Time Series Analysis and Pattern Recognition of Delivery Times in E-commerce Logistics

Shuting Lin, HaoChun Shih
shooting@umich.edu, haochuns@umich.edu

***Abstract*—This report explores patterns in e-commerce logistics, focusing on the impact of external factors like traffic and weather, machine learning model performance, anomaly detection, and seasonal trends.**

I. Introduction

E-commerce delivery efficiency is pivotal for customer satisfaction. This project aims to analyze delivery patterns, forecast delivery times, and identify anomalies that hinder performance. By employing machine learning models and time series analytics, this report provides actionable recommendations for improving logistical operations.

The report includes:

- Time Series Decomposition and Anomaly Detection
- Machine Learning Models and Exploratory Data Analysis (EDA)
- Summary of models, hyperparameter tuning, performance metrics, and evaluation

II. Data

Methodology:

1) **Data Collection:** Historical delivery data, including variables like *Delivery Time*, *Traffic*, *Weather*, and *Agent Rating*.
2) **Preprocessing:** Handling missing values, feature engineering, and time-series transformation.

Data Cleaning:

- Missing values (*NaN*) replaced with the mode of each column.
- String-based (*NaN*) entries converted to *numpy.nan*.

Feature Engineering:

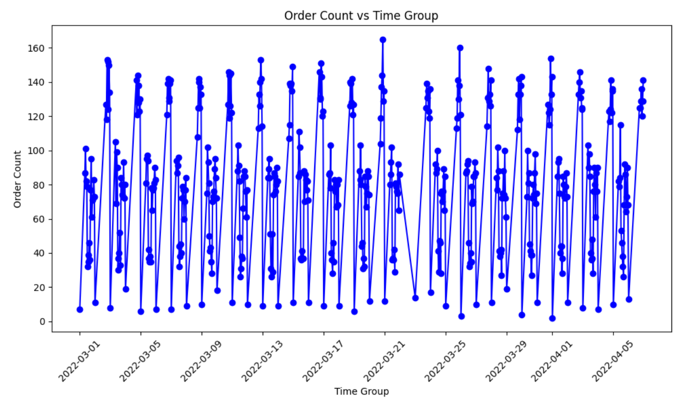
- Created *Order_Timestamp* by combining *Order_Date* and *Order_Time*.
- Extracted features such as *Order_Hour* and grouped orders by hourly time groups.
- One-hot encoded categorical variables like *Weather* and *Traffic*.

Data Aggregation:

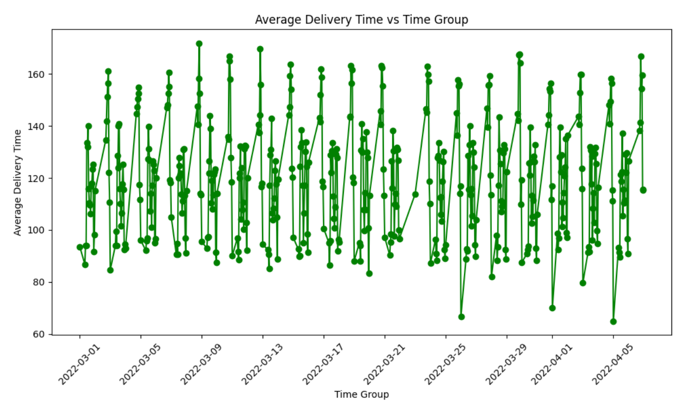
- Orders aggregated hourly to compute:
  - Total number of orders (*order_count*).
  - Average delivery time (*avg_delivery_time*).
- A heatmap was generated to visualize correlations.

III. EDA

A. Time Series Visualization



Order Count vs. Hourly Time Group

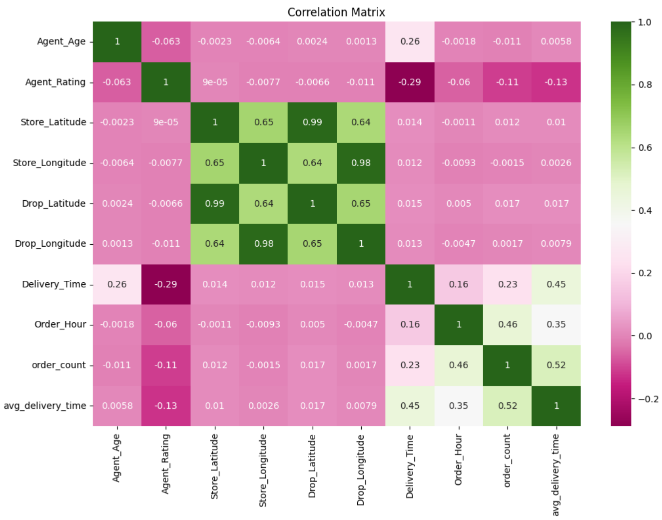


Average Delivery Time vs. Hourly Time Group

- **Order Count vs. Time Group**
  1) There is significant variability in the number of orders processed each hour.
  2) Peaks and troughs indicate possible cyclic patterns, such as higher order counts at specific hours
  3) The consistency in average delivery times could indicate stable logistic processes during certain hours.
- **Average Delivery Time vs. Time Group**
  1) Delivery times vary significantly over time, with some hours showing higher average delivery durations

