Pet Food Market Analysis and Prediction Based on Multiple Regression and Machine Learning Technology

Huawei Huang [1, #], Jiarui Huang [2, #], Ali Zulfiqar [3, *]

[1] School of Social Science, The University of Manchester, Manchester, United Kingdom

[2] School of mathematics and statistics, Nanjing University of Science and Technology, Nanjing, China

[3] Department of Mechanical Engineering, The University of Hong Kong, Hong Kong China

* Corresponding Author Email: zulfiali@hku.hk

[#]These authors contributed equally

Abstract. This study focuses on the pet food market, aiming to analyze the development trend and influencing factors of the Chinese and global markets, and make accurate forecasts. For the Chinese market, we collected historical data and used linear regression and multiple regression models incorporating social factors to analyze changes in market size and pet number in the past five years and forecast trends in the next three years. In the global market research, data conversion, feature extraction, random forest model and other technologies are used to predict the demand and market size of pet food in different countries by combining population characteristics and pet consumption data. The study found that the scale of China's pet market will grow but the growth rate will be slow, the number of cats will increase, and the number of dogs will be relatively stable; Global markets have performed differently. This study provides decision support for enterprises and policy makers, helps the industry cope with market changes and policy adjustments, and explores sustainable development paths.

Keywords: Pet food market; Multiple regression; Machine learning; Market forecast; Sustainable development.

1. Introduction

Under the background of today's economic and social development, the pet industry is booming, and the pet food market has become a research hotspot [1-3]. With the improvement of residents' living standards and changes in family structure, pet raising has become increasingly popular. Globally, the pet market is also showing strong momentum [4-6].

Previous studies have mostly focused on one aspect of the pet market, for example, some studies only focused on simple statistical analysis of pet population growth or market size. In this study, multiple regression and machine learning techniques are integrated to explore the Chinese and global pet food market [7-10]. By integrating multi-source information such as historical data, social factors and demographic characteristics, a systematic analysis and prediction model is built to make up for the shortcomings of existing research in comprehensiveness and depth, aiming to provide a more scientific and accurate basis for corporate strategy formulation and policy decision-making in the pet food industry, and help the industry achieve sustainable development in a dynamic market environment.

2. Research problem

2.1. Study on the dynamic change trend of market size and pet quantity in China's pet industry

It is necessary to analyze the changes in the market size and number of pets in China's pet industry in the past five years, focus on distinguishing pet types such as cats and dogs, combine social factors such as youth, women and single people's pet preferences, and predict the market size in the next

three years based on key development factors. It involves data collation, trend and factor analysis and prediction calculation, providing scientific prediction for the development of the industry.

2.2. Analysis and forecast of regional differences in global pet food market demand and size

It is necessary to analyze the development trend of the global pet food market, collect data on the number of global pets and market size over the years, and pay attention to the changes in the number of cats and dogs and the impact of social factors such as youth and women on market demand. Construct a multiple regression model, including social factors and historical pet data, to predict the market size and trend in different regions and countries in the next three years.

3. Results

3.1. Research on the dynamic change trend of pet industry market size and pet number in China based on linear regression and social factors

3.1.1 Terms, Definitions and Symbols

The terms, definitions, and symbols of this section are shown in Table 1

Table 1. Terms, Definitions and Symbols 1

Symbols	Instruction
t	year，represents the time
M(t)	The market size of pet industry in Year t (unit: RMB 100 million))
C(t)	Number of pet cats in year t (unit: 10,000)
D(t)	Number of pet dogs in year t (unit: 10,000)
Y (t)	Percentage of young people's preference for pets in year t (unit: %)
F (t)	Percentage of female preference for pets in year t (unit: %)
S(t)	Percentage of single people's preference for pets in year t (unit: %)
$\hat{M}(t)$	Predicted market size of the pet industry in year t
$\hat{C}(t)$	Predicted number of pet cats in year t
$\hat{D}(t)$	Predicted number of pet dogs in year t t

3.1.2 Assumptions

• The growth trend of market size is linear and influenced primarily by pet population, consumption levels, and social trends.

