Logo

Description automatically generated

22CSAI01I

Introduction to Data Science

# **Final Phase**

**Part 1 Contribution Table:**

|  |  |  |
| --- | --- | --- |
| **ID** | **Name** | **All Of Us Equally Contributed** |
| **211392** | **Ahmed Sameh** | **Question 4 Question 6, and Question 5** |
| **211896** | **Abdulrahman Ayman** | **Question 1, Question 6, and Question 5** |
| **212071** | **Sara Anabtawi** | **Question 3, Question 6, and Question 5** |
| **206879** | **Ezz Yasser** | **Question 2 and Question 5** |

**Part 2 Contribution Table:**

|  |  |  |
| --- | --- | --- |
| **ID** | **Name** | **All Of Us Equally Contributed** |
| **211392** | **Ahmed Sameh** | **Question 4** |
| **211896** | **Abdulrahman Ayman** | **Question 1** |
| **212071** | **Sara Anabtawi** | **Question 2 and Question 3** |
| **206879** | **Ezz Yasser** | **Question 5 and Question 6** |

**CORONA-VIRUS**

**Question list:**

**Q1) Does the season affect the number of cases?**

(\*hint: State whether the seasons affect the spreading of the virus by tracking the change of number of cases depending on each season. The seasons will be predicted through the given month of each reported case.

Dataset used: **(Owid) Covid dataset.**

**Chart, bar chart

Description automatically generated**

**Q1 Analysis:**

According to our dataset, the seasons affect the number of cases where the number of cases reported in the winter is the highest and the lowest in fall.

**Q2) Where did the corona originate using the first case that was recorded according to the dataset?**

(\*hint: Determine where the corona virus originated through comparing the date of each case as well as its location to find the first time a case was reported and state where it was.)

Dataset used: **(Owid) Covid dataset.**

Graphical user interface, text, application, email

Description automatically generated

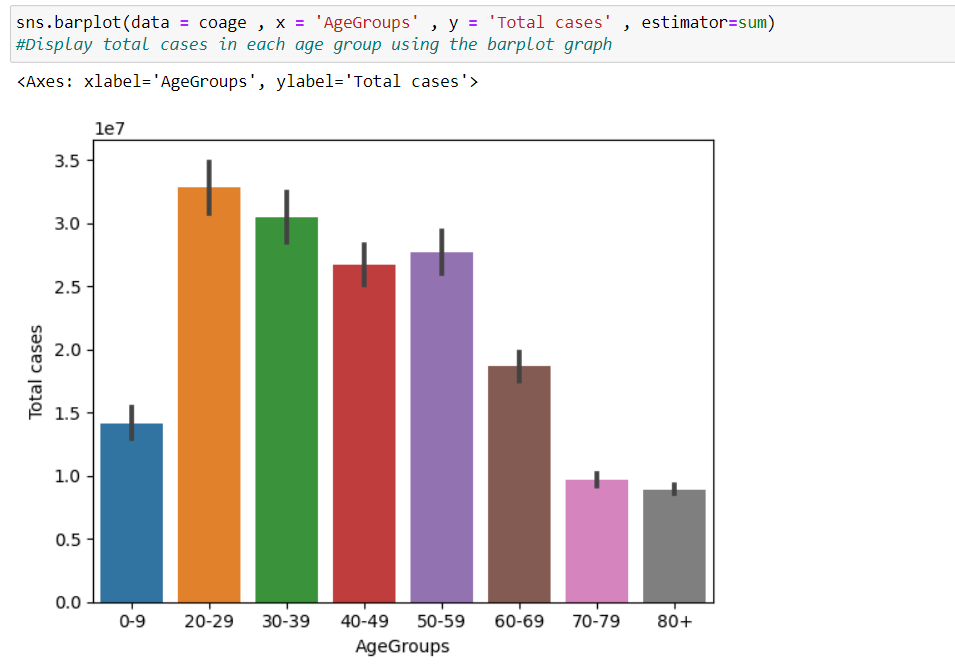
**Q2 Analysis:**

Based on our dataset, the first reported case of covid-19 was in South Korea.

