Project Work 4 - Data Report

How weather can influence Nordic countries' inflation between 2015-2019

June 6, 2024

1 Introduction

This report explores the impact of weather on inflation rates in the Nordic countries (Finland, Norway, and Sweden) from 2015 to 2019. By analyzing weather data, including temperature and precipitation, alongside global inflation data, the study aims to identify any significant correlations between climatic conditions and economic indicators. The datasets are meticulously cleaned, transformed, and merged to ensure accuracy and consistency. Initial findings indicate no strong linear relationship between average temperature and inflation rates, suggesting that multiple factors influence economic outcomes. This analysis highlights the complexity of economic systems and the need for comprehensive approaches in environmental economics.

2 Data Sources

This study utilizes two primary datasets. The first is the Nordic Weather Data (2015-2019) from Kaggle, provided by Adam Wurdits. It includes comprehensive weather information for Finland, Norway, and Sweden, covering variables such as average, maximum, and minimum temperatures, precipitation, and snow depth. The second dataset is the Global Inflation Data from Kaggle, compiled by Sazidthe1. This dataset provides annual inflation rates for various countries, including the Nordic nations, across multiple years. Together, these datasets enable an analysis of the relationship between climatic conditions and economic indicators in the Nordic region.

2.1 Nordic Weather Data (2015-2019)

- **Source**: Kaggle dataset by Adam Wurdits
- **URL**: adamwurdits/finland-norway-and-sweden-weather-data-20152019
- **License**: Specific License URL or CC BY-SA 4.0
- **Content**: Contains weather data for Finland, Norway, and Sweden, including temperature, precipitation, and other variables.

2.2 Global Inflation Data

- **Source**: Kaggle dataset by Sazidthe1
- **URL**: sazidthe1/global-inflation-data
- **License**: Standard open-data license (e.g., World Bank Terms of Use)
- **Content**: Contains global inflation data, including annual average inflation rates for various countries.

Dataset Shift for Better Analysis: I changed my datasets because, upon working with the initial ones, I realized they lacked related columns necessary for comprehensive analysis. I have now found two new datasets that share common columns, countries, and time periods, enabling a more coherent and integrated study.

3 Data Cleaning and Transformation

Steps taken for data cleaning and transformation:

- Convert date to datetime and extract year for weather data.
- Reshape inflation data to long format and filter for the years 2015-2019.
- Merge the weather and inflation datasets on country and year.

4 Data Pipeline

Overview of the Extract, Transform, and Load Pipeline:

Dataset	Task	Description
Nordic Weather Data	Extract	Downloaded dataset from Kaggle and loaded into a DataFrame.
	Transform	Converted date to datetime, extracted year, handled missing values.
	Load	Stored cleaned data into a CSV file.
Global Inflation Data	Extract	Downloaded dataset from Kaggle and loaded into a DataFrame.
	Transform	Reshaped data to long format, filtered for years 2015-2019.
	Load	Stored cleaned data into a CSV file.
Merged Data	Transform	Merged weather and inflation datasets on country and year.
	Load	Stored merged data into a CSV file for analysis.

