| Team Number: | apmcm24207711 |
|---|---|
| Problem Chosen: | C |

2024 APMCM summary sheet

The paper mainly used Python to analyze and predict the development trend of pet markets in China and overseas, and focused on the development of pet food industry, aiming to provide strategies for the future development of pet food industry in China.

For problem 1, China's pet industry development was analyzed from two perspectives: (1) the structural changes in cat and dog populations, and (2) an evaluation system using three indicators (market size, household penetration rate, and investment amount). Then, six factors were tested using Pearson's correlation coefficient to assess their impact on industry development. Finally, Holt-Winter model and Linear Regression were applied to predict the three indicators of evaluation system for the next three years, indicating positive future development trends of China's pet industry.

For problem 2, overseas pet industry analysis followed a similar approach to problem 1, with total industry expenditure replacing the investment amount indicator. To forecast global pet food demand, historical sales data from 2010 to 2023 was collected to measure its demand. The Holt-Winter model predicted global pet food sales would reach 141.2, 150.0, and 163.7 billion dollars in the next three years.

For problem 3, China's pet food industry was analyzed through trends and structural proportions of production and export values. Using data from attachments 1 and 2 along with global pet food sales, a VAR model predicted future values of production and exports. Production is expected to reach 445.8, 431.1, and 761.7 billion dollars, while exports will reach 34.7, 63.1, and 65.2 billion dollars in the next three years.

For question 4, a DID model quantified the impact of overseas economic policies on China's pet food industry. The US, France, and Germany formed the treatment group, with Malaysia as the control group. By analyzing export data (volume and values) before and after recent economic suppression events, results revealed significant negative effects of these countries' policies on China's pet food exports.

The paper finally evaluated all the time series models used, proving their reliability. The paper stands out by assessing the pet industry's development through a diversified evaluation system and applying suitable time series models to predict future trends in different scenarios.

**Keywords:** Pet Industry Analysis   Pet Food Industry   Time Series   Python

# Contents

# I. Introduction

## 1.1 Question Background

On a global scale, the pet industry has shown significant growth, and its related industries, such as pet food, pet healthcare and pet supplies, play a positive role in promoting the economic development.

In recent years, many Chinese families have adopted pets such as cats and dogs and shown an upward trend, directly leading to the rapid development of pet food industry. Due to the larger and larger market environment, many other economically developed regions also show a positive industry development.

In China, the production and export of pet food have occupied a crucial position in the global pet industry. However, because of the changes in international economic policies such as tariff policies, the pet food industry in China may suffer from negative effects.

In this article, we will analyze and forecast the development of the pet industry and related industries in China and globally, and provide Chinese pet industry with development strategies.

## 1.2 Problem Restatement

Before delving into the solution of the problem, it is necessary to sort out and restate the core problem.

**Question 1:** Based on the two different pet types(cat and dog), collect and combine the relevant data and analyze the development of China's pet industry in the past five years. In addition, analyze the factors which affect the growth of China's pet industry, establish a mathematical model to forecast the next-three-years development of China's pet industry.

**Question 2:** Apart from the data in China's pet industry, collect the data of the development of the global pet industry by the pet types mentioned above and analyze its future global demand for pet food in the next three years by making a mathematical model.

**Question 3:** According to China's pet food production and export values in the recent five years, evaluate the recent development of China's pet food industry. And combine with the trend of global demand of pet food market and China's development, foresee its future production and export of pet food in the next three years.

**Question 4:** Conduct a mathematical model to quantitatively analyze the influence of the new foreign economic policies promulgated by the European countries and America, combining with the data collected and the calculation results. And give practical strategies for the development of China's pet food industry.

# II. Problem analysis

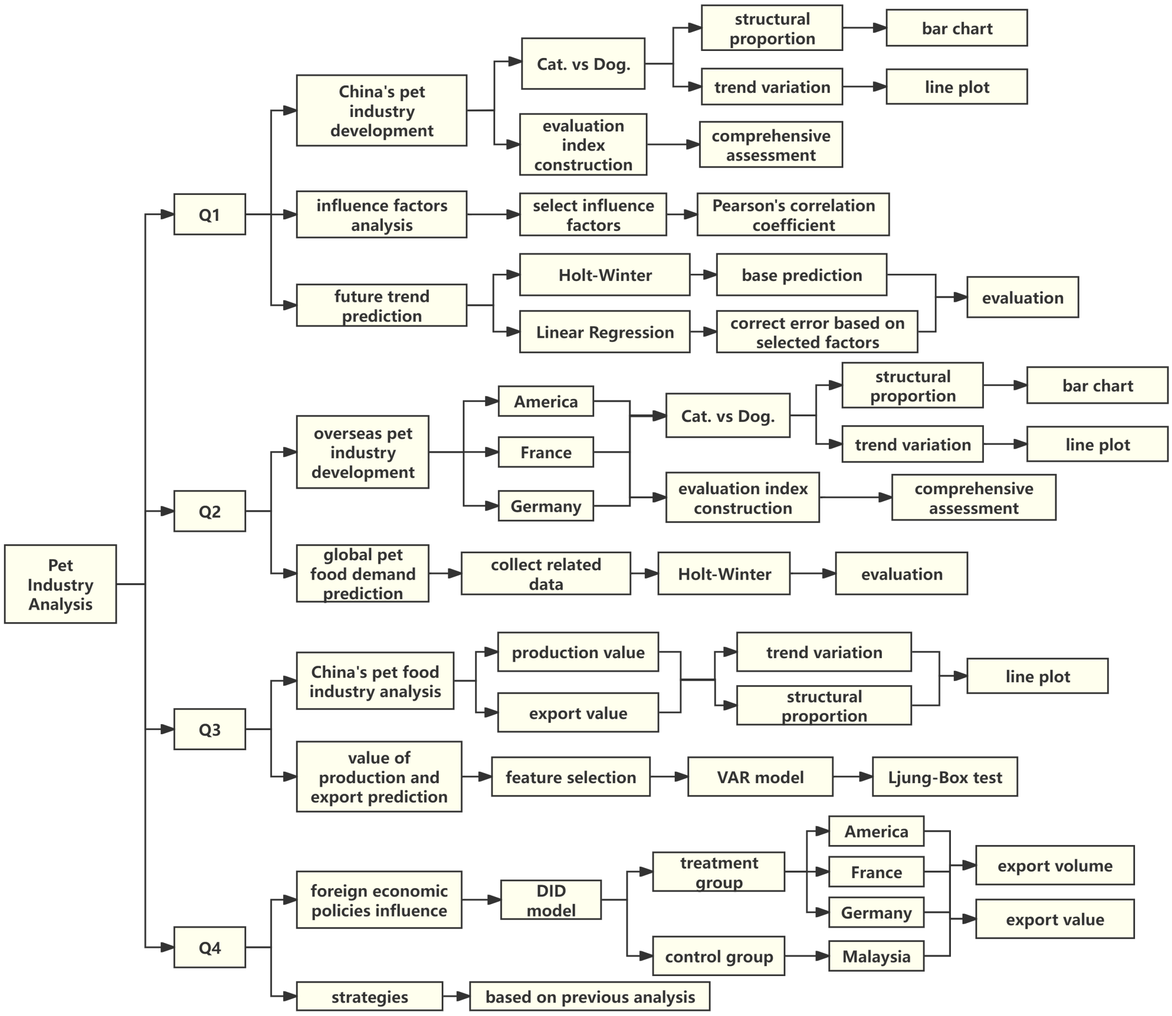


**Figure 1 Overall Process of Solution**

## 2.1 Question 1

The first problem is to analyze China's pet industry development and influence factors and predict future trends. The development of China's pet industry is comprehensively analyzed by the number of cats and dogs and the construction of an indicator system.

Calculate the annual growth rate of cats and dogs and draw the histogram and growth rate line chart of cats and dogs. Combined with the comprehensive evaluation indicators, the calculated comprehensive score can be more comprehensively analyzed. As for the analysis of its influencing factors, choose the most comprehensive influencing factors, and analyze which influencing factors are the most important by calculating Pearson's correlation coefficient. To predict the development trend, use Holt-Winter preliminary forecast and then use multiple linear regression to correct errors based on selected factors. Finally, evaluate the result.

## 2.2 Question 2

Analyze the development of the global pet industry and the future global demand for pet food in the next three years. We firstly visualize the number and structure proportion of cats and dogs in the attachment 2 in the last five years. In addition, we introduce additional indicators to make a comprehensive assessment of their development of pet industry which reflects the globe's. Finally, we use related data to establish a Holt-Winter model to predict future global demand for pet food in the next three years, with model evaluation.

## 2.3 Question 3

This question requires to make an analysis of the development of China's pet food industry and predict its production and export of pet food in the next three years. We regard the values of China's pet food production and export as indicators to assess the development of China's pet food industry and draw a line plot to show its trend variation and structural proportion. To perform values of production and export prediction, we select features to establish the VAR model and use Ljung-Box to test it.

## 2.4 Question 4

In this question, we need to quantitatively analyze the impact of economic policies in Europe, America and other countries on the development of the pet food industry, and provide strategies for the sustainable development of China's pet food industry based on the quantitative results. We firstly select a trade turning point event for France, the United States and Germany to evaluate the impact of the event on the development of China's pet food industry using the DID model. The processing group is the the previous three countries, and the control group is Malaysia. The quantitative data (pet food export volume and export value) are collected according to the time of the event, and then the impact of the policy is evaluated. Finally, based on the previous analysis, we provide strategies for the sustainable development of China's pet food industry.

# III. Model Hypothesis

## 3.1 Assumption 1

The pet industry discussed in our solution only considers two pet types: cats and dogs.

**Explanation:** According to the Apa report, cats and dogs account for 74% of the pet market and are the main objects of pet ownership [1]. More abundant and sufficient industry chain is related to cats and dogs. Although there are other pet types such as aquariums, reptiles, and birds, they create far less economic value than cats and dogs.

## 3.2 Assumption 2

There will be no major public health events in China or other countries in the next three years, and there will be no serious international conflicts between China and other countries.

**Explanation:** The COVID-19 pandemic has proved that major public health events can significantly affect the development of the pet industry. In addition, the stability of international relations is crucial to the normal operation of the pet industry chain. Only ignoring these possibility, the prediction can be reliable. The three-year forecast horizon is relatively short, and major disruptive events are unlikely to occur under normal circumstances.

## 3.3 Assumption 3

The development of pet industry in the United States, France and Germany reflects the development of pet industry in overseas countries.

**Explanation:** The U.S., as the world's largest pet market, along with France and Germany representing the European market, provide comprehensive insights into global industry trends and consumer behaviors in developed economies. Their combined market data serves as a reliable indicator for understanding international pet industry patterns.

## 3.4 Assumption 4

The exchange rate is set at EUR/USD = 1.1 and CNY/USD = 0.14.

**Explanation:** This exchange rate assumption is essential for standardizing financial data across different markets in our analysis. Using fixed exchange rates helps eliminate currency fluctuation impacts when comparing pet industry data between China, the United States, and European countries (France and Germany).

# IV. Notations

| Symbol | Description |
|---|---|
| $MSE$ | Mean squared error: MSE = $\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ |
| $RMSE$ | Root mean squared error: RMSE = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$ |
| $MAE$ | Mean absolute error: MAE = $\frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$ |
| $R^2$ | R-squared: $R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$ |
| $lb\_stat$ | Ljung-Box statistics: $Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$ |
| $lb\_pvalue$ | Ljung-Box's p-value: $p$-value = 1 − $\mathrm{CDF}_{\chi^2}(Q, \mathrm{df} = h)$ |

# V. Models and Solutions for Question 1

## 5.1 Model Establishment

### 5.1.1 *comprehensive Evaluation Model*

In this paper, we construct a comprehensive evaluation system, assign different weights to the indicators, and use the weighted sum method to calculate the comprehensive score to analyze the development status. The formula is as follows. In question 1, it is used to analyze the development of China's pet industry in the past five years. And in question 2, it is used to analyze the development of the global pet industry.

Suppose we have a set of data $x_1, x_2, \cdots, x_n$, and the corresponding weight is $w_1, w_2, \cdots, w_n$. Then the formula for the weighted sum is:

$$S = w_1x_1 + w_2x_2 + \cdots + w_nx_n = \sum_{i=1}^{n} w_ix_i \tag{1}$$

### 5.1.2 *Pearson Correlation Coefficient*

The Pearson correlation coefficient is a statistical measure used to reflect the degree of linear correlation between two variables. Its calculation formula is:

$$r_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2}$$

where $n$ is the number of data points (or observations) in both variables $X$ and $Y$. $x_i$ is the value of the $i$-th observation in variable $X$. $y_i$ is the value of the $i$-th observation in variable $Y$.

