Applications of R in a Big Data company

Deeplink Interaction Clickflow

The sankey chart provides a general understanding of the user behavior specifically for the trivago search bar. Each vertical line represents one interaction step of the user. Starting from our homepage the user can choose different search contexts, for instance selecting a destination and performing a clickout afterwards or refining the search in order to find the ideal hotel.

R packages: shiny, googleVis

Dealform Interaction Clickflow

The sankey chart provides a general understanding of the user behavior specifically for the trivago search bar. Each vertical line represents one interaction step of the user. Starting from our homepage the user can choose different search contexts, for instance selecting a destination and performing a clickout afterwards or refining the search in order to find the ideal hotel.

R packages: shiny, googleVis

Average Booking Value (ABV) per state in the US

The upper map shows how the average value of a booking depends on the location of the user. The bottom map depicts the difference of the average booking value of each state when compared to the national ABV.

R packages: maps, gridExtra, ggplot2

Sherlock Anomaly detection

- Monitors a broad range of Data and Metrics 3x daily
- Notifies relevant person when values fall outside a dynamic expected range.

R packages: shiny, googleVis, RSQLite, assertthat, impaler, methods, tidyr

Receiver Operating Characteristic (ROC) curve for user return predictions

We made a prediction of the likelihood that a user might return to our website in the future using a Boosted Regression Trees (BRT) prediction model. This chart shows a ROC curve for the expected return rate of different users in the test set. It shows how the False Positive Rate (1 - Specificity) increases with the True Positive Rate (Sensitivity) when lower cutoff values are chosen.

R packages: RImpala, dplyrimpaladb, rpivotTable, gbm, dismo, ICEbox, Epi, SDMTools

trivago Facts

- Hotels: 1.3 million
- Employees: 1200 +
- 2018 visits per year: 1.4 billion
- 2018 data processed per day: 500GB
- 2018 data stored: 2 Petabyte
- 2018 data analyzed: 3.5 Petabyte

Crazy about data? company.trivago.com/jobs

July 5th
6:30 pm
User R! 2017 Conference
Brussels (BE)

team members

Alex Dophin, Data Analyst
Michael Frings, Data Analyst
Peter Brejcak, Happy Data Analyst making crawlers sad
Toni Linnenbruegger, Data Analyst