

Exam

Business Intelligence and Process Modelling

Universiteit Leiden — Informatica & Economie

Friday June 8, 2018, 14:00–17:00

This exam consists of **20 questions** divided over four sections. Your answer can be in **Dutch or English**. Always give a precise, well-motivated but to-the-point answer. Write down any non-trivial assumptions. The number of points awarded for each perfectly answered question is listed in front of the question, and sums to **100 points**. Your grade is computed by dividing the number of points by 10. Good luck!

(17p) Visual Analytics & Event Data

- 1. (3p) What is “codeless reporting”?
- 2. (4p) Explain how Anscombe’s Quartet highlights the importance of data visualization.
- 3. (6p) Explain three different strategies to detect anomalies in event data.
Hint: examples of event data were amongst others discussed in the guest lecture about Eventpad.
- 4. (4p) Describe at least two things that are wrong about the visualization shown in Figure 1, and explain which visualization concepts or visualization metaphors they relate to.

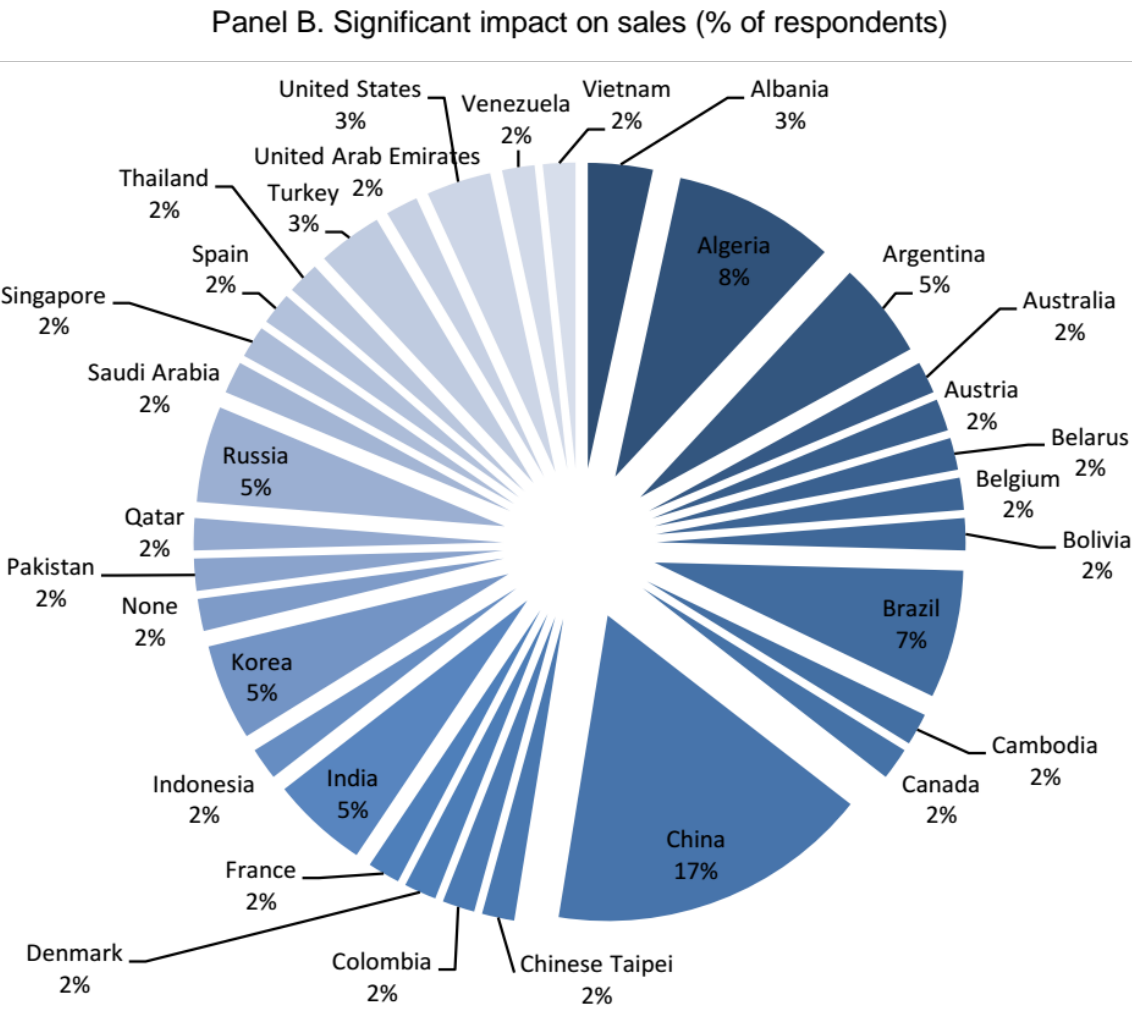


Figure 1: Visualization of the impact of an economic measure on the sales of a certain type of goods.