**Q3) which age range has the most deaths? Which age range has the most number of cases?**

(\*hint: compare the age ranges by their total number of cases and deaths to find the age range with most death and the age range with the highest number of cases.)

Dataset used: **Age dataset.**



**Q3 part1 Analysis:**

According to our dataset, the age range with the highest number of total cases is from 20-29

Chart, histogram

Description automatically generated

**Q3 part2 Analysis:**

According to our dataset, the age range with the highest number of total deaths is 80+.

**Q4) what is the number of deaths of smokers against normal people(non-smokers) in each continent where both were affected by the coronavirus according to the dataset?**

(\*hint: by comparing the death rates of the smokers and the normal people.)

Dataset used: **(Owid) Covid dataset.**

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

**Q4 Analysis:**

According to our dataset, The Continent that has the highest number of the death of normal people(non-smokers) reaching around (50000) and the total death number of smokers reaching around (17500) is located in South America .

**Q5) which continent has the highest sudden increase in the number of cases?**

(\*hints: By comparing the continents based on the number of cases each day and monitoring a sudden increase in the cases)

Dataset used: **(Owid) Covid dataset.**

Chart

Description automatically generated

**Q5 Analysis:**

According to our dataset, the continent that has the highest sudden increase in the total number of cases is South America.

**Q6) is the number of cases in the countries who took early precautions any different from the rest of the countries? If so, how?**

(\*hint: Compare the dates between the countries to figure out which country took early precautions in order to find out whether the total cases of these countries is anyhow different from the rest of the countries.)

Dataset used: **(Owid) Covid dataset, Containment measure dataset**

**A picture containing graphical user interface

Description automatically generated**

**Q6 Analysis:**

According to our dataset, The First Country that took early containment and mitigation measures against covid is Macau. The Second Country that took early containment and mitigation measures against covid is China.

The Third Country that took early containment and mitigation measures against covid is Trinidad and Tobago.

The Last Country that took early containment and mitigation measures against covid is US.

The Second Last Country that took early containment and mitigation measures against covid is Brazil.

The Third Last Country that took early containment and mitigation measures against covid is Czechia.

As you can see below in the figure, the countries that took early precautions has less number of cases than the countries who were late in taking measures against the covid.

**Chart

Description automatically generated**

Dataset Sizes & Descriptions:

**1) (Owid) Covid dataset**-

Dataset Size:

67 columns: 260568 rows

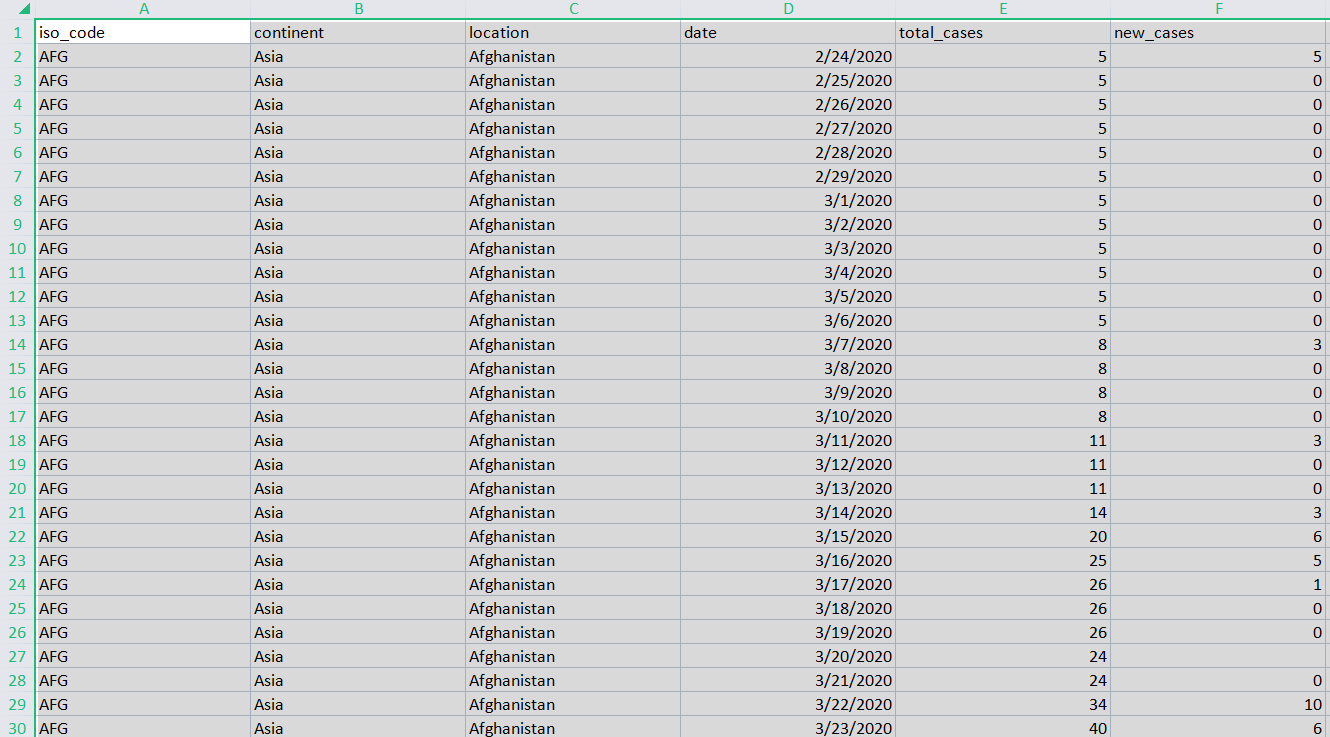
Dataset Source:

