Data Recognition

The current system in the Planing and agricultural economic department

The Department of Agricultural Planning and Economics consists of five administration. In each administration, a partial system is used to deal with data, as most of it handles data in Excel format or separated values. Some also use regular text files. In addition to the presence of a database using MySQL, but this distribution leads to an impact on the efficiency of analyzes in management as a whole.

1. Text files: The department receives an amount of information from several ministries and other departments in the form of text files, whether in Excel or CSV format. The department is located in the city of Khartoum, and it receives these data from departments and ministries located in the same city, with branches in different cities of Sudan. These ministries receive their information from their branches on weekly and monthly grounds while it provides the Department of Agricultural Planning and Economy on a monthly bases.

These data are collected monthly and prepared to generate reports for the administration. The agricultural statistics administration of the department supervises this. We discussed the data sources in the first chapter in our discussion of background according to the figure ??.  
  
diagram showing PAE reporting flow

1. Database: The Department also has a database that uses MySQL, which collects this information from several sources in cooperation with the relevant departments. These data are then provided to the Agricultural Statistics ِAdministration and other offices that need reports. The task of managing and feeding this database depends on the Information Technology Administration of the Department.  
     
   Reporting Flow in Information Technology Office
2. Weather Data: The administration also receives weather data from the national meteorology Authority. Here we preferred to receive data from an international weather website using the API for the accuracy and speed of the data it provides if compared to the infrastructure owned by the Sudan Meteorological Authority.  
   Reporting Flow for Wether API

Data structure

1. 1. CSV data structure   
   We chose four of the primary crops that can be considered as cash crops for Sudan, namely: wheat, sorghum, millet, and cotton because Wheat, Millet, and Sorghum are introduced as a fundamental component of food in Sudan, mainly wheat in the north and cities, while millet is more concentrated in the west, and it and sorghum share consumption in all Sudan to varying degrees. All values ​​are for the years from 1960 to 2019. Some properties may differ depending on the type of crop. The basic features of these crops are as follows:
   1. Area Harvested: Indicates the area from which the crop is collected. It is the harvested area only, as the uncultivated or un harvested area is excluded and execlud also the unharvested area without cause or due to damage. [http://www.fao.org/waicent/faostat/agricult/pr\_ele-e.htm]
   2. Beginning Stocks: Existing supplies of a crops commodity that consist of remaining stock carried over from the previous year's production. [https://www.ers.usda.gov/data-products/food-availability-per-capita-data-system/glossary/]  
      the ending stocks carried into the new marketing year from the previous year
   3. Domestic Consumption: the quantity that is harvested or imported and used in the country[https://dictionary.cambridge.org/us/dictionary/english/domestic-consumption]

all possible uses of the commodity: food, feed, seed, waste, and industrial processing

* 1. Ending Stocks: The remainder of current crop production carried over into the next crop year. [https://www.ers.usda.gov/data-products/food-availability-per-capita-data-system/glossary/]  
     The unused commodity was remaining at the end of the marketing year for use in the next year.
  2. Exports: The quantity transferred from the crop to another country or region for trade purposes
  3. Imports: The quantity brought from another country for use within Sudan
  4. Production: It is the actual harvested output of the crop, with the exception of the harvest and threshold losses, and the portion that is not harvested for any reason. Therefore, production includes both the quantity sold in the market and consumed by the producor. If the production data refer to two calendar years and it is not possible to divide the production between them, then we attribute it to the year in which the largest amount of production occurred. We use the metric ton to calculate the output (MT).[ http://www.fao.org/waicent/faostat/agricult/pr\_ele-e.htm]
  5. Total Supply: mainly it calculated by the following formula  
     beginning stocks + domestic production + imports
  6. Yield: The data represents the harvested output per unit. This value is usually calculated by dividing the production by the harvested area. Data are recorded in a hectare (100 kilograms per hectare) (HG/HA).[ http://www.fao.org/waicent/faostat/agricult/pr\_ele-e.htm]

