**Portfolio Project - Option #1: Business Intelligence Solution for U.S. Organization**

Ana Moser

Colorado State University - Global Campus

MIS480 – Capstone – Business Analytics & Information Systems

Dr. Justin Bateh

February 7, 2021

**Portfolio Project - Option #1: Business Intelligence Solution for U.S. Organization**

The New York Times is a company dedicated to reporting stories worldwide, believing that transmitting information to the public about different topics will enrich people's lives. It was established in 1851, and it has a reputation for being objective and focused on the cultured and intellectual audience. The New York Times has covered and reported many events worldwide that have positioned it as the largest newspaper in circulation (Augustyn, 2020). This organization should count on an efficient and reliable system that allows it to gather, analyze, and deliver data to comply with its business objective. The New York Times has access to different data sources, making it an excellent candidate to implement business intelligence tools to enhance its processes.

The New York Times is working with public datasets about the COVID-19 outbreak since January 2020. The data gathered by health care organizations need to be compiled and organized to understand any patterns and take decisions to avoid the spread. The first COVID-19 case confirmed in the U.S. was on January 21, 2020, and since then, the New York Times has been tracking the numbers of cases (Allen et al., 2020). This study's variables are date, county, State, fips, cases, and deaths for COVID-19 from the first case until December 31, 2020 (The New York Times, 2020). Many researches and companies rely on the information that circulates online and The New York Times should focus on providing excellent coverage of important events to show unbiased and objective reports for its users.

The public dataset is stored in GitHub, where many other entities could access and analyze the impact of COVID-19 cases in different industries. Tracking the spread of this virus could help understand the effect in some sectors, and it could help increase the measures of safety and find secondary plans to reinforce the economy. Some of the industries affected by COVID-19 are airlines, leisure facilities, oil and gas drilling, auto parts and equipment, and restaurants (Haydon & Kumar, 2020). The analysis of this dataset is done by SAS, a statistical software that can handle big data and represent it in graphs. It is an efficient and reliable BI solution tool that will support the New York Times' work covering the pandemic (Why Business Intelligence Is Important, 2020).

The ongoing repository of COVID-19 cases and deaths updated by The New York Times on GitHub has been compiled from state and local governments, and health departments, thus, this data is reliable and it could be accessed by the public (The New York Times, 2020). By August 2020, there were 79,611,982 tests reported from which 6,873,739 tested positive (Previous U.S. Viral Testing Data, 2020). The active COVID-19 testing within the U.S. has became challenging to gather and process the data due to its extensive size. However, this is a perfect opportunity to implement business intelligence tools to report data through the analysis of summaries and graphs that will present a better detail of trends that eventually will help to predict or understand behaviors (Pratt & Fruhlinger, 2019).

SAS (Statistical Analysis Software) or statistical software package is handy for predictive, descriptive, and prescriptive analysis. It is also used for data management and graphical reports due to its versatility in the acceptance of several input data sources. It has a unique graphical user interface (GUI) that grants the facility to use this software. SAS provides high-quality machine learning features for data analytics that can handle a high volume of data. SAS is a competitive tool with both analytical and graphical aid to inform and represent useful patterns that will provide better insight into the responsible entities to make decisions. It offers high security and excellent customer service levels; however, it is not open-source and is highly-priced (Advantages of SAS, n.d.). This business intelligence solution will provide to The New York Times the interface to import, analyze, and graph data in an efficient way to release reliable reports.

Some of the graphical data representations that SAS provides are histograms, bar charts, pie charts, scatter plots, box plots, and heat maps. Data visualization plays an essential role in organizations that want to describe patterns or find out anomalies. Each graphical representation has specific uses depending on the nature of the variables; for instance, a histogram represents the probability of distribution of a continuous variable, whereas a scatter plot shows the relationship between two variables (Kumari, 2020). The New York Times currently is using histograms, bar charts, and heat maps to represent COVID-19 cases within the U.S.