• The growth trend of pet populations (cats and dogs) follows historical patterns.

• Market size is linearly correlated with pet population and social factors such as youth, female, and single preferences.

• Input data is accurate and representative, unaffected by anomalies or sudden events.

3.1.3 The Foundation of Model

When constructing the prediction model for the market size of the pet industry, two primary steps are involved: data preprocessing and regression models election. The model consists of the following parts: a linear regression model for market size, linear regression models for pet populations, and a multivariate regression model incorporating social factors.

(1) Data Collection and Preprocessing:Collect historical data on market size, pet populations (cats and dogs), and social preference factors (preferences of youth, women, and single individuals). The data is sorted by year, with missing values handled and unit conversions made as needed.

(2) Linear Regression Model for Market Size: Based on historical market size data, a linear regression model is established. It is assumed that market size M(t) grows linearly with time t:

$$\hat{M}(t) = \alpha_1 t + \beta_1 \quad (1)$$

where $\alpha 1$ and $\beta 1$ are model parameters solved using the least squares method.

(3) Linear Regression Models for Pet Populations:Separate linear regression models are constructed for cat and dog populations, assuming their respective numbers grow linearly with time:

$$\hat{\mathbf{C}}(\mathbf{t}) = \alpha_2 \mathbf{t} + \beta_2 \tag{2}$$

$$\widehat{D}(t) = \alpha_3 t + \beta_{3\leftarrow} \tag{3}$$

where $\alpha 2$, $\beta 2$ and $\alpha 3$, $\beta 3$ are parameters.

(4) Multivariate Regression Model Incorporating Social Factors:By considering the preferences of youth, women, and single individuals for pets, it is assumed that market size is linearly related to these factors. A multivariate regression model is established as follows:

$$\widehat{\mathrm{M}}(\mathrm{t}) = \alpha_4 \mathrm{Y}(\mathrm{t}) + \beta_4 \mathrm{F}(\mathrm{t}) + \gamma_4 \mathrm{S}(\mathrm{t}) + \delta_4 \tag{4}$$

where $Y(t)$、$F(t)$ and $S(t)$ represent the proportions of youth, women, and single individuals, respectively, with $\alpha 4$ $.\beta 4$ $.\gamma 4$ and $\delta 4$ as model parameters.

3.1.4 Solution and Result

Using the scikit-learn library in Python, the models were trained and evaluated. The steps are as follows:

(1) Training the Market Size Prediction Model:A linear regression model was trained using historical market size data, and predictions for the next three years were obtained.

(2) Training the Pet Population Prediction Models:Separate linear regression models were built for cat and dog populations using historical data to predict changes in population trends over the next three years.

(3) Training the Market Size Model Incorporating Social Factors:Social trend factors (preferences of youth, women, and single individuals) were introduced to predict market size through a multivariate regression model, better reflecting the influence of social trends on market size.

(4) Model Evaluation:Metrics such as mean squared error (MSE) and(R2)were calculated to evaluate the prediction accuracy and fit of each model.

(5) Visualization of Results:Using the Matplotlib library, the trends and predictions for market size and pet populations were plotted for an intuitive representation of model performance.

1) Predicted Results for Market Size

Figure 1 shows the historical data and predicted data for the market size of China's pet industry. Over the past decade, the market size has demonstrated a clear upward trend, especially with accelerated growth after 2016. Predictions indicate that the market size will continue to grow steadily over the next four years (2023–2026). When additional factors are considered, the growth rate slows slightly, but the overall trend remains upward.

