Assessment Brief

Student Name/ID Number	Tran Duy Hung/GCS210483
Unit Number and Title	**14: Business Intelligence**
Academic Year	2023-2024
Unit Tutor	Sam Nguyen
Assignment Title	**Assignment 2: Apply BI tools & techniques and their impact**
Issue Date	
Submission Date	
IV Name & Date	

Submission Format
Part I: Project submission. This should be a zip / rar folder of your project, including all necessary files to run your project. There should be a link to your Tableau work on Tableau Public cloud. Part II: The submission is in the form of a group written report. This should be written in a concise, formal business style using single spacing and font size 12. You are required to make use of headings, paragraphs and subsections as appropriate, and all work must be supported with research and referenced using the Harvard referencing system. Please also provide a bibliography using the Harvard referencing system. Part III: Team needs to present their point of view about how business intelligence tools can contribute to effective decision-making as well as the legal issues involved in exploiting user data for business intelligence. You may need to research for specific examples of organizations that use BI tools to enhance or improve their business and evaluate how they can use BI tools for extend their target audience and make them more competitive within the market.

Unit Learning Outcomes
LO3 Demonstrate the use of business intelligence tools and technologies
Assignment Brief
(Continued from previous scenario) Your next task is to demonstrate to the board of directors about the ability of applying business intelligence in the company's current business processes. To demonstrate BI, you need to prepare a presentation about BI and related tools & techniques and a demonstration on real company dataset. For the presentation, you need: - Explain general concept of what is BI - Introduction to some tools / techniques for BI and their application in general For the demonstration, you need: - A (some) data set(s) extracted from the company's business processes. Explain the dataset. - Show how you pre-process data for later analysis, explain each step and it purpose - Design dashboards to show your analysis on pre-processed data. Explain clearly purpose of dashboards and charts. Suggestions should be made after analysis During the demonstration, you need collect feed-back and comments from users to review how well your dashboards design meet user or business requirement and what customization needed for future use. Team needs to present their point of view about how business intelligence tools can contribute to effective decision-making as well as the legal issues involved in exploiting user data for business intelligence. You may need to research for specific examples of organizations that use BI tools to enhance or improve their business and evaluate how they can use BI tools for extend their target audience and make them more competitive within the market. To summary, you need to submit a report in PDF includes 4 parts: your presentation, result of demonstration and review of user feedback, point of view on BI contribution and legal issues.

<table>
<tr><th colspan="3">Learning Outcomes and Assessment Criteria</th></tr>
<tr><th>Pass</th><th>Merit</th><th>Distinction</th></tr>
<tr><td colspan="2">LO3 Demonstrate the use of business intelligence tools and technologies</td><td rowspan="2">D3 Provide a critical review of the design in terms of how it meets a specific user or business requirement and identify what customisation has been integrated into the design.</td></tr>
<tr><td>P3 Determine, with examples, what business intelligence is and the tools and techniques associated with it.

P4 Design a business intelligence tool, application or interface that can perform a specific task to support problem-solving or decision-making at an advanced level.</td><td>M3 Customise the design to ensure that it is user friendly and has a functional interface.</td></tr>
<tr><td colspan="2">LO4 Discuss the impact of business intelligence tools and technologies for effective decision-making purposes and the legal/regulatory context in which they are used</td><td rowspan="2">D4 Evaluate how organisations could use business intelligence to extend their target audience and make them more competitive within the market, taking security legislation into consideration</td></tr>
<tr><td>P5 Discuss how business intelligence tools can contribute to effective decision-making.

P6 Explore the legal issues involved in the secure exploitation of business intelligence tools</td><td>M4 Conduct research to identify specific examples of organisations that have used business intelligence tools to enhance or improve operations.</td></tr>
</table>

Table Of Content

Topic: Predicting the earthquakes, magnitude of the earthquakes and depth of the earthquakes

1. Introduction.

Business intelligence (BI) is all about turning an organization's data into insights that can be used to inform business decisions. BI analysts will use BI tools, software or services to access and analyze datasets and translate their findings into reports, summaries, dashboards, graphs, charts or maps. In recent years, the advent of modern data visualization and reporting tools has transformed the discipline, empowering businesses to use big data insights to identify, develop and create new business opportunities. (Negash, 2004)

1.1. Overview of problems

Predicting earthquakes is an ongoing and important area of research and monitoring, as it can help to improve public safety and reduce the potential damage and loss of life caused by earthquakes. However, it is important to note down every variable from each earthquake to get a much more precise prediction. The crucial variable of predicting is using trusted data and accounting all the natural variables. (Su et al., 2020)

1.2. Motivations.

Organizations that leverage business intelligence have better control and visibility over their processes. It helps them identify and rectify errors or inefficiencies in existing processes. It also allows them to predict unforeseen challenges and act accordingly.

1.3. Objective.

Predicting earthquakes is a complex and challenging task that involves the analysis of various geological, seismological, and other relevant data. The objective of earthquake prediction is to provide advanced warning to people in areas that are likely to be affected by an earthquake, allowing them to take appropriate measures to minimize damage and protect themselves from harm.(Gokhberg et al., 1982)

Scenario#1: Linear regression predicting the depth and level of magnitude of earthquakes.