**Figure 1.**

*Code to import raw dataset into SAS.*

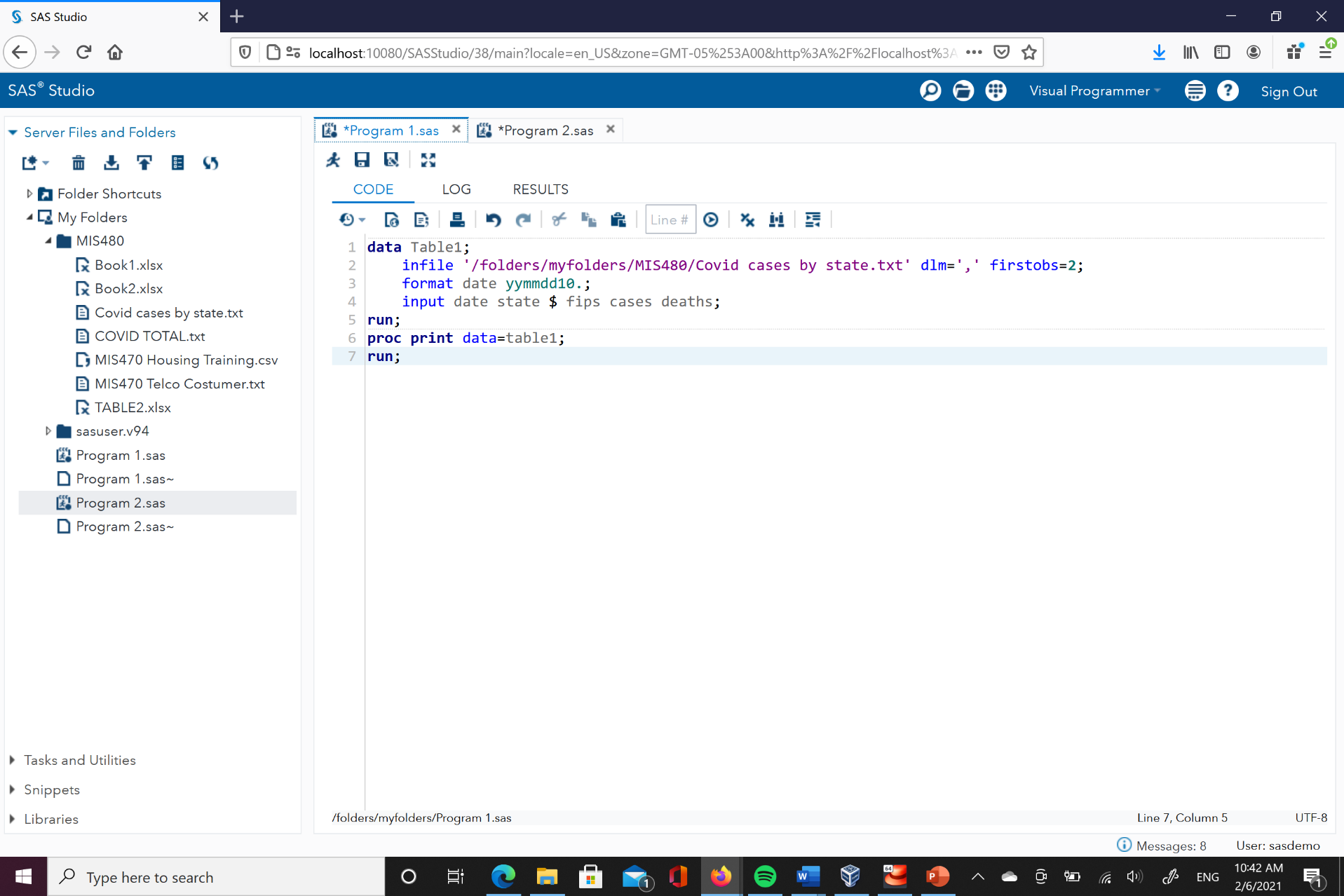


Figure 1 shows the code and format necessary to import the raw dataset into SAS. It is a csv file with 16735 observations and five variables. Once the raw file is uploaded to SAS, it has to be imported, then, a file is created