Dec520-101C-Team11-Project Final Report

Names(NetIds): Eisenberg Xie(jx113), Jenny Park(yp114), Muhammad Umer Khan(mk544), Shiven Ahuja(ska21), Ye Yang(yy328), Yijie Wang(yw484), Yuge Jin(yj162)

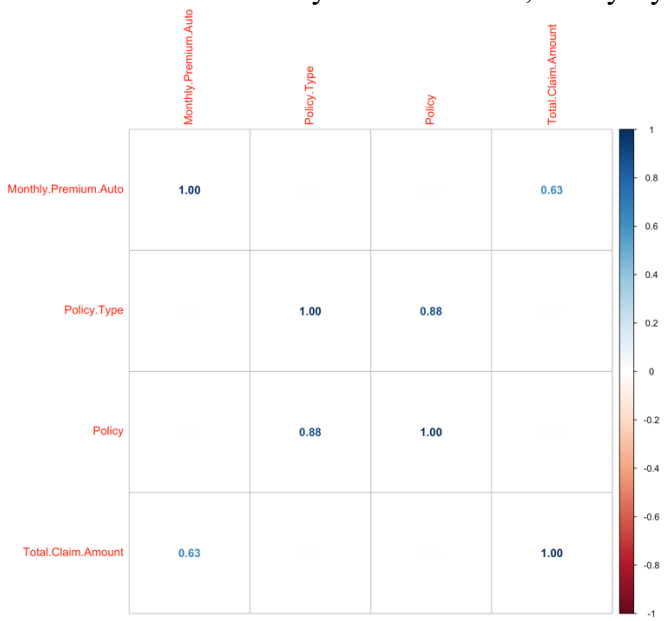
Appendix

1. Variable Dictionary

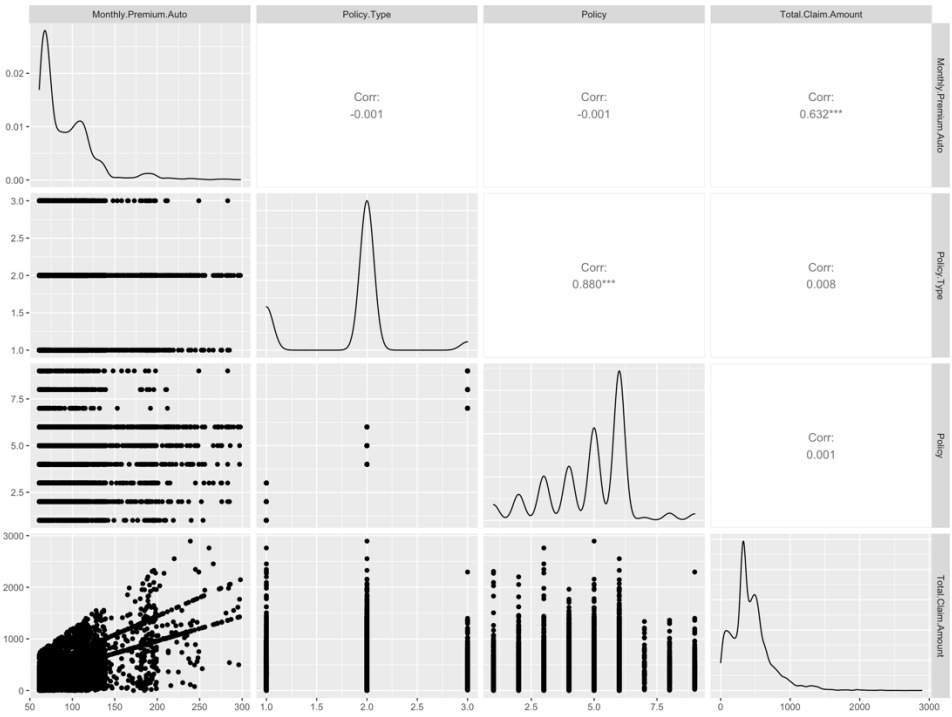
Column Name	Description	Type
Customer	Unique Customer ID Identifier	Categorical
State	State of Residence in US (Arizona/California/Nevada/Oregon/Washington)	Categorical
Customer Lifetime Value	Value expected from customer, measuring a customer's profitability over the long term	Numerical
Response	Whether accept the cross-selling policy offer or not (Yes/No)	Boolean
Coverage	Type of coverage (Basic/Extended/Premium)	Categorical
Education	Education Level (High School or Below/College/Bachelor/Master/Doctor)	Categorical
Effective To Date	Policy Last Date	Date
EmploymentStatus	Employment Status (Disabled/Employed/Medical Leave/Retired/Unemployed)	Categorical
Gender	Gender (F/M)	Categorical
Income	Income	Numerical
Location Code	City Tier (Rural/Suburban/Urban)	Categorical
Marital Status	Marital Status (Single/Married/Divorced)	Categorical
Monthly Premium Auto	Monthly Premium Auto	Numerical
Months Since Last Claim	Months Since Last Claim	Numerical
Months Since Policy Inception	Months Since Policy Inception	Numerical
Number of Open Complaints	Number of Open Complaints	Numerical
Number of Policies	Number of Policies	Numerical
Policy Type	Policy Type (Corporate Auto/Personal Auto/Special Auto)	Categorical

Policy	Type of policy (Corporate L1/Corporate L2/Corporate L3/Personal L1/Personal L2/Personal L3/Special L1/Special L2/Special L3)	Categorical
Renew Offer Type	Type of offer used (Offer1/Offer2/Offer3/Offer4)	Categorical
Sales Channel	Sales Channel through which lead was acquired (Agent/Branch/Call Center/Web)	Categorical
Total Claim Amount	Total Claim Amount	Numerical
Vehicle Class	Vehicle Class (Four-Door Car/Luxury Car/Luxury SUV/Sports Car/SUV/Two-Door Car)	Categorical
Vehicle Size	Vehicle Size (Large/Medsize/Small)	Categorical

