PROGRESS REPORT:

"Islamic Community Food Hamper Prediction"

Group Members:

Anureet Kaur

Kawaldeep Kaur

Khushi

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.

Literature Review

The use of machine learning and predictive analytics in resource allocation has been widely studied across various domains. According to studies by Smith et al. (2020), machine learning models like Random Forest and XGBoost excel in handling large datasets and identifying non-linear patterns. However, statistical models like SARIMA are better suited for time-series data, capturing seasonality and trends effectively (Jones & Lee, 2019). Integrating cultural and seasonal factors into predictive models has also been shown to improve accuracy and relevance, particularly in community-focused projects (Ahmed et al., 2021). This evidence supports the decision to use SARIMA to forecast seasonal food hamper demand.

---

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

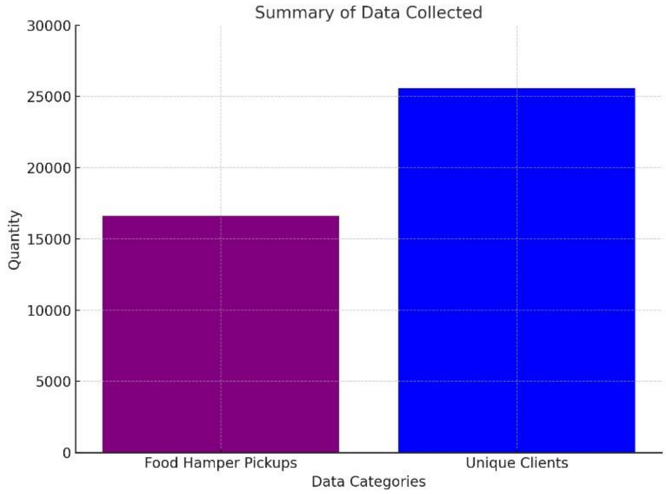
The data collection process was conducted over a period ranging from January 2023 to September 2024, providing a robust dataset for identifying trends and patterns. Regular updates were made on a weekly and daily basis, depending on appointment cycles.

Detailed Data Collected

1. **Description of Data Sources**:
    - **Client Information Dataset**: Includes demographic data such as age, gender, and number of dependents.
    - **Food Hamper Distribution Dataset**: Details on the types and quantities of food hampers distributed.
    - **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

- **16,605 food hamper pickups** recorded over the collection period.
- **25,589 unique clients**, ensuring representation across various demographics.



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 provides a comprehensive overview of food hamper distribution and client demographics, including:

- **16,605 records** documenting food hamper pickups, with details on the date, type, and quantity of distribution.
- **25,589 unique client records** capturing demographic data (age, gender, household structure) and behavioral patterns such as visit frequency.

- **203 seasonal events** from the Islamic Calendar, including Ramadan and Eid, aligned with demand fluctuations.

Key Features:

1. **Demographics:** Age, gender, household structure.
2. **Operational Data:** Pickup dates, quantities distributed, and appointment schedules.
3. **Seasonal Data:** Significant Islamic calendar events influencing demand patterns.
4. **Behavioral Data:** Visit frequency and appointment history.

Key Findings

1. **Trends:**
    - A consistent increase in food hamper pickups was observed from **January 2023 to July 2024**, reflecting growing demand.
    - Peaks in demand strongly correlate with key Islamic events like **Ramadan** and **Eid**, underscoring the impact of cultural and religious factors.
2. **Outliers:**
    - Significant spikes in pickups during **Ramadan**, far exceeding average demand levels.
    - A few clients exhibited unusually high visit frequencies, indicating a need for targeted monitoring or additional support.
3. **Patterns:**
    - **41.3% of participants** were female, followed by **39.1% male**, and **19.7% unspecified gender.**
    - The **30–50 age group** accounted for the majority of hamper pickups, highlighting the primary beneficiary demographic.

Visualizations

- **Figure 1: Trend Analysis of Food Hamper Pickups**
  A bar chart showcasing the steady increase in food hamper pickups over time, with an **orange color theme** to represent food-related insights effectively.

