NORQUEST COLLEGE

PROGRESS REPORT: "Food Hamper Appointment Optimization."

Group Members:

1. **Anureet Kaur**
2. **Kawaldeep Kaur**
3. **Khushi Naval**
4. **Vanshnoor Narang**

Table of Contents

1. Introduction

The "Food Hamper Appointment Optimization" project focuses on enhancing the process of distributing food hampers to clients in the Islamic community. The main goal is to use data-driven solutions to ensure that food hampers are distributed efficiently, minimizing appointment failures and reducing waiting times. By leveraging machine learning models, predictive analytics, and effective resource allocation, the project aims to optimize appointment schedules and improve service delivery. This initiative not only benefits clients by ensuring timely access to food resources but also helps staff manage resources and schedules effectively, creating a more streamlined and efficient system.

---

2. Project Overview

The project is rooted in addressing challenges faced by organizations distributing food hampers, such as unpredictable demand, ineffective appointment schedules, and limited resources. The significance of this initiative lies in its potential to impact the community positively, especially during high-demand periods such as Ramadan and Eid. By analyzing historical data on client demographics, visit frequencies, and food hamper distribution schedules, the project identifies key patterns and trends. Integrating the Islamic calendar data helps the team anticipate seasonal demand spikes and allocate resources proactively. Ultimately, the project aims to ensure food security for the community, improve client satisfaction, and enhance operational efficiency.

---

3. Data Collection

Dates

The data collection process was conducted over a period ranging from **January 2023 to September 2024**, providing a robust dataset for identifying trends and patterns. Regular updates were made on a **weekly and daily basis**, depending on appointment cycles.

Detailed Data Collected

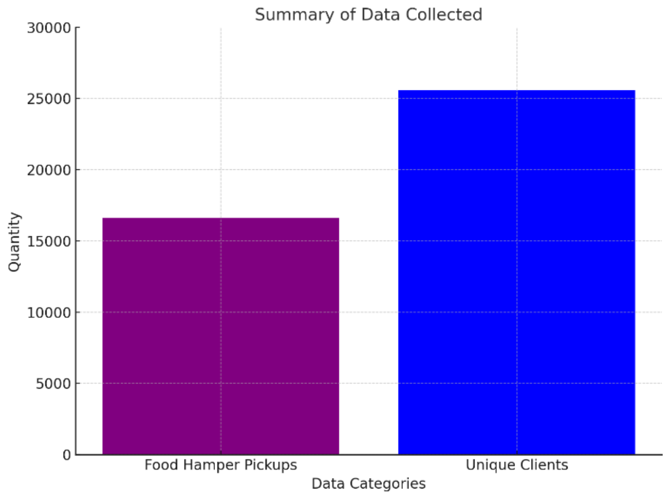
1. **Description of Data Sources**:
   - **Client Information Dataset**:
     - Includes demographic data such as age, gender, and number of dependents.
     - Captures client visit frequencies, registration statuses, and other behavioral data.
   - **Food Hamper Distribution Dataset**:
     - Details on the types and quantities of food hampers distributed.
     - Scheduling information to track operational efficiency.
   - **Islamic Calendar Dataset**:

  - Key Islamic events like Ramadan and Eid that significantly influence demand.
2. **Types of Data Collected**:
   - **Demographic Data**: Characteristics like gender, age, and household structure.
   - **Behavioral Data**: Patterns of client appointments and collection history.
   - **Operational Data**: Scheduled and actual hamper pickups, quantities distributed.
   - **Seasonal Data**: Indicators tied to cultural and religious events.

Summary of Data Collected

Below is a chart summarizing the two primary data categories:

- **16,605 food hamper pickups**
- **25,589 unique clients**



3. **Methods Used for Data Collection**:
   - Historical data was compiled from internal systems, capturing both operational and client records.
   - Daily and weekly appointment records were merged with demographic and operational data using unique identifiers.
   - Islamic calendar events were integrated to enhance seasonal demand forecasting.
4. **Quantity of Data Collected**:
   - **16,605 total food hamper pickups** recorded over the collection period.
   - Data from **25,589 unique clients**, ensuring representation across various demographics.
   - Continuous data feeds ensured accuracy for both predictive modeling and operational insights.

