

The International Transport Energy Modeling (iTEM) Open Data
Project
Rule Book

Sonia Yeh, Sandra Alnajjar, Dima Faraji, Hanna Habtemariam Morde, and Serah Rashmi (15-12-2023)
Humberto Linero, Sonia Yeh, and Paul Kishimoto (15-04-2020)

September 11, 2024

Recommended Citation:

International Transport Energy Modeling (iTEM) Group. (2023). The iTEM Open Data Project Rule Book. Zenodo. <https://doi.org/10.5281/zenodo.4121179>

Contents

General Information	3
How Tos	5
S012 De facto population (both sexes) in a country	6
T000 Passenger transport: Inland passenger transport	8
T001 Costal Transport	10
T002 Container Transport	11
T003 Inland Freight Transport	12
T004 New Road Vehicle Registrations by Vehicle Category and Fuel Type	14
T005 Direct CO2 emissions from global (and regional) transport scenarios	16
T006 Modal split of freight transport	18
T007 Modal split of passenger transport	19
T008 Passenger Road Vehicle Fleet and rate per thousand inhabitants by Vehicle Category	20
T009 Passenger Car	21
T010 Commercial Vehicle	22
T011 Aviation Total Passenger Kilometers	24
T012 Freight Transport - Railways	25
T013 Passengers Kilometer Travel - Roads	26
T014 LDV Sales	27
T015 Freight Transport - for Railways	28
T016 LDV Sales	29
T017 Freight Vehicle Registration	30
T018 Freight Transport - Tonne-km for Roads	31
T019 Freight Transport - Aviation (Domestic)	32
T020 Vehicle registration (Bus)	33
T021 Passengers Kilometer Travel - Roads	34
T022 Vehicle registration (LDV)	35
T023 Freight Transport -Aviation (Domestic)	36

General Information

This documentation contains detailed information of the **iTEM Open Database**, a harmonized transport data set of historical values, 1970 - 2018. It aims to create transparency through two key features:

Open-Data: Assembling a comprehensive collection of publicly-available transportation data

Open-Code: All code and documentation will be publicly accessible and open for modification and extension.

The iTEM Open Database is comprised of individual datasets collected from public sources. Each dataset is downloaded, cleaned, and harmonised to the common region and technology definitions defined by the iTEM consortium <https://transportenergy.org>. For each dataset, we describe the name of the dataset, the web link to the original source, the web link to the cleaning script (in python), variables, and explain the data cleaning steps (which explains the data cleaning script in plain English).

Nomenclature

- Datasets are numbered by the order they were collected and processed. Names that start with **T** stand for *Transport*, names that start with **S** stand for *sociodemographic*, and names that start with **A** stand for *Analysis* in which variables are derived for validation purposes. An "_iTEM" identifier is added to the end of the variable names indicating that **iTEM** is the data source, meaning that these variables are calculated by iTEM for validation instead of being collected from any of the original sources.

Definitions of regions

Unless otherwise specified, all the iTEM regions are obtained from the following file: <https://github.com/transportenergy/metadata/blob/master/model/regions.yaml>. The ISO code of each country is obtained according to the library *PyCountry*. However, certain countries in the dataset do not have the exact names as those appearing in the library; therefore, the section *Country and ISO Code* indicates what name is used for the countries that are not found in *PyCountry*.

Data quality flagging system

To enhance transparency and ensure the quality of our datasets, we have instituted a flagging system named **Data quality flag** within our database. Each dataset entry includes a **Data quality flag** column. A symbol "!" or "!!" in this column denotes the possibility of data quality issues. We detail below the various concerns this flag may indicate and strongly advise our users to consult the associated documentation.

Data quality flag explanation The presence of exclamation mark(s) "!" or "!!" in the **Data quality flag** column suggests potential data quality issues, which can encompass:

1. Tiered Confidence Levels:

- "!!" (Critical): The data's credibility is in question due to limited checks or substantial inconsistencies with other sources.
- "!" (Caution): The data appears plausible but is lacking in cross-verification or has been suggested for manual review.
- (Blank): The data aligns with established benchmarks or has been corroborated with trustworthy sources.

2. Data Source Reliability:

- Data derived from **Verified Sources** is deemed dependable and is not flagged.
- Data originating from **Trusted** and **Unverified Sources** may be marked with a "!" to recommend additional scrutiny.
- Data from **Suspect Sources** will be flagged with "!!".

3. Cross-Referencing Indicators:

- **Consistent with External Data** will not be flagged.
- **Partially Consistent** data will bear a "!".
- **Inconsistent** data will be flagged with "!!".
- **Not Cross-Referenced** data may be flagged with a "!" depending on the situation.

4. Automated Check Indicators:

- **Automated Check Passed** data will not carry a flag.
- **Manual Check Recommended** and **Automated Check Failed** data will be flagged with a "!!".
- **Not Checked** data may be flagged with a "!" as a cautionary step.

Community Feedback and Issue Reporting We highly regard our user community's involvement in upholding the quality of our datasets. Should any data quality issues arise or if you have any feedback, we urge you to report these instances through our GitHub issue tracker at <https://github.com/transportenergy/database/issues>. We welcome users to identify potential inaccuracies, which our team will examine and rectify in due course.

Version Control and Updates Our datasets undergo version control and are updated on a regular basis. We invite users to consult the version history to observe the continual enhancements and modifications.

How Tos

Forthcoming

How to navigate the Open Data

The input file used in each script is located at <https://github.com/transportenergy/metadata/tree/master/historical/input>. Detailed instructions on how to generate the latest iTEM Open Database (the merged file of individual datasets) is forthcoming.

How to navigate the Open Code

The scripts for cleaning the data is located at <https://github.com/transportenergy/database/tree/master/item/historical/scripts>.

