Reducing CO_2 Emissions in Transportation using Machine Learning

Presented by:

Ade Barkah & Patrick Robert

Advisors:

Dr. Josué C. Velázquez-Martínez

Dr. Karla M. Gámez-Pérez

MIT Supply Chain Management ResearchFest 2018 – May 22, 2018

MIT Center for Transportation & Logistics

Sustainable Logistics Initiative

Agenda

Introduction

- Motivation
- Company

Methodology

- Overall process
- Data sources
- Route analysis
- Utilization factors

Results

- *k*-means clustering
- Qualitative analysis of clusters
- Ranking of vehicles in clusters

Conclusions

Motivation

- **Greenhouse gas emissions** are the main drivers of climate change[1]
- **Transportation is the largest contributor to CO_2 emissions** in the US[2]
- **Increasing fuel prices** drive companies to look for operational efficiencies
- **Growth in transportation needs from e-commerce**

Objective:

- **Explore ways to reduce of CO_2 emissions** by analyzing truck-route assignments based on **vehicle types, road topology and traffic conditions**

In our project, we estimate that we could reduce up to 7.2% in fuel consumption and CO_2 emissions by that using the best vehicle type in each cluster of routes.

Sources: [1]EPA, 2016. [2]EIA, 2016.

Company Overview

- Leading retailer in Mexico, Revenue $6.1B (est. Deloitte)
- 1,300 retail stores
- 19 Regional DCs, 600 Warehouses
- 1,200 last mile delivery vehicles
- Customer profile: low and mid-level household income
- The company also offers its products through their online channel
- Product lines: appliances, electronics, apparel & accessories, baby & kids, home & patio, office, toys




Methodology: Overall Process

Route Analysis

- Quality check
- Position normalization
- Elevation correction
- Segmentation analysis
- Distance calculation
- Gradient calculation
- Calculate route statistics

Fuel Analysis

- Quality check
- Statistical analysis
- Outlier removal
- Route matching
- Compute consumption

Load Analysis

- Quality check
- Create utilization bins
- Compile delivery data
- Reference SKU master
- Calculate trip weight
- Calculate utilization
- Assign utilization bin

Clustering

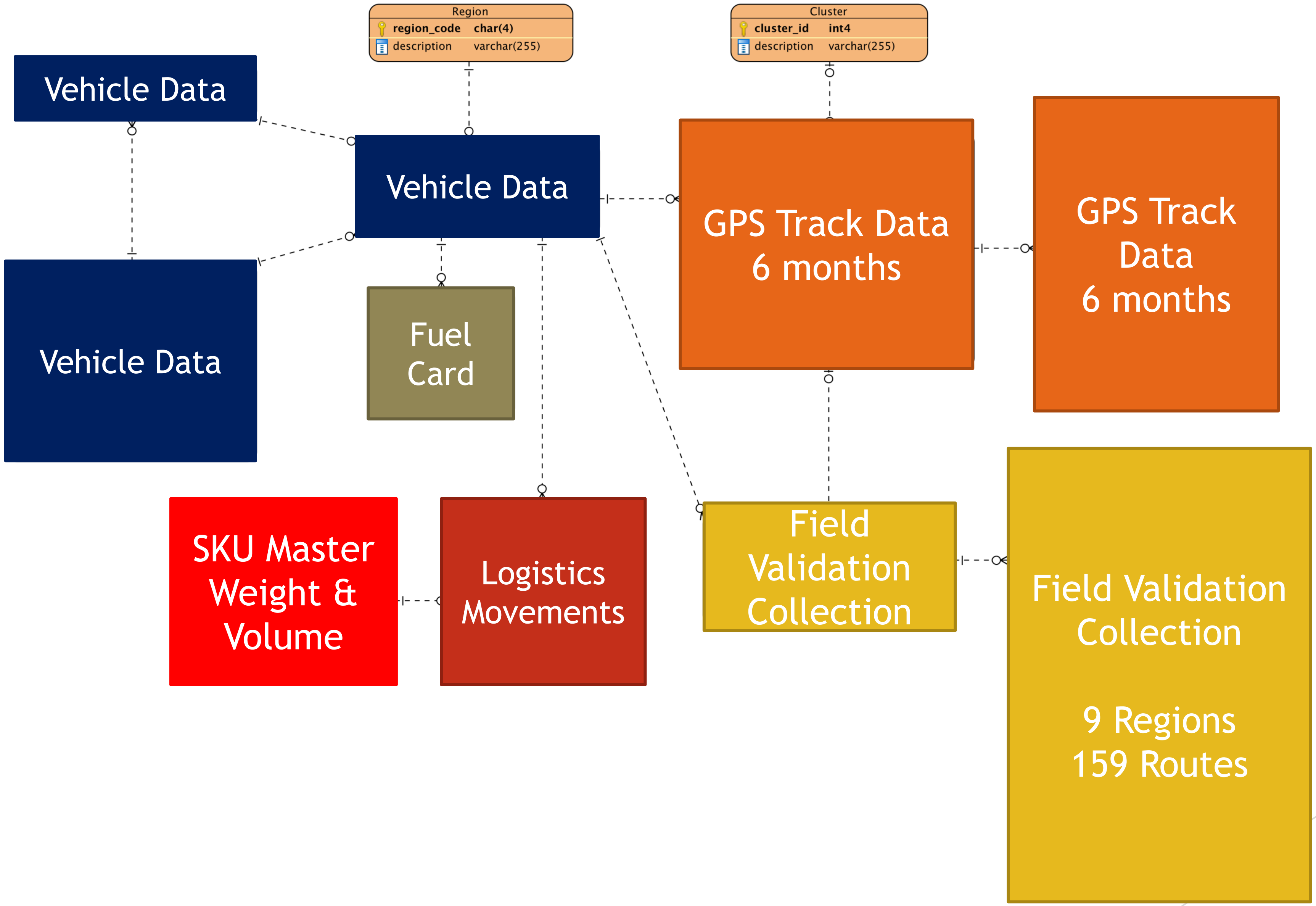
- Collate route factors
- Determine # of clusters
- Cluster each load bin
- Quantitative analysis
- Qualitative analysis

