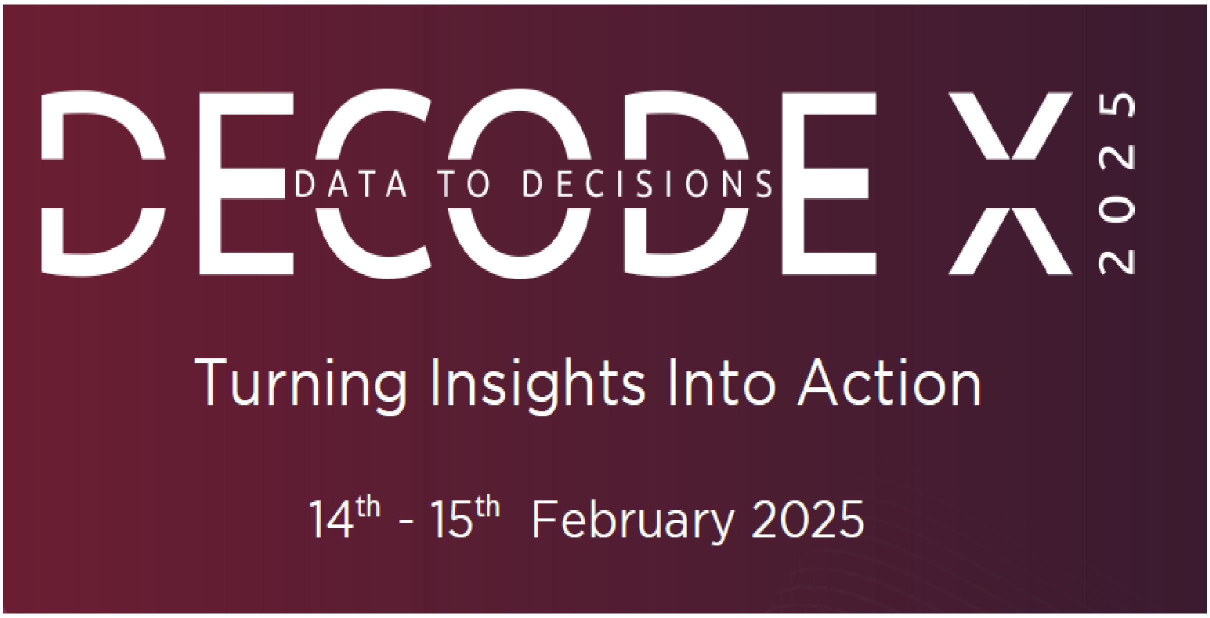


DecodeX Hackathon

N. L. Dalmia Institute of Management Studies and Research

Lost in Transit: The PalletSense Challenge

Team Name: Analytica

Team Member:

1. **Shruti Khapekar (Team Leader)**
2. **Sandesh Tayde**
3. **Keyur Parkhi**

Identification of Business Problem

Background

PalletSense provides smart pallets embedded with IoT sensors to track shipments in real-time. However, due to limited GPS ping frequency (only 2-3 times per day per pallet), there are visibility gaps in tracking the pallets' movement. This results in potential losses, inefficiencies, and operational delays.

The Challenge

PalletSense is looking for an innovative approach to improve tracking without increasing the GPS ping frequency. So, Develop an approach that:

- Enhances visibility into pallet journeys using analytical techniques to ensure more regular updates on locations of its pallets.
- Detects deviations and generates alerts for off-route scenarios to allow adoption of pro-active intervention measures.

Key Objectives:

- Ability to track all pallets on-route with greater regularity and reliability.
- While ensuring regular updates on pallets whereabouts, identify how a deviation can be conclusively established so that alerts can be triggered.

To Access Code: **Appendix**

Data Preparation and Availability

PalletSense has provided two Datasets:

1. Known Route Repository

A collection of commonly travelled routes based on historical data, including known locations.

- 3541 records, representing known route locations.
- 24 unique routes, with the most frequent "1_Portland_Atlanta" (296 occurrences).