<https://github.com/owid/covid-19-data/tree/master/public/data>

Dataset description:

This dataset shows the number of corona-virus cases in many countries ranging between the year 2020 and 2023 with their corresponding months and days as well as the total number of death and the patients with other long-term diseases.

Dataset Screenshot:



**2) Vaccine dataset**-

Dataset Size:

15 columns: 86513 rows

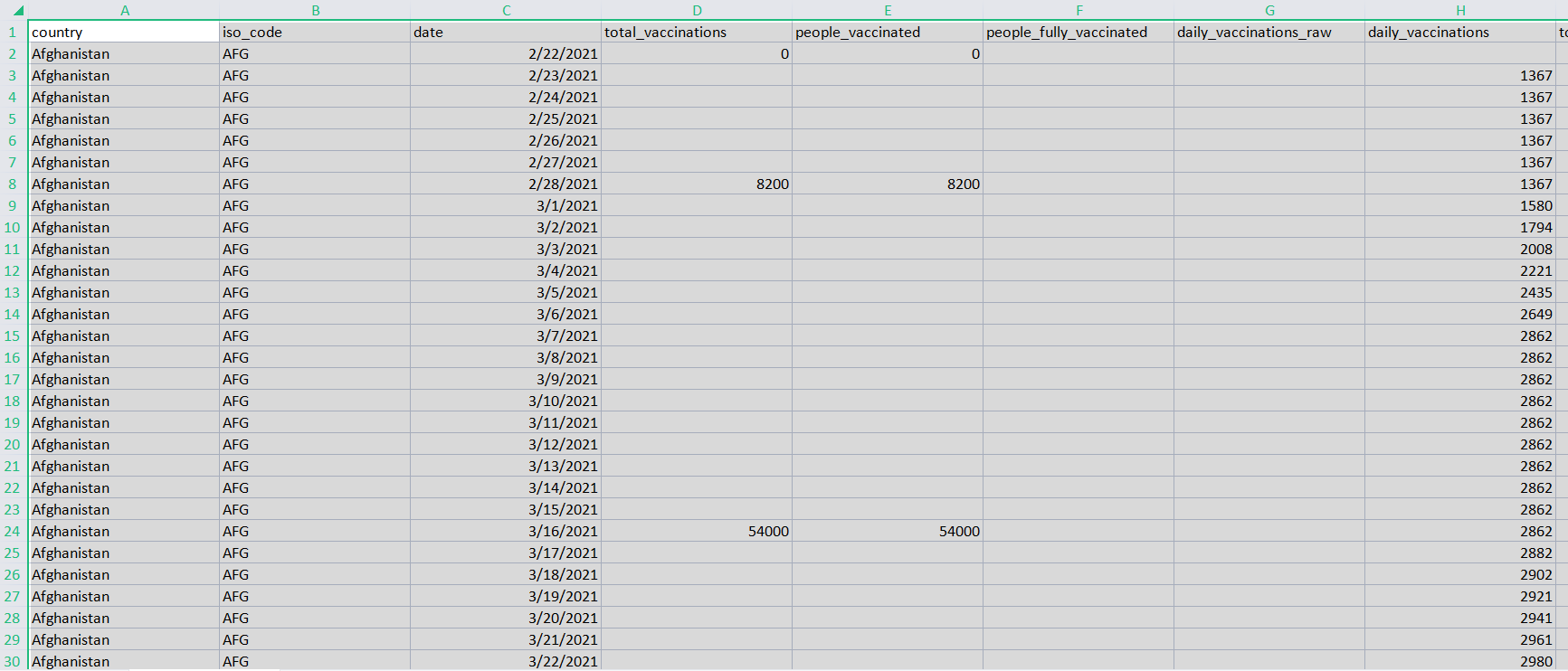
Dataset Source:

<https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress?resource=download>

Dataset Descriptions:

This dataset gives us information about many countries ranging from the year 2021 and 2022 with people that are vaccinated in total and the vaccine type.

Dataset Screenshot:



**3) Age groups dataset**-

Dataset Size:

9 columns: 5338 rows

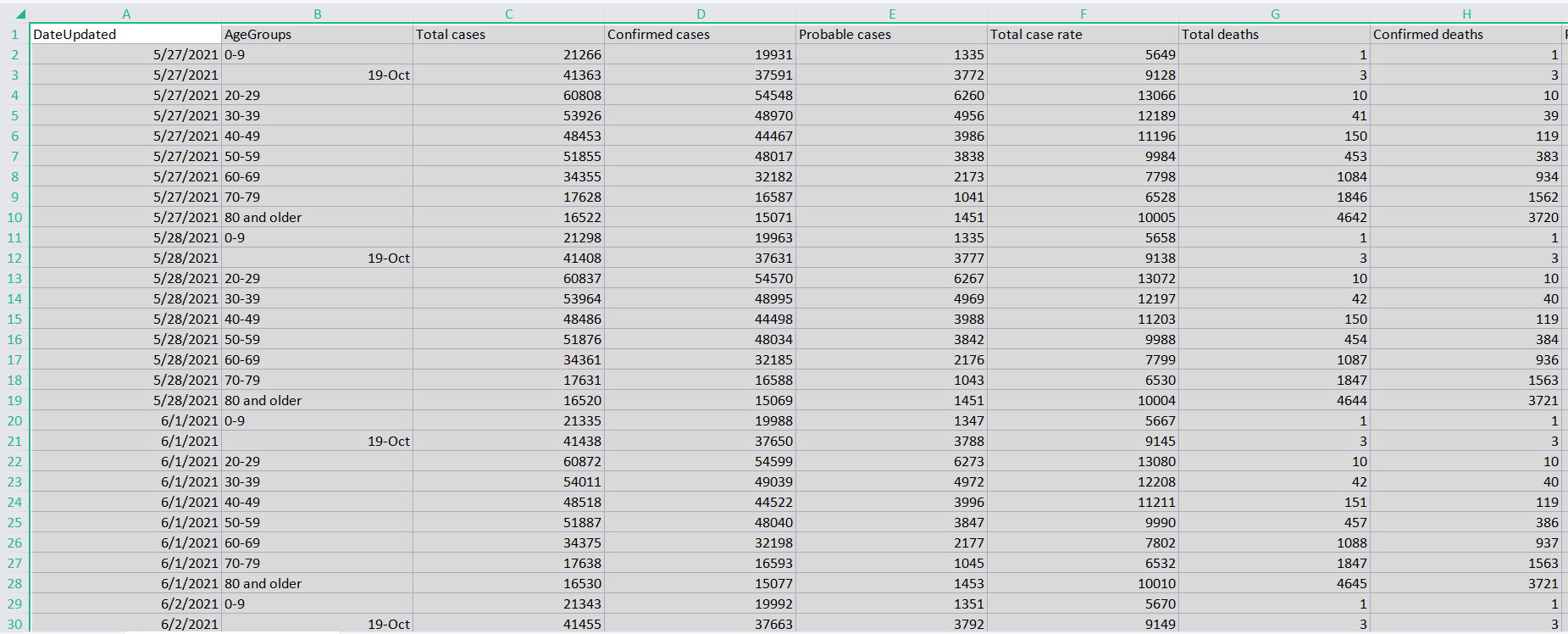
Dataset Source:

<https://catalog.data.gov/dataset/covid-19-cases-and-deaths-by-age-group>

Dataset Description:

The dataset gives us information about the total cases and number of deaths in each age range from the years 2020 and 2021.

Dataset Screenshot:



**4) Containment measure dataset-**

Dataset Size:

4 columns: 1704 rows

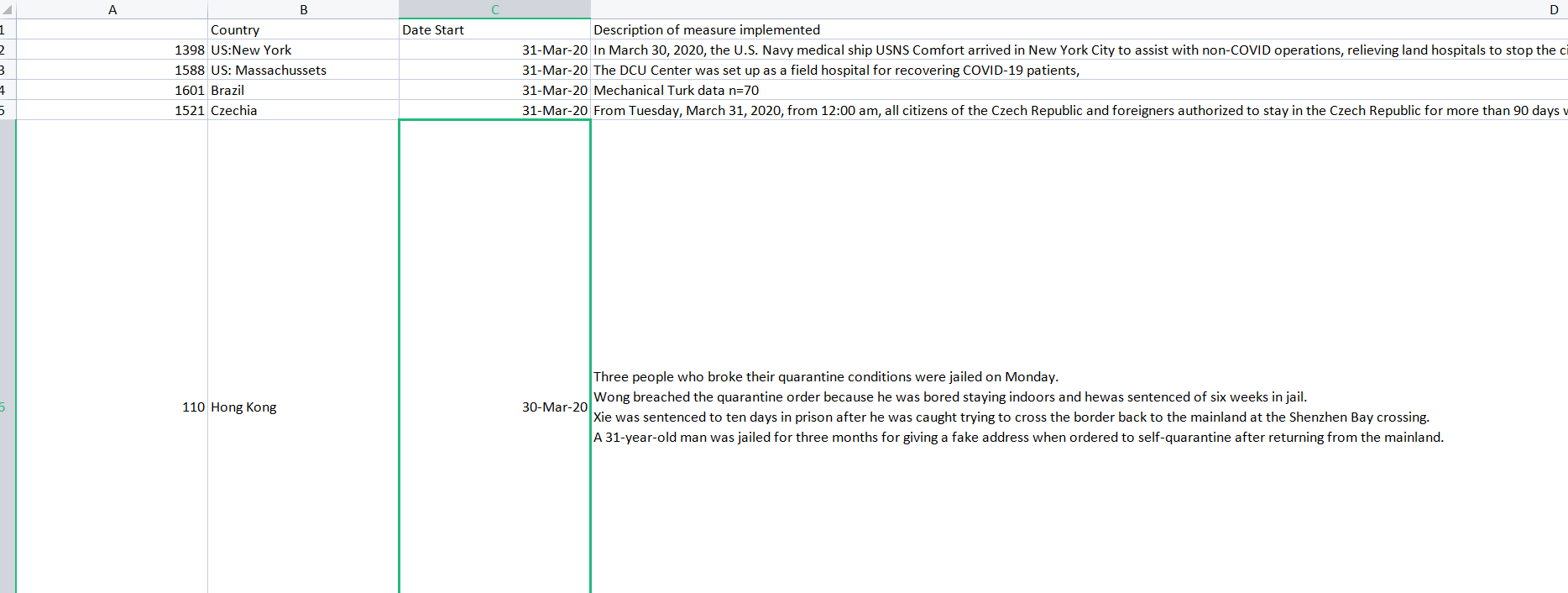
Dataset Source:

<https://www.kaggle.com/code/paultimothymooney/covid-19-containment-and-mitigation-measures/output>

Dataset Description:

The datasets gives us information about the countries and the measures they took when they knew about the corona-virus outbreak from the year 2019 and 2020.