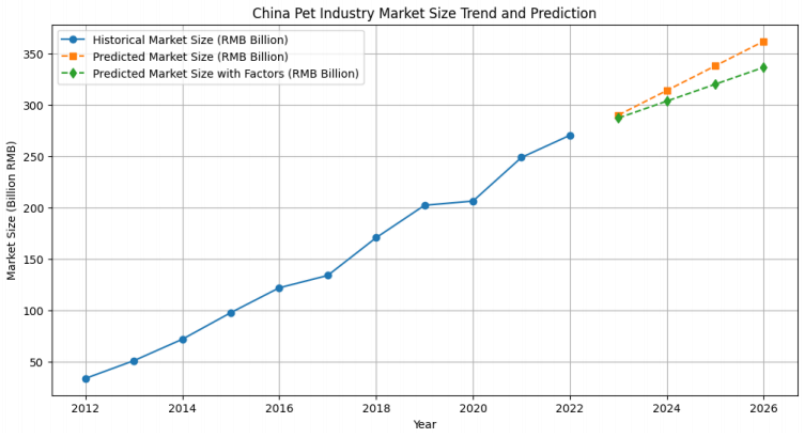


Figure 1. Pet industry market size trend and forecast in China

2) Predicted Results for Pet Populations

Figure 2 illustrates the historical and predicted data for the populations of cats and dogs in China. It can be observed that the number of cats has continued to grow since 2019 and is expected to reach nearly 100 million by 2026. In contrast, the dog population has remained relatively stable, with slight fluctuations in 2020, and is predicted to decrease slightly in the coming years. These changes may be attributed to factors such as lifestyle, economic conditions, and preferences for different types of pets.

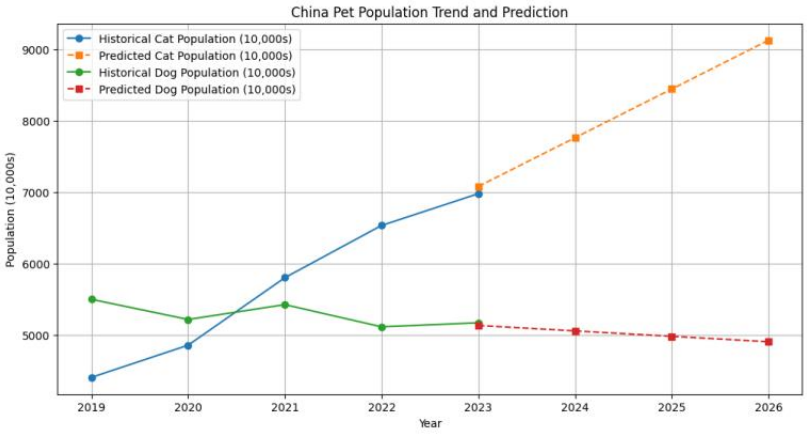


Figure 2. Pet population trend and forecast in China

3)Visual Comparison of Predictions

Figure 3 provides a comparative analysis of predicted data, including market size and pet population trends. The pet market in China is expected to continue expanding, with a more pronounced growth in the cat segment. In contrast, the dog population is predicted to gradually decrease while maintaining a high level overall. Predictions incorporating additional factors show slightly lower growth rates than those based solely on historical data, indicating that socioeconomic factors have a certain impact on market size.

	Year	Predicted_Market_Size	Predicted_Market_Size_with_Factors	\
0	2023	290.243636	287.425	
1	2024	314.228182	303.885	
2	2025	338.212727	320.345	
3	2026	362.197273	336.805	

	Predicted_Cat_Population	Predicted_Dog_Population
0	7081.2	5137.8
1	7762.2	5061.9
2	8443.2	4986.0
3	9124.2	4910.1

Figure 3. Comparison of predicted data

3.1.5 Analysis of the Result

From the above analysis, we can find that the past five points of China's pet industry have shown a stable and sustained growth trend. In addition, according to our forecasting model, we can conclude that the pet industry will continue to grow in China.

As the quality of life of the Chinese people has improved in recent years, people's disposable incomes have also risen dramatically, so people are more willing to spend their money on leisure and entertainment. Keeping pets is a good example, pets can bring pleasure to people, but also can be accompanied when people are lonely. Pet owners, mainly educated and affluent 80s and 90s-born individuals, dominate over 0.7 of the market, pushing demand beyond the traditional pet industry. Therefore, we can see that in the overall environment in China, the pet industry will have great potential and prospects for development in the future.

