Data science Project:

Project (80%) This page sets out your data science project.

It will consist of between 3 and 8 charts, tables or visualisations. You must also briefly discuss four topics: 1. The aims of your project.

The aim of the project is to show the effect of organized crime on aspects of quality of life. Such as Average cash expenditure, University Probability, Private/Grammar school concentration.

2. The data you used, how you accessed it, including notes on automation/replication;

3. Challenges in data cleaning and/or analysis, and the tools you used to overcome them;

4. Your conclusions.

The written text that accompanies your work must not exceed 800 words in total. Key marking criteria Your site (portfolio and project) will be marked against the follow cross cutting criteria:

1. Accessibility

a. Hosted on a GitHub page – open data approach [Document title]

b. Clear links to data and notes in you code

c. Mobile first: site should be clear on a smartphone screen

2. Empirical design. a. Clarity of question b. Choice of data sets c. Empirical approach

3. Automation.

a. Use of APIs

b. Replicability of project

4. Impact.

a. Key takeaways from the project

b. Interactivity of the data.

5. Description.

a. Clarity of write up

b. Logical steps from data to conclusions

c. Does the project stimulate further work?

Criminal Organization,

Part 1: Heat map of UK counties showing number of organised crime arrests.

Part 2:

Machine Learning regression model.

Regression 1:

X= Number of organised crime arrests.

Y = Purchases of items over £1000 made in cash across county. (Trying to understand the effect on cash income of people in areas of varying organized crime levels.

Regression 2:

Region Level

X= Number of organised crime arrests in a county annually year end June 2022, as percentage of county population.

Y= Percentage of Enrolled students who achieve at least 2 A level equivalents in 16-18 studies.

Can create an additional map chart and colour co-ordinate the regions that follow the linear regression model the best (error term size-based colour code).

County Level

Pie chart showing the change in crime type highlighting what could be considered organised. Between 2021 to 2022.

References:

Map1 =

[JOBS05: Workforce jobs by region and industry - Office for National Statistics (ons.gov.uk)](https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/workforcejobsbyregionandindustryjobs05)

Regressionworld =

<https://www.bing.com/ck/a?!&&p=b024a7abd6df160aJmltdHM9MTY3Mjk2MzIwMCZpZ3VpZD0yMTRkZjQ3OS1lNjJiLTZlM2MtMWQ5Yy1lNjBiZTdkMDZmMTgmaW5zaWQ9NTQxNQ&ptn=3&hsh=3&fclid=214df479-e62b-6e3c-1d9c-e60be7d06f18&psq=global+organised+crime+index&u=a1aHR0cHM6Ly9nbG9iYWxpbml0aWF0aXZlLm5ldC9pbml0aWF0aXZlcy9vY2luZGV4LyM6fjp0ZXh0PVRoZSUyMEdsb2JhbCUyME9yZ2FuaXplZCUyMENyaW1lJTIwSW5kZXglMjBpcyUyMGElMjBtdWx0aS1kaW1lbnNpb25hbCxhc3Nlc3NtZW50cyUyMGFuZCUyMGV2YWx1YXRpb25zJTIwYnklMjB0aGUlMjBHSS1UT0MlRTIlODAlOTlzJTIwcmVnaW9uYWwlMjBvYnNlcnZhdG9yaWVzLg&ntb=1>

[Education Index - Wikipedia](https://en.wikipedia.org/wiki/Education_Index)

Words

This Webpage is designed to explore the impact of Organized Crime in the UK and understand its significance to education/employment outcomes relative to other countries.

Map:

Our UK Map is a visual representation of the number of Weapon Offences occurring between 2021-22 additionally, the share of specialised job roles in each region is presented this has been taken from the ONS website. The visualisation is designed to help make connections between the level of violent crime and regional job outcomes. While no direct relationship between violent crime and percentage of jobs considered ‘specialised’ is obvious, Urban regions such as London show higher concentrations of both crime and job specialisation oppose to somewhere such as Wales which is considerably less dense.

Construction of ‘percentage of specialised jobs’ was calculated by summing the included industries divided by total jobs per region.

Regressions:

The regression analysis is broken down into three categories, County, Regional, and Country. The County and regional regressions use the same axis, for the X-axis an estimation of the proportion of organised crimes is used. This has been constructed using the data.police.uk series, in which crimes Homicide, Violence (with and without injury), Robbery, Theft Offences, Burglary, Criminal damage, Drug offences, and Possession of weapons were logged as ‘Potentially Organised crime related’ and used in data series.

https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/policeforceareadatatables

While the validity of this model is questionable as not all the mentioned crimes are 100% associated with Organisational behaviour the mentioned crimes make up over 90% of all convictions related to organised crime in the UK.

For our measurement of the effect on education, data from Gov.UK was utilised taking each individual school average attainment percentage of at least 2-A levels or equivalent and subgrouping at the Region and County Level of average attainment.

County/Region Results:

Our Regression models at the county and regional level suggest almost no relationships between ‘Potential organised crimes’ and ‘Average Pass Attainment’, and our line of best fit suggest a positive relationship in which an additional 10% increase in ‘Organised crime’ would lead to a 1.6% increase in Education attainment (County).

Limitations: Due to the limitation of data on organised crime its possible the model is over biased towards the portion of Organised crimes, a reason could be noise errors in the data or perhaps the lack of sample points, particularly for County level.

To Overcome this an Additional Country level regression is presented to see if any relationship between organised crime and education levels occurs, which is not being explained at the UK level.

A more reasonable regression output comes from the Country level regression with a higher and a negative correlation. From this we can interpret that a relationship does exist between the two indicators however, the UK level of Organized crime is small relative to countries such as the Philippine’s and can’t be accurately measured with the level of noise in the data.

Charts:

The first chart measures different employment levels for Age range and the Drug use amongst 16–24-year-olds. Wile no causality can be stated here it does appear that the unemployment spike between 16-17 year olds from 2008-2014 is also associated with a decline in drug use amongst both 16-20 and 20-24. Intuitively this makes sense as less disposable income post the financial crisis would reduce the demand for discretionary items.

Conclusion:

While organised crime seems to affect the level of education on a global level, in the UK less prominent effects are seen, this is likely due to very little major crime organisations in the UK compared to other countries.

Challenges Cleaning and Accessing data:

Due to the disparity of information amongst organised crime it made it challenging to find suitable datasets additionally data was compiled by annual reports which made it limiting in terms of API extraction.

Multi Line Chart: Used an API to extract data from ONS as well as 2 CSV files for the share of jobs in an industry.

The Country level regression utilises ML techniques and, The scraped data was done using beautiful soup and Wiki Plug in.