New year resolution time series analysis using Python

Fakhra Akhtar

Faculty of Computer Science an Information Technology, RIPHA University Lahore, Pakistan

30th -June-2020

hanaali829@gmail.com

Faizan Ahmed Khan

Faculty of Computer Science COMSATS University Islamabad, Lahore-Campus

geofaizan.khan@gmail.com

Muhammad Tauseef Hanif

Faculty of Computer Science COMSATS University Islamabad, Lahore-Campus

tauseefhanif05@gmail.com

Asad Imran Malik

Superior University

malik_asad75@hotmail.com

Waqar Younis

University of Gujrat

viki_meo@yahoo.com

Abstract — New year goals are promises we make to ourselves to improve as an individual Welcoming in the New Year for many people implies making a New Year's goals. Beginning without any preparation toward the start of the year offers a new beginning and a fresh start, many people accept this open door to make goals. Many hold onto the opportunity to set another objective as an endeavour to dispose of a negative behaviour pattern or start a more advantageous way of life, for example, stopping smoking or getting more fit. Needing to make goals is something worth being thankful for and exceptionally supported. It gives individuals something to anticipate and continue progressing in the direction of. Regardless of whether they aren't finished, making goals shows that you have the conviction and expectation in your capacity to change your habits and become a better person The dataset used in this paper relates to new year Resolutions. The objective of this paper was to conduct a time series analysis of various attributes of new year resolutions to formulate effective result for proper decision making The results are visualized in the form of graphs and the analysis was done using python.

The results further explain the fact that in the time frame of 2015-2017 the person became more focused and started developing proper interest in doing gym and focused less on diet The best result was achieved in the timeline of 2015-2017 as he managed both gym and his diet in an efficient way to build up a healthy balanced lifestyle. This research recommends that staying active is a crucial part of maintaining good health and wellness. if a person is making new year resolutions based on health he should remain focus on them according to the years he has planned as it requires a lot of motivation and dedication to follow all the resolutions. one should be capable enough to manage his finance and diet accordingly Make exercise and physical activity a permanent part of your daily routine as well as an important part of your new year resolutions.

Keywords- New Year Resolution, Python, Timeseries

*****

INTRODUCTION

New Year's resolution is a guarantee an individual makes for the new year. Notwithstanding what goal you focus on, the objective is to improve life in the coming year. Goals can come in numerous structures. A few people make a guarantee to change an unfortunate propensity, for example, stopping smoking or eating less lousy nourishment. Others make a guarantee to build up a positive propensity, for example, beginning an activity program, and improving their lifestyle. The convention of New Year's goals goes back to 153 B.C.The antiquated Babylonians are said to have been the principal

individuals to make New Year's goals, somewhere in the range of 4,000 years prior. They were likewise the first to hold recorded festivals to pay tribute to the new year however for them the year started not in January but rather in mid-March when the harvests were planted. Regardless of the convention's strict roots, New Year's goals today are a generally mainstream practice. Rather than making promises to the divine beings, the vast majority make goals just for themselves and spotlight absolutely on personal development which may clarify why such goals appear to be so difficult to finish on. As indicated by an ongoing examination, while upwards of 45% of Americans state they for the most part make New Year's goals, just 8% are effective in accomplishing their objectives. However, that terrible record likely won't prevent individuals from making goals at any point shortly, all things considered, we've had around 4,000 years of training. In light of a review directed by specialists at the University of Scranton, in 2014 they found that lone 77% of individuals clung to their New Year's goals inside the main week, that figure plunges to 46% following a half year. A progression of studies, distributed in the Personality and Social Psychology Bulletin in 2016, found that 55% of goals are identified with wellbeing, while 20% included taking care of monetary obligations. The most well-known goals are: getting more fit, accomplishing more exercise, stopping smoking, and setting aside cash. The fundamental explanation that individuals don't adhere to their goals is that they set an excessive number of or they're unreasonable to accomplish. Creators of (1) consider Time arrangement anticipating is a significant region of determining in which past perceptions of a similar variable are gathered and dissected to build up a model depicting the basic relationship. The model is then used to extrapolate the time arrangement into what's to come. This demonstrating approach is especially valuable when little information is free on the basic information producing measure or when there is no agreeable illustrative model that relates the expectation variable to other logical factors. Much exertion has been committed in the course of recent a long time to the turn of events and improvement of time arrangement gauging models. Specialists of (2) portray the way that there is variety in exertion relying upon how testing the objective is contrasted with an individual's capacity to complete the objective. Objectives that are respectably testing invigorate the most exertion contrasted and objectives that are either excessively simple or excessively hard. On the off chance that an objective is too simple, numerous individuals basically won't make a decent attempt to achieve it since they figure they don't have to. On the off chance that an objective is too hard, the thinking is "The reason trouble?" since achieving the objective appears to be unimaginable. (3) Making New Year's goals is a typical convention.

