Blood Donation Management System in SQL

JOO Jung Eun (3035552246)
GUN Chun Ho (3035466320)
SHI Daniel Xing-Han (3035435890)

Table of Contents

Abstract

This report will detail the workings of an SQL blood donation management system and how it can be used to gather information and statistics about trends in the data. We use operations in SQL to discover the relationship between blood types. We will then use the trends to deduce and discuss certain health-related patterns in the world. To be specific, the six questions we will attempt to answer are:

1. Which blood type is contributed the most?
2. Which blood type is needed the most?
3. Which age group people donate blood the most?
4. Which age group people require blood the most?
5. Which gender donate blood the most?
6. Which gender require blood the most?

While the four questions seem trivial, we believe there will be additional trends revealed as we analyze the data. We will address any additional details in the Discussion and Conclusion sections of this report.

We will refer to the aforementioned six questions by their respective question numbers (Questions 1-6) when needed.

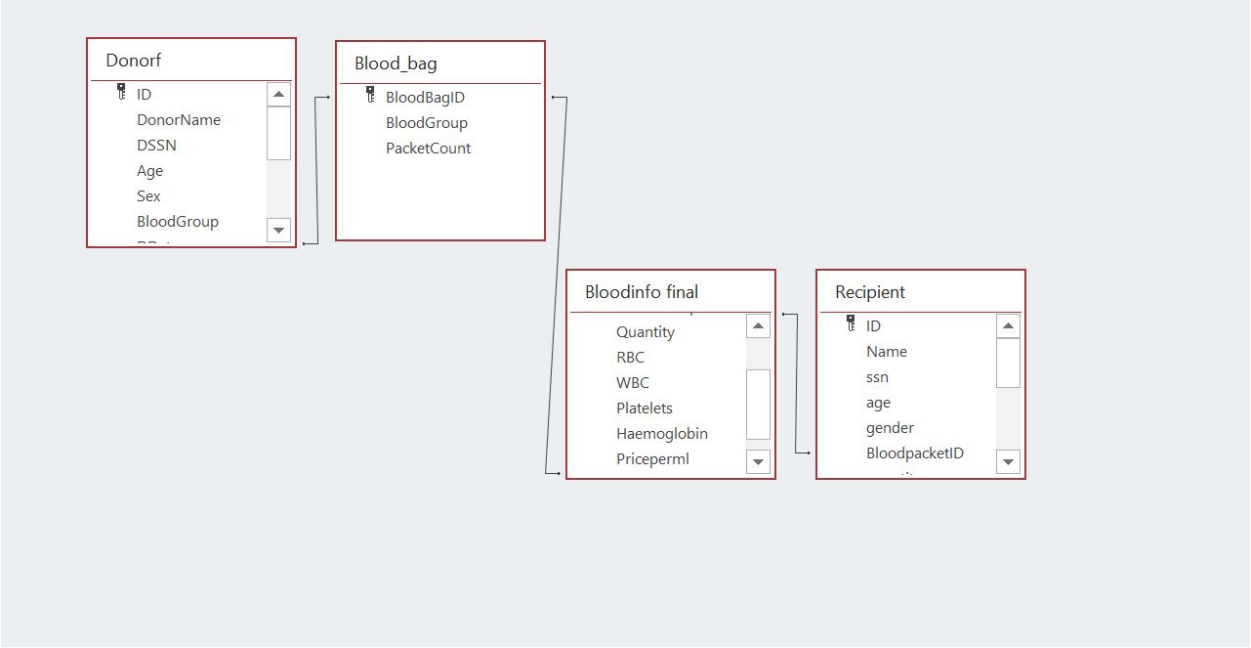
Motivation

Given the current health crisis which is occurring in conjunction with COVID-19, we have decided to focus on a project which can improve the efficiency of health related activities. Blood donation is a great example in which a database should be implemented because of the large number of donors and other variables that need to be tracked (donor information, equipment, blood type, compensation, etc.). It is also important that we discover certain trends within the vast amount of data so that resources in the health community can be directed to the right people in an accurate manner.

Model

As you can see from the Entity-Relationship diagram, our core information is the Blood_bag table. From the donor, it can contribute to the blood_bad table. For example, when a person who donated the blood, his/her information like name, age, sex, bloodgroup and diseases.etc will be stored into the donor table and then the blood he/she donated will be assigned the blood packet_ID, the same blood group will be stored inside the same BloodBag, having the same bloodbag_ID. For example, blood packets 1001 and 1002 are A+ blood then both of them will be stored in the BloodBag 501.

