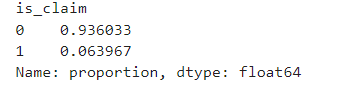
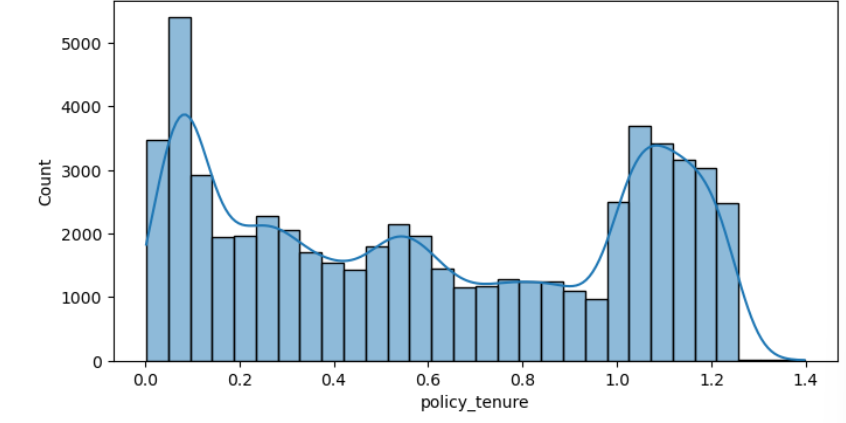
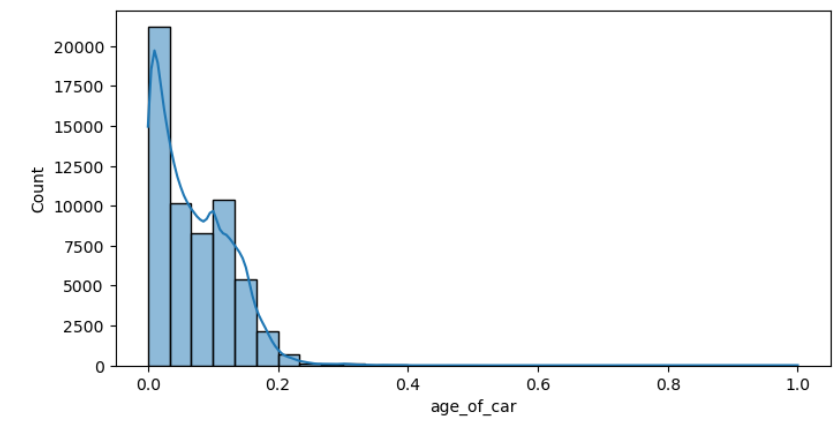
General Observations:-



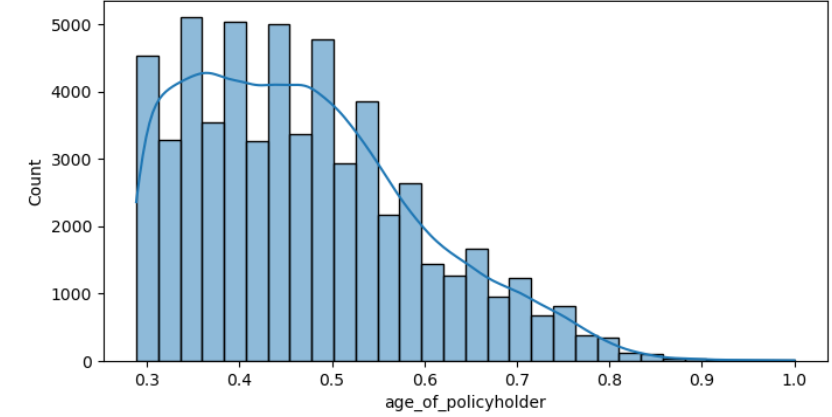
1. According to the provided Dataset, **93% of the people don't claim insurance**, while **7% of the people claim insurance.**



