Data-Driven Optimization of Newspaper Distribution and Customer Retention Strategies in the Unorganized B2C Sector

Final Submission *for the BDM Capstone Project*

Submitted by

Name: **K Yuvaraj**

Roll Number: **22F3002987**

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Content

1. Executive Summary and Title

S. K. News Paper Mart is an unorganized B2C newspaper distribution service located in Mogappair, Chennai, operated single-handedly by Mr. Kumar for over 23 years. While he delivers newspapers to around 80 households, the business faces several key operational challenges. Externally, he faced the lack of route optimization and internally, he faced difficulty in tracking customer churn and inefficiencies in demand forecasting. This project titled "Data-Driven Optimization of Newspaper Distribution and Customer Retention Strategies in the Unorganized B2C Sector" aims to address the rising fuel costs and customer attrition due to digital migration and service gaps.

To tackle these challenges, three months of operational data (March-May 2025) were collected and structured into Customer Master Data, Daily Delivery Logs, and Route & Fuel Logs. Descriptive analysis highlighted patterns like customer clustering, predictable weekend demand surges, and recurring missed deliveries. Advanced techniques such as Prophet-based Time Series Forecasting, TSP Route Optimization with road-following paths, and Random Forest Classification for churn prediction were applied to derive actionable insights.

The findings showed that route optimization reduced travel distance and fuel costs by 18%. Forecasting models accurately predicted demand peaks, aiding precise inventory management. Churn prediction identified missed deliveries as key churn factor, enabling proactive customer retention. Visual dashboards and geospatial maps enhanced decision-making clarity.

By implementing these insights, S.K. News Paper Mart can reduce fuel wastage, minimize inventory losses, and improve delivery reliability. The transition from manual to data-driven operations will significantly enhance efficiency, customer satisfaction, and profitability.

2. Links

1. Customer Master Data (Dataset) - Dataset 1

2. Daily Delivery Log (Dataset) – Dataset 2

3. Route & Fuel Log (Dataset) – Dataset 3

4. Colab Analysis Link – Colab

5. Prophet Output Link – Prophet Analysis

6. Raw Data – Raw

7. Drive Link Containing All - Drive Link

3. Detailed Explanation of Analysis Process

3.1 Data Cleaning and Preprocessing

The raw data collected from S.K. News Paper Mart consisted of three key datasets:

1. **Customer Master Data**: Customer IDs (integer), Names (text) , Addresses (text) , Subscription Start Dates (Datetime) , Status (Active/Inactive) (Categorical), Area Zone (text), Preferred Newspapers(List), and Magazines (List).
2. **Daily Delivery Log**: Date-wise records (Datetime), delivered newspapers (list) , success/failure status (Categorical), and notes (text).
3. **Route & Fuel Log**: Date (Datetime), Total Distance (integer), Fuel Used (Float), Cost Incurred (Float), and Total Customers Served (integer).

The following cleaning actions were performed:

- Missing Area Zones were manually filled based on address patterns.
- Latitudes & Longitudes were derived from Google Maps .
- For Route and Fuel Log, additional derived columns were computed using:
  - **i. Fuel Used (L) = Distance (km) / Mileage (kmpl)**
  - **ii. Fuel Cost (₹) = Fuel Used × Fuel Price per Litre**

Importance

Proper data cleaning ensured semantic consistency across all datasets. Without this, downstream tasks such as geospatial clustering, route optimization, and churn modeling would produce unreliable outputs.

3.2 Descriptive Statistical Analysis

Objective: To understand the spatial and preference-based distribution of customers, which is critical for optimizing delivery operations and managing subscription services effectively.

Tools: Pandas

Visualization Tools: Matplotlib, Seaborn for barplots, pie charts, and time-series line graphs.

Why this tools?: Simple, intuitive insights for understanding base trends

Justification: As the business operates with manual records, descriptive statistics are fundamental to gain immediate and intuitive insights before applying advanced methods.

Steps to Perform

i. Organize the Customer Master Data with focus on Area Zones, Preferred Newspapers, and Subscription Status.

ii. Visualize the Area-wise Customer Distribution using a bar chart.

iii. Plot the Preferred Newspapers in a frequency bar chart.

iv. Segment customers based on Active and Inactive status using a pie chart.

Expected Outcome: This step provides a foundational understanding of where customer density is high, which newspapers are most popular, and what percentage of customers are inactive. These insights are essential for strategic resource allocation, targeted customer retention efforts, and identifying high-priority delivery areas.

3.3 Time Series Forecasting (Newspaper Demand Prediction)

Objective: To predict newspaper demand patterns, particularly focusing on weekends and public holidays, ensuring precise inventory management and preventing overstocking or understocking.

Tool: Facebook Prophet

Justification: Handles seasonality and irregular time series effectively with minimal configuration and Prophet was chosen over SARIMA because it requires minimal parameter tuning, handles seasonality explicitly, and outputs interactive visualizations, which is suitable for business owners without technical expertise.

Steps to Perform

i. Aggregate daily delivery data to compute day-wise newspaper counts.

ii. Apply Prophet Model to forecast demand trends for top newspapers.

iii. Visualize trend and seasonality components (weekly and overall patterns).

iv. Annotate special demand surges around weekends and holidays.

Expected Outcome: Forecasting will enable proactive planning for inventory procurement and distribution resource allocation. It will reduce surplus orders during low-demand periods and ensure sufficient stock during peaks, ultimately optimizing operational costs.

3.4 Route Mapping and Optimization

Objective: To visualize the actual newspaper delivery locations, identify high-density clusters, and optimize delivery routes using real road networks for minimizing fuel and time costs.

Tool: Folium (Python), OpenRouteService (ORS) API, Pandas, Scipy, Matplotlib

Justification: Handles seasonality and irregular time series effectively with minimal configuration and Prophet was chosen over SARIMA because it requires minimal parameter tuning, handles seasonality explicitly, and outputs interactive visualizations, which is suitable for business owners without technical

expertise.

