Data Storm 6.0 - Helloworld 2.0

1. Introduction

The challenge focuses on improving insurance agent performance by predicting agents who are at risk of not making any sales ("NILL") in the following month, and then creating actionable strategies to help improve performance for all agents. The key objectives of the challenge were to:

- **Predict at-risk agents ("NILL")**: Identify agents who are likely to fail and make no sales in the upcoming month.

- **Enhance agent performance**: By categorizing agents based on performance levels (High, Medium, Low), we can suggest personalized improvement plans tailored to their needs.

The methods applied to achieve these objectives included:

- **Exploratory Data Analysis (EDA)**: To identify patterns, trends, and relationships within the dataset.

- **Predictive Modeling**: Building and training a machine learning model to predict "NILL" agents.

- **Performance Interventions**: Using the classification to recommend actions to help agents improve based on their performance.

2. Exploratory Data Analysis (EDA)

1. Key Metrics and Distributions

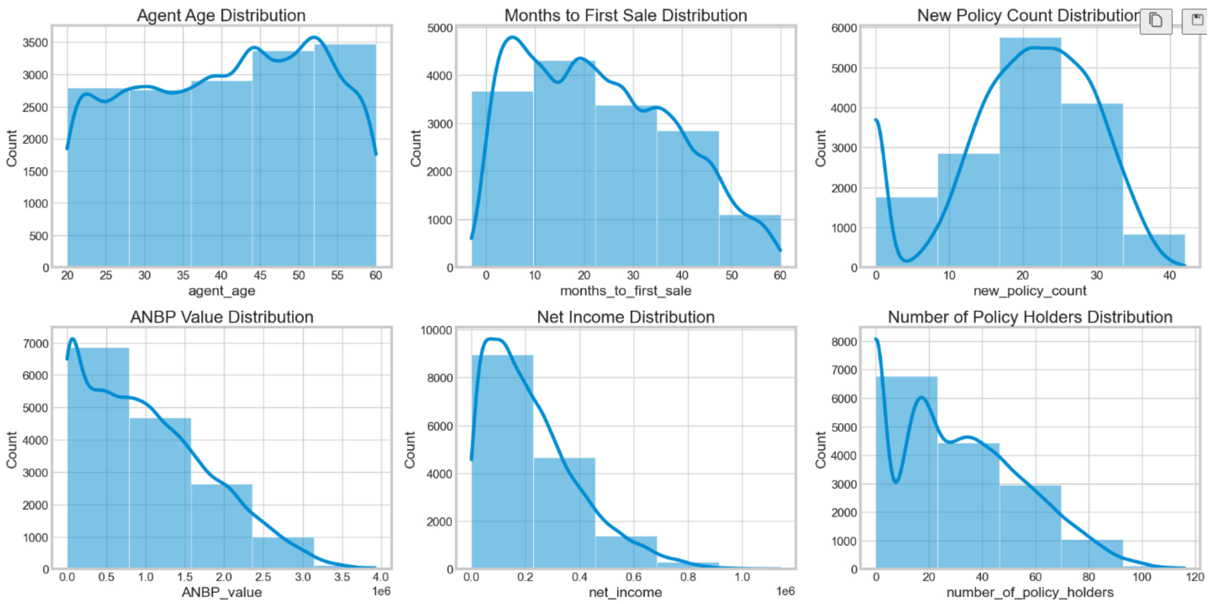
The first step was to explore the dataset’s key metrics. Using summary statistics, the mean, median, and standard deviation for critical variables such as sales and performance indicators were calculated. The distributions were analyzed using visualization techniques like histograms to understand the spread and identify any patterns or outliers.

