

AUDIT ANALYSIS ASSIGNMENT

DHRUVIN RAKESH SHAH
(NUID: 001087062)

TABLE OF CONTENT

PART 1

AUDIT STATUS

IN USA or OUSA

GXP AREA

AUDIT TYPE

AUDIT METHOD

PROPOSED QUARTER

CONCLUSION

PART 2

DAYS INTAKE QSENT

DAYS QSENT QRECEIVED

DAYS_ONSITESCHEDULED_AUDITSTARTDATE

DAYS_STARTDATE_ENDDATE

DAYS_AUDITEND_FINALREPORTDUE

DAYS_FINALREPORTDUE_COMPLETIONDUE

PART 3

PART I

a) Audit Status Variable

Business Question:

Comparison between Computerized System Quality Assurance audits (CSQA Audits Sheet) and Good Working Practice audits (GxP Audits Sheet) 2017.

Analysis:

Frequency and Percentage Distribution Table

> CSQA_AuditStatus_pie

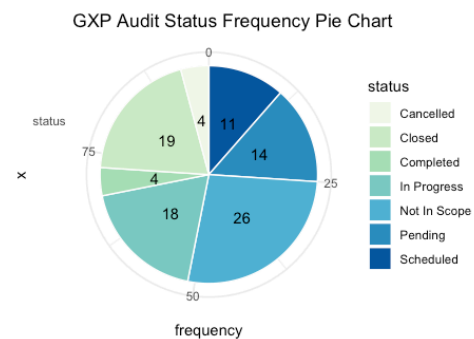
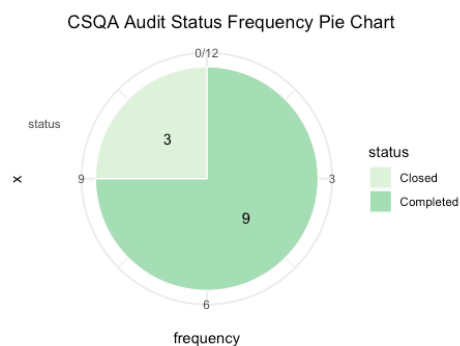
	status	frequency	percent
1	Closed	3	25
2	Completed	9	75

> GXP_auditstatus_pie

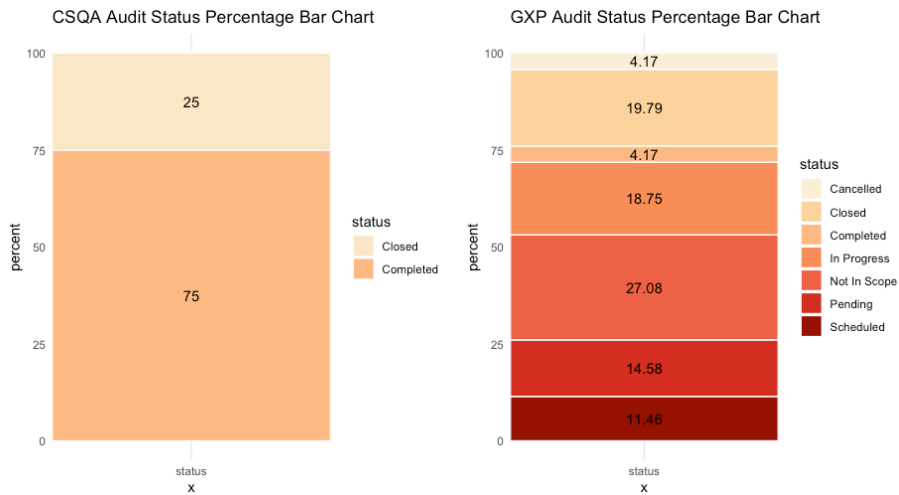
	status	frequency	percent
1	Cancelled	4	4.17
2	Closed	19	19.79
3	Completed	4	4.17
4	In Progress	18	18.75
5	Not In Scope	26	27.08
6	Pending	14	14.58
7	Scheduled	11	11.46

Graphical Representation:

Frequency:



Percentage:



b) In USA or OUSA Variable

Business Question:

Comparison between Computerized System Quality Assurance audits (CSQA Audits Sheet) and Good Working Practice audits (GxP Audits Sheet) 2017.