Next, the recipient who received the blood will be stored in the recipient table which will contain their information like name, age, gender. Etc. The recipient will also be linked to the bloodinfo table which contains the blood group, quantity, amount of red blood cell and white blood cell in each blood packet. The bloodinfo table will also contribute to the information of blood contained in the blood bag.



Data

The dataset is obtained from https://github.com/vishakh2494/DataBase . This dataset contains 15 donors and 12 recipients and then we added on some donors to the number of donors reached 46. The downloaded dataset is in SQL code. The structure of the data has been re-designed to fit in a relational database schema as shown in the Figure below. That is, normalization has to be done and primary keys are to set up correctly to link up each database table.

The columns and data types of each table in the database are listed below:

Table: blood_bag

Name	Type
Bloodbag ID	Number
Blood Group	Short Text
Packetcount	Number

Table: Bloodinfo-final

Name	Type
Bloodpacket	Number
Blood group	Short Text
Quantity	Number
Red Blood cell	Number
White Blood cell	Number
Platelets	Number
Haemoglobin	Number
Priceperml	Number
BloodBagID	Number

Table: Donorf

Name	Type
DonorName	Short text
ID	Number
DSSN	Number
Age	Number
Gender	Long Integer
BloodGroup	Short text
DDate	Number
Quantity	Number
PreownDisease	Short text
Address	Short text
Contact	Number
Bloodbagid	Number
BloodID	Number

Table: Recipient

Name	Type
ID	Number
Name	Short Text
Ssn	Number
Age	Number
Gender	Short Text
Quantity	Number
Hospital	Short Text
Doctor	Short Text

Address	Short Text
Contact	Number
Notice	Short Text
BloodBagID	Number
BloodGroup	Number

Methodology

In terms of the work we have done, there are two main components to our project: identifying the variables necessary to answer the questions we have put forth and creating a summary of the answers to our 6 questions.

The database management system that we have modelled our questions after has dozens of variables; however, the questions we have put forth only require a few of these variables. Hence, we have chosen the variables which we see fit to analyze and discuss our data. Variables that we need include blood type, age, patient name. The variable blood bag ID is chosen in order to specify a primary key.

To answer our questions, we first specify the specific blood types and age ranges. There are four main blood types (O, A, B, AB) and two subtypes (+, -) for each main type for a total of eight blood types. For age range, we will set the age ranges to be 16-26, 27-36, etc.

To investigate the characteristics of the blood banks, 7 queries are created. There are 46 donors in total in the blood bank with different Blood groups. The first query aims to find out which blood group has contributed the most to the blood bank. In order to count different blood types which was defined as characters, we decided to create a blood bag ID for each blood types. For example A+ is blood bag ID 501 ..etc. Since the original data didn't have this column of blood bag ID, we moved the data to microsoft excel and used VBA code to create a column which fills the blood bag ID corresponding to the blood type.

VBA coding function

```
Sub fill()

Dim outRow, rowNum, outSheet, i
Dim value As String

outRow = 6
outSheet = "Donor"

Worksheets(outSheet).Activate

For outRow = 6 To 51

value = Cell(outRow, 8).value
Cell(outRow, 14) = Switch(value = "A+", 501, value = "A-", 502, value = "B+", 503, value = "B-", 504, value = "O+", 505, value = "O-", 506, value = "AB+", 507, value = "AB-", 508)
Next


End Sub
```

Then we imported the data to ms access Since one of the main aim for our project is to find out the blood storage situation. So we create two other queries which count the number of people in each blood group by the SQL code below:

SQL code for storage system for donor:

SELECT Blood_bag.BloodGroup, Blood_bag.BloodBagID, Count(Donorf.bloodbagid) AS bloodbagidcount
FROM Blood_bag INNER JOIN Donorf ON Blood_bag.BloodBagID = Donorf.bloodbagid

GROUP BY Blood_bag.BloodGroup, Blood_bag.BloodBagID;

BloodBagID	BloodGroup	bloodbagidc
502	A-	10
501	A+	9
504	AB-	4
503	AB+	7
508	B-	3
507	B+	4
506	O-	3
505	O+	6

