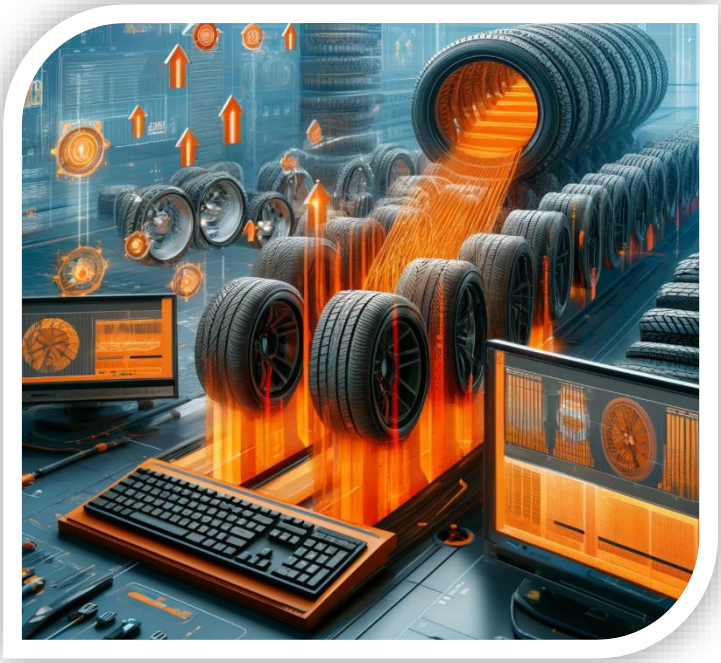
Normalization for Tire Data from Multiple Suppliers and Matching Techniques



ABSTRACT

This research focuses on improving the way tire data from various suppliers, is organized and matched. I aim to address the challenge of inconsistent data formats across suppliers. The paper explains how I standardized **Suppliers tires** data and matched it with **the Bac dataset**, which includes detailed tire information. I evaluate my methods based on how accurately they match the data and handle missing information. This study helps make tire data more usable for truck companies, enhancing their decision-making processes in the automotive industry.

INTRODUCTION

Accurate tire data is crucial for truck companies to manage tire purchases and maintenance effectively. However, the challenge lies in the diverse data formats provided by different tire suppliers, like Kosson, which hinder efficient data integration and usage. This research project specifically addresses the normalization and matching of tire data from Kosson and other suppliers to a standardized database, the Bac dataset. By standardizing and aligning Kosson's tire data with established parameters such as width, aspect ratio, rim diameter, manufacturer code, and model, this study aims to improve data usability and decision-making in the trucking industry. The subsequent sections will detail the methodologies employed, the effectiveness of these methods, and their practical implications.

DATA ANALYSIS

To effectively normalize and match tire data, a thorough understanding of the data provided by suppliers is crucial. The data from Kosson, one of the primary suppliers for this study, exhibits several key properties that need to be analysed for effective integration.

Merk	Omschrijving	Afmetingen		Prijs	Voorraad
Binnenbanden	BI 825R20 BINNENBAND (TUBE) 000	825R20	€	19,00	3
	BM 205/75R17.5 BF14 124/122M TL DOT2019 EC071	205/75R175	€	149,00	12
Bridgestone	BR 11R22.5 R168 DOT21 TL 148L 145L	11R225	€	259,00	24
Bridgestone	BR 12R22.5 M840 DOT21 TL 152K 150L 16PR	12R225	€	279,00	20
Bridgestone	BR 8.25R15 R187 143J141J TT 18PR (NO SETS) DCB71	825R15	€	265,00	12
Bridgestone	BR 205/65 R17.5 RT1 M+S 3PMSF 132/130J(133F) CCA69	205/65R175	€	214,00	24
Bridgestone	BR 205/75 R17.5 RD2 124/122M M+S 3PMSF CBA71	205/75R175	€	204,00	2
Bridgestone	BR 205/75 R17.5 RS2 124/122M M+S 3PMSF CBA69	205/75R175	€	184,00	20
Bridgestone	BR 225/75 R17.5 RD2 129M127M M+S 3PMSF CBB74	225/75R175	€	196,00	30

Key Properties of the Data:

- **Width:** The width of the tire, measured in millimeters, is a critical dimension that affects the tire's fit and performance.
- **Aspect Ratio:** This is the ratio of the height of the tire's cross-section to its width. Changes in the aspect ratio can impact the vehicle's handling characteristics.
- **Rim Diameter:** The diameter of the rim on which the tire can be mounted, measured in inches. This needs to match with the vehicle's wheels for proper installation.
- **Manufacturer Code:** Each tire has a unique code assigned by the manufacturer, which helps in identifying the tire and tracking its specifications.
- **Model:** The model's name or number provided by the manufacturer, which can be used to ascertain detailed characteristics and quality standards.