Field Validation

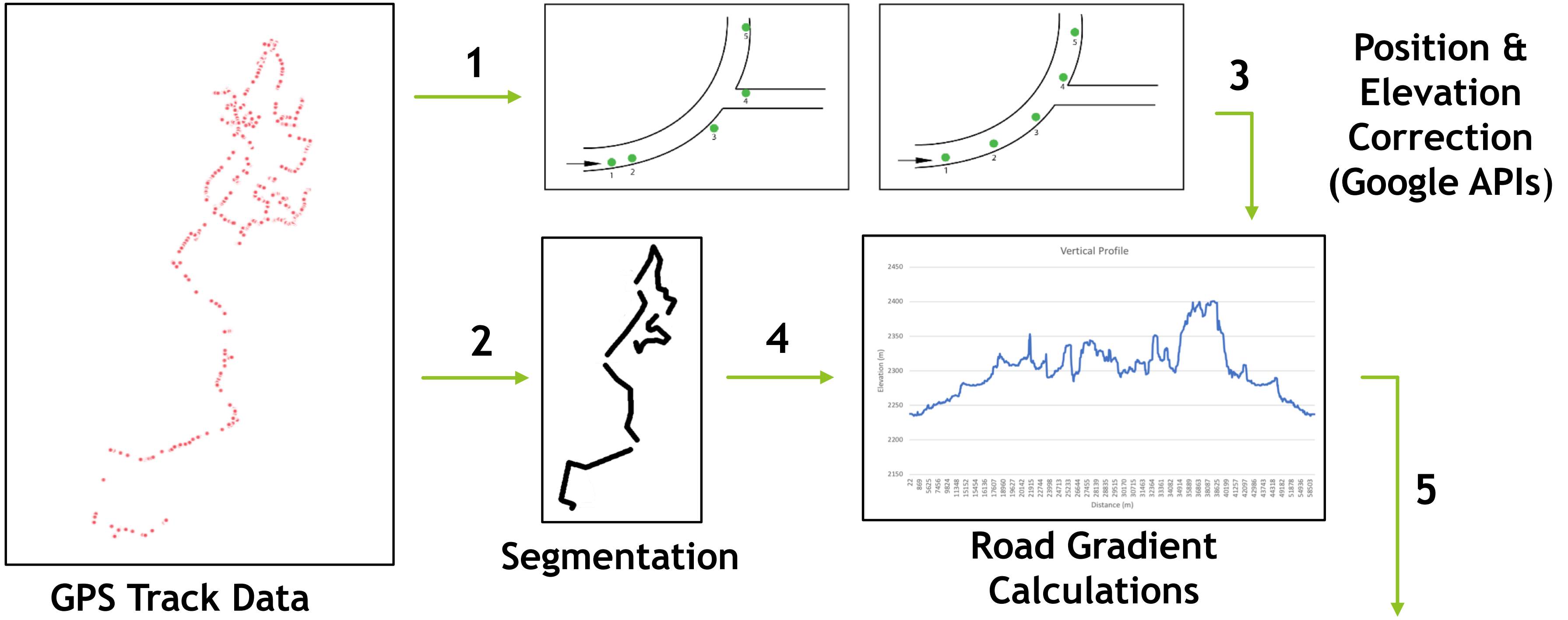
- Design survey
- Create Fulcrum app
- Run survey
- Collect survey results
- Collect truck GPS data
- Validate against clusters

Ranking & Insight Generation

Data Sources



Route Analysis Illustration



Mean Gradient	Median Gradient	Max Gradient	SD Gradient	Var Gradient	Mean Velocity	Median Velocity	Max Velocity	SD Velocity	Var Velocity	Mean Elevation	Median Elevation	Max Elevation	SD Elevation	Var Elevation	Average Segment Length	p_flat	p_2	p_4	GPS Dist

