

# Call Center Regression Data Analysis

Michael Marcaccio

November 14 2022

## Abstract

With over 15 million people employed, a predicted compound annual growth rate (CAGR) of 5.6 percent between 2020 and 2027, and with over 28,000 locations in the United States alone, call centers play a pivotal role in a business' success. In this paper, call center volume was forecasted and models were created to predict the number of agents needed to meet critical attributes such as waittime, calltime, and holdtime. Forecasting gives businesses the ability to make informed business decision and develop data-driven strategies. In the literature it is very common to see call center volume being predicted using different techniques, but there is very limited studies on attempting to correlate the number of agents needed. Through Regression techniques, there are 8 models proposed to model the number of agents needed based on waittime, calltime, goaltime, and the amount of calls handled.

## Introduction

With over 15 million people employed, a predicted compound annual growth rate (CAGR) of 5.6 percent between 2020 and 2027, and with over 28,000 locations in the United States alone, call centers play a pivotal role in a business' success. Call centers are a key part of customer service that will save a company time, money, and unnecessary obstacles. In the banking industry, calls can range from inquires, transfers, payments, reporting, to processing. This means members could call about their account balance, credit card bills, loan applications, or unauthorized transactions. It is crucial that a bank is prepared for spikes in calls and have agents knowledgeable in all aspects of banking.

Many studies have been done working with call center volumes as presented in Modeling and Forecasting Call Center Arrivals (Ibrahim et al., 2016). Here an Autoregressive moving average (ARIMA) model is used, which is depicted in the paper as:

$$\Phi(B)(x_i - \mu) = \theta_q(B)\varepsilon_t \quad (1)$$

Ibrahim uses multiple ARIMA models to depict seasonality with a combination of exponential smoothing. Holt-Winters smoothing is also used, with three equations of:

$$M_t = \alpha_0(X_t - S_{t-s}) + (1 - \alpha_0)(M_{t-1} + B_{t-1}), \quad (2)$$

$$B_t = \alpha_1(M_t - M_{t-1}) + (1 - \alpha_1)B_{t-1}, \quad (3)$$

$$S_t = \alpha_2(X_t - M_t) + (1 - \alpha_2)S_{t-s}, \quad (4)$$

where  $B_t$  is the slope component,  $M_t$  is the level component,  $S_t$  is the seasonal component, and  $s$  is the period of seasonality. ARIMA models are a great model to create as “they allow the representation of a wide array of potentially useful predictor functions in models which contain relatively few parameters” (Newbold, 1983). With the combination of Holt-Winters smoothing, Ibrahim creates a great model for forecasting call volume, however the paper never goes into detail about how many agents there should be working at a set time.

(Evensen et al., 1999) goes into detail about effective service delivery stating the expectations of customers are a “function of the customer’s own experiences...and when judging their own service quality, financial institutions need to evaluate themselves on objective measures which span across industries”. When taking this into consideration, I focused my models on using a calls per agent approach, so agents are not overworked and can treat each caller with respect while answering their concerns in a timely manner to represent the business in a good light. (Avramidis and L’Ecuyer, 2005) confirms this as it is described as the “call-to-agent assignment problem”, and that an efficiency-driven call center is important.

The contributions of the paper creates models on the basis to maximize quality of service in relation to the number of agents. Based on this, models were created to minimize:

1. Wait Time
2. Call Time
3. Reduce the Amount of Agent Handle Time
4. Reduce the Amount of Calls not being handled

The remainder of the paper will present the dataset and break it into a clear view on the historical data. I have also created a forecast on the projected call volume throughout a week on 15 minute intervals.

## Data

The data comes from an undisclosed bank where I wrote two SQL queries to obtain. This comes directly from the banks records dating back from February 2020 to September 2022, and I was lucky enough to do a data analysis on real-life data. This dataset consists of 466,565 observations of 31 variables. Some significant variables consist of Calldate, Caller\_ID, Caller\_ID, QueuedYN, AnsweredYN, AbandonedYN, QueuedName, InteractionOutcome, AnsweredGoalMet, StartTime, EndTime, WaitTime, Holdtime, AgentTalkDuration, AgentHandleTime, WrapUpTime, DayoftheYear, WeekoftheYear, and NbrofAgents.

