Proposed Database Design for Sale of Vegetable Starts at Pattee Creek Farm

Danielle Wyenberg
GEOG 574: Geospatial Database Design and Development
University of Wisconsin-Madison

Background

Farmers Markets are becoming increasingly popular across the country, and there is a growing interest in eating sustainably and locally among consumers. Many people are looking to boost food security and develop at-home hobbies, causing a shortage of seeds and seedlings for the past two growing seasons. Due to this perceived demand and past experience with gardening, a local, family-owned business supplying vegetable seedlings to the surrounding community is explored and presented in the following database design.

Seedlings will be grown in a greenhouse on Pattee Creek Farm in South Dakota and will be primarily sold at farmers markets in the surrounding area. Orders may be taken from customers at any time, but the pickup location will be a local farmers market. This concept needs some planning, however, and a deeper look into the components of this type of business and how they relate to each other.

Objective

A spatial database will be designed in order to store and access plant, customer, and market information. Several research questions will be identified and explored, including the following: 1) How many farmers markets can be feasibly participated in as a vendor, based on location? 2) How many customers are located within each market radius? And 3) Which varieties of plants sell best during specified time periods?

DATA COLLECTION

In order to create a table of local farmers markets and their locations, data was gathered from the Iowa State University Extension Office's *Local Food and Flavors Guide* along with personal research on local farmers market options. A table was created to organize the data, consisting of Name and location information as well as day and time, and any fee associated with registering for each respective market.

Since the vegetable start business is still in the concept phase, data was gathered on plants and pricing, and a mock dataset was created for customers and orders. TIGER line data was utilized for county shapefiles, filtered to extract only those counties of interest near Pattee Creek Farm (*Figure 1*).

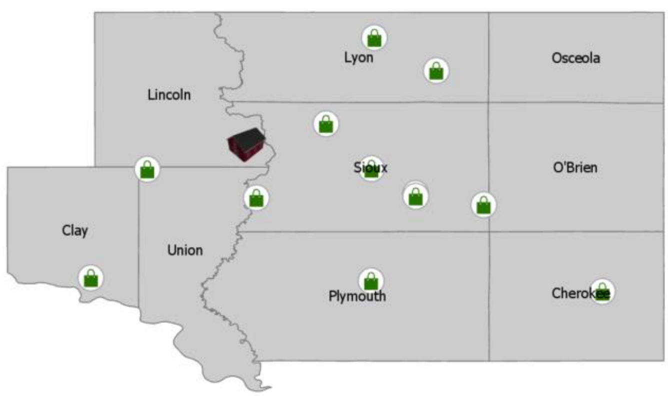

Figure 1: Area of Interest

DATABASE DESIGN

In order to determine which information to include in the database, an E-R (Entity-Relationship) model was created (*Figure 2*). An E-R model helps to organize the features and information that will be utilized in the database. Each entity, or data object, is represented with a rectangle in the E-R model. Plants, orders, customers, markets, and order details are each defined as entities in this database. Each entity has attributes attached to it, and entities are linked to one another using relationships. The E-R diagram also helps to organize data types and primary keys in order to identify unique information.

A logical schema was generated from the E-R diagram (*Figure 3*) to further describe the relational database for the Pattee Creek Farm vegetable starts company concept. From here, it was determined that a spatial database would be used for the implementation, specifically, an ArcGIS geodatabase. Feature classes and their field names were organized and defined. Customers and Farmers Market locations can be geocoded by address using a locator, and other information (ex: orders, plants) will be stored in tables that can be accessed and processed as needed.