Dataset Overview:

- Number of agents in sample: 15308
- Number of features: 24

```
Agent Demographics:
          agent_age  months_to_first_sale
count  15308.000000          15308.000000
mean      40.585772             23.034165
std       11.688204             15.256349
min       20.000000             -3.000000
25%       31.000000             10.000000
50%       41.000000             22.000000
75%       51.000000             35.000000
max       60.000000             60.000000
```

```
Activity Metrics:
       unique_proposals_last_7_days  unique_proposals_last_15_days  \
count                  15308.000000                   15308.000000
mean                       1.491246                       2.991116
std                        1.119554                       2.001613
min                        0.000000                       0.000000
25%                        0.000000                       1.000000
50%                        1.000000                       3.000000
75%                        2.000000                       5.000000
max                        3.000000                       6.000000
```

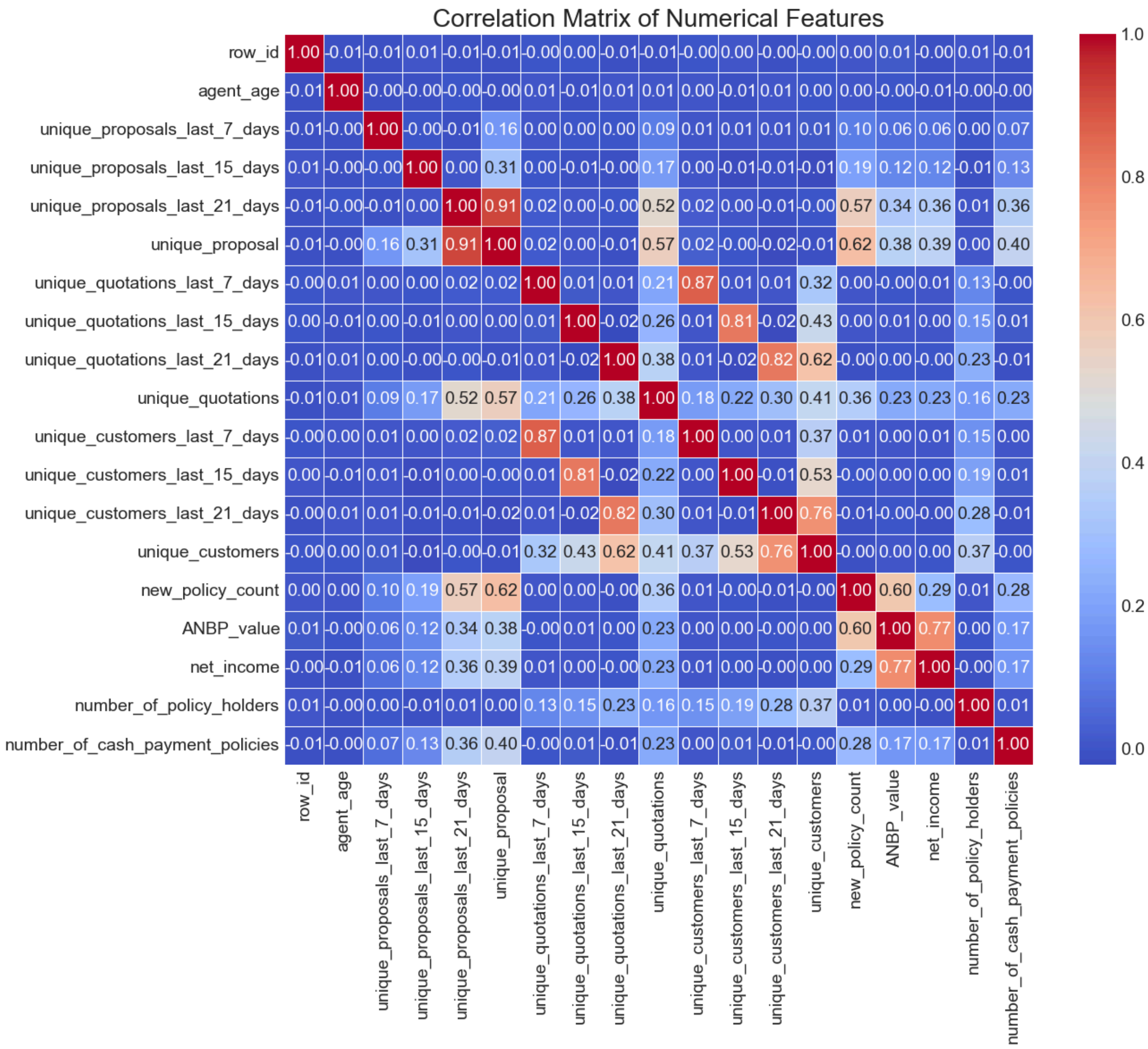


2.Sales Patterns by Month

Time-series analysis was performed to assess how the sales performance of agents fluctuated by month. This allowed us to identify periods of unexpected drops or spikes in agent performance. For example, if agents were showing a significant drop in performance during certain months, this could signal factors like external market conditions, lack of training, or even seasonal trends affecting the sales pipeline. The monthly trends were visualized in line graphs to easily track the agent's performance trajectory.

