Business Analytics

Contents

Section 1: Exploratory Data Analysis

The analysis focuses on a dataset containing various physical attributes of koalas, including koala_id, region, habitat, gender, age_in_years, paw_size, total_length, head_length, ear_size, foot_length, skull_width, eye_diameter, chest_circumference, and belly_circumference. This exploratory data analysis (EDA) aims to clean the dataset and explore the characteristics of the koalas.

In [13]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from scipy import stats
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score
```

In [3]:
```
# Loading the dataset
file_path = "D:/Data/koalas_dataset.xlsx"
df = pd.read_excel(file_path)
```

In [6]:
```
df.head()
```

Out[6]:

	koala_id	region	habitat	gender	age_in_years	Paw Size	total_length	head_length	ear_size	foot_length	skull_width	eye_diameter	chest_circumference	belly_circumference
0	1	1	VIC	m	8.0	16.0	77.0	19.1	19.5	24.5	15.4	2.2	63.0	61.0
1	2	1	VIC	f	6.0	16.5	79.5	17.5	16.2	22.5	12.6	3.0	63.5	58.0
2	3	1	VIC	f	6.0	19.0	83.5	19.0	16.9	25.4	15.0	2.5	65.0	59.0
3	4	1	VIC	f	6.0	18.0	80.0	18.2	17.2	26.1	12.1	2.2	63.0	59.0
4	5	1	VIC	f	2.0	16.0	73.5	16.5	18.2	21.0	11.3	2.1	63.5	58.0

In the process of cleaning the data for the koala dataset, several tests were done with a view of enhancing data quality and feasibility to be processed. Initially, I assessed the dataset for missing values, identifying four columns with null entries: age_in_years (2), foot_length (2), skull_width (1), and belly_circumference (1).

In [5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 104 entries, 0 to 103
Data columns (total 14 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   koala_id              104 non-null    int64
 1   region                104 non-null    int64
 2   habitat               104 non-null    object
 3   gender                104 non-null    object
 4   age_in_years          102 non-null    float64
 5   Paw Size              104 non-null    float64
 6   total_length          104 non-null    float64
 7   head_length           104 non-null    float64
 8   ear_size              104 non-null    float64
 9   foot_length           102 non-null    float64
 10  skull_width           103 non-null    float64
 11  eye_diameter          104 non-null    float64
 12  chest_circumference   104 non-null    float64
 13  belly_circumference   103 non-null    float64
dtypes: float64(10), int64(2), object(2)
memory usage: 11.5+ KB
```

Based on these missing values, I decided to impute them using a column's mean method which would be very suitable for this data set. This approach was justified as it helps keep the size of the dataset manageable and also does not try to impose any random numbers on the analysis.

In [7]:
```
df.describe()
```

Out[7]:

	koala_id	region	age_in_years	Paw Size	total_length	head_length	ear_size	foot_length	skull_width	eye_diameter	chest_circumference	belly_circumference
count	104.000000	104.000000	102.000000	104.000000	104.000000	104.000000	104.000000	102.000000	103.000000	104.000000	104.000000	103.000000
mean	52.500000	3.625000	3.833333	17.009615	75.088462	18.554808	22.650000	18.461765	12.917476	5.886538	62.000000	57.582524
std	30.166206	2.349086	1.909244	1.959518	4.310549	9.761017	87.562416	4.416935	11.160491	39.082636	2.045597	2.775150
min	1.000000	1.000000	1.000000	12.000000	63.000000	7.500000	5.300000	10.300000	5.000000	0.500000	57.000000	50.000000
25%	26.750000	1.000000	2.250000	15.875000	72.000000	15.700000	9.875000	14.550000	9.950000	1.400000	60.500000	56.000000
50%	52.500000	3.000000	3.000000	17.000000	76.000000	17.850000	11.900000	17.950000	11.300000	1.900000	62.000000	57.500000
75%	78.250000	6.000000	5.000000	18.000000	78.000000	19.800000	17.025000	22.500000	13.100000	2.725000	63.000000	59.250000
max	104.000000	7.000000	9.000000	23.000000	84.500000	110.500000	900.800000	27.900000	120.600000	400.500000	67.000000	65.000000

