FACULTY OF SCIENCE & TECHNOLOGY

BSc (Hons) Forensic Computing and Security

May 2021

An explorative study into vulnerabilities in countries and smart, circular and plain cities.

by

Paul-David Jarvis

Bournemouth University, Department of Computing and Informatics, Final Year Project

Faculty of Science & Technology

Department of Computing and Informatics

Final Year Project

Abstract

Over half of the world's population will be living in cities in the near future and research shows this will cause problems across the board in many sectors including the draining of the city's finite resources as well as an increase cyber attack surface. Many initiates have been taken to mitigate these two and other inevitable problems and cities are now adopting an CE agenda or a smart city approach. This study looks at the recent threat landscape and attempts to provide some context when it comes to cities and their attack surface with an ultimate goal of providing accurate and valuable actionable threat intelligence. With the use of automation, we scrape data from Shodan.io and combine with secondary research to provide contextualization. Our most significant finding is cities that have adopted a CE or smart approach, have a wider attack surface and have a larger exposure profile than cities that haven't made the transition.

Dissertation Declaration

I agree that, should the University wish to retain it for reference purposes, a copy of my dissertation may be held by Bournemouth University normally for a period of 3 academic years. I understand that once the retention period has expired my dissertation will be destroyed.

Confidentiality

I confirm that this dissertation does not contain information of a commercial or confidential nature or include personal information other than that which would normally be in the public domain unless the relevant permissions have been obtained. In particular any information which identifies a particular individual's religious or political beliefs, information relating to their health, ethnicity, criminal history or sex life has been anonymised unless permission has been granted for its publication from the person to whom it relates.

Copyright

The copyright for this dissertation remains with me.

Requests for Information

I agree that this dissertation may be made available as the result of a request for information under the Freedom of Information Act.

p.jarvis

Signed: ______________________________

Name: Paul-David Jarvis

Date: 01/05/2021

Programme: BSc (Hons) Forensic Computing and Security

Original Work Declaration

This dissertation and the project that it is based on are my own work, except where stated, in accordance with University regulations.

p.jarvis

Signed: ______________________________

Name: Paul-David Jarvis

Date: 01/05/2021

Acknowledgements

The first person who I would like to acknowledge and give my gratitude is to Professor Katos who has supported and offered his guidance not only as a supervisor for this project but also as an academic advisor during my first year at Bournemouth University and who offered me the opportunities which I never thought I would get; most notability involved me in DTRAP. I have found a side passion in the security of IoT devices and I hope to work together again on many other publications. I would also like to acknowledge my family who have offered me support in every way possible and the many astonishing and lifelong friends and people I have met during university, especially Carlo, who ended up supporting me in a way that I would not have believed possible.

Publications

It should be noted that some of the work featured in this project was also used in an academic paper submitted to Digital Threats: Research and Practice (DTRAP) in the ACM Digital Library. The paper submitted labelled ”Vulnerability Exposure Driven Intelligence in Smart, Circular Cities” was written by Jarvis, P. and Katos, V. and Damianou, A. and later joined by Ciobanu, C. from ENISA. Unfortunately the paper was returned by the editor in chief from DTRAP with major revisions required before re-submission.

Contents

List of Figures

List of Tables

Abbreviation	Meaning
APT	Advanced Persistent Threats
CVE	Common Vulnerabilities and Exposures
CPE	Common Platform Enumerations
CWE	Common Weakness Enumeration
CAPEC	Common Attack Pattern Enumeration and Classification
CVSS	Common Vulnerability Scoring System
CIMI	Cities in Motion Index
CLI	Command Line Interface
DoS	Denial of Service
ENISA	European Union Agency For CyberSecurity
IoT	Internet of Things
IDEAL-CITIES	Intelligence-Driven Urban Internet-of-Things Ecosystems for Circular, Safe and Inclusive Smart CITIES
IMD	Institute for Management Development
IESE	The Business School of University of Navarra
ISC	Impact Sub-Score
ICS	Industrial Control System
WDCI	World Digital Competitiveness Index
WCSS	within-cluster sum of squares
NIST	National Institute of Standards and Technology
NVD	National Vulnerability Database
NCSC	National Cyber Security Centre
SUTD	Singapore University for Technology and Design
SCI	Smart City Index
XSS	Cross-Site Scripting

Table 1: Table of Abbreviations

1 INTRODUCTION & MOTIVATION

It is predicted that 68% of the world population will live in cities by 2050 (Nations 2018) and this introduces a much bigger challenge on cities and governments to manage their resources efficiently. As it stands, cities consume over 75% of natural resources, produce over 50% of global waste, and emit between 60-80% of greenhouse gases (Foundation 2017). To address the accelerating depletion of infinite resources, cities are adopting a diversity of approaches across various dimensions, specifically technological, which introduces IT and network systems interconnected to achieve sustainability. The Cities' starting points to achieve sustainability may differ, but in all cases, the end goal is to leverage technological enablers such as 5G and Software Defined Networks, IoT, and Industrial Systems in general, as well as the adoption of cloud and edge computing paradigms.

The term smart city is continuously evolving, however, it is clear that being ”smart” requires a city to have a certain intellectual ability that addresses several innovative socio-technical and socio-economic aspects of growth (Zaharis (nd)). These aspects lead to smart city conceptions as “green” referring to urban infrastructure for environmental protection and reduction of CO2 emission, “interconnected” related to the revolution of broadband economy, “intelligent” declaring the capacity to produce added value information from the processing of city's real-time data from sensors and activators, whereas the terms “innovating”, “knowledge” cities interchangeably refer to the city's ability to raise innovation based on knowledgeable and creative human capital (Zygiaris (2012)).

The term Circular Economy is used to describe the economic model where everything has value and nothing is wasted. This means that assets like materials, devices, services and general resources are not disposed of after the first usage, but, they are used again and again for a variety of purposes. The main goal of a Circular Economy is to maintain its utility without producing new assets or wasting them before the end of its life cycle. CE has recently been added to the agenda of city officials and the Danish government using platforms like Gate21 that provides information about the living labs of the city. Furthermore, the CE agenda has been embedded in primary and secondary schools programs, and in refresher and vocational training in order to emphasise the importance of it as a significant part of citizens' life (Intelligence-Driven Urban Internet-of-Things Ecosystems for Circular and CITIES (2019)).

Copenhagen is a great European city that has adopted a CE approach. The city has received numerous awards and recognitions for one of its greatest achievements in its reduction of carbon emissions and transforming its city into the first-ever city that will be carbon neutral until 2025 globally (Intelligence-Driven Urban Internet-of-Things Ecosystems for Circular and CITIES (2019)).

A data-driven Circular Economy is the utilisation of reactive, adaptive, autonomous, or collaborative objects and systems for economic and environmental value creation and appropriation through closing material and energy loops, minimizing natural resource depletion, and restoring natural biospheric balances (Langley et al. (2021)).

Full sustainability and cities becoming more data-driven is a work in progress, and cities that adopt a smart approach are in-between to the end goal of full sustainability which is now often thought to be delivered through the concept of circular economy (Intelligence-Driven Urban Internet-of-Things Ecosystems for Circular and CITIES (nd). However, now that more cities are adopting a more smart approach to be data-driven and adopt a circular economy agenda, this would inevitably introduce significant cyber security risks if a city has not implemented a suitable risk mitigation plan. The aforementioned cities to be more data-driven, smart, and circular means that they become more dependent on the cyber-physical and socio-technical systems that they have implemented, and this means that risk can occur between cyberspace and physical space. The IoT devices like sensors which are implemented onto the physical infrastructure that allows the communication between two nodes are at risk as well as risk from the human nodes. The more systems a city implements, the wider their potential attack surface increases and the ripple effect of this is that the impact scope is increased and can affect the information security triad; confidentiality, integrity, and availability as well as the privacy and safety of the citizens.

To the best of our knowledge, studies that focus on vulnerabilities in smart cities are predominantly theoretical, so the purpose of this work is to explore empirically that which has only been theoretical. We focus on applying context, common vulnerabilities and exposures in various geographic areas linked to specific cities and countries, and additional relevant data. It is clear that cities that plan on becoming smart or circular cities need to sustain an adequate plan as not doing so can result in threats to not only their cities and infrastructure but their citizens.

This work looked at over 250 cities which contained a mix of smart cities from the SCI and CIMI, circular cities from C40, and plain cities that haven't adopted a smart or circular agenda and over 200 different countries to explore their exposure to common vulnerabilities. These classes will be defined and touched upon later on in Chapter 4.2.

1.1 Hypotheses

Throughout this project, we developed a variety of hypotheses and evolve current ones with the more data we collected. We believed that more populated cities like Singapore which has a 2020 population of over five and a half million (Nations 2020) would have more IoT enabled devices than a city with less than a million citizens and would mean that they're more likely vulnerable to hacking and CVE assigned security flaws; cities that are smarter and circular that depends on technology to make their citizens' life more efficient would be more vulnerable to hacking or CVE assigned security flaws than cities that haven't taken that initiative yet; Cities and countries scored higher in indexes like the SCI from the IMD or the CIMI from the IESE meaning they're more advanced than other cities scored lower would be more vulnerable as they have implemented more technology, however, it could also be suggested because these cities scored higher in the indexes should mean they're less vulnerable due to their more advanced nature.

To follow up on this we look at the statistics we scraped from our sources to either confirm or reject our hypotheses. We look at the CIMI factors like `Economy`, `Human Captial`, `Social Cohesion`, `Technology`, and others. We also look at the SCI for their overall ratings and questions asking the citizens about their perception of smartness within their city. We performed factor analysis like Cronbach's Alpha and K-Means clustering. Each hypothesis previously mentioned will be looked at more statistically and scientifically.

1.2 Aims and objectives

This project aims to conduct an empirical evaluation of the cities exposure to software vulnerabilities. As such, this project came with a series of technical challenges which were mapped to the following project objectives:

- Collect enough data to allow the analysis of normal, smart, and circular cities and their exposure to vulnerabilities. This would require the collection of primary vulnerability data and secondary data from independent and published research capturing the smart city parameters.
- Develop several bash scripts that implement Shodan's CLI to automate the process of scraping and collecting data for specific queries.
- Develop a series of Jupypter notebooks to investigate the developed hypotheses.

The process of analysing the primary data would follow a two-phased approach. First, an initial Exploratory Data Analysis would be conducted to get a feel of the domain. Second, a deeper analysis will be conducted. In this stage, the correlations and associations between the vulnerabilities and smart cities data sets are studied.

1.3 Success criteria for each objective

- The scripts that were written in bash using Shodan's CLI must compile correctly and return the accurate data. We are mainly looking for numbers greater than 0 as this provides us with the means to analyse the specific data, however, we are not writing out data that returned 0 as it could still provide valuable information and maintain unbiasedness.
- The aforementioned scripts must return a good amount of data which allows us to have a variety of analyses. This isn't a worry as we will have over 200,000 individual pieces of data in our data sets.
- We have to create Jupyter notebooks with a satisfying amount of python code dedicated to statistical analysis and produces valuable results, plots, and graphs that provide an insight into vulnerabilities in a city or country.

1.4 Methodology Approach

Chapter 3 is where we go into detail about the methodology approach taken throughout this project. We collected both primary and secondary research from various sources. Our primary research came from Shodan.io, and we go into extreme detail on how and what was collected from their search engine and our second data is constructed for several indexes like the CIMI from IESE and the SCI from IMD. All this data will be combined to understand exposure and correlate together and validate.

1.5 Overview of the remaining chapters

Chapter 2 is the literature review is where we study and evaluate the already available literature that we used for this project, this includes information regarding vulnerabilities, information regarding cities, and information regarding countries. Chapter 3 is the methodology, in this chapter we describe the proposed approach and then discuss the data sets we used. Chapter 4 includes our analysis and discussion on the findings. Chapter 5 summarizes the main findings and outlines areas for future research.

2 LITERATURE REVIEW

2.1 Vulnerabilities

2.1.1 Shodan

Shodan is an online search engine created by John Matherly that indexes devices like routers, computers, and servers that are connected to the internet (Shodan 2013). Shodan doesn't only index these devices. We can also use queries to geolocate the IP addresses of the indexed devices, which means we can provide an overview of countries and cities that are more vulnerable than others. However, the accuracy of geolocation is limited and bounded by Shodan's geo-mapping process. Shodan also specifies if a device it has indexed is vulnerable to a particular vulnerability like ”MS17-010” which would allow local authority and governments of cities that may be adopting a smart or circular city approach to gather actionable threat intelligence and can provide help in conducting threat assessments. Knowing this information can also help cities understand the threat landscape and the specific threats that threaten them and potentially open up collaboration efforts.

2.1.2 ENISA

ENISA is an EU agency dedicated to achieving a high common level of cybersecurity across Europe (for Cybersecurity 2005). They have produced a report and a data set regarding the state of vulnerabilities in the year 2018 to 2019 which we have used, improved and expanded. Although ENISA's data set contains information originally sourced from Shodan, our improved and expanded data set contains individual exposure information on devices and their potential vulnerability to a particular CVE or Metasploit module as well as the cities and countries that are vulnerable to a specific vulnerability.

2.1.3 MITRE CVE

CVE has publicly disclosed security flaws discovered by individual researchers or companies. These discovered security flaws are assigned a CVE ID which allows them to be easily identifiable through all external resources that discuss a particular CVE like the NVD and other third-party websites like Red Hat or Oracle. A brief description is also included with every CVE as well as references to the aforementioned third-party websites. The limit with CVE is that these are the only three things that they share with the user so we have to combine this information with another source like the NVD.

2.1.4 NIST NVD

NVD is a U.S. government repository of vulnerabilities (MITRE (nde)) and includes more technical information about a particular CVE. The NVD shares the scores from the CVSS calculated by the CVE attack vector, attack complexity, privileges required, user interaction, scope, confidentiality, integrity, and availability. They also share the CVE's CWE which is what flaw or weakness the CVE uses and its CPE which is the known affected software.

The Base Score is a function of the Impact and Exploitability sub score equations. Where the base score is defined below (MITRE (ndc)).

$$If(Impact\ sub\ score <= 0)\ 0\ else,$$

$$scopeunchanged\ Roundup(Minimum[(Impact + Exploitability),\ 10])$$

$$scopechanged\ Roundup(Minimum[1.08\ \times\ (Impact + Exploitability),\ 10])$$

and the Impact Sub Score (ISC) is defined as,

$$Scope\ Unchanged\ 6.42\ \times\ ISC_{Base}$$

$$Scope\ Changed\ 7.52\ \times [ISC_{Base}\ -\ 0.029]\ -\ 3.25\ \times\ [ISC_{Base}\ -\ 0.02]^{15}$$

Where,

$$ISC_{Base}\ =\ 1 - [(1\ - Impact_{Conf}) \times AttackComplexity \times PrivilegeRequired \times UserInteraction$$

And the Exploitability sub score is,

$$8.22\ \times\ AttackVector\ \times\ AttackComplexity\ \times\ PrivilegeRequired\ \times\ UserInteraction$$

It should be noted that CVSS is founded on widely accepted methodologies and considered a good and trustworthy method to classify vulnerabilities. However, the data quality in these databases has received some criticism, and alternative ways to calculate vulnerability scores have been suggested. A Bayesian investigation was carried out on CVSS and performed well for confidentiality, integrity and availability. However, it performed poorer when attempting to classify vulnerabilities with a low access complexity and vulnerabilities requiring multiple authentications (Johnson et al. (2016).

2.1.5 MITRE CWE

CVE is a community-developed list of common software and hardware weakness types that have potential security ramifications. These weaknesses are flaws, faults, bugs, vulnerabilities, or other errors in software or hardware implementation, code, design, or architecture. If these weaknesses are left unaddressed, they could result in systems, networks, or hardware being vulnerable to attackers. The CWE List and associated classification taxonomy serve as a language that can be used to identify and describe these weaknesses in terms of CWEs. The main goal of CWE is to stop vulnerabilities at the source by educating software and hardware, architects, designers, programmers, and acquires on how to eliminate the most common mistakes before software and hardware are delivered. Ultimately, using CWE helps prevent the kinds of security vulnerabilities that have plagued the software and hardware industries and put enterprises at risk (MITRE ndd).

2.1.6 MITRE CAPEC

CAPEC attempts to provide a publicly available catalogue of common attack patterns that helps users understand how adversaries exploit weaknesses in applications and other cyber-enabled capabilities.

”Attack Patterns” are descriptions of the common attributes and approaches employed by adversaries to exploit known weaknesses in cyber-enabled capabilities. Attack patterns define the challenges that an adversary may face and how they go about solving them. They derive from the concept of design patterns applied in a destructive rather than constructive context and are generated from in-depth analysis of specific real-world exploit examples.

Each attack pattern captures knowledge about how specific parts of an attack are designed and executed and gives guidance on ways to mitigate the attack’s effectiveness. Attack patterns help those developing applications or administrating cyber-enabled capabilities to better understand the specific elements of an attack and how to stop them from succeeding (MITRE nda).

2.1.7 MITRE ATT&CK

ATT&CK is a globally accessible knowledge base of adversary tactics and techniques from real-world observations. The ATT&CK knowledge base is used as a foundation for the development of specific threat models and methodologies in the private sector, in government, and the cybersecurity product and service community. One of the main goals of ATT&CK is to bring communities together to develop more effective cybersecurity (MITRE 2015).

2.2 Cities

2.2.1 Ideal Cities

IDEAL-CITIES is a project funded by the Marie Skłodowska-Curie RISE action to improve the well-being and inclusively of citizens (especially vulnerable citizen groups), produce a more effective response to crime or other emergencies, and make smart cities feel more secure and safe to the citizens living in them (Intelligence-Driven Urban Internet-of-Things Ecosystems for Circular and CITIES nd).

Maturity models are one of the widespread areas in the field of improving organisational performance. They identify organizational strengths and weaknesses as well as providing bench marking information. There are many maturity models like OPM3, CMMI, P3M3, PRINCE, BPMM, and Kerzner's project management maturity model. These models may differ in terms of factors, number of levels as well as application domains. (Khoshgoftar and Osman (2009)).

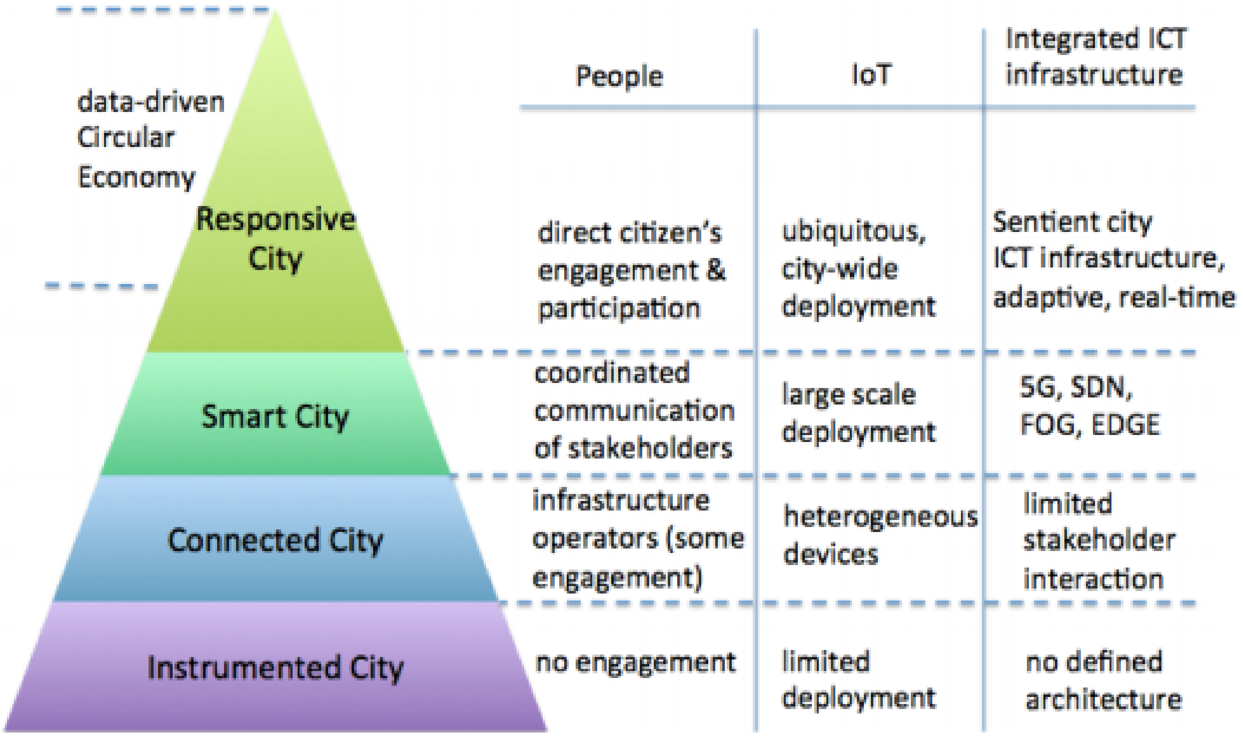
When looking at how smart a city is, we used a maturity model described in a report released by Ideal Cities called "D2.1: Circular Economy models for smart city assets". This maturity model will help in establishing the stage(s) or CE readiness of a city against the set of agreed dimensions a smart city is composed of (Intelligence-Driven Urban Internet-of-Things Ecosystems for Circular and CITIES 2019).

The maturity model in Figure 1 outlines the following stages that a city takes to make the transition from an instrumented city to a responsive city.