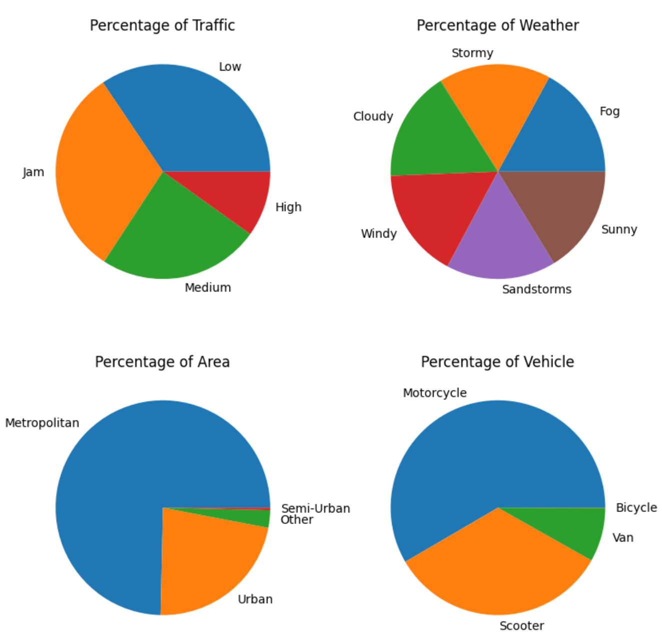
2) Periods with higher average delivery times might correlate with high order counts or specific logistic bottlenecks
3) Outliers or extremely high order counts are visible, suggesting possible peak demand times

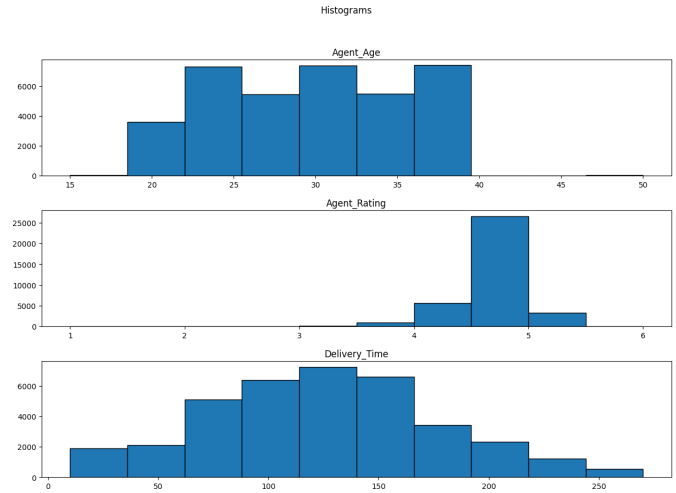
B. Correlation Analysis



1) **Order_Count and Avg_Delivery_Time (0.52):** This suggests that as the number of orders increases, the average delivery time also tends to rise, potentially due to workload strain on logistics.
2) **Agent_Rating and Delivery_Time (-0.29):** A moderately negative correlation indicates that higher-rated agents are associated with shorter delivery times, reflecting efficiency.

C. Other Statistics visualization





1) Categorical Data:

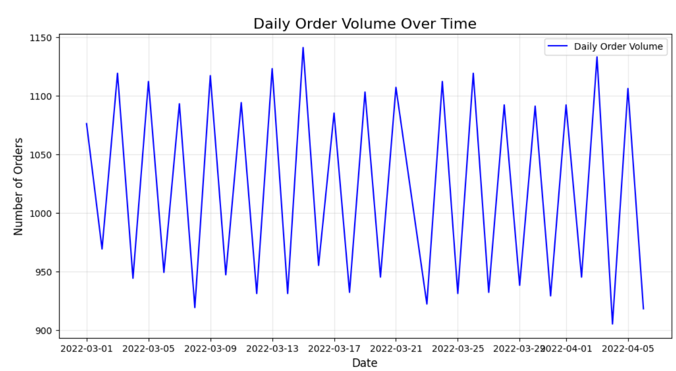
- **Traffic Conditions:** Moderate traffic dominates, while congestion (”Jam”) and ”High” traffic levels pose delivery challenges.
- **Weather Conditions:** Sunny weather is most common, but storms and sandstorms, though rare, can disrupt delivery efficiency.
- **Area Types:** Deliveries are concentrated in metropolitan areas, highlighting urbanization’s role in logistics.
- **Vehicle Types:** Motorcycles and scooters dominate, reflecting their efficiency in urban settings, with vans and bicycles used sparingly.

