Team Members	No	Full Name	Student ID
	1.	Aitore Issadykova	201275926
	2.	Assem Kussainova	201535798

1. **Project Title:** Prediction of the peak of COVID-19 cases.

2. **Problem Description:** The aim of this work is to study the conditions of the spread of the COVID-19 (temperature and humidity) to predict the peak of coronavirus cases in countries where this problem has not yet been encountered or where the infection develops at the initial stage. Coronavirus infection COVID-19 is an infectious disease caused by a new coronavirus that has not previously been detected in humans. According to recent publications, the optimal temperature for pathogen transmission is up to 10 degrees, and air humidity up to 50%. The work will check previously published statistical trends and analyze the relationship between the number of infected and temperature conditions.

3. **Significance of the Problem:** Having spread outside of China, the coronavirus epidemic has spread to over 150 countries. Since the advent of the new coronavirus has taken the world by surprise. In record time, the mortality rate increased due to lack of training and necessary measures. In connection with the foregoing, the motivation for this study was the determination of the nearest peaks of the disease for timely adoption of measures to enhance the safety of public health.

4. **Data Requirements:** Key data will be collected from countries where the peak of the spread of the virus has already passed, such as China, Italy and South Korea. Historical data on temperature and humidity during outbreaks will be used along with data on morbidity and mortality among the population of the above countries. Weather data in selected countries will be collected from the World Weather website [1], and data on the prevalence of coronavirus infection will be taken from statistics maintained by the website Worldometer.info [2].

5. **Infrastructure Requirements:** All work will be done on the personal computers of team members. The work will be divided in such a way that it will be possible to carry out tasks in parallel without any conflicts. The Python programming language and Anaconda component from a Python distribution will be used to process and analyze data.

6. **Existing Solutions and Their Shortcomings:** COVID-19 is a relatively new disease, as a result of which there are no earlier publications on it. Researchers from Beihang University and Tsinghua University conducted research that concluded that there is a possibility that high temperature and high humidity reduce the transmission of COVID-19 [3]. The study was based on data from 100 Chinese cities with more than 40 cases. In the proposed work, the amount of data will expand from 1 country to 3, which may contribute to more accurate and important results. In addition, the criteria for the most dangerous weather conditions with which the virus spreads will be clarified.

7. **Technical Roadmap:** The work includes the stages of data visualization, analysis, building a data model, checking for accuracy and conclusions. The first stage is data visualization, during which temperature maps, graphs of temperature and humidity changes over time, the rate and growth trends of the number of infected among the population can be built. If there are any patterns or trends in the first stage, the project will continue with building models and testing them on existing data from China, Italy and South Korea. Several machine learning algorithms will be chosen to build the models, and for prediction purposes the Mean Squared Errors (MSE) will be used as a metric and accuracy will be testing by Mean Absolute Percentage Errors (MAPE). Further, the resulting behavior model can be designed for countries at the initial stage of infection, including Kazakhstan. On the other hand, if there no connection between COVID-19 cases and air temperature and humidity will found, the project will aim to predict market stock prices during pandemic times such as MERS, SARS, Swine Flu etc. The forecasting model will be trained on previous pandemic times and will tested on the latest two pandemic eras. The same metrics: MSE and MAPE – will be applied for this project.
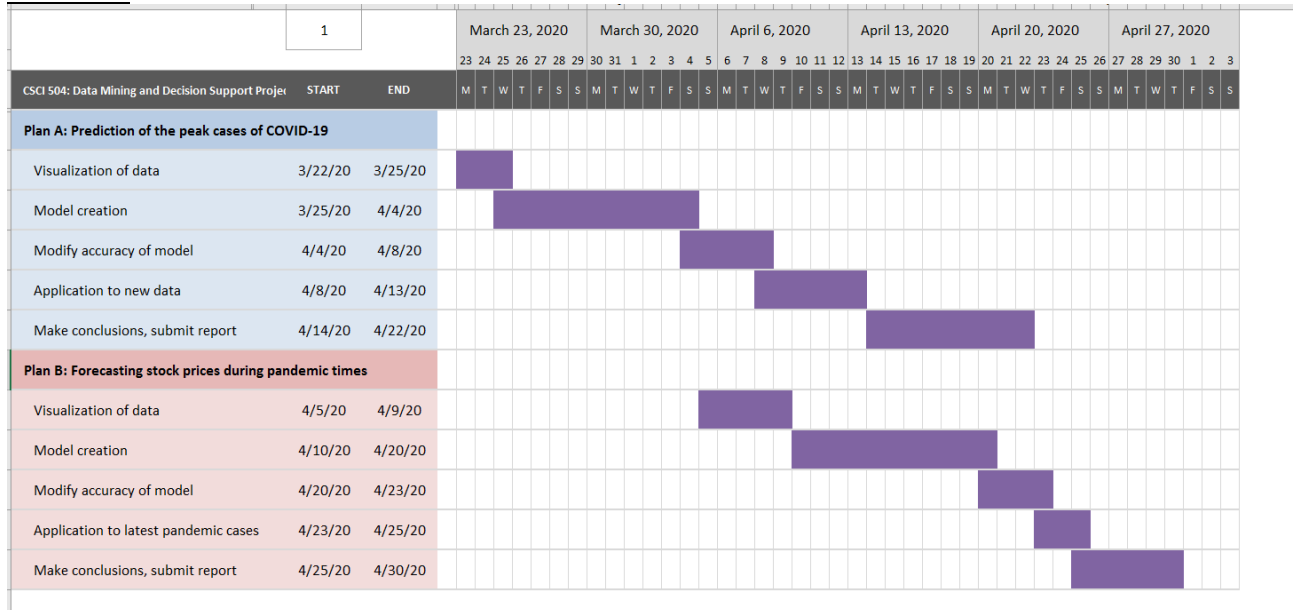
8. **Schedule:**

CSCI 504: Data Mining and Decision Support Project	START	END	March 23, 2020	March 30, 2020	April 6, 2020	April 13, 2020	April 20, 2020	April 27, 2020
Plan A: Prediction of the peak cases of COVID-19								
Visualization of data	3/22/20	3/25/20						
Model creation	3/25/20	4/4/20						
Modify accuracy of model	4/4/20	4/8/20						
Application to new data	4/8/20	4/13/20						
Make conclusions, submit report	4/14/20	4/22/20						
Plan B: Forecasting stock prices during pandemic times								
Visualization of data	4/5/20	4/9/20						
Model creation	4/10/20	4/20/20						
Modify accuracy of model	4/20/20	4/23/20						
Application to latest pandemic cases	4/23/20	4/25/20						
Make conclusions, submit report	4/25/20	4/30/20						

Figure 1: Gannt Chart

9. <u>**References:**</u>

[1] Tutiempo Network, S.L. (2020) World Weather - Local Weather Forecast. Retrieved from https://en.tutiempo.net/

[2] Worldometer (2020). COVID-19 coronavirus pandemic. Retrieved from https://www.worldometers.info/coronavirus/

[3] Wang, Jingyuan and Tang, Ke and Feng, Kai and Lv, Weifeng, High Temperature and High Humidity Reduce the Transmission of COVID-19 (March 9, 2020). Available at SSRN: https://ssrn.com/abstract=3551767