Data Description:

Route_ID: Unique identifier for the route

Timestamp: Time at which location was pinged

Latitude: Latitude of the pinged location

Longitude: Longitude of the pinged location

Summary Statistics:

```
                  Route_ID      Time        Date timestamp
count                 3541      3541        3541      3541
unique                  24      3302           3      3234
top     1_Portland_Atlanta  04:27:31  09-07-2024   27:31.3
freq                   296        26        2319        26
Duplicate Records (Known Route): 153
```

2. Ongoing Trip Data

Real-time GPS pings from pallets currently in transit.

Data Description:

Pallet_ID: Unique identifier of pallet

Timestamp: Timestamp at which ping was recorded

Latitude: Latitude of the ping

Longitude: Longitude of the ping

- 851 unique pallet pings, representing ongoing shipments.
- 625 unique pallet IDs, meaning some pallets have multiple pings.

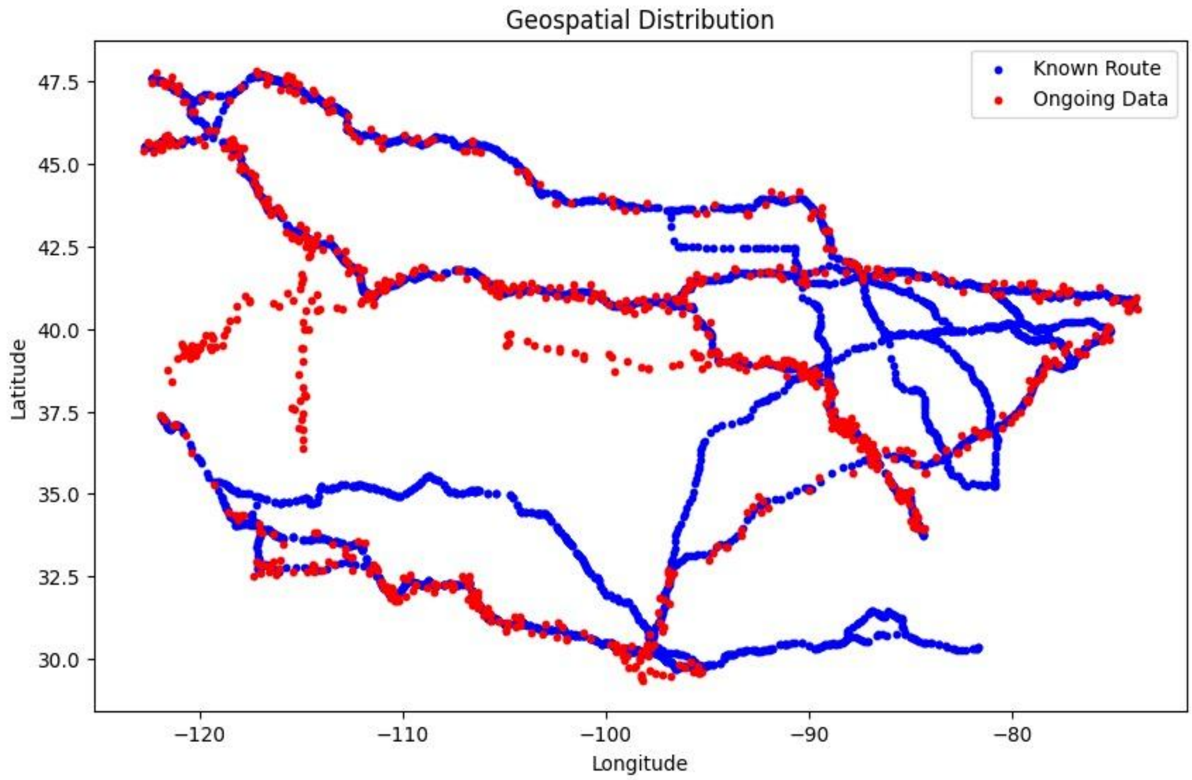
Summary Statistics:

```
       Pallet_ID      Time        Date timestamp
count        851       851         851       851
unique       625       848           3       842
top        Y0623  14:16:50  06-09-2024   00:00.0
freq           2         2         424         2
Duplicate Records (Ongoing Data): 0
```

Data Preprocessing Steps:

- Converted timestamps to datetime format for proper analysis of the data.
- Merged ongoing trip data with historical known routes for route adherence analysis.

