5405 Final Project: World Tree

Lexi Dingle, Jozie Wille,
Charlie Rudy & Drake Swezey

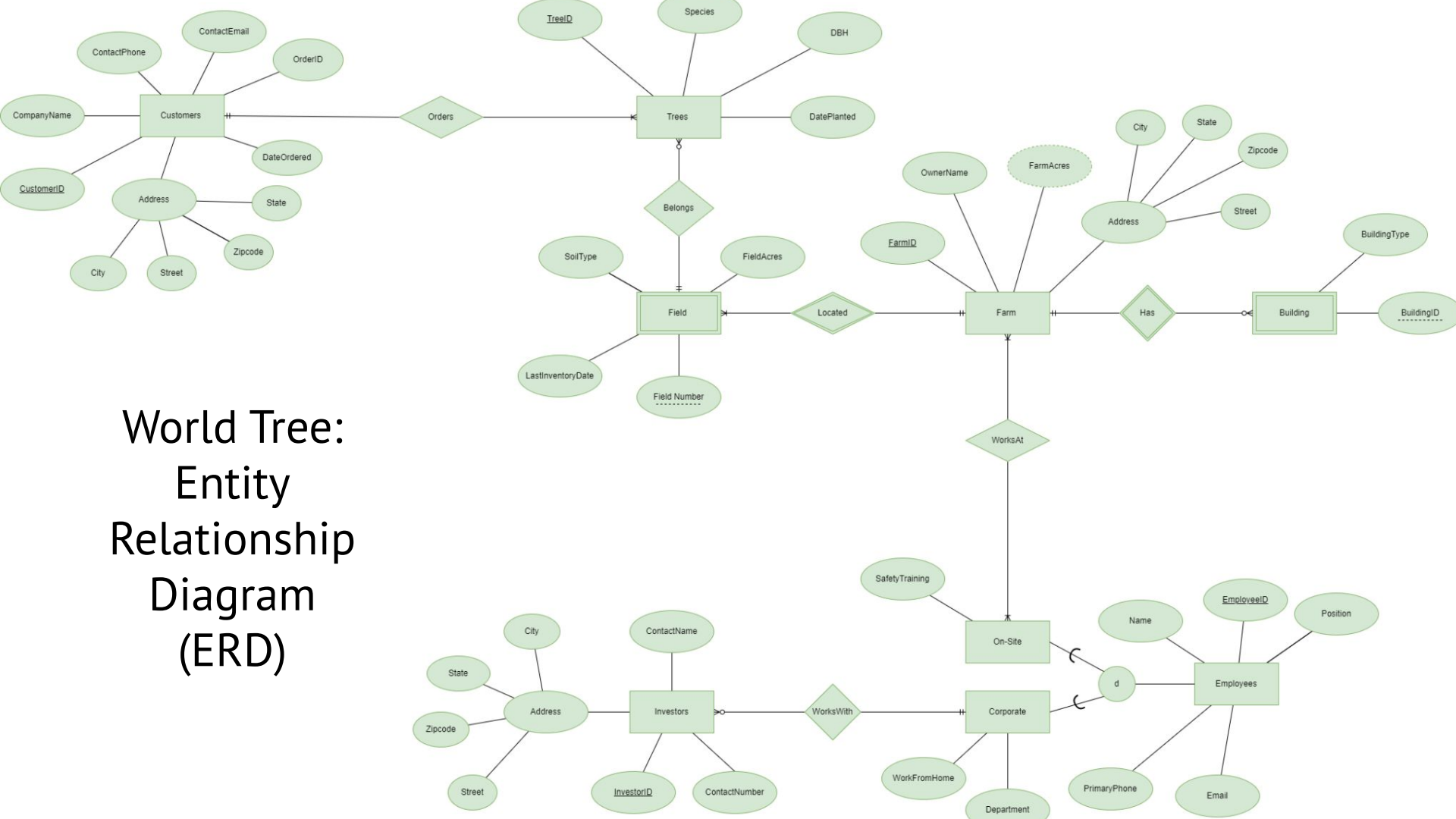
World Tree's Problem

- World Tree uses a fast-growing Paulownia Sp. type of hardwood tree to generate high-quality lumber while removing carbon dioxide from the environment.
- The purpose of World Tree is moving to remote sensing technologies so that precise data can be acquired for each individual tree in the various fields due to the time-consuming and labor-intensive nature of inventory keeping.
- This transition will require the implementation of a much more robust database than the excel files that are currently being used by the company, which is why we have created this new database.

ERD Assumptions and Non-Standard Decisions

- **Assumptions:**
  - 1 acre produces 162 trees, there is 196 board feet per tree
  - No trees die in the field while they are growing, each tree that is planted is sold eventually
  - **Sales Price per Tree:** 196 board feet x $5.89 per board foot = $1,154.44 per tree
  - The only trees that are recorded in our database are trees that have been sold to customers.
- **Non-Standard Decisions:**
  - Not every farm has employees, assuming these farms are “out of season” or not open yet
- **Species Key:**
  - 1 = Paulownia Sp. Tree
  - 2 = Oak Tree
  - 3 = Pine Tree
- **Soil Key:**
  - 1 = Alluvial Soil