SQL code for storage system for recipient:

SELECT Count(Recipient.BloodGroup) AS CountOfBloodGroup, Blood_bag.BloodBagID, Recipient.BloodGroup
FROM Recipient INNER JOIN Blood_bag ON Recipient.BloodBagID = Blood_bag.BloodBagID
GROUP BY Blood_bag.BloodBagID, Recipient.BloodGroup;

BloodBagID	BloodGroup	CountOfBloc
501	A+	3
502	A-	1
503	AB+	1
504	AB-	2
505	O+	2
506	O-	1
507	B+	1
508	B-	1

In order to find the total number of blood, we need to subtract two queries by building one more queries, using the builder function.

Builder function:

Total: StorageDbloodtype![bloodbagidcount]-StorageRbloodtype!CountOfBloodGroup

SQL code:

SELECT StorageDbloodtype.BloodGroup, StorageDbloodtype.BloodBagID, StorageDbloodtype.bloodbagidcount, StorageRbloodtype.CountOfBloodGroup, StorageRbloodtype.BloodBagID, StorageRbloodtype.BloodGroup, StorageDbloodtype![bloodbagidcount]-StorageRbloodtype!CountOfBloodGroup AS Total
FROM StorageDbloodtype INNER JOIN StorageRbloodtype ON StorageDbloodtype.BloodBagID = StorageRbloodtype.BloodBagID;

Result:

Total_Number_Storage

StorageDblo	StorageDblo	bloodbagidc	CountOfBloc	StorageRblo	StorageRblo	Total
A-	502	10	1	502	A-	9
A+	501	9	3	501	A+	6
AB-	504	4	2	504	AB-	2
AB+	503	7	1	503	AB+	6
B-	508	3	1	508	B-	2
B+	507	4	1	507	B+	3
O-	506	3	1	506	O-	2
O+	505	6	2	505	O+	4

Another aim of our project is to answer the questions 3,4. Therefore, we created query each from the recipient and the donor and categorized them into age group and count the number of donor and recipient in the following age group

The SQL code is listed below:

```
SELECT [%$##@_Alias].AgeGroup, Count(*) AS CountOfAgeGroup
FROM (SELECT Switch([Age] Between 16 And 26, "16-26", [Age] Between 27 And 36, "27-36", [Age] Between 37 And 46, "37-46", [Age] > 46, "Over 46") AS AgeGroup FROM Donorf)  AS [%$##@_Alias]
GROUP BY [%$##@_Alias].AgeGroup;
```

We divided the blood group as 16-26, 27-36, 37-46 and older than 46. And the result is in the figure below.

Query: Agegroup_donor

AgeGroup	CountOfAge
16-26	12
27-36	9
37-46	6
Over 46	19

Similarly, we want to find out which age blood received the most amount of blood. The query AgeGroup_recipient is to count which blood group received the most blood.

SQL code:

```
SELECT [%$##@_Alias].AgeGroup, Count(*) AS CountOfAgeGroup
FROM (SELECT Switch([Age] Between 16 And 26, "16-26", [Age] Between 27 And 36, "27-36", [Age] Between 37 And 46, "37-46", [Age] > 46, "Over 46") AS AgeGroup FROM Recipient)  AS [%$##@_Alias]
GROUP BY [%$##@_Alias].AgeGroup;
```

We define the same age group as the donor.