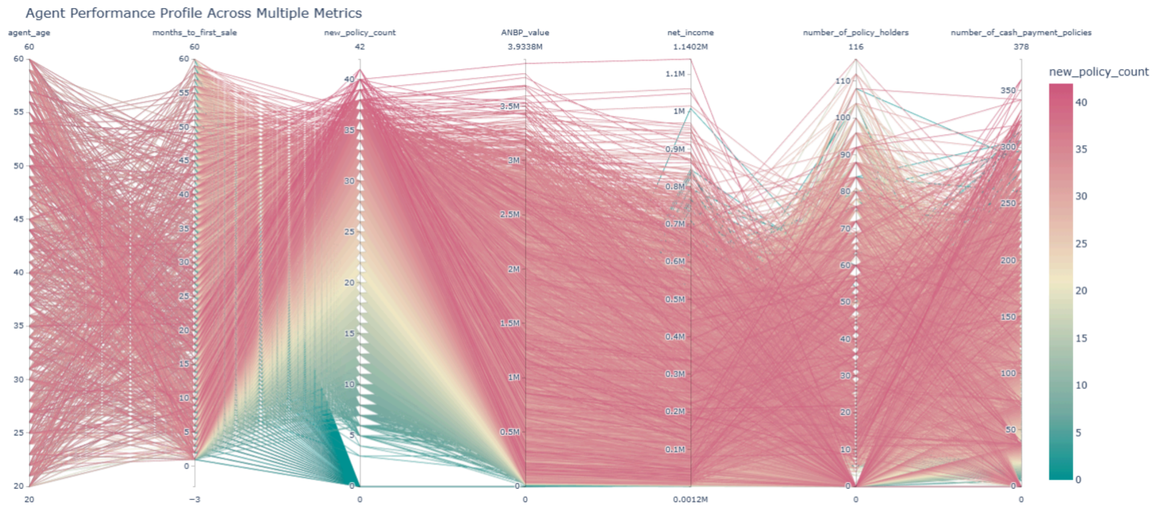
3.Multivariate Analysis

Next, a multivariate analysis was carried out to understand the interactions between various numerical features. Key relationships explored included how sales performance interacted with factors like the number of training hours. The correlation between sales performance and other factors, such as the number of policies sold or past performance, was examined using correlation matrices and scatter plots. These analyses helped to uncover any hidden patterns that may explain why some agents perform better than others.



4.Individual Agent Trajectories

The performance trajectories of individual agents were also analyzed. By tracking agent performance over time, we observed whether agents consistently performed well or had sporadic sales patterns. This analysis helped to identify agents who may be at risk of going "NILL" based on their historical sales patterns. For instance, agents with consistently low performance over multiple months were flagged for further intervention.



5.Innovative Insights

The EDA also led to several innovative insights:

- **Training Hours Impact**: A strong correlation was found between the number of training hours an agent received and their initial sales performance. Agents who spent more time in training appeared to perform better during their early months.

- **Performance Consistency**: Agents with a steady performance trajectory over time were less likely to go "NILL." Identifying such agents early allowed for proactive interventions.