Analysis Approach:

To begin with, I will assess the consistency and completeness of data across these properties. This involves checking for missing values, outliers, or inconsistent entries that may affect the subsequent stages of data normalization and matching. The goal is to identify common patterns or discrepancies that could influence the accuracy of the dataset when integrated with the Bac dataset.

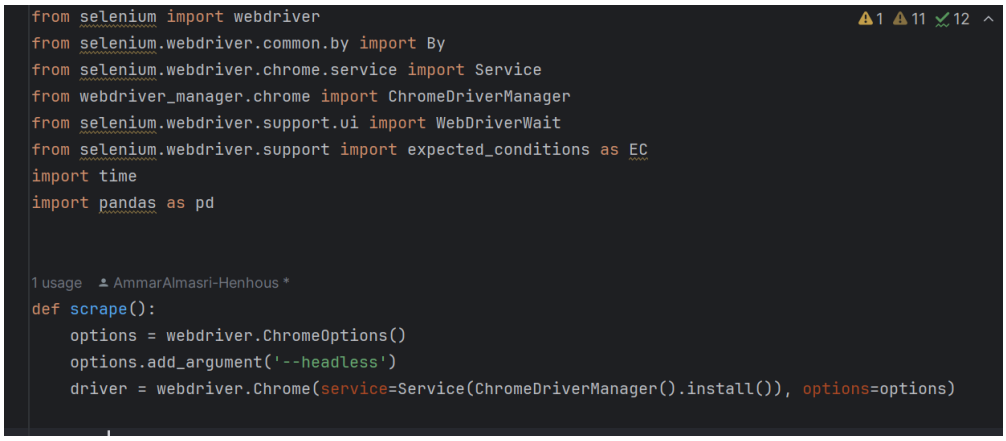
This preliminary analysis sets the stage for developing tailored normalization techniques that address the specific issues identified in Kosson's tire data, ensuring that the integration process is as seamless and efficient as possible.

DATA PREPARATION

The data preparation stage is crucial for ensuring that tire data is clean, consistent, and ready for analysis and integration. This process involves several automated steps executed through Python scripts, Flask services, and n8n.

- **Python Scripts**

Python scripts perform the essential tasks of extracting, cleaning, and standardizing the tire data from Kosson. These scripts automatically handle discrepancies in data formats and ensure that all entries meet the required quality standards.

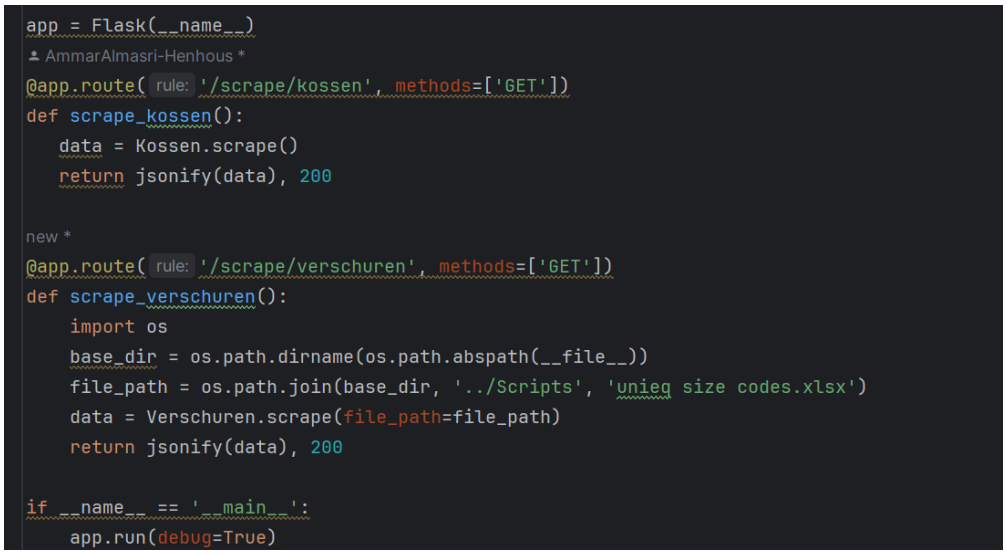


```python
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.chrome.service import Service
from webdriver_manager.chrome import ChromeDriverManager
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
import time
import pandas as pd


1 usage  AmmarAlmasri-Henhous *
def scrape():
    options = webdriver.ChromeOptions()
    options.add_argument('--headless')
    driver = webdriver.Chrome(service=Service(ChromeDriverManager().install()), options=options)
```

