Database Systems Course NMIMS Summer/Fall 2021

Final Project

Business Scenario:

Two children started a lemonade stand to earn some extra money at a local country club/golf course over the summer (June, July, and August). The golf course is open every day (weather permitting) except Mondays. The golf course built a lemonade stand structure with an ice-making machine and secure storing of inventory. Then, the golf course signed an agreement allowing the children to operate the lemonade stand on the golf course for a rental fee of 5% of sales. In the agreement, the golf course provides electricity and water (for their computer and ice machine).

The children were excited to offer 19 flavors of lemonade. They purchase the lemonade in individual containers (12 ounce cans) from the local grocery store. For weekdays (Tuesday through Friday), one child works for a fixed wage ($75 per day). On Saturday and Sunday, two children are scheduled to work each earning a fixed wage ($75 each per day).

One of the children has a parent who had recently gone back to school at the time the children were starting the Lemonade Stand and was taking a database course. One of the assignments for the course was to design a database. In order to complete the assignment and help the children, she offered to design a database so they could keep track of inventory and sales.

To track sales by employee, she designed an invoice for the children to enter sales (see Figure 1 below). They would like to be able to track sales by employee. They would also like to be able to track contact information (name, phone, address, city) for each employee.

The sales are recorded at the time of purchase. For each sale, the employee enters the member number, flavor(s), and the quantity of each flavor for the sale. A paper sales receipt is signed by the member to acknowledge receipt of the goods and kept on file. The country club/golf course has members (these members are customers of the lemonade stand) who have a membership card that works similar to a credit card for purchases at the club house, pro shop, and at the lemonade stand. The members are billed at the end of the month for purchases that month. Therefore, the lemonade stand must submit a monthly report detailing sales to each member to the club. The members remit a check directly to the lemonade stand.

Cash receipts are recorded by an employee when the payments are received. The record of receipt includes the date, employee ID, amount of the check, which bank account the check will be deposited in, the golf course id (only one location at the current time, but they hope to grow their business), and the member ID.

They would also like to track information about customers (members) including name, phone, address, city.

Right now, they only have one lemonade stand, but in the future, they hope to expand operations to other golf courses. As such, they would also like to track information about each location for the lemonade stand. Specifically, they would like to track the name of the golf course on which

the lemonade stand is located, the address of the golf course, and the name and phone number for the person in charge of the golf course.

Figure 1 – Lemonade Stand Sales Invoice

Figure 1: Example form

Every Monday, the children's parent takes them to the grocery store to purchase their inventory. Currently, the children mentally estimate both the items and quantity of inventory needed (create a grocery list). As needed, they buy cups and paper towels. The weekly grocery bill is paid by the parent. The trip to the store causes some frustration for the parent as the trip typically includes a long wait while the children decide what and how much to buy. At the end of the month, the children pay their parent for the sum of the month's purchases of inventory and provide a list of total sales and cash receipts recorded/deposited for the month.

Part 1: Understand the business from a narrative description

Reflecting on the above business scenario, demonstrate an understanding of the business processes and reporting needs of the business both internally and externally by answering the following thought questions.

1) What are the reports specifically required of the system?
2) What reports would management want to have from this system?

Part 2: Database Normalization and setup in Microsoft Access

1) Create an ER model for this business situation. Your model should include only entities and relationships, no attributes. Please use Crow's feet notation.
2) Based on the narrative given, you are required to implement a database that satisfies the above requirements using Microsoft Access.
3) Create a data dictionary for your ER model in a separate Microsoft Excel spreadsheet. All attributes including keys and foreign keys must be identified. Use these naming conventions for tables and attributes in your data dictionary. Entity names should match problem the domain and be understandable to users of the system. Attributes uses as keys or foreign keys being with c_, other attributes begin with letters that indicate their data type including, t_ for text d_ for date, n_ for numbers with decimal places, i_for integers, b_ for boolean, etc. Attributes should also have a two character suffix derived from the table name, e.g., _cu for a customer table. The primary key of the customer table would be something like c_customer_cu, the date of first order by the customer might by d_firstOrder_cu, and the current balance would be n_currentBalance_cu. I have included a sample data dictionary.
4) Add at least ten fictional (and appropriate) records to each table in the Microsoft Access database.

Part 3 - Understand the database structure

Using the solution ERD with the data provided, answer questions about where and how we can extract data/information from the system.