- **Sales Drop Patterns**: The time-series analysis revealed recurring periods when agent performance dropped, signaling the need for additional support or interventions during those months.

These insights formed the basis for predicting "NILL" agents and understanding the factors influencing their performance.

Additional Insights:

- Average age of agents: 40.6 years
- Average time to first sale: 23.0 months
- Average new policy count: 20.3 policies
- Average ANBP value: $1,025,337.79
- Average net income: $228,041.39

	agent_code	performance_index	performance_category
1696	68198d93	0.882554	High
6041	bc913ebe	0.859864	High
9077	f2c082c3	0.838250	High
3143	3c26fc5a	0.828460	High
6727	1a1279f8	0.820719	High
...	...	...	...
4045	5da2a068	0.001782	Low
2813	61b67baf	0.001550	Low
7303	57480056	0.001170	Low
4414	328e4201	0.000801	Low
2334	c724fffa	0.000670	Low

3. Predicting "NILL" Agents

In this part of the project, the goal is to predict which agents are at risk of making no sales ("NILL") in the following month. This is critical because identifying these agents early allows for targeted intervention, such as training or motivation, before they fall behind in performance.

1. Data Preparation

The first step involved loading and preparing the training data. The following transformations were applied:

- **Date Columns Conversion**: The relevant date columns (`agent_join_month`, `first_policy_sold_month`, and `year_month`) were converted to datetime format to allow for time-based analysis.

- **Agent Experience**: The experience of each agent was calculated in months, which was derived from the difference between their joining date and the current month.

- **Time to First Sale**: For agents who had not made their first sale, a placeholder value (999 months) was used. For others, the time to the first sale in months was calculated.

- **First Sale Indicator**: A binary indicator was created to identify whether an agent had made their first sale.

- **Conversion Rates**: Proposal-to-quotation and quotation-to-policy conversion rates were calculated to measure the efficiency of agents in converting proposals into sales.

- **Policy and Income Metrics**: The average policy value and average income per policy were computed for each agent.

- **Proposal Activity**: A series of proposal activity metrics over different time periods (7, 15, 21 days) were created to measure recent proposal efforts.

2. Feature Engineering

After the initial data preparation, additional features were engineered to better capture agent behaviors and performance:

- **Proposal and Quotation Ratios**: Ratios of proposals and quotations per customer were calculated to measure how efficiently agents were interacting with customers.

- **Recent Customer Activity**: The ratio of recent customer interactions (last 7 days) to total customers was calculated to gauge how up-to-date the agent's sales efforts were.

- **Policyholder Ratio**: A ratio of the number of policyholders to the number of policies sold was created to identify how well agents were retaining customers.

- **Activity Decay**: Decay metrics were calculated to measure the decline in proposal and quotation activity over time, helping to spot agents whose performance was declining.

- **Age and Experience Categories**: Age and experience were categorized into groups (e.g., '<25', '25-35', etc.) to make it easier to identify patterns by demographic or experience level.

3. Target Variable

The target variable for the prediction model was created by classifying agents who had made no sales (i.e., `new_policy_count == 0`) in the current month as "NILL". This binary target variable (1,0) was then used to train the predictive model.

4. Data Preprocessing

- **Handling Missing Data**: Missing values in numeric features were filled with the median of the column, while missing categorical values were filled with the mode.

- **Feature Scaling and Encoding**: Numerical features were scaled using standardization (`StandardScaler`), and categorical features were one-hot encoded using `OneHotEncoder` to ensure they could be used in machine learning models.

5. Model Training

A series of machine learning models were trained to predict which agents are likely to go "NILL":

- **Models Used**:
    - **Random Forest Classifier**: An ensemble model that builds multiple decision trees and aggregates their predictions.
    - **Gradient Boosting Classifier**: A boosting method that combines weak learners (decision trees) in a sequential manner to improve prediction accuracy.
    - **XGBoost**: An optimized version of gradient boosting that is known for its efficiency and high performance.

