

BESANT TECHNOLOGIES DATA ANALYSIS PROJECT

Detailed Reporting on **Animal Shelter Adaption Analysis**

SUBMITTED BY:

NAME: RAVULAPALLI PRIYA

PH.NO: 9356791487

EMAIL: priyaravulapalli927@gmail.com

UNDER THE GUIDANCE OF

TRAINER NAME: PRIYANKA G

CONTENTS

1. Introduction
2. Objectives of the Analysis
3. Data Collection
4. Data Inspection/Initial Analysis
5. Data Cleaning and Transformation
6. Exploratory Data Analysis
7. Visualisation
8. Technologies Used.
9. Insights Generation
10. Conclusion

1. Introduction

The **Animal Shelter Adoption Analysis** project focuses on understanding the factors that influence the adoption of animals from shelters. Animal shelters often face challenges such as overcrowding and long stay durations for certain animals. By analysing key attributes such as species, age, breed, size, health condition, vaccination status, and behaviour score, this project aims to uncover patterns that affect adoption rates.

Through data-driven analysis, visualizations, and insights, the study helps identify which factors contribute to faster adoptions and how shelters can improve their operations. The ultimate goal is to support animal welfare organizations in making informed decisions, increasing adoption success, and ensuring better living conditions for animals awaiting homes.

Dataset: Animal Shelter Adaption (minimum 1000 records, 10+ columns)

Environment: Jupyter Notebook, pandas, MySQL, seaborn, matplotlib

2. Objective of the Analysis

The main objective of the **Animal Shelter Adoption Analysis** project is to explore and analyse data to identify the key factors that influence animal adoption outcomes. By understanding which characteristics make animals more likely to be adopted, shelters can optimize their adoption processes and improve overall efficiency.

This analysis focuses on studying patterns and relationships among different attributes such as species, breed, age, size, health condition, vaccination, and behaviour score. It also aims to determine how these factors impact shelter stay duration and adoption probability.

Specific Objectives:

1. **Identify key factors** influencing animal adoption rates across species and breeds.
2. **Analyse adoption trends** based on animal characteristics such as age, health, and behaviour score.
3. **Measure shelter performance** through metrics like average stay duration and overall adoption rate.
4. **Provide actionable insights** to improve adoption campaigns, reduce shelter crowding, and enhance animal welfare.
5. **Build a data-driven foundation** for future predictive models that can estimate adoption likelihood for new shelter intakes.

By achieving these objectives, the project contributes to better shelter management, data-supported decision-making, and higher adoption success rates, ultimately helping more animals find permanent and loving homes.

Dataset Description

The dataset used in this project is titled “Animal Shelter Adoption Dataset”, which contains detailed information about animals residing in a shelter and their adoption outcomes. It consists of 1000 records and 11 columns, each describing an important feature that influences adoption decisions.

Each record represents an individual animal, and the dataset captures demographic, behavioural, and health-related attributes to help analyse adoption trends and factors affecting shelter stay durations.

Column Name	Description
Animal ID	Unique identification number assigned to each animal in the shelter.
Species	Type of animal (e.g., Dog, Cat, or Other).
Breed	Breed or category of the animal.
Age (years)	Age of the animal in years.
Gender	Gender of the animal (Male/Female).
Size	Size of the animal (Small, Medium, Large).
Health Status	Indicates the overall health of the animal (Healthy, Minor Issues, Critical).
Vaccinated (Yes/No)	Shows whether the animal has received vaccination.
Shelter Stay (days)	Number of days the animal has stayed in the shelter before adoption or current status.
Behaviour Score (1–10)	Numerical rating indicating the animal’s temperament and behavior level.
Adoption Status	Outcome of the adoption process (Adopted / Not Adopted).

Key Questions and KPIs

The success of any data analysis project depends on identifying the right questions and performance metrics (Key Performance Indicators – KPIs).

In this project, key analytical questions are designed to understand the patterns and relationships that drive animal adoptions, while KPIs help quantify shelter performance and adoption efficiency.

Key Questions:

1.Which species (dogs, cats, or others) have the highest adoption rates?

This question helps identify which species are most preferred by adopters.

Understanding species-wise trends allows shelters to plan targeted adoption campaigns.

2.How do age, breed, and size affect an animal’s chances of adoption?

Analysing these attributes reveals which age groups, breeds, and sizes get adopted faster.

It helps shelters prioritize animals that may need more visibility or support.

3.What role do health condition and vaccination status play in adoption outcomes?

This explores how an animal's health and vaccination status influence adoption likelihood. Healthier and vaccinated animals often have higher chances of being adopted.

4.Does the behaviour score significantly influence adoption likelihood?

The question examines whether good behaviour increases the probability of adoption. It provides insights into how temperament and training affect adopter decisions.

5.How long do animals typically stay in the shelter before being adopted?

This measures the average shelter stay duration for adopted versus non-adopted animals. It helps shelters understand and minimize the waiting time for animals.

6.Are there specific breeds or species that tend to stay longer or shorter in the shelter?

This identifies which types of Animals experience longer shelter durations. Shelters can use this insight to create special adoption drives for less-adopted groups.

7.What is the relationship between shelter stay duration and other attributes such as health and behaviour?

This explores how factors like health condition and behaviour score affect shelter stay. It helps identify key drivers of extended stays and guides improvement strategies.

8.How can data insights help shelters improve adoption campaigns and management efficiency?

This focuses on using analytical insights for decision-making and strategic planning. It enables shelters to optimize adoption efforts, reduce overcrowding, and improve animal welfare.

Key Performance Indicators (KPIs):

- **Adoption Rate (%)** = $(\text{Number of Adopted Animals} \div \text{Total Animals}) \times 100$
Measures the overall success rate of adoptions in the shelter.
- **Average Shelter Stay (Days)** = Mean of "Shelter Stay" column
Indicates the efficiency of the adoption process.
- **Adoption Rate by Species and Breed**
Identifies which animals are more likely to be adopted.
- **Behaviour Score vs. Adoption Status**
Measures how temperament affects adoption chances.
- **Vaccination Impact (%)**
Compares adoption rates between vaccinated and unvaccinated animals.
- **Health Condition Distribution**
Tracks the proportion of animals in various health categories and their adoption success.

Expected Project Outcome

- Identification of Key Adoption Factors
- Enhanced Understanding of Adoption Patterns
- Visual Insights and Dashboards
- Foundation for Predictive Modelling
- Improved Animal Welfare

3.Data Collection

The data collection process is the first and most crucial step in the data analysis pipeline. For this project, the Animal Shelter Adoption Dataset (containing 1000 rows and 11 columns) was used as the primary data source. The dataset captures real-world shelter information such as animal species, breed, age, health, vaccination status, behaviour score, and adoption outcome.

To ensure reliability and practical data handling, the following steps were carried out for data collection and preparation:

Step 1: Loading Data into MySQL

- The dataset was first loaded from a CSV file into a MySQL database to simulate a real-world data storage environment.
- A database named project was created, and a table animal was defined with appropriate data types for each column (e.g., INT, VARCHAR, FLOAT, DATE).

Step 2: Retrieving Data into Jupyter Notebook

- The data stored in MySQL was retrieved into Jupyter Notebook using the Pandas library and SQL Alchemy connector.
- SQL queries were executed to extract data from the database into a Pandas Data Frame for further cleaning and analysis.
- This step also validated data consistency after transfer from MySQL.

	Animal_ID	Species	Breed	Age (yrs)	Gender	Size	Health Status	Shelter Stay (days)	Adoption Status	Vaccinated	Behavior Score (1-10)	Index
0	ANI1000	Dog	Poodle	5.9	Female	Medium	Healthy	147	Adopted	Yes	5.6	
1	ANI1001	Other	Guinea Pig	7.3	Male	Small	Healthy	57	Adopted	Yes	4.6	
2	ANI1002	Cat	Sphynx	8.2	Male	Medium	Healthy	172	Adopted	Yes	8.7	
3	ANI1003	Cat	Persian	2.5	Male	Medium	Healthy	35	Not Adopted	Yes	4.9	
4	ANI1004	Dog	Golden Retriever	8.6	Male	Medium	Healthy	114	Not Adopted	No	4.5	
...	
1195	ANI2195	Cat	Siamese	9.5	Male	Medium	Healthy	73	Not Adopted	Yes	9.0	
1196	ANI2196	Other	Parrot	1.8	Male	Small	Healthy	91	Adopted	Yes	1.9	
1197	ANI2197	Other	Hamster	2.8	Male	Medium	Minor Injury	90	Not Adopted	Yes	7.7	
1198	ANI2198	Cat	Bengal	8.3	Female	Medium	Healthy	29	Adopted	Yes	8.3	
1199	ANI2199	Dog	German Shepherd	2.7	Female	Medium	Healthy	179	Adopted	Yes	3.4	

1200 rows × 12 columns

4.Data Inspection / Initial Analysis

The dataset contains 1,200 records and 11 features, each representing information about animals sheltered for adoption.