Which tables would you find the data needed to answer the following questions?

1) How many members/customers does the lemonade stand serve?
2) How many flavors of inventory are there? What are the names of the inventory items?
3) Based on the tables, do you think all lemonade sells for the same price? Why?
4) Does the system allow for discounts? For example, do high volume customers receive discounts?
5) Will the system allow the business to expand to multiple locations?

Part 4 – SQL to answer business questions

Develop queries to answer the following business questions. Save the queries in the format "qry_Question#" for example, the first query would be saved as "qry_Question1".

1) The children would like to be able to look up and quickly find a member's number. List all members' first and last names, member ID number, and order them alphabetically by last name (Lname).
2) How many (count of) cash receipts were received in August 2013?
3) What was the total cash received in August 2013?
4) What was the total cash received for the entire year of 2013?
5) How many flavors of inventory are in stock?
6) List inventory flavors on-hand from highest to lowest.
7) What are the total sales in dollars for the company for the month of June 2013?
8) What are the total sales in dollars for the company for the entire year of 2013?
9) What amount was sold to each member for the month of June 2013 (list from the highest to lowest amount)?
10) What were the cash receipts from Members in July 2013 (presumably for June sales) (list from the highest to lowest amount)?
11) What is the balance owed by each Member (Accounts Receivable balance) at July 31, 2013 (before billing for July) (list from the highest to lowest balance)?
12) What is the 2015 ending Acct. Rec. balance for every member listing the highest first?
13) Who is the best customer by volume (display only the best customer)?
14) How much did we sell (qty and $) of each product in June 2013?
15) What were the top 3 flavors in units sold for the month of June 2013?
16) What were the 3 lowest flavors in units sold (and list those not sold)?
17) Are there any members who have not purchased any lemonade? If so, list their name(s).
18) List the total sales units and dollars by shift (employee) (for all 3 years).
19) List the daily sales units and dollars (for all 3 years).
20) List the minimum, maximum, and average daily sales by employee.

Part 5: Application Development in Microsoft Access

Part 5a: Sales data entry form

Develop a form in Microsoft Access for users to enter in sales invoices. Your form should appear like Figure 2 below:

Figure 2: Sales Invoice form

Part 5b: Report in Microsoft Access. Develop a report in Microsoft Access that shows the Accounts Receivable balance by customer at September 2015. Your report should appear like Figure 3 below:

Member ID	First Name	Last Name	Balance
77	Carolyn	Turner	$22.00
20	Jack	Haas	$19.00
27	Bobby	Tway	$19.00
72	John	Cuellar	$16.00
43	Jacob	Sluman	$16.00
66	Othoniel	Ames	$16.00
79	Jerilyn	Chereta	$16.00
94	George	Goodelman	$16.00
51	Barry	Nauman	$12.00
53	Remedios	Brenner	$11.00
74	Rita	Startzel	$10.00
62	Adrienne	Bogo	$10.00
184	Cody	Hastings	$10.00
180	Jamie	Douglas	$8.00
8	Frank	Couples	$7.00
87	Tatiana	Cechova	$7.00
189	Abby	Mackelprang	$6.00

Figure 3

Part 6-Basic data analytics

Part 6a: Clean and import external data (weather data).

Table1 below shows some descriptive statistics of the children's' performance. EID 1 represents one child working Tuesday/Thursday alone, EID 2 represents one child working Wednesday/Friday alone, and EID3 represents both children working together on the weekends.

Answer the following thought questions.

1) What possible explanations might exist that lead to the variability in sales for each shift?
2) How might daily events/tournaments/weather impact sales?

Table 1 – Daily Sales Data By Employee

EID	Avg Daily quantity	Max Daily quantity	Min Daily quantity	Avg Daily Dollars	Max Daily Dollars	Min Daily Dollars
1	96.06	268	12	242.46	681	35
2	48.66	120	3	122.72	301	9
3	105.35	240	15	265.95	615	38

The owners of the lemonade stand want to know if weather affects daily sales. Many items like temperature, rain, or even cloud cover could influence sales. This knowledge is important for sales forecasting and budgeting purposes. For example, knowing that cloudy weather decreases sales would help the owners determine how much lemonade to purchase for resale.

In order to determine if weather affects sales, perform the following steps:

1 Acquire external data (weather). In practice, similar to what was pulled from the National Weather Service, data typically requires making sure we understand what is provided and that it is in a format usable for our purpose.