The first stage is the instrumented city, where the city embeds constellations of sensors and devices on the physical infrastructure (e.g. bridges, street lights, gas pipes, the grid). During this stage, these devices perform basic tasks for a specific example. These devices are also not able to communicate with any other devices built for a different purpose. The intelligence gathered by stakeholders would be used primarily as an advice source.

The second stage is the connected city which is where connectors will be installed between the different constellations. This added capability doesn't necessarily involve any actual exploitation or systematic utilization of the available data. The stakeholders are still concentrated in consuming the data within their respective sectors or domains, despite the increased heterogeneity and availability of the data.

The third stage is the smart city which all assets are fully interconnected and actionable information is available to the stakeholders which allows them to reach high levels of situational awareness. Infrastructure operators, utility, and service providers are also known as the "back office" opera-



	People	IoT	Integrated ICT infrastructure
Responsive City	direct citizen's engagement & participation	ubiquitous, city-wide deployment	Sentient city ICT infrastructure, adaptive, real-time
Smart City	coordinated communication of stakeholders	large scale deployment	5G, SDN, FOG, EDGE
Connected City	infrastructure operators (some engagement)	heterogeneous devices	limited stakeholder interaction
Instrumented City	no engagement	limited deployment	no defined architecture

Figure 1: A maturity model for a smart, circular city (Sourced from Intelligence-Driven Urban Internet-of-Things Ecosystems for Circular and CITIES (2019))

tions where almost all activities are performed, can perform intelligent processing, and the local government stakeholders can perform global processing and maintain an overview of the city's operations.

Stage four is the responsive city stage which is where everything prior comes together and the citizens, intelligent assets, and the rest of the components are in complete sync. Information gathered is available in real-time and in an appropriate and accessible format. These city's assets and infrastructure can dynamically adjust to cater to the short-term and ad-hoc needs of its citizens by collaborative decision-making. With the use of predictive analytic and proactive computing, the city can anticipate changes and respond. The goal of a responsive city is to reach the ultimate level of sustainability by ensuring that finite resources are available when needed and waste is virtually eliminated.

2.2.2 IMD Smart City Index 2020

IMD, in collaboration with SUTD, has released the 2020 SCI which takes a look at 109 cities and ranks them based on their economic and technological data, as well as how their citizens' perceptions of how ”smart” their cities are (IMD 2020c). The index shows the results from a survey conducted on citizens from these specific cities and shows their attitudes toward personal data being used to improve their cities, potentially invasive technology like facial recognition to lower crime, their trust in authorities, and their use of technology for day-to-day transactions. In addition to showing the attitudes of citizens, they also released survey answers for five key areas: health and safety, mobility, activities, opportunities, and governance on structure and technology (IMD 2020b).

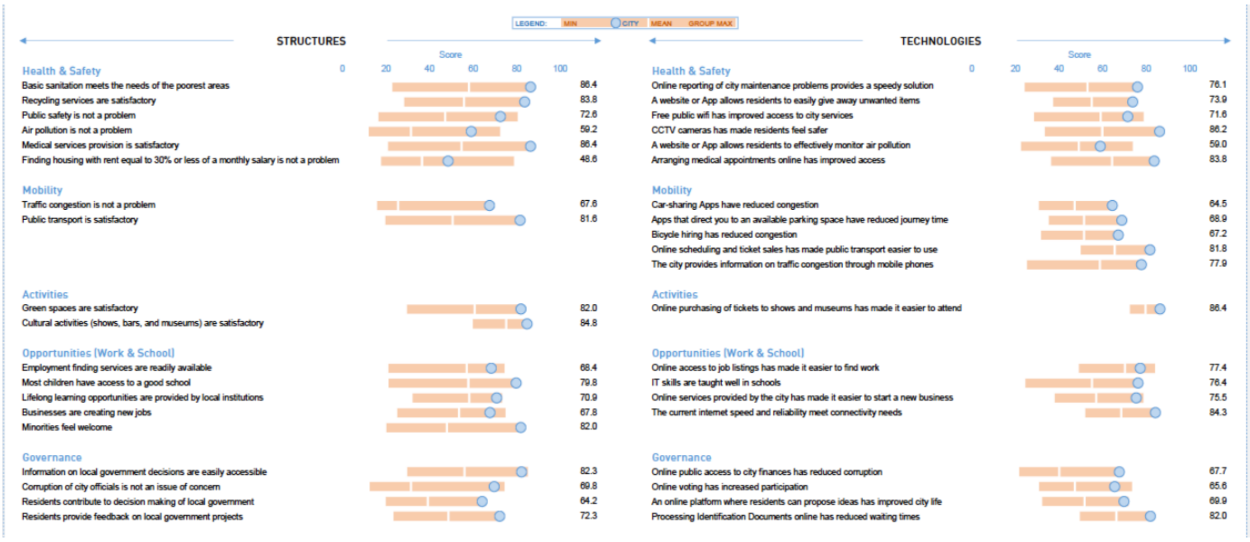


Figure 2: Structures and Technology (Sourced from IMD (2020b))

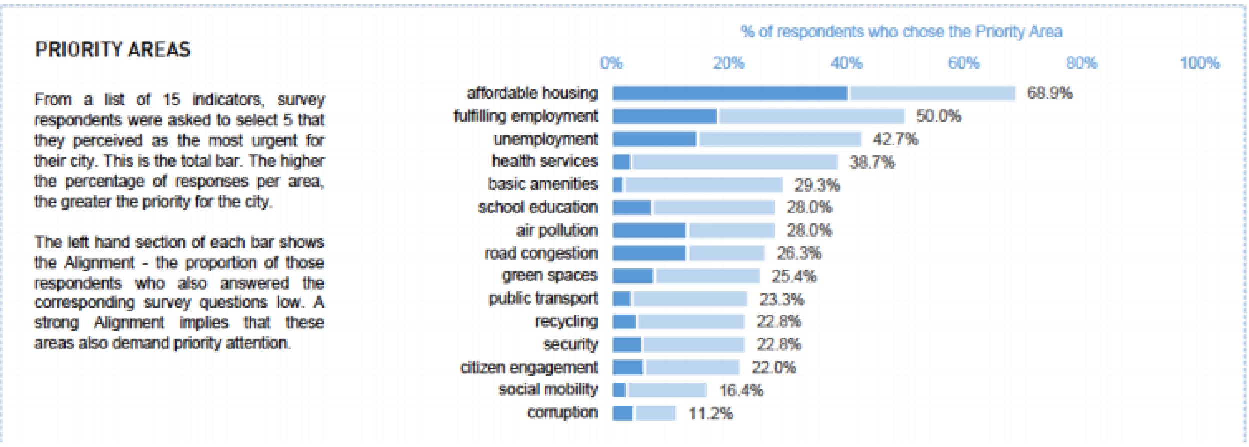


Figure 3: Priority Areas (Sourced from IMD (2020b))

2.2.3 NCSC

Another important piece of literature from a credible source was an article released recently from the NCSC where they offer guidance to try and help authorities build awareness and understanding of the security considerations needed to design, build, and manage their smart cities. More specifically, it recommends a set of cyber security principles that will help ensure the security of a smart city and its underlying infrastructure, so that it is both more resilient to cyber attack and easier to manage (Centre (2020)).

They offered three important principles which were understanding, designing and managing a smart city as well as offering several more precise sub principles for each main principle. Our project mainly collided with principles from designing your smart city. They specifically mentioned the ATT&CK framework from MITRE, one that we used to classify our types of vulnerabilities. Sub principle 7 is all about reducing exposure where they offer guidance on what authorities should do such as removing default credentials, switching off unnecessary ports or services, and limiting the places where people are allowed to access the system. These are interesting principles as throughout this project we found that many of these cities had critical infrastructure and sensors with default credentials and some times no credentials, unnecessary ports open leading to nothing and we were able to access the critical infrastructure and sensors from a remote location over Shodan.io if we wanted to.

2.2.4 IESE Cities in Motion Index

IESE publishes annually an index that aims to evaluate the development of a substantial group of cities. This year is noteworthy as this index was published during the COVID-19 health crisis which allows us to comprehend how a pandemic affects the cities studied. This year's index analysed 101 indicators across 9 key dimensions: human capital, social cohesion, the economy, governance, the environment, mobility and transportation, urban planning, international projection, and technology which gave a comprehensive view of the cities indexed (School 2020).

2.2.5 C40 Cities

C40 Cities is a network of the world's megacities that come together to address the ongoing threat of climate change. There are 97 of the world's greatest cities and over 700 million citizens that work together to take action, leading the way towards a healthier and more sustainable future. This initiate has done a lot to improve the city's use of resources and has seen more than 66,000 electric buses on the streets and 82 of the 97 cities have implemented cycle hire schemes (Cities 2021).

2.3 Countries

2.3.1 IMD World Digital Competitiveness

The IMD has also observed 63 countries and has measured their capacity and readiness to adopt and explore digital technologies as a key driver for economic transformation in business, government, and wider society (IMD 2020a). The report looks at three factors Knowledge, which captures the intangible infrastructure necessary for the learning and discovery dimensions of technology; Technology, which quantifies the landscape of developing digital technologies; and Future Readiness, which examines the level of preparedness of an economy to assume its digital transformation (IMD 2021).

3 METHODOLOGY

3.1 Data Gathering

3.1.1 MITRE CVE, CWE, CAPEC and ATT&CK

Almost all our data regarding vulnerabilities came from MITRE, which is an organization attempting to solve problems in several sectors to make the world safer.

Our list of CVE was sourced from the CVE website. The list used contained over 200,000 security flaws dating from 1999 to 2021. However, this project is only looking at the recent threat landscape which means we are discarding the security flaws before 2017. We are left now with 100,000 security flaws which will be checked through Shodan's CLI to identify if any has been indexed by a shodan attached to an IoT device. Once that is done, we are left with our official CVE list which will be used in our scripts to check our other data.

Once we pass all our vulnerabilities through the NVD we can now scrape from the CWE website which is all about the common weaknesses that are associated with vulnerabilities. One piece of information that we scraped from the CWE website is the CAPECs linked to our CVEs. The CAPECs data scraping involves the common attack patterns as well as ATT&CK tactics, techniques, and sub techniques linked to the CAPECs.

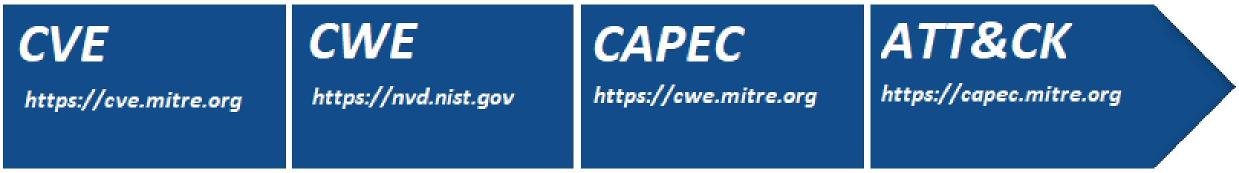


Figure 4: Vulnerabilities Linking Methodology

3.1.2 NVD

The information scraped from the NVD about our specific CVEs includes their CVSS scores and severity, CPE, and CWE. The linking of this data means we can now conclude on how severe a vulnerability, the vulnerable platforms, and the weaknesses our vulnerabilities exploit.

Severity	Base Score Range
CVSS2	
Low	0.0 - 3.9
Medium	4.0 - 6.9
High	7.0 - 10.0
CVSS3	
None	0.0
Low	0.1 - 3.9
Medium	4.0 - 6.9
High	7.0 - 8.9
Critical	9.0 - 10.0

Table 3.1: CVSS v2.0 & v3.0 Rankings

3.1.3 Shodan.io

We can use Shodan and their CLI feature to automate the scraping of all data. These data can be combined which each other in certain ways to provide a much better look at the state of vulnerabilities. We can identify web servers that are vulnerable to a specific vulnerability by extracting that data from the port data set; we can combine this with the product data set and identify how many of those exposed web servers are running Apache httpd or LightSpeed httpd. See Appendix D for the scripts used.

The following data was collected:

- **Operating Systems (OS):** The host OS for the device that was vulnerable to a particular CVE was collected. E.g. Windows 10 Home 18362 or Windows 8.1 Pro 9600.
- **City:** The city that the device which was vulnerable to a particular vulnerability was geolocated to. E.g. Singapore or London.
- **Country:** The country that the device which was vulnerable to a particular vulnerability was geolocated to. E.g. Greece or the United Kingdom.
- **Port:** The number of exposed devices running with a particular port open that is vulnerable to a particular vulnerability. E.g. There were just under 5 million devices with port 22 open vulnerable to CVE-2018-15919 captured in February 2021.
- **Device:** The number of exposed devices that have a particular device associated with it by Shodan vulnerable to a particular vulnerability. E.g. There were just over a thousand devices

assigned ”firewall” by Shodan vulnerable to CVE-2017-15906 captured in February 2021.

- **Tag:** The number of exposed devices that have a particular tag associated with it by Shodan vulnerable to a particular vulnerability. E.g. There are 47 million devices that have the tag ”cloud” vulnerable to CVE-2017-1000369 captured in November 2020.
- **Product:** The number of devices running a particular product exposed to a particular vulnerability. E.g. There was a total of 130 million devices running ”Apache httpd” captured in February 2021.
- **Protocol:** The number of ICS exposed to a particular vulnerability and running a particular protocol.

3.1.4 IESE Cities in Motion index

We collected plenty of interesting statistics from the CIMI like their overall ranking and ratings in several sectors like economy, technology, governance, and environment. This data will allow us to draw conclusions and determine whether a more advanced city like London which was considered the top city and scored higher in everything is more vulnerable than cities like Caracas in Venezuela which scored the lowest.

3.1.5 IMD Smart City Index 2020

The SCI provided us with the population of every city it indexed alongside overall ratings and rankings on how smart they are and survey data that shows the percentage of respondents who agree with four questions regarding the willingness to concede personal data, whether they feel comfortable with face recognition, and whether online information has increased trust in authorities.

3.1.6 IMD World Digital Competitiveness

IMD also provided insight into our list of countries with their IMD WDCI report that reported on how advanced they are and how future-readiness they are. They provided statistics on technology, science, training and education, and more. When possible, we collected population data from the United Nations to ensure accuracy and a credible source. We also attempt to collect data on their Gini coefficient to determine if more wealth is a factor.

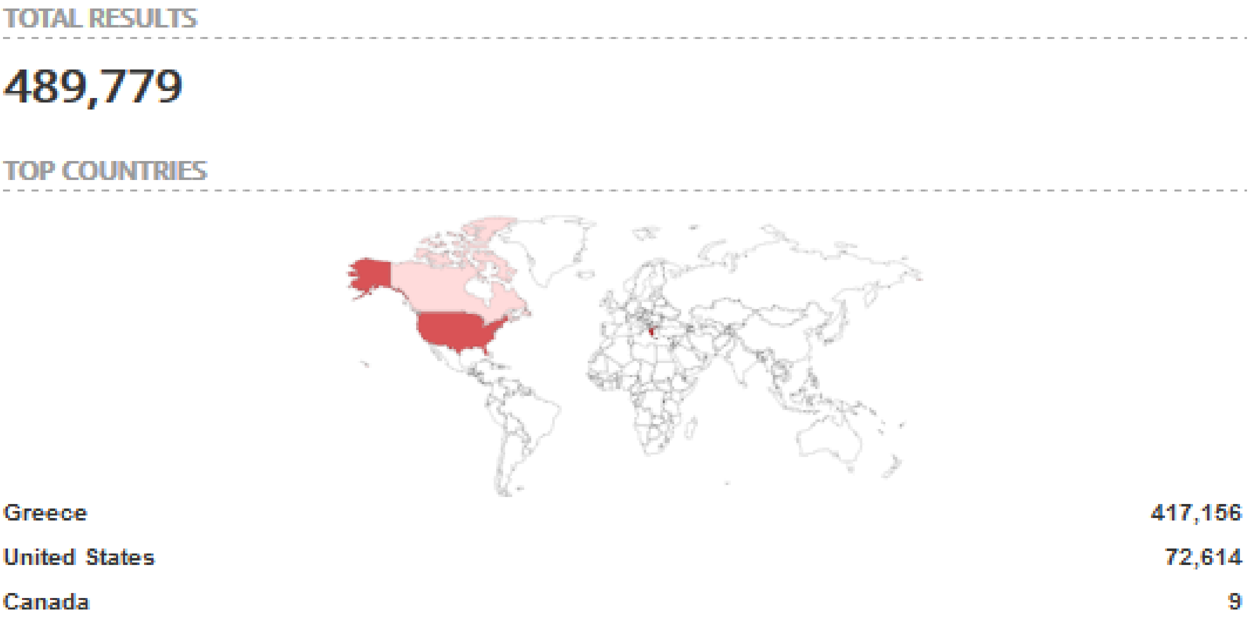
3.2 Bash Script Creation

Bash is a Unix shell and command language on Linux operating systems. Shodan CLI allows users to develop scripts in bash language that allows the automation of scraping and collection of results from Shodan.

The scripts that were created to automatically scrape the data contained several queries needed to collect the data we wanted. These four queries were the most important. We also had several other queries to retrieve port, device type, product type and what protocol is running.

- **vuln:** This query allowed us to request devices vulnerable to a given vulnerability such as MS17-010 or CVE-2017-15906.
- **count:** This query allowed us to request the total amount of devices. In November 2020, 7,647,885 devices are vulnerable to CVE-2017-15906.
- **country:** This query allowed us to request the number of devices from a specific country.
- **city:** This query allowed us to request the number of devices from a specific city.

The scripts were later revised to include both country and city together to ensure that the correct results were returned. An important point was brought up when discussing our academic paper ”Vulnerability Exposure Driven Intelligence in Smart, Circular Cities” which was that countries may have a city with the exact same name. Figure 5 shows the results of a query ran in Shodan.io to check the exposure numbers in Athens. However, there is an Athens in Greece, Canada, and the United States. To fix this potential problem, I changed my code and included additional hard-coded country codes to make sure that the cities which are being queried are in the correct country.



TOTAL RESULTS

489,779

TOP COUNTRIES

Greece	417,156
United States	72,614
Canada	9

Figure 5: Athens Query (Sourced from Shodan (2013))

3.3 Jupyter Notebook

This project is largely analysing the data scraped from Shodan, CVE, NVD, and other sources and will produce an extremely large data set which means we will be using Jupyter Notebook and uses usually include: data cleaning and transformation, numerical simulation, statistical modelling, data visualization, machine learning, and much more. Jupyter also supports over 40 programming languages, and we will be using Python to create code to analyse our data set.

We are working alongside ENISA and their 2018-2019 vulnerabilities notebooks and furthering their research by including a longer variety of dates as well as other vulnerability information like cities and country exposure to provide a much larger analysis.

3.4 Data Sets

The data set used was a single XLSX file that included multiple spreadsheets for specific areas of our data research. The most notable and important data sets are those that relate to vulnerabilities, cities, and countries. A full data legend can be found within the first jupyter notebook.

Source	Features
Vulnerabilities	
NIST NVD	CVSS, CWE-ID, CPE-ID
MITRE CVE	CVE-ID
MITRE CWE	CAPEC-ID
MITRE CAPEC	ATT&CK Data
MITRE ATT&CK	Techniques, Sub Techniques & Tactics
Shodan.io	Exposure Numbers
Cities	
IESE Cities in Motion	96 variables, City In Motion Ranking, Cities
IMD Smart City Index	Adoption of Digital Tech, Citizen's Perceptions, Ranking, Cities
C40	Leading CE Cities
Countries	
IMD World Digital Competitiveness	Factors: Knowledge, Tech, Future Readiness, Score

Table 3.2: Vulnerabilities, Cities and Countries Dataset Description

4 ANALYSIS AND FINDINGS

4.1 Vulnerabilities

4.1.1 MITRE CVE

Looking at the actual vulnerabilities is our first step to understanding the threat landscape and the state of vulnerabilities in 2021. Figure 6 plots the top 10 CVEs that have the highest exposure rate. It is important to be aware of what CVEs have the largest exposure and consequently the most common vulnerability as cities can check their exposure rates with this list of CVEs. CVE-2017-15906 allows malicious threat actors to bypass security in OpenSSH and knowing this and knowing what versions of OpenSSH are vulnerable means that we can mitigate our attack surface.

