We’re Here to Get You There: A Statistical Analysis of Bridgewater State University’s Transit System

Abigail Adams

Follow this and additional works at: https://vc.bridgew.edu/honors_proj

Part of the Mathematics Commons, and the Statistics and Probability Commons

Recommended Citation
Copyright © 2021 Abigail Adams

This item is available as part of Virtual Commons, the open-access institutional repository of Bridgewater State University, Bridgewater, Massachusetts.
We’re Here to Get You There: A Statistical Analysis of Bridgewater State University’s Transit System

Abigail Adams

Submitted in Partial Completion of the
Requirements for Commonwealth Honors in Mathematics

Bridgewater State University

May 11, 2021

Dr. Wanchunzi Yu, Thesis Advisor
Dr. Laura K. Gross, Committee Member
Dr. Kevin Rion, Committee Member
We’re Here to Get You There: 
A Statistical Analysis of Bridgewater State University’s Transit System

Abigail Adams*    Dr. Wanchunzi Yu †

Abstract

Bridgewater State University first established its on-campus transportation service in January of 1984. While it began only running as an on-campus service for students throughout the day, the service grew to expand by offering an off-campus connection to the neighboring city of Brockton and absorbed the night service system from the campus safety team. As BSU Transit continues to grow, the organization is seeking ways to improve their overall service and better prepare their fleet and driver pool to accommodate this growth. The purpose of this research is to analyze trends among the data collected by BSU Transit and assist in making educated predictions for future trends within the organization. With the collection of about 5-10 years of data, descriptive statistics and statistical modeling were utilized to analyze important factors such as passenger counts and total mileage reporting to produce informative results for BSU Transit. Time series analysis and forecasting was implemented in order to predict future observations of passenger counts and total mileage counts, while the descriptive statistics breakdown provided the overall view of the system’s growth through the years. Several seasonal ARIMA models were fit to the data in an attempt to forecast monthly predictions for each metric of focus. Through the analysis of residual auto-correlation function (ACF), partial auto-correlation function (PACF) plots, and Ljung-Box Statistics test, all assumptions were met for each model. Normality and stationarity of the data was also checked through diagnostics in order to clarify the model’s true fit and accuracy. The 95% Confidence intervals of each model’s forecasts were constructed and utilized as well as mean squared errors to ensure accuracy of such predictions. Extensive data cleaning and organizing from former BSU Transit monthly reports was completed to achieve such results. A description of the data was also performed in order to highlight trends and understand patterns among the data, and further explain these trends recognized in the time series models.

*Dept. of Mathematics, Bridgewater State University, 24 Park Avenue, Bridgewater, MA 02325
†Dept. of Mathematics, Bridgewater State University, 24 Park Avenue, Bridgewater, MA 02325
1 Introduction

BSU Transit is a student-run on-campus transit service which serves the Bridgewater State University Campus and neighboring city of Brockton 9 months out of the year. From September to May, daily and nightly service occurs in line with the University’s academic calendar in order to accommodate students’ abilities to travel back and forth across the University’s 278 acres of land. Whether it be to and from dorm halls, catching the bus to class on a cold New England morning, or taking the Route 28 service line into the city, BSU Transit runs Monday through Friday through all hours of the day and most of the night, and strictly as a night service on weekends and holidays. The main purpose of the University’s transit system is to assist commuting students in traveling to campus from Brockton Massachusetts and allow for those on campus to feel safe in traveling from one side of campus to the other, as the alternative may sometimes result in upwards of a 20 minute walk.

1.1 BSU Transit

BSU Transit was first established on January 17th, 1984. It began and continues to operate with assistance from Brockton Area Transit (BAT). Originally funded through federal grants and yearly student fees, BSU Transit has since then worked itself into an entirely free service to students and community members with the establishment of budgeted state and federal funding, based on BSU Transit’s continued growth in service.

As a result of the CoViD-19 pandemic, which is still present among us as this paper is written, ridership and service has decreased rapidly due to the switch to online classes, less students being present on campus, and a lack of funding. Unfortunately, this also means that BSU Transit’s driver pool has been greatly affected in this process and the years to come for Transit will require tremendous rebuilding. Let it be noted that this paper will assess what a non-CoViD-19 impacted future would look like for BSU Transit based on its past, as well as describe how Transit was structured and performing just before the Spring of 2020.

BSU Transit was at its peak in ridership in the 2019 Fall semester. With Day Service running Monday-Friday 7AM-7PM alongside Night Service Sunday-Saturday 7PM-2AM, BSU Transit’s fleet was collecting upwards of 35,000 passengers in its busiest months, and a grand total of almost 115,000 riders. Day Service includes any route that travels between the hours of 7AM-7PM Monday-Friday. These routes consist of the Red, Green, and Blue lines which are strategically coordinated routes that run continuously on no specific time interval through each bus stop that falls upon that path, ensuring that all bus stops are being serviced at all times. Passengers may only be picked up and dropped off at bus stops, and all service is completely free of charge.

