Introduction for Python 6022 Assignment

After the end of WW2, many countries’ governments around the world believed cars and airplanes were the future of transport and invested heavily in airports and highways. While cars and planes are indeed useful forms of transportation and certainly have their places in the world economy, there exists another form of transportation which can transform and revitalize the cities and regions in which it’s built: high-speed rail. As a concept, high-speed rail (HSR) is not a new technology, yet the environment benefits and positive socio-economic externalities it can bring to communities are nothing short of extraordinary. Large reductions of greenhouse gas emissions, less congestion on roads and airplanes alike, and the spurring of development along corridors are just a few of them.

Within the last few decades, many countries in Europe have chosen to create and expand their HSR networks, connecting populations with a faster, greener, and more convenient mode of travel. However, there are still major gaps in those networks, especially across international borders. And given the EU goal to become a climate-neutral economy by 2050, it is crucial for investments in high-speed rail to be carried out in a timely manner. The only question left to answer is where such systems should be built. As such, this report was created to try and answer this question:

**Which city-pairs in Europe could stand to benefit the most from a high-speed rail connection?**

The answer to this question would be based upon three main factors:

* 1. The number of current air passengers.
  2. The distance between city centers.
  3. The metropolitan population of both cities.

In addition, two sub-questions were devised to place and give better context to the main question:

1. Which city pairs already have the highest number of air passengers?
2. Which city pairs fall within the set distance range?

This report only focuses on cities which do *not* already have a HSR line between them. What constitutes a high-speed rail line was defined according to Category 1 of the European Union Directive 96/48/EC, Annex 1, which requires tracks to be specially constructed for high speeds, permitting a maximum speed of at least 250 km/h. For the purposes of this report, it was decided that if a route between two cities only has a segment where 250 km/h is reached, and not along the whole length of the line, it would be considered a normal-speed line, and thus eligible for inclusion in the study. [change to: reaches at one point]

It was decided that the geographical area of the report was to be constrained to six of the largest economies in the EU: France, Germany, Spain, Belgium, Luxembourg, and the Netherlands. These countries were chosen due to their economic and political strength within the EU, their well-developed rail networks, and their likeliness to invest further to improve said networks.

High-speed rail networks are expensive to build, and there are a few reasons for not building an HSR connection between cities, the first and foremost of which being distance. After consulting multiple sources, the ideal range for a HSR connection between two cities was determined to be between 150 – 1200 km. This range was considered appropriate because it is where the door-to-door travel time for HSR will be most competitive with other options. For shorter distances, the time differences between HSR and normal speed rail don’t justify the higher costs of a HSR line, and longer distances will be in heavy competition with airplanes, especially so if those routes go over large bodies of water and/or mountain ranges.

Data from Eurostat was utilized due to its breadth, good formatting, and ease of access. A four-year timespan, stretching from 2016 – 2019, was decided upon so passenger number data could be compared from year to year and trends could be extrapolated. Data from after 2020 was not used due to the large impacts of the COVID-19 pandemic and the subsequent lockdowns on the global economy and passenger traffic between countries.

To Add(?)

[There was no lower population limit placed on cities, simply sorted biggest pop to smallest]

[Geographic features can make a huge difference with capital costs, and so were only considered on a case-by case basis, e.g Frankfurt-Mallorca is huge, but a bridge across the Med would be financially ill-advised]

**Data Pipeline**

Add in exploratory data raison d'etre