Causal_impact

Project Overview

The Causal Impact model uses a Bayesian structural time-series framework to estimate the hypothetical outcome of a time series in the absence of an intervention, allowing it to measure the causal effect of the intervention by comparing observed data to this hypothetical scenario.

Data Sources

The model runs on preprocessed data which is aggregated from the following data files:

- sector2021-2024.xlsx

- hdm nor daily 2023.csv

- audit_new_database_merged_all_2024_oct_update.xlsx

These files should be placed in the appropriate directories as referenced in the config.json.

Configuration

The config.json file contains the paths to the input data and configurable parameters for the pipeline. Key sections include:

- Data: Specifies paths to input data files and output paths.

- Dates: Dates for treatment and observation periods.

- Filtering: Filtering criteria for TINs and LICENSE_NUMBERs.

Installation

Follow these steps to set up the project on your local machine:

1. Clone the repository:

```
git clone https://github.com/your-repo/Causal_impact.git
```

2. Ensure Python 3.8.10 is installed:

This project requires Python 3.8.10 You can download and install it from the official Python website: https://www.python.org/downloads/.

3. Create a virtual environment

After installing Python 3.8, create a virtual environment for the project:

For Windows:

```
python -m venv causal-env
causal-env/Scripts/activate
```

For macOS/Linux:

```
python3.8 -m venv causal-env
source causal-env/bin/activate
```

4. Install dependencies

After activating the virtual environment, install the required dependencies:

```
pip install -r requirements.txt
```

5. Run the main script

To run the main script, execute the following command:

```
python main.py
```

This will execute the model codes according to the settings defined in config.json.

Output

The model scripts will generate the following outputs:

1. Charts for each individual TIN

- Observed vs Predicted Behavior Chart: This chart will compare the observed sales to the predicted counterfactual scenario.

- Difference Chart: This chart will depict the differences between observed and predicted sales behaviors.

2. Excel Report

The Excel file will contain detailed data on the daily and cumulative monetary impact of audits for each TIN, presented in AMD, along with p-values to indicate the statistical significance of the impact.