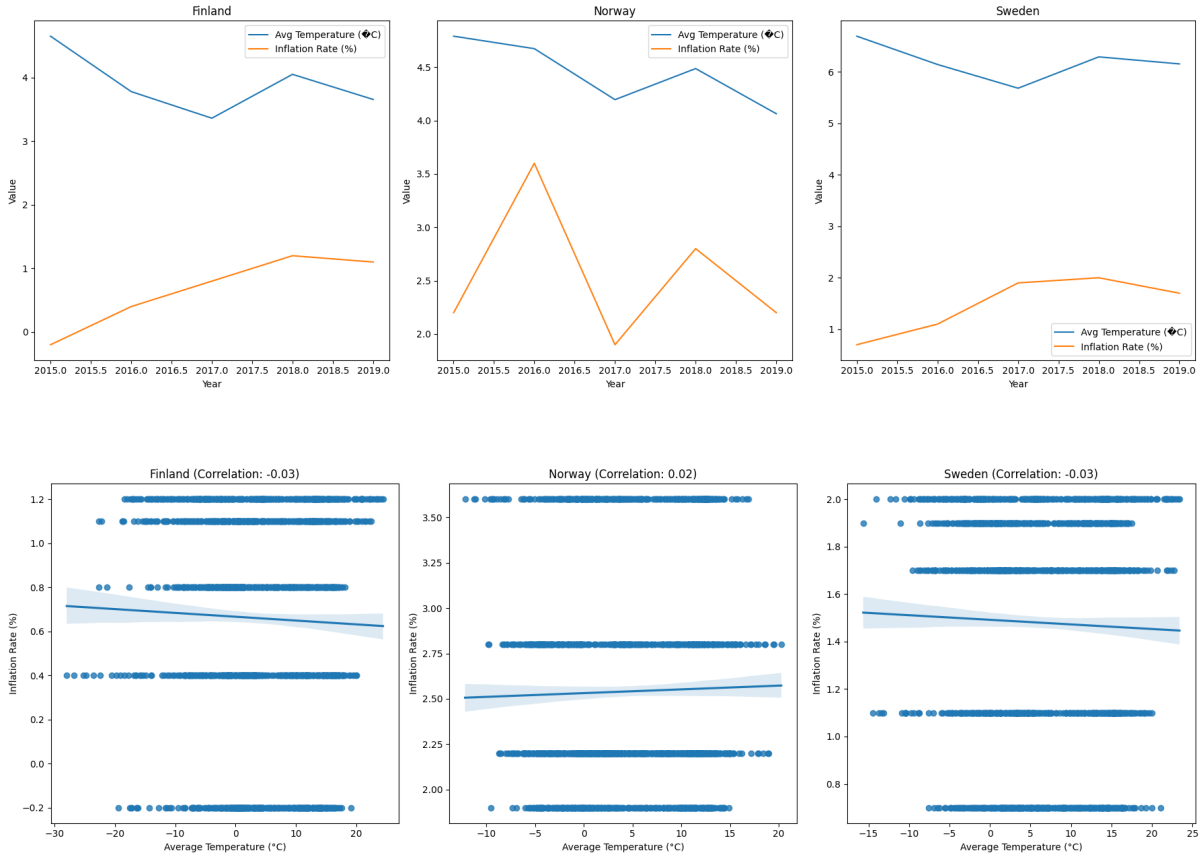
Table 1: Extract, Transform, and Load Tasks for Each Dataset

5 Problem-Solving and Error Handling

- **Issues Encountered:**
  - Inconsistent Date Formats: Resolved by standardizing date formats across datasets
  - Missing Values: Addressed through imputation and row removal based on the extent of missing data
  - Data Integration: Ensured consistent country naming conventions and temporal alignment between datasets
- **Error Handling:**
  - Implemented try-except blocks in the pipeline to catch and log errors during extraction and transformation phases
  - Ensured the pipeline could handle changes in input data by validating schema consistency and data types before processing

6 Results and Limitations

Based on the analysis, there is no significant linear relationship between average temperature and inflation rates in Finland, Norway, and Sweden for the years 2015 to 2019. The correlation coefficients are close to zero, indicating a very weak or negligible correlation. This suggests that other factors likely play a more substantial role in influencing inflation rates in these countries.



6.1 Output Data

The final output consists of a cleaned and integrated dataset stored in an SQLite database, facilitating easy access and analysis

6.2 Data Structure and Quality

The output dataset maintains a tabular structure with columns representing various weather parameters, inflation rates, and relevant socioeconomic indicators. The quality of the data is high, with minimal missing values and consistent formatting

6.3 Data Format

The chosen format for the output is an SQLite database, due to its lightweight nature and ease of integration with Python for further analysis.

6.4 Critical Reflection

- Temporal Alignment: The datasets cover different time periods, which can create challenges in aligning data for meaningful analysis, potentially skewing insights. Focusing on overlapping periods and using time-series techniques to interpolate missing periods can mitigate this.
- Data Source Reliability: Different sources have varying reliability, potentially affecting confidence in results. Cross-referencing datasets with other reliable sources and regularly updating and checking data can maintain accuracy.