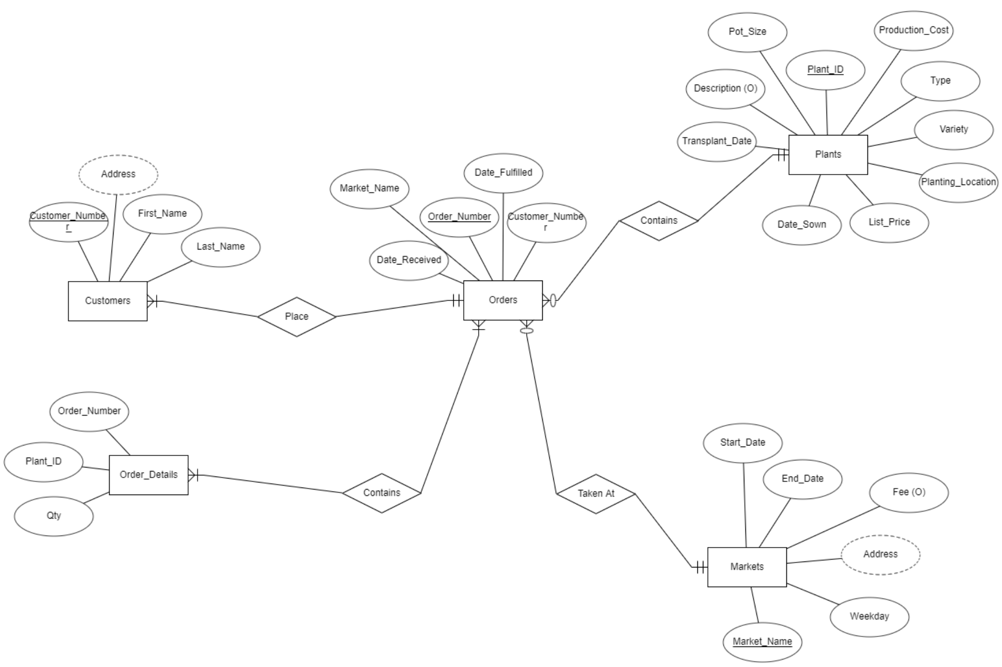

Figure 2: E-R Model

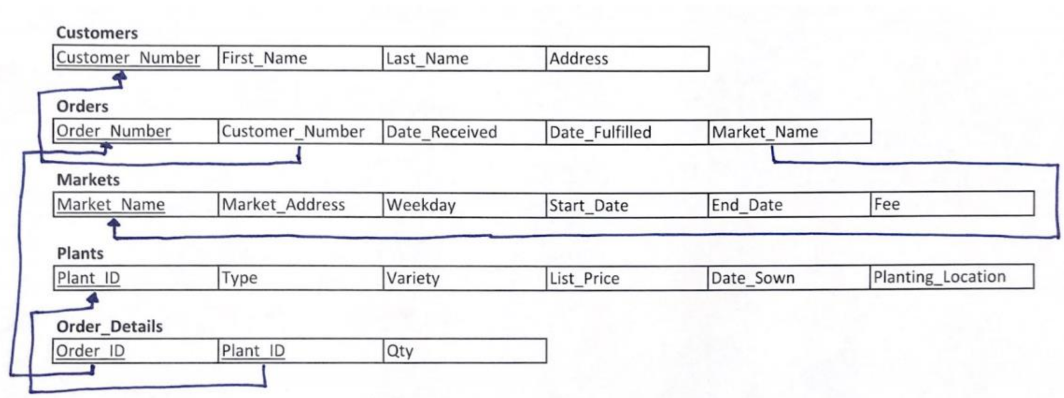

Figure 3: Logical Schema

Database Implementation

Based on the database design, tables were built according the logical schema seen above (*Figure 3*), containing fields for each attribute. The shapefiles and attribute tables were imported into the geodatabase (*Figure 5*), taking into account the primary keys, foreign keys, and other relationship constraints. Appropriate data types were identified and utilized. Datasets were imported into the database. The farmers market and customer tables were both geocoded by address to create spatial data that can be visualized in ArcGIS Pro (*Figure 4*). A domain was created for pot size (*Figure 6)*, as these are coded values.