It is important to acknowledge that AgentTalkDuration refers to the actual amount of time the agent the agent was talking while AbandonedYN refers to if the caller ended the call before speaking to a representative. The QueuedYN is if the member was put in a queue and had to wait to speak to an agent. The queue is an automated service or an IVR (Interactive Voice Response). The latest generation of speech-recognition technology allows IVRs to interpret complex user commands, so customers may be able to “self-serve”, i.e.,

complete the service interaction at the IVR. (Avramidis and L'Ecuyer, 2005). Holdtime is how long the member waits when an agent places on them hold, while waittime is how long they have to wait before someone answers there call. WrapUpTime refers to how long the member takes to hang up the phone after the agent stops talk. AgentHandleTime is the total of Holdtime, WrapUpTime, and AgentTalkDuration combined.

This data is allows me to easily answer the research question as I am using real-life data. With a huge sample size and many variables, there is a great opportunity for data analysis and modeling. With this dataset, I know when doing calculations such as finding means and creating plots, they will able to accurately give a truly image of call center volume and agent information to a small period of time such as 15 minutes. I did do some exploratory data analysis by creating figures that represent call center volume by the year, month, day of week, and time. I also created a table that shows the average wait time, talk time, hold time, wrap up time, and count of every interaction Outcome.

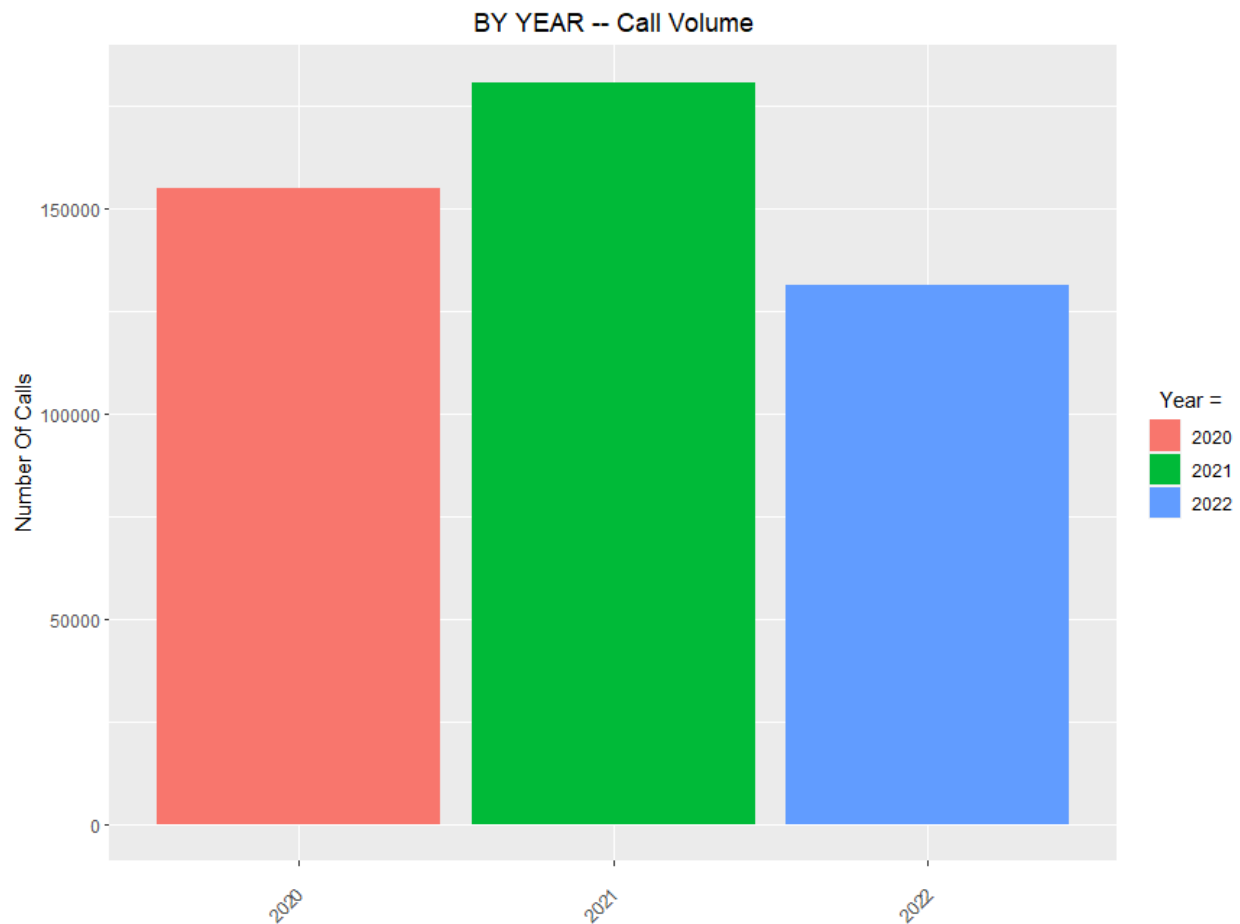


Figure 1: Call Volume Grouped by Year.

Here we can see that 2021 has the highest call volume. Please Note that 2020 only has 11 months, while 2021 only has 10. This may be why the data appears skewed.

