SCHEDULED TRIP DEMAND FORECASTING FY 2025 – FY 2034

PREPARED FOR

ACCESS SERVICES

Date
April 10, 2024

hollingsworthconsulting.net

Table of Contents

1. Executive Summary

The scheduled trip demand and the number of new applicants drive the need for funding and resources at Access Services. Reliable forecasts and projections are essential to plan sufficient budgets and adequate operational resources to fully fund the expected demand as required by ADA regulations. Through an initial analysis, followed by training and testing, the best fit forecasting models were selected and used to develop projections for fiscal years 2025 through 2034.

The initial analysis of the essential factors led to several discoveries. The analysis of scheduled trip demand revealed a distinct difference in trends before and after the COVID-19 pandemic. In other words, the number of trip requests prior to the pandemic lacks relevance to explain the current demand, which limits the value of this historic data. The post-pandemic data set begins April 2020, limiting the size of the data set available to predict future values. The same is true for the number of new applicants.

Training and testing of different models were performed to identify and select the best fit. The hold-out method for training and testing typically splits historical data sets into two groups with 70% for training the model and 30% for testing it. Our data set split ranged from 91% training and 9% testing to 72% for training and 28% for testing. Testing several potential forecasting models led to findings including:

- The Long Short-Term Memory (LSTM) model produces the most accurate projections for scheduled trip demand in the testing period with a Mean Adjusted Percentage Error (MAPE) of 0.93% while the MAPE of previous forecasts were 7.31% for the same testing period.
- The ARIMA model best fits the new applicant data.
- Previous forecasts tended to under forecast, a negative bias, for scheduled trip demand, which resulted in significant practical errors for scheduled trip demand because they were too low.

The LSTM model was applied using a historical data set, January 2021 through October 2023, to develop projections for scheduled trip demand from November 2023 through June 2031. Linear regression was merged with the LSTM model forming the hybrid model used to create scheduled trip demand for the remaining three years of the ten (10) year forecast. The ARIMA model used the historical data set from April 2020 through November 2023 to develop projections of new applicants for the entire ten (10) year forecast (Fiscal Years 2025 through 2034).

Occasionally, scheduled trips are canceled by the eligible rider, or the eligible rider is a no-show. An analysis of the completed trip ratio indicates that 80.21% of scheduled trips are completed and this ratio is relatively constant.

Training and testing results demonstrate the LSTM model performed better than previous forecasts in the testing phase and is likely to outperform previous forecasts. The accuracy of projections for scheduled trip demand is expected to further represent improved reliable data forecasting as the number of observations in the post-pandemic data set increases. Additional steps for the next iteration of forecasting include testing additional detrending techniques and transformers, as well as considering additional forecasting models such as LSTM multivariate and the Meta (Facebook) Prophet.

2. Introduction

Access Services provides Americans with Disabilities Act (ADA) mandated paratransit service for eligible persons in Los Angeles County, California. Its services are available to any location within three quarters of a mile of any public bus fixed route and the same distance around Metro rail stations during its operating hours. Its service area is divided into six regions and extends into portions of the surrounding counties of San Bernardino, Orange, and Ventura. Independent ridership estimates are necessary to fully fund the expected ADA paratransit demand. Reliable forecasts and projections are essential to plan sufficient budgets and adequate operational resources for these critical services.

The first step to developing the projections includes an initial analysis, a peer review, and an assessment of pandemic effects. It is necessary to understand the state of paratransit, internally and nationally, to guide the course of projections for scheduled trip demand and new applicants.

The second step is identifying potential tools and models to create forecasts. Training and testing prospective models reveal the best choice to select for projections. The selected model(s) are then applied to develop forecasts of scheduled trip demand and new applicants. The third step is an evaluation to understand risks associated with the forecasted values and the confidence in utilizing the values. The final step of the approach includes reflection on the steps, and evaluation of them, to identify ways to improve the forecasting model for projections in the next iteration.

2.1. Purpose

Previous paratransit demand forecasting relied on historical data that involved scientific analysis and review. Under predictable circumstances that ebb and flow over time, traditional projections that utilize trends and shifts in scheduled trip demand were relatively effective. However, the COVID-19 global pandemic that occurred in March 2020 disrupted the trends and shifts causing them to be less predictable and continue to be such.

While pre-pandemic trends and shifts in demand may return over time, the need for a more robust and dynamic forecasting approach has emerged. The purpose of this project is to develop an approach for training and testing prospective models that reveal the best choice(s) to select for projections. Doing so will enhance the reliability and sustainability of the projections used to make critical resource and budget decisions required to serve the paratransit riders, their families, and the region.

Ongoing evaluation of the applied models along with their values for scheduled trip demand and new applicants provide for understanding of risks associated with the forecasted values and the confidence in utilizing these values. Reflection and evaluation of the approach creates a continuous improvement cycle of the forecasting model for projections in future iterations, increasing both their relevance and precision. Future iterations of forecasting models that utilize machine learning and advanced time series models better position Access Services to test additional or alternative factors, further enhancing the utility of projections and return on investment.

3. Scheduled Trip Demand

3.1. Initial Analysis – Scheduled Trip Demand

Patterns of the past help plan for the uncertainty of the future. Essential factors that provide insight into the paratransit needs in Los Angeles County, California, include the scheduled trip demand. A review of peer paratransit services also provides an opportunity for insights on services in other regions of the country.

The focus of the initial analysis includes a historical analysis on scheduled trip demand, a brief evaluation of the global pandemic effect, and a peer review. Together, the historical analysis, pandemic evaluation, and peer review provide direction for the types of forecasting models and the variables to include (exclude), consider and evaluate.

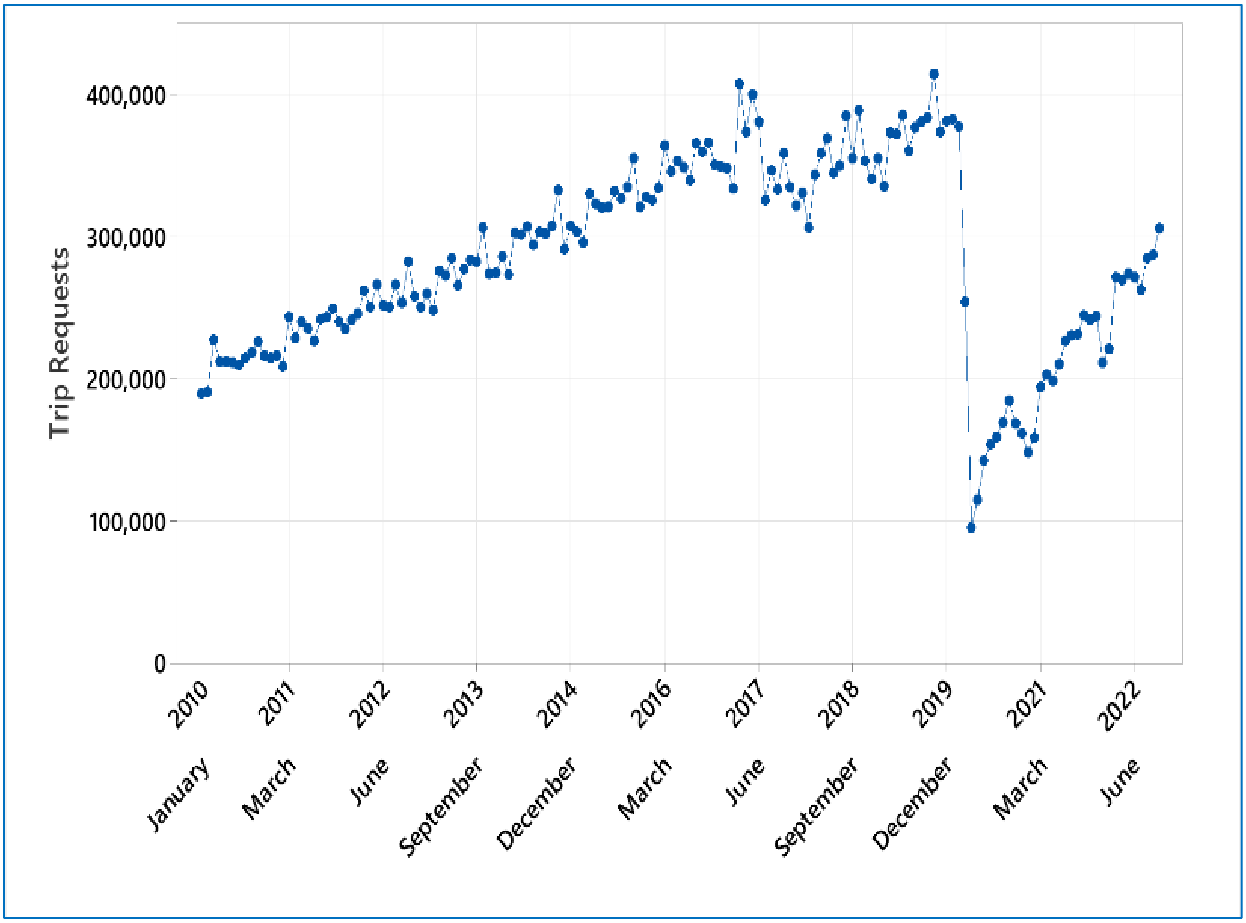
3.1.1. Historical Analysis – Scheduled Trip Demand

The examination of evidence from the past helps form a more coherent story. The focus of this examination includes an analysis of events in time series to identify patterns, trends, and changes over time. The analysis identifies the presence of (or lack of) seasonal patterns, cyclical patterns, stationarity, and autocorrelation along with trends. These components are key for model identification and selection.

Access Services needs to understand both ridership and the number of eligible customers to develop an accurate, effective budget and plan future fiscal year(s). The historical analysis includes scheduled trip demand and the number of new applicants.

The initial data for the historical analysis of scheduled trip demand includes the number of trip requests from January 2010 through October 2022. The data file provided to Hollingworth Consulting included the count of trip requests per month for each service region and the count of trip requests per month system wide. Visualization was the first step to begin to understand this variable. The time series plot for the monthly scheduled trip demand is shown below in Figure 1.

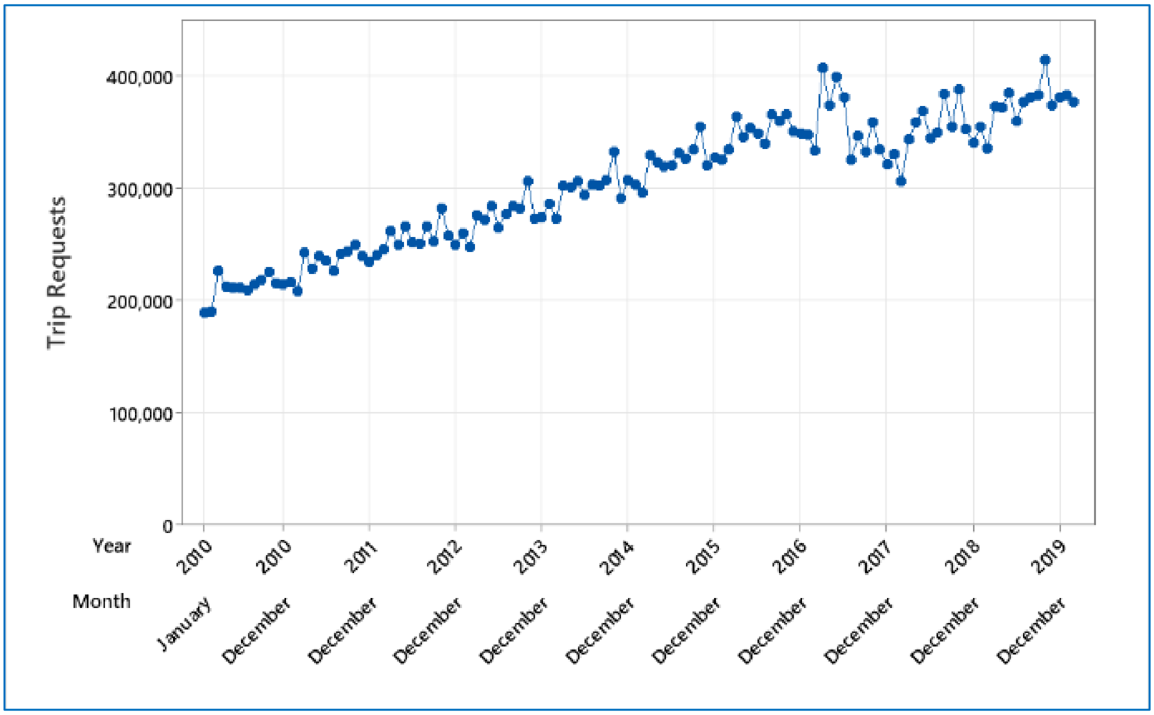
Figure 1 Monthly Scheduled Trip Demand Time Series Plot

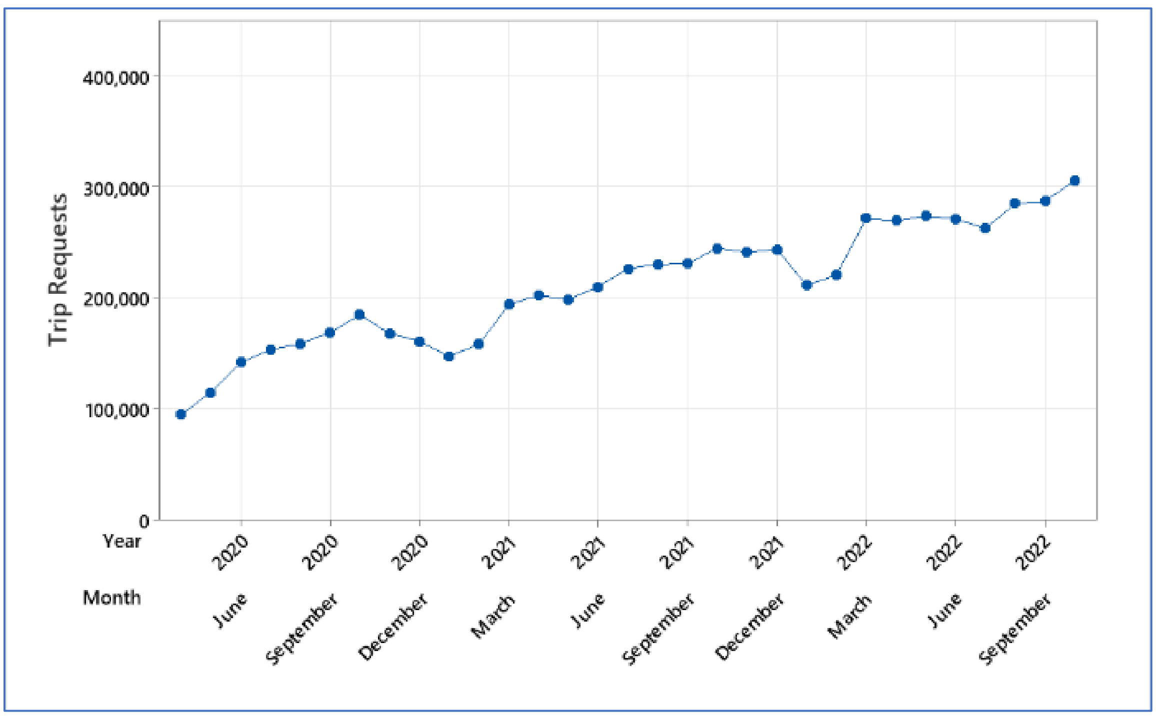


The graph reveals a story with a twist. There is a 75% drop in the number of trip requests per month from February 2020 to April 2020, equal to 281,522 fewer scheduled trips per month. The scheduled trip demand does not return to calendar year 2019 levels until October 2022.

The intercept and the slope of the scheduled trip demand appear to be different comparing February 2020 and the months prior to it with April 2020 and the months following. The time series plot shown in Figure 2 displays monthly scheduled trip demand prior to the pandemic above the same metric after the pandemic began through October 2022.

Figure 2 Monthly Scheduled Trip Demand Time Series Plot – Pre and Post-pandemic





Visual inspection of these two Figures illustrates a clear difference in the observed values. There are only two (2) observations in the plot on the top in Figure 2 below the value of 200,000 trip requests while there are thirteen (13) observations below the same amount shown on the bottom. The slope of the lines in each figure looks different, however a trend analysis is necessary to identify, quantify, and explain the validity of any differences. Further discussion and details of this topic will be addressed later in this report.

Using the trend analysis tool in Minitab Statistical Software (version 21.4.2) to examine the monthly scheduled trip demand, four (4) different trend models were identified and utilized to compare time series data and determine the general trend model that best fit the observations. The four trend models include: (1) linear, (2) quadratic, (3) exponential growth (or decay), and (4) the S-curve. The tool calculates three metrics to identify and choose the model that fits best: (1) Mean Absolute Percent Error (MAPE), (2) Mean Absolute Deviation (MAD), and (3) Mean Standard Deviation (MSD). The definitions and equations for the metrics are shown in Appendix A-1: Definitions. The lower the value for the metric the better the observations fit the model compared to the other models.

The quadratic model is the best fit of a general trend model for the monthly scheduled trip demand prior to the pandemic as shown in Figure 3. The quadratic model scored a lower MAPE, MAD, and MSD than the linear and exponential growth (or decay) trend models while tying the S-curve model for MAPE but outperforming it on MAD and MSD as shown in Table 1.

Figure 3 Trend Analysis Monthly Scheduled Trip Demand – Pre-Pandemic

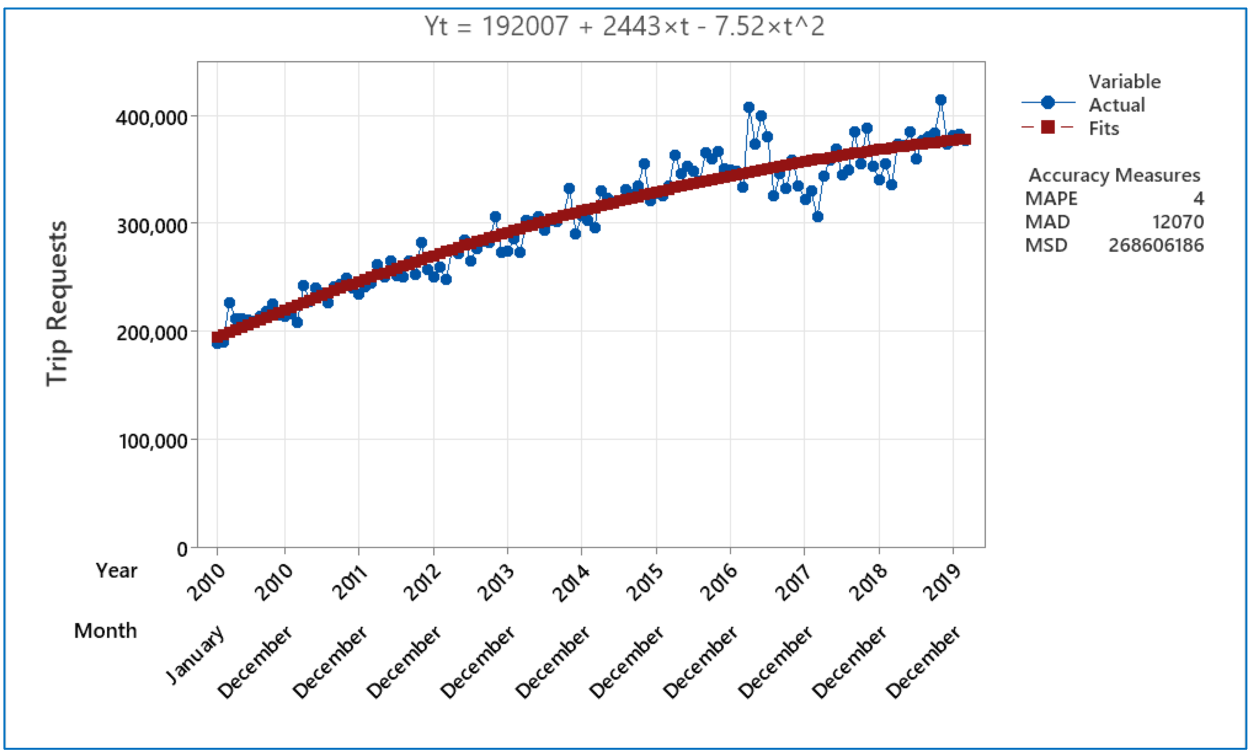


Table 1 Pre-Pandemic Scheduled Trip Demand Trend Model Scores

TREND MODEL	MAPE	MAD	MSD
Linear	5	14,292	338,159,170
Quadratic	4	12,070	268,606,186
Exponential Growth (Decay)	5	16,766	438,473,616
S-Curve	4	12,370	271,225,954

The quadratic model is the best fit of a general trend model for the monthly scheduled trip demand after the pandemic as shown in Figure 4. The quadratic model tied the linear and S-curve models for MAPE but scored (better) on all three other models for both MAD and MSD for this data set as shown in Table 2.