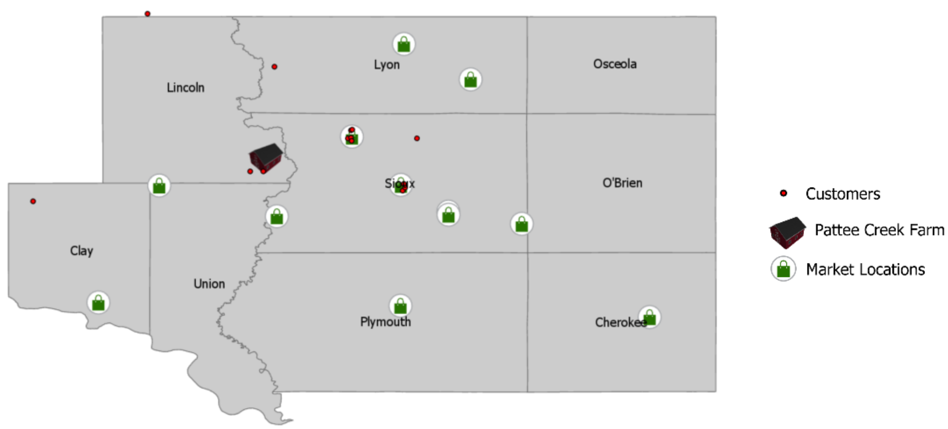

Figure 4: Geocoded Locations for Markets and Customers

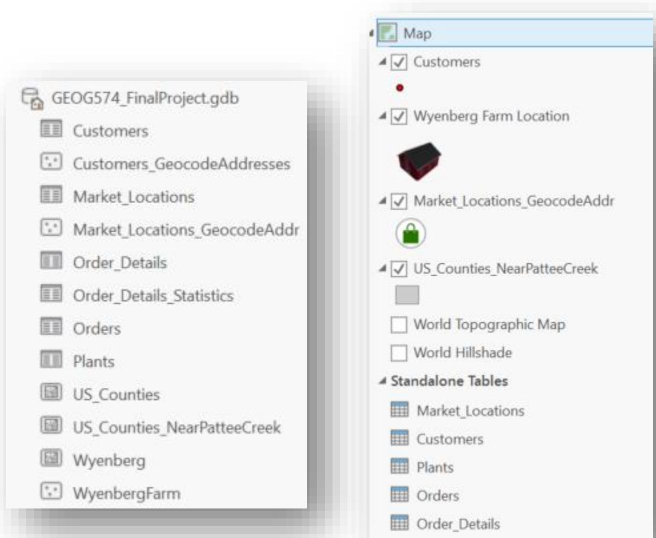

Figure 5: Geodatabase and Map Contents

Code	Description
3	3in Diameter, Pint
5	5in Diameter, Quart
6	6in Diameter, 1 Gal
9	9in Diamter, 2 Gal
10	10in Diameter, 3 Gal
12	12in Diameter, 5 Gal
14	14in Diameter, 7 Gal

Figure 6: Pot_Size Domain

DATABASE MANIPULATIONS

In order to manipulate the database to answer questions, the user will employ queries and geoprocessing tools available in ArcGIS Pro. To simplify commonly utilized processes, workflows were created using ModelBuilder in ArcGIS Pro. The first workflow (*Figure 7*) automates the process for selecting markets within a user defined radius of Pattee Creek Ranch using the select by location geoprocessing tool. This allows the user to easily filter through which markets are closest in proximity. The second workflow was developed to solve the problem of identifying which plants were sold on a user-specified date, information that would be helpful for planning and analysis for future seasons of planting and selling. This process is performed by utilizing a table join and then a select by attribute tool in which the user can input a query with drop-down parameter options for simplicity (*Figure 8*). Many other research questions could be answered by applying similar methods developing queries and geoprocessing tools.