Each model was trained using the prepared features and target variable, and the performance was evaluated using accuracy. The model that achieved the best accuracy was selected as the final model - `RandomForest`

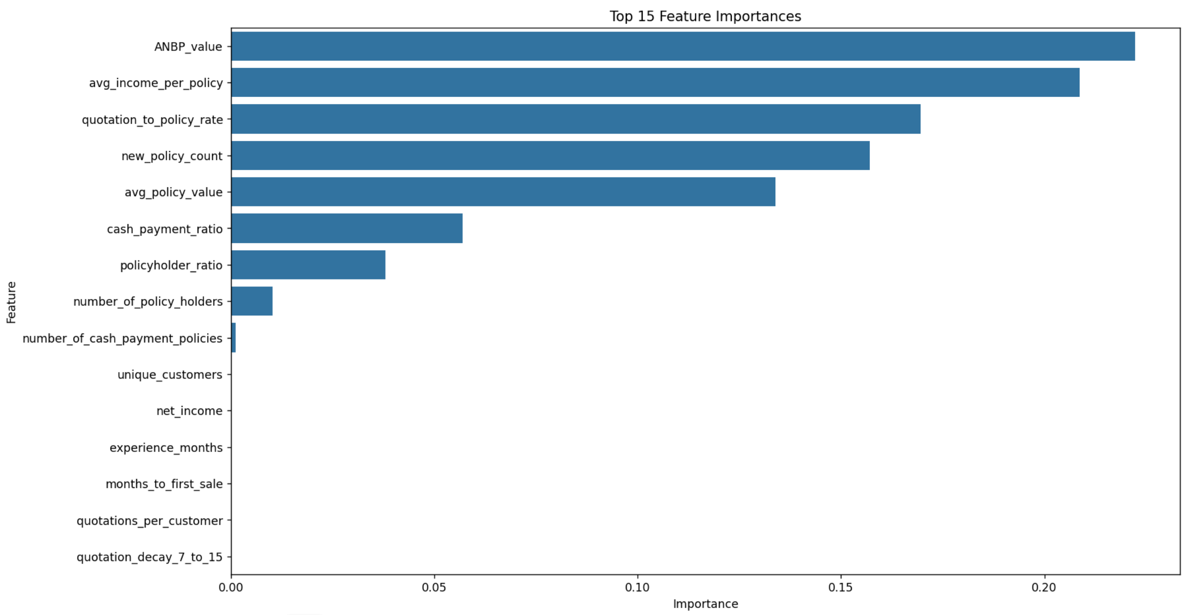
6. Feature Importance

Once the best model was selected, the importance of each feature in predicting "NILL" agents was extracted. This step is crucial for understanding which factors contribute most to the prediction, enabling the identification of key drivers of agent success or failure.

The features were ranked by their importance, and the top features were highlighted to provide actionable insights into which behaviors and characteristics matter most for agent performance. These insights can inform targeted interventions and help predict future "NILL" agents.

Top 10 factors affecting agent performance:

1. ANBP_value: 0.2222
2. avg_income_per_policy: 0.2087
3. quotation_to_policy_rate: 0.1695
4. new_policy_count: 0.1571
5. avg_policy_value: 0.1339
6. cash_payment_ratio: 0.0570
7. policyholder_ratio: 0.0380
8. number_of_policy_holders: 0.0102
9. number_of_cash_payment_policies: 0.0012
10. unique_customers: 0.0002