2. Correlation of Monthly Premium Auto, Policy Type, Policy, and Total Claim Amount



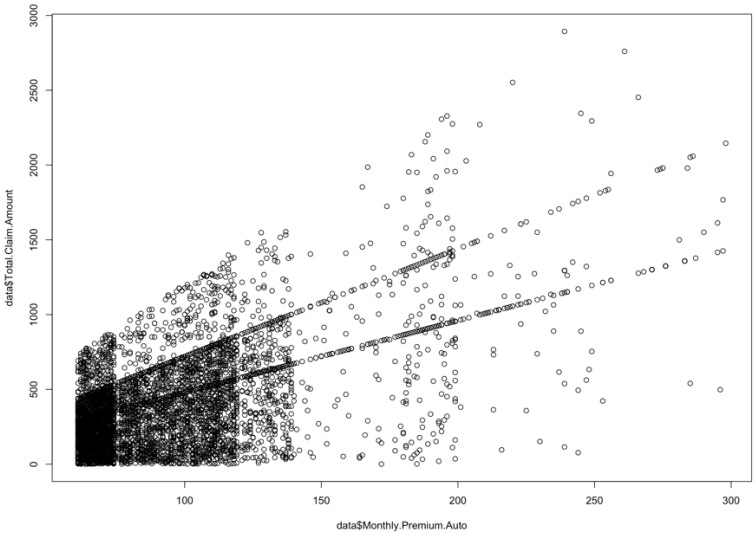
3. GG-Pairs plot of Monthly Premium Auto, Policy Type, Policy, and Total Claim Amount



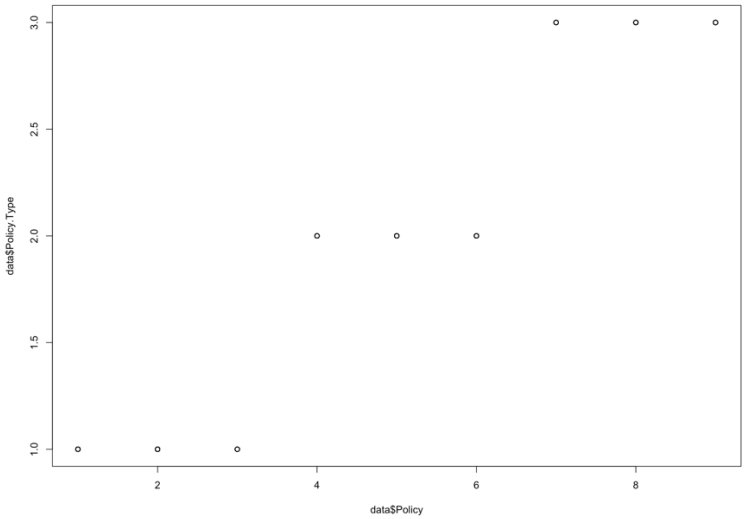
4. VIF Score Table

Customer.Lifetime.Value	Response	Coverage
1.197765	1.050747	1.294022
Education	EmploymentStatus	Gender
1.008516	2.275622	1.016488
Income	Location.Code	Marital.Status
2.262277	1.152958	1.148016
Monthly.Premium.Auto	Months.Since.Last.Claim	Months.Since.Policy.Inception
2.579536	1.005424	1.013921
Number.of.Open.Complaints	Number.of.Policies	Policy.Type
1.004493	1.011093	4.438234
Policy	Renew.Offer.Type	Sales.Channel
4.436449	1.104205	1.026067
Total.Claim.Amount	Vehicle.Class	Vehicle.Size
2.557651	1.069115	1.016025