Night Service, while fairly similar to Day Service, does have its slight differences. Between the hours of 7PM-2AM 7 days a week, passengers may wait at bus stops on the specified Gold Line route, or they may wait at bus stops elsewhere on campus and call in for their pickup location to the dispatcher throughout the night. The purpose of a reduced line and only consistently covering the most popular bus stops is to cut back on riderless miles,
as commuter lots and academic buildings are closed during night service hours. Night Service was not adopted by BSU Transit until 2014, when the shuttle system absorbed the previously known Campus Safety Escort’s responsibilities. In years past, BSU Transit had run just until dusk, when the BSU Campus Safety Escort team would take over and transport students in vans from parking lots to dorms on campus. With the growth of BSU Transit though, it was only a matter of time before it would assume all nightly duties to ensure campus safety and convenience.

As a portion of Day Service, the Route 28 service is of utmost importance. The Route 28s, or Brockton Runs, are off-campus shuttles that travel to and from the BAT Bus Centre in the city of Brockton, MA. Initially, these runs did not make the complete trip into Brockton and instead made connections on the West Bridgewater and Brockton line. Once The Route 28 began servicing all the way into the Centre, ridership increased tremendously, and continues to increase each year. This service is utilized by both BSU commuter students and those residing in either Bridgewater or Brockton who are not students at the University. This service is also free of charge, but instead runs on a set timed schedule throughout the day, and is a “flag down” service where riders may be picked up at any point along route 28 before entering Brockton. Once the driver enters Brockton, the only point of pick up or drop off is at the BAT Center at the end of the line. This is BSU Transit’s most important route, as it is one that gives back to their partners at Brockton Area Transit. While increasing commuter interest and convenience, this service also attracts and increases ridership for BAT and their transit authority. In order to continue and increase service and funding for BSU Transit, the reporting of passengers and expanding upon their Route 28 service is crucial.

2 Data Description

The data utilized in this analysis consists of observations of all BSU Transit functions. Some of the data was collected back to the earliest year of 2011, and continues to the current times which are constantly being updated. Some of the data collected consist of each select service’s total passenger counts. For example, how many riders were collected during Day Service, how many were collected during Night Service, and more specifically how many were collected on both BSU owned vehicles and BAT owned vehicles. Other observations such as miles traveled and total number of gallons of fuel pumped are also recorded to help monitor BSU Transit’s fuel economy status and the usage of each specific vehicle, as both of these are important factors when considering BSU Transit’s need for upgrades among their fleet in order to be cost effective and environmentally conscious.

2.1 Data Source

The 2011-2020 monthly total passenger counts, monthly route 28 passenger counts, and total miles traveled per month, were all collected from BSU Transit’s Total Ridership Archive that is updated at the completion of every Spring semester at Bridgewater State
University. This master spreadsheet holds the less-detailed breakdown of every year to compile a clean and concise record of BSU Transit’s progression through the months.

The more detailed data, which describes specific vehicle passenger counts, projected mileage counts, gallons of fuel pumped, and ridership per service, is compiled into monthly reports constructed by student managers at BSU Transit from the years 2015-present. These reports are FTA (Federal Transit Authority) mandated, and are specifically structured to provide useful information to BAT and the FTA in order to receive continued funding. Mileage counts are recorded by month for every single vehicle, as well as total monthly fuel usage for each vehicle, in order to monitor the fuel economy for the fleet.

In these reports there is a projected amount of miles that the FTA expects BSU Transit to travel each month. This number is calculated by using the total amount of miles per route, and the average amount of hours each bus drives on that route based on the semester’s current schedule. This estimation is not quite accurate as it strictly utilizes estimates for what should be happening, not through the observations of previous events. This analysis will help adjust these estimates and provide more accurate predictions for BSU Transit’s future mileages, as well as passenger reports.

2.2 Data Organization

Because these data were collected in report form, a substantial amount of data cleaning and organizing was required. First the ridership and mileage archives from 2011 were exported to a general table that contained each month’s observations for all 9 years. To check for any possibilities of human error, time series plots were constructed and it was noticed in the mileage counts that in 2016, mileages were incredibly high. After cross referencing with the monthly reports it was found that whomever input the mileages for 2016 had miscalculated the total miles and this was fixed. This overall data was utilized for the time series analysis performed for passenger counts and total miles traveled.

Once this had been completed, the monthly reports dating back to 2015 were utilized to collect all expected miles and intricate details of passenger counts for those 5 years. This monthly data over the course of 5 years was utilized for descriptive statistics and data visualizations.

2.3 Descriptive Statistics and Data Visualizations

Tableau software was utilized in order to visualize service comparisons and time series of each the data categories. These graphics help describe intricate trends among the data that will help further explain the construction of time series models later in this paper.