for it (Table 1). The proc print function will display the dataset organized into the columns that were identified in the code, this output is shown in Figure 2.

**Figure 2.**

*Output from the raw dataset imported (First 15 rows).*

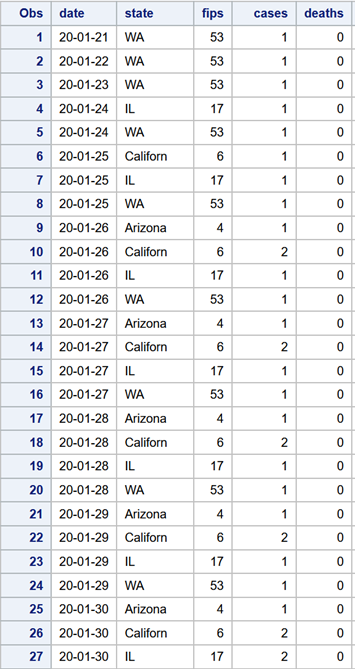
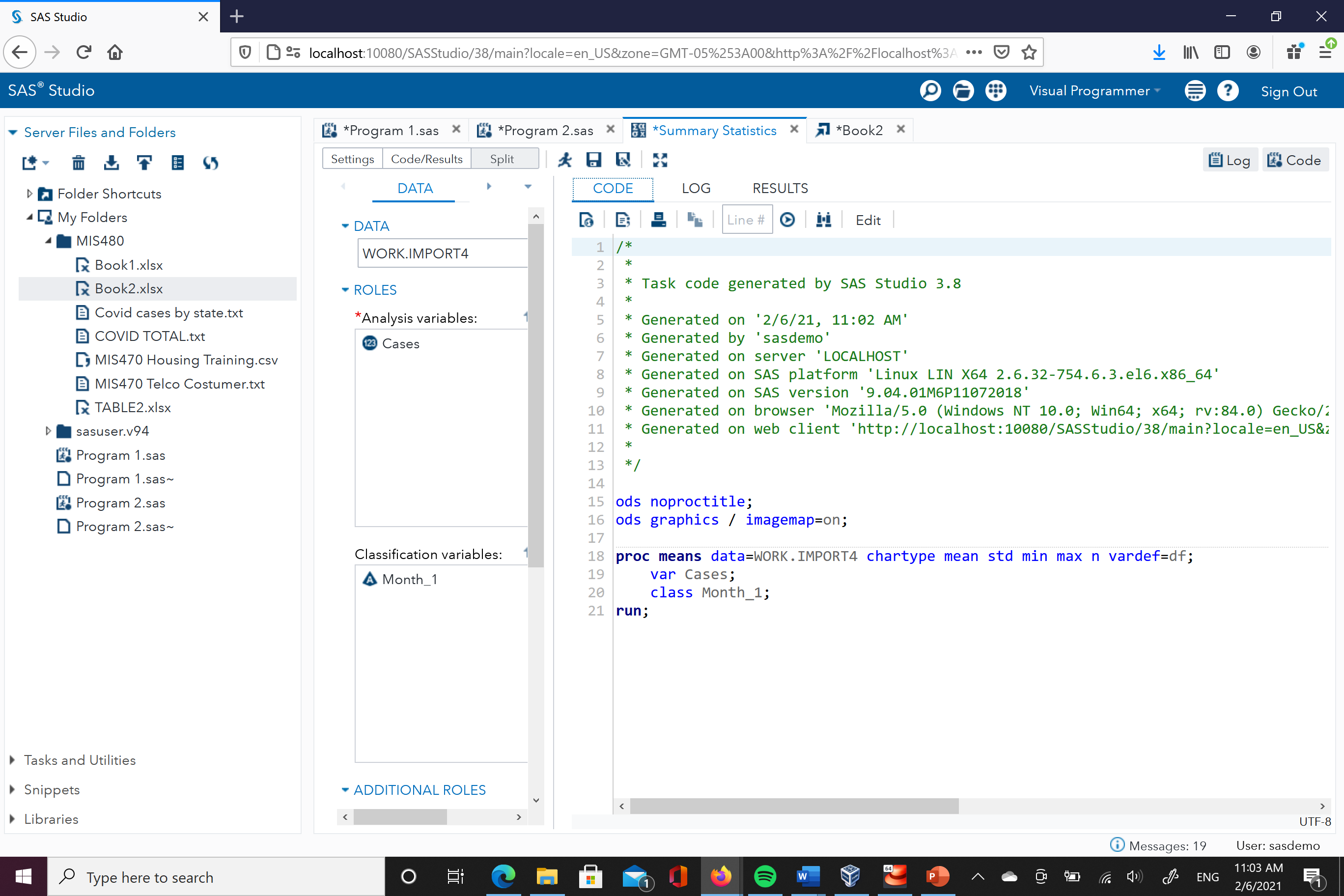


