A picture containing tree, outdoor, hillside, lush

Description automatically generated

Student Project: Seton Hill’s Response to COVID-19

Data Analytics Capstone Project

SDT 300

Matthew Fluet

**Abstract**

Coronavirus 2019 (COVID-19) has developed into a large-scale health issue causing millions of deaths around the globe. It has rapidly been transmitted over the whole world since its first detection in Wuhan, China in December 2019. The fast spread of the virus from person-to-person contact has created a public health emergency, being declared an international pandemic in March 2020 by the World Health Organization (WHO). In late September 2020 COVID-19 rose to 6,960,152 cases (WHO) causing self-isolation and social distancing in attempt to slow the spread of the virus. In the absence of having a set plan on how to react to a virus of this scale it has made it very difficult for businesses and schools to return to normalcy. In this review, I highlight Seton Hill University’s procedure and how it relates to feeling safe, respecting freedoms, mental health, student athletes and if topics were adequately learned in the classroom.

**Introduction**

When deciding on a process how Universities will handle the large amounts of students who are coming in from across the country there was no right answer. Especially at such an early stage of the COVID-19 virus, deciding confidently on protocols was an extremely difficult task. Schools across America all acted independently basing decisions off of their personal conditions. Some of those conditions being the population of the school and if they have accommodating facilities for the 2-meter distancing recommendation from the centers for disease control and prevention (CDC).

State by State also acted independently with Covid-19 related restrictions depending on how many cases were in the area. With the major impact of COVID-19 the majority of higher education was moved to a remote learning system. Harvard, Rutgers, Princeton and Georgetown Universities all adapted online learning. With these major schools having a range between 5,500-50,000 undergrad students enrolled it made sense not to even attempt to return to in-person classes.

With Seton Hill having just over 2,000 undergrad students enrolled it gave the university some flexibility on creating a conclusive decision on what steps should be taken in order to keep the lives of students as traditional and safe as possible. Seton Hill ultimately decided to have students return to in-person classes for the fall 2020 semester with an option to do remote learning for special occasions. There were several new regulations made with this decision in order to keep the students, faculty and staff as safe as possible. It was made mandatory for all students to abide by these regulations by signing a social contract. Regulations included making it mandatory to wear a mask while on campus and implementing a gathering limit of no more than 10 people when off of campus. The university has also introduced contact tracing done by the SaferMe app in an attempt to stop the rapid spread of the virus. Rules were also put in place regarding the capacity of classrooms, restrooms and the cafeteria. The cafeteria was reduced to a fraction of the traditional size however continuing to offer full menu. The athletes created a unique challenge. With the constant contact within groups mandatory screening was mandated.Table

Description automatically generated A daily screening in the form of a questionnaire and temperature checks were completed before any student athlete could participate in practice or games. For non-student athletes random covid tests were administered in order to oversee the whole student body. In an attempt to have the return to traditional in-person classes and with all of the procedures being implemented, it raises the question to whether all of these steps were enough to adequately respond to the novel Coronavirus.

**Survey**

In order to have an accurate gauge of the student bodies individual views and experiences our data analytics capstone class created a study on the Seton Hill community as to how the university handled COVID-19. This survey was not sponsored or at the request of Seton Hill University.

Our survey was approved and was given a consent as it was deemed that the research did not pose a risk to the subjects by the International Review Board (IRB).

**Initial Review**

After collecting the data, the initial review was very positive. There were 95 responses to the survey from students. The responses from the students even though they are a small sample size of our full student body it was accurate in representing our complete demographic.

Table.1

Table.1 shows how some attributes collected from the survey and how well it matches the full school’s demographic. When looking at the credits taken during the semester the majority respond with credits in-between a 12-18 range which is very standard. The residency feature is also inline, as the majority of students are on-campus residents our survey has them representing 66% of the survey leaving off-campus residents and commuters from home to represent the remaining 44% of the survey which I believe accurately represents the whole body of Seton Hill. Finally, when looking at academic year we can see how evenly spread the count is across the sophomore, junior and senior class while the first-year class is representing 31% of the data. All these factors of the small sample recorded from the survey coincide very well with the full body at Seton Hill which helps validate the survey.

**Data Cleaning**

When it comes to cleaning to data it is important to keep the data as honest and true as possible to accurately represent the analysis. It is important to not exclude any entities unless there is justification.