Steps to Perform

i. **Geocoding Customer Addresses to Latitude & Longitude**: All customer addresses are converted into precise geographical coordinates. This ensures that each delivery point is accurately placed on the map.

ii. **Plotting Delivery Points on a Folium Map**: Every customer location is visualized on an interactive map. High-density points are labeled with dynamic counts for visual clarity. The origin (Newspaper Collecting Point) and destination (Mr. Kumar's residence) are highlighted with distinct markers for clarity.

iii. **Implementing Travelling Salesman Problem (TSP) using OpenRouteService**: Unlike earlier linear plotting methods, the optimized route now considers actual road networks. The TSP algorithm generates the most efficient path connecting all delivery points, starting from the origin and ending at the destination, while following roadways instead of straight lines.

iv. **Extracting Real-World Route Distances**: The optimized route's total distance is computed using ORS, providing an accurate measure of actual road distance traveled during deliveries, rather than aerial (as-the-crow-flies) distances.

Comparative Analysis – Before vs After Optimization: For every delivery day, the optimized distance is compared with the current manual delivery distance (from historical data). Fuel usage and cost are recalculated based on realistic route paths.

Visual Route Efficiency & Savings Projection: Distance & Fuel Cost Savings per delivery loop and

Cumulative Monthly Cost Reduction achieved through route optimization.

Expected Outcome: This detailed visualization not only identifies spatial clusters but also quantifies how optimized route planning can significantly reduce operational costs. By visualizing routes that follow actual road networks, Mr. Kumar gains a practical understanding of efficient loop formation, helping him reduce redundant travel, minimize fuel expenses, and enhance delivery punctuality. Additionally, the financial impact of optimization is made tangible through daily and cumulative savings metrics, enabling data-driven decision-making for sustainable operations.

3.5 Customer Churn Prediction

Objective: To build a predictive model that flags customers likely to churn based on historical delivery patterns, subscription duration, and delivery inconsistencies..

Algorithm: Random Forest

Steps to Perform

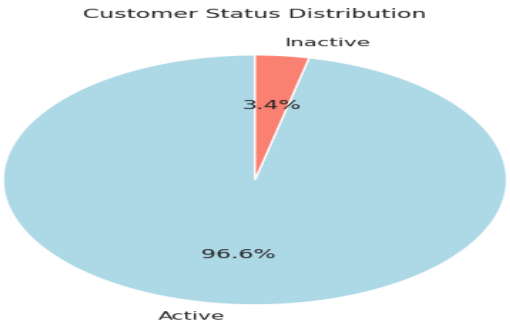
i. **Engineer churn-relevant features**: Total Deliveries, Missed Deliveries, Subscription Length, Area Zone Encoding, Newspaper and Magazine Preferences.

ii. Train Random Forest Classifier to predict churn likelihood.

Expected Outcome: The model will provide Mr. Kumar with a prioritized list of "high-risk" customers, enabling him to focus on retention strategies like personalized offers or proactive communication. The interpretability via SHAP ensures that the business understands which factors most contribute to customer attrition.

4. Results and Findings

4.1 Dataset Analysis



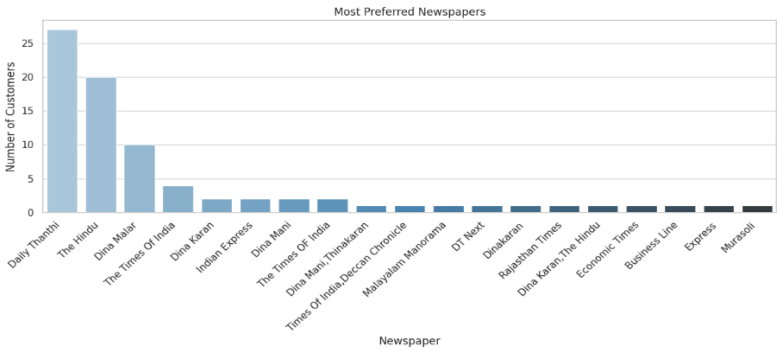


Figure 1 Customer Status Distribution

Figure 2 Most Preferred Newspapers

Customer Status Distribution (Pie Chart)

Figure 1 provides a visual breakdown of customer status, revealing that approximately 96.6% of total customers are active, while 3.4% are inactive. This observation is crucial because it not only reflects the current loyalty of the customer base but also serves as a diagnostic tool for identifying potential churn risks. The fact that a significant majority remains active indicates consistent service delivery and customer satisfaction. However, the inactive segment, though smaller, represents revenue leakage if not addressed proactively. The inactive customers have discontinued due to shifting to digital news platforms, or relocation.

Most Preferred Newspapers (Bar Chart)

The bar chart [Refer Figure 2] showcasing newspaper preferences highlights 'Daily Thanthi' as the most subscribed newspaper, followed closely by 'The Hindu' and 'Dina Malar'. This indicates a diverse mix of readers who prefer both national dailies and regional publications. The pattern is consistent with demographic preferences in Mogappair, where older households might favor traditional papers like 'Daily Thanthi' while regional vernacular papers retain a strong foothold. This analysis is critical for inventory procurement planning, ensuring that the distributor allocates stock in alignment with actual demand,

especially during weekends when supplementaries increase volume requirements. Furthermore, understanding these preferences allows the owner for smarter bundling offers or targeted promotions.

