☆ Retail Case Study [Forecasting]

☆ Key players in retail
- Manufacturers
- wholesaler
- Retailers
- Consumers

☆ Types of Retail Stores
- Departmental [Spar]
- Supermarket [Costco]
- Convenience [7-11]
- Speciality [Apple]
- Discount [Walmart]
- E-tailer [Flipkart]

☆ Retailers be it conventional / Ecomerce use data as leverage to build models for consumer engagement, Supply chain optimization, to target desired customer & see some influential trends.

☆ Purchasing Behaviours can be analyzed by : Association Analysis
Market-Basket Analysis

☆ Alongwith this use : Recommendation Engine
Collaborative Filtering techniques

- Customer Segmentation using predictive modelling techniques like (k-means) is done.

- Web & Social Media Analytics is another technique which uses (NLP) & (Sentiment analysis)

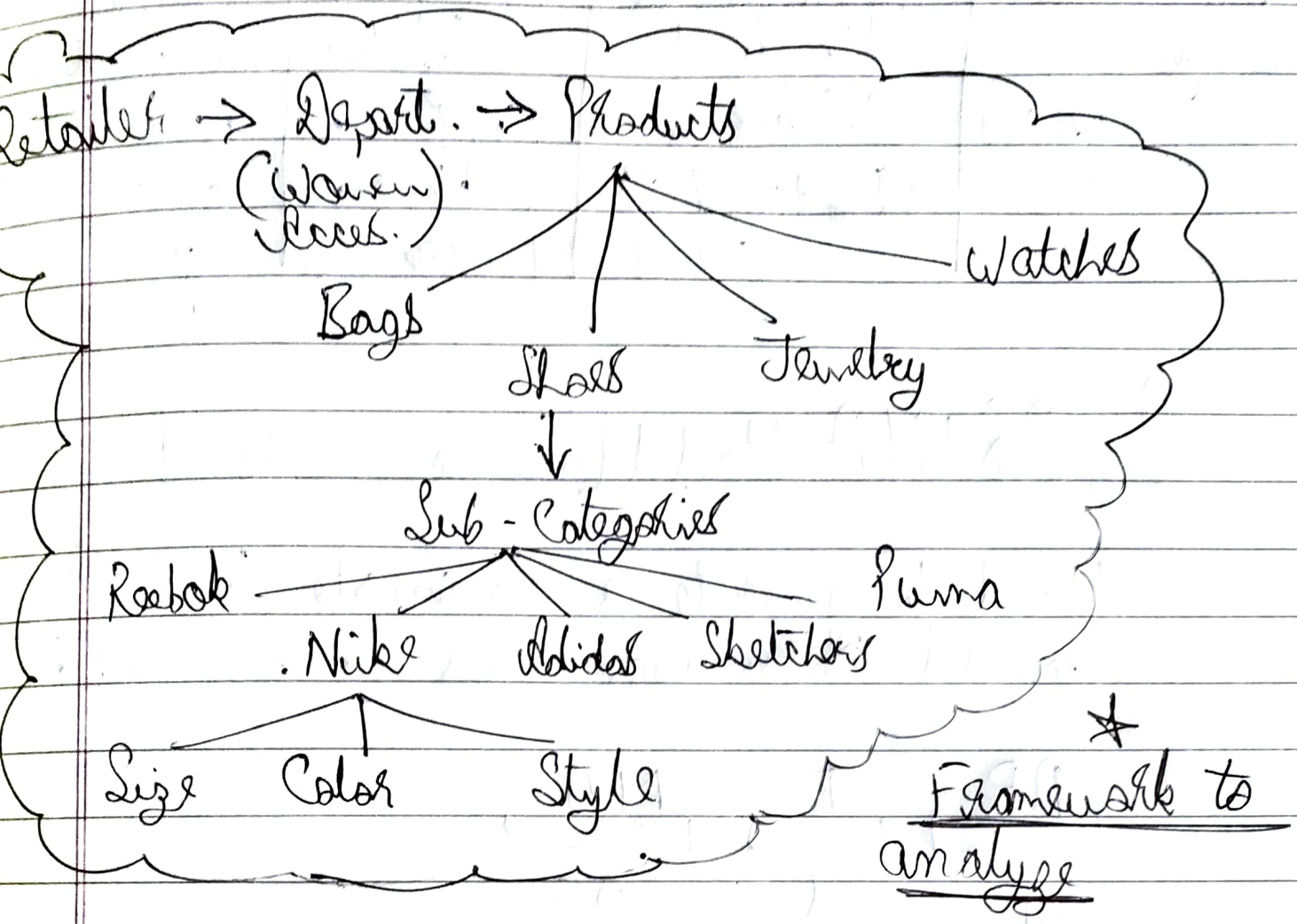
- Supply Chain is movement of final goods from manufacturers to suppliers & finally to the consumer.

- Co.'s can use predictive analytics classification Models like (decision tree) (logistic regn.) to predict the delay of shipment.

- Price Optimization
  - Price Elasticity
  - Price Level Optimization [Max^z Rev.]

- Advance Analytics like (Decision tree), (Logistic Regn.), (Demand Model) & (Price El.) are used in deciding optimum prices

★ Space Optimization & Assortment Planning



★ Case Study : SARIMA (forecasting)

* Data is related to a retailer known as Liberty which contains sales of foods & beverages.