S012

Information

- Dataset name: De facto population (both sexes) in a country as of 1 July of the year indicated
- Link to cleaning script: <https://github.com/transportenergy/database/blob/master/item/historical/scripts/S012.ipynb>

Source United Nations [https://population.un.org/wpp/Download/Files/1_Indicators%20\(Standard\)/EXCEL_FILES/1_Population/WPP2019_POP_F01_1_TOTAL_POPULATION_BOTH_SEXES.xlsx](https://population.un.org/wpp/Download/Files/1_Indicators%20(Standard)/EXCEL_FILES/1_Population/WPP2019_POP_F01_1_TOTAL_POPULATION_BOTH_SEXES.xlsx)

Country and ISO Code The following name changes were performed:

- Venezuela (Bolivarian Republic of) →Venezuela
- Holy See →Holy See (Vatican City State)
- China, Taiwan Province of China →Taiwan
- State of Palestine →Palestine
- Wallis and Futuna Islands →Wallis and Futuna
- Saint Helena →Saint Helena, Ascension and Tristan da Cunha
- United States Virgin Islands →Virgin Islands, U.S.
- Iran (Islamic Republic of) →Iran, Islamic Republic of
- Dem. People's Republic of Korea →Korea, Democratic People's Republic of
- Democratic Republic of the Congo →Congo, The Democratic Republic of the
- China, Macao SAR →Macao
- Bolivia (Plurinational State of) →Bolivia
- Republic of Korea →Korea, Republic of
- China, Hong Kong SAR →Hong Kong
- Micronesia (Fed. States of) →Micronesia, Federated States of

The only country we could not assign an ISO code was *Channel Islands*

ITEM Region The following countries were not assigned an ITEM region:

1. *Channel Islands*
2. *Saint Martin (French part)*
3. *Sint Maarten (Dutch part)*
4. *South Sudan*
5. *Bonaire, Sint Eustatius and Saba*
6. *Saint Barthelemy*
7. *Curaçao*

Variable *Population.*

Unit 10^3 *people.*

Service Null.

Mode Null.

Vehicle Type Null.

Technology Null.

Fuel : Null.

T000

Information

- Dataset name: Passenger transport: Inland passenger transport
- Link to cleaning script: <https://github.com/transportenergy/database/blob/master/item/historical/scripts/T000.py>

Source International Transport Forum <https://stats.oecd.org/index.aspx?queryid=79863>.

Country and ISO Code The following name changes were performed:

- Montenegro, Republic of →Montenegro
- Bosnia-Herzegovina →Bosnia and Herzegovina
- Korea →Korea, Republic of
- Serbia, Republic of →Serbia

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Passenger Activity*.

Unit The unit is changed from *Passenger-kilometres, Millions* to 10^9 *passenger-km / yr*.

Service *Passenger*.

Mode

- The mode for *Total inland passenger transport* is *All*.
- The mode for *Rail passenger transport* is *Rail*.
- The mode for *Road passenger transport by passenger cars* is *Road*.
- The mode for *Road passenger transport by buses and coaches* is *Road*.

Vehicle Type

- The Vehicle Type for *Total inland passenger transport* is *All*.
- The Vehicle Type for *Rail passenger transport* is *All*.
- The Vehicle Type for *Road passenger transport by passenger cars* is *LDV*.
- The Vehicle Type for *Road passenger transport by buses and coaches* is *Bus*.

Technology *All*.

Fuel *All*.

Data Cleaning

- The variable *Road Passenger Transport* is the sum of *Road passenger transport by passenger cars* and *Road passenger transport by buses and coaches*. In other words, Mode *Road Vehicle Type All* is the sum of Mode *Road Vehicle Type LDV* and Mode *Road Vehicle Type Bus*.
- There are 22 countries that have missing data for *Road passenger transport by passenger cars* or *Road passenger transport by buses and coaches* for certain years (we call it "problematic time periods" below), therefore the total sum *Road Passenger Transport* is incorrectly reported. Below are the rules on how we handle these cases:
 - Albania: Remove *Road passenger transport* (Mode *Road Vehicle Type All*) & *Road passenger transport by buses and coaches* (Mode *Road Vehicle Type Bus*) during the problematic time periods.
 - Armenia: Remove *Road passenger transport* (Mode *Road Vehicle Type All*) during the problematic time periods.
 - Azerbaijan: Remove *Road passenger transport* (Mode *Road Vehicle Type All*) during the problematic time periods.
 - Belarus: Remove *Road passenger transport* (Mode *Road Vehicle Type All*) during the problematic time periods.
 - Bulgaria: Remove *Road passenger transport* (Mode *Road Vehicle Type All*) during the problematic time periods.
 - Canada: Remove *Road passenger transport* (Mode *Road Vehicle Type All*) during the problematic time periods.
 - Russian Federation: Remove *Road passenger transport* (Mode *Road Vehicle Type All*) during the problematic time periods.
 - Switzerland: Remove *Road passenger transport* (Mode *Road Vehicle Type All*) during the problematic time periods.
 - United States: Remove *Road passenger transport* (Mode *Road Vehicle Type All*) during the problematic time periods.

T001

Information

- Dataset name: Coastal Transport
- Link to cleaning script: <https://github.com/transportenergy/database/blob/master/item/historical/scripts/T001.py>

Source International Transport Forum https://stats.oecd.org/Index.aspx?DataSetCode=ITF_GOODS_TRANSPORT#.

Country and ISO Code The following name changes were performed:

- Montenegro, Republic of →Montenegro
- Korea →Korea, Republic of
- Serbia, Republic of →Serbia

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Freight Activity*

Unit The unit is changed from *Tonnes-Kilometer* to 10^9 *tonne-km / yr*.

Service *Freight*

Mode *Shipping*

Vehicle Type *Coastal*

Technology *All*

Fuel *All*

T002

Information

- Dataset name: Container Transport
- Link to cleaning script: <https://github.com/transportenergy/database/blob/master/item/historical/scripts/T002.ipynb>

Source International Transport Forum https://stats.oecd.org/Index.aspx?DataSetCode=ITF_GOODS_TRANSPORT#.

Country and ISO Code The following name changes were performed:

- Korea →Korea, Republic of

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable The variable is set to either *Freight (TEU)* or *Freight (Weight)*.