Using the data from the scenario to create a scatter diagram to show a relationship between two variables of the earthquakes. Combined with a simple linear regression line used to fit a model between the two variables. predict the value of the target variable based on given data.

Scenario#2: Pie charts show the people injured and total of casualties

From the data of the scenario, it's used to create a create a pie chart to show the injured and cassettes of natural causes. Each slice represents the percentage of accidents in each state.

Scenario#3: Bar chart demonstrate the level of magnitude of the earthquakes and numbers of cases

Using the data to create the bar chart to demonstrate the level and number of earthquakes. The highest bar shows the most common type of earthquakes from 5 to 5.8 is the most common type. The highest level was 9.1 magnitude, there were fewer cases in the past year. From the chart we can clearly see the most common type of earthquakes and rarest level of magnitude.

Scenario#4: Bubble chart shows the level of magnitude and numbers of cases

Using the data to create the bubble chart to demonstrate the level and number of earthquakes. The bigger the bubble the more frequent type of earthquake, this shows the numeric relationship between each variable.

Scenario#5: Map with circle and color separation show the location and level of magnitude of the earthquakes

Using latitude and longitude data to determine the location of the earthquakes. Each of the circles on the map show the level and depth of the earthquakes. Level ranging from 5.1 to 9.

Scenario#6: Choropleth maps show the people injured because of natural causes

Using Choropleth map to demonstrate the number of natural accidents that happened in the US. The brighter color, the greater number of accidents occurred in that state. But there were some problems with the data imported, there were some states missing, the data wasn't reported, or the application could not identify the state.

Scenario#7: Dashboard showing data of earthquakes location, ID and Magnitude of the earthquakes

Using dashboard to show many sheets of data from different sources. Dashboards are useful for monitoring, measuring, and analyzing relevant data. Using it to see information of earthquakes like ID and type of magnitude. Many sheets showing their unique data in different ways like maps, tables and bubble charts. All the charts must have a relation to each other's.

1.4. Summary

The overall scenario has shown the data used to display all the chart and map above is precise in displaying the sales and when predicting the value of a variable based on the value of another variable. The coordinate when used in map was geometrically accurate when displayed on the map.

2. Related works and Dataset.

2.1. Related works

Business Intelligence Challenges for Independent Game Publishing

Recent observations in the USSR appear to suggest that wide-band electromagnetic radiation occurs just prior to earthquakes. To apply this phenomenon to earthquake prediction, measurements have been carried out at the Sugadaira Space Radiowave Observatory in Japan under the USSR-Japan Cooperation Program in 1980. The recorded noise level at 81 kHz is comparatively quiet throughout the day and night. However, about one-half hour before the main shock of a magnitude 7 earthquake at 0733 UT on March 31, 1980, the instrument recorded an anomalous amplitude increase to 15 dB higher than the normal level. VLF data recorded synoptically at Sugadaira suggests that unusual impulsive radiation at frequencies below 1.5 kHz also occurred shortly before the earthquake. Similar 81-kHz emissions were observed prior to magnitude 5 and 6 earthquakes on September 25, 1980, and January 28, 1981. (Foshay and Kuziemsky, 2014)

2.2. Dataset

A data set is a structured collection of data points related to a particular subject. A collection of related data sets is called a database. Data sets can be tabular or non-tabular. Tabular data sets contain structured data that is organized by rows and columns. Non-tabular data sets contain unstructured data contained by brackets. (Techopedia, 2011)

Data sets can also be categorized by the type of information they contain. Popular types of data sets include:

- Numerical - data is expressed in numbers rather than natural language.
- Bivariate - contains two types of related data.
- Multivariate - contains three or more than three types of related data.
- Categorical - data variables can have one of two values.
- Correlation - values in the data set have a relationship with each other.

2.3. Scenarios Dataset

Using dataset from the scenario we can filter and categorize by data type. Data needs to be sorted out before using. The table below shows the data that will be used to create a map to locate all the earthquakes and level magnitude of the earthquakes and this dataset has 23413 rows and 7 columns. There are some data that is not accounted for when measuring these are the most crucial data for making charts and maps.

database.csv Latitude	database.csv Longitude	database.csv Type	database.csv Depth	database.csv Depth Error	database.csv Depth Seismic Stations	database.csv Magnitude
19.246	145.616	Earthquake	131.600	*null*	*null*	6.00000
1.863	127.352	Earthquake	80.000	*null*	*null*	5.80000
-20.579	-173.972	Earthquake	20.000	*null*	*null*	6.20000
-59.076	-23.557	Earthquake	15.000	*null*	*null*	5.80000
11.938	126.427	Earthquake	15.000	*null*	*null*	5.80000
-13.405	166.629	Earthquake	35.000	*null*	*null*	6.70000
27.357	87.867	Earthquake	20.000	*null*	*null*	5.90000
-13.309	166.212	Earthquake	35.000	*null*	*null*	6.00000
-56.452	-27.043	Earthquake	95.000	*null*	*null*	6.00000
-24.563	178.487	Earthquake	565.000	*null*	*null*	5.80000

Table 1: Earthquakes dataset

3. Proposed model.

3.1. Correlation

A correlation coefficient is a number between -1 and 1 that tells you the strength and direction of a relationship between variables. In other words, it reflects how similar the measurements of two or more variables are across a dataset.