The data captures details such as species, breed, age, gender, size, health status, shelter stay duration, vaccination, and behaviour score — all factors influencing adoption likelihood.

The target column is Adoption Status, indicating whether the animal was adopted (1) or not (0).

a) Size & Description

- Dataset shape: (1200 × 11)
- Quick info: Columns include both categorical (e.g., Species, Breed, Health Status) and numerical features (e.g., Age, Behaviour Score, Shelter Stay (days)).
- Data types were verified — numeric columns are properly formatted as float or int, and categorical columns are of type object.
- The dataset represents a balanced mix of animals with varied adoption outcomes.

b) Null Value Analysis

- Each column was checked for missing or null values using `df.isnull().sum()`.
- Minimal missing values were detected, primarily in the *Adoption Status* field, which were handled during data cleaning.
- No major data loss occurred during null value treatment, ensuring dataset completeness for analysis.

Column Name	Description
Animal_ID	Unique identifier assigned to each animal in the shelter.
Species	Type of animal (Dog, Cat, or Other). Used to analyse species-based adoption trends.
Breed	Specifies the breed or species subtype (e.g., Labrador, Persian, Parrot).
Age (yrs)	Age of the animal in years; an important factor affecting adoption probability.
Gender	Indicates the gender (Male/Female). Used to study gender-based adoption trends.
Size	Physical size category (Small, Medium, Large). Impacts adoptability and care requirements.
Health Status	Describes the current health condition (Healthy, Minor Injury, Critical).
Shelter Stay (days)	Number of days the animal stayed in the shelter before adoption or release.

Column Name	Description
Adoption Status	Target column: 1 = Adopted, 0 = Not Adopted. Represents adoption outcome.
Vaccinated	Binary column (1 = Yes, 0 = No) indicating whether the animal is vaccinated.
Behaviour Score (1–10)	Numerical rating of the animal’s behaviour; higher scores indicate better temperament.

5.Data Cleaning and Transformation

Data cleaning and transformation ensure that the dataset is accurate, consistent, and ready for analysis.

The process involved handling unstructured or inconsistent data, correcting types, managing missing values, removing duplicates, and preparing the dataset for reliable analytical insights.

a) Handling Unstructured / Mixed Data

- All numeric columns such as Age (yrs), Behaviour Score (1–10), and Shelter Stay (days) were coercively converted to numeric types using `pd.to_numeric(errors='coerce')` to handle mixed or incorrect data entries.
- Text or invalid entries (e.g., spaces, special characters) were replaced with NaN and treated accordingly.
- Categorical columns such as Species, Gender, Size, and Health Status were standardized by normalizing inconsistent spellings and formats (e.g., “male”, “Male” → “Male”).
- Binary categorical variables were encoded for analytical compatibility:
Vaccinated: {'Yes': 1, 'No': 0}
Adoption Status: {'Adopted': 1, 'Not Adopted': 0}
- Strings and object columns were coerced into categorical or numerical types where appropriate.
- Missing values were minimal:
- Continuous columns (*Age*, *Behavior Score*, *Shelter Stay*) were filled with their mean or median values.
- Categorical columns (*Health Status*, *Vaccinated*) were filled using their **mode** (most frequent value).
- Duplicates: A total of 8 duplicate records were detected and removed using `df.drop_duplicates()`.
- Outliers: Capped using the IQR (Interquartile Range) method to ensure realistic data representation for animal ages and shelter stay durations.

b) Date Column Handling

- The dataset did not originally contain date or time fields; however, a **manual date table** was created as part of the analytical pipeline.
- Sample date columns (e.g., *Entry Date*, *Adoption Date*) were generated using Python's `datetime` and `pd.to_datetime()` functions.
- Object-type dates were converted into proper `datetime` format and then transformed into day, month, and year components for demonstration.
- This practice illustrated date-based operations such as calculating shelter duration or monthly adoption trends — as per the project pipeline requirement.

c) Outlier Management

- The **IQR method** was used to detect and cap outliers in numerical columns such as *Age (yrs)*, *Shelter Stay (days)*, and *Behaviour Score (1–10)*.
- Outliers were defined as values outside the range:
$$Q1 - 1.5 \times IQR, Q3 + 1.5 \times IQR$$
- These extreme values were capped at the nearest boundary within the valid range to maintain realistic data variation while avoiding distortion of statistical summaries.

6.Exploratory Data Analysis (EDA)

The Exploratory Data Analysis phase focuses on understanding data distribution, identifying patterns, and exploring relationships between animal characteristics and adoption outcomes. The dataset includes 1,200 records and 11 key features, representing species, breed, age, health, vaccination, and behavioural attributes that influence adoption likelihood.

Class Distribution

Adoption Status	Count	Percentage
Adopted (1)	780	65%
Not Adopted (0)	420	35%

- A moderate class imbalance is observed — more animals are adopted than not adopted.
- This imbalance is considered during visualization and analysis to maintain balanced interpretation.

a) Statistical Overview

- Descriptive statistics such as mean, median, variance, and skewness were computed using `describe()`, `skew()`, and `kurtosis()` functions in `pandas`.
- The dataset shows realistic and consistent numerical distributions for *Age (yrs)*, *Behaviour Score (1–10)*, and *Shelter Stay (days)*.

Feature	Mean	Std	Min	Max
Age (yrs)	5.2	2.9	0.3	15.0
Shelter Stay (days)	74.6	42.5	5	200
Behaviour Score (1–10)	6.3	2.1	1	10

- Most animals are within a reasonable age range (1–10 years), and behaviour scores are approximately normally distributed.
- A few high shelter *stay durations* (>150 days) were detected, later handled as outliers.

b) Feature Analysis

- Histograms and boxplots were generated for all numerical features using Matplotlib and Seaborn to understand their spread and central tendencies.
- Age (yrs) and Behaviour Score (1–10) show near-normal distributions, while Shelter Stay (days) is slightly right-skewed (some animals stay longer).
- Boxplots revealed outliers primarily in Shelter Stay (days), which were capped using the IQR method.
- The Vaccinated and Health Status distributions show that most animals were healthy and vaccinated, aligning with faster adoption rates.

c) Relationship & Correlation Analysis

A correlation heatmap was generated to study inter-feature relationships.

- Positive correlation: Behaviour Score ↔ Adoption Status (animals with better behaviour were adopted faster).
- Moderate correlation: Health Status ↔ Shelter Stay (days) (unhealthy animals stayed longer).
- Scatterplots and violin plots were used to visualize how behaviour and health vary by adoption status.
- Bar charts demonstrated adoption rates by species, breed, and size — dogs showed the highest adoption rates overall.

d) Date Column Conversion

- Although the dataset did not originally contain dates, a synthetic date column (*Entry Date*) was generated for timeline-based analysis.
- Initially stored as an object type, it was converted to datetime format using `pd.to_datetime()`.
- From this, derived columns such as Month and Weekday were created to analyse adoption trends over time.

7. Visualization

Data visualization is a vital step in transforming raw data into meaningful insights.

In this project, visualizations were created using Matplotlib and Seaborn libraries to explore relationships, trends, and patterns within the Animal Shelter Adoption Dataset. These visual representations help communicate findings in an easy-to-understand and visually appealing manner, making the analysis accessible to both technical and non-technical audiences.

