Problem Identification Overview and Context

Our client is the Urban Design Group, an organization that raises and allocates funds designed to improve the quality of life in cities. The Urban Design Group is currently working on a project to improve the well-being of the citizens in London. With the current population of London at around 9,648,000, and a trend of ~1-1.5% population increase per year over the past few years[1], the importance of urban design has become an important consideration for the city's urban planning commissions. The Urban Design Group has been awarded a government grant of 5 million pounds, and the company wants to know where to allocate these funds over the next 5 years so as to increase the well-being of citizens in the London boroughs.

Over the past few years, London's local government has been making a push for "greening London"; the London Plan 2021 outlines an integrated environmental strategy for increasing urban greenspace, as well as improving air quality and mitigating the effects of climate change. The Urban Design Group would therefore like to know whether allocating funds towards increasing natural areas and urban greenspace would be the best way to spend its grant money, or whether the funds would be better used if put towards programs that decrease childhood obesity rates, crime rates, or unemployment. The Urban Design Group has asked our data science team to predict what factors are most important for predicting the longevity (as a measure of well-being) of London citizens.

The focus of this project will be on developing a model to predict life-expectancy (within an error margin of 3 years) and identifying which factors are most important for predicting longevity within the London boroughs. We will use data from the London Ward Well-Being Scores dataset, collected by the Greater London Authority over the period of 2009-2013. This dataset includes information on life-expectancy, as well as 12 different well-being indicators such as childhood obesity rates, crime rates, unemployment rates, and access to greenspace and public transport. We will also use a second well-being dataset from the Greater London Authority that uses a different metric (total area) for urban greenspace so we can assess whether different greenspace metrics affect our model. We will use these combined these metrics to produce the best model for predicting life-expectancy, and produce a project report and slide deck presentation to present to the Urban Design Group.

We anticipate that the key stakeholders for this project include the CEO and vice-president of the Urban Design Group.

---

1. Data Source: United Nations - World Population Prospects