In addition to analyzing the relationship between two variables graphically, we can also measure the strength of the linear relationship between two variables using a measure called the correlation coefficient. The correlation coefficient of two variables can be estimated from sample data using Equation 1 or the algebraic equivalent, Equation 2.

Two-Variable Relationships

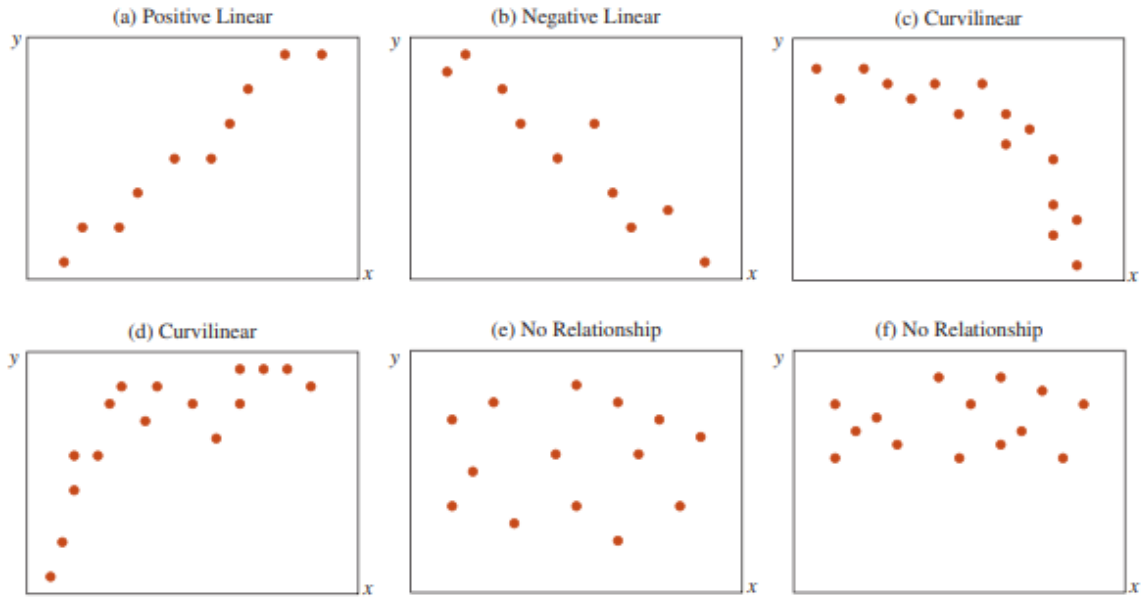

Figure 1: Example of Two-Variable Relationships

$$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{[\Sigma(x - \bar{x})^2][\Sigma(y - \bar{y})^2]}} \quad (1)$$

Correlation between Two Variables

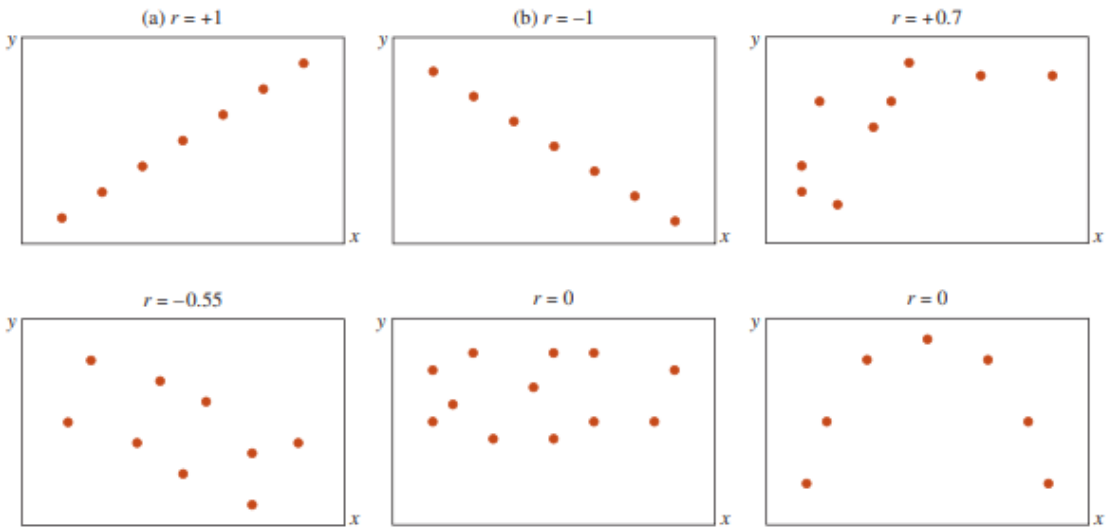

Figure 2: Example of Correlation between Two Variables

$$r = \frac{n\Sigma xy - \Sigma x\Sigma y}{\sqrt{[n(\Sigma x^2) - (\Sigma x)^2][n(\Sigma y^2) - (\Sigma} \qquad (2)$$

The sample correlation coefficient computed using Equations 14.1 and 14.2 is called the Pearson product moment correlation. The sample correlation coefficient, r, can range from a perfect positive correlation, +1.0, to a perfect negative correlation, -1.0. A perfect correlation occurs if all points on the scatter plot fall on a straight line. If two variables have no linear relationship, the correlation between them is 0 and there is no linear relationship between the x and y variables. Consequently, the more the correlation differs from 0.0, the stronger the linear relationship between the two variables. The sign of the correlation coefficient indicates the direction of the relationship. Figure 14.2 illustrates some examples of correlation between two variables. Once again, for the correlation coefficient to equal plus or minus 1.0, all the (x, y) points must form a perfectly straight line. The more the points depart from a straight line, the weaker (closer to 0.0) the correlation is between the two variables.