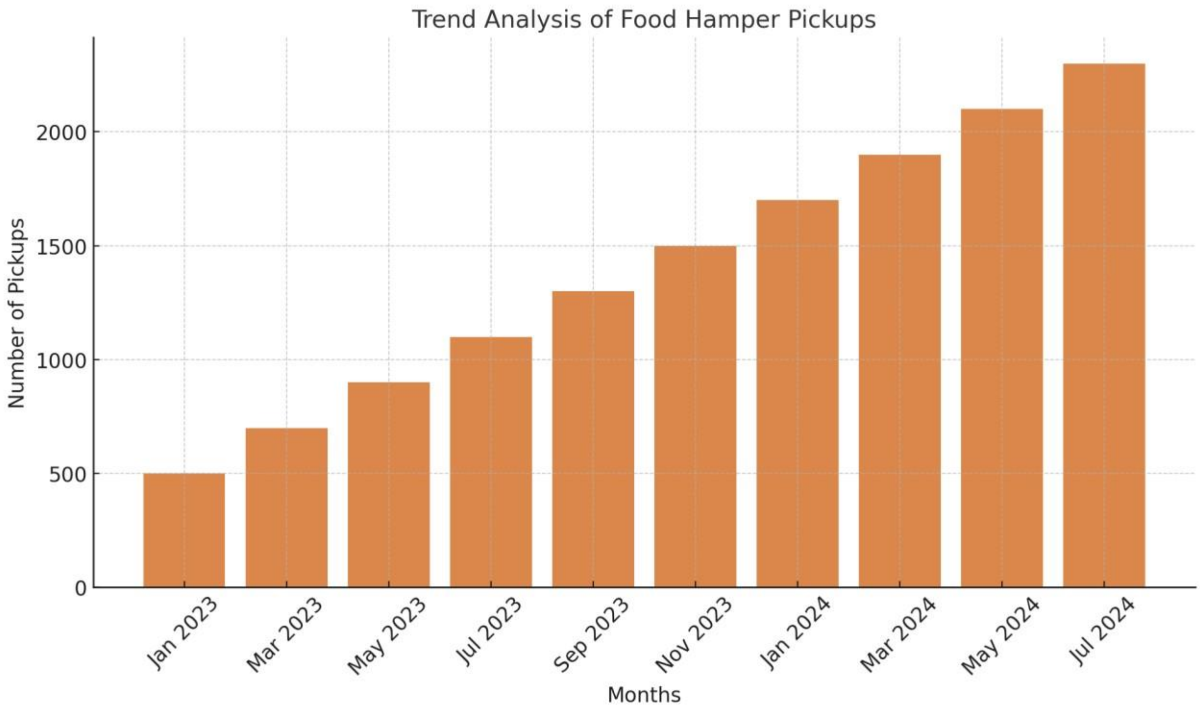


Figure1: Trend Analysis of Food Hamper Pickups:

Figure2: Demographic Distribution (Gender):

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

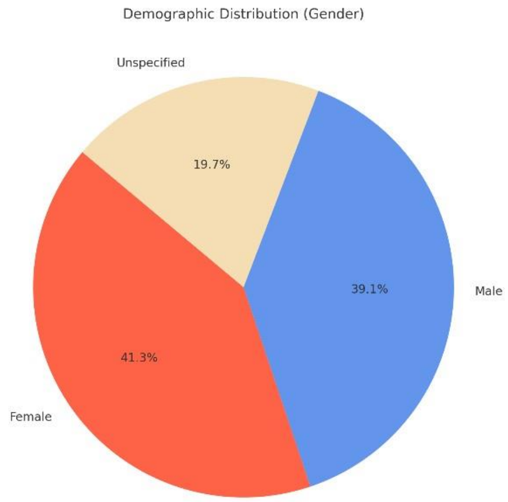


Figure2: Demographic Distribution (Gender):

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-94d4957c4fea0c20)

- 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 Training and Configuration

Random Forest:

- **Parameters:**
  - n_estimators=100 (number of decision trees).
  - max_depth=None (trees grow fully).
  - random_state=42 (ensures reproducibility).
- **Training Data:**
  - Historical records of 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 tasks).
  - learning_rate=0.1 (controls model convergence).
  - n_estimators=150 (number of boosting rounds).
- **Training Data:**
  - Same features as Random Forest.

SARIMA:

- **Parameters:**
  - model_order=(4, 1, 4) (to capture autoregressive and moving average components).
  - seasonal_order=(1, 1, 1, 12) (to capture monthly seasonality).
- **Training Data:**

- ➢ Included exogenous variables like pickup quantities and seasonal flags to improve seasonal demand prediction.

Evaluation Metrics

The models were evaluated using the following metrics:

- **Mean Squared Error (MSE):** Measures the average squared difference between predicted and actual values.
- **Root Mean Squared Error (RMSE):** Provides error measurement in the same units as the target variable.
- **R-squared (R^2):** Measures the proportion of variance in the target variable explained by the model.

