

Sustainable Insuring and Developing Strategies Based on Evaluative and Predicting Model

Summary

Affected by climate change and extreme weather, the environment of the insurance market has changed significantly, leading many insurers to withdraw from covering certain regions. While the change in the insurance landscape has forced communities and property developers to be more advisable when building real estate, it has also resulted in historic cultural landmarks in some communities facing greater risks. By combining data with actual situations, we have developed **decision making mathematical models** to offer targeted advice to insurers, developers, and community leaders.

For insurers, their profitability in a region primarily depends on the per capita payment capability of that region and the risk of meteorological disasters occurring there. To calculate and predict the profitability of insurance in a region, we establish a **insuring model** by defining a **profitability index** by dividing the per capita GDP of the selected region by the weighted sum of data related to meteorological disasters in that region. The weights of various meteorological disaster data are derived using the **Analytic Hierarchy Process (AHP)**. After this, we use the **ARIMA time series model** to predict changes in the profitability index for the region in the coming years. By comparing these predictions with previous indexes and considering the region's prior insurance profitability, we propose strategies for whether or not to provide coverage.

When it comes to the recommendation for communities and property developers, the model established for insurers may be slightly adapted. We adjusted some variables and weights to fit for different priorities of insurers and developers in decision-making scenarios. Furthermore, we optimized for specific areas, calculating a new **communities and property development index**, providing developers with suggestions for site selection.

In the last problem, we established a new preserving model for historic or cultural landmarks, which are considered not worth insuring, to assess the protection value of them, offering community leaders appropriate preservation recommendations. Seven indexes are put forward to measure factors such as the age, type, economic, reputational benefits, risk level and community influence. We then established a **Multilayer Perceptron model**, a type of neural network model to assess the protection value of a building and provide suggestions.

Although this model has some drawbacks including weak cross-regional comparison capabilities and looser calculation of some variable indexes, it can still provide relatively accurate advice to insurers, developers, and community leaders against the backdrop of continually worsening extreme weather conditions.

Keywords: Decision Making, Analytic Hierarchy Process, ARIMA Model, Multilayer Perception Model

Contents

1	Introduction	3
1.1	Background	3
1.2	Restatement of the Problem	4
1.3	Our Work	4
2	Assumptions and Justification	5
3	Notations	5
4	Model Preparation	5
4.1	Data Cleaning and Integration	6
4.2	Missing Data Processing	6
5	Taks 1: Underwriting Strategy Prediction	6
5.1	Overview	6
5.2	Calculation of Profit Index Based on Analytic Hierarchy Process	7
5.2.1	Data Preprocessing	8
5.2.2	Priority Setting	8
5.2.3	Synthesis of Priorities	8
5.2.4	Consistency Verification	9
5.3	Profit Index Prediction Based on ARIMA Model	9
5.3.1	The Fundamental of ARIMA Model	9
5.3.2	Application to the Calculated Index	10
6	Task 2: Adapted Model for Community and Property Developer	11
6.1	Drawbacks of the Current Model in Developer Scenario	12
6.2	Adaptation of the Model	12
6.2.1	Modification of the Variables	12
6.2.2	Modification of the Concrete Model	12
6.2.3	Evaluation of construction conditions in the United States	14
7	Task 3: Developing a Preservation Model	16
7.1	Problem Analysis	16
7.2	Model Construction	16
7.2.1	Data Collection	16
7.2.2	Data Preprocessing	18
7.2.3	Model Selection & Training	18
7.2.4	Model Evaluation	19
8	Task 4: Recommendation Letter for Community Leader	19
9	Strength and Weakness	19
9.0.1	Strength of Our Research	19
9.0.2	Weakness of Our Research	20
References		20

1 Introduction

1.1 Background

In recent years, influenced by climate change, the frequency of extreme-weather events has increased worldwide. Extreme weather poses a headache for both insurers and property owners. For insurers, an increase in the frequency of extreme weather inevitably leads to higher claim payouts, which in turn result in a reduction in insurers' willingness to launch insurance services. On the other hand, for property owners in specific parts of the world, the diminishing availability of insurance services means their properties being less protected. The shared dilemma of insurance company profitability and property owners' affordability is becoming more pronounced against the backdrop of climate change.

For insurance companies, underwriting too little leads to a lack of customer volume, while underwriting too much results in excessively high claim payouts, both of which can lead to a business crisis. For property owners, it becomes particularly important to conduct real estate construction and other developments more advisably under a deteriorating insurance environment. Moreover, beyond commercial properties, how properties with cultural or community significance in a certain district can better face the insurance crisis is a further question worth pondering. In the following thesis, we will be dedicated to developing mathematical models to provide recommendations for insurers, property owners, and community leaders.

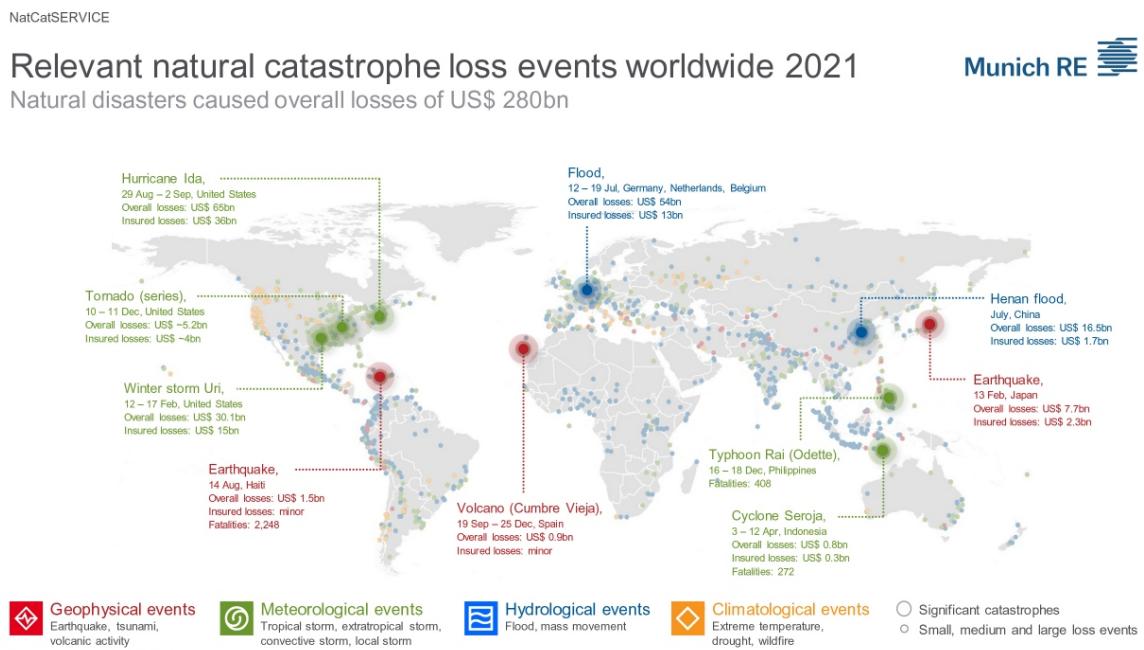


Figure 1: Natural Catastrophe-Related Loss Worldwide 2021

1.2 Restatement of the Problem

Extreme-weather events post new challenges on both insurers and property owners. Through in-depth analysis and research on the background of the problem, combined with the specific constraints given, the restate of the problem can be expressed as follows:

- Establish a mathematical model to quantify the impact of extreme weather on insurance claims. Take other factors into consideration, and provide underwriting recommendations for insurers.
- Adapt the previous model to the communities and property developers' scenario. Comprehensively assess the risk and summarize strategies for communities and property developers during development and construction.
- For regions where the result in Problem 1 is "not to underwrite", establish a model to assess the value of retaining specific buildings and propose corresponding preserving measures.
- Select a historical landmark. Use the previous insurance and preserving models to evaluate the value of the building, and write a letter to the local community about preserving the landmark.