Geospatial Distribution:



Intepretation:

Red and Blue points overlap closely, indicating the vehicle is following the planned route.

Proposed Approach

To enhance pallet tracking and detect off-route movements without increasing GPS ping frequency, we propose a hybrid approach combining spatial analysis, machine learning, and visualization techniques.

1. Route Adherence Analysis (On-Route vs. Off-Route Detection)

- Use of cKDTree (Spatial Nearest Neighbour Search) to check if a pallet's location is close to a known route.
- Threshold-Based Distance Calculation (10,20 and 30 km) to classify pallets as "On-Route" or "Off-Route". **To Identify Known Routes Followed by Pallet.**

2. Anomaly Detection Using Machine Learning

- Isolation Forest: Identifies unusual movements based on GPS patterns.
- Haversine Distance Calculation: Ensures detected anomalies are truly far from known routes. **To Get Pallet Deviate from Their Route**.

  .

3. To Get Pallets Travelling Together

- Find out the first occurrence of each unique Pallet by unique Date.
- Identify Pallet travelling on the same Route.

4. Detecting Stationary Pallets

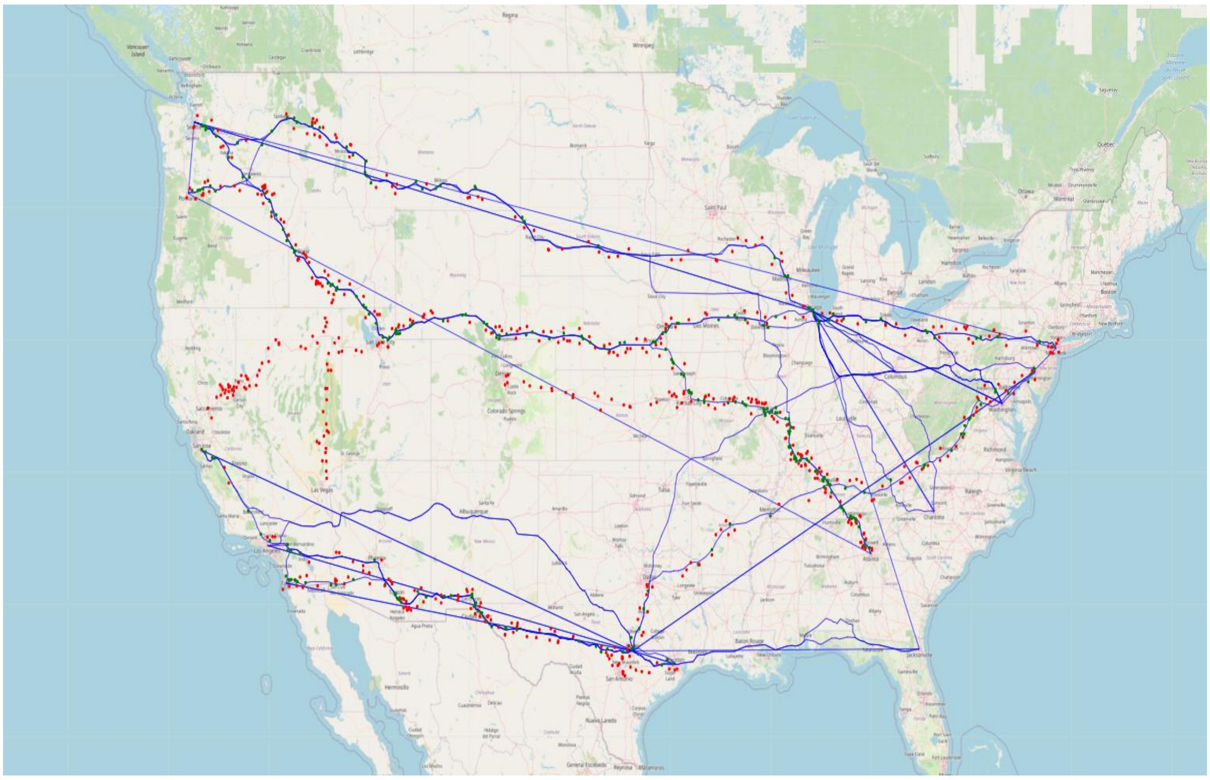
- Detection of Stationary Pallets using Analytical Method.