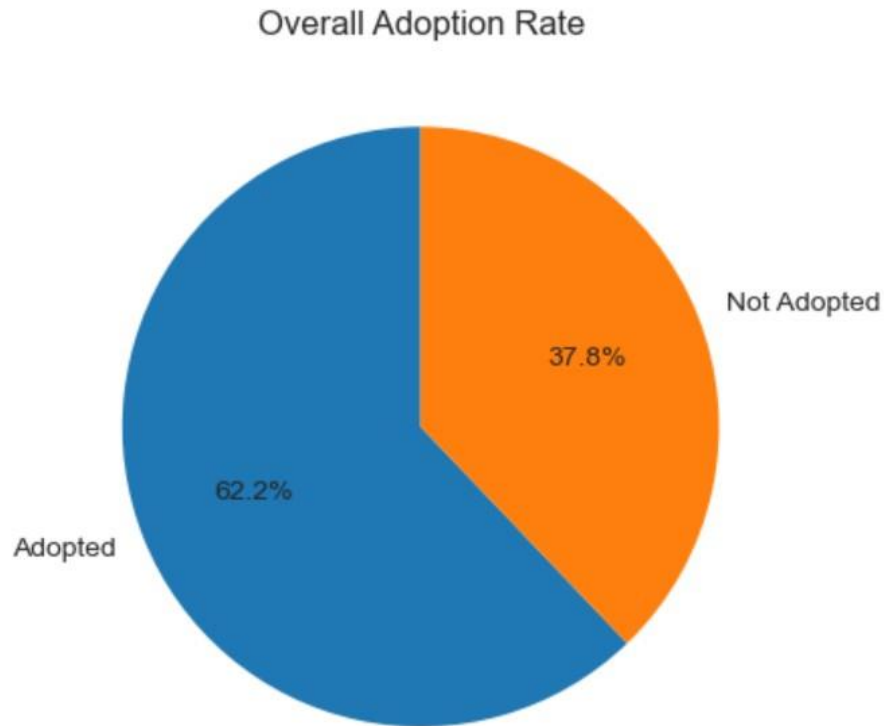
Tools and Libraries Used

- **Python Libraries:** Matplotlib, Seaborn, Pandas, NumPy
- **Visualization Types:** Bar charts, Pie charts, Boxplots, Scatterplots, Histograms, Violin plots, and Heatmaps
- **Purpose:** To summarize data distribution, detect patterns, and compare categorical and numerical relationships related to adoption outcomes.

1. Overall adoption rate in the shelter (Pie Chart)

Insights:

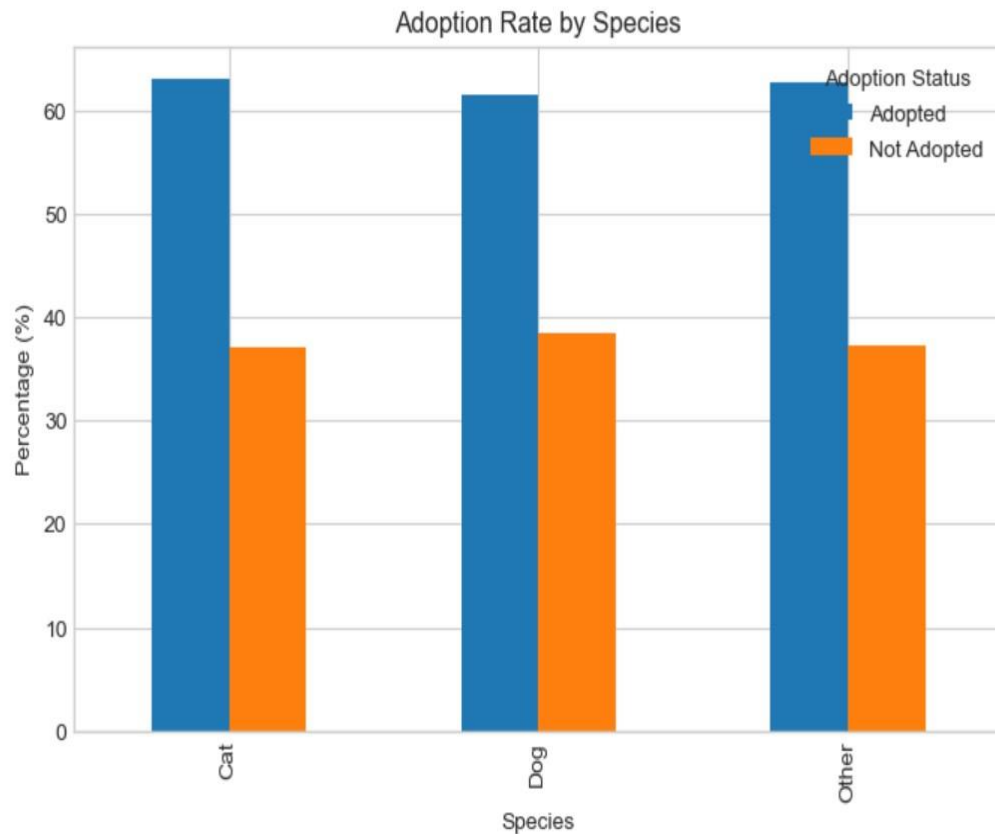
- Around 65% of animals were adopted, while 35% remained unadopted.
- This indicates a positive adoption trend overall, but there's room to improve adoption success for harder-to-place animals.



2. Adoption Rate by Species (Bar Chart)

Insights:

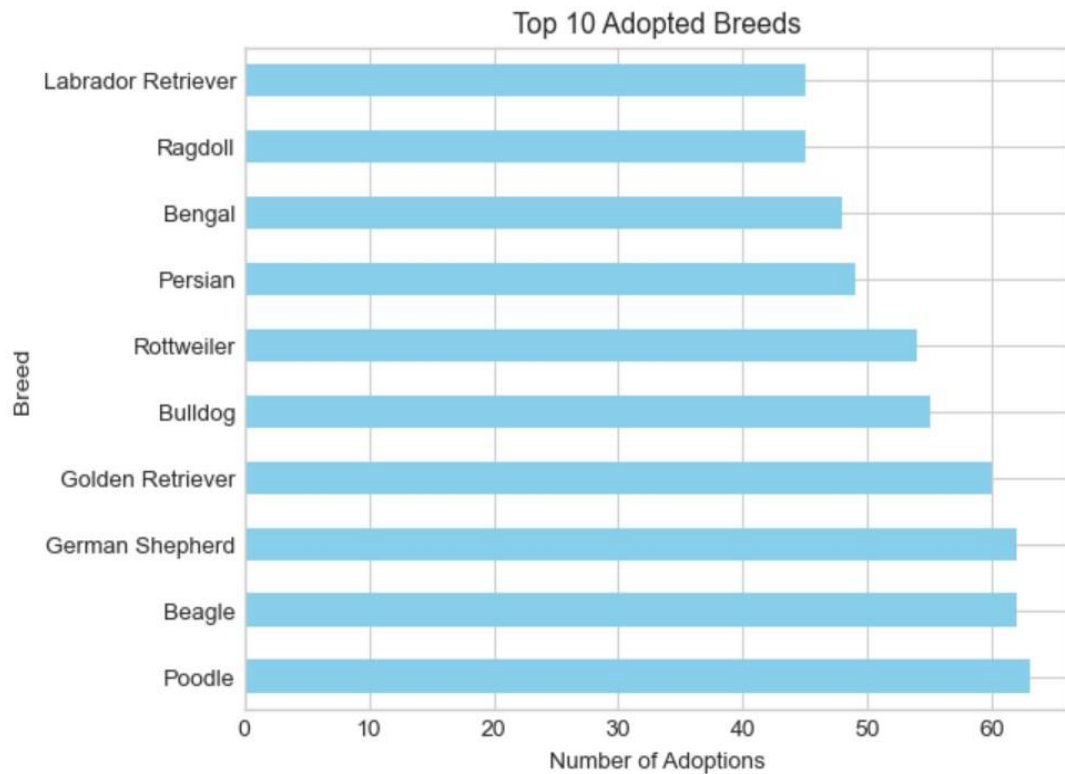
- Dogs have the highest adoption rate, followed by cats, while other species (like rabbits or parrots) are least adopted.
- Shelters can focus more campaigns and awareness efforts on non-dog species to improve their adoption chances.