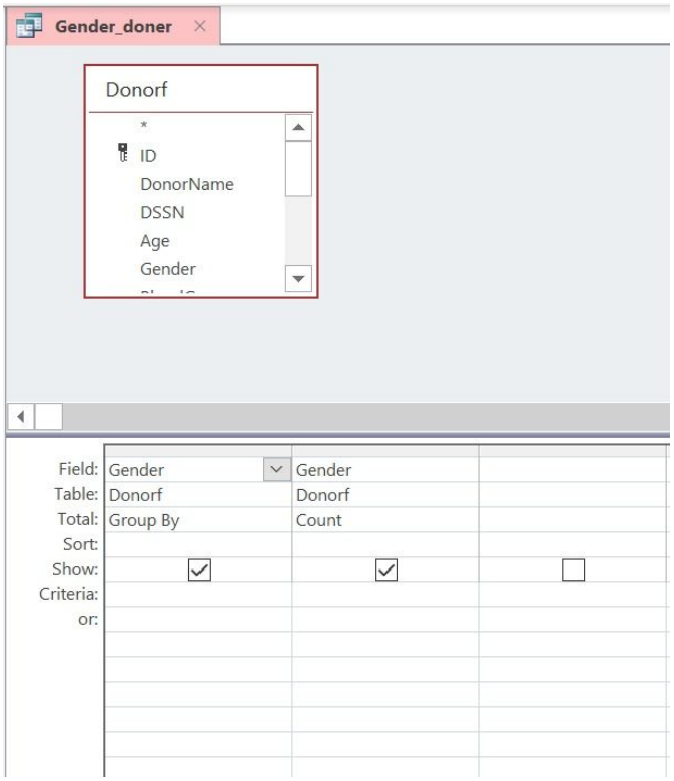
Query: AgeGroup_recipient

AgeGroup	CountOfAgeGroup
16-26	2
27-36	1
37-46	4
Over 46	5

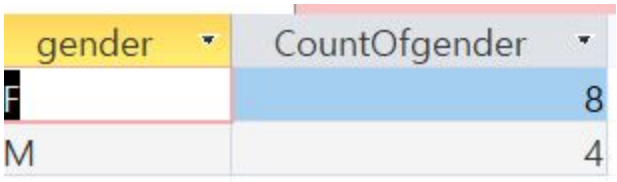
Furthermore, the final aim was to also answer question 5,6 which was related to gender group. This was done similarly like the age group. We created two queries for recipient and donor. Thereby, we used ‘summary function’ to count the number of ‘F’ which stands for female, and ‘M’ which stand for male, from ‘donor’ and ‘recipient’ tables. Through these we created a ‘Gender’ column and a ’Count of gender’ column.