Figure 4 Trend Analysis Monthly Scheduled Trip Demand – Post-Pandemic

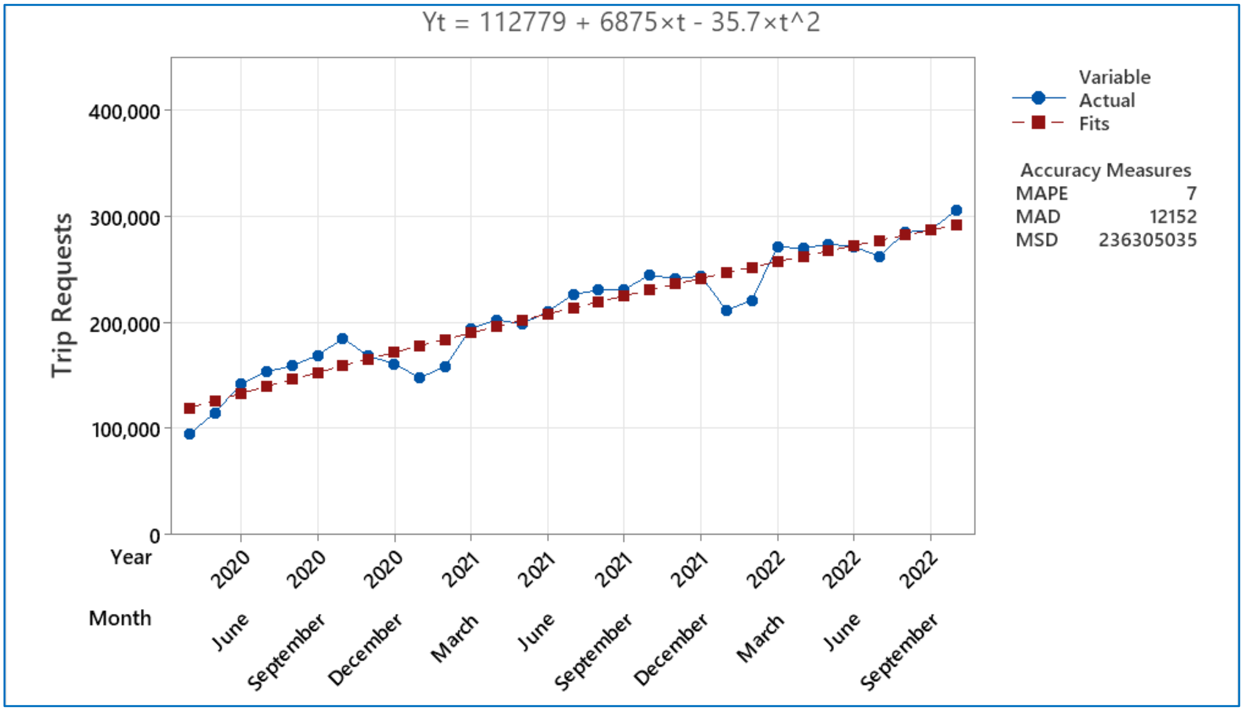


Table 2 Post-Pandemic Scheduled Trip Demand Trend Model Scores

TREND MODEL	MAPE	MAD	MSD
Linear	7	12,637	242,802,185
Quadratic	7	12,152	236,305,035
Exponential Growth (Decay)	8	15,636	317,258,470
S-Curve	7	13,027	283,597,201

Figure 3 illustrates the polynomial equation representing the trend of the monthly scheduled trip demand prior to the pandemic, $Y_t = 192{,}007 + 2{,}443t - 7.52t^2$. The equation illustrated in Figure 4, Y_t =

$112{,}779 + 6{,}875t - 35.2t^2$, represents the trend post-pandemic. The equations in Figure 3 and Figure 4 are different indicating there is a mathematical difference between the trend for the monthly scheduled trip demand before the pandemic and after the pandemic. This is important because pre-pandemic scheduled trip demand does not reflect current scheduled trip demand limiting its value to forecast scheduled trips.

3.1.2. Global Pandemic Effect

A new disease, COVID-19 (coronavirus disease 2019), spread worldwide causing a global pandemic. The World Health Organization (WHO) declared the virus a pandemic on March 11, 2020. State shutdowns began in March 2020 to prevent the spread of the virus. The virus and the shutdowns along with risk mitigation tactics such as requiring face coverings and social distancing changed the behavior, choices, and consumption of most people in the United States.

"The COVID-19 pandemic has disrupted everything from consumer behavior to supply chains, and the economic fallout is causing further changes," reported Sara Brown in a webinar hosted by MIT Sloan Management Review in January 2021 (para 1).

Several data sources document the disruption and change in behavior patterns that continue across the United States with uncertainty about whether and when pre-pandemic patterns will re-emerge. One example is retail sales, which increased as much as 30% (inflation adjusted) from Quarter 4 in 2019 to Quarter 4 of 2022. Restaurant visits in 2022 were down 12.2% compared to 2019, while grocery store visits increased 5% during the same period. Telehealth, an alternative to in-person healthcare, is double the pre-pandemic levels for Medicare recipients (Gilbert, et al., 2022).

"Pre-pandemic data is now unreliable, or even obsolete in predicting new trends," according to Seddik Cherif, Strategic Insights Manager at Google (2021, para 1). Examples from *The Washington Post* provide evidence to support this theory (Gilbert, et al., 2022).

"The simplest predictive model is what happened yesterday," Jeffrey Camm, a professor and associate dean of business analytics at Wake Forest University, posits. "That's what we're going to use to predict what's going to happen today" (Camm, 2020). The pandemic changed the paradigm for utilizing historical data to the point where pre-pandemic data only provides value in certain, limited context (Ivy Professional School, 2022). Further, there is evidence the pandemic changed the dynamics of demand at this time. However, it is plausible the demand dynamics will revert to original values and patterns (Ahmed & Sarkodie, 2021).

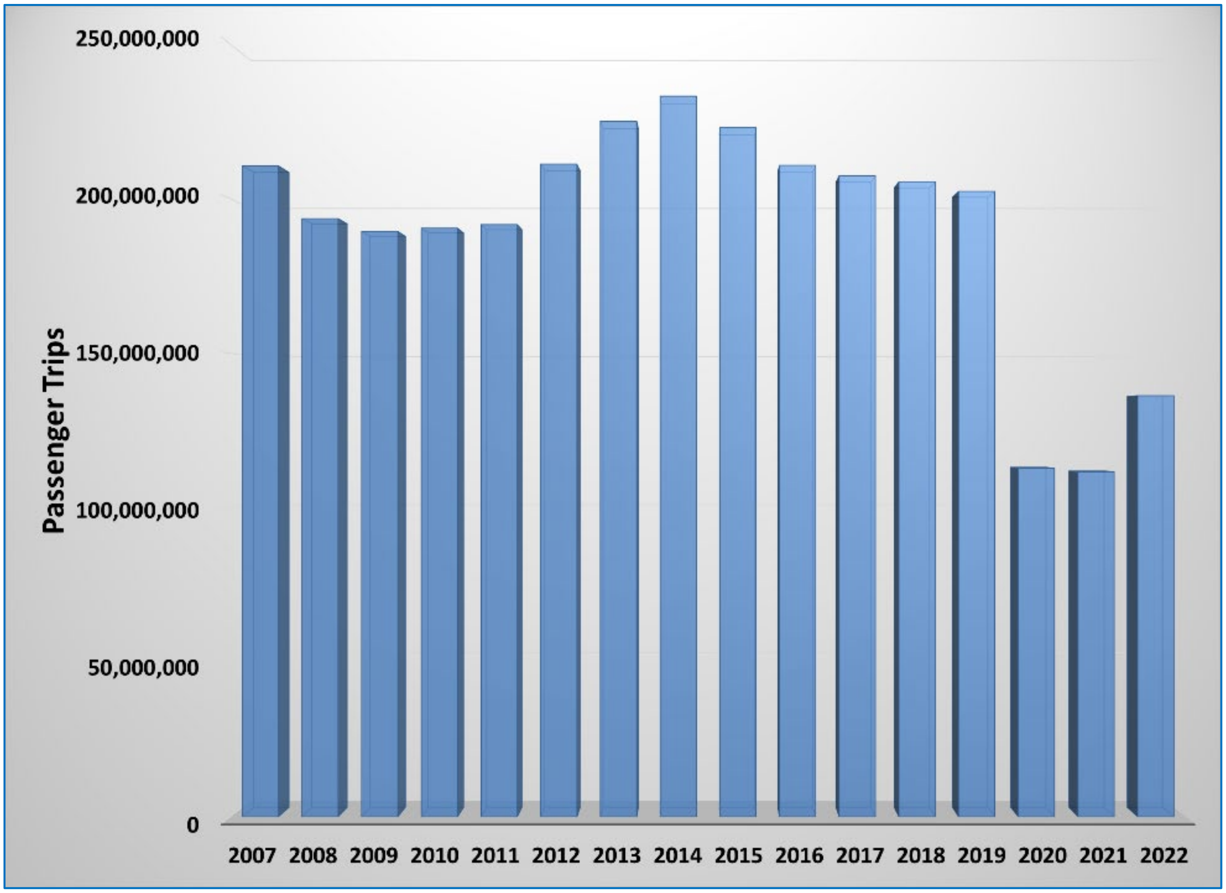
3.1.3. Peer Review

It is always important to understand practices and the trends that result or occur. This project includes a brief peer review that considers an overview of national trends and practices along with a review of comparable agencies.

The focus of the peer review is the demand response mode of service that reflects the Access Services model. The data analysis and comparisons in the review are based on calendar year data and quarterly data from two primary sources, the American Public Transportation Association (APTA) and the Florida Transit Information System (FTIS). These sources primarily use the National Transit Database (NTD) to collect and compile data. These sources provide the calendar year data up through 2022 limiting the ability to analyze annual post-pandemic trends and patterns.

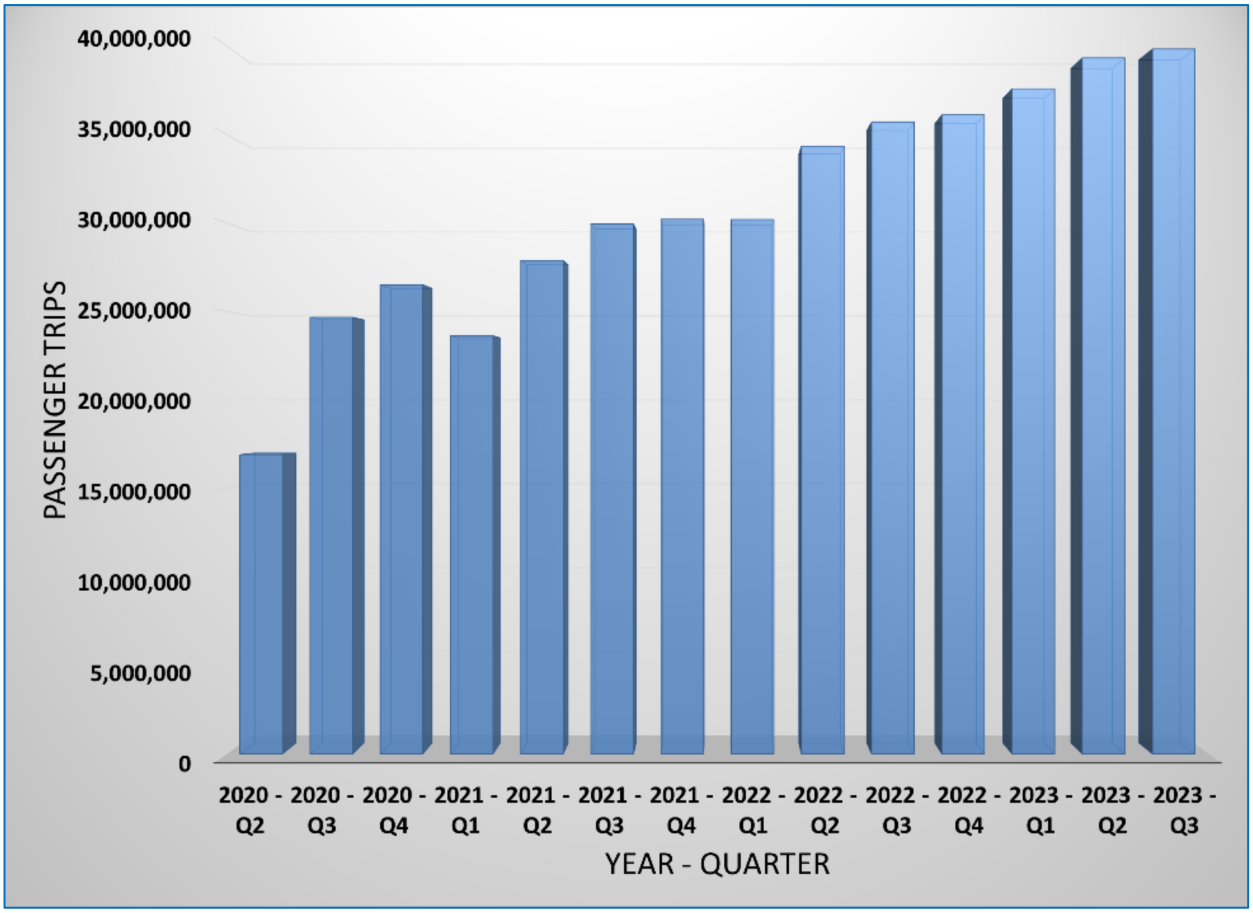
The APTA ridership report calculates ridership based upon the number of unlinked passenger trips. The bar graph in Figure 5 shows the national annual ridership for the most recent 16 years.

Figure 5 National Annual Ridership – Demand Response Model



The national annual ridership does not reveal specific patterns or trends other than the significant decrease in ridership after the COVID-19 pandemic. The APTA also compiles national quarterly ridership data. Figure 6 displays the national quarterly ridership for the post-pandemic period.

Figure 6 National Quarterly Ridership – Demand Response Mode



The national quarterly ridership indicates a clear upward trend in the number of demand response passengers for the post-pandemic period. The trend analysis tool in Minitab Statistical Software (version 21.4.2) was used to examine the national quarterly ridership. The analysis, shown in Appendix A-2: National Quarterly Ridership Trend Analysis, identified quadratic model as the best fit of a general trend model. The quadratic model is the same general model also identified as the best fit for Access Services scheduled trip demand. Access Services is experiencing a similar trend of increasing passenger trips as the national post-pandemic trend.

The Florida Transit Information System (FTIS) uses a likeness score to identify similar transit agencies, peers. Population, service type, and percent demand response, are among the factors used for the calculations found in the Guide to FTIS Peer Selection, https://ftis.org/iNTD-Urban/quickguidev2.0.pdf. The FTIS Peer Selection tool helped identify six (6) peers[1] of Access Services:

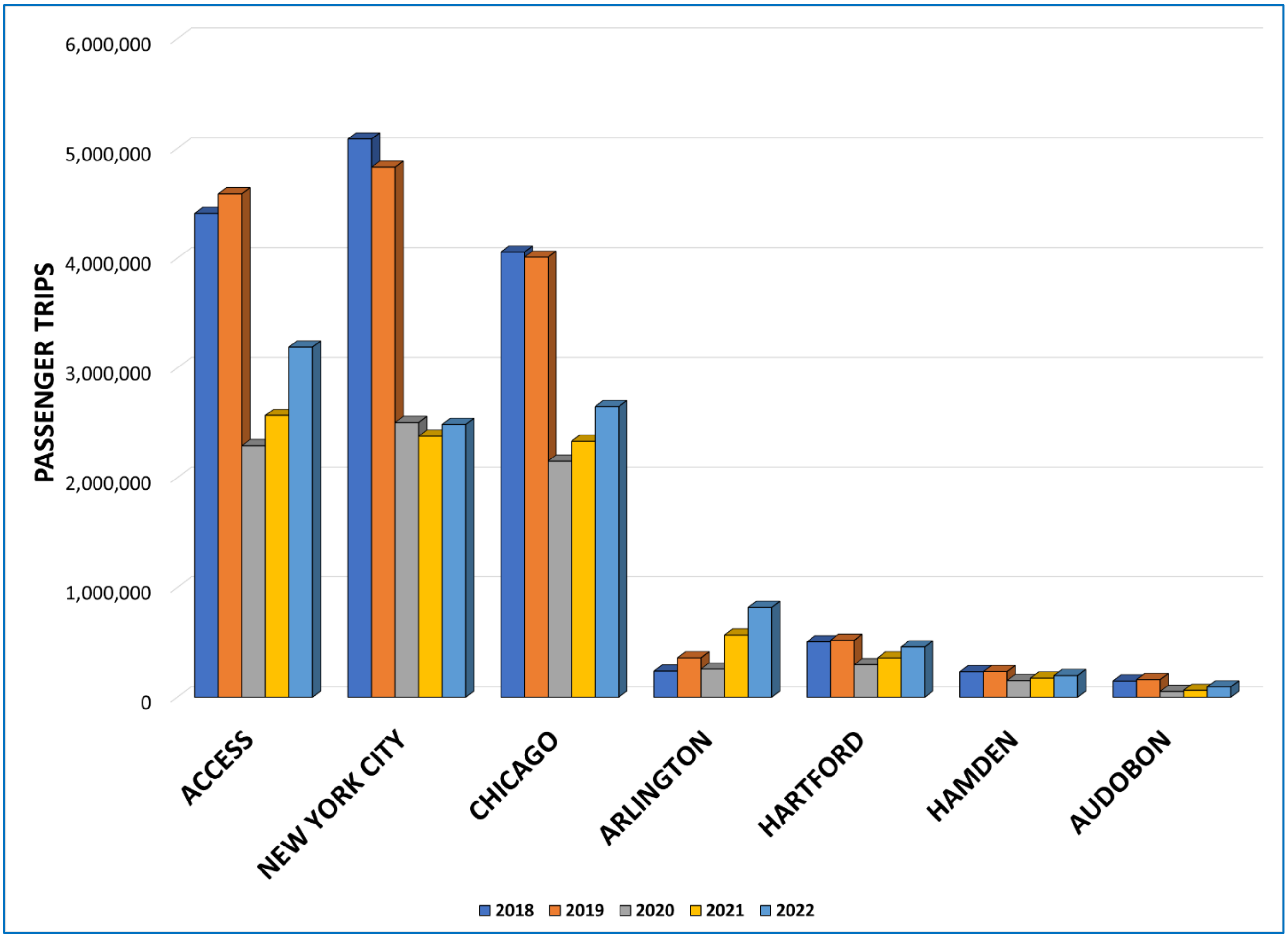
- MTA New York City Transit, New York, NY (New York City)
- Pace-Suburban Bus Division, ADA Paratransit Services, Arlington Heights, IL (Chicago)
- City of Arlington, Arlington, TX (Arlington)

[1] All six peers have identical scores to Access Services of 1.0 for the "Percent Service Demand Response" in the FTIS Peer Selection tool.

- Greater Hartford Transit District, Hartford, CT (Hartford)
- Greater New Haven Transit District, Hamden, CT (Hamden)
- Senior Citizens United Community Services of Camden County, Inc., Audubon, NJ (Audubon)

The Complete Monthly Ridership (with adjustments and estimates) report in the National Transit Database (NTD) was used to collect the annual ridership to compare Access Services with its peers. The annual ridership, passenger trips, are shown in Figure 7 for five (5) calendar years, 2018 through 2022.

Figure 7 Peer Comparison – Annual Ridership 2018-2022



Five of the six peers experienced a similar decline in annual ridership as Access Services at the onset of the pandemic. All peers experienced a similar increase in annual demand response ridership post-pandemic except New York City.

FTIS provides reports to compare peers on both efficiency and effectiveness measures. Table 3 shows the operating expense per passenger trip and the passenger trips per revenue hour for Access Services and its peers in 2022.

Table 3 Peer Comparison – Efficiency and Effectiveness

	MEASURES	ACCESS	NYC	CHICAGO	ARLINGTON	HARTFORD	HAMDEN	AUDOBON
EFFICIENCY	Operating Expense Per Passenger Trip	$57.16	$188.00	$ 64.57	$ 15.53	$ 46.60	$ 61.03	$ 41.09
EFFECTIVENESS	Passenger Trips Per Revenue Hour	1.86	1.20	1.66	3.45	1.85	1.56	1.95

A brief analysis of Table 3 indicates Access Services' operating expenses are lower than the average of its peers[2]. Access Services' passenger trips per revenue hour is greater than the median value of its peers indicating its effectiveness is on par with peers or better than the peers.