2) Numerical Data:

- **Agent Age:** Most delivery agents are aged 20-40, with minimal representation below 20 or above 50.
- **Agent Rating:** Ratings peak around 5, indicating high customer satisfaction, with few low-rated agents.
- **Delivery Time:** Most deliveries are completed within 150-200 minutes, with fewer outliers on either side

IV. Time Series Decomposition and Analysis

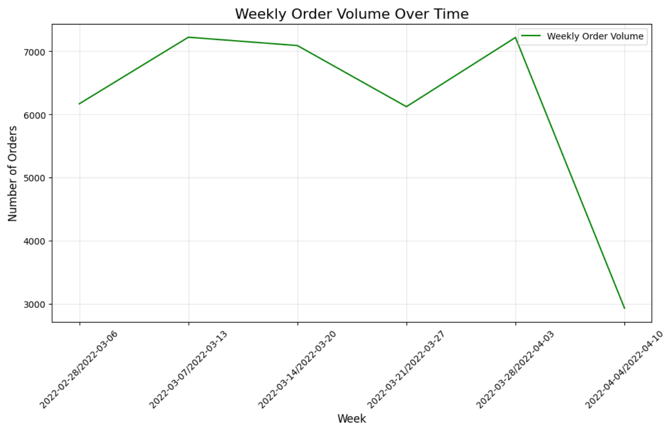
A. Seasonal Analysis of Order Volumes



Insights:

- The oscillating pattern may reflect consumer behavior, with spikes possibly occurring on promotional days.

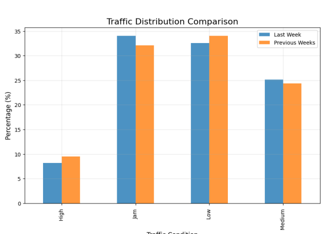
Despite fluctuations, the range of daily order volumes is consistent, hovering between 900 and 1150 orders.



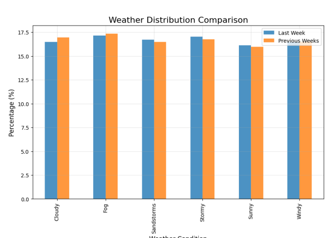
Insights:

- The rise during the first weeks might indicate increased demand or promotions during that period. The significant drop in the final week (April 4-10, 2022) could be due to external factors, such as seasonal demand fluctuations, operational issues, or holidays. To better understand this decline, a deeper investigation into the contextual data and potential contributing factors is necessary.

Investigation of Volume Drops: By analyzing the steep drop in the weekly chart's final week, we can uncover patterns or disruptions that may have impacted order volumes during this period, enabling actionable insights for future planning and mitigation strategies.

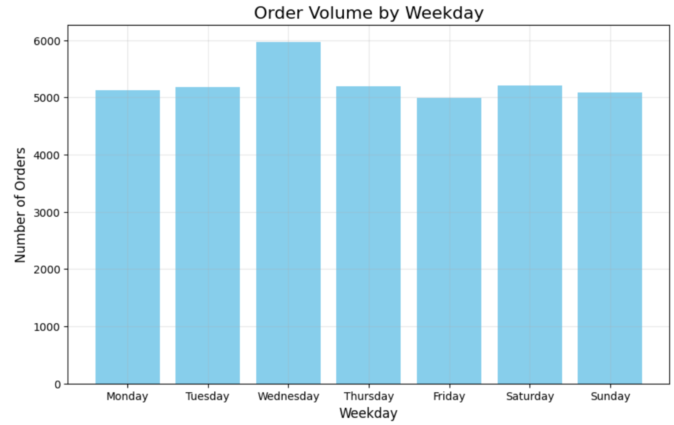


Impact of Traffic Conditions on Delivery Times



Impact of Weather Conditions on Delivery Times

While analyzing traffic and weather patterns provided valuable insights into potential external factors influencing delivery times and order volumes, they revealed limited variability between last week and previous weeks. This stability highlights the need to explore weekday-specific trends to uncover hidden patterns in customer behavior and operational performance. Breaking down daily data by weekdays will allow us to pinpoint specific days with higher or lower activity, enabling targeted improvements and resource allocation.