Route Statistics (Features) for clustering

Vehicle Utilization (Load)

Home deliveries – Histogram Vehicle Utilization

For each trip, we estimate weight utilization from the trip's cargo manifest

- For **home deliveries**, the average utilization was 55%
- For **store deliveries**, the average utilization was 76%
- For **all deliveries**, the combined average utilization was 64%

- To control the load effect on fuel consumption calculations, we compare trucks with similar load values to each other
- We define 4 utilization bins to analyze fuel performance

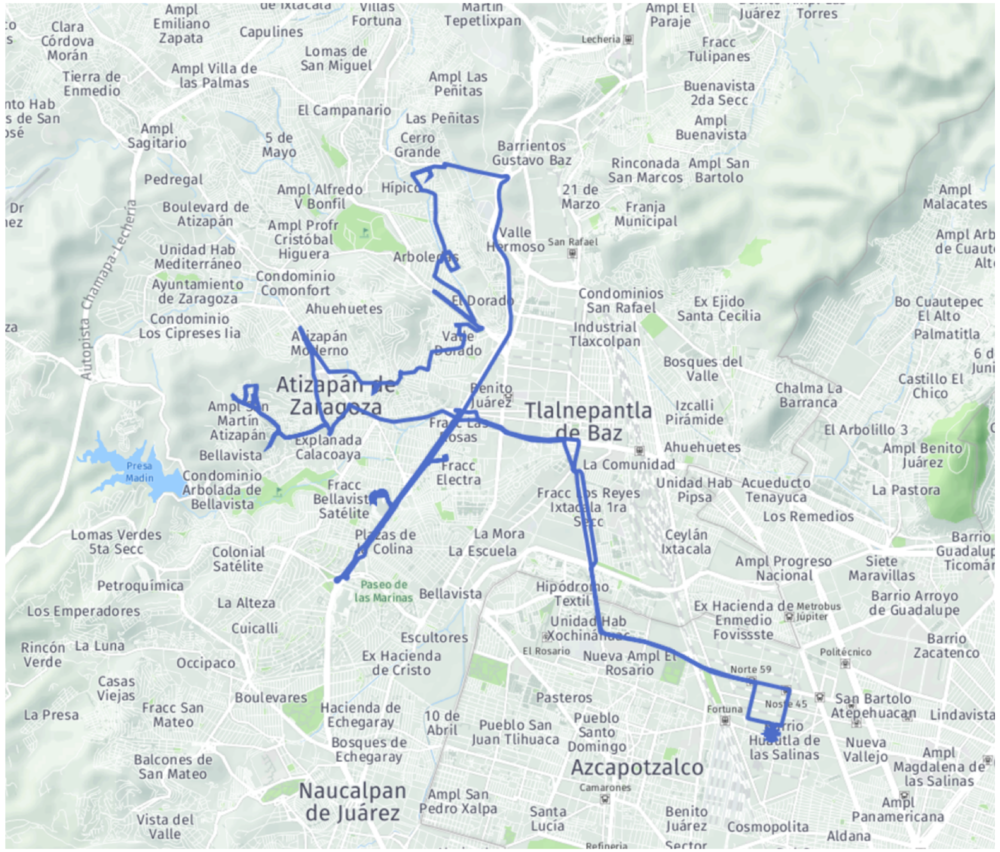
Utilization Bin	Utilization Range (Weight)		Average Utilization
Low Utilization	0	33%	14%
Medium Utilization	33%	66%	48%
High Utilization	66%	100%	82%
Overload	100%	150%	120%

Clustering Factors

From the 24 parameters (factors) computed in the GPS processing steps, we choose six to form our clusters with:

- Gradient variability (proxy for hilly conditions)
- Mean velocity
- Mean elevation
- Average segment length
- Percent of the route that's flat (road gradient is less than +/- 1%)
- Percent of the route that's steep (road gradient is 4% or greater)