  - 2 = Black Cotton Soil
  - 3 = Laterite Soil
  - 4 = Alkaline Soil
  - 5 = Saline Soil
  - 6 = Marshy Soil

World Tree: Entity Relationship Diagram (ERD)



Relational Schema Mapping of the ERD

- BUILDING(BuildingID, BuildingType, FarmID[FK])
- EMPLOYEES(EmployeeID, Position, Name, PrimaryPhone, Email)
- FARM(FarmID, OwnerName, City, Street, State, ZipCode)
- FIELD(FieldNumber, LastInventoryDate, SoilType, FieldAcres, FarmID[FK])
- TREES(TreeID, CustomerID[FK], Species, DBH, DatePlanted, FieldNumber[FK], FarmID[FK])
- INVESTORS(InvestorID, ContactName, ContactNumber, City, Street, State, ZipCode, EmployeeID[FK])
- CUSTOMERS(CustomerID, OrderID, DateOrdered, ContactEmail, ContactPhone, CompanyName, City, Street, State, ZipCode)
- ONSITE(EmployeeID[FK], SafetyTraining)
- WORKSAT(EmployeeID[FK], FarmID[FK])
- CORPORATE(EmployeeID[FK], Department, WorkFromHome)

Data Sources

The dataset we used for our project is a fictitious data set that we generated using both Mockaroo and Excel. We chose to focus on the topic of trees because Drake works in this industry therefore, it is a real world example. Due to it's real world applicability, we used this project to improve our SQL skills and explore scenarios where this dataset could be implemented.

Tables:

- Building
- Employees
- Farm
- Field
- Trees
- Investors
- Customers
- On-Site
- Works At
- Corporate