Analysis:

Frequency and Percentage Distribution Table

> CSQA_USAstatus_pie

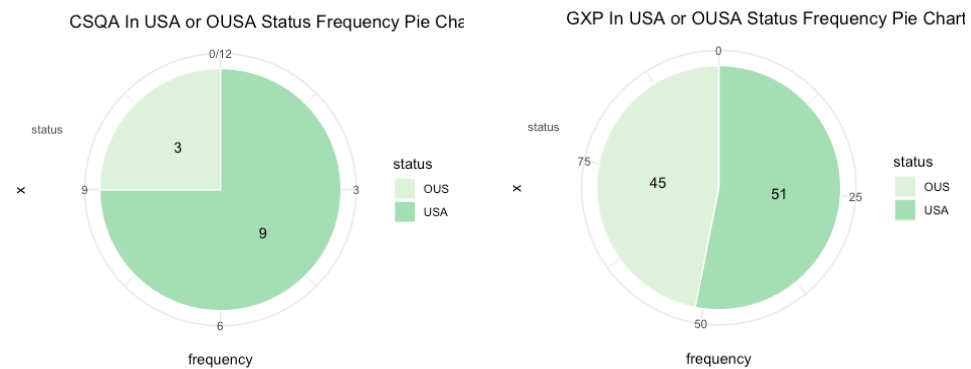
	status	frequency	percent
1	OUS	3	25
2	USA	9	75

> GXP_USAstatus_pie

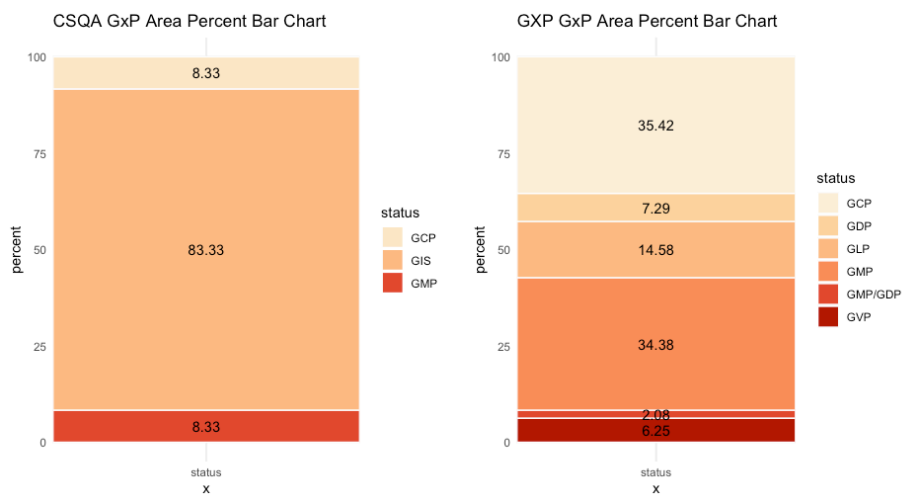
	status	frequency	percent
1	OUS	45	46.88
2	USA	51	53.12

Graphical Representation:

Frequency:



Percentage:



c) GxP Area Variable

Business Question:

Comparison between Computerized System Quality Assurance audits (CSQA Audits Sheet) and Good Working Practice audits (GxP Audits Sheet) 2017.

Analysis:

Frequency and Percentage Distribution Table

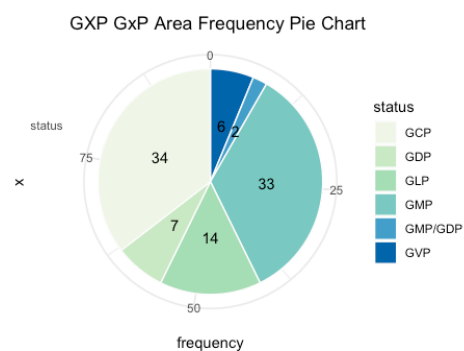
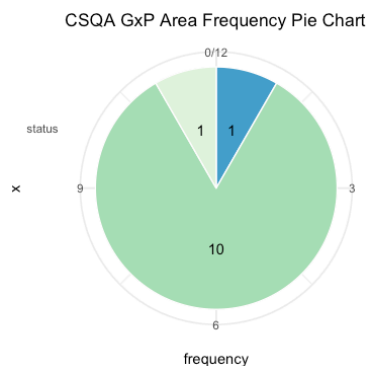
> CSQA_GxParea_pie

