Case Study 23

You are an employee of a fictitious company “106”, which sells gearshift systems to car manufacturers. In order to improve the product of the automatic gearshift system “K3AG1”, your company plans to analyze the damage cases of the last years. The automatic gearshift “K3AG1” is a standard component of the car brand “OEM1” and can be selected by the customer as an equipment feature of the car types “Type11” or “Type12”. From the management of your department you get the task to analyze the production data and parts lists of the customer and your suppliers. Key performance indicators such as service life, mileage, failure rates and usage data are of great importance.

Task and record information:

All information, including when a vehicle was produced in which plant and whether a vehicle is defective, can be found in the group’s production data. A vehicle is always considered to be defective if an installed individual part, an installed component or the entire vehicle is marked as defective. This logic also applies accordingly to components that contain defective individual parts.

You are given access to the group’s own database and must decide for yourself which data you require for your analysis. Furthermore, the German Federal Motor Transport Authority (KBA) sent you registration data and geodata, which are also stored on the database. The categories of data sets are listed below:

- **Single Part = Einzelteil**
- **Component = Komponente**
- **Vehicle = Fahrzeug**
- **Geodata = Geodaten**
- **Registrations = Zulassungen**
- **Logistics delay = Logistikverzug**

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

For all instances of the supply chain, i.e. for single parts, components and vehicles, production data with information on ID number, manufacturer, manufacturing plant, production date and error entries are available. The ID numbers consist of part designation, manufacturer, plant and sequential number. Example: 1-201-2011-3, component T1, produced by manufacturer “201” in plant “2011”, 3rd component from this series.

For components and vehicles there are also parts lists available, which are declared by the naming convention Components_Name_Abbreviation. They contain information about all parts or components installed in each case.

Notes on the procedure for your case study:

Create an R-Project (.Rproj) to insert all of the necessary files. The basic target is the development of an application to analyze and present your problem. It should be possible to manipulate certain settings interactively. Hence, the reactive updates are automatically processed while presenting. To evaluate the results, it is important to document your approach with an R Markdown file. Please write down every step you are proceeding and follow the instructions below:

1. Import relevant datasets from the provided tubcloud folder. Start this task by listing all documents which are required.
2. Prepare the data according to the principles of tidy data and put them together in a single dataset.
3. Develop a shiny app which meets the following criteria:
    - All used packages must be loadable by the current R-version.
    - The Application must be executable from the submission folder without any additional accommodations at any desired machine (e.g. at the examiners). It is advisable to test for this behavior before submitting.
    - The application should only refer to the dataset you created in step two.
    - The layout of the application should be in accordance with the target group and the corporate color, which is light-blue. Furthermore, a logo should be integrated in the layout. You can either design your own logo or you can use the logo of the department for quality science (Fachgebiet für Qualitätswissenschaften). In addition, the font has to be changed into “Source Sans Pro”.
4. Visualize the following in the application
    a. An interactive graph showing the registration history over time of the affected vehicles in the communities. Additionally, in a separate tab, use a heat map display to visualize damage hotspots. Incorporate popups into the map display that provide information on registered vehicles.
    b. Also, implement the ability to animate the registration history over time in the heat map over a specified period of time.
    c. Implement meaningful interactive displays in a separate tab for each named metric.
    d. Your underlying data set as a table, so you can prove visuals. Again, remember to show only the necessary attributes.
5. Document the results of your analysis in an R Markdown file. Describe your data analysis process step-by-step and discuss the result using meaningful graphs from your developed app. Describing your process will help you understand your solution steps. If there are any problems with your R code, documentation can be used to acknowledge a fundamentally correct approach. Therefore, comment your code carefully and follow the tidyverse style guide. Create a table of contents that is displayed in the html file and is clickable. Integrate meaningful screenshots of your app in the results chapter.
   The following bullet points should be listed in the table of contents as a minimum:
    - Importing the data
    - Data preparation
    - Creation of the final data set
    - Evaluation
    - Result

6. The submission folder (name: Case_Study_IDA_Group_XX) has to contain the following files (be aware of the file formats). XX needs to be replaced with the group number (e.g. Group 01 takes “01” as replacement):

Sub-folders to put into submission folder:

```
- Additional_files
- Data (please leave empty for submission)
- www
```

Files to put into submission folder:

```
- Final_dataset_group_XX.csv
- General_Tasks_Group_XX.rmd
- General_Tasks_Group_XX.html
- Case_Study_Group_XX.rmd
- Case_Study_Group_XX.html
- Case_Study_App_XX.r
- Case_Study_Group_XX.Rproj
```

7. Upload deadline for the submission folder is **15.09.2023, 23:59**.

General tasks

In addition to the case study, there are further general tasks to accomplish. Remember the documentation of your code, which helps your answers to interpret and if necessary, points for incomplete solutions can be given.

1. Logistics plays 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 detailed way for a professional application. Therefore, create a distribution for the logistics delay of component „K7”. Use the production date (“Produktionsdatum”) from the data set “Komponente_K7.csv” and the receiving date of incoming goods (“Wareneingang”) from “Logistikverzug_K7.csv” (logistics delay). You can assume that produced goods are issued one day after production date. For the model design in R, create a new data set “Logistics delay” that contains the required information from both data sets.
    a. How is the logistics delay distributed? Justify your choice with statistical tests and briefly describe your approach.
    b. Determine the mean of the logistics delay (watch out for weekends). Please interpret this number and discuss possible alternatives.
    c. Visualize the distribution in an appropriate way by displaying the histogram and the density function using “plotly”. Please describe how you selected the size of the bins.
    d. Please describe how you proceed, if you have to create a decision tree, which is describing the classification problem to classify whether the component (K7) is defective (Fehlerhaft) or not? (Hint: You might want to work with visualizations.)
2. Why does it make sense to store the available data in separate files instead of saving everything in a huge table? Name at least four benefits. The available tables represent a typical data base structure. How is it called?

3. How many of the parts T16 ended up in vehicles registered in Adelshofen?
4. Which data types do the attributes of the registration table “Zulassungen_aller_Fahrzeuge” have? Put your answers into a table which is integrated into your Markdown document and describe the characteristics of the data type(s).
5. Create a linear model from the table “Fahrzeuge_OEM1_Typ11_Fehleranalyse”, which sets the mileage in relation to suitable variables. Derive recommendations for OEM1 from this.
6. On 11.08.2010 there was a hit and run accident. There is no trace of the license plate of the car involved in the accident. The police asks for your help, as you work for the Federal Motor Transport Authority, and asks where the vehicle with the body part number “K5-112-1122-79” was registered.