Based on our guesses and surveys, we analyze that in China, there are several reasons why more people own cats than dogs. This is because cats are better suited to city life, with limited space in apartments, lower activity needs and independence, and relatively less daily management and expense. Additionally, cities have more restrictions on dog ownership, whereas the barriers to cat ownership are lower. The quiet nature of cats and the popularity of social media have also contributed to more people choosing to own cats, especially young people and empty nesters.

3.1.6 Strength and Weakness

```
Market Model MSE: 43.734471074362325
Market Model R^2: 0.9924547849236229
Cat Model MSE: 14053.36
Cat Model R^2: 0.9850746306700294
Dog Model MSE: 10834.220000000952
Dog Model R^2: 0.5153740588588507
Factor Model MSE: 5.94050000000013
Factor Model R^2: 0.9952957057367919
```

Figure 4 Model evaluation index

Figure 4 shows the evaluation metrics for each model:

The market size prediction model exhibits a low mean squared error (MSE) and an R^2 value close to 1, indicating high prediction accuracy.

The cat population prediction model has an R^2 value close to 1, suggesting a good fit and high stability. dog population prediction model has an R^2 value of 0.515, indicating moderate performance and lower prediction accuracy.

The market size prediction model that incorporates social factors has an R^2 value close to 1, demonstrating the significant influence of social factors on market size.

3.2. Data transformation and random forest model are used to analyze and forecast the regional differences in global pet food market demand and size

3.2.1 Terms, Definitions and Symbols

The terms, definitions, and symbols of this section are shown in Table 2.

Table 2. Terms, Definitions and Symbols 2

Symbols	Instruction
$Cat^{c,\ y}_{pop}$	Number of cats in country c in year y
$Dog^{c}_{po}{}^{y}_{p}$	Number of dogs in country c in year y
$YPR^{c,\ y}$	Young population ratio in country c in year y
$FPR^{c,\ y}$	Female preference rate in country c in year y
$MPR^{c,\ y}$	MBTI preference rate in country c in year y
$Revenue^{c,\ y}$	Total pet food revenue in country c in year y
P_{cat}	Average annual pet food expenditure per cat
P_{dog}	Average annual pet food expenditure per dog
$n_{estimators}$	Number of decision trees
max_depth	Maximum depth of trees
min_samples_split	Minimum number of samples required to split a node
min_samples_leaf	Minimum number of samples required to be at a leaf node

3.2.2 Assumptions

The trends in pet ownership and pet food demand in different regions will remain stable in the coming years, with minimal annual fluctuations.

The demographic characteristics selected (e.g., young population ratio, female preference rate, MBTI preference rate) are significant predictors of pet population and pet food demand.

The average annual expenditure per pet (cat or dog) will remain constant over the next three years, simplifying revenue predictions.

3.2.3 The Foundation of Model

(1) Feature Selection and Data Preprocessing

Data Transformation:Annual data is converted into monthly data using linear interpolation to better capture seasonal fluctuations and other influencing factors. Seasonal adjustments are made for specific months to reflect actual market changes.

Feature Extraction: Feature Extraction: New features are generated from time-series data, including:

–Temporal features: Month (1 to 12), quarter (1 to 4).

–Demographic and preference features: young population ratio, female preference rate, MBTI preference rate.

For predicting the future populations of cats and dogs (and thereby pet food demand), the following demographic features are used as model inputs:

$$X = \{YPR^{c,y}, FPR^{c,y}, MPR^{c,y}\} \tag{5}$$

Target variables:

$$Y_{cat} = Cat_{pop}^{c,y} \tag{6}$$

$$Y_{dog} = Dog_{pop}^{c,y\leftarrow} \tag{7}$$

Data is split into training (80%) and testing (20%) sets to evaluate model generalization.