	status	frequency	percent
1	GCP	1	8.33
2	GIS	10	83.33
3	GMP	1	8.33

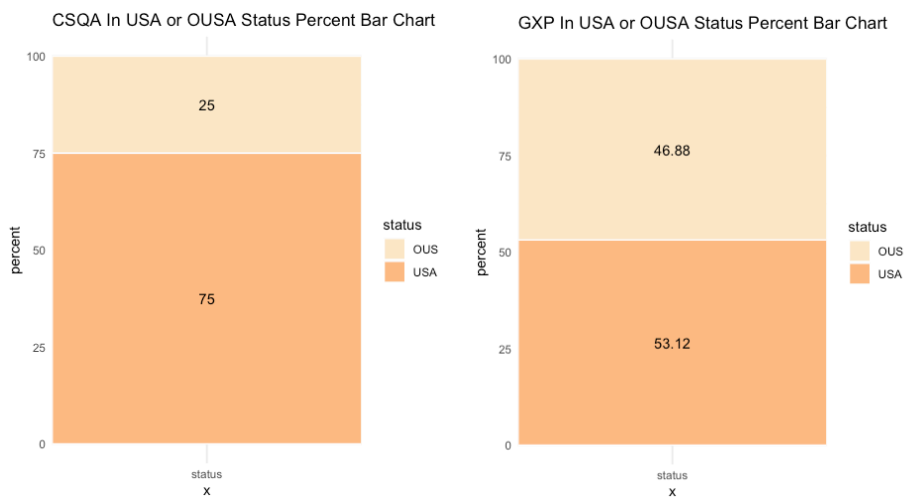
> GXP_GxParea_pie

	status	frequency	percent
1	GCP	34	35.42
2	GDP	7	7.29
3	GLP	14	14.58
4	GMP	33	34.38
5	GMP/GDP	2	2.08
6	GVP	6	6.25

Graphical Representation: Frequency



Percentage



d) Audit Type Variable

Business Question:

Comparison between Computerized System Quality Assurance audits (CSQA Audits Sheet) and Good Working Practice audits (GxP Audits Sheet) 2017.

Analysis:

Frequency and Percentage Distribution Table

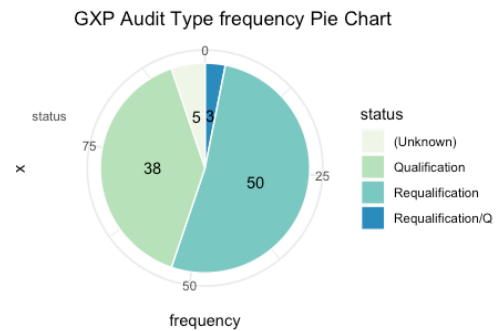
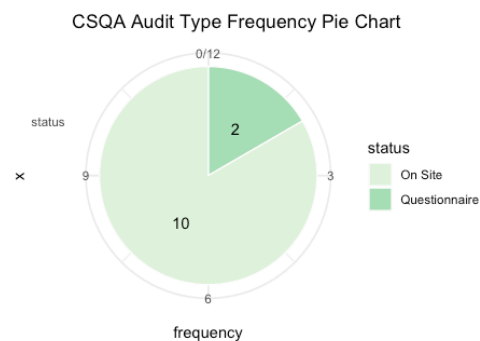
> CSQA_AuditType_pie

