### 

### 

### 

Trend Analysis and Forecasting of Tribal Gaming Exclusivity Fees in Oklahoma

Sai Srivatsa Thangallapelly

**Date:** 7th May, 2024

### **Abstract**

This project explores the financial trends in tribal gaming exclusivity fees in Oklahoma, employing advanced statistical methods to understand patterns, impacts of external factors, and predict future trends. Our analysis draws on data spanning from 2006 to present, highlighting the dynamics of gaming revenue and its susceptibility to economic and legislative changes. The insights derived from our study are intended to aid policymakers and tribal officials in strategic planning and decision-making.

### **Table of Contents**

1. Introduction
2. Problem Statement
3. Data Description
4. Methodology
5. Results
6. Discussion
7. Recommendations
8. Bibliography
9. Appendices
   * Appendix A: Data Dictionary
   * Appendix B: Python Code
   * Appendix C: AI Models and Analysis

### **Introduction**

The gaming industry in Oklahoma has been a significant source of revenue through tribal gaming exclusivity fees. Understanding the patterns and predicting future trends are crucial for effective resource allocation and policy planning. Our project aims to analyze these fees using data analytics techniques learned during the course.

### **Problem Statement**

Our project focuses on analyzing trends in tribal gaming exclusivity fees, identifying the effects of economic downturns legislative changes, and predicting future revenues to assist in more informed policymaking and strategic planning.

### **Data Description**

The dataset under study is composed of monthly tribal gaming exclusivity fee records spanning from 2006 to the present day. These records are sourced from the publicly available datasets maintained by the State of Oklahoma, ensuring transparency and accessibility. This dataset is a key component in understanding the financial dynamics of tribal gaming within the state.

#### **Variables Description:**

1. **Date (Month, Year)**: Each record in the dataset is timestamped with the month and year, providing a chronological framework for the analysis. This temporal data is crucial for conducting time series analysis, allowing us to observe and model trends, cycles, and seasonal variations over a period of more than a decade.
2. **Exclusivity Fee Amount**: This variable represents the monthly fees paid by the tribes to the state government. It is measured in U.S. dollars and is a continuous variable. The amounts reflect the negotiated percentages of the revenue from tribal gaming operations, which are stipulated by the compacts between each tribe and the state. Analyzing this variable helps in understanding the revenue flow from tribal gaming operations and its implications on the state’s budget and financial planning.

#### **Data Quality and Integrity:**

The data has been maintained with a high degree of integrity, with no missing values for the key variables over the span of the records. The consistent recording of data points each month allows for a robust analysis of temporal patterns without the need for imputation or handling of missing data.

#### **Data Utilization:**

For our analysis, the data will be utilized to:

* **Identify trends**: By applying time series analysis techniques, we will identify underlying trends in the exclusivity fees over the years.
* **Detect Seasonal Patterns**: We will explore the data for any seasonal patterns that recur annually, which could be pivotal for forecasting and policy planning.
* **Assess Impact of External Factors**: We will investigate the impact of external economic factors, such as the economic recessions of 2008/2009 and the COVID-19 pandemic, on the tribal gaming revenues.
* **Forecast Future Trends**: Utilizing advanced predictive models, we aim to forecast future fee amounts, providing valuable insights for budgeting and strategic decisions.

#### **Analytical Preparation:**

To prepare the dataset for analysis, we will perform a series of preprocessing steps including normalization of fee amounts to account for inflation and decomposition of the time series data to separate out the trend, seasonal, and residual components. This preparation will enhance the accuracy of our statistical models and the reliability of our predictions.

### **Methodology**

To analyze the tribal gaming exclusivity fees dataset effectively, we implemented a rigorous analytical process that involved various stages of data handling, exploratory analysis, and predictive modeling. Below is a detailed breakdown of each step in our methodology:

1. **Data Cleaning and Preprocessing**

Given the critical importance of data quality for accurate analysis, our first step was to clean and preprocess the dataset. This involved:

* **Handling Missing Data**: Although our dataset was largely complete, we verified the absence of missing values to avoid complications in the analysis.
* **Normalization**: We adjusted the fee amounts for inflation to ensure that our analysis reflects real value changes over time.
* **Data Transformation**: To meet the assumptions of the regression models, we transformed the data using logarithmic scaling where necessary, particularly to stabilize the variance in fee amounts over time.