### 5.1.3 *Holt-Winters Exponential Smoothing Model*

The Holt-Winters model is an exponentially smoothed time series forecasting method for capturing trends and seasonal features. In this code, an Additive Model is used, without considering seasonal components. The formula is as follows:

**Smooth trend formula:**

$$l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \tag{3}$$

where $l_t$ is Smooth level value at the current moment. $y_t$ is current Actual Observations. $b_{t-1}$ is trends from the previous moment. $\alpha$ is smoothing parameters (update weights for control levels).

**Trend formula:**

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \tag{4}$$

where $b_t$ is updated values for current trends. $\beta$ is smoothing parameters (update weights that control trends).

**Forecast formula:**

$$\hat{y}_{t+h} = l_t + hb_t \tag{5}$$

where $\hat{y}_{t+h}$is $h$ Step predicted value. $h$ is predicted steps.

This model adapts to historical data and generates future predicted values by repeatedly optimizing the parameters $\alpha$ and $\beta$.

#### 5.1.4 *Multiple Linear Regression Model*

The multiple linear regression model is used to analyze the influence of multiple feature variables on the target variable and predict the future changes of the target variable.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \tag{6}$$

where $y$ is target variables (e.g. annual growth rate). $x_1, x_2, \ldots, x_k$ is characteristic variables (e.g. GDP, per capita income, etc.). $\beta_0$ is intercept term (indicates predicted value without feature). $\beta_1, \beta_2, \ldots, \beta_k$ is regression coefficients for features (representing the weight of each feature's impact on the target). $\epsilon$ is random error term.

### 5.2 Solution

#### 5.2.1 *China Pet Industry Analysis*

In order to more intuitively see the changes in the number of cats and dogs in China, after calculated the rate of change in the number of cats and dogs from 2019 to 2023, we created a stacked bar chart to visualize the number and growth rate of cats and dogs.

In order to analyze the development of China's pet industry, According to the 2024 Pet Industry Market Trend Change Report [2], we use three indicators, including Market Size, Household Penetration Rate and Number of Pet Track Investments and Financing in China, as Table 1 shown, to build a comprehensive evaluation system.

**Table 1 Chinese Pet Industry Evaluation Index**

| Indicator | Weight |
|---|---|
| Market Size | 0.4 |
| Household Penetration Rate | 0.3 |
| Number of Pet Track Investments and Financing | 0.3 |

The market size of China's pet industry is more relatively important for analyzing the development of the pet industry. Therefore, its weight is 0.4, and other indicators are 0.3.

After standardizing the data for these three indicators, the composite score and the Cat and Dog Market Share for each year from 2019 to 2023 was calculated according to their weights. In this way, we can analyze the development of the China's pet industry.

To analyze the influence factors of pet industry, we collected data from the China Economic and Social Big Data Research Platform [3]. We mainly frame a system of influence factors based on three types of factors: economic factors, population structure, and social development, as shown in Table 2

**Table 2 Influencing Factors System**

| Category | Factors |
|---|---|
| Economic Factors | GDP per capita (CNY) |
| | Disposable income per capita (CNY) |
| Demographic Structure | Change in national marriage registrations (10,000 couples) |
| | Proportion of population aged 65 and above (%) |
| Social Development | Urbanization rate (%) |
| | E-commerce transaction value (trillion CNY) |

After that, using the development score of China's pet industry and three evaluation indicators as target variables, the Pearson correlation coefficient is calculated and a heat map is drawn to visualize the data lake and evaluate which factor has the greatest impact on the industry.

Based on the number and trends of pet cats and dogs in China and the comprehensive evaluation score, the development of the country's pet industry over the past five years can be analyzed. A heat map of influencing factors is used to examine the key elements driving this development.

### 5.2.2 *Future Trend Prediction*

We forecast the development of China's pet industry in the next three years by analyzing the data of three indicators from 2019 to 2023. Since the data is time series data and exists a trend, we use the Holt-Winters exponential smoothing model to preliminarily predict the market size, penetration rate, and number of investment and financing for the next three years.

However, only five years of data is not accurate enough for prediction. We introduce a multiple linear regression model to correct the error. In the multiple linear regression model, we take six influencing factors mentioned earlier as features to predict the three indicators respectively. Finally, combined with Holt-Winters and Linear Regression results, we predict the market size, penetration rate, and investment and financing quantity in the next three years.

## 5.3 Results

### 5.3.1 *the Development of China's Pet Industry analysis*

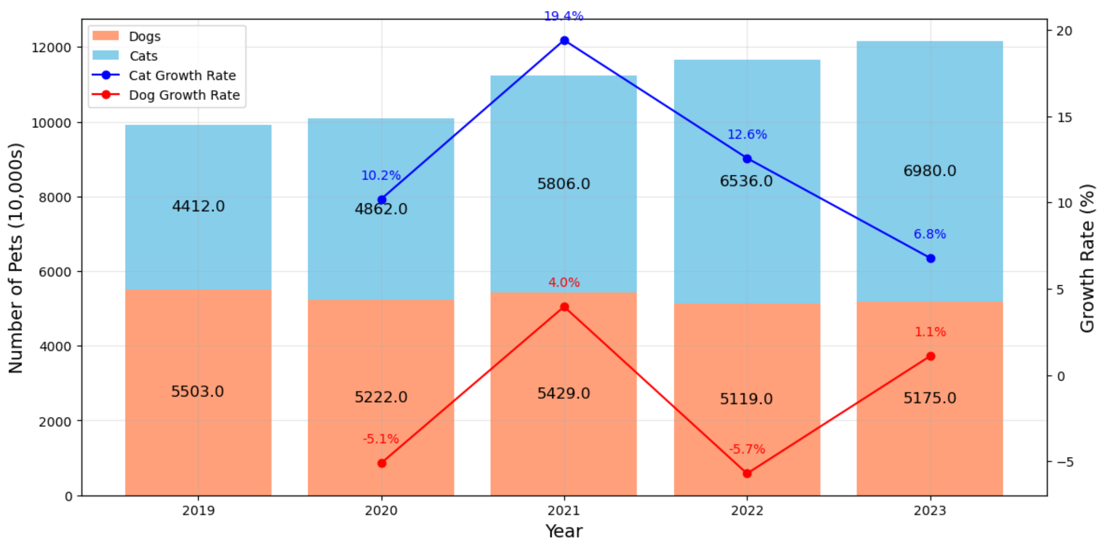


**Figure 2 The Growth Rate and the Number of Pets in China (2019-2023)**

As shown in Figure 2, from 2019 to 2023, the number of cats has increased significantly, but the number of dogs has stabilized and even declined. In 2021, the growth rate of both categories of pets is relatively high, with the growth rate of cats reaching 19.4% and dogs reaching 4%. However, the sum of the number of cats and dogs will increase year by year in 2019-2020.

As shown by Figure 3, the comprehensive score calculated according to the three indicators is basically increasing year by year, except that the comprehensive score drops to 0.113 in 2020. The upward trend of the cat's market share is also obvious, rising from about 0.45 to nearly 0.58 in four years. On the contrary, the market share of pet

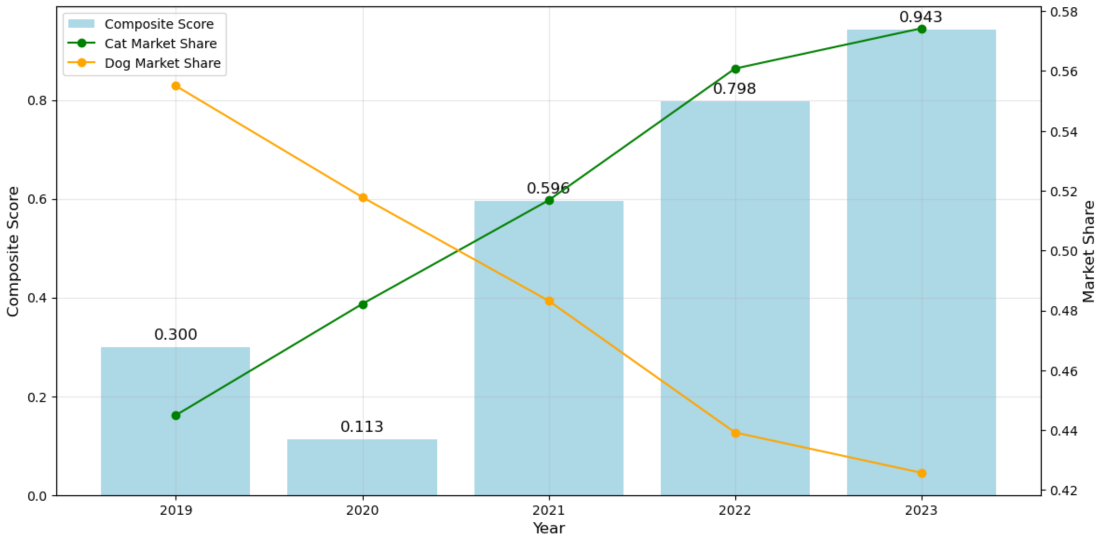


**Figure 3 Composite Score and Pet's Market Share in China (2019-2023)**

dogs is decreasing year by year.

In the past five years, consumer preferences for pet types have changed significantly in China's pet industry. The popularity of cats has increased year by year, showing strong momentum in both population growth and market share. The market for dogs is relatively stable but slightly shrinking, as evidenced by the negative growth in dog numbers and continued decline in market share in 2022.

In general, the entire pet industry has shown a positive development trend in the past five years. The increase in comprehensive scores indicates that the industry is constantly progressing. This may be due to the improvement of people's living standards, the increased demand for pet companionship, and the diversified development of pet-related products and services [4].

According to the Pearson correlation index, in addition to the influencing factors heat map, as Figure 4, we can see that economic factors (such as per capita GDP, per capita disposable income) and social structural factors (such as urbanization rate, the proportion of the elderly population) have a significant positive impact on the development of China's pet industry. At the same time, the development of e-commerce has also provided strong support for the prosperity of the pet industry. However, marital status is negatively correlated with the pet industry, which may reflect the social phenomenon that single people are more inclined to keep pets.


**Figure 4 Influencing Factors Heat Map**

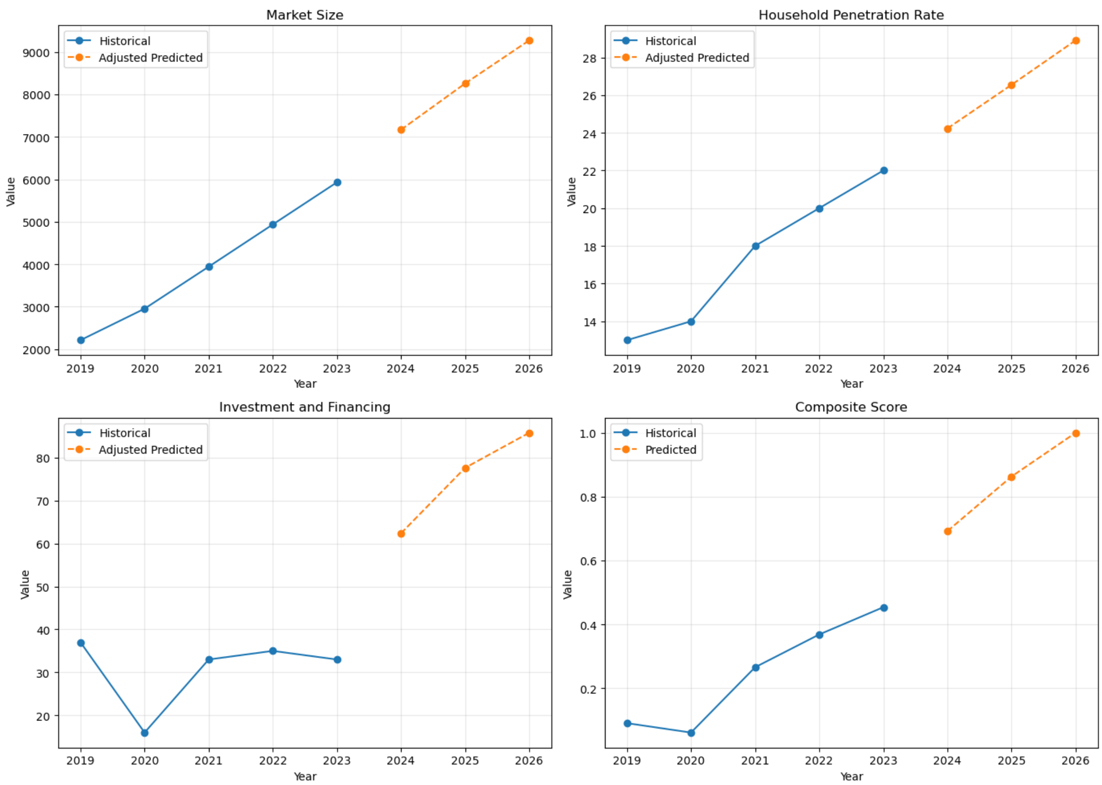


**Figure 5 2024-2026 Indicator Predicted Value**

### 5.3.2 *China's Pet Industry Prediction*

We use the Holt-Winter model to forecast and multiple linear regression models to predict evaluation indicators (including Market Size, Household Penetration Rate, Investment and Financing), then calculate the composite score, as shown in Figure 5.