1.3 Our Work

In this paper, we offer advice to people in different industries through the establishment and operation of mathematical models. We use the Analytic Hierarchy Process (AHP) and ARIMA time series models to provide insurers and developers with advice on underwriting and real estate construction. We use the Multilayer Perception neural network model to offer landmark protection advice to community leaders. Additionally, we communicate our recommendations to the community through a letter. Below is a picture showcasing the workflow.

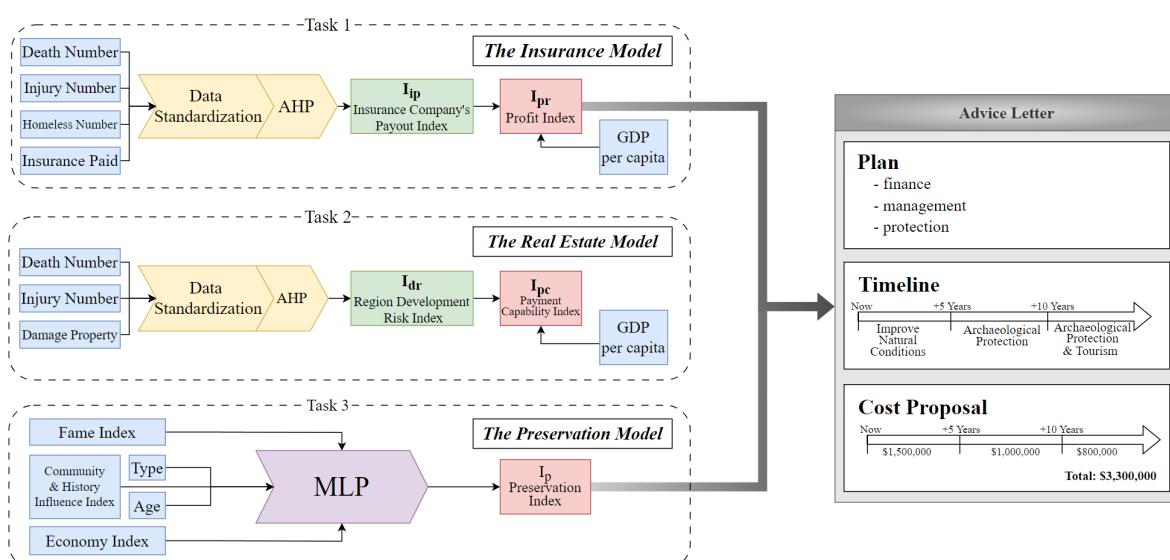


Figure 2: Workflow

2 Assumptions and Justification

Assumption 1: The impacts of natural disasters and other factors impacting the profitability of insurers and developers will change steadily such that past data can be used to infer future trends.

Justification: As there is no evidence to suggest that the trends in global climate and economic development will undergo significant changes in the coming years, we assume that past natural disaster and economic data can be used to infer future changes [1].

Assumption 2: The overall situation of the insurance industry will not be affected by other uncontrollable factors in the coming years.

Justification: Since this paper only evaluates and predicts the situation in the coming years, the impact of external factors such as policy changes and inflation on insurance underwriting and payouts is ignored.

Assumption 3: The overall economic development of a region can reflect the purchasing capability of the residents for insurance in that area.

Justification: Due to the inability to accurately obtain the overall insurance costs for all regions, and given that research indicates that the average insurance premiums paid by residents of a region have a nearly linear non-linear correlation with the per capita income of that region [2], we believe that the insurance expense situation in the region can be quantified by per capita income or GDP per capita.

Additional assumptions are made to simplify analysis for individual sections. These assumptions will be discussed at the appropriate locations.

3 Notations

Table 1: Notations

Symbol	Description	Symbol	Description
I_{pc}	Payment Capability Index	s_5/x_5	Original/standardized Property Loss
I_{ip}	Insurance Payout Index	K	Adjustment parameter of I_{pc}
I_{pr}	Profit Index	I_a	The age of a building
I_{dr}	Development Risk Index	I_t	The type of a building
I_{ps}	Preservation Index	I_e	Economy index of a building
s_1/x_1	Original/standardized Death data	I_f	Fame index of a building
s_2/x_2	Original/standardized Injury data	I_c	Community & history index of a building
s_3/x_3	Original/standardized Homelessness data	I_s	Severity index of a building
s_4/x_4	Original/standardized Insurance Payout	I_p	Preservation index of a building

*There are some variables that are not listed here and will be discussed in detail in each section.

4 Model Preparation

We have gathered extensive data over the past three to five decades concerning natural disasters and their consequential impacts in various countries, sourced through online research. The dataset encompasses figures representing individuals who have been directly or indirectly afflicted by a myriad of disaster categories. Additionally,

it documents the financial detriments spawned by such calamities, complemented by select annual records of insurance reimbursements. Considering the completeness of the data, we chose the United States and India as the subjects of problem 1. Following is the data preprocessing approach of problem 1, which is also applied similarly for problem 2 and 3. Other data preprocessing approach will be mentioned in the corresponding section.

4.1 Data Cleaning and Integration

We optimized the data we obtained. First, we cleaned up data not related to meteorological disasters, such as earthquakes. In addition, we eliminated some invalid data, such as various ratio data, costs of post-disaster reconstruction, etc. At the same time, we integrated the existing data. For research needs and to reduce workload, we directly calculated the total casualties and economic losses caused by all meteorological disasters.

4.2 Missing Data Processing

In the collected data, there are inevitably some missing values. For missing data in the number of casualties, we directly interpolated them. As for insured economic losses, due to their more significant absence and their relevance to total economic losses, with substantial annual variability, we calculated the ratio of insured losses to total economic losses for the years with available data. We then interpolated these ratios and multiplied the interpolated values by the total economic losses in the years with missing insurance data, obtaining a more complete dataset.

