Introduction

Take the 515K Hotel Review Data in Europe, explore it, analyse it, and tell a nuanced story about it using visualisation.

- The dataset includes 515,000 customer reviews and scoring of 1,493 luxury hotels across Europe, which were collected by booking.com
- The data set (.csv file 238MB) can be downloaded from here.

Work in your teams (already assigned).

Task: each team needs to compile and submit:

1. A .zip file containing:
   a. an **executive summary** of your findings and key messages (250 words, 3 figures max) and
   b. an **R Markdown** file or a **Jupyter Notebook** to show the process and how to reproduce your key figures from the data.

2. A **Youtube/Vimeo link** (either public or hidden, but not private!) to a recorded presentation detailing your analysis process and justifications for visualisation design (video of 8~10 mins). **Please submit this in the comments section of the submission area.**

Instructions

Download the dataset

https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe

Data wrangling and transformation

- The data set is large and will probably not open well in Excel, so you will need to load the CSV file into aR or another platform of your choice like R or Python (e.g. Python)
- Keep a record of your analysis process so that the analysis is reproducible

Find a story

- Explore the story and make sure it is true and insightful
- You can make a story around a business argument

Create the visualisation and refine

- You must present at least 3 different chart types (i.e. don't just put 3 scatter plots in report)
- Consider **Why**, **How**, and **What** of data visualization

Deliverables

1 – Zip file with the following two files:

a. Executive summary

The goal of an executive summary is to communicate the key findings concisely with visualisation. Think of a company's executive as the target audience. They are not data analysts, but they are interested in making business decisions based on your findings.

- 250 words and 3 static figures (max)
- Include your group ID in the header and student names in the footer

File naming convention: BS2307Visualisation_Group#-Summary.pdf

b. R markdown **OR** Jupyter Notebook

The goal of this part is reproducibility.

- Select up to 3 key figures
- Create an R Markdown document with HTML template or a Jupyter Notebook and explain how you produced the key figures
- Briefly comment on each step, so other analysts (or future self) can understand what each step does
- You only need to submit the document/notebook as HTML file, which should show R / Python code as HTML output from loading the data to creating a visualisation.

File naming convention: BS2307Visualisation_Group#-KeyFigure.html=

2 – Recorded presentation

Please submit this as a Youtube or Vimeo link in the comments section of the submission area. Make sure your video link can be accessed and it's not set as 'private'.

The goal of a recorded presentation is to explain your analysis process.

- Explain how you used visualisation to find new insights
- Explain why, how, and what of data visualisation for your key figures
- Explain how you refined selected visualisations for communication
- Justify your design choices where applicable
- If interactive visualisation is used, explain how interactivity was used in analysis

8~10 min presentation

- You may choose to be in the video in person or just do a voice over against a series of presented visuals, or a slide deck that you run through with narration
- All voices of members of the group should appear across the recording.
- You may do a recording from camera phones, it doesn't need to have high production value. Just make sure the key parts of your presentation are visible to the eye and audible to the ear.

File naming convention: BS2307Visualisation_Group#-Presentation.mov