Unit

- *Freight (TEU)* is *Number*
- *Freight (Weight)* is 10^3 *tonne / yr*

Service *Freight*

Mode

- The mode for *Rail containers transport (TEU)* is *Rail*
- The mode for *Maritime containers transport (weight)* is *Shipping*

Vehicle Type *Container*

Technology *All*

Fuel *All*

T003

Information

- Dataset name: Inland Freight Transport
- Link to cleaning script: <https://github.com/transportenergy/database/blob/master/item/historical/scripts/T003.py>

Source International Transport Forum https://stats.oecd.org/Index.aspx?DataSetCode=ITF_GOODS_TRANSPORT#.

Country and ISO Code The following name changes were performed:

- Montenegro, Republic of → Montenegro
- Bosnia-Herzegovina → Bosnia and Herzegovina
- Korea → Korea, Republic of
- Serbia, Republic of → Serbia

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Freight Activity*

Unit The unit is changed from *Million Tonnes-kilometers* to 10^9 *tonne-km / yr.*

Service

- The service for *Road freight transport on own account* is *Freight*.
- The service for *Inland waterways freight transport* is *Freight*.
- The service for *Rail freight transport* is *Freight*.
- The service for *Road freight transport* is *Freight*.
- The service for *Road freight transport for hire and reward* is *Freight*.
- The service for *Total inland freight transport* is *Freight*.
- The service for *Pipelines transport* is *Pipeline*.

Mode

- The mode for *Road freight transport* is *Road*.
- The mode for *Road freight transport for hire and reward* is *Road*.
- The mode for *Road freight transport on own account* is *Road*.
- The mode for *Rail freight transport* is *Rail*.
- The mode for *Pipelines transport* is *Pipeline*.
- The mode for *Inland waterways freight transport* is *Shipping*.
- The mode for *Total inland freight transport* is *Inland*.

We also created a new Mode called *Inland (exl. Pipeline)*, which is the result represent the sum of all services except *Pipeline*.

Vehicle Type

- The vehicle type for *Road freight transport* is *All*.
- The vehicle type for *Road freight transport for hire and reward* is *For Hire and Reward*.
- The vehicle type for *Road freight transport on own account* is *For Own Account*.
- The vehicle type for *Rail freight transport* is *All*.
- The vehicle type for *Pipelines transport* is *Pipeline*.
- The vehicle type for *Inland waterways freight transport* is *Inland*.
- The vehicle type for *Total inland freight transport* is *All*.
- The vehicle type for *Inland (exl. Pipeline)* is *All*.

Technology *All*

Fuel *All*

T004

Information

- Dataset name: New Road Vehicle Registrations by Vehicle Category and Fuel Type
- Link to cleaning script: <https://github.com/transportenergy/database/blob/master/item/historical/scripts/T004.py>

Source United Nations Economic Commission for Europe https://datasource.kapsarc.org/explore/dataset/new-road-vehicle-registrations-by-vehicle-category-and-fuel-type/export/?disjunctive.country_name&disjunctive.date&disjunctive.frequency&disjunctive.fuel_type_name&disjunctive.type_of_vehicle_name.

Country and ISO Code The following name changes were performed:

- The former Yugoslav Republic of Macedonia →North Macedonia

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Sales (New Vehicles)*

Unit $10^6 \text{ vehicle / yr}$

Service

- The service for *New lorries (vehicle wt over 3500 kg)* is *Freight*
- The service for *New road tractors* is *Freight*
- The service for *New passenger cars* is *Passenger*
- The service for *New motor coaches, buses and trolley buses* is *Freight*
- The service for *New light goods vehicles* is *Freight*

Mode *Road*

Vehicle Type

- The Vehicle Type for *New lorries (vehicle wt over 3500 kg)* is *Heavy Truck*
- The Vehicle Type for *New road tractors* is *Medium Truck*
- The Vehicle Type for *New passenger cars* is *LDV*
- The Vehicle Type for *New motor coaches, buses and trolley buses* is *Bus*
- The Vehicle Type for *New light goods vehicles* is *Light Truck*

Technology

- The Technology for *LPG* is *Conventional*
- The Technology for *Compressed natural gas (CNG)* is *Natural Gas Vehicle*
- The Technology for *Liquefied natural gas (LNG)* is *Natural Gas Vehicle*
- The Technology for *Bioethanol* is *Conventional*
- The Technology for *Bi-fuel vehicles* is *Conventional*
- The Technology for *Biodiesel* is *Conventional*
- The Technology for *Diesel (excluding hybrids)* is *Conventional*
- The Technology for *Hybrid electric-diesel* is *Conventional*
- The Technology for *Hybrid electric-petrol* is *Conventional*
- The Technology for *Diesel* is *Conventional*
- The Technology for *Petrol* is *Conventional*
- The Technology for *Petrol (excluding hybrids)* is *Conventional*
- The Technology for *Plug-in hybrid diesel-electric* is *PHEV*
- The Technology for *Plug-in hybrid petrol-electric* is *PHEV*
- The Technology for *Hydrogen and fuel cells* is *Fuel Cell*
- The Technology for *Electricity* is *BEV*
- The Technology for *Total* is *All*
- The Technology for *Alternative (total)* is *Alternative*

Fuel

- The Fuel for *LPG* is *Liquid - Fossil*
- The Fuel for *Compressed natural gas (CNG)* is *Natural gas*
- The Fuel for *Liquefied natural gas (LNG)* is *Natural gas*
- The Fuel for *Bioethanol* is *Liquid-Bio*
- The Fuel for *Bi-fuel vehicles* is *Liquid-Bio*
- The Fuel for *Biodiesel* is *Liquid-Bio*
- The Fuel for *Diesel (excluding hybrids)* is *Liquid - Fossil*
- The Fuel for *Hybrid electric-diesel* is *Liquid - Fossil*
- The Fuel for *Hybrid electric-petrol* is *Liquid - Fossil*
- The Fuel for *Diesel* is *Liquid - Fossil*
- The Fuel for *Petrol* is *Liquid - Fossil*
- The Fuel for *Petrol (excluding hybrids)* is *Liquid - Fossil*
- The Fuel for *Plug-in hybrid diesel-electric* is *Electricity*
- The Fuel for *Plug-in hybrid petrol-electric* is *Electricity*
- The Fuel for *Hydrogen and fuel cells* is *Hydrogen*
- The Fuel for *Electricity* is *Electricity*
- The Fuel for *Total* is *All*
- The Fuel for *Alternative (total)* is *Alternative*

T005

Information

- Dataset name: Direct CO2 emissions from global (and regional) transport scenarios
- Link to cleaning script: <https://github.com/transportenergy/database/blob/master/item/historical/scripts/T005.ipynb>

Source Joint Research Center https://edgar.jrc.ec.europa.eu/overview.php?v=50_GHG.