(2) Random Forest Model Construction

To predict pet populations, random forest regression models are used for their flexibility and ability to capture non-linear relationships between features and target variables.

Model parameters are optimized via grid search over the following hyperparameter space:

$$\mathbf{n}_{\text{estimators}^{\cdot\cdot}} =\cdot \{100,\cdot 200,\cdot 300\} \tag{8}$$

$$\max\_depth\cdot =\cdot \{\text{None},'10,\cdot 20,\cdot 30\} \tag{9}$$

$$\text{min\_samples\_split.} =.\{2,\cdot 5,\cdot 10\}^{\leftarrow} \tag{10}$$

$$\text{min\_ samples\_leaf}\cdot =\cdot \{1,2,4\} \tag{11}$$

(3) Revenue Prediction Model

After predicting the populations of cats and dogs for 2024–2026, pet food revenue is calculated as follows:

$$\text{Revenue}^{c,y} =\cdot P_{cat}Cat_{pop}^{c,y} +\cdot P_{dog} \cdot Dog_{pop}^{c,y} \tag{12}$$

Where $Pcat$ and $Pdog$ are the average annual expenditures per cat and dog, respectively, assumed to be constant.

3.2.4 Solution and Result

To address the nature of the problem, regression models suitable for time series forecasting are selected, such as Random Forest (RF), Support Vector Regression (SVR), and Gradient Boosting Trees (GBT). Grid search is employed to optimize model hyperparameters, and cross-validation (e.g., K-fold) is used to evaluate performance. The steps are as follows:

(1) Training Market Size Prediction Models:Using the historical data of the market size, the regression model (such as linear regression and random forest regression) in scikit-learn is trained to obtain the forecast results of the market size in the next three years.

(20 Training Pet Population Prediction Models:Historical data on cat and dog populations is used to build regression models (e.g., SVR, Gradient Boosting Trees), predicting changes in pet populations over the next three years.

(3) Training Market Size Models Incorporating Additional Factors:Social trend factors (preferences of youth, women, and single individuals) are included to predict market size using multivariate regression models, reflecting the impact of social trends.

(4) Model Evaluation:Metrics such as mean squared error (MSE) and(R2)are calculated to evaluate prediction accuracy and model fit. Cross-validation ensures generalization.

(5) Result Visualization:Using Matplotlib，historical and predicted data for market size and pet populations are visualized to intuitively demonstrate model performance.

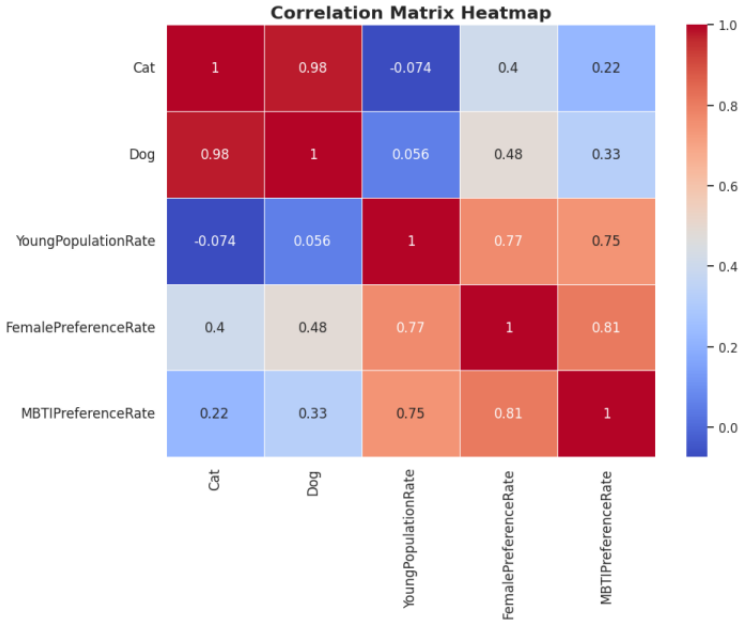


Figure 5. Correlation matrix heat map

Figure 5 shows the correlation between different variables. The darker the color, the higher the correlation, and the color from blue to red represents the range of correlations. The following is a specific analysis and summary of the correlation between each variable:

•Correlation between the number of cats and dogs: There is a strong positive correlation between the number of cats and the number of dogs, i.e., when the number of cats increases, the number of dogs also tends to increase.

