



Semester 1 – 2021/2022

Course Code	DS620
Course Name	Data Visualization
Assignment type	Critical Thinking
Module	11

Student ID	G200007615
Student Name	Abdulaziz Alqumayzi
CRN	15059

Solutions:

Data Scientist Job Market in the U.S.

Introduction

In this critical thinking, our mission is to create Python graphics for Covid-19 dataset. In this assignment, we will complete the following tasks: Examine the information provided. Examine the dataset for missing values and, if any exist, fill them with 0s. Each column's data should be converted to a suitable data type. Create interactive maps that illustrate where each geographical column is located (latitude and longitude). Sort the information into categories that make sense to you. Using the relevant chart, choose and visualize the data. The dataset used for this critical thinking was retrieved from the **Kaggle** called Coronavirus Lat/Long Dataset. The following is the dataset's processed version from 2019 Coronavirus dataset (January - February 2020) dataset. To make geographic analysis easier, author replaced NAs with 0 and added longitude and latitude columns.

Tasks Processing Using Python

Task 1 - Read the data: in the following figure 1, two datasets were imported. The original data for the critical thinking and a support dataset to unify the latitude and longitude.

Figure 1

Importing datasets

```
df = pd.read_csv('CV_LatLon_21Jan_12Mar.csv') # import the dataset
df_loc = pd.read_csv('countries.csv') # import location of countries dataset
```

Task 2 - Explore the dataset, make sure there are no missing values, and if there are, they are filled with 0s: data was investigated for how many columns and rows, null values, data types of columns, number of countries exists to create new dataframe we will use in creating interactive map visualization. the datasets were joined but there were many countries missing, and one irrelevant value was found as a country, which is Cruise Ship.

Task 3 - Group the data as you see appropriate: the new dataframe has aggregate to sum the values of confirmed, recovered and death columns in figure 2.

Figure 2

Grouping

```
df_final = df3.groupby(['country', 'latitude', 'longitude'])['confirmed', 'recovered', 'death'].sum().reset_index()
```

Task 4 - Convert data in each column into an appropriate data type: the final dataset (df_final) contains the right data types.

Task 5 - Create interactive visualizations that show the location of each geographical column (latitude and longitude): plotly package was used to create the interactive map visualization. In figure 3 is the code to create the map and figure 4 is the map showing data points of confirmed covid-19 injured and the color shows the density.

Figure 3

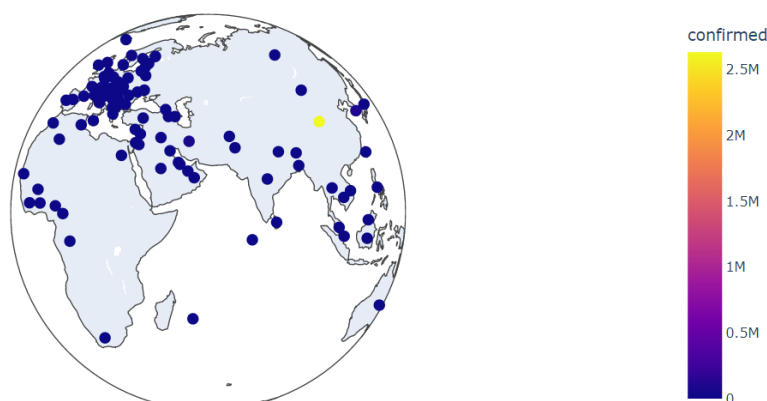
Code for creating the map of confirmed covid-19 injury

```
map_fig_1 = px.scatter_geo(df_final, lat='latitude',
                           lon='longitude',
                           color='confirmed',
                           projection='orthographic',
                           title='Confirmd Covid-19 Map',
                           hover_name='country',
                           hover_data=['confirmed', 'recovered', 'death'])
map_fig_1.update_traces(marker=dict(size=10))
map_fig_1.show()
```

Figure 4

Map of confirmed covid-19 injury

Confirmd Covid-19 Map



In addition, two maps were created for recovered and death.

Task 6 - Choose and visualize the data using the appropriate chart: bar chart was used to view the confirmed, recovered and death of covid-19. In figure 5 shows the coding that create the interactive bar chart.

