Cairo University
Faculty of Economics and Political Science
Statistics Department
English Section

The Impact of Climate Change on Covid-19 Spread-out:

A Case Study on G-20's Countries.

A Graduation Project Thesis

Submitted to Statistics Department, English Section, Faculty of Economics and Political Science

By:

Nadine Essam Farouk

Alyiaa Assem Ramadan Morsy

Yomna Abdullatif Ibrahim Osman

Ghofran Talaat Sayed Muhammad

Supervised by:

Dr. Rasha Elsouda

June-2022

Table of contents

Contents

Table of Abbreviations

Abbreviation	Explanation
PM10	Particulate Matter (10)
PM2.5	Particulate Matter (2.5)
NO2	Nitrogen dioxide
SO2	Sulfur dioxide
CO	Carbon monoxide
O3	Ozone

Chapter One

Introduction

Abstract

Over the past decades, the whole world was suffering from climate change and considered it a disastrous crisis. Meanwhile, the world has recently suffered from the repercussions of COVID-19, some people claimed that climate change has played a role in the COVID-19 emerging disaster, and others say the opposite. This study aims to show and determine the effect of three main variables of Global Warming and climate change (Temperature, Humidity, and Ozone gas) on the COVID-19 spread and its Daily New Cases. In this context, a case study is made on eight countries from G-20 countries using panel data analysis techniques. The paper begins by addressing the above-mentioned research question and its main precise objective, then moving on to the literature review part where it discusses the previously made studies concerning this issue; afterward the data collection procedure has been tackled to proceed on to the descriptive part. Thereafter, panel data analysis has been made to demonstrate the results and answer the research question. The results of the analysis yield that the climate change variables don't have any impact on the spread of COVID-19. Therefore, the study has reached a conclusion that leads to the denial of the effect of the climate change variables on the increase of the COVID-19 daily new cases.

Overview

Covid-19 is defined as a severe acute respiratory syndrome and with the rise in the number of confirmed cases global economy, social networks, global transportation and air quality are affected by this pandemic. The number of confirmed cases has grown in most affected countries on March 11, 2020, WHO announced the outbreak as a pandemic situation.

Air pollution also is responsible for the third number of deaths in the world from stroke, lung cancer and heart disease while ⅔ of exposures is caused by burning of fossil fuels that also cause climate change WHO (2022). We don't have enough evidence that climate change is influencing the spread of Covid-19 but we know that it matters for our health**.** Covid-19 produce major changes in air quality as some environmental factors like temperature and humidity contributes significantly to decrease growth rate infection C-CHANGE | Harvard T.H. Chan School of Public Health. **(**2022**)**.

On a global average basis, a 34% reduction in NO_2 concentration were estimated during lockdown period (until April 30, 2020) and global average of O_3 concentration increased by 86% at the same period, NO_2 was the pollutant most affected by covid-19 a pandemic as its emissions were from the sources that are restricted by lock down. Torkmahalleh et al. (2021).

Exposure to these pollutants can cause oxidative stress and systemic inflammation that reduces the host's immunity against any kind of infections which provide quality care thus it is essential to study the changes in atmosphere during Covid-19. It is reported that the cold climates are more favorable for the spread of the virus which helps it to make similar cluster patterns in colder region that it temperature lies between 5-11 degree Celsius and low absolute humidity (4-7 g/m^3) as rising temperature by 1 degree lead to a 0.86 decrease in cumulative cases that equals 13-17 cases per day which proves that climate conditions could regulate the transmission of it rapidly spreading**.** Wang et al. (2020)

So, to control Covid-19 spread, countries reduce their economic activity that lead to temporary improvement in air quality as the percentage of pollutants decreases in the atmosphere. To help limit the risk of infectious diseases we should reduce air pollution caused by burning fossil fuels like coal, oil and natural gas which also helps keep our lungs healthy, and protect us from respiratory infections like coronavirus. Climate change and pollution is one of the main causes of problems related to the respiratory system as a new Harvard T.H. Chan School of Public Health study confirms.

We know that African American communities are disproportionately exposed to air pollution and we're now seeing this pollution driving higher mortality rates from COVID-19. We owe it to everyone to improve health, and we do that by reducing the sources of pollution that drive a large burden of disease around the world.

The actions we need to combat climate change are the same actions we need to make people healthier right now, especially for diseases causing huge burdens on our health like Covid-19 and its mutants. C-CHANGE | Harvard T.H. Chan School of Public Health. (2022).

Many parallels exist between climate change and the Covid-19 pandemic. However, in recent months, concerns have grown about the fact that the health emergency has put many climate adaptation and mitigation policies on hold during the pandemic. In countries where climate action is critical, the current pandemic has had a negative impact on short-term climate policy planning.

Climate change is expected to be the most dangerous problem humanity has ever faced. However, some authors argue that, in recent months, concerns about climate change have been put on hold in some countries due to the urgency of the ongoing Covid-19 pandemic. For example, during the second wave of Covid-19 in Brazil, a significant number of pieces of legislation that relaxes environmental laws, ranging from easing forest protections to declassifying pesticide toxicity, were approved .During this time, the US has reversed some environmental regulations and appears to be directing stimulus funds toward reviving the fossil fuel industry. Meanwhile, the German Council of Economic Experts published a report on the coronavirus crisis that omitted any mention of environmental concerns or the terms climate change or sustainability. The Columbia Climate School also mentions the postponement of COP-26 and international environmental negotiations, which could allow countries to shift their actions away from combating climate change. Delays in climate-friendly policies and related investments have been noted by the International Energy Agency. Furthermore, as a result of the heavy use of plastics and private transportation, some private companies have increased waste and pollution production.

Many people noticed immediate reductions in carbon emissions around the world during the first months of the Covid-19 pandemic, which turned out to be the short-term effects of the rapid decline in economic activity. By June 2020, China's air pollution levels had returned to pre-pandemic levels, though the World Meteorological Organization reported that overall carbon dioxide levels in the atmosphere had increased in 2020 compared to 2019.

Covid-19 pandemic is likely to disrupt efforts to address climate-related challenges for a long time, especially in middle- and low-income countries that have been hit the hardest by the effects of climate change. Extreme weather events have been more frequent, intense, and widespread than ever before Climate-related disasters threaten to overwhelm local health systems, which are already under severe strain. When a major disaster is combined with a pandemic, the costs of damage and recovery are estimated to be up to 20% higher than normal, and annual averages of disaster-related losses could nearly double in a worst-case climate scenario. Covid-19 and climate change have affected more than 50 million people around the world, according to the Red Cross. Madagascar's recent famine, the first solely due to climate change, has magnified the pandemic's health and economic effects, resulting in a humanitarian crisis. Extreme weather events have dominated headlines around the world in the spring and summer of 2021. In Turkey, Greece, the western U.S and Canada record-breaking heat waves have sparked devastating wildfires, which are worsening the Covid-19 cases due to the plummets of the air quality. USGLC (2022). Covid-19's economic costs are likely to make it impossible to meet current climate financing targets. The World Economic Forum estimates that $5.7 trillion in annual funding is required for effective climate change mitigation and adaptation, but the International Monetary Fund estimates that Covid-19 will cost the global economy more than $28 trillion in output over the next five years. Covid-19, according to the Eurasia Group, will cause a shift in attention and resources, which, while necessary for ending the pandemic, will put climate change "on the backburner". USGLC (2022).

The 2020 UN Climate Change Conference (COP26) was postponed by a full year to November 2021, after Covid-19 derailed global climate diplomacy. The talks will serve as a deadline for countries to submit more stringent emission reduction proposals. The battles against climate change and the pandemic are inextricably linked.

Climate change mitigation is also a public health strategy because many of the root causes of climate change also increase the risk of pandemics. Many countries, including the United Kingdom, have restricted travel and other activities in response to the Covid-19 pandemic in 2020 and 2021. As a result, global greenhouse gas emissions and local air pollution were temporarily reduced. Since the mid-nineteenth century, greenhouse gas concentrations in the atmosphere have been rising. These gasses form a blanket around the planet once they reach the atmosphere, trapping heat from the sun and raising global temperatures. Aerosols, which are microscopic particles suspended in the atmosphere, have the potential to influence the Earth's climate. Man-made aerosols are primarily produced by pollution from automobiles and factories, and they're often produced naturally. We looked at results from several studies to see if the temporary reduction in greenhouse gasses and aerosol pollutants during the pandemic had any impact on our climate.

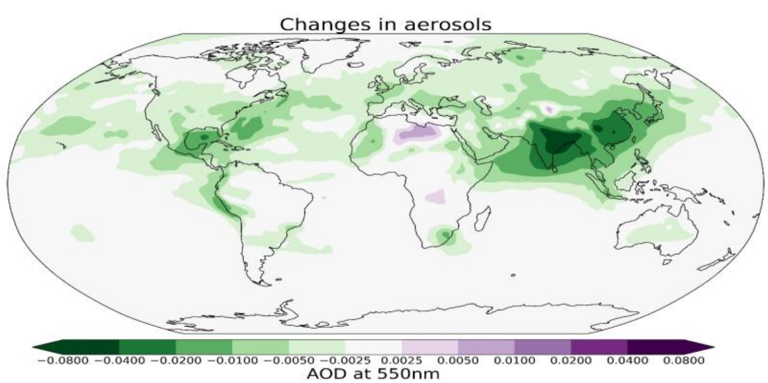

Figure 1: The map shows the average change in aerosols across the Earth system Met Office (2022).

Regions with lower levels of aerosols in the atmosphere are shaded green, while regions with higher levels of aerosols are shaded purple.

The findings show that aerosol amounts were reduced for 2020, especially over southern and eastern Asia. As a result, the amount of solar radiation reaching the surface in that area increased; however, this was insufficient to alter the climate. The annual temperature and rainfall in that region, as well as globally, did not change significantly. Because gasses like carbon dioxide have such a long lifetime in the atmosphere, changes in emissions only have a minor impact. While a 7% reduction in emissions is remarkable, it still means that 93% of our normal emissions were released into the atmosphere, and carbon dioxide levels continued to rise. The models were used to simulate the climate for the five-year period from 2020 to 2024, assuming that emissions reductions would last for two years before returning to previous levels. All the previous mentioned results can be seen through the next figures. Met Office (2022).

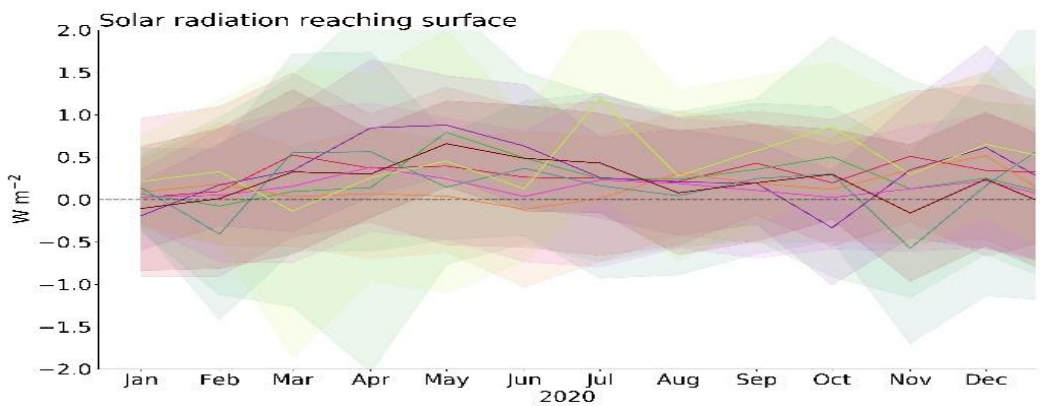

Figure 2: solar radiation reaching the surface. Met Office (2022).

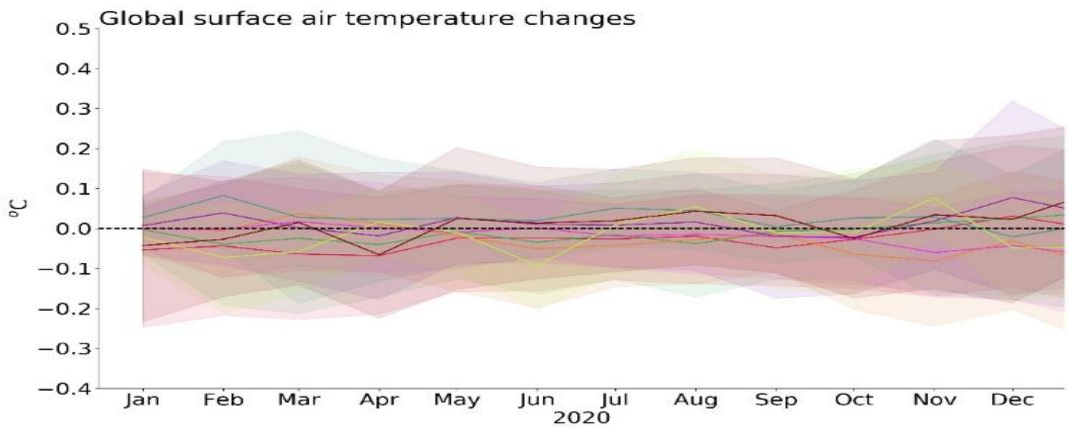

Figure 3: global surface air temperature changes. Met Office (2022).

The shaded regions show the spread of results when each model is run multiple times, and each colored line shows the results from one model.

Objectives and Research Questions

Our main objective from this study is to examine the relationship between Covid-19 and climate change in some of the G-20's countries in a way that helps us to represent and examine almost most of the continents. We are going to do that in a way that helps us to answer our main research question which is:

Can Climate affect Covid-19?

Study Problem

Nowadays , we are being threatened not just by the pandemic and its spread and quick evolution but also by the quick climate changes and its impact on the environment , where we found that all were caused by the precautions taken and how humans are dealing with both crises, so we are going to analyze the relationship between the Covid-19 and climate change during the period 2020-2022 on eight countries which are: United kingdom, United states, Saudi Arabia, China , Brazil , Italy, South Africa and Australia to be representatives of the most of the world's continents, that also found to be members of the G20 which is a 20-country intergovernmental forum that includes the European Union (EU). It focuses on major global economic issues such as international financial stability, climate change mitigation, and long-term development. We will try to study the case of the relationship between the climate changes and Covid-19 reaching findings that could clear out the situation so we could make recommendations in order to have a better future for us and our environment.

Chapter Two

Data Collection and Methodology

Literature review

Addas and Maghrabi (2021), in an integrative review paper aimed to systematically assess the bibliographic review on the impact of lockdowns on air quality on the global level. At first, 237 studies related to the impact of COVID-19 lockdowns on air quality were collected, and only 144 studies were finally taken into account for this literature review since they met the review criterion which was “The Preferred Reporting Items for Systematic Reviews and Meta-Analyses” (PRISMA). The literature was extracted from different places like, Scopus, Google Scholar, PubMed, Web of Science, and the Google search engine. According to Addas and Maghrabi core findings, it appears that there was substantial improvement in air quality due to lockdowns happening due to Covid-19 across the world that provided us with an opportunity to realize the impact of anthropogenic pressures on the environment. Thus, the findings of this review assist the planners and policymakers to understand that the implementation of lockdowns may be an effective and important measure to restore the environment, and to build quality ecosystems in urban environments.

In a study conducted by Masum and Pal (2020) the aim was to evaluate statistically the impacts of COVID-19 lockdown on selected air quality pollutants and air quality indexes on Chittagong city (City in Bangladesh) from 26 March to 26 April 2020. The daily average concentrations of air pollutants PM10, PM2.5, NO2, SO2 and CO Chittagong city during COVID-19 lockdown were evaluated statistically and then compared with dry season data averaging over 8 years from 2012 to 2019. During lockdown, all pollutants studied showed a statistically significant decreasing trend except NO2; a reduction of 40%, 32% and 13% compared to the daily mean concentrations of these previous dry seasons were observed for PM2.5, PM10 and NO2 respectively. The improvement in air quality index value was 26% in comparison to the previous dry season because of the less human activities during this period. The factor analysis approved that AQI in Chittagong city is largely influenced by PM10 and PM2.5 during the period of COVID-19 shutdown.

There is a paper discussing a study made by Torkmahalleh et al. (2021), it basically tackles the assessments of the impacts of COVID-19 lockdowns on ground-level PM2.5, NO2, and O3 concentrations on a global scale. The obtained data was from 34 countries, 141 cities, and 458 air monitoring stations on 5 continents where on a global average basis, a 34.0% reduction in NO2 concentration and a 15.0% reduction in PM2.5 were estimated during the strict lockdown period (until April 30, 2020). Global average O3 concentration increased by 86.0% during this same period. Worldwide, NO2 was the pollutant most affected by the COVID-19 pandemic and this because its emissions were from sources that were typically restricted by the lockdowns and due to its short lifetime in the atmosphere. The results show that lockout measures and consequent lower emissions lowered exposure to the majority of dangerous pollutants, potentially yielding in global health benefits. However, the increased O3 could have significantly diminished those benefits, necessitating more extensive health examinations to correctly measure the benefits. At the same time, these limits came at a high cost in terms of money and other health concerns (depression, suicide, spousal abuse, drug overdoses, etc.). As a result, any corresponding reductions in air pollution would have to be achieved without the imposed activity cutbacks' substantial economic and social effects.

Another study was conducted by Dang (2020), discussing cross national evidence on the causal impacts of COVID-19 on air pollution. This happened through assembling a rich database consisting of daily, sub-national level data of air quality for 178 countries before and after the COVID-19 lockdowns. Using a Regression Discontinuity Design technique, the paper also investigates their effects on air quality. The lockdowns resulted in huge reductions in global air pollution. These findings are similar across air quality indicators and data sources, as well as model specifications. Some limited evidence suggests that countries with a higher percentage of commerce and manufacturing in their economies or with lower levels of air pollution at the start of the lockdowns experience lower levels of air pollution after the lockdowns, but the converse is true for countries near the equator. Improved air quality could also be explained by mobility constraints imposed as a result of the lockdowns.

According to a research paper made by Sarmadi et al. (2021), in spite of the world's economic, social, and health disruptions, the Coronavirus Disease 2019 (COVID-19) pandemic gave a chance for the environment to minimize ambient pollution. The goal of this research was to look at how air quality indexes (AQI) changed in industrial, heavily populated, and capital cities around the world before and after 2020. The World Air Quality Index and other free resources were used to calculate the AQI (WAQI). The relationships between meteorological and AQI variables were investigated using bivariate correlation analysis. The paired-sample t-test or Wilcoxon signed-rank test were used to compare mean differences (standard deviation: SD) of AQI parameters from different years. To identify meteorological variables that affect the AQI values, a multivariable linear regression analysis was used. The data showed that air quality increased in 2020 for all pollutants except carbon monoxide and ozone; however, the trend was reversed in 2021, possibly due to the relaxation of some countries' limitations. Although this improvement in quality was only transient, it is an important result for pollution management plans.

In a study conducted by Roy (2020), it was discussed whether the global temperature had any role in the spread and vulnerability to COVID-19, and how to use the knowledge demonstrated in overcoming the spread of the coronavirus disease. The study highlighted that for transmitting the virus, global temperature played an important role and a moderately cool environment was the most favorable one, the risk was significantly lesser for warm places and countries. The maximum reported case, as well as death, was found in countries when the temperature was between around 275 K (2 °C) to 290 K (17 °C). The study identified that USA and UK belong to the previously mentioned countries category. For temperature 300 K (27 °C) and above, a significantly lesser degree of vulnerability was reported. Countries from Asia as Saudi Arabia and China belong to this category. This work discusses how countries can switch from one fragile state to another based on changes in temperature. The research that was done into how temperature affects viruses provided valuable insights that suggest controlling the temperature can be very effective in stopping an outbreak. With that knowledge, some simple and effective solutions are proposed, which are without side effects and very cost-effective too.

A review was published by Mecenas et al. (2020), it mainly addresses the description of the current knowledge about the emergence and replicability of the coronavirus and its relationship with different weather factors such as temperature and relative humidity. The review stated that based on the limited data and evidence, it seems that the spread of COVID-19 is less common in warm and wet climates. Temperature and humidity alone do not usually cause such great variability in outbreaks of COVID-19. Public isolation policies, herd immunity, migration patterns, population density, and cultural aspects may play a role in how the spread of this disease occurs. The understanding of the health policies' effects on the weather is invaluable for the benefit of humanity in this time of crisis.

There is a study conducted previously by Evans et al. (2020). The study stated the social-distancing measures taken to combat the COVID-19 pandemic have led to a significant reduction in air pollutant emissions. It is difficult to quantify the changes due to air pollution because it varies depending on the synoptic and seasonal patterns. They used a machine learning algorithm to assess changes in nitrogen dioxide (NO2) and ozone (O3) at 5,756 observation sites in 46 countries from January through June 2020. The response of surface ozone levels is complicated by competing influences of atmospheric chemistry. While surface O3 levels in some locations increased by up to 50%, they found the overall net impact on daily average O3 levels between February - June 2020 to be small. The analysis showed that the O3 diurnal cycle is flattening, with more nighttime ozone and less daytime ozone. This is likely due to a reduction in photochemical production. The O3 response is dependent on season, time scale, and environment, with declines in surface O3 forecasted if NO2 emission reductions continue.

Data Collection

It is worth mentioning that we are going to work on two datasets; one will be related to the air quality and climate change during the Covid-19 period and the other is mainly related to Covid-19 spread-out globally. As a result, we are going to mention the data collection procedures we went through and how our data were collected.

A. Air Quality Data

Data were collected from The World Air Quality Index project which is a non-profit project started in 2007. Its main mission is to promote air pollution awareness for citizens and provide unified and worldwide air quality information and with the Covid-19 spreading out all over the world, the World Air Quality Index (WAQI) project team saw a surge in requests for global data covering the whole world map. As a result, they, now, are providing a new dedicated data-set, updated 3 times a day, and covering about 380 major cities in the world, from January 2020 until now. As a result, we requested their data to work on it in our project.

Mainly, we are working on some of the G-20's countries; to be exact only 8 major countries represent most of the world's continents. The countries are:

- **The United States**: a country in North America.
- **United Kingdom**: a country in Europe.
- **China**: is a country in East Asia.
- **Saudi Arabia**: a country in Western Asia.
- **Brazil**: is a country in South America.
- **Australia**: a continent/ country in Oceania.
- **Italy**: a country in Europe.
- **South Africa**: a country in Africa.

Why have we chosen these 8 countries specifically?
This is an intuitive question that has to be answered first.

➢ **China/ KSA**
(In Asia)
Since Asia is Earth's largest and most populous continent, we took two countries a try to represent it

➢ **South Africa**
(In Africa)
Since Africa is the world's second-largest and second-most-populous continent, we have chosen South Africa to represent it. Taking into consideration that we have chosen only one country from Africa since we have to choose a country that has two characteristics as follows:

1. The country has to have a membership in the G-20.
2. The country has to be in Africa.

- **Brazil**

  (In South America)

  Since South America is a continent entirely in the Western Hemisphere and mostly in the Southern Hemisphere, with a relatively small portion in the Northern Hemisphere; we have chosen the largest country in South America (Brazil) to represent it.

- **USA**

  (In North America)

  Since North America is a continent in the Northern Hemisphere and almost entirely within the Western Hemisphere, it can be described as the northern subcontinent of a single continent called America; we have chosen the second largest country on the continent to represent it.

- **UK/ Italy**

  (In Europe)

  Since Europe as a continent recognized as a part of Eurasia, located entirely in the Northern Hemisphere and mostly in the Eastern Hemisphere. Comprising the westernmost peninsulas of Eurasia, it shares the continental landmass of Afro-Eurasia with both Asia and Africa, we have chosen UK and Italy in order to represent it.

- **Australia**

  (In Oceania or sometimes considered as a continent)

  Since Australia is the largest country by area in Oceania and the world's sixth-largest country, we have chosen it to represent Oceania.

The WAQI project collected the data from each of the eight countries separately and by city. Hence, we are going to mention the data provider of each of these countries /cities. It is worth mentioning that we are going to work on all the countries' cities to fully represent them.
From the WAQI dataset, we have determined three variables for the eight countries to be detected and analyzed along with their relationships with COVID-19's spread to fully answer our research question, and they are as follows:

1. Temperature.
2. Humidity.
3. Ozone gas (O3).

It was observed that the dataset includes various descriptive values for each of the previously mentioned variables i.e., max, min, count, median, and variance. So, we have taken the median as a representative value for each of them (the reason will be mentioned later in this chapter; in the data description section). Then, it was highly observed that there is more than one value for the same date of the daily recorded data for different cities within the same country, hence, we were subject to use the R software to compute the median of the medians for each variable within the same country for the same date to have only one value for each variable on each day from the 1st of March, 2020 till that of October, 2021. This resulted in having 580 values for each variable, with unrepeated daily data in each country. This is a better form for the data to be well-representative of the whole country and be easily used, analyzed, and interpreted in the form of panel data along with that of the two variables (Daily New Cases and Active Cases) from the COVID-19 dataset.