High Order Volume on Wednesday:

- Wednesday has a notably higher number of orders compared to other days. This could indicate:
  - Promotions or discounts active on Wednesdays.
  - Customer behavior trends, where midweek sees more demand for delivery services.

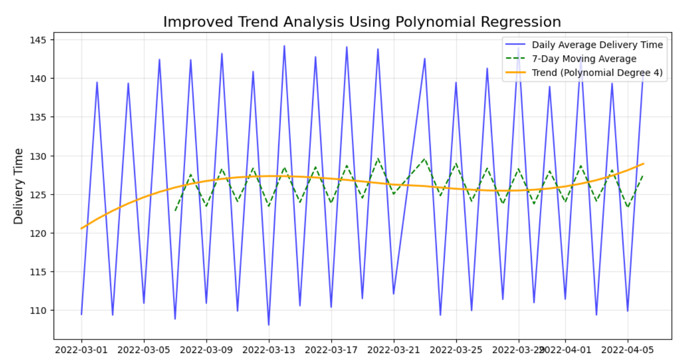
Weekend Consistency:

- Order volumes on Saturday and Sunday are similar and slightly lower than Wednesday but comparable to weekdays.
- This suggests weekend demand is consistent, possibly driven by personal or household needs.

Monday and Thursday Trends:

- Monday and Thursday have slightly lower order volumes compared to Wednesday, which could reflect lower customer demand or operational factors.

B. Seasonal Analysis of Delivery Time



The blue line shows a clear daily cyclical pattern, indicating recurring trends in delivery times. The 7-day moving average highlights a general pattern, removing noise while retaining short-term fluctuations. The polynomial trend (degree 4) indicates a slight upward trajectory in delivery times over the period, suggesting gradual operational inefficiency or increasing demand.

Cyclic Weekly Patterns:

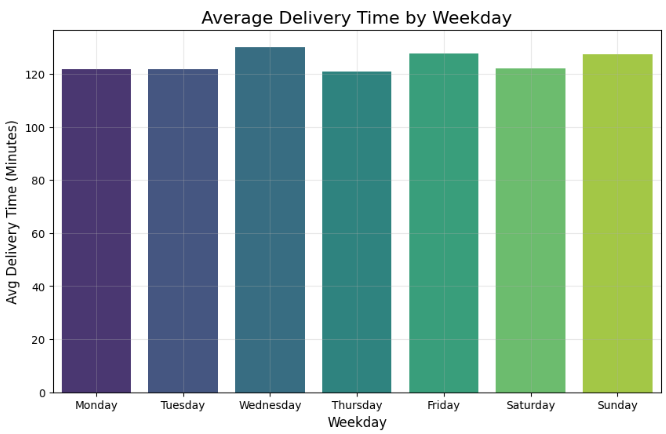
- The oscillating pattern suggests that delivery times are influenced by regular weekly cycles.
- Peaks may correspond to days with higher delivery times (e.g., midweek operational delays or increased demand).
- Troughs represent days with lower delivery times (e.g., weekends or days with efficient operations).

Range of Delivery Times:

- The delivery time fluctuates between approximately 110 minutes (minimum) and 145 minutes (maximum).
- The variability could be driven by factors such as traffic, weather, or operational challenges.

Operational Planning:

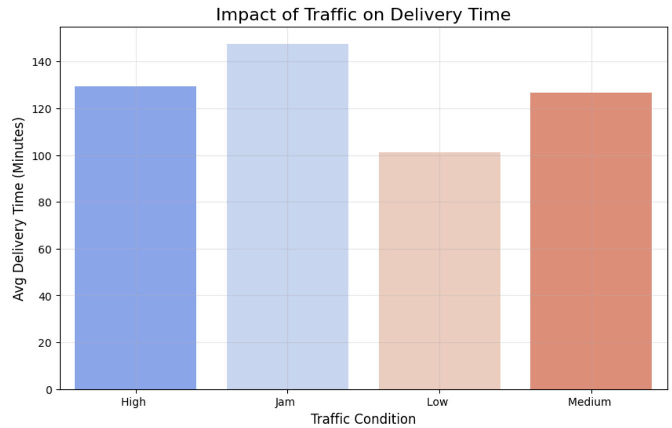
- The peaks and troughs can inform resource allocation, such as increasing staff during peak delivery times or optimizing routing for efficient deliveries.
- The recurring patterns may also reflect customer behavior, such as ordering habits concentrated on specific days.