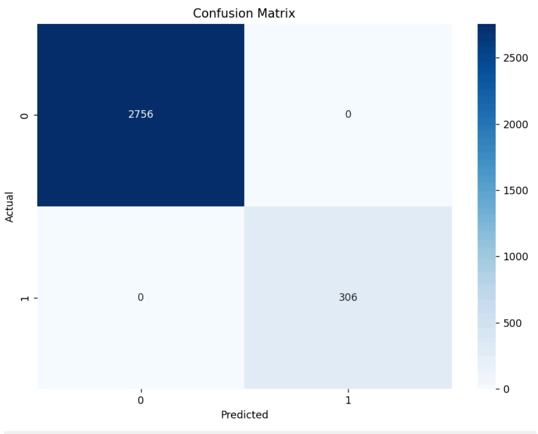


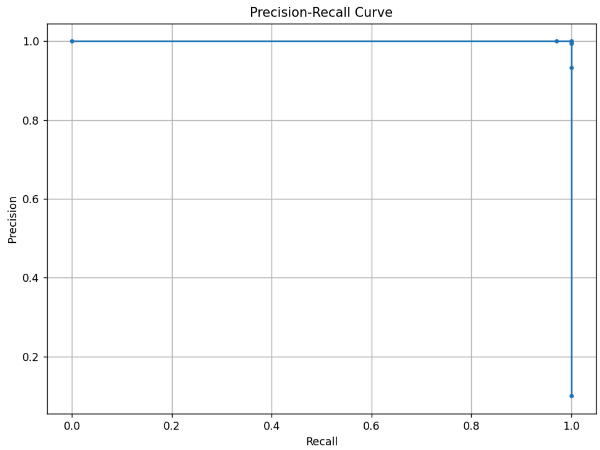
7. Model Evaluation

To evaluate model performance, the following metrics were used:

- **Confusion Matrix**: This showed the true positive, true negative, false positive, and false negative predictions, allowing us to assess the model's accuracy.

- **Precision-Recall Curve**: The precision-recall curve provided insight into the model's performance at various thresholds and was especially useful in imbalanced classification problems, where the number of "NILL" agents may be much smaller than the non-"NILL" agents.





8. Personalized SMART action plans

1. For Agents with Low Proposal Activity

These agents need to focus on increasing their outreach to potential customers.

Action Plan:

- **Daily Goal**: Contact at least 5 potential customers.
- **Weekly Training**: Attend training on prospecting techniques.
- **Mentorship**: Pair with a high-performing mentor for shadowing to learn best practices.

---

2. For Agents with Good Proposals but Low Conversion to Quotations

These agents are good at generating proposals but struggle with converting them into quotations.

Action Plan:

- **Presentation Skills**: Focus on improving presentation skills to create more compelling proposals.
- **Product Knowledge**: Review product knowledge training to ensure they are able to highlight the right features.
- **Objection Handling**: Practice objection handling scenarios to close more proposals.

---

3. For Agents with Good Quotations but Low Policy Conversion

These agents have strong quotations but face challenges in converting those to policies.

Action Plan:

- **Closing Techniques**: Provide training on advanced closing techniques to help agents seal the deal.
- **Review Product Fit**: Revisit pricing and product fit to ensure that the offering is appealing to the clients.
- **Follow-up System**: Implement a system for regular follow-ups on pending quotations to convert more leads.

---

4. For Agents with Inconsistent Activity (High Decay Metrics)

These agents demonstrate fluctuating performance, and more consistent effort is needed to stabilize their output.

Action Plan:

- **Daily Activity Tracking**: Implement daily activity tracking to monitor progress and maintain consistent performance.
- **Set Weekly Targets**: Set weekly rather than monthly targets to keep agents motivated and on track.
- **Regular Check-ins**: Schedule regular check-ins with managers to offer guidance and encouragement.

4. Performance Interventions

Once agents have been classified into different performance categories (Low, Medium, High), personalized intervention plans are developed to help each group improve. These interventions are based on the performance tier of each agent, ensuring that the support is tailored to the needs of the agent in question. The classification and intervention strategy also utilize **K-Means clustering** and to ensure the approach is robust and based on actual performance patterns.

1. Method to Classify Current Agent Performance (Low, Medium, High)

To classify agents into performance categories, we first analyze their past and current performance data. The performance metrics considered include:

- **Sales Metrics**: Number of policies sold, total income, and average policy value.
- **Activity Metrics**: Proposal activity, quotation activity, and conversion rates.
- **Engagement Metrics**: Engagement with clients and proposals, as well as any trends or decays in their activity over time.