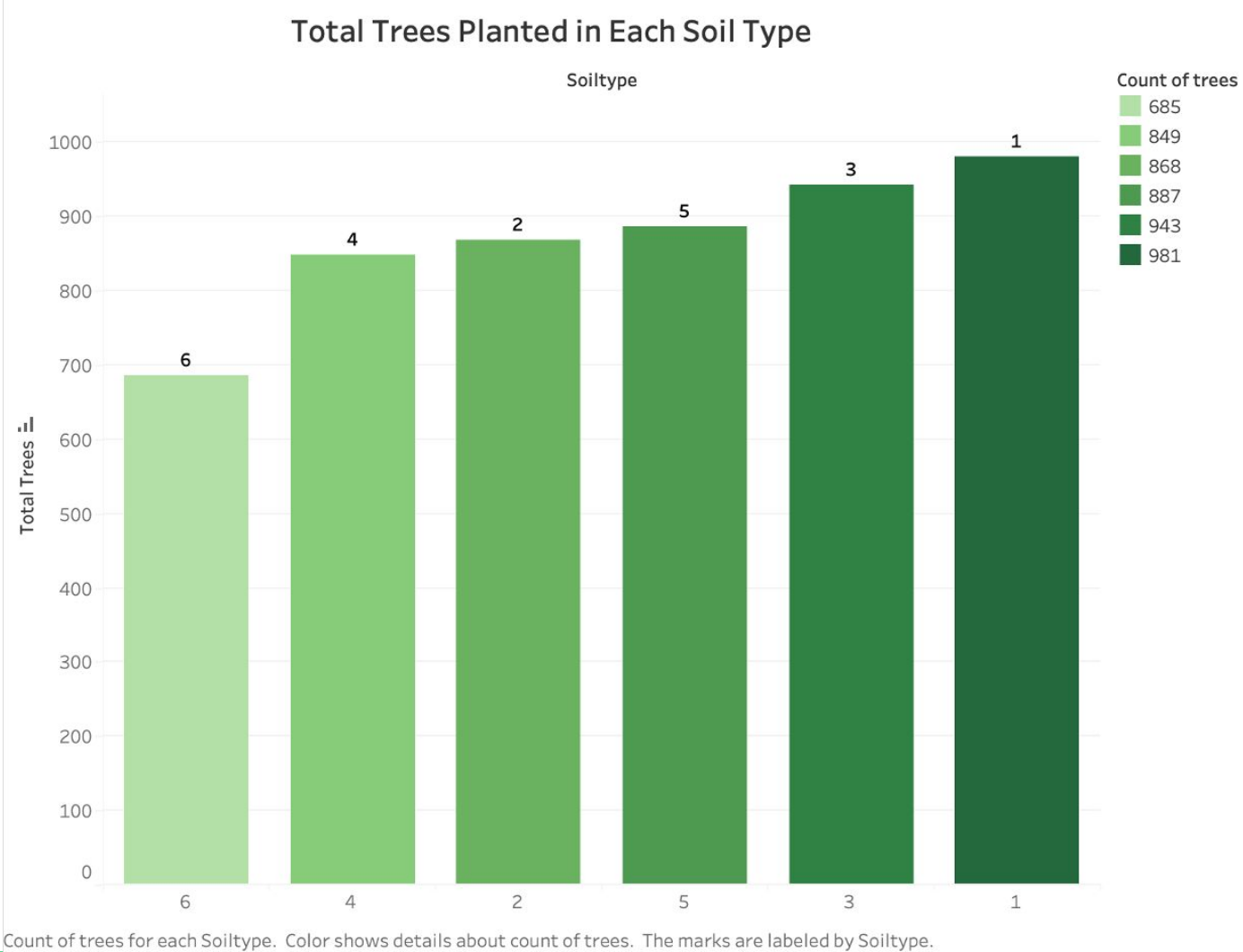
Query 1: The number of trees planted in each soil type by species

Explanation of Purpose: We want to see which tree species and soil type combination is selling best to customers. If one stands out, we can target customers to see why they wanted those trees. If one combination is selling at a much lower rate, we can discontinue the combination and work to grow a more desirable tree.

SQL Commands: *SELECT COUNT(TREEID) AS 'Number of Trees', SOILTYPE AS 'Soil Type', SPECIES FROM TREES JOIN FIELD GROUP BY SOILTYPE;*

Why report is helpful: Once we know what combination is selling the best, we will be able to increase that tree and soil combo to help increase overall sales. We could also look at the combinations of tree and soil in different geographic regions to see if place is affecting growth.

Number of Trees	Soil Type	SPECIES
9000	1	2
9000	2	2
11000	3	2
11000	4	2
8000	5	2
6000	6	2



Count of trees for each Soiltype. Color shows details about count of trees. The marks are labeled by Soiltype.