•Correlation of cat and dog numbers with other variables:

–Ratio to young population: There is a strong positive correlation between the number of cats and the number of dogs, i.e., when the number of cats increases, the number of dogs also tends to increase.

–Ratio to young population: The number of cats is weakly negatively correlated with the proportion of young people, and the number of dogs is weakly positively correlated with the proportion of young people, suggesting that countries with a higher proportion of young people are likely to have more dogs but fewer cats.

–Preference rate with MBTI: There is a weak positive correlation between the number of cats and dogs and MBTI preference rates, with countries with high MBTI preference rates likely to have slightly more cats and dogs.

•Correlations between variables:

–Proportion of young population and female preference rate: There is a strong positive correlation between the two, with countries with a higher proportion of young people generally also having higher rates of female preference

– Proportion of young population and MBTI preference rate: There is a strong positive correlation between the two, with countries with a higher proportion of young people generally having higher MBTI preference rates.

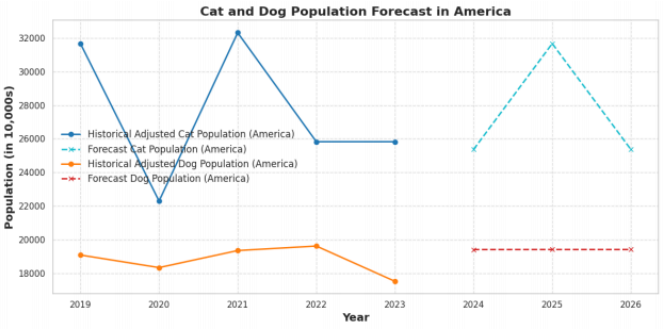
– Female preference rate and MBTI preference rate: There is a very strong positive correlation between the two, and countries with high rates of female preference generally also have higher rates of MBTI preference.

•There is a strong correlation between the number of cats and dogs.

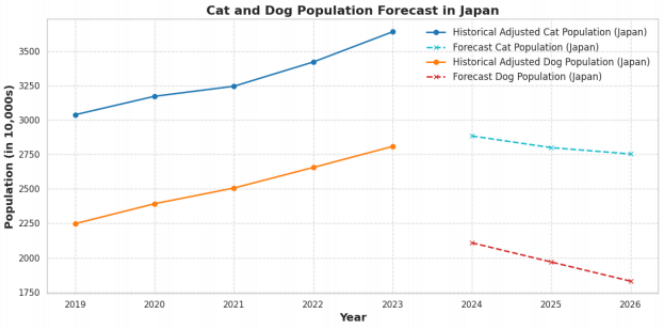
•There is a strong positive correlation between the proportion of young population, female preference rate and MBTI preference rate.

•The correlation between the number of cats and dogs and the proportion of the young population was weak or almost non-existent.

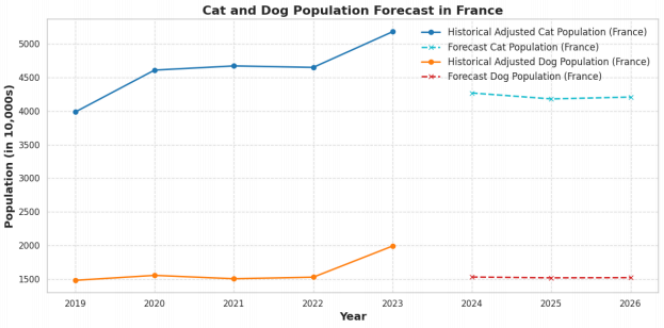
•There was a positive correlation between the number of cats and dogs and the female preference rate and the MBTI preference rate.



