**Final Case Assignment**

**Introduction:**

Based on the rich dataset provided, I’ve made use of the EDA techniques to derive meaningful insights from the data that it contains to the target variable of matching. Then I’ve built models to incorporate these insights and draw meaningful aspects from these variables for better matching.

**Exploratory Data Analysis & Recommendations based on results:**

**Observations form the variables and box plots:**

The distributions of several of the categorical variables, including relationship history, food preferences, and reading habits, are balanced for both matched and unpaired users. This implies that while they could still have a role when combined with other factors, these factors might not have a significant direct impact on matching success.Since they reveal significant disparities between matched and unmatched users, variables including time spent online, user engagement, and response behavior may be more important.

**Correlation Matrix:**

1. **Focus on Engagement and Interaction:** Users who engage more are likely to stay on the platform longer. Add interactive features like daily challenges, badges, or gamified incentives to boost activity. This will improve retention and increase match chances by providing more exposure to potential connections.
2. **Encourage Profile Completeness:** Full profiles slightly increase match chances. Add a completion feature with reminders or incentives for users to fill in missing details, especially about preferences like activities, dietary needs, or cooking skills. Over time, this will improve profile richness and match rates.
3. **Leverage Activity-Based Matching:** Users with similar activity preferences tend to have better match outcomes. Increase the weight of activity preferences in the matching algorithm by suggesting matches based on shared online activity patterns to enhance compatibility.
4. **Incorporate Behavioral Nudges:** Behavioral nudges can influence user interactions and profile updates. Implement reminders for inactive users or encourage frequent activity participants to connect with others who share similar behaviors, keeping profiles relevant and promoting meaningful interactions.
5. **Refine the Role of Cooking Skills and Dietary Preferences:** While cooking skills and dietary preferences have weak correlations with matches, they can be used in niche segments. Introduce interest groups or recipe-sharing communities to help users connect over shared dietary restrictions or cooking skills, fostering meaningful connections.

**A graph with text on it

Description automatically generatedModel based Analysis and Recommendations:**

**PCA Analysis:** Cooking skill, age, and online activity are key factors in matching algorithms due to their strong correlation with principal components (PC1 and PC2). Pairing users with similar cooking skills and age ranges could improve match quality by aligning common interests and compatibility**.**

**Random Forest:** Random Forest models typically provide better predictive performance by combining multiple decision trees to reduce variance and avoid overfitting.

**Logistic Regression:** The logistic regression model shows significant sensitivity but poor specificity, which means it accurately identifies negative matches ("No") but fails to predict positive matches ("Yes").

**KNN (K-Nearest Neighbors):** The KNN model has a moderate accuracy of 60% and performs well in predicting "No" matches but struggles with predicting "Yes".

**Recommendations:**

1. **Refine Matching Algorithm with Key Features:** Refine the matching algorithm by prioritizing key features like Cooking Skill, Age, and Online Activity, which have the most influence in PCA analysis and models. Implement a model-based approach, such as Random Forest or Logistic Regression, to focus on these features for more relevant matches, and adapt the system over time using ensemble models like Random Forest or Gradient Boosting as new user behavior data is collected.
2. **Incorporate Interaction Effects into the Matching Process:** Incorporate the interaction between Cooking Skill and Age into the matching process, as it plays a significant role in improving match quality. Modify the algorithm to give higher weight to this interaction.
3. **Personalize User Experience Based on Evolving Behavior:** Personalize the user experience by adapting to evolving behavior patterns, as cooking habits and online activity influence match outcomes. Immediately, introduce personalized notifications and match suggestions based on users' cooking preferences and activity. In the long term, implement reinforcement learning models that adapt to user interactions, continually improving recommendations and match suggestions.
4. **Feature Selection and Refinement:** Use feature importance from Random Forest to identify and prioritize the most influential variables in the matching algorithm. Immediately, focus on key features like Cooking Skill, Age, and Online Activity, while removing weakly contributing ones like User ID or Profile Completeness. Over time, continuously reassess feature importance as more data is collected, refining the algorithm to maintain its predictive accuracy.