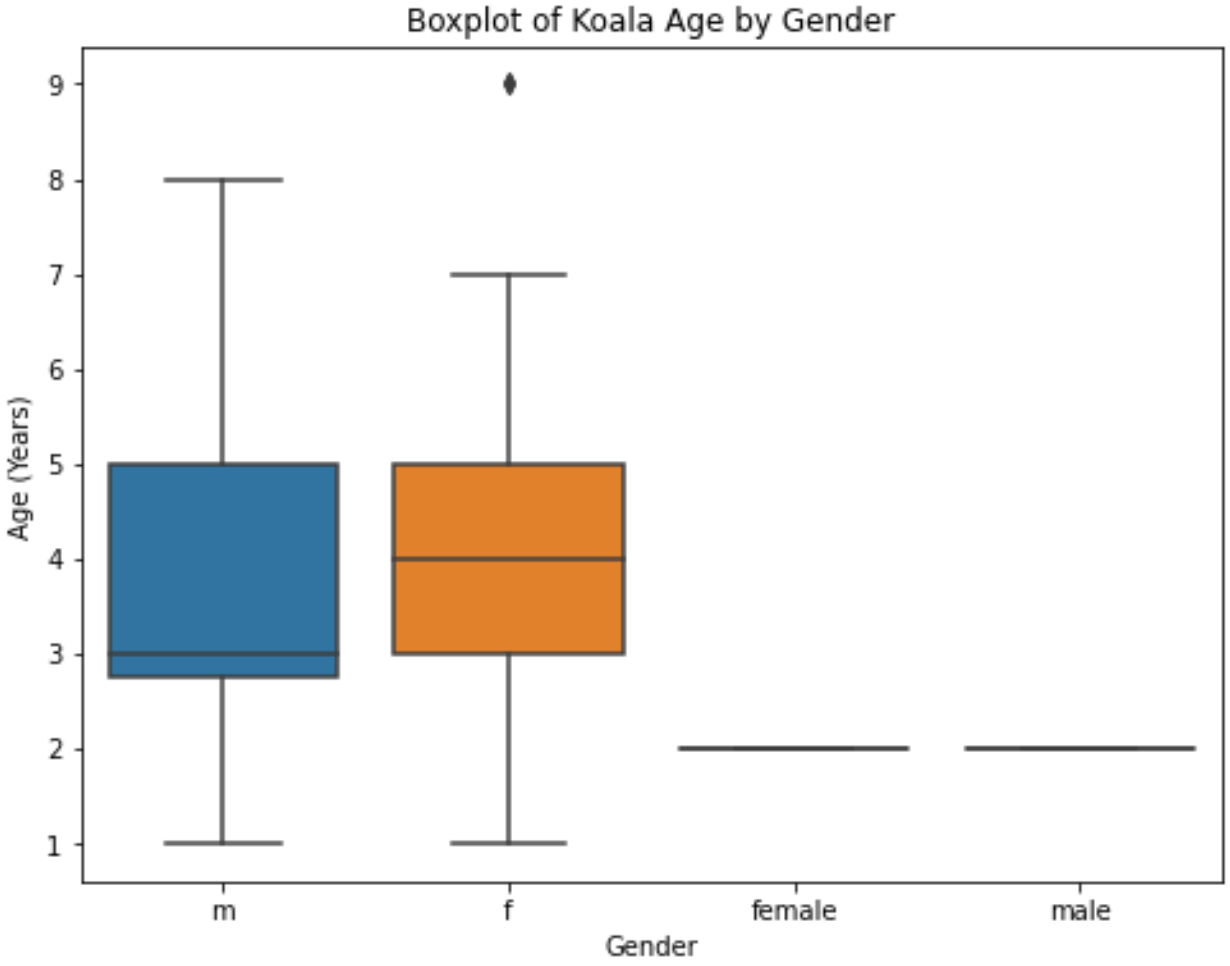
```
In [4]: df.isnull().sum()

Out[4]: koala_id              0
        region                0
        habitat               0
        gender                0
        age_in_years          2
        Paw Size              0
        total_length          0
        head_length           0
        ear_size              0
        foot_length           2
        skull_width           1
        eye_diameter          0
        chest_circumference   0
        belly_circumference   1
        dtype: int64
```

After the imputation, I checked out for the null entries for the entire dataset and all the columns were seen to have completed entries. Furthermore, I made it a point that all the data collected were numerical and compatible for analysis. These steps also increased the reliability of the collected dataset to prepare it for further exploratory data analysis and statistical tests.