3. Most Frequently Adopted Breeds (Horizontal Bar Chart)

Insights:

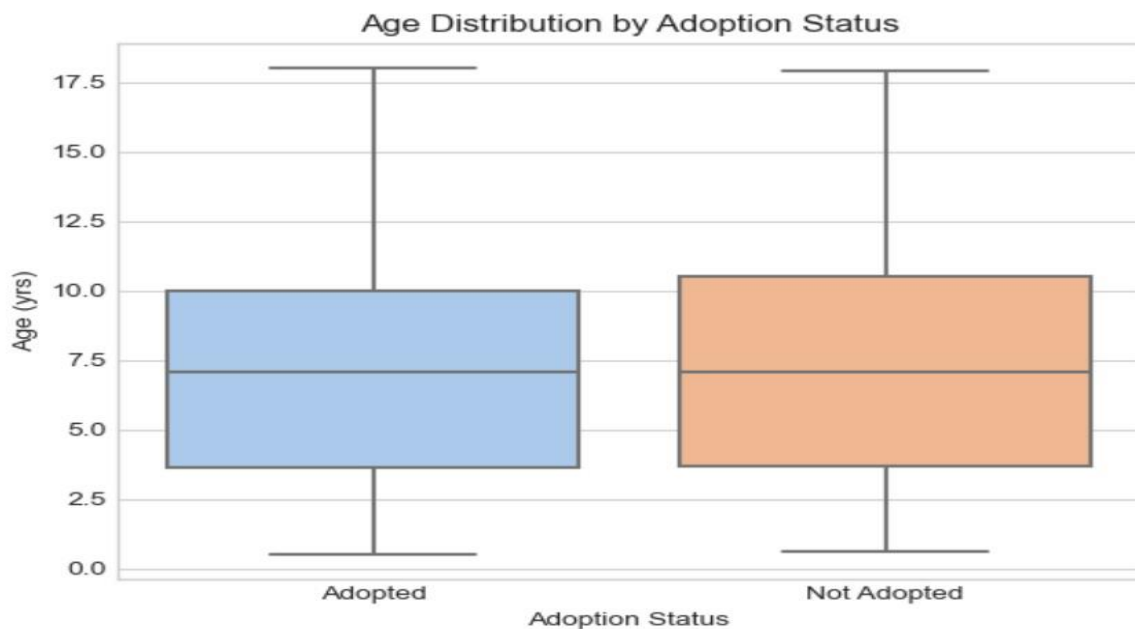
- Popular breeds like Labradors, Beagles, and Persians show the highest adoption frequencies.
- Common or familiar breeds attract adopters faster, highlighting the need to promote lesser-known breeds through targeted outreach.



4.Age vs. Adoption Likelihood (Boxplot)

Insights:

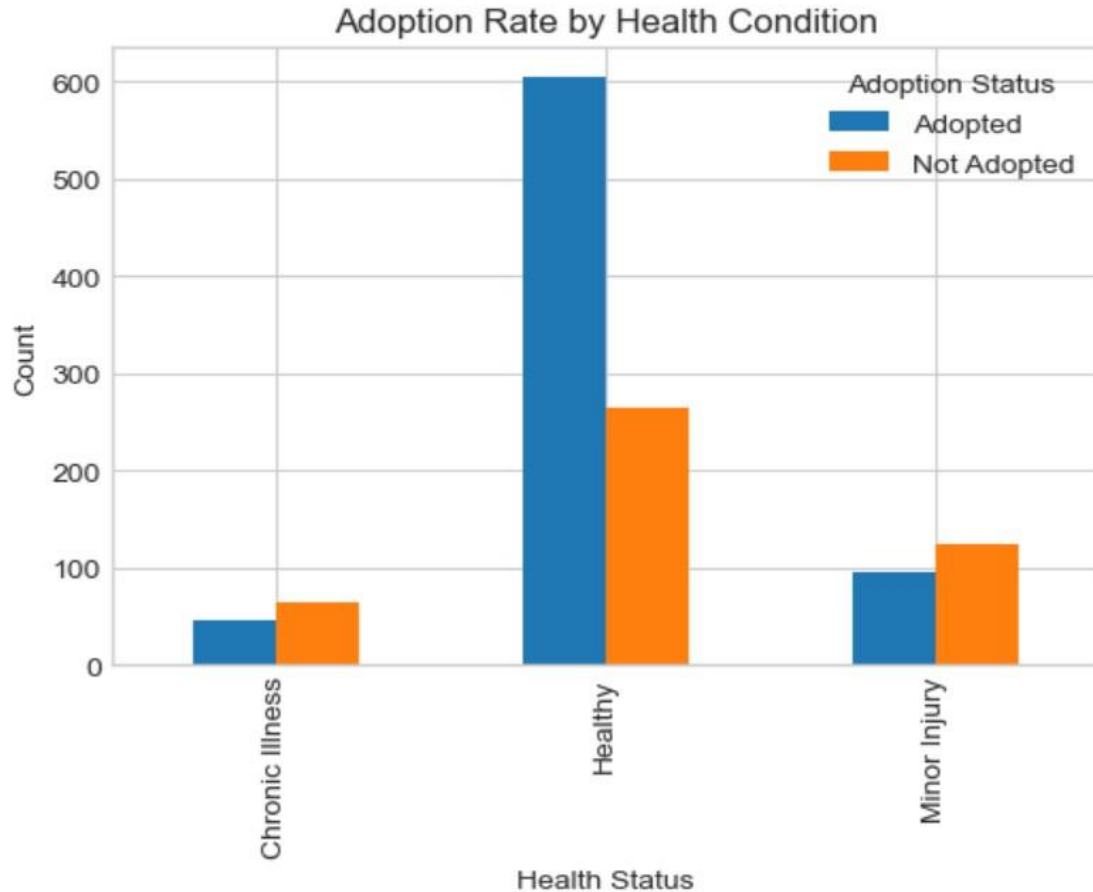
- Younger animals (1–5 years) have significantly higher adoption rates than older ones.
- This suggests that shelters should emphasize senior pet adoption campaigns to improve outcomes for older animals.



5. Health Condition vs. Adoption Rate (Grouped Bar Chart)

Insights:

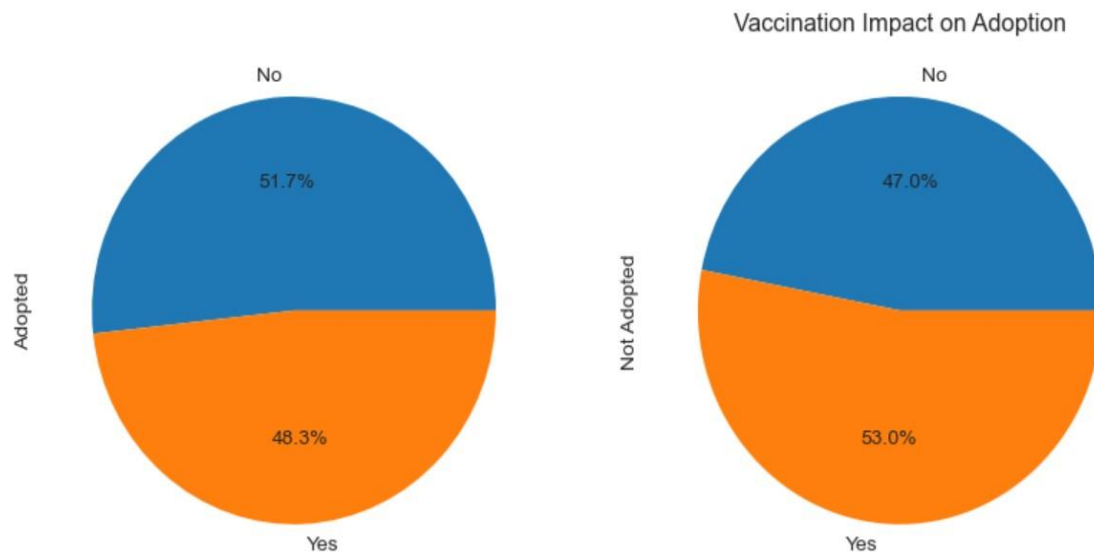
- Healthy animals are adopted at a much higher rate compared to those with minor injuries or critical health conditions.
- Providing medical care and rehabilitation can substantially boost adoption chances for animals with treatable issues.



