Forecasting Temperature for Baku, Azerbaijan

By Monica Stettler, Cynthia Cho, and Rafael Leon Espinoza

Introduction

Temperature is just one facet of what makes up the weather, but it is the main measure people use to set their expectations for their day, based on how hot or cold the day is forecasted to be. Temperature not only affects everyday life, but it impacts industries, such as farming and insurance. It is undeniable the significance and impact that temperature has on individuals, and companies' business decisions. Given the important role of temperature, the objective of this project is to create a time series model to forecast temperature. We obtained data for cities all over the world and chose to train and evaluate our models with data from Baku, Azerbaijan.

To find the best forecasting method, multiple time series techniques were tested: SARIMA, time series regression, and a neural network using an LSTM approach. We had initially proposed to test ARMA, ARIMA, SARIMA, and GARCH. However, once we completed our exploratory analysis and having recently learned in class about GARCH, we determined that SARIMA was the only one of those that would be appropriate. For relevance to the course, SARIMA and time series regression were chosen because both are capable of accounting for seasonality in the data. The choice to use neural network was because LSTM, a member of the recurrent neural network family, can handle sequential data by creating both long-term and short-term memory parts which are especially applicable to time series forecasting. Our expectation was that all three methods would be effective for time series data and perform comparably. Our goal was to develop the best temperature time series forecasting model we can. The best model from each technique was compared against each other and a preferred model was selected based on AIC and RMSE.

Literature Review

Time series analysis has been applied to varying topics and bodies of work when the data is dependent on the chronological order of time, and its relationship to the variable(s) in the dataset. Areas that have applied time series analysis include finance, energy, health, environmental studies, and more. Previous research allows us to examine the application of various time series approaches and the strengths of those techniques. While the application of time series may be wide in scope, the objective is nonetheless the same: to model the data with respect to the effect that values from past periods have on future periods.

The Autoregressive Integrated Moving Average (ARIMA) is a commonly used time series approach for forecasting where the data has a trend component. SARIMA is an extension of the ARIMA model technique that handles the seasonal effects of the data by using past period values. Peng Chen et al (2018) [2] explore the application of SARIMA on modeling the monthly mean temperature of Nanjing, China with data collected from 1951 to 2017, and forecast for the years 2015-2017. Considering both the non-seasonal and seasonal components of the data from their exploratory analysis, they create a statistically significant model that is successfully able to forecast future monthly temperature with a low MSE. Another application of SARIMA is demonstrated through research in detecting future influenza outbreaks. Zhang et al (2018) [5] model SARIMA to data collected on influenza's notification, temperature, and Google Trends (GT) for the years 2011 to 2016 in Brisbane and Gold Coast, Australia. They report on the significance that lags have on models and the reduction in prediction errors. They cite that the important determinant in their model is the autoregressive (AR) component for modeling influenza outbreaks. They conclude that their SARIMA model was able to forecast influenza outbreaks based on the features they used to build their SARIMA model.

Another approach often used in time series analysis are additive and multiplicative models that accommodate for multiple patterns that include the decomposition components and/or varying polynomial degrees of the components. In their research, Gould et al (2008) [3] apply a new forecasting time series model capable of accommodating multiple seasonal patterns to assess hourly as well as daily patterns for utility loads and traffic flows. By allowing for multiple seasonal updates during a given period, they apply multiplicative seasonality such that seasonal effects increase when the time series values are higher. When comparing their multiple seasonal process model with their exponential smoothing baseline models, they find that their model returned more accurate predictions because of its flexibility to

accommodate multiple seasonal data. Additive and multiplicative models are an alternative to the traditional techniques applied to time series that can be modeled with the help of time series decomposition.

In addition to the methods mentioned above, popular machine learning approaches have been implemented where traditional time series applications have typically been the primary choice. One such approach is neural networks and deep learning. In their research, Baboo and Shereef (2010) [1] apply the Back Propagation Neural Network (BPN) algorithm to make temperature forecasts. They state that the main advantage of BPN is the ability to capture the complex relationships that contribute to forecasting temperatures. In addition, it handles linear and non-linear data relationships. What distinguishes a BPN from a traditional time series technique is the propagation of the errors from the output nodes back to the input nodes through multiple iterations which has the neural network learning. When their model is applied to 12 months of forecasting, they conclude that a BPN model can be considered as an alternative approach to traditional weather forecasting approaches using time series methods.

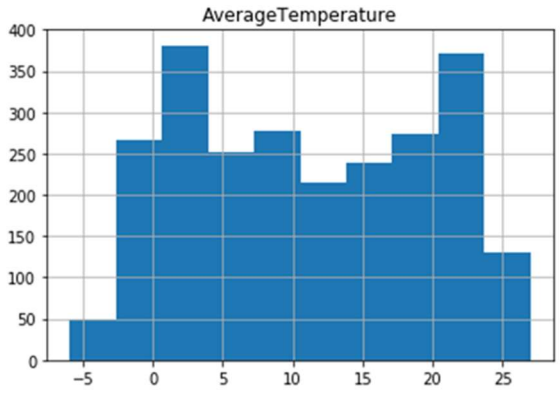
Hippert and Pedreira (2004) [4] apply linear time series models and neural networks to forecast temperature profiles for short-term loads – predicting the next day's 24-hour temperature. They use a few linear models, a linear-neural network hybrid model, and a neural network only model with a feed-forward method and multilayer perceptron with 24 outputs to predict 24-hour values at once. What they find is that their neural network only model is able to outperform their other models. The reason why the neural network is able to perform so well is due to the network being robust enough to handle noisy data. Additionally, the network is able to accommodate all 24 hours as input allowing for the structure to explore the autocorrelation of the series, and therefore actually creating a "profile shape" rather than just running a technique.

Data and Preprocessing and Methods

The data was sourced from Kaggle. The data we obtained comprised 5 datasets: GlobalTemperatures between 1750-2015, GlobalLandTemperaturesbyCountry -243 countries from 1849-2013, GlobalLandTemperaturesbyState - 241 states from various countries from 1855-2013, GlobalLandTemperaturesbyMajorCity – 100 cities from 1849-2013, and GlobalLandTemperaturesbyCity – 417 cities between 1833-2013. The data was provided to Kaggle by Berkeley Earth and consisted of *monthly* average temperatures.

Prior to beginning data exploratory analysis, it was necessary to understand the data and how it was collected. We were surprised to learn that there are a lot of issues with temperature collection, especially over the past 200 years. In the early days, mercury thermometers were used which were later exchanged for digital, which gives somewhat different readings. The time of day that the temperature was collected was inconsistent - where even small differences in time can make a big difference in temperature. Temperature stations have moved locations, which affect readings; some were shut down for long periods of time; some were moved to airport locations which are much hotter; some were badly managed. Given all these issues, trying to predict global temperature would include bias and complications making it difficult to make accurate predictions. For this reason, we chose to go with a single city.

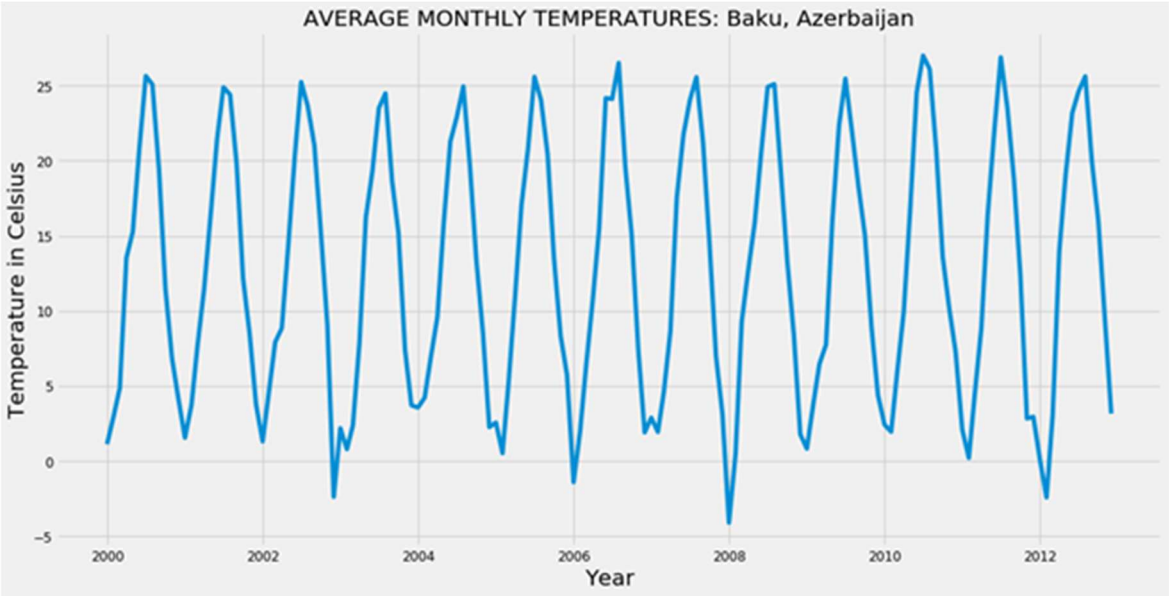
While checking for missing data, a total of six missing temperatures were found in 1816, 1918, 1919, and 2013; because the total number was not significant, they were omitted from the dataset. Upon further analysis, the histogram of the data was not normal and exhibited a bimodal quality as shown in the histogram. Below is the histogram along with summary statistics for the subset of the data.