The paper is organised in the following context The first phase of the paper explains the data set used and literature review regarding the concerned topic. The second phase deals with the methodology data analysis tool. The third part discusses the research results based on the detailed analysisFourth phase deals with Discussion, conclusion and references.

DATASET DESCRIPTION

The dataset used in this research paper is basically on new year resolutions and is denoted by .csv. It has attributes relating to finance, gym and diet. No algorithms are being used as all the entire results are visualized in the form of various graphs it deals with the new year resolutions of a single person who is diet cautious and is managing all his activities relating to his finance, diet and gym based on various new year resolutions.

LITERATURE REVIEW

(4) Masum, at el. (2018) in their exploration consider Multi-step estimating extremely testing and there is an absence of studies accessible that comprise of AI calculations and procedures for multi-step anticipating. They additionally found that the absence of coordinated efforts between these various fields is making an obstruction to additional turns of events. They played out a multi-step time arrangement determining on three nonlinear electric burden datasets removed from Open-Power-System-Data.org utilizing two AI models. They likewise analyzed the accompanying models like Multi-step determining execution of Auto-Regressive Integrated Moving Average (ARIMA) and Long-Short-Term-Memory (LSTM) based Recurrent Neural Networks (RNN) models. The consequences of Comparative investigation of determining execution of the two models uncovered that the LSTM model has better execution in correlation than the ARIMA model for multi-step electric burden gauging. (5) Shabib at el (2015) in their paper answered search and my huge informational indexes of time arrangement information all the more productively when contrasted with the already existing technique for utilizing Euclidean Distance. The UCR DTW calculation was created for a solitary CPU centre. They thought about two techniques for parallelizing the DTW calculation. To begin with, they considered a multi-centre usage, trailed by a group execution utilizing Spark. From the multi-centre execution, they accomplished almost straight speedup. In the Spark execution, they found that direct usage of DTW doesn't perform well. Their outcomes finished up the realities that parallelizing the consecutive DTW code improves the exhibition of the calculation while keeping up 100% precision to give a straight speedup in the two cases. (6) Lippi at. el (2013) in their examination depicts the way that the writing on momentary traffic stream anticipating has gone through extraordinary advancement as of late. The target of their examination paper is twofold. To begin with, they inspected existing ways to deal with momentary traffic stream gauging strategies under the basic perspective on probabilistic graphical models, introducing a broad trial correlation, which proposes a typical standard for their presentation examination and gives the foundation to work on an openly accessible informational index. Besides, they introduced two new help vector relapse models, which are explicitly conceived to profit by normal traffic stream irregularity and are appeared to speak to a fascinating trade-off between expectation precision and computational proficiency. The outcomes demonstrated that irregularity is a critical element in accomplishing high exactness; in any case, the most precise models regularly require high computational assets both during the preparation stage and at forecast time. As a future course of exploration, it was portrayed that a critical improvement in time-arrangement determining may come from social realizing, where interrelations between various time arrangements are considered. (7) Arandia et al (2014) in their exploration inspects the down-to-earth ramifications of utilizing telemetry information at worldly goals going from best to