WAQI project and its data sources:

Firstly, USA's Data:

There are about 57 cities in USA, all the daily data for each of them will be considered, and in the following, the data centres and sources for the cities' data in USA will be displayed:

- Oklahoma Department of Environmental Quality: www.deq.state.ok.us/
- Air Now - US EPA: www.airnow.gov/
- North Carolina Department of Environmental Quality: deq.nc.gov/
- Tennessee Department of Environment & Conservation: tn.gov/ENVIRONMENT
- Arkansas Department of Environmental Quality (ADEQ): www.adeq.state.ar.us/
- Mississippi Department of Environmental Quality: www.deq.state.ms.us/
- Louisiana Department of Environmental Quality: deq.louisiana.gov/
- Massachusett Office of Energy and Environmental Affair: www.mass.gov/eea/agencies/massdep/air/
- Air Quality in Virginia: www.deq.virginia.gov/
- Washington State Department of Ecology: www.ecy.wa.gov/
- Oregon Department of Environmental Quality (DEQ): www.oregon.gov/DEQ/
- Idaho Department of Environmental Quality: airquality.deq.idaho.gov/
- Texas Commission on Environmental Quality (TCEQ): www.tceq.texas.gov/
- State of Hawaii, Department of Health, Clean Air Branch: health.hawaii.gov/cab/
- CARB - California Air Resources Board: www.arb.ca.gov/
- Illinois Environmental Protection Agency: www.epa.illinois.gov/
- Wisconsin Department Of Natural Resources: airquality.wi.gov/
- South Carolina Department of Health and Environmental Control: www.scdhec.gov/
- Connecticut Department of Energy and Environmental Protection (DEEP): ct.gov/deep/cwp/view.asp?a=2684&q=321758
- Air Now - US EPA - Maryland state: www.mde.state.md.us/Pages/Home.aspx
- Indiana Department of Environmental Management: www.in.gov/idem/
- Georgia Department of Natural Resources, Environmental Protection Division, Air Protection Branch: epd.georgia.gov/air/
- State of Rhode Island - Department of Environmental Management: www.dem.ri.gov/
- New York State Department of Environmental Conservation (NYSDEC): www.dec.ny.gov/
- NJDEP/DAQ - New Jersey Department of Environmental Protection - Division of Air Quality: www.nj.gov/dep/daq/
- South Coast Air Quality Management District (AQMD): www.aqmd.gov/
- Michigan Department of Environmental Quality: www.michigan.gov/deq

- Bureau of Air Quality, Pennsylvania's Department of Environmental Protection: www.dep.pa.gov/
- Arizona Department of Environmental Quality - Air Quality Division: www.azdeq.gov/
- Air Monitoring - Maricopa County Air Quality Department: www.maricopa.gov/
- Nebraska Department of Environmental Quality: www.deq.state.ne.us/
- New Mexico Environment Department: www.env.nm.gov/aqb/
- Florida Department of Environmental Protection: floridadep.gov/air/
- Air Quality in Clark County: airquality.clarkcountynv.gov/
- Ohio Environmental Protection Agency: www.epa.state.oh.us/
- Minnesota Pollution Control Agency: www.pca.state.mn.us/
- Colorado Department of Environmental Protection: www.colorado.gov/airquality/
- UTAH department of environmental quality: deq.utah.gov/division-air-quality
- Alabama Department of Environmental Management: www.adem.state.al.us/default.cnt
- Pima County Department of Environmental Quality Air: pima.gov/

Secondly, UK's Data:

There are about 21 cities in UK, all the daily data for each of them will be considered, and in the following, the data centres and sources for the cities' data in UK will be displayed:

- Air Quality in Scotland - latest data, forecasts and air quality information: www.scottishairquality.scot/
- UK-AIR, air quality information resource - Defra, UK: uk-air.defra.gov.uk/
- Norfolk Air Quality Monitoring Network: www.norfolkairquality.net/
- Northern Ireland Air: www.airqualityni.co.uk/
- Air Quality in Wales: airquality.gov.wales/
- London Air Quality Network - Environmental Research Group, King's College London: londonair.org.uk/
- Herts & Bed Air Quality Monitoring Network: www.hertsbedsair.net/
- Heathrow Airwatch: www.heathrowairwatch.org.uk/
- Great Air Manchester: www.greatairmanchester.org.uk/
- Air Quality in Kent and Medway: www.kentair.org.uk/

Thirdly, China's Data:

There are about 52 cities in China, all the daily data for each of them will be considered, and in the following, the data centres and sources for the cities' data in China will be displayed:

- Beijing Environmental Protection Monitoring Centre: www.bjmemc.com.cn/
- Hebei Province Environment Protection Agency: www.hebei.gov.cn/

- U.S Embassy Beijing Air Quality Monitor: china.usembassy-china.org.cn/embassy-consulates/beijing/air-quality-monitor/
- Guangdong Environmental Protection public network: gdee.gd.gov.cn/
- Yunnan Environmental Protection Agency: sthjt.yn.gov.cn/
- Zhejiang Environmental Protection Bureau: sthjt.zj.gov.cn/
- Chongqing Environmental Protection Bureau: www.cepb.gov.cn/
- Environmental Protection Department of Shandong Province: sthj.shandong.gov.cn/
- Hainan Environmental Protection Agency: www.dloer.gov.cn/
- Urumqi Environmental Protection Agency: www.wlmqhb.gov.cn/
- Helongjiang Environmental Protection Agency: www.hljdep.gov.cn/
- Guiyang Municipal Environmental Protection Bureau: www.ghb.gov.cn/
- Shenzhen Environment Network: meeb.sz.gov.cn/
- Henan Environmental Protection Agency: sthjt.henan.gov.cn/
- Ningxia Environmental Protection Agency: sthjt.nx.gov.cn/
- China National Urban air quality real-time publishing platform: 113.108.142.147:20035/emcpublish/
- Liaoning Provincial Environmental Monitoring: sthj.ln.gov.cn
- Tibet Autonomous Region Environmental Protection Agency: ee.xizang.gov.cn/
- Shanghai Environment Monitoring Centre: sthj.sh.gov.cn/
- U.S. Consulate Shanghai Air Quality Monitor: china.usembassy-china.org.cn/embassy-consulates/shanghai/air-quality-monitor-stateair/
- Environmental Protection Agency of Jilin Province: sthjt.jl.gov.cn/
- Guangxi Zhuang Autonomous Region Environmental Protection Agency: sthjt.gxzf.gov.cn/
- Anhui Environmental Protection Agency: sthjt.ah.gov.cn/
- Chengdu Environmental Protection Agency: sthj.chengdu.gov.cn/
- Sichuan Province Environmental Protection Agency: sthjt.sc.gov.cn/
- U.S. Consulate Chengdu Air Quality Monitor: china.usembassy-china.org.cn/embassy-consulates/chengdu/air-quality-monitor/
- Jiangsu Environmental Protection public network: hbt.jiangsu.gov.cn/
- Nanjing Air Quality Distribution System: 222.190.111.117:8023/
- Shaanxi Provincial Environmental Protection Office: sthjt.shaanxi.gov.cn
- Xi'an Environmental Protection Agency: xaepb.xa.gov.cn/
- Hunan Environmental Protection Agency: sthjt.hunan.gov.cn/
- Wuhan Environmental Protection Bureau: www.whepb.gov.cn/
- Hubei Environmental Protection Agency: www.hbepb.gov.cn/
- Tianjin Environmental Monitoring Centre: www.tjemc.org.cn/
- Jiangsu Province PM2.5 Air Monitoring Commission: www.jshb.gov.cn/jshbw/
- Jiangxi Province Environmental Protection Agency: www.jxepb.gov.cn/
- Gangsu Environmental Protection Agency: sthj.gansu.gov.cn/
- Gangsu Environmental Protection Agency: www.gsep.gansu.gov.cn/

- Shanxi Province Environmental Monitoring Centre: sthjt.shanxi.gov.cn/
- U.S. Consulate Guangzhou Air Quality Monitor: china.usembassy-china.org.cn/embassy-consulates/guangzhou/u-s-consulate-air-quality-monitor-stateair/
- Guangzhou Environmental Protection Bureau: sthjj.gz.gov.cn/
- Environment Monitoring Centre of Ningbo: hbj.ningbo.gov.cn/
- Environment Monitoring Centre of Ningbo: www.nbemc.gov.cn/
- Department of Ecology and Environment of Fujian Province: sthjt.fujian.gov.cn/

Fourthly, Saudi Arabia's data:

There are about 7 cities in KSA, all the daily data for each of them will be considered, and in the following, the data centers and sources for the cities' data in KSA will be displayed:

- Saudi Arabia General Authority for Meteorology and Environmental Protection: www.pme.gov.sa/ar/.
- Dhahran Air Quality Monitor - US Embassy: sa.usembassy.gov/.

Fifthly, Brazil's data:

Brazil is composed of 3 major cities, which from the data sources have obtained the daily data for the weather status. The two data sources are as follows:

- CETESB - Companhia Ambiental do Estado de São Paulo: www.cetesb.sp.gov.br/.
- IEMA - Instituto Estadual de Meio Ambiente Recursos Hídricos: iema.es.gov.br/qualidadedoar/indicedequalidadedoar.

Sixthly, Australia's data:

Australia is well-known to be a continent and a country at the same time. There are 11 major cities in Australia, and the data sources work there in collecting the data from each of these cities on a daily basis. The data sources can demonstrated as the following:

- Office of Environment and Heritage - NSW: www.environment.nsw.gov.au/
- EPA Tasmania - Tasmania Environment Protection Authority: epa.tas.gov.au/epa/air
- NTEPA - Northern Territory Environment Protection Authority: www.ntepa.nt.gov.au/waste-pollution/air
- Environment Protection Authority | EPA Victoria: epa.vic.gov.au/
- Australian Capital Territory (ACT)
- Health: www.health.act.gov.au/about-our-health-system/population-health/environmental-monitoring/monitoring-and-regulating-air
- EPA South Australia:: Air quality in South Australia: www.epa.sa.gov.au/data_and_publications/air_quality_monitoring

- Air quality | Environment, land and water | Queensland Government: www.qld.gov.au/environment/pollution/monitoring/air
- Western Australia Air Quality Management Branch: www.der.wa.gov.au/

Seventhly: Italy's data:

Italy is a country in Europe composed of 12 cities from which the data was collected. The data sources are distributed among the cities, and they can be displayed as follows:

- Arpa Emilia-Romagna: www.arpae.it/
- ARPAT - Agenzia regionale per la protezione ambientale della Toscana: www.arpat.toscana.it/
- ARAP Piemonte - Agenzia Regionale per la Protezione dell'aria in Piemonte): www.sistemapiemonte.it/ambiente/srqa/
- Agenzia Regionale per la Protezione Ambientale in Campania: www.arpacampania.it/
- European Environment Agency: www.eea.europa.eu/themes/air/
- ARAP Lazio - Agenzia Regionale per la Protezione dell'Ambiente del Lazio): www.arpalazio.net/
- Agenzia Regionale per la Protezione dell'Ambiente della Lombardia): ita.arpalombardia.it/ITA/

Eighthly, South Africa's data:

South Africa is the only country in Africa to be enrolled the G-20 countries, it has 12 major cities. The weather data is collected throughout the following two main data sources:

- South African Air Quality Information System - SAAQIS: saaqis.environment.gov.za.
- RBCAA - Richards Bay Clean Air Association: www.rbcaa.org.za/.

B. Covid-19 Spread-out Data

The data set was obtained from "Kaggle" and this data was mainly scraped from "Worldometers" on (2022-01-05) by Joseph Assaker. It includes a data for 218 countries and all countries have records dating from 2020-2-15 to 2022-01-05, but we are going to deal with the data of only eight countries of them; USA, UK, China, KSA, Brazil, Australia, Italy, and South Africa on the same period that was mentioned for the climate data. It is worth mentioning that Covid-19 data have many variables but we are going to deal with only one of them; the Daily New Cases. Therefore, the usage of panel data and its analysis will be a milestone since we have to differentiate between climate change through time and its variations among the previously mentioned countries along with their COVID-19's Daily New Cases.

Methodology

Through the data that was collected on the Covid-19 pandemic from “Worldometer” which is the best free reference websites by the American Library Association (ALA) and trough the data that was collected on the climate from The World Air Quality Index project which is a non-profit project, we are going to apply panel data analysis over 19 months starting from March, 1 2020 till October, 1 2020 (i.e., the reasons behind conducing a panel data analysis will be mentioned later on chapter 4). Along with the panel data analysis we will conduct some descriptive statistics for each country by variable. The variables description will be divided into two major parts; **time series plots** as well as the **usual descriptive statistics** (mean, variance, covariances… etc.).

The analysis was conducted using many softwares like: MINITAB and R-Programming Language and STATA.

Data Description

With the spreading of Covid-19 all over the world, World Air Quality Index Project Team decided to make a project that provides new dedicated dataset for the major cities in the world. In the scope of our project we get our data from Air quality open data platform as mentioned before. The data was provided on a daily bases for 380 city all over the world and it is updated three times a day. So, the data was provided on a daily base but by city not by country.

So, in order to merge the climate data and Covid-19 dataset which was provided on a daily bases by country not by city, we used the R-Programming Language.

The climate data had many variables but we have chosen only three on them in order to represent the climate change and these variables are:

1. Temperature
2. Humidity
3. O3

Variables Choosing Criteria:

Choosing those variables especially to represent the climate changes was depending on: Although it is well recognized that temperature is a key factor in ozone episodes, it is not yet understood how rising global temperatures may affect the intensity and frequency of surface level ozone. More respiratory infections could result from this rise, which would be detrimental for children, the elderly, and those with asthma.

Climate change also involves water vapor. More water vapor is held in a warmer environment, which raises the possibility of more ozone forming. However, additional cloud cover, particularly early in the day, could slow down reactions and hence slow down the creation of ozone. Weak winds are also present when temperatures are high, which causes the atmosphere to become stagnant. Thus, the ozone levels might increase as the air simply cooks.

Ozone concentrations above a certain point can worsen chronic respiratory conditions and potentially raise mortality rates and this would certainly affect their immune response to the pandemic, the likelihood of more intense rainfall and hazardous heat waves is governed by humidity. Just like with temperature, we need to keep an eye on and comprehend fluctuations in surface humidity. The two pillars of climate change, humidity and temperature, operate in tandem. Ozone at ground level (O3) is a dangerous air pollutant that has negative impacts on both people and the environment. When sunlight is present, pollutants released from factories, power plants, industrial boilers, refineries, and other sources undergo chemical reactions.

Ozone can still reach high levels throughout the colder months, although it is most likely to do so on hot, bright days in urban areas. Even in remote regions, high ozone levels are possible due to the wind's ability to transport ozone across great distances.

Countries Choosing Criteria from different aspects:

Choosing the countries was based on:

Firstly, the idea of representing each continent so we could cover each part of the world with different climate conditions , where each continent were represented by two countries except for Australia which represent its continent

Secondly, the countries’ membership in the G20 which previously mentioned.

Since our data was provided for each city, we had duplicates (i.e. on 10-3-2020 in USA the values of humidity was measured for each city so we had around 50 different values for the same day) and also each variable was measured in terms of its minimum, maximum ,median ,variance and count.

So, we found that working on the median of the variable will be the most suitable to represent each variable, then we have taken the median of the medians of all of these values of each variable for all of the cities in each of the chosen countries in order to avoid the duplicates and have at the end only one value that represent that country in that day. And by the end of this process our data was ready to be used and to be merged with the Covid-19 data that was represented by only one variable which is Daily new cases since each of them is now provided for each country by day.

That was the first manipulation we have done in order to conduct our analysis. The reason behind the next manipulation in our data is that when it come for the analysis, we have chosen to conduct a panel data analysis for different reasons that will be mentioned later on chapter four, however one of the main reasons behind conducting a panel data analysis is that it is more efficient for studying different cross-sectional units (countries) over the time, it is even more efficient than the time series analysis; since in time series analysis we will be forced to conduct the same analysis eight times (i.e., since we have eight countries).

The second manipulation, in fact was to transform our data into a panel data, but for doing that we found that we have two types of panel data:

1. Short panel data
2. Long panel data

We found also that our data after transforming it into a panel data will be a long panel data. Long panel data refers to a dataset where the number of time units is greater than the cross-sectional units. Long panel data analysis is very complicated and was beyond our studying scope. And as a result we solved this problem by changing the daily panel data into a seasonal data by taking the median for each variable based on the **Meteorological Seasons** definition (i.e., will be explained by details in chapter 4). By that we will have 7 seasons that covers the studied period and 8 countries. Hence, the number of cross sectional units is greater than the number of time points and this is the definition of short panel data. By that short panel data analysis was easily applied.

Chapter Three

Descriptive Analysis

Overview

In this chapter we are going to conduct some descriptive statistics for the USA, UK, KSA and China, Brazil, South Africa, Italy and Australia variables. The variables description will be divided into two major parts; time series plots as well as the usual descriptive statistics (mean, variance, covariances… etc.).

United States' Descriptive Analysis

1- Temperature

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
16.725	55.297	-2.200	28.000	30.200	0	580

Table (1) : Descriptive statistics for United States Temperature

★ **Comments :**

- The maximum median temperature was approximately 28 while the minimum was -2.2.
- It is well known that the range is the difference between the largest and smallest data values in the sample. Hence; it represents the interval that contains all the data values. For temperature the range was about 30.2.
- It is well known that the variance measures how spread out the data are about their mean. For temperature the variability of our data was about 55.297 approximately.
- It seems like the number of non-missing values in the temperature data (N) is 580 and the number of missing values (N^*) is zero.

● The average temperature median value was about 16.725 approximately.

❖ **Time series plot**

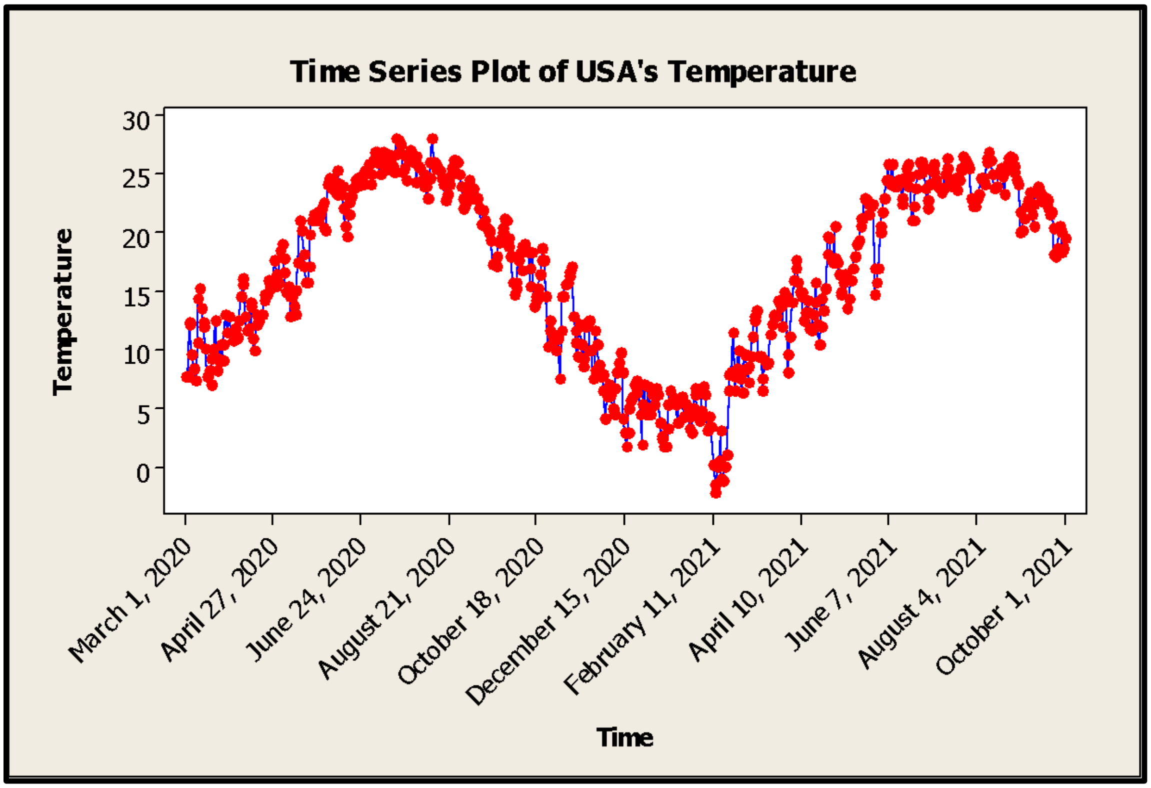

Figure (1): Time series plot for United States's Temperature

★ **Comment :**

It seems that there is no trend in the temperature during the 21 months we are studying. It is well known that a seasonal pattern is a rise and fall in the data values that repeats regularly over the same time period. Hence, we can find out that the temperature has a seasonal pattern with obvious fluctuations over time without any kind of outliers (outliers inspection was made using R-programming language).

2- Humidity

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
69.021	**60.984**	**43.450**	**85.350**	**41.900**	**0**	**580**

Table (2) : Descriptive statistics for United States' Humidity

★ **Comments :**

- The maximum median value of humidity was 85.350 approximately while the minimum was 43.45.
- The range was about 41.9 approximately.
- Humidity dispersion was about 60.984 approximately.
- It seems like the number of non-missing values in the humidity data (N) is 580 and the number of missing values (N^*) is zero.
- The average of humidity median values was about 69.021 approximately.

❖ Time series plot

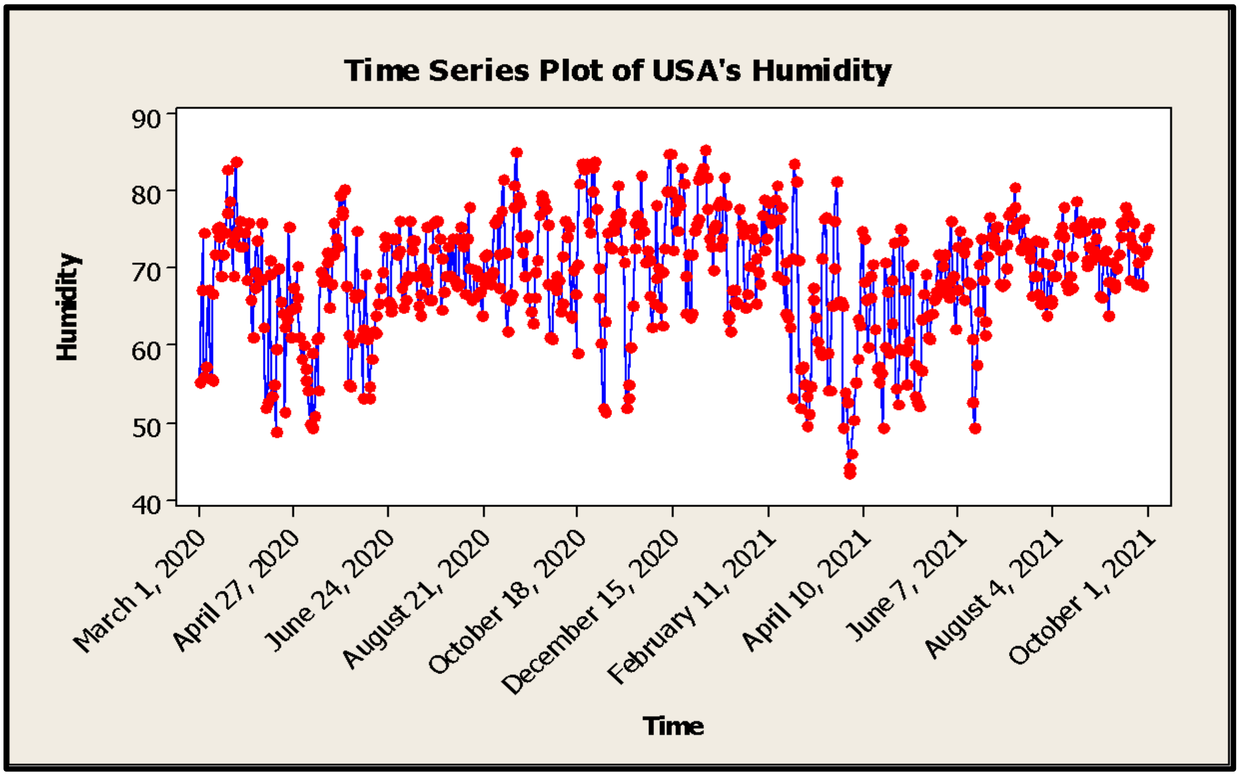

Figure (2): Time series plot for United States' Humidity

★ **Comment :**

It seems like there is a huge amount of fluctuations without any kind of trend or saddle points with a noticeable number of outliers but the most irregular one was 50.50 on April 4, 2021. And the lowest median value of humidity was 43.45 on April 2, 2021.

3- O3

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
21.299	22.307	8.600	33.600	25.000	0	580

Table (3) : Descriptive statistics for United States' O3

★ **Comments:**

- The maximum median value of O3 was 33.6 approximately while the minimum was 8.6
- The range was about 25 approximately.
- O3 dispersion was about 22.307 approximately.
- It seems like the number of non-missing values in the O3 data (N) is 580 and the number of missing values (N^*) is zero.
- The average of O3 median values was about 21.299 approximately.

❖ **Time Series Plot**

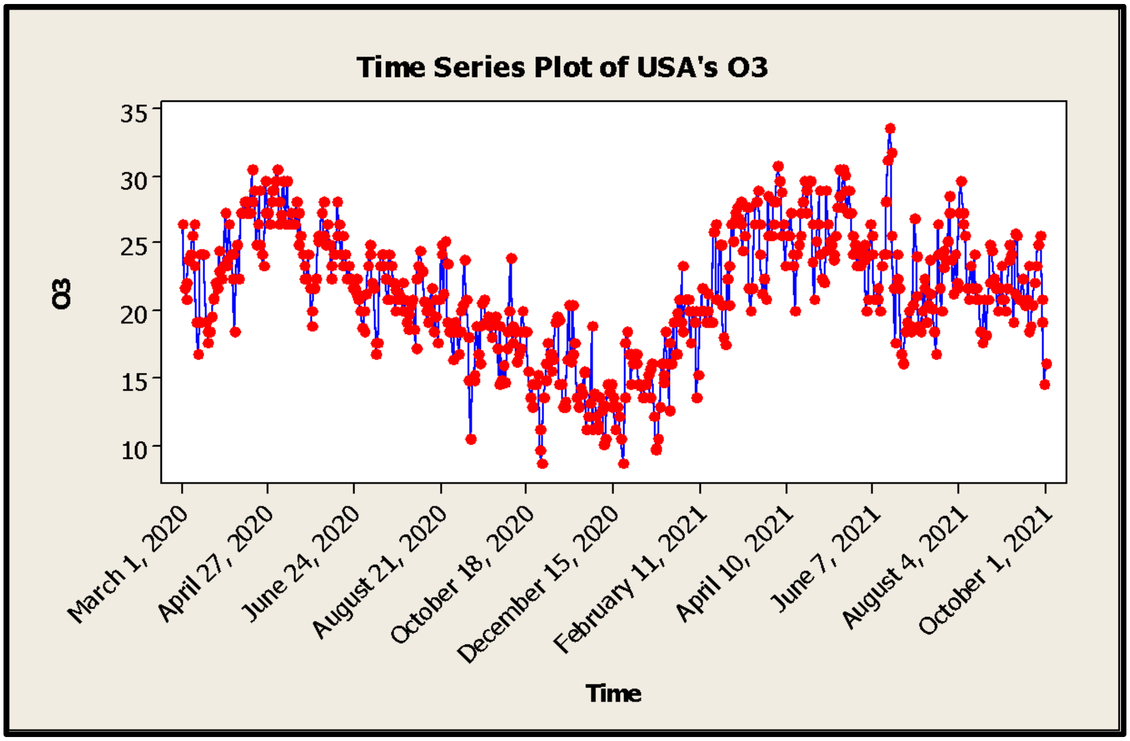

Figure (3): Time series plot for United States' O3

★ **Comment:**

A huge amount of fluctuations exist in O3 data without any kind of trend with a noticeable pattern of seasonality in the values. It seems like there are two outliers in O3 data and the most irregular one was 8.7 on October 28, 2020.

4- Daily new cases

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
76720	4089860492	7	306271	306264	0	580

Table (4) : Descriptive statistics for United States' Daily new cases

★ **Comments:**

- The maximum number of the daily new infected cases with Covid-19 was 306271 while the minimum was 7.
- The range was about 306264 cases.
- The dispersion was very high; about 4089860492.
- It seems like the number of non-missing values in the daily new cases data (N) is 580 and the number of missing values (N^*) is zero.
- The average number of daily new cases was about 76720 cases.

❖ **Time Series Plot**

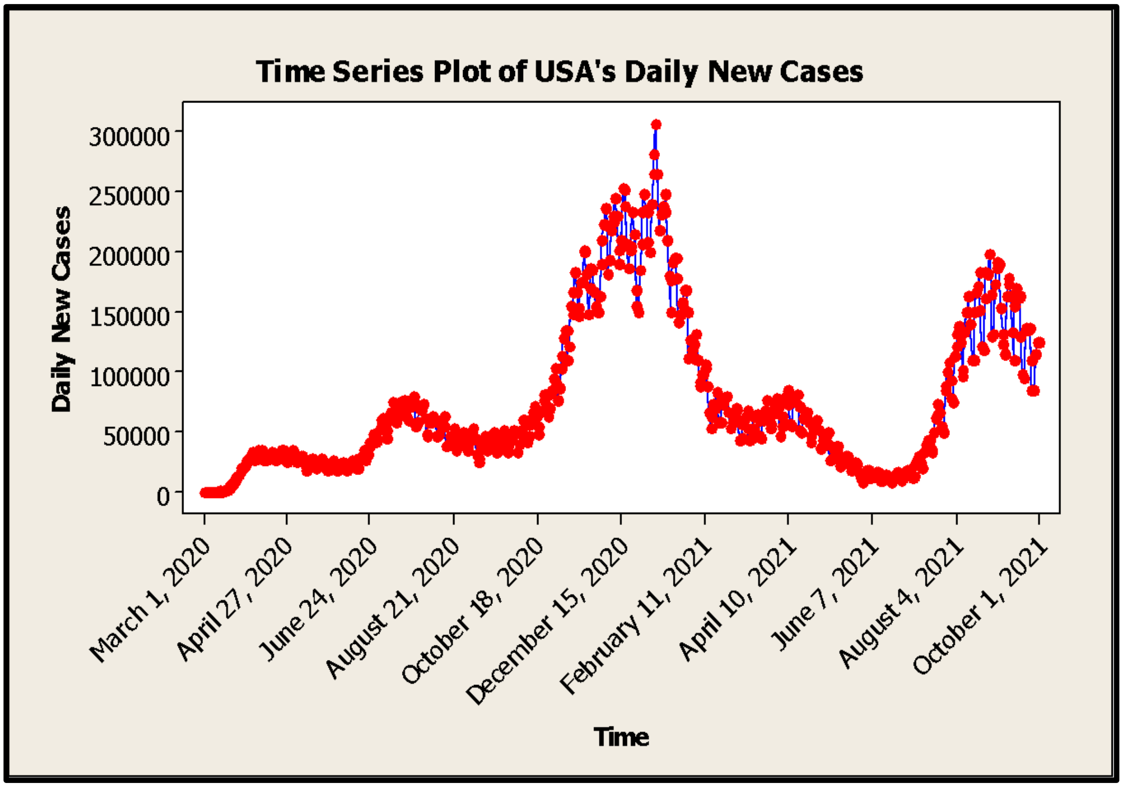

Figure (4): Time series plot for United States' Daily new cases

★ **Comment:**

This plot shows a strong level of fluctuations in the daily new infected cases with Covid-19, the number of cases keeps on increasing then decreasing and sometimes it shows a kind of stability status. Hence, there is no pattern of trend or seasonality but it is obvious that there are some outliers and we can say that most of these outliers appears on the time interval from December 2020 to the mid of January 2021 and the most anomaly value was 306271 cases on January 8, 2021.

❖ Covariances

Variables	Temperature	Humidity	O3	Daily new cases
Temperature	55			
Humidity	2	61		
O3	10	-22	22	
Daily new cases	-191035	131693	-169420	4089860492

Table (5) : Covariances between the variables

★ Comments:

It is well known that we can use the covariance to determine the direction of a linear relationship between two variables as follows:

- As Temperature increases Humidity tends to increase as well.
- As Temperature increases O3 tends to increase as well.
- As Temperature increases Daily New Cases tend to decrease.
- As Temperature increases Active Cases tend to decrease.
- As Humidity increases O3 tends to decrease.
- As Humidity increases the Daily New Cases tend to increase as well.
- As Humidity increases the Active Cases tend to increase as well.
- As O3 increases, Daily New Cases tend to decrease.
- As O3 increases Active Cases tend to decrease.
- As Daily New Cases increase Active Cases tend to increase as well.

