Experiment - 3 Report

1. In terms of skewness, how do you explain the histogram obtained in step 2 ?

* Histogram shows right (positive) skewness.
* Most data values are low, with a long tail to the right.
* Mean > Median, indicating skewness.
* Common in count or rate-based data.
* Suggests need for log transformation for normalization.

2. In step 4, which model is more appropriate and why ?

* Poisson regression (GLM) suits count data.
* Linear model assumes normal and continuous data — not valid here.
* Poisson handles non-negative integers and skewed distributions.
* Models mean-variance relationship correctly.
* Hence, GLM (Poisson) is more appropriate.

3. Which model (step 5, step 6, step 7) is best (best in what metric?)?

* Step 5: Poisson → assumes variance = mean.
* Step 6: Quasi-Poisson → allows overdispersion but no AIC.
* Step 7: Negative Binomial → best for overdispersed data.
* Metric: AIC (lower is better) or residual deviance.
* Best: Negative Binomial (handles variance flexibly).

4. Differentiate between linear model (lm) and generalized linear model (glm)

GLM is a generalized form of LM.

| Linear Model   * assumes normal errors * uses identity link * for continuous data * constant variance | Generalized Linear Model   * supports various distributions. * uses log, logit, etc. * for count/binary data. * variance depends on the mean. |
| --- | --- |

5. Why is log transformation beneficial in step 8 ?

* Reduces right skewness in data.
* Stabilizes variance and normalizes residuals.
* Makes relationships more linear.
* Handles large values/outliers better.
* Improves model fit and interpretability.