Dataset Screenshot:



**Part2:**

1. How did the coronavirus affect the GDP (according to the dataset)?

Dataset used: (Owid) Covid dataset.

Website table scrape: <https://www.macrotrends.net/countries/ranking/gdp-per-capita>

Excel after scrape: gdp.csv

Q1 Analysis:

In the year 2020 the GDP was lower than the GDP in 2021 because the corona cases got lower except North America and Oceania where the GDP in 2020 was higher.

Chart, bar chart

Description automatically generated

1. Did Covid cause a drastic change in the Obesity Rate of each Continent? If so, how? Which Continent has the highest Obesity rate before covid? And after Covid?

Dataset used: (Owid) Covid dataset, Vaccine dataset.

Website table scrape: <https://data.worldobesity.org/tables/prevalence-of-adult-overweight-obesity-2/>

<https://en.m.wikipedia.org/wiki/List_of_countries_by_obesity_rate>

Excel after scrape: ObesityData.xlsx, ObesityData2.xlsx.

Q2 Analysis:

in the following bar plot the Obesity Rate Before Covid and Obesity Rate After Covid of each continent are displayed in order to compare and specify whether the Obesity rate is affected by Covid or not through comparing the obesity rate before and after covid of each continent and according to the plot and our dataset the Obesity rates went through a drastic change after covid in some of the continent such as Europe since it highly increased after covid as well as North America since it highly decreased after covid. moreover, all the continents had a change after covid. The Continent with the highest obesity rate before Covid is Asia however after covid the continent with the highest obesity rate became Europe.

Chart, bar chart

Description automatically generated

Q3- Does the Overweight rate affect the Diabetes rate of each continent? If so, how?

Dataset used: (Owid) Covid dataset, Vaccine dataset.

Website table scrape: <https://data.worldobesity.org/tables/prevalence-of-adult-overweight-obesity-2/>

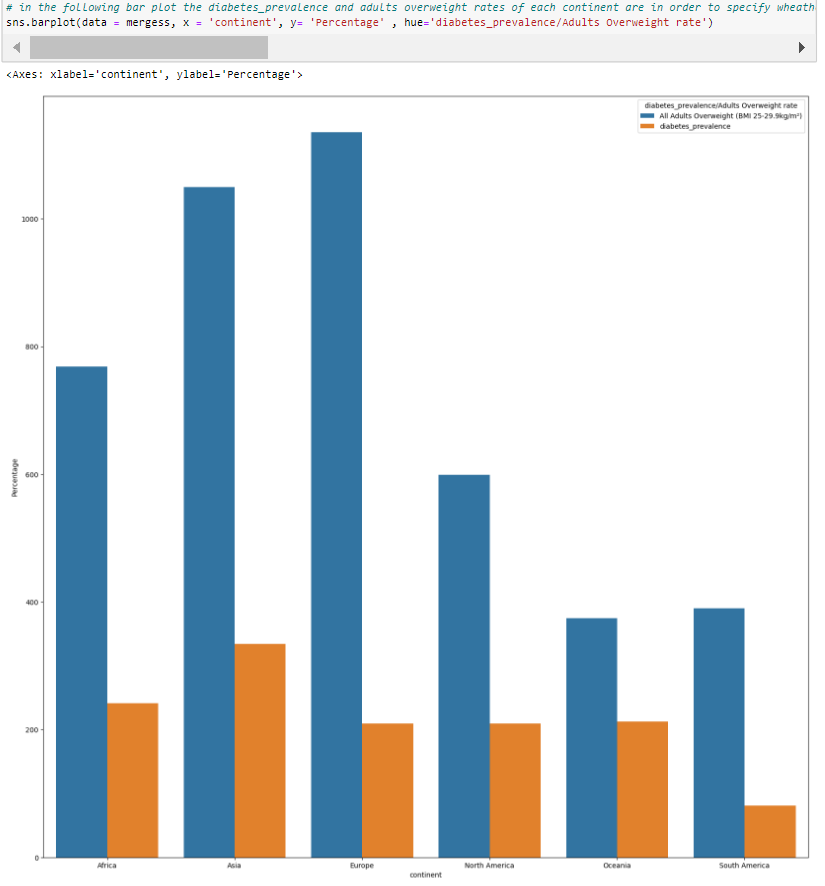
<https://en.m.wikipedia.org/wiki/List_of_countries_by_obesity_rate>