Analysis

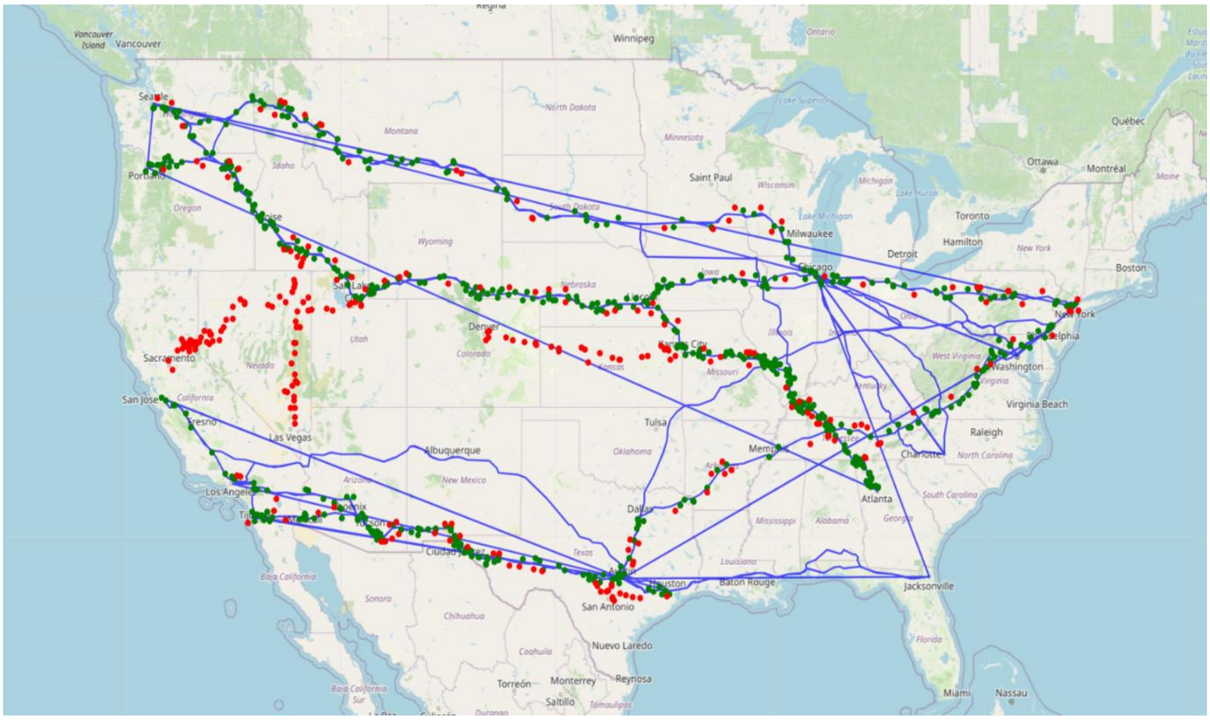
1. Route Adherence Analysis

The pallet route analysis employs a nearest-neighbour search algorithm using a cKDTree spatial index. This index is built from the latitude and longitude coordinates of the established route. For each pallet in the ongoing data, the algorithm identifies the closest point on the known route and calculates the distance between them. If this distance exceeds a predefined threshold (approximately 10,20 and 30km), the pallet is flagged as off-route, indicating a potential deviation. The results are visualized on a Folium map, distinguishing on-route and off-route pallets with distinct markers.