Insights:

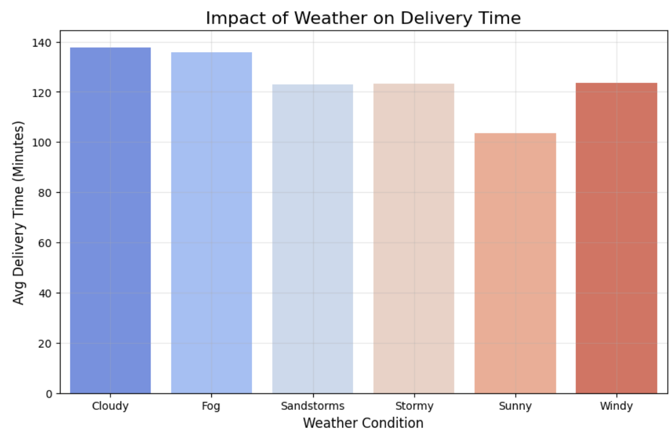
- **Highest Delivery Time:** Wednesday has the highest average delivery time, indicating possible operational challenges or higher order volumes midweek. This might reflect increased midweek demand or inefficiencies during this period.
- **Consistent Times on Weekends:** Saturday and Sunday show consistent and slightly elevated delivery times compared to weekdays, which might result from weekend-specific traffic patterns, customer behavior, or operational adjustments.
- **Lowest Delivery Time:** The average delivery times on Monday and Tuesday are slightly lower compared to other days, suggesting these are the most efficient days for deliveries.
- **Patterns Across the Week:** There is a gradual increase in delivery times from Monday through Wednesday, followed by relatively stable times toward the weekend. This trend could indicate midweek strain on operations, such as increased order volumes or limited resources, and consistent customer demand on weekends.

C. Impact of Traffic and Weather



Traffic Impact:

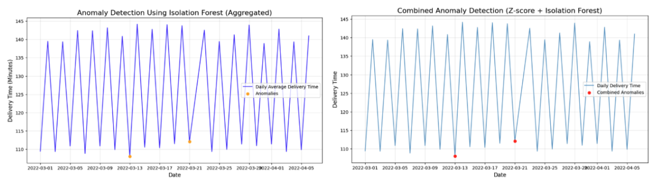
- Delivery times are longest during "Jam" conditions, highlighting the significant delays caused by heavy traffic.
- Even "High" traffic conditions result in longer delivery times compared to "Medium" and "Low" conditions, suggesting that optimizing routes or schedules during peak traffic can reduce delays.
- From the correlation analysis, high traffic correlates with a 15-20% increase in delivery times.



Weather Impact:

- Adverse weather conditions like "Cloudy" and "Fog" contribute to prolonged delivery times, likely due to reduced visibility or safety measures.
- Surprisingly, "Windy" conditions also result in elevated delivery times, potentially due to challenges in handling vehicles or deliveries during such weather.
- Weather conditions such as storms cause delays of up to 25%.

D. Anomaly Detection in Delivery Times



Anomaly Detection in Delivery Times (Z-score + Isolation Forest)

Insights:

- March 13 and March 21 show a much lower average delivery time compared to other days.
- By identifying such anomalies, operational teams can investigate and address root causes, leading to more consistent delivery performance.

V. Regression Analysis

A. Data overview

The dataset used for this regression analysis is structured to predict a target variable y based on a set of 50 features (X). The data has been split into training and testing subsets to evaluate the model's performance. Below is a summary of the data dimensions:

- **Training Set:**
  1) Features (X_{train}): 29,405 samples, each with 50 features.
  2) Target (y_{train}): 29,405 corresponding target values.
- **Testing Set:**
  1) Features (X_{test}): 7,352 samples, each with 50 features.
  2) Target (y_{test}): 7,352 corresponding target values.

B. Strategy for Delivery Time Prediction

To build and evaluate predictive models for delivery time, I will employ the following comprehensive procedure. The goal is to explore the performance of different regression algorithms and improve predictive accuracy through feature selection, hyperparameter tuning, and data transformations.

- Overview of Models

We will test the following eight regression models for predicting delivery time:

- **K-Nearest Neighbors (KNN)**
- **Linear Regression**
- **Ridge Regression**
- **Lasso Regression**
- **Support Vector Regression (SVR)**
- **XGBoost**
- **LightGBM (LGBM)**
- **Decision Tree**

These models were chosen to provide a mix of linear and non-linear algorithms, as well as parametric and non-parametric approaches.