Country and ISO Code The following name changes were performed:

- Swaziland →Eswatini
- Saint Helena →Saint Helena, Ascension and Tristan da Cunha
- Libyan Arab Jamahiriya →Libya
- Congo_the Democratic Republic of the →Congo, The Democratic Republic of the
- Reunion →Réunion
- Int. Aviation →World
- Int. Shipping →World
- Virgin Islands_British →Virgin Islands, British
- Cote d'Ivoire →Côte d'Ivoire
- Taiwan_Province of China →Taiwan, Province of China
- Cape Verde →Cabo Verde
- Tanzania_United Republic of →Tanzania, United Republic of
- The former Yugoslav Republic of Macedonia →North Macedonia

ITEM Region To the following countries we assigned the ITEM region manually:

- Serbia and Montenegro →SCG
- World →WLD
- Netherlands Antilles →ANT

Variable *CO2 Emission (ttw)*

Unit $10^6 t CO_2 / yr$

Service *All*

Mode The mapping done for countries is the following:

- The mode for *Railways* is *Rail*
- The mode for *Road Transportation* is *Road*
- The mode for *Civil Aviation* is *Air*
- The mode for *Other Transportation* is *Other*
- The mode for *Water-borne Navigation* is *Shipping*

The mapping done for the Int. Aviation country is the following:

- The mode for *Civil Aviation* is *Domestic Aviation*

The mapping done for the Int. Shipping country is the following:

- The mode for *Water-Borne Navigation* is *Domestic Shipping*

Vehicle Type *All*

Technology *All*

Fuel *All*

T006

Information

- Dataset name: Modal split of freight transport
- Link to cleaning script: <https://github.com/transportenergy/database/blob/master/item/historical/scripts/T006.ipynb>

Source Eurostat https://datasource.kapsarc.org/explore/dataset/modal-split-of-freight-transport/information/?disjunctive.date&disjunctive.frequency&disjunctive.geo_name&disjunctive.measure_name&disjunctive.tra_mode_name.

Country and ISO Code The following name changes were performed:

- European Union (current composition) →EU28

ITEM Region To the following countries, the ITEM region was assigned manually as follows:

- European Union (current composition) →EU-28

Variable *Freight Activity*

Unit *% tonne-kilometres / yr*

Service *Freight*

Mode

- The mode for *Railways* is *Rail*
- The mode for *Roads* is *Road*
- The mode for *Inland waterways* is *Shipping*

Vehicle Type

- The mode for *Railways* is *All*
- The mode for *Roads* is *All*
- The mode for *Inland waterways* is *Inland Waterway*

Technology *All*

Fuel *All*

T007

Information

- Dataset name: Modal split of passenger transport
- Link to cleaning script: <https://github.com/transportenergy/database/blob/master/item/historical/scripts/T007.ipynb>

Source Eurostat https://datasource.kapsarc.org/explore/dataset/modal-split-of-passenger-transport/export/?disjunctive.date&disjunctive.frequency&disjunctive.geo_name&disjunctive.measure_name&disjunctive.vehicle_name.

Country and ISO Code The following name changes were performed:

- European Union (28 countries) →EU28
- European Union (27 countries) →EU27
- The former Yugoslav Republic of Macedonia →North Macedonia

ITEM Region To the following countries, the ITEM region was assigned manually as follows:

- EU27 →EU-27
- EU28 →EU-28

Variable *Passenger Activity*

Unit *% in total inland passenger-km / yr*

Service *Passenger*

Mode

- The mode for *Trains* is *Rail*
- The mode for *Passenger cars* is *Road*
- The mode for *Motor coaches, buses and trolley buses* is *Road*

Vehicle Type

- The vehicle type for *Trains* is *All*
- The vehicle type for *Passenger cars* is *LDV*
- The vehicle type for *Motor coaches, buses and trolley buses* is *Bus*

Technology *All*

Fuel *All*

T008

Information

- Dataset name: Passenger Road Vehicle Fleet and rate per thousand inhabitants by Vehicle Category
- Link to cleaning script: <https://github.com/transportenergy/database/blob/master/item/historical/scripts/T008.ipynb>

Source United Nations Economic Commission for Europe https://datasource.kapsarc.org/explore/dataset/passenger-road-vehicle-fleet-and-rate-per-thousand-inhabitants-by-vehicle-category/information/?disjunctive.country_name&disjunctive.date&disjunctive.frequency&disjunctive.measurement_name&disjunctive.vehicle_category_name.

Country and ISO Code All countries have an assigned ISO code.

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Stock*

Unit For the given variable there are two units:

- 10^6 *vehicle*.
- *Vehicles per 1000 inhabitants*.

Service *Passenger*

Mode *Road*

Vehicle Type

- The Vehicle Type for *Special purpose vehicles* is *Special purpose vehicles*
- The Vehicle Type for *Passenger cars* is *LDV*
- The Vehicle Type for *Trams* is *Trams*
- The Vehicle Type for *Motorcycles* is *Motorcycles*
- The Vehicle Type for *Motor coaches, buses and trolley bus* is *Bus*
- The Vehicle Type for *Mopeds* is *Mopeds*

Technology *All*

Fuel *All*

T009

Information

- Dataset name: Passenger Car
- Link to cleaning script: <https://github.com/transportenergy/database/blob/master/item/historical/scripts/T009.py>

Source International Organization of Motor Vehicle Manufacturers
<http://www.oica.net/category/vehicles-in-use/>

Country and ISO Code The following name changes were performed:

- RUSSIA →Russian Federation
- SYRIA →Syrian Arab Republic
- IRAN →Iran, Islamic Republic of
- BOSNIA →Bosnia and Herzegovina
- HONG-KONG →Hong Kong
- IVORY COAST →Côte d'Ivoire
- BRUNEI →Brunei Darussalam
- MOLDAVIA →Moldova, Republic of
- SOUTH KOREA →Korea, Republic of
- CONGO KINSHASA →Congo, The Democratic Republic of the
- PALESTINE →Palestine, State of
- MACEDONIA →North Macedonia