Figure 5

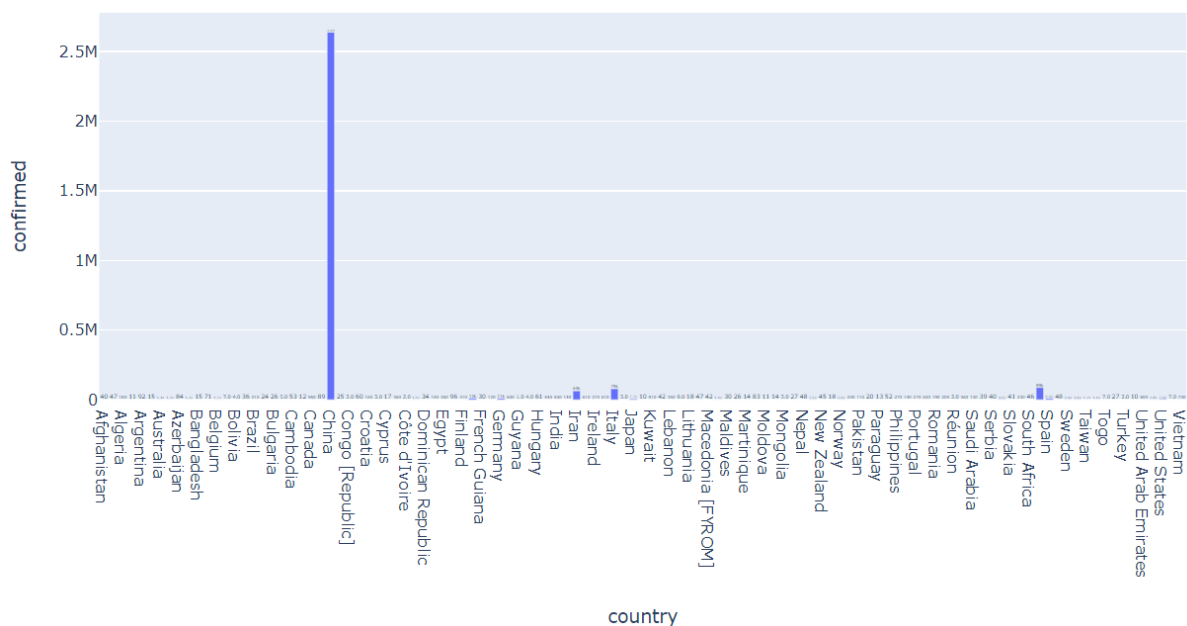
Code for creating the bar chart of confirmed covid-19 injury

```
fig_1 = px.bar(df_final, y='confirmed', x='country', text='confirmed')
fig_1.update_traces(texttemplate='%{text:.2s}', textposition='outside')
fig_1.show()
```

And the figure 6 shows the chart of confirmed covid-19 injuries. The bar chart interactive, so we can zoom in and out to see more details.

Figure 6

Bar chart of confirmed covid-19 injury



Summarizes of the Results Statistical and Visual Analysis

China results make other countries seems are very low. In fact, this dataset are collected before March 2020, in that time China was only the affected country by covid-19. Both map and bar chart show only the most affected country, but when we hover on ca country on the map or zoom-in in the bar chart we can see more details.

References

- Belorkar, A., Guntuku, S. C., Hora, S., & Kumar, A. (2020). *Interactive data visualization with python: Present your data as an effective and compelling story*. Packt Publishing.
- So, B. (2020, February 6). *2019 coronavirus dataset (January - February 2020)*. Kaggle. Retrieved November 27, 2021, from <https://www.kaggle.com/brendaso/2019-coronavirus-dataset-01212020-01262020>.
- Maliarenko, M. (2020, March 13). Coronavirus Lat/Lon Dataset. Kaggle. Retrieved November 27, 2021, from <https://www.kaggle.com/grebublin/coronavirus-latlon-dataset>.
- Google | dataset publishing language . (2021). Countries.csv . google developers. Retrieved November 27, 2021, from https://developers.google.com/public-data/docs/canonical/countries_csv.