Note: It is well known that covariance is similar to correlation but when the covariance is calculated, the data are not standardized. Therefore, the covariance is expressed in units that vary with the data and is not converted to a standardized scale of −1 to +1 as a result we cannot use the covariance to assess the strength of a linear relationship. To assess the strength of the relationship between two variables we will use the correlation we have obtained using Minitab.

❖ Correlations

Variables	Temperature	Humidity	O3	Daily new cases
Temperature				
Humidity	0.028			
O3	0.281	-0.588		
Daily new cases	-0.402	0.264	-0.561	

Table (6) : Correlation between the variables

➢ **Note** : cell contents are Pearson correlation

★ Comments:

It is well known that the correlation between any variable and itself is a strong perfect relation = 1.

- There is a very weak positive linear relation between Temperature and Humidity.
- There is an approximate weak positive linear relation between Temperature and O3.
- There is a moderate negative linear relation between Temperature and Daily New Cases.
- There is an approximate weak negative linear relation between Temperature and Active Cases.
- There is a moderate negative linear relation between Humidity and O3.
- There is an approximate weak positive linear relation between Humidity and Daily New Cases.
- There is a very weak positive linear relation between Humidity and Active Cases.
- There is a moderate negative linear relation between O3 and Daily New Cases.
- There is a very weak negative linear relation between O3 and Active Cases.
- There is a moderate positive linear relation between Daily New Cases and Active Cases.

China’s Descriptive Analysis

1- Temperature

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
18.427	**64.363**	**-4.150**	**30.000**	**34.150**	**0**	**580**

Table (7) : Descriptive statistics for China’s Temperature

★ **Comments :**

- The average degrees for the temperature were around 18.427 during the period 2020-2021.
- For temperature, the variability of our data is about 64.363 approximately.
- The maximum temperature was approximately 30 while the minimum was -4.150.
- For temperature the range was about 34.150.
- It seems like the number of non-missing values in the temperature data (N) is 580 and the number of missing values ($N*$) is zero.

❖ Time series plot

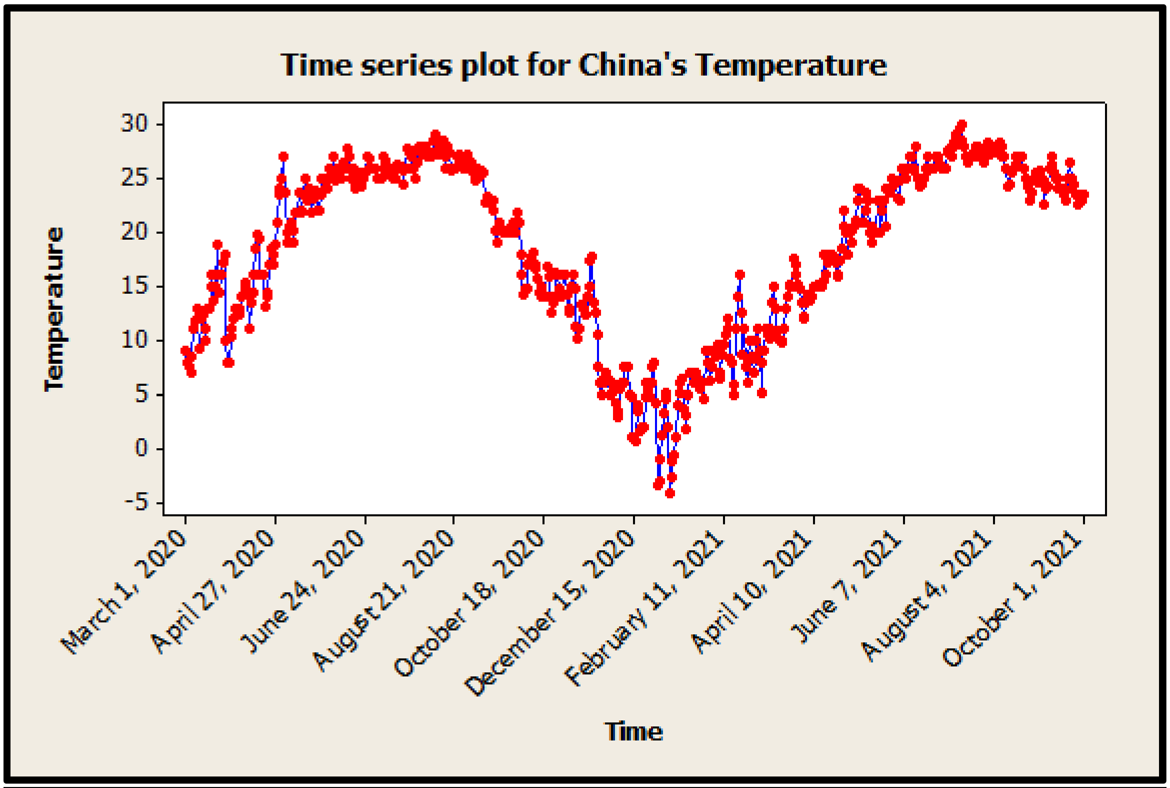

Figure (5): Time series plot for China's Temperature

★ **Comment :**

It seems that there is neither trend nor outliers over the period 2020-2021.It's clear that the temperature has obvious fluctuations regularly over time representing a pattern of seasonality.

2- Humidity

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
69.130	101.586	38.000	86.700	48.700	0	580

Table (8) : Descriptive statistics for China's Humidity

★ **Comments :**

- The average value of the humidity was around 69.130 during the period 2020-2021.
- For humidity the variability of our data is approximately 101.586.
- The maximum humidity was approximately 86.70 while the minimum was 38..
- For humidity the range was about 48.70.
- It seems like the number of non-missing values in the humidity data (N) is 580 and the number of missing values ($N*$) is zero.

❖ Time series plot

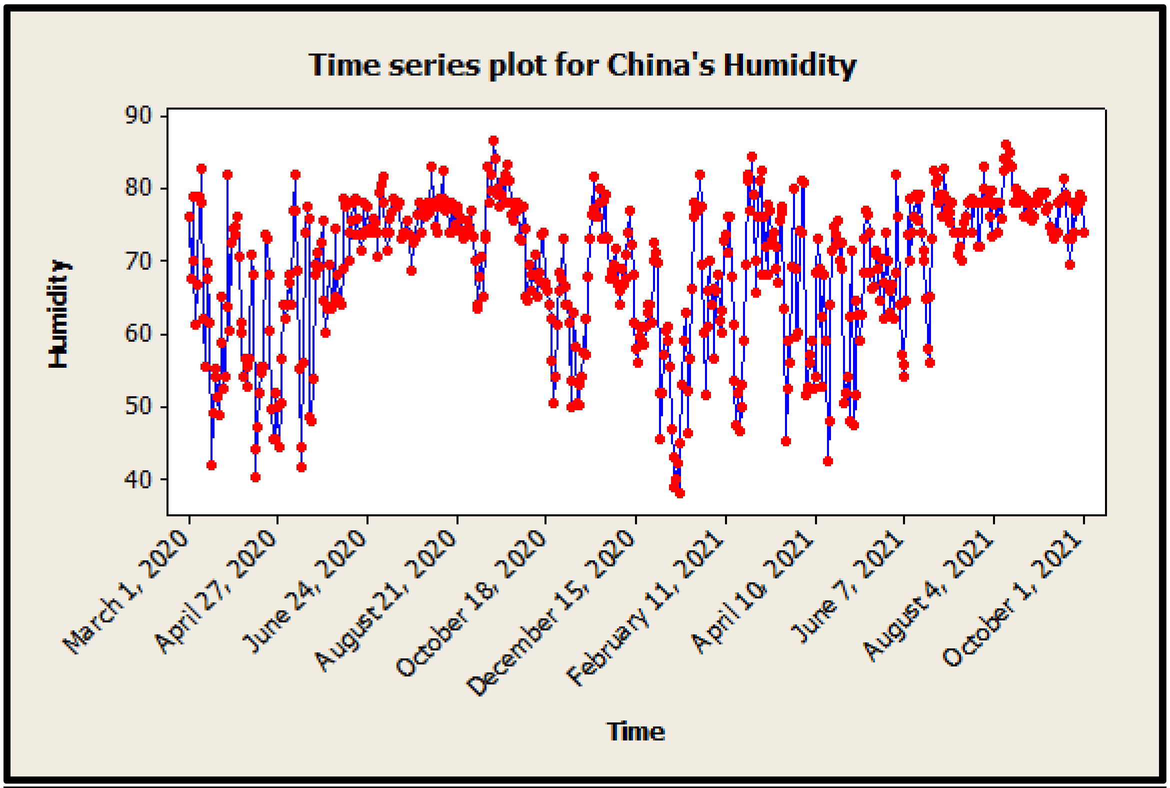

Figure (6): Time series plot for China's Humidity

★ **Comment :**

There is a large amount of fluctuations without any kind of trend or saddle points and not even a seasonality effect in the humidity values over time. There are some outliers and we can say that those outliers appear in the time interval from March 14, 2020 to April 18, 2021 and the most anomalous values were 38 on January 12, 2021 and 41.55 on April 18, 2021.

<u>3- O3</u>

❖ **<u>Descriptive statistics</u>**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
25.277	47.656	5.300	42.100	36.800	0	580

Table (9) : Descriptive statistics for China's O3

★ **<u>Comments</u> :**

- The average value of the O3 was around 25.277 during the period 2020-2021.
- For O3 the variability of our data is approximately 47.656.
- The maximum humidity was approximately 42.10 while the minimum was 5.30.
- For O3 the range was about 36.80.
- It seems like the number of non-missing values in the O3 data (N) is 580 and the number of missing values ($N*$) is zero.

- ❖ **<u>Time series plot</u>**

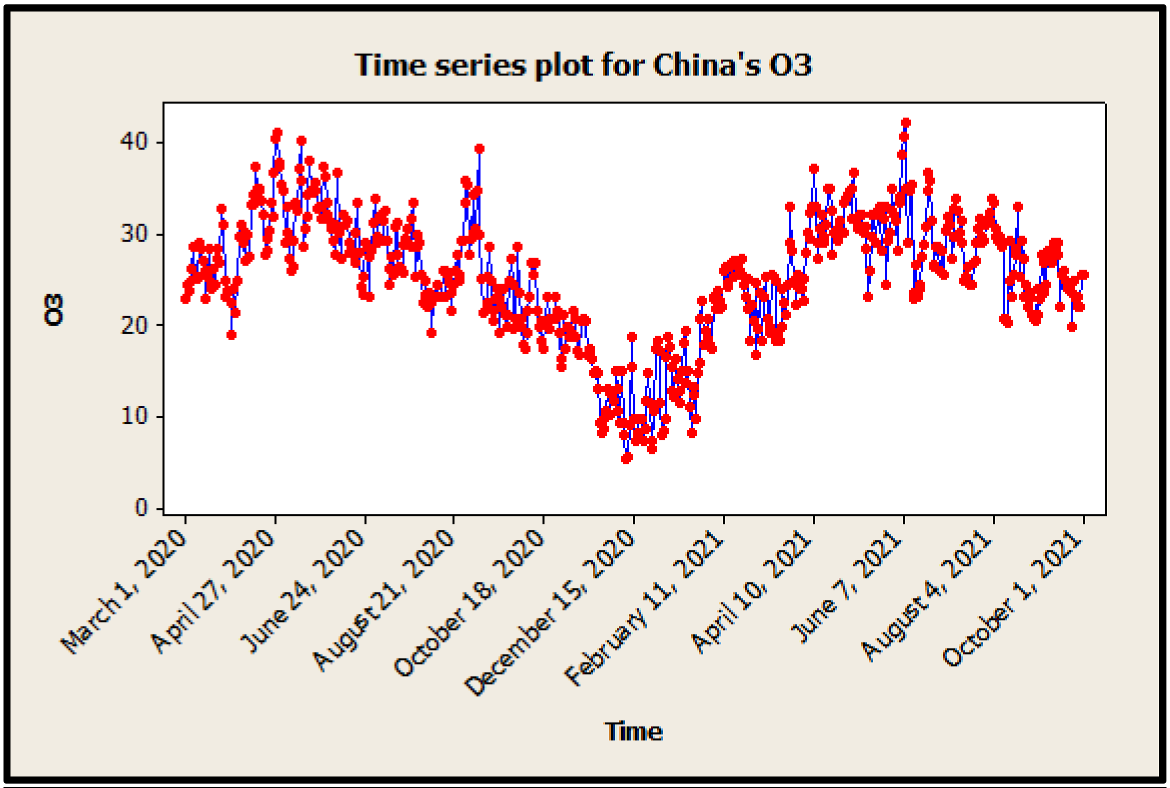

Figure (7): Time series plot for China's O3

★ **<u>Comment</u> :**

There is a large amount of fluctuations in O3 data without any kind of trend but with an obvious pattern of seasonality. There are some outliers and we can say that those outliers appear in the time interval from November 24, 2020 to January 21, 2021 and the most anomalous values were 5.3 on December 10, 2020 and 8.4 on January 3, 2021.

4-Daily new cases

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
28.17	978.64	0.00	325.00	325.00	0	580

Table (10) : Descriptive statistics for China's Daily new cases

★ **Comments :**

- The average number of Daily new cases was around 28.17 during the period 2020-2021.
- For Daily new cases the variability of our data is approximately 978.64.
- The maximum number of Daily new cases was approximately 325 while the minimum was 0.
- For Daily new cases the range was about 325.
- It seems like the number of non-missing values in Daily new cases data (N) is 580 and the number of missing values ($N*$) is zero.

❖ **Time series plot**

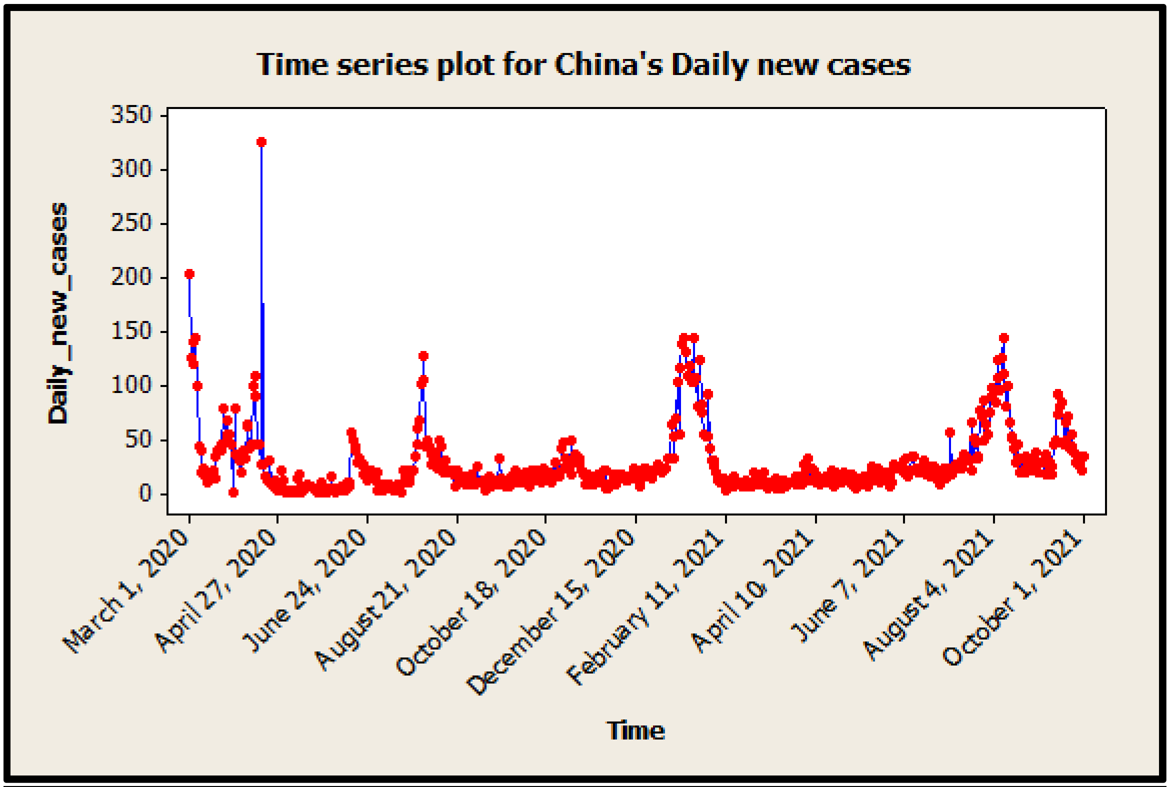

Figure (8): Time series plot for China's Daily new cases

★ **Comment :**

There is a moderate level of fluctuations in the daily new infected cases with Covid-19, the number of new cases are so close almost constant showing a kind of stability status but at some points of time there was a sudden high increase then decreasing again to nearly its previous status. Hence, there is no pattern of trend or seasonality .There are some outliers and we can say that those outliers appear in the time interval from March 1, 2020 to September 21, 2021 and the most anomalous values were 202 cases on March 1, 2020 and 85 cases on July 20, 2021.

❖ **Covariances**

Variables	Temperature	Humidity	O3	Daily new cases
Temperature	**64**			
Humidity	**36**	**102**		
O3	**36**	**-3**	**48**	
Daily new cases	**-19**	**-1**	**-22**	**979**

Table (11) : Covariances between the variables

★ **Comments :**

To determine the direction of a linear relationship between two variables we have used the covariances from where we conclude that :

- As Temperature increases, Humidity tends to increase as well.
- As Temperature increases,O3 tends to increase as well.
- As Temperature increases, Daily New Cases tend to decrease.
- As Humidity increases, O3 tends to decrease.
- As Humidity increases, the Daily New Cases tend to decrease.
- As O3 increases, Daily New Cases tend to decrease.

❖ **Correlations**

Variables	Temperature	Humidity	O3	Daily new cases
Temperature				
Humidity	**0.446**			
O3	**0.650**	**-0.050**		
Daily new cases	**-0.074**	**-0.004**	**-0.102**	

Table (12) : Correlation between the variables

★ **Comments :**

To assess the strength of the relationship between two variables we have used the correlation between two variables from where we concluded that:

(Note that: the correlation between any variable and itself is a strong perfect relation = 1).

- There is a moderate positive linear relation between Temperature and Humidity.
- There is a moderate positive linear relation between Temperature and O3.
- There is a very weak negative linear relation between Temperature and Daily New Cases.
- There is a very weak negative linear relation between Humidity and O3.
- There is a very weak negative linear relation between Humidity and Daily New Cases.
- There is a very weak negative linear relation between O3 and Daily New Cases.

United Kingdom's Descriptive Analysis

1- Temperature

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
11.483	25.845	-1.1	23.9	25	0	580

Table (13) : Descriptive statistics for United Kingdom's Temperature

★ **Comments:**

- Average temperature of the UK during this period was 11.483.
- Lowest temperature in the UK reaches -1.1 and the highest temperature reaches 25.
- The variability of temperature data was 25.845 approximately.
- Number of non-missing values was 580 and the number of missing equals zero.
- The range of our temperature variable reaches 25.

- **<u>Time series plot:</u>**

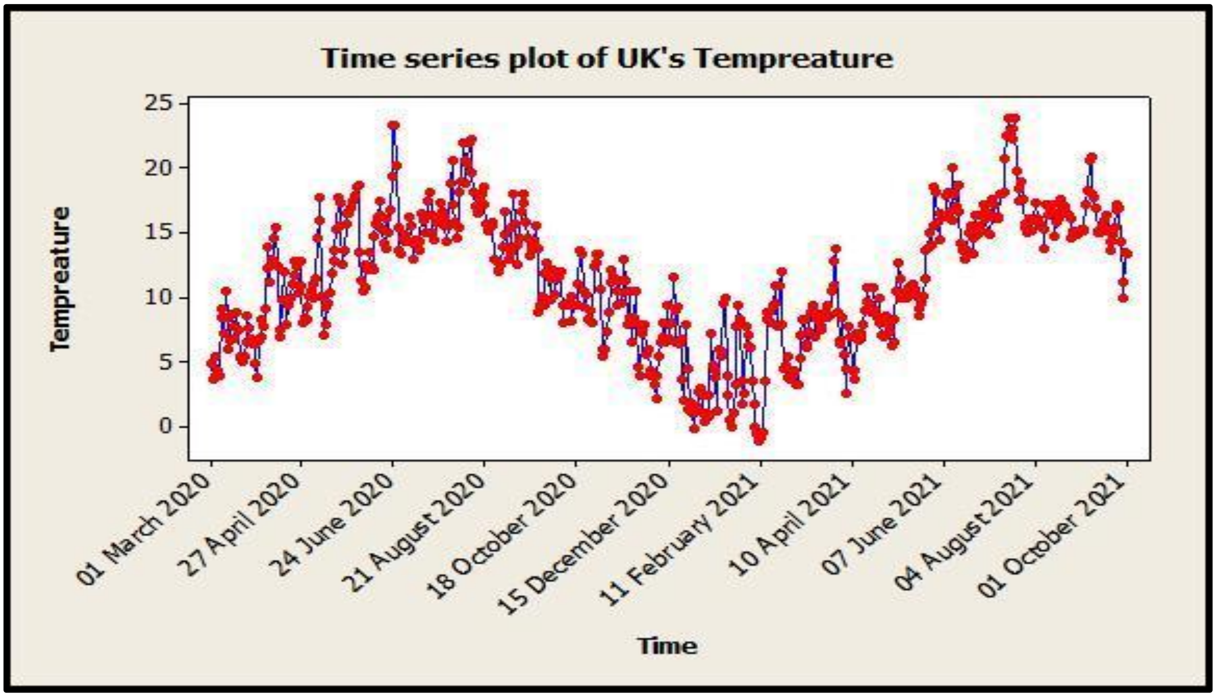

Figure (9): time series plot for United Kingdom's Temperature

★ **<u>Comment:</u>**

This plot shows from March till August and from Feb 11, 2021 till October 1, 2021 an increasing level of temperature while from Aug 2020 till Feb 2021 shows a decreasing level of temperature so, we conclude that there's no trend as data of temperature is fluctuating. There are no outliers.

2- Humidity

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
78.75	84.041	48.9	95.4	46	0	580

Table (14) : Descriptive statistics for United Kingdom's Humidity

★ **Comments:**

- Average value for humidity in the UK during this period was 78.75.
- Lowest value of humidity in the UK reached 48.9 and the highest value was 95.4.
- The variability of humidity data was 84.041 approximately.
- Number of non-missing values was 580 and the number of missing equals zero.
- The range of humidity was 46.

❖ Time series plot:

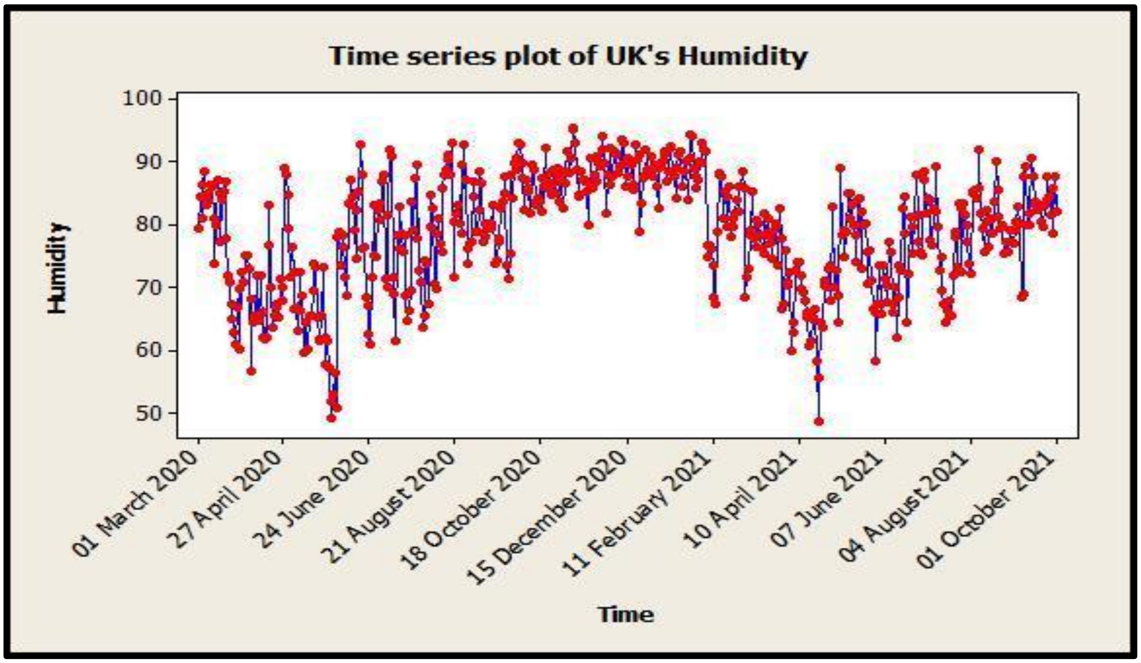

Figure (10): Time series plot for United Kingdom's Humidity

★ Comment:

In this plot we can see that there's no trend as the data is fluctuating, it decreased then increased then decreased again. There are three outliers whose values equal 49.25, 50.80, and 48.90 on May 29, 2020, June 2, 2020, and April 23, 2021 respectively.

3- O3

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
21.189	38.086	2.5	46	43.5	0	580

Table (15) : Descriptive statistics for United Kingdom's O3

★ **Comments:**

- Average value of O3 in the UK during this period was approximately 21.
- Lowest value of O3 during this period reached 2.5 and the highest value was 46.
- The variability of O3 data was approximately 38 which indicate a high dispersion from mean.
- Number of non-missing values was 580 and the number of missing equals zero.
- The range of humidity was 43.5.

❖ Time series plot:

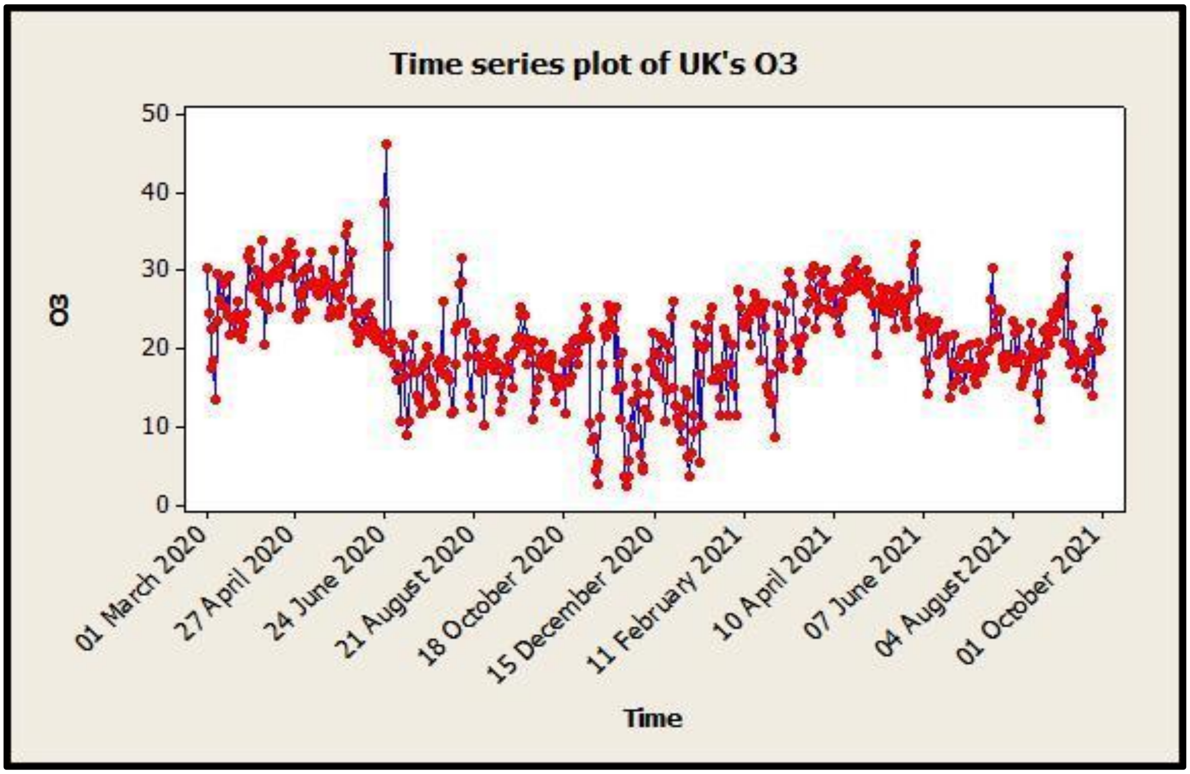

Figure (11): Time series plot for United Kingdom's O3

★ Comment:

This plot shows slight fluctuations in a common range as most of the values lie between 10 and 30 but there are a lot of outliers as well where maximum outlier equals 46.00 on July 25, 2020 while minimum outlier equals 2.5 on Nov 27, 2020.