- **Flask Services**

Flask, a web framework, is used to host these Python scripts as accessible services. This setup allows the scripts to be triggered remotely, ensuring flexibility and integration with other systems in our data workflow.



```python
app = Flask(__name__)
AmmarAlmasri-Henhous *
@app.route( rule: '/scrape/kossen', methods=['GET'])
def scrape_kossen():
    data = Kossen.scrape()
    return jsonify(data), 200

new *
@app.route( rule: '/scrape/verschuren', methods=['GET'])
def scrape_verschuren():
    import os
    base_dir = os.path.dirname(os.path.abspath(__file__))
    file_path = os.path.join(base_dir, '../Scripts', 'unieq size codes.xlsx')
    data = Verschuren.scrape(file_path=file_path)
    return jsonify(data), 200

if __name__ == '__main__':
    app.run(debug=True)
```

- **n8n Automation**

n8n coordinates the flow of data from Flask to the final normalization and matching stages. It automates the process, ensuring that data moves seamlessly through the pipeline and that each step is executed accurately.

DATA NORMALIZATION:

Normalization is the first critical step to ensure uniformity across the collected tire data from Kosson. This process involves:

- **Extraction of Properties:** Using n8n, I automate the extraction of all relevant tire properties such as width, aspect ratio, rim diameter, manufacturer code, and model. This ensures that every piece of essential data is captured and formatted correctly.
- **Standardization:** Each property is standardized to conform to the Bac dataset's specifications. This includes converting measurements to a uniform system (e.g., all dimensions to millimeters), ensuring consistent formatting across data entries, and aligning categorical data like manufacturer names to predefined categories.

Before:

0	1	2	3	4
Binnenbanden	BI 825R20 BINNENBAND (TUBE) 000	825R20	€\n19,00	3
[empty]	BM 205/75R17.5 BF14 124/122M TL DOT2019 EC071	205/75R175	€\n149,00	12
Bridgestone	BR 11R22.5 R168 DOT21 TL 148L 145L	11R225	€\n259,00	24
Bridgestone	BR 12R22.5 M840 DOT21 TL 152K 150L 16PR	12R225	€\n279,00	20
Bridgestone	BR 8.25R15 R187 143J141J TT 18PR (NO SETS) DCB71	825R15	€\n265,00	12
Bridgestone	BR 205/65 R17.5 RT1 M+S 3PMSF 132/130J(133F) CCA69	205/65R175	€\n214,00	24
Bridgestone	BR 205/75 R17.5 RD2 124/122M M+S 3PMSF CBA71	205/75R175	€\n204,00	2
Bridgestone	BR 205/75 R17.5 RS2 124/122M M+S 3PMSF CBA69	205/75R175	€\n184,00	20
Bridgestone	BR 225/75 R17.5 RD2 129M127M M+S 3PMSF CBB74	225/75R175	€\n196,00	30
Bridgestone	BR 225/75 R17.5 RS2 129/127M M+S 3PMSF CBA70	225/75R175	€\n209,00	2

After:

Manufacturer	DefinitionKossen	size	Voorraad	Width	RimDiameter	AspectRatio	price	ManufacturerCode	extractedMerk
Binnenbanden	BI 825R20 BINNENBAND (TUBE) 000	825R20	3	825	20	0	€19,00	BI	BINNENBAND
[empty]	BM 205/75R17.5 BF14 124/122M TL DOT2019 EC071	205/75R175	12	205	175	75	€149,00	BM	BF14
Bridgestone	BR 11R22.5 R168 DOT21 TL 148L 145L	11R225	24	11	225	0	€259,00	BR	R168
Bridgestone	BR 12R22.5 M840 DOT21 TL 152K 150L 16PR	12R225	20	12	225	0	€279,00	BR	M840
Bridgestone	BR 8.25R15 R187 143J141J TT 18PR (NO SETS) DCB71	825R15	12	825	15	0	€265,00	BR	R187
Bridgestone	BR 205/65 R17.5 RT1 M+S 3PMSF 132/130J(133F) CCA69	205/65R175	24	205	175	65	€214,00	BR	RT1
Bridgestone	BR 205/75 R17.5 RD2 124/122M M+S 3PMSF CBA71	205/75R175	2	205	175	75	€204,00	BR	RD2
Bridgestone	BR 205/75 R17.5 RS2 124/122M M+S 3PMSF CBA69	205/75R175	20	205	175	75	€184,00	BR	RS2
Bridgestone	BR 225/75 R17.5 RD2 129M127M M+S 3PMSF CBB74	225/75R175	30	225	175	75	€196,00	BR	RD2
Bridgestone	BR 225/75 R17.5 RS2 129/127M M+S 3PMSF CBA70	225/75R175	2	225	175	75	€209,00	BR	RS2

DATA MATCHING:

After normalization, the next step is to match the standardized data with the Bac dataset. This is done through:

Attribute Comparison: n8n workflows compare each tire's properties from Kosson with those in the Bac dataset to find exact or close matches.