China's pet industry will show multi-faceted growth in 2024-2026. Market size on upward trend 2019-2023, will continue growing with large growth rate and reach higher value in 2026. Household penetration rate increasing from 2019-2023, will continue rising in 2024-2026, approaching 28% by 2026. Investment and financing data fluctuated in 2019-2023, but will grow in 2024-2026. Comprehensive score slow upward trend in 2019-2023, will continue rising in 2024-2026.

# VI. Models and Solutions for Question 2

## 6.1 Model Establishment

In this question, our team use a weighted average model and a Holt-Winter model, which is the same as Question 1.

When computing the annual growth rates of cats and dogs in the three countries, we use the growth rate calculation formula. In mathematics, this formula can be expressed as:

$$\text{Growth Rate} = \left(\frac{\text{Current Year Count} - \text{Previous Year Count}}{\text{Previous Year Count}}\right) \times 100\% \tag{7}$$

When normalizing the data, we use Max-Min normalization:

$$\text{Normalized Value} = \frac{\text{Value} - \text{Min}}{\text{Max} - \text{Min}} \tag{8}$$

## 6.2 Solution

### 6.2.1 *Overseas Pet Industry Analysis*

Considering that there are too many countries in the world, we selected America, France and Germany as outstanding representatives to reflect the development of the global pet industry.

To analyze the development of the global pet industry in recent years, firstly, we compare the changes of the number of cats and dogs and their structural proportions, and show the rate of changes. To make data visible, we draw two charts. The fist one is a stacked bar chart, which represents the changes of the number of cats and dogs and their structural proportions in the United States, France and Germany over the past five years. We use colors with significant difference to show the difference of the number of cats and dogs in the three countries in the same year. In the same country, we still use colors with small difference to show the structural proportions of cats and dogs and their changes in the five years. The second is a line chart, displaying the annual growth rates of cats and dogs in the United States, France, and Germany from 2019 to 2023.

Secondly, we use a weighted average mathematical model and introduce three evaluation indicators, which are the market size of global pet industry, the global penetration rate of pet households and the global total expenditure of the pet industry respectively. Among them, the first indicator accounts for 0.4, while the last two indicators account for 0.3, as Table 3 shows.

**Table 3 Global Pet industry Evaluation Indicator**

| Indicator | Weight |
|---|---|
| Market Size | 0.4 |
| Household Penetration Rate | 0.3 |
| Total Expenditure | 0.3 |

Since that the numerical differences between indicators can reach up to a hundred times, it is necessary to normalize them and make them between zero to one. During the process, a Max-Min normalization is used to make different magnitudes of data between zero to one, so they are convenient for data analysis and model establishment. Then we calculate the annual composite score of each country using predefined weights and formula. In addition, we perform the market share calculation of pet cats and dogs for each country each year according to the total number of cats and dogs each year. After that, we draw the figure to visualize the data. In the graph, the composite scores are represented by the bar chart, while the market shares are represented by the line chart.

### 6.2.2 *The forecast of the global demand for pet food*

Since that pet food sales growth reflects market demand growth, so we can regard the global pet food sales revenue as the global demand for pet food. Then we collect the global pet food sales from 2010 to 2023 in the Pet industry database [5]. When forecasting the future global demand for pet food in the next three years, we found that the data measured in billion dollars of global pet food sales revenue in 2021 was lost, so we perform interpolation on this data. We create the interpolation function, and choose the cubic interpolation. According to the data that is not missing, it fits the data points using a cubic polynomial and it is more accurate than linear interpolation. After that, we use the Holt-Winters model to perform time series forecasting on the sales data. The seasonal period, trend computation and seasonal computation are set to 4, additive, additive respectively.

## 6.3 Results

### 6.3.1 *Overseas Pet Industry Analysis by Pet Type*

In the Figure 6, it can be seen from the chart that the number of pets in the US far exceeds that of the other two countries for almost 9 times. But the number in the US shows a downward trend while those in France and Germany remain stable from 2019 to 2023. In France and Germany, people tend to keep a cat rather than a dog, while in the US, there are more dogs than cats most of the time. Looking at the line chart, the US's cat growth rate shows the most significant fluctuation, followed by the France dog growth rate, and others remain relatively stable.

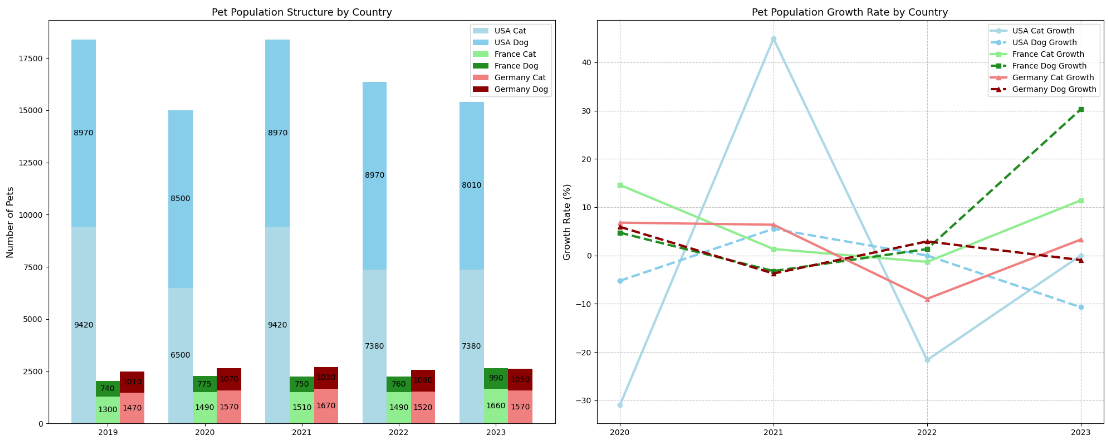


**Figure 6 Pet Population Analysis among Three Countries**

Therefore, according to the analysis result and the calculation of the three indicators in the Figure 7, we find that the USA enjoys a much higher composite score and the cat and dog market share than other countries. And the cat market share is lower than dog's. In conclusion, the development level of pet industry varies among overseas countries, and the development in the US goes well.

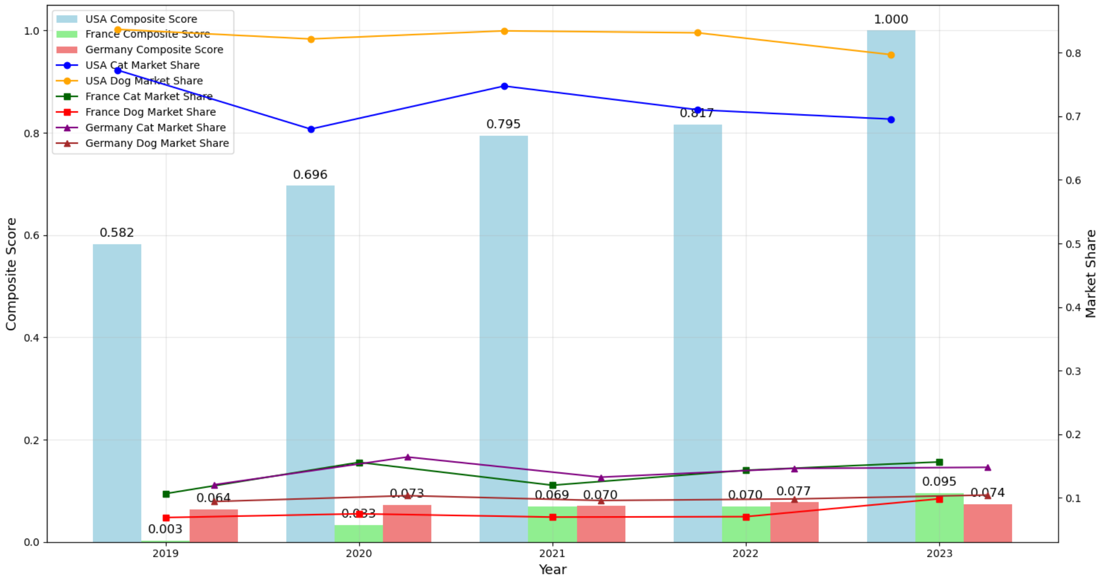


**Figure 7 The Composite Score and Pet Market Share**

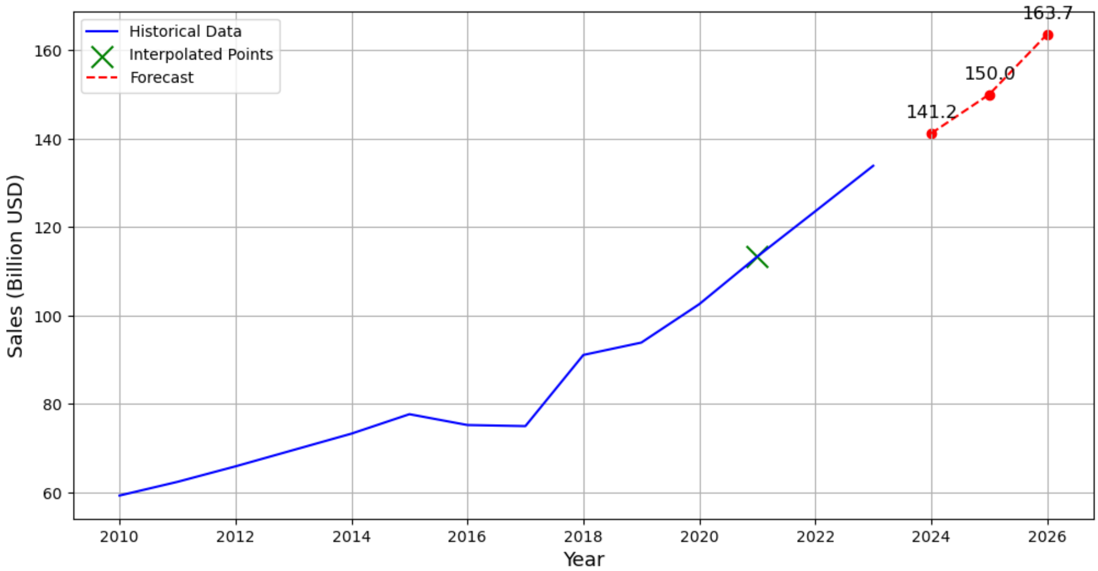


**Figure 8 The Forecast of Global Pet Food Sales**

### 6.3.2 *The forecast of the global demand for pet food*

In the Figure 8, the global pet food sales increased steadily in the past 13 years, except for a slight decline in 2017. According to the Holt-Winter model results, the sales will continue to grow in the next three years with 141.2, 150.0, 163.7 billions dollars respectively, meaning the global demand for pet food is also expected to increase.

# VII. Models and Solutions for Question 3

## 7.1 Model Establishment

### 7.1.1 *Vector Autoregression Model (VAR)*

The VAR model is a statistical model used to analyze the relationships between multiple time series variables. Compared to Autoregressive (AR) model, it allows to analyze multiple variables at the same time and are able to capture their dynamic interactions. And its formula is:

$$\mathbf{Y}_t = \mathbf{c} + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{Y}_{t-i} + \varepsilon_t \tag{9}$$

where $\mathbf{Y}_t$ is a $k$-dimensional vector of endogenous variables at time t. $\mathbf{c}$ is a $k$-dimensional constance, which is a intercept term. $\mathbf{A}_i$ is a $k \times k$ coefficient matrix, which represents the influence of the $i$-th lag on the current variable. $\varepsilon_t$ is a $k$-dimensional error vector (white noise), satisfying $\mathbb{E}[\varepsilon_t] = \mathbf{0}$ and $\mathrm{Var}(\varepsilon_t) = \mathbf{\Sigma}$, where $\mathbf{\Sigma}$ is the covariance matrix of the error terms.

## 7.2 Solution

### 7.2.1 *The Analysis of China's Pet Food Industry*

The value of China's pet food production is a direct indicator of the industry's growth, for if the demand of pet food increases, the production values will rise. For China's pet food export values which is measured in 100 millions, it reflects the international demand for Chinese pet food products. So the two values can be considered as the indicators to analyze and forecast the development of China's pet food industry.

