

Group 11 - BI project document

Project Goals

This project aims to create an overview of the agriculture and production of critical food crops across the United States of America. The end users envisaged for this project are federal government agencies mandated with ensuring food and crop security and crop-based manufacturers who would like to locate their industries close to supply locations.

Dashboards allow users to track individual crops, view cultivation over states as well as track acres under cultivation. We are confident these metrics will allow users to track cultivation and production and view any sharp reduction in yields or crop failure. Such business intelligence will allow federal agencies to see current trends, look at possible forecasts and compare these forecasts with actual values. Doing so will enable the agencies to see any concerning or uplifting patterns and position their policies accordingly like creating contingencies in the event of possible yield reduction or failure or boosting production by providing support and incentive to grow a specific type of crop during a period.

A good example of the above point could be - How the federal government in the past has seen a growing demand for corn and provided incentives to farmers to grow more of it making the U.S. currently the largest exporter. Similarly, the federal government has not provided incentives to oats for the last several years now. Oats which were one of the staple food sources and heavily produced crop in the U.S. has seen a steady decline.

Coming to businesses who would want to locate close to supply, our dashboards would allow them to identify states to locate themselves. They can do so by identifying states which have been consistently performing well in terms of a particular crop over the recent periods and by viewing the forecast for the crop on a national level aggregate in the coming years.

A good example of the above point could be - The ethanol industry has successfully identified high-growing states across the U.S. and positioned itself nearby supply, which enabled them to grow at a faster rate over the last few years. Similarly, new industries similar to exiting industries like Quaker Oats or some company that could create a new product altogether from a crop like Sorghum Syrup could position themselves better. Not only new industries, but current growing industries like ethanol can position themselves better with these visualizations

Team Roles

Our team worked through the project by sharing the responsibilities in the following areas

- Workflow creation - Dilip Pinnamreddy, Akram
 - For creating the workflow, Dilip and Akram took up the primary responsibilities. Dilip came up with the vision of the project and data source. The team initially discussed all the data sets to make sure everyone was aware of the data structure, potential pitfalls, and relevant columns. Post that the data was cleaned via a tableau prep flow.

- Dashboard creation - Dilip Pinnamreddy, Srikanth Tamima
 - For creating the dashboards, initially, the entire team brainstormed ideas together on how to present the current data, which charts to show, and what possible insights could be drawn from the dashboards. Once the ideas had been discussed in the initial stage Dilip and Srikanth worked in tandem to create the dashboards. Once the initial dashboards were created the entire team again discussed what changes could be made and how the dashboards could be improved. Then the final edits were again incorporated.
- Report and Presentation creation - Ananthu, Saksham
 - The reports were primarily by Ananthu and Saksham. The report and presentation creation was the final process after the dashboards had been created. Both the documents focus on gathering the relevant information from all the earlier stages and presenting it coherently. The report was reviewed by the entire team before submission.

Data Sources and Preparation

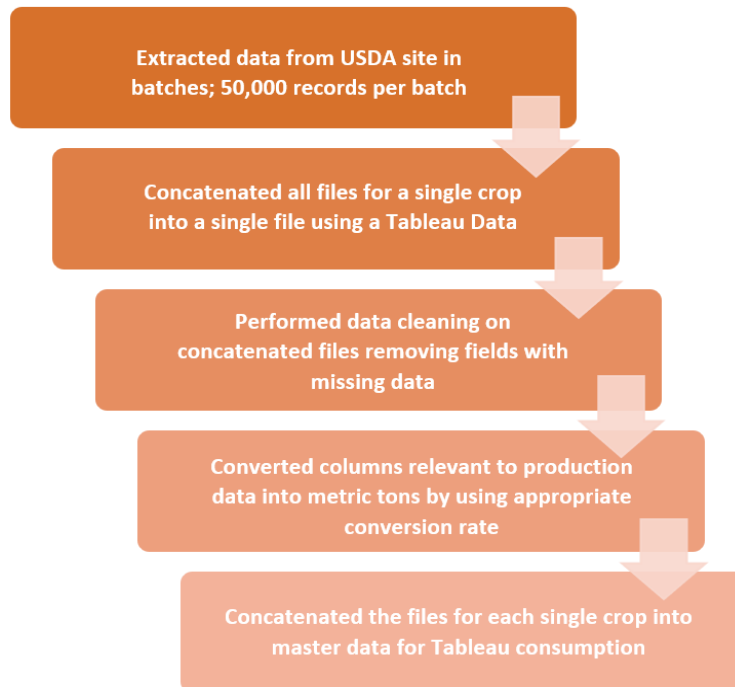
Our data has been collected from the **United States Department of Agriculture - National Agricultural Statistics Service**. We have focused particularly on Field Crops grown in the United States. We primarily focused on the major U.S. Feed Grains like Corn, Barley, Oats, Rice, Sorghum, Soybeans, and Wheat.

