

# COMPARATIVE ANALYSIS OF ECONOMIC FACTORS IN SELECT STATES— PROJECT OUTLINE

## OBJECTIVE STATEMENT

The objective of this project is to analyze and compare key economic factors, such as salary ranges for data analysts, median home prices, and job growth rates, in specific states. By examining these factors, we aim to provide insights into the economic landscape and opportunities within each state, aiding decision-making processes for individuals, businesses, and policymakers.

## QUESTIONS

1. "What is the average salary range for data analysts in a specific state?":
  - Factors: Average salary, job market demand, cost of living index, and industry growth.
  - Analysis approach: Gather data on average salaries for data analysts in various states and compare them. Consider other factors such as job market demand and industry growth to understand the overall opportunities and potential for career growth in each state. Additionally, take into account the cost of living index to determine the relative affordability of each state.
2. "What is the median home price in a specific state?":
  - Factors: Median home price, housing market trends, population growth, and local economy.
  - Analysis approach: Collect data on median home sell prices in different states and compare them. Analyze housing market trends and consider factors such as population growth and the strength of the local economy to understand the overall housing market conditions in each state. This analysis will help provide insights into the affordability and potential investment opportunities in the real estate market.
3. "Which states have the best K-12 schools?":
  - Factors: Academic performance, graduation rates, teacher-to-student ratios, school funding, and resources.
  - Analysis approach: Gather data on academic performance, graduation rates, teacher-to-student ratios, school funding, and resources for K-12 schools in various states. Compare these factors to determine which states have the best overall K-12 education systems. Additionally, consider any specific criteria or rankings such as standardized test scores or national education rankings to help identify the states with the highest-quality K-12 schools.

## METHOD AND LANGUAGES

### 1. Data Collection:

- Gather data on the 9 key measures mentioned in the methodology, such as home price appreciation, high-tech and small business growth, job market and salary growth, amenities, cultural and recreational opportunities, and crime rate. Ensure to collect this data for each US state instead of metropolitan areas.

### 2. Data Preprocessing:

- Clean and preprocess the collected data for analysis.
- Normalize and transform the data to a common range of 0-100 points using the min-max normalization method.
- Utilize **Python Pandas** for data manipulation and preprocessing tasks.

### 3. Feature Selection:

- Review the 9 key measures and determine the most relevant features for the economic performance of a state.
- Consider factors such as GDP growth, employment rates, income levels, business growth, and other economic indicators.

### 4. Model Selection:

- Choose an appropriate machine learning algorithm for regression prediction based on the available data.
- Consider algorithms such as linear regression, decision trees, or gradient boosting algorithms.
- Utilize **Scikit-learn** for machine learning tasks.

### 5. Training and Evaluation:

- Split the data into training and testing sets.
- Train the machine learning model using the training set and evaluate its performance using the testing set.
- Adjust model parameters and evaluate different algorithms to find the best performing model.

### 6. Predictions:

- Apply the trained model to the selected features and make predictions for the best economic US state ranking in 2024.
- Use the methodology's weighting scheme to assign importance to each metric.
- Generate the final rankings.

### 7. Interpretation:

- Analyze the results and interpret the predictions based on the model's performance and the significance of the selected features.
- Compare the predicted rankings with historical rankings to assess the accuracy of the model.

#### 8. Visualization:

- Utilize **JavaScript Plotly** to create interactive and visually appealing charts and graphs for displaying the model's predictions, trends, and other relevant data visualizations.

#### 9. User Interface:

- Use **HTML, CSS, and Bootstrap** to design and structure the user interface of the project, allowing users to interact with the data and explore the results.
  - State Selection:
    1. Provide a dropdown menu with all the states listed alphabetically.
    2. Users can select a specific state from the dropdown to retrieve data and predictions for that state.
  - Metric Selection:
    1. Provide a list of checkboxes with different economic factors such as GDP, unemployment rate, inflation, etc.
    2. Users can select one or more checkboxes to choose the specific metrics they want to analyze.
  - Time Period Selection:
    1. Include a slider with a range of years or a dropdown menu with pre-defined time periods (e.g., last 5 years, last 10 years).
    2. Users can adjust the slider or select the desired time period from the dropdown to view data and predictions for that specific period.
  - Comparison Options:
    1. Provide checkboxes next to each state in a list, allowing users to select multiple states for comparison.
    2. Users can select one or more checkboxes to compare the selected states and view their respective data and predictions side by side.
  - Visualization Options:
    1. Offer different visualization options such as bar charts, line graphs, or pie charts.

2. Users can select the type of visualization they prefer to display the data and predictions.

- Predictive Analysis:

1. Include input fields where users can enter their own data or values for specific metrics.
2. Users can input their data, and the model will generate predictions or rankings based on the provided values.