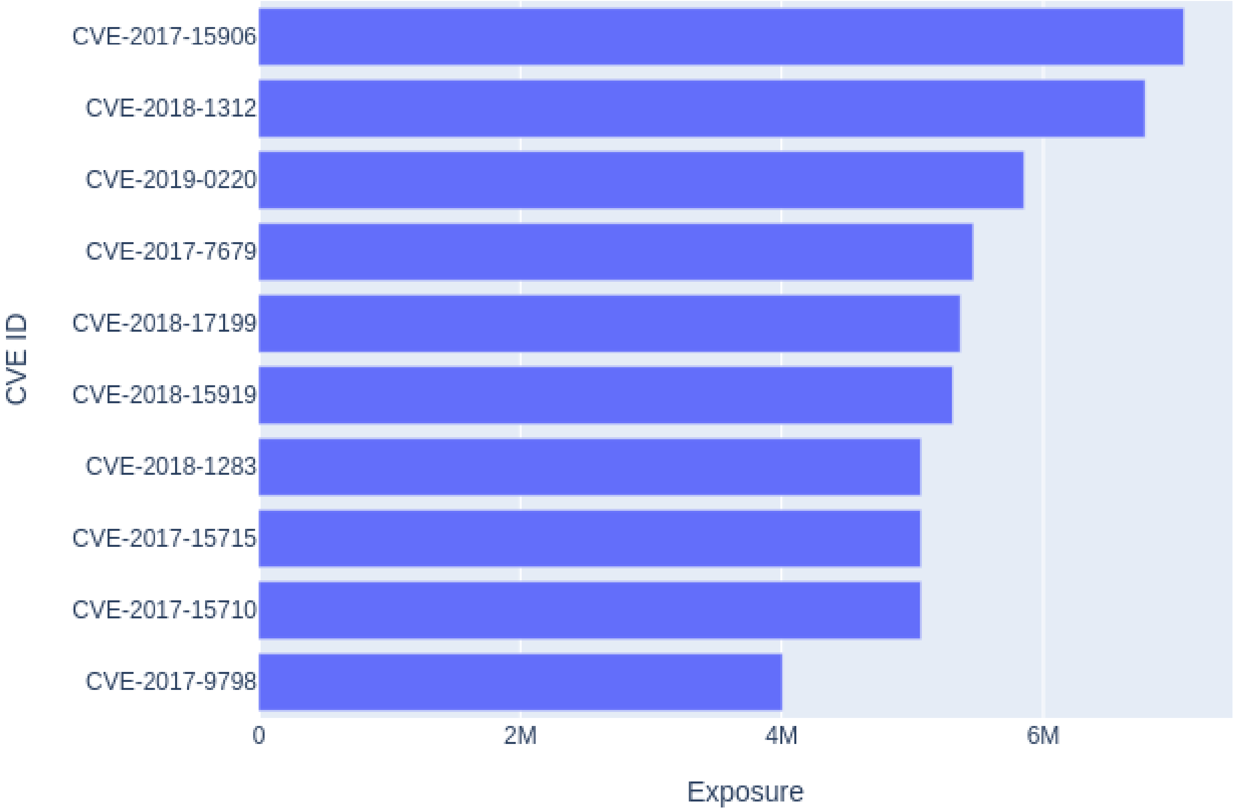


Figure 6: Top 10 CVEs

Throughout the start of this project towards the end of this project, I made sure to keep the total exposure numbers for our CVEs to see if we can identify a trend. we used seaborn which is a python data visualisation library based on matplotlib to plot the top 10 CVEs that expose the most devices. Figure 7 shows our line plot of our top 10 CVEs. We can see that the CVEs keep relatively consistent exposure numbers apart from **CVE-2017-15906** and **CVE-2018-15919** where during March of 2021 roughly 500,000 devices are patched and now exposed again within April. We can see that during March all our top 10 CVEs actually decrease slightly and increase again after.

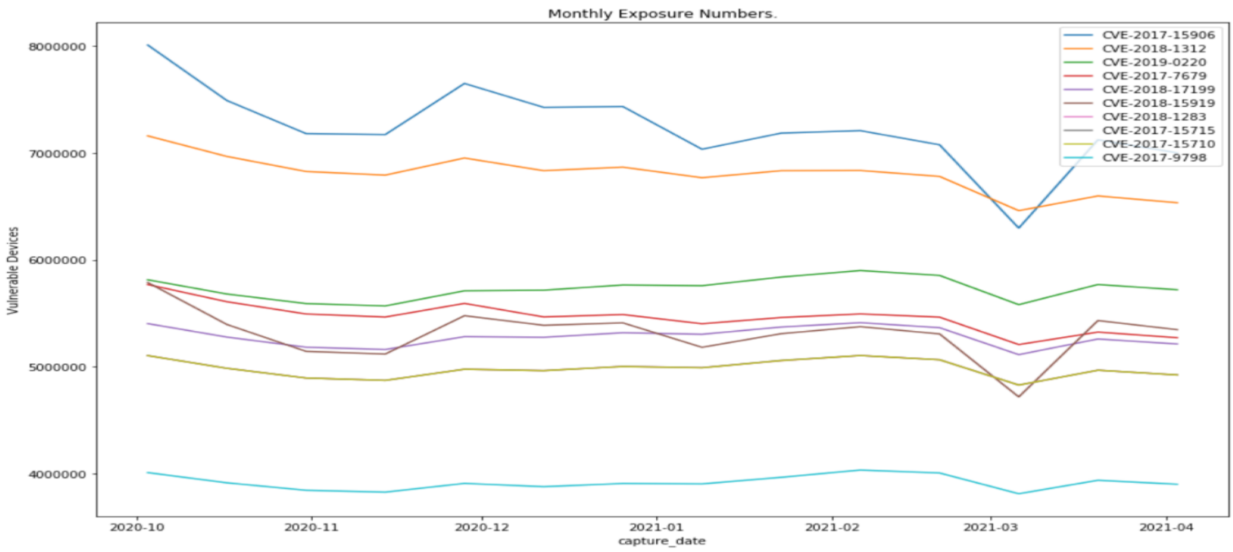


Figure 7: Top 10 CVEs and Historical Numbers

4.1.2 NIST NVD & CVSS

The information sourced from the NVD combined with their correlated CVEs means we can now analyse our vulnerabilities based on CVSS information. Figure 8 shows a histogram of our vulnerabilities based on their CVSS2 and CVSS3 severity ratings. We can observe that:

- CVSS2 have rated a small number of our vulnerabilities more than CVSS3 in **LOW**.
- There are more **MEDIUM** vulnerabilities rated by CVSS2 than there is CVSS3.
- CVSS3 have rated more vulnerabilities **HIGH** than CVSS2.
- CVSS2 have no category for **CRITICAL**.

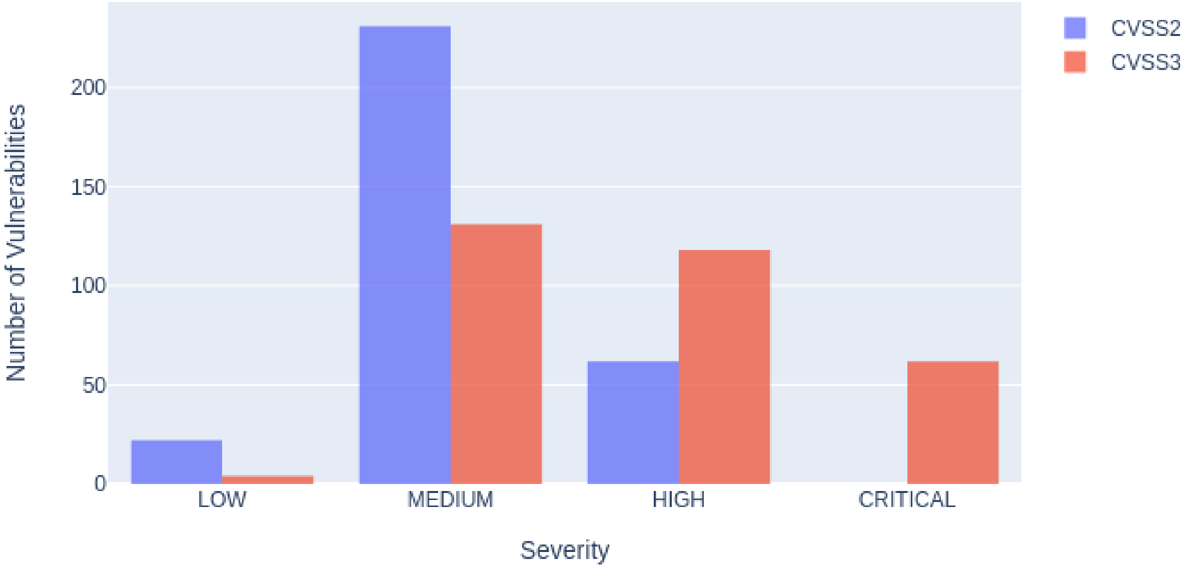


Figure 8: Number of vulnerabilities based on severity ratings

Figure 9 shows a histogram of all our vulnerabilities and both their CVSS2 and CVSS3 base scores. We can see from this histogram that:

- CVSS3 has a much higher number of vulnerabilities (**54**) than CVSS2 (**4**) with a base score between **9.2 - 10.8**.
- CVSS2 has a higher presence in the lower end of the base score scale.
- Most vulnerabilities are ranked between **3.8 - 7.7**, which makes sense considering that most of our vulnerabilities were scored with a severity rating of **MEDIUM** from both CVSS3 and CVSS2.

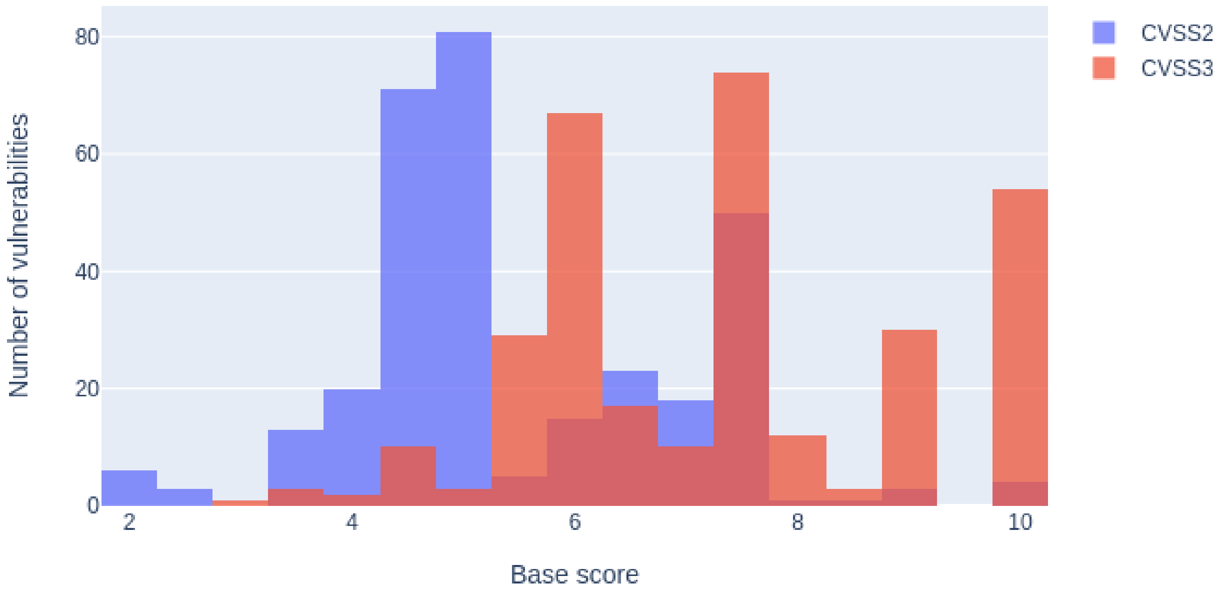


Figure 9: Number of vulnerabilities according on base scores

4.1.3 MITRE CWE

The additional information scraped from the CWE website when combined with our CVEs allows us to further enrich our data and allows us to paint a more detailed and complex picture into the current threat landscape and investigate how and which common weaknesses adversaries attack. Figure 10 shows the top 10 CWEs that are the most recurring in our data set, we can identify some common weaknesses like path traversal, SQL injection, or Cross-Site Scripting.

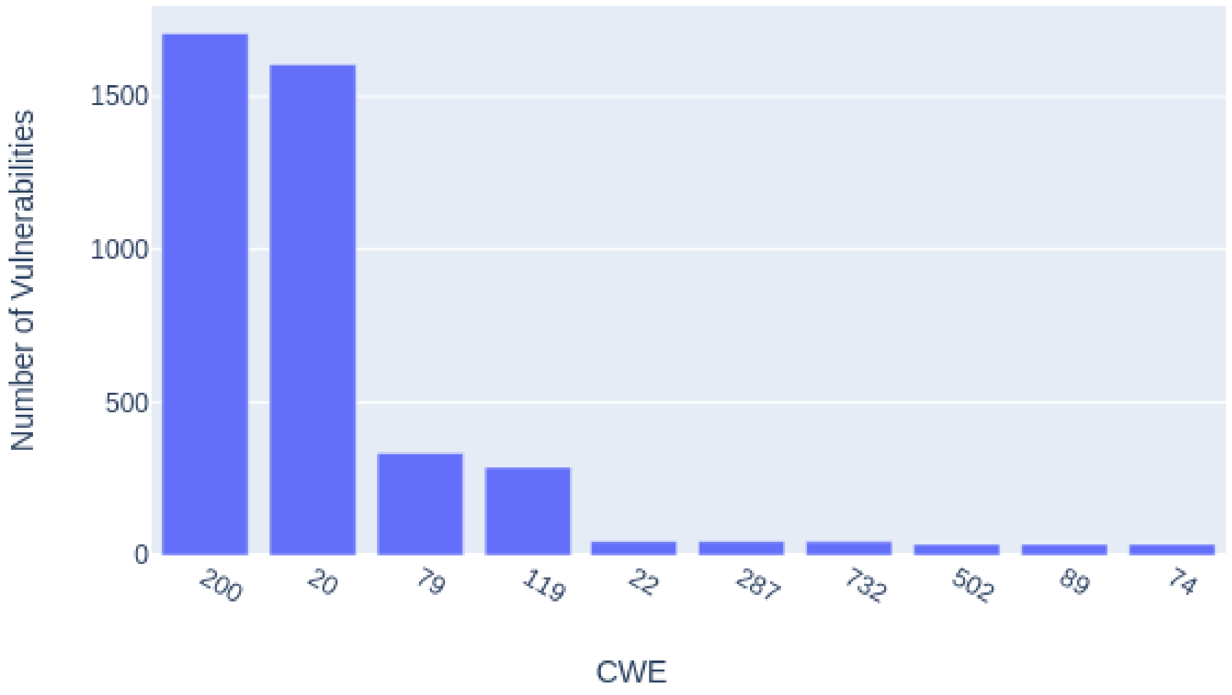


Figure 10: Top 10 CWEs

CWE ID	CWE
200:	Exposure of Sensitive Information to an Unauthorized Actor
20:	Improper Input Validation
79:	Improper Neutralization of Input During Web Page Generation ('XSS')
119:	Improper Restriction of Operations within the Bounds of a Memory Buffer
22:	Improper Limitation of a Pathname to a Restricted Directory ('Path Traversal')
287:	Improper Authentication
732:	Incorrect Permission Assignment for Critical Resource
502:	Deserialization of Untrusted Data
89:	Improper Neutralization of Special Elements used in an SQL Command ('SQLI')
74:	Improper Neutralization of Special Elements in Output Used by a Downstream Component

Table 4.1: Top 10 CWEs

Additionally, when these CWEs are linked with our vulnerability data set means we can now look at them in other contexts like cities, ports, services or, protocols. We've used seaborn and their heat maps to provide a more physical representation of CWEs and how dangerous they can be by linking them to the corresponding CVSS3 base score.

There isn't much difference within these heat maps, but the minuscule differences here and there help tell a story within the given context. Figure 11 shows all our heat maps relating to CWEs. One of the heat maps is for the top 10 cities to the top CWEs and their associated CVSS3 base score; we can recognize CVEs linked to CWE-ID 22 and takes advantage of a path traversal weakness that is more severe in Singapore than it is in Paris. Another one shows that the protocol PCWorx is less vulnerable than Ethernet/IP, Modbus, ProConOS, and S7 Communication. We can also recognize that all weaknesses affect Apache but at different scores.

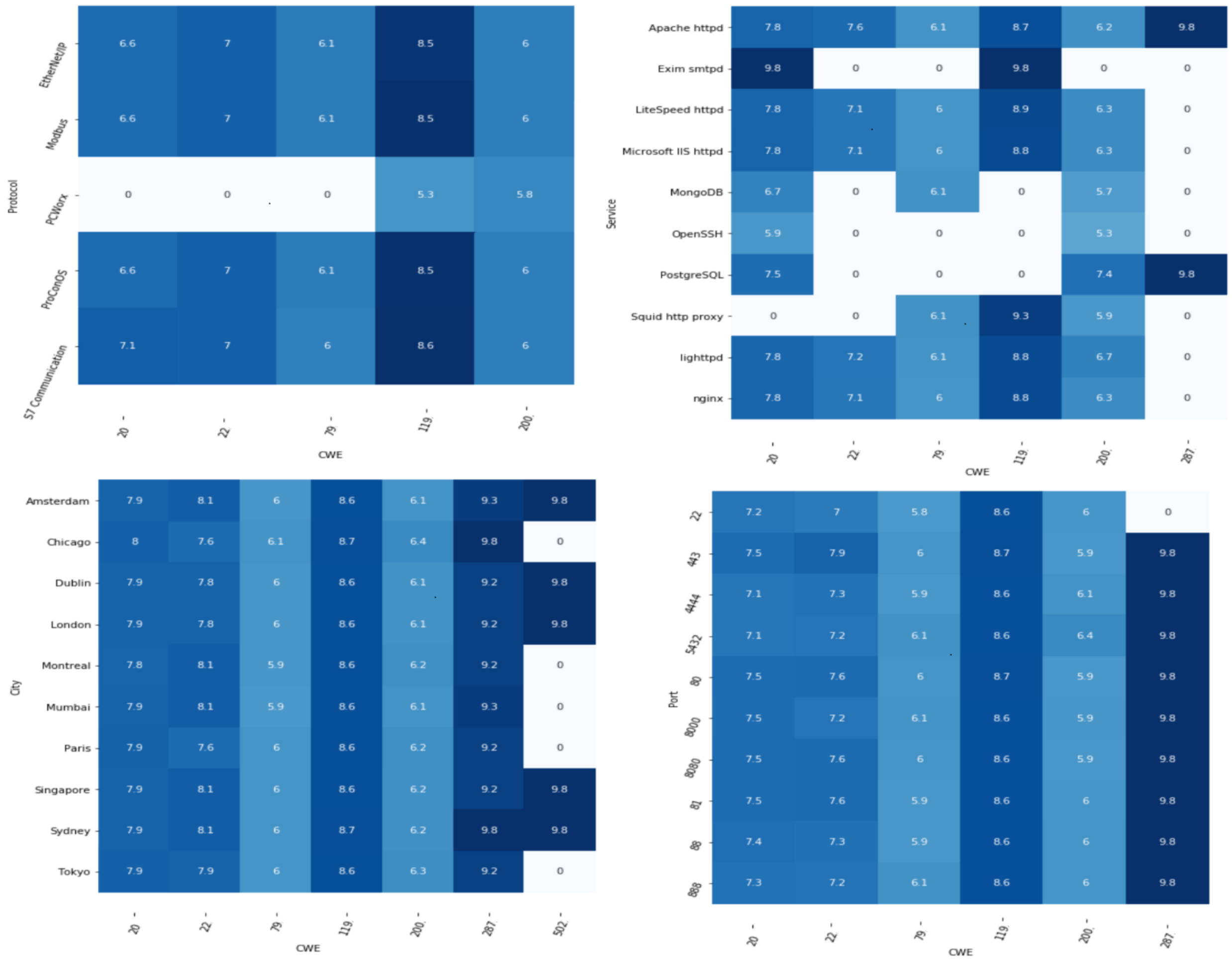


Figure 11: Protocols, Services, Cities and Ports Heat map for CWE

4.1.4 MITRE CAPEC

Each vulnerability within our list had roughly a single weakness associated with them and each weakness had multiple attack patterns associated with them. This means during the analysis stage we had to separate each vulnerability into single attack patterns using the utility file provided with the Jupyter notebooks we got ENISA and joined the separate data set to a list of common attack patterns. The CAPECs are now easier to understand and put into context on how it helps us look at the threat landscape. CAPECs can help us understand how adversaries operate by providing us with a comprehensive dictionary of known attack patterns that adversaries use to exploit known weaknesses (MITRE (nda)).

Figure 12 shows the top 10 most common CAPECs that are vulnerabilities are linked to. CAPEC-22 was the top CAPEC that occurred within our data set. Exploiting Trust in the client is an attack that exploits vulnerabilities in client/server communication channel authentication and data integrity. It leverages the implicit trust a server places in the client, or more importantly, that which the server believes is the client. An attacker executes this type of attack by communicating directly with the server where the server believes it is communicating only with a valid client. There are many variations of this type of attack (MITRE (ndb)).