3.1.4. Initial Analysis Summary – Scheduled Trip Demand

The historical analysis reveals insights for the scheduled trip demand and paratransit ridership nationwide. The results provide guidance on the types of models to select, and the periods of data utilized to forecast these variables.

The results document that there is a difference in trends for scheduled trip demand pre-pandemic and post-pandemic. The pandemic disrupted consumer behavior including paratransit trip requests. A review of the results of the historical analysis with Access Services validated the findings based on agency knowledge and experience with the day-to-day operations.

The pre-pandemic data does not reflect current behavior; post-pandemic data captures the altered dynamics more accurately. Forecasting models need to use post-pandemic data to ensure relevance to the current economic, social, and environmental context and to generate projections.

There is limited data available post-pandemic, 44 months for this project (April 2020 through November 2023). There is uncertainty about whether, or when, pre-pandemic patterns will re-emerge. Models to forecast scheduled trip demand need to have the capacity to predict accurately, with a limited amount of historical data, and be able to respond quickly to changes given the anticipated uncertainty.

3.2. Conceptual Models

3.2.1. Overview of model selection

There are several models that can be utilized to forecast the time series data of trip requests of the potential Access Services customers. There are several phenomena that data (particularly time series) can exhibit, and any model should take these phenomena into account including:

- Trend, which occurs when the data has increasing or decreasing values as time progresses from one period to another,
- Seasonality, which refers to recurring patterns that follow a regular and predictable interval, often associated with calendar seasons or other periodic occurrences.
- Cyclicality, which encompasses fluctuations that occur over an extended duration, typically not as rigidly defined as seasonal patterns, and often are influenced by economic or external factors.

[2] The City of Arlington and New York City appear to be outliers for the operating expense per passenger trip, with the Arlington value significantly lower than peers and the New York City value significantly higher.

- Autocorrelation, which happens when the next value of data item is dependent on some previous data point, either an immediate predecessor or a predecessor with some lag (distance between related data items).

Making a model from real data, to be able to predict the future behavior of the time series, is critical. Such a model should account for all mentioned phenomena. Traditional models use statistical analysis in order to predict the model parameters, such that some measure of mean square error between the data and the model will be minimized. This will be explained in the section on linear regression. From the linear regression, as the simplest model for time series prediction, many other models were developed in classical statistical analysis, and other models were developed utilizing neural network.

Several methods were explored in order to provide a reliable estimate of data trends in the future for this project. In the literature Korstanje (2021) mentions some of them for univariate time series analysis:

1. AR – autoregression
2. MA – moving average
3. ARMA – combination of autoregression and moving average
4. ARIMA – adding differences to ARMA model
5. SARIMA – adding seasonality to ARIMA
6. SARIMAX – introducing additional, external variable(s) to SARIMA model

In addition to these simple methods, advances in supervised machine learning in Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), Deep Learning (DL), and others, developed methods such as Simple RNN with Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM), and DeepAR.

Following is a description of three methods that were utilized to develop the trip request estimates: linear regression, ARIMA and LSTM.

3.2.2. Linear regression

Linear regression is the required content in the first college-level course on statistics and this section provides a brief overview (Montgomery & Runger, 2024).

Given a data set *{(y_i, x_i), for i = 1,… n}* of *n* statistical units (data measurements), a linear regression can be represented as

$$\hat{Y} = \beta_0 + \beta_1 * X$$

Such that

$$E(x) = \mu_{y\backslash x} = \beta_0 + \beta_1 * x$$

Actually, for every y_i measured error of difference was introduced between the measured (real value) and calculated value:

$$y_i = \beta_0 + \beta_1 * x_i + \varepsilon$$

Where ε is random error (variable) with mean 0 (zero) and unknown variance σ^2. These errors correspond to the difference between real values and calculated values from the equation (see Figure 8 and Figure 9). The linear regression method calculates parameters β_0 and β_1 such that the error random variable has minimal variance, so called least square estimates. The formulas for those parameters, based on data values, can be found in any statistical handbook.

Figure 8 Graph of Data Points and the Regression Line

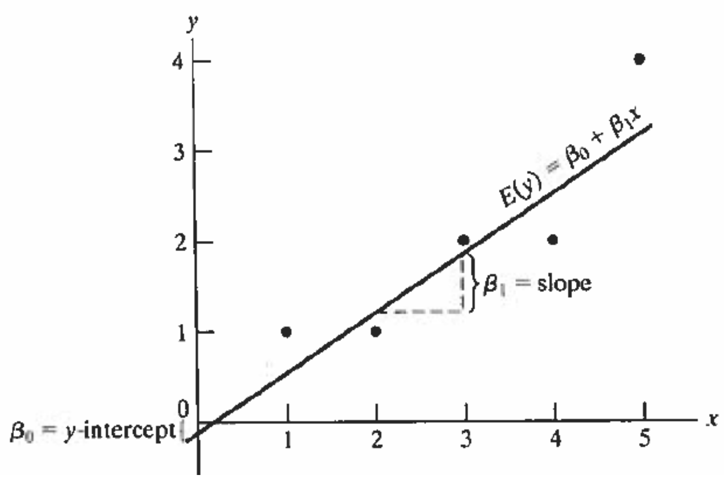
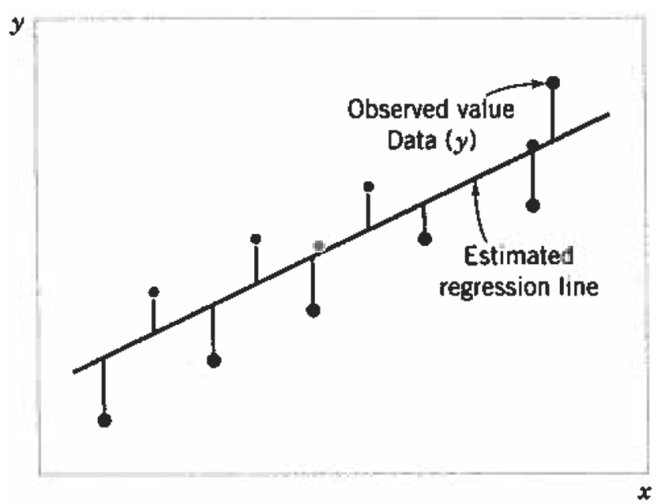


Figure 9 Deviation of the Data from the Estimated Regression Model (Montgomery & Runger, n.d.)



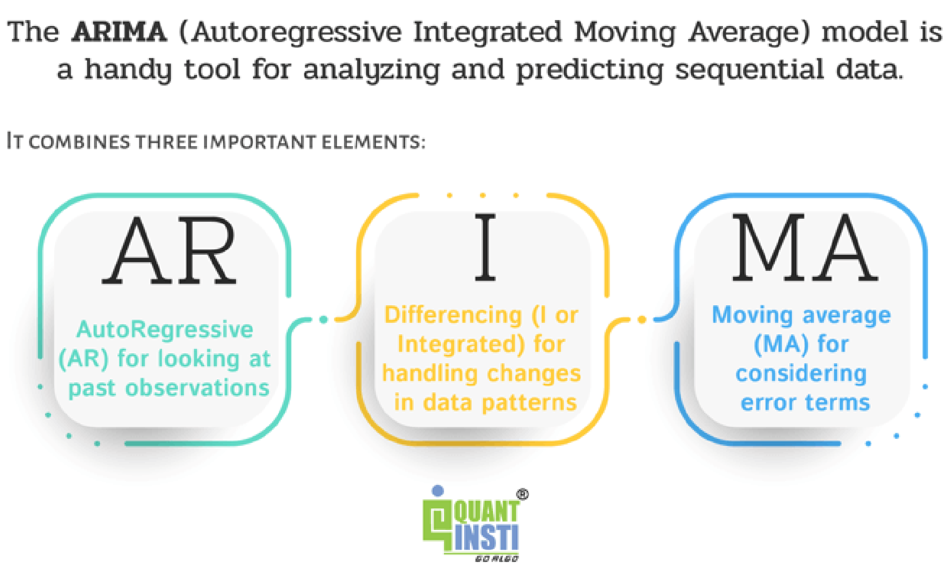
Therefore, in its simplest form, one variable linear regression of a time series takes dependent variable y_i recorded in a series of time points x_i, and produces a line based on the data points, which minimizes deviations of calculated value $\hat{y}_i$ from actual values y_i. This deviation can be linear (using absolute values of differences), square (squaring differences), or similar. By simple algebraic transformations, linear regression may also be used to model some other curves, like exponential $y = ae^{bx}$, logarithmic $y = a * \log x + b$, or power curves, $y = a * b^x$.

It is obvious that the simple linear regression cannot detect many phenomena mentioned above in the data; it is good for detecting only trends. Therefore, other advanced statistical models have been developed to address those phenomena, some mentioned above, and some that are described in the next two sections.

3.2.3. Autoregressive Integrated Moving Average (ARIMA)

Autoregressive integrated moving average (ARIMA) is a method for the time series analysis that can handle both stationary and non-stationary time series by introducing a differencing step to a non-stationary series, and addresses seasonality by seasonal differencing. It is the most extended classical statistical analysis model to obtain forecasts from time series data. It combines (see Figure 10) autoregression (AR) in order to look at past observations and moving averages (MA) to minimize the error terms, and differencing (or integration) for handling changes in data patterns. ARIMA is built into most statistical software and provides several methods for specifying execution parameters; the software also provides the method to forecast with Best ARIMA model which automatically selects the best model from a set of candidates. The forecasts using ARIMA model are calculated recursively, using the developed model and parameter estimates.

Figure 10 The Elements of ARIMA Model



3.2.4. Long Short-Term Memory (LSTM) Model

LSTM model is the type of recurrent neural network model that was developed by Hochreiter and Schmidhuber (1997) to address the problem of vanishing gradient in traditional RNN models, which were sensitive to the gap length when modeling seasonal data. LSTM model, like other neural network models, consists of one input layer, several internal (hidden) layers, and one output layer, as shown in the example on Figure 11 (Surakhi et al., 2021). The LSTM builds its neuronal units as sets of cells, controlled by input gates, output gates, and forget gates all built with appropriate sigmoid functions (see Figure 12, Calzone, 2022). The cell remembers values over any time intervals, while gates regulate the flow of information into the cell and from the cell to other cells. The forget gates are responsible for a decision if the information from a previous state will be forgotten or not. All these selections allow the

LSTM model to maintain long-term dependencies while propagating short-term variations between consecutive states.

Figure 11 Layered Architecture of the LSTM Model (Surakhi et al., 2021)

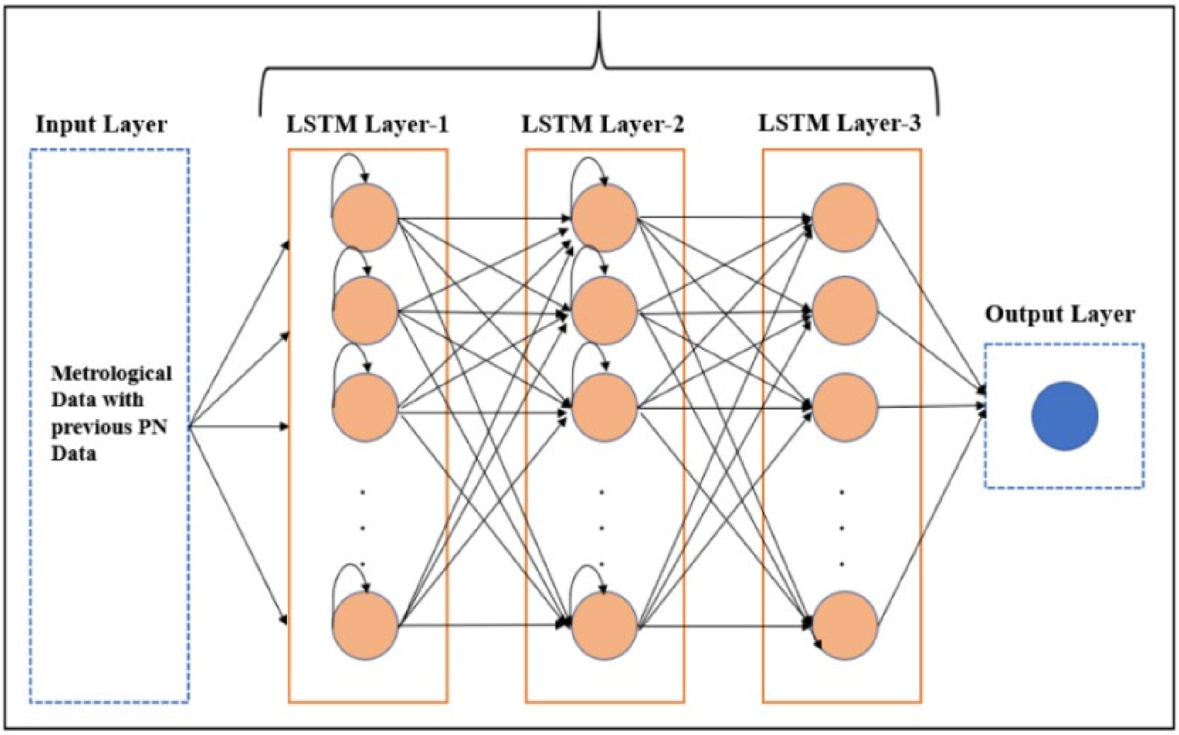
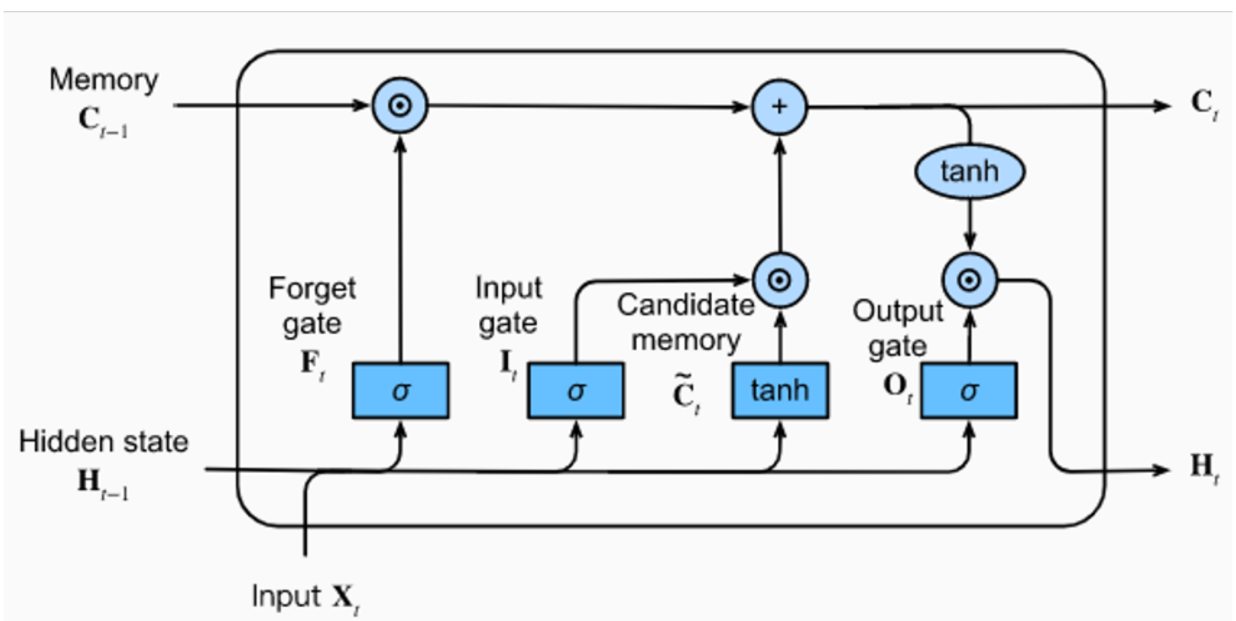


Figure 12 The Structure of the LSTM Cell with Three Gates: Input Gate, Forget Gate, and Output Gate (Calzone, 2022)



The dynamics of an LSTM cell can be described by the following equations, from Bedi and Toshniwal (2019) and Bordoni and Giagu (2023):

$$i_t = \sigma\left(x_t U^i + h_{t-1} W^i\right)$$

$$f_t = \sigma\left(x_t U^f + h_{t-1} W^f\right)$$

$$o_t = \sigma\left(x_t U^o + h^o_{\{t-1\}W}\right)$$

$$\tilde{c} = Tanh\left(x_t U^g + h^g_{\{t-1\}W}\right)$$

$$c_t = \sigma\left(f_t \otimes c_{\{t-1\}} \oplus i_t \otimes \backslash tilde\{c\}_t\right)$$

$$h_t = Tanh(c_t) \otimes o_t$$

Where

x_t is the input vector at time step *t*,

h_t is the output vector at time step *t*,

c_t is the cell state vector at time step *t*,

c~ is a candidate cell state that is computed based on the current input and the previous hidden state,

U is the weight matrix that connects the inputs to the hidden layer,

W is the recurrent connection between the previously hidden layer and current hidden layer,

i_t, f_t, and o_t are the input gate, forget gate, and output gate vectors respectively.

Input to the LSTM model is a time series data set (for this project actual trip requests from the past, either monthly or daily), which is used for both training of the model and its testing to minimize the statistical errors. Usually 70-80% of the data set is used for training, while the remaining 20-30% is used for testing of the model. Once an appropriate model is obtained through training and testing, it is used to predict the future data in the continuation of the time series.

Python Libraries was employed to perform LSTM model utilizing the LSTM model three stages:

Transformation, which is the preparation of the data set for the LSTM model that first removes trend in the data (detrending), then removes seasonality, and finally normalizes the data into the range of (0,…1), which is suitable for the next stage, forecasting.

Forecasting, is the essence of the LSTM Recurrent Neural Net learning model with specifying necessary parameters for the model, such as number of *epochs*, *lag* specification (layers in the model), and number of *units* (cells in the model).

Revert, which is the last phase used to revert the normalized data (in testing and in the prediction) to actual values and to reintroduce trend and seasonality into predicted values.

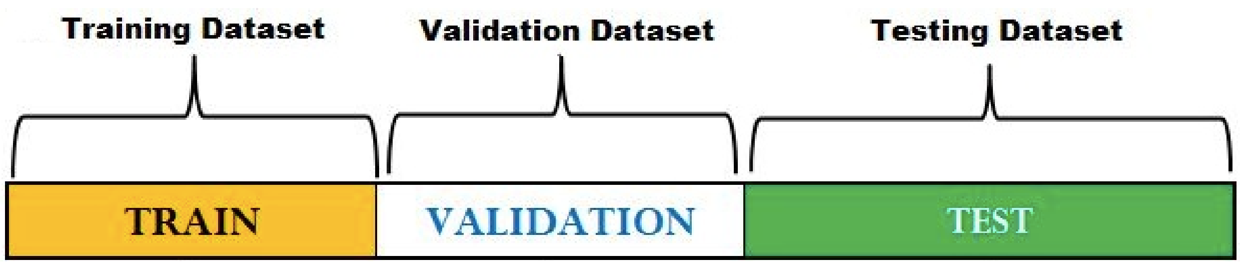
The model is trained and tested on real data (past observations) and then used to predict time series into the future periods.

For the trained and tested models, daily and monthly data are utilized to make the LSTM model of trip requests, as will be explained later.

3.3. Tests and Results – Scheduled Trip Demand

Several models were tested to evaluate the results and select the model(s) with the best likelihood to produce accurate forecasts. Common practice in data science is to separate historical data into three (3) data sets shown in Figure 13. The data is split, in chronological sequence for time series data sets, with the first 70% designated for training, 10% allocated for validation, and the remaining 20% set aside for testing.