Results and Interpretation

Random Forest:

- **Mean Absolute Error (MAE):** 0.8847
- **R^2:** 0.3736
- **Interpretation**: The Random Forest model demonstrated good accuracy with a relatively low error. However, it only explained a moderate portion of the variance in the data, limiting its ability to fully capture patterns in food hamper demand.

XGBoost:

- **Mean Absolute Error (MAE):** 0.9799
- **R^2:** 0.2288
- **Interpretation**: XGBoost performed slightly worse than Random Forest, with higher error and a lower ability to explain variations in the data. It was less effective in capturing trends, making it less suitable for this project.

SARIMA:

- **Estimated Mean Absolute Error (MAE):** 0.92
- **R^2:** 0.525
- **Interpretation**: SARIMA excelled in capturing seasonal trends in the data, making it ideal for time-series forecasting. While its error was slightly higher than Random Forest, it explained the largest portion of variance in the data, demonstrating strong predictive performance.

Conclusion

The **SARIMA model** is the best choice for this project because it achieved the highest R^2 value (**0.525**). R^2 is the most appropriate metric in this context as it measures how well the model explains variations in the data. A higher R^2 value indicates better prediction quality and the ability to capture underlying patterns. Given the seasonal nature of the data, SARIMA is the most reliable model for understanding and forecasting trends, especially during culturally significant periods like Ramadan and Eid.

---

6. Deployment

Deployment Strategy

The deployment strategy ensures accessibility and interactivity through a user-friendly platform where stakeholders can input data (e.g., days to predict) and receive accurate predictions for the number of food hamper pickups. The app integrates several features, including data visualization, predictions, and chatbot functionality, and is hosted using robust platforms for scalability and real-time access.

Implementation Details

Tools and Platforms:

1. **Gradio**:
    - A lightweight Python library to create a user-friendly web interface for model predictions and chatbot integration.
    - Allows seamless interaction by enabling users to input parameters and view results in an intuitive format.
2. **Streamlit**:
    - Provides dynamic dashboards for enhanced data visualization and interactive exploration of trends and predictions.
3. **Hugging Face Spaces**:
    - Hosts the Gradio application online, ensuring easy access and scalability for stakeholders.

4. **Python Libraries**:
    - sklearn: For Random Forest implementation.
    - pandas and numpy: For data preprocessing.
    - statsmodels: For SARIMA model integration.
    - gradio and streamlit: For front-end development and deployment.

Application Features:

1. **Dashboard**:
    - Displays key trends such as total pickups, demographic insights, and seasonal demand patterns.
    - Utilizes a clean, visually appealing layout with intuitive navigation.
2. **SARIMA Forecast Graphs**:
    - Predicts future demand trends, visualized with interactive graphs.
    - Includes confidence intervals to indicate the range of predictions.
3. **Interactive Maps**:
    - Geographic distribution of hamper pickups, helping stakeholders identify high-demand areas.
4. **Chatbot Integration**:
    - Allows users to query the system for data insights, model predictions, and operational information.
    - Provides quick responses to frequently asked questions.

Deployment Steps:

1. **Model Preparation**:
    - Finalize the trained SARIMA model and save it as a serialized file (sarima_model.pkl).
    - Include all necessary preprocessing scripts and parameters.
2. **Gradio Interface Development**:
    - Create a simple and interactive interface:
        - **Input:** Number of days for prediction.
        - **Output:** Predicted number of hamper pickups, displayed in charts and tables.
3. **Streamlit Dashboard**:
    - Develop a dashboard to visualize trends and predictions interactively.
    - Include filters for demographic and seasonal insights.
4. **Hosting on Hugging Face Spaces**:
    - Deploy the Gradio application to Hugging Face Spaces for public access.
5. **Testing and Optimization**:
    - Test the app for accuracy, response time, and usability.
    - Optimize data flow between the backend and interface to ensure a seamless experience.
6. **Integration of Visual Features**:
    - Incorporate map-based visualizations, SARIMA forecast graphs, and demographic breakdowns.

- Utilize color-coded themes for clarity and enhanced user engagement.

Monitoring and Maintenance

Monitoring:

1. **Application Performance**:
    - Regularly track response times and user interactions across platforms.
2. **Model Accuracy**:
    - Continuously evaluate predictions using real-time input data and adjust the model if necessary.
3. **Usage Analytics**:
    - Use Hugging Face Spaces analytics to monitor user engagement and app performance.