Why use Correlation?

Not only can we measure this relationship, but we can also use one variable to predict the other. For example, if we know how much we're planning to increase our spending on advertising then we can use correlation to accurately predict what the increase in visitors to the website is likely to be.

3.2. Linear regression

Linear regression plays a fundamental role in statistical modeling. This article provides a step-by-step coverage of linear models in the order of model specification, model estimation, statistical inference, variable selection, model diagnosis, and prediction. Computation issues in linear regression and intimately relevant extensions of linear models are also discussed. (Su et al., 2012)

Coefficient b1, the regression slope coefficient of the population regression line, measures the average change in the value of the dependent variable, y, for each unit change in x. The regression slope can be positive, zero, or negative, depending on the relationship between x and y. For example, a positive population slope of 12 (β1 = 12) means that for a 1-unit increase in x, we can expect an average 12-unit increase in y. Correspondingly, if the population slope is negative 12 (β1 = -12), we can expect an average decrease of 12 units in y for a 1-unit increase in x. The population's y intercept, b0, indicates the mean value of y when x is 0. However, this interpretation holds only if the population could have x values equal to 0. When this cannot occur, b0 does not have a meaningful interpretation in the regression model.

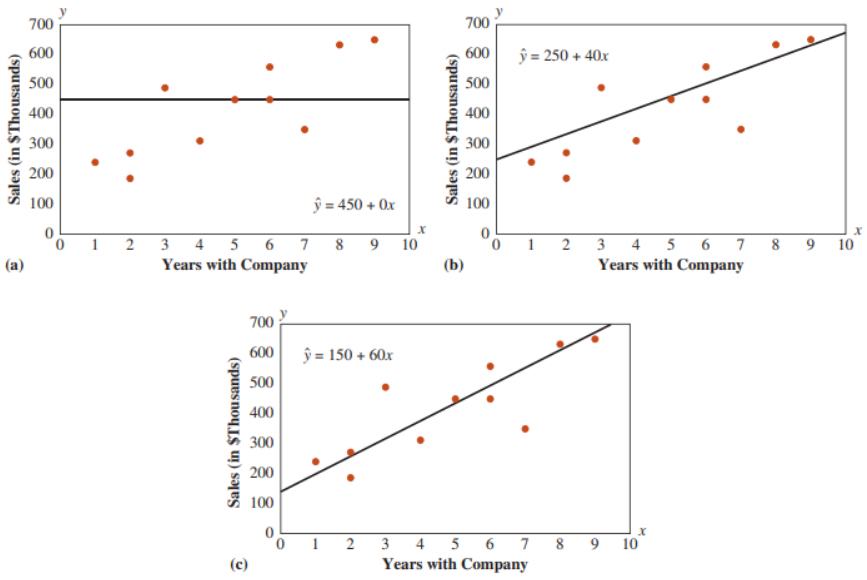

Figure 3: Example of linear regression

$$\hat{y} = b_0 + b_1 x \quad (3)$$

The regression line through the sample data is the best estimate of the population regression line. However, there are an infinite number of possible regression lines for a set of points. For example, Figure 14.7 shows three of the possible different lines that pass through the Midwest Distribution data. Which line should be used to estimate the true regression model? We must establish a criterion for selecting the best line. We use the least squares criterion. To understand the least squares criterion, you need to know about prediction error, or residual, which is the difference between the actual y coordinate of an (x, y) point and the predicted value of that y coordinate produced by the regression line. Figure 14.8 shows how the prediction error is calculated for the employee who was with Midwest for four years (x = 4) using one possible regression line (where yn is the predicted sales value). (Tosi, 2009)

Why use Linear regression?

More precisely, linear regression is used to determine the character and strength of the association between a dependent variable and a series of other independent variables. It helps create models to make predictions.

4. Simulating scenarios and Results.

4.1. Package installation.

Tableau

Figure 6: Tableau application

It helps users create different charts, graphs, maps, dashboards, and stories for visualizing and analyzing data, to help in making business decisions. Tableau has a lot of unique, exciting features that make it one of the most popular tools in business intelligence (BI).