Statistics table for AverageTemperature:	
count	156
mean	12.647442
std	8.625335
min	-4.105
25%	4.45425
50%	12.568
75%	20.64675
max	27

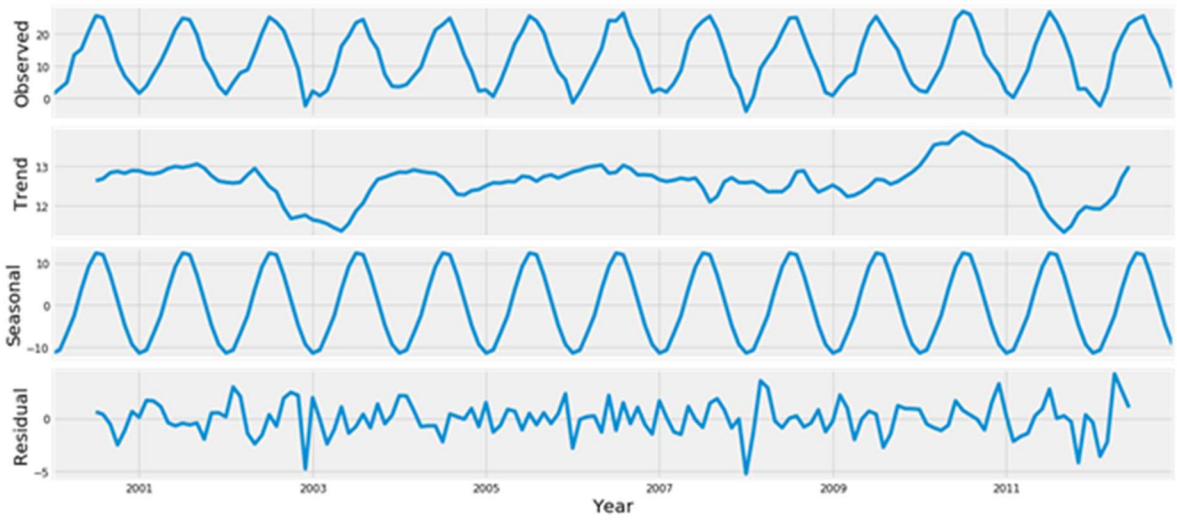
The initial plot of the data did not provide information as 192 years of monthly data was excessive. We then took various slices of the dataset to see if it would provide any insight: the last 100 years (most recent 100 years), the last 50 years (most recent 50 years), and most recent 12 years (since 2000). With the 50-year slices, we could clearly see the seasonal pattern and with the 12-year slice, it was even more evident. *For full graph of 192 years, please see Appendix - Data and Preprocessing and Methods Figure 1.*

Most recent 12 years (2000-2012)



A decomposition of the data was also rendered, which separated any trend from seasonality and the residuals. From the decomposition, there is no trend, but there is clear seasonality and that the residuals are random.

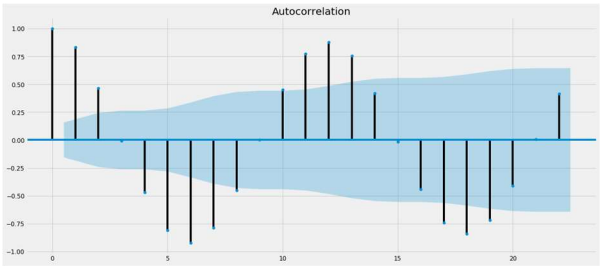
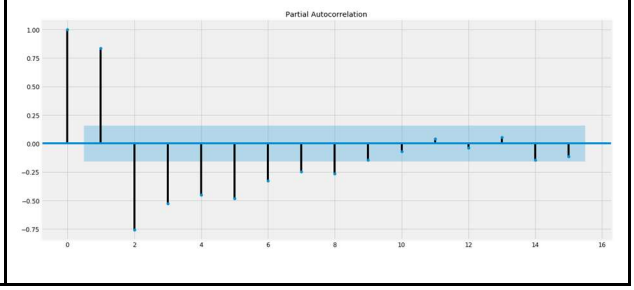
Decomposition of Baku, Azerbaijan for the recent 12 years



In comparing the different slices, it appeared that the behavior of temperature was consistent and that the most recent temperature recordings would be sufficient input for the forecasting models. By choosing to work with the data collected between 200-2012, we estimated that some of data collection issues identified would be less of a factor.

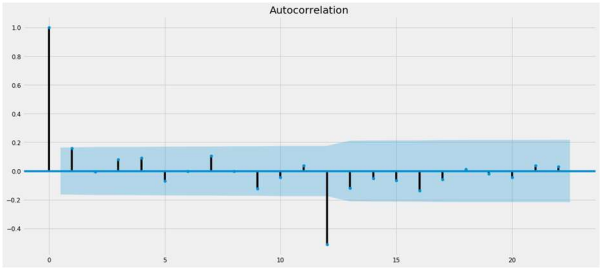
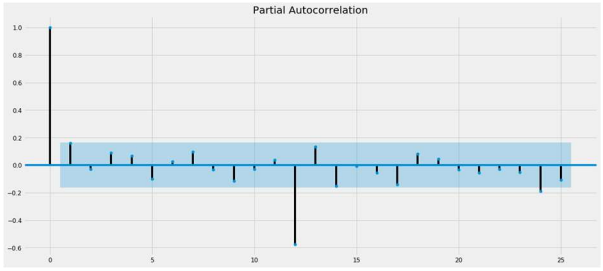
The next approach was to look at the ACF and PACF plots. Given that there was no identifiable trend in the time plot nor in the decomposition but there was seasonality, it was the expectation that there would not be a need to do trend differencing, but that seasonal differencing was required in order to obtain stationarity with our data. The ACF plot shows a distinctive seasonal pattern, while the PACF plot decays pretty quickly.

Raw data

ACF-raw	**PACF-raw**
	

Based on all the analysis to this point, it is clear that seasonal differencing is necessary in order to obtain stationarity with the data.