(a) Cat and Dog Population Forecast in America



(b) Cat and Dog Population Forecast in Japan



(c) Cat and Dog Population Forecast in France

Figure 6. Projected trends in cat and dog populations by country

Figure 6 analyzes historical and projected trends in cat and dog populations in the United States, Japan, and France.

•United States: The cat population declined sharply after peaking in 2019 and rose again in

2021 before remaining stable in 2022-2023. Projections show a significant increase in the cat population starting in 2024, peaking in 2025, but declining in 2026. The dog population is on the decline, especially in 2020, with projections showing a slight recovery and leveling off from 2024.

•Japan: The cat population has continued to rise since 2019, especially in 2021-2023, and is expected to continue to grow from 2024, but the growth rate is expected to slow down and reach a new high in 2026. The dog population also rose, but after peaking in 2023, projections show a decline from 2024 onwards.

•France: The cat population continues to grow in 2019-2023, with faster growth in 2021-2023, and is forecast to continue to grow after 2024, but at a slower rate, reaching a new high in 2026. The dog population is also on a slow upward trend, peaking in 2023 and projected to decline from 2024 onwards.

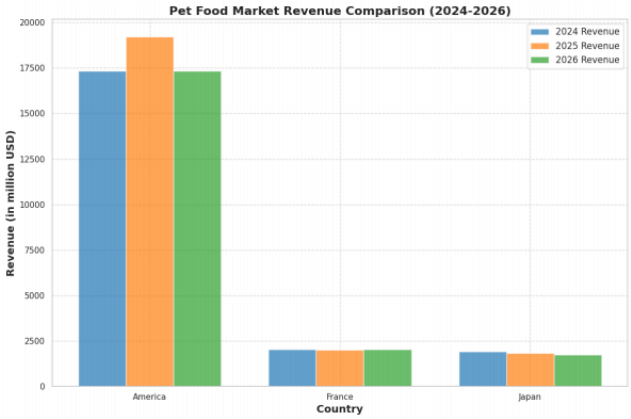


Figure 7. Comparison of pet food market revenue in 2024-2026

• America:

– In 2024: The pet food market in the United States is about $17.5 billion in revenue.

– In 2025: That figure grew to nearly $19 billion.

– In 2026: The pet food market in the United States is expected to further grow to approxi- mately $17.5 billion in revenue by 2026

• Japan:

– In 2024: The pet food market in Japan generates about $2 billion in revenue.

– In 2025: In the following two years, the figure fell slightly, to about $2 billion in 2025

– In 2026: Japan's pet food market revenue is expected to further decline to about $1.5 billion by 2026.

• France:

– In 2024: The pet food market in France is about $2.5 billion in revenue.

– In 2025: That figure remained largely unchanged for the next two years, at about $2.5 billion in 2025.

– In 2026: The pet food market in France is expected to decline slightly to about $2 billion in revenue by 2026.

Conclusion:Cat populations in all three countries have continued to grow over the past few years and are expected to continue to do so in the coming years, albeit at a slower rate. The dog population in the United States has been on a downward trend for the past few years, and is predicted to pick up slightly and level off in the future. Dog populations in Japan and France have risen in the past few years, but are expected to decline from 2024 onwards. Overall, the pet food market in the United States is much larger than in France and Japan, and has shown a steady growth trend over the past three years.

3.2.5 Analysis of the Result

All of these three countries are developed countries. In developed countries, people normally have more disposable income spending on leisure. The popularity but high cost of pets in these societies

reflects their strong affordability. The reason why the U.S. market is much larger than the other two countries in terms of market share is because the U.S. market is much larger than the other two countries and has a higher population than the sum of other two countries. Therefore, the U.S. has a greater potential for growth.

There may be several reasons why people prefer to have both types of pets. One is to have variety in the types of pets in one's family and to reduce the tedium of having a single pet. The second is that different types of pets can bring different feelings to people, but they can also get along very well together.