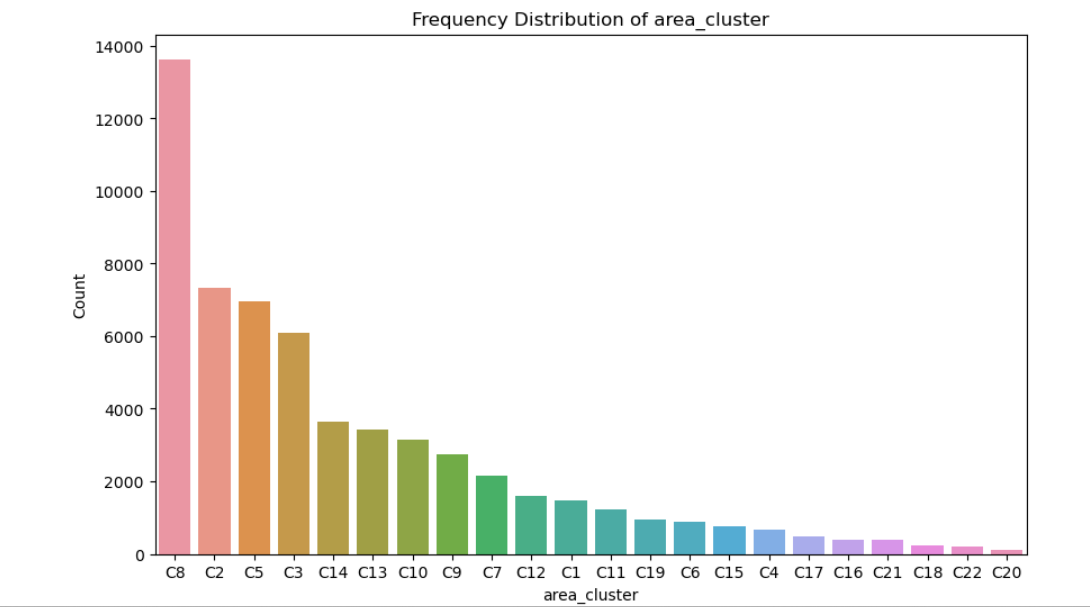
1. There is a significant peak at the beginning of the policy tenure range (0.0467 - 0.0933). **This indicates that a substantial number of new policies have been added, suggesting a considerable influx of new customers**. (histogram-policy\_tenure)[Number of Bins=30]
2. **The presence of two significant peaks suggests a bimodal distribution, indicating two distinct groups of policyholders based on policy tenure.** One group prefers shorter-term insurance policies, while the other opts for longer-term commitments. The **notable peak around the policy tenure of 1.0 to 1.2.**



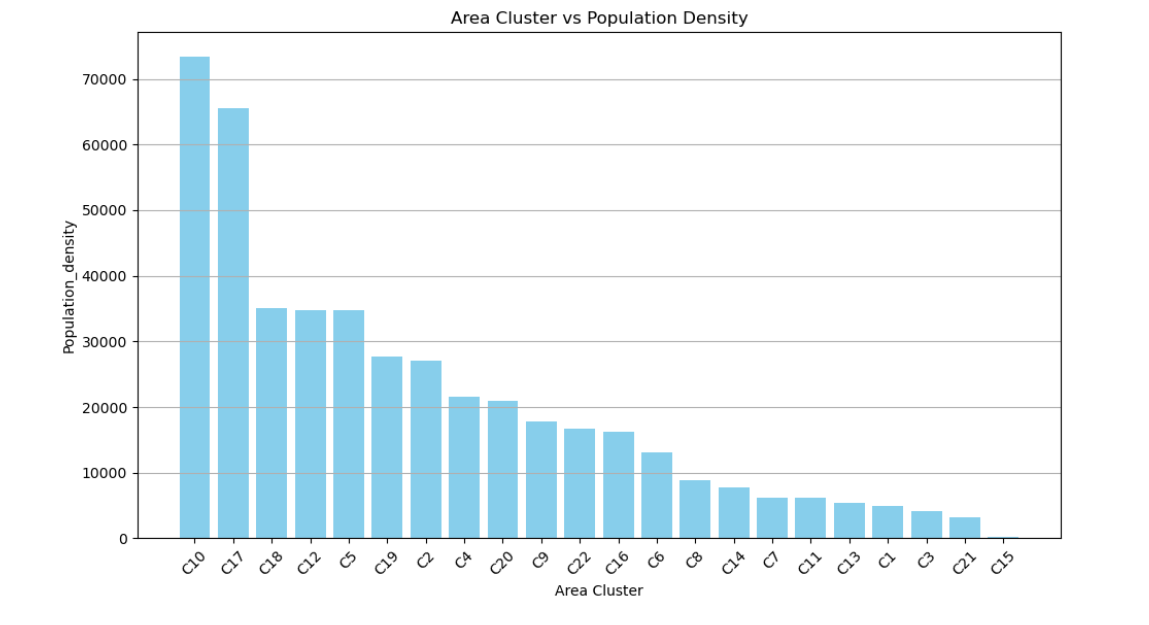
1. The **peak of the distribution is near zero,** indicating that a significant number of new insurance policies are **purchased for newer cars**.
2. **As cars age, fewer people tend to opt for insurance policies**. Generally, individuals are more likely to **forgo insurance for older vehicles.**