To eliminate the differences caused by exchange rates between currencies, we assume that 1 CNY is equal to 0.14 USD (US Dollars) and convert CNY to USD. According to the data given in attachment 3, our team draw a line graph to display the changes of China's pet food production and export values from 2019 to 2023 and the ratio of export to production.

#### 7.2.2 *The prediction of the production and exports of pet food in China*

Production and Exports value will be influenced by the domestic and foreign situations. In order to consider these impacts, we introduce the data of attachment 1 and 2 which give the number of cats and dogs among four countries and global pet food sales as features.

To apply the VAR model, it is required to ensure each feature stationary. Firstly, our team perform Augmented Dickey Fuller (ADF) test to check the stationarity of each feature of the data. If the result of a feature is larger than 0.05, meaning that it is not stationary enough, then it should be applied first-order differencing. After the data processing, we use them to fit the VAR model. Then we extract the residuals to perform the Ljung-Box test check the autocorrelation. Finally, we use the fitted VAR model to forecast the differenced values in the next three years and convert them to actual values.

### 7.3 Results

#### 7.3.1 *The Analysis of the China's Pet Food Industry*

From the Figure 9, it is readily apparent that there was an considerable incline in the number of China's pet food production value, although there was a slight decrease in the year 2022. And the number ascended from just over 50 to almost 400 billions dollars in the five years. The export values climbed at a low speed and a decline can be seen in 2020, remaining between 0 and 50 billions. For the ratio of export to production, there was a noticeable decline, though it showed an incline in 2022. The percentage descended from 0.35 in 2019 to 0.10 in 2023. In conclusion, there are more and more pet food produced but not for export. The China's pet food industry developed at a relative fast speed positively.

#### 7.3.2 *The prediction of the production and exports of pet food in China*

The Figure 10 illustrates that in the next three years, the production value of pet food will suffer from a decline in 2025, but it will rise to 761.7 billions dollars in 2026 finally. For the export value, it will ascend stably to 65.2 billions dollars in 2026. As can be seen, the two values will increase in the next three years.

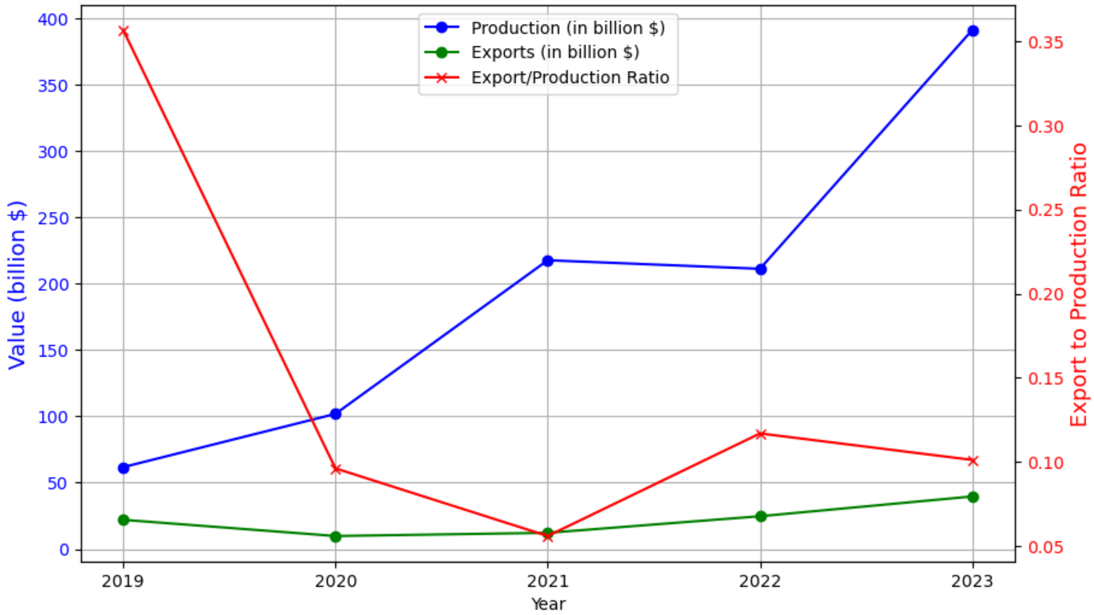


**Figure 9 Pet Population in Three Countries**

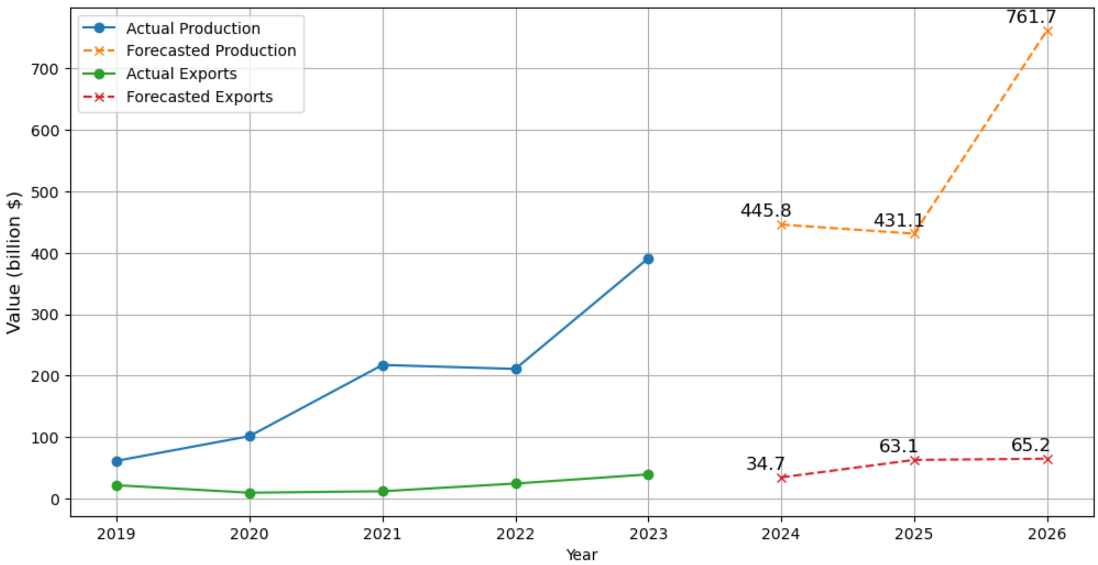


**Figure 10 Pet Population in Three Countries**

# VIII. Models and Solutions for Question 4

## 8.1 Model Establishment

### 8.1.1 *Differences-in-Differences (DID) Model*

The DID model, also known as the quasi - experimental research design, is a statistical technique used to estimate the causal effect of a treatment or policy intervention. Basic

DID model formula is:

$$Y_{it} = \beta_0 + \beta_1 \text{treated}_i + \beta_2 \text{policy}_t + \beta_3 (\text{treated}_i \times \text{policy}_t) + \epsilon_{it} \tag{10}$$

where $Y_{it}$ is the dependent variable (such as export volume or export value) of the $i$-th group at the $t$-th time point. $\beta_0$ is the intercept term. $treated_i$ is the group variable, where 1 represents the treatment group and 0 represents the control group $policy_t$ is the time variable, where 1 represents the post-policy period and 0 represents the pre-policy period. $(treated_i \times policy_t)$ is the interaction term, used to measure the impact of the policy on the treatment group. $\epsilon_{it}$ is the error term.

## 8.2 Solution

In order to quantitatively analyze the impact of the economic policies of European and American countries on the development of China's pet food industry, we took Germany, France and the United States as the treatment group. A trade turning point was selected for these three countries.

During the Trump administration, the United States began to impose a 10% tariff on pet food exported from China. From May 2019, the tariff increased from 10% to 25% [6]. In September 2024, France supported the European Union to impose tariffs on Chinese electric vehicles, and the Sino-French trade friction [7]. In July 2023, Germany, the economic core of the European Union, issued its first China strategy in the form of an official Germany document, actively taking measures to reduce economic dependence on China. This policy adjustment may have a certain impact on Sino-Germany trade [8]. We also chose Malaysia, a country with friendly trade relations with China, as the control group.

The DID model was then used to assess the impact of policy interventions on Chinese pet food exports among three countries. Based on China customs data, we collected the volume and value of Chinese exports each month for about one year before and after the policy event for the treatment group (France, Germany, the United States) and the control group (Malaysia) [9].

The overall data construction is as Table 4.

## 8.3 Results

The DID model compares the changes in export volume and value of the treatment group (France, USA, Germany) and the control group (Malaysia) before and after the

**Table 4 The Data Construction of DID**

| Control | Treatment | Time | Outcome |
|---|---|---|---|
| | America | May, 2019 | exports volume (kg) |
| Malaysia | France | September, 2024 | |
| | Germany | July, 2023 | exports value ($) |

implementation of the policy, thereby excluding the common effects of time trends or other external factors. The following results were obtained:

**Table 5 Post-Treatment Export Data for Different Countries**

| Outcome | France (Treated) | America (Treated) | Germany (Treated) |
|---|---|---|---|
| Volume | -14200 | -832000 | 58340 |
| Value | -321000 | -4090000 | -496000 |

For the United States, both export value and export volume showed a significant downward trend, indicating that the high tariff policy has significantly suppressed China's pet food exports to the United States. For France, the impact on export volume and export value is smaller than that of the United States, but it is still negative, indicating that the policy intervention has still had a restraining effect on exports. The decrease in export value to Germany but the increase in export volume is contrary to the Sino-German trade friction and the policy intention of reducing dependence on the Chinese economy. The increase in export volume may indicate that in the early stage of the policy, Chinese companies responded to the policy change by reducing price competition or increasing export volume, thereby maintaining market share.

Based on the above findings, it is evident that external economic policies significantly impact China's pet food exports. To promote sustainable development in this industry, it is crucial to devise strategies that enhance resilience against such external shocks and capitalize on emerging opportunities. Combining some references and calculated data, we have initially developed some strategies that can promote the sustainable development of China's pet food industry and resist the impact of external economic policies.

**1. Strengthen branding and enhance product additional value:** Our country's pet food exports have problems such as weak international competitiveness, low export prices, and weak brand effect [10]. Therefore, enterprises should enhance the added value of products by improving product quality, increasing product innovation and adopting high-end packaging, while strengthening branding and improving branding impression and influence.

**2. Increase productivity and reduce costs:** Pet food processing and export trade is facing unfavorable factors such as rising raw material and energy prices [11]. Enterprises should improve production efficiency and reduce production costs through technological innovation and management optimization in order to maintain price competitiveness.

**3. Enhance product quality and environmental standards:** According to the RCEP agreement, China should raise the quality and environmental standards of feed export products to deal with green trade barriers [12]. This will not only help to improve the competitiveness of products in the international market, but also reduce trade friction caused by environmental protection issues.

# IX. Model Evaluation

**Question 1:** Based on the history data and predicted values in the last five years, the performance of three forecasting models combining Holt-Winters and linear regression for three metrics: market size, penetration, and amount of investment financing was comprehensively evaluated using four metrics: mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and R square ($R^2$). The detailed results are as Table 6.

**Table 6 Model Performance Metrics in Q1**

| Models | MSE | RMSE | MAE | $R^2$ |
|---|---|---|---|---|
| Market size | 5081.26 | 71.28 | 60.76 | 0.9971 |
| Penetration rate | 0.32 | 0.57 | 0.48 | 0.9730 |
| Investments and financing | 54.54 | 7.39 | 5.48 | 0.0425 |

The Market Size Prediction Model and the Penetration Rate Prediction Model both perform well, with the market size model demonstrating extremely high accuracy. However, The Number of Investments and Financing Prediction Model performs poorly, likely due to high variability in the data or the absence of critical features. In the future, we can make an effort to improve the performance of predictive model of investment number.

**Question 2:** We utilizes the Q1 evaluation metrics to assess the Holt-Winters model's performance in this question based on the historical data from 2010 to 2023. The Table 7 shows the result:

**Table 7 Model Performance Metrics in Q2**

| MSE | RMSE | MAE | $R^2$ |
|---|---|---|---|
| 15.19 | 3.90 | 2.71 | 0.9705 |

The result shows that the Holt-Winters model fit the historical data well with 0.9705 of $R^2$ and its prediction is reliable.

**Question 3:** We calculate the residuals of all variables between the historical values and the predicted values, use them to perform Ljung-box test. The result of the Ljung-Box test shows that the p value of each variable is larger than 0.05, indicating that the VAR model can fit the data well.