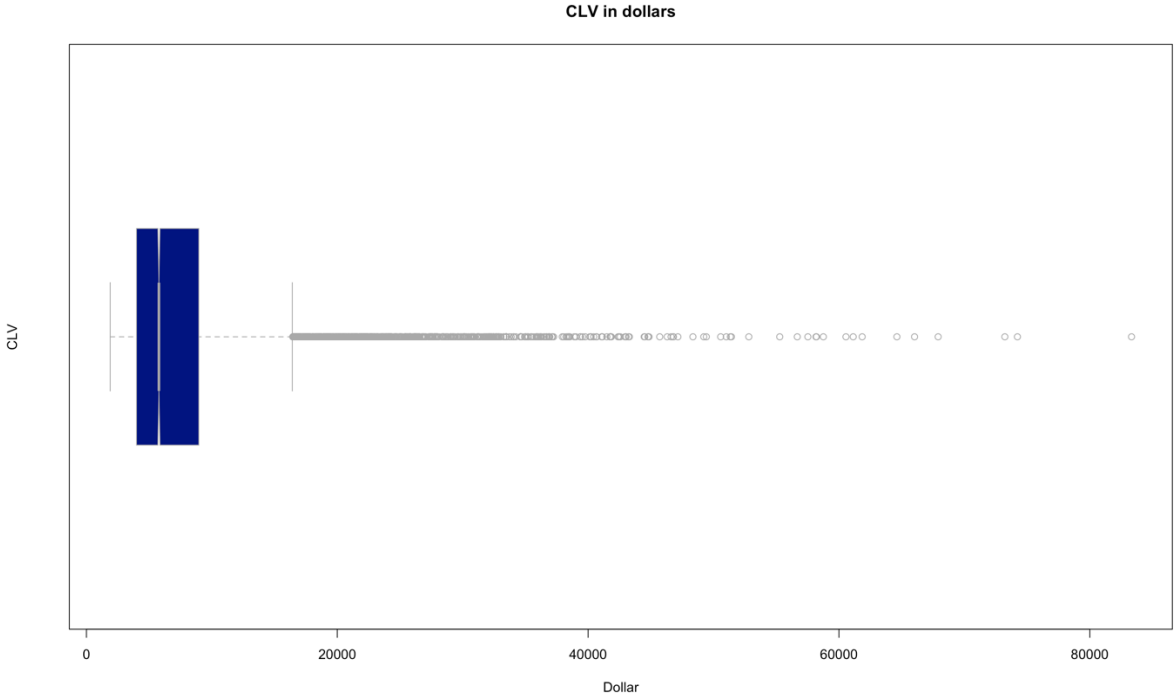
5. Scatter Plot of MPA and TCA



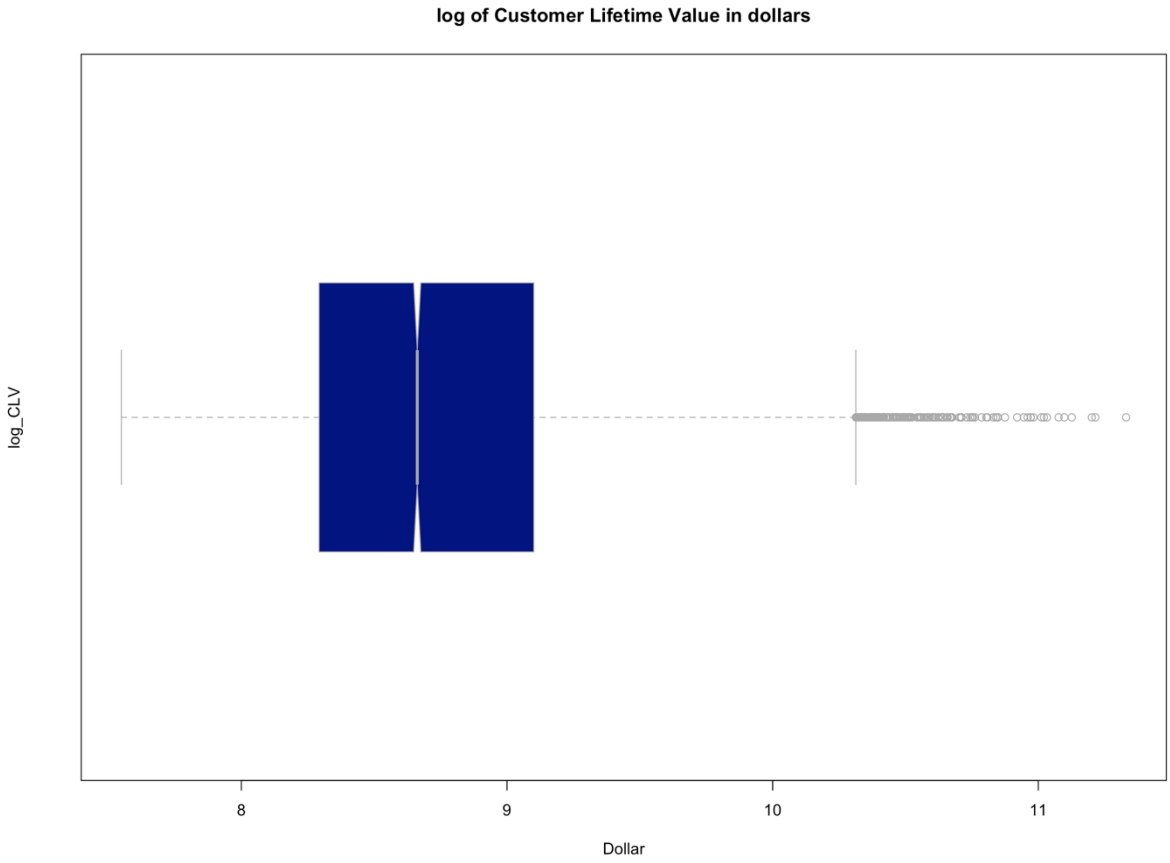
6. Scatter Plot of Policy.Type and Policy



