CIS 2334 Semester Project

Part 2

The marine biologists research team are satisfied with the databased you have built. Now, the data is safely stored in database. The research team wants a professional excel application to perform some complex analysis. Your excellent excel skills will be required to build such an application.

The research team has specified the functionalities of such application. To satisfy the research team's requirements, you must use your excel skills to perform the 6 following tasks:

1. Get the sample data (the data you will use for semester project part 2) from the database.

First you need to acquire the sample data from the database you have previously built.

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; "MYTABLE" is the table name you have used to store the abalone data in the database.

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 all the next tasks. Note – The above query generates a <u>different</u> data file for every student.

c. Using Excel, save the "Firstname_Lastname_Personal_Data.csv" file as "Firstname_Lastname_ExcData.xlsx" file. Note – only .xlsx file allows you to perform various excel functions in the next tasks.

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

2. Categorize the abalones. (2% of the final grade)

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 abalone with negligible research values. The following are the standards that the marine biologists are using to categorize abalones. Each category contains several different rules.

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, it belongs to Category III.

What you need to do:

a. Create a new work sheet in the "Firstname_Lastname_ExcData.xlsx" document.
b. Name your work sheet "Categorization".
c. Copy the Personal Data into the "Categorization" worksheet/tab.
d. Implement the criteria defined above and separate the three Categories. For each Category create a data table in the "Categorization" worksheet/tab.
d. 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. (2% of the final grade)

The marine biologists are interested in the economic value of the abalones and would like to investigate several problems. Firstly, they would like to know the values of every single abalone that has been collected. They have provided you with the market standards used to estimate an abalone value:

Abalone value =

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

In addition, If the abalone belongs to Category I, its value will be multiplied by 1.5; If the abalone belongs to Category II, its value will be multiplied by 0.8.
Secondly, they would like to know the average value for each gender.
Lastly, 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 region to compute the average abalone value for each gender; for each gender find the most expensive and the less expensive abalone.
e. In the same work sheet, use another region to compute the average abalone value for each water region; for each water region find the most expensive and the less 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. (3% of the final grade)

It is of great importance to conduct statistical analysis because statistics review the fundamental characteristics of the data. The research team asked you to create a work sheet 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.
e. Use the proper chart to review the relationship between "Shell_weight" and "Shucked_weight".
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 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.
i. Design the charts/plots using all the design principles that you have learned in module 4. For instance: add proper titles for each chart, add proper titles for each axis, add legend, use different color scheme for different data series and so on.
   Note – 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. (4% of the final grade, 5% if you do the bonus question)

The research team constantly ran into situations when they need to locate certain abalones in the data set. It is inefficient to eyeballing such huge data set. Therefore, they ask you to create a few search functions. Firstly, 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. Lastly, 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 displays all the information about the corresponding abalone. 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	0.5514	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. **(Bonus 1% extra on top of the final grad)** Create a search function that use 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 West Coast are skipped.

g. 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.

6. Staff new scientists. (2% of the final grade)

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 region. The table below is showing the costs in dollar 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 new hired scientists must be allocated in such a way that all water regions are covered. Each scientist can work at most in two 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.
d. 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 in blackboard:

"Firstname_Lastname_ExcData.xlsx".