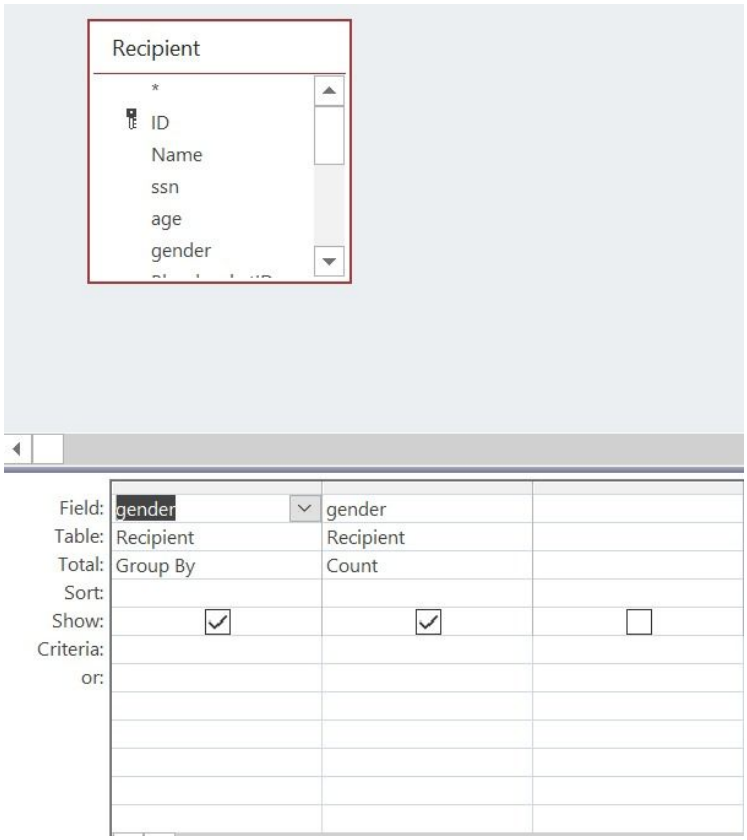
Query: Gender_donor

Gender	CountOfGe
F	20
M	26



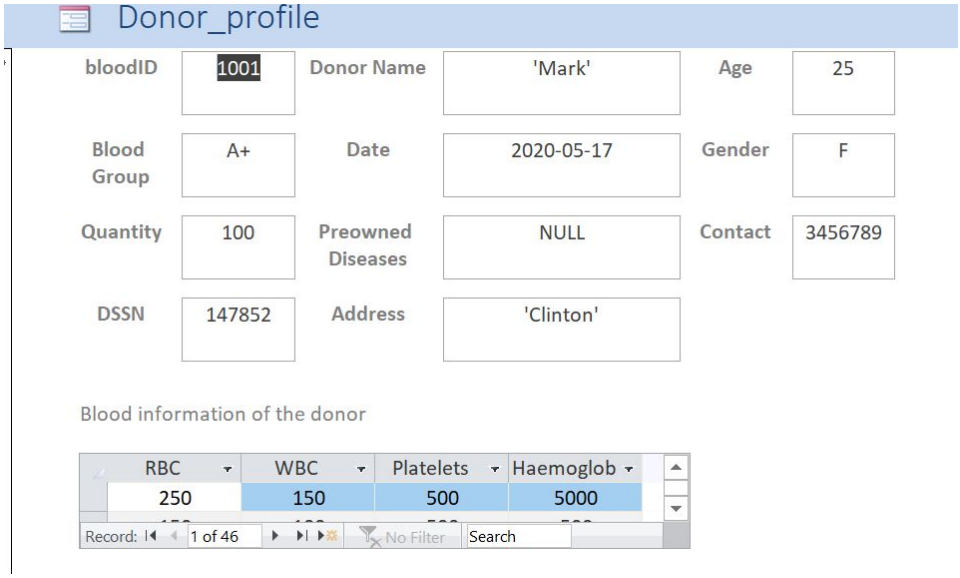
Query: Gender_recipient





To facilitate data entry and browsing the database, 2 forms are created and linked together. In the donor from, users can browse the details of each donor and the related blood information of each host.

Form:Donor_profile



Analysis & Discussion

From the data above, we can see that more people in the age ranges 16-26 and 46+ are willing to donate blood. This may be due to the fact that people in the age range 26-46 are busier with their careers and may not have the spare time to donate blood. On the contrary, there seems to be a positive correlation between age and need for blood. This can be explained by the fact that younger people tend to be healthier, while older people may need blood after medical operations due to health concerns.

5 reports are created to provide the information of the Blood bank storage, the donation statistics by age group, the reception statistics by age group and the donor and recipient statistics by gender. Each of them is provided together with the bar chart to demonstrate the data.

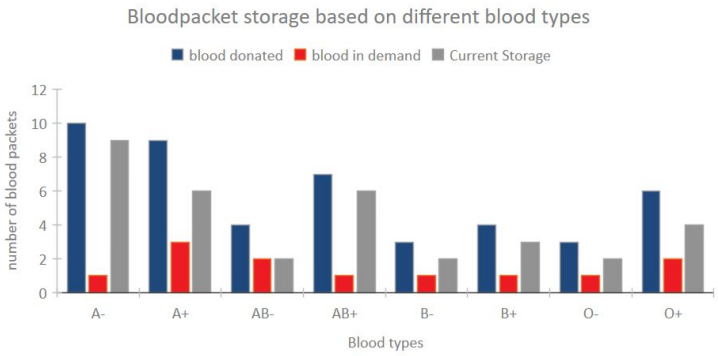
Report: Blood_Bank_Storage

By using the query Total_Number_Storage, a report is created. The report presents the amount of blood we currently have in each blood group. A screenshot of the report is shown below.

Blood Bank Storage

Thursday, 18 June 2020
8:58:42 pm

Bloodtype	Numer of Donors	Number of Recepient	blood_bagid	Total
A-	10	1	502	9
A+	9	3	501	6
AB-	4	2	504	2
AB+	7	1	503	6
B-	3	1	508	2
B+	4	1	507	3
O-	3	1	506	2
O+	6	2	505	4



