

Supplementary Documentation: Data Acquisition and Cleaning

Data Acquisition:

Railway station and track geospatial data were acquired using the Python `osmnx` package, querying OpenStreetMap (OSM) for relevant geographic data on railway infrastructure.

The electrification status data for the railway tracks was retrieved from OpenRailwayMap (<https://openrailwaymap.org/>), providing detailed information about electrified segments of the network.

Data Cleaning and Validation:

The initial dataset of railway stations obtained from OSM was cross-verified against the latest official list of railway stations available on the Uzbekistan Railways (Uzrailways) website. Discrepancies and inconsistencies were identified and corrected using Python and R scripts, with some manual verification performed in Excel.

Railway track data sourced from OSM were validated against the most up-to-date static map provided officially. Discrepancies in the railway track data were addressed using Python scripts for initial automation and further refined in QGIS software.

Regional Division:

OSM data lacked administrative regional boundaries, requiring manual division of railway tracks. This was resolved by extracting and parsing railway track segments and programmatically labeling them according to the official regional information derived from the official station list using Python.

Creation of Shapefiles:

Shapefiles incorporating corrected railway tracks, regional divisions, and electrification status were created and edited within the QGIS environment, culminating in detailed spatial representations used in the project's main visualization.

Final Cleaning Steps:

Final adjustments and validation steps were documented and executed through Jupyter notebooks:

AUX_electrification.ipynb: Addressed and finalized the electrification data integration.

AUX_mtu_correction.ipynb: Managed corrections to Maximum Train Unit (MTU) data.

Data Storage:

All cleaned datasets, shapefiles, and documented outcomes are systematically organized and stored within the project's directory structure, specifically:

3.qgis project: Contains core shapefiles used for map visualization.

4.other_shapefiles: Stores supplementary shapefiles, including those detailing electrification statuses.

Tools used:

Python packages used include osmnx, geopandas, pandas, and shapely.

R packages utilized primarily include sf for spatial data manipulation.

QGIS (Quantum Geographic Information System) is an open-source geographic information system software for editing, analyzing, and visualizing geospatial data.