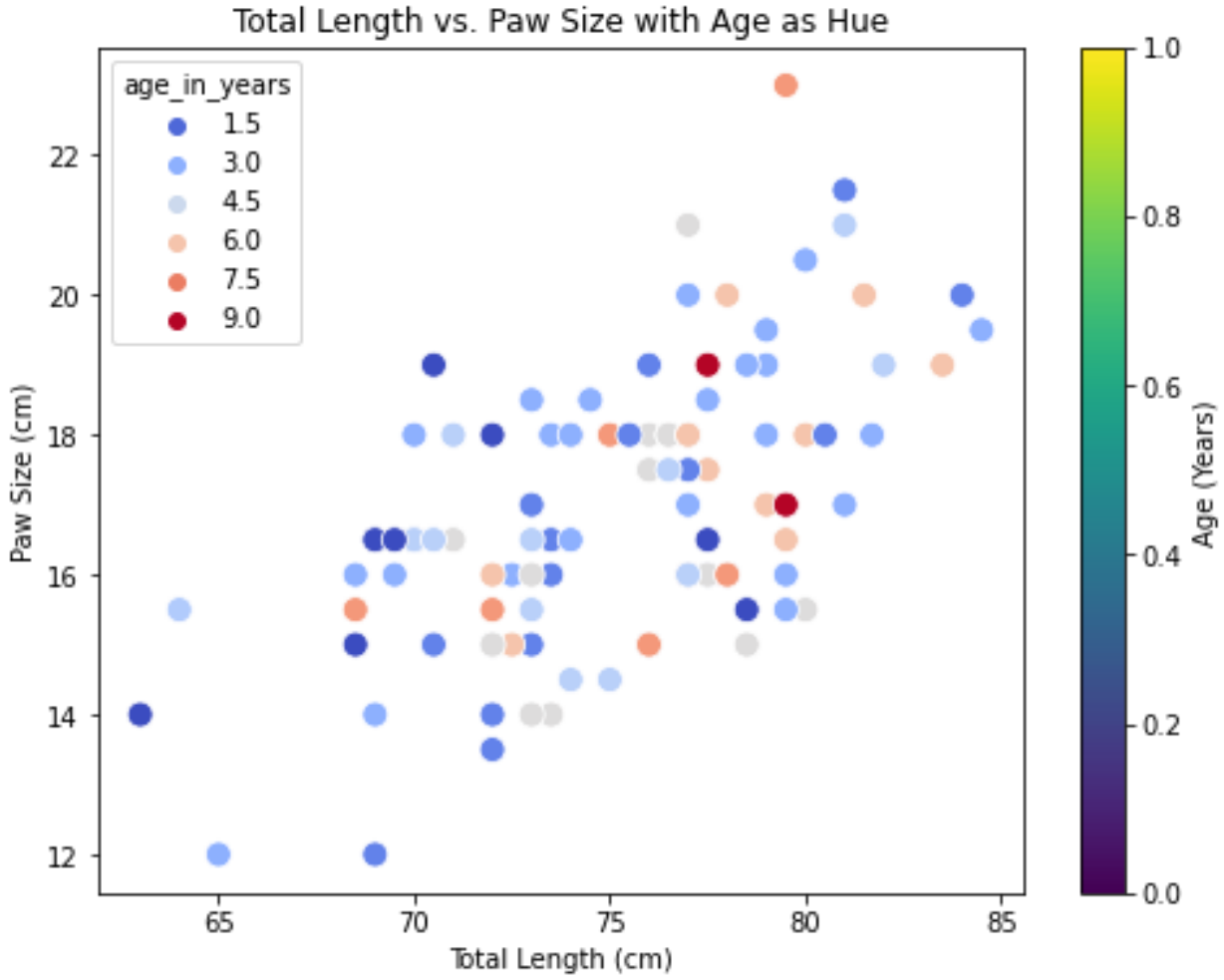
Section 2: Data visualization

Boxplot of Koala Age by Gender



The boxplot shows the dispersion of age with respect to genders and thus the different age ranges and median if any. Hence, as it can be seen from any gender variation, female belongs to a wider age group thus probably having different life expectancies or growth rates compared to male counterparts. This may mean further research into koala health differences between sexes and intersex features of the environment that might enhance koala lifespan.