First order seasonal differencing:

ACF - seasonalDiff=1	PACF - seasonalDiff=1
	

After a first order of seasonal differencing, the plots were not able to reach a clean state of stationarity as there is still some autocorrelation at lag 12, additional differencing was tested up to the fourth order. With each increased order of differencing, the ACF plot remained consistent reflecting no change, however, the PACF plot worsened with random lags pushing past the confidence threshold. *(See Appendix - Data and Preprocessing and Methods, Figures 2 and 3 for additional ACF and PACF plots)*

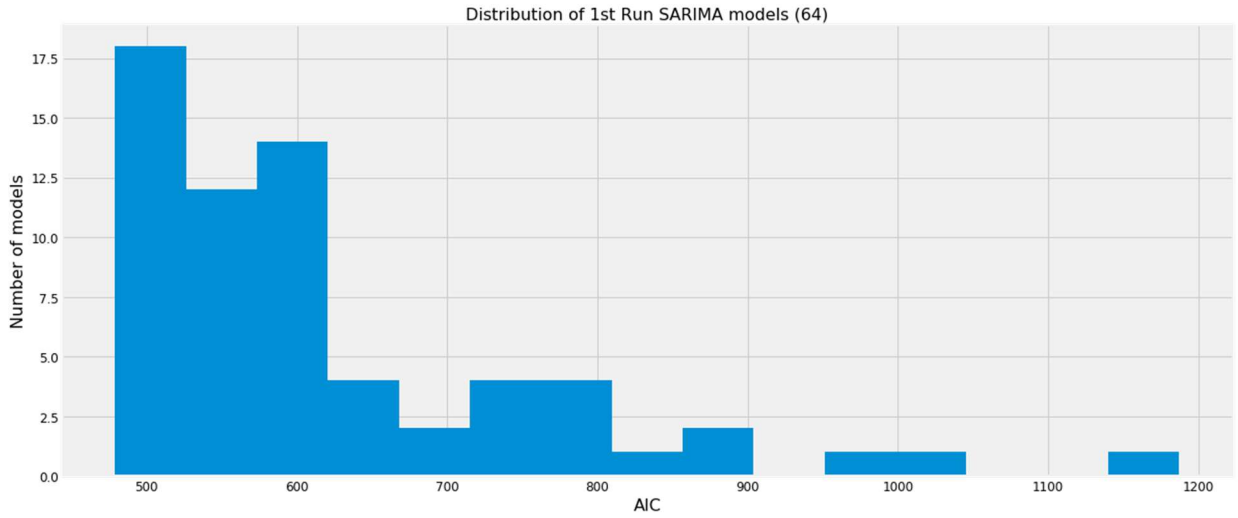
Although the initial analysis indicated that no trend differencing was needed, trend differencing was performed, with no benefit. Given these results, it was determined that order 1 seasonal differencing, while not perfect, was the best choice. Going into the model building/experimentation phase based off of the first order seasonal differenced ACF and PACF plots, it was predicted that the SARIMA (0,0,0) x (1,1,1)12 model would work the best. The predicted configurations for the SARIMA model were based on the following:

Non-Seasonal:	Process:	Value:
p:	Determined by the number of lags in the PACF starting from 1. Here we do not see any pushing past the confidence threshold.	0
d:	Trend differencing: None	0
q:	Determined by the number of lags in the ACF starting from 1. Here we do not see any pushing past the confidence threshold.	0
Seasonal:	**Process:**	**Value:**
P:	For the seasonal component, the number of lags are determined from the frequency value (12), and assessed in multiples of 12 using the PACF. At the 12th lag (representing a year), the lag is significant.	1
D:	Seasonal differencing	1
Q:	For the seasonal component, the number of lags are determined from the frequency value (12), and assessed in multiples of 12 using the ACF. At the 12th lag (representing a year), the lag is significant.	1

Models

SARIMA

Prior to the class presentation, 64 models were tested with combinations of 0 and 1 for our pdq and PDQ in our SARIMA model. Values for the configuration was left at 0 and 1 because neither the ACF nor the PACF plot showed lag 2 reaching beyond the threshold. It required too much space to show the chart of all 64 models, so provided below is the histogram, which shows the distribution of the AIC values as well as a chart of the top 5 models. The best model was chosen based on the lowest AIC value.



Top 5 SARIMA models - (pre presentation)

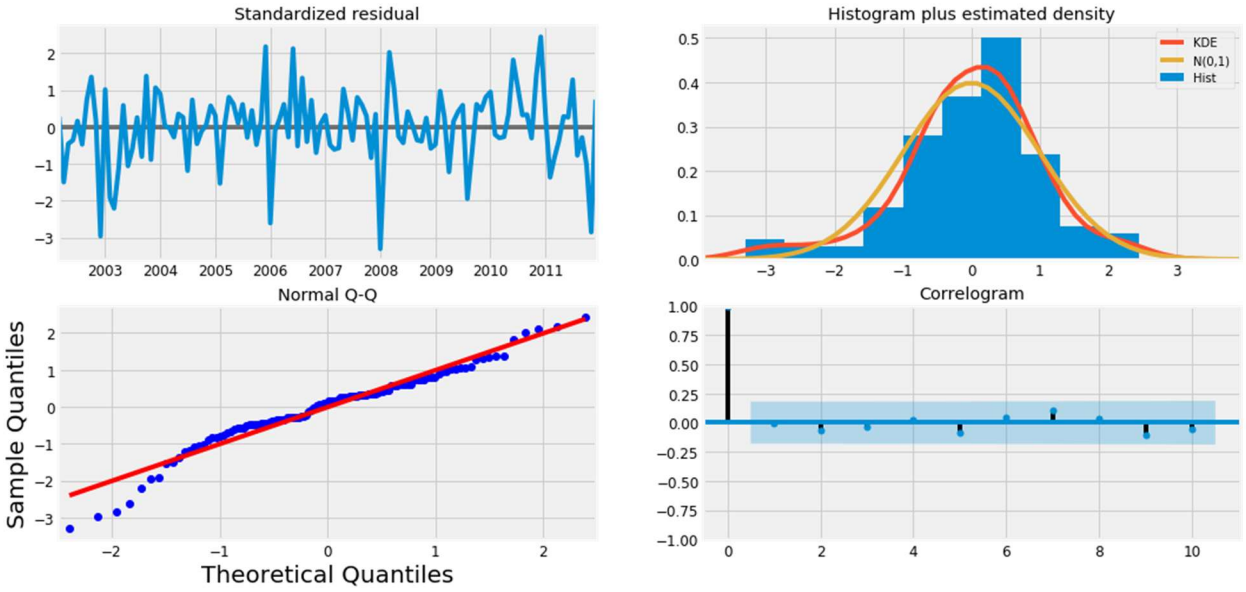
pdq	PDQ	AIC
(0, 0, 1)	(1, 1, 1, 12)	478.998043
(1, 0, 1)	(1, 1, 1, 12)	480.751334
(0, 0, 1)	(0, 1, 1, 12)	481.666227
(1, 0, 1)	(0, 1, 1, 12)	482.123017
(1, 0, 0)	(1, 1, 1, 12)	483.80321

The best model's configuration is: SARIMA(0, 0, 1), (1, 1, 1, 12)12, with an AIC of 478.998043. The best model was only slightly off from our initial prediction of what the optimal model would be based only on our analysis of the ACF and PACF plot. The best model out of the 64 that we tested included a non-seasonal MA lag of 1. But overall, the predicted model and best model is close in configuration.

SARIMA Residual Analysis:

From the residual plots, the standardized residual plot shows that the residuals are random, and the correlogram shows that the residuals are independent. This is further confirmed by the plot of the Ljung Box Test with 20 lags where the lags are all above the 0.05 significance threshold; therefore, the null hypothesis that the residuals are independent cannot be rejected. Our residuals are independent. The histogram and the QQ-plot both show that the residuals may not be normally distributed. The QQ-plot shows that the plots are deviating from the line on the lower left side of the graph, indicating a tail. This is further supported by the Jarque-Bera Normality test, with a probability value of less than 0.05, signifying that the normality assumption is rejected for the residuals.