Field Validation of the Cluster Factors

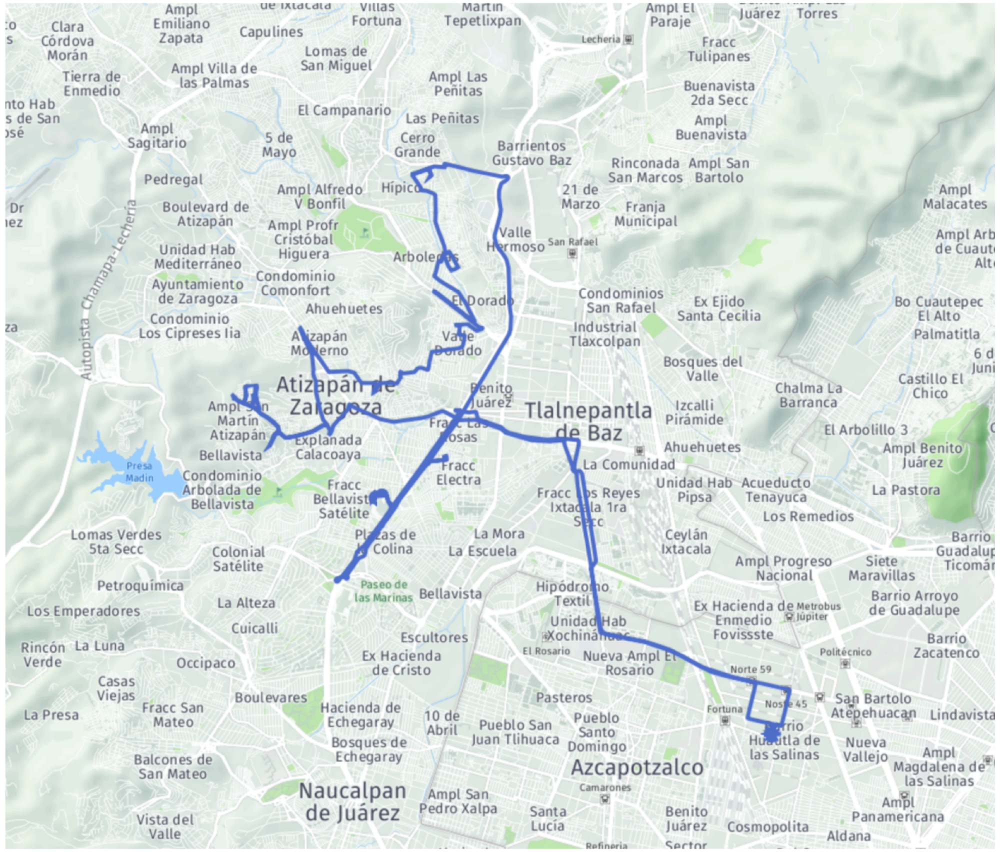


A route in Mexico City

Total GPS distance	59816.60 m
Odometer distance	56613.00 m
Max distance delta	1045.77 m
Max time delta	481.00 s
Total duration	17.84 h
Gradient mean	-0.01%
Gradient median	-0.03%
Gradient max	12.43%
Gradient sd	3.95
Gradient var	15.63
Velocity mean	4.38 m/s
Velocity median	3.55 m/s
Velocity max	21.30 m/s
Velocity sd	3.39
Velocity var	11.48
Elevation mean	2300.55 m/s
Elevation median	2284.00 m/s
Elevation max	2474.00 m/s
Elevation sd	53.12
Elevation var	2821.38
Average segment length	8450.94 m
Percent flat	64.61%
Percent > 2%	13.24%
Percent > 4%	7.65%

Computed Statistics

Field Validation of the Cluster Factors



A route in Mexico City

Total GPS distance	59816.60 m
Odometer distance	56613.00 m
Max distance delta	1045.77 m
Max time delta	481.00 s
Total duration	17.84 h
Gradient mean	-0.01%
Gradient median	-0.03%
Gradient max	12.43%
Gradient sd	3.95
Gradient var	15.63
Velocity mean	4.38 m/s
Velocity median	3.55 m/s
Velocity max	21.30 m/s
Velocity sd	3.39
Velocity var	11.48
Elevation mean	2300.55 m/s
Elevation median	2284.00 m/s
Elevation max	2474.00 m/s
Elevation sd	53.12
Elevation var	2821.38
Average segment length	8450.94 m
Percent flat	64.61%
Percent > 2%	13.24%
Percent > 4%	7.65%

Computed Statistics

Field Validation of the Cluster Factors

A route in Mexico City

Total GPS distance	59816.60 m
Odometer distance	56613.00 m
Max distance delta	1045.77 m
Max time delta	481.00 s
Total duration	17.84 h
Gradient mean	-0.01%
Gradient median	-0.03%
Gradient max	12.43%
Gradient sd	3.95
Gradient var	15.63
Velocity mean	4.38 m/s
Velocity median	3.55 m/s
Velocity max	21.30 m/s
Velocity sd	3.39
Velocity var	11.48
Elevation mean	2300.55 m/s
Elevation median	2284.00 m/s
Elevation max	2474.00 m/s
Elevation sd	53.12
Elevation var	2821.38
Average segment length	8450.94 m
Percent flat	64.61%
Percent > 2%	13.24%
Percent > 4%	7.65%

