

Exploring Breed-Specific Rehoming Timelines of Dogs at Animal Shelters

ABSTRACT

People own a companion animal, which is adopted from shelter homes. The animals in the shelter homes are either stray or abandoned and the rehoming time depends on many factors. This research will delve into valuable insights of rehoming durations and provide some important information for understanding the factors influencing them.

INTRODUCTION

Owner looks at different parameters while adopting from the animal shelter, making the parameter of rehoming time an important one; hence, it is very important to delve into this parameter. Researchers suggest that the average rehoming time for dogs is 27 weeks, with a variance of around 74 for all dogs, irrespective of the breed. Whereas other researchers believe that the rehoming time will differ from 27 weeks according to a specific breed, Hence, by conducting a detailed analysis of the timelines for the rehoming times for different breeds, we will have the ability to comment on whether they are specific to a dog breed or the same for all the dogs.

RESULTS

The data uses 82 records of three dog breeds: 'Dobermann', 'Shih Tzu', and 'West Highland White Terrier' ('WHWT') with features: 'Visited', 'Rehomed', 'Health', 'Breed', 'Age', 'Reason', 'Returned'. Six missing values of 'Breed' were dropped as the analysis investigates timelines based on breeds. Nine invalid values of '99999' in the 'Rehomed' column are imputed with the mean of the rehomed times of that breed for which this value is associated. Two missing values are replaced by 'Unknown' in the 'Neglect' column, preventing significant data loss. Six rows, 7.32% data was dropped.

Data Exploration

Puppies have shorter rehoming times than fully grown regardless of the breed (Figure 1).

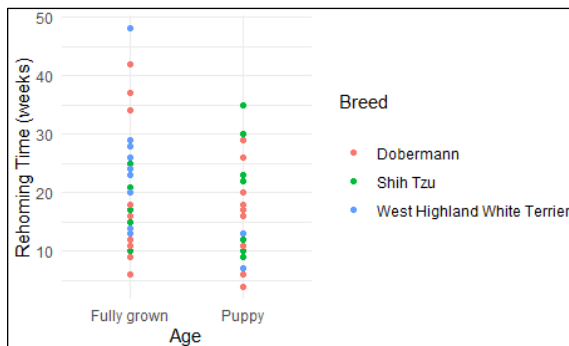


Figure 1: A Scatter plot illustrating distribution of age against the rehoming times in weeks for 82 records of three different breeds listed.

Breed	Age	SD Rehomed (weeks)
Dobermann	Puppy	7.438
	Fully grown	12.710
West Highland White Terrier	Puppy	4.242
	Fully grown	8.911
Shih tzu	Puppy	8.834
	Fully grown	8.468

Table 1: Rehoming Duration Variation Across Dog Breeds and Age Groups

Looking at the standard deviation (Table_1) fully grown Dobermanns have longer rehoming times than the mean, while others have shorter, indicating the data is more spread out. WHWT spends the longest in shelter (Table_2), followed by Shih Tzu and Dobermann. 50% of dogs are rehomed

Breed	Rehomed (weeks)					
	Min	1 st quartile	Median	Mean	3 rd quartile	IQR
Dobermann	4	11	17	17.8695	21	10
West Highland White Terrier	7	13	20	20.1851	24.5	12.25
Shih tzu	9	12.25	18.5	19.5384	24.5	11.50

Table 2: Table presenting key statistical measures of the rehoming durations in weeks indicating central tendencies and variability in the time. This was calculated for 82 entries of which includes all the three breeds.