4.2. Performance of scenarios.

Scenario#1: Linear regression predicting the depth and level of magnitude of earthquakes

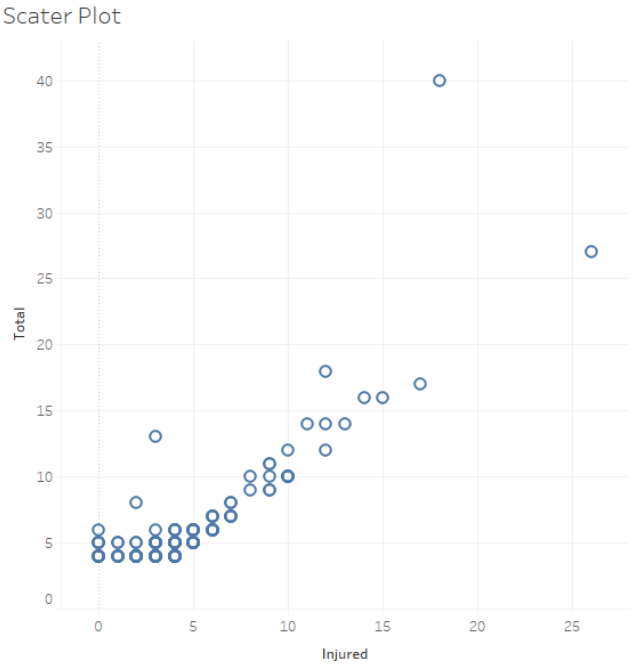

Figure 7: Scatter plot

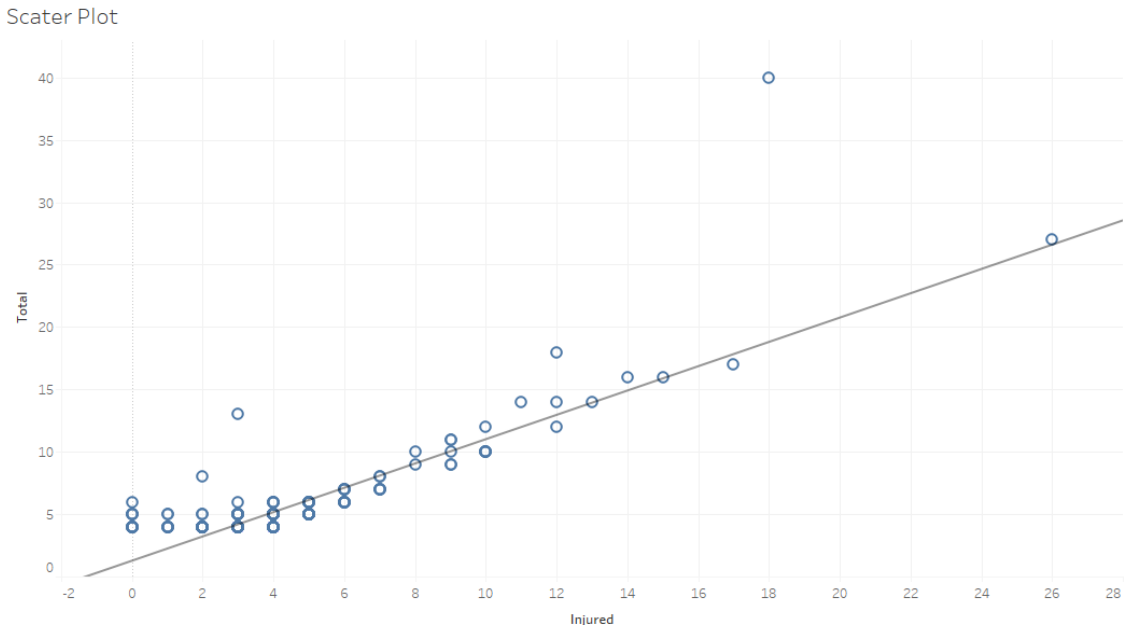

Figure 8: Linear regression

Scenario#2: Pie charts show the people injured and total of casualties

Figure 9: Pie charts

Scenario#3: Bar chart demonstrate the level of magnitude of the earthquakes and numbers of cases