	status	frequency	percent
1	On Site	10	83.33
2	Questionnaire	2	16.67

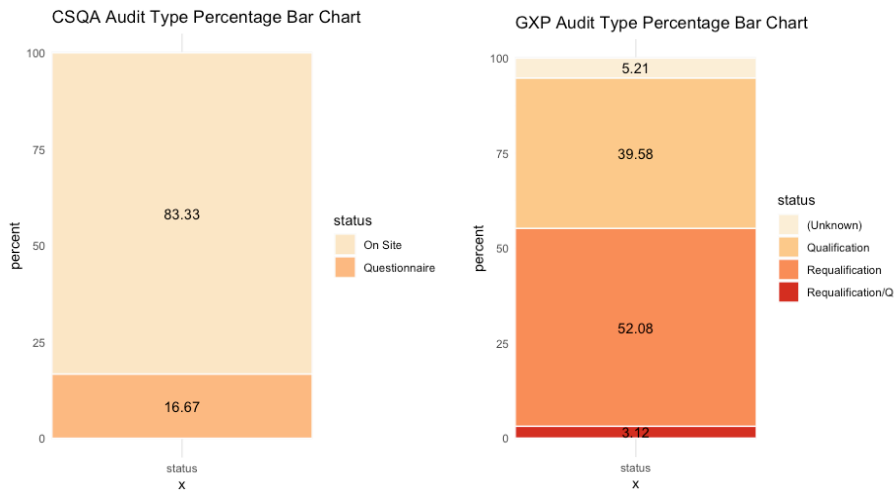
> GXP_AuditType_pie

	status	frequency	percent
1	(Unknown)	5	5.21
2	Qualification	38	39.58
3	Requalification	50	52.08
4	Requalification/Q	3	3.12

Graphical Representation: Frequency



Percentage



e) Audit Method Variable

Business Question:

Comparison between Computerized System Quality Assurance audits (CSQA Audits Sheet) and Good Working Practice audits (GxP Audits Sheet) 2017.

Analysis:

Frequency and Percentage Distribution Table

> CSQA_AuditMethod_pie

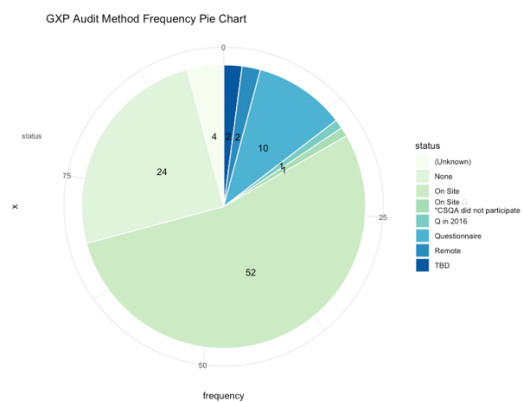
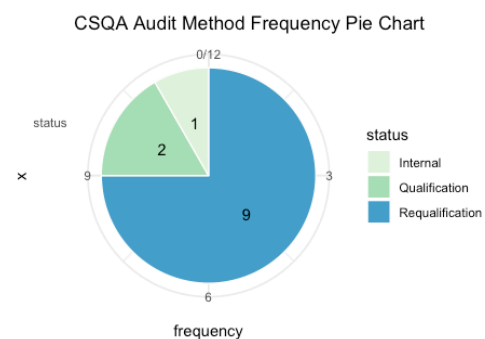
	status	frequency	percent
1	Internal	1	8.33
2	Qualification	2	16.67
3	Requalification	9	75.00