Residuals

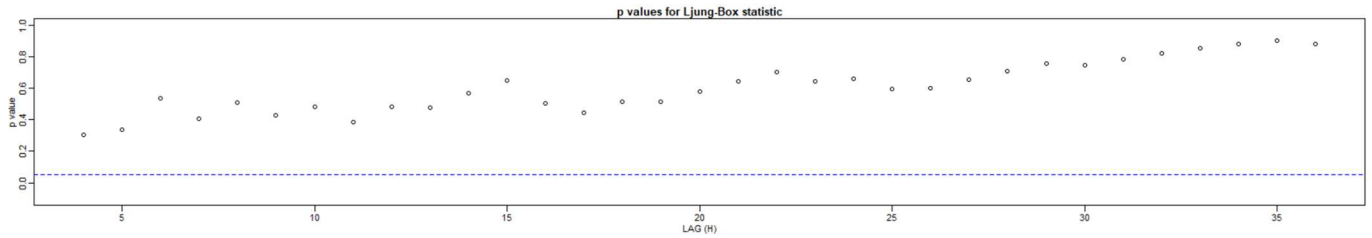


Coefficient Testing, Jarque-Bera Test, and Ljung-Box Test

Dep. Variable:	AverageTemperature	**No. Observations:**	144
Model:	SARIMAX(0, 0, 1)x(1, 1, 1, 12)	**Log Likelihood**	-235.499
Date:	Sun, 17 Nov 2019	**AIC**	478.998
Time:	14:00:17	**BIC**	490.081
Sample:	01-01-2000	**HQIC**	483.498
	- 12-01-2011		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	0.2399	0.085	2.824	0.005	0.073	0.406
ar.S.L12	-0.2386	0.117	-2.035	0.042	-0.468	-0.009
ma.S.L12	-0.7920	0.124	-6.412	0.000	-1.034	-0.550
sigma2	2.8926	0.324	8.932	0.000	2.258	3.527

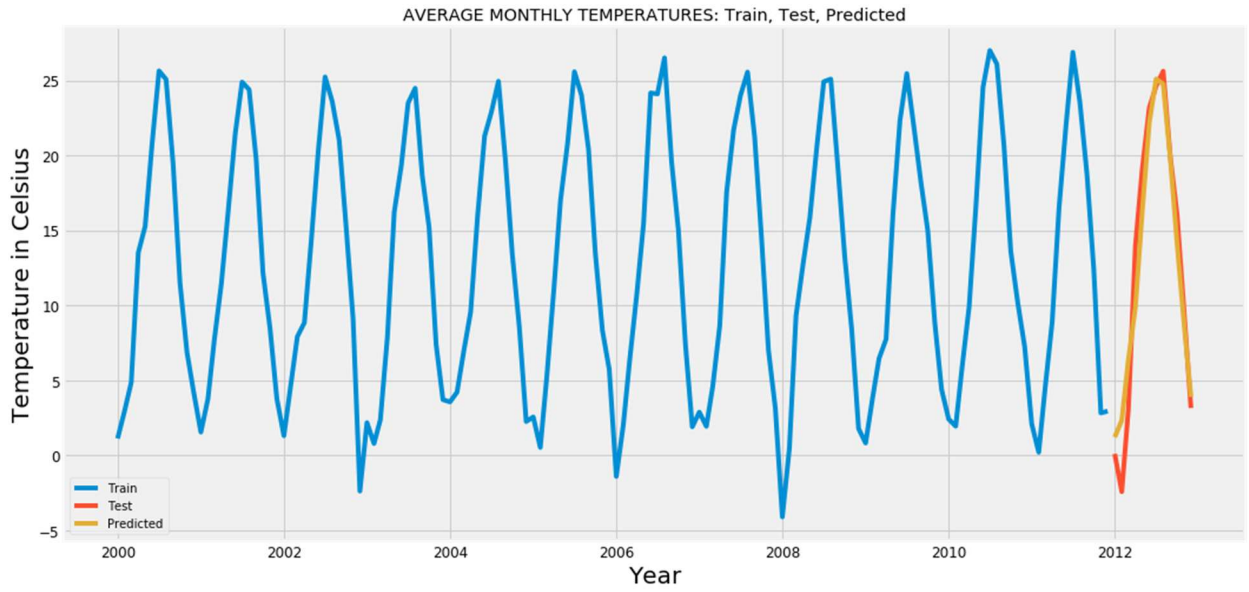
Ljung-Box (Q):	21.35	**Jarque-Bera (JB):**	16.54
Prob(Q):	0.99	**Prob(JB):**	0.00
Heteroskedasticity (H):	1.00	**Skew:**	-0.61
Prob(H) (two-sided):	1.00	**Kurtosis:**	4.37



SARIMA Predictions:

As mentioned earlier, we split our training and test data into 2000-2011 for the training and 2012 as the test. We trained our SARIMA model on the training data and then used that to predict the monthly temperatures for 2012. We then compared our predicted to the actual temperature (the test set). Below is a graph of the results where the red line indicates the test values, and the yellow indicates the predicted values from the best fitting SARIMA model. The resulting RMSE is 2.35.

Based on these results and the graph, we think our model did an excellent job.



PHASE II - post presentation

After the class presentations, we wondered whether there were better models. What if we expanded the combinations? We re-ran our code to evaluate all the possible combinations of pdq and PDQ using 0,1,2,3,4. After 24 hours of running, we had to break the run in order to finish our paper. We generated 9,537 models. The top 25 models, based on lowest AIC scores all had combinations of PDQ of (4,4,0) or (4,4,1) which was surprising. The best result of AIC = 251 beat our first best model which had AIC=479. Below are the top 5 results with the lowest AIC. However, even though the AIC was 48% lower than our previously best model, the RMSE was 7.8, which was 70% worse. Interesting results. We think that the order of 4 model is too complex and is probably unstable. It is too many derivatives aways from the raw data to perform reliably. *For further analysis, please see Appendix - Model Figure 1. SARIMA Model.*

Top 5 SARIMA models - (post presentation)
(2nd trun)

pdq	PDQ	AIC
(4,1,0)	**(4,4,1)**	**251.041127**
(4,0,4)	(4,4,0)	251.757435
(4,0,4)	(4,4,1)	251.963962
(4,2,2)	(4,4,0)	251.973176
(4,0,1)	(4,4,1)	252.224422

Time Series Regression

A time series regression additive model is a regression model where the systematic and non-systematic components sum to the model; these systematic and non-systematic parts of the model are essentially the decomposed time series components. The method is appropriate if the magnitude of the seasonal oscillations or the variance around the trend-cycle, does not vary with time. The additive model is an extension of linear regression with the addition of the level, which is the average value of the series, the trend that incorporates the increasing or decreasing values of the series, the seasonality or repeating cycles, and the inherit random variations. The general form of the equation is shown below.

Time Series Additive Regression Model General Structure:

$$y_t = Level + Trend + Seasonality + Noise$$

After performing exploratory analysis from the time series decomposition, the conclusion was that we did not see any trend. To verify this, a full-scale additive model accounting for both trend and seasonality was created to test this hypothesis resulting in trend not being statistically significant to the model. Therefore, the trend component was left out of the model building process.

The time series regression additive model has 12 dummy variables capturing the seasonality component of data, which is represented by the months: January to December; season 1, January, is removed from the model to avoid multicollinearity. The seasonal coefficients are positive and between 1.04 and 23.76; all but season 2 are shown to be statistically significant to the model. Below is a proposed model's equation:

Proposed Time Series Additive Regression Model General Structure:

$$y_t = 1.25 + 1.05S_2 + 4.75S_3 + 8.83S_4 + 14.83S_5 + 20.37S_6 + 23.76S_7 + 23.35S_8 + 18.47S_9 + 12.51S_{10} + 6.52S_{11} + 1.98S_{12} + \varepsilon_t$$

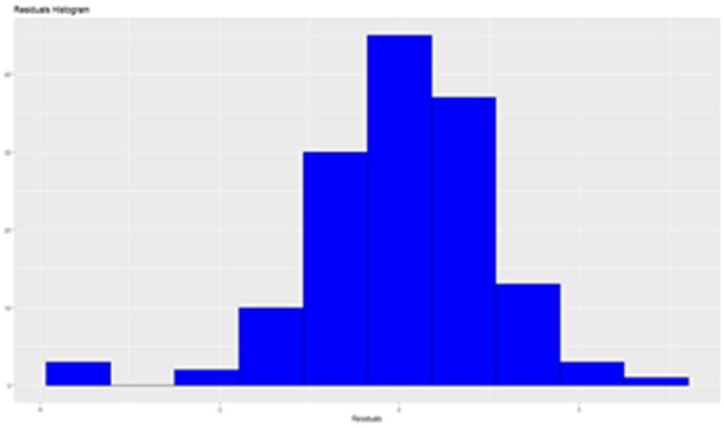
Coefficients	Estimate	Std. Error	T Value	PR (>|t|)	Significance
(Intercept)	1.2503	0.4676	2.674	0.00845	**
Season 2	1.046	0.6613	1.582	0.11613	
Season 3	4.7455	0.6613	7.176	4.66 e-11	***
Season 4	8.8298	0.6613	13.351	<2e-16	***
Season 5	14.8348	0.6613	22.431	<2e-16	***
Season 6	20.3714	0.6613	30.803	<2e-16	***
Season 7	23.7611	0.6613	35.929	<2e-16	***
Season 8	23.3549	0.6613	35.314	<2e-16	***
Season 9	18.4686	0.6613	27.926	<2e-16	***
Season 10	12.5051	0.6613	18.909	<2e-16	***
Season 11	6.524	0.6613	9.865	<2e-16	***
Season 12	1.9751	0.6613	2.986	0.00336	**

