Solution Series 3

1. Possible key metrics of good performance and explanation why to choose this metric / these metrics:
    a. An elevator manufacturer is analysing its whole supply chain about the supply chain risk the manufacturer has.

      Key metrics: Avoidance or reduction of possible business interruption e.g. to a certain number of days; certain level of manufacturing material available; price stability of manufacturing material;

      Rationale: The elevator manufacturer wants to minimise the risk that a break in the supply of material incurs. A break could have different implications ranging from higher prices in materials to an interruption of the business.

    b. An insurance company is analysing the claims for fraud cases.

      Key metrics: Amount of detected fraud cases where the insurance company has not to pay claims.

      Rationale: Cost saving / avoidance of paying not justified claims.

    c. A watch manufacturer wants to reduce the amount of chemicals in the production process without loss of quality and efficiency.

      Key metrics: The amount of chemicals used per quantity compared to a certain threshold; processing time per quantity; number of quantity which does not pass the quality check;

      Rationale: Reducing the chemicals without any impact to the production process.

    d. A cargo firm wants to improve the transport routes and efficiency. The payload-distance is typically measured in tonne-kilometre.

      Key metrics: Reduction in tonne-kilometre by delivering the same or even more services / transports and/or delivering the same tonne-kilometre with less resources (trucks, people).

      Rationale: By optimising the cargo delivery costs (infrastructure and/or resources) can be decreased.

    e. A bank is analysing the non-compliant product sales for products which should not be sold to a person based on the person's risk profile.

      Key metrics: Number of non-compliant product sales found per X customer; number of non-compliant product sales found per client advisor Y;

      Rationale: Avoidance of fines and possible law-suits;

f. A machine manufacturing company is analysing how to improve the production process with regards of processing time, needed resources and required raw material.

Key metrics: Production cost per quantity and production time per quantity; production capacity;

Rationale: Increasing the production capacity with the same resource results in more revenue and higher margin; or production of the same amount of quantity with the same resources, time and material result in lower production costs and thus, higher margin.

2. Which data type do you have in the following examples:
    a. You would like to analyse the movement over time of book prices on Amazon for statistic books.

    Book prices: numerical, discrete, time series data, structured,
    Metadata: time point, "items left"
    Primary data: one is collecting them now for the analysis

    b. You are filling out a multiple choice test with four possible answers.

    Categorical, ordinal, cross-sectional
    Primary data: is collecting the chosen answers
    Maybe metadata e.g. gender, age, place, time point

    c. You are analysing tweeds on Twitter.

    Tweets: Unstructured, secondary data
    Metadata: time point, e.g. tweets regarding a special topic "tennis", "airlines"

    d. For the pricing of an insurance product you are analysing the claims and customer data of an insurance company.

    Claims data: nominal, time series data, structured, "continuous"
    Customer data: ordinal, cross-sectional data, but maybe also binary data if customers are labelled by "0" no claim, "1" claim;
    Meta data: time points of the claims
    Claims: primary data
    Customer data: secondary data

    e. You are supporting in a clinical trial as a statistical expert in testing a new drug against a control group. The probands are examined during the trials and the information collected and analysed afterwards.

    Medical information: nominal and/or ordinal, could be categorical, cross-sectional, but if data about the progress of a disease is collected: time series data, structured
    Primary data from the test and control group
    Metadata: gender, age

f. An online hotel booking web page is using the data when a customer has accessed the web page and with which device for determining the price range of offered rooms to that customer.

Depends on the data collected: nominal (time spend on a web site) and/or ordinal (which theme, type of device), discrete (search terms) or "continuous", cross-sectional or time series data (order in navigating through the page),
Primary data if real-time analysed

3. Sentiment Analysis:

a) Before reading the text, do you reckon that the bloggers like Coke Life?

Based on the feeling when reading the blog, it seems that the majority has something against Coke Life maybe 60% whereas the other are supportive to Coke in general and Coke Life.

b) By reading the randomly chosen blog, assess the quality of the sentiment analysis. Please note that certain comments might be rude:

The result is actually the other way around. E.g. the first blog posting shows a net result of 1 positive. Overall, there is a small positive tendencies which is against the feeling out from reading through the blog.

```
> output
positive negative net
1         4        3   1
2         9       15  -6
3         2        3  -1
4        10        6   4
5        14        2  12
6         0        0   0
7         0        2  -2
8         9        8   1
9         4        1   3
10       10        6   4
11        3        3   0
12        1        0   1
13        1        2  -1
14        5        1   4
15        3        3   0
16        2        1   1
17        1        0   1
18        0        1  -1
19       10       14  -4
20        0        0   0
21        0        1  -1
> summary
positive negative      net
88           72        16
```

c) Based on b) or your own examples, find expressions or idioms for which "sentiment_analysis" provides an erroneous sentiment.

E.g. denials "I don't like", "don't get me wrong" will not be detected and categorised as positive e.g. "like" or negative "wrong" or irony "well done" if somebody did a failure.