4-Daily new cases

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
13521	213249555	3	67794	67791	0	580

Table (16) : Descriptive statistics for United Kingdom's Daily new cases

★ **Comments:**

- Average number of daily new cases in the UK during the studied period was 13521 cases.
- Lowest number of daily new cases during this period was 3 and the highest was 67794.
- The dispersion of daily new cases was approximately 213249555.
- Number of non-missing values was 580 and the number of missing equals zero.
- The range of O3 was 67791.

❖ **Time series plot:**

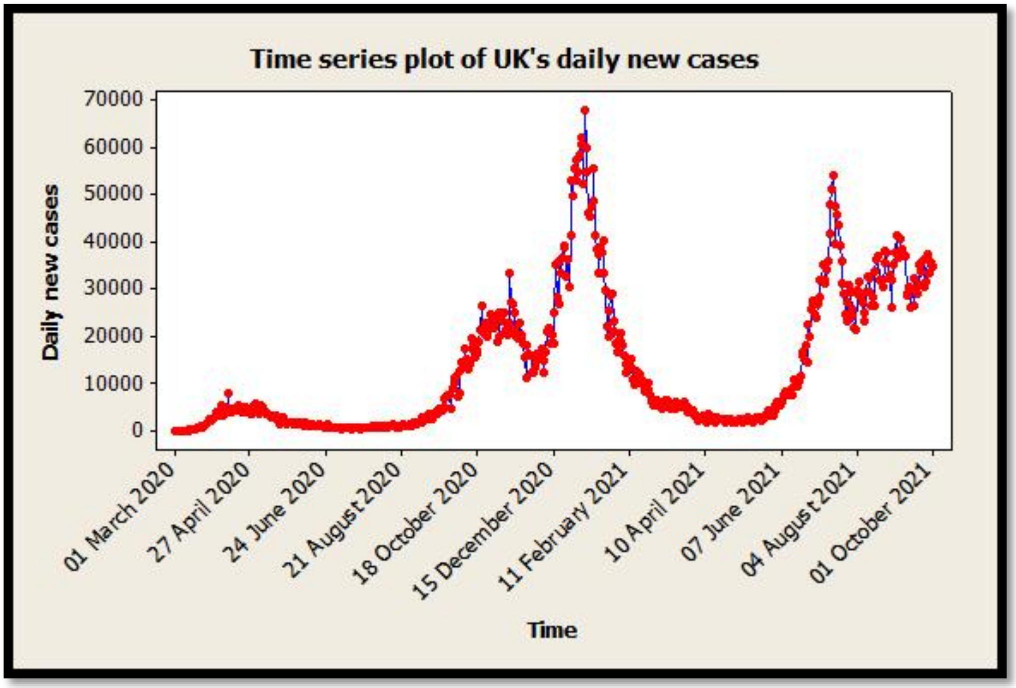

Figure (12): Time series plot for United Kingdom's Daily new cases

★ **Comment:**

The plot shows that daily new cases of Covid-19 was approximately stable from March to October 2020 (for 8 months), then an increasing trend appeared then decreasing then increasing again. Fluctuations do exist and the peak (highest number of new cases) lies between 15 Dec 2020 and 11 Feb 2021(at the mid of this period there's an outlier its value was 67794).

❖ **Covariances:**

Variables	Temperature	Humidity	O3	Daily new cases
Temperature	**26**			
Humidity	**-12**	**84**		
O3	**4**	**-36**	**38**	
Daily new cases	**-5780**	**-5780**	**-34740**	**213249555**

Table (17) : Covariances between the variables

★ **Comments:**

- When the number of daily new cases increases, the number of active cases increases.
- When temperature increases number of active cases decrease
- When temperature increases, the number of daily new cases decreases.
- When temperature increases, the level of humidity decreases.
- When humidity increases, the number of daily new cases increases.
- When humidity increases number of active cases increases
- When percent of O3 gas increases, the number of daily new cases decreases.
- When percent of O3 gas increases, the humidity level decreases.
- When percent of O3 gas increases, the number of active cases decreases.
- When percent of O3 gas increases, temperature increases.

❖ **Correlations**

Variables	Temperature	Humidity	O3	Daily new cases
Temperature				
Humidity	**-0.255**			
O3	**0.128**	**-0.644**		
Daily new cases	**-0.078**	**0.390**	**-0.385**	

Table (18) : Correlation between the variables

★ **Comment:**

- There is a moderate positive linear relationship between daily new cases and active cases equals (0.69)
- There is a very weak negative linear relationship between temperature and daily new cases equals(-0.078)
- There is an approximate moderate positive linear relationship between humidity and daily new cases equals(0.39)
- There is an approximate moderate negative linear relationship between O3 and daily new cases equals(-0.385)
- There is a moderate negative linear relationship between Active cases and Temperature equals (-0.425)
- There is a moderate positive linear relationship between Active cases and humidity equals (0.446)
- There is an approximate moderate negative linear relationship between Active cases and O3 equals (-0.335)
- There is a weak negative linear relationship between temperature and humidity equals (-0.255)
- There is a very weak positive linear relationship between temperature and O3 equals (0.128).
- There is a moderate negative linear relationship between O3 gas and humidity equals (-0.644).

Saudi Arabia's Descriptive Analysis

1- Temperature

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
22.231	55.347	5.000	39.000	34.000	0	580

Table (19) : Descriptive statistics for Saudi Arabia's Temperature

★ **Comments:**

- The average temperature in the KSA throughout the period from March 2020 till October 2021 was 22.231.
- The variability of the data (its deviation from the mean) was found to be about 55.347.
- The highest temperature recorded was 39, while the lowest one was 5.
- The temperature range equaled around 34.
- The number of non-missings was 580, and this was reflected on the missings to be zero.

❖ Time Series Plot:

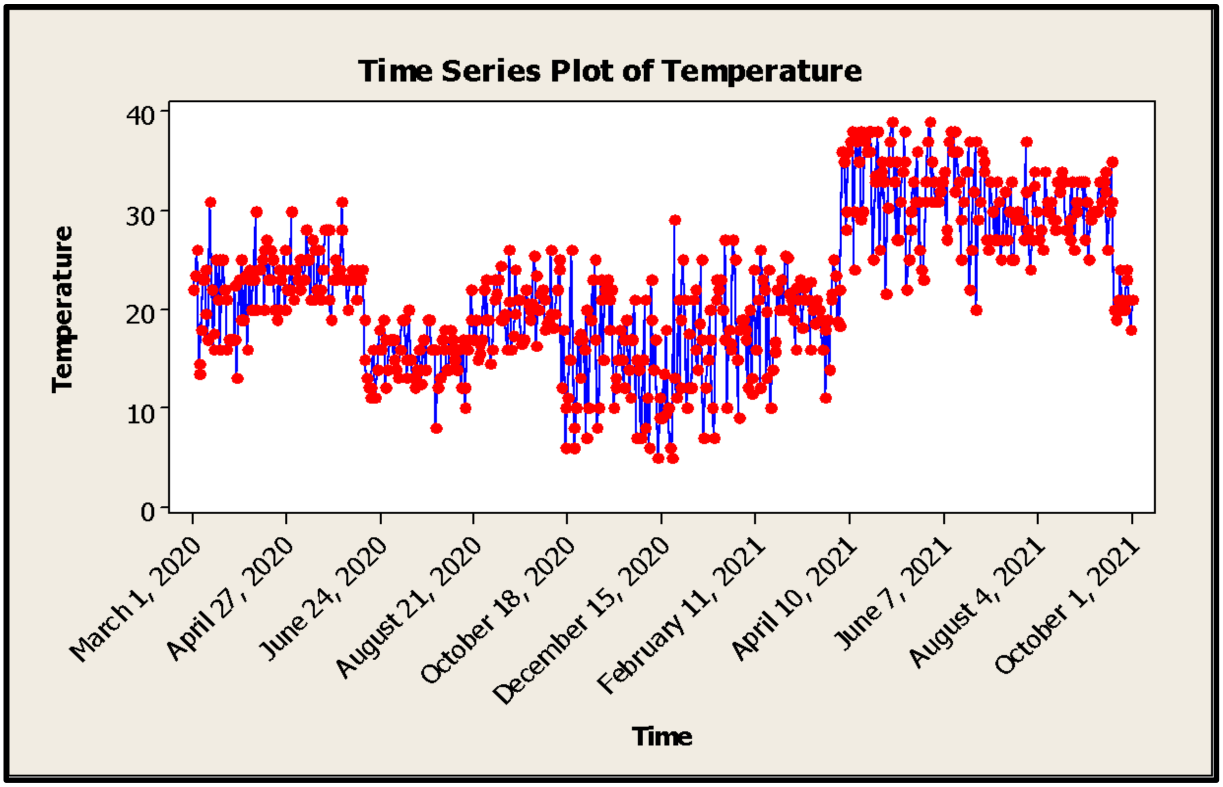

Figure (13): Time series plot for Saudi Arabia's Temperature

★ Comment:

It's observed that there are no sudden shifts, trends, outliers, seasonal effects or cyclic movements. The temperature data shows strong random variations or fluctuations. The period from the beginning of March 2020 till that of October 2021 experienced relatively average temperature measures till October 202, and then it decreased till April 2021, ending with high measures till August 2021, and a gradual decrease mode.

2- Humidity

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
43.564	328.921	7.000	93.000	86.000	0	580

Table (20) : Descriptive statistics for Saudi Arabia's Humidity

★ **Comments:**

- The average humidity in the KSA throughout the period from March 2020 till October 2021 was 43.564.
- The variability of the humidity data was found to be equal 328.921approximately.
- The highest humidity record that hit the KSA was 93, while the lowest one was 7.
- The humidity variable's range was 86.
- The Non-Missings was 580, and this was reflected on the Missings to be zero.

❖ **Time Series Plot:**

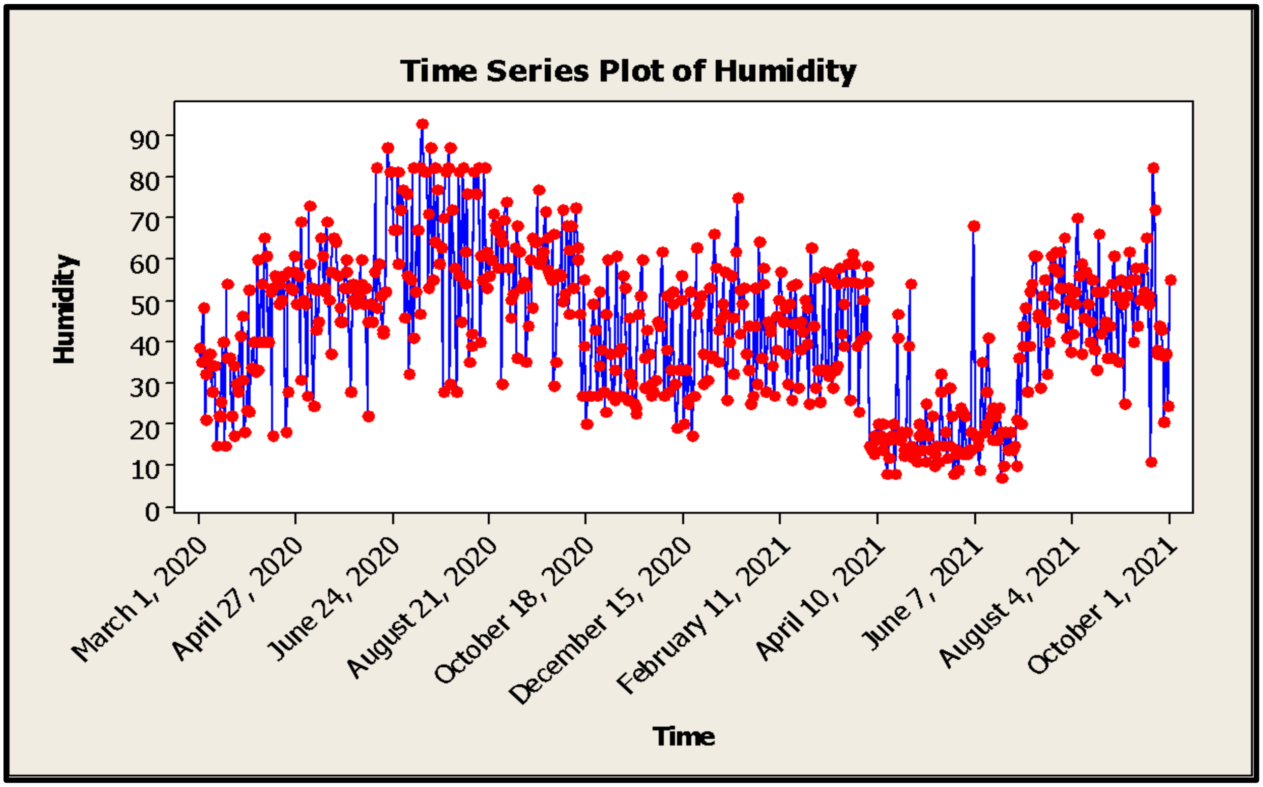

Figure (14): Time series plot for Saudi Arabia's Humidity

★ **Comment:**

It's observed that there's neither trend nor seasonal effect or cyclic movements over the humidity's pattern in the whole 21 month, however, there are noticeable extreme variations/fluctuations. There are no noticed extreme outliers for this variable in KSA. The humidity was at its highest values in the same months' temperature had its ultimate measures.

3- O3

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
22.427	**140.323**	**1.000**	**67.000**	**66.000**	**0**	**580**

Table (21) : Descriptive statistics for Saudi Arabia's O3

★ **Comments:**

★ The average O3 value in the KSA throughout the period from March 2020 till October 2021 was 22.427.

★ The variance of the O3 was 140.323 approximately.

★ The highest O3 level was 67, while the lowest one was 1.

★ The O3 variable's range equaled around 66.

★ The Non-Missings was 580, and this was reflected on the Missings to be zero.

❖ **Time Series Plot:**

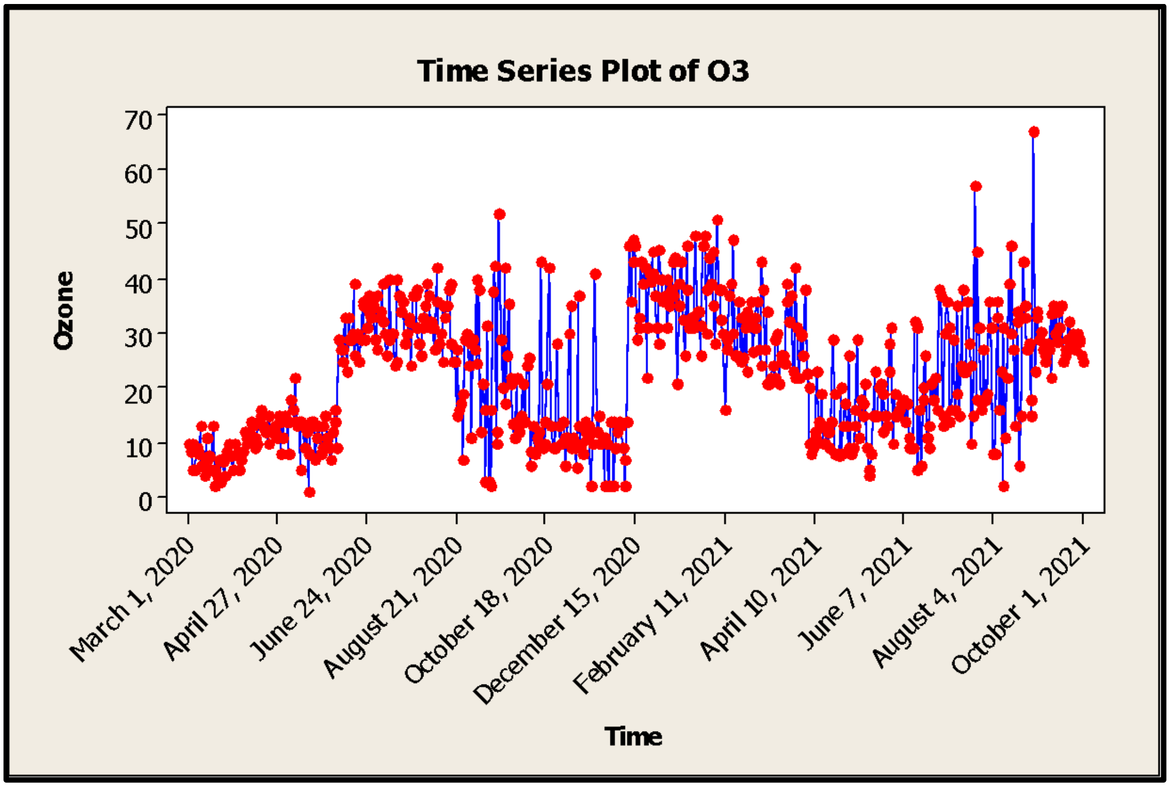

Figure (15): Time series plot for Saudi Arabia's O3

★ **Comment:**

It's noticed the existence of an extreme outlier in the O3 data on December 30, 2021. The data of O3 shows very strong random variations over the months of the mentioned period, as it reached its peak in August 2021. Also, it is observed that it had the lowest values on May 17, 2020, Nov 25, 2020, Sept 13 2020, Dec 9 2020, Dec 1 2020, Mar 18 2020, Nov 28 2020, Nov 16 2020, Aug 10 2021. There isn't any indication for a trend, seasonal effect, or cyclic movements.

4-Daily new cases

- **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
943.4	914235.8	0.0	4919.0	4919.0	0	580

Table (22) : Descriptive statistics for Saudi Arabia's Daily new cases

★ **Comments:**

- The average number of the Daily New Cases in the KSA from March 2020 till October 2021 was 943.4.
- The variability of the Daily New Cases was found to be about 914235.8.
- The highest number of the Daily New Cases recorded was 4919, while the lowest one was 0.
- The range of the Daily New Cases was equal to 4919.
- The Non-Missings was 580, and this was reflected on the Missings to be zero.

❖ **Time Series Plot:**

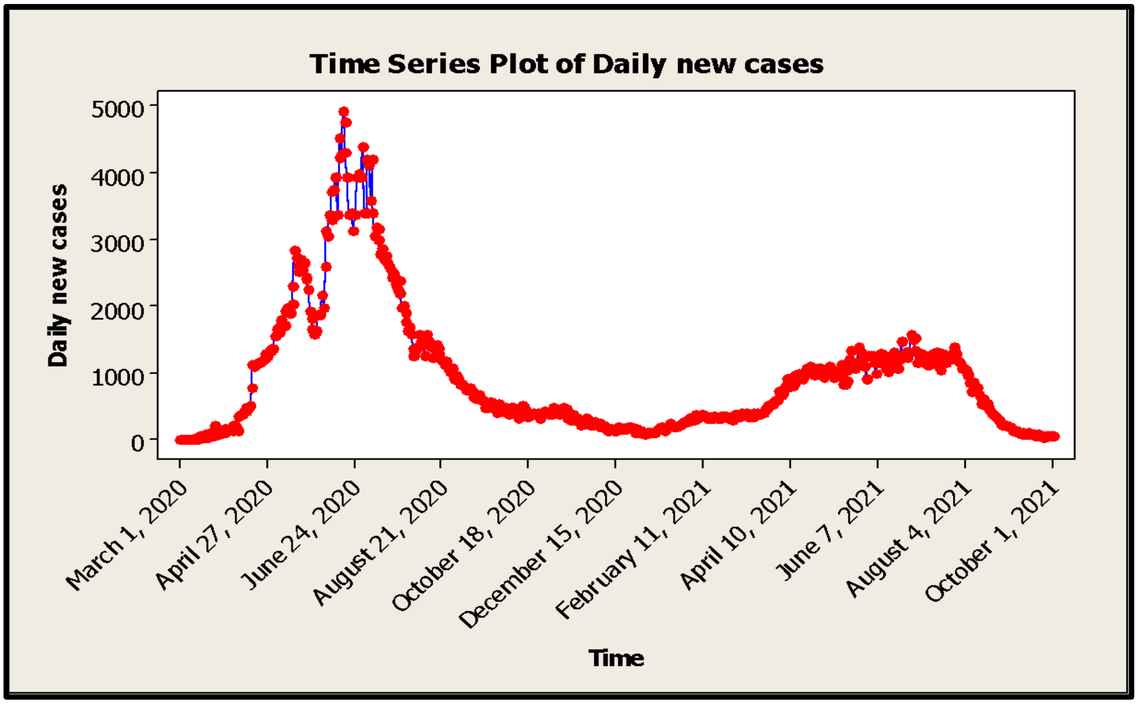

Figure (16): Time series plot for Saudi Arabia's Daily new cases

★ **Comment:**

There are observed saddle points, fluctuations and variations for the number of the daily new cases over the time mentioned, the plot shows an increasing mode till May 2020 then it has experienced a saddle point where it began to decrease, then increased again in the second quarter of 2020, until it reached its peak on June 17, 2020 to be equal 4919 cases, and it has experienced some slight ups and downs in 2021. There were neither trends nor cyclic movements or seasonal patterns.

❖ **Covariances**

Variables	Temperature	Humidity	O3	Daily new cases
Temperature	**55**			
Humidity	**-44**	**329**		
O3	**-17**	**53**	**140**	
Daily new cases	**-130**	**4691**	**880**	**914236**

Table (23) : Covariances between the variables

★ **Comments:**

- As the temperature increases in the KSA throughout the previously mentioned period, the humidity tends to decrease.
- As the temperature increases, the O3 in the KSA tends to decrease.
- As the temperature increases, the number of Daily New Cases in the KSA tends to decrease.
- As the temperature increases, the number of the Active Cases in the KSA tends to decrease.
- As the humidity increases, the O3 tends to decrease.
- As the humidity increases, the number of the Daily New Cases tends to decrease.
- As the humidity increases, the number of the Active Cases tends to decrease.
- As the O3 increases, the number of the Daily New Cases tends to decrease.
- As the O3 increases, the number of the Active Cases tends to decrease.
- As the number of the Daily New Cases in the KSA increases, the number of the Active Cases tends to increase.

❖ **Correlations**

Variables	Temperature	Humidity	O3	Daily new cases
Temperature				
Humidity	**-0.328**			
O3	**-0.188**	**0.245**		
Daily new cases	**-0.018**	**0.271**	**0.078**	

Table (24) : Correlation between the variables

★ **Comments:**

- There is an approximate moderate negative linear relationship between the temperature and the humidity.
- There is a very weak negative linear relationship between the temperature and the O3 in the KSA.
- There is a very weak negative linear relationship between the temperature and the number of the Daily New Cases of COVID-19.
- There is an approximate weak negative linear relationship between the temperature and the number of the Active Cases of COVID-19.
- There is an approximate weak positive linear relationship between the humidity and the O3 in the KSA.
- There is an approximate weak positive linear relationship between the humidity and the number of COVID-19 Daily New Cases.
- There is a moderate positive linear relationship between the humidity and the number of COVID-19 Active Cases.
- There is a very weak positive linear relationship between the O3 level and the number of the Daily New Cases.
- There is a very weak positive linear relationship between the O3 level and the number of the Active Cases.
- There is an approximate strong positive linear relationship between the number of the Daily New Cases of COVID-19 and that of the Active Cases.

Brazil's Descriptive Analysis

1- Temperature:

❖ **Descriptive Statistics:**

Mean	**Variance**	**Minimum**	**Maximum**	**Range**	**Missings**	**Non missings**
19.271	**10.461**	**8.000**	**27.000**	**19.000**	**0**	**580**

Table (25) : Descriptive statistics for Brazil's Temperature

★ **Comments:**

- The average temperature in Brazil throughout the period from March 2020 till October 2021 was 19.271.
- The variability of the data (its deviation from the mean) was found to be about 10.46.
- The highest temperature recorded was 27, while the lowest one was 8.
- The temperature range equaled around 19.
- The Non-Missings was 580, and this was reflected on the Missings to be zero.

❖ Time Series Plot:

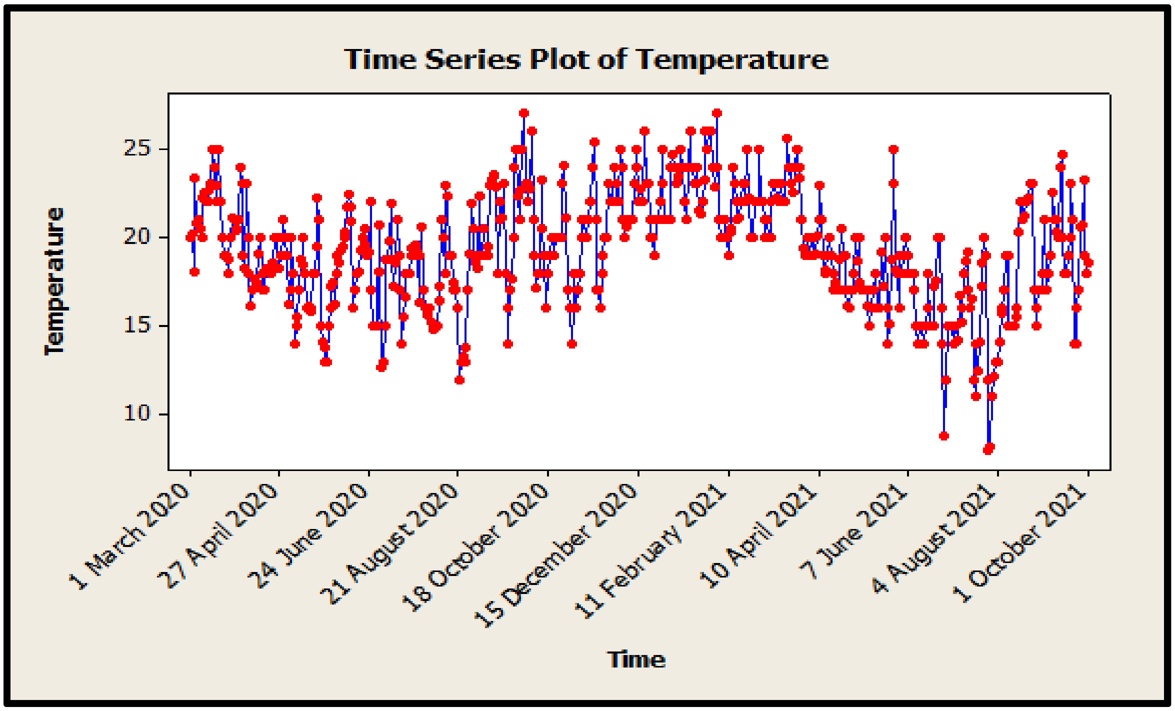

Figure (17): Time series plot for Brazil's Temperature

★ Comment:

The graph demonstrates that there are no sudden shifts, trends, outliers, seasonal effects or cyclic movements. The data for Brazil shows strong random variations. The period from the beginning of March 2020 till that of October 2021 has experienced a relative average temperature measure till October 2021, and then it decreased till April 2021, ending with the lowest measures in July, 2021 and August 2021, and a gradual increase mode till the beginning of October, 2021. There are a number of outliers found in the temperature data for Brazil but the extremist one was on July 29, 2021 to be 8.

2- Humidity

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
78.328	69.973	42.000	94.000	52.000	0	580

Table (26) : Descriptive statistics for Brazil's Humidity

★ **Comments:**

- The average humidity in Brazil throughout the period from March 2020 till October 2021 was 78.328.
- The variability of the humidity data was found to be approximately 69.973 .
- The highest humidity record that hit Brazil was 94, while the lowest one was 42.
- The humidity variable's range was 52.
- The Non-Missings was 580, and this was reflected on the Missings to be zero.