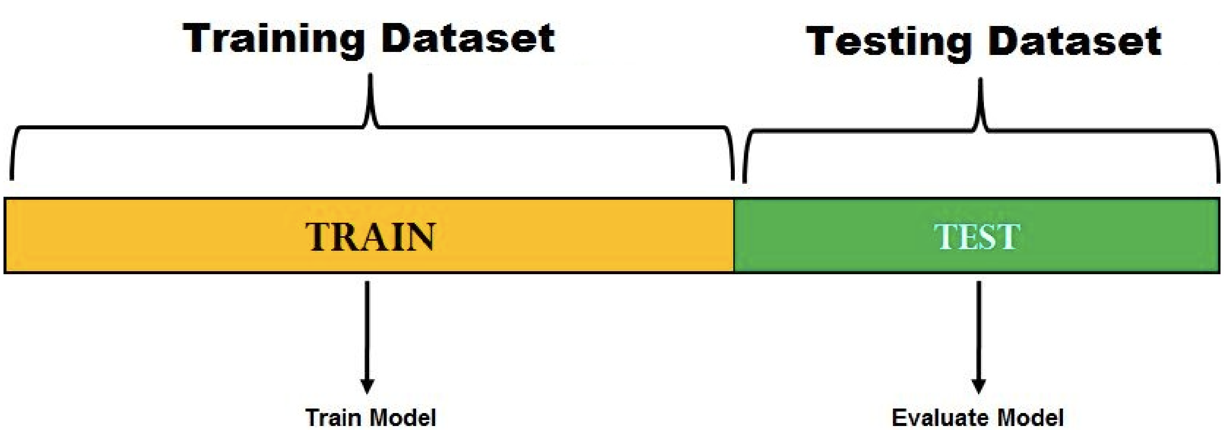
Figure 13 Training Validation Testing Model



* *Image credit to Analytics Yogi, https://vitalflux.com/hold-out-method-for-training-machine-learning-model/*

The historical analysis reveals the need to utilize post-pandemic data to reflect current trends most accurately. There is a limited amount of post-pandemic data available, especially considering potential seasonality and day-to-day variation. The hold-out method is common practice in data science and machine learning when there is too little data to break it into three (3) traditional sets. Figure 14 provides a visual representation of the hold-out method.

Figure 14 The Hold-Out Method Data Sets



* *Image credit to Analytics Yogi, https://vitalflux.com/hold-out-method-for-training-machine-learning-model/*

For the hold-out method, the data is grouped into the training set (70%) and the test set (30%) in chronological sequence for time series data. The training set is used to create the model. The test set is compared to the results of the model trained on the training data set to evaluate the model performance.

3.3.1. Training and Testing – Scheduled Trip Demand

The initial training and testing were conducted with monthly scheduled trip demand, the number of trip requests, from April 2020 through August 2023. The training data consisted of 31 months (April 2020 through October 2022), approximately 76% of the available data. The testing data set consisted of 10 months (November 2022 through August 2023). Variations of three models were tested: Linear Regression, ARIMA, and LSTM.

The LSTM model with parameters of 18 lags (layers), 36 units (cells), and 200 epochs provided better accuracy than ARIMA and Linear Regression for the testing period. The comparison of the LSTM model using monthly data to the actual data and the previous forecast is shown in Table 4.

Table 4 LSTM Model - Monthly Data Comparison

YEAR	MONTH	ACTUAL TRIP REQUESTS	LSTM MODEL FORECAST (MONTHLY)	ERROR	ABSOLUTE ERROR (DEVIATION)	ABSOLUTE PERCENT ERROR	SQUARED ERROR	PREVIOUS FORECAST	ERROR	ABSOLUTE ERROR (DEVIATION)	ABSOLUTE PERCENT ERROR	SQUARED ERROR
2022	NOVEMBER	287,324	300,145	-12,821	12,821	4.46%	164,382,472	278,696	8,628	8,628	3.00%	74,442,384
2022	DECEMBER	288,377	302,111	-13,734	13,734	4.76%	188,614,838	290,223	-1,846	1,846	0.64%	3,407,117
2023	JANUARY	293,119	269,089	24,030	24,030	8.20%	577,434,186	281,816	11,303	11,303	3.86%	127,752,849
2023	FEBRUARY	283,804	277,789	6,015	6,015	2.12%	36,183,579	267,915	15,889	15,889	5.60%	252,467,059
2023	MARCH	331,851	326,660	5,191	5,191	1.56%	26,941,535	297,776	34,075	34,075	10.27%	1,161,118,077
2023	APRIL	315,771	323,935	-8,164	8,164	2.59%	66,656,286	300,363	15,408	15,408	4.88%	237,397,570
2023	MAY	337,826	328,145	9,681	9,681	2.87%	93,722,359	333,624	4,202	4,202	1.24%	17,657,551
2023	JUNE	326,934	324,827	2,107	2,107	0.64%	4,439,131	300,804	26,130	26,130	7.99%	682,800,954
2023	JULY	322,123	315,461	6,662	6,662	2.07%	44,384,341	321,762	361	361	0.11%	130,224
2023	AUGUST	351,654	337,123	14,531	14,531	4.13%	211,157,008	327,198	24,456	24,456	6.95%	598,115,544
OVERALL		3,138,783	3,105,285	33,498	33,498	1.07%	1,413,915,735	3,000,176	138,607	138,607	4.42%	3,155,289,328
MEAN ABSOLUTE DEVIATION (MAD)					10,294					14,230		
MEAN ABSOLUTE PERCENTAGE ERROR						3.34%					4.45%	
MEAN SQUARED ERROR							141,391,574					315,528,933

The LSTM model, utilizing monthly data, outperformed previous forecasts by more than 105,000 trip requests for the sum of the testing data, however, it still presented a negative bias. In other words, the projected values were regularly lower than the actual (observed) values. The negative bias is a limitation when budgeting for scheduled trip demand because the budget does not fulfill actual need.

To improve the accuracy of the forecast of scheduled trip demand, the analysis shifted to the number of daily trip requests recorded from January 1, 2021, through October 31, 2023. The change in measurement increased the size of the data set 2,422%, from 41 data points (historical observations) to 1,034. The training data set, January 1, 2021, through July 31, 2023, was 91% of the data while the testing data set, August 31, 2023, through October 31, 2023, was 9% of the data set. Several parameters of the LSTM model were tested, resulting in the discovery of a model with 28 lags, 30 units, and 200 epochs, which provided the best accuracy.

The results of the LSTM model were compiled into monthly forecasts and compared to the actual observations along with the previous forecasts. Table 5 showcases the results and comparisons.

Table 5 LSTM Model - Daily Data Comparison

YEAR	MONTH	ACTUAL TRIP REQUESTS	LSTM MODEL FORECAST (DAILY)	ERROR	ABSOLUTE ERROR (DEVIATION)	ABSOLUTE PERCENT ERROR	SQUARED ERROR	PREVIOUS FORECAST	ERROR	ABSOLUTE ERROR (DEVIATION)	ABSOLUTE PERCENT ERROR	SQUARED ERROR
2023	AUGUST	351,654	353,518	-1,864	1,864	0.53%	3,475,577	327,198	24,456	24,456	6.95%	598,115,544
2023	SEPTEMBER	342,300	346,031	-3,731	3,731	1.09%	13,924,040	323,869	18,431	18,431	5.38%	339,684,023
2023	OCTOBER	361,847	357,663	4,184	4,184	1.16%	17,504,070	327,113	34,734	34,734	9.60%	1,206,468,318
OVERALL		1,055,801	1,057,213	-1,412	1,412	0.13%	1,993,734	978,180	77,621	77,621	7.35%	6,025,046,416
MEAN ABSOLUTE DEVIATION (MAD)					3,260					25,874		
MEAN ABSOLUTE PERCENTAGE ERROR						0.93%					7.31%	
MEAN SQUARED ERROR							11,634,562					714,755,962

The LSTM model for **daily data** outperformed the LSTM model for monthly data with an overall absolute percentage error of 0.13% compared to 1.07%. Essentially the daily data proved to be 88% more accurate for the LSTM model than monthly data when comparing the overall deviation (difference) from the overall actual observations, in alignment with how Access Services performs a budget comparison. The LSTM model utilizing the daily number of trips predicted the actual (observed) number of trip requests within 1,412 requests of the total while previous forecasts projected a deficit of 77,621 too few requests. Projections from the LSTM model are seven (7) times more accurate than previous forecasts in initial tests.

Based on these results, the LSTM model, utilizing the daily number of trip requests with parameters of 28 lags, 30 units, and 200 epochs, was selected to forecast scheduled trip demand.

3.3.2. Test Results Summary – Scheduled Trip Demand

Testing of several potential forecasting models led to findings including:

- The LSTM model produces the most accurate projections for scheduled trip demand.
- Previous forecasts had a tendency to under forecast, a negative bias, for scheduled trip demand.
- Previous forecasts were deficient and resulted in significant practical error for scheduled trip demand.

The findings were shared with Access Services in a virtual meeting, where the discoveries were validated and confirmed. Then projections were developed for Fiscal Years 2025 through 2034 using the LSTM model for scheduled trip demand.

Scheduled trip demand forecasts are based upon the historical trip requests made post-pandemic, from January 2021 through October 2023. The limited amount of historical data creates a predicament for the deliverable requested by Access Services. Essentially, the analysts are attempting to forecast ten years of scheduled trip demand with less than three full years of historical post-pandemic data. The LSTM model is a deep learning, recurrent neural network capable of learning the patterns of trip requests and projecting the demand for this uncertain situation.

3.4. Methodology

The LSTM model utilizes historical data to pass through the remember and forget gates to the learn patterns. The model is physically limited to the number of periods it can project into the future based upon the amount of data available to pass through these gates. Two separate phases, short-term and long-term, were utilized to address the lack of historical data and limitations of the model.

3.4.1. Short-Term

The short-term forecast is defined as the next five (5) fiscal years. The short-term forecast is most critical to ensure the budget and operational resources are in place for Access Services to deliver its services to eligible riders.

The LSTM model produced over two and a half Fiscal Years of forecasts using historical daily trip requests from January 1, 2021, through October 31, 2023. The daily forecasts were aggregated to calculate the monthly scheduled trip demand. Forecasts for the remainder of Fiscal Year 2024, along with projections for Fiscal Years 2025 and 2026, were based solely on historical, post-pandemic, trip requests.

The LSTM model reached its limit using historical trip requests after Fiscal Year 2026. A data set comprised of historical daily trip requests from January 1, 2021, through October 31, 2023, and trip request forecasts for November 1, 2023, through June 30, 2026, were utilized to create projections for Fiscal Years 2027 through 2031.

Various parameters for the LSTM model were tested using data combined with historical data and forecasted. The results were compared to previous forecasts utilized by Access Services for Fiscal Years 2027 through 2031. The LSTM model projections ranged from 15.18% to 21.77% greater than Access Services' previous forecasts. The parameters of seven (7) lags, 30 units, and 200 epochs were selected to create projections 15.18% greater than the previous forecasts, which addresses the issue of previous forecasts being too low, with a negative bias.

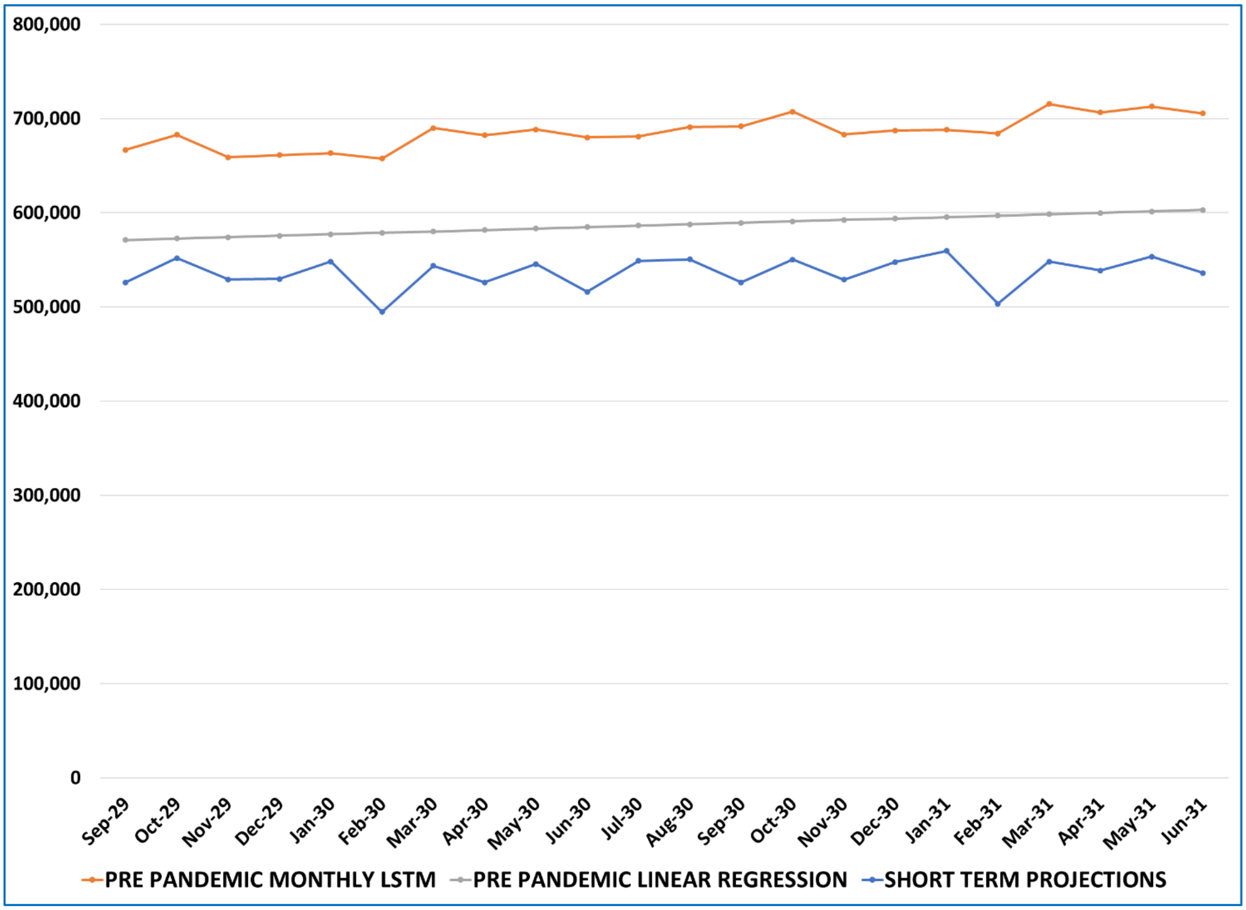
The LSTM model reached its limit again in terms of having enough data to forecast for Fiscal Years 2032 through 2034. A data set comprised of historical daily trip requests from January 1, 2021, through October 31, 2023, and trip request forecasts for November 1, 2023, through June 30, 2031, were used to create projections for Fiscal Years 2032 through 2034. This model and data set resulted in a slightly downward trend in scheduled trip demand. The result is likely a weakness of the use of the quadratic trend in the LSTM model. It is important to note that neither the historical analysis of scheduled trip demand pre-pandemic, nor post-pandemic, suggest a downward trend. The methodology for the short-term provided seven (7) full fiscal years of forecasts, extending into the long-term period. A different approach was needed to address the projections for later fiscal years.

3.4.2. Long-Term

Models with a different data set were evaluated to address the slightly downward trend in scheduled trip demand projections produced by the LSTM model using post-pandemic historical and forecasted data. For these models, historical monthly trip request data prior to the pandemic, from January 2010 through February 2020 were utilized.

Short-term forecasts for September 2029 through June 2031 were used as a baseline to evaluate the accuracy of long-term models compared to models with pre-pandemic historical data. The time series plot in Figure 15 shows the performance of the LSTM model and the linear regression models using the monthly pre pandemic historical data set compared to short term projections. The results of both models are greater than the short-term projections, with the linear regression being the closest.

Figure 15 Pre-Pandemic Data Set Model Comparison



The average distance between the short-term projections and the Linear Regression model with pre-pandemic data is 50,527 trip requests per month. This distance equates to an average absolute percentage error of 9.5%. A hybrid approach was developed to address this error between the baseline LSTM model and the Linear Regression model with pre-pandemic data to adjust the forecast trend for Fiscal Years 2032 through 2034.

The average number of trip requests from each model, the baseline LSTM model, and the Linear Regression model with pre-pandemic data, were calculated to create a hybrid model to forecast Fiscal Years 2032 through 2034. For example, the short-term LSTM model projected 6,431,822 trip requests in Fiscal Year 2032, while the Linear Regression model with pre-pandemic data forecasted 7,353,629 trip requests. The average of these two data points is 6,892,726 trip requests, which is the forecast for Fiscal Year 2032. The hybrid model preserves the approach of utilizing post-pandemic data to forecast scheduled trip demand while preventing the unlikely downward trend in projections.

3.5. Forecasting Assumptions and Risk Analysis – Scheduled Trip Demand

There are different approaches to evaluate risk for the results of forecasting models. The Long Short-Term Memory (LSTM) model provides the best overall forecasts of scheduled trip demand in the testing phase. Predictors used to evaluate the error of the model during this phase include the Mean Average

Deviation (MAD), Mean Average Percentage Error (MAPE), and the Mean Squared Error (MSE). Definitions and formulas for these metrics are shown in Appendix A-1: Definitions.

The value of the error is an inverse relationship with the accuracy of the model. The lower the value of the result of the error formula, the greater the accuracy of the model.

The LSTM Model, using the daily number of trip requests, exhibited high accuracy in the testing phase as indicated by the error measures in Table 3. Forecasts from this model are compiled into monthly projections or trip requests. Of note, the MAPE provides a gauge of the relative accuracy, which means projections *were within 0.93% of the actual observations in the testing phase*. Previous forecasts were substantially less accurate with a MAPE of 7.31% in the testing phase.

The MAD measures the practical accuracy of the predictions. The LSTM model experienced an MAD of 3,260 in the testing phase, meaning the average error was 3,260 trip requests per month. This equates to a *potential error of 39,120 trip requests per year or about one tenth the average error from prior forecasts*. The MAD for the previous forecast in the testing phase was 25,874 per month, and an error of 310,488 trips per year.

The forecasted values of the LSTM model are subject to error; however, this error is relatively minimal. The projections provide significantly better accuracy than previous forecasts. The error could have a slight negative bias, meaning the actual observed values could be greater than the forecasted values.

3.5.1. Assumptions – Scheduled Trip Demand

Projections for scheduled trip demand have inherent uncertainty, which is typical of any predictions. The forecasted values one (1), five (5), and ten (10) fiscal years into the future are greater than the most recently completed fiscal year, 2023 (3,605,481). Two factors explain the feasibility of these projections and the possibility of actual observations reaching these values.

The first factor to explain the feasibility of forecasted values is the linear trend analysis of pre-pandemic trip requests, January 2010 through February 2020. The analysis creates forecasted values greater than the scheduled demand projections from the model as shown in Figure 15. In some ways, this serves as a reference for the expected increase in scheduled trip demand. If the COVID-19 pandemic had not occurred, then it is plausible Access Services could have observed actual numbers of trip requests of this magnitude, greater than the forecasted values herein.

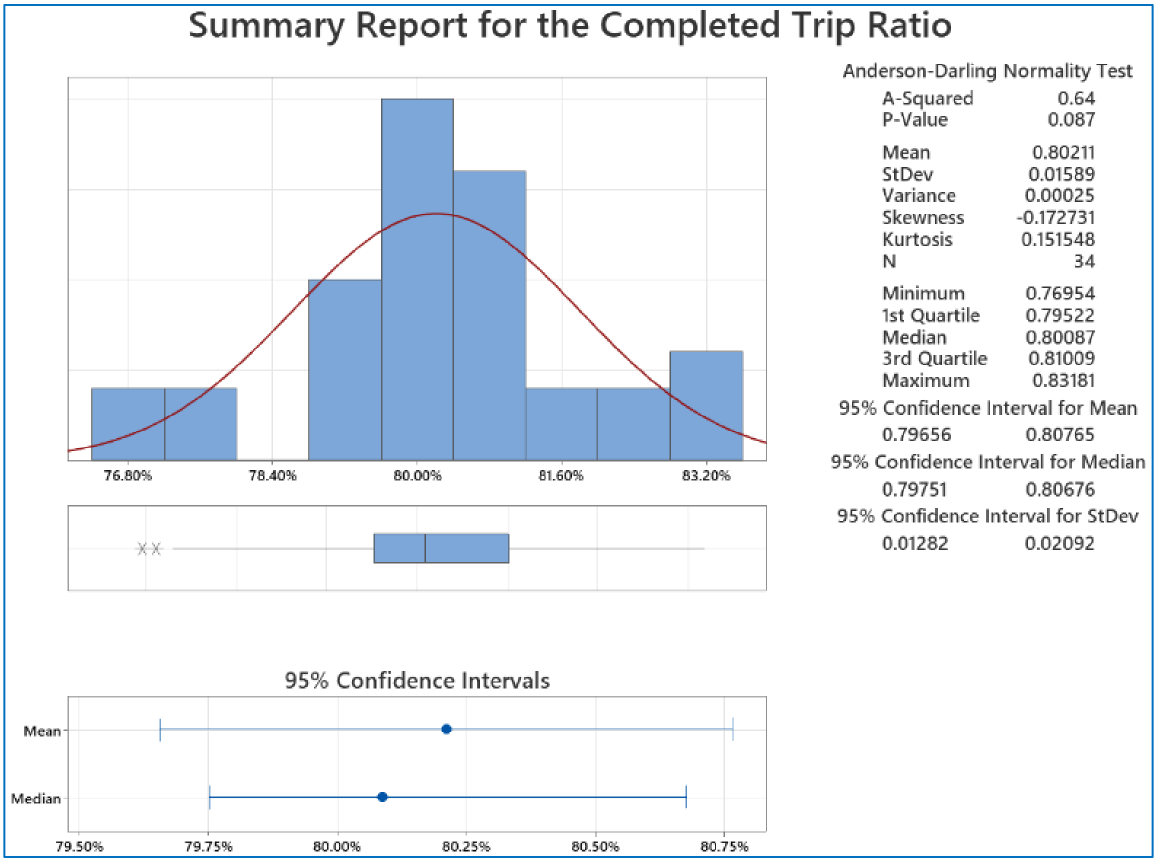
Secondly, Access Services is experiencing an increase in the number of unique riders each month as shown in Appendix A-5: Analysis of Unique Riders. This trend suggests more eligible riders are requesting trips, which leads to an increase in scheduled trip demand.