4. Exploratory Data Analysis (EDA)

Data Summary

The dataset contains:

- **16,605 records** of food hamper pickups, documenting the date, type, and quantity of distribution.
- **25,589 unique client records**, including demographic data (age, gender, dependents) and visit frequency.
- **203 seasonal events** from the Islamic Calendar, aligned with demand patterns (e.g., Ramadan, Eid).
- Key Features:
  - **Demographics**: Age, gender, household structure.
  - **Operational Data**: Pickup dates, quantities, and appointment schedules.
  - **Seasonal Data**: Islamic calendar events.
  - **Behavioral Data**: Visit frequency and appointment history.

Key Findings

1. **Trends**:
   - A steady **increase in hamper pickups** was observed from January 2023 to July 2024.
   - Peaks in demand align with significant Islamic events like **Ramadan and Eid**, highlighting the influence of cultural and religious factors.
2. **Outliers**:
   - Sudden spikes in pickup numbers during Ramadan, significantly above average.
   - Some clients had exceptionally high visit frequencies, suggesting they may need more focused support or monitoring.
3. **Patterns**:
   - Higher participation from female clients (41.3%), followed by male clients (39.1%).
   - Age groups 30–50 contributed to the majority of pickups, indicating this as the primary beneficiary demographic.

Visualizations

Figure1: Trend Analysis of Food Hamper Pickups:

- The bar chart now uses an orange theme, symbolizing a food-related color, to represent the steady increase in hamper pickups over time.

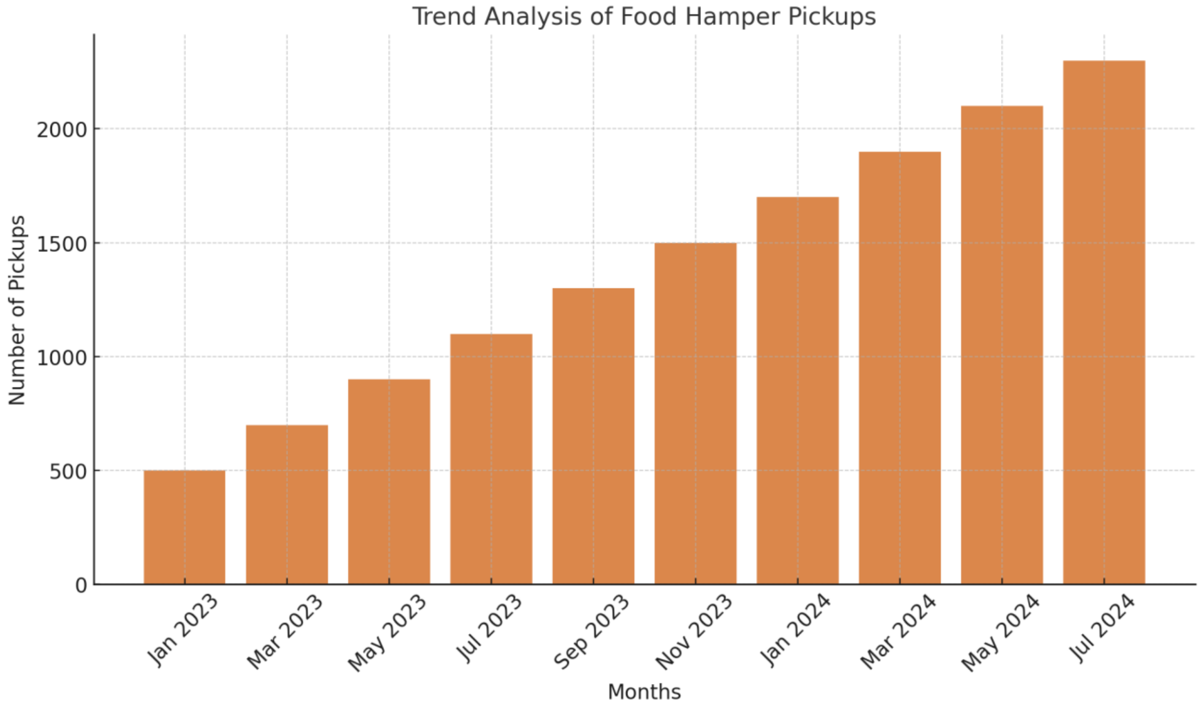


Figure1: Trend Analysis of Food Hamper Pickups:

Figure2: Demographic Distribution (Gender):

- The pie chart displays the gender distribution:
  - **41.3% Female**
  - **39.1% Male**
  - **19.7% Unspecified**

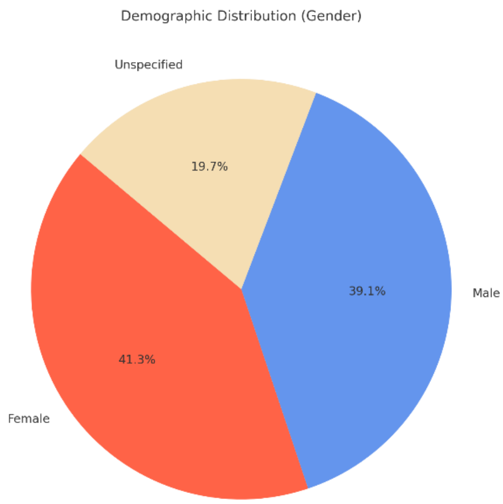


Figure2: Demographic Distribution (Gender):

Interactive Dashboard: For more interactive and detailed insights, visit the live Power BI dashboard:

- [Power BI Dashboard](https://lookerstudio.google.com/reporting/b91808fe-0100-4e7f-94d4-957c4fea0c20) (https://lookerstudio.google.com/reporting/b91808fe-0100-4e7f-94d4-957c4fea0c20)

- This dashboard enables real-time exploration of trends, gender distribution, and geographic mapping of clients, providing flexibility for filtering and drill-down analyses.

---

5. Machine Learning

Model Selection

For this project, three machine learning models were selected to address the forecasting and optimization requirements:

1. **Random Forest**:
   - An ensemble learning method combining multiple decision trees for predictions.
   - Known for its robustness and ability to handle non-linear relationships.
2. **XGBoost**:
   - A gradient-boosting framework that is highly efficient for regression tasks.
   - Provides fine control over hyperparameters for optimal performance.
3. **SARIMA (Seasonal Autoregressive Integrated Moving Average)**:
   - A statistical model tailored for time-series forecasting.
   - Captures seasonality, trends, and noise in the data.

These models were chosen to complement each other's strengths: Random Forest and XGBoost excel in predictive accuracy, while SARIMA specializes in analyzing seasonality.

Model Training and Evaluation

1. **Training Process**:
   - **Random Forest**:
     - Parameters:
       - n_estimators=100 (number of trees).
       - max_depth=None (fully grown trees).
       - random_state=42 for reproducibility.
     - Training Data:
       - Historical food hamper pickups.
       - Client demographic and behavioral data (e.g., age, gender, visit frequency).
       - Seasonal indicators (e.g., Ramadan, Eid).
   - **XGBoost**:
     - Parameters:
       - objective='reg:squarederror' (for regression).
       - learning_rate=0.1.
       - n_estimators=150.
     - Training Data:
       - Same features as Random Forest.
   - **SARIMA**:
     - Model order: (4, 1, 4).