**Table 8 Ljung-Box Test Results**

| Variable | lb_stat | lb_pvalue |
|---|---|---|
| Production in dollars | 2.840236686390533 | 0.09193067161929441 |
| Exports | 2.2280557584611627 | 0.13552428918205203 |
| Global Food | 1.3856186518928901 | 0.2391464001513451 |
| China Cat | 3.2461911357340725 | 0.07158963678923984 |
| China Dog | 0.9534062383165093 | 0.32885378155405437 |
| USA Cat | 1.5217920623326027 | 0.21734850210820406 |
| USA Dog | 3.2335883773978766 | 0.07214245049747338 |
| France Cat | 2.228055758461164 | 0.1355242891820516 |
| France Dog | 3.1419863637791807 | 0.07630083040582475 |
| German Cat | 0.0 | 1.0 |
| German Dog | 2.657312925170069 | 0.1030747303555349 |

# X. References

[1] Apa Institute, CHINA PET INDUSTRY ANNUAL REPORT (2023), https://www.fxbaogao.com/view?id=4518785&query=%7B%22keywords%22:%22%E5%AE%A0%E7%89%A9%E6%95%B0%E9%87%8F%22%7D&index=0&pid=343&xid=.

[2] Kurun Data, ”2024 Pet Industry Market Trends Report,” https://www.fxbaogao.com/view?id=4487794, 2024-11-24.

[3] CNKI, ”China Economic and Social Big Data Research Platform,” https://data.cnki.net/, 2024-11-24.

[4] Wang, Xinzhu, ”A Brief Discussion on the Development Status and Trends of China's Pet Industry,” Modern Animal Husbandry and Veterinary Science, 2018(2): 57-59.

[5] Pet Industry Database, https://www.petslib.cn/insight/market/504621029876826112.html.

[6] Sun, Ruizhe, ”Rationally Handling Trade Frictions: Joint Development is Our Social Responsibility,” Dyeing and Finishing Technology, 2019, 41(5): 64.

[7] China Metallurgical News Agency, ”EU Consultative Vote on Tariffs for Chinese Electric Vehicles: 12 Countries Support, 4 Oppose, 11 Abstain,” Sichuan Metallurgy, 2024, 46(4): 71.

[8] Xiong, Wei, and Zhao, Doudou, ”Germany's First 'China Strategy' and Policy Adjustments Towards China,” Border and Ocean Studies, 2023(5): 42-57.

[9] General Administration of Customs of the People's Republic of China, Customs statistical data online query platform, http://stats.customs.gov.cn/.

[10] Liu, Yongquan, Sun, Qi, and Shi, Yutao, ”Measurement and Analysis of the International Competitiveness of China's Pet Food,” China Feed, 2023(11): 114-118. DOI:10.15906/j.cnki.cn11-2975/s.2023050101-05.

[11] Zhang, Yunnuan, ”Research on International Marketing Strategy of W Company's Pet Food,” Master's Thesis, Ocean University of China, Shandong, 2011. DOI:10.7666/d.y2159564.

[12] Ma, Huiying, and Zheng, Junrui, ”Reasons and Solutions for Green Trade Barriers in Feed Export under the RCEP Perspective,” China Feed, 2024(18): 153-156. DOI:10.15906/j.cnki.cn11-2975/s.20241839.

[13] Lütkepohl, H, New Introduction to Multiple Time Series Analysis, Springer, 2005.

# XI. Appendix

**Table 9 Chinese Pet Industry Related Features (2019-2023)**

| Feature | 2019 | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|---|
| Cat | 4412 | 4862 | 5806 | 6536 | 6980 |
| Dog | 5503 | 5222 | 5429 | 5119 | 5175 |
| Market Size | 2212 | 2953 | 3942 | 4936 | 5928 |
| Household Penetration Rate | 13 | 14 | 18 | 20 | 22 |
| Number of Pet Track Investments and Financing | 37 | 16 | 33 | 35 | 33 |
| GDP per capita | 70892 | 71965 | 80976 | 85698 | 89358 |
| Per capita disposable income | 30733 | 32189 | 35128 | 36883 | 39218 |
| Proportion of the population aged 65 and over | 12.6 | 13.5 | 14.2 | 14.9 | 15.4 |
| National Marriage Registration | 927.3 | 814.3 | 764.3 | 683.5 | 768 |
| National Urbanization Rate | 62.71 | 63.89 | 64.72 | 65.22 | 66.16 |
| E-commerce Transaction Volume | 34.81 | 37.2 | 42.3 | 44.5 | 58.38 |

**Table 10 Global Pet Food Sales Data for Question 2**

| Year | Pet food sales worldwide |
|---|---|
| 2010 | 59.3 |
| 2011 | 62.4 |
| 2012 | 65.9 |
| 2013 | 69.6 |
| 2014 | 73.3 |
| 2015 | 77.7 |
| 2016 | 75.25 |
| 2017 | 75 |
| 2018 | 91.1 |
| 2019 | 93.9 |
| 2020 | 102.6 |
| 2021 | - |
| 2022 | 123.6 |
| 2023 | 133.9 |

**Table 11 Pet Industry Data in the US (2019 - 2023)**

| Feature | 2019 | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|---|
| **Cat** | 9420 | 6500 | 9420 | 7380 | 7380 |
| **Dog** | 8970 | 8500 | 8970 | 8970 | 8010 |
| **Household Penetration Rate** | 0.67 | 0.7 | 0.7 | 0.66 | 0.85 |
| **Total expenditures in the pet industry** | 957 | 1036 | 1236 | 1368 | 1470 |
| **Market Size** | 753.8 | 1036 | 1236 | 1300 | 1470 |

**Table 12 Pet Industry Data in France (2019 - 2023)**

| Feature | 2019 | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|---|
| **Cat** | 1300 | 1490 | 1510 | 1490 | 1660 |
| **Dog** | 740 | 775 | 750 | 760 | 990 |
| **Household Penetration Rate** | 0.35 | 0.4 | 0.46 | 0.46 | 0.5 |
| **Total expenditures in the pet industry** | 48 | 49 | 49 | 42 | 47 |
| **Market Size** | 49 | 48 | 49 | 58 | 59 |

**Table 13 Pet Industry Data in Germany (2019 - 2023)**

| Feature | 2019 | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|---|
| **Cat** | 1470 | 1570 | 1670 | 1520 | 1570 |
| **Dog** | 1010 | 1070 | 1030 | 1060 | 1050 |
| **Household Penetration Rate** | 0.45 | 0.47 | 0.46 | 0.46 | 0.45 |
| **Total expenditures in the pet industry** | 51.55 | 45.13 | 60 | 65 | 70 |
| **Market Size** | 51.55 | 45.13 | 45.13 | 65 | 70 |

**Table 14 The data used in question 3**

| Feature | 2019 | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|---|
| Production | 440.7 | 727.3 | 1554 | 1508 | 2793 |
| Exports | 22.0 | 9.8 | 12.2 | 24.7 | 39.6 |
| Global_Food | 93.9 | 102.6 | 113.4 | 123.6 | 133.9 |
| China_Cat | 4412 | 4862 | 5806 | 6536 | 6980 |
| China_Dog | 5503 | 5222 | 5429 | 5119 | 5175 |

| Feature | 2019 | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|---|
| USA_Cat | 9420 | 6500 | 9420 | 7380 | 7380 |
| USA_Dog | 8970 | 8500 | 8970 | 8970 | 8010 |
| France_Cat | 1300 | 1490 | 1510 | 1490 | 1660 |
| France_Dog | 740 | 775 | 750 | 760 | 990 |
| German_Cat | 1470 | 1570 | 1670 | 1520 | 1570 |
| German_Dog | 1010 | 1070 | 1030 | 1060 | 1050 |

**Table 15 Export Data of America (2018-2020)**

| id | time | group | export_amount | export_value | policy |
|---|---|---|---|---|---|
| 1 | 201805 | control | 229485 | 232,182 | 0 |
| 1 | 201806 | control | 168001 | 175,850 | 0 |
| 1 | 201807 | control | 183517 | 162,991 | 0 |
| 1 | 201808 | control | 200635 | 238,057 | 0 |
| 1 | 201809 | control | 486935 | 393,504 | 0 |
| 1 | 201810 | control | 478671 | 360,929 | 0 |
| 1 | 201811 | control | 652302 | 510,804 | 0 |
| 1 | 201812 | control | 881389 | 657,886 | 0 |
| 1 | 201901 | control | 470843 | 360,953 | 0 |
| 1 | 201902 | control | 618754 | 451,501 | 0 |
| 1 | 201903 | control | 429113 | 334,255 | 0 |
| 1 | 201904 | control | 469888 | 423,110 | 0 |
| 1 | 201905 | control | 1148311 | 949,566 | 1 |
| 1 | 201906 | control | 535004 | 457,785 | 1 |
| 1 | 201907 | control | 723822 | 626,080 | 1 |
| 1 | 201908 | control | 597170 | 464,644 | 1 |
| 1 | 201909 | control | 604825 | 556,009 | 1 |
| 1 | 201910 | control | 739212 | 670,868 | 1 |
| 1 | 201911 | control | 683376 | 605,946 | 1 |
| 1 | 201912 | control | 1117057 | 971,436 | 1 |
| 1 | 202001 | control | 1021088 | 889,994 | 1 |
| 1 | 202002 | control | 398102 | 313,349 | 1 |
| 1 | 202003 | control | 1562822 | 1,310,566 | 1 |

| id | time | group | export_amount | export_value | policy |
|---|---|---|---|---|---|
| 1 | 202004 | control | 1198649 | 1,070,288 | 1 |
| 2 | 201805 | treatment | 4057367 | 29,343,123 | 0 |
| 2 | 201806 | treatment | 4114431 | 29,793,095 | 0 |
| 2 | 201807 | treatment | 3329708 | 23,593,288 | 0 |
| 2 | 201808 | treatment | 3437309 | 24,913,787 | 0 |
| 2 | 201809 | treatment | 4728256 | 32,585,480 | 0 |
| 2 | 201810 | treatment | 3640432 | 25,421,275 | 0 |
| 2 | 201811 | treatment | 4008247 | 27,044,416 | 0 |
| 2 | 201812 | treatment | 4219797 | 29,318,638 | 0 |
| 2 | 201901 | treatment | 4129937 | 28,389,155 | 0 |
| 2 | 201902 | treatment | 1585274 | 10,877,224 | 0 |
| 2 | 201903 | treatment | 3045560 | 21,088,212 | 0 |
| 2 | 201904 | treatment | 2973389 | 20,943,389 | 0 |
| 2 | 201905 | treatment | 3596569 | 24,292,715 | 1 |
| 2 | 201906 | treatment | 2806454 | 19,516,285 | 1 |
| 2 | 201907 | treatment | 3960754 | 27,506,278 | 1 |
| 2 | 201908 | treatment | 3622491 | 24,560,418 | 1 |
| 2 | 201909 | treatment | 3793236 | 25,295,800 | 1 |
| 2 | 201910 | treatment | 3568949 | 23,743,166 | 1 |
| 2 | 201911 | treatment | 3283150 | 21,753,658 | 1 |
| 2 | 201912 | treatment | 3191274 | 20,757,476 | 1 |
| 2 | 202001 | treatment | 4152799 | 28,146,068 | 1 |
| 2 | 202002 | treatment | 617015 | 4,619,440 | 1 |
| 2 | 202003 | treatment | 2288223 | 15,292,325 | 1 |
| 2 | 202004 | treatment | 3463209 | 23,309,160 | 1 |

**Table 16 Export Data of France (2024**

| id | time | group | export_amount | export_value | policy |
|---|---|---|---|---|---|
| 1.00 | 202404 | control | 1118914.00 | 1134152.00 | 0.00 |
| 1.00 | 202405 | control | 602437.00 | 874438.00 | 0.00 |
| 1.00 | 202406 | control | 908910.00 | 1208208.00 | 0.00 |
| 1.00 | 202407 | control | 1160742.00 | 1587648.00 | 0.00 |

| id | time | group | export_amount | export_value | policy |
|---|---|---|---|---|---|
| 1.00 | 202408 | control | 519185.00 | 808084.00 | 0.00 |
| 1.00 | 202409 | control | 899390.00 | 1617639.00 | 0.00 |
| 1.00 | 202410 | control | 824024.00 | 1171190.00 | 1.00 |
| 2.00 | 202404 | treatment | 139897.00 | 876004.00 | 0.00 |
| 2.00 | 202405 | treatment | 106732.00 | 701465.00 | 0.00 |
| 2.00 | 202406 | treatment | 192826.00 | 1249665.00 | 0.00 |
| 2.00 | 202407 | treatment | 178674.00 | 1155664.00 | 0.00 |
| 2.00 | 202408 | treatment | 200727.00 | 1258387.00 | 0.00 |
| 2.00 | 202409 | treatment | 146585.00 | 964186.00 | 0.00 |
| 2.00 | 202410 | treatment | 102463.00 | 679092.00 | 1.00 |