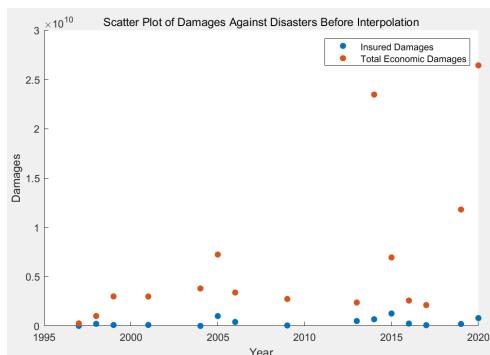


Figure 3: Data Before Processing

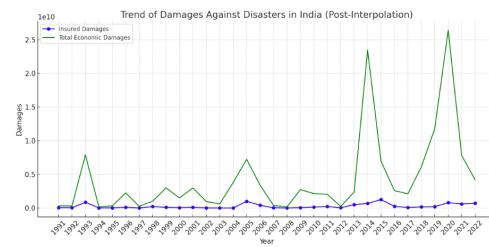


Figure 4: Data After Processing

5 Tasks 1: Underwriting Strategy Prediction

5.1 Overview

To address the first problem, we first use the Analytic Hierarchy Process (AHP) to derive an insurance payout index, determined by the local payment capability and the local meteorological disaster situation. After calculating an area's insurance payout index for the past several decades, we use the ARIMA model to predict the index for

the area in the coming years. By comparing this with previous insurance payout indexes and observing the profit situation of insurance companies in the area in previous years, we can then formulate an underwriting strategy. The process of this section is showcased in the below figure:

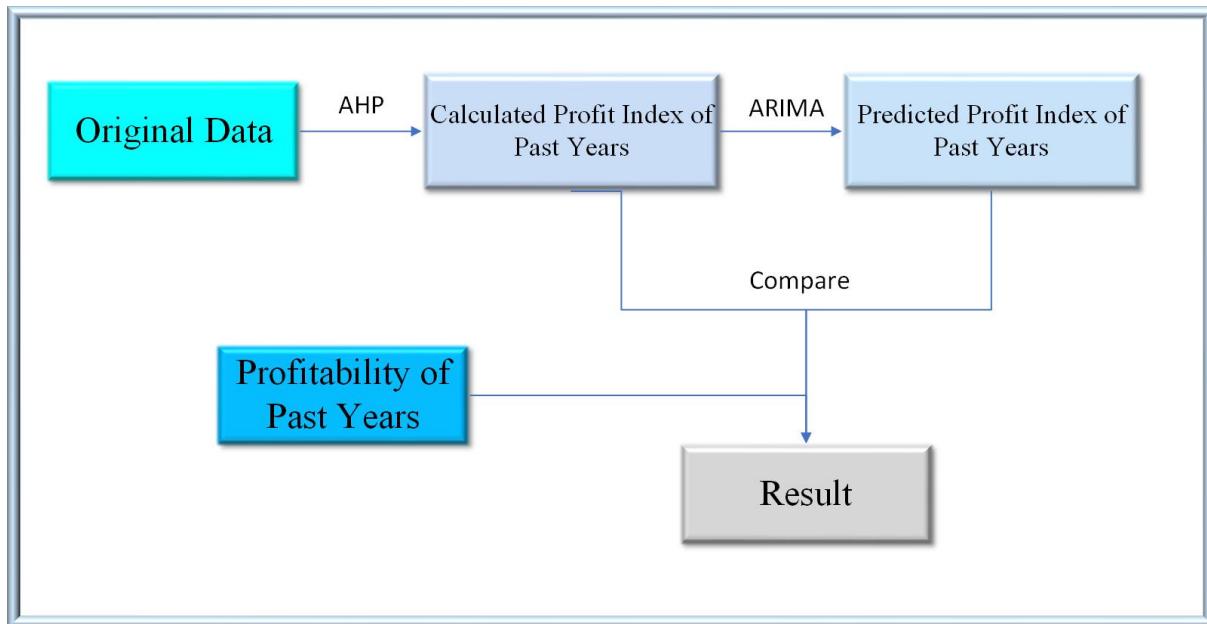


Figure 5: Overview of Section 5

5.2 Calculation of Profit Index Based on Analytic Hierarchy Process

The Analytic Hierarchy Process (AHP), is a structured technique for organizing and analyzing complex decisions. It represents an accurate approach for quantifying the weights of decision criteria and evaluating the alternatives that those criteria affect. AHP involves decomposing a complex multi-criteria decision problem into a hierarchy of more easily comprehended sub-problems, each of which can be analyzed independently. The process of employing AHP can be broken down into the following steps:

1. **Data Preprocessing:** Before entering the hierarchical analysis process, the data needs to be preprocessed to ensure its magnitudes are on the same level.
2. **Priority Setting:** Pairwise comparisons are made among the elements at each level of the hierarchy with respect to their impact on an element above them. These comparisons are used to establish priorities for the elements, often using a scale of relative importance from one to nine.
3. **Synthesis of Priorities:** The priorities are combined across the hierarchy levels to produce a set of overall priorities for the decision alternatives. This involves calculating eigenvectors for each comparison matrix to derive the local priorities, and then synthesizing these to determine global priorities.
4. **Consistency Verification:** The consistency of the pairwise comparisons is evaluated to ensure that the judgments have not been randomly assigned. AHP includes a consistency ratio that should typically not exceed 10%. If the ratio is higher, the pairwise comparisons should be reviewed and adjusted.

In this research, we will define a profit index. The profit index is determined by dividing the region's payment capability index (represented by per capita GDP) by the insurance payout index (determined by the Analytic Hierarchy Process). The insurance payout index will be calculated at first as below.

5.2.1 Data Preprocessing

For the data we use to calculate the profitability index, first, we need to standardize their magnitudes. To achieve this goal, data preprocessing is necessary. Therefore, we convert each piece of data into the ratio of the original data to the sum of all data in its column, thereby ensuring all data fall between 0 and 1.

Define (s_1, s_2, s_3, s_4) as original (Death, Injury, Homelessness, Insurance Payout) data

Define (x_1, x_2, x_3, x_4) as standardized (Death, Injury, Homelessness, Insurance Payout) data

(s_1, s_2, s_3, s_4) and (x_1, x_2, x_3, x_4) satisfying

$$x_i = \frac{s_i}{\sum_{y=1991}^{2023} s_{iy}} \quad (x_i \in [0, 1]) \quad (i = 1, 2, 3, 4)$$

5.2.2 Priority Setting

In pairwise priority setting, higher value stands for higher importance in comparison to the other factor. The priority comparison is shown as below:

Table 2: Pairwise Priority 1

	Death	Injury	Homelessness	Insurance Payout
Death	1	3	10	1/3
Injury	1/3	1	2	1/5
Homelessness	1/10	1/2	1	1/7
Insurance Payout	3	5	7	1

5.2.3 Synthesis of Priorities

1. Normalize the pairwise comparison matrix \mathbf{A} by dividing each element by the sum of its column, resulting in a normalized matrix \mathbf{N} .

$$n_{ij} = \frac{a_{ij}}{\sum_{k=1}^n a_{kj}}$$

2. Calculate the priority vector \mathbf{w} by averaging over the rows of the normalized matrix \mathbf{N} .

$$w_i = \frac{1}{n} \sum_{j=1}^n n_{ij}$$

3. Alternatively, the principal eigenvector of \mathbf{A} can be used as the priority vector. This is found by solving the equation $\mathbf{Aw} = \lambda_{\max} \mathbf{w}$, where λ_{\max} is the largest eigenvalue of \mathbf{A} .