2) Open the Excel "weather" file and become familiar with the data (See Table 2).

Table 2 – Description (Data Dictionary) of Weather Variables

	Column	Description adapted from National Weather Service
1	DY	The day of the month.
2	MAX	The highest temperature for the day in degrees Fahrenheit (F).
3	MIN	The lowest temperature for the day in degrees Fahrenheit (F).
4	AVG	The average temperature for the day, computed by finding the average of the values in columns 2 and 3, then rounding
5	DEP	Departure from normal. The difference between column 4 and the 30-year normal temperature for this date. A minus (-) is number of degrees below normal. A zero (0) indicates that the average for that day was the Normal.
6a & 6b	HDD & CDD	Degree Day: A gauge of the amount of heating or cooling needed for a building using 65 degrees as a baseline. To compute heating/cooling degree-days, take the average temperature for a day and subtract the reference temperature of 65 degrees. If the difference is positive, it is called a "Cooling Degree Day". If the difference is negative, it is called a "Heating Degree Day".
7	WTR	Total precipitation for the day to the nearest .01". This includes all forms of precipitation (T = Trace, some precipitation fell but not enough to measure *for data file T=.01*).
8	SNW	Total snowfall for the day to the nearest tenth of an inch.
9	DPTH	Snow depth on the ground to the nearest inch at 1200UTC.
10	AVG SPD	Average wind speed for the day in miles per hour (mph).
11	MX SPD	The highest wind speed in mph averaged over a 2-minute period.
12	2MIN DIR	The direction (in compass degrees divided by 10) from which the wind speed in column 11 came from. (N=36, S=18, W=27, E=09, etc.)
13	MIN	The number of minutes of sunshine received at the station.
14	PSBL	The percentage of possible sunshine. Computed by dividing the minutes of sunshine in column 13 by the total possible minutes. Not reported at all locations.
15	S-S	The average sky cover between sunrise and sunset in tenths of sky covered. The minimum of "0" means no clouds observed, "10" means clouds covered the entire sky for that day.
16	WX	A coded number representing certain types of weather observed during the day. 1=Fog, 2=Fog reducing visibility to 1/4 mile or less, 3=Thunder, 4=Ice pellets, 5=Hail, 6=Glaze or rime, 7=Blowing dust or sand: visibility 1/2 mile or less, 8=Smoke or haze, 9=Blowing snow, X = Tornado

3) In practice, data files typically need to be cleaned before use. The weather file has several issues that need to be addressed before the data can be used. First, the date column is in numeric format. Second, rainfall has a "T" for a trace of rain.

    a) Date column issue: the number (ex. 10613 = 1st day of June, 2013) will not be read by Excel or Access as the desired date. If we use the Excel cell format to change from numeric (general) to a date format, it will incorrectly change the date to 01-29-1929. To fix the date column, we suggest inserting a column. Title the column heading in row 1 ("date"), and then set the format of rows 2, 3, and 4 to date (ex. 3/14/2013). Then, in rows 2, 3, and 4 of that column, type 06/01/2013, 06/02/2013, and 06/03/2013 respectively. Highlight those three cells and place your cursor at the bottom right of row 4 cell. Then, click and hold the left mouse button and drag down slowly to the end of the August 2013. Repeat for 2014 and 2015 summer dates.

    b) For the rainfall (WTR) column, scroll down the column and replace all "T" values with .01. (The "replace" command can be used for this step. Alternatively, a student could write a formula to convert.)

    c) Note the WX column identifies special weather events. See coding in Figure 4. In the data the most common are 1, 2, 3 and 8. If a day has a 138, the weather service noted 1 = fog, 3 = thunder, and 8 = haze. If a person wants to compare weather features to sales, the coding would need to be split into multiple columns. (This is not needed for our analysis.)

After cleaning the weather data, the weather data in your spreadsheet should appear like Figure 4:

Figure 4: Cleaned weather data

4. Save the spreadsheet for importing into MS-Access, and close the file. In MS Access, import the external data file from Microsoft Excel. Allow Access to assign the primary key.