The USDA site hosts both CENSUS and SURVEY data for these crops with details like Production in BU/CWT, Acres Planted, Acres Harvested, and Yield in BU/LBS per Acre for field crops over the years. We extracted this data for our analysis. We specifically took these categories to get a holistic flow process from the germination period to the production period of the crops so that the end-users would be able to make decisions about policies in case of the federal government or to identify potential in states in the U.S. for crop-based businesses.

Conversions used are:

- 1 Bushel (BU) of Barley = 0.021772 Metric Tons
- 1 BU of Corn = 0.0254 Metric Tons
- 1 BU of Oats = 0.0145 Metric Tons
- 1 Hundredweight (CWT) of Rice = 0.0508023 Metric Tons
- 1 BU of Sorghum = 0.0254 Metric Tons
- 1 BU of Soybeans = 0.0272155 Metric Tons
- 1 BU of Wheat = 0.0272155 Metric Tons

In the below flowchart we have highlighted a high-level design of the flow utilized for Crop Production data from the USDA in **Tableau Prep**.



In the initial clean step for each crop, after merging all files for a particular crop:

1. We changed the data type of production values into a numerical value from a text field. The data could have initially been by Tableau as a text field because of the missing values which were denoted by '(D)'.
2. Then, we filtered out null values to keep rows only with production values. Tableau has a limitation of not being able to fill in missing values intuitively based on moving averages and nearest neighbors.
3. Finally, we created a calculated field, which is the conversion from individual values to a single unit - Metric Tons.

In the second and final clean step, after merging the file generated in the previous step for each crop:

1. We dropped unnecessary table-based columns generated due to the union of files.
2. We dropped columns like **Zip Code**, **Region**, **watershed_code**, **Watershed**, **Domain Category**, and **CV (%)**.
3. Also, we changed the data-role of State to State/Province as recommended by the prep software.

For the other areas of focus - Acres planted, harvested, and yield, due to a state-level aggregate of data extraction, we were able to obtain the data in a single file. Hence, we were able to skip the first union step. However, after the second union step, i.e., merging all the crops, we performed all the clean steps mentioned above in a single clean step.

There was one problem with the data collected on state-level aggregates. In some instances, the USDA data classifies the less producing states together as 'OTHER STATES'. We ignored this detail in our dashboards using geographical/map-based features.

Dashboard Interactions and Findings

The final file to refer to is **Final Grain Analysis Tableau.twbx**

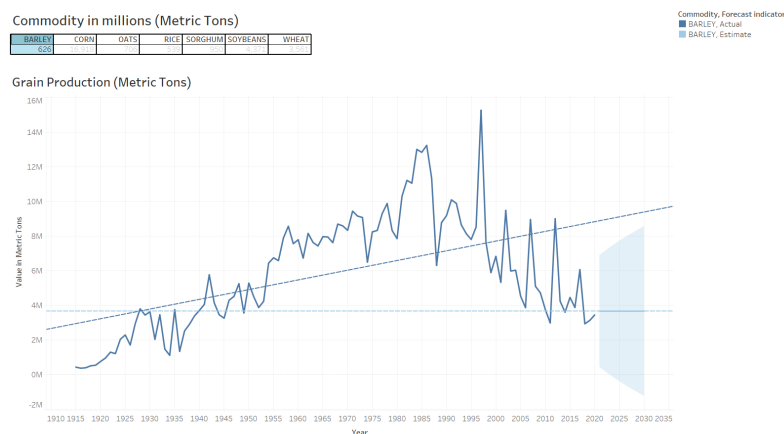
Please clear all the filters before starting the interactions with the dashboard. The save file state is on Dashboard 1 with no filters, however, we would suggest users navigate to this dashboard if required. Our dashboards are a combination of seven crops and a view of all of them together is not an ideal way to view them. We recommend filtering the data to a specific crop, year, and state wherever applicable for better visualizations and insights.

