

Management Summary: Modal split, Group 3, Question 17

Authors: Fardokhtsadat Mohammadi, Dragos-Andrei Brasoveanu, Mihai Cata, Mehdi Masinovic

The purpose of this document is to summarize and present our findings related to question 17 of the third exercise in Data Oriented Programming Paradigms. The analysis discussed below aims to explain potential trends in modal split across multiple countries, i.e. with which means of transport people tend to move, how this tendency has changed over time, and what factors could be influencing the modal split behavior.

In order to allow for better comparisons among countries, we chose to focus on countries centered in and around Europe only, because transportation related data is recorded on a regular basis as part of the continuous monitoring efforts of the European Union. The Eurostat statistical office offers such data on their official website, from which we have analyzed and studied several transport-related datasets in the time period starting from 1990 up until 2020.

The benefit of using these datasets is that the information extracted can be studied on a yearly basis and grouped together on the country level. Nevertheless, it was necessary to develop a standardized way for preprocessing all datasets, which included transforming data columns, dealing with missing data or special data formats, and aggregating information per country and year.

The biggest challenge in the preprocessing step was the missing data that several countries have for the earlier years of this analysis (1990-2009). Unfortunately, not all countries have had the appropriate infrastructure to record transport related data in the past, which is why many of them have missing information up until 1994 (Bulgaria, Hungary), 1998 (Estonia, North Macedonia), 2000 (Lithuania, Netherlands), and even 2009 (Montenegro, Serbia). Therefore, for more complex investigations - such as correlation analysis or feature importance - we decided to only focus on countries that had data available for the full time period studied.

We considered to extrapolate over the missing years using SPLINE extrapolation or linear models, but we concluded that it will be difficult to predict estimates for such long time periods, because for those countries, there is no data available for 4, 10, or even 15 consecutive years. Additionally, the relatively large number of 17+ remaining countries - that don't have any missing information - still allows us to extensively analyze the transportation data even when removing rows with missing information.

We were able to distinguish different kinds of modal split tendencies that differ in transportation type usage. Firstly, countries mainly located in Western Europe – such as the United Kingdom, Norway, Netherlands, France, Austria, or Germany – heavily rely on and consistently use car transportation as their main form of transport, where transportation by car makes up around 80% of usage, and bus and train transport usually revolve around 10% in proportions. This trend remains steady over the course of the last 30 years. Secondly, countries located in Eastern Europe – such as Bulgaria, Czech Republic, North Macedonia, Poland, or Slovenia – tend to have a considerably lower car transportation usage, with bus transport proportions consistently reaching up to ~30% in the majority of years studied. Lastly, there are countries that exhibit very strong changes in modal split tendencies – such as Slovakia, Romania, and Turkey. Over the course of 25 years, Slovakia went from a bus-transportation proportion of more than 40% in 1993-1998, to predominantly using cars for transportation from 2000-2019. Romania's main transport usage, on the other hand, is by car over the whole study period (1995-2019), but has experienced a consistent decrease in train transport usage from almost 30% in 1995 to less than 5% in 2019. Similarly, Turkey's most extensively used form of transportation was bus-transport from 1990-2001 with more than 50% usage over the whole time period, but this tendency has changed to bus transportation decreasing down to 30% and car transportation continuously increasing from ~28% in 1990 to over 70% in 2019. The modal split behavior for Austria (Figure A) and Romania (Figure B) is shown on page 2, where the proportions of transportation type usage (Blue: Bus, Orange: Car, Green: Train) is shown from 1990 to 2019.

We further analyzed the correlation structure among countries, to see whether the above described separation in countries centered in or around Western Europe, Eastern Europe, and a third unrelated group, would also be visible in the form of graphs. We can observe a split in two groups based on the correlation in train transportation, with one group mainly forming Western European countries, and the other largely filled with countries located in Eastern Europe. Figure C presents countries as nodes and depicts the similarity of countries by placing them into a group of nodes. The group structure could be affected by the smaller sample size studied for this plot, as we have only considered countries with fully available data across the whole study period (1990-2020).

In order to assess the effect of transportation infrastructure on the modal split behavior in countries, we investigated several factors related to transport. From our correlation studies, we found that the more a country invests in train transportation infrastructure, i.e. by expanding the railway length, the higher is the usage of trains. The increase in railway length and train-transport opportunities effectively decreases and negatively correlates with bus and car usage – meaning that as people tend to use trains more, they travel less by car or bus. In contrast, there is a small positive correlation between the GDP of a country and the transportation type usage: the higher the GDP of a country, the greater is the usage of buses and cars, but a smaller train usage. We argue that a larger GDP indicates more wealth in a population, and therefore could favor the usage of cars over trains. Figure D presents a correlation plot for Slovakia, which depicts the correlation structure of train usage compared to several factors such as motorway and railway length, road traffic deaths, GDP of a country, and air transport usage.

In addition to correlation analysis, we investigated which one of the transport related factors we studied have the greatest influence on the modal split of a country. Here, we trained a random forest model and used the tree impurity within the tree to assess the contribution of each feature studied. In figure E, the feature importance for car transport is illustrated, which shows that the railway length, air transportation usage, and the GDP of a country have a great impact on how frequently cars are used as a form of transportation. We observed similar results in the feature importance analysis for bus and train usage.