Apply the algorithm to the previous matrix, we can conclude the weight of each factor.

Table 3: Weight of Factors

Death	Injury	Homelessness	Insurance Payout
0.3045378	0.0974034	0.04946848	0.54859031

5.2.4 Consistency Verification

The consistency of the matrix should be checked to ensure the reliability of the results. This involves calculating the Consistency Index (CI) and the Consistency Ratio (CR).

$$CI = \frac{\lambda_{\max} - n}{n - 1}$$

$$CR = \frac{CI}{RI}$$

where RI is the Random Index, a value that depends on the order of the matrix A . Apply the algorithm to the result, we can conclude the consistency factor as below:

$$CI = 0.06145380504128865$$

$$CR = 0.06828200560143183$$

which satisfy the consistency verification.

As we mentioned before, we define (x_1, x_2, x_3, x_4) as standardized (Death, Injury, Homelessness, Insurance Payout) data

Then the insurance payout index $I_{ip} = 0.30 * x_1 + 0.10 * x_2 + 0.05 * x_3 + 0.55 * x_4$

For the payment capability index, we similarly define $I_{pc} = K * GDP_{year}$ where K can be adjusted to accommodate the varying conditions of different regions.

$$I_{pr} = \frac{I_{pc}}{I_{ip}}$$

Apply this algorithm to the dataset including the relevant condition of India and the United States, we can obtain the profit index of past years of the two countries as showcased in the tabular 4 on the next page.

*the index can only reflect the condition in its specific region, which cannot be compared among regions

5.3 Profit Index Prediction Based on ARIMA Model

5.3.1 The Fundamental of ARIMA Model

Autoregressive Integrated Moving Average (ARIMA) models are widely used in time series analysis to forecast future points in the series. They are applied to univariate time series data, which is a series of measurements of the same variable collected over time. An ARIMA model is characterized by three terms: p , d , and q :

- p is the order of the Autoregressive (AR) part,
- d is the degree of differencing (the number of times the data have had past values subtracted),

Table 4: profit index of past years

Year	India	United States	Year	India	United States
1991	2.864820322	11.21039793	2007	6.807459526	14.16867303
1992	6.08143515	1.979539193	2008	9.287036557	3.509554829
1993	0.326049531	10.91624434	2009	14.04642012	12.68425497
1994	5.123417735	4.094575035	2010	9.479707768	16.78476521
1995	3.435417541	7.456435596	2011	8.075468361	3.845592268
1996	3.044080431	13.11353727	2012	28.74436885	3.256779208
1997	6.152173855	22.05878508	2013	2.852411005	9.238898668
1998	1.174388557	7.839101181	2014	3.320368031	11.47594657
1999	1.235511755	10.92630678	2015	1.805248035	8.792754043
2000	4.582168466	57.71465501	2016	7.425676779	4.965653138
2001	0.352673254	16.85571305	2017	18.78646699	1.268761522
2002	6.410478798	14.84620790	2018	13.72137113	3.367955104
2003	6.694110872	7.714675714	2019	11.26790885	16.52487294
2004	1.425600652	2.368312371	2020	3.432435304	3.447958418
2005	0.925193881	0.960023639	2021	5.212994196	2.321304838
2006	2.030563775	19.12436422	2022	4.851617458	2.143619283

- q is the order of the Moving Average (MA) part.

The AR part of ARIMA indicates that the evolving variable of interest is regressed on its own lagged (i.e., prior observed) values. The autoregressive part of an ARIMA model is defined as:

$$AR(p) : Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + \epsilon_t$$

where Y_t is the time series, $\alpha_1, \alpha_2, \dots, \alpha_p$ are the parameters of the model, and ϵ_t is white noise.

The integrated part of ARIMA represents the differencing of raw observations to make the time series stationary, i.e., data values are replaced by the difference between the data values and their previous values:

$$I(d) : \nabla^d Y_t = (1 - L)^d Y_t$$

where L is the lag operator, d is the order of differencing required to make the series stationary, and ∇^d is the differencing operator.

The moving average model is used to model the error term of the time series as a linear combination of error terms occurring contemporaneously and at various times in the past:

$$MA(q) : Y_t = \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + \cdots + \beta_q \epsilon_{t-q}$$

5.3.2 Application to the Calculated Index

The first step in ARIMA modeling is to ensure that the time series is stationary. This involves determining the optimal degree of differencing (d) to remove any trends or seasonality in the data, which can be done by ADF, ACF and PACF process. In order to obtain the most accurate predictive parameters p, d, q , we conducted an exhaustive search for these parameters. We utilized the first 75% of the data to construct an

ARIMA model for forecasting the remaining 25%, subsequently calculating the forecast variance. We selected the set of p, d, q parameters that resulted in the minimum for. After selection, the (p, d, q) values for the Indian dataset are $(2, 0, 1)$, while for the United States, they are $(5, 1, 2)$. Having determined the p, d, q values for India and the United States, we conducted an ARIMA forecast, the results of which are shown in the following table and figure:

Table 5: profit index of past years

Year	India	United States
2023	6.48762	6.52499
2024	6.11031	2.21465
2025	6.67457	1.19450