**Table 17 Export Data of Germany(2022-2024)**

| id | time | group | export_amount | export_value | policy |
|---|---|---|---|---|---|
| 1.00 | 202207 | control | 1351664.00 | 1418261.00 | 0.00 |
| 1.00 | 202208 | control | 1766407.00 | 1800661.00 | 0.00 |
| 1.00 | 202209 | control | 1259195.00 | 1358148.00 | 0.00 |
| 1.00 | 202210 | control | 1180058.00 | 1193690.00 | 0.00 |
| 1.00 | 202211 | control | 695739.00 | 845955.00 | 0.00 |
| 1.00 | 202212 | control | 794191.00 | 1059363.00 | 0.00 |
| 1.00 | 202301 | control | 434026.00 | 576940.00 | 0.00 |
| 1.00 | 202302 | control | 592935.00 | 670592.00 | 0.00 |
| 1.00 | 202303 | control | 1405374.00 | 1542644.00 | 0.00 |
| 1.00 | 202304 | control | 1218820.00 | 1367974.00 | 0.00 |
| 1.00 | 202305 | control | 1207260.00 | 1411496.00 | 0.00 |
| 1.00 | 202306 | control | 1211904.00 | 1631701.00 | 0.00 |
| 1.00 | 202307 | control | 686081.00 | 907565.00 | 1.00 |
| 1.00 | 202308 | control | 620624.00 | 822784.00 | 1.00 |
| 1.00 | 202309 | control | 840387.00 | 1036866.00 | 1.00 |
| 1.00 | 202310 | control | 752533.00 | 1060573.00 | 1.00 |
| 1.00 | 202311 | control | 1601192.00 | 1893017.00 | 1.00 |
| 1.00 | 202312 | control | 1516265.00 | 1855113.00 | 1.00 |
| 1.00 | 202401 | control | 787587.00 | 1255124.00 | 1.00 |

| id | time | group | export_amount | export_value | policy |
|---|---|---|---|---|---|
| 1.00 | 202402 | control | 1825421.00 | 2250128.00 | 1.00 |
| 1.00 | 202403 | control | 1350678.00 | 1765510.00 | 1.00 |
| 1.00 | 202404 | control | 1118914.00 | 1134152.00 | 1.00 |
| 1.00 | 202405 | control | 602437.00 | 874438.00 | 1.00 |
| 1.00 | 202406 | control | 908910.00 | 1208208.00 | 1.00 |
| 2.00 | 202207 | treatment | 2001721.00 | 14056068.00 | 0.00 |
| 2.00 | 202208 | treatment | 2067072.00 | 14943381.00 | 0.00 |
| 2.00 | 202209 | treatment | 1945445.00 | 14057823.00 | 0.00 |
| 2.00 | 202210 | treatment | 1940547.00 | 14170147.00 | 0.00 |
| 2.00 | 202211 | treatment | 2377708.00 | 16824274.00 | 0.00 |
| 2.00 | 202212 | treatment | 2603592.00 | 18813967.00 | 0.00 |
| 2.00 | 202301 | treatment | 2028721.00 | 14797140.00 | 0.00 |
| 2.00 | 202302 | treatment | 1486664.00 | 10820076.00 | 0.00 |
| 2.00 | 202303 | treatment | 3294467.00 | 24148221.00 | 0.00 |
| 2.00 | 202304 | treatment | 2684301.00 | 19249433.00 | 0.00 |
| 2.00 | 202305 | treatment | 2525140.00 | 17844883.00 | 0.00 |
| 2.00 | 202306 | treatment | 2719589.00 | 18722533.00 | 0.00 |
| 2.00 | 202307 | treatment | 1865272.00 | 12990025.00 | 1.00 |
| 2.00 | 202308 | treatment | 1924160.00 | 13281461.00 | 1.00 |
| 2.00 | 202309 | treatment | 2333731.00 | 16601772.00 | 1.00 |
| 2.00 | 202310 | treatment | 2294462.00 | 16080641.00 | 1.00 |
| 2.00 | 202311 | treatment | 2078785.00 | 14424502.00 | 1.00 |
| 2.00 | 202312 | treatment | 2400858.00 | 16890020.00 | 1.00 |
| 2.00 | 202401 | treatment | 2288985.00 | 15978891.00 | 1.00 |
| 2.00 | 202402 | treatment | 1833156.00 | 12730243.00 | 1.00 |
| 2.00 | 202403 | treatment | 2387644.00 | 16311077.00 | 1.00 |
| 2.00 | 202404 | treatment | 2556908.00 | 17406531.00 | 1.00 |
| 2.00 | 202405 | treatment | 2669439.00 | 18596881.00 | 1.00 |
| 2.00 | 202406 | treatment | 3235134.00 | 22393037.00 | 1.00 |

Listing 1: The code of question 1

```
import pandas as pd
import matplotlib.pyplot as plt
```

```
import numpy as np
import seaborn as sns
from statsmodels.tsa.holtwinters import ExponentialSmoothing
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error,
    r2_score

# Read data
data = pd.read_excel('./data/attachment1.xlsx')
data.set_index('feature', inplace=True)

data.head()

# Extract data
years = data.columns
cats = data.loc['Cat']
dogs = data.loc['Dog']

# Calculate total number and change rate
total_pets = cats + dogs
cat_change_rate = cats.pct_change() * 100
dog_change_rate = dogs.pct_change() * 100

# Draw stacked bar chart
fig, ax1 = plt.subplots(figsize=(12, 6))

# Draw stacked bar chart for cats and dogs
bars_dogs = ax1.bar(years, dogs, label='Dogs', color='#FFA07A')
bars_cats = ax1.bar(years, cats, bottom=dogs, label='Cats',
    color='#87CEEB')

# Mark the number of cats and dogs
for bar, dog, cat in zip(bars_dogs, dogs, cats):
    ax1.text(bar.get_x() + bar.get_width() / 2, dog / 2, f'{dog}',
        ha='center', va='center', color='black', fontsize=12)
    ax1.text(bar.get_x() + bar.get_width() / 2, dog + cat / 2, f'{cat}',
        ha='center', va='center', color='black', fontsize=12)
```

```
# Set left Y-axis
ax1.set_xlabel('Year', fontsize=14)
ax1.set_ylabel('Number of Pets (10,000s)', fontsize=14)
# ax1.set_title('Pet Population and Growth Rate (2019-2023)')

# Create right Y-axis
ax2 = ax1.twinx()

# Draw growth rate line chart
line1, = ax2.plot(years, cat_change_rate, color='blue', marker='o',
    linestyle='-', label='Cat Growth Rate')
line2, = ax2.plot(years, dog_change_rate, color='red', marker='o',
    linestyle='-', label='Dog Growth Rate')
ax2.set_ylabel('Growth Rate (%)', fontsize=14)

# Mark growth rate values and move them up
for x, y in zip(years, cat_change_rate):
    ax2.text(x, y + 1, f'{y:.1f}%', ha='center', va='bottom',
        color='blue') # Move up 1 unit
for x, y in zip(years, dog_change_rate):
    ax2.text(x, y + 1, f'{y:.1f}%', ha='center', va='bottom',
        color='red') # Move up 1 unit

# Add legend
lines1, labels1 = ax1.get_legend_handles_labels()
lines2, labels2 = ax2.get_legend_handles_labels()
ax1.legend(lines1 + lines2, labels1 + labels2, loc='upper left')

# Add grid lines
ax1.grid(True, alpha=0.3)

# Show the chart
plt.tight_layout()
plt.show()

data = data.transpose()
```

```
data.head()

# Extract required data
market_size = data['Market Size']
penetration_rate = data['Household Penetration Rate']
investment = data['Number of Pet Track Investments and Financing']

# Normalization function
def normalize(series):
    return (series - series.min()) / (series.max() - series.min())

# Normalize data
norm_market_size = normalize(market_size)
norm_penetration_rate = normalize(penetration_rate)
norm_investment = normalize(investment)

# Calculate composite score
weights = {'market_size': 0.4, 'penetration_rate': 0.3, 'investment':
    0.3}
data['Composite Score'] = (norm_market_size * weights['market_size'] +
                           norm_penetration_rate *
                               weights['penetration_rate'] +
                           norm_investment * weights['investment'])

# Analyze the market development of cats and dogs
cats = data['Cat']
dogs = data['Dog']
total_pets = cats + dogs

data['Cat Market Share'] = cats / total_pets
data['Dog Market Share'] = dogs / total_pets

# Print results
print(data[['Composite Score', 'Cat Market Share', 'Dog Market Share']])

# Visualization
fig, ax1 = plt.subplots(figsize=(12, 6))
```

```
# Draw bar chart of composite score
bars = ax1.bar(data.index, data['Composite Score'], color='lightblue',
   label='Composite Score')
ax1.set_xlabel('Year', fontsize=12)
ax1.set_ylabel('Composite Score', fontsize=12)
# ax1.set_title('Pet Industry Composite Score and Market Share
   (2019-2023)')

# Mark the composite score values on the bar chart
for bar in bars:
   yval = bar.get_height()
   ax1.text(bar.get_x() + bar.get_width() / 2, yval + 0.01,
      f'{yval:.3f}', ha='center', va='bottom', fontsize=12)

# Create second y-axis
ax2 = ax1.twinx()

# Draw line chart of cat and dog market shares
line1, = ax2.plot(data.index, data['Cat Market Share'], color='green',
   marker='o', linestyle='-', label='Cat Market Share')
line2, = ax2.plot(data.index, data['Dog Market Share'], color='orange',
   marker='o', linestyle='-', label='Dog Market Share')
ax2.set_ylabel('Market Share', fontsize=12)

# Combine legends
lines, labels = ax1.get_legend_handles_labels()
lines2, labels2 = ax2.get_legend_handles_labels()
ax1.legend(lines + [line1, line2], labels + labels2, loc='upper left')

# Add grid lines
ax1.grid(True, alpha=0.3)

# Show the chart
plt.tight_layout()
plt.show()
```

```
# Extract influencing factors
factors = data[['GDP per capita', 'per capita disposable income',
              'National Marriage Registration', 'Proportion of the
                 population aged 65 and over',
              'National Urbanization Rate', 'E-commerce Transaction
                 Volume']]

# Create name mapping dictionary
name_mapping = {
   'Market Size': 'Market',
   'Household Penetration Rate': 'Penetration',
   'Number of Pet Track Investments and Financing': 'Investment',
   'Composite Score': 'Score'
}

# Select target variables to analyze
target_columns = ['Market Size', 'Household Penetration Rate',
               'Number of Pet Track Investments and Financing',
                  'Composite Score']

# Calculate correlation
correlation_matrix = pd.DataFrame(index=factors.columns)
for col in target_columns:
   correlation_matrix[name_mapping[col]] = factors.corrwith(data[col])

# Visualization
plt.figure(figsize=(12, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
# plt.title('Correlation between Factors and Target Variables')
plt.xlabel('Target Variables')
plt.ylabel('Factors')
plt.tight_layout()
plt.show()

# Time series prediction
def predict_with_time_series(data, steps=3):
   model = ExponentialSmoothing(data, trend='add', seasonal=None).fit()
```

```
    forecast = model.forecast(steps)
    future_years = pd.Index(range(2024, 2027), name='Year')
    return pd.Series(forecast.values, index=future_years), model

# Feature adjustment
def adjust_with_features(base_prediction, features, target):
    # Calculate annual growth of historical data
    historical_growth = np.diff(target)

    # Use features to predict the adjustment of growth rate
    model = LinearRegression().fit(features[:-1], historical_growth)

    adjusted_predictions = []
    current_value = target[-1]

    for i, base_value in enumerate(base_prediction):
        # Dynamically get new growth_adjustment
        if i < len(features) - 1:
            growth_adjustment = model.predict(features[i].reshape(1,
                -1))[0]
        else:
            growth_adjustment = model.predict(features[-1].reshape(1,
                -1))[0]

        # Combine time series prediction and feature adjustment
        base_growth = base_value - current_value
        adjusted_growth = (base_growth + growth_adjustment) / 2 # The
            weight can be adjusted
        next_value = current_value + adjusted_growth
        adjusted_predictions.append(next_value)
        current_value = next_value

    return pd.Series(adjusted_predictions, index=base_prediction.index)

# Predict market size for the next three years
ts_prediction_market_size, ms_model =
    predict_with_time_series(market_size)
```

```
adjusted_market_size = adjust_with_features(ts_prediction_market_size,
    factors.values, market_size.values)

# Predict penetration rate for the next three years
ts_prediction_penetration_rate, pr_model =
    predict_with_time_series(penetration_rate)
adjusted_penetration_rate =
    adjust_with_features(ts_prediction_penetration_rate,
    factors.values, penetration_rate.values)

# Predict the number of investments and financings for the next three
    years
ts_prediction_investment, inves_model =
    predict_with_time_series(investment)
adjusted_investment = adjust_with_features(ts_prediction_investment,
    factors.values, investment.values)

# Print adjusted prediction results
print("\nAdjusted Market Size Prediction (2024-2026):")
print(adjusted_market_size.round(2))
print("\nAdjusted Penetration Rate Prediction (2024-2026):")
print(adjusted_penetration_rate.round(2))
print("\nAdjusted Investment Prediction (2024-2026):")
print(adjusted_investment.round(2))

def calculate_composite_score(market_size, penetration_rate,
    investment):
    # Normalization function
    def normalize(series):
        return (series - series.min()) / (series.max() - series.min())

    # Normalize each indicator
    norm_market_size = normalize(market_size)
    norm_penetration_rate = normalize(penetration_rate)
    norm_investment = normalize(investment)

    # Set weights
```

```
    weights = {
        'market_size': 0.4,
        'penetration_rate': 0.3,
        'investment': 0.3
    }

    # Calculate composite score
    composite_score = (norm_market_size * weights['market_size'] +
                      norm_penetration_rate * weights['penetration_rate'] +
                      norm_investment * weights['investment'])

    # Ensure the returned Series has the correct index
    if isinstance(market_size, pd.Series):
        composite_score.index = market_size.index

    return composite_score

# Visualize data for the next three years
years = np.array(range(2019, 2027))
fig, axs = plt.subplots(2, 2, figsize=(14, 10))

# Draw market size
axs[0, 0].plot(years[:5], market_size, marker='o', label='Historical')
axs[0, 0].plot(years[5:], adjusted_market_size,
             marker='o', linestyle='--', label='Adjusted Predicted')
axs[0, 0].set_title('Market Size')
axs[0, 0].legend()

# Draw penetration rate
axs[0, 1].plot(years[:5], penetration_rate, marker='o',
    label='Historical')
axs[0, 1].plot(years[5:], adjusted_penetration_rate,
             marker='o', linestyle='--', label='Adjusted Predicted')
axs[0, 1].set_title('Household Penetration Rate')
axs[0, 1].legend()

# Draw the number of investments and financings
```

```
axs[1, 0].plot(years[:5], investment, marker='o', label='Historical')
axs[1, 0].plot(years[5:], adjusted_investment,
             marker='o', linestyle='--', label='Adjusted Predicted')
axs[1, 0].set_title('Investment and Financing')
axs[1, 0].legend()

future_scores = calculate_composite_score(
   pd.concat([market_size, adjusted_market_size]),
   pd.concat([penetration_rate, adjusted_penetration_rate]),
   pd.concat([investment, adjusted_investment])
)

axs[1, 1].plot(years[:5], future_scores[:5], marker='o',
   label='Historical')
axs[1, 1].plot(years[5:], future_scores[5:],
             marker='o', linestyle='--', label='Predicted')
axs[1, 1].set_title('Composite Score')
axs[1, 1].legend()

# Add grid and labels
for ax in axs.flat:
   ax.set_xlabel('Year')
   ax.set_ylabel('Value')
   ax.grid(True, alpha=0.3)

plt.tight_layout()
plt.show()

# Model evaluation

def evaluate_model_performance(actual_values, fitted_values):
   """
   Calculate and print the evaluation metrics of the model.

   Parameters:
   actual_values (array-like): Actual values.
   fitted_values (array-like): Model fitted values.
```

```
    Returns:
    dict: A dictionary containing various evaluation metrics.
    """
    mse = mean_squared_error(actual_values, fitted_values)
    mae = mean_absolute_error(actual_values, fitted_values)
    rmse = np.sqrt(mse)
    r2 = r2_score(actual_values, fitted_values)

    print("\nModel Performance Metrics:")
    print(f"Mean Squared Error (MSE): {mse:.2f}")
    print(f"Root Mean Squared Error (RMSE): {rmse:.2f}")
    print(f"Mean Absolute Error (MAE): {mae:.2f}")
    print(f"R-squared (R): {r2:.4f}")

    return {
        "MSE": mse,
        "RMSE": rmse,
        "MAE": mae,
        "R-squared": r2
    }

ms_eval = evaluate_model_performance(market_size.values,
    ms_model.fittedvalues.values)
pr_eval = evaluate_model_performance(penetration_rate.values,
    pr_model.fittedvalues.values)
inves_eval = evaluate_model_performance(investment.values,
    inves_model.fittedvalues.values)

print(ms_eval)
print(pr_eval)
print(inves_eval)
```