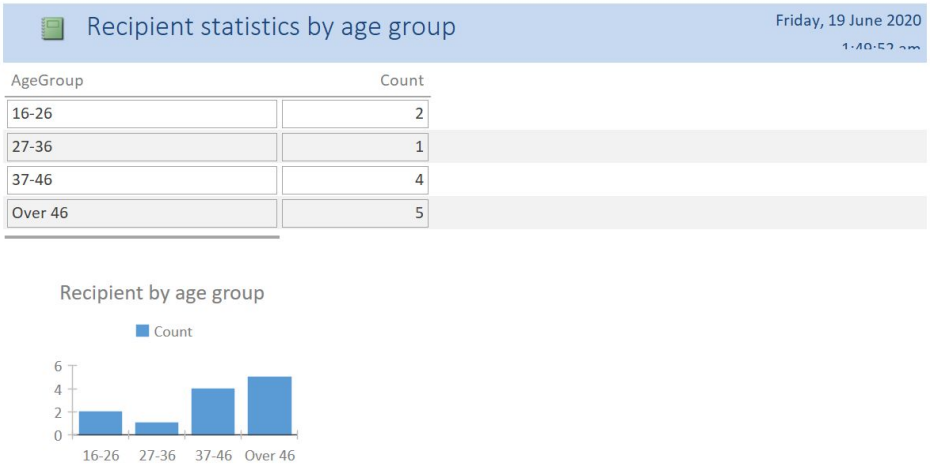
Interestingly, AB+ is the third most commonly donated blood type-subtype combination, even though it is rarer than most of the other blood types, and only other AB+ blood types can receive AB+ blood.

We also see that A+ is the blood type with the highest demand, with 3 out of 12 recipients receiving A+ type blood.

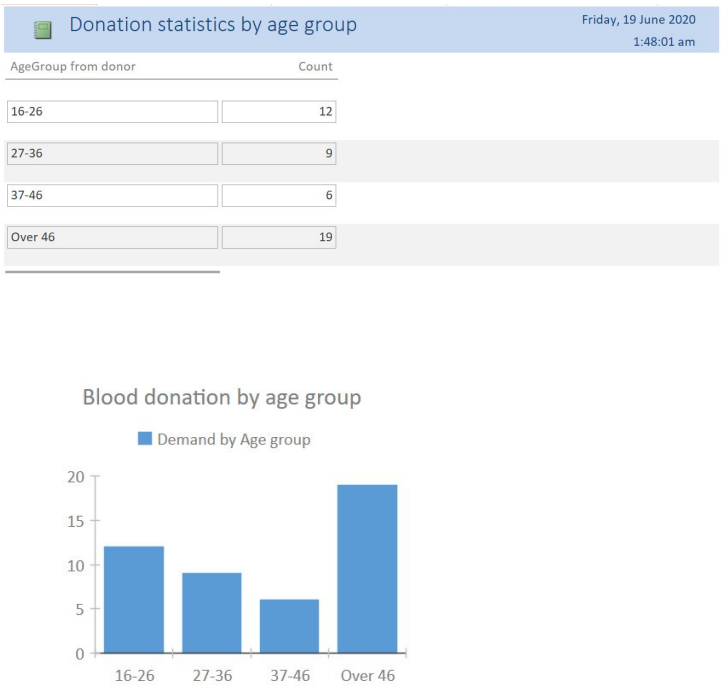
There seems to be no clear correlation between donor and blood subtype (+, -).

Report: Recipient_statistics_by_agegroup

By using the query agegroup_Recipient, a report is created. This report shows which age group received the most amount of blood. A screenshot of the report is shown below.