Computed Statistics

Field Validation of the Cluster Factors

Steep Hills on Route

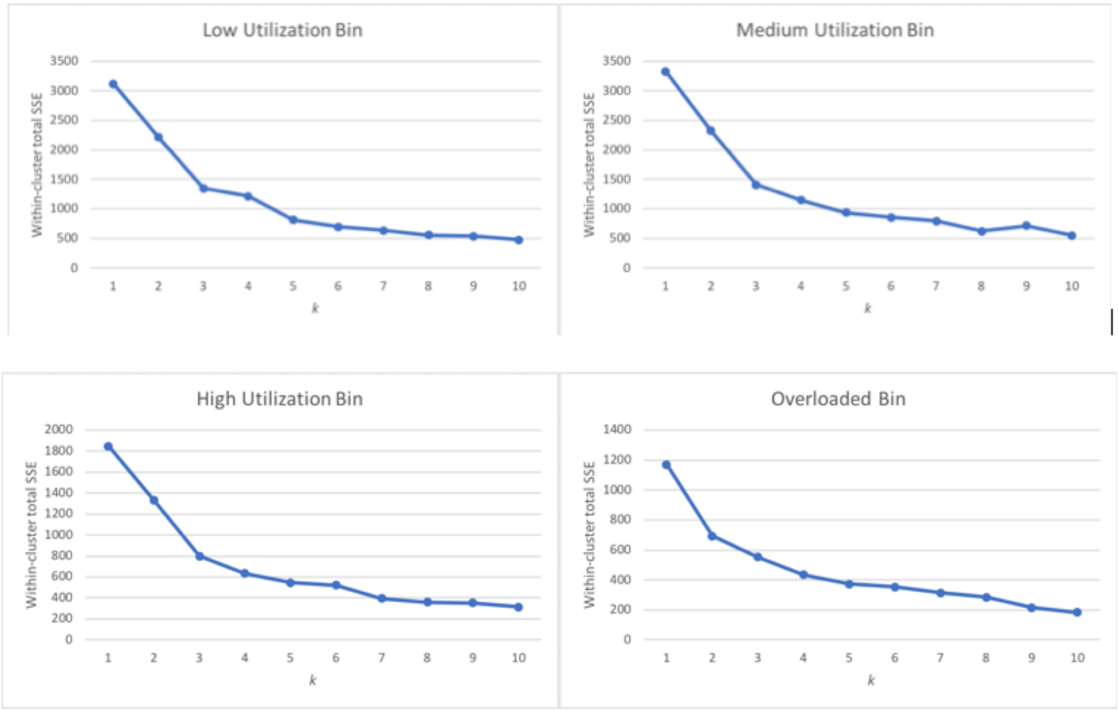
Field Validation of the Cluster Factors

Record #	Trip Time (s)	Cumulative Distance (m)	GPS Elevation (m)	Corrected Elevation (m)	GPS Velocity (m/s)	GPS Latitude	GPS Longitude	Estimated Gradient %
762	54587	41470.12	2322	2315	3.6	19.54	-99.24	9.70
763	54621	41654.84	2341	2330	6.9	19.54	-99.24	8.18
764	54646	41805.44	2349	2343	4.1	19.54	-99.24	8.24
765	54649	41820.28	2349	2342	5.8	19.54	-99.24	7.42
766	54650	41826.66	2350	2343	5.5	19.54	-99.24	7.32
767	54651	41832.02	2350	2343	4.7	19.54	-99.24	7.27
768	54653	41842.07	2351	2345	4.7	19.54	-99.24	7.78

Computed Slopes at Observed Location

Cluster Formation

- We used "*k*-means clustering" to group similar routes together
- Groupings are based on topographical features (e.g., road gradients) and traffic conditions (e.g., average speed) affecting fuel consumption
- The number of groups determined empirically, best = 4 groups



(Plots of "within-cluster SSE distances" to determine the best number of clusters)

Qualitative Analysis of Clusters

We classify four clusters to study similarities and differences in the routes *Selected k=4*

	Cluster A	Cluster B	Cluster C	Cluster D
Elevation	High	Low	Low	High
Topology	Hilly	Flat	Flat	Flat
Average Speed	Low	Medium	High	Low
Segment Lengths	Short	Medium	Long	Short