> GXP_AuditMethod_pie

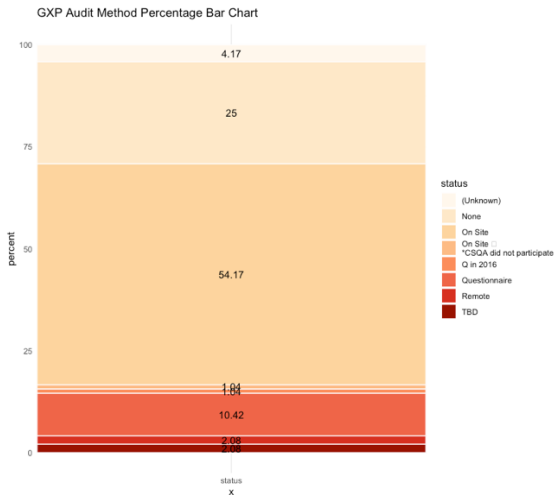
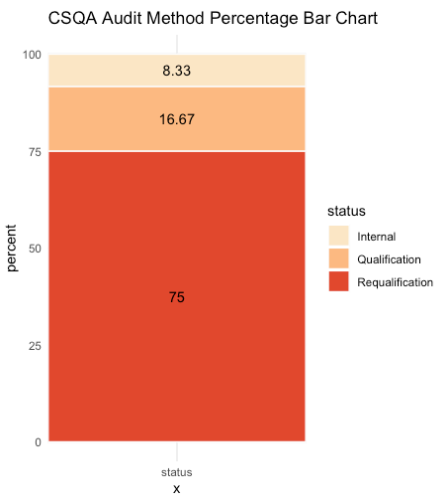
	status	frequency	percent
1	(Unknown)	4	4.17
2	None	24	25.00
3	On Site	52	54.17
4	On Site \r\n*CSQA did not participate	1	1.04
5	Q in 2016	1	1.04
6	Questionnaire	10	10.42
7	Remote	2	2.08
8	TBD	2	2.08

Graphical Representation:

Frequency



Percentage



f) Proposal Quarter Variable

Business Question:

Comparison between Computerized System Quality Assurance audits (CSQA Audits Sheet) and Good Working Practice audits (GxP Audits Sheet) 2017.

Analysis:

Frequency and Percentage Distribution Table

> CSQA_proposemethod_pie

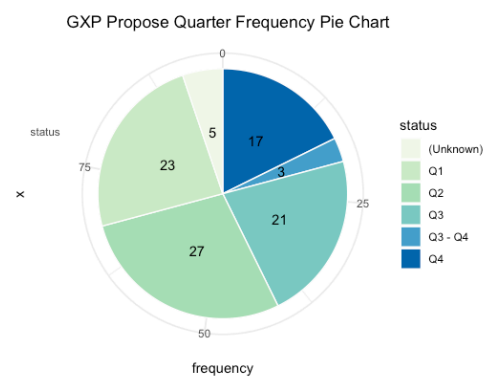
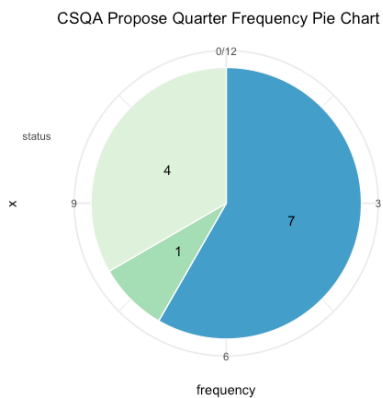
	status	frequency	percent
1	Q1	4	33.33
2	Q2	1	8.33
3	Q3	7	58.33

> GXP_proposemethod_pie

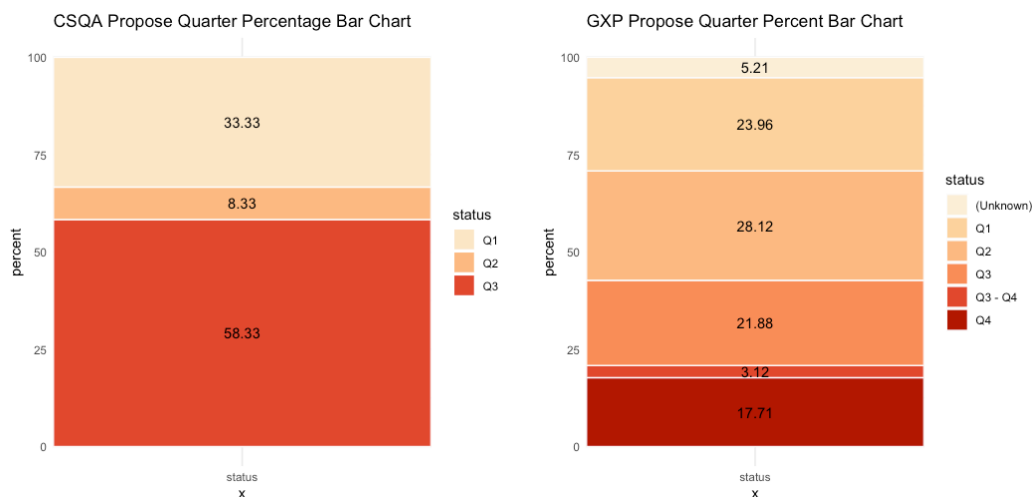
	status	frequency	percent
1	(Unknown)	5	5.21
2	Q1	23	23.96
3	Q2	27	28.12
4	Q3	21	21.88
5	Q3 - Q4	3	3.12
6	Q4	17	17.71