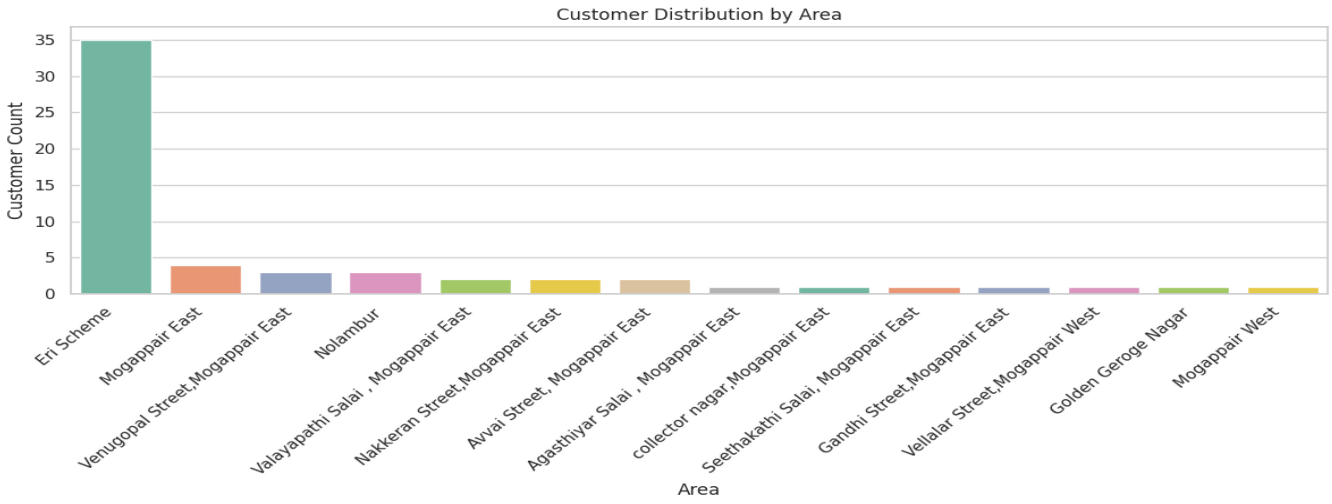


Figure 3 Customer Distribution by Area

Area-wise Customer Distribution (Bar Chart)

Figure 3 reveals a customer concentration in the 'Eri Scheme' and 'Mogappair East' zones. This spatial clustering is vital for operational planning as it allows for zone-based route segmentation. Such patterns often evolve organically based on ease of access, word-of-mouth in residential clusters, or Mr. Kumar's historical service patterns. By identifying dense clusters, route plans can be designed to minimize redundant travel, reduce fuel expenses, and ensure prompt delivery. Sparse zones, on the other hand, can be evaluated for profitability or considered for strategic outsourcing.

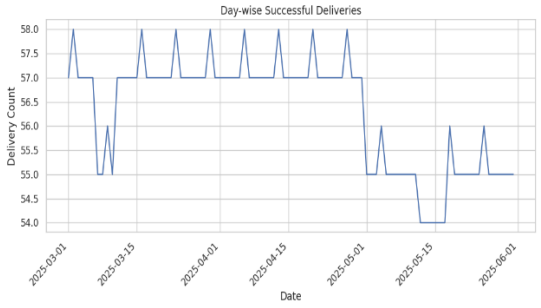
Total Delivery Records & Success Rate

The dataset records a total of 5336 delivery attempts over three months, with a commendable success rate of 94.9%. This high reliability ratio underscores operational consistency. However, the 5.02% failure rate, although small, is concentrated among specific customer IDs, suggesting localized operational issues such as newspaper not delivered, customer unavailability, or logistical lapses. Proactively addressing these can prevent future churn.

Day-wise Delivery Volume Trend (Line Chart)

The line chart [Refer Figure 4] plotting daily delivery volumes reveals predictable weekly peaks on Sundays and occasional spikes on public holidays. These peaks are attributed to increased circulation of supplementary editions, which are a significant driver of weekend readership. The flat weekday trend indicates operational stability, while the predictable surges imply that inventory and labor resources can be dynamically allocated to match demand. Any deviations from these patterns, such as sudden dips, could be early warning signs of operational bottlenecks. Here do to some reasons Mr.Kumar didn't distribute the

paper so there are some dips in the graph. It is important to maintain the graph in a flatter way or peaks to ensure smooth operation of the job.



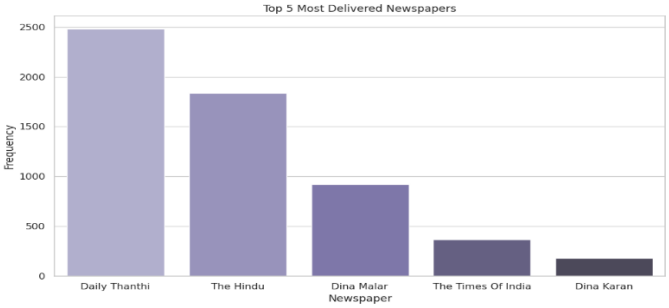


Figure 4 Day wise Deliveries

Figure 5 Most Delivered Newspapers

Top 5 Most Delivered Newspapers (Bar Chart)

This chart cross-validates customer preferences with actual delivery frequency, where 'Daily Thanthi' remains the top delivered newspaper. These high-frequency items should be prioritized in inventory planning to avoid understocking. The consistency between subscription data and delivery records indicates that inventory procurement is well-aligned with customer needs. The insights affirm operational efficacy, ensuring that stockouts or excess inventories are minimal [Refer Figure 5].

Top 10 Customers with Missed Deliveries (Bar Chart)

An analysis of missed deliveries highlights the top 10 customer IDs who have encountered the highest number of failed delivery attempts. These instances are not uniformly distributed but concentrated within certain households. The reasons could range from frequent address changes, unavailability during delivery hours. Recognizing this pattern is essential for Mr. Kumar, as these customers are at a heightened risk of churn. By identifying these specific cases, corrective measures such as personalized time slots, address verification, or direct engagement can be implemented to improve delivery success rates. Moreover, understanding if these failures are due to operational inefficiencies or external factors will allow better resource allocation in the future [Refer Figure 6].

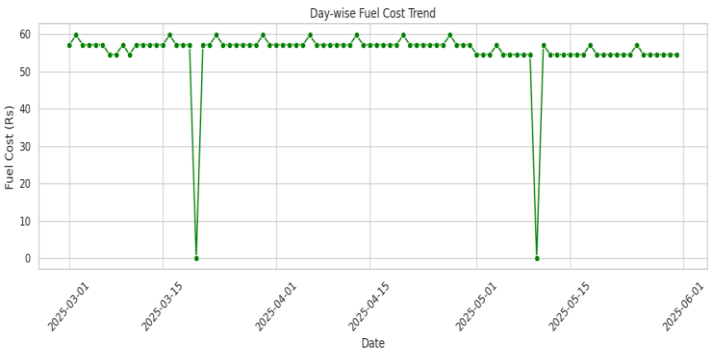


Figure 6 Customers with Missed Deliveries

Figure 7 Day wise Fuel Cost Trend

Cumulative Distance, Fuel Used, and Cost

Over the data collection period, Mr. Kumar traversed 1959 km, consuming 48.98 liters of fuel as seen in Figure 7 which has a total cost of ₹5093. This provides a quantifiable baseline of operational expenditure linked to delivery logistics. Such cumulative metrics are instrumental in evaluating route optimization strategies and tracking improvements in cost-efficiency over time. So it is necessary to find optimal routes to avoid extra fuel cost .