(30p) Business Intelligence

4. (5p) Name five characteristics of a good KPI, and give an example of each in the context of a modern taxi company such as Uber.
5. (2p) Good data quality is important in order to derive meaningful results from data mining. Name and briefly explain two important aspects related to data quality.
6. (3p) Give three reasons why data analysis is more efficient when done from a data warehouse, rather than from a transactional system.
7. (2p) Explain the concepts “knowledge gap” and “execution gap” in the context of the DIKW process.
8. (2p) Although hierarchical clustering has worse time complexity than k -means clustering, sometimes hierarchical clustering is still preferred. Explain two reasons why this could be the case.
9. (6p) Explain how outlier detection can be done in a supervised, unsupervised, and semi-supervised context. Also give examples.
10. (4p) What types of problems can not be learned by a perceptron? Also give an example of such a problem.
11. (6p) Name three parameters that one typically tunes when training a neural network, and explain each of them briefly.

(13p) Network Analytics

Case: Network of Wikipedia

Wikipedia is a large digital collection of information, organized into different articles on particular topics. This data can be modelled as a network, in which each node is a Wikipedia article, and a link between article A and B denotes the fact that we can click a link in article A that leads to article B. For example, the page on “Leiden University” links to the page on the “Prince of Orange”. We only consider the links within the article itself; not links in for example navigation menu’s, and restrict ourselves to the version of Wikipedia in one language, namely English (en).

11. a) (2p) What do the network metrics of indegree and outdegree indicate in general?
b) (2p) What do these two metrics mean in the considered Wikipedia network?
12. (6p) This network has one giant component, a power law degree distribution and exhibits the small-world phenomenon. Explain these three concepts.
13. (3p) Pages on certain countries such as India and China have a very high value for indegree centrality, but not a high value for closeness centrality (ignoring link directionality). What does this indicate in the network, and what may it hint at in the context of Wikipedia and its contents?

(50p) Process Modelling

- 14. (6p) Name three key differences between process models drawn using Petri Nets and using Business Process Modelling Notation (BPMN).
- 15. (8p) Consider a small airport with two runways. A runway is a strip of land on an airport, on which aircrafts can take off and land. Aircrafts that enter the airspace of the airport need to wait until they get permission to land. Similarly, aircrafts need permission to take off. At any point in time only one aircraft can use a runway but both runways can be used at the same time. The total number of aircrafts the airport can hold is ten, so an aircraft cannot land if there are already ten aircrafts at the airport (including runways). After an aircraft has landed, it unloads its passengers, loads new passengers, and then is ready for take-off.
 Model this situation airport as a Petri net.
- 16. (6p) Explain the terms discovery, conformance and enhancement in the context of business process modelling.
- 17. (4p) Name and explain four criteria for judging the quality of a discovered process model.
- 18. (7p) Give the reachability graph of the Petri net in Figure 2.

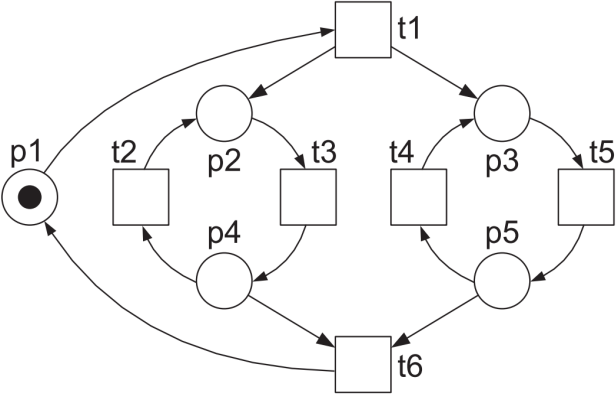


Figure 2: A Petri net.

- 19. a) (4p) When is a Petri net *sound*?
 b) (3p) What is a WorkFlow net (WF-net), and why is distinguishing this type of net so important?
- 20. (12p) Apply the α -algorithm to derive the footprint and final corresponding Petri net of event log L .

$$L = [\langle a, b, c, d \rangle, \langle a, c, b, d \rangle, \langle a, b, c, e, f, b, c, d \rangle, \langle a, b, c, e, f, c, b, d \rangle, \langle a, c, b, e, f, b, c, d \rangle, \langle a, c, b, e, f, b, c, e, f, c, b, d \rangle]$$

Explain your steps in detail.

End of exam. Please do not forget to fill in the evaluation form!