Continuous Assessment 2

Author: Ciara O'Hara

Programme: MSc in Data Analytics May 2023

Student ID: sbs23015

Email: sbs23015@student.cct.ie

May 24, 2023

Contents

0.1 brief

Abstract

Introduction

First steps were to determine and identify an appropriate Irish dataset under the theme of Constrction. There were two datasets found: A House Construction Cost Index from 1975 - 2017, and a social housing construction status reports from 2017 - 2021. Homelessness in Ireland is an issue of major significance and public importance at the moment. In fact, this is an issue across Europe to varying degrees, with the excpetion of Finland (*Finland's Zero Homeless Strategy: Lessons from a Success Story* 2021). It was decided to try to explore - from publicly available data - what Finland has done differently, the factors that may have impacted that, and to attempt a sentiment analysis around the topic in both Ireland and Finland. As such, the next step was to gather appropriate and complementary Finnish data. Statistics Finland's free-of-charge statistical databases, Tilastokeskus was found, which contains various construction related data (as well as data on many other aspects of Finnish society).

https://www.linesight.com/insights/regional-report/europe-2021/

https://www.geeksforgeeks.org/newspaper-scraping-using-python-and-news-api/

https://towardsdatascience.com/web-scraping-news-articles-in-python-9dd605799558

https://www.geeksforgeeks.org/newspaper-article-scraping-curation-python/

Materials and Methods

3.1 Data sources

Social housing construction status reports for Ireland were taken for 2018 - 2021 from the publicly available Department of Housing, Local Government, and Heritage data on data.gov.ie (*Department of Housing, Local Government, and Heritage* 2023) with the house construction cost index data for Ireland from the same source (*HSM09 - House Construction Cost Index 2023* 2023). The Finnish data was taken from Tilastokesks, Statistics Finland's free-of-charge statistical databases. Building and dwelling production (*12fy – Building and dwelling production, 1995M01-2023M02* 2023) and Building cost index (*11na – Building cost index by type of cost, annual data, 1990-2022* 2023) data were used. This varies from the Irish data in that it is residential building production in general, not social housing production exclusively, so the comparisons made between countries will be more general. However, it was deemed sufficient for this project to provide a demonstrable example. Finally, web scraping was carried out to gather newspaper article data for sentiment analysis. Articles from various newspapers were used; The Irish Independent (*Finnish model shows how a more radical approach could solve homeless problem* 2018), The Irish Times (*Homeless response 'should focus on needs of children'* 2023), The Journal (*The government is trying to reduce chronic homelessness ... Here's how Finland ended it* 2018), The Helsinki Times (*Homelessness can be eradicated by 2027 with close cooperation: Report* 2023), YLE (*Has Finland really solved homelessness?* 2022), the Guardian (*'It's a miracle': Helsinki's radical solution to homelessness* 2019; *What can the UK learn from how Finland solved homelessness?* 2017), Politico (*To help the homeless, close a shelter* 2019), CBC (*London wants to eradicate homelessness. Here's how Finland is doing it* 2023; *Housing is a human right: How Finland is eradicating homelessness* 2020), The Toronto Star (*How Finland managed to virtually end homelessness* 2023) and The Christian Science Monitor (*Finland's homeless crisis nearly solved. How? By giving homes to all who need.* 2018). Finally population statistics came from eurostat (*Population on 1 January* 2023).

3.2 Programming

Data was gathered from three sources, as above. For the cost index and residential building completion data, the Irish data was downloaded in .csv format, though it was also available in .xlsx (Excel), JSON and .px formats. Had JSON format been used, similar techniques to those used for the Finnish data would have been appropriate. The Finnish data was gathered using APIs and JSON, though it was also available in .xlsx as well as .csv and other delimited formats, which would have allowed the use of pandas with the pyarrow parser, for more efficient and faster loading (*The fastest way to read a CSV in Pandas* 2023). Once gathered and imported in JSON format (*PxWeb API* 2023), a for loop was used to iterate through the data and extract the variables of interest for the analysis. An alternative to this would have been to use pandas' JSON functions to read `(pandas.read_json)` (*pandas.read json* 2023) and then to tabulate `(pandas.json_normalize)` the data into a pandas dataframe format (*(pandas.json normalize)* 2023).

3.2.1 Python libraries used

3.3 Data preparation

3.3.1 Cleaning the data

3.3.2 Data visualisation

The colours chosen for the sentiment analysis visualisations were designed to reflect the traditional colours associated with political leanings, for example, red for left-leaning politics, magenta for centrist, and blue for conservative politics.