coarsest to precisely conjecture day by day water utilization. The calculation executed plays out another fit for each model after a sliding window approach where new boundaries are assessed and gauges are produced each 24 h. Models with week after week intermittent structures are found to all the more productively eliminate the autocorrelations concerning models of the day by day occasional sort it is seen that more modest assessment windows emphatically influence the capacity of the models to adjust to unexpected changes in the water request time arrangement. The examination gives bits of knowledge on the time arrangement models definition and execution just as on the impacts of the transient goal while assessing day by day water creation. It was demonstrated that the decision of the model assessment window may considerably affect the estimate exactness. (8) Ransom at. el (2002) In their examination introduced a grouping of both norm and progressed Fourier procedures that are helpful in the investigation of astrophysical time arrangement of long term where the perception time is a lot more noteworthy than the time They evaluated the operational attributes of Fourier changes of time-arrangement information, including power-otherworldly insights, talking about a portion of the contrasts between investigations of binned information, inspected information, and occasion information, and momentarily examined calculations for ascertaining discrete Fourier changes (DFTs) of long time arrangement. They additionally examined the reaction of DFTs to intermittent signals and present methods to recuperate Fourier sufficiency "lost" during straightforward customary examinations if the periodicities change recurrence during the perception. These strategies incorporate Fourier introduction, which permitted them to address the reaction for signals that happen between Fourier recurrence containers. They consider the investigation of time-arrangement information is a significant apparatus in numerous zones of astronomy, including research including white midgets, dark openings, and neutron stars. Besides, their exploration recommends the way that in the investigation of neutron stars, the time-arrangement examination has had specific significance for pulsar research due to the high cognizance of pulsar periodicities (9) Stallard at. el (2007) in their exploration presents a period space model of a hurling float wave-energy converter and examines the tuning issue is sporadic oceans. The tuning issue is tended to by utilizing both fixed (inactive) and versatile (dynamic) power-take-off settings. The fixed force adopts off tuning strategy incorporates models dependent on tuning the gadget normal recurrence to either the energy recurrence or pinnacle recurrence of the ocean state or a weighted normal of a few pinnacle frequencies. The versatile tuning approaches utilize a sliding discrete Fourier change recurrence investigation, or a period arrangement examination of the deliberate wave height and gadget speed to gauge a restricted predominant wave recurrence and thus figure power-take-off settings. The paper presents subtleties of these tuning procedures by examining issues identified with the displaying, reenactment, and anticipated force catches for every strategy. A near investigation of every technique alongside the commonsense ramifications of the outcomes and suggestions are likewise introduced. The outcomes present that time arrangement examination based tuning procedures can be utilized to build the force caught from unpredictable waves. The most noteworthy force catch is accomplished by ceaselessly tuning the regular recurrence of the gadget to a wave recurrence got from the estimation of the nearby wave time frame.

METHODOLOGY

Quantitative work often involves working with time-series data in various activities. A time series is an ordered sequence of data that typically represents how some quantity changes over time. Examples of such quantities could be high-frequency measurements from a seismometer over a few days, to yearly temperature averages measured at a range of locations across a city, to population changes of different species, but we can use the same software tools to work with them. For decision making in this research, the dataset is interpreted and analyzed using Python. Due to its flexibility and simple coding language, it can support different styles of programming including structural and object-oriented. In Python, it is very popular to use the pandas' package to work with time series. It offers a powerful suite of optimized tools that can produce useful analyses in just a few lines of code. A panda. Data Frame object can contain several quantities, each of which can be extracted as individual pandas. Series object and these objects have several useful methods specifically for working with time-series data.

A. PYTHON

Python is a suitable language for both learning and real-world programming. It supports a wide variety of third-party tools which makes it much easier to use and motivates the users to continue with, Having an elegant syntax. Furthermore, Python is also more forgiving of errors, so you'll still be able to compile and run your program until you hit the problematic part.

A. Time Series Analysis

Time series is a collection of data points collected at constant time intervals. These are analyzed to determine the long term trends to forecast the future it can be briefly differentiated as it is time-dependent. So in case of the basic assumption of a linear regression model, it specifies that observations are independent. Furthermore, along with an increasing or decreasing trend, most time series have some form of seasonality trends, i.e. variations specific to a particular time frame. For example, if you see the sales of a refreshing drink over time, you will invariably find higher sales in summer season as compared to winter".

RESULTS

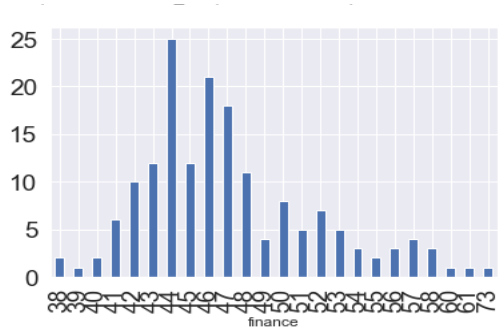
Python and Time series analysis was used to analyze the results the data set relates to new year resolutions having three attributes i.e. gym, finance and diet. The results are shown in the form of various table and graphs.

Table 1 Ratio of attributes on a monthly basis

S.No	Month	Gym	Diet	Finance
0	2004-02	75	26	49
1	2004-03	67	24	47
2	2004-04	70	22	48
3	2004-05	72	22	43
4	2004-06	64	24	45