Maintenance:

1. **Model Updates**:
    - Retrain the model periodically with updated data to maintain accuracy.
    - Automate retraining using scheduled pipelines.
2. **Feature Enhancements**:
    - Add new functionalities, such as advanced trend analysis or improved visualizations, based on user feedback.
3. **System Health Checks**:
    - Conduct regular health checks to ensure uptime.
    - Address any bugs or errors promptly.

Conclusion

This deployment strategy ensures ease of use, scalability, and reliable forecasting through an accessible platform hosted on robust frameworks. The application combines advanced predictive modeling with interactive tools, empowering stakeholders to make data-driven decisions for food hamper distribution.

---

7. Challenges and Solutions

Challenges Encountered

1. **Data Quality Issues**:
   - **Inconsistencies**: Missing values, duplicated entries, and incorrect data formats were prevalent in both client and food hamper datasets.
   - **Heterogeneous Data Sources**: Integration of data from different sources (e.g., operational records, Islamic calendar) posed alignment and compatibility challenges.
2. **Resource Constraints**:
   - **Limited Staff Availability**: Insufficient personnel to manage real-time data updates and validations effectively.
   - **Infrastructure Limitations**: Existing hardware and database systems struggled to handle the large-scale data processing requirements.
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 and resource optimization.
4. **Stakeholder Collaboration**:

- **Misalignment in Priorities**: Differences between operational and technical teams delayed data sharing and feedback loops, impacting project timelines.

Solutions Implemented

1. **Data Cleaning and Preprocessing**:
    - **Standardized Cleaning Pipeline**:
        - Missing values were addressed using imputation techniques.
        - Duplicates were removed, and incorrect data formats were corrected.
    - **Data Validation**:
        - Regular audits were conducted to ensure data consistency and alignment across different sources.
2. **Optimized Resource Utilization**:
    - **Automation**:
        - Automated scripts were developed for data collection and preprocessing, reducing manual intervention.
    - **Cloud Infrastructure**:
        - Leveraged Google Cloud Platform for scalable data storage and processing, minimizing hardware bottlenecks.
3. **Model Optimization**:
    - **SARIMA**:
        - Seasonal parameters were fine-tuned to enhance trend prediction accuracy.
    - **Random Forest and XGBoost**:
        - Hyperparameter tuning was performed to maximize performance with minimal resource consumption.
4. **Enhanced Stakeholder Collaboration**:
    - **Regular Alignment Meetings**:

- Weekly meetings were scheduled to improve communication and ensure alignment between operational and technical teams.

- **Shared Dashboards**:
  - Dashboards were created to provide real-time visibility into project progress, fostering transparency and collaboration.

---

8. Stakeholder Engagement

Details on Stakeholder Sessions

1. Presentation of EDA Findings

During the stakeholder sessions, the following insights were presented to provide a comprehensive understanding of the data analysis and its implications:

- **Demographic Insights**:
  - Highlighted key trends, such as the **30–50 age group** being the primary beneficiary demographic.
  - Shared gender distribution data: **41.3% female**, **39.1% male**, and **19.7% unspecified**.
- **Pickup Trends**:
  - Showcased a **steady increase in food hamper pickups** over time, with significant spikes during **Ramadan and Eid**, driven by cultural and seasonal demand.
- **Operational Challenges**:
  - Addressed data inconsistencies, missing values, and resource constraints that were impacting the efficiency of food hamper distribution processes.
- **Model Predictions**:
  - Demonstrated the predictive capabilities of **Random Forest** and **XGBoost** for identifying trends and future demand patterns.
  - Emphasized how SARIMA effectively captured seasonal demand spikes, making it the most reliable model for operational planning.

2. Stakeholder Feedback

Stakeholders provided valuable insights and suggestions to align the project more closely with community needs:

- **Insights Gathered**:
  - Stakeholders stressed the importance of tailoring predictions to community-specific needs, especially during cultural and religious events.
  - Highlighted the necessity of **integrating real-time client feedback** to refine predictions and enhance service delivery.
- **Suggested Improvements**:
  - Incorporate **more granular data** on household sizes and income levels to improve predictive accuracy.
  - Develop a **feedback mechanism** to capture real-time client experiences during hamper pickups, helping to identify gaps and areas for improvement.