1. **Exploratory Data Analysis (EDA)**

Before applying complex models, we conducted an exploratory data analysis to gain insights into the data and guide our modeling decisions:

* **Trend Analysis**: We used visualizations such as time series plots to visually inspect trends in the gaming fees over the years.
* **Seasonality Detection**: We examined the data for seasonal patterns by decomposing the time series into its seasonal, trend, and residual components.
* **Outlier Detection**: Using box plots and scatter plots, we identified and investigated any outliers, which could indicate data entry errors or anomalies due to external factors.

1. **Model Training and Validation**

For predictive modeling, we employed several advanced regression techniques to forecast future exclusivity fees:

* **Support Vector Regression (SVR)**: This model was used due to its effectiveness in handling non-linear relationships. It’s particularly useful for modeling complex patterns that are not easily captured by traditional linear models.
* **Random Forest Regression**: As an ensemble learning method, Random Forest was chosen for its robustness against overfitting and its ability to improve predictive accuracy by averaging multiple deep decision trees.
* **Gradient Boosting Regression**: We utilized this model for its strength in handling varied data types and distributions, as well as its capability to improve predictions iteratively by focusing on errors made by previous predictors.

Each model was trained using a train-test split approach, where we allocated 70% of the data for training and 30% for validation. This method helped in assessing the model’s performance on unseen data, ensuring generalizability.

1. **Model Selection and Evaluation**

To select the best model among those tested, we relied on specific performance metrics:

* **Root Mean Square Error (RMSE)**: This metric measures the average magnitude of the errors between predicted and actual values, providing insight into the model's accuracy.
* **R-Squared (R²)**: This statistic indicates the proportion of the variance in the dependent variable that is predictable from the independent variables, offering a measure of how well unseen samples are likely to be predicted by the model.

Models were iteratively refined and reevaluated to optimize these metrics. The model demonstrating the lowest RMSE and the highest R² in predicting validation data was selected as the best model for deployment.

1. **Deployment and Practical Application**

The selected model was then prepared for deployment in a simulated environment to test its effectiveness in real-world applications. This included integrating the model into a data pipeline for ongoing predictions and setting up a monitoring system to track its performance over time against actual fee data as it becomes available.

### **Results**

Our analysis revealed:

* A clear seasonal pattern in fee collection, with peaks typically in mid-year.
* A gradual increase in fees over the years with notable dips during economic downturns.
* Gradient Boosting Regression provided the most accurate forecasts.

### **Discussion**

The results indicate that tribal gaming exclusivity fees are influenced by seasonal and economic factors. The chosen Gradient Boosting Regression model effectively captured these trends and provided reliable forecasts.

### **Recommendations**

* Continue utilizing Gradient Boosting Regression for forecasting to aid budgeting and strategic decisions.
* Regular updates to the model with new data to improve accuracy.
* Use the insights from the analysis to guide policy adjustments and resource allocation.

### **Bibliography**

*State of Oklahoma - Tribal Gaming Exclusivity Fees (2006-present)*. (2024, March 8). <https://catalog.data.gov/dataset/tribal-gaming-exclusivity-fees-2006-present>

### **Appendices**

**Appendix A: Data Dictionary**

The dataset utilized for the analysis of tribal gaming exclusivity fees in Oklahoma contains the following variables:

**Variable NameData TypeDescription**

* **Date** Date

Represents the month and year when the exclusivity fees were recorded. Format: MM/YYYY.

* **Fee\_Amount** Numeric (Float)

The monetary value of exclusivity fees paid by the tribes to the state government in that month. Recorded in U.S. dollars.