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Stock*

Unit 10^6 *vehicle*

Service *Passenger*

Mode *Road*

Vehicle Type *All*

Technology *All*

Fuel *All*

T010

Information

- Dataset name: Commercial Vehicle
- Link to cleaning script: <https://github.com/transportenergy/database/blob/master/item/historical/scripts/T010.ipynb>

Source

International Organization of Motor Vehicle Manufacturers
<http://www.oica.net/category/vehicles-in-use/>

Country and ISO Code

The following name changes were performed:

- RUSSIA →Russian Federation
- SYRIA →Syrian Arab Republic
- IRAN →Iran, Islamic Republic of
- BOSNIA →Bosnia and Herzegovina
- HONG-KONG →Hong Kong
- IVORY COAST →Côte d'Ivoire
- BRUNEI →Brunei Darussalam
- MOLDAVIA →Moldova, Republic of
- SOUTH KOREA →Korea, Republic of
- CONGO KINSHASA →Congo, The Democratic Republic of the
- PALESTINE →Palestine, State of
- MACEDONIA →North Macedonia

The only country we could not assign an ISO code was Azerbaijan.

ITEM Region All countries in the dataset are assigned to an iTEM region, except for Azerbaijan.

Variable *Stock*

Unit 10^6 *vehicle*

Service *Freight*

Mode Road

Vehicle Type *All*

Technology *All*

Fuel *All*

T011

Information

- Dataset name: Aviation Total Passenger Kilometers
- Description: A passenger-kilometre, abbreviated as pkm, is the unit of measurement representing the transport of one passenger by a defined mode of transport i.e. aviation over one kilometre. It includes both domestic and international.
- Link to cleaning script: [https://github.com/linero-tech/iteminternship/blob/main/code/dima/T011_TAS-PAT-017\(1\).ipynb](https://github.com/linero-tech/iteminternship/blob/main/code/dima/T011_TAS-PAT-017(1).ipynb)

Source Data was compiled from the Asian Transport Outlook Database (ATO2023 TAS-PAT-017(1)), accessible at <https://data.adb.org/dataset/asian-transport-outlook-database>. The original data was sourced from Country Official Statistics. The dataset was downloaded on November 6, 2023.

Country and ISO Code All countries have an assigned ISO code.

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Passenger Activity*

Unit The unit is changed from *Million passenger kilometers* to 10^9 *passenger-km / yr*.

Service *Aviation*

Mode *Aviation*

Vehicle Type *All*

Technology *All*

Fuel *All*

Data Cleaning The following column names are removed from the datasets

- Remarks
- Source (2022-04)
- Source (2021-10)
- Source (2023-03)

Data Quality flag "!" (Caution): The data appears plausible but is lacking in cross-verification or has been suggested for manual review.

T012

Information

- Dataset name: Freight Transport - Tonne-km for Railways
- Description: A tonne-kilometre, is a unit of measure of freight transport which represents the transport of one tonne of goods (including packaging and tare weights of intermodal transport units) by railways over a distance of one kilometre. Only the distance on the national territory of the reporting country is taken into account for national, international and transit transport.
- Link to cleaning script: [https://github.com/linero-tech/item-internship/blob/main/code/dima/T012_TAS-FRA-005\(2\).ipynb](https://github.com/linero-tech/item-internship/blob/main/code/dima/T012_TAS-FRA-005(2).ipynb)

Source Data was compiled from the Asian Transport Outlook Database (ATO2023 TAS-FRA-005(2)), accessible at <https://data.adb.org/dataset/asian-transport-outlook-database>. The original data was sourced from International Union of Railways (UIC) <https://uic.org/special-groups/raildata/>. The dataset was downloaded on November 6, 2023.

Country and ISO Code All countries have an assigned ISO code.

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Freight Activity*

Unit The unit is changed from *Million tonne kilometers* to 10^9 *tonne-km / yr*.

Service *Freight*

Mode *Rail*

Vehicle Type *All*

Technology *All*

Fuel *All*

Data Quality flag "!!" (Caution): Inconsistent with External Data.

T013

Information

- Dataset name: Passengers Kilometer Travel - Roads
- Description: A passenger-kilometre, abbreviated as pkm, is the unit of measurement representing the transport of one passenger by a defined mode of transport i.e. roads over one kilometre.
- Link to cleaning script: [https://github.com/linero-tech/item-internship/blob/main/code/dima/T013_TAS-PAG-005\(2\).ipynb](https://github.com/linero-tech/item-internship/blob/main/code/dima/T013_TAS-PAG-005(2).ipynb)

Source Data was compiled from the Asian Transport Outlook Database (ATO2023 TAS-PAG-005(2)), accessible at <https://data.adb.org/dataset/asian-transport-outlook-database>. The original data was sourced from World Bank (WB). The dataset was downloaded on November 6, 2023.

Country and ISO Code All countries have an assigned ISO code.

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Passenger Activity*

Unit The unit is changed from *Million passenger kilometers* to 10^9 *passenger-km / yr*.

Service *Passenger*

Mode *Road*

Vehicle Type *All*

Technology *All*

Fuel *All*

Data Quality flag "!" (Caution): The data appears plausible but is lacking in cross-verification or has been suggested for manual review.

T014

Information

- Dataset name: LDV Sales
- Description: Sales data for LDV. LDV refers to road motor vehicle, other than two and three wheelers intended for the carriage of passengers and designed to seat no more than nine persons (including the driver). Refers to category M1 of the UN Consolidated Resolution on the Construction of Vehicles (R.E.3). It includes all types of fuel.
- Link to cleaning script: [https://github.com/linero-tech/item-internship/blob/main/code/dima/T014_TAS-VEP-005\(2\).ipynb](https://github.com/linero-tech/item-internship/blob/main/code/dima/T014_TAS-VEP-005(2).ipynb)

Source Data was compiled from the Asian Transport Outlook Database (ATO2023 TAS-VEP-005(2)), accessible at <https://data.adb.org/dataset/asian-transport-outlook-database>. The original data was sourced from Carsalesbase <https://carsalesbase.com/>. The dataset was downloaded on November 6, 2023.