Below, we have highlighted interactions with each of our dashboards with explanation, and an example for easier navigation. In case of any issue during navigation, we would request users to clear filters.

We created 4 dashboards with interactions

1. Field Crop Production in the U.S. with a Tableau forecast for the next 10 years.

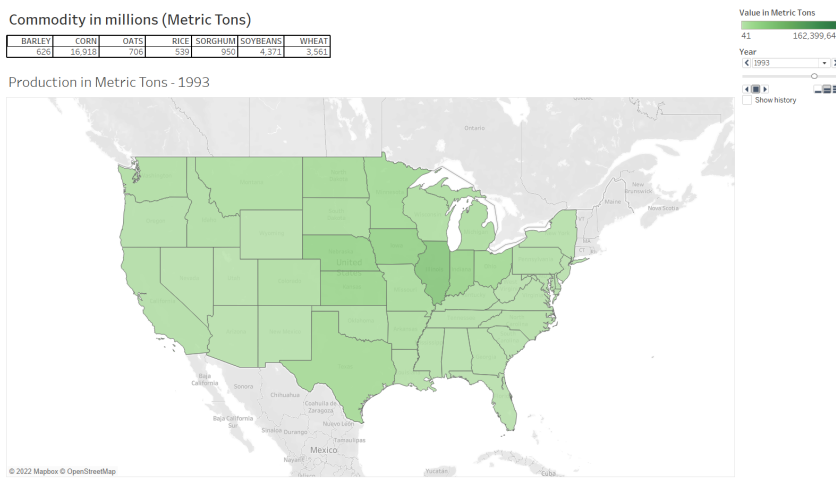
To interact with the dashboard you would have to click on the specific crop at the top with the crop totals to filter down the results below. Please find an example below



This dashboard enabled us to see the overall trends for each crop. Overall, Corn, Rice, Soybeans, and Wheat see an upward trend in the future. Crops like Barley and Sorghum have seen a partial decline/stationary growth in the recent years, however, their past production has led to an overall upward trend. Oats have seen a steep decline for quite some years now.

2. Spread of field crops across the U.S. terrains. This helps us visualize the spread of field crops in the U.S. over years.

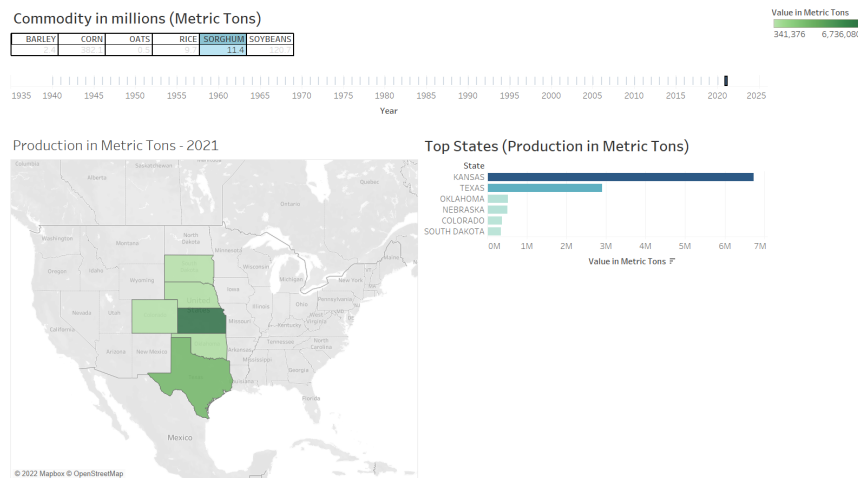
To interact with the dashboard, first, select a crop at the top and towards the right side legend, you can select the year to start, and clicking on the play would show the progression over the years. Please find an image example below. A short video has been displayed in our presentation.



This dashboard enables us to see the states that have consistently over the years produced the crop, and states that do it periodically. The depth of the color in the map indicates the amount of production.