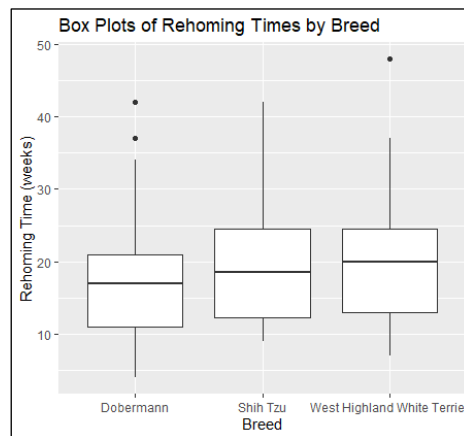


Figure 2: A boxplot illustrating the distribution of rehoming durations in weeks against the breeds.

within 20 weeks, with the Dobermann breed being quickly rehomed than others (Table_2). IQR for Dobermann is the lowest, indicating most data falls within a narrow range (Figure_2) followed by Shih Tzu and WHWT. Outliers are present in Dobermann and WHWT, with Shih Tzu being a right-skewed breed. There is no significant difference between the mean and median of breeds.

Modelling and Estimation

Rehoming data, a continuous variable that, in theory, can range from 0 to ∞ , indicating that dog can be rehomed as soon as it arrives to the animal shelter, or a dog cannot be rehomed at all. The data (Figure_3.1) is right-skewed with a tail to the left, indicating fewer dogs with greater number of

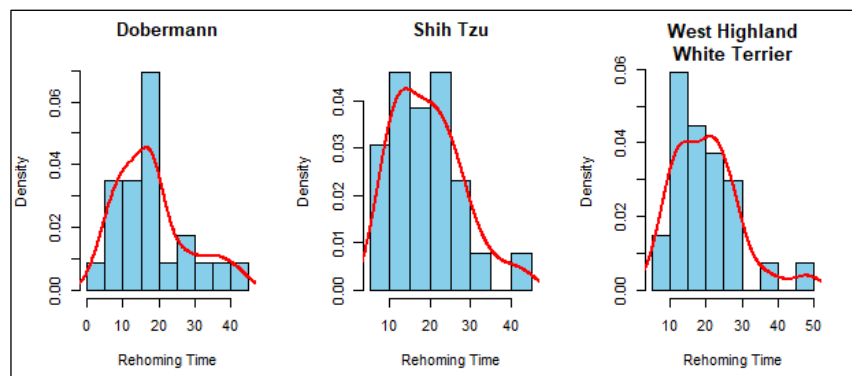


Figure 3.1: Histograms of rehoming times in weeks against density for each breed.

rehoming weeks. Thorough analysis determines the data follow a log-normal distribution.

Lognormal distribution is that when taking the natural log of data, the log data shows a normal distribution of an approximate bell-shaped [1] (Figure_3.2).

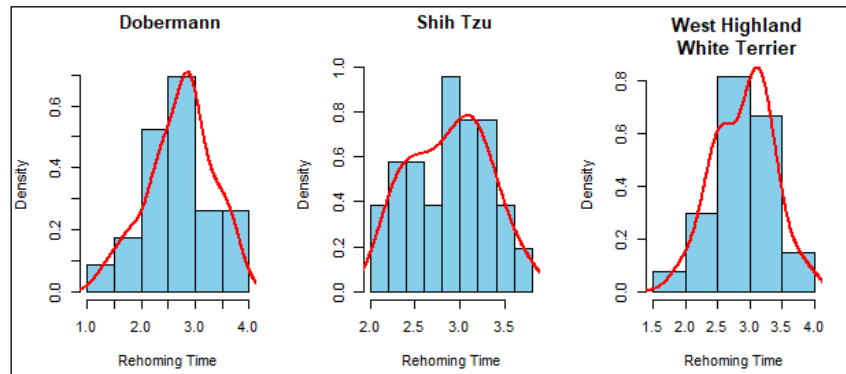


Figure 3.2: Histograms of rehoming times in weeks against density (log-transformed) for each breed

For further analysis, the data will be log transformed, and conclusions will be drawn by retransforming the data. Shapiro test determines the normality of distribution.