Graphical Representation:

Frequency



Percentage



Conclusion Part I:

From the above analysis we can observe that Computerized System Quality Assurance audits (2017 CSQA Audits) have less variables in the desired column of analysis than Good Working Practice audits (2017 GxP Audits). Moreover, the data in “2017 GxP Audits” is more lucid than “2017 CSQA Audits”. The entries in the “2017 GxP Audits” are greater than “2017 CSQA Audits”, hence it is easy to predict the data which has more tuples. The Data in “2017 GxP Audits” is perfectly categorized than the other sheet.

PART II

- 1) “Date of Intake” and “Date Q Sent”. Name that variable
“Days_Intake_QSent”

Business Question:

Comparison between Computerized System Quality Assurance audits (CSQA Audits Sheet) and Good Working Practice audits (GxP Audits Sheet) 2017.

Analysis:

Days Lapsed:

```
> (CSQA_Days_Intake_QSent <- ymd(CSQA_Date_Qsent) - ymd(CSQA_Date_Intake))
Time differences in days
[1] NA NA NA NA NA NA NA NA 74 NA 27 NA
> (GXP_Days_Intake_QSent <- ymd(GXP_Date_Qsent) - ymd(GXP_Date_Intake))
Time differences in days
[1] NA NA NA NA NA 6 6 NA NA NA NA 1 1 2 1 1 1 NA NA NA NA NA NA 71 NA NA
[27] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
[53] NA NA 88 NA NA NA NA NA NA NA NA NA NA NA NA 13 NA NA NA NA NA NA NA NA
[79] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
```

Mean:

```
> (mean(CSQA_Days_Intake_QSent,na.rm = TRUE))
Time difference of 50.5 days
> (mean(GXP_Days_Intake_QSent,na.rm = TRUE))
Time difference of 17.36364 days
```

Median:

```
> (median(CSQA_Days_Intake_QSent,na.rm = TRUE))
Time difference of 50.5 days
> (median(GXP_Days_Intake_QSent,na.rm = TRUE))
Time difference of 2 days
```

- 2) “Date Q Sent” and “Date Q Received”. Name that variable “Days_QSent_QReceived. Based on the name of the variables, what do you think that variable means? Does it apply to all audits? Why?

Business Question:

Comparison between Computerized System Quality Assurance audits (CSQA Audits Sheet) and Good Working Practice audits (GxP Audits Sheet) 2017.

Analysis:

Days Lapsed:

```
> (CSQA_Days_QSent_QReceived <- ymd(CSQA_Date_Qrecieved) - ymd(CSQA_Date_Qsent))
Time differences in days
[1] NA NA NA NA NA NA NA NA 14 NA 22 NA
> (GXP_Days_QSent_QReceived <- ymd(GXP_Date_Qrecieved) - ymd(GXP_Date_Qsent))
Time differences in days
[1] NA NA NA NA NA 13 14 NA NA NA NA 56 34 32 35 41 41 NA NA NA NA NA NA NA NA
[27] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
[53] NA NA 11 NA NA NA NA NA NA NA NA NA NA NA 15 NA NA NA NA NA NA NA NA NA
[79] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
```

Mean:

```
> (mean(CSQA_Days_QSent_QReceived,na.rm = TRUE))
Time difference of 18 days
> (mean(GXP_Days_QSent_QReceived,na.rm = TRUE))
Time difference of 29.2 days
```

Median:

```
> (median(CSQA_Days_QSent_QReceived,na.rm = TRUE))
Time difference of 18 days
> (median(GXP_Days_QSent_QReceived,na.rm = TRUE))
Time difference of 33 days
```


- 3) “Date On Site Scheduled” and “Audit Start Date”. Name that variable “Days_OnSiteScheduled_AuditStartDate. Does this variable apply to all audits? Why?

