Project Group 3

Members: Youri Beijer Ecem Tyurkay Rico de Jong Saumitra Deo Pranshu Sharma

Student numbers: 4965027 6301207 5348870 6434223 6505015

Research Objective

Requires data modeling and quantitative research in Transport, Infrastructure & Logistics

The aim of this report is to answer the following research question:

"How is rail freight distributed across European countries by consignment type and commodity group, and how does the railway network influence these patterns/trends?"

Contribution Statement

Be specific. Some of the tasks can be coding (expect everyone to do this), background research, conceptualisation, visualisation, data analysis, data modelling

Each author is responsible for answering their own sub-question.

Author 1: Saumitra Deo

"How do different countries relate to different shares of consignment types?"

Tasks: Extract relevant Eurostat dataset (rail_go_consgmt), Clean and preprocess country–consignment data, Build contingency tables (country vs. consignment type), Perform chi-square test of independence, Create visualizations: bar charts, heatmaps of shares by country, Write results section discussing differences across countries.

Author 2: Youri Beijer

"How does network length influence consignment?"

Tasks: Retrieve Eurostat dataset (rail_if_line_na) for network length, Merge network length data with consignment dataset (rail_go_consgmt), Perform correlation analysis (Pearson/Spearman) between network length and consignment types, Run regression models to test influence of network length, Create scatter plots with regression lines, Document methodology and statistical findings.

Author 3: Rico de Jong

"How does the network length influence the modal share?"

Tasks: Collect data on modal share of rail freight from Eurostat, Link modal share with network length & density indicators, Conduct regression analysis (network length & density vs. modal share), Create scatter plots, regression plots, and comparative tables, Interpret whether denser networks show higher modal shares, Draft results and add visual evidence to the report.

Author 4: Pranshu Sharma

"Which 5 EU countries have the most- and least train freight?"

Tasks: Extract rail freight volume dataset from Eurostat, Rank countries and identify top 5 and bottom 5, Compare structural differences (total volume, growth trends), Produce tables, rankings, and bar plots of volumes, Write results section interpreting patterns among high vs. low performers.

Author 5: Ecem Tyurkay

"How has the trend of full train consignments changed over time in the top- and bottom-five EU freight countries?"

Tasks: Filter dataset for full train consignments, Perform time-series analysis (2008–present), Plot trends for top 5 and bottom 5 countries (line charts, percentage shares), Test hypotheses using regression or trend analysis.

"How do commodity groups (NST 2007 classification) influence the distribution of consignment types?"

Tasks: Retrieve dataset (rail_go_grpgood), Combine with consignment dataset, Build contingency tables (commodity group × consignment type), Run chi-square independence tests, Visualize with stacked bar charts and heatmaps, Interpret which commodity groups are linked to certain consignment types.

Data Used

rail_go_consgmt – Goods transported by type of consignment (Eurostat).

rail_go_grpgood – Goods transported by group of goods (Eurostat, NST 2007 classification, from 2008 onwards).

rail_if_line_na – Length of electric and non-electric railway lines, by nature of transport (Eurostat).

Rail freight volumes (tonnes and tonne-kilometres) – Eurostat indicators for total freight transported.

Rail modal share indicators – Eurostat statistics on the modal split of freight transport.

Data Pipeline

Data Collection

Download datasets from Eurostat (rail_go_consgmt, rail_go_grpgood, rail_if_line_na, freight volumes, modal share).

Ensure consistent formats (CSV/Excel).

Data Cleaning & Pre-processing

Remove missing or inconsistent entries.

Standardize country names, years, units (tonnes, tonne-km, percentages).

Filter to the required timeframe (2008 onwards).

Merge datasets where needed (e.g., network length with consignment shares).

Data Storage & Versioning

Store cleaned datasets in your GitHub repository.

Each member works on a separate branch, commits changes, and documents preprocessing steps.

Analysis & Testing

Perform statistical tests (chi-square, correlation, regression).

Generate descriptive statistics (means, shares, rankings).

Run time-series analysis for trends.

Visualization & Reporting

Create charts (scatter plots, bar charts, line graphs, heatmaps).

Export results for inclusion in the final report.

Each member writes their analysis section, integrated into the final document.