* Since forecasting can only be done for a T.S data, we would have to convert irregular data to T.S format.

* Time Series data follows diff. patterns such as trend, seasonal, cyclical & random components.

Trend	Seasonal	Cyclic	Random
Upward/Down. Usually for longer periods	Recurring sales like due to Christmas .. [for fixed period]	Eco. Recession or Inflation [not fixed & Uncertain]	Unexplained & unpredictable left when all other patterns removed from T.S

★ ARIMA Model [Box-Jenkins]

Used to generate a univariate time series model to analyze sales & forecast it.

$$\underbrace{AR}_{p}\ \underbrace{I}_{d}\ \underbrace{MA}_{q}$$

★ **AR Model**: where dept. is regressed against past values.

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \varepsilon_t$$

★★ Value at time (t) is predicted from values at time (t-1)

★ **MA Model**: where dept. is regressed against past errors terms.

$$Y_t = \beta_0 + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \dots \phi_q \varepsilon_{t-q}$$

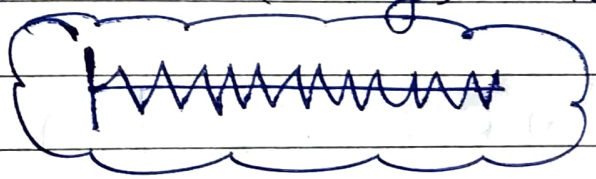
Error at time (t) depends on past values of error term (ε_{t-1})

* When both combined we get ARMA Model.

* ARMA show how my future sales of food & beverages is depended on my previous sales of food & " & errors associated with it. It fails to show any patterns.

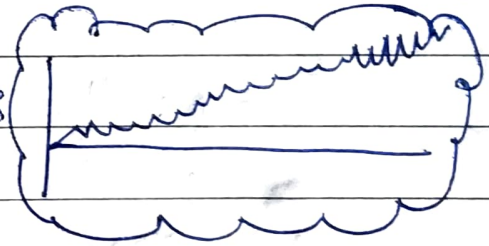
* ARIMA tells the same infor. but also considers the trends & patterns along the time series.

* We assume (always) : Time Series = Stationary



[Constant]
[ups & downs]

* However, Non Stationary T.S : Show trends or patterns. & (Y_t) does depend on time.



* We would have to convert irregular T.S to a regular T.S by differencing method where,

$$T.S \sim (0, \sigma^2)$$

mean, var.^{n} constant [log Transformation]

3 - Steps of ARIMA Modelling

- Check if series stationary / not
- If yes proceed / If not, start differencing
- Check if series has white noise [Autocorr.]
- If present, ARIMA not legal.

1) **Identification:** Autocorrel. factor (ACF) & Partial " " (PACF).

* Autocorrel. = Data is correlated to itself

$$\text{If } Cov.(e_i, e_j) \neq 0$$

* Ex: If temp. on 1st & 2nd day of month is more similar than temp. that occured on 31st, the data is autocorrel.
* Ex: Regrn. Residuals Correl. to each other.

Correl. b/w Y_t & Y_{t-1}

* ACF helps in identifying no. of lags for (q)
* PACF helps in " " " " " (p)

Defined as correl. b/w Y_t & Y_{t-1} for adjusted time lags

* Model with lowest AIC = Best

* II) Estimation [of parameters]

* Methods like Yule Walkers, Method of Moments, Cond.^{n} least Sq., Max.^{n} Likelihood

* Check adequacy of model with t-values

* Check Goodness of fit & choose model based on lowest AIC & BIC.

* Chi-Sq. for residuals : Helps check white-noise

If residuals autocorrelated : NO white noise
If residuals uncorrelated : White Noise

III) Forecasting
Where predicted / future values are forecasted along with Std. error estimate & (95%) C.I.

★ SARIMA

Represents both seasonal / non seasonal factors present in multiplicative model.

$$ARIMA(p, d, q) \times (P, D, Q)_s$$

★ Prediction Accuracy

1) **Mean Abs. Error** : Scale dependent & cannot compare series with diff. units.

$$MAE = mean(|\hat{x} - x|)$$

abs. forecast error (for - obs.)

2) **Root Mean Sq. Error** : Measures of how spread the prediction errors are.

Scale dependent & compare forecasts with diff models.

- The smaller the error, the better the prediction accuracy

$$RMSE = \sqrt{mean(e_i^2)}$$

* **Mean Abs (%) Error** : Scale Indep. & helps comparing diff. forecast models with diff. units.

- Here, Accuracy = (%) terms.

$$MAPE = \left(\frac{100}{n}\right) \sum_{t=1}^{n} \left| \frac{(x_t - \hat{x})}{x_t} \right|$$

- Lower the MAPE, better the prediction accuracy.

(Not recommended)

* Use PACF to determine terms used in AR model

* Use ACF to determine terms used in MA model

* If PACF has more sig. values then we chose that model (AR) of order = no. of sig. values = No. of Indept. variables

* If ACF has sig. values then we chose MA (model) of order = no. of sig. value = No. of Indept value

Chose the simpler model with less Indept. var.'s

Horz. lines = Sig. threshold

* If PACF & ACF both have the 1st order term to be sig. then we can choose both AR(1) & MA(1) in our model.