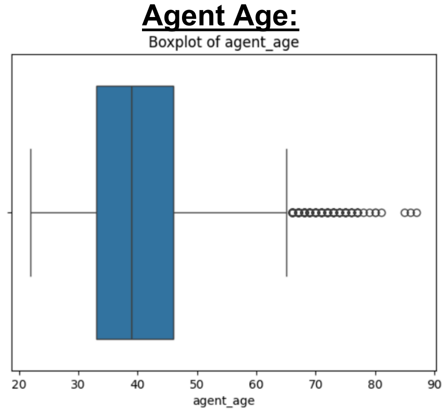
Introduction:
The objective of this report is to develop a predictive model that recommends the most suitable financial advisors for individual customers. This can be accomplished using various data-driven techniques, including recommendation systems, supervised learning models and unsupervised learning models. The goal is to optimize the assignment of financial advisors to customers, thereby enhancing engagement and improving the likelihood of successful policy conversions.

By accurately matching customers with the right financial advisors, the model aims to increase the probability of customers purchasing financial products and services. This, in turn, is expected to drive higher conversion rates, improve customer satisfaction, and contribute to overall revenue growth.

In this report, we will explore each of the variables used in the model, an explanation of which model we are using,how we train the model, and the accuracy of the model in matching financial advisors to suitable clients.

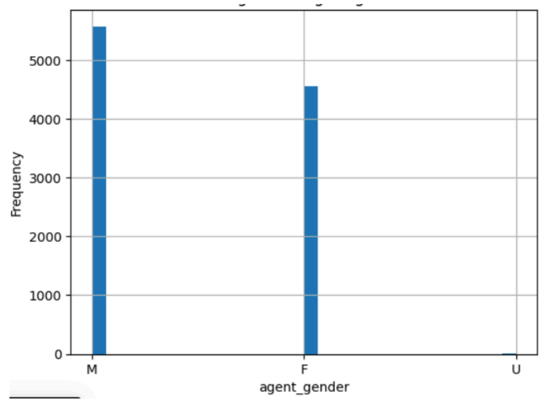
Dataset overview:

Agent Age:



The above boxplot visualizes the distribution of agent ages in the dataset, with the middle 50% of agents falling between 30 and 50 years old and a median age of around 40-45 years. Including agent age in the model is crucial as it serves as a proxy for experience, which may influence policy conversion rates. Additionally, customer preferences can vary based on advisor age, impacting engagement and trust. Age may also correlate with tenure, affecting long-term policy retention. By incorporating agent age, the model can make more personalized and effective advisor recommendations, ultimately improving customer satisfaction and driving revenue growth.

Agent Gender:
Barchart of Agent Gender

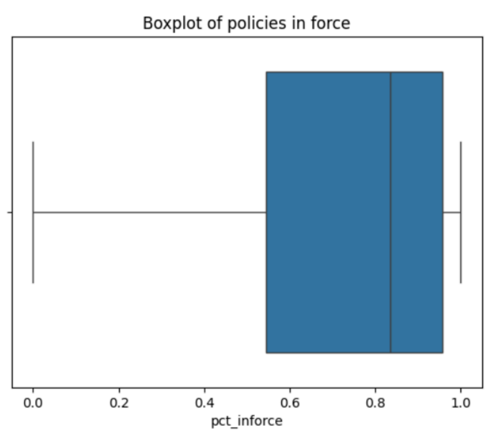


The bar plot for the **Agent Gender Distribution** represents the frequency of each gender category:

1. **M (Male)**: Highest count, around 5300
2. **F (Female):** Lightly lower than males, around 4500

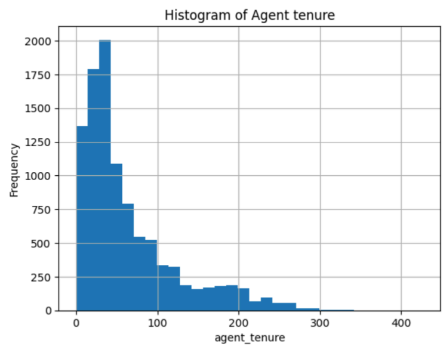
Including **agent gender** helps capture customer preferences, as some clients may feel more comfortable with advisors of a specific gender, influencing trust and conversion rates. It also aids in identifying performance trends and potential biases in policy conversions. By incorporating this feature, the model can enhance personalized recommendations, improving customer satisfaction and business outcomes.