❖ **<u>Time Series Plot:</u>**

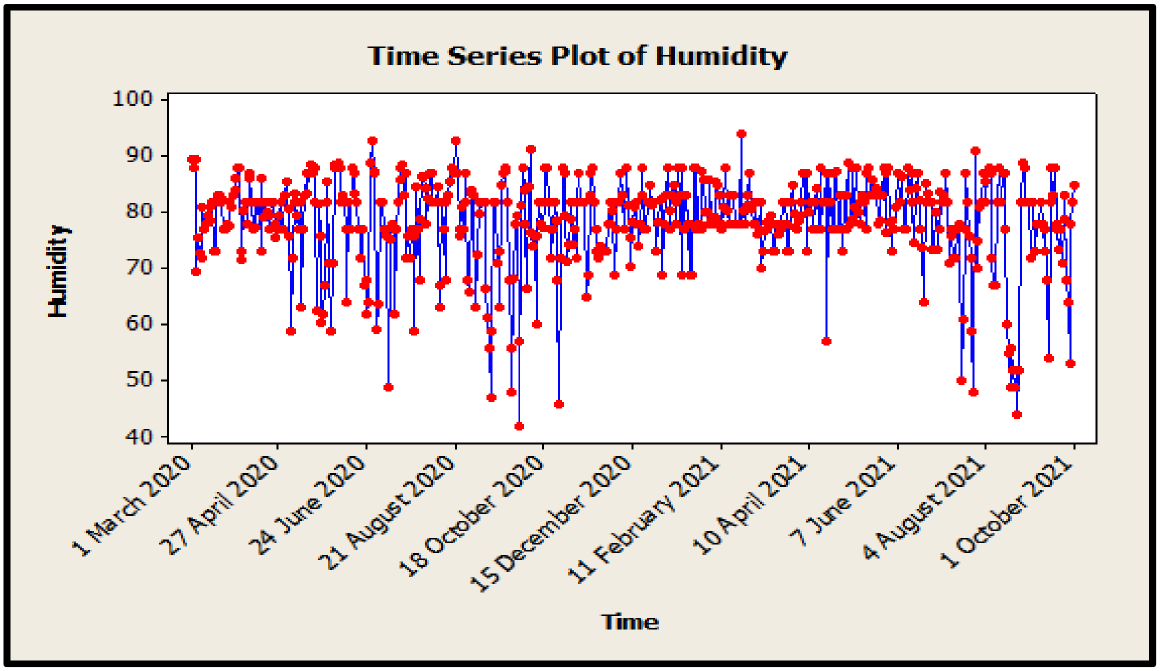

Figure (18): Time series plot for Brazil's Humidity

★ **<u>Comment:</u>**

It's highly found that there's neither trend nor seasonal effect or cyclic movements over the humidity's graph in the whole 21 month. However, there are noticeable extreme variations/fluctuations. There are many noticed extreme outliers for this variable in Brazil as it's strongly observed on October 2, 2020 to be 42 and on February 24, 2021 to be 94.

3- O3

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
14.582	18.468	4.900	32.900	28.000	0	580

Table (27) : Descriptive statistics for Brazil's O3

★ **Comments:**

- The average O3 value in Brazil throughout the period from March 2020 till October 2021 was 14.582.
- The variance of the O3 was 18.468 approximately.
- The highest O3 level was 32.9, while the lowest one was 4.9.
- The O3 variable's range equaled around 28.
- The Non-Missings was 580, and this was reflected on the Missings to be zero.

❖ <u>Time Series Plot:</u>

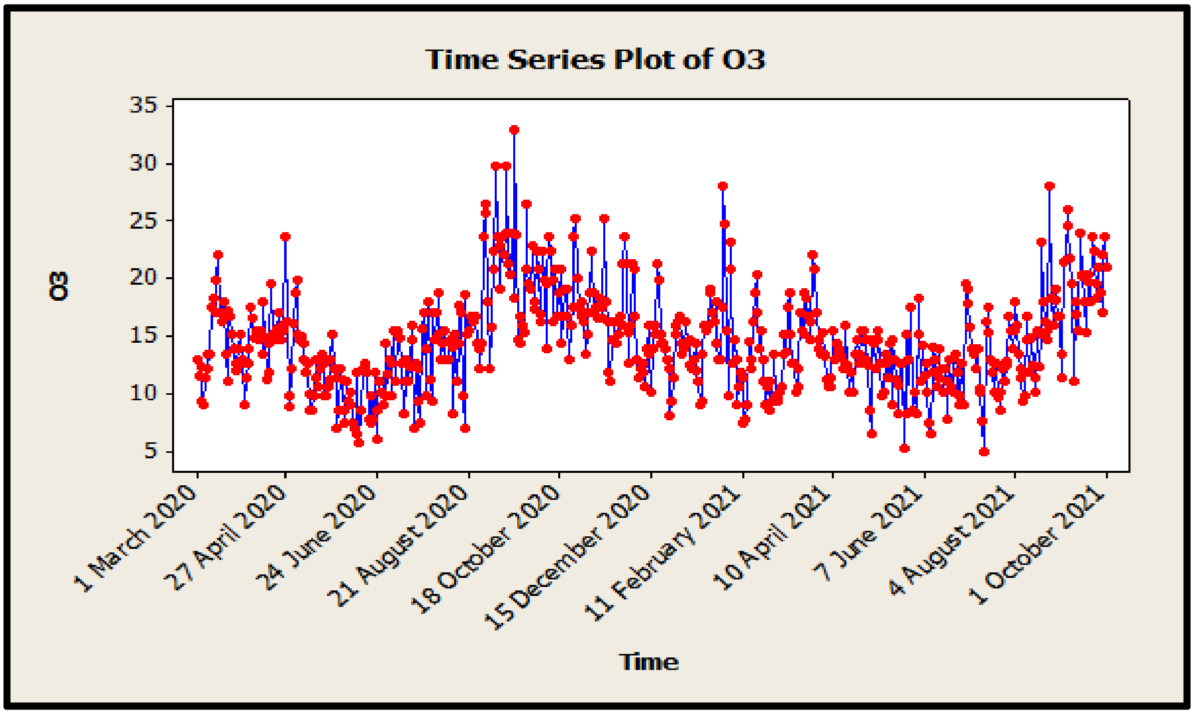

Figure (19): Time series plot for Brazil's O3

★ <u>Comment:</u>

It's noticed the existence of 7 extreme outliers in the O3 data and the most observed one was found to be on 19th September, 2020. The data of O3 shows random variations over the months of the mentioned period, besides a slight indication of the presence of a seasonal effect pattern. There isn't any indication for a trend, or cyclic movements.

4-Daily new cases

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
36975	598521300	0	115041	115041	0	580

Table (28) : Descriptive statistics for Brazil's Daily new cases

★ **Comments:**

- The average number of the Daily New Cases in Brazil from March 2020 till October 2021 was 36975.
- The variability of the Daily New Cases was found to be about 598521300.
- The highest number of the Daily New Cases recorded was 115041, while the lowest one was 0.
- The range of the Daily New Cases was equal to 11504.
- The Non-Missings was 580, and this was reflected on the Missings to be zero.

❖ Time Series Plot:

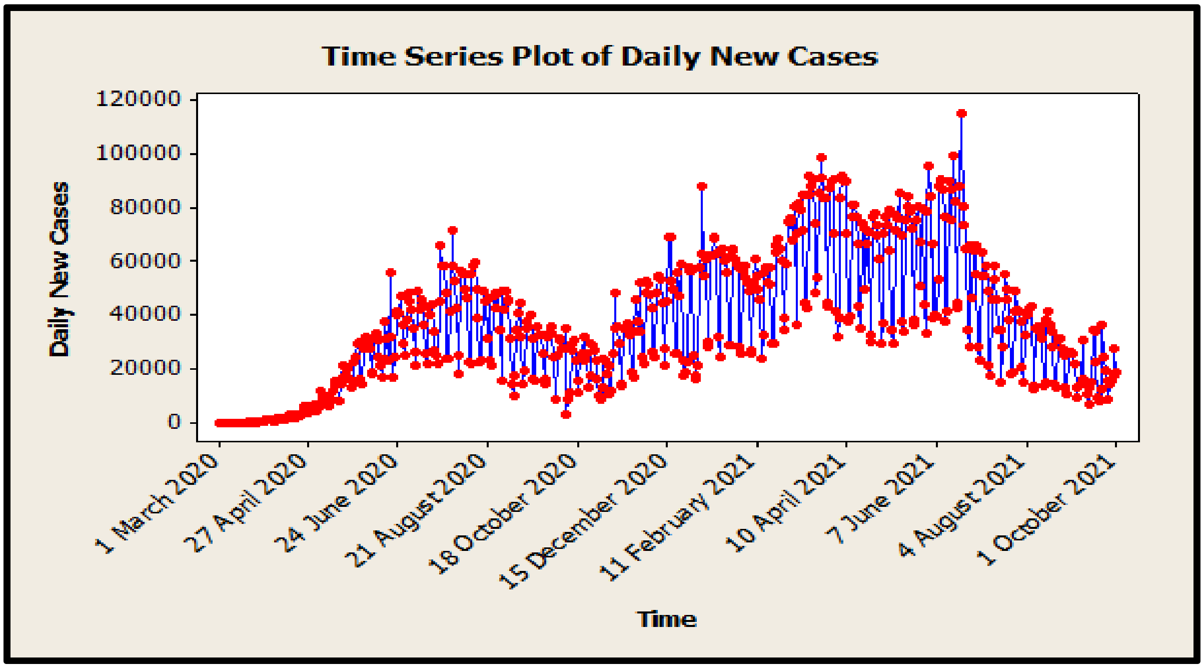

Figure (20): Time series plot for Brazil's Daily new cases

★ Comment:

There are noticed fluctuations and variations for the number of the daily new cases over the time mentioned, the plot shows an increasing mode till November 2020 then it has experienced a saddle point where it began to decrease, then increased again in the first quarter of 2020, until it reached its peak on June 23, 2020 to be equal 115041 cases, and it has experienced some slight ups and downs in 2021 till it decreased in Brazil at the beginning of October 2021. There were neither trends nor cyclic movements or seasonal patterns.

❖ Covariances

Variables	Temperature	Humidity	O3	Daily new cases
Temperature	**10**			
Humidity	**-7**	**70**		
O3	**3**	**-7**	**18**	
Daily new cases	**-196**	**21700**	**-23299**	**598521300**

Table (29) : Covariances between the variables

★ **Comments:**

- As the temperature increases in Brazil throughout the previously mentioned period, the humidity tends to decrease.
- As the temperature increases, the O3 in Brazil tends to increase.
- As the temperature increases, the number of Daily New Cases in Brazil tends to decrease.
- As the temperature increases, the number of the Active Cases in Brazil tends to decrease.
- As the humidity increases, the O3 tends to decrease.
- As the humidity increases, the number of the Daily New Cases tends to increase.
- As the humidity increases, the number of the Active Cases tends to increase.
- As the O3 increases, the number of the Daily New Cases tends to decrease.
- As the O3 increases, the number of the Active Cases tends to decrease.
- As the number of the Daily New Cases in Brazil increases, the number of the Active Cases tends to increase.

❖ **Correlations**

Variables	Temperature	Humidity	O3	Daily new cases
Temperature				
Humidity	-0.268			
O3	0.234	-0.184		
Daily new cases	-0.002	0.106	-0.222	

Table (30) : Correlation between the variables

★ **Comments:**

- There is an approximate weak negative linear relationship between the temperature and the humidity.
- There is an approximate weak positive linear relationship between the temperature and the O3 in Brazil.
- There is a very weak negative linear relationship between the temperature and the number of the Daily New Cases of COVID-19.
- There is a very weak negative linear relationship between the temperature and the number of the Active Cases of COVID-19.
- There is a very weak negative linear relationship between the humidity and the O3 in Brazil.
- There is a very weak positive linear relationship between the humidity and the number of COVID-19 Daily New Cases.
- There is a very weak positive linear relationship between the humidity and the number of COVID-19 Active Cases.
- There is an approximate weak negative linear relationship between the O3 level and the number of the Daily New Cases.
- There is a very weak negative linear relationship between the O3 level and the number of the Active Cases.
- There is an approximate strong positive linear relationship between the number of the Daily New Cases of COVID-19 and that of the Active Cases.

Australia's Descriptive Analysis

1- Temperature

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
16.150	13.156	9.4	25	15.6	0	580

Table (31) : Descriptive statistics for Australia's Temperature

★ **Comments:**

- The maximum median temperature was 25 while the minimum was 9.4.
- The range was about 15.6.
- Temperature dispersion was about 13.2 approximately.
- It seems like the number of non-missing values in the temperature data (N) is 580 and the number of missing values (N^*) is zero.
- The average temperature median value was about 16.2 approximately.

❖ <u>Time Series Plot</u>

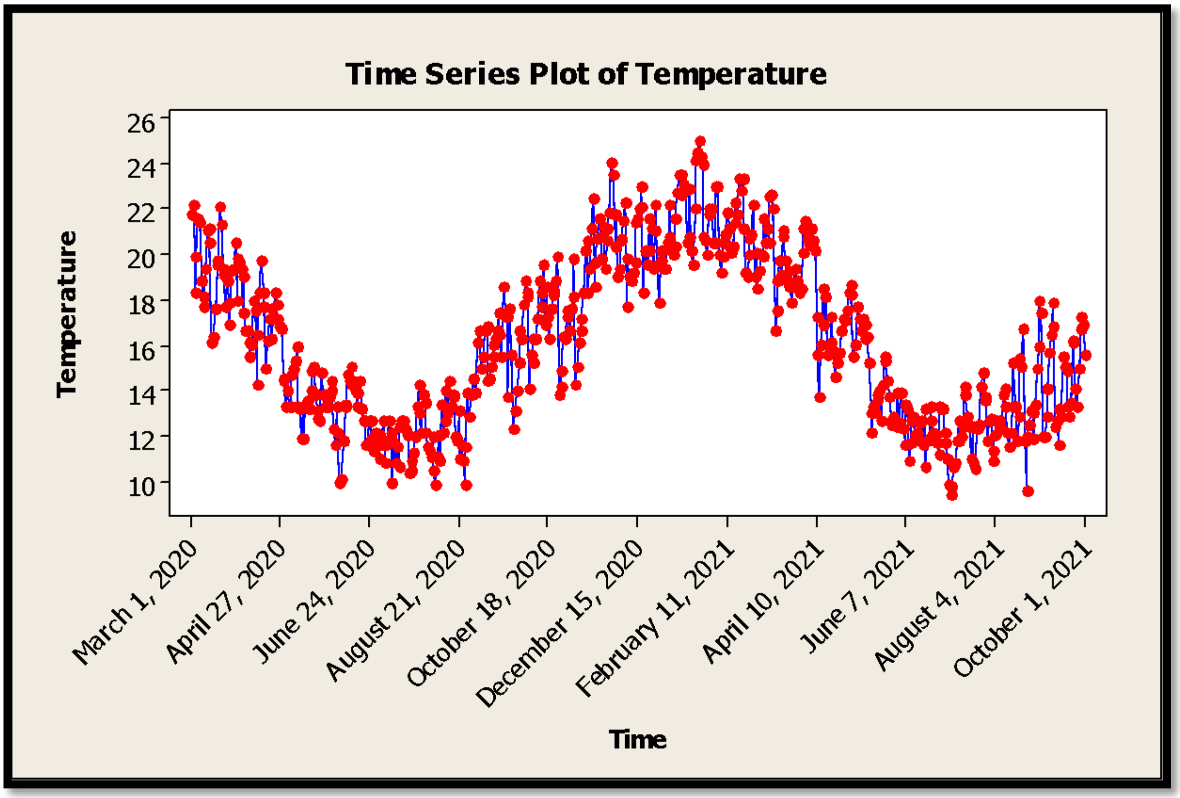

Figure (21): Time Series plot for Australia's Temperature

★ <u>Comment:</u>

It seems that there is no trend in the temperature during the 21 months we are studying, but it seems like there is a seasonal pattern with obvious fluctuations over time without any kind of outliers.

2- Humidity

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
73.554	48.835	49.2	91	41.8	0	580

Table (32) : Descriptive statistics for Australia's Humidity

★ **Comments:**

- The maximum median value of humidity was 91 while the minimum was 49.2.
- The range was about 41.8 approximately.
- Humidity dispersion was about 48.835 approximately.
- It seems like the number of non-missing values in the humidity data (N) is 580 and the number of missing values (N^*) is zero.
- The average of humidity median values was about 73.554 approximately.

❖ **Time Series Plot**

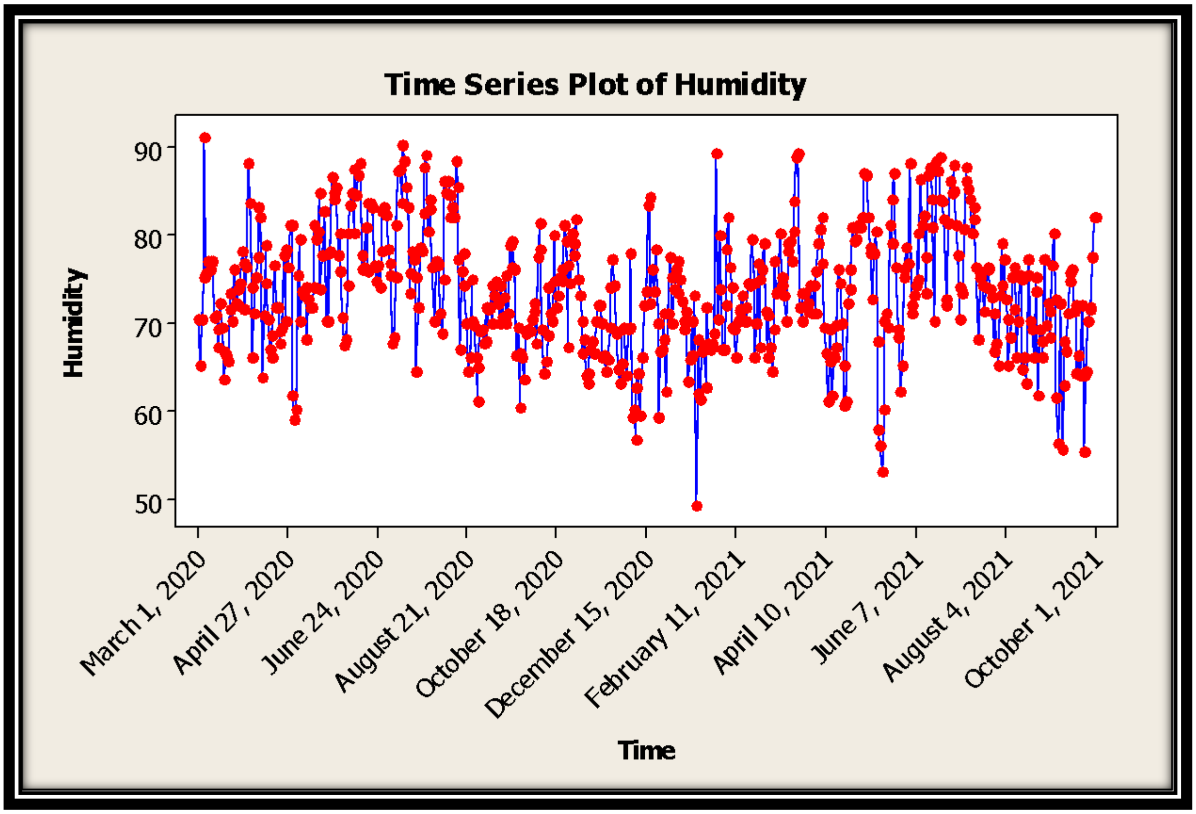

Figure (22): Time Series plot for Australia's Humidity

★ **Comment:**

It seems like there is a huge amount of fluctuations without any kind of trend or saddle points with a noticeable number of outliers but the most irregular one was 55.5 on September 9, 2021. And the lowest median value of humidity was 49.2 on January 16, 2021.

3- O3

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
10.107	**7.806**	**1.6**	**17.7**	**16.1**	**0**	**580**

Table (33) : Descriptive statistics for Australia's O3

★ **Comments:**

- The maximum median value of O3 was 17.7 approximately while the minimum was 1.6
- The range was about 16.1 approximately.
- O3 dispersion was about 7.81 approximately.
- It seems like the number of non-missing values in the O3 data (N) is 580 and the number of missing values (N^*) is zero.
- The average of O3 median values was about 10.11 approximately.

❖ <u>Time Series Plot</u>

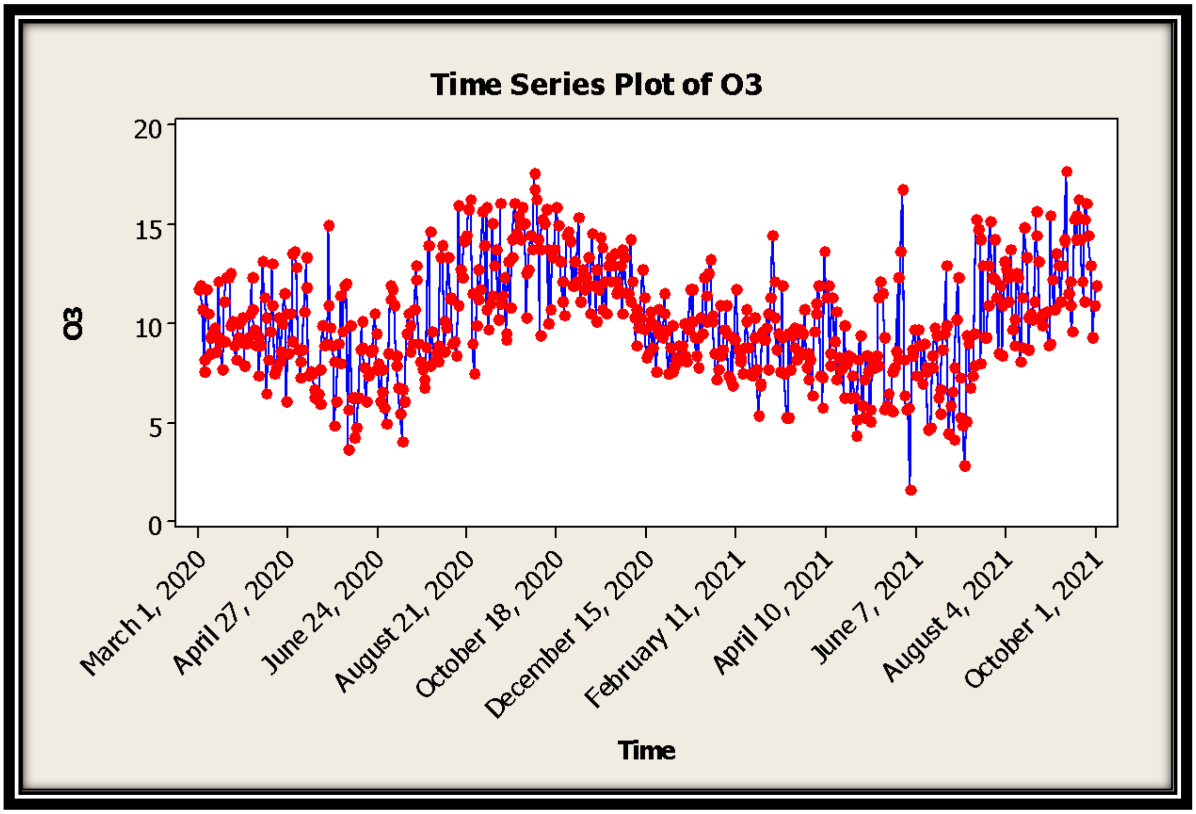

Figure (23): Time series plot for Australia's O3

★ <u>Comment:</u>

A huge amount of fluctuations exist in O3 data without any kind of trend with a pattern that seems to be a seasonality pattern. It seems like there is only one outlier (which is the lowest value in the O3 data) with a value of 1.6 on June 3, 2021.

4-Daily new cases

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
184.7	170711.1	1	2400	2399	0	580

Table (34) : Descriptive statistics for Australia's Daily new cases

★ **Comments:**

- The maximum number of the daily new infected cases with Covid-19 was 2400 while the minimum was 1 case.
- The range was about 2399 cases.
- The dispersion was not very high compared to other countries' situations; about 170711 approximately.
- It seems like the number of non-missing values in the daily new cases data (N) is 580 and the number of missing values (N^*) is zero.
- The average number of daily new cases was about 185 cases, which is considered to be a low average compared to other countries.

❖ **Time Series Plot**

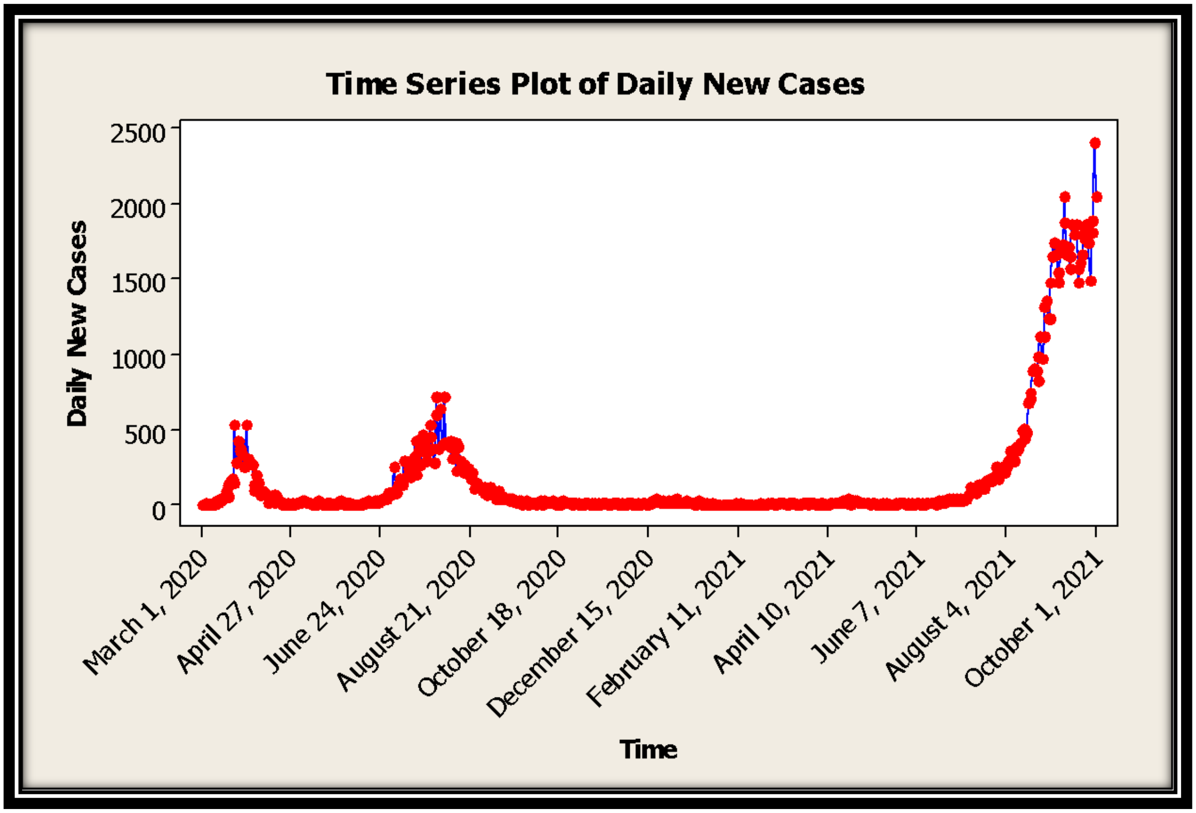

Figure (24): Time Series plot for Australia's Daily New Cases

★ **Comment:**

In this plot we can notice that the period from August 2020 till June 2021 had stability in the number of new daily infected Covid-19 cases, however the period from March 2020 till July 2020 had three different patterns; increasing, decreasing and stability pattern and some outliers do exist in this period. The most fruitful period of Covid-19 daily new cases in Australia was at the end of 2021 where the maximum number of the infected people in Australia was 2400 on September 30, 2021. This figure shows no sign of seasonality. However, fluctuations and outliers do exist.

❖ **Covariances**

Variables	Temperature	Humidity	O3	Daily new cases
Temperature	**13.16**			
Humidity	**-6.19**	**48.84**		
O3	**0.68**	**-9.09**	**7.81**	
Daily new cases	**-337.83**	**-302.86**	**306.93**	**170711.13**

Table (35) : Covariances between the variables

★ **Comments:**

It is well known that we can use the covariance to determine the direction of a linear relationship between two variables as follows:

- As Temperature increases Humidity tends to decrease.
- As Temperature increases O3 tends to increase as well.
- As Temperature increases Daily New Cases tend to decrease.
- As Humidity increases O3 tends to decrease.
- As Humidity increases the Daily New Cases tend to decrease.
- As O3 increases Daily New Cases tend to increase as well.

❖ Correlations

Variables	Temperature	Humidity	O3	Daily new cases
Temperature				
Humidity	**-0.244**			
O3	**0.067**	**-0.466**		
Daily new cases	**-0.225**	**-0.105**	**0.266**	

Table (36) : Correlation between the variables

★ **Comments:**

- It is well known that the correlation between any variable and itself is a strong perfect relation = 1.
- There is a weak negative linear relation between Temperature and Humidity.
- There is a very weak positive linear relation between Temperature and O3.
- There is a weak negative linear relation between Temperature and Daily New Cases.
- There is a moderate negative linear relation between Humidity and O3.
- There is a weak negative linear relation between Humidity and Daily New Cases.
- There is a weak positive linear relation between O3 and Daily New Cases.

Italy’s Descriptive Analysis

1- Temperature

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
16.606	52.979	-0.100	29.700	29.800	0	580

Table (37) : Descriptive statistics for Italy’s Temperature

★ **Comments :**

- The average degrees for the temperature were around 16.606 during the period 2020-2021.
- For temperature, the variability of our data is about 52.979 approximately.
- The maximum temperature was approximately 29.700 while the minimum was -0.100.
- For temperature the range was about 29.800.
- It seems like the number of non-missing values in the temperature data (N) is 580 and the number of missing values (N*) is zero.