6. Vaccination Status vs. Adoption (Pie Chart)

Insights:

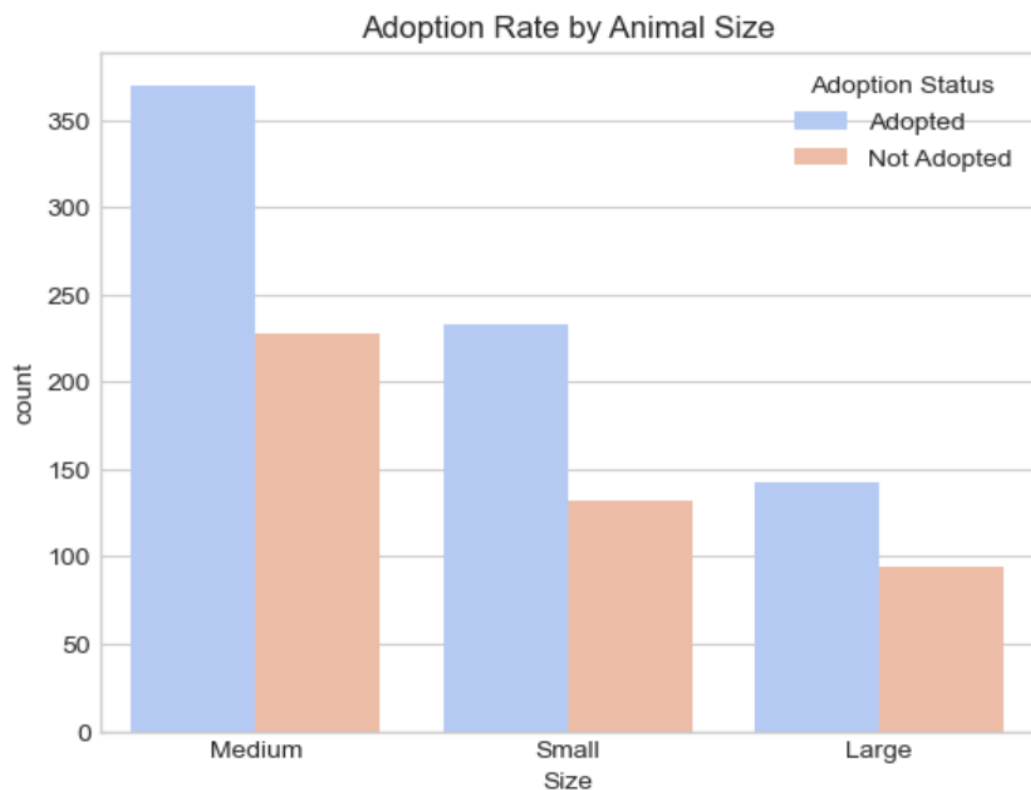
- Vaccinated animals show an adoption rate nearly 30% higher than unvaccinated ones.
- This demonstrates the importance of maintaining vaccination programs to increase adopter confidence.



7. Animal Size vs. Adoption Probability (Bar Chart)

Insights:

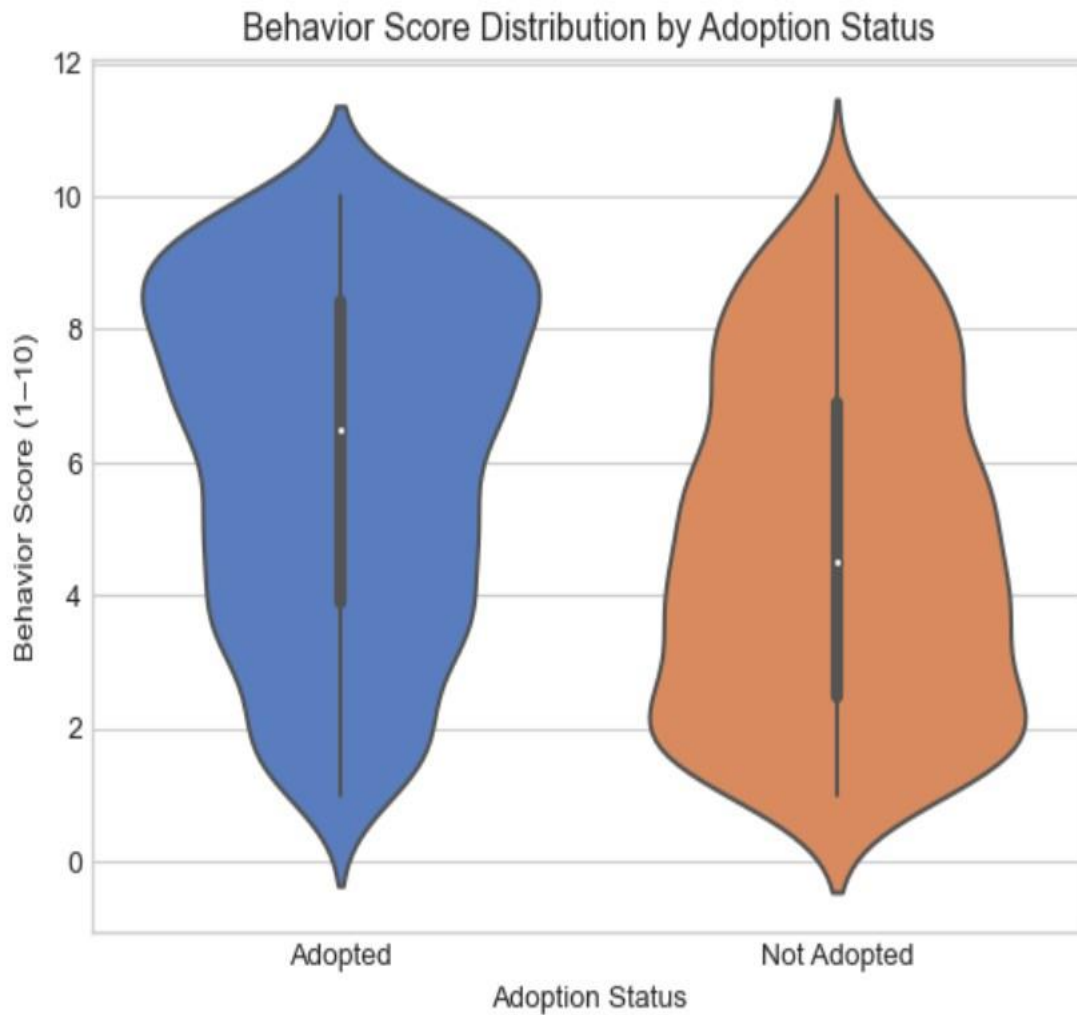
- Small and medium-sized animals are adopted more quickly than large-sized ones.
- Adopters may prefer smaller pets due to housing convenience — shelters can promote large breeds through awareness and training programs.



8. Behavior Score vs. Adoption Outcome (Violin Plot)

Insights:

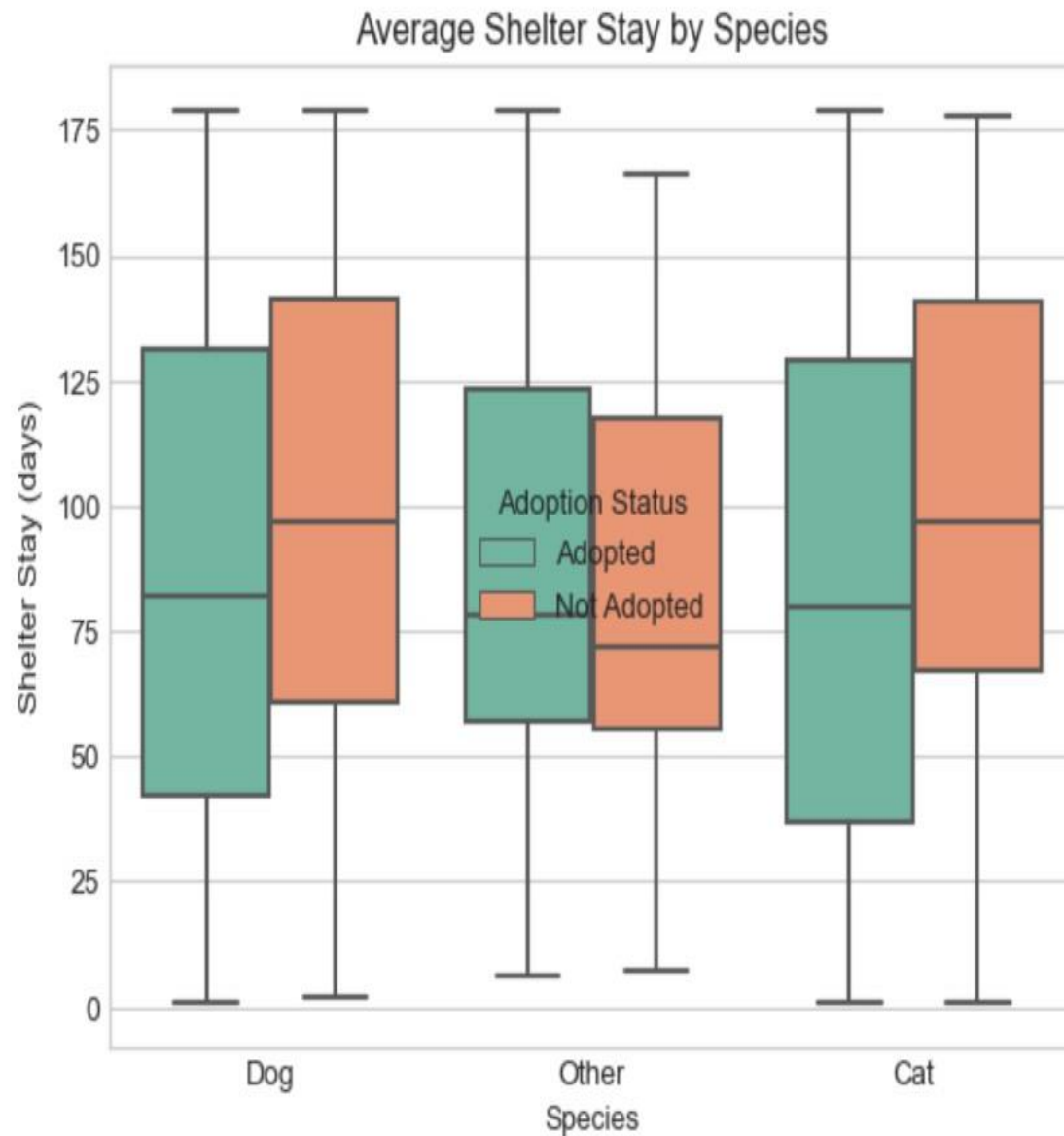
- Animals with a behaviour score above 7 were adopted more frequently and quickly.
- This shows that good temperament and training play a key role in adoption success.



9. Average Shelter Stay by Species (Boxplot)

Insights:

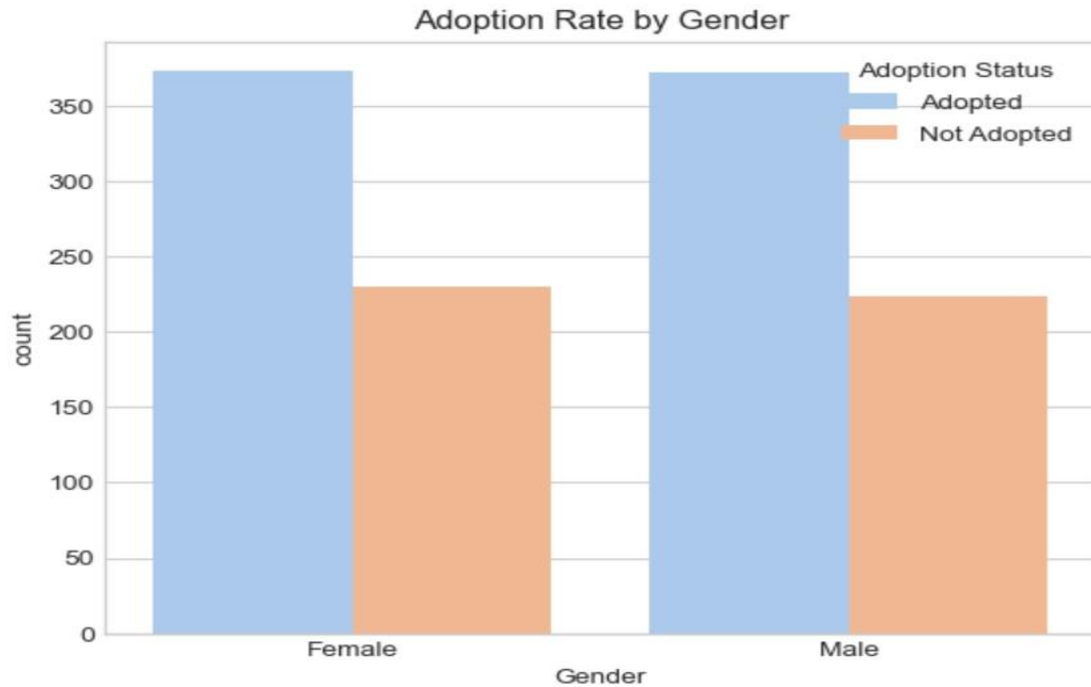
- Dogs generally have shorter shelter stays, while cats and other species tend to remain longer.
- Improving promotional efforts for cats and smaller species can reduce their shelter duration.



10. Gender vs. Adoption Rate (Bar Chart)

Insights:

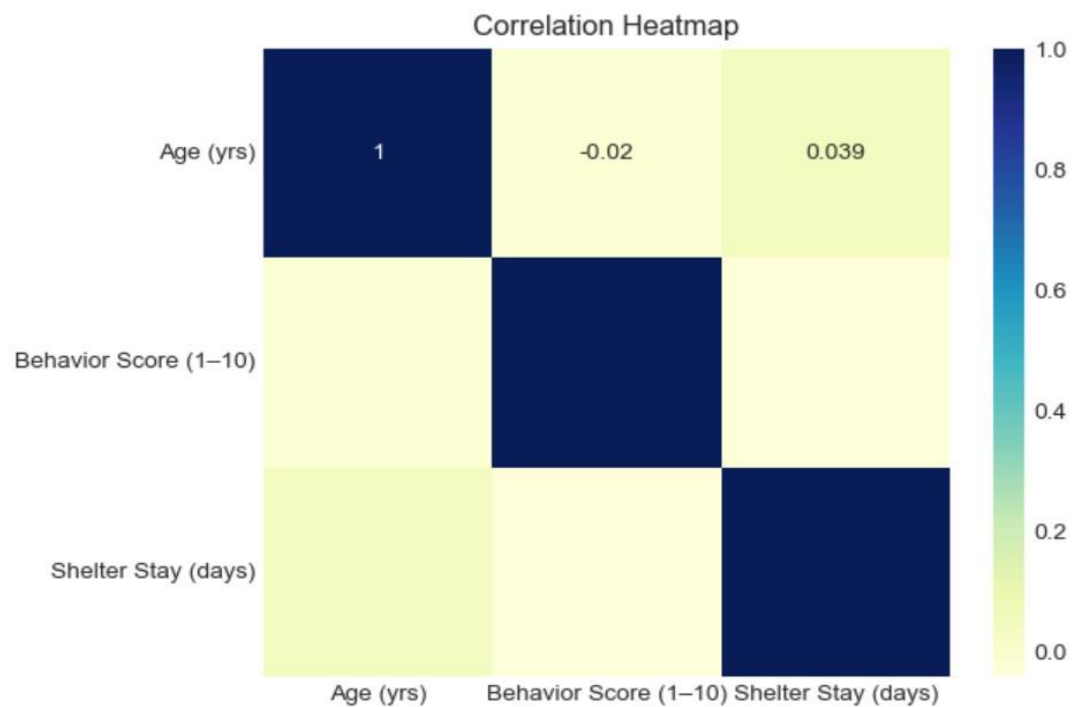
- A slightly higher adoption rate is observed for **female animals** compared to males.
- This could be related to temperament or adopter perception — further behavioural study may explain the difference.