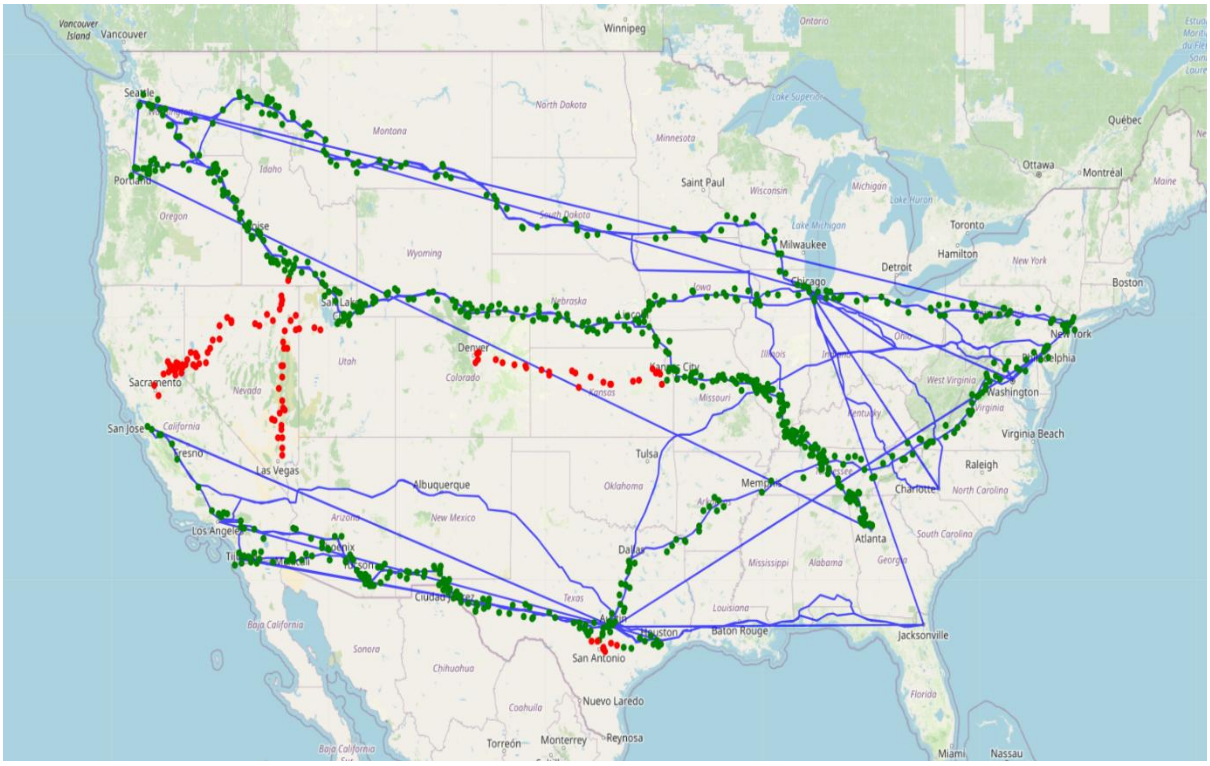
Off-Route Pallets(10km): 558



Off-Route Pallets(20km): 264



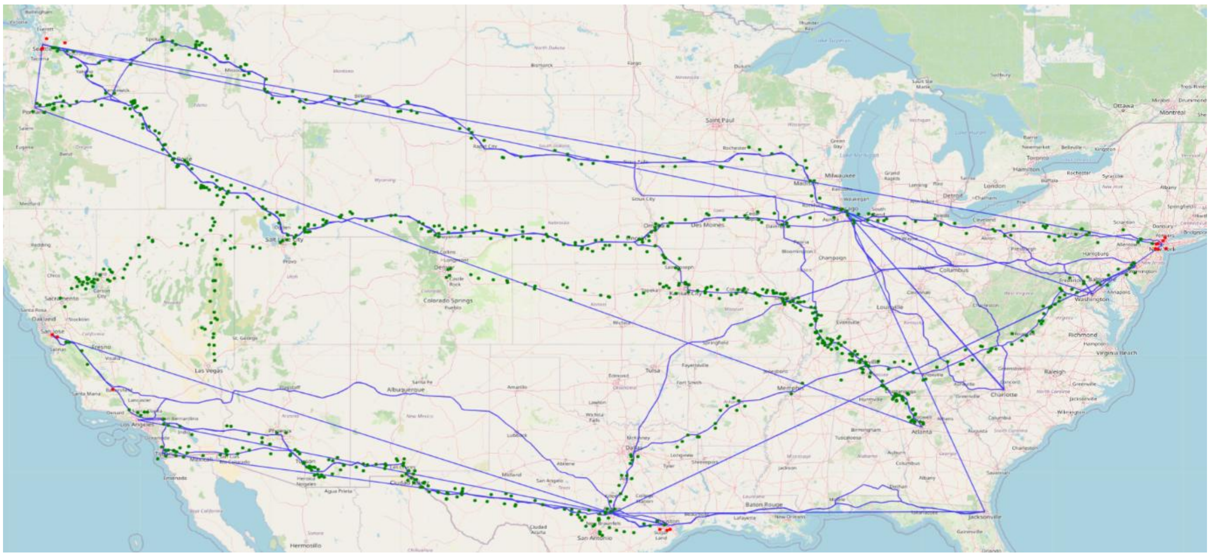
Off-Route Pallets(50km): 106



2. Anomaly Detection Using Machine Learning

This approach combines Isolation Forest and Haversine distance. Isolation Forest detects unusual GPS patterns, flagging potential anomalies. Haversine distance verifies these anomalies by measuring their distance from known routes. Pallets flagged as anomalous and exceeding a distance threshold are identified as true route deviations, enhancing detection accuracy.

Total Anomaly Detection: 69



Known Routes Followed by the Pallet

To identify the if the given pallet follows a route, check if the coordinates of a particular pallet lie in some route. We have set a threshold of 0.05, i.e. we found the absolute difference between the coordinate of route and pallet is less than or equal to 0.05, we allocate that pallet to that corresponding route.

```
array(['0_Chicago_Columbus', '0_Chicago_Seattle', '0_Portland_Atlanta',
       '1_Austin_Philadelphia', '0_Austin_San Diego',
       '0_Austin_Philadelphia', '1_New York City_Seattle',
       '0_New York City_Seattle', '1_Chicago_Seattle',
       '1_Chicago_Washington, D.C.', '1_Portland_Atlanta',
       '0_Austin_Los Angeles', '1_Austin_San Diego',
       '0_Chicago_Washington, D.C.'], dtype=object)
