARS,2003,Whole Year ,Arecanut,1261,1525),(Andaman
and Nicobar Islands,NICOBARS,2003,Kharif   ,Rice,52,90),(Andaman and Nicobar Islands,NICOBARS,2002,Whole Year ,Sugarcane,5,40),(Andaman and
Nicobar Islands,NICOBARS,2002,Whole Year ,Dry ginger,47,133),(Andaman and Nicobar Islands,NICOBARS,2002,Whole Year ,Dry chillies,413,28),(Anda
man and Nicobar Islands,NICOBARS,2002,Whole Year ,Coconut ,18240,67490000),(Andaman and Nicobar Islands,NICOBARS,2002,Whole Year ,Cashewnut,71
9,208),(Andaman and Nicobar Islands,NICOBARS,2002,Whole Year ,Black pepper,63,13),(Andaman and Nicobar Islands,NICOBARS,2002,Whole Year ,Banan
a,213,1278),(Andaman and Nicobar Islands,NICOBARS,2002,Whole Year ,Arecanut,1258,2083),(Andaman and Nicobar Islands,NICOBARS,2002,Kharif   ,
Rice,189,510),(Andaman and Nicobar Islands,NICOBARS,2001,Whole Year ,Sweet potato,11,33),(Andaman and Nicobar Islands,NICOBARS,2001,Whole Year
,Sugarcane,1,1),(Andaman and Nicobar Islands,NICOBARS,2001,Whole Year ,Dry ginger,46,100),(Andaman and Nicobar Islands,NICOBARS,2001,Whole Ye
ar ,Coconut ,18190,64430000),(Andaman and Nicobar Islands,NICOBARS,2001,Whole Year ,Cashewnut,719,192),(Andaman and Nicobar Islands,NICOBARS,2
001,Kharif   ,Rice,83,300),(Andaman and Nicobar Islands,NICOBARS,2001,Kharif   ,Other Kharif pulses,2,1),(Andaman and Nicobar Islands,NICO
BARS,2001,Kharif   ,Arecanut,1254,2061),(Andaman and Nicobar Islands,NICOBARS,2000,Whole Year ,Tapioca,40,169),(Andaman and Nicobar Islands,
NICOBARS,2000,Whole Year ,Sweet potato,5,15),(Andaman and Nicobar Islands,NICOBARS,2000,Whole Year ,Sugarcane,1,2),(Andaman and Nicobar Island
s,NICOBARS,2000,Whole Year ,Dry ginger,36,100),(Andaman and Nicobar Islands,NICOBARS,2000,Whole Year ,Coconut ,18168,65100000),(Andaman and Ni
cobar Islands,NICOBARS,2000,Whole Year ,Cashewnut,720,165),(Andaman and Nicobar Islands,NICOBARS,2000,Whole Year ,Banana,176,641),(Andaman and
 Nicobar Islands,NICOBARS,2000,Kharif   ,Rice,102,321),(Andaman and Nicobar Islands,NICOBARS,2000,Kharif   ,Other Kharif pulses,2,1),(Anda
man and Nicobar Islands,NICOBARS,2000,Kharif   ,Arecanut,1254,2000),(Andaman and Nicobar Islands,SOUTH ANDAMANS,2010,Whole Year ,Tapioca,22,
220)})
grunt> []