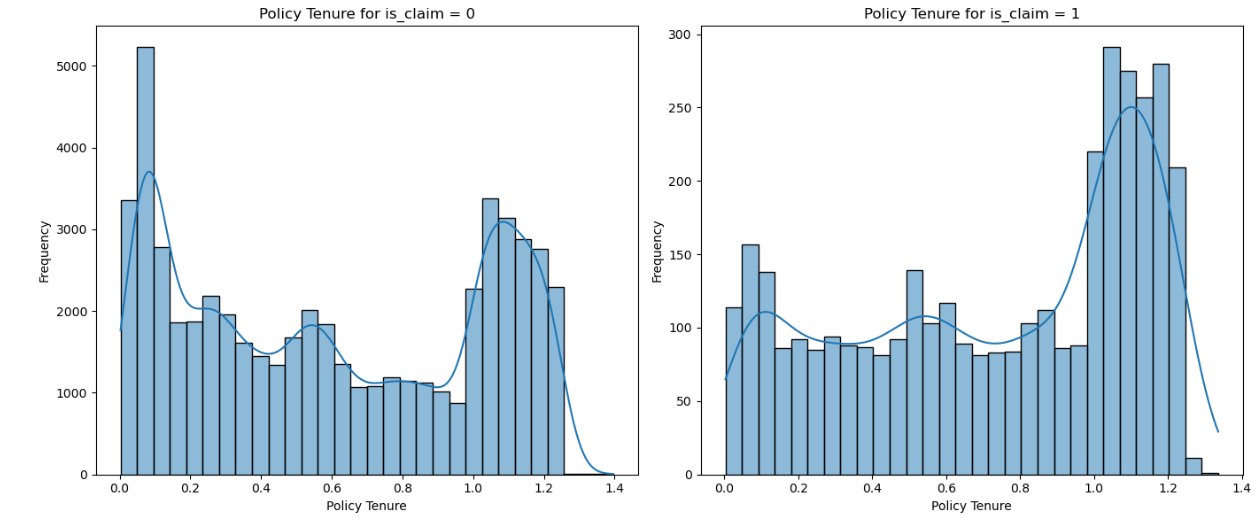


1. **Individuals in their late 20s to late 50s or early 60s as we can see in the histogram are purchasing insurance in this class because they represent the working population**, possessing both the purchasing power **and the need for a vehicle.**



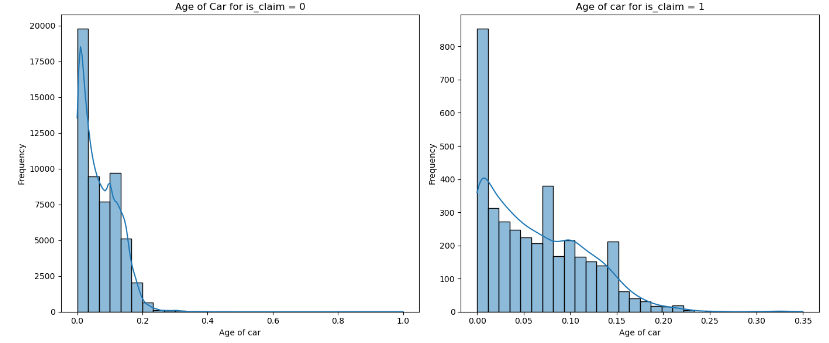
1. **In C8, there are most number of people purchasing insurance policy**, it has a comparatively **lower population density** as seen in the graph below, but features the **most number of insurance policies**.
2. **In C2 to C3 the insurance policies are moderately popular there is a gradual decrease in popularity in the number of insurance polici.es**





