Note: Data is grouped!

1. Model **lm.main** with all predictor variables, Linear Model
2. Initial checks of lm.main:
    a. Adj. R^2 = 0.3582; Not the best
    b. Residuals v. Fitted Plot; Not linear, fanning
    c. QQPlot & Shapiro Wilk Test; Not normal, Failed with p-value = 2.013e-10

3. Model **glm.p.main** with all predictor variables, Poisson with log link
4. Initial checks of glm.p.main:
    a. Goodness-of-Fit; Failed with p-value = 4.278675e-34
    b. Check Overdispersion; check_overdispersion → Overdispersion detected

5. plot(fitdist(Store$customers, ‘pois’)); Shows evidence of more lower values than expected → Try different distribution
    a. Only 1 zero count in response so not zero-inflated
6. plot(fitdist(Store$customers, ‘nbinom’)); Looks okay

7. Model **glm.nb.model** with all predictor variables; Negative Binomial with log link
8. Initial checks of glm.nb.main:
    a. Goodness-of-Fit; Passed with p-value = 0.2737936
    b. Check Outliers; abs(studentized residuals) > 3 → None
    c. Check Leverage; hatvalue > (3 * 5) / 110 → Obs. 15, 94
    d. Check Influence; Cook's Distance > F(.5, 5, 105) → None
    e. Check VIF; all predictors okay (< 4)
    f. Check Overdispersion; check_overdispersion → No overdispersion detected

9. Confirm NB model fits better than Poisson:
    a. LR test statistic = -2*(logLik(glm.p.main) - logLik(glm.nb.main)) = 148.4815
    b. 0.5 * chi-square (df=1) → 1.861465e-34 → Strong evidence that the negative binomial model fits the data better than the Poisson model

10. Used stepwise regression (both directions) on all predictors and 2-way interactions
    a. **reduced.step** model: units, income, compdist, storedist, compdist:storedist
11. Bonferroni-adjusted p-value from LR test for each predictor at 5% significance (.01)
    a. compdist:storedist not significant (p-value = 0.0879754)
12. LRT between reduced.step and glm.nb.model; Significant with p-value = 0 → keep reduced.step

13. Model **reduced.manual.m1** with: units, income, compdist, storedist
14. Bonferroni-adjusted p-value from LR test for each predictor at 5% significance (.0125)
    a. all predictors significant
15. LRT between reduced.manual.m1 and reduced.step; Not significant with p-value = 0.08905502 → keep reduced.manual.m1
16. Compare AIC values; reduced.step smallest, but reduced.manual.m1 almost as small and simpler model
    a. glm.nb.main = 693.2414
    b. reduced.step = 691.0015
    c. reduced.manual.m1 = 691.8929

17. Final model candidates
    a. **final.m1** = reduced.step = units, income, compdist, storedist, compdist:storedist
    b. **final.m2** = reduced.manual.m1 = units, income, compdist, storedist

18. Model requirement checks for final.m1:
    a. Goodness-of-Fit; Passed with p-value = 0.2899502
    b. Check Outliers; abs(studentized residuals) > 3 → None
    c. Check Leverage; hatvalue > (3 * 5) / 110 → Obs. 3, 15, 43, 94
    d. Check Influence; Cook's Distance > F(.5, 5, 105) → None
    e. Check VIF; compdist (10.44), storedist (10.07) not okay
    f. Check Overdispersion; check_overdispersion → No overdispersion detected
19. Scale compdist and storedist to create model **final.m1.2**
20. Model requirement checks for final.m1.2:
    a. Goodness-of-Fit; Passed with p-value = 0.2899502
    b. Check Outliers; abs(studentized residuals) > 3 → None
    c. Check Leverage; hatvalue > (3 * 5) / 110 → Obs. 3, 15, 43, 94
    d. Check Influence; Cook's Distance > F(.5, 5, 105) → None
    e. Check VIF; all predictors okay (< 4)
    f. Check Overdispersion; check_overdispersion → No overdispersion detected
21. Model requirement checks for final.m2:
    a. Goodness-of-Fit; Passed with p-value = 0.2906951
    b. Check Outliers; abs(studentized residuals) > 3 → None
    c. Check Leverage; hatvalue > (3 * 4) / 110 → Obs. 15, 30, 94
    d. Check Influence; Cook's Distance > F(.5, 4, 106) → None
    e. Check VIF; all predictors okay (< 4)
    f. Check Overdispersion; check_overdispersion → No overdispersion detected

22. Compare final.m1.2 and final.m2:
    a. LRT; Keep simpler model with p-value = 0.08905502
    b. AIC(final.m1.2) = 691.0015
    c. AIC(final.m2) = 691.8929
    d. AIC for final.m1.2 slightly smaller, but this model is more complex → final model will be final.m2 = units, income, compdist, storedist
23. LRT between final model with saturated model; keep simpler model (final model) with p-value = 0.8882921
    a. AIC(fill.model) = 708.1234 → larger than AIC of final model (final.m2)

24. Scale all predictors in final model for useful interpretation of intercept
25. Interpret intercept and effects of all significant predictors (all the predictors)
26. Create a table of predictions for all min, mean, max predictor combinations (all are quantitative) → 3^4 = 81 predictions
27. Comment on unresolved problems