2.3.1 Day Service

About 70% of all passengers each year are accounted for through BSU Transit’s Day Service. The busiest month on average each year is October, due to the high demand
for service during Homecoming and Halloween festivities on campus. On average, BSU Transit services about 26,000 passengers during this month. The slowest month of the year, May, which only consists of about 18 days of operation each year, averages a total of only 5,000 passengers. Plotting a time series, as seen in Figure 1, allows for the visualization of these descriptive statistics. Notice the peaks that occur in October and April each year, as well as the rapid declines that occur in January and May.

Figure 1: Total Day Service Passengers 2015 - 2020

5% of all Day service Passengers are made up of those who utilize BSU Transit’s Route 28 Brockton Runs in and out of the city. A representation of these proportions can be seen in Figure 2 below. The dark blue bars indicate the total Route 28 ridership by month, and the lighter blue bars show the total ridership for Day Service by month from 2016-2020. While this may not seem like an applicable representation, the idea is to recognize the increase in the Rt. 28 ridership makeup of Day Service over time.
2.3.2 Night Service

The remaining 30% of passengers accounted for were those who utilized Night Service. In Figure 3, a time series from 2015 - 2020 of the Nightly ridership is presented.

As seen in the time series, the busiest months for night service tend to follow Day Service trends, where October and April are the high traffic months, and January and May are consistently low. Night Service has its highest total ridership in October 2016, just 2 years after its start up, with 13,505 total passengers.

A visual representation of that 30%/70% split between Day and Night Service can be seen below. This can simply be explained by the fact that a majority of BSU Transit’s riders are going to be utilizing Day Service over Night Service, as classes are in session...
during the day hours and there is a substantial amount of foot traffic, including residents and commuters, on campus each day. Figure 4 shows the comparison of proportions for Night passengers and Day passengers. The purple bars represent the total Night Service passengers by month from 2015-2020, and the orange bars are the total Day Service Passengers.

![Figure 4: Total Night and Day Service Passengers Comparison 2015 - 2020](image)

2.3.3 Fuel Economy

Each month, student employees keep track of how many miles all individual buses travel in a week, then compile these miles for the month to monitor the usage of each vehicle. This is done in order to plan for maintenance and be diligent about distributing work throughout the fleet to avoid breakdowns. Figure 5 is a time series visualization that shows the total miles traveled by all BSU Transit vehicles from 2015 - 2020.
As pictured in the time series, the average amount of miles traveled is a total of 12,352 with a maximum of 21,688 miles in April of 2016 and a minimum of 2,835 in May 2018. This is consistent with the trends found in passenger counts. This may be an indication that 2016 was a busy year for BSU Transit, or even that BSU as a whole had an increase in the number of students attending the University, while the Spring of 2018 may have been quite the opposite.

While monitoring mileages per vehicle, BSU Transit also consistently observes each vehicle’s monthly fuel economy with miles traveled and total gallons of fuel pumped. The time series plot in Figure 6 represents the total monthly fuel economy for the years 2015 - 2020, which visibly levels out as the years progress. The large fuel economy levels in the earlier years may be explained by an upgraded fleet, where new vehicles requiring more fuel than previous vehicles were introduced. The average fuel economy for the past 5 years is 6.3 MPG, with a low of 3.44MPG in February 2017.
3 Time Series Analysis

Univariate time series data is data that contains a single variable that holds numerous values over a set amount of time. An example of this, in BSU Transit’s case, would be observing how many passengers they transport in a month over the course of several years. Time series data can be used not only to visualize simple trends over time, but to also assist in predicting future trends that haven’t occurred yet. In order to do this, there is a time series model known as an ARIMA model, or an Auto Regressive Integrated Moving Average Model, which utilizes previous observations of a data set over time to make accurate and applicable predictions for future observations.

An auto regressive integrated moving average model is a form of analysis that gauges the strength of one dependent variable relative to other changing variables. The models examine the differences between values in the series as opposed to using the actual values. An ARIMA model is made up of three components. The AR component, or auto regressive component, refers to a model that shows a changing variable that regresses on its own lagged, or previously recorded, values. The integrated, or I component, in an ARIMA model represents the differencing of observations to allow for the time series to become stationary. This means that the data values are being replaced by the difference between the current data values and the previous values. Finally, the MA component of an ARIMA model is the moving average term. This term incorporates the dependency between an observation and a residual error from a moving average model applied to previous observations.

Each term has a corresponding integer value that determines the level at which each model is applied. These integer values are represented in parameters with a standard notation of p(AR), d(I), and q(MA). Table 1 displays how to interpret each integer value.