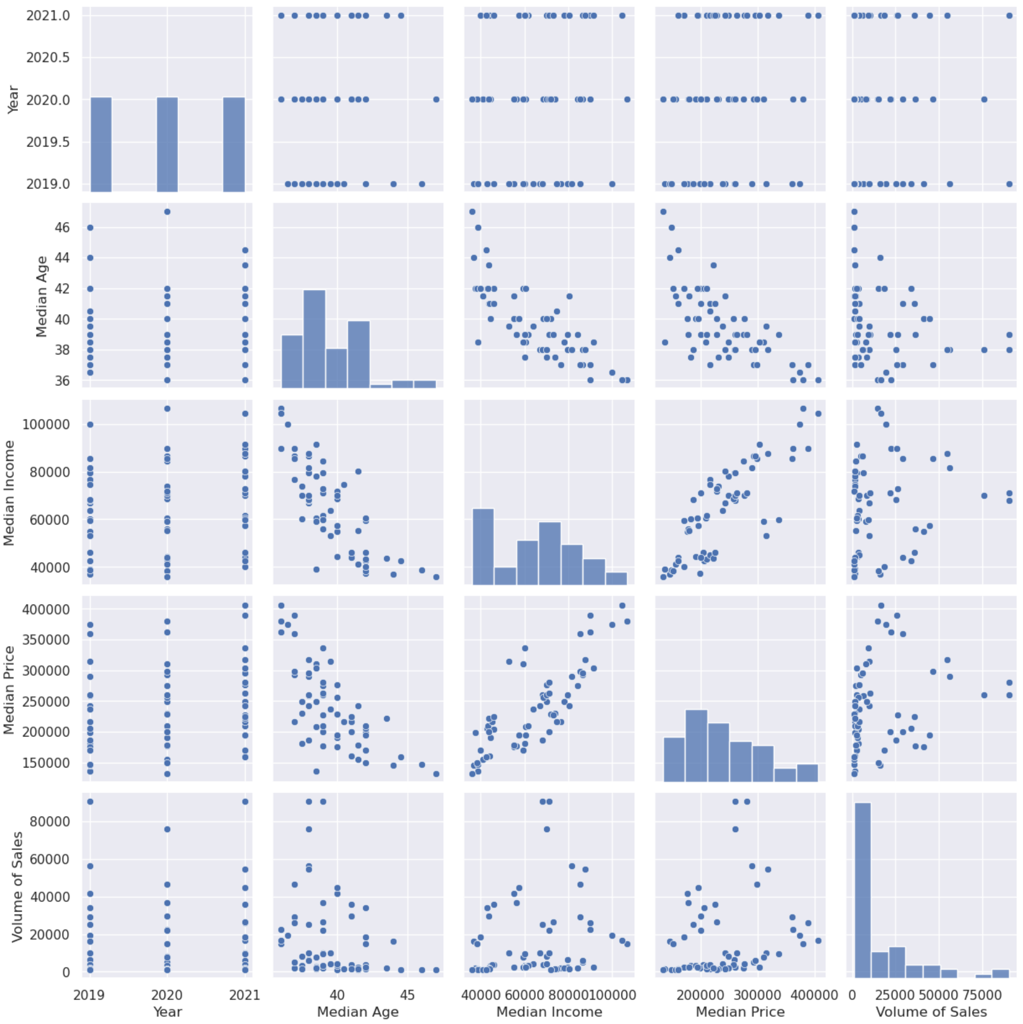

Figure 3.1: *Sample figure.*

3.4 Statistics

3.4.1 Python libraries

3.4.2 Statistical analysis

ARIMA is an appropriate tool for modelling regular-interval, non-seasonal, time-series data, and may be used to predict data in the series into the future (*What Is ARIMA Modeling?* 2023). Univariate analysis was employed here, in that only previous values in the time series were used to predict future ones. First an augmented Dickey-Fuller test was carried out to determine whether the data was stationary. This test showed that it was not stationary (p = 0.6, DFτ = -1.33), i.e. it had some time-dependent structure (*Augmented Dickey-Fuller Test in Python (With Example)* 2023). Autocorrelation and augmented Dickey-Fuller were used to determine the order of differencing (Raval, 2023), which was 1 (augmented Dickey-Fuller: p = 1.0 x 10^{-5}, DFτ = -5.17), and the most significant degree of lag was determined from the autocorrelation plot (taken to be 50) (Raval, 2023; Nadeem, 2021). Maximising the goodness-of-fit of the ARIMA model is done by determining the best values for p, d and q and this was done by minimising the RMSE (Raval, 2023; *How to do cross validation for time series?* 2023) using cross-validation on a rolling basis, which ensures that no 'future' data is used in the training of the model (Shrivastava, 2020; Hyndman and Athanasopoulos, 2018). An order of differencing of 2 was later found to give better results (based on RMSE) than a value of 1.