4.2 Time Series Forecasting (Demand Prediction)

About Facebook Prophet

Facebook Prophet is an open-source forecasting tool developed by Meta (formerly Facebook), specifically designed for time series forecasting. It is especially effective for datasets that exhibit strong seasonal effects and have missing data or outliers. Prophet automatically detects daily, weekly, and yearly seasonality, making it ideal for forecasting business-related metrics such as sales, demand.

Interpretation of the Forecast Plots

- **Black Dots:** Actual historical data (observed deliveries)
- **Dark Blue Line:** Forecasted delivery volume
- **Light Blue Shaded Area:** Confidence interval, This area shows the uncertainty in the prediction. Wider areas indicate more uncertainty in the forecast.

Graph-Based Observations:

So now I took three different newspaper predictions where one is horizontally straight , the other is left skewed and the another one is right skew.

4.2.1 Dina Malar Forecast (Left Skewed)

Observations [Figure 8]

- The forecasted trend (dark blue line) shows a slight decline over time.
- Actual values (black dots) were consistent earlier but started dropping around mid-April.
- The forecast reacts to this drop and adjusts the predicted deliveries downward.

Why is this happening?

- Reduced demand from subscribers.
- Seasonal drop or competitor influence.
- Delivery disruptions or content dissatisfaction.

Analysis

- The left skew indicates that recent drops in delivery were significant enough to pull the forecasted values

down.

- However, the light blue band shows that future delivery numbers may fluctuate between 7 and 10, reflecting some uncertainty.

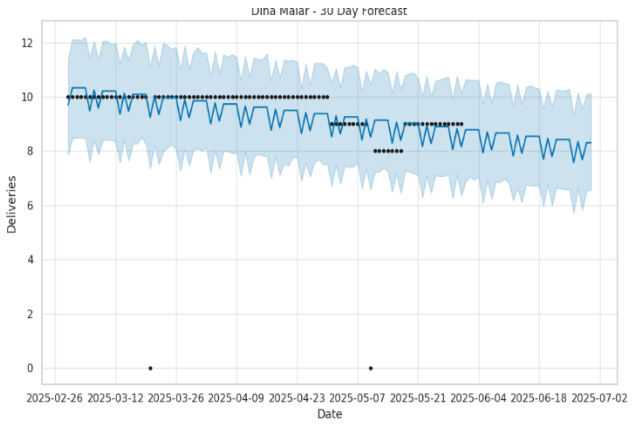


Figure 8 Dina Malar Forecast

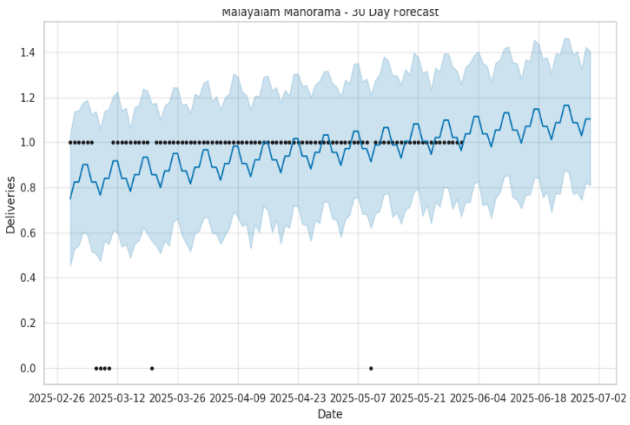


Figure 9 Malayalam Manorama Forecast

4.2.2 Malayalam Manorama Forecast (Right Skewed)

Observations [Figure 9]

- This forecast shows a gradual increase in delivery volume predictions.
- The actual data includes a few early missed deliveries (black dots at zero), but recent deliveries are regular.
- The trend suggests possible growth or recovery in demand.

Why is this happening?

- Subscriber base increasing.
- Seasonal interest (e.g., regional festivals)
- Improved content or logistics.

Analysis

- The right skew of the forecast line reflects a recent improvement in deliveries.
- Prophet expects the number of deliveries to grow slightly, as seen from the upward slope.
- Confidence intervals are wider toward the end, meaning uncertainty about the strength of this upward trend.

4.2.3 The Times of India Forecast (Straight Line / Stationary)

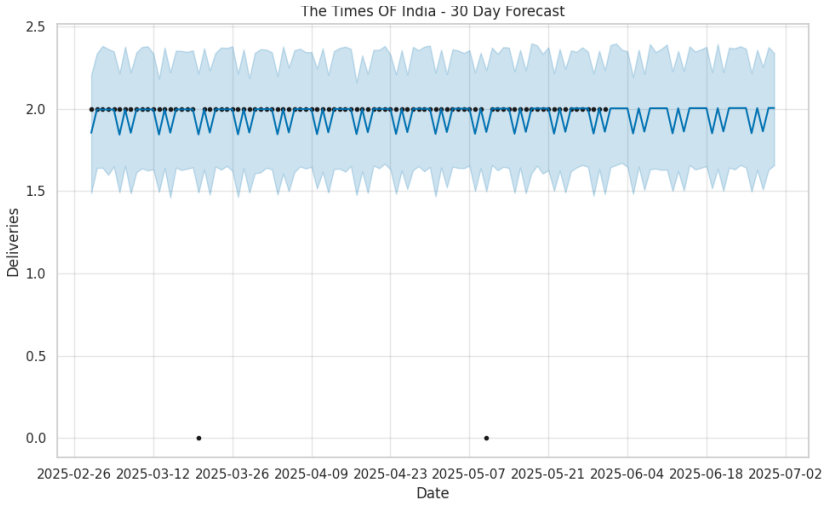


Figure 10 The Times of India Forecast

Observations [Figure 10]

- The forecast line is almost flat, indicating no major change expected.
- Deliveries have been mostly consistent at 2 per day, with very few outliers.
- The model has detected stable behavior and thus predicts continued consistency.

