IDA Case Study 14

As head of the quality department of a big automotive group, one of your tasks is to ensure the quality of your suppliers. Since supplier inspections require much time and funding, it is especially important that these inspections are only carried out unscheduled in absolutely necessary cases.

For your decision, where the next inspection should be carried out, analyze the data from your group. It makes sense to carry out the inspection at the factories of the supplier with the highest failure rate. For this, use the data from the component suppliers for each factory in relation to the annual volume produced for each year of the available data. Also specify whether there are dominant product groups for the failures. Finally, decide whether a product group or a comprehensive process inspection is necessary.

Information about tasks and data sets:

All information, including when a vehicle was produced in which plant can be found in the Group's production data. You gain access to the group's own database and must decide for yourself which data you need for your analysis. In addition, the Federal Motor Transport Authority (KBA) sent you registration data and geodata, which are also stored in the database. In the following the categories of the data sets are enumerated:

- Part = Einzelteil
- Component = Komponente
- Vehicle = Fahrzeug
- Geodata = Geodata
- Registrations = Zulassungen
- Logistics delays = Lieferverzüge

It is also recommended to take a close look at the structure of the respective tables before starting the analysis.

For all instances of the supply chain, i.e. for individual parts, components and vehicles, production data with information on ID number, manufacturer, manufacturing plant, production date and failure entries are available.

The ID numbers consist of the part designation, manufacturer, plant and consecutive number. Example: 1-201-2011-3, component T1, produced by manufacturer "201" in plant "2011", 3rd component from this series.

For components and vehicles, you can also use BOMs (Bestandteile) that are declared using the naming convention Components_Name_Abbreviation. They contain information about all installed parts or components.

Hints for the approach to your case study:

The basic target is the development of an application to analyze your problem. It should be possible to manipulate certain settings interactively, that are automatically taken into account while analyzing. To evaluate the results, it is important to document your approach with an R Markdown file.

1. Import relevant datasets from the provided tubcloud folder.
2. Prepare the data according to the principles of *tidy data* and put them together in a single dataset.
3. Develop a Shiny app and visualize the following:
    a. A Pareto diagram for each component supplier factory that displays the relative and absolute error frequencies, broken down into the different components for any selected year. Afterwards, select a component factory for which an inspection is to be carried out and give reasons for your decision about the type of test.
    b. A line diagram should be generated by selecting a component from the Pareto diagram. The line diagram has to display the absolute failure figures of the component over the selected year, whereby each producing factory has to be displayed by a separate line. Calculate the total failure figures for all factories and display them as a further line in the diagram.
    c. Your basic dataset so you can prove your visualization. Remember: Only show necessary attributes.
4. Document the results of your analysis in an R Markdown file. Describe your data analysis process step by step and discuss the result by referencing convincing graphics from your developed app. The description of your approach is necessary for the traceability of your solution steps. If there are problems with your R code, the documentation can be used to acknowledge a correct approach.
5. Add a short manual for your application and conclude your documentation with analyzing your results. Remember to write your code according to the Tidyverse Style Guide and add comments to your code for explanation.

General tasks

In addition to the case study, there are further general tasks to do. Remember the documentation of your code, so that your solutions can be understood and if necessary, points for incomplete solutions can be given.

1. Logistics play a more and more important role in the product development of the automobile industry. Parts produced by the supplier must first be delivered to the OEM before they can be installed. What seems logical at first sight should be analyzed in more detail for a professional organization. Create a distribution for the logistics delay of component "K7". Use the production date from the data set "Komponente_K7.csv" and the goods in date from "Logistics_delay_K7.csv". You can assume that the manufacturer delivers the item to the goods out on the production day.
   For the model design in R, create a new data set "Logistics delay" that contains the required information from both data sets.
   Answer the following questions:
    - How is the logistics delay distributed? Justify your selection and briefly describe your approach.
    - What is the minimum/maximum time between goods out and goods in?
    - Determine the mean of the logistics delay.

- Visualize the distribution in an appropriate way.

2. Why does it make sense to store the available data in separate files instead of saving everything in a huge table? How do you call the underlying data base concept?

3. How many of the components K7 ended up in vehicles registered in the city of Dortmund?

4. Which data types have the attributes of the registration table "Zulassungen_alle_Fahrzeuge"?

5. You want to publish your application. Why does it make sense to store the data sets in a data base on a server and what is the many difference to storing the data sets on your personal computer?

6. On 11 August 2010 was an accident occured. The driver left the scene without a trace. The license plate of the car which caused the accident is still missing. Since you work for the Federal Motor Transport Authority, the police ask for your help to find out where the vehicle with the engine code "K1BE2-104-1041-32049" (corresponds the engine ID number) was registered.