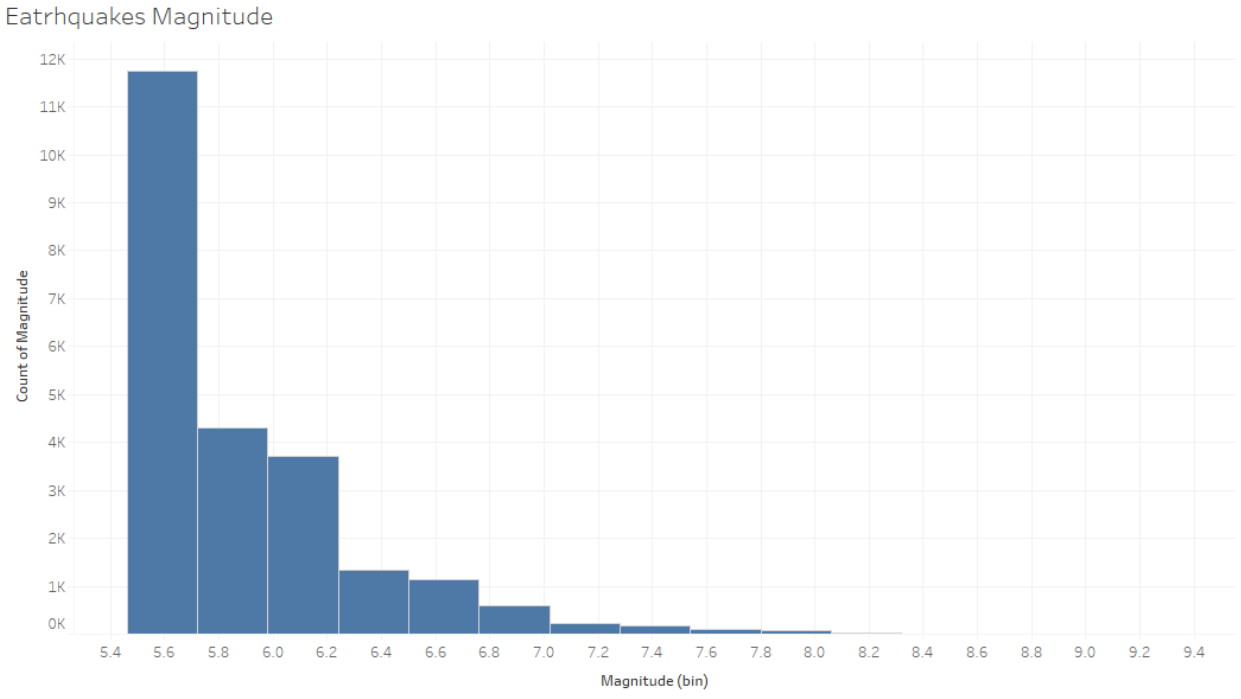

Figure 10: Bar chart

Scenario#4: Bubble chart shows the level of magnitude and numbers of cases

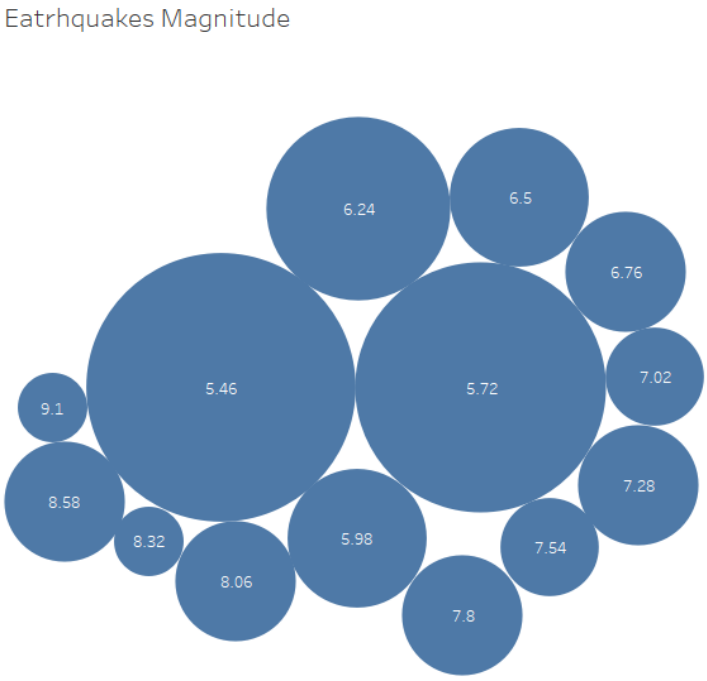

Figure 11: Bubble chart

Scenario#5: Map with circle and color separation show the location and level of magnitude of the earthquakes