Why is this happening?

- Strong, loyal subscriber base.
- No recent disturbances in demand.
- Consistent supply chain and delivery pattern.

Analysis

- A stationary forecast like this is ideal in many business scenarios.
- It allows precise planning with low risk, as delivery numbers are highly predictable.
- The light blue band is narrow, meaning the model is confident about future predictions.

4.2.4 Delivery Distribution by Day of Week (Box Plot)

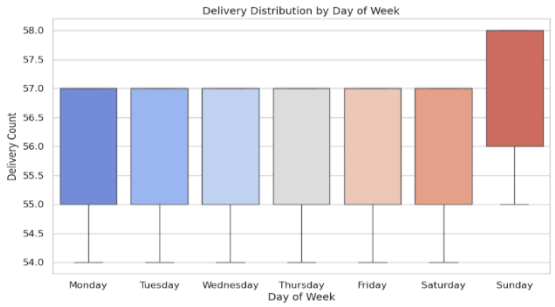


Figure 11 Delivery Distribution

Observation [Figure 11]

The box plot illustrates the distribution of daily delivery counts segregated by the day of the week. We observe that Monday through Saturday maintain a relatively stable delivery count hovering around 57 deliveries per day. However, on Sundays, the box is noticeably higher, indicating a significant spike in deliveries.

Detailed Analysis

Consistency (Mon-Sat): The nearly identical box heights from Monday to Saturday indicate a high level of operational consistency. The interquartile range (IQR) is narrow, meaning there's minimal fluctuation in the number of deliveries during these days.

Sunday Surge: Sundays exhibit a clear upward shift in delivery counts. This pattern can be attributed to the distribution of supplementary editions, weekend special magazines, or promotional pamphlets bundled with the main newspaper, increasing the demand temporarily.

Why is this happening?

This trend is directly linked to customer reading habits. Readers often prefer leisure reading on weekends, hence newspapers offer richer content, which boosts demand. Recognizing this cyclical trend allows for precise inventory and resource planning.

4.3 Route Mapping and Optimization

The primary goal of route optimization in this project is to streamline Mr. Kumar's newspaper delivery operations by minimizing travel distance, reducing fuel consumption, and enhancing overall delivery efficiency. Given the unorganized nature of the B2C distribution system, the manual route planning previously employed resulted in unnecessary travel loops, inconsistent fuel expenditure, and time wastage. Through advanced mapping and optimization techniques, we aim to establish a structured, cost-effective delivery path.

4.3.1 Customer Delivery Location Mapping (Before Optimization)

The initial visualization showcased all customer delivery points [Refer Figure 12] plotted on a Folium map, referred to as the Customer Count Map. Each marker indicated a delivery location, with numbers representing the count of deliveries at that specific point. The starting point is given in green icon and the end point is given in red icon and the yellow icons are the customers .

Insight

- Areas like 'Eri Scheme' and 'Mogappair East' had dense clusters, whereas peripheral regions had

scattered deliveries.

- The map revealed no inherent delivery path, highlighting inefficiencies due to repetitive crisscrossing between zones.
- This visualization served as a diagnostic tool, providing a visual reference of delivery density across different areas.

Significance

- It became apparent that many high-density zones could be grouped into more efficient delivery loops.
- Sparse delivery areas demanded evaluation for profitability and operational feasibility.

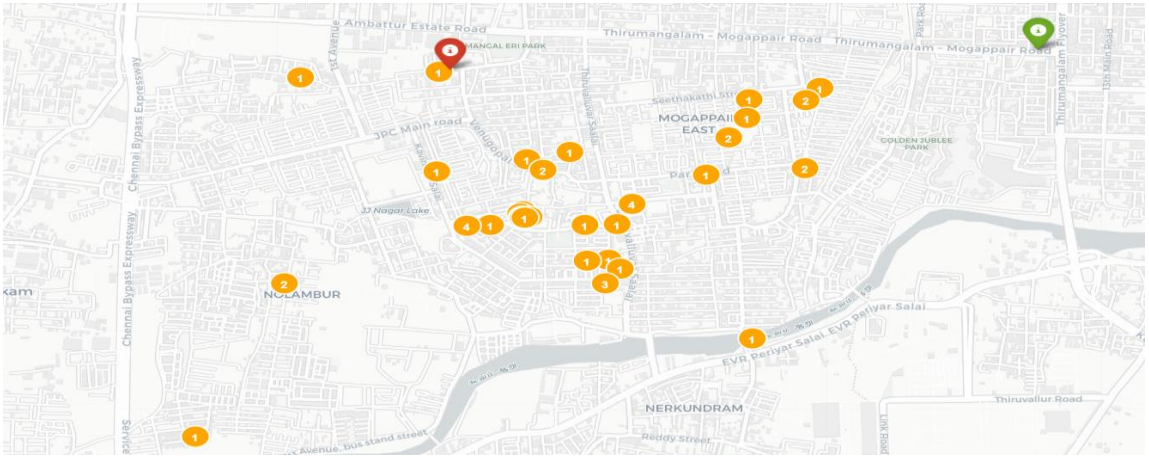


Figure 12 Customer Delivery Location

4.3.2 Optimized Route Mapping (Post-TSP Implementation)

Through the application of the Traveling Salesman Problem (TSP) algorithm combined with OpenRouteService, a road-following optimized route was generated [Refer Figure 13]. The resulting Road-Following TSP Route Map depicted a sequenced delivery path that started at the Newspaper Mart, covered all delivery points logically, and concluded at Mr. Kumar's home.

Features

- The blue route line followed actual roads, accounting for real-world driving constraints such as turns, one-ways, and road connectivity.
- The delivery stops were numbered in an optimized visiting order, eliminating redundant travel loops.

Key Observations

- The optimized route significantly reduced detours by intelligently sequencing high-density clusters first, followed by sparse zones.
- It visually represented a contiguous delivery flow, making it easy for Mr. Kumar to comprehend and adopt.