5) Perform a query to match the daily weather data to the daily sales data.:
  a. Create a query that includes total daily sales ($ and units)
  b. Create a query using a "daily sales query" and the "weather table" joined on date.
  c. The query should have sales data and weather data for all days the business was open for all three years.
6) Extract your query in part 6a to a spreadsheet.
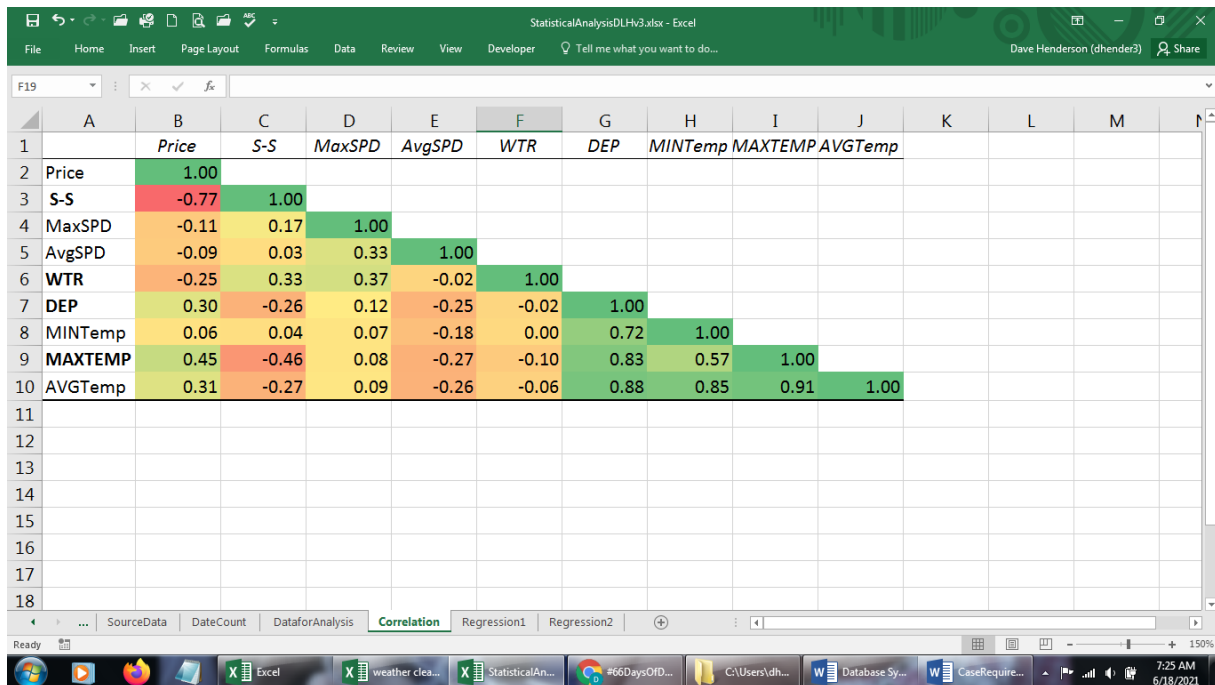
Part 6b: Correlation analysis

This deliverable requires you to conduct a correlation based on the sales weather spreadsheet you created in step 6a. The following steps are required for this case:

1) Rename your worksheet with your sales and weather data to "SourceData".
2) Copy the data to another spreadsheet tab and rename this tab "DataforAnalysis". You need to arrange the data on this tab so that the data appear like Figure 5. You can arrange the data manually or you could use a pivot table. To use a pivot table in Excel:
   a. Create a pivot table as prep for the correlation analysis.
   b. Insert the pivot table on a separate worksheet labeled "DataforAnalysis". Once you have created your pivot table, Edit→Paste special values. You can also complete this step without a pivot table by removing certain fields from your source data (Qty for example). Your pivot table should look like Figure 5:

Date	Price	S-S	MaxSPD	AvgSPD	WTR	DEP	MINTemp	MAXTEMP	AVGTemp
6/1/2013	275	6	16	8.9	0.01	3	68	85	77
6/2/2013	133	8	22	8.2	0.33	4	70	85	78
6/4/2013	261	5	14	5.8	0	4	69	89	79
6/5/2013	91	8	18	7	4.14	3	70	85	78
6/6/2013	54	9	12	4.9	0.01	-3	69	77	73
6/7/2013	57	8	24	6.4	0.01	0	68	83	76
6/8/2013	61	9	10	4.6	0	-1	67	83	75
6/9/2013	95	8	20	6.2	0.81	-1	70	80	75
6/11/2013	298	5	16	9.4	0	4	71	90	81
6/12/2013	152	6	15	7.3	0.01	6	74	91	83
6/13/2013	491	3	43	11.2	0.85	2	67	91	79
6/14/2013	131	4	15	7.3	0	0	68	86	77
6/15/2013	535	3	14	6.3	0	-1	67	84	76
6/16/2013	377	5	22	5.4	0.45	-1	67	86	77
6/18/2013	149	7	22	8.1	0.11	-1	68	86	77
6/19/2013	99	7	14	6.9	0	-2	69	83	76
6/20/2013	378	3	17	9.3	0	-1	69	84	77
6/21/2013	224	3	14	7.3	0	-1	69	85	77
6/22/2013	544	2	12	5.3	0	-2	68	85	77
6/23/2013	193	7	10	4.2	0	-2	69	84	77
6/25/2013	198	7	14	4.6	0.01	1	71	88	80
6/26/2013	100	6	15	7.2	0	1	71	88	80
6/27/2013	198	7	24	9.9	0.13	2	74	88	81
6/28/2013	88	7	20	8.1	0.01	4	73	92	83

Figure 5: Data ready for correlation analysis

3) Create a correlation matrix using the data analysis toolpack and save this matrix on a separate worksheet. Use conditional formatting (color scales) to highlight the correlations. Your correlation matrix should appear as shown in Figure 6:

	Price	S-S	MaxSPD	AvgSPD	WTR	DEP	MINTemp	MAXTEMP	AVGTemp
Price	1.00								
S-S	-0.77	1.00							
MaxSPD	-0.11	0.17	1.00						
AvgSPD	-0.09	0.03	0.33	1.00					
WTR	-0.25	0.33	0.37	-0.02	1.00				
DEP	0.30	-0.26	0.12	-0.25	-0.02	1.00			
MINTemp	0.06	0.04	0.07	-0.18	0.00	0.72	1.00		
MAXTEMP	0.45	-0.46	0.08	-0.27	-0.10	0.83	0.57	1.00	
AVGTemp	0.31	-0.27	0.09	-0.26	-0.06	0.88	0.85	0.91	1.00

Figure 6: Sample Correlation

4) Using the correlation analysis in figure 6, which variables do you think would best predict price? Why?

Part 7-Presentation

1) Develop a presentation using PowerPoint (or other software) that addresses the following. You will present to the class virtually on either October 12th or October 14th.

2) Your presentation should be **NO longer than 10 minutes** (allow 5 minutes for questions) and should address the following points:

   a. Provide an overview of the business scenario

   b. Using the information provided by the queries in parts 3 and 4, provide an analysis of how well the lemonade stand is performing. Your analysis should answer the following questions:

      i. How good are sales and cash receipts for 2013?

      ii. Is the lemonade stand doing a good job collecting cash on their sales? Why is this an issue the lemonade stand needs to think about?

      iii. Which lemonade flavors should the lemonade stand focus on in the future? Are there some flavors that maybe they shouldn't sell anymore?

      iv. Which employees appear to generate the most sales? What about the least sales?

v. What is the effect of weather on sales? How does understanding the effect of weather on sales improve our understanding of the business?

vi. Who are the lemonade stand's best customers? Who are the worst customers? How should the lemonade stand use this knowledge?

vii. Can you think of any IT security or control risks that exist in the database or the sales process? Would there be ways for employees to steal inventory using the database for example?

c. Demonstrate your form and sample report (part 5).

Deadline for project: October 11ᵗʰ 2021. Answers to questions in part 1 in a document (e.g., Microssoft Word document)

1) **ER model for part 2.**
2) **Microsoft Access database with sample records.**
3) **Data dictionary in a spreadsheet (e.g., Microsoft Excel)**
4) **Microsoft Access database with queries for parts 3 and 4.**
5) **Microsoft Access database with sample form and report for part 5.**
6) **Microsoft Excel spreadsheet with cleaned weather data (part 6). The cleaned weather data should also be imported into your Microsoft Access database (part 6a)**
7) **Microsoft Excel spreadsheet with correlation analysis (part 6b).**
8) **Microsoft PowerPoint presentation addressing part 7.**

References:

This case is based on the following:

Schafer, B., and Bee, S. (2018). The Lemonade Stand: An Elementary Case for Introducing Data Analytics. AIS Educator Journal. Volume 13, Number 1, pages 30-43.