Upon initial review of the data there were two instances of referential integrity where two attributes contradicted each other. The first instance consisted of a student inputting that they have never been in quarantine since the fall semester in 2020, while at the same time when selecting how many times they have quarantined they inputted a selection of 1-2 times. The second occurrence had a student select that they have been in quarantine/isolation however when asked how many times they selected 0. In order to keep the data as true and accurate as possible it was my decision to remove these inputs leaving our survey to have 93 responses.

One mistake that was made when creating our survey was that an additional question was added after releasing the survey to the public. The column added was a categorical question that asked if the student believed Seton Hill University had an adequate response to COIVD-19. Because of the late addition we had 16 entries that were missing this value, this raises concern as to what should be done with these entries, elimination of the whole row would not be suitable as it would remove 16.8% of our data from our survey. I will discuss the work around used to solve this problem later in the report.

**Recoding data**

After initial review of the data, it was important to take a deeper look at some of the attributes. Using a python script, I recoded the data from a string variable to an integer to make it available to gain some descriptive statistics from our data.

With the survey data now recoded into integers it makes it possible to create some simple histograms and gain information through descriptive statistics of our main basic model questions I used a program called PSPP to acquire the descriptive statistics and create histograms, the program is very similar to the well know SPSS.

Chart, histogram

Description automatically generatedFigure.1

Figure.1 displays a histogram of the satisfaction students felt with the way topics were covered in the classroom during the COVID-19 pandemic. An important thing to notice when validating the data is that it follows the normal distribution curve with a nice arch in the middle and the tail ends of the histogram not spilling up the sides. When looking how satisfied students are with the way Seton Hill handled COVID-19 it is important to look at the mean of the histogram with a value of 1 representing least satisfied and a value of 5 representing the most satisfied, the middle point will be a mean of 3.0. Any mean above the middle point will indicate that the student body is more satisfied than dissatisfied.

Chart, histogram

Description automatically generatedFigure.2

Figure.2 demonstrates how satisfied the student body is when it comes to how Seton Hill handled mental health regarding COVID-19. A positive point is that the mean is above the middle point as a 3.2 showing that there is more of a following of satisfaction for the care of mental health. For the most part this graph follows the normal distribution line however there was a much larger group of students who felt disappointed with how mental health was considered during the pandemic. This is something that will be watched when exploring the analysis phase.

Table

Description automatically generated(Table.2)

Table.2 shows the rest of the descriptive statistics of the basic model questions. A crucial value to observe is the skewness of the data, similar to the mean it shows how the data is more susceptible to either more satisfied or less satisfied in relation to the question. With respect to how Seton Hill handled these topics we are looking for a negative value as it will show that the data is leaning towards the right which equates to higher values. This will signify that the student body is more satisfied compared to less satisfied. 4 out of the 5 questions all have a negative skew which help show that Seton Hill university adequately handled the COVID-19 pandemic.

**Overall Satisfaction ranking**

After seeing how positive the basic statistics I decided a good way to rank every entity individually would be by creating a new column in the data to track the overall experience score by summing the five main basic model questions ranked on their levels of satisfaction. With 5 representing very satisfied, the highest possible score an entity could have would be 25 and the lowest being 5. This way we will be able to gauge each individual and give them a score on how they satisfied they are with how Seton Hill handled covid-19. Using these scores will be very beneficial in the analysis phase being able to relate this number to the other categorical data I will be able to determine which groups of students where the most satisfied and vice versa.

Chart, histogram

Description automatically generated(Figure. 4)

Figure .4 exhibits the distribution graph of the calculated satisfaction score. The results came back super positive with the values following the trends from the normal distribution curve. This helps validate the use of the overall satisfaction score so that these values can be used when analyzing attributes to view how it reacts with other questions to determine if Seton Hill adequately handled the COVID-19 virus.

(Figure.5)

Chart

Description automatically generated with medium confidence

Figure 5. is a boxplot representation of the overall satisfaction score. Because of the normalized distribution of the Histogram (fig.4) I felt it was sufficient to when using these values to solve the problem of our missing data in the Adequate response column. Using the mean of 17.18 as our middle point I will be able to confidently classify if the student would have responded with a yes or no. I understand this is not a 100% accurate solution to the missing data problem. However, I felt this would be a better step than completely deleting all the useful data all because of one missing entity in 16 rows.

**Analysis**

When determining if Seton hill handled COVID-19 in a satisfactory manner most of the data to this point is indicating that it has. With almost all of the Histograms to this point showing a normalized distribution I am confident that the data collected from the survey is valid. As well as taking a closer look at some of the descriptive statistics, in Table.2 we were able to see that 4/5 basic model questions had a negative skew showing that the majority of people responding to the survey are satisfied.