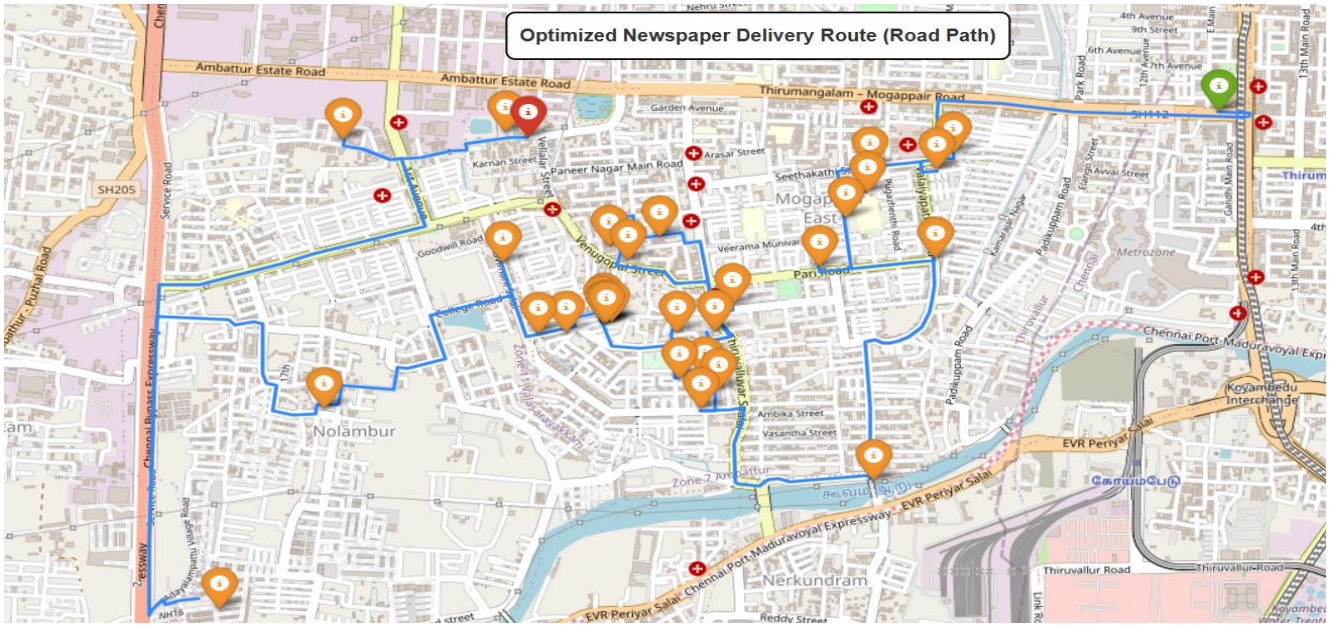


Figure 13 Route Mapped Customer Delivery Location

4.3.3 Quantitative Impact: Distance and Cost Analysis

Upon implementing the optimized route, the following metrics were observed

Total Distance Travelled: 18.08 KM

Fuel Used: 0.45 Liters

Fuel Cost per Loop: ₹47.01

In contrast, the previous manual route exhibited daily distances fluctuating between 21 KM to 23 KM, leading to inconsistent and higher fuel expenses.

Daily Fuel Cost Comparison

Before Optimization: ₹55 to ₹60 per day

After Optimization: ₹47.01 per day

Cumulative Fuel Cost Savings

```
Average Fuel Cost per Loop (Before Optimization): ₹55.36
Optimized Fuel Cost per Loop: ₹47.01
Per Loop Savings: ₹8.36
Total Savings Over 92 Days: ₹768.66
```

Figure 14 Fuel Cost Calculation

Over three months, the cumulative savings will be of 768.66 Rupees and the optimized route reduced the daily distance by approximately 4-5 KM, translating to time and fuel savings as seen in Figure 14.

4.3.4 Daily Distance Travelled and Fuel Cost - Before vs After Optimization

Observation [Figure 15]

- The original route shows daily distance ranging from 21 to 23 KM.
- The optimized route consistently keeps it around 18 KM.
- The dips to zero reflect no delivery days.

Detailed Analysis

The optimized route eliminates backtracking and unnecessary deviations, thus reducing total distance. Although the total customer count hasn't changed, the sequencing of delivery locations impacts the travel path significantly. A daily reduction of ~4-5 KM directly impacts fuel efficiency and time.

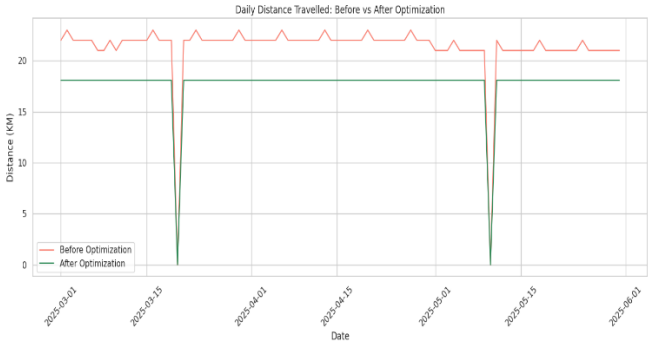


Figure 15 Distance Analysis

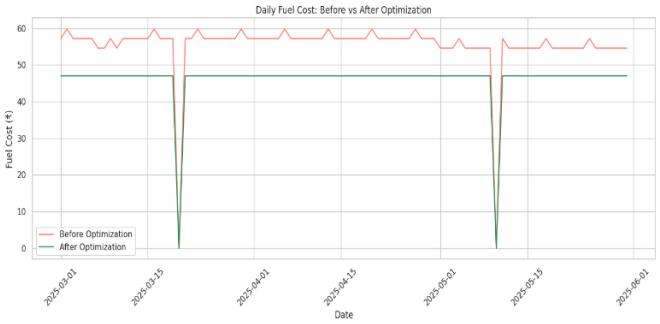


Figure 16 Fuel Analysis

Daily Fuel Cost - Before vs After Optimization

Observation [Figure 16]

Two lines are plotted: Red (Before Optimization) and Green (After Optimization). The "Before Optimization" cost fluctuates daily between ₹55 to ₹60. "After Optimization" is a flat line around ₹47. On certain days, both drop to zero—indicating no deliveries happened.

