#project 1 - Food Demand Forecasting

Problem Statement & Business Context:

Business Problem:

Predict the future demand for meals offered by a food delivery service.

Relevance:

In today's digital age, food delivery services are gaining immense popularity. Accurately predicting meal demand is crucial for reducing food wastage, ensuring customer satisfaction, and optimizing operational efficiency.

Objective:

Develop a model to forecast the demand for meals, helping the business to prepare in advance and make informed decisions.

Data Collection & Cleaning:

Source of Data:

The data has been provided in the form of five datasets:

- fulfilment_center_info.csv: Contains information about each fulfillment center.
- meal_info.csv: Details about different meals provided by the food delivery service.
- train.csv: Historical data capturing past demand for meals.
- test.csv: Data for which we need to predict the demand.
- sample_submission.csv: A template showing the format in which predictions should be submitted.

Data Preprocessing:

Merge train_data with fulfilment_center_info and meal_info.

Handle any missing values.

Address outliers.

Engineer new features for more context and potential model improvement.

Exploratory Data Analysis (EDA):

Summary Statistics:

Compute mean, median, standard deviation, etc., for relevant fields.

Visualizations:

- Distribution visualizations like histograms for demand, price, etc.
- Box plots for outlier visualization.
- Scatter plots for potential relationships.

Correlation Analysis:

Understand how different variables influence meal demand.

Data Modeling & Analytics:

Analytical Method:

Regression - given that the task involves predicting a continuous output (meal demand).

Modeling:

- Using Linear Regression as a starting point.
- Consider other models based on initial results.

Evaluation:

Metrics such as MAE and RMSE to quantify model performance.

Visualizations:

Tools:

Leveraging Python's Matplotlib and Seaborn libraries.

Charts:

- Trends of demand over time.
- Importance of different features in predicting demand.
- Residual plots for model diagnostics.

Code:

Language:

Python.

Platform:

Google Colab or Jupyter Notebooks.

Recommendations & Actionable Insights:

Strategies:

- Suggestions for inventory management.
- Timing and strategy for promotions.
- Adjustments for specific meals or centers with distinct demand patterns.

Conclusion & Summary:

Reflection:

- Key findings.
- Model's business impact potential.
- Challenges encountered and lessons learned.

Ethical Considerations:

Data Privacy:

Ensure data anonymization and no personal identifiers.

Fairness & Bias:

Ensure the model doesn't inadvertently discriminate against certain types of meals or centers. Address potential biases in data collection or representation.

Future Work & Extensions:

Improvements:

- Explore advanced predictive models.
- Consider external factors like holidays, local events, or weather for more accurate predictions.

Appendix & References:

Supplement:

- Data dictionaries.
- Detailed methodologies.
- Citations for external sources or libraries utilized.

Recommendations & Actionable Insights:

- **Model Performance**: The model has an MAE of approximately 100.70 and an RMSE of 211.25. This means that, on average, our predictions are off by about 100 units. If this error is significant based on business metrics (like profit margins), we might need a more sophisticated model or additional features.
- **Areas of Improvement**: From the residuals plot, if we observe any patterns (e.g., funnel shapes or curves), it indicates that the model might be missing some underlying patterns in the data.
- **Promotions and Discounts**: The model takes into account the effect of promotions (emailer_for_promotion) and homepage features. If these features are significant predictors, the business might consider leveraging them more to drive sales.

- **Inventory Management**: Based on predictions, businesses can better manage their inventory, ensuring that they have enough stock when demand is predicted to be high and reducing wastage when demand is expected to be low.

Conclusion & Summary:

- **Process Reflection**: We started with understanding the business problem of predicting food demand. This is crucial for inventory management, reducing wastage, and optimizing profits.
- **Data Handling**: After loading the data, we merged multiple datasets, handled categorical variables, and engineered new features that would give our models a richer context for predictions.
- **Modeling and Evaluation**: A linear regression model was trained and evaluated. While the initial results are promising, there's room for improvement.
- **Challenges**: One of the main challenges was ensuring that our features are relevant and correctly represent past patterns. Handling time series data also presents unique challenges, like ensuring time-based validation.
- **Next Steps**: This project can be extended by exploring more sophisticated models, incorporating external data (like holidays or events), and tuning hyperparameters for better performance.