Figure 3 shows the code to calculate the mean, standard deviation, minimum and maximum values, and number of observations of the dataset (summary statistics). *Proc means* is a procedure or function in SAS that computes descriptive statistics. The variable that is being analyzed is cases and it will be classified by month. This same code could be used to analyze the variable deaths, also, it could be classified by state or fips. Figure 4 shows the output, however, numbers are good for some, but sometimes data visualization could provide better insights, even for no SAS experts. From this report, we can see that December was the month with more cases, the rate increased along the year.

**Figure 3.**

*Code to calculate summary statistics in SAS – cases variable.*



**Figure 4.**

*Output summary statistics in SAS – cases variable.*

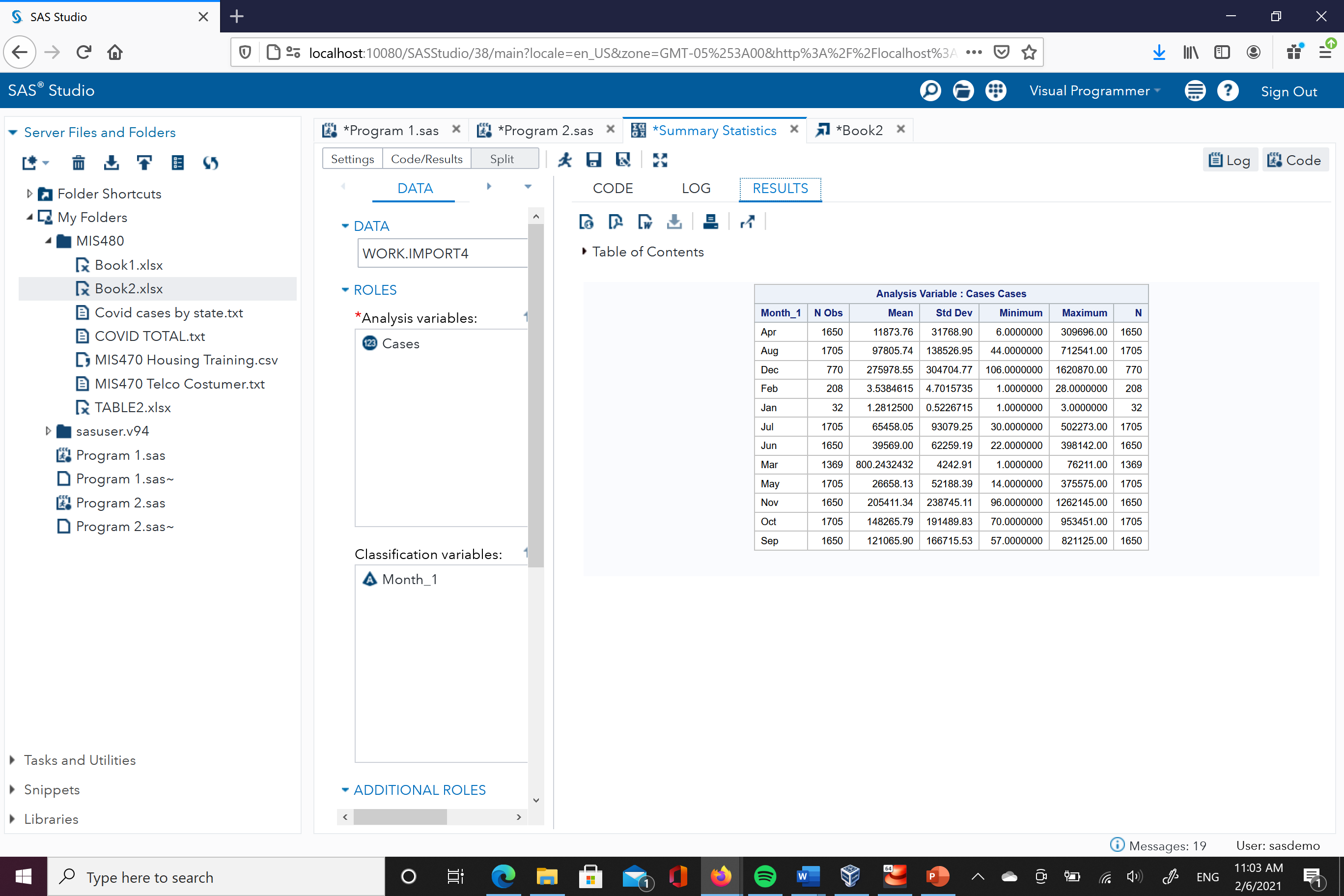
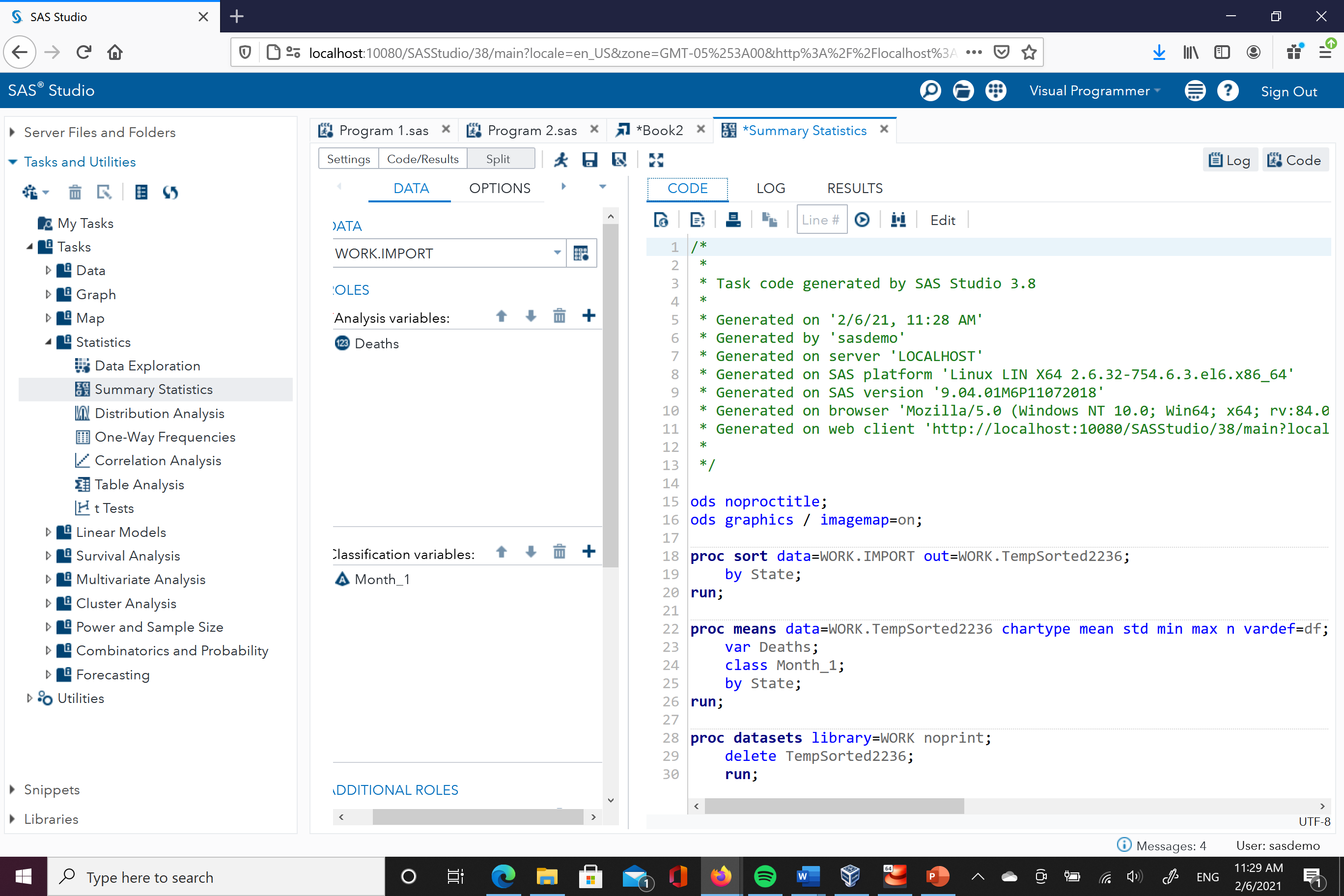


Figure 5 shows the code to compute the summary statistics for the variable deaths classified by month and by state, and Figure 6 display the output. SAS uses procedures (PROC) as functions to run it through the software. For each State, we can see in which month there were more deaths.