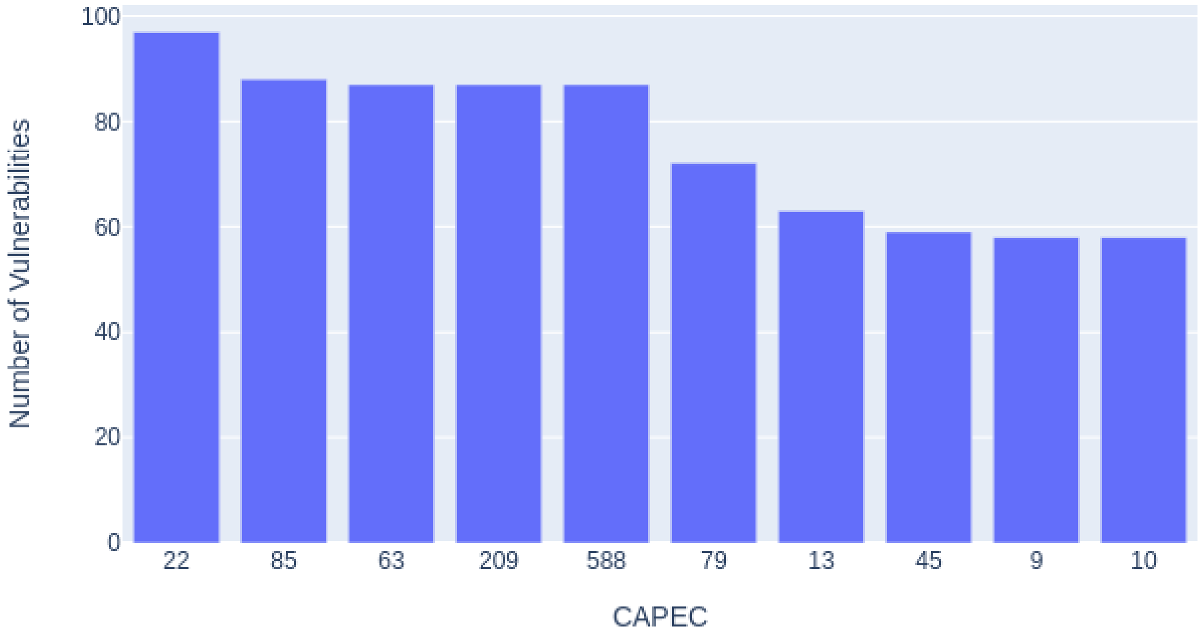


Figure 12: Top 10 CAPECs

CAPEC ID	CAPEC
22:	Exploiting Trust in Client
85:	AJAX Footprinting
63:	Cross-Site Scripting (XSS)
209:	XSS Using MIME Type Mismatch
588:	DOM-Based XSS
79:	Using Slashes in Alternate Encoding
13:	Subverting Environment Variable Values
45:	Buffer Overflow via Symbolic Links
9:	Buffer Overflow in Local Command-Line Utilities
10:	Buffer Overflow via Environment Variables

Table 4.2: Top 10 CAPECs

As before with CWEs, we can combine data scraped from CAPEC with our list of vulnerabilities to enrich the data sets. Figure 13 shows all our heat maps relating to CAPEC data. The most obvious thing that jumps out from these heat maps is CAPEC-ID 9 and 10 appear to have a higher base score across the board, and both are common attack patterns relating to a buffer overflow attack. Both these CAPECs appear to affect each of the top 10 cities, ports, and services.

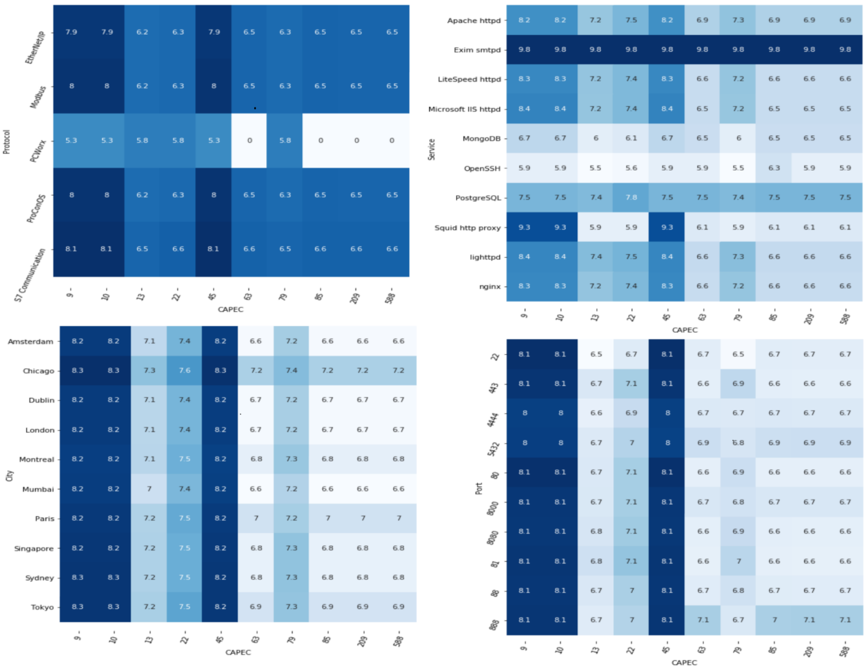


Figure 13: Protocols, Services, Cities and Ports Heat map for CAPEC

4.1.5 MITRE ATT&CK

Linking ATT&CK data to our vulnerabilities is a brilliant way to provide cities with more actionable threat intelligence and help improve their security operations and architecture as well as vulnerability contextualization as ATT&CK data is based on real-world observations.

Tactics represent the “why” of an ATT&CK technique or sub-technique. It is the adversary’s tactical goal: the reason for acting (MITRE (2015)). Figure 14 shows the top 10 ATT&CK tactics linked to our CVEs through their associated CAPECs and table 4.3 tells us what each of the tactics is.

The top tactic associated with our CVEs is ”Discovery”, where an adversary uses techniques assigned by ATT&CK to gain knowledge about the system or internal network they’re attacking. The techniques used help the attackers observe the environment and take note of what they control to ensure the most efficient entry point.

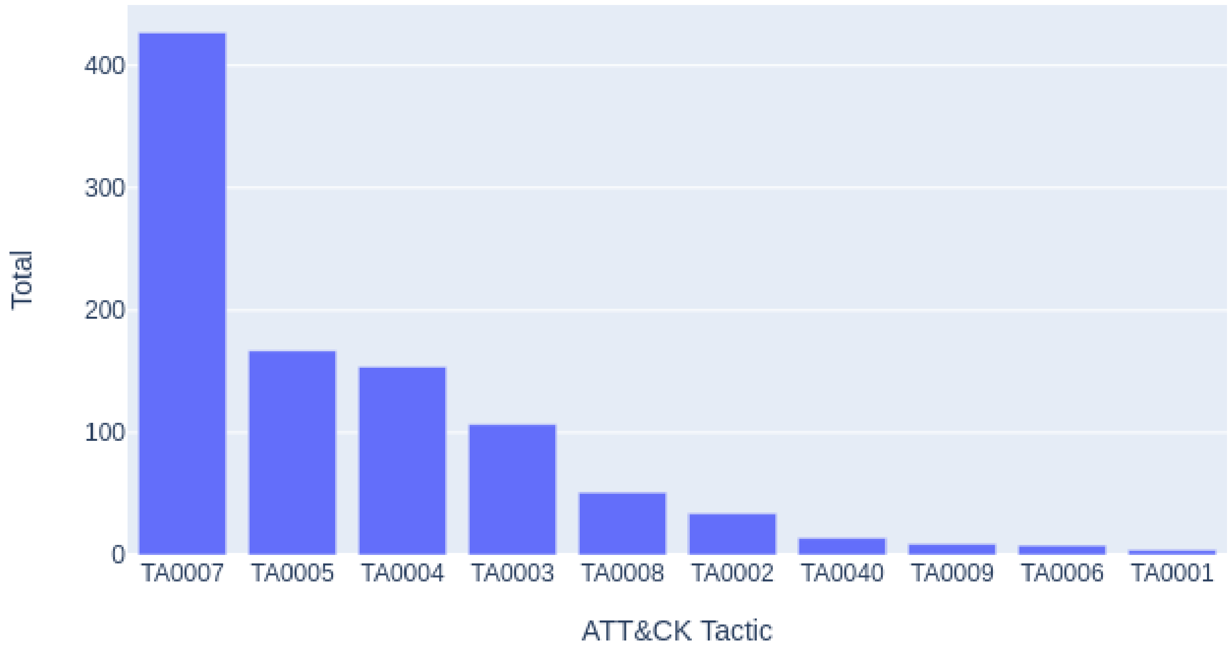


Figure 14: Top 10 ATT&CK Tactic

ATT&CK Tactic ID	Tactic
TA0007:	Discovery
TA0005:	Defense Evasion
TA0004:	Privilege Escalation
TA0003:	Persistence
TA0008:	Lateral Movement
TA0002:	Execution
TA0040:	Impact
TA0009:	Collection
TA0006:	Credential Access
TA0001:	Initial Access

Table 4.3: Top 10 ATT&CK Tactic

Box plot diagrams were also used to quickly contextualize two specific data points in our case it was both CVSS3 and CVSS2 base scores for our ATT&CK data. Figure 15 shows a box plot of the ATT&CK tactics that our CVEs were linked to. We can see that across the board CVSS3 scored our tactics mostly higher than CVSS2 in their minimum, median, and maximum as well as their first and third quartile.

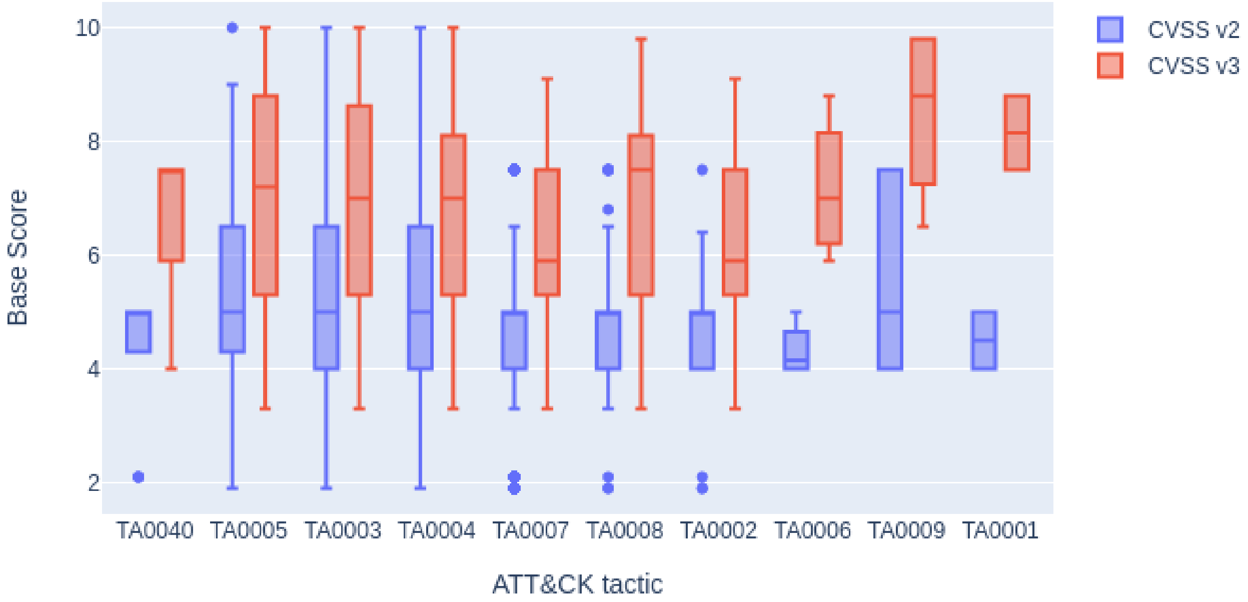


Figure 15: ATT&CK Tactics Box plot

These techniques represent a view of “how” an adversary achieves his tactic by acting. Figure 16 shows the top 10 ATT&CK techniques linked to our CVEs through their associated CAPECs and table 4.4 tells us what each of the techniques is.

The top technique associated with our CVEs is ”Hijack Execution Flow”, where an adversary can execute their malicious payloads by hijacking how programs are running on their victim’s device. This can lead to advanced persistent threats (APT), privilege escalation, or defence evasion.

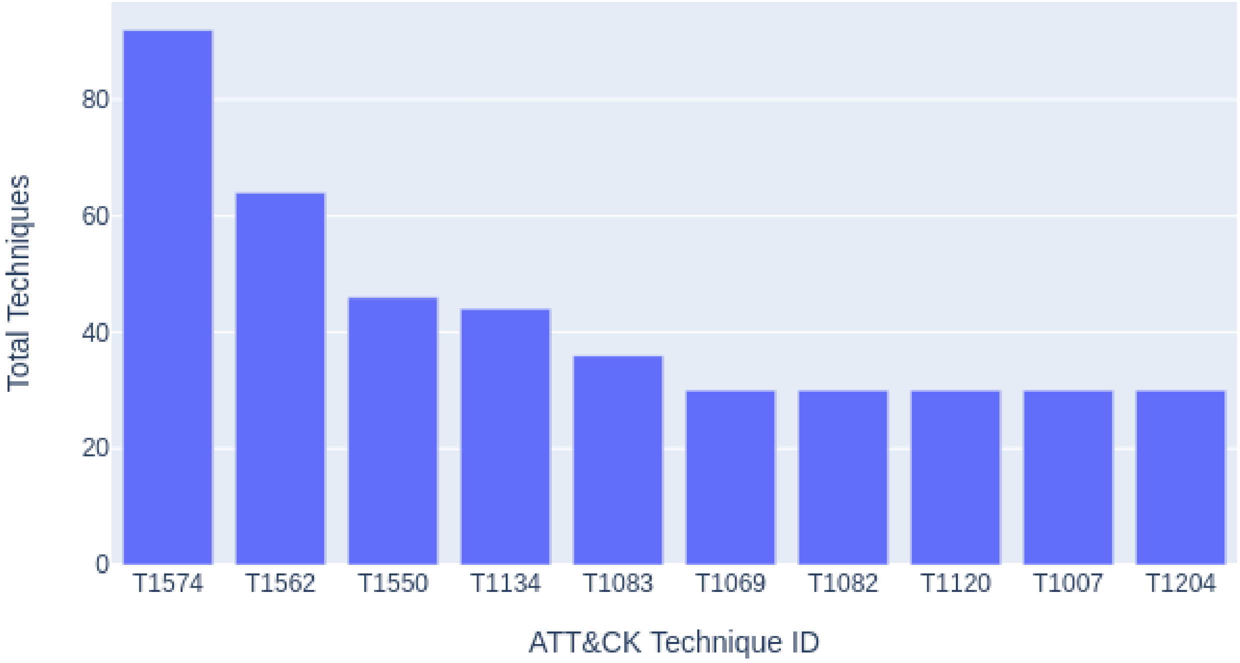


Figure 16: Top 10 ATT&CK Techniques

ATT&CK Technique ID	Technique
T1574:	Hijack Execution Flow
T1562:	Impair Defenses
T1550:	Use Alternate Authentication Material
T1134:	Access Token Manipulation
T1083:	File and Directory Permissions Modification
T1069:	Permission Groups Discovery
T1082:	System Information Discovery
T1120:	Peripheral Device Discovery
T1007:	System Service Discovery
T1204:	User Execution

Table 4.4: Top 10 ATT&CK Techniques

Sub-techniques are a more specific description of the adversarial behaviour used to achieve a goal. Figure 17 shows the top 10 ATT&CK sub techniques linked to our CVEs through their associated CAPECs and table 4.5 tells us what each of the sub techniques is.

The top sub technique is ”LD_PRELOAD” where an adversary manipulates the Linux library loader LD_PRELOAD to point to a malicious library which when the target’s device requests the library it will execute the adversary’s malicious code and executes it causing potential detection evasion, privilege escalation, unauthorized access to network resources.

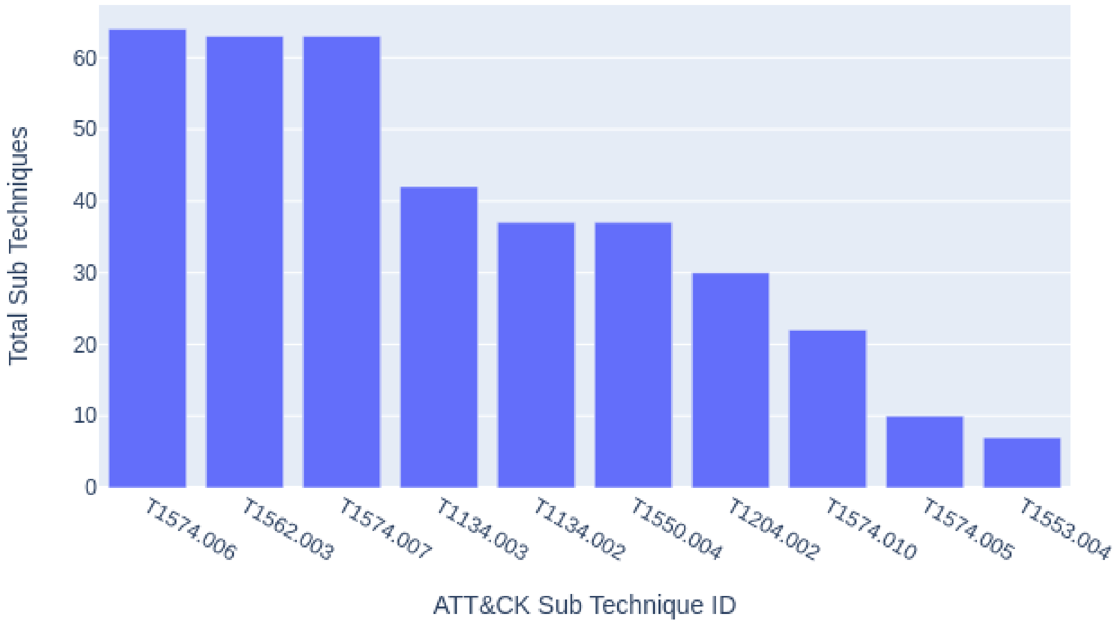


Figure 17: Top 10 ATT&CK Sub Techniques

ATT&CK Sub Technique ID	Sub Technique
T1574.006:	LD_PRELOAD
T1562.003:	Impair Command History Logging
T1574.007:	Path Interception by PATH Environment Variable
T1134.003:	Make and Impersonate Token
T1134.002:	Create Process with Token
T1550.004:	Web Session Cookie
T1204.002:	Malicious File
T1574.010:	Services File Permissions Weakness
T1574.005:	Executable Installer File Permissions Weakness
T1553.004:	Install Root Certificate

Table 4.5: Top 10 ATT&CK Sub Techniques

4.1.6 Exploits

Simply knowing what CVE ID has the largest exposure rate isn't enough; which is why I scraped the web for the specific exploits associated with our vulnerabilities. Almost all this data was sourced from the vendor's website identified by MITRE CVE. This information when combined with other information scraped allows us to part putting the pieces together. Figure 18 shows the type of attacks and figure 19 shows us which platform is vulnerable to the attack associated with a specific CVE. The results aren't surprising with DoS and XSS being the most common attack type and Joomla!, PHP, Wordpress, and Apache being the most common platform.

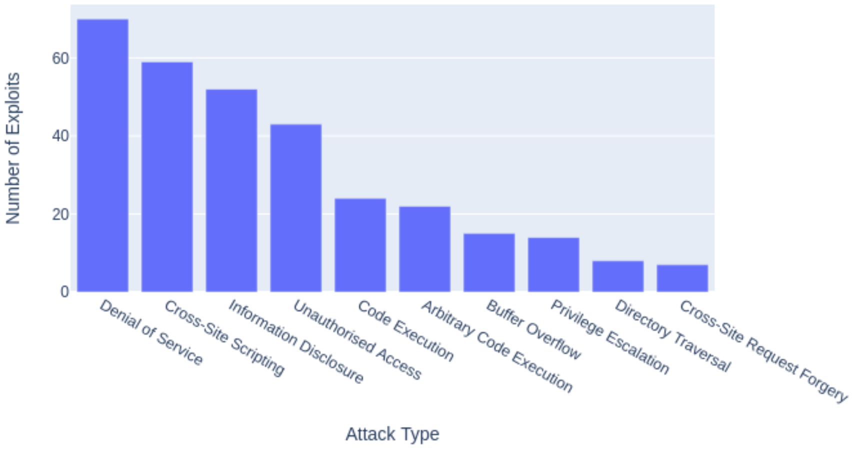


Figure 18: Exploit Attack

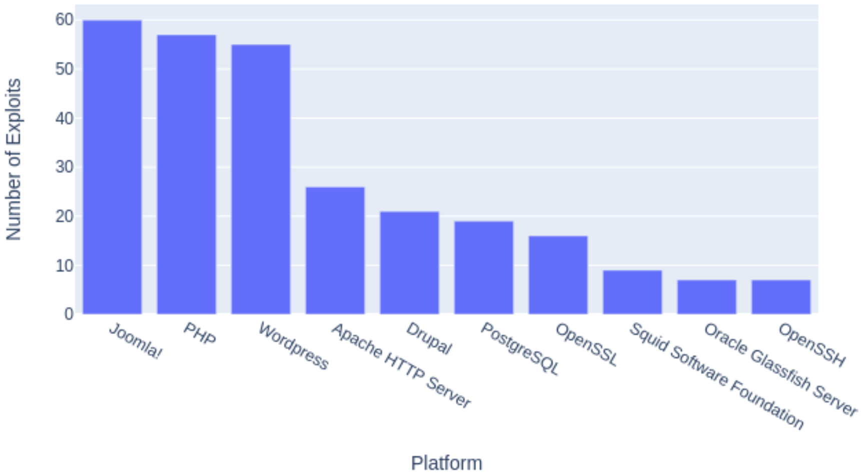


Figure 19: Exploit Platform

We split up the exploits based on whether they were web or not and enriched them with base scores and severities from CVSS. Figure 20 shows how many web and non-web exploits we have and their severity. We can discern CVSS2 has given more web and non-web exploits a **LOW** and **MEDIUM** than CVSS3 which has given more exploits the **HIGH** and **CRITICAL** rating. We can also note there are more than three times the web exploits than non-web. This makes sense when you take into consideration that the most common exploit attacks and platforms are mainly web-based.

