The Effect of Education on Crime Rate in Bristol, UK.

A DATA SCIENCE CAPSTONE PROJECT

Rhys Morgan | IBM Applied Data Science Capstone | July 30, 2020

Contents

Introduction

Background

Bristol is a bustling city enveloped in the hills of South West England and straddling the River Avon. The centuries-old heritage and classical architecture give this place its own unmistakable identity.

The city dwellers of Bristol grew in population from 535,907 in 2011, to 679,000 in 2019 as its high street markets and shopping districts continued to prosper. Hence, the original plan for this project was to investigate the growth of population against the growth of retail stores and businesses. However, with the current state of COVID19 in the UK, future growth has become much harder to predict.

As an aerospace engineer, Bristol is a key area for me – it is home to the British Ministry of Defence, BAE Systems, Airbus, MBDA, to name a few. Therefore, I would personally like to know what areas are considered better than others, and why. To kickstart this journey, I will look at education and crime – two opposites that can either make or break a district.

Figure 1: Houses in Clifton.

The Problem

Bristol is generally a safe place – especially in the key tourist areas. However, there is still moderate risk of petty crimes such as pickpocketing and robbery. The poorer districts of the city have a generally higher rate of crime compared to wealthier areas. This is due to many reasons, but my question for this report is:

"Does education directly affect crime rate in Bristol?"

Yes, search engines would most certainly have multiple answers, accompanied by many arguments. However, I wish to explore this question thoroughly with unbiased, publicly available data, and then go even deeper by looking at neighbourhoods individually and pinpointing locations that would greatly benefit having a better education system.

A safer city will most definitely result in a better economy. Thus, being able to identify weak points in the system to better educate the population is a very powerful tool indeed. This investigation will cover all ages from nursery to university, as different age groups may provide insight into which stage(s) of education may be struggling the most.

Aims

The main goal of this report is to explore each district in Bristol, comparing their education systems with their crime rates each year from 2016 to 2020. With these findings, I hope to make recommendations for each district on what they should do next, so that the population may feel safer and more secure in their respective neighbourhoods.

As a secondary goal, I hope to show that there is a strong correlation between the two, as that could reinforce the belief that people who commit crime generally have poor childhoods.

Objectives

The investigation will begin by determining the most suitable method of comparing education and crime. Then gathering that information and turning it into data that can be manipulated and analysed.

By using machine learning and visualisation techniques, I will discuss predictions and any anomalies that occur. Following this, I will come up with suggestions based on the data from each district.

Throughout this report, I must use Foursquare's API and Python's data science packages to be able to meet the requirements of the IBM capstone project.

Data

How?

Data can be used to solve this problem by providing valuable information on crime rate and education. Through visualisation techniques, the data can show different trends and patterns that otherwise would be very difficult to see.

When using data science tools such as Python's Pandas, Numpy, and sci-kit learn packages, the data can be extracted from multiple sources, collated, manipulated and analysed to provide future predictions and recommendations.

What?

The data that this project is going to need includes crime records of every district in Bristol from 2016-2020, along with all local schools ranging from nursery to university.

For choropleth charts and data mapping, I will also require geographical coordinates of the districts and crime reports.

Where?

The crime data must come from a reliable, unbiased source. Preferably from local councils or government websites. Therefore, the crime data will be taken from Bristol's government website (opendata.bristol.gov.uk). Foursquare's API will be used to gather information on the schools in each district.

Geographical data can be obtained either from the same sources as the other data, or by using Python packages to turn addresses into geographical coordinates (latitude, longitude). Fortunately, the same site where the crime data is found also provides GeoJSON data.

Preferably, the data should be in CSV and JSON format for ease of use. However, Python can be used to scrape tabular data from HTML pages (websites) should the need arise.