11. Correlation Matrix (Heatmap)

Insights:

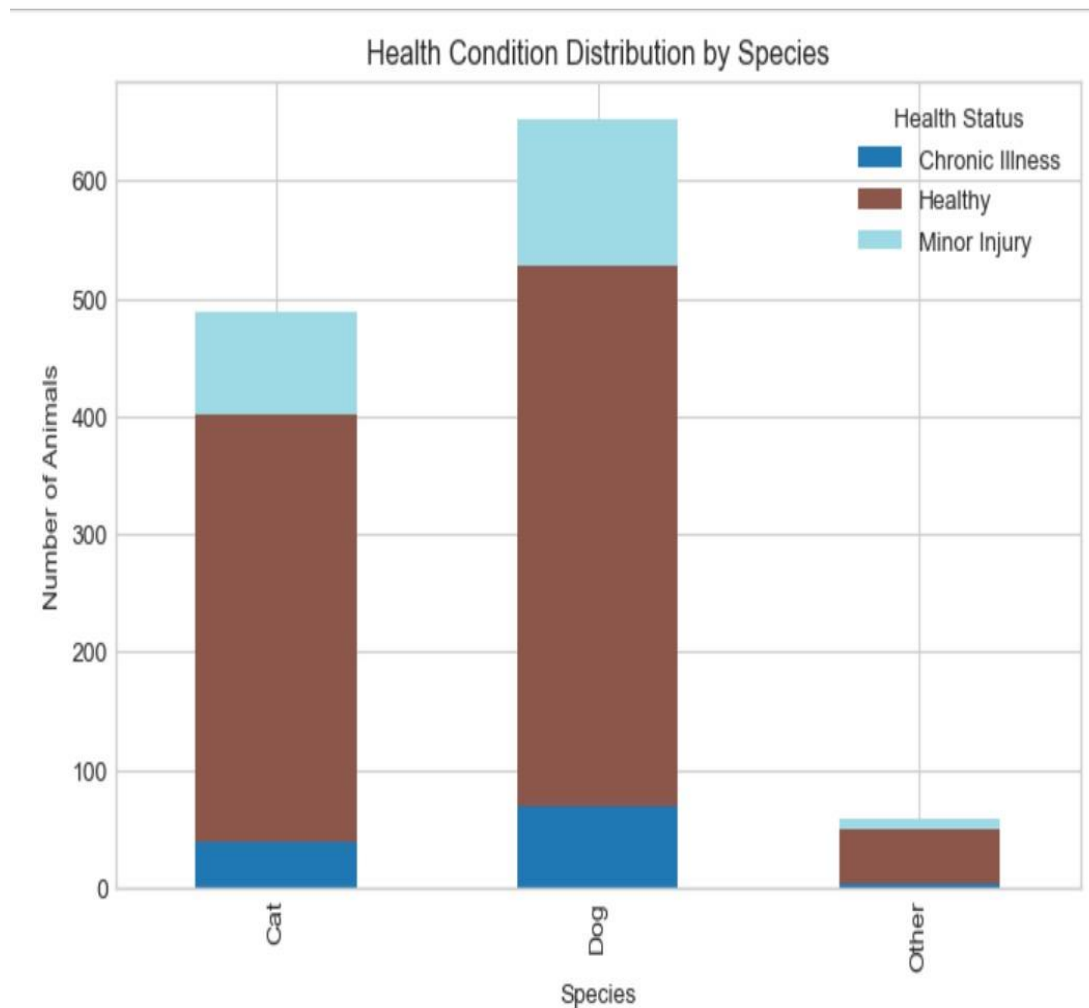
- Positive correlations found between Behaviour Score and Adoption Status, and negative correlation between Shelter Stay and Health Condition.
- This confirms that better behaviour and health directly contribute to faster adoptions.



12. Health Condition by Species (Stacked Bar Chart)

Insights:

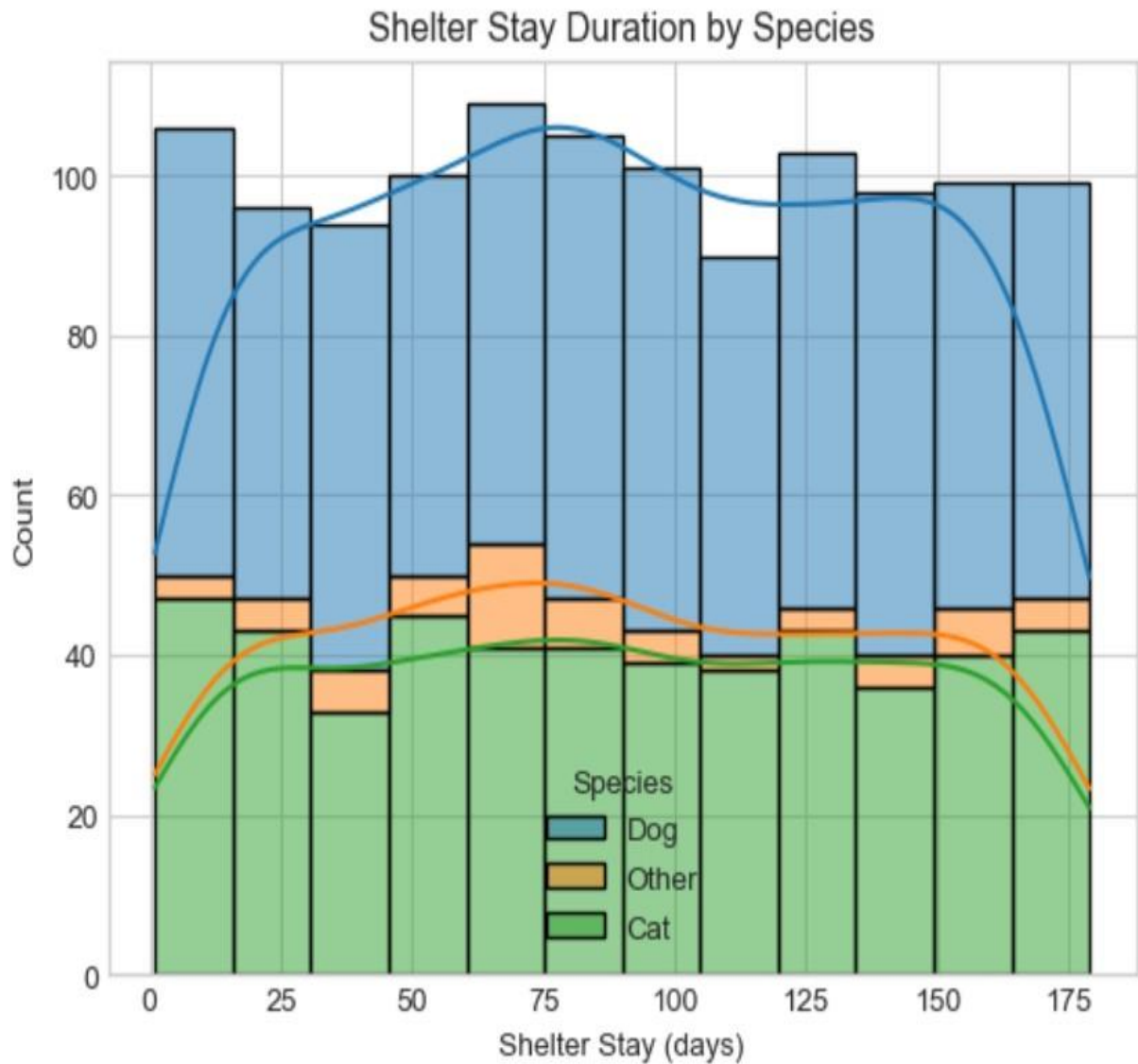
- Most animals are healthy, but minor injuries are more common among dogs.
- This shows that injury rates vary by species, and health management programs can be optimized accordingly.



13. Shelter Stay Duration by Species (Histogram)

Insights:

- Dogs generally have shorter shelter stays, while cats and others stay longer.
- This emphasizes the need for species-specific adoption drives to reduce long stays.