<table>
<thead>
<tr>
<th>ARIMA Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>The lag order, or number of parameters in the model for the AR component</td>
</tr>
<tr>
<td>d</td>
<td>The number of times the raw observations are differenced</td>
</tr>
<tr>
<td>q</td>
<td>The number of parameters in the model for the MA component</td>
</tr>
</tbody>
</table>

In order to determine what terms are necessary to produce an accurate model, the original time series of the data and the ACF and PACF plots are utilized to determine seasonality and whether autoregressive terms or differencing is required. A seasonal pattern occurs when a time series is affected by seasonal factors such as the time of the year or the day of the week. Seasonality is always of a known frequency. When seasonality is present in data, the data is not stationary and therefore must undergo seasonal differencing in order to achieve stationarity, or a constant mean among the data. This can be done by adding a differencing, or I, term in the seasonal portion of a seasonal ARIMA model. For example, a model that contains only seasonal differencing would follow this order (0,0,0)x(0,1,0), and a model containing only non-seasonal differencing would follow this order (0,1,0)x(0,0,0).
An ACF plot is an auto correlation function plot, and a PACF plot is a partial auto correlation function plot. An ACF plot describes how related present values are to its past terms. The significance of lags within the confidence intervals tell us whether an auto regressive or moving average term is needed, and also indicates whether seasonal or non-seasonal differencing may be necessary. A partial auto correlation function tells us about the correlation of the observations, which remains after removing the effects which are already explained by the earlier lags, with the next lag value. So, if any hidden information lies within the residuals, we might get a good correlation to keep that next lag as a feature while modeling.

Once a model that produces no significant lags in the ACF or PACF plot is constructed, there are a number of diagnostics and tests that are performed to assess the accuracy and fit of that model to the data. Constant mean and constant variance are checked and assured true, and the model is also tested for normality and randomness among its residuals. With a model that fits the data, the next step is to apply that model and assess its ability to accurately predict observations.

When a forecast is constructed, the number of data points in your original time series will help determine the most effective length of your forecast. In BSU Transit’s case the data is collected monthly over the course of about 7 years, resulting in around 63 data points per model. In order for our results to be as accurate as possible, predicting 1 year at a time was the most effective. The ARIMA model utilizes the time series data to predict following years data, along with a 95% confidence interval to account for error in the model. This section will walk you through our model selection process for BSU Transit’s total miles traveled, total passenger counts, and route 28 passenger counts, as well as explain the results of the 1-year forecasts constructed from these models.

### 3.1 Total Miles Traveled

In order to evaluate the effect of an increase in service, or to assist in better preparing BSU Transit’s fleet for future growth, a seasonal ARIMA Model was created from the total mileage data for all vehicles collected each month over a 7 year period (2011-2018). The original time series of these 7 years is shown in Figure 7.
3.1.1 Model Selection

From the original time series shown in Figure 7, there is a positively increasing trend over time, as well as a seasonal trend present as the values peak in April and drop significantly in May each year. These observations lead to the assumption that both seasonal and non-seasonal differencing may be required in our model. Figure 8 displays the original ACF and PACF plots for the time series.

These plots helped solidify these assumptions and allowed us to conclude that our first attempted ARIMA model would be of order (0,0,0)x(0,1,0), to apply seasonal differencing first. Figure 9 displays the ACF and PACF of the differenced observations for our first attempted model.
Next, we attempted an ARIMA model with order $(0,1,0) \times (0,1,0)$ to account for non-seasonal differencing to counter the increasing trend present in the original time series. Figure 10 shows the ACF and PACF of the differenced observations for this second attempted model.

Our ACF and PACF plots carried less and less significant lags, or lags that lie past the interval denoted by the blue dashed line, so this is an indication that we are on the right track for our model. The next attempted model utilized the addition of a seasonal auto regressive term leaving us with the $(0,1,0) \times (1,1,0)$ ARIMA model. Figure 11 displays the ACF and PACF of this third attempted model’s differenced observations.
Our fourth and final model attempt was a seasonal ARIMA model of order (0,1,1)x(1,1,0) with an auto regressive term, a seasonal moving average term, and both seasonal and non-seasonal differencing. The final model’s ACF and PACF are seen below in Figure 12.

With all lags insignificant and no obvious trends present among the ACF and PACF plots, we were able to continue in confirming our model’s accuracy and checking all diagnostics.

### 3.1.2 Diagnostics

In order to assess how good of a fit our model was to the data, we utilized diagnostic plots and Ljung-Box Statistics tests. The Ljung-Box test is utilized to assess whether auto correlation exists in a time series or not. The null and alternative hypotheses for this test
are as follows:

\( H_0 \): The residuals are independently distributed
\( H_a \): The residuals are not independently distributed; they exhibit serial correlation

Ideally, we would like to fail to reject the null hypothesis. That is, we would like to see the p-value of the test be greater than 0.05 because this means the residuals for our time series model are independent, which is often an assumption we make when creating a model. In the case of our model, the Ljung-Box test returns a p-value of 0.4448. Therefore we do not reject the null hypothesis, and we can conclude that our model is a good fit by the Ljung-Box test, and that our residuals are independently distributed.

Next, some of the diagnostic plots checked were the histogram of the residuals, a plot of the residuals versus the fitted values, a normal probability plot (or normal QQ plot), and an ordered residual plot.