Country and ISO Code All countries have an assigned ISO code.

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Sales (New Vehicles)*

Unit The unit is changed from *Number* to 10^6 *vehicle / yr*.

Service *Passenger*

Mode *Road*

Vehicle Type *LDV*

Technology *All*

Fuel *All*

Data Quality flag "!" (Caution): The data appears plausible but is lacking in cross-verification or has been suggested for manual review.

T015

Information

- Dataset name: Freight Transport - Tonne-km for Railways
- Description: A tonne-kilometre, is a unit of measure of freight transport which represents the transport of one tonne of goods (including packaging and tare weights of intermodal transport units) by railways over a distance of one kilometre. Only the distance on the national territory of the reporting country is taken into account for national, international and transit transport.
- Link to cleaning script: [https://github.com/linero-tech/item-internship/blob/main/code/hanna/T015_TAS-FRA-005\(3\).ipynb](https://github.com/linero-tech/item-internship/blob/main/code/hanna/T015_TAS-FRA-005(3).ipynb)

Source Data was compiled from the Asian Transport Outlook Database (ATO2023 TAS-FRA-005(3)), accessible at <https://data.adb.org/dataset/asian-transport-outlook-database>. The original data was sourced from Country Official Statistics . The dataset was downloaded on November 6, 2023.

Country and ISO Code All countries have an assigned ISO code.

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Freight Activity*

Unit The unit is changed from *Million tonne kilometers* to 10^9 *tonne-km / yr*.

Service *Freight*

Mode *Rail*

Vehicle Type *All*

Technology *All*

Fuel *All*

Data Cleaning The following column names are removed from the datasets

- Remarks
- Source (2022-04)
- Source (2021-10)
- Source (2023-03)

Data Quality flag "!" (Caution): The data appears plausible but is lacking in cross-verification or has been suggested for manual review.

T016

Information

- Dataset name: LDV Sales
- Description: Sales data for LDV. LDV refers to road motor vehicle, other than two and three wheelers intended for the carriage of passengers and designed to seat no more than nine persons (including the driver). Refers to category M1 of the UN Consolidated Resolution on the Construction of Vehicles (R.E.3). It includes all types of fuel.
- Link to cleaning script: [https://github.com/linero-tech/item-internship/blob/main/code/hanna/T016_TAS-VEP-005\(1\).ipynb](https://github.com/linero-tech/item-internship/blob/main/code/hanna/T016_TAS-VEP-005(1).ipynb).

Source Data was compiled from the Asian Transport Outlook Database (ATO2023 TAS-VEP-005(1), accessible at <https://data.adb.org/dataset/asian-transport-outlook-database>. The original data was sourced from OICA <http://www.oica.net/category/sales-statistics/>. The dataset was downloaded on November 6, 2023.

Country and ISO Code All countries have an assigned ISO code.

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Sales (New Vehicles)*

Unit The unit is changed from *Number* to 10^6 *vehicle/year*.

Service *Passenger*

Mode *Road*

Vehicle Type *LDV*

Technology *All*

Fuel *All*

Data Quality flag "!" (Caution): The data appears plausible but is lacking in cross-verification or has been suggested for manual review.

T017

Information

- Dataset name: Freight Vehicle Registration
- Description: We report cumulative number of goods vehicle registrations i.e. LCVs and trucks on road.
- Link to cleaning script: https://github.com/linero-tech/item-internship/blob/main/code/hanna/T017_TAS-VEP-020.ipynb

Source Data was compiled from the Asian Transport Outlook Database (ATO2023 TAS-VEP-020), accessible at <https://data.adb.org/dataset/asian-transport-outlook-database>. The original data was sourced from Country Official Statistics . The dataset was downloaded on November 6, 2023.

Country and ISO Code All countries have an assigned ISO code.

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Stock*

Unit The unit is changed from *Number* to 10^6 *vehicle/year*.

Service *Freight*

Mode *Road*

Vehicle Type *All*

Technology *All*

Fuel *All*

Data Quality flag "!" (Caution): The data appears plausible but is lacking in cross-verification or has been suggested for manual review.

T018

Information

- Freight Transport - Tonne-km for Roads
- Description: A tonne-kilometre, is a unit of measure of freight transport which represents the transport of one tonne of goods (including packaging and tare weights of intermodal transport units) by roads over a distance of one kilometre. Only the distance on the national territory of the reporting country is taken into account for national, international and transit transport.
- Link to cleaning script: [https://github.com/linero-tech/item-internship/blob/main/code/serah/T018_TAS-FRA-004\(2\).ipynb](https://github.com/linero-tech/item-internship/blob/main/code/serah/T018_TAS-FRA-004(2).ipynb)

Source Data was compiled from the Asian Transport Outlook Database (ATO2023 TAS-FRA-004(2)), accessible at <https://data.adb.org/dataset/asian-transport-outlook-database>. The original data was sourced from Country Official Statistics . The dataset was downloaded on November 6, 2023.

Country and ISO Code All countries have an assigned ISO code.

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Freight Activity*

Unit The unit is changed from *Million tonne kilometers* to 10^9 *tonne-km / yr*.

Service *Freight*

Mode *Road*

Vehicle Type *All*

Technology *All*

Fuel *All*

Data Cleaning The following column names are removed from the datasets

- Remarks
- Source (2022-04)
- Source (2021-10)
- Source (2023-03)

Data Quality flag "!" (Caution): The data appears plausible but is lacking in cross-verification or has been suggested for manual review.

T019

Information

- Dataset name: Freight Transport - Tonne-km for Aviation (Domestic)
- Description: Domestic tonne-kilometre, is a unit of measure of freight transport which represents the transport of one tonne of goods (including packaging and tare weights of intermodal transport units) by a given transport mode (road, rail, air, sea, inland waterways, pipeline etc.) over a distance of one kilometre. Only the distance on the national territory of the reporting country is taken into account for national, international and transit transport.
- Link to cleaning script: [https://github.com/linero-tech/item-internship/blob/main/code/serah/T019_TAS-FRA-007\(3\).ipynb](https://github.com/linero-tech/item-internship/blob/main/code/serah/T019_TAS-FRA-007(3).ipynb)

Source Data was compiled from the Asian Transport Outlook Database (ATO2023 TAS-FRA-007(3)), accessible at <https://data.adb.org/dataset/asian-transport-outlook-database>. The original data was sourced from Country Official Statistics . The dataset was downloaded on November 6, 2023.