Using the overall satisfaction score it was necessary to create a ranking of attributes in order to determine which have a stronger correlation. Using Orange, an open source data visualization and machine learning program, I calculated the relief F score. The relief F score is an algorithm developed in 1992 that uses a filter method approach to feature selection.

Table

Description automatically generated(Table.3)

Table 3. shows the results from the relief F score, giving me a ranking of attributes with the overall satisfaction score as my target. This helps classify my columns in order to gain insight of conditional dependencies with attributes and give me a value on attribute estimation in classification and regression.

Timeline

Description automatically generated(Figure.6)

Figure 6. shows the relationship between if the students felt an adequate response to COVID-19 and their ranking on if they believe masks are an effective means of slowing the spread of the virus. The ranks go as 1-5 with 1 being the lowest rank. When interpreting the boxplot there are a couple key points to take note of. First of all when looking at the mean the individuals who do not believe Seton Hill adequately handle covid-19 had a value of 2.37 while the individuals who did had a mean of 3.43 in relation to if they feel masks are effective. This is something to be expected as the students who don’t think masks are effective may be upset with all of the regulations the University has implemented, perhaps feeling that their freedoms have been taken away.

(Table.4)

Chart

Description automatically generated

This is validated with table 4 showing the average score of students ranking if masks are effective against COVID-19. Students who disagreed that Seton Hill respected their freedoms have much lower mask effectiveness average with ranks of 2.182 for strongly disagree and 1.929 for disagree. This is a large difference compared to students who believe that their freedoms respected with both the values of agree and strongly agree being approximately a value of 4.

Another point that was brought to my attention when looking at figure.6 was larger highlighted blue box which represents the spread of the data in between the first and third quartiles. There is a major difference between the spreads of students who felt Seton Hill had an adequate response to COVID-19. The size of the spread when students replied with a no with regards to an adequate response have a wide range of feelings to if masks are effective. This is compared to a value of yes to seton hill having an adequate response with a much smaller spread with mask effectiveness. This indicates that students who think there was an adequate response to covid are much more probable to believe that masks are an effective means of slowing the COVID-19 virus.

A picture containing text, outdoor, tree, red

Description automatically generated

Timeline

Description automatically generated(Figue.7)

Figure 7 shows another box plot in relation to if students trust that Seton Hill adequately handled Covid-19, associated with how many times the student was placed in quarantine. The values for showing how many times a student was in quarantine/isolation were ranked from 1-3, 1 signifying that they have never been placed in quarantine. 2 conveys the student has been in quarantine 1-3 times and the value of three means they have been placed in quarantine 3+ times. Right way I noticed how students who have been placed in quarantine more times have a stronger view that Seton Hill did not adequately handle COVID-19. The mean helps validate this as students who replied no had a mean of 2.11 showing that the majority who have been placed in quarantine more than once have a negative view on how COVID-19 was handled. Comparing this to students who replied yes having a mean value of 1.78, signifying that the majority of students who are happy with the actions Seton Hill took to manage COVID-19 have been placed in quarantine a lot less. This is also shown by the spread of the first and third quartile the students who believe the actions taken by Seton Hill were adequate to handle the virus are in between 1-2 while the students who believed the opposite had a spread in-between 2-3. Further proving that for students the more times you were in quarantine the more you became dissatisfied with how Seton Hill handled COVID-19. This could mean that the quarantine process that Seton Hill implemented had flaws.

I wanted to revisit and review the data in association to how students felt their mental health was cared for during the return of traditional class in fall of 2020 and spring of 2021. I speculated the best way to represent this was if students considered withdrawing from the school throughout the year.

Chart, histogram

Description automatically generated(Figure.8)

Figure 8. serves as a distribution visual of how their satisfaction levels in relation to if they considered withdrawing from the university during the fall and spring semester of 2020-2021. The values were ranked from 1-5 with 1 representing least satisfied to 5 being very satisfied. The two points I wanted to highlight were the values of 1 and 2, these two points did not follow a normalized distribution having much higher values than expected. This makes me confident in saying that I do not think Seton Hill took proper actions when considering mental health. Something that really grabbed my attention was the distribution of if students considered withdrawal from the school. From what the visual shows the only spikes where students considered withdrawal were also students that happen to feel dissatisfied with Seton Hills care for mental health. This is a red flag for the university as it indicates that because of lack of care for the student’s mental health during these hard times of COVID-19 it has cause students to consider withdrawing from the university.