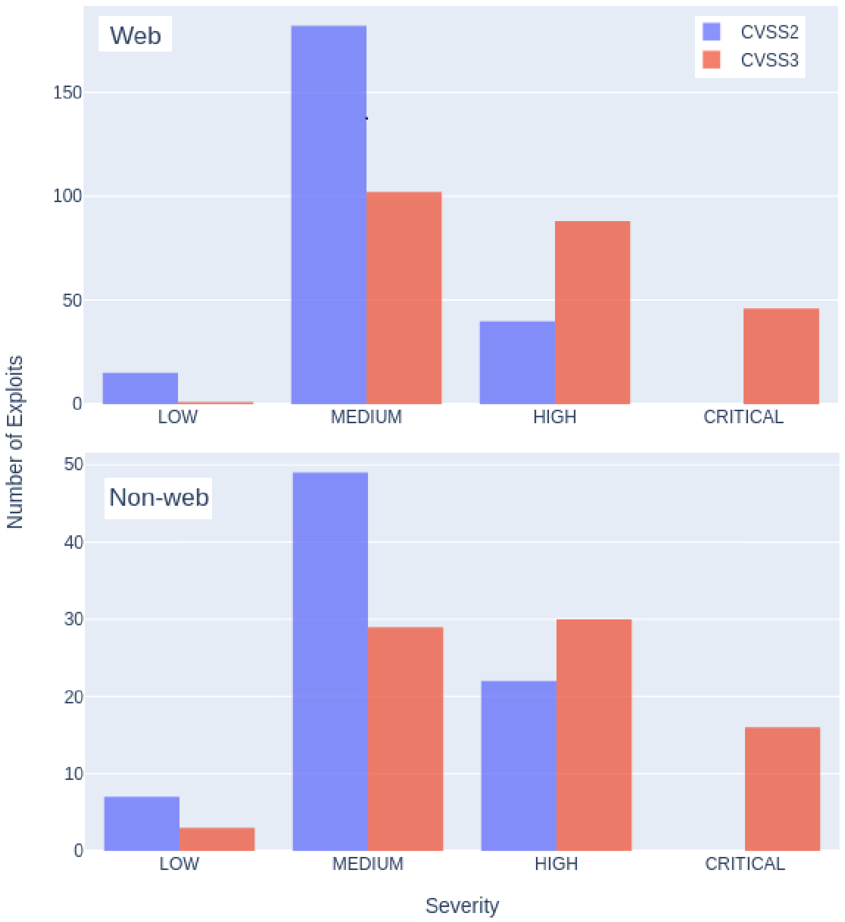


Figure 20: Web & Non-web Severity

4.2 Cities

When analysing our data sets and focusing on cities, we take into consideration the maturity level of a city and whether they have a circular or smart agenda. We examined four classes of cities:

Class 1: the **plain** city. This class refers to the cities that have not shown a smart or circular city agenda or not showing any intention of doing so;

Class 2: the **circular** city. This class refers to cities that have or are implementing a CE approach and making the transition into data-driven;

Class 3: the **smart** city. This class refers to the cities that are considered smart and are included in the IMD SCI or the IESE CIMI;

Class 4: the **both** city. This class is referring to cities that are considered smart and are included in either index, also have implemented a CE approach, and are included in the C40 Index.

As mentioned in the introduction chapter we go into this project with some hypotheses:

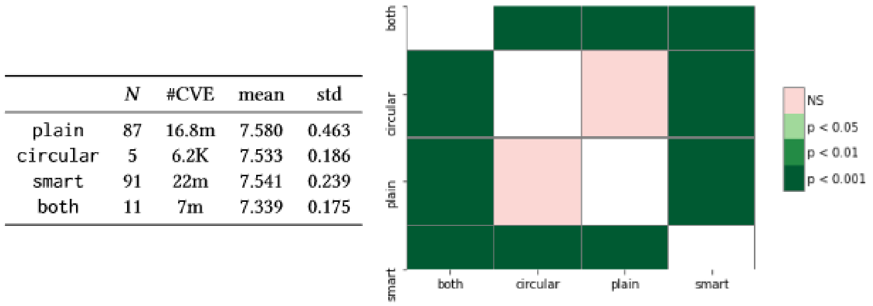
H_1 Different types of cities have different exposure rates.

H_2 Cities have a much higher exposure rate with an increased population.

H_3 Cities ranked higher in either the SCI or CIMI have less exposure than those that ranked lower.

Investigating H_1 focuses mainly on the total exposure numbers rather than anything else. We expect that smart cities or cities with a data-driven approach have a much higher attack surface than cities that haven't adopted a smart city or CE agenda. We have to look at every class of cities when investigating H_1 and at the end, we can compare the total exposure rates for the top cities ($N > 10$ in classes 1, 3, and 4 and $N = 5$ in CE).

Figure 21 summarises the pairwise vulnerability means comparisons of the different classes of cities. In every case, the average CVSS base score is around the mid 7s. The **plain** city class has the highest mean score than the other classes of cities.



	N	#CVE	mean	std
plain	87	16.8m	7.580	0.463
circular	5	6.2K	7.533	0.186
smart	91	22m	7.541	0.239
both	11	7m	7.339	0.175

Figure 21: Pairwise comparisons of the four city classes (Sourced from Katos et al. (2020))

It should be noted that this exploratory analysis is limited as the comparison is a static approach. Also, our **circular** class of cities sample were relatively small ($N = 5$), which is why we considered class **motion** as a further class of cities that contains all cities from the CIMI and are the cities that are on the low end of the maturity scale but are showing signs of smartness and higher than the cities from the class **plain** who have not declared or adopted any smart or circular agenda or initiative.

Because of the large number of CVEs that affect our different classes of cities, we normalise the data to develop meaningful comparisons. We achieved this by dividing our vulnerability variable by our population variable and storing that integer into a separate variable that was later used for our regression models.

We can see from figure 22 that Ashburn is more vulnerable than the other top 9 cities sitting at an exposure rate of 3.6 million devices vulnerable to our list of CVEs as of 19th February. This is not a surprise as Ashburn, Virginia is a major hub for internet traffic with a vast number of data centres within the city which increases its attack surface. So considering this we can consider that the other normal cities would be a more accurate indicator of how vulnerable plain cities are. We can also see from figure 23 that the exposure distribution is more like what we would expect with highly smart cities being more exposed. 24 displays at the top exposed cities that have implemented a circular agenda.

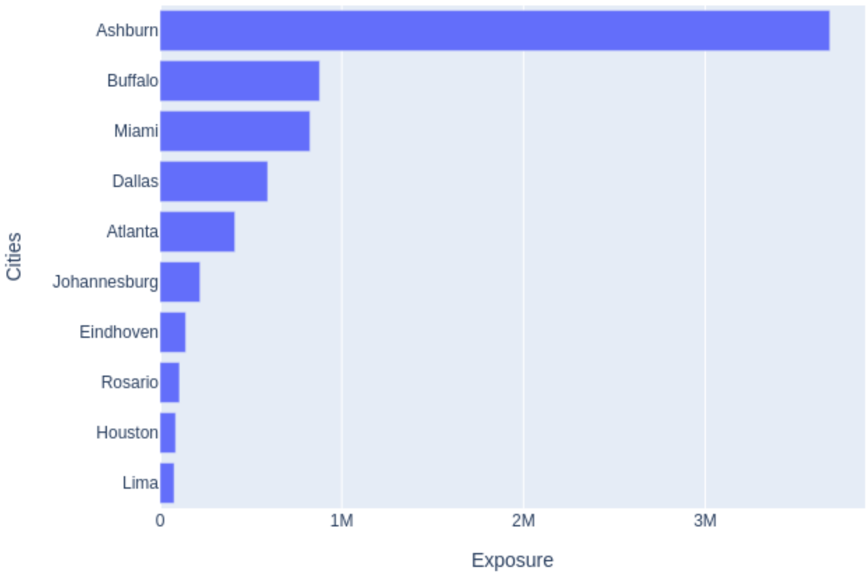


Figure 22: Top 10 Exposed Plain Cities

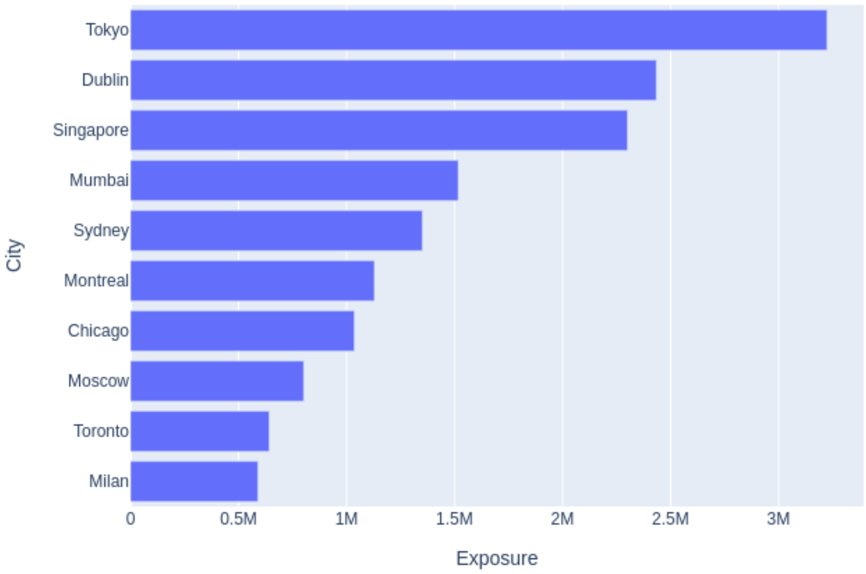


Figure 23: Top 10 Exposed Smart Cities

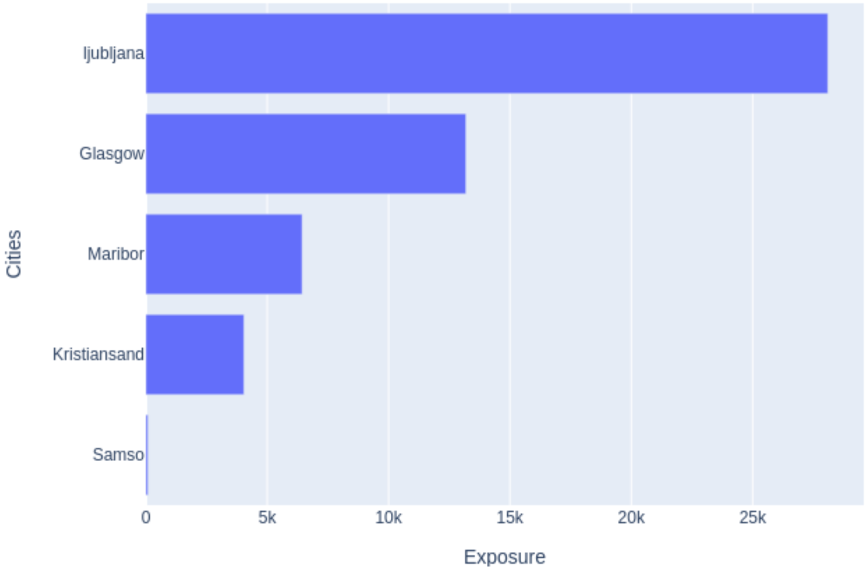


Figure 24: Top 5 Exposed Circular Cities

When looking into H_2, we can look at the exposure numbers and population for N cities. We expect cities that have a higher population have a much higher rate of exposure due to the increased number of devices per individual citizen. Figure 25 shows the top 10 cities with the highest population from all city classes. We can notice that Tokyo is our highest populated city with a population of 38 million people and is also the smart city that has the highest exposure rate of 3.2 million devices vulnerable to our list of CVEs which can be seen in figure 23.

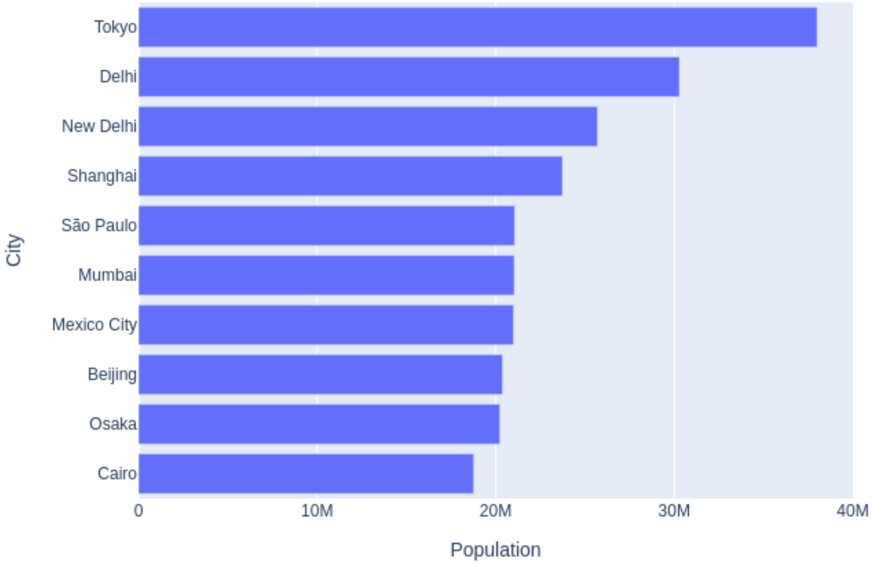


Figure 25: Top 10 Populated Cities

When looking into H_3, we can look at comparing the exposure rates for cities that ranked higher in both the SCI and the CIMI with cities that ranked lower. Cities that ranked higher in either index are considered more advanced and would have scored higher in factors like technology. Figure 26 shows Singapore which is considered the most advanced smart city and ranked first with Cairo, Nairobi, Manila, Lagos, Abuja, and Rabat which were all ranked last in the SCI. We can note sophisticated cities like Singapore which scored the highest in either index have a much higher exposure rate than those considered less advanced. This makes H_3 an intriguing hypothesis as we expect that a more advanced smart city that is smarter about the technology deployed into their cities should have a negative relationship with their total exposure. However, the more technology that a city implements and depends on, the wider the city attack surface is, meaning a potentially bigger exposure rate.

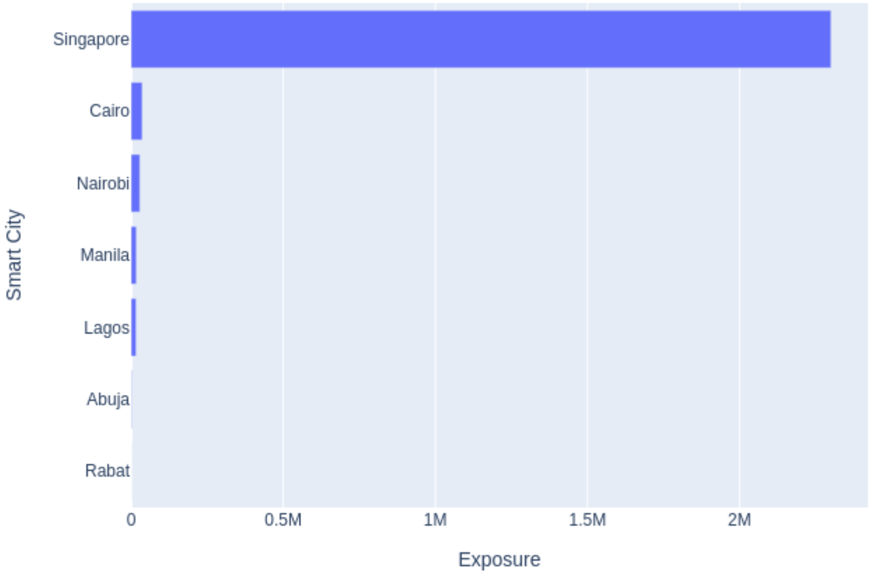


Figure 26: Comparison between Singapore, Cairo, Nairobi, Manila, Lagos, Abuja and Rabat

Figure 27 is a hierarchical clustering (Ward's method) of vulnerabilities in a list of cities. The information contextualised can be used as actionable threat intelligence in two different ways. Firstly, cities, local authorities and governance with similar vulnerabilities or exposure profiles could work together to protect their cities. The second way is cities with similar vulnerabilities and exposure profiles that were attacked could act as an early warning side for the other city and means these cities can increase their situational awareness with information sharing and improve the city's response to any future incident.

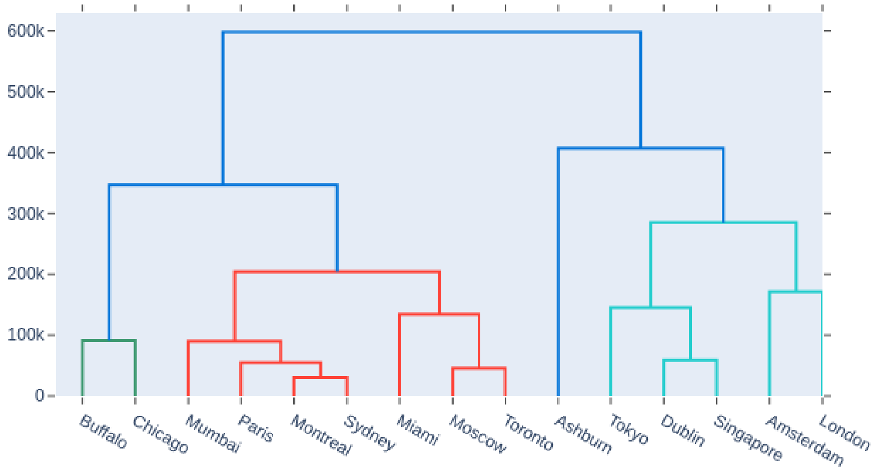


Figure 27: Hierarchical clustering of cities (class:`All`)

When looking at the clusters produced on the dendrogram, we took a conventional approach that used a halfway threshold and in this case, it was three hundred thousand. Figure 28 shows our threshold and we can see that four different clusters are emerging.

$Cluster1$ Buffalo and Chicago.

$Cluster2$ Mumbai, Paris, Sydney, Montreal, Miami, Toronto and Moscow.

$Cluster3$ Ashburn.

$Cluster4$ Tokyo, Dublin, Singapore, Amsterdam, and London.

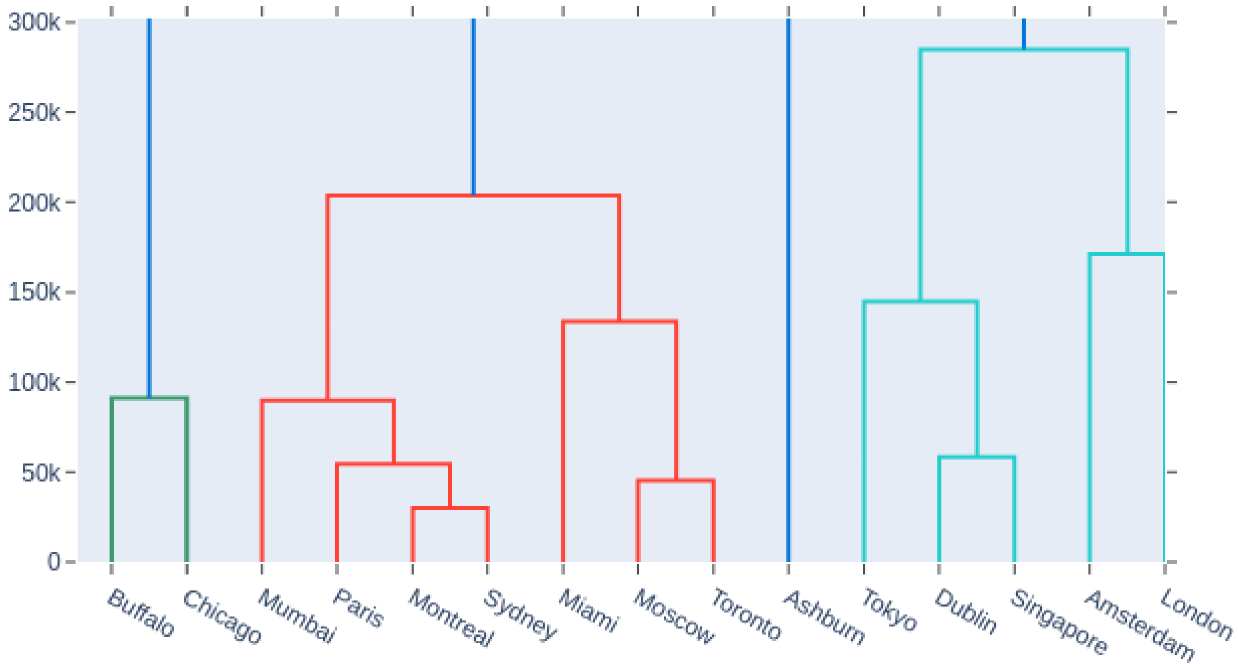


Figure 28: Zoomed hierarchical clustering of cities (class:`All`)

To ensure that we chose the correct amount of clusters when looking at the distance between cities we adopted the popular elbow method. Figure 29 shows our results from running this method. We found the lowest WCSS with the smallest amount of clusters. In our case, the optimal amount of clusters was three; any more clusters would have offered little to no improvement.

