CIS 2334 Semester Project

Part 2, Due November 4th, 2024, at 6 pm

Background Information

The Marine Biologists research team are satisfied with the data that you have loaded into your database from the Semester Project Part 1. Now that the data is safely stored in a database, the research team wants an excel worksheet application to perform some complex data analysis.

Project Tasks

To satisfy the research team's requirements, you must use your Excel skills to perform these 6 tasks.

1. Get the required sample data from the database you created for your Semester Project 1. (10 points)

   What you need to do:

   a. Use MySQL Workbench, select the database and run the following query:

   SET @StudentNumber = YOUR_COUGAR_ID;

   SELECT * FROM MYTABLE
   ORDER BY RAND(@StudentNumber)
   LIMIT 1000;

   In this query, "YOUR_COUGAR_ID" is your own Cougar ID; and "MYTABLE" is the table name you used to store the abalone data in the database for Project 1.

   NOTE: For this task you need to use the Abalone database after you deleted the data described in Task 6 from Project 1.

   If you had a wrong answer in Task 6 (Please check the IA feedback comment for information), then you need to redo Project 1's Task# 6 with the original database.

   Also, check question 5, only 1 new distinct abalone row should have been added. ('4177', 'M', '0.71', '0.555', '0.195', '1.9485', '0.9455', '0.3765', '0.495', '12', '1', 'Jan', 'Odom', 'University of Houston', '1', 'Florida Atlantic Coast', '22.1')

   *** In case the last digit of your student ID is 0, you can change it to 4. So, 4/4 = 1.**

   *** In case your last digit is not 0 but your database does not have enough data to generate 1000 rows of data after the deletion described in Project 1, Task #6, then please redo the steps needed in Project 1's, Task #6 using the last digit of 4.**

b. Export the results from this query and save the results as a ".csv" file using the name "Firstname_Lastname_Personal_Data.csv". This csv file will be the data you will use for every task in this project. **Note** – The above query generates a unique data file for each student.
c. Using Excel, save the "Firstname_Lastname_Personal_Data.csv" file as "Firstname_Lastname_ExcData.xlsx" file.

   Note – only the .xlsx file type allows you to perform various excel functions for the rest of the tasks for this project.

d. Name your current worksheet/tab "Personal Data".

2. Categorize the abalones. (10 points)

The Marine Biologists discovered that certain abalones have some rare *qualities*. To further investigate these rare qualities, the research team divided the abalone data into three categories: **Category I**, **Category II**, and **Category III**. They believe that **Category I** and **Category II** abalones have important research values. **Category III** abalones are normal abalones with negligible research values. The following are the standards that this group of marine biologists are using to categorize abalones. Each category contains several different rules.

At the top of your worksheet, you will display all the rows of data belonging to **Category 1**. Then, you will next place all the rows of data belonging to **Category 2**. Finally, you will display all the rows of data belonging to **Category 3**.

a. If an abalone sample satisfies one of the following criteria, it belongs to **Category I**:
   1) Ring > 15 AND Infant.
   2) Male AND Length > 0.75.
   3) (Shell weight > 0.8 AND Shucked weight > 0.5) OR (Shucked weight > 1.2).

b. If an abalone sample **does not belong to Category I and satisfies** one of the following criteria, it belongs to **Category II**:
   1) (Shell weight < 0.4) AND (Shucked weight < 0.4).
   2) (Ring > 15 AND Male) OR (**RING** > 18 AND Female).
   3) Length < 0.36.

c. If an abalone sample doesn't belong to **Category I or Category II**, then the sample belongs to **Category III.**

What you need to do:

1) Create a new worksheet in the "Firstname_Lastname_ExcData.xlsx" document.
2) Name your worksheet "Categorization".
3) Copy the Personal Data into the "Categorization" worksheet/tab.
4) Implement the criteria defined above and separate the data into three Categories. For each Category create a data table in the "Categorization" worksheet/tab.

Make sure that you design the worksheet using all the design tools that you have learned in module 1 and module 2 of the textbook. For instance, add explanations and descriptions for each Category, add proper titles for each category, use different color scheme, apply proper cell format and so on.

Note – your work sheet should be professional, easy to read, and easy to understand.