![Figure 13: Total Miles Traveled Final Model Diagnostic Plots](image)

The histogram of residuals can be seen in Figure 13(a). This histogram can be used to check whether the variance is normally distributed. A symmetric bell-shaped histogram
which is evenly distributed around zero indicates that the normality assumption is likely to be true. Here in Figure 13 you will notice that our histogram is in fact symmetrical which indicates normality with a little heavy tail.

The next plot, just a simple residual plot, is utilized to determine our model’s fit. Figure 13(b) shows the residual plot which appears to have no obvious patterns and therefore is an indication that our model is of constant variance.

Figure 13(c) displays the ordered residual plot. This plot allows us to visualize whether or not our model presents a constant mean. The fluctuation about the red line indicates that we do in fact have a constant mean, as there is no increasing trend as time progresses.

The final plot, Figure 13(d), is our normal probability plot. This plot should display the residuals all falling along the red trend line to indicate that our model is normal. Unfortunately, you may notice that these points do not all fall along the trend line, and they are a bit skewed at the tails. This is okay, as normality is a robust assumption and may not always be completely met when working with real data.

After checking diagnostics and ensuring our model is capable of accurately predicting future trends, we were able to confidently select the seasonal ARIMA model of order $(0,1,1)x(1,1,0)$ to utilize in predicting future miles traveled by BSU Transit.

### 3.2 Total Passenger Counts

The next time series model we constructed was one that would have the ability to predict BSU Transit’s future total passenger counts. These counts include all day service and all night service passengers. This will be helpful for BSU Transit in better preparing their driver pool and fleet to accommodate for the growth in ridership rates. Figure 14 shows the original time series for the 2011-2018 data collection of passenger counts.

![Figure 14: Total Passenger Count Time Series 2011-2018](image)
3.2.1 Model Selection

Like our total miles traveled model, Figure 14 shows an increasing trend over time, as well as seasonality. This indicates that our model may require seasonal and non-seasonal differencing, with the possible addition of auto regressive or moving average terms. Pictured in Figure 15 are the time series’ original ACF and PACF plots.

![Figure 15: Total Passengers Original ACF and PACF](image)

Presenting similar trends again as our Total Miles time series data, a number of seasonal ARIMA models were formulated in order to select the most accurately presented model. Steps very similar to those found in section 3.1.1 were performed, and we selected a seasonal ARIMA model of order \((1,1,0) \times (1,1,0)\) as our final model. The final ACF and PACF plots of this model are displayed below in Figure 16.

![Figure 16: Total Predicted Miles Final Model ACF and PACF](image)

Both the ACF and PACF plots show that the addition of seasonal and non-seasonal auto regressive terms helped achieve insignificant lags and seasonal and non-seasonal differencing assisted in providing our model with a constant mean and constant variance.
3.2.2 Diagnostics

After selecting our final model, we followed through with performing the same tests and checking the same diagnostics as with our mileage model to ensure our model was a good fit and would accurately predict future ridership.

The Ljung-Box test for this model returned a p-value of 0.5357 which is much larger than the 0.05 significance level, and therefore we do not reject the null hypothesis and we can conclude that our model’s residuals are independently distributed.

![Figure 17: Total Passengers Final Model Diagnostic Plots](image)

You will notice in Figure 17(a) that the histogram is not symmetrical about zero, and appears quite skewed. This can be explained by the addition of Night Service increasing passenger counts significantly in the year 2014. This large jump in ridership made the previous year’s data skewed. Occurrences such as this are common in data that experience large changes over time. This does not prevent our model from being accurate or a good fit.

The residual plot in Figure 17(b) presents no patterns, and is an indication of constant variance. The ordered residual plot in Figure 17(c) shows that our model has a constant
mean. The normal probability plot in Figure 17(d) is a clear indication that our model has achieved normality, as almost all of the residuals fall upon the trend line.

3.3 Route 28 Passenger Counts

The third and final model formulated for this research was created to take a closer look at BSU Transit’s ridership over time, by being able to predict future numbers for their Route 28 service. This will help benefit BSU Transit in continuing their federal funding through the FTA, and to better plan for expanding their Route 28 service, for example how many runs are performed per day in order to accommodate the increase in passengers. The time series graph for Route 28 passengers from 2012 - 2019 is shown below in Figure 18.

![Figure 18: Total Rt. 28 Passenger Count Time Series 2012-2019](image)

3.3.1 Model Selection

Like our total passenger count time series, the Route 28 ridership time series presents seasonal and increasing trends. The same steps taken in sections 3.1.1 and 3.2.1 were performed for the Route 28 model, and our final selected model was a seasonal ARIMA model with order \((1,0,0)\times(1,1,0)\) with seasonal differencing, and auto regressive terms both seasonal and non-seasonal. Figure 19 displays the original ACF and PACF plots of the time series, and Figure 20 displays our final model’s ACF and PACF plots of its residuals.
The Final Model Residual ACF and PACF plots indicate that the seasonal differencing and auto regressive terms were enough to form a model that fits our data.