3.5.2. Scheduled Trip Cancellation Risk

Scheduled trip demand reflects the number of trip requests from eligible riders. As with any reservation or appointment, there are cancellations by the rider and rider no-shows. An analysis of the number of trip requests, scheduled trip demand, provides guidance to Access Services for translating the scheduled trip demand into the number of completed trips.

The completed trip analysis uses monthly data for the period from January 2021 through October 2023. The completed trip ratio is the number of completed trips divided by the number of scheduled trips (trip requests) during the month. This ratio is the percentage (%) complete. The analysis shown in Figure 16 shows the descriptive statistics along with a histogram that visually represents the amount of variation typically experienced in the ratio.

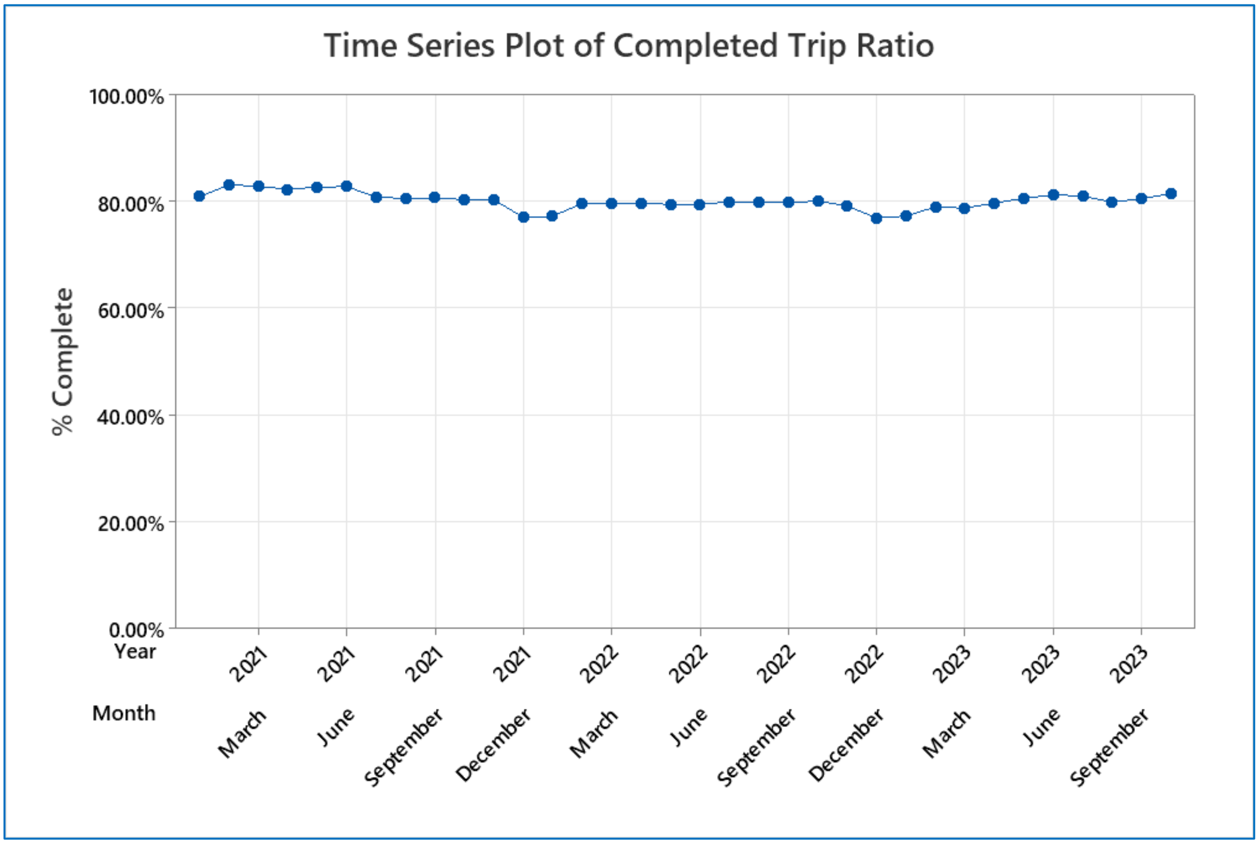
Figure 16 Completed Trip Ratio Statistical Analysis



The analysis illustrates the average completed trip ratio is 80.21%. In other words, scheduled trips are completed 80.21% of the time. The analysis also shows the completed trip ratio is normally distributed. A second normality test, the Ryan-Joiner, was performed to confirm this data is normally distributed. There is 95% confidence that the average completed trip ratio is between 79.75% and 80.77%.

The histogram in Figure 16 also displays the variation of the completed trip ratio from month to month. The ratio ranges from 76.95% to 83.18%. While it experiences slight fluctuation month to month, the ratio of completed trips is relatively constant based on it being normally distributed and as shown in the time series graph in Figure 17.

Figure 17 Time Series Plot – Completed Trip Ratio



The scheduled trip demand forecast, along with the completed trip ratio analysis, provides Access Services with insight for planning budgets and operational resources along with confident long-range planning for the next ten (10) fiscal years.

3.6. Conclusion and Next Steps – Scheduled Trip Demand

The initial analysis of the scheduled trip demand reveals essential insight into current trends. The trend is significantly different after the COVID-19 pandemic than the trend prior to the pandemic. Forecasting models need to utilize post-pandemic historical data for projections to reflect current trends.

Training and testing of several forecasting models led to the selection of the models that best fit current trends for the scheduled trip demand. The Long Short-Term Memory (LSTM) model best fits the scheduled trip demand for the next seven (7) calendar years. Testing results demonstrate LSTM performed better than previous forecasts in the testing phase and is likely to outperform previous forecasts.

Due to the limited amount of post-pandemic historical data and the use of a quadratic trend in the forecasting model, a hybrid approach between the LSTM model and linear regression of pre-pandemic data were used to forecast scheduled trip demand for Fiscal Years 2032 through 2034.

An analysis of the completed trip ratio, the percentage of scheduled trips completed, is stable with an average value of 80.21%.

The national trend exhibits increasing demand response trips as discovered in the peer review for this project. The trend indicates Access Services' scheduled trip demand will increase in the future, supporting the projections provided in this report.

The accuracy of projections for scheduled trip demand during the next iteration of forecasting is expected to further improve as additional data becomes available. These projections are currently based on historical daily trip requests from January 2021 through October 2023. The number of observations of historical daily trip requests will increase 35% by October 2024. This increase creates both additional training and testing data, which will enhance confidence and accuracy in the projection model.

Other steps to improve the accuracy of projections for scheduled trip demand during the next iteration involves incorporating detrending methods such as Empirical Mode Decomposition (EMD) in the LSTM forecasting model as well as transformers. EMD provides both detrending and feature construction to better address underlying patterns and capture essential characteristics that improve forecasting accuracy. Transformers strengthen the LSTM model by further differentiating the importance of past observations and focus attention on more relevant observations to improve forecasting accuracy.

Next steps include considering additional forecasting and machine learning models and testing them as appropriate. The Long Short-Term Memory (LSTM) model selected to develop schedule trip demand is univariate time series forecasting. This means the only factors considered in the model of eligible rider behavior are the dependent variable, number of trip requests, and the time.

The LSTM model also has the capability for multivariate time series forecasting, which means independent variables and their effect on the number of trip requests are considered. Examples of other independent variables that could be included in a multivariate LSTM model include the unemployment rate, fuel prices, population, and more. This model will be considered based upon the performance of the univariate model along with an initial analysis of other independent variables.

The Meta (Facebook) Prophet model is a modular regression model for forecasting with features such as the ability to model multiple seasonalities and the ability to identify changepoints, when a change occurs in the data (Taylor, 2017). There are options to choose the growth function (linear, logistic, flat) in the Prophet model as well as a function to handle drastic changes in values for holidays and events (Taylor, 2017). The Prophet model will be considered based upon the performance of the LSTM univariate forecasting model along with an initial analysis of other independent variables.

Additional forecasting models to be evaluated and considered include the Time-Series Foundation Model developed by Google for the scheduled trip demand.

Monitoring of the demand dynamics on future iterations will continue to determine if the data sets revert to pre-pandemic values and patterns.

4. New Applicants

4.1. Initial Analysis – New Applicants

Patterns of the past help plan for the uncertainty of the future. Essential factors that provide insight about the paratransit needs in Los Angeles County, California, include the number of new applicants for paratransit service. A study on the COVID-19 pandemic, shown in Section 3.1.2, provides further insight. A review of peer paratransit services, shown in Section 3.1.3, also provides an opportunity for insights on services in other regions of the country.

The focus of the initial analysis includes a historical analysis on the number of new applicants for paratransit service, a brief evaluation of the global pandemic effect, and peer review. Together, the historical analysis, pandemic evaluation, and peer review provide direction for the types of forecasting models and the variables to include (exclude), consider and evaluate.

4.1.1. Historical Analysis – New Applicants

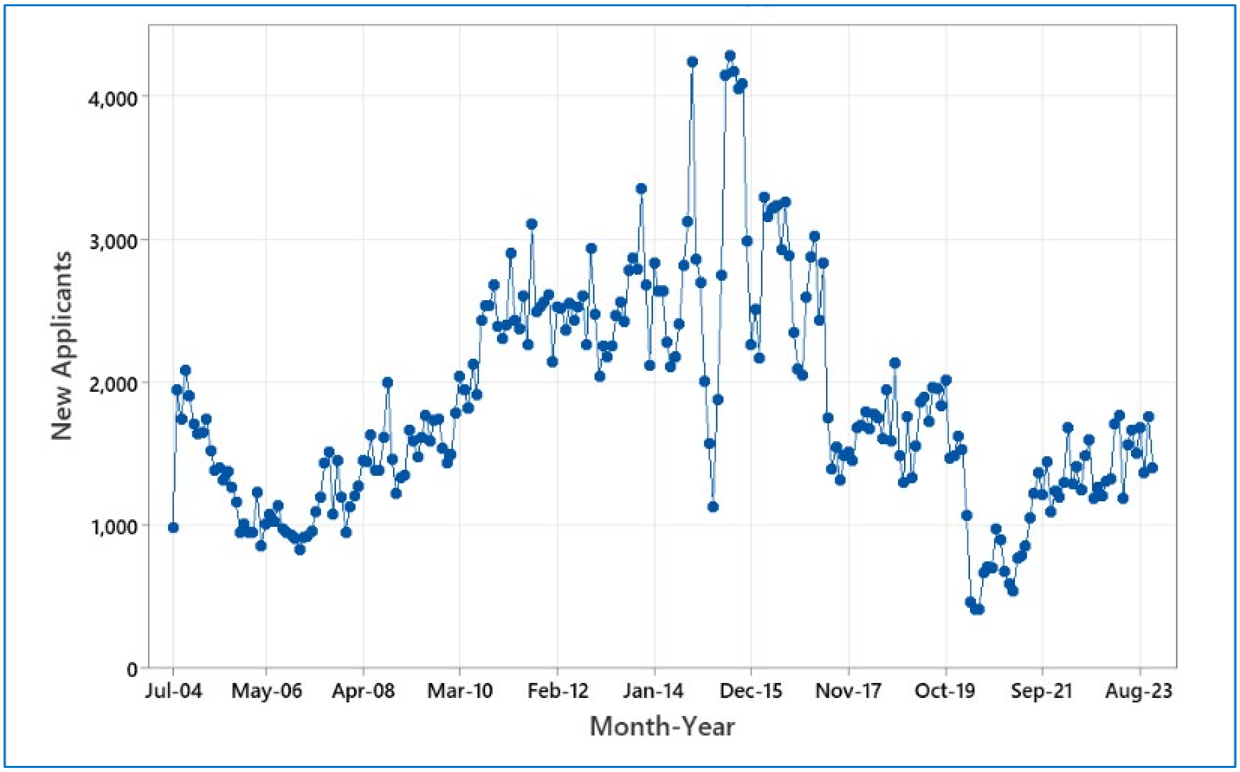
The examination of evidence from the past helps form a more coherent story. The focus of this examination includes an analysis of events in time series to identify patterns, trends, and changes over time. The analysis identifies the presence of (or lack of) seasonal patterns, cyclical patterns, stationarity, and autocorrelation along with trends. These components are key for model identification and selection.

Access Services needs to understand the number of eligible customers to develop an accurate, effective budget and plan future fiscal year(s). The historical analysis includes the number of new applicants.

The initial data for the historical analysis of new applicants to become eligible customers for Access Services includes the number of new applicants, from July 2004 through November 2023. The data file provided to Hollingworth Consulting included a count of certification evaluations (new applicants) per month for each service region[3]. Visualization was the first step to begin to understand this variable. The time series plot for monthly new applicants is shown below in Figure 18.

[3] Recertification evaluations were excluded since these are for existing eligible riders.

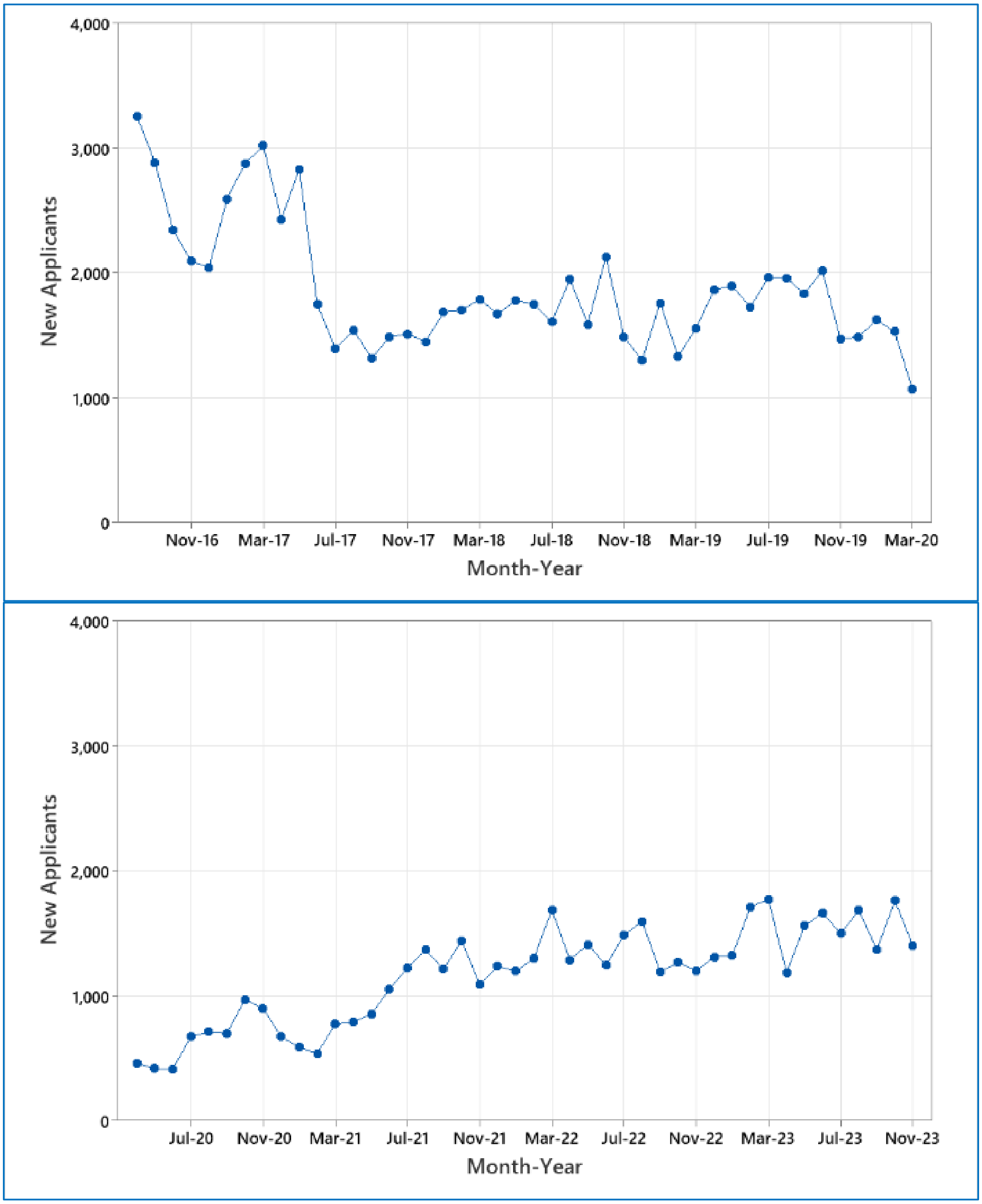
Figure 18 Monthly New Applicants Time Series Plot



The time series plot reveals several fluctuations with upward and downward shifts during the twenty (20) year period. These shifts are as numerous as 3,000 new applicants in a month. There is a large downward shift from February 2020 to June 2020, with 73% fewer new applicants (1,119) in just five (5) months. There is an upward shift beginning in July 2020 continuing through November 2023. The sudden, drastic change from a downward shift to an upward shift suggests a new or different trend.

Due to the significant historical shifts in the number of new applicants, there was further study of the time after the pandemic began, a span of 44 months. The post-pandemic data was compared to the historical number of new applicants for the same time span prior to the pandemic. Figure 19 shows the pre-pandemic number of new applicants on the left and the post-pandemic applicant numbers on the right.

Figure 19 Monthly New Applicants Time Series Plot Pre and Post-pandemic



Visual inspection of Figure 19 reveals the number of new applicants appeared to be decreasing during the 44 months leading up to the pandemic, while the number of applicants appears to be increasing after the pandemic. The trend analysis tool in Minitab Statistical Software (version 21.4.2) was used to examine the number of monthly new applicants as was done for the historical analysis of scheduled trip demand.

The quadratic model is the best fit of a general trend model for the monthly number of new applicants before the pandemic as shown in Figure 20. The quadratic model scored a lower MAPE, MAD, and MSD than the linear and the exponential growth (or decay) models for this data set[4] as shown in Table 6.

Figure 20 Trend Analysis of New Applicants – Pre-Pandemic

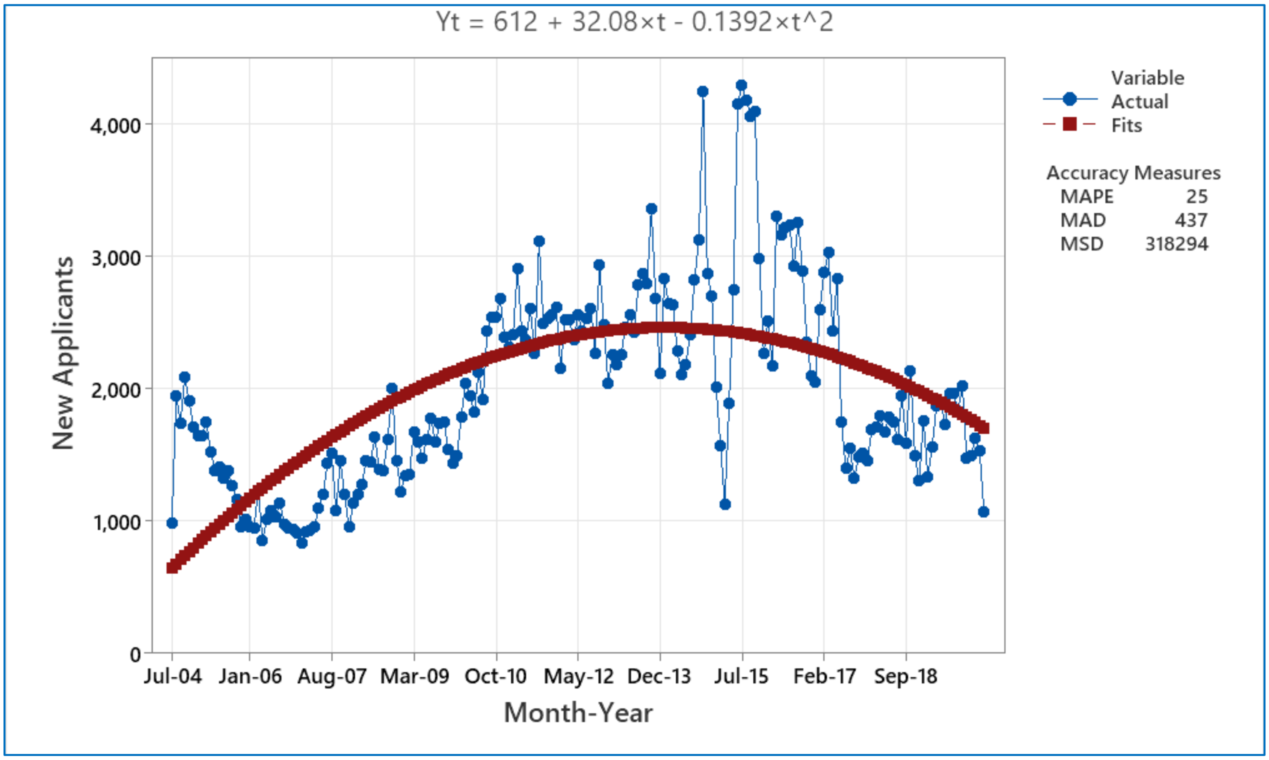


Table 6 Pre-Pandemic New Applicants Trend Model Scores

TREND MODEL	MAPE	MAD	MSD
Linear	31	554	455,684
Quadratic	25	437	318,294
Exponential Growth (Decay)	30	564	486,423
S-Curve	-	-	-

The S-curve model is the best fit of a general trend model for the monthly number of new applicants in the 44 months pre-pandemic as shown in Figure 21. The S-curve model scored a lower MAPE, MAD, and MSD than the linear, exponential growth (or decay), and quadratic models for this data set as shown in Table 7.