For each of the above properties, we will study the following features

* Market Year: It is the year in which the production process took place
* • Quantity: it varies according to the property. If we were studying, for example, the export, then it would be the exported quantity
* • Unit of Measure: The unit of measurement varies according to the crop and the property; for example, with cotton will be Bales, which is equal 480 lb. With the harvested area, it will be one thousand hectares, which equals 10 square kilometers, and it is shown in the table ??:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| property | Measure unit |  | equivalent in the metric system |  |
|  | Cotton | Cereals | Cotton | cereals |
| Area Harvested | HA | HA | 0.01 km2 | 0.01 km2 |
| Beginning Stocks | 480 lb. Bales | MT | 217.724 KG | 1000 KG |
| Domestic Consumption | 480 lb. Bales | MT | 217.724 KG | 1000 KG |
| Ending Stocks | 480 lb. Bales | MT | 217.724 KG | 1000 KG |
| Exports | 480 lb. Bales | MT | 217.724 KG | 1000 KG |
| Imports | 480 lb. Bales | MT | 217.724 KG | 1000 KG |
| Production | 480 lb. Bales | MT | 217.724 KG | 1000 KG |
| Total Supply | 480 lb. Bales | MT | 217.724 KG | 1000 KG |
| Yield | KG/HA | MT/HA | KG/KM2 | KG/KM2 |

The measuring unit and equivalent in metric system for each property

* Growth Rate: Growth rates refer to the percentage change of production within a specific time period. And we calculated it by the formula

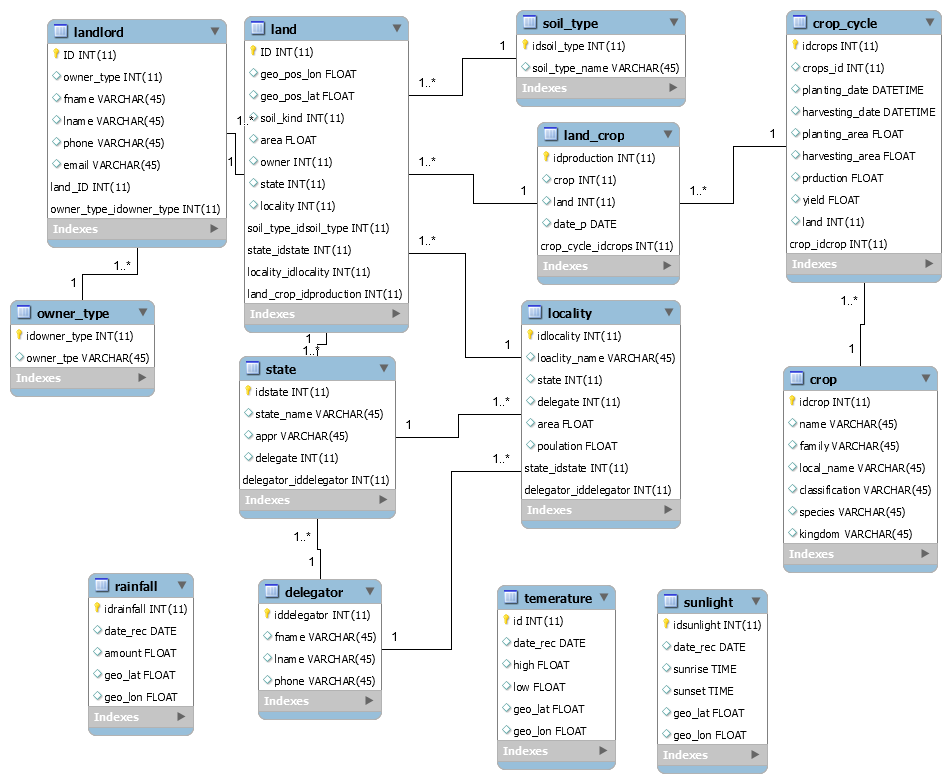


Where:

PR = Percent Rate  
VPresent = Present or Future Value  
VPast = Past or Present Value

The relational database data structure

The domestic production database is a small database designed using MySQL to collect some essential data. According to the Administration of Agricultural Statistics, they are working on a more comprehensive database, initiated in cooperation with the Information Technology Administration to implement. It is well known that the MySQL database is a widely used database worldwide and supported by almost all hosting providers, was open source, and free before they became owners of Oracle.

In this research, the following tables will be used from the database that we display the ERD for it in figure ??

Entity relation diagram for PAE\_DB database

landlord table: In this table, information is provided about landlords and their contact information.

owner\_type table: This table stores information about the type of landowner who is divided into individuals, companies, and projects owned by the state. In order to help explain the contribution of each sector to domestic production.

Land table: Contains important information related to the cultivated land, such as its geographical location, area, soil type, landlord, and administrative belonging.