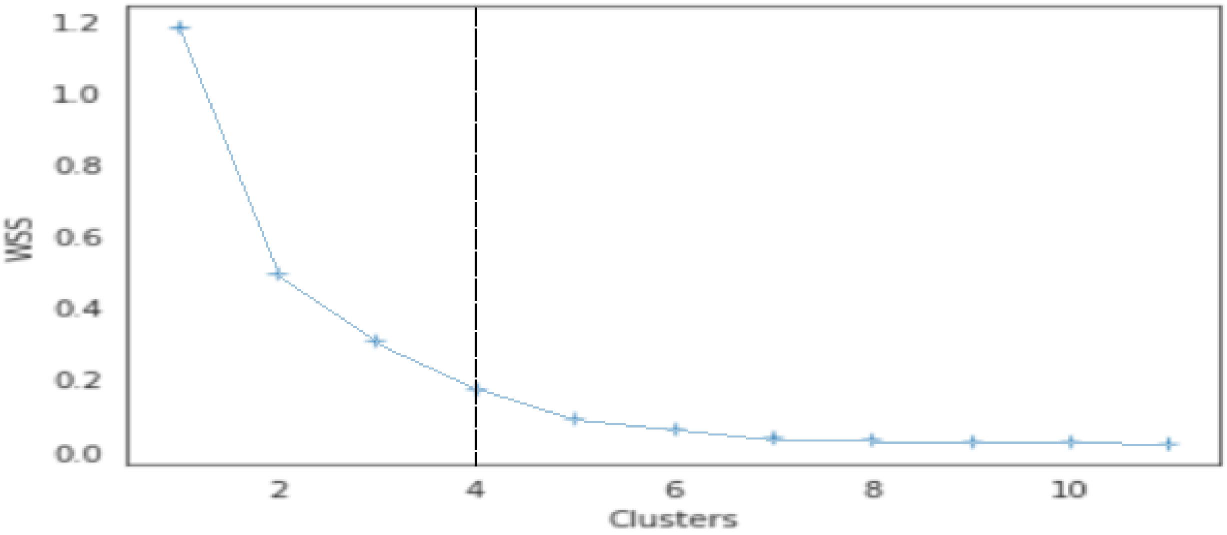


Figure 29: Elbow Plot

Table 4.6 contains the results of a factor analysis performed on the top 15 cities from all classes with the highest exposure profile and the cities that are most vulnerable to attack based on our list of CVEs. The factor analysis produced three factors (groups) containing the list of cities that the analysis grouped together. We ran the Cronbach's Alpha test on each of the groups, and we had significantly high alphas (over 0.7) across the board which shows a strong internal consistency and reliability. This means, as well as using a dendrogram to potentially show a city's vulnerability footprint, we can look at the cities within the assigned factors, and the assigned factors can represent them.

Factor 1		Factor 2		Factor 3	
Ashburn	0.7868	Buffalo	0.9614	Amsterdam	0.8005
Dublin	0.8855	Chicago	0.9704	Toronto	0.7408
London	0.7171	Miami	0.8848		
Montreal	0.9545				
Moscow	0.8153				
Mumbai	0.8762				
Paris	0.8934				
Singapore	0.9063				
Sydney	0.9587				
Tokyo	0.9504				
Cronbach's alpha	0.9607		0.9330		0.7671

Table 4.6: Factor Analysis of top 15 cities with the most potential vulnerabilities

To ensure that our factor analysis produces the correct number of factors, we perform a Barlett Sphericity test to check whether our data returns a significant p-value and in our case, it is 0.0. Figure 30 shows our scree plot that plots our eigenvalues and our factors, and then we can use the same method as before with our clusters to check what would be the correct number of factors, in this case, it was three.

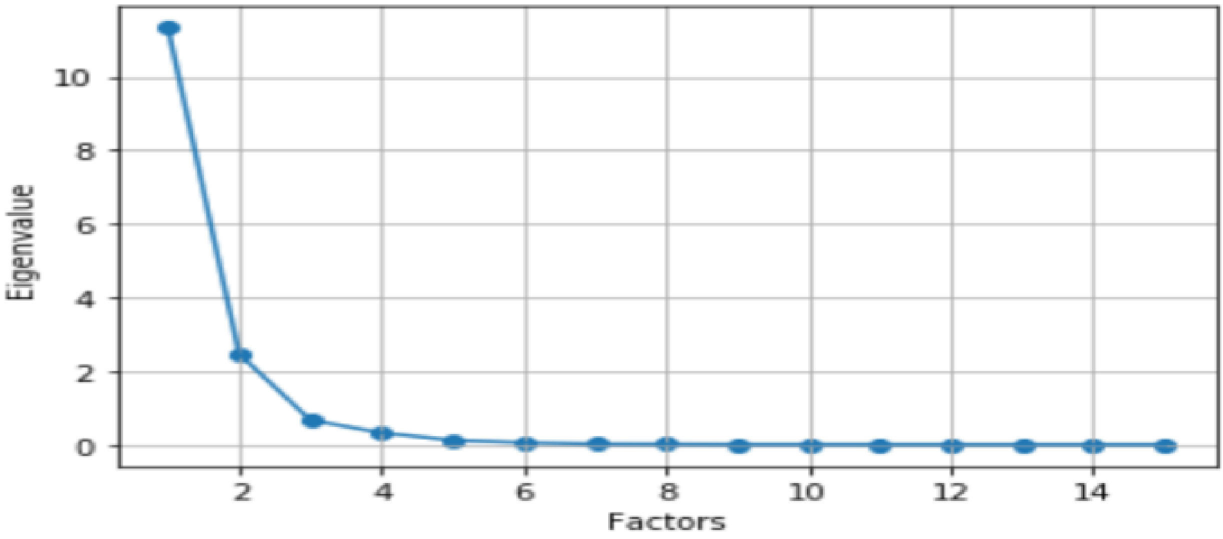
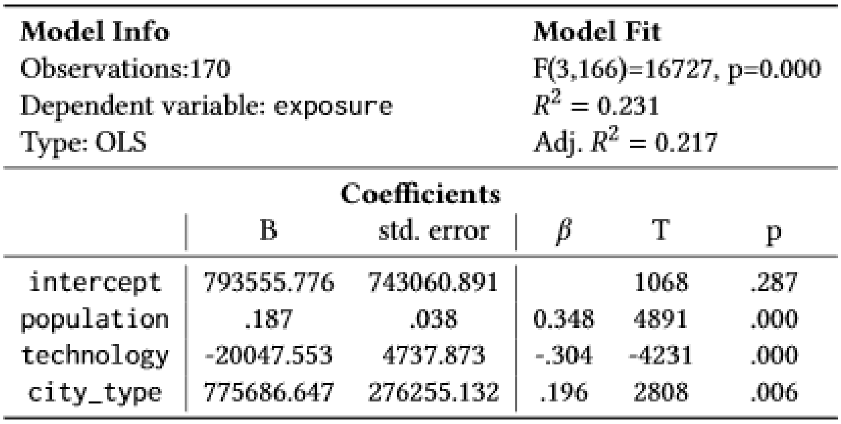


Figure 30: Eigenvalue and Factors Scree Plot

In addition to Ward's method and factor analysis, we also perform OLS regression to explore and study the factors that influence our dependent variable, exposure. Figure 31 shows our first regression model, within this model the independent variable `city_type` refers to our four different city classes with `circular` being replaced by cities in the CIMI. Figure 32 is another regression model where our variable `city_type` is exploded into their different classes to help investigate our first hypothesis H_1. We can observe the β coefficient which shows that smarter cities are more exposed than the cities that haven't adopted a smart or circular approach. It also shows technology as another independent factor which shows the fact that technology alone is insufficient to address the potential issues that could arise within a city.



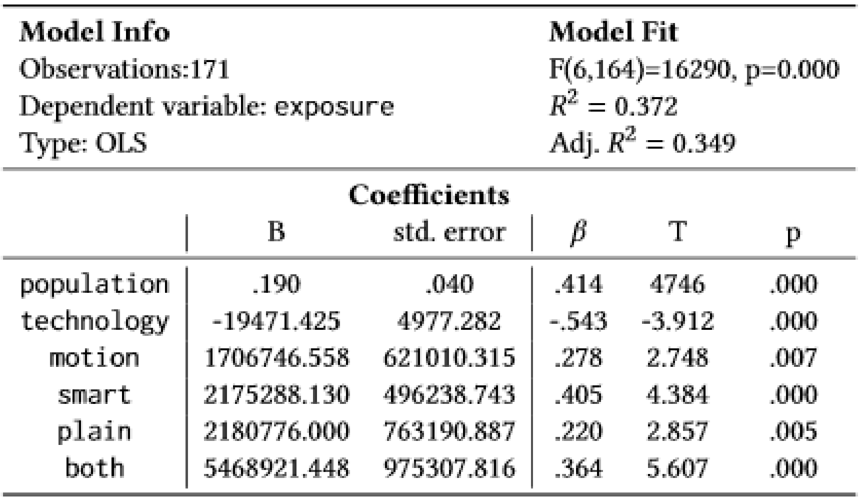
Model Info
Observations:170
Dependent variable: exposure
Type: OLS

Model Fit
F(3,166)=16727, p=0.000
$R^2 = 0.231$
Adj. $R^2 = 0.217$

Coefficients

	B	std. error	β	T	p
intercept	793555.776	743060.891		1068	.287
population	.187	.038	0.348	4891	.000
technology	-20047.553	4737.873	-.304	-4231	.000
city_type	775686.647	276255.132	.196	2808	.006

Figure 31: Regression Model 1 (Sourced from Katos et al. (2020))



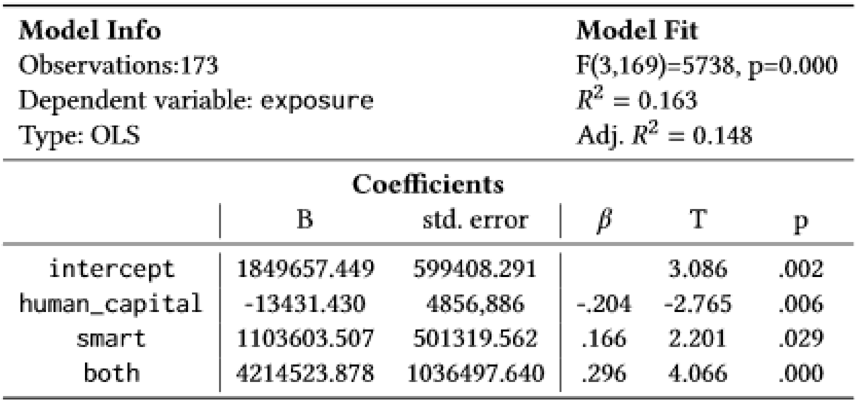
Model Info	**Model Fit**
Observations:171	F(6,164)=16290, p=0.000
Dependent variable: exposure	$R^2 = 0.372$
Type: OLS	Adj. $R^2 = 0.349$

Coefficients

	B	std. error	β	T	p
population	.190	.040	.414	4746	.000
technology	-19471.425	4977.282	-.543	-3.912	.000
motion	1706746.558	621010.315	.278	2.748	.007
smart	2175288.130	496238.743	.405	4.384	.000
plain	2180776.000	763190.887	.220	2.857	.005
both	5468921.448	975307.816	.364	5.607	.000

Figure 32: Regression Model 2 (Sourced from Katos et al. (2020))

Figure 33 shows the results of a backward stepwise regression with exposure being the dependent variable; the initial dependent variables are `human_capital`, `social_cohesion`, `technology` defined in the CIMI as well as three classes of cities: **plain, smart** and **both**. The **Circular** class was omitted due to the small number of cities. This OLS model was run four times when every time one of the variables was removed. The final models included the variables `human_capital`, **smart** and **both**. The results show that smart cities, cities that adopt a smart or circular agenda and human capital are significant in predicting vulnerability exposure. We can learn from observing the β coefficients that smart and circular cities are more vulnerable than smart cities that are currently not adopting any circular approach showing that smart and circular cities are more likely to be technologically "hungry" which highlights the data-driven nature of CE but simultaneously being more exposed.



Model Info	**Model Fit**
Observations:173	F(3,169)=5738, p=0.000
Dependent variable: exposure	$R^2 = 0.163$
Type: OLS	Adj. $R^2 = 0.148$

Coefficients

	B	std. error	β	T	p
intercept	1849657.449	599408.291		3.086	.002
human_capital	-13431.430	4856,886	-.204	-2.765	.006
smart	1103603.507	501319.562	.166	2.201	.029
both	4214523.878	1036497.640	.296	4.066	.000

Figure 33: Regression Model 3 (Sourced from Katos et al. (2020))

We can also provide additional context and actionable threat intelligence by combining our CVEs CVSS scores to the list of the top 10 cities with the biggest exposure rate from smart and both classes of cities. This data allows cities to prioritize the severity of their vulnerabilities and mitigate them.

Figure 34 shows the top 10 cities which are most vulnerable by their severity rating and it can be seen that there are a few vulnerabilities that are rated low, there appears to be a similar amount of vulnerabilities rated medium and high and a smaller amount of critical vulnerabilities which should be prioritized. Figure 35 shows the same as figure 34 but instead is CVSS 2.0 rather than 3.0. There are more vulnerabilities rated lower within this figure and more medium than there is high which cities can decide whether to prioritize the lesser number of vulnerabilities but more severe or more vulnerabilities but less severe.

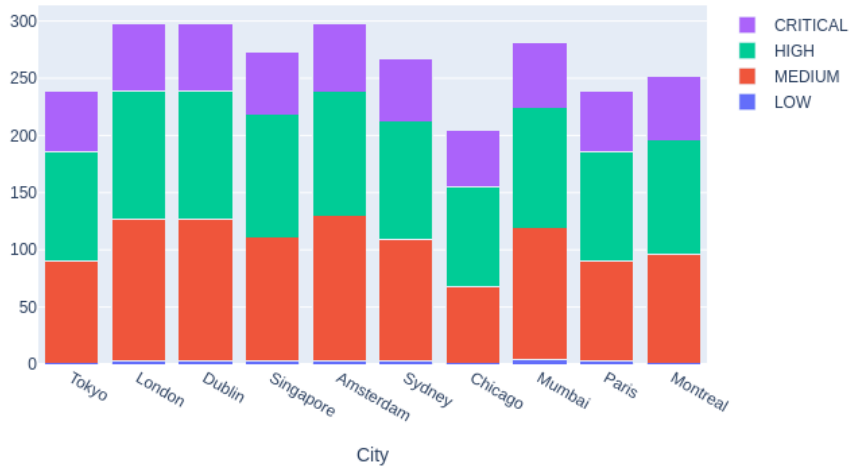


Figure 34: Top 10 Vulnerable Cities by CVSS3.0 Severity (class:`Smart & Both`)

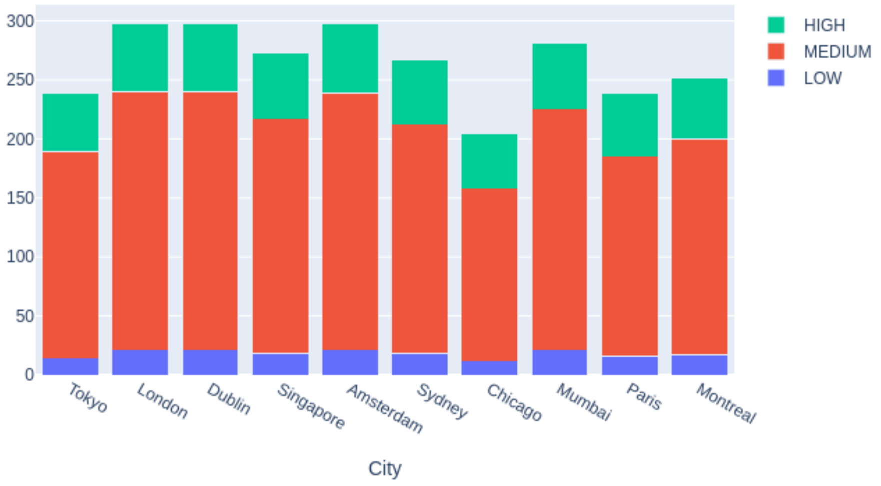


Figure 35: Top 10 Vulnerable Cities by CVSS2.0 Severity (class:`Smart & Both`)

4.3 Countries

We also had a substantial data set involving exposure data for specific countries, this included data from the IMD WDCI as well as other external pieces of information like the Gini coefficient values, this is because we deduced that a countries GDP and wealth has a positive relationship in a countries vulnerability footprint and relates to a city's cybersecurity investment and capacity (Creese et al. (2020)). This data helps facilitate a contextualize the state of vulnerabilities and can an excellent reference and baseline to build a pragmatic cyber situational awareness capacity. We explore the following hypotheses:

H_1 Countries with a higher population are more vulnerable than those with a lower population.

H_2 Countries that scored higher in the WDCI should be less exposed than cities that scored lower.

H_3 Exposure rates for countries decrease with GDP per capita.

H_3 Vulnerability severity decreases with GDP per capita.

When investigating H1 we look at two factors: the country's population and the exposure rate. Figure 36 shows the top 10 populated countries and figure 37 shows the top 10 countries with the highest exposure rates. We can recognize from both charts that China, the USA, Brazil, and Russia are on both the top populated and top exposed list.

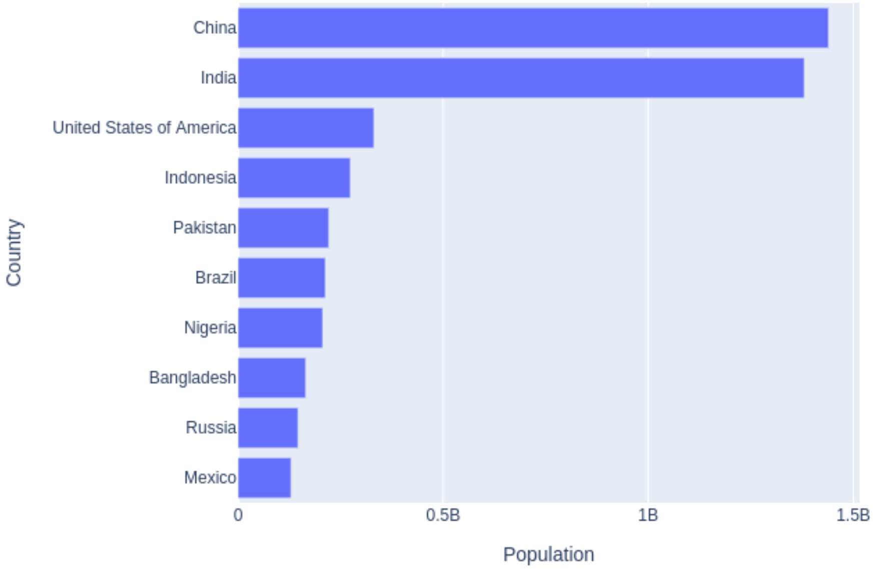


Figure 36: Top 10 Populated Countries

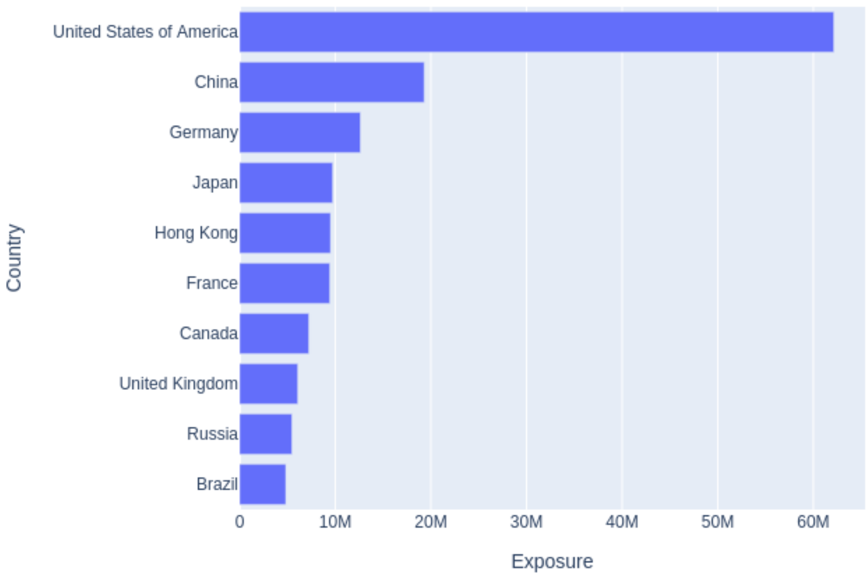


Figure 37: Top 10 Exposed Countries

When investigating the hypothesis H_2 we look at countries included in the WDCI and the factors it provides us like the country's overall performance and any factor relating to technology. Figure 38 shows both the exposure rate for countries which scored less than and greater than **25.0** from the WDCI. Countries that scored lower than **25.0** have significantly higher exposure rates than countries that scored higher than **25.0**. From this, we can see that more advanced countries in every factor from the WDCI are more vulnerable to hacking than those less advanced. The fact that more advanced countries are more dependent on technology might be the cause of almost double the exposure rates.

