**LAST MILE LOGISTICS FOR RURAL INDIA**

Submitted By - Abhinav Kumar

**#### Executive summary**

This project investigates data-driven, multi-modal logistics solutions for rural India, where infrastructure gaps and connectivity issues continue to hinder growth. By integrating machine learning with optimisation techniques, the study demonstrates how villages can be clustered, depots strategically placed, and delivery routes optimised for cost-effectiveness and efficiency. Rural India holds significant untapped potential for e-commerce, public services, and development initiatives, and improving logistics in these areas is essential for unlocking that value.

**#### Rationale**

Why should anyone care about this question?

Rural India’s logistics systems face critical bottlenecks, including poor road infrastructure, incomplete mapping, fragmented transport options, and unreliable last-mile delivery. These constraints make the distribution of goods, healthcare supplies, and government benefits both slow and inefficient. By applying data science and optimisation frameworks, this project highlights pathways to bridge these gaps, enabling equitable access, strengthening public distribution systems, and expanding e-commerce penetration. The effective integration of rural supply chains with national logistics networks can drive inclusive socio-economic growth.

**#### Research Question**

What are you trying to answer?

This study addresses key challenges such as:

* Clustering and optimisation using pincode datasets.
* Mapping transport routes and resources to reduce or avoid multiple mode changes.
* Predicting delivery timelines with higher accuracy.
* Identifying inefficiencies and minimising logistical bottlenecks.
* Enhancing government outreach and ensuring last-mile delivery of benefits in rural areas.

**#### Data Sources**

What data will you use to answer you question?

Sample village coordinates for Nuh district (10 representative villages).

Sample Delhivery delivery dataset: Delhivery Kaggle Dataset

Village and road network data via OSMnx (OpenStreetMap)

**#### Methodology**

What methods are you using to answer the question?

Following methods were used:-

* Clustering: KMeans clustering on village coordinates to identify optimal delivery hubs.
* For Routing - NetworkX shortest-path analysis for intra-village delivery; OR-Tools optional for advanced routing.
* Random Forest regression on sample delivery data to predict delivery times based on distance and package weight.
* For Visualization - Folium maps for village locations, Matplotlib/Seaborn for clusters, paths, feature importance, and correlation heatmaps.

**#### Results**

What did your research find?

The study demonstrates the importance of adopting flexible and reliable machine learning models for complex rural logistics networks. The insights highlight how predictive modelling can streamline last-mile delivery, reduce inefficiencies, and improve the reliability of logistics operations in rural India.

**#### Next steps**

What suggestions do you have for next steps?

This project used a relatively small dataset compared to the scale of India. To achieve more accurate and generalisable results, future work need to involve:

* Expand dataset to include all villages of India for higher accuracy.
* Integrate real Delhivery shipment data for improved machine learning predictions.
* Explore hybrid routing methods combining NetworkX and OR-Tools for scalable solutions.
* Implement a full Streamlit dashboard for interactive analysis and stakeholder presentation.

**#### Outline of project**

<https://github.com/abhi86031/Last-Mile-Logistics-for-Rural-India/blob/main/Last%20Mile%20Logistics%20-%20Nuh%20Village.ipynb>

**##### Contact and Further Information**

Abhinav Kumar

email id - singh2abhi@gmail.com