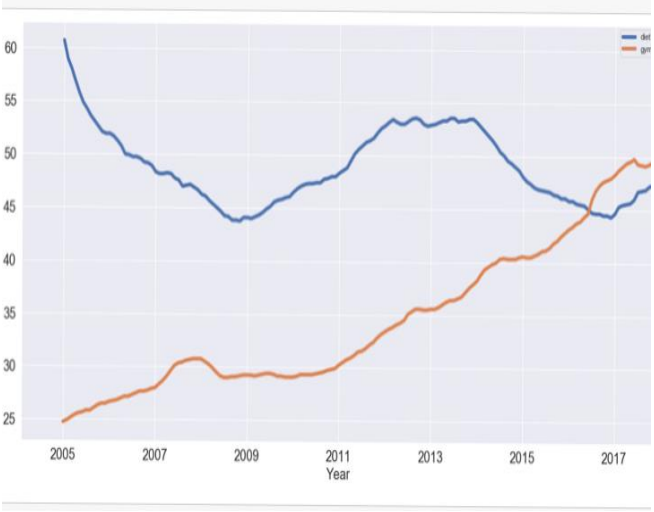
Table 1 shows the ratio the attributes of a corresponding year on monthly basis. In February 2004 the ratios of the gym were 75, whereas diet had 26 and finance had 49. The highest result was observed in the second month of 2004 as the person focused more on the gym as compared to diet and finance. Whereas in the next months the ratios started to drop as the interest of the person began to decline.

Figure 1 Bar Chart Interpretation



Following bar chart shows the results of finance and diet on x-axis finance ratios are shown whereas y-axis shows ratios of diet and gym are shown it is observed that the ratio of finance increased at 43 along with gym & diet.

Figure 2 Comparison chart



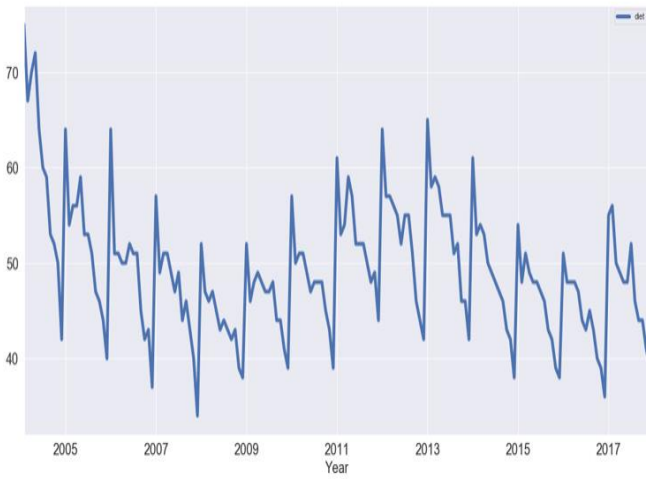
Following graph shows the comparison between gym and diet attribute on yearly basis on x-axis years are shown whereas y-axis represents ratios. It has been observed that in 2005 the ratio of the gym was extremely low i.e. 25 gradually it started to increase with time Whereas in 2005 the person was more diet cautious and neglect gym

These results explain the fact that in the time frame of 2015-2017 the person became more focused and started developing proper interest in doing gym and focused less on diet

The best result was achieved in the timeline of 2015-2017 as he managed both gym and his diet in an efficient way to build up a healthy balanced lifestyle.

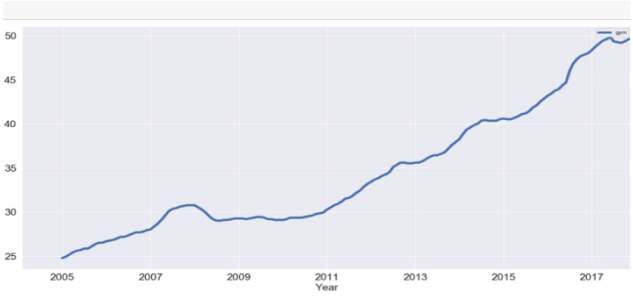
.

Figure 3 graphical representation of diet Attribute



Following graph shows the result of diet attribute on yearly basis on x-axis years are shown whereas y-axis represents ratios. It has been observed that in 2005 the ratio of diet was high i.e.70, gradually it started to decrease with time as the person started to lose interest the lowest ratio was observed in 2017.

Figure 4 graphical representation of gym Attribute



Following graph shows the results of gym attribute on yearly basis on x-axis years are shown whereas y-axis represents ratios. It has been observed that in 2005 the ratio of the gym was extremely low i.e. 25 gradually it started to increase with time as the person became more focused and started developing proper interest in doing gym and made a proper resolution for it in 2011 the highest value recorded was 50 in 2017.

Figure 5 combined graphical representation of attributes

Following graph shows the combined result of the attributes on yearly basis On x-axis years are shown with the interval gap of 2 years whereas on y-axis ratios are shown It is observed that in 2005 the diet had the highest ratio i.e. 70 The result also showed that with the increase in the quantity of diet finance also increased whereas the results of the gym remain consistent proving the fact that the person is more focused on his diet as compared to the gym.