Figure A

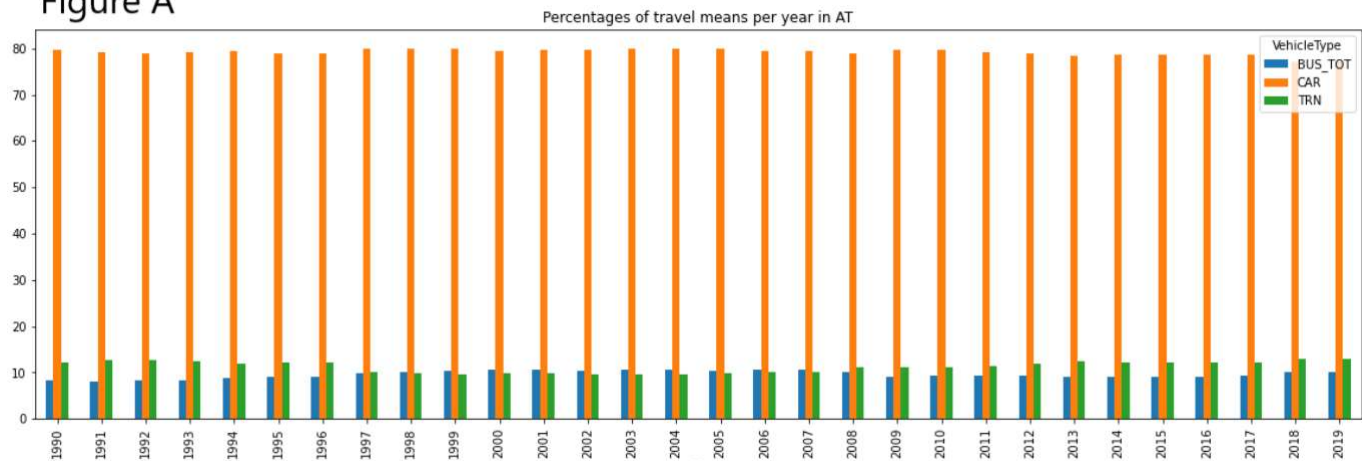


Figure B

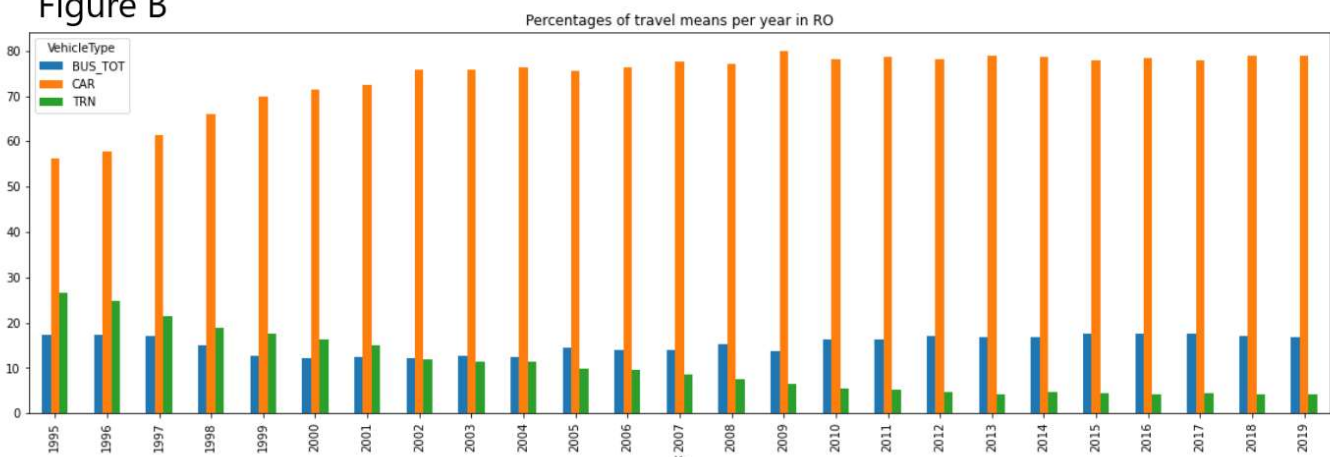


Figure C

Correlation between country trends in train transportation (heatmap)

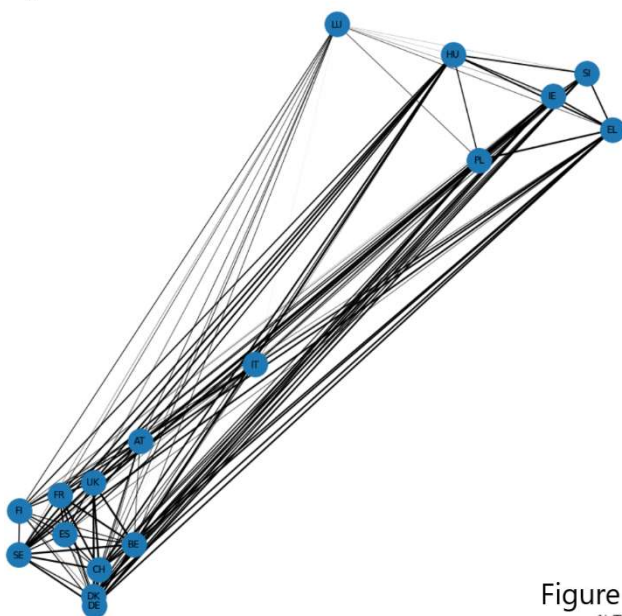


Figure D

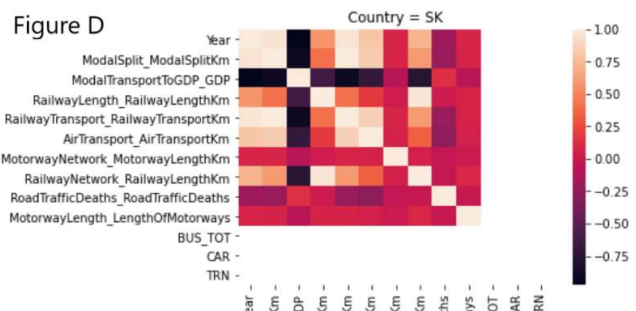


Figure E