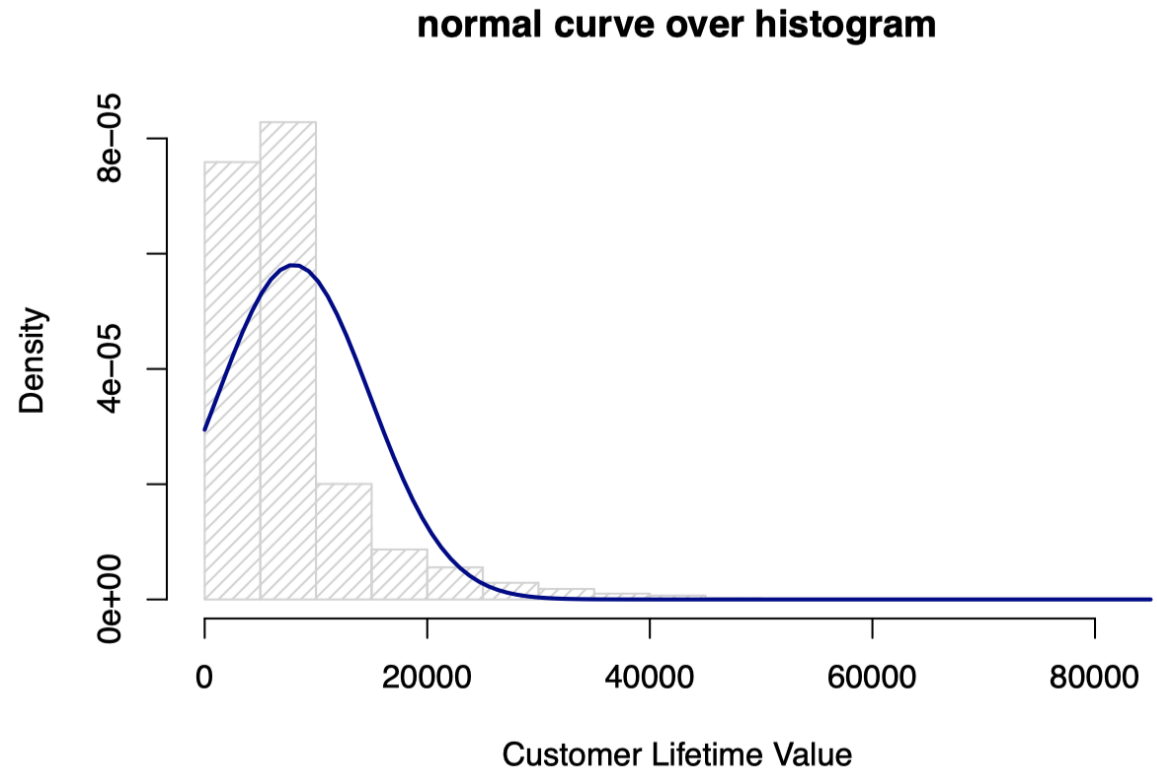
7. Boxplot of Customer Lifetime Value



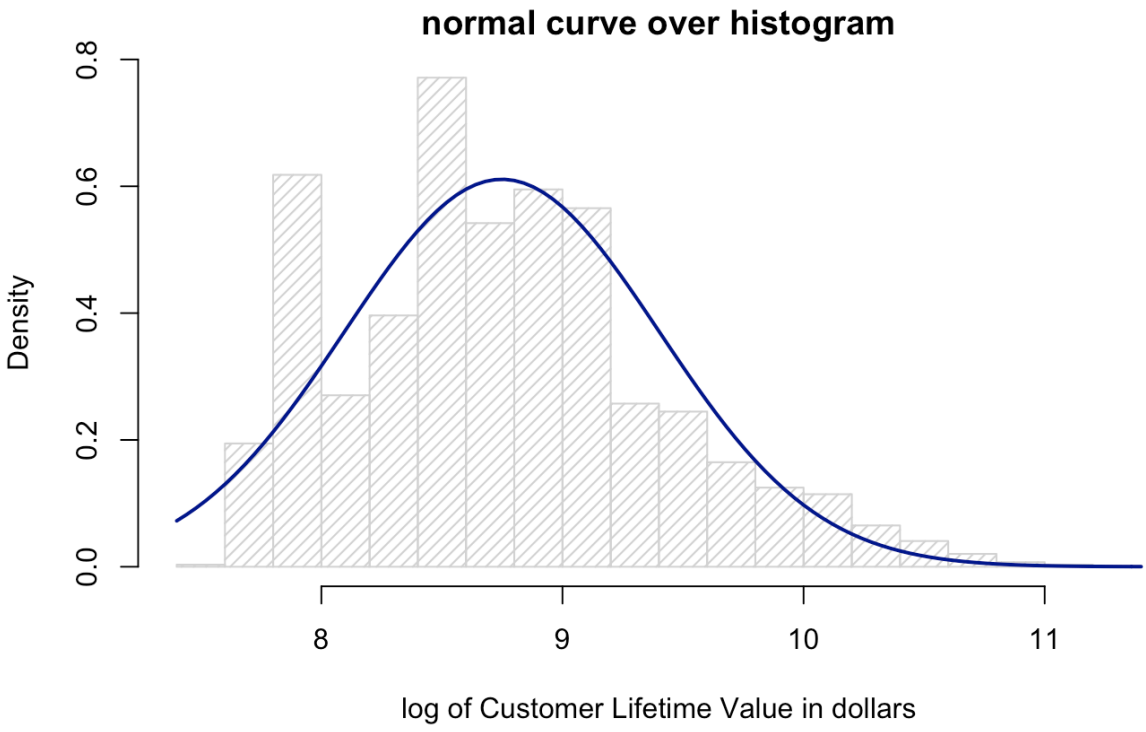
8. Boxplot of Log of Customer Lifetime Value



9. Normal Curve over Histogram of CLV

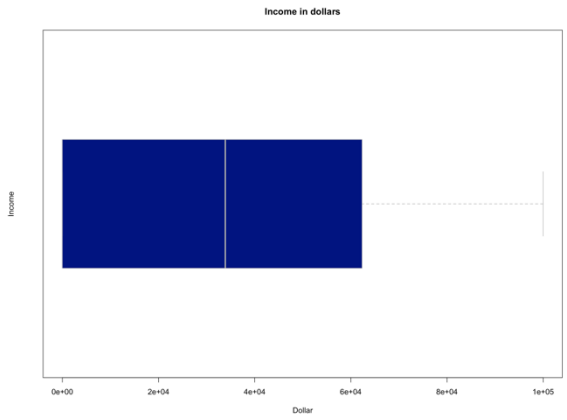


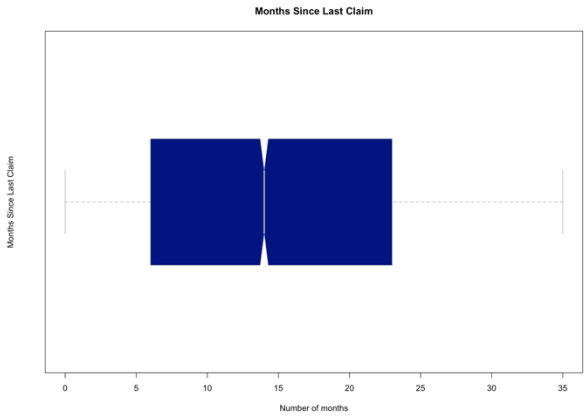
10. Normal Curve over Histogram of Log of Customer Lifetime Value



