**COULD A TIME SERIES OF REVENUE FROM THE INITIAL TWO YEARS OF BUSINESS INDICATE SIGNIFICANT TRENDS?**

**TASK 1**

**D213**

**STUDENT NAME: IBRAHIM SULEIMAN**

**DATA ANALYTICS**

**STUDENT NUMBER: 001429984**

**COLLEGE OF INFORMATION TECHNOLOGY, WESTERN GOVERNORS UNIVERSITY**

**NAME OF INSTRUCTOR: DANIEL SMITH**

**NOVEMBER 25TH, 2022**

**Part 1: Research Question**

Section A: Description of the Report

Section A1:

Could a time series of revenue from the initial two years of business indicate significant trends?

Using a time-series model to project future sales, we will respond to this research topic. Data points collected over a period of time may be analyzed using time series analysis. The main goal is to use statistical methods to examine the data and make it possible to make future forecasts.

Section A2: Goal of the data analysis

The goal of this investigation is to employ time series modeling on this collection of days and income data to look for any potential trends. Future results might then be predicted using this data.

Part II: Method Justification

The assumption that the data are stationary is one that many time series approaches make. The mean, variance, and autocorrelation structures do not alter over time in a stationary process. How closely a time series is associated with its earlier values is known as autocorrelation (Nao R2018).

Part III: Data Preparation

Data transformation and column type conversion were the main components of the data preparation step. A line graph representation of the time series is attached:Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

Chart

Description automatically generated with medium confidence

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, Word

Description automatically generated

Graphical user interface, text, application, Word, email

Description automatically generated

Section C5: Copy of cleaned dataset

The cleaned dataset was provided as csv file in the submission named;

We exported our prepared dataset as

df.to\_csv('prepared\_d212task1.csv', index = False)

Part IV: Model Identification and Analysis

Graphical user interface, text, application, email

Description automatically generated

Timeline

Description automatically generated

Graphical user interface

Description automatically generated with medium confidence

We can see that there is seasonality in the data by analyzing the time series dataset. Through the years, the tendency seems to be good.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

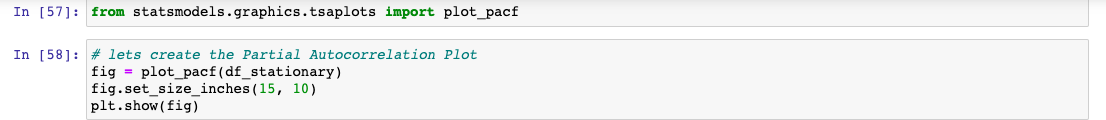
Description automatically generated



The auto correlation function demonstrates the strong association between each observation and its most recent history.

Graphical user interface

Description automatically generated with medium confidence



Table

Description automatically generated

Graphical user interface, text, application

Description automatically generated

To suit the dataset, a model was developed and tested. The dataset was also predicted to determine if a positive or negative trend would emerge.

Table

Description automatically generated

Chart, line chart

Description automatically generated

Chart, histogram

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

A picture containing shape

Description automatically generated

A picture containing shape

Description automatically generated

A picture containing graphical user interface

Description automatically generated

A picture containing shape

Description automatically generatedBackground pattern

Description automatically generated with low confidence

Graphical user interface, chart, line chart

Description automatically generated

Part V: Data Summary and Implications

We used a prediction algorithm to process every value from the original dataset. As a result, for each original value, we were given a different predicted value. The duration of the forecast and predictions is equal to that of the first data set.

For each anticipated vs. real value, the model assessment is shown, and the RMSE is 0.517. The above final visualization shows the final model in comparison to the test set. One course of action a business may take based on this is to be aware that time series testing can be utilized to detect patterns and values.

One course of action a business may take based on this knowledge is understanding that time series testing can be utilized to uncover patterns and values. Although the RMSE for this test is just 0.517, a business might conduct many more tests on a larger volume of data to obtain far more definitive conclusions. The information might then be used to inform business choices.

Section G: Sources

Hyndman, R. et al. (2008). Forecasting with Exponential Smoothing: The State Space Approach. Springer Science & Business Media.

Nao, Robert. (2018). Introduction to ARIMA models. <https://people.duke.edu/~rnau/411home.htm>

Shanan, S. (2018, October 3). Time series Forecasting — ARIMA models. Retrieved from Towards Data Science: <https://towardsdatascience.com/time-series-forecasting-arima-models-7f221e9eee06>

Evgeniy R.(2020). 5 Stages of Data Preprocessing. Medium. <https://medium.com/@evgen.ryzhkov/5-stages-of-data-preprocessing-for-k-means-clustering-b755426f9932>