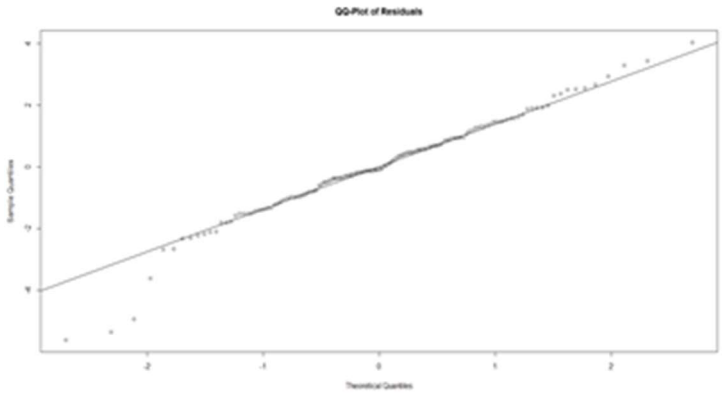
The analysis of the seasonal coefficients and the variation among the seasons is as expected. The coefficients increasingly increment from Season 5 (May) through 9 (September), which can be interpreted as a direct association with the high peaks in temperature in Baku from May to September. Therefore, the model is accurately capturing the seasonal component. In terms of coefficient importance for the model, only Season 2 coefficient with P- value of 0.11613 is not significant, but as it does not have an impact on the overall model and the residuals analysis, the coefficient is retained in the model.

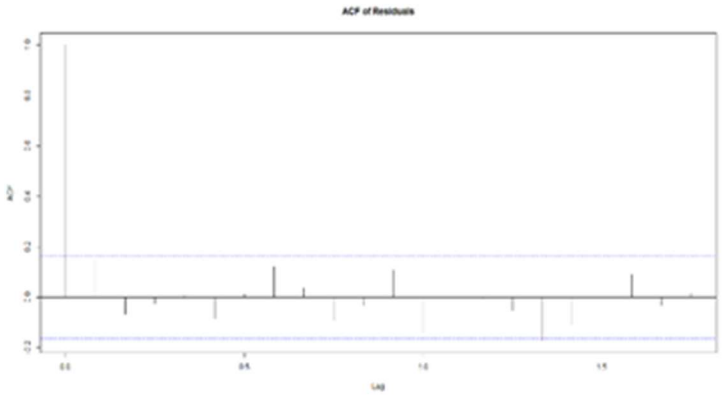
Time Series Regression Residual Analysis:

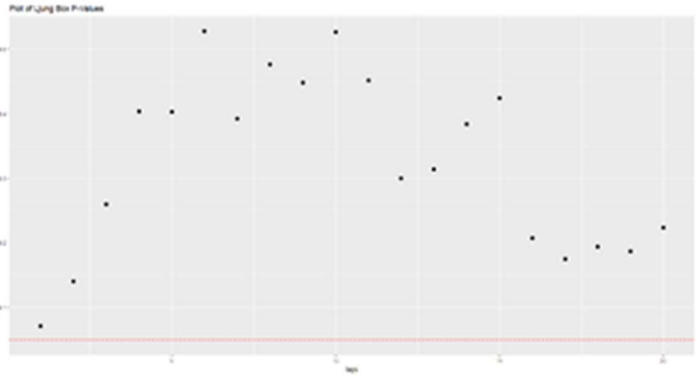
The residual analysis evaluates two important aspects: normality and independence. From the histogram of the residual is not shown as symmetric, which is supported by what is shown on the QQ plot, where the left tail is deviating from the QQ plot line. The Jarque Bera Normality Test was performed resulting in a p-value of 3.391e-6; the null hypothesis was rejected confirming the non-normality of the residuals. To measure the independence, the ACF of the residuals confirms that there is no autocorrelation within the residuals. To further test for independence amongst the residuals, a Ljung-Box test was performed for 20 lags and plotted as shown below. All plots are above the 0.05 significant threshold represented by the red line, concluding that the residuals are indeed independent.