DISCUSSION

Based on the assumptions of results it was discovered that Based on results it is observed that in 2005 the diet had the highest ratio i.e. 70. It has also been observed that with the increase in the quantity of diet finance also increased whereas the results of the gym remain consistent proving the fact that the person is more focused on his diet as compared to the gym. These result further explain the fact that in the time frame of 2015-2017 the person became more focused and started developing proper interest in doing gym and focused less on diet.

The best result was achieved in the timeline of 2015-2017 as he managed both gym and his diet efficiently to build up a healthy balanced lifestyle.

CONCLUSION AND RECOMMENDATION

The results further explain the fact that in the time frame of 2015-2017 the person became more focused and started developing proper interest in doing gym and focused less on diet The best result was achieved in the timeline of 2015-2017 as he managed both gym and his diet efficiently to build up a healthy balanced lifestyle. This research recommends that staying active is a crucial part of maintaining good health and wellness. if a person is making new year resolutions based on health he should remain focus on them according to the years he has planned as it requires a lot of motivation and dedication to follow all the resolutions. one should be capable enough to manage his finance and diet accordingly Make exercise and physical activity a permanent part of your daily routine as well as an important part of your new year resolutions.

REFERENCES

1) Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159-175.
2) Schrager, S. B. (2020). Goal-Setting Theory for the New Year. Family Practice Management, 27(1), 3-3.
3) Munro, C. L., & Hope, A. A. (2020). Healthy Work Environment: Resolutions for 2020.
4) Masum, S., Liu, Y., & Chiverton, J. (2018, June). Multi-step time series forecasting of electric load using machine learning models. In International Conference on Artificial Intelligence and Soft Computing (pp. 148-159). Springer, Cham.
5) Shabib, A., Narang, A., Niddodi, C. P., Das, M., Pradeep, R., Shenoy, V., ... & Sitaram, D. (2015, August). Parallelization of searching and mining time series data using dynamic time warping. In 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 343-348). IEEE.
6) Lippi, M., Bertini, M., & Frasconi, P. (2013). Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning. IEEE Transactions on Intelligent Transportation Systems, 14(2), 871-882.
7) Arandia, E., Eck, B., & McKenna, S. (2014). The effect of temporal resolution on the accuracy of forecasting models for total system demand. Procedia Engineering, 89, 916-925.
8) Srinath, K. R. (2017). Python–The Fastest Growing Programming Language. International Research Journal of Engineering and Technology (IRJET), 4(12), 354-357.
9) Yavuz, H., Stallard, T. J., McCabe, A. P., & Aggidis, G. A. (2007). Time series analysis-based adaptive tuning techniques for a heaving wave energy converter in irregular seas. Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy, 221(1), 77-90.
10) Shumway, R. H., & Stoffer, D. S. (2017). Time series analysis and its applications: with R examples. Springer.
11) McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. " O'Reilly Media, Inc.".
12) Hirsch, R. M., Slack, J. R., & Smith, R. A. (1982). Techniques of trend analysis for monthly water quality data. Water resources research, 18(1), 107-121.
13) Schelter, B., Winterhalder, M., Maiwald, T., Brandt, A., Schad, A., Schulze-Bonhage, A., & Timmer, J. (2006). Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *16*(1), 013108
14) Salisbury, J. I., & Wimbush, M. (2002). Using modern time series analysis techniques to predict ENSO events from the SOI time series.
15) Malyarenko, D. I., Cooke, W. E., Adam, B. L., Malik, G., Chen, H., Tracy, E. R., ... & Manos, D. M. (2005). Enhancement of sensitivity and resolution of surface-enhanced laser desorption/ionization time-of-flight mass spectrometric records for serum peptides using time-series analysis techniques. *Clinical chemistry*, *51*(1), 65-74.
16) Strozzi, F., Zaldívar, J. M., & Zbilut, J. P. (2002). Application of nonlinear time series analysis techniques to high-frequency currency exchange data. *Physica A: Statistical Mechanics and its Applications*, *312*(3-4), 520-538.
17) Filkov, V., Skiena, S., & Zhi, J. (2002). Analysis techniques for microarray time-series data. *Journal of Computational Biology*, *9*(2), 317-330.
18) Rehfeld, K., Marwan, N., Heitzig, J., & Kurths, J. (2011). Comparison of correlation analysis techniques for irregularly sampled time series. *Nonlinear Processes in Geophysics*, *18*(3), 389-404.
19) Oliphant, T. E. (2007). Python for scientific computing. *Computing in Science & Engineering*, *9*(3), 10-20.
20) Millman, K. J., & Aivazis, M. (2011). Python for scientists and engineers. *Computing in Science & Engineering*, *13*(2), 9-12.
21) Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830