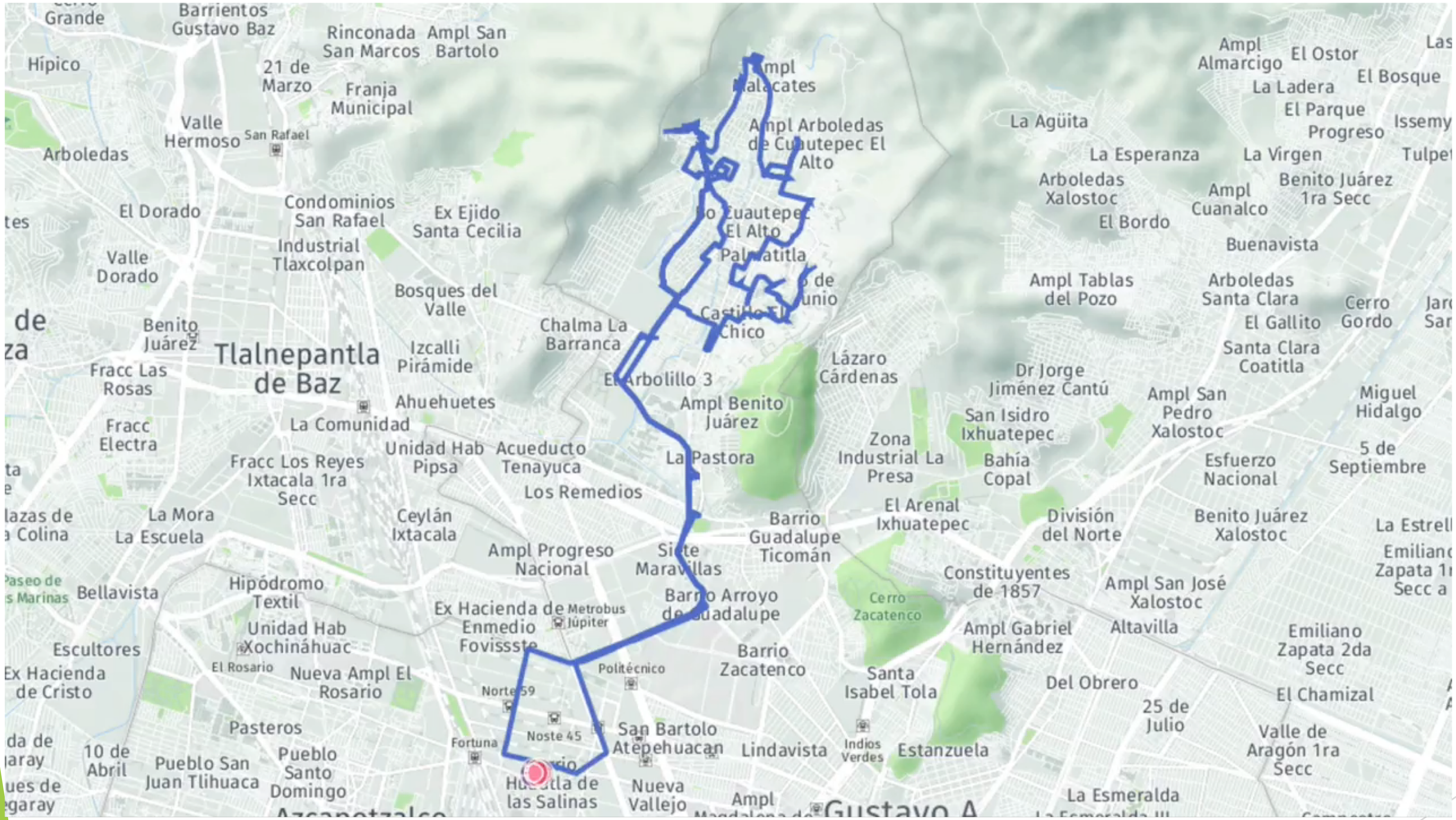
Cluster A primarily describes high altitude urban areas near Mexico City.

Cluster B denotes small and medium-sized cities with low elevation.

Cluster C is indicative of rural areas.

Cluster D mainly describes outskirts areas of Mexico City.