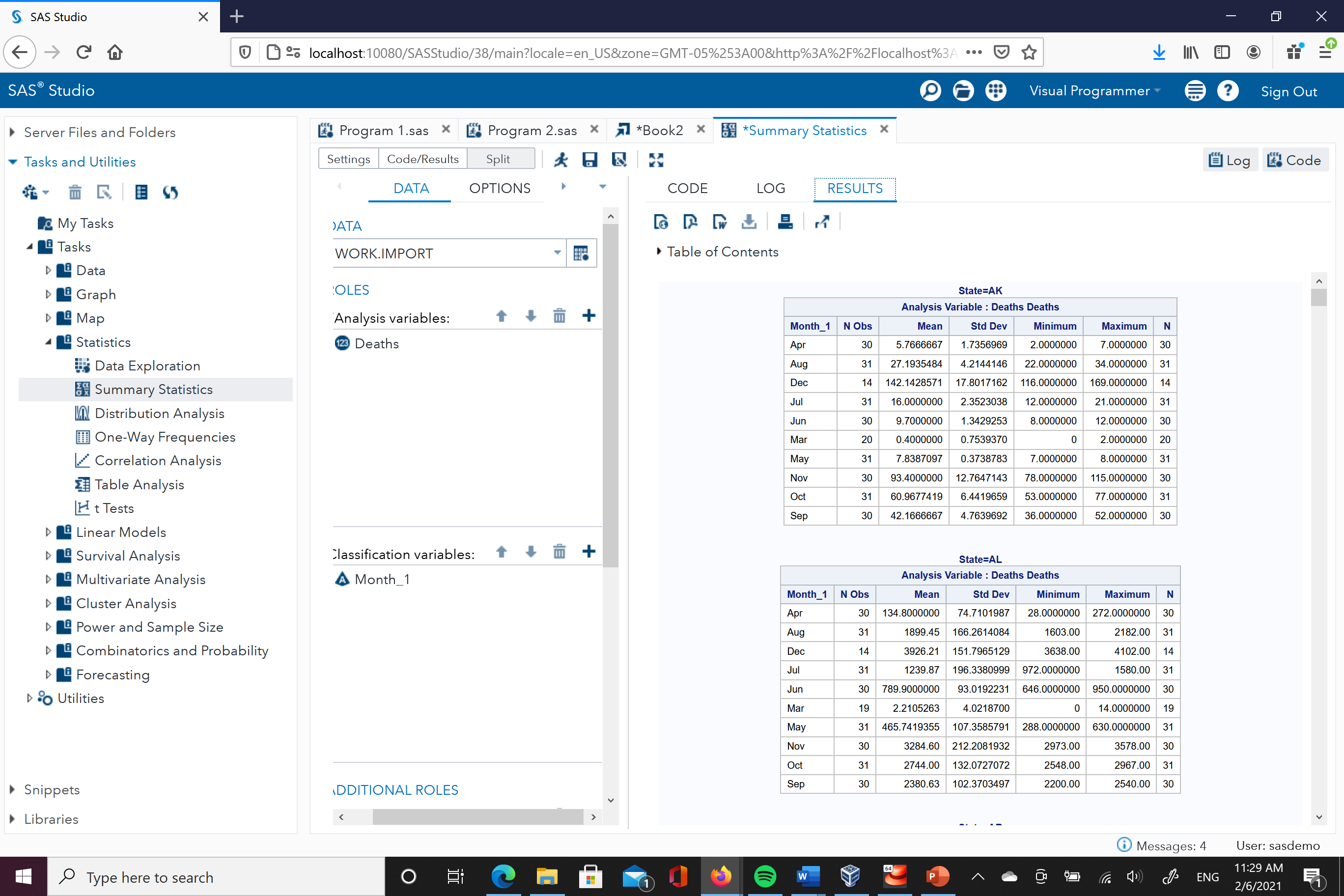
**Figure 5.**

*Code to calculate summary statistics in SAS – deaths variable (classified by month and State).*



**Figure 6.**

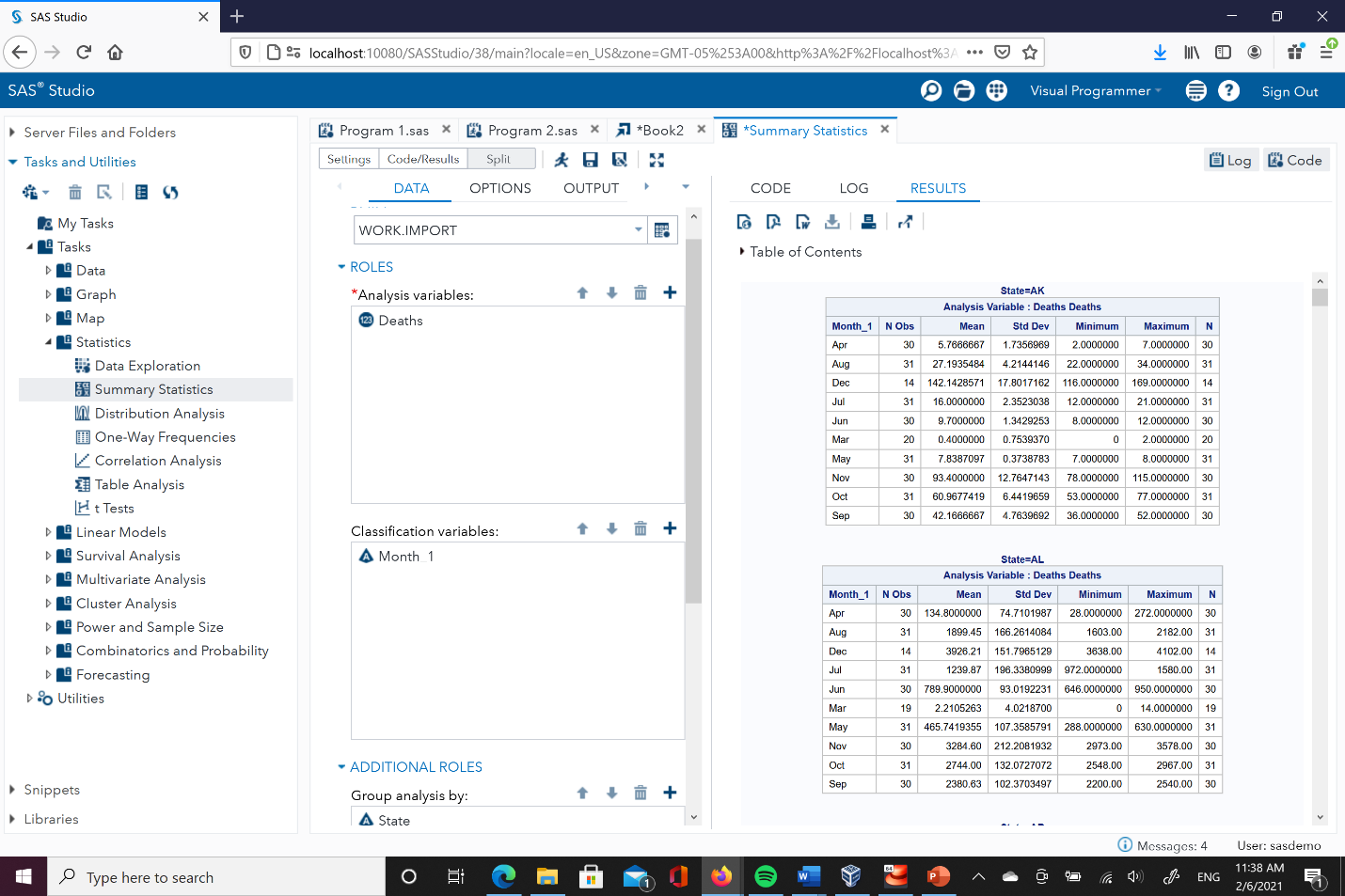
*Output summary statistics in SAS – deaths variable (classified by month and State).*