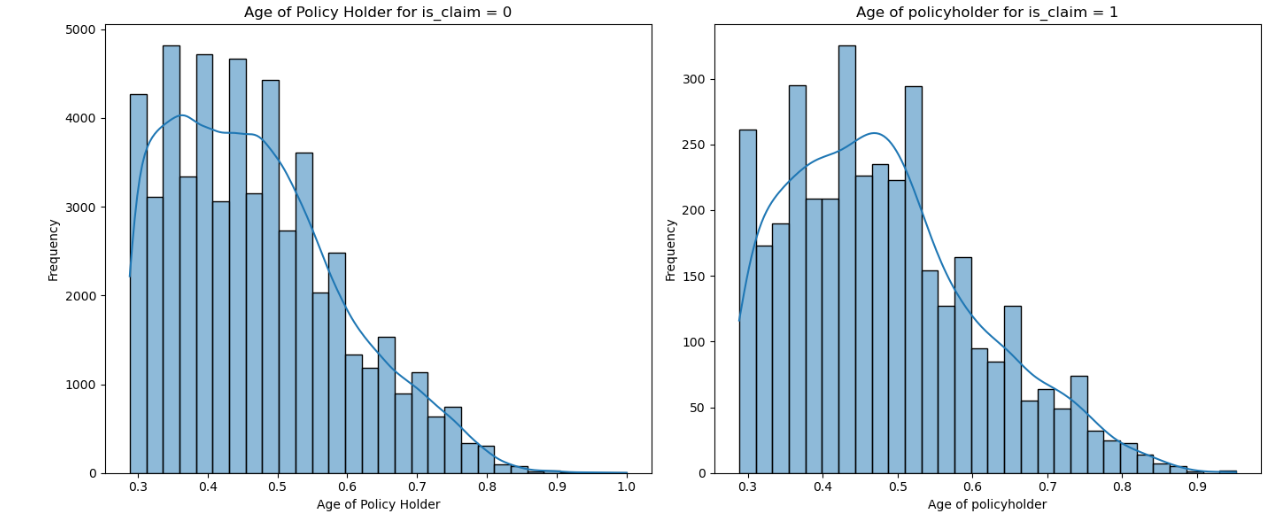
**Insights from Policy Tenure Distribution by Claim Occurrence**

1. The distribution of **policy tenure for is\_claim = 0 (left plot) and is\_claim = 1 (right plot)** shows distinct patterns. This indicates that the tenure distribution is different for those who filed claims compared to those who did not.
2. **For is\_claim = 0,** there is a **significant peak at the beginning of the policy tenure range (around 0.0).** This suggests **that a large number of policyholders do not file claims within a short duration of their policy**. The frequency decreases as the policy tenure increases.
3. **For is\_claim = 1, the distribution shows a noticeable increase in the frequency of claims for policy tenures around 1.0 to 1.2.** This suggests that policyholders are more **likely to file claims after having the policy for a longer duration.**
4. **Both plots show lower frequencies for mid-range policy tenures (around 0.4 to 0.8),** which could indicate **a transitional period where fewer claims occur**.



**Insights from Age of Car Distribution by Claim Occurrence**

1. **For is\_claim = 0 (left plot), the distribution shows a significant peak at the beginning of the car age range (around 0.0).** This indicates that **newer cars have a higher frequency of not filing claims.** The frequency decreases rapidly as the car age increases.
2. **For is\_claim = 1 (right plot), the distribution also shows a high frequency for newer cars (around 0.0), but the decline in frequency is less steep compared to is\_claim = 0.** This suggests that **slightly older cars (up to 0.1 normalized years) have a relatively higher frequency of filing claims.**
3. **Both plots exhibit a high frequency for very new cars (close to 0.0), indicating that very new cars have a significant proportion of both claim and non-claim instances. However, the non-claim frequency is much higher.**
4. **The distribution for is\_claim = 1 shows a longer tail, indicating that claims are spread out more evenly across a wider range of car ages,** whereas the **non-claim distribution is more concentrated towards newer car**s.
5. **The is\_claim = 1 plot shows multiple peaks**, with secondary peaks around 0.05, 0.1, and 0.15. **This could suggest specific age intervals where cars are more prone to filing claims.**
6. The comparison of both plots suggests that newer cars have a higher likelihood of not filing claims, but there is a significant proportion of claims in slightly older cars. This could imply that car age is a relevant factor in predicting claim occurrence.



**Insights from Age of Policy Holder by Claim Occurrence**

1. Non-claim Policyholders (Left Plot):

* The distribution is right-skewed.
* Most policyholders who have not filed a claim fall within the 0.3 to 0.6 range of normalized age.
* The frequency decreases significantly after 0.6.