Breed	W	p
Dobermann	0.970	0.710
Shih tzu	0.966	0.547
West Highland White Terrier	0.976	0.771

Table 3: Computed Shapiro test values on 76 records for three breeds

Values for W approaches 1 and p (Table_3) are above the significance level of 0.05[2] indicating that the log-normal distribution is a good fit for normal distribution.

Breed	μ (mean)	σ	Exp(μ)
Dobermann	2.729	0.586	15.244
Shih Tzu	2.883	0.423	17.878
West Highland White Terrier	2.911	0.437	18.377

Table 4: Computed estimates via the Maximum Likelihood function of 76 records for three breeds

Values of μ (average size), σ (spread out) are estimated using the Maximum Likelihood function. The goal is to try different combinations of μ and σ to see which ones make the observed numbers most probable. A lower σ suggests that the data points are tightly clustered around the mean with less variability of data (Table_4). The least rehoming time is for Dobermann, followed by Shih Tzu and West Highland, as indicated by the original scaled data (Exp(μ)).

Inference

Confidence interval will tell us how accurate our estimates (Table_4) are likely to be, with a range bounded above and below the estimated mean containing the true average of the rehoming time for a specific dog breed. To determine whether the mean of 27 weeks lies within the confidence intervals, a

t-test is done as the sample size (n) is quite small [3] with confidence level of 0.95, meaning we will be 95% confident that the true mean lies within the intervals. The Central Limit Theorem states, the distribution of the sample mean drawn from the sample taken from population will approximately be normal if the sample is large enough, even if the original population mean is not.

Breed	t value	p value	df
Dobermann (n = 23)	- 4.5741	0.0001484	22
Shih Tzu (n = 26)	- 4.8647	5.291e-05	25
West Highland White Terrier (n = 27)	- 4.4879	0.0001299	26

Table 5: Computed t - test values on 76 records for three breeds, n is the sample size of breed

The small sample size with degrees of freedom ($df = n - 1$) for each breed (Table_5), creates a distortion in the distribution; the sample mean will be a less precise estimate, contributing to additional uncertainty. To compensate, t-test uses t values from the t distributions [4]. One sampled t-test is done to compare the mean of a single sample (each breed) to a known hypothesized mean (27 weeks). Null hypotheses state no difference between the specified value (27) and the sample mean (Table_4). Two-tailed t-test to conclude about the comparison.

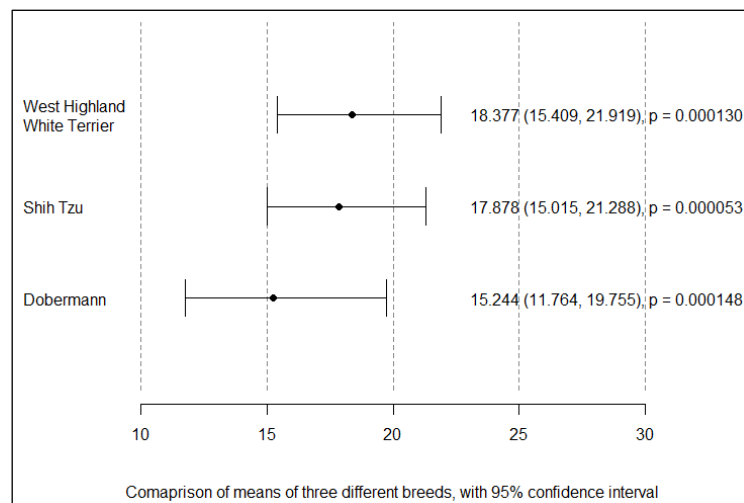


Figure 4: A caterpillar/forest plot of 95% confidence intervals for three different breeds via a two-sample t - test

27 does not lie in the confidence intervals of any breed (Figure_4). Negative t values indicate that the true mean is below the hypothesized mean. The confidence intervals for all the breeds indicate that none have a mean rehoming time of 27 weeks. The lower (11.763) and the upper bound (19.755) intervals are the shortest and longest plausible rehoming times. Dobermann has a wide range of plausible times as compared to the other two breeds. The confidence intervals for all the breeds notably deviated from the population mean of 27 weeks. These highlight the fact that mean rehoming time will deviate from 27 weeks.