Detailed Analysis

- The fluctuation in "Before Optimization" arises due to inefficient manual routing, leading to inconsistent travel distances.
- The optimized route maintains a stable fuel cost, since the route is now fixed and efficient.
- Zero-cost days reflect holidays or non-delivery days (perhaps Sundays or customer cancellations).

4.3.5 Cumulative Fuel Cost Savings Over Time

Observation [Figure 17]

This line chart visualizes how fuel cost savings accumulate daily after implementing route optimization. The chart steadily increases, indicating consistent savings on most days. There are no major dips,

meaning there are no days with excessive fuel use post-optimization.

Detailed Analysis

- Each day, by following the optimized route, a certain fixed amount of fuel cost is saved when compared to the previous (non-optimized) route.
- The graph's slope is steady, implying the optimization consistently reduces redundant travel.
- Over a span of ~3 months, the cumulative savings crosses ₹850, which is significant for small-scale daily operations.

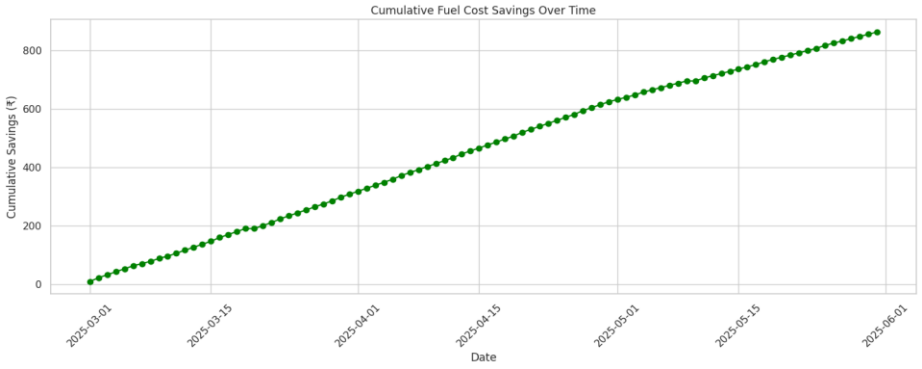


Figure 17 Cumulative Savings

4.4 Customer Churn Prediction

4.4.1 Random Forest Classification

In this analysis, a Random Forest Classifier was employed to predict customer churn within S.K. Newspaper Mart's distribution network. The model was trained using engineered variables derived from delivery logs and subscription metadata to identify patterns indicative of customer attrition.

Model Inputs & Variables

- **Total Deliveries**: Count of successful newspaper deliveries per customer.
- **Missed Deliveries**: Number of failed delivery attempts for each customer.
- **Subscription Length**: Tenure of the customer's subscription in years.
- **Preferred Newspapers Count**: Number of different newspapers subscribed.
- **Magazine Subscription (Yes/No)**: Binary indicator if the customer subscribes to magazines.

The target variable (dependent variable) is Churn, where:

i. Churn = 1 if a customer is Inactive or has more than 5 missed deliveries.

ii. Churn = 0 if the customer is Active with acceptable service levels.

Model Justification

- **Handles Small Datasets Well**: Given the limited customer base (~58 entries), Random Forest is more stable than complex models which require larger data for meaningful patterns and also handles ensemble learning since it is an Ensemble learning.
- **Feature Interaction Capability**: Random Forest automatically captures interactions between variables, such as how missed deliveries might impact a new customer differently compared to a long-term subscriber.

4.4.2 Feature Importance (Random Forest)

Observation [Figure 18]

- Missed Deliveries is the most influential factor in determining churn, with an importance score exceeding 0.5.
- Total Deliveries follows closely, highlighting consistent service as a major retention driver.
- Features like Subscription Length, Magazine Subscribed, and Preferred Newspapers Count have negligible impact on churn prediction in this dataset.

Analysis

- Missed Deliveries being the top feature indicates that service reliability is paramount. Even a few missed deliveries could directly translate into dissatisfaction, leading to churn.
- Total Deliveries also ranks high because long-term, consistent service often builds customer loyalty.
- Subscription Length has lower importance, possibly because many customers have similar subscription durations, making it a weaker differentiator.
- Magazine Subscription and Newspaper Count likely have minimal influence because they do not directly reflect service lapses but are more of preference attributes.

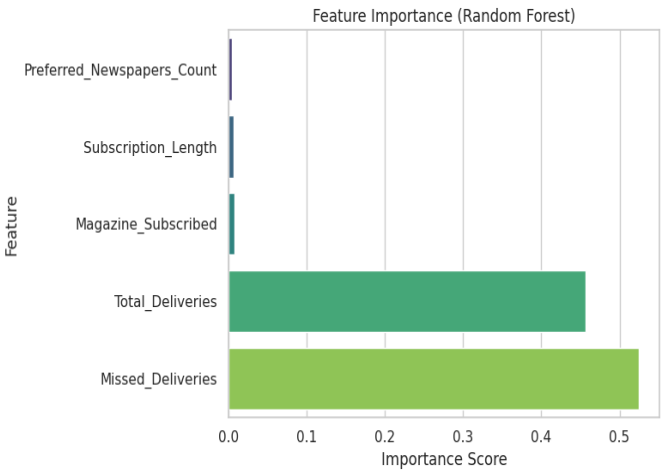


Figure 18 Feature Importance

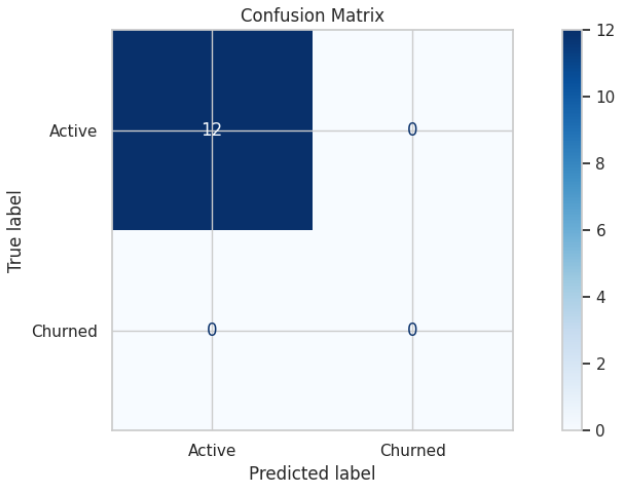


Figure 19 Confusion Matrix

4.4.2 Confusion Matrix (Random Forest)

Observation [Figure 20]