Business Question:

Comparison between Computerized System Quality Assurance audits (CSQA Audits Sheet) and Good Working Practice audits (GxP Audits Sheet) 2017.

Analysis:

Days Lapsed:

```
> (CSQA_Days_OnSiteScheduled_AuditStartDate <- ymd(CSQA_Date_AuditStartDate) - ymd(CSQA_Date_OnSiteScheduled))
Time differences in days
[1] 30 35 13 55 63 55 60 48 NA 48 NA NA
> (GXP_Days_OnSiteScheduled_AuditStartDate <- ymd(GXP_Date_AuditStartDate) - ymd(GXP_Date_OnSiteScheduled))
Time differences in days
[1] NA NA NA 37 NA NA NA 38 39 13 NA NA NA NA NA NA NA 14 NA 6
[21] 39 NA 10 47 53 NA NA NA NA NA NA NA 28 102 61 28 22 56 146 91
[41] NA 67 80 83 85 NA NA 40 NA 47 NA 54 NA NA 54 NA NA NA 82 152
[61] 151 153 60 33 33 39 17 20 0 NA NA NA NA NA NA 22 NA 39 20 NA
[81] NA 17 NA NA NA 55 NA 67 NA NA NA NA NA NA 84 68
```

Mean:

```
> (mean(CSQA_Days_OnSiteScheduled_AuditStartDate,na.rm = TRUE))
Time difference of 45.22222 days
> (mean(GXP_Days_OnSiteScheduled_AuditStartDate,na.rm = TRUE))
Time difference of 54.48889 days
```

Median:

```
> (median(CSQA_Days_OnSiteScheduled_AuditStartDate,na.rm = TRUE))
Time difference of 48 days
> (median(GXP_Days_OnSiteScheduled_AuditStartDate,na.rm = TRUE))
Time difference of 47 days
```

- 4) “Audit Start Date and “Audit End Date”. Name that variable “Days_StartDate_EndDate”.

Business Question:

Comparison between Computerized System Quality Assurance audits (CSQA Audits Sheet) and Good Working Practice audits (GxP Audits Sheet) 2017.

Analysis:

Days Lapsed:

```
> (CSQA_Days_StartDate_EndDate <- ymd(CSQA_Date_AuditEndDate) - ymd(CSQA_Date_AuditStartDate))
Time differences in days
[1] 1 1 1 1 1 1 1 0 NA 0 NA NA
> (GXP_Days_StartDate_EndDate <- ymd(GXP_Date_AuditEndDate) - ymd(GXP_Date_AuditStartDate))
Time differences in days
[1] NA 2 NA 1 1 NA NA 3 2 2 NA NA NA NA NA NA 3 NA 0 1 1 1 1 NA NA NA NA NA NA 1 1 1
[36] 0 1 1 2 1 NA 0 1 1 0 NA NA 1 NA 1 NA 1 NA NA 0 NA NA NA 2 1 0 2 1 2 1 1 1 0 0 NA
[71] NA NA NA NA 2 1 NA 1 1 NA NA 1 NA NA NA 1 NA 1 NA NA NA NA 0 1 1
```

Mean:

```
> (mean(CSQA_Days_StartDate_EndDate,na.rm = TRUE))
Time difference of 0.7777778 days
> (mean(GXP_Days_StartDate_EndDate,na.rm = TRUE))
Time difference of 1.06 days
```

Median:

```
> (median(CSQA_Days_StartDate_EndDate,na.rm = TRUE))
Time difference of 1 days
> (median(GXP_Days_StartDate_EndDate,na.rm = TRUE))
Time difference of 1 days
```

5) “Audit End Date” and “Date Final Report Due”. Name that variable “Days_AuditEnd_FinalReportDue”

Business Question:

Comparison between Computerized System Quality Assurance audits (CSQA Audits Sheet) and Good Working Practice audits (GxP Audits Sheet) 2017.