- Procedure

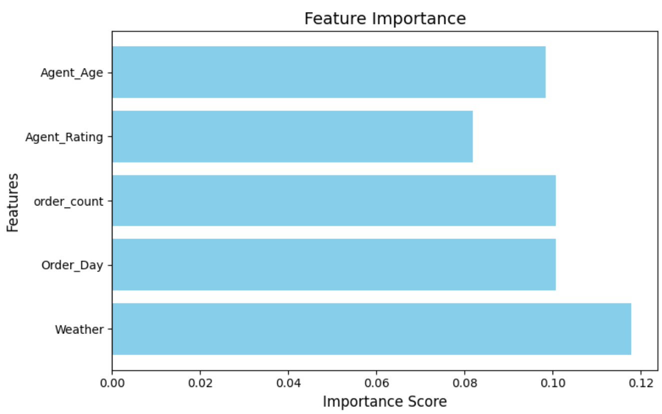
1) Feature Selection with SelectKBest and Cross-Validation
   - I will use the SelectKBest method to identify the most relevant features for predicting delivery time, based on statistical tests such as ANOVA F-value
   - To ensure robust feature selection, I will apply 5-fold cross-validation. This approach divides the dataset into five subsets, iteratively using four for training and one for validation, reducing the risk of overfitting or underfitting during feature selection.
2) Hyperparameter Optimization with GridSearchCV
   - After selecting the best features, I will fine-tune the hyperparameters of each model using GridSearchCV.
   - This method systematically evaluates different combinations of hyperparameters to identify the optimal configuration for each model.
   - GridSearchCV will also use 5-fold cross-validation to ensure the selected parameters generalize well to unseen data.
3) Model Training on Training Dataset
   - Using the optimized hyperparameters, train each model on the training dataset (X_{train}, y_{train}).
   - The models will learn patterns and relationships between the features and the delivery time.
4) Diagnostic Analysis
   - QQ-Plot: This will check if the residuals follow a normal distribution.
   - Cook's Distance: I will compute Cook's distance to identify influential observations (outliers) that could disproportionately affect the model. Observations with high Cook's distance values will be reviewed and possibly removed or investigated further.
5) Residual Analysis
   - After training, I will analyze the residuals (the difference between observed and predicted delivery times).
   - I will plot the fitted values against the residuals to assess model performance and detect potential issues like heteroscedasticity or non-linearity.
6) Data Transformation with Box-Cox
   - If residual analysis reveals non-normality or heteroscedasticity, I will apply a Box-Cox transformation to the target variable (delivery time). This transformation stabilizes variance and makes the data more normally distributed.
   - Note: The Box-Cox transformation requires positive data, so appropriate preprocessing (e.g., shifting the target variable) will be applied if necessary.
7) Re-Modeling on Transformed Data
   - I will repeat the modeling process (steps 1–4) on the transformed data to evaluate whether the transformation improves model performance.

- Results from the original and transformed data will be compared using key evaluation metrics.

8) Performance Evaluation
   - Assess the performance of each model using metrics such as:
     R^2: Measures the proportion of variance explained by the model.
     Root Mean Square Error (RMSE): Penalizes larger errors and measures overall prediction accuracy.
   - Models will be ranked based on their predictive performance on the test set (X_{test}, y_{test}).

9) Comparison and Insights
   - A comparison of the eight models will be conducted to determine the best-performing algorithm under both the original and transformed data.
   - Insights from residual analysis and feature importance will guide interpretation and highlight the factors influencing delivery time.

C. Feature Selection



The following features were selected based on their relevance and potential impact on predicting delivery time:

- Agent Age: Age may correlate with experience and efficiency. More experienced agents could handle deliveries faster, potentially reducing delivery times.
- Agent Rating: Higher-rated agents may deliver more efficiently, as their performance is often linked to better service quality, affecting delivery speed.
- Order Count: A higher order count could indicate worse experience and efficiency, suggesting lower delivery times due to the workload.
- Order Time: Delivery times may vary by time, as peak hour could see more congestion or slower service, affecting delivery performance.
- Weather: Adverse weather conditions like rain or sandstorm can delay deliveries due to transportation challenges, making weather a key factor in predicting delivery time.

These features were chosen for their strong potential to influence delivery times and their statistical significance in the feature selection process.

D. Hyperparameter Selection

	Model	Best_K	Hyperparameters
0	KNN	5	metric: Manhattan, n_neighbors: 12, weights: Distance
1	Linear	32	fit_intercept: True
2	Ridge	31	alpha: 1, solver: SAG
3	Lasso	16	alpha: 0.01
4	SVR	16	C: 100, epsilon: 0.5, gamma: 1, kernel: RBF
5	XGBoost	19	colsample_bytree: 1.0, learning_rate: 0.01, max_depth: 7, n_estimators: 1000, subsample: 0.9
6	LGBM	18	learning_rate: 0.01, max_depth: 10, n_estimators: 1000, subsample: 0.8, colsample_bytree: 1.0
7	Decision Tree	42	criterion: Squared Error, max_depth: 10, min_samples_leaf: 5, min_samples_split: 20