Query 2: Revenue per Customer with Order Date, Quantity and Shipping Address by State

Explanation of Purpose: The purpose of this query is to produce an output including revenue per customer, the date on which they placed their order, the state to which we would have to ship, and the quantity we would have to send.

Why this report is helpful: This report is helpful because the organization can use it as context to compare its current sales to its prior sales because it displays sales patterns by customer, state, and date of purchase. This report can also be used to forecast future sales, calculate the amount of trees that will need to be stocked at each farm, and estimate the number of sales that will be made there, all of which will increase earnings. This report can be used to comprehend client purchasing trends and provide guidance on how to increase customer retention.

SQL Commands: *select format(count(treeid)*196*5.89,2) as revenue, count(treeid)*196 as 'BoardFeet', customerid, DateOrdered, customers.state from trees natural join customers group by customerid;*

Sales Price per Tree: 196 board feet x $5.89 per board foot = $1,154.44 per tree (used for profit calculations) renamed 'BoardFeet' in SQL Query

revenue	'BoardFeet'	customerid	DateOrdered	state
130,451.72	22148	1	2022-07-27	GA
110,826.24	18816	10	2021-11-24	WA
192,791.48	32732	2	2022-07-13	FL
46,177.60	7840	3	2022-05-22	TN
218,189.16	37044	4	2022-01-21	FL
73,884.16	12544	5	2022-05-17	VA
55,413.12	9408	6	2021-11-21	DC
115,444.00	19600	7	2022-07-06	NY
113,135.12	19208	8	2021-11-29	WA
98,127.40	16660	9	2022-10-24	TN

Query 3: Customer Demographics and Investor Demographics

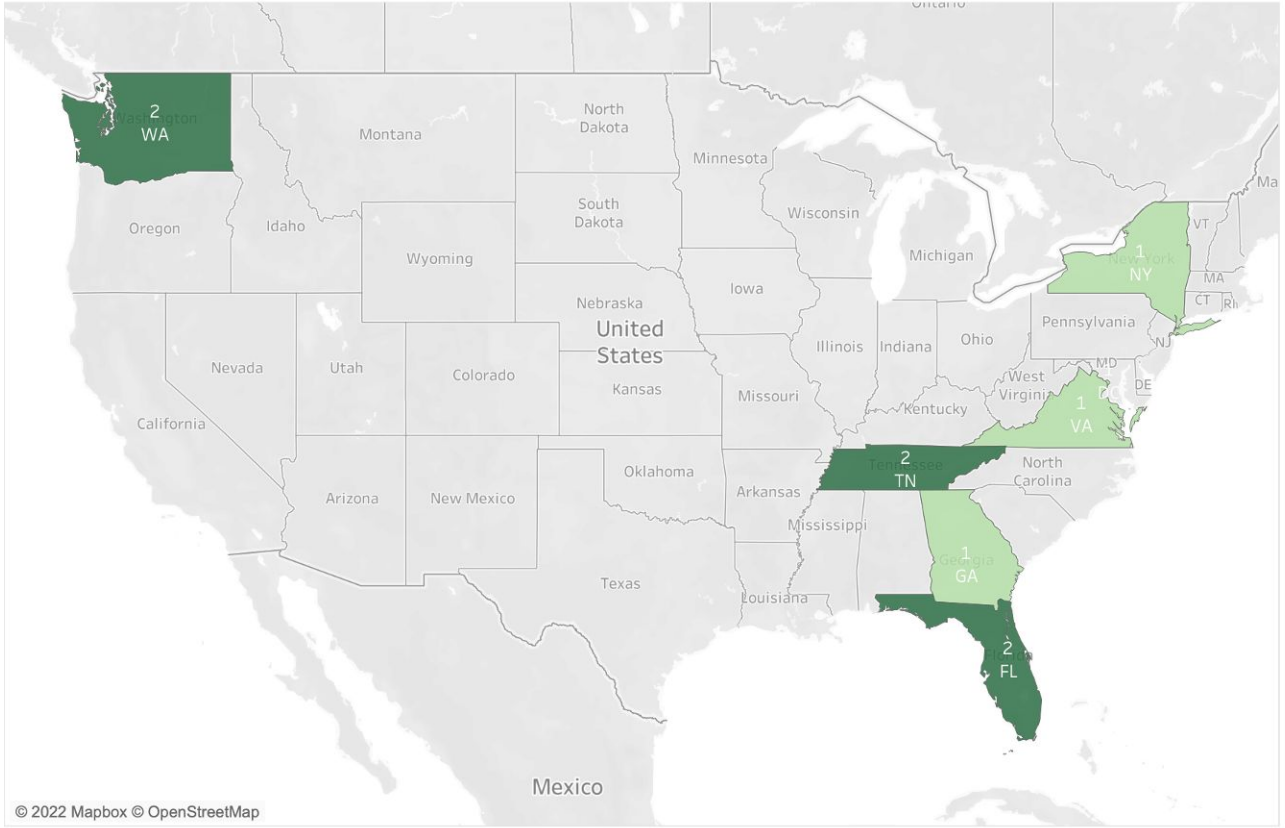
Explanation of Purpose: The purpose of creating a query that identifies demographics or purchasing behavior about customers or investors is so the marketing department can better understand their customers and investors in order to be proactive in marketing to and building relationships with them.