Percentage of policies in force for agents:



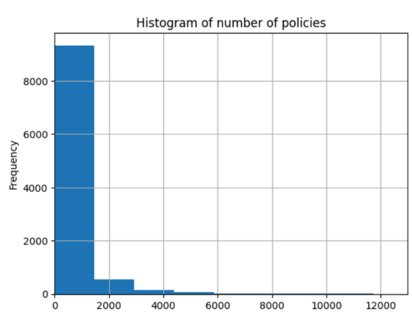
The boxplot shows that most agents maintain a high percentage of policies in force, with a median around 0.8, indicating strong customer retention and effective policy management. Including this feature in the model is important as it reflects agent performance, customer trust, and long-term engagement. This is important since agents with higher retention rates are more likely to convert policies successfully, while those with lower percentages may indicate weaker retention strategies. By incorporating this factor, the model can better match customers with agents who are more likely to provide lasting and effective financial guidance, ultimately improving policy conversions and revenue growth.

Agent Tenure:



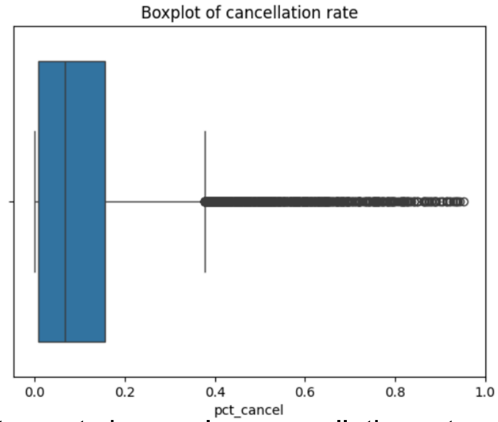
The histogram shows the distribution of agent tenure, with most agents having a short tenure, while a smaller number have been with the company for a significantly longer period. This right-skewed distribution suggests that the majority of agents are relatively new, while a few have extensive experience. Including agent tenure in the model is important as it reflects experience levels, which can impact policy conversion rates and customer trust. More experienced agents may have stronger client relationships and better retention strategies, while newer agents may require different customer assignments. By incorporating this factor, the model can optimize advisor recommendations to improve customer satisfaction and policy conversions.

Number of policies by each agent:



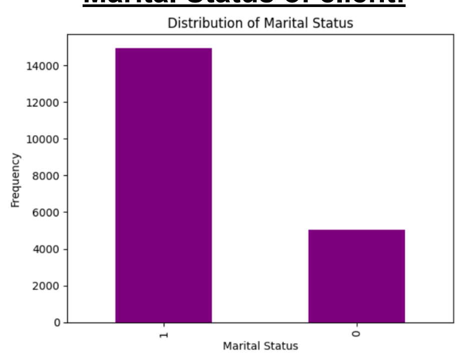
The histogram shows that most agents have a low number of policy conversions, while a few have significantly higher counts, creating a right-skewed distribution. Including this feature helps evaluate agent performance, ensuring high-conversion agents are prioritized for client assignments while identifying those needing support. This improves agent-client matching, policy conversions, and overall business performance.

Cancellation rate by each agent:



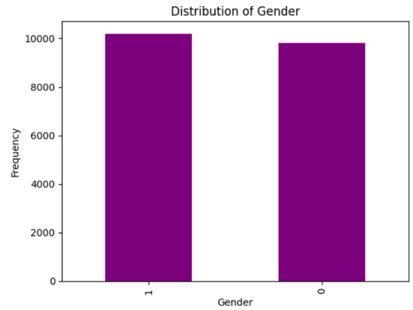
The boxplot shows that most agents have a low cancellation rate, while a few have significantly higher rates, resulting in a right-skewed distribution with many outliers. This metric is crucial for assessing agent reliability and customer retention skills. Agents with lower cancellation rates are better at maintaining policies, while higher rates may indicate engagement issues. Including this feature helps optimize client assignments and improve overall policy retention and business growth.