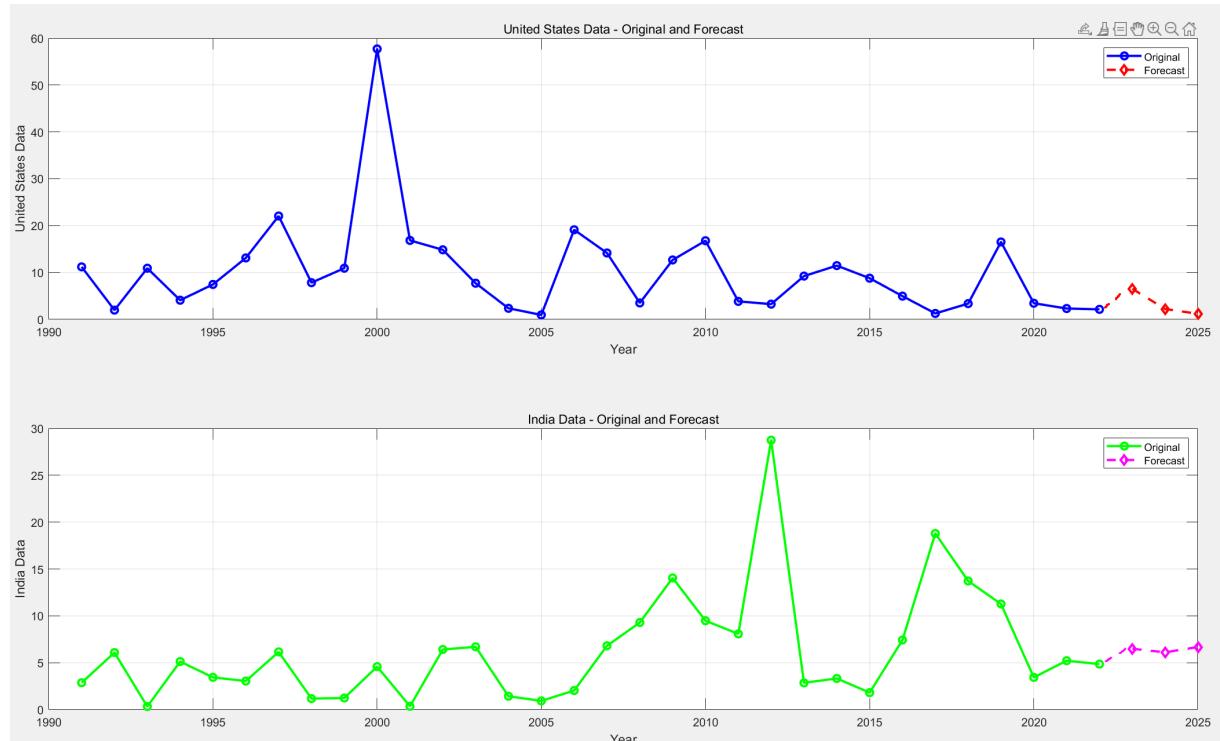


Figure 6: Predicted Profit Index

As shown in the chart, the insurance profitability index for India is steadily increasing. Given the previous profit situation, insurance companies can continue underwriting. On the other hand, the insurance profitability index for the United States is on a downward trend, which may be due to higher insurance payouts. Insurance companies could consider underwriting with appropriately adjusted payouts, or choose not to underwrite in some regions.

6 Task 2: Adapted Model for Community and Property Developer

In the background introduction of the paper, it was mentioned that the worsening meteorological disaster conditions pose challenges not only to insurance compa-

nies but also to communities and property developers. The troubles faced by the developers are to some extent similar to those of insurance companies. However, the challenges faced by developers differ in focus compared to those faced by insurers..Therefore, the model previously established can also be adapted to help the developers choose their cites. In addition, communities and real estate developers also face some other challenges, including changes in the insurance landscape, which also present new requirements for modifications to the model.

6.1 Drawbacks of the Current Model in Developer Scenario

As has been mentioned before, the challenges faced by the developers is slightly different from those faced by the insurers. Therefore, adapting and modifying the model to fit it into developers' scenario is necessary. Followings are the specific drawbacks of the current model

- The previous model was from the perspective of insurers, while the current model demands the perspective of developers, which naturally leads to shifts in variables that need to be considered. For example, in terms of property loss, the total loss amount rather than the insured loss needs to be considered.
- As perspectives and research objectives change, the weight of each variable factor as well need to be modified. For instance, the developers will lay more emphasis on the loss amount rather than the human casualty
- The previous model was designed to analyze the overall situation of a country (such as the United States, India), whereas the requirement for the second question is to focus on a specific region (In this new model, we particularly target every state in the US) and provide suggestions for real estate and other construction projects. Therefore, detailed adjustments are necessary.

6.2 Adaptation of the Model

6.2.1 Modification of the Variables

First, we need to make some changes to the types of variables. For insurance companies, what they need to care about is the specific amount of compensation they pay, while for communities and property developers, they should be more concerned about the total economic loss. Therefore, we need to change the variable of insurance payout loss in the first model to total loss. Secondly, for developers, the number of homeless people only represents the loss of their real estate, which is already included in the total economic loss, which does not need to be calculated again. However, the data on casualties will have a more significant impact on the reputation of developers, so it also needs to be considered in the calculation.

After the modification of variables, Total Deaths, total injuries and total economic damage become the factors taken into consideration when modelling.

6.2.2 Modification of the Concrete Model

To make the model more applicable to developers' situations, we have reconstructed a paired priority matrix as below:

Table 6: Pairwise Priority 2

	Death	Injury	Property Loss
Death	1	5	1/2
Injury	1/5	1	1/8
Property Loss	2	8	1

The weights of (Death, Injury, Property Loss) turn out to be (0.325512460.070129530.604358)

$$CI = 0.0027675558692492608$$

$$CR = 0.00477164805042976$$

Because the data this time includes a large number of state data over the years, if the standardization method of the previous model is continued, the Property Loss will become very small after standardization (even less than 10^{-5}). In order to make the model more accurate, the standardization method of the three parameters is uniformly optimized.

Define (s_1, s_2, s_5) as original (Death, Injury, Property Loss) data

Define (x_1, x_2, x_5) as standardized (Death, Injury, Property Loss) data

(s_1, s_2, s_5) and (x_1, x_2, x_5) satisfying new relationship

$$x_i = \frac{s_i - \min(s_i)}{\max(s_i) - \min(s_i)} \times (R_{\max} - R_{\min}) + R_{\min} (x_i \in [0, 100]) (i = 1, 2, 5)$$

To avoid extreme tiny number here, we take $R_{\max} = 100$ $R_{\min} = 0$

Then the Development Risk Index can be calculated

$$I_{dr} = 0.33 * x_1 + 0.07 * x_2 + 0.60 * x_3$$

The payment capability of this region can be calculated

$$I_{pc} = \text{Average Individual Income} \times K$$

Average personal income can reflect both population and economic conditions. The greater the per capita personal income, the stronger the payout capability, the bigger the I_{pc} [3].

$$K = 1 + \text{Regional per capita income growth rate}$$

K theoretically varies from region to region, which can overcome some unknown affection. Here we can simply consider defining K as 1+Regional per capita income growth rate, which reflect regional economic development potentialities, thus perhaps effectively adjusting the results.

The Community and Property Development Potential Index can be calculated

$$I_{dp} = \frac{I_{pc}}{I_{dr}}$$

6.2.3 Evaluation of construction conditions in the United States

In order to meet the assessment needs of the communities and property developers regarding the development capability of a region, we have collected data on all the property losses, injuries, and fatalities caused by natural disasters such as droughts, hurricanes, extreme temperatures, etc. that occurred in the United States from 1950 to 2021.

After completing the standardization of data and using the evaluative model mentioned above, we can calculate the Development Risk Index for each state in the United States on an annual basis. We have extracted data for five states: Texas, California, Florida, New York, and Illinois, as depicted in the figure.