Listing 2: The code of question2

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
# read the data from the three countries
usa_data = pd.read_excel('./data/attachment2_America.xlsx')

fra_data = pd.read_excel('./data/attachment2_France.xlsx')

ger_data = pd.read_excel('./data/attachment2_Germany.xlsx')

usa_data

usa_data['Year']

# Create two side-by-side subplots
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(20, 8))

# Set the width and position of the bars
bar_width = 0.25
years = usa_data['Year']
r1 = np.arange(len(years))
r2 = [x + bar_width for x in r1]
r3 = [x + bar_width for x in r2]

# First subplot: Plot stacked bar charts
# USA data
usa_cats = ax1.bar(r1, usa_data['Cat'], bar_width, label='USA Cat',
    color='lightblue')
usa_dogs = ax1.bar(r1, usa_data['Dog'], bar_width,
    bottom=usa_data['Cat'], label='USA Dog', color='skyblue')

# France data
fra_cats = ax1.bar(r2, fra_data['Cat'], bar_width, label='France Cat',
    color='lightgreen')
fra_dogs = ax1.bar(r2, fra_data['Dog'], bar_width,
    bottom=fra_data['Cat'], label='France Dog', color='forestgreen')

# Germany data
ger_cats = ax1.bar(r3, ger_data['Cat'], bar_width, label='Germany Cat',
```

```
    color='lightcoral')
ger_dogs = ax1.bar(r3, ger_data['Dog'], bar_width,
    bottom=ger_data['Cat'], label='Germany Dog', color='darkred')

# Add numerical labels
def add_labels(ax, bars, bottoms=None):
    for i, bar in enumerate(bars):
        height = bar.get_height()
        if bottoms is not None:
            y_pos = bottoms[i] + height/2
        else:
            y_pos = height/2
        ax.text(bar.get_x() + bar.get_width()/2., y_pos,
                f'{int(height)}',
                ha='center', va='center', fontsize=10)

# Add labels for each group of bars
add_labels(ax1, usa_cats)
add_labels(ax1, usa_dogs, usa_data['Cat'].values)
add_labels(ax1, fra_cats)
add_labels(ax1, fra_dogs, fra_data['Cat'].values)
add_labels(ax1, ger_cats)
add_labels(ax1, ger_dogs, ger_data['Cat'].values)

# Set the labels and title for the first subplot
ax1.set_xticks([r + bar_width for r in range(len(years))])
ax1.set_xticklabels(years)
ax1.set_ylabel('Number of Pets', fontsize=12)
ax1.set_title('Pet Population Structure by Country', fontsize=13)
ax1.legend(loc='upper right')

# Second subplot: Plot growth rate line chart
# Calculate the growth rate for cats and dogs separately
def calculate_growth_rate(data):
    return [(data.iloc[i] - data.iloc[i-1])/data.iloc[i-1] * 100 for i
        in range(1, len(data))]
```

```
# Calculate the growth rate for cats and dogs in each country
usa_cat_growth = calculate_growth_rate(usa_data['Cat'])
usa_dog_growth = calculate_growth_rate(usa_data['Dog'])
fra_cat_growth = calculate_growth_rate(fra_data['Cat'])
fra_dog_growth = calculate_growth_rate(fra_data['Dog'])
ger_cat_growth = calculate_growth_rate(ger_data['Cat'])
ger_dog_growth = calculate_growth_rate(ger_data['Dog'])

# Plot the growth rate line chart
years_growth = years[1:]
ax2.plot(years_growth, usa_cat_growth, 'o-', label='USA Cat Growth',
    color='lightblue', linewidth=3)
ax2.plot(years_growth, usa_dog_growth, 'o--', label='USA Dog Growth',
    color='skyblue', linewidth=3)
ax2.plot(years_growth, fra_cat_growth, 's-', label='France Cat Growth',
    color='lightgreen', linewidth=3)
ax2.plot(years_growth, fra_dog_growth, 's--', label='France Dog
    Growth', color='forestgreen', linewidth=3)
ax2.plot(years_growth, ger_cat_growth, '^-', label='Germany Cat
    Growth', color='lightcoral', linewidth=3)
ax2.plot(years_growth, ger_dog_growth, '^--', label='Germany Dog
    Growth', color='darkred', linewidth=3)

# Set labels and title for the second subplot
ax2.set_xticks([2020,2021,2022,2023])
ax2.set_ylabel('Growth Rate (%)', fontsize=12)
ax2.set_title('Pet Population Growth Rate by Country', fontsize=13)
ax2.legend(loc='upper right')
ax2.grid(True, linestyle='--', alpha=0.7)

# Adjust layout
plt.tight_layout()
plt.show()

# Print growth rate data
print("\nPet population annual change rates by country:")
print("\nUSA Cat Growth Rate:", [f"{x:.2f}%" for x in usa_cat_growth])
```

```
print("USA Dog Growth Rate:", [f"{x:.2f}%" for x in usa_dog_growth])
print("\nFrance Cat Growth Rate:", [f"{x:.2f}%" for x in
    fra_cat_growth])
print("France Dog Growth Rate:", [f"{x:.2f}%" for x in fra_dog_growth])
print("\nGermany Cat Growth Rate:", [f"{x:.2f}%" for x in
    ger_cat_growth])
print("Germany Dog Growth Rate:", [f"{x:.2f}%" for x in ger_dog_growth])

# Uniform currency unit
exchange_rate = 1.1 # Exchange rate of Euro to US Dollar
for df in [fra_data, ger_data]:
    df['Market Size'] = df['Market Size'] * exchange_rate
    df['Total expenditures in the pet industry'] = df['Total
        expenditures in the pet industry'] * exchange_rate

# Merge data
all_data = pd.concat([usa_data, fra_data, ger_data], keys=['USA',
    'France', 'Germany'])
all_data

# Normalization function
def normalize(series):
    return (series - series.min()) / (series.max() - series.min())

# Calculate composite score
def calculate_composite_score(data):

    # Normalize three indicators
    norm_market_size = normalize(data['Market Size'])
    norm_penetration_rate = normalize(data['Household Penetration Rate'])
    norm_expenditure = normalize(data['Total expenditures in the pet
        industry'])

    # Calculate composite score
    weights = {'market_size': 0.4, 'penetration_rate': 0.3,
        'expenditure': 0.3}
    data['Composite Score'] = (
```

```
        norm_market_size * weights['market_size'] +
        norm_penetration_rate * weights['penetration_rate'] +
        norm_expenditure * weights['expenditure']
    )

    # Group by year to calculate market share
    yearly_totals = data.groupby('Year').agg({
        'Cat': 'sum',
        'Dog': 'sum'
    })

    # Calculate the market share for each country in each year
    for year in data['Year'].unique():
        year_mask = data['Year'] == year
        total_cats = yearly_totals.loc[year, 'Cat']
        total_dogs = yearly_totals.loc[year, 'Dog']

        data.loc[year_mask, 'Cat Market Share'] = data.loc[year_mask,
            'Cat'] / total_cats
        data.loc[year_mask, 'Dog Market Share'] = data.loc[year_mask,
            'Dog'] / total_dogs

    return data

# Calculate scores
all_scores = calculate_composite_score(all_data)

# View results
all_scores

# Extract data for each country from all_scores
usa_scores = all_scores.xs('USA')
fra_scores = all_scores.xs('France')
ger_scores = all_scores.xs('Germany')

# Set up the plot
fig, ax1 = plt.subplots(figsize=(15, 8))
```

```
# Set the width and position of the bar chart
bar_width = 0.25
r1 = np.arange(len(usa_scores))
r2 = [x + bar_width for x in r1]
r3 = [x + bar_width for x in r2]

# Plot the composite score bar chart
bars1 = ax1.bar(r1, usa_scores['Composite Score'], bar_width,
    label='USA Composite Score', color='lightblue')
bars2 = ax1.bar(r2, fra_scores['Composite Score'], bar_width,
    label='France Composite Score', color='lightgreen')
bars3 = ax1.bar(r3, ger_scores['Composite Score'], bar_width,
    label='Germany Composite Score', color='lightcoral')

ax1.set_xlabel('Year', fontsize=13)
ax1.set_ylabel('Composite Score', fontsize=13)
ax1.set_xticks([r + bar_width for r in range(len(usa_scores))])
ax1.set_xticklabels(usa_scores['Year'])
# ax1.set_title('Pet Industry Composite Score and Market Share
    (2019-2023)')