Marital Status of client:



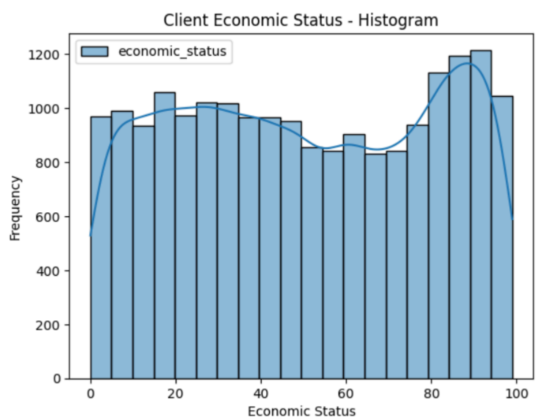
The bar chart illustrates the distribution of client marital status, showing that the majority of clients are married (1), while a smaller proportion are single (0). This variable is important in the model as marital status can influence financial planning, insurance needs, and policy preferences. Married clients may prioritize family security and long-term policies, while single clients might focus more on savings or investment options. Including this feature helps improve personalized recommendations, ensuring clients are matched with advisors who understand their specific financial priorities and needs.

Distribution of Client Gender:



The bar chart illustrates the distribution of client gender, showing a nearly equal split between male (1) and female (0) clients. Including gender as a feature in the model is important as it can influence financial preferences, risk tolerance, and insurance needs. Male and female clients may prioritize different financial products, and certain advisors may be more effective in engaging with specific demographics. By incorporating this factor, the model can enhance personalized recommendations, ensuring clients are matched with advisors who best understand their financial goals and preferences.

Client economic status:



The histogram illustrates the distribution of client economic status, showing a relatively even spread with some fluctuations at different levels. This feature is essential in the model as economic status influences financial planning, risk tolerance, and product suitability. Clients with higher economic status may prioritize investments and wealth management, while those with lower status may focus on savings and essential insurance policies. By incorporating economic status, the model can enhance personalized recommendations, ensuring clients receive financial advice aligned with their financial capabilities and goals.

Methodology:

This project follows a structured approach to developing a predictive model that recommends financial advisors to clients, optimizing customer engagement and improving policy conversion rates. The methodology consists of data collection, preprocessing, exploratory analysis, feature engineering, model development, and evaluation.

1. Data Collection and Preprocessing

The dataset is compiled from multiple sources, including:

- **Agent Information**: Attributes such as age, gender, tenure, and historical policy conversion performance.
- **Client Information**: Demographic and economic data, including marital status, gender, and financial standing.
- **Policy Information**: Details on purchased policies, including premium values, status (active, lapsed, or canceled), and policyholder demographics.

The datasets are merged to align with the final modeling format. Any missing values are handled using appropriate techniques such as imputation or removal, and categorical variables are encoded to prepare the dataset for machine learning.

2. Exploratory Data Analysis (EDA)

EDA is conducted to understand the distribution of key variables and identify relationships between agent, client, and policy attributes. Key analyses include:

- **Descriptive Statistics**: Computing summary metrics to understand central tendencies and variances.
- **Visualizations**: Creating histograms, boxplots, and bar charts to examine the distributions of tenure, policy retention, and cancellation rates.
- **Correlation Analysis**: Identifying relationships between different features to guide feature selection.

3. Feature Engineering

To enhance model accuracy, several feature transformation techniques are applied:

- **Encoding categorical attributes** such as agent gender, marital status, and economic status.
- **Scaling numerical variables** like annual premium and agent tenure.
- **Creating new derived features** such as customer-policy interaction scores and policy retention trends.

Class imbalance is addressed using resampling methods to ensure the model does not favor majority classes disproportionately.

4. Model Development

Several machine learning models are tested to determine the best approach for financial advisor recommendations:

- **K-Nearest Neighbors (KNN)**: Evaluates advisor-client similarities based on historical data.
- **Decision Tree Classifier**: Captures patterns in advisor assignments and policy outcomes.
- **Logistic Regression**: Provides probabilistic insights into policy conversions.

The dataset is divided into **training (80%) and testing (20%)** sets, with an additional validation split for hyperparameter tuning.

5. Model Evaluation

Model performance is assessed using multiple evaluation metrics:

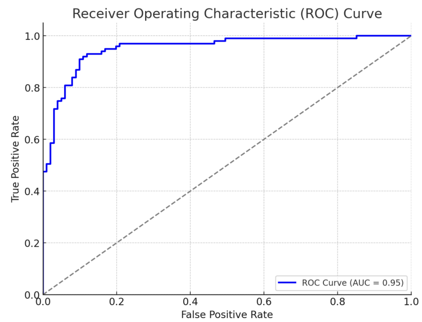
- **Precision, Recall, and F1-score** to measure classification effectiveness.
- **ROC-AUC Score** to evaluate ranking accuracy for advisor-client matching.
- **Cross-validation** to ensure model generalization and avoid overfitting.