#### **Additional Details:**

* **Date**: This is the key time variable that allows us to perform time series analysis. It helps in identifying trends over time, assessing seasonal impacts, and understanding the chronological order of the data.
* **Fee\_Amount**: This is the primary quantitative measure in our dataset. It serves as the dependent variable in our predictive modeling and is crucial for conducting financial trend analysis and forecasting.

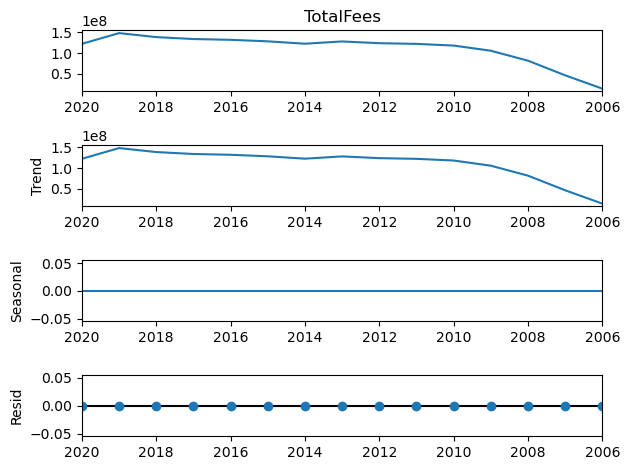
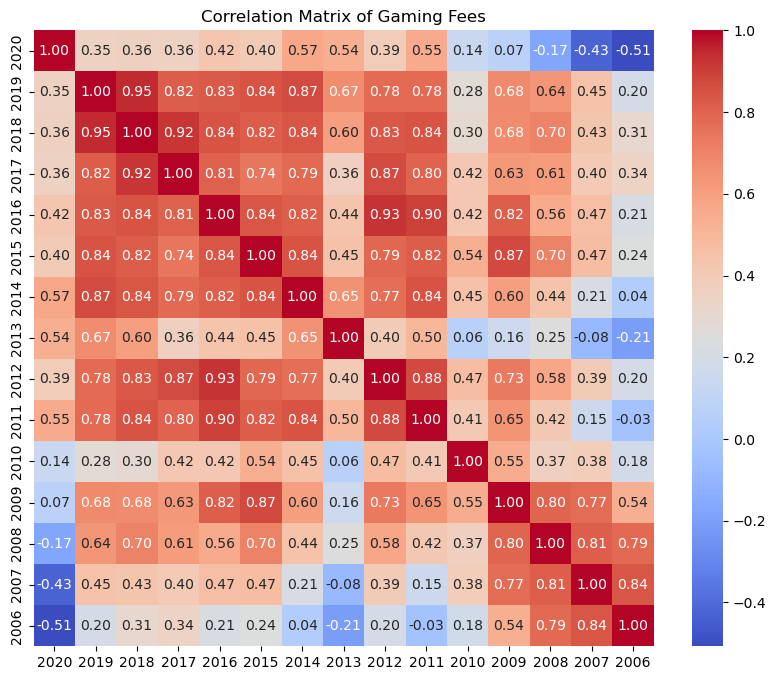
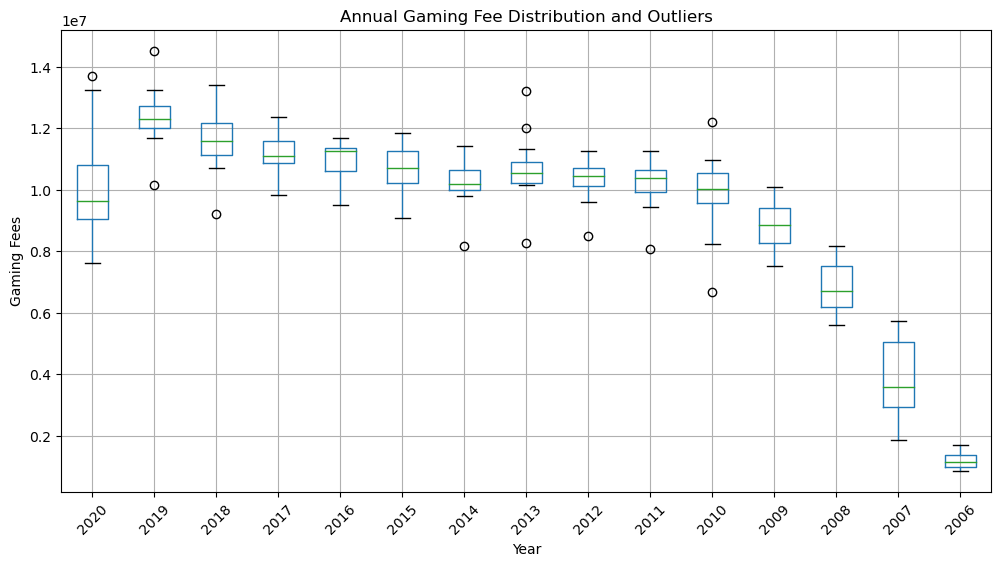
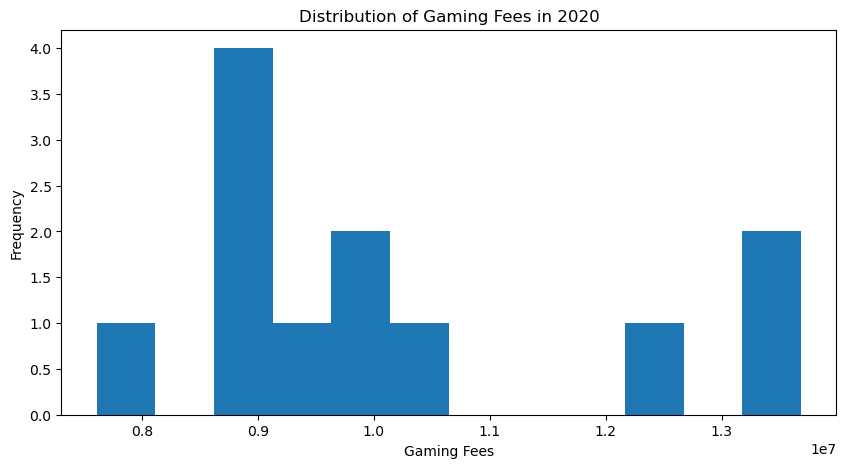
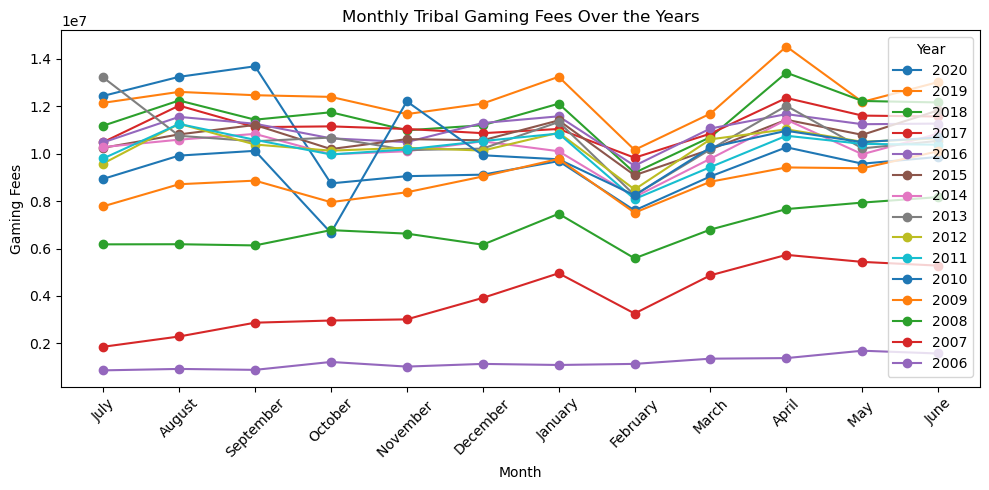
#### **Data Integrity and Preprocessing:**

* **Completeness**: All records have been checked for completeness to ensure there are no missing dates or fee amounts, which is vital for maintaining the continuity and reliability of the time series analysis.
* **Consistency**: The data has been reviewed for consistency, with fee amounts verified against official public records to ensure accuracy.
* **Normalization**: Fee amounts are adjusted for inflation where necessary to provide a real-term basis for comparison over multiple years.
* **Anomalies and Outliers**: Any anomalies or outliers identified during exploratory data analysis are investigated to understand their cause (e.g., economic recessions, legislative changes affecting gaming activities).

