Treyce Kangas-Simon

Jean-Marc Romain

Georgette Powley

Elias Pennell

NTEC365 46182

25 October 2022

**Project Proposal**

Our team proposes to design a project that is two-fold in nature. We intend to follow the full life cycle of data, from generation through interpretation. The project will employ the techniques of Python automated web scraping in conjunction with traditional data analysis and correlation calculation. We will be using the generated data to draw a comparison between listed home pricing and crime statistics for Portland, OR. Specifically, we will be looking to see if there's any correlation between how a house is priced and how much crime has been committed in the nearby vicinity.

Zillow or Realtor (Realtor is easier in term of design and finding the variables) will play an integral role in this project. It will be the primary resource that generates data of house addresses and prices. A web scraping program will be written in Python that will scrape the necessary data from the website. This will be the data collection and processing steps in data life cycle. Additionally, the web scraping program will assist in data storage and management as it will auto-write into .csv files.

Our crime statistics will be gathered from City of Portland’s crime data, provided through the Portland Police Bureau’s “Crime Dashboard”. The plan is to operate basic statistics with this data to answer some questions. Questions such as, where is the crime the more prominent? Does it reflect on the price of houses in the same location? Is it only violent crime or all crime? Can we fine tune it for burglaries vs car jackings vs murder? We believe every crime has a negative impact but without the basic statistics, we don’t know for sure.

These two data sets will be analyzed in order to find if there is a correlation between crimes committed and housing prices. The analysis will seek to determine if there is any linear correlation between home price and committed crime using the Pearson correlation coefficient. This data analysis can be further visualized using simple graphing techniques. We would like to go a step further and actually map the data that is generated and will do so if time permits. Finally, the generated data will be interpreted to finalize and succinctly clarify if any correlation was found between the two data sets.

**References**

Ceccato, V., & Wilhelmsson, M. (2020). Do crime hot spots affect housing prices? *Nordic Journal of Criminology*, *21*(1), 84–102. <https://doi.org/10.1080/2578983x.2019.1662595>

GeeksforGeeks. “How to Scrape Multiple Pages of a Website Using Python?” *GeeksforGeeks*, 30 Nov. 2021, [www.geeksforgeeks.org/how-to-scrape-multiple-pages-of-a-website-using-python](http://www.geeksforgeeks.org/how-to-scrape-multiple-pages-of-a-website-using-python).

*How to Solve 403 Forbidden Errors When Web Scraping | ScrapeOps*. scrapeops.io/web-scraping-playbook/403-forbidden-error-web-scraping. Accessed 24 Oct. 2022.

Lynch, A. K., & Rasmussen, D. W. (2001). Measuring the impact of crime on house prices. *Applied Economics*, *33*(15), 1981–1989. [doi.org/10.1080/00036840110021735](https://doi.org/10.1080/00036840110021735). Retrieved October 24, 2022, from [clarkcollege.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=5563101](https://clarkcollege.idm.oclc.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=sih&AN=5563101&site=ehost-live&scope=site)

Munzert, Simon., and Simon Munzert. Automated Data Collection with R : a Practical Guide to Web Scraping and Text Mining. Chichester, England: Wiley, 2015. Print.

“Portland, OR Real Estate - Portland Homes for Sale.” *realtor.com®*, 25 Oct. 2022, [www.realtor.com/realestateandhomes-search/Portland\_OR](http://www.realtor.com/realestateandhomes-search/Portland_OR)

Portland Police Bureau. (n.d.). *Crime statistics*. Crime Statistics RSS. Retrieved October 24, 2022, from [www.portlandoregon.gov/police/71978](http://www.portlandoregon.gov/police/71978)

*ScrapeOps*. scrapeops.io/app/proxy. Accessed 24 Oct. 2022.

Taylor, R. B. (1995). The impact of crime on communities. *The Annals of the American Academy of Political and Social Science*, *539*(1), 28–45. [doi.org/10.1177/0002716295539001003](https://doi.org/10.1177/0002716295539001003). Retrieved October 25, 2022, from [www.researchgate.net/publication/249666367\_The\_Impact\_of\_Crime\_on\_Communities](http://www.researchgate.net/publication/249666367_The_Impact_of_Crime_on_Communities)

Vanden Broucke, Seppe., and Bart. Baesens. Practical Web Scraping for Data Science Best Practices and Examples with Python. 1st ed. 2018. Berkeley, CA: Apress, 2018. Web.