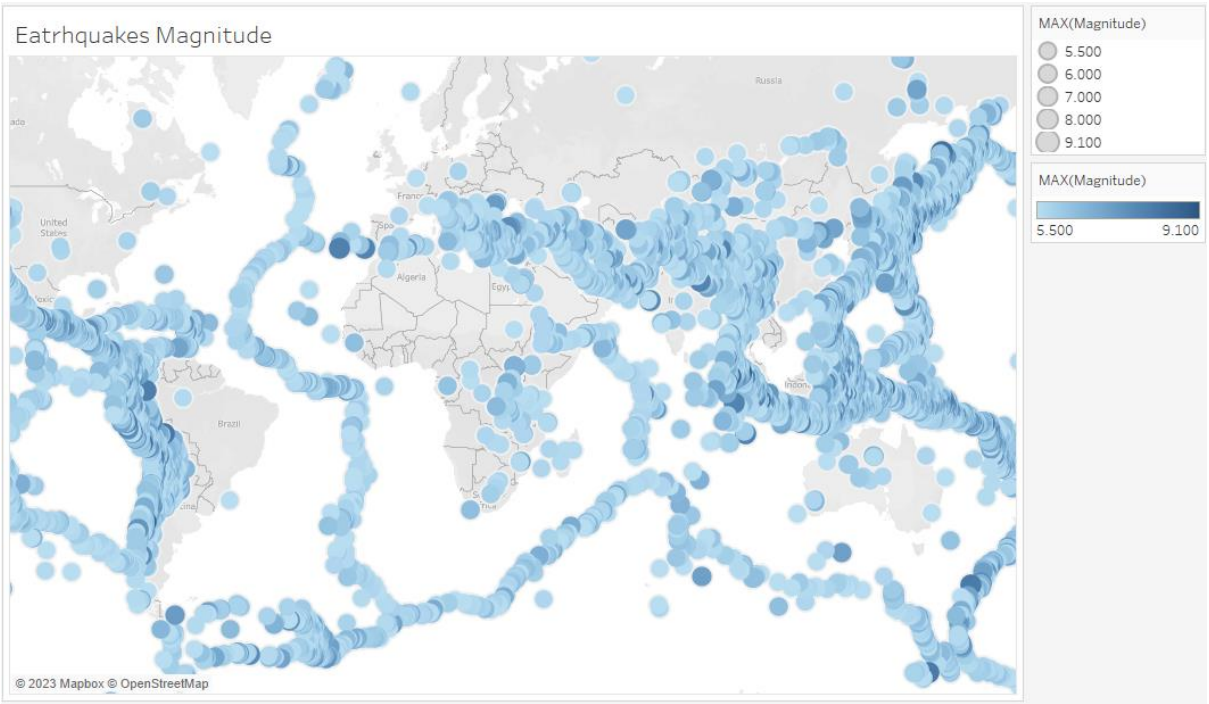

Figure 12: Marked map for earthquakes activities

Scenario#6: Choropleth maps show the people injured because of natural causes

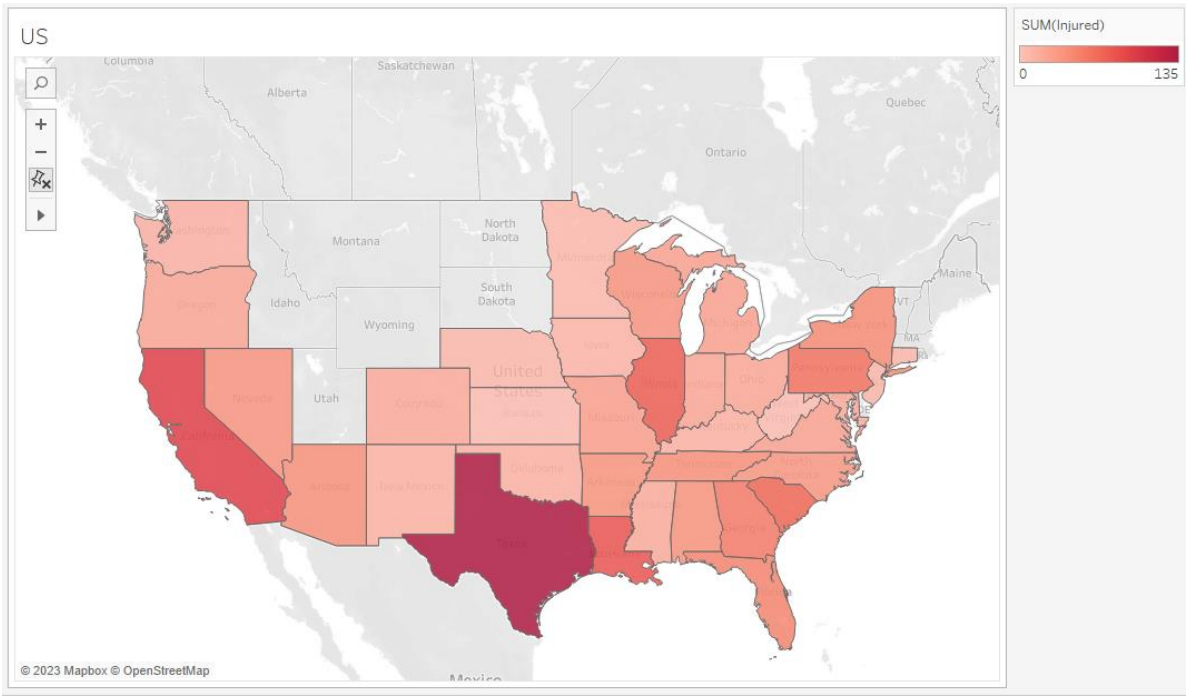

Figure 13: Choropleth map show accidents in US states

Scenario#7: Dashboard showing location, ID and Magnitude of the earthquakes

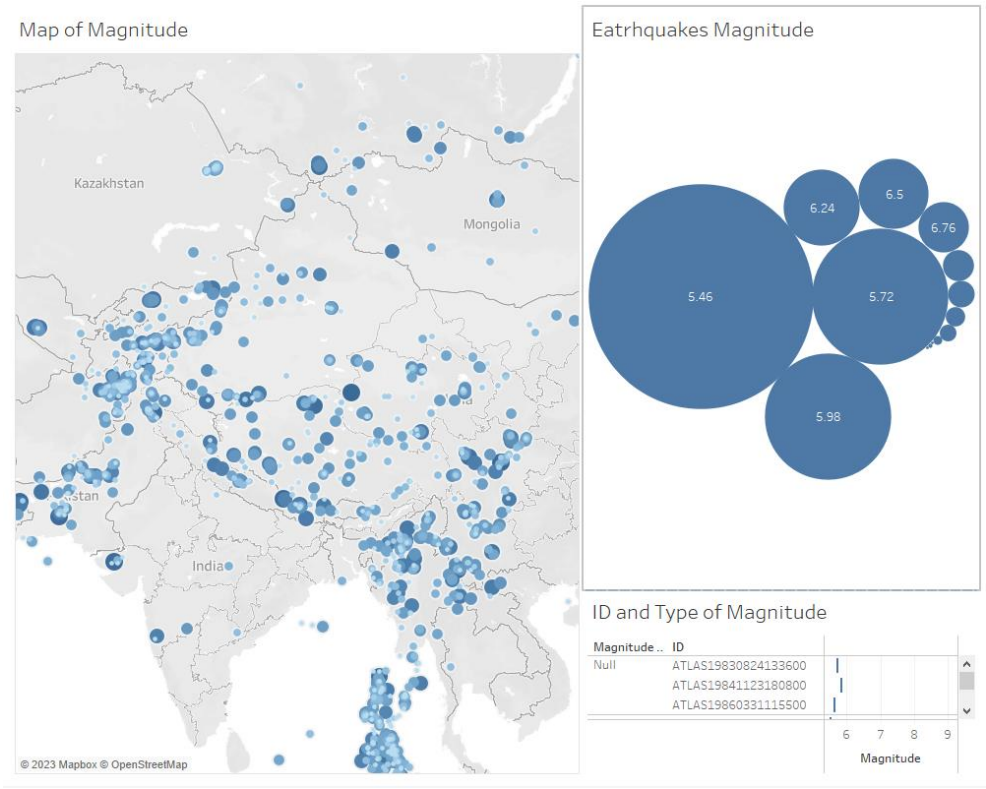

Figure 14: Overall of dashboard

Map have different size and color for each circle marker for each Magnitude of earthquakes when choose a circle on the map marker. All the information about that marker will show up. Each sheet in the dashboard helps support each other's like bubble chart in charge of the level of magnitude and number of earthquakes, the bigger the bubble the more frequent earthquake of that level of magnitude. Using Gantt view to show the ID, type of earthquakes and magnitude of that specific earthquakes.

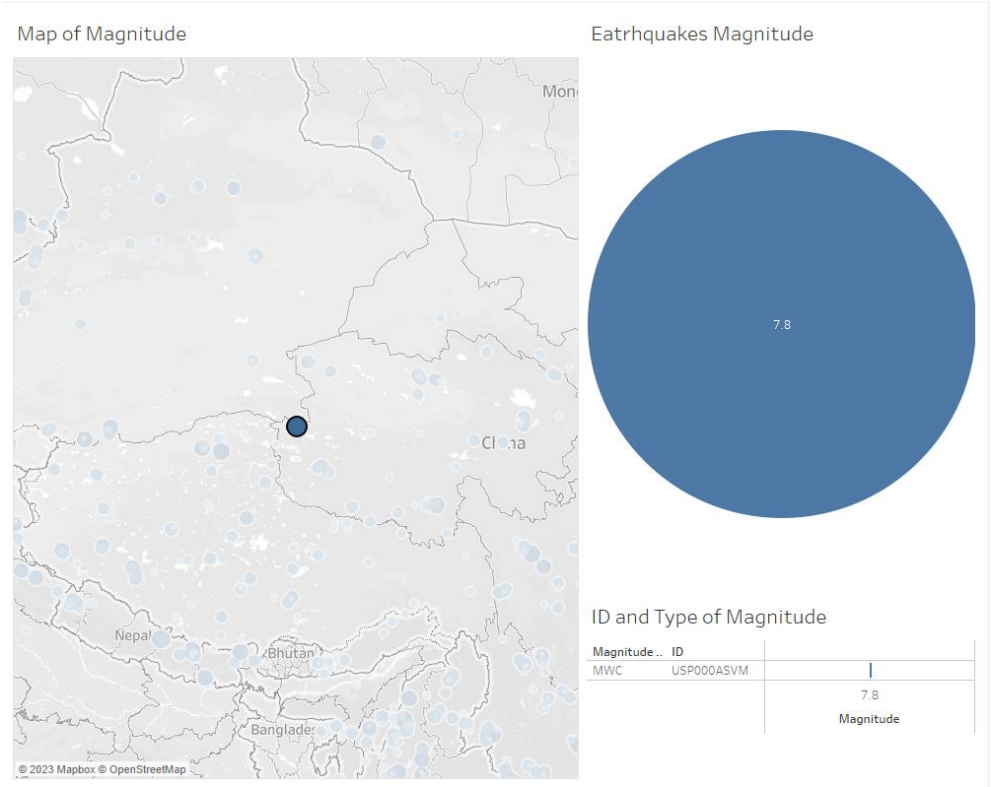

Figure 15: Focused data of dashboard

5. Conclusion of the scenario.