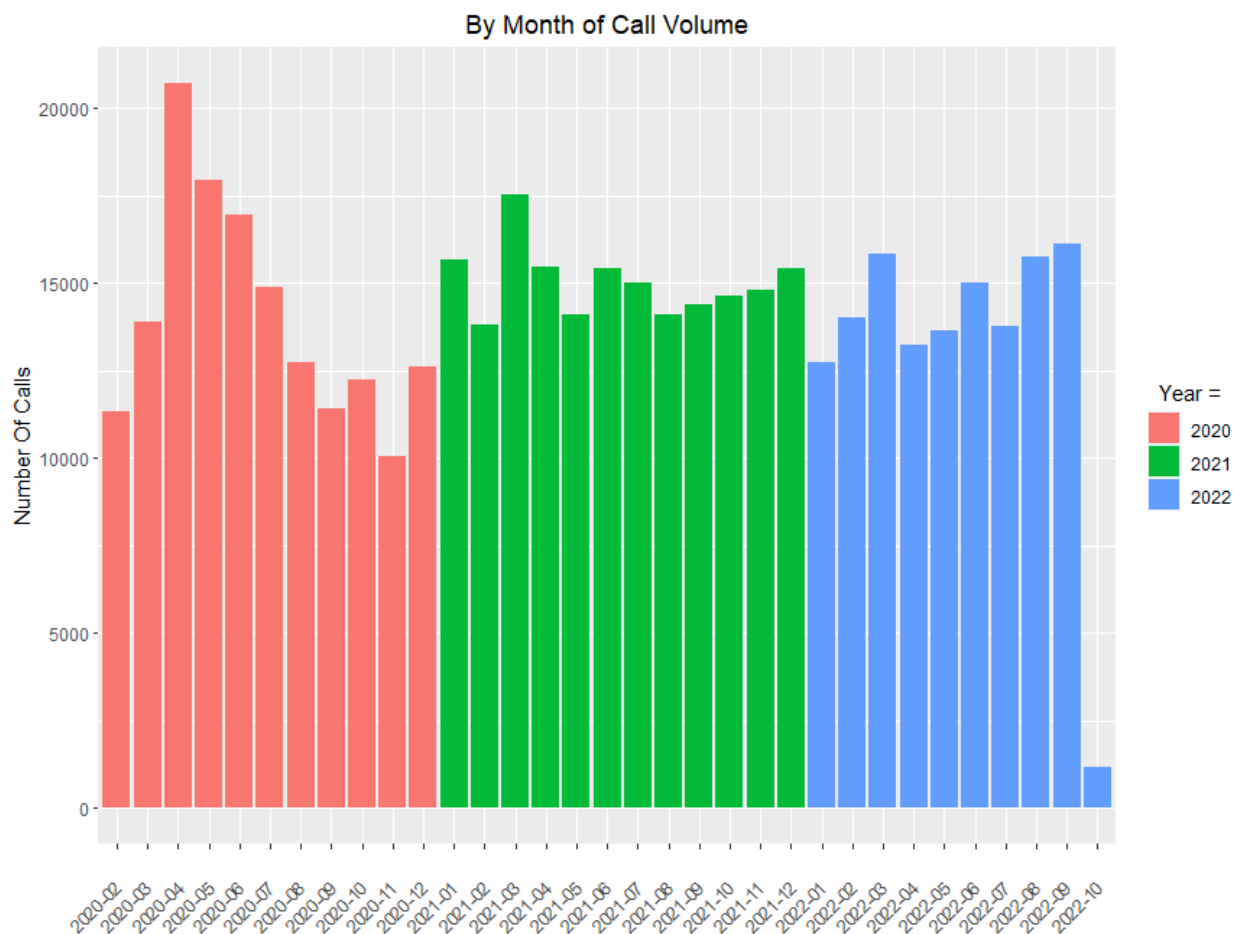


Figure 2: Call Volume Grouped by Month.

When grouped by month, we can see an interesting peak around April. This may be because that is tax season and members are calling for tax information. Also April 2020 may have a higher peak than April 2021, due to COVID-19. During that time, many individuals were eligible for a 1,200 dollar stimulus check.

