

# Dataset Access and Reproduction Instructions

This document outlines the sources and preparation steps for the datasets used in our **Visualizing Airfare Trends** application. It ensures that anyone reviewing or rerunning our system can access the appropriate data, either by downloading public datasets or by reproducing our processed files.

## 1. U.S. Airline Fare Dataset

### Download Location:

<https://www.kaggle.com/datasets/bhavikjikadara/us-airline-flight-routes-and-fares-1993-2024?source=download>

### File to Save As:

`airline_data.csv`

### Directory:

`/visualizing-airfare-trends-backend`

### Instructions:

- You must have a free Kaggle account to download this dataset.
- Once downloaded, place the CSV file in the backend folder of the project and rename it as `airline_data.csv`.

## 2. Macroeconomic Dataset

We use a combined macroeconomic dataset derived from three publicly available sources. These metrics are used to display relevant economic context during fare predictions.

### Sources and Download Links

Metric	Frequency	Source URL	Save As
--------	-----------	------------	---------

Unemployment Rate	Monthly	<a href="https://fred.stlouisfed.org/series/UNRATE">https://fred.stlouisfed.org/series/UNRATE</a>	<b>UNRATE.csv</b>
Global Oil Price	Quarterly	<a href="https://fred.stlouisfed.org/series/POILBREUSDQ">https://fred.stlouisfed.org/series/POILBREUSDQ</a>	<b>POILBREUSDQ.csv</b>
Gross Domestic Product	Quarterly	<a href="https://fred.stlouisfed.org/series/GDP">https://fred.stlouisfed.org/series/GDP</a>	<b>GDP.csv</b>

## Reproduction Instructions

If you'd like to reproduce the combined macroeconomic dataset:

- Download the three datasets linked above into `/visualizing-airfare-trends-backend/data`. They should be named as indicated.
- Use the provided Python script (see `/visualizing-airfare-trends-backend/scripts/combine_macro_data.py`) to join and align the datasets by date.
- The script outputs `macro_data.csv`, which should be placed in the `/visualizing-airfare-trends-backend` folder.

If needed, update the script paths or column names depending on the file format downloaded from FRED.

## Direct Access to Combined Dataset (Optional)

If you would prefer to use our pre-combined macroeconomic dataset, you may download it directly from:

<https://drive.google.com/file/d/1CS0JBsn1fIMVvbt1ppQFOwAGwAwM5E6L/view?usp=sharing>

Save the file as `macro_data.csv` and place it in `/visualizing-airfare-trends-backend`.