# Getsafe Data Case

For Getsafe, it’s vital to understand customers as precisely as possible.

To give you an impression, we provide you with 2 resources:

* reponses.csv: stores existing replies to a recent survey
  + information about lifestyle
  + information about existing insurance products
  + Preferences for possible future Getsafe features
* users.csv: basic user-information from our database (age, gender)

Some hints:

* Columns “user\_id” in responses.csv and column “id” in users.csv can be used to join data from the two sources
* Not all follow-up questions (in columns) in responses.csv (e.g. “Wann wirst du voraussichtlich deinen ersten Job anfangen?”) were shown to all survey-participants, only if it made sense from previous results
* “Amount of bought insurance products” is a good enough proxy for value that was realized already for the purpose of this case (in reality, this differs by avg. price per product, margins and some other details)
* People tend not to switch insurance products once they bought them, so it is safe to assume once people have a certain product, they won’t purchase it ever again (so there is also no customer value to be expected from selling them sth. customers already have/won’t switch)

**The Case**

Now, we ask you to take your time and come up with exciting insights. You can develop your own questions or use the following ones for inspiration:

* Which are the relevant customer clusters? What makes them a good fit for an insurance company?
* What’s your idea for a “sweet spot” to acquire new customers and how is it characterised (age, customer acquisition channel, certain events in life or lifestyle situations)? In other words, what describes our most valuable customers at time of acquisition?
* Is “travel health insurance” a good product to acquire especially valuable customers?

Don’t waste your time on slide-asthetics (powerpoint, …) too much, we only want to follow your idea (code) in a presentation meeting and see results in basic charts. We prefer great charts though (titles, labels, proper choice of chart type), because communication to non-data peeople is key. Pick any environment you feel comfortable with to do your analysis (Python, R, SQL, SPSS, Stata, … ).

Some thoughts can lead you down the rabbit hole and we ourselves get overly excited about some things and spend hours analysing things. Because of that it’s totally ok for the purpose of this case to limit your explorations to 2 hours max. Let’s rather have a great discussion together :-)