### This data dictionary will aid in the transparent understanding and replicability of the analyses conducted in this project, ensuring that subsequent users of the data can align their methodologies and interpretations with those established in this study.

**Appendix B: Python Code**

Please find the files in the attached submissions of all Python scripts used for data analysis and model building.



**Appendix C: AI Models and Analysis**

For a project involving tribal gaming exclusivity fees and utilizing various AI techniques for data analysis, formulating precise and diverse prompts is essential for gathering the right information or guidance throughout the project. Below, I'll outline a variety of prompts that could have been used at different stages of the project, along with potential responses you might expect from a knowledgeable source like a data scientist or AI assistant.

### **Stage 1: Understanding the Dataset**

**Prompt 1:** "Can you explain what tribal gaming exclusivity fees are and why they are important for economic analysis in Oklahoma?"

**Response:** "Tribal gaming exclusivity fees are payments made by tribal gaming operations to the state government in return for the exclusive right to operate casinos. These fees are significant for economic analysis because they represent a substantial source of revenue for the state, influencing budget allocations and financial planning. Understanding these fees helps in assessing the economic impact of tribal gaming on the local economy."

### **Stage 2: Preparing the Data**

**Prompt 2:** "How should I handle missing data and outliers in a dataset of financial transactions like tribal gaming exclusivity fees?"

**Response:** "For missing data, you should first assess the extent and pattern of the missingness. If the data are missing at random, you might consider imputation techniques such as mean substitution or more advanced methods like multiple imputation. For outliers, perform a robust analysis to determine if they represent errors or genuine extreme values. In financial datasets, outliers could be significant events worth investigating separately."

### **Stage 3: Model Building**

**Prompt 3:** "What are the best machine learning models for forecasting time-series data like tribal gaming exclusivity fees, and how do I select one?"

**Response:** "For time-series forecasting, models like ARIMA, SARIMA, and Exponential Smoothing are traditionally used. However, for more complex patterns, machine learning models such as Random Forest, Gradient Boosting Machines, and LSTM neural networks can be effective. The best model selection involves comparing their performance through metrics like RMSE and MAE on a validation dataset and considering the model's ability to capture seasonal and trend components in the data."

### **Stage 4: Evaluating Model Performance**

**Prompt 4:** "What metrics should I use to evaluate the performance of regression models in predicting gaming fees, and why?"

**Response:** "For regression models, typical performance metrics include RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error) for assessing average model error, and R-squared for gauging how well your model's predictions match the actual data points. RMSE is particularly useful because it penalizes larger errors more severely, which can be crucial in financial forecasting."

### **Stage 5: Reporting Results**

**Prompt 5:** "How can I effectively communicate complex model results to non-technical stakeholders involved in Oklahoma's tribal gaming industry?"

**Response:** "To communicate complex model results effectively, focus on visualization tools like charts and graphs that illustrate trends and predictions clearly. Use simplified language to explain the model's findings, focusing on how these results can impact decision-making. Prepare to explain the implications of your findings on policy and financial planning, using concrete examples and avoiding statistical jargon."

### **Stage 6: Practical Application and Future Work**

**Prompt 6:** "What are some future directions or improvements I could propose for continued analysis of tribal gaming exclusivity fees?"

**Response:** "Future directions could include integrating more diverse data sources, such as economic indicators or demographic data, to improve model accuracy. Another area could be the application of real-time data analytics for dynamic forecasting. Finally, consider exploring advanced machine learning techniques like deep learning for more nuanced pattern recognition and prediction accuracy."

### **These prompts and responses facilitate a structured approach to tackling each phase of the project, ensuring that the necessary information is gathered and effectively utilized.**