Comparison

Two sample Welch's t-test, due to the data's unequal variance (Table_6) is done to compare rehoming mean time for each breed to find confidence interval for the difference in mean rehoming between each breed. This test assumes that the data is approximately normally distributed, as demonstrated in previous

sections. The null hypothesis states that the mean difference is 0. The confidence level is 95%, resulting in 5% (0.05) significance level, which is the risk of rejecting a true null hypothesis.

Breed A	Breed B	Confidence intervals	t value	p value	df
Dobermann	Shih Tzu	0.628, 1.157	- 1.055	0.297	39.529
Dobermann	West Highland White Terrier	0.610, 1.126	- 1.233	0.2246	40.066
Shih Tzu	West Highland White Terrier	0.763, 1.239	- 0.228	0.820	50.997

Table 6: Computed two sample t – test values for comparing the difference in mean for 76 records of three breeds.

The negative t values (Table_6) indicate the mean of Breed A is less than Breed B, and all the p values are beyond 0.05, indicating we don't have enough evidence to fail to reject the null hypothesis. There isn't a statistically significant difference in the mean rehoming times between all breed pairs, despite numerically different.

DISCUSSIONS

The study reveals that the three dog breeds, namely Dobermann, Shih Tzu, and West Highland White Terrier, show that the rehoming time is breed-dependent and is less than 27 weeks. However, the data is too small to draw a strong conclusion. Age and health should also be considered. The findings can be used to design adoption strategies for animal shelters, target marketing, and understand patterns. This helps match the potential adopters, reducing the rehoming time. Understanding rehoming times helps animal shelter staff understand and adapt daily operations, resulting in more efficiency according to breed. Considering location can provide insights, and transferring specific breeds to different locations ultimately reduces rehoming times and increases adoption rates. Prediction model could be developed by considering various variables to understand rehoming times.

REFERENCES

- [1] DeGroot, Morris, and Mark Schervish. Probability and Statistics: Pearson New International Edition, Pearson Education, Limited, 2013. ProQuest Ebook Central, pg. 312, definition 5.6.4.
- [2] Selcuk Korkmaz, Dincer Goksuluk and Gokmen Zararsiz: MVN: An R Package for Assessing Multivariate Normality, The R Journal Vol. 6/2, December 2014, pg. 157, para 1.
- [3] https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_confidence_intervals/bs704_confidence_intervals_print.html#:~:text=Consequently%2C%20the%2095%25%20CI%20is,to%20include%20the%20unknown%20parameter. Author: Lisa Sullivan, Date accessed: 9 December 2023
- [4] <https://www.cs.cmu.edu/afs/cs/project/jair/pub/volume20/long03a-html/node64.html>
Author: Derek Long, date written: 06/11/2003, Date accessed: 11 December 2023
- [5] Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
- [6] Venables, W. N. & Ripley Wickham H, François R, Henry L, Müller K, Vaughan D (2023). `_dplyr: A Grammar of Data Manipulation_`. R package version 1.1.3, <<https://CRAN.R-project.org/package=dplyr>>., B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
- [7] H. Wickham. `ggplot2: Elegant Graphics for Data Analysis`. Springer-Verlag New York, 2016.
- [8] R Core Team (2023). `_R: A Language and Environment for Statistical Computing_`. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>.

APPENDIX

The following sections contain the R code using R packages [5][6][7][8] scripted to carry out all the tasks in this report.