Timeline

Description automatically generated with medium confidence(Figure.9)

Figure.9 was a dashboard created in Tableau to visualize other attributes from the survey using the created overall satisfaction scores. When looking at the visuals for me there was nothing surprising or out of ordinary. When looking at the academic year the overall satisfaction values are similar across the classes. The credits show a linear change but are all within a similar range. For me it makes sense the way the linear direction is going as students who take more credits in the semester are more likely to be stressed out and have a lower satisfaction score while students who have less credits are more stress free and have a better score.

The Student athlete shows a table however while there is some difference with the overall satisfaction score with student athletes and regular students this result is justified. Being a student athlete myself I understand why there may be a lower score result. I believe this is due to a zoom call that happened with the president of the school in the fall semester. Although I was not able to attend this meeting, from what I understand it left a lot of the athletes very unhappy as they felt they were being victimized. The zoom meeting was a warning explaining that if anyone broke the rules, they would lose their scholarship and would be suspended from sport. The majority of student athletes did not appreciate this warning as they felt it shouldn’t have been directed specifically to the athletes but rather encompass the entire student body. For this reason, I understand why the student athletes have a lower overall satisfaction score compared to non-student athletes. Lastly, when looking at the residency satisfaction scores there is some difference. Commuters from home have the highest score and this may reflect the idea that they are not around campus a lot and are not as affected by the rules and regulations that Seton Hill implemented. Off campus students have the lowest satisfaction score. Gaining freedom is what the majority of students are wanting to achieve when moving off of campus. When Seton Hill mandated no more than 10-people at a gathering and was enforcing it outside of campus boundaries it was viewed as an infringement of basic rights and freedoms. For this reason, I can understand why off-campus students may be less satisfied with how the school handled COVID-19

**Results**

After reviewing the survey, I was satisfied at how clean our data turned out. With the majority of all of our main basic model questions showing normalized distribution. With these basic model questions, I was able to create an overall experience score, making a new attribute available to use in connection to the other questions asked in our survey. With this it was difficult to find anything out of the ordinary, with most of our data having a justification why the this was the results. Will almost all of the data showing that Seton Hill had an adequate response to COVID-19, there was one feature that did show cause for some attention. That being the results of how students felt their mental health was cared for during the pandemic at Seton Hill. Even though the results showed a negative skew meaning the majority of students were satisfied in this area there were a large number of students who were dissatisfied with the actions taken. I furthered this by showing the results in connection with students who considered withdrawing from the school. The majority of students who considered withdrawing also felt that mental health was not considered enough by Seton Hill during the pandemic.

After presenting all of the data I can confidently say that the majority of students who attend Seton Hill felt that there was an adequate response to the COVID-19 virus with the return of traditional classes.

Citations

“COVID-19 and Your Health.” *Centers for Disease Control and Prevention*, 11 Feb. 2020, www.cdc.gov/coronavirus/2019- ncov/community/schools- childcare/decision-tool.html.

Crisp, Erin. “Leveraging Feedback Experiences in Online Learning.” *EDUCAUSE*, 1 June 2020, er.educause.edu/articles/2020/6/leveraging-feedback-experiences-in-online-learning.

“How Diverse Is Seton Hill University?” *College Factual*, 15 Apr. 2021, www.collegefactual.com/colleges/seton-hill-university/student-life/diversity.

Inside Higher Ed. “Coronavirus Roundup: More Universities Announce Plans for Largely.” *Inside Higher Ed*, 7 July 2020, www.insidehighered.com/news/2020/07/07/coronavirus-roundup-more-universities-announce-plans-largely-online-fall-terms.

Sawchuk, Stephen Catherine Gewertz. “Schools Are Retreating to Remote Learning as COVID-19 Surges. Do They Have To?” *Education Week*, 23 Feb. 2021, www.edweek.org/leadership/schools-are-retreating-to-remote-learning-as-covid-19-surges-do-they-have-to/2020/11.

“The Pandemic Pushed Universities Online. The Change Was Long Overdue.” *Harvard Business Review*, 29 Sept. 2020, hbr.org/2020/09/the-pandemic-pushed-universities-online-the-change-was-long-overdue.

Wikipedia contributors. “Relief (Feature Selection).” *Wikipedia*, 17 Jan. 2021, en.wikipedia.org/wiki/Relief\_(feature\_selection).