Two csv files are attached: trends1.csv and predictions.csv

These have information for 3 types of tops.
In trends1.csv, each top has a score for popularity of those types of items, for the date.

predictions.csv has a fit that has then been extrapolated to the next 365 days.

Notes :
- Figure 1 – 9 : *Trends* sheet
- Figure 10 – 15 : *Given Predictions* sheet
- Figure 16 – 17 : *My Predictions* sheet
- Figure 18 – 23 : *Prediction Comparison* sheet
- *Trend Value* sheet and *Growth Value* sheet used for Question 3

Please complete the following tasks:
1) Assess the three trends in trends1.csv, which is performing best. Elaborate on how you defined "best performance".
   The trends for Top 1, 2 and 3 are plotted as shown in Figure 1. The linear trendlines for all three datasets given are plotted as shown in Figure 2.

   The linear trendlines for Top 1 and 3 in Figure 2 look quite similar as they seem to have the same gradient. Thus, the growth trendlines for all three datasets are plotted as shown in Figure 3 for better data analysis.

   Figure 3 shows that the gradient of the growth trendline for Top 1 is increasing while it is constant for Top 3. The trendline gradient represents the change in popularity score of the item. The higher the trendline gradient, the higher the popularity score of the items in future. This tell us that Top 1 trends performs better than Top 3. As a result, even though the popularity scores of Top 1 data are mostly lower than that of Top 3, it will exceeds that of Top 3 at one point later. However, according to Figure 2 and 3, it is noted that the trendline gradient for Top 2 is the steepest compared to others.

   There are lots of factor that can affect the popularity score of the items such as the item pricing, brand, item availability, packaging and product placement. These factors affect the change in popularity score of the items with respect to time.

   Defining the "best performance" as the steepest gradient of the trendline, this implies that Top 2 trend performs the best compared to trends Top 1 and 3.

2) Quantify the performance over the last year - to establish what proportion of demand has changed.
   Over the last year, the popularity score of items for Top 2 data has been the highest compared to that of Top 1 and 3. This is probably due to the a better branding and packaging. A good brand name and logo can both impact the popularity score. Same thing applies for the item packaging as a good item packaging indicates a good representation of selling item towards the customer. Besides that, a higher popularity score of items for Top 2 compared to others

can also because of a better product placement. For example, a clear product placement such as on an end cap in a highly trafficked area of the store, or on the front page or in the side bar of an ecommerce business like Amazon, eBay can boost the popularity score. The fluctuation of the popularity score of the items over time for all datasets given are probably because of factors like pricing, reputation and availability of the items.

For the case of pricing, there would probably be sales at certain times where the item price would be reduced, resulting in a rising in popularity score. The popularity score would then be dropped after the sale ends. Besides that, the reputation factor can also be considered in this case as a better advertisement of the item can bring a greater awareness to the customers regarding the item and thus raising the popularity score. A lesser advertisement induces a drop in the popularity score. Same thing applies for the availability of the item. If the item is out of stock, the popularity score would definitely be decreases until it is available again.

3) Look at the predictions.csv file - there are three fits for each trend, comment on which you find to be best.
Every three trendlines given for each trend are plotted with their corresponding trend, as shown in Figure 10 to 12 for linear trendline and Figure 13 to 15 for growth trendline. The best fit for each trends can be roughly identified by observing the Figure 10 to 15.

For Top 1, Figure 10 and 13 show that the best trendline should be either Fit 2 or Fit 3. By comparing the trendline value obtained for Fit 2 and 3(see *Trend Value* sheet and *Growth Value* sheet), it is found that Fit 3 is the best fit as it has a greater popularity score prediction than that of Fit 2 for every same date (dates for those next 365 days).

For Top 2, Figure 11 and 14 show that the best trendline is clearly Fit 1. This can be confirmed by making a comparison between the trendline values for Top 2 obtained for each fits (see *Trend Value* sheet and *Growth Value* sheet) to compare which one has a higher popularity score prediction for the next 365 days after the final date of the initial trend.

By using the same method above, the best trendline for Top 3 is Fit 1.

4) Quantify these predictions into an assessment of how you would expect the demand to change over the next 3, 6, 12 months.
The prediction tell us that the raise in the popularity score for the items will keep growing for the next 365 days and also over next 3, 6, 12 months. This could be probably due to the same factors mentioned in part 1 which are good branding, packaging and product placement.

However, if we take a consideration of external factor, it is not impossible for the demand to decrease either gradually or rapidly, depending on the factor. The example of external factor would be such as the change in style or fashion trend, the strength of currency, and many more.

The fluctuation of the popularity score would always occurs due to the mentioned factor such as pricing, reputation and availability.

5) Please create your own predictions based of the three types of tops.
   Refer to *Data Analysis.xlsx,* under *My Prediction* sheet.

6) Briefly explain your choice of model, and how it performs better or worse than the provided predictions.

   In Figure 2 and 3, since there is no zero or negative values of popularity scores, the data can be assessed using growth trendlines. The score value for trend Top 3 seems to be increasing at a steady rate, so the trendline mode can be chosen to be either linear or exponential. However, since the score values for trends Top 1 and 2 increase at a non-steady(increasing) rate, the most suitable trendline mode for Top 1 and 2 data are probably exponential. This can be seen by comparing the solid lines of the same colour between in Figure 2 and 3.

   This can be further investigated by comparing the residual plots between linear and exponential modes for each trend(see Figure 4 to 9). For Top 1, it turned out that the residual plots for both linear and exponential are not differ significantly (refer to Figure 4 and 7) so both modes are acceptable to use. Same thing applies for Top 3 (refer to Figure 6 and 9).

   However, for Top 2, most of the residuals for exponential mode are not deviated from x-axis as much as that of linear mode(refer to Figure 5 and 8). This implies that the most suitable trendline mode for Top 2 is exponential.

   My choice of model would probably be the nonlinear regression since all data in Figure 3 (nonlinear regression) has a greater value of R^2 (Coefficient Of Determination) compared to the corresponding data in Figure 2, which is the linear regression. R^2 is a statistical measure of how close the data are to the fitted regression line. In general, the higher the value of R^2, the better the model fits the given data.

   Based on Figure 21 to 23, it is noted that the self-made prediction for each trends perform better than the provided one. However, should the linear regression is chosen as the choice of model, only the self-made prediction for trend Top 1 performs better than the provided one but not for trends Top 2 and 3. This can be seen in Figure 18 to 20.

   End.