MACHINE LEARNING EXAM 1

Tursunov Akhmadjon

**1 Model Reporting**

* Ushbu bo'lim loyihaning qisqacha tavsifi, maqsadi va asosiy natijalarini taqdim etadi.
* Maqsad:
* Ushbu loyiha maqsadi sug’urta mijozlaridan olingan ma’lumotlarga asoslanib, asoslanib Polisy Status ni aniqlashdir.

**Ma'lumotlar Haqida Umumiy Ma'lumot**

Ma'lumotlar to'plami 1 200 000 tani tashkil etib, 20 ta xususiyat(features)dan iborat. Target feature “Policy Type” boʻlib, turli siyosat toifalarini ifodalaydi.

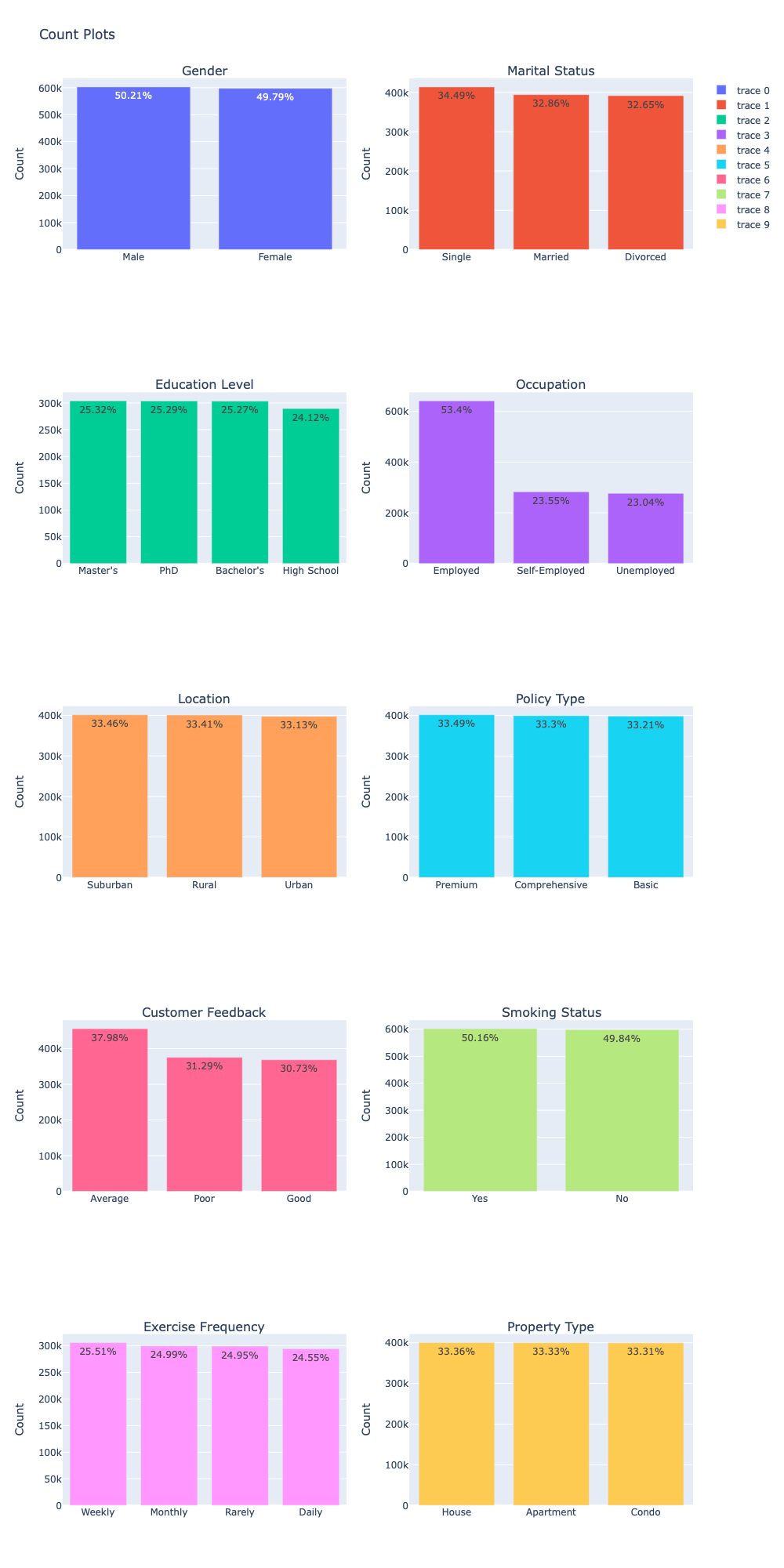
* Demografik xususiyatlar: Yoshi, yillik daromadi, qaramog'idagilar soni, jinsi, oilaviy ahvoli, ta'lim darajasi va kasbi.
* Salomatlik va turmush tarzi xususiyatlari: Salomatlik reytingi, chekish holati, mashqlar chastotasi va mijozlarning fikr-mulohazalari.
* Siyosat va daʼvolar xususiyatlari: Oldingi da'volar, sug'urta muddati, mukofot miqdori va siyosatning boshlanish sanasi (yil, oy, kun).
* Mulk va joylashuv xususiyatlari: Avtomobil yoshi, mulk turi va joylashuvi turi (qishloq, shahar atrofi, shahar).