[4] The tool utilized for trend analysis could not fit the data to an S-curve model.

Figure 21 Trend Analysis of New Applicants – 44 Months Pre-Pandemic

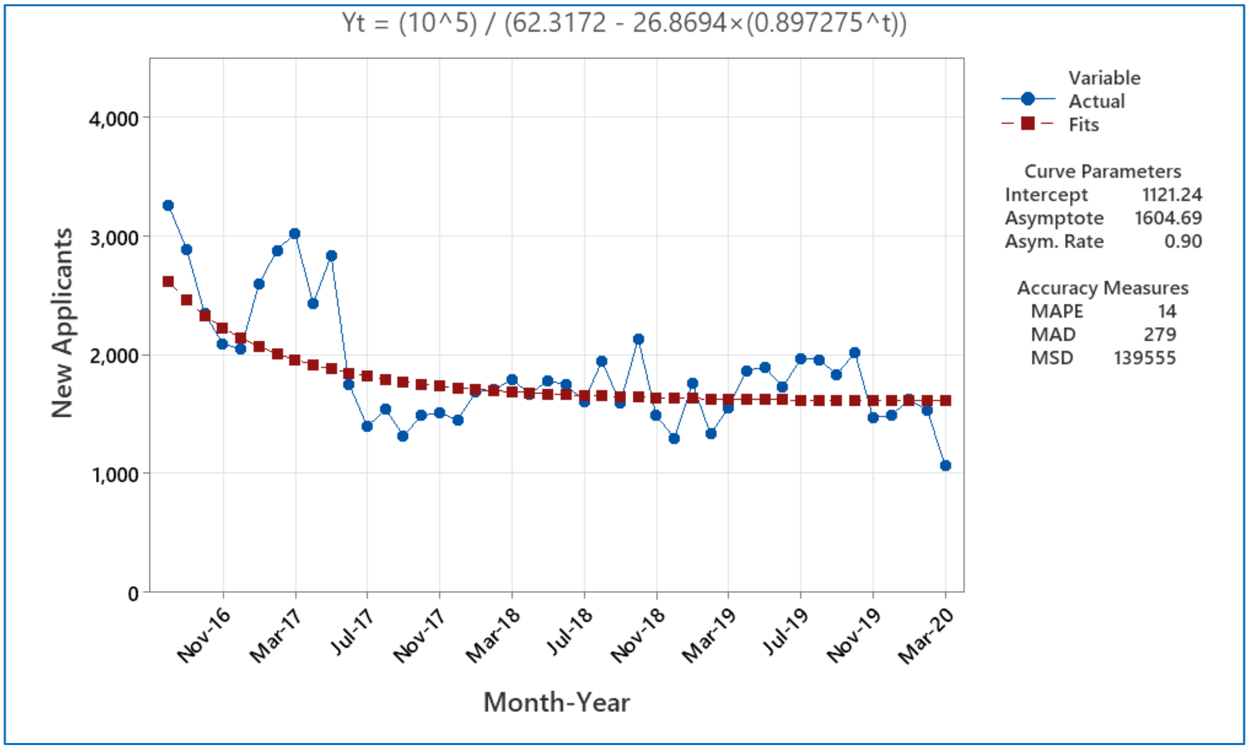


Table 7 44 months Pre-Pandemic New Applicants Trend Model Scores

TREND MODEL	MAPE	MAD	MSD
Linear	19	338	163,471
Quadratic	17	298	128,658
Exponential Growth (Decay)	17	324	160,789
S-Curve	14	279	139,555

The quadratic model is the best fit of a general trend model for the monthly number of new applicants post-pandemic as shown in Figure 22. The quadratic model scored a lower MAPE, MAD, and MSD than the linear, exponential growth (or decay), and S-curve models for this data set as shown in Table 8.

Figure 22 Trend Analysis of New Applicants – Post-Pandemic

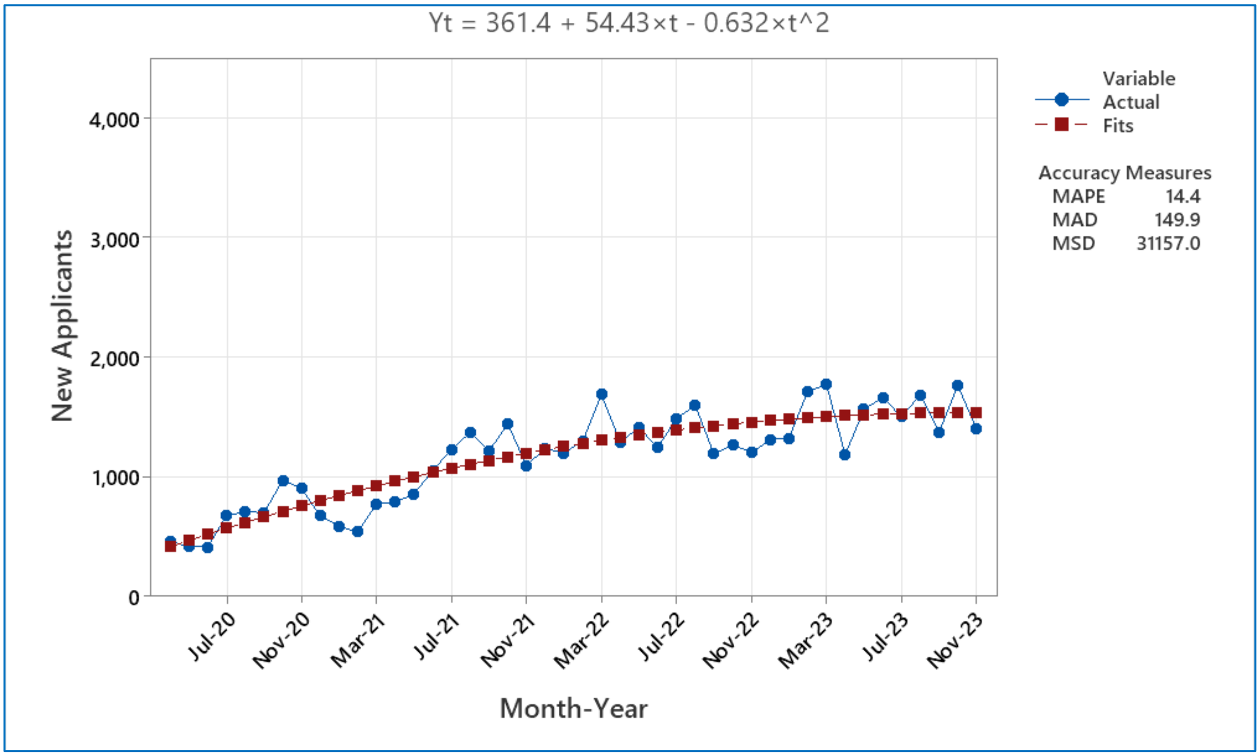


Table 8 Post-Pandemic New Applicant Trend Model Scores

TREND MODEL	MAPE	MAD	MSD
Linear	16	164	39,447
Quadratic	14	150	31,157
Exponential Growth (Decay)	18	189	54,757
S-Curve	15	157	33,063

The trend analysis indicates that there is a clear difference between the trends prior to the pandemic to that of the trends after the pandemic. The equations for the trend models in Figure 20, Figure 21, and Figure 22 are all mathematically different. The best fit of a general trend model for the number of new applicants post-pandemic is a quadratic model while the best fit for the equivalent period prior to the pandemic is a different model, the S-curve. The mathematical equations validate that there is a difference between the trends prior to the pandemic to that of the trends after the pandemic.

The historical data for new applicants is based upon the number of monthly evaluations for certifications. Forecasts for the number of new applicants provide Access Services with the knowledge to plan budgets and operational resources necessary to perform certification evaluations. The projections also reveal the potential impact on future scheduled trip demand.

4.1.2. Initial Analysis Summary – New Applicants

The historical analysis reveals insights into the number of new applicants for paratransit service and paratransit ridership nationwide. The results provide guidance on the types of models to select, and the periods of data utilized to forecast these variables.

The results document that there is a difference in trends for the number of new applicants pre-pandemic and post-pandemic. The pandemic disrupted consumer behavior including the number of new applicants. A review of the results of the historical analysis with Access Services validated the findings based on agency knowledge and experience with the day-to-day operations.

The pre-pandemic data does not reflect current behavior; post-pandemic data captures the altered dynamics more accurately. Forecasting models need to use post-pandemic data to ensure relevance to the current economic, social, and environmental context and to generate projections.

There is limited data available post-pandemic, 44 months for this project (April 2020 through November 2023). There is uncertainty about whether, or when, pre-pandemic patterns will re-emerge. Models to forecast scheduled trip demand and the number of new applicants need to have the capacity to predict accurately, with a limited amount of historical data, and be able to respond quickly to changes given the anticipated uncertainty.

4.2. Tests and Results – New Applicants

Forecasting models such as linear regression, ARIMA, and LSTM discussed in Section 3.2 were considered for the number of new applicants. The hold-out method discussed in Section 3.3, Figure 14, was used to train and test models to forecast the number of new applicants.

4.2.1. Training and Testing – New Applicants

The historical analysis reveals the need to utilize post-pandemic data to reflect current trends most accurately. There is a limited amount of post-pandemic data available, especially considering potential seasonality and day-to-day variation.

The initial training and testing were conducted for the number of new applicants with monthly historical data from April 2020 through November 2023. With the limited number of post-pandemic data points, the training set consisted of April 2020 through November 2022 and the testing set included December 2022 through November 2023. The training set included 72% of the data set while the testing set included the remaining 28%. Variations of three models were tested: Linear Regression, ARIMA, and LSTM.

The post-pandemic data for the monthly number of new applicants is normally distributed as shown in Appendix A-3: Normality Test Results – Post-pandemic New Applicants. This indicates this data set can be used directly in ARIMA and Linear Regression without the need for a technique such as a Box-Cox Transformation to convert it to the shape of a normal distribution.

An ARIMA model and an LSTM model created the most accurate results of the models tested. The parameters for the ARIMA model, with the most accurate test results, were a differencing value of 2, an autoregressive term (lags) value of 1, and a moving average term (lags of the forecast errors) value of 1. Both terms in the model, autoregressive and moving average, meet assumptions and fit the data well as indicated by the hypothesis testing results shown in Appendix A-4: ARIMA Model Parameters – Results and Analysis. Test results also show the residuals are independent, validating the selection of the model parameters.

The forecast testing results are shown in Table 9. The parameters for the LSTM model with the most accurate test results were four (4) lags, 30 units, and 200 epochs.

Table 9 New Applicant Model Testing Results

MONTH	ACTUAL NEW APPLICANTS		ARIMA FORECAST	ERROR	ABSOLUTE ERROR (DEVIATION)	ABSOLUTE PERCENT ERROR	SQUARED ERROR		LSTM FORECAST	ERROR	ABSOLUTE ERROR (DEVIATION)	ABSOLUTE PERCENT ERROR	SQUARED ERROR
Dec-22	1,310		1,280	30	30	2.29%	902		1,347	-37	37	2.80%	1,342
Jan-23	1,323		1,302	21	21	1.61%	455		1,378	-55	55	4.13%	2,989
Feb-23	1,713		1,345	368	368	21.47%	135,269		1,367	346	346	20.20%	119,766
Mar-23	1,772		1,380	392	392	22.11%	153,452		1,711	61	61	3.46%	3,765
Apr-23	1,185		1,419	-234	234	19.71%	54,578		1,487	-302	302	25.50%	91,309
May-23	1,563		1,456	107	107	6.87%	11,515		1,531	32	32	2.03%	1,009
Jun-23	1,664		1,493	171	171	10.26%	29,149		1,397	267	267	16.06%	71,399
Jul-23	1,500		1,531	-31	31	2.04%	939		1,663	-163	163	10.85%	26,496
Aug-23	1,685		1,568	117	117	6.94%	13,666		1,539	146	146	8.65%	21,246
Sep-23	1,371		1,606	-235	235	17.11%	55,000		1,329	42	42	3.10%	1,805
Oct-23	1,762		1,643	119	119	6.76%	14,173		1,477	285	285	16.16%	81,048
Nov-23	1,402		1,680	-278	278	19.86%	77,495		1,315	87	87	6.17%	7,483
TOTAL TESTING PERIOD	18,250		17,702	548	548	3.00%	299,986		17,540	710	710	3.89%	503,517
MEAN ABSOLUTE DEVIATION (MAD)					175						152		
MEAN ABSOLUTE PERCENTAGE ERROR						11.42%						9.93%	
MEAN SQUARED ERROR							45,549						35,805

The LSTM model predicts the month-to-month variation of new applicants better than the ARIMA model with the lower MAPE and MSE shown in Table 9. One limitation of the LSTM model is the lack of ability to forecast ten (10) fiscal years using solely historical data due to the limited number of post-pandemic data points.

The deliverable of this forecasting project is projections for Fiscal Years 2025 through 2034 of the number of new applicants per year, a composite of the monthly data, which reduces the emphasis on understanding the month-to-month variation. The ARIMA model predicts the overall composite testing period better than the LSTM model. The ARIMA model forecasts Fiscal Years 2025 through 2034 using solely historical data. The ARIMA model is also linear and less susceptible to the risk of a negative bias over the period of ten fiscal years.

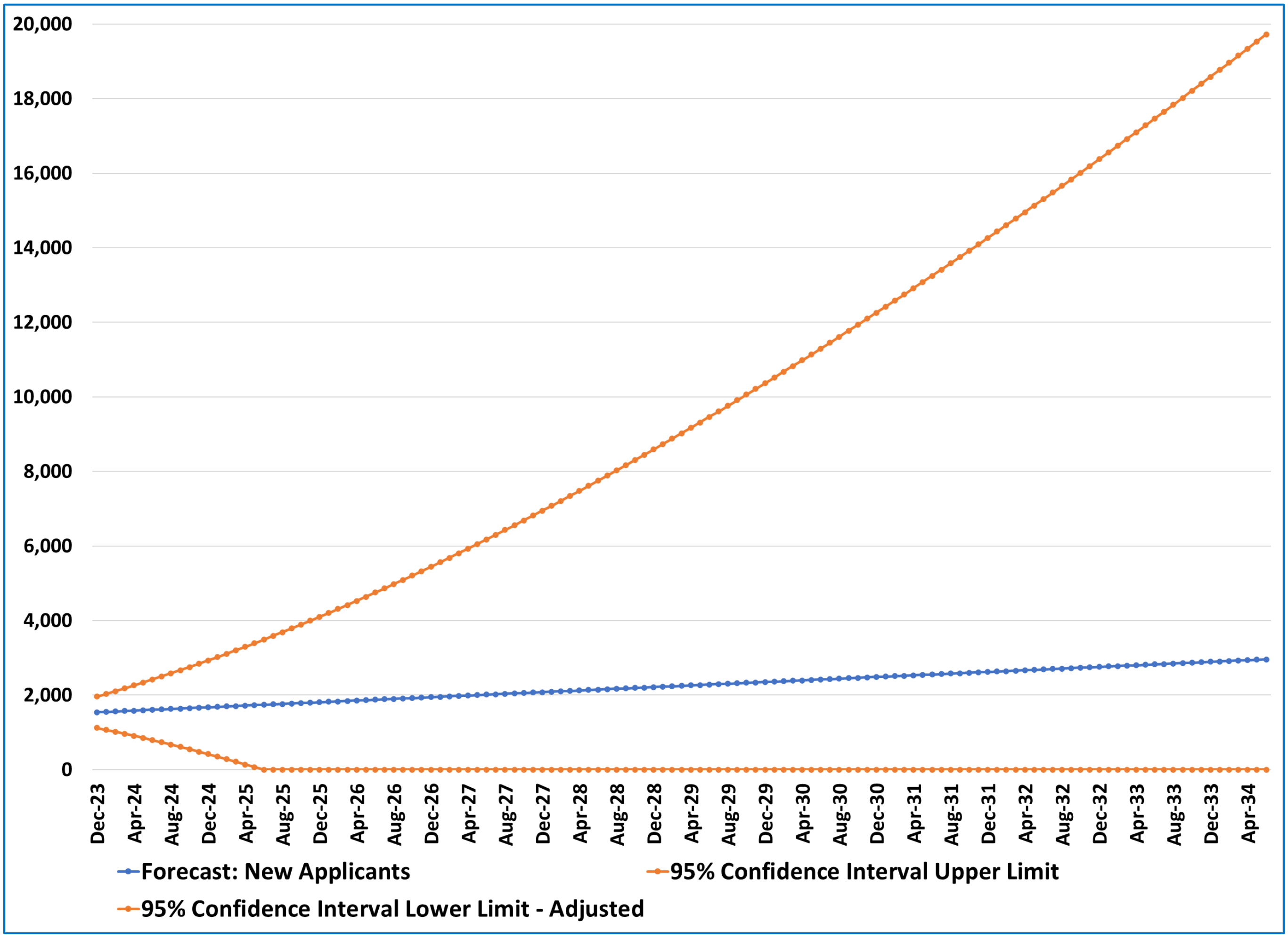
4.2.2. Test Results Summary – New Applicants

Testing of several potential forecasting models led to finding the ARIMA model is the best fit for the new applicant data. The finding was shared with Access Services in a virtual meeting, where the discoveries were validated and confirmed. Then projections were developed for Fiscal Years 2025 through 2034 using the ARIMA model for new applicants.

4.3. Forecasting Assumptions and Risk Analysis – New Applicants

The ARIMA model used to forecast the number of new applicants calculates upper and lower limits for the projected values. These limits form an interval of confidence the actual observations will fit within during the specific period. The monthly forecast of new applicants along with the upper and lower limits for the 95% confidence interval is shown in Figure 23.

Figure 23 Monthly New Applicant 95% Confidence Intervals



The 95% confidence interval provides significantly more certainty than a 90% confidence interval used in previous forecasts, however the interval results in a much larger range between the upper and lower limits to accommodate the sureness. Due to the proximity of the forecast of monthly new applicants to the value of zero, and the wide range to accommodate the 95% confidence interval, the lower limit results in negative values beginning in June 2025. There are no observed number of monthly new applicants with a negative number, hence the lower limit was adjusted to zero beginning in June 2025 since a negative value is not possible.

4.4. Conclusion and Next Steps – New Applicants

The initial analysis of the number of new applicants reveals essential insight into current trends. The trend is significantly different after the COVID-19 pandemic than the trend prior to the pandemic.

Forecasting models need to utilize post-pandemic historical data for projections to reflect current trends.

Training and testing of several forecasting models led to the selection of the models that best fit current trends for the number of new applicants. The ARIMA model best fits the number of new applicants.

The accuracy of projections for the number of new applicants during the next iteration of forecasting is expected to further improve. These projections are currently based on historical monthly number of new applicants from April 2020 through November 2023. The number of observations of historical monthly number of new applicants will increase 27% by November 2024. This increase creates both additional training and testing data, which will enhance confidence and accuracy in the projection model.