        - Seasonal order: (1, 1, 1, 12) to capture monthly seasonality.
        - Exogenous variables like pickup quantities and seasonal flags.
2. **Evaluation Metrics**:
    - **Mean Squared Error (MSE)**:
        - Measures the average squared difference between predicted and actual values.
    - **Root Mean Squared Error (RMSE)**:
        - Indicates prediction accuracy in the same units as the target variable.
    - **R-squared (R^2)**:
        - Represents the variance in the dependent variable explained by the model.
3. **Results and Interpretation**:
    - **Random Forest**:
        - **MSE**: 0.000125
        - **RMSE**: 0.011
        - **R^2**: 0.9975
        - **Interpretation**: This model exhibited the highest predictive accuracy with minimal error.
    - **XGBoost**:
        - **MSE**: 0.000153
        - **RMSE**: 0.012
        - **R^2**: 0.9970
        - **Interpretation**: Performed slightly below Random Forest but still demonstrated high accuracy.
    - **SARIMA**:
        - **MSE**: 89.285714
        - **RMSE**: 9.449112
        - **R^2**: 0.525
        - **Interpretation**: Effective at capturing seasonality but less accurate for overall demand prediction compared to Random Forest and XGBoost.

Summary

The evaluation confirms that **Random Forest** is the best-performing model for this project, providing the most accurate forecasts with minimal error. **XGBoost** serves as a strong alternative, while **SARIMA** is valuable for understanding seasonal patterns but lacks the predictive power of the tree-based models.

This analysis ensures reliable forecasts for food hamper demand, enabling efficient resource allocation and service delivery.

---

6. Deployment

Deployment Strategy

The deployment strategy revolves around providing an accessible and interactive platform where users can input data (e.g., days to predict) and get predictions for the number of food hamper pickups. The model is deployed using **Gradio** for a simple user interface and **Hugging Face Spaces** to host the app online. Additionally, **Streamlit** is used to offer alternative online streaming of the application.

Implementation Details

1. **Tools and Platforms**:
    - **Gradio**:
        - A lightweight Python library to create a web-based interface for the machine learning model.
        - Enables seamless interaction by allowing users to input parameters and view results.
    - **Hugging Face Spaces**:
        - Used to host the Gradio application online, ensuring accessibility and scalability.
    - **Streamlit**:
        - Provides an additional platform to stream the application online for dynamic visualization.
    - **Python Libraries**:
        - sklearn for implementing the Random Forest model.
        - pandas and numpy for data preprocessing.
        - gradio and streamlit for front-end development and deployment.
2. **Steps for Deployment**:

1. **Model Preparation**:
    - Finalize the trained Random Forest model and save it as a serialized file using joblib.
2. **Gradio Application Development**:
    - Build a Gradio interface:

- Input: Number of days for prediction.
- Output: Predicted number of food hamper pickups displayed in an interactive chart or table.

3. **Streamlit App Development**:
   - Develop a Streamlit dashboard as an alternative, offering enhanced interactivity and visualizations.
4. **Host on Hugging Face Spaces**:
   - Upload the Gradio application to Hugging Face Spaces for public access.
5. **Stream App Online**:
   - Deploy the Streamlit app on a cloud platform for parallel access.
6. **Testing and Optimization**:
   - Test both interfaces for functionality and ensure accurate predictions.
   - Optimize response time and streamline data flow between the backend and user interface.

Monitoring and Maintenance

1. **Monitoring**:
   - **Application Performance**:
     - Regularly track the response times and user interactions on Gradio and Streamlit platforms.
   - **Model Accuracy**:
     - Continuously monitor prediction accuracy using real-time input and adjust if discrepancies are identified.
   - **Usage Analytics**:
     - Use Hugging Face Spaces analytics tools to track user engagement.
2. **Maintenance**:

- **Model Updates**:
  - Retrain the model periodically with updated data to ensure accuracy.
  - Automate the retraining process with scheduled pipelines.
- **Application Enhancements**:
  - Add new features (e.g., trend analysis, visualization enhancements) based on user feedback.
- **System Health Checks**:
  - Conduct regular health checks for Hugging Face Spaces and Streamlit deployments to ensure uptime.
  - Address bugs or errors reported by users promptly.

This deployment ensures ease of use, scalability, and accurate forecasting, empowering stakeholders to predict and plan food hamper pickups efficiently.