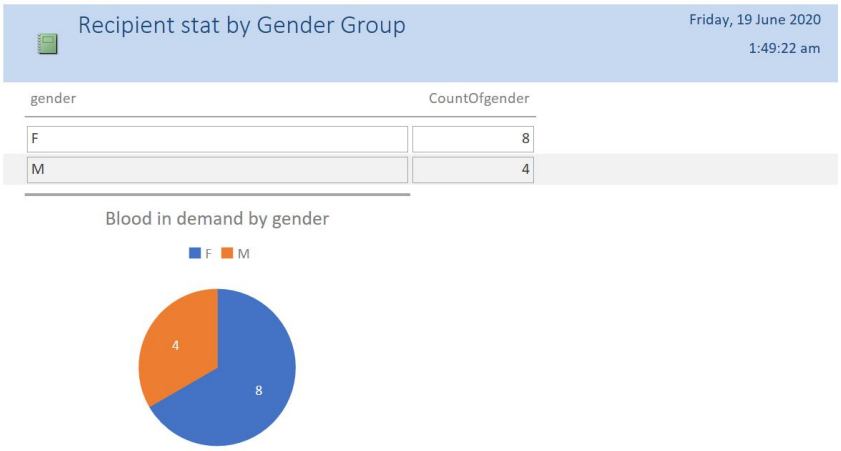
Report: Donor_statistics_by_agegroup



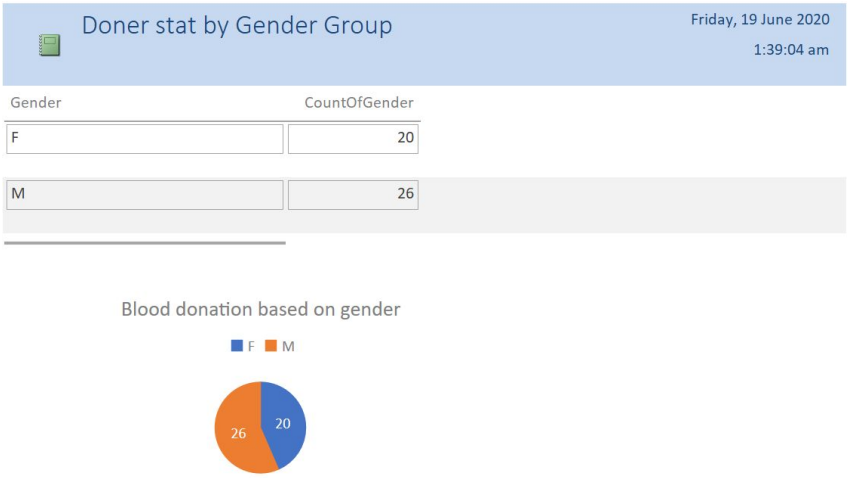
From the donors report, it shows that the age group 16-26 and over 46 donated the most blood. But surprisingly not the younger people donated more than the older people. This shows that there is no relationship between the age and the frequency they donated. According to the research from the National Center for Biotechnology Information, it shows that the age group

40-49 and 30-39 donate blood more frequently than other age groups (Beth. S, Adelbert. J, Krista. H, 2011). It matches with our result. The reason behind is, these two age group of people have a healthy body and they have more free time for their schedule. For the Recipient report, it shows that the age group 37-46 and over 46 receive blood more frequently than the other age group. It shows that the frequency of blood use is related to age because the older the people are, the higher chance they get injuries as their body system starts to get weaker, leading to the higher chance they need the blood.

Report: Recipient stat by gender



Report: Donor_statistics_by_agegroup



By the result from the graph, it shows that females actually received much more blood than male. This may show that females may need more blood than male because of pregnancy or some female diseases.

Conclusion

Our project seeks to organize blood donations in a database effectively. From the data, we have managed to discover certain trends in the data, such as frequent blood types and common age ranges of donors/recipients. We also utilized SQL operations to calculate the remaining blood bags for each blood type (number of donor blood bags - number of recipient blood bags). SQL is an ideal programming language to carry out less complicated operations such as searching for an age or finding the difference between two columns because it does not require complicated code.

With the data, we have confirmed some health-related trends. For example, elder people may have more health related concerns and may need more blood. There are also some anomalies. The distribution of blood types donated differs from the distribution of blood types received. In some cases, the distribution of blood types in the data differs from the distribution of blood types in the general population.

Due to some time and knowledge limitations, we could not create a more detailed analysis. Also, some of the data we used for the dataset is simulated. Further investigation could be done regarding the trends mentioned in this report.

References

Blood Types. (2020). Retrieved 19 June 2020, from https://www.redcrossblood.org/donate-blood/blood-types.html

vishakh2494/DataBase. (2020). Retrieved 19 June 2020, from https://github.com/vishakh2494/DataBase

Demographic Patterns of Blood Donors and Donations in a Large Metropolitan Area. (2011). Retrieved 19 June 2020, from https://pubmed.ncbi.nlm.nih.gov/21805814/

Gender disparities in red blood cell transfusion in elective surgery: a post hoc multicentre cohort study. (2016). Retrieved 18 June 2020, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2957494/

Gender differences in giving blood: a review of the literature. (2010). Retrieved 19 June 2020, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5168603/

Computer Code and file:

Notice:

All the tables, forms, queries, and reports in this project are created and code by ourselves.

No codes or tables are from others project or other reference sources.

File submitted:blood_bank.accdb (MS Access database created by MS office 2017), the database include:

- *4 tables*
- *7 queries*
- *3 forms*
- *5 reports*

Subject Index