Although executing this code is one way to run programs in SAS, it also has a graphical user interface (GUI) that grants the facility to use this software. Figure 7 shows a screenshot that was used to compute the summary statistics for the variable deaths classified by month and by state. SAS allows the user to interact with several options available to analyze data. Instead of using code, the user can click in the task that will execute the command; however, the software will generate the code automatically. Many other tasks can be executed by SAS, as it is shown in Figure 7, some of them are cluster analysis, descriptive analysis, model selection, multivariate analysis, and predictive modeling among others (SAS Features, n.d.).

**Figure 7.**

*Graphical User Interface (GUI) in SAS.*



**Figure 8.**

*Visualization Tools in SAS.*

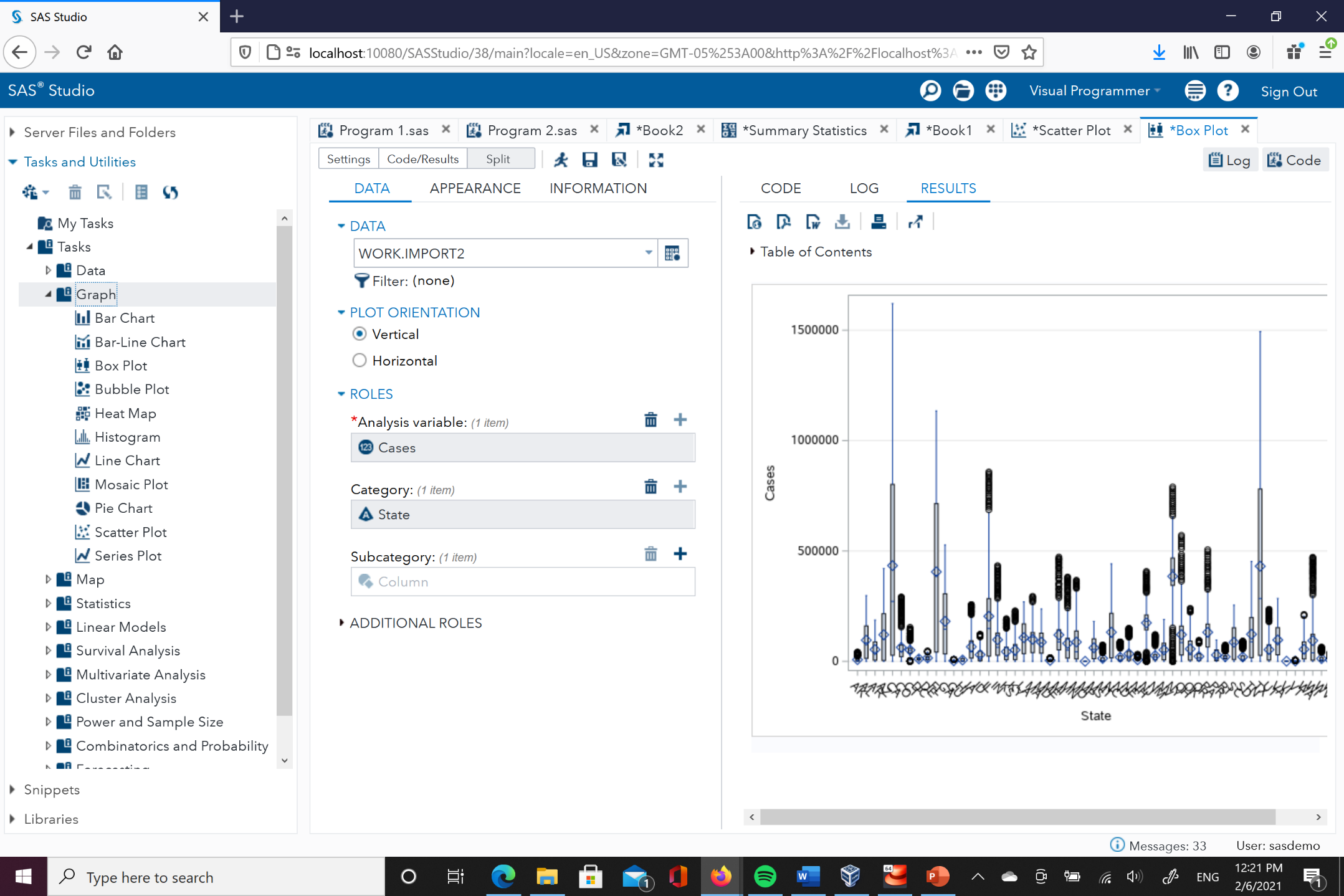
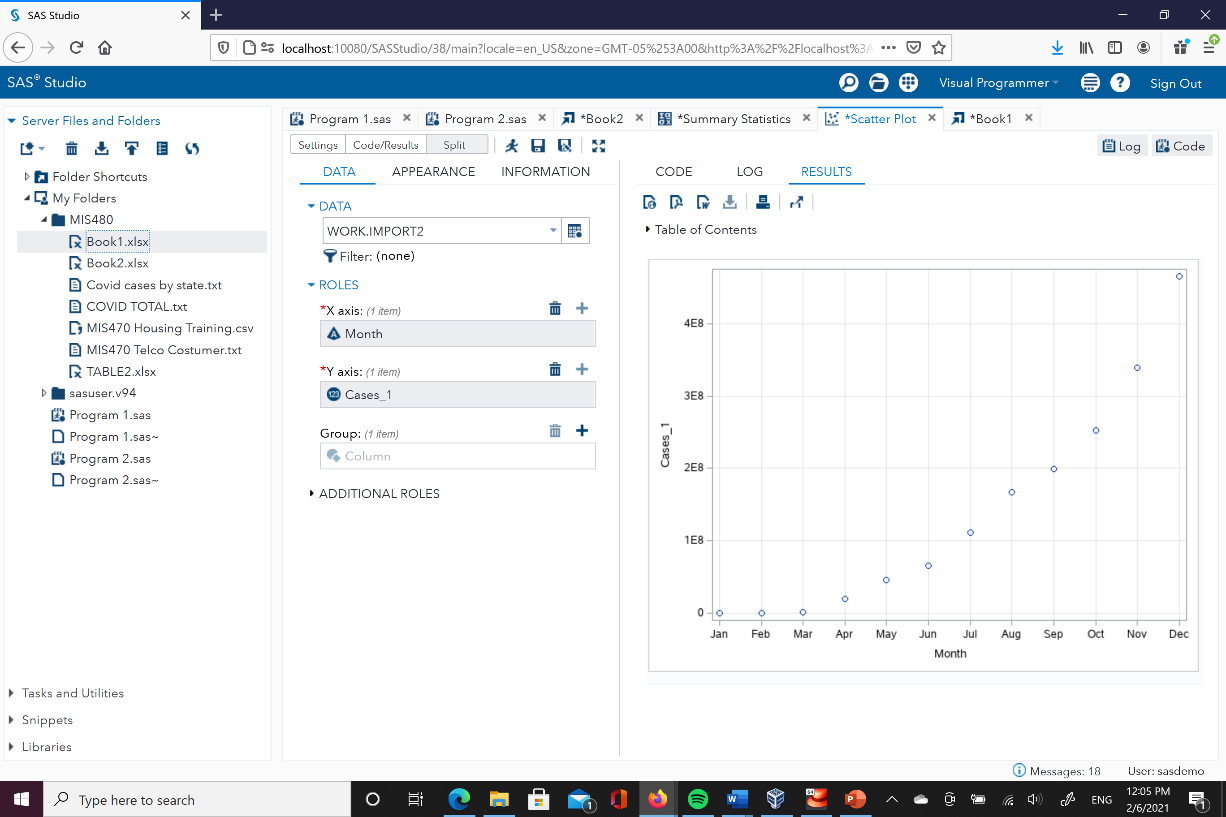