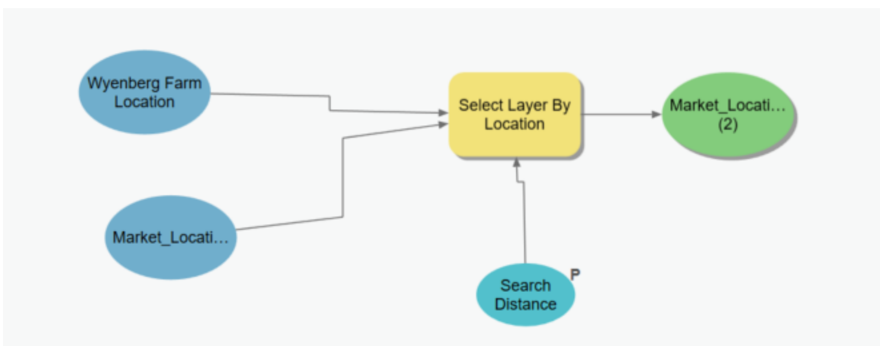

Figure 7: Select by Location in ModelBuilder

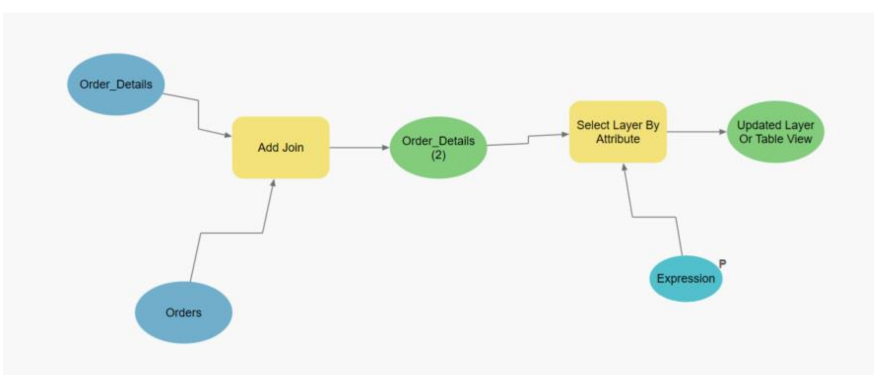

Figure 8: Add Join and Select by Attribute in ModelBuilder

Case Studies

In addition to the examples above, several other research questions were explored to demonstrate the functionality of the database. A combination of queries, geoprocessing and analysis tools give flexibility to answer a countless variety of questions pertaining to the vegetable starts business. Two examples and their respective results are presented below.

Example 1:

The user would like to take orders from customers and identify convenient farmers market pick-up locations based on location and proximity. To solve this problem, the user could apply a query to the Orders table by date, and then utilize the customer numbers from those orders to query the customer table with the select by attribute tool. Next, a buffer tool could be employed to create buffers around each farmer's market. Finally, a select by location or summarize within tool would be appropriate for identifying the number of customers within each buffer.

Example 2:

The user would like to identify which variety of plants are best-sellers. The Order Details table contains information on plant identification number and quantity sold in each order. To find the data needed, the Summary Statistics analysis tool could be applied, with parameters outlined as seen in Figure 9. Running this application results in a table of information organized by Plant_ID with a field identifying the frequency of that plant occurring in an order (regardless of quantity in each order) and a sum total of the quantity sold.

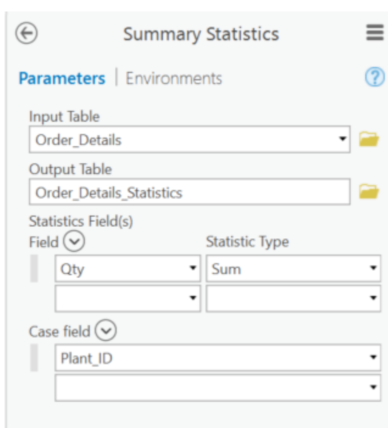

Figure 9: Summary Statistics Tool

Conclusion and Results

A functioning spatial database was designed that met all of the specifications outlined. This database is flexible, organized, and accessible to users. In the future, the database will be put into practice and will undergo transformations and improvements as the business grows and develops. Because the design was logical and methodical, future amendments or expansions will be straightforward.

REFERENCES

Cronin, Dana 2021, *Suppliers Field Growing Demand for Seeds from Pandemic Gardeners*, National Public Radio, accessed May 3, 2021, <https://www.npr.org/2021/02/05/964043089/suppliers-field-growing-demand-for-seeds-from-pandemic-gardeners>

Local Food and Flavors Guide, Iowa State University Extension and Outreach, accessed May 3, 2021, < https://www.extension.iastate.edu/flavorsofnorthwestiowa/farm-directory?field_geofield_distance%255Bdistance%255D=&field_geofield_distance%255Bunit%255D=3959&field_geofield_distance%255Borigin%255D=&field_county_tid=All&field_adventure_type_tid=228>