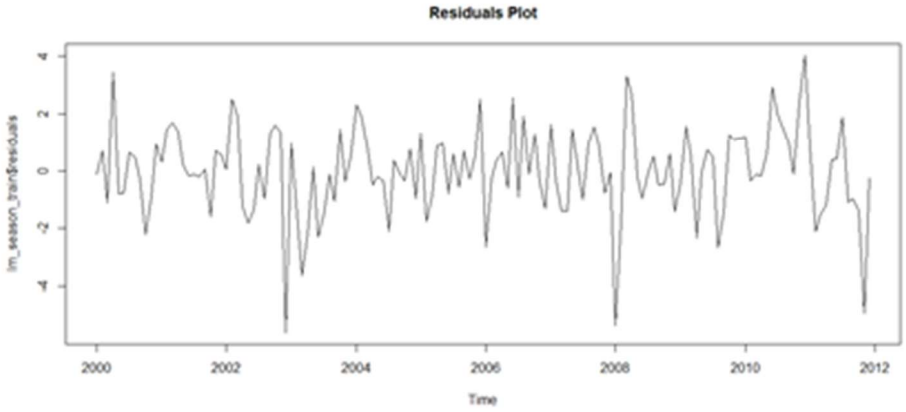
Residuals Analysis Plot



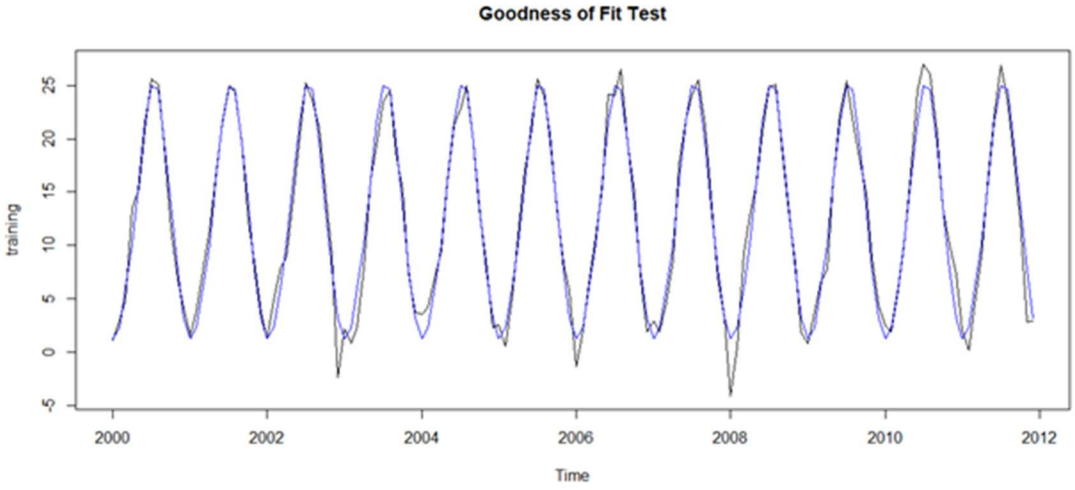






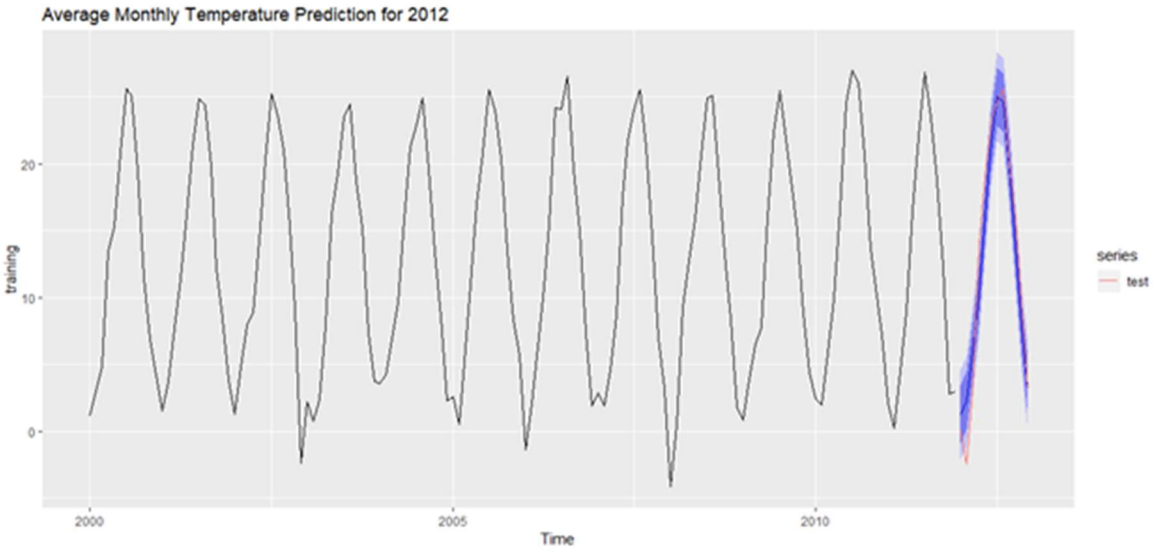


Though the proposed additive model violates the normality assumption, the model will be compared against the other models for its forecasting performance. The Goodness of Fit Test plot shows that the fitted values from the trained model overlaid on top of the actual data is tracking the true points closely, where there is no significant over or underestimation of the data. The goodness of fit of the model during the training returns an AIC of 561.05.



Time Series Regression Predictions:

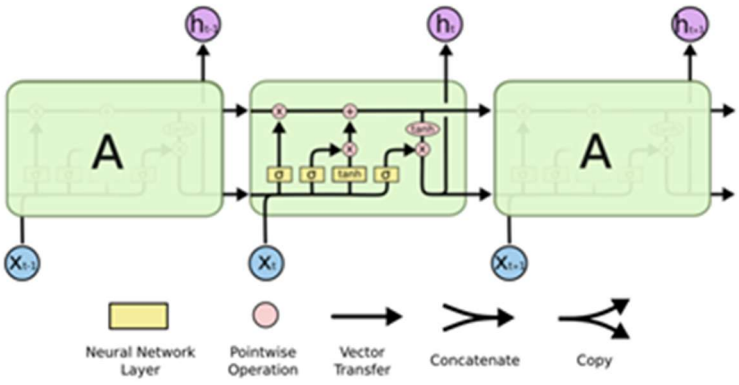
The time series regression additive model was evaluated using the testing set for the year 2012. The red line indicates the test values, and the confidence interval of the predicted values are shown within the blue shading along of the predicted forecast line shown in black. Visually, based on graph, the forecasted values for 2012 are shown to be close to the actual test values. The evaluation of the model on the testing set is measured using the Root Mean Squared Error (RMSE), which is 2.387.



Neural Network

Long Short-Term Memory networks (LSTM) are a special type of Recurrent Neural Network capable of learning long-term dependencies by relying on a chain like structure of interacting layers within each module as seen in the LSTM Structure. A general structure contains 4 sequence layers set up to execute the core tasks. First, the forget gate layer incorporates a sigmoid function to decide which input information is discarded. Second, the input gate layer stores the new information with a sigmoid function. Third, the update layer incorporates a Tanh function to combine stored with the new information and update the current state and lastly the output layer filters the information that is going to be yield relying on a sigmoid/tanh combination. The LSTM structure by having this continuous flow of information can step back n-number of times and formed accurately connections to continuously learn and improves its prediction.

LSTM Structure



The increase in the application of neural networks as a forecasting technique for time series is that in contrast to traditional methods like ARIMA and Regression, neural networks do not make any type of assumptions in terms of the

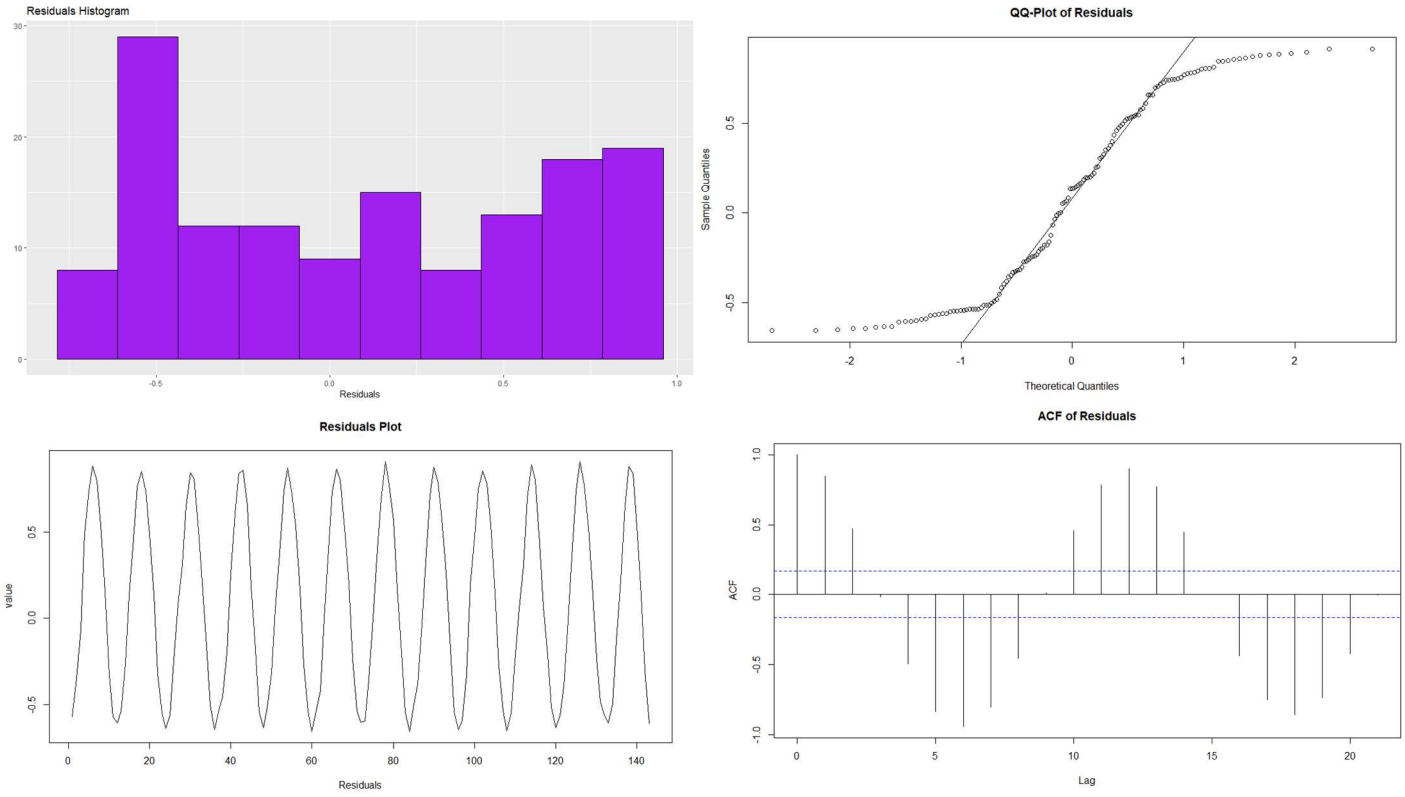
data (Trend, Seasonality, Normality). The objective is to capture the existing pattern in the series by using the activation functions to incorporate not only the linear but also the nonlinear behavior.

The LSTM model used to forecast the average temperatures in Baku is a simple but reliable stacked structure. The model structure was created by 2 LSTM layers with 50 neurons each and 2 dropout layers to temporarily remove 20% of the neurons to avoid overfitting. The model was compiled using 300 epochs passing a single batch, an Adam Optimizer to control the learning rate and the Mean Squared Error as the loss function. The LSTM Model generated over 30 thousand parameters as shown below in the LSTM Neural Model Summary.