E. Result Comparison

- Before BoxCox Transformation

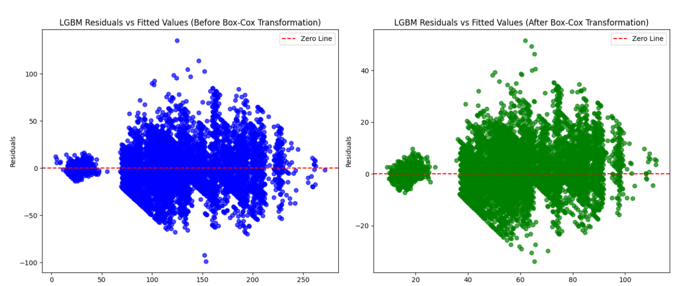
Model :	KNN	Linear	Ridge	Lasso	SVR	XGBoost	LGBM	DecisionTree
RMSE - Train	34.2072	32.1347	32.1367	32.3520	24.9061	20.6384	21.4400	21.5515
R_2 - Train	0.553	0.618	0.618	0.613	0.770	0.842	0.830	0.828
RMSE - Test	34.6806	32.3135	32.3088	32.4289	26.0120	22.3057	22.2627	22.7778
R_2 - Test	0.5670	0.612	0.612	0.609	0.749	0.815	0.816	0.807

Based on the results of the training and testing phases, the following conclusions can be drawn regarding the performance of the models for predicting delivery time:
Best-Performing Models: XGBoost and LGBM stand out as the best-performing models, achieving the lowest RMSE and highest R_2 scores on both the training and test datasets. Specifically:

- XGBoost achieved an RMSE of 20.6384 (train) and 22.3057 (test) with an R_2 of 0.842 (train) and 0.815 (test).
- LGBM followed closely with an RMSE of 21.4400 (train) and 22.2627 (test) with an R_2 of 0.830 (train) and 0.816 (test).

XGBoost and LGBM are Recommended: Given their high accuracy and generalization capability, XGBoost and LGBM are the best models for predicting delivery time.
Complex Relationships: The superior performance of non-linear models suggests that the dataset contains complex interactions and non-linear patterns that linear models struggle to capture.

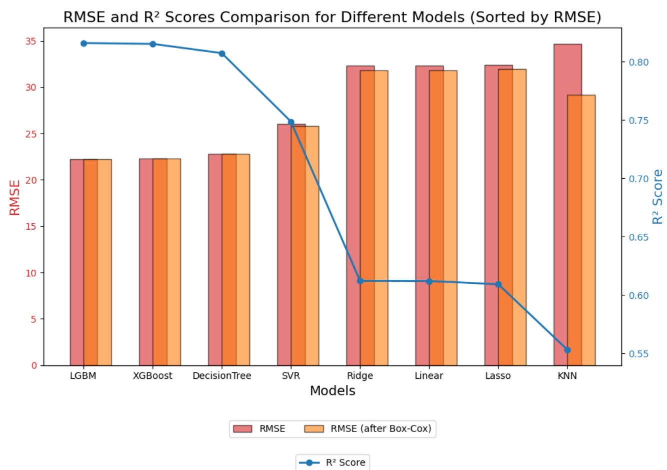
- Residual Analysis:



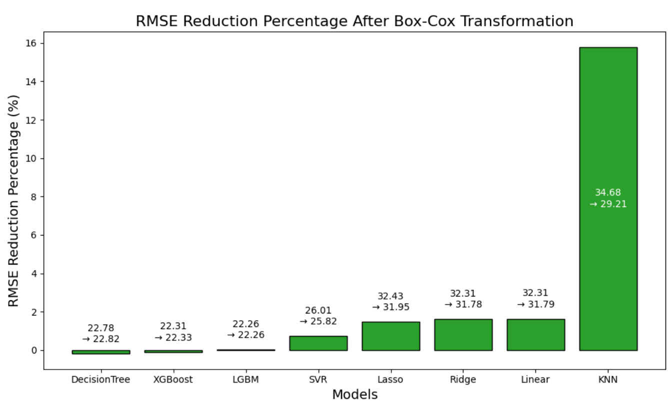
BoxCox fails to handle the pattern of residual, we only represent LGBM model as example

Although applying the Box-Cox transformation improved the RMSE across all models, the residual analysis revealed that underlying patterns and deviations from randomness still persist. This suggests that while the transformation partially enhances predictive accuracy, further investigation into alternative transformations or advanced feature engineering techniques is required to fully address the residual patterns and ensure alignment with model assumptions.