Data cleaning

```
mysample_copy <- mysample
mysample_copy <- mysample_copy[!is.na(mysample_copy$Breed),]
rownames(mysample_copy) <- NULL
rows_to_impute <- which(mysample_copy$Rehomed == 99999)
for (row_index in rows_to_impute) {
  breed <- mysample_copy$Breed[row_index]
  breed_mean <- mean(mysample_copy$Rehomed[mysample_copy$Breed == breed &
mysample_copy$Rehomed != 99999], na.rm = TRUE)
  mysample_copy$Rehomed[row_index] <- breed_mean
}
```

Data splitting

```
dobermann_data <- mysample_copy %>% filter(Breed == "Dobermann")
shihtzu_data <- mysample_copy %>% filter(Breed == "Shih Tzu")
westhighland_data <- mysample_copy %>% filter(Breed == "West Highland White Terrier")
```

Numerical summaries for each breed

```
summary_dobermann <- breed_dfs[[1]] %>%
  summarise(
    Mean_Rehomed = mean(Rehomed, na.rm = TRUE),
    Median_Rehomed = median(Rehomed, na.rm = TRUE),
    SD_Rehomed = sd(Rehomed, na.rm = TRUE)
  )

summary_shihtzu <- breed_dfs[[2]] %>%
  summarise(
    Mean_Rehomed = mean(Rehomed, na.rm = TRUE),
    Median_Rehomed = median(Rehomed, na.rm = TRUE),
    SD_Rehomed = sd(Rehomed, na.rm = TRUE)
  )

summary_westhighland <- breed_dfs[[3]] %>%
  summarise(
    Mean_Rehomed = mean(Rehomed, na.rm = TRUE),
    Median_Rehomed = median(Rehomed, na.rm = TRUE),
    SD_Rehomed = sd(Rehomed, na.rm = TRUE)
  )
```

Scatter plot

```
scatter_plot <- ggplot(mysample_copy, aes(x = Age, y = Rehomed, color = Breed)) +
  geom_point() +
  labs(title = "Relationship Between Age and Rehoming Time by Breed",
    x = "Age",
    y = "Rehoming Time (weeks)",
    color = "Breed") +
  theme_minimal()
```

Box plot

```
boxplot <- ggplot(mysample_copy, aes(x = Breed, y = Rehomed)) +
  geom_boxplot() +
  labs(title = "Box Plots of Rehoming Times by Breed",
    x = "Breed",
    y = "Rehoming Time (weeks)") +
  theme_minimal(base_size = 10) +
```

Calculate IQR

```
iqr_data <- mysample_copy %>%
  group_by(Breed) %>%
  summarise(
    Q1 = quantile(Rehomed, 0.25, na.rm = TRUE),
    Median = median(Rehomed, na.rm = TRUE),
    Q3 = quantile(Rehomed, 0.75, na.rm = TRUE),
    IQR = IQR(Rehomed, na.rm = TRUE)
  )
```

Plot histograms

```
hist_data_d <- dobermann_data$Rehomed
hist_data_s <- shihtzu_data$Rehomed
hist_data_w <- westhighland_data$Rehomed

hist(hist_data_d, probability = TRUE, col = "skyblue", border = "black", xlab = "Rehoming Time", main =
"Dobermann")
lines(density(hist_data_d), col = "red", lwd = 2)

hist(hist_data_s, probability = TRUE, col = "skyblue", border = "black", xlab = "Rehoming Time", main =
"Shih Tzu")
lines(density(hist_data_s), col = "red", lwd = 2)

hist(hist_data_w, probability = TRUE, col = "skyblue", border = "black", xlab = "Rehoming Time", main =
"West Highland\nWhite Terrier")
lines(density(hist_data_w), col = "red", lwd = 2)

hist(log(hist_data_d), probability = TRUE, col = "skyblue", border = "black", xlab = "Rehoming Time", main =
" Dobermann ")
lines(density(log(hist_data_d)), col = "red", lwd = 2)

hist(log(hist_data_s), probability = TRUE, col = "skyblue", border = "black", xlab = "Rehoming Time", main =
" Shih Tzu ")
lines(density(log(hist_data_s)), col = "red", lwd = 2)

hist(log(hist_data_w), probability = TRUE, col = "skyblue", border = "black", xlab = "Rehoming Time",
main = " West Highland\nWhite Terrier ")
lines(density(log(hist_data_w)), col = "red", lwd = 2)
```