```

3. To Get Pallets Travelling Together

To establish if the pallets travel together, we will first make different data sets for different values of first occurrence of Pallet by Date. We need to compare the coordinates of the historic and the real time data. The coordinates may not be exactly same.

For that, we set a threshold of 0.05 for the acceptable variation in the coordinate pairs. Then we make groups of pallets based on the routes of historic column. We then combine pallets with same route.

Date: 05-09-2024

	Matched_Route_ID	Pallet_ID
0	0_Austin_Los Angeles	[R81212, F6284, F91222, J4308, W21223]
1	0_Austin_Philadelphia	[P31634, C71205, P91652, G51655, K81656, A91512, X91661]
2	0_Austin_San Diego	[Z5254, N8256, Z5258, S1259, D8260]
3	0_Chicago_Columbus	[Y0623]
4	0_Chicago_Seattle	[A8624]
5	0_Chicago_Washington, D.C.	[N61030]
6	0_New York City_Seattle	[U4781, M11008, C11010, U31019, U51031]
7	0_Portland_Atlanta	[B11353, T21358, B7883, E3888, S3907, J51519, Y11520]
8	1_Austin_Philadelphia	[D51632, A11633]
9	1_Austin_San Diego	[I4280, Z1281, P91219, T1313]
10	1_Chicago_Seattle	[P4640, O2645, O61367, T0649, G4652, G6656, C7675]
11	1_Chicago_Washington, D.C.	[I81643]
12	1_New York City_Seattle	[Y0780, J1633, P9791, V5793, X01369, I3797, H3801, J8664, P21381]
13	1_Portland_Atlanta	[L3792, S2884, L8886, S31504, O61508, F31510, L0899, P01518]