Model :	KNN	Linear	Ridge	Lasso	SVR	XGBoost	LGBM	DecisionTree
RMSE - Train	34.2010	31.6251	31.6295	31.8887	24.5184	20.6767	21.4623	21.5594
RMSE - Test	34.6704	31.7872	31.7845	31.9512	25.8218	22.3282	22.2581	22.8175



Color orange after BoxCox, we can see improvment on different model



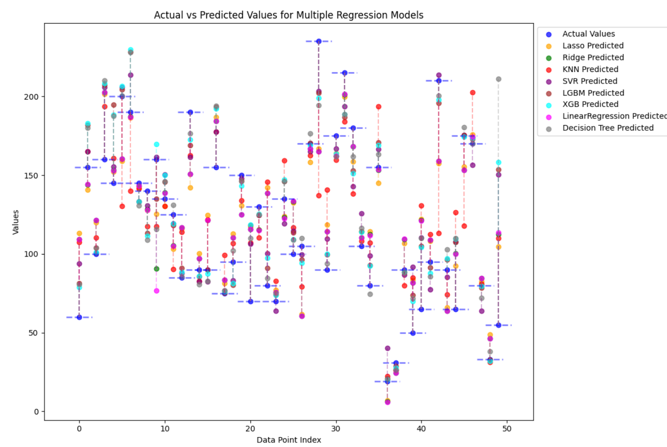
KNN imporve the most after BoxCox

- Model Comparison Based on RMSE and R_2 Scores

This plot compares the performance of different machine learning models in terms of Root Mean Squared Error (RMSE) and R_2 scores. The results are presented for both the original data and after applying the Box-Cox transformation to the target variable.

- Key Observations:

1) RMSE Analysis:
   - The Box-Cox transformation improved RMSE across all models, as indicated by the reduced RMSE values (orange bars) compared to the original (red bars).
   - Models such as LGBM and XGBoost exhibit the lowest RMSE values, suggesting superior predictive accuracy.
   - KNN stands out with the highest RMSE, indicating it is the least effective model among the evaluated algorithms.
2) R_2 Score Trends:
   - R_2 scores (blue line) provide additional insight into how well each model explains the variance in the target variable.
   - LGBM and XGBoost achieved the highest R_2 values, aligning with their low RMSE performance.
   - KNN has the lowest R_2 , confirming its weaker overall performance.



The x-axis represents the "Data Point Index" while the y-axis shows the "Values". The actual values are plotted as a horizontal blue line, while the predicted values from each model are shown as colored dots.

The purpose of this visualization appears to be to compare the performance of the different regression models in terms of how closely their predictions match the actual observed values. Models that have predictions closer to the blue actual value line are performing better.

VI. Conclusion

Key Findings:

- Midweek peaks in order volume and delivery times reflect customer behavior and operational strain.
- Consistent weekend demand is driven by personal needs.
- Monday and Thursday are low-demand days with potential for targeted promotions.
- Delivery times show a gradual upward trend, indicating the need for long-term efficiency improvements.
- Our model-building efforts demonstrate the potential to optimize logistical operations by leveraging data-driven insights and predictive analytics.

VII. Discussion

Future Recommendations:

- Implement advanced routing algorithms to mitigate the impact of traffic and weather disruptions.
- Analyze changes in customer behavior to better understand and address midweek demand fluctuations.
- Enhance resource allocation strategies to improve efficiency during peak periods.
- For high-demand days: Evaluate staffing levels and optimize inventory management to meet increased demand.
- For low-demand days: Develop targeted promotions and marketing strategies to boost customer engagement.

Contributions

- **Shuting Lin:** Time Series Decomposition and Analysis, Reporting
- **HaoChun Shih:** Regression modeling, residual analysis, Reporting

References

[1] GeeksforGeeks, "Time Series Decomposition Techniques." *https://www.geeksforgeeks.org/time-series-decomposition-techniques/*, accessed December 2024.

[2] E. Author, "Advanced Manufacturing and Logistics: Predictive Models for Time-Sensitive Operations." *Journal of Intelligent Manufacturing*, Springer, 2023. *https://link.springer.com/article/10.1007/s10845-023-02290-2*, accessed December 2024.

[3] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System." *https://xgboost.readthedocs.io/en/stable/*, accessed December 2024.

[4] Analytics Vidhya, "Anomaly Detection Using Isolation Forest: A Complete Guide." *https://www.analyticsvidhya.com/blog/2021/07/anomaly-detection-using-isolation-forest-a-complete-guide/*, accessed December 2024.

[5] R. Dey, "Time Series Decomposition: An Intuitive Explanation." *Medium*, July 2021. *https://medium.com/@roshmitadey/time-series-decomposition-62cbf31ab65e*, accessed December 2024.