11. Boxplot of Income

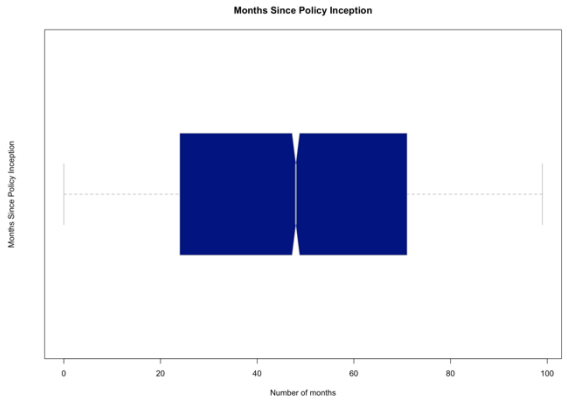
12. Boxplot of Months Since Last Claim

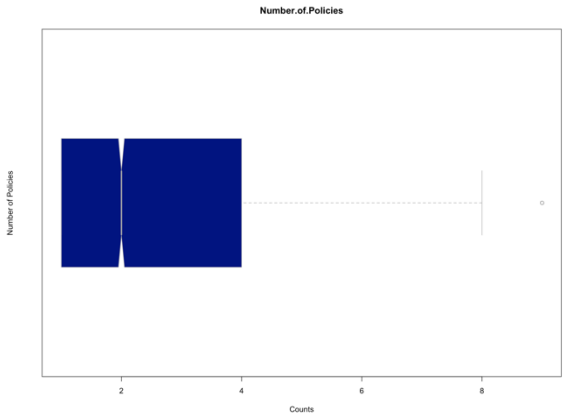




13. Boxplot of Months Since Policy Inception

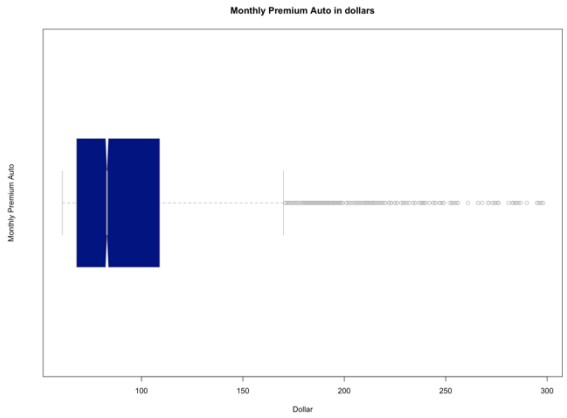
14. Boxplot of Number of Policies

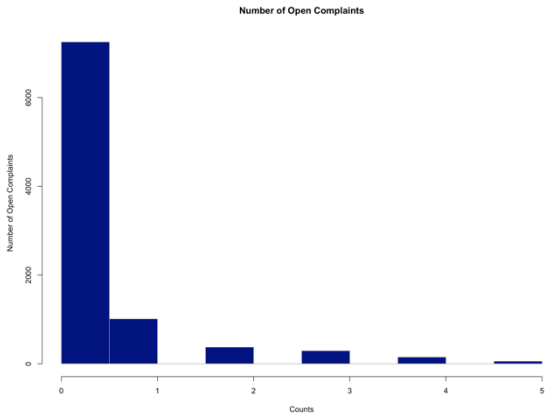