Figure 8 shows the visualization tools available in SAS to display data analytics results. Some of the most used graphs are histograms, scatter plots, and bar charts.

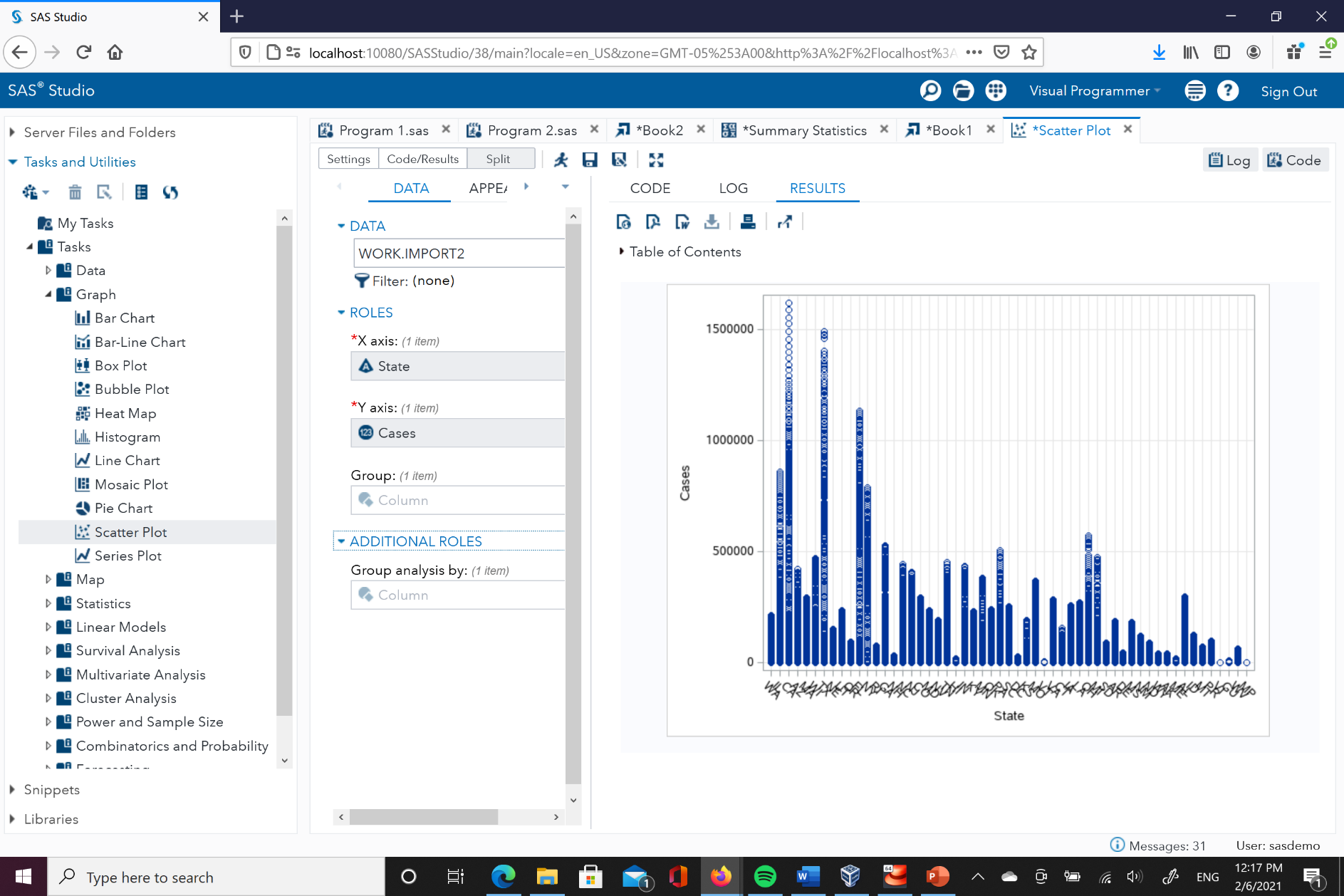
**Figure 9.**

*Scatter Plot in SAS – Month vs. Total Cases.*



**Figure 10.**

*Scatter Plot in SAS – State vs. Cases.*



Figures 9 and 10 displays the analysis outcome for the variable cases, month, and State. The possibilities to combine different variables and inspect their correlation could be useful to understand patterns of this pandemic outbreak.

The benefits of using Business Intelligence tools provide users the ability to make data-driven decisions. The variety of features such as dashboards, visualizations, and reporting give an advantage to handle vast amounts of data. Including the advanced technology and the implementation of techniques will enhance the expected outcomes. The New York Times should invest in this business intelligence tool to succeed in its COVID-19 coverage. Its mission is to inform and report the American public about the event that has impacted the most not only the people, but also the economy and other factors. Having the ability to gather, analyze, and report efficiently is the best benefit that a BI tool could offer to improve the performance of this organization (Pratt & Fruhlinger, 2019).

**Conclusion**

The implementation of business intelligence tools for The New York Times optimizes and enhances its purpose to keep informed the American public. Having an excellent background and the trust of millions of readers mean that the expectations are always high, and the opportunity to keep improving is open. Business intelligence allows manage data in an efficient and competent way to actively analyze data and draw strategies based on it. The pandemic that started in 2020 has surprised the world since no one was prepared to handle it, and the price has been paid with human lives. The rapid generation and spread of data are becoming a priority to monitor the constant release of data from the health care organizations to gather, analyze, and inform the public (Allen et al., 2020). Researchers, companies, and the public in general are being informed with reliable and trustworthy information provided by The New York Times and it is a plus to transform these software services from SAS into valuable insights to achieve the goal of enriching people’s life through the transmission of information.

**References**

Advantages of SAS. (n.d.). EDUCBA. Retrieved from https://www.educba.com/advantages-of- sas/

Augustyn, A. (2020). The New York Times. Encyclopædia Britannica. Retrieved from https://www.britannica.com /topic/The-New-York-Times

Allen, J., Almukhtar, S., Aufrichtig, A., Barnard, A., Bloch, M., Cahalan, S., Cai, W., Calderone, J., Collins, K., Conlen, M., Cook, L., Gianordoli, G., Harmon, A., Harris, R., Hassan, A., Huang, J., Issawi, D., Ivory, D., Lai, R., Lemonides, A., McCann, A., Oppel, R., Patel, J., Saldanha, A., Semple, K., Walton, J., Singhvi, A., Smart, C., Smith, M., Sun, A., Watkins, D., Williams, T., Wu, J & Yourish, K. (2020). Coronavirus in the U.S.: Latest Map and Case Count. *The New York Times*. Retrieved from https://www.nytimes.com/ interactive/2020/us/coronavirus-us-cases.html

Haydon, D & Kumar, N. (2020). Industries Most and Least Impacted by COVID-19 from a Probability of Default Perspective – September 2020 Update. *S&P Global*. Retrieved from https://www.spglobal.com/marketintelligence/en/news-insights/blog/industries- most-and-least-impacted-by-covid19-from-a-probability-of-default-perspective- september-2020-update

Kumari, N. (2020). SAS Data Visualization. *Medium.* Retrieved from https://medium.com/swlh/sas-data-visualisation-9223dc30e039

Pratt, M., & Fruhlinger, J. (2019). What is business intelligence? Transforming data into business insights. CIO. Retrieved from https://www.cio.com/article/2439504/business- intelligence-definition-and-solutions.html

Previous U.S. Viral Testing Data. (2020). Centers for Disease Control and Prevention (CDC). Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/previous- testing-in-us.html

SAS Features. (n.d.). SAS. Retrieved from https://www.sas.com/en\_us/software/stat. html#m=features

The New York Times. (2020). Coronavirus (Covid-19) Data in the United States. Retrieved from https://github.com/nytimes/covid-19-data.

Why Business Intelligence Is Important. (2020). *Analytics.* Retrieved from https://analytiks.co/why-business-intelligence-is-important/