Covid dataset

* 2 different sets: till 2021, from 2021
* Both load in 2 lists 🡪 all the data is made in safety/security regions, but we have to convert them to province / province group

Formatting data

* merge the datasets together
* group the data by year and month so that we can later compare this to the other data set on the same time scale
* we filter all the relevant columns out of the data 🡪 (speaks for itself) 🡪 security region code (year, month, security region code (to see what province it actually happens)
* drop irrelevant data to make the code more efficient

1st real process:

* making the provinces and provincegroup 🡪 foud through looping through the whole dataset, i2 & i4 are the columns we want to change, we put the new data there which we will use. We take the code from the certain row, region code, we strip the text VR from it (so it only contains numbers, then we change that into an integer and do that number -1 so we get an index number for a list). (we could have also created a dictionary but we realized that later, in that way we didn t need to strip the data).
* We then rename the column for name clarity
* At last we grouped the values by province so that we can determine the
* The outcome is an overview of the hospitalized covid cases per month per province.

2nd process – travel dataset

* The dataset is different because it contains metadata, code that stands for outcomes so we first have to convert that into usable data.
* 1st scrap all the irrelevant data: data that is not about provincegroups
* We do that by checking if a certain region cchteriatic has LD in it.
  + If LD is not in the unit we append a new array: true, if it is: false. By that we filter the data.
* We have the data frame with true/false as outcome, of which we deleted the true ones.
* The same is applied for every trip characteristic with the ID 2031, which means that it is talking about months/days: we want to keep the monthly data (so days will be removed).
* After that we also remove the data of which after 2031 the number lower than 160 follows: because that stands for days (instead of months).
* Remove all the data from margins that have no code: MW00000, because that means that this data doesn’t include all the data we need (upper/lowerbound etc.).
* We also remove populationkey: A048709, because this
* Lastly we also remove all travel keys that do not include T001080 \*\* Incorrect
* And we rename values in region characteristics and trip char. So they state the province group & month instead of values.
* Lastly we remove the columns of ID, population, and margins.

Even if the Netherlands is a small country, there are still quite some differences between

the regions. While the north is somewhat sparsely populated, the western part with nearly

the same area contains 5 times more people. Due to that, the average age of people living

in the western region which is also often referred to as „Randstad “, is somewhat lower

than the average age of people in the Netherlands in general. It is reasonable to assume,

that those differences in population density and age distribution also had an influence on

mobility behaviour during the corona pandemic.

To examine this a little further, we grouped the provinces of the Netherlands in four

regions: North, East, South and West. A short characterisation is found in the picture

below. Our approach is to compare the change in average travelled time and average

travelled distance throughout the years between those different groups, in correlation with

the total hospitalisation rate during that time. Furthermore, we also want to have a closer

look on how the severity of corona infections could have changed the travel behaviour in

terms of different travel purposes. The investigated travel purposes are listed below.

In order to find a connection, we have formulated the following hypotheses in advance,

which we would like to answer within this report:

The graph above shows the relationship between the number of hospital admissions and the average travel distance within the Netherlands. Observing this relationship, one can say that the average travel distance decreases when the number of hospitalizations increases. Even though this graph doesn’t offset the data against a duration axis specifically, making it difficult to see when changes happened. The decrease in this graph can still be compared to the real-life circumstances of the Covid-19 period 2020/2021. For example, since the estimated incubation time of Covid-19 was set to approximately a week, it was able to spread quickly before people realized they were infected (e.g., Carnal, Football matches, big gatherings). Hence, due the government imposing restrictions (closing stores, restaurants, events, and more) people were indirectly forced to stay closer to home. However, this data shows a relationship for The Netherlands as a country, it doesn’t show the relation per region and travel purpose. Nor does it take the average travel time into account. Therefore, the following graphs will focus on those parts, trying to uncover which regions and purposes faced the largest change and why this could be the case.

The graph above shows the relationship between the number of hospital admissions and the average travel distance within the Netherlands per region. By doing so, it is slightly more detailed than the graph showing the relationship for The Netherlands as a whole. Hence, there are some obvious differences when looking at the regions individually. As can be noted, all regions face a decrease in the average travel distance, whereas the most Covid-19 hospitalizations occur in the south and west. This could be explained by the population density within these regions and several big gatherings such as Carnaval in the South. When looking at the graphs, we could state that there is some relation between the number of hospitalizations and average travel distance within The Netherlands, but also per region.

The graph above, a scatter plot with a trend line, shows the correlation of the average travel time per trip per month and the covid-severity measured by the number of hospital admission per month. The different Province-groups are displayed with different colours.

It can be clearly seen that the western provinces had the highest number of hospital admissions, followed by the southern and eastern provinces. The northern provinces have the lowest number of hospital admissions, which corresponds well with the population size of these regions.

Even if the average distance travelled per trip decreased with higher hospitalisation rates as seen in graph 2, the hypothesis Statement 4 that the average travel time would also be reduced, cannot be confirmed. The overall trendline shows that the average travel time per trip stays at about 27 minutes, no matter how many hospital admissions there are. This can be explained very well with Marchetti's constant, which is the average time spent by a person travelling per day. Marchetti stated this average time is approximately half an hour for a one-way trip (Marchetti, 1994).

A related theory is the one from Zahavi, which states that people have a stable daily travel time budget (TTB) and even with increased transport speed people tend to use this saving in travel time to travel further (Zahavi, 1974). Even though the average travel distance decreased, people still used the same budget travel time budget to travel every day. With shorter trips but consistent travel time it can be concluded that the travel speed must have dropped, which can be explained by a shift to slower transport modes like biking or walking instead of using a car. This supports the thesis that can be drawn from Zahavis theory, that there is a basic human appetite for travel.

Zahavi, Y., Travetime budgets and mobility in urban areas, Washington, DC United States 20590, Publication Date: 1974-5, https://rosap.ntl.bts.gov/view/dot/12144

Marchetti, C. (September 1994). "Anthropological invariants in travel behavior" (PDF). Technological Forecasting and Social Change. 47 (1): 75–88. doi:10.1016/0040-1625(94)90041-8.