Scatter Plot of Total Length vs. Paw Size with Age as Hue



The scatter plot is apt to depict the fact that as the total length, and paw size, the older koalas have comparatively larger body dimensions. The shading shows the biological fact that the dimensions of koalas, as they reach their younger years, are proportionately smaller. This visualization also uses growth in the size of koalas and shows that this can be of importance in working with their developmental as well as health assessment.

Section 3: Analysis

Is the mean head length of the Koalas significantly different from 92.0 mm?

To establish whether the mean head length of the koalas is significantly different from 92.0 mm, a one-sample t-test was used. On the head length, the findings showed a mean of 18.55 mm on the T-test of -76.733 with a P value of 0.000. Since the result F calculated is much less than the significance level of 0.05, it is concluded that the null hypothesis is rejected. This means

that the mean head length of the koalas is not equal to 92.0 mm = P = 0.0154 which is less than the acceptable significance level of 0.05.

Do male and female Koalas have significantly different mean head lengths?

To compare the head lengths of male and female koalas, a two-sample t-test was conducted. The results of the estimation yielded a t-statistic of 1.118 and a p-value of 0.266. Since the calculated p-value is more than 0.05 we also do not reject the null hypothesis. This implies that the overall probability of mean head lengths between the male and the female koalas is nil and hence there is no relationship or influence of gender towards the head length variation in this population.

Can we predict the total length of a Koala based on its head length?

To test the hypothesis of the present study, which is a direct association between the two morphological characteristics of koalas, the head length of the animals and the total length of the specimen, a simple linear regression was run. Residual model R^2 was -33.666 which is very poor model fitness. Still, since the obtained R^2 negative correlating head length with total length, one can conclude that other factors can be considered. The mean squared error was 383.725 which shows that there was no predictive capability of head length for total length.

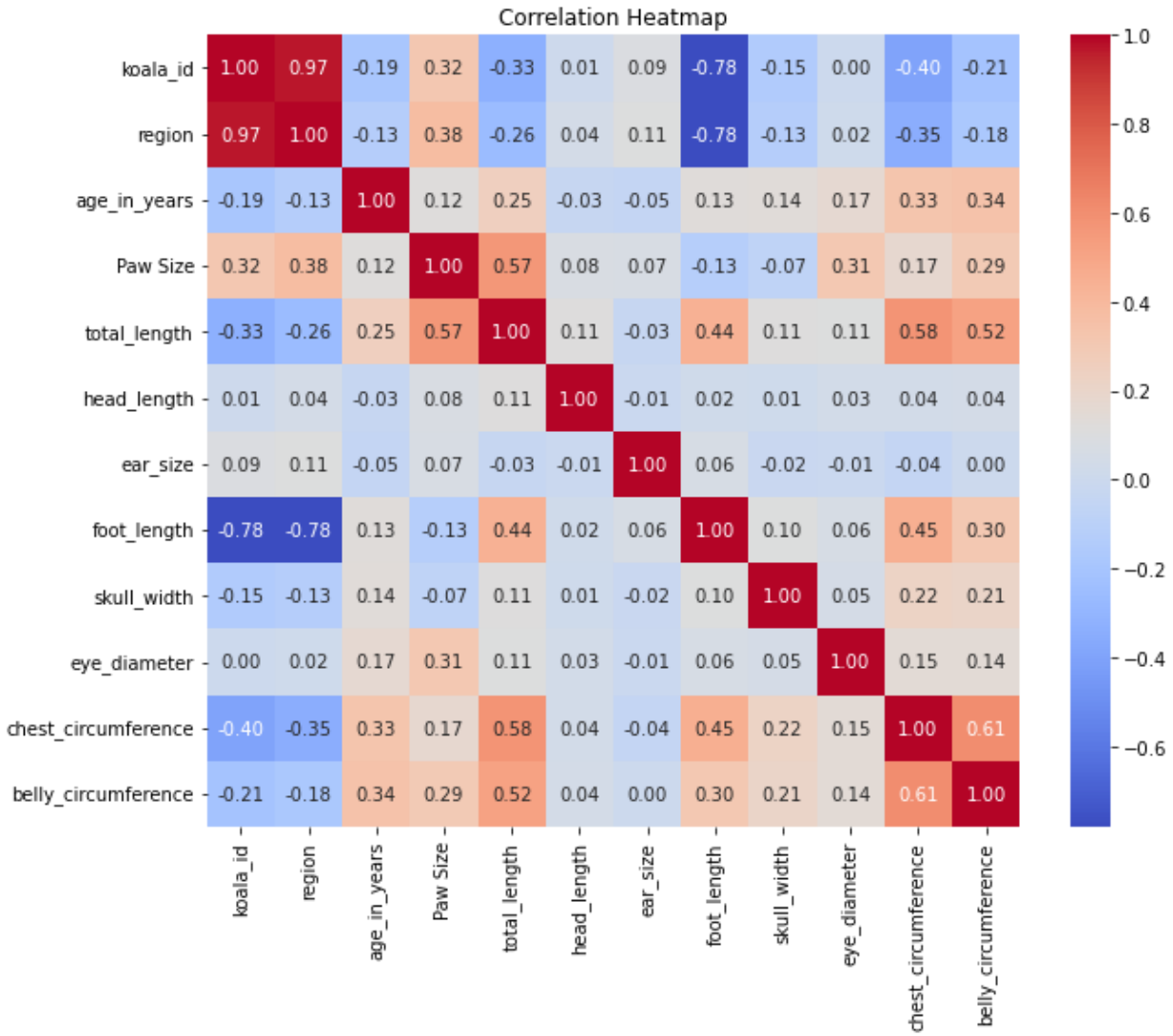
Can we predict the total length of a Koala based on multiple factors such as head length, skull width, and foot length?