❖ **Time series plot**

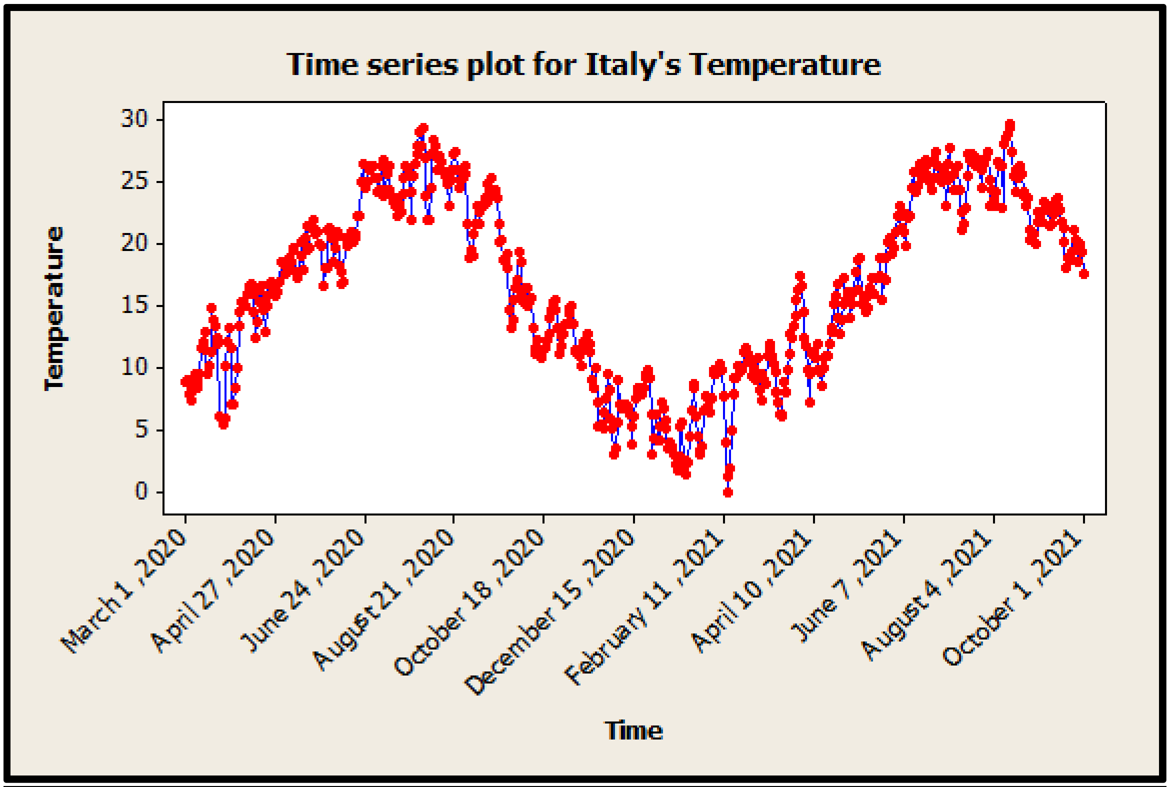

Figure (25): Time series plot for Italy's Temperature

★ **Comment :**

It seems that there is neither trend nor outliers over the period 2020-2021.It's clear that the temperature has obvious fluctuations regularly over time representing a pattern of seasonality.

2- Humidity

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
67.281	193.724	30.000	93.250	63.250	0	580

Table (38) : Descriptive statistics for Italy's Humidity

★ **Comments :**

- The average value of the humidity was around 67.281 during the period 2020-2021.
- For humidity the variability of our data is approximately 193.724.
- The maximum humidity was approximately 93.250 while the minimum was 30..
- For humidity the range was about 63.250.
- It seems like the number of non-missing values in the humidity data (N) is 580 and the number of missing values (N*) is zero.

❖ **<u>Time series plot</u>**

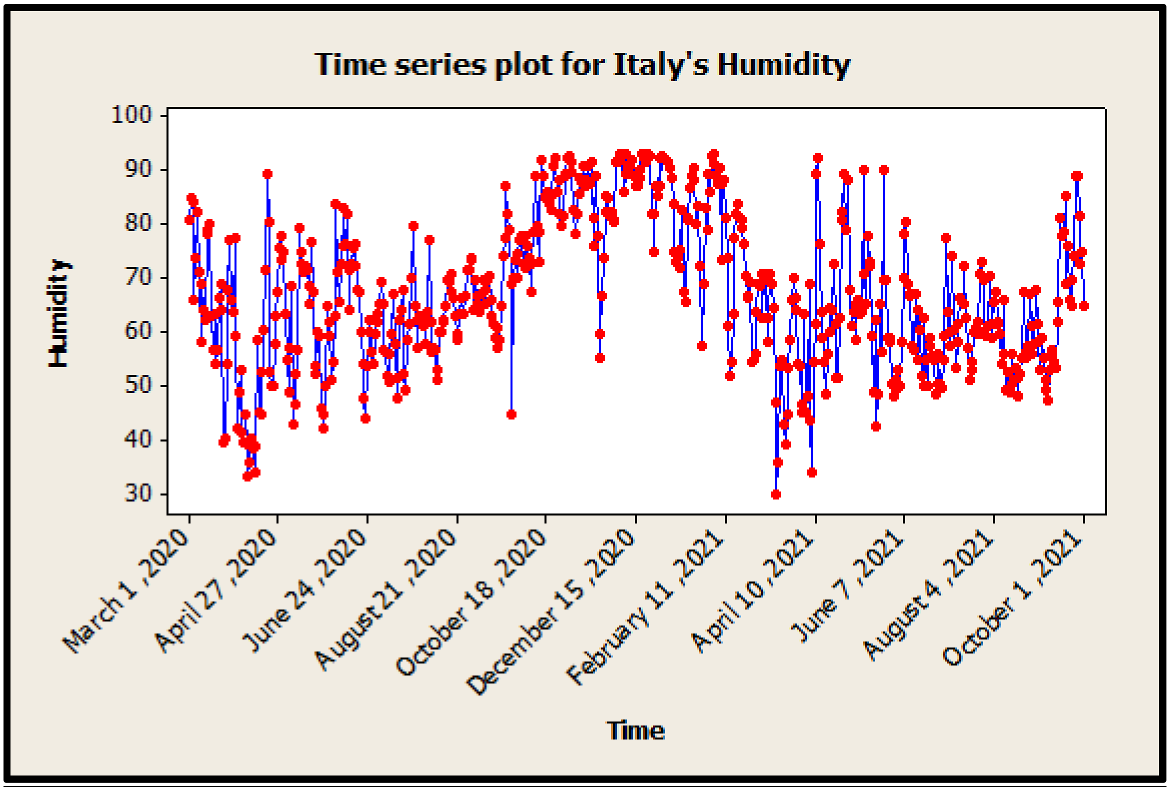

Figure (26): Time series plot for Italy's Humidity

★ **<u>Comment</u> :**

There is a large amount of fluctuations without any kind of trend or saddle points and not even a seasonality effect in the humidity values over time. There are no outliers.

3- O3

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
32.191	234.300	3.300	78.000	74.700	0	580

Table (39) : Descriptive statistics for Italy's O3

★ **Comments :**

- The average value of the O3 was around 32.191 during the period 2020-2021.
- For O3 the variability of our data is approximately 234.300.
- The maximum humidity was approximately 78 while the minimum was 3.30.
- For O3 the range was about 74.700.
- It seems like the number of non-missing values in the O3 data (N) is 580 and the number of missing values (N*) is zero.

❖ <u>Time series plot</u>

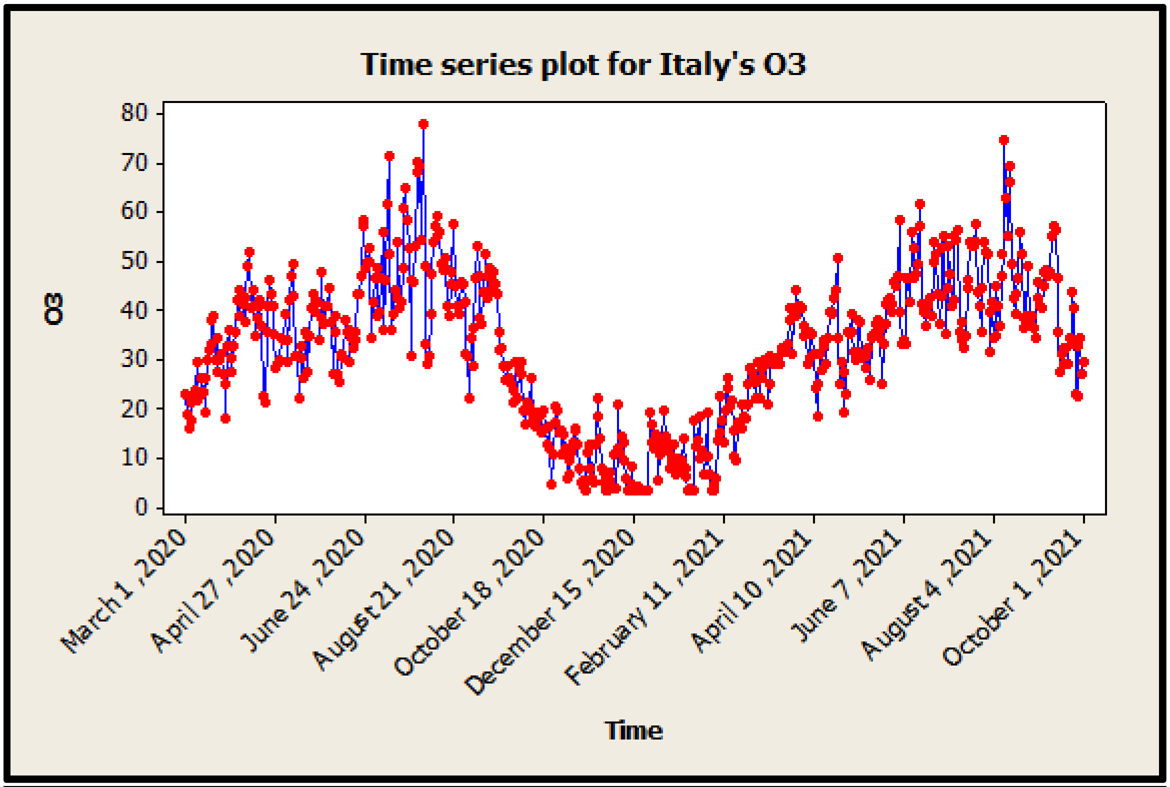

Figure (27): Time series plot for Italy's O3

★ **<u>Comment</u> :**

There is a large amount of fluctuations in O3 data with almost a seasonal pattern without any kind of trend. There are some outliers and we can say that there is an outlier appearing on 1st of August 2020 with a value of 78.

4-Daily new cases

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
8060	**78382881**	**113**	**41198**	**41085**	**0**	**580**

Table (40) : Descriptive statistics for Italy's Daily new cases

★ **Comments :**

- The average number of daily new cases was around 8060 during the period 2020-2021.
- For Daily new cases the variability of our data is approximately 78382881.
- The maximum number of daily new cases was approximately 41198 while the minimum was 133.
- For daily new cases the range was about 41085.
- It seems like the number of non-missing values in Daily new cases data (N) is 580 and the number of missing values (N*) is zero.

❖ **Time series plot**

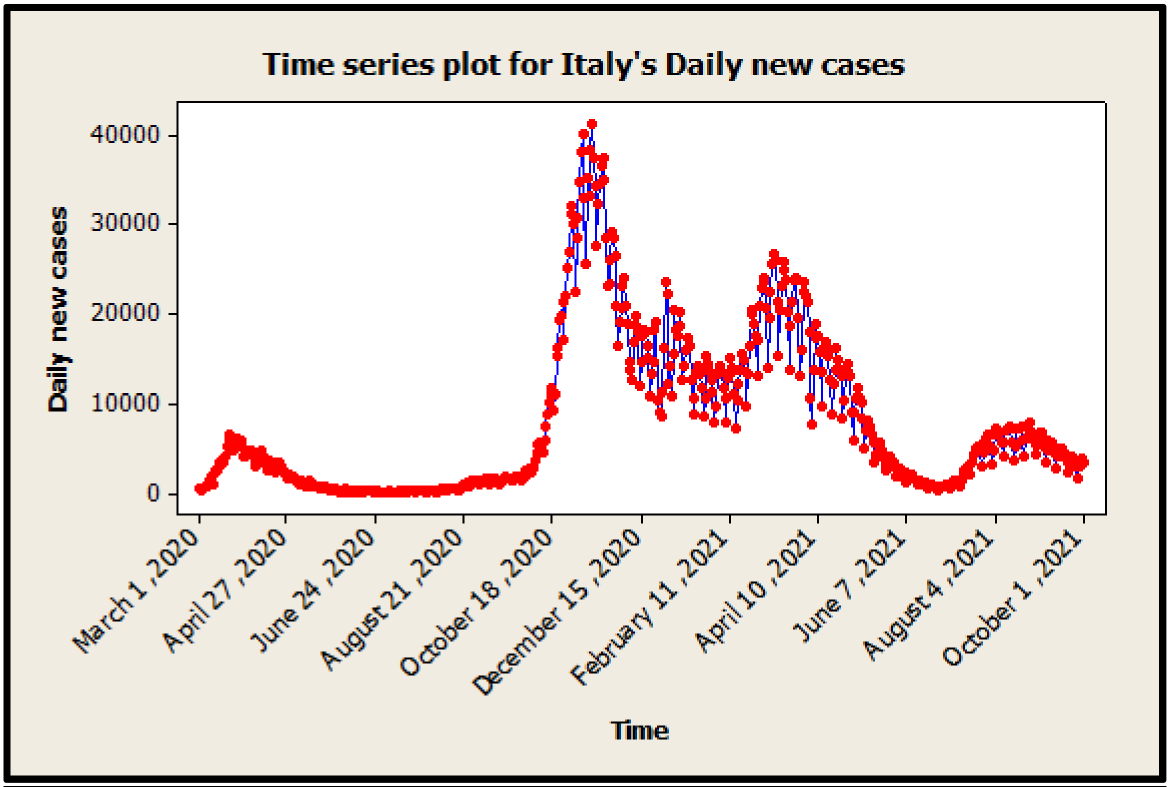

Figure (28): Time series plot for Italy's Daily new cases

★ **Comment :**

There is a moderate level of fluctuations in the daily new infected cases with Covid-19, at first, the number of new cases were increasing slightly then decreased also slightly till almost constant showing a kind of stability status but from 18 October 2020 till the mid of April 2021 it showed a high increase at some points of time then decreases d after and it continued to decrease showing still average fluctuations and then slightly increased one more time at the beginning of august 2021 and decreased again. Hence, there is no pattern of trend or seasonality. There are some outliers and we can say that those outliers appear in the time interval from the 31th October 2020 to 21th November 2020 and the most anomalous values were 31959 cases on 31th October 2020 and 41198 cases on 13th November 2020.

❖ **Covariances**

Variables	Temperature	Humidity	O3	Daily new cases
Temperature	**53**			
Humidity	**-44**	**194**		
O3	**92**	**-154**	**234**	
Daily new cases	**-40907**	**47532**	**-83969**	**78382881**

Table (41) : Covariances between the variables

★ **Comments :**

To determine the direction of a linear relationship between two variables we have used the covariances from where we conclude that:

- As Temperature increases, Humidity tends to decrease.
- As Temperature increases, O3 tends to increase as well.
- As Temperature increases, Daily New Cases tend to decrease.
- As Humidity increases, O3 tends to decrease.
- As Humidity increases, the Daily New Cases tend to increase.
- As O3 increases, Daily New Cases tend to decrease.

❖ Correlations

Variables	Temperature	Humidity	O3	Daily new cases
Temperature				
Humidity	**-0.433**			
O3	**0.822**	**-0.721**		
Daily new cases	**-0.635**	**0.386**	**-0.620**	

Table (42) : Correlation between the variables

★ **Comments :**

To assess the strength of the relationship between two variables we have used the correlation between two variables from where we concluded that:

(Note that: the correlation between any variable and itself is a strong perfect relation = 1).

- There is a moderate negative linear relation between Temperature and Humidity.
- There is a very strong positive linear relation between Temperature and O3.
- There is a strong negative linear relation between Temperature and Daily New Cases.
- There is a strong negative linear relation between Humidity and O3.
- There is a weak positive linear relation between Humidity and Daily New Cases.
- There is a strong negative linear relation between O3 and Daily New Cases.

South Africa's Descriptive Analysis

1- Temperature

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
16.565	17.023	6.200	24.6000	18.400	0	580

Table (43) : Descriptive statistics for South Africa's Temperature

★ **Comments :**

- Average temperature of South Africa during this period was 16.565.
- Lowest temperature in South Africa reaches 6.2 and the highest temperature reaches 24.6.
- Variability of temperature data was approximately 17.023.
- Number of non-missing values was 580 and the number of missing equals zero.
- Range of our temperature variable reaches 18.4.

❖ <u>Time series plot</u>

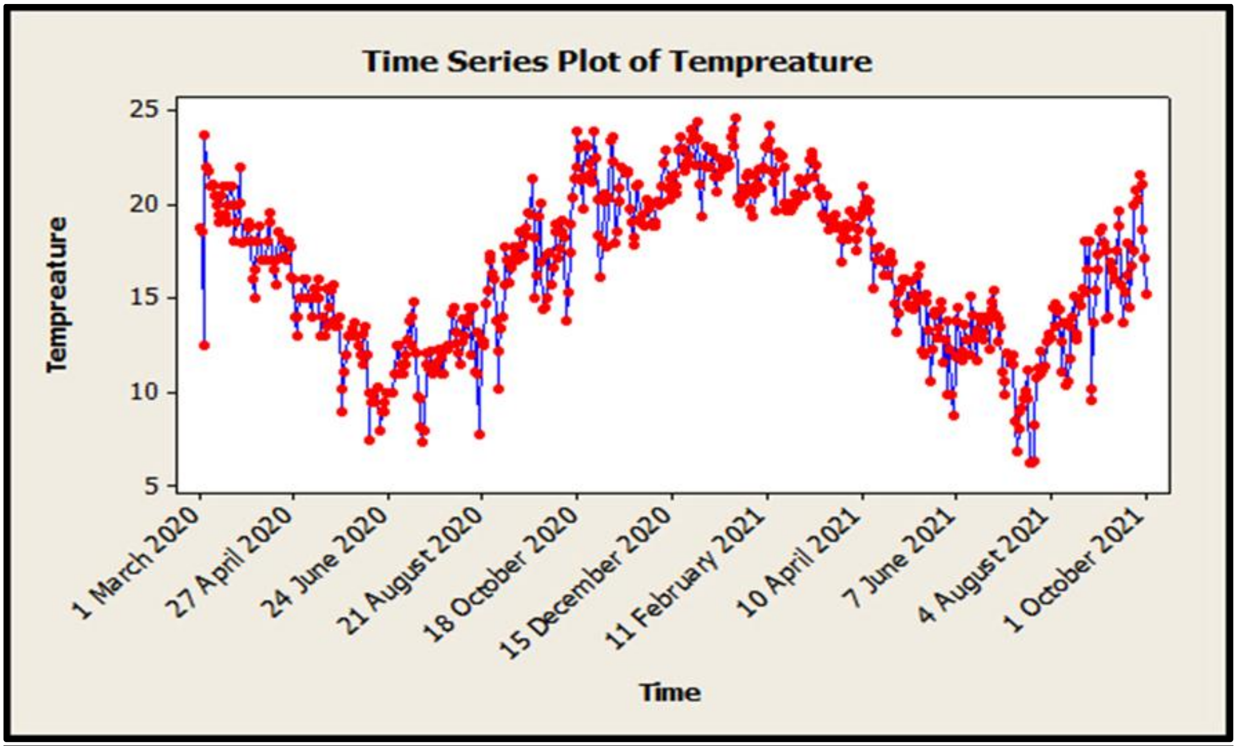

Figure (29): Time series plot for South Africa's Temperature

★ **<u>Comment</u> :**

In this plot we observe that there are some fluctuations and there's seasonality, and it seems like that there are no outliers and nor trend.

2- Humidity

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
66.181	135.866	33.150	90.750	57.6	0	580

Table (44) : Descriptive statistics for South Africa's Humidity

★ **Comments :**

- Average value for humidity in South Africa during this period was 66.181.
- Lowest value of humidity in South Africa reached 33.15 and the highest value was 90.75.
- Variability of South Africa data was 135.866 approximately.
- Number of non-missing values was 580 and the number of missing equals zero.
- Range of humidity was 57.6.

❖ **Time series plot**

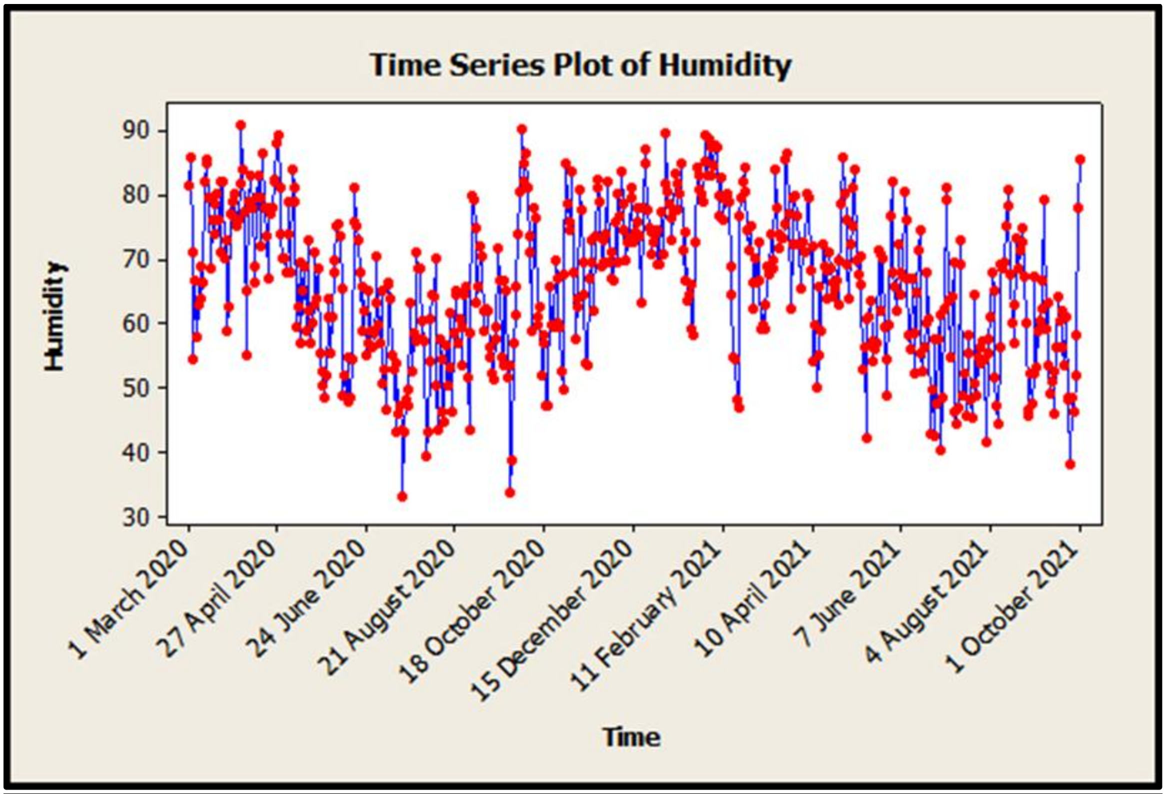

Figure (30): Time series plot for South Africa's Humidity

★ **Comment :**

In this plot we observe that there is a large amount of fluctuations and most values range from 40 up to 90. There are no outliers and nor trend is observed.

3- O3

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
7.3952	4.6788	1.900	15.600	13.7	0	580

Table (45) : Descriptive statistics for South Africa's O3

★ **Comments :**

- Average value of O3 in South Africa during this period was approximately 7.
- Lowest value of O3 during this period reached 1.9 and the highest value was 15.6.
- Variability of O3 data was 5 approximately.
- Number of non-missing values was 580 and the number of missing equals zero.
- Range of humidity was 13.7.

❖ **Time series plot**

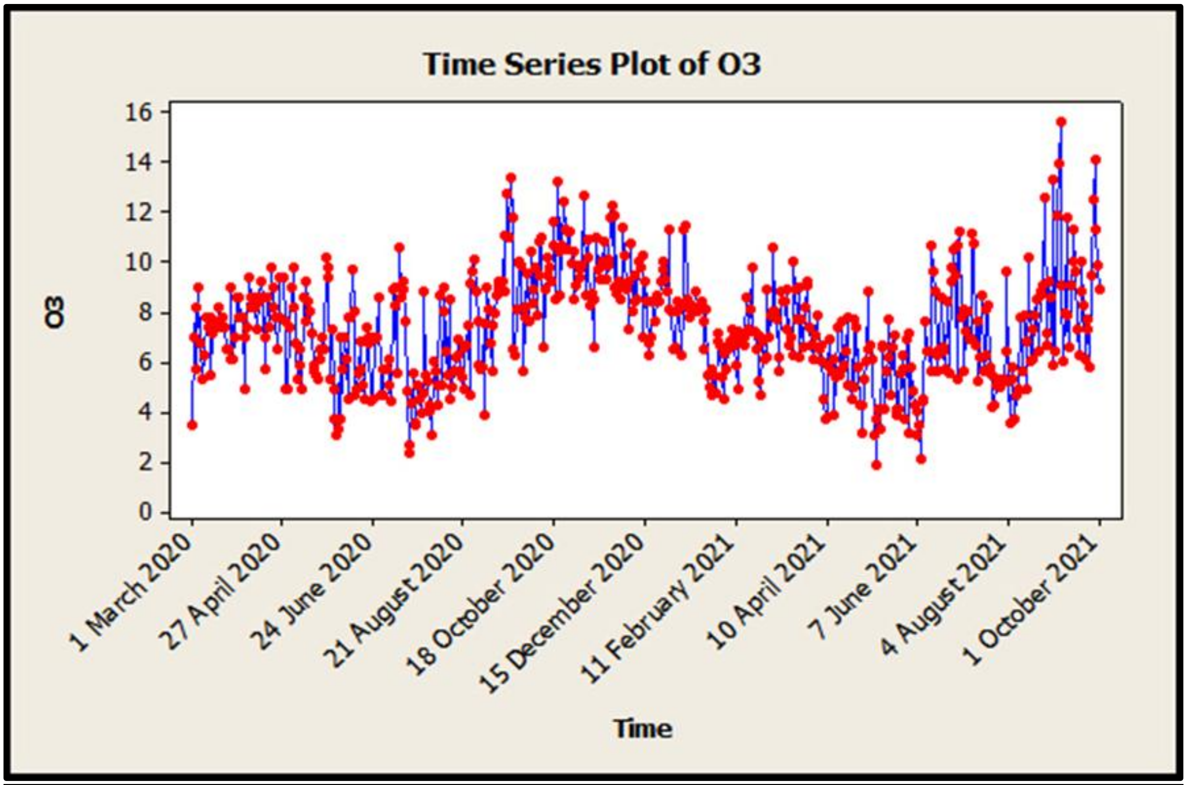

Figure (31): Time series plot for South Africa's O3

★ **Comment :**

In this plot we observe that there are some fluctuations and are varying in constant range from 2 up to 12. There are three outliers whose values equal to 13.95, 15.60, and 14.10 on Sep, 5 2021, Sep, 6 2021 and Sep, 28 2021 respectively. No trend was observed.

4-Daily new cases

❖ **Descriptive statistics**

Mean	Variance	Minimum	Maximum	Range	Missings	Non missings
5082	28976998	0	26645	26645	0	580

Table (46) : Descriptive statistics for South Africa's Daily new cases

★ **Comments :**

- Average number of daily new cases in South Africa during this period was 5082 cases.
- Lowest number of daily new cases during this period was 0 and the highest was 26645.
- The dispersion of the new cases was approximately 28976998.
- Number of non-missing values was 575 and the number of missing equals 0.
- Range of daily new cases was 26645.

