

# Intro\_to\_DS\_2022\_Assignments

git REPO : [https://git.arts.ac.uk/20002791/Intro\\_to\\_DS\\_2022\\_Assignments.git](https://git.arts.ac.uk/20002791/Intro_to_DS_2022_Assignments.git)

### **Assignment 1 Decolonizing Wikipedia Page on Limehouse London and Opium Den**

a report on discrimination found on the selected Wikipedia page

### **Assignment 2 Linear Regression and other Analysis on the Selected Dataset**

Exploring the cause of the gender pay gap in Taiwan's creative industry

1. a data file that I manually downloaded and cleaned from the Taiwanese government's open data source
2. a pdf file of the analysis report
3. a notebook that contains the code

### **Assignment 3 Clustering and Interpretation Analysis**

two files contain two projects of clustering

#### 1. Customer Segmentation (clustering a cleaned dataset from Kaggle)

- 1.1 Segmentation data (csv file)
- 1.2 Description of the data (xlsx. file)
- 1.3 a notebook contains the code used

#### 2. Taiwan National Health Insurance (clustering of a unorganised dataset from the government's website)

- 2.1 data used - NHI\_Diarrhea (csv. file)
- 2.2 NHI Data Description
- 2.3 a notebook contains the code used