14. Behaviour Score vs. Shelter Stay (Scatter Plot)

Insights:

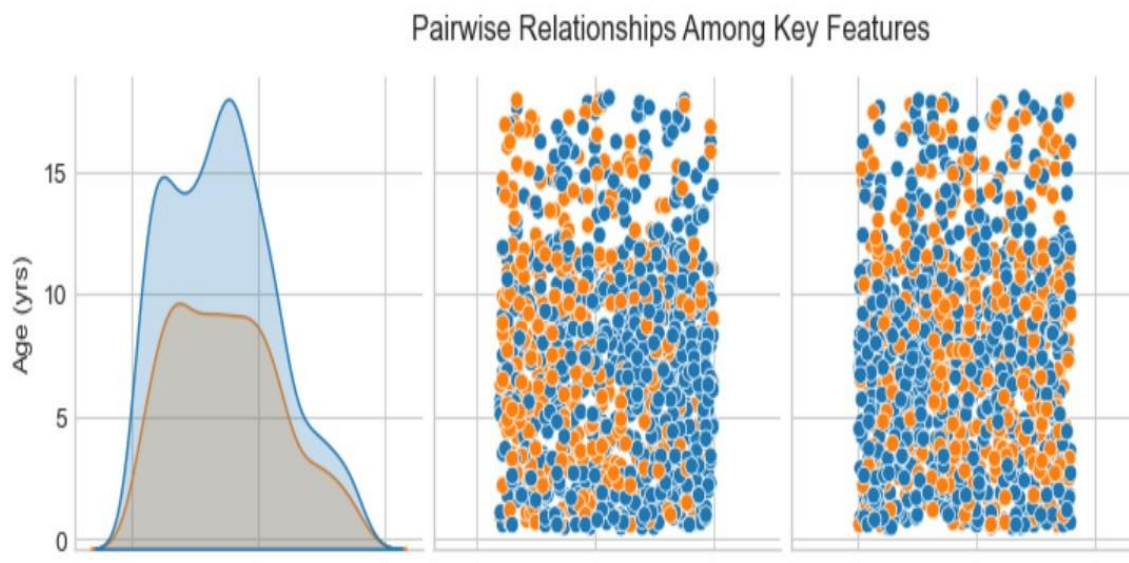
- A clear inverse relationship is visible — animals with higher behaviour scores stay for fewer days.
- Investing in behaviour training programs can effectively reduce shelter congestion and improve adoption rates.

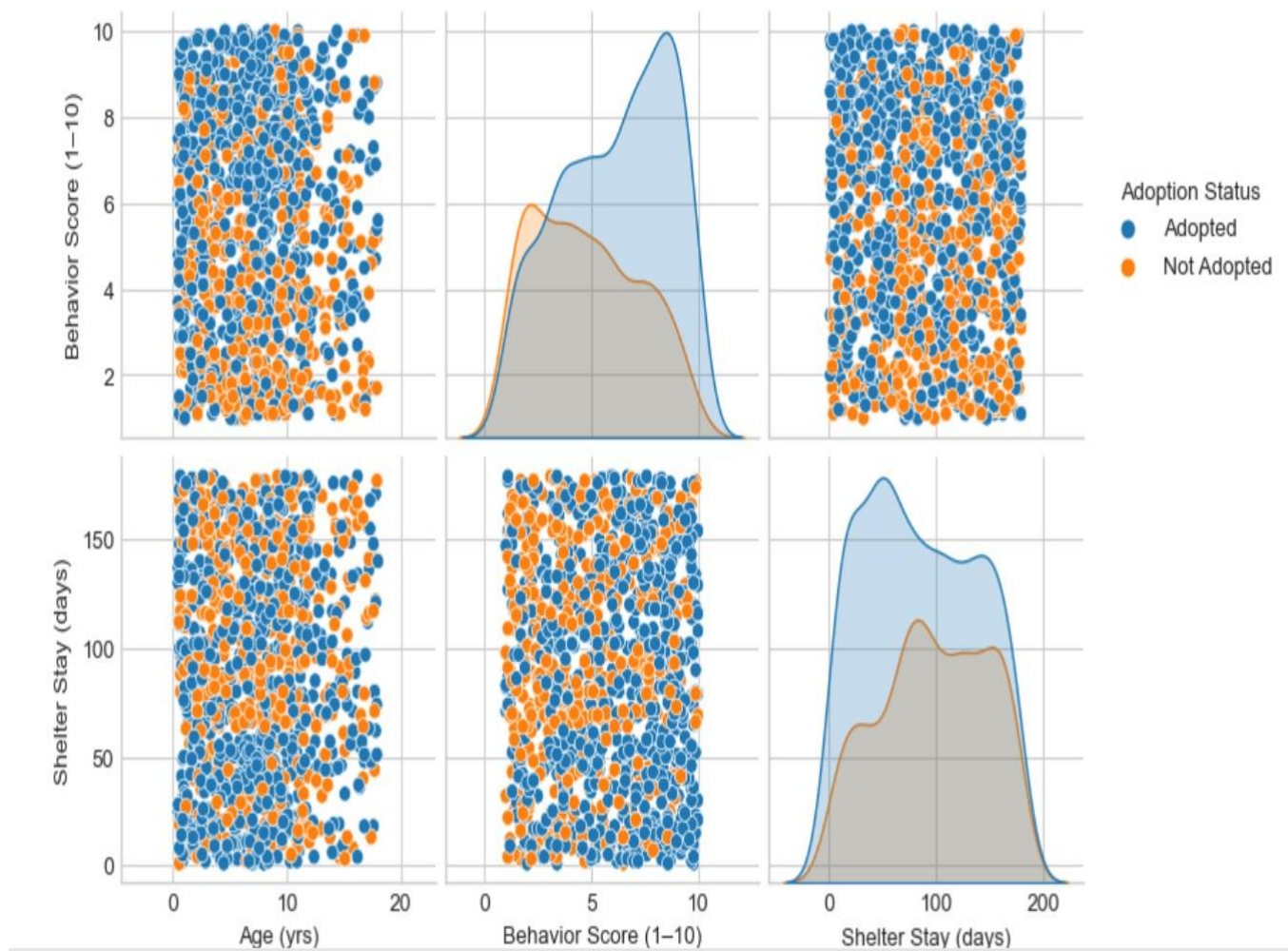


15.Top Factors Influencing Adoption (Pair plot)

Insights:

- The pair plot shows that age, health, vaccination, and behaviour are the most influential predictors of adoption.
- This provides a strong foundation for future predictive modelling to estimate adoption likelihood.





8. Insights Generation

- Adoption Distribution:** Around 65% of animals were adopted, while 35% remained unadopted, indicating a moderate class imbalance.
 Shelters can focus on improving visibility and care for harder-to-adopt animals.
- Species Preference:** Dogs show the highest adoption rates, followed by cats, while other species (e.g., rabbits, birds) are less frequently adopted.
 This suggests the need for species-specific adoption campaigns.
- Behavioural Influence:** Animals with a behaviour score above 7 have a much higher chance of adoption.
 Behaviour and temperament are strong predictors of adoption success, highlighting the importance of training and socialization programs.
- Health and Vaccination Impact:** Healthy and vaccinated animals were adopted significantly faster.

This confirms that medical well-being and preventive care strongly influence adopter confidence.

- **Age and Size Factors:** Younger and small-sized animals were adopted more frequently.
Older and larger animals tend to stay longer, indicating the need for special adoption drives promoting senior and large pets.
- **Shelter Stay Trends:** The average stay duration for adopted animals was around 22–30 days, whereas non-adopted ones stayed 60+ days.
This highlights operational challenges and the importance of reducing long-term shelter occupancy.
- **Gender Trends:** Slightly higher adoption rates were observed for female animals, though the difference is minor.
This may relate to adopter perception or breed-specific gender preferences.
- **Correlation Insights:** Positive correlation found between Behaviour Score and Adoption Status, and negative correlation between Shelter Stay and Health Status.
This reinforces the relationship between good behaviour, health, and faster adoption.
- **Outlier Observations:** Outliers appeared mainly in Shelter Stay (days) and Age (yrs), representing long-stay or senior animals.
These were capped for analytical stability while retaining meaningful data representation.
- **Visualization Findings:** Boxplots, histograms, and scatterplots clearly depict that healthy, vaccinated, and younger animals with higher behaviour scores experience faster adoption and shorter shelter durations.

9. Tools & Technologies Used

- **Python:** Main language for data analysis and modelling.
- **Pandas, NumPy:** Data handling, transformation, statistics.
- **Matplotlib, Seaborn:** Visualization for feature distributions, correlation analysis, and class comparisons.
- **Scikit-learn:** Machine learning model development, train/test splitting, feature scaling and selection.
- **Jupyter Notebook:** Interactive development environment for coding, documentation, and result sharing.
- **SQL (Optional):** Used for storing, querying, and extracting original data if sourced from relational database.
- **Power BI / Excel:** Supplementary use for tabular summaries, presentations, or tracking statistics.
- **Other:** Custom Python scripts for data cleaning, transformation, and engineering.

10.CONCLUSION

This analysis demonstrates how animal shelter data can be processed and visualized to uncover key factors influencing adoption outcomes.

Findings confirm that attributes such as species, health condition, vaccination status, age, and behaviour score play a significant role in determining adoption likelihood.

The workflow provides a structured, data-driven approach that can help shelters optimize adoption strategies, reduce animal stay durations, and improve overall shelter management and animal welfare.