Country and ISO Code All countries have an assigned ISO code.

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Freight Activity*

Unit The unit is changed from *Million tonne kilometers* to 10^9 *tonne-km / yr*.

Service *Freight*

Mode The mode is changed from *Aviation* to *Aviation (Domestic)*.

Vehicle Type *All*

Technology *All*

Fuel *All*

Data Cleaning The following column names are removed from the datasets

- Remarks
- Source (2022-04)
- Source (2021-10)
- Source (2023-03)

Data Quality flag "!" (Caution): The data appears plausible but is lacking in cross-verification or has been suggested for manual review.

T020

Information

- Dataset name: Vehicle registration (Bus)
- Description: We report cumulative number of vehicle registrations i.e. Bus on road. Statistics on domestic production of Bus which includes both mini bus and conventional bus. The conventional bus is Passenger road motor vehicle designed to carry more than 24 persons (including the driver), and with provision to carry seated as well as standing passengers. Refers to class I and class II of categories M2 and M3 of the UN Consolidated Resolution on the Construction of Vehicles (R.E.3). The mini bus is a Passenger road motor vehicle designed to carry 10-23 seated or standing persons (including the driver). Refers to class A and class B of categories M2 and M3 of the UN Consolidated Resolution on the Construction of Vehicles (R.E.3).
- Link to cleaning script: https://github.com/linero-tech/item-internship/blob/main/code/serah/T020_TAS-VEP-018.ipynb

Source Data was compiled from the Asian Transport Outlook Database (ATO2023 TAS-VEP-018), accessible at <https://data.adb.org/dataset/asian-transport-outlook-database>. The original data was sourced from Country Official Statistics . The dataset was downloaded on November 6, 2023.

Country and ISO Code All countries have an assigned ISO code.

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Stock*

Unit The unit is changed from *Number* to 10^6 *vehicle*.

Service *Passenger*

Mode *Road*

Vehicle Type *Bus*

Technology *All*

Fuel *All*

Data Cleaning The following column names are removed from the datasets

- Remarks
- Source (2022-04)
- Source (2021-10)
- Source (2023-03)

Data Quality flag "!" (Caution): The data appears plausible but is lacking in cross-verification or has been suggested for manual review.

T021

Information

- Dataset name: Passengers Kilometer Travel - Roads
- Description: A passenger-kilometre, abbreviated as pkm, is the unit of measurement representing the transport of one passenger by a defined mode of transport i.e. roads over one kilometre. Data may or may not include public transport and private transport.
- Link to cleaning script: [https://github.com/linero-tech/item-internship/blob/main/code/sandra/T021_TAS_PAG_005\(3\).ipynb](https://github.com/linero-tech/item-internship/blob/main/code/sandra/T021_TAS_PAG_005(3).ipynb)

Source Data was compiled from the Asian Transport Outlook Database (ATO2023 TAS-PAG-005(3)), accessible at <https://data.adb.org/dataset/asian-transport-outlook-database>. The original data was sourced from Country Official Statistics . The dataset was downloaded on November 6, 2023.

Country and ISO Code All countries have an assigned ISO code.

ITEM Region All countries in the dataset are assigned to an iTEM region.
subsection*Variable *Passenger Activity*

Unit The unit is changed from *Million passenger kilometers* to 10^9 *passenger-km / yr*.

Service *Passenger*

Mode *Road*

Vehicle Type *All*

Technology *All*

Fuel *All*

Data Quality flag "!" (Caution): The data appears plausible but is lacking in cross-verification or has been suggested for manual review.

T022

Information

- Dataset name: Vehicle registration (LDV)
- Description: We report cumulative number of vehicle registrations i.e. LDVs on road. LDV refers to road motor vehicle, other than two and three wheelers intended for the carriage of passengers and designed to seat no more than nine persons (including the driver). Refers to category M1 of the UN Consolidated Resolution on the Construction of Vehicles (R.E.3). It includes all types of fuel.
- Link to cleaning script: https://github.com/linero-tech/item-internship/blob/main/code/sandra/T022_TAS_VEP_017.ipynb

Source Data was compiled from the Asian Transport Outlook Database (ATO2023 TAS-VEP-017), accessible at <https://data.adb.org/dataset/asian-transport-outlook-database>. The original data was sourced from Country Official Statistics . The dataset was downloaded on November 6, 2023.

Country and ISO Code All countries have an assigned ISO code.

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Stock*

Unit The unit is changed from Number to 10^6 *vehicle*.

Service *Passenger*

Mode *Road*

Vehicle Type *LDV*

Technology *All*

Fuel *All*

Data Quality flag "!" (Caution): The data appears plausible but is lacking in cross-verification or has been suggested for manual review.

T023

Information

- Dataset name: Freight Transport - Tonne-km for Aviation (Domestic)
- Description: Domestic tonne-kilometre, is a unit of measure of freight transport which represents the transport of one tonne of goods (including packaging and tare weights of intermodal transport units) by a given transport mode (road, rail, air, sea, inland waterways, pipeline etc.) over a distance of one kilometre. Only the distance on the national territory of the reporting country is taken into account for national, international and transit transport.
- Link to cleaning script: [https://github.com/linero-tech/item-internship/blob/main/code/sandra/T023_TAS_FRA_007\(2\).ipynb](https://github.com/linero-tech/item-internship/blob/main/code/sandra/T023_TAS_FRA_007(2).ipynb)

Source Data was compiled from the Asian Transport Outlook Database (ATO2023 TAS-FRA-007(2)), accessible at <https://data.adb.org/dataset/asian-transport-outlook-database>. The original data was sourced from World Bank (WB) <https://data.worldbank.org/indicator/IS.AIR.GOOD.MT.K1->. The dataset was downloaded on November 6, 2023.