Excel after scrape: ObesityData.xlsx, ObesityData2.xlsx.

**Q3 Analysis:**

in the following bar plot the diabetes\_prevalence and adult's overweight rates of each continent are displayed in order to specify whether the adult overweight rate affects the diabetes\_prevalence rate and as you can see in the bar plot according to our datasets having a high overweight rate affects the diabetes\_prevalence but not fully since there are other factors that affects the diabetes\_prevalence rate.

Question 3 Plot:



Q4- Which Age Group was the most affected by depression after the year 2020?

Dataset used: covidagegroup.

Website table scrape: <https://www.cdc.gov/mmwr/volumes/70/wr/mm7013e2.htm>

Excel after scrape: DepressionmeeAge.csv, DepressionAge.csv.

Q4 Analysis:

According to our dataset the year after 2020 is 2021 where in the figure below it is seen that the Age range of 20-29 has the Highest depression rate.

Chart, bar chart

Description automatically generated

Q5 -Calculate the profit of 2019 from travel and tourism.

Dataset used : (owid) covid dataset

Website table scrape:

<https://www.economicsobservatory.com/how-is-covid-19-affecting-international-travel-and-tourism>

excel after scrape: international\_travel\_and\_tourism1.csv

Q5- analysis

A picture containing text, screenshot, diagram, line

Description automatically generated

Q6 - Check which content has the highest losses in 2020 from travel and tourism also check which content has the highest losses in 2021 from travel and tourism.

Dataset used: (owid) covid dataset

Website table scrape:

<https://www.economicsobservatory.com/how-is-covid-19-affecting-international-travel-and-tourism>

excel after scrape: international\_travel\_and\_tourism1.csv.

Q6- analysis

2020

According to the dataset Africa has greater number of loses than other continent and north America has lowest number of loses.

Question 6 Plot:

A picture containing text, screenshot, diagram, line

Description automatically generated

2021

According to dataset Africa still the one who has greatest number of loses and Europe become the one who has lowest number of loses.

A picture containing text, screenshot, diagram, plot

Description automatically generated

Conclusion:

The Covid has affected the Obesity rate, Depression rate, GDP, and Airline Profit. Regarding the obesity rate according to the plot and our dataset the Obesity rates went through a drastic change after covid in some of the continent such as Europe since it highly increased after covid as well as North America since it highly decreased after covid. moreover, all the continents had a change after covid. The Continent with the highest obesity rate before Covid is Asia however after covid the continent with the highest obesity rate became Europe. Moreover, according to our datasets having a high overweight rate affects the diabetes prevalence but not fully since there are other factors that affect the diabetes prevalence rate. Moreover, regarding the depression, the relationship between The Age groups , the depression values and the Year where in age ranges 0-9 there was no depression value where in 20-29 it is seen that in 2020 and 2021 has hit its highest depression value between the rest of the age ranges with this information we can reach a conclusion that the most age-range with depression is the ages 20-29.Furthermore, regarding the GDP In the year 2020 the GDP was lower than the GDP in 2021 because the corona cases got lower except North America and Oceania where the GDP in 2020 was higher. Moreover, regarding the airline profits According to the dataset in 2020 Africa has greater number of loses than other continent and north America has lowest number of loses. And in 2021 Africa is still the one who has greatest number of losses and Europe become the one who has lowest number of losses.