upGrad

#LifeKoKaroLift

$\{ \begin{array}{l} \rightarrow \text{Step} \\ \rightarrow \text{Python} \end{array}$

Time Series: Assignment

upGrad

Course : Data Science
Lecture On : Assignment
Instructor : Sumit Shukla

What we will cover in this session?

1   Understanding Coefficient of Variation (CV)

2   Assignment walkthrough

3   QnA

upGrad

Idea behind Coefficient of Variation(CV)

The coefficient of variation represents the ratio of the standard deviation to the mean, and it is a useful statistic for comparing the degree of variation from one data series to another, even if the means are drastically different from one another.(source: investopedia.com)

$$\overset{\times 3}{\overbrace{X = \quad 1, \quad 2, \quad 3}}$$

$$\overline{x} = 2 \qquad S_x = 1$$

$$\underset{+2}{\underbrace{Y = \quad 101, \quad 102, \quad 103}}$$

$$\overline{y} = 102 \qquad S_y = 1$$

$$cv(x) = \frac{S_x}{\overline{x}} = \frac{1}{2} = 0.5$$

$$cv(y) = \frac{S_y}{\overline{y}} = \frac{1}{102} = 0.0098$$

Lower the cv value, more that series is stable.

4

Question-1: Refer to the data with mean and standard deviation for two samples. Which sample is having more fluctuations or higher spread?

A • Sample-1
B • Sample-2 ✓
C • Can't say

$$CV(S-1) = \frac{0.51}{7.2} = 0.07 \quad (\text{stable})$$

$$CV(S-2) = \frac{1.92}{7.3} = 0.26 \quad (\text{fluctating})$$

	sample1	sample2
	6.5	4.2
	6.7	5.4
	6.8	6.2
	7.2	7.7
	7.3	7.7
	7.4	8.4
	7.8	9.2
	7.9	9.8
Mean	7.2	7.3
Standard Deviation	0.51	1.92

Global Mart is an online supergiant store that has worldwide operations. This store takes orders and delivers across the globe and deals with all the major product categories — consumer, corporate & home office.As a sales manager for this store, you have to forecast the sales of the products for the next 6 months, so that you have an estimate of the demand and the sales of the products in those months and can plan your inventory and business processes accordingly.

Handwritten annotations (left side): Daily basis; forecast 2012-07; sales for the next 2013-05; 6-month for that 2013-09; particular 2013-01; market-segment 2013-11; which is most consistency profitable =

Order Date	Segment	Market	Sales	Quantity	Profit
31-07-2012	Consumer	US	2309.65	7	762.1845
05-02-2013	Corporate	APAC	3709.395	9	-288.765
17-10-2013	Consumer	APAC	5175.171	9	919.971
28-01-2013	Home Office	EU	2892.51	5	-96.54
05-11-2013	Consumer	Africa	2832.96	8	311.52
28-06-2013	Corporate	APAC	2862.675	5	763.275
07-11-2011	Consumer	APAC	1822.08	4	564.84
14-04-2012	Consumer	APAC	5244.84	6	996.48
14-10-2014	Corporate	US	5083.96	5	1906.485
28-01-2012	Consumer	US	4297.644	13	-1862.3124
05-04-2011	Corporate	US	4164.05	5	83.281

Handwritten annotations (right side): market -segment; US - Consumer; APAC - corporate; APAC - Consumer; :

Two stage

① stage → find out most consistently
Profitable market segment
(Look for that market-segment with
least COV value)

filter

② stage → forecast sales for the next 6 month
for that particular market-segment.

Stage-2 :

i) Use the (stage-2, step-II) df
and filter it for the chosen
market-segment

ii) agg. the data on order-date to
find sum of sales

iii) split the data into train-test
(42) (6)

iv) forecasting.

Stage-1

df →

i) Convert order-date to year-month
ii) Create a new column — market-segment
iii) Aggregate the data in the following way
iv) Split your data into train-test
(42) (6)

v) use train data and find out the most
consistently profitable market segment. (cov)

market-segments

order
date

→ Sum of
profit
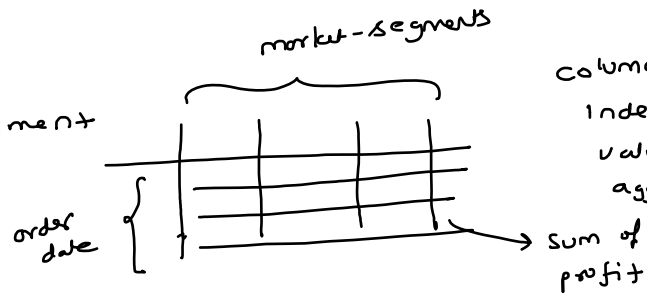
column = market-segment
Index = order-date
values = profit
agg-func = sum

Now, due to certain unpredictable circumstances in the market, as a company, you are prioritizing only the best and most consistent market segment in terms of profitability. You want to see which market segment is the most consistently profitable and accordingly, you would want to check how the sales forecasts look for the next 6 months.

Market	Segment
Africa	Consumer
APAC (Asia Pacific)	Corporate
Canada	Home Office
EMEA(Middle East)	
EU (European Union)	
LATAM (Latin America)	
US (United States)	

Next Step:

- After selecting the best market-segment, filter the data for that particular market-segment.
- Group the data based on order_date and find the Sales

Forecasting:

- Split the data into train and test. Last six month will be your test data.
- Perform following time series forecasting method for Sales.
  - ~~Naive Method~~
  - Simple average
  - Simple Moving Average
  - Simple Exponential Smoothing Technique
  - Holt Method
  - Holt Winters' additive method
  - Holt Winters' Multiplicative method
- Calculate MAPE for all the models and keep the track of it in a DataFrame as shown in the lecture videos.

upGrad

Forecasting:

- Perform Stationarity test and make the time series stationary.
- Perform following time series forecasting method for Sales.
  - AR
  - MA
  - ARMA
  - ARIMA
  - SARIMA
- Calculate MAPE for all the models and keep the track of it in a DataFrame as shown in the lecture videos.

Presentation:

- The presentation asked in the submission explains all the required details. The coefficient of variation is explained well for the 21 market segments and which market segment is the most profitable and why it is most profitable is also explained.
- Explaining well the optimum technique from the flow chart that might work best for the sales forecasts.
- Explaining well the insights from the plots and the MAPE values derived from the 3 smoothing techniques and ARIMA set of techniques.
- Concluding with reasons which smoothing technique and ARIMA technique works best for sales forecasts.

on an-average how much
% the actual values
differ from forecasted
values
$$50\%$$

$$MAPE = \frac{100}{n} \times \sum \left\| \frac{y - \hat{y}}{y} \right\|$$

$$(100\%) = \begin{bmatrix} 10 & \text{actual } y = 10 \\ 20 & \text{predicted } y = 5 \end{bmatrix} = 50\%.$$

$$\frac{10 - 5}{10} = \frac{5}{10} = 0$$

Thank You!

23/05/19