Why report is helpful: We wanted to run this query in order to have a better understanding of how we can help market to current customers as well as identify potential new customers. By understanding which states customers who purchase the most are located we may be able to identify marketing opportunities within those states. For example, we see FL, WA, and TN have higher counts of customers, the marketing team should look into why this is, and potentially focus efforts in these states. We are also able to see which customers are purchasing a lot from us which can offer some insight into which of our current customers deserve a lot of our attention. Ideally, we would also run some queries on where our investors are located, how much each invests in the company and other demographics when the database is more mature with this information.

SQL Commands:
SELECT COUNT(TREEID), COUNT(DISTINCT), STATE FROM TREES NATURAL LEFT JOIN CUSTOMERS GROUP BY STATE;

COUNT(TREEID)	COUNT(distinct((CUSTOMERID)))	STATE
48	1	DC
356	2	FL
113	1	GA
100	1	NY
125	2	TN
64	1	VA
194	2	WA

Total World Tree Customers Per State



Map based on Longitude (generated) and Latitude (generated). Color shows count of customers. The marks are labeled by State and count of customers. Details are shown for State.

Query 4: Field and Employee Top Performers

Explanation of Purpose: We wanted to determine the best performing fields and the employees responsible for them. To do this we found the average DBH and Marketable Height of our trees based on the field they are located in. Then we filtered only the fields in the 90th percentile for both measurements. Finally, we joined the worksat table in order to see which employees are responsible for these high performing fields

Why report is helpful: As a result of this query we can go to the high performing fields to gain insight on what is causing this effect. We also determined which employees are responsible for these fields in case they are one of the variables generating the performance.

SQL Commands: Farm Acres is derived from field; on field table *select Round(avg(DBH_CM), 2), Round(AVG(Marketableheight_meters),2), farmid, fieldnumber, employeeid from trees natural join worksat group by farmid, fieldnumber having avg(DBH_CM) >47*

union select Round(avg(DBH_CM), 2), Round(AVG(Marketableheight_meters),2), farmid, fieldnumber, employeeid from trees natural join worksat group by farmid, fieldnumber having avg(marketableheight_meters) >5.7

Round(avg(DBH_CM), 2)	Round(AVG(Marketableheight_meters),2)	farmid	fieldnumber	employeeid
47.21	5.52	C	2	003
47.09	5.39	D	1	004
47.68	5.55	H	2	008
46.74	6.03	D	2	004
43.98	5.73	G	1	007
44.71	5.76	J	3	010