Cluster A example: a route in Mexico City



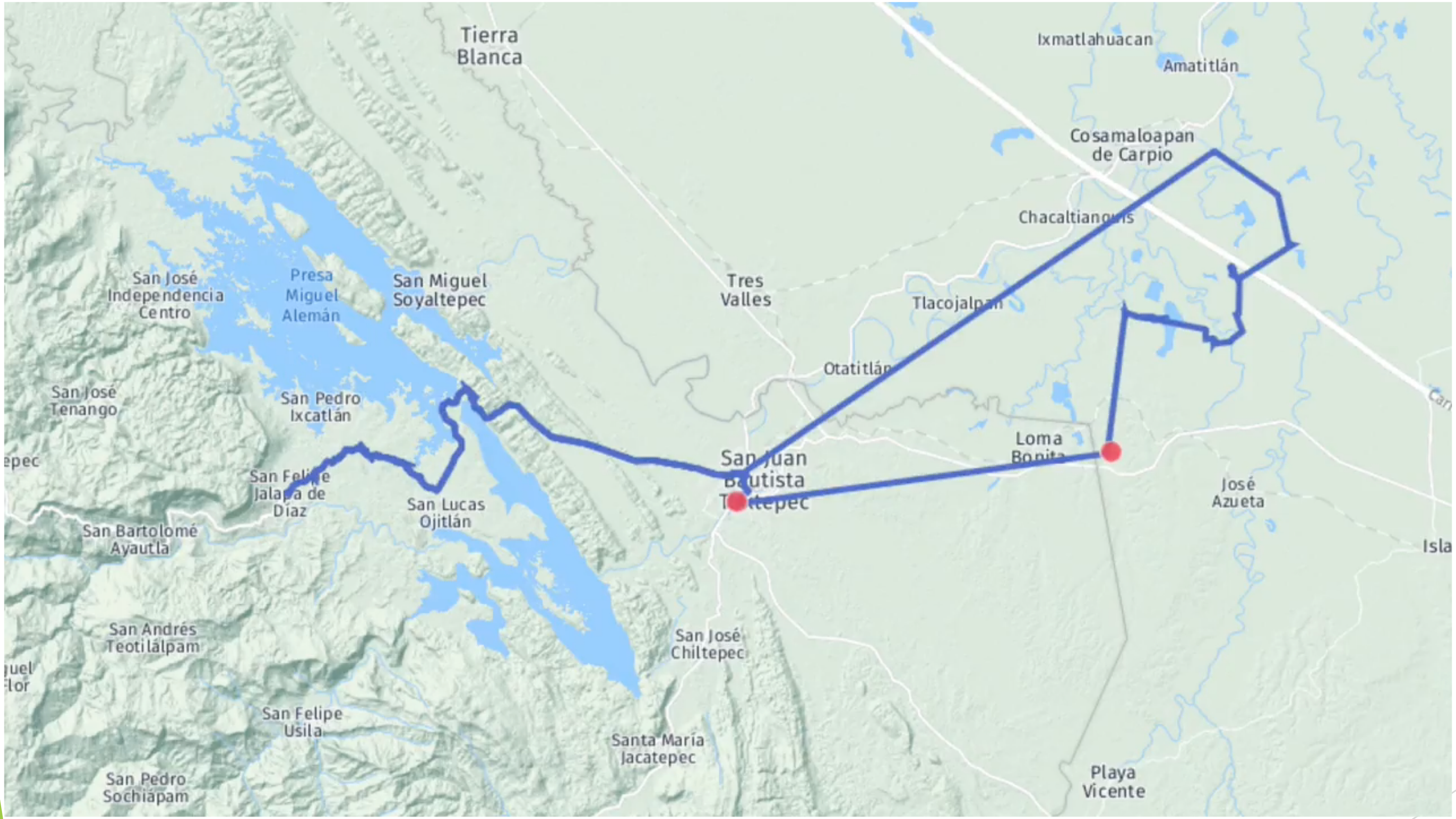
```
AZCP1804 2017-06-19 SUMMARY

Total GPS distance= 47391.20 m
Odometer distance= 43645.00 m
Max distance delta= 1248.37 m
Max time delta= 501.00 s
Total duration= 16.97 h
Gradient mean= 0.53%
Gradient median= 0.33%
Gradient max= 28.80%
Gradient sd= 7.96
Gradient var= 63.32
Velocity mean= 3.21 m/s
Velocity median= 2.90 m/s
Velocity max= 18.00 m/s
Velocity sd= 2.07
Velocity var= 4.29
Average segment length= 9382.81 m
Percent flat= 39.05%
```

Characteristics:

- High elevation, hilly
- Low average velocity
- Short segment lengths

Cluster C example: a route in Veracruz



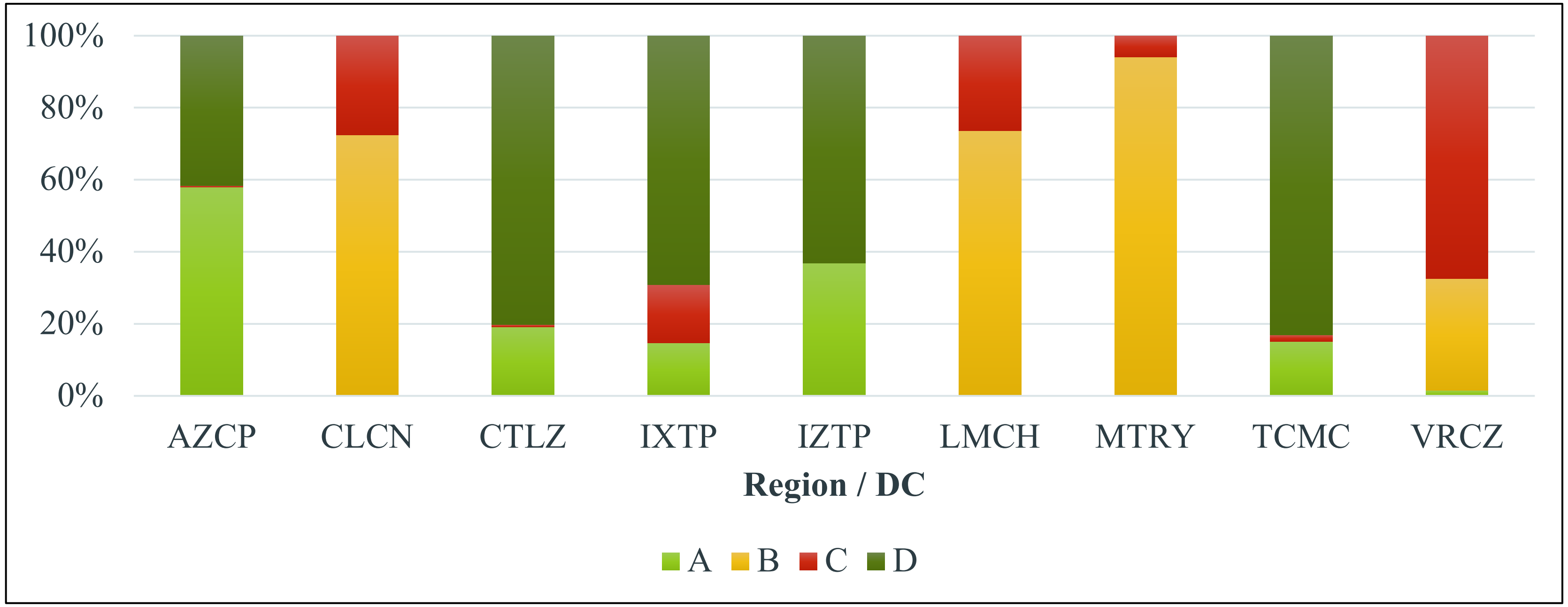
VRCZ4943 2017-11-23 SUMMARY

Total GPS distance= 278394.80 m
Odometer distance= 137519.00 m
Max distance delta= 52402.11 m
Max time delta= 8655.00 s
Total duration= 20.90 h
Gradient mean= 0.03%
Gradient median= -0.03%
Gradient max= 14.25%
Gradient sd= 3.36
Gradient var= 11.30
Velocity mean= 9.99 m/s
Velocity median= 10.30 m/s
Velocity max= 29.40 m/s
Velocity sd= 5.85
Velocity var= 34.23
Average segment length= 11529.34 m
Percent flat= 61.84%

Characteristics:

- Low elevation, flat
- High average velocity
- Long segment lengths

Distribution of clusters by region

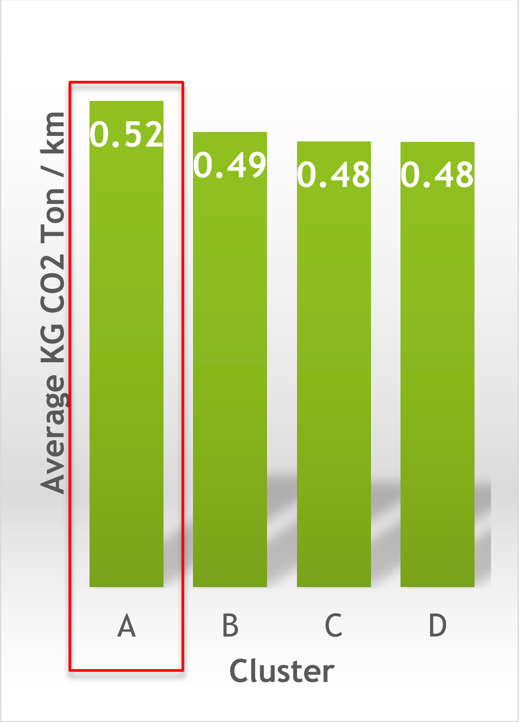


Most of the urban areas belong primarily to Clusters A and D, while Clusters B and C are more representative of suburban and rural regions.

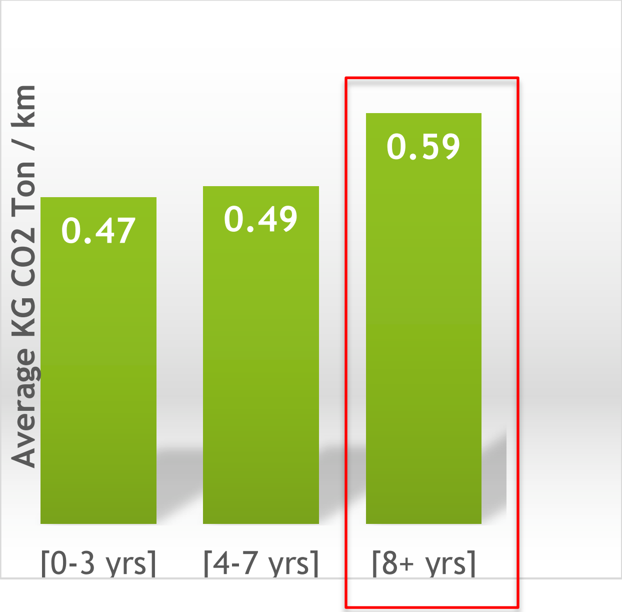
Ranking of Vehicles in Clusters

Medium Utilization Example

Clusters



Vehicle Age



Cluster A (high elevation, hilly, short segments, low velocity) has the greatest impact on CO_2 emissions, being approximately 10% larger than the other clusters

General Ranking - Emission Factor (Kg CO2 / Ton-Km)

9 vehicle types available

	Cluster A	Cluster B	Cluster C	Cluster D
Best Models	VehicleType5	VehicleType4	VehicleType1	VehicleType6
	VehicleType4	VehicleType1	VehicleType4	VehicleType1

Conclusions

Delivery routes can be meaningfully clustered based on factors such as topography and traffic conditions.

Some clusters demand more managerial attention compared to others.

For Coppel, Cluster A due shows an increased fuel consumption factor relative to the other clusters.

We can exchange vehicles between regions to reduce CO_2 consumption by assigning the best vehicle types to delivery areas.

In the case of Coppel, yielding a potential 7.2% reduction.

Replace your old vehicles.

We found older vehicles emit significantly more greenhouse gases than newer vehicles.


Q
&
A