Ireland and Finland have roughly similar populations (*Population on 1 January* 2023) but to ensure that the

As there was only four years of Irish social housing construction data available, the irish and finnish cost indexes and construction activity could only be compared over 4 years, which is not a long enough dataset. The inferential analysis was still carried out, but in reality, any results based on such a small dataset would be meaningless. In order to make valid comparisons across both countries the cost index and number of houses built were defined on a per-capita basis. A Shapiro-Wilk test showed that the data could be considered normally distributed (Irish cost index per Capita: W=0.870, pvalue=0.298, Irish houses built per Capita: W=0.827, pvalue=0.161, Finnish cost index per Capita: W=0.763, pvalue=0.051, Finnish houses built per Capita: W=0.839, pvalue=0.193).

Table 3.1: *Results of Shapiro-Wilk tests for normality on the Irish and Finnish construction costs and housing units completed per-capita.*

Variable	W Test statistic	p-value
Irish cost index per Capita	0.870	0.298
Irish houses built per Capita	0.827	0.161
Finnish cost index per Capita	0.763	0.051
Finnish houses built per Capita	0.839	0.193

3.5 Machine learning

The irish data was not split into test and train sets as the dataset was too small.

In splitting the finnish data into test and train a split of 0.3 was taken, as the dataset was small, and it was important to have sufficient data to test with. However, with only ... records, the dataset was too small for meaningful analysis and in practice ...

Five newspaper articles on the subject of Finland's approach to the homelessness crises were compared

Media Bias / Fact Check was used to determine the political leanings of the various media outlets used (*Media Bias / Fact Check* 2023).

3.5.1 Machine learning methods used

Results

Conclusion

References

Finland's Zero Homeless Strategy: Lessons from a Success Story (2021). Ecoscope. URL: https://oecdecoscope.blog/2021/12/13/finlands-zero-homeless-strategy-lessons-from-a-success-story/comment-page-1/ (visited on 05/06/2023).

Department of Housing, Local Government, and Heritage (2023). https://data.gov.ie. URL: https://data.gov.ie/organization/department-of-housing-local-government-and-heritage (visited on 05/10/2023).

HSM09 - House Construction Cost Index 2023 (2023). https://data.gov.ie. URL: https://data.gov.ie/dataset/hsm09-house-construction-cost-index?package_type=dataset (visited on 05/06/2023).

12fy – Building and dwelling production, 1995M01-2023M02 (2023). Tilastokeskus. URL: https://pxdata.stat.fi/PxWeb/pxweb/en/StatFin/StatFin__ras/statfin_ras_pxt_12fy.px/ (visited on 05/10/2023).

11na – Building cost index by type of cost, annual data, 1990-2022 (2023). Tilastokeskus. URL: https://pxdata.stat.fi/PxWeb/pxweb/en/StatFin/StatFin__rki/statfin_rki_pxt_11na.px/ (visited on 05/10/2023).

Finnish model shows how a more radical approach could solve homeless problem (Oct. 25, 2018). The Irish Independent. URL: https://www.independent.ie/irish-news/finnish-model-shows-how-a-more-radical-approach-could-solve-homeless-problem/37456401.html (visited on 05/16/2023).

Homeless response 'should focus on needs of children' (May 9, 2023). The Irish Times. URL: https://www.irishtimes.com/ireland/social-affairs/2023/05/09/homeless-response-should-focus-on-needs-of-children/ (visited on 05/16/2023).

The government is trying to reduce chronic homelessness ... Here's how Finland ended it (Oct. 25, 2018). The Journal. URL: https://www.thejournal.ie/finland-homeless-housing-first-ireland-4303419-Oct2018/ (visited on 05/16/2023).

Homelessness can be eradicated by 2027 with close cooperation: Report (Feb. 10, 2023). The Helsinki Times. URL: https://www.helsinkitimes.fi/finland/finland-news/domestic/22934-homelessness-can-be-eradicated-by-2027-with-close-cooperation-report.html (visited on 05/16/2023).