Analysis:

Days Lapsed:

```
> (CSQA_Days_AuditEnd_FinalReportDue <- ymd(CSQA_Date_FinalReportDue) - ymd(CSQA_Date_AuditEndDate))
Time differences in days
[1] 30 30 30 30 30 30 30 30 30 30 NA NA
> (GXP_Days_AuditEnd_FinalReportDue <- ymd(GXP_Date_FinalReportDue) - ymd(GXP_Date_AuditEndDate))
Time differences in days
[1] NA 27 NA 30 30 NA NA 30 30 30 NA NA NA NA NA NA NA 30 NA 30 30 30 30 30 NA NA NA NA NA NA NA 30 30 30
[36] 30 30 30 30 30 NA 30 30 30 30 NA NA 30 NA 30 NA 30 NA NA NA NA NA 30 30 30 30 30 30 30 30 30 NA
[71] NA NA NA 30 30 NA 30 30 NA 30 NA 30 NA NA 30 NA NA 30 NA NA NA 30 30 30
```

Mean:

```
> (mean(CSQA_Days_AuditEnd_FinalReportDue,na.rm = TRUE))
Time difference of 30 days
> (mean(GXP_Days_AuditEnd_FinalReportDue,na.rm = TRUE))
Time difference of 29.94 days
```

Median:

```
> (median(CSQA_Days_AuditEnd_FinalReportDue,na.rm = TRUE))
Time difference of 30 days
> (median(GXP_Days_AuditEnd_FinalReportDue,na.rm = TRUE))
Time difference of 30 days
```

- 6) “Date Final Report Due” and “Date of Completion”. Name that variable “Days_FinalReportDue_CompletionDate”

Business Question:

Comparison between Computerized System Quality Assurance audits (CSQA Audits Sheet) and Good Working Practice audits (GxP Audits Sheet) 2017.

Analysis:

Days Lapsed:

```
> (CSQA_Days_FinalReportDue_CompletionDate <- ymd(CSQA_Date_CompletionDate) - ymd(CSQA_Date_FinalReportDue))
Time differences in days
[1] 9 -1 0 40 5 0 13 -1 11 0 0 NA
> (GXP_Days_FinalReportDue_CompletionDate <- ymd(GXP_Date_CompletionDate) - ymd(GXP_Date_FinalReportDue))
Time differences in days
[1] NA 1 NA -1 0 0 0 -1 -8 -1 NA -12 3 5 9 -15 20 NA NA NA NA 0 0 NA NA NA
[27] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
[53] NA NA 0 NA NA NA NA NA NA NA NA NA -1 -1 -7 NA 0 NA NA NA NA NA 4 -2 NA 0
[79] 3 NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
```

Mean:

```
> (mean(CSQA_Days_FinalReportDue_CompletionDate, na.rm = TRUE))
Time difference of 6.909091 days
> (mean(GXP_Days_FinalReportDue_CompletionDate, na.rm = TRUE))
Time difference of -0.16 days
```

Median:

```
> (median(CSQA_Days_FinalReportDue_CompletionDate, na.rm = TRUE))
Time difference of 0 days
> (median(GXP_Days_FinalReportDue_CompletionDate, na.rm = TRUE))
Time difference of 0 days
```

PART III

CONCLUSION:

No, I do not recommend merging the audit sheets “2017 GxP Audits” and “2017 CSQA Audits”. In the above analyzed data, it is witnessed that both audit sheets have some common variable attributes in columns, but it is even observed that “2017 CSQA Audits” does not contain some variables which are present in “2017 GxP Audits” and vice versa. So if we merge “Good Working Practice” as well as “Computerized System Quality Assurance” then there will be issues with the frequencies of common variable columns or there is a possibility that in the given data there are redundant entries which are calculated in both the audits, and if sheets are merged then the entries will get calculated twice hence there can be an error in the calculation which will mess up the cleaned data. So it is advisable to analyzed them individually.