Date: 06-09-2024

	Matched_Route_ID	Pallet_ID
0	0_Austin_Los Angeles	[P1329, M31237, O91238]
1	0_Austin_Philadelphia	[U41680]
2	0_Austin_San Jose	[S21256, Q51260, U71262]
3	0_Chicago_Columbus	[H61047]
4	0_Chicago_Seattle	[T01048, B51053, Q51079]
5	0_New York City_Seattle	[L91046, Y21049, D41056]
6	0_Portland_Atlanta	[S3679, E61523, F9914, N4915, O6922, O7924, C1927, R61533, J51538, T61541, K91412, W31418, X01561]
7	1_Austin_Philadelphia	[I91526, M51529]
8	1_Austin_San Diego	[C3318, H31228, R1321, Y81248]
9	1_Austin_San Jose	[J01261]
10	1_Chicago_Charlotte	[T91676]
11	1_Chicago_Seattle	[T8684, T31399, N9689]
12	1_New York City_Seattle	[B41397, M6686, W41402, T51403, H71550, N31562]
13	1_Portland_Atlanta	[V91687]

Date: 07-09-2024

	Matched_Route_ID	Pallet_ID
0	0_Portland_Atlanta	[R11572]
1	1_New York City_Seattle	[K11574]

4. Detecting Stationary Pallet

For the historic data, each route is available for 3 days. We analysed each route individually, by calculating difference between distance of two consecutive points. So that if the difference between the two route is same that will help to conclude stationarity. But while considering the distance time doesn't come under consideration so we make use of speed.

We assumed if the speed is less than 20km/hr then at some point the pallet was stationary on that route. i.e. if it is travelling this much slow it is possible that it has to identify pallets that are stuck at a location.

Most Frequent Route:

1. 0_Chicago Seattle (Frequent)
   Pallets are **A8624, T01048, B51053, Q51079**

2. 0_New York City_Seattle
   Pallets are **L91046, Y21049, D41056, U4781, M11008, C11010, U31019, U51031**

3. 0_Austin_Philadelphia
   Pallets are **P31634, C71205, P91652, G51655, K81656, A91512, X91661, U41680**

4. 0_Austin_Jacksonville

The PalletSense Company should take care when the GPS pallets that are stuck at a location, when it is travelling from "0_Chicago_Seattle".

Conclusion and Findings

Results:

1. **Route Adherence Analysis**

    - Using a spatial nearest-neighbour approach (cKDTree), off-route pallets were identified based on distance thresholds.

    - Comparing number of Off-route pallet counts based of 10km, 20km & 50km.

2. **Anomaly Detection with Machine Learning**

    - The Isolation Forest algorithm identified **69 anomalies**, with Haversine distance which will help improving accuracy.
    - These anomalies represent significant route deviations and possible misroutes.

3. **Pallets Traveling Together:**

    - A threshold-based grouping approach identified pallets moving on the same route based on approximation criterion.

4. **Detecting Stationary Pallets:**

    - By analysing speed and distance variations, identified stationary pallets.
    - The most frequent routes with stationary pallets are 0_Chicago-Seattle

Conclusion:

- The give analytical framework enhances accuracy of tracking pallet without increasing GPS ping frequency.

- The route adherence model effectively identifies deviations, allowing for proactive measures.

- Grouping pallets that are traveling together can improve planning and reduce operational inefficiencies.

- Identifying stationary pallets helps prevent delays and ensures timely delivery.