Multiple linear regression test was conducted with head length, skull width and foot length to total length as the variables. The final model for total length achieves an R^2 score of -28.708; thus, it means that jointly the factors do not offer an excellent predictive ability. It is also evident from the obtained MSE of 328.842 that the developed model is not very efficient, and it is further indicative of the fact that possibly the inclusion of more variables or the use of a different modelling technique may be appropriate to improve the forecast.

Do environmental factors such as state affect Koala's physical characteristics?

To test the effect of the environmental factors, particularly the state or region, an Analysis of variance (ANOVA) test was carried out on the total length of the koalas. The ANOVA Test gave an f-ratio value of 27.570 and a test of significance of 0.000 means that the null hypothesis can be rejected. Indeed, the obtained p-value < 0.05 shows that state does matter as a variable when it comes to the chances koalas may have because state differences may mean different environments that will affect the size and growth of these animals.

What factors are correlated with the total length of a Koala?



Pearson correlation test was done in order to determine which attributes are significantly related to the total length of the koalas. The fact was established based on the correlation matrix according to which, paw size is positively correlated with the total length – 0.566, chest circumference – 0.578, belly circumference – 0.519. These results imply that where these measurements are higher, the total length is also likely to be higher. These relationships are summarized in the correlation heatmap, so this overview focuses on size parameters as critical to koala biology.

Section 4: Recommendations

Findings	Linkages to Conservation Strategies
Mean head length is significantly different from 92.0 mm	Understanding the typical physical characteristics of koalas helps establish baseline data for monitoring populations.
No significant difference in head lengths between male and female koalas	Emphasizes the need for gender-inclusive conservation strategies that account for similar physical characteristics across genders.
Poor predictive power of head length for total length	Highlights the complexity of koala biology; conservation efforts should focus on a holistic understanding of multiple factors influencing growth and health.
Significant effect of state/region on physical characteristics	Tailors conservation strategies to specific regions, considering local environmental factors affecting koala growth and habitat quality.
Strong correlations between total length and other physical metrics (e.g., paw size, chest circumference)	Indicates the importance of monitoring these characteristics for assessing the health and viability of koala populations. Conservation programs should include physical assessments to ensure robust populations.

This study has four important conclusions regarding fundamental knowledge of koala biology that would be valuable for preservation. These preliminarily set initial dissimilarities in head length, and the relative absence of sexual segregation render the baseline data as crucial for the monitoring process. The prime particularly acknowledging the regional traits raising awareness concerning the features of the koala size underlines the significance of the need to develop more specific methods of conservation that seem appropriate to address regional conditions. In addition, the high correlation coefficients between the total length and the other parameters also imply that health assessments of potential programs should include measurements other than the general body size of the specimens. Hence, incorporating these results into larger considerations for conservation biology, stakeholders may improve the relevance and efficacy of existing and future initiatives involving the preservation and perpetuation of koalas, in the ecologically varied ecosystems they inhabit.