3.3.2 Diagnostics

To complete our final model, we once more checked all diagnostics and performed the Ljung-Box test to ensure that our model was a good fit for our data. The Ljung-Box test returned a p-value of 0.09138 for our final model, which is larger than our 0.05 threshold, and therefore allows us to fail to reject the null hypothesis and conclude that our residuals are independently distributed.

The histogram for the Route 28 final model shown in Figure 21(a) is symmetrical about the point of zero, and therefore indicates normality. The residual plot in Figure 21(b) presents no trends or patterns to solidify the constant variance. Figure 21(c) shows the ordered residual plot of our model, and clearly shows that the model achieves a constant mean. Finally, our normal probability plot in Figure 21(d) presents non-normality among our model’s residuals, as a majority of our residuals do not fall upon the trend line. Again,
this is okay, as normality is a robust assumption and does not deter our model from being accurate.

With the checking and passing of our diagnostics, we can confidently perform forecasts for predicting future Route 28 monthly ridership trends for BSU Transit utilizing our final seasonal ARIMA (1,0,0)x(1,1,0) model.

4 Model Forecasting

4.1 Total Miles Forecast

Following the selection of seasonal ARIMA with order (0,1,1)x(1,1,0) as the final model for predicting future monthly mileages for BSU Transit, a one year forecast for the 2018-2019 year was calculated. These predictions were made with 95% confidence, and were compared to the true values of total miles of that year to ensure accuracy.

Figure 22 below shows the forecasted points for each month of the 2018-2019 school year. The grey shaded area represents the 95% confidence interval. This interval indicates
that our model can predict with 95% confidence that the true value of each specific point lies within that interval. For example, the predicted number of total miles for November 2018 is 12,654 miles, and if the true value is not 12,654 then we are 95% confident that the November 2018 total mileage lies somewhere within that shaded region.

![Figure 22: Total Predicted Miles 2018-2019](image)

Table 2 shows the predicted values for total miles in 2018-2019, the true values for the 2018-2019 year, and the difference between the real and forecasted values.

<table>
<thead>
<tr>
<th>Month</th>
<th>Predicted Miles</th>
<th>Actual Miles</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sept-18</td>
<td>12,897</td>
<td>12,055</td>
<td>-841</td>
</tr>
<tr>
<td>Oct-18</td>
<td>13,524</td>
<td>18,785</td>
<td>5,261</td>
</tr>
<tr>
<td>Nov-18</td>
<td>12,654</td>
<td>13,961</td>
<td>1,306</td>
</tr>
<tr>
<td>Dec-18</td>
<td>9,021</td>
<td>12,125</td>
<td>3,104</td>
</tr>
<tr>
<td>Jan-19</td>
<td>7,942</td>
<td>6,094</td>
<td>-1,848</td>
</tr>
<tr>
<td>Feb-19</td>
<td>8,816</td>
<td>17,364</td>
<td>8,548</td>
</tr>
<tr>
<td>Mar-19</td>
<td>11,853</td>
<td>13,202</td>
<td>1,349</td>
</tr>
<tr>
<td>Apr-19</td>
<td>14,905</td>
<td>19,529</td>
<td>4,624</td>
</tr>
<tr>
<td>May-19</td>
<td>3,702</td>
<td>6,820</td>
<td>3,117</td>
</tr>
</tbody>
</table>

A visualization of this table is represented in Figure 23 below. This graph is the same as the previous in Figure 22, with the addition of the true total miles for the 2018-2019 year represented on the red trend line.
You will notice that all but 3 of the months of the year fall within the 95% confidence interval. At first glance, this may seem to be an indication that our model is in fact not accurately representative of the data and is not a good fit, but take into consideration the reality of the data. These obscure differences in October, February, and April, can be explained. In the 2018-2019 year, Bridgewater State University had one of its largest Homecoming celebrations and the University’s athletic teams participated in some of the most competitions that year than they had before. BSU Transit is involved in supplying transportation for athletic teams for away games all over New England and are also responsible for providing transportation during events such as Homecoming, Accepted Students Days, and other event shuttles that take place for various departments of the University. The 2018-2019 year just happened to be one of BSU Transit’s busiest years in athletic trips and event shuttles, which resulted in higher than usual mileage counts in October, February, and April of this year.

Furthermore, this model is a step in the right direction for BSU Transit. Previously, the student-run organization utilized an estimated formula to calculate predicted monthly mileages from factors such as shifts per week, hours per shift, and the total estimated miles a bus would travel during those hours. This formula was unfortunately not presenting BSU Transit with results accurate enough to make reliable predictions, but the use of this time series analysis could assist BSU Transit in making better predictions for their future years of service. Figure 24 below shows the previously predicted trend line in comparison to the new predictions made by the seasonal ARIMA model.
While it may appear that the previous predictions that lie on the blue trend line are closer in accuracy than the model’s prediction, let Table 3 be a clear indication that this is in fact not true by showing that the difference in the previous predictions and the true values are much greater than the difference between the model’s prediction and the true values.