Table of soil\_type: data stored on this table is the type of soils, according to relation to the diversity of soil types in Sudan, which helps in providing advice to owners on the most suitable crop type for cultivation. It also helps in studying the productivity of each type of land for the specific crop.

crop\_cycle table: This table stores information related to the agricultural cycle, such as the dates of planting and harvesting, and the cultivated and harvested areas for the farming season. You will use here the same measures mentioned in the table ??? which is a hectare and metric ton

crop table: Here information is stored about the agricultural crop, such as local and scientific name, classification, family, and kingdom. This helps in subsequent scientific research on a specific family or specific crop.

State table: Important data for Sudan's 17 states: two states in the north, three in Kordofan, five in Darfur, three in the east, and five in the center, including Khartoum state, which contains the country's capital.

Locality table: Sudan consists of about 189 localities that differ in their areas and their number in each state, as it reaches 21 in some states, while they are only 3 in other states. They also differ in the prevailing climate, from desert to subtropical. This table stores information about the state, its area, and its population, in addition to the delegator.

The Delegator table: The person responsible for communicating agricultural statistics, in the specific locality or state, to the Department of Planning and Agricultural Economics.

Land\_crop table: This table is used as a link table to provide us with a many-to-many relationship between the lands and crops tables because it is known that every piece of land owned by an individual or group may produce more than one crop and that a single crop can be grown on more than one land. We also define the irrigation method used, which is divided into two sectors: rain and irrigated.

The following three tables are for recording weather conditions, temperature, amount of rain and sunlight, and the geographical location is also recorded, even if it lacks wind intensity and humidity. It is used to set forecasts for the amount of rain and also helps to predict productivity in the event of favorable weather.

Temperature table: In this table, the information related to the temperature is stored as the lowest and highest degree with the determination of the location

Rainfall table: We record data related to the amount of rain, date, and location

Sunlight Table: It stores sunlight data such as sunrise and sunset.

1. Weather API data structure

To get weather data, we use the API method to get our data from OpenWeatherMap. [OpenWeatherMap](http://www.openweathermap.com/) gives current weather information and forecasts. It is very simple to use. The [OpenWeatherMap weather API](http://www.openweathermap.com/api)uses JSON to exchange data.

The weather API returns current weather, forecast, historical weather information, and weather station data. Moreover, it also has UV index data.

OpenWeatherMap provides different two modes to look for a city. One uses name pattern and another using geo-coordinates.

All these APIs are free to use, but we have to create a dev key so that you can make requests.

when we call using the following API call

<https://samples.openweathermap.org/data/2.5/weather?id=379252&appid=b6907d289e10d714a6e88b30761fae22>

we get the following result:

{

"coord":

{

"lon":145.77,

"lat":-16.92

},

"weather":

[{

"id":802,

"main":"Clouds",

"description":"scattered clouds",

"icon":"03n"

}],

"base":"stations",

"main":

{

"temp":300.15,

"pressure":1007,

"humidity":74,

"temp\_min":300.15,

"temp\_max":300.15

},

"visibility":10000,

"wind":

{

"speed":3.6,

"deg":160

},

"clouds":

{

"all":40

},

"dt":1485790200,

"sys":