Validation: Each potential match is validated to ensure accuracy, with discrepancies or mismatches flagged for manual review or reprocessing.

Manufacturer	Definition Kossen	Definition BAC	ID BAC
Bridgestone	BR 12R22.5 M840 DOT21 TL 152K 150L 16PR	12 R22.5 M840 152K150L 16 TL	309675
Bridgestone	BR 8.25R15 R187 143J141J TT 18PR (NO SETS) DCB71	8.25 R15 R187 143/141J 18PR TT SET 14	339459
Bridgestone	BR 205/65 R17.5 RT1 M+S 3PMSF 132/130J(133F) CCA69	205/65 R17.5 RT1 132/130J (133F) TL	339488
Bridgestone	BR 205/75 R17.5 RS2 124/122M M+S 3PMSF CBA69	205/75 R17.5 RS2 124M122M 12 TL	284477
Bridgestone	BR 205/75 R17.5 RD2 124/122M M+S 3PMSF CBA71	205/75 R17.5 RD2 124M122M 12 TL	284478
Bridgestone	BR 225/75 R17.5 RD2 129M127M M+S 3PMSF CBB74	225/75 R17.5 RD2 129M127M 14 TL	267078
Bridgestone	BR 225/75 R17.5 RS2 129/127M M+S 3PMSF CBA70	225/75 R17.5 RS2 Z 129M127M 14 TL	267081
Bridgestone	BR 235/75 R17.5 RS2 132/130M J+S 3PMSF CBA70	235/75 R17.5 RS2 Z 132M130M TL	267091
Bridgestone	BR 235/75 R17.5 RD2 132/130M M+S 3PMSF CBB74	235/75 R17.5 RD2 132M130M TL	267092

DATABASE INTEGRATION:

The final step is the integration of the matched data into a MySQL database, which allows for:

- **Data Storage:** All matched and validated tire data is stored in structured tables within the MySQL database, making it easily accessible for queries and analysis.
- **Data Accessibility:** The integration provides a centralized location for accessing the consolidated tire data, enhancing data-driven decision-making processes within the company.

RESULTS

The primary goal of this project was to integrate tire data from Supplier into the Bac dataset effectively, enhancing data usability for truck companies. **Out of 748 tire entries from Kosson**, **495 were successfully matched to the Bac dataset**, yielding **a match rate of approximately 66%.**

MATCHING SUCCESS:

- **Successful Matches:** The **495 successfully** matched tires reflect the robustness of the normalization and matching techniques implemented. These tires were accurately aligned with the Bac dataset based on key properties such as width, aspect ratio, rim diameter, manufacturer code, and model.

- **High-Quality Data Integration:** The matched entries have been integrated into a MySQL database, ensuring that high-quality, consistent tire data is available for operational use by truck companies.

CHALLENGES IN MATCHING:

- **Manufacturer Code Issues: 26 tires could not be matched** due to the absence of their manufacturer codes in the Bac dataset. These included tires from lesser-known manufacturers like **Wielen, Taurus, and Samson,** suggesting a gap in the Bac dataset's coverage.

- **Model Discrepancies:** The majority of **unmatched tires (227 entries)** were due to missing or unrecognizable model information. This discrepancy likely stemmed from variations in model naming, or new models not yet included in the Bac dataset.

Recommendations for Improvement:

- **Dataset Expansion:** Regularly updating the Bac dataset to include new manufacturers and models will likely increase future match rates.

- **Enhanced Data Cleaning:** Improving data cleaning techniques to ensure more accurate extraction and standardization of tire model information.

- **Uniform Naming Conventions:** Establishing clearer naming conventions could help mitigate mismatches caused by different model naming practices among suppliers.

CONCLUSION:

The project demonstrates significant success in matching supplier tire data with the Bac dataset, with two-thirds of the tire entries successfully integrated into the operational database. The challenges identified provide a clear direction for further enhancing the data integration tool to support more comprehensive decision-making for truck companies.