Shapiro Test

```
shapiro.test(log(dobermann_data$Rehomed))
shapiro.test(log(shihtzu_data$Rehomed))
shapiro.test(log(westhighland_data$Rehomed))
```

Maximum Likelihood function

```
fit <- fitdistr(dobermann_data$Rehomed, densfun = "lognormal")
estimated_mu_log <- fit$estimate[1]
estimated_sigma_log <- fit$estimate[2]
estimated_mu = exp(estimated_mu_log)

fit <- fitdistr(shihtzu_data$Rehomed, densfun = "lognormal")
estimated_mu_log <- fit$estimate[1]
estimated_sigma_log <- fit$estimate[2]
estimated_mu = exp(estimated_mu_log)

fit <- fitdistr(westhighland_data$Rehomed, densfun = "lognormal")
estimated_mu_log <- fit$estimate[1]
estimated_sigma_log <- fit$estimate[2]
estimated_mu = exp(estimated_mu_log)
```

t – tests

```
t.test(log(dobermann_data$Rehomed), alternative = "two.sided", mu = 3.295837, conf.level = 0.95)
t.test(log(shihtzu_data$Rehomed), alternative = "two.sided", mu = 3.295837, conf.level = 0.95)
t.test(log(westhighland_data$Rehomed), alternative = "two.sided", mu = 3.295837, conf.level = 0.95)
```


Caterpillar/forest plot

```
analysis = c("Dobermann","Shih Tzu","West Highland \nWhite Terrier")

estimate = c(15.244, 17.878, 18.377)
upper = c(19.75457, 21.28758, 21.91868)
lower = c(11.76393, 15.01519, 15.40892)
pval = c(0.0001484,5.291e-05,0.0001299)

par(mar = c(6, 6, 1, 6))
plot(x = 0,
     xlim = c(10, 30), ylim=c(0, 4),
     type = "n", xaxt = "n", yaxt="n",
     xlab = NULL, ylab= NULL, ann = FALSE,
     bty="n")

axis(side = 1, cex.axis = 1)
mtext("Comaprison of means of three different breeds, with 95% confidence interval",
     side = 1, line = 4)

for(i in c(10, 15, 20, 25, 30)){
  lines(c(i, i), c(0, 4), lty = 2, col = "gray53")
}

verticalpos = 1:3
mtext(text = analysis, at = verticalpos,
     side = 2, line = 5, outer = FALSE, las = 1, adj = 0)

points(estimate, verticalpos, pch = 16)

for(i in 1:3 ){
  lines(c(lower[i], upper[i]), c(verticalpos[i], verticalpos[i]))
  lines(c(lower[i], lower[i]), c(verticalpos[i] + 0.2, verticalpos[i] - 0.2))
  lines(c(upper[i], upper[i]), c(verticalpos[i] + 0.2, verticalpos[i] - 0.2))
}

est <- formatC(estimate, format='f', digits = 3)
P <- formatC(pval , format = 'f', digits = 6)
pval <- paste("p =", P)

L <- formatC(lower, format = 'f', digits = 3)
U <- formatC(upper, format = 'f', digits = 3)
interval <- paste("(", L, ", ", U, ")", sep = "")

results <- paste(est, interval, pval)
mtext(text = results, at = verticalpos,
     side = 4, line = 4, outer = FALSE, las = 1, adj = 1)
```

Welch's two sampled t – test

```
t.test(log(dobermann_data$Rehomed), log(shihtzu_data$Rehomed))
t.test(log(dobermann_data$Rehomed), log(westhighland_data$Rehomed))
t.test(log(shihtzu_data$Rehomed), log(westhighland_data$Rehomed))
```