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.

Solution-
All three csv file merge and create new file with name "insurance_employee_vendor".

Task # 2

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

Solution-
Top 3 Insurance Type:

- Property with 1692 claims
- Mobile with 1692 claims
- Health with 1690 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?

Solution-
Top 5 States, claims
1. CA   148
2. AZ   105
3. FL   104
4. TN   100
5. AR   97

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.

Solution-
The mean of COLOCATION new column is 0.0044.

Assignment

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.

Solution-

Updated the AUTHORITY_CONTACTED column with 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.

Solution-

There are 9216 claims with deviation less than 0.5 and 784 claims with transaction history of less than 30 days.

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.

Solution-

All Agents who have worked on more than 2 types of Insurance Claims
Few agents list is attached below

AGENT00807	Don Filkins	528800
AGENT00679	Clara Barnett	489000
AGENT00771	Roger Burns	422100
AGENT00125	Salvador Soriano	400400
AGENT00789	Alison Baron	392900

Assignment

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?

Solution-

Overall change in % of the Premium Amount Collected for all these Customer is 2.68%.

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.

Solution-

Final mean is 0.0299 and customer eligibility for discount is 2.29%.

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.

Solution-

	NO_OF_CLAIMS_IN_LAST30DAYS	NO_OF_CLAIMS_IN_LAST3DAYS	CLAIM_VELOCITY
Health	136	14	9.714286
Life	140	14	10.000000
Mobile	134	10	13.400000
Motor	122	13	9.384615
Property	129	13	9.923077
Travel	125	14	8.928571

Assignment

Task #11

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

Solution-

Few low performing agents are in table below

AGENT_ID	CLAIM_AMOUNT
AGENT00006	22400
AGENT00014	7400
AGENT00016	27000
AGENT00085	25700
AGENT00088	27700
AGENT00091	30000
AGENT00098	28200

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.

Solution-

The mean of this Suspicious column is 0.12416666.