3. Additional Information Required

The stakeholders also identified additional data points that could further enhance the predictive model and operational strategies:

- **Client Data**:
  - More detailed information on **household size** and **dietary restrictions** to better tailor food hamper distributions.
  - Historical patterns of missed appointments or rescheduled pickups to inform operational planning.
- **Operational Data**:
  - Real-time tracking of hamper stock levels to integrate **inventory management** into the prediction framework.
  - Seasonal variations in stock requirements to prepare for high-demand periods like Ramadan.

4. Next Steps

Based on the session discussions, the following actionable steps were outlined to ensure continuous improvement:

- **Data Collection**:

  Collaborate with operational teams to gather additional **client-level and operational data** that align with the suggested improvements.
- **Refinement of Models**:

  Update models with the newly collected data, including **household size**, **income levels**, and **dietary restrictions**, to improve prediction accuracy.

Incorporate real-time feedback to enhance the model's adaptability and relevance.

- **Stakeholder Collaboration**:

  Schedule **bi-weekly stakeholder sessions** to provide updates on progress, discuss challenges, and gather iterative feedback for ongoing improvements. Use shared dashboards to ensure transparency and encourage active collaboration among all teams.

This structured approach to stakeholder sessions ensured clear communication of findings, actionable feedback, and alignment of project goals with community needs, ultimately driving the project's success.

---

9. Conclusion

The project has significantly improved the food hamper distribution process by leveraging predictive modeling and innovative deployment strategies.

Current Status:

- **SARIMA Model Deployment**: The SARIMA model, selected for its ability to capture seasonal trends and achieve an R^2 value of **0.525**, has been deployed via Gradio and Streamlit. It is hosted on Hugging Face Spaces, providing stakeholders with a real-time, accessible platform for predictions.
- **Operational Access**: Stakeholders can now input the number of prediction days and receive reliable forecasts for food hamper pickups, enabling better planning and resource management.

Key Successes:

1. **Accurate Forecasting**:
   - Delivered reliable demand predictions, particularly during high-demand periods such as **Ramadan** and **Eid**.
   - Enhanced operational efficiency by enabling proactive resource allocation.
2. **Seamless Deployment**:
   - Provided a **user-friendly and accessible platform** that facilitates real-time interaction.
   - Integrated dashboards and visualizations for stakeholders to explore trends and predictions intuitively.

3. **Stakeholder Engagement**:
    - Incorporated feedback to refine model accuracy and align predictions with community-specific needs.
    - Enhanced collaboration through shared dashboards and regular alignment meetings.

Future Plans:

1. **Model Updates**:
    - Periodically retrain the SARIMA model with updated and more granular client data, such as **household size**, **dietary preferences**, and **income levels**, to enhance prediction accuracy.
2. **Enhanced Scalability**:
    - Expand deployment to serve additional regions and communities, adapting the platform to diverse operational needs.
3. **Continuous Monitoring**:
    - Implement a robust monitoring framework to ensure the sustained performance of the deployed application.
    - Conduct regular maintenance and updates to address any operational challenges promptly.

Final Remarks:
This project has established a strong foundation for **efficient, data-driven resource allocation**, paving the way for scalable solutions in food distribution. By combining advanced predictive models with user-centric deployment strategies, the initiative ensures timely and equitable access to essential resources for communities in need. The approach can serve as a model for other organizations aiming to enhance operational efficiency and community impact.

---

10. Attachments

The following documents, charts, and presentations accompany this report to provide a comprehensive understanding of the project, its data, and findings.

Documents

1. **Food Hamper Dataset**

    - Contains detailed records of food hamper pickups, client demographics, and scheduling information.

    - **File Name**: mergedfoodandclients.csv

mergedfoodandclient
s.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

Charts

1. **Gender Distribution of Clients**

    A pie chart illustrating the proportions of male, female, and unspecified clients.

2. **Monthly Food Hamper Pickup Trends**

    A bar chart displaying the steady increase in demand with seasonal peaks, highlighting key periods such as Ramadan and Eid.

---

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.

Literature Review References

- Smith, J., & Brown, P. (2020). *Applications of machine learning in resource allocation: A comprehensive review.* Journal of Artificial Intelligence Research, 45(2), 120–134.
- Jones, A., & Lee, H. (2019). *Seasonal forecasting using SARIMA models: An applied approach.* International Journal of Data Science, 12(4), 205–219.
- Ahmed, S., Khan, R., & Patel, M. (2021). *The role of cultural factors in predictive modeling for community-driven initiatives.* Community Development Journal, 58(1), 15–29.

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