Using **K-Means clustering**, agents are grouped into three clusters based on their performance metrics. These clusters are mapped to performance levels as follows:

- **Low Performers**: These agents consistently underperform. Their sales figures are low, and they show little engagement with the sales pipeline (e.g., low proposal and quotation rates).

- **Medium Performers**: Agents in this category show moderate performance. They may have inconsistent sales patterns or show potential but lack the necessary support or guidance to break into higher performance levels.

- **High Performers**: These agents consistently perform well, surpassing targets and demonstrating high engagement with the sales process. They are top performers who are highly productive and may be more experienced or better equipped with sales techniques..

```
Agent Performance Classification Summary:
                   count  new_policy_count    ANBP_value     net_income
performance_level
High                6994         27.700029  1.553551e+06  314265.227910
Low                 1528          0.000000  0.000000e+00  224314.337042
Medium              6786         17.176393  7.118090e+05  140013.897878
```

2. Intervention Strategy Based on Performance Category

For each performance category, a tailored intervention strategy `agent_performance_classified.csv`is developed to help agents improve. The specific interventions are designed to address the challenges faced by agents in each category:

- **Low Performers**:
  - **Additional Training**: Low performers often struggle due to a lack of knowledge or skills. Additional training can include both product knowledge and sales techniques.
  - **Mentoring**: Pairing low performers with more experienced agents can help them learn from their peers and receive guidance on overcoming challenges.
  - **Motivational Support**: These agents may be disengaged or lack confidence. Providing motivational support and setting small, achievable goals can help them rebuild their confidence.
  - **Close Monitoring**: Regular check-ins and performance reviews can help track progress and provide necessary adjustments to their training or workload.
- **Medium Performers**:
  - **Targeted Feedback**: Medium performers may benefit from specific, actionable feedback on areas they need to improve. This feedback should be data-driven, pointing to where they fall short (e.g., conversion rates)
  - **Coaching**: Personalized coaching sessions focusing on specific skills or behaviors can help agents transition from medium to high performance.
  - **Increased Responsibility**: Providing medium performers with additional responsibilities or opportunities for growth can motivate them to push themselves further and improve their performance.
  - **Motivation Programs**: Recognizing medium performers and offering incentives or rewards for improved performance can also help boost morale and increase productivity.
- **High Performers**:
  - **Career Development Opportunities**: High performers should be given opportunities to further advance their careers, such as leadership training or promotional opportunities. They should be empowered to mentor others and take on more responsibility.
  - **Reward Systems**: Offering performance-based rewards such as bonuses, recognition, or exclusive incentives can help retain top talent and keep them motivated.
  - **Skills Enhancement**: Even high performers can benefit from further skill development. Offering advanced sales techniques or strategic training can help them maintain their edge.

- **Autonomy and Leadership Roles**: High performers may prefer autonomy in their work. Allowing them to take more ownership of their projects or lead a team can enhance their job satisfaction and performance.

```
Example Intervention Strategies:

Agent Code: 455ca878 (Performance Level: Low)
1. Sales Training: Focus on improving conversion techniques and objection handling.
2. Activity Management: Set daily prospecting targets and provide closer supervision.
4. Weekly Performance Review: Schedule weekly one-on-one sessions to review metrics and provide feedback.
5. Product Knowledge: Complete refresher course on core products.


Agent Code: 62154055 (Performance Level: Medium)
4. Specialized Product Training: Advanced training on high-margin products.
5. Monthly Group Coaching: Join peer group sessions to share best practices.
...
3. Advanced Sales Techniques: Training on complex products and high-net-worth client acquisition.
4. Recognition Program: Highlight achievements in company communications and events.
5. Career Path Planning: Set long-term goals and development plan for advancement.
```

3. Progress Tracker to Measure Changes Over Time

To ensure the interventions are effective, a **progress tracker** `agent_progress_tracker.csv` is implemented. This tool helps monitor agents' performance before and after the intervention, allowing for real-time adjustments to strategies based on actual performance changes.