Assignment-

Task # 1

Merge the 3 dataset and create 1 view of data.

You can merge insurance_data.csv and employee_data.csv on

AGENT_ID You can merge insurance_data.csv and vendor_data.csv on

VENDOR_ID Note: Use left Outer join as not all claims require Vendor

Task #2

Business Leader wants to find Top 3 Insurance Type where we are getting most insurance claims?

Task #3

Business Leader wants to find Top 5 States where we are getting most insurance claims for customer belonging to HIGH(H) risk segment?

Task #4

Business wants to create a new variable “COLOCATION” which will have following values IF

Customer State == Incident State == Agent Address State THEN 1 ELSE 0

Find the mean of this new column

Task #5

Data entry error was detected in the data and you are required to correct it. If for any claim transaction “AUTHORITY_CONTACTED” is NOT “Police” and POLICE_AVAILABLE == 1 Then Update “AUTHORITY_CONTACTED” to Police.

Task #6

Business wants to check the Claim Amount for deviation for each transaction, they would like you to calculate as follow

CLAIM_DEVIATION = AVG_CLAIM_AMOUNT_FOR_LAST_30DAYS (same insurance type) / CURRENT_CLAIM_AMOUNT

If the value < 0.5 THEN CLAIM_DEVIATION = 1 ELSE 0

If there is less than 30 days of transaction history THEN

-1 Note: LAST_30DAYS does not include current day

Task #7

Find All Agents who have worked on more than 2 types of Insurance Claims. Sort them by Total Claim Amount Approved under them in descending order

Task #8

Mobile & Travel Insurance premium are discounted by 10%

Health and Property Insurance premium are increased by 7%

Life and Motor Insurance premium are marginally increased by 2%

What will be overall change in % of the Premium Amount Collected for all these Customer?

Task #9

Business wants to give discount to customer who are loyal and under stress due to Covid 19. They have laid down an eligibility Criteria as follow

IF CUSTOMER_TENURE > 60 AND EMPLOYMENT_STATUS = “N”
AND NO_OF_FAMILY_MEMBERS >=4 THEN 1 ELSE 0

Create a new column “ELIGIBLE_FOR_DISCOUNT” and find it

mean.

Task #10

Business wants to check Claim Velocity which is defined as follow

CLAIM_VELOCITY = NO_OF_CLAIMS_IN_LAST30DAYS (for the current insurance type) / NO_OF_CLAIMS_IN_LAST3DAYS (for the current insurance type)

Note: LAST30DAYS & LAST3DAYS does not include current

day

Task #11

Find all low performing agents i.e. employees who are in the bottom 5 percentile based on Claims worked by them.

Task #12

Business wants to find all Suspicious Employees (Agents).

IF TOTAL CLAIM AMOUNT which meet below criteria is >= 15000 THEN AGENT is classified as Suspicious ELSE Not

CLAIM_STATUS = Approved AND CUSTOMER_RISK_SEGMENTATION = High AND INCIDENT_SEVERITY = “Major Loss”

If Suspicious, then 1 ELSE 0. Find mean of this column.