**Data Preprocessing jarayoni**

* Feature Engineering:

Data setimizdagi yetishmaydigan qiymatlarni to’ldirib oldik. So’ng Dataset dagi malummot turlarini to’g’irlab chiqdik. Policy start date -> object ko’rinishidan, policy start yil, oy, kun ko’rinishidagi 3 ta featurega bo’lib numerical qilib oldik. Age, 'Number of Dependents, Previous Claims, Vehicle Age, Insurance Duration featurelarni “int” ko’rinishiga o’tkazib.

Feature larimizning bir biri o’rtasidagi munosabatlarni organdik.



* Coding:

Target featuremizini label encoding bilan, qolgan feature larni get dummies bilan kodlab oldik. Bu data setimizni toliqligicha numerical o’rinishga o’tkazishga yordam beradi.

* Maʼlumotlarni boʻlish:

Dataset imizni 80% ini traingn uchun qolgan 20% izini test uchun bo’lib oldik

**Modelni Rivojlantirish**

* Tayyor data set imizni bir nechta modellar ga o’qitib, ularni o’zaro solishtiramiz.

***Decision Tree Classifier natijasi***

**ROC AUC Score: 0.49975036226696945**

**precision recall f1-score support**

**0 0.33 0.33 0.33 79412**

**1 0.34 0.34 0.34 80253**

**2 0.33 0.33 0.33 80335**

**accuracy 0.33 240000**

**macro avg 0.33 0.33 0.33 240000**

**weighted avg 0.33 0.33 0.33 240000**

**Ada Boost Classifier natijasi**

**ROC AUC Score: 0.5015809231383136**

**precision recall f1-score support**

**0 0.33 0.30 0.32 79412**

**1 0.33 0.37 0.35 80253**

**2 0.34 0.33 0.33 80335**

**accuracy 0.34 240000**

**macro avg 0.34 0.33 0.33 240000**

**weighted avg 0.34 0.34 0.33 240000**

**GaussianNB natijasi**

**ROC AUC Score: 0.5027014908033233**

**precision recall f1-score support**

**0 0.34 0.36 0.35 79412**

**1 0.33 0.27 0.30 80253**

**2 0.34 0.37 0.35 