There is no clear link between the number of cats and dogs and the proportion of the population that is young, so we can presume that owning a pet is something that people of every age may enjoy, and that this preference does not belong to a particular group.

There is a positive correlation between the number of cats and dogs and female preference rates and MBTI preference rates. Psychologically speaking, girls are by nature more caring and compassionate, and at the same time they are able to be more attentive when caring for pets. People with different MBTIs also have different personality traits, so they have different attitudes toward pet ownership.

3.2.6 Strength and Weakness

```
R-squared for Cat (Overall): 0.97
R-squared for Dog (Overall): 1.00
Mean Squared Error Summary:
   Country       MSE_Cat     MSE_Dog
0  Overall  2231.271796  220.090392
R-squared Summary:
   Country    R2_Cat    R2_Dog
0  Overall  0.966876  0.997207
```

Figure 8. Model evaluation index

Fig 8 shows the evaluation metrics of the model:

• The R^2 of cat is 0.97, the R^2 of dog is 1.00, it shows that the model fits the historical data well, especially the dog data almost exactly match.

• The MSE of cat is 2231.27， the MSE of dog is 220.09， it shows that the model has a large error in predicting the number of cats, but the prediction of dogs is more accurate.

• When country factors are considered, R^2 value of cat and dog are 0.9688 and 0.9972, respectively, indicating that the model also has high accuracy in different country situations.

4. Conclusions

This study analyzes and forecasts the Chinese and global pet food market by constructing a variety of models. In the Chinese market, the growth trend and the changing trend of the number of pets are clarified, and it is found that social factors have an important impact on the market size. At the same time, the problems in nonlinear relationship processing and overfitting of the model are pointed out, and the improvement direction is given. The global market research reveals the differences of pet food markets in different countries and the role of related factors. In general, this study provides a valuable reference for the development of the industry, but the model still needs to be further improved to cope with the complex and changeable market environment. In the future, the research can be deepened in the aspects of technology integration and data expansion to promote the sustainable development of the pet food industry.

References

[1] Global pet industry forecast to grow[J]. Veterinary Record,2024,195(5):191-191.

[2] Du C, Dai X, Wang Z, et al.Plant-wide optimal scheduling of multi-grade PET production with time window constraints: A hybrid discrete/continuous-time optimization formulation[J].Computers and Chemical Engineering,2024,186108682-.

[3] Misun J, Kiseol Y, Maria H K, et al.Curation subscription box services: Implications for the pet industry[J].Journal of Retailing and Consumer Services,2024,76

[4] Edoardo T, Souza J P D, Samir B, et al.Multilayer plastic film chemical recycling via sequential hydrothermal liquefaction[J].Resources, Conservation & Recycling,2023,197

[5] Samantha K. Of companionship, curfew, and conflict: multispecies leisure in the age of COVID[J]. Leisure Studies,2022,41(3):301-309.

[6] Emma G, Heiner S. Bringing a governance perspective to plastic litter: A structural analysis of the German PET industry[J]. Sustainable Production and Consumption,2022,31630-641.

[7] Jianwen C, Ya Z, Yanpeng C, et al.A life-cycle perspective for analyzing carbon neutrality potential of polyethylene terephthalate (PET) plastics in China[J].Journal of Cleaner Production,2022,330

[8] Shu T, Hongrui T, Qingsong W, et al.Evaluation and optimization of blanket production from recycled polyethylene terephthalate based on the coordination of environment, economy, and society[J].Science of the Total Environment,2021,772145049-145049.

[9] Lia K, Emilio D, E H R, et al.An investigation of Salmonella Fluntern illnesses linked to leopard geckos-United States, 2018.[J].Zoonoses and public health,2019,66(8):974-977.

[10] E. M L, R. C F, Megin N. Time to get serious about antimicrobial stewardship in the commercial pet industry[J]. JAVMA-JOURNAL OF THE AMERICAN VETERINARY MEDICAL ASSOCIATION,2018,253(2):155-156.