Table 3: Total Miles Prediction Comparison 2018-2019

<table>
<thead>
<tr>
<th>Month</th>
<th>Previous Prediction Difference</th>
<th>Model Prediction Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sept-18</td>
<td>6,268</td>
<td>-841</td>
</tr>
<tr>
<td>Oct-18</td>
<td>9,932</td>
<td>5,261</td>
</tr>
<tr>
<td>Nov-18</td>
<td>6,658</td>
<td>1,306</td>
</tr>
<tr>
<td>Dec-18</td>
<td>2,513</td>
<td>3,104</td>
</tr>
<tr>
<td>Jan-19</td>
<td>5,989</td>
<td>-1,848</td>
</tr>
<tr>
<td>Feb-19</td>
<td>5,099</td>
<td>8,548</td>
</tr>
<tr>
<td>Mar-19</td>
<td>7,379</td>
<td>1,349</td>
</tr>
<tr>
<td>Apr-19</td>
<td>7,626</td>
<td>4,624</td>
</tr>
<tr>
<td>May-19</td>
<td>3,179</td>
<td>3,117</td>
</tr>
</tbody>
</table>

It is clear here that the previous predictions were significantly inaccurate a majority of the time, while the model only presented issues during those months previously explained. Overall, the model selected poses as an accurate model to utilize in making future predictions for the total miles traveled each month by BSU Transit.

4.2 Total Passengers Forecast

The final model chosen for predicting BSU Transit’s total monthly passenger count was a seasonal ARIMA model with order $(1,1,0)(1,1,0)$. This model was utilized to accurately
predict the total monthly ridership for the 2018-2019 year. These predictions were also made with 95% confidence, and were compared to the true values of total passengers of that year to ensure accuracy. Figure 25 shows the forecasted ridership along with the 95% confidence interval.

![Figure 25: Total Predicted Passenger Counts 2018-2019](image)

Table 4 shows the true number of riders per month for the 2018-2019 year along with the predicted values and the difference among them.

<table>
<thead>
<tr>
<th>Month</th>
<th>Predicted Ridership</th>
<th>Actual Ridership</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sept-18</td>
<td>23,301</td>
<td>28,180</td>
<td>4,879</td>
</tr>
<tr>
<td>Oct-18</td>
<td>30,551</td>
<td>39,115</td>
<td>8,564</td>
</tr>
<tr>
<td>Nov-18</td>
<td>26,724</td>
<td>36,134</td>
<td>9,410</td>
</tr>
<tr>
<td>Dec-18</td>
<td>20,544</td>
<td>21,004</td>
<td>460</td>
</tr>
<tr>
<td>Jan-19</td>
<td>18,896</td>
<td>16,677</td>
<td>-2,219</td>
</tr>
<tr>
<td>Feb-19</td>
<td>31,442</td>
<td>39,263</td>
<td>7,821</td>
</tr>
<tr>
<td>Mar-19</td>
<td>27,869</td>
<td>30,779</td>
<td>2,910</td>
</tr>
<tr>
<td>Apr-19</td>
<td>33,247</td>
<td>41,085</td>
<td>7,838</td>
</tr>
<tr>
<td>May-19</td>
<td>2,718</td>
<td>11,401</td>
<td>8,683</td>
</tr>
</tbody>
</table>

You may notice that the difference in the predictions and actual passenger counts are substantially higher than the difference in mileages. This can be simply explained by the range in observations themselves. Ridership numbers can be upwards of 41,000 passengers in a month, like in April of 2019, whereas the highest actual mileage count was only about 18,000. This results in larger differences, which are proportionally similar to the differences from the mileage model.
To assess the accuracy of the passenger count model, the forecasted values were plotted with the actual values. Here we check to see if the actual values lie on the predicted trend line, or within the 95% confidence interval.

Figure 26: Total Predicted Passenger Counts vs. Actual Passenger Counts 2018-2019

In Figure 26, all of the true ridership values for all of the months of the year either lie right on the forecasted trend line or within the confidence interval. This is an obvious indication that the model selected was in fact a good fit for the data, and the application of this model may be further used to accurately predict future years of total monthly passenger counts for BSU Transit.

4.3 Route 28 Passengers Forecast

After concluding our model selection process for the Route 28 Passenger Counts, we decided upon a seasonal ARIMA with order (1,0,0)x(1,1,0) as our final model. This model was implemented to forecast the total monthly Route 28 ridership for the 2019-2020 year. These predictions were also made with 95% confidence, and were compared to the true values of total Rt. 28 passengers of that year to ensure accuracy. The predictions assess all 9 months of the semester, but only January and February of the spring semester are included in the accuracy assessment, as there are no passenger counts for the months of March, April, or May due to the CoViD-19 pandemic. Figure 27 shows the forecasted ridership along with the 95% confidence interval.
Table 5 lists the predicted values, the actual values, and their differences.