# Mark the score values on the bar chart
def add_bar_labels(bars):
    for bar in bars:
        yval = bar.get_height()
        ax1.text(bar.get_x() + bar.get_width()/2, yval + 0.01,
            f'{yval:.3f}', ha='center', va='bottom', fontsize=12)

add_bar_labels(bars1)
add_bar_labels(bars2)
add_bar_labels(bars3)

# Create a second y-axis
ax2 = ax1.twinx()

# Plot the line chart for cat and dog market share
```

```
line1, = ax2.plot(r1, usa_scores['Cat Market Share'], color='blue',
    marker='o', linestyle='-', label='USA Cat Market Share')
line2, = ax2.plot(r1, usa_scores['Dog Market Share'], color='orange',
    marker='o', linestyle='-', label='USA Dog Market Share')
line3, = ax2.plot(r2, fra_scores['Cat Market Share'],
    color='darkgreen', marker='s', linestyle='-', label='France Cat
    Market Share')
line4, = ax2.plot(r2, fra_scores['Dog Market Share'], color='red',
    marker='s', linestyle='-', label='France Dog Market Share')
line5, = ax2.plot(r3, ger_scores['Cat Market Share'], color='purple',
    marker='^', linestyle='-', label='Germany Cat Market Share')
line6, = ax2.plot(r3, ger_scores['Dog Market Share'], color='brown',
    marker='^', linestyle='-', label='Germany Dog Market Share')

ax2.set_ylabel('Market Share', fontsize=13)

# Combine legends
lines, labels = ax1.get_legend_handles_labels()
lines2, labels2 = ax2.get_legend_handles_labels()
ax1.legend(lines + [line1, line2, line3, line4, line5, line6], labels +
    labels2, loc='upper left')

# Add grid lines
ax1.grid(True, alpha=0.3)

# Display the chart
plt.tight_layout()
plt.show()

data = pd.read_excel('./data/attachment2_world.xlsx')
data.rename(columns={'Pet food sales worldwide': 'Sales'}, inplace=True)
data

from scipy import interpolate
import matplotlib.pyplot as plt
from statsmodels.tsa.holtwinters import ExponentialSmoothing
# Interpolation processing
```

```python
def interpolate_missing(data):
    # Create interpolation function
    f = interpolate.interp1d(
        data[data['Sales'].notna()]['Year'],
        data[data['Sales'].notna()]['Sales'],
        kind='cubic'
    )

    # Interpolate missing values
    missing_years = data[data['Sales'].isna()]['Year']
    interpolated_values = f(missing_years)
    data.loc[data['Year'].isin(missing_years), 'Sales'] =
        interpolated_values

    return data, missing_years, interpolated_values

# Perform interpolation
data, missing_years, interpolated_values = interpolate_missing(data)

print(missing_years, interpolated_values)

# Forecast using the Holt-Winters model
def forecast_sales(data, forecast_periods=3):
    # Create time series model
    model = ExponentialSmoothing(
        data['Sales'],
        seasonal_periods=4,
        trend='add',
        seasonal='add'
    )

    # Fit the model
    fitted_model = model.fit()

    # Forecast future values
    forecast = fitted_model.forecast(forecast_periods)
```

```
    return forecast, fitted_model

# Make predictions
future_years = [2024, 2025, 2026]
forecast, fitted_model = forecast_sales(data)

# Visualize interpolation and forecast results
plt.figure(figsize=(12, 6))
plt.plot(data['Year'], data['Sales'], 'b-', label='Historical Data')
plt.scatter(missing_years, interpolated_values, color='green',
    marker='x', s=200, label='Interpolated Points')
plt.plot(future_years, forecast, 'r--', label='Forecast')
plt.scatter(future_years, forecast, color='red')

# Add forecast value labels
for i, year in enumerate(future_years):
    plt.annotate(f'{forecast.iloc[i]:.1f}',
               (year, forecast.iloc[i]),
               textcoords="offset points",
               xytext=(0,10),
               ha='center', fontsize=12)

# plt.title('Pet Food Sales Worldwide: Interpolation and Forecast')
plt.xlabel('Year', fontsize=13)
plt.ylabel('Sales (Billion USD)', fontsize=13)
plt.legend()
plt.grid(True)
plt.show()

# Print forecast results
print("\nForecast Results:")
for year, value in zip(future_years, forecast):
    print(f"Year {year}: {value:.1f} Billion USD")

from sklearn.metrics import mean_squared_error, mean_absolute_error,
    r2_score
# Calculate model evaluation metrics
```

```
actual_values = data['Sales'].values
fitted_values = fitted_model.fittedvalues.values

mse = mean_squared_error(actual_values, fitted_values)
mae = mean_absolute_error(actual_values, fitted_values)
rmse = np.sqrt(mse)
r2 = r2_score(actual_values, fitted_values)

print("\nModel Performance Metrics:")
print(f"Mean Squared Error (MSE): {mse:.2f}")
print(f"Root Mean Squared Error (RMSE): {rmse:.2f}")
print(f"Mean Absolute Error (MAE): {mae:.2f}")
print(f"R-squared (R): {r2:.4f}")
```

Listing 3: The code of question3

```
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_excel('./data/attachment3.xlsx')
data

# Convert production value from RMB to USD
data['Production_in_dollars'] = data['Production'] * 0.14

# Calculate the export to production ratio
data['Export_to_Production_Ratio'] = data['Exports'] /
    data['Production_in_dollars']

# Create the chart
fig, ax1 = plt.subplots(figsize=(10, 6))

# Plot the trends of production value and export value
line1, = ax1.plot(data['Year'], data['Production_in_dollars'],
    marker='o', color='b', label='Production (in billion $)')
line2, = ax1.plot(data['Year'], data['Exports'], marker='o', color='g',
    label='Exports (in billion $)')
ax1.set_xlabel('Year')
```

```
ax1.set_xticks(range(2019, 2024))
ax1.set_ylabel('Value (billion $)', color='b', fontsize=13)
ax1.tick_params(axis='y', labelcolor='b')
# ax1.set_title('Trend of Production and Exports & Export to Production
    Ratio')

# Create a second y-axis for plotting the ratio
ax2 = ax1.twinx()
line3, = ax2.plot(data['Year'], data['Export_to_Production_Ratio'],
    marker='x', color='r', label='Export/Production Ratio')
ax2.set_ylabel('Export to Production Ratio', color='r', fontsize=13)
ax2.tick_params(axis='y', labelcolor='r')

# Combine legends
lines = [line1, line2, line3]
labels = [line.get_label() for line in lines]
ax1.legend(lines, labels, loc='upper center')

# Show grid
ax1.grid(True)

# Show the chart
plt.show()

from statsmodels.tsa.api import VAR
from statsmodels.tsa.stattools import adfuller
from statsmodels.stats.diagnostic import acorr_ljungbox
import numpy as np

# Set time series index
data.set_index('Year', inplace=True)

features = ['Production_in_dollars', 'Exports', 'Global_Food',
    'China_Cat', 'China_Dog', 'USA_Cat', 'USA_Dog', 'France_Cat',
    'France_Dog', 'German_Cat', 'German_Dog']
model_data = data[features]
```

```
# ADF test function
def adf_test(series, title=''):
    """Perform ADF test and print results."""
    result = adfuller(series.dropna(), autolag='AIC')
    return result[1] <= 0.05 # Return True if the series is stationary

# Check the stationarity of each feature and apply differencing to
    non-stationary series
data_diff = model_data.copy()
for column in model_data.columns:
    is_stationary = adf_test(model_data[column], title=column)
    if not is_stationary:
        data_diff[column] = model_data[column].diff().dropna()

# Drop the first row with NaN values (due to differencing)
data_diff = data_diff.dropna()

data_diff

# Fit VAR model using differenced data
model = VAR(data_diff)
model_fitted = model.fit(maxlags=1)

# Extract residuals
residuals = model_fitted.resid

residuals

# Model evaluation
# Use Ljung-Box test to check autocorrelation of residuals for each
    variable
ljung_box_results = {}
print("Ljung-Box Test Results:")
for column in residuals.columns:
    lb_test_results = acorr_ljungbox(residuals[column], lags=[1],
        return_df=True)
    ljung_box_results[column] = lb_test_results.iloc[0].to_dict()
```

```
    print(f"\n{column} residuals:")
    print(lb_test_results)

# Save Ljung-Box test results to a DataFrame
ljung_box_df = pd.DataFrame(ljung_box_results).T
ljung_box_df.to_csv('./results/Q3/ljung_box_results.csv')
print("\nLjung-Box Test Results DataFrame:")
print(ljung_box_df)

# Forecast the next three years of differenced values
lag_order = model_fitted.k_ar
forecast_diff = model_fitted.forecast(data_diff.values[-lag_order:],
    steps=3)

# Convert the forecasted differenced values to actual values
forecast = model_data.values[-1] + np.cumsum(forecast_diff, axis=0)

# Convert the forecast results to a DataFrame
forecast_df = pd.DataFrame(forecast, index=[2024, 2025, 2026],
    columns=model_data.columns)

# Print the forecast results
print("\nFuture Production Predictions (in billions RMB):")
print(forecast_df['Production_in_dollars'])
print("\nFuture Exports Predictions (in billions USD):")
print(forecast_df['Exports'])

# Visualize the forecast results
plt.figure(figsize=(12, 6))
plt.plot(data.index, data['Production_in_dollars'], label='Actual
    Production', marker='o')
plt.plot(forecast_df.index, forecast_df['Production_in_dollars'],
    label='Forecasted Production', marker='x', linestyle='--')

plt.plot(data.index, data['Exports'], label='Actual Exports',
    marker='o')
plt.plot(forecast_df.index, forecast_df['Exports'], label='Forecasted
```

```
    Exports', marker='x', linestyle='--')

# Annotate forecast values
for year in forecast_df.index:
    plt.annotate(f"{forecast_df['Production_in_dollars'].loc[year]:.1f}",
                 (year, forecast_df['Production_in_dollars'].loc[year]),
                 textcoords="offset points", xytext=(-10,5), ha='center',
                    fontsize=12)
    plt.annotate(f"{forecast_df['Exports'].loc[year]:.1f}",
                 (year, forecast_df['Exports'].loc[year]),
                 textcoords="offset points", xytext=(-10,5), ha='center',
                    fontsize=12)

plt.xlabel('Year')
plt.ylabel('Value (billion $)', fontsize=12)
# plt.title('Production and Exports: Actual vs Forecasted')
plt.legend()
plt.grid(True)
plt.show()
```

Listing 4: The code of question 4

```
import pandas as pd
import statsmodels.api as sm
import statsmodels.formula.api as smf

# Import data of the United States and Germany
usa = pd.read_excel("./data/q4.xlsx", sheet_name="America")
germany = pd.read_excel("./data/q4.xlsx", sheet_name="German")
france = pd.read_excel("./data/q4.xlsx", sheet_name="France")

# Define a function to process data and build models
def process_data(df, country_name):
    # Convert the time to date format
    df['time'] = pd.to_datetime(df['time'], format='%Y%m')

    # Add interaction term
    df['treated'] = df['group'].apply(lambda x: 1 if x == 'treatment'
```

```
    else 0)
df['treated_post'] = df['treated'] * df['policy']

# Build DID models for export amount
model_amount = smf.ols('export_amount ~ treated + policy +
    treated_post', data=df).fit()

# Build DID models for export value
model_value = smf.ols('export_value ~ treated + policy +
    treated_post', data=df).fit()

# Output model results
print(f"Results for {country_name}:")
print(model_amount.summary())
print(model_value.summary())

# Export model results to CSV and LaTeX
# Extract table data from summary
results_table_amount = model_amount.summary().tables[1]
results_table_value = model_value.summary().tables[1]

# Convert to DataFrame
df_results_amount = pd.DataFrame(results_table_amount.data[1:],
    columns=results_table_amount.data[0])
df_results_value = pd.DataFrame(results_table_value.data[1:],
    columns=results_table_value.data[0])

# Write DataFrame to CSV file
df_results_amount.to_csv(f'./results/Q4/{country_name}_regression_results_amount.csv',
    index=False)
df_results_value.to_csv(f'./results/Q4/{country_name}_regression_results_value.csv',
    index=False)

# Write summary to LaTeX file
summary_latex_amount = model_amount.summary().as_latex()
summary_latex_value = model_value.summary().as_latex()
```

```
    with
        open(f'./results/Q4/{country_name}_regression_summary_amount.tex',
        'w') as file:
        file.write(summary_latex_amount)
    with
        open(f'./results/Q4/{country_name}_regression_summary_value.tex',
        'w') as file:
        file.write(summary_latex_value)

# Process data of the United States
process_data(usa, "USA")

# Process data of Germany
process_data(germany, "Germany")

# Process data of France
process_data(france, "France")
```