

Exercise Sheet 1

1. Name three different types of data to which Visual Analytics can be applied. For each of the examples, name a question that might be relevant to the evaluation of the data and justify your decision.

a. Hotel Bookings dataset of a country/region

We can try to identify which type of hotels are popular in a particular area.

b. Bank dataset

We can try to find transactions which results in fraud.

b. Gaming data

Try to find the player's position which results in maximum goals.

2. Using an example, explain why a purely statistical evaluation of data is not sufficient to derive hypotheses.

Anscombe-Quartett has 4 data set whose statistical characteristics are same but when plotted its total different.

https://en.wikipedia.org/wiki/Anscombe%27s_quartet

3. What are the 4 core objectives of Visual Analytics and explain them using an example.

Information fusion- gathering all the data available and preprocessing

Simulation and analysis- initial analysis on the graphs regarding correlation between number of booking and area or season or weather

Data mining – finding hidden patterns in booking data about user booking

Visualization and Interaction- providing the results in a nice dashboard to users where they can even drill down the information.

4. Explain using an example except weather maps where Visual Analytics makes little sense.

Any small dataset like for revenue of a quarter given in a table.

5. Explain the interaction of the four components of Visual Analytics system using an example.
 - a. Analytical reasoning
 - b. Visual representations and interaction techniques
 - c. Data representation and transformation
 - d. Production, presentation and dissemination

To get the maximum revenue for a hotel booking website, first component can involve the refining the question that what the user really wants out of the data and this process. Where all managers can focus to maximize the revenue and hence ask the right questions.

On the raw data, preprocessing and cleansing the data so that it can be used to answer the questions is data representation and transformation part. Which can also be used later when we have the analysis ready and the output needs to be shown appropriately.

Once the data is available and after the analysis the questions has been asked the visual presentation of the available data and the answers should be plotted in graphs. Choosing the correct kind of graph is important here.

Finally once we have the answers to make what changes in the booking systems or target certain kind of customers, production and presentation part helps us communicate it with managers to give them the right picture.

6. Explain with an example why reproducibility in a Visual Analytics system is important.

For a booking system an analyst may recommend to managers to buy more hotels in an area based on the analysis performed once. If managers choose to invest there but later on when same results are not reproduced based on same analysis the invested time and money is a waste.

7. Imagine that we had been running a pedestrian survey in the USA. We had let pedestrians write their height and sex on a sheet of paper. Afterwards, the handwritten statements were digitalized and saved in a csv file. Our

task is, to summarize and describe the heights of the survey participants to a person not involved to the study. How could the heights be visualized? Give reasons for your decision.

Visual analytics might not be required here at all because by statistical methods max, min, average heights of participants can be summarized which can again be summarized by sex of the participants.