2022-11-05 10:47:51,920 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.
bytes-per-checksum
grunt> filter_district = filter agriculture by District_Name == 'MADURAI';
grunt> dump filter_district;
2022-11-05 10:48:53,584 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.
bytes-per-checksum
2022-11-05 10:48:53,597 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER
2022-11-05 10:48:53,611 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.
bytes-per-checksum
2022-11-05 10:48:53,635 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyP
rune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, NestedLimitOptimizer, P
artitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2022-11-05 10:48:53,682 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (PS Old Gen) of size 699400192 to monitor
. collectionUsageThreshold = 489580128, usageThreshold = 489580128
2022-11-05 10:48:53,720 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 1
00 optimistic? false
2022-11-05 10:48:53,734 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before op
timization: 1
2022-11-05 10:48:53,734 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after opt
imization: 1
2022-11-05 10:48:53,749 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.
bytes-per-checksum
2022-11-05 10:48:53,800 [main] INFO org.apache.hadoop.metrics2.impl.MetricsConfig - Loaded properties from hadoop-metrics2.properties
2022-11-05 10:48:53,892 [main] INFO org.apache.hadoop.metrics2.impl.MetricsSystemImpl - Scheduled Metric snapshot period at 10 second(s).
2022-11-05 10:48:53,892 [main] INFO org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system started
2022-11-05 10:48:53,909 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2022-11-05 10:48:53,912 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.reduce.markreset.buffer.percent is deprecat
ed. Instead, use mapreduce.reduce.markreset.buffer.percent
2022-11-05 10:48:53,912 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markr
eset.buffer.percent is not set, set to default 0.3
2022-11-05 10:48:53,914 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.output.compress is deprecated. Instead, use map
reduce.output.fileoutputformat.compress
2022-11-05 10:48:53,922 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - Setting up single store
 job
2022-11-05 10:48:53,931 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Key [pig.schematuple] is false, will not generate code.
2022-11-05 10:48:53,931 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Starting process to move generated code to distributed cacche
2022-11-05 10:48:53,931 [main] INFO org.apache.pig.data.SchemaTupleFrontend - Distributed cache not supported or needed in local mode. Settin
g key [pig.schematuple.local.dir] with code temp directory: /tmp/1667625533931-0
2022-11-05 10:48:53,961 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 1 map-reduce job(s) wait

```
(Tamil Nadu,MADURAI,1997,Kharif        ,Horse-gram,117,60)
(Tamil Nadu,MADURAI,1997,Kharif        ,Onion,1109,7011)
(Tamil Nadu,MADURAI,1997,Kharif        ,Sesamum,2403,550)
(Tamil Nadu,MADURAI,1997,Kharif        ,Small millets,1105,1150)
(Tamil Nadu,MADURAI,1997,Whole Year ,Arhar/Tur,1498,750)
(Tamil Nadu,MADURAI,1997,Whole Year ,Bajra,1020,1290)
(Tamil Nadu,MADURAI,1997,Whole Year ,Banana,2077,139430)
(Tamil Nadu,MADURAI,1997,Whole Year ,Cashewnut,232,140)
(Tamil Nadu,MADURAI,1997,Whole Year ,Coriander,253,120)
(Tamil Nadu,MADURAI,1997,Whole Year ,Cotton(lint),13061,8580)
(Tamil Nadu,MADURAI,1997,Whole Year ,Dry chillies,505,370)
(Tamil Nadu,MADURAI,1997,Whole Year ,Gram,19,10)
(Tamil Nadu,MADURAI,1997,Whole Year ,Groundnut,11983,16580)
(Tamil Nadu,MADURAI,1997,Whole Year ,Jowar,9626,8860)
(Tamil Nadu,MADURAI,1997,Whole Year ,Maize,306,620)
(Tamil Nadu,MADURAI,1997,Whole Year ,Moong(Green Gram),5158,1610)
(Tamil Nadu,MADURAI,1997,Whole Year ,Pulses total,10273,3400)
(Tamil Nadu,MADURAI,1997,Whole Year ,Ragi,298,720)
(Tamil Nadu,MADURAI,1997,Whole Year ,Rice,83061,314380)
(Tamil Nadu,MADURAI,1997,Whole Year ,Small millets,720,20)
(Tamil Nadu,MADURAI,1997,Whole Year ,Sugarcane,8108,18223000)
(Tamil Nadu,MADURAI,1997,Whole Year ,Sunflower,967,860)
(Tamil Nadu,MADURAI,1997,Whole Year ,Sweet potato,93,1350)
(Tamil Nadu,MADURAI,1997,Whole Year ,Tapioca,56,2100)
(Tamil Nadu,MADURAI,1997,Whole Year ,Total foodgrain,109708,332710)
(Tamil Nadu,MADURAI,1997,Whole Year ,Turmeric,7,40)
(Tamil Nadu,MADURAI,1997,Whole Year ,Urad,1248,580)
(Tamil Nadu,MADURAI,1998,Kharif        ,Arhar/Tur,4055,4465)
(Tamil Nadu,MADURAI,1998,Kharif        ,Bajra,3412,6225)
(Tamil Nadu,MADURAI,1998,Kharif        ,Cotton(lint),17147,26970)
(Tamil Nadu,MADURAI,1998,Kharif        ,Groundnut,14722,25578)
(Tamil Nadu,MADURAI,1998,Kharif        ,Horse-gram,153,77)
(Tamil Nadu,MADURAI,1998,Kharif        ,Jowar,12744,14981)
(Tamil Nadu,MADURAI,1998,Kharif        ,Maize,1025,2066)
(Tamil Nadu,MADURAI,1998,Kharif        ,Moong(Green Gram),13040,7159)
(Tamil Nadu,MADURAI,1998,Kharif        ,Pulses total,26613,14426)
(Tamil Nadu,MADURAI,1998,Kharif        ,Ragi,229,408)
(Tamil Nadu,MADURAI,1998,Kharif        ,Rice,88338,303471)
```

```
(Pulses total)
(Rice)
(Sesamum)
(Soyabean)
(Sugarcane)
(Total foodgrain)
(Urad)
(Barley)
(Gram)
(Lentil)
(Masoor)
(other oilseeds)
(Peas & beans (Pulses))
(Potato)
(Pulses total)
(Rapeseed &Mustard)
(Total foodgrain)
(Wheat)
(Maize)
(Moong(Green Gram))
(Pulses total)
(Rice)
(Total foodgrain)
(Urad)
(Pulses total)
(Rice)
(Sunflower)
(Total foodgrain)
(Turmeric)
(Arhar/Tur)
(Horse-gram)
(Maize)
(Other Cereals & Millets)
(Other Kharif pulses)
(Potato)
(Ragi)
(Rice)
(Sesamum)
```