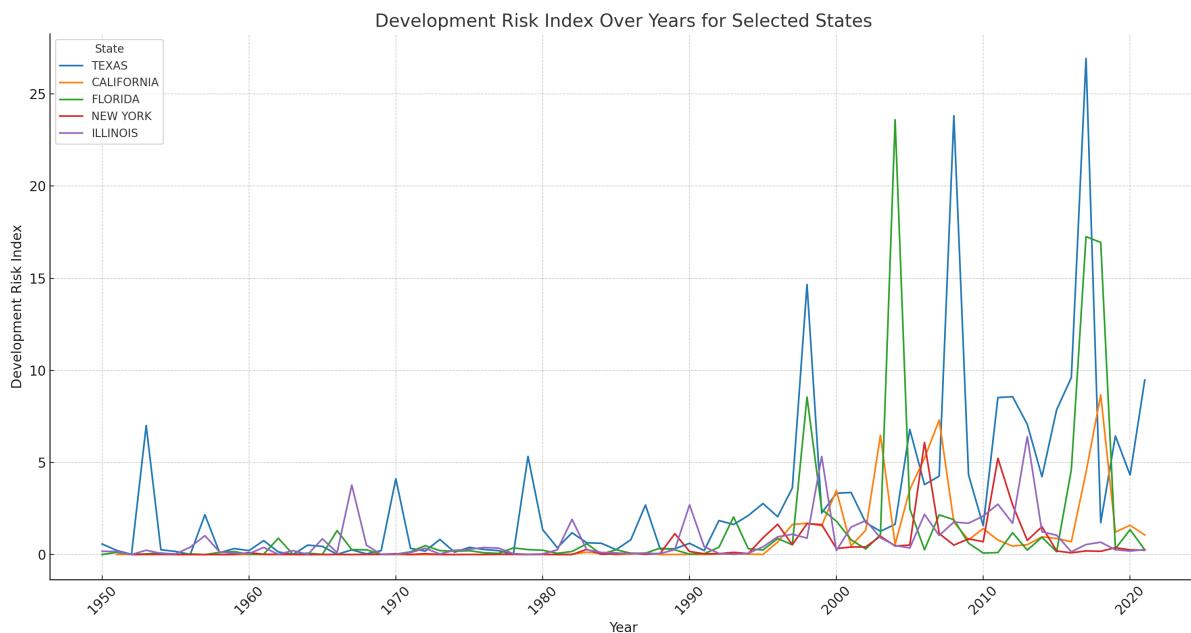


Figure 7: Development Risk Index over Years for Selected States

We have observed that the needs of real estate developers and insurance companies differ. Insurance companies only need to consider whether to provide coverage in the short term, whereas real estate developers must contemplate long-term development risks and returns. Therefore, in contrast to the original model, which predicted the Insurance Payout Index for the next three years, in our adjusted model, we have decided to forecast the Development Risk Index for the next ten years. The approach remains the same as the original model, using the ARIMA model and the same parameter adjustments. Furthermore, we select the maximum forecasted value of the I_{dr} as the assessment of the Development Risk Index for the next ten years.

After prediction of ARIMA model, we can print two bar charts of the post-forecast I_{dr} for some states, which can be used to assess the development risk of this region. Additionally, we attached some (p,d,q) parameters used by different states' model and concrete I_{dr} index to the top of the bar.

We used the model above, gathering the average disposable income of 2023 people in each U.S. state and divide this by I_{dr} and multiply it by the growth in per capita income for the region to get I_{dp} . It is used to measure the development potential of real estate and communities in the area. We chose to convert the I_{dp} values into colors to draw a thermodynamic diagram. (Converted RGB value $R = 255, G, B \propto \ln(I_{dp})$) The

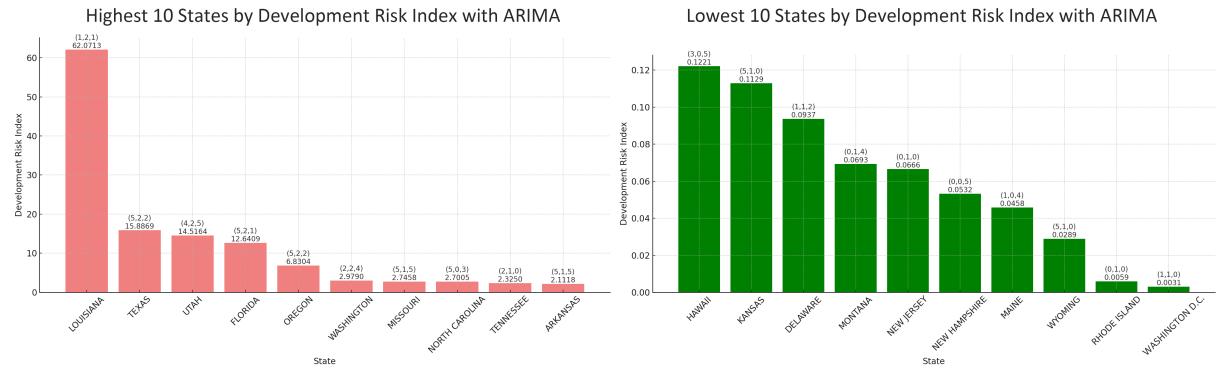


Figure 8: The 10 States with the Largest I_{dr} Value and the 10 States With the Smallest I_{dr} Value

closer you are to white, the less potential for development, and the closer you are to red, the greater the potential for development. This presentation is intuitive and easy to evaluate.