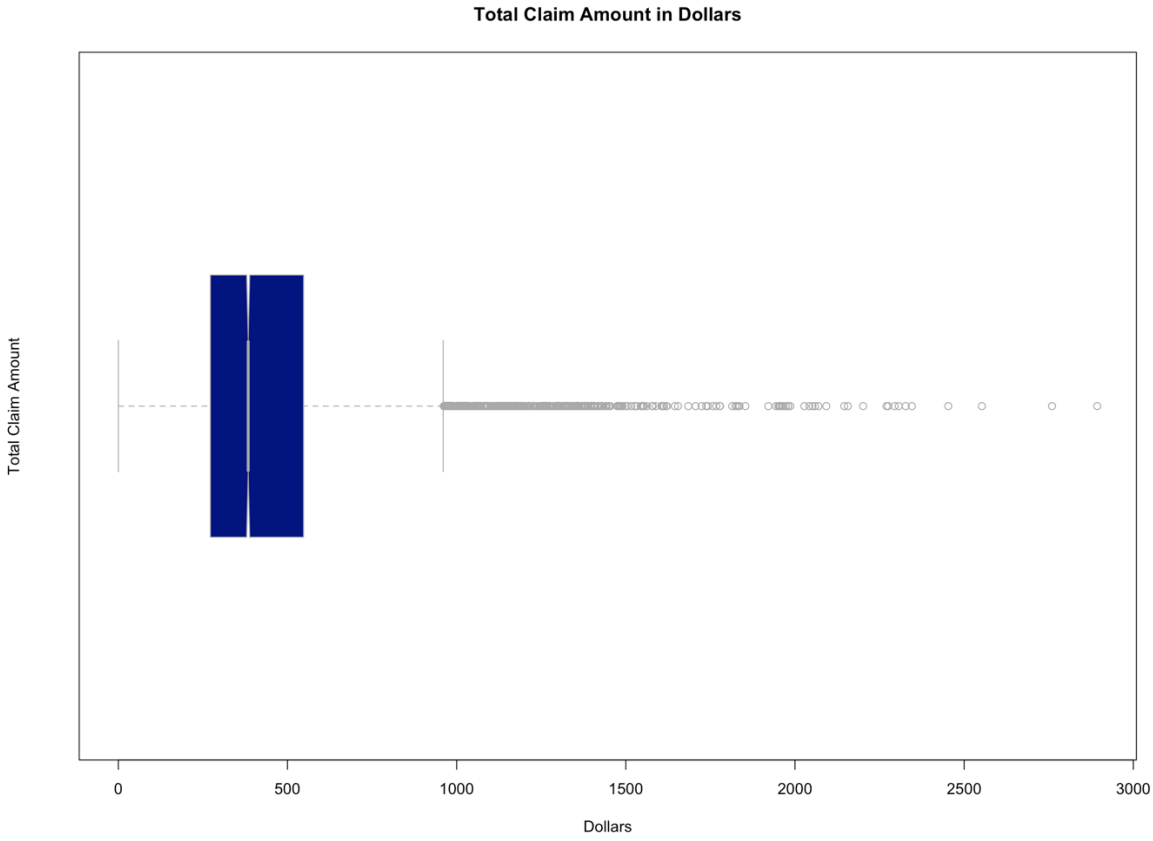
15. Boxplot of Monthly Premium Auto

16. Histogram of Number of Open Claims

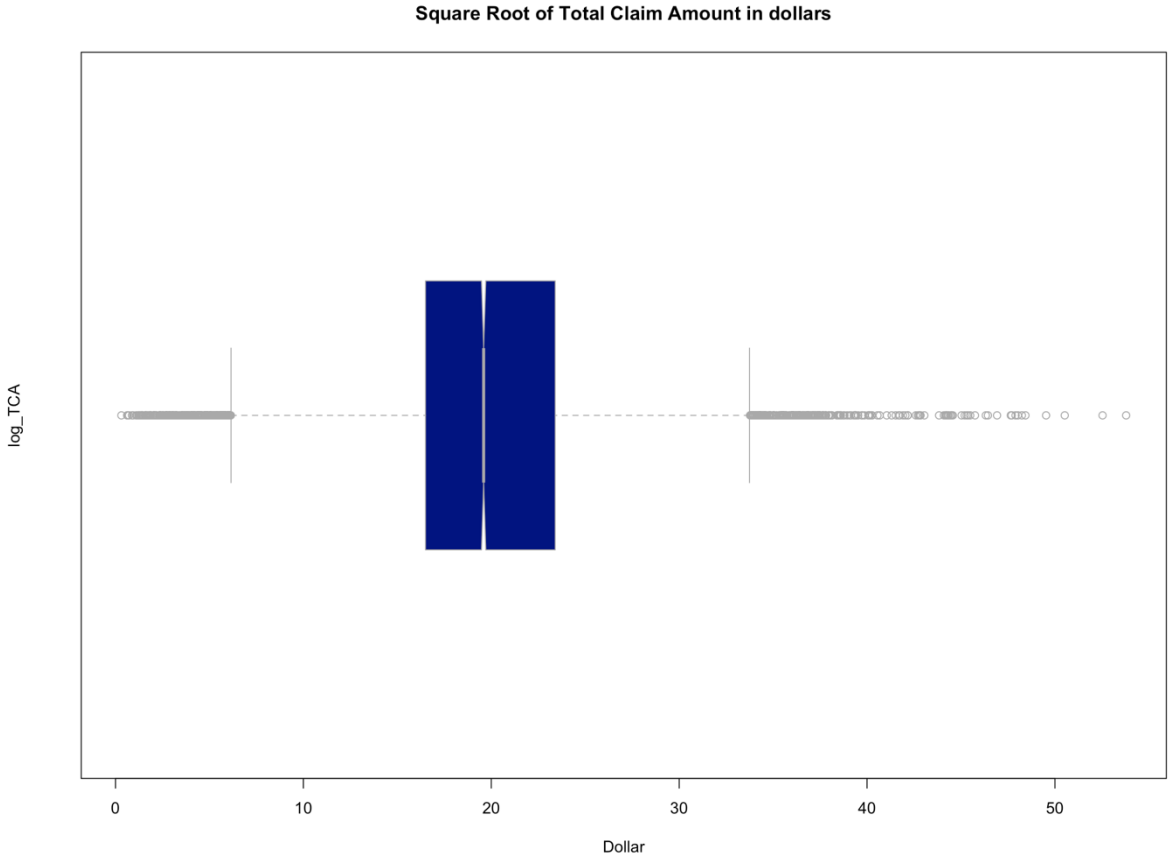




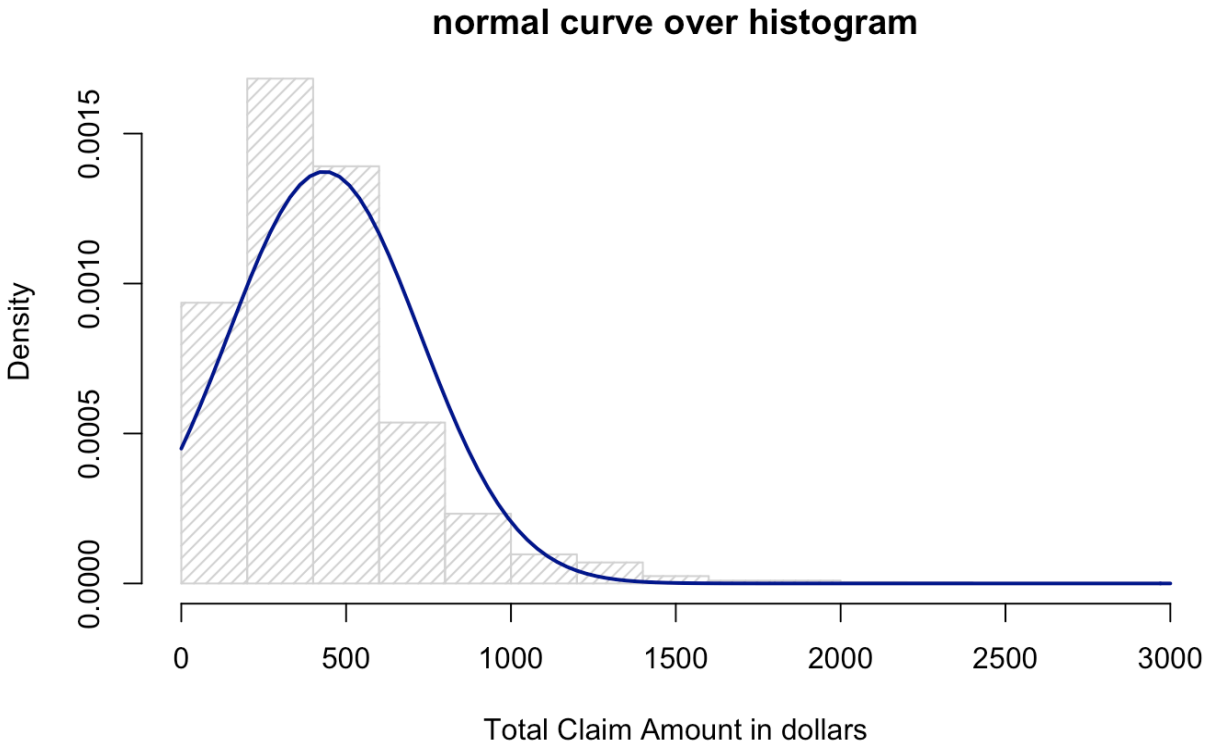
17. Boxplot of Total Claim Amount



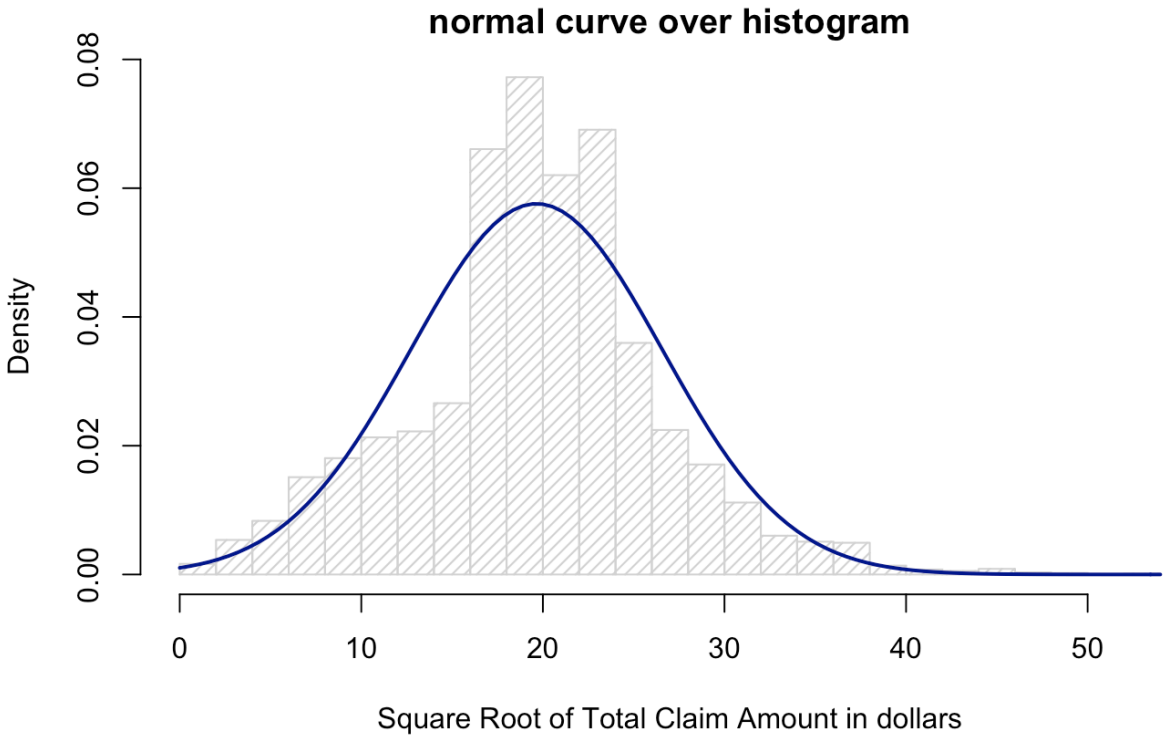
18. Boxplot of Square Root of Total Claim Amount