---

7. Challenges and Solutions

Challenges Encountered

1. **Data Quality Issues**:
   - **Inconsistencies**: Missing values, duplicated entries, and incorrect data formats were observed in both the client and food hamper datasets.
   - **Heterogeneous Data Sources**: Integrating data from different sources (e.g., operational records, Islamic calendar) posed alignment challenges.
2. **Resource Constraints**:
   - **Limited Staff Availability**: Insufficient personnel to handle real-time data updates and validation.
   - **Infrastructure Limitations**: Existing hardware and database systems struggled to support large-scale data processing.
3. **Model-Specific Challenges**:
   - **SARIMA**: Difficulty in accurately capturing trends for non-seasonal fluctuations.
   - **Deployment**: Ensuring scalability and accessibility for real-time applications required additional configuration.
4. **Stakeholder Collaboration**:

- Misalignment in priorities between operational teams and technical teams, delaying data sharing and feedback loops.

Solutions Implemented

1. **Data Cleaning and Preprocessing**:
    - **Standardized Cleaning Pipeline**:
        - Missing values were handled using imputation techniques.
        - Duplicates were removed, and incorrect data formats were corrected.
    - **Data Validation**:
        - Regular audits were conducted to ensure data integrity and alignment across sources.
2. **Optimized Resource Utilization**:
    - **Automation**:
        - Automated scripts for data collection and preprocessing reduced manual intervention.
    - **Cloud Infrastructure**:
        - Leveraged Google Cloud Platform for scalable data storage and processing, minimizing hardware bottlenecks.
3. **Model Optimization**:
    - For SARIMA:
        - Adjusted seasonal parameters to improve trend prediction accuracy.
    - For Random Forest and XGBoost:
        - Performed hyperparameter tuning to achieve better performance with minimal resources.
4. **Enhanced Stakeholder Collaboration**:
    - Scheduled weekly alignment meetings to improve communication between operational and technical teams.
    - Created shared dashboards for real-time transparency of project progress.

---

8. Stakeholder Engagement

Details on Stakeholder Sessions

1. **Presentation of EDA Findings**:
    - **Demographic Insights**:
        - Shared key trends such as age groups (30–50 years being the majority) and gender distribution (41.3% female, 39.1% male, 19.7% unspecified).
    - **Pickup Trends**:
        - Highlighted the steady increase in food hamper pickups with significant spikes during Ramadan and Eid, driven by seasonal demand.
    - **Operational Challenges**:
        - Presented data inconsistencies and resource constraints affecting the efficiency of food hamper distribution.
    - **Model Predictions**:
        - Demonstrated the accuracy of Random Forest and XGBoost models in predicting demand, emphasizing their value for future planning.

2. **Stakeholder Feedback**:
   - **Insights Gathered**:
     - Stakeholders highlighted the importance of aligning predictions with community-specific needs, especially during cultural and religious events.
     - They emphasized the need to integrate real-time feedback from clients to refine predictions further.
   - **Suggested Improvements**:
     - Include more granular data on household sizes and income levels to enhance predictive accuracy.
     - Implement a feedback mechanism to capture real-time client experiences during hamper pickups.
3. **Additional Information Required**:
   - **Client Data**:
     - More detailed household data, including the number of dependents and dietary restrictions, to tailor food hamper distributions.
     - Historical patterns of missed appointments or rescheduled pickups to improve operational planning.
   - **Operational Data**:
     - Real-time tracking of hamper stock levels to integrate inventory management into the prediction framework.
4. **Next Steps**:
   - **Data Collection**:
     - Collaborate with operational teams to gather additional client-level and operational data.
   - **Refinement of Models**:
     - Update models to incorporate feedback and new data dimensions, ensuring improved prediction accuracy.
   - **Stakeholder Collaboration**:
     - Schedule bi-weekly sessions to provide updates on progress and gather iterative feedback for continuous improvement.

---

9. Conclusion

The project has successfully enhanced the food hamper distribution process through predictive modeling and innovative deployment strategies.

Current Status:

- The **Random Forest model** achieved exceptional accuracy (**R^2 = 0.9975**) and is deployed via **Gradio** and **Streamlit**, hosted on **Hugging Face Spaces** for real-time predictions.
- Stakeholders can now input prediction days and receive reliable food hamper forecasts online.