❖ **Time series plot**

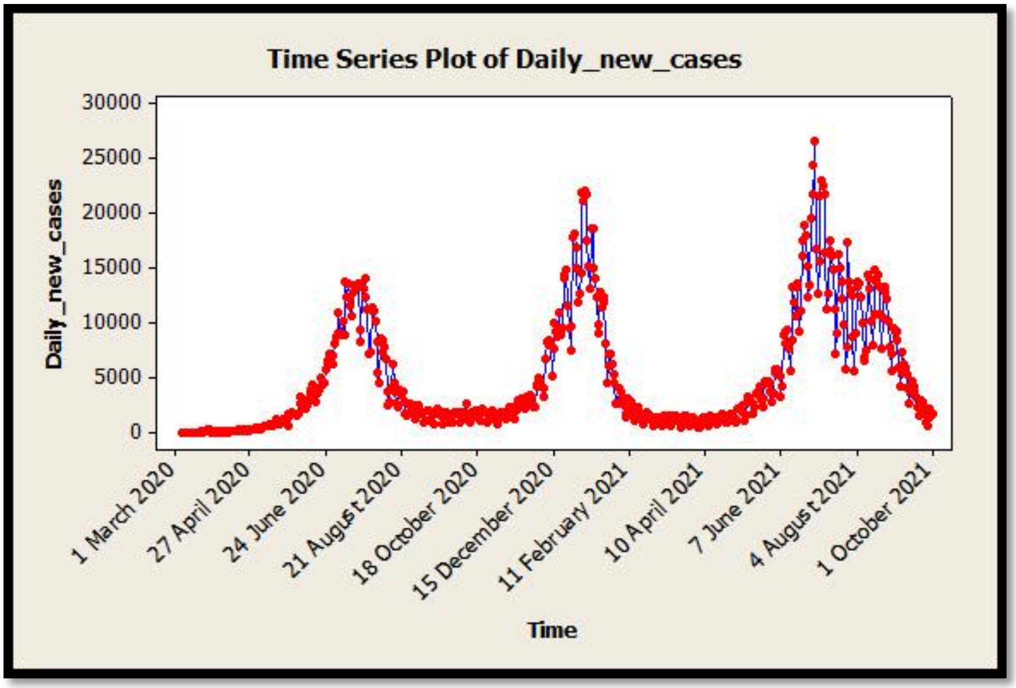

Figure (32): Time series plot for South Africa's Daily new cases

★ **Comment :**

In this plot we observed that there's an obvious repeated fluctuations as the daily new cases increase then decrease through the same time periods approximately which indicates presence of seasonality. We have 14 outliers however the highest one takes the value 26645 on July 3 2021 which is the maximum value of daily new cases.

- **<u>Covariances</u>**

Variables	Temperature	Humidity	O3	Daily new cases
Temperature	**17**			
Humidity	**17**	**136**		
O3	**4**	**0**	**5**	
Daily new cases	**-5812**	**-11990**	**-565**	**28976998**

Table (47) : Covariances between the variables

★ **<u>Comments</u> :**

- When the number of daily new cases increases, the number of active cases increases.
- When temperature increases number of active cases decrease
- When temperature increases, the number of daily new cases decreases.
- When temperature increases, the level of humidity increases.
- When humidity increases, the number of daily new cases decreases.
- When humidity increases, the number of active cases decreases.
- When O3 increases, the number of daily new cases decreases.
- O3 and humidity don't affect each other.
- When O3 increases, the number of active cases decreases.
- When O3 increases, temperature increases.

❖ **Correlations**

Variables	Temperature	Humidity	O3	Daily new cases
Temperature				
Humidity	**0.36**			
O3	**0.393**	**0.019**		
Daily new cases	**-0.262**	**-0.191**	**-0.049**	

Table (48) : Correlation between the variables

★ **Comments :**

- There is a strong positive linear relationship between daily new cases and active cases equals (0.885)
- There is a weak negative linear relationship between temperature and daily new cases equals(-0.262)
- There is a weak negative linear relationship between humidity and daily new cases equals(-0.191)
- There is a very weak negative linear relationship between O3 and daily new cases equals(-0.049)
- There is weak negative linear relationship between Active cases and Temperature equals (-0.238)
- There is a very weak negative linear relationship between Active cases and humidity equals (-0.259)
- There is an approximate moderate negative linear relationship between Active cases and O3 equals (-0.015)
- There is a moderate positive linear relationship between temperature and humidity equals (0.36)
- There is a moderate positive equals ٣linear relationship between temperature and O .(٠,٣٩٣)
- There is a very weak negative linear relationship between O3 gas and humidity equals (0.019).

Chapter Four

Panel Data Analysis

Overview

It is well known that in time series data we observe the values of one or more variables over some time. However, in cross-section data, the values of one or more variables are collected for several sample units, or subjects, at the same time point. In panel data, the same cross-sectional unit is surveyed over time. Hence, Panel data analysis is at the watershed of time series and cross-section econometrics. Panel datasets have enriched the set of possible identification arrangements, and forced economists to think more carefully about the nature and sources of identification of parameters of potential interest.

There are other names for panel data, such as pooled data, a combination of time series and cross-section data, micro-panel data, longitudinal data …etc.

Since panel data relate to the cross-sectional units over time, there is bound to be heterogeneity in these units. The techniques of panel data estimation can take such heterogeneity explicitly into account by allowing for subject-specific variables. Panel data combines time series of cross-section observations; hence it gives "more informative data, more variability, less collinearity among variables, more degrees of freedom and more efficiency." For these reasons, we have decided to use Panel Data Analysis instead of the usual Time Series Analysis to have an efficient analysis that accounts for the same country over time.

Balanced and Unbalanced Panel Data:

Panel data might be balanced or unbalanced. A panel is considered to be balanced if each subject (country) has the same number of observations. However, if each entity has a different number of observations, then we have an unbalanced panel.

In the panel data literature there are two common terms: short panel and long panel;

1. In a short panel the number of cross-sectional units (N) is greater than the number of periods (T).
2. In a long panel, T is greater than N.

Long panel data analysis is very complicated and hence we have converted our daily data into seven seasons by taking the median of the data according to the Meteorological Seasons and using R-Programming Language.

What are the Meteorological Seasons? Meteorologists, divide the seasons according to weather rhythms and the Gregorian calendar. According to the Meteorological Seasons, the best way to split the seasons up is to divide the year into quarters, each consisting of three complete calendar months. Spring is then defined as March, April, and May, summer as June, July and August, autumn as September, October, and November, and winter as December, January, and February. These months never change and so enable meteorologists to collate data easily and compare seasonal statistics.

Converting the daily data into seven seasons data and working on eight cross-sectional units (Countries) helped in dealing with the short panel data analysis instead of the long panel data analysis and this is because now we have N=8 > T=7.

Panel Data Analysis/ How to Estimate Panel Data?

This section considers the two most popular panel data models on which our analysis was based, in addition to the Simple Regression Analysis

Panel Data Analysis Techniques:

1. Simple Regression Analysis (not preferable)
2. Random Intercept Model; sometimes in literature called Fixed Effect Model including:
   - Least Squares Dummy Variable Estimator
   - Between Estimator
   - Within Estimator
   - First Difference Estimator
3. Random Effect Model

Simple regression analysis done over the pooled data using the ordinary Least Square (OLS) technique:

One way to estimate panel data is to ignore the fact that we have got panel data and this is called the simple linear regression analysis. This model is as follows:

$$y_{it} = \alpha + \beta' x_{it} + \varepsilon_{it} \quad i = 1,2, \dots N \quad t = 1,2, \dots T$$

In this analysis, we do actually lose too much information (by ignoring the possible existing heterogeneity). We simply pool all the observations and estimate a regression model, ignoring the cross-section and time-series nature of the data.

Random Intercept Model or Fixed Effect Model:

This model assumes that differences between individuals can be accommodated from a different intercept. To estimate the random Intercept model, we can go through different approaches/estimation techniques:

- **The fixed effects least squares dummy variable (LSDV) model,** here, all the observations are pooled, but at the same time, we allow for each cross-section unit (country) to have its own (intercept) dummy variable.

- **The fixed effects within-group model,** in which, we pool all the observations, and for each country, we express each variable as a deviation from its mean value and then estimate an OLS regression on such **"de-meaned" values**.

- **The between estimator,** in which we run a cross-sectional regression in the time-averaged values of the variables.

- **The First Difference Estimator,** in which we are modeling the change in y_{it}rather than modelingy_{it} itself.

Random Effect Model:

The **random-effects model (REM).** Here, the intercept values are a random drawing from a much bigger population of cross-sectional units "countries".

In this chapter we are going to make a panel data analysis on our Covid-19 Climate data to get a clear answer for our main research question in this paper which is "Can Climate affect Covid-19?" in a try to answer the research question, we have considered the daily new cases of Covid-19 as the dependent variable, and all of O3, Humidity and Temperature as the independent variables.

Exploring the Panel Data

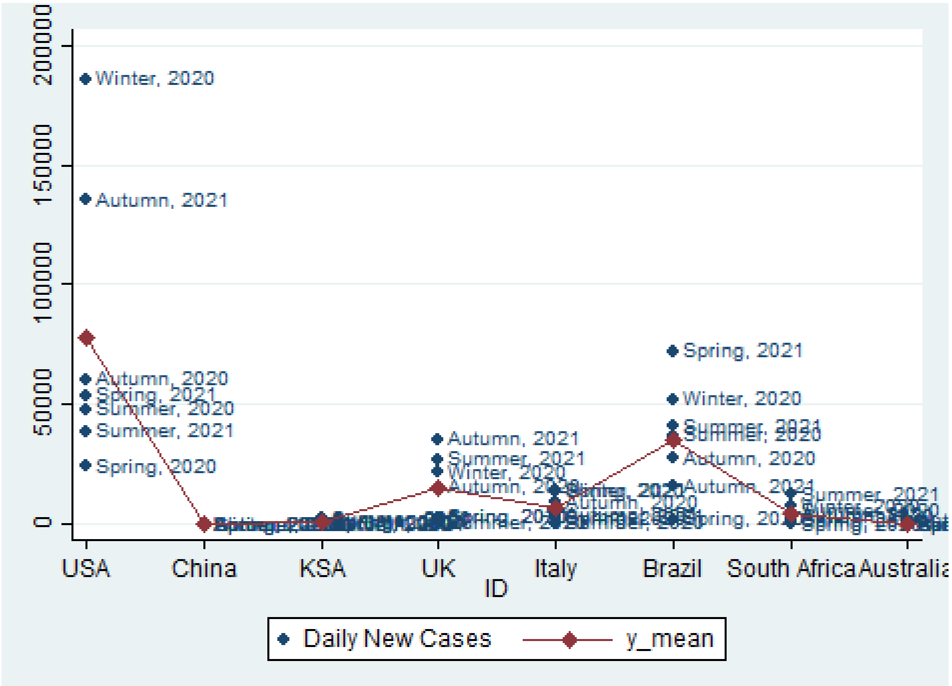

Figure (1): Heterogeneity across Countries

<u>Comment:</u>

It seems like the data has a very obvious nature of heterogeneity. The variations in the number of daily new cases varies within and across countries. As it appears China, UKA, Italy, South Africa, and Australia have very low variations; meaning that the number of daily new cases over time does not vary too much. However, for the USA, UK, and Brazil there were noticeable variations in the numbers of the daily new cases over time. We can also notice that the USA has the highest average of the daily new cases, while China has the lowest and that was expected; although the starting point for the outbreak of COVID-19 was China however China managed to control and limit Covid-19 break-out.

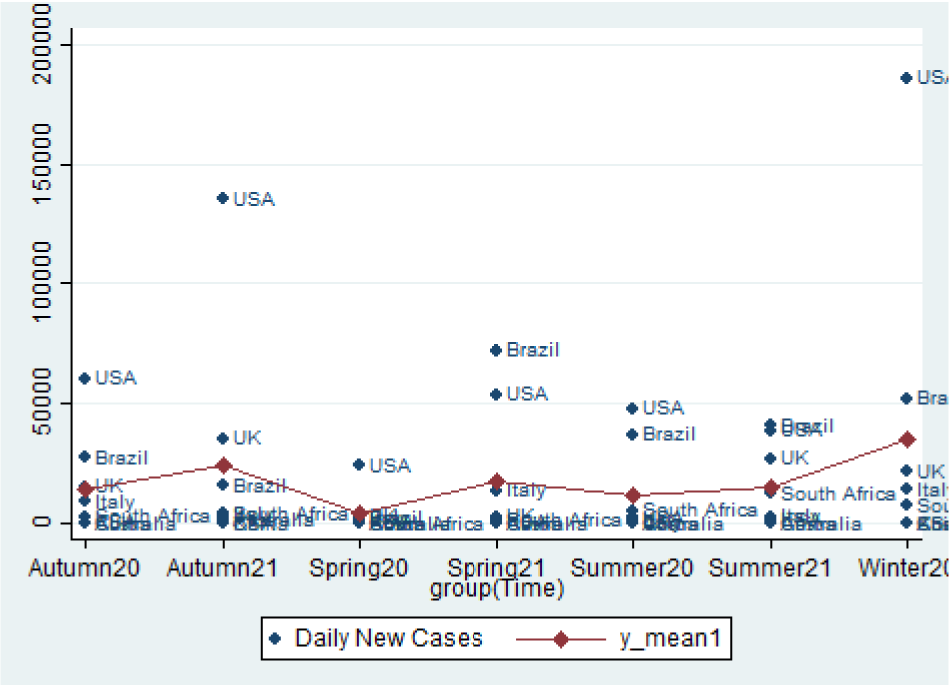

Figure (2): Heterogeneity across Seasons

Comment:

As it seems, there are variations in the average of the daily new cases among the 8 countries over time. In the Spring of 2020; the lowest average number of COVID-19 cases was recorded, while the highest was in the Winter of 2020. The USA occupied first place in the number of daily new cases in Autumn 2021, Autumn 2021, Spring 2020, Summer 2021, and Winter 2020. The highest number of daily new cases was in the USA in Winter 2020 and it can be considered an outlier. Most of the countries managed to limit or control the daily new cases over time except the USA followed by Brazil.

The Main Analysis

Simple Regression Analysis

```
. * Pooled OLS estimator
. reg DailyNewCases Temperature Humidity O3

      Source |       SS           df       MS      Number of obs   =        56
-------------+----------------------------------   F(3, 52)        =      1.15
       Model |  3.8335e+09         3  1.2778e+09   Prob > F        =    0.3384
    Residual |  5.7873e+10        52  1.1129e+09   R-squared       =    0.0621
-------------+----------------------------------   Adj R-squared   =    0.0080
       Total |  6.1706e+10        55  1.1219e+09   Root MSE        =     33361

------------------------------------------------------------------------------
DailyNewCa~s |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 Temperature |  -444.8387    839.513    -0.53   0.598    -2129.445    1239.767
    Humidity |   553.4688    392.639     1.41   0.165    -234.4189    1341.357
          O3 |  -23.51589   483.7575    -0.05   0.961    -994.2464    947.2147
       _cons |   -12278.3   36677.78    -0.33   0.739    -85877.65    61321.05
------------------------------------------------------------------------------
```

Output (1): Simple Regression Model

➔ **Insights from Stata Output:**

The F-value represents the Mean Square Model (1.2778e+09) divided by the Mean Square Residual (1.1129e+09), yielding F=1.15. It is well known that the p-value associated with this F value is used to answer the question "Do the independent variables reliably predict the dependent variable?" Hence, according to the simple regression model output, it seems like the p-value compared to the alpha level (0.05) indicates the insignificance of the simple regression model; the independent variables do not show a statistically significant relationship with the dependent variable.

As we know that the R-Squared is the proportion of variance in the dependent variable (**daily new cases**) which can be predicted/explained from/by the independent variables (**Temperature, Humidity, and O3**). The R-squared value indicates that 6.21% only of the variance in the number of daily new cases can be predicted from the variables Temperature, Humidity, and O3. This indicates that the independent variables are not explaining much in the variation of the dependent variable. Hence, the simple regression model does not fit our data well, and that was expected from the beginning since the simple regression model neglects the cross-section and time-series nature of our data.

Random Intercept Model; Least Squares Dummy Variable (LSDV)

```
. *Fixed Effect MODELS
. * LSDV Estimator, Fixed Effect Model
. reg DailyNewCases Temperature Humidity O3 i.ID
```

Source	SS	df	MS			
Model	3.8063e+10	10	3.8063e+09	Number of obs	=	56
Residual	2.3643e+10	45	525402986	F(10, 45)	=	7.24
				Prob > F	=	0.0000
				R-squared	=	0.6168
				Adj R-squared	=	0.5317
Total	6.1706e+10	55	1.1219e+09	Root MSE	=	22922

DailyNewCa~s	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Temperature	-738.2236	662.6176	-1.11	0.271	-2072.804	596.3568
Humidity	293.4224	491.7129	0.60	0.554	-696.9382	1283.783
O3	-431.9173	593.2894	-0.73	0.470	-1626.863	763.0289
ID						
2	-75319.58	12506.58	-6.02	0.000	-100509.1	-50130.04
3	-64821.3	17865.48	-3.63	0.001	-100804.2	-28838.38
4	-69694.69	13647.91	-5.11	0.000	-97182.99	-42206.39
5	-65900.93	13800.72	-4.78	0.000	-93697.01	-38104.84
6	-46949.43	13398.86	-3.50	0.001	-73936.12	-19962.73
7	-78166.52	15166.86	-5.15	0.000	-108714.2	-47618.89
8	-84190.74	13562.03	-6.21	0.000	-111506.1	-56875.42
_cons	79270.55	41622.8	1.90	0.063	-4562.075	163103.2

Output (2): LSDV Model

➔ **Insights from Stata Output:**

The F-value represents the Mean Square Model (3.8063e+09) divided by the Mean Square Residual (525402986), yielding F=7.24. The p-value that is associated with this F value is less than the significance level (0.05). Hence, it seems like the LSDV is a significant model; the independent variables show a statistically significant relationship with the dependent variable.

R-Squared indicates that 61.68 % of the variation in the number of **daily new cases** can be predicted from the variables (**Temperature, Humidity, and O3**); indicating that more variability was explained by the LSDV model compared to the simple regression model.

It is well known that fixed effects assume that differences between countries (cross-sectional units) can be accommodated from different intercepts.

Hence, the regression equation of the LSDV fixed effects model is as follows: $y_{it} = \beta_0 + \beta_1 Temp_{it} + \beta_2 Hum_{it} + \beta_3 O3_{it} + \gamma_1 D_2 + \gamma_2 D_3 + \gamma_3 D_4 + \gamma_4 D_5 + \gamma_5 D_6 + \gamma_6 D_7 + \gamma_7 D_8 + \varepsilon_{it} \quad \forall\ i = 1,2, \dots 8 \quad t = 1,2, \dots 7$

Where:

$$D_2 = \begin{cases} 1 & for\ country\ 2; China \\ 0 & O.W \end{cases}$$

$$D_3 = \begin{cases} 1 & for\ country\ 3; KSA \\ 0 & O.W \end{cases}$$

$$D_4 = \begin{cases} 1 & for\ country\ 4; UK \\ 0 & O.W \end{cases}$$

$$D_5 = \begin{cases} 1 & for\ country\ 5; Italy \\ 0 & O.W \end{cases}$$

$$D_6 = \begin{cases} 1 & for\ country\ 6; Brazil \\ 0 & O.W \end{cases}$$

$$D_7 = \begin{cases} 1 & for\ country\ 7; South\ Africa \\ 0 & O.W \end{cases}$$

$$D_8 = \begin{cases} 1 & for\ country\ 8; \text{Australia} \\ 0 & O.W \end{cases}$$

$$And\ USA\ is\ considered\ to\ be\ the\ base/reference\ category$$

And hence, according to the Stata output, the model interpretation will be as follows:

	Coefficient	p-value	Interpretation
Constant: Base Category: USA	79270.55	0.063	Since p-value= 0.063 which is insignificant for significance level (0.05) however it is significant at (0.1) significance level. Then we can say that the average number of daily new cases in the USA was 79271 cases approximately at (0.1) significance level.
ID 2: China	-75319.58	0.000	Since p-value= 0.000 which is < 0.05, then we can say that the average number of daily new cases in China is lower than that of the USA by approximately 75320 cases.
ID 3: KSA	-64821.3	0.001	Since p-value= 0.001 which is < 0.05, then we can say that the average number of daily new cases in KSA is lower than that of the USA by

			approximately 64821 cases.
ID 4: UK	-69694.69	0.000	Since p-value= 0.000 which is < 0.05, then we can say that the average number of daily new cases in the UK is lower than that of the USA by 69695 cases approximately.
ID 5: Italy	-65900.93	0.000	Since p-value= 0.000 which is < 0.05, then we can say that the average number of daily new cases in Italy is lower than that of the USA by 65901 cases approximately.
ID 6: Brazil	-46949.43	0.001	Since p-value= 0.001 which is < 0.05, then we can say that the average number of daily new cases in Brazil is lower than that of the USA by approximately 46949 cases.
ID 7: South Africa	-78166.52	0.000	Since p-value= 0.000 which is < 0.05, then we can say that the average number of daily new cases in South Africa is lower than that of the USA by approximately 78167 cases.
ID 8: Australia	-84190.74	0.000	Since p-value= 0.000 which is < 0.05, then we can say that the average number of daily new cases in Australia is lower than that of the USA by 84191case approximately.
Temperature	-738.2236	0.271	Since p-value= 0.271 which is > 0.05, then we can say that Temperature has an insignificant impact on the daily news cases.
Humidity	293.4224	0.554	Since p-value= 0. 554 which is > 0.05, then we can say that Humidity has an insignificant impact on the daily new cases.
O3	-431.9173	0.470	Since p-value= 0. 470 which is > 0.05, then we can say that O3 has an insignificant impact on the daily new cases.

Table (1): LSDV Coefficients' interpretation

The Between Estimator

The Between Estimator of the panel data means that we regress the averages of the explanatory variables of the subjects against the averages of the outcome variables of the subjects. This estimator is a cross-section regression with N data points. The between estimator uses just the cross-sectional variation in the data. The between estimator uses only information on how each individual differs from the global average, ignoring the variation over time for each individual in the sample. An interesting feature of the Between estimator is that it tends to reduce the effect of measurement errors since it uses time averages. It would be consistent with $T \to \infty$ but that is an unlikely condition in most panel data sets. Under the usual assumptions, pooled OLS using the between transformation is consistent and unbiased.

So, based on the previous, the model for our dataset can be shown as follows:

$$\overline{y}_i = \beta_1 + \sum_{j=4}^{k} \beta_j \overline{x}_{ji} + \alpha + \overline{\varepsilon}_i \qquad where\ j = 1, 2, 3, 4 \qquad i = 1, 2, 3$$

Using STATA as a package for our analysis, and mentioning the explanatory variables (Temperature, Humidity, and O3) and the response variable (Daily New Cases), the output for the previously mentioned estimator was obtained to indicate the following:

```
.
. * Between estimator

.
. xtreg DailyNewCases Temperature Humidity O3, be

Between regression (regression on group means)  Number of obs      =        56
Group variable: ID                              Number of groups   =         8

R-sq:                                           Obs per group:
     within  = 0.0154                                          min =         7
     between = 0.1210                                          avg =       7.0
     overall = 0.0063                                          max =         7

                                                F(3,4)             =      0.18
sd(u_i + avg(e_i.))=  33619.11                  Prob > F           =    0.9025

------------------------------------------------------------------------------
DailyNewCa~s |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 Temperature |   2308.193   5773.692     0.40   0.710    -13722.15    18338.53
    Humidity |   1206.566   1680.149     0.72   0.512    -3458.275    5871.407
          O3 |   212.7117   1588.651     0.13   0.900    -4198.092    4623.515
       _cons |  -109935.7   205122.4    -0.54   0.620    -679446.7    459575.4
------------------------------------------------------------------------------
```

Output (3): Between Estimator Model

➔ **Insights from Stata Output:**

- The **R-squared value** for the "between" is equal to 0.121, which refers to a percentage of 12.1% from the variance in the averages of the dependent variable that can be explained by the averages of independent variables.
- The **p-values** of the coefficients of the explanatory variables are as follows:
  *Temperature's p-value= 0.710

  *Humidity's p-value= 0.512

  *O3's p-value= 0.900

  Hence, it is concluded that the coefficients of the three explanatory variables in this model are all *insignificant*, and this reflects the fact that we won't reject the null hypothesis. Also, this brought the fact that the averages of the explanatory variables have no significant effect on the average of the dependent variable.

- The **Prob> F = 0.9025** which is a value greater than **0.05**, shows that the model is inadequate to be used.
- The **t-values** for the coefficients of the explanatory variables are lower than **1.96** hence the regressors (explanatory variables) are irrelevant to the dependent variable (Daily New Cases).
- The coefficients of the variables have insignificant p-values and hence, it can be concluded that they have no impact on the daily new cases variable.

Therefore, it is concluded that the Between Estimator isn't a suitable model to represent and fit the data well.

The Within estimator or Fixed effects estimator

The Within estimator of the vector of the parameters in a model with panel data is computed as an ordinary least squares estimator using the deviations from the time averages of the data for each cross-section unit (deviations from group means). It is equivalent to estimating a model with dummy variables for each unit (the least squares dummy variable, or LSDV, model). The usual consistent estimator is most commonly called the within groups estimator.

<u>There are several ways to obtain this estimator:</u>

1. OLS on a model with dummy variables for each of the n fixed effects.
2. OLS after a differencing transformation that eliminates α_i: **OLS of** $(y_{it} - \bar{y}_i)$ **on** $(X_{it} - \bar{y}_i)$
3. GLS after this same differencing transformation.

<u>The model that represents our data can be stated as follows:</u>

$$y_{it} - \overline{y}_i = \left(X_{it} - \overline{X}_i\right)\beta + (\alpha_i - \overline{\alpha}_i) + (u_{it} - \overline{u}_i) \implies \ddot{y}_{it} = \ddot{X}_{it}\beta + \ddot{u}_{it}$$

where $\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\overline{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$, and $\overline{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_{it}$.

#

$i = 1, 2,...., 8$ $\qquad$ $t = 1, 2,..... 7$

**** The terms within parentheses are to emphasize that they are unobserved.***

The transformation of the original equation, known as the within transformation, has eliminated α_i from the equation. So, we can estimate β consistently by using OLS on $\Delta y_{it} = \Delta x_{it}\beta + \Delta u_{it}$ and this is called the within estimator or the Fixed Effects estimator.

By Using STATA to obtain this estimator for our panel data, the following was noticed taking into consideration that the explanatory variables are (Temperature, Humidity, and O3) and the dependent one is the (Daily New Cases):

```
.
. * Fixed effects or within estimator

.
. xtreg DailyNewCases Temperature Humidity O3, fe

Fixed-effects (within) regression               Number of obs     =         56
Group variable: ID                              Number of groups  =          8

R-sq:                                           Obs per group:
     within  = 0.0802                                         min =          7
     between = 0.0236                                         avg =        7.0
     overall = 0.0446                                         max =          7

                                                F(3,45)           =       1.31
corr(u_i, Xb)  = -0.0701                        Prob > F          =     0.2835

------------------------------------------------------------------------------
DailyNewCa~s |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 Temperature |  -738.2236   662.6176    -1.11   0.271    -2072.804    596.3568
    Humidity |   293.4224   491.7129     0.60   0.554    -696.9382    1283.783
          O3 |  -431.9173   593.2894    -0.73   0.470    -1626.863    763.0289
       _cons |   18640.15   39796.44     0.47   0.642       -61514    98794.31
-------------+----------------------------------------------------------------
     sigma_u |  26910.155
     sigma_e |  22921.671
         rho |  .57952942   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(7, 45) = 9.31                       Prob > F = 0.0000
```

Output (4): The Within Estimator Model

➔ **Insights from Stata Output:**

- **The R-squared value** for the within estimator equals 0.0802 which is approximately a percentage of 8%, and this somehow suggests that the goodness of fit for the model is not that good. Hence, the panel data doesn't fit this within the estimator model where the proportion of variance in the dependent variable that can be explained by the independent variables is relatively low.
- **The p-values** for all the coefficients of the explanatory variables are greater than 0.05 which is considered an indication of the insignificancy of the coefficients; hence we won't reject the null hypothesis. So, the explanatory variables won't have a significant influence on the dependent variable.
- **The Prob> F** = 0.2835 which is greater than 0.05 where which is a strong indication of the inadequacy of using this model.
- **The corr (u_i, Xb) =** -0.0701 refers to the correlation between the errors and the regressors in the fixed effect model with a negative relationship approximately equal to -0.07.
- **'Rho'** or the intraclass correlation indicates that about 58% of the variance in the model is due to differences across panels.

- **The t-values** for the coefficients of the explanatory variables are lower than 1.96, and in return show the less relevance of the explanatory variables (temperature, Humidity, and O3) to the dependent variable (Daily New Cases).
- The within-estimator shows that the coefficients of the explanatory variables are with insignificant p-values and we can conclude that the p-values yield to the insignificancy of those coefficients.

It can be concluded from the above results that the within estimator model is neither suitable nor efficient to be used for our data.

First Differences estimator model

The First-Differenced (FD) estimator is an approach that is used to address the problem of omitted variables in econometrics and statistics by using panel data. The estimator is obtained by running a pooled OLS estimation for a regression of the differenced variables. To eliminate the fixed effects in this model, you use first-differenced methods to difference them out instead of using the within the transformation. Because the intercept is differenced out, the intercept cannot be estimated by first-differenced methods. First-differencing is the easiest way of dealing with fixed effects.

<u>The First-Differenced Estimator can be represented as follows to suit our panel data:</u>

$$y_{it} = x_{it}\beta + (\alpha_i + u_{it})$$

$$y_{it} - y_{i,t-1} = (x_{it} - x_{i,t-1})\beta + (\alpha_i - \alpha_i + u_{it} - u_{i,t-1}), \Delta y_{it} = \Delta x_{it}\beta + \Delta u_{it}$$

$$i = 1, 2, \ldots, 8 \qquad t = 1, 2, \ldots, 7, \qquad where\ \alpha_i,\ u_{it}\ are\ unobserved$$

This removes the individual fixed effect, so we can obtain consistent estimates

of β by estimating the equation in first differences by OLS.