The performance of the 7 scenarios below worked as expected, using the Tableau application to create charts and maps. The chart's accuracy is determined by the correctness of the data it displays. The data must be collected from reliable sources and presented without distortion. Overall, the performance of charts can greatly impact the effectiveness in communicating information to chart intended audience. It is important to carefully consider the design and content of the chart to ensure that it is accurate, clear, simple, and relevant to its intended purpose and audience.

References

FOSHAY, N. & KUZIEMSKY, C. J. I. J. O. I. M. 2014. Towards an implementation framework for business intelligence in healthcare. 34**,** 20-27.

GOKHBERG, M., MORGOUNOV, V., YOSHINO, T. & TOMIZAWA, I. J. J. O. G. R. S. E. 1982. Experimental measurement of electromagnetic emissions possibly related to earthquakes in Japan. 87**,** 7824-7828.

NEGASH, S. J. C. O. T. A. F. I. S. 2004. Business intelligence. 13**,** 15.

SU, X., YAN, X. & TSAI, C. L. J. W. I. R. C. S. 2012. Linear regression. 4**,** 275-294.

SU, Y., BACKLUND, P. & ENGSTRÖM, H. J. I. J. O. C. G. T. 2020. Business intelligence challenges for independent game publishing. 2020.

TOSI, S. 2009. *Matplotlib for Python developers*, Packt Publishing Ltd.