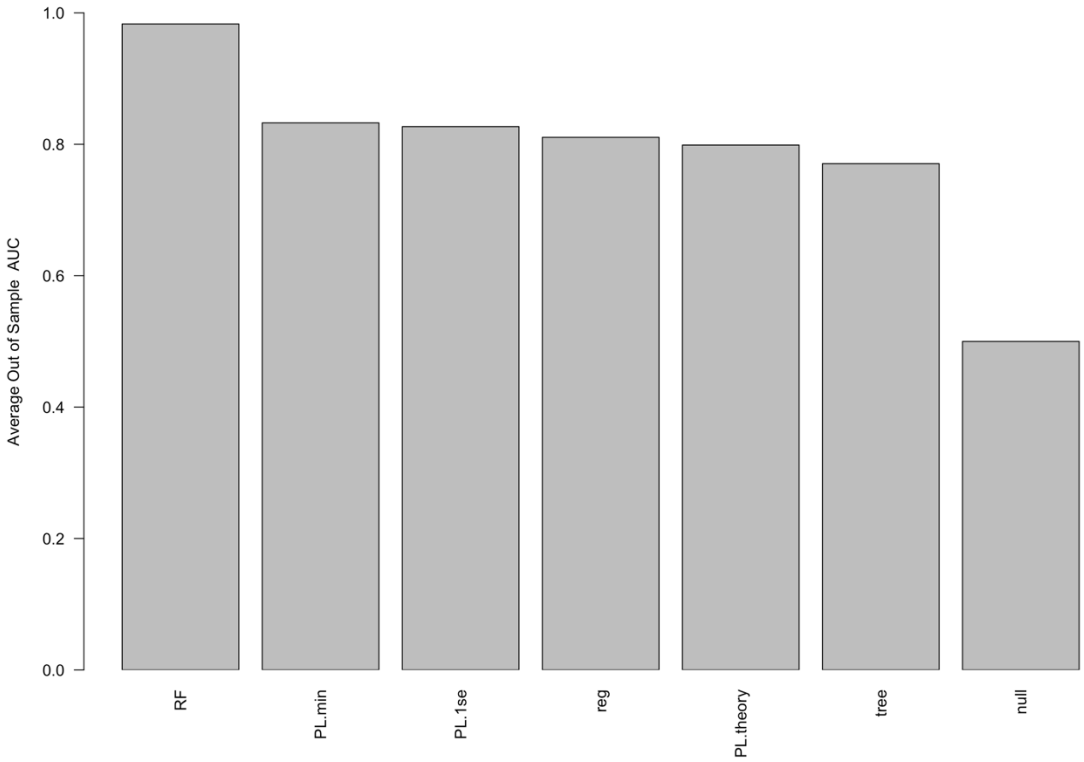
19. Normal Curve over Histogram of Total Claim Amount



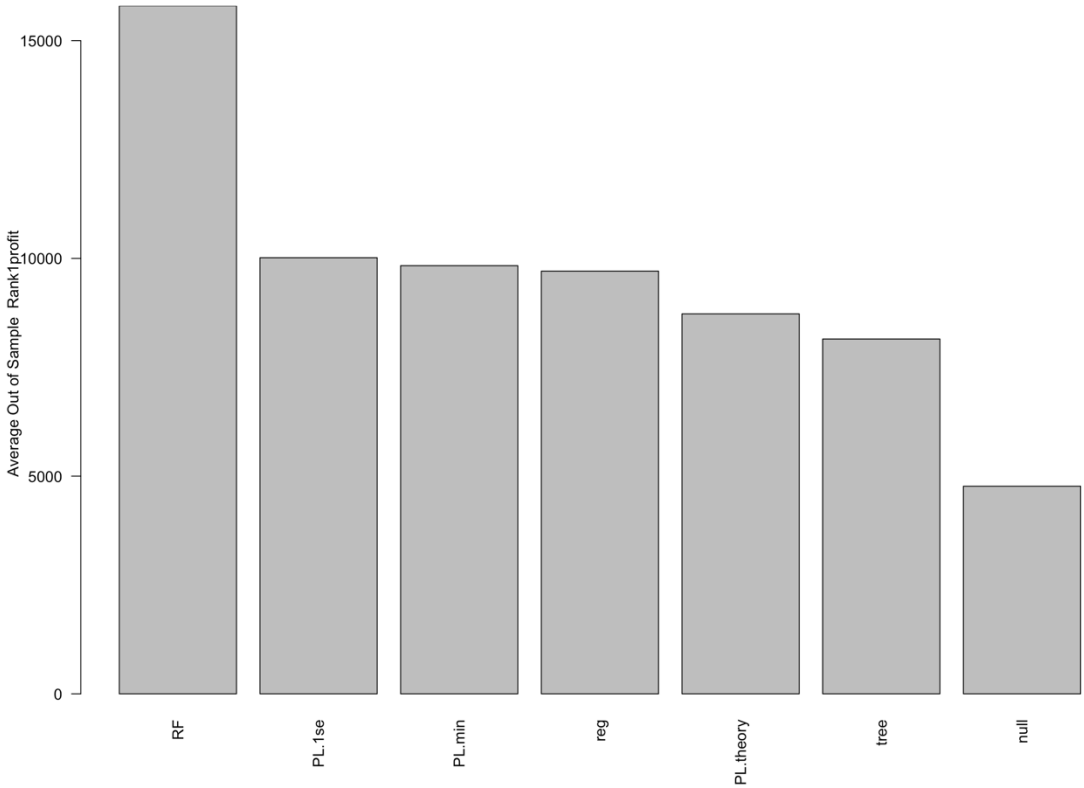
20. Normal Curve over Histogram of Square Root of Total Claim Amount