LSTM Neural Model Summary

Layer	Type	Std. Error	# Parameters
Lstm_1	LSTM	(1,1,50)	10400
Dropout 1-0.20	DROPOUT	(1,1,50)	0
Lstm_2	LSTM	(1,50)	20200
Dropout 2-0.20	DROPOUT	(1,50)	0
Dense_1	DENSE	(1,1)	51
Total Parameters	30651		
Trainable Parameters	30651		
Non-Trainable Parameters	0		

Neural Network Residual Analysis:



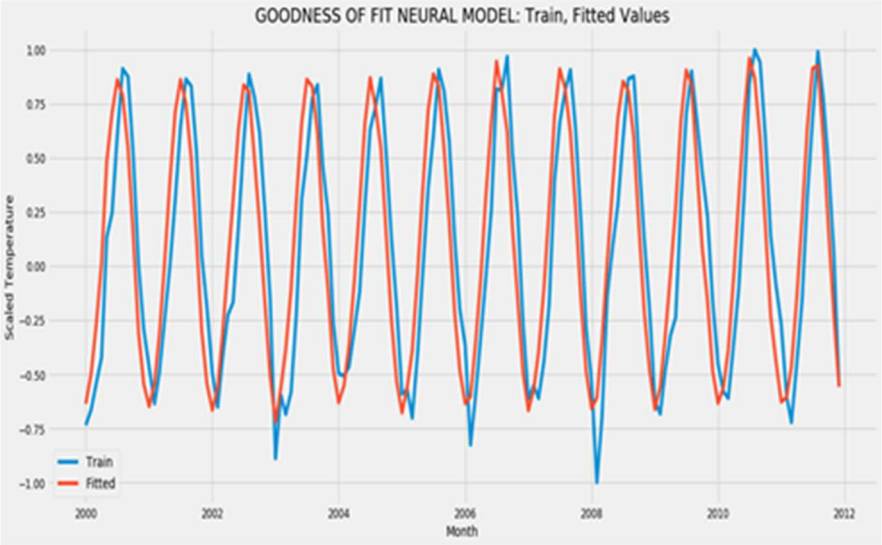
The histogram and the QQ-plot for the LSTM model shows that the distribution of the residuals is not normal, which confirmed by the Jarque-Bera test. The p-value of 0.001271 means that the null hypothesis of normality is rejected. The ACF plot of the residuals and the residuals plot shows that there is still seasonality to the residuals. However, this

is expected as unlike time series, neural network models are “model free” and do not make any assumptions regarding the data on independence or normality. Because neural networks makes no assumptions, the residual analysis was performed for completeness and less for model performance assessment. The residual plots were added for the sake of completeness.

Neural Network Predictions:

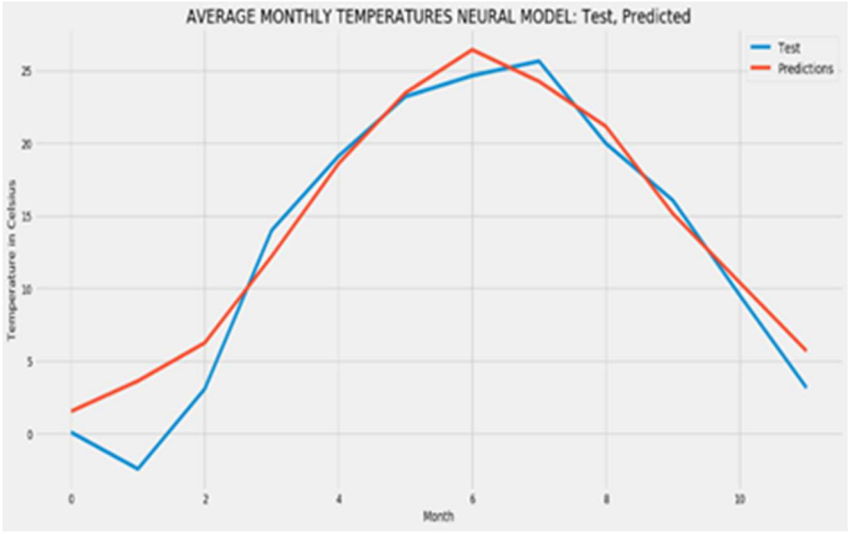
The LSTM model was trained and tested using the same average monthly temperature data sets. Visually the goodness of fit plot shows the model captured the overall behavior of the training set except for 3 major downward trends shown below.

Goodness of Fit Plot



The forecast and the test plot confirm the predictions replicate the test set’s overall pattern, but interestingly, for forecast temperatures in the range of 10 to 20 degrees with a gap of 2 months between the segments are quite accurate in comparison with the test set.

LSTM Model Forecast Plot



To evaluate the overall performance of the model in terms of metrics, the Root Mean Squared Error was calculated and averaged in 10 trial testing run to minimize the effect of the inherent randomness introduced by the dropout layers and guarantee the reproducibility of model results. The RMSE achieved by the models was 2.45 as reflected in the Testing Set Validation RSME table.

Testing Set Validation RMSE

Testing Run	RMSE
Trail 1	2.193
Trail 2	2.259
Trail 3	2.69
Trail 4	2.534
Trail 5	2.379
Trail 6	2.047
Trail 7	2.508
Trail 8	3.04
Trail 9	2.513
Trail 10	2.373
Average	**2.4572**

Conclusion and future work

We were able to successfully predict the temperature for 12-months (2012) for Baku. All three models performed similarly. To compare our top models, we used the RMSE evaluation metric. Based on each model's RMSE, our best model was the SARIMA (0,0,1)(1,1,1)12 model which had an RMSE of 2.35. Our expectation prior to experimentation was that the SARIMA model would outperform the Time Series Regression Additive model, but that the Neural Networks LSTM model would come close.

Summary of best models :

Model	AIC	RMSE	
SARIMA	**479**	**2.35**	best model
TS Regression	561	2.39	
LSTM	NA *	2.46	

* Comments for the Comparison: The AIC takes into account the number of parameters in the model to calculate the metric, but as neural network models incorporates so many parameters, the AIC metric can only be used to compare different neural structures and not models obtained by other methods like ARIMA or Regression.

For future work, it would be interesting to obtain additional weather variables such as air pressure and density to create a multivariate model and compare its performance. It is very possible that with added features, the prediction may be more accurate resulting in a smaller RMSE score.

This project allowed us to apply what we learned in class. Additionally, it provided an opportunity to compare different time series techniques against one another as well as a machine learning technique. Whether it is a machine learning problem or a time series problem, it is always good practice to apply various applicable approaches to see how the results compare allowing for deeper analysis.

References

[1] Baboo, S. S., & Shereef, I. K. (2010). An efficient weather forecasting system using artificial neural network. *International journal of environmental science and development*, *1*(4), 321.

[2] Chen, P., Niu, A., Liu, D., Jiang, W., & Ma, B. (2018, July). Time series forecasting of temperatures using SARIMA: An example from Nanjing. In *IOP Conference Series: Materials Science and Engineering* (Vol. 394, No. 5, p. 052024). IOP Publishing.

[3] Gould, P. G., Koehler, A. B., Ord, J. K., Snyder, R. D., Hyndman, R. J., & Vahid-Araghi, F. (2008). Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research*, *191*(1), 207-222.

[4] Hippert, H. S., & Pedreira, C. E. (2004). Estimating temperature profiles for short-term load forecasting: neural networks compared to linear models. *IEE Proceedings-Generation, Transmission and Distribution*, *151*(4), 543-547.

[5] Zhang, Y., Bambrick, H., Mengersen, K., Tong, S., & Hu, W. (2018). Using Google Trends and ambient temperature to predict seasonal influenza outbreaks. *Environment international*, *117*, 284-291.