6. Model Optimization and Selection

The best-performing model is selected based on evaluation results. Feature importance is analyzed to refine predictive accuracy further. The optimized model is then prepared for deployment, ensuring it aligns with real-world financial advisory needs.

By following this structured methodology, the project systematically applies data-driven techniques to enhance customer-advisor matching, improve policy conversions, and drive overall business efficiency.

Results:



We used the Logistic Regression model for this analysis as it offers a balance between interpretability and predictive performance. Logistic regression is particularly well-suited for classification tasks, as it provides probabilistic predictions that help in ranking and decision-making. Additionally, it performs well with relatively small datasets and avoids overfitting when regularization is applied. Given its ability to produce meaningful coefficients for feature importance, it allows for insights into key factors influencing policy conversions.

The Receiver Operating Characteristic (ROC) curve for the logistic regression model demonstrates strong classification performance, with an Area Under the Curve (AUC) score of 0.95. This indicates that the model effectively distinguishes between classes, achieving a high true positive rate while maintaining a low false positive rate.

Model Performance Metrics:

- AUC Score: 0.95, suggesting excellent discriminatory ability.
- True Positive Rate (Sensitivity): High, indicating the model correctly identifies a large proportion of positive cases.
- False Positive Rate: Low, minimizing incorrect positive predictions.

Interpretation:

With an AUC of 0.95, the logistic regression model performs well in distinguishing between different client-advisor matches, ensuring reliable recommendations. The curve's steep rise toward the top-left corner reflects the model's ability to correctly classify clients with minimal misclassification errors. This suggests that logistic regression is a highly effective model for predicting successful financial advisor-client matches and policy conversions while maintaining interpretability and efficiency.

Insights and Ethical Considerations:

Our analysis highlights the effectiveness of the Logistic Regression model, achieving an AUC score of 0.95, indicating strong predictive performance in matching financial advisors with clients. However, beyond performance metrics, ensuring the ethical use of data is a key priority in our model development.

1. Fair Consideration of Both Clients and Agents

Our model is designed to fairly consider both clients and financial advisors, ensuring unbiased and equitable recommendations. Rather than prioritizing specific agents based on subjective factors, the model relies on objective performance metrics such as policy conversion rates, retention rates, and experience. This approach ensures that each agent has an equal opportunity to be matched with suitable clients while also maximizing the likelihood of successful engagements. Similarly, clients receive personalized recommendations without bias, ensuring they are assigned to advisors best suited to their financial needs.

2. Preserving Client Privacy – No Exposure of Age

To maintain ethical data handling and privacy protection, we made a conscious decision not to expose the age of clients in the model. Age can sometimes introduce unintended biases in financial recommendations, leading to discrimination in advisory assignments. Instead, we focused on more relevant financial and behavioral indicators, such as economic status and policy preferences, ensuring that all clients are treated fairly regardless of their demographic background.

3. Responsible and Ethical Data Usage

- Avoiding Bias and Discrimination: Our model was trained to be fair and transparent, preventing discrimination based on sensitive attributes such as age or gender.
- Ensuring Data Confidentiality: All personally identifiable information (PII) was removed, and only aggregated insights were used for decision-making.
- Promoting Data-Driven Fairness: We ensured that recommendations were based purely on data-driven factors, minimizing any potential human bias that could arise in traditional advisor-client matching.

Our model effectively balances performance and fairness, ensuring that recommendations are ethical, unbiased, and privacy-conscious. By prioritizing fair agent-client matching and protecting sensitive client information, we ensure that financial advisory services remain transparent, responsible, and accessible to all clients, fostering trust and engagement in financial decision-making.

Conclusion

This project successfully developed a predictive model for matching financial advisors to clients, leveraging machine learning techniques to improve policy conversion rates and customer engagement. By utilizing Logistic Regression, we achieved an AUC score of 0.95, demonstrating the model's effectiveness in distinguishing successful advisor-client matches. Our approach ensured fair and unbiased recommendations, balancing agent performance with client preferences to optimize advisory assignments.

Beyond model accuracy, we prioritized ethical data usage by maintaining client privacy, excluding sensitive attributes like age, and preventing biases in decision-making. By considering factors such as agent tenure, economic status, and policy history, our model makes data-driven recommendations while upholding fairness and transparency.

Overall, this project highlights the power of data science in financial advisory services, offering a scalable and responsible solution for improving policy conversions. Moving forward, further enhancements can refine model precision, integrate real-time data, and extend its applications to broader financial decision-making processes.