Country and ISO Code All countries have an assigned ISO code.

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable *Freight Activity*

Unit The unit is changed from *Million tonne kilometers* to 10^9 *tonne-km / yr*.

Service *Freight*

Mode The mode is changed from *Aviation* to *Aviation (Domestic)*.

Vehicle Type *All*

Technology *All*

Fuel *All*

Data Quality flag "!" (Caution): The data appears plausible but is lacking in cross-verification or has been suggested for manual review.

T024

Information

- Dataset name: Trends in global fuel economy of new vehicles: 2005 - 2022
- Description: Yearly vehicle sales are defined by segment (small car, medium car, large car, small SUV, large SUV, Light Commercial Vehicle) and powertrain (internal combustion engine, mild hybrid, hybrid, plug-in hybrid, battery electric, fuel cell hydrogen). For each combination of segment and powertrain the data contains information on sales, average weight (kg), average footprint (m²), and average specific energy consumption (lge/100 km). The data was obtained from a set of sources and processed to gain the best possible estimate of global trends in energy consumption for light duty vehicles.
- Link to cleaning script: ??

Source Cazzola, P., Paoli, L., & Teter, J. (2023). Trends in global fuel economy of new vehicles: 2005 - 2022 [Data set]. Zenodo (GFEI.2023-1.0).<https://doi.org/10.5281/zenodo.10148349>.

Country and ISO Code All countries have an assigned ISO code with the exception of the following:

- `CountryISO3 = "ROM"`, there was no country associated according to the `pycountry` module. Referring to the ISO 3166 Country Code Standard, “ROM” was not listed as a Country Code; instead, “ROU” was listed as the Country Code for Romania. Therefore, `CountryISO3` values containing “ROM” were replaced with “ROU”.

ITEM Region All countries in the dataset are assigned to an iTEM region.

Variable

- "registrations" = *Stock*
- "specific_energy_consumption_l_100km" = *Fuel Economy*

Unit

- *Stock*: 10⁶ vehicle
- *Fuel Economy*: l per 100km

Service *Passenger*

Mode *Road*

Vehicle Type *LDV, LDV - small car, LDV - medium car, LDV - large car, LDV - small suv, LDV - large suv, LDV - lcv, LDV - unclassified.*

Technology

- *All*: include the following powertrain: ev, hv, ice, phev, fcv, mhv, unclassified.
- *BEV*: include the following powertrain: ev.
- *Conventional*: include the following powertrain: ice, hv, mhv.
- *Fuel Cell*: include the following powertrain: fcv.
- *PHEV*: include the following powertrain: phev.

Fuel

- When *Technology* = *All*, *Fuel* = *All*.
- When *Technology* = *BEV*, *Fuel* = *Electricity*.
- When *Technology* = *Conventional*, *Fuel* = *Liquid*.
- When *Technology* = *Fuel Cell*, *Fuel* = *Hydrogen*.
- When *Technology* = *PHEV*, *Fuel* = *Liquid+Electricity*.

Data Cleaning

1. Before data cleaning, the total number of rows was 11,325. The number of rows containing “NaN” and/or empty strings totaled 1,114, and these were removed from the dataset. Thus, the remaining 10,211 rows were considered as the input dataset.
2. For “registration”, all numerical values are rounded to the nearest whole number (1). This ensures consistency and accuracy in the registration data.

Data Processing While importing, two different approaches were taken:

1. Importing data as it is. “segment“ are assigned to “Vehicle Type” in the following categories. “Source” is “GFEI 2023-1.0”:
 - *segment* = *large car* \Rightarrow *Vehicle Type* = *LDV - large car*.
 - *segment* = *large suv* \Rightarrow *Vehicle Type* = *LDV - large suv*.
 - *segment* = *lcu* \Rightarrow *Vehicle Type* = *LDV - lcu*.
 - *segment* = *medium car* \Rightarrow *Vehicle Type* = *LDV - medium car*.
 - *segment* = *small car* \Rightarrow *Vehicle Type* = *LDV - small car*.
 - *segment* = *unclassified* \Rightarrow *Vehicle Type* = *LDV - unclassified*.
2. Aggregating all *segment* by creating a new “Vehicle Type” = “LDV.” Change “Source” to “iTEM”, indicating that these values are based on our calculations. See more details in the next section.

Data Aggregation

Stock The variable *stock* for LDV represents the total number of vehicles within each country, summed by vehicle class and by technology class across all segments, and then normalized to millions of vehicles. The calculation of *stock* for LDV is formalized as:

$$stock = \frac{1}{10^6} \left(\sum_{\substack{\text{vehicle class} \\ \text{technology class}}} \text{registration}_{\text{large car}} \right. \\ + \text{registration}_{\text{large suv}} \\ + \text{registration}_{\text{lcu}} \\ + \text{registration}_{\text{medium car}} \\ + \text{registration}_{\text{small car}} \\ + \text{registration}_{\text{small suv}} \\ \left. + \text{registration}_{\text{unclassified}} \right) \quad (1)$$

Where the sum accumulates the total registrations for each combination of vehicle and technology class across all segments. The final stock value is expressed in units of millions of vehicles.

Fuel Economy The variable *Fuel Economy* for LDV quantifies the average energy efficiency of vehicles within a country, differentiated by *Technology* type ("powertrain"). It is calculated as a weighted sum of the variable `specific_energy_consumption_l_100km`, with the weighting given by the `registration` variable for each segment. The calculation is formalized as:

$$\text{fuel economy}_{\text{technology type}} = \frac{\sum_{i=1}^n (\text{specific_energy_consumption_l_100km}_i \times \text{registration}_i)}{\sum_{i=1}^n \text{registration}_i} \quad (2)$$

Where:

- `specific_energy_consumption_l_100kmi` is the specific energy consumption for the *i*-th segment, measured in liters per 100 kilometers,
- `registrationi` is the number of registered vehicles for the *i*-th segment,
- The summation runs over all segments within the dataset for a given technology type.

This results in `fuel_economy` for LDV being expressed as the average liters per 100 kilometers (l/100km) for each technology type within a country, providing an indicator of the overall fuel efficiency.