3. Calculate the economic values of abalones. (10 points)

The Marine Biologists are also interested in the economic value of the abalones and would like to investigate several problems. First, they would like to know the values of every single abalone that has been collected.

They have provided you with the current market standards used to estimate an abalone's monetary value:

Normal Abalone value = [1 + 1/3 (Length – 0.5) + 1/3 (Diameter– 0.4) + 1/3 (Height– 0.4)] * Whole_weight * $0.5.

a) In addition, If the abalone belongs to **Category I**, its value will be multiplied by 1.5.
b) If the abalone belongs to **Category II**, its value will be multiplied by 0.8.
c) If the abalone does not belong to **Category 1** or **Category 2**, then the abalone belongs to **Category 3** which has a normal abalone value.
d) The Marine Biologists would like to know the average value for each gender.
e) Finally, they would like to know the average value for each water region.

What you need to do:

a. Create a new worksheet/tab in the "Firstname_Lastname_ExcData.xlsx".
b. Name your worksheet "Economic Values".
c. Compute the economic value for each abalone.
d. In the same worksheet, use another section of the worksheet to compute the average abalone value for each gender; for each gender find the most expensive and the least expensive abalone.
e. In the same worksheet, use another section of the worksheet to compute the average abalone value for each water region; for each water region find the most expensive and the least expensive abalone.
f. Design the worksheet using all the design tools that you have learned in module 1 and module 2 of the textbook. For instance: add explanations and descriptions for each table, add proper titles for each table, use different color scheme, apply proper cell format and so on.

Note – your work sheet should be professional, easy to read, and easy to understand.

4. Conduct a basic statistical analysis on the abalone data. (25 points)

It is important to conduct statistical analysis because statistics will help explore the fundamental characteristics of the data. As a result, the research team has asked you to create a worksheet to do the statistical analysis.

What you need to do:

a. Create a new worksheet/tab in the "Firstname_Lastname_ExcData.xlsx".
b. Name your worksheet/tab "Statistics".
c. Compute the mean, variance, minimum, and maximum values for "Length", "Diameter", "Height", "Whole_weight", "Shucked_weight", "Viscera_weight", "Shell_weight" and "Rings".
d. Create a box-and-whisker plot for "Whole_weight" for each gender. Use the data from your Personal Data worksheet.
e. Use the proper chart to review the relationship between "Shell_weight" and "Shucked_weight". Use the data from your Personal Data worksheet.
f. Use the proper chart to show the distribution of the "Diameter" for each gender.
g. Use the proper chart to show the average "Height" for each "Rings" value.
h. Design the worksheet using all the design tools that you have learned in module 1 and module 3 of the textbook. For instance: add explanations and descriptions for each table, add proper titles for each table, use different color scheme, apply proper cell format and so on.

Note – your worksheet should be professional, easy to read, and easy to understand. You will need to design the charts/plots using all the design principles that you have learned in Module 4. For example, please add proper titles for each chart, add proper titles for each axis, add a legend, and use a different color scheme for different data series and so on.

You will not get credit if you don't use the proper type of chart. Your charts should be easy to read, understand and have all the necessary elements.

5. Create fast search functions for the data. (30 points)

The research team constantly ran into situations when they needed to locate certain abalones in the data set. It is inefficient to eyeball or manually scan such a huge data set. Therefore, they ask you to create a few search functions. First, they want to be able to retrieve a particular abalone if its ID is given. Secondly, they want to know the *row number* of an abalone if all the following attributes are given together: "Length", "Diameter", "Height", "Whole_weight", "Shucked_weight", "Viscera_weight", "Shell_weight" and "Rings". Thirdly, they asked for a pivot table showing summary information for every water region. Finally, if a collector's name is given, and the water region is given, the team wants to locate all the abalones collected by this given collector from this given water region.