Table 5: Total Rt.28 Riders Predicted vs. Actual

<table>
<thead>
<tr>
<th>Month</th>
<th>Predicted Rt.28 Ridership</th>
<th>Actual Rt.28 Ridership</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sept-18</td>
<td>1,466</td>
<td>1,380</td>
<td>-86</td>
</tr>
<tr>
<td>Oct-18</td>
<td>1,712</td>
<td>1,785</td>
<td>73</td>
</tr>
<tr>
<td>Nov-18</td>
<td>1,469</td>
<td>1,518</td>
<td>49</td>
</tr>
<tr>
<td>Dec-18</td>
<td>827</td>
<td>824</td>
<td>-3</td>
</tr>
<tr>
<td>Jan-19</td>
<td>532</td>
<td>651</td>
<td>119</td>
</tr>
<tr>
<td>Feb-19</td>
<td>1,321</td>
<td>1,756</td>
<td>435</td>
</tr>
</tbody>
</table>

You will notice that these differences are significantly small at some points. This is a good indication that our model is a good fit and is accurately predicting ridership for the Route 28 service. This can be further assessed by plotting the true ridership for 2019-2020 and comparing that to the forecasted values formulated from our model.
Figure 28: Predicted Rt.28 Passenger Counts vs. Actual Rt.28 Passenger Counts 2019-2020

Figure 28 allows us to visualize just how closely these predictions were to the actual values recorded for Rt.28 passengers that year. This again indicates that the model selected is capable of forecasting accurate monthly results. This model can be utilized by BSU Transit in the future to predict Route 28 ridership trends and observe BSU Transit’s continued growth.

5 Summary

To conclude, the three final seasonal ARIMA models chosen for BSU Transit’s are as follows:

- Total Miles Traveled: SARIMA (0,1,1)x(1,1,0)
- Total Passenger Counts: SARIMA (1,1,0)x(1,1,0)
- Total Rt. 28 Passenger Counts: SARIMA (1,0,0)x(1,1,0)

Each model was successful in accurately predicting either 2018-2019 monthly observations, or the 2019-2020 year observations. This indicates that our models could accurately predict future years going forward. Unfortunately, the CoViD-19 pandemic has had a great effect on BSU Transit’s ability to run as a daily campus function. Ridership has decreased by over 90%, where the largest monthly passenger count for this current year has not been much over 1,000 total passengers. This will very clearly have an effect on BSU Transit’s ability to utilize these SARIMA models just yet. It is expected that the 2021-2022 year will consist of significant rebuilding and restructure of BSU’s student-run transit authority, and these model predictions will not be applicable to this reconstructive year. We can confidently say though that going forward, as Bridgewater State returns to full capacity on campus with in-person classes, events, and more as the pandemic becomes what used to be, passenger counts and total traveled miles will continue to grow and surpass limits before.
The goal of this research was to provide BSU Transit with the ability to equip their fleet with vehicles large enough to account for the capacity in which they would be transporting, as well as prepare for future maintenance and fleet upgrades with the rise in total mileages per month. With the applications of these two models, BSU Transit may consider this growth when making large budget decisions to either expand or upgrade their fleet, which they did not have before. They may also utilize the total miles predictions for future monthly reports in order to present more accurate predictions than their current estimations. This could provide the FTA with the ability to see that BSU Transit’s mileages are on track and growing over time with the expansion of service.

The Route 28 passenger model predictions will be exceptionally beneficial in providing the FTA and BAT with more reason to continue funding to BSU Transit and assist in expanding their service. This may include providing upgraded vehicles and allowing for more Route 28s to run throughout the day, to provide more service to the community and offer more money making opportunities for BSU Transit’s student workers.

Furthermore, the visible growth of BSU Transit over time in both the forecasts and descriptive statistics provides the University with the confidence that investing in the development of BSU Transit will be beneficial to the student workers, the overall campus community, commuter students, and the town of Bridgewater and the city of Brockton. BSU Transit provides incredible opportunities for students to experience work, as well as management experience in their college career. The evolvement of this campus transit authority will encourage the University to continue providing jobs for their ambitious students and giving back to these hard working students by offering higher pay, and other student benefits at Bridgewater State University. With the support from Brockton Area Transit, the Federal Transit Authority, and the University as a whole, Bridgewater State University’s Transit system will continue to exceed all expectations and they may now have the grounds to sustain that assumption through the use of statistics.
6 Appendix

Figure 29: Original Total Miles Traveled Time Series with 1-Year Forecast

Figure 30: Original Total Passenger Time Series with 1-Year Forecast
7 References


*All Bridgewater State University Transit Archives*