1. Claim Policyholders (Right Plot):

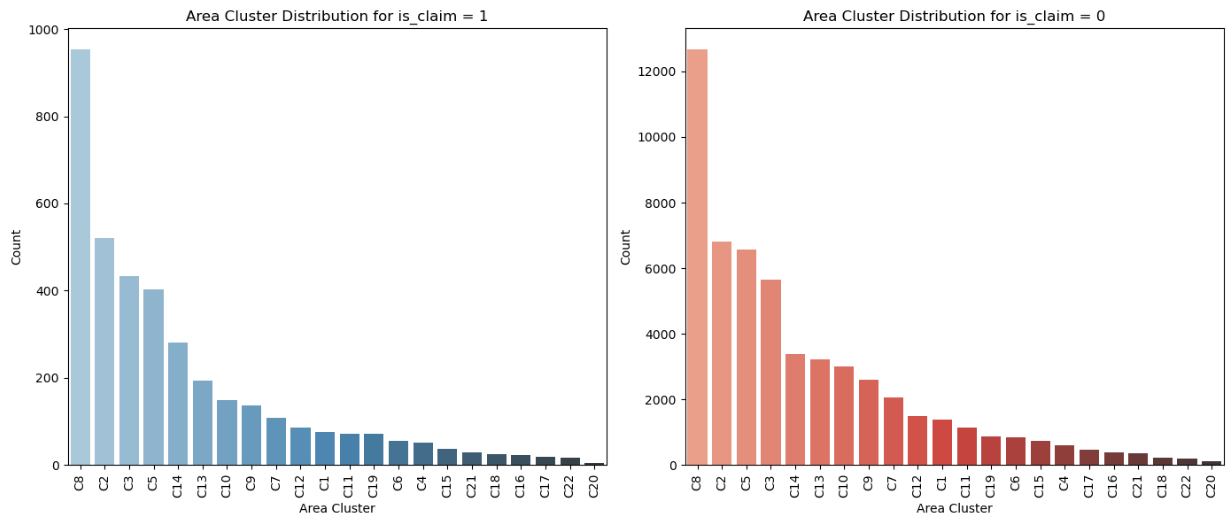
* The distribution is also right-skewed but has fewer overall counts compared to non-claim policyholders.
* The majority of policyholders who filed a claim are within the 0.3 to 0.5 range of normalized age.
* There is a noticeable drop in frequency after 0.6.

Insights:

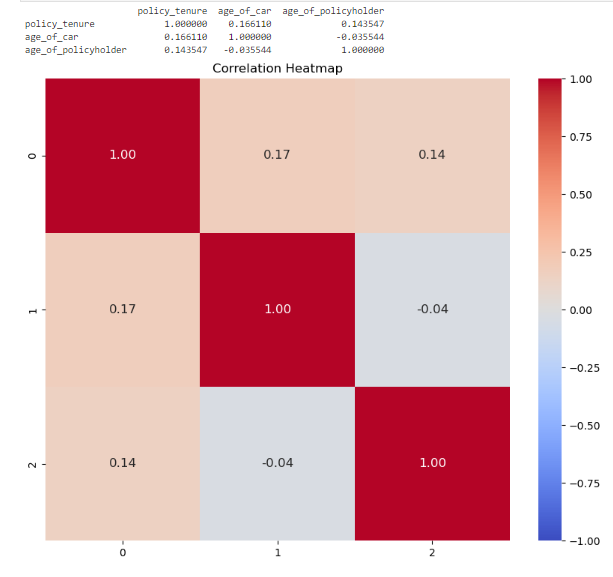
**Age of Policyholder: Younger policyholders (normalized age 0.3 to 0.5) are more common among both claim and non-claim groups. There is a slightly higher proportion of policyholders in the younger age brackets among those who have filed a claim compared to those who haven't.**

Potential Risk Factor:

Younger policyholders might have a higher risk of filing a claim.



1. **In C8, there are most number of people purchasing insurance policy**, and since it has the most number of people it denotes that C8 has the highest percentage of people claiming and not claiming the insurance and the values follow a the same general trend after that.
2. **There is still a significant decrease** in the number of people claiming the insurance and people not claiming the insurance.



**The correlation Matrix denotes a weak correlation between the numerical attributes** hence we can concur **that numerical attributes are independent of each other**.