Solution Student Handbook – Chapter 3

Objective Type Questions

Please choose the correct option in the questions below:

(1) Data can be visualized using:
(a) Graphs
(b) Maps
(c) Charts
(d) All of the above

(2) Which of the following statement is false?
(a) Data visualization can absorb information quickly.
(b) Data visualization decreases the insights and takes a slower decision.
(c) Data visualization is a type of visual art.
(d) None of the above.

(3) Which of the following is a use case of data visualization?
(a) Healthcare
(b) Sales and Marketing
(c) Politics/Campaigning
(d) All of the above

(4) Which format of data is easiest for analysis?
(a) Tabular data
(b) Text data in a PDF
(c) Data in an image
(d) Speech data

(5) Which visualization is best for representing a relation between two variables?
(a) Scatter plot
(b) Histogram
(c) Pie chart
(d) Gantt Chart

Standard Questions

Please answer the questions below in no less than 100 words:

(1) What are the steps to make sure that the correct data is collected for analysis?

The steps to make sure that the correct data is collected for analysis are as followed:

a. **Quality of data**: The data which we are going to collect should be complete and of top quality. If we make an analysis or try to build an unreliable database on the basis bad quality or incomplete data which is been collected, then definitely we are going to get the wrong result which will not be as per our expectation. First of all we must give a preference to the quality of the data as the top most quality that is to be collected for analysis.
b. **Completeness of data**: We should even make sure that whatever the data is been collected should be complete set. As if incomplete set of data will lead to occur of many wrong results.
c. **Format of data**: The data which is been collected for the analysis, its format should be good and right. The collected data should be readable and accessible for analysis, and if incase if the data is not in the right format then the data needs to be upgraded and changed into the right format which is required for analysis.

(2) Write short note on the statistical techniques which can be used for data analysis.

There are various techniques based on statistics which are used for analyzing of data. Mostly three statistical techniques are used for analyzing of data which are as followed:

a.   **Regression**: It is a technique of finding out the relation or co relations between the various variables in the data. With the help of the Regression analysis we can come to know about how the value of the dependent variable changes as per the change in the independent variable. Hence regression analysis makes it easier in identifying that how dependent variable is been effected by the change in the independent variable.
b.   **Cohort Analysis**: This analysis helps us to observe and make easily comparison on how the different customer's groups behaves the overtime.
c.   **Predictive Analysis**: This analysis involves the prediction of future consequences that will occur based on the historical datasets. This analysis can be made use for creating the alternative options and preventive measures against the risk assessments.

(3) Is it important to assess the end-users for a visualization? Explain in your own words.

Yes, it is more important to access the end-users for a visualization which will help them to decide how they can present the insights. If the insights are been presented in a very proper manner which will helps the decision makers to take the accurate decision on the basis of the gathered and analyzed data.

High Order Thinking Skills (HOTS)

Please answer the questions below in no less than 200 words.

(1) *What are the things to consider before deciding on an appropriate visualization for your data?*

The things to consider before deciding on an appropriate visualization for your data are as followed:

a.   If we want to get the correct result, then we should make sure that we had collected the correct and good quality data.
b.   If the collected data is not correct or its quality is not good then there won't be any impact of the created algorithm. By making use of the irrelevant data, it take us to the wrong path/conclusion and misguide us.
c.   You cannot correctly find out the pre-processed raw data which could not be used directly for your requirements. You need to understand that from where and how the data was been collected.

   i.   **Quality of data**: The data which we are going to collect should be complete and of top quality. If we make an analysis or try to build an unreliable database on the basis bad quality or incomplete data which is been collected, then definitely we are going to get the wrong result which will not be as per our expectation. First of all we must give a preference to the quality of the data as the top most quality that is to be collected for analysis.
   ii. **Completeness of data**: We should even make sure that whatever the data is been collected should be complete set. As if incomplete set of data will lead to occur of many wrong results.
   iii. **Format of data**: The data which is been collected for the analysis, its format should be good and right. The collected data should be readable and accessible for analysis, and if incase if the data is not in the right format then the data needs to be upgraded and changed into the right format which is required for analysis.

(2) If you find that the data collected has outliers, what steps can you take to ensure that your analysis is still accurate?

If you find that the data collected has outliers, following steps we are going to make use for ensuring that our analysis is still accurate:

a.   **Quality of data**: The data which we are going to collect should be complete and of top quality. If we make an analysis or try to build an unreliable database on the basis bad quality or incomplete data which is been collected, then definitely we are going to get the wrong result which will not be as per our expectation. First

of all we must give a preference to the quality of the data as the top most quality that is to be collected for analysis.

b. **Completeness of data**: We should even make sure that whatever the data is been collected should be complete set. As if incomplete set of data will lead to occur of many wrong results.

c. **Format of data**: The data which is been collected for the analysis, its format should be good and right. The collected data should be readable and accessible for analysis, and if incase if the data is not in the right format then the data needs to be upgraded and changed into the right format which is required for analysis.

Applied Project

Each student should write down the marks he/she had received in the examination for the subjects studied in the previous grade. Use these marks to plot on the paper.
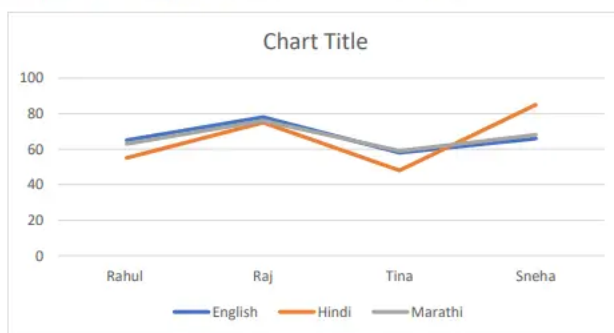
Data is related to the marks of each students received for the subjects studied in the previous grade.

Students Names	English	Hindi	Marathi
Rahul	65	55	63
Raj	78	75	76
Tina	58	48	59
Sneha	66	85	68

a. Bar graph to display marks of each individual subject.
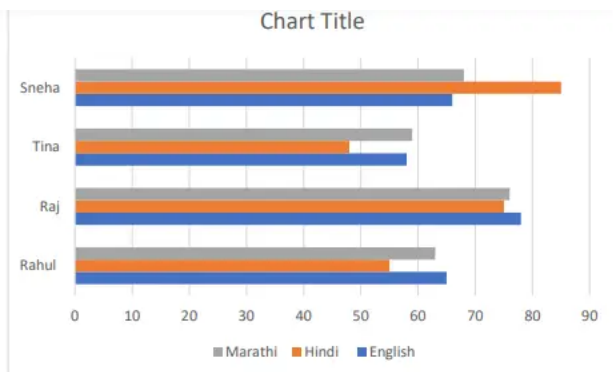
b) Line graph to display marks of each individual subject.

(c) Pie chart to show percentage contribution of marks of each subject in the total marks obtained.

English Marks

Rahul   Raj   Tina   Sneha

Hindi Marks

Rahul   Raj   Tina   Sneha

Marathi Marks

Rahul   Raj   Tina   Sneha