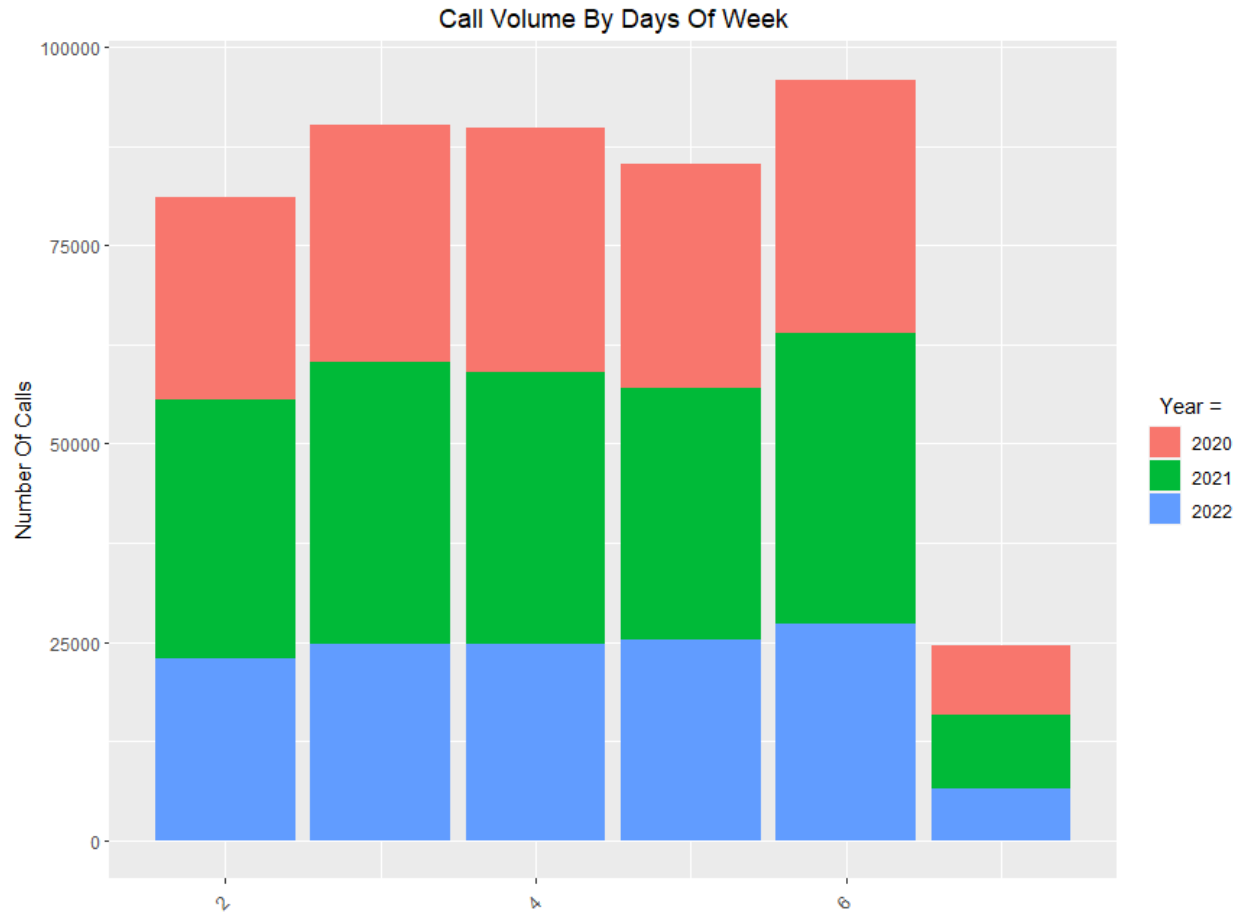


Figure 3: Call Volume Grouped by Day.

With Sunday representing 1, we can see that most calls happen on a Friday. Saturday has an extremely low count. This can be explained due to the varying hours of the call center. Monday through Thursday, the call center is open 8am to 5pm, a total of 9 hours. Friday the call center is open 8am-7pm, a total of 11 hours. Saturday the call center is only open 8:30 am to 12pm, a total of 3 and a half hours.