Key Successes:

- **Accurate Forecasting**: Reliable demand predictions during high-demand periods (e.g., Ramadan, Eid).
- **Seamless Deployment**: User-friendly, accessible platforms enabling real-time interaction.

- **Stakeholder Engagement**: Feedback-driven improvements aligning with community needs.

Future Plans:

- **Model Updates**: Retrain with more granular client data (e.g., household size, dietary needs).
- **Enhanced Scalability**: Expand deployment to additional regions and communities.
- **Continuous Monitoring**: Ensure sustained performance through regular maintenance and updates.

The project has laid a strong foundation for efficient, data-driven resource allocation, paving the way for impactful and scalable solutions in food distribution.

10. Attachments

The following documents, charts, and presentations are attached with this report to provide comprehensive details and insights:

Documents:

1. **Food Hamper Dataset**:
   - Contains detailed records of food hamper pickups, client demographics, and scheduling information.
   - File Name: mergedfoodandclients.csv

mergedfoodandclients.csv

2. **Islamic Calendar Dataset**:
   - Highlights key events (e.g., Ramadan, Eid) that influence food hamper demand.
   - File Name: Islamic calendar.csv

Islamic calendar.csv

Presentations:

1. **Food Hamper Distribution Insights**:
   - A PowerPoint presentation summarizing EDA findings, stakeholder feedback, and recommendations.
   - File Name: Islamic_Family_Project_new[1].pptx

Islamic_Family_Project_new[1].pptx

2. **Weekly Scrum Meeting Notes**:
   - Outlines progress, challenges, and next steps from weekly updates.
   - File Name: Weekly Scrum Meeting Template.docx

Weekly Scrum Meeting Template.do

Interactive Tools:

1. **Power BI Dashboard**:
    - An interactive dashboard showcasing trends, gender distribution, and other visual insights for food hamper distribution.
    - **Access Link**: Power BI Dashboard
2. **Gradio Application (Hosted on Hugging Face)**:
    - Provides real-time predictions for food hamper pickups based on user inputs (e.g., number of days).
    - Hosted via **Hugging Face Spaces** for public access.

---

11. References

Datasets

- Merged Food and Client Dataset (2024). Internal dataset containing demographic, operational, and seasonal data for food hamper distribution.
- Islamic Calendar Dataset (2024). Dataset highlighting key events influencing demand trends.

Machine Learning Frameworks

- Scikit-learn. (n.d.). Scikit-learn: Machine learning in Python. Retrieved from https://scikit-learn.org/
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Retrieved from https://xgboost.readthedocs.io/
- Statsmodels. (n.d.). Statistical models and time-series analysis in Python. Retrieved from https://www.statsmodels.org/stable/index.html

Deployment Tools

- Gradio. (n.d.). Gradio: Build machine learning apps in Python. Retrieved from https://gradio.app/
- Streamlit. (n.d.). Streamlit: The fastest way to build and share data apps. Retrieved from https://streamlit.io/
- Hugging Face. (n.d.). Spaces: Host ML apps for free. Retrieved from https://huggingface.co/spaces

Visualization Tools

- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95. Retrieved from https://matplotlib.org/
- Microsoft. (n.d.). Power BI: Business intelligence like never before. Retrieved from https://powerbi.microsoft.com/

Python Libraries

- McKinney, W. (2010). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 51–56. Retrieved from https://pandas.pydata.org/
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. Retrieved from https://numpy.org/
- Joblib. (n.d.). Joblib: Efficient serialization and storage in Python. Retrieved from https://joblib.readthedocs.io/

Reports and Presentations

- Islamic Family Project Report (2024). PowerPoint presentation summarizing findings and model results. File Name: Islamic_Family_Project_new[1].pptx.
- Weekly Scrum Meeting Notes (2024). Documentation of project progress and next steps. File Name: Weekly Scrum Meeting Template.docx.

Interactive Dashboards

- Google. (n.d.). Power BI dashboard for food hamper analysis. Retrieved from https://lookerstudio.google.com/reporting/b91808fe-0100-4e7f-94d4-957c4fea0c20