{

"type":1,

"id":8166,

"message":0.2064,

"country":"AU",

"sunrise":1485720272,

"sunset":1485766550

},

"id":2172797,

"name":"Cairns",

"cod":200

}

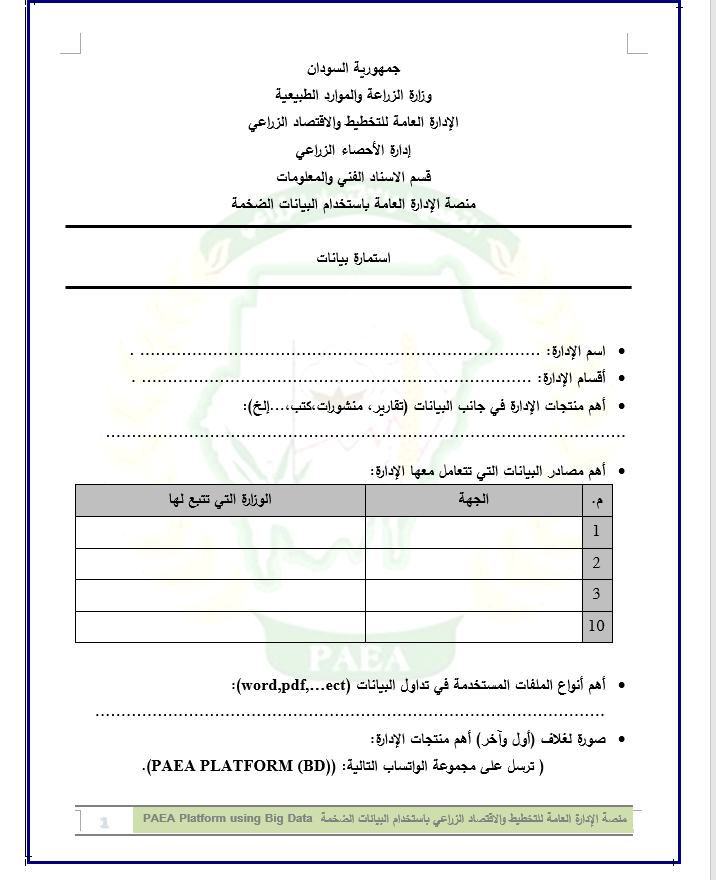
It responds in JSON file providing the following attributes:

* coord
  + coord.lon City geolocation, longitude
  + coord.lat City geolocation, latitude
* weather (more info Weather condition codes)
  + weather.id Weather condition id. It divided into seven groups
  + weather.main it is the group that content weather parameters (for example Rain, Snow, Extreme, etc.)
  + weather.description Weather conditions. You may prefer to get the description in your language.
  + weather.icon the id for weather icon
* base base present the Internal parameter
* main
  + main.temp the temperature. We can use any unit, but the default one is Kelvin, Metric or Celsius, Imperial, and Fahrenheit.
  + main.feels\_like for temperature also. It represents the human perception about the weather.
  + main.pressure this parameter gives us the atmospheric pressure (measure on the sea level), it uses hPa
  + main.humidity: Humidity, %
  + main.temp\_min the amount of minimum temperature.
  + main.temp\_max the amount of maximum temperature.
  + main.sea\_level it measures the atmospheric pressure that toke on the sea level, it measures using hPa
  + main.grnd\_level: it measures the atmospheric pressure that toke on the ground level; it measures using hPa.
* wind
  + wind.speed parameter for wind speed. The default unit for the metric is meter/sec, and for imperial are miles/hour.
  + wind.deg it gives the wind direction.
* clouds
  + clouds.all Cloudiness, %
* rain
  + rain.1h it retrieves the rain amount for the last 1 hour, mm
  + rain.3h it retrieves the rain amount for the last 3 hour, with millimeter mm
* snow
  + snow.1h it retrieves the snow amount for the last 1 hour, mm
  + snow.3h it retrieves the snow amount for the last 3 hour, with millimeter mm
* dt Time of data calculation, UNIX, UTC
* sys
  + sys.type Internal parameter
  + sys.id Internal parameter
  + sys.message Internal parameter
  + sys.country Country code (GB, JP, etc.)
  + sys.sunrise Sunrise time, Unix, UTC
  + sys.sunset: Sunset time, Unix, UTC
* timezone Shift in seconds from UTC
* id City ID
* name City name

cod Internal parameter

Data Sources:

The primary data source in this research is the data we obtain from the Department of Planning and Agricultural Economics. It was very difficult to obtain for a variety of reasons, most notably the bureaucratic complications and the lack of good infrastructure to store and retrieve data. Where we initially made a questionnaire distributed to the various administration in the department to find out the type of data available in each administration it appears in figure ??



This helps us to know about the type of data available in each administration; most of them were in doc format in addition to Xls with some data in SPSS format. We benefited from it in the first type of data after cleaning the data and converting it to CSV format.

Also, available data in the database, although incomplete in the absence of data for some years.

We had to resort to global databases, so we got our data from the FAO, which offers the FAOSTAT database, which gives you free access to food and agricultural data from 1961. And the good thing about the FAO site is that they also provide the data in CSV format.

As for the weather data, we tried to communicate with the Sudanese Meteorological Authority, but we did not obtain data from them, so we turned to the Internet and obtained access to the openweathermap.com website, as mentioned above. It provides free access to the current weather forecast with some restrictions. With the provision of access to historical data if you are a subscriber.