Developing Potential Thermodynamic Diagrams

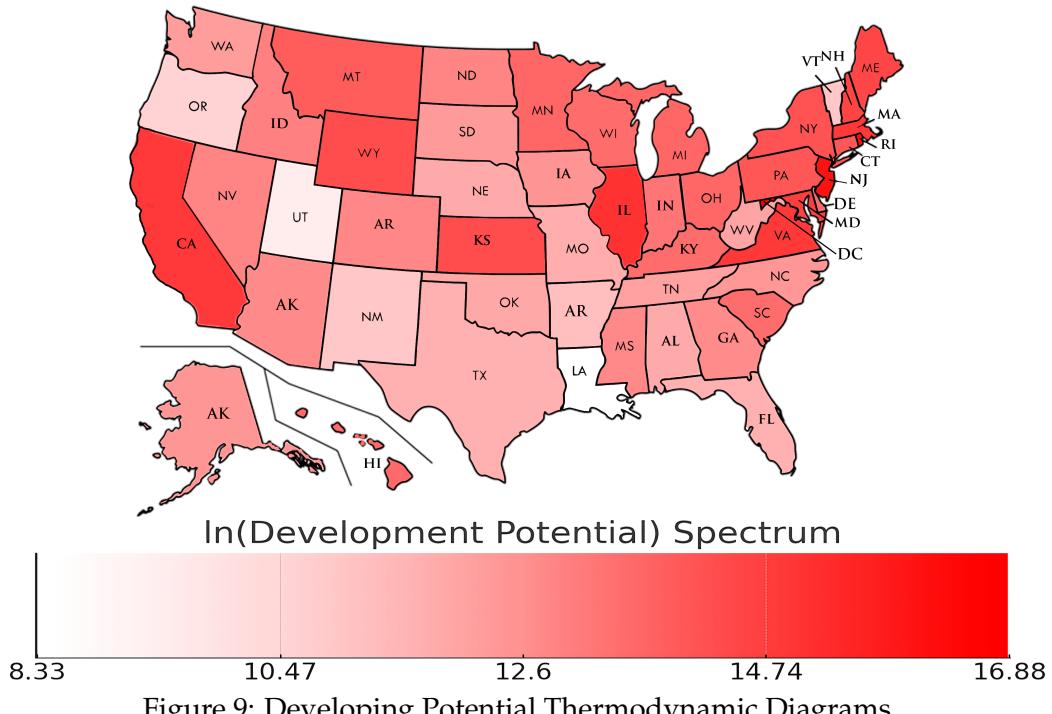


Figure 9: Developing Potential Thermodynamic Diagrams

In short, through the above adapted model, the calculated results can be directly provided to real estate and community investment evaluation. Using model-assisted evaluation, developers can pay attention to risk areas to avoid construction, identify high-risk areas, and avoid large-scale real estate development in these areas, while not slack in strengthening building standards, adjusting insurance rate adjustments, developing disaster response and recovery plans and taking responsibilities when disasters strike.

7 Task 3: Developing a Preservation Model

7.1 Problem Analysis

In the previous two models, we provided insurers, communities, and property developers with advice on underwriting and construction. However, changes in the insurance environment have introduced a new problem: many landmarks of historical and cultural significance in communities affected by meteorological disasters may be abandoned by insurers due to insufficient profitability. At this point, assessing the value of a building and designating a protection plan for it becomes very important.

7.2 Model Construction

7.2.1 Data Collection

When evaluating a building is worth preserving or not, it is important to quantify its cultural, historical, economic, and community significance. We collected 800 entries of data and processed them into following format:

Table 7: Example Data

Age	Type	Economy Index	Fame Index	Community & History Influence Index	Severity Index	Preserve
47	2	1	3	1	0	1
82	3	2	2	2	1	3
38	1	1	1	1	0	3
42	2	4	1	4	0	5
51	2	2	3	1	1	4
20	4	3	1	1	0	2
34	2	3	3	2	1	5
43	2	3	4	3	1	5

Here are explanations of different parameters:

- AGE I_a : The age of a building can reflect its historical value. It is determined by:

$$Age = 2024 - BuiltTime$$

- TYPE I_t : The type of a building can partially show the influence of this building.

Table 8: Building Type

Type	Notation
1	Apartment
2	Office Building
3	Tools: store/parks/church/government/hospital/...
4	Education: school/university/museum/...
5	Pure tourist attractions/landmarks/others/...

- ECONOMY INDEX I_e : This index of a building indicates its influence on the local economy.

Table 9: Economy Index

Index	Meaning	Notation	Examples
1	No Influence	This building cannot contribute to the local economy.	Free parks/Museums
2	Little Influence	This building contributes a little to the local economy.	Small stores/Restaurants
3	Moderate Influence	This building contributes a little to the local economy moderately.	Schools/Shopping malls
4	Significant Influence	This building contributes a little to the local economy significantly.	Office buildings
5	Top Influence	This building plays a leading role in the local economy.	Stock exchange/Factories

- FAME INDEX I_f : This index shows the building's historical or cultural value to the general public.

Table 10: Fame Index

Index	Meaning	Notation
1	Not Famous	This building is not known to the general public.
2	A little famous	This building is known only by a few people.
3	Somewhat famous	This building is known by some people.
4	Famous	This building is known by many people.
5	Very famous	This building is known by the whole general public.

- COMMUNITY & HISTORY INFLUENCE INDEX I_c : This index shows whether the building is important to the local or historical community.

Table 11: Community & History Influence Index

Index	Meaning	Notation
1	Not Important	This building is not important to the locals or historians. Most people don't need it.
2	A little important	This building is important only to a few locals or historians. Only a few people use it.
3	Somewhat important	This building is important to some locals or historians. Some people use it.
4	Important	This building is important to many locals or historians. Many people use it, or it has a historical value.
5	Very Important	This building is very important to most locals or historians. Most people cannot live without it, or it plays an important role in human history

- SEVERITY INDEX I_s : This index explains the condition of a building.

Table 12: Severity Index

Index	Meaning	Notation
0	Good condition	This building is well-maintained and is in good condition.
1	Bad condition	This building is in bad condition without proper maintenance. It might collapse if no attention is given to it.

- PRESERVATION INDEX I_p : This index shows the extent to which the building should be preserved.

Table 13: Preservation Index

Index	Meaning	Notation
1	Minimal attention	This building requires little attention. Extra protection isn't necessary.
2	Limited attention	Only limited attention should be paid to this building.
3	Moderate attention	Moderate attention should be paid to this building. Protection is important to its future development.
4	Important	The protection of this building should be emphasized. Appropriate preservation is important to its future development.
5	Emergency	This building urgently needs careful preservation. It might be destroyed without proper protection.

7.2.2 Data Preprocessing

Data standardization is performed here to ensure that the scale or range of different feature values does not adversely affect the training of machine learning models. During standardization, feature values (I_a, I_t, I_e, I_f, I_c) are rescaled to have a mean of 0 and a standard deviation of 1 group by group using this formula:

$$z = \frac{x - \mu}{\epsilon}$$

in which x is the raw data, μ is the mean of this group of data, and ϵ is the standard deviation of this group of data.

This helps to ensure that all features have similar scales, avoiding situations where some features dominate the model training due to large numerical values, or are ignored due to small numerical values. This also helps the model converge better, having a more smooth optimization process.

7.2.3 Model Selection & Training

Given that this task requires finding a relationship between the input combination I_a, I_t, I_e, I_f, I_c and the single output I_p , we choose to use MLP, Multilayer Perceptron model, to complete this task. It is a feedforward artificial neural network which contains fully connected neurons with a nonlinear activation function [3]. It is often utilized to distinguish data that is not linearly separable, finding a connection between multiple inputs and outputs.

After several adjustments, our MLP model contains 1 input layer, 4 hidden layers and 1 output layer. The number of neurons in each hidden layer is 50, 100, 100 and 100 respectively. We use "tanh" function as the activator of each neuron, and we choose "LBFGS" method as to be the optimizer because our database is relatively small.

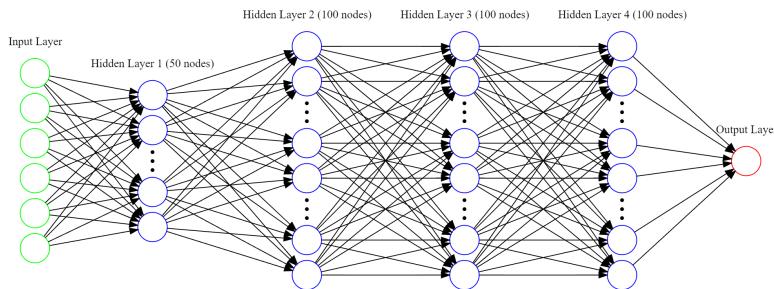


Figure 10: MLP Structure

Below are the benchmark results of our model:

TestAccuracy : 0.947 TestPrecision : 0.958 TestRecall : 0.947 TestF1Score : 0.947

7.2.4 Model Evaluation

Although our model has the ability to calculate the I_p of a building or a landmark, it still has some problems.