Has Finland really solved homelessness? (Apr. 19, 2022). YLE. URL: https://yle.fi/a/3-12409059 (visited on 05/16/2023).

'It's a miracle': Helsinki's radical solution to homelessness (June 3, 2019). The Guardian. URL: https://www.theguardian.com/cities/2019/jun/03/its-a-miracle-helsinkis-radical-solution-to-homelessness (visited on 05/16/2023).

What can the UK learn from how Finland solved homelessness? (Mar. 22, 2017). The Guardian. URL: https://www.theguardian.com/housing-network/2017/mar/22/finland-solved-homelessness-eu-crisis-housing-first (visited on 05/24/2023).

To help the homeless, close a shelter (July 18, 2019). Politico. URL: https://www.politico.eu/article/to-help-the-homeless-helsinki-finland-close-a-shelter/ (visited on 05/24/2023).

London wants to eradicate homelessness. Here's how Finland is doing it (Jan. 28, 2023). CBC. URL: https://www.cbc.ca/news/canada/london/london-wants-to-eradicate-homelessness-here-s-how-finland-is-doing-it-1.6728398 (visited on 05/24/2023).

Housing is a human right: How Finland is eradicating homelessness (Aug. 19, 2020). CBC. URL: https://www.cbc.ca/radio/sunday/the-sunday-edition-for-january-26-2020-1.5429251/housing-is-a-human-right-how-finland-is-eradicating-homelessness-1.5437402 (visited on 05/24/2023).

How Finland managed to virtually end homelessness (Apr. 20, 2023). The Toronto Star. URL: https://www.thestar.com/opinion/contributors/2023/04/20/how-finland-managed-to-virtually-end-homelessness.html?rf (visited on 05/24/2023).

Finland's homeless crisis nearly solved. How? By giving homes to all who need. (Mar. 21, 2018). Christian Science Monitor. URL: https://www.csmonitor.com/World/Europe/2018/0321/Finland-s-homeless-crisis-nearly-solved.-How-By-giving-homes-to-all-who-need (visited on 05/24/2023).

Population on 1 January (2023). eurostat. URL: https://ec.europa.eu/eurostat/databrowser/view/TPS00001/default/table?lang=en (visited on 05/16/2023).

The fastest way to read a CSV in Pandas (2023). pythonspeed.com. URL: https://pythonspeed.com/articles/pandas-read-csv-fast/ (visited on 05/24/2023).

PxWeb API (2023). stat.fi. URL: https://pxdata.stat.fi/api1.html (visited on 05/10/2023).

pandas.read json (2023). NumFOCUS, Inc. URL: https://pandas.pydata.org/docs/reference/api/pandas.read_json.html (visited on 05/24/2023).

(pandas.json normalize) (2023). NumFOCUS, Inc. URL: https://pandas.pydata.org/docs/reference/api/pandas.json_normalize.html (visited on 05/24/2023).

What Is ARIMA Modeling? (2023). Master's in Data Science. URL: https://www.mastersindatascience.org/learning/statistics-data-science/what-is-arima-modeling/ (visited on 05/11/2023).

Augmented Dickey-Fuller Test in Python (With Example) (2023). Statology. URL: https://www.statology.org/dickey-fuller-test-python/ (visited on 05/11/2023).

Raval, P. (2023). *How to Build ARIMA Model in Python for time series forecasting?* ProjectPro. URL: https://www.projectpro.io/article/how-to-build-arima-model-in-python/544 (visited on 05/11/2023).

Nadeem (2021). *ARIMA: Advanced Time Series Methods: Auto Regression Integrated Moving Average.* Medium. URL: https://medium.com/analytics-vidhya/arima-fc1f962c22d4 (visited on 05/11/2023).

How to do cross validation for time series? (2023). ProjectPro. URL: https://www.projectpro.io/recipes/do-cross-validation-for-time-series (visited on 05/11/2023).

Shrivastava, S. (2020). *Cross Validation in Time Series*. Medium. URL: https://medium.com/@soumyachess1496/cross-validation-in-time-series-566ae4981ce4 (visited on 05/11/2023).

Hyndman, R. J. and G. Athanasopoulos (2018). *Forecasting: Principles and Practice*. Melbourne, Australia: OTexts.

Media Bias / Fact Check (2023). https://mediabiasfactcheck.com/. URL: https://mediabiasfactcheck.com/ (visited on 05/10/2023).