

Document: Reproducing Code for Traffic Volume Estimation

Introduction

This document provides a structured guide to reproduce the code from three Jupyter notebooks related to a Traffic Volume Estimation project. These notebooks focus on data gathering, preprocessing, and model development for estimating traffic volumes. To facilitate ease of use, the CSV file data produced from Daria's work is also included, allowing users to directly use this data without re-running Daria's code or the "Data Preprocessing.ipynb" file repeatedly.

Overview of Notebooks

1. **Additional Data:** Sets up the environment and installs necessary packages.
2. **Data Preprocessing:** Handles loading, cleaning, and transforming the data.
3. **Traffic Volume Estimation Model:** Concentrates on building and evaluating the traffic volume estimation model.

Step-by-Step Guide

Step 1: Run "Additional Data.ipynb"

1. **Required Software:** Ensure Python is installed on your system.
2. **Install Necessary Packages:**
 - Execute `!pip install --upgrade holidays` to install the 'holidays' package.
 - Additional packages, such as requests, may be required, as indicated in the first notebook.

Step 2: Run "Data Preprocessing.ipynb"

- **Important:** Load the data from "Merged flows DB ALL HEADERS.csv", which is an output from Daria's work. This file is included for convenience and to streamline the process.
- **Preprocessing Steps:** Follow the code to preprocess the data. This includes converting date columns to DateTime objects and other necessary data transformations.

- **Output:** At the end of the script, the main data used for the traffic estimation model will be saved as "Main Data.csv".

Step 3: Run "Traffic Volume Estimation Model.ipynb"

- **Load Data:** Begin by loading the "Main Data.csv" created in the previous step.
- **Model Development:**
 - Follow the notebook to build the traffic volume estimation model. This includes data cleaning specific to the model and handling missing values.
 - Proceed with training and evaluating the model as per the instructions in the notebook.

Notes and Recommendations

- **Data Consistency:** Ensure the datasets used are consistent with the specified files for accurate result reproduction.
- **Sequential Execution:** Execute the code cells in the order they appear, as they often have dependencies.
- **Troubleshooting:** If errors arise, check for common issues like missing packages, incorrect file paths, or data format issues.

Conclusion

This guide is designed to assist you in accurately reproducing the steps for the Traffic Volume Estimation project. It's important to follow the instructions in sequence and understand each step's purpose for successful implementation. The inclusion of the preprocessed CSV file streamlines the process, allowing you to focus on the traffic volume estimation model development. If you encounter any specific challenges or have questions about certain parts of the notebooks, feel free to ask for more detailed explanations or assistance.