The dataset collects all the information about the population and location of defined London boroughs, customer zones, and CLCs, about the distances between each CLC to each customer zone, about the unit price of AVs (Autonomous Vehicles) and TVs (Traditional Vehicles), about the delivery cost, and about the CO2 emission and its cost. All these data are used as inputs for the mathematical model, and the information about the population is also used for calculating the infection risk during pandemic situations. The detailed data and all references can be found in the Excel file ‘Data for research’.

First, the sheet 'Customer zones information' lists the 91 customer zones within the Great London and each zone's details that include the population as well as the latitudes and longitudes of the geographical center of each zone. Each zone’s population is gained by multiplying the population data in the 1960s by a coefficient, which is calculated by dividing the current population of a larger zone by the sum of populations of the small zones in the 1960s that comprises the large zone. Finally, the total population () is calculated by adding populations of the 91 zones together.

Second, in the 'CLCs information' sheet, 20 selected CLCs are presented with their addresses as well as their latitudes and longitudes. The CLCs are the 20 largest distribution centers of a leading grosser brand within the Great London, and their locations are extracted through Google Map.

Third, the sheet 'Linear distances' presents all the linear distances from th CLC to th customer zones that are calculated according to the latitudes and longitudes of the 20 CLCs and 91 customer zones. However, to better represent the real distance of non-linear routes vehicles actually run, this research uses the adjusted distances that are calculated by multiplying the linear distances by an adjustment coefficient, as presented in sheet 'Adjusted distances'. Real running distances of 91 routes, from the first CLC to 91 zones separately, are found by Google Map, then the adjustment coefficient is calculated as the mean value of the results obtained by dividing real distances of the first 61 routes by their linear distances. The real running distances of the remaining 30 routed are used to test the degree of confidence of the transfer coefficient by MATLAB. Fourth, the result shows that this adjustment coefficient is efficient at the 95% confidence level.

Fifth, in the 'Vehicle price' sheet, the price per TV is calculated as the mean of the unit prices of the top 7 best-selling vans in the UK according to a UK van market research conducted by Commercial Fleet (2014). While the price per AV is calculated in the 'Vehicle price' sheet according to the finding of Fagnant and Kockelman (2015) that the price per AV will keep ₤10,000 higher than the price per TV in the future 15 years.

Sixth, the sheet 'CO2 emission summarizes the CO2 price and the CO2 emission for AVs using electricity and TVs using gasoline.

Finally, the sheet 'Delivery cost' summarizes the relative costs for delivery for AVs using electricity and TVs using gasoline.

**References**

Commercial Fleet, 2014. The UK van market. <<https://www.commercialfleet.org/sales-figures/the-uk-van-market>>

Fagnant, D.J. and Kockelman, K., 2015. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transportation Research Part A: Policy and Practice, 77, pp.167-181.