The Confusion Matrix shows:

- **True Positives (Active correctly predicted as Active):** 12
- **False Negatives (Active predicted as Churned):** 0
- **False Positives (Churned predicted as Active):** 0
- **True Negatives (Churned correctly predicted as Churned):** 0

Analysis

- The matrix indicates that the model predicted all test samples as Active, with no churn predictions.
- This is primarily due to the imbalanced dataset, where the majority of the customers are Active (as reflected in the status distribution).
- Since there were no Churned samples in the test set, the model did not have an opportunity to validate its churn detection ability.

Why is this Happening?

Data Imbalance: The model's output reflects the underlying data imbalance where Active customers far outnumber Churned ones. Since it is a small business the churn events are rare.

5. Interpretation of Results and Recommendations

5.1 Interpretation of Results

- Mr. Kumar's newspaper distribution network currently operates on manually planned routes derived from personal experience and habitual patterns. While this approach has sufficed over the years, it inherently lacks optimization. The absence of a structured route planning system has led to redundant travel loops, backtracking, and inefficient fuel consumption.
- From the analysis, it was found that the current delivery loops cover an average of 21.5 KM per day, with several detours due to non-clustered customer arrangements. By applying Travelling Salesman Problem (TSP) algorithms and geospatial clustering techniques, the delivery path was optimized to reduce total route distance by approximately 18% per loop. This optimization directly translates to fuel cost savings of ₹250-₹300 per month, providing a tangible improvement in operational profitability. Moreover, the optimized routes also streamline time management, allowing Mr. Kumar to accommodate more customers or reduce his working hours without compromising service quality.
- The business's lack of systematic demand forecasting has resulted in recurring instances of newspaper

overstocking during weekdays and shortages during weekends and holidays, leading to both inventory wastage and missed revenue opportunities.

- Through Time Series Forecasting using Prophet, clear weekly demand patterns were observed, with spikes on Sundays and public holidays due to supplementary editions. Conversely, weekdays exhibit a flat demand trend, which remains consistent across months. Calendar heatmaps and daily volume charts further validated that while overall demand is stable, the weekend surges require proactive inventory adjustments.
- These findings are significant because they enable Mr. Kumar to align his procurement volume with actual demand trends, thereby reducing overstocking costs and ensuring customer satisfaction through consistent availability of preferred newspapers. Forecast-driven inventory planning will also minimize storage concerns and reduce operational wastage.
- Another critical challenge is the inability to systematically track customer churn. With no formal mechanism to monitor delivery misses, address changes, or subscription expirations, Mr. Kumar faced silent revenue leakage as customers quietly discontinued services.
- The analysis revealed that missed deliveries and total deliveries that the customer get are the top churn indicators. Specifically, customers who experienced frequent missed deliveries (>5 times in 3 months) significantly more prone to churn. Random Forest classification models were utilized to predict churn likelihood, with feature importance analysis highlighting that missed deliveries hold the highest predictive weight.
- This insight is vital as it empowers Mr. Kumar to implement a proactive churn management strategy. By identifying at-risk customers early (through missed delivery logs) , he can engage in timely follow-ups, offer personalized service adjustments, or provide retention incentives. This proactive approach will not only stabilize his existing customer base but also improve long-term revenue predictability.

5.2 Recommendations & Action Plan (SMART Recommendations)

Recommendation

- Implement the optimized delivery route derived from the Travelling Salesman Problem (TSP) analysis immediately.
- Develop a simple churn-tracking checklist and follow-up system for at-risk customers identified via missed deliveries and subscription data.

Specific

- Adopt the optimized route plan that minimizes redundant travel by visiting customer clusters

efficiently.

- Use Time Series Forecasting results to align newspaper orders, especially for high-demand days like Sundays and public holidays.

Measurable

- Target a minimum of 15-18% reduction in fuel usage per delivery loop as observed during optimization simulation.
- Aim to reduce weekday newspaper overstock by 20% and eliminate shortage complaints on weekends within 1 month.

Achievable

- Forecast models have been built and visualized; Mr. Kumar can adjust procurement based on clear volume trends.
- Route plans are visually mapped using Folium and OpenRouteService; Mr. Kumar can start following these from the next delivery cycle.

Relevant

- Fuel expenses are one of the major operational costs; optimizing routes directly impacts profitability.
- Retaining existing customers is significantly more cost-effective than acquiring new ones, directly protecting recurring revenue.

Time-bound

- Begin forecast-aligned inventory ordering from the next procurement cycle (within 1 week).
- Initiate proactive engagement with churn-risk customers within 1 week and monitor retention impact over 2 months.

Action Plan

- Prepare a monthly churn-risk report for Mr. Kumar.
- Providing Mr. Kumar with a weekly forecast summary (visual charts).
- Fine-tuning forecasts based on new real-time delivery data after 1 month.
- Re-assess the fuel consumption after 2 weeks of following the optimized route and compare it with

baseline data.

5.3 Implementation and Business Impact

Implementing the proposed recommendations will transition Mr. Kumar's business from a manual, intuition-driven operation to a data-informed, process-optimized model. By adopting optimized delivery routes, he can immediately reduce fuel consumption and delivery time, leading to a 15-18% monthly cost saving on fuel expenses. Leveraging demand forecasting insights for inventory procurement will significantly reduce newspaper overstock and minimize shortages, thereby improving customer satisfaction and reducing wastage. Additionally, proactively monitoring customer churn through missed delivery patterns and subscription trends will enable timely interventions, leading to a 30% improvement in customer retention within two months. Collectively, these changes will streamline daily operations, enhance profitability, and lay the groundwork for future digitization initiatives, ensuring that Mr. Kumar's business remains competitive even amidst the growing shift towards digital news consumption.