Hive commands:

create database Rahul;

use Rahul;

create table agriculture(State_Name String,District_Name String,Crop_Year int,Season String,Crop String,Area float,Production int) row format delimited fields terminated by ',';

load data local inpath '/home/hadoop/Documents/crop.csv' into table agriculture;

select*from agriculture;

create view crop_year as select Crop from agriculture where Crop_Year>2002;

select*from crop_year;

select count(*) from agriculture;

select State_Name,lower(District_Name) from agriculture;

select Crop,sqrt(Production) from agriculture;

select max(Production) from agriculture;

Output:


```
hive> create table agriculture(State_Name String,District_Name String,Crop_Year int,Season String,Crop String,Area float,Production int) row f
ormat delimited fields terminated by ',';
OK
Time taken: 0.595 seconds
hive> load data inpath '/home/hadoop/Documents/crop.csv' into table agriculture;
FAILED: SemanticException Line 1:17 Invalid path ''/home/hadoop/Documents/crop.csv'': No files matching path hdfs://127.0.0.1:9000/home/hadoop
/Documents/crop.csv
hive> load data local inpath '/home/hadoop/Documents/crop.csv' into table agriculture;
Loading data to table rahul.agriculture
OK
Time taken: 1.03 seconds
hive> select*from agriculture;
OK
State_Name      District_Name   NULL    Season  Crop    NULL    NULL
Tamil Nadu      MADURAI 2002    Whole Year      Pump Kin        15.0    0
Tamil Nadu      MADURAI 2002    Whole Year      Ribed Guard     4.0     0
Tamil Nadu      MADURAI 2002    Whole Year      Snak Guard      6.0     0
Tamil Nadu      MADURAI 2002    Whole Year      Sugarcane       5639.0  601070
Tamil Nadu      MADURAI 2002    Whole Year      Sweet potato    63.0    861
Tamil Nadu      MADURAI 2002    Whole Year      Tapioca 98.0    3096
Tamil Nadu      MADURAI 2002    Whole Year      Tomato  376.0   4632
Tamil Nadu      MADURAI 2002    Whole Year      Turmeric        29.0    108
Tamil Nadu      MADURAI 2002    Whole Year      Water Melon     2.0     0
Tamil Nadu      MADURAI 2002    Whole Year      Yam     17.0    0
Tamil Nadu      MADURAI 2003    Kharif  Arhar/Tur       1420.0  869
Tamil Nadu      MADURAI 2003    Kharif  Bajra   3338.0  3757
Tamil Nadu      MADURAI 2003    Kharif  Castor seed     36.0    12
Tamil Nadu      MADURAI 2003    Kharif  Cotton(lint)    10651.0 6570
Tamil Nadu      MADURAI 2003    Kharif  Groundnut       5742.0  8446
Tamil Nadu      MADURAI 2003    Kharif  Horse-gram      168.0   45
Tamil Nadu      MADURAI 2003    Kharif  Jowar   11319.0 13154
Tamil Nadu      MADURAI 2003    Kharif  Korra   7.0     2
Tamil Nadu      MADURAI 2003    Kharif  Maize   794.0   1591
Tamil Nadu      MADURAI 2003    Kharif  Moong(Green Gram)       4644.0  391
Tamil Nadu      MADURAI 2003    Kharif  Other Kharif pulses     2436.0  423
Tamil Nadu      MADURAI 2003    Kharif  Ragi    178.0   274
```



```
hive> drop view crop_year;
OK
Time taken: 0.454 seconds
hive> create view crop_year as select Crop from agriculture where Crop_Year>2002;
OK
Time taken: 0.156 seconds
hive> select*from crop_year;
OK
Arhar/Tur
Bajra
Castor seed
Cotton(lint)
Groundnut
Horse-gram
Jowar
Korra
Maize
Moong(Green Gram)
Other Kharif pulses
Ragi
Rice
Sesamum
Small millets
Sunflower
Urad
Gram
Ash Gourd
Banana
Bhindi
Bitter Gourd
Bottle Gourd
Brinjal
Cabbage
Cashewnut