The first-differenced estimator requires strict exogeneity as the last equation contains the residuals u_{it} and $u_{i,t-1}$, whereas the vector of transformed explanatory variable contains x_{it} and $x_{i,t-1}$. Hence, we need $E(x_{it}\, u_{is}) = 0$ for $s = t,\ t - 1$.

Fixed effects and First Differences are two alternative ways of removing the fixed effect.

Then, which method should we use?

- When T= 2, FE and FD are exactly equivalent and so in this case it does not matter which one we use.
- But when T is greater than or equal to 3, FE and FD are not the same. Under the null hypothesis that the model is correctly specified, FE and FD will differ only because of sampling error whenever T is greater than or equal to 3. Hence, if FE and FD are significantly different - so that the differences in the estimates cannot be attributed to sampling error - we should worry about the validity of the strict exogeneity assumption. (That is our case).

<u>The following STATA output is for the first differences estimator, and it can be explained as follows with (Temperature, Humidity, and O3) as explanatory variables and (Daily New Cases) as a dependent variable:</u>

```
.
. * First-differences estimator

.
. reg D.(DailyNewCases Temperature Humidity O3), noconstant
```

Source	SS	df	MS			
				Number of obs	=	48
				F(3, 45)	=	0.40
Model	1.2756e+09	3	425194296	Prob > F	=	0.7570
Residual	4.8413e+10	45	1.0759e+09	R-squared	=	0.0257
				Adj R-squared	=	-0.0393
Total	4.9689e+10	48	1.0352e+09	Root MSE	=	32800

D. DailyNewCa~s	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Temperature D1.	-386.1261	634.062	-0.61	0.546	-1663.192	890.9402
Humidity D1.	46.57949	461.2532	0.10	0.920	-882.4322	975.5912
O3 D1.	-359.6607	597.8097	-0.60	0.550	-1563.711	844.3898

Output (5): First Difference Estimator Model

➔ **Insights from Stata Output:**

- The sub-table is in the upper left section of the readout. This sub-table is called the **ANOVA**. The Root MSE is essentially the standard deviation of the residual in this model equals 32800.
- **The model sum of squares** has a moderate residual sum of squares equals 4.8413e+10.

- **The R-squared value** is 0.0257 which is approximately 2% and this returns a result of not having a good fit for the data where the proportion of variance in the dependent variable that can be explained by the independent variables is relatively low.
- **The p-values** for all the coefficients of the explanatory variables are greater than 0.05 which reflects the insignificancy of those coefficients. So, the explanatory variables won't have a significant influence on the dependent variable.
- **The Prob> F** = 0.757 which is greater than 0.05 where this is a strong indication of the inadequacy of using this model.
- **The t-values** for the coefficients of the explanatory variables are lower than 1.96, and in return show the less relevance of the explanatory variables (temperature, Humidity, and O3) to the dependent variable (Daily New Cases) in this model.
- **The intercept** is considered as the time trend in the first-differenced estimator.
- The first-differenced coefficients have no significant impact on the daily new cases.

We can get from the above results that the first- differenced estimator model is neither suitable nor efficient to be used for our data.

The Random-Effects estimator Model

In the random-effects model, the individual-specific effect is a random variable that is uncorrelated to the explanatory variables. The random-effects model can be consistently estimated by both the RE estimator and the FE estimator. We would prefer the RE estimator if we can be sure that the individual-specific effect is unrelated. This is usually tested by a (Durbin-Wu-) Hausman test. However, the Hausman test is only valid under homoscedasticity and cannot include time-fixed effects. The Random Effects Model allows for consistent and efficient estimates of β and allows you to identify the effect of the gun control laws by exploiting the variation of these laws across states. This model will estimate panel data where interference variables may be interconnected between time and between individuals. In the Random Effect model, the difference between intercepts is accommodated by the error terms of each company. The advantage of using the Random Effect model is to eliminate heteroscedasticity. This model is also called the Error Component Model (ECM) or Generalized Least Square (GLS) technique. In principle, the random effect model is different from the common effect and fixed effect, especially since this model does not use the principle of ordinary least squares, but instead uses the principle of maximum likelihood or general least square.

Random effects assume that:

- The entity's error term is not correlated with the predictors which allows for time-invariant variables to play a role as explanatory variables.
- In random effects, you need to specify those individual characteristics that may or may not influence the predictor variables. The problem with this is that some variables may not be available therefore leading to omitted variable bias in the model.

- Random effects allow generalizing of the inferences beyond the sample used in the model.

The Random Effects Estimator Model for our panel dataset is as follows:

$$y_{it} = \alpha + \beta' X_{it} + u_i + \varepsilon_{it} \qquad where\ i = 1, 2, ..., 8 \qquad t = 1, 2, ..., 7$$

ε_{it}: The residual as a whole where the residual is a combination of cross section and time series. (**Within-entity error**)

u_i: The individual residual which is the random characteristic of unit observation the i-thand remains at all times. (**Between-entity error**)

By using STATA, the next output showed the results for the Random Effects Estimator model was stated with (Temperature, Humidity, and O3) as explanatory variables and (Daily New Cases) as a dependent variable:

```
. * Random effects estimator

.
. *Default, GLS Estimator

.
. xtreg DailyNewCases Temperature Humidity O3, re

Random-effects GLS regression                   Number of obs     =         56
Group variable: ID                              Number of groups  =          8

R-sq:                                           Obs per group:
     within  = 0.0798                                         min =          7
     between = 0.0313                                         avg =        7.0
     overall = 0.0497                                         max =          7

                                                Wald chi2(3)      =       4.22
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.2384

------------------------------------------------------------------------------
DailyNewCa~s |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 Temperature |  -739.9227   634.3918    -1.17   0.243    -1983.308    503.4624
    Humidity |   343.4333   440.8411     0.78   0.436    -520.5994    1207.466
          O3 |  -352.2841   538.2306    -0.65   0.513    -1407.197    702.6285
       _cons |   13688.32   37838.25     0.36   0.718    -60473.28    87849.92
-------------+----------------------------------------------------------------
     sigma_u |  32483.646
     sigma_e |  22921.671
         rho |  .66759065   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Output (6): Random Effect Model

➔ **Insights from Stata Output:**

- It's observed from the above output that **the corr (u_i, X) = 0** which means that the differences across units (error terms) are uncorrelated with the regressors, as if u_i is correlated with any of the predictors, we will run into a problem of endogeneity where the regression coefficients will be potentially biased and inconsistent and this is considered the major random effect assumption.

- **The Wald chi2 (3) = 4.22** which is greater than 0.05, and reflects the fact that the coefficients for the explanatory variables are not different from zero.
- The **p-values** are higher than 0.05 indicating the insignificancy of the coefficients of the explanatory variables, and that the coefficients have no explicit influence on the dependent variable (Daily New Cases).
- **'Rho'** or the intraclass correlation indicates that about 67% of the variance in the model is due to differences across panels.
- **The R-squared value** for the within estimator equals 0.0497 which is approximately a percentage of 5%, and then we should consider this model to be of low goodness of fit to model the previously mentioned panel datasets.

Hausman Test

The Hausman test is sometimes described as a test for model misspecification. In panel data analysis, the Hausman test can help in choosing between a fixed-effects model and a random-effects model.

Where the null hypothesis is that the preferred model is random effects and the alternative hypothesis is that the model has fixed effects.

$$H_0: REM$$

$$H_1: FEM$$

```
. hausman fixed random

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed        random       Difference          S.E.
-------------+----------------------------------------------------------------
 Temperature |   -738.2236    -739.9227        1.699039        191.3351
    Humidity |    293.4224     343.4333       -50.0109         217.8089
          O3 |   -431.9173    -352.2841       -79.63312        249.5998
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        0.16
                Prob>chi2 =      0.9835
```

Output (7): Hausman Test

➔ Insights from Stata Output:

As it appears, the p-value (0.9835) is greater than the significance level (0.05) meaning that we don't have enough evidence to reject the null hypothesis; meaning that the suitable model is the Random Effect Model.

Breusch-Pagan Lagrange multiplier (LM)

The LM test helps in deciding between a random-effects regression and a simple OLS regression. The null hypothesis in the LM test is that variance across countries is zero. This is no significant difference across countries (i.e. no panel effect).

```
Breusch and Pagan Lagrangian multiplier test for random effects

        DailyNewCases[ID,t] = Xb + u[ID] + e[ID,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                DailyNe~s|   1.12e+09       33495.22
                        e|   5.25e+08       22921.67
                        u|   1.06e+09       32483.65

        Test:   Var(u) = 0
                             chibar2(01) =    44.04
                          Prob > chibar2 =   0.0000
.
```

Output (8): LM Test

➔ Insights from Stata Output:

As it appears, the p-value (0.0000) is less than the significance level (0.05) meaning that we do have enough evidence to reject the null hypothesis; meaning that the suitable model is the Random Effect Model (there is a significant difference across countries).

Chapter Five

Main Findings and Conclusion

Main Findings of the Analysis

Throughout the analysis, we found that heterogeneity does exist in our data among and within countries and seasons. USA and Brazil had the main leading rule in the average number of the daily new cases of Covid-19, however, China had the lowest average number of the daily new cases although the starting point for the outbreak of COVID-19 was China however China managed to control and limit Covid-19 break-out. And winter 2020 was the most fruitful season with the average number of daily new cases over the whole studied period. Due to the existing heterogeneity Simple regression model was not a preferable/ suitable model since it neglected the fact of heterogeneity existence. According to what we have found we can say that the average number of daily new cases changes heavily from one country to another. And our main result was that climate change represented by; "Temperature", "Humidity" and "O3" had no impact on the daily new cases of Covid-19. What people were thinking of is that as the weather gets colder the number of infected cases by Covid-19 will increase; however, we were able to prove the opposite. The quote that says "Don't Judge a Book by Its Cover" can be applied here, we can't judge and make a conclusion on two large different phenomena by looking on the micro-level – the number of Covid-19 new cases day by day- instead we have to look and analyze on the macro level.

Conclusion

According to the United Nations Climate Change definition, Climate change refers to long-term shifts in temperatures and weather patterns. These shifts may be natural, such as through variations in the solar cycle. But since the 1800s, human activities have been the main driver of climate change, primarily due to burning fossil fuels like coal, oil, and gas. In this paper, we were interested in studying if Climate Change can affect Covid-19 spread-out as it was assumed. We have made a focus on some of the G-20 countries and the choice of the countries resulted from many different reasons as mentioned before. By modeling and studying different eight countries' climate and Covid-19 cases daily, we found that what we were seeing as a fact isn't a real fact; meaning that climate change has no real impact on the average number of the daily new cases of Covid-19. The heterogeneity among countries regarding the Covid-19 cases may have resulted from other reasons aside from climate change; for example, The precautionary measures that countries have followed, the awareness of the people, or even adherence to the guidelines of the World Health Organization… etc.

References

1. Addas, A. and Maghrabi, A., (2021). The Impact of COVID-19 Lockdowns on Air Quality—A Global Review. *Sustainability*, *13*(18), p.10212.
2. Ahmad, D., (2020). *Panel Data Analysis using STATA*. [online] Youtube. Available at: <https://www.youtube.com/watch?v=POyR9TYNTUA&t=2625s> [Accessed 27 March 2022].
3. Allen, J., (2004). *Tango in Atmosphere: Ozone and Climate Change*. [online] NASA. Available at: <https://www.giss.nasa.gov/research/features/200402_tango/#:~:text=As%20energy%20demand%20and%20production,potential%20for%20greater%20ozone%20formation> [Accessed 1 June 2022].
4. Burrows, L., (2016). *The complex relationship between heat and ozone*. [online] Harvard Gazette. Available at: <https://news.harvard.edu/gazette/story/2016/04/the-complex-relationship-between-heat-and-ozone/> [Accessed 1 June 2022].
5. C-CHANGE | Harvard T.H. Chan School of Public Health. (n.d). *Coronavirus and Climate Change.* [online] Available at: <https://www.hsph.harvard.edu/c-change/subtopics/coronavirus-and-climate-change/> [Accessed 3 March 2022].
6. Damodar, N., (2009). Basic Econometrics Fifth Edition. McGraw-Hill.
7. Dang, H.A. and Trinh, T.A., (2020). Does the COVID-19 pandemic improve global air quality? New cross-national evidence on its unintended consequences.
8. G20 Foundation. (n.d). *What is the G20 | G20 Foundation*. [Online] Available at: <https://www.g20foundation.org/g20/what-is-the-g20.> [Accessed 2 March 2022].
9. Hedges, L. and Salkind, N., (2022). *Random-Effects Models*. [online] SAGA. Available at: <https://methods.sagepub.com/reference/encyc-of-research-design/n360.xml#:~:text=Random%2Deffects%20models%20are%20statistical> [Accessed 8 June 2022].
10. Hermoso, J.C.M., Sadang, J.M.M., Evangelista, N.C.B. and Rosete, M.A.L., (2021). The Effects of Infrastructure Investments on Services and Wholesale/Retail Productivity: A Panel-Data Evidence from the Philippines. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, *6*(12), pp.233-246.

11. Kaggle. (2022). *Covid-19 Global Dataset*. [Online] Available at: <https://www.kaggle.com/josephassaker/covid19-global-dataset> [Accessed 28 February 2022].
12. Lambert, B., (2013). *Least Squares Dummy Variables estimators*. [online] Youtube. Available at: <https://www.youtube.com/watch?v=i7vYh1kCEOY&t=161s> [Accessed 12 June 2022].
13. Masum, M.H. and Pal, S.K., (2020). Statistical evaluation of selected air quality parameters influenced by COVID-19 lockdown. *Global Journal of Environmental Science and Management*, *6*(Special Issue (Covid-19)), pp.85-94.
14. Met Office. (2021). *How did COVID-19 lockdowns affect the climate?*. [online] Available at: <https://www.metoffice.gov.uk/research/news/2021/how-did-covid-19-lockdowns-affect-the-climate> [Accessed 4 March 2022].
15. National Centers for Environmental Information (NCEI). (2022). *Meteorological Versus Astronomical Seasons*. [online] Available at: <https://www.ncei.noaa.gov/news/meteorological-versus-astronomical-seasons> [Accessed 14 May 2022].
16. Project, T., (2022). *COVID-19 Worldwide Air Quality data*. [online] aqicn.org. Available at: <https://aqicn.org/data-platform/covid19/verify/86165cf5-f151-4abf-baf2-20632073107f> [Accessed 1 March 2022].
17. Rönkkö, M., (2019). *Random effects assumption and its tests*. [online] Youtube. Available at: <https://www.youtube.com/watch?v=bbQ8amG6LMY&t=107s> [Accessed 16 May 2022].
18. Sarmadi, M., Rahimi, S., Rezaei, M., Sanaei, D. and Dianatinasab, M., (2021). Air quality index variation before and after the onset of COVID-19 pandemic: a comprehensive study on 87 capital, industrial and polluted cities of the world. *Environmental Sciences Europe*, *33*(1), pp.1-17.
19. Schmidheiny, K.,(n.d). *Panel Data: Fixed and Random Effects*. [ebook] Kurt Schmidheiny. Available at: <https://www.schmidheiny.name/teaching/panel2up.pdf> [Accessed 10 June 2022].
20. Sil, A. and Kumar, V.N., (2020). Does weather affect the growth rate of COVID-19, a study to comprehend transmission dynamics on human health. *Journal of Safety Science and Resilience*, *1*(1), pp.3-11.
21. Torkmahalleh, M.A., Akhmetvaliyeva, Z., Darvishi Omran, A., Darvish Omran, F., Kazemitabar, M., Naseri, M., Naseri, M., Sharifi, H., Malekipirbazari, M., Kwasi Adotey, E. and Gorjinezhad, S., (2021). Global air quality and covid-19 pandemic: Do we breathe cleaner air?.
22. Torres-Reyna, O., (2007). Panel data analysis fixed and random effects using Stata (v. 4.2). *Data & Statistical Services, Priceton University*, *112*, p.49.

23. UCLA. (2022). *Introduction to the features of SAS | SAS Learning Modules*. [online] Available at: <https://stats.oarc.ucla.edu/sas/modules/introduction-to-the-features-of-sas/> [Accessed 16 May 2022].
24. United Nations. (2022). *What Is Climate Change? | United Nations*. [online] Available at: <https://www.un.org/en/climatechange/what-is-climate-change> [Accessed 12 April 2022].
25. US EPA. (2022). *Ground-level Ozone Basics*. [online] Available at: <https://www.epa.gov/ground-level-ozone-pollution/ground-level-ozone-basics> [Accessed 3 June 2022].
26. USGLC. (2021). *COVID-19 Brief: Impact on Climate*. [online] Available at: <https://www.usglc.org/coronavirus/climate/> [Accessed 5 March 2022].
27. Wang, M., Jiang, A., Gong, L., Lu, L., Guo, W., Li, C., Zheng, J., Li, C., Yang, B., Zeng, J. and Chen, Y., (2020). Temperature significantly change COVID-19 transmission in 429 cities. *medrxiv*.
28. WHO. (2020). *Coronavirus disease (COVID-19): Climate change*. [online] Available at: <https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-climate-change> [Accessed 2 March 2022].
29. Youtube. (2017). *Lecture 7 Panel Data Models (Part I)*. [online] Available at: <https://www.youtube.com/watch?v=RQOos3YlyMU&t=2400s> [Accessed 2 April 2022].
30. Youtube. (2017). *Lecture 8 Panel Data Models (Part II)*. [online] Available at: <https://www.youtube.com/watch?v=Et5CRL51yEo&t=3229s> [Accessed 5 April 2022].
31. Zulfikar, R. and STp, M.M., (2019). Estimation model and selection method of panel data regression: an overview of common effect, fixed effect, and random effect model. *INA-Rxiv 9qe2b, Center for Open Science*.

Appendix

Brazil Minitab Output

Descriptive Statistics: Temperature

Variable	N	N*	Mean	Variance	Minimum	Maximum	Range
Temperature	580	0	19.271	10.461	8.000	27.000	19.000

Descriptive Statistics: Humidity

Variable	N	N*	Mean	Variance	Minimum	Maximum	Range
Humidity	580	0	78.328	69.973	42.000	94.000	52.000

Descriptive Statistics: O3

Variable	N	N*	Mean	Variance	Minimum	Maximum	Range
O3	580	0	14.582	18.468	4.900	32.900	28.000

Correlations: Temperature, Humidity, O3, Daily New Cases

	Temperature	Humidity	O3
Humidity	-0.268		
	0.000		
O3	0.234	-0.184	
	0.000	0.000	
Daily New Cases	-0.002	0.106	-0.222
	0.953	0.011	0.000

Cell Contents: Pearson correlation
P-Value

Covariances: Temperature, Humidity, O3, Daily New Cases

	Temperature	Humidity	O3
Temperature	10		
Humidity	-7	70	
O3	3	-7	18
Daily New Cases	-196	21700	-23299

	Daily New Cases
Daily New Cases	598521300

Saudi Arabia Minitab Output

Descriptive Statistics: Temperature

Variable	N	N*	Mean	Variance	Minimum	Maximum	Range
Temperature	580	0	22.231	55.347	5.000	39.000	34.000

Descriptive Statistics: Humidity

Variable	N	N*	Mean	Variance	Minimum	Maximum	Range
Humidity	580	0	43.564	328.921	7.000	93.000	86.000

Descriptive Statistics: O3

Variable	N	N*	Mean	Variance	Minimum	Maximum	Range
O3	580	0	22.427	140.323	1.000	67.000	66.000

Correlations: Temperature, Humidity, O3, Daily New Cases

	Temperature	Humidity	O3
Humidity	-0.328 0.000		
O3	-0.188 0.000	0.245 0.000	
Daily New Cases	-0.018 0.660	0.271 0.000	0.078 0.062

Cell Contents: Pearson correlation
P-Value

Covariances: Temperature, Humidity, O3, Daily New Cases

	Temperature	Humidity	O3
Temperature	55.35		
Humidity	-44.22	328.92	
O3	-16.60	52.72	140.32
Daily New Cases	-130.32	4691.45	879.95

	Daily New Cases
Daily New Cases	914235.82

China Minitab Output

Descriptive Statistics: Temperature, Humidity, O3, Daily_new_cases

Variable	N	N*	Mean	Variance	Minimum	Maximum	Range
Temperature	580	0	18.427	64.363	-4.150	30.000	34.150
Humidity	580	0	69.130	101.586	38.000	86.700	48.700
O3	580	0	25.277	47.656	5.300	42.100	36.800
Daily_new_cases	580	0	28.17	978.64	0.00	325.00	325.00

Correlations: Temperature, Humidity, O3, Daily_new_cases

	Temperature	Humidity	O3
Humidity	0.446		
O3	0.650	-0.050	
Daily_new_cases	-0.074	-0.004	-0.102

Cell Contents: Pearson correlation

Covariances: Temperature, Humidity, O3, Daily_new_cases

	Temperature	Humidity	O3
Temperature	64.3627		
Humidity	36.0311	101.5855	
O3	36.0185	-3.4496	47.6556
Daily_new_cases	-18.5966	-1.4017	-22.0068

	Daily_new_cases
Daily_new_cases	978.6450

Italy Minitab Output

Descriptive Statistics: Temperature, Humidity, O3, Daily new cases

Variable	N	N*	Mean	Variance	Minimum	Maximum	Range
Temperature	580	0	16.606	52.979	-0.100	29.700	29.800
Humidity	580	0	67.281	193.724	30.000	93.250	63.250
O3	580	0	32.191	234.300	3.300	78.000	74.700
Daily new cases	580	0	8060	78382881	113	41198	41085

Covariances: Temperature, Humidity, O3, Daily new cases

	Temperature	Humidity	O3
Temperature	53		
Humidity	-44	194	
O3	92	-154	234
Daily new cases	-40907	47532	-83969

	Daily new cases
Daily new cases	78382881

Correlations: Temperature, Humidity, O3, Daily new cases

	Temperature	Humidity	O3
Humidity	-0.433		
O3	0.822	-0.721	
Daily new cases	-0.635	0.386	-0.620

Cell Contents: Pearson correlation

UK Minitab Output

Correlations: Daily new cases, Tempreature, Humidity, O3

	Daily new cases	Tempreature	Humidity
Tempreature	-0.078		
Humidity	0.390	-0.255	
O3	-0.385	0.128	-0.644

Cell Contents: Pearson correlation

Covariances: Daily new cases, Tempreature, Humidity, O3

	Daily new cases	Tempreature	Humidity
Daily new cases	213249555		
Tempreature	-5780	26	
Humidity	52251	-12	84
O3	-34740	4	-36

	O3
O3	38

Descriptive Statistics: Daily new cases, Tempreature, Humidity, O3

Variable	N	N*	Mean	Variance	Minimum	Maximum	Range
Daily new cases	580	0	13521	213249555	3	67794	67791
Tempreature	580	0	11.483	25.845	-1.100	23.900	25.000
Humidity	580	0	78.750	84.041	48.900	95.400	46.500
O3	580	0	21.189	38.086	2.500	46.000	43.500

South Africa Minitab Output

Descriptive Statistics: Daily_new_cases, Tempreature, Humidity, O3

Variable	N	N*	Mean	Variance	Minimum	Maximum	Range
Daily_new_cases	575	5	5082	28976998	0	26645	26645
Tempreature	580	0	16.565	17.023	6.200	24.600	18.400
Humidity	580	0	66.181	135.866	33.150	90.750	57.600
O3	580	0	7.3952	4.6788	1.9000	15.6000	13.7000

Correlations: Daily_new_cases, Tempreature, Humidity, O3

	Daily_new_cases	Tempreature	Humidity
Tempreature	-0.262		
Humidity	-0.191	0.360	
O3	-0.049	0.393	0.019

Cell Contents: Pearson correlation

Covariances: Daily_new_cases, Tempreature, Humidity, O3

	Daily_new_cases	Tempreature	Humidity
Daily_new_cases	28976998		
Tempreature	-5812	17	
Humidity	-11990	17	136
O3	-565	4	0

	O3
O3	5

Australia Minitab Output

Descriptive Statistics: Temperature, Humidity, O3, Daily New Cases

Variable	N	N*	Mean	Variance	Minimum	Maximum	Range
Temperature	580	0	16.150	13.156	9.400	25.000	15.600
Humidity	580	0	73.554	48.835	49.200	91.000	41.800
O3	580	0	10.107	7.806	1.600	17.700	16.100
Daily New Cases	580	0	184.7	170711.1	1.0	2400.0	2399.0

Correlations: Temperature, Humidity, O3, Daily New Cases

	Temperature	Humidity	O3
Humidity	-0.244		
O3	0.067	-0.466	
Daily New Cases	-0.225	-0.105	0.266

Cell Contents: Pearson correlation

Covariances: Temperature, Humidity, O3, Daily New Cases

	Temperature	Humidity	O3
Temperature	13.16		
Humidity	-6.19	48.84	
O3	0.68	-9.09	7.81
Daily New Cases	-337.83	-302.86	306.93

	Daily New Cases
Daily New Cases	170711.13

USA Minitab Output

Descriptive Statistics: Temperature, Humidity, O3, Daily New Cases

Variable	N	N*	Mean	Variance	Minimum	Maximum	Range
Temperature	580	0	16.725	55.297	-2.200	28.000	30.200
Humidity	580	0	69.021	60.984	43.450	85.350	41.900
O3	580	0	21.299	22.307	8.600	33.600	25.000
Daily New Cases	580	0	76720	4089860492	7	306271	306264

Correlations: Temperature, Humidity, O3, Daily New Cases

	Temperature	Humidity	O3
Humidity	0.028		
O3	0.281	-0.588	
Daily New Cases	-0.402	0.264	-0.561

Cell Contents: Pearson correlation

Covariances: Temperature, Humidity, O3, Daily New Cases

	Temperature	Humidity	O3
Temperature	55		
Humidity	2	61	
O3	10	-22	22
Daily New Cases	-191035	131693	-169420

	Daily New Cases
Daily New Cases	4089860492

Setting up the Data R codes

```r
setwd("E:/ghofran")

# importing your dataset
t<- read.csv("tem.csv")
h<- read.csv("hum.csv")
o<- read.csv("o3.csv")
c<- read.csv("covid.csv")

library(dplyr)

# create a new modified dataset & then merge
tem<- t %>% group_by(Time) %>% summarise(Temperature= median(Temperature))
hum<- h %>% group_by(Time) %>% summarise(Humidity= median(Humidity))
o3<- o %>% group_by(Time) %>% summarise(O3= median(O3))

#merge
climate_data0 <- merge(tem, hum, by="Time")

climate_data <- merge(climate_data0, o3, by="Time")
data<- merge(climate_data, c, by="Time")


#export your data
?write.table
install.packages("writexl")
install.packages("writecsv")

write.table(data, file = "Final Data R.CSV", sep = ",")
```

Checking for the Outliers R codes

```r
boxplot(df$Temperature,
        ylab = "TEMP")

boxplot.stats(df$Temperature)$out
out <- boxplot.stats(df$Temperature)$out
out_ind <- which(df$Temperature %in% c(out))
out_ind
```

Panel Data Analysis STATA commands

```
bysort ID: egen y_mean=mean(DailyNewCases)
twoway scatter DailyNewCases ID, mlabel(Time) || connected y_mean ID, msymbol(diamond) || ,xlabel(1 "USA" 2 "China" 3 "KSA" 4 "UK" 5 "Italy" 6 "Brazil" 7 "South Af

bysort Season: egen y_mean1=mean(DailyNewCases)
twoway scatter DailyNewCases Season, mlabel(Country)|| connected y_mean1 Season, msymbol(diamond) || ,xlabel(1 "Autumn20" 2 "Autumn21" 3 "Spring20" 4 "Spring21" 5 '
```

```
* Set data as panel data
xtset ID Season

* Pooled OLS estimator
reg DailyNewCases Temperature Humidity O3
estat hettest

*Fixed Effect MODELS
* LSDV Estimator, Fixed Effect Model
reg DailyNewCases Temperature Humidity O3 i.ID

* Between estimator
xtreg DailyNewCases Temperature Humidity O3, be

* Fixed effects or within estimator
xtreg DailyNewCases Temperature Humidity O3, fe

* First-differences estimator
reg D.(DailyNewCases Temperature Humidity O3), noconstant

* Random effects estimator
*Default, GLS Estimator
xtreg DailyNewCases Temperature Humidity O3, re

* Hausman test for fixed versus random effects model
xtreg DailyNewCases Temperature Humidity O3, fe
estimates store fixed
xtreg DailyNewCases Temperature Humidity O3, re
estimates store random
hausman fixed random

* Breusch-Pagan LM test for random effects versus OLS
xtreg DailyNewCases Temperature Humidity O3, re
xttest0
```