- Export and Share:

1. Provide buttons or icons for users to download the analysis results as a CSV file or share them via email or social media platforms.
2. Users can click on the export or share buttons to perform the respective actions.

#### 10. Data Storage:

- Utilize **SQL** databases for storing and managing structured data efficiently.
- Consider **MongoDB** for handling unstructured or semi-structured data if required.

### OVERALL LIST OF LANGUAGES TO BE USED

- Python Pandas
- Scikit-learn
- JavaScript Plotly
- HTML, CSS, and Bootstrap
- SQL
- MongoDB

### NORMALIZATION PROCESS

To set the range of the data from 0-100, use a technique called normalization.

Normalization transforms the data so that it falls within a specific range, such as 0-100.

Step-by-step process:

1. Identify the minimum and maximum values for each metric. For example, a metric for home price appreciation, find the minimum and maximum values of this metric across all the states.

2. Apply the min-max normalization formula to rescale the values to the desired range.

The formula is as follows:

$$\text{normalized\_value} = (\text{original\_value} - \text{min\_value}) / (\text{max\_value} - \text{min\_value}) * 100$$

For each data point, subtract the minimum value from the original value, divide it by the range (max\_value - min\_value), and then multiply it by 100. This will produce the normalized value between 0 and 100.

## **ASSIGNING WEIGHTS PROCESS**

Assigning weights to each metric and incorporating them into the analysis will be done using the following steps:

1. Define the metrics: Identify the specific metrics or factors that will be used for analysis, such as average salary, median home price, academic performance, etc.
2. Normalize and transform the data: Convert each metric to a common range of 0-100 points. This can be achieved by scaling the data using a formula like min-max normalization. For example, if the minimum value for a metric is 50 and the maximum value is 200, you can transform a data point  $x$  to a normalized score  $y$  using the formula:  $y = (x - \text{min}) / (\text{max} - \text{min}) * 100$ .
3. Assign weights: Determine the importance or relevance of each metric to the overall analysis. Assign weights to each metric based on their relative significance. For example, if the average salary is more important than the median home price, assign a higher weight to the average salary metric. Weights must add up to 100%.
4. Apply the weights: Multiply the normalized scores of each metric by their respective weights. This will give each metric a weighted score. For example, if the weight for the average salary metric is 0.6, and the normalized score for a specific state is 80, the weighted score for that state would be  $0.6 * 80 = 48$ .
5. Sum the weighted scores: Add up the weighted scores for each metric to obtain a total weighted score for each state or city.
6. Sort and rank: Sort the states or cities based on their total weighted scores in descending order. The higher the score, the better the ranking.

METRIC ▼	WEIGHT ▼	REASON ▼	SOURCE ▼
Job Growth	20%	Job growth is a crucial indicator of economic health as it signifies increased employment opportunities, higher wages, and improved consumer spending. It has a significant impact on various sectors and drives overall economic growth.	Bureau of Labor Statistics
High-Tech GDP Growth	15%	High-tech industries are known for their innovation, productivity, and ability to drive economic growth. By assigning a significant weight to high-tech GDP growth, we acknowledge the role of technology-driven sectors in boosting the economy.	Bureau of Labor Statistics
Salary Growth	15%	Rising salaries indicate improved economic conditions and increased consumer purchasing power. It stimulates demand, fuels economic activity, and contributes to overall growth.	Bureau of Labor Statistics
Small Business Growth	12%	Small businesses are key drivers of job creation and innovation. By assigning weight to their growth, we recognize their importance in fostering economic development and diversification.	Census Bureau
One-year Home Appreciation	10%	Home appreciation reflects a healthy real estate market, which has ripple effects on consumer wealth, construction activity, and related industries. It contributes to economic growth, although to a slightly lesser extent.	Fed Housing Finance Agency
Net Migration	10%	Net migration, the difference between people moving in and out of an area, can impact economic growth. It affects the labor market, demand for goods and services, and overall population growth.	Census Bureau
Unemployment Rate	9%	The unemployment rate is an important economic indicator as it reflects the availability of jobs and the health of the labor market. Lower unemployment rates generally indicate a stronger economy.	Bureau of Labor Statistics
Personal Income	8%	Personal income levels affect consumer spending, savings, and investment. By assigning weight to personal income, we consider its influence on economic growth through increased purchasing power.	Bureau of Economic Analysis
Violent Crime Rate	1%	While important for overall societal well-being, the direct impact of the violent crime rate on economic growth is relatively limited compared to the other factors mentioned. Hence, it is assigned a lower weight.	FBI Uniform Crime Statistics
Total Weight	100%		