Other steps to improve the accuracy of projections for the number of new applicants during the next iteration involve further testing of other forecasting models such as LSTM. Testing the LSTM model will incorporate detrending methods such as Empirical Mode Decomposition (EMD) as well as transformers.

Next steps include considering additional forecasting and machine learning models and testing them as appropriate. The ARIMA model selected to develop projections for the number of new applicants is univariate time series forecasting. This means the only factors considered in the model of eligible rider behavior are the dependent variable, number of trip requests, and the time.

The LSTM model also has the capability for multivariate time series forecasting, which means independent variables and their effect on the number of trip requests are considered. Examples of other independent variables that could be included in a multivariate LSTM model include the unemployment rate, fuel prices, population, and more. This model will be considered based upon the performance of the ARIMA model along with an initial analysis of other independent variables.

The Meta (Facebook) Prophet model is a modular regression model with features such as the ability to model multiple seasonalities and the ability to identify changepoints, when a change occurs in the data (Taylor, 2017). There are options to choose the growth function (linear, logistic, flat) in the Prophet model as well as a function to handle drastic changes in values for holidays and events (Taylor, 2017). The Prophet model will be considered based upon the performance of the ARIMA univariate model along with an initial analysis of other independent variables.

Additional models to be evaluated and considered include the Koopman Filter for predicting the number of new applicants.

Monitoring of the demand dynamics on future iterations will continue to determine if the data sets revert to pre-pandemic values and patterns.

5. Scheduled Trip Demand Forecasts

Access Services uses scheduled trip demand to plan budgets and operations for upcoming fiscal years. Table 10 shows the monthly forecasts through Fiscal Year 2025.

Table 10 Scheduled Trip Demand Monthly Forecast - Fiscal Years 2024 through 2025

MONTH	TOTAL	Antelope Valley	Eastern	Santa Clarita	SF Valley	Southern	West/ Central
Jul-22	262,392	13,310	70,078	3,176	38,668	93,509	43,651
Aug-22	284,786	15,199	76,546	3,374	41,992	100,417	47,258
Sep-22	287,099	14,644	77,985	2,973	41,953	102,024	47,520
Oct-22	305,537	15,733	84,250	3,330	44,544	106,104	51,576
Nov-22	287,324	14,182	79,728	3,035	41,955	99,595	48,829
Dec-22	288,377	14,403	78,046	3,291	42,910	101,515	48,212
Jan-23	293,119	14,607	78,887	3,223	42,443	104,312	49,647
Feb-23	283,804	14,056	77,594	2,885	41,579	99,759	47,931
Mar-23	331,851	15,881	91,703	3,416	48,914	115,951	55,986
Apr-23	315,771	15,338	88,395	3,094	46,046	110,309	52,589
May-23	337,826	15,822	94,669	3,321	49,191	117,834	56,989
Jun-23	326,934	15,173	93,649	3,193	46,845	112,866	55,208
Jul-23	322,123	14,530	92,622	3,604	46,907	109,963	54,497
Aug-23	351,654	16,553	100,282	3,776	51,050	121,205	58,788
Sep-23	342,300	15,276	98,144	3,611	50,140	119,148	55,981
Oct-23	361,847	16,342	105,882	1,765	52,966	125,755	59,137
Nov-23	342,149	16,703	93,084	3,848	47,652	123,446	57,416
Dec-23	359,307	17,540	97,752	4,041	50,042	129,637	60,295
Jan-24	371,089	18,115	100,957	4,174	51,683	133,887	62,272
Feb-24	345,660	16,874	94,039	3,888	48,141	124,713	58,005
Mar-24	376,416	18,375	102,406	4,234	52,425	135,809	63,166
Apr-24	377,577	18,432	102,722	4,247	52,587	136,228	63,361
May-24	398,095	19,434	108,304	4,478	55,444	143,631	66,804
Jun-24	382,208	18,658	103,982	4,299	53,231	137,899	64,138
Jul-24	413,944	20,207	112,616	4,656	57,651	149,349	69,464
Aug-24	412,858	20,154	112,320	4,644	57,500	148,958	69,281
Sep-24	362,429	17,693	98,601	4,076	50,477	130,763	60,819
Oct-24	385,511	18,819	104,881	4,336	53,692	139,091	64,692
Nov-24	393,089	19,189	106,942	4,421	54,747	141,825	65,964
Dec-24	418,208	20,416	113,776	4,704	58,245	150,888	70,179
Jan-25	428,833	20,934	116,666	4,823	59,725	154,721	71,962
Feb-25	379,110	18,507	103,139	4,264	52,800	136,781	63,618
Mar-25	410,355	20,032	111,640	4,616	57,152	148,055	68,862
Apr-25	403,572	19,701	109,794	4,539	56,207	145,607	67,723
May-25	425,819	20,787	115,847	4,789	59,305	153,634	71,456
Jun-25	423,729	20,685	115,278	4,766	59,014	152,880	71,106

* Actual Scheduled Trips (Trip Requests)

Access Services uses ten (10) year projections for scheduled trip demand for long-range planning activities such as strategic planning, capital purchase planning, and other operational decisions. The scheduled trip demand annual forecasts through Fiscal Year 2034 are shown in Table 11.

Table 11 Scheduled Trip Demand Annual Forecast - Fiscal Years 2024 through 2034

Fiscal Year	TOTAL	Antelope Valley	Eastern	Santa Clarita	SF Valley	Southern	West/ Central
2023*	3,604,820	178,348	991,530	38,311	527,040	1,264,195	605,396
2024**	4,330,423	206,833	1,200,175	45,965	612,268	1,541,322	723,860
2025	4,857,458	237,126	1,321,500	54,635	676,516	1,752,553	815,127
2026	5,386,553	262,955	1,465,444	60,586	750,205	1,943,448	903,914
2027	5,660,881	276,347	1,540,076	63,672	788,412	2,042,425	949,949
2028	5,994,824	292,649	1,630,927	67,428	834,921	2,162,910	1,005,988
2029	6,259,951	305,592	1,703,057	70,410	871,846	2,258,567	1,050,479
2030	6,411,324	312,982	1,744,239	72,112	892,929	2,313,182	1,075,881
2031	6,491,570	316,899	1,766,070	73,015	904,105	2,342,134	1,089,347
2032	6,892,726	336,482	1,875,207	77,527	959,975	2,486,870	1,156,664
2033	6,942,227	338,899	1,888,674	78,084	966,870	2,504,730	1,164,971
2034	6,950,682	339,311	1,890,974	78,179	968,047	2,507,780	1,166,390

* Actual Scheduled Trips (Trip Requests)
** Projections Include Actual Scheduled Trips (Trip Requests) through October 31, 2023

Scheduled trip demand has a positive trend as the demand increases from one fiscal year to the next. The percentage increase (or decrease) for the scheduled trip demand from one year to the next are shown in Table 12.

Table 12 Scheduled Trip Demand Annual Forecast % Increase (Decrease) From Prior Year

Fiscal Year	TOTAL	Antelope Valley	Eastern	Santa Clarita	SF Valley	Southern	West/ Central
2023							
2024	20.1%	16.0%	21.0%	20.0%	16.2%	21.9%	19.6%
2025	12.2%	14.6%	10.1%	18.9%	10.5%	13.7%	12.6%
2026	10.9%	10.9%	10.9%	10.9%	10.9%	10.9%	10.9%
2027	5.1%	5.1%	5.1%	5.1%	5.1%	5.1%	5.1%
2028	5.9%	5.9%	5.9%	5.9%	5.9%	5.9%	5.9%
2029	4.4%	4.4%	4.4%	4.4%	4.4%	4.4%	4.4%
2030	2.4%	2.4%	2.4%	2.4%	2.4%	2.4%	2.4%
2031	1.3%	1.3%	1.3%	1.3%	1.3%	1.3%	1.3%
2032	6.2%	6.2%	6.2%	6.2%	6.2%	6.2%	6.2%
2033	0.7%	0.7%	0.7%	0.7%	0.7%	0.7%	0.7%
2034	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%

6. New Applicant Forecasts

Projections for the annual number of new applicants are shown in Table 13. These projections include ten (10) fiscal years from Fiscal Year 2025 through 2034. It should be noted that projections include the remainder of Fiscal Year 2024.

Table 13 New Applicant Annual Forecast - Fiscal Years 2024 through 2034

Fiscal Year	TOTAL	Antelope Valley	Eastern	Santa Clarita	SF Valley	Southern	West/ Central
2023*	17,277	779	5,370	160	2,390	5,779	2,799
2024**	18,737	898	5,591	193	2,595	6,359	3,102
2025	20,173	966	6,019	208	2,794	6,846	3,340
2026	21,799	1,044	6,504	225	3,019	7,398	3,609
2027	23,424	1,122	6,989	241	3,244	7,949	3,878
2028	25,049	1,200	7,474	258	3,469	8,501	4,147
2029	26,674	1,278	7,959	275	3,694	9,052	4,416
2030	28,299	1,356	8,444	292	3,919	9,604	4,685
2031	29,924	1,434	8,929	308	4,144	10,155	4,954
2032	31,549	1,511	9,414	325	4,369	10,707	5,223
2033	33,174	1,589	9,899	342	4,594	11,258	5,492
2034	34,799	1,667	10,384	359	4,819	11,810	5,761

* Actual New Applicants
** Projections Include Actual New Applicants through November 30, 2023

The annual number of new applicants has a positive trend as the demand increases from one fiscal year to the next. The percentage increase (or decrease) for the number of new applicants from one year to the next are shown in Table 14.

Table 14 New Applicant Annual Forecast % Increase (Decrease) From Prior Year

Fiscal Year	TOTAL	Antelope Valley	Eastern	Santa Clarita	SF Valley	Southern	West/ Central
2023							
2024	8.5%	15.2%	4.1%	20.7%	8.6%	10.0%	10.8%
2025	7.7%	7.7%	7.7%	7.7%	7.7%	7.7%	7.7%
2026	8.1%	8.1%	8.1%	8.1%	8.1%	8.1%	8.1%
2027	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%	7.5%
2028	6.9%	6.9%	6.9%	6.9%	6.9%	6.9%	6.9%
2029	6.5%	6.5%	6.5%	6.5%	6.5%	6.5%	6.5%
2030	6.1%	6.1%	6.1%	6.1%	6.1%	6.1%	6.1%
2031	5.7%	5.7%	5.7%	5.7%	5.7%	5.7%	5.7%
2032	5.4%	5.4%	5.4%	5.4%	5.4%	5.4%	5.4%
2033	5.2%	5.2%	5.2%	5.2%	5.2%	5.2%	5.2%
2034	4.9%	4.9%	4.9%	4.9%	4.9%	4.9%	4.9%

The mean average annual increase in the projected number of new applicants is 6.6%.

7. Reference List

Ahmed, M.Y., & Sarkodie, S.A. (December 2021) COVID-19 pandemic and economic policy uncertainty regimes affect commodity market volatility. *Resources Policy, 74*, Article 102393. https://doi.org/10.1016/j.resourpol.2021.102303

Bedi, J., and Toshniwal, D. (2018) Empirical Mode Decomposition Based Deep Learning for Electricity Demand Forecasting, in IEEE Access, vol. 6, pp. 49144-49156

Bordoni, S., & Giagu, S. (2023). Convolutional neural network based decoders for surface codes. *Quantum Information Processing*, *22*(3), 151.

Brown, S. (2021, January 5). *How COVID-19 is disrupting data analytics strategies.* MIT Management Sloan School. https://mitsloan.mit.edu/ideas-made-to-matter/how-COVID-19-disrupting-data-analytics-strategies#:~:text=COVID%2D19%20is%20upending%20data,fallout%20is%20causing%20further%20changes.

Calzone, O. (2022, April 10). An Intuitive Explanation of LSTM. Medium. https://medium.com/@ottaviocalzone/an-intuitive-explanation-of-lstm-a035eb6ab42c

Camm, J.D., Davenport, T.H., & Heichler, E. 2020. *Machine learning and data analytics in the pandemic era.* [Webinar]. https://sloanreview.mit.edu/video/machine-learning-and-data-analytics-in-the-pandemic-era/?utm_source=newsletter&utm_medium=email&utm_content=Watch%20on%20demand%20%C2%BB&utm_campaign=WebinarSAS1120

Cherif, S. (2021, March 29). *Post covid world: can historical data predict the future?* LinkedIn. https://www.linkedin.com/pulse/post-covid-world-can-historical-data-predict-future-seddik-cherif/

Gilbert, C., Bever, L. (2022, March 11). *Gyms, pets and takeout: How the pandemic has shifted daily life.* Washington Post. https://www.washingtonpost.com/wellness/interactive/2023/pandemic-changes-daily-life/

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Ivy Professional School. (2022, July 27). *Important impact of data science post covid we should know.* https://ivyproschool.com/blog/data-science-post-covid/

Korstanje, J. (2021). Advanced Forecasting with Python: With State-of-the-Art-Models Including LSTMs, Facebook's Prophet, and Amazon's DeepAR. Apress. https://doi.org/10.1007/978-1-4842-7150-6

Montgomery, D., & Runger, G. (n.d.). Applied Statistics and Probability for Engineers, 7th Edition | Wiley. Wiley.Com. Retrieved February 21, 2024, from https://www.wiley.com/en-us/Applied+Statistics+and+Probability+for+Engineers%2C+7th+Edition-p-9781119400363

Surakhi, O., Zaidan, M. A., Fung, P. L., Hossein Motlagh, N., Serhan, S., AlKhanafseh, M., Ghoniem, R. M., & Hussein, T. (2021). Time-Lag Selection for Time-Series Forecasting Using Neural Network and Heuristic Algorithm. *Electronics, 10*(20), Article 20. https://doi.org/10.3390/electronics10202518

Taylor SJ, Letham B. 2017. Forecasting at scale. PeerJ Preprints 5:e3190v2 https://doi.org/10.7287/peerj.preprints.3190v2

8. Appendix A

8.1.Appendix A-1: Definitions

<u>Dependent variable</u> A Dependent variable is what happens as a result of the independent variable. In other words, a variable (often denoted by y) whose value depends on that of another.

<u>Exponential Growth Trend</u> A time series where values increase by a consistent relative rate (eg. 10% per year on previous year value)

<u>Generalized linear model</u> (**GLM**) flexible generalization of ordinary linear regression. The GLM generalizes linear regression by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its predicted value.

<u>Independent Variable</u> variables that stand on their own and aren't affected by anything that you do. A variable (often denoted by x) whose variation does not depend on that of another.

For example, the weather (rain, snow, temperature, etc.) is independent of fares. Regardless of any increases or decreases in the fare, the temperature will not be affected.

<u>Intercept</u> the distance from the origin to a point where a graph crosses a *y* coordinate axis.

<u>Lags</u> number of layers in an LSTM model.

<u>Linear regression</u> linear approach for modeling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). The case of one explanatory variable is called simple linear regression.

<u>Linear Trend</u> A time series where each data point increases (decreases) by a consistent value and forms a straight line.

<u>Mean absolute deviation (MAD)</u> the absolute difference between the observed and forecasted values.

$$\text{MAD} = \frac{\sum \left| y_i - \hat{y}_i \right|}{N}$$

<u>Mean absolute percent error (MAPE)</u> average error for the absolute difference between the observed and forecasted values.

$$\text{MAPE} = \frac{\sum \left| \frac{y_i - \hat{y}_i}{y_i} \right|}{N}$$

<u>Mean squared error (MSD)</u> the average squared difference between the observed and forecasted values.

$$\frac{\sum_{t=1}^{n} \left| y_t - \hat{y}_t \right|^2}{n}$$

Multivariate analysis of variance (**MANOVA**) extends the analysis of variance to cover cases where there is more than one dependent variable to be analyzed simultaneously; see also Multivariate analysis of covariance (MANCOVA).

Multivariate regression attempts to determine a formula that can describe how elements in a vector of variables respond simultaneously to changes in others. For linear relations, regression analyses here are based on forms of the general linear model. Some suggest that multivariate regression is distinct from multivariable regression, however, that is debated and not consistently true across scientific fields.

Polynomial equation an equation comprised of variables, exponents, and coefficients. The degree of the equation is the value of the largest exponent.

Polynomial regression form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modeled as an n^{th} degree polynomial in x.

Quadratic Trend A time series where values increase (decrease) at a rate that is not constant.

S-Curve (Pearl-Reed Logistic) Trend A time series where values increase exponentially until the saturation causes growth to switch to a linear trend and growth stops at maturity.

Slope a number that describes both the direction and steepness of a line.

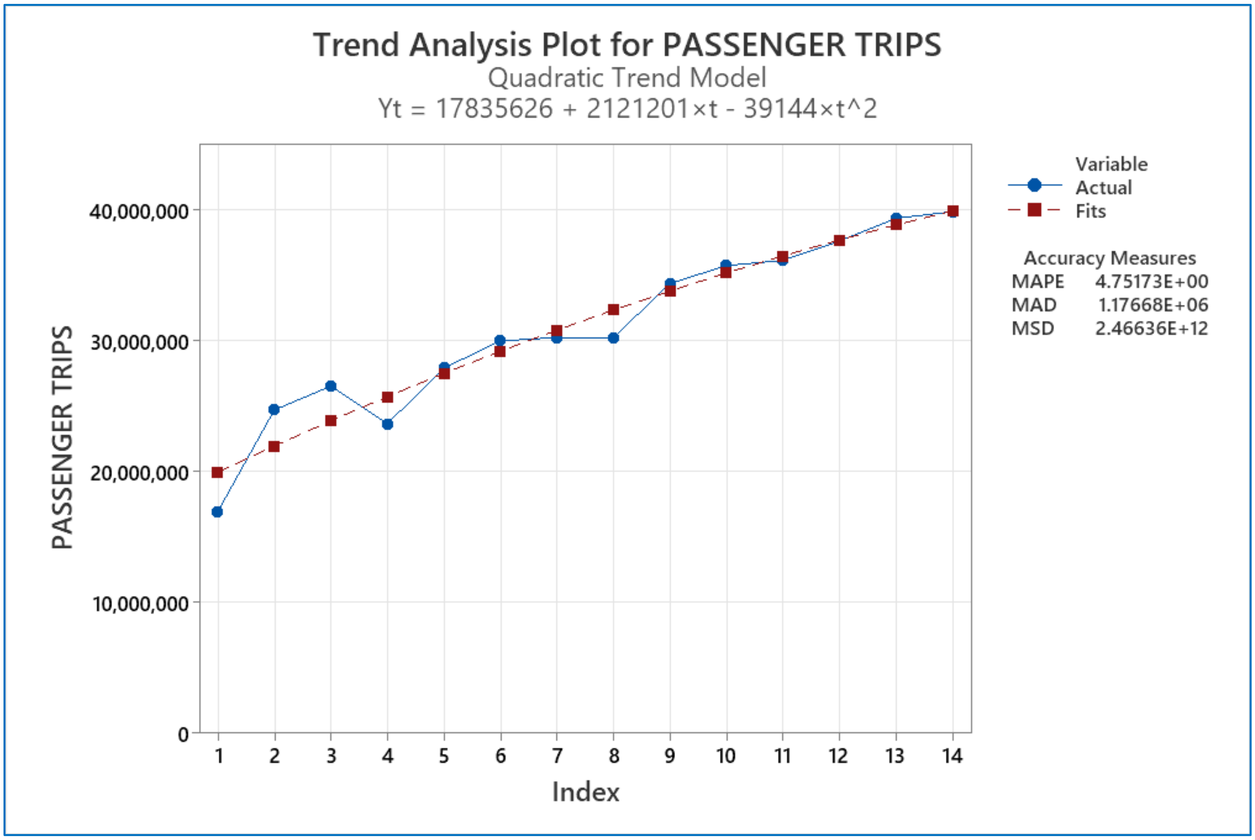
Trend an upwards or downwards shift in a data set over time.

Units number of cells in an LSTM model.