80335**

**accuracy 0.34 240000**

**macro avg 0.34 0.34 0.33 240000**

**weighted avg 0.34 0.34 0.33 240000**

**Logistic Regression natijasi**

**ROC AUC Score: 0.5021117381127073**

**precision recall f1-score support**

**0 0.33 0.27 0.30 79412**

**1 0.33 0.20 0.25 80253**

**2 0.34 0.53 0.41 80335**

**accuracy 0.33 240000**

**macro avg 0.33 0.33 0.32 240000**

**weighted avg 0.33 0.33 0.32 240000**

**Linear Discriminant Analysis natijasi**

**ROC AUC Score: 0.5021207506780441**

**precision recall f1-score support**

**0 0.33 0.29 0.31 79412**

**1 0.33 0.23 0.27 80253**

**2 0.33 0.49 0.40 80335**

**accuracy 0.33 240000**

**macro avg 0.33 0.33 0.33 240000**

**weighted avg 0.33 0.33 0.33 240000**

**MLP Classifier natijasi**

**ROC AUC Score: 0.4998684345390849**

**precision recall f1-score support**

**0 0.33 0.23 0.27 79412**

**1 0.33 0.21 0.26 80253**

**2 0.34 0.56 0.42 80335**

**accuracy 0.33 240000**

**macro avg 0.33 0.33 0.32 240000**

**weighted avg 0.33 0.33 0.32 240000**

**XGB Classifier natijasi**

**ROC AUC Score: 0.5035922819286331**

**precision recall f1-score support**

**0 0.33 0.33 0.33 79412**

**1 0.34 0.31 0.32 80253**

**2 0.34 0.37 0.35 80335**

**accuracy 0.34 240000**

**macro avg 0.34 0.34 0.34 240000**

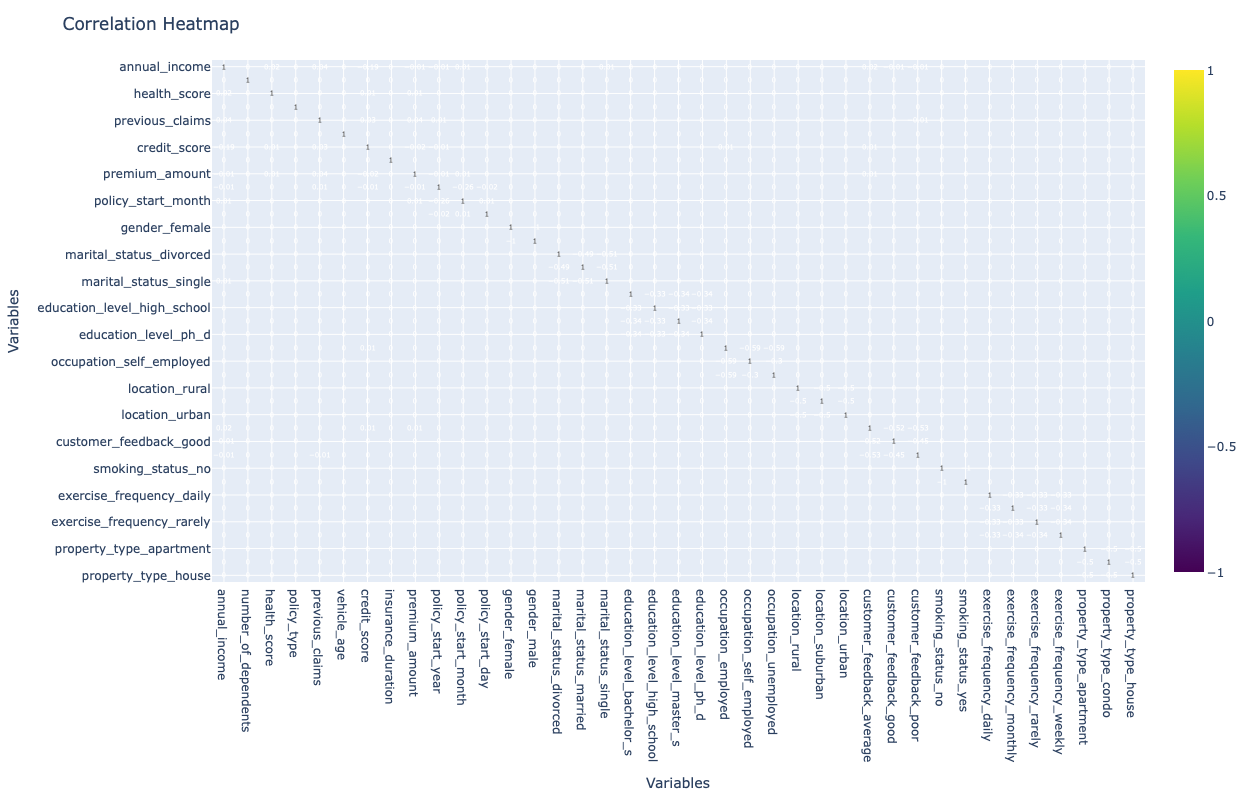
**weighted avg 0.34 0.34 0.34 240000**

Modelni Baholash

* Accuracy Score: [Qiymat].

Xulosa

* Natijamiz juda yomon xatolik juda katta, model aniqligi juda past. Bunga sabab dataset dagi feature lardir.



**Correlation matrix dan ham ko’rinib turibdiki feature larimiz orasidagi munosabatlar 0 ga yaqin. Buni to’g’irlash uchun keyinchalik to’g’ri featurelar yaratishimiz kerak.**