PW SKILLS

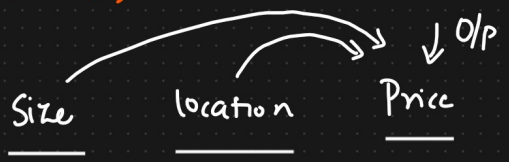
① Covariance And Correlation

Covariance indicates the relationship of two variables whenever one variable changes. If an increase in one variable results in an increase in the other variable, both variables are said to have a positive covariance. Decreases in one variable also cause a decrease in the other.

X	y
2	3
4	5
6	7
8	9

{Relationship between X and Y}

X ↑	y ↑
X ↑	y ↓
X ↓	y ↑
X ↓	y ↓



X ↑	y ↑
X ↓	y ↓

X ↓	y ↑
X ↑	y ↓

$$\text{Covariance}(X,Y) = \sum_{i=1}^{n} \frac{(x_i-\bar{x})(y_i-\bar{y})}{n-1}$$

$$\text{Cov}(x,y)$$

$$\text{Var}(x) = \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n-1}$$

⇓

X ↑	y ↑
X ↓	y ↓

+ve Cov

Spread of the data $\Leftarrow$ $Cov(x,x) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})}{n-1}$

X ↑	y ↓
X ↓	y ↑

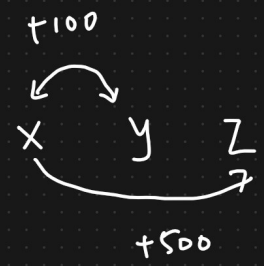
-ve Cov

$$Cov(x,y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

X	y
2	3
4	5
6	7
$\bar{x} = 4$	$\bar{y} = 5$

$$= \frac{(2-4)(3-5) + (4-4)(5-5) + (6-4)(7-5)}{2}$$

$$= \frac{4 + 0 + 4}{2} = 4 \text{ +ve}$$



X & y are having a positive variance

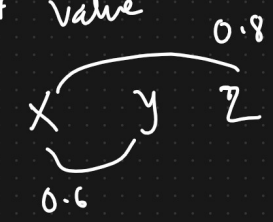
Advantages

① Relationship between X & y

+ve or -ve value

Disadvantage

① Covariance does not specific limit value



② Pearson Correlation Coefficient [-1 to +1]

The Pearson coefficient is a type of correlation coefficient that represents the relationship between two variables that are measured on the same interval or ratio scale. The Pearson coefficient is a measure of the strength of the association between two continuous variables.

$$\rho_{x,y} = \frac{cov(x,y)}{\sigma_x \sigma_y} = [-1 \text{ to } 1]$$

① The more the value towards +1 the more +ve correlated it is (x,y)

② The more the value toward -1 the more -ve correlated it is (x,y)

Dataset : 1000 features (ML Models) O/p or Dependent feature

Independent features ⇓ ⇑

Haunted	Size of the house	No. of Rooms	Location	No. of people staying	Price
-	-	-	-	-	-

Feature Selection : near to 0 ⇒ DROP No of people staying

better than person

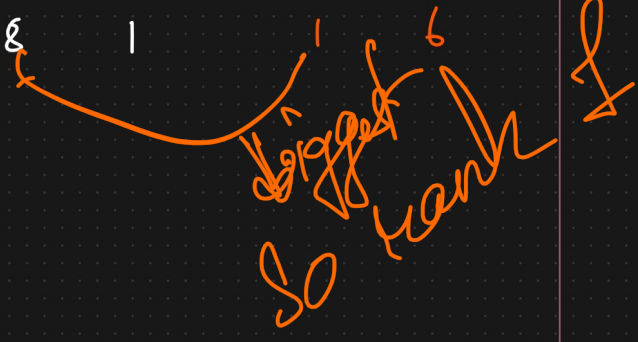
③ Spearman Rank Correlation

Rank

$$r_s = \frac{cov(R(x), R(y))}{\sigma_{R(x)} * \sigma_{R(y)}}$$

Range

X	Y	R(X)	R(Y)
1	2	5	5
3	4	4	4
5	6	3	3
7	8	2	1
0	7	6	2


8
1
1
6
biggest
so rank 1