- Seldom Low Accuracy: Due to the relatively small size of our dataset, the model may not be adequately trained, potentially leading to inaccurate results.
- Limited Application Scenarios: The background of our dataset is the buildings and landmarks in New York. Although it is enough for common usage, special attention should be paid when using it to analyse presevation index of particular landmarks which are in special situations. Results obtained should be adjusted according to the actual situation.
- Generalized Results: The output of this model is only an index. Detailed preservation extent and methods need extra work.

8 Task 4: Recommendation Letter for Community Leader

After concluding the results of our previous research, we have written a recommendation letter proposing measures for the protection of the Mohenjo-Daro.

The letter is shown on the last but one page. 9.0.2

9 Strength and Weakness

9.0.1 Strength of Our Research

- Our model encompasses a broad time range and a large volume of data, employing the Analytic Hierarchy Process, ARIMA time series model, and Multilayer Perceptron neural network model for evaluation and prediction to make recommendations.
- Our model has a wide range of applications. It can be adapted for use by individuals in many different industries with slight adaptation and modification.

9.0.2 Weakness of Our Research

- Due to the use of the Analytic Hierarchy Process to determine weights, the calculated index weights may have a strong subjective component.
- Due to the short duration of the research, the sensitivity of the model has not been fully verified.

References

- [1] Yuchuan Lai and David A Dzombak. Use of the autoregressive integrated moving average (arima) model to forecast near-term regional temperature and precipitation. *Weather and Forecasting*, 35(3):959–976, 2020.
- [2] Chien-Chiang Lee and Yi-Bin Chiu. The impact of real income on insurance premiums: Evidence from panel data. *International Review of Economics & Finance*, 21(1):246–260, 2012.
- [3] Lea Bregar, Jože Rovan, and Matija Pavšič. Validity of gdp per capita for international development comparisons. *Economic & Business Review*, 10(3), 2008.

Dear community leaders of Larkana District, Sindh Pakistan,

Hello!

We are Team #2405335 of MCM/ICM 2024. It's our honor to write to you to discuss the preservation of Mohenjo-Daro, the famous archaeological site in your community. We have developed two models – the Insurance Model and the Preservation Model – to help you develop a new way to preserve this unique and precious historical heritage. Our latter model indicates that your building possesses value worth and severely demanding preserving across multiple dimensions. Unfortunately, the former model found that this building might face an embarrassing situation where no insurers are willing to provide coverage. Below is the challenges facing the Mohenjo-Daro

- Natural Factors:
 - Floods, river erosion and soil damage
- Human Factors:
 - Tourists' bad behavior and poor management
 - Insufficient funds and limited resources

Given the circumstances, we believe that measures should be taken as soon as possible. Here is our plan:

- In **finance**, we propose seeking government financial support and advocating for increased economic backing for safeguarding historical sites like Mohenjo-Daro. Meanwhile, enhance the development of local tourism to obtain more finance is also important.
- In terms of **management**, the first thing is to enhance the awareness of the government and the people on the protection of Mohenjo-Daro. At the same time, the management team and methods of the Mohenjo-Daro site should be optimized to reduce costs and increase efficiency. Also, the rational management of tourism and other industries should be applied to minimize the negative impact of these industries on this site.
- In terms of **protection**, it's crucial to engage professional teams and experts to devise the most effective preservation strategies tailored to local conditions. Addressing severe challenges, such as flooding and saltwater erosion leading to structural deterioration, may involve implementing solutions like constructing embankments.

We have worked out a timeline for this project. In the next five years, the local community and the preservation team should focus on improving the soil and water conditions around Mohenjo-Daro, and build protective facilities to reduce the natural threats. In the second five-year period, promote the synchronous and sustainable development of archaeological protection and tourism. However, protection work should always be the main job. After that, develop the local tourism industry, increase the popularity of the site to attract more visitors stimulating local economic development.

Below is the cost proposal for this project:

- Cost Estimate:
 - Phase One (5 Years): Improving Soil and Water Conditions, Constructing Protective Facilities
 - > \$1,500,000
 - Phase Two (5 Years): Promoting Archaeological Protection and Tourism Development
 - > \$1,000,000
 - Phase Three (Afterward): Developing Local Tourism and Increasing Site Popularity
 - > \$800,000
- Grand Total: \$3,300,000

We do hope this letter could be of some help to you about the preservation of Mohenjo-Daro!

Report on Use of AI

1. OpenAI ChatGPT (April, 2023 version, ChatGPT-4)

Query 1: Please give me a website where you can check the GDP data and population growth rate of the states in the United States?

Output 1: For GDP data, you can visit the U.S. Bureau of Economic Analysis (BEA) website at: <https://www.bea.gov/>. For population data and growth rates, you can visit the U.S. Census Bureau website at: <https://www.census.gov/>. Both of these websites provide comprehensive and regularly updated economic and demographic information at the national and state levels in the United States.

Query 2: Where can I find the profits and losses of American insurance companies in the past 30 years?

Output 2: To find information about the financial performance of American insurance companies over the last 30 years, you have several resources at your disposal. The National Association of Insurance Commissioners (NAIC) provides comprehensive data and analysis reports on the insurance industry, which include information about the financial condition and performance of insurance companies in the United States. Their reports cover various sectors like Property & Casualty, Life, Health, and others, providing a detailed view of the industry's performance over the years.

Query 3: If data is missing, how do I fill it up with interpolation?

Output 3: When data is missing, interpolation is a commonly used technique for data completion. Different interpolation methods can be chosen based on the specific situation and data characteristics. Here are some common interpolation methods:
1. Linear Interpolation
2. Polynomial Interpolation
3. Spline Interpolation
4. Nearest Neighbor Interpolation ...

2. Microsoft Copilot (Feb 6, 2024 version)

Query 1: Briefly introduce MLP.

Output 1: A Multilayer Perceptron (MLP) is a type of artificial neural network that is widely used for various machine learning tasks such as classification and regression. It is called a multilayer perceptron because it has many layers of nodes, known as artificial neurons, that connect to each other.

Query 2: Briefly introduce how Moenjo-Daro is preserved right now

Output 2: The preservation of **Moenjo-Daro** has faced challenges. Here's a summary of its current conservation status:

1. Recent Challenges: In 2022, disastrous monsoon flooding impacted Sindh, including Moenjo-Daro. The floods caused damage to the site, resulting in partial deterioration and structural issues¹.

2. Climate Change and Hazards: A framework for preservation includes strategies to cope with natural hazards and potential effects of climate change¹.