3. In this dashboard, we included a year slider as well along with the crop totals. Selecting a crop and a specific year would show the spread of the field crop produced, along with the descending order of the production across the states

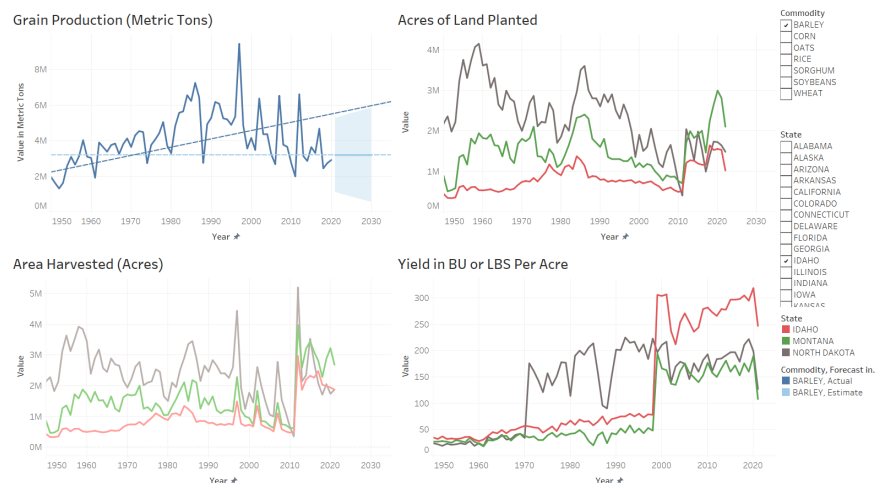
To interact with this dashboard, click on the crop, and then select the year to see the details. Please find an example below.



Using this dashboard we were able to find leading producers of a specific field crop across the U.S in recent years.

- a. Barley - Idaho, Montana, North Dakota
 - b. Rice - Arkansas, California, Louisiana
 - c. Oats - Minnesota, Iowa, North Dakota
 - d. Corn - Iowa, Illinois, Nebraska
 - e. Sorghum - Kansas, Texas, Oklahoma
 - f. Soybeans - Illinois, Iowa, Minnesota
 - g. Wheat - Kansas, North Dakota, Washington
4. In this dashboard, we see the acres planted, acres harvested and the yield per acre for the top states for each crop.

To interact with this dashboard, select a specific crop from the filter on the right-hand side, and then select the top three states for the crop as detailed above. Please find an example below



Note: We would only be detailing one crop in this section to stick to the limit of 5 pages in the document.

Trends for each crop

a. Barley

- i. Barley is overall seeing an upward trend in terms of the production, however, this seems to be offset by large productions in the late nineties followed by a decline.
- ii. Coming to the top 3 leading states in production, Acres planted have a slight decline in North Dakota, a recent increase in Montana and Idaho.
- iii. Area Harvested has been mostly constant but has seen an increase around 2010 followed again by a constant trend.

North Dakota can produce more from less harvested land. This occurrence should be explored further and the best practices should be shared with others for better crop yields. The federal government should incentivize other field crop producers to implement these best practices.

Also, due to a stable or a growing crop production in this state, a business with their core ingredient being Barley should locate in these states. However, the crop decline over the years overall in the U.S. is a concern. If there is a possibility of further decline in the crop, businesses should locate themselves close to the port cities pulling in the imports of the crop, rather than close to the current supply locations.

Changes from the Original Design Document

Quite a few changes were incorporated from the initial design document we prepared based on the feedback provided.

1. Detailed End Users/Target Audience: We detailed our end users with examples to have a firmer understanding of the target audience in the Introduction section of this document.
2. Expanded Data Sources: Initially, we had only opted for the “Production” category under each of the seven commodities we had chosen. However, we felt the need to incorporate more data points like Acres harvested, planted, and yield for a better understanding of the plantation. This helped us provide more insights to the end-users.
3. Deliberated on Goodness of Measure: To address the genericness in the design document, we expanded on how each principle would apply to our project.

Future Improvements

We can deep-dive into data even further. Currently, we worked on a state-level aggregate but depending on the data availability, we can go into a U.S. county-level analysis.

Doing so will help federal agencies better create their policies for the different geographical needs of the field-crop producers and also help crop-based businesses to pinpoint a location for production to a more granular level.

A popup dashboard could be created which can host all the details in a single layer but during the dashboard walkthrough, we could zoom into individual elements and proceed one by one to create a story.