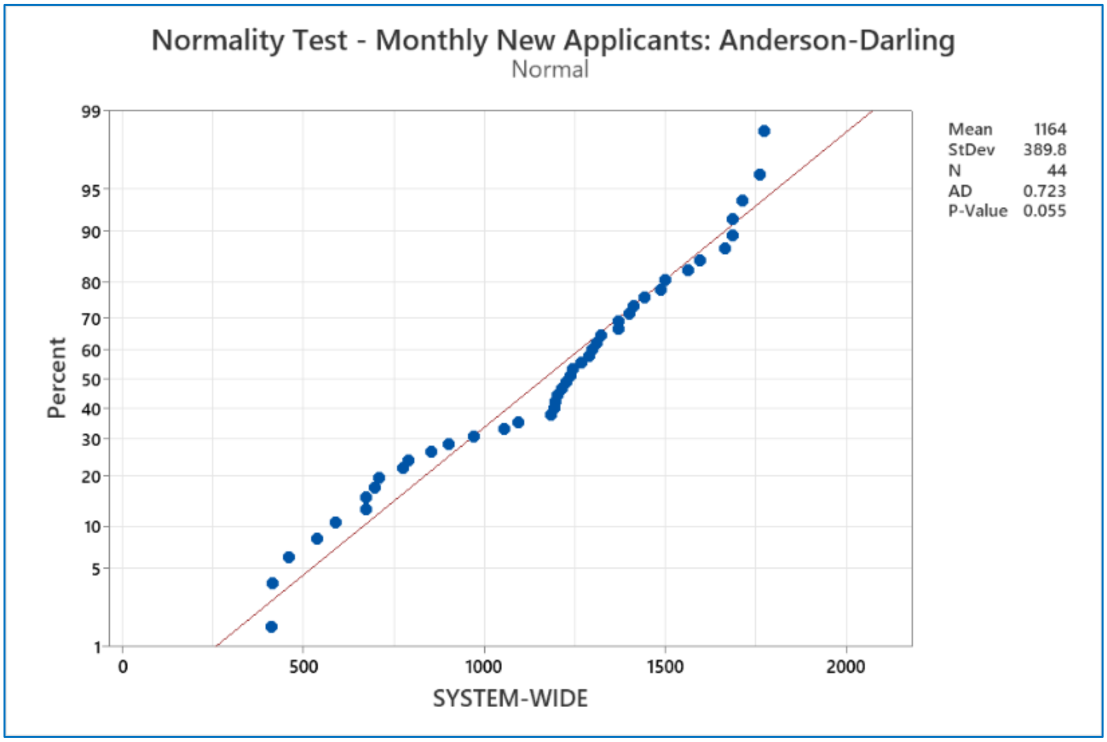
Equation Terms

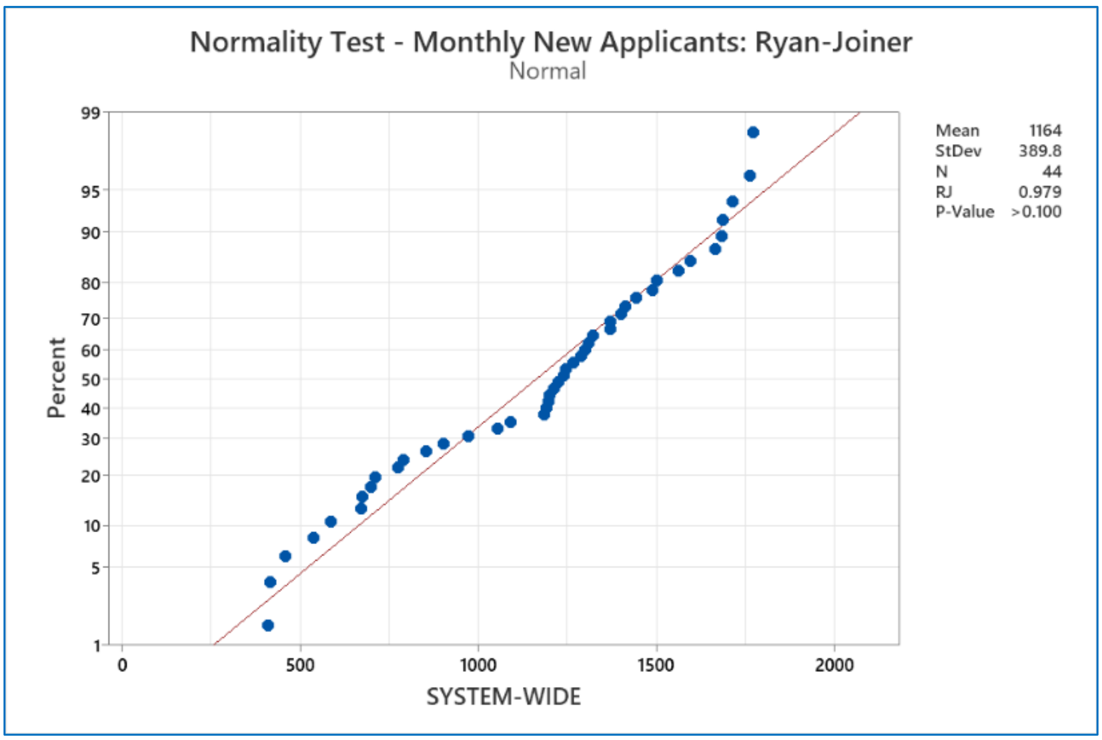
y_i	i^{th} observed response value
$\bar{y}$	mean response
$\hat{y}_i$	i^{th} fitted response
N	number of rows

8.2. Appendix A-2: National Quarterly Ridership Trend Analysis



8.3. Appendix A-3: Normality Test Results – Post-pandemic New Applicants





8.4. Appendix A-4: ARIMA Model Parameters – Results and Analysis

Method

Criterion for best model	Minimum AICc
Rows used	32
Rows unused	0

Model Selection

Model (d = 2)	LogLikelihood	AICc	AIC	BIC
p = 1, q = 1*	-201.401	409.725	408.802	413.005
p = 0, q = 2	-201.554	410.032	409.109	413.313
p = 0, q = 1	-202.947	410.338	409.893	412.696
p = 2, q = 2	-199.685	411.870	409.370	416.376
p = 2, q = 1	-201.540	412.680	411.080	416.685
p = 1, q = 2	-202.679	414.957	413.357	418.962
p = 2, q = 0	-205.077	417.078	416.155	420.358
p = 1, q = 0	-207.705	419.855	419.410	422.213

** Best model with minimum AICc. Output for the best model follows.*

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	-0.388	0.186	-2.09	0.046
MA 1	1.001	0.166	6.04	0.000

Differencing: 2 Regular
Number of observations after differencing: 30

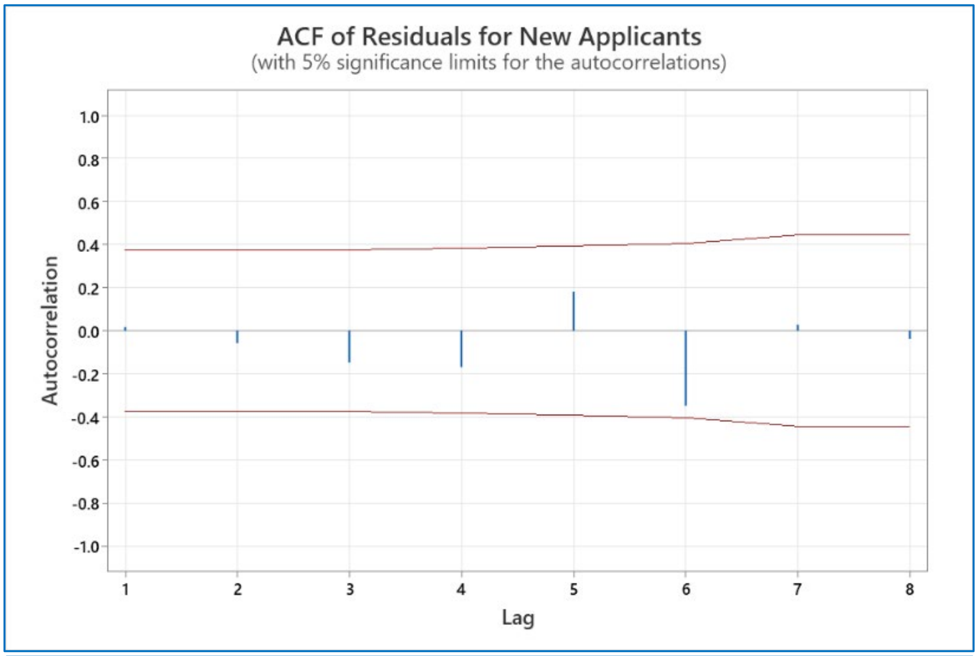
Model Summary

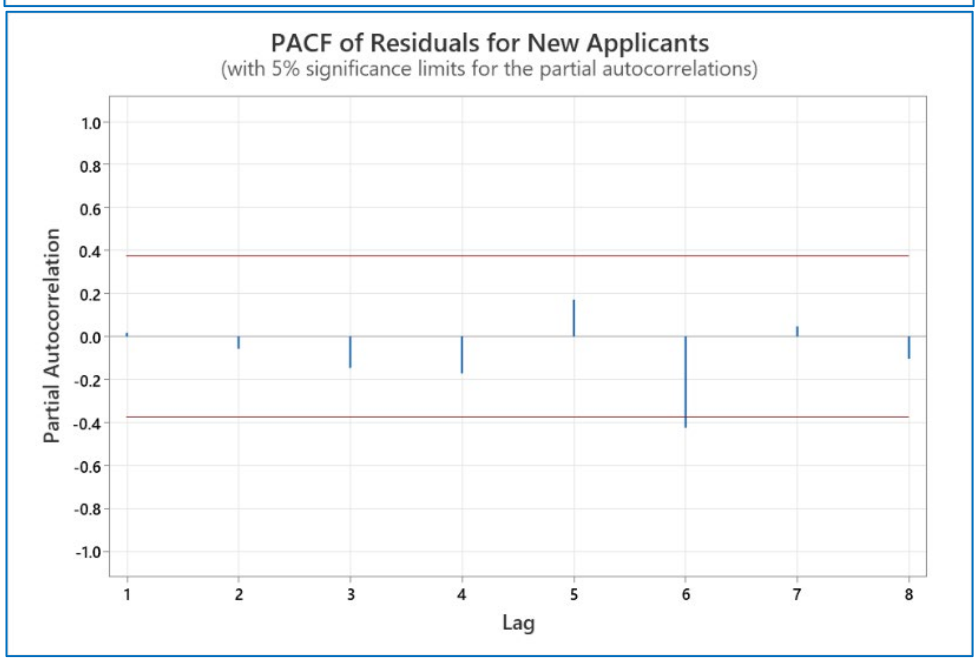
DF	SS	MS	MSD	AICc	AIC	BIC
28	1032941	36890.8	34431.4	409.725	408.802	413.005

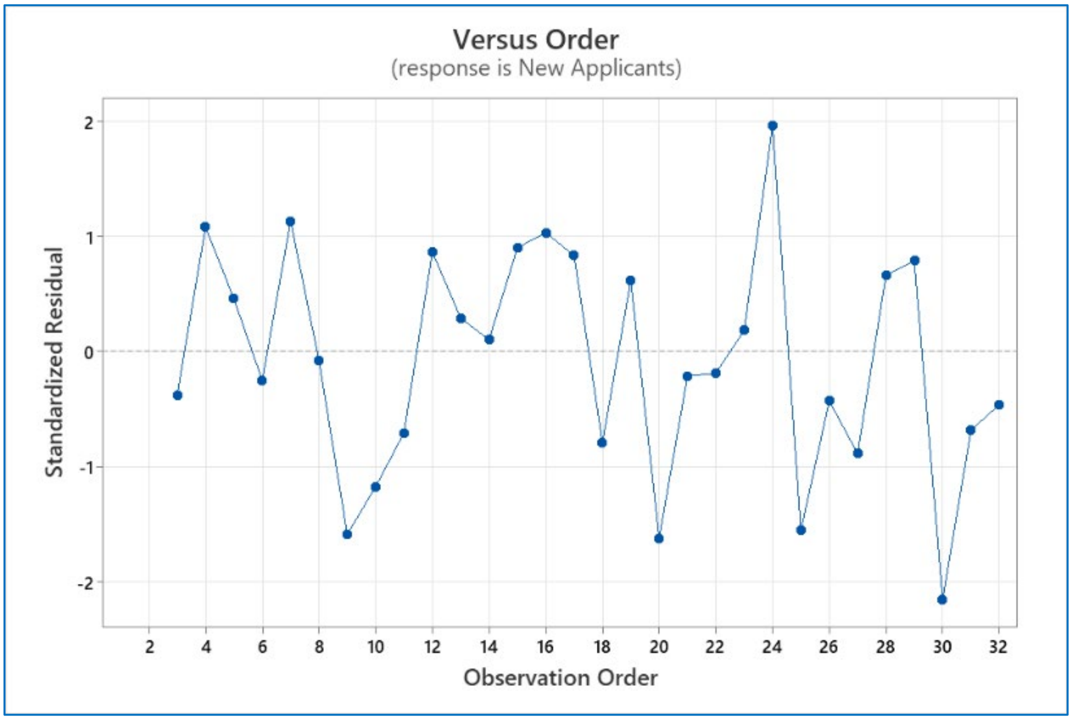
MS = variance of the white noise series

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

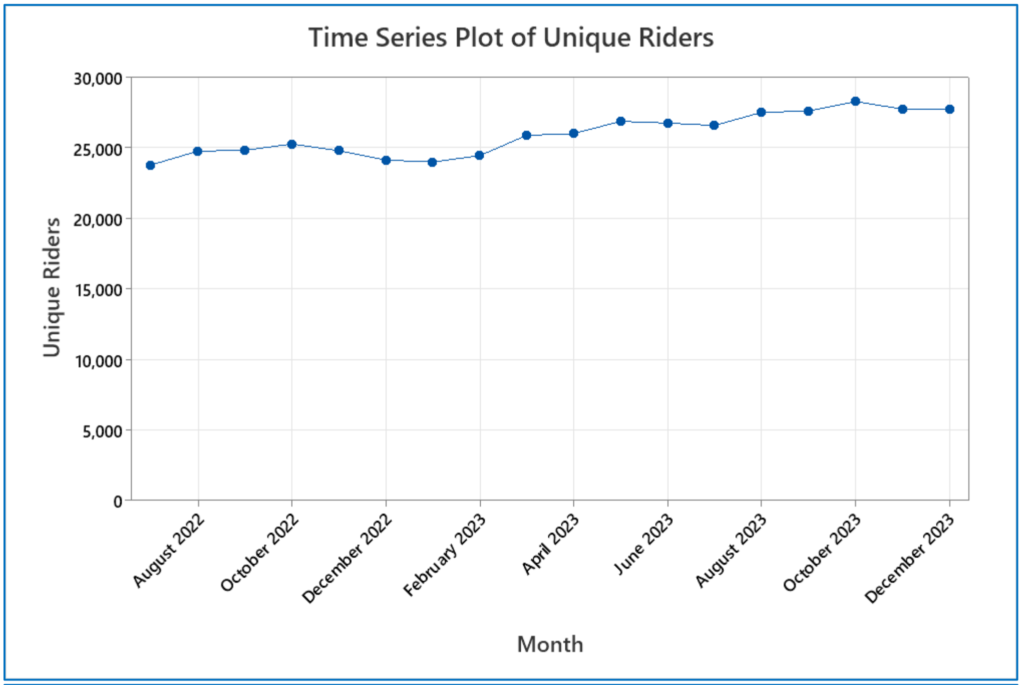
Lag	12	24	36	48
Chi-Square	12.44	23.67	*	*
DF	10	22	*	*
P-Value	0.256	0.365	*	*

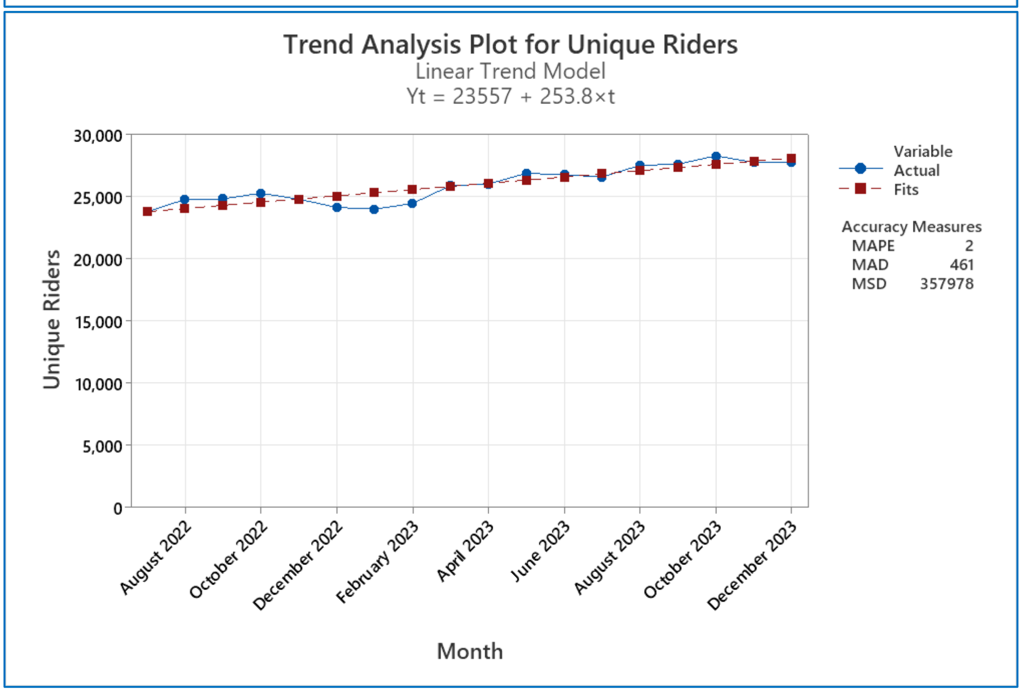

ACF of Residuals for New Applicants
(with 5% significance limits for the autocorrelations)
Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
1
2
3
4
5
6
7
8
Lag

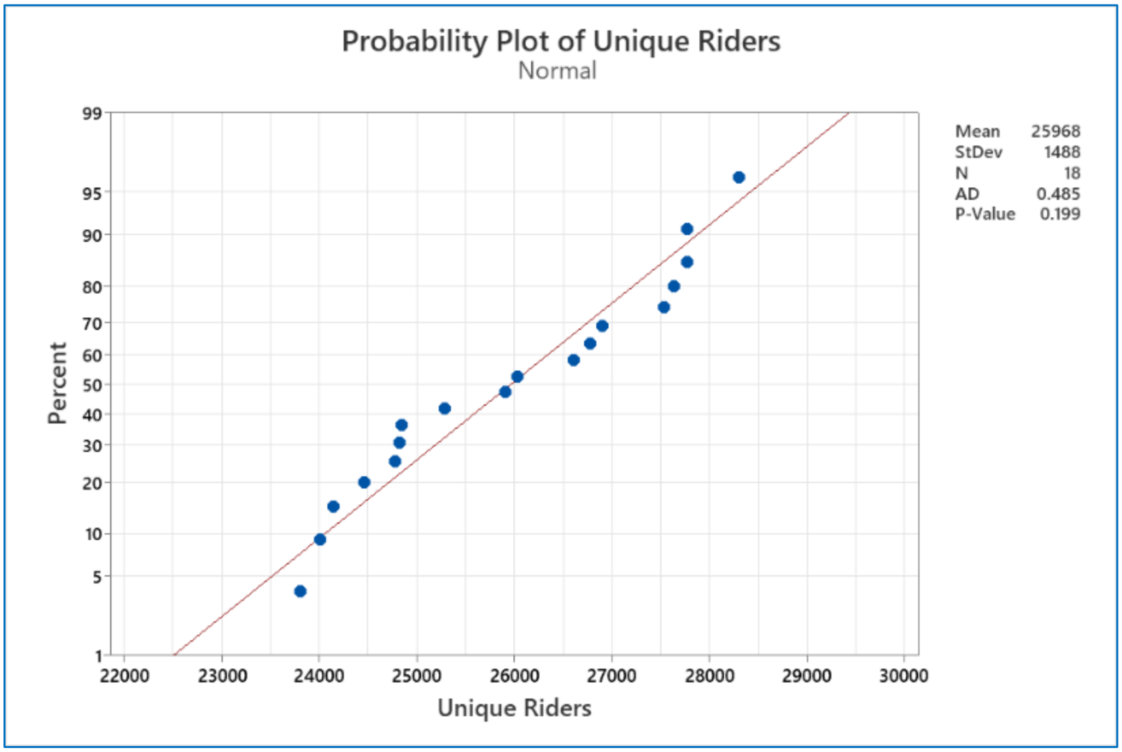

PACF of Residuals for New Applicants
(with 5% significance limits for the partial autocorrelations)
Partial Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
1
2
3
4
5
6
7
8
Lag


Versus Order
(response is New Applicants)
Standardized Residual
Observation Order
2
1
0
-1
-2
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32

8.5.Appendix A-5: Analysis of Unique Riders






Probability Plot of Unique Riders
Normal
Mean 25968
StDev 1488
N 18
AD 0.485
P-Value 0.199
Percent
Unique Riders
99
95
90
80
70
60
50
40
30
20
10
5
1
22000
23000
24000
25000
26000
27000
28000
29000
30000