What you need to do:

a. Create a new worksheet/tab in the "Firstname_Lastname_ExcData.xlsx".
b. Name your worksheet/tab "Search functions".
c. Create a search function that has an input cell where the abalone ID is entered and has some output cells that display all the information about the corresponding abalone. Use the data from your Personal Data worksheet. For example, if you type "1" in the input cell, the output cell will display something like:

Gender	Length	Diameter	Height	Whole_weight	Shucked_weight	Viscera_weight	Shell_weight	Rings	Collector	Collector_First	Collector_last	Collector_organization	Water_region	Feb_temp
M	0.455	0.365	0.095	0.514	0.2245	0.101	0.15	15	15	Conan	Dalton	University of Houston	West Coast	12.3°C

d. Create a search function that has several input cells where the following information is entered: "Length", "Diameter", "Height", "Whole_weight", "Shucked_weight", "Viscera_weight", "Shell_weight" and "Rings. The function should have as an output the abalone's *row number* in the Personal Data worksheet/tab. For example, in my own Personal Data, the output cell will return "2", if I enter the following information in the input cells.

Gender	Length	Diameter	Height	Whole_weight	Shucked_weight	Viscera_weight	Shell_weight	Rings
M	0.455	0.365	0.095	0.514	0.2245	0.101	0.15	15

e. Create a pivot table, showing the "Average of Shucked_weight", "Max of Viscera_weight", and "Min of Shell_weight" of the abalones for each water region.

f. **(Optional Task – Possible 5 Extra Credit Points)** Create a search function that uses as input cells: "Collector_First", "Collector_last" and "Water_region". The output should be all the abalones collected by that collector from that water region. For example, if I enter "Kali", "Metcalfe", and "West Coast", the output cell will return the following data.

132	F	0.44	0.35	0.125	0.4035	0.175	0.063	0.129	9
144	M	0.56	0.455	0.155	0.797	0.34	0.19	0.2425	11
…	…	…	…	…	…	…	…	…	…

Note – "…" means that the rest of the records that Kali Metcalf collected from the West Coast water region are not printed out on these instructions.

Design the worksheet using all the design tools that you have learned in module 1 and module 2 of the textbook. For instance: add explanations and descriptions for each table, add proper titles for each table, use different color scheme, apply proper cell format and so on. Your work sheet should be professional, easy to read, and easy to understand.

6. Staff new scientists. (15 points)

The research team is expanding and has hired 5 new scientists to collect abalone samples. Each new scientist has variable costs when working in different water regions. The table below is showing the costs in dollars per minute:

	Florida Atlantic Coast	Florida Gulf Coast	Northeast	Mid-Atlantic	Southeast	Gulf of Mexico	Great Lakes	West Coast	North Pacific	Pacific
Beatrice Roberts	7	4	4	7	8	3	9	6	8	4
Alivia Allen	7	3	6	4	5	7	4	3	3	7
Tori Terry	4	2	10	7	3	2	9	2	2	5
Angela Crawford	4	8	5	6	9	4	3	2	3	7
Axel Anthony	8	10	4	8	8	6	4	5	9	10

The newly hired scientists must be allocated in such a way that all water regions are covered. Each scientist can work at most in two different water regions. The following table is an example of such scientists' allocation.

	Florida Atlantic Coast	Florida Gulf Coast	Northeast	Mid-Atlantic	Southeast	Gulf of Mexico	Great Lakes	West Coast	North Pacific	Pacific	sum scientists
Beatrice Roberts	1	0	0	0	0	0	0	0	1	0	2
Alivia Allen	0	1	0	0	0	0	0	1	0	0	2
Tori Terry	0	0	1	0	0	0	1	0	0	0	2
Angela Crawford	0	0	0	1	0	1	0	0	0		2
Axel Anthony	0	0	0	0	1	0	0	0	0	1	2
sum water	1	1	1	1	1	1	1	1	1	1	

Your task is to figure out the optimal allocation for the new scientists, minimizing the total cost when allocating the 5 new scientists.

What you need to do:

a. Create a new worksheet/tab in the "Firstname_Lastname_ExcData.xlsx".
b. Name your worksheet/tab "Optimal Allocation".
c. Solve the above problem and find the optimal solution. **HINT**: Use Solver and set Solving Method to Simplex LP.

Design the worksheet using all the design tools that you have learned in module 1 and module 2 of the textbook. For instance: add explanations and descriptions for each table, add proper titles for each table, use different color scheme, apply proper cell format and so on.
Note – your work sheet should be professional, easy to read, and easy to understand.

What you need to submit on Canvas: 1 file

"Firstname_Lastname_ExcData.xlsx".