Data Analyst Sample task & questions for discussion

Sample task – simple data aggregation and visualisation:

The attached CSV file contains data about the daily number of downloads from CRAN for several selected R packages in 2022. The CSV is based on data obtained using the *cranlogs* package for R (for details see https://github.com/r-hub/cranlogs).

Process the data and create an aesthetically appealing visualisation of the evolution of download numbers for each package throughout the year. Feel free to pick any data wrangling and visualisation library and any type of chart that you deem appropriate. You can get as creative as you want with colours, fonts, labels etc.

In addition, create a chart showing the average number of downloads of the *shiny* package per day of the week, as well as the minimum and maximum single-day value obtained, for each day of the week (i. e. Mon-Sun), across the timespan covered in the dataset. For the sake of simplicity, ignore or remove dates without any downloads.

Feel free to reach out if anything is unclear or you would like to know additional information!

Predictive analysis

Imagine that you are operating a business that is open 7 days a week. From long-term experience, you know that Mondays are the days with the highest sales volume, and usually about 45-50% of weekly revenue is made on Mondays and Tuesdays. Each Wednesday morning, you would like to forecast the current week's final revenue.

Which methods and what kind of statistical model could you use for this task? Which data do you think might be useful here? Which challenges and problems would you expect to face?

Data visualisation

Pie charts and donut charts are viewed as bad practice by many data visualisation professionals. Often, bar charts are cited as a better alternative for visualising certain data, for example, in the following blog post by data storytelling expert Cole Nussbaumer: https://www.storytellingwithdata.com/blog/2011/07/death-to-pie-charts.

What do you personally think about pie charts? Do you agree with the statements in the abovementioned blog post? Can you think of cases where pie charts might be a good way to visualise data nevertheless?