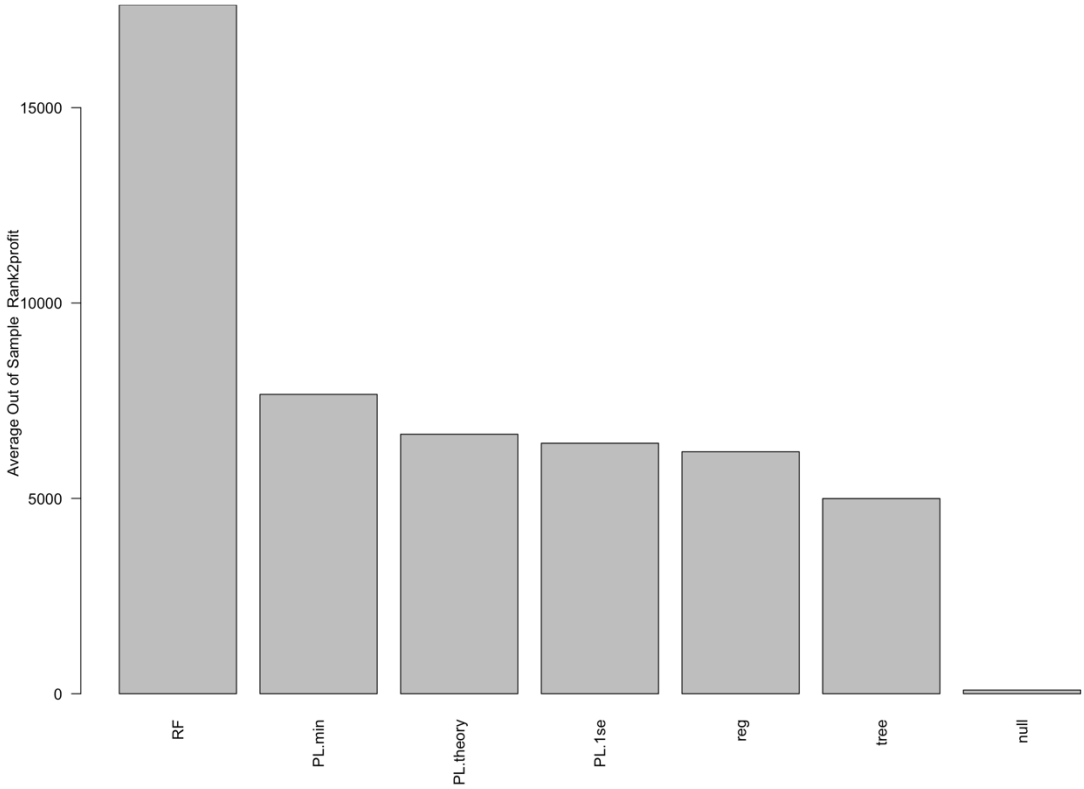
21. Bar plot of Average Out of Sample AUC



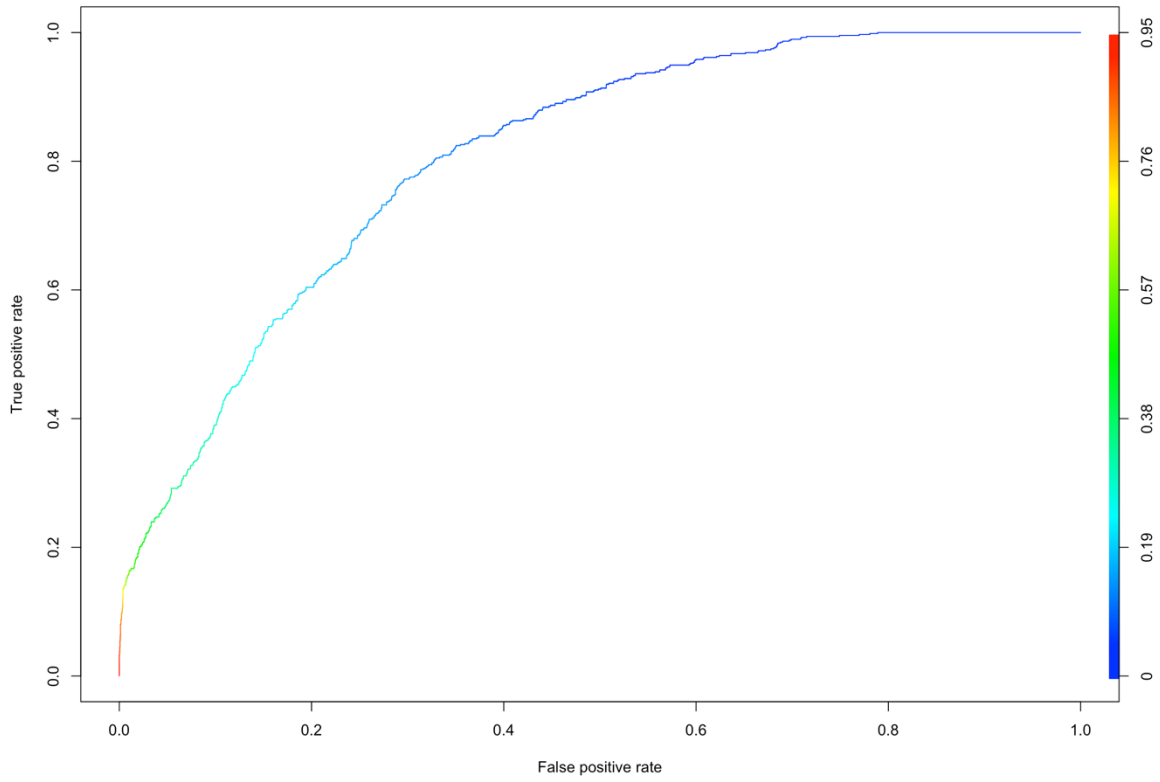
22. Bar plot of Average Out of Sample Rank1profit



23. Bar plot of Average Out of Sample Rank2profit



24. ROC curve of Post Lasso Minimum Choice λ



25. Contribution List:

Umer: Business Understanding, Interpretation of Unsupervised and Deployment of Supervised;

Ye Yang: Data understanding, deployment of the unsupervised model;

Jenny Park: Data Understanding, Supervised model evaluation;

Yuge Jin: Data Understanding, Data Preparation, Modeling, Appendix, Partial Coding;

Eisenberg Xie: Data Understanding, Data Preparation;

Shiven Ahuja: Data Understanding, Unsupervised Modeling, Partial Coding, Documentation;

Yijie Wang: Business assumptions, Supervised model training, Cross-validation and Evaluation, Profit curve plot.