Time taken: 0.136 seconds, Fetched: 26 row(s)
hive>
```

```
Gram
Ash Gourd
Banana
Bhindi
Bitter Gourd
Bottle Gourd
Brinjal
Cabbage
Cashewnut
Time taken: 0.136 seconds, Fetched: 26 row(s)
hive> select count(*) from agriculture;
Query ID = hadoop_20221105114105_cb92a262-898a-44e0-a4a8-4be9335473af
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1667627862975_0001, Tracking URL = http://ab1-cse107:8088/proxy/application_1667627862975_0001/
Kill Command = /home/hadoop/hadoop-3.2.3/bin/mapred job  -kill job_1667627862975_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-11-05 11:41:18,121 Stage-1 map = 0%,  reduce = 0%
2022-11-05 11:41:24,286 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.46 sec
2022-11-05 11:41:29,416 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.4 sec
MapReduce Total cumulative CPU time: 3 seconds 400 msec
Ended Job = job_1667627862975_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.4 sec   HDFS Read: 15367 HDFS Write: 102 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 400 msec
OK
37
Time taken: 25.629 seconds, Fetched: 1 row(s)
hive>
```

```
hive> select Crop,sqrt(Production) from agriculture;
OK
Crop    NULL
Pump Kin        0.0
Ribed Guard     0.0
Snak Guard      0.0
Sugarcane       775.2870436167498
Sweet potato    29.34280150224242
Tapioca 55.641710972974224
Tomato  68.05879810869422
Turmeric        10.392304845413264
Water Melon     0.0
Yam     0.0
Arhar/Tur       29.478805945967352
Bajra   61.29437168288782
Castor seed     3.4641016151377544
Cotton(lint)    81.05553651663777
Groundnut       91.90212184710427
Horse-gram      6.708203932499369
Jowar   114.69088891450794
Korra   1.4142135623730951
Maize   39.88734135035826
Moong(Green Gram)       19.77371993328519
Other Kharif pulses     20.566963801203133
Ragi    16.55294535724685
Rice    254.99803920814765
Sesamum 25.199206336708304
Small millets   16.0
Sunflower       17.4928556845359
Urad    23.853720883753127
Gram    6.928203230275509
Ash Gourd       0.0
Banana  293.5523803344132
Bhindi  39.344631145812
Bitter Gourd    0.0
Bottle Gourd    0.0
```

```
Banana  293.5523803344132
Bhindi  39.344631145812
Bitter Gourd    0.0
Bottle Gourd    0.0
Brinjal 40.149719799769464
Cabbage 8.0
Cashewnut       9.38083151964686
Time taken: 0.132 seconds, Fetched: 37 row(s)
hive> select Crop,max(Production) from agriculture;
FAILED: SemanticException [Error 10025]: Line 1:7 Expression not in GROUP BY key 'Crop'
hive> select max(Production) from agriculture;
Query ID = hadoop_20221105114751_09d33945-9731-4f22-9f91-02b9728b4fdb
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1667627862975_0002, Tracking URL = http://ab1-cse107:8088/proxy/application_1667627862975_0002/
Kill Command = /home/hadoop/hadoop-3.2.3/bin/mapred job  -kill job_1667627862975_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-11-05 11:47:57,919 Stage-1 map = 0%,  reduce = 0%
2022-11-05 11:48:03,043 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.52 sec
2022-11-05 11:48:08,176 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.41 sec
MapReduce Total cumulative CPU time: 3 seconds 410 msec
Ended Job = job_1667627862975_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.41 sec   HDFS Read: 15376 HDFS Write: 106 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 410 msec
OK
601070
Time taken: 18.099 seconds, Fetched: 1 row(s)
hive>
```