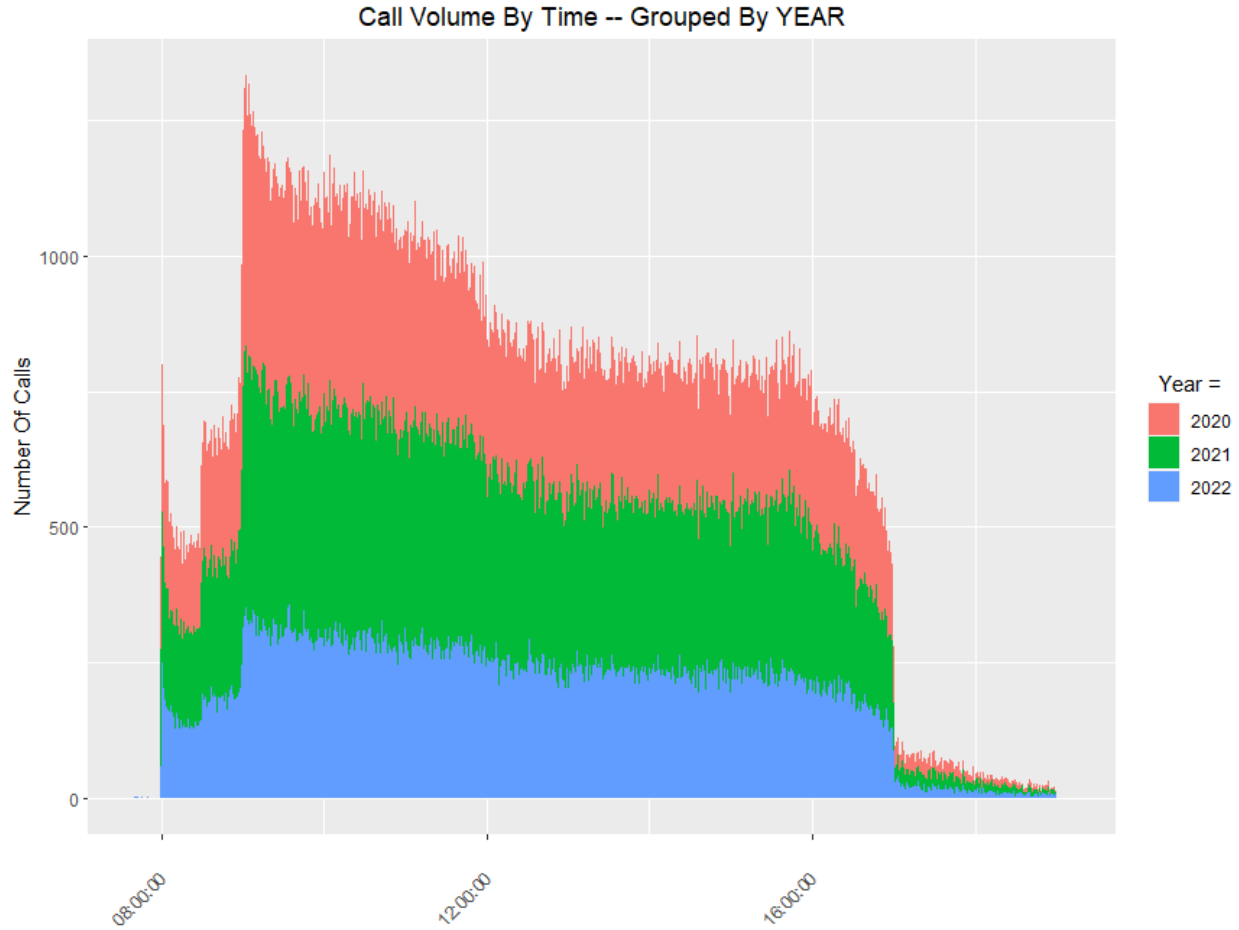


Figure 4: Call Volume Grouped by Time.

The call center consistently has a spike around 9am throughout the 3 years. This is suprising as the call center opens at 8am. This is important to be aware of when creating a forecasting model and having the banks employees be at their station at 9am. The decreasing trend throughout the day is not suprising as most people are working throughout the day.

Outcome	Count	Wait Avg	Talk Avg	Hold Avg	Wrap Avg	Total Talk Avg
Abandoned	2364	100	11.3	0.06	2.40	13.8
Handled	460047	107	253	26.2	16.3	295
Leave Number	411	214	66	6.18	7.80	80
Not Specified	299	1.21	120	0.278	3.24	123
Service Unavailable	1	0	0	0	0	0
Wrong Number	3	357	0	0	12	12

Here it is prevalent that a majority of calls are handled. These variables played a key part in my models for the agents.

## Methods

I created linear models to show the relationship between calls per agent and average waittime, holdtime, average goal met, and average interaction value. Average goal met was a variable created in relationship to the AnsweredGoalMet. AnsweredGoalMet was given in 1 and 0s, 1 if the goal was hit and 0 if it was not hit. This variable is very fluid, as each company has its own reasoning for what they want their goal time to be. Average Interaction Value was based off a variable that was created based on the InteractionOutcome. “Handled ”got assigned 1 as it was a positive outcome and -1 for “Abandoned” and “Leave Number” as they were deemed negative. The others were considered neutral, for example “Not Specified”, was given a 0. Linear Regression is based off five key assumptions, there is a linear relationship, multivariate normality, No or little multicollinearity, No auto-correlation, and Homoscedasticity. multicollinearity refers to

## Simulation

## Discussion

## References

- Avramidis, A. N. and P. L’Ecuyer (2005). Modeling and simulation of call centers. In *Proceedings of the Winter Simulation Conference, 2005.*, pp. 9–pp. IEEE.
- Evensen, A., F. X. Frei, and P. T. Harker (1999). *Effective call center management: evidence from financial services*. Citeseer.
- Ibrahim, R., H. Ye, P. L’Ecuyer, and H. Shen (2016). Modeling and forecasting call center arrivals: A literature survey and a case study. *International Journal of Forecasting* 32(3), 865–874.
- Newbold, P. (1983). Arima model building and the time series analysis approach to forecasting. *Journal of forecasting* 2(1), 23–35.