Appendix

Data and Preprocessing and Methods

Figure 1

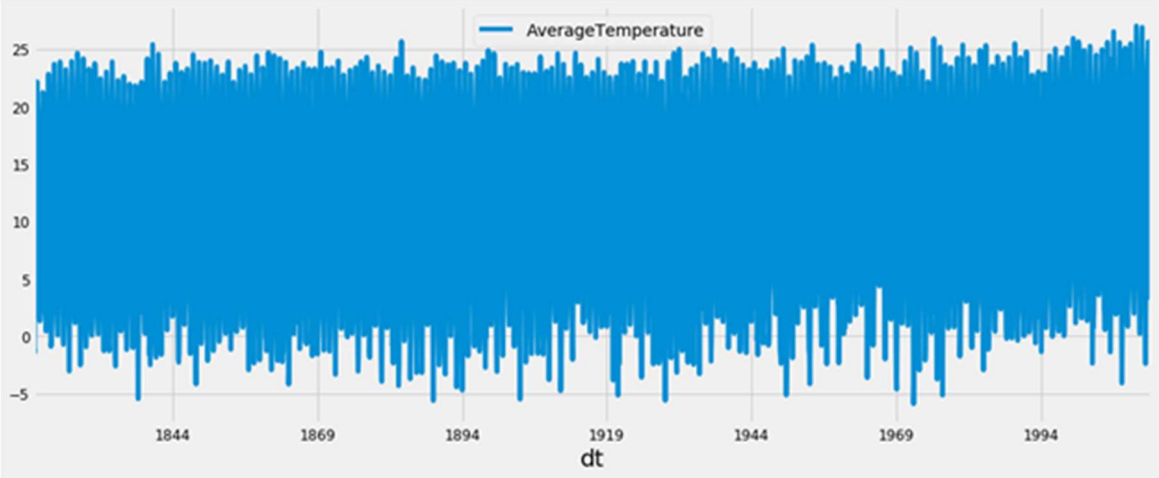


Figure 2

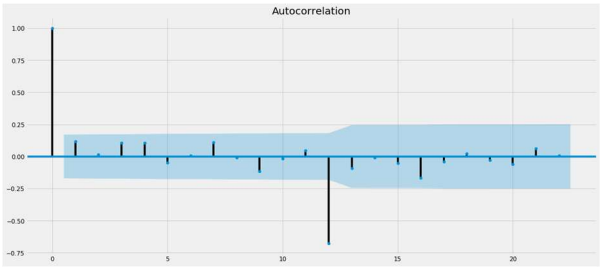
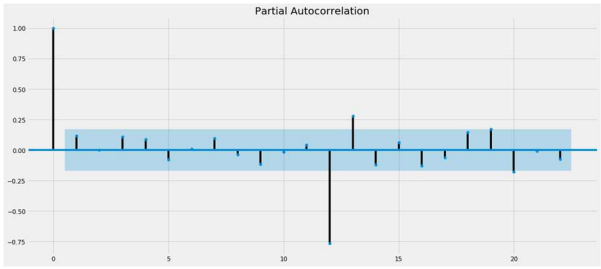
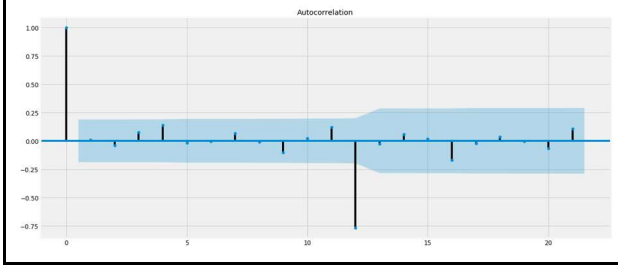
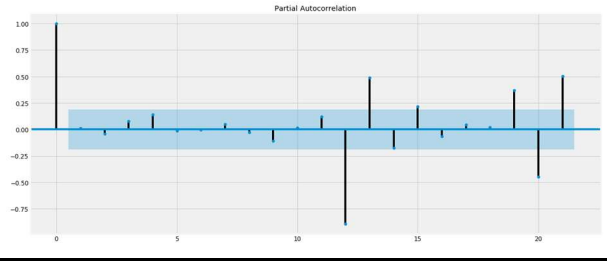
ACF - seasonalDiff=2	PACF - seasonalDiff=2
	

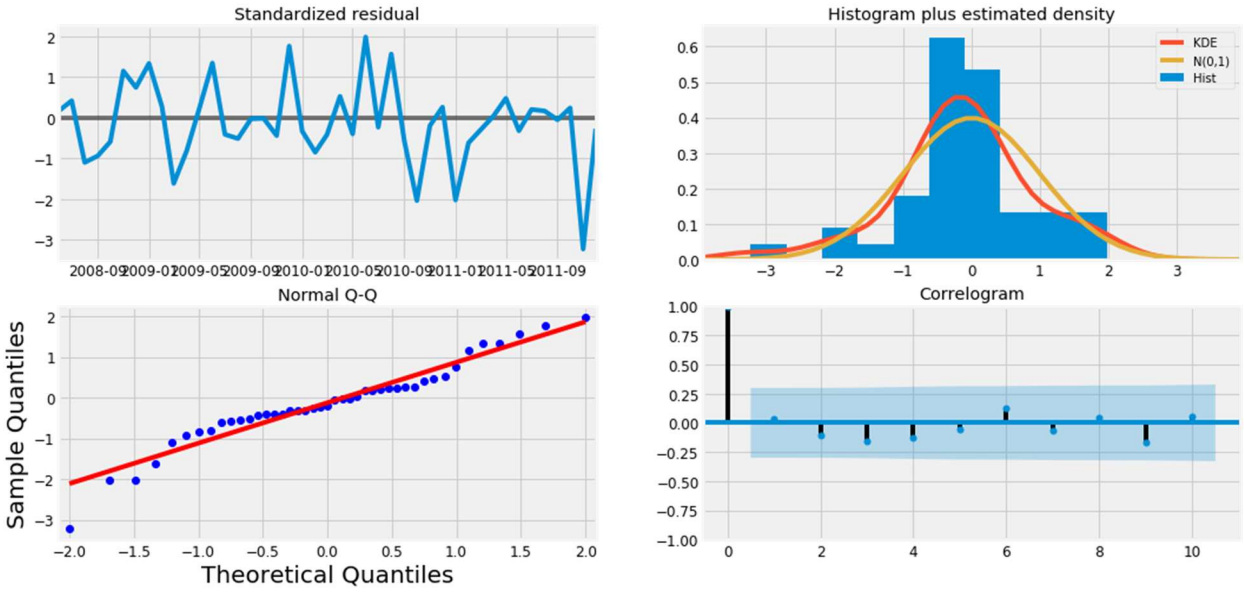
Figure 3

ACF - seasonalDiff=4	PACF - seasonalDiff=4
	

Model

Figure 1. SARIMA Model

Even though the AIC was so significantly lower than our previous best model the RMSE and residual tests were not good. The new 'best' model had an RMSE of 7.8 vs. 2.35 for the previous best. Below are the residual plots. We observe that the residuals are not white noise, nor are they normal. The PCF or the residuals did look good.



COEFFICIENT TESTS, JARQUE-BERA, LJUNG-BOX:

Dep. Variable:	AverageTemperature	**No. Observations:**	144
Model:	SARIMAX(4, 1, 0)x(4, 4, 1, 12)	**Log Likelihood**	-115.521
Date:	Sat, 30 Nov 2019	**AIC**	251.041
Time:	15:03:56	**BIC**	268.653
Sample:	01-01-2000	**HQIC**	257.536
	- 12-01-2011		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.4549	0.313	-1.451	0.147	-1.069	0.159
ar.L2	-0.4369	0.303	-1.442	0.149	-1.031	0.157
ar.L3	-0.3484	0.235	-1.482	0.138	-0.809	0.112
ar.L4	-0.1871	0.210	-0.889	0.374	-0.600	0.225
ar.S.L12	-2.0126	0.165	-12.200	0.000	-2.336	-1.689
ar.S.L24	-2.0144	0.310	-6.503	0.000	-2.622	-1.407
ar.S.L36	-1.1974	0.316	-3.795	0.000	-1.816	-0.579
ar.S.L48	-0.3774	0.160	-2.356	0.018	-0.691	-0.063
ma.S.L12	0.9935	29.822	0.033	0.973	-57.456	59.443
sigma2	0.7488	21.989	0.034	0.973	-42.349	43.847

Ljung-Box (Q):	32.70	**Jarque-Bera (JB):**	4.28
Prob(Q):	0.79	**Prob(JB):**	0.12
Heteroskedasticity (H):	1.34	**Skew:**	-0.44
Prob(H) (two-sided):	0.59	**Kurtosis:**	4.27

PREDICTIONS

As stated earlier, our new RMSE was 7.8 which is evident in the graph of the new predictions below. In the previous graph of our predictions, it was difficult to see the line of the actual as the predicted lay almost perfectly on top. Here you can clearly see both lines, indicating less than perfect predictions.