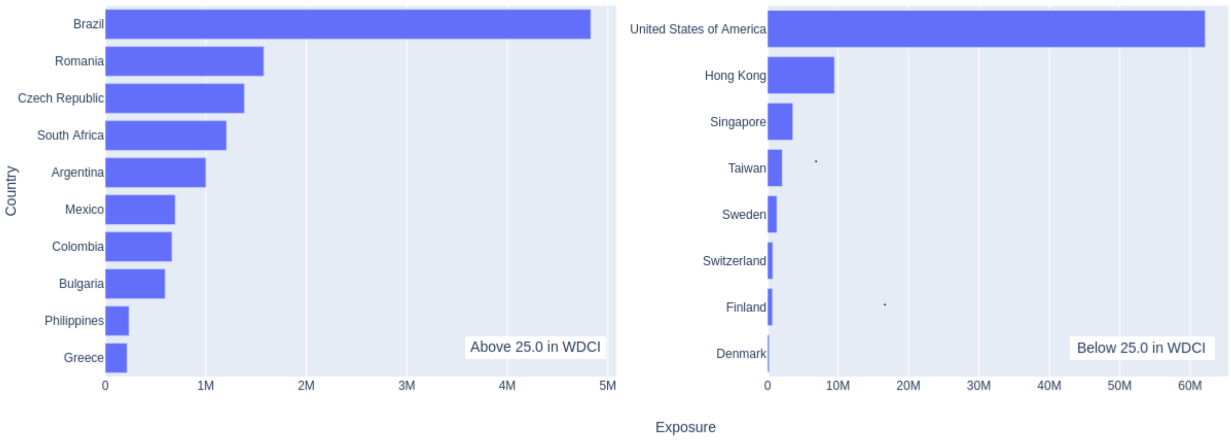


Figure 38: Exposure for Countries Scored Below and above 25.0

When investigating H_3, we look at the IMD WDCI and their data for countries with a greater and less than $20,000 GDP per capita. We take the cities and put both lists into separate data sets and produce two bar graphs that show the exposure rates for cities greater than and less than to show if there is any correlation between cities GDP per capita and how vulnerable they are to hacking. Figure 39 shows the top 10 countries with the biggest exposure profile with a GDP per capita greater and lesser than $20,000. We can observe that there is a positive relationship between a higher GDP per capita and how exposed they are.

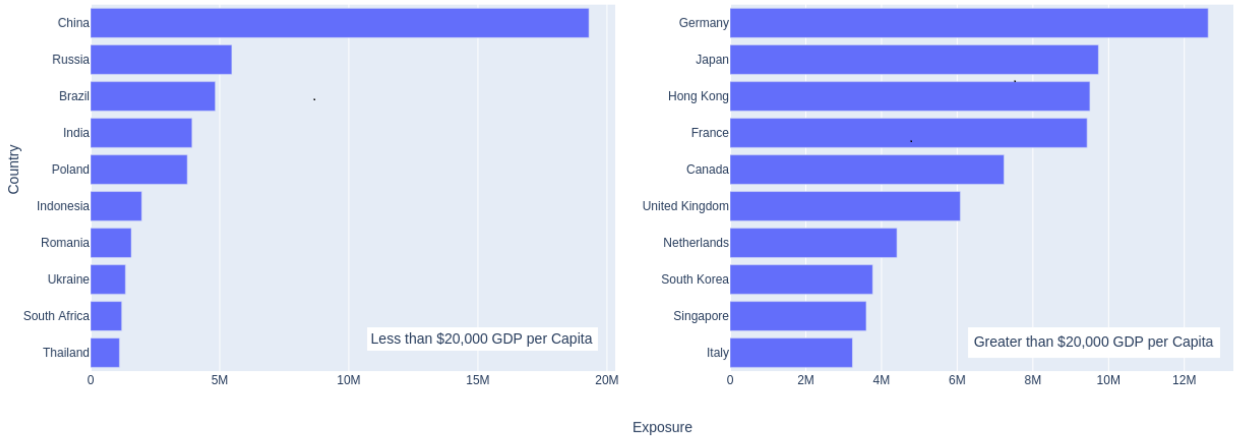
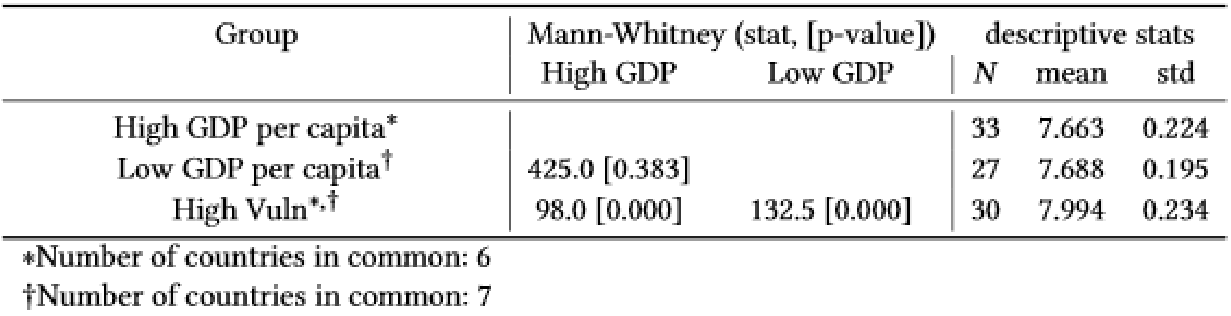


Figure 39: Top 10 Exposed Countries with GDP per capita greater and lesser than $20,000

Figure 40 shows a pairwise comparison of means of three different groups of countries: countries with high earning GDP per capita, countries with low earning GDP per capita and the top 30 countries with the highest average CVSSv3 base score. We included the latter group to explore the positioning of the low GDP ranked countries in finer granularity. From the results, we can see that countries with a low GDP per capita have a slightly higher average base score than countries with a high GDP per capita. On the other hand, the most statistically significant fact is that the countries with the highest average base score have a higher mean than the countries with low and high-income GDP. We can also observe that low GDP per capita countries has a mean score of 7.688 whereas countries with a high GDP per capita have a mean score of 7.663 showing that there is a slightly higher severity in lower GDP per capita countries.



Group	Mann-Whitney (stat, [p-value])		descriptive stats		
	High GDP	Low GDP	N	mean	std
High GDP per capita*			33	7.663	0.224
Low GDP per capita†	425.0 [0.383]		27	7.688	0.195
High Vuln*,†	98.0 [0.000]	132.5 [0.000]	30	7.994	0.234

∗Number of countries in common: 6

†Number of countries in common: 7

Figure 40: Comparison of means of GDP groups (Sourced from Katos et al. (2020))

5 CONCLUSION

5.1 Hypotheses

We discovered that cities that declared solely a circular agenda or had not shown any smart or circular initiate had a relatively smaller exposure rate to those declared smart by the CIMI or the SCI. We also discovered that trend that with an increase of a population followed by an increased exposure profile. Our hypothesis H_3 was a fascinating one, we anticipated that a city declared smarter and advanced by either the CIMI or SCI would have a much lower exposure profile than those that scored lower. However, it was found these cities had a tremendous amount more exposure than cities that scored lower; this makes sense though, as even though we expected a smarter and more advanced city to protect itself more against potential cyber-attacks and malicious threat actors, a city which depends on more technology has a wider attack surface than those that don't.

As mentioned in the previous hypothesis section, we expected that a higher population would mean a higher exposure than those with a smaller population; in this case, it was the same when analysing countries. We found almost all countries with a huge population were also those that had the highest exposure profile. Just like our city hypothesis H_3, our country hypothesis H_2 is also alike. We looked at how well a country performed in the WDCI and expected that a country considered more advanced could protect itself more. However, the same thing happened, a country more advanced had a bigger exposure profile than those considered less advanced. We also expected a country's exposure profile would decrease with an increase of GDP per capita, which again was an intriguing result. We spotted an outlier which is China which is four times the amount of the nearest country. Other than China, it seems countries have a greater exposure profile with a greater than $20,000 GDP per capita which could be chalked up to a much greater GDP meaning a country is more advanced and from this research, we know that more advanced countries or cities are more exposed.

5.2 Future Work

If we were to carry on this project in the future, there would be a few changes that I would like to make regarding the way I went about this project. The first one would be the method of collection; Shodan's CLI limited us to execute a single script at one time, so if we were able to increase the number of scripts executed then we could cut down the amount of time it takes for the data to be collected. Another improvement we could make is to make the swap over to Python from Bash in our collection scripts as mentioned in the limitation of our project section, our scripts can take days to complete however I believe Python could be a much more effective and faster method to collect our data. I would also like to make an addition to our data set and collect certain pieces of data from Shodan that might provide valuable information like linking protocols to specific cities.

5.3 Limitations

A noteworthy limitation that we faced during this project was the extreme volatility of the data. It was clear that when we first started to collect our data, the volatility would be an issue which we would have to consider during our analysis. So for the purpose of this project, our data is considered as a snapshot of time which isn't an issue as this data could still be used to explore a relationship. A prime example of this limitation is CVE-2019-9637 which affected up to four million devices then suddenly only a couple thousand and back up to millions.

A second noteworthy limitation of this research project was the fact that there we were unable to gather any information that sheds a light on the type of ownership for a vulnerable device. This means we observe a city as a whole rather than distinguishing the different sectors like critical infrastructure or simple households. This information is crucial when looking at a city's vulnerability footprint as cities with their critical infrastructure vulnerable would suffer more than a city with more households vulnerable in the case of a cyber-attack.

A third limitation that we faced was some nonexistent or obsolescent statistics. This shows mostly when collecting the Gini coefficient data for countries; smaller countries like Trinidad and Tobago Gini values were from 1992 with no recent number. Other smaller countries like Wallis and Futuna have provided no official value for their Gini. This also occurs when looking at data related to vulnerabilities, some common attack patterns and weaknesses have not been decided for our specific common vulnerabilities which suggest they have not been around long, but they are accepted as the standard industry when talking about a vulnerability in the cyber security community.

A fourth limitation that we faced in this project was our **plain** class of cities. We had good reasons why certain cities were placed in the other types of classes such as cities within the CIMI and SCI were placed in the **smart** or **both** and cities that declared a circular economy agenda was placed within the **circular** class but when it comes to our **plain** class we had no proper criteria or reason why we placed cities within this class so we ended up removing over 100 plain cities that were not as known as the ones we kept in.

Word count: 9400

References

Centre, N. C. S., 2020. Connected places cyber security principles. Available From: https://www.ncsc.gov.uk/collection/connected-places-security-principles [Accessed 10 May 2021].

Intelligence-Driven Urban Internet-of-Things Ecosystems for Circular, S. and CITIES, I. S., 2019. D2.1: Circular economy models for smart city assets. Available From: https://www.ideal-cities.eu/wp-content/uploads/2019/10/IDEAL-CITIES-D2.1.pdf [Accessed 4 February 2021]].

Intelligence-Driven Urban Internet-of-Things Ecosystems for Circular, S. and CITIES, I. S., nd. Ideal-cities. Available From: https://www.ideal-cities.eu/ [Accessed 2 February 2021].

Cities, C., 2021. C40. Available From: https://www.c40.org/about [Accessed 2 February 2021].

Creese, S., Dutton, W., Esteve-Gonzalez, P. and Shillair, R., 2020. Cybersecurity capacity building: Cross-national benefits and international divides. URL AvailableFrom:https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3658350.

for Cybersecurity, E. U. A., 2005. About enisa - the european union agency for cybersecurity. Available From: https://www.enisa.europa.eu/about-enisa [Accessed 13 January 2021].

Foundation, E. M., 2017. How can we embed circular economy principles to build thriving, liveable, and resilient cities?. Available From: https://www.ellenmacarthurfoundation.org/our-work/activities/circular-economy-in-cities [Accessed 1 February 2021].

IMD, 2020a. Imd world digital competitiveness ranking 2020. Available From: https://www.imd.org/wcc/world-competitiveness-center-rankings/world-digital-competitiveness-rankings-2020/ [Accessed 2 February 2021].

IMD, 2020b. Smart city index 2020. Available From: https://www.imd.org/globalassets/wcc/docs/smart_-city/smartcityindex_2020.pdf [Accessed 1 February 2021].

IMD, 2020c. Smart city index 2020 by imd business school. Available From: https://www.imd.org/smart-city-observatory/smart-city-index/ [Accessed 1 February 2021].

IMD, 2021. Imd world digital competitiveness ranking 2020. Available From: https://www.imd.org/smart-city-observatory/smart-city-index/ [Accessed 2 February 2021].

Johnson, P., Lagerstrom, R., Ekstedt, M. and Franke, U., 2016. Can the common vulnerability scoring system be trusted? a bayesian analysis. Available From: https://ieeexplore.ieee.org/document/7797152 [Accessed 7 May 2021].

Jupyter, nd. Project jupyter. Available From: https://jupyter.org/ [Accessed 4 February 2021].

Katos, V., Jarvis, P. and Damianou, A., 2020. Vulnerability exposure driven intelligence in smart, circular cities. Available From: DTRAP [Accessed 13 April 2021].

Khoshgoftar, M. and Osman, O., 2009. Comparison of maturity models. *2009 2nd IEEE International Conference on Computer Science and Information Technology*.

Langley, D., Doorn, J., Ng, I., Stieglitz, S., Lazovik, A. and Boonstra, A., 2021. The internet of everything: Smart things and their impact on business models. *Journal of Business Research*, 122. URL `https://www.sciencedirect.com/science/article/pii/S014829631930801X`.

MITRE, 2015. Att&ck. Available From: https://attack.mitre.org/ [Accessed 10 February 2021].

MITRE, nda. Capec. Available From: https://capec.mitre.org/ [Accessed 10 February 2021].

MITRE, ndb. Capec - capec-22: Exploiting trust in client (version 3.4). Available From: https://capec.mitre.org/data/definitions/22.html [Accessed 28 February 2021].

MITRE, ndc. Cvss v3 equations. Available From: https://nvd.nist.gov/vuln-metrics/cvss/v3-calculator/v30/equations [Accessed 13 April 2021].

MITRE, ndd. Cwe. Available From: https://cwe.mitre.org/about/index.html [Accessed 10 February 2021].

MITRE, nde. National vulnerability database. Available From: https://nvd.nist.gov/ [Accessed 1 February 2021].

Nations, U., 2018. United nations department of economic and social affairs. Available From: https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html [Accessed 13 January 2021].

Nations, U., 2020. Undata app. Available From: https://data.un.org/en/iso/sg.html [Accessed 2 February 2021].

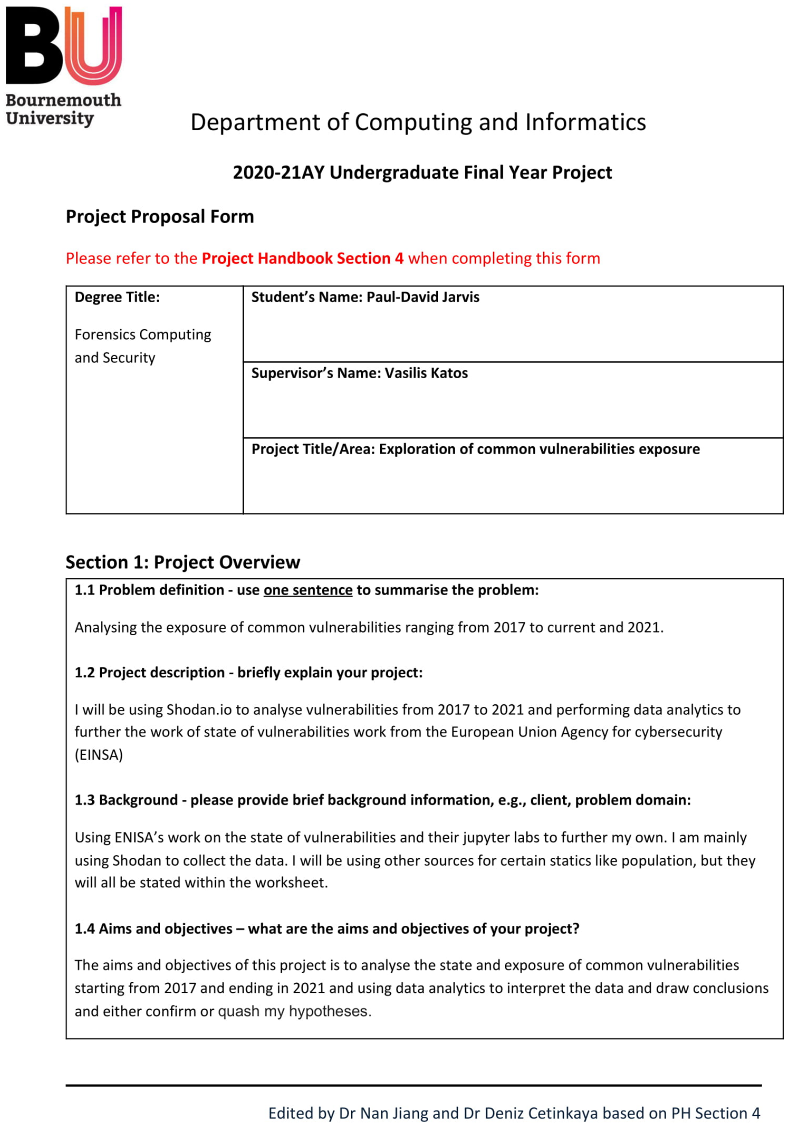
School, I. B., 2020. Iese cities in motion index 2020 — cities in motion. Available From: https://blog.iese.edu/cities-challenges-and-management/2020/10/27/iese-cities-in-motion-index-2020 [Access 2 February 2021].

Shodan, 2013. Shodan.io. Available From: https://www.shodan.io/ [Accessed 13 January 2021].

Zaharis, N., nd. Part b: Valorisation plan. Available From: http://intervalue.urenio.org/valorisation/smartctmodel/ip-protection/ [Accessed 1 February 2021].

Zygiaris, S., 2012. Smart city reference model: Assisting planners to conceptualize the building of smart city innovation ecosystems. *Journal of the Knowledge Economy*, 4.

Appendix A - Project Proposal



Bournemouth University

Department of Computing and Informatics

2020-21AY Undergraduate Final Year Project

Project Proposal Form

Please refer to the **Project Handbook Section 4** when completing this form

Degree Title: Forensics Computing and Security	**Student's Name: Paul-David Jarvis**
	Supervisor's Name: Vasilis Katos
	Project Title/Area: Exploration of common vulnerabilities exposure

Section 1: Project Overview

1.1 Problem definition - use <u>one sentence</u> to summarise the problem:

Analysing the exposure of common vulnerabilities ranging from 2017 to current and 2021.

1.2 Project description - briefly explain your project:

I will be using Shodan.io to analyse vulnerabilities from 2017 to 2021 and performing data analytics to further the work of state of vulnerabilities work from the European Union Agency for cybersecurity (EINSA)

1.3 Background - please provide brief background information, e.g., client, problem domain:

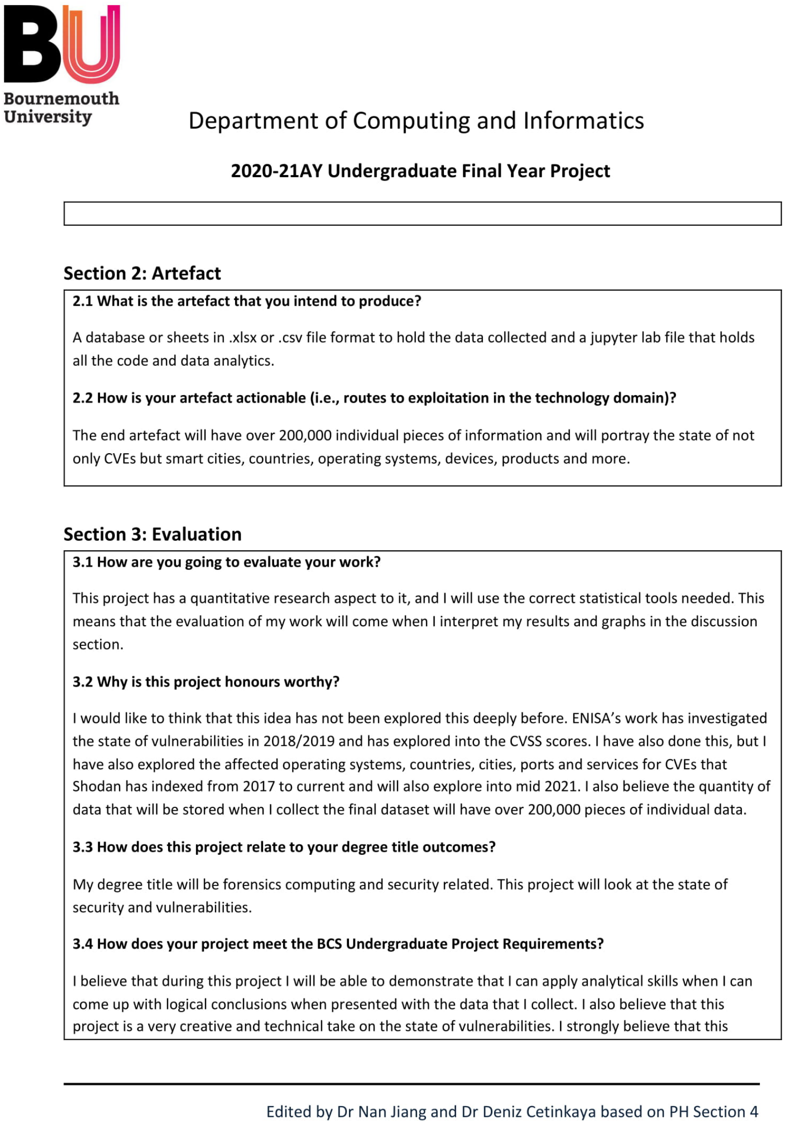
Using ENISA's work on the state of vulnerabilities and their jupyter labs to further my own. I am mainly using Shodan to collect the data. I will be using other sources for certain statics like population, but they will all be stated within the worksheet.

1.4 Aims and objectives – what are the aims and objectives of your project?

The aims and objectives of this project is to analyse the state and exposure of common vulnerabilities starting from 2017 and ending in 2021 and using data analytics to interpret the data and draw conclusions and either confirm or quash my hypotheses.

Edited by Dr Nan Jiang and Dr Deniz Cetinkaya based on PH Section 4

Figure 41: First Page of Project Proposal.



Bournemouth University

Department of Computing and Informatics

2020-21AY Undergraduate Final Year Project

Section 2: Artefact

2.1 What is the artefact that you intend to produce?

A database or sheets in .xlsx or .csv file format to hold the data collected and a jupyter lab file that holds all the code and data analytics.

2.2 How is your artefact actionable (i.e., routes to exploitation in the technology domain)?

The end artefact will have over 200,000 individual pieces of information and will portray the state of not only CVEs but smart cities, countries, operating systems, devices, products and more.

Section 3: Evaluation

3.1 How are you going to evaluate your work?

This project has a quantitative research aspect to it, and I will use the correct statistical tools needed. This means that the evaluation of my work will come when I interpret my results and graphs in the discussion section.

3.2 Why is this project honours worthy?

I would like to think that this idea has not been explored this deeply before. ENISA's work has investigated the state of vulnerabilities in 2018/2019 and has explored into the CVSS scores. I have also done this, but I have also explored the affected operating systems, countries, cities, ports and services for CVEs that Shodan has indexed from 2017 to current and will also explore into mid 2021. I also believe the quantity of data that will be stored when I collect the final dataset will have over 200,000 pieces of individual data.

3.3 How does this project relate to your degree title outcomes?

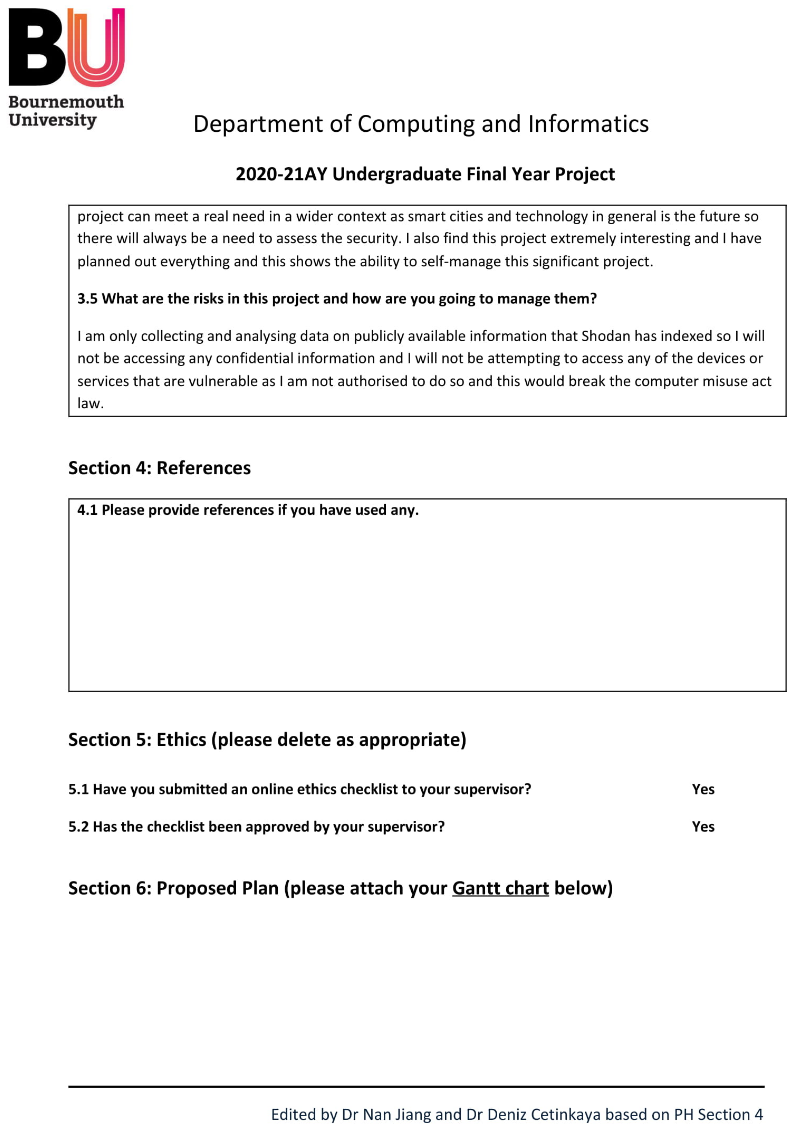
My degree title will be forensics computing and security related. This project will look at the state of security and vulnerabilities.

3.4 How does your project meet the BCS Undergraduate Project Requirements?

I believe that during this project I will be able to demonstrate that I can apply analytical skills when I can come up with logical conclusions when presented with the data that I collect. I also believe that this project is a very creative and technical take on the state of vulnerabilities. I strongly believe that this

Edited by Dr Nan Jiang and Dr Deniz Cetinkaya based on PH Section 4

Figure 42: Second Page of Project Proposal.



Department of Computing and Informatics

2020-21AY Undergraduate Final Year Project

project can meet a real need in a wider context as smart cities and technology in general is the future so there will always be a need to assess the security. I also find this project extremely interesting and I have planned out everything and this shows the ability to self-manage this significant project.

3.5 What are the risks in this project and how are you going to manage them?

I am only collecting and analysing data on publicly available information that Shodan has indexed so I will not be accessing any confidential information and I will not be attempting to access any of the devices or services that are vulnerable as I am not authorised to do so and this would break the computer misuse act law.

Section 4: References

4.1 Please provide references if you have used any.

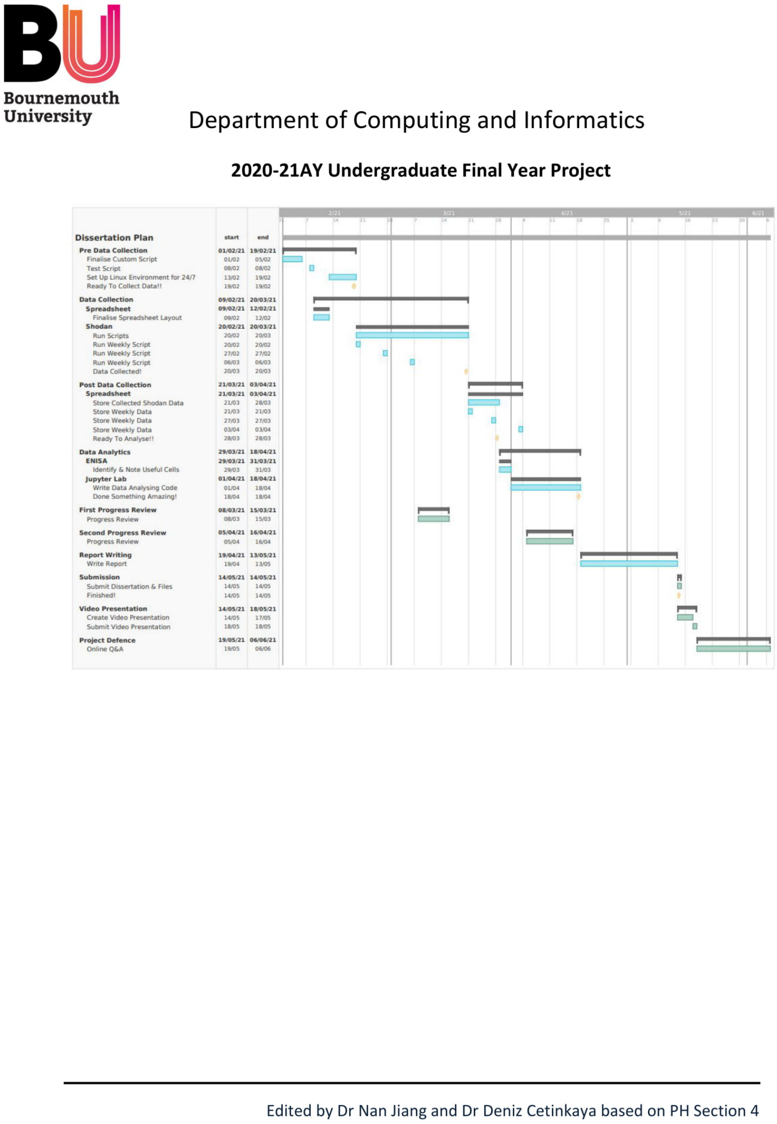
Section 5: Ethics (please delete as appropriate)

5.1 Have you submitted an online ethics checklist to your supervisor?	**Yes**
5.2 Has the checklist been approved by your supervisor?	**Yes**

Section 6: Proposed Plan (please attach your Gantt chart below)

Edited by Dr Nan Jiang and Dr Deniz Cetinkaya based on PH Section 4

Figure 43: Third Page of Project Proposal.



Bournemouth University

Department of Computing and Informatics

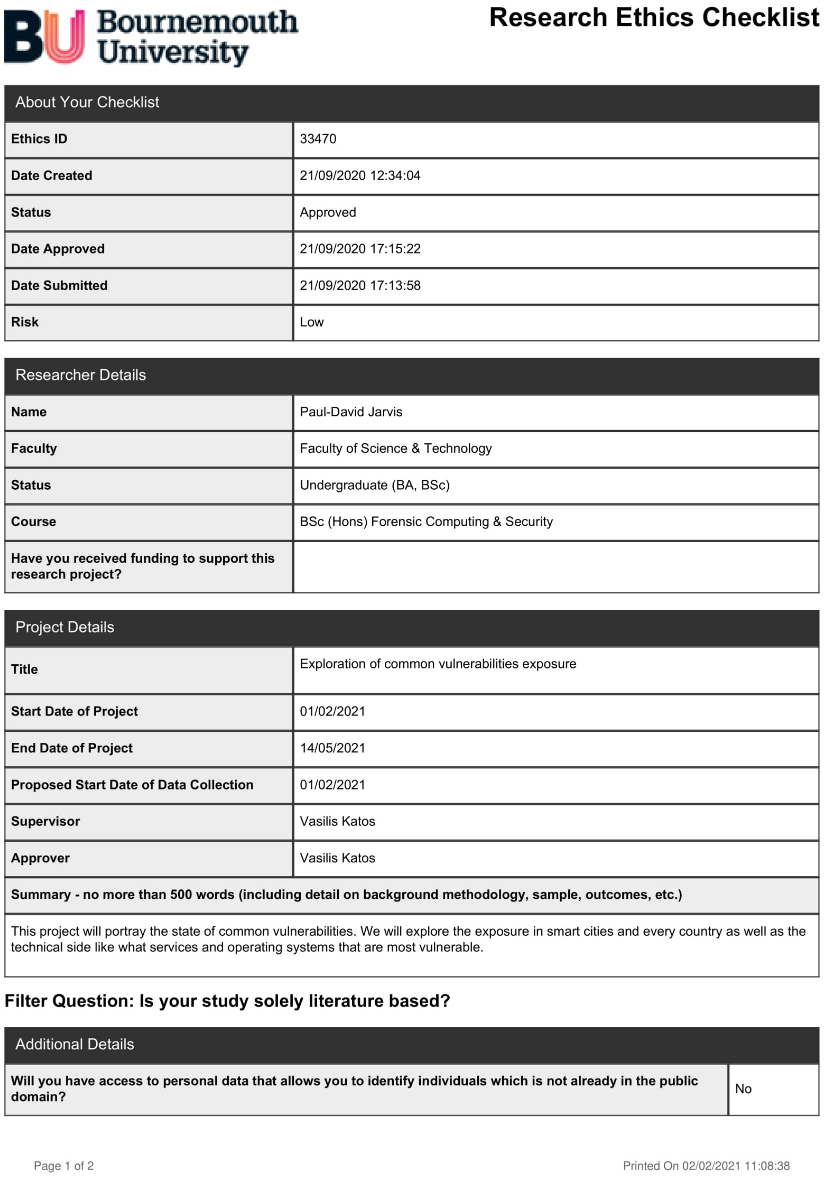
2020-21AY Undergraduate Final Year Project

Dissertation Plan	start	end
Pre Data Collection	**01/02/21**	**19/02/21**
Finalise Custom Script	01/02	05/02
Test Script	08/02	08/02
Set Up Linux Environment for 24/7	13/02	19/02
Ready To Collect Data!!	19/02	19/02
Data Collection	**09/02/21**	**20/03/21**
Spreadsheet	**09/02/21**	**12/02/21**
Finalise Spreadsheet Layout	09/02	12/02
Shodan	**20/02/21**	**20/03/21**
Run Scripts	20/02	20/03
Run Weekly Script	20/02	20/02
Run Weekly Script	27/02	27/02
Run Weekly Script	06/03	06/03
Data Collected!	20/03	20/03
Post Data Collection	**21/03/21**	**03/04/21**
Spreadsheet	**21/03/21**	**03/04/21**
Store Collected Shodan Data	21/03	28/03
Store Weekly Data	21/03	21/03
Store Weekly Data	27/03	27/03
Store Weekly Data	03/04	03/04
Ready To Analyse!!	28/03	28/03
Data Analytics	**29/03/21**	**18/04/21**
ENISA	**29/03/21**	**31/03/21**
Identify & Note Useful Cells	29/03	31/03
Jupyter Lab	**01/04/21**	**18/04/21**
Write Data Analysing Code	01/04	18/04
Done Something Amazing!	18/04	18/04
First Progress Review	**08/03/21**	**15/03/21**
Progress Review	08/03	15/03
Second Progress Review	**05/04/21**	**16/04/21**
Progress Review	05/04	16/04
Report Writing	**19/04/21**	**13/05/21**
Write Report	19/04	13/05
Submission	**14/05/21**	**14/05/21**
Submit Dissertation & Files	14/05	14/05
Finished!	14/05	14/05
Video Presentation	**14/05/21**	**18/05/21**
Create Video Presentation	14/05	17/05
Submit Video Presentation	18/05	18/05
Project Defence	**19/05/21**	**06/06/21**
Online Q&A	19/05	06/06

Edited by Dr Nan Jiang and Dr Deniz Cetinkaya based on PH Section 4

Figure 44: Fourth Page of Project Proposal.

Appendix B - Ethics Checklist



Bournemouth University

Research Ethics Checklist

About Your Checklist	
Ethics ID	33470
Date Created	21/09/2020 12:34:04
Status	Approved
Date Approved	21/09/2020 17:15:22
Date Submitted	21/09/2020 17:13:58
Risk	Low

Researcher Details	
Name	Paul-David Jarvis
Faculty	Faculty of Science & Technology
Status	Undergraduate (BA, BSc)
Course	BSc (Hons) Forensic Computing & Security
Have you received funding to support this research project?	

Project Details	
Title	Exploration of common vulnerabilities exposure
Start Date of Project	01/02/2021
End Date of Project	14/05/2021
Proposed Start Date of Data Collection	01/02/2021
Supervisor	Vasilis Katos
Approver	Vasilis Katos

Summary - no more than 500 words (including detail on background methodology, sample, outcomes, etc.)

This project will portray the state of common vulnerabilities. We will explore the exposure in smart cities and every country as well as the technical side like what services and operating systems that are most vulnerable.

Filter Question: Is your study solely literature based?

Additional Details	
Will you have access to personal data that allows you to identify individuals which is not already in the public domain?	No

Page 1 of 2

Printed On 02/02/2021 11:08:38

Figure 45: First Page of Ethics Checklist.

Will you have access to confidential corporate or company data (that is not covered by confidentiality terms within an agreement or separate confidentiality agreement)?	No

Storage, Access and Disposal of Research Data	
Once your project completes, will any anonymised research data be stored on BU's Online Research Data Repository "BORDaR"?	Yes

Page 2 of 2

Printed On 02/02/2021 11:08:38

Figure 46: Second Page of Ethics Checklist.

Appendix C - Artifacts

- **table_of_contents.ipynb**: Includes our table of contents, data legend, related work and capture dates.
- **1_vulnerabilities.ipynb**: Includes all work involving our list of CVEs.
- **2_cities.ipynb**: Includes all work involving our list of cities.
- **3_countries.ipynb**: Includes all work involving our list of countries
- **4_operatingsystems.ipynb**: Includes all work involving our list of operating systems.
- **5_ports.ipynb**: Includes all work involving our list of ports.
- **6_products.ipynb**: Includes all work involving our list of products.
- **7_tags.ipynb**: Includes all work involving our list of tags.
- **8_protocols.ipynb**: Includes all work involving our list of protocols.
- **9_historical_numbers.ipynb**: Includes all work involving our list of CVEs and their historical numbers.
- **Data Set**: A Excel spreadsheet including all our data we collected.
- **Scripts**: Shodan scripts written in bash used to scrape our data automatically.

Appendix D - Shodan Scripts

D.1 Facets Scripts

These Scripts return the top 10 results for a specific CVE. These are ran instead of a count script as these queries are very niche so to save time it will return definite results meaning it executes a single time for a single CVE rather than executing ten times for a specific CVE. This just means that you have to manually go through the extract the data.

D.1.1 Tag

```
#!/bin/bash
while read CVEs
    do echo $CVEs
        shodan stats --facets tag vuln:$CVEs
        sleep 1
done < CVEs
```

D.1.2 Products

```
#!/bin/bash
while read CVEs
    do echo $CVEs
        shodan stats --facets product vuln:$CVEs
        sleep 1
done < CVEs
```

D.1.3 Devices

```
#!/bin/bash
while read CVEs
    do echo $CVEs
        shodan stats --facets device vuln:$CVEs
        sleep 1
done < CVEs
```

D.1.4 Operating Systems

```
#!/bin/bash
while read CVEs
    do echo $CVEs
        shodan stats --facets os vuln:$CVEs
        sleep 1
done < CVEs
```

D.2 Count Scripts

These scripts return a total count for the written query. These scripts require the list of security flaws identified with CVE numbers and other input files depending on what script is being executed. These scripts will return a number higher than zero.

The cities script is extremely long as it reads a file for every country that we analysed to ensure that the correct cities exposure data is being returned instead of another city in a completely different country such as Athens in Greece, United States or Canada. It also has to be executed within the same folder that includes all the cities input and the list of CVEs.

D.2.1 CVEs

```
#!/bin/bash
while read CVEs; do
  echo $CVEs
  shodan count vuln:$CVEs
  sleep 1
done < CVEs
```

D.2.2 Countries

```
#!/bin/bash
while read CVEs
    do echo $CVEs
    while read CountryCodes
        do echo $CountryCodes
        shodan count Country:$CountryCodes vuln:$CVEs
        sleep 1
        done < CountryCodes
done < CVEs
```

D.2.3 Ports

```
#!/bin/bash
while read CVEs
    do echo $CVEs
    while read Ports
        do echo "Port: $Ports"
        shodan count port:"$Ports" vuln:$CVEs
        sleep 1
        done < Ports
done < CVEs
```

D.2.4 Protocols

```
#!/bin/bash
echo "BACnet"
while read CVEs
        do echo $CVEs
        shodan count port:47808 vuln:$CVEs
done < CVEs

echo "CODESYS"
while read CVEs
        do echo $CVEs
        shodan count port:2455 operating system vuln:$CVEs
done < CVEs

echo "Crimson V3.0"
while read CVEs
        do echo $CVEs
        shodan count port:789 product:"Red Lion Controls" vuln:$CVEs
done <  CVEs
```

```
echo "DNP3"
while read CVEs
        do echo $CVEs
        shodan count port:20000 source address vuln:$CVEs
done < CVEs

echo "EtherNet/IP"
while read CVEs
        do echo $CVEs
        shodan count port:44818 vuln:$CVEs
done < CVEs

echo "Factory Interface Network Service"
while read CVEs
        do echo $CVEs
        shodan count port:9600 response code vuln:$CVEs
done < CVEs

echo "Fox"
while read CVEs
        do echo $CVEs
        shodan count port:1911,4911 product:Niagara vuln:$CVEs
done < CVEs

echo "Highway Addressable Remote Transducer Protocol"
while read CVEs
        do echo $CVEs
        shodan count port:5094 hart-ip vuln:$CVEs
done < CVEs

echo "IEC 60870"
while read CVEs
        do echo $CVEs
        shodan count port:2404 asdu address vuln:$CVEs
done < CVEs
```

```
echo "MELSEC-Q"
while read CVEs
        do echo $CVEs
        shodan count port:5006,5007 product:mitsubishi vuln:$CVEs
done < CVEs

echo "Modbus"
while read CVEs
        do echo $CVEs
        shodan count port:502 vuln:$CVEs
done < CVEs

echo "PCWorx"
while read CVEs
        do echo $CVEs
        shodan count port:1962 PLC vuln:$CVEs
done < CVEs

echo "ProConOS"
while read CVEs
        do echo $CVEs
        shodan count port:20547 vuln:$CVEs
done < CVEs

echo "S7 Communication"
while read CVEs
        do echo $CVEs
        shodan count port:102 vuln:$CVEs
done < CVEs

echo "Service Request Transport Protocol"
while read CVEs
        do echo $CVEs
        shodan count port:18245,18246 product:"general eletric" vuln:$CVEs
```

```
done < CVEs
```

D.2.5 Cities

```
#!/bin/bash
while read CVEs
    do echo $CVEs

    while read Argentina
        do echo $Argentina
        shodan count city:$Argentina vuln:$CVEs country:"AR"
        sleep 1
        done < Argentina

    while read Australia
        do echo $Australia
        shodan count city:$Australia vuln:$CVEs country:"AU"
        sleep 1
        done < Australia

    while read Austria
        